

# Editors

Paul Worsfold, *University of Plymouth, Plymouth, UK*

Alan Townshend, *University of Hull, Hull, UK*

Colin Poole, *Wayne State University, Detroit, MI, USA*

# Editorial Advisory Board

Freddy Adams, *University of Antwerp, Antwerp, Belgium*

Jonathan W Aylott, *University of Hull, Hull, UK*

Damia Barcelo, *Institut d' Investigacions Químiques i Ambientals de Barcelona, Consejo Superior de Investigaciones Científicas, Barcelona, Spain*

Alan M Bond, *Monash University, Melbourne, VIC, Australia*

Chris Burgess, *Burgess Consultancy Co., Durham, UK*

Marcela Burguera, *Universidad de los Andes, Merida, Venezuela*

John Chalmers, *VS Consulting, Stokesley, UK*

Richard Dams, *Instituut voor Nucleaire Wetenschappen (INW), Gent, Belgium*

Cees Gooijer, *Free University, Amsterdam, The Netherlands*

John D Green, *Beverley, UK*

Elo H Hansen, *Technical University of Denmark, Lyngby, Denmark*

Steve J Hill, *University of Plymouth, Plymouth, UK*

Jerry W King, *Los Alamos National Laboratory, Los Alamos, NM, USA*

Larry J Kricka, *Medicine Centre University of Pennsylvania, Philadelphia, PA, USA*

Maria D Luque de Castro, *Universidad de Córdoba, Córdoba, Spain*

Pierre Margot, *Université de Lausanne, Lausanne, Switzerland*

Philip J Marriott, *RMIT University, Melbourne, VIC, Australia*

Ian D McKelvie, *Monash University, Melbourne, VIC, Australia*

Veronika R Meyer, *EMPA St Gallen, St Gallen, Switzerland*

Heinrich H D Meyer, *Institut für Physiology, München, Germany*

Jim N Miller, *Loughborough University of Technology, Loughborough, UK*

Pavel N Nesterenko, *Lomonosov Moscow State University, Moscow, Russia*

Reinhard Niessner, *Technische Universität München, München, Germany*

Matthias Otto, *Institute for Analytical Chemistry, Freiberg, Germany*



Pier G Righetti, *University of Verona, Verona, Italy*

Alfredo Sanz-Medel, *University of Oviedo, Oviedo, Spain*

Malcolm R Smyth, *Dublin City University, Dublin, Republic of Ireland*

Robert E Synovec, *University of Washington, Seattle, WA, USA*

Peter C Uden, *University of Massachusetts at Amherst, Amherst, MA, USA*

Miguel Valcárcel, *Universidad de Córdoba, Córdoba, Spain*

Wolfhard Wegscheider, *University of Leoben, Leoben, Austria*

Ian D Wilson, *Astra Zeneca Pharmaceuticals, Macclesfield, UK*

Otto S Wolfbeis, *University of Regensburg, Regensburg, Germany*

Elias Zagatto, *Universidade de São Paulo, Piracicaba, Brazil*

Yury Zolotov, *USSR Academy of Sciences, Moscow, Russia*

# Preface

Analytical science impacts on all aspects of life in the twenty-first century. Reliable, high-quality analytical data are essential prerequisites for monitoring health (and disease), for enhancing the efficiency of industrial processes, improving product quality and reducing emissions, and for studying complex biogeochemical interactions in the environment. New analytical techniques and methods are key drivers for advances in drug discovery, forensic science, and life sciences, for monitoring the quality of foodstuffs, pharmaceuticals, and other consumer products, for furthering our understanding of environmental processes, and for monitoring compliance with legislation.

The means by which analyses are achieved varies from simple color tests for the qualitative identification of anions and cations through to complex and expensive computer-controlled instrumentation for quantitative determination of trace amounts of a single organic compound or element in a complex matrix. Increasingly, such instrumentation is a hybrid of techniques for separation and detection that requires extensive data processing. So wide has the subject of Analytical Science become that complete coverage, providing information that is comprehensible to an interested scientist, can only be achieved in a multi-volume encyclopedia such as this. Even then, the length of each of the approximately 550 articles needs to be limited in order to keep the size of the encyclopedia manageable.

The encyclopedia covers all facets of modern analytical science, with articles from an international authorship of experts in their specialist fields. The articles cover three broad areas: analytical techniques (e.g., mass spectrometry, liquid chromatography, atomic spectrometry); areas of application (e.g., forensic, environmental, clinical); and analytes (e.g., arsenic, nucleic acids, polycyclic aromatic hydrocarbons). The authors and Editorial Advisory Board members are drawn from all continents and we are grateful to the great majority who met their deadlines.

The boundaries of Analytical Sciences are constantly pushing into new areas and we have taken a broad view of what material should be included. Comprehensive indexing and cross-referencing are important features that should allow rapid access to relevant information for users of the encyclopedia.

The first edition of this encyclopedia, published in 1995, was the inspiration of Robert Macrae. Following the success of the Encyclopedia of Food Science, Food Technology and Nutrition, of which he was a leading editor, he realized how valuable a similar Encyclopedia of Analytical Science would be. Dr. Macrae served as managing editor of the first edition until his unexpected death in November 1993. Without him this encyclopedia would never even have begun.

**Paul Worsfold, Alan Townshend and Colin Poole**

# Introduction

It is increasingly appreciated that knowledge of the nature and composition of materials gives a greater control of their properties. As the range of materials becomes more diverse and valuable, analytical science, which determines this nature and composition, also achieves greater recognition and attention.

Many attempts have been made to provide a satisfactory definition of analytical science. The most recent is that proposed by the Working Party on Analytical Chemistry of the Federation of European Chemical Societies (*Analytical Chemistry* (1994) 66; 98A–101A); it reads:

Analytical chemistry is a scientific discipline that develops and applies methods, instruments, and strategies to obtain information on the composition and nature of matter in space and time.

Thus, analytical science includes within its remit not only a considerable amount of chemistry, but also an increasing proportion of biochemistry, physics and electronics, computer science, mathematics and chemometrics, and even management and economics. But these are combined into a distinct area of science with its own philosophy, procedures, and objectives.

The increasing scope and the impressive rate of change of analytical science over the 10 years that have elapsed since the publication of the first edition of this encyclopedia are reflected in the extensive changes that have been made to the contents in producing this second edition. The majority of the articles are new or have been extensively rewritten, and all topics have been selected on the basis of their relevance to analytical science at the beginning of the twenty-first century. New articles include DNA sequencing, endocrine disrupting chemicals, ‘lab-on-a-chip’ technologies, field flow fractionation, nitric oxide, prions, and solid-phase microextraction, again giving a flavor of the breadth and relevance of modern analytical science. In a similar vein, subjects now considered to be less appropriate have not been included in this second edition.

The large number of articles in this encyclopedia and the wide variety of the subject matter emphasize the considerable scope of modern analytical science. The articles fall mainly into three classes:

- Analysis for particular analytes
- Analysis of particular types of samples
- Analytical techniques

Particular analytes that are the subjects of articles in the encyclopedia include a wide range of classes of organic compounds (e.g., amino acids, dioxins, humic and fulvic compounds, lipids, nucleic acids, polycyclic aromatic hydrocarbons, proteins) as well as specific compounds (e.g., ethanol, glucose). There is also an extensive selection of compounds having particular types of function (e.g., antioxidants, neurotoxins, pesticides, vitamins). Inorganic elements are not assigned individual articles, except for those elements where their speciation provides significant analytical challenges (e.g., arsenic, carbon, chromium, selenium, sulfur). The concentrations at which such analytes can be determined range from per cent levels, through trace concentrations ( $\mu\text{g ml}^{-1}$ ) to ultratrace levels ( $\text{ng ml}^{-1}$ ,  $\text{pg ml}^{-1}$ , and even less). Such is the sensitivity of some modern analytical techniques and procedures that the detection of individual molecules is now possible in some instances.

The types of sample that must be analyzed are numerous. They include raw materials, intermediates, products, and effluents of industrial processes. Analysis is essential for controlling the manufacturing process, the quality of the product, and the hazards of any discharges into the environment. Articles are included, therefore, on such diverse products as adhesives, building materials, ceramics, glasses, and paints, as well as on process analysis per se. There is a section on food analysis and on pharmaceutical compounds. Other groups of materials that are the subject of many articles are clinical samples and forensic specimens. Specific materials that merit individual articles include blood, coal, fertilizers, and meat. Particular importance is placed on the means of obtaining representative samples, and the processes to which they may be subjected before the analytical measurement is made. Equally, the quality of the analytical process is a matter that is dealt with in some depth, including standards, traceability, accreditation, and interlaboratory studies.

A considerable proportion of the encyclopedia is dedicated to descriptions of techniques and to the wide range of applications for which they are used. These include the instrumentation available for making the analytical measurement, for example, atomic absorption and emission spectrometry, chromatography and

electrophoresis, fluorimetry, mass spectrometry, nuclear magnetic resonance spectroscopy, X-ray fluorescence spectrometry, and the various surface analysis techniques. Other techniques of great utility are described, such as immunoassays, amplification reactions (including the polymerase chain reaction), and radiochemical methods.

With such a diversity of topics, some overlap between articles is inevitable and, indeed, desirable. Each article is intended to be self-contained, but extensive cross-references are included to enable further information on particular topics to be found elsewhere in the encyclopedia. Even in articles where there might, at first sight, scope for duplication, it will be seen that each article has its own distinct perspective. For example, there are articles on ethanol, on forensic sciences, determination of alcohol in body fluids, and on food and nutritional analysis, alcoholic beverages, but it can readily be seen that the emphasis in each article is very different.

A glance at the table of contents, in volume 10, will show that some topics merit a large number of articles, a reflection of their importance in current analytical science. Several techniques, for example, mass spectrometry, nuclear magnetic resonance spectroscopy, atomic emission spectrometry, microscopy, the various chromatographic techniques (e.g., gas, liquid and thin-layer), and electrophoresis, merit a series of articles, as do areas such as food and nutritional analysis, forensic sciences, archaeometry, pharmaceutical analysis, sensors, and surface analysis. Each of these collections of articles, written by experts in their fields, provides at least as much up-to-date information on that particular subject as a complete textbook.

In conclusion, this encyclopedia provides detailed information by acknowledged experts on most aspects of modern analytical science. It is designed to be easy to access and, if further information is required, bibliographies are provided. The grouping of subjects and the cross-referencing should emphasize both the variety and the unity of analytical science; that there is a thread that links what at first sight are very diverse topics, but which in fact demand a common philosophy. This ‘analytical approach’ is what the encyclopedia is all about.

**Paul Worsfold, Alan Townshend and Colin Poole**

## **Dedication**

The original idea for the *Encyclopedia of Analytical Science* came from Dr. Robert Macrae, who played a large part in its realization of the first edition, and scientific editor, until his sudden, untimely death in November 1993. This work is dedicated to him; we hope that it will serve as a lasting memorial to his enthusiasm for publishing, and commitment to scientific endeavor.

## **Disclaimer**

This encyclopedia is a guide providing general information concerning its subject matter; it is not a procedural manual. The readers should consult current procedural manuals for state-of-the-art instructions and applicable government safety regulations. The publisher and the authors do not accept responsibility for any misuse of this encyclopedia, including its use as a procedural manual or as a source of specific instructions.

# Permission Acknowledgments

The following material is reproduced with kind permission of the American Association for the Advancement of Science

Figure 6 of LIQUID CHROMATOGRAPHY | Chiral Analysis of Amino Acids

<http://www.sciencemag.org>

The following material is reproduced with kind permission of Oxford University Press

Figure 2 of DNA SEQUENCING

Figure 1 of ELECTRON SPIN RESONANCE SPECTROSCOPY | Principles and Instrumentation

Table 1 of ELECTROPHORESIS | Principles

Figures 1, 2, and 5 of MICROSCOPY | Overview

Figure 2 of SPECTROPHOTOMETRY | Biochemical Applications

<http://www.oup.com>

# A

## ACCREDITATION

See **QUALITY ASSURANCE: Accreditation**

## ACID-BASE INDICATORS

See **INDICATORS: Acid-Base**

## ACID-BASE TITRATION

See **TITRIMETRY: Overview; Potentiometric; Photometric**

## ACTIVATION ANALYSIS

Contents

**Neutron Activation**

**Charged-Particle Activation**

**Photon Activation**

### Neutron Activation

**S J Parry**, Imperial College of London, Ascot, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 8–16,  
© 1995, Elsevier Ltd.

### Introduction

Neutron activation analysis is a powerful nondestructive multielement technique that can be applied to the determination of over 60 elements in a broad range of matrices. The method involves activation of the sample in a source of neutrons followed by  $\gamma$ -ray spectrometry to identify and quantify the induced

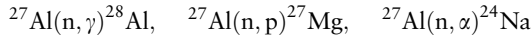
activity, and there from the concentration of the element in the sample. Although access to a reactor is required for the irradiation, the samples may then be transferred to another laboratory for analysis, in contrast to prompt activation where analysis is carried out while the sample is being irradiated. The advantages of the technique over other multielement methods for trace elements are that: (1) solids can be analyzed without any sample dissolution, and consequently the method is particularly useful for the analysis of insoluble materials like plastics, carbon fiber, and boron compounds, and in the determination of volatile elements that may be lost on dissolution, such as arsenic, selenium, and mercury; (2) extremely low detection limits may be attained, for example, the detection limit for gold in a 10 g sample of silicon is 6 pg per g; and (3) the method is



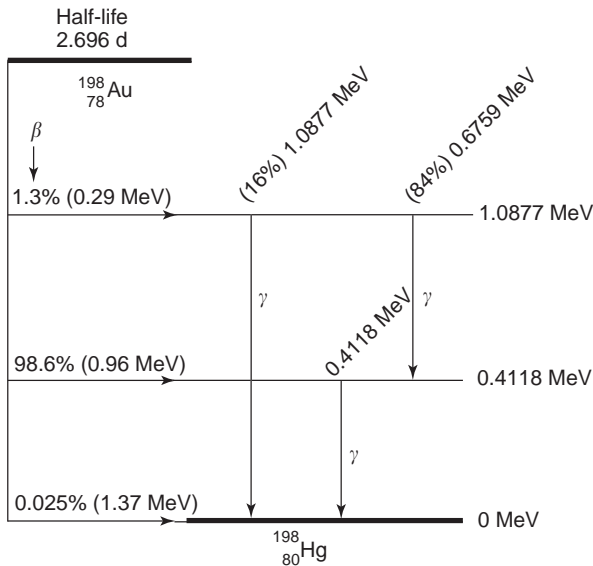
nondestructive and therefore the sample may be re-analyzed by the same, or an alternative, method.

## Basic Theory

The method is based on the fact that most elements have one or more stable isotopes that can be made radioactive on interaction with a neutron. For example, reactions that can occur with aluminum are



The  $(n, \gamma)$  interaction normally occurs in a thermal neutron flux, while the  $(n, p)$  and  $(n, \alpha)$  reactions are induced with fast neutrons. Many of these products are radioactive and decay with the production of  $\beta$  and  $\gamma$  radiation. **Figure 1** is the decay scheme for  $^{198}\text{Au}$ , showing the main  $\gamma$ -ray energy at 0.4118 MeV.  $\gamma$ -Rays are readily identified by their characteristic energies and so the activity of the product radionuclide is measured using  $\gamma$ -ray spectrometry in multielement neutron activation analysis. **Figure 2** shows schematically how the induced activity of the sample changes as it is irradiated for a time,  $t_i$ , and counted for a time,  $t_{\text{meas}}$ , after a decay period,  $t_d$ . The  $\gamma$ -ray intensity of the radionuclide,  $A_p$ , measured over time,  $t_{\text{meas}}$ , is quantified from the accumulated spectrum and expressed as counts per second (cps).



**Figure 1** The decay scheme for  $^{198}\text{Au}$ . (Based on data from Browne E and Firestone RB (1986) *Table of Radioactive Isotopes*. New York: Wiley; and Hammed MA, Lowles IM, and MacMahon TD (1992) Decay scheme data for  $^{154}\text{Eu}$ ,  $^{198}\text{Au}$  and  $^{239}\text{Np}$ . *Nuclear Instrumentation and Methods in Physics Research A* 312: 308–316.)

This count rate can be corrected for decay since irradiation ceased ( $\exp(-\lambda t_d)$ ), detector efficiency ( $E$ ), and branching ratio ( $P$ ), which is the number of  $\gamma$ -rays emitted per disintegration of the nucleus, in order to calculate the activity of the radionuclide,  $A_{t_i}$

$$A_{t_i} = A_p \exp(-\lambda t_d) / EP \quad [1]$$

where  $\lambda$  is the decay constant ( $= \ln 2 / T_{1/2}$ , where  $T_{1/2}$  is the half-life) of the product radionuclide and  $t_d$  is as shown in **Figure 2**.

The probability of an interaction taking place, called the cross-section, is highest for  $(n, \gamma)$  reactions and so these activation products are generally most useful. The activation rate or activity generated ( $A$ ) of the generated radionuclide is dependent on the cross-section for the reaction ( $\sigma$ ), the number of target nuclei ( $N$ ), and the neutron flux ( $\phi$ ):

$$A = \sigma N \phi \quad [2]$$

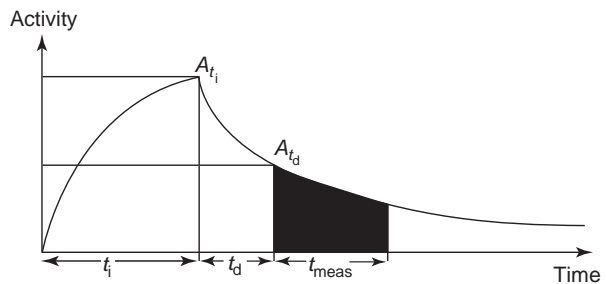
The number of target nuclei is dependent on the amount of element X in the sample and the abundance of the target isotope. If we take 1 g of element X, we have  $N_A / A_w$  atoms, where  $N_A$  is Avogadro's number, and  $A_w$  is the relative atomic mass of element X. If the isotopic abundance of the target is  $\theta$ , and the amount of the element X in grams is  $w$ , the number of target nuclei  $N$  is given by

$$N = N_A \theta w / A_w \quad [3]$$

Substituting [3] into [2]

$$A = \sigma N_A \theta \phi w / A_w \quad [4]$$

The net rate of production of the radionuclide ( $dN^*/dt$ ) is dependent on the activation rate



**Figure 2** Activity,  $A$ , as a function of time during irradiation and decay. The integrated activity measured is indicated by the shaded area.

(Eqn [2]) and the decay rate of the product ( $\lambda N^*$ ):

$$dN^*/dt = \sigma N\phi - \lambda N^* \quad [5]$$

where  $N^*$  is the number of radioactive nuclei. Therefore,

$$N^* = \sigma\phi N(1 - \exp(-\lambda t))/\lambda \quad [6]$$

So the disintegration rate, or activity, after irradiation for time  $t_i$  ( $A_{t_i}$ ) is given by

$$A_{t_i} = \sigma N_A \theta \phi w (1 - \exp(-\lambda t_i))/A_w \quad [7]$$

Equation [7] is the activation equation.

Substituting the activation equation [7] into eqn [1] gives

$$A_p = \frac{\sigma N_A \theta \phi EP w (1 - \exp(-\lambda t_i))}{A_w \exp(-\lambda t_d)} \quad [8]$$

The radionuclide will continue to decay during the measuring time ( $t_{\text{meas}}$ ), as shown in **Figure 2**, so it is necessary to correct  $A_p$  by the factor  $\lambda t_{\text{meas}}/(1 - \exp(-\lambda t_{\text{meas}}))$ , unless  $t_{\text{meas}}$  is short compared to the half-life of the radionuclide (and therefore to  $t_d$ ).

The amount of element X in the sample can be deduced from

$$w = \frac{A_p A_w \exp(-\lambda t_d)}{\sigma N_A \theta \phi EP (1 - \exp(-\lambda t_i))} \quad [9]$$

This expression is readily simplified when standards of known composition are irradiated and counted with the samples. If the irradiation and decay conditions are kept constant and all measurements are made with a constant neutron flux and detector geometry, the concentration by weight of element X becomes the simple ratio:

$$\frac{w_{(\text{standard})}}{w_{(\text{sample})}} = \frac{A_{p(\text{standard})}}{A_{p(\text{sample})}}$$

where  $w_{(\text{standard})}$  and  $w_{(\text{sample})}$  are the amounts of element X in the standard and the sample, respectively, and  $A_{p(\text{standard})}$  and  $A_{p(\text{sample})}$  are the measured count rates in the standard and sample, respectively.

In most analytical problems this is the most simple and straightforward procedure.

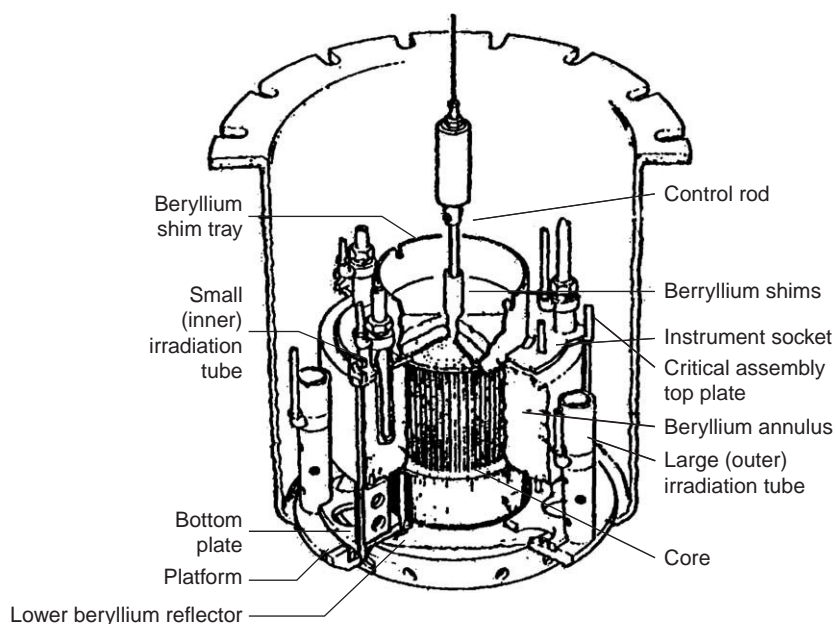
## Instrumentation

### Irradiation Devices

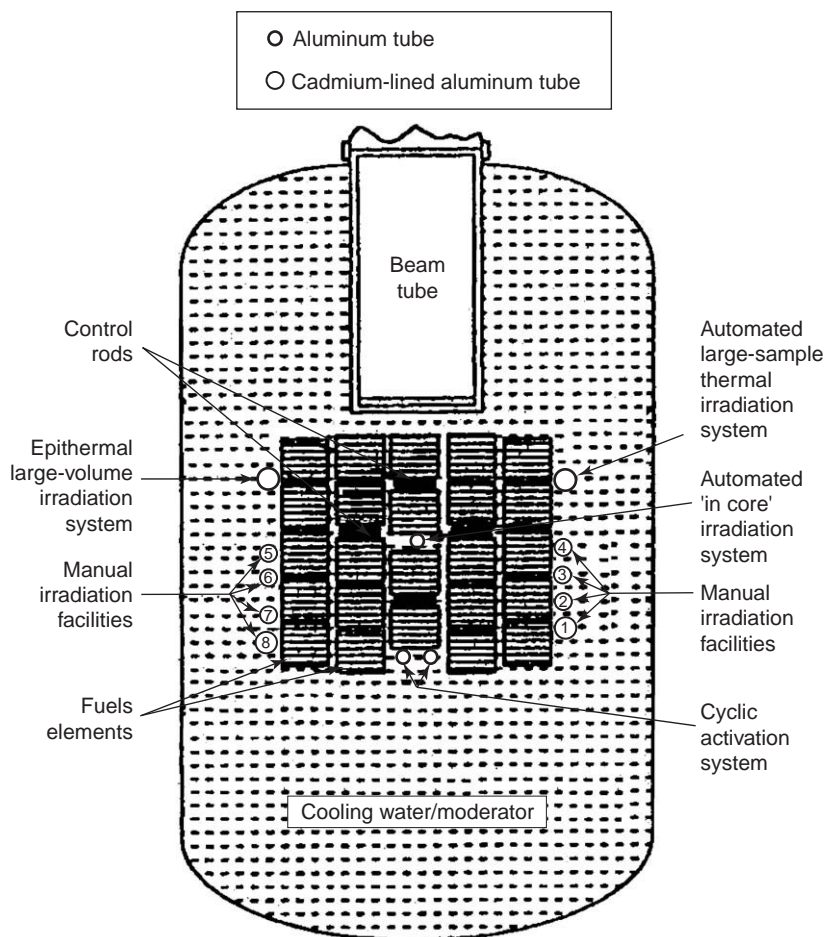
There are a number of different neutron sources available, including the 14 MeV neutron generator

and the smaller  $^{252}\text{Cf}$  or Am/Be sources, but the source most commonly used for activation analysis is the nuclear reactor. The reactors are based on a core of uranium enriched in  $^{235}\text{U}$ . The uranium core is surrounded with moderator of water or graphite to slow the neutrons emitted on fission. The heat generated during fission is dissipated with a coolant such as water. The rate of production of neutrons is controlled by rods made of neutron-absorbing material such as cadmium. Facilities for neutron activation range from low-power research reactors of 100–300 kW, to the larger reactors generating megawatts of power. The thermal neutron fluxes available in these sources range from  $10^{15}$  to  $10^{18} \text{ m}^{-2} \text{ s}^{-1}$ . Neutron fluxes of  $10^{15}$  to  $10^{16} \text{ m}^{-2} \text{ s}^{-1}$ , available in the lower-power research reactors, provide adequate sensitivity for most applications and these research reactors often provide better access and irradiation devices more suited to the activation of samples for analysis. The American TRIGA and Canadian Slowpoke reactors are typical pool-type facilities with light-water cooling and moderation. Most of the neutrons in the pool-type reactor are thermalized and have the most suitable energy for the  $(n, \gamma)$  reactions used in activation analysis. They also have a significant epithermal neutron component that may be used in the absence of thermal neutrons by employing a cadmium filter to remove neutrons with energies below 0.5 eV. The Slowpoke reactor in particular was designed as a low-power device suitable for operation in a university. The design, in **Figure 3**, shows the compact core, the central control rod, and the irradiation tubes.

Each reactor has its own design of irradiation facilities, but where activation analysis is carried out routinely it would be expected that there would be both manually loaded irradiation tubes for long irradiations, and pneumatic systems for irradiations of a few minutes. In addition there may be irradiation tubes with thermal neutron filters for epithermal neutron activation and a cyclic activation system for repeated activation of the same sample. The layout of the irradiation devices in the Imperial College (UK) research reactor is shown in **Figure 4**. The degree of automation can be quite variable and in some facilities hundreds of samples are irradiated totally automatically, in others even short irradiations with a pneumatic system may be timed manually. The most important consideration in activation analysis is that the timings on irradiation are reproducible. It will not matter if the length of irradiation cannot be measured very accurately provided it is repeatable and both standard and sample have the same length of irradiation.



**Figure 3** The Slowpoke reactor: a view of the critical assembly and reactor container, lower section. (Reprinted with permission from Atomic Energy of Canada Limited (AECL) Research, Canada.)



**Figure 4** A schematic diagram of the Imperial College reactor showing the pneumatic and manual irradiation facilities.

### Detection by $\gamma$ -Ray Spectrometry

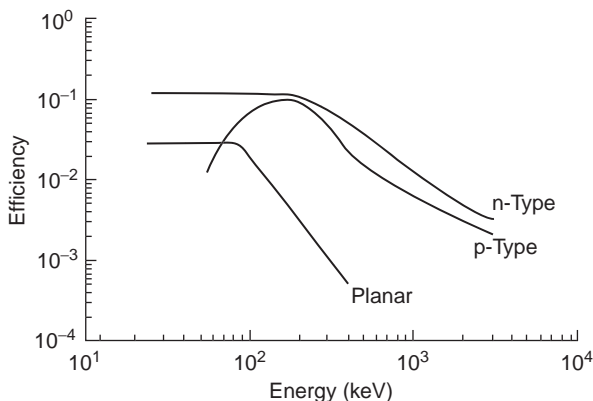
High-resolution semiconductor germanium detectors, operated at liquid nitrogen temperature, are used for  $\gamma$ -ray spectrometry when neutron activation analysis is used to determine a number of elements. Photoelectric absorption in the germanium crystal, leading to ionization, is used to detect and measure the energy of the  $\gamma$ -ray in the detector. The electric charge created by the ionization is collected by applying several thousand volts across the crystal through the preamplifier. The preamplifier amplifies the small pulse from the detector ( $\sim 0.05 \text{ pC MeV}^{-1}$ ) with a gain of  $\sim 2 \text{ V pC}^{-1}$  to  $\sim 0.1 \text{ V MeV}^{-1}$ . These pulses are shaped and amplified to a maximum of  $\sim 10 \text{ V}$ , depending on the gain setting. The analog-to-digital converter (ADC) changes the amplified pulse into a digital signal that is deposited as a count in the appropriate channel number in the analyzer. The analysis equipment required has been simplified by modern electronics, and the power supply and high voltage for the detector, the amplifier, the ADC, and the analyzer are now combined. This single unit, plus the detector and a personal computer is all the equipment required for analyzing the samples.

Several different types of  $\gamma$ -ray detectors are used for activation analysis. The most generally useful one is about a 40-ml volume crystal of n-type or p-type germanium, which covers a range of energies from 50 keV to several MeV. The resolution of the detectors is important and for adequate resolution of overlapping peaks it is usually necessary to use a detector that gives a full-width at half-maximum of 1.8–2.0 keV for the  $^{60}\text{Co}$  peak at 1.33 MeV. A thin beryllium window in the casing facilitates the measurement of low-energy  $\gamma$ -rays and X-rays down to a few keV. A planar germanium crystal is used to optimize

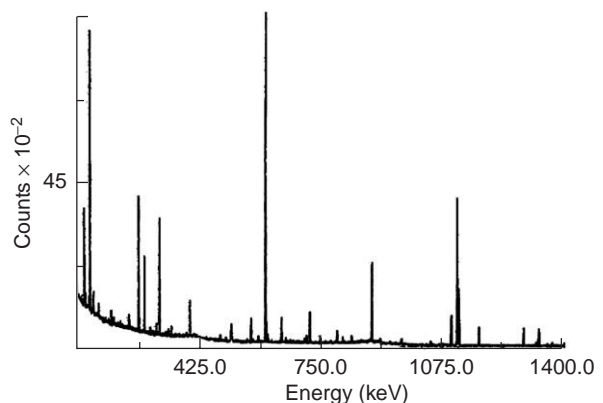
the resolution of low-energy  $\gamma$ -rays but the efficiency is poor at higher energies, as demonstrated by the plot of detector efficiency in Figure 5. Compton scattering occurs when a  $\gamma$ -ray interacts with the detector material and transfers some energy by inelastic scattering to the recoil electron, resulting in lower-energy background in the spectrum. A detector with a high peak-to-Compton ratio will give best signal-to-background ratio and hence detection limits.

Figure 6 shows a typical  $\gamma$ -ray spectrum, in this case for an air particulate sample on filter paper, Spectra, which are often complex, are processed to locate, identify, and evaluate the peaks using a personal computer or workstation running software purchased from the manufacturer of the electronics or with programs written in-house. Peak-search programs work through the spectrum, loading those channels where the counts are above the background using a preset value for sensitivity. The spectrometry system is calibrated prior to use using standard sources with precisely known energies and the peak locations are calculated from the centroid channel number using the energy calibration.

The  $\gamma$ -ray spectral lines are identified and assigned to the appropriate radionuclide using a library of known  $\gamma$ -ray energies. Interferences from radionuclides emitting a  $\gamma$ -ray of similar energy can lead to peaks being wrongly identified. The library of radionuclides can be comprehensive and the relative intensities for lines from a particular radionuclide can be used as a further check on its identity, provided that the efficiency of the detector is known. The same radionuclide can be produced by competing nuclear reactions from different elements and so it is also important to check the source of the radioactivity. The experienced analyst uses standards and reference materials to check for sources of these interferences.



**Figure 5** The efficiency curves for a coaxial p-type, a coaxial n-type, and a planar germanium detector. (Reprinted with permission from Parry SJ (1991) Activation spectrometry in chemical analysis.)



**Figure 6** The  $\gamma$ -ray spectrum of air particulate on filter paper, irradiated for 35 h, and counted for 6 h, 25 days after irradiation.

## Methodology

### Quantitative Analysis

The total number of counts under the peak is measured for each  $\gamma$ -ray energy, by subtracting the estimated counts in the background from the total peak area. The background under the peak is estimated by averaging the counts in the channels on either side of the peak. Multiplets are separated using Gaussian fitting routines to estimate the contributions from each peak. The activity in counts per second is compared to a known standard irradiated and counted under the same conditions to determine the concentration of the element in the sample.

### Use of Standards

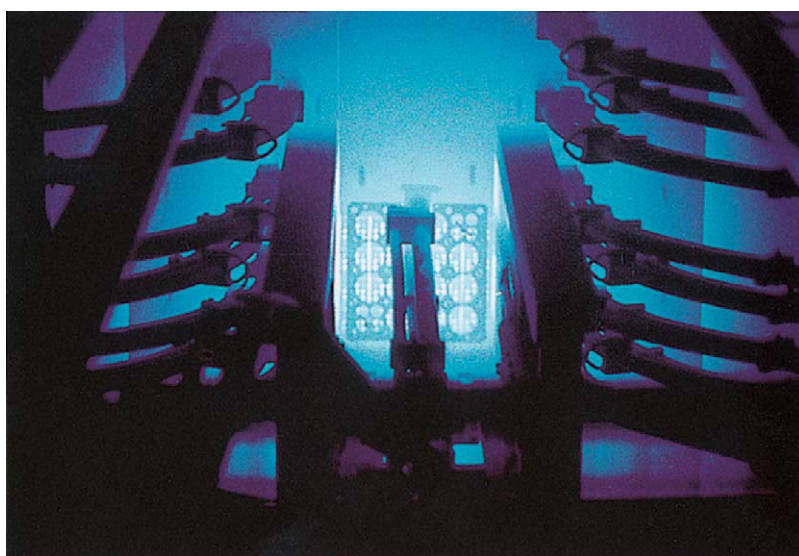
Standards are prepared as solids or in solution from pure elements or compounds. Standard solutions sold for inductively coupled plasma (ICP) spectrometric techniques are ideal. These are prepared as multi-element standards for use as liquids or dried on a suitable solid substrate. The standard should have a similar geometry to that of the sample. The composition is not so critical, however, and standards can be used over several orders of magnitude. If irradiation and counting conditions are kept constant during analysis, the standard data may be stored in a database for use on future occasions, with inclusion of a monitor to check the irradiation and counting conditions. In variable conditions of irradiation, decay and  $\gamma$ -ray spectrometry standards are included with the samples during analysis, and this is the common

approach. Reference materials of suitable composition are used to check the validity of the whole procedure.

### Irradiation

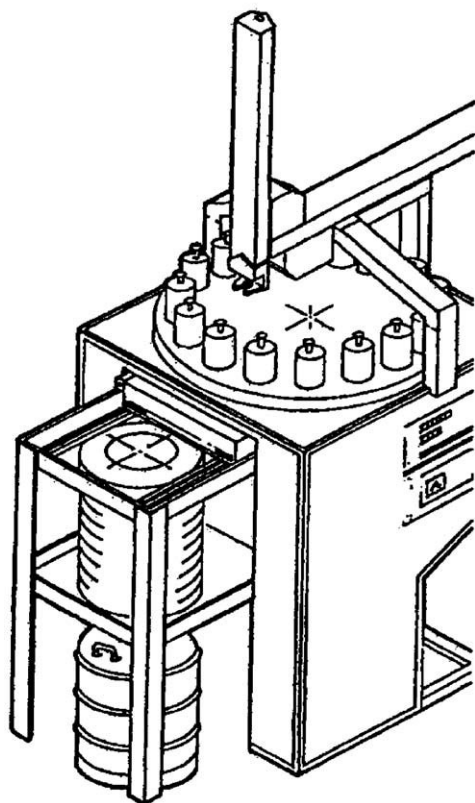
The simplicity of sample preparation is one of the main advantages of the technique. Solids, liquids, and even gases are analyzed directly as received with no prior treatment. The sample is sealed in a suitable container, usually of polyethylene or quartz glass, and placed in the irradiation position in the reactor. When quantitative analysis is carried out with chemical standards, it is necessary to ensure that the samples have a constant geometry for irradiation and  $\gamma$ -ray spectrometry. Samples are therefore used in powder form or as chunks of uniform shape and thickness. In the case of liquids, identical volumes must be used to provide consistent geometry and great care is required to ensure that no leaks occur in the irradiation site. Reactor operators have strict control over samples allowed in the irradiation devices and hazardous materials will not be permitted.

Once encapsulated, the samples are usually placed in an outer irradiation can of polyethylene or aluminum. The irradiation can is then placed for the required time in the irradiation site. When irradiation is complete the sample is taken out of the reactor, either manually or through a pneumatic system. At this stage the samples may be sent directly to a counting device as part of a totally automated system on site, or to an analyst who could be working on the other side of the world (Figure 7).



**Figure 7** Nuclear reactor Thetis dedicated to neutron activation analysis. View in the 'swimming pool' with nuclear fuel element, Cerenkov radiation, safety and control plates and pneumatic transport tubes. (Reproduced with permission from Institute of Nuclear Sciences, Gent, Belgium.)





**Figure 8** A carousel-type sample changer for large samples, designed for a standard lead castle. (Reprinted with permission from EG & G Instruments, Wokingham; L.E. Pink Engineering Ltd, Reading and Fab Cast Engineering Ltd, Dartford.)

### Counting

Detector systems have a variety of configurations and the endcap of the detector casing may be vertical, horizontal, or even at an angle. The sample is positioned in a sample holder, which is usually made of plastic to avoid attenuation of the  $\gamma$ -rays. The efficiency of the detector is dependent on the distance between the radioactive source and the detector endcap, so the sample holder is designed to give reproducible geometry, at a fixed distance. Sample changers are used for automated counting and they can be purchased commercially or manufactured in house. The designs vary from carousel-like devices, shown in Figure 8, which move the samples over a vertical dipstick detector on a circular turntable, to pneumatic systems that blow the samples sequentially through a tube to the counting position, usually with a detector in a horizontal configuration.

### Choice of Irradiation and Counting Conditions

The choice of the analytical procedure is dependent on the elements to be determined and the sample

matrix. Neutron activation analysis is a multielement method of analysis that is usually applied to the determination of more than one element. However, because of the minimal sample preparation required, it is also sometimes useful for just one element in the analysis of difficult materials. If there is one particular element to be determined, the conditions can be optimized for that element. The choice of irradiation time will depend on the half-life of the radionuclide of interest, but as a general rule an irradiation of one to two half-lives will be sufficient. In the case of radionuclides such as  $^{60}\text{Co}$ , which has a half-life of over 5 years, it is not possible to irradiate for so long. The decay period before analyzing the samples will depend on many unavoidable factors, including accessibility and transportation, but ideally they should be counted before the radionuclide of interest has decayed significantly. The sample is counted for sufficient time to allow good statistics on the peak of interest, and there is little point in counting after the radionuclide has gone through two half-lives.

Normally the sample is analyzed for more than one radionuclide, so the choice of irradiation, decay, and counting times is a compromise. However, there are a small number of procedures that are typically used for the determination of the whole suite of elements, and these are shown in Table 1 for a variety of sample types. In general, sodium, magnesium, aluminum, silicon, sulfur, chlorine, potassium, titanium, vanadium, manganese, copper, iodine, and dysprosium are measured after a short irradiation of between 10 s and 10 min and a thermal neutron flux of between  $5 \times 10^{15}$  and  $5 \times 10^{17} \text{ m}^{-2} \text{ s}^{-1}$ . Decay periods of 10 s to 2.5 h are allowed, before counting for 20–1000 s. The remaining elements are determined with a long irradiation for 1–72 h, with decay periods of 20 h to 30 days, and counting times of 2000 s to 12 h.

### Sources of Error

The major interferences in the  $\gamma$ -ray spectrum result from high backgrounds due to elements in the sample matrix. Detection limits are restricted by the background continuum of energy produced from other  $\gamma$ -emitters due to the Compton effect. For example, sodium in blood, scandium in rocks, and manganese in steel will all activate well and cause background activities, thus giving poorer detection limits for other elements. This background effect is limited by using specialized equipment such as a Compton suppression detector or by activation with epithermal neutrons to reduce the thermal neutron activation of the matrix.

**Table 1** Typical conditions for multielement neutron activation analysis

Type of sample	Neutron flux ( $m^{-2}s^{-1}$ )	Irradiation time	Decay period	Count time	Elements determined
Hair <sup>a</sup>	$5 \times 10^{15}$	30 s	10 s	20 s	Ag, Cl, F, Se
		10 min	5 min	5 min	Al, Ba, Ca, Cu, I, K, Mn, Na, S, Zn
		16 h	2 d	3000 s	As, Au, Sb
Air particulates <sup>b</sup>	$2 \times 10^{16}$	5 min	21 d	3000 s	Co, Hg, Ag, Ba, Se, Zn
			3 min	400 s	Al, V, Ti, S, Ca, Cu
	$1.5 \times 10^{17}$	2–5 h	15 min	15 min	Mg, Br, Cl, I, Mn, Na, In
			20–30 h	2000 s	K, Zn, Br, As, Ga, Sb, La, Sm, Eu, W, Au
Marine biological sample <sup>c</sup>	$5 \times 10^{16}$	10 s	150 s	10 min	Sc, Cr, Fe, Co, Ni, Zn, Se, Ag, Sb, Ce, Hg, Th
					Ca, Cu, Cl, Mg, Na, V
	$7 \times 10^{16}$	1 h	2.5 h	1000 s	K, Mn
			7 d	1 h	Sb, As, La, Sm
Leafy vegetables <sup>d</sup>	$4.9 \times 10^{17}$	15 s	14 d	1 h	Ag, Ce, Cr, Co, Eu, Fe, Hf, Rb, Sc, Se, Zn
			2 min	10 min	Ca, Cl, Cu, K, Mg, Mn, Na, V
			4 h	6–8 h	Ag, As, Br, Cd, Co, Cr, Cs, Eu, Fe, Rb, Sb, Sc, Se, Zn
Deep sea sediment <sup>e</sup>	$5 \times 10^{15}$	4 min	7 min	10 min	Al, Ca, Cl, I, Mg, Mn, Na, Si, Ti, U, V
		7 h	3–4 d	30 min	As, Br, K, La, Sb, Sm, Yb
			25 d	12 h	Ba, Ce, Co, Cr, Cs, Eu, Fe, Gd, Hf, In, Lu, Nd, Rb, Sc, Se, Sr, Ta, Tb, Th, Tm, Zn, Zr

<sup>a</sup>Chatt A, Sayjad M, De Silva KN, and Secord CA (1985) Human scalp hair as an epidemiologic monitor of environmental exposure to elemental pollutants. *IAEA, Health-related Monitoring of Trace Element Pollutants using Nuclear Techniques*, pp. 33–49. IAEA TECDOC-330. Vienna: IAEA.

<sup>b</sup>Dams R, De Corte F, Hertogen J, *et al.* (1976) Activation analysis—Part 1. In: West TS (ed.), *Physical Chemistry Series Two*, pp. 24–38. International Review of Science. London: Butterworths.

<sup>c</sup>Simmons A and Landsberger S (1987) Analysis of marine biological certified reference material by various nondestructive neutron activation methods. *Journal of Radioanalytical and Nuclear Chemistry* 110(2): 555–564.

<sup>d</sup>Cunningham WC and Stroube WB Jr (1987) Application of an instrumental neutron activation analysis procedure to analysis of food. *The Science of the Total Environment* 63: 29–43.

<sup>e</sup>Fong BB and Chatt A (1987) Characterization of deep sea sediments by INAA for radioactive waste management purposes. *Journal of Radioanalytical and Nuclear Chemistry* 110(1): 135–145.

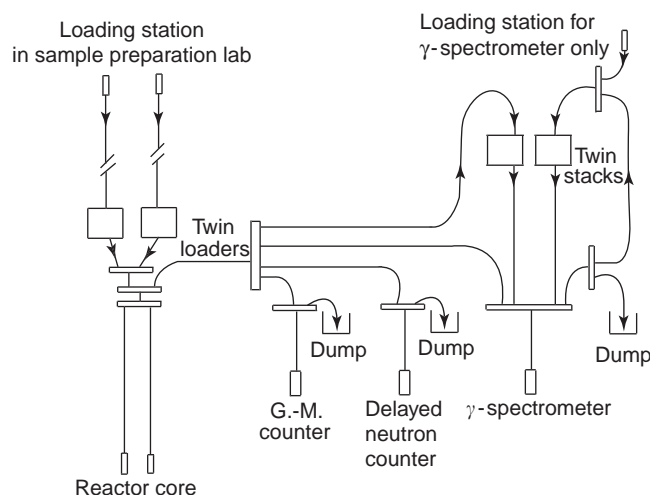
Interferences that can lead to erroneous results include overlapping peaks from radionuclides with similar  $\gamma$ -ray energies and the production of the same radionuclide from two different isotopes as the result of different nuclear reactions, for example, the production of  $^{28}\text{Al}$  from the  $^{27}\text{Al}(n, \gamma)$  reaction and the  $^{28}\text{Si}(n, p)$  reaction. There is also the possibility of neutron self-shielding, where the activation of the isotope is diminished by the absorption of neutrons within the matrix of the sample; in the case of very high concentrations of the element, self-shielding by the element of interest itself may occur.

Based on counting statistics, the  $3\sigma$  detection limit is given by  $3B^{0.5}$ , where  $B$  is the background area. The error associated with a net signal area  $A$  on a background  $B$  is  $(A + 2B)^{0.5}$ , and the error for the sample is combined with the error for the standard to give total counting error. In addition, errors associated with counting geometry and neutron

flux variation on irradiation have to be taken into account.

### Safety

In common with all techniques involving radioactivity, particular care is taken to ensure that the hazard from the irradiated samples is minimized. The electronics of the  $\gamma$ -ray spectrometry systems used in this technique cannot work at very high count rates and so it is not usually necessary or even advisable to induce high levels of radioactivity in samples for activation analysis. There is now a greater stress on the principle of keeping risk as low as possible, and so use is made of automation to reduce the contact time for radiation workers. **Figure 9** shows a totally automated system for the irradiation and analysis of samples, avoiding any exposure to radiation for the operator.



**Figure 9** A totally automated system, based on pneumatic transfer, for neutron activation analysis with  $\gamma$ -ray spectrometry and delayed neutron counting. (Reprinted with the permission from Atomic Energy of Canada Limited (AECL) Research, Canada.)

**Table 2**  $3\sigma$  Detection limits for neutron activation analysis

Element	Ideal conditions <sup>a</sup> ( $\mu\text{g kg}^{-1}$ )	Air particulates <sup>b</sup> ( $\text{ng m}^{-3}$ )	Sediments ( $\text{mg kg}^{-1}$ ) <sup>c</sup>
Aluminum	1	8	130
Antimony	0.05	1	0.18
Arsenic	0.01	4	10
Barium	1	—	90
Bromine	0.01	2.5	—
Cadmium	0.5	—	—
Calcium	100	200	3000
Cerium	0.1	0.25	0.26
Cesium	0.1	—	0.24
Chlorine	1	100	280
Chromium	3	0.25	1.2
Cobalt	0.01	0.025	0.14
Copper	0.01	20	—
Dysprosium	0.001	—	—
Erbium	0.1	—	—
Europium	0.01	0.01	0.059
Gadolinium	0.1	—	2.4
Gallium	0.05	1	—
Germanium	10	—	—
Gold	0.0001	0.1	—
Hafnium	0.1	—	0.33
Holmium	0.2	—	—
Indium	0.1	0.04	0.60
Iodine	0.01	20	3.2
Iridium	0.001	—	—
Iron	100	20	105
Lanthanum	0.01	0.2	0.85
Lead	1000	—	—
Lutetium	0.01	—	0.03
Magnesium	10	600	812
Manganese	0.01	0.6	16
Mercury	0.1	0.1	—
Molybdenum	1	—	—
Neodymium	0.5	—	2.4
Nickel	10	20	—
Niobium	1	—	—
Osmium	1	—	—

**Table 2** Continued

Element	Ideal conditions <sup>a</sup> ( $\mu\text{g kg}^{-1}$ )	Air particulates <sup>b</sup> ( $\text{ng m}^{-3}$ )	Sediments ( $\text{mg kg}^{-1}$ ) <sup>c</sup>
Palladium	0.05	—	—
Phosphorus	1	—	—
Platinum	0.5	0.5	—
Potassium	0.2	7.5	907
Praseodymium	0.01	—	—
Rhenium	0.01	—	—
Rhodium	0.1	—	—
Rubidium	0.1	—	1.1
Ruthenium	1	—	—
Samarium	0.005	0.005	0.15
Scandium	0.01	0.004	0.019
Selenium	5	0.1	4.3
Silicon	10	—	—
Silver	1	1	—
Sodium	0.1	5	360
Strontium	10	—	100
Sulphur	10	5000	—
Tantalum	0.05	—	0.077
Terbium	0.1	—	0.11
Thorium	0.1	0.04	1.6
Thulium	0.01	—	0.070
Tin	1	—	—
Titanium	3	40	430
Tungsten	0.01	0.5	—
Uranium	0.01	—	0.39
Vanadium	0.01	2	4.8
Ytterbium	0.01	—	0.69
Zinc	10	1	220
Zirconium	100	—	90

<sup>a</sup>Revet G (1987) *Comparison of Nuclear Analytical Methods with Competitive Methods*, pp. 147–162. IAEA-TECDOC-435. Vienna: IAEA.

<sup>b</sup>Dams R, Robbins JA, Rahn KA, and Winchester JW (1970) *Analytical Chemistry* 42(8): 861–868.

<sup>c</sup>Fong BB and Chatt A (1987) *Journal of Radioanalytical and Nuclear Chemistry* 110(1): 135–145.



## Applications

Activation analysis can be applied to most types of material, including food products, urine, feces, air pollution particulates, river water, marine samples, vegetation, soils, sediments, ores, plastics, petroleum products, pharmaceuticals, coal, metals, alloys, semiconductor materials, clays, ceramics, and glasses. Variations in the analytical procedures for multielemental analysis are dictated by the matrix and so are detection limits. Table 2 compares the detection limits for an air particulate sample on filter paper and those for a deep-sea sediment to the detection limits in the ideal situation. Petroleum products, pharmaceuticals, plastics, carbon products, and air filters are not activated to produce  $\gamma$ -rays, so the spectrum is free of any interferences from the matrix. This means that detection limits are low and over 60 elements can be determined simultaneously. Soils, sediments, clays, and glasses have elements such as sodium and scandium in their matrices that can produce background activities that result in poorer detection limits for most elements.

Metals generally activate well, which means that they are not easily analyzed for trace elements. However, metals such as vanadium and titanium can be left after irradiation to allow the short-lived activity to decay before counting the longer-lived

radionuclides of interest. Semiconductor materials such as pure silicon are ideal for analysis once the activity from the matrix has decayed, and this is a growing area of application. The fact that neutron activation analysis does not require sample dissolution is of particular value in all the applications that have been mentioned here.

*See also:* **Quality Assurance:** Spectroscopic Standards; Reference Materials; Production of Reference Materials.

## Further Reading

- Alfassi ZB (ed.) (1989) *Activation Analysis*. Boca Raton: CRC Press.
- Amiel S (ed.) (1981) *Nondestructive Activation Analysis*. Amsterdam: Elsevier.
- Das HA, Faanhof A, and van der Sloot HA (1989) *Environmental Radioanalysis*. Amsterdam: Elsevier.
- Das HA, Faanhof A, and van der Sloot HA (1989) *Radioanalysis in Geochemistry*. Amsterdam: Elsevier.
- Ehmann WD and Vance DE (1991) *Radiochemistry and Nuclear Methods of Analysis*. New York: Wiley.
- Heydorn K, (1984) *Neutron Activation Analysis for Clinical Trace Elements Research*, vols. I and II. Boca Raton: CRC Press.
- Parry SJ (2003) *Handbook of Neutron Activation Analysis*. Woking: Viridian Publishing.
- Tolgyessy J and Klehr BH (1987) *Nuclear Environmental Chemical Analysis*. Chichester: Ellis Horwood.

## Charged Particle Activation

**K Strijckmans**, Ghent University, Ghent, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Activation analysis (AA) is an analytical method for determining the (stable isotope) concentration of trace elements in the bulk of a sample. Speciation (i.e., discrimination between different oxidation states of the analyte element or between different compounds) is not possible with AA. AA is based on nuclear reactions leading to radionuclides specific for a given element. After irradiation these radionuclides are measured. Identification of the radiation emitted by measuring its energy and/or half-life provides qualitative analysis, while measuring the number of particles or photons emitted per second (i.e., (radio) activity) yields quantitative analysis. Charged particle activation analysis (CPAA) is based on charged

particle-induced reactions, while neutron activation analysis (NAA) and photon activation analysis (PAA) make use of neutron- and photon-induced reactions, respectively. Although CPAA, NAA, and PAA are all based on the same principle of AA, charged particle-induced reactions (and consequently CPAA) are fundamentally different from neutron- or photon-induced reactions.

## Stopping Power and Range

When a sample is irradiated with charged particles, only a small fraction of them induce nuclear reactions. Charged particles passing through matter lose energy mainly by interaction with electrons, leading to excitation or ionization of the target atoms or molecules (at least for the energy interval of interest for CPAA).

The stopping power of a target for a charged particle is defined as the energy loss per unit of thickness

and is commonly expressed in CPAA in terms of  $\text{MeV g}^{-1} \text{cm}^2$ :

$$S = -\frac{1}{\rho} \frac{dE}{dl} \quad [1]$$

where  $E$  is the energy (MeV),  $\rho$  is the mass density ( $\text{g cm}^{-3}$ ), and  $l$  is the thickness (cm).

For a sufficiently high energy ( $> 1$  MeV) the stopping power of the elements can be calculated by the Bethe formula. The stopping power depends on the charged particle and its energy and the target material. The stopping power is roughly proportional to  $Z_a^2 m_a / E_a$ , where  $Z_a$  is the atomic number,  $m_a$  is the

mass, and  $E_a$  is the energy of the charged particle. The stopping power increases with decreasing atomic number ( $Z_a$ ) of the target material. Tables 1 and 2 summarize some stopping power data, which can be calculated using the SRIM (stopping and ranges of ions in matter) program of Ziegler.

For mixtures and compounds the stopping power can be calculated from stopping power data for the elemental components by the additivity rule of Bragg and Kleeman:

$$S = \sum w_i \cdot S_i \quad [2]$$

**Table 1** Stopping power in ( $\text{MeV g}^{-1} \text{cm}^{-2}$ ) of 10 elements for proton energies ( $E_p$ ) between 1 and 20 MeV and deuteron energies ( $E_d$ ) between 2 and 40 MeV

$E_p$	$E_d$	$^1H$	$^2He$	$^3Li$	$^{13}Al$	$^{22}Ti$	$^{32}Ge$	$^{42}Mo$	$^{53}I$	$^{73}Ta$	$^{92}U$
1	2	683	281	229	176	142	116	104	89.3	67.0	60.4
2	4	389	168	138	110	93.2	76.5	68.2	58.0	48.7	43.1
3	6	279	122	101	82.9	71.1	59.4	53.5	45.9	39.3	34.7
4	8	220	96.6	80.5	67.4	58.2	49.1	44.6	38.6	33.3	29.6
5	10	183	80.6	67.5	57.2	49.7	42.3	38.6	33.6	29.2	26.0
6	12	157	69.5	58.3	50.0	43.5	37.3	34.2	29.9	26.1	23.3
7	14	138	61.3	51.5	44.5	38.9	33.5	30.8	27.0	23.6	21.1
8	16	123	54.9	46.2	40.2	35.2	30.4	28.1	24.7	21.7	19.4
9	18	112	49.8	42.0	36.7	32.3	28.0	25.9	22.8	20.1	18.0
10	20	102	45.7	38.5	33.9	29.8	25.9	24.2	21.2	18.8	16.9
12	24	87.6	39.3	33.2	29.4	26.0	22.7	21.1	18.7	16.6	15.0
14	28	76.9	34.6	29.2	26.1	23.1	20.3	18.9	16.8	15.0	13.5
16	32	68.7	30.9	26.2	23.5	20.9	18.3	17.2	15.3	13.6	12.3
18	36	62.2	28.0	23.8	21.4	19.1	16.8	15.8	14.1	12.6	11.4
20	40	56.9	25.7	21.8	19.7	17.6	15.5	14.6	13.1	11.7	10.6

**Table 2** Stopping power ( $\text{MeV g}^{-1} \text{cm}^2$ ) of 10 elements for helium-3 energies ( $E_{3He}$ ) between 1 and 30 MeV and helium-4 energies ( $E_{4He}$ ) between 1.3 and 40 MeV

$E_{3He}$	$E_{4He}$	$^1H$	$^2He$	$^3Li$	$^{13}Al$	$^{22}Ti$	$^{32}Ge$	$^{42}Mo$	$^{53}I$	$^{73}Ta$	$^{92}U$
1	1.3	4482	1729	1707	1140	975	652	679	535	389	417
2	2.7	2996	1285	1159	863	721	547	505	417	310	311
3	4.0	2337	987	889	692	579	460	405	349	263	255
4	5.3	1957	817	731	580	488	395	346	302	232	220
5	6.7	1700	711	625	502	423	347	306	268	210	195
6	8.0	1511	639	548	443	375	309	275	241	192	176
7	9.3	1366	586	489	398	337	280	252	220	178	161
8	10.7	1249	544	443	362	307	255	233	202	166	148
9	12.0	1152	511	405	332	282	235	216	187	156	138
10	13.3	1071	483	374	308	261	218	203	175	147	129
12	16.0	885	388	322	271	233	197	180	159	133	119
14	18.7	779	343	286	242	209	178	163	144	121	109
16	21.3	696	307	257	219	190	162	149	132	112	101
18	24.0	631	279	234	201	174	150	138	122	104	93.7
20	26.7	577	256	214	186	161	139	128	114	97.5	87.8
22	29.3	532	237	199	172	150	130	120	107	91.7	82.6
24	32.0	494	220	185	161	141	122	113	101	86.7	78.2
26	34.7	462	206	173	152	133	115	107	95.3	82.3	74.2
28	37.3	434	194	163	143	126	109	101	90.5	78.4	70.7
30	40.0	409	183	154	136	119	104	96.6	86.3	74.9	67.6

where  $w_i$  is the mass fraction of component  $i$ , and  $S_i$  is the stopping power of component  $i$ .

A charged particle loses only a very small fraction of its energy in a single interaction with an electron. As it is barely deflected from its original direction, charged particle paths are straight lines. Charged particles with the same incident energy will be stopped at a given depth in the target, the range. The relation between the range ( $R$ ) of a charged particle with incident energy  $E_i$  in a particular target and the stopping power ( $S$ ) of that target for that particular charged particle is given by

$$R = \int_{E_i}^0 \left( \frac{1}{\rho} \frac{dE}{dL} \right)^{-1} dE = \int_0^{E_i} \left( \frac{1}{S} \right) dE \quad [3]$$

The range is commonly expressed in CPAA as mass thickness ( $\text{g cm}^{-2}$ ). Ranges of protons or deuterons are larger than ranges of helium-3 or helium-4 particles and ranges in the very low atomic number targets (e.g., hydrogen) are smaller than for high atomic number targets.

## Nuclear Reactions

CPAA is based on a nuclear reaction  $A(a,b)B$  where a stable target nuclide (at rest)  $A$  is irradiated by accelerated charged particles  $a$  to form a radionuclide  $B$  with emission of one (or several or a cluster of) particle(s)  $b$ . Important characteristics are the minimum particle energy required to induce a nuclear reaction and the probability that this reaction will take place. As a charged particle is slowed down when traversing matter, this probability (cross-section, see below) should be known as a function of the energy.

A nuclear reaction  $A + a \rightarrow B + b + Q$  or  $A(a,b)B$  that releases energy is called an exoergic reaction:  $Q > 0$ . A nuclear reaction where energy must be provided (as kinetic energy of the charged particle) is called an endoergic reaction:  $Q < 0$ . In an exoergic (endoergic) reaction, mass-energy is thus lost (created). Considering the equivalence between mass and energy,

$$E = mc^2 \quad [4]$$

where  $E$  is energy in joules,  $m$  is mass in kilograms, and  $c$  is the velocity of light ( $2.997\,92 \times 10^8 \text{ m s}^{-1}$ ), the  $Q$ -value in joules is given by

$$Q = (m_A + m_a - m_B - m_b)c^2 \quad [5]$$

where  $m_i$  are the masses of the nuclides  $A$ ,  $a$ ,  $B$ , and  $b$  in kilograms.

More practically, the  $Q$ -value in MeV is given by

$$Q = 931.5(m_A + m_a - m_B - m_b) \quad [6]$$

where  $m_i$  are the masses of the nuclides  $A$ ,  $a$ ,  $B$ , and  $b$  in unified atomic mass units (u) because, according to eqn [4],  $m = 1 \text{ u}$  corresponds to  $E = 931.5 \text{ MeV}$ . Mass in unified atomic mass units  $1 \text{ u} \approx 1.66054 \times 10^{-27} \text{ kg}$ . Energy in electronvolts:  $1 \text{ eV} \approx 1.602 \times 10^{-19} \text{ J}$ .  $Q$ -values (in MeV) have been compiled by Nuclear Data Centers.

Endoergic reactions ( $Q < 0$ ) are only energetically possible for charged particles with a minimum kinetic energy, the threshold energy  $E_t$ . The threshold energy is slightly higher than  $-Q$ . The compound nuclide, formed by collision of the charged particle  $a$  and the target nuclide  $A$ , retains a fraction of the kinetic energy of the charged particle as kinetic energy. This fraction is approximately

$$\frac{A_a}{A_A + A_a} \quad [7]$$

where  $A_i$  are the nucleon numbers of the nuclides  $A$  and  $a$ , and is thus 'lost' to compensate the shortage of mass-energy for an endoergic reaction. Nucleon (or mass) number  $A = Z + N$ , where  $Z$  is the proton (or atomic) number and  $N$  is the neutron number. Thus, for endoergic reactions ( $Q < 0$ ) the threshold energy is given by

$$E_t \approx \frac{-Q(A_A + A_a)}{A_A} \quad [8]$$

For exoergic reactions ( $Q > 0$ ) the threshold energy is zero by definition.

The Coulomb barrier also determines the minimum energy of the charged particle needed to induce a nuclear reaction. The Coulomb repulsive force between the target nucleus and the charged particle dominates at 'large' distances and increases with decreasing distance of separation between charged particle and target nucleus, until the charged particle comes within the range of the attractive nuclear forces of the target nucleus. At some particular distance (i.e., the sum of the charged particle and target nucleus radii) the forces balance each other, and at shorter distances the attractive nuclear force dominates. The decrease in kinetic energy is given by the Coulomb barrier

$$E_C(\text{MeV}) = 1.02 \frac{A_A + A_a}{A_A} \frac{Z_A Z_a}{\sqrt[3]{A_A} + \sqrt[3]{A_a}} \quad [9]$$

The kinetic energy that the charged particle has lost through the Coulomb barrier is released again when a nuclear reaction occurs. Consequently, the Coulomb barrier does not influence the energetics

of a nuclear reaction except that a charged particle must have a kinetic energy higher than the Coulomb barrier in order that the reaction can occur. Table 3 gives approximate values of the Coulomb barrier for protons, deuterons, helium-3, and helium-4 particles as a function of the atomic number of the target. Quantum mechanical treatment of the problem indicated that the reaction probability for particles with kinetic energies lower than the Coulomb barrier is not exactly zero but is very low and drops rapidly as the energy of the charged particle decreases.

The probability of a nuclear reaction occurring is expressed as the (nuclear reaction) cross-section; it has the dimensions of area. This originates from the simple picture that the probability for a reaction between the target nucleus and the incident charged particles is proportional to the geometric cross-section that the target nucleus presents to a beam of charged particles. The geometric cross-section is of the order of magnitude of  $10^{-28} \text{ m}^2$ . The barn ( $1 \text{ b} = 10^{-28} \text{ m}^2$ ) is used as a unit for reaction probability. The cross-section for a particular reaction (also called partial reaction cross-section) depends on the energy of the charged particle. It is zero for charged particles below the threshold energy of that reaction and very low below the Coulomb barrier. At higher energies the cross-section increases up to a maximum (typically 1 b) and decreases as more complex reactions become competitive. Experimental or calculated cross-section data (i.e., excitation functions) have been compiled by Nuclear Data Centers.

## Interferences

For the determination of an analyte element A, based on the reaction  $A(a,b)B$  where B is a radionuclide, three different types of interferences can be distinguished: nuclear, spectral, and matrix interference.

If the radionuclide B is formed out of another element C than the analyte element A, the determination of A by reaction  $A(a,b)B$  can be subject to nuclear interference from C by the reaction  $C(c,d)B$ . Generally, particles a and c are identical. However, nuclear interference by secondary particles (e.g., fast or thermalized neutrons produced by  $(x,n)$  reactions where x is the charged particle) should also be considered. Only if the threshold energy of the interfering reaction is higher than the threshold energy of the analyte reaction, can nuclear interference be avoided by a proper choice of the charged particle energy. To obtain high sensitivity the charged particle energy is chosen just below the threshold energy of the interfering reaction.

If a radionuclide D is formed out of another element C than the analyte element A, the determination of A by reaction  $A(a,b)B$  can be subject to spectral interference from C by the reaction  $C(c,d)D$ . Three methods can be applied to avoid spectral interference:

1. Proper choice of the charged particle energy (see Nuclear Interference).
2. Selective measurement of radionuclide B besides radionuclide D by spectrometry or decay

**Table 3** Coulomb barrier (MeV) for protons (p), deuterons (d),  $^3\text{He}$ , and  $^4\text{He}$  particles as a function of the atomic number ( $Z_A$ ) of the target nuclide

$Z_A$	P	d	$^3\text{He}$	$^4\text{He}$	$Z_A$	p	d	$^3\text{He}$	$^4\text{He}$
4	1.5	1.5	3.1	3.2	46	8.3	8.0	15.6	15.4
6	2.0	2.0	4.1	4.2	48	8.5	8.2	16.1	15.8
8	2.5	2.4	4.9	5.0	50	8.7	8.4	16.5	16.2
10	2.9	2.8	5.6	5.7	52	8.9	8.6	16.8	16.5
12	3.3	3.2	6.3	6.4	54	9.1	8.8	17.3	17.0
14	3.7	3.6	7.1	7.1	56	9.4	9.0	17.7	17.4
16	4.0	3.9	7.7	7.7	58	9.6	9.3	18.2	18.0
18	4.3	4.1	8.1	8.1	60	9.9	9.6	18.7	18.4
20	4.7	4.6	9.0	9.0	62	10.1	9.8	19.1	18.8
22	5.0	4.8	9.4	9.3	64	10.3	9.9	19.5	19.2
24	5.3	5.1	10.0	9.9	66	10.5	10.2	19.9	19.6
26	5.6	5.4	10.6	10.5	68	10.7	10.4	20.3	20.0
28	6.0	5.7	11.3	11.2	70	10.9	10.6	20.7	20.4
30	6.2	6.0	11.7	11.6	72	11.2	10.8	21.1	20.8
32	6.4	6.2	12.1	12.0	74	11.4	11.0	21.6	21.2
34	6.6	6.4	12.6	12.4	76	11.6	11.2	21.9	21.6
36	6.9	6.7	13.1	12.9	78	11.8	11.4	22.3	22.0
38	7.2	7.0	13.6	13.5	80	12.0	11.6	22.7	22.4
40	7.5	7.2	14.2	14.0	82	12.2	11.8	23.1	22.8
42	7.8	7.5	14.7	14.5	90	12.9	12.5	24.5	24.2
44	8.0	7.7	15.2	15.0	92	13.1	12.7	24.9	24.6

curve analysis. The former is possible if the energies of the  $\gamma$ -lines are more different than the energy resolution of the  $\gamma$ -spectrometer used (typically 2 keV for an HPGe detector). The latter is possible if the half-lives of the radionuclides C and D are different. For considerably different half-lives, a sufficiently short (long) decay time makes it possible to measure the short- (long-) lived radionuclide selectively.

3. Chemical separation of B from D. This is only possible if B and D are not radioisotopes of the same element.

A special case is spectral interference from the matrix. CPAA being primarily a method for trace element determinations, the concentration of the analyte element is much lower (typically  $10^6$  times) than the concentration of the major matrix element(s). If a major matrix element (C) is activated, it often happens that the activity of D is much higher (also  $10^6$  times) than that of B. Even for quite different  $\gamma$ -energies or half-lives, selective measurement becomes quite difficult. The interference can only be avoided by a proper choice of the incident energy or by chemical separation of B from D. In the former case an instrumental analysis (i.e., without chemical separation) is possible.

## Standardization

‘Standardization’ matches the concept of ‘calibration’ generally used in analytical chemistry. The number of radionuclides produced per unit time by irradiation of an infinitesimally thin target is determined by the balance of the increase due to nuclear reactions and the decrease due to radioactive decay:

$$\frac{dN}{dt} = I \cdot \sigma \cdot n \cdot dl - \lambda \cdot N \quad [10]$$

where  $N$  is the number of radionuclides for the nuclear reaction considered,  $t$  is the time (s),  $I$  is the number of charged particles per unit time, or beam intensity ( $s^{-1}$ ),  $\sigma$  is the (partial) nuclear reaction cross-section for the nuclear reaction considered ( $cm^2$ ),  $n$  is the number of target nuclides for the nuclear reaction considered per unit volume ( $cm^{-3}$ ),  $l$  is the thickness of the target (cm), and  $\lambda$  is the decay constant of the radionuclide formed for the nuclear reaction considered ( $s^{-1}$ ).

The number of radionuclides formed after an irradiation time  $t_i$  is found by integration of eqn [10]. On condition that the beam intensity  $I$  is constant during irradiation, the activity is then

$$A = \lambda \cdot N = I \cdot \sigma \cdot n \cdot dl \cdot (1 - \exp(-\lambda t_i)) \quad [11]$$

For a ‘thick’ target, i.e., thicker than the range  $R$  of the charged particles used, the activity is obtained by integration of eqn [11] between 0 and  $R$ . It can be demonstrated that the beam intensity  $I$  is nearly constant as a function of the depth  $l$ . On condition that the analyte element is homogeneously distributed in the sample, one obtains

$$A = I \cdot n \cdot \rho^{-1} \cdot (1 - \exp(-\lambda t_i)) \int_0^R (\sigma \cdot \rho) dl \quad [12]$$

Using the stopping power  $S$  (eqn [1]) the integral in eqn [12] can be expressed as a function of the energy,

$$\int_0^R (\sigma \cdot \rho) dl = \int_{E_t}^{E_i} \left( \frac{\sigma}{S} \right) dE \quad [13]$$

where  $E_i$  is the incident energy and the integration limit  $E = 0$  (corresponding to  $l = R$ ) may be replaced by the threshold energy  $E_t$ , as the cross-section is zero for an energy below the threshold energy of the nuclear reaction.

Although, in principle, the concentration of the analyte element can be calculated by eqns [12] and [13], in CPAA a relative method is usually applied whereby a standard (s) and a sample (x) are irradiated separately but with the same incident energy  $E_i$ . From eqns [12] and [13], the concentration of the analyte element in the sample is then

$$c_x = c_s \cdot \frac{A_x \cdot I_s \cdot [1 - \exp(-\lambda \cdot t_{i_s})]}{A_s \cdot I_x \cdot [1 - \exp(-\lambda \cdot t_{i_x})]} \cdot F \quad [14]$$

$$F = \int_{E_t}^{E_i} \left( \frac{\sigma}{S_s} \right) dE / \int_{E_t}^{E_i} \left( \frac{\sigma}{S_x} \right) dE \quad [15]$$

The number of target nuclides per unit volume ( $n$ ) can be replaced by the concentration of the analyte element ( $c$ , e.g., in mg per kg), at least if the isotopic abundance of the target nuclide is equal for sample and standard (which is mostly the case).

The concentration (‘concentration’  $c$  matches the SI quantity mass fraction  $w$ ) of the analyte element in the sample ( $c_x$ ) can thus be calculated by eqn [14] from the concentration of the analyte element in the standard,  $c_s$ ; the ratio of the activity in sample and standard at the end of irradiation,  $A_x/A_s$ ; the ratio of the beam intensity for standard and sample,  $I_s/I_x$ ; the irradiation time,  $t_i$ ; the decay constant,  $\lambda = \ln 2/t_{1/2}$  where  $t_{1/2}$  is the half-life; and the  $F$ -factor (eqn [15]) representing the correction for the difference in stopping power of sample and standard.

The  $F$ -factor can be calculated according to eqn [15] if (1) the stopping power of sample ( $S_x$ ) and standard ( $S_s$ ) and (2) the nuclear reaction cross-section ( $\sigma$ ) are known, both as a function of energy in the energy interval between the threshold energy ( $E_t$ )

and the incident energy ( $E_i$ ). The first condition can be fulfilled in a large number of cases. If the major element concentration of the sample is known and if the matrix does not change during irradiation, the stopping power of the sample can be calculated. It is obvious that the stopping power of the standard can be calculated. If the second condition is also fulfilled, the most obvious method for calculation of the  $F$ -factor is numerical integration of eqn [15].

As the second condition is not always fulfilled, two approximative standardization methods have been proposed that do not require knowledge of the nuclear reaction cross-section. The first approximative method makes use of stopping power data for sample and standard,

$$F \approx S_x(E_M)/S_s(E_M) \quad [16]$$

at an energy  $E_M$ , i.e., the average of incident and threshold energy,

$$E_M = (E_i + E_t)/2 \quad [17]$$

The second approximative method makes use of range data for sample and standard at both the incident and threshold energies:

$$F \approx \frac{R_s(E_i) - R_s(E_t)}{R_x(E_i) - R_x(E_t)} \quad [18]$$

These two approximative methods are equivalent and yield accurate results if one or more of the following conditions are fulfilled: (1) the atomic number of (the elemental components of) sample and standard are comparable; (2) the threshold energy of the nuclear reaction used is high; or (3) the incident energy is high as compared with the threshold energy.

The average stopping power method requires knowledge of the cross-section and is in principle also an approximative method, but with a negligible systematic error compared with the previous two methods. The  $F$ -factor is calculated using stopping power data for sample and standard,

$$F \approx S_x(E_m)/S_s(E_m) \quad [19]$$

at an 'average energy'  $E_m$  calculated as

$$E_m = \int_0^{E_i} (E \cdot \sigma) dE / \int_0^{E_i} \sigma \cdot dE \quad [20]$$

Up to now, all methods have required knowledge of the stopping power of the sample, which can be calculated for all elements and for any mixture or compound with known composition. Many samples, however, have a very complex matrix composition that is not always accurately known. Irradiation with

charged particles can cause substantial heating in materials of low thermal conductivity, e.g., silicates, such as environmental and geological materials. This can give rise to evaporation of, for example, the organic fraction present in the sample. The introduction of systematic positive errors in quantitative analysis is to be feared from enrichment of the analyte element during irradiation and the fact that the stopping power calculated for the sample is too high. Indeed, the stopping power of a material, taking into account the organic fraction also, is higher than the stopping power of the inorganic fraction of that material (the stopping power of hydrogen for protons is about three times higher than the stopping power of, for example, silicon dioxide). If the organic fraction (or part of it) evaporates during irradiation, a too high stopping power is used in the calculations and consequently a too high concentration is obtained.

Consequently, an internal standard method has been proposed for materials with unknown or variable matrix composition. The idea is that by using the known concentration of one element (i.e., the internal standard) and by measuring the induced activity of the internal standard, a correction can be calculated for all elements to be determined (i.e., the analyte elements) without knowledge of the matrix composition of the sample. Approximative methods not requiring knowledge of the reaction cross-section have also been proposed.

The 'two reactions' method is derived from the average stopping power method and also makes use of an internal standard of known concentration. The main advantage is that no stopping power data at all are necessary, so that the accuracy of the analytical method is not influenced by the accuracy of the stopping power data.

So far, a relative method has been applied, whereby a standard and a sample are irradiated separately but with the same incident energy. In the standard addition method, however, a sample is irradiated, and separately the same sample to which an accurately known amount of the analyte element is added. As both matrices are almost identical (only trace amounts of analyte are added), the  $F$ -factor is unity and consequently the method does not make use of any stopping power data.

## Experimental

### Irradiation

For CPAA charged particles have to be accelerated up to an energy higher than the Coulomb barrier (to obtain high reaction cross-section and thus detection limit) and lower than the threshold energy of

reactions more complex than  $(x,n)$  or  $(x,\alpha)$  reactions, where  $x$  is the charged particle (to avoid nuclear interferences). Energies between a few MeV and 20 MeV should be available for protons and deuterons and up to 30 MeV for helium-3 or helium-4 particles. Accelerators that cover this energy range of interest for CPAA are the isochronous cyclotron, the 'tandem Van de Graaf' accelerator or the linear accelerator (linac). An isochronous cyclotron with variable energy is generally used.

Irradiation with a charged particle beam can cause substantial heating of the sample. Therefore, efficient cooling of the target is the major concern in target design. The simplest target design is possible for solid samples with good thermal conductivity, not powdered and available as a foil or sheet with a thickness slightly larger than the range. Irradiation under vacuum is then possible with the sample mounted on a water-cooled sample holder. For powdered samples or samples with poor thermal conductivity, irradiation in a helium atmosphere has been developed. Applications of CPAA for the analysis of liquids are less obvious, as many optical (AAS, OES) or mass spectrometric (ICP-MS) methods of analysis exist for diluted aqueous solutions.

When a relative method is used in CPAA, the ratio of the beam intensity for standard and sample has to be determined experimentally. As knowledge of the absolute beam intensity is not required, it is common practice to cover sample and standard with a thin (much less than the range) I-monitor foil. The induced activity in the I-monitor foil is then a measure of the beam intensity. Pure metal foils are the obvious choice: they are good thermal conductors, monoelemental, and available in different thicknesses.

To avoid recoil nuclides from the I-monitor foil in the sample (or the standard) a recoil foil is inserted between the I-monitor foil and the sample (or the standard). As the range of the recoil nuclides is much lower than the range of the charged particles commonly used in CPAA, a few micrometers of aluminum foil is sufficient to stop the recoil nuclides completely, while the energy of the charged particles is reduced by a minor (but not negligible) fraction.

### **Chemical Etch**

Although charged particles have a limited range in matter, CPAA is considered as an analytical method for determination of the concentration in the bulk of a sample rather than at the surface. Accordingly, interference from the analyte element at the surface has to be avoided. This is possible for solid samples that are not powdered simply by etching the sample before irradiation, at least if further contamination of

the surface can be avoided until the end of irradiation (contamination after irradiation gives no interference). For the determination of light elements (boron, carbon, nitrogen, and oxygen) at low concentrations, however, the latter condition cannot always be fulfilled. This is the case, for example, for the determination of bulk oxygen concentration in aluminum. A freshly etched aluminum sample is exposed to air again and has an aluminum oxide layer that will also be activated. Chemical etch after irradiation is then the appropriate method. A few micrometers of the activated surface will be removed.

### **Radiochemical Separation**

Instrumental analysis is possible if spectral interference can be avoided by a proper choice of the incident energy or the measuring conditions (see Interferences). If not, the radionuclide B, formed from the analyte element A, has to be separated radiochemically from interfering radionuclide(s) D, formed out of interfering element(s) C. The latter case is called radiochemical analysis, in contrast to instrumental analysis for the former. This section deals with some major differences between radiochemical separation and common chemical separation, used for non-nuclear methods of analysis.

For most charged particle-induced reactions the atomic number of the radionuclide B is different from that of the analyte element A. This is the case for  $(p,n)$ ,  $(p,\alpha)$ ,  $(d,n)$ ,  $(d,\alpha)$ ,  $(^3\text{He},n)$ ,  $(^3\text{He},d)$ ,  $(\alpha,n)$ , and  $(\alpha,d)$  reactions, but not for  $(p,d)$  and  $(^3\text{He},\alpha)$  reactions. The radiochemical separation to be developed for CPAA is thus different from that for all non-nuclear analytical methods and for some other methods based on activation analysis, such as thermal- and fast-NAA (using the  $(n,\gamma)$  and  $(n,2n)$  reactions, respectively) and PAA (using the  $(\gamma,n)$  reaction). Also, in principle, it is not necessary to separate the matrix, but rather the radionuclide(s) formed out of the matrix element(s). Again, the atomic number of the radionuclide(s) is usually different from that of the matrix element(s). Owing to the chemical separation involved, CPAA is considered to be an independent analytical method, not subject to the same systematic errors as other analytical methods.

A radiochemical separation has three important advantages compared with a common chemical separation: (1) Inactive carriers can be added for the elements to be separated (B and D). This avoids the difficulties of a chemical separation at the trace level. (2) Reagent impurities (or blanks) do not influence the detection limit capabilities of the analytical method. (3) Separations may not be quantitative and even not reproducible (see below).

In the choice and development of a radiochemical separation the following seven points have to be considered.

1. *Selectivity of the separation.* As already pointed out (see Interferences), the induced matrix activity (D) can be up to  $10^6$  times higher than the activity to be measured (B). It is not possible to resolve such a low activity in the presence of such a high activity. In such unfavorable cases the decontamination factor (ratio of D before to after separation) should be  $\sim 10^4$  or better.
2. *Quantitativeness of the separation.* A separation that provides quantitative recovery of B is to be preferred. If the separation is not quantitative and not even reproducible, determination of the yield for each individual separation is possible. Two pathways can be followed: addition before separation of an accurately known amount of inactive (active) carrier and measurement of the mass (activity) of the carrier after separation.
3. *Speed of the separation.* To obtain an optimum detection limit, the time needed to perform the whole separation procedure should not exceed a few half-lives of the radionuclide to be measured.
4. *Use of inactive carrier.* It is not only an advantage but usually also a necessity to add an inactive carrier of the elements to be separated. Indeed, if the atomic number of the radionuclide B is different from that of the analyte element A (which is mostly the case, see above), this radionuclide B is produced 'carrier free'. The number of radionuclides (N) can be calculated from the activity (A) and the half-life ( $t_{1/2}$ ):

$$N = A \cdot t_{1/2} / \ln 2 \quad [21]$$

For  $A = 1 \text{ kBq}$  and  $t_{1/2} = 1 \text{ day}$ , the amount of substance is  $10^{-16} \text{ mol}$ . It is clear that, if there is no accidental addition of inactive element B (e.g., impurity in the matrix or reagents), it can be very difficult to separate chemically such a low mass.

5. *Detection efficiency of the measurement.* To obtain the optimum detection limit, the activity should be measured with the highest possible detection efficiency. The latter is also determined by the geometry (size and distance from the detector) of the source. Separation procedures ending with large volumes of solution are less favorable than those ending with a precipitate.
6. *Self-absorption of the source.* Self-absorption of the source to be measured is, in principle, not a problem, as long as the absorption is identical for all samples and standards. Self-absorption can be important for low-energy  $\gamma$ -rays or sources containing components of high atomic number.

7. *Health risk.* Health risk is quantified as dose expressed in the SI unit sievert ( $1 \text{ Sv} = 1 \text{ J kg}^{-1}$ ). The dose equals the amount of absorbed radiation energy per unit mass, multiplied by a weighing factor for radiation  $w_r$ , to take into account the biological effects of different radiations (for  $\gamma$ -radiation,  $w_r = 1$ ). The maximum permissible dose for occupational workers is 20 mSv per year. Natural sources provide 2–6 mSv per year. The activity level required for precise  $\gamma$ -measurements (kilobecquerels) provides no health risk for external radiation ( $\text{nSv h}^{-1}$  at 30 cm). The induced matrix activity, however, may be much higher. It is good practice first to separate the main matrix activity, choosing a procedure that limits manual intervention to the strict minimum. Ion exchange chromatography, for example, is a better choice than (manual) solvent extraction. Ingestion of radioactive material is to be avoided by appropriate laboratory practice.

The experimental development of a radiochemical separation can be supported by the use of radiotracers. The chemical separation is simulated step by step with addition of radiotracers (and inactive tracers also) for the element to be separated (B) and for the interfering element(s) (D), in order to check quantitiveness and selectivity, respectively. The tracer has to be brought to the same chemical form as the radionuclide to be separated. This is often not possible for the dissolution step preceding the actual separation procedure. Therefore, the dissolution procedure selected should guarantee quantitative recovery of the radionuclide to be measured.

There is no need to apply the radiochemical separation to the standards, at least if monoelemental standards are used (i.e., the pure analyte element or a compound of it). However, standards and samples have to be measured with identical detection efficiency. Therefore, the standards are brought into exactly the same geometrical form as the samples after separation. If the last step of a separation is ion chromatography, the standards are dissolved and the total volume is brought exactly to the volume of the eluate. Possible differences in self-absorption have to be considered (e.g., because the acid used for the elution differs from the one used to dissolve the standard).

### Measurement of Radioactivity

Charged particle-induced reactions, such as (p,n) and (p, $\alpha$ ) reactions, yield radionuclides that are too rich in protons and consequently decay by positron ( $\beta^+$ ) emission or electron capture (EC), mostly followed by gamma ( $\gamma$ ) emission. Positrons are emitted with



energy between zero and a characteristic maximum energy. Electron capture involves measurement of X-rays or Auger electrons and is of no importance for CPAA.  $\gamma$ -Rays are monoenergetic. A radionuclide (and the analyte element the radionuclide is formed from) can be identified by its characteristic  $\gamma$ -photon energy (qualitative analysis). The (radio)activity measured (i.e., the number of positrons or  $\gamma$ -photons measured per second) provides quantitative information. A limited number of radionuclides are pure positron emitters. Positron emitters are measured indirectly by the annihilation radiation: a positron loses its kinetic energy, it annihilates with an electron (i.e., both electron masses are converted into energy) and two  $\gamma$ -photons are emitted in opposite directions. The energy of both  $\gamma$ -photons is 511 keV, i.e., the rest mass of the electron. It is clear that information about the characteristic maximum energy is lost during the annihilation process. A positron emitter is then to be identified by its half-life. It is clear that  $\gamma$ -spectrometry is the method of choice, except for pure positron emitters.

## Applications

The most obvious application of CPAA is the determination of traces of light elements such as carbon, nitrogen, and oxygen in the bulk of very pure metals and semiconductors. The analytical methods routinely used are combustion in an oxygen stream and measurement of the carbon dioxide formed (for carbon), the Kjeldahl method (for nitrogen), and reducing fusion (for nitrogen and oxygen). The detection limit is of the order of mg per kg and systematic errors are to be feared because of the effects of surface contamination, reagent blanks, and non-quantitative recovery of the gases. Activation analysis is not subjected to the same kind of systematic errors. Surface contamination can be avoided by chemical etch after irradiation (see Experimental); reagent blank errors do not occur; if instrumental analysis is not possible, the yield of a radiochemical

separation can be determined, even if not reproducible (see Experimental). Because of unfavorable nuclear characteristics, thermal neutron activation cannot be applied for the determination of these light elements. Activation analysis with fast neutrons or photons has lower sensitivity (higher detection limit) than CPAA.

Nuclear data for the determination of boron, carbon, nitrogen, and oxygen by CPAA are summarized in Table 4. For short-lived radionuclides such as oxygen-14 and oxygen-15, only instrumental analysis is possible. For longer-lived radionuclides, a chemical separation can be applied if necessary. Beryllium can be separated by anion chromatography; carbon by combustion in an oxygen stream or dissolution in an oxidizing acid followed by trapping of the carbon dioxide formed; and nitrogen and fluorine by steam-distillation as ammonia and hexafluoro-silicate, respectively. Beryllium-7 and oxygen-14 are  $\gamma$ -emitters and qualitative and quantitative measurement is performed with a  $\gamma$ -spectrometer. The other listed nuclides are pure positron emitters to be identified by their half-lives. Nuclear interference cannot always be avoided. As the concentration of the interfering elements is generally lower than that of the analyte elements, this interference can usually be neglected or corrected for. As a monoelemental standard, either the pure element is used or a pure compound with well-known stoichiometry and good stability during irradiation. Correction is necessary for different stopping powers of sample and standard (see Standardization).

## CPAA for Surface Characterization

Although CPAA has been widely applied for bulk analysis, it is also applicable for surface characterization within strict limitations.

CPAA for surface characterization is based on irradiation of a 'thin' surface layer, i.e., the thickness of the surface layer ( $\sim$  nm to  $\mu$ m) is much lower than the range of the charged particles used ( $\sim$  mm). Moreover, the substrate (i.e., the layer on which the

**Table 4** Nuclear data for the determination of boron, carbon, nitrogen, and oxygen by CPAA

	Reaction	$E_t$ (MeV)	$t_{1/2}$	Radiation ( $E_\gamma$ , keV)	Nuclear interference	Standard
Boron	$^{10}\text{B}(p,\alpha)^7\text{Be}$	0	53.4 days	$\beta^+$ , $\gamma$ (478)	Li	B, $\text{H}_3\text{BO}_3$
	$^{10}\text{B}(d,\alpha n)^7\text{Be}$	1.3			Li	
	$^{10}\text{B}(d,n)^{11}\text{C}$	0	20.4 min	$\beta^+$	None	B
Carbon	$^{12}\text{C}(d,n)^{13}\text{N}$	0.3	9.97 min	$\beta^+$	None	C, polyethylene
Nitrogen	$^{14}\text{N}(p,n)^{14}\text{O}$	6.3	70.5 s	$\beta^+$ , $\gamma$ (2313)	None	$\text{NaNO}_3$ , nylon-6
	$^{14}\text{N}(p,\alpha)^{11}\text{C}$	3.1	20.4 min	$\beta^+$	B	
	$^{14}\text{N}(d,n)^{15}\text{O}$	0	2.05 min	$\beta^+$	None	
Oxygen	$^{16}\text{O}(^3\text{He},p)^{18}\text{F}$	0	109.8 min	$\beta^+$	F, Na	$\text{SiO}_2$
	$^{16}\text{O}(t,n)^{18}\text{F}$	0				None

surface layer is deposited) should not contain the analyte element(s). The goal is to determine the partial mass thickness, i.e., mass of the analyte element per surface unit, e.g. mg Ti per  $\text{cm}^2$  for a  $\text{TiO}_2$  surface layer. The induced activity is proportional to the mass thickness as it equals  $c_x \times$  the denominator in eqn [18],  $E_t$  being replaced by  $E_o$ , i.e., the outgoing energy, slightly lower than  $E_t$ .

The partial mass thickness(es) determined can be used to calculate (1) the elemental composition of a surface layer or (2) the total mass thickness knowing the stoichiometry of the surface layer. Moreover, the thickness in nm or  $\mu\text{m}$  can be calculated knowing the density of the surface layer material.

Examples are the determination of (1) the Cu/Y and Ba/Y ratios in YBCO ( $\text{YBa}_2\text{Cu}_3\text{O}_{6+\delta}$ ), which strongly influence the resistivity of this high-temperature super conductor and (2) the mass thickness of Al,  $\text{Al}_2\text{O}_3$ , or  $\text{TiO}_2$  layers on PET foil, assuming the stoichiometry mentioned, which improve the water- and airproof characteristics of this packing material.

## Analytical Capability

CPAA is an analytical method for the determination of trace elements in the bulk of a sample. CPAA can also be applied at higher concentration levels, but then other, more common analytical methods can yield even better results more conveniently. Determination of a specific compound or the oxidation state of an element (i.e., speciation) is excluded. CPAA is considered to analyze the bulk of a sample, as the activated depth (slightly less than the range) is between 0.1 and 1 mm and because surface contamination can be avoided by chemical etch after irradiation. The sample mass analyzed depends on the activated depth and the beam diameter (typically 0.5–2 cm).

No general statement can be made about the elements that can be determined and the samples that can be analyzed, because these depend on the nuclear characteristics of the target nuclide (isotopic abundance), the nuclear reaction (cross-section and related parameters such as threshold energy and Coulomb barrier), and the radionuclide induced (half-life, radiation emitted, energy, and its intensity) for the analyte element, the possible interfering elements and the major components of the sample. CPAA can solve a number of important analytical problems in material science (e.g., determination of boron, carbon, nitrogen, and oxygen impurities in very pure materials such as copper or silicon) and environmental science (e.g., determination of the toxic elements cadmium, thallium, and lead in solid environmental samples). As these problems cannot be solved by NAA, CPAA and NAA are complementary to each other.

CPAA can easily be applied to solid, massive samples with good thermal conductivity and available as foil or sheet slightly thicker than the range (e.g., metals). Special precautions have to be taken for powdered materials and poor thermal conductors. Analysis of liquids and gases is inconvenient.

If spectral interferences can be avoided (multielemental), instrumental analysis is possible (i.e., without radiochemical separation between irradiation of the sample and measurement of the induced activity). In general, an instrumental analysis is possible for analyte elements with low atomic number in a matrix with high atomic number, because the charged particle energy can be chosen just below the Coulomb barrier for the matrix element.

Detection limits down to the  $\mu\text{g}$  per kg level have been attained for the most favorable instrumental analyses (e.g., carbon and nitrogen in molybdenum and tungsten) and for radiochemical analyses (e.g., cadmium and thallium in zinc) at least if no nuclear interferences occur. This is below the concentration levels at which these impurities influence the material characteristics and below the detection limit attainable by more common methods of analysis. A precision (reproducibility) of a few percent is possible at the mg per kg concentration level in the most favorable cases. However, at higher concentration levels the precision will not improve significantly. Many systematic errors can be checked experimentally (e.g., interferences, yield of a radiochemical separation); others can be avoided experimentally (e.g., surface contamination). Systematic errors due to reagent blanks do not arise.

To apply CPAA one needs access to a cyclotron, a radiochemical laboratory, and  $\gamma$ -spectrometry and  $\beta^+$ -counting equipment. The time needed for one analysis is determined by the half-life of the radionuclide induced and whether a radiochemical separation is necessary or not. The inherent complexity and costs are the major drawbacks of CPAA.

CPAA is an independent analytical method (with respect even to NAA or PAA) that is not subject to the same potential systematic errors as optical (AAS, OES) or mass-spectrometric (ICP-MS) analytical methods, which often require sample dissolution prior to measurement (CPAA does not require sample dissolution prior to irradiation). Therefore, CPAA can provide a significant contribution to certification analyses of reference materials (e.g., BCR, NIST).

**See also:** Activation Analysis: Neutron Activation; Photon Activation. **Quality Assurance:** Reference Materials; Production of Reference Materials. **Radiochemical Methods:** Radiotracers.

## Further Reading

- Blondiaux G, Debrun JL, and Maggiore CJ (1995) Charged particle activation analysis. In: Tesmer JR, Nastasi M, Barbour JC, Maggiore CJ, and Mayer JW (eds.) *Handbook of Modern Ion Beam Analysis*. Pittsburgh: Material Research Society.
- Casteleyn K, Strijckmans K, and Dams R (1992) Internal standardization in charged particle activation analysis. *Nuclear Instruments and Methods B* 68: 161–165.
- De Neve K, Strijckmans K, and Dams R (2000) Feasibility study on the characterisation of thin layers by charged particle-activation analysis. *Analytical Chemistry* 72: 2814–2820.
- Strijckmans K (1997) Charged-particle activation analysis. In: Brune D, Hellborg R, Whitlow HJ, and Hunderi O (eds.) *Surface Characterisation: A User's Sourcebook*, pp. 169–175. Weinheim: Wiley-VCH.
- Strijckmans K (2000) Charged particle activation analysis. In: Meyers RA (ed.) *Encyclopedia of Analytical Chemistry*, pp. 12424–12447. New York: Wiley.
- Strijckmans K (2001) The isochronous cyclotron: principles and recent developments. *Computerized Medical Imaging and Graphics* 25: 69–78.
- Strijckmans K (2003) *Nuclear Data for Charged Particle Activation*, <http://allserv.UGent.be/~kstrykmn/nuclear/>.
- Vandecasteele C and Strijckmans K (1980) Standardization in charged particle activation analysis. *Journal of Radioanalytical Chemistry* 57: 121–136.

## Photon Activation

**D Schulze**, Sächsisches Institut für gebietliche Umwelt- und Sanierungstechnologieforschung i.G., Leipzig, Germany

**C Segebade**, Bundesanstalt für Materialforschung und -prüfung (BAM), Berlin, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 25–32, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

## Introduction

Photon activation analysis is based upon the reaction of nuclei with high-energy photons such as  $\gamma$ -rays or X-rays. Analogously to neutron activation analysis — which is better known and much more frequently applied than photon activation analysis — the reaction products are normally radioactive. The induced radioactivity, preferably the emitted gamma radiation and characteristic X-rays, is measured with appropriate spectrometers, in basically the same procedure as for neutron activation analysis and activation analysis with charged particles. One difference between the techniques relates to the different natures of the respective product nuclides; after photon activation it is advantageous to employ, in addition to classical gamma-ray spectrometry, low-energy photon spectroscopy using special semiconductor diodes designed for measurement of ‘soft’ photons. This will be explained further in the context of radiation measurement below.

Another difference is in the yield distribution between the product activities: during exposure to thermal neutrons a single reaction type occurs almost

exclusively, namely ( $n, \gamma$ ). In activation with high-energy (say 30 MeV) bremsstrahlung photons several reaction types can occur, the ( $\gamma, n$ )-type being favored, whereas during charged particle bombardment several reaction types can be induced with comparable activity yields.

Data processing is performed in the same way as for neutron activation analysis, although the spectra of photon activated samples look slightly different from those resulting from after neutron or charged-particle exposure.

Photon activation analysis can be regarded as a complementary technique to thermal neutron activation analysis, with some advantages and several drawbacks. Particularly advantageous features of photon activation analysis as compared to neutron activation include:

- The activating radiation source (usually a linear or circular electron accelerator) is smaller, less expensive, and easier to operate than a nuclear reactor normally used for neutron activation.
- Instrumental multielement analyses can be performed, nondestructively in advantageous cases (i.e., large objects).
- Large objects – up to above the size of the human body – can be analyzed nondestructively, where either bulk analyses comprising the entire object or analyses of selected limited spots can be performed.
- Elements can be analyzed with high detection power that cannot be detected by neutron activation, e.g., light elements (deuterium, beryllium, carbon, nitrogen, oxygen, fluorine, silicon, phosphorus) or elements of environmental interest (nickel, thallium, lead, bismuth).

- Matrix problems frequently encountered in neutron activation analysis (excessive matrix activity, inhomogeneous activation due to large absorption cross-sections, etc.) are significantly reduced.

One disadvantage of photon activation analysis is the comparatively poor detection power for several elements, e.g., sodium, vanadium, cobalt, the rare-earth elements. Thus, ultratrace determinations of these elements are barely possible using photon activation. As in other instrumental analytical techniques, photon activation analysis is a relative quantification method and hence needs reference materials with known compositions. (Reference materials are dealt with below.)

A description of photon activation analysis including historical development can be found in the literature (see Further Reading section).

## History

The first photonuclear activation for analytical purposes was performed with radionuclides as the activating radiation source. These applications were reported in the early 1950s, although apparently the first beryllium determinations by photodisintegration were performed in the late 1930s in the Soviet Union. The analytical detection power of photon activation analysis using radionuclide sources is poor and restricted to the analysis of deuterium, beryllium, several fissile nuclides, and a few nuclides that have low-lying isomeric states. Nonetheless, nuclide excitation is still in use.

Later, small static accelerators were used for the analysis of these elements. With the advent of high-energy cyclic electron accelerators (betatron, microtron, linac), the high-energy bremsstrahlung produced by these machines has been used for photon activation. As a result, the list of determinable elements has increased dramatically and, with a few exceptions, now covers the entire periodic table. Some of the light elements (carbon, nitrogen, oxygen) were analyzed in the early 1950s using photonuclear reactions induced by bremsstrahlung from a betatron. A chemical separation is normally involved in the analysis of these elements. A great deal of fundamental work was devoted to this problem during the 1960s (see later).

Along with the maturity of gamma spectroscopy equipment, the extension of the method to heavier elements, as happened in the 1960s, was straightforward.

## Photonuclear Reactions

Shortly after the beginning of photonuclear research in 1934, photoreactions were exploited for analytical purposes as well as for radionuclide production.

### Absorption of Photons by Nuclei

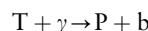
Nuclear reactions induced by electromagnetic radiation can be described – at least for sufficiently low-photon energies – in terms of a two-stage process. Absorption of a photon leads to an intermediate highly excited state of the nucleus. The excitation energy of this so-called compound nucleus can then be released by emission of photons, neutrons, or charged particles.

### De-Excitation of the Nucleus after Absorption of a Photon

The excitation energy may be released by the following processes: (1) Re-emission of a photon with the same energy as the incident photon. This process is called elastic scattering or  $(\gamma, \gamma)$ -reaction. (2) Emission of photons with lower energy. This type of reaction is known as inelastic scattering or  $(\gamma, \gamma')$ -reaction. (3) Emission of neutrons, protons, or composite particles if the excitation energy exceeds the particle separation threshold.

### Photonuclear Reactions Used for Activation Analysis

According to the processes described above, a photonuclear reaction is described by



or symbolically, according to convention,



where T is the target nucleus, P is the activation product, and b is the promptly emitted particle, usually  $\gamma$  or n-type being the most relevant for analytical application. At higher incident photon energies,  $(\gamma, 2n)$ ,  $(\gamma, 3n)$ ,  $(\gamma, p)$ ,  $(\gamma, \alpha)$ , and other reactions can also be induced during photon bombardment. However, these reactions mostly lead to undesirable interference and thus should be minimized through proper selection of the activating energy. In daily laboratory practice, 30 MeV bremsstrahlung has proved optimal for photon activation analysis.

As mentioned above, the first photoactivations were performed with radionuclide sources. Consequently, analytical application was then restricted to nuclides with low neutron binding energies (photodisintegration of  $^2\text{H}$ ,  $^9\text{Be}$ ); nuclides that undergo low-energy photofission (thorium, uranium), being detected by prompt neutron counting; and nuclides that have sufficiently long-lived isomeric states to be counted with some delay after activation.

With the help of high-energy accelerators, photon activation was soon extended to the analysis of

carbon, nitrogen, oxygen, and fluorine, all these being elements with nuclear properties unfavorable for neutron activation analysis. These require chemical separation after photon exposure since their activation products are pure positron emitters. Activation products of heavier elements mostly emit characteristic  $\gamma$ -ray spectra, hence they can be analyzed instrumentally, particularly since high-resolution semiconductor photon spectrometers are now generally available. In some cases, however, a radiochemical separation step must also be included in the analytical procedure for heavier elements, e.g., when the desired analytical signal is interfered with by signals of other components, or in the case of excessive matrix activity background. In this case it can be of advantage to use scintillation spectrometry.

As mentioned earlier, the use of low-energy photon spectrometry is frequently advisable. Among the activation products resulting from photon exposure, neutron-deficient nuclides predominate via the most probable reaction types (see eqn [1])—( $\gamma, n$ ) in particular. While most product nuclides with low atomic numbers decay by positron emission, the electron-capture decay mode dominates among activation products of medium to high  $Z$ , resulting in emission of intense characteristic X-ray spectra. These can be measured conveniently with low-energy photon detectors (see below, Photon spectrometry). In many cases low-energy photon spectrometry offers several advantages when compared to classical  $\gamma$ -measurement.

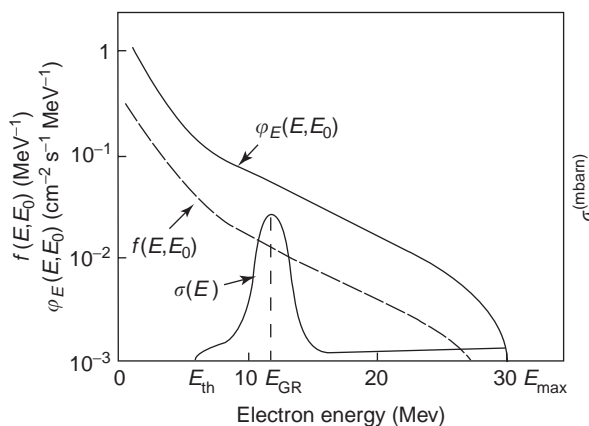
### Induced Activity and Analytical Sensitivity

The general radioactivation eqn [2] is

$$A = \frac{mN_A h}{A_r} \cdot \left( \int_{E_{th}}^{E_{max}} \sigma(E) \phi_E(E) dE \right) (1 - e^{-\lambda T_i}) e^{-\lambda T_D} \quad [2]$$

where  $A$  is the induced activity,  $m$  the target mass,  $N_A$  Avogadro's constant,  $h$  the isotopic abundance,  $A_r$  the relative atomic mass,  $E_{th}$  the threshold energy,  $E_{max}$  the maximum energy of the bremsstrahlung continuum (equals incident electron energy),  $\sigma(E)$  the energy-differential activation cross-section,  $\phi_E(E)$  the energy-differential photon flux density,  $\lambda$  the decay constant,  $T_i$  the exposure period, and  $T_D$  the decay time.

The achievable product activity depends on the photon energy and flux density of the activating source (Figure 1). Generally, the difference in detection limits among the elements is nowhere near as large as in neutron activation analysis. Typically, the detection limits, assuming purely instrumental analysis, lie between 0.001 and 1  $\mu\text{g}$ , whereas in thermal



**Figure 1** Schematic representation of bremsstrahlung flux ( $\phi(E)$ ), normalized bremsstrahlung continuum ( $f(E)$ ), and photo-cross-section ( $\sigma(E)$ ).

**Table 1** Achievable absolute detection limits in instrumental photon activation analysis

Element	Detection limit ( $\mu\text{g}$ )
Carbon	0.02
Nitrogen	0.02
Oxygen	0.05
Fluorine	0.001
Calcium	0.5
Titanium	0.005
Nickel	0.06
Rubidium	0.2
Zirconium	0.03
Cadmium	0.02
Tin	0.2
Barium	0.1
Samarium	0.005
Gold	0.005
Thallium	0.04
Lead	0.1
Bismuth	10.0
Uranium	0.001

neutron activation analysis the range covers many orders of magnitude. The practical detection limit is often restricted by high matrix radiation background. Furthermore, the large variety of reaction types that may be induced means the analyst has to select carefully the appropriate incident energy for the given task. This is a compromise between maximum analytical activity and minimum interfering activity. For example,

Analytical reaction:  $^{59}\text{Co}(\gamma, n) ^{58}\text{Co}$

Interfering reaction:  $^{63}\text{Cu}(\gamma, xn) ^{58}\text{Co}$

Table 1 lists the achievable analytical detection limits of several elements.

## Radiation Sources

In radionuclide photoactivation,  $^{124}\text{Sb}$   $\gamma$ -ray sources have mostly been used, but others have also been reported, e.g.,  $^{24}\text{Na}$  or  $^{60}\text{Co}$ . The use of isotopic sources for photon activation analysis is of limited value and is restricted to a few appropriate cases.

## Electron Accelerators

The disadvantages of isotope sources mentioned above can be circumvented by using bremsstrahlung for photoactivation. Accelerator-produced high-energy electrons are directed on to a heavy-metal target, preferably tantalum, tungsten, gold, or platinum. In this target, high-energy bremsstrahlung is produced as the electrons are decelerated in the Coulomb fields of the target nuclei.

The achievable photon flux densities usually exceed those of radionuclide sources by orders of magnitude. Moreover, the effective cross-section is significantly enlarged since the bremsstrahlung is continuous with electron energy (see **Figure 1**). Finally, much higher photon energies are produced than are obtainable with any isotope source. Hence, photonuclear reactions can be induced with bremsstrahlung, whereas  $\gamma$ -rays, with a few exceptions, can achieve only isomeric-state excitation.

One has to distinguish between linear and circular electron accelerators according to the geometry of the particle trajectories. The different accelerators are further defined by their operation mode, namely static and cyclic devices. In static accelerators electrons are accelerated by a constant high-voltage potential. The maximum achievable particle energy is directly dependent upon the maximum high voltage of the individual machine. Among static machines, only van de Graaff accelerators sometimes have been used for photon activation.

In cyclic accelerators (linacs, betatrons, microtrons) electron energies are achieved by multiple application of comparatively low voltage to the electrons. The maximum achievable energy is dependent on various parameters. Practical experience suggests that accelerators producing energies higher than, say, 50 MeV are unnecessary. Moreover, as explained above, these energies lead to frequent interfering reactions. Analytical requirements are best met by machines that provide  $\sim 30$  MeV bremsstrahlung at an average electron beam current of at least  $100\text{ }\mu\text{A}$ . The output energy should be freely selectable so that interfering higher-order reactions can be discarded by adjustment of the incident energy below the respective threshold energy.

Of these accelerators, two types have mostly been used for photon activation analysis, namely the linear accelerator (also called linac) and the microtron. These and other accelerators will not be described in detail here since normally the analyst is engaged in the sample handling, activity measurement, and data processing rather than in the operation of the radiation sources, which is usually done by separate operating personnel.

## Activation with Photoneutrons

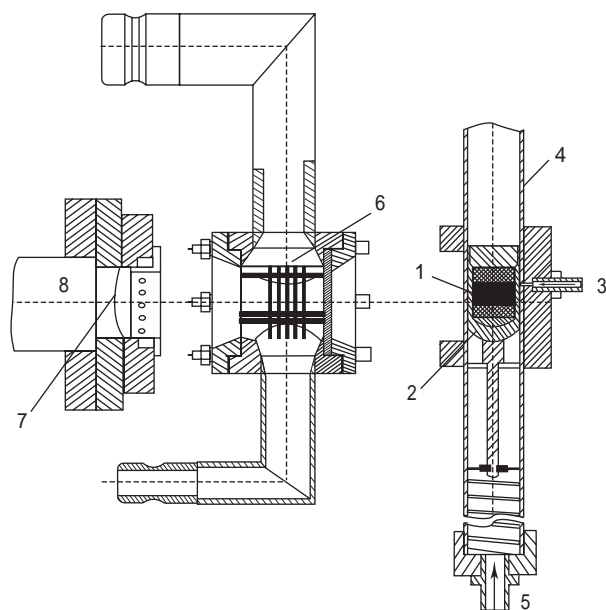
Photoneutrons are produced at considerable flux densities in the heavy metal of the bremsstrahlung converter. For example, in the linear accelerator used by the authors, a thermal neutron flux density of some  $10^{10}\text{ cm}^{-2}\text{ s}^{-1}$  has been produced in the tantalum converter during normal operation (30 MeV,  $150\text{ }\mu\text{A}$ ). This appears a relatively poor flux density compared to that of a standard nuclear research reactor. Nevertheless, in advantageous cases trace analyses can be performed. For example, routine analyses of several elements in air-dust samples were carried out by activation with photoneutrons (see below). Normally, however, the photoneutron flux is analytically negligible, whether as an analytical tool or as a source of interference.

## Sample Preparation and Irradiation

Basically, any kind of material – solid, liquid, or gaseous, organic or inorganic – can be analyzed by photoactivation. However, in preparing analytical samples, precautions are necessary against damage to the material, e.g., during long-period bremsstrahlung bombardment of organic matrices. Losses of components radiolytically volatilized during bremsstrahlung exposure sometimes have to be taken into account, e.g., mercury evaporation.

Typically, a set of samples of a few milligrams up to gram amounts are individually wrapped in aluminum foil and packed into the irradiation container. Pneumatic transfer systems are usually used for transport of samples to the irradiation position. **Figure 2** shows a schematic representation of a typical irradiation position. The sample rabbit is rotated by air-blow so as to obtain a quasihomogeneous activity distribution. Excessive heat is removed from the sample by the air-jet. Exposure periods depend upon the half-lives of the expected product activities; for a typical multielement trace analysis, irradiation times of a few hours are sufficient.

Other modes of sample preparation will be described below in the application examples.

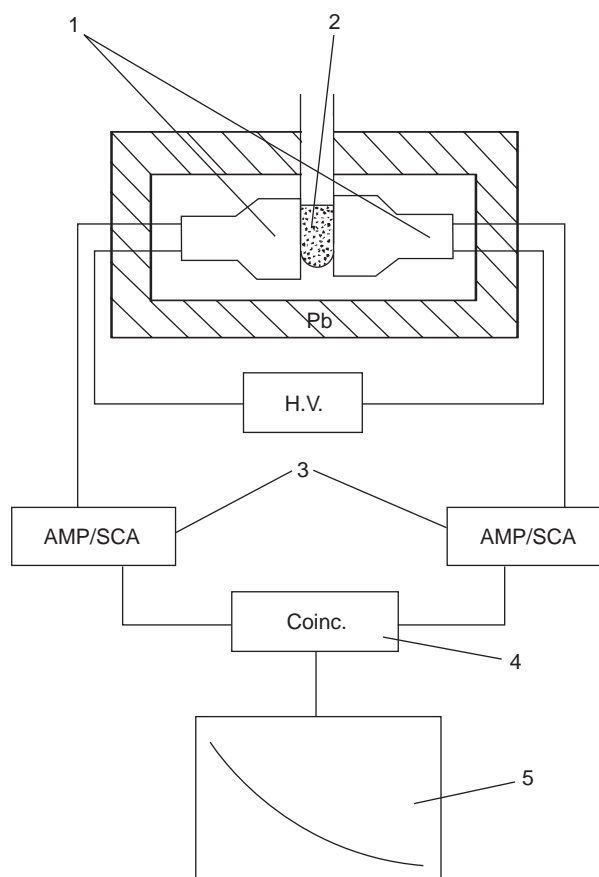


**Figure 2** Typical bremsstrahlung exposure position; schematic representation. 1 = Sample and reference material; 2 = sample rabbit; 3 = tangential airflow injection inlet; 4 = pneumatic tube; 5 = compressed air inlet for sample rabbit retransfer; 6 = bremsstrahlung converter assembly; 7 = electron beam window (Ti); 8 = accelerator tube.

## Photon Spectrometry

Activity measurements using radiation other than electromagnetic have been performed in only a few cases; only photon spectrometry is discussed here.

Two basic principles of photon counting have been used: detection by scintillators exploiting the radio-luminescence effect, and detection by semiconductor crystals. These detectors and the subsequent spectrometry electronics will not be discussed here. Spectrometry during neutron and photon activation analysis is performed in basically the same way. One difference is due to the different nature of the product nuclides in the respective activation techniques.  $\beta^-$ -Emitters are mainly produced by neutron activation, whereas neutron-deficient nuclei predominate after photon exposure. Thus, scintillation spectrometry after neutron activation has now more or less been abandoned. In photon activation analysis of the light elements, however,  $\gamma$ - $\gamma$  coincidence spectrometry using two NaI(Tl) crystals in sandwich geometry is used (see Figure 3). Here the background radiation can be suppressed to negligible level by measuring the  $\beta^+$  annihilation quanta simultaneously. Furthermore, as briefly mentioned earlier, low-energy photon spectrometry is recommended for multicomponent spectra measurement after photon activation. Flat ( $\sim 10$  mm thick) planar germanium or silicon diodes are used, equipped with



**Figure 3** Schematic representation of a  $\gamma$ - $\gamma$  coincidence spectrometer. 1 = NaI(Tl) crystal detector; 2 = irradiated sample; 3 = spectroscopy amplifier plus single channel analyzer (adjusted to screen the 511 keV annihilation peak); 4 = coincidence unit; 5 = recording instrument, e.g., multichannel analyzer.

thin beryllium entrance windows so as to allow measurement of soft photon radiation, e.g., characteristic X-rays. In several cases, the analysis of this energy region, say from 5 to 190 keV, offers some advantage compared with classical  $\gamma$ -ray spectrometry.

## Analytical Procedure

### Reference Materials

Photon activation analysis, as well as other instrumental techniques, is generally quantified by comparison of activities in the sample with those in a reference material of known elemental composition that was irradiated simultaneously. In photon activation analysis this is necessary particularly because some accelerator parameters and nuclear data for the photoreactions involved are either unknown or not precisely determinable. In addition, some machine parameters of the accelerator cannot be assumed constant throughout the exposure period and might show uncontrolled variation. The use of

reference materials irradiated simultaneously with the samples implicitly accounts for these parameters. Primary standards, i.e., pure elements or substances synthesized from stoichiometrically well-determined compounds, are used for quantification through comparison of the activities after activation. Certified reference materials are also analyzed to ascertain the total accuracy of the results. Frequently it is useful to provide a matrix-inherent or additive internal standard that serves as a flux monitor. The integral quality of the analytical data (in terms of accuracy and precision) is thereby significantly improved.

### Analysis of Light Elements

One special feature of photon activation analysis is the analysis of light elements such as carbon, nitrogen, oxygen, and fluorine. The analytical procedure in this case is essentially different from that in instrumental multielement analysis. This is due to the fact that the prominent photonuclear reactions of these elements, e.g.,  $^{12}\text{C}(\gamma, n)^{11}\text{C}$ , yield pure  $\beta^+$ -emitters exclusively and thus emit no characteristic photon radiation by which they might be detected. Hence, after activation these elements have to be separated radiochemically from the sample matrix prior to annihilation radiation measurement. In the case of carbon, nitrogen, and oxygen, this is usually performed by heat extraction. Fluorine separations are normally carried out by distillation as hexafluorosilicic acid. Nondestructive or instrumental analyses of the light elements are possible only in exceptionally favorable cases, namely if the activated matrix does not emit positron radiation at significant level at the time of measurement. However, currently photon activation techniques are under development that, with help of improved high-performance hardware and software, will enable instrumental analyses of light elements in multicomponent samples through photon activation.

### Analysis of Heavier Elements; Multielement Analysis

High-resolution photon spectrometry with semiconductor detectors has normally been used for multielement photon activation analysis. One generally strives for a purely instrumental or, if possible, nondestructive procedure. In several cases, however, a radiochemical separation step is unavoidable and sometimes advantageous.

## Examples of Analytical Procedures

### Radiochemical Analysis of Carbon in Molybdenum

A typical analytical procedure, such as that for carbon in molybdenum, can be summarized as follows.

About 1 g of material was placed in the rabbit. Pieces of graphite foil were put on both sides of the sample. These served as primary reference material. The rabbit was then transferred to the irradiation position through a pneumatic tube system. After 20 min exposure to 35 MeV bremsstrahlung (mean electron beam current  $80\text{ }\mu\text{A}$ ) the sample was transported back to the radiochemical laboratory and unpacked. The surface contamination was then removed by etching with aqua regia. The sample was placed in an alumina crucible that contained a mixture of 87.5%  $\text{Pb}_3\text{O}_4$  and 12.5%  $\text{B}_2\text{O}_3$  as oxidizing agent plus a piece of iron to initiate and support the melting process. The crucible was heated in a high-frequency furnace with an output power of 1 kW at a frequency of 10 MHz. The sample-borne carbon thus volatilized as  $^{11}\text{CO}_2/^{11}\text{CO}$  was transported by an argon flow from the melted sample. Inactive carbon dioxide originating from the iron flux served as an inactive carrier. The gas mixture was cleaned by a glass wool filter and led through a tube containing Schütze reagent (iodide pentoxide on silica gel) to oxidize radiocarbon monoxide to the desired dioxide. Finally,  $^{11}\text{CO}_2$  was absorbed with potassium hydroxide and then, after a total decay period of 30 min, counted in the coincidence spectrometer described above. A decay curve was recorded to ascertain the radiochemical purity of the sample. The reference material was then counted without a separation step. The detection limit was  $0.02\text{ }\mu\text{g}$  under the described conditions.

### Fluorine in Seawater

This procedure is fundamentally different from that described above not only because of the water matrix but because fluorine cannot be mobilized by heat extraction, but requires distillation. A typical procedure is as follows.

Three milliliters of seawater was freeze-dried. Artificially synthesized seawater containing known amounts of NaF was used as reference material and treated accordingly. In order to avoid first-order interference by sodium through the reaction  $^{23}\text{Na}(\gamma, \alpha n)^{18}\text{F}$ , the linear accelerator bremsstrahlung radiation was used at 22 MeV maximum energy so as to stay below the threshold energy of this reaction. The exposure period was 20 min. After dissolution of the sample in distilled water, the other halogens were removed by silver nitrate precipitation. Radiofluorine was then distilled as  $\text{H}_2\text{Si}^{18}\text{F}_6$  with some inactive sodium fluoride added as a carrier for quantitative vapor distillation. Finally,  $\text{PbCl}^{18}\text{F}$  was precipitated and then counted with a  $\gamma$ - $\gamma$  coincidence sodium iodide spectrometer. Data evaluation was performed



by a decay-curve analysis program. The overall detection limit under the experimental conditions described was found to be 0.03  $\mu\text{g}$  of fluorine.

### **Radiochemical Multielement Analysis of River Water**

Photon activation analysis is suited to water analysis because practically unlimited volumes can be irradiated, thus enhancing the total sensitivity and avoiding solidification prior to analysis; any pretreatment of water, e.g., evaporation, entails the danger of contamination or undesirable loss of components. Water can be analyzed without pretreatment using photon activation as follows.

Water (1.5 l) was brought in a storage vessel and 200  $\mu\text{g}$  of each of scandium and samarium were added as nitrate solution to serve as internal standards. The sample was then irradiated with 30 MeV bremsstrahlung (mean electron beam current 150  $\mu\text{A}$ ) from an electron linear accelerator for 5 h in a 0.7-l flow-chamber under slow continuous cyclic flow. Thereafter, the components of interest (Sc, Cr, Mn, Co, Ni, Cu, Zn, As, Y, Zr, Nb, Mo, Cd, Sn, Sb, I, Ba, Tl, Pb, and U) were separated by a sulfur-substituted cellulose-based sorbant filter switched into the flow cycle after activation. This sorbant was then dried in a microwave furnace and pressed to a thin (1 mm thick) pellet. Product activities were counted with a large coaxial germanium detector (where  $^{44\text{m}}\text{Sc}$  was used as an internal bremsstrahlung flux monitor), and with a planar germanium low-energy photon diode (where 103 keV of  $^{153}\text{Sm}$  served as an internal standard). Results were obtained by automatic spectrum processing; data evaluation was achieved with personal computer programs. The results were in good agreement with analyses performed in parallel using different procedures.

### **Radiochemical Analysis of Rock Material**

Owing to the nature of the matrix analyzed, a radiochemical separation step is sometimes required in the analytical procedure, e.g., in the case of severe interference of the analyzed signals by matrix activity. The analytical quality of the results (in terms of accuracy, precision, and sensitivity) is severely degraded if lower-energy gamma peaks (say less than 511 keV) to be evaluated are superimposed upon a huge Compton background continuum, or if the signal is interfered by other strong peaks in the close vicinity, in the worst case overlapping the desired signal. In the rock material analysis described the analyzed  $\gamma$ -ray line of  $^{57}\text{Ni}$  (1378 keV) originating from trace amounts of nickel suffered from the closeness of the 1369-keV line of  $^{24}\text{Na}$  due to

magnesium present as a major component in the sample. Therefore, radionickel was separated from the matrices of the samples (USGS standard rocks with recommended element contents). After exposure to 30 MeV bremsstrahlung at a mean electron beam current of 70  $\mu\text{A}$  for 2 h the samples (400 mg each) were transferred to platinum crucibles, together with 10 mg nickel carrier. This mixture was fused with 4 g sodium carbonate, and then dissolved in concentrated hydrochloric acid; the hydroxides were then precipitated with ammonia. After purification by redissolution and reprecipitation nickel was precipitated as glyoximate and formed into a small disk that was then subjected to high-resolution gamma spectrometry. Primary standards served as reference materials. The agreement with the comparison data was satisfactory. An overall sensitivity of  $\sim 50$  ng of nickel under the described conditions was found.

### **Instrumental Air-Dust Analysis**

Review of the analytical literature indicates the analysis of airborne particulate matter to be one of the most frequent applications for photon activation analysis.

Air-dust samples from different locations in the Federal Republic of Germany were taken in high-volume samplers. Cellulose nitrate membrane filters were used for particulate collection. The filter was divided into several subsamples to be analyzed comparatively by 12 laboratories using eight different techniques, photon activation being one of them. Internal standard solutions of 50  $\mu\text{g}$  scandium and 10  $\mu\text{g}$  samarium (see above, River water analysis) were dropped onto the sample, which was then dried and dissolved in acetone. Cellulose powder (150 mg) was then added to serve as a binding matrix for the pellet pressed after stirring to dryness. These were placed in a rabbit, transported to the irradiation position of a linear electron accelerator by pneumatic transfer, and irradiated with 30-MeV bremsstrahlung. After 4 h exposure and a cooling period of 2 days, the activation products of the sample were measured with the spectrometers described above; the resulting spectra were processed as already described (see above, River water analysis). Na, Mg, Si, Cl, K, Ca, Ti, Cr, Fe, Co, Ni, Zn, As, Se, Br, Rb, Sr, Y, Zr, Nb, Mo, Cd, Sn, Sb, I, Cs, Ba, Ce, Nd, Tl, Pb, and U were analyzed by their respective photoreaction products; vanadium and manganese were determined using  $^{52}\text{V}$  and  $^{56}\text{Mn}$ , respectively, produced through activation by photoneutrons. In comparing the results of all techniques applied, good agreement of the photon activation data with the joint consensus values was reported.

### **In vivo Analysis of Trace Metals in Humans**

During cancer radiotherapy using a betatron, several metal components were analyzed *in vivo* in the treated body regions. Patients were irradiated repeatedly (each exposure period was 10 min) at a distance of 1 m from the bremsstrahlung converter. Immediately thereafter, the treated body regions were measured with a large-volume germanium detector. Mg, Cl, Ca, Cr, Fe, Ni, Zn, Sr, Sn, and Pb were analyzed. No significant differences in the element concentrations were apparent during the treatment, which lasted for several weeks. However, qualitative and quantitative differences became discernible regarding the type of disease and the mode of treatment.

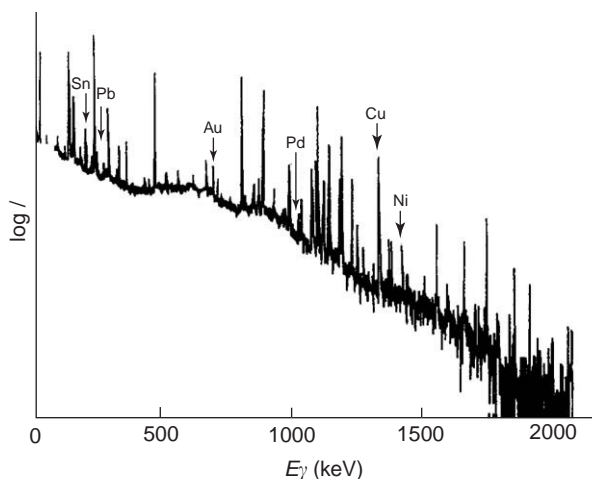
### **Nondestructive Analysis of an Ancient Musical Instrument**

In studying antiquities and objects of art, nondestructiveness is generally essential. A baroque silver trumpet (early eighteenth century) from a collection of 12 in total was analyzed nondestructively by photon activation. The primary objective was the characterization of the basic alloy (Ag/Cu) since several parts of the instrument had to be replaced. An alloy had to be used for restoration that was as similar as possible to the original material since it was intended to make the trumpet chorus playable, which necessitated providing compatible acoustic properties of all instruments. Copper sheet disks that served as flux monitors were attached to the areas of interest, which were then activated with bremsstrahlung from an electron linear accelerator. Maximum energy of 16 MeV was selected in order to avoid the production of undesirably long-lived  $^{105}\text{Ag}$  activity in the matrix alloy. Within the selected irradiated areas, spots of several square millimeters were screened with a lead collimator and measured with a coaxial germanium detector (see Figure 4). Thereafter, a piece of silver alloy sheet with known composition was activated and measured as a primary reference material. Processing of the resulting  $\gamma$ -ray spectra and quantitative data calculation were performed with appropriate personal computer programs.

With these analytical results, a silver alloy with a very similar composition could be found. After restoration, acoustic examination yielded excellent homogeneity of the chorus. The collection can now serve its original purpose of being played at special ceremonial events.

### **Fields of Application**

Since there is no limitation of the method with respect to the matrix studied, there is a broad spectrum of applications in the field of material science.



**Figure 4**  $\gamma$ -Ray spectrum of an ancient silver trumpet after exposure to 30 MeV bremsstrahlung with several component signals marked; most of the others are due to the matrix silver.

Photon activation analysis has primarily been applied in the areas of geo- and cosmochemistry, oceanography, environmental science, industrial raw- and end-product analysis, high-purity material studies, organic material/medical and biological material analysis, forensic science, art and archaeology, certification of candidate reference materials.

**See also:** **Activation Analysis:** Neutron Activation. **Air Analysis:** Sampling; Outdoor Air. **Archaeometry and Antique Analysis:** Metallic and Ceramic Objects. **Carbon. Geochemistry:** Inorganic; Soil, Major Inorganic Components. **History of Analytical Science. Nitrogen. Quality Assurance:** Reference Materials. **Water Analysis:** Freshwater; Seawater – Inorganic Compounds.

### **Further Reading**

- Ebihara M, Oura Y, Ishii T, *et al.* (2000) How effectively is the photon activation analysis applied to meteorite samples?. *Journal of Radioanalytical and Nuclear Chemistry* 244(3): 491–496.
- Kucera J and Randa Z (2001) Possibilities of low-level determination of silicon in biological materials by activation analysis. *Fresenius Journal of Analytical Chemistry* 370(2–3): 241–245.
- Lutz GL, Segebade C, and Weise P (1987) *Photon Activation Analysis*. Berlin: Walter De Gruyter Inc.
- Miyamoto Y, Haba H, Kajikawa A, *et al.* (1999) Interferences in neutron and photon activation analysis. *Journal of Radioanalytical and Nuclear Chemistry* 240(1): 165–175.
- Randa Z, Kucera J, and Soukal L (2003) Elemental characterization of the new Czech meteorite ‘Moravka’ by neutron and photon activation analysis. *Journal of Radioanalytical and Nuclear Chemistry* 257(2): 275–283.
- Weise HP, Gerner W, and Hedrich M (2001) Determination of elements by nuclear analytical methods. *Fresenius Journal of Analytical Chemistry* 369(1): 8–14.

# ADHESIVES AND SEALANTS

**R G Leonard and K Quigley**, Henkel-Loctite, Tallaght, Dublin, Republic of Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The use of adhesives/sealants in both the industrial and consumer spheres has increased dramatically in the past 20 years. In the industrial segment, both reactive and nonreactive systems are used in a wide variety of applications. The increasing use of reactive systems has, however, tended to overshadow that of the nonreactive systems. Most prominent amongst the reactive systems favored in industrial applications are anaerobic sealants (methacrylate ester based), instant adhesives (alkyl cyanoacrylate ester based), acrylic (toughened) adhesives, epoxy resin adhesives, polyurethane/isocyanate-based adhesives, silicone adhesives/sealants, and phenolic resin adhesives.

In the case of nonreactive type of systems favored for the industrial market, the most widely used types are (1) hot melt adhesives, (2) solvent-based adhesives, and (3) pressure-sensitive adhesives.

In the hot-melt type, two principal polymer types are used: polyolefin and ethylene co-polymer based; and polyester and polyamide type. In the solvent-based type, the most prominent are neoprene (polychloroprene)-based solvent and latex types, and polyvinyl acetate emulsions. For pressure-sensitive adhesives, the most favored are acrylic adhesives and butyl rubber/polyisobutylene types.

For the analytical laboratory required to analyze a sealant/adhesive formulation a considerable challenge is posed by the wide range and considerable complexity of many of the adhesive types. Undoubtedly, the reactive formulations pose the greatest challenge. This is particularly evident in the case of the redox-cured methacrylate resin-type formulations (anaerobic products), which are widely used in threadlocking, sealing, and retaining applications in the engineering and machinery industries.

## Analysis of Anaerobic Sealants/Adhesives

These anaerobic sealants are essentially liquids that remain stable (i.e., in the original liquid form) in the presence of oxygen. When confined between two closely fitting metal parts (e.g., nut and bolt), thereby excluding oxygen, the monomeric constituents in the

liquid polymerize. This polymerization is rendered especially favorable if the metal is able to catalyze the redox-based 'cure' process. These sealants consist of a complex mixture of methacrylate (or less commonly, acrylate) esters, free-radical initiators, accelerators, free-radical inhibitors, metal-chelating agents, plasticizers, thickeners, inert fillers, pigments, and dyestuffs. These constituents are very carefully adjusted to give the correct balance between speed of cure and stability.

A rather simplified schematic of the redox-based cure chemistry of anaerobic sealants is shown below in **Scheme 1**.

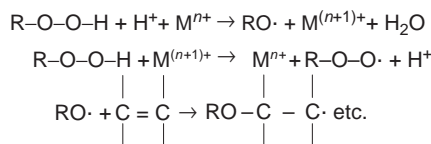
The reaction scheme highlights the key catalytic role played by the active metal surface. In essence, the metal functions as an intrinsic component of the cure system.

### Typical Reactive Monomers and Resins

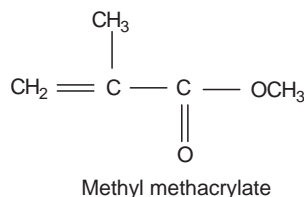
The polymerizable methacrylate monomer is typically the dominant component in the composition. The monomer may be present as a single component or in a mixture. Some examples of commonly used methacrylate esters are shown in **Schemes 2–5**.

### Typical Cure Promoters and Stabilizers

The cure systems typically incorporate an organic peroxide, an organic reducing agent, and an organic acid. The preferred peroxides include cumene hydroperoxide (CHP), *t*-butyl hydroperoxide, *t*-butyl perbenzoate, and dibenzoyl peroxide. The organic reducing agents are typically aromatic amines such



**Scheme 1**



**Scheme 2**

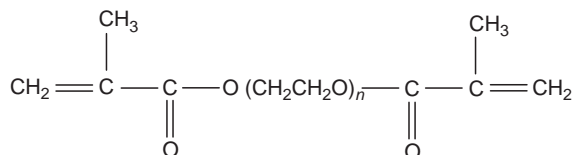
as *N,N*-dimethyl-*p*-toluidine and 1,2,3,4-tetrahydroquinoline. Occasionally, other reducing agents including acylhydrazines and mercaptans may also be used.

A variety of organic acids including *p*-toluenesulfonic acid and maleic acid have been used as cure accelerators. The other essential components in any anaerobic or acrylic formulation are the stabilizing components that are required to prevent any premature polymerization, either during manufacture or in subsequent storage of the formulations prior to use. The free-radical inhibitors may include phenolic compounds such as hydroquinone, *p*-methoxyphenol, pyrogallol, etc. Cure retarders may also be used as typified by 1,4-benzoquinone and 1,4-naphthoquinone.

Metal-sequestering agents are sometimes incorporated at low levels (10–100 mg per kg) to preclude any problems arising from trace metal contamination. Ethylenediaminetetraacetic acid (EDTA) and its various analogs tend to be the preferred type.

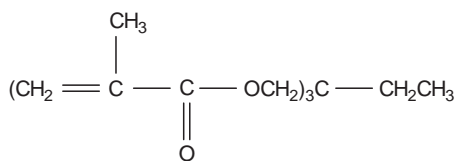
### Characterization of Monomer/Resin Components

The presence and/or identity of the polymerizable component(s) in a formulation are routinely



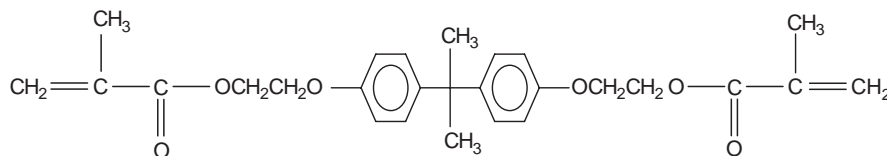
Poly(ethylene glycol) dimethacrylate

Scheme 3



Trimethylolpropane trimethacrylate

Scheme 4

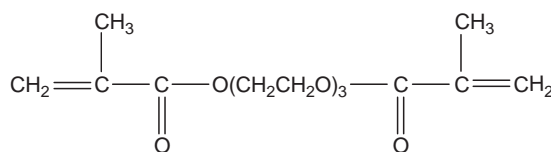


Ethoxylated bisphenol (A) dimethacrylate

Scheme 5

established by a combination of nuclear magnetic resonance (NMR) and infrared (IR) spectroscopy together with gas chromatographic (GC) and liquid chromatographic (LC) techniques. In the case of anaerobic sealants, which cure through a room-temperature redox-initiated mechanism, the monomeric components are typically mono-, di-, or tri-functional methacrylate esters or a combination of these. The  $^1\text{H}$  and proton decoupled  $^{13}\text{C}$  NMR spectra of triethyleneglycol dimethacrylate (TRIEGMA) Scheme 6 recorded on a 270 MHz FT-NMR instrument are shown in Figures 1 and 2, respectively. This monomer is used extensively in anaerobic products.

The proton NMR spectrum shows the methyl substituents from the methacrylate end groups at 1.9 ppm (relative to tetramethylsilane (TMS)). The methylene groups bonded to ether oxygen appear in



Triethyleneglycol dimethacrylate

Scheme 6

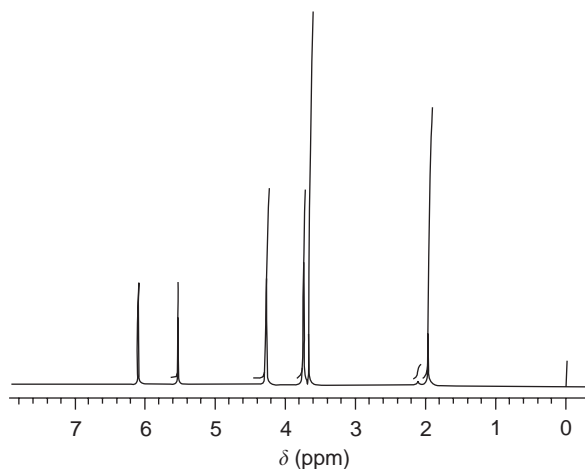


Figure 1  $^1\text{H}$  NMR spectrum (270 MHz) of triethyleneglycol dimethacrylate in  $\text{CDCl}_3$ .

the range 3.6–3.8 ppm, while the methylene groups bonded to ester oxygen are found at 4.3 ppm (relative to TMS). The terminal olefinic protons give resonances at 5.6–6.1 ppm.

The  $^{13}\text{C}$  spectrum of the TRIEGMA monomer shows the following resonances: the methyl substituents at 18 ppm, ester methylene groups at 64 ppm, ether methylene groups at 69 and 71 ppm, olefinic carbon atoms at 126 and 137 ppm (terminal carbon atom), and carboxyl carbon atoms at 168 ppm.

Such multifunctional esters are used to promote the formation of a rigid cross-linked structure on polymerization. Quantitation of these esters is normally based on GC analysis using an internal standard. Either narrow or wide bore capillary columns may be used. In the case of ultraviolet cure-initiated formulations, the use of higher-molecular mass urethane–methacrylate resins such as **Scheme 7** is involved.

The 270 MHz  $^1\text{H}$  NMR spectrum of the urethane resin includes resonances typical of methacrylate esters as outlined above. In addition to these, bands due to alkyl groups (1.0–1.4 ppm range), methyl substituents on aromatic rings (2.2 ppm), and

aromatic protons (7–8 ppm) are present. The  $^{13}\text{C}$  spectrum of the same urethane resin shows, in addition to the methacrylate resonances, resonances associated with alkyl substituents (17 ppm), aromatic rings (110–140 ppm), and the urethane carboxyl moiety (154 ppm).

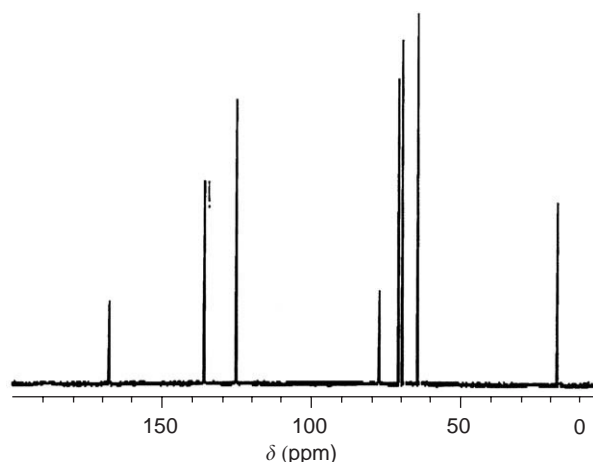
Such resins may be used in combination with the simpler methacrylate monomers to improve the performance of the cured product. Their relative involatility and the presence of a UV-absorbing chromophore make LC the preferred approach for analysis. Again the use of an internal standard is favored for quantitation purposes.

### Characterization of Cure/Stabilizer Systems

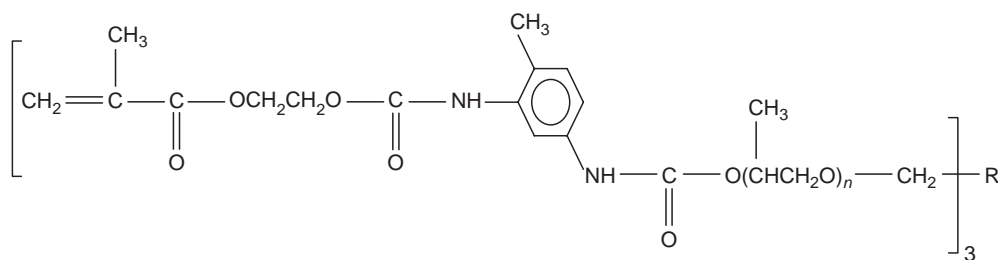
The performance of modern adhesive/sealant formulations is critically dependent on the cure system used. This must be judiciously balanced in order to give the desired performance, especially with respect to cure speed, together with a viable stability for the packaged product. The cure promoters are present at relatively low concentration ( $\sim 1\%$  w/w). It is essential therefore that the appropriate analytical methods and techniques are available to identify and quantify these reactive components.

Thin-layer chromatography (TLC) is a fast, inexpensive technique well suited to adhesive/sealant analysis. Unlike many analytical techniques, little advanced instrumentation is required, apart from a UV/visible light source for some tests. In addition, sample preparation is generally minimal. Furthermore, in conjunction with specific spray reagents, TLC can be rapidly used to confirm the presence of organic peroxides, reducing agents, organic acids, and photoinitiators (if a UV-curing formulation is involved). The presence of certain free-radical inhibitors can also be established by TLC.

A typical TLC procedure involves the use of commercially available precoated plates (silica gel F254). Direct application of dilute solutions of the adhesive products is followed by development of



**Figure 2**  $^{13}\text{C}$  NMR spectrum (270 MHz) of triethyleneglycol dimethacrylate in  $\text{CDCl}_3$ .



Urethane–methacrylate resin

**Scheme 7**

the plate in an appropriate solvent system. The curatives in the formulation will separate from the other components (and from each other) as the chromatogram is developed. Subsequent analysis of the TLC plate is carried out with the use of specific spray reagents for visual detection of the components. A common curative used in adhesive formulations is the peroxy initiator CHP. A TLC plate containing the substance is sprayed with an aqueous methanolic solution of *N,N*-dimethyl-*p*-phenylenediamine dihydrochloride and heated gently. The presence of CHP is confirmed by the appearance of purple spots on the plate, corresponding in  $R_f$  value with a standard solution of CHP on the same plate. Aromatic amines are also components in adhesive/sealant catalyst systems. The spray reagent used to detect these compounds consists of a dilute aqueous solution of iron(III) chloride and potassium hexacyanoferrate(III). The presence and chemical identity of the amines (and other appropriate reducing agents) can be determined by the appearance of variously colored spots with different  $R_f$  values on the chromatogram.

In a similar fashion, acidic species present in some catalyst systems can be determined on TLC plates by the use of a dilute ethanolic solution of 2,6-dichlorophenolindophenol sodium salt.

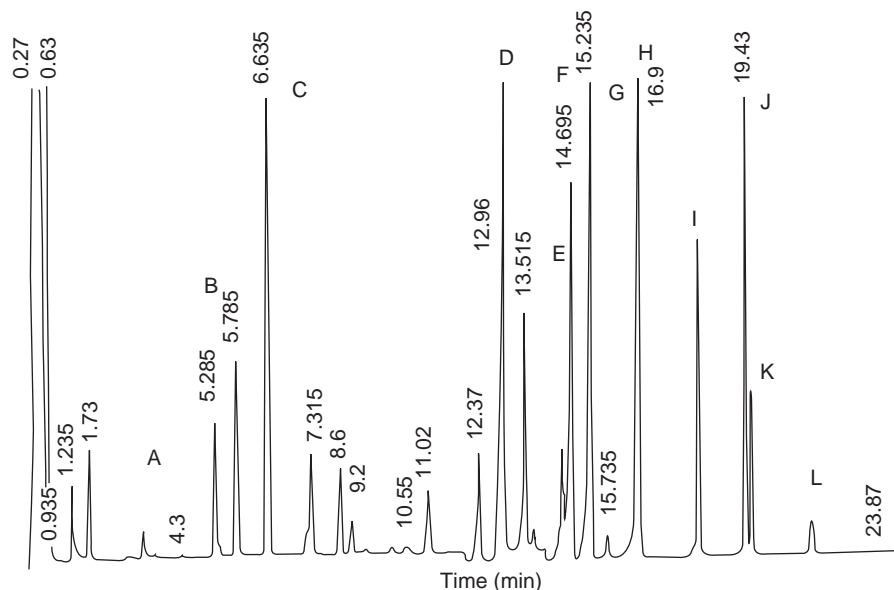
Free-radical stabilizers such as hydroquinone, *p*-benzoquinone, *p*-toluquinone, chloranil, and 1,4-naphthoquinone can be detected by reaction with rhodanine and ammonia (spot test or TLC method) to give a colored complex.

While the above procedures are useful for qualitative work, commercially available TLC plate scanners (with varying levels of automation) can be used for quantitative assays.

In the case of radiation-curable (e.g., UV) adhesives, TLC represents a very useful screening test for the presence of several characteristic photo-initiators, e.g., benzophenone, 2,2'-diethoxyacetophenone, 1-benzoylcyclohexanol, and 4,4'-dimethylaminobenzophenone (Michler's ketone).

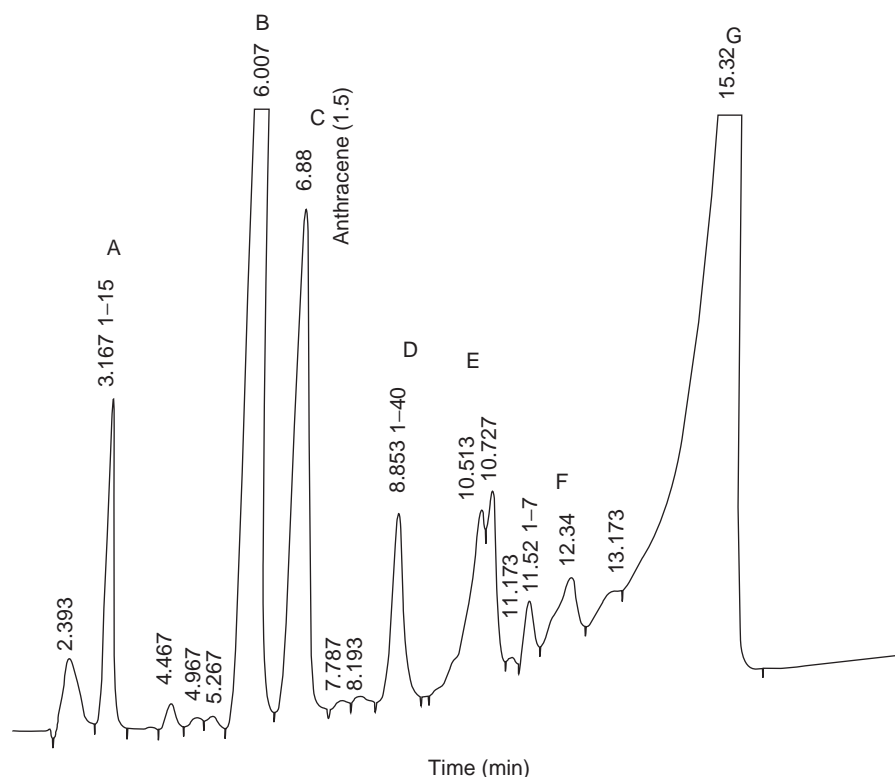
GC and LC procedures may then be used to confirm the TLC results, as well as being used to quantify the level of each component. These chromatographic techniques, moreover, have the added advantage that they may also be used to identify and quantify the monomeric and plasticizer components of the formulation (Figures 3 and 4). As stated previously, either narrow or wide bore capillary columns may be used for GC analysis. Preferred stationary phases are OVI (poly(dimethylsiloxane)) and OV17 (poly(phenylmethylsiloxane)). Flame ionization is the usual mode of detection, but where unambiguous confirmation of peak identity is required, a gas chromatograph coupled to a bench-top mass spectrometer may be used (GC-MS).

In LC analysis, the usual approach adopted is to employ the reversed-phase mode using a  $C_{18}$ -type stationary phase. A typical eluent would be a binary phase based on methanol or tetrahydrofuran with water. The latter solvent mixture has the advantage that the strong solvent capability of the tetrahydrofuran readily dissolves the polymeric components



**Figure 3** GC trace of a typical anaerobic sealant: A, carboxylic acid; B, organic peroxides; C, internal standard (I); D, methacrylate ester; E, aryl hydrazine; F, methacrylate ester; G, internal standard; H, methacrylate ester; I, methacrylate ester; J, plasticizer; K, methacrylate ester; L, methacrylate ester.





**Figure 4** LC trace of typical UV-curing adhesive. A, adhesion promoter; B, photoinitiator; C, internal standard; D, methacrylate ester (I); E, urethane-methacrylate oligomers; F, methacrylate ester (II); G, urethane-methacrylate resin.

usually incorporated in these formulations. A variable wavelength UV/visible detector is used, and when quantitation is required the compound is monitored at its wavelength of maximum absorption.

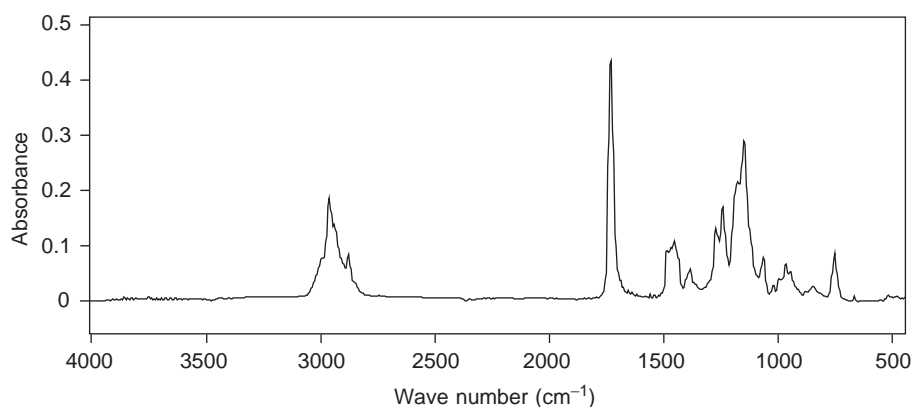
The peroxide content of sealants/adhesives is normally established by an iodometric titration approach that involves reducing the peroxy group with iodide ion in an acidic medium and titrating the liberated iodine with standard sodium thiosulfate. The endpoint is established by potentiometric detection with a platinum-reference electrode combination. Aromatic amine content is determined by titration with perchloric acid in acetic acid as solvent. The endpoint is established using potentiometric detection with a glass-reference electrode combination. The latter electrode combination is also favored for quantitation of the acidic components of the formulation. Dilute alkali hydroxide or tetrabutylammonium hydroxide are used as titrants in aqueous or nonaqueous media, respectively. In the case of the organic peroxides, an important test applied to the incoming materials is an assay procedure based on their measured active oxygen content. This is routinely determined using a titrimetric procedure involving iodometric analysis. This procedure although well established, is relatively time consuming. Recent studies have confirmed that several of the

more frequently used commercial organic peroxides may be assayed by chromatographic techniques, e.g., GC and reverse-phase liquid chromatography (RPLC) as appropriate. Accuracy and precision compare favorably to the classical titrimetric approaches. The principal advantages are the relatively short analysis time as compared to the conventional approach. In the case of the GC assay, the use of an on-column injection technique is necessary to prevent thermal decomposition of the peroxide in the injection port.

### Viscosity Modifiers and Fillers

Viscosity modifiers, usually referred to as thickeners, are used in varying concentrations throughout the range of sealants and adhesives. They span a diverse range of chemical structures. Some of the more commonly encountered thickeners include poly(alkyl methacrylates) – homo and/co-polymers; poly(alkyl acrylates) – homo and/co-polymers; polystyrene; acrylonitrile-butadiene-styrene co-polymers; poly(vinyl acetate); fumaric or maleic acid-based polyesters.

The chemical identity of such materials may readily be confirmed by IR spectroscopy. The initial step involves isolation of the thickener from the formulation matrix. Typically, the adhesive sample is diluted with chloroform to reduce the viscosity. This



**Figure 5** FTIR spectrum of co-polymer of methyl methacrylate and butyl methacrylate.

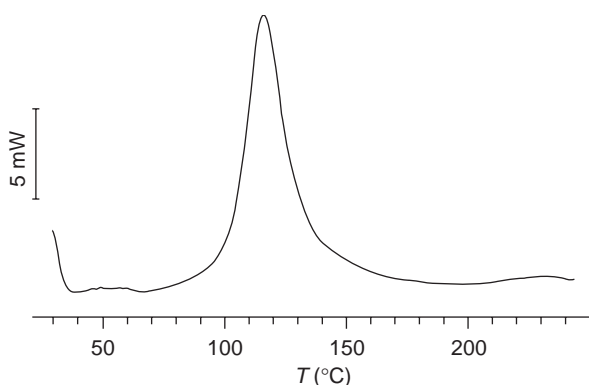
solution is then subjected to dropwise addition of a polar solvent such as methanol until the thickener is fully precipitated out of solution. After careful drying to remove all traces of solvent, the sample can be analyzed by Fourier transform IR (FTIR) or dispersive IR spectroscopy. The IR spectrum of an acrylic thickener identified as a methyl methacrylate/butyl methacrylate co-polymer is reproduced in **Figure 5**.

Fillers and thixotropic agents are used widely throughout the various acrylic sealant and adhesive products. Fillers also form a significant component of epoxy-based adhesives. For example, if it is necessary to identify the filler used in the hardener component of an epoxy product, the isolation sequence involves initial centrifugation of the product to isolate the crude filler component. The latter is then washed repeatedly with acetone and dried thoroughly. The fine dry powder may then be used to prepare a standard KBr disk for IR analysis.

## Some Useful Analytical Techniques in the Analysis of Industrial Adhesives and Sealants

### Thermal Analysis

Differential scanning calorimetry (DSC) can be used to generate cure profiles of certain adhesive producers. For example, in the electronics assembly business, epoxy-based chipbonding products are used to hold components on to printed circuit boards prior to the automated soldering process. These adhesives are formulated with heat curatives, i.e., they are designed to harden on application of heat. As such, the heat curing profile of these materials can be examined by DSC. The polymerization exotherm detected for a typical epoxy-based chipbonder is shown in **Figure 6**. The heat of reaction, together with peak temperature and other details, can be obtained from this trace. This type of analysis is useful,



**Figure 6** Cure profile of typical chipbonder product (DSC scan).

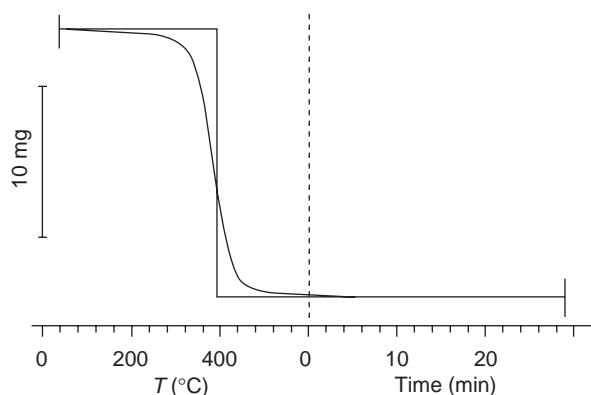
not only for product monitoring, but as an aid to product development work where different cure profiles may be required to match specific applications.

A further use of DSC in this area is the detection of incomplete (residual) curing in solid samples. A thermally curable epoxy adhesive that has not been fully cured will polymerize as the sample is heated in the calorimeter. This polymerization will give rise to a small exotherm that can be used to estimate the degree of cure in the original sample.

Finally, DSC is used in the broader area of polymer evaluation to measure glass-transition temperature, degree of crystallinity, decomposition profile, etc. Current computer control and storage facilities available with thermal analysis equipment allow data to be comprehensively evaluated, reanalyzed, and compared, reducing the need for repeat analyses.

Thermogravimetric analysis (TGA) is a technique whereby weight changes in a sample can be continuously monitored with respect to temperature. TGA can be used to look at weight losses associated with the polymerization of an adhesive product, or to determine percentage volatiles released from a cured product during thermal cycling.



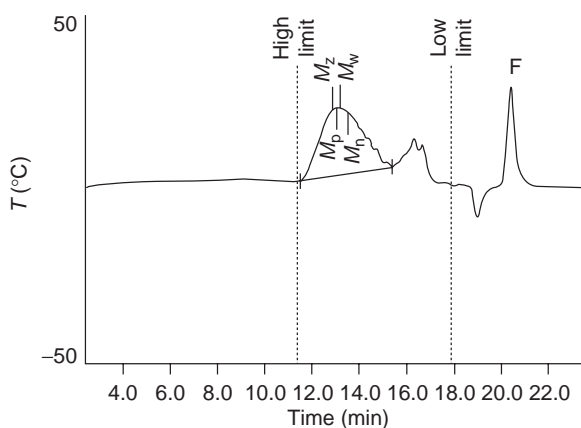


**Figure 7** TG curve (room temperature to 600°C) for a chipbonding product to determine filler content.

The technique can also be used to determine mineral filler contents of certain products. In this case, a sample is heated gradually to a high temperature and maintained at this temperature for a period to ensure complete ashing. A typical thermogram obtained for a chipbonding product is shown in **Figure 7**. The sample was heated from room temperature to 600°C, and held at the final temperature for 30 min. The resultant stable baseline obtained after complete combustion is used in the calculation of filler content for the sample (13.7% w/w).

### Gel Permeation Chromatography

Gel permeation chromatography (GPC) or size-exclusion chromatography (SEC) is a favored technique for the analysis of polymers and oligomeric compounds. GPC separates sample molecules in the mobile phase on the basis of their effective size. In most cases molecular size can be correlated directly to molecular mass. For the analysis of organic materials the column(s) is packed with a highly cross-linked spherical polystyrene/divinyl benzene matrix with a tightly controlled pore diameter. Any sample molecules that are smaller than the pore size can diffuse into and out of the pores, whereas sample molecules larger than the pores are effectively excluded. As a result, larger molecules elute more rapidly than small molecules. The separation is thus essentially mechanical in nature, rather than chemical as in other forms of chromatography. All molecules larger than the pores elute first and at the same retention time. The smallest molecule that cannot penetrate the pore defines the exclusion limit of the column. Molecules smaller than a certain size have equal access to the pores and so they elute together at the total permeation volume or dead volume. Molecules between these extremes are separated in order of their respective molecular sizes.



**Figure 8** GPC trace of acrylate process oligomer.  $M_p = 2029$ ;  $M_n = 1534$ ;  $M_w = 1862$ ;  $M_z = 2272$ ;  $M_w/M_n$  (polydispersity) = 1.213.

In the analysis of a sample of unknown molecular mass or molecular mass distribution, the first step is to generate a calibration plot ( $\log M_r$  versus elution volume or time) using a set of narrow disperse standards of known molecular masses. The function giving the best fit to the plot is computed and the sample is then analyzed against this calibration function.

In this manner, the number average molecular mass ( $M_n$ ), the mass average molecular mass ( $M_w$ ), the peak molecular mass ( $M_p$ ), and the polydispersity ( $M_w/M_n$ ) may be calculated.

Polystyrene calibration standards are most frequently used, and with polymers of different chemical structure, molecular masses are reported relative to polystyrene.

The GPC trace and computed molecular masses obtained for an acrylate process oligomer are reproduced in **Figure 8**.

GPC is also a very useful technique for characterizing the molecular mass range/distribution of the various organic polymers and co-polymers used as thickeners in anaerobic, acrylic, and radiation-curable adhesives. The base polymer, tackifying resin, and petroleum wax components of hot-melt type adhesives may also be characterized in a similar manner.

The principal limiting factor is solubility, but the use of strong solvents, e.g., tetrahydrofuran, extends the applicability of this technique. Solubility problems may also be overcome by the use of elevated temperatures. Effective high-temperature GPC requires that all the key components – injector, columns, and detector – be temperature controlled. Commercial systems embodying these concepts are available.

### Pyrolysis Gas Chromatography

Pyrolysis GC is an analytical technique whereby complex involatile materials are broken down into smaller volatile constituent molecules by the use of

very high temperatures. Polymeric materials including cured sealants and adhesives lend themselves very readily to analysis by this technique. Essentially, a fingerprint uniquely characteristic of the original material is obtained.

In combination with IR analysis, pyrolysis GC is particularly useful in the following areas:

1. Characterization of polymeric thickeners used in the formulation of sealants and adhesives. Polymers of this type are frequently used to provide protective coatings – in rust-preventative products for example.
2. Characterization of plastic packaging materials.
3. Identification of cured adhesives, by generating pyrogram ‘fingerprints’ for selected adhesives.

Pyrolysis GC is an especially valuable and informative technique when coupled with mass spectrometric detection, which facilitates identification of the pyrolysis products.

### Atomic Spectroscopy

Atomic spectroscopy (including atomic absorption spectrometry, atomic emission spectrometry, and atomic fluorescence spectrometry) is of use across the span of reactive adhesive technologies. For example, the cure of anaerobic adhesives on non-reactive surfaces is usually assisted by the use of an active metal-based primer. Similarly, the cross-linking of silicone adhesives is promoted by the use of organometallic salts of cobalt, tin, iron, lead, and platinum. In the case of polyurethane adhesives, the key condensation reactions are catalyzed by tin salts (e.g., dibutyl tin dilaurate and stannous octoate).

The adhesive or primer is usually dissolved in an appropriate solvent and analyzed. A calibration curve is generated using standard solutions of the metals prepared in the same solvent and covering the anticipated concentration range. Where the matrix is difficult to solubilize or where very low levels (low ppm to ppb) are involved, the sample should be ashed prior to analysis. A low level of a binder component, e.g., *p*-toluenesulfonic acid should be added to the sample prior to heating to prevent loss of the metal through volatilization. The residue obtained following the ashing process is dissolved in a dilute mineral acid and analyzed by atomic spectroscopy. Where matrix interference is suspected, a standard addition approach may be used.

Packaging plastics may also be prepared for trace metals analysis by dry ashing prior to atomic spectroscopic analysis.

For matrices containing a high level of inorganic fillers, microwave-assisted acid digestion in the presence of concentrated mineral acids may be

required to facilitate the analysis of the sample by atomic spectroscopy.

Where the formulations embody significant levels of silicon dioxide or titanium dioxide the sequential use of nitric acid followed by hydrofluoric acid may be necessary to dissolve these oxides prior to analysis.

### Polarography

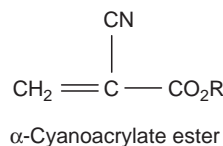
Where specification of the metal is considered to be important, polarographic analysis may be utilized. In the case of primers based on oil-soluble iron salts for use with anaerobic adhesives, the  $\text{Fe}^{2+}/\text{Fe}^{3+}$  ratio may be determined by differential pulse polarography in a supporting electrolyte based on ammonium pyrophosphate buffer adjusted to pH 9.0. The iron(II) form is known to be significantly more effective than the iron(III) form in the redox-based curing process.

Polarographic analysis may also be used to determine the level of selected chelating agents incorporated in an adhesive formulation. In the case of EDTA-based chelators, a standard addition approach using copper(II) is preferred. The supporting electrolyte is ammonium acetate at pH 5.0. This approach takes advantage of the fact that metal-chelate are typically reduced at more negative potentials than the metals themselves.

In passing it should be noted that voltammetric analysis is also very useful for determining many of the typical cure components used in the engineering adhesives. Thus, organic peroxides, aromatic amines, substituted hydrazines, thiols, etc., may also be determined at the mercury electrode.

### Other Adhesives Technologies

**Instant (cyanoacrylate) adhesives** The use of ‘instant adhesives’ based on  $\alpha$ -cyanoacrylate esters (**Scheme 8**) in both the industrial and consumer markets continues to increase. Typically, the methyl, ethyl, and butyl esters are used, but interest is also developing in the new ‘low-odor’ types based on alkoxyalkyl esters. The primary mechanism of polymerization is anionic. Weak bases including water can initiate polymerization to yield very high molecular mass homopolymers. Key additives are selected anionic stabilizers including both Lewis and protonic acids. Other additives may include organic polymers (e.g., poly(methyl methacrylate) functioning as thickening agents) and silica-based thixotropic agents.



**Scheme 8**

The nature of the ester may readily be confirmed by NMR ( $^1\text{H}/^{13}\text{C}$ ) and IR analysis. Protonic acid stabilizers typically present at the low mg per kg level may be quantified by nonaqueous titrimetry. Lewis acid stabilizers may be determined by an indirect approach such as atomic spectroscopy (emission/absorption).

The nature and concentration of the thickener may be determined using the following approach. The cyanoacrylate (CA) sample is initially diluted (1:1) with chloroform. Hexane is then added dropwise to the CA solution, with constant stirring. The thickening agent of high relative molecular mass will normally precipitate from solution as a white coagulated mass. The supernatant liquid (containing CA monomer) is decanted, and chloroform is added to the crude thickener to redissolve it. Dropwise addition of hexane is repeated, to effect slow precipitation of the thickener component. Further dissolution/precipitation of the solid, followed by a final hexane wash, yields a relatively pure sample of the thickening agent. The isolated polymer may then be characterized by IR, NMR, GPC, pyrolysis GC-MS, etc.

Although the preferred route of polymerization is anionic, free-radical polymerization is also possible, particularly at higher temperatures. Thus, low levels of suitable inhibitors, e.g., hydroquinone, *p*-methoxyphenol, 2,6-di-*t*-butyl-4-methylphenol, etc., may also be incorporated in the adhesive formulation. The type and concentration of such stabilizers may be determined by RPLC with UV detection.

Gas chromatographic analysis of cyanoacrylate esters is complicated by their highly reactive nature. The problems posed by premature polymerization were previously addressed by first reacting the cyanoacrylate ester to form the Diels-Alder adduct with an appropriate diene, e.g., isoprene. The resulting Diels-Alder complex was then analyzed by GC.

Problems presented by this approach included the presence of impurities contributed by the diene, etc. The approach currently used in our laboratory involves direct GC/flame ionization detector (FID) analysis on medium- or wide-bore cyanopropyl/phenyl/methyl silicone bonded-phase columns. The key prerequisite is to stabilize the sample against anionic polymerization by prior addition of a low level of an appropriate protonic acid stabilizer.

### Polyurethane Adhesives

A typical polyurethane sealant formulation may contain a urethane prepolymer, an isocyanate, plasticizer, fillers (including carbon black), and several other components. An important criterion is the concentration of residual isocyanate functionality available for further reaction. This may be established by reaction with

di-*n*-butylamine and back-titration of the unreacted amine with standardized hydrochloric acid. The identification of the base isocyanate may be determined by NMR ( $^1\text{H}/^{13}\text{C}$ ) spectrometry and/or pyrolysis GC. Identification of the plasticizer can be achieved if this component is isolated from the sealant matrix. A small quantity of sealant is allowed to moisture-cure (exposure to atmosphere for 24 h). The black, rubbery solid is then divided into small pieces with a knife, and extracted with 5 ml of hexane. The hexane solution is dried with a suitable drying agent (e.g., anhydrous sodium sulfate) and filtered to remove any solid particles. The solvent is removed from the sample by evaporation prior to analysis by, e.g., IR or NMR spectrometry.

The procedure yields a relatively pure extract of the plasticizer component, since the other major components in the polyurethane formulation are either physically or chemically immobilized in the cured polymer matrix.

### Epoxy Adhesives

The base resin in epoxy adhesive products can be readily identified by IR and NMR spectrometry. The most commonly employed resin in these products (both one-part and two-part) is that based on the diglycidyl ether of bisphenol A (bisphenol A epoxy resin). Other resins of importance include the cycloaliphatic epoxies and phenolic epoxy novolaks.

Epoxy products may contain high loadings of inorganic fillers (e.g., silica, talc, calcium carbonate). These can be removed by centrifugation of a sample that has been dissolved in a suitable solvent (e.g., acetone), prior to spectrometric analysis of the epoxy resin. The fillers may be analysed separately by IR spectrometry.

Hardener components in two-part products can also be analysed by IR/NMR spectrometry, provided fillers, pigments, etc., have been removed from the bulk sample.

Common curing agents in such products include anhydrides (e.g., phthalic anhydride, hexahydrophthalic anhydride, pyromellitic dianhydride), and also polyamine and polyamide-based hardeners.

One-part, heat-curing epoxies require no mixing, since polymerization is initiated by heating the product to trigger the curative in the formulation. Dicyandiamide is a high-temperature curing agent that may be detected in an epoxy matrix by either IR or LC analysis, the latter being carried out on an aqueous methanolic extract of the sample.

Thermal analysis (DSC) of heat-curing epoxides (Figure 6), such as specialist chipbonding products for the electronics industry, yields additional information about these materials, including heat of polymerization, peak temperature, onset of heat cure, and polymer conversion data.

The electronics industry has a particular concern regarding the level of 'ionics', such as the halides, which can contribute to metallization corrosion. Following defined extraction protocols involving the use of de-ionized water the extracts of both cured and noncured formulations may be analyzed by ion chromatography (anion and cation) with conductivity as the favored detection mode. Low  $\mu\text{g ml}^{-1}$  levels of extractable ions may be measured using this approach.

Where total halide, e.g., chlorine, is required then total combustion in a high-pressure oxygen bomb should precede the anion ion chromatography analysis.

### Cured Adhesives

In the case of a fully cured adhesive the analytical options are considerably narrowed. Should the curing reaction involve any significant degree of cross-linking, a relatively insoluble matrix will result. Appropriate analytical techniques under these circumstances would include IR (reflective techniques such as specular and diffuse reflectance), pyrolysis GC/MS, and photoelectron

spectroscopy. A computer-assisted search against appropriate library collections is a useful asset, particularly in the case of the first two techniques.

*See also:* **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Principles and Instrumentation. **Atomic Fluorescence Spectrometry.** **Gas Chromatography:** Pyrolysis; Mass Spectrometry. **Liquid Chromatography:** Normal Phase; Reversed Phase; Size-Exclusion. **Polarography:** Inorganic Applications; Organic Applications. **Polymers:** Synthetic. **Thin-Layer Chromatography:** Overview. **Voltammetry:** Organic Compounds.

### Further Reading

Leonard RG (1992) In: Smyth MR (ed.) *Chemical Analysis in Complex Matrices*, pp. 149–191. London: Ellis Horwood.  
Leonard RG, Brennan MC, Quigley K, and Heatley DF (1992) Analysis of anaerobic sealants and adhesives. *Analytical Proceedings* 29: 393–396.  
Skeist I (1989) *Handbook of adhesives*. New York: Von Nostrand Reinhold.

## AFFINITY CHROMATOGRAPHY

*See* LIQUID CHROMATOGRAPHY: Affinity Chromatography

## AFM

*See* MICROSCOPY TECHNIQUES: Atomic Force and Scanning Tunneling Microscopy

## AIR ANALYSIS

Contents

**Sampling**

**Outdoor Air**

**Workplace Air**

### Sampling

**J McQuaid**, University of Leeds, Leeds, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Surekha Devi, pp. 50–55, © 1995, Elsevier Ltd.

### Introduction

For many years, man has been concerned with the composition of the air, particularly when anthropogenic activities perturb air quality leading to detrimental health effects. In modern times there has been considerable effort to study the atmospheric processes which occur following emission of these

pollutants. Perhaps the most straightforward method of measuring the components in air is to ‘capture’ a sample and return it to the laboratory for analysis. More complex instruments are also available for deployment to provide on-site measurements.

In the field of air-quality standards, particulate matter (PM) has been divided into a number of size ranges; these are based on their health effects. PM10 are the largest particles (<10 μm) and PM2.5 and PM1.0 getting smaller, as the size of these particles decreases their ability to penetrate deeper into the body results in a serious escalation in their toxicity. Often it is not primarily the particles that are harmful but the organic materials that are often absorbed onto the particle surface, particularly when the particle is carbonaceous in character.

Sampling techniques vary according to the media to be monitored, e.g., gases or aerosols (aerosols are particles and/or droplets suspended in a bulk gas). Aerosols pose the more challenging to determine since there are significant difficulties to be overcome to ensure quantitative sampling of the full range of size fractions. The range of particle size and corresponding methods of sampling are given in Table 1.

Primary Considerations

Sampling of the atmosphere involves making decisions about the following:

- 1. Volume of the sample, variable from 1 ml to several m<sup>3</sup>.
- 2. Rate of sampling, variable from milliliters per minute to liters per minute.
- 3. Duration of sampling, jointly determined by the volume and the rate.
- 4. Collection limitations, specific for the target species under study.

Table 1 Methods of air sampling

Type of sample	Particle size range (μm)	Method
Particulates and aerosols	> 10	Gravitation
	> 5	Centrifugal
	> 1.0	Filtration
	100–1.0	Impaction
	80–0.2	Electrostatic precipitation
	10–0.1	Thermal precipitation
Gaseous	–	Adsorption
	–	Absorption
	–	Grab/whole air

As a general rule, collection efficiency above 80% is considered desirable.

The sampling technique should avoid any modification (physical or chemical) of the collected sample, by way of coalescence or chemical and/or physical destruction. Additional care should be taken to ensure preservation of thermal or photolytically labile species. Similarly, it is necessary to prevent errors caused by factors related to transportation, storage, and time-lag between collection and analysis. Collected samples must be stored in containers made from a suitably inert material such as glass or stainless steel. Plastic containers should not be used for samples containing oil or grease, phosphorus, dissolved oxygen, or organic compounds are to be analyzed. Glass containers should be avoided if silica or fluoro compounds are to be analyzed. It is preferable to follow accepted protocols for sample storage and transportation (e.g., US EPA-TO17).

Sampling Methods and Equipment

Sampling methods are of two types – (1) ambient sampling and (2) at-source sampling – which in turn can be further subdivided into continuous sampling and sequential sampling. Various instruments have been developed, based on the needs of these four types. Ambient air sampling is done at locations exposed to the atmosphere, whereas at-source sampling is carried out at the point of discharge before pollutants are diluted in the atmosphere.

Depending upon the objectives of the analysis, the choice of static sampling sites may vary from a busy street at a height of 0.5 m to a free-floating balloon at an altitude of several hundred meters. It is common to locate the sampling apparatus at a height of 2 m above ground, and it is essential that there are no obstructions, although there are no standard practices. Sampling sites that are accessible but are secure from tampering and have a reliable power supply for the equipment are favored by analysts. During the act of sampling, the sample intake should not be exposed to contamination from specific localized sources. Sites for stack sampling, for example, need to be selected keeping in view the safety of the analyst, easy access to the stack interior, satisfactory flow distribution, and convenience in adjusting the equipment. Generally, a location in a vertical flue (stack) sited eight flue equivalent diameters upstream from a flow disturbance caused by a bend, inlet, or outlet is considered ideal. For a rectangular cross-section, sample points are located at the centroids of smaller equal-area rectangles, and for a circular cross-section at the centroids of equal-area circular segments.

The source sampling can be carried out either when the source is stationary or when it is mobile. Stationary sources are of four types: (1) steady and uniform, (2) steady but not uniform, (3) unsteady but uniform, and (4) unsteady and nonuniform. The first type requires only one sample; the second requires composite sampling at several location; the third demands one sample spread over the entire operation; and the last type calls for a composite and repeated sampling.

Sampling of exhaust emissions from mobile vehicular and aircraft sources is conditioned by the mode of operation of the engine; and exhaust emission tests are usually performed by operating the vehicle on a dynamometer. For vehicular emission sample acquisition generally two types of engine dynamometers and one chassis dynamometer are used. Engine dynamometers generally used are either electric or eddy current absorption dynamometers. Other types such as water brake and fan brake dynamometers are not satisfactory and hence not commonly used. For stationary sources high-volume samplers can be used for sample collection.

The four components that form a sampling system (train) are the inlet/transmission, collection, flow measurement, and pump/air-moving. This allows the 'cleanest' route for the air avoiding as many potential sources of contamination/modification. Where the air is being analyzed at the sampling location, the analytical system is ideally placed between the collection and flow measurement sections. Exceptions to this are the collection of air by grab sampling and the collection of pressurized air samples using a clean bellows pump that is placed upstream of the sample canister.

### The Inlet/Transmission Components

The intake device may range from a thin-walled probe to a free vertical tube. Modifications are necessitated by the difficulties caused by adhesion of aerosols to the tube walls, condensation of volatile components within transfer tubes, or reaction of gaseous components with the material of the transfer tube and with the collected particles.

**Flow-measurement components** There are two types of flow-measurement components used in air sampling: volume meters and rate meters. Volume meters, which are of the dry gas meter, wet gas meter, or cycloid type, have an advantage as frequent checks are required in the use of rate meters to ensure accurate volumes. The advantage of rate meters is their compact size. They can be of several types, ranging

from venturimeters, orifices, nozzles, rotameters, pitot tubes, or turbines to hot-wire anemometers.

**Pumps/air-moving components** The air-moving component of the sampling system draws the sample through the system by creating a vacuum using a mechanical blower, aspirator, or hand-operated pump. Mechanical blowers are used for continuous flow at medium to high rates. Electrically powered pumps are preferable for avoiding contamination, likely to be caused by petrol-driven pumps. Another type of air-moving device is based on the principle of the centrifugal action of an impeller. In addition to these, liquid displacement vessels, evacuated containers, automobile intake manifolds, and expanding plastic bags and syringes are used for gaseous sample collection. In some applications air is pumped into the sampling system: this results in a positive pressure within the sample train, and any leaks are outward. This is particularly important in trace level sampling. In some applications air is pumped into the sampling system; this results in a positive pressure within the sample train and any leaks are outward, thus avoiding any possible contamination. This is particularly important when sampling at trace levels.

**Collectors** The components of sampling systems used for collection of aerosols and gaseous materials differ from each other.

*Collection components for aerosols and particulates* Isokinetic sampling is desirable in the case of aerosols. Researchers have studied errors occurring in anisokinetic sampling and have developed formulae for correcting the results. Anisokinetic sampling may lead to an inaccurate particle size distribution. When the sampled particles are smaller than 3–5  $\mu\text{m}$  in diameter, the requirements for isokinetic sampling can be relaxed. Isokinetic sampling is not required for stagnant and low-velocity masses. The flow rates involved, irrespective of iso- and anisokinetic sampling, vary between  $0.5 \text{ l min}^{-1}$  and  $750 \text{ l min}^{-1}$ . Difficulties in the sampling of aerosols arise from electrostatic attraction of particles toward non-conducting glass and plastic tubing, as well as the evaporation and hygroscopic nature of liquid aerosols or their reaction with solid particles. The following types of collection device are commonly used for the collection of aerosols and particulates.

**Filtration** Air is filtered for collection of aerosols by drawing the air sample through glass fiber, cellulose, sintered glass, or membrane or granular filters. On accumulation of the aerosol the flow through the filter is restricted, and collection is influenced by

**Table 2** Collection characteristics of selected filters

Filter	Composition	Fibre/pore diameter ( $\mu\text{m}$ )	DOP <sup>a</sup> efficiency (%)
Whatman 41	Cellulose (fibre)	3–20	72
Nuclepore	Polycarbonate (capillary pore membrane)	0.8	90
Microsorban	Polystyrene (fibre)	0.7	99.5
MSA1106B	Glass fibre	0.1–4	99.93
Millipore AA	Cellulose ester (membrane)	0.8	99.98

<sup>a</sup>Based on retention of dioctyl phthalate particles, 0.3  $\mu\text{m}$  diameter, at a face velocity of 0.27  $\text{m s}^{-1}$ .

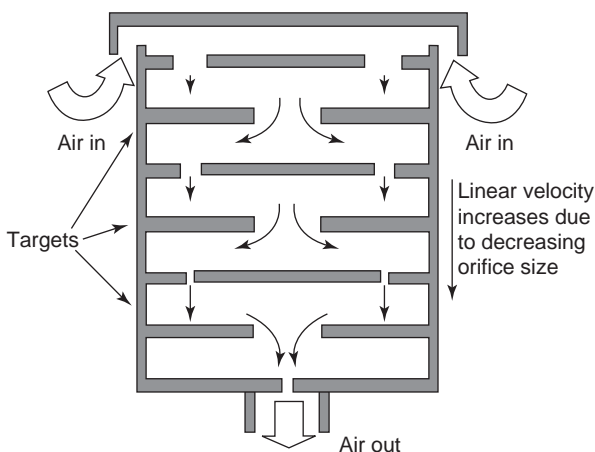
electrostatic forces. The appropriate collection technique is determined by the nature of the filter, its collection efficiency, the pressure drop developed, the background concentration of trace constituents within the filter, and the structural supports used for the filters. Table 2 gives the types of commonly used filters, their composition, and efficiency.

A sample train consisting of a number of filters of decreasing pore size can be used to obtain a rough size distribution of particles. It is imperative that the collection efficiencies of such systems are determined and routinely monitored.

**Impingement/Impaction** In this technique, the air flow is forced through a jet, which increases the velocity of the particles and the carrier gas and changes their direction abruptly. There are wet and dry impingers. In the standard impinger, an orifice having a bore of 2–3 mm is used, through which the tube discharges the sample onto a flat plate of 5–10 mm width. In the midjet impinger, which is generally wet, a 3  $\text{l min}^{-1}$  flow rate is maintained. The Greenburg–Smith impinger is used as a wet or dry impinger, and it operates using a high-velocity air jet of 30  $\text{l min}^{-1}$  that impinges on a glass plate immersed in water or alcohol. The particles are collected in the surrounding fluid. The collection efficiency of this technique is  $\sim 100\%$  for particles greater than 1  $\mu\text{m}$  in diameter.

The mechanism of dry cascade impactors involves a series of sequential jets and collection plates, with increasing jet velocities and decreasing gaps between the collecting plates and the jets, as shown in Figure 1.

**Precipitation** Precipitation of aerosols is achieved using thermal or electrostatic precipitators. Thermal precipitation is effected through a sharp temperature gradient by suspending a wire across a cylindrical collection chamber. This device is especially effective for collection of particles smaller than 5  $\mu\text{m}$ . The suspended particles migrate from a zone of high temperature to a low-temperature one. The thermal force is negligible for a less than 750  $^{\circ}\text{C cm}^{-1}$



**Figure 1** Schematic of typical cascade impactor. (Taken from Reeve R (2002) *Introduction to Environmental Analysis*. Wiley; © Wiley.)

temperature gradient. The flow rate used is 7–20  $\text{ml min}^{-1}$  and is useful in microscopic investigations.

In electrostatic precipitators a high potential difference of 12 kV DC or 20 kV AC is maintained across the plates of a capacitor. The ions produced collide with aerosol particles and charge them; the particles are then collected at the respective electrodes. The device is efficient for the collection of particles ranging from 0.2–80  $\mu\text{m}$  in size at flow rates up to 85  $\text{l min}^{-1}$ . However, it is unsuitable for use in explosive atmospheres.

**Gravitational and centrifugal collection** The Earth's gravitational force of attraction causes a particle to settle at a rate dependent on the fluid drag forces opposing gravity. The rate of settling of the mass is dependent on the momentum of the particle which, in turn, is governed by the size and the flow velocity; in general, those of <7  $\mu\text{m}$  diameter will remain suspended.

In a centrifugal collector the air stream moves in a cyclonic manner, resulting in a net centrifugal effect on the particle, causing it to move outward to the collection surface, from it drops into a dust collection



chamber. The denuded gas leaves through a pipe extending through the top to the center of the cyclone. The collection device is efficient for particles having a  $>5\mu\text{m}$  diameter size. However, for collection of smaller particles the efficiency of collection can be improved by increasing the inlet velocity. Both the methods are useful for preliminary separation of large-sized particles.

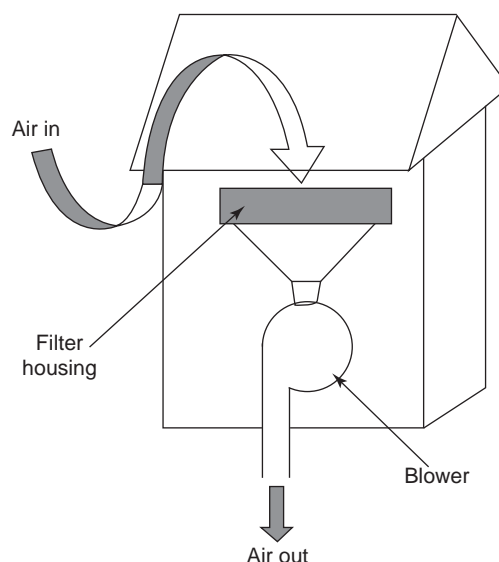
Naturally settling dust particles are collected using a dustfall canister with a protection guard called a bird's eye.

### Samplers in common use

**British suspended particulate sampler** In recent years the old British standard method has been superseded by European directives such as EN12341 'Air Quality – Field test procedure to demonstrate reference equivalence of sampling methods for the PM<sub>10</sub> fraction of particulate matter. The determination of smaller size fractions (PM<sub>2.5</sub>) is also covered by a further directive. The US EPA have similar standard reference methods for particulate material (USEPA 40 CFR part 50). An early method was to simply compare the color of a filter paper through which a volume of air was drawn to an incremental gray scale (16 shades from white to black); this was then converted into an integrated particle loading with reference to the size cut-off offered by the pore size of the filter used. This was known as the 'black smoke index' method.

**High volume sampler for suspended particulates** A high flow rate ( $10\text{--}15\text{ l s}^{-1}$ ) is used to draw air through a  $20\text{ cm} \times 25\text{ cm}$  quartz fibre filter; this results in a typical sample volume of the order of  $1000\text{ m}^3$ . This sampler is useful for particles  $<100\mu\text{m}$  diameter and requires a protection guard. **Figure 2** illustrates a simple schematic of a typical high volume sampler. It is common for such samplers to include an impactor inlet to provide access to a specific size fraction such as PM<sub>10</sub> or PM<sub>2.5</sub> which are both monitored under modern air-quality legislation. Glass filters are used for their gradual headloss buildup and nonhygroscopic characteristics. In order to obtain good quantitation, high volume samplers require regular calibration of the flow system. It is quite common for a polyurethane foam (PUF) plug to be inserted downstream of the filter in order to collect smaller particles and, to some extent, semivolatiles.

**Tape filter samplers** This is useful for taking a series of sequential samples for predetermined sampling periods. A common sampler consists of an intake nozzle, a sampling head to hold and seal the filter tape, spools to collect the filter tape, an interval timer



**Figure 2** Typical-high volume sampler.

and clock, a flow meter, and a vacuum blower. It is operated at a  $7\text{ l min}^{-1}$  rate with cellulose filter tape for sequential sampling periods of 1–4 h. This allows the collection of samples for the black smoke index.

**High-volume cascade impactors** This technique consists of a stack of five 30 cm-diameter plates, each with 300 jet orifices, mounted over a standard high-volume glass fiber filter. The collection medium is perforated sheets of aluminum foil or glass fiber mounted on aluminum plates. Particle size fractions greater than  $7\mu\text{m}$  and less than  $1.1\mu\text{m}$  are collected at a flow rate of  $0.56\text{ l min}^{-1}$ .

### Collection components for gases

**General precautions** Air contains gases and vapors in addition to particulates. The collection techniques for gases are different from those used for aerosols. Several basic techniques are used for collecting gases and vapors, but their usefulness depends on the specific conditions of the source. A knowledge of the composition and history of the sample can help in proper selection and modification of the sampling technique to be used.

Most of the general considerations for aerosol sampling are also applicable to gaseous sampling; and an additional consideration relates to the temperature and pressure of the material at the source. Among the important methods used for collection of gaseous materials are the adsorption, absorption, condensation grab, and cryogenic processes.

**Adsorption** This is a surface phenomenon. Under equilibrium conditions at constant temperature, the



volume of gas adsorbed on the collection phase is governed by the relative surface area. The Langmuir adsorption isotherm, which is based upon homogeneous surface characteristics or for more complex multilayer surfaces the Brunauer, Emmet, and Teller isotherm, provides the collection parameters. The rate of adsorption is approximately inversely proportional to the volatility of a gaseous mixture. The commonly used adsorbents are activated carbon, silica gel, alumina, and various chromatographic support phases. Ideally, the adsorbent should have a large surface area, an affinity for polar or nonpolar compounds, a high and predictable retention capacity, desirable desorption properties, and a relative selectivity for atmospheric gases in selective applications. On the other hand it should have no chemical reactivity with the gases to be collected and should not be prone to fracturing, crushing, and flaking.

The release of the collected analytes is achieved either by solvent extraction (e.g., carbon disulfide) or more commonly nowadays by thermal desorption. Since the adsorbent is effectively acting as a chromatographic column, there are breakthrough limits that must be adhered to. This is generally taken as 66% of the volume of sample after which the least retained (most volatile) begins to elute from the outlet of the adsorbent trap. Specifications for the use of adsorbent tubes for sampling of volatile organic compounds (VOCs) are given in US EPA Compendium Method TO17.

**Gas chromatographic phase adsorption** Conventional gas chromatographic support materials coated with a liquid phase and porous polymeric materials are used for adsorption of gaseous samples. This is a versatile and rapidly evolving technique and is used for a variety of gaseous samples: chromosorb 101 and 102 have been used successfully for collection of organic gases and hydrocarbons at sampling rates of  $20 \text{ min}^{-1}$ . In this process, desorption of the adsorbed

phase is achieved by shaking it with hexane or by similar solvent extractions.

Adsorbent tubes can also be used for passive/diffusive sampling. Airborne pollutants diffuse along a short tube at the end of which is the adsorbent material. The rate of diffusion along this tube is governed by Fick's law of diffusion. This is a low cost and effective method of long-term monitoring, where tubes are typically deployed for a week. Diffusion tubes also have benefits since they require no power and are simple to use. This method is not limited to VOCs, however as inorganic species such as ozone, nitrogen dioxide, and sulfur dioxide can also be monitored using diffusion tubes.

**Absorption** In absorption collection, gas molecules are dissolved in a liquid collecting phase. The efficiency of the process depends on the contact surface area and is achieved by transforming the air stream into small, finely dispersed bubbles with a relatively long duration of travel through the absorbent. Absorption, like adsorption, is efficient at lower temperatures for volatile components. Table 3 shows the efficiency of absorption for various gaseous components. The commonly used absorption devices are fritted glass scrubbers, impingers, packed columns, countercurrent scrubbers, and atomizing scrubbers. Over the years these devices have undergone a number of modifications. Other frit materials are plastics and ceramics. Glass tubing with small holes can also be used. The absorbers are designed to detain for the rising bubbles for longer durations. Impingers are less efficient and are of the type discussed earlier.

**Condensation** The condensation method (cryotrapping) faces the primary problem of condensation of water containing components of interest.

Sequential traps at progressively lower temperatures, with the first designed for water vapor, can overcome this difficulty. To avoid combustion hazards,

**Table 3** Absorption efficiencies for various media and gaseous samples

Chemical	Absorption efficiency		
	Absorbent in gas washing bottle	Gas washing bottle efficiency (%)	Stack concentration ( $\text{mg l}^{-1}$ )
Sulfuric acid <sup>a</sup>	Water	99.8	72
Hydrochloric acid	Water	100	47
Perchloric acid	Water	99.8	41
Acetic acid	Water	99.5	1065
Nitric acid	Water	100	182
Ammonia	Water	84.3	103
Chlorine	2% $\text{Na}_2\text{CO}_3$	77.0	103
Ethylenediamine	Water	99.6	211

<sup>a</sup>Although depends on the particle size of aerosol.

the process is carried out at liquid oxygen temperature ( $-183^{\circ}\text{C}$ ). Coolants are placed in wide-mouthed Dewar vessels, and double-walled collector flasks are immersed in them. Coolants used are mixtures of water, ice, salt, dry ice, liquid air, and liquid oxygen.

**Grab/whole-air sampling** There is a wide range of materials from which sampling vessels are made: from polymeric bags to glass and metal containers. Bags are filled either by air being pumped into them or they are placed in an airtight container and the air around them is removed by a pump; when the inlet of the bag is opened to the atmosphere, air is drawn into the bag. Rigid containers can be evacuated, and upon opening the inlet, air is sucked in. This is a useful method, as it requires no power. The canister may also be fitted with a fixed orifice to prolong the sampling time. Vessels with an inlet and outlet can have the sample flushed through and closing the taps seals the sample. A final method to fill canisters is to pump air into them using a clean bellows pump.

It must be remembered, however, that whole air samples by their nature contain all the constituents from the air and thus reactions may occur after sampling; in particular, ozone reacts readily with many atmospheric species. Care must be taken to avoid low temperatures during storage/transportation since less volatile components may condense onto the internal walls. Often when trace analysis is to be conducted, internal surfaces are passivated either by electropolishing or by application of an inert silica coating (e.g., Silcosteel<sup>®</sup>, Restek Corp., US) in order to reduce wall losses. Evacuated canisters can be stored in preparation for use in case of gas release from chemical/industrial facilities. Emergency response services have even equipped remote-control helicopters with single

grab sample canisters that can be flown into highly hazardous areas following gas leaks.

**Source sampling** For collection of gaseous air from mobile (vehicular) sources, an exhaust gas sampling system has been developed in the United States in which the entire exhaust cycle is sampled and diluted to prevent condensation.

Sample collection equipment (pumps, flow meters, etc.) are similar to those discussed for aerosol collection.

**See also:** **Gas Chromatography:** Overview; Column Technology; Physicochemical Measurements. **Sampling:** Theory.

## Further Reading

- British Standard BS174 part 11 (1993) Determination of a black smoke index in ambient air. ISO 9835: 1993.
- Cohen B and McCammon C (eds.) (2001) *Air sampling Instruments*, 9th edn. Cincinnati: ACGIH.
- Colbeck I (ed.) (1998) *Physical and Chemical Properties of Aerosols*. UK: Blackie.
- Faith WL and Atkisson AA (1972) *Air Pollution (Environmental Science and Technology)*. Wiley.
- Hinds WC (1999) *Aerosol Technology: Properties, Behaviour and Measurement of Airborne Particles*, 2nd edn. New York: Wiley.
- <http://www.wmo.ch/>
- <http://www.epa.gov/>
- <http://www.skinc.com/>
- Lodge JP Jr., (ed.) (1998) *Methods of Air Sampling and Analysis*. 3rd edn. CRC.
- Reeve R (2002) *Introduction to Environmental Analysis*. Wiley.
- Wight GD (1994) *Fundamentals of Air Sampling*, Chapman and Hall.
- Young RJ and Perry R (1997) *Handbook of Air Pollution Analysis*. Chapman and Hall.

## Outdoor Air

**R M Harrison**, The University of Birmingham, Birmingham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by R M Harrison and J D Peak, pp. 55–59, © 1995, Elsevier Ltd.

## Introduction

The general atmosphere comprises nitrogen (78%), oxygen (21%), and argon (1%). In addition to these

gases there are a number of other species present at trace levels. These species can be of natural and/or anthropogenic (manmade) origin and in recent years significant changes in the composition of the atmosphere have been occurring as a result of large-scale emissions of certain pollutants.

These emissions are giving rise to problems on a global scale, such as the enhanced greenhouse effect and depletion of ozone from the stratosphere. In addition, regional problems such as localized ozone creation and smog formation have also been

observed. Although the concentrations of various specific pollutants vary with both time and location, it is generally true that they are highest in urban and industrial areas, although there are exceptions. As an example, certain 'secondary' pollutants such as ozone ( $O_3$ ) that are derived chemically from primary pollutants such as nitrogen oxides ( $NO_x$ ) and volatile organic compounds take time to form within the atmosphere and thus tend to be most concentrated at an appreciable downwind distance from urban and industrial areas.

The need to develop reliable analytical methods to determine the levels of both gaseous and particulate pollutants has gathered considerable impetus in the last 20 years or so and reliable, specific, rapid-response techniques are now available for many pollutants. In order to review the available methodology for the analysis of the various priority pollutants, it is necessary to classify them according to the physical form in which they are generally present in the atmosphere. The two classes of pollutant form are gaseous pollutants and particulate species.

## Gases

While it is not practicable to describe methods for all trace gases, some of the more important analytes and methods have been selected.

### Nitrogen Oxides

Nitrogen oxide (NO) and nitrogen dioxide ( $NO_2$ ) are collectively known as nitrogen oxides ( $NO_x$ ). They are generally grouped together because most anthropogenic  $NO_2$  derives from emissions of NO and they interconvert readily within the atmosphere. Major anthropogenic sources are vehicular emissions and stationary combustion of fossil fuels. Of the two compounds,  $NO_2$  has been implicated in a variety of respiratory ailments in humans and is generally regarded as the more important of the two species.

There are several methods available for the analysis of  $NO_x$ . All chemical methods involve the oxidation of NO to  $NO_2$ , while the chemiluminescence method involves the thermal decomposition of  $NO_2$  to NO. Three methods that are widely used are the Christie arsenite procedure, the diffusion tube method, and chemiluminescence. The principles behind these methods will be described briefly.

The arsenite method involves bubbling the air sample through a mixture of sodium hydroxide and sodium arsenite solution to form a stable solution of sodium nitrite. The nitrite ion produced is then reacted with phosphoric acid, sulfanilamide, and N-1-(naphthyl) ethylenediamine dihydrochloride to form an azo dye.

The analysis method can be modified for NO analysis if, after the  $NO_2$  is first removed from the gas stream, the remaining NO is then oxidized to  $NO_2$  with solid chromic oxide. The  $NO_2$  is then determined with arsenite as described above. This method can suffer an interference with  $SO_2$  if it is not first converted to sulfate ion ( $SO_4^{2-}$ ) by hydrogen peroxide prior to analysis. The method is relatively simple and inexpensive and has a reported detection limit of  $\sim 2 \mu g m^{-3}$ .

The diffusion tube method, which is highly selective toward  $NO_2$  but not entirely free of interferences, involves the absorption of  $NO_2$  by triethanolamine (the  $NO_2$  is converted to the nitrite ion in the process) after the sample gas has diffused passively along the plastic tube. The nitrite ion is then determined colorimetrically using a modified Griess-Saltzman reagent. This is a relatively simple method that, despite being potentially subject to a number of interferences from species such as  $SO_2$ , has been shown to give fairly reliable measurements, although it is best regarded as semiquantitative. One disadvantage of the method is that long sampling periods (up to 2 weeks) may be required to achieve reasonable sensitivity.

The chemiluminescence method is accepted as the most reliable and precise method currently available for the analysis of oxides of nitrogen. The basis of the method is the reaction between atmospheric NO and  $O_3$  (generated within the instrument). This reaction generates light (chemiluminescence), which is detected by a photomultiplier and converted to a concentration by calibration of the instrument with standard gas mixtures. To determine the  $NO_2$  content of the air,  $NO_2$  must be converted by thermal decomposition to NO and the instrument then measures  $NO_x$ .  $NO_2$  is then determined by difference between the  $NO_x$  and NO channels.

There is some potential for interference in the  $NO_x$  channel (and hence in the  $NO_2$  measurement) from species such as nitric acid. Detection limits for most commercial instruments are of the order of  $1 \mu g m^{-3}$ , although higher-sensitivity instruments are available. In recent years, instruments capable of measuring NO and  $NO_2$  at background ( $ng m^{-3}$ ) levels have been developed, based upon the NO/ $O_3$  reaction and also the  $NO_2$ /luminol chemiluminescent reaction.

### Sulfur Dioxide

Sources of sulfur dioxide ( $SO_2$ ) include fossil fuel combustion and smelting of nonferrous ores. It has been implicated in various respiratory ailments in humans and is known to play a major role in the formation of acid rain through its conversion to sulfuric acid in the atmosphere. There are a number of

methods available for the analysis of gaseous SO<sub>2</sub>; some of these suffer from serious interferences. Several of the more widely used methods are described below.

The methods involving collection in hydrogen peroxide are relatively simple to carry out. The basis of these methods is to absorb SO<sub>2</sub> in a solution of dilute H<sub>2</sub>O<sub>2</sub>, which converts it to sulfuric acid. The atmospheric concentration of SO<sub>2</sub> is then determined indirectly either by determining the free acid (H<sup>+</sup> ion) concentration by conductivity or pH measurements or by titration (acidimetric methods), or, alternatively, by determining the sulfate concentration by reaction with color-forming reagent followed by spectrophotometric measurement (colorimetric method), or by ion chromatography. One major problem with the acidimetric procedure is that any gas that can give an acid in solution and any alkaline gaseous species that will neutralize the acid (i.e., NH<sub>3</sub>) will interfere. The colorimetric and ion-chromatographic methods are more specific. The limits of detection for both methods are  $\sim 5 \mu\text{g m}^{-3}$  for a 24 h sample under ideal conditions.

A better colorimetric procedure involves bubbling air containing SO<sub>2</sub> through a solution of potassium tetrachloromercurate. The stable dichlorosulfite-mercurate ion that is formed in this reaction is then reacted with formaldehyde and bleached pararosaniline to form the intensely colored (red-purple) species pararosaniline methylsulfonic acid. The pH of the final solution is adjusted to 1.6 by the addition of phosphoric acid to the color reagent and the concentration of SO<sub>2</sub> is determined by measuring the product colorimetrically at  $560 \pm 2 \text{ nm}$ . This method has the advantage of being quite specific and has a quoted detection limit of  $5 \mu\text{g m}^{-3}$  for a 1 h sample.

The widely used continuous instrumental method for determination of SO<sub>2</sub> is based upon gas-phase fluorescence. Pulsed ultraviolet (UV) light (214 nm) is used to irradiate the air sample, which flows continuously through an optical cell. The SO<sub>2</sub> re-emits fluorescent radiation at 340 nm that is detected by a photomultiplier tube (PM), with the signal obtained being converted to concentration. This method is almost specific with a detection limit of  $\sim 1 \mu\text{g m}^{-3}$  in commercially available analyzers.

A second physical method that has been used for analysis of SO<sub>2</sub> in air is flame photometry. The sample is burned in a hydrogen-rich flame with light emission from the S<sub>2</sub> species at 394 nm being detected by a PM tube. However, since this wavelength is characteristic of sulfur, any other sulfur-containing species can interfere, and this method has largely been superseded by the gas-phase fluorescence method described above.

Sulfur dioxide (also nitrogen oxides) can also be determined by the differential optical absorption

spectroscopy (DOAS) method. This will be described later in the section on remote sensing.

### Carbon Monoxide

Carbon monoxide (CO) is predominantly produced from vehicular emissions, although it can be emitted in smaller amounts from other processes that involve the combustion of organic material. Its effects on health are primarily from its ability to displace O<sub>2</sub> from hemoglobin, and give rise to both morbidity and mortality due to the body's deprivation of oxygen.

There are two methods that are predominantly used to analyze CO; these are based on infrared (IR) absorption and electrochemistry. The IR technique is based on the fact that CO will absorb light at  $4.67 \mu\text{m}$  ( $2165 \text{ cm}^{-1}$ ). The CO concentration is then determined from the extent of absorption of the sample. There are two types of analyzer design, known as nondispersive and gas filter correlation analyzers. Interferences from CO<sub>2</sub> and water vapor can be overcome by instrumental design and are generally not significant. Reported detection limits are  $<0.5 \text{ mg m}^{-3}$  for this technique. The electrochemical cell technique is based on the electrochemical detection of CO as it is oxidized to CO<sub>2</sub>. Interferences from other oxidizable gases can be minimized by the use of special inlet filters and the detection limits obtained are comparable to the IR technique described earlier.

### Other Gas-Phase Species

In addition to the three important species already described, many other gas-phase pollutants are also considered important. These include nitric acid (HNO<sub>3</sub>), nitrous oxide (N<sub>2</sub>O) – which is a known greenhouse gas – as well as secondary pollutants such as ozone and peroxyacetylnitrate (PAN). A method that has been used widely for the determination of nitric acid and other species such as nitrous acid (HONO) and SO<sub>2</sub> is the annular denuder, in which the sample air is drawn through an annular glass tube (or series of tubes) that have been coated internally with a compound designed to trap the analyte species. After sampling, the denuder is extracted and the extract is analyzed by ion chromatography or other relevant procedure. Thus, for example, gaseous acids (HNO<sub>3</sub>, HCl, HONO, etc.) are collected in sodium carbonate, and ammonia is collected in oxalic or phosphorous acid. Nitrous oxide analysis is generally achieved by gas chromatography (GC) with electron capture detection. Only physical methods are feasible owing to the inert nature of this species and the lack of any chemical reactions that can be used.

Ozone, an important secondary pollutant, can be analyzed by DOAS (see Remote Sensing) and also by chemiluminescence, where the sample air is mixed with ethylene in a special flow cell at atmospheric pressure. The two gases react to produce an emission in the region of 400–600 nm and the light produced is monitored with a PM tube; the concentration of  $O_3$  can then be determined from the signal intensity. Interferences are negligible and the observed detection limits are  $\sim 1$  ppb in commercial instruments. The preferred technique for ambient air is UV light absorption in a folded path cell at 254 nm. Detection limits are  $\sim 1$  ppb. There is a potential interference from other gases absorbing at this wavelength (e.g., aromatic hydrocarbons). This is overcome by cycling the instrument between two modes: in one the absorbance of air passing through an ozone scrubber is measured, while in the other mode the absorbance of unscrubbed air is determined. The difference is due to the absorption by ozone. Calibration is via the absorption coefficient for ozone, as standard atmospheres are difficult to generate repeatedly. PAN can be determined by several methods. One such uses GC separation followed by an electron-capture detector. There are no known interferences to this method and the detection limits are  $\sim 1 \mu\text{g m}^{-3}$ , or lower if the PAN is cryogenically preconcentrated. In an alternative procedure, after chromatographic separation, PAN is thermally converted to  $NO_2$  and determined by the chemiluminescent reaction of  $NO_2$  with luminol.

## Particulate Matter

Particulate pollutants are emitted from many sources. Additionally, particles are formed in the atmosphere by both chemical and physical conversions from natural and anthropogenic gaseous substances. Particulate pollutants cover a size range from  $<10$  nm to  $>100 \mu\text{m}$ . The major proportion of the aerosol below  $1 \mu\text{m}$  is generally man-made, including sulfates from  $SO_2$  oxidation and carbon from vehicle exhausts, for example. Particles of a greater size are frequently natural (e.g., soil-derived, marine aerosol) but this division cannot be regarded as absolute.

Particulate matter is collected (sampled) either from suspension in the air (usually by filtration) or by collection of depositing particles as they fall out of the atmosphere under gravity or by turbulent or diffusive deposition. In general, airborne particles collected by air filtration are  $10 \mu\text{m}$  or less in diameter, while particles greater than this size predominate in deposit gauge collectors. In this section, the analysis of several important pollutants

that are generally present in a particulate from in the air will be described, together with a brief description of the method used to collect them.

### Mass

Most developed countries have legislated limit values of the concentrations of the mass of particulate matter in the atmosphere. These regulations are generally framed in terms of particles within specific size ranges termed  $PM_{10}$  and  $PM_{2.5}$ . The former refers to the mass of particles sampled through an inlet with a 50% cut-off efficiency at  $10 \mu\text{m}$  aerodynamic diameter (i.e., effectively all particles smaller than  $10 \mu\text{m}$  diameter), while for  $PM_{2.5}$  the 50% cut point for the sampler inlet is at  $2.5 \mu\text{m}$ . Both the US and European (CEN) reference methods for measuring particle mass depend upon sampling a known volume of air through a high-efficiency filter that is weighed under controlled conditions of temperature and humidity before and after sampling. Continuously reading instruments include the  $\beta$ -absorption gauge (based on absorption of  $\beta$ -radiation by collected particles) and the tapered element oscillating microbalance. Because these instruments often precondition the air stream (e.g., by heating) before sampling, the measurements may differ from those made by the reference methods which do not permit such conditioning.

### Metals

Metal-containing particles are generally collected by drawing air through a filter and subsequently analyzing the filter on which the particles are trapped. The material used for the filter is dependent on the species that is to be collected. The choice of filter material is also related to the requirements of the subsequent analysis step (i.e., low blank content). Two important metals that have environmental significance are lead (Pb) and cadmium (Cd). There are several methods available for the analysis of these two species.

The most commonly used technique is atomic absorption spectroscopy (AAS). Lead- and cadmium-containing particles are collected on membrane filters (usually Teflon or cellulose ester), extracted into strong acid solution, and analyzed by AAS. If the procedures are optimized there are no interferences and detection limits are of the order of  $10 \text{ ng m}^{-3}$  for lead and  $1 \text{ ng m}^{-3}$  for cadmium in a 24 h sample. The figures given are for electrothermal atomization using a graphite furnace; atomization using a flame results in reduced sensitivity and consequently higher limits of detection. Multielement analysis of metals

dissolved from airborne particles is often achieved by use of inductively coupled plasma-optical emission spectroscopy, or if high sensitivity is required, inductively coupled plasma-mass spectrometry.

For X-ray fluorescence the samples are collected by filtration and placed in the X-ray spectrometer without sample pretreatment. When irradiated with X-rays, the metals re-emit secondary X-rays at wavelengths characteristic of the individual metals. There are little interferences and good limits of detection can be obtained.

### Organic Species

Methods for the analysis of particulate organic air pollutants are of necessity complex. The majority of methods require filter collection of the sample followed by extraction and a chromatographic separation. For example, in the analysis of polycyclic aromatic hydrocarbons (PAHs) the extraction and analysis procedure involves a Soxhlet (or other) solvent extraction off the glass fiber or Teflon filter used to collect the sample, followed by analysis of the extracted sample by liquid chromatography. The detectors are either UV absorption or fluorescence monitors. Alternatively, the extracts of PAHs may be analyzed using gas chromatography-mass spectrometry techniques. Detection limits in the picogram ( $10^{-12}$  g) range have been reported.

### Inorganic Ions

Inorganic particulate ions such as nitrate and sulfate, which are attracting considerable interest owing to the involvement of these species in acid rain studies, are generally collected using filters. The use of polytetrafluoroethylene (PTFE) membrane filters is generally recommended, especially for nitrates, owing to the low background of soluble nitrate ion in this paper. A number of colorimetric methods may be used for nitrate analysis, but ion chromatography is nowadays almost universally used. PTFE filters are also appropriate for sulfate collection. It has been shown that other media such as glass fiber are unsuitable for sulfates owing to artifact formation from  $\text{SO}_2$  reactions on the filter surface. The filter is extracted with a known quantity of aqueous sodium carbonate-sodium bicarbonate buffer containing 10% (v/v) isopropanol and either mechanically shaken or placed in an ultrasonic bath for 1–2 h to complete the removal of the analytes from the filter. The extracts are then analyzed by ion chromatography. Detection limits of well below  $1 \mu\text{g m}^{-3}$  for 3 h samples can readily be obtained using this procedure.

## Remote Sensing

The term 'remote sensing' is often associated with observations of the earth from orbiting satellites. However, in recent years the use of remote sensing instrumentation has been applied to the monitoring of air pollution in the atmosphere. The term itself refers to the fact that the instruments used to make measurements can do so not only at the site at which the instrument is located but also at points remote from it. Several types of instruments have been used for air pollution monitoring and some of the more important ones will be discussed briefly in this section.

### Single-Wavelength Lidar

The principle of lidar is that a laser pulse is fired into the atmosphere and as it proceeds along its path, radiation that is scattered by aerosol and other particles is directed back toward the laser where it is collected with a telescope and measured with a detector. The lidar system can be operated either in single- or multi-wavelength mode (by using a tuneable laser) so as to detect a number of different species according to the attenuation they cause at specific wavelengths. Several practical systems have been developed to monitor species such as  $\text{SO}_2$  and  $\text{O}_3$  but the instrumentation required is complex and very expensive. Analysis of the signal also requires considerable expertise.

### Differential Optical Absorption Spectroscopy

The DOAS technique originally developed by Platt, Perner, and co-workers uses a broadband light source with an emission in the wavelength region from 200 to  $\sim 700$  nm. This light, which is collimated by a parabolic mirror, passes through the outside air over a pathlength that can be up to 10 km before it is recaptured with a second parabolic mirror and focused to the detector using an optical fiber. By varying the wavelength of the light source the concentrations of specific species can be determined as the light beam is absorbed by trace species that can absorb light at that particular wavelength. The concentrations of these species can then be derived from Beer's law.

Several instruments have been developed to monitor air pollutants such as  $\text{SO}_2$ ,  $\text{NO}_2$ , HONO, and  $\text{O}_3$  and these are now commercially available. Although these once seemed likely to revolutionize air quality monitoring, their uptake in practice has been very limited. Most statutory air quality networks require point monitors, and long-path instruments are not accepted.

See also: **Atomic Absorption Spectrometry:** Electrothermal. **Atomic Emission Spectrometry:** Flame Photometry. **Cadmium. Carbon. Chemiluminescence:** Overview. **Fluorescence:** Environmental Applications. **Gas Chromatography:** Environmental Applications. **Laser-Based Techniques. Lead. Nitrogen. Ozone. Polycyclic Aromatic Hydrocarbons:** Environmental Applications. **Remote Gas Sensing:** Overview. **Spectrophotometry:** Inorganic Compounds. **Sulfur. X-Ray Fluorescence and Emission:** X-Ray Fluorescence Theory.

## Further Reading

- Baron PA and Willeke K (2001) *Aerosol Measurement*, 2nd edn. New York: Wiley-Interscience.
- Butterwick L, Harrison RM, and Merritt Q (1991) *Handbook for Urban Air Improvement*. Brussels: Commission of the European Communities.

- Harrison RM and Perry R (eds.) (1986) *Handbook of Air Pollution Analysis*, 2nd edn. London: Chapman and Hall.
- Keith LH and Walker MM (1995) *Handbook of Air Toxics: Sampling Analysis and Properties*. Boca Raton, FL: CRC Press.
- Lodge JP Jr.(ed.) (ed.) (1989) *Methods of Air Sampling and Analysis*, 3rd edn. Michigan: Lewis Inc.
- Platt U and Perner D (1980) Direct measurements of atmospheric CH<sub>2</sub>O, HNO<sub>2</sub>, O<sub>3</sub>, NO<sub>2</sub>, and SO<sub>2</sub> by differential absorption in the near UV. *Journal of Geophysical Research* 85: 7453–7458.
- Possanzini M, Febo A, and Liberti A (1983) New design of a high performance denuder for the sampling of atmospheric pollutants. *Atmospheric Environment* 17: 2605–2610.
- Robinson R (2003) Atmospheric monitoring techniques. In: Hewitt CN and Jackson A (eds.) *Handbook of Atmospheric Science*. Oxford: Blackwell.

## Workplace Air

**J Yu and R M A Hahne**, University of Washington, Seattle, WA, USA

© 2005, Dow Chemical Company. Published by Elsevier Ltd. All Rights Reserved.

## Introduction

As a basis for the prevention and control of occupational hazards, workplace air analysis has undergone rapid development in the field of analytical science during the last two decades. It paralleled the establishment and development of occupational exposure guidelines for hazardous materials set by government/advisory agencies.

Workplace air analysis can be distinguished from other types of analysis: the selection of sampling medium, time, and location are crucial steps in workplace air analysis; transportation and storage of the samples may also influence the measurement results; and, almost always, real-time monitoring or quick turnaround analysis time is required. Several international organizations have issued guidelines and standards on methods for measuring air contaminants in the working environments, such as the World Health Organization, the International Union of Pure and Applied Chemistry, and the International Organization for Standardization. Most national government agencies/organizations, such as Occupational Safety and Health Administration (OSHA), USA; National Institute of Occupational Safety and Health (NIOSH), USA; The Commission of the European Union; and the Health and Safety

Executive (UK) have also established more specific standard methods for the sampling and measurement of airborne contaminants in workplace environments.

In this article, regulatory requirements for workplace air quality will be reviewed, as will general principles of workplace air analysis of both real-time experiments and those performed in the laboratory. Analytical methods for contaminants concentrated from air samples will be presented and discussed.

## Legislation

Although it has long since been recognized that workplace air contaminants can be hazardous to workers' health, a significant driving force for workplace air analysis is governmental regulations, apart from the nonregulatory guidelines and recommendations such as the threshold limit values (TLVs) of the American Conference of Governmental Industrial Hygienists (ACGIH), which specifies concentrations of airborne contaminants in the workplace air to which nearly all workers may be repeatedly exposed without adverse health effects. TLVs may be expressed as full-shift, time-weighted average (TWA) concentrations in the breathing zone of a worker (TLV-TWA), as short-duration concentrations (~15 min), or as 'instantaneous' ceiling concentrations that are not to exceed TLV-C at any time.

In the United States, permissible exposure limits (PELs) of the Occupational Safety and Health Administration (US Department of Labor) exist for over 400 different organic and inorganic materials.

In Canada, individual provinces have jurisdiction for the establishment and enforcement of workplace airborne contaminant levels in private industries, and most of the provinces have established their own occupational exposure levels.

In Germany, the MAK (Maximale Arbeitsplatz Konzentration – maximum workplace concentration) values established for workplace air have been adopted by the Ministry of Labor and are mandatory. MAKs are the maximum concentrations allowed for exposures of repeated or extended periods of time. There are currently ~150 MAK values; concentrations exceeding the MAKs are allowed for reduced durations or reduced frequencies. In addition, technical guidance concentrations (Technische Richtkonzentrationen) have been established for substances identified as carcinogens and potential carcinogens.

In the United Kingdom, there are two types of occupational exposure limits: occupational exposure standards (OESs) and maximum exposure limits (MELs). An OES is set at a level at which there is no indication of risk to health, while the MELs are set for materials for which serious health implications exist, but which due to socioeconomic factors must have a numerically higher value in order for the controls associated with certain uses to be regarded as reasonably practicable.

Many other countries have OELs with varying degrees of enforcements, updated annually.

Measurements of workplace air are frequently made to demonstrate compliance with legal or other standards, but may also be made for other purposes, including identifying unknown potential hazards, obtaining data for epidemiological (exposure-response) studies, and demonstrating the efficiency of control measures.

## General Principles

The type of air contaminants that occur in the workplace depends on the raw materials used and the processes involved. Air contaminants can be classified into two groups based on their physical properties: (1) aerosols (a suspension of liquid or solid particles in the air), and (2) gases/vapors.

The analysis of workplace airborne contaminants depends on the specificity, sensitivity, accuracy, reproducibility, and stability of the analytical methods, similar to that in other chemical and physical analyses.

As OELs continue to decline, the analytical method of choice must be sensitive enough to reliably quantitate the workplace air contaminant at or below the designated exposure level. As an example of potential problems associated with lack of sensitivity, when the ACGIH announced its recommended

change in the short-term exposure limit for formaldehyde in 1990, there was only a single device commercially available that had the sensitivity to measure formaldehyde at that concentration for such short durations.

The need for accuracy in a workplace monitoring method is vital, since the measurements need to reflect the actual conditions in the workplace. On the other hand, given the temporal and spatial variability in most airborne contaminant concentrations in the workplace, it is generally not required that the monitoring method be highly accurate, i.e. within a few percent of the 'true value'. The NIOSH (USA) recommends that (1) the overall bias of a monitoring method be <10% of the values determined by a well-characterized independent method, and (2) that the overall precision of sampling and analysis should be such that the total error is <25% in at least 95% of the samples analyzed, based on the analysis of 6–10 samples at each of three or four concentration levels.

The analytical methods employed for workplace air analysis can be categorized into two general categories:

1. Direct reading instruments for use in the field.
2. Sample collection with subsequent laboratory analysis.

The direct reading instrument can provide data on exposures in the field. There are several reasons why one chooses direct reading instruments:

1. Quick turnaround time; there is no need to wait for long, often costly, laboratory analyses.
2. Can identify high short-term exposure that may lead to acute effects, e.g., facilitate confined space entry, and also provide estimates of long-term exposures with correct sampling procedures.
3. Provide some alarm functions for evacuation or remedial action.

On the other hand, field sampling followed by laboratory analysis provides the kind of precision and accuracy that no field/onsite measurement can consistently provide.

## Aerosol (Particulate Matter)

Airborne particulates can be either solids or liquids. Dust, fumes, smoke, and fibers are dispersed solids; mists and fogs are dispersed liquids.

### Sample Collection

Particulate matter ranges in size from visible to microscopic. Particles can be sampled to determine the



total quantity, or only the respirable fraction. The respirable particles are those that are retained in the lung and are generally considered to be of an aerodynamic size below 10  $\mu\text{m}$ .

**Filters** For total particulate/aerosol sampling, the open- or close-faced 37 mm filter cassette is the most popular in many countries. A newer type of sampler for inhalable dust is the IOM sampler with a 25 mm filter. Many studies have shown that the IOM sampler provides the best reference for inhalable aerosol. Other samplers that are emerging for this purpose are Respicon, and the Button samplers. There are several types of filters commonly used for this purpose: glass fiber, mixed cellulose ester fiber (MCE), and poly(vinyl chloride) (PVC) filters. They are selected based on their ability to collect material and their suitability for laboratory analysis. For example, PVC filters are normally used for gravimetric analysis and MCEs are used when a chemical analysis is needed followed by digestion or extraction.

**Cyclones** A cyclone is used to collect particles of respirable size, which separates out the large particles from an air stream by the centrifugal motion of the airflow, with the remaining respirable particles being directed onto a collection device such as a filter. Several types of cyclones are commercially available: Dorr–Oliver cyclone, Higgins–Dewell, SKC Aluminum, and SKC GS.

**Multistage Impactors** Separation in this device is based on the momentum of particles of various masses. All the particulates in the air are collected, but are separated by size on a series of multiple stages. **Figure 1** shows four commercially available devices for the size separation and collection of airborne particulates.



**Figure 1** Four commercially available devices for the size separation and/or collection of airborne particulates. From left to right: IOM sampler, button sampler, cascade multistage impactor, cyclone.

## Physical Analysis

There are two common physical analyses that can be performed on the particulate matter collected on filters or other media.

Gravimetric analysis is the most commonly used physical method, in which the filter is weighed both pre- and postsampling using a microbalance. The quotient of the increase in filter weight divided by the air volume drawn through the filter yields the average concentration of the airborne particulate matter during the sampling period.

The second physical method of analysis for particulates is microscopy. Although there are a number of microscopic techniques used for the qualitative and quantitative analysis, the most common technique used for the analysis for workplace air is phase-contrast microscopy (PCM), used in the quantitation of fibers, especially asbestos trapped on a filter. Another type of microscopic method, electron microscopy, including scattering and transmission electron modes, has seen increased popularity due to its superior resolution and magnification power, despite the high cost and other disadvantages. In scanning electron microscopy (SEM), a very fine ‘probe’ of high-energy electrons scans across the surface of the specimen, inducing the emission of secondary electrons or reflection of the backscattered electrons from the primary beam. The intensity of emission of both secondary and backscattered electrons is very sensitive to the angle at which the electron beam strikes the surface, i.e., to topographical features on the specimen. The emitted electron current is collected, amplified, and displayed as a fluorescent image on a cathode ray tube.

## Chemical Analysis

In many cases, aerosol/particulate matter trapped on a filter paper is extracted from the filter using solvents or acids, after which the extract is analyzed. This technique is used for the analysis of: polynuclear aromatic hydrocarbons trapped on membrane filters or solid sorbents, various pesticides from polyurethane foam filters, many metals and their salts, and crystalline silica on PVC filter samples. The polynuclear aromatic compounds and pesticides can be analyzed using either gas chromatography (GC) or high-performance liquid chromatography (HPLC); the metals are analyzed using atomic absorption/emission spectrophotometry (AAS/AES, either flame or flameless, depending on the metal), inductively coupled plasma spectrometry, or ion chromatography. The analysis of crystalline silica is done using X-ray diffraction (XRD), infrared (IR), or visible spectrophotometry.

Ion chromatography can be used for the quantitation of a mixture of anions collected as salts, which might include such substances as fluoride or nitrate ion. Specific ion-selective electrode measurements are also used for the analysis of numerous anions and cations. Special attention should be given to elimination/minimization of interferences from the matrix.

In some cases, the filter is used as both particulate trapping medium and as a substrate for holding a derivatizing reagent that reacts with and stabilizes the particulate matter. For example, 1-(2-pyridyl)piperazine-coated glass fiber filters are used to trap and react with particulate methylene bisphenyl isocyanate to form a derivative that is then analyzed using HPLC coupled with a fluorescence detector.

The direct-on-filter technique for either XRD or IR, in which sampling and analysis is accomplished using the same filter, has been accepted widely as an alternative to the routine analysis following desorption. This is more efficient as it bypasses the intensive sample preparation; less contamination and less sample loss are also added advantages.

There are also other techniques that are used frequently to characterize particulate matter collected from workplace air, for example, X-ray fluorescence (XRF) analysis – both lab-based wavelength-dispersive and hand-held energy-dispersive XRF spectrometry – are used for identification and quantitation of either bulk samples or air filters for numerous chemical elements. The latter is now available with both tube-excitation and radionuclide-excitation sources.

### Direct Reading Aerosol Monitors

Several types of direct reading monitors are available for measuring airborne aerosol concentrations.

The most popular direct reading aerosol monitors are light-scattering devices, which operate by illuminating the sample as it passes through a chamber and measuring the scattered light at a certain angle. This type of monitor is sensitive to the size, shape, and refractive index of the particles. It is used for determining aerosol size and count.

Another type of aerosol monitoring device relies on the behavior of a piezoelectric crystal, whose oscillation frequency changes with the mass of aerosol deposited on it. It is called piezoelectric mass sensor. After each sampling period, the concentration of the aerosol is displayed and the crystal is automatically cleaned and ready for the next measurement. Sampling efficiency is affected by both the mass and the size of the particles. Very low sensitivity is observed when the particle size is larger than 10  $\mu\text{m}$  in diameter or larger masses of particles are collected.

## Gases/Vapors

A large number of gases and organic vapors occur in various workplaces, and many different techniques are required to assess them. Gas and vapor sample collection for monitoring is relatively simple compared to dust sampling. There is no size sampling requirement – monitoring simply requires that one collects and measures the concentration in the atmosphere. Exposure is based on the total amount available for inhalation.

### Sample Collection

Personal air sampling is preferred over area sampling for evaluating worker exposure to contaminants in workplace air. The worker wears a sampling device that is placed very close to the breathing zone to collect an air sample. Area air samples are normally used to evaluate background concentrations and locate sources of exposures. The sampling device is placed in a fixed location in the area of interest.

**Active sampling/sorbent tubes** Probably the most common and versatile type of sample collection device is the solid sorbent tube. These small cylindrical tubes, 3–8 mm internal diameter, containing tens to thousands of milligrams of one or a combination of sorbents, have the workplace air drawn through them to facilitate trapping the contaminant(s) of interest. Solid sorbents commonly used are various forms of activated charcoal or carbon for stable compounds; silica gel for collecting polar compounds such as alcohols, amines, and phenols; synthetic polymers such as various chromsorb, ambersorb, XAD, and Tenax for a wide range of low level of higher molecular weight organics compounds.

Active sampling requires a sampling pump to draw air through the sorbent tubes. The sampling pump needs to be calibrated (with the entire sampling train assembled as it will be used in the field) to within  $\pm 5\%$  of the recommended flow rate according to OSHA analytical methods.

**Passive monitors** Passive monitors provide personal sampling without the use of sampling pumps. They rely on the movement of contaminant molecules across a concentration gradient, which for steady-state conditions can be defined by Fick's first law of diffusion. In simple terms, contaminants move from an area of high to low concentration at defined rates.

Common types of passive monitor involve the trapping of the air contaminant of interest using a solid sorbent or a reactive material contained behind

**Table 1** Passive monitors for occupational air sampling

<i>Sampler model</i>	<i>Manufacture</i>	<i>Target contaminants</i>	<i>Sampling media</i>	<i>Analytical method</i>
3M 3551	3M	Ethylene oxide	Hydrogen bromide treated charcoal disk	GC-ECD
3M 3721	3M	Formaldehyde	Bisulfite-impregnated paper	Visible spectrometry
GMD 570	GMD	Formaldehyde	Glass fiber filter impregnated with 2,4-DNPH	HPLC-UV
526-200/201	SKC	Formaldehyde	Bisulfite-impregnated paper	Visible spectrometry
AT571	Assay technologies	Formaldehyde	Glass fiber filter impregnated with 2,4-DNPH	HPLC-UV
575-Series	SKC	VOCs	Charcoal/Anasorb 747/727	GC-FID
SKC 520	SKC	Inorganic mercury	Anasorb C300	AA
NITROX	Landauer	Nitrous oxide	Molecular sieve	IR
3M3500	3M	VOC	Charcoal	GC-FID, GC-MS

a diffusive barrier. After use, the sorbent is removed and desorbed with a solvent and then analyzed. Another type of passive monitors involves the diffusion of the gas or vapor along an open path into a solution or onto a treated sorbent with which the contaminant reacts. **Table 1** lists manufacturer and model details for both types of passive monitors.

Passive monitors are relatively inexpensive and easy to use, and most commercially available monitors meet or exceed NIOSH accuracy requirements.

**Grab samplers** Grab samples are collected to measure gas and vapor concentrations at certain time and place, which can be used to evaluate 'peak' or ceiling exposures. Grab samples are collected mostly in the following situations: onsite analysis for certain field applications; emergency situations, like spills, leaks, when instantaneous sampling and analysis is critical; or when other sampling media/methods are not available.

Flexible sampling bags made of different materials or evacuated rigid containers, such as SUMMA canisters, are used for grab sample collection.

### Laboratory Analysis

The analysis of material trapped on solid sorbents is done in a number of ways. Arguably the most common analytical technique is desorption in a suitable solvent with gas chromatographic analysis of the resulting solution. There has also been a considerable body of work done on thermal desorption of the trapped material, with or without preconcentration of the desorbed material prior to gas chromatographic analysis. Thermal desorption introduces the potential for pyrolysis of the material of interest, and thus its use is somewhat limited, in comparison to solvent desorption. Mass-selective (MS) detectors and the flame-ionization detectors (FIDs) are the most commonly used detectors for this purpose. Both detectors are considered to be universal detectors for organic compounds; the FID is used mostly for

quantitation while MS can be used for both qualitative identification and quantitative analysis. Other detectors such electron-capture, thermal conductivity, Hall electrolytic conductivity, nitrogen-phosphorus, and alkali flame ionization detectors have all been used for the analysis of workplace air contaminants.

While the gas chromatographic technique is used for most of the analysis of volatile and semivolatile contaminants in workplace air for compounds with relative low volatility, or compounds that are derivatized for better stability or increased sensitivity, HPLC is the analytical method of choice. UV – visible and fluorescence detectors are the two common types of detectors that are used.

Other analytical methods used for the analysis of solid sorbent samples include ion chromatography, ion-selective electrodes, and polarography measurements.

**Table 2** lists the common analytical methods discussed with their respective detection limits and dynamic range.

### Direct Reading Analysis

Direct reading instruments are one of the most important tools that are available to occupational hygienists to detect and quantify workplace air contaminants. These instruments allow real-time or near real-time measurements in the field, thus eliminating the lag-time if the samples have to be sent for laboratory analysis.

**Electrochemical sensors** Electrochemical sensors are widely used for the detection of toxic gases at the parts per million (ppm) level and for oxygen in levels of percent of volume (% vol). Toxic gas sensors are available for a wide range of gases, including carbon monoxide, hydrogen sulfide, sulfur dioxide, nitrogen dioxide, chlorine, and many others.

Although the sensors are designed to be specific to each gas, there are often some cross-interferences with

**Table 2** List of analytical methods for gas/vapor analysis

<i>Analytical method</i>	<i>Typical analytes</i>	<i>Typical detection limit</i>	<i>Linear dynamic range</i>
GC-FID	Hydrocarbons	10–100 pg	10 <sup>7</sup>
GC-ECD	Organohalogens, chlorinated solvents, and pesticides	0.05–1 pg	10 <sup>4</sup>
GC-FPD	Sulfur and phosphorus compounds	10–100 pg	10 <sup>3</sup> –10 <sup>4</sup>
GC-NPD	Organonitrogen and organophosphorus compounds	0.1–10 pg	10 <sup>3</sup>
GC-PID	Compounds ionized by UV	2 pg C per s	10 <sup>7</sup>
GC-MS	Tunable for any compound	10 pg–10 ng (depending on SIM or scan)	10 <sup>5</sup>
GC-AED	Tunable for any compound	0.1–20 pg s <sup>-1</sup>	10 <sup>3</sup> –10 <sup>4</sup>
HPLC-UV	Conjugated organic compounds	0.1–1.0 ng	10 <sup>4</sup> –10 <sup>5</sup>
HPLC-fluorescence	Highly conjugated compounds like PAHs, isocyanate derivatives, etc.	1–10 pg	10 <sup>3</sup> –10 <sup>4</sup>
Ion chromatograph	Anions	0.1–1.0 ng	10 <sup>3</sup> –10 <sup>4</sup>

other gases present. Overall, electrochemical sensors offer very good performance for the routine monitoring of toxic gases and percent of volume oxygen present in both portable and fixed gas monitors.

**Combustible gas instruments (CGIs)** Three types of CGIs are available. The first is the catalytic sensor, which is commonly used to detect and quantify combustible gases and vapors from 0% to 100% LEL (lower explosive limit). The sensor consists of two elements: detector and reference. Both elements consist of metal coils operating in a Wheatstone bridge circuit. The burning gas increases the detector's temperature, resulting in an increase in resistance in the element. The sensor's response to a combustible gas depends on the chemical composition, the molecular weight, and vapor pressure of the gas.

The catalytic sensor is less sensitive to temperature and humidity effects, offers repeatable performance, and is relatively stable. However, it is susceptible to poisoning or inhibition from some gases, which may decrease its sensitivity or damage the sensor beyond recovery.

The second type is the thermal conductivity sensor, which has been used in instruments for measuring combustible gases above the % LEL range and for leak detection for many years. Like the catalytic sensor, this one also consists of two wire elements of detector and reference. The response depends on the thermal conductivity of the gas being detected. Some advantages of the thermal conductivity sensor are that it does not require oxygen to operate and is not susceptible to poisons. One drawback is that it cannot measure gases with thermal conductivities similar to the reference gas (nitrogen). Thermal conductivity sensors are used primarily in portable gas leak detectors.

The third type is the metal oxide sensor, which is described in more detail later in this article.

**Photoionization/flame-ionization detectors (PIDs/FIDs)** PIDs and FIDs detectors are often used in situations where high sensitivity (sub-ppm levels) and limited selectivity (broad-range coverage) are desired. The normal working range is between 0.1 and 100 000 ppm for FID and 0.2–2000 ppm for PID. PID/FIDs are commonly used for detecting volatile organic compounds (VOCs) such as benzene/toluene/xylene, vinyl chloride, and hexane, and provide quick response for this growing concern.

A PID operates by ionizing components of a sample stream with high-energy ultraviolet UV light (while FID operates by ionizing components with high-temperature hydrogen/air flame) and detecting the resulting charged particles collected at an electrode within the detector. Advantages of this technology include the fast response time and excellent shelf-life. PIDs suffer from sensor drift and humidity effects; therefore, calibration requirements are more demanding than other common gas detectors.

**IR gas analyzer (nondispersive IR (NDIR) absorption)** The NDIR sensor, commonly referred to as the IR sensor, is based on the principle that many gases absorb light energy at a specific wavelength, typically in the IR range.

The limitation of NDIR technology for gas detection depends on the uniqueness of the absorption spectrum of a particular gas. IR sensors can detect gases in inert atmospheres (little or no oxygen present), are not susceptible to poisons, and can be made very specific to a particular target gas.

NDIR sensors are also extremely stable, quick to respond to gas, can tolerate long calibration intervals, and have a wide working range (sub-ppm to

low percentage level). IR sensors are commonly used to detect methane, carbon dioxide, and nitric oxides in both portable and fixed gas detection instrumentation.

**Photoacoustic sensors (PAS)** Quantification of air contaminants by PAS involves the use of UV or IR radiation to quantify air contaminants. It involves the absorption of a pulse of light energy by a molecule and the subsequent detection of a pressure wave generated by heat energy released by the molecule upon its return to the ground state. Photoacoustic spectroscopy analysis is nondestructive, can be done in real time, and can be a few orders of magnitude more sensitive than conventional UV-vis spectroscopy. Detection limits are chemical-specific and are reported to be between 0.001 and 1 ppm.

**Fiber optic sensors** The use of fiber optic technology is relatively new and fiber optic sensors are relatively inexpensive to manufacture.

One proven method of optical sensing uses a fiber coated with chemical reagents. A beam of light is guided over the fiber probe as the analyte gas is introduced. The sensor detects the luminescence changes resulting from the reaction of the chemicals with the gas. This type of optical sensor can be formulated for a variety of gases such as oxygen, ammonia, chlorine, and hydrogen, where the chemical principles are well known and suitable laser and light emitting diodes sources are available.

**Solid-state sensors** Solid-state sensors, also referred to as metal oxide sensors, are best used as general survey instruments because of their lack of selectivity/specificity. This type of sensor has a rather narrow working range of 1–50 ppm. A variety of MOS sensors are available for the detection of combustible gases, chlorinated solvents, and some toxic gases, such as carbon monoxide and hydrogen sulfide.

During operation, the sensing element is heated to 250–350°C. When gas enters the sensor, it reacts with the oxide coating, causing a decrease in resistance between the two electrodes, which is then measured.

**Portable GCs** GC and GC-MS remain dominant techniques for the determination of multiple volatile and semivolatile organic compounds in workplace air analysis. Portable gas chromatographs were introduced in the late 1970s and the technology has continued to evolve ever since. Most field GCs consist of an injection system with either syringe



**Figure 2** Photovac Voyager portable GC.

mode or loop mode, capillary or packed columns, a programmable oven, and one or more detectors. Most common detectors used in portable GCs are FID, PID, or ECD (electron capture detector). A field GC can be as accurate as  $\pm 5\%$  of the reading and it also offers a wide dynamic range, from very low-ppb to hundreds of ppm concentration levels for FID/PID.

The Photovac Voyager, a fifth-generation portable GC, is shown in **Figure 2**. Features of the Photovac Voyager include a three-column system and dual PID-ECD detectors. The Voyager also includes a factory-programmed assay that provides preset instrument operating parameters and compound library to simplify operation.

Sixth-generation portable GCs, which are high-speed portable gas chromatographs, have been introduced recently. They can analyze single or multiple compounds within as little as a few seconds up to a couple of minutes.

**Non-instrumental methods/detector tubes** Chemical indicator tubes, using noninstrumental technology, are widely used for the analysis of workplace air in a real-time basis. Sealed glass tubes are filled with a reagent specifically sensitive to a

target gas/vapor. If the target gas is present in an air sample drawn through the tube, a color change will occur in the reagent layer of the tube that depends on the contaminant's concentration. The sample is pulled via either a manual piston pump or a battery-operated motorized pump. There is a wide range of different tubes, specific to different compounds or groups of compounds, and having various effective concentration ranges. The chemistries relied upon to effect the color changes in the tubes are specific to the individual compounds and the reactive reagent.

For example, the Draeger tube for styrene is based on the reaction of styrene with formaldehyde in the presence of sulfuric acid to form a red-brown color stain in the tube.

A few of these devices rely on air contact with the tube through diffusion, and thus operate passively. Tubes used with battery-operated pump or those that sample the air through diffusion are used to determine time-weighted average concentrations of the workplace air contaminant of interest, while all the tubes used with a hand-operated vacuum pump are intended for measuring more or less instantaneous concentrations. The Safety Equipment Institute (USA) currently certifies chemical indicator tubes based on American National Standards Institute/International Safety Equipment Association standard 102 (1996).

## Quality Assurance Aspects of Workplace Air Analysis

Quality assurance is a set of operating principles that, if strictly followed during sample collection and analysis, will produce data of known and defensible quality. The reliability of analytical results depends on many factors such as the personnel performing the tests, environmental conditions, test methods, validation of test methods, equipment functioning, measurement traceability, sampling, and the storage and handling of samples. To minimize all the negative impacts of these factors, a properly validated, documented standard operating procedure must be used for all workplace air analysis. This standard operating procedure should include, but are not limited to, the following elements: acceptable sampling/

analysis devices and procedures, proper transportation and storage conditions, calibration requirements for both field and laboratory instrumentation, estimation of measurement uncertainties, and acceptable ranges for spike recovery efficiencies.

There are other routinely practiced procedures in terms of intralaboratory and interlaboratory quality control, such as the control chart, round-robin interlaboratory testing, and various proficiency tests offered by different agencies, such as the PATs (Proficiency Analytical Testing) by AIHA in the US and WASPs (Workplace Analysis Scheme for Proficiency) by the Health and Safety Laboratory in Great Britain. In normal circumstances, a combination of these quality control measures should be used in performing workplace air analysis.

*See also:* **Air Analysis:** Outdoor Air. **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Principles and Instrumentation. **Environmental Analysis. Gas Chromatography:** Overview; Principles; Instrumentation. **Liquid Chromatography:** Overview; Principles; Instrumentation. **Personal Monitoring:** Active; Passive. **Quality Assurance:** Quality Control; Instrument Calibration. **Spectrophotometry:** Overview; Inorganic Compounds; Organic Compounds.

## Further Reading

- AIHA Sampling and Laboratory Analysis Committee (1997) *Laboratory Quality Assurance Manual*, 2nd edn. Fairfax, VA: AIHA Press.
- Cassinelli ME and O'Connor PF (1994) *NIOSH Manual of Analytical Methods (NMAM®)*, 4th edn., pp. 94–113. USA: DHHS (NIOSH).
- Koester CJ, Simonich SL, and Esser BK (2003) Environmental analysis. *Analytical Chemistry* 75(12): 2813–2829.
- Plog BA and Quinlan PJ (2002) *Fundamentals of Industrial Hygiene*, 5th edn. chs 16 and 17. USA: National Safety Council.
- Salvatore RD (1997) *The Occupational Environment – its Evaluation and Control*, 1st edn., Fairfax, VA: AIHA Press.
- Sipin MF, Guazzotti SA, and Prather KA (2003) Recent advances and some remaining challenges in analytical chemistry of the atmosphere. *Analytical Chemistry* 75(12): 2929–2940.

# ALCOHOL

See **ETHANOL. FORENSIC SCIENCES: Alcohol in Body Fluids**

# ALKALOIDS

**R Verpoorte**, Gorlaeus Laboratories, Leiden,  
The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

## Definition and Classification of Alkaloids

An alkaloid has been defined by Pelletier as: “a cyclic organic compound containing nitrogen in a negative oxidation state which is of limited distribution among living organisms”.

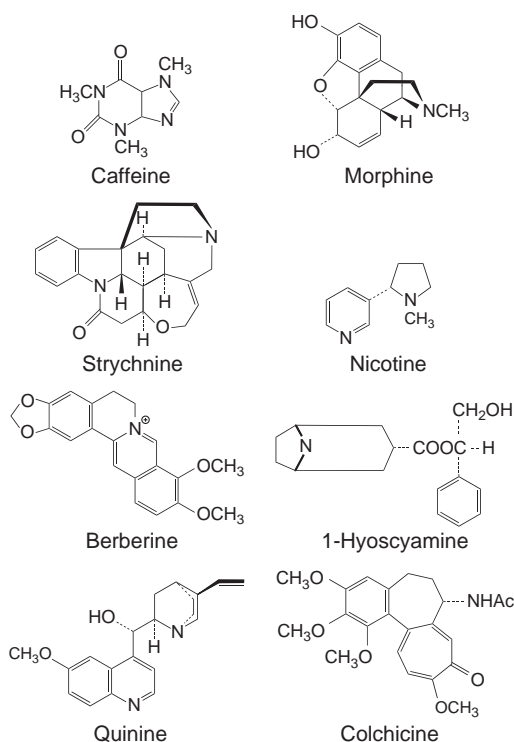
Most alkaloids have basic properties connected with a heterocyclic tertiary nitrogen. Notable exceptions are colchicine, caffeine, and paclitaxel. Most alkaloids are biosynthetically derived from amino acids such as phenylalanine, tyrosine, tryptophan, ornithine, and lysine. Alkaloids represent a wide variety of chemical structures. About 20 000 alkaloids are known, most being isolated from plants. But alkaloids have also been found in microorganisms, marine organisms such as algae, dinoflagellates, and puffer fish, and terrestrial animals such as insects, salamanders, and toads.

Alkaloids are often classified according to their molecular skeleton, e.g., the two largest groups are the indole alkaloids and isoquinoline alkaloids (each more than 4000 compounds). Other important groups are tropane alkaloids (~300 compounds), steroidal alkaloids (~450 compounds), and pyridine and pyrrolizidine alkaloids (respectively, ~250 and 570 compounds).

Classification based on botanical origin of the alkaloids are also used, e.g., *Papaver* (opium) alkaloids, *Cinchona* alkaloids, *Rauvolfia* alkaloids, *Catharanthus* alkaloids, *Strychnos* alkaloids, Ergot alkaloids, cactus alkaloids, and *Solanum* alkaloids. The structures of some alkaloids are shown in **Figure 1**.

Many alkaloids have strong biological activities in man. In part this can be explained by structural relationship with important signal compounds (neurotransmitters) such as dopamine, acetylcholine, noradrenaline, and serotonin. The fact that alkaloids are water soluble under acidic conditions and lipid soluble under neutral and basic conditions give them unique properties for medicinal use, as they can be transported in the protonated form, and can pass membranes in the neutral form. In fact most synthetic medicines do contain one or more tertiary nitrogens.

A number of alkaloids have commercial interest as medicines or tools in pharmacological studies (see **Table 1**). Both pure compounds and plants (or extracts



**Figure 1** Structures of some different types of alkaloids.

**Table 1** Some alkaloids of pharmaceutical interest

Ajmalicine	Physostigmine
Ajmaline	Pilocarpine
Quinine	Veratrine
Quinidine	Solasodine
Strychnine	Harringtonine
Reserpine	Ephedrine
Rescinamine	Mescaline
Yohimbine	Aconitine
Vincamine	Nicotine
Vinblastine	Tetrodotoxin
Vincristine	Saxitoxin
9-Hydroxyellipticine	Sparteine
Camptothecin	Lobeline
Emetine	Muscarine
Atropine (1-hyoscyamine)	Serotonin
Scopolamine	Harmane
Cocaine	Psilocybine
Codeine	Caffeine
Morphine	Theophylline
Thebaine	Theobromine
Papaverine	Taxol
Narceine	Ergotamine
Narcotine	Ergonovine
Berberine	Ergosine
Sanguinarine	Ergocristine
Tubocurarine	Ergocryptine
Boldine	Ergocornine
Colchicine	Lysergic acid



thereof) containing alkaloids are used. Furthermore some alkaloids are widely found as drugs of abuse (e.g., mescaline, cocaine, psilocybin, psilocin, morphine, and its semisynthetic derivative heroin), as doping compounds (e.g., strychnine, ephedrine, caffeine), and as poisons (e.g., strychnine, pyrrolizidine alkaloids, coniine, nicotine, tetrodotoxin). Consequently, methods for the determination of alkaloids can be found in quite different contexts, mainly dependent on the matrices in which the alkaloids are found.

## Chemical Properties and Artifact Formation

Many alkaloids are difficult to crystallize as free base, but do crystallize as a salt. Most alkaloids have basic properties. The  $pK_a$  values vary from 6 to 12, with most alkaloids in the range of 7–9. In general the free base is soluble in organic solvents and not in water. Protonation of the amine function under acidic conditions usually results in a water-soluble compound. This characteristic is widely used in the selective isolation of alkaloids. Quaternary alkaloids are poorly soluble in organic solvents but are soluble in water at any pH.

Most alkaloids are colorless—only a few highly conjugated compounds are colored (e.g., berberine, sanguinarine, serpentine) or show strong fluorescence (e.g., quinine).

Alkaloids are often quite unstable, and N-oxidation in particular is quite common. Besides the effect of heat and light, their stability is influenced by solvents. Halogenated solvents are widely used in alkaloid research, and chloroform in particular is one of the most suitable because of its relatively strong proton donor character. However, these solvents are very reactive in terms of artifact formation. In chloroform (N-)oxidations occur easily. Furthermore peroxides in ethers may cause N-oxidations. With dichloromethane quaternary N-dichlorometho compounds may easily be formed. Similar compounds are formed with minor impurities present in chloroform. But also chloroform itself may react with an alkaloid to give artifacts. The decomposition product formed from chloroform by oxidation, phosgene, reacts with ethanol that is always present in this solvent. This product ethyl chloroformate reacts with tertiary nitrogens to yield ethylcarbamates. Alkaloids are in general more stable in toluene, ethyl acetate, and alcoholic solutions. In case of alkaloids containing a carbinolamine function, reactions with alcohols will occur (e.g., O-methyl pseudostrychnine is formed from pseudostrychnine with methanol). The 1–2% of ethanol in chloroform may also result in ethyl-derivatives. Ketones such as acetone and methyl ethyl ketone are well-known artifact formers, as

they can react both with carbinolamines and amines. Berberine for example may react with acetone. Ammonia and acetone may react during column chromatography on silica, yielding conjugates that give a Dragendorff-positive reaction. Ammonia may also react with aldehydes present in plant materials. Giving rise to artificial alkaloids, e.g., gentianine, which is formed from sweroside during extraction.

## Extraction

Alkaloids can be extracted under neutral or basic conditions (after basification of the plant material or biofluid to pH 7–9 with ammonia, sodium carbonate, or sodium hydrogencarbonate) as free base with organic solvents (e.g., dichloromethane, chloroform, ethers, ethyl acetate, alcohols). Some alkaloids can only be extracted at higher pH (>10) e.g., tryptamine. On the other hand alkaloids containing phenolic groups (e.g., morphine) are deprotonated at higher pH, and are thus not extracted by organic solvents under such conditions. Alkaloids can also be extracted in the protonated form (after acidification to pH 2–4 with dilute acids such as phosphoric acid, sulfuric acid, or citric acid) with water or alcohols (e.g., methanol). Also acetic acid and trifluoro acetic acid can be used for acidification. However, because of their lipophilic properties they will lead to ion-pair formation with the alkaloids, which in case of liquid–liquid extraction of the acidic aqueous extract with an organic solvent will pass into the organic phase (see below). A 0.1% trifluoroacetic acid solution in water as such is a very efficient solvent for alkaloid extraction from plant material which can directly be used for high-performance liquid chromatography (HPLC) analysis.

Further purification of alkaloids can be done by liquid–liquid extraction or liquid–solid extraction.

In liquid–liquid extraction the alkaloids are, after basification, extracted from an aqueous solution with an immiscible organic solvent (e.g., dichloromethane, diethyl ether, ethyl acetate, chloroform), or from an organic solvent with an aqueous acid. With the aid of ion-pairing agents (e.g., alkylsulfonic acids), alkaloids can be extracted from an acidic aqueous solution with organic solvents. It should be noted that common ions such as  $Cl^-$ ,  $Br^-$ ,  $I^-$ , acetate, and trifluoroacetate also result in the formation of ion-pairs which are readily soluble in organic solvents.

For liquid–solid extractions there are several possibilities. Reversed-phase materials such as chemically bonded  $C_8$  and  $C_{18}$  on silica are widely used. Also ion-exchange materials are used for the selective extraction of alkaloids.

For preparative purposes purifications based on precipitation of alkaloids are employed. A crude



extract of the alkaloids is made with aqueous acid. Subsequently the alkaloids are precipitated with reagents such as Mayer's reagent (1 mol l<sup>-1</sup> mercury(II) chloride in 5% aqueous potassium iodide) or Reinecke's salt (5% aqueous ammoniumtetrathio-cyanatodiammonochromate in 30% acetic acid) at pH 2, or picric acid (saturated aqueous solution) at pH 5–6. After collection by filtration or centrifugation, the precipitate is dissolved in an organic solvent (e.g., acetone–methanol–water, 6:2:1 (v/v/v)). The complexing substance is then removed by means of an anion exchanger. This method is particularly suited for the purification of quaternary alkaloids.

## Detection Reactions

Numerous reactions have been described for the detection of alkaloids in aqueous solutions. The most common are precipitation reactions such as Dragendorff's reagent (bismuth iodide), Mayer's reagent, picric acid, and Reinecke's salt (see above). All these reagents are more or less specific for protonated tertiary nitrogens (and quaternary nitrogens). Dragendorff's reagent is also used for the detection of alkaloids on thin-layer chromatography (TLC) plates. It may cause false-positive reactions with, for example, compounds containing conjugated carbonyl or lactone functions.

## Chromatographic Analysis

### Thin-Layer Chromatography

TLC is widely used in the analysis of alkaloid extracts. Particularly in the analysis of plant materials it offers the advantage of molecules with a broad range of polarities being separated in a single analysis. Also it does not require extensive sample cleanup of crude

extracts as in case of gas chromatography (GC) and HPLC. The most widely used stationary phase is silica. Alumina plates have been applied in the past, but are nowadays rarely employed. Reversed-phase materials such as chemically bonded C<sub>18</sub> on silica are also applied, but because of problems of tailing, the much cheaper silica plates are still widely employed.

In the case of strongly basic alkaloids, severe tailing may occur on silica gel plates because of the acidic properties of this adsorbent. Therefore, mobile phases containing bases like ammonia or diethylamine are widely used. Alternatively, TLC plates impregnated with basic solutions have been employed. For the detection of highly polar quaternary alkaloids and N-oxides, solvent systems consisting of methanol and aqueous salt solutions are useful. In Table 2 some widely used TLC systems are summarized.

Many methods have been reported for the detection of alkaloids on TLC plates. Besides quenching of ultraviolet (UV) radiation on fluorescent plates, general reagents for selectively detecting alkaloids are Dragendorff's reagent (orange-brown spots) and potassium iodoplatinate reagent (violet-purple spots) (Table 3). There is less risk of false-positive reactions with the latter (see above). Highly selective reagents have been reported for the various classes of alkaloids (see Table 4). These are based on different colorations under strongly oxidizing conditions.

### Gas Chromatography

Although alkaloids are nonvolatile, many alkaloid mixtures have been analyzed successfully by GC, especially capillary GC. Generally stationary phases with a low or sometimes intermediate polarity are used. Though hardly used anymore, for packed columns basic deactivation of the supporting material will improve peak shapes. For capillary columns

**Table 2** Some common for alkaloid analysis

<i>Solvent system (all with silica plates)</i>	<i>Commonly used ratios (v/v)</i>	<i>Polarity range<sup>a</sup></i>
Cyclohexane–chloroform–diethylamine	5:4:1–(0):9:1	lp–mp
Chloroform–acetone–diethylamine	5:4:1	mp
Chloroform–methanol–ammonia	8:1:1	mp
Chloroform–methanol/ethanol	99:1–1:1	lp–mp, wb
Ethyl acetate–isopropanol–25% aq. ammonia	100:2:1, 80:15:5, 45:35:5	lp–mp
Ethyl acetate–methanol	9:1–1:1	lp–mp, wb
Toluene–ethyl acetate–diethylamine	7:2:1	lp–mp
Toluene–acetone–ethanol–25% aq. ammonia	20:20:3:1	mp
Dichloromethane–diethyl ether–diethylamine	20:15:5	mp
Acetone–methanol–25% aq. ammonia	40:10:2, 95:(0):5	mp–p
Methanol–25% aq. ammonia	95:5	lp–p
n-Butanol–acetic acid–water	4:1:1	lp–p
Methanol–1 mol l <sup>-1</sup> aq. NH <sub>4</sub> NO <sub>3</sub> –2 mol l <sup>-1</sup> aq. ammonia	7:1:2	lp–p
Methanol–0.2 mol l <sup>-1</sup> aq. NH <sub>4</sub> NO <sub>3</sub>	3:2	lp–p

<sup>a</sup>lp, low polar compounds; mp, medium polar compounds; p, polar compounds; wb, weakly basic compounds.

**Table 3** Detection reagents for alkaloids on TLC plates

*Dragendorff's reagent* (modification according to Munier)  
 (A) 1.7% bismuth subnitrate in 20% aq. tartaric acid solution.  
 (B) 40% Potassium iodide in water.  
 (A) and (B) are mixed (5:2) and the spray reagent is prepared by mixing 50 ml of the stock solution with 100 g tartaric acid and 500 ml water.  
 Colors observed after spraying: orange–brown spots for alkaloids.

*Potassium iodoplatinate reagent*

The reagent prepared freshly by mixing 3 ml of 10% aq. hexachloroplatinic acid solution with 97 ml of water and 100 ml of 6% aqueous potassium iodide solution.

Colors observed after spraying: violet–purple spots for alkaloids.

**Table 4** Some selective color reactions for the detection of alkaloids on TLC plates

Treatment	Alkaloids detected
0.2 mol l <sup>-1</sup> iron(III) chloride in 35% perchloric acid: heat	Indole alkaloids, isoquinoline alkaloids
1% Cerium(IV) sulfate in 10% sulfuric acid	Indole alkaloids
1% p-dimethylaminobenzaldehyde in ethanol, followed by exposure to hydrochloric acid vapor	Ergot alkaloids
Sulfuric acid: heat	Various alkaloids

usually a length of 10–25 m is chosen. High temperatures for injection (200–300°C) and column (temperature gradients from 100°C to 250°C) are generally required. By using nitrogen specific detectors high selectivity can be obtained. The combination of GC with mass spectrometry (MS) has been fruitfully used for alkaloid analysis.

### Liquid Chromatography

For alkaloid analysis, liquid chromatography (LC) has developed into an important tool. Most separations are done on reversed-phase materials (C<sub>8</sub>-, C<sub>18</sub>-, and phenyl-bonded phases on silica), a major problem being extensive tailing arising from the interaction of the basic nitrogen and residual acidic silanol groups in the reversed-phase materials. Strong bases in particular show this problem. By adding long-chain alkylamines (e.g., hexylamine) in low concentrations (typically 1 mmol l<sup>-1</sup>) to the mobile phase, tailing can be reduced considerably. Also addition of amines like triethylamine or tetramethylammonium can be helpful in reducing tailing. Presently many special treated columns are available that show improved peak shapes for basic compounds. For these special treated columns selectivity and efficiency can vary considerably between brands,

**Table 5** General outlines of reversed-phase LC systems for the separation of alkaloids

Stationary phase	Mobile phase
C <sub>8</sub> -, C <sub>18</sub> - or phenyl-bonded phase with low percentage of free silanol groups.	<p><i>Ion-suppression mode</i>: methanol (acetonitrile)–water containing ~0.01–0.1 mol l<sup>-1</sup> phosphate buffer, ammonium carbonate or sodium acetate (pH 4–7).</p> <p><i>Ion-pair mode</i>: methanol (acetonitrile)–water containing ~0.005 mol l<sup>-1</sup> alkylsulfonate and 1% acid (e.g. acetic acid), pH 2–4.</p>

**Table 6** Mobile phases used for isocratic LC separation of alkaloids, with silica gel as the stationary phase

Dichloromethane Chloroform Diethyl or isopropyl ether Tetrahydrofuran Ethyl acetate	–	<table> <tr> <td>Methanol</td><td>–</td><td>Ammonia</td></tr> <tr> <td>Isopropanol</td><td>–</td><td>Diethylamine</td></tr> <tr> <td></td><td></td><td>Triethylamine</td></tr> <tr> <td></td><td></td><td>(~1% of the mobile phase)</td></tr> </table>	Methanol	–	Ammonia	Isopropanol	–	Diethylamine			Triethylamine			(~1% of the mobile phase)
Methanol	–	Ammonia												
Isopropanol	–	Diethylamine												
		Triethylamine												
		(~1% of the mobile phase)												

and plate numbers are usually lower than obtained for neutral test compounds on the same column. Less common is the use of polymeric-, ion-exchange, and aminopropyl-type columns.

Because of the equilibrium between free base and protonated alkaloid, the pH of the mobile phase has to be controlled. As silica-based stationary phases are unstable at higher pH (above 8), usually a pH between 2 and 4 is used. Both ion-suppression and ion-pairing modes are employed. Some general features of reversed-phase LC systems are given in **Table 5**. Isocratic separation using silica has been reported (**Table 6**), but despite the large selectivity offered, they are not commonly used anymore. The systems usually are similar to those applied in TLC.

For detection, UV absorption is most widely used. Some alkaloids can also be detected by means of their fluorescence. Some type of alkaloids have poor UV-absorptive properties, e.g., tropane alkaloids, pyrrolizidine alkaloids, and steroidal alkaloids, necessitating detection at a lower wavelength (200–220 nm). Electrochemical detection has been applied for alkaloids, enabling selective detection in the presence of interfering compounds.

MS coupled with LC is the most selective detection for alkaloids. As all-volatile eluents are needed for LC–MS, the acidic eluents often contain formic acid, acetic acid, or trifluoroacetic acid.

### Capillary Electrophoresis

Because of their ionic nature alkaloids can be analyzed by capillary zone electrophoresis. Low

pH-buffers can be used for the analysis of alkaloids (e.g., a pH 2.5 phosphate buffer). For closely related compounds capillary electrophoresis (CE) selectivity might be a problem as the charge-to-mass ratios are very similar. Capillary electrochromatography can also be used, but residual silanol groups in the stationary phase may cause tailing, similarly as in HPLC.

### Countercurrent Chromatography

The preparative isolation of alkaloids can be done with modern countercurrent chromatography. Because of the ionic nature of alkaloids, ion-pair gradients can be used to improve separation. Possible two-phase solvent systems are chloroform-methanol-aqueous phosphate or citrate buffer, pH ~4, containing perchlorate, acetate, or chloride as ion-pairing agent. Also pH-zone refining is a powerful tool for the large-scale separation of alkaloids.

## Spectrometric Techniques

### Nuclear Magnetic Resonance Spectrometry

For identification and structure elucidation,  $^1\text{H}$  nuclear magnetic resonance (NMR) spectrometry is a very useful tool in alkaloid research. With modern NMR high-field NMR spectrometers (300–800 MHz)  $^1\text{H}$  NMR-spectra can be obtained from 0.01 to 1 mg quantities of alkaloid. For identification such spectra are usually conclusive in most cases. For more complex two-dimensional spectra, e.g., in the case of complete assignment of the spectra of new compounds, larger amounts of sample are necessary. The usual solvent is deuterated chloroform. Small amounts of acid (e.g., as impurity caused by decomposition of chloroform) cause large shifts (up to 1–2 ppm) for protons close to the nitrogen. The shift is dependent on the degree of protonation of the nitrogen. This phenomenon can be used to improve spectral resolution and to identify signals close to the protonated nitrogen, i.e., not only adjacent protons are affected, but also those which are spatially close. Trifluoroacetic acid can thus be used as shift reagent in the analysis of  $^1\text{H}$  NMR-spectra of alkaloids.  $^1\text{H}$  NMR can also be used for the quantitation of alkaloids in crude extracts (e.g., caffeine, ephedrine, and pyrrolizidine alkaloids) or pharmaceutical formulations.

$^{13}\text{C}$  NMR spectrometry is likewise a major tool in structure elucidation of alkaloids. However, larger amounts of material are needed (1–10 mg). Shifts caused by acidic impurities may also occur, but are relatively less dramatic than in  $^1\text{H}$  NMR spectrometry.  $^{15}\text{N}$  NMR spectrometry has been applied to

alkaloids, but due to problems of sensitivity, its application is limited.

### Mass Spectrometry

The fragmentation of alkaloids has been studied extensively. MS is therefore a major tool in the identification and structure elucidation of alkaloids. In combination with GC and LC it is very useful in the quantitative analysis in complex mixtures of alkaloids, e.g., in plant or other biological materials. Solvent systems suited for LC-MS should only contain volatile compounds, i.e., the salts used in reversed-phase eluents should also be volatile (e.g., ammonium acetate, ammonium carbonate, ammonium formate, trifluoroacetic acid).

**See also:** **Extraction:** Solvent Extraction Principles. **Gas Chromatography:** Mass Spectrometry; High-Temperature Techniques. **Liquid Chromatography:** Normal Phase; Reversed Phase; Liquid Chromatography–Mass Spectrometry. **Nuclear Magnetic Resonance Spectroscopy – Applicable Elements:** Nitrogen-15.

### Further Reading

- Baerheim SA and Verpoorte R (1983) *Chromatography of Alkaloids. Part A: Thin-Layer Chromatography*, *Journal of Chromatography Library*, vol. 23A. Amsterdam: Elsevier.
- Brossi A (ed.) (1983–1992) *The Alkaloids*, vols. 21–40. New York: Academic Press.
- Cordell GA (1981) *Introduction to Alkaloids. A Biogenetic Approach*. New York: Wiley.
- Cordell GA (ed.) (1992) *The Alkaloids*, vol. 40. New York: Academic Press.
- Hesse M (1974) *Progress in Mass Spectrometry*, vol. 1, *Parts 1 and 2, Mass Spectrometry of Indole Alkaloids*. Weinheim: Verlag Chemie.
- Hesse M and Bernhard HO (1975) *Progress in Mass Spectrometry*, vol. 3, *Mass Spectrometry of Alkaloids*. Weinheim: Verlag Chemie.
- Pelletier SW (ed.) (1983) *Alkaloids: Chemical and Biological Perspectives*, vols. 1–6. New York: Wiley.
- Popl M, Fährnich J, and Tatar V (1990) *Chromatographic Analysis of Alkaloids*. New York: Dekker.
- Sangster AW and Stuart KL (1965) Ultra-violet spectra of alkaloids. *Chemical Reviews* 65: 69–130.
- Southon IW and Buckingham J (1989) *Dictionary of Alkaloids*. London: Chapman and Hall.
- Stoeckigt J, Sheludko Y, Unger M, *et al.* (2002) High-performance liquid chromatography, capillary electrophoretic and capillary electrophoretic-electrospray ionization mass spectrometric analysis of selected alkaloid groups. *Journal of Chromatography A* 967: 85–113.
- Verpoorte R and Baerheim Svendsen A (1984) *Chromatography of Alkaloids. Part B: Gas-Liquid Chromatography and High-Performance Liquid Chromatography*.

*Journal of Chromatography Library*, vol. 23B. Amsterdam: Elsevier.

Verpoorte R (1986) Methods for the structure elucidation of alkaloids. *Journal of Natural Products* 49: 1–25.

## ALKYLLEAD SPECIES

See LEAD

## AMINO ACIDS

**S M Lunte**, University of Kansas, Kansas, KS, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 81–90, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

### Introduction

Amino acid analysis is one of the most widely employed analytical procedures. It is utilized in the elucidation of the structure of proteins, the diagnosis of disease states, and the determination of the nutritional value of foods, to name just a few examples. Sample types can be divided into two broad categories, protein hydrolysates and biological samples.

For the determination of the amino acid composition of a protein, a highly purified sample is hydrolyzed under acidic conditions to its constituent amino acids. The amino acid composition of the protein is determined from the concentrations of the individual amino acids and from the molecular weight of the protein. Because the protein sample is normally very pure, the composition of the hydrolysate is limited to amino acids commonly found in proteins. One does not have to be concerned about other endogenous amines that may interfere with the analysis. Most commercial amino acid analyzers quantify only the 20 amino acids commonly found in proteins. Therefore, the biggest challenge for this application is to produce a method with adequate sensitivity to accurately determine the amino acid composition of very small amounts ( $<1\text{ }\mu\text{g}$ ) of protein isolated from gels or nitrocellulose membranes.

The determination of amino acids in physiological samples presents a different set of challenges. Blood, urine, cerebrospinal fluid, and tissue samples are the most common types in this category. In these cases the amino acid composition of the mixture is usually not well defined. The total number of amino acids to

be determined may be smaller or larger than for protein hydrolysates, depending on the sample type and the goals of the analysis. Many nonprotein amino acids, such as taurine and  $\gamma$ -aminobutyrate (GABA), are included in the assay of physiological amino acids. In contrast to protein hydrolysates, which are fairly 'clean' samples, most biological fluids and tissues contain endogenous amines that may interfere with the determination of the amino acids.

### Qualitative Spot Tests

Amino acids lack a distinguishing chromophore; therefore, detection is almost always based on the reaction of the amine with a derivatizing reagent. For a simple spot test, it is necessary to choose a reagent that reacts quickly to give a product possessing different spectroscopic properties from the parent reagent. Examples of this type of reagent include ninhydrin, 1-fluoro-2,4-dinitrobenzene, fluorescamine, and *o*-phthalaldehyde (OPA). The resulting absorbance or fluorescence reading will provide an indication of the total amount of amino acid present in the sample.

### Colorimetry

The simplest method for the determination of amino acids is reaction with ninhydrin. Ninhydrin reacts with both primary and secondary amino acids to produce Ruhemann's purple, which can be detected by ultraviolet (UV)–visible spectroscopy. The reaction requires heat, and a reducing agent is generally added to stabilize the color formation. Primary amines are detected with the greatest sensitivity at 570 nm, while the absorption maximum for secondary amines is 440 nm. If both primary and secondary amines are to be determined, a common absorption wavelength of 500 nm is employed; however, this leads to decreased sensitivity. Under optimal

conditions, a minimum of 5–10 nmol of material is required for detection.

Sanger's reagent, 2,4-dinitrofluorobenzene, can also be employed for the colormetric detection of amino acids. The reagent reacts with both primary and secondary amines and derivatives are detected at 420 nm. The molar absorptivity of the reagent is  $10^4$ , allowing quantification of amino acids at the micromolar level. The reagent also reacts with thiols and alcohols, but these products are relatively unstable.

### Fluorimetry

Fluorescamine is employed for more sensitive determination of amino acids by fluorescence spectroscopy. Since it undergoes rapid hydrolysis under aqueous conditions, the reagent is normally dissolved in acetone. It is commonly employed for the detection of amino acids on surfaces, such as thin-layer chromatography (TLC) plates. Amino acids are detected based on emission of light at 470 nm following excitation at 390 nm. Detection limits for amino acids are in the low picomole range.

OPA can also be employed for the detection of mixtures of amino acids. It reacts with primary amines in the presence of mercaptoethanol or other thiols to produce highly fluorescent isoindole derivatives (excitation wavelength 340 nm, emission 455 nm). In contrast to that with ninhydrin, this

reaction proceeds rapidly at room temperature and is complete in less than 1 min. The OPA and the thiol can be added to a basic solution of amino acids to provide a measurement of the total amino acid content based on a fluorescence measurement.

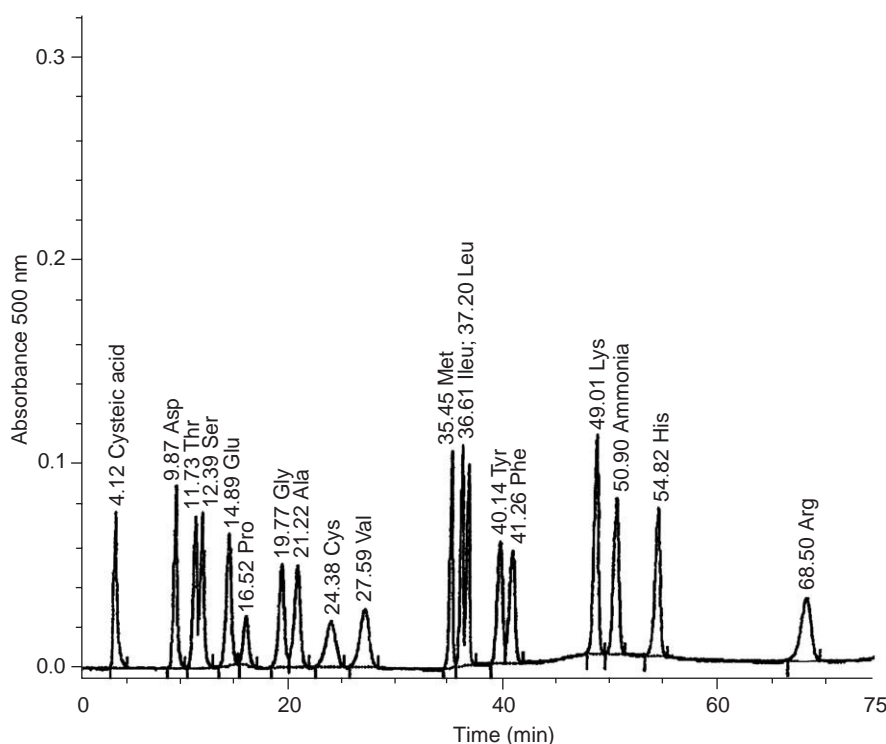
## Chromatographic Methods

### Ion-Exchange Chromatography

The determination of concentrations of individual amino acids in mixtures demands a separation step. The classical method employs cation-exchange chromatography using a strongly acidic sulfonated polystyrene/divinylbenzene copolymer as a support. These columns are very rugged and can withstand concentrated acids, bases, and large amounts of organic solvents. Amino acids are normally eluted with an acidic mobile phase. A typical chromatogram using cation-exchange chromatography is illustrated in **Figure 1**.

### Direct detection

*UV spectroscopy* Amino acids can be detected directly in chromatographic eluents by absorbance spectroscopy at 200 nm with detection limits between 1 and 10 pmol. However, there are several drawbacks to this method of detection for liquid chromatography (LC). The first is a lack of selectivity



**Figure 1** Separation of amino acids by cation-exchange chromatography and postcolumn reaction with ninhydrin. (Courtesy of Biochemical Research Service Laboratory, University of Kansas; used by permission.)

in biological samples, since there are many other endogenous compounds that absorb at this wavelength. Second, many of the solvents employed in LC also absorb at 200 nm, leading to a high background signal if they are present in the mobile phase.

**Pulsed amperometric detection (PAD)** Amino acids are not generally considered to be electrochemically active because products of the oxidation accumulate on the electrode surface and prevent it from participating in any further electrochemical processes. This problem can be overcome if PAD is employed. Amino acids are generally detected using a platinum electrode under alkaline conditions ( $0.25 \text{ mol l}^{-1}$  NaOH) using a triple-pulse waveform with  $E_1$ ,  $E_2$ , and  $E_3$  at 0.50,  $-0.89$ , and  $0.70 \text{ V}$ , respectively. Due to the basic conditions required for the detection of amino acids, a base-stable anion-exchange column must be employed. Detection limits of  $50 \text{ pmol}$  have been obtained for phenylalanine and methionine using this technique.

**Direct amperometry** Direct detection of amino acids can be accomplished with nickel and copper electrodes. The oxidation occurs under only neutral or basic conditions; therefore, the separation is limited to silica- or polymer-based columns. Cation-exchange chromatography, which requires acidic conditions, cannot be used unless base is added postcolumn.

In the case of the nickel electrode, detection is performed at  $+0.49 \text{ V}$  versus the standard calomel electrode. At this potential,  $\text{Ni}^{3+}$  is present on the electrode surface. The  $\text{Ni}^{3+}$  is reduced in the presence of the amino acids to  $\text{Ni}^{2+}$ . The current produced is proportional to the concentration of the amino acid. Detection limits for glycine have been reported to be in the low nanogram range.

Amino acids can also be determined by amperometric detection with a copper electrode under neutral or alkaline conditions. This method is very selective, with the working electrode potential set at only  $+0.10 \text{ mV}$  versus  $\text{Ag}/\text{AgCl}$ . In this detection method, amino acids complex with  $\text{Cu}^{2+}$  ions present on the electrode surface, producing a response that is proportional to the concentration of the amino acid. Detection limits are between  $10$  and  $100 \text{ pmol}$  and can be improved if microbore chromatography is employed. The slower flow rates utilized with microbore chromatography columns allow more time for complexation to take place, thus improving the detection limits.

**Postcolumn derivatization** Although amino acids can be determined directly following cation-exchange

chromatography using the detection methods discussed above, the most prevalent technique for increasing the sensitivity is postcolumn addition of a derivatizing reagent. The resulting derivatives can then be detected by UV or fluorescence detection. Several of the reagents employed for postcolumn derivatization such as ninhydrin, fluorescamine, and OPA are also used to determine total amino acids.

Ninhydrin was the first reagent utilized for the detection of amino acids following cation-exchange chromatography, and it is still very popular today. For liquid chromatographic determinations, dual wavelength and diode array absorbance detectors are often used to detect both primary and secondary amino acids with good sensitivity. However, if sensitivity is not important, both species can be determined at a wavelength of  $500 \text{ nm}$ . Detection limits using the optimum wavelengths are  $\sim 50 \text{ pmol}$ .

Because it is based on UV absorption, the use of ninhydrin is the least sensitive method for the determination of amino acids. It also has the disadvantage that the reaction coil must be heated to  $130^\circ\text{C}$  in order for the postcolumn chemistry to occur in the minimum amount of time. In addition, the reagent reacts with ammonia, which leads to baseline instability if buffers containing trace ammonia are employed to make up the mobile phase.

For increased sensitivity, reagents that yield fluorescent products are employed. These include OPA, fluorescamine, and 4-fluoro-7-nitro-benzo-2-oxa-1,3-diazole (NBD-F). Derivatization with OPA provides the best limits of detection for most amino acids. Although the products of OPA derivatization are unstable and decompose fairly rapidly to produce nonfluorescent 2,3-dihydro-1H-isoindol-1-ones, this is not a problem with postcolumn derivatization since the products are detected almost instantaneously.

One of the drawbacks of OPA is that it reacts only with primary amines. Proline can be determined if sodium hypochlorite or chloramine T is added to the postcolumn reagent. Under these conditions, proline is converted to aminobutyraldehyde, which reacts with OPA. The addition of hypochlorite can be accomplished in two ways. It can be pulsed in at the time of elution of proline. This provides the highest sensitivity for the primary amino acids, but causes a baseline disturbance around the elution time of proline. It also requires computerized equipment for precise control of the timing of the pulse. The second possibility is to add hypochlorite to the mobile phase, but this causes a decrease in response for the primary amino acids due to their conversion to chloramines. Another problem that occurs with OPA is that the derivatives of lysine and cysteine exhibit fluorescence

quenching. However, the fluorescence response can be revived by addition of Brij 35 to the mobile phase. Despite these drawbacks, the sensitivity of this method is better than those of the competing postcolumn methods, with detection limits for amino acids in the 1–2 pmol range.

Fluorescamine has also been employed for postcolumn derivatization with fluorescence detection. As discussed previously, a fluorescent product is generated in the presence of primary amines within a few seconds. Fluorescamine undergoes hydrolysis, but the hydrolysis reaction is much slower than the reaction with primary amines and the product of that reaction does not fluoresce. For postcolumn addition, the reagent is added as a solution in acetone, acetonitrile, or methanol. As with OPA, proline can be detected through conversion to aminobutyraldehyde by oxidation with *N*-chlorosuccinimide. Unlike OPA, fluorescamine reacts poorly with ammonia. Therefore, problems with baseline drift are not as prominent. Since the reaction of fluorescamine with primary amines produces an equilibrated mixture of two products, it is used most frequently as a postcolumn reagent.

4-Chloro-7-nitrobenzo-2-oxa-1,3-diazole (NBD-Cl) and 4-fluoro-7-nitrobenzo-2-oxa-1,3-diazole (NBD-F) can also be employed for postcolumn detection of amino acids. The reagents also react with alcohols and thiols, but are most reactive toward amines. NBD-F is 50–100 times more reactive than NBD-Cl and is therefore the preferred form for postcolumn derivatization. Unlike OPA and fluorescamine, NBD-F can be used for the determination of both primary and secondary amines. The excitation and emission wavelengths for the NBD-amino acids are 470 and 530 nm, respectively. The longer emission wavelength characteristic of this reagent makes it more selective for the analysis of biological materials. The fluorescence is pH dependent and is highest at low pH. Therefore, to obtain the best sensitivity, HCl must also be added postcolumn to enhance fluorescence quantum yield.

Although there is an increasing trend toward precolumn derivatization methods, there are several advantages to an ion-exchange separation with postcolumn derivatization. The first is that there is little sample preparation; protein hydrolysates can be injected directly into the column for analysis by LC. Unlike precolumn methods, it is not necessary to separate reaction side-products from the reaction mixture; instead, they become a source of background. With postcolumn methods there is more sample-to-sample consistency because of the robustness of the ion-exchange separation and the fact that the reaction time is determined by the column size.

When using postcolumn derivatization, the reaction does not have to go to completion and can produce multiple products. These do not affect the precision of the method as long as the mixing time and product ratio are reproducible. Finally, the derivatives do not have to be stable; they can start to degrade immediately after detection.

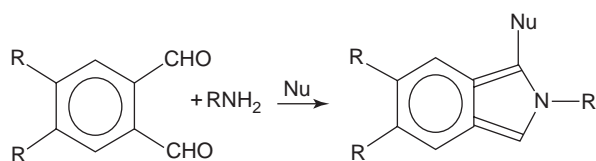
Two of the problems inherent in postcolumn derivatization are difficulties with baseline stability due to impurities in the mobile phase and late-eluting peaks. Ammonia, in particular, can interfere with the detection of the basic amino acids. Also with postcolumn derivatization, the reaction conditions are restricted by the mobile phase used for the separation. Frequently, the column effluent must be adjusted to achieve optimal reaction conditions, including pH. Loss of resolution due to band broadening produced by the reaction coil employed for mixing the reagent with the column effluent can also be a problem. However, 'knitted coils' are now available that produce very little or no band broadening. Finally, equipment and reagent costs are much higher than with precolumn derivatization.

### Reversed-Phase Liquid Chromatography

Reversed-phase LC is becoming increasingly popular for the separation of amino acids. Reversed-phase columns are readily available commercially and exhibit higher efficiencies than most commercially available ion-exchange columns. They are also compatible with aqueous samples, since water is generally a major component of the mobile phase. Therefore, it is not necessary to employ additional sample preparation steps in order to produce a sample in a nonaqueous environment.

**Precolumn derivatization** For separation by reversed-phase chromatography, precolumn derivatization is a necessity. The derivatization step increases not only detectability of the analyte but also its hydrophobicity, which makes the separation of the amino acids by reversed-phase chromatography possible. Several precolumn derivatization reagents are available. These include OPA, naphthalenedialdehyde (NDA), dimethylaminonaphthalene sulphonyl chloride (DANSYL), phenyl isothiocyanate (PITC), and *N*-(9-fluorenyl)methoxycarbonate (Fmoc).

Reagents based on dialdehyde chemistry have been shown to be highly sensitive for fluorescence and/or electrochemical detection of amino acids. They are both fluorogenic and electrogenic; that is, the compounds themselves are not fluorescent or electroactive but produce derivatives that exhibit both properties. The overall reaction scheme for these



**Figure 2** Reaction of aromatic dialdehydes with primary amines to form isoindole derivatives. Nu, nucleophile.

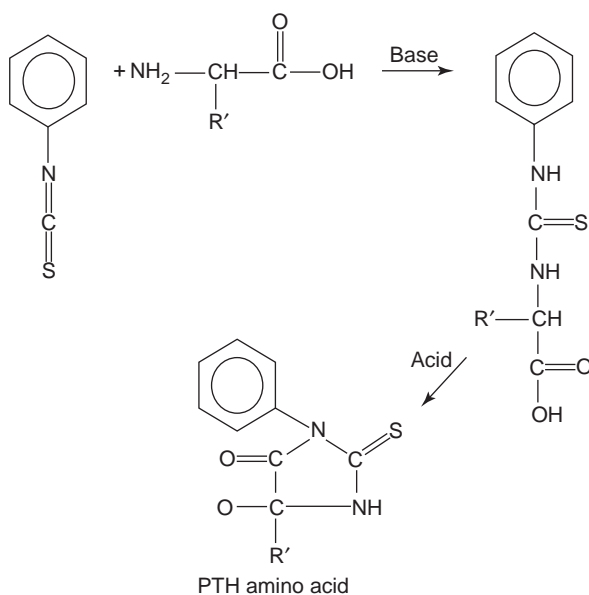
types of compounds is shown in **Figure 2**. OPA reacts very quickly with primary amines, but the derivatives produced are not very stable. Therefore, almost all precolumn derivatization employing OPA is done with an automated sample preparation system that can add reagents to the sample and precisely monitor the reaction time. More stable derivatives can be produced if either *t*-butyl thiol or thiosulfate is employed as the nucleophile. However, these derivatives are not fluorogenic and thus can only be assayed by electrochemical detection. Detection limits for OPA derivatives of amino acids are frequently reported to be less than 1 pmol.

Recently, an analog of OPA, NDA, which yields more stable derivatives, was introduced. NDA reacts with primary amines in the presence of cyanide to produce highly fluorescent cyanobenz-[f]isoindole (CBI) derivatives. The reaction time is ~20 min, as opposed to 2 min with OPA. However, the resulting CBI derivatives are considerably more stable than their OPA/mercaptoethanol counterparts; most derivatives exhibit little or no degradation after 16 h in the dark. Detection limits as low as 100 fmol have been reported.

Because both of these reagents exhibit fluorescence quenching, problems arise in the determination of lysine and cysteine. Since they also react only with primary amines, special steps must be undertaken to detect proline.

The DANSYL reagent reacts with both primary and secondary amines to produce fluorescent derivatives. These derivatives are very stable when stored in the dark and authentic standards are commercially available. One of the problems with this derivatization chemistry is hydrolysis of the reagent and the formation of multiple derivatives for several of the amino acids. In addition, there is fluorescence quenching in polar solvents, so detection limits using reversed-phase chromatography with UV detection are often comparable to that obtained by fluorescence. The limit of detection using UV detection is ~100 pmol.

PITC chemistry is becoming increasingly popular for the analysis of amino acids. This is extracted from the Edman degradation chemistry. The reaction scheme is illustrated in **Figure 3**. The most popular



**Figure 3** Reaction of primary and secondary amines with PITC. Formation of PTC and PTH amino acids.

way to detect amino acids using this chemistry is to perform only the first step of the Edman degradation reaction. The reaction conditions consist of derivatization of the amino acids for 5 min in a mixture of acetonitrile, pyridine, triethylamine, and water in the ratio 10:5:3:2 (v/v/v/v). Excess reagent is removed by vacuum and the sample is dissolved in mobile phase prior to injection. The phenylthiocarbamyl (PTC) derivatives that are produced can be determined by UV detection. Detection limits using this method have been reported to be as low as 1 pmol. PTC derivatives can be converted into phenylthiohydantoin (PTH) derivatives. However, this procedure adds a considerable amount of time to the assay procedure.

FMOC is a reagent that reacts quickly with both primary and secondary amino acids. Excess reagent must be removed rapidly because it undergoes hydrolysis. The excitation and emission wavelengths are 270 and 315 nm, respectively. An increasingly popular method for amino acid analysis is to measure the primary amino acids with OPA and the secondary amino acids with FMOC. Samples are derivatized first with OPA and then with FMOC. Primary amines are detected using the fluorescence properties of OPA while the secondary amines are detected using those of FMOC.

The advantages of precolumn over postcolumn derivatization include the ability to optimize reaction conditions without concern for their compatibility with the chromatographic conditions. Reaction time, pH, solvent, and reagent concentration can all be optimized to obtain the maximum yield of product.



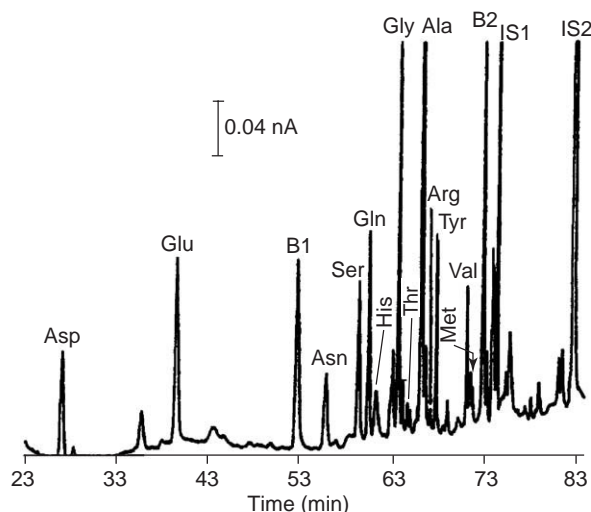
Side-products (which create baseline noise in post-column methods) can be eliminated via a cleanup step or separated chromatographically from the peaks of interest.

Precolumn derivatization methods are not free from disadvantages. One is that side-products from the reaction can sometimes be difficult to separate from the peaks of interest. Also with precolumn derivatization, as its name implies, each sample must be derivatized prior to injection. This can add to the total analysis time. However, instrumentation is now available to derivatize several samples simultaneously in a closed system, which not only solves the sample preparation time problem but also results in highly reproducible chemistry. Since the reaction may not go to 100%, an internal standard is frequently necessary. Finally, the addition of a similar 'tag' to all the amino acids may make them more difficult to separate. Table 1 summarizes the properties of several pre- and postcolumn derivatizing reagents employed for amino acid analysis.

**Microbore and open-tubular chromatography** In order to increase the sensitivity of LC for the analysis of amino acids, the techniques of microbore and open-tubular chromatography are becoming increasingly important. These techniques provide greater mass sensitivity because there is less dilution of the sample as it moves down the chromatographic column. Separations in which capillaries of 100  $\mu\text{m}$  or less ID were used provided the capability of analyzing very small samples, including single cells. Figure 4 shows the determination of several NDA-derivatized amino acids in a single neuron by open-tubular LC.

### Capillary Electrophoresis

Capillary electrophoresis (CE) was first introduced in 1983. Since that time there have been numerous reports of the determination of amino acids by CE. The simplest system consists of a high-voltage power supply, two buffer reservoirs, a transparent plastic safety box, two platinum wires, a detector, and a fused silica capillary. The capillaries employed have very small diameters, generally between 5 and 100  $\mu\text{m}$ . In general, the same types of detectors are used as for liquid



**Figure 4** Chromatogram of NDA-tagged amino acids from a single E4 neuron from *Helix aspersa*. B1 and B2 are peaks present in blank runs. Trp is obscured by B2 and Ile, Phe, and Leu are the three peaks immediately following B2. IS1 and IS2 are internal standards norleucine and normetanephrine, respectively. All unlabeled peaks are unknown. (Reproduced with permission from Oates and Jorgensen (1990) *Analytical Chemistry* 62: 1575.)

**Table 1** Comparison of reagents for analysis of amines

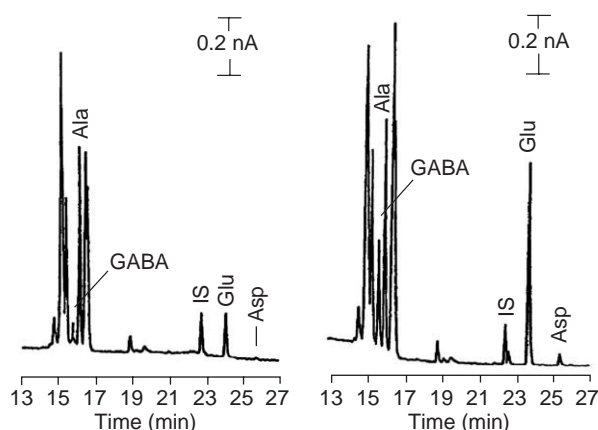
	Ninhydrin	OPA	FMOF	PITC	DANSYL	NDA/cyanide
Primary amines	Yes	Yes	Yes	Yes	Yes	Yes
Secondary amines	Yes	No	Yes	Yes	Yes	No
Stable derivative	NA	No	Yes	Yes	Yes	Yes
Fluorescence detection	No	Yes	Yes	No	Yes	Yes
Electrochemistry detection	No	Yes	No	Yes	Yes	Yes
Chemiluminescence detection	No	NA	NA	No	Yes	Yes
UV-vis detection	Yes	Yes	Yes	Yes	Yes	Yes
Reagent interferes <sup>a</sup>	No	No	Yes	Yes	Yes	No
Aqueous stability	Yes	Yes	No	No	No	Yes
Postcolumn	Yes	Yes	No	No	No	No
Precolumn	No	Yes	Yes	Yes	Yes	Yes
Simple derivatization <sup>b</sup>	Yes	Yes	Yes	No	No	Yes
Limits of detection <sup>c</sup>	High pmol	fmol	fmol	Low pmol	fmol	Low fmol

<sup>a</sup> Reagent itself is fluorescent, chemiluminescence, UV-vis, or electrochemistry detectable under the same conditions as the product.

<sup>b</sup> Number of steps, length of incubation.

<sup>c</sup> Optimal conditions (not routine).

Reproduced with permission from Lunte and Wong (1990) *Current Separations* 10: 20.



**Figure 5** Detection of aspartate, glutamate, alanine, and GABA in a brain microdialysate sample by capillary electrophoresis/electrochemistry. (Reproduced with permission from O'Shea *et al.* (1992) *Journal of Chromatography* 608: 189.)

chromatographic separations with some modifications. UV detectors are modified versions of commercially available instruments. Fluorescence detectors usually rely on a laser as the excitation source because it is easy to focus on the capillary wall. Electrochemical detectors employ carbon fibers rather than the conventional glassy carbon disk electrodes.

In CE, the primary separation mechanism is based on charge. The different amino acids exhibit different electrophoretic mobilities, when a potential is applied across the capillary. However, due to the presence of ionized silanol groups on the fused silica surface, an electroosmotic flow is set up, which is greater than the electrophoretic mobility of the analytes. Therefore, all analytes, regardless of charge, migrate toward the cathode. Many of the amino acids are separated by CE using a fused-silica capillary and a simple alkaline buffer such as borate. However, in order to resolve the neutral amino acids completely, micellar CE is frequently employed. In this case, sodium octyl sulfate or another surfactant is added to the mobile phase. Derivatized amino acids partition in and out of the micelle, yielding a separation that is based both on sample hydrophobicity and charge.

One of the advantages of CE is that it can be employed for the analysis of very small samples. Generally, only a few nanoliters of sample are required for analysis. Hence, CE has been employed for the analysis of amino acids in single cells, small tissue extracts, and microdialysis samples. **Figure 5** shows a separation of NDA amino acids from a rat brain dialysate using CE with electrochemical detection.

### Thin-Layer Chromatography

If qualitative information is desired, TLC can be employed for amino acid analysis. Paper, silica gel,

and cellulose have all been employed for the separation of amino acids. Spots are normally detected by spraying the plate with a 0.1% solution of ninhydrin dissolved in acetone. The plate is heated for 10 min at 105°C. An alternative method of development is the use of ninhydrin containing cadmium acetate. This plate can be developed by soaking overnight in sulfuric acid. Quantification by scanning densitometry is also possible.

Combined TLC electrophoresis enhances the separation of amino acids. Amino acids are separated in one dimension on the TLC plate and then a voltage is applied across the other direction.

### Gas Chromatography

Gas chromatography (GC) has also been employed for the analysis of amino acids. The primary advantage of GC is that it can be combined with mass spectrometry (MS) for the identification of rare or unusual amino acids in biological samples.

**Derivatization** Most amino acids are nonvolatile and must be derivatized prior to separation by GC. Amino acids are both esterified and acylated to produce *N*(*O,S*)-acylamino acid esters, with the most common being *N*(*O,S*)-trifluoroacetyl-*n*-butylesters. Other alcohols of low relative molecular mass have also been employed for the esterification including methanol, propanol, isopropyl alcohol, butanol, or isobutyl alcohol. Several different acylating reagents have been explored as well, including acetic anhydride, trifluoroacetic anhydride, pentafluoropropion anhydride, or heptafluorobutyric anhydride.

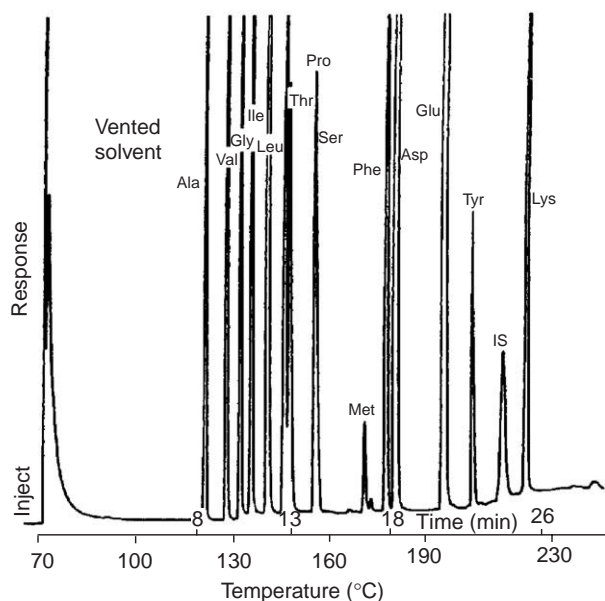
To guarantee quantitative yields of ester derivatives, the amino acid extract must be free of water. Following derivatization, the reaction solvent is evaporated under nitrogen at low temperature to ensure that there is no loss of volatile esters. The amino acids are then acylated. Reaction conditions vary from 2 h at room temperature to 10 min at 150°C. Prior to injection, the solution is evaporated to dryness and reconstituted in a solvent compatible with GC. Arginine can cause problems in GC analysis due to its low volatility.

**Detection** Femtomole detection limits can be achieved using electron capture detection (ECD). Nitrogen-phosphorus and flame ionization detectors have also been employed. However, the primary advantage of GC is its compatibility with MS. Using a mass-selective detector, amino acids can be determined at the low femtomole level. **Figure 6** shows the

separation of amino acids as their *N*-TFA *n*-butyl esters present in a soy bean meal hydrolysate by GC-ECD.

### Mass Spectrometry

There are some cases in which MS has been employed for the determination of amino acids without a separation step. Amino acid samples derivatized with Sanger's reagent can be 'separated' based

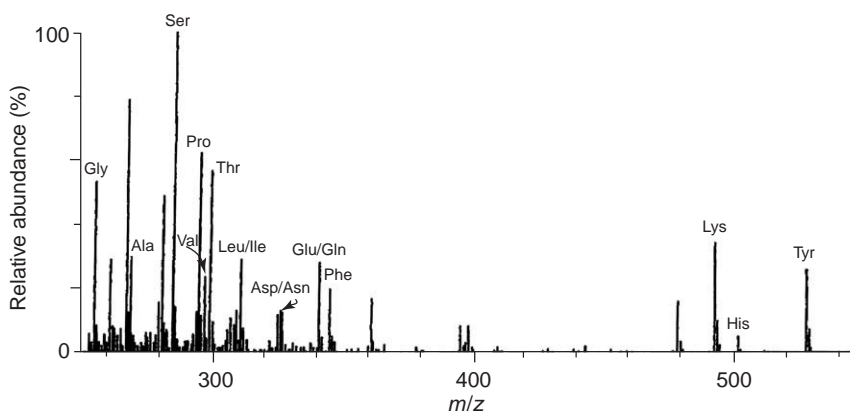


**Figure 6** GLC analysis of soy bean meal hydrolysate. Experimental conditions: 25 mg of soy bean meal hydrolyzed for 22 h at 110°C; cleaned by cation exchange; approximately 18 µg total amino acid injected. (Reprinted with permission from Zumwalt RW, Kuo KCT, and Gehrke CW (eds.) (1987) *Amino Acid Analysis by Gas Chromatography*, vol. 1, p. 38. Boca Raton, FL: CRC Press; © CRC Press, Boca Raton, Florida.)

on mass using negative chemical ionization MS. **Figure 7** shows an example of a plasma sample derivatized by this reagent and analyzed by MS. Despite this illustration, MS is rarely used for the determination of mixtures of amino acids without being coupled to a separation step. Generally, *N*-perfluoroacetylaminobutyl esters are determined by GC-MS. Due to the high molecular weights of the *N*(*O,S*)-acylalkyl esters, chemical ionization (CI) is the preferred method of ionization. In CI, a reaction gas such as methane is introduced into the mass spectrometer producing  $\text{CH}_5$  and  $\text{C}_2\text{H}_5^+$  ions. These ions then react with the *N*(*O,S*)-acyl esters to produce an  $M+1$  ion through the addition of a proton. Methane is frequently employed as the carrier gas for GC experiments that employ MS as a detector.

In GC-MS, compounds can be analyzed in one of two ways. If compound identification is the object of the assay the detector is run in the scanning mode and a mass spectrum of the compound is obtained as the compound elutes from the gas chromatograph. Alternatively, selective ion monitoring (SIM) can be employed for increased sensitivity. In this case, only compounds containing a characteristic ion in their mass are detected. Using SIM, detection limits are 2–3 orders of magnitude lower than those with scanning detection (frequently as low as 0.1 pg for negative ion CI detection of an electronegative group).

Liquid chromatography-mass spectrometry (LC-MS) is also becoming an increasingly popular analytical tool. Derivatives employed for LC analysis with UV or fluorescence detection have also been studied by LC-MS. It is particularly useful for the identification of rare amino acids and verification of peak purity.



**Figure 7** Detection of amino acids in plasma as their dinitrophenyl amino acid methyl esters by negative ion chemical ionization mass spectrometry. (Reproduced with permission from the Ph.D. Dissertation of Todd Williams, University of Maine, December 1987; Figure 11, p. 99.)

## Special Problems

One of the biggest problems in trace amino acid analysis is contamination of the sample by environmental amino acids. A fingerprint placed on the side of glassware can consist of up to 10 pmol of several amino acids. When performing amino acid analysis of proteins, special care must be taken to avoid contamination of the sample. Glassware should be cleaned thoroughly and frequently pyrolyzed to assure no amino acid contamination. Gloves should be worn when preparing solutions and performing the derivatization. One should be certain to use high-purity reagents and chemicals.

For the detection of amino acids in biological samples, a protein precipitation step is generally necessary. For ion-exchange-based separations, sulfosalicylic acid is the most popular reagent. Because precolumn derivatization is almost always employed for reversed-phase systems, methanol (which does not affect the pH) is usually utilized as the precipitation reagent. For GC studies, picric acid is generally used. Once the protein has been removed, samples can be injected directly into a cation-exchange column or derivatized for reversed-phase LC and GC analysis. Further purification prior to injection is often accomplished using an ion-exchange cartridge.

In the case of amino acid analysis, the quantification of cysteine can be difficult because it is oxidized to cystine during acid hydrolysis. To circumvent this problem, cysteine can be oxidized to cysteic acid with performic acid prior to analysis. Alternatively, cysteine can be converted to the pyridyl ethyl derivative and subsequently detected by postcolumn reaction with ninhydrin. Still another method involves the production of carboxymethyl cysteine following alkylation with iodoacetic acid. All of these cysteine derivatives can be separated by either reversed-phase precolumn or ion-exchange postcolumn methods.

Overall, the analysis of amino acids at trace levels continues to be a challenging problem. Reversed-phase LC methods based on derivatization of the amino acids are still not as robust as classical ion-exchange methods, although they are frequently

more sensitive. The presence of contaminating amino acids in buffers and glassware continues to be a source of frustration for the analyst interested in quantifying amino acids at the subpicomole level. Further development of automated equipment for amino acid analysis should eliminate or greatly reduce contact with human skin and, thus, many of these contaminants.

**See also:** **Amperometry.** **Capillary Electrophoresis:** Clinical Applications. **Clinical Analysis:** Overview; Sample Handling. **Derivatization of Analytes.** **Fluorescence:** Derivatization. **Gas Chromatography:** Mass Spectrometry. **Ion Exchange:** Overview. **Liquid Chromatography:** Packed Capillary; Normal Phase; Reversed Phase; Liquid Chromatography–Mass Spectrometry. **Mass Spectrometry:** Selected Ion Monitoring. **Sampling:** Practice. **Spectrophotometry:** Organic Compounds.

## Further Reading

- Beesley TF and Scott RPW (1998) *Chiral Chromatography*. New York: Wiley.
- Blau K and Halket JM (eds.) (1993) *Handbook of Derivatives for Chromatography*. Chichester: Wiley.
- Deyl Z, Hyanek J, and Horakova M (1986) Profiling of amino acids in body fluids and tissues by means of liquid chromatography. *Journal of Chromatography* 379: 177–201.
- Gehrke CW, Wixom RL, and Bayer E (2001) *Chromatography a Century of Discovery 1900–2000, the Bridge to the Sciences/Technology*. Amsterdam: Elsevier.
- Grunau JA and Swiader JM (1992) Chromatography of 99 amino acids and other ninhydrin reactive compounds in the Pickering lithium gradient system. *Journal of Chromatography* 594: 165–175.
- Sherma J and Fried B (eds.) (1996) *Handbook of Thin-Layer Chromatography*. New York: Dekker.
- Smith JT (1997) Developments in amino acid analysis using capillary electrophoresis. *Electrophoresis* 18: 2377–2378.
- Zhou J, Hefta S, and Lee TD (1997) High sensitivity analysis of phenylthiohydantoin amino acid derivatives by electrospray mass spectrometry. *Journal of the American Society for Mass Spectrometry* 8: 1165–1168.

# AMPEROMETRIC OXYGEN SENSORS

See **SENSORS: Amperometric Oxygen Sensors**

# AMPEROMETRY

**S B Adeloju**, Monash University, Gippsland, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Amperometry is an electroanalytical technique that involves the application of a constant reducing or oxidizing potential to an indicator (working) electrode and the subsequent measurement of the resulting steady-state current. Thus, the term ‘amperometry’ is derived from the current measurement unit, ‘ampere,’ and the measurement device, ‘meter,’ used in this method. Usually, the magnitude of the measured current is dependent on the concentration of the reduced or oxidized substance, and hence this method can be used for various analytical applications.

This method offers the ability to distinguish selectively between a number of electroactive species in solutions by judicious selection of the applied potential and/or choice of electrode material. In cases where a nonspecific potential is applied, the resulting current may be contributed by several electroactive species. The careful choice of the composition of the supporting electrolyte may also be useful in improving the selectivity of amperometric methods. In cases where oxygen interferes, its removal by purging the solution with an inert gas such as nitrogen or argon may be necessary.

Various analytical methods now employ amperometric measurements as part of their procedures. In particular, amperometric titrations have been widely used for the analysis of various substances in samples ranging from water to radioactive materials. Also, amperometric sensors, such as the dissolved oxygen probe and various amperometric biosensors, are widely used for clinical, environmental, and industrial monitoring. Furthermore, amperometric detectors have gained considerable use since the 1970s in high-performance liquid chromatographic determination of various substances and in flow injection analysis.

## Principles

The electrochemical oxidation or reduction of an electroactive species on a suitable electrode material by the application of the desired potential results in

either a steady-state anodic or cathodic current. (It is important to note that the current measured in such a system is distinctly different from the current usually measured in the so-called ‘galvanic cell’, which does not require the application of a potential.) According to the Cottrell equation, the resulting current can be related to the concentration of the electroactive species as follows:

$$I = \frac{nFAD^{1/2}C_b}{\pi t} \quad [1]$$

where  $I$  is the diffusion current ( $\mu\text{A}$ ),  $t$  is the electrolysis time (s),  $n$  is the number of electrons involved in the electrode reaction,  $F$  is the Faraday constant ( $96\,487\text{ C equiv}^{-1}$ ),  $A$  is the electrode area ( $\text{m}^2$ ),  $D$  is the diffusion coefficient of the electroactive species ( $\text{m}^2\text{s}^{-1}$ ), and  $C_b$  is the bulk concentration of the electroactive species ( $\text{mmol l}^{-1}$ ). The magnitude of the resulting current will thus be directly proportional to the concentration of the reduced or oxidized substance. This technique therefore provides a simple and reliable method for the quantification of various substances based on the relationship between the measured current and concentration of the analyte.

Owing to the time dependence of the resulting anodic or cathodic current, the reliable analytical utilization of amperometry requires measurement at a fixed time interval. However, the variable nature of the resulting anodic and cathodic currents can be useful in elucidating the electrochemical process(es) on various electrode surfaces. This has indeed led to the development of a variant of this technique, known as ‘chronoamperometry’, which involves the application of the desired anodic or cathodic potential and the subsequent measurement of the resulting current versus time. The main requirements in chronoamperometry are that:

1. The transport of the reactants and products to or from the electrode surface must be governed by diffusion without stirring of the solution.
2. The electrode surface must remain constant and must not be involved in the electrode process during the measurement.

The application of the desired potential to the electrode in amperometry results in a decrease in the concentration of the electroactive species on the electrode surface to zero due to the anodic or cathodic reaction(s) taking place at the vicinity of the electrode. In general, the rate of current change will vary

depending on the geometry of the chosen electrode and the diffusion properties of the substance. With a flat type of electrode, where linear diffusion is assumed, the rate of current change can also be expressed by the Cottrell equation as

$$\Delta I = \frac{nFAD^{1/2}C_b}{(\pi t)^{1/2}} \quad [2]$$

According to this equation, the current decreases toward zero with an increase in the electrolysis time and, hence, indicates that the concentration of the electroactive substance in the vicinity of the electrode decreases progressively with time. In situations where nonlinear diffusion of the electroactive substance to the electrode surface (i.e., spherical diffusion) is involved, the relationship between the rate of current change and other parameters are more complex.

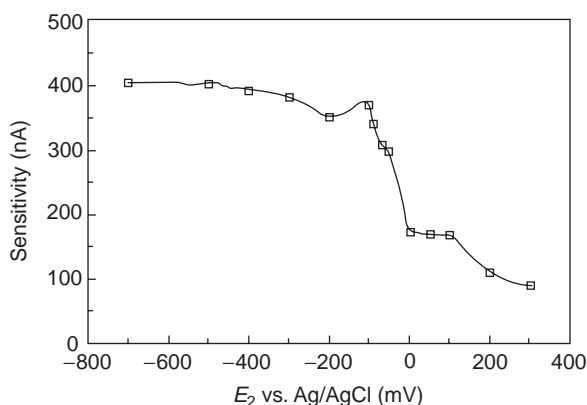
The concept of chronoamperometry has been further extended to develop a more advanced variant known as 'double-potential-step chronoamperometry.' This mode of amperometry is of considerable interest in electrochemical research, and in pulsed amperometric detection in chromatography and flow injection analysis. In double-potential-step chronoamperometry, the applied potential is changed twice during the measurement, first by applying a potential to induce oxidation or reduction at the electrode surface, as in the normal amperometric mode. The potential is then changed again, after a short electrolysis time, to either the initial value or a new value. The second potential step is often chosen to enable the product of the initial electrode process to be converted back to either its original form or to another product. **Figure 1** shows how the choice of the second potential ( $E_2$ ) is successfully used to influence the sensitivity of a pulsed amperometric polypyrrole-based biosensor for urea. Evidently, improved sensitivity can be obtained for pulsed amperometric detection of urea with the biosensor by careful selection of the second applied potential.

The various modes in which amperometry can be employed have resulted in its extensive utilization in analytical chemistry – in amperometric titration, amperometric sensing, and amperometric detection in flowing systems. The various applications of amperometry are discussed below.

## Amperometric Titrations

### Basic Theory

The most important feature in amperometry is the linear relationship between the diffusion current and the concentration of the electroactive species (this also applies to other electroanalytical techniques

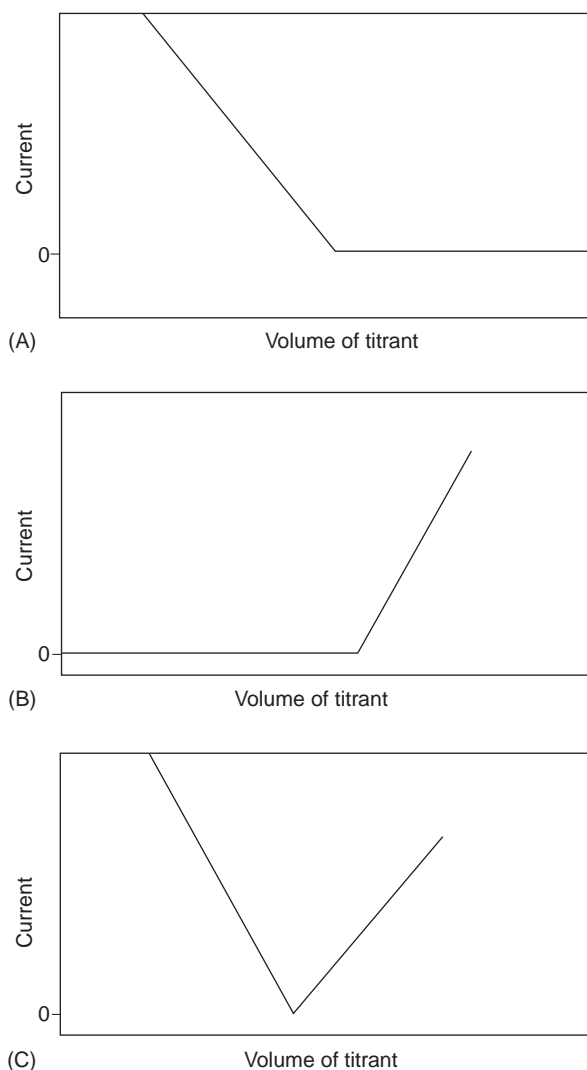


**Figure 1** Influence of the applied second potential on the pulsed amperometric detection of urea with a polypyrrole – urase biosensor.  $E_1$  was set  $-70$  mV, while  $E_2$  was varied. Pulse width was 120 ms.

such as polarography and voltammetry). Consequently, as demonstrated in eqn [2], the measurement of current in amperometry offers a simple and reliable approach for the monitoring of the changes in concentration of the electroactive species that occur during a titration. The main requirement in such a measurement is that at least one of the substances (reactant or product) involved in the titration is electroactive. This ensures that the current flow during the titration is due to either the oxidation or reduction of the reactant or product. Thus, the plot of the resulting current from the titration against the titrant volume will enable the location of the equivalence point. The observed current changes during such a titration is proportional, as suggested by eqn [2], to changes in the concentration of the electroactive species. **Figure 2** shows that the titration curves usually consist of two straight lines and their point of intersection gives the equivalence point.

The method described above is known as 'amperometric titration' and it is widely used in analytical chemistry. Such titrations can be performed in two ways: (1) by adjusting the potential of a single microelectrode within the range of the limiting current of the electroactive species, as in polarography, and subsequently recording the current flow as a function of increasing titrant volume; and (2) by applying a fixed potential to the indicator electrode (i.e., holding the potential difference between a pair of microelectrodes constant), while the current is measured as a function of the titrant volume. The mode in which such a titration is performed can involve either the use of one or two indicator electrode(s), depending on the nature of the titration. These possibilities give amperometric titration an excellent and unique flexibility that, in some cases,





**Figure 2** Amperometric titration curves for systems where (A) only analyte is reduced; (B) only the reagent is reduced; and (C) both analyte and reagent are reduced.

enables the monitoring of the reactive product or reagent, rather than of the analyte.

### Instrumentation

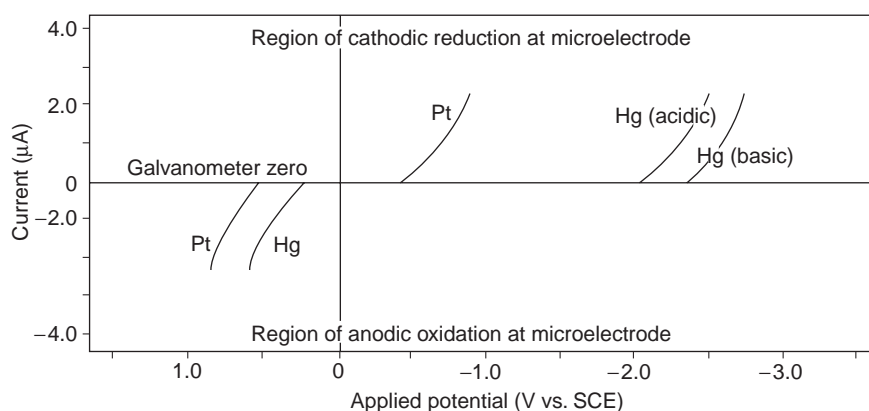
A simple two-electrode electrochemical cell consisting of either single or dual polarizable electrode(s) is normally required for amperometric titrations of various organic and inorganic substances. By definition, a polarizable electrode is a suitable electronic conductor whose potential changes even with the passage of relatively small current. In contrast, the potential of a nonpolarizable electrode, such as the saturated calomel and silver – silver chloride electrodes that are commonly employed as reference electrodes, remains reasonably constant even when a large current is passed.

Amperometric titrations with single polarizable electrodes are based on earlier work that evolve from the use of polarographic techniques. Most of these involved the application of a constant potential to a dropping mercury electrode (DME) and the subsequent monitoring of the current changes as a function of the titrant volume. In cases where the electroactive species are oxidized at potentials beyond the anodic potential range of mercury, a rotating platinum electrode (RPE) may be employed instead of the DME. **Figure 3** illustrates the anodic and cathodic potential limits for mercury and platinum microelectrodes. As can be seen, the use of mercury microelectrodes is limited in the positive (anodic) potential region by the oxidation of water at  $\sim +0.25$  V versus the SCE and in the negative region by the liberation of hydrogen at  $\sim -1.8$  V in acid or  $\sim -2.3$  V in basic media. In contrast, the use of platinum microelectrodes is limited at  $\sim +0.65$  and  $\sim -0.45$  V versus the SCE by the oxidation of water and hydrogen liberation, respectively. Evidently, mercury microelectrodes offer a more substantially negative potential limit for the reduction of electroactive species, whilst platinum microelectrodes permit oxidation at a more positive potential than can be accomplished with mercury microelectrodes.

An RPE can be more easily constructed by using a length of platinum wire sealed in a glass tube and mounted on a synchronous motor to permit rotation at a constant speed. The rotation of the electrode reduces the thickness of the diffusion layer and improves the current densities by a factor of 20–30 times over those obtained with a DME. Despite the higher sensitivity attainable with the RPE, the shape of the amperogram is often similar for both electrodes. However, the magnitude of the measured current with the RPE is sensitive to changes on the electrode surface and, hence, it is more variable from one titration to the other than with a DME. Also, unlike with DME, there is little or no charging current associated with the use of RPE.

The removal of oxygen from solution is almost always mandatory when a DME is employed for amperometric titration, as this substance can be readily reduced on this electrode. In contrast, this does not usually create a problem when an RPE is used for such a titration and therefore it is often possible to use a simple open cell for the titration. In cases where oxygen is a problem, this can be overcome by bubbling nitrogen or argon through the sample solution for 5–20 min and over the solution during the titration.

The choice of titrant concentration is another important consideration in amperometric titration. Substantial dilution of the titrant may result in a



**Figure 3** Attainable anodic and cathodic potential limits for platinum and mercury microelectrodes. (Reproduced with permission from Ewing GW (1985) *Instrumental Methods of Chemical Analysis*, 5th edn., p. 290. New York: McGraw-Hill.)

nonlinear variation of the resulting current, before and after equivalence point, with the titrant volume. The curvature of the straight-line portions of the curve can be corrected by multiplying each current reading by a factor  $(V_1 + V_2)/V_1$ . Prior to the equivalence point,  $V_1$  is the initial volume of the sample and  $V_2$  is the titrant volume. After equivalence,  $V_1$  is the equivalence point volume and  $V_2$  is the titrant volume added beyond the equivalence point. The effect of dilution can be reduced or eliminated by using more concentrated titrant.

Apart from the selection of an appropriate electrode material and titrant concentration, the main requirement in amperometric titration is the choice of either a single or dual polarizable electrode system. The basic differences and characteristics of these two titration systems are discussed below.

#### Titration with a Single Polarizable Electrode

Amperometric titration with a single polarizable electrode utilizes an indicator electrode (polarizable) and a nonpolarizable electrode in the form of a simple working-reference electrode system in which the potential of the former (working/indicator electrode) is held constant relative to the latter (nonpolarizable electrode). As previously illustrated in **Figure 2**, the current flowing due to the oxidation or reduction of the electroactive species in such a system is recorded as a function of titrant volume. For a precipitation titration in which the current flow is caused by the reduction of the electroactive species, increasing addition of the titrant results in the precipitation of the analyte as an insoluble substance and, subsequently, in a decrease in the current flow. **Figure 2A** shows that the current for such a titration will drop to zero and thereafter remain unchanged even when excess titrant is added. The equivalence point for such a titration is determined by drawing the two straight

lines and extrapolating to the intersection to locate the end point. In contrast, **Figure 2B** shows the titration curve for a system in which the titrant is reducible, but the analyte is unreactive. In this case, the potential of the electrode is held at a value where only the electroactive species can be reduced. **Figure 2C** illustrates a titration curve for a system in which both the analyte and the titrant are reducible within the same potential range. In other cases where the electroactive titrant may undergo oxidation at an RPE, the direction of the titration curve will be opposite to that shown in **Figure 2B** and, hence, the resulting current beyond the equivalence point will decrease.

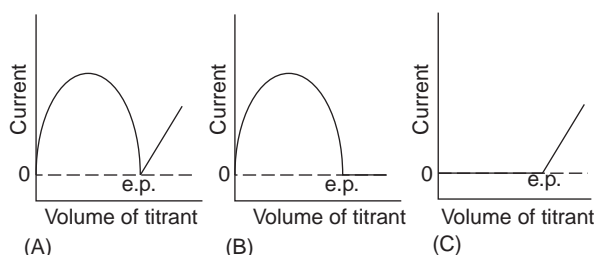
#### Titration with Dual Polarizable Electrodes

This type of amperometric titration utilizes two identical polarizable (indicator) electrodes, in the form of a two working (working-working) electrode system with no reference electrode. A constant potential difference of 10–500 mV is maintained between the two electrodes, but the potential of each electrode may change during the titration and may be outside the limiting current region for the electroactive species. However, under controlled experimental conditions a plot of the resulting current versus titrant volume gives the equivalence point by an abrupt current change, as illustrated in **Figure 4**. Titrations with dual polarizable electrodes are often used for equivalence point detection in redox titration, where one of the electrodes functions as a cathode and the other as an anode.

#### Potential Sources

The polarizable and nonpolarizable electrodes may be connected to a potential source such as a power supply or a potentiostat, and the resulting current from the titration may be recorded with the aid of a





**Figure 4** Amperometric titration curves for redox reactions on dual polarizable electrodes: (A) reversible half-reactions for titrant and sample; (B) reversible half-reaction for sample only; and (C) reversible half-reaction for titrant only. e.p., equivalent point (Reprinted with permission from Peters GD, Hayes JM, and Hieftje GM (1974) *Chemical Separations and Measurements*; © Brooks/Cole, a division of Thomson Learning; www.thomson-rights.com. fax: 800 730-2215.)

microammeter or a damped galvanometer connected in series with the cell.

## Applications

The use of amperometry as an endpoint detection method has been widely adopted for the titration of substances at the millimole per liter level, with excellent analytical precision ( $\pm 1\%$  RSD at  $0.01 \text{ mmol l}^{-1}$  level). The majority of these titrations involve the formation of precipitates, while others involve complexometric and redox titrations. At the millimole per liter level, the accuracy of this method is better than that achievable with other electroanalytical methods and as good as those of spectrophotometric titration.

Amperometric titrations have been more widely applied for the titration of inorganic and organic substances. The titration of inorganic substances usually involves the determination of various anions, cations, or compounds and may involve the formation of precipitates, complexation, or redox reactions of the titrant or the analyte, or both. In this case the analyte, titrant, or both can be electroinactive, and hence can demonstrate all the features of the amperograms shown in Figure 2. Table 1 provides a list of some inorganic substances that can be determined by amperometric titration. Other inorganic substances that can be determined by amperometric titrations include boron, bismuth, calcium, cadmium, copper, hafnium, holmium, indium, magnesium, neodymium, nickel, scandium, samarium, terbium, thorium, thallium, tungsten, zinc bromide, fluoride, hyponitrite, iodate, sulfate, sulfite, and thiosulfate.

In contrast, the amperometric titration of organic substances usually involves the use of an electroreducible titrant in which the organic compound is often unreactive. The flow of current in such a system remains essentially at zero until the equivalence point is

reached, after which the current increases linearly due to the reduction of the excess titrant. Even in cases where both the titrant and titration product are electroactive, the potential can be adjusted to enable the reduction of only the titrant. Table 1 also provides a list of some organic substances that can be determined by amperometric titrations. Other organic substances that can be determined by amperometric titrations include aromatic amines, aromatic nitrosocompounds, ascorbic acid, benzotriazole, cysteine, dithiocarbamates, dodecylammonium bromide, dodecylpyridinium bromide, ethylenediaminetetraacetic acid (EDTA), formaldehyde, hydroxylamine salts, hydroquinone, ketones, thiazoles, organic mercapto groups, opium, salicylates, quaternary ammonium salts, sulfanilamide, alkaloids, sulfenamides, isothiocyanates, urea, and other amides.

## Chlorine in Water and Wastewater

Amperometric titrations have been successfully used for accurate determination of residual chlorine in water. Different species of chlorine have also been determined, with suitable modification of the method, as free available chlorine, chloramine, chlorine dioxide, and chlorite. Various dual polarizable electrodes have been used for the amperometric titration of low concentrations of residual chlorine in water. In particular, amperometric (iodometric) titrations with dual platinum electrodes have gained considerable interest in the determination of total residual chlorine in water. Selective iodometric titrations with dual polarizable electrodes have also been found to be useful for the speciation of chlorine in water, providing the ability to distinguish between free and combined residual chlorine, and between monochloramine and dichloramine in water. For example, in the absence of iodide only the residual chlorine is detected at the neutral pH range in water and, hence, with the addition of iodide at this pH range monochloramine can also be detected. Consequently, by bringing the pH down to 4 and adding more iodide, dichloramine can be detected as well. To obtain the free available chlorine concentration, chlorine dioxide must be removed by hydrolysis at a high pH prior to the titration of the neutral sample solution in the absence of iodide. If not removed, chlorine dioxide is reduced by iodide to chlorite quantitatively and 20% of this contributes to the concentrations obtained for free available chlorine and chloramine. The chlorite generated from the reduction of chlorine dioxide by iodide can be reduced quantitatively to chloride in acid solution and, hence, subsequent titration in neutral solution will measure chlorite, chlorine dioxide, free available chlorine, and chloramines. The presence of nitrogen trichloride has been found to

**Table 1** Analytical applications of amperometric titrations

<i>Sample</i>	<i>Analyte</i>	<i>Reagent/titrant</i>	<i>Polarizable electrode</i>
Insecticides	Total chlorine	Silver nitrate (after decomposition with Na and xylene)	Sliver
Water–petroleum refining	Chloride	Isoamyl alcohol–ethanol	RPE
Water	Arsenic	Potassium bromate	RPE
	Chloride	Silver ion	Graphite or RPE
Serum	Chloride	Silver ion	Graphite
Plating baths	Thiourea	Mercuric nitrate	Mercury
Spent sulfite liquor	Total sulfur	Lead nitrate (after oxidation to sulfate)	RPE
Catalysts	Palladium	Potassium hexacyanoferrate	RPE
		Potassium thiocyanate	Platinum
	Platinum	Cerium(IV)	Graphite or RPE
		Sodium thiosulfate after treatment with KI	RPE
Drugs	Aluminum	EDTA	RPE
Quartz glasses/glass	Vanadium	Glycinethymol blue	RPE
		EDTA	Rotating
			Graphite
Nuclear fuels	Uranium	Mohr's salt	RPE
Iron ores	Iron	Manganese(III)	RPE
Ores	Mercury	Thiosalicylic acid	RPE
Steels	Cobalt	Potassium hexacyanoferrate in ammoniacal medium	RPE
	Phosphorus	Conversion to molybdophosphate followed by the titration of Mo	Graphite or RPE
	Vanadium	Ferrocene	Graphite or RPE
		Copper(I)	RPE
		Mohr's salt	RPE
	Vanadate	Potassium iodide	RPE
		Naphthylamine	RPE
	Chromate	Potassium iodide	RPE
		Naphthylamine	RPE
	Chromium	Copper(I)	RPE
		Calcium caboxymethyldithiocarbamate	
	Manganese	Copper(I)	RPE
	Selenium	Ascorbic acid	RPE
	Tellurium	Ascorbic acid	RPE
Alloys	Antimony	Potassium dichromate	RPE
	Iron	Potassium dichromate	RPE
	Molybdenum	Chromium(II)	RPE
	Rhenium	Chromium(II)	RPE
Brass	Manganese	Copper(I)	RPE
Mineral waters	Iodide	Potassium dichromate	Graphite or RPE
	Chloride	Mercurous nitrate	RPE
Brines and seawater	Iodide	Sodium thiosulfate in the presence of potassium iodate	RPE
		Palladium chloride	Graphite
Wastewater	Hydrazine	Bromine	RPE
	Phenol	Bromine	RPE
Rubber	Sulfur	Iodine (after conversion by sulfite to thiosulfate)	RPE
Biological samples	Mercury	Iodine	Graphite or RPE
Silicate materials	Zirconium	Xylenol orange (after fusion with sodium carbonate)	RPE
Ores	Gold	2-Imino-2-mercapto-thioacetamide	RPE
	Manganese	Thiourea	RPE
		Thiosalicylic acid	RPE
Oxalate baths	Iron	Cerium(IV)	RPE
Steels, slags, and other metallurgical products	Cerium	Mohr's salt	RPE
Iron and steel	Cerium	Methyldimercapto-thiopyrone	Graphite
Glasses and spinels	Vanadium(III)	Potassium dichromate	Graphite
	Vanadium(IV)	Chromium sulfate	Graphite

*Continued*

Table 1 Continued

<i>Sample</i>	<i>Analyte</i>	<i>Reagent/titrant</i>	<i>Polarizable electrode</i>
Bricks and cements	Chromium(VI)	Sodium arsenite	RPE
Steel, bricks, and clinkers	Manganese	Thiopiperidone	RPE
Binders	Chromium	Potassium iodide (after treatment with sodium chlorite)	RPE
Aluminum-based catalyst	Iridium and palladium	Potassium iodide	RPE
Organic substances	Selenium	Iodine (after treatment with KI and sodium thiosulfate)	RPE
Industrial materials	Selenium	2,4-dithioburet	RPE
Silver solder	Silver	Thiooxime	RPE
Gasoline	Lead	Thiosalicylamide	RPE or DME
Carbonate ore	Manganese	Potassium hexacyanoferrate	RPE
Agglomerates and ores	Silicon and titanium	Diantipyrylmethane	RPE
Copper and antimony alloys	Gallium	<i>N</i> -3-styrylacrylophenyl-hydroxylamine	Graphite
Magnesium alloys	Silver	<i>N</i> -cinnamoyl- <i>p</i> -tolyl hydroxylamine	Graphite
Electrolysis brine	Chlorate, chlorine, chlorite, hypochlorite	Ferrocence	RPE
		Arsenite-chloramine-T	DME or RPE
Propellants	Nitroglycerin	Sodium thiosulfate (after treatment with Mohr's salts and addition of KI)	RPE
Radiopharmaceuticals	Technetium	Tin(II)	RPE
Limestone	Calcium	EGTA	DME
Alloys and glasses	Yttrium and lanthanum	DTPA	Graphitic
Water	Total heavy metals	EDTA	DME
Steels and other materials	Vanadium and chromium	Eriochrome cyanine R	RPE
Iron ore concentrates	Iron	Potassium dichromate	RPE
Bauxite	Titanium	Methylene blue	RPE
Low grade ore	Uranium	Potassium dichromate	RPE
Tablets	Ascorbic acid	Dichloramine-T in presence of potassium iodide	Platinum-graphite
Water	Chlorine species	Phenylarsine oxide	RPE
Chinese herb drugs	Paeonol and catechin	Bromine	RPE
Brines and seawater evaporates	Iodide	Potassium iodate (after treatment with KI and thiosulfate)	RPE
Quartz, pyrites, and jewelery	Gold	Hydroquinone	RPE

DTPA, diethylenetriaminepentaacetic acid; DME, dropping mercury electrode; EDTA, ethylenediaminetetraacetic acid; EGTA, bis(aminoethyl)glycol ether-*N,N,N',N'*-tetraacetic acid; RPE, rotating platinum electrode.

contribute mainly to the free chlorine fraction and partially to the chloramine fraction. In these titrations, the endpoint is indicated by a decrease in the current to zero. The sample must be stirred thoroughly during the titration and the entrainment of air by cavitation must be avoided.

The presence of other oxidizing agents such as manganese(IV) can interfere with the chlorine determination if the titration is performed at  $\text{pH} < 3.5$ . Under this condition, the organic chloramines are often converted to either monochloramine or dichloramine, resulting in positive errors for these substances. Also, the volatilization of chlorine compounds can result in low recoveries. This can occur due to the violent agitation or aeration of some of the chloramines prior to the addition of iodide.

These losses can be avoided by using two portions of the sample: one for the determination of the free chlorine and the other treated with the titrant and iodide prior to the determination of the chloramines. Ferric and chlorate ions do not interfere, while nitrites and manganese dioxide are present only as chlorites.

As the residual chlorine in wastewater is usually combined, iodide is often added to liberate iodine, which is more active than the combined chlorine. Such liberation is necessary in view of the presence of amines, ammonia, organic nitrogen, and other organic compounds in wastewaters that tend to combine with a large proportion of the free residual chlorine. The losses of the liberated iodine at  $\text{pH} \leq 4$  are relatively small, but it is preferable to add excess

titrant before adding iodide and consequently back-titrating with standard iodine. This often involves the mixing of the standard titrant (potassium iodide) and the buffer, prior to adding the sample and subsequently titrating with a standard iodine solution. The use of a titrant, such as sodium arsenite, at this low pH is not possible and therefore either phenylarsine oxide or sodium thiosulfate is often employed. The use of phenylarsine oxide as a titrant is preferred over others due to its much higher stability. When employed as a titrant, the interferences from other oxidizing agents, such as manganese(IV), is avoided by performing the titration at pH 3.5. Chromate, silver, and cuprous ions are known to interfere with the amperometric titration of chlorine in wastewater. Also, this method cannot be performed in the presence of high concentrations of cupric ions. Problems with color and turbidity are easily avoided by the use of amperometric endpoint instead of the starch endpoint approach. Chlorine dioxide can interfere with the amperometric titration of both the free residual and combined chlorine. In cases where the organic chloramines react very slowly at  $\text{pH} \geq 3.5$ , the use of lower pH (1.0–2.5) is recommended.

The amperometric titrations of chlorine in water and wastewater are particularly useful in establishing the correlation between the degree of disinfection and residual chlorine. This titration is also often used as a standard method of comparison for the determination of free or combined chlorine. It is relatively more sensitive and has a detection limit of  $0.01 \text{ mg l}^{-1}$ , which is superior to the limit of  $\geq 1 \text{ mg l}^{-1}$  achievable by most colorimetric and titrimetric methods. Furthermore, the amperometric determination of chlorine is less affected by common oxidizing agents, color, turbidity, and temperature variations. However, the skill and care required for reliable amperometric titration is much higher than that required for other methods.

### Total Oxygen Demand

Two of the most useful parameters in the assessment of water quality are biochemical oxygen demand (BOD) and chemical oxygen demand (COD). Unfortunately, the measurement of these parameters is time consuming and suffers from some interferences. The BOD method requires a 5-day incubation period, while the COD measurement requires a complex decomposition of the sample solution.

An alternative method known as total oxygen demand (TOD) is based on the use of amperometry. It involves the passage of a carrier gas containing a small, but constant amount of oxygen through a heated tube containing a catalyst and then into a

galvanic type cell that consists of a silver cathode and a lead anode, covered with a layer of potassium hydroxide. After attaining a steady-state current, a small aliquot (20  $\mu\text{l}$ ) of water sample is injected into the carrier gas. Consequently, the oxidizable substances in the water react with oxygen in the catalyst furnace, resulting in an oxygen depletion that is detected amperometrically by a decrease in the current. TOD results often agree favorably with COD values when the latter are not affected by interferences. Each TOD measurement takes  $\sim 3$  min and is amenable to remote and/or intermittent monitoring by the use of an appropriate sample injection device.

### Total Available Heavy Metals

A square-wave amperometric titration has been used for the determination of the total available heavy metals in water samples. This method involves the direct anodic oxidation of mercury in the presence of excess EDTA. The resulting mercury wave is used to detect the endpoint of the amperometric titration by running a polarogram after each successive addition of an aliquot of EDTA. The successful utilization of this method lies in the ability to discriminate between  $\text{Ca(II)}$  and heavy metals, such as  $\text{Cu(II)}$  and  $\text{Zn(II)}$ . Thus, in practice, it involves 1:1 dilution of samples with  $0.2 \text{ mol l}^{-1}$  acetate buffer (pH 4.8), prior to the amperometric titration. At this pH, heavy metals, such as  $\text{Fe(III)}$ ,  $\text{Hg(II)}$ ,  $\text{Ni(II)}$ ,  $\text{Cu(II)}$ ,  $\text{Pb(II)}$ ,  $\text{Zn(II)}$ ,  $\text{Cd(II)}$ ,  $\text{Co(II)}$ , and  $\text{Al(III)}$ , are completely ( $\geq 99\%$ ) converted to EDTA complexes. Furthermore, the presence of  $\text{Ca(II)}$  does not interfere with the determination of the available heavy metals under these conditions. As little as  $1 \mu\text{mol l}^{-1}$  of available heavy metals has been successfully determined in water samples by this method.

## Amperometric Sensors

### Origin

Unlike other sensors, most of the available chemical amperometric sensors evolved from earlier investigation of various electrochemical processes on electrode surfaces, long before the use of the generic term of 'chemical sensor' became acceptable. Of course, as the emphasis of such earlier work was directed mainly toward the prediction and understanding of electrode mechanisms rather than for analytical purposes, most of these investigations were done under conditions that were not fully comparable with a modern sensing device. However, the considerable knowledge gained from the earlier work has been most beneficial in the design and development of amperometric sensors. This has led to the

development of reliable amperometric sensors for oxygen and chlorine, as well as to the development of biochemical sensors for various substances.

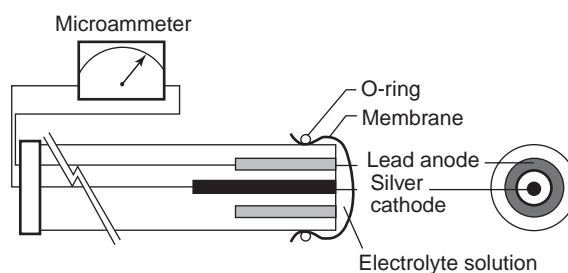
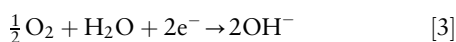
### Dissolved Oxygen Sensor

By far the most important application of amperometry is the determination of dissolved oxygen. An earlier approach for oxygen measurement made use of the galvanic type cell, which consists of a zinc anode and a carbon, gold or platinum cathode, and hence does not require the application of an external polarizing potential. Unfortunately, this cell arrangement was prone to poisoning when used in direct contact with samples containing surface-active substances and/or suspended solids. The prevention of the accumulation of deleterious substances on the electrode surfaces was achieved by covering them with a membrane. As well as keeping the electrode clean, the membrane establishes a finite diffusion layer that stabilizes the electrode performance. In effect, the membrane-covered electrodes become independent of the hydrodynamic properties of the sample solution.

The main requirement in the use of membranes in the construction of oxygen sensors is that the membrane must be permeable to oxygen, while restricting the passage of water and electrolyte. On this basis, the use of a single point calibration method based on the use of air, with the assumed linearity of response for other measurement, is feasible. Another calibration method involves the use of solutions previously verified by the Winkler oxygen titrimetric method.

One oxygen sensor design utilizes a concentrically mounted cathode and anode embedded in a plastic material, with only the surfaces of the electrodes exposed to the sample solution. In operation, a thin layer of the supporting electrolyte, such as potassium chloride or potassium hydroxide, is brought in contact with the electrodes by confinement within the outer membrane. The dissolved oxygen, which diffuses through the membrane, undergoes electrochemical reduction at the cathode within the confined layer. The resulting current from the reduction process is directly proportional to the oxygen concentration in solution.

Another galvanic oxygen sensor utilizes a lead anode, a silver cathode, and  $4\text{ mol l}^{-1}$  potassium hydroxide as electrolyte. In addition, a microammeter is connected in between the anode and cathode to provide the current readings. The flow of current, as in other oxygen sensors, is due to the diffusion of oxygen from the sample solution through the membrane to the silver cathode, where it is reduced:



**Figure 5** A galvanic oxygen sensor. The electrolyte is  $4\text{ mol l}^{-1}$  potassium hydroxide.

and the lead anode is subsequently oxidized in the presence of the hydroxide:

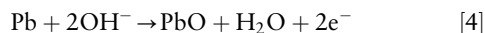


Figure 5 illustrates a typical cell arrangement for a galvanic oxygen sensor. The electrodes are insulated from each other with nonconducting plastic and the surface is covered with a permeable membrane. In practice, a few drops of the electrolyte is added between the membrane and the electrodes. The amount of oxygen that travels through the membrane is proportional to the dissolved oxygen concentration in the sample. The main requirement in using this sensor is to construct and calibrate the meter in such a way that the resulting current can be directly related to the oxygen concentration in sample solutions.

There are also nongalvanic oxygen sensors that are based on the same principles. This type of oxygen sensor typically consists of an inert metallic cathode covered with a gas-permeable membrane and a silver anode. These sensors are usually connected in series with a 1.5 V battery and an external potential of 0.8 V is applied between the electrodes to initiate the amperometric measurement of oxygen. One such sensor employs a gold cathode, a silver anode, and a potassium chloride gel as the electrolyte.

Owing to the realization in the 1970s that the current of a microelectrode ( $< 10\text{ }\mu\text{m}$ ) rapidly reaches a steady-state value that is not sensitive to mass transport, there has also been some considerable interest in developing microelectrode-based oxygen sensors. The construction and utilization of a microelectrode-based Clark-type oxygen sensor has gained wider use in the past two decades. In addition, the incorporation of such a micro-oxygen sensor on a chip for the multiampereometric sensing of oxygen and other substances, such as glucose and glutamate, with the aid of immobilized enzyme layers has been reported. Also, miniaturized oxygen sensors, placed either on the tip of a catheter or mounted in a microcell through which the sample is pumped or stirred into, are commonly used in clinical studies.

The available amperometric oxygen sensors are capable of determining dissolved oxygen concentration in aqueous media reliably when present between 0.1 and 50.0 mg l<sup>-1</sup> and they are readily amenable to field studies. Due to the membrane coverage of the electrodes, the sensors can be readily used in solutions containing both ionic and organic contaminants, as well as in gaseous environments. When used for the determination of oxygen in blood and other biological fluids, the covering membrane has also been found to be useful in preventing contamination and fouling by other blood constituents, such as erythrocytes. Potential interferences include sulfur dioxide, hydrogen sulfide, and other gaseous substances that can permeate the outer membrane.

A DME has also been used for the amperometric determination of oxygen in river water and effluent samples. This involves the application of a constant potential of -1.5 V versus the SCE. This approach has been found to be useful in eliminating interferences from cyanide and sulfide, as well as correcting for contributions from metal ions, such as iron, copper, zinc, and nickel. The method can be reliably used for oxygen determination when present within the concentration range of 0–15 mg l<sup>-1</sup>.

The various oxygen sensors have not only gained wider use in laboratory and in-field analysis, they have also become popular for online monitoring of industrial processes.

### Chlorine Sensor

Amperometric sensors, of similar design to the oxygen sensors, are also available for chlorine determination in water and wastewaters. In this case, the sensor employs a silver wire cathode, immersed in a sodium chloride solution in the inner tube, and a platinum wire anode wound around a porous thimble on a glass tube. The electrode assembly, which functions as a galvanic cell, is employed in a similar approach to the dissolved oxygen probe, giving a current reading upon immersion into sample solution. The resulting current is also directly proportional to the concentration of the active chlorine in the sample. This type of sensor has gained considerable use for domestic and industrial purposes.

### Biosensors

The immobilization of enzymes, cell cultures, tissues, and other biologically active substances on various electrode substrates have been explored extensively for the development of amperometric biosensors for various substances. This approach has been useful in improving the selectivity of amperometric measurements and hence for developing new amperometric

biosensors. Earlier work in this area concentrated on the construction of enzyme electrodes that employ oxygen sensors as transducers. The most well known and widely used enzyme electrode is the glucose electrode, which has even been used as an implantable sensor in clinical studies/monitoring. Several other transducers that are not based on the use of oxygen sensors are now widely used in this area and many new amperometric biosensors are reported regularly in the literature.

### Amperometric Detectors for Chromatography and Flow Injection Analysis

Two other areas where amperometry has played a major role since the 1970s are the detection of various substances in high-performance liquid chromatography and flow injection analysis. To a lesser extent, amperometric detection has also been applied to ion chromatography of anions and cations.

**See also:** **Enzymes:** Enzyme-Based Electrodes. **Flow Injection Analysis:** Detection Techniques. **Liquid Chromatography:** Principles. **Process Analysis:** Sensors. **Sensors:** Amperometric Oxygen Sensors; Tissue-Based. **Titrimetry:** Overview. **Water Analysis:** Sewage; Biochemical Oxygen Demand; Chemical Oxygen Demand.

### Further Reading

- Adeloju SB, Shaw SJ, and Wallace GG (1994) Polypyrrole-based amperometric biosensor for sulphite determination. *Electroanalysis* 6: 865–870.
- Adeloju SB, Shaw SJ, and Wallace GG (1997) Pulsed amperometric detection of urea in blood samples on a conducting polypyrrole – urease biosensor. *Analytica Chimica Acta* 341: 155–160.
- Bard AJ and Faulkner LR (2001) *Electrochemical Methods: Fundamentals and Applications*. New York: Wiley.
- Harris DC (1999) *Quantitative Chemical Analysis*, 5th edn. New York: W.H. Freeman and Company.
- Janata J (1990) Chemical sensors. *Analytical Chemistry* 62: 33R–44R.
- Janata J and Bezegh A (1988) Chemical sensors. *Analytical Chemistry* 60: 62R–74R.
- Rubinson KA and Rubinson JF (2000) *Contemporary Instrumental Analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Stock JT (1965) *Amperometric Titrations*. New York: Interscience.
- Stock JT (1982) Amperometric, bipotentiometric, and coulometric titration. *Analytical Chemistry* 54: 1R–9R.
- Zhao HJ, Yuan YJ, Adeloju SB, and Wallace GG (2002) Study on the formation of the prussian blue films on polypyrrole surface as a potential mediator system for biosensing applications. *Analytica Chimica Acta* 472: 113–121.

# AMPHETAMINES

J T Cody, San Antonio, TX, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Amphetamines and related compounds have a long and storied history extending literally thousands of years in the past with the use of the drug 'Ma Huang' in China (identified as ephedrine in the late 1800s). Amphetamine is a synthetic drug first produced in 1887 followed several years later by methamphetamine in 1914 (structures given in **Figure 1**). The simple nature of these molecules is deceptive when compared with the numerous and varied activities they have within the body. Their central nervous system (CNS) activity is that of an indirect acting sympathomimetic and it is their CNS stimulatory activity that leads to their potential for abuse. In addition to their CNS activity, amphetamines also have peripheral activity. Their site of action is dependent

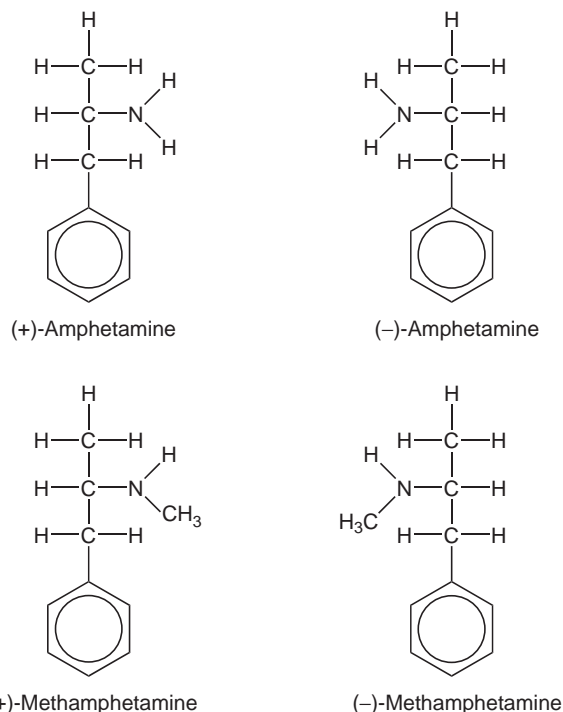
on the enantiomeric form of the drug present with the (*S*)-enantiomer having the greatest CNS activity. Several of the drugs in this class are on the schedule of controlled substances while others are commonly found in over-the-counter medications. Clinically, amphetamine and methamphetamine are used for treatment of narcolepsy, attention deficit disorder with hyperactivity, and as an anorexic. Tolerance and dependence can rapidly develop therefore leading to strict prescribing guidelines.

Following administration of methamphetamine, both methamphetamine and amphetamine are excreted. The excretion profile of these drugs is dependent on both metabolism and excretion; the longer the drugs remain in the body, the more metabolic degradation takes place. The 'average' excretion of unchanged drug is ~30% for amphetamine and 43% for methamphetamine. The actual amount of unchanged amphetamine excreted in urine is greatly affected by pH and can vary from as little as 1% in highly alkaline urine to as much as 74% in strongly acidic urine. Methamphetamine also shows widely differing excretion rates that range from 2% in alkaline urine to 76% in acidic urine. The amount of methamphetamine excreted in the form of amphetamine also varies from 7% in acidic urine to as low as 0.1% in strongly alkaline urine reflecting the metabolism of methamphetamine and amphetamine as long as the drug remains in the body. Methamphetamine is demethylated to amphetamine and subsequently to a number of different compounds prior to excretion. The metabolic pathways of amphetamine and methamphetamine are summarized in **Figure 2**.

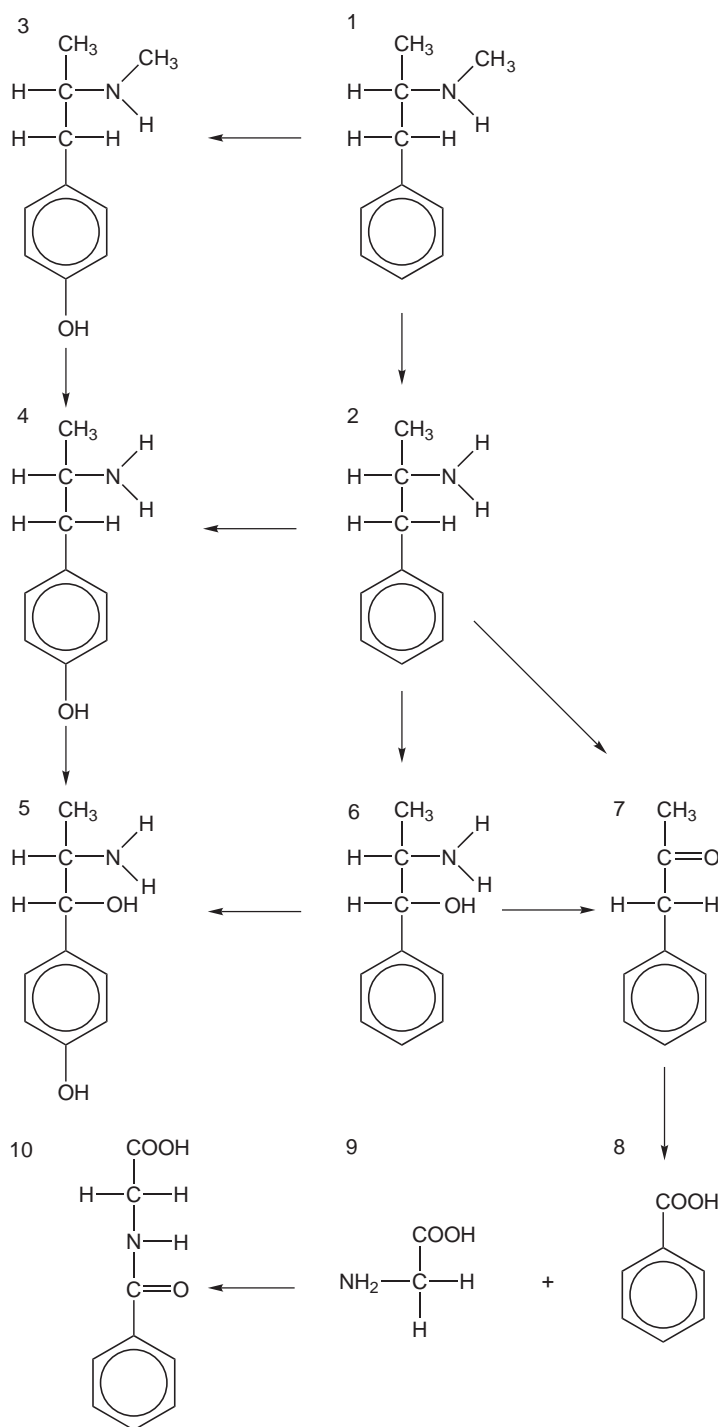
Here, the use of amphetamines, common analytical methods used in their analysis, their analysis in various biological fluids, and the unique aspects related to forensic analysis of amphetamines are reviewed. A glimpse at the future trends in analysis of the amphetamines is also provided.

## Methods

Analysis of amphetamines is commonly accomplished using a number of different methodologies. These include techniques such as thin-layer chromatography (TLC), gas chromatography (GC), liquid chromatography (LC), infrared spectroscopy, and mass spectrometry (MS). In addition, several different immunoassays are commonly used for the analysis of amphetamines.



**Figure 1** Structures of stereoisomer of amphetamine and methamphetamine. Note: The references to the stereoisomer of amphetamine and methamphetamine are often designated using various different indications of the orientation. Common designations for the dextro form of the drugs include *d*, *d*, *D*, *S*, and (+) and similarly for the levo form the designations *l*, *l*, *L*, *R*, and (–) are used.



**Figure 2** Structures of major metabolites of amphetamine and methamphetamine. 1, Methamphetamine; 2, amphetamine; 3, hydroxymethamphetamine; 4, hydroxyamphetamine; 5, hydroxynorephedrine; 6, norephedrine; 7, phenylacetone; 8, benzoic acid; 9, glycine; 10, hippuric acid. Note: stereochemistry and ionic state are not indicated in these structures and conjugates, other than hippuric acid, are not shown.

All immunoassays are based on the principle of competitive displacement of a labeled drug from an antigen–antibody complex by unlabeled drug in the sample. The fundamental difference in the currently available immunoassays is the

detection method employed. Commonly used methods include: fluorescence polarization, enzyme immunoassay, cloned enzyme donor immunoassay, enzyme-linked immunosorbent assay, and radioimmunoassay.



In all of the immunoassay methods, the antibodies are targeted to the amphetamines of interest and designed to minimize cross-reactivity with related compounds. While seemingly simple, achieving specificity is a significant challenge. Since there are so many compounds found in biological fluids that are structurally similar to the amphetamines, specificity is difficult to attain. Tremendous advances have been made over the past several years in this area, however, and the newer assay systems are far better at properly identifying the presence of amphetamine and methamphetamine while significantly diminishing cross-reactivity to related compounds. An advantage of immunoassays over other methods is they are rapid and require no sample preparation. A sample can be added directly to the reagents and the presence of the drugs determined in a relatively short time. Immunoassays are the most popular screening method because they lend themselves to automation thus allowing laboratories to process large numbers of samples. Their biggest disadvantage, however, is lack of specificity. Quantification based on immunoassay is possible, but owing to differing binding affinities to substances, including metabolites, the accuracy of this approach is generally limited. The most common application of immunoassays is as a screening tool for the identification of presumptive positives that are then confirmed by another more specific confirmation assay.

TLC is more specific in the identification of amphetamines but is a far more time-consuming procedure than are immunoassays. With few exceptions, in order to analyze a sample by methods other than immunoassay, the drugs are extracted from the biological matrix prior to analysis (see later). In the case of TLC, the development (chromatographing the drugs up the plate) is a relatively slow process that adds to the time required for analysis. TLC does have the advantage of being able to identify a wide variety of compounds in a sample and is not limited to testing for a single compound or class of compounds as are the immunoassays. Because of the structural similarities, separation of the amphetamines from the many different closely related amines is often not possible with a single solvent system. Typically, using a number of different solvent systems does allow for separation and identification of the individual compounds. Another disadvantage of TLC is it does not have the sensitivity of the other methods. Quantification using TLC is possible only with substantial effort; therefore, it is typically used only as a qualitative tool although some high-performance thin-layer chromatography methods have been shown to be quite specific and precise.

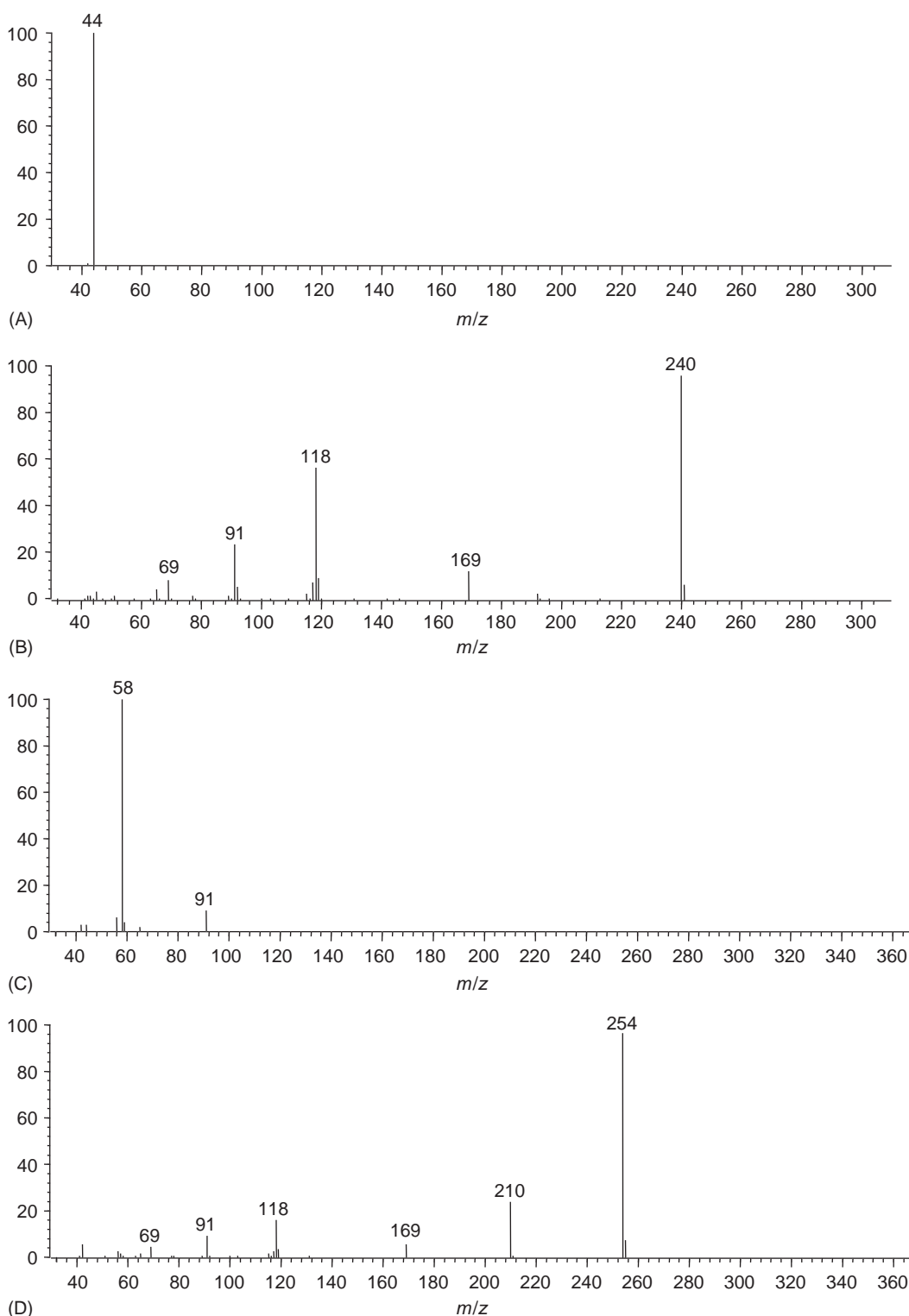
LC is a powerful analytical tool and has been used for the identification of amphetamines in virtually all biological samples. The resolving power of LC allows for specific identification of the drugs and often requires less sample preparation than other chromatographic methods. The natural ultraviolet absorbance of the amphetamines allows for identification by LC, and if derivatized, fluorescent and electrochemical detectors yield even more sensitive and specific detection of these drugs. Derivatization of the amphetamines also enhances their chromatographic behavior. Use of appropriate chiral stationary phases or chiral derivatizing reagents also allows chromatographic separation of the enantiomeric forms of amphetamine and methamphetamine.

Typically, GC has even greater sensitivity than LC and is used with a number of different detectors including flame ionization (FID), nitrogen-phosphorus (NPD), electron capture (EC), and MS. The FID is essentially a universal detector because virtually any flammable compound generates a response.

The NPD, as the name implies, is sensitive for compounds containing nitrogen. It therefore works well with amphetamines because of the presence of the amine group. This detector is also a selective technique owing to the specificity of detecting only those compounds containing nitrogen or phosphorus, thus eliminating many potentially interfering compounds in the sample extract. Electron capture is also a very sensitive detection technique that has been effectively used for amphetamines. However, unlike the NPD, EC detection requires derivatization of the amphetamine with a strongly electronegative group such as a perfluoroacyl group prior to analysis.

Mass spectrometers are commonly used in combination with a gas chromatograph. Most GC-MS analysis of amphetamines is accomplished in the electron ionization mode. This is typically because of the wider availability and ease of use of this ionization method. Chemical ionization is also used for analysis of amphetamines, taking advantage of the relatively high ionization efficiency of the amphetamines when using a reagent gas such as ammonia. A disadvantage of using chemical ionization, however, is the higher degree of complexity in using the instrument and the lessened degree of characteristic fragmentation when compared with electron ionization of derivatized amphetamines.

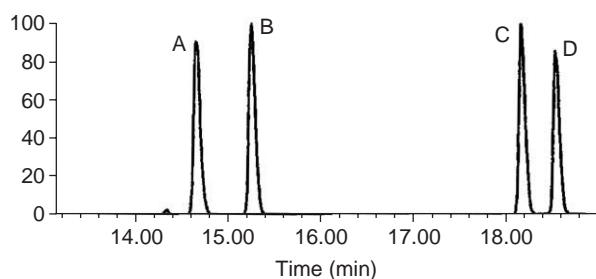
Derivatization of the amphetamines is not required for GC-MS analysis. However, because the mass spectra of underivatized amphetamines are very simple and essentially composed of ions at  $m/z$  44 for amphetamine and  $m/z$  58 for methamphetamine, derivatization is used to not only yield higher mass ions but also to provide multiple characteristic ions



**Figure 3** Mass spectra of (A) amphetamine, (B) heptafluorobutyryl derivative of amphetamine, (C) methamphetamine, (D) heptafluorobutyryl derivative of methamphetamine.

for unequivocal identification of the analyte. The spectra of underivatized amphetamine and methamphetamine and the spectra of the corresponding heptafluorobutyryl derivatives are illustrated in **Figure 3**.

The gas chromatographic behavior of underivatized amphetamines, which tends to show tailing asymmetrical peaks, is also dramatically improved by derivatization. As is true for LC, appropriate



**Figure 4** Gas chromatography of trifluoroacetyl-(L)-prolyl derivatives of amphetamine and methamphetamine enantiomers. Peaks: A, (*R*)-amphetamine; B, (*S*)-amphetamine; C, (*R*)-methamphetamine; D, (*S*)-methamphetamine. GC conditions: 120°C for 2 min, rising to 200°C at 4°C min<sup>-1</sup>. Ions monitored at *m/z* 237 for amphetamine and *m/z* 251 for methamphetamine.

chiral stationary phases or chiral derivatizing reagents allow chromatographic separation of the enantiomeric forms of amphetamine and methamphetamine by GC. The most common of these methods involves use of trifluoroacetyl-L-prolyl chloride (TPC). On standard achiral GC columns, the enantiomeric forms of amphetamine and methamphetamine co-chromatograph unless derivatized with a chiral reagent such as TPC. When derivatized with a chiral reagent, they can then be readily separated. The gas chromatographic separation of the enantiomers of amphetamine and methamphetamine is shown in Figure 4.

## Sample Preparation

In most analytical procedures, other than immunoassays and a few LC procedures, the amphetamines must be extracted from their biological matrix. There are numerous methods for the extraction of amphetamines from biological samples; the most common include liquid-liquid and solid phase. The *pK* of the amine group on these drugs is ~9.9, thus making the sample alkaline significantly enhances their extraction into organic solvents. Simply by adjusting the sample pH to greater than the *pK*, the amphetamines can be efficiently extracted into a small amount of organic solvent. To provide a cleaner extract, samples are often back-extracted from the organic solvent into an acid solution, leaving the purely lipophilic compounds behind. After adjusting the pH again to a high level, the amphetamines are re-extracted into the organic solvent, yielding a much cleaner extract than is achieved by the single-step extraction.

Solid-phase sorbent materials used for the extraction of amphetamines allow for high recovery and selectivity. Binding the amphetamines to the column packing material allows for washing away of the

unwanted components in the sample. Several column sorbents can be used for extraction including those based simply on hydrophobic interactions such as C<sub>18</sub>. These use the hydrophobic interactions between the drugs and the sorbent to extract the compounds of interest. In the case of the amphetamines, this requires adjustment of the pH to a level that can cause problems with silica-based columns. Another method is to use ion exchange where the sample pH is slightly acidic (commonly 6.0) leaving the amine group positively charged. The ionic interaction between the analyte and the column is quite strong allowing the use of strong wash solvents to eliminate potentially interfering compounds. The drugs are then eluted by increasing the pH causing the loss of charge and therefore strong interaction with the column. Due to their versatility, the most popular columns are actually bifunctional in that they contain both ionic and hydrophobic binding capabilities. Such columns can therefore be used to eliminate more potentially interfering compounds in the sample than liquid-liquid extraction or single-mode solid-phase extraction and generally have the potential to produce a cleaner extract.

Protein precipitation is common prior to extraction of drugs from blood samples. Whole blood can also cause problems with solid-phase extraction by plugging the column, although these difficulties can be overcome by precipitation or dilution. Although both liquid-liquid and solid-phase extractions give excellent recoveries of amphetamine and methamphetamine, other closely related compounds are often extracted under the same conditions. In some cases, the initial concentrations of these compounds are so high that, even if their recovery is relatively low, they can pose significant problems in the analysis. One method of eliminating this interference is to destroy the unwanted compounds using an oxidizing agent such as periodate. Hydroxylated compounds such as ephedrine, pseudoephedrine, and phenylpropanolamine are oxidatively cleaved by periodate oxidation, thus they do not interfere with the assay.

## Biological Samples

Urine is the most commonly used biological fluid for the analysis of amphetamines. Blood is also often used for this purpose and the choice of specimen is, to a large extent, dictated by the purpose of the analysis. Other biological samples used in the post-mortem setting include vitreous humor and various tissue homogenates. When testing workplace or sports testing samples, urine holds a variety of advantages over blood. The most obvious is that urine collection is physically less intrusive than

collecting blood samples. In addition, detection of amphetamine use is easier in urine for a number of reasons, including higher concentration of the drugs and extended detection times compared with those in blood. As with all urine drug testing, however, the measured concentration of the drug in urine does not correlate well with the activity of the drug on the body at the time of collection. While interpretation of urine drug testing data is limited, it is further hampered in the case of amphetamines by the fact that the excretion rate is greatly affected by the urinary pH. The differences in the rates of excretion can yield half-lives from 7 to 34 h for amphetamine and from 12 to 34 h for methamphetamine. An advantage of measuring drug concentrations in blood is that it allows the determination of the amount of the drug available to various tissues. However, the interpretation of the total dose of a drug or time since dose is complicated by the fact that blood levels are also significantly influenced by the variable excretion rates of the drug. Since blood concentration represents the amount of drug available in the body at the time of collection, the level can be helpful in determining the influence, if any, the drug had on the individual at that time. From a clinical viewpoint, these questions are often less important than in a forensic investigation.

The use of hair as a sample for detection of drug has been used for a number of years. Hair has the advantage of serving as a record of past drug use. Since drugs deposited in hair are placed there shortly following use, as the hair grows, it represents a sequential history of use over time. Deposition of drugs into hair has been investigated and several important studies have shown the specific drugs and drug metabolites incorporated into hair after use. Issues of differences in hair composition and external contamination have to be considered when analyzing hair and extensive protocols have been developed to ensure the drug detected was actually incorporated into the hair rather than being contamination by environmental contact. A significant challenge with analysis of drugs in hair samples is the low levels found compared to those seen in other biological samples. Modern instrumental methods, particularly MS-MS, have allowed hair analysis to become a more commonly used technique for the detection of drug use. Sweat has also been used as a sample for the detection of drug use. Like hair, sweat can be used to detect use for a longer period of time than the other sample types. Unlike the other samples, however, sweat collection is more time consuming. An absorbent patch is placed on the subject and remains in place for a period of time (commonly a week), then the patch is removed and analyzed for

the presence of drugs. Evaluation of contamination has been conducted and the technique has been shown to identify drug use during the time the patch was in place. Another sample matrix gaining in popularity is oral fluid. Drugs are excreted into the saliva which can be relatively quickly collected and tested. Oral fluid is collected either as an expectorated fluid which is collected into a small sample container or, more commonly, using a device with an absorbent pad that collects the fluid. The pad is then placed in a buffer, allowed to equilibrate, and is then tested for the presence of drugs. This procedure holds some advantages in that the sample collection is less invasive than collection of blood and less intrusive than collection of urine. It also avoids many of the potential problems of sample substitution and adulteration possible, particularly with unobserved collection of urine samples. Several studies have evaluated the extent of drug found in oral fluid following use, but more work needs to be accomplished to fully characterize the details.

## Forensic Samples

Forensic analysis of samples poses challenges not encountered with analysis for research or clinical purposes. Some of the challenges are related to the forensic nature of the analysis itself and sometimes to the fact the sample itself may be far different from those seen in typical analysis (e.g., clotted blood in postmortem analysis or urine that is several days old rather than fresh). Almost any sample can be 'forensic' since the forensic nature of a sample comes from the use to which the analytical result is put rather than from the actual nature of the sample itself. As a result, virtually any biological fluid or tissue may be a forensic sample. One critical component of forensic analysis is the need to document the origin and handling of the sample from collection until analysis is completed, maintaining a clear, well documented 'chain-of-custody'. Urine is the most commonly analyzed sample for amphetamines and the results of that testing are often used as evidence in legal proceedings. Blood is also a good sample for detecting the presence of amphetamines. Since a considerable amount is known about amphetamines in terms of their pharmacology and pharmacokinetics, blood is a more valuable tool in the determination of influence of the amphetamines on an individual. Ante-mortem blood may be analyzed to determine whether or not amphetamines were used and can be important in determining the pharmacological influence of the amphetamines on the individual. This information can be helpful in determining whether the individual was intoxicated or under the influence

of the drugs at the time the sample was collected. Postmortem analysis of blood is also useful in the determination of amphetamines and can be of significant value in interpreting the cause or potential contribution of the drugs to the death. Of particular interest in postmortem analysis is evaluation of the sample for other compounds that may interact with amphetamines to produce a potentially lethal combination where neither compound by itself would have caused the effect. Another complication in postmortem analysis is redistribution of drugs after death which complicates interpretation.

Interpretation of analytical results is a critical part of forensic analyses. Owing to the complex nature of the pharmacokinetics of the amphetamines, interpretation can often be quite difficult. A critical issue in forensic analysis is to separate legitimate use of drugs from illicit use. Another issue, particularly in postmortem analysis, is showing the effect of the drug on the body even if used with valid medical prescription. Several factors make the interpretation of amphetamine results problematic. One of these is the formulation of the Vicks Inhaler in the USA which contains 50 mg of (*R*)-methamphetamine (levmetphetamine). Although this formulation is not found in all parts of the world, it is readily available throughout the USA. This over-the-counter medication is commonly used to treat the nasal congestion associated with colds and allergies due to its peripheral vasoconstrictive effects. Methamphetamine used in prescription medications is (*S*)-methamphetamine and illicit methamphetamine is almost always (*S*)-methamphetamine or racemic (*R,S*)-methamphetamine. In commonly used analytical methods, enantiomers cannot be differentiated from one another. Therefore, use of controlled methamphetamine or the over-the-counter Vicks Inhaler would lead to the same result. However, the use of LC or GC with either chiral columns or a chiral derivatizing reagent will allow the enantiomers to be separated and indicate the difference between use of a Vicks Inhaler and a controlled form of the drug.

Another complicating factor for interpretation of positive test results is the fact that there are a number of other drugs that are metabolized by the body to either amphetamine or methamphetamine. This situation is difficult to determine in many cases because the parent drug is either not excreted or is excreted for only a short period of time in detectable amounts. Drugs that are metabolized to methamphetamine and/or amphetamine include: amphetaminil, benzphetamine, clobenzorex, deprenyl, dimethylamphetamine, ethylamphetamine, famprofazone, fencamine, fenethylline, fenproporex, furfenorex, mefenorex, mesocarb, and prenylamine.

## Emerging Techniques in Amphetamine Analysis

In recent years, a novel extraction technique has been developed and several investigators have implemented this procedure in the analysis of amphetamines. Solid-phase microextraction (SPME) exploits the long-used principle of volatilizing analytes of interest into the headspace over a liquid sample. Headspace analysis has been employed for many years for a variety of volatile compounds, including the amphetamines. The uniqueness of SPME (and a related technique; solid-phase dynamic extraction) is that it allows concentration of the analyte(s) onto the sorbent needle. Following absorption of the analyte(s), the fiber is placed into the injection port of the GC and the compounds are thermally desorbed and analyzed. Modifications of this fundamental technique are to extend the fiber into the liquid itself thus allowing analysis of compounds that are not volatile and can therefore not be collected from the headspace. Several techniques have also been developed to derivatize the extracted analytes. A significant advantage of SPME is the fact that a sample need only be placed into the instrument and no other intervention is required to analyze samples. Currently, the most significant drawback to widespread use of SPME is the time required to analyze each sample.

Analyses of amphetamines, like most other drugs, are most commonly accomplished using GC-MS, which remains the gold standard for confirmation analysis. Concern about the (*R*)-enantiomer has also led to chiral separation techniques being used for definitive analysis of methamphetamine to establish whether or not the drug was an over-the-counter form or a controlled form of the drug.

The coupling of GC-MS with other analytical techniques is being used to assist in the unequivocal identification of amphetamines and various related compounds. One example of this is the use of Fourier transform infrared spectroscopy. The use of multi-stage mass analysis has arrived in the form of mass spectrometers that are capable of doing mass analysis on a compound followed by further fragmentation and analysis in a second mass analysis. Coupling this capability with GC (or LC) has dramatically increased the specificity of analysis and affords greater sensitivity due to the decreased potential for interference and lower background signal. The dramatic advantage of specificity gained by MS over other less specific techniques is dramatically enhanced by tandem mass spectrometry (MS/MS). MS/MS analysis is generally referred to as being accomplished in time or space. MS/MS in space is used to describe MS/MS in the linear triple quadrupole analyzer.

Ionization in the source is accomplished and the first quadrupole is set to isolate a single precursor ion. This ion is then fragmented in the second quadrupole that is filled with a collision gas that causes the ion to fragment. The product ions thus formed can then be scanned or selectively monitored in the third quadrupole.

MS/MS in time is used to describe MS/MS in the ion trap. Basically, MS/MS analysis in the ion trap is accomplished by external ionization of the analytes with the ions then being introduced into the trap. The ion to be used as the precursor is 'trapped' while all others are ejected from the trap. Energy is then added to cause fragmentation of the trapped ion and the resulting product ions are scanned out of the trap and detected.

As described earlier, high-performance liquid chromatography (HPLC) has been used for many years for the analysis of a wide variety of analytes. Numerous different detectors have been used with HPLC. Coupling MS with HPLC dramatically expanded following the development of atmospheric pressure ionization techniques. This was quickly followed by development of benchtop instruments making them less expensive and more widely available to laboratories. The ability to use less extensive extraction procedures and no need for derivatization offers significant advantage to this technique. Other advantages of HPLC over GC including the ability to analyze compounds with low volatility, poor thermal stability, and allowing analysis of conjugates of drugs and metabolites remain with the analytical advantage of MS with the technique.

Capillary electrophoresis has been used for the analysis of a wide variety of analytes including the amphetamines. Practical use of this technique has grown significantly over the last few years. Capillary electrophoresis uses either electrophoretic or electrokinetic separation, or both. This technique allows analysis of analytes from various matrices with relatively little sample preparation and often no requirement for derivatization. Even separation of underivatized enantiomers has been accomplished with capillary electrophoresis. Capillary zone electrophoresis and micellar electrokinetic capillary chromatography are the two most commonly used methods of capillary electrophoresis. Studies evaluating parameters such as pH, stirring, temperature, addition of salts, and selection of sorbents has lead to significant improvements. Another advantage of capillary electrophoresis is the low sample volume requirements (commonly less than 0.1 ml). Disadvantages of this technique include the difficulty of combining with MS and automation allowing routine analysis of multiple samples in a reasonable

time. Advances in nanospray LC interfaces and sample stacking techniques together with improved automation may bring capillary electrophoresis into the mainstream as an analytical tool.

**See also:** **Derivatization of Analytes.** **Extraction:** Solvent Extraction Principles. **Forensic Sciences:** Drug Screening in Sport; Systematic Drug Identification. **Gas Chromatography:** Mass Spectrometry; Chiral Separations. **Immunoassays, Techniques:** Enzyme Immunoassays; Luminescence Immunoassays. **Liquid Chromatography:** Column Technology. **Pharmaceutical Analysis:** Sample Preparation.

## Further Reading

- Caldwell J (ed.) (1980) *Amphetamines and Related Stimulants: Chemical, Biological, Clinical and Sociological Aspects*. Boca Raton, FL: CRC Press.
- Al-Dirbashi O, Kuroda N, Inuduka S, Menichini F, and Nakashima K (1999) HPLC with fluorescence detection of methamphetamine and amphetamine in segmentally analyzed human hair. *Analyst* 124(4): 493–497.
- Bogusz MJ, Kruger KD, and Maier RD (2000) Analysis of underivatized amphetamines and related phenethylamines with high-performance liquid chromatography-atmospheric pressure chemical ionization mass spectrometry. *Journal of Analytical Toxicology* 24(2): 77–84.
- Cody JT (2002) Precursor medications as a source of methamphetamine and/or amphetamine positive drug testing results. *Journal of Occupational and Environmental Medicine* 44(5): 435–450.
- Geiser L, Cherkaoui S, and Veuthey JL (2000) Simultaneous analysis of some amphetamine derivatives in urine by nonaqueous capillary electrophoresis coupled to electrospray ionization mass spectrometry. *Journal of Chromatography A* 895(1–2): 111–121.
- Jurado C, Gimenez MP, Soriano T, Menendez M, and Repetto M (2000) Rapid analysis of amphetamine, methamphetamine, MDA, and MDMA in urine using solid-phase microextraction, direct on-fiber derivatization, and analysis by GC-MS. *Journal of Analytical Toxicology* 24(1): 11–16.
- Kintz P and Samyn N (2000) Unconventional samples and alternative matrices. In: Bogusz MJ (ed.) *Handbook of Analytical Separations – Forensic Science*, 1st edn., pp. 459–488. Amsterdam: Elsevier Science.
- Nakahara Y and Kikura R (1996) Hair analysis for drugs of abuse. XIII. Effect of structural factors on incorporation of drugs into hair: The incorporation rates of amphetamine analogs. *Archives of Toxicology* 70(12): 841–849.
- Peters FT, Schaefer S, Staack RF, Kraemer T, and Maurer HH (2003) Screening for and validated quantification of amphetamines and of amphetamine- and piperazine-derived designer drugs in human blood plasma by gas chromatography-mass spectrometry. *Journal of Mass Spectrometry* 38(6): 659–676.

- Samyn N, De Boeck G, and Verstraete AG (2002) The use of oral fluid and sweat wipes for the detection of drugs of abuse in drivers. *Journal of Forensic Science* 47(6): 1380–1387.
- Uhl M (2000) Tandem mass spectrometry: A helpful tool in hair analysis for the forensic expert. *Forensic Science International* 107(1–3): 169–179.
- Van Bocxlaer JF, Clauwaert KM, Lambert WE, *et al.* (2000) Liquid chromatography-mass spectrometry in forensic toxicology. *Mass Spectrometry Review* 19(4): 165–214.
- Yonamine M, Tawil N, Moreau RL, and Alves Silva O (2003) Solid-phase micro-extraction–gas chromatography–mass spectrometry and headspace–gas chromatography of tetrahydrocannabinol, amphetamine, methamphetamine, cocaine and ethanol in saliva samples. *Journal of Chromatography B* 789(1): 73–78.
- Zambonin CG (2003) Coupling solid-phase microextraction to liquid chromatography. A review. *Analytical and Bioanalytical Chemistry* 375: 73–80.

## AMPLIFICATION REACTIONS

**J M Wages Jr**, Palmetto Consulting and Research, Tupelo, MS, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by M. de la Guardia Cirugeda, pp. 100–108, © 1995, Elsevier Ltd.

### Introduction

The term amplification usually refers to a process whereby the power of a signal is increased without altering its basic information-carrying characteristics. The signal may be electronic, hydrodynamic, acoustic, or, in our case, chemical.

The Commission on Analytical Reactions and Reagents of the International Union of Pure and Applied Chemistry (IUPAC) defines the term amplification as follows: ‘An amplification reaction is one which replaces the conventional reaction used in a particular determination so that a more favourable measurement can be made. The sequence can be repeated to provide a further favourable increase in measurement’. This definition avoids the ambiguity of the terms multiplication and magnification, previously employed, and it has definitively introduced the term amplification in the analytical literature.

A number of changes are necessary to adapt this definition to the present state of analytical methods, not only in inorganic and organic analysis but also in biochemical analysis. However, two basic ideas must be retained: the fact that the use of amplification reactions is a way of increasing the sensitivity of the analytical measurements, and that these reactions can be carried out in an iterative form.

Historically, as has been indicated by Friedrich Mohr, the first amplification method was introduced by Bunsen and Dupré and independently by Golfier-Besseyre for the determination of iodide by oxidation to iodate. After elimination of the excess of oxidant, the reaction of iodate with iodide provided a sixfold enhancement of the original amount of iodide as iodine.

This method was also employed by Winkler in 1900 and Hunter in 1909, the latter using sodium hypochlorite as oxidant.

The application of these procedures for the determination of iodine in organic compounds on the microscale, carried out in 1929 by Leipert, promoted the popularization of amplification reactions in analytical laboratory practice.

In the literature published during the last 25 years, several examples of amplification reactions can be found that enhance the analytical sensitivity obtainable in the direct determination of cations and anions, and also in the elemental analysis and the selective determination of organic molecules.

This article reviews the principles of amplification reactions and their use for analytical purposes, with particular reference to the mechanisms involved in the methods discussed, in order to systematize present knowledge about these reactions. On the other hand, new aspects, concerning enzymatic amplification reactions and the polymerase chain reaction, have been included to establish the similarities and differences between these reactions and the classical amplification procedures.

### Principles and Classification

According to the definition of amplification, the increase in the number of moles of the compound to be measured by an amplification reaction must be different from that obtained normally by a conventional stoichiometric reaction. Thus, the simple weighting effect obtained in the gravimetric analysis of cations with complex organic molecules cannot be considered amplification reactions. In amplification reactions, the mass of the compound is increased by a selected series of reactions.

There are two basic forms of amplification: (1) direct amplification, where the compound to be determined is amplified directly and is subsequently measured; (2) indirect amplification, where the

compound is associated with some other, which is then amplified and measured.

Both direct and indirect amplification processes can be implemented by various mechanisms:

1. Change of oxidation state. The amplification is based on the change of valency or oxidation state of the compound to be measured (direct) or that associated with a preceding one (indirect).
2. Amplification by alternative precipitation. The precipitation of two salts of the same compound, with different stoichiometries, is employed as a basis for cyclic amplification of an anion or a cation.
3. Amplification with a complex of the ion to be amplified. This combines precipitation and valency-change processes.
4. Transportation. The compound to be amplified is employed for the transport of an equivalent amount of a reagent from one place to another. Amplification by the use of ion exchangers is a particular case of this general type of amplification process.
5. Alternate addition of equivalent amounts of two reagents. This involves a cyclic process in which a bare excess of reactant is alternately added to the specimens to be determined and the stoichiometric precipitates of the starting concentration collected after each cycle.

The two last cases are actually examples of iterative amplifications but *sensu stricto* do not suppose the use of any alternative reaction.

The parameter that defines the amplification procedures is the amplification factor,  $f$ , the quotient between the amplified amount of a compound that is measured,  $y$ , and the starting amount,  $x$ , to be amplified (eqn [1]):

$$f = y/x \quad [1]$$

Each amplification reaction provides a given factor  $f$  as a function of the reactions carried out. However, all the above-mentioned procedures can be applied several times before the measurement step, providing a cyclic amplification whenever the product of an amplification reaction can be transformed by whatever series of reactions into the original amplifiable compound.

Cyclic amplification complicates the experimental procedures but increases extraordinarily the amplified number of moles obtained.

Cyclic processes can be classified as arithmetic and geometric, depending on the relation between the amplified number of moles and the amplification factor obtained in each cycle. When at each single cycle the same amount of reagent is obtained, this can be termed arithmetic amplification. Thus, the following

relationship holds between the amplified number of moles,  $y$ , and the starting number,  $x$ , as a function of amplification factor  $f$  and the number of cycles,  $n$ :

$$y = fnx \quad [2]$$

Weisz has used the term multiplication to describe this kind of cyclic amplification, but the term arithmetic cyclic amplification is preferable to avoid misinterpretation.

In other cases, the amount of substance in circulation does not remain constant in each step but increases following a geometric progression and the amplified number of moles increases exponentially with the number of cycles as follows:

$$y = f^n x \quad [3]$$

Emich has suggested the term exponential method for these cases, but it might equally well be called cyclic geometric amplification.

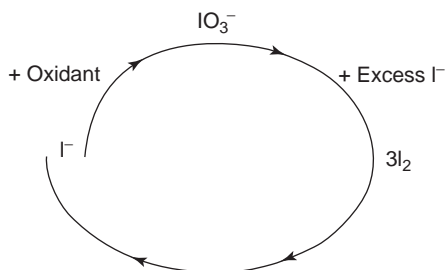
In the following sections, several applications of each type of amplification mechanism are discussed.

## Applications

### Change of Oxidation State

The determination of iodide, after preceding oxidation to  $\text{IO}_3^-$  and reaction of the iodate formed with an excess of  $\text{I}^-$ , is an example of direct valency exchange process. **Scheme 1** summarizes the experimental procedure involved.

This method, very common in the analytical literature, gives an amplification factor of 6 and is the basis of a large number of amplification reactions. It can be applied using different oxidants, such as chlorine or bromine water, or sodium hypochlorite. A method has been proposed for the simultaneous determination of iodide and bromide ions, and also for the determination of iodine and bromine in organic compounds. The method is based on the oxidation of both ions to their halates and titration of the iodate and bromate after reaction with iodide for



**Scheme 1**



iodate at pH 3.8–4.1 and at a lower pH for the bromate–iodide reaction.

On the other hand, when  $I^-$  is oxidized by  $IO_4^-$ , a factor of 24 can be achieved as shown in Scheme 2.

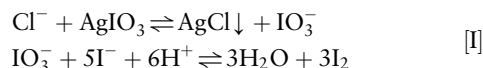
Another example of direct valency exchange amplification is the method suggested by Emich in 1933 and developed by Schöniger in 1966 for the amplification of microgram amounts of  $CO_2$  by a cyclic geometric amplification procedure (Scheme 3).

The consecutive use of reductions with Pt/C and oxidations on CuO can also provide water amplification by the series of reactions shown in Scheme 4, which is one example of an indirect amplification procedure.

Reactions based on the equilibrium of the iodide/iodate system have been used extensively for indirect amplification, and a large number of mechanisms can be differentiated, such as metathesis of anions with  $IO_3^-$  or  $I^-$ , precipitation of cations with  $I^-$ , reduction with

$I^-$ , oxidation with  $I_2$ , oxidation with  $IO_4^-$ , oxidation of thiocyanate–metal complexes, bromination and reduction with  $I^-$ ,  $S_N2$  substitution and oxidation with  $Br_2$ , and complexation of cations with ethylenediaminetetraacetic acid (EDTA) in the presence of  $Pb(IO_3)_2$ .

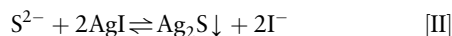
The reaction of anions with  $AgIO_3$  is a typical example of ‘indirect amplification by metathesis’ and, as can be seen in reaction [I], it gives sixfold amplification:



As summarized in Table 1, anions other than  $Cl^-$  can be amplified in the same way.

Hassan and Thoria have developed an amplification procedure for the determination of organophosphorus compounds (triphenyl phosphine, tritolylphosphate, diethylbenzyl phosphates, tri-n-butylphosphine oxide, and triphenyl phosphate). The method is based on combustion of the organic compound in an oxygen-filled flask and the formation of the secondary form phosphate by adjustment of the pH value of the combustion product. Calcium iodate is then added and the  $IO_3^-$  released is treated with  $I^-$  to obtain an amplification factor of 12.

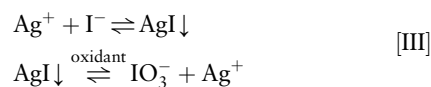
One example of indirect amplification by metathesis of anions with  $I^-$  is the determination of  $S^{2-}$  after reaction with  $AgI$  (reaction [II]):



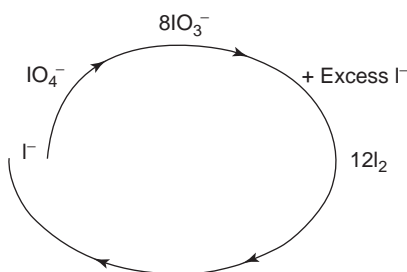
$I^-$  is then amplified by oxidation to  $IO_3^-$  and in this case an amplification factor of 12 is obtained.

Precipitation of cations with iodide and further oxidation of the precipitates to obtain  $IO_3^-$  involves an indirect method of amplification by valency exchange.

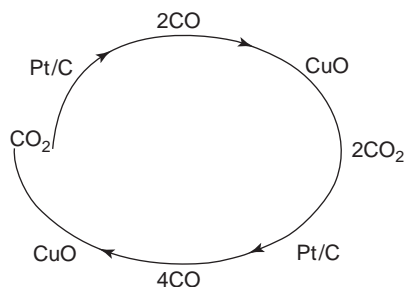
$Ag^+$  can be amplified six times by the reaction [III]:



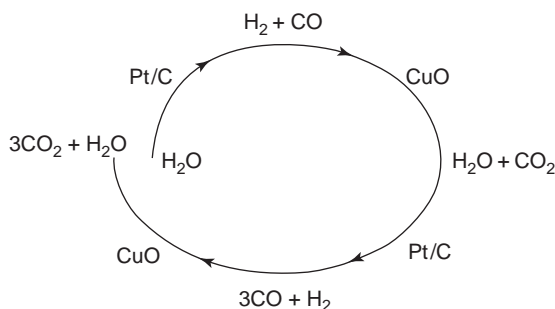
$IO_3^-$  is then amplified according to Scheme 1.



Scheme 2



Scheme 3



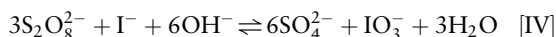
Scheme 4

**Table 1** Applications of indirect amplification by metathesis of anions

Anion	Compound added	Amplification factor, <i>f</i>
SCN <sup>-</sup>	AgIO <sub>3</sub>	6
CN <sup>-</sup>	AgIO <sub>3</sub>	6
SO <sub>4</sub> <sup>2-</sup>	Ba(IO <sub>3</sub> ) <sub>2</sub>	12
F <sup>-</sup>	Ca(IO <sub>3</sub> ) <sub>2</sub>	6
HPO <sub>4</sub> <sup>2-</sup>	Ca(IO <sub>3</sub> ) <sub>2</sub>	12
PO <sub>4</sub> <sup>3-</sup>	Ca(IO <sub>3</sub> ) <sub>2</sub>	18
AsO <sub>4</sub> <sup>3-</sup>	Ca(IO <sub>3</sub> ) <sub>2</sub>	12
EDTA	Pb(IO <sub>3</sub> ) <sub>2</sub>	12

Other cations, such as mercury, bismuth, lead, and thallium, can be precipitated with iodide and amplified in the same way as silver with an amplification factor of 6. In the case of thallium iodide, 4 mol of  $I_2$  are formed because the change in oxidation state from Tl(III) to Tl(I) produces an additional mole of  $I_2$ .

Reduction of oxidizing compounds with  $I^-$  provides an indirect way for their amplification via **Scheme 1**. A typical example is the polarographic determination of peroxodisulfate (reaction [IV]):

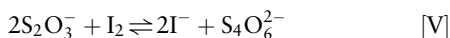


In this case,  $IO_3^-$  is reduced to iodine and iodide and then amplified by **Scheme 1**, providing an amplification factor of 6. Alternatively, the use of periodate provides a 24-fold amplification procedure (see **Scheme 2**).

When the iodate formed, at the end of the amplification process, is reduced at a dropping mercury electrode, a single wave with six times the magnitude of that found for iodide is obtained, and so a 36-fold amplification is achieved. The polarographic reduction of the iodate formed after oxidation of iodide with periodate involves a 144-fold enhancement of the sensitivity.

Other oxidizing compounds, such as quinone, chloramine T, and nitrite, can be amplified in a similar way with an amplification factor of 6, or 36 when a two-step cyclic reaction is carried out.

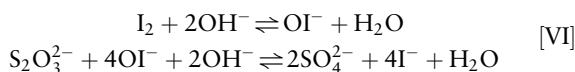
Reducing substances can be indirectly amplified after oxidation with  $I_2$ . An example of this type of reaction is amplification of thiosulfate as follows (reaction [V]):



The  $I^-$  formed is amplified six times by **Scheme 1**.

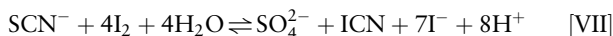
Thiosulfate can be oxidized to sulfate in the presence of hydroxide and the iodide obtained treated with bromine water after neutralization of the excess alkali with  $H_2SO_4$ , providing an amplification factor of 48.

At pH 12, iodide forms hypoiodite, which oxidizes  $S_2O_3^{2-}$  to  $SO_4^{2-}$  (reaction [VI]):

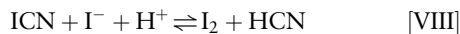


After acidification, the excess of  $OI^-$  is converted to  $I_2$ , which is then extracted. The oxidation of the remaining  $I^-$  with bromine water produces  $IO_3^-$ , which reacts with an excess of  $I^-$  to form 24 mol of iodine.

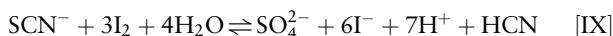
Thiocyanate is oxidized by  $I_2$  in a hydrogen carbonate medium (pH 8.2) to sulfate (reaction [VII]):



Acidification of the solution of pH 2.5 with  $H_2SO_4$  causes the reaction between iodine cyanide and iodide as shown in reaction [VIII]:

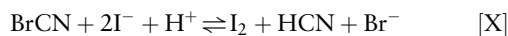


Thus, the overall reaction [IX]:



The oxidation of the remaining  $I^-$  with bromine after removal of iodine with chloroform results in the formation of  $IO_3^-$  from  $I^-$  and the oxidation of HCN to BrCN.

After destruction of the excess bromine with formic acid, the addition of potassium iodide provides 38 mol of iodine, 36 mol from the 6 mol of  $IO_3^-$  and 2 additional moles from the  $I_2$  formed in reaction [X]:



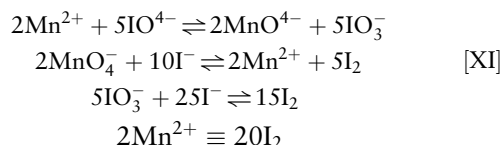
Other examples of indirect amplification after oxidation with  $I_2$  are summarized in **Table 2**. Especially interesting is the method developed by Amin and Al-Allaf for the amplification of the iodometric determination of organolead compounds, using **Scheme 1**, because in this case alkyllead and phenyllead compounds show different behavior. Phenyllead reaction products with  $I_2$  are extracted quantitatively into an organic phase, whereas in the determination of ethyllead compounds the reaction products

**Table 2** Examples of indirect amplification after oxidation with  $I_2$

Compound amplified	Amplification factor, <i>f</i>
Arsenite	14
Isonicotinic acid hydrazide	6
Semicarbazide	6
Thiosemicarbazide	6
1-Acetylthiosemicarbazide	6
1-Phenylthiosemicarbazide	6
4-Phenylthiosemicarbazide	6
Hydroquinone	6
Thioglycolic acid	6
Uric acid	12
Chloral hydrate	48
Glucose	12
Galactose	12
Arabinose	12
Lactose	12
Maltose	12
Organolead compounds	3–12

containing iodine are extracted into an aqueous phase. Thus, both types of compound can be determined in a mixture. Amplification factors from 3 to 12 have been obtained in each case.

Reducing compounds react with periodate to give an equivalent amount of iodate that can be amplified, and so oxidation with  $\text{IO}_4^-$  at pH 7 provides another path for indirect amplification. A typical example is the 1:20 amplification of  $\text{Mn}^{2+}$  by reaction [XI], in which

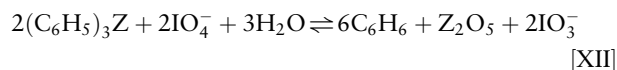


Chromium(III) can be amplified in the same way with  $f=12$ ; antimony(III) with  $f=6$ ; and sulfur dioxide with  $f=6$ .

Other organic compounds, aldoses (such as glucose, mannose, or galactose), and glucose-1-phosphate have been amplified after oxidation with  $\text{IO}_4^-$ .

Besade and Gawargious have carried out the iodometric submicro determination of  $\alpha$ -aminoalcohols with amplification factors of 6, 12, and 18 for primary, secondary, and tertiary amino groups, respectively, owing to the stoichiometry of the oxidation reactions with  $\text{IO}_4^-$ .

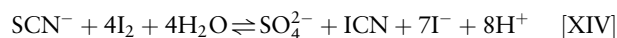
Triphenylphosphine, triphenylarsine, and triphenylstibine can be amplified by oxidation with periodate in acidic or alkaline medium and iodometric titration of the iodate ion released. Two reaction routes (reactions [XII] and [XIII]) can be employed, where  $Z = \text{P}, \text{As}, \text{or Sb}$ :



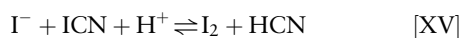
In both cases an actual amplification factor of 3 is achieved.

All the above reactions can be carried out using molybdate at pH 3, as masking agent of the excess of periodate, as first proposed by Burnel in 1965.

One special case of indirect amplification via iodine oxidation is offered by the reaction with thiocyanate in sodium bicarbonate alkaline medium (reaction [XIV]):



One of the seven iodide ions formed reacts, at acid pH, with the iodine cyanide as shown in reaction [XV]:

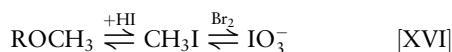


After extraction of iodine, reaction with an excess of bromine provides 6 mol of  $\text{IO}_3^-$  per mole of  $\text{SCN}^-$ , which are amplified in the classical way. On the other hand, bromine oxidizes HCN providing BrCN, which reacts with 2 mol of  $\text{I}^-$  to form an additional mole of iodine, and so an amplification factor of 38 is obtained for  $\text{SCN}^-$ .

The amplification of  $\text{SCN}^-$  can be used as a basis for the amplification of metals by oxidation of thiocyanate metal complexes and several examples of this have been published, such as the amplification of Pb after formation of  $\text{Pb}(\text{SCN})_3[\text{N}(\text{C}_4\text{H}_9)_4]$ , in which  $1\text{Pb} \equiv 114\text{I}$ , or after forming  $\text{Pb}_3[\text{Cr}(\text{SCN})_6]_2$ , in which  $1\text{Pb} = 152\text{I}$ , and the amplification of bismuth via formation of  $\text{Bi}[\text{Cr}(\text{SCN})_6]$  or  $\text{Bi}(\text{SCN})_6[\text{N}(\text{C}_4\text{H}_9)_4]_3$ , in which cases an amplification factor of 228 is obtained.

Bromination of organic compounds and subsequent reaction with iodide provides an indirect amplification based on the extraction of the iodine liberated, reduction to iodide, and iodometric titration of the iodate formed by Scheme 1. By this method phenol is amplified 12 times and resorcinol and phloroglucinol 24 times. Other compounds, such as salicylic acid, acetylsalicylic acid, and *p*-hydroxybenzoic acid, have been also amplified.

Nucleophilic bimolecular substitution of alkoxyl and oxidation of the iodide liberated from the distilled alkyl iodide provides a sixfold indirect amplification procedure as seen in reaction [XVI], where  $\text{IO}_3^- \equiv 3\text{I}_2$ :



As previously described, the reaction of ethylenediaminetetraacetic acid (EDTA) with  $\text{Pb}(\text{IO}_3)_2$  provides an indirect amplification procedure based on the formation of 3 mol of  $\text{I}_2$  per mole of  $\text{IO}_3^-$ .

Complexation of cations with EDTA in the presence of  $\text{Pb}(\text{IO}_3)_2$  provides indirect amplification, which has been applied to the determination of Bi, Fe, In, Hg, Th, Cu, Ni, Pb, Zn, Cd, Co, and Al.

The general procedure consists of the addition to the sample of a known excess of EDTA and some lead iodate. The fraction of EDTA that remains unreacted with the cation liberates an equivalent amount of iodate, which is then amplified by reaction with iodide, after filtration.

In a similar way to that reported for the iodide-iodate system, the bromine-bromate can be used to provide amplification of several species.

Although the indirect determination of phosphorus, as molybdophosphoric acid, is not considered an example of an amplification reaction by several authors, because it is based on a favorable stoichiometry more than on the use of an alternative reaction, Belcher and Uden have proposed an actual amplification

procedure, based on the precipitation of molybdenum with oxine and titration with potassium bromate. In this case, each mole of oxine consumes 4 equivalents of standard bromate solution and, taking into account that  $\text{MoO}_2(\text{C}_9\text{H}_6\text{ON})_2$  is formed between molybdate and oxine, a final amplification factor of 96 for phosphorus can be obtained.

### Alternative Precipitation

The direct amplification of silver can be carried out by precipitation with chromate.  $\text{Ag}_2\text{CrO}_4$  is filtered and then treated with barium chloride for the precipitation of both  $\text{BaCrO}_4$  and  $\text{AgCl}$ . These precipitates are treated with  $\text{Ag}^+$  and thus 1 mol of  $\text{Ag}_2\text{CrO}_4$  and 2 mol of  $\text{AgCl}$  are formed.

The process can be repeated, and on each treatment with  $\text{BaCl}_2$  an amount of silver chloride equivalent to twice the original silver chromate is produced; after  $n$  cycles  $2(n+1)$  moles  $\text{AgCl}$  can be found.

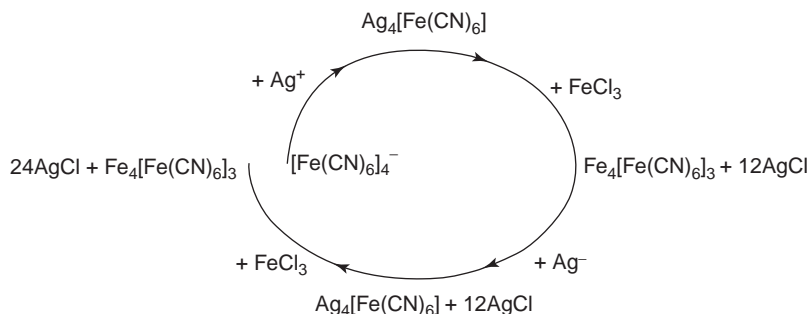
This procedure can also be applied for the indirect determination of  $\text{CrO}_4^{2-}$ , by measuring the  $\text{AgCl}$  obtained.

One other indirect amplification, similar to those mentioned, is the amplification of phosphate by precipitation of silver phosphate and its conversion to silver chromate.

A particular case of the alternative precipitation method, which Weisz called cyclic amplification by fixation of both ions of an amplifiable compound, requires that the substance to be amplified be subjected to a series of stoichiometric reactions in order to enhance its mass.

**Scheme 5** for the amplification of hexacyanoferrate(II), via precipitation of the silver salt and formation of silver chloride and Prussian Blue by reaction with iron chloride, shows that it is a particular case of the alternative precipitation method in which the amplification takes place by a cyclic procedure.

This is an indirect procedure, but silver can also be amplified in a direct way by measuring the final  $\text{AgCl}$  formed.



**Scheme 5**

Weisz has applied this type of amplification reaction to the determination of phosphate, hexacyanoferrate(II), chromate, zinc, iron, and silver, and has obtained amplification factors up to 153 in a series of cyclic processes carried out in an automated form.

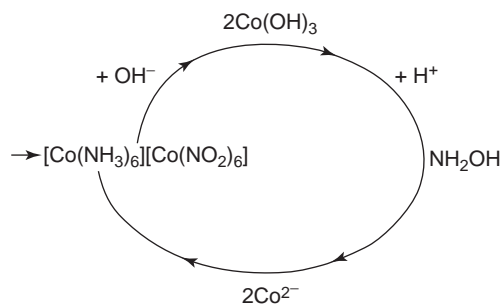
### Amplification with a Complex of the Ion to be Amplified

In this method the ion to be amplified is precipitated with a ligand containing the same ion, and then a cyclic procedure can be carried out. In general, this kind of amplification provides a geometric amplification factor.

Cobalt can be amplified by precipitation of hexaamminecobalt(III)–hexanitritocobaltate(III); the precipitate is then dissolved in  $\text{NaOH}$  and reduced to cobalt(II) with hydroxylamine. The doubled amount of cobalt produced in this step is converted again to hexanitritocobaltate(III) and precipitated with hexaammine cobaltate(III) following the cyclic procedure indicated in **Scheme 6**.

This process provides a twofold amplification in each cycle. The same system can provide a fourfold amplification if cobalt is converted to  $\text{Co}(\text{NH}_3)_6^{3+}$  and then precipitated as  $[\text{Co}(\text{NH}_3)_6][\text{Co}(\text{NH}_3)_2(\text{NO}_2)_4]_3$ .

Other examples of the same type are the amplification of  $\text{Ag}^+$  by reaction with  $(\text{Ag}(\text{CN})_2)^-$  to form



**Scheme 6**

2 mol of AgCN, and the reaction between Hg(II) and  $(\text{HgI}_4)^2$  to form 2 mol of  $\text{HgI}_2$ .

### Amplification by Transportation

Based on the principles of cyclic iteration, Weisz proposed the term transportation to define an amplification method in which an equivalent amount of a reagent is transported, by the compound to be amplified, from one place to another.

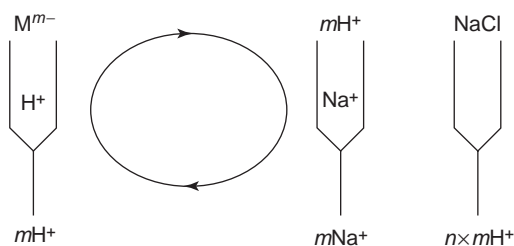
The transportation method involves an arithmetic increase of the amount of the substance to be measured, but it is not based on an alternative reaction and so cannot be considered an example of amplification *sensu stricto*, considering the definition of the term by the IUPAC.

The reaction between micromolar amounts of EDTA and solid mercury iodate is an example of the transportation process based on alternative precipitation. EDTA provides the solubilization of an equivalent amount of iodate and the formation of an Hg(II)-EDTA complex. The filtrate is treated with  $\text{Ag}_2\text{S}$  and the EDTA liberated is added to the original sample and brought back into the cycle.

With the aid of ion-exchange columns the amplification of cations and anions by transportation can be carried out using the following procedure. An aqueous solution containing a cation such as  $\text{Na}^+$ ,  $\text{Ca}^{2+}$ , or  $\text{Mg}^{2+}$  is applied to a cation-exchange column (in the acid form) and there it liberates an equivalent amount of protons. The solution obtained is introduced to another column (in the basic form) and liberates an equivalent amount of sodium ions, which are again passed through the acid form column, and so on. After  $n$  cycles all the protons on the cation exchangers are eluted with NaCl and titrated providing  $n \times m$  times the starting amount of the ion to be amplified,  $m$  being the charge of the ion considered (Scheme 7).

In a similar way, anions such as  $\text{Cl}^-$  and  $\text{Br}^-$  can be amplified using anion-exchange resins in the  $\text{OH}^-$  and  $\text{Cl}^-$  forms.

In all these cases an arithmetic cyclic amplification takes place and, experimentally, up to five cycles have been performed with good results.



Scheme 7

### Amplification by Alternate Addition of Equivalent Amounts of Two Reagents

Weisz has proposed a general methodology that permits cyclic amplification of the mass of a compound to be determined gravimetrically. This method, called alternate addition of equivalent amounts of reagents, is based on the addition, to the compound to be amplified (A), of a slight excess of another compound (B), which forms a stoichiometric product (AB). The reaction product is then collected and an amount of A equivalent to that initially employed of B is added; the product obtained is discarded and the procedure is reinitialized. After  $n$  cycles the mass of the product to be measured has been amplified  $n$  times.

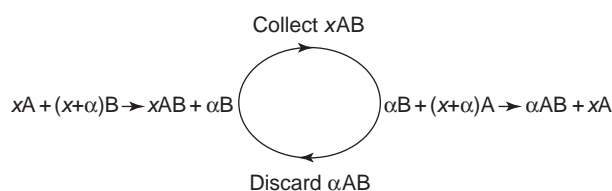
Scheme 8 describes the experimental procedure. The process has been employed for the amplification of silver (using  $\text{Cl}^-$ ), lead (using  $\text{CrO}_4^{2-}$ ), nickel (using dimethylglyoxime), and copper (using benzoin oxime).

This procedure demonstrates that chemical amplification can be obtained without seeking an alternative process, but rather only including a 'touch of imagination' in our current chemical process. It has been included in the present article for that reason. However, neither alternate addition of equivalent amounts of two reagents nor the transportation method is an amplification reaction.

### Enzymatic Amplification Reactions

Catalytic reactions constitute a special type of reaction mechanism in which a chemical amplification takes place. With very low catalyst concentrations a significant amount of reaction product can be measured, enhancing the analytical sensitivity of the measurements. In general, these reactions are not amplification reactions. However, in some enzymatic reactions, a series of enzymes and other co-factors act on each other in a sequential fashion, providing a cascade reaction or, in a cyclic process, providing a rapid, amplified response of the small initial signal, and these processes can be considered examples of amplification reactions.

Metabolic pathways provide interesting examples of cyclic amplification processes and there have been described more than 140 enzymes and other proteins



Scheme 8

whose biological activity can produce a chemical amplification. In general, their activity is controlled by reversible covalent modification and two particular mechanisms can be found: substrate cycles and interconvertible enzyme cycles.

However, the sense of the term amplification in enzymology is not exactly the same as that previously considered for inorganic and organic reactions. In some cases, it is employed to indicate the enhancement of the total rate of a reaction by increase in the activity of one enzyme and decrease in the activity of another that acts on the same substrate, but in an opposite way. On the other hand, the synthesis of an active form of an enzyme from another can cause improvement in the effect of an external factor. The analytical chemist can exploit these situations in order to improve the sensitivity of determinations.

The processes mentioned can be applied in the analytical laboratory; for example, the determination of nicotinamide adenine dinucleotide phosphate (NADP) has been amplified by enzymatic reaction. The process consists of the addition to the sample of a mixture of  $\alpha$ -ketoglutarate, ammonium ion, and glucose-6-phosphate, which act as substrates, and controlled amounts of glutamate dehydrogenase and glucose-6-phosphate dehydrogenase enzymes. It initializes the cyclic reactions (Scheme 9), which provide increasing amounts of glutamate and 6-phosphogluconate. After the cycling sequence, the mixture is treated to destroy the enzymes and the 6-phosphogluconate is measured.

With large amounts of the two enzymes, more than 20 000 cycles per hour can be carried out and an overall amplification of the order of  $10^8$  can be achieved, enabling the assay of an amount as small as  $10^{-19}$  mol of NADP.

This example indicates the large number of amplification processes in biological systems that can be translated to laboratory scale and provide alternative

techniques to solve sensitivity problems in laboratory practice.

## Nucleic Acid Amplification Technologies

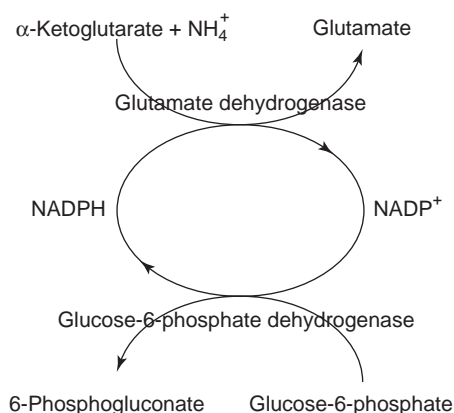
Detection and analysis of DNA and RNA present particular problems for the analytical biochemist because of the very low concentration of analyte. Often only one or two copies of a given sequence are present in the sample. Without either target or signal amplification, detection of such rare sequences is not feasible.

### Polymerase Chain Reaction

Nucleic acid amplification enables detection of nucleic acid sequences using laboratory methods analogous to processes that occur *in vivo*. The first such technology to be described was the polymerase chain reaction (PCR), which uses sequence-specific oligonucleotide primers and a thermostable DNA polymerase to carry out *in vitro* DNA replication. Repeated cycles of heat denaturation, primer annealing or hybridization, and primer extension by the polymerase facilitate synthesis of millions of copies of the original target sequence. DNA amplification by PCR can be coupled with reverse transcription of an RNA template to a complementary DNA (cDNA), thus enabling RNA detection by closely similar methods (reverse transcription PCR). Quantitative PCR methods provide information on gene expression levels. *In situ* PCR enables spatial localization of particular RNA transcripts in cells on a microscope slide. As a result of its preparative and analytical utility and the ease of assay configuration, PCR is almost universally used in basic research, clinical diagnosis, molecular genetics, veterinary science, human identification, microbial testing of food and environmental samples, archeochemistry, and other fields where there is a need for sensitive analysis of nucleic acids.

### Other Target Amplification Methods

In addition to PCR, there are many other technologies to amplify nucleic acids. For example, ligation-based amplification or ligase chain reaction uses sequence-directed oligonucleotide primers and thermostable DNA ligase to assay point mutations, deletions, or insertions in DNA. Strand-displacement amplification uses the inherent strand-displacement activity of DNA polymerases to conduct DNA amplification at a constant temperature. Transcription-based methods such as nucleic acid sequence-based amplification (NASBA) involve *in vitro* RNA transcription. NASBA and most other transcription-based



**Scheme 9**

methods have three major advantages over PCR: (1) RNA templates do not have to be converted to cDNA before amplification; (2) RNA transcription is isothermal and thus does not require thermal cycling instrumentation as does PCR; and (3) RNA products are less stable than the DNA products of PCR, making contamination of reagents with previously amplified products less likely.

### Signal Amplification in Nucleic Acid Analysis

Various signal amplification approaches have been developed to allow sequence-specific detection without a target amplification step. For example, branched DNA probes contain an analyte-specific sequence and a sequence tag that is used to bind probes for signal detection. The assay culminates in the attachment of an enzyme-labeled reporter probe to facilitate ultrasensitive chemiluminescent detection. Reporter probes tagged with a replicable sequence provide another means to amplify signal from an analyte. Thus far, signal amplification has been limited to relatively abundant targets, while PCR detection of single molecules is performed routinely.

*See also:* **Forensic Sciences:** Blood Analysis; DNA Profiling. **Lead. Phosphorus. Polymerase Chain Reaction. Sulfur.**

### Further Reading

- Abramson RD and Myers TW (1993) Nucleic acid amplification technologies. *Current Opinion in Biotechnology* 4: 41–47.
- Andras SC, Power JB, Cocking EC, and Davey MR (2001) Strategies for signal amplification in nucleic acid detection. *Molecular Biotechnology* 19: 29–44.
- Belcher R (1977) Amplification reactions. *Talanta* 24: 533–534.
- Belcher R and Stephen WI (1984) Recommendations on use of the term amplification reactions. *Pure and Applied Chemistry* 54: 2554–2556.
- Blaedel WJ and Bogoslaski RC (1978) Chemical amplification in analysis: A review. *Analytical Chemistry* 50: 1026–1032.
- Burns DT and Townshend A (1992) Amplification reactions; origins and definitions – progress and present status. *Talanta* 39: 715–735.
- Deiman B, van Aarle P, and Sillekens P (2002) Characteristics and applications of nucleic acid sequence-based amplification (NASBA). *Molecular Biotechnology* 20: 163–179.
- Lisby G (1999) Application of nucleic acid amplification in clinical microbiology. *Molecular Biotechnology* 12: 75–99.
- Matsumura H and Miyachi S (1980) Cycling assay for nicotinamide adenine dinucleotides. *Methods in Enzymology* 69: 465–470.
- Price NC and Stevens X (1989) *Fundamentals of Enzymology*, 2nd edn. New York: Oxford University Press.
- Thiele D (1990) The technique of polymerase chain reaction. A new diagnostic tool in microbiology and other scientific fields. *Zentralblatt für Bakteriologie* 273: 431–454.
- Weisz H (1988) Amplification reactions in microanalysis. A survey. *Mikrochimica Acta* III: 315–326.
- White TJ, Arnheim N, and Erlich HA (1989) The polymerase chain reaction. *Trends in Genetics* 5: 185–188.

## ANALYTICAL REAGENTS

Contents

**Specification**

**Purification**

### Specification

**E J Newman**, Newman, Bucknell Associates, Wimborne, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 108–114, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

### Introduction

During the last half-century, analytical chemistry has undergone a complete revolution. There are clear needs to establish compositions at forever-decreasing low levels of concentration and mass, in an ever-widening range of matrices for purposes as diverse as health and safety, materials science, the environment, and microchip technology. At the same time, laboratories have had to learn to cope with the analysis



of increasingly large numbers of samples. These pressures, together with improved understanding of the physicochemical principles of chemical separation, sensors, and spectroscopy, and the opportunities offered by progress in computing, electronics, and automation, have presented analysts with a huge choice of powerful, sensitive, and productive techniques for tackling their problems.

Some of these techniques permit direct examination of the sample and have little or no requirement for reagents. Most, however, are applied to solutions of samples, frequently after separation or preconcentration. As the role of classical analysis has declined from being dominant to relatively unimportant, there has been a reduction in demand for many reagents that were formerly widely used, such as the many organic chelating reagents for metals that were applied in gravimetric and colorimetric analysis, and used as spot-test reagents and metal indicators.

The techniques referred to above are still with us, of course, and it seems that no analytical approach ever completely disappears, but the range of classical methods that are still used has decreased very significantly, and suppliers of reagents have reacted by withdrawing many formerly popular reagent chemicals from their catalogs.

At the same time, new techniques have created demands for new reagents, as may be illustrated by mentioning derivatization reagents for gas and liquid chromatography, chemicals for electrophoresis, and nuclear magnetic resonance (NMR) shift reagents. More will be said about these developments in later sections of this article.

The present-day practitioner must have a very broad and detailed knowledge of available techniques, including areas of application, benefits such as speed, selectivity, and sensitivity, and limitations such as costs and sources of error, in order to make skilled judgements about methods and to take advantage of the range and diversity of the problem-solving capabilities of analytical chemistry.

It is also important that the chemist should make an educated selection of the other major tools of his profession, including chemical reagents. In many cases, advice is readily available from suppliers, especially with regard to the choice of materials for well-established analytical methods. However, the rate of change in analytical chemistry is fast, and when a new technique is employed there may not be chemicals readily at hand that are suited to the purpose, and they may not become available for many years until the demand is of sufficient commercial interest. In the meantime, the analyst is faced with the prospect of purifying materials. This is the subject of the second article in this entry on analytical reagents.

## The Concept of Purity Applied to Reagents

A pure compound is one that contains no molecules or ions other than those indicated in its chemical formula. In reality no such compound exists, of course, and so it is common practice in the chemical, pharmaceutical, and related industries to express the closeness with which substances approach the ideal by means of specifications controlling the impurities they contain to maximum permissible concentrations.

Other properties of the substance giving an indication of its purity level will be included, such as (depending on its nature) maximum allowable quantities of insoluble matter, nonvolatile matter or residue on ignition, the pH of solution, density, melting point or boiling range, refractive index, and specific rotation.

Another indication of degree of purity that is often used for an organic liquid is to establish that only one peak is produced under examination by gas or liquid chromatography using one or more sets of conditions.

It is possible for specifications expressing maximum limits of impurities to give a misleading view of purity if it is assumed that the supplier is aware of all the possible impurities that could be present in a given chemical and has taken steps to control them.

It is not advisable to assume that all significant impurities have been controlled, even when the specification is supported by an emission spectroscopic examination certificate stating that other impurities have not been detected. For example, a metal oxide might contain a significant amount of the carbonate, and a pure metal could contain a significant content of occluded oxygen. In general, a supplier will usually not be aware of the presence of an impurity that has never been specifically sought. Commercial considerations might also result in the omission from a specification of an impurity whose limiting concentration would detract from the high standard implied by the other requirements of the specification.

The analytical chemist's own knowledge of chemistry taken in conjunction with the published specification does, however, give a good guide to the purity of materials. First there are common associations, caused by similarities in chemical behavior, that are to be expected and indeed do occur in chemicals. In chlorides, for example, bromides would be expected to be present, and magnesium would be expected as an impurity in calcium salts. Low concentrations of bromide in a chloride and of magnesium in a calcium compound are good indicators of a high degree of purification.



All other expected relationships of elements or ions may be used in assessing the extent of purification of inorganic substances. Common associations include sodium and potassium; calcium, strontium, and barium; phosphorus and arsenic; arsenic and antimony; and iron, cobalt, and nickel. There are a number of less well known but useful associations that are often described in textbooks of geochemistry, including silver with lead and its compounds; arsenic with copper and its compounds; and manganese in magnesium compounds.

The degree of purity of an organic compound is less easy to judge from its specification, but a few very simple relationships can, with experience, be quite helpful. For example, control of isomers in substituted aromatic compounds, of homologs, of decomposition products, of products of simple oxidation such as corresponding acids in aldehydes, or of reduction such as corresponding alcohols in ketones, and of by-products from known synthetic routes.

Finally, it is important to mention the widely distributed and abundant elements sodium, potassium, calcium, magnesium, aluminum, silicon, and iron. These are present to some extent in the vast majority of chemical products and, although not an infallible guide, their concentrations indicate the degree of purification of a chemical and the care with which it has been handled, stored, and packaged.

## Classification of General Reagents

The fitness for purpose of a reagent may be impossible to judge unless it has a small number of applications. Chemicals such as nitric acid or ammonia solution have such a vast range of uses that it is unusual to find them specially prepared for any particular application. Suppliers will usually address this situation by offering a range of, say, three or four quality standards of the reagents, with specifications representing different degrees of purification, thus leaving analysts to select the brands most suitable for their work.

Thus, for one of the substances mentioned above that is available in, say, four grades, as a general rule, the cheapest grade will have the fewest 'maximum levels of impurities', controlled at mg per kg or even tens of mg per kg. The most expensive will have a long list of impurities controlled at sub- $\mu$ g per kg, possibly ng per kg concentrations.

It is important for the analyst to exercise good judgement in the selection of his reagents and this requires an understanding of the usefulness and possible limitations of chemical specifications. The use of very high purity reagents may add unnecessary

costs to an analysis, yet the failure to use an appropriate high purity chemical is often a false economy that leads to a greater increase in indirect costs than is saved by reagent costs being lower.

It is encouraging that, in recent years, many high purity liquids, ammonia solution, acetic acid, and the mineral acids, hydrogen peroxide solution, and a range of organic solvents, have become available mainly because of the responses of prime manufacturers to the stringent requirements of the microelectronics companies for high-purity processing chemicals for use in microchip production. As a consequence, not only have improved materials become available, but the use of ion chromatography, inductively coupled plasma (ICP) emission spectroscopy, and ICP mass spectrometry has enabled impurity profiles of these chemicals to be established for a wide range of substances, often for as many as 50 impurities, and most of them at  $\mu$ g per kg level or below.

This advance has had a knock-on effect for the long-established analytical reagents grade, in which significant improvements have been made to the materials and their specifications. Limits for most impurities in analytical grade acids, solvents, ammonia solutions, and hydrogen peroxide solutions are now quoted at  $\mu$ g per kg levels.

The quality control analyses of these chemicals are performed using almost the whole range of trace analysis techniques available. Among the most important are atomic absorption spectrophotometry in all its forms, ICP emission spectrometry, and ICP mass spectroscopy, ion chromatography, gas and liquid chromatography, ultraviolet and visible absorption spectrophotometry, voltammetry, and spectrofluorimetry.

These reagents, and many of those mentioned later, are also often assayed. The determination of the assay of a substance may be useful for many reasons: it may serve as a confirmation of identity via gross composition, to ensure that for a hydrated compound the correct hydrate has been manufactured, to provide concentration values for reagents that are supplied as solutions, such as ammonia solutions and mineral acids, and to assay materials that always contain quantities of water but not as hydrates, of which perhaps the commonest is sodium hydroxide.

Specifications are guarantees of minimum standards of purity and are not to be regarded as actual analyses or typical analyses. This point is frequently misunderstood and has caused some suppliers to quote typical or actual batch analyses. Neither of these conventions has any practical value to the user unless also backed by published specifications which ensure that the ranges of typical or actual batch data

do not include concentrations of impurities that would be unacceptable for the intended use.

It is also necessary for the user to take into account that fine chemicals for laboratory use are usually produced by batch processes rather than by continuous methods of manufacture. Therefore, because specifications set minimum standards of purity, the analysis of one or two batches of a given reagent from a supplier will not give reliable general information concerning the impurity profile of that product.

The convention of operating to specifications also introduces the possibility that a reagent supplied as a lower grade will meet the specification of a higher grade of the product. This occurs because of batch-to-batch variation and it would be unwise to draw any other conclusion. As a general rule, analysts should select as reagents for regular use only those materials whose specifications (either published or agreed with the supplier) meet the requirements.

### High-Purity Reagents

Reagents of somewhat higher purity than conventional analytical reagent grade are now increasingly required for a wide range of analyses in fields as diverse as the environmental, materials, and life sciences.

These were first required in the early 1950s when the UK manufacturers were asked by the Society of Public Analysts to provide high-purity acids and ammonia solutions for use in regulatory analyses of lead in foodstuffs. These 'low in lead' reagents were soon prepared with guarantees that they contained less than an agreed limit of 0.005 ppm (i.e., less than 5 µg per kg) of lead as determined by a specific colorimetric method based on dithizone. This was a significant achievement for the time and, as expected, as analytical sensitivities improved, the suppliers were able to confirm that the purification processes made significant overall improvements to the products. Low-in-lead reagents were manufactured in small batches but the lessons learned were transferable to a much larger scale when demand came within the next decade for high-purity processing chemicals for semiconductor manufacture.

Reagents of high overall purity include the acids, ammonia solution, hydrogen peroxide solution, and organic solvents previously discussed. Several high-purity solid substances having wide-ranging applications are also available to standards rather higher than those of the corresponding analytical grade materials. These include sodium hydroxide; a number of pH buffer components; ashing aids used when the resulting ash will be tested for trace elements; biochemical reagents such as ammonium sulfate,

sucrose, and urea; and matrix modifiers employed in graphite furnace atomic absorption spectrometry. The testing of these items usually involves multielement techniques such as ICP spectrometry and ion chromatography plus electrothermal atomic absorption spectrometry for selected metallic impurities, spectrofluorimetry, and gas chromatography or liquid chromatography as appropriate.

Ultrahigh-purity acetic acid, ammonia solutions, and mineral acids prepared by double subboiling distillation, having ng per kg impurity levels, are available commercially for use in the most demanding trace analyses.

### Standard Reference Materials

The International Standards Organization (ISO) has defined a reference material as a material or substance one or more properties of which are sufficiently well established to be used for the calibration of a method, or for assigning values to materials. The ISO definition of a certified reference material is a reference material one or more of whose property values are certified by a technical procedure, accompanied by or traceable to a certificate or other documentation that is issued by a certifying body.

Certified reference materials are used to provide reference values to facilitate the development and validation of analytical methods, and for the calibration, verification, and quality control of analytical measurement systems.

A standard reference material may therefore be either a highly pure reference compound or a well-characterized substance or a calibration standard.

Primary, secondary, and working chemical standards are high-purity chemicals. They have been defined by the Analytical Chemistry Division of the International Union of Pure and Applied Chemistry (IUPAC) as follows:

- A primary standard is commercially available and has a purity of 99.98–100.02%.
- A working standard is commercially available and has a purity of 99.95–100.05%.
- A secondary standard is a substance that may be of lower purity and which can be standardized against a primary standard.

From the earlier discussion on purity, it will be appreciated that an assay alone will not be sufficient to establish a primary standard unless it is certified for use in similar assays. It will generally be necessary to select material that can be established as highly pure by examination for all the impurities that may be

present. The tolerance values placed by IUPAC on the purity for a primary standard, for example, may then be interpreted as meaning that the sum of the impurities is not greater than 0.02% and that on assaying by a suitable high-precision method the calculated result falls between 99.98% and 100.02%.

High-purity materials are available as pure solids for chemical uses such as metals used as reference substances for metallurgical analysis, and compounds used as primary standards for many types of titrimetry, and as standards for elemental microanalysis. Other available high-purity substances intended for diverse physical properties include ion activity standards for the calibration of pH and ion-selective electrodes; standards for various thermodynamic uses including melting point determinations, differential scanning calorimetry and bomb calorimetry; and standards for the calibration of spectrophotometers.

The majority of chemical reference materials (CRMs) are widely analyzed materials having one or more chemical or physical properties sufficiently well established to be used as reference values for calibration or performance assessment. Items as diverse as gas chromatography-mass spectrometry system performance standards, human or bovine serum, fly-ash, soils, estuarine sediments, rice flour, and stainless steels are included in the vast range of CRMs available.

Standard solutions and gas mixtures are also available for many single- and multisubstance techniques; for example, drug assay standards, standard solutions of metals in single- or multielement form, and carbon monoxide and nitrogen dioxide in air.

Finally, there are many authentic specimens and fairly high-purity organic materials that are used to confirm chromatographic retention times and infrared, NMR, and mass spectra.

There are three methods of establishing a certified reference value. The first is to use an established reference method that is based on fundamental principles and that has high precision and negligible bias so that the uncertainty limits of the certified value are small. This work may be performed by one laboratory or a small number of laboratories. Another method often used for organic substances and carried out by one laboratory is to use two or more methods that have been shown to be reliable and that are based on different principles. An important criterion for acceptance of data when using this approach is that the values obtained by each of the methods employed must fall within the permissible uncertainty interval of the end-use of the substance. The third method is to use the statistically established 'consensus' value obtained from a cooperative study in

which many laboratories conduct analyses of the substance, preferably using a number of different analytical methods, including any established reference method.

## Organic Reagents for Inorganic Analysis

A generation ago, when much of analysis still consisted of classical and colorimetric methods, several hundred organic reagents were available to the analyst. Although the use of these methods has declined very considerably, they are still used and several of the better known organic reagents remain listed in the suppliers' catalogs.

Organic reagents found applications as precipitants for gravimetric analysis, extractants for separating small quantities of metals, colorimetric reagents, titrants (principally ethylenediaminetetraacetic acid (EDTA)) and metal indicators for complexometry, and masking agents to prevent undesirable reactions from occurring during analysis.

Organic reagents are weak acids or bases, and in their conjugate base forms they provide ligands that combine with metal ions to form coordination complexes. The most stable complexes are generally formed when the ligands are chelating, that is they have two or more combining functional groups so that each ligand group occupies two or more coordination positions.

Some organic reagents are selective and react with very few metals, but the majority are far from selective. However, highly selective conditions may often be derived using pH control and masking and sometimes careful oxidation or reduction of potential interferents.

Some of the most successful and widely used chelating reagents include dimethylglyoxime for the gravimetric determination of nickel; 1,10-phenanthroline and its derivatives for the colorimetric determination of iron and copper; dithizone for the separation and colorimetric determination of a number of metals but particularly lead, silver, zinc, cadmium, and mercury; the dithiocarbamates such as diethylammonium diethyldithiocarbamate and ammonium pyrrolidinedithiocarbamate, used for colorimetry but more widely applied now as selective extractants; and the most successful titrant, EDTA.

Two flow techniques in particular, the use of liquid chromatography or ion-exchange chromatography to separate metal ions followed by postcolumn derivatization of the metals in the effluent, and the determination of metals by flow-injection analysis, have created further applications for suitable organic

chemical reagents. Colorimetric methods using organic reagents are also employed in automatic high-productivity techniques, such as in discrete analyzers used in clinical biochemistry and water-industry laboratories. Colorimetric reagents will also be considered in the next section of this article.

Other types of organic reagent include those used to form colored ion-association systems, precipitants forming insoluble normal salts, and the insoluble salts of chloroanilic acid.

Some dyestuffs form ion-association systems with large inorganic anions. Brilliant green co-extracts stoichiometrically with hexachloroantimonate ion, providing a method for the determination of antimony.

The normal salts of several carboxylic acids such as oxalic, benzoic, and mandelic acids are insoluble and useful gravimetric methods were based on this property. It is also worth mentioning that sodium tetraphenylmetaborate is soluble, whereas its potassium salt is insoluble, and that selective gravimetric and titrimetric methods are thus possible.

The insoluble salts of chloroanilic acid may be exemplified by barium chloroanilate, which is still used for the automated monitoring of sulfur dioxide emissions, after conversion to sulfate with hydrogen peroxide. Solid barium chloroanilate reacts with solutions containing sulfate to form insoluble barium sulfate, thus releasing an equivalent concentration of chloroanilic acid, which is measured by ultraviolet spectrophotometry.

The standards of organic reagents vary widely. Several of the better-known reagents are available as analytical reagent grade products, but the majority has lower and variable purity, especially those that are produced primarily as commercial dyestuffs. They are usually tested only to ensure that they give the required color reactions or precipitates and, as they are normally used to a large extent to the determination of this type of examination is insensitive to batch-to-batch variation.

## Chromogenic Reagents, Including Indicators

Chromogenic reagents include the various kinds of chemicals used as pH indicators (for both aqueous and nonaqueous systems), redox indicators, metal indicators, giving changes usually in color but also in fluorescence or luminescence, and reagents used for visual and/or absorptiometric colorimetry.

The indicator properties result from the different colors of the conjugate acid–base forms of the materials used as pH indicators, or of the oxidized and reduced forms of redox indicators, and the

metal-complexed and uncomplexed forms of the metal indicators. Fluorescent or chemiluminescent indicators have advantages for use in colored or turbid solutions, and for remote sensing.

The most common acid–base indicators are either azo dyes: for example, methyl orange and methyl red; nitrophenols; phthaleins such as phenolphthalein or thymolphthalein; or sulfonephthaleins like bromophenol blue or bromocresol green. Acid–base indicators are available that cover visual transitions usually expressed in intervals of 2 pH units ranging from pH 0.0 to 2.0 in small increments up to pH 12.0–14.0.

Redox indicators include indophenols: for example, 2,6-dichlorophenolindophenol; azine dyes such as the well-known thiazine dye methylene blue; indigo carmine and other indigo derivatives; derivatives of diphenylamine including diphenylamine-4-sulfonic acid and variamine blue; and the 1,10-phenanthroline–ferrous complex.

Metal indicators include a number of well-known organic reagents: for example arsenazo III, catechol violet, dithizone, 1-(2'-pyridylazo)-2-naphthol, and 4-(2'-pyridylazo)resorcinol. Also included among metal indicators are several commercial dyes such as chrome azurol S, eriochrome black T, eriochrome blue-black B, and pyrogallol red. Finally, there are the 'designer' metal indicators containing Mannich reaction-substituted iminodiacetic acid groups, exemplified by alizarin complexone, methylthymol blue, and xylenol orange.

As a general rule, these indicators are, with a few exceptions, somewhat impure and full chemical analysis shows large batch-to-batch variations. They are, however, normally tested for their suitability as indicators under standard conditions of use, and their measured visual or instrumental transition intervals must usually comply with standards set to ensure parity of performance between batches.

## Other Developments in Analytical Reagents

### Special-Purpose Reagents

Since the 1960s the trend in reagent supply has been increasingly toward the introduction of chemicals intended for specific applications. Reagents were first provided for techniques such as ultraviolet spectroscopy and later for liquid chromatography, electrophoresis and so on, and for special areas of application such as histology, hematology, and electron microscopy, and also for specific purposes like the NMR shift reagents.

Other special-purpose reagents include many of the convenience reagents discussed in the next section and, of course, a huge range of materials for chromatography including materials for supports, stationary phases, silylating reagents, and derivatization reagents.

In general, the quality control of these items consists of ensuring that they are suitable for the intended purpose. Some, such as the ultraviolet and liquid chromatography solvents are of intrinsically high standard, but others, particularly stains for histology, may be chemically far from pure. The analyst must therefore be particularly cautious if using such a reagent for any purpose other than that for which it has been listed by the supplier.

### Convenience Preparations

The most familiar convenience reagents are acid–base and pH indicator papers. Several other types of indicator papers exist for redox and more specific test purposes; for example, the estimation of glucose in blood and urine. The principle has been extended using a range of colorimetric reactions into the use of test kits for rapid and semiquantitative analyses, particularly of waters and soils. Reagents impregnated onto inert support materials in tubes through which measured amounts of air are drawn are used in large quantities to monitor workplace environments. These tubes are available for many different substances. Color reactions in the tubes give estimates of the concentrations of the tested species that are present in the air.

Convenience reagents also include ready-to-use buffers, bench reagents, and standard solutions and indicators for titrimetry. Buffers and standard volumetric solutions are also available in concentrated form so that a simple volumetric dilution of the contents of a container will produce the required reagent concentration.

Standard solutions of inorganic species, both non-metals and metals, are particularly widely used in laboratories. Multielement metal standards, for use in ICP emission spectroscopy and other multielement techniques, must be backed with a guarantee that high-purity components are used in their formulation; otherwise, trace impurities in the individual ingredients may contribute significantly toward total quantities of the very elements it is required to measure. Undesirable ‘blanks’ of this kind are easily overlooked and can seriously degrade analytical performance.

It is now common for manufacturers to offer ready-to-use reagents, often as complicated mixtures, for use in automated analytical procedures such as in clinical biochemistry and for online process control.

Methods of quality control (QC) of convenience reagents vary according to type. Simple solutions are controlled by analysis and for standard solutions an accuracy of 0.1% is required wherever attainable. Multielement metal standards are prepared from high-purity starting materials that have been well analyzed, and are tested by a multielement method to ensure that no extraneous contamination has occurred. Prepared reagent mixtures for specific applications and test kits are subject to both QC analysis of the individual components and then, after preparation, have to meet performance criteria.

**See also:** **Analytical Reagents:** Purification. **Indicators:** Acid–Base; Redox; Complexometric, Adsorption, and Luminescence Indicators. **Quality Assurance:** Quality Control; Primary Standards; Reference Materials; Production of Reference Materials.

### Further Reading

- American Chemical Society (2000) *Reagent Chemicals*, 9th edn. Oxford: Oxford University Press.
- Analytical Methods Committee, Society for Analytical Chemistry (1972) Reference substances and reagents for use in organic microanalysis. *Analyst* 97: 740.
- Bucknell DJ (ed.) (1984) *‘AnalalR’ Standards for Laboratory Chemicals*. Poole: BDH Chemicals Ltd.
- International Standards Organization (1983–1989) *Reagents for Chemical Analysis*, ISO 6353, Parts 1–3. Geneva: ISO.
- LGC Teddington, *Reference Materials*, online catalog, at [www.lgc.co.uk/ref.asp](http://www.lgc.co.uk/ref.asp)
- Merck E (1972) *Merck Standards*. Darmstadt: Merck.
- National Institute of Standards and Technology, *Standard Reference Materials*, available on-line at [www.nist.gov](http://www.nist.gov)
- Prolabo (1982) *Prolabo Analytical Standards Reagents*. Paris: Prolabo.
- Ross E, Kothe J, Naumann R, *et al.* (2003) Indicator reagents. In: *Ullmann’s Encyclopedia of Industrial Chemistry*, 7th edn., vol. 17, pp. 645–671. Weinheim: Wiley-VCH.
- Stephen WI and Townshend A (1984) A 50-year quest for purer reagents. *Chemistry in Britain* 20: 1008–1015.
- Townshend A, Burns DT, Guilbault GG, *et al.* (1993) *Dictionary of Analytical Reagents*. London: Chapman and Hall.

## Purification

**E J Newman**, Newman, Bucknell Associates, Wimborne, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 114–119, © 1995, Elsevier Ltd.

### Introduction

Methods for the purification of chemicals belong to the realm of separation science. Every process for the separation of one species from others has been employed for purifying substances. The traditional techniques of distillation, precipitation, recrystallization, and, for many metals, refining and electrolysis, are the basic processes of purification and are, for most industrial and technical purposes, quite sufficient.

Most chemicals for analytical use are purified adequately by the application of traditional methods to selected starting materials, particularly for the regular analytical-reagent grades in which impurities are controlled at mg per kg levels. The success of these methods depends on the nature of the substances concerned and the degree of purity of the starting materials in respect both of overall purity and of specific impurities that may be difficult to remove.

The important relationship that generally exists between the careful study and selection of raw materials and the purity of the finished products cannot be too strongly emphasized, and may be assumed to apply to the methods presented throughout this article.

Another important general consideration for the improvement of purity is the choice of materials for the construction of the apparatus in which purification operations are to be performed. The most appropriate materials to withstand chemical attack may nevertheless introduce undesirable impurities into the product. For this reason it is frequently necessary to run process trials in the equipment until available impurities have been leached from the apparatus.

For the highest grades of purity and for certain chemicals, the well-known techniques of distillation and recrystallization cannot be used, except as the first stages of purification. This is particularly true either when exceptional purity is needed or when an undesirable impurity is not separated by the physical process applied. In such cases two or more techniques may be successively applied to achieve purification. Alternatively, it may be necessary to use

different physical processes. The more common of these include sublimation, chromatography, and zone refining.

In many liquids, such as the mineral acids, impurities are known to be present in vast numbers of minute particles and significant purification is achieved by ultrafiltration under clean-room conditions.

The major impurity in a solid chemical is often water or the solvent from which it was crystallized, so that drying, or solvent removal, is another purification process.

Finally, there are numerous chemical methods for the removal of impurities. Such methods are aimed at reducing the concentrations of particular impurities and so it is not possible to describe them comprehensively. Instead, several examples are outlined to indicate the range of approaches that have been used successfully.

The variety of techniques available for the purification of small amounts of substance is particularly wide, and those that are used for very narrow ranges of chemicals or for submilligram quantities are not considered here.

### Distillation and Sublimation

#### Distillation

Distillation is one of the two most widely used techniques for the purification of liquids and low-melting-point solids. Simple distillation will only effect separation from solid impurities, of course, so that fractional distillation is more common, particularly for the purification of organic compounds, and with modern apparatus equipped with computer control it is possible to achieve very high levels of purity. Fractional distillation is applicable over a very wide range of quantities of substance from subgram amounts to the tonnages manufactured by the heavy organic chemicals industry.

Organic liquids with high boiling points or those that are liable to undergo partial decomposition when distilled at atmospheric pressure should be distilled or fractionally distilled at reduced pressure, the process being commonly called vacuum distillation. Some practical precautions that must be taken in vacuum distillation are dealt with in the first reference in Further Reading. It is important also to consider safety aspects that apply to particular chemicals, for example, that ethers may contain explosive, thermally labile peroxides, and should be tested

before use and chemically treated if necessary. Another important safety precaution to consider is that stills accommodating even a few liters of organic liquid may well need earthing and blanketing with inert gas to ensure safe operation.

Constant-boiling mixtures, known as azeotropes, are formed between several mixtures of two or more compounds. The commoner form of azeotrope has a constant boiling point lower than that of either component, and familiar examples include ethanol with water, benzene with cyclohexane, and methanol with methyl acetate. The less common azeotropes have boiling points higher than those of the individual components, and these include the well-known mineral acids, and mixtures of acetic acid with pyridine, and of acetone with chloroform. Methods for the separation of the components of an azeotropic mixture include chemical treatment, such as removing water with a molecular sieve or by 'salting out' the organic phase from an azeotrope with water; fractional crystallization of the mixture; and distillation with a third substance that forms a ternary azeotropic mixture. For example, benzene added to the ethanol-water azeotrope produces a lower-boiling-point azeotrope that can be distilled off to remove the water.

The technique of steam distillation is sometimes used in purification because it is fairly selective in that relatively few water-insoluble substances are steam-volatile, and because it facilitates the distillation of some compounds at temperatures well below their normal boiling points. The codistilled water is usually easily removed. Examples of its use include the distillation of nitrobenzene and naphthalene.

Acids and ammonia solutions intended for use in trace analysis present particular problems. They are widely needed, often in significant quantities, for a wide range of metal determinations. Their chemical reactivity presents the problem that during purification the apparatus employed must be carefully selected and conditioned to minimize its contamination of the product, and after it has been purified the containers for its storage present similar difficulties.

Many years ago, in response to the requirement for specially purified acids for use in the determination of lead in foodstuffs, suppliers provided 'low-in-lead' reagents that had been redistilled two or three times in well-conditioned glass or quartz apparatus that had been specially selected as being essentially lead-free. Later demands for larger quantities and a wide range of controlled impurities could be met by scaling up the process somewhat, but production by this means is expensive and does not yield product that meets the needs of current trace analysis capabilities: it delivers product that has roughly  $\mu\text{g kg}^{-1}$

impurity levels but not lower, because of carryover during distillation.

Laboratory-scale preparations of ammonia and volatile acid solutions intended for use in trace metal analysis have been successfully performed by isothermal distillation, usually at room temperatures. In this technique, two beakers, one containing the chemical being purified and the other holding pure water, are placed together in a sealed closed container for a few days. The volatile component becomes uniformly shared between the contents of the two vessels and nonvolatile impurities remain in the original beaker.

The most successful method for purifying volatile acids, ammonia solutions, and solvents, and which is operated on a small commercial scale, is that of sub-boiling distillation. The apparatus is constructed from quartz or, for hydrofluoric acid, of polytetrafluoroethylene (PTFE), and infrared radiators vaporize the surface of the liquid without bringing it to the boil. The vapor is condensed on a tapered 'cold-finger condenser' and the liquid is collected in a suitable container, itself constructed from quartz or PTFE. Apparatus is available that will deliver 20 or more liters per still per day.

Finally, mention must be made of so-called 'molecular distillation', which has been used to purify small amounts of compounds of low volatility, often for use as reference materials. At very high vacuum, the mean free path of a molecule becomes quite large, for example, several centimeters. Molecular distillation makes use of this fact. The liquid of low vapor pressure is evaporated under high vacuum with only a short straight-line distance between its surface and that of the condensing surface. This form of distillation uses much lower temperatures than ordinary vacuum distillation and significantly reduces thermal degradation of the product. The distance between the surfaces of the liquid and the condenser may be adjusted, for a given temperature of operation, to be within the mean free path of molecules of given relative molecular mass.

### Sublimation

In sublimation, a solid substance is volatilized by heating and the vapor is condensed back to the solid at a cooled surface. The distance between the surface of the vaporized solid and the collecting surface is short compared with distances used in distillation. Sublimation may be conducted at atmospheric pressure but reduced pressure is often employed to enhance sublimation and to speed up the process. An atmosphere of inert gas at low pressure is advisable for sensitive compounds.

Purification through sublimation is applicable to a number of organic and inorganic substances and is useful for the purification of many simple inorganic compounds used as working standards in analysis, including ammonium halides, arsenic(III) oxide, phosphorus(V) oxide, and iodine.

## Recrystallization

Recrystallization is the second of the two most widely used methods of purification. The normal procedure is to dissolve the substance to be purified in a suitable solvent, at an evaluated temperature, to form an almost saturated solution. It is then allowed to cool so that the dissolved substance crystallizes out. The crystalline product is then separated by filtration or centrifugation, leaving impurities behind in the solution or 'mother liquor', and is then washed free of mother liquor with successive small volumes of the cold solvent, and dried.

It is good practice to remove insoluble impurities from the original hot solution by filtration, so that they will not be included in the product, and this requires the use of a filter in which the high temperature can be maintained.

Organic compounds in particular may contain colored impurities and it is frequently possible to remove these by thoroughly mixing the original hot solution with activated charcoal before filtration and crystallization.

When chemical treatments are employed to remove selected impurities, they are usually also applied to the first stage of recrystallization. Chemical treatments are the subject of a later section of this article.

Second and subsequent recrystallization will usually improve purity, and for greater effectiveness two or more recrystallizations from different solvents should be used if possible.

Two considerations are worth noting. The first is that rapid crystallization is often recommended because slow crystallization gives larger crystals, which, although they look more perfect to the naked eye, quite often contain occlusions of mother liquor. However, trace chemical impurities tend to concentrate at crystal surfaces and so a high surface area is not compatible with best chemical purity. Preliminary recrystallization with slow cooling to produce well-formed crystals followed by a second recrystallization with rapid cooling has often been employed to overcome both problems.

Some substances have solubilities in certain solvents that do not vary significantly with temperature: a common example is sodium chloride in water. But because the equilibrium between a solid and

its saturated solution is dynamic, prolonged digestion of such a substance with the solvent has the effect of recrystallizing it.

Recrystallization does not always achieve the desired degree of purification. The author has experience of a few compounds that attained higher concentrations of certain impurities on successive recrystallizations. This is to be expected if an impurity has a crystalline form isomorphous with that of the bulk material or can in some way enter the host lattice, and this is often not possible to predict.

## Zone Recrystallization

Zone recrystallization incorporates practices from both recrystallization and column chromatography but is a purification technique in its own right. It is simple to operate, economical with both the substance to be purified and the solvents used, and is almost invariably conducted at room temperature, although in principle it can be operated at any temperature. It is very suitable for thermally labile substances and works satisfactorily for compounds of wide-ranging solubilities.

The chemical to be purified is evenly packed in comminuted form into a vertical column. A small volume of solvent liquid is introduced at the top of the column and allowed to percolate through under gravity. The solvent addition is controlled in quantity and rate so that after traveling a short distance along the column it has filled the interstices and has become a saturated solution of the bulk substance. Further small volumes of the solvent are added so that a number of solvent zones are present before the leading zone emerges from the column. The impurities are removed in the liquid effluent. Another solvent in which the host material is insoluble may be used to remove the original solvent.

## Chromatography

A variety of chromatographic methods are used successfully for the purification of small quantities of organic chemicals and biochemicals, based on both gas chromatography (GC) and liquid chromatography (LC).

Gas-liquid chromatography may be used, in principle, to separate the components of any volatile liquid in a column similar to, but somewhat larger than, an analytical column operated at a temperature a little greater than the boiling point of the substance being purified. It is normal practice to use partition (i.e., gas-liquid) chromatography with a packed column containing an inert support material coated with a suitable high-boiling-point liquid, of which a



number are available with different polarities. The separations are monitored using a nondestructive detector, such as a thermal conductivity detector, and the resulting chromatogram is displayed. As the components emerge from the column they are captured by condensation in high-efficiency strongly cooled traps.

Liquid chromatography is rather more widely applicable than gas-liquid chromatography and in some forms has the capability to handle larger amounts of substance. All the main forms of LC are used, namely adsorption, partition, ion exchange, and gel filtration, and some newer forms of chromatography including metal-chelate adsorption, hydrophobic adsorption, and, particularly for biochemicals, affinity chromatography.

The variety of types of adsorbents, solvents, stationary phases, ion exchangers and eluents, gels and special phases, and the numerous applications of chromatographic methods, have given rise to a vast literature on the subject, the review of which is beyond the scope of this article. Instead, some of the most widely used methods and a few of the more interesting applications are presented, and readers are advised to consult the chromatographic literature for more information.

Column chromatography has been practiced on quite significant amounts of material, and adsorption chromatography has been used on a commercial scale for the purification of an extremely wide range of organic compounds. Among the analytical reagents purified by this technique are solvents for ultraviolet spectrophotometry and liquid chromatography (LC), in processes that may also involve chemical treatment and distillation.

In principle, LC is similar to GC, but is more versatile because, whereas in GC only the polarity of the stationary phase can be varied to bring about or improve separation, in LC the polarities of both the stationary phase and the solvent can be varied widely, and also the elution can be performed with solvent programming by means of which the polarity of the eluent can be varied throughout the process as a further aid to separation.

LC uses high pressures to elute small samples through high-efficiency columns fairly quickly, and this combination of pressure and column efficiency results in much better separations than are possible with atmospheric-pressure chromatography systems. The technique can be used only for the purification of relatively small amounts of substance.

LC using chiral stationary phases has been used for the separation and purification of enantiomers and is attracting great interest, particularly because of the biological significance of chirality.

The technique of flash chromatography, in which a larger amount of substance is purified on a fairly high-efficiency column at a few atmospheres pressure, has proved quite useful and is employed, by the pharmaceutical industry for example, for the purification of small amounts of reference materials.

## Zone Refining

Impurities generally lower the melting points of solids or the freezing points of liquids. When an impure liquid is gradually cooled, the first crystals to form are usually purer than the remaining liquid. This principle is the basis of purification by the techniques of fractional solidification, recrystallization from the molten substance, and zone refining. There are instances in which impurities form eutectics with the host material that have higher melting points than the pure substance, so that it is good practice to reject the first crop of crystals as well as the last fraction of the melt. Fractional freezing has been used to purify many organic substances, often on a large scale, including benzene, substituted benzenes, and other aromatic compounds of relatively low molar mass.

Recrystallization from the melt is used to purify several organic compounds for use as reference materials and working standards. Column crystallization is the name given to the technique of filling a column (usually glass) with the molten substance to be purified, and then gradually lowering it into a cooling bath held at a temperature slightly below the expected melting point. Cooling is stopped before the material has fully solidified and the remaining melt is drained off. The technique of recrystallization from the melt has also been employed for the purification of a wide variety of organic compounds including benzoic acid, phenols, aromatic hydrocarbons, and higher aliphatic acids and alcohols. The technique gives the most satisfactory results on substances that are already substantially pure and is therefore often applied as the final stage of purification.

Zone refining or zone melting is a particular form of fractional crystallization applicable to any solid. The apparatus consists of a long tube with an annular heater that is arranged to pass very slowly along the tube (or in smaller devices through which the tube passes) so that a slowly traveling molten zone is created in the column of material. Several passages of the molten zone are made and impurities concentrate at the ends of the column. After the passage of several molten zones, the substance is allowed to cool and the ends containing the impurities are discarded.

The technique is carried out on quantities ranging from the semimicro, for laboratory investigations, to the macro commercial scale. It is used in the production of highest-purity silicon and semiconductors, single-crystal compounds of technical importance such as alkali metal halides, and for many organic compounds. As mentioned earlier, the technique works best for substances that are already fairly pure, and in the field of analytical reagents is particularly useful for converting analytical-grade substances into superpure grade for use as working and reference standards. Reagents and reference substances purified in this way include benzoic acid; alkali halides; naphthalene, anthracene, chrysene, and pyrene; morphine; dibenzyl and biphenyl; benzo-phenone; and 4-aminoacetanilide.

## Drying and Solvent Removal

A solvent is usually removed simply by heating the material at a temperature above the boiling point of the solvent. If heating to this temperature will cause decomposition of the substance, heating under vacuum is the next alternative. In cases where no heating can be used, the substance is sealed in an evacuated vessel containing a suitable absorbent; for example, sodium hydroxide pellets are suitable for removing acetic acid.

Drying can present particular problems. Inorganic substances are often hydrates and it is important not to over dry. These materials are usually left to equilibrate in a sealed vessel containing a suitable hygrostat, which is a strong aqueous solution having a known and suitable vapor pressure. Tables of hygrosats are available in the well-known chemical data books. Labile organic chemicals and biochemicals are freeze-dried under vacuum, and sometimes are allowed to just melt and then refrozen with the vacuum maintained throughout the cycle, which may be repeated as necessary.

The fractional distillation of organic liquids and the use of azeotropes to remove water has been mentioned earlier. In many instances, organic liquids may be dried directly by adding desiccants. Traces of water may be removed from some solvents by treatment with activated alumina or, on the laboratory scale, with sodium, sodium borohydride, or calcium hydride. Molecular sieves of type 4A, i.e., with a porosity of  $\sim 4 \text{ \AA}$ , are also extremely useful for the removal of water and a number of other small molecules from organic liquids.

Drying agents vary greatly in the intensity of their action, and their suitability to particular substances, especially with regard to any hazard and the likelihood of causing other contamination.

## Filtration

Filtration is employed to remove insoluble impurities from solutions and melts. The techniques of filtration of the laboratory and small manufacturing scales need not be repeated here. It is appropriate to remember the wide range of filtration media and aids that are now available and that filtration may be performed hot or cold and under temperature control. For the efficient removal of relatively small quantities of fine suspended matter, centrifugation provides a suitable approach.

All liquids contain vast numbers of submicrometer-sized particles that are too small to be visible but that can be detected and counted with high-sensitivity apparatus. Chemicals used on industrial processing of ultralarge-scale integration devices must be essentially free from these particles and suitable filtration membranes are available with porosities as low as  $0.01 \mu\text{m}$ . It is to be expected that use of these filters will make significant improvement to the chemical purity of the product. In acids and ammonia solution, many trace metal impurities are present in the form of submicrometer particles and ultrafiltration in a suitable clean environment is useful for improving reagents intended for trace analysis at levels of  $\mu\text{g}$  per kg and lower.

## Chemical Treatment

Solvent extraction is used for the separation of substances of different polarity and is employed in freeing organic liquids of inorganic impurities. Examples include the washing of acids from insoluble solvents with sodium carbonate solution and the purification of dithizone solutions by transferring its anionic form into dilute alkali solution, discarding the organic phase, and then reverting into fresh organic phase by acidification of the aqueous solution and extraction. It is also used to remove organic impurities such as grease from inorganic compounds, and generally for separating substances of sufficiently different solubilities. In the laboratory it is often employed to free reagent solutions of troublesome impurities; for example, those that could cause high blank values. This is usually combined with an appropriate chemical reaction. Thus, some reagents intended for use in a test for iron can be purified by pretreatment with 4,7-diphenyl-1,10-phenanthroline and solvent extraction to remove the complexed metal.

Precipitation has been used widely for selective purification in much the same way, but in the author's experience it is less useful because only substantial amounts of impurities give filterable precipitates, and so 'collectors' must be added.

In removing one impurity, therefore, the general purity of a reagent solution will be decreased.

Ion exchange is also used quite widely to purify reagent solutions, particularly to remove trace metals and sometimes anions. Alumina, for example, is a particularly effective anion exchanger for sulfate ion and removes it from acidic and neutral solutions.

Controlled-potential electrolysis has also been used for removing some metals from inorganic solutions, and reagents for use in voltammetry have been purified to remove traces of cadmium, copper, lead, nickel, and zinc.

It is sometimes possible to remove a metal such as copper from an inorganic solution by stirring in the presence of a more reactive metal such as zinc. The copper displaced from solution plates out on the surface of the zinc. An equivalent amount of zinc passes into solution, of course.

There are no useful guidelines for chemical processing because of the range of problems encountered, but a few examples will now be given of approaches that have been used. Lithium aluminum hydride is used to reduce traces of carbonyl compounds, including carboxylic acids and anhydrides, and peroxides to alcohols. It is also useful for reducing amides and nitriles to amines. Precautions must be taken when using the reagent, such as ensuring that the substance to be purified is dry and that reactions are conducted in the cold.

Sodium borohydride is less active but has the property of reducing aldehydes, ketones, and peroxides, and a number of other commonly encountered impurities. Aldehydes are eliminated from acetic acid and other lower monocarboxylic acids by refluxing with sodium or potassium dichromate solution and fractionation. Alcohols and esters for use as UV solvents may be treated with sodium borohydride or with 2,4-dinitrophenylhydrazine before distilling, to remove aldehydes and ketones.

Traces of nitric acid in hydrofluoric acid have been prevented from distilling by pretreatment with aluminum powder to reduce nitrate to ammonium, and

traces of sulfide in the same acid have been removed by oxidizing to sulfate with hydrogen peroxide before distillation.

**See also:** **Distillation. Liquid Chromatography:** Overview; Chiral; Pharmaceutical Applications. **Quality Assurance:** Primary Standards; Reference Materials; Production of Reference Materials. **Solvents. Sublimation.**

## Further Reading

- Breck DW (1974) *Zeolite Molecular Sieves*. New York: Wiley.
- Coatsee JF (ed.) (1982) *Purification of Solvents*. Oxford: Pergamon.
- Diamond PS and Denman RF (1973) *Laboratory Techniques in Chemistry and Biochemistry*, 2nd edn. London: Butterworths.
- Gaylord NG (1956) *Reductions with Complex Metal Hydrides*. New York: Interscience.
- Herington EFG (1963) *Zone Melting of Organic Compounds*. New York: Wiley.
- Jablonski WZ (1970) British Patent Appl., Sept. 1967. *Chemical Abstracts* 72: 33716c.
- Krell E (1963) *Handbook of Laboratory Distillation*. Amsterdam: Elsevier.
- Kuehner EC, Alvarez R, Paulsen PJ, and Murphy TJ (1972) Production and analysis of special high-purity acids purified by sub-boiling distillation. *Analytical Chemistry* 44: 2050–2056.
- Maas RP and Dressing SA (1983) Purification of nitric acid at trace metal levels. *Analytical Chemistry* 55: 808–809.
- Perrin DD and Armarego WLF (1988) *Purification of Laboratory Chemicals*, 3rd edn. Oxford: Pergamon Press.
- Pfann W (1966) *Zone Melting*. New York: Wiley.
- Riddick JA, Bunger WB, Sakano TK (1970) Organic solvents: Properties and methods of purification. In: Weissburger (ed.) *Techniques of Chemistry*, 4th edn., vol. II. New York: Wiley-Interscience.
- Weissberger A (ed.) (1956) *Techniques of Organic Chemistry*, 2nd edn., vol. III, part 1. New York: Interscience.
- Zief M and Horvath J (1976) Contamination in trace analysis. *Laboratory Equipment Digest*, October 1976, p. 47.

# ANODIC STRIPPING VOLTAMMETRY

See **VOLTAMMETRY: Anodic Stripping**

# ANTIOXIDANTS

See **FOOD AND NUTRITIONAL ANALYSIS: Antioxidants and Preservatives**

# ANTIQUES

See **ARCHAEOMETRY AND ANTIQUE ANALYSIS: Dating of Artifacts; Metallic and Ceramic Objects; Art and Conservation; Organic and Biological Materials**

## ARCHAEOMETRY AND ANTIQUE ANALYSIS

### Contents

#### Dating of Artifacts

#### Metallic and Ceramic Objects

#### Art and Conservation

#### Organic and Biological Materials

### Dating of Artifacts

**J Ambers**, The British Museum, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Considering the huge variety of surviving archaeological artifacts and contexts, the range of scientific dating techniques available to the archaeologist is remarkably small. All are material specific and seldom (with the exception of dendrochronology) can an accuracy of better than a few hundred years be achieved. There are also some obvious lacunae: there is no method of dating the manufacture of purely metallic objects, or of the vast majority of lithics, and in such cases, archaeological dating depends on the traditional methods of stratigraphy and typology (the importance of which cannot be underestimated, but which do not fall within the remit of this article).

Even where material suitable for scientific dating is found, the event that can be dated may not have direct significance to artifacts or archaeology. Most scientific dating methods relate to material, not manufacture: only in a few cases like thermoluminescence is the age directly related to a point in a production process generated. For wood, for example, dates produced, whether by radiocarbon or dendrochronology, can only relate to the time at which the plant tissues present were formed, not when the tree was felled or used; seasoning, storage, reuse, or even deliberate use of old materials to deceive cannot be detected.

Equally, a sample can only be used to date an associated context or material if that association is real

and direct; if a context has been disturbed after formation, or if the sample is residual, as, for example, bone from an earlier midden included in the fill of a later ditch, the use of scientific dating methods, no matter how precise and accurate the measurements, has no realistic meaning. Despite all these caveats, the introduction of scientific dating methods into archaeology has had a revolutionary effect on our perceptions of the past, has provided us with a firm chronological framework, and, particularly in the case of radiocarbon, has led to major reevaluations of previously accepted chronologies.

This article is intended as a general introduction to some of the methods used. Radiocarbon is deliberately discussed in most detail as it is probably the most frequently exploited of the techniques. Many of the more general points made in the radiocarbon section do, however, relate equally to other dating methods.

### Requirements for a Functional Dating Technique

Devising a usable dating technique requires a characteristic that alters in a known fashion with time (usually by steady increase or decrease), the ability to measure that change, a means of deriving a clean sample unaltered in other ways, and a zeroing mechanism to effectively start the clock. Not surprisingly, most dating methods are dependant in some way on naturally occurring radioactivity, although there are a number of other mechanisms that have proved usable.

Besides those methods that give a fixed time point (with or without calibration by other means), there are some 'relative' methods that, whilst not yielding a real 'date', can give information on the sequence of events.

## Radiocarbon Dating

### Theory

Radiocarbon dating is possible because of the existence in nature of the radioactive isotope  $^{14}\text{C}$  (albeit in small quantities; the vast majority of natural carbon is composed of the stable isotopes  $^{13}\text{C}$  and  $^{12}\text{C}$ ). This isotope has the advantages for the study of the human past of a conveniently long half-life (of  $\sim 5730$  years, although by convention radiocarbon results are calculated on the 'Libby' half-life of 5568, the accepted figure at the time that the radiocarbon technique was devised; the necessary correction is made as part of the calibration process described below) and of the ubiquity of carbon throughout the biosphere. The theoretical basis of the method is illustrated schematically in Figure 1.  $^{14}\text{C}$  is formed in the upper atmosphere by the action of cosmic rays on  $^{14}\text{N}$ . The resultant  $^{14}\text{C}$  rapidly oxidizes to  $^{14}\text{CO}_2$  and is swept into the general carbon cycle, mixing rapidly and thoroughly throughout the atmosphere, and via photosynthesis and the food chain into the biosphere. There is also an exchange into bodies of water, notably the deep oceans, which hold the vast majority of the world's carbon, although oceanic carbon distribution is complex. In terrestrial contexts the consumption of plant material by animals ensures the rapid spread of  $^{14}\text{C}$  throughout living organisms. During life food consumption constantly replenishes the  $^{14}\text{C}$  lost by decay. On death, however, this replacement ceases and the level of  $^{14}\text{C}$  falls following the general formula for radioactive decay given in eqn [1]:

$$A = A_0 e^{-\lambda t} \quad [1]$$

where  $A$  and  $A_0$  are the present and original isotopic quantities, respectively,  $\lambda$  is the reciprocal of the mean life (the average lifetime of a  $^{14}\text{C}$  atom), and  $t$  is the elapsed time.

If it can be assumed that the rate of  $^{14}\text{C}$  production has not varied over time, and thus that a dynamic equilibrium has formed, and if it is possible to extract clean sample carbon, unaltered apart from the decline in  $^{14}\text{C}$ , and to measure its current  $^{14}\text{C}$  concentration, it is possible using eqn [1] to calculate the elapsed time since the death of the organism. In practice, the process is far more complicated than this brief description indicates. Principally, one of the basic assumptions, that the rate of  $^{14}\text{C}$  formation is constant, is known to be incorrect. The rate has, in fact, varied over time in response to a number of effects, principally fluctuations in the cosmic-ray flux with changes in the geomagnetic field and in solar activity. Because of this, no radiocarbon measurement equates directly with a calendar date, and all such measurements must be calibrated before use.

### Radiocarbon Calibration

Unfortunately, the temporal changes in  $^{14}\text{C}$  flux have been too complex for simple modeling and it has been necessary instead to find a way of making direct measurements. Dendrochronologically dated wood (see section below for an explanation of this procedure) has proved ideal for this purpose; the tissue of an individual tree ring stops exchanging carbon within a very short time of being formed, thus

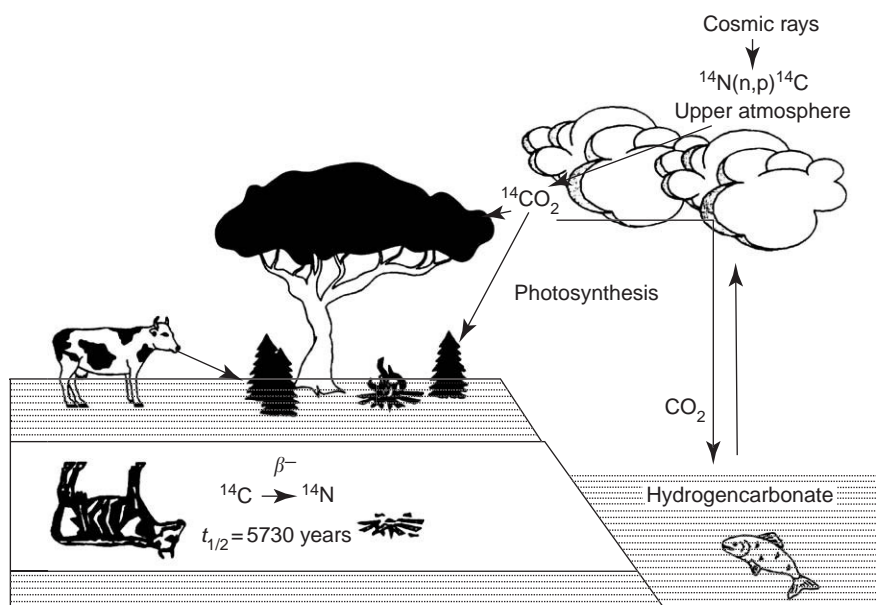


Figure 1 Schematic depiction of the  $^{14}\text{C}$  cycle.

preserving a measure of the  $^{14}\text{C}$  concentration in that year. As discussed below, this can present difficulties in dating, but it has allowed the production of long graphs (usually termed calibration curves). Currently there are internationally agreed and accepted high-precision calibration curves, based on 10- and 20-year sections of ancient oak preserved in the peat bogs of Ireland and lowland Europe, back to 10 300 BP. This is further extended back to 24 000 BP by uranium/thorium dating (see below) of corals (although with considerably larger error terms). The calibration curve has proved to have two main characteristics; an overall, and apparently sinusoidal, curve (period  $\sim 9000$  years), with a large number of short-term variations (wiggles) superimposed. These wiggles are the main bugbear in the practical usage of radiocarbon as a dating method. They mean that a single radiocarbon measurement can calibrate to a calendar year range larger or smaller than the measurement precision, depending on the slope of the calibration plot at that precise period, and even that a single result can relate to two, or more, perfectly possible calendar age ranges.

While the original radiocarbon analysis has the Gaussian probability distribution typical of any such measurements, the necessity to calibrate it against a nonmonotonic graph destroys this property, producing a probability distribution of unpredictable, but always irregular, form. As a consequence, the application of most standard mathematical techniques to the calibrated date ranges is impossible. This problem, which once made radiocarbon data extremely difficult to fully exploit, has now been largely overcome by the use of Bayesian statistics. Use of this technique has permitted radiocarbon measurements to be easily modeled, compared and combined with each other, and with results from other dating methods.

In addition to cosmogenic  $^{14}\text{C}$  variations there are also two distinct readily recognizable anthropogenic variations in  $^{14}\text{C}$  level, at the most recent end of the curve. These are the fossil fuel effect, a rapid lowering of atmospheric  $^{14}\text{C}$  levels at the end of the nineteenth century, due to the widespread burning of fossil fuels which released 'old' carbon depleted in  $^{14}\text{C}$  into the atmosphere; and the bomb effect, a massive increase in  $^{14}\text{C}$  in the 1960s following the atmospheric testing of nuclear weapons. Both have implications for radiocarbon dating, with the fossil fuel effect making precise dating of much of the last two centuries impossible. The bomb effect has had at least one positive side-effect; the resultant spike has been widely used as an isotopic marker, allowing the dynamics of the carbon cycle to be more readily explored.

## Measurement Techniques and Sample Sizes

Until the 1980s, measurement of  $^{14}\text{C}$  activity could only be made by decay counting (usually termed conventional dating), in either gas proportional or liquid scintillation counters, necessitating the use of large samples (of  $\sim 4$  g of elemental carbon, and commensurately higher weights of raw material), but more recently the use of accelerator mass spectrometry (AMS) has permitted the direct measurement of  $^{14}\text{C}$  ions and has led to a dramatic decline in sample size to 3 mg or less, thus greatly extending the range of materials which can be dated.

Since all the counting methods involve the conversion of sample carbon to suitable chemical forms (usually methane or carbon dioxide for gas counting, benzene for liquid scintillation, and graphite or carbon dioxide for AMS), the method is always destructive. Additionally, extensive cleaning of the sample is always required to remove extraneous carbon. Fortunately, most radiocarbon samples are either of wood or charcoal, both of which are remarkably stable and do not exchange carbon with their surroundings, although a prolonged residency in the soil will obviously result in an amount of intrusive material. Bone presents more of a problem as the inorganic component (mainly hydroxyapatite) is not a closed system and is liable to isotopic exchange, particularly with groundwater. For this reason only the organic (protein) component of bone, mostly present as collagen, is analyzed. The advent of AMS has allowed more specific fractions, down to individual amino acids, to be extracted and dated. The inorganic fraction of shell is subject to a similar exchange, but here there are further complications as both terrestrial and marine mollusca are capable of directly reworking geological carbonates, and marine samples are subject to the marine effect discussed below.

Chemical methods are available to cope with most natural contaminants, but the use of some modern materials, particularly of petroleum-derived coatings as consolidants, can destroy the integrity of materials as radiocarbon samples.

## Sample Materials

Although in theory it is possible to date the death of any living organism by radiocarbon, there are a number of practical limitations on the materials that can be meaningfully analyzed. Firstly, as with all dating techniques, there is the question of sheer durability; it is only possible to date those materials that have survived from antiquity. These seldom include soft tissue, or the more transitory plant fibers, and even given the extended range of sample types



allowed by the small sample requirements of AMS, ~90% of archaeological radiocarbon samples are wood, charcoal, or bone.

Equally important, but frequently overlooked, is the necessity to analyze only materials where the time of cessation of carbon exchange relates in some way to an archaeological event. The most obvious example of the dangers involved here relates to wood. As already discussed, the tissues of individual tree rings cease to exchange carbon within a few years of their formation. This has been invaluable for the production of calibration curves, but is an enormous problem in practical dating terms. The majority of archaeological recovered wood or charcoal comes from mature trees. For long-lived species, such materials can have an apparent age of several hundred years ( $1000 +$  years for European yew) even before felling. Moreover, the potential offset between the date measured by radiocarbon and the actual date of use can be huge, and more importantly unquantifiable. The only answer is to avoid the problem by identifying all woods before dating and only analyzing young roundwood or short-lived species.

Obviously there are occasions where this is impossible, particularly where the question is the date of a wooden object. All that can be hoped for is a *terminus post quem*, an oldest possible date, before which the object cannot have existed. An offset also occurs with marine materials. Here, because of the time taken for atmospheric carbon dioxide to absorb into ocean waters, and poor oceanic mixing, all marine organisms, and all animals dependant on marine sources for food (including whale and walrus, both frequently exploited for raw material), are subject to an age offset in the region of a few hundred years that varies with exact location, making it effectively unquantifiable. Hence, precise dating of the purported medieval whalebone plaque illustrated in **Figure 2** is impossible. An AMS date on a small sample measured at the Oxford laboratory gave an uncalibrated figure of  $1480 \pm 80$  BP (OxA-1164). Adding the expected marine offset of a few hundred years gives a final result in accord with the style, although this figure can only relate to the whalebone itself; the possibility still remains that old material was used for a more carving.

### Age Range

The maximum age definable by radiocarbon depends on minimum measurement levels. Using conventional techniques this was limited to around eight half-lives, extended to ~60 000 years by enrichment techniques. With the advent of AMS it was hoped to extend this to 70 000 or 80 000 years, but such figures have yet to be routinely obtained. Minimum age is complicated to define; the fossil fuel effect means that



**Figure 2** Purportedly medieval Spanish whalebone plaque. (© Trustees of the British Museum).

results for the past 200 years are virtually indistinguishable, and such results are quoted as modern by convention. However, the bomb effect makes the years immediately after 1950 easily identifiable. Precision varies both with measurement error and with the shape of the calibration curve. The so-called high-precision measurements can give results with error terms of  $\pm 18$  radiocarbon years or less, whilst results for full-sized samples measured to normal precision will give figures of  $\sim \pm 40$  years for the past 10 000 years, with the error term increasing slightly for older materials or for less than ideal sample sizes. What these error terms equate to in real calendar years is entirely dependant on the shape of the calibration curve at the period in question.

### Radiocarbon Conventions

The presentation of radiocarbon measurements has been rigidly formalized. All measurements are made relative to a 'modern' standard of oxalic acid, the activity of which has been related to the theoretical

activity of wood of AD 1950 had no fossil fuel effect occurred. Measurements are quoted in years BP, with present fixed at AD 1950, with an error term equal to  $\pm 1\sigma$ , and using the Libby half-life of 5568 years. Calibrated results are quoted as cal AD or cal BC as appropriate, with the confidence level noted. There are few such formalized rules for the formulation of results by other dating techniques, and individual publications have to be consulted for such details.

### Case Studies

Practical use of  $^{14}\text{C}$  is best illustrated with some examples. Probably the most publicized recent case has been the dating of the Turin Shroud, a length of linen bearing the shadowy imprint of a man, believed by many to be the shroud of Christ. Independent analysis of small samples of the linen of the shroud by three AMS dating laboratories produced concordant results, with the weighted mean calibrating to a possible calendar date range of 1260–1390 cal AD (at the 68% confidence level), which ties in closely with the first known historical appearance of the object in ca. AD 1350. Given the contentious nature of the material, it is not surprising that there has been some controversy about this result, but the scientific evidence certainly indicates the linen to be medieval.

More typical is a project run by English Heritage to securely date the various phases of activity at Stonehenge, a massive megalithic monument in Southern England. This site has long been acknowledged to be complex and multiperiod, with probable origins in the Neolithic, but the actual sequence of construction has been far from clear. Here, 52 radiocarbon determinations were measured to date the three phases of the site itself, plus pre- and postmonument use of the area. The dating of Phase 1, the cutting of the main ditches and banks, was made easier by the presence of large numbers of tools made from the antler of red deer, used during the construction work, in the base of the ditches themselves. Many of these came from slaughtered animals rather than being naturally shed, and can therefore be expected to date to within a few years of ditch construction. It was more difficult to obtain samples directly related to other phases (particularly the construction of the main stone circles), but nonetheless a combination of radiocarbon measurements and Bayesian statistical analysis has produced a definitive chronology for this important site.

## Luminescence Dating

### Theory

There are two forms of luminescence dating of importance to archaeology; thermoluminescence (TL)

dating and optically stimulated luminescence (OSL) dating. TL was the earliest to be widely used, although OSL is rapidly gaining in importance. Both methods are dependant on the damage done to some nonconducting crystalline materials (notably quartz and feldspar) by naturally produced ionizing radiation, although in neither case is the process fully understood. The theory behind the techniques is probably most easily explained in terms of TL.

Thermoluminescence is the light emitted in addition to the normal incandescent glow when some crystalline materials are heated. Although the process is in reality much more complex, it is most conveniently explained in terms of a simplified model involving 'traps', small imperfections in the crystal lattice. When the crystal is exposed to ionizing radiation, electrons are released. Most of these will recombine almost immediately, but some fall into traps, where they remain until sufficient energy is input to release them. The depth of the trap (i.e., the energy required to escape from it) varies, but can be substantial, equivalent to a long electron residence period at ambient temperatures. It is usually these relatively stable traps that are used for dating purposes. On release the electrons can take several paths, one of which is to recombine with an ion within the structure. If this occurs at a specific type of defect (a luminescence center) light (i.e., the TL) is emitted. Since the number of trapped electrons is proportional to the exposure of the material to ionizing radiation, and the TL is proportional to the number of electrons released, if the material has since been exposed to a measurable and constant radiation flux it is possible to relate TL directly to the time since the traps were last emptied. If the date of this clearing of the traps has some archaeological significance, it is possible to use this property as a meaningful dating technique. In the case of TL dating, emptying of the traps is caused by heating (most commonly when clay is fired to produce ceramic). After zeroing, the material will be exposed to radiation from naturally occurring uranium, thorium, and potassium ( $^{40}\text{K}$ ) present in the sample itself and its surroundings. The concentrations of these isotopes are highly variable, so that values must be determined separately for each sample and context, but all are very long-lived, and providing the material has remained in the same environment, the radiation flux can be treated as being constant. It is therefore possible to use eqn [2]:

$$\text{Age} = \frac{\text{Archaeological TL}}{\text{Annual dose} \times \text{TL per unit dose}} \quad [2]$$

where archaeological TL is the naturally accumulated TL, annual dose is the annual dose rate to which



the sample has been exposed, and TL per unit dose is the sensitivity of the sample material to that radiation.

There are many complications to the method, including the anomalous fading exhibited by some materials and the variability of response by different materials, which means that the individual sensitivity of each sample to radiation must be determined. A number of different measurement techniques, involving the use of different fractions of the sample, have therefore been developed; the measuring laboratory will select the most appropriate.

The theory underlying OSL dating is analogous to that of TL, but here the zeroing process is the exposure of fine soil particles to light. This means it is most useful when looking at sediments which have been bleached by sunlight during deposition, and which relate to an archaeological event in some way (such as the first soil deposited in a ditch).

### **Sample Materials**

In archaeological contexts TL is most commonly used to date the firing of ceramics (including the clay cores sometimes found in metal castings) and burnt stone (particularly flint). OSL is used mostly to date sediment deposits.

### **Dose Determination**

The radiation dose received from the environment is determined by onsite measurement either with a gamma-spectrometer or by burying capsules of a highly sensitive material (usually calcium fluoride) for a known period. The radiation dose from the material itself (the internal dose) is found by direct measurements of the portions of the samples. For situations where the find spot is not known, or where ceramics have been constantly used and never buried, the environmental dose rate can only be estimated. This is obviously subject to a much greater error, and such measurements are a form of authenticity testing rather than dating.

### **Age Range**

The age range for which TL can be used depends on specific characteristics of sample sensitivity and environment; for ceramics the lower limit can be a few hundred years, whilst for flint it is seldom possible to date anything younger than a few thousand years. The upper limit is also variable, depending on the point at which all the available traps are filled (saturation), and on the stability of the traps, but ages of 100 000–500 000 years should be possible. Precision depends on exact circumstances but can be as good as  $\sim \pm 5\text{--}10\%$  of the age for dating and

$\pm 25\%$  for authenticity testing, where the environmental dose-rate is estimated.

The age range for OSL is again dependant heavily on site conditions, but is basically similar, as is the precision, although new measurement methods, involving the use of lasers, can give precision as good as  $\pm 1.5\%$  for young ( $< 2000$  years) sites, and multiple sampling can allow the resolution of events on the scale of a few decades for the last 100 000 years.

### **Case Studies**

An example of the use of TL as a dating method is provided by the excavations at Pontnewydd Cave in north Wales, where despite the difficulties caused by variations in the environmental dose rate because of the nature of the deposits, TL has established an early hominid occupation  $\sim 200\,000$  years ago.

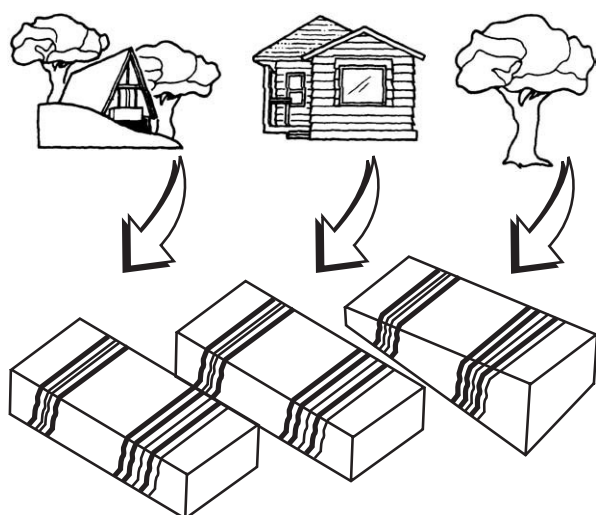
OSL has been used to at least partially solve one of the continuing questions of British archaeology; the age of the hill figures cut into the chalk hillsides of Britain. The White Horse of Uffington is one such figure, dated variously to the Iron Age, Roman occupation, or later. Excavation of the figure has proved that, rather than being manufactured by simply clearing areas of the chalk surface, trenches were dug and chalk packed into them to form the lines. Silt from the earliest of these has been dated by OSL to 1400–600 BC, placing the figure firmly in the late Bronze Age.

### **Electron Spin Resonance**

A related dating technique is electron spin resonance (ESR) where the presence of trapped electrons is detected by their response to electromagnetic radiation. ESR is particularly useful as it can be applied to a complementary range of materials, including tooth enamel and shell, which decompose on heating and are therefore not suitable for TL work.

### **Dendrochronology**

Dendrochronology, or tree-ring dating, as implied by the name, is only applicable to timber. It exploits the characteristic of certain species of tree (notably oak and pine) of producing annual rings the width of which varies with local weather and water supply conditions. These patterns of rings can be compiled by cross-matching measurements from trees of known felling date, backwards through time, with wood from other sources, such as building timbers, to build up master curves of ring width pattern against age, as illustrated in **Figure 3**. Samples of unknown date can then be dated by comparison with



**Figure 3** Schematic demonstration of the construction of a dendrochronological master curve.

this master. Such curves are normally specific to species and to region, although there has been some limited success with interspecies cross-dating.

Where material from a site with extensive timber survival cannot be tied directly to one of the existing master chronologies, it is sometimes possible to construct a 'floating' chronology that can be used as a relative dating tool. There are a few periods where the growth ring pattern is so distinctive as to produce a 'signature', almost instantly datable on the basis of a small number of rings, but for the most part dendrochronology requires large samples (90+ rings). Even given this amount of material, and a suitable local master curve, the method is not always successful, but where it is it yields a very precise date, giving not only the exact year, but even the season of growth of the outermost ring present.

In real terms, this apparent precision may not mean much; the figure can only apply to the outermost ring present, which is not necessarily the original surface of the tree. In fact, in most circumstances, the more vulnerable sapwood at least will have been trimmed before use. In building timbers it is frequently only this layer which is removed, leaving the 'waney edge', the boundary between sapwood and heartwood. There are well-defined formulae for the numbers of sapwood rings in different species, which can be applied as a correction if the waney edge is present, but if it is absent it is impossible to quantify how many years of growth have been lost.

In addition, as with radiocarbon, dendrochronology can only date the formation of the wood, not its use. A dendro date gives at best the year of felling. It can give no indication of time elapsed in

seasoning, storage, or possible repeated reuse (a frequent event with large timbers). Furthermore, any self-respecting forger will select wood of a suitable age as a raw material.

Dating is possible over the whole calendar range for which master curves are available, but the size of sample required means that the method is seldom applied to samples more than a few thousand years old.

Another important property of the long dendrochronological curves now available (see section on radiocarbon for details) is that they provide a proxy climatic record for their region; whilst not always easy to interpret, they do give direct evidence of the nature of climatic change over time for thousands of years.

## Decay Series Dating (Other than Radiocarbon)

These methods relate, in general, to the formation of geological deposits, and are seldom of direct use in archaeology. They do serve, however, as a means of defining time periods, for example, for strata trapped between lava flows, and play an important part in the understanding of early man.

### Potassium–Argon/Argon–Argon Dating

The weakly radioactive  $^{40}\text{K}$ , naturally present in most of the earth's crust, decays to two daughters,  $^{40}\text{Ar}$  and  $^{40}\text{Ca}$ , with a half-life of 1250 million years.  $^{40}\text{Ar}$  is insoluble in molten rock giving a zeroing mechanism for lava flows. After resolidification of the rock, argon gas generated can be trapped within the glassy matrix. Given a closed system, collection and measurement (by mass spectrometry) of trapped  $^{40}\text{Ar}$  can therefore be used to date the formation of the lava flow. The long half-life and minute original quantities of potassium mean that the method is fairly imprecise (with error terms ranging approximately from  $\pm 1000$  to  $\pm 10\,000$  years depending on age), but it does extend far back into geological history. A variation on this technique,  $^{40}\text{Ar}/^{39}\text{Ar}$  dating, involves the artificial generation of  $^{39}\text{Ar}$  from the stable  $^{39}\text{K}$  by neutron irradiation. Potassium concentration can then be determined by measurement of the  $^{39}\text{Ar}$ . This has the advantage of requiring only one measurement procedure, of the isotopic ratios of argon, but more importantly allows a controlled release of argon by gradual temperature increase, which can be used to detect sample contamination. In an uncontaminated sample the  $^{40}\text{Ar}/^{39}\text{Ar}$  ratio will remain a constant; any variation demonstrates a problem with the sample, due either to a partial

reheating postformation releasing some argon gas, or the presence of atmospheric argon that survived the initial melting.

### **Uranium Series Dating**

Uranium series dating utilizes the two natural uranium decay chains (from parent  $^{238}\text{U}$  and  $^{235}\text{U}$ ) and their progeny. The zeroing process here is usually the formation of calcite from carbonates (with uranium present as an impurity) carried in solution. Since the primary long-lived daughter,  $^{230}\text{Th}$ , is largely water insoluble, in any process involving dissolution and redeposition of uranium there will be effectively no thorium present in the newly formed crystal matrix, and, provided the system remains closed, the proportions of generated thorium and other daughters can be used as a clock. The method is most frequently applied to geological formations of calcite (spellotherms or travertine) but has occasionally been applied to deposition of calcite within fossil bone, although with limited success owing to variation in uranium uptake and the open nature of the system.

Age ranges and precision vary with uranium content of each sample, but in general the method can cover a period between 1000 and 100 000 years BP to error limits of  $\sim \pm 1\%$  of age.

## **Other Dating Techniques**

### **Amino Acid Racemization**

It is possible for asymmetric compounds to exist in two mirror image but nonsuperimposable forms, known as optical isomers. The amino acids that build up proteins are invariably formed in living tissue as only one form, the L-isomer. After formation, a slow racemization process begins in which the L-form is converted to the alternative D-form, until a dynamic equilibrium is reached in which the amounts of the two forms are equal. Provided that the equilibrium point has not been reached, the portion of the D-form present can be used as an indication of age. The rate of conversion is heavily dependant on a number of factors, principally temperature, humidity level, and degree of chemical breakdown of the tissue involved (diagenesis).

Amino acid dating has been applied with varying degrees of success; because the rate constant varies so much with local conditions, it can only be determined by calibration using similar materials from the same area, dated by other independent means. The assumption has to be made that both sample and calibration material have been subjected to the same conditions (for example, boiling of bone will

radically alter the rate constant for that time) and there have been some heavily publicized incorrect ages generated as a result of wrongly estimated rate constants, notably with regard to the first human occupation of the Americas. Despite these drawbacks, the method, which requires sample sizes of less than 1 g, can be useful.

The timescale that can be covered depends on the amino acid used and the environmental conditions, but ages in excess of 50 000 years are certainly theoretically possible. Precision is heavily dependant on that of the calibration results and no general figures can be quoted.

### **Obsidian Hydration**

This technique is only applicable to the volcanic glass obsidian, and depends on measuring the thickness of the opaque hydration layer that slowly develops on any fresh surface. As this is a diffusion process, the rate is heavily dependant on temperature and results must be calibrated for particular areas using dates determined by other methods, making estimates of precision impossible to generalize. Efforts have been made to apply similar procedures to surface change on other lithics and on glasses, but never with the same degree of success.

### **Archaeomagnetism**

It is well known, even if the process is not well understood, that there have been extensive temporal changes in both the strength and direction of the earth's magnetic field (including occasional total polarity reversal). Most clays and soils contain a small percentage of finely dispersed particles of iron oxide. These grains are small and frequently have a single magnetic axis. In raw clay these particles are randomly aligned, but on firing above a critical temperature (the blocking temperature) some acquire sufficient energy to adjust direction to that closest to the external magnetic field, hence producing a very small but permanent thermoremanent magnetism with an intensity proportional to that of the current geomagnetic field, and in the same direction. This property is only useful as a means of dating if the sample is found *in situ* (so that the original magnetic direction can be determined) and if there is an existing reference graph of changes in field direction and strength against time for the locality, dated by other means.

Large samples are required, in the form of blocks of baked material several cubic centimeters, but as the method is useful mainly for dating kilns and hearths this is seldom a problem. Sedimentation processes can produce a similar remnant magnetism,

although with a field strength an order of magnitude less, which can be used for dating provided there has been no postdepositional disturbance.

Precision is highly variable, depending on the rapidity of field change and on the precision of the calibration samples.

## Selection of a Suitable Dating Technique

In general, the constraints of methods and materials are such that there is little conflict about dating methods. The preferred technique will depend on the context, the materials available, and the prospective error term with each method. For many sites the ideal answer will be a combination of methods, followed by the use of Bayesian statistical techniques to analyze the results.

See also: **Electron Spin Resonance Spectroscopy:** Principles and Instrumentation. **Glasses.** **Liquid**

**Chromatography:** Amino Acids. **Mass Spectrometry:** Overview; Stable Isotope Ratio.

## Further Reading

- Aitken MJ (1990) *Science-based Dating in Archaeology*. London: Longman.
- Aitken MJ (1998) *Introduction to Optical Dating: The Dating of Quaternary Sediments by the Use of Photon-stimulated Luminescence*. Oxford: Clarendon Press.
- Baillie MGL (1982) *Tree-ring Dating and Archaeology*. London: Croom Helm.
- Bowman SGE (1990) *Radiocarbon Dating*. London: British Museum Press.
- Eighmy JL and Sternberg RS (1991) *Archaeomagnetic Dating*. Tucson: University of Arizona Press.
- Taylor RE and Aitken MJ (eds.) (1997) *Chronometric Dating in Archaeology*. New York: Plenum Press.
- Wagner GA (1998) *Age Determination of Young Rocks and Artifacts: Physical and Chemical Clocks in Quaternary Geology and Archaeology (Natural Science in Archaeology)*. Berlin: Springer-Verlag.

## Metallic and Ceramic Objects

**L T Gibson**, University of Strathclyde, Glasgow, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The application of scientific knowledge and the use of analytical techniques have been implemented in archaeometric and conservation science researches for many years. Specimens are often subjected to analytical scrutiny to help provide a fundamental understanding of the materials used in the production of ancient objects, and to assess the stability of artifacts. The goal is to inform our generation of past technologies, social history, and to ensure the continued protection of items important to our cultural heritage. Each year many hundreds (if not thousands) of research papers report the results of archaeological and conservation research that utilize modern analytical instruments. The numbers of articles increase each year, and as such it is impossible to list all areas of research in this short section. Therefore, the main aim of this article is to describe, briefly, some of the many analytical techniques currently used (1) in the archaeological study of metals and ceramics, (2) in the conservation and restoration of metals and ceramics, and (3) to identify corrosion and/or deterioration products.

## Archaeological Examination of Metallic and Ceramic Specimens

Metals and ceramics are the focus of archaeological research as they are often thought to be among the most rudimentary substances in humanity's technological and social development. There are many reasons to initiate archaeological research; however, it can be argued that the main aims of study of metallic and ceramic specimens are to determine their origin, authenticity, and function, and to explore past manufacturing processes. Perhaps the largest single reason for the study of these specimens is to ascertain their provenance (the geographic origin of the raw materials used). The data obtained from provenance studies are used to identify the raw materials used during production, to date and/or authenticate specimens, and to provide information about ancient trading patterns. The basic premise in provenance determination is that each material source can be distinguished by its chemical composition. Consequently, one of the major applications of analytical science to archaeology has been to determine the major, minor, and trace element composition of archaeological specimens. Coupled with the analysis is the need for the analytical scientist to interpret the analytical data correctly. Difficulties associated with this task should not be underestimated, as imitations

of recognized ceramics can cause subtle problems in provenance studies. Moreover, archaeometallurgists note that chemical provenancing of metal specimens can only be achieved when factors such as the addition or contamination of trace elements in metal alloys, and the high temperatures and extreme redox conditions used when processing the ores, are considered. Indeed, analytical interpretation can be further complicated if the composition of the surfaces of specimens are altered as a result of chemical weathering, corrosion, restoration or previous treatments.

## **Conservation of Metals and Ceramics**

When archaeological specimens are buried, it is not uncommon for a stable equilibrium to be established between the specimens and their surroundings. After excavation, many specimens rapidly deteriorate as a result of disturbing their equilibrium conditions. Conservators, and conservation scientists, examine artifacts to assess their stability, or to restore materials that have undergone, or are likely to undergo, a chemical or physical change. Clearly, the analytical scientist examining the specimen must have some knowledge of the types of material under study and be able to recognize known, stable specimen compositions.

After excavation and study, it is common to house ceramic and metallic specimens in museums, or to ensure appropriate environments are created to store or display specimens. Here, the goal is longevity, in other words, to safeguard specimens for future generations. This is a formidable challenge as all materials degrade naturally and at best we can only hope to prolong the lifetime of such specimens. The rate of change of materials can be slow (over hundreds of years), or indeed it can be very rapid (changes have been observed on metallic and ceramic surfaces over a matter of months). The surfaces of specimens commonly undergo significant chemical and physical alterations as a result of exposure to pollution and to changes in environmental factors such as light, temperature, and relative humidity. The synergistic effects of such factors are often very difficult to predict, and the chemical processes invoked, or exacerbated, on the surfaces of materials are difficult to elucidate. For example, a rather insidious problem has been observed when storing metals and ceramics in certain wooden cabinets or drawers. Acids and aldehydes (carbonyl pollutants) emitted from the materials used to make the furniture are known to react with the surface of ceramic and metal specimens. Since the first documented report of carbonyl pollution-induced deterioration on shells in 1899, a wide range

of multifarious, complex corrosion products have been observed on thousands of metal and calcareous specimens. Efflorescences of various compositions have been observed, and identified, on porous ceramics, and complex corrosion crusts have formed on the surface of metals such as lead, copper, and bronze. Analytical science has a tremendous role to play when attempting to understand the mechanisms of deterioration of such specimens. First, the deterioration product has to be identified before the chemical deterioration processes can be elucidated. Preventive conservation strategies are also implemented to examine the stability of materials and their surrounding environmental conditions in a bid to retard further decay.

Analytical investigations also focus on new ways of protecting susceptible materials that have not yet shown signs of degradation. The continued efforts to replicate archaeological artifacts and antiques and subject them to accelerated deterioration experiments provide useful results in the development of new treatments. Much can be learned from understanding how compositions deteriorate due to their inherent incompatibility or instability to external agents.

## **Instrumental Considerations**

As outlined above, there are a number of difficulties associated with the analysis of archaeological specimens. In summary, many specimens are:

1. unique or valuable, hence the analysis needs to be nondestructive or the sample taken from the specimen needs to be extremely small;
2. inherently inhomogeneous and the surface of the materials can differ substantially from their bulk compositions (for example, if they are glazed or decorated);
3. unstable due to removal from their burial environment; and
4. deteriorated or corroded.

As a result, a compliment of techniques will often be chosen to obtain solutions to the questions posed by archaeologists or conservators. The choice of instruments used should be assessed on a case-by-case basis, considering factors such as the amount of sample preparation required, cost and availability of instrumentation, sample size availability, complexity and sensitivity of instrumentation, and the ability to perform nondestructive testing.

The mode of analysis must also be considered. Qualitative analysis is performed to provide chemical information about the elements or compounds contained in a specimen whereas quantitative analysis

permits determination of the exact concentration of analytes in a sample. Should quantitative information be required standard reference materials should be prepared, or need to be commercially available, to calibrate instruments. Matrix matching between standards and samples should be encouraged, where possible, to prevent matrix interferences during analysis. However, this can be difficult when specimens of unknown origin and composition are examined. The need for a high level of accuracy and precision in the analytical method is a prerequisite, especially in provenance studies when sources of materials are distinguished from one another by examination of trace elements. The introduction of an artificial spread of data by using a technique with poor precision must be avoided.

Interestingly, much of the debate about the 'best' analytical technique to employ, and also the most appropriate chemometric approach to be used on the data is fruitless if the analytical scientist does not consider the data in the context of the archaeological or conservation questions to be answered. To solve complex problems in this area of research collaboration between analytical scientists, archaeologists, conservators, restorers, and art historians is crucial.

## Instrumental Techniques

It is perhaps not a surprise to learn that most modern analytical instruments have their place in archaeometric and/or conservation research. Many techniques are used extensively to study ceramic and metallic specimens or to identify pitting, weathering crusts, inclusions, efflorescence, and corrosion products on the surface of samples taken from specimens. In addition, the homogeneity of materials of mixed composition is examined, the results of previous restorations are assessed, and the major, minor, and trace element compositions of samples are recorded. A selection of instruments commonly used in archaeology and conservation research is given below. Detailed descriptions of the instruments can be found in the relevant articles in this encyclopedia.

## Visual Examination of Surface Topography

### Optical Microscopy

The optical microscope is the primary analytical technique of conservators and archaeologists to investigate surface topography, and to assess deterioration products of ceramic and metallic specimens. Although not a sophisticated analytical technique,

the optical microscope is widely available in museum and conservation laboratories. Automatic scanning interference microscopes (for example, Nomarski interference microscopy) are also used to obtain surface maps when estimations of the vertical height differences of surface features are required.

### Scanning Electron Microscopy

The use of electrons instead of a light source provides much higher magnification ( $>10\,000\times$ ) of specimens, unique imaging, and the opportunity to perform elemental analysis and phase identification. Scanning electron microscope (SEM), which provides a maximum lateral resolution of 10 nm, is used extensively to examine closely spaced topographical features of metallic and ceramic samples. The increased depth of field allows a large amount of the sample to be in focus at any one time, thus several areas of a sample can be examined without alteration of the field of view. Weathered and unaltered areas of a surface can be compared in one sample. Its main disadvantage is that samples are required for analysis and ceramic materials need to be coated to prevent charging. However, the combination of high magnification and resolution together with the large depth of field make it possible to study samples less than 0.1 mm in size.

### Environmental Scanning Electron Microscopy

When surface coating of ceramic materials is not appropriate environmental scanning electron microscope (ESEM) can be used. Whereas conventional scanning electron microscopy requires a relatively high vacuum in the specimen chamber to prevent atmospheric interference with primary or secondary electrons, an ESEM instrument uses a thin membrane to separate the specimen chamber from the vacuum system. Thus, nonconducting samples can be analyzed without charging, and hydrated specimens can be analyzed without fear of dehydration in the vacuum.

## Elemental Determination of Surfaces by Electron Beam Methods

### Energy Dispersive Spectroscopy

Elemental analysis of surfaces in SEM is performed using energy dispersive spectroscopy (EDS), which measures the energy and intensity distribution of X-ray signals generated by the electron beam striking the surface of the specimen. The elemental composition at a point, along a line, or in a defined area, can be easily determined to a high degree of precision. Elemental maps are often used to locate

areas rich in elements present in concentrations greater than 1% by weight. The electrons penetrate a depth of 0.02–1.0  $\mu\text{m}$  into the sample and so care must be taken to prepare samples appropriately if information about the bulk is required. EDS is used extensively for examination of metallic and ceramic specimens, and their associated weathering crusts or corrosion products.

### **Electron Probe Microanalysis**

Electron probe microanalysis (EPMA) (also known as wavelength-dispersive electron microprobe analysis) is used in the analysis of artifact composition; this technique is similar to X-ray fluorescence (see below) and is useful for studying small changes in composition within the body of an artifact as the penetration depth is up to 5  $\mu\text{m}$ . Excellent qualitative or quantitative determination of elements can be obtained with an atomic number greater than 5, although instruments that can detect elements only above oxygen are common in conservation laboratories. Lower detection limits (down to 100  $\text{mg kg}^{-1}$ ) than EDS can be achieved by EPMA, which can detect minor and trace elements in a sample. Standards can be used to obtain quantitative information. Samples prepared for analysis are generally flat, and the microprobe analyzer is designed to allow small amount of  $x$  and  $y$  translation rather than the full specimen manipulation that is typical of SEM (for example, full rotation and tilt). It is extremely useful for identification of elements in metallic or ceramic specimens, and to examine elements used to color, opacify, and clarify glass. Care must be taken with the analysis of glasses as the beam can heat the sample and cause the migration of alkali ions at the surface of the glass. As with EDS, the process is nondestructive for metallic samples, although nonconducting materials need to be coated and prepared as polished sections before insertion into the vacuum.

### **Auger Electron Spectrometry**

Auger electron spectrometry (AES) is not used as widely as SEM–EDS or EPMA. It has a limited use in the field of archaeology and conservation, possibly due to the limited penetration depth of the sample (less than 100  $\text{\AA}$  is not unusual), although research has been successfully undertaken with glass specimens. The sample (maximum size  $\sim 18 \text{ mm} \times 12 \text{ mm}$ ) can be scanned to survey the elemental composition of the analyzed surface, or to quantify elements present to detection limits of 0.1%. The lateral distribution of elements on the surface can be resolved to 0.3  $\mu\text{m}$ .

## **Elemental Determination of Surfaces by X-Ray Methods**

### **X-Ray Fluorescence Spectrometry**

The composition of a specimen is often determined by X-ray fluorescence (XRF) spectrometry, which performs rapid, qualitative, and semiquantitative determination of major and minor surface elements. Although both wavelength- and energy-dispersive (ED) analyzers can be used to detect the secondary X-rays, ED-XRF instruments are more common for the compositional determination of archaeological and conservation samples. Detection limits of  $\sim 0.1\%$  are expected; therefore, the analysis is difficult for trace elements. A laboratory XRF system, commonly used to quantify elements in metal and ceramic samples (noninsulating materials need to be coated), is considered to be an indispensable tool. As with all these surface analytical techniques, care has to be taken that weathering products (thick patinas or corrosion crusts) do not obscure bulk analysis results. Thus, samples are normally prepared to provide a flat polished surface to produce quantitative results.

Whole objects can be analyzed at ambient conditions using either a radioactive source (used for field investigations) or an X-ray tube. The advantage of the radioactive source is that it reduces the overall weight of the instrument, but the sensitivity is relatively poor due to the low photon output of the source. Legislation also restricts the transport of radioactive sources. In contrast, an instrument utilizing an X-ray tube as the source has found much greater use in conservation as the sensitivity is increased by an order of magnitude. This set up is commonly used for the analysis of whole archaeological objects, although care must be taken as radiation burns have been observed, and lead-containing glass has been known to discolor.

### **X-Ray Photoelectron Spectrometry**

X-ray photoelectron spectrometry (XPS) is used to determine major and minor element compositions of metallic and ceramic surfaces. It can also be used to determine the oxidation states of ions on the surface of a sample. However, it has a limited depth penetration of 2–20 atomic layers and so the measurements taken will be greatly influenced by the method of sample preparation. The ability to characterize elements with atomic numbers less than 10, coupled with the ability to analyze samples smaller than 1.5  $\text{cm}^2$ , make this technique particularly useful for the colorants and clarifying agents used in glasses and ceramics. For example, it has been used to

identify the oxidation states of copper on blue color Egyptian glass.

## Elemental Determination of Surfaces and Bulk Materials by Particle Methods

### Particle-Induced X-Ray Emission

Particle-induced X-ray emission (PIXE) is a technique that induces characteristic XRF by bombarding the surface of the sample with photons or helium ions. As a result of its low detection limits (between 1 and 100 ppm) for samples weights of a few milligrams, and the higher sensitivity obtained compared to XRF, PIXE is used to detect trace elements as well as major and minor elements. Thus, it is a technique of great importance in provenance studies. Indeed, it is often described as one of the most powerful methods for identifying the technology processes in archaeometallurgical investigations. Normally, samples are placed in a vacuum chamber and most instruments can accommodate samples sizes up to 3 mm<sup>2</sup>. Further advantages of this technique are that the instrument can be modified, separating the vacuum system from the sample chamber, permitting examination of specimens without sampling or charging. The penetration depth is ~10 µm, and as such it can indicate the bulk sample composition if surfaces are polished to remove tarnish, patinas, or weathering crusts prior to analysis. However, ion beams produced by accelerators in PIXE are expensive, and thus there are limited instruments available in conservation laboratories.

### Neutron Activation Analysis

Neutron activation analysis (NAA) is an important technique for performing both qualitative and quantitative multielement analysis of major, minor, and trace elements in the surfaces and interiors of a sample. For many elements NAA offers sensitivities that are superior to those attainable by other methods, on the order of parts per billion or better. In this technique neutrons interact with the target nucleus via a nonelastic collision, and a compound nucleus forms in an excited state. The excitation energy of the compound nucleus, due to the binding energy of the neutron with the nucleus, will be lost almost instantaneously as it de-excites to a more stable configuration through emission of one or more characteristic prompt gamma rays. In many cases, this new configuration yields a radioactive nucleus that also de-excites (or decays) by emission of one or more characteristic delayed gamma rays, but at a

much slower rate according to the unique half-life of the radioactive nucleus. The process is nondestructive (although the tube placed into the reactor only has a diameter of ~5 cm), if radiation times are chosen carefully so that samples do not remain radioactive for long periods of time. Further advantages of this technique are that minimal sample preparation is required and very small samples can be examined to obtain low detection limits. One disadvantage is perhaps accessibility as nuclear research reactors are required. Despite this, NAA is frequently applied to study the provenance, colorants, and opacifying agents in glass and ceramics, and metal specimens.

### Secondary Ion Mass Spectrometry

The ultrahigh vacuum technique of secondary ion mass spectrometry (SIMS) is the most sensitive of all the commonly employed surface analytical techniques. There are a number of variants of the technique; for example, static SIMS is used to examine submonolayer elemental analysis, dynamic SIMS is used to obtain compositional information as a function of depth below the surface, and imaging SIMS is used for spatially resolved elemental analysis. Specimens commonly examined by SIMS are ~2.5 cm in diameter and 1 cm thick. When ceramic materials are examined, sample charging is prevented by flooding the surface of the sample with an electron beam. Dynamic SIMS is of great interest and has been used to study changes in elemental composition from weathered or corroded surfaces of samples into the bulk matrix.

## Compositional Analysis of the Bulk Using Powdered Samples or Solutions

### X-Ray Diffraction

X-ray diffraction (XRD) is frequently used to study the crystalline phases present in powdered ceramic samples, or to examine their weathering products, efflorescence samples on the surface of deteriorated ceramics, and metallic corrosion products. The characteristic patterns produced are identified by comparison with those held in a database of known inorganic and organic phases. Although a wide range of diffractometers are available in museum conservation laboratories from the Debye-Scherrer camera (for single crystal analysis) to diffractometers that allow analyses of whole objects, most instruments used in archaeological or conservation labs are designed to analyze small samples of powders (destructive analysis).



### Atomic Absorption Spectrometry

Atomic absorption spectrometry (AAS) (in combination with XRF) has been used to examine a variety of metal alloys including European Medieval brass objects, glass specimens (Egyptian, Renaissance Venetian, and medieval Scottish cathedral glass) and to determine the composition of corrosion crusts on metal surfaces. It is also used to track changes in the composition of specimens after accelerated aging experiments. The specimens need to be drilled into the bulk to produce sample weights between 0.015 and 0.5 g. Samples are digested to produce solutions that are analyzed to determine the concentration of metal ions in solution. The technique offers excellent detection limits permitting trace element analysis ( $\sim 0.1$ – $1$  ng ml<sup>-1</sup> for a furnace and  $10$  ng ml<sup>-1</sup>– $1$   $\mu$ g ml<sup>-1</sup> for a flame atomizer). However, a disadvantage of the system is that elements are detected sequentially as the light source used is changed for each element. Thus, it is not suitable for screening tests, and some prior knowledge of the elements expected is desirable.

### Inductively Coupled Plasma-Atomic Emission Spectrometry

The sample being analyzed is introduced into the plasma as a fine droplet aerosol. Light from the different elements is separated into different wavelengths by means of a grating and is captured by light-sensitive detectors. This permits simultaneous analysis of up to 40 elements and inductively coupled plasma-atomic emission spectrometry (ICP-AES) is consequently a multielement technique. In terms of sensitivity, ICP-AES is generally comparable to flame atomic absorption, for digestion of sample sizes between 0.1 and 0.5 g. The disadvantage is similar to AAS in that it is a destructive technique; however, it has been extensively used for the determination of deterioration products removed from the surface of specimens, and to assess the effect of accelerated aging experiments. To reduce detection limits further, and for the analysis of complex unknowns ICP-mass spectrometry (MS) is used. Laser ablation ICP-MS has been used for the direct sampling of metals. The sample preparation required is reduced significantly; however, the penetration depth is only 10  $\mu$ m, so care has to be taken not to misinterpret the results.

### Lead Isotope Ratio Analysis

The ratios of the four lead isotopes ( $\text{Pb}^{204}$ ,  $\text{Pb}^{206}$ ,  $\text{Pb}^{207}$ , and  $\text{Pb}^{208}$ ) in a lead ore specimen will depend upon the geological age of deposition of the ore. Thus, the lead isotope ratios are used to source the deposit used to produce archaeological objects containing lead. Destructive sampling of the specimen is

necessary, although specimen sizes can be as small as 1 mm<sup>3</sup>. The sample preparation is lengthy and involves the initial dissolution of samples in hydrofluoric and perchloric acids and dissolution of residual solids in dilute nitric acid. The lead is then precipitated from the nitric acid as lead sulfide, which is evaporated onto a filament coated with saturated boric acid, and subsequently heated to produce a fused boric oxide/lead sulfide bead. The concentration of the lead isotopes in the specimen are then determined using ICP-MS. The technique has been applied to study lead alloys, and ancient glass materials with high lead oxide contents such as red and yellow opaque glasses.

### Ion Chromatography

Water-soluble salts in stone, ceramics, and pottery are commonly assessed using ion chromatography (IC). The technique has also been used extensively to determine the composition of efflorescence salts on the surface of deteriorated ceramics and glass specimens, and to investigate the change in composition of specimens after accelerated aging experiments. Sample quantities range from 1 mg to a few grams, depending on the concentration of soluble ions in the sample. Detection limits are in the low parts per million range for most anions and cations (better detection limits can be achieved if ultraclean water sources are used to prepare the eluent, standards, and samples). Thus, the technique is commonly used to study differences in trace element compositions.

*See also:* **Activation Analysis:** Neutron Activation. **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Mass Spectrometry:** Archaeological Applications. **Microscopy Techniques:** Scanning Electron Microscopy. **Surface Analysis:** X-Ray Photoelectron Spectroscopy; Particle-Induced X-Ray Emission; Auger Electron Spectroscopy. **X-Ray Absorption and Diffraction:** X-Ray Diffraction – Powder. **X-Ray Fluorescence and Emission:** X-Ray Fluorescence Theory.

### Further Reading

- Bradley S (ed.) (1997) *The Interface between Science and Conservation*. British Museum Occasional Paper No. 116. London: The British Museum.
- Brommelle NS and Thomson G (eds.) (1982) *Science and Technology in the Service of Conservation*. Preprints of the Contributions to the Washington Congress, 3–9 September 1982, The International Institute for Conservation of Historic and Artistic Works, London.
- Demortier G (1991) Review of the recent applications of high energy microprobes in art and archaeology. *Nuclear*

- Instruments & Methods in Physics Research. Section B, Beam Interactions with Materials and Atoms* 54: 334–345.
- Ferrenti M (1993) *Scientific Investigations of Works of Art*. Rome: ICCROM.
- Hatchfield PB (ed.) (2002) *Pollutants in the Museum Environment. Practical Strategies for Problem Solving in Design, Exhibition, and Storage*. London: Archetype Publications.
- Lambert JB (ed.) (1984) *Archaeological Chemistry III*. Advances in Chemistry Series 205. Washington DC: American Chemical Society.
- Leute U (1990) Spectrometry and archaeometry. *Spectrochimica Acta Reviews* 13(2): 167–190.
- Mosk JA and Tennent NH (eds.) (2002) *Contributions to Conservation, Research in Conservation at the Netherlands Institute for Cultural Heritage (ICM)*. London: James and James (Science Publishers) Ltd.
- Pollard AM and Heron C (eds.) (1996) *Archaeological Chemistry*. Cambridge: The Royal Society of Chemistry.
- Roy A and Smith P (eds.) (1994) *Preventive Conservation Practice, Theory, and Research*. Preprints of the Contributions to the Ottawa Congress, 12–16 September, The International Institute for Conservation of Historic and Artistic Works, London.
- Tennent NH (ed.) (1993) *Conservation Science in the U.K.* Preprints of the meeting held in Glasgow, May 1993. London: James and James (Science Publishers) Ltd.
- Tennent NH (ed.) (1999) *The Conservation of Glass and Ceramics. Research, Practice, and Training*. London: James and James (Science Publishers) Ltd.
- Virginia M (ed.) (1996) *Archaeological Chemistry: Organic, Inorganic and Biochemical Analysis*. ACS Symposium Series 625. Orna: American Chemical Society.

## Art and Conservation

**L T Gibson**, University of Strathclyde, Glasgow, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Examination, conservation, and restoration of paintings (easel painting, wall paintings, and murals) have been undertaken almost as long as painting has been practiced. Since early articles reported the result of analyses of artists' materials in the late eighteenth century, the subject has emerged as a serious scientific area of research. It is perhaps fair to comment that the expansion of knowledge of artists' materials correlates with the development of improved analytical instrumentation capable of determining the composition of materials on a microscopic level with accuracy and precision. As the level of sophistication of instruments, and data manipulation techniques, increase so does our interpretation of analytical data providing a greater understanding of the materials used, their stabilities, and past technologies used to produce artworks. The main emphasis of this article is to discuss, briefly, many of the analytical instruments involved in the conservation of art works such as easel and wall paintings, murals, and illustrated manuscripts. This short article is not intended to be an exhaustive list of techniques used, but to indicate the great importance of analytical science in this area of research. Thus, in addition to the introductory sections, the article contains five short sections dedicated to (1) microchemical tests, (2) optical investigations, (3) the identification

of traditional and synthetic pigments, (4) identification of resins, waxes, and glues, and (5) the use of lasers in modern day studies of art. Further reading is recommended to explore the many detailed case studies that employ specific analytical techniques.

### Why Do We Study Works of Art?

Many studies implemented focus on identification of pigments and binding media used in an artwork, or they study the artwork as a whole. It is often argued that the most important reason for identification of pigments is for the purpose of restoration. Where possible, original pigments (or a combination of safer alternative pigments) should be used to in-paint damaged areas of a painting. However, the types of pigments used in paintings also provide information to:

1. establish its origin, thus aiding provenance studies;
2. detect forgeries (for example, the year in which a synthetic pigment was synthesized can act as a marker, or the pigments identified are assessed to see whether or not they would have been used on a work of art by a particular artist at a certain period of time);
3. identify which pigments should be used in conservation of a painting;
4. assess the stability of the painting should it experience increased exposure to light, heat, relative humidity, or pollution (for example, inorganic

pigments are generally more photostable and can be exposed to a higher light lux than organic pigments, which are prone to photochemical degradation);

5. understand the development of artistic styles.

Organic materials used as painting media (such as drying oils, waxes, gums, resins, egg, milk, and animal glues) are also subject to analytical scrutiny. Before the development of synthetic pigments artists' techniques differed more in the binding media than in the pigments used, thus examination of binding media has considerable importance for the study of paintings and their associated techniques. Identification of the exact organic component used is difficult even when the artwork is in a pristine condition. The analysis is made even more difficult when the materials examined have undergone some form of deterioration over time. In addition, artists frequently experimented with different, less common, formulations and additives. For example, in the late eighteenth and early nineteenth centuries, terpenoids (balsams, resins, turpentine), prepolymerized plant oils, and waxes were commonly added to modify oil paint. Many of these oil paintings are now showing evidence of failure such as shrinkage, darkening, and flaking. Identification of the presence of additives that may have exacerbated the deterioration in small paint samples is a difficult analytical task. For these reasons identification of binding media is one of the most difficult questions addressed by analytical scientists working in this area of research.

Preventive conservation of the full artwork is becoming increasingly more important in recent years. Unfortunately, damage to artwork is common and forms of deterioration include structural damage such as detachment and loss of paint layers, distortions in the canvas, and structural decay of the painting's support. Chemical changes can cause discoloration of paint and varnish layers, and changes in the chemical composition of the support. The changes can occur naturally over time or they can be invoked or exacerbated by changing environmental factors such as temperature, relative humidity, light, and pollution. The full painting is studied to determine whether or not the artwork will be stable under certain environmental conditions imposed during storage or display. Often, test panels are prepared, dried under normal and accelerated aging conditions, and the materials are studied before and after aging to help predict the future stability of materials, or to select appropriate conditions to protect the artwork for years to come.

## **What Are the Analytical Considerations?**

There are different categories of examination that must be considered in any conservation or restoration treatment. First, information is gathered about the history surrounding the artwork from the materials and techniques used in its construction, to previous conservation or restoration treatments. The analytical procedure is initiated at the next step, which involves the technical examination of the artwork. One of the main considerations is in the selection of appropriate instrumentation. The choice of instruments used in the study will be based upon consideration of factors such as the ability to perform nondestructive analysis, the instrument's detection limits, the sensitivity of the response, sampling requirements (including the preparations involved), availability of instrumentation, and cost. To date, a large range of equipment has been used to examine works of art and often more than one technique is used to solve a particular problem. Knowledge of the limitations of each instrument used is imperative as, for example, the analytical scientist must be able to distinguish between the absence of an expected analyte due to loss or deterioration rather than detection limitations. It is equally important to realize that the selection criteria must also consider the form of analytical data required to answer the carefully derived, relevant problem.

Sampling is a controversial topic when working with valuable works of art, or where extremely small amounts of a sample are taken from a work of art. Nondestructive methods of analyses are usually the goal of the restorer or conservator. However, even when artworks can be examined noninvasively there are handling, storage, transport, and security issues to be considered if the artwork needs to be removed from its location and taken to a laboratory for analysis. For these reasons, where possible, samples are removed *in situ* from artworks to carry out the identification. In general, sampling is often permitted for paintings (microscopic samples are removed from damaged areas); however, it is almost never permitted for manuscripts owing to their fragility. Finally, it is important to ensure that the microscopic sample taken from the artwork is representative of the materials used in the painting, and that the samples are retained with the conservation documentation.

## **Microchemical Tests**

In conservation laboratories microchemical tests are used to identify metal ions in pigments. Before the test is applied it is necessary to remove insoluble

organic resins that have encapsulated the pigment by wet-ashing the sample with hot perchloric acid. If the residue is not water soluble at this stage a drop of aqua regia is added and the wet-ash stage is repeated. Metal ions present in the digested samples can then be determined. For example, lead ions precipitate as thin yellow hexagonal plates when a very dilute drop of aqueous potassium iodide is mixed with a very dilute solution containing lead ions. For those metal-containing pigments that do not dissolve in either perchloric acid or aqua regia (for example, chromium oxide, cobalt blue, quartz, smalt, titanium white), insolubility is in itself an indication of the pigment used, although further tests are required for conclusive identification. A wide range of screening tests are also available for the study of oil or proteinaceous binding material in paint cross-sections.

### Optical Examination of Pigments

Because of its low cost and widespread availability, polarized light microscopy (PLM) is one of the most common and useful instrumental aids for the identification of pigments from detached paint samples, their cross-sections, or dispersions. A number of characteristics are used in combination to identify pigments including shape, size, transparency, association, homogeneity, color, pleochroism, refractive index, birefringence, sign of elongation, and extinction. The technique is particularly useful for identifying amorphous substances such as smalt, cobalt blue, Van Dyke brown, and charcoal, and to differentiate between pigments that have the same composition but differ in source or processing, e.g., whiting as chalk, limestone, or precipitated calcium carbonate. It should be emphasized, however, that correct answers will only be obtained after the analytical scientist masters the skill of PLM.

### The Use of Modern Analytical Equipment to Identify Traditional and Synthetic Pigments

A wide range of analytical techniques is necessary to provide an unambiguous identification of pigments in a sample. Elemental techniques are often used, such as scanning electron microscopy (SEM) with energy-dispersive spectroscopy (EDS), X-ray fluorescence (XRF) spectrometry, scanning electron microprobe analysis (EPMA), X-ray photoelectron spectroscopy (XPS), particle-induced X-ray emission (PIXE), neutron activation analysis (NAA), atomic absorption spectrometry (AAS), inductively coupled

plasma-atomic emission spectrometry (ICP-AES), and inductively coupled plasma-mass spectrometry (ICP-MS). Further information on the use of these techniques in archaeological or conservation research is given elsewhere. Often, more than one technique is required to confirm the range of pigments contained in any one sample, for example, where inorganic pigments can be dissolved they are examined using AAS, ICP-AES, or ICP-MS. Solid samples are more commonly examined by techniques such as XRF, PIXE, or SEM-EDS. Although less common, NAA has been used to examine underlying paint layers of a sample. Immediately after irradiation a series of photographic films are placed in contact with the surface of a painting. Upon development the films show the distribution of pigments that contain radioactive materials at the time of exposure. Different autoradiographs are obtained depending on the decay times of the radioactive elements in the sample. This form of analysis is used to study elements in various pigments used in the seventeenth century such as Mn, Cu, Na, As, P, Hg, Co.

Other techniques are used to identify pigments based on their molecular fragments. These techniques include UV-visible reflectance spectroscopy, infrared (IR) spectroscopy, mass spectrometry, and X-ray diffraction (XRD). Some thermal methods exploit differences in physical characteristics that are sensitive to the bulk properties of a paint sample. Differential scanning calorimetry (DSC) and thermal gravimetric analysis (TGA) are often used. Finally, separation techniques such as liquid chromatography (LC) have been used to identify a wide range of dyes.

Overall, it is important that the chosen techniques produce informative data from microscopic samples, even a single grain of pigment if necessary, and that signals from other materials do not interfere during the analysis. The choice of which technique to use depends on whether or not it can be applied *in situ*, its specificity, sensitivity, spatial resolution, and immunity to interference. For example, IR is highly specific and particularly useful for organic pigments, but it is prone to interference from binders and supports. SEM-EDS provides good spatial resolution but is not normally applied *in situ*, or without coating the pigment with a good electrical conductor. XRF is not as useful for organic pigments as they contain light elements. Noncrystalline pigments such as van dyke brown, bitumen, lamp black, and smalt are amorphous and cannot be analyzed by XRD. The experiences and knowledge of both the analytical and conservation scientists are therefore important when deciding on the most appropriate analytical technique to employ.

## Separation Techniques for the Analysis of Organic Binding Media

Materials used as ancient painting media are complex mixtures of organic compounds including resins, oils, waxes, animal glues, egg, and milk employed alone or in combination. The molecular structures of many of these individual components invariably become even more complex with age, as a result of oxidative changes, the joining of molecular units by polymerization, and the cross-linking and molecular scission of some components. As the identification of binding media in paintings is one of the most difficult questions addressed by analytical scientists, the procedures used to elucidate the composition of such mixtures must include separation steps. Thin-layer chromatography is used to identify oils, resins, and waxes using silica stationary phases. For oils (such as linseed, walnut, and poppy seed) and egg yolk, a liquid phase containing hexane and diethyl ether is used to separate the analytes in the sample and the spots are developed using iodine vapor. For the study of resins such as shellac, copal, and seedlac, chloroform is used as the liquid phase and spots are developed with 20% antimony pentachloride in chloroform. A 9:2:1 mixture of chloroform–diethylether–formic acid solution is used to resolve waxes such as beeswax and paraffin wax after development of the plate using 20% antimony pentachloride in chloroform.

Analysis of oils and proteinaceous binding materials are often performed by gas chromatography (GC) with flame ionization detection. Before analysis the oils are first transformed into their corresponding free fatty acids, and the proteins into their amino acids. The hydrolyzed samples are then derivatized to produce volatile products before they are injected into the GC instrument. With its lower detection limits, higher sensitivity, and ability to identify unknowns, GC–mass spectrometry (MS) is the most widely accepted technique for the analysis of traditional binding media in oil paintings. It is also used for the determination of modern synthetic binders that contain high molecular weight polymeric materials such as polyvinyl acetate; thermally labile fragments are introduced to the system via a pyrolysis analyzer. Considerable success has also been achieved for the analysis of proteinaceous media (collagen, animal, and fish glues) by LC after hydrolysis of the samples and derivatization with phenyl isothiocyanate. For example, using this method high quantities of hydroxyproline are used to indicate the presence of collagen, animal or fish glues, whereas trace quantities of this analyte indicate the presence of egg white or egg yolk.

To assess the stability of, and the effect of changing environmental factors on, binding media reconstructed paint formulations are often subjected to accelerated aging tests. Chemical and physical changes in the components are studied to help determine how, for example, the mechanisms of oxidation and cross-linking of an oil paint are affected by increased temperature and humidity over time. Changes in the molecular weight distribution of additives have been assessed using size-exclusion chromatography with refractive index and ultraviolet detection. To assess subtle changes in composition it is common to use more sophisticated (often coupled) techniques. For example, direct temperature resolved mass spectrometry and electrospray Fourier transform MS have been used to examine chemical changes in the triglyceride composition of oils as a result of oxygenation and cross-linking. Thermally assisted hydrolysis and methylation pyrolysis GC–MS has been successfully used for the analysis of traditional oils and natural resins, a technique that is not only very sensitive but also involves minimal sample preparation.

## Pigment and Media Analysis Using Laser-Based Technologies

Laser-based analytical techniques – such as laser induced fluorescence (LIF), laser induced breakdown spectroscopy (LIBS), and Raman spectroscopy – are also used for the identification of pigments, binding media and varnishes. In all techniques a low-intensity laser is used to excite an extremely small area of a paint sample. The signals generated provide information that can be directly related to reference samples of the molecular structure of pigments, inorganic and organic binding media.

Using LIF both inorganic and organic fluorescent species can be identified by interpretation of the broadband fluorescence spectra they produce. The spectrum obtained is characteristic of the species under examination. However, one limitation of LIF is the analysis of pigments that have low fluorescence quantum yields, they are almost impossible to detect especially in the presence of fluorescent impurities.

LIBS is strongly recommended for elemental analysis of cross-sections as the majority of encrustations and overpaint layers, as well as pigments, contain metals. One of the main advantages of this technique is that extremely small samples can be examined (0.1 nm) compared to traditional methods of analysis by Fourier transform infrared where a few milligrams are needed, even with a diamond cell instrument. No further sample preparation is required as

the sample is ablated creating a microscopic crater only several tenths of a microgram in diameter and 0.1  $\mu\text{m}$  in depth. Thus, the technique is normally attributed as noninvasive. LIBS offers exceptional selectivity as a result of sharp line emission spectra. The individual frequencies and relative proportions of the atomic emission lines provide characteristic fingerprints for most elements providing unambiguous identification. Although there may be interference from background emission from the organic matrix in which the pigment is dispersed.

Raman microscopy has become a powerful tool in the investigation of pigments and organic media in paintings. The high spatial resolution of the microscope makes it possible to examine individual grains of pigment in a sample. Fluorescence from binders can be minimized using an aperture at the secondary focus of the scattered radiation. If the pigment itself fluoresces the excitation wavelength of the laser can be altered to a lower frequency to significantly reduce the fluorescence. On excitation of the sample the inelastic, Raman, scattering gives rise to many bands, which are known collectively as a Raman spectrum. The spectra are highly specific in wavenumber, intensity, and bandwidth to the sample, and thus provide a unique fingerprint of the sample. In a typical setup a microscope is coupled to a spectrometer with a sensitive diode array or CCD detector. There are some limitations; for example, certain organic pigments are photosensitive, or fail to yield a Raman spectrum owing to their small particle size and/or high degree of dilution (such as a lake). The nondestructive nature of Raman microscopy, together with the possibility of *in situ*, highly specific analysis, makes it especially useful for analysis of illustrated manuscripts where sampling is not normally permitted. Moreover, illustrators, even in medieval times, preferred to use inorganic pigments since these colors were known to be less fugitive.

See also: **Activation Analysis:** Neutron Activation. **Archaeometry and Antique Analysis:** Metallic and Ceramic Objects; Organic and Biological Materials. **Mass Spectrometry:** Archaeological Applications. **Paints:** Water-Based; Organic Solvent-Based.

## Further Reading

- Bacci M *et al.* (1997) Non-destructive spectroscopic techniques for monitoring the ageing effects on works of art. *European Cultural Heritage Newsletter on Research* 10: 145–146.
- Brommelle NS and Thomson G (eds.) (1982) *Science and Technology in the Service of Conservation*. Preprints of the Contributions to the Washington Congress, 3–9 September 1982, The International Institute for Conservation of Historic and Artistic Works, London.
- Clark RJH (1998) An arts/science interface: medieval manuscripts pigments and spectroscopy. *Proceedings of the Royal Institution of Great Britain* 69.
- Cooper M (ed.) (1993) *Laser Cleaning in Conservation: An Introduction*. Oxford: Butterworth Hienemann.
- Ferretti M (1993) *Scientific Investigations of Works of Art*. Rome: ICCROM.
- FitzHugh EW (ed.) (1997) *Artists' Pigments, a Handbook of their History and Characteristics*, vol. 3. London: Oxford University Press.
- Lahanier C (1991) Scientific methods applied to the study of art objects. *Mikrochimica Acta* 2: 245–254.
- Malmqvist KG (1995) Applications in art and archaeology. *Chemical Analysis: A Series of Monographs on Analytical Chemistry and its Applications* 133: 367–417.
- McCrone W, Chartier DR, and Weiss RJ (eds.) (1998) Scientific detection of fakery in art. International Society for Optical Engineering, *Proceedings of SPIE*, 29–30 January, San Jose, California.
- McCrone WC (1997) The microscopical identification of artists' pigments. *Journal of the International Institute for Conservation – Canadian Group* 7(1 & 2).
- Spoto G (2000) Secondary ion mass spectrometry in art and archaeology. *Thermochimica Acta* 365(1–2): 157–166.
- Spoto G, Torrisi A, and Contino A (2000) Probing archaeological and artistic solid materials by spatially resolved analytical techniques. *Chemical Society Reviews* 29(6): 429–439.
- Striegel MF and Hill J (1994) *Methods in Scientific Examination of Works of Art, Volume 1, Thin-Layer Chromatography Course Book*. California: The Getty Conservation Institute.
- Striegel MF and Hill J (1994) *Methods in Scientific Examination of Works of Art, Volume 2, Thin-Layer Chromatography Protocols*. California: The Getty Conservation Institute.

## Organic and Biological Materials

**M Richards**, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

This article concerns the application of analytical biochemical techniques to preserved organic remains

**Table 1** Summary of the main methods used in the identification and characterization of a selection of ancient organic materials

<i>Material</i>	<i>Methods used</i>	<i>Purpose</i>
Bone and teeth	aDNA	Identification of species of origin, phylogenetic analysis, sex determination, pathogen identification
Bone and teeth	Stable isotope analysis	Determining diet and place of origin
Parchment	aDNA	Identification of species of origin
Parchment	Immunochemistry	Identification of species of origin of proteins
Pottery residues	GC/MS, GC–IR–MS, Pyr–GC–MS	Identification of lipid origin, especially plant and animal fats
Pottery residues	Immunochemistry	Identification of species of origin of proteins
Stone tool residues	Immunochemistry	Identification of species of origin of proteins
Resins	GC/MS, GC–C–IR–MS, Pyr–GC–MS	Identification of plant of origin
Soil lipids	GC/MS, GC–C–IR–MS, Pyr–GC–MS	Identification of lipid origin, especially manure

recovered from archaeological or historical contexts. The preserved material is often greatly altered compared to its modern counterparts; therefore, modified extraction procedures are required, as are a number of tests to determine the degree of preservation. This research area has grown rapidly in recent years, driven by advances in technology, especially in the ability to measure very small samples, which is necessary as many of these methods are destructive, and the material of interest is often rare. The goal of research in this area is to both correctly identify and isolate biomolecules of interest, and then derive information of archaeological or historical interest from them (Table 1). Because of this focus, this area of archaeological science research is often termed ‘biomolecular archaeology’.

## Ancient DNA (aDNA)

DNA has been successfully recovered from a range of ancient organic materials, including tissue, bone, and teeth. The methods used are derived from standard biomolecular techniques of DNA extraction, amplification, cloning, and sequencing, but are modified to account for the degraded state of the DNA. The goal of this area of research is twofold; to first recover intact DNA, and to then sequence the DNA. The may be done to simply determine biological species, or it can be used to determine the phylogenetic affinities of ancient samples.

### aDNA Application: Human Bone and Teeth

aDNA extracted from human bone and teeth can potentially be used to determine the sex of an individual, provide phylogenetic information, indicate familial relationships, and identify the presence of disease pathogens. However, much of this potential is yet to be realized as this area has been greatly hindered by the problem of contamination by modern human and pathogen DNA. aDNA is often present (if at all) in significantly smaller quantities

than in modern bone and teeth, and is often fragmentary. As there is so little aDNA the sequences need to be amplified using polymerase chain reaction (PCR), which is where the errors can occur. PCR is designed for modern DNA, and therefore can better amplify modern intact molecules than fragmentary ancient ones. Contamination of samples can occur in both the field and the museum, as modern DNA can be introduced to the bone or tooth through handling the samples, or during cleaning of the samples. Contamination in laboratories is also a problem, although most modern laboratories have adequate procedures in place to limit this.

Researchers have endeavored to extract and amplify mitochondrial DNA (mtDNA) as well as sections of nuclear DNA from ancient samples. mtDNA is inherited maternally, so the sequence can show maternal lineage, and modern (living) human mtDNA sequences have been used to attempt to reconstruct the genetic history of Europe, focusing on whether extant peoples are descendents of Palaeolithic or Neolithic peoples, or even later migrations. This area of research is problematic even with modern humans, and with the problems of contamination it is nearly impossible with ancient samples. Recent research has shown, for example, that damage to the DNA can cause changes in the sequences that mimic other mtDNA sequences, so for example, a ‘European’ mtDNA sequence could be altered upon burial to resemble a ‘Near Eastern’ sequence. The great success of this research has been the successful extraction and sequencing of a number of Neanderthal mtDNA sequences, which showed that Neanderthals had significantly different mtDNA sequences than all of the living humans currently sequenced, and were therefore our ‘cousins’ and not direct ancestors.

Nuclear DNA sequences could provide significant information on phylogeny, familial relationships, and genetic disorders, but aDNA is often so fragmentary that it is difficult to sequence. Also, researchers are currently working on understanding the modern

human genome, so we cannot hope to understand the functions of past gene sequences until we understand modern ones. Some work has been undertaken on trying to use DNA to sex individuals, looking for the presence of the Y chromosome to indicate a male sequence. Again, contamination is a very significant problem here and this method is controversial.

A third area of aDNA research is the attempts to identify and amplify pathogen DNA from bone. Again, this method is in its early stages and almost all of the results published so far have been challenged. A major problem with this analysis is that the pathogen DNA is likely to be present in extremely small concentrations, if it has survived at all. A number of researchers have attempted to identify the pathogen that causes tuberculosis (*Mycobacterium tuberculosis*) with some success. However, it is important to note that the presence of this pathogen does not mean that the person actually had the disease, but could simply have been a carrier.

Due to contamination problems, most aDNA researchers will not use curated human skeletal material for analysis, but will only work on currently or recently excavated material. Preferably, bones will be excavated by people whose DNA sequences are known, either someone from the DNA laboratory or by a designated excavator.

#### **aDNA Application: Animal Bone and Teeth**

Following the same procedures outlined above for human bone and teeth, it is possible to identify the species of unidentified bone fragments, as well as produce phylogenetic information on certain species. For example, the sequences of living and ancient cattle samples from Europe and the Near East have been sequenced to determine when and where cattle were domesticated. If these analyses are undertaken in laboratories that have not analyzed modern animal samples, then there is much less concern about contamination. For this reason aDNA analysis of faunal samples has been among the most successful applications of this method.

#### **aDNA Application: Species Identification of Organics**

By extracting and then amplifying intact DNA from organic materials it is possible to determine the species of origin of that organic material. This is done using a range of species-specific PCR primers. These primers, as well as methods to characterize the DNA extract, have to be modified to account for the very fragmentary nature of the aDNA. With these methods, and providing that the DNA is still reasonably

intact, it is possible to identify the species used, for example, in producing parchment. Researchers have also been able to extract aDNA from animal coprolites, providing information on the range of plant and animal species the organism consumed.

#### **Stable Isotope Analysis**

Stable isotope analysis (SIA) method uses the stable isotope ratios of various elements as biological tracers. Human and animal body tissues are composed of elements that are ingested through food or water. The isotope ratios of stable isotopes of some of these elements (e.g., C, N, S, O, H, Sr, and Pb) from food and water are preserved in bone and teeth. Therefore, we can measure the isotope ratios of those elements in bone and teeth and then, by comparing those values to known isotope values of possible foods and water sources it is possible to determine aspects of past diets (palaeodiet) as well as the location where that human or animal consumed the food and water (migration).

Extraction procedures vary depending on the element of interest, and focus on purifying a specific component of bone and teeth. As with any study of preserved organics, contamination and degradation of the materials of interest is a problem. Therefore, extraction procedures are designed to both extract as much intact molecules as possible, and then characterize that component to determine the extent of degradation and contamination.

Measurements of the stable isotope ratios are conducted using mass spectrometers. For the light stable isotopes of carbon, nitrogen, oxygen, sulfur, and hydrogen it is possible to use continuous flow isotope ratio mass spectrometry. Here, the sample is combusted (C, N, S) in the presence of oxygen, or pyrolyzed (O, H) using glassy carbon, in an elemental analyzer in a stream of helium. The resultant gases are then passed through an isotope ratio mass spectrometer, which measures the relative amounts of the two stable isotopes of interest. For the heavier isotopes of strontium and lead it is necessary to purify the sample in a clean laboratory, as the concentrations in bone and teeth are on the order of parts per million or parts per billion, and then the solution is analyzed using inductively couple plasma-mass spectrometry (ICP-MS). It is also possible to use a laser, coupled to an ICP-mass spectrometer, to ablate bone and tooth samples, which reduces the needs for extensive sample preparation and allows targeted sampling of very small areas of bone and teeth.

By convention, light stable isotope values are reported using the delta notation, in parts per thousand. For example, the isotope ratio of the stable



isotopes of carbon,  $^{13}\text{C}$  and  $^{12}\text{C}$ , are reported as  $\delta^{13}\text{C}$  values according to the formula:

$$\delta^{13}\text{C} = (((^{13}\text{C}/^{12}\text{C}_{\text{sample}})/(^{13}\text{C}/^{12}\text{C}_{\text{standard}})) - 1) \cdot 1000$$

### SIA Application: Bone and Tooth Protein to Reconstruct Palaeodiet

Stable isotope analyses as used in modern and archaeological dietary studies endeavors to determine the dietary sources of carbon and nitrogen found in body tissues by measuring the ratios of the two stable carbon isotopes,  $^{13}\text{C}$  and  $^{12}\text{C}$ , and the ratios of the two stable nitrogen isotopes,  $^{15}\text{N}$  and  $^{14}\text{N}$ , in foods as well as in the body tissues of interest. The ratios of these isotopes are compared to known standards and are presented as  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values. Most of this research focuses on isotope measurements of the best preserved organic component of bone, the protein collagen (which comprises  $\sim 20\%$  of modern bone by weight). Collagen carbon and nitrogen is largely derived from dietary protein and likely reflects dietary inputs from approximately the last 10 years of life. Carbon isotope values indicate whether dietary protein came from marine or terrestrial sources, and it can also distinguish between  $\text{C}_3$  and  $\text{C}_4$  photosynthetic-pathway plants. Mammal collagen  $\delta^{15}\text{N}$  values indicate the trophic level of an organism in a food web, as there is an increase in the  $\delta^{15}\text{N}$  value of  $\sim 2\text{--}4\%$  each step up the food chain. Collagen  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values are specific to regions and ecosystems, and can also vary through time, related possibly to

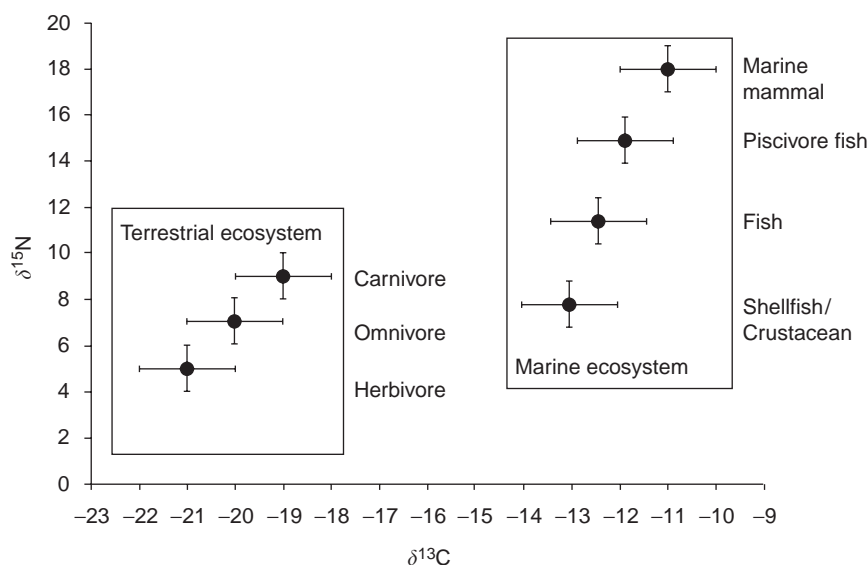
climatic effects. Therefore, it is important to take the ecosystem approach to isotope analyses and measure the  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values of fauna associated temporally, as well as spatially, with the humans of interest (Figure 1).

Additionally, studies of infant and juvenile bone and teeth  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values can tell us about the age of weaning in past populations, as infants that are breastfeeding have  $\delta^{15}\text{N}$  values that are higher than their mothers, which then drop to lower values when the child is weaned onto solid food.

Sampling protocols for carbon and nitrogen isotope analysis are fairly simple, as is the extraction procedure. This method requires the extraction of the protein collagen from bone, and then the further purification of this collagen for isotope analysis. SIA often requires only a few hundred milligrams of bone, less if the bone is well preserved.

### SIA Application: Bone and Tooth Mineral to Study Migration

Bone mineral and dental enamel oxygen isotope values reflect the oxygen isotope values of the water that a mammal consumes. If that mammal is migratory between climatic zones that have very different oxygen isotope ratios then the different values may be recorded in the bone or enamel. Therefore, oxygen isotopes have the potential to identify migrating species or humans. There are many exciting possibilities with oxygen isotopes of bone and enamel, but there are also serious concerns over contamination



**Figure 1** Typical carbon and nitrogen stable isotope ratios of collagen extracted from various species. In this example the terrestrial animals consume only  $\text{C}_3$  plants (herbivores and omnivores) or animals that consume only  $\text{C}_3$  plants (carnivores and omnivores). The marine organisms are distinguished from the terrestrial organisms through the  $\delta^{13}\text{C}$  value, which is less depleted in  $^{13}\text{C}$ . The  $\delta^{15}\text{N}$  value indicates the trophic level of an organism in the food web.

of bone and enamel by soil and groundwater oxygen. Generally, enamel has been shown to be much more immune to contamination than bone.

The measurements of the concentrations of various trace elements in bone and teeth has been used to determine past diets, but this area has largely been discredited in recent years, due to probably insurmountable problems with diagenesis and the uptake of new elements from the soil into the bone. However, promising advances are being made using the isotopes of some of these elements, like lead and strontium, to determine geographical place of origin of individuals. Currently, bone is not an appropriate material for this analysis, due to contamination problems, but tooth enamel is more resistant to diagenetic changes, and in some cases, can be used for this analysis. Usually, the whole tooth is used for these analyses, as trace element concentrations across the tooth need to be measured to test whether there has been soil contamination.

## Organic Residue Analysis

Organic residue analysis (ORA) focuses on the characterization, primarily for identification purposes, of ancient organic residues that have been preserved, often because of an association with an inorganic material. As with the other areas discussed here, researchers in ORA need to modify their analytical techniques for application to degraded and damaged surviving organics. There has been much erroneous identification published in the literature stemming from an inadequate appreciation of the degraded state of preserved organic molecules, and not recognizing that modern techniques need to be modified for ancient samples.

This area uses a variety of analytical methods. To identify and characterize lipids the main analytical technique employed is gas chromatography (GC), often in association with mass spectrometry (GC-MS). Newer advancements include the use of pyrolysis-GC-MS and the measurement of the stable isotope values of individual compounds in GC-C-IR-MS. To identify proteins a range of immunochemical identification techniques are used, including enzyme-linked immunosorbent assay and radioimmunoassay.

### ORA Application: Lipid Residues in Pottery

The most successful use of these identification methods is in the identification of preserved organics in pottery. It is likely that these residues are incorporated into the pottery through cooking, and survive as they are associated with the mineral structure of the pot, although the mechanisms are poorly understood.

GC-MS analysis of pottery can identify the presence of lipids derived from plant or animal sources, using the presence and ratios of various fatty acids. It is not possible to further define the source of the fats using reference values for these ratios from modern fats, as archaeological samples are greatly degraded, which changes the relative amounts of fatty acids. However, further identification of the source of the lipid fatty acids is possible through the use of GC-C-IR-MS. This significant advancement in the field measures the  $\delta^{13}\text{C}$  values of fatty acids of interest, allowing the identification of ruminant versus non-ruminant fats, as well as lipids derived from milk. A great success of this method has been the identification of the earliest use of milk in Britain.

Other compounds in pottery that can be identified using GC-MS are beeswax, identified by the distribution of *n*-alkanes and long-chain palmitic wax esters. Plant oils can also be possibly identified, and identification is often enhanced through the use of GC-C-IR-MS.

Sample preparation is relatively simple. A subsample of the potsherd is ground to powder and the lipids are then removed with the use of solvents such as chloroform and methanol. The resulting lipids are then derivatized before being introduced into the GC-MS.

### ORA Application: Protein Residues in Pottery

A significant recent advancement in this field is the application of immunochemical methods to identify proteins preserved in pottery. As with the lipids, the proteins appear to bind to the pottery and are protected and preserved. The proteins are degraded, however, so it is not possible to use antibodies prepared using modern proteins. Instead, new antibodies must be created using degraded proteins, either taken from archaeological samples, or artificially degraded in the laboratory. There is great potential in this method, as it could allow the identification of specific species, whereas GC-MS and GC-C-IR-MS only identify to broad categories. Unfortunately, relatively few pots have preserved proteins.

### ORA Application: Organic Residues on Stone Tools

Residues from plants and animals can be preserved on tools made from stone such as flint or obsidian. However, the processes involved in preserving the organics are also poorly understood, and indeed there has been some debate that stone is incapable of preserving residues, apart from exceptional circumstances. Theoretically, the residues are deposited during the use of the tool, whether that is through the processing of plant or animal foods, or both. Most of the published literature in this area relies on

the use of immunochemistry to identify species of origin of preserved proteins (e.g., blood proteins). This has been confounded by the use of antibodies made from modern materials as well as the use of general antibodies that can give false positive results.

#### ORA Application: Other Residues

It is also possible to characterize other residues found in the archaeological record. Two such residues of interest are natural products such as resins and bitumens and the other are lipids preserved in soils in archaeological sites. Resins from birch bark, pine, and spruce trees have been identified in the archaeological record using GC–MS. These resins have been used to, for example, haft stone tools, and seal and repair pottery. Identification of lipids in soils is a relatively new area, and the use of GC–C–IR–MS has been a major advancement in this area. It is possible to identify the use of manure on fields, which can pinpoint the first farming in a region, as well as help identify farming methods. Additionally, it has been possible to characterize the embalming materials used to preserve Egyptian mummies.

See also: **Archaeometry and Antique Analysis:** Dating of Artifacts; Metallic and Ceramic Objects; Art and Conservation.

#### Further Reading

- Ambrose SH (1993) Isotopic analysis of paleodiets: Methodological and interpretive considerations. In: Sandford MK (ed.) *Investigations of Ancient Human Tissue: Chemical Analyses in Anthropology*, pp. 59–130. Langhorne, PA: Gordon and Breach Science Publishers.
- Brothwell DR and Pollard AM (2001) *Handbook of Archaeological Sciences*. Chichester: Wiley.
- Craig OE and Collins MJ (2000) An improved method for the immunological detection of mineral bound protein using hydrofluoric acid and direct capture. *Journal of Immunological Methods* 236: 89–97.
- Evershed RP (2000) Biomolecular analysis by organic mass spectrometry. In: Ciliberto E and Spoto G (eds.) *Modern Analytical Methods in Art and Archaeology*, pp. 177–239. New York: Wiley.
- Hofreiter M, Serre D, Poinar HN, Kuch M, and Pääbo S (2001) Ancient DNA. *Nature Reviews Genetics* 2: 353–359.
- Katzenberg MA (2000) Stable isotope analysis: a tool for studying past diet, demography and life history. In: Katzenberg MA and Saunders SR (eds.) *The Biological Anthropology of the Human Skeleton*, pp. 305–327. New York: Wiley.
- Pollard AM and Heron CP (1996) *Archaeological Chemistry*. Cambridge: Royal Society of Chemistry.

## ARSENIC

**A Raab and J Feldmann**, University of Aberdeen, Aberdeen, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The metalloid arsenic belongs to the same group as the essential elements nitrogen and phosphorus. Like nitrogen and phosphorous, it forms stable oxygen bounds, but in anaerobe environment it prefers sulfur as binding partner. It forms stable penta- and trivalent compounds, whereby inorganic arsenic compounds are often found together with other chalcophilic elements like copper or iron. Organic arsenic compounds, containing one to four arsenic–carbon bounds, are widespread in biological samples. Inorganic compounds of arsenic are widely used in the metallurgy and in the semiconductor industry, organo-arsenic compounds are among others used as pesticides and herbicides.

Arsenic speciation is the analytical activity that leads to the identification and quantification of a

specific form (molecular or complex structure or oxidation state) of arsenic in a given environment (as derived from the IUPAC definition on speciation). It includes sampling, sample preparation, identification, and quantification of the species, under study, without the induction of changes to this species. In the following only such techniques are considered that allow a characterization of functionally defined species. Other techniques defining operational characterized species, like the differentiation of species into soluble and insoluble in a given solvent, are not considered.

The knowledge of the arsenic species present in a given matrix, e.g., water, has a large impact on the health of living beings, since the toxicity of arsenic is strongly depending on its molecular form. Inorganic arsenic species were used as early as 500 BC for their toxic properties and their ascribed beneficial health effects. Arsenic compounds were among others used for the treatment of syphilis. They are still used in some traditional medicines, like Chinese ones, and arsenic oxide has recently been under trial for the treatment of acute promyelocytic leukemia.

Nevertheless, the research into the negative health effects of different arsenic compounds is predominant, especially into the toxicity and carcinogenicity of different arsenic species for humans.

## Arsenic Species

The inorganic arsenious acid ( $\text{As}^{\text{III}}$ ) and arsenic acid ( $\text{As}^{\text{V}}$ ) and their salts are water soluble and present with their high toxic and carcinogenic potential (Table 1) a widespread problem for the water industry trying to provide safe drinking water. Both species are metabolized by most organisms to methylated organo-arsenicals, with often a lower toxic, but not necessarily a lower carcinogenic, potential. Organo-arsenicals exist as do the inorganic species in the penta- and trivalent oxidation state, whereby the pentavalent state is the more stable one in oxidative environments, like dimethylarsinic acid ( $\text{DMA}^{\text{V}}$ ) and methylarsonic acid ( $\text{MA}^{\text{V}}$ ). Recently, the trivalent methylated intermediates ( $\text{MA}^{\text{III}}$ ,  $\text{DMA}^{\text{III}}$ ) enjoyed major attention, since these species show higher toxicity than inorganic precursors. The metabolic pathway of inorganic arsenic to  $\text{DMA}^{\text{V}}$  was first described in 1945 by Challenger. It includes the alternating reduction of pentavalent to trivalent arsenic followed by oxidative methylation. Whereas in humans and most mammals the arsenic metabolism ends with the excretion of the stable pentavalent  $\text{DMA}^{\text{V}}$ , fish and others produce and/or store arsenobetaine, containing four arsenic-carbon bonds, and marine algae among others produce arsenic compounds containing a ribose moiety. Recently, labile arsenic-peptide complexes, tri-(glutamylcysteinylglycyl)trithioarsenite ( $\text{As}^{\text{III}}(\text{GS})_3$ ) and arsenite-phytochelatin(3) ( $\text{As}^{\text{III}}\text{-PC}_3$ ) were identified in mammals and plants, respectively

(Figure 1). These different arsenic species have a wide range of acute toxicity and carcinogenicity toward humans, which makes the determination and quantification of the arsenic species present in, e.g., food or water, mandatory for health and safety reasons.

## Analytical Strategies

Arsenic speciation can be divided into three subtasks each depending on the others:

- sample preparation (collection, preparation, and storage);
- separation of the arsenic species from other molecules (species) present in the sample (by hydride generation (HG), gas chromatography (GC) or liquid chromatography (LC)); and
- detection (element specific and/or molecule specific).

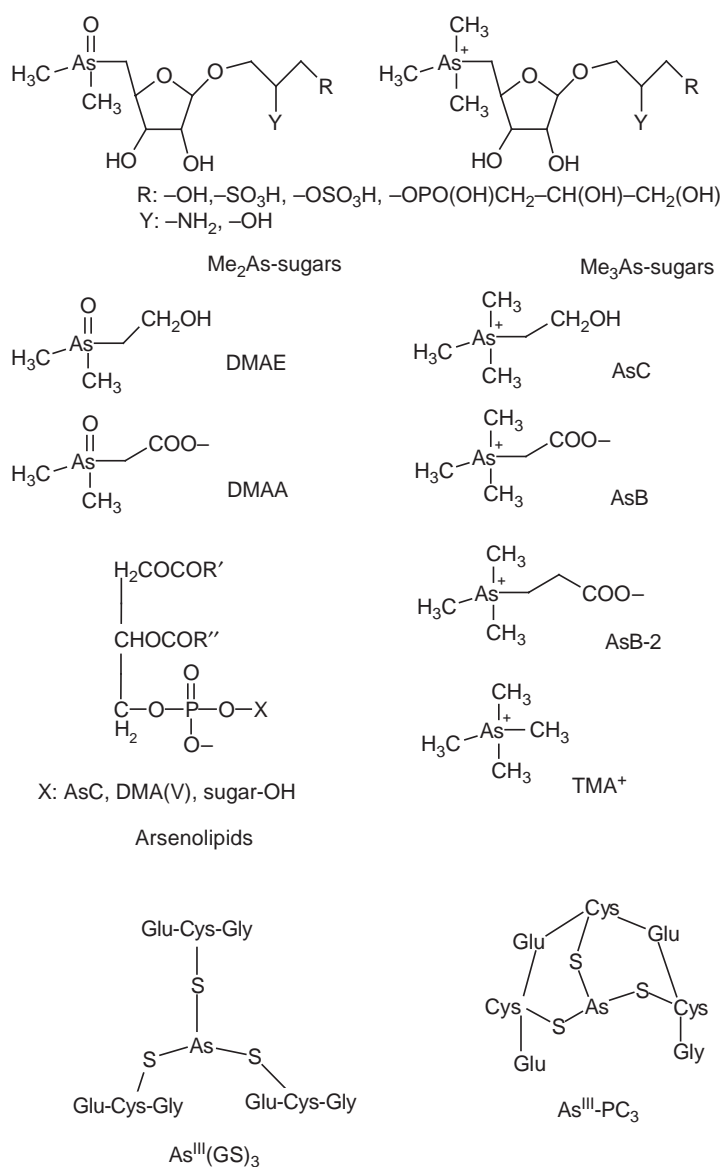
These tasks are described in the following in more detail.

## Sample Preparation

The main focus during sample collection, preparation, and storage has to be the preservation of the arsenic species present in the original or at least to conserve the major functional group information of the species. Most of the pentavalent arsenic species present in biological or environmental samples identified so far, in, for example, urine, are quite stable. Samples containing them can be collected, prepared, and stored without risk of species changes as long as contamination and a change of pH and salinity is avoided and the storage conditions are not reductive.

**Table 1** Selected arsenic species occurring in the environment (E) and biota (B) and their determined acute toxicity for male mice in mg per kg body weight

No. of R	Trivalent species	B	E	LD <sub>50</sub>	Pentavalent species	B	E	LD <sub>50</sub>
0	As(OH) <sub>3</sub>	✓	✓	4.5 (rat)	H <sub>3</sub> AsO <sub>4</sub>	✓	✓	
	AsH <sub>3</sub>	✓	✓	3				
1	MA <sup>III</sup>	✓	✓		MA <sup>V</sup>	✓	✓	1800
1	MeAsH <sub>2</sub>	✓	✓					
2	DMA <sup>III</sup>	✓	✓		DMA <sup>V</sup>	✓	✓	1200
2	Me <sub>2</sub> AsH	✓	✓					
3	Me <sub>3</sub> As (TMA <sup>III</sup> )	✓	✓	8000	TMAO	✓	✓	> 10 000
3					DMAE		✓	?
3					DMAA	✓	✓	
3					Me <sub>2</sub> As-sugar-OH, -SO <sub>4</sub> , etc.	✓	(✓)	> 12 000
3					Arsenolipids	✓		?
4					TMA <sup>+</sup>	✓	✓	900
4					Me <sub>3</sub> As-sugars	✓		
4					AsC	✓	✓	6540
4					AsB	✓	✓	> 10 000
4					AsB-2	✓		?



**Figure 1** Molecular structures of some selected organoarsenic species occurring with some commonly used abbreviations.

Trivalent arsenic species need a reductive sample environment; this is especially true for methylated species, since these are extremely labile and sensitive to oxygen. A similar problem is presented by arsenic species containing sulfur, whether it is dimethylarsinothioic acid (DMA<sup>V</sup>), where there is a direct arsenic sulfur bond, or arsenic species complexed by sulfur containing amino acids (e.g., As<sup>III</sup>(GS)<sub>3</sub>). Gas samples containing volatile arsenic species (Me<sub>x</sub>AsH<sub>3-x</sub>) can be stored for weeks in the dark, but they oxidize in the presence of ultraviolet (UV) light within hours.

Depending on the intended technique used for separation of the arsenic species most samples need

to be transferred into liquid form without disturbance of the arsenic species. Sample matrices like water or urine do not require much sample preparation except maybe filtration, but other matrices like plant or tissue samples require the use of extraction methods prior to the separation. The extraction method(s) used have to be tested whether they enable the extraction of the arsenic species without changing the oxidation state or in case of complexes binding partners. Many arsenic species, like DMA<sup>V</sup>, arsenobetaine, or arsenosugars are stable enough to survive relatively harsh extraction conditions. For these methanol/water in combination with ultrasonication is an often-used extraction method.

Other arsenic species, like the trivalent methylated intermediates of arsenic metabolism predicted by Challenger, were for many years fairly elusive, because of their instability against oxygen and harsh extraction conditions. The recently identified arsenic–glutathione and arsenite–phytochelatin complexes suffer from the same problem, and can only be stabilized in acidic extracts.

Generally, the sample should be measured after sample collection as quickly as possible with as limited a sample preparation as possible. If sample storage is necessary for some time, subsamples should be stored under different conditions to allow the identification of species transformation during storage (e.g., oxidation, methylation, or demethylation). Lower temperatures,  $-20^{\circ}\text{C}$  to  $4^{\circ}\text{C}$ , are generally more useful than the addition of preservatives and acid, which may have an influence on the speciation.

Over the years a wide range of analytical techniques were employed for the speciation of arsenic, including different separation techniques. Which technique is used depends among others on the question to be answered and the species to be identified.

In addition, it should be mentioned that the highest emphasis until now is on water-soluble arsenic species, which is reflected in the remainder of the article. But it should not be forgotten that lipid-soluble arsenic species (like arsenolipids) exist. However, they need different extraction protocols using mainly lipophilic solvents like methanol/chloroform and different separation techniques.

## Separation of Arsenic Species

The separation technique suited depends on the complexity of the sample matrix and the number of arsenic species present in the sample.

### Hydride Generation Coupled to Gas Chromatography (HG–GC)

HG in its simplest form allows the speciation of inorganic  $\text{As}^{\text{III}}$  and  $\text{As}^{\text{V}}$ . For the speciation of samples containing more than these two hydride-active species, a combination with GC is necessary. Every arsenic species, with the exception of those carrying four carbon bonds, can be transformed into arsines by borohydride. Most of the arsines are volatile and can be separated by GC. Depending on the pH used during hydride formation a distinction between tri- and pentavalent arsenic is in most cases possible. Trivalent arsenic species form hydrides already at pH 7, whereas pentavalent ones are only reactive at pH 1 (Table 2). Measuring the same sample with both pHs

**Table 2** Arsenic species and the arsines formed by treatment with borohydride

Species	Solution pH 1	Solution pH 7 (buffered)
$\text{As}(\text{OH})_3$	$\text{AsH}_3$	$\text{AsH}_3$
$\text{H}_3\text{AsO}_3$	$\text{AsH}_3$	–
$\text{MA}^{\text{III}}$	$\text{MeAsH}_2$	$\text{MeAsH}_2$
$\text{MA}^{\text{V}}$	$\text{MeAsH}_2$	–
$\text{DMA}^{\text{III}}$	$\text{Me}_2\text{AsH}$	$\text{Me}_2\text{AsH}$
$\text{DMA}^{\text{V}}$	$\text{Me}_2\text{AsH}$	–
TMAO	$\text{Me}_3\text{As}$	$\text{Me}_3\text{As}$
$\text{DMAS}^{\text{V}}$	$\text{Me}_2\text{AsH}$	$\text{Me}_2\text{AsH}$

gives therefore the ratio of penta- to trivalent arsenic. It has to be kept in mind that the efficiency of the hydride formation is species dependent and there might be pentavalent arsenic species present in the sample that form hydrides already at pH 7, like  $\text{DMAS}^{\text{V}}$ . HG is routinely used for the determination of inorganic arsenic species and in combination with cryotrapping and GC for the separation of mono-methylated and dimethylated arsenic, since these species are easy to volatilise. Dimethylated arsenosugars, which can as well form volatile arsenic species during HG, have not yet been successfully separated by this technique.

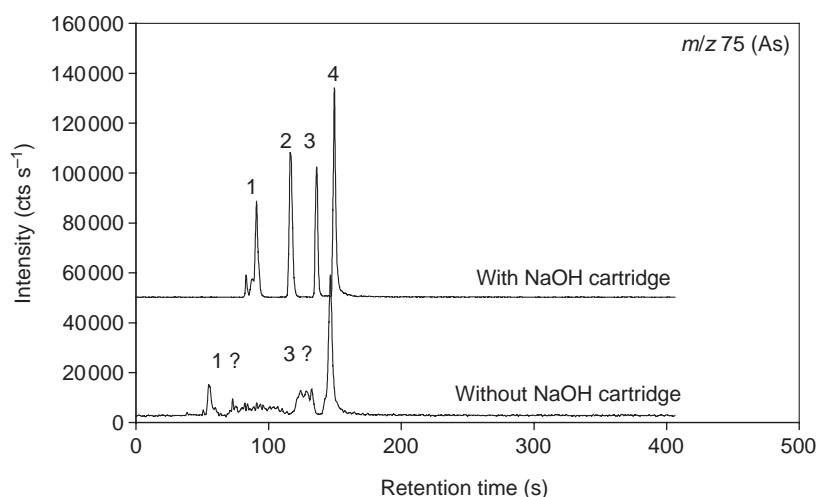
### Gas Chromatography

The volatile hydrides formed during HG with borohydride can be separated on a nonpolar GC column depending on their boiling point. Additionally, gaseous arsines can be trapped cryogenically and separated directly (Figure 2).

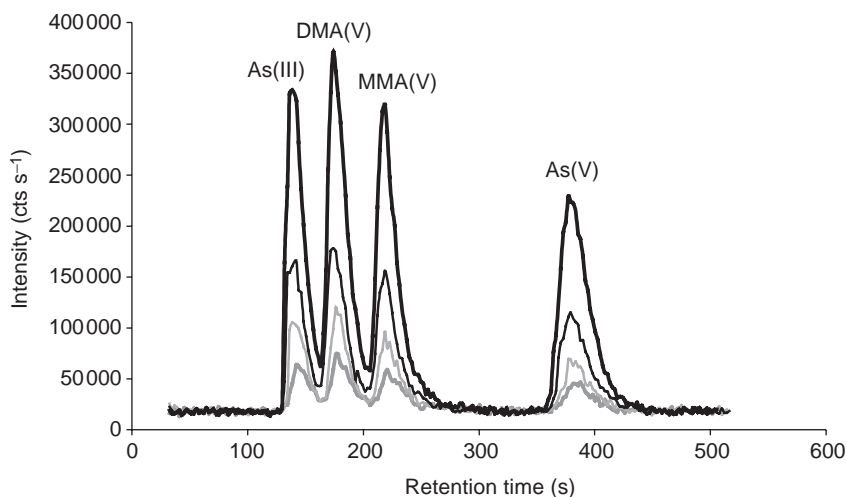
### Liquid Chromatography

Over the years especially anion and cation exchange chromatography, ion-pair chromatography, and size exclusion chromatography showed their value in the speciation of arsenic-containing compounds. All these chromatographic methods have the additional advantage that they are well suited for the online combination with most detectors used to detect arsenic species. One example for the separation of the most important arsenic species is shown in Figure 3.

Most arsenic species are either cationic or anionic at a given pH and can therefore be separated by ion chromatography. A two-method approach (anion and cation exchange chromatography) can result in a 2D map of retention times for more than 10 arsenic species (Figure 4) that assists in identification. Using carefully chosen separation conditions the main arsenic species present in biological samples can be separated. More labile arsenic complexes are mostly separated by size exclusion chromatography, which



**Figure 2** CT-GC-ICP-MS chromatogram of an air sample containing large amounts of  $\text{CO}_2$  with and without NaOH cartridges before cryotrapping of 1:  $\text{AsH}_3$ , 2:  $\text{MeAsH}_2$ , 3:  $\text{Me}_2\text{AsH}$ , and 4:  $\text{Me}_3\text{As}$  of a concentration of  $10 \mu\text{g}$  arsenic per  $\text{m}^3$  each.



**Figure 3** Separation of multispecies standard of  $\text{As(III)}$ ,  $\text{DMA(V)}$ ,  $\text{MA(V)}$ , and  $\text{As(V)}$  on a short strong anion exchange column (150 mm PRP X-100, Hamilton,  $30 \text{ mol l}^{-1}$  phosphate buffer adjusted with ammonia to pH 6.0). Concentration of each species 5, 10, 15, and  $30 \mu\text{g}$  arsenic per liter.

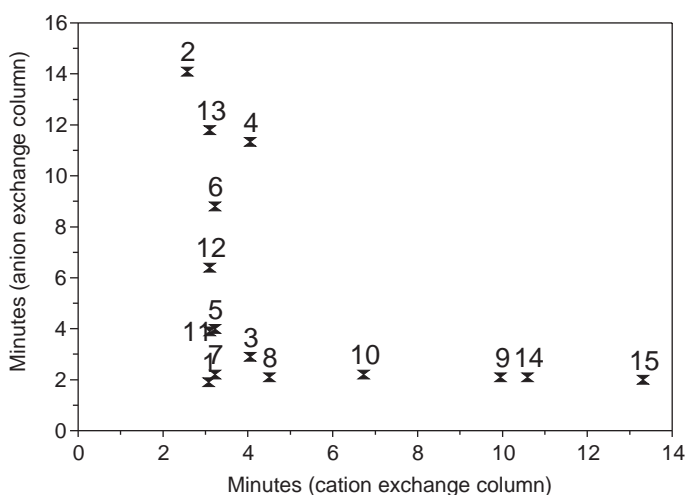
is a gentler separation method but suffers from low resolution especially since most arsenic species are present as relatively small molecules. Reverse-phase chromatography can be used successfully for the separation of these labile arsenic complexes (e.g.,  $\text{As}^{\text{III}}(\text{GS})_3$ ) as well. This kind of separation has a much better resolution power than size exclusion chromatography, but is sometimes difficult to combine with elemental detectors like inductively coupled plasma-mass spectrometers, since it uses organic solvents as mobile phase.

Several points need consideration in addition to the ability to separate the arsenic species of interest from other species present in the sample with the chosen chromatographic conditions. One of them is the stability of the species in the eluent and on the

column; another is the recovery rate of the species from the column. A good example for these potential problems is the separation of  $\text{As}^{\text{III}}(\text{GS})_3$ , which is unstable in alkaline conditions and does not elute from a strong anion exchange column except after disintegration to  $\text{As}^{\text{III}}$ .

It is not only necessary to have an inside knowledge of the separation power of the different techniques, more importantly, it is also necessary to know which arsenic species will be hidden by using the separation technique chosen.

Capillary electrophoresis (CE) has been used to separate a number of arsenic species, but up to now problems with matrix effects and the high detection limits do not permit its application to biological and environmental samples.



**Figure 4** Comparison of retention times (min) of standard arsenic species on strong anion and cation exchange columns (Hamilton PRP X-100 (20 mmol l<sup>-1</sup> carbonate pH 8.0) and Sulpelcosil LC-SCX (20 mmol l<sup>-1</sup> pyridine pH 3.0)). Arsenic species: 1: As<sup>III</sup>, 2: As<sup>V</sup>, 3: DMA<sup>V</sup>, 4: DMAS<sup>V</sup>, 5: MA<sup>V</sup>, 6: MAS<sup>V</sup>, 7: MA<sup>III</sup>, 8: AsB, 9: AsC, 10: Me<sub>2</sub>As-sugar-OH, 11: Me<sub>2</sub>As-sugar-PO<sub>4</sub>, 12: Me<sub>2</sub>As-sugar-SO<sub>3</sub>, 13: Me<sub>2</sub>As-sugar-SO<sub>4</sub>, 14: DMAE, 15: TMA<sup>+</sup>.

## Detectors for Arsenic Species

There are two different approaches to detect arsenic species; one is the use of an element-specific detector and the other the use of a molecule-specific detector. Both have their positive and negative points. Using an arsenic-specific detector (ICP-MS) in combination with the chromatography enables the detection and identification of arsenic-containing species by comparing the retention times with those of known standards, but not the identification of the molecules per se which is possible by using a molecule-specific detector, like ES-MS (electrospray mass spectrometer). The use of ICP-MS and ES-MS, the most often used detectors for arsenic speciation, is described in some detail in the following paragraph, with some additional information on the other often used detectors.

### ICP-MS

Most chromatographic techniques can be coupled online to an ICP-MS without any problems. In the harsh ionization conditions of the plasma, molecules disintegrate rapidly and form readily single charged element ions that are separated by their mass charge ratio and detected. The signal intensity of the element is mostly independent of the molecular species prior to ionization, which enables quantification without the need to use identical standard species. Especially in the quantification of arsenic compounds from biological samples, which often contain unknown species, this feature is very useful. ICP-MS is also a quite sensitive technique that often achieves detection limits of a few nanograms or picograms per

sample volume/mass. Using ICP-MS as detector for arsenic speciation with its low detection limits has also some considerable drawbacks. Among these is the complete loss of molecular information, which allows the identification of species only by retention time and comparison with a known standard when coupled with GC or high-performance liquid chromatography (HPLC). This presents the main problem when biological samples are analyzed, since known and well-characterized arsenic species are rarely available.

### ES-MS

ES-MS gives in contrast to the ICP-MS molecular information, mainly the molecular mass of the protonated molecule (in positive mode), and depending on the instrument used it can give information about molecule fragments formed in the high vacuum of the instrument. ES-MS is a so-called soft ionization method during which normally one proton is transferred onto the neutral sample molecule producing, therefore, a single charged cation for small molecules while larger molecules can add multiple protons resulting in multiple charged cations. They are separated by their mass-to-charge ratio ( $m/z$ ) from other molecules. Using either an instrument able to work in the MS/MS mode or repeatedly measuring the same sample with different fragmentor voltages (FVs) does often give useful information about the species. It has to be kept in mind that most biological and environmental sample matrices often contain a number of species that show up at the same  $m/z$  ratio as the arsenic species, therefore making the task of identifying the species in the sample quite difficult to



impossible without additional information. The signal intensity in ES-MS spectra depends on the molecule itself, the exact separation conditions (especially pH and buffer concentration), and on the concentration of co-eluting molecules. Since these factors have an unpredictable influence, quantification by ES-MS in raw sample (extracts) is only possible in standard addition mode when the species is available as pure standard compound.

### Other Detectors

Atomic fluorescence spectrometers (AFS) are sometimes used online for the detection of hydride-forming arsenic species. Nonhydride-forming species, like arsenosugars, can only be determined by AFS after online digestion, although recently a report has revealed that even these arsenosugars may form volatile arsine species. X-ray techniques, like XANES or EXAFS, can be used to determine the electronic environment of the arsenic. These techniques are often applied to determine whether arsenic is bound to sulfur or oxygen atoms. Proton and carbon nuclear magnetic resonance (NMR) can be used for the identification of the organic part of an arsenic species, when the sample is concentrated enough. Samples intended for NMR measurements must be pure; otherwise the assignment of the signals to the hydrogen and carbon atoms in the molecule is not possible. The arsenic nucleus itself is not useful for NMR measurements, since its large quadrupole moment causes extensive line broadening even in highly symmetric molecules like  $\text{AsF}_6^-$ .

Arsenic species might not be separable and detectable due to:

- co-elution with other more abundant arsenic species;
- decomposition of labile compounds during extraction or separation;
- irreversible reaction with the stationary phase; or
- formation of macromolecules that stick to the columns.

## Arsenic Speciation in Environmental and Biological Samples

### Determination of Arsines in Landfill Gas

For example, bacteria living in landfill sites can metabolize the arsenic present in such sites to volatile arsines. For the determination of these highly toxic ultratrace compounds by GC it is necessary to have a series of cleanup steps if GC is coupled to electron impact mass spectrometry (GC-MS) or ICP-MS, since the samples contain dozens of different organic

**Table 3** Estimates atmospheric lifetime of volatile arsenic compounds in moisturized air at a starting concentration of  $10 \mu\text{g m}^{-3}$

	Dark, 20°C	Dark, 50°C	UV, 30°C (5000 lux)
$\text{AsH}_3$	>2000 h	>600 h	150.0 h
$\text{MeAsH}_2$	1440 h	600 h	1.5 h
$\text{Me}_2\text{AsH}$	480 h	84 h	0.1 h
$\text{Me}_3\text{As}$	—	30 h	0.1 h

compounds in high concentrations. Due to the low concentration of volatile arsenic compounds a pre-concentration method is necessary. Cryotrapping (CT) on a packed column filled with nonpolar chromatographic material (SP-2100 on Chromosorb) with subsequent separation of the different species according to their boiling point can be used to pre-concentrate and select the fraction containing arsenic species. The fraction is collected and injected again onto a second column. This time the column has a higher separation power and a different polarity, so that an unequivocal identification of the volatile metal species can be made by fragmentation patterns of the eluting molecules using GS-MS. Once the structures are fully established using this method, capillary-GC can be coupled directly to ICP-MS without the use of a molecule-specific detector. ICP-MS is much more sensitive for the different metal species, which makes the cleanup steps redundant. **Figure 2** illustrates the sensitivity of the CT-GC-ICP-MS with detection limits at the subpicogram level. Using a NaOH cartridge,  $\text{CO}_2$  can be absorbed before the gas sample is cryotrapped (**Figure 2**). This technique was routinely used for the stability testing of volatile arsenic species in moisturized air at room temperature and at 50°C in the dark and under intense UV radiation. Concentrations of  $\sim 10 \mu\text{g}$  arsenic per  $\text{m}^3$  are a realistic concentration in order to find out whether volatile species have a high enough atmospheric lifetime to diffuse in the vicinity of point sources such as landfill sites. The stability of arsines depends on the methylation grade, temperature, and more significantly UV radiation in the environment (**Table 3**).

### Arsenic Metabolites of Inorganic Arsenic in Mammals

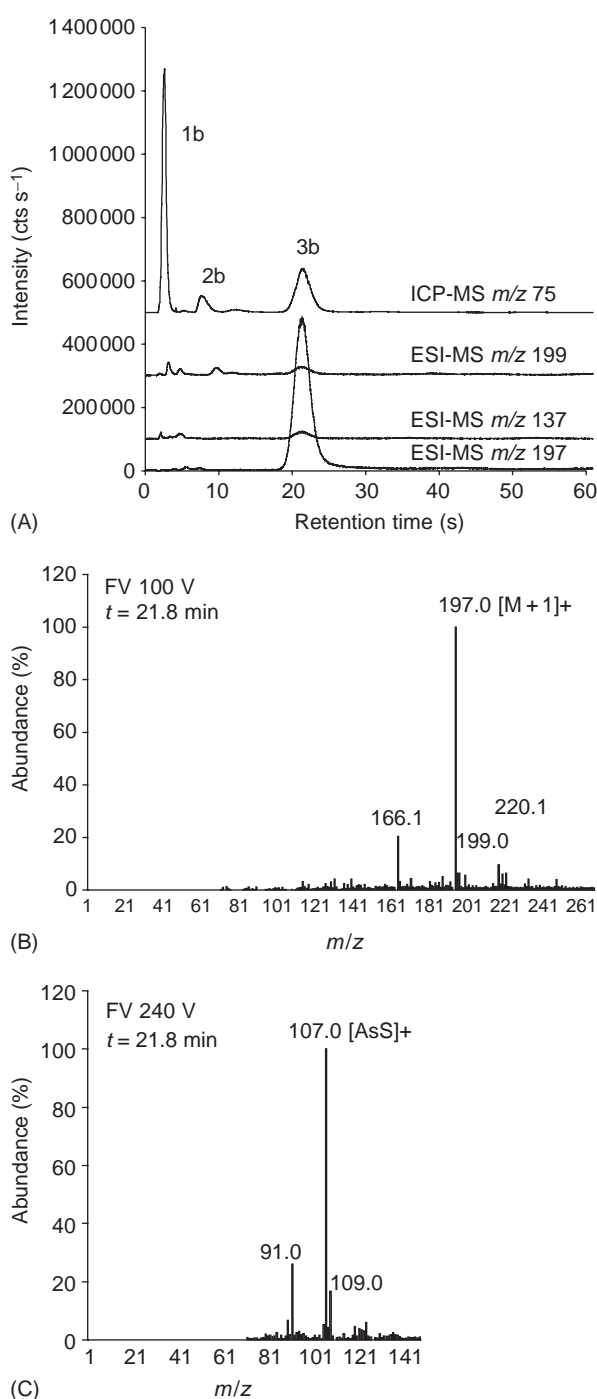
Ingested arsenate and arsenite are taken up in the intestine and transported by the blood into the liver, which is the major site of arsenic metabolism before excretion by urine.

The main excreted forms are dimethylarsinic acid ( $\text{DMA}^{\text{V}}$ ) and methylarsonic acid ( $\text{MA}^{\text{V}}$ ); in addition, several minor arsenic compounds can be found in urine depending on the amount of ingested arsenic

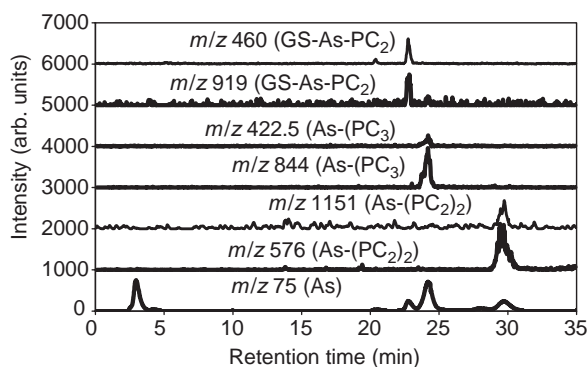
and duration of exposure. LC coupled to ICP-MS is the most sensitive way to study arsenic species in urine with detection limits of less than 1 ng As per ml. Anion exchange and cation exchange chromatography are the main separation techniques. Since ICP-MS is an element-specific detector the availability of pure, precisely characterized standard compounds for identification is essential as the careful check, even if a chromatographic signal contains more than one arsenic compound. Using only anion exchange chromatography can lead to misidentification of cationic species such as arsenobetaine as arsenite, since both compounds are very difficult to separate on such a column (Figure 4). For the identification of co-eluting cationic species the use of cation exchange chromatography is essential. Nevertheless, misidentification can be anticipated using only ICP-MS as detector. Unknown species like dimethylarsenothioyl acetic acid, recently found in urine, can only be identified using elemental and molecular detectors and the characteristic fragmentation pattern of the species (Figure 5).

### Arsenic in Biota

Plants have developed a whole range of ways to deal with high arsenic concentrations in their environment; some even actively accumulate arsenic to extremely high levels. To avoid toxicity they reduce either the uptake of arsenic (in soil/water mostly present as arsenate) by reducing the transporters for phosphate that are used for the uptake of arsenate. Another way to deal with the high toxicity is to metabolize the inorganic arsenic into less acutely toxic organic species or they sequester arsenic into the vacuole with the aid of specific peptides (phytochelatins, PCs) and store it there. The arsenic complexes of these peptides synthesized by plants in response to arsenic exposure were only recently identified. Since these arsenic species are not covalently bound and contain the more labile trivalent arsenic their stability under standard extraction and storage conditions is limited. Arsenic-PC complexes are unstable under the influence of oxygen and disintegrate into the peptide and arsenite, which is then partially oxidized to arsenate. A successful extraction of these complexes is possible under acidic conditions at 1°C using fresh plant material. The separation of As-PCs is also hampered by their low stability under most separation conditions. Using size exclusion chromatography with ICP-MS detection does not provide conclusive proof, even ES-MS is not too helpful since the separation power of size exclusion chromatography is very limited. The complexes are successfully separated using reverse-phase chromatography with organic



**Figure 5** (A) Separation of newly identified arsenic species (3b, dimethylarsenothioyl acetic acid) in urine by using strong anion exchange chromatography (PRP X-100 Hamilton) coupled to ICP-MS for arsenic selective detection ( $m/z$  75) and parallel to ESI-MS for the protonated molecular mass ( $m/z$  197, 199) and major fragments ( $m/z$  137). (B) Gentle ionization at 100 V. (C) Molecular fragments of  $m/z$  107, 109 (As=S) and  $m/z$  91 (As=O) at 240 V fragmentor voltage. (Reproduced with permission from Hansen HR, Pickford R, Thomas-Oates J, Jaspars M, and Feldmann J (2004) 2-Dimethylarsenothioyl acetic acid identified in a biological sample: The first occurrence of a mammalian arsinothioyl metabolite. *Angewandte Chemie International Edition* 43(3): 337–340; © Wiley VCH.)



**Figure 6** Separation of As-PC complexes, ESI-MS data of As-PC species,  $m/z$  75 (arsenic) measured by parallel use of ICP-MS, separation with MeOH/formic acid gradient on a C18 ODS2 column (Reprinted with permission from Raab A, Feldmann J, and Meharg AA (2004) The nature of arsenic-phytochelatin complexes in *Holcus lanatus* and *Pteris cretica*. *Plant Physiology* 134: 1113–1122; © American Society of Plant Biologists.)

solvent (Figure 6). This separation method is well suited for the combination with ES-MS; it even improves the sensitivity of this technique somewhat, but is difficult to combine with ICP-MS. Combining reverse-phase chromatography online to an ICP-MS is possible, as done in this example, by addition of oxygen to the plasma and using a low-flow nebulizer for the ICP-MS with a flow splitter after the column to avoid too high amounts of organic solvent reaching the plasma of the ICP-MS.

## Emerging Techniques

Arsenic speciation in biological and environmental samples is anything but routine analysis. This is indicated in the small number of certified reference materials available with certified concentration of arsenic species. Attempt so far to produce standard reference materials for arsenosugars have failed and these are compounds that are considered as relatively stable arsenic species during extraction and separation procedures. The real challenge today is to identify unknown arsenic-containing compounds detected by LC-ICP-MS. For this purpose other techniques such as ES-MS or MALDI-MS are used to identify the molecular mass of the species. Multiple dimensional chromatographic separations (size exclusion chromatography + ion exchange + reverse phase) can help to solve the problem of co-eluting species for stable arsenic species as has been demonstrated successfully for the identification of a series of complex organoarsenicals in seaweed. FT-ICR-MS and qTOF-MS may bring a new dimension in the field of arsenic speciation, due to their accurate mass measurements using the mass defect of arsenic ( $m = 74.92159 \text{ g mol}^{-1}$ ) as an indicator of an

arsenic-containing compound. It is expected that traditional biochemical methods such as gel electrophoresis will find more use in this field, especially for the determination of arsenic-protein complexes. In addition to this new fleet of conventional techniques, X-ray absorption methods such as XANES or EXAFS can be utilized for the detection of very labile arsenic species directly in the unchanged organism, when they are present in concentrations of more than 20 mg per kg. It is just emerging that besides the wealth of organoarsenic species, many labile coordination species may occur, which may play a key role in the complex biochemistry of arsenic.

**See also:** **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Liquid Chromatography:** Overview. **Mass Spectrometry:** Electrospray.

## Further Reading

- Cornelis R (ed.) (2003) *Handbook of Elemental Speciation, Parts 1 and 2*. Chichester: Wiley.
- Cullen WR and Reimer KJ (1989) Arsenic speciation in the environment. *Chemical Review* 89: 713–764.
- Devalla S and Feldmann J (2003) Determination of lipid-soluble arsenic species in seaweed-eating sheep from Orkney. *Applied Organometallic Chemistry* 17: 906–912.
- Feldmann J, Lai VW-M, Cullen WR, *et al.* (1999) Sample preparation and storage can change arsenic speciation in human urine. *Clinical Chemistry* 45(11): 1988–1997.
- Kala SV, Neely MW, Kala G, *et al.* (2000) The MRP2/cMOAT transporter and arsenic-glutathione complex formation are required for biliary excretion of arsenic. *Journal of Biological Chemistry* 275(43): 33404–33408.
- Kuehnelt D, Irgolic KJ, and Goessler W (2001) Comparison of three methods for the extraction of arsenic compounds from the NRCC standard reference material DORM-2 and the brown alga *Hijiki fuziforme*. *Applied Organometallic Chemistry* 15(6): 445–456.
- Le XC, Lu XF, Ma MS, *et al.* (2000) Speciation of key arsenic metabolic intermediates in human urine. *Analytical Chemistry* 72(21): 5172–5177.
- McSheehy S, Pohl P, Lobinski R, and Szpunar J (2001) Complementarity of multidimensional HPLC-ICP-MS and electrospray MS-MS for speciation analysis of arsenic in algae. *Analytica Chimica Acta* 440(1): 3–16.
- McSheehy S, Szpunar J, Morabito R, and Quevauviller P (2003) The speciation of arsenic in biological tissues and the certification of reference materials for quality control. *Trac – Trends in Analytical Chemistry* 22(4): 191–209.
- Pedersen SN and Francesconi KA (2000) Liquid chromatography electrospray mass spectrometry with variable fragmentor voltages gives simultaneous elemental and molecular detection of arsenic compounds. *Rapid Communications in Mass Spectrometry* 14(8): 641–645.
- Pergantis SA, Wangkarn S, Francesconi KA, and Thomas-Oates JE (2000) Identification of arsenosugars at the

picogram level using nanoelectrospray quadrupole time-of-flight mass spectrometry. *Analytical Chemistry* 72(2): 357–366.

Stybło M, Del Razo LM, Vega L, *et al.* (2000) Comparative toxicity of trivalent and pentavalent inorganic and

methyated arsenicals in rat and human cells. *Archives of Toxicology* 74(6): 289–299.

Wilkin RT, Wallschläger D, and Ford RG (2003) Speciation of arsenic in sulfidic waters. *Geochemical Transactions* 4: 1–7.

## ASBESTOS

**H Janik and M Wrona**, Technical University of Gdansk, Gdansk, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Asbestos minerals are subdivided into two classes, with six distinct types of fiber (Table 1). The serpentine group consists solely of chrysotile, popularly known as ‘white asbestos’, which once accounted for more than 95% of the asbestos used worldwide. Its crystal morphology is snake-like, with a tendency to form bundles. It is softer and more flexible than the other types of fiber. Chrysotile has a low iron content (~1.5%) and limited durability in the lung. The amphibole group consists of five chemically and morphologically diverse fibers primarily represented by crocidolite and amosite asbestos. Crocidolite, sometimes called ‘blue asbestos’, occurs as needle-like fibers that are high in iron content (36%) and are more durable than chrysotile in human lung.

Because of both natural and industrial processes, asbestos fibers have been widely dispersed and are present in the ambient atmosphere, the hydrosphere, in the soil of many areas, and in lung tissue after

occupational exposure. The establishment of asbestos fibers as carcinogenic and causing pulmonary and pleural diseases has led to the development of increasingly sensitive analyses for their detection and identification.

Optical microscopy (OM), polarized light microscopy (PLM), phase contrast microscopy, scanning electron microscopy (SEM), transmission electron microscopy (TEM), and scanning transmission electron microscopy (STEM) are the methods normally used for identification and quantification of the trace amounts of asbestos fibers that are encountered in the environment and lung tissue. Energy-dispersive X-ray spectrometry (EDXS) is used in both SEM and TEM for chemical analysis of individual particles, while selected-area electron diffraction (SAED) pattern analysis in TEM can provide details of the cell unit of individual particles of mass down to  $10^{-15}$  g. It helps to differentiate between antigorite and chrysotile. Secondary ion mass spectrometry, laser microprobe mass spectrometry (LMMS), electron probe X-ray microanalysis (EPXMA), and X-ray photoelectron spectroscopy (XPS) are also analytical techniques used for asbestos chemical characterization.

### Sampling Methods

Sampling should normally be undertaken using procedures described in International Standards Organization (ISO) documents.

Moreover, careful cleaning of all glass and plastic equipment like glassware or bottles is very important, especially in the preparation of suitable blank control filters. There may be serious social and economic consequences from the finding of a few fine fibrils.

### Airborne Dust Sampling

Airborne dust is placed on a filter by drawing through it measured volumes (minimum  $2\text{ m}^3$ ) of air. Usually filters with a pore diameter of 0.2 to  $1, 2\text{ }\mu\text{m}$  are used for collecting airborne dust during

**Table 1** Mineral types and chemical formulae of the asbestos varieties

Mineral group	Asbestos variety	Nominal chemical formulae <sup>a</sup>
Serpentine	Chrysotile	$\text{Mg}_3\text{Si}_4\text{O}_{16}(\text{OH})_8^b$
Amphibole <sup>c</sup>	Amosite	$(\text{Fe}^{2+}\text{Mg})_7\text{Si}_8\text{O}_{22}(\text{OH})_2$
Amphibole	Crocidolite	$\text{Na}_2\text{Fe}^{3+}(\text{Fe}^{2+}\text{Mg})_3\text{Si}_8\text{O}_{22}(\text{OH})_2$
Amphibole	Anthophyllite	$\text{Mg}_7\text{Si}_8\text{O}_{22}(\text{OH})_2$
Amphibole	Actinolite	$\text{Ca}_2\text{Fe}_2^{2+}\text{MgSi}_8\text{O}_{22}(\text{OH})_2$
Amphibole	Tremolite	$\text{Ca}_2\text{Mg}_5\text{Si}_8\text{O}_{22}(\text{OH})_2$

<sup>a</sup>Most natural asbestos can deviate a little from this nominal composition.

<sup>b</sup>Minor substitution of silicon by  $\text{Al}^{3+}$  may occur, and minor substitution of magnesium by  $\text{Al}^{3+}$ ,  $\text{Fe}^{2+}$ ,  $\text{Fe}^{3+}$ ,  $\text{Ni}^{2+}$ ,  $\text{Mn}^{2+}$ , and  $\text{Co}^{2+}$  may also be present.

<sup>c</sup>In some varieties of amphibole the particular elements can be partially substituted by K, Li, Mn, Al, Cr, Ti, Pb, and Zn.

filtration. The type of filter used depends on the microscope used for further analysis.

### Water Sampling

Water samples are normally collected in clean, sterile, pretreated plastic bottles. If there is an excessive amount of suspended organic matter in the water, pretreatment of the samples using ultraviolet light and an ozone bubbler or sodium hypochlorite is necessary. Filtration is carried out using standard laboratory glassware, usually through a 25 mm-diameter filter of an appropriate material. The filter type depends on the microscope used for further analysis.

### Bulk Material Sampling

A solution or suspension must be prepared from bulk material or soil contaminated with asbestos fibers in a suitable liquid. In the case of contaminated soil a few kilograms should be taken as the minimum for analysis. Insoluble bulk samples of 2–3 g are ground to fine powder and mixed with distilled water. The way of filtration is the same as for simple water samples.

### Lung Tissue Sampling

Four to six grams of tissue blocks of formalin-fixed lung parenchyma are subjected to two different ashing techniques. To exclude any sampling errors, each tissue sample of peripheral disease-free lung is cut into 4–6 mm<sup>3</sup> cubes, which are mixed afterwards and randomly divided into two parts. One part is processed by wet ashing, the second one by hot ashing. In wet ashing 2–3 g lung tissue is digested in 45 ml of 13% solution with sodium hypochlorite (NaClO). In hot ashing it is heated in an oven at 600°C for 15 min. The ash is then suspended with distilled/deionized water and filtered as in water sampling. The hot ashing may have some drawback due to the fact that not all asbestos fibers can withstand a temperature of 600°C for 15 min and might break. Wet digestion cannot completely digest the tissue.

## Optical Microscopy

For OM examination tortuous-pore cellulose nitrate (CN), acetate, or mixed ester (CE) membrane filters with a pore size of 0.8 or 1.2 µm are recommended.

A low-power stereo binocular microscope is used to characterize and select suspected asbestos types fibers for further examination. Evaluation of the fibrous nature, color, luster, elasticity, flexibility, and other handling properties allows the detection and subsequent identification of asbestos in bulk materials at levels down to  $\sim 10 \text{ mg kg}^{-1}$ .

Microscope slides (72 mm × 25 mm × 0.8 mm approximately) and coverslips should be of the best optical quality and of the correct thickness for obtaining quality images (usually 0.17 mm). In the case of counting a particular fiber, an assumption has to be made about the cross-sectional profile of the fiber type; chrysotile is usually assumed to be cylindrical in shape, while the amphiboles are considered to have a thickness-to-width ratio of  $\sim 1.6:1$ . Discrimination of asbestos fibers using the morphology and refractive index in such cases is aided by prior identification of the fiber types present in the bulk material using PLM or other methods.

### Polarized-Light Microscopy

PLM is used to assess optical properties such as birefringence, pleochroism, and the extinction position. PLM can be used to differentiate asbestos from nonasbestos fibers.

### Phase-Contrast Optical Microscopy

Phase-contrast optical microscopy (PCOM) is used mainly for assessment of the refractive index in comparison with liquids of standard index. Next PCOM is widely used to measure fiber concentration. Before examination of the sample, acid treatment may be used to remove carbonates from cement matrices, or organic solvents may be necessary to remove plastic binder materials. OM operates at magnifications of only  $400\times$  and will not resolve fibers below 0.25 µm in diameter. Furthermore, PCOM cannot distinguish asbestos fibers from other fibers (e.g., gypsum, mineral wool, fiberglass, cellulose, etc.).

**Preparation of filters** Membrane filters of CE are placed on clean microscope slides and left to a short burst of acetone vapor. Then the filter is cleared by collapsing the filter structure. A drop or two of triacetin is applied to the filter surface, and a cover slip is carefully placed on it. The filter is ready for evaluation after a brief period of warming.

An optional procedure is to use an aqueous solution of dimethylformamide and acetic acid (so called Eukitt) for clearance of the filter structure, after which it is dried by warming at  $\sim 65^\circ\text{C}$  for a few minutes before a drop of Euparal resin is applied and the coverslip is gently placed on top.

**Fiber counting** Airborne fiber concentrations determined using OM refer strictly to those fibers that are longer than 5 µm, have diameters less than 3 µm, and have an aspect ratio of less than 3:1. The aspect ratio may be slightly different in some countries. All the

fibers are counted that are encountered during a systematic microscopic search of a small proportion of the filter. Thus PCOM is nonspecific for asbestos. The graticule (e.g., Walton–Beckett) is used to define a counting field and count fibers. Rules for split fibers, touching particles, fibers partially within the graticule, and other artifacts are also specified in the documentation of the method. Analysis using PCOM is limited to morphological observations and to some extent to refractive index assessment. Airborne fiber concentrations are calculated proportionately from the number of fibers counted, the effective area of the filter, the area searched, and the volume of air filtered and are usually expressed in fibers per milliliter or fibers per liter.

## Scanning Electron Microscopy

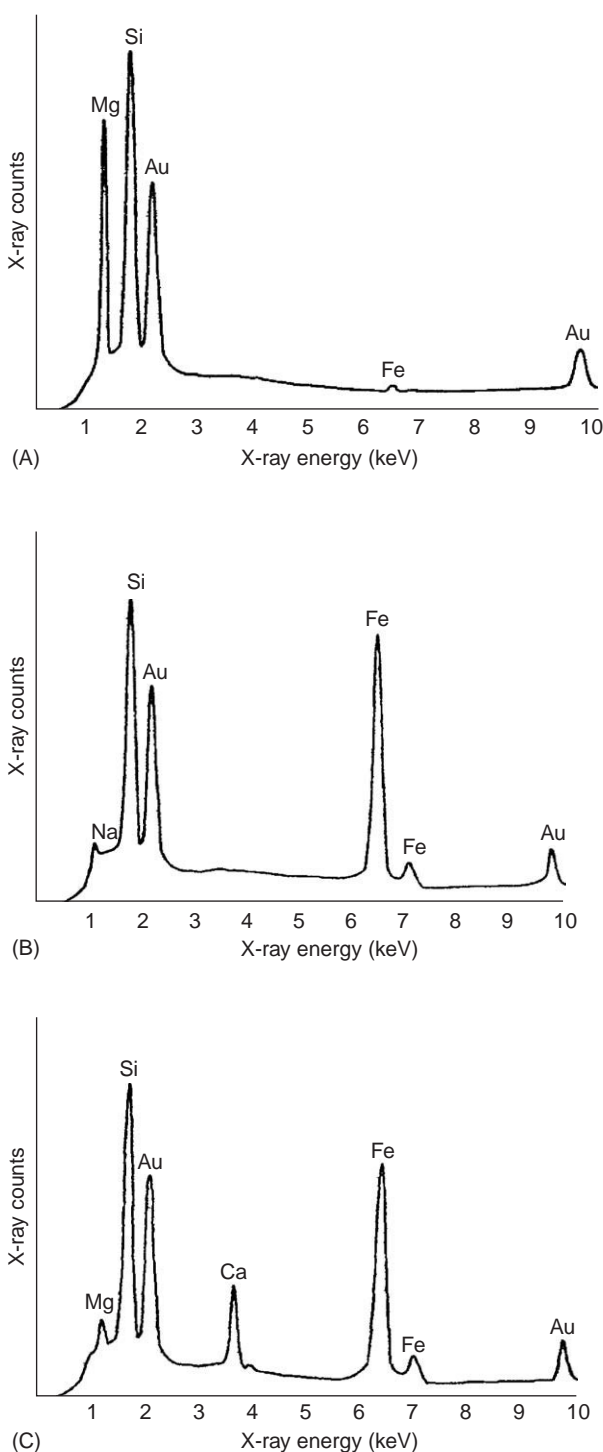
For SEM analysis a Nucleopore poly(carbonate) (PC) membrane filter (diameter 37 mm, pore size 0.2  $\mu\text{m}$ ) is used. After filtration each piece of the filter is coated with a very thin layer of gold. Routine SEM viewing is combined with an energy-dispersive X-ray microanalyzer (EDX). The particles and fibers are identified by their relative elemental peak intensities (based on magnesium, silicon, iron, and calcium), compared with their respective mineral standards placed in the computer bank (Figures 1–3).

### SEM Operating Conditions, Counting, Fiber Type Determination

SEM should be operated at magnifications of 2000–100 000 $\times$  with an acceleration voltage of 20–40 kV. Mineral fibers are counted on the screen. All inorganic particles having a length-to-width ratio at least 3:1 and parallel sides are defined as perfect fibers and counted. Cleavage fragments should not be counted. At a magnification of 5000 $\times$ , 13–25 inorganic fibers are counted or a minimum of 400 viewing fields are analyzed in order to reach an analytical sensitivity of 0.1 million fibers per gram for chrysotile. The fiber dimensions are measured with magnifications up to 100 000 $\times$ , and all inorganic fibers longer than 1  $\mu\text{m}$  are recorded.

EDX analysis is used to determine the fiber type. It should be carried out at higher magnifications (up to 50 000 $\times$ ). The absence of silicon or the presence of other elements such as aluminum or potassium can be used to demonstrate that a fiber is not asbestos, although care must be taken that there is no contamination of the fiber or of the EDX spectrum.

The proportions of the X-ray counts in the peaks of the spectra are not the same as the elemental proportions in the minerals.

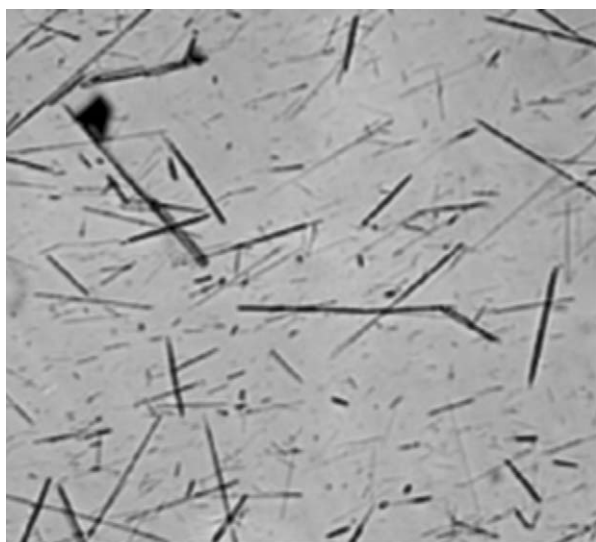


**Figure 1** EDX spectrum of different types of asbestos fibers: (A) chrysotile; (B) crocidolite; (C) actinolite. The presence of an Au peak is the result of the sample preparation – filter coating with gold. The conditions of measurement: SEM with EDX; magnification, 2000–50 000 $\times$ , acceleration voltage, 20–30 kV.

## Transmission Electron Microscopy

Capillary PC membrane filters of maximum pore size 0.4  $\mu\text{m}$  or CE or CN membrane filters of maximum





**Figure 2** An example of an OM micrograph of asbestos fibers.

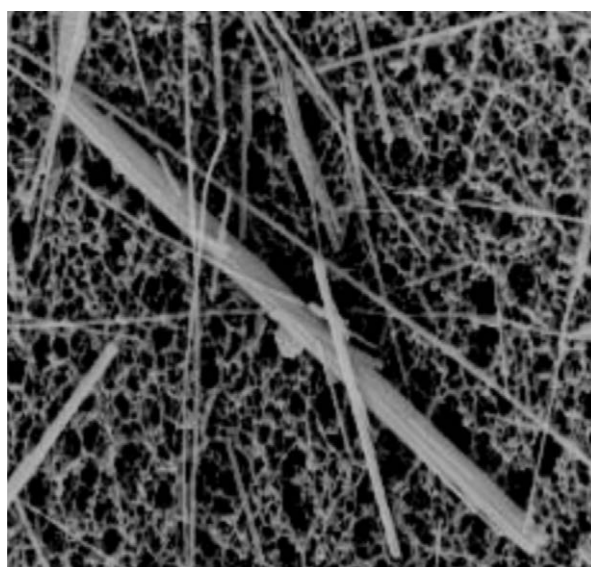
pore size  $0.8\text{ }\mu\text{m}$  are used for TEM analysis. Details of TEM are given in ISO standards.

#### TEM Filter Preparation

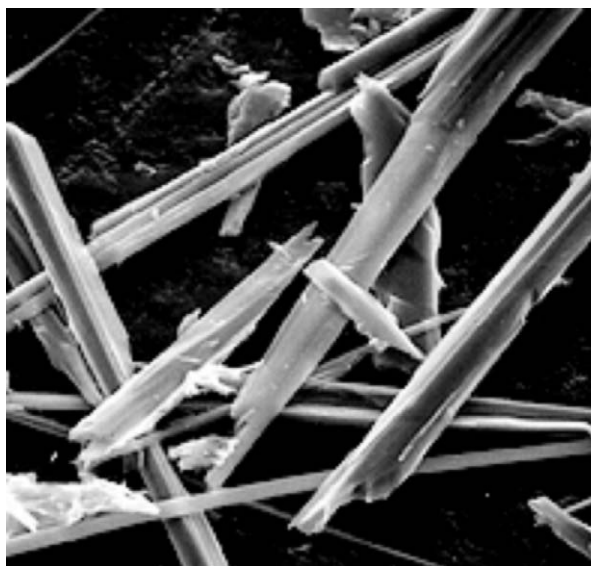
There are two ways of transferring the fibers from the filter to an electron-transparent membrane on a suitable TEM grid.

**Direct methods** In direct methods a PC filter, with fibers on it, is simply coated with carbon in a high-voltage carbon evaporation device. When CE or CN filters are used, they must be cleared and etched before coating with carbon. Care must be taken not to overheat the filter in the process by prolonging the arcing or by placing the filters too close to the arc as the filter may not then dissolve in the later stages of preparation. Small segments of carbon-coated filter,  $\sim 2$  or  $3\text{ mm}$  square, are cut from the filter and placed carbon side up on a TEM standard grid. Either this is placed on a Jaffe washer with a suitable solvent or a single drop of the solvent is carefully placed on the filter, where it is kept for  $\sim 30\text{ min}$ . Typical solvents are acetone for CE filters and chloroform or *N*-methyl-2-pyrrolidone for PC filters. TEM grids with the carbon film less than 75% of the grid squares should be rejected from the analysis.

**Indirect methods** Indirect methods are used where there is an excessive amount of organic fiber along with the asbestos, or some other soluble mineral fiber in an airborne sample. In these cases, treatment of the sample in a low-temperature oxygen plasma incinerator, or dissolution of the sample in a suitable solvent, followed by redeposition of the cleaned sample on a second filter can make the analysis



(A)



(B)

**Figure 3** Examples of SEM micrographs of asbestos fibers: (A) amphibole; (B) tremolite.

considerably easier. Subsequent preparation for TEM analysis follows the procedures of the direct method. Fiber counts for determination of regulatory asbestos fiber concentrations in the assessment of health risk will not usually be obtainable using indirect analysis methods. On the other hand, the indirect methods may be the only option for some types of environmental samples.

#### TEM Operating Conditions

TEM operating at an acceleration voltage of between 80 and 120 kV with a resolution of  $<1.0\text{ nm}$  and with a range of magnification from  $\sim 500\times$  to

$100\,000\times$  is routinely used. The viewing screen should have suitable calibrated markings to allow measurement of fibers directly on screen. The TEM should be capable of producing an electron diffraction (ED) pattern from a mineral particle of area  $0.6\,\mu\text{m}^2$  or less, selected from an image at  $20\,000\times$  magnification. A SAED aperture may be used to limit the area producing the pattern. A specimen holder that is capable of being tilted through  $+30^\circ$  to  $-30^\circ$  about an axis in the specimen plane combined with a rotation of  $360^\circ$  in the vertical axis is required for zone-axis ED.

The EDX system will normally be similar to that described for SEM for energy range and resolution. Additional specifications are required for asbestos analysis in that a background-subtracted sodium  $K\alpha$  peak from a 50 nm-diameter crocidolite fiber should have an integrated count rate of at least one count per second when irradiated by a probe of dimension 250 nm or smaller at an acceleration potential of 80 kV. The peak-to-ground ratio in this analysis should also be at least 1.0.

Analysis and inspection of SAED directly on screen is often enough for identification for chrysotile and a 'possible amphibole', but computer-aided measurement and analysis are required for detailed diffraction pattern indexing.

### Fiber Counting

The method of counting of asbestos fibers in TEM preparations differs from those of SEM or PCOM in that it does not rely on counting fibers within individual fields of view but depends on the systematic searching of complete grid openings in the TEM grid at a high magnification. The fiber lengths of interest are usually those longer than  $0.5\,\mu\text{m}$ , which for chrysotile fibrils implies a diameter range from 30 nm upward.

The rules for fiber or structure counting in TEM have gradually become more complicated with the recognition of the needs of different types, and matrices of different types containing various arrangements of fibers. In addition, there has been a need to recognize that the fiber numbers in such arrangements need to be counted carefully to provide reliable comparisons between samples and laboratories. It is necessary to read the ISO draft documents to count fibers properly.

### TEM Identification of Fibers

The identification of fibers as asbestos using TEM can be almost unambiguous, but it involves the most complex, difficult, and costly procedures. The process of identification involves three aspects of fiber

characteristics: morphology, crystal structure, and composition.

Fiber morphology is a very distinctive feature for chrysotile. Chrysotile fibrils are generally uniform in diameter, at  $\sim 25\,\text{nm}$ , and have a tubular structure that is almost unique and that appears, in a clear TEM image, as a lighter core to each dark fiber. Chrysotile is very susceptible to being damaged by the high energy of the electron beam, and so the tubular structure is easily destroyed and may not be visible. Similarly, chemical attack, even from quite dilute acid solutions, can leach magnesium from the fibrils and damage the tubular structure. The amphibole asbestos minerals are generally lath-shaped, parallel-sided, needle-like fibers. These usually show features such as extinction contours, multiple twin plane dislocations, and multiple chain defects (Wadsley defects), which are all characteristics of asbestos amphiboles rather than the more normal prismatic amphiboles. The amphiboles are very stable under an electron beam and are chemically very stable, and so they do not usually show any serious signs of damage.

EDX analysis in TEM is more definitive than in SEM and can be used to identify fiber chemistry quantitatively. The particles or fibers can be treated as thin films for the purposes of the calculation of atomic or mass fractions. This means that the effects of mass absorption and fluorescence can be ignored, although simple particle size effects need correction. In general the net integrated EDX peak areas can be treated as directly related to atomic or mass fractions with a proportionality factor derived from specific size ranges of similar reference minerals in an individual electron microscope. Such exact analysis is not always necessary, and simple comparison of spectra with those of reference minerals may be sufficient in many cases, although in this case the identification is recognized as being of a lower order of certainty than quantitative TEM analysis. Calibration of EDX equipment, as for SEM, is essential.

### TEM ED

Qualitative SAED consists of observation of the pattern of diffraction spots obtained on the TEM viewing screen from a randomly oriented fiber or particle. Such a pattern indicates that the material is crystalline. Chrysotile fibrils, with their cylindrical form, will usually give the same characteristic pattern, corresponding to a  $0.73\,\text{nm}$  spacing for (002) planes, and a layer line repeat of  $0.53\,\text{nm}$ , as well as streaking of the (110) and (130) reflections. These observations and measurements can be made directly from the screen if the appropriate calibrated screen



markings are available, but records should be made of a certain number of fibers in any sample. A range of other nonasbestos minerals can display a tubular morphology, including halloysite, palygorskite, talc, and vermiculite, and although their tubular forms are rare, their absence can only be proved using quantitative SAED or EDX.

SAED identification of amphibole asbestos fibers is only achieved in a formal sense by quantitative zone-axis interpretation, and even then it may not be possible in every case. Most randomly oriented fibers will show a diffraction pattern with a 0.53 nm layer spacing, but this is not by itself diagnostic for an amphibole. If multiple twinning, parallel to the *c* crystallographic axis, is present in the fiber, effectively resulting in several parallel crystals with different axial orientations, then apparently random spot distributions along the layer lines will be visible. If these observations can be made, then the fiber may be identified as an amphibole asbestos. If accurate measurement of the SAED pattern is to be made, then it is necessary to use an internal calibration standard such as a thin coating of gold on the underside of the TEM specimen. An Energy-dispersive spectrometer or electron probe X-ray microanalyzer, if present in the TEM, can be used as well.

## Scanning Transmission Electron Microscopy

STEM is a combination of scanning-transmission and transmission modes of operation in an electron microscope. Fibers are sought in the transmission mode at a magnification of 10 000–60 000 $\times$ ; this allows a maximum speed for the searching operation. Additionally, the ED pattern is generated in that mode. The scanning-transmission mode is used to measure the fiber dimensions at a magnification of up to 20 000 $\times$  on the monitor screen and to obtain the X-ray spectrum for elemental analysis.

## Secondary Ion Mass Analysis

The asbestos samples are prepared from a suspension in hexane spotted on indium foil. Ion images (from  $\sim 150\ \mu\text{m}$ ), mass spectra and the energy distribution of positive secondary ions are obtained. Either silicon or magnesium positive ion images may be used to indicate the asbestos fibers.

## X-Ray Photoelectron Spectroscopy

The asbestos samples are mounted with double-sided tape. The area of analysis is 600  $\mu\text{m}$ . An electron

flood gun operating at a few electronvolts and a nickel mesh covering the sample/sample holder arrangement are used for charge compensation. The Mg/Si concentration is calculated to recognize the fiber type.

## Laser Microprobe Mass Spectrometry

LMMS uses a pulsed beam of photons to evaporate a sample in a small region, as small as 0.5  $\mu\text{m}$  in diameter. A fraction of the evaporated atoms is ionized by the laser beam, accelerated to a kinetic energy of 3 keV, and then analyzed using a time-of-flight mass spectrometer. The instrument provides detection of all elements and isotopes, with detection limits in microgram per gram range from a total sample consumption of 0.1 pg per spectrum. Absolute detection limits are thus in the femtogram range. Quantitative analysis is based on the use of sensitivity factors, and the accuracy is limited to  $\sim 20\%$  relative. A sample for transmission geometry LMMS must be in the form of a particle or a thin film. Typical applications include analysis of particles as small as 200 nm in diameter.

## General Analytical Considerations

### Confidence Level of Identification

As is clear from the foregoing sections, there is a wide array of different observations and measurements available for every single fiber. Electron microscopy can be used to detect very thin fibers and allows us to identify different types of fibers. Therefore, nonasbestos fibers can be eliminated from the measurement. The disadvantage of the STEM method lies in the time and cost consideration. OM has more drawbacks, for example, a limited resolution. To be identified as chrysotile, fibers must exhibit their characteristic morphology and also contain magnesium and silicon, whereas anthophyllites have additional peaks of calcium or iron in their X-ray spectra. In several samples the identity of chrysotile and amphibole fibers is confirmed qualitatively using TEM (at 80 kV, magnification 20 000 $\times$ ) and SAED.

In counting thin chrysotile fibers, the SEM method may produce lower results than the TEM technique, but for coarser amphibole fibers such an effect is not likely. A large number of amphibole and nearly all chrysotile fibers are undetectable using OM. Definitive identification of a chrysotile can only be made if a characteristic SAED pattern is recorded for a tubular morphology fiber and the EDX spectrum is quantitatively consistent. Not all these elements may be either necessary or desirable for any or all fibers in every sample. This means that the level to which each fiber has been identified needs to be recorded to

**Table 2** Classification of fibers with tubular morphology

TM	Tubular morphology, not sufficiently characteristic for classification as chrysotile
CM	Characteristic chrysotile morphology
CD	Chrysotile SAED pattern
CQ	Chrysotile composition by quantitative EDXS
CMQ	Chrysotile morphology and composition by quantitative EDXS
CDQ	Chrysotile SAED pattern and composition by quantitative EDXS
NAM	Nonasbestos mineral

**Table 3** Classification of fibers without tubular morphology

AD	Amphibole fiber by random orientation SAED (shows layer pattern of 0.53 nm spacing)
AX	Amphibole by qualitative EDX. Spectrum has elemental components consistent with amphibole
ADX	Amphibole by random orientation SAED and qualitative EDX
AQ	Amphibole quantitative EDX
AZ	Amphibole by one zone-axis SAED pattern
ADQ	Amphibole by random orientation SAED and quantitative EDX
AZQ	Amphibole by one zone-axis SAED pattern and quantitative EDX
AZZ	Amphibole by two zone-axis SAED patterns with consistent interaxial angle
AZZQ	Amphibole by two zone-axis SAED patterns with consistent interaxial and quantitative EDX
NAM	Nonasbestos mineral
UF	Unidentified fiber

avoid differences and ambiguities in results and to allow interlaboratory comparisons to be made. **Table 2**, from the draft ISO standard, illustrates the different levels of possible identification that may be made for tubular fibers found in environment samples. It is important to recognize also the influence of the chemical history of each fiber in the identification since the chemistry of chrysotile is not stable, even in the mildest of acid conditions. The amphibole asbestos minerals have an even more complicated hierarchy of identifications based upon four levels of SAED analysis and three levels of EDX analysis (**Table 3**). The resulting 11 possible classifications for any fiber are made even more complicated by the need to incorporate the distinction between asbestos and nonasbestos varieties of the amphiboles.

XPS, EPXMA, and LMMS are especially used to characterize the surface of modified asbestos fibers. Alteration of asbestos surfaces by either acid leaching or deposition of a chemical coating reduces the hazards associated with asbestos exposure.

### Fiber Counting: Accuracy and Precision

There is no independent method available to determine the accuracy of fiber counts, although mass

estimates based upon fiber counts may be traceable to other mass estimates of standards. All the quantitative analyses described here rely upon procedures of fiber counting that have been shown to involve a series of uncertainties that require careful consideration in the interpretation of results.

If it is assumed that both the deposition of fibers on the filter and the selection of fields for counting are random, then there will be a variability in the results that can be described by a Poisson distribution. For fiber counts of  $\sim 50$  this leads to 95% confidence intervals in the results of  $\sim \pm 15$  (30%) and a coefficient of variation of  $\sim 15\%$ , although for smaller fiber counts the variability will be much larger. In addition to this variability there is an unavoidable degree of nonuniformity of the asbestos deposit on the filter. There is also subjective variation between microscopists in interpretation of fiber structures and in their ability to detect and identify fibers. Overall counting performances cannot then be expected to produce a coefficient of variation better than  $\sim 25\%$  for counts of  $\sim 50$  fibers.

Good quality control procedures are essential to keep these subjective variations to acceptably low levels, and satisfactory performance in both internal and external fiber counting exchange schemes is usually recommended by regulatory authorities.

### Sensitivity and Detection Limits

The analytical sensitivity of fiber counting methods is usually accepted as being a single fiber found in the course of a standard search. With conventional methods, this may translate to airborne fiber concentrations of  $\sim 0.1$  fiber per liter in ambient samples of  $\sim 1 \text{ m}^3$ , or  $\sim 1000$  fibers per liter for water samples of  $\sim 1 \text{ l}$ .

The detection limits are determined by the number of fibers in suitably prepared blank filters and by the random nature of the filter and of the search. In some electron microscopy preparations, where no fibers are found in the blank controls, the detection limit has been taken as 3.69 times the sensitivity (the upper 95% confidence limit for zero in a Poisson distribution). Otherwise, as in PCOM, where a nonzero blank control is common, a higher detection limit prevails. Thus a series of blank filters might produce an average fiber count of five fibers in a standard analysis, which for routine air samples ( $\sim 0.5 \text{ m}^3$ ) could translate into a detection limit of  $\sim 0.01$  fibers per milliliter.

*See also:* **Air Analysis:** Sampling. **Microscopy Applications:** Environmental. **Microscopy Techniques:** Light Microscopy; Specimen Preparation for Electron Microscopy; Scanning Electron Microscopy. **Surface Analysis:** Low

Energy Electron Diffraction. **Water Analysis:** Overview. **X-Ray Fluorescence and Emission:** Energy Dispersive X-Ray Fluorescence.

## Further Reading

- Addison J, Davies LST, McIntosh C, and Porteous RH (1992) *Bulk Asbestos Reference Minerals for Optical Microscope Identification: Preparation, Evaluation and Validation*, IOM report to client, HSE contract no. R2490/R48.44. Edinburgh: Institute of Occupational Medicine.
- Asbestos International Association (1984) *Method for the Determination of Airborne Asbestos Fibres and Other Inorganic Fibres by Scanning Electron Microscopy*, AIA Health and Safety Publication, Recommended Technical Method No. 2 (RTM2). London: AIA.
- BeruBe K, Mossman B, Quinlan T, and Taatjes D (1998) Diverse microscopy imaging techniques for studies of asbestos-induced lung disease. *Microscopy and Analysis*, European ed., 21–23, March.
- Chatfield EJ (1987) Asbestos measurements in building atmospheres. Sheridan park research community. Mississauga, Ontario: Ontario Research Foundation.
- Chatfield EJ and Dillon MJ (1985) Analytical method for the determination of asbestos fibres in water. *US Environmental Protection Agency*. Report on contract 68-03-2717. Athens GA: Environmental Research Laboratory.
- International Organization for Standardization (1982) *Determination of Airborne Inorganic Fiber Concentration in Workplace by Light Microscopy (Membrane Filter Method)*. ISO 8672.
- International Organization for Standardization (1991) *Ambient Air – Determination of Asbestos Fibres – Direct Transfer Transmission Electron Microscopy Procedure*. ISO/TC 146/SC 3/WG1 N28.
- International Organization for Standardization (1999) *Ambient Air – Determination of Asbestos Fibres – Indirect Transfer Transmission Electron Microscopy Method*. ISO 13794: 1999(E).
- Michaels L and Chissick SS (eds.) (1979) Properties, applications and hazards. *Asbestos*, vol. 1, New York: Wiley.

## Glossary

Amphibole	Group of rock-forming ferromagnesium silicate minerals, closely related in crystal form and composition.
The asbestos minerals	Naturally occurring fibrous, crystalline, silicate minerals belonging to the serpentine and amphibole groups.
Membrane filters	Filters of cellulose esters or polycarbonate with pore sizes of 0.2–1.2 µm. Used for filtration of asbestos-bearing solutions or suspensions.
Aspect ratio	Ratio of length to width of a particle.
Electron diffraction	Technique in electron microscopy by which the crystal structure of a specimen is examined.
Energy-dispersive X-ray analysis	Measurement of the energies and intensities of X-rays using a solid-state detector and multichannel analyzer system.
Fiber	Elongated particle that has parallel or stepped sides; for the purpose of ISO, a fiber is defined to have an aspect ratio equal to or greater than 5:1 and a minimum length of 0.5 µm.
Selected-area electron diffraction	Technique in electron microscopy in which the crystal structure of a small area of a sample is examined.
Serpentine	Group of common rock-forming minerals having the nominal formula.
Twinning	Occurrence of crystals of the same species joined together at a particular mutual orientation, such that the relative orientations are related by a definite law.
Zone-axis	Line or crystallographic direction through the center of a crystal that is parallel to the intersection edges of the crystal defining the crystal zone.

## ASV

See **VOLTAMMETRY: Anodic Stripping**

## ATMOSPHERIC ANALYSIS

See **AIR ANALYSIS: Sampling; Outdoor Air; Workplace Air**

# ATOMIC ABSORPTION SPECTROMETRY

Contents

**Principles and Instrumentation**

**Interferences and Background Correction**

**Flame**

**Electrothermal**

**Vapor Generation**

## Principles and Instrumentation

**E H Ivanova**, Institute of General and Inorganic Chemistry, Sofia, Bulgaria

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Atomic absorption spectrometry (AAS) is a technique in which free gaseous atoms absorb electromagnetic radiation at a specific wavelength to produce a corresponding measurable signal. The absorption signal is proportional to the concentration of the free atoms present in the optical path.

The physical phenomenon, first observed by Foucault in 1849, was converted into a popular analytical technique by Walsh essentially with the development of the hollow cathode lamp as the source of spectral energy for absorbance measurements. This permitted relatively cheap atomic absorption spectrometers to become available.

This article examines the basic theory of energy states of atoms, the quantitative analysis by atomic absorption, and the main components of the atomic absorption spectrometer.

### Atomic Structure and Spectra

For a strong absorption of electromagnetic radiation the lower energy state of the analyte atom must be highly populated and all selection rules must be observed. In order to examine these criteria, both energy states involved in the transition of the atom should be known. The upper energy state is not known for all elements, but even in such cases important conclusions may be drawn from the spectral term of the lower energy state, which is usually the ground state of the atoms. Therefore, the derivation

of the electronic configuration of the atoms will be discussed.

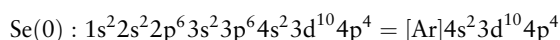
The distribution of electrons in an atom is governed by two atom-building principles:

- The orbitals (energy levels) are occupied in order of increasing orbital energy.
- All electrons in an atom must be present in a different microstate, i.e., the electrons must be distinguishable at least in one of their quantum numbers (Pauli's exclusion principle).

The principal quantum number  $n$  defines the shell in which the electron is located; the maximum number of electrons taken up in a shell is  $2n^2$ . Within the shell the electrons reside in orbitals of different symmetry, described by the angular momentum quantum number  $l$ , which can take values of 0, 1, 2, ...,  $n$ . An orbital can accommodate up to two electrons having opposite spins. Each of the electrons is characterized by the inner quantum number  $j$  that can take values of  $l \pm 1/2$ .

The general atom-building principles (a) and (b) are supplemented by rules that depend on the magnetic and chemical behavior of the elements:

- The special stability of the electronic configuration of the noble (inert) gases allows the classification of electrons into two main groups: core electrons, present in a noble gas shell, and valence electrons, present in subshells. Core electrons are not involved in chemical reactions and do not contribute to the generation of spectroscopic terms, which means that the elucidation of electronic configurations is greatly simplified, e.g.,



- The special stability of the half-filled and completely filled d and f subshells has the consequence that the configurations predicted by the rules (a) and (b) only are less stable than the configuration

predicted by the supplementary rule (d):

$$\text{Cr}(0) = [\text{Ar}]4s^23d^4 = [\text{Ar}]4s^13d^5 = [\text{Ar}]3d^54s^1$$

$$\text{Cu}(0) = [\text{Ar}]4s^23d^9 = [\text{Ar}]4s^13d^{10} = [\text{Ar}]3d^{10}4s^1$$

$$\text{Gd}(0) = [\text{Xe}]6s^24f^8 = [\text{Xe}]6s^24f^75d^1 = [\text{Xe}]4f^75d^16s^2$$

After applying the atom-building principles and the additional rules (c) and (d), the configuration obtained often does not contain the subshells in order of increasing principal quantum number. Standardization is easily achieved by altering the position of the subshells in the written configurations. A physicochemical reason for the natural order of principal quantum number  $n$  is given by the fact that the electrons with higher  $n$ -values ionize first.

In order to elucidate the spectroscopic terms belonging to a given electronic configuration, the first step is to ascertain the electron configuration using the atom-building principles (a), (b), and the additional rules (c), (d). The number of Pauli-allowed combinations  $N_{\text{ms}}$  (number of microstates) should be also calculated so that the correctness of the terms obtained can be checked. For a subshell partly filled with electrons,  $N_{\text{ms}}$  is given by the expressions:

$$N_{\text{ms}}(nl^N) = \frac{1 \times 2 \times \dots \times (4l+2)}{(1 \times 2 \times \dots \times N)(1 \times 2 \times \dots \times (\langle 4l+2 \rangle - N))} \quad \text{for } 1 \leq N \leq (4l+1) \quad [1]$$

where  $N$  is the number of electrons in the orbitals.

$$N_{\text{ms}}(nl^{(4l+2)}) = 1 \quad \text{for completely filled shells} \quad [2]$$

If more than one partially filled subshell is present, the product of the  $N_{\text{ms}}$  values calculated for all these subshells gives the number of microstates of the configuration:

$$N_{\text{ms}}(nl^N, n'l'^{N'}, \dots) = N_{\text{ms}}(nl^N) \times N_{\text{ms}}(n'l'^{N'}) \times N_{\text{ms}}(\dots) \quad [3]$$

An example is the excited configuration of Be, ..., Ba ( $s^1p^1$ ):

$$N_{\text{ms}} = \frac{1 \times 2}{1 \times (1)} \times \frac{1 \times 2 \times \dots \times 6}{1 \times (1 \times 2 \times \dots \times 5)} = 2 \times 6 = 12$$

Another example is the ground configuration of Gd ( $6s^24f^75d^1$ ):

$$\begin{aligned} N_{\text{ms}} &= 1 \times \frac{1 \times 2 \times \dots \times 14}{1 \times 2 \times \dots \times 7 \times (1 \times 2 \times \dots \times 7)} \\ &\quad \times \frac{1 \times 2 \times \dots \times 10}{1 \times (1 \times 2 \times \dots \times 9)} \\ &= 1 \times 3432 \times 10 = 34\,320 \end{aligned}$$

These examples show clearly that the number of microstates is small for a few valence electrons in the  $s$  and  $p$  subshells but becomes large for more electrons in the  $d$  and  $f$  subshells.

**Table 1** lists the electronic configurations and some related properties of the elements. In the selection of the elements, two criteria were observed:

- The resonance line, i.e., the line due to the transition between the ground state and the lowest excited state, must be situated within the spectral range of standard atomic absorption spectrometers (190–860 nm).
- The characteristic concentration, i.e., the concentration yielding 1% absorption (or 0.0044 absorbance), must be lower than  $100 \text{ mg l}^{-1}$ .

The excitation energies of the noble gases, the halogens, and sulfur are so high that the corresponding resonance lines are situated in the vacuum ultraviolet (UV) region, where oxygen intensively absorbs. In some cases this nonspecific absorption can be reduced by the use of a shielding gas (e.g., Ar or  $\text{N}_2$ ).

The elements can be classified into four spectrochemical groups on the basis of  $E_1$  and  $D_0$ : (a)  $E_1 \geq 7.0$  and  $D_0 \leq 4.2$ ; (b)  $E_1 \leq 7.0$  and  $D_0 \leq 4.2$ ; (c)  $E_1 \leq 7.0$  and  $D_0 \geq 4.2$ , and (d)  $E_1 \geq 7.0$  and  $D_0 \geq 4.2$ .

The number of microstates  $N_{\text{ms}}$  resulting from the corresponding electron configuration varies between one for electron configurations with closed valence shells and 34 320 for Gd.  $N_{\text{ms}}$  can serve as a measure of the complexity of the atomic absorption (and emission) spectra. An element with a large number of microstates also has a large number of spectroscopic terms and atomic lines, because of the many different term combinations possible.

According to the selection rules, transitions are allowed for which the angular momentum quantum number  $l$  increases by one unit, while the principal quantum number can change by any amount. The allowed transitions of the electrons can be compiled in term series in which the principal quantum number  $n$  precedes the term symbol as a number. The series in which  $l = 0, 1, 2, 3$  are designed by the letters  $s, p, d, f$ . Through the spin of the electrons and the associated magnetic field, splitting of the energy levels takes place, described by the inner quantum number  $j$ , which results in a fine multiplet structure of the spectral lines.

The energy levels can only absorb well-defined amounts of energy, i.e., they are quantized according to the symmetry rules. The most stable electronic configuration of an atom that has the lowest energy is the ground state. For example, the electronic configuration of the sodium atom is  $1s^22s^22p^63s^1$  (ground state with energy  $E_0$ ). The transition

**Table 1** Some spectrochemical characteristics of the elements

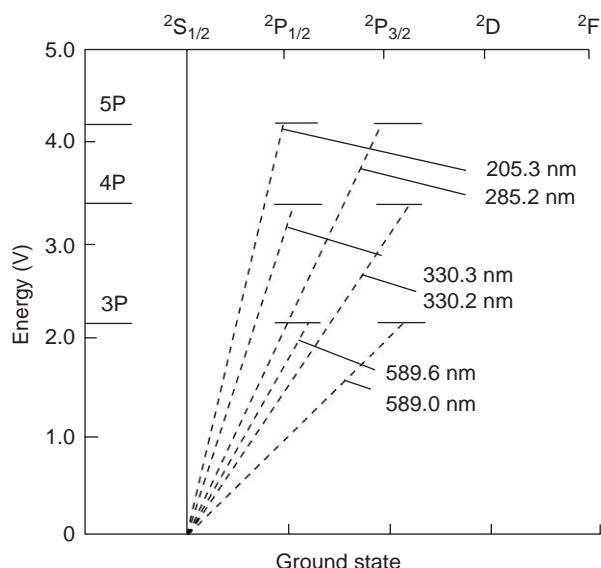
<i>Z</i>	<i>Symbol</i>	<i>Electron configuration</i>	<i>N<sub>ms</sub></i>	<i>E<sub>i</sub></i> (eV)	<i>D<sub>0</sub></i> (eV)
3	Li	1s <sup>2</sup> 2s <sup>1</sup>	2	5.392	3.46
4	Be	[He]2s <sup>2</sup>	1	9.322	4.51
5	B	[He]2s <sup>2</sup> 2p <sup>1</sup>	6	8.298	8.38
11	Na	[Ne]3s <sup>1</sup>	2	5.139	2.65
12	Mg	[Ne]3s <sup>2</sup>	1	7.646	3.76
13	Al	[Ne]3s <sup>2</sup> 3p <sup>1</sup>	6	5.986	5.31
14	Si	[Ne]3s <sup>2</sup> 3p <sup>2</sup>	15	8.151	8.29
15	P	[Ne]3s <sup>2</sup> 3p <sup>3</sup>	20	10.49	6.21
19	K	[Ar]4s <sup>1</sup>	2	4.341	2.89
20	Ca	[Ar]4s <sup>2</sup>	1	6.113	4.17
21	Sc	[Ar]3d <sup>1</sup> 4s <sup>2</sup>	10	6.54	7.06
22	Ti	[Ar]3d <sup>2</sup> 4s <sup>2</sup>	45	6.82	6.97
23	V	[Ar]3d <sup>3</sup> 4s <sup>2</sup>	120	6.74	4.45
24	Cr	[Ar]3d <sup>5</sup> 4s <sup>1</sup>	504	6.766	4.45
25	Mn	[Ar]3d <sup>5</sup> 4s <sup>2</sup>	252	7.435	4.18
26	Fe	[Ar]3d <sup>6</sup> 4s <sup>2</sup>	210	7.870	4.05
27	Co	[Ar]3d <sup>7</sup> 4s <sup>2</sup>	240	7.86	3.99
28	Ni	[Ar]3d <sup>8</sup> 4s <sup>2</sup>	45	7.635	3.96
29	Cu	[Ar]3d <sup>10</sup> 4s <sup>2</sup>	2	7.726	2.79
30	Zn	[Ar]3d <sup>10</sup> 4s <sup>2</sup>	1	9.394	<2.3
31	Ga	[Ar]3d <sup>10</sup> 4s <sup>2</sup> 4p <sup>1</sup>	6	5.999	3.66
32	Ge	[Ar]3d <sup>10</sup> 4s <sup>2</sup> 4p <sup>2</sup>	15	7.899	6.83
33	As	[Ar]3d <sup>10</sup> 4s <sup>2</sup> 4p <sup>3</sup>	20	9.81	4.99
34	Se	[Ar]3d <sup>10</sup> 4s <sup>2</sup> 4p <sup>4</sup>	15	9.752	4.82
37	Rb	[Kr]5s <sup>1</sup>	2	4.177	2.65
38	Sr	[Kr]5s <sup>2</sup>	1	5.695	4.41
39	Y	[Kr]4d <sup>1</sup> 5s <sup>2</sup>	10	6.38	7.46
40	Zr	[Kr]4d <sup>2</sup> 5s <sup>2</sup>	45	6.84	8.04
41	Nb	[Kr]4d <sup>4</sup> 5s <sup>1</sup>	420	6.88	8.00
42	Mo	[Kr]4d <sup>5</sup> 5s <sup>1</sup>	504	7.099	5.81
43	Tc	[Kr]4d <sup>5</sup> 5s <sup>2</sup>	252	7.28	—
44	Ru	[Kr]4d <sup>7</sup> 5s <sup>1</sup>	240	7.37	5.48
45	Rh	[Kr]4d <sup>8</sup> 5s <sup>1</sup>	90	7.46	4.20
46	Pd	[Kr]4d <sup>10</sup>	1	8.34	3.95
47	Ag	[Kr]4d <sup>10</sup> 5s <sup>1</sup>	2	7.576	2.28
48	Cd	[Kr]4d <sup>10</sup> 5s <sup>2</sup>	1	8.993	2.44
49	In	[Kr]4d <sup>10</sup> 5s <sup>2</sup> 5p <sup>1</sup>	6	5.786	<3.3
50	Sn	[Kr]4d <sup>10</sup> 5s <sup>2</sup> 5p <sup>2</sup>	15	7.344	5.51
51	Sb	[Kr]4d <sup>10</sup> 5s <sup>2</sup> 5p <sup>3</sup>	20	8.641	4.50
52	Te	[Kr]4d <sup>10</sup> 5s <sup>2</sup> 5p <sup>4</sup>	15	9.009	3.90
55	Cs	[Xe]6s <sup>1</sup>	2	3.894	2.31
56	Ba	[Xe]6s <sup>2</sup>	1	5.212	6.21
57	La	[Xe]5d <sup>1</sup> 6s <sup>2</sup>	10	5.577	8.28
59	Pr	[Xe]4f <sup>3</sup> 6s <sup>2</sup>	364	5.42	7.82
60	Nd	[Xe]4f <sup>4</sup> 6s <sup>2</sup>	1001	5.49	7.29
62	Sm	[Xe]4f <sup>6</sup> 6s <sup>2</sup>	3003	5.63	5.85
63	Eu	[Xe]4f <sup>7</sup> 6s <sup>2</sup>	3432	5.67	4.99
64	Gd	[Xe]4f <sup>7</sup> 5d <sup>1</sup> 6s <sup>2</sup>	34320	6.14	7.46
65	Tb	[Xe]4f <sup>9</sup> 6s <sup>2</sup>	2002	5.85	7.39
66	Dy	[Xe]4f <sup>10</sup> 6s <sup>2</sup>	1001	5.93	6.29
67	Ho	[Xe]4f <sup>11</sup> 6s <sup>2</sup>	364	6.02	6.33
68	Er	[Xe]4f <sup>12</sup> 6s <sup>2</sup>	91	6.10	6.37
69	Tm	[Xe]4f <sup>13</sup> 6s <sup>2</sup>	14	6.18	5.20
70	Yb	[Xe]4f <sup>14</sup> 6s <sup>2</sup>	1	6.254	4.12
71	Lu	[Xe]4f <sup>14</sup> 5d <sup>1</sup> 6s <sup>2</sup>	10	5.426	7.03
72	Hf	[Xe]4f <sup>14</sup> 5d <sup>2</sup> 6s <sup>2</sup>	45	7.0	8.31
73	Ta	[Xe]4f <sup>14</sup> 5d <sup>3</sup> 6s <sup>2</sup>	120	7.89	8.28
74	W	[Xe]4f <sup>14</sup> 5d <sup>4</sup> 6s <sup>2</sup>	210	7.98	6.96
75	Re	[Xe]4f <sup>14</sup> 5d <sup>5</sup> 6s <sup>2</sup>	252	7.88	6.50
76	Os	[Xe]4f <sup>14</sup> 5d <sup>6</sup> 6s <sup>2</sup>	210	8.7	6.20
77	Ir	[Xe]4f <sup>14</sup> 5d <sup>7</sup> 6s <sup>2</sup>	120	9.1	4.30

Continued

Table 1 Continued

Z	Symbol	Electron configuration	$N_{ms}$	$E_1$ (eV)	$D_0$ (eV)
78	Pt	[Xe]4f <sup>14</sup> 5d <sup>9</sup> 6s <sup>1</sup>	20	9.0	4.09
79	Au	[Xe]4f <sup>14</sup> 5d <sup>10</sup> 6s <sup>1</sup>	2	9.225	2.30
80	Hg	[Xe]4f <sup>14</sup> 5d <sup>10</sup> 6s <sup>2</sup>	1	10.44	2.29
81	Tl	[Xe]4f <sup>14</sup> 5d <sup>10</sup> 6s <sup>2</sup> 6p <sup>1</sup>	6	6.108	—
82	Pb	[Xe]4f <sup>14</sup> 5d <sup>10</sup> 6s <sup>2</sup> 6p <sup>2</sup>	15	7.416	3.96
83	Bi	[Xe]4f <sup>14</sup> 5d <sup>10</sup> 6s <sup>2</sup> 6p <sup>3</sup>	20	7.289	3.50
92	U	[Rn]5f <sup>3</sup> 6d <sup>1</sup> 7s <sup>1</sup>	3640	—	7.87

Elements with resonance lines between 193.70 nm (As) and 852.11 nm (Cs) and characteristic concentrations  $c_{1\%} < 100 \text{ mg l}^{-1}$  were selected. Z is the number of electrons,  $N_{ms}$  the number of microstates,  $E_1$  the first ionization energy, and  $D_0$  the bond dissociation energy of the corresponding monoxide MO.



**Figure 1** Partial Grotrian diagram for sodium. (Reprinted from Robinson JW (1990) *Atomic Absorption Spectroscopy* by courtesy of Marcel Dekker Inc.)

between the 3s orbital and a p orbital can be realized by absorption of light of definite wavelength, as illustrated in the partial term (Grotrian) diagram of sodium shown in **Figure 1**. For the sake of clarity, many of the upper-state transitions are omitted.

The absorption lines due to the transition between the ground state and the lowest excited state of the atom are referred to as the resonance lines. Sometimes alternate lines are used in AAS to extend the working range to higher concentrations or to avoid spectral interferences (seldom in AAS).

To realize the first transition 3s → 3p, the frequency of the incident light ( $\nu_1$ ) should correspond to:

$$\nu_1 = \frac{E_1^* - E_0}{h}$$

where  $E_1^*$  is the energy of the first excited state,  $E_0$  is the energy of the ground state, and  $h$  is Planck's constant.

The intensity of the incident light will decrease as part of it will be absorbed by the atoms. The electron

resides in the excited state  $E_1^*$  for only  $10^{-7}$  to  $10^{-9}$  s, after which it returns to its stable ground state by emitting radiant energy of the same frequency  $\nu_1$ .

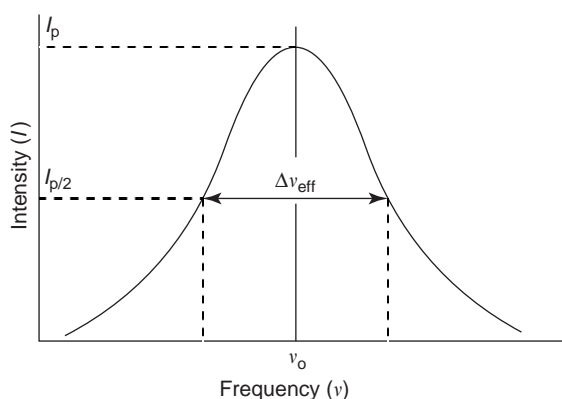
## Spectral Line Width

It follows from the above discussions that the spectral width of the light emitted by the atoms should be infinitely small. Correspondingly, the atoms should absorb only light of a definite frequency. Actually, a band of intensity distribution close to the Gaussian (**Figure 2**) is emitted or absorbed, respectively. The profile of the line is characterized by the central frequency  $\nu_0$ , the peak intensity  $I_p$ , and the frequency distribution with a width of  $\Delta\nu_{\text{eff}}$ , generally quoted as the full-width at half-maximum  $I_{p/2}$  (FWHM).

The broadening of the spectral lines is due to three reasons:

1. The natural broadening depends on the lifetime of the excited state and is usually in the range  $10^{-5}$  to  $10^{-3}$  nm.
2. The Doppler broadening ( $D$ ) is proportional to the absolute temperature ( $T$ ):  $D \sim \sqrt{T}$ . It is due to the random movement of the atoms and leads to a symmetrical broadening of the spectral line. Within the range 20–3000°C, the Doppler broadening varies between  $10^{-5}$  and  $10^{-2}$  nm.
3. The collisional broadening depends on the number of collisions of the emitting or absorbing atoms with other particles. After each collision the energy of the atom changes resulting in line broadening. In most cases the energy of the atom decreases after collision, leading to an additional shift of the line profile. The shift is proportional to the pressure of the surrounding particles. A pressure change in the range 0.01–760 mm Hg leads to a collisional broadening of  $10^{-6}$  to  $10^{-2}$  nm.

The spectral line broadening is a sum of the natural, Doppler, and collisional broadening; however, the natural broadening is relatively small and can be neglected.



**Figure 2** Profile and characteristics of a spectral line.

## Quantitative Analysis by Atomic Absorption

The atomic absorption process consists of the following steps: light at a specific wavelength of initial intensity  $I_0$  passes through the absorbing layer containing ground-state analyte atoms. The initial light intensity decreases by an amount determined by the atom concentration in the absorbing layer and the reduced intensity  $I$  is measured. The absorbance  $A$ , i.e., the logarithm of the ratio between the initial light intensity and the reduced intensity, is proportional to the concentration of free analyte atoms in the absorbing layer according to the Beer–Lambert law:

$$A = \log I_0/I = abC$$

where  $a$  is the absorption coefficient, a constant characteristic for the absorbing species at a specific wavelength,  $b$  is the length of the absorbing layer, and  $C$  is the concentration of the absorbing species.

Several related terms are used to define the amount of light absorption: the transmittance  $T = I/I_0$ , also expressed in percentage terms,  $T\% = 100 \times I/I_0$  and the percent absorption  $A\% = 100 - T\%$ , which is the complement of percent transmission.

For analytical purposes the absorbance  $A$  is mostly used, which is proportional to the concentration of free gaseous atoms of the analyte under definite experimental conditions. The formation of free gaseous atoms from the analyte species present in the sample solution involves several processes, e.g., desolvation, volatilization, dissociation of the chemical compounds. The efficiency of these processes for the given experimental conditions is included in the calibration equation. The calibration relationship is established by measuring the absorbances of standard solutions containing known amounts of analyte and plotting the absorbance data against concentration. Over the region where the Beer–Lambert law is

obeyed, the calibration yields a straight line. A deviation from linearity is usually observed at absorbances greater than 0.5–1.0.

After calibration is established, the absorbance of solutions of unknown analyte concentrations may be measured and the concentration may be determined from the calibration curve.

## Typical Performance Characteristics of AAS

### Sensitivity and Characteristic Concentration

The sensitivity of the AAS determination is defined by the slope of the calibration curve in its initial straight part. A convenient characteristic of the sensitivity is the characteristic concentration – the analyte concentration that produces 1% absorption (or 0.0044 absorbance) signal. The characteristic concentrations of the elements for flame atomizers are between 0.01 and 10 mg l<sup>-1</sup>, for electrothermal atomizers – ~2–3 orders of magnitude lower. Factors affecting the sensitivity are: the oscillator strength (the probability) of the corresponding electron transition, the type of the atomizer – flame or electrothermal – and the efficiency of atomization.

### Detection Limits

The term detection limit incorporates a consideration of both signal size and baseline noise to give an indication of the lowest concentration of an element that can be measured with a given statistical certainty. According to IUPAC, the detection limit is defined as the concentration that will give an absorbance signal three times the magnitude of the baseline noise. The latter can be statistically quantitated typically by making 10 or more replicate measurements of the baseline absorbance signal of an analytical blank, and determining the standard deviation of the measurements. The detection limit is then defined as the concentration that will produce an absorbance signal three times the standard deviation of the blank.

Practically, routine analytical work should be limited to concentrations several times higher than the detection limit.

### Precision and Accuracy

The precision is typically 0.3–1% at absorbances larger than 0.1 or 0.2 for flame atomization and 1–5% for electrothermal atomization. The accuracy is generally limited by random error and noise to 0.5–5%, but may be affected by systematic errors due to spectral and chemical interferences.



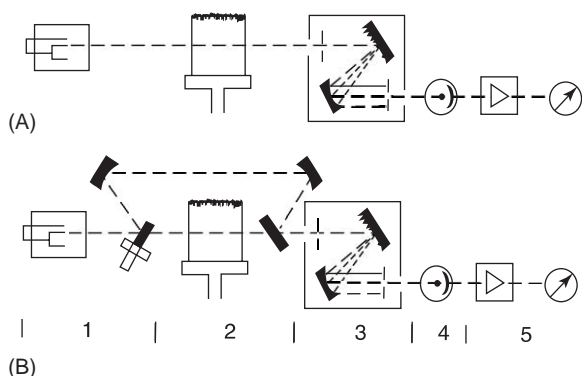
## Instrumentation

The basic components of an atomic absorption spectrometer are schematically presented in Figure 3 and are discussed below:

1. The light source emits the atomic lines of the analyte element. The source radiation is modulated to provide a means of selectively amplifying the light emitted by the source lamp and discriminating the emission from the sample cell.
2. In the sample cell an atomic vapor must be generated. To this purpose the sample is introduced into the atomizer – a flame system or electrically heated furnace aligned in the optical path of the spectrometer.
3. A monochromator is used to disperse the various wavelengths of light that are emitted from the source and to isolate the particular line of interest. This allows a selected element to be determined in the presence of others.
4. The wavelength isolated by the monochromator is directed onto the detector. This is normally a photomultiplier tube producing electrical current proportional to the light intensity.
5. The electrical current is then amplified and processed by the instrument electronics to produce the readout signal which is a measure of the light attenuation occurring in the sample cell.

## Optical Systems

Optical systems corresponding to single-beam or double-beam spectrometers can be distinguished in



**Figure 3** Principle of construction of atomic absorption spectrometers. (A) Single-beam spectrometer with electrically modulated lamp radiation; (B) double-beam spectrometer with reflection and splitting of the primary radiation by a rotating, partially mirrored quartz disk (chopper). 1 – radiation source, 2 – sample cell (atomizer), 3 – monochromator, 4 – detector, 5 – electronics and readout (by permission of Wiley-VCH from Welz B and Sperling M (1999) *Atomic Absorption Spectrometry*, 3rd, completely revised edition. Weinheim: Wiley-VCH).

AAS. In single-beam instruments (Figure 3A), all measurements are based on the varying intensity of a single beam of light in a single optical path. These instruments have fewer optical components and thus have the advantage of low light attenuation. Therefore, they do not need a high electronic amplification of the signal. Single-beam atomic absorption spectrometers provide, however, no means to compensate for instrumental variations during the analysis, such as drift in source intensity.

In double-beam systems (Figure 3B), the light from the radiation source is split by a beam splitter into two beams – a sample beam directed through the sample cell and a reference beam directed around the sample cell. The reference beam serves as a monitor of the lamp intensity and the response characteristics of electronic circuitry. The observed absorbance, determined from the ratio of sample beam and reference beam readings, is thus free of effects due to drifting lamp intensities.

There are several alternative system designs that combine the advantages of double-beam and single-beam optical systems. One such design uses two mechanically adjusted mirrors to alternately direct the entire output of the source through either the sample path or a reference path.

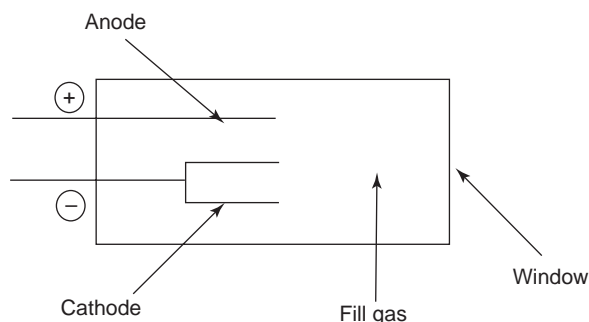
By utilizing the Zeeman effect it is also possible to obtain an (optical) single-beam instrument with double-beam characteristics. When the magnetic field at the atomizer is switched on, the analyte ions cannot absorb the radiation from the source, while they can absorb normally when the magnetic field is switched off. Since the radiation with and without magnetic field has the same intensity, drifts in the source or detector are eliminated, as with a double-beam instrument.

## Radiation Sources

As radiation sources in AAS, those line sources are mainly used that emit the spectral lines of one or more elements. Line sources make it possible to use conventional instead of high-resolution monochromators, as the monochromator only has to isolate the line of interest from other lines (mainly lamp fill gas lines). Hollow cathode lamps and electrodeless discharge lamps are the main types of lamps employed.

Hollow cathode lamps (HCLs) are available for most of the elements determinable by AAS. Figure 4 shows the schematic diagram of an HCL.

The cathode of the lamp is a hollow cylinder of the metal to be determined. The anode and cathode are sealed in a glass cylinder normally filled with either Ne or Ar at low pressure. At the end of the glass



**Figure 4** Schematic diagram of a hollow cathode lamp. (Reprinted from Robinson JW (1990) *Atomic Absorption Spectroscopy* by courtesy of Marcel Dekker Inc.)

cylinder is a window transparent to the emitted radiation. When an electrical potential is applied between the anode and cathode, some of the fill gas atoms are ionized. The positively charged fill gas ions, accelerated through the electrical field, bombard the negatively charged cathode by dislodging (sputtering) metal atoms from it. Sputtered metal atoms are then excited to emit their spectrum through kinetic transfer by impact with fill gas ions.

In electrodeless discharge lamps (EDLs) a small amount of the metal to be determined or its volatile compound is sealed inside a quartz bulb placed inside a small radio-frequency (RF) generator. When power is applied to the generator, an RF field is created and a low-pressure plasma is generated which vaporizes and excites the atoms inside the bulb, causing them to emit their spectrum. EDLs are particularly advantageous for volatile elements and for elements emitting in the far UV range.

Continuum sources emit radiation continuously distributed over a greater wavelength range. In AAS, continuum sources are used mainly for background measurements and correction. The deuterium lamp emits radiant power in the short wavelength range from  $\sim 190$  to  $330$  nm, while the halogen lamp covers the spectral range above  $300$  nm.

### The Monochromator System

The major components of the monochromator are the slits and the dispersion element. The source radiation falls on the entrance slit and is directed to the dispersion element that is based on either reflection or refraction. Only the light of the desired wavelength passes through the exit slit to fall onto the detector. The monochromator plays an important role in determining the baseline noise in an atomic absorption spectrometer since it defines the amount of light energy reaching the detector.

### Detection of Radiation

The photomultiplier tube is a radiation detector using the photoelectric principle. The radiation falling on a photocathode causes the emission of primary electrons that are released into the surrounding vacuum. As a result of the applied voltage between the electrodes (dynodes), the primary electron is accelerated so that when it strikes a dynode several secondary electrons are emitted, leading to a cascade effect.

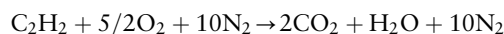
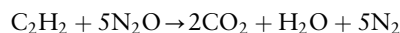
### Atomizers

As discussed above, for the measurement of atomic absorbance free ground-state atoms must be present in the sample cell. With the exception of mercury, free atoms do not exist in a stable form at room temperature. The atomizer is the place where free atoms of the analyte are created. The number of free atoms depends on the dynamic equilibrium between the number created and the number lost, e.g., by compound formation and ionization. The bond dissociation energy of the corresponding monoxide  $D_0$  and the first ionization energy  $E_1$  (shown in Table 1) are among the factors affecting this equilibrium.

The choice of the suitable atomizer for a given analytical task depends on the concentration of the analyte in the sample, the analyte itself, and the amount of sample available. In the following, the main atomizer types will be briefly presented.

**Flame atomizer** The flame is the oldest method of obtaining a population of free atoms in an atomic absorption instrument. A liquid sample is nebulized to form an aerosol of fine liquid droplets, which is introduced into the base of the flame where the solvent evaporates. Higher in the flame the sample residue decomposes and free atoms are generated.

In routine flame AAS work only the air/acetylene flame and the nitrous oxide/acetylene flame, both premixed laminar flames, are used. The combustion of acetylene with nitrogen oxide gives a considerably higher temperature ( $3000$  K) than the air/acetylene reaction ( $2500$  K).  $N_2O$  is kinetically very stable at room temperature but reacts with  $C_2H_2$  very fast at elevated temperatures by releasing its own chemical energy of  $83.8 \text{ kJ mol}^{-1}$ , thus giving a higher flame temperature. Furthermore, only  $5 \text{ mol } N_2$  per mol of  $C_2H_2$  must be heated to the corresponding temperature, compared to  $10$  for the air/acetylene mixture as detailed in the reactions below:



The main restriction of the flame atomizer is the sensitivity, generally limited to the  $\text{mg l}^{-1}$  range. There are several attempts to increase the sensitivity of flame AAS, e.g., by locating a quartz, silica, or metal tube in the flame. An increase in the detection limits of more than one order of magnitude may be achieved in this way. This strongly increased power of detection is mainly due to the prolonged residence time of the analyte atoms in the absorbing layer.

**Graphite tube atomizer** The main type of electrothermal atomizer is the graphite tube. A definite volume of a liquid sample is introduced into the graphite tube which is subsequently gradually and step-wise heated to increasing temperatures according to a preselected program normally including drying, pyrolysis, atomization, clean out, and cooling steps. During the drying step low-temperature heating is applied to evaporate the solvents from the sample. The ideal temperature is just below the boiling point of the solvent to prevent sputtering of the sample. The aim of the pyrolysis step is to break down or evaporate the matrix prior to atomization by heating to the highest possible temperature without losses of the analyte element. The atomization step aims at the creation of free gaseous atoms of the analyte. The atomization temperature depends on the type of both analyte and sample. The clean-out and cooling steps serve to prepare the graphite tube for the next run.

The graphite tube atomizer operates under spatially and temporally nonisothermal conditions. The analyte atoms volatilized from the tube wall come into a cooler gas, so that molecular species may be formed, leading to chemical interferences. L'vov suggested the sample to be deposited onto a small graphite platform aligned in the central part of the graphite tube. The platform is heated primarily by radiation from the walls, so that there is a time lag between heating of the tube and the platform. The platform reaches the atomization temperature when the tube wall and the gas have already reached (or almost reached) equilibrium. The analyte atoms are thus volatilized in a gas with a temperature higher than the equilibrium, so that chemical interferences are diminished.

**Quartz tube atomizer** Some elements like antimony, arsenic, bismuth, lead, tin, can be vaporized as molecules (e.g., hydrides) by chemical reaction at room temperature. For this purpose a reductant solution ( $\text{NaBH}_4$ ,  $\text{SnCl}_2$ ) is added to the sample solution and the obtained gaseous hydrides are transported to the quartz tube atomizer (hydride

generation AAS). The same arrangement can be used for mercury, which upon reductant addition is released as atomic vapor. In this case the quartz tube acts as a sample cell only (cold vapor AAS).

### Sample Introduction and Sample Pretreatment

Sample introduction and sample pretreatment are still limiting factors in AAS, determining the selectivity and sensitivity of the AA analysis. In AAS, typically liquid samples are analyzed. Nevertheless, several sample introduction systems have been developed for the direct atomic absorption analysis of solid samples avoiding sample dissolution, which is labor intensive and suffers risks of both contamination and losses. Solid samples in the form of powders, slurries, etc., are mostly analyzed in electrothermal atomizers.

Another approach offering a variety of rapid, reliable, and economic ways for sample pretreatment and sample introduction in AAS is the online coupling of flow injection systems with both flame and electrothermal AAS techniques. Such integrated systems have made a great progress in recent years by providing automated operation avoiding time-consuming manual work and enhancing accuracy and precision.

*See also:* **Atomic Absorption Spectrometry:** Interferences and Background Correction; Flame; Electrothermal; Vapor Generation. **Atomic Spectrometry:** Overview. **Flow Injection Analysis:** Principles.

### Further Reading

- Beatty RD and Kerber JD (1993) *Concepts, Instrumentation and Techniques in Atomic Absorption Spectrometry*. Norwalk: The Perkin-Elmer Corporation.
- Bendicho C and de Loos-Vollebregt MTC (1991) Solid sampling in electrothermal atomic absorption spectrometry using commercial atomizers. *Journal of Analytical Atomic Spectrometry* 6: 353–374.
- Ebdon L, Evans EH, Fisher AS, and Hill SJ (1998) Chapter 2 “Flame atomic absorption spectrometry”, Chapter 3 “Electrothermal atomic absorption spectrometry”. In: *An Introduction to Analytical Atomic Spectrometry*. Chichester: Wiley.
- Fang Z (1995) *Flow Injection Atomic Absorption Spectrometry*. Chichester: Wiley.
- Robinson JW (1990) *Atomic Absorption Spectroscopy*. New York: Dekker.
- Vandecasteele C and Block CB (1993) *Modern Methods for Trace Element Determination*. Chichester: Wiley.
- Welz B and Sperling M (1999) *Atomic Absorption Spectrometry*. Third Completely Revised Edition Weinheim: Wiley-VCH.

## Interferences and Background Correction

**D J Butcher**, Western Carolina University, Cullowhee, NC, USA

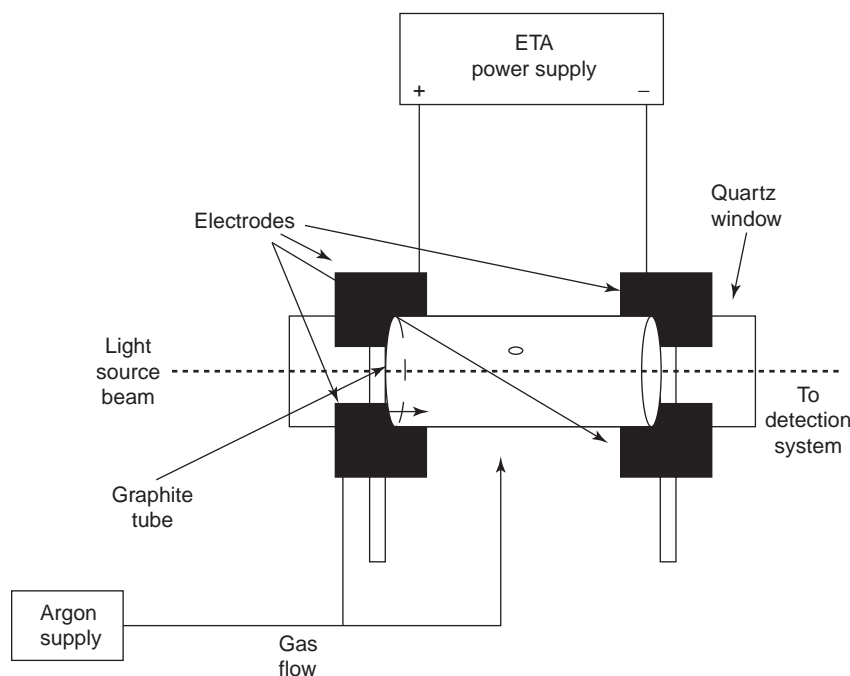
© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Atomic absorption spectrometry (AAS) is a very sensitive method of elemental analysis, allowing the determination of metals in a variety of samples at the picogram level. It has been used for thousands of applications involving a wide diversity of samples. The atomic absorption phenomenon involves a measurement of the reduction of the intensity of optical radiation subsequent to its passage through a cell containing gaseous atoms. Modern instrumentation for AAS typically consists of a light source called a hollow cathode lamp (HCL), which emits specific wavelengths of light that are ideally only absorbable by the analyte; an 'atom cell', which serves to convert the samples into gaseous atoms that can absorb light from the HCL; a 'detection system' that serves to isolate and quantify the wavelengths of interest; and a computer system to control instrument operation and collect and process data. Most commonly, samples are converted to aqueous solutions by digestion procedures to minimize interferences and provide optimal precision and accuracy.

The most common atom cells employed for AAS are flames and electrothermal atomizers (ETAs). Flames, which employ a carefully controlled combustion environment to produce atoms, have the advantages of speed, ease of use, and continuous operation, which allows simple interface with chromatographic systems for speciation. However, the efficiency of sample introduction is typically 5%, and the atoms are dispersed across a relatively large volume, causing the sensitivity of flame AAS to be relatively poor (typically parts per million level). ETAs, also called graphite furnaces, employ a small graphite tube whose temperature can be accurately controlled by a power supply (**Figure 1**). Argon gas flows around the tube to prevent its combustion at elevated temperatures. Each ETA-AAS measurement involves a process called the 'atomization cycle'. ETA-AAS provides much higher sensitivity (typically parts per billion level, or picograms on a mass basis) compared to flame AAS because the atoms are concentrated in a relatively small volume and its high-(essentially 100%) atomization efficiency. In addition, ETA-AAS is capable of analyzing small volumes (typically 20  $\mu$ l per measurement).

However, the practical use of ETA-AAS for real sample analysis may be hampered by interferences, induced by nonanalyte components of sample (the sample matrix), that induce a change, either positive



**Figure 1** Schematic diagram of an electrothermal atomizer.

or negative, in the analyte absorption and prevent accurate analysis. Interferences are a much greater problem in ETA-AAS than flame AAS, and hence this article will focus on the former. A considerable amount of scientific effort was focused in the 1980s and 1990s to create instrumental developments and analytical protocols, called 'modern furnace technology' (MFT), to eliminate or at least minimize ETA-AAS interferences.

This article outlines the major types of interferences that are of significance in ETA-AAS, and the major components of MFT are outlined to describe ways to prevent interferences. Finally, an outline of procedures to develop methods for practical ETA-AAS analysis is provided.

## Types of Interferences for ETA-AAS

The three major types of interferences encountered in ETA-AAS are listed in Table 1. These include spectral, chemical, and physical differences in analytical standards and samples that may lead to inaccurate analysis by ETA-AAS. Detailed descriptions of these interferences are given below.

### Spectral Interferences

Spectral interferences involve a change in the amount of light that reaches the detection system by concomitants in samples. They can be classified as spectral overlaps, scatter, and molecular absorption.

Spectral overlaps arise in the relatively rare case in which a nonanalyte element absorbs light from the HCL. These interferences are relatively uncommon because the emission profile of an HCL is considerably

narrower than the absorption profile of the atoms in the ETA, and the interferent must be present at high concentrations. The most significant analytical spectral overlap occurs between gallium and manganese at 403.3 nm.

Scatter of source light occurs if particles are formed in the graphite tube that attenuates the source radiation. Particles may be formed from inorganic salts (e.g., sodium chloride), decomposition of organic compounds, or from the graphite tube at very high temperatures. The magnitude of the scattering coefficient  $\tau$  is given by Rayleigh's law of scattering:

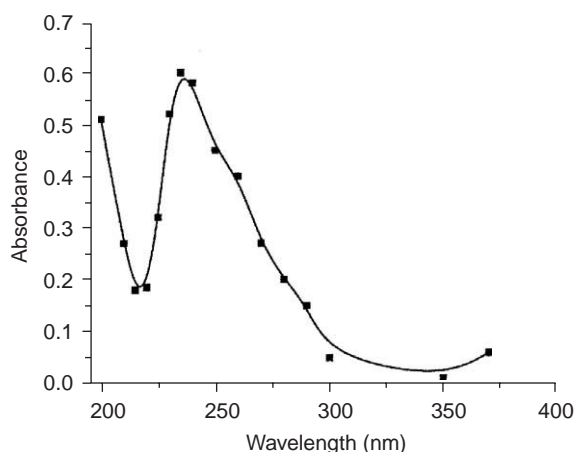
$$\tau = 24\pi^3 \frac{N\nu^2}{\lambda^4} \quad [1]$$

where  $\nu$  is the volume of the scatterers ( $\text{cm}^3$ ).  $N$  is the density of scatterers (number  $\text{cm}^{-3}$ ), and  $\lambda$  is the wavelength of light (cm). This relationship shows that scatter has a maximum value in the presence of large scattering particles and at low-source wavelengths. Scatter may be a significant source of error for ETA-AAS. However, it is characterized by relatively constant magnitude as a function of wavelength, and hence its effects can usually be removed by the use of MFT.

Molecular absorption involves the absorption of HCL radiation by small molecules produced by components of the sample matrix. Alkali metal halide compounds absorb radiation throughout the ultraviolet region where most ETA-AAS measurements are performed, as shown in Figure 2. These broad spectra are caused by absorption of source emission that induces photodissociation of the molecules. In addition to photodissociation continuum spectra, other molecules are characterized by electronic band spectra, which are typically 0.5–10 nm in width and may be further split into vibrational and rotational

**Table 1** Types of interferences in ETA-AAS

Interference	Subclassification	Description
Spectral	Spectral overlaps	Absorption by nonanalyte atoms
	Scatter	Source light attenuated by particles
	Molecular absorption	Small molecules from sample matrix absorb light
Chemical	Volatile compound formation	Vaporization of analyte before atomization step
	Involatile compound formation	Incomplete vaporization of analyte during atomization step
	Gas-phase interferences	Gas chemical reactions between analyte and matrix components during atomization step
Physical		Differences in viscosity and surface tension of standards and samples



**Figure 2** Molecular spectrum of sodium chloride (5  $\mu\text{g}$ ) in an electrothermal atomizer. (Data from Culver BR and Surles T (1975) *Analytical Chemistry* 47: 920.

lines. Electronic band spectra are particularly difficult to correct for because of the variation in intensity across the absorption profile of the analyte.

### Chemical Interferences

Chemical interferences involve a chemical reaction between the analyte and components of the sample matrix that reduces the formation of gaseous atoms during the atomization step. These interferences may be further characterized based on the type of chemical reaction involved. Volatile compound formation involves a reaction between the analyte and a matrix component that produces a volatile molecule that is vaporized out of the furnace during the pyrolysis step. Involatile compound formation involves a reaction between the analyte and a matrix component that produces a molecule that is insufficiently volatile or sufficiently stable to reduce the formation of atoms during the atomization step. Gas-phase interferences refer to a chemical reaction between the analyte and a concomitant in the vapor phase.

Chemical interferences can be eliminated in most cases by the use of MFT (see below). Perhaps the most widely studied interference is the chloride interference for volatile elements such as lead and thallium, which causes vaporization of these elements at low temperatures (500°C). Although the mechanism of this interference has been widely studied, controversy still exists about its exact cause. Chemical interferences are also problematic for involatile elements. For example, metals that form involatile carbides, such as silicon, tantalum, and tungsten, are very difficult to determine by ETA-AAS because of their reactivity with the graphite tube.

### Physical Interferences

Physical interferences involve differences between physical properties of standard and sample solutions,

such as viscosity and surface tension, which result in changes in the ETA-AAS analytical signal. These differences typically do not affect the quantity of material introduced into the graphite tube, but may affect the degree of spreading of the sample on the tube surface. Increased spreading may increase analyte-graphite interactions and decrease the atomization efficiency of the analyte. As an example, physical interferences may be encountered using a sample introduction technique called slurry sampling, in which an aqueous slurry of a powdered sample is directly introduced into the tube. Surfactants are commonly added to these samples to assist in wetting and dispersing biological powders. However, these chemicals also increase spreading on the graphite and potentially induce a physical interference. These interferences may also be eliminated by the use of MFT.

## Modern Furnace Technology for Practical Analysis by ETA-AAS

The components of MFT for ETA-AAS are outlined in Table 2. These include instrumental developments, the use of high-quality graphite materials, a modern method of background correction, and chemical modifiers. The combination of these instrumental developments and protocols allows the routine use of ETA-AAS for real sample analysis.

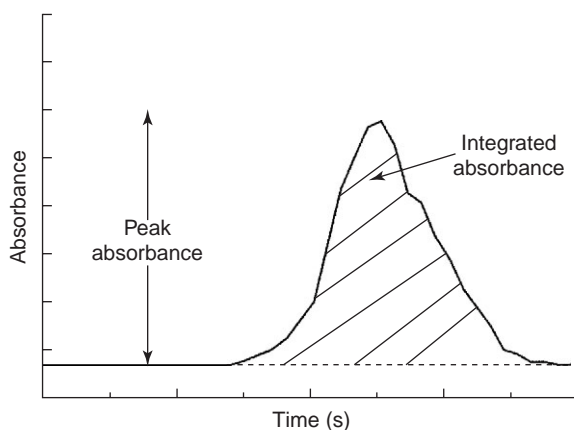
### Instrumental Developments for ETA-AAS

In flame AAS, the sample is continuously introduced into the atom cell, producing a steady-state signal. Consequently, flame AAS instrumentation is designed to measure the peak height of this continuous signal, and this practice employed into early instrumentation for ETA-AAS. As described above, a discrete volume of sample (typically 20  $\mu$ l) is

**Table 2** Instrumental developments and analytical protocols involved in MFT

<i>MFT category</i>	<i>MFT component</i>	<i>Advantage</i>
Instrumental developments	Integrated absorbance	Better accuracy and precision than peak absorbance
	Fast electronics	Accurate characterization of transient signal
	Autosampler	Improved precision compared to manual pipetting
High-quality graphite	Pyrolytically coated graphite tubes	Reduced analyte-graphite reactions
	Transversely heated tubes	Reduced chemical interferences by vaporization into a uniformly hot tube
	Platform atomization with high-atomization heating rates	Atomization into a hot environment that reduces chemical interferences
Background correction	Zeeman effect	More accurate characterization of spectral interferences
Chemical modifiers		Reduced spectral and chemical interferences





**Figure 3** Electrothermal atomizer signal from a single furnace firing.

introduced into the graphite tube that is converted into gaseous atoms during the atomization step, resulting in a transient signal, as illustrated in **Figure 3**. The use of integrated absorbance, rather than peak absorbance, has been shown to provide superior accuracy and precision compared to peak absorbance. This is particularly true for real sample analysis, where the sample matrix may reduce the height of the analytical signal, but not its area.

The electronics for instrumentation designed flame AAS typically have long time constants (0.1–1 s) to smooth the steady-state signals obtained in this technique. However, the use of these electronics in early ETA-AAS instrumentation caused distortions of the transient ETA-AAS signals, which are typically a few seconds in width. Modern instrumentation for ETA-AAS employs time constants on the orders of milliseconds to prevent these errors.

Although autosamplers for flame AAS allow automated collection of data, they do not typically improve the accuracy and precision of analysis. Consequently, much of the early ETA-AAS literature employed manual pipetting. However, it is very difficult to reproducibly manually inject the small volumes employed in this technique. An autosampler should therefore be used to obtain the best accuracy and precision.

### Graphite for ETA-AAS

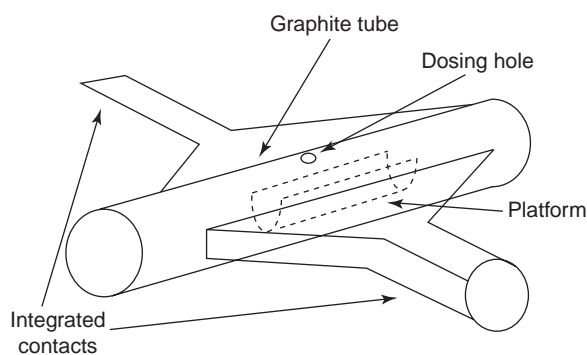
The types of graphite substrates may significantly affect the analytical performance of ETA-AAS (**Figure 3**). Early commercial instrumentation in the 1970s employed relatively large graphite tubes (50 mm in length by 8 mm diameter) composed of polycrystalline graphite. Although these tubes

accommodated large sample volumes, they were highly susceptible to chemical interferences that can be minimized by the methodology described below.

First, modern instrumentation uses smaller tubes (20–30 mm in length by 3–6 mm in diameter) that can be rapidly heated at more than  $1000^{\circ}\text{C s}^{-1}$ . High-heating rates allow the analyte to be rapidly atomized, improving precision. Second, chemical interferences between graphite and analytes were shown to be minimized by the addition of a 50  $\mu\text{m}$  layer of pyrolytic graphite on the surface of the tube, forming pyrolytically coated graphite. Pyrolytic graphite has a relatively high density and serves to reduce the chemical reactivity of the graphite. Pyrolytically coated graphite should always be employed for the determination of elements that may form carbides (e.g., vanadium, titanium, molybdenum). It is also effective at minimizing physical interferences due to increased analyte-graphite interactions.

Another approach to minimize interferences for volatile elements (e.g., lead, thallium) involves atomization into a tube that is at a relatively high temperature. Early work involved sample introduction directly onto the wall of the graphite tube. Under these circumstances, atomization occurred while the tube was still heating to the atomization temperature, which may allow gas-phase reactions to occur between the analyte and components of the sample matrix. In order to alleviate this problem, a small graphite shelf, called an 'L'vov platform', is inserted in the bottom of the graphite tube. Sample is introduced on to the platform. Since electrical current does not pass through the platform, it is heated radiatively by the tube walls, and hence its temperature is lower than that of the walls. This ensures that atomization occurs after the tube and gas inside it has reached a relatively high and constant temperature, minimizing chemical interferences.

Most furnace designs involve the passage of current through the length of the furnace using a longitudinal heating. This furnace design may cause severe temperature gradients, reported to be as large as  $1200^{\circ}\text{C}$ , between the center of the tube and its ends. These relatively cool regions may allow the analyte to condense or react with matrix components, resulting in interferences. An alternative to longitudinal heating is 'transverse heating', where the flow of current is perpendicular to the length of the tube (**Figure 4**). In this approach, the electrode contacts and L'vov platform are integrated as part of the graphite tube. This design has been shown to reduce interferences compared to longitudinal heating schemes.

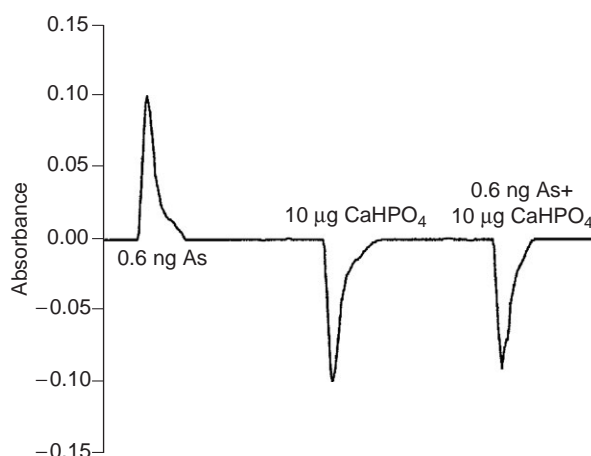


**Figure 4** Transversely heated graphite tube with integrated platform and contacts.

### Background Correction for ETA-AAS

Spectral interferences may have a deleterious effect on the precision and accuracy of practical analysis performed by ETA-AAS. The effects of spectral backgrounds may be minimized by the use of methods of background correction. The basic goal of any method of background correction is the accurate measurement of the background absorbance, so it can be subtracted from the uncorrected signal (signal plus background) to give a background corrected signal. The ideal method of background correction would offer ease of use, low cost, applicability to all elements and wavelengths, measurement of background at the analytical wavelength, the use of one source, minimal effect on the detection limit or linear dynamic range, compatibility with a high-sampling frequency ( $>60$  Hz), and widespread use involving a variety of sample applications. As one might expect, no one method meets all of these criteria, and hence compromises must be made on some of these analytical parameters. In this article, the two most widely used methods of background correction are discussed: continuum source and Zeeman.

**Continuum source background correction** Continuum source background correction involves the use of a continuum source, such as a deuterium arc in the ultraviolet, or a tungsten halide lamp in the visible, to measure background attenuation. The spectral output from a HCL and the continuum source are alternatively transmitted through the graphite tube. The HCL has a narrow band emission profile (typically  $0.003$  nm) that may be attenuated by the analyte and molecules or particles responsible for the background signal. The continuum source has a relatively wide emission profile (typically  $0.2$ – $1$  nm), which is defined by the spectral bandpass of the monochromator, and may be absorbed by the sources



**Figure 5** Determination of arsenic ( $193.7$  nm) in the presence of  $\text{CaHPO}_4$  with continuum source background correction and a  $1$  nm spectral bandpass.

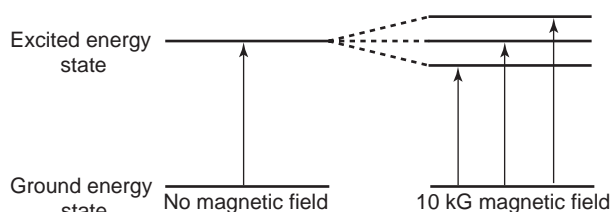
of background but not by the analyte. Therefore, electronic subtraction of the HCL measurement (signal plus background) minus the continuum source measurement (background) gives a background corrected measurement.

Continuum source background correction has two major limitations for practical analysis. First, it is difficult to exactly align the two light sources, which leads to inaccurate analyses, particularly at high-background levels. Second, it is unable to correct for background signals whose magnitude varies across the bandpass of the monochromator, which are called 'structured backgrounds'. Under these circumstances, continuum source background correction cannot provide accurate analyses. An example of a continuum source background correction error is shown in **Figure 5**, in which the measured background at the analytical wavelength is higher than the signal, resulting in a negative signal called an 'overcorrection error'.

In spite of these limitations, continuum source background correction may be used with good accuracy for many analyses. It offers low cost, wide applicability, operation at high frequencies, and little degradation in detection limits or linear dynamic range. It is commonly found in commercial instrumentation alone or with other methods of correction.

**Zeeman effect background correction** Zeeman effect background correction involves the use of the Zeeman effect, which involves the splitting of atomic lines into two or more components in the presence of an intense magnetic field (**Figure 6**). In general, these components only absorb one direction of polarized light, and hence the combination of a magnetic field and a polarizer may be used to make the background





**Figure 6** Normal Zeeman splitting pattern in a transverse magnetic field.

measurement. It is assumed that the background signals are not affected by the magnetic field, which is usually (but not always) a correct assumption. The most common arrangement for Zeeman ETA-AAS uses an electromagnet arranged around the graphite tube. Measurement when the magnetic field is zero gives signal plus background, measurement at the maximum field strength gives the background, and, hence, subtraction of the two gives a background corrected measurement. A number of instrumental configurations are commercially available for Zeeman ETA-AAS. These are beyond the scope of this article; readers are referred to the section Further Reading.

Zeeman effect background correction is generally regarded as the most accurate method, and it can be employed with almost all elements and transitions. However, it requires an expensive magnet system and typically degrades the analytical signal and linear dynamic range. In addition, a few reports have demonstrated interferences due to a molecule or second element with an absorption line very close to the analytical wavelength. Ideally, if cost is not an option, it may be preferable to be able to utilize both methods of correction, in case interferences limit the use of one of these methods.

### Chemical Modifiers for ETA-AAS

A very important facet of modern furnace technology involves the addition of chemical reagents, to standards and samples, to minimize interferences. These reagents are called chemical (matrix) modifiers. Modifiers serve to separate the analyte from the sample matrix to reduce spectral and chemical interferences. For volatile elements, such as lead, modifiers typically reduce the volatility of the analyte, allowing the use of a higher pyrolysis temperature that may allow removal of concomitants before the analytical measurement during the atomization step.

An ideal modifier would be available in high purity, minimize interferences, cause no reduction in the lifetime of the graphite tube, not be commonly

**Table 3** Chemical modifiers used for ETA-AAS

<i>Element</i>	<i>Recommended chemical modifiers</i>
Arsenic	Palladium nitrate and magnesium nitrate; nickel nitrate
Bismuth	Palladium nitrate and magnesium nitrate; nickel nitrate
Cadmium	Palladium nitrate and magnesium nitrate; diammonium hydrogen phosphate and magnesium nitrate
Chromium	Magnesium nitrate
Copper	Palladium nitrate and magnesium nitrate
Lead	Palladium nitrate and magnesium nitrate; diammonium hydrogen phosphate and magnesium nitrate
Manganese	Magnesium nitrate
Selenium	Palladium nitrate and magnesium nitrate; nickel nitrate and magnesium nitrate
Thallium	Palladium nitrate and magnesium nitrate; diammonium hydrogen phosphate
Zinc	Nickel nitrate

determined by ETA-AAS, and be applicable to many elements. A summary of commonly used modifiers is provided in Table 3. Commonly used modifiers include palladium nitrate, magnesium nitrate, diammonium hydrogen phosphate, and nickel nitrate. Palladium nitrate, either alone or with other compounds, best meets these criteria and is the most widely used modifier at this time. The introduction of oxygen as a modifier during the pyrolysis step, which is called oxygen ashing, has been used for the analysis of biological samples. Oxygen ashing has been shown to reduce formation of carbonaceous residue produced by these samples and minimize interferences.

**See also:** Atomic Absorption Spectrometry: Principles and Instrumentation; Flame; Electrothermal; Vapor Generation. Elemental Speciation: Overview.

### Further Reading

- Butcher DJ and Sneddon J (1998) *A Practical Guide to Atomic Absorption Spectrometry*. New York: Wiley.
- Butcher DJ (1998) Recent developments in graphite furnace atomic absorption spectrometry. In: Sneddon J (ed.) *Advances in Atomic Spectroscopy*, vol. 4, p. 151. Greenwich, CT: JAI Press.
- Butcher DJ (2002) Advances in electrothermal atomization for atomic absorption and atomic fluorescence. *Applied Spectroscopy Reviews* 37: 305–320.
- Dean JR (1997) *Atomic Absorption and Plasma Spectroscopy*. Chichester: Wiley.
- Dulude G (1992) Background correction techniques in atomic absorption spectrometry. In: Joseph S (ed.) *Advances in Atomic Spectroscopy*, vol. 1. Amsterdam: Elsevier.

- Ebdon L, Evans EH, Fisher A, and Hill SJ (1998) *An introduction to analytical atomic spectrometry*. Chichester: Wiley.
- Haswell SJ (1991) *Atomic Absorption Spectrometry: Theory, Design, and Applications*. Analytical Spectroscopy Library Ed., vol. 5. Amsterdam: Elsevier.
- Jackson KW (ed.) (1999) *Electrothermal Atomization for Analytical Atomic Spectrometry*. Chichester: Wiley.
- Lajunen LHJ (1992) *Spectrochemical Analysis by Atomic Absorption and Emission*. Cambridge: Royal Society of Chemistry.
- Minoia C and Caroli S (1992) *Applications of Zeeman Graphite Furnace Atomic Absorption Spectrometry in the Chemical Laboratory and Toxicology*. Oxford, UK: Pergamon.
- Slavin W (1984) *Graphite AAS: A Source Book*. Norwalk, CT: Perkin-Elmer Corporation.
- Sneddon J and Butcher DJ (2002) Application of graphite furnace atomic absorption spectrometry to biological and clinical samples. In: Sneddon J (ed.) *Advances in Atomic Spectroscopy*, vol. 7. Amsterdam: Elsevier.
- Varma A (1990) *CRC Handbook of Furnace Atomic Absorption Spectrometry*. Boca Raton, FL: CRC Press.
- Welz B (1999) *Atomic Absorption Spectrometry*, 3rd edn. New York: Wiley-VCH.

## Flame

**S J Hill**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Flame atomic absorption was until recently the most widely used techniques for trace metal analysis, reflecting its ease of use and relative freedom from interferences. Although now superseded in many laboratories by inductively coupled plasma atomic emission spectrometry and inductively coupled plasma mass spectrometry, flame atomic absorption spectrometry still is a very valid option for many applications. The sample, usually in solution, is sprayed into the flame following the generation of an aerosol by means of a nebulizer. The theory of atomic absorption spectrometry (AAS) and details of the basic instrumentation required are described in a previous article. This article briefly reviews the nature of the flames employed in AAS, the specific requirements of the instrumentation for use with flame AAS, and the atomization processes that take place within the flame. An overview is given of possible interferences and various modifications that may provide some practical advantage over conventional flame cells. Finally, a number of application notes for common matrices are given.

### Types of Flame

In AAS, the flame is only required to produce ground state atoms. Two types of flame are employed to achieve this: the premixed combustion flame consisting of a fuel and oxidant gas, and the diffusion flame where the fuel is also the carrier gas that burns on contact with air. Premixed flames commonly employ either air or dinitrogen oxide as the oxidant, and either acetylene, propane, or hydrogen as the fuel gas.

The temperature of diffusion flames is lower than that of premixed flames, although it must be remembered that atomization occurs as a result of both the high enthalpy and temperature of the flame and through chemical effects (i.e., chemical compounds and radicals within the flame), as discussed below. Thus, although the burning temperatures of two different premixed gas flames may be similar, the analytical characteristics can be very different. The characteristics (indicative values) of some commonly used flames are shown in **Table 1**.

Of the various flame types available, the air–acetylene flame is the most widely used. This flame is stable, simple to operate, and produces sufficient atomization to enable good sensitivity and freedom from many interelement interferences. Over 30 elements may be determined using an air–acetylene flame, although the flame conditions may have to be adjusted to create a suitable environment for some elements. For example, the alkaline earth metals require a fuel rich (reducing flame), whilst the noble metals are determined using a lean (oxidizing) flame. In many cases, however, increasing the oxygen content of the flame may be counterproductive since although this will produce a hotter flame it may also promote the formation of refractory oxides. Ionization will occur in an air–acetylene flame for a number of easily ionized elements such as the alkali metals. In such cases, an ionization suppressor or buffer consisting of a large excess of an easily ionized element (such as cesium or potassium) may be added. This has the effect of suppressing the ionization (e.g., sodium) by a simple mass action effect resulting from the excess cesium in the flame. The alkali metals may be determined using an air–propane flame, although such flames are seldom used these days, atomic emission spectrometry being the preferred technique for such elements.

**Table 1** Typical characteristics of premixed flames used in AAS

Flame type	Flow rates ( $\text{l min}^{-1}$ )		Approximate temperature (K)	Maximum burning velocity ( $\text{cm s}^{-1}$ )
	Fuel	Oxidant		
<i>Air-propane</i>				
Lean	0.3	8	2200	45
Stoichiometric	0.3–0.45	8		
Rich	0.45	8		
<i>Air-acetylene</i>				
Lean	1.2	8		
Stoichiometric	1.2–1.5	8	2450	160
Luminous	1.5–1.7	8		
Rich	1.7–2.2	8	2300	
<i>N<sub>2</sub>O-acetylene</i>				
Lean	3.5	10		
Stoichiometric	3.5–4.5	10	3200	285
Rich	4.5	10		
<i>Air-hydrogen</i>				
Stoichiometric	6	8	3200	320
<i>N<sub>2</sub>O-hydrogen</i>				
Stoichiometric	10	10	2900	380
<i>N<sub>2</sub>O-propane</i>				
Stoichiometric	4	10	2900	250

Another premixed flame commonly employed is the dinitrogen oxide (nitrous oxide)–acetylene flame. This flame is both hot and reducing and commonly used for the determination of elements such as Al, B, Ba, Be, Mo, Nb, Re, Sc, Si, Ta, V, W, Zr, the lanthanoids, and actinoids. This type of flame has a characteristic red inter-conal zone due to the presence of the cyanogen radical that is produced under slightly fuel-rich conditions. This radical is a very efficient scavenger of oxygen and the higher temperature aids in promoting dissociation. For safety reasons, the nitrous oxide–acetylene flame should never be run fuel-lean and carbon deposits should not be allowed to build up on the burner (a 5 cm slot burner is used). To establish the flame, an air–acetylene flame is lit first, made very fuel-rich, and then switched to nitrous oxide–acetylene using a two-way valve. The flame is shut down using the reverse procedure.

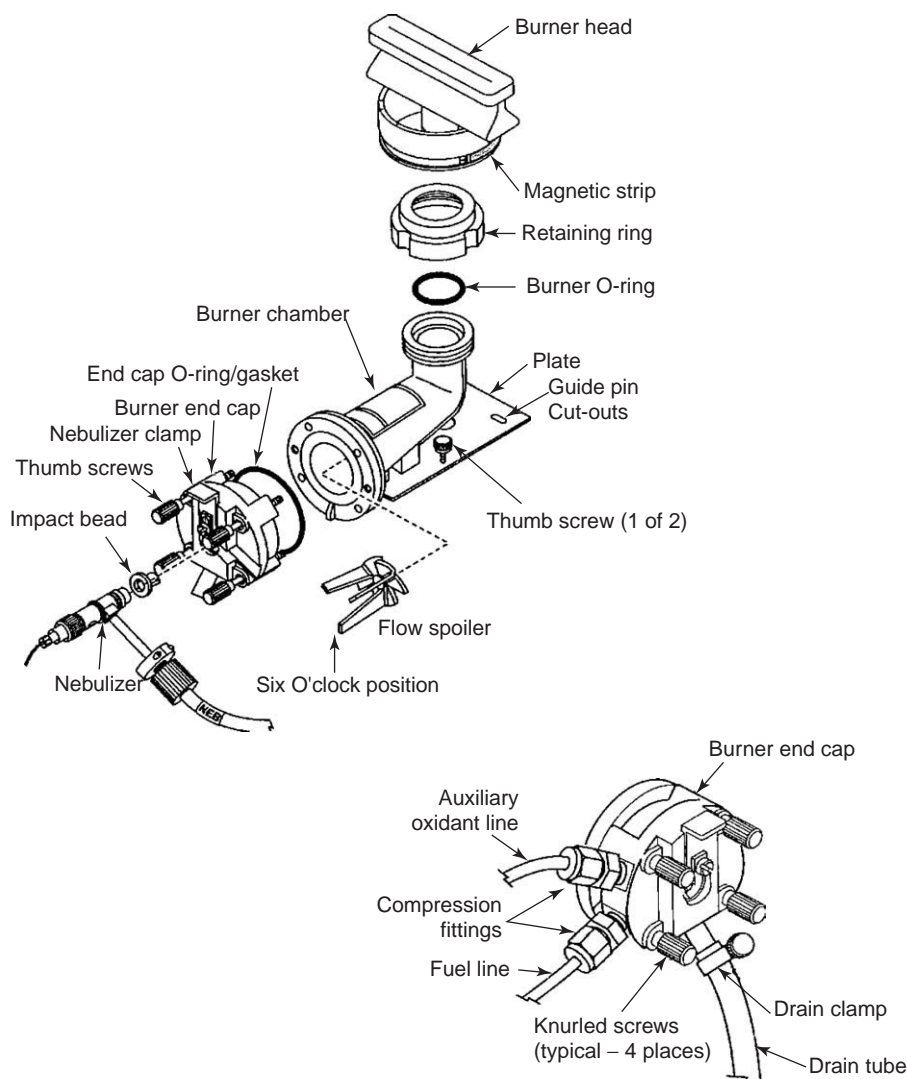
Hydrogen diffusion flames are used to determine easily atomized elements such as arsenic and selenium. Often such flames, for example, the argon–hydrogen flame, are used in conjunction with hydride generation techniques. The flame itself is transparent over a wide spectral range and so has some benefit over hydrocarbon flames that will absorb strongly at low wavelengths.

## Sample Introduction and Burner System

In order for the atom cell, i.e., the flame, to function effectively and produce an atomic vapor, the sample

must be presented in the form of a fine aerosol. This is achieved using a nebulizer and spray chamber prior to the burner. A typical pneumatic nebulizer system for a premixed flame is shown in **Figure 1**. The sample solution is sucked up a plastic capillary tube by the reduced pressure created at the end of the tube by the oxidant or carrier gas flow, i.e., the Venturi effect. During this process the sample is shattered into tiny droplets as it exits the capillary tube to produce an aerosol. This process may be optimized with respect to sample uptake and drop size by adjusting the position of the nebulizer capillary. Additionally, an impact bead may be placed in the path of the initial aerosol to provide secondary fragmentation. Both the nebulizer and impact bead must be made of corrosion-resistant material such as a platinum–iridium alloy (90:10) or stainless steel coated with an inert plastic.

Once formed, the aerosol passes into the burner or spray chamber (sometimes also called the cloud chamber). The role of the spray chamber is to homogenize both the aerosol and gases that tend to dampen fluctuations in nebulizer efficiency, and to remove any large droplets before they reach the flame. Large droplets (diameter  $> 10 \mu\text{m}$ ) collect on the sides of the chamber and then drain to waste. Spoilers and baffles placed at the end of the spray chamber aid this process. Because the spray chamber will fill with flammable gas, modern instruments will also incorporate some form of antiflashback protection from the flame.



**Figure 1** Diagram of nebulizer, spray chamber, and burner system components as used in the PerkinElmer AAnalyst 200/400 atomic absorption spectrometer. (Reproduced by kind permission of PerkinElmer Inc. All rights reserved.)

Although the nebulizer and spray chamber arrangement described allows only small droplets to reach the flame, the efficiency in terms of sample transport is only  $\sim 10\%$ . Ultrasonic nebulizers offering higher efficiency are available, but these are more complicated and have a tendency to suffer from memory effects, i.e., contamination from previous samples. Organic solvents also improve the efficiency, probably forming smaller droplets as a result of their lower surface tension. However, organic solvents may also affect nebulization because of their different density, viscosity, and saturated vapor pressure. In addition, they may also act as secondary fuels, thus affecting the flame.

The introduction of solid powder samples is possible using pneumatic sampling devices. Various graphite capsules and vibration tubes have also been used on this form of sample. In both cases the

precision is very poor and the introduction of powders is thus seldom attempted these days.

The design of the burner employed in AAS depends on the oxidant-fuel gas mixture. Slot burners are now used almost exclusively, the length of the slot being 10 cm for use with air-acetylene and multishot burners, and 5 cm for nitrous oxide-acetylene burners, reflecting the higher burning velocity of the flame as shown in Table 1. The slot width and the conductivity of the metal used for the construction of the burner are also important in terms of stability of operation and prevention of clogging. The burner should also be constructed or coated in an inert material to avoid corrosion.

### Atomization Processes in the Flame

Once the aerosol reaches the flame the droplets are desolvated to form a mist of salt clotlets, which

then fuses and evaporates or sublimates. This process is critically dependent on the size and number of the particles, their composition, and the flame mixture. As the absolute concentration of analyte in the flame is very small ( $<10^{-3}$  atm), the saturated vapor pressure may not be exceeded even at temperatures below the melting point. The vaporization is fast; the molecules that make up the vapor decomposing into individual atoms almost instantaneously. In fact, the distance that the sample travels through the flame before atomization may be very short, for example, less than 1 cm for sodium chloride in an air-acetylene flame. Higher in the flame the sodium concentration will slowly decrease due to cooling of the flame and the composition of the flame has little effect on the atomization process. Other species, however, are more refractory (e.g., the alkaline-earth and rare-earth oxides) and may require a hotter flame and reducing environment to aid dissociation, thus reflecting the importance of flame chemistry (both ground state and excited atoms may be produced by radical reactions in the primary reaction zone of the flame). Since dissociation and the subsequent ionization of a diatomic compound are similar processes, it is possible to calculate the degree of ionization of atoms into ions and electrons for different elements as a function of temperature and partial pressure using the Saha equation.

For many elements the atomization efficiency, defined as the ratio of the number of atoms to the total number of analyte species, atoms, ions, and molecules in the flame, is 1, although for other elements (e.g., the lanthanoids) it is less than 1, even in a nitrous oxide-acetylene flame. However, the formation of atoms is not the end of the story since once formed they may be lost through compound formation or ionization. Ionization increases exponentially with temperature and is a particular problem for the elements on the left of the periodic table, i.e., the alkali and the alkaline-earth elements. It is also a problem with Al, Ga, In, Sc, Ti, and Tl in the nitrous oxide-acetylene flame. A summary of atomization in flames is presented in Figure 2.

## Interferences and Errors

The use of a line source and the ratio method (i.e.,  $I^\circ/I$ ) tend to minimize errors in AAS, since many instrumental errors such as long-term source drift, small monochromator drifts, should be cancelled out. However, a stable uptake rate, or aspiration rate, is required. There are principally three types of interference that may be identified in flame AAS: chemical, ionization, and spectral interferences. Both

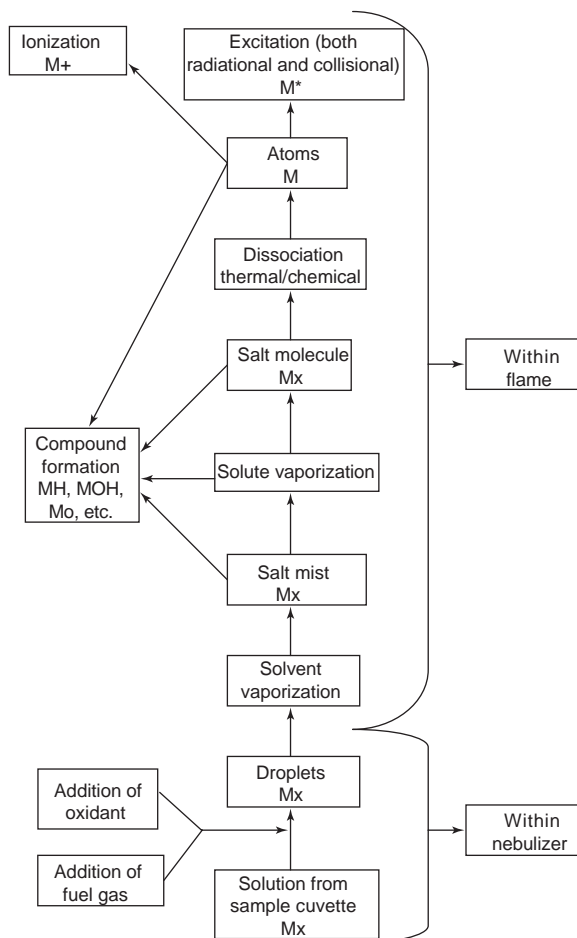


Figure 2 Summary of atomization in flames.

chemical and ionization interferences require the presence of an analyte.

## Chemical Interferences

Chemical effects may originate from either the sample solution or the flame and represent the major type of interference in flame-AAS. The mechanisms behind chemical interferences can be split into two types. The first of these is where the atomization of the analyte is not complete through occlusion into refractory compounds. Stable compound formation of this type will cause a depression of the signal through physical entrapment of small amounts of analyte in clots of matrix oxide in the flame. In premixed laminar flames the volatilization of solid particles begins as soon as they enter the primary reaction zone. The time this process takes is dependent on the size of the particle and so the occurrence of such interferences will depend critically on the observation height in the flame. In the second type, the analyte atoms may react with other atoms or

radicals, leading to the formation of less volatile (or, occasionally, more volatile) compounds. The best-known example of this type of interference is that of phosphate on calcium, although sulfate and silicate have similar effects. The interference manifests itself as a pronounced 'knee' in plots of increasing phosphate concentration against calcium signal. The constant level of interference following the 'knee' suggests formation of a compound that is less volatile than calcium chloride (probably calcium phosphate), which restricts the formation of calcium atoms. There are many other examples of this type of interference, all showing this pronounced 'knee', which distinguishes them from nonspecific occlusions. To overcome this type of interference it is possible to use a hotter flame (the interference of phosphate on calcium is not observed in a nitrous oxide–acetylene flame), make observations higher in the flame, use a 'releasing agent' (an excess of lanthanum or strontium releases calcium from phosphate interference), or finally to use a protective chelating agent (excess ethylenediaminetetraacetic acid (EDTA) protects calcium from the phosphate interference).

### **Ionization Interferences**

This is a vapor-phase interference that may be particularly troublesome for elements with low ionization potentials such as the alkali metals and some alkaline-earth metals when using an air–acetylene flame, and elements having an ionization energy of 7.5 eV or lower using a nitrous oxide–acetylene flame. In such cases, the analyte is partly ionized in the flame, leading to a decrease in the absorption signal. However, the extent of the ionization for a specific element and temperature is dependent on the concentration element. For example, the ionization of barium is strongest at low concentrations and results in curvature of the analytical curve precluding trace analysis. The effect can be explained in terms of the self-suppression of the ionization (see above).

Ionization interferences may be suppressed in two ways. First, a cooler flame may be employed, for example, the alkali metals are little ionized in the cooler air–hydrogen flame. However, this approach is not suitable for the majority of the elements since they are either not determined in cool flames (e.g., the lanthanoids) or subject to solute-volatilization interferences (e.g., barium). The second approach is to shift the ionization equilibrium on the basis of the law of mass action by producing a large excess of electrons in the flame or by charge transfer. In practice, this is simply achieved by adding a large excess of an easily ionized element (e.g., potassium) to both the sample and reference solutions. The effect of this

is to ionize the buffer element in the flame, whilst suppressing the ionization of the analyte.

### **Spectral Interferences**

Spectral interferences in AAS due to direct overlapping of the emission lines from the primary radiation source and the adsorption line of another element in the atom cell are very rare. To give rise to a spectral interference, the lines must not merely be within the band pass of the monochromator, but must actually overlap with each other's spectral profile (i.e., be within 0.01 nm). Spectral interferences resulting from the overlap of molecular bands and lines are more of a problem in AAS. Examples of this type of interference are the nonspecific absorption at 217.0 nm which affects lead (e.g., sodium chloride gives a strong molecular absorption at this wavelength), and the calcium hydroxide absorption band on barium at 553.55 nm. Spectral interferences may usually be eliminated by the use of background correction.

### **Modifications to Conventional Flame Cells**

The use of flame atom cells has many advantages for routine analytical determination. These include the fact that most elements can be readily atomized by the appropriate flame; flame cells are easily optimized and simple to use; and due to their long history much is known about their fundamental behavior. In addition, flames give a steady signal and offer signal-to-background and signal-to-noise ratios that facilitate good sensitivity and precision (0.4–2% r.s.d.) over a wide wavelength range (200–800 nm). However, there are also a number of practical disadvantages that may be encountered when using conventional flame cells. The first of these is that conventional indirect flame systems require relatively large volumes of solution to operate, reflecting the fact that only ~10% of solution uptake is delivered to the flame. Samples also have short transit times in the flames, giving rise to the possibility of incomplete vaporization as discussed above, and once the atoms are formed they are subject to dilution effects from the relatively high flow rate of unburnt gas used to support the flame. It has been estimated that atoms spend only  $10^{-4}$  s in the analysis volume – much less than is required to give a stable signal. Finally, although the sample introduction works well for aqueous solutions, difficulties may be encountered when trying to nebulize organic solvents (which may extinguish the flame) or introduce solids. To overcome these shortfalls, a number of modifications to the flame cell have been proposed.

Small samples ( $25\text{--}200\text{ mm}^3$ ) may be introduced using the technique of pulse nebulization (also known as discrete sample nebulization, direct-injection cup nebulization, gulp sampling, and Hoescht cup nebulization). This technique may also be employed for higher concentrations than normally nebulized. A cup or funnel made of an inert material (e.g., polytetrafluoroethylene) is attached to the nebulizer tubing and the sample is put into the cup as a discrete aliquot using a micropipette. The sample is totally consumed and the transient peak signal recorded.

The use of branched uptake capillaries, connected to the nebulizer using a T-piece, may be advantageous when a buffer or ionization suppressor is required. In addition to avoiding time-consuming solution preparation, it is also possible to calibrate organic extracts using aqueous standards in this way. The approach may also be extended to couple more complex flow injection systems employing novel chemistries in the same way.

The final modification commonly employed is the use of sampling boats and cups. One of the first examples of such a device was the Kahn sampling boat, where the sample was evaporated from a tantalum boat that was simply pushed into the flame. An improvement in sensitivity may be achieved for the more easily atomized elements, although the reproducibility is often poor. A modification to this approach was later (1970) reported by Delves, who replaced the tantalum boat with a nickel microcrucible, the so-called Delves cup. The cup itself is mounted onto a device that allows it to be positioned near the flame to char the sample before insertion into the flame to allow atomization. A nickel absorption tube was also positioned in the flame (aligned with the hollow cathode lamp in such a way as to allow the light to pass through the tube unhindered), the atoms entering through a hole half-way along its length. In this way the residence time of the atoms in the flame could be increased. Such devices are now seldom used.

The use of tubes to increase the residence time of atoms in the analytical zone, and hence improve detection limits have more recently been reported for a variety of applications. Such tubes are often fabricated from silica and employ slots, one directly above the burner slot and the other usually at  $180^\circ$ , to decrease the turbulence of the hot gases. The improvement in sensitivity associated with these tubes is generally confined to those elements readily dissociated to their ground state atoms in the flame. Elements with relatively high metal-oxide dissociation energies such as some of the transition metals that

are normally best determined using the nitrous oxide–acetylene flame are precluded because of the excessive thermal shock this hotter flame would impose on the quartz tube.

Although very useful for many applications, it should be stressed that the above devices will not overcome all of the problems associated with the use of flames. For example, they will not help alleviate the banded and continuous spectra that give rise to background radiation in flames. The banded spectra arise from the excited molecules and radicals in the flame gases, whilst the dissociation, ionization, and recombination of these species give rise to the continuous spectra. Such background radiation is a particular problem with flames when using low wavelengths (i.e., below  $200\text{ nm}$ ). Other problems associated with the use of flames include scatter radiation resulting from particulate matter in the light path, and various safety requirements, particularly with regard to explosion hazards (always present with flames of high burning velocity) and toxic flame products (necessitating the use of extraction systems).

## Selected Applications

There are many applications for flame atomic absorption spectroscopy, most requiring the sample to be in solution. Table 2 lists typical detection limits that may be obtained on a modern instrument, together with details of the flame type recommended for each element. In practice, the concentration of the analyte in solution should be at least 10 times greater than the detection limit.

### Liquid Samples

Information on total analyte levels in aqueous samples may be obtained directly following acidification of the sample, providing the analyte is present in sufficient concentration. Samples may also be filtered to provide information on the dissolved fraction, although losses may occur during the filtering process. Samples with analyte concentrations below the detection limits may be preconcentrated using either evaporation or ion exchange resins prior to analysis. Scale expansion may also be employed. On the other hand, if the analyte is in high concentration, the solution must be diluted to within the linear working range, or some other means of reducing the sensitivity employed such as removal of the impact bead, burner rotation, or the use of alternative less-sensitive absorption lines. Releasing agents and ionization buffers may be required for some samples, although these should be used in

**Table 2** Typical detection limits for flame atomic absorption spectrometry

<i>Element</i>	<i>Characteristic concentration (<math>\mu\text{g ml}^{-1}</math>)</i>	<i>Detection limit (<math>\mu\text{g ml}^{-1}</math>)</i>	<i>Normal range (<math>\mu\text{g ml}^{-1}</math>)</i>	<i>Flame type</i>
Ag	0.03	0.002	0.02–10	Air–C <sub>2</sub> H <sub>2</sub>
Al	0.8	0.03	0.3–200	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
As	0.5	0.3	3–150	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Au	0.1	0.01	0.1–30	Air–C <sub>2</sub> H <sub>2</sub>
B	8.0	0.5	5–2000	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Ba	0.2	0.02	0.2–50	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Be	0.015	0.001	0.01–4	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Bi	0.2	0.05	0.5–50	Air–C <sub>2</sub> H <sub>2</sub>
Ca	0.01	0.001	0.01–3	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Cd	0.01	0.0015	0.02–3	Air–C <sub>2</sub> H <sub>2</sub>
Co	0.05	0.005	0.05–15	Air–C <sub>2</sub> H <sub>2</sub>
Cr	0.05	0.006	0.06–15	Air–C <sub>2</sub> H <sub>2</sub>
Cs	0.02	0.004	0.04–5	Air–C <sub>2</sub> H <sub>2</sub>
Cu	0.03	0.003	0.03–10	Air–C <sub>2</sub> H <sub>2</sub>
Dy	0.6	0.03	3.0–150	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Er	0.5	0.03	3.0–150	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Eu	0.3	0.02	0.2–100	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Fe	0.05	0.006	0.06–15	Air–C <sub>2</sub> H <sub>2</sub>
Ga	0.8	0.08	1–200	Air–C <sub>2</sub> H <sub>2</sub>
Gd	20	2.0	20–6000	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Ge	1.0	0.2	2–300	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Hf	10	2.0	20–3000	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Hg	1.5	0.15	2–400	Air–C <sub>2</sub> H <sub>2</sub>
Ho	0.7	0.04	0.4–200	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
In	0.15	0.04	0.4–40	Air–C <sub>2</sub> H <sub>2</sub>
Ir	0.8	0.5	5–200	Air–C <sub>2</sub> H <sub>2</sub>
K	0.007	0.003	0.03–2	Air–C <sub>2</sub> H <sub>2</sub>
La	40	2.0	20–10 000	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Li	0.02	0.002	0.02–5	Air–C <sub>2</sub> H <sub>2</sub>
Lu	7.0	0.3	3–2000	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Mg	0.003	0.0003	0.003–1	Air–C <sub>2</sub> H <sub>2</sub>
Mn	0.02	0.002	0.02–5	Air–C <sub>2</sub> H <sub>2</sub>
Mo	0.3	0.02	0.2–100	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Na	0.003	0.0002	0.002–1	Air–C <sub>2</sub> H <sub>2</sub>
Nb	20	2.0	20–6000	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Nd	6.0	1.0	10–1500	Air–C <sub>2</sub> H <sub>2</sub>
Ni	0.07	0.01	0.1–20	Air–C <sub>2</sub> H <sub>2</sub>
Os	1.0	0.1	1–300	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
P	120	40	400–30 000	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Pb	0.1	0.01	0.1–30	Air–C <sub>2</sub> H <sub>2</sub>
Pd	0.05	0.01	0.1–15	Air–C <sub>2</sub> H <sub>2</sub>
Pr	20	8.0	100–5000	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Pt	1.0	0.1	1–300	Air–C <sub>2</sub> H <sub>2</sub>
Rb	0.05	0.009	0.1–15	Air–C <sub>2</sub> H <sub>2</sub>
Re	8.0	0.8	10–2000	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Rh	0.1	0.005	0.05–30	Air–C <sub>2</sub> H <sub>2</sub>
Ru	0.4	0.08	1–150	Air–C <sub>2</sub> H <sub>2</sub>
Sb	0.3	0.04	0.4–100	Air–C <sub>2</sub> H <sub>2</sub>
Sc	0.3	0.05	0.5–80	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Se	1.0	0.5	5–250	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Sm	6.0	1.0	10–1500	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Si	1.5	0.25	3–400	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Sn	0.7	0.1	1–200	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Sr	0.04	0.002	0.02–10	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Ta	10	2.0	20–3000	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Tb	7.0	0.7	7–2000	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Te	0.2	0.03	0.3–60	Air–C <sub>2</sub> H <sub>2</sub>
Ti	1.0	0.08	1–300	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
Tl	0.2	0.02	0.2–50	Air–C <sub>2</sub> H <sub>2</sub>
Tm	0.3	0.02	0.2–100	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>
U	100	40	400–30 000	N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>

*Continued*



Table 2 Continued

Element	Characteristic concentration ( $\mu\text{g ml}^{-1}$ )	Detection limit ( $\mu\text{g ml}^{-1}$ )	Normal range ( $\mu\text{g ml}^{-1}$ )	Flame type
V	0.7	0.07	1–200	$\text{N}_2\text{O}-\text{C}_2\text{H}_2$
W	5.0	1.0	10–1500	$\text{N}_2\text{O}-\text{C}_2\text{H}_2$
Y	2.0	0.2	2–500	$\text{N}_2\text{O}-\text{C}_2\text{H}_2$
Yb	0.06	0.004	0.04–15	$\text{N}_2\text{O}-\text{C}_2\text{H}_2$
Zn	0.008	0.0008	0.01–2	$\text{N}_2\text{O}-\text{C}_2\text{H}_2$
Zr	9.0	1.0	10–2000	$\text{Air}-\text{C}_2\text{H}_2$

Both the characteristic concentrations and detection limits are quoted for the most sensitive line. Data based on information supplied by Varian Ltd., Walton-on-Thames, UK, for the SpectrAA series of spectrometers.

concentrated form in order to avoid any undue dilution of the sample where trace elements are to be determined.

Nonaqueous solutions may also be nebulized directly providing the viscosity is similar to that of water for which most nebulizers are designed. Both white spirit and methyl isobutyl ketone fulfill this requirement and these may be used as diluents for other organic liquids. Standards may be prepared in the pure basic solvent.

### Soils and Sediments

Clearly some form of sample pretreatment is required for soils and sediments. Total levels may be obtained following sodium carbonate–boric acid fusion and the dissolution in hydrochloric acid employing lanthanum as a buffer and releasing agent. If the determination of silicon is not required, it may be volatilized as silicon tetrafluoride using hydrofluoric acid, although some calcium may also be lost as calcium fluoride. For many samples, however, it may be more appropriate to determine the exchangeable cation content of the sample. Here, the sample may be shaken with an extractant solution, for example,  $1 \text{ mol l}^{-1}$  ammonium chloride, ammonium acetate, or disodium EDTA, prior to filtration and analysis. Where final solutions contain more than  $\sim 0.5\%$  of dissolved material, the standards should also contain the major constituents, even where no chemical interference is expected, in order to match the viscosity and surface tension and avoid ‘matrix’ effects.

### Plant Material

Plant material should be dried at  $40^\circ\text{C}$ , crushed to pass through a  $0.5\text{--}0.7 \text{ mm}$  sieve, and then oven-dried at  $105^\circ\text{C}$ . The dried material must then undergo dissolution using a suitable digestion procedure such as treatment with hydrogen peroxide–sulfuric acid mixture, or sulfuric, perchloric, nitric acid mixtures. Care must always be taken to avoid the loss of volatile elements such as lead and cadmium during digestion. Contamination from reagents should also be avoided. It is also possible to extract a number of elements quantitatively from plant material and organic residues by boiling in hydrochloric acid. Both ashing and acid extraction methods may be used as a preliminary step to concentration by solvent extraction of elements that are normally too low in concentration to be measured directly in a flame.

See also: **Atomic Absorption Spectrometry: Principles and Instrumentation.**

### Further Reading

- Cresser MS (1994) *Flame Spectrometry in Environmental Chemical Analysis: A Practical Guide*. Cambridge: Royal Society of Chemistry.
- Cullen M (ed.) (2004) *Atomic Spectrometry in Elemental Analysis*. Oxford: Blackwell.
- Ebdon L, Evans EH, Fisher A, and Hill SJ (1998) *An Introduction to Analytical Atomic Spectrometry*. Chichester: Wiley.
- Haswell SJ (ed.) (1991) *Atomic Absorption Spectrometry*. Amsterdam: Elsevier.

## Electrothermal

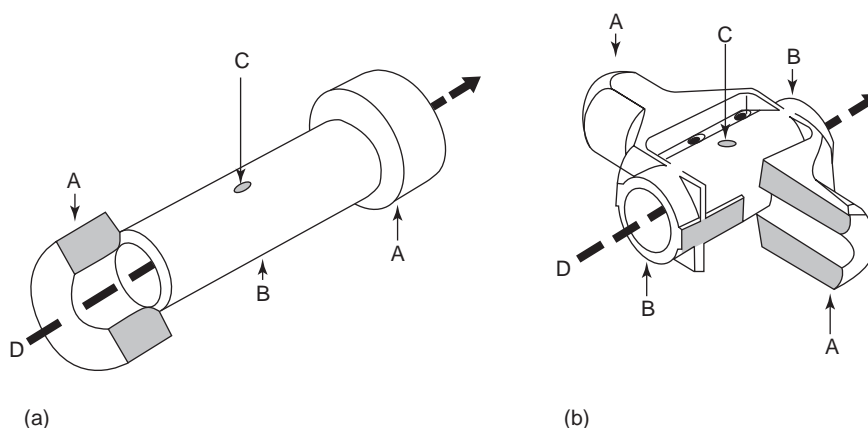
**P Fodor and I Ipolyi**, Budapest Corvinus University, Budapest, Hungary

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by W Frech and D C Baxtes, pp. 228–236, © 1995, Elsevier Ltd.

### Introduction

Electrothermal (ET) atomizers, which were first used for analytical work by L'vov in the late 1950s, typically consist of a tube of electrically conducting material, usually manufactured from graphite (hence the



**Figure 1** End-heated (a) and side-heated (b) electrothermal atomizer configurations. A, water-cooled graphite electrical contact cylinders; B, graphite tube; C, sample injection port; D, light path of spectrometer.

name graphite furnace), mounted in the light path of a spectrometer (see **Figure 1**). Metals such as tungsten and tantalum are also used to construct ET atomizers, in particular for the determination of carbide-forming elements. After the sample to be analyzed has been introduced, via a small aperture in the upper tube wall, the atomizer is heated resistively through a series of controlled temperature stages. Nowadays computer-controlled programs regulate the starting and final temperature, the rate and timing of the graphite tube heating, and even the gas (Ar or N<sub>2</sub>) flush of the tube. During the initial steps in this procedure, the sample is dried and thermally pretreated (pyrolyzed), facilitating the selective volatilization of matrix components, which are purged from the atomizer by a flow of argon gas. The tube is then rapidly heated (1000–2000°C s<sup>-1</sup>) to a sufficiently high temperature (1000–2700°C) to atomize the element of interest, during which time the transient absorbance signal is recorded.

Because the material that is used to form the tube surface is practically impermeable to gas (pyrolytic graphite), the atomic vapor is contained in the atomizer for a relatively long time, normally for a few tenths of a second. This long residence time, together with the high degree of analyte atomization caused by the reducing environment, leads to a factor of  $\sim 10^3$  increase in sensitivity compared with flame atomic absorption spectrometry (AAS). In contrast to the latter, solutions, slurries, and solids can be conveniently analyzed and the possibility for *in situ* removal of matrix components significantly reduces the risk from interferences. The high optical transparency of the argon atmosphere in the ET atomizer favors light transmission, which is important for measurements in the wavelength region 190–230 nm.

## Atomizer Design and Temperature Characteristics

Equation [1] is a simplified expression for the dependence of the peak area or integrated absorbance signal,  $A_{\text{int}}(s)$ , on the atomizer geometry:

$$A_{\text{int}} = k(T, P) N \frac{\tau_R}{\sigma} = k(T, P) N \times \frac{l^2}{8D(T, P)} \times \frac{1}{\pi r^2} \quad [1]$$

where  $k$  (cm<sup>2</sup>) is the temperature ( $T$  (K))- and pressure ( $P$  (Pa))-dependent absorption coefficient,  $N$  is the total number of analyte atoms in the sample,  $\tau_R$  (s) is the mean residence time of analyte atoms in the absorption volume,  $\sigma$  (cm<sup>2</sup>) is the cross-sectional area of the ET atomizer, and  $l$  (cm) and  $r$  (cm) are the length and inner radius of the tube, respectively, and  $D$  (cm<sup>2</sup> s<sup>-1</sup>) is the diffusion coefficient or ‘interdiffusivity’ in terms of the Chapman–Enskog theory. This equation is valid if the concentration of atoms decreases linearly from the center to both ends of the tube, and assumes that all the analyte is converted to free atoms. Furthermore, the temperature (and pressure) in the atomizer should be constant to avoid changes in  $k$  and  $D$  during the development of the absorbance signal. The peak height shows a more complicated dependence on the atomizer heating rate and rate of atom formation, in addition to the factors included in eqn [1].

For optimum sensitivity, the tube should be as long as possible and have a rather small diameter. However, practical considerations dictate that the tube radius should be large enough to transmit a sufficiently high-radiation flux and to introduce large sample volumes or masses. Limitation on the tube length is placed by the need to provide high heating rates at reasonable power consumptions. Thus, the

tube dimensions must constitute a workable compromise between the aforementioned aspects and high sensitivity.

The most important feature of ET atomizers is their heating characteristics. In ET-AAS they determine factors such as the magnitude of interferences, background absorbance, and memory effects, as well as sensitivity. Thus, a clear understanding of the temperature development in various atomizers is a prerequisite to understanding atomizer possibilities and limitations.

Figure 2a shows the change of the wall temperature with time of an end-heated atomizer at the center and two off-center positions. During the initial heating phase the atomizer is isothermal over a substantial tube length (i.e., spatially isothermal) but a temperature gradient develops with time because of heat losses to the water-cooled contacts. It should be mentioned that the onset of this gradient will be shifted toward earlier times if tube heating commences from a higher temperature, i.e., when a thermal pretreatment step immediately precedes atomization, or when a more efficient water cooling is used. Here, the temperature at which atomization/vaporization from the wall takes place cannot be controlled. Depending on its physical properties, the sample normally starts to vaporize at relatively low temperatures often insufficient for complete atomization. The fact that the temperature of the tube may be rising during measurements (i.e., the tube is temporally nonisothermal) introduces various problems, in particular, when the integration method of peak evaluation is used since both the absorption

coefficient and atomic residence time in the atomizer are temperature dependent, as seen in eqn [1]. Elements of low volatility that are vaporized at the relatively high temperatures associated with thermal gradient development may either condense or adsorb on cooler wall regions located away from the tube center. Revaporization may follow during a subsequent atomization sequence (memory effects), since these regions will then initially attain a higher temperature. For the high heating rates used, and in the absence of convection, the atomizer gas-phase temperature closely follows the wall temperature for each cross-section of the tube.

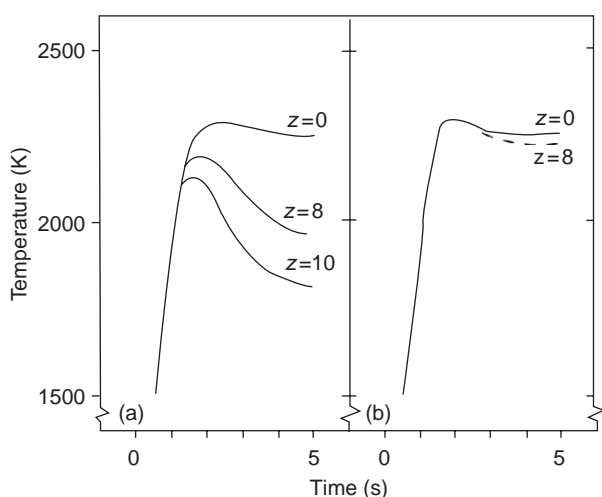
As can be seen in Figure 2b, the side-heated atomizer provides spatial isothermality, which substantially reduces condensation and memory effects.

To overcome problems with analyte vaporization during tube heating, L'vov proposed introducing the sample onto a small platform of graphite (see Figure 3). The heating characteristics of platform-equipped tubes are only slightly different from those without a platform. However, since the platform has a finite heat capacity and is heated primarily by tube radiation, the platform temperature will lag significantly behind that of the tube wall. The sample will therefore be volatilized later in time (relative to wall atomization), and at higher tube and gas-phase temperatures, which normally favor atom formation. Thus, the platform simply reduces interference effects arising from temporal nonisothermality.

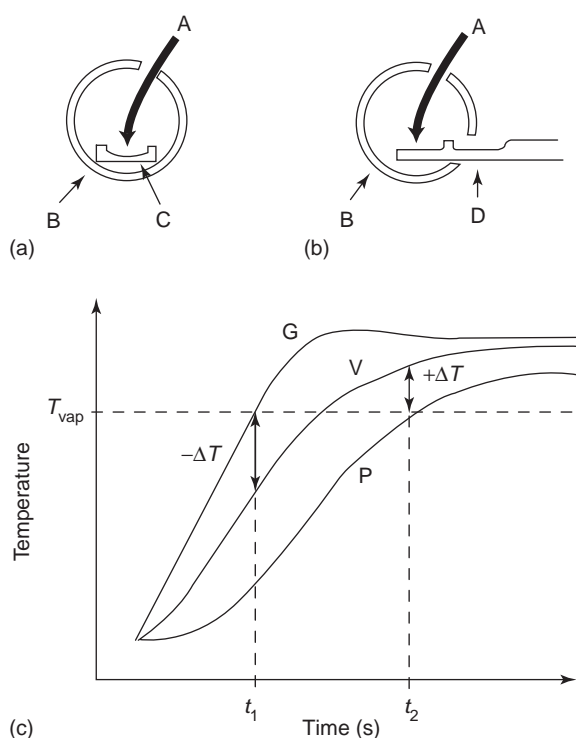
An alternative way to achieve atomization under temporally isothermal conditions is to place the sample on a graphite probe (see Figure 3). The probe is removed from the tube prior to the atomization stage, and then reintroduced once the tube has reached the set temperature. This system is more complex than the platform, and requires the presence of an additional aperture in the tube, which reduces sensitivity. The use of platforms and probes is very efficient in reducing interference effects, but suffers from the disadvantage that the processes of vaporization and atomization are coupled, which may sometimes require compromise conditions.

The temperature characteristics of a constant temperature two-step atomizer are shown in Figure 4. The main advantage of such constant temperature atomizers is their ability to introduce the sample into the measuring zone at any selected temperature and at variable heating rates up to at least  $3000^{\circ}\text{C s}^{-1}$ . This allows selective volatilization of the analyte separately from the matrix, or very slow vaporization of the sample to keep the peak absorbance signal within the linear range.

The description of temperature characteristics in atomizers is further complicated by the fact that the

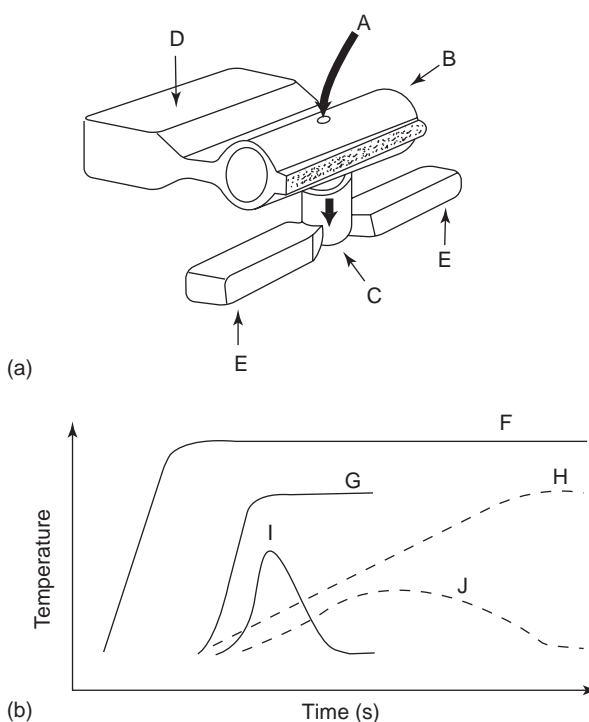


**Figure 2** Temperature distribution as a function of time in the end-heated (a) and side-heated (b) atomizers shown in Figure 1. The graphite tubes are 28 and 17.5 mm long in (a) and (b), respectively, and  $z$  is the distance from the tube center (mm) where the temperature measurements were made.



**Figure 3** Schematic illustration of platform (a) and probe (b) atomization for ET-AAS. The sample is injected (A) through the injection port in the graphite tube (B) onto the platform (C) or probe (D). In (c), the heating profiles for the graphite tube wall (G), the vapor phase (V), and the platform (P) are shown. For atomization from the wall, the analyte vaporizes ( $T_{\text{vap}}$ ) at a time point ( $t_1$ ) when the vapor phase is at a much lower temperature than the graphite tube ( $-\Delta T$ ). Volatilization from the platform commences after a substantial time delay ( $t_2$ ) when the temperature of the vapor phase is higher ( $+\Delta T$ ) than that of the platform ( $T_{\text{vap}}$ ). Furthermore, the vapor-phase temperature is almost constant (temporally isothermal), a prerequisite for the use of peak area signal evaluation, as discussed in connection with eqn [1]. The higher vapor-phase temperature obtained using volatilization from the platform rather than from the wall favors dissociation or atomization of analyte species. Using the probe, the delay time for sample introduction into the vapor phase can be more exactly controlled.

platform or sample compartment is, by design, at a lower temperature than the tube wall during the initial stages of the atomization step. The resulting thermal gradient between the point where the sample is deposited and the tube wall decreases with time. Although this radial temperature gradient is not as pronounced as that along the axis in end-heated furnaces, spatially resolved measurements have indicated that interferences are greater in the immediate vicinity of the platform than at the opposite tube wall. Strictly speaking, none of the atomizers discussed is completely void of temperature gradients, but they differ greatly in the degree of nonisothermality. Toward isothermal atomization the transversely heated graphite atomizer (HGA) has definitely been a



**Figure 4** Schematic illustration of a constant temperature two-step atomizer (a). The sample is injected (A) through the injection port in the side-heated graphite tube (B) and deposited in the sample cup (C). Heating of the tube (at D) and the cup (at E) is achieved by means of two power supplies connected at the water-cooled contact points (D, E). Heating profiles for the tube (F) and cup (G, H) are shown in (b). By reducing the cup heating rate, the rate of sample introduction into the measurement zone (graphite tube) can be controlled to allow large sample masses to be volatilized without exceeding the linear response range. Traces (I) and (J) show the effect of high (G) and low (H) heating rates, respectively, on the signal shape.

major step. The most pronounced advantages in comparison with the longitudinally HGA are the isothermal wall temperature and the very small temperature gradient in the gas phase along the tube axis that can be further reduced by using end-capped tubes.

Integrated platform is an efficient tool to achieve isothermal atomization with respect to time. The platform acts as a thermal ballast with reducing the temporal heating rate during the applied rapid heating and thus creating radial temperature gradients between platform and tube wall. While temperature gradients are always in the direction toward hotter temperatures, therefore beneficial with respect to efficient atomization and the reduction of gas-phase interference, they cannot be ignored regarding their effect on detection. As long as there are temperature gradients, the atomic absorption signal of atoms released from the platform will depend on their location inside the atomizer volume, i.e., on their distance above the platform.

## Mechanisms of Atomization

The understanding of mechanisms leading to atom formation in ET-AAS is still superficial and, to some extent speculative, because of the complex dynamic system under consideration. It should be emphasized that it is difficult to characterize condensed and gaseous phase species in systems where fast reactions of small masses (picograms to nanograms) of analyte are studied in a high-temperature environment.

### Characterization of the Condensed Phase

Since graphite tubes are used almost exclusively in ET-AAS, the following discussion will be confined to this substrate. The base material in graphite tubes for ET-AAS, polycrystalline electrographite, is composed of a large number of disordered crystals, irregularly aggregated to form macroscopic grains, which are held together by carbon binder bridges. This material has anisotropic properties and fulfills the analytical requirements of being resistant to thermal shock and having the appropriate electrical conductivity. The surface is coated with a layer ( $\sim 50\ \mu\text{m}$ ) of high-density pyrolytic graphite, vapor-deposited under carefully controlled conditions of temperature, pressure, and type of hydrocarbon gas. As well as having a low permeability to gases, the pyrolytic graphite coating is relatively resistant toward oxidation due to the high degree of crystalline order. Nevertheless, different types of active sites are present, particularly at the edges of the basal graphite planes, where gases may be adsorbed (analyte,  $\text{H}_2\text{O}$ ,  $\text{CO}_2$ ,  $\text{CO}$ ,  $\text{O}_2$ ,  $\text{H}_2$ ). Some of these species may react with the carbon atoms at active sites, the reaction products being released at various temperatures during heating, depending on the nature of the initial reactants. For example, the reaction between carbon and molecular oxygen proceeds rapidly at temperatures above 1200 K and is kinetically controlled by the reactivity of the surface below 2000 K. Loss of carbon atoms from the lattice creates new active sites, and so the presence of oxidants may change the surface properties.

Oxygen has nevertheless been used successfully during thermal pretreatment of biological sample matrices to both facilitate decomposition of organic material and stabilize the analyte (aluminum, cadmium, lead, selenium) in oxide format active sites. However, the pretreatment temperature must be accurately controlled to prevent oxidation of graphite.

The analyte can be distributed in different ways on the surface of the graphite (tube wall, platform, or probe) as a monolayer, spheres, crystal caps, intercalated between the pyrolytic graphite layers, or adsorbed on active sites. The physical form of the analyte, as well as its chemical properties, helps to

decide the order of atom release as well as the activation energy of the atomization process.

### Characterization of the Gaseous Phase

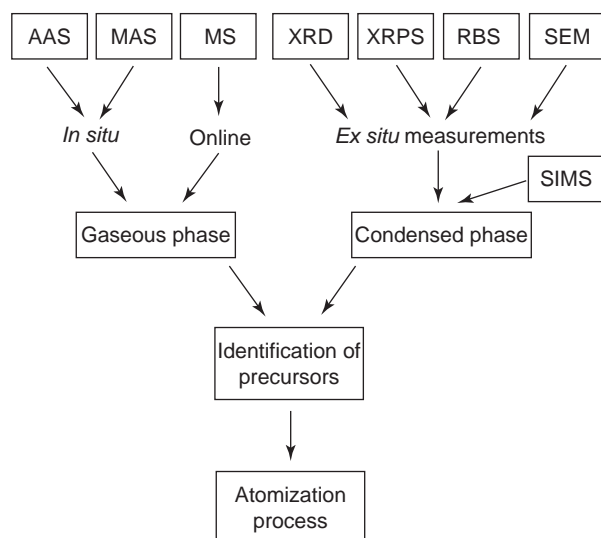
The gaseous phase in an ET atomizer, unlike that in a flame, has a very low-buffer capacity and thus can be influenced dramatically by sample constituents. The chemical composition changes rapidly with temperature and time, depending on the sample matrix, graphite surface properties, thermal pretreatment temperature, impurities in the inert purge gas, and ingress of air through the injection port. For example, in an argon atmosphere of 2000 K, using the type of atomizer shown in **Figure 1a** and without sample addition, the partial pressure of carbon monoxide,  $p(\text{CO})$ , is  $\sim 10^2\ \text{Pa}$ , and at 3000 K  $p(\text{C}_2)$  and  $p(\text{CN})$  are  $\sim 4$  and  $0.6\ \text{Pa}$ , respectively, the latter figure referring to a nitrogen atmosphere in the atomizer. During atomization of a sample pretreated at a high temperature, the partial pressures of carbon monoxide and other oxidant species (such as  $\text{O}_2$ ,  $\text{NO}$ ,  $\text{NO}_2$ , and  $\text{HNO}_3$  from nitrate decomposition) are often increased, while  $p(\text{CN})$  is reduced. This means that oxidants generated during sample matrix decomposition can evolve into the gas phase without complete reduction by the graphite at temperatures below 2000 K.

### Analytical Techniques for the Investigation of Atomization Mechanisms

To establish the processes leading to atom formation it is essential to monitor and identify, *in situ*, the precursors to free atoms in the condensed and gaseous phases. As can be seen from **Figure 5**, no single analytical technique fulfills this requirement.

There is as yet no readily available technique capable of *in situ* registration of analyte species on the atomizer surface during heating. Furthermore, most of the techniques available are not sensitive enough for measurements using typical analytical masses (picograms to nanograms), although secondary ion mass spectrometry has the potential sensitivity. Therefore, for the studies performed so far, microgram masses of analyte have been used and thus there is an uncertainty involved in interpreting the results. Condensation effects and associated back-reactions on the surface make extrapolation of the results to the pre-quenched, high-temperature conditions unreliable.

The most frequently used analytical technique for the identification of gaseous precursors to free analyte atoms is mass spectrometry (MS). Although *in situ* measurements are not possible, the ET atomizer vapor phase can be readily sampled and transferred to the MS detector online, allowing documentation of the actual vaporization- and atomization-related



**Figure 5** Some analytical techniques for the investigation of atomization mechanisms. MAS, molecular absorption spectrometry; MS, mass spectrometry; XRD, X-ray diffraction; XRPS, X-ray photoelectron spectrometry; RBS, Rutherford backscattering spectrometry; SEM, scanning electron microscopy; SIMS, secondary ion mass spectrometry.

processes, using analytically relevant masses, to be provided on the same timescale, but after a certain delay. Initial studies were performed with the ET atomizer held in vacuum, yielding information on the analyte species vaporized from the sampling site. Due to the extremely long mean free paths associated with vacuum conditions, homogeneous and heterogeneous interactions of gaseous analyte species are restricted. Measurements made at atmospheric pressure in the ET atomizer provide additional information on reactions occurring in the gas phase and following collisions with the graphite surface. A selection of the elements studied and the atomization mechanisms elucidated using MS is given in Table 1.

The simplest and most straightforward technique for *in situ* measurement of gaseous species is molecular absorption spectrometry (MAS). However, application of MAS is hindered by the poor sensitivity of the technique and the requirement for high-resolution spectrometers for the identification of the molecular species on the basis of their band structure. In addition, for many of the potentially interesting species suitable absorption lines and reference data are lacking.

Both AAS and MAS have been used to study the radial distribution of analyte species, as a function of temperature and time, in the ET atomizer. Such measurements provide data on the interaction of analyte species with the tube wall and on the effect of air ingress through the injection port.

## Modeling of Atomization Processes

The energy involved in the rate-limiting step of the atomization process can be derived from the initial part of the transient absorbance versus time plot, according to the Arrhenius equation

$$\ln A_t = E_a/RT + A_0 \quad [2]$$

where  $A_t$  is the absorbance at time  $t$ ,  $E_a$  is the activation energy involved in the formation of free atoms ( $\text{kJ mol}^{-1}$ ),  $R$  is the gas constant ( $\text{kJ mol}^{-1} \text{K}^{-1}$ ),  $T$  is the temperature (K), and  $A_0$  is a constant.  $E_a$  is then compared with tabulated thermodynamic data, for example, the enthalpy of evaporation of the material. Similarities in these energy values are often used in postulating on the mechanism governing the appearance of atoms in the gas phase. Several models have been evaluated to determine the order of release in the analyte atom formation process. One of them is a visual method to derive the order of release by comparing various absorbance profiles obtained for different concentrations of analyte assuming that the activation energy is not dependent upon the covered surface in the atomization time interval and from changes of the temperature at the peak maximum. However, conclusions were that the activation energy and the order of release derived from the method are often subjective as the order of release providing the most extended linear range on data sets has to be selected.

In addition, the order of release and the activation energy may significantly vary with analyte concentration; and the temperatures with the time of maximum absorbance are not suitable parameters to correlate with the order of release for the atomization process.

The use of various absorbance profiles for different analyte concentrations has been applied, with activation energy independent of the amount of analyte. In this case, the main problem was due to the fact that the order of release and the activation energy were deduced from different and in many cases very short temperature ranges. Beside the adsorption/desorption processes, a second modeling approach based on condensation/evaporation processes have been proposed where some features of the atomization process are also interpreted – double values of the activation energy, values of  $E_a < \Delta H_T$ , the origin of the empirical relationship between  $E_a$  and  $T_{\text{app}}$  (appearance) as well as the theoretical activation energies and preexponential factors.

A method based on Maxwell's theory with broad temperature range of applicability has also been proposed assuming that the vaporization rate of material

**Table 1** Selection of proposed atomization mechanism from the literature

Proposed mechanism <sup>a</sup>		Precursor	
$M(NO_3)_{(s,l)}$	$\rightarrow$	$MO_{(ad)}$	$Cu(NO_3)_2$
$MO_{(s,l)}$	$\rightarrow$	$MO_{(ad)}$	$BeO$
$MO_{(s,l)}$	$\rightarrow$	$M_{(s)}$	$Ag_2O$
$MO_{(g)}$	$\rightarrow$	$MO_{(ad)}$	$AsO$
$MO_{(ad)}$	$\rightarrow$	$M_{(ad)}$	$CuO$
$MO_{(s,l)} + C_{(s)}$	$\rightarrow$	$M_{(s,l)}$	$PdO, CoO, NiO, CuO, PbO$
$MO_{(s,l)} + C_{(s)}$	$\rightarrow$	$M_{(g)}$	$MnO, MgO$
$M_{(s,l)}$	$\rightarrow$	$M_{(g)}$	$Ag, Pd, Co, Ni, Cu, Pb, Au$
$M_{(ad)}$	$\rightarrow$	$M_{(g)}$	$Cu, AS$
$MO_{(ad)}$	$\rightarrow$	$M_{(g)}$	$BeO$
$MO_{(s,l)} + C$	$\rightarrow$	$MC_{(ad)}$	$SeO_2$
$MC_{(ad)}$	$\rightarrow$	$M_{(g)}$	$SeC$
$MC_{(s,l)}$	$\rightarrow$	$M_{(g)}$	$CuC_2, Mo_2C_2, VC, BaC_2, SrC_2, CaC_2, Al_2C_2$
$MO_{(s,l)}$	$\rightarrow$	$M_{(g)}$	$Al_2O_3^b, In_2O_3, PbO, CdO$
$MO_{(s,l)}$	$\rightarrow$	$MO_{(g)}$	$Al_2O_3, In_2O_3, As_4O_6, PbO, Ga_2O_3$
$MO_{(g)}$	$\rightarrow$	$M_{(g)}$	$AlO, Al_2O, In_2O, As_4O_6, PbO, ZnO, Ga_2O, MnO, CdO$
$M_{(s,l)}$	$\rightarrow$	$M_{2(g)}$	$Se, Co$
$M_{2(g)}$	$\rightarrow$	$M_{(g)}$	$Se_2, CO_2$
$MOH_{(s,l)}$	$\rightarrow$	$M_{(g)}$	$RbOH$
$MO_{(s)} + MC_{(g)}$	$\rightarrow$	$M_{(g)}$	$Al_2O_3$
$M_{(g)} + C_{(s)}$	$\rightarrow$	$MC_{(g)}$	$Al$

<sup>a</sup>s, l, g, and ad are solid, liquid, gaseous, and adsorbed species, respectively.<sup>b</sup>AlO and Al<sub>2</sub>O are formed as well.

from graphite surface depends on the number of atoms having an higher energy than the energy of vaporization process. As Maxwell's theory, the model has been developed on the basis of probability arguments, using one single absorbance signal under nonisothermal conditions and considers the dissipation of the analyte atoms from the simultaneous contribution of the diffusion, thermal expansion, and redeposition processes. If more than one precursor generates the atomic vapor resulting in two or more resolved pulses, the method can independently be applied to each pulse. The main advantages of the method are (1) the activation energy can be evaluated without the need to know the order of release of the

analyte atom formation process and (2) the calculation of the activation energy can be possible in any moment of the atomization process, which means that the method uses virtually ~100% of the rising edge of the peak and not necessarily at the first moment of the absorbance–time profile.

High-temperature equilibrium calculations (HTECs) are useful for studying complex chemical systems, and they enable the simultaneous investigation of condensed and gaseous phases. For the calculations general computer programs are used, and the equilibrium compositions in the solid, liquid, and ideal gas phases are calculated for the given amounts of the elements assumed to be present in the system (input



amounts) using the free-energy minimization principle. Reliable equilibrium compositions for a multi-component system can be obtained provided that accurate thermodynamic data are available for all possible species that may be formed, and that reactions are sufficiently rapid to attain a state close to equilibrium. The latter condition is mainly determined by temperature, time, and the concentrations of the elements present. In ET atomizers, reaction products are continuously removed during heating by convection and diffusion, which complicates the selection of the initial input amounts of the elements constituting the system. A second problem is that some reactions, in particular heterogeneous reactions, may be kinetically controlled, as described above for the reaction between carbon and oxygen below 2000 K. Allowance for such reactions can be made in the program. Advantages of HTECs are that predictions of the effects of varying input amounts on the occurrence of interferences can be made, and complicated chemical systems can be modeled. However, spatial concentration gradients or kinetically driven reactions cannot be accounted for.

A model for the reduction of oxides by gaseous carbides (ROC model) has been proposed. The model involves the formation of metal carbides at the graphite surface, gas-phase transport of these carbides to the metal oxides, reduction of the oxides, and generation of free atoms and carbon monoxide. The reaction is self-sustained through further reactions between the free atoms and carbon.

**Table 1** summarizes a selection of proposed atomization mechanisms. That there is no consensus of opinion is a reflection on the weaknesses of the models and techniques used to investigate the complex reactions involved in atom formation. There is a clear need to develop methods that would help to identify *in situ* the condensed or adsorbed species that are present on the graphite surface.

## Occurrence and Elimination of Nonspectral Interference Effects

By definition, an interference effect occurs when the analytical signal is changed by the sample matrix compared with the reference or calibration standard, typically an acidified aqueous solution. This article is only concerned with nonspectral interferences in ET-AAS; spectral interferences are considered elsewhere. It has been demonstrated in ET-AAS that the atomization efficiency (conversion of analyte to free atoms) is  $\sim 100\%$  for the majority of elements in simple solutions, which means that, in most cases, only negative nonspectral interferences can occur, i.e., the signal can only be reduced by the presence of

a matrix. The types of interference and their elimination are discussed below. It is assumed that standards and samples are quantitatively transferred to the intended location and are dried smoothly without sputtering and irreproducible spreading. Otherwise problems may arise from the resultant variation in the distribution of sample within the atomizer, particularly when the tube is not heated homogeneously over its length. The influence of the magnetic field on complex absorption patterns has also been recorded with high temporal and spatial resolutions, which enables the further investigation of potential spectral interferences in Zeeman-AAS measurements.

## Recent Instrumental Developments in Electrothermal Atomic Absorption Spectrometry

Although instrumental developments are mainly in the hands of manufacturers, research field has proposed the use of continuum light sources instead of hollow cathode lamps to develop multielement systems. Coupling xenon short-arc lamp, double-echelle monochromator or CCD detector to electrothermal vaporization (ETV) proved to extend spectral resolution main as atomic absorption can be measured not only at the center of the absorption line (with maximum sensitivity), but also in its wings (with reduced sensitivity). The dynamic range of determination increased to five to six orders of magnitude, thus eliminating the classical disadvantage of the technique – the limited dynamic range. CCD detectors also provide extra information about the spectral neighborhood of the analytical line in question, resulting in more reliable and accurate background correction than in the case of applying deuterium lamp or Zeeman background correction, especially for difficult spectra. Applying a continuum source atomic absorption system the Zeeman-splitting of the analytical lines can be also be measured in the range of 190–900 nm, providing a good agreement between the theoretical and experimental values of the spilling lines. The capabilities of AAS have undoubtedly been extended with continuum light sources – turning a traditionally single-element technique into a multielement technique. With the replacement of the one-dimensional multiarray detector by a two-dimensional multiarray detector the simultaneous multielement version is to follow and gain acceptance.

## Stabilized Temperature Platform Furnace Concept

This concept should probably be referred to as the stabilized temperature platform atomizer, in line with recent International Union of Pure and Applied Chemists nomenclature recommendations, but will



be called stabilized temperature platform furnace concept (STPF) in the following to conform to the current literature.

Other than the use of the platform (see above), the STPF concept incorporates a number of methodological and instrumental improvements that give conditions approximating to those of a constant temperature atomizer:

1. Rapid heating of the atomizer to make sure that the sample is vaporized from the platform only after the tube has reached its final temperature.
2. Use of modifiers (see below) to stabilize the analyte during thermal pretreatment, to delay volatilization until the temperature in the atomizer is sufficiently high and stable for efficient atomization, and to buffer the composition of the gaseous phase.
3. Stopping the gas flow during the atomization step to attain maximum sensitivity (gas stop conditions).
4. Quantification only with integrated absorbances. Provided that the temperature of the vapor phase is constant, this eliminates the effects of matrix-induced changes in the rate of atom formation, which alter the peak height. Constant temperature is most nearly achieved by using the platform.
5. A short response time of the measurement system to monitor accurately the rapidly changing signal profile.
6. An efficient background correction system.

#### Losses of Analyte During Thermal Pretreatment

Many analytes may form volatile species with other elements in the sample, for example, halides, or be present in compounds that exhibit high-vapor pressures at relatively low temperatures (mercury, arsenic, selenium, organometallics). Such compounds may be volatilized and swept from the tube, in molecular form, prior to the atomization step. These losses can be dealt with by adding a large excess of a reagent (a modifier) to change, *in situ*, the thermochemical behavior of the analyte and the matrix.

**Matrix removal** Volynsky grouped the chemical modifiers applied in ETV-AAS analyses as follows:

- Pd containing modifiers, e.g.,  $\text{Pd}(\text{NO}_3)_2$ ;
- organic acids, e.g., ascorbic, oxalic, and tartaric acids;
- organic-metal complexing reagents, e.g., ethylenediaminetetraacetic acid, cupferron, 1-(2-pyridylazo)naphthol, 1-(2-pyridylazo)resorcinol, 2-(5-bromo-2-pyridylazo)-5-(diethyl-amino-phenol);
- phosphates, e.g.,  $\text{NH}_4\text{H}_2\text{PO}_4$  and  $(\text{NH}_4)_2\text{PO}_4$ ;

- nitrates, e.g.,  $\text{NH}_4\text{NO}_3$ ,  $\text{AgNO}_3$ ,  $\text{KNO}_3$ ,  $\text{Sr}(\text{NO}_3)_2$ , and  $\text{Mg}(\text{NO}_3)_2$ .

However, matrix modification has always been a topic attracting attention and providing a wide variety of choices and not necessarily consistent results.

#### Changes in Atomization Efficiency and Atomic Residence Time

Despite the use of modifiers and stabilized temperature conditions, it is not always possible to remove the matrix completely or to achieve sufficiently high temperatures to ensure complete atomization. Some analytes form extremely thermally stable oxides, carbides, hydrides, halides and sulfides; intermetallic molecules can also be formed during the vaporization of metal matrices, thus reducing the atomization efficiency. In end-heated ET atomizers (see Figure 1a), it is possible for matrix species to condense at the water-cooled tube ends. The analytical signal may then decrease because the analyte residence time in the vapor phase is diminished through interaction between these trapped matrix species and the gaseous analyte atoms. This interference effect is eliminated in spatially isothermal, side-heated atomizers (Figure 1b).

The rate of diffusional losses of atoms can be changed by matrix vapors, if present at sufficiently high pressures at the same time as analyte, and if the diffusion coefficient of the analyte in the inert gas (argon) and in the matrix containing vapor differ appreciably. Since the diffusion coefficient in most matrix vapors is likely to be less than that in argon, this will result in a positive interference. The effect of this is rather small, however, since the gas phase will consist of an argon-matrix vapor mixture.

It should be noted that the interference effects discussed above can be corrected for using the standard additions method combined with the STPF concept.

#### ET-AAS Performance and Applications

Using STPF conditions, the instrumental response is reasonably stable, both between atomizers of the same type and on a day-to-day basis. As a measure of the response the term 'characteristic mass', symbolized by  $m_o$  (pg), has been introduced. This is defined as the absolute mass of analyte yielding an integrated absorbance of 0.0044 s. Characteristic mass values are provided by the instrument manufacturers to specify the performance of the ET-AAS equipment, and can be used to check that certain vital parts of the system (atomizer, light source) are functioning

**Table 2** Characteristics of the ET-AAS technique using STPF conditions

Sample requirement	5–100 $\mu\text{l}$ per determination
Time per determination	30–200 s
Detection limits	
Absolute	Picogram to nanogram
Relative	0.01–100 $\mu\text{g l}^{-1}$
Number of elements	~ 65 (single-element technique)
Analytical working range	2–3 orders of magnitude (nonlinear)
Interference effects	Small if matrix <0.1%

properly. It must be stressed that the characteristic mass should not be used as the only criterion for assessing ET-AAS performance and comparing different instrument systems. For practical ET-AAS work, requiring accurate and precise results, the atomization efficiency, low susceptibility to interference effects, and good detection limits are decisive factors.

Table 2 summarizes some of the important features of the ET-AAS technique using STPF conditions. For end-heated atomizers (see Figure 1a), the platform cannot be used in the determination of elements of low volatility like molybdenum, titanium, and vanadium. However, all elements can be determined using STPF conditions in side-heated atomizers (Figure 1b). Other than the characteristics shown in Table 2, it should be mentioned that modern ET-AAS instruments are fully automated and can work unattended, but experienced personnel are required to optimize parameters (particularly the temperature program) and ensure the quality of the results.

Although the analytical working range may extend to three orders of magnitude, a direct linear relationship between concentration and peak area is typically limited to one or two decades. This obviously leads to problems in applying the standard additions method. However, the replacement of hollow cathode lamps with continuum light sources has brought about improvements, extending the dynamic range of determination to five to six orders of magnitude.

Since ET-AAS is a trace element technique, special equipment is required to minimize the risks for contamination or analyte loss at the critical sample preparation stage of the analytical procedure. To minimize blank levels, additions of reagents to the sample should be limited, and for this reason dissolution techniques such as oxygen combustion and microwave-assisted wet digestion in closed cells are to be recommended. Sulfur- and halide-containing reagents may cause interferences in the determination of certain elements by ET-AAS and should thus be avoided.

ET-AAS is most often applied to the analysis of liquid samples. However, it is also possible to analyze

solids directly, by making use of the specially designed graphite tubes that are available. This approach avoids potentially difficult and time-consuming sample dissolution, and allows the study of the distribution of analyte in the parent material. Limitations include the presence of high concentrations of matrix that may exacerbate interference effects, problems with calibration, and labor-intensive sample introduction, even though attempts have been made to automate solid sampling ET-AAS for routine applications.

Introduction of slurried samples into ET-AAS is widely used for a range of sample types; automatic systems for this sampling procedure are also available.

Advantages of this approach include the possibility of diluting the sample if required, ease of automation, and greater flexibility in the choice of calibration technique.

### ***In Situ* Enrichment and Hybrid Techniques**

Graphite tubes can be used as both the trapping medium and the atomization cell for analytes that can be converted to a volatile species. This approach has a number of attractive features. First, very large sample volumes can be employed for one determination, giving very low detection limits. Second, the analyte is separated from the sample matrix, reducing potential interference problems. Third, atomization can be performed under gas-stop conditions maximizing sensitivity, unlike conventional flow-through cells, and, finally, relatively high temperatures can be used to ensure efficient atomization and conditioning of the tube, which is not possible with quartz tube atomizers.

The combination of hydride generation with ET-AAS has been utilized for arsenic, bismuth, germanium, antimony, selenium, tin, and tellurium, enabling reliable trace element determinations to be performed at the nanogram per liter level. Recent work has demonstrated the improved hydride collection efficiency obtained using palladium-treated graphite surfaces. Attempts to extend the range of analytes amenable to *in situ* enrichment in a graphite tube have resulted in procedures for the generation of nickel carbonyl and tetraethyl lead. It is also possible to reduce mercury species and collect the liberated mercury vapor in tubes coated with gold or lined with platinum gauze.

Flow-injection online preconcentration and separation with ion exchange or sorbent extraction in packed microcolumns and/or precipitation and collection in knotted reactors has proved to extend the capabilities of ET-AAS by allowing relative detection limits to be lowered by two to three orders of magnitude and troublesome matrices to be removed.

ET-AAS has been used as an element-specific detector for gas chromatography using a continuously heated atomizer. One problem with this approach is that the gaseous effluent purges the atomizer volume, decreasing sensitivity. For liquid chromatography (LC), there is an acute requirement for sensitive and specific detectors. However, attempts to couple LC to ET-AAS are difficult, due to the continuous liquid flow supplied by the former, and the discrete sampling mode of the latter. The coupling can be made using a thermospray interface, but vaporization and thermal expansion of the mobile phase entering the heated atomizer again reduces sensitivity. Improvements in LC coupled to ET-AAS may be envisaged using desolvation devices.

Obviously the use of ET-AAS detectors in chromatographic applications is limited to the determination of metal- and metalloid-containing species.

ETV may also serve as sample introduction for inductively coupled plasma (ICP)-atomic emission spectroscopy (AES)/MS providing the possibility of *in situ* sample preparation by selective vaporization of different sample components, using appropriate heating programs. By the reduction/elimination of matrix components, spectral interferences can be minimized and matrix effects in the plasma decreased.

ETV with solid sampling has shown its potential in speciation analysis by separating species according to their volatility. The suitability for the direct and simultaneous determination of methylmercury and inorganic mercury in ETV-ICP-ID-MS coupling has been presented. However, the drawback of the application is the difficulty of achieving proper calibration of all the species involved, and thus, less straightforward two-step approaches based on the use of two different temperature programs (one for the determination of the less volatile species after sample pretreatment and selective removal at low temperatures of the more volatile one) have also been proposed.

**See also:** **Atomic Absorption Spectrometry:** Principles and Instrumentation; Interferences and Background Correction. **Atomic Mass Spectrometry:** Inductively Coupled Plasma; Laser Microprobe. **Liquid Chromatography:** Column Technology.

## Further Reading

- Aller AJ (2001) A model for the determination of the activation energy and the order of release of the atom formation in electrothermal atomization atomic absorption spectroscopy. *Spectrochimica Acta B* 56: 1441–1457.
- Batley GE (1989) *Trace Element Speciation: Analytical Methods and Problems*. Boca Raton, FL: CRC Press.
- Bendicho C and de Loos-Vollebregt MTC (1991) Solid sampling in electrothermal atomic absorption spectrometry using commercial atomizers: A review. *Journal of Analytical Atomic Spectrometry* 6: 353–374.
- L'vov BV (1970) *Atomic Absorption Spectrochemical Analysis*. London: Adam Hilger.
- Schuetz M, *et al.* (2000) Continuum source-atomic spectrometry using a two-dimensional charge coupled device. *Spectrochimica Acta B* 55: 1895–1912.
- Sturgeon RE, Willie SN, Sproule GI, Robinson PT, and Berman SS (1989) Sequestration of volatile element hydrides by platinum group elements for graphite furnace atomic absorption. *Spectrochimica Acta* 44B: 667–682.
- Styril DL and Redfield DA (1993) Perspectives on mechanisms of electrothermal atomization. *Spectrochimica Acta Reviews* 15: 71–123.
- Tsarev DL, Slaveykova VI, and Mandjukov PB (1990) Chemical modification in graphite furnace atomic absorption spectrometry. *Spectrochimica Acta Reviews* 13: 225–274.
- Volynsky AB (1998) Application of graphite tubes modified with high-melting point carbides in electrothermal atomic absorption spectrometry. I. General approach. *Spectrochimica Acta B* 53: 509–535.
- Volynsky AB (1998) Graphite atomizers modified with high-melting carbides for electrothermal atomic absorption spectrometry. II. Practical aspects. *Spectrochimica Acta B* 53: 1607–1645.
- Welz B (1999) Atomic absorption spectrometry – pregnant again after 45 years. *Spectrochimica Acta B* 54: 2081–2094.

## Vapor Generation

**R Sturgeon**, Institute for National Measurement Standards, Ottawa, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The generation of gaseous analytes and their introduction into atomization cells offers several

significant advantages over conventional solution-phase nebulization of samples. These include the elimination of the need for a nebulizer, enhancement of transport efficiency (approaching 100%), and the presentation of a homogeneous vapor to the atomizer. In addition to the widely practiced techniques of covalent hydride generation and mercury cold-vapor formation, which comprise the subject of this article, mention must also be made of the successful

generation and introduction of volatile chlorides (Bi, Cd, Ge, Mo, Pb, Sn, Tl, As, and Zn),  $\beta$ -diketonates (Cr, Fe, Zn, Co, Mn, Cu, Ni, and Pb), and dithiocarbamates (Co and Cu) into atomization cells. Additionally, ethylation, propylation, and phenylation (Sn, Pb, Se, and Hg) as well as carbonylation (Ni) reactions have been implemented.

Hydride generation (and cold-vapor) techniques significantly improve atomic absorption spectrometry (AAS) concentration detection limits while offering several advantages: (1) separation of the analyte from the matrix is achieved which invariably leads to improved accuracy of determination; (2) preconcentration is easily implemented; (3) simple chemical speciation may be discerned in many cases; and (4) the procedures are amenable to automation. Disadvantages with the approach that are frequently cited include interferences from concomitant elements (notably transition metals), pH effects, oxidation state influences (which may be advantageously used for speciation) and gas-phase atomization interferences (mutual effect from other hydrides).

Hydride generation in AAS became popular after 1970 in response to a study by Holak, which demonstrated the analytical potential of this approach for arsenic. Since then, elements of groups IVA, VA, and VIA of the periodic table have been shown to form volatile covalent hydrogen compounds with sufficient efficiency to be of practical analytical use. These include arsenic, bismuth, germanium, lead, antimony, selenium, tin, tellurium, and to some extent, indium and thallium.

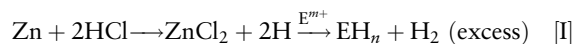
Although the determination of mercury in air by absorption spectroscopy was practiced before the advent of AAS, significant utilization of the cold-vapor technique arose during the 1960s (following the work of Hatch and Ott) and has continued, essentially unaltered in procedure, to the present.

Hydride generation and cold-vapor techniques may be conveniently characterized by three steps: (1) generation of the volatile analyte; (2) its collection (if necessary) and transfer to the atomizer; and (3) decomposition to the gaseous metal atoms (unnecessary for mercury) with measurement of the AA response. Each of these steps will be briefly reviewed prior to considering the analytical performance of these techniques.

## Formation of Covalent Hydrides and Mercury Vapor

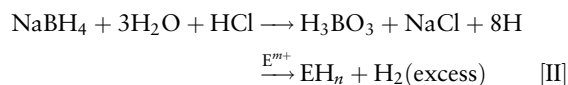
Although several reactions have been utilized for the production of the hydrides, all rely on the formation of atomic hydrogen as a reductant. The classical metal-acid reaction employing  $\text{Zn-HCl}$  is limited to

the generation of  $\text{AsH}_3$ ,  $\text{SbH}_3$ , and  $\text{SeH}_2$ :



where E is the analyte element and  $m$  may or may not equal  $n$ . This system requires that these analytes be present in their lower oxidation states prior to reaction (otherwise they must be prereduced by addition of  $\text{SnCl}_2$  and KI to the acidified sample). Zinc metal is then added and the hydrides, along with excess hydrogen, are evolved. The reaction is slow, difficult to automate, suffers from large blanks due to impurities in the zinc, and is inefficient as a result of incomplete reaction and/or entrapment of the hydride in the precipitated zinc sludge. These factors, along with the availability of more effective reducing agents (such as sodium tetrahydroborate) have served to all but eliminate use of this approach.

Reduction to the hydride with sodium tetrahydroborate via reaction [II] is considerably more efficient:



and can be used to generate the hydrides of antimony, arsenic, bismuth, germanium, lead, selenium, tellurium, and tin. It should be noted that the participation of 'nascent' hydrogen in reaction [II] is currently considered improbable. In addition to increased elemental coverage, this reduction method is superior with respect to efficiency, speed, and reduced contamination. Reaction times are believed to be in the range of 10–30 s, although decomposition of the tetrahydroborate reagent is thought to be complete in a fraction of a second under acidic conditions. Although initially added to acidified samples in the form of solid pellets, this reagent is now almost exclusively dosed into acidified samples as a 0.1–10% (m/v) solution, lending itself to ease of automation. The reagent is somewhat unstable and it is often recommended that it be prepared for storage by the addition of NaOH or KOH at concentrations of 0.1–2%. Others suggest that it be prepared fresh daily, filtered through membrane filters or stored in a refrigerator.

Most recently, amineboranes of the type  $\text{L-BH}_3$  (where  $\text{L} = \text{NH}_3$ ; *tert*- $\text{BuNH}_2$ ;  $\text{Me}_2\text{NH}$ ;  $\text{Me}_3\text{N}$ ) and sodium cyanotrihydroborate(III) ( $\text{NaBH}_3\text{CN}$ ) have been tested for efficacy of generation of elemental mercury and volatile hydrides of As(III), As(V), Sb(III), Sb(V), Bi(III), Se(IV), Se(VI), Te(IV), and Te(VI). All of the reductants are suitable for efficient generation of cold-vapor mercury but only some of the amineboranes are suitable for hydride generation,

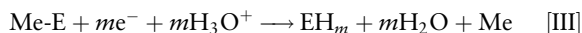
with Se(VI) and Te(VI) remaining unreactive. Reducing power follows the order:  $\text{NaBH}_4 > \text{H}_3\text{N}-\text{BH}_3 > \text{tert-BuNH}_2-\text{BH}_3 > \text{NaBH}_3\text{CN} \geq \text{Me}_2\text{HN}-\text{BH}_3 > \text{Me}_3\text{N}-\text{BH}_3$ . Of note is that the amineboranes and cyanotrihydroborate(III) provide for better control over interferences from Fe(III), Ni(II), Co(II), and Cu(II) than  $\text{NaBH}_4$ . These systems remain to be more thoroughly investigated, especially from an applications viewpoint.

Although aqueous systems are the most frequently encountered, the hydrides of some elements (antimony, lead, and tin) have also been generated directly from nonaqueous media (*N,N*-dimethylformamide, DMF) by the addition of  $\text{NaBH}_4$ /DMF solution as reductant.

Hydrochloric acid remains the acid of choice, although  $\text{H}_2\text{SO}_4$  and  $\text{HNO}_3$  have also been used in generation media. Optimum acidity ranges depend on the element and are also often tailored to suit specific matrices, i.e.,  $1\text{--}9\text{ mol l}^{-1}$  for antimony, arsenic, and bismuth;  $1\text{--}3\text{ mol l}^{-1}$  for germanium;  $0.1\text{--}0.2\text{ mol l}^{-1}$  for tin and lead;  $2.5\text{--}5\text{ mol l}^{-1}$  for selenium; and  $2.5\text{--}3.6\text{ mol l}^{-1}$  for tellurium. Additional constraints enter into the generation of  $\text{PbH}_2$  wherein oxidizing agents such as  $\text{KCr}_2\text{O}_7$ ,  $\text{H}_2\text{O}_2$ ,  $(\text{NH}_4)_2\text{S}_2\text{O}_8$ ,  $\text{Ce}(\text{SO}_4)_2$ , or  $\text{KMnO}_4$  are frequently added to the solution to create lead(IV) intermediate in order to enhance the efficiency of conversion.

The above reduction technique is not used with impunity, including the introduction of contamination (notably tin), interferences from concomitant transition metal ions and the evolution of excessive hydrogen (not considered a problem for AAS detection but potentially detrimental to plasma source atomic spectrometric techniques).

Although infrequently used, electrochemical generation of the hydrides is also possible and has been applied to the determination of arsenic and tin in a batch approach and to antimony, arsenic, germanium, selenium, and tin using a flow-through electrolytic cell. The hydride is generated in the cathodic space of an electrolytic cell, with concurrent oxidation of water in the anodic compartment, as illustrated by the reaction below. Here, Me-E represents the reduced analyte element on the metallic cathode surface (Me):

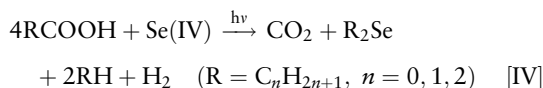


A high mass transfer rate of the analyte to the cathode surface is required for optimum efficiency. Batch sampling, continuous flow, and flow injection (FI) solution delivery coupled with batch reactors, thin-layer membrane separator designs, and tubular cells have been examined. Although no significant

gain is evident with respect to interferences, minimization of contamination through elimination of chemical reducing reagent is possible, offering the lure of ever lower detection limits. Electrochemical generation is reported to be less subject to oxidation state influence, exhibit greater freedom from interferences arising from concomitant elements, and liberate less excess hydrogen than homogeneous generation reactions based on tetrahydroborate(III).

Thermochemical generation of hydrides appears feasible and has been utilized for the determination of arsenic species in the effluent from a liquid chromatograph. It is based on the injection of a thermospray aerosol into a methanol/oxygen flame where pyrolysis of the eluate occurs with subsequent thermochemical derivatization of the analytes and their transfer to a cool hydrogen-rich  $\text{H}_2/\text{O}_2$  diffusion flame for atomization/AAS detection.

More recently, photo-induced generation of the hydrides has been reported utilizing UV-irradiation of the sample in an aqueous medium spiked with low molecular weight acids such as formic, acetic, malonic, etc. The efficiency of the process is greater than 70% for Se and, depending on the species in solution, radical reactions can give rise to various products, including the simple hydrides or alkylated derivatives, as outlined below:



It is self-evident that use of formic acid gives rise to the hydride species. Early results suggest that a number of elements are amenable to such a reaction, including As, Pb, Hg, etc.

Cold-vapor AAS (CV-AAS) occurs in a manner similar to hydride generation AAS in that elemental mercury is formed in solution by reduction to  $\text{Hg}^0$  followed by transport to and detection in an absorption cell. Simple first-order speciation of this element may be achieved using  $\text{SnCl}_2$  under acidic conditions to selectively reduce inorganic mercury, and under basic conditions in the presence of copper(II) ion to reduce both inorganic and organomercury compounds. Until recently, tin(II) chloride solution (5–10% m/v) was almost exclusively used for the reduction of mercury. Currently,  $\text{NaBH}_4$  is finding increasing application for this purpose and it has been suggested that selective reduction of inorganic mercury can be achieved using  $\text{SnCl}_2$  at basic pH followed by subsequent reduction of organomercury compounds in the same solution by addition of  $\text{NaBH}_4$ . This latter reagent also presents the advantage of faster liberation of the mercury from the solution phase and, in some cases, reduced interferences.

Sodium tetrahydroborate is also attractive because of its comprehensive nature in that generation of mercury as well as all other covalent hydrides of interest can be accomplished in the same reaction vessel.

## Vapor Generator

### Hydrides

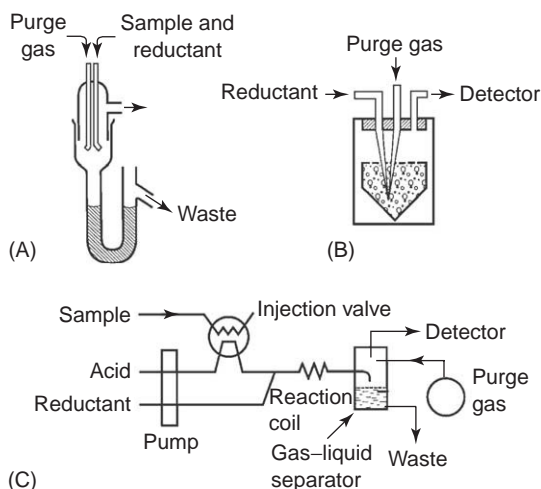
The hydride that is formed is flushed from the generating chamber with argon, helium, or nitrogen. There are essentially three methods used to generate hydrides with  $\text{NaBH}_4$ : (1) continuous systems where sample and reagent are pumped and mixed in a continuous fashion and then passed to some type of gas-liquid separation device; (2) batch systems, wherein a pellet of  $\text{NaBH}_4$  or an aliquot of the reductant solution is added to the sample from which the volatile products are purged with a flow of transfer gas; and (3) FI systems in which discrete volumes of sample are merged with flowing streams of acid and/or reductant. These three arrangements are illustrated in **Figure 1**. Although automation is easily implemented and precision of measurement is often improved with continuous sampling as a result of the steady-state signals, this approach is at present little used with the AAS detection mode. Mixing coils, frequently inserted into flowing reaction systems, are useful only in that they aid in the phase separation of the hydride prior to entry into the actual gas-liquid separator. Such coils are generally not needed to increase the reaction time and in some cases have been shown to have a detrimental effect in that an enhancement in interference is observed. Gas-liquid phase separators have included simple arrangements, such as that illustrated in **Figure 1**, in addition to various

hydrophobic membrane-based systems which provide for dry gas streams as well as shorter response and memory times.

The hydride technique is an absolute procedure in that the measured response is directly proportional to the absolute mass of the analyte element and not to its concentration in the solution. In practice, an aliquot of acid is usually added to the batch vessel followed by an accurately dispensed volume of sample. Depending on the apparatus used, peak-height quantitation may exhibit some sample volume dependence but this is absent when signal integration is used. Most batch reaction vessels for the hydride technique accept a relatively large sample volume (1–50 ml) whereas online FI techniques are flexible in that variable sample loop volumes or timed sample pumping can be used to deliver an almost continuous distribution of desired volumes (typically 0.1–10 ml). FI offers a number of distinct improvements over the batch technique, such as high absolute sensitivity, reduced sample and reagent consumption, reduced interference effects, ease of incorporation of elaborate sample pretreatment, high sampling frequency, and ease of automation. A singular disadvantage remains the small volumes that, compared with batch systems, results in inferior relative sensitivity. As this is more than compensated for by the greater freedom from interferences and high absolute sensitivity, FI techniques are to be preferred for the analysis of real samples.

The transient hydride plume produced by either the batch or FI approach may be directly transferred to the atomization cell because reduction reactions are sufficiently rapid. Alternatively, the hydride is frequently collected in a cold trap (usually a cryogenic U-tube filled with a suitable adsorbent that is subsequently warmed to desorb the analyte) connected to the generator via a  $\text{CaCl}_2$  drying tube, absorbed in solutions of  $\text{AgNO}_3$ , Ag-diethyldithiocarbamate,  $\text{KI/I}_2$ ,  $\text{Ce(IV)/KI}$ , and  $\text{I}_2$  (for subsequent quantification by graphite furnace AAS), or sequestered directly in a preheated graphite furnace prior to atomization/quantitation. These latter schemes permit significant additional preconcentration factors to be achieved while often eliminating or minimizing interferences associated with the generation step (generation kinetics). Most recently, the application of solid-phase microextraction fibers has permitted sampling of the evolved vapors for subsequent desorption and introduction into suitable detectors (such as ICPs and EI-MS sources).

Several new approaches to the generation of the hydrides have recently come to the attention of the analytical community. One utilizes an anion exchanger, in the tetrahydroborate form, packed as a



**Figure 1** Hydride generation approaches: (A) continuous generation; (B) batch generation; and (C) flow injection.

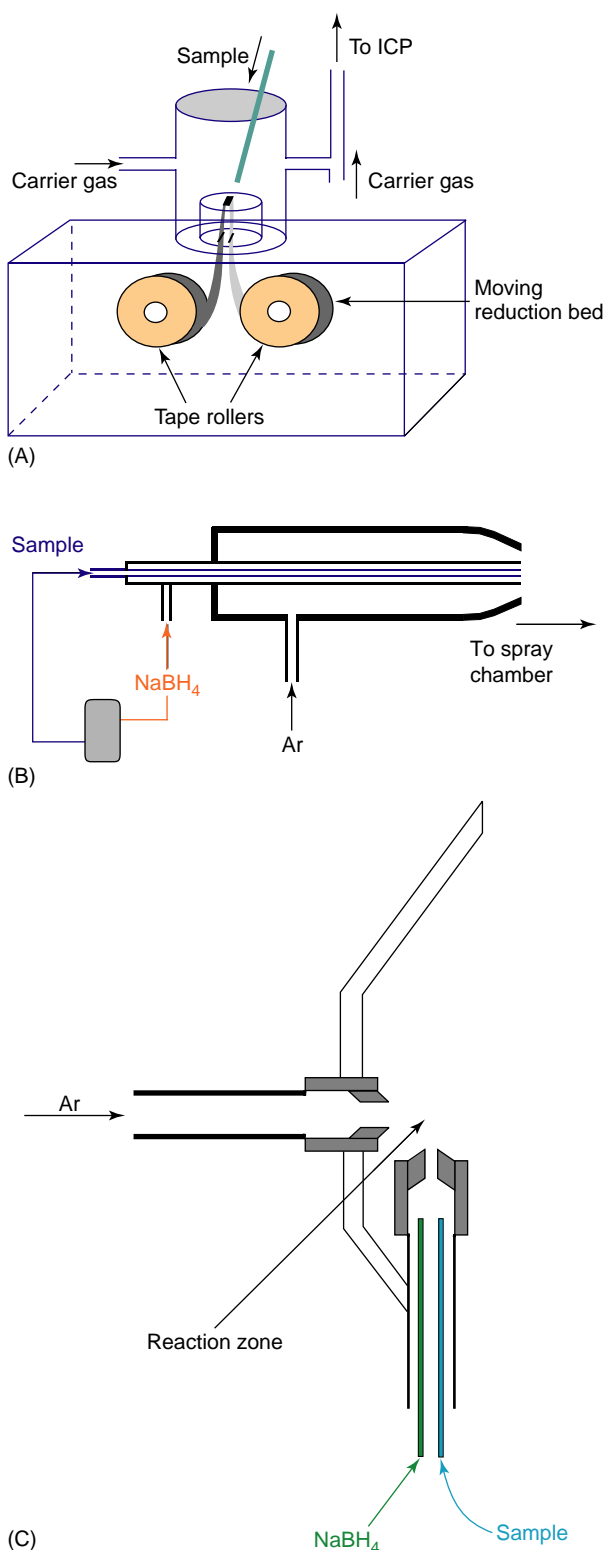
bed in the liquid channel of a gas–liquid separation membrane cell. FI-pumping of an acidic sample through the cell results in the heterogeneous production of hydrides that are rapidly transferred to the measurement cell. The extent of some interference effects is reduced with this arrangement.

Recently, attention has focused on taking advantage of the rapid generation kinetics of the tetrahydroborate–acid reduction reaction to expand the suite of elements amenable to hydride generation. These include the use of a moving bed generator as well as modified parallel path concentric and cross-flow nebulizers, as illustrated in **Figure 2**. Rapid interaction of sample with reductant, while providing for instantaneous gas–liquid separation achieved with these systems, offer the potential for reduction or elimination of interferences from concomitant species in solution. Successful generation of a number of ‘unconventional’ volatile species by reaction of acidified sample solutions with sodium tetrahydroborate has now been reported to include (in addition to Cd and Hg and the classical hydride forming elements): Cu, Ag, Au, Zn, Ni, Pd, Rh, Pt, Ti, Ir, Mn, Co, Fe, and Cr. Little is known of these species, except that they are relatively unstable, requiring rapid gas–liquid or gas–solid separation techniques, and they appear to be molecular in nature. The full scope of elements amenable to such reactions is currently unknown. Reported generation efficiencies range from greater than 50% to less than 1%. Only minimal analytical use of these approaches has been reported to date.

Combined generators, which are designed to take advantage of conventional introduction of liquid samples into atomization sources (such as the inductively coupled argon plasma, or emission or mass spectrometric detection of the analyte) along with simultaneous generation/introduction of the hydride forming elements, have also been proposed. These are usually based on a conventional or modified spray chamber that permits introduction of reductant through an additional channel, permitting its interaction with the acidified sample aerosol.

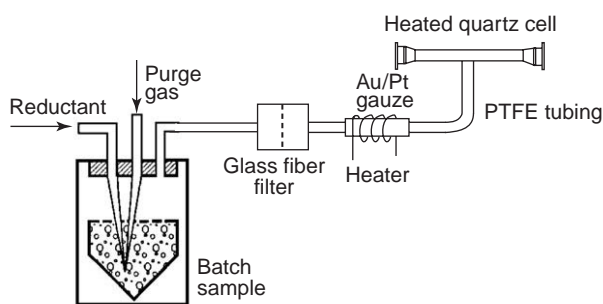
### Cold Vapor

Mercury is the only metallic element that is liquid at room temperature and possesses a significant vapor pressure. As a result of these unique properties, mercury can be determined without an atomization cell simply by reducing it to the elemental state and transferring it to the vapor phase within the optical path of a suitable detection system. Absorption is usually measured at the 253.7 nm resonance line. Similar to hydride generation, the majority of such



**Figure 2** Hydride generation systems: (A) movable bed generator (Reproduced with permission from Tian X-D, Zhuang Z-X, Chen B, and Wang W-R (1998) Movable reduction bed hydride generator coupled with ICP-OES for the determination of some hydride forming elements. *Analysts* 123: 627–632.); (B) modified Meinhard concentric nebulizer; and (C) cross flow nebulizer.





**Figure 3** Typical cold-vapor mercury generation system.

determinations is usually accomplished using batch procedures, although FI systems are now enjoying more widespread usage. **Figure 3** illustrates a typical mercury generation system. Tin(II) chloride historically remains the preferred reductant; although  $\text{NaBH}_4$  is a more powerful reducing agent, its use may increase the risk of certain interferences. These include excessive water droplet formation and carry-over as potential poisoning of amalgamation traps due to cogenerated hydrides if certain precautions are not undertaken. Use of  $5 \text{ mol l}^{-1}$  HCl in the presence of  $200 \text{ mg l}^{-1}$  Fe(III), an alkaline gas wash bottle to eliminate cogenerated  $\text{SeH}_2$ , a drying agent ( $\text{Mg}(\text{ClO}_4)_2$ ) and, most importantly, ensuring that the amalgamation trap is below  $100^\circ\text{C}$  during collection (otherwise it suffers gradual poisoning) serve to minimize such problems. Helium is preferred over other gases as the most efficient sparging and transfer medium and an inline glass fiber filter is often used in place of the chemical desiccant (when  $\text{SnCl}_2$  is used as the reductant) to remove any fine water droplets from the carrier stream (water vapor does not interfere with the absorption detection of mercury). Particular care must be taken in the selection of vessels and tubing in such apparatus due to the mobility and reactivity of mercury. Quartz ware and fluorinated ethylene propylene are often advised because they can be easily cleaned, are nonpermeable, and exhibit little affinity for adsorption of mercury. Polytetrafluoroethylene (PTFE) tubing is superior to other materials for the transfer of mercury vapor.

As with hydride generation, the evolved mercury vapor can either be transferred directly to the optical cell for measurement or trapped for later release. As the mercury is normally liberated slowly from solution over a period of 1–2 min, low intensity signals are generated; consequently, vapor trapping is frequently used to improve concentration detection limits for solution analysis. Enhanced separation and concentration of mercury is usually achieved by amalgamation on a noble metal trap, from which it is subsequently thermally desorbed (at  $500\text{--}700^\circ\text{C}$ ).

Currently, a gold–platinum gauze (90% Au, 10% Pt) is found to be most useful for this purpose because of its performance characteristics, which include high surface area, physical stability, thermal conductivity, and ease of cleaning. These properties lead to optimal signal reproducibility and generally an increase in sensitivity by 10-fold. Occasionally, two-stage gold amalgamation procedures have been utilized wherein the thermally desorbed mercury is recollected on a ‘standardized’ gold trap, which has been carefully calibrated, and is then desorbed into the spectrometer/detector. Water vapor that has collected on the first trap can be evaporated and interferences due to organic compounds (if air sampling has been used) that are oxidized during the first desorption, are thereby avoided. Alternatively, the generated mercury can be directed into a standard graphite tube atomizer (for GF-AAS), which has been lined with Pt gauze to effect *in situ* preconcentration. Subsequent high temperature heating desorbs the mercury directly into the optical beam of the spectrometer, serving to further enhance sensitivity.

## Atomization and Detection

### Hydrides

Conventional acetylene-based flame systems have found little use as atomization cells for hydrides. The relatively cool argon (entrained air)–hydrogen or simple air–hydrogen flame is advantageous as it exhibits low background absorption at lower wavelengths (15% at  $193.7 \text{ nm}$ ). However, the excess hydrogen generated along with the hydrides often perturbs the flame, changing its composition and absorption characteristics, and the low kinetic temperature makes it more susceptible to interferences.

Currently, the most popular atomization source for hydride generation AAS is the heated quartz T-tube (typically  $10 \text{ mm}$  diameter  $\times$   $100\text{--}150 \text{ mm}$  length). Both argon–hydrogen and air–acetylene flames have been utilized to heat open-ended silica tubes to which the hydrides are delivered in a stream of carrier gas from the generator via the central arm of the T. The quartz tube can also be heated electrically ( $700\text{--}1000^\circ\text{C}$ ) with the advantage of longer analyte residence time in the optical path and the possibility of obtaining the optimum atomization temperature for each element. Often, the tube is sealed with removable quartz windows at either end and fitted with nipples at the extreme ends as exits for the gas flow. These features result in improved sensitivity over the flame heated cells. Deterioration and aging of the interior surface of the quartz tube invariably occurs, leading to a decrease in sensitivity



and precision. This has been attributed both to an irreversible devitrification of the quartz to a less satisfactory  $\beta$ -cristobalite structure and to contamination of the atomizer surface caused by deposition of liquid particles of the sample or small droplets that are carried from the generator into the tube. In the latter case, careful cleaning by soaking in 40% (m/v) HF often restores the sensitivity.

Also used for atomization sources are the flame-in-tube devices, in which the excess hydrogen generated during the reduction step is used to carry the hydrides to a T-shaped quartz tube. A small amount of oxygen or air is added to support the combustion of a small flame. Although more complicated than the simple quartz tube, this system does not exhibit any significant analytical advantages over the latter but has been found useful for mechanistic studies relating to atomization processes.

A recent, second generation modified flame-in-tube atomizer consisting of multiple miniflame ports located along the axis of the quartz tube provides for an extended region of active atomization to realize an enhanced absorption path length and reduced mutual interference effects from other cogenerated hydrides, as discussed later.

Continuously heated graphite furnaces have seen limited use as atomizers in which the hydride enters a preheated furnace (1800–2300°C) and is atomized during its transit time through the device. Although more bulky and difficult to operate, these systems offer a continuously clean and reproducible atomization surface, permit variable atomization temperatures to be used, and result in slightly improved limits of detection than those arising with the heated quartz tube. Tin, however, can suffer adsorption on any graphite transfer lines and reactions between water vapor, hydrogen, and graphite at high temperatures give rise to increased background absorption.

An alternative system that is attractive for batch generation techniques, particularly FI-based approaches, is the use of the graphite furnace as both

the preconcentration cell and atomizer. The liberated hydrides are swept by a stream of inert gas into a warm (300°C) graphite furnace and directed onto the surface of a small (4  $\mu$ g) previously reduced deposit of palladium which serves to catalytically decompose and trap the hydride. Following completion of the generation phase, the sequestered sample is atomized by cycling the furnace through a conventional high temperature heating program and the signal recorded. This approach permits additional preconcentration to be achieved, is less prone to interferences, can be optimized for individual elements, and readily permits Zeeman-based background correction to be implemented. Concentration detection limits are one to two orders of magnitude superior to all other hydride generation approaches.

In addition to these 'online' atomizers, the graphite furnace has also been used in its traditional role as an offline device for the analysis of conventional liquid samples of sequestered hydrides. Aliquots of absorbing solutions used to tap the generated hydrides are injected into the furnace, atomized and quantitated in the usual manner.

Detection systems employed with hydride generation approaches are conventional AA spectrometers, usually fitted with intense electrodeless discharge or hollow cathode lamp sources. Quartz tube cells are of suitable dimensions to be compatible with the optical systems of all modern spectrometers. Background correction is usually achieved in double-beam optics using deuterium sources, and Zeeman-effect background correction can be implemented when the graphite furnace is used as the atomization cell.

Table 1 summarizes the limits of detection of a number of hydride forming elements reported for the various generation–detection methods. For comparison with continuous sampling techniques, a 10 ml sample volume has been assumed for *in situ* trapping in the graphite furnace (although larger volumes are easily accommodated) and a 500  $\mu$ l volume for FI approaches. It is clear that, despite the small sample

**Table 1** Concentration detection limits ( $3\sigma$ ) given in nanogram per milliliter for hydride generation and cold-vapor AAS techniques

Element	$\lambda$ (nm)	GF-AAS (20 $\mu$ l)	Quartz tube, continuous	CF-AAS <sup>a</sup> , <i>in situ</i>	Quartz tube, FI (500 $\mu$ l)
As	193.7	1.0	1.2	0.0043	0.40
Bi	223.1	0.5	0.3	0.0032	0.10
Ge	265.1	0.8	3.8	0.0017	—
Pb	283.3	0.3	0.6	0.0240	0.40
Se	196.0	1.5	2.7	0.0008	0.60
Sn	224.6	1.0	0.7	0.0037	0.30
Sb	217.6	0.8	0.7	0.0029	0.08
Te	214.3	0.5	2.3	—	0.20
Hg	253.7	1.0	0.02 (0.001) <sup>b</sup>	0.002	—

<sup>a</sup>Based on a 10  $\mu$ l sample volume.

<sup>b</sup>Gold trap amalgamation, 10  $\mu$ l sample volume.

consumed for FI work, relative concentration detection limits are as good as those for continuous sampling. Additionally, 1000-fold improvements in detection power are readily achieved using *in situ* trapping.

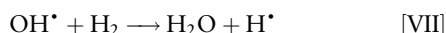
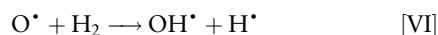
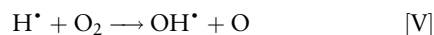
### Cold Vapor

As the cold-vapor mercury sample is already in the atomic state, there is no need of an atomizer, *per se*. The vapor, transferred directly from the cell or desorbed as a plug from a heated amalgamation trap, is commonly swept into a moderately heated (resistance wound heating to 200°C) 10 cm quartz T-tube located within the optical beam of a conventional AA spectrometer. Attenuation of an intense electrodeless discharge lamp line source at 253.7 nm is used as a measure of the absorption. Alternatively, dedicated continuum source AA-based spectrometers fitted with long path absorption cells (30 cm) are frequently used to increase sensitivity and detection limit.

Similar to hydride generation, graphite furnaces may be used both as the trapping medium and the atomizer for mercury. This may be accomplished by directing the evolved mercury from the generation cell onto an amalgamation medium (a reduced solution of gold or a gold-palladium gauze) inserted into the graphite tube for *in situ* preconcentration. Detection limits are generally improved threefold over conventional amalgamation trapping-quartz tube detection systems. Table 1 summarizes the reported detection capabilities for mercury using these various approaches.

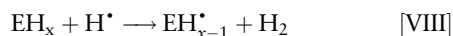
### Atomization Mechanisms

Atomization of the hydrides is currently believed to proceed via interaction with free hydrogen radicals; oxygen also plays an active role. In argon/hydrogen diffusion flames and quartz tube atomizers, a cloud of hydrogen radicals is formed by reactions between hydrogen and oxygen:



The concentration of hydrogen radicals is several orders of magnitude higher than that of hydroxy radicals. In the quartz tube, this occurs either in a flame burning at the end of an oxygen delivery capillary (for a flame-in-tube device) or at the beginning of the hot zone for an externally heated tube (above 600°C). Only a small portion of the volume of the atomizer is filled by the cloud, as determined by the gas dynamics, geometry of the tube and

its temperature profile. The hydride is atomized within the radical cloud in accord with sequential collisions:



The number of hydrogen radicals is primarily determined by the oxygen supply to the atomizer and, if insufficient, thermal decomposition of the hydride may occur (if the temperature is high) and lead to the formation of dimeric (and tetrameric for arsenic) species or, in the absence of hydrogen, to oxides, with consequent loss of sensitivity. Decreasing the inner diameter of the quartz tube serves to increase the radical cloud density but, as analyte atom attrition occurs predominantly by reaction with the quartz wall, the consequent less favorable surface area-to-volume ratio precludes any major enhancement in performance. Deterioration in sensitivity noted with tube aging may be accounted for by the formation of active sites, which serve to deplete the hydrogen radical concentration and/or scavenge analyte atoms.

Use of the multiple microflame quartz tube atomizer provides for a significant improvement in performance of this device. Recurrent analyte atomization occurs over the whole optical tube length, achieved by production of H-radicals at multiple points within the tube by oxygen microflames burning in the hydrogen-containing atmosphere. This feature serves to enhance the range of linearity in response to 200 µg/l for Se and Sb and 100 µg/l for As. A 10–100-fold improvement in tolerance to interferences from cogenerated analytes is also achieved.

It is believed that similar mechanisms are involved in the atomization from a graphite furnace operating in the direct transfer continuous mode of measurement. However, the probability of thermal dissociation increases substantially as the temperature of the source increases, with the result that radical mechanisms probably decline in significance above 2000°C. Atomization of the hydride elements trapped on reduced palladium within the graphite furnace appears to proceed at high temperature as a simple first-order desorption process from the surface of the deposit.

### Interferences

#### Hydrides

Chemical interferences may occur in the liquid phase during formation and release of the hydride or in the gas phase during its transport to the atomizer or

within the atomizer. The extent and severity of these interference effects vary widely and are dependent on the instrumentation and hydride generation system used. A 'physical' interference, often referred to as a kinetic interference, may arise as a result of a difference in the release rate of the hydride from solution due to a volume effect or perhaps to sample foaming. These interferences are only encountered in direct systems where the measurement is performed online; they do not occur when the hydride is collected.

Spectral interferences are essentially absent in hydride generation AAS because the analyte is completely separated from the sample matrix. Minor fluctuations that may occur in the baseline during sample introduction are easily compensated for with conventional background correction systems.

Although strictly 'chemical' in nature, oxidation state interferences are well recognized and generally easily dealt with. For elements of group IVA (germanium, tin, and lead), little is reported as to the influence of oxidation state on sensitivity, although it is clear that generation efficiency is remarkably enhanced for lead if reduction is performed in the presence of strong oxidizing agents (suggesting the formation of an intermediate Pb(IV) species). For elements of group VA (antimony, arsenic, and bismuth), the sensitivity difference in peak-height mode of quantitation is less than twofold for the +3 and +5 oxidation states. As bismuth exists virtually in the +3 state, there is no concern. Under sufficiently acidic conditions, response from both arsenic(III) and arsenic(V) is the same. On- or offline prereduction of the sample with KI-ascorbic acid or KI-HCl is easily implemented for both arsenic and antimony if the sample preparation or dilution steps preceding measurement did not leave the analyte in the trivalent state. For the elements of group VIA (selenium and tellurium), only the +4 oxidation state is reactive and prereduction is required. This is often accomplished using hot 4–6 mol l<sup>-1</sup> HCl. Advantage is frequently taken of the differential reactivity of the separate oxidation states of antimony, arsenic, and selenium to effect a first-order speciation of the element in the sample with respect to this parameter. This is easily implemented for antimony and arsenic simply by selectively generating the hydride from the +3 state at high pH and from both forms at low pH.

It is also important that the various chemical species of the analyte be converted to a common form. Some of the organoforms of these elements commonly found in biological fluids and tissues are inert toward hydride formation (e.g., selenomethionine and arsenobetaine) and care must be taken to ensure that the sample has been completely oxidized by,

for example: treatment with peroxydisulfate at pH 2 for organoselenium compounds; alkaline peroxydisulfate for arsenic; and a mixture of KBrO<sub>3</sub>-KBr-HCl for tin and bismuth (note that in all cases it is necessary to reduce all oxidation states 'back' to the lower valence state prior to analysis). Alternatively, sample analysis before and after extensive oxidative treatment again affords a further route for the methodologically defined speciation of the element. Additionally, methylarsonate and dimethylarsinate are reduced to methylarsine and dimethylarsine, respectively, and may be chromatographically determined following cryogenic trapping.

Few reports of the interference of acids (except HF) on the determination of the hydride forming elements have appeared. Only tin and lead exhibit a relatively strong pH dependence so that they are normally determined in buffered media. In most cases, dilute hydrochloric acid (<1 mol l<sup>-1</sup>) is preferred, as higher concentrations give rise to increased contamination.

Of the common metallic elements, significant interferences are encountered in the presence of high concentrations of other hydride forming elements, iron, copper, nickel, cobalt, and the platinum group elements. This effect is not dependent on the analyte-to-interferent ratio, but rather on the concentration of the interferent in the analysis solution. Although several theories exist to account for these effects, the consensus is that preferential reduction of the interfering ion occurs and the resulting finely dispersed precipitate scavenges (and decomposes) the hydride formed in the secondary reaction. This explanation is supported experimentally, in that when concomitant element precipitation is avoided through the use of more concentrated acid solutions (5 mol l<sup>-1</sup> HCl) and/or reduced concentration of tetrahydroborate reagent (0.5 versus 3%), the range of interference free determination can be enhanced 10–100 fold. Other theories suggest that, even in the absence of precipitation, signal suppression can occur due to the formation of a soluble complex between a lower than normal oxidation state of the interferent and the analyte. Excess NaBH<sub>4</sub> serves to stabilize this intermediate. In the presence of high concentrations of other hydride-forming elements, competitive reactions between the different hydrides and possible loss of compound on metal precipitates may occur.

Gas-phase interference effects have been noted due to the mutual interference from the presence of other hydride-forming elements. As above, the extent of the interference is only dependent on the concentration of the interferent and not on the analyte-to-interferent ratio. As atomization of the hydrides occurs by collisions with hydrogen radicals, a 'radical

population interference' may arise when an excess of cogenerated hydride depletes the hydrogen radical concentration to a level insufficient to completely atomize the analyte hydride. This theory is consistent with all observations made using heated quartz T-tube atomizers. Indeed, use of the multiple micro-flame quartz tube atomizer significantly reduces this problem, as noted earlier. Gas-phase interferences in the graphite furnace atomizer have been attributed to the formation of diatomic molecules between the analyte and cogenerated interferent and can be eliminated by use of high temperature.

A number of approaches have been taken to minimize or eliminate interferences. These include: increasing the acidity of the sample solution; reducing the concentration of the tetrahydroborate reagent; adding chelating or complexing agents to the solution (L-cysteine and L-cystine are particularly attractive); the addition of interference releasing elements such as iron, tellurium, and copper, and the preseparation of the analyte from the matrix.

FI techniques are particularly attractive for implementing such procedures, as they conveniently allow for rapid, elaborate sample pretreatment with minimal risk of increased contamination. This includes the online oxidation and prereduction of various chemical forms of the analyte, the addition of releasing or masking agents, as well as the separation of the analyte from major matrix components by ion exchange.

### Cold Vapor

No spectral interferences occur with the cold-vapor technique; water vapor does not absorb at the 253.7 nm line but water droplets carried into the cell may result in source attenuation. As a consequence, the cell is often operated at 200°C to prevent condensation.

Chemical interferences occur infrequently but include silver, arsenic, copper, iodine, antimony, and selenium. Tolerance to all these species except iodine and selenium is higher when  $\text{SnCl}_2$  is used as the reductant as compared to  $\text{NaBH}_4$ . Interference by platinum group metals using  $\text{NaBH}_4$  reductant is similar to that encountered with the hydride forming elements. Species containing sulfhydryl groups retard the release of mercury from the sample but this interference effect is eliminated with amalgamation and/or signal integration. Organomercury compounds can be conveniently preoxidized using  $\text{KBrO}_3$ - $\text{KBr}$ - $\text{HCl}$  reagent prior to tetrahydroborate reduction.

Partial or complete poisoning of the amalgamation trap is more likely to occur when  $\text{NaBH}_4$  is used as a reductant and the precautions outlined earlier are not

adhered to. Frequent cleaning in nitric acid minimizes this problem.

### Applications

Hydride generation techniques are superior to direct solution analysis in several ways. However, the attraction offered by enhanced detection limits is offset by the relatively few elements to which the technique can be applied, potential interferences, as well as limitations imposed on the sample preparation procedures in that strict adherence to valence states and chemical form must be maintained. Cold-vapor generation of mercury currently provides the most desirable means of quantitation of this element, although detection limits lower than AAS can be achieved when it is coupled to other means of detection (e.g., nondispersive atomic fluorescence or microwave induced plasma atomic emission spectrometry).

Applications of both vapor generation techniques have been widespread in that waters and effluents, metallurgical, clinical, biological, agricultural, geological, and environmental samples have all been analyzed at both the trace and ultratrace levels for these analytes. The reader is referred to the Further Reading section for an extensive compilation of specific applications.

Currently, detection power is primarily limited by reagent contamination. Progress in the widespread implementation of FI techniques, which feature online sample preparation and pretreatment capabilities as well as capabilities for rapid automation, should facilitate a further revolution in the use of vapor generation techniques in atomic spectroscopy.

*See also: Atomic Absorption Spectrometry: Principles and Instrumentation; Interferences and Background Correction; Flame; Electrothermal.*

### Further Reading

- Chen H, Brindle ID, and Zheng S (1992) Stibine generation combined with flow injection for the determination of antimony in metal samples by atomic emission spectrometry. *Analyst* 117: 1603–1608.
- Dedina J (1988) Evaluation of hydride generation an atomization for AAS. *Progress in Analytical Spectroscopy* 11: 251–360.
- D'Ulivo A, Loreti V, Onor M, Pitzalis E, and Zamboni R (2003) Chemical vapor generation atomic spectrometry using amineboranes and cyanotrihydroborate(III) reagents. *Analytical Chemistry* 75: 2591–2600.
- Guo M-M, Sturgeon RE, Mester Z, and Gardner G (2003) UV vapor generation for determination of selenium by heated quartz tube atomic absorption spectrometry. *Analytical Chemistry* 75: 2092–2099.

- Hatch WR and Ott WL (1968) Determination of submicrogram quantities of mercury by atomic absorption spectrometry. *Analytical Chemistry* 40: 2085–2087.
- Holak W (1969) Gas sampling technique for arsenic determination by atomic absorption spectrometry. *Analytical Chemistry* 41: 1712–1713.
- MacLaughlin RL and Brindle ID (2002) A new sample introduction system for atomic spectrometry combining vapor generation and nebulization capacities. *Journal of Analytical Atomic Spectrometry* 17: 1540–1548.
- Matousek T, Dedina J, and Selecka A (2002) Multiple microflame quartz tube atomizer – further development towards the ideal hydride atomizer for atomic absorption spectrometry. *Spectrochimica Acta* 57B: 451–462.
- Nakahara T (1983) Applications of hydride generation techniques in atomic absorption, atomic fluorescence and plasma atomic emission spectrometry. *Progress in Analytical Atomic Spectrometry* 6: 163–223.
- Sturgeon R and Mester Z (2002) Analytical applications of volatile metal derivatives. *Applied Spectroscopy* 56: 2002A–213A.
- Sturgeon RE, Willie SN, Sproule GI, Robinson PT, and Berman SS (1989) Sequestration of volatile element hydrides by platinum group elements for graphite furnace atomic absorption. *Spectrochimica Acta* 44B: 667–682.
- Welz B and Schubert-Jacobs M (1991) Evaluation of a flow injection system and optimization of parameters for hydride generation atomic absorption spectroscopy. *Atomic Spectroscopy* 12: 91–104.
- Welz B and Sperling M (1999) *Atomic Absorption Spectrometry*, 3rd edn. Weinheim: Wiley-VCH Publishers.
- Welz B, Melcher M, Sinemus HW, and Maier D (1984) Picotrace determination of mercury using the amalgamation technique. *Atomic Spectroscopy* 5: 37–42.

# ATOMIC EMISSION SPECTROMETRY

## Contents

### Principles and Instrumentation

### Interferences and Background Correction

### Flame Photometry

### Inductively Coupled Plasma

### Microwave-Induced Plasma

## Principles and Instrumentation

**R M Twyman**, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Atomic (or optical) emission spectrometry (AES, OES) is an important technique for the multielement analysis of a wide range of materials. Many elements have been discovered using emission spectrometry and it is the most commonly used procedure for the measurement of trace elements in rocks, water, soil, manufactured goods, and biological specimens. The technique is used to monitor the levels of different chemicals and trace elements in the environment and to determine the compositions of solids, liquids, and gases. In geoanalysis, emission spectrometry has been instrumental in the exploration of economic mineral deposits. In metallurgy and in the semiconductor industry, emission

spectrometry is of prominent importance in the production control of both raw materials and finished products. Finally, emission spectrometry allows the elements present in the sun and stars to be identified, helping us to understand better the nature of the universe. These are only a few examples of scientific and technical disciplines in which the technique of emission spectrometry has made a significant contribution.

## Theory and Signal Generation

### Atomic Spectra

AES involves the measurement of electromagnetic radiation emitted from atoms. Both qualitative and quantitative data can be obtained from this type of analysis. In the former case, the identity of different elements reflects the spectral wavelengths that are produced, while in the latter case, the intensity of the emitted radiation is related to the concentration of each element. Atomic spectra are derived from the transition of electrons from one discrete electron orbital in an atom to another. These spectra can be understood in terms of the Bohr atomic model.

In the Bohr model, the atom is depicted as a nucleus surrounded by discrete electron orbits, each associated with energy of the order  $h\nu$ . Every atom has a certain number of electron orbitals, and each electron orbital has a particular energy level. When all the electrons are present in the orbitals, the atoms are in the most stable form (the ground state). When energy (either thermal, resulting from collision, or radiational, resulting from the absorption of electromagnetic radiation) is applied to an atom and is sufficient to lift an electron from a shell with energy  $E_i$  to one with  $E_j$ , the atom is said to be in the excited state. The state of excitation is unstable and decays rapidly. The residence time of the unstable excited state is very short, in the order of  $10^{-8}$  s. When electrons return to the stable ground state, energy is emitted and that energy is equal to the difference in the energies between the ground and excited states. The energy is released in the form of electromagnetic radiation and defines the wavelength of the transition. The relationship between the energy and wavelength is described by the Planck equation:

$$E_j - E_i = h\nu = hc/\lambda$$

where  $E_j - E_i$  is the energy difference between the two levels (and  $E_j > E_i$ );  $h$  is Planck's constant,  $6.624 \times 10^{-34}$  J s $^{-1}$ ;  $\nu$  is the frequency of the radiation;  $c$  is the velocity of light in a vacuum,  $2.9979 \times 10^8$  m s $^{-1}$ ; and  $\lambda$  is the wavelength of the radiation in meters.

If enough energy is absorbed by the atom, electrons may escape completely, leaving the atom in the ionized state. The energy required for ionization is called the ionization potential. Ions also possess ground and excited states, through which they can absorb and emit energy by the same processes described for an atom.

Figure 1 illustrates the electron shell configuration in terms of energy levels. Horizontal lines represent

energy levels and vertical lines depict permissible transitions between them. The arrows show the direction of energy input or output (ascending arrows show the absorption of energy while descending arrows show energy radiation). When an electron in a quantum level  $j$  is captured by an ionized atom, energy is liberated according to the following equation:

$$h\nu = h\nu_j + mv^2/2 = E_j + mv^2/2$$

The wavelength for the emitted radiation due to a transition  $E_j - E_i$  is

$$\lambda = hc/E$$

where  $E$  is the energy difference and  $\lambda$  is the wavelength of the emitted radiation.

Spectra of neutral excited atoms are denoted as I, and correspond to those deexciting to the ground state (resonance lines) or close to the ground state (near-resonance lines). They are observed in low-energy sources such as flames. Spectra of singly ionized atoms are denoted as II, and they are observed in high-energy sources such as electrical sparks, inductively coupled plasmas (ICPs), and glow discharges. Every element has a characteristic emission spectrum, which is the basis of spectrochemical analysis.

## Molecular Spectra

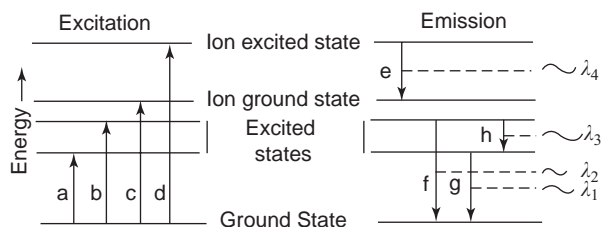
Molecular spectra consist of numerous densely grouped lines. These are called band spectra because they appear as luminous bands. The fine structure can only be observed with high-resolution instruments. Molecular spectra of excited molecules are related to the energy states of a molecule rotating around the principal axes of inertia. Band spectra in the near-infrared are produced by energy transitions related to oscillatory vibrations of individual molecules.

## Continuum

This is radiation distributed continuously over the wavelength range and can be attributed to recombination processes and other background factors. The intensity of the background increases with temperature.

## Instrument Design – Overview

A spectrometer consists of three main parts: (1) an emission source, which produces the spectrum; (2) an optical system, which scatters the spectrum; and (3) a device to measure the emitted lines. The two major types of instrument for the analysis of emission spectra are sequential and simultaneous spectrometers, although there are many variants of each in terms of mechanical and optical characteristics. The spectral



**Figure 1** Electron shell configurations in terms of energy levels. Arrows depict permissible transitions by absorption and excitation (ascending) or radiation and photon emission (descending), a and b represent excitation, c is ionization, d is ionization plus excitation, e is ion emission, and f, g, and h are atom emissions. (Reproduced with permission from Boss CB and Freedman KJ (1989) Concepts, Instrumentation and Techniques. *Inductively Coupled Plasma Atomic Emission Spectrometry*, Perkin Elmer Corp.)

wavelength range of interest is  $\sim 160\text{--}900\text{ nm}$ , but not all instruments are capable of covering this range, and the resolution may vary with wavelength. Oxygen and water vapor absorb short-wave UV emissions ( $< 190\text{ nm}$ ) and this obscures the emission lines of some common and important elements (e.g., hydrogen, carbon, oxygen, nitrogen, chlorine, phosphorus, and sulfur). Therefore, oxygen and water vapor must be eliminated from the instrument, either by evacuation or by flushing with nitrogen or argon. Nitrogen flushing does not interfere with the analysis of samples containing nitrogen compounds since molecular nitrogen ( $\text{N}_2$ ) does not obscure the emission lines from atomic nitrogen in the sample.

## Instrument Design – Emission Sources

In spectrochemical analysis, atomization and excitation can be achieved using various different emission sources. The spectra derived from low-energy sources such as flames are simpler than those from electrical discharges, although the temperature of flames and furnaces ( $2000\text{--}4000\text{ K}$ ) is inadequate to excite many of the elements. Nevertheless, flame emission spectrometry is widely used for the determination of the alkali elements (lithium, sodium, and potassium), whose excitation states are low enough to be populated at flame temperatures. Higher-energy sources produce higher temperatures and therefore more emission lines. In electrical discharges, arcs and sparks are created by applying currents and potentials across conducting electrodes, and a large quantity of the sample surface is evaporated in this process. Better quantitative analysis is achieved using plasma sources: ICP, direct current plasma (DCP), and microwave induced plasma (MIP), which generally achieve temperatures of  $7000\text{--}8000\text{ K}$ . Glow discharge sources, which use high-energy argon atoms and ions to excite atoms ejected from the analyte surface, are often used for the analysis of metals.

The degree of excitation by a thermal source can be described by the Boltzmann distribution equation. If  $N_1$  is the number of atoms in the excited state and  $N_0$  is the number in the ground state, then the excited fraction is given by

$$N_1/N_0 = (g_1/g_0) \exp(-E/kT)$$

where  $E$  is the energy difference between the ground and excited states,  $T$  is the absolute temperature (K),  $k$  is the Boltzmann constant ( $1.38 \times 10^{-23} \text{ J K}^{-1}$ ), and  $g_1$  and  $g_0$  are quantum statistical weighting factors.

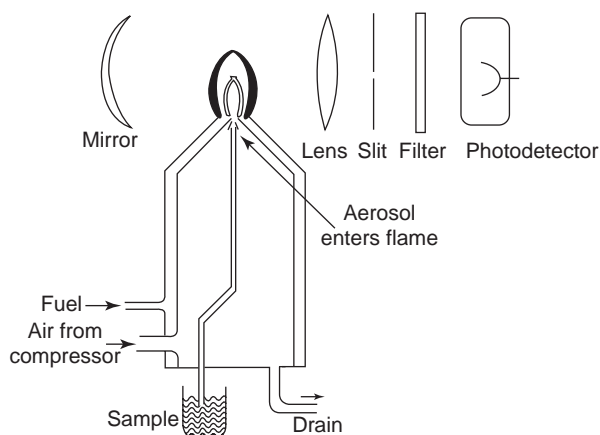
## Flame Sources

In flame emission spectrometry, the sample solution is sprayed or aspirated into a flame as a fine mist or aerosol. The sample is vaporized in the flame, and atomized by a combination of heat and the action of a reducing gas. The atoms are excited into higher electronic states by the heat, and as they revert to the ground state they emit photons, which are measured by the detector. The layout of a flame photometer is shown in Figure 2.

## Discharge Sources (Arcs and Sparks)

The first discharge sources produced direct current (DC) arcs by electrically heating the sample in an electrode cup and vaporizing the analytes into a low-voltage, high-current discharge. The temperature of the arc plasma varies from  $4000$  to  $5000\text{ K}$ . The limits of detection are good and the entire sample can be consumed. However, due to variations in the volatilization process, the accuracy is usually poor. Reproducibility can be improved by using internal references and optimizing the conditions of vaporization (graphite electrode designs, addition of modifiers). The spectral range of the source is limited due to the presence of cyanogen bands with heads at  $421.6$ ,  $388.3$ , and  $359.0\text{ nm}$  when graphite is used. Argon can be injected into the arc to minimize these effects. In the so-called cathode region, intensity is relatively higher ( $10\text{--}50$  times) than in the central and anode regions of the arc.

Spark sources produce lower average temperature than arcs, but the local temperature can be as high as  $40\,000\text{ K}$ . Like arc sources, sparks produce atomic lines, but also more pronounced lines for ions, which are known as spark lines. The emission source consists of a sparking stand and a spark generator. A spark forms between the cathode and the sample (which acts as the anode). The adjustable gap



**Figure 2** Schematic of a flame emission spectrophotometer.



between the cathode and the sample is sometimes termed the entrode, and is filled with argon. The generator initially produces a brief low-energy discharge that ionizes the argon and creates the plasma. A second, high-energy discharge then causes the sample to vaporize at the sparking point, exciting the atoms and generating the emission spectrum. A large number of sparks is generated, each lasting only a few microseconds. Due to the lower sample consumption, the limits of detections are poorer than in the arc. The high-voltage alternating current (AC) arc is intermediate in analytical performance between the DC arc and the high-voltage spark. These sources are used mainly in the metallurgical industry.

### Direct Current Plasmas

The DCP evolved from DC arcs and can be classified into two types: discharges confined within a chamber (wall-stabilized) and unconfined plasmas. Various designs and configurations of the electrode exist for injecting the sample aerosol-carried gas into the plume, one of which is shown in **Figure 3**. Magnetic fields can be used to enhance the coupling of the sample into the plasma, as is useful for the analysis of very complex materials. It is characterized by ease of operation, robustness, and optimization for a great variety of complex matrices. Analyte signals are observed in the tail flame or close to the region where the sample is injected into the discharge and where the density of excited species is greatest. The discharge has been shown to be suitable for the analysis of slurries and solutions containing very high salt concentrations.

### Microwave-Induced Plasmas

MIPs have been used widely in gas chromatography (GC) detection. Low-power MIPs (50–150 W) do not accept liquid aerosols efficiently. One way of overcoming this disadvantage is to desolvate the sample solution or to employ electrothermal vaporization. A commercial system is available consisting of a

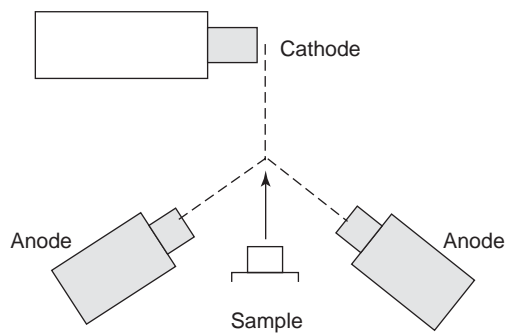
water-cooled discharge tube inserted in a reentrant-type cavity that does not require tuning. It is generally considered that a laminar flow torch is more suitable than tangential flow for GC detection. In its usual form, the MIP operates with a low helium flow rate ( $1 \text{ l min}^{-1}$ ). MIPs are very suitable for the determination of halides and other nonmetals. A microwave plasma torch (MPT) has been developed consisting of concentric copper tubes. The carrier gas and aerosol enter the inner tube while the outer tube serves as the microwave cavity. The plasma is formed at the top of the MPT and extends out like a flame, but with a central channel for the introduction of the aerosol. The MPT is superior to the conventional MIP since it allows the introduction of wet aerosols at lower power.

### Inductively Coupled Plasmas

ICP sources have brought about a revolution in multi-element analysis. ICPs are generated from radiofrequency (RF) magnetic fields induced by a water- or air-cooled copper coil looped around a quartz tube. The RF magnetic field oscillates at 27.12 or 40.68 MHz, at incident powers ranging from 0.5 to 2.5 kW. Higher powers are usually applied when organic solvents are aspirated. Argon gas flows through a torch, which consists of three concentric tubes usually constructed from fused silica. The plasma is initiated by seeding the argon stream with electrons provided from a Tesla coil. The electrons, detached from the argon atoms, collide with further argon atoms and populate the coil region with positive and negative charges. Because of the magnetic field, the particles flow in a closed annular path. Due to the conductance of the gases in the coil region, the charged particles are heated by inductive coupling to a temperature equaling the ionization temperature of the support gases  $\sim 7000\text{--}8000 \text{ K}$  in the case of argon. A chain reaction of collisional ionization occurs, resulting in the formation of the ICP. In practice, the plasma impedance is monitored along with the tube grid current, grid voltage, plate current, and voltage. These data are fed back to a loop to control the plasma power. The configuration of an ICP-AES system is shown in **Figure 4**.

### Glow Discharge Sources

Glow discharge is based on a phenomenon called sputtering, where atoms ejected from the surface of the analyte by high-energy argon atoms and ions achieve the excited state in the resulting plasma. A copper tube filled with argon is juxtaposed with the sample and a potential difference applied across the gap, with either a direct current or a RF alternating current. Electrons jump from the negatively charged



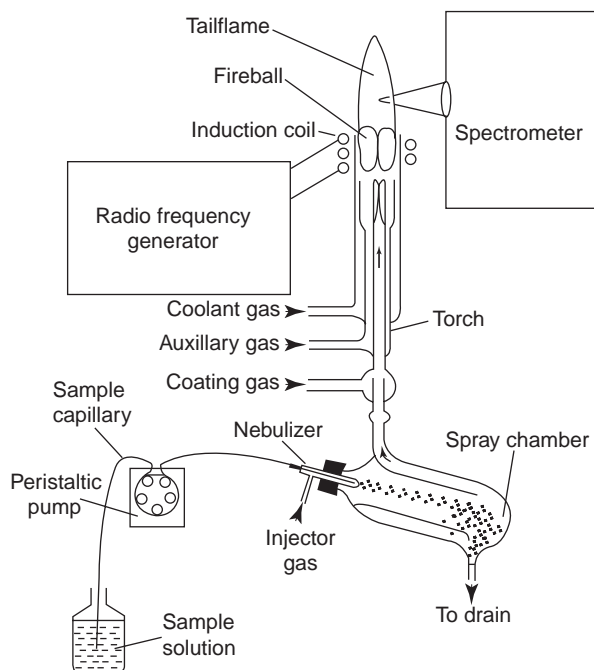
**Figure 3** A schematic design of a direct current plasma source.



sample toward the positively charged copper electrode and collide with argon atoms, creating positively charged argon ions that are attracted toward the sample surface. En route, they collide with other

argon atoms and ions, and then strike the analyte surface with sufficient energy to displace electrons and atoms from the sample (sputtering). These analyte atoms also collide with the electrons and high-energy argon atoms/ions, causing them to be excited to higher energy states. As they deexcite, they emit photons resulting in a 'glow' extending 2–3 mm from the sample (Figure 5).

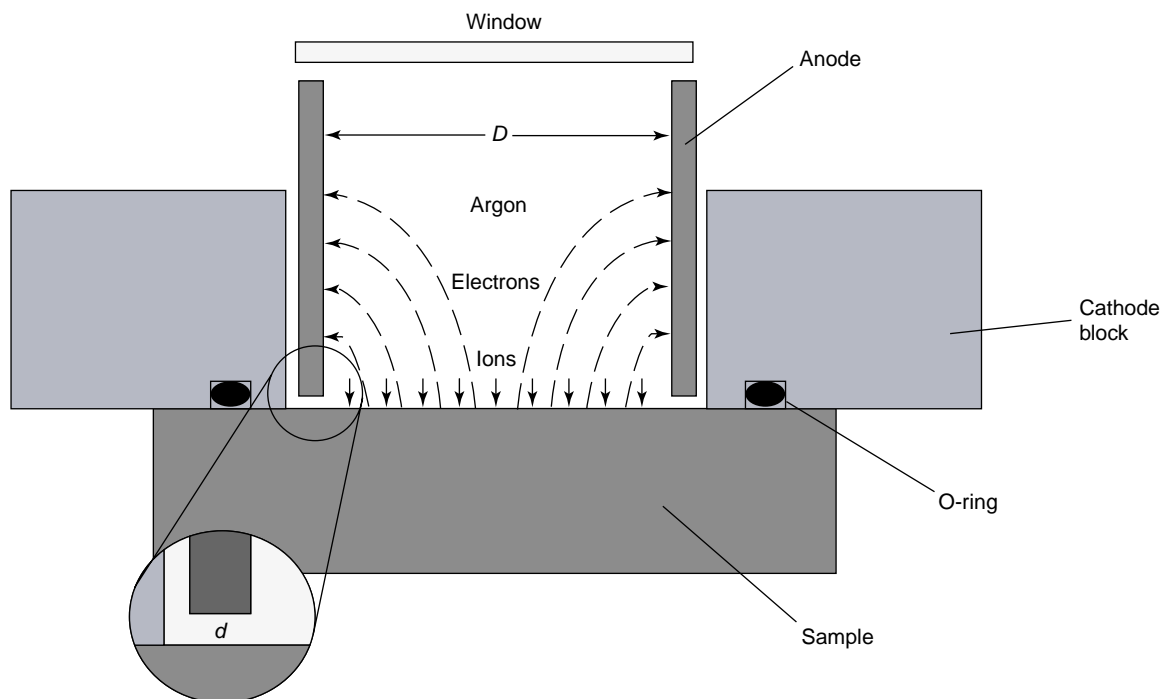
Glow discharge sources are based on three principal designs. The Grimm source, which consists of a copper cathode block in direct contact with the metal sample, is used with DC voltage. The Renault source is based on the Grimm source, but utilizes a ceramic cathode block that allows the use of RF voltages. The most recent development is the Marcus source, which also operates in RF mode. It has a ceramic cathode block and a very short anode tube to facilitate rapid plasma expansion. Although DC and RF plasmas are similar, RF plasmas are more stable and show a greater sputtering depth. The most important difference is that RF glow discharges can be used to analyze both conducting and nonconducting analytes, while DC glow discharges are restricted to conductors.



**Figure 4** Schematic of an inductively coupled plasma source for atomic emission spectrometry.

## Instrument Design – Sample Introduction

Some of the emission sources discussed above are designed for use with solid samples, which can be



**Figure 5** Principle of glow discharge atomic emission spectrometry.  $D$  = diameter of the anode and  $d$  = distance from anode to sample.

attached to or used as electrodes. For plasma sources, solid samples can be converted into slurries if the particle size is small, but an alternative is to use spark ablation or laser ablation to explode the sample, generating a small amount of vapor that is carried into the plasma in a gas stream. Gaseous samples can be analyzed directly, but liquid samples are first converted into aerosols, so that they can desolvate and atomize completely. The most common sample delivery system consists of a peristaltic pump and capillary tube to deliver a constant flow of analyte liquid into a nebulizer. The nebulizer turns the analyte liquid into fine droplets, which are carried by gas into the plasma. Larger droplets ( $> 5 \mu\text{m}$ ) are captured in the spray chamber, and are drained from the instrument.

### Analysis of Solids in Plasma Sources

**Spark ablation** In this method, conducting samples are vaporized with an electrical discharge, while nonconducting samples are first modified by mixing with copper or carbon powder. The dry aerosol is carried by an argon stream into the plasma. The system is calibrated with samples of similar physical and chemical composition. Slight differences can be compensated through the use of internal standards. Large particles can be eliminated using traps, as long as internal references are available.

**Laser ablation** This method is used as a micro-chemical sampling procedure for localized determinations. A pulsed neodymium–yttrium aluminum garnet (Nd–YAG) laser is used to ablate material from solid samples. Repetitive laser pulses and sample translation can be used to improve the precision and accuracy of the analysis. Refractory materials and geological samples can be analyzed for trace and major elements. Powdered samples can be pelleted under high pressure for bulk analysis.

### Nebulizers

Many different types of nebulizer are available, some of which are suitable for general analytical purposes while others have more specialized uses. Pneumatic nebulizers are general-purpose devices in which the aerosol is formed by the shattering effect of a high-velocity jet of gas. The most commonly used pneumatic nebulizer is the Meinhard glass concentric nebulizer, in which the sample is introduced along a narrow capillary tube located within a larger glass or quartz tube. The outer tube contains argon gas, which flows to a 10–20 mm gap surrounding the sample capillary. The aerosol is formed by the

venturi effect produced by the gas stream as it is forced through the annulus, which fragments the liquid into fine droplets. The Meinhard nebulizer is sensitive to clogging, so particulates and solutions with high salt concentrations should be avoided.

Cross-flow nebulizers consist of a capillary tube that directs a stream of argon gas at  $90^\circ$  to the sample delivery tube, creating an aerosol due to its shearing effect over the sample tube. Again, these nebulizers are designed for general-purpose use, although they are less prone to blockage where the sample has a high salt concentration. Parallel flow nebulizers, such as the Burgener nebulizer, consist of parallel sample and gas capillaries with adjacent exits, so that the liquid sample is drawn into the gas stream. These are designed specifically to deal with inert samples with a high concentration of dissolved solids. In the Babington nebulizer, the liquid sample flows over a smooth surface containing a small orifice, through which argon flows at a high velocity, shearing the liquid into tiny droplets. As above, this device is not susceptible to clogging and can handle viscous solutions and suspensions. Another nebulizer designed for these difficult samples is the V-groove nebulizer, a special type of Babington nebulizer in which the liquid flows down a vertical V-shaped groove toward the gas orifice, from which argon flows at a pressure of 210–1050 kPa. A further Babington-type nebulizer, the Frit nebulizer, produces very fine aerosols but has very long washout times.

Ultrasonic nebulizers offer enhanced sensitivity in detection limits by using a vibrating piezoelectric transducer to set up standing waves in the liquid to produce uniformly sized droplets. The efficiency of nebulization is so high that the solvent loading of the aerosol needs to be reduced by thermal desolvation, otherwise cooling of the plasma takes place. Direct injection nebulizers (DINs) are effective devices for introducing liquids directly into the plasma when flows are slower than  $0.1 \text{ ml min}^{-1}$ . By avoiding the use of a spray chamber, DINs allow 100% transport into the plasma. However, solvent loading in the plasma can exceed the optimum level and can cause a reduction in the plasma temperature, resulting in lower intensities.

### Instrument Design – Optical Systems

Once the sample has been introduced into the emission source, atomized, and excited, the emitted photons are diffracted by an optical system consisting of slits, mirrors, and gratings, which focus the spectral lines onto a detector. This section discusses the types of gratings and spectrophotometer designs that can be used in AES.

## Gratings

Gratings are reflective surfaces containing parallel, equally spaced lines. Their resolving power is proportional to the number of lines, which in turn depends on the line spacing. For radiation arriving with an angle of incidence  $\alpha$ , the angle  $\beta$  at which it will be diffracted by a grating of  $N$  lines per millimeter depends on the wavelength  $\lambda$  of the radiation, and is defined by the equation

$$\sin \alpha + \sin \beta = kN\lambda$$

In this equation,  $k$  is the order of diffraction, meaning that constructive interference occurs at wavelength  $\lambda$  (first order),  $\lambda/2$  (second order),  $\lambda/3$  (third order), and so on. The spectrum produced by a grating thus consists of several superimposed orders, with the first and second orders carrying most of the energy. The blaze angle of a diffraction grating is the mirroring angle of each line. The blaze angle is not arbitrarily selected; radiation can be concentrated into the first order, rather than in multiple orders and the performance of the grating for a particular spectral range can be optimized by varying the blaze angle. To cover the entire analytical spectral range, two gratings may be employed, or the spectra can be observed in several orders.

Ruled gratings have been replaced by holographic gratings produced from the interference patterns generated by two synchronous lasers. The interference pattern is etched into a photosensitive film, and replicas of these gratings can be manufactured readily. The grating pattern created in this way is greatly superior to those produced mechanically, being much more accurate, having higher linearity, and being free from imperfections and distortions that can give rise to ghosts and stray light. However, the brightness achieved by the conventional holographic gratings is not as high as that of ruled gratings because of the sinusoidal groove profile.

The echelle grating is designed to operate in multiple orders to produce high-resolution spectra. In contrast to normal gratings, echelles are ruled at 30–300 grooves per millimeter by ion bombardment, and are frequently in the range of 30–120 orders. The grating consists of saw-like grooves varying in height and 6–13  $\mu\text{m}$  in depth. The echelle grating is designed to separate one order of diffraction from another. This is achieved by inserting a quartz prism, which acts as an order sorter, in front of the grating. Successive spectra are diffracted to different degrees, so that the various orders appear as a two-dimensional array of wavelengths. These gratings are installed in polychromators, sequential systems, and in charge-injected and charge-transfer detector instruments (see below).

The resolving power of a grating is a measure of its ability to separate to the baseline two adjacent spectral wavelengths  $\lambda_1$  and  $\lambda_2$ . The resolution  $R$  is expressed as

$$R = \lambda/(\lambda_1 - \lambda_2)$$

## Monochromators

A monochromator measures a single wavelength, but can be scanned through a wide wavelength range. Polychromatic light passes through an entrance slit and is dispersed by diffraction gratings. These instruments are used in sequential mode, element determinations being performed one after the other. Two designs are usually employed in commercial instrumentation: Czerny–Turner and echelle.

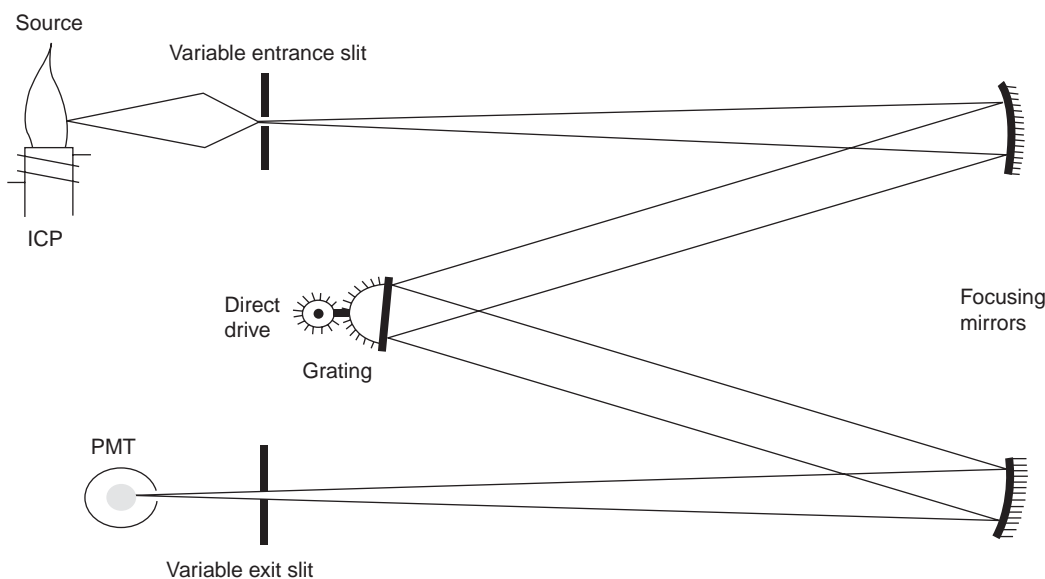
**Czerny–Turner monochromators** Two mirrors are used to reflect and focus the polychromatic and diffracted beams. Wavelengths are selected by using a computer to rotate the grating in various ways. This design is shown in Figure 6. As the grating rotates, a different wavelength is focused onto the exit slit.

**Echelle monochromators** These are high-resolution instruments that readily achieve resolutions of 5 pm in contrast to the 10–20 pm that is normal for conventional sequential instruments. Because the spectra are recorded one above the other, such instruments can be very compact.

## Simultaneous Spectrometers

There are several types of simultaneous instruments, which carry out elemental determinations in parallel rather than in series. They differ in optical design and the type of detector that is used.

**Polychromators** These instruments can measure numerous spectral lines simultaneously, and several approaches can be used to achieve this. In the classical Paschen–Runge design, light is directed onto a diffraction grating that diffracts the polychromatic radiation into its individual wavelengths. Exit slits are located at predetermined positions on a Rowland circle to focus the diffracted wavelengths onto a specific photomultiplier tube (PMT). In this design, the number of elements that can be determined simultaneously is limited by the geometric configuration of the exit optic area of the instrument. This is a disadvantage compared with the state-of-the-art charge-coupled and charge-injection devices discussed below. Geometric alignment is also controlled by the space requirements of the PMTs, and as a



**Figure 6** Czerny–Turner monochromator. ICP = inductively coupled plasma and PMT = photomultiplier tube. Reproduced from [www.thespectroscopynet.com](http://www.thespectroscopynet.com)

result a series of mirrors is employed to direct the spectral radiation to the measuring surface of the detector. For the determination of the spectral lines in the low part of the UV spectrum ( $<190$  nm), as is necessary in the detection of aluminum and phosphorus, the spectrometer must be contained in a vacuum or purged of oxygen using nitrogen or argon gas.

Polychromators have several advantages, including high sample throughput, lower running costs, and the ability to measure more than 20 elements, in duplicate, with background correction in less than 5 min using only 5 ml of solution. However, disadvantages include the high costs associated with the individual electronic systems required for each spectral line measurement, and the inflexibility of the instrument due to the static optical system.

## Instrument Design – Detection

### Photomultiplier Tubes

Photons emerging from the exit slits of the spectrophotometer are detected by one of two types of device, a PMT or a solid-state component. Photomultipliers are often used as detectors in AES. Incident photons emerging from the exit slit fall on the photocathode, liberating electrons, and the current is amplified by a set of dynodes. The final anode current is proportional to the incident photon signal received by the photocathode. The measurement dynamic range is very broad, i.e.,  $10^{15}$ , and sensitivity is high. These detectors allow the detection of low intensities emitted by trace elements, as well as strong signals from major elements. They have very fast response times,

typically 1–2 ns for a 10–90% change in signal. The main inconvenience of photomultipliers is their cost.

There are several types of photomultipliers, which differ in the nature of the entrance window, either crystal or fluoride, and in the nature of the sensitive layer on the photocathode. Some are only sensitive in the far-UV while others are more sensitive in the visible. The type of photomultiplier to be used is selected according to the wavelength of the line to be detected. A fatigue lamp (a small incandescent light source) is often used with photomultipliers to keep the temperature of the tube and its associated electronics constant. The fatigue lamp is switched on when the emission source is off and vice versa.

**Charge-injection devices (CIDs)** CIDs can be used in combination with an echelle spectrometer to produce a flexible detection system for multielement analysis using a direct current arc, AC spark, or inductively coupled or direct current argon plasmas. The CID consists of a two-dimensional array of detector elements and when it is coupled to a polychromatic dispersive system, simultaneous multielement analysis is provided over the spectral range 170–800 nm. In addition to the multielement capability and large dynamic range (eight orders of magnitude), the system allows random access integration, where each detector element can be nondestructively processed until an appropriate signal-to-noise ratio is attained. Background can be read simultaneously, and alternate spectral lines can be selected for the analysis of spectrally complex materials. The detection limits are similar to those obtained by conventional

detection systems. The system is ideally suited to semiquantitative screening analysis.

**Charge coupled devices (CCDs) and echelle polychromators** CCDs and echelle polychromators are now available, and in such systems the high-energy echelle spectrometer utilizes two detector focal planes and two cross-dispersers. At 200 nm, the resolution is 0.007 nm but this degrades substantially in the visible region. The cross-disperser for the UV region (167–375 nm) is a grating (374 lines per millimeter) with a Schmidt correction incorporated into its surface. Aberration is also corrected for a 400 mm radius focal plane. The cross-disperser for the visible region is a fused 60° quartz prism. A segmented array CCD consists of 224 addressable subarrays with over 6000 pixels on a 13 × 18 mm silicon substrate. One of the drawbacks of the CCD is the inability to read nondestructively. When the charge is read on a subarray, it is destroyed and cannot be monitored to adjust the integration time.

**Photodiode arrays (PDAs)** A spectrally segmented PDA spectrophotometer has also been developed. Radiation from the plasma is predispersed and transmitted through an optical mask prior to dispersion by an echelle grating onto a PDA. Limits of detection are similar to those obtained by conventional detection systems, except at low wavelengths where degradation has been observed.

*See also: Atomic Emission Spectrometry: Interferences and Background Correction; Flame Photometry; Inductively Coupled Plasma; Microwave-Induced Plasma.*

## Further Reading

Beauchemin D, Le Blanc JCY, Peters GR, and Craig JM (1992) Plasma emission spectrometry. *Analytical Chemistry Reviews* 64: 442R–467R.

- Bings NH, Bogaerts A, and Broekaert JAC (2002) Atomic spectroscopy. *Analytical Chemistry* 74: 2691–2711.
- Bogaerts A and Gijbels R (1998) Fundamental aspects and applications of glow discharge spectrometric techniques. *Spectrochimica Acta B* 53: 1–42.
- Cave MR, Butler O, Chenery SRN, *et al.* (2001) Atomic spectrometry update. Environmental Analysis. *Journal of Analytical Atomic Spectrometry* 16: 194–235.
- Evans EH, Day JA, Price WJ, *et al.* (2002) Atomic spectrometry update. Advances in atomic emission, absorption and fluorescence spectrometry, and related techniques. *Journal of Analytical Atomic Spectrometry* 17: 622–651.
- Evans EH, Day JA, Price WJ, *et al.* (2003) Atomic spectrometry update. Advances in atomic emission, absorption and fluorescence spectrometry, and related techniques. *Journal of Analytical Atomic Spectrometry* 18: 808–833.
- Evans EH, Fisher A, and Hill S (1998) *An Introduction to Analytical Atomic Spectrometry*. New York: Wiley.
- Fisher A, Hinds MW, Nelms SN, Penny DM, and Goodall P (2002) Atomic spectrometry update. Industrial analysis: Metals, chemicals and advanced materials. *Journal of Analytical Atomic Spectrometry* 17: 1624–1649.
- Hill SJ (1999) *Inductively Coupled Plasma Spectrometry and its Applications*. Boca Raton: CRC Press.
- Marcus RK and Broekaert JAC (2003) *Glow Discharge Plasmas in Analytical Spectroscopy*. New York: Wiley.
- Montaser A and Golightly DW (1992) *Inductively Coupled Plasmas in Analytical Atomic Spectrometry*. Weinheim: VCH.
- Payling R (1998) Glow discharge optical emission spectrometry. *Spectroscopy* 133: 36–48.
- Payling R and Larkins P (2000) *Optical Emission Lines of the Elements*. New York: Wiley.
- Taylor A, Branch S, Halls D, Patriarca M, and White M (2003) Atomic spectrometry update. Clinical and biological materials, foods and beverages. *Journal of Analytical Atomic Spectrometry* 18: 385–427.
- Thomsen VBE (1996) *Modern Spectrochemical Analysis of Metals: An Introduction for Users of Arc/Spark Instrumentation*. Ohio: ASM International.

## Interferences and Background Correction

**R M Twyman**, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

When a spectral line produced by one element is very close to a line produced by another element, interference between the two lines may take place. If this

happens, the measured intensity is a combination of the intensities of the two emission lines and does not give a true indication of the abundance of either element. Spectral line interference in atomic emission spectrometry (AES) can severely affect the accuracy of trace and minor element determinations. Therefore, measurement of the analytical line of interest must take into account the possibility of interference from adjacent spectral lines and the interference

should be corrected to obtain the true intensity of the element being measured. The level of interference is dependent upon the composition of the sample, the gases used in the instrument, and the resolution of the spectrometer. Carry-over, contaminated reagents, and matrix interference effects may also occur. Interference becomes worse as the concentration of the interfering element in the sample increases and as the resolution of the instrument decreases. There are several different forms of interference, including the background continuum, stray light, wing overlap, molecular bands, and direct overlap. Table 1 shows examples of interference types commonly observed when using inductively coupled plasma (ICP) sources.

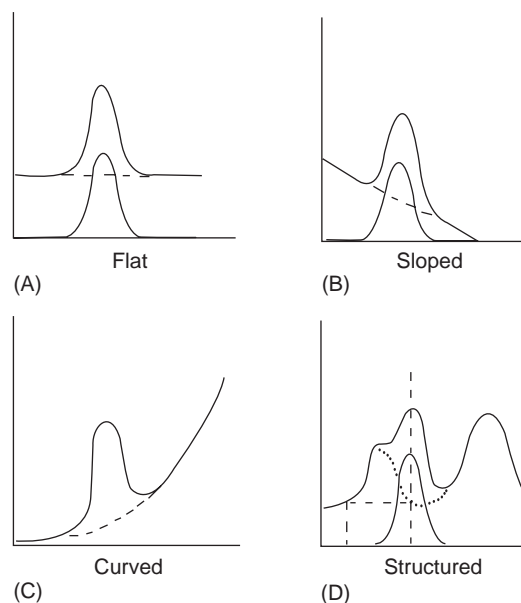
## Types of Spectral Interference

### Background Continuum

The background continuum is radiation distributed continuously over the wavelength range of the spectrometer. This type of interference originates from several sources, including recombination, Brehmsstrahlung, line-broadening, stray light within the spectrometer, and the presence of molecular species. Its intensity increases with temperature. Several typical forms of background are shown in Figure 1.

A common source of background is the support gas plasma. The effect is minimal with air and nitrogen but very intense when argon is used. About 200 argon lines have been observed predominantly between 350 and 450 nm (there appear to be none in the ultraviolet (UV) range). Very intense and broad argon lines are produced in the near-infrared range

when ICP sources are used, limiting the applicability of this technique in the detection of halides and other nonmetals. Where the analyte is dissolved in an organic solvent, further interference occurs in the form of intense carbon lines.



**Figure 1** Types of background emission in atomic emission spectrometry. (A) Flat background, which can be addressed using one-point correction. (B) Linear sloping background, which can be addressed using three-point correction. (C) Curved background, still with continual variation. This is best addressed using multivariate techniques. (D) Complex, structured background, comprising interferences from line-rich elements. May be overcome with higher resolution analysis although multivariate statistical techniques such as Kalman filtering can also be applied (see main text).

**Table 1** Examples of interference types in analytical ICP-AES

Type of interference	Interferent	Analyte	Application
Spectral overlap	Ca 180.74 Mn 180.747	S I 180.73 S I 180.73	Sulfate in water
Spectral overlap	Ti 228.618	Co II 228.616	Co in Ti alloys and rocks
Spectral overlap	Cu 213.853	Zn 213.856	Zn in Cu alloys
Spectral overlap	Fe II 234.8	Be 234.861	Be in geological materials
Spectral overlap	Be II 313.042	Ti II 313.08	
Resonance broadening	Ca 393.3 Ca 396.8	Al 396.1	Al in drinking water and serum
Far scatter	Ca 393.37	Zn I 213.68	Zn in drinking water
OH bands	OH 310.327 OH 310.236 OH 310.214 OH 310.123	V II 310.23	Trace determination of V
OH bands	OH bands	Be II 313	
Cross-talk	Mg 279.5	Cr 267	
Broadening	Wings Al 214.54	Al 220.467 Cd II 214.438	Pb II 220.353 in geological samples
Carbon lines	C 193.09	As 193.7	
C2 Mullikan system	232.5	Ni 231.604	

Recombination (the capture of electrons by ions) is another source of background, and while the continuum is present throughout the optical range of the spectrometer, it increases with decreasing wavelength and is most troublesome at  $\sim 200$  nm. The light emitted by aluminum in the range 190–220 nm is due to recombination, and this interferes with numerous of other elements in the same region of the spectrum. In ICP-AES, background enhancements for calcium, magnesium, and aluminum can seriously affect analytical performance, and in the case of aluminum this is due to the broadening of the most intense lines.

As the sensitivity of detection approaches the limit of the spectrometer, noise factors in the instrument should also be taken into account. Sources of noise influencing the signal-to-noise ratio in ICP-AES have been published extensively, and include fluctuations in the nebulizer-spray chamber system, plasma flicker, and pulsation. In high-voltage spark dischargers, the light emitted during the first 5–10  $\mu$ s consists primarily of continuum and band spectra that fade, while the emissions of analyte spark lines persist for  $\sim 50$   $\mu$ s. Therefore, the time resolution can be used to distinguish genuine signals from background.

### Stray Light

Stray light is radiation inside the spectrometer that has a wavelength outside the spectral band pass. It arises mostly due to imperfections in the optical design of the instrument, and its impact has been reduced as instruments have evolved and become more sophisticated. The magnitude of the effect is sample-dependent, and is particularly noticeable when solutions containing high concentrations of calcium and magnesium are analyzed. There is evidence that stray light is derived from the strong Ca II 393.366 and 396.847 nm lines, and that the limits of determination of aluminum in calcium-rich solutions are degraded unless holographic gratings are used.

Cross-talk is the interception of unwanted light in the region of the slit and photomultiplier tube (PMT). This occurs when the radiation from a strong emitter falls upon the slits, mirrors, and detectors of adjacent analytes. For example, if the detector system for the Cr 267.716 nm line is located next to the very intense Mg 279.553 nm line in the focal plane, cross-talk can occur if detection optics have not been adequately masked.

### Line-Broadening and Wing Interference

The broadening of spectral lines can be due to Doppler effects, foreign gas, or resonance effects. Doppler broadening is due to the random motion of atoms. The atoms move in different directions and

at slightly different speeds, so individual atoms emit light at slightly different wavelengths depending on whether they are moving toward or away from the detector. Doppler widths can be in the order of 0.001–0.01 nm, depending on the element. Doppler broadening produces a Gaussian profile, and the wing intensities decrease rapidly, therefore contributing nothing to the background more than 0.01 nm away from the broadened spectral line.

Foreign gas broadening is due to collisions between unlike atoms, and includes a related phenomenon called Stark broadening which is caused by collisions between excited atoms and charged particles, i.e., electrons and ions. Collisions between atoms and other particles cause the energy levels in the atoms to change, and thus the energy required to excite the atoms also changes by a small amount. At 5000–6000 K, the effect of foreign gas broadening is even smaller than that of Doppler broadening, and usually can be ignored.

Resonance broadening is caused by collisions between like atoms, and this has a significant impact on wing intensities because its effect is to broaden the wings and suppress the central emission line. The effect occurs because, at higher pressures where atoms are closer together, photons emitted at the peak frequency are likely to be absorbed by surrounding atoms, while those at the wings are less likely to be emitted and trapped by surrounding atoms. As an example, resonance broadening of the Ca I 393.3 nm and Ca II 396.8 nm lines results in interference with the 396.1 and 394.4 nm lines of aluminum, such that the limits of detection for aluminum are degraded in the presence of high concentrations of calcium.

### Molecular Bands

Molecular bands occur due to the energy states of a molecule rotating around the principal axes of inertia and the energy transitions related to oscillatory vibrations of individual molecules. Examples of molecular bands interfering with atomic lines include the hydroxyl radical and organic compounds. Hydroxyl radicals are produced both in dry plasmas (from the entrainment of hydrogen and oxygen) and in wet plasmas (from the dissociation of water molecules). Rotational spectra are produced in the 281–295 and 306–325 nm ranges.

Significant carbon-based molecular bands have been identified when samples containing high concentrations of carbon are excited. The C2 Swan bands are located at 593–620, 527–547, and 467–474 nm. The Mullikan C2 system occurs at 232.5 nm and degrades the limits of detection of lines such as Ni 231.604 nm. Bands due to cyanogen (CN) are

predominant in direct current (DC) arcs where nitrogen is entrained in the arc column. These molecular bands can be reduced by adding oxygen to the plasma (ICP sources) or argon in the case of DC arcs. As a result, the limits of detection are improved, especially for the rare earth elements in DC arc sources.

### Direct Spectral Overlap

Elements such as tungsten, zirconium, uranium, and the rare earth elements have multiple spectral lines, which make line selection a difficult task. The degree of interference and sample composition is related to what is called the critical concentration ratio (CCR), which is defined as the ratio of the concentration of interferent  $i$  to that of the analyte  $a$  at which the ratio of the line intensities  $I_i/I_a$  is equal to unity. If the measured concentration ratio exceeds the CCR, the intensity of the interferent line will be higher than that of the analyte line and will be detrimental to accuracy. In some spectrometers, optical cross-talk in the region of the exit slit and detector will present itself as a direct overlap.

## Techniques to Correct for Spectral Interferences

### Line Selection

In classical emission spectrometry, there are numerous listings of spectral lines. The Massachusetts Institute of Technology (MIT) Wavelength Tables list 110 000 spectral lines, and the intensities quoted are those observed in DC arcs or alternating current (AC) sparks. The National Institute of Standards and Technology (NIST) Atomic Spectra Database contains ~72 000 spectral lines with intensities observed in a copper arc. The database is accessible over the Internet ([http://physics.nist.gov/cgi-bin/AtData/main\\_asd](http://physics.nist.gov/cgi-bin/AtData/main_asd)) and can be searched using either lines or energy levels. The values of the upper and lower energy transitions are listed. Other lists are available for spectral lines located in the vacuum UV.

Although the relative intensities of spectral lines in the ICP differ from those observed in the DC arc and AC spark, the published tables are invaluable for the selection of analyte lines in ICP sources, and the identification of spectral interferences in the spectrometer bandwidths. However, spectral lines are emitted by ICP sources that are not emitted by DC arcs and sparks. In order to facilitate spectral line selection in ICP-AES, numerous spectral line atlases are now available which list the best analytical lines and the potential interferences due to coincidences from major and minor constituents. Simulated

interference allows spectral lines to be selected for various materials.

### High-Resolution Spectrometers

Direct line interferences can only be avoided by selecting alternative lines, or performing a line interference correction. The interference from adjacent lines can be minimized by using high-resolution spectrometers, decreasing the width of the slits or using higher spectral orders. A high-resolution spectrometer is a definite advantage for the analysis of spectrally complex materials. However, there is a practical limit to the resolution that can be achieved. This is determined by the natural line widths and the extent of broadening. Furthermore, even with a high-resolution monochromator, there may not be sensitive lines free from spectral interferences that can be used for trace element determinations in complex materials.

### Background Correction

Background correction is used to compensate for continuum interferences. In the simplest cases (where the background is static across the spectral wavelengths of interest, as shown in **Figure 1A**), this can be performed by measuring the background at a single point and subtracting the intensity from the gross signal to derive the analyte concentration. This one-point background correction method also facilitates the measurement of the background when there are sample-to-sample variations in background levels. The nature of the background can be determined rapidly by scanning a representative sample, reference materials, or synthetically prepared solutions and solids in the region of the analyte line.

Background effects are rarely as simple as discussed above. In **Figure 1B**, the background is not constant, but declines with increasing wavelength. Such a situation could occur when the analyte line is close to a broadened interferent. In this case, the use of a single point would produce an erroneous result and two points at approximately equal distances from the center of the profile should be used to calculate the background. Since one signal is measured at the analyte line and two signals are measured in the adjacent wavelengths to determine the background, this is called three-point background correction. The average of these measurements, or their weighted average if unequal distances are used, is subtracted from the peak intensity.

The case shown in **Figure 1C** is a variant of **Figure 1B** where the background shows a curved rather than sloping profile, and the procedure described above would produce only an approximate correction.



However, the collection of multiple readings across a spectral window centered on the analyte wavelength allows correlation or digital filtering methods to be used. These transform the data and fit the curve to a mathematical function, which can be applied to the unimproved spectrograph and used to eliminate the smoothly varying background. Heuristic and statistical algorithms have been used in this manner, as well as convolution methods, e.g., square wave convolution using a zero area symmetrical one-period square wave.

In Figure 1D, background compensation in the vicinity of the analyte line is difficult because the background is structured. This type of interference can only be overcome with a high-resolution spectrophotometer or with a multivariate convolution technique. When photodiode arrays or solid-state devices such as charge coupled devices (CCDs) and charge injection devices (CIDs) are used, it is possible to measure the background at the same time as the analyte. Separate channels are used for the selection of analyte spectral lines and regions for background comparison. Background can also be reduced by adjusting the operating conditions of the source to enhance the signal-to-background ratio. For example, in DC arcs the background can be reduced by operating the plasma in an inert argon atmosphere. In ICP sources, the use of a 40.68 MHz generator results in significantly lower background due to the reduced temperatures.

### Spectral Line Interference Coefficients

The contribution of interference elements can be estimated by performing spectral line interference corrections. Calibrators are prepared in which mutually interferent elements are not present in the same solution. These solutions are then used to calibrate the system. Apparent concentrations are obtained by analyzing the ultrapure single element solutions (or solids). The interference coefficients are calculated by dividing the apparent concentration by the concentration of the interferent. In ICP-AES, the corrections are generally linear and thus a single element solution suffices to determine the correction factor. In spark and DC arc emission spectrometry, several samples are required. In practice, the determination of an element may be influenced by several other sample concomitants, and the final corrected concentration must be the summation of all the interferents. To complicate matters further, an iterative procedure must be used to deal with mutual interferences.

It should be emphasized that correction coefficients can vary as a function of ICP operating conditions.

A spectroscopic diagnostic has been developed to maintain the optimum conditions of ICP operation, in particular the flow rates of the aerosol carrier. The ratio between an atomic line of copper and an ion line of manganese is used to adjust divergent intensity responses because of unfavorable operating conditions. When this diagnostic is applied, the variability of the interference coefficients is small. The larger the correction factor, the larger the error in the quantitation of the analyte line subject to interference.

### Multivariate Procedures

Mathematical procedures have been developed to automatically extract the analyte line from a structured background comprising continuum, instrument drift, and line interferences. Convolution techniques and Fourier deconvolution can be used to recognize individual peaks and resolve analyte and interference lines by width reduction, often resulting in two- to three-fold improvements in resolution. However, the greatest impact has been made by the use of multivariate statistical methods, including Kalman filtering, multiple linear regression (MLR), partial least squares (PLS) analysis, the generalized standard additions method (GSAM) and methods for pre-reducing the number of spectral lines used. Such methods have been shown to correct for spectral line interferences separated by as little as 1–2 pm from the analyte line, even using a medium resolution spectrometer. Analyte/interferent intensity ratios are as low as 1:10 and the analyte intensity is close to the limit of detection. Kalman filtering also corrects for the noise adjacent to the line, corrects for spectral drift, calculates the concentration of unknowns, and compensates for spectral line interferences if the lines of interferents do not overlap. Since the noise averaged over the spectral window and peak area is used rather than height, true limits of detection in the sample are between one and three orders of magnitude better than conventional limits of detection obtained using three-point background compensations. All these techniques are based on the recognition of the spectral forms, i.e., a complete model of the spectral interferences has to be made. Therefore, the system requires the scanning of the sample and pure component solutions of the analytes and the interferences in the bandwidths of interest. Thus, a large amount of work is needed in order to identify the interference in individual scans before the technique can be applied.

*See also:* **Atomic Absorption Spectrometry:** Interferences and Background Correction. **Atomic Emission Spectrometry:** Principles and Instrumentation.

## Further Reading

- Bings NH, Bogaerts A, and Broekaert JAC (2002) Atomic spectroscopy. *Analytical Chemistry* 74: 2691–2711.
- Boumans PWJM (ed.) (1987) *Inductively Coupled Plasma Emission Spectroscopy: Part I. Methodology, Instrumentation and Performance*. New York: Wiley.
- Dawson JB, Snook RD, and Price WJ (1993) Background and background correction in analytical atomic spectrometry: Part I. Emission spectrometry, a tutorial review. *Journal of Analytical Atomic Spectrometry* 8: 517–537.
- Evans EH, Fisher A, and Hill S (1998) *An Introduction to Analytical Atomic Spectrometry*. New York: Wiley.
- Griffiths ML, Svozil D, Worsfold PJ, Denham S, and Evans EH (2002) Variable reduction algorithm for atomic emission spectra: Application to multivariate calibration and quantitative analysis of industrial samples. *Journal of Analytical Atomic Spectrometry* 17: 800–812.
- Montaser A and Golightly DW (1992) *Inductively Coupled Plasmas in Analytical Atomic Spectrometry*. Weinheim: VCH.
- Payling R and Larkins P (2000) *Optical Emission Lines of the Elements*. New York: Wiley.
- Van Veen EH and de Loos-Vollebregt MTC (1998) Application of mathematical procedures to background correction and multivariate analysis in inductively coupled plasma-optical emission spectrometry. *Spectrochimica Acta B* 53: 639–669.

## Flame Photometry

**E E Pickett**, University of Missouri-Columbia, Columbia, MO, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 263–269, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

## Introduction

The history of spectroscopy is closely associated with the use of flames as light sources from its beginnings more than two hundred years ago. Qualitative analyses for several elements were being carried out over the last one hundred years. Quantitative analyses evolved in the 1930s as spectrometers and light-measuring devices were improved. The peak of interest in flame emission was reached in about 1960. Thereafter atomic absorption took over much of the field: absorption measurements are easier to perform, and the range of elements that can be determined is broader. The somewhat more complex instrumentation needed was quickly put in reliable form. Graphite furnace methods extended the range of usefulness of atomic absorption. More recently, high-energy plasmas with very sophisticated instrumentation have further replaced flame emission. In this article the basic processes of flame emission will be described and the range of its usefulness re-examined. It is suggested that modern developments make it a useful method still.

## Signal Generation

Compared with other common excitation sources – the arcs, sparks, and plasmas – flames have rather

low temperatures, up to 3000 K. Most atomic transitions thus occur between the ground state and a few low-lying excited states, even with elements having great numbers of possible states.

The temperatures of several common analytically useful flames are given in **Table 1**. These are the so-called theoretical temperatures, calculated for stoichiometric fuel–oxidant gas mixtures by Snelleman. They are roughly one hundred degrees higher than most measured temperatures. Moreover, the stoichiometric mixtures do not give the highest attainable temperatures: these are reached at somewhat higher fuel-to-oxidant ratios, especially for the air–acetylene flame, due at least in part to air entrainment. Fuel richness also alters rates and extents of chemical reactions in flames. In any case, the tabulated values show the relative temperatures of useful flames.

The distribution of atoms of an element among available states of energy  $\varepsilon_i$  above the ground state is governed by their energy and the absolute temperature according to the familiar Boltzmann factor,  $\exp(-\Delta\varepsilon_i/kT)$ . The most intense lines, often called

**Table 1** Theoretical flame temperatures and maximum burning velocities of some useful flames

Flame	Volume ratio	Temperature (K)	Maximum burning velocity ( $\text{cm s}^{-1}$ )
Air–C <sub>2</sub> H <sub>2</sub>	12.5	2537	158–266
Air–C <sub>3</sub> H <sub>8</sub>	25	2267	39–43
O <sub>2</sub> –C <sub>2</sub> H <sub>2</sub>	2.5	3343	1100–2480
O <sub>2</sub> –C <sub>3</sub> H <sub>8</sub>	5	3094	370–390
N <sub>2</sub> O–C <sub>2</sub> H <sub>2</sub>	5	3148	285

Reprinted from Alkemade CThJ and Herrmann R (1979) *Fundamentals of Analytical Flame Spectroscopy* (trans. Auerback R and Gilbert PT, Jr). New York: Halsted Press.

**Table 2** Boltzmann factors for the practical range of resonance line wavelengths and flame temperatures (T)

Wavelength of resonance line (nm)	Boltzmann factors, $\exp(-\Delta\epsilon_i/kT)$	
	2000 K	3000 K
200	$2.44 \times 10^{-16}$	$3.89 \times 10^{-11}$
300	$3.89 \times 10^{-11}$	$1.15 \times 10^{-7}$
400	$1.55 \times 10^{-8}$	$6.27 \times 10^{-6}$
500	$5.69 \times 10^{-7}$	$6.84 \times 10^{-5}$
600	$6.27 \times 10^{-6}$	$3.39 \times 10^{-4}$
700	$3.46 \times 10^{-5}$	$1.06 \times 10^{-3}$
800	$1.25 \times 10^{-4}$	$2.50 \times 10^{-3}$
900	$3.39 \times 10^{-4}$	$4.8 \times 10^{-3}$

resonance lines, are those arising from permitted transitions between the ground states and certain low-energy excited states. Values of the factors are given for 2000 K and 3000 K, the useful flame temperature range, at intervals throughout the spectrum (Table 2). When corrected for their multiplicity (small whole number factors), the values give the fractions in the excited states directly. It is apparent that the vast majority of atoms in flames are in their ground states.

These considerations have several practical results.

1. Flame atomic absorption is more sensitive for those elements having their resonance lines at shorter wavelengths. (However, the line from the primary light source, the hollow cathode lamp, must also be sufficiently intense, a condition that is now readily attained.)
2. In hotter flames, elements having resonance lines at longer wavelengths may show greater sensitivity (lower detection limits) by emission than by atomic absorption. In an absorption measurement, one must always measure the difference between the incident and transmitted intensities: the detection limit is reached when the difference approaches the combined variations in these two intensities. Flame emission has no such restraint.
3. Flame emission spectra are relatively simple, and small spectrometers were thus considered adequate. The more versatile and larger modern instruments have greater resolving power so that more spectral lines are resolved and background continuum is made weaker by their greater dispersion.
4. Flame-burning velocities (Table 1) determine the minimum gas flow rates needed to keep the flame burning smoothly above the burner slot. The 'residence time', during which a given atom passes through the observed flame region, is thereby established –  $\sim 1$  ms in common flames. The slower-burning nitrous oxide–acetylene (NOA) flame is thus

preferred to the hotter but much faster-burning oxy-acetylene flame, which is also inclined to explode.

5. Excited species may be produced by mechanisms other than thermal excitation, e.g. chemiluminescence, so that the intensity is greatly increased.

## Atom Formation

Samples generally are introduced into flames in the form of solutions, by spraying or 'nebulizing' either directly into the flame along with the fuel or indirectly into a chamber to allow coarser droplets to settle out, with the finer aerosol and fuel passing on to the burner. The direct 'total consumption' approach was widely used in emission until the advent of atomic absorption, for which it was less suitable. It was turbulent, audibly noisy, and gave less complete vaporization of solutes. This type of sample introduction is still used in some clinical flame photometers.

The indirect 'premixed' nebulizer–burner combination is inherently quieter, more stable, and gives less trouble with chemical interferences. The residual flicker or 'flame noise' may often be the principal obstacle to improving detection limits. The dropwise nature of the aerosol also causes statistical fluctuation of the signal, similar to the 'shot noise' of photoelectric measurements, in both flame emission and atomic absorption.

The efficiency of transport of sample solution to the premixed flame is usually between 5% and 10%. Other more elaborate devices, such as the ultrasonic type, do somewhat better and are preferred in some applications, but are seldom used in flame photometry.

As each aerosol droplet enters the flame, many processes must occur in  $\sim 1$  ms to convert it into a cloud of excited atoms for measurement in the region just above the primary reaction zone. The measured signal depends on aerosol droplet size distribution and its rate of evaporation, volatilization of the resulting solid particles, cooling by solvent evaporation, chemical reactions among the sample and flame species, and especially any combination and dissociation reactions involving analyte atoms, including ionization and electron recombination. Some of these may be manipulated to advantage by the analyst; flame richness is especially important. Undesirable reactions involving sample constituents also require attention: these sample-dependent chemical interferences, discussed below, have received much study.

In spite of the many processes occurring, flames are relatively stable and analyte atom formation and excitation may be made reproducible to within 1% for a sufficient length of time. In this regard, they are superior to high-voltage sparks and d.c. arcs and compare well with various plasmas and low-pressure gas discharges.

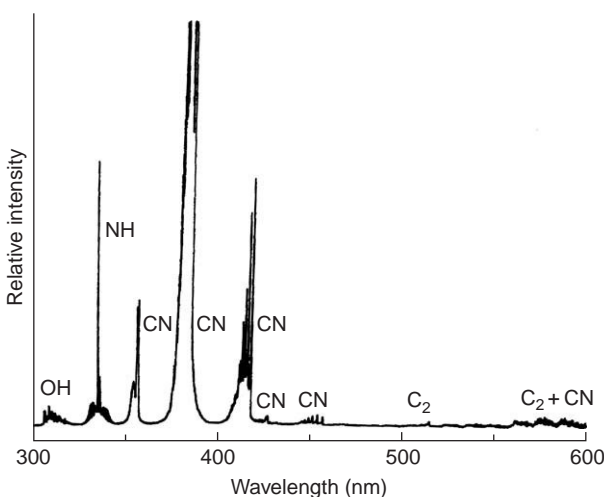
## Flame Spectrometers and Filter Photometers

The simple and inexpensive flame photometers now widely used in clinical laboratories are designed to determine sodium and potassium in blood plasma and urine, with lithium as an added internal standard element. At least two of these instruments may be used to determine calcium in plasma and also lithium when it has been administered at high levels for therapeutic purposes. All these elements give intense spectral lines in such samples, even after 100-fold dilution. Their lines are well separated in wavelength from each other. All other metals are present in much smaller amounts, except magnesium, which is very poorly excited in the cooler air–propane flames employed. Interference filters with transmission peak half-widths of several nanometers, sometimes with colored glass filters added, serve to isolate desired spectral lines without using dispersive optics. Read-out is by panel meter or digital display. Recorder connections are provided on some models. Precision is at least as good as 2% RSD; accuracy has been shown to be adequate for intended purposes.

Simpler forms of these instruments were in common use in Europe by 1940. Some could be used for analysis of plant ash solutions and soil extracts for potassium and calcium. The sodium lines were interfered with by the red–orange CaO molecular emission in such samples. A quite remarkable instrument was developed at this time by the Swedish botanist H. Lundegardh. He employed a premixed air–acetylene flame and a small quartz prism spectrograph having the desired wavelength coverage for most purposes on a 5 in. ( $\sim 12.7$  cm) photographic plate. The spectral line images were measured with a microphotometer in such a way as to achieve an approximate correction for spectral background. Up to 40 elements could be determined. He then automated the entire process!

A few of these instruments were imported and used in the USA. But flame photometry was uncommon there until the introduction of the Beckman total–consumption nebulizer–burner, producing a turbulent oxyacetylene flame. It was used with a high-quality silica prism spectrometer, a photomultiplier detector, and very simple electronic null-balancing circuitry. Flame emission analysis for many elements was thus widely practised until atomic absorption equipment became available.

Following the introduction of the NOA flame in atomic absorption by J B Willis in 1965, certain elements could be determined much more sensitively than before. The use of this flame in emission was soon investigated and found to be quite worthwhile.



**Figure 1** Emission spectrum of the nitrous oxide–acetylene flame, somewhat rich (red zone 3 mm high). Scan rate  $12.5 \text{ nm min}^{-1}$ ,  $25 \mu\text{m}$  slits. The principal molecular emission bands are labeled.

This hotter flame has a more complex and intense spectrum than the air–acetylene flame (**Figure 1**). When made fuel-rich, the NOA flame is more strongly reducing than other flames, enhancing its usefulness. To exploit its advantages in emission, a spectrometer with higher resolving power than that of the absorption instruments is preferred, so that the flame background may be more widely spread out and weakened relative to the atomic lines. Line intensity in a spectrometer varies directly with the slit width whereas continuous background intensity varies with the square of the slit width.

A half-metre grating spectrometer with  $25 \mu\text{m}$  slits typically can give a spectral bandpass or effective resolving power of  $0.04 \text{ nm}$  in the first order. With the NOA flame, quite useful detection limits for many elements may be obtained (**Table 3**). Slits of width  $10 \mu\text{m}$  may be used for further line-to-background discrimination when needed, as for the aluminium  $396.1 \text{ nm}$  measurement. To correct for background emission by scanning across a line profile, a wavelength drive at least as slow as  $0.2 \text{ nm per minute}$  is required.

Several commercial atomic absorption instruments contain monochromators with similar dispersion but are seldom equipped with slits giving less than  $0.1 \text{ nm}$  bandpass. Even then the emission mode is more sensitive than absorption, at least for the alkali metals, using the air–acetylene flame. However, their scan speeds, when provided, are too fast for more general use in emission.

The nebulizer–burner combinations perfected for atomic absorption serve equally well in emission. Gas flow controllers must permit igniting and

**Table 3** Detection limits by flame emission and atomic absorption spectrometry. Concentrations, in  $\text{mg l}^{-1}$ , giving signals twice the r.m.s. noise in the flame background

Element	Wavelength (nm)	Emission		Absorption	
		$\text{N}_2\text{O}-\text{C}_2\text{H}_2$	$\text{Air}-\text{C}_2\text{H}_2$	$\text{N}_2\text{O}-\text{C}_2\text{H}_2$	$\text{Air}-\text{C}_2\text{H}_2$
Al	396.1	0.02L			
	309.2			0.1	
Ba	553.6	0.001L		0.05	
Ca	422.7	0.0001L	0.005		0.002
Cr	425.4	0.005I			
	357.9				0.005
Dy	404.6	0.07R			
	421.2			0.2	
Er	400.8	0.04R		0.1	
Eu	459.4	0.001R		0.08	
Fe	372.0	0.05L			
	248.3				0.005
Ga	403.3	0.01L			
	287.4				0.07
Ho	405.4	0.02R			
	410.4			0.1	
In	451.1	0.002L			
	303.9				0.05
K	766.5		0.0001		0.005
Li	670.8	0.00001L	0.0001		0.005
Mg	285.2	0.005L			0.0003
Mn	403.1	0.005I			
	279.5				0.002
Na	589.0		<0.0005		0.002
Rb	780.0		0.0002		0.005
Ru	372.8	0.02			
	349.9				0.3
Sc	402.0	0.03R			
	391.2			0.1	
Tl	535.0	0.02L			
	276.8				0.2
Tm	371.8	0.02R			
	410.6			1	
Yb	398.8	0.002R		0.04	

R, strong oxide-forming elements; L, nonoxide-forming elements; I, other elements.

extinguishing the NOA flame as an air-acetylene flame. The 5 cm slot burner is perhaps most useful. A silica lens is placed so as to form an image of the flame on the face of the slit, which should be masked so that  $\sim 3$  mm of the flame above the primary reaction zone is accepted, including the red zone of the rich NOA flame.

Commercially available d.c. amplifiers with several sensitivity ranges are satisfactory, as are photomultiplier power supplies with stepwise and continuous voltage settings. Meter readout is desirable, but a recorder is more convenient and is often used for background correction. Limitations on sensitivity generally are imposed by flame phenomena, not by the stability or gain of the electronics. Although all these components are readily available, apparently no manufacturer offers a complete instrument optimized for high-quality flame emission work. The spectrometers and polychromators used with

plasmas could of course be used for flame emission, to advantage in certain instances, although this is seldom done: optimum spectral line choices largely differ for the two methods.

## Further Aspects of Flame Emission Work

### Effect of Stoichiometry in the NOA Flame

For best results, somewhat greater attention to detail is needed in flame emission than in atomic absorption spectrometry. Flame richness greatly influences background intensity and, for many elements, line intensity as well, generally with both increasing together but at different rates. The strong oxide-forming elements, denoted by R in Table 3, are more sensitive in rich flames, by as much as 50-fold for certain rare-earth elements. For these, the red CN

molecular emission needs to be increased to at least 1 cm flame height. For nonoxide formers (L in the table) lean flames are suitable; other elements (I) are best determined with a 3–5 mm red flame.

### Linearity of Response

Flame emission intensity often varies linearly with concentration over several orders of magnitude, with the upper limit set by curvature due to self-absorption. With the 5 cm slot burner, linearity is good below  $1 \text{ mg l}^{-1}$  for most elements and nearly perfect below  $0.1 \text{ mg l}^{-1}$ . For potassium this means four orders of magnitude of linearity and for lithium five. The upper limit may be raised by setting the slot burner cross-wise.

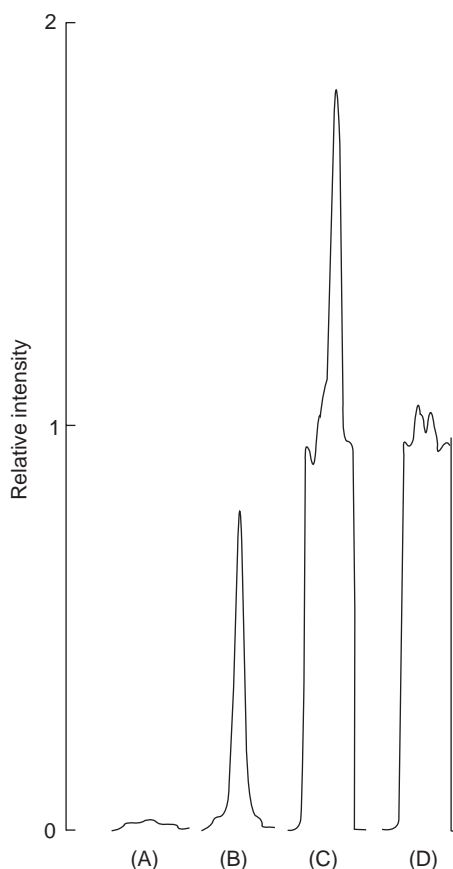
### Background Correction

Discussion of this topic is limited here to the aspects most important in flame work. Another article in this Encyclopedia includes a more general treatment.

The wavelength scan of **Figure 1** was made at high speed and does not show the complex fine structure present. With few exceptions, windows with adequate transmission can be found for any analyte. However, the sample matrix often contributes additional background. The molecular spectra of metal oxides are essentially continuous, even at high resolving power. The commonest offender probably is calcium. The familiar strong red emission of CaO spreads over several hundred nanometres; there is a strong narrow band beneath the barium 553.5 nm line and there are weaker bands in the blue and ultraviolet.

With the rather unsophisticated instruments under discussion, background may be corrected for by scanning or by successive online and offline measurements. In general, scanning is needed when background is intense whereas the other method suffices for weak background.

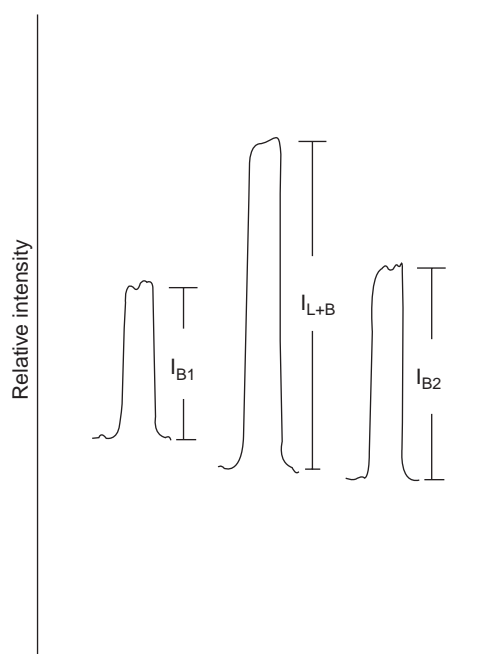
1. *Scanning.* The spectrum is scanned and recorded across the complete line profile at  $\sim 0.2 \text{ nm}$  per minute for perhaps  $0.4 \text{ nm}$  overall (**Figure 2**). The height of the lithium peak above the small CaO fine structure peak on the right is to be compared with that of the calcium standard scan. The correction is accurate to within 2% even though the Ca:Li ratio is  $1.6 \times 10^6$ . The method is slower and requires several millilitres of sample solution. (Chemical interferences obviously are minimal in this example.)
2. *Two- or three-wavelength method.* **Figure 3** illustrates the procedure for the most general case, when the flame background and sample background



**Figure 2** Four 0.3 nm scans of the lithium 670.8 nm line; scan speed  $0.2 \text{ nm min}^{-1}$ . (A) NOA flame only; (B)  $2 \mu\text{g l}^{-1}$  lithium standard solution; (C)  $3.2 \text{ g l}^{-1}$  calcium, as purified  $\text{CaCO}_3$  in  $0.1 \text{ mol l}^{-1}$  HCl, with  $2 \mu\text{g l}^{-1}$  lithium added; (D)  $\text{CaCO}_3$  solution only, same concentration as in (C). All are actual recordings, starting and ending with zero light intensity (beam blocked), showing low electrical noise.

both change with wavelength close to the line. The corrected line intensity is obtained by subtracting the average of  $I_{B1}$  and  $I_{B2}$  from  $I_{L+B}$ . If it is known by trial with typical samples that the background is flat, only one offline reading is needed. The readings should be made at wavelengths removed from the line by perhaps five times the spectral bandpass, where the line profile virtually disappears. The operator must always study the background problem whenever an unfamiliar element-sample combination appears. The two-wavelength method is often required, at least, but is easily accomplished.

A reacting or 'wobbling' quartz refractor plate was introduced for automatic background correction. Placed just behind the entrance slit, its vibration caused the line image to sweep across the exit slit,  $\sim 0.1 \mu\text{m}$  overall. Its motion was synchronized with



**Figure 3** Diagram illustrating the three-wavelength background correction method, when both flame and sample background contributions vary near the line. Online reading in the centre; readings offline by 0.2 nm, one on each side of line. Zero intensity is at the low end of the axis. Each trace starts and ends with flame background intensity.

readout so as to subtract intensities on the sides from that of the line peak.

### Interference Effects

Both chemical and physical effects can cause line intensities for samples to differ from those for simple standard solutions at the same concentration. The chemical effects can be more severe and are treated more fully.

**Chemical effects** The classic example is the suppression of calcium by phosphate, first pointed out in 1877. Such effects arise whenever analytes combine with other sample constituents to alter their rates of vaporization compared with those of the standards. The chief methods of overcoming this and other major effects are presented.

1. Prepare calibration graphs by using standard additions. In effect, each sample is used in preparing its own calibration graph. It is very slow and has various pitfalls: it is rarely used in flame emission in the NOA flame.
2. Add a 'releasing agent', a salt of another metal added in large amount to displace the analyte from its state of different, usually lower, volatility. Lanthanum is most often used, at one hundred-fold or higher concentration than the analyte.

3. Go to a hotter flame. The calcium-phosphate effect virtually disappears in the NOA flame.
4. Alter flame richness. Effects for several elements are diminished in leaner flames. It cannot be used for analytes requiring rich flames, such as rare-earth elements.
5. Separate analyte from matrix. This is most helpful when it is also used to concentrate or enrich the analyte, that is, to transfer it to a smaller volume of solution, perhaps along with other analytes at the same time. A great many methods have been described.
6. Add a bulkier, more volatile matrix. Certain high-boiling point acids, and many organic compounds that decompose without melting when heated, have recently been found to suppress several elements in flames even at concentrations as low as a few milligrams per litre. Rare-earth elements show it most strongly, but several common transition elements also suffer, e.g. cobalt and nickel. When large amounts of other salts are present, as in most practical samples, the effect is much reduced or absent.
7. Suppress ionization by adding high concentrations of salts of easily ionized metals, notably the alkali metals. Analytes having low ionization energies may be partially ionized in hotter flames. Manganese, with a fairly typical ionization potential of 7.4 eV, is ~10% ionized in the NOA flame. Alkali metal salts are added at a few hundred times greater concentration than analytes so as to overwhelm sample-to-sample variation. The same amounts are added to standards. Errors due to ionization may easily be reduced to 1 or 2%.

**Physical effects** Several physical properties mentioned under atom formation are associated with the samples themselves and may need to be controlled. Solution temperature, viscosity, surface tension, and vapor pressure all influence nebulization and transport to the flame. Many of the effects are interdependent and not easily isolated for study. For example, enhancement of signals by the addition of flammable organic solvents such as ethanol was commonly practised at one time. This caused changes in aerosol formation, transport, droplet size distribution, evaporation rate, and flame temperature. Although the solvents decreased flame temperature by partially replacing higher energy acetylene fuel, other processes were accelerated. Stability generally suffered somewhat and signals were often slow to reach their final average values. Organic solvents now are most often used incidentally, in the solvent extraction of chelates of the analytes, to enrich them and separate them from interferences.

Physical effects may be controlled in practice by frequent measurement of standards, rinsing nebulizers with solvent after each sample and performing operations at a constant rate. Automatic sample changers can improve results by performing these operations reproducibly. Partial clogging of the nebulizer uptake tube may be the main source of error remaining, as it may not be noticed until the next standard measurement.

## Practical Applications of Flame Photometry

The widely used clinical flame photometers have been described in the section above on flame spectrometers and filter photometers. Analyte concentrations and their inherent line intensities are such that samples may be diluted greatly and various difficulties avoided. The burners are small and round so that self-absorption is slight and calibration curves nearly linear. Use of lithium added at a high concentration as the internal standard helps to correct for some of the uncontrolled variables.

At higher levels of sensitivity and versatility, flame atomic absorption has replaced flame emission spectrometry chiefly for two reasons. It is easier to build absorption instruments that can be operated reliably by untrained workers than is the case for the equivalent flame emission equipment. And the range of elements determinable in many important sample types is greater by atomic absorption. For agricultural workers, the ability to analyze for zinc, copper and magnesium in their samples led to rapid acceptance of the method. In the environmental and health-related sciences, the toxic elements lead, cadmium, selenium, and others suddenly became much easier to determine. The development of electrothermal atomizers, as attachments for the basic absorption instruments, made them still more valuable.

At present, flame emission still retains advantages in approximately the same area in which it began, determination of alkali metals. However, sodium and potassium are so abundant that they are readily determined by absorption in most sample types. Rubidium has been used as a tracer in studies of insect-plant relationships, taking advantage of its low natural abundance, low toxicity, and great sensitivity by flame emission. The higher red-sensitive photomultipliers such as RCA 4840 should be used beyond 600 nm.

Lithium, however, is unique in its suitability for determination by flame emission, especially at very low levels. Even the plasmas are inferior for lithium because of the very high ionization they produce:  $2\text{ }\mu\text{g l}^{-1}$ , the detection limit often quoted for lithium, is poorer by at least 100-fold.

Lithium is a fairly common element: its abundance in the earth's crust is greater than those of copper and zinc. It too may be determined in many samples by flame atomic absorption. However, flame emission in the NOA flame was shown to be valuable for determining lithium in very small animal tissues at normal physiological levels. The  $10\text{ ng l}^{-1}$  detection limit permitted analysis for lithium in individual rat pituitaries at  $\sim 30\text{ }\mu\text{g Li per g dry weight}$ , and eventually enabled establishing the essentiality of lithium in the rat.

Beyond the alkali metals, occasional examples of the use of flame emission have been presented. Ytterbium has been used as a tracer in digestibility studies in large animals. Flame atomic absorption for this element was beset by many interferences when the wet-ashed excreta were analyzed directly. A simple rapid carrier precipitation technique was devised, with lanthanum oxalate as the carrier and the phases separated by centrifuging. A small aliquot of the digests could be taken and all steps performed in a 15 ml centrifuge tube. When combined with flame emission in the NOA flame, the method was sufficiently sensitive, reliable, and rapid. To employ atomic absorption, the method has to be scaled up 10-fold or more.

Table 3 may suggest further applications, especially when analysts need to avoid the high costs of plasma methods. With four exceptions, all elements listed give at least three times lower detection limits by flame emission than by flame atomic absorption. All may be considered sensitive in an absolute sense, with detection limits below  $0.1\text{ mg l}^{-1}$ . The four exceptions are chromium, iron, magnesium, and manganese. These elements are widespread and abundant and their determination by flame emission is generally trouble-free.

The great sensitivities for aluminium, calcium, strontium, and barium by emission should be valuable in geological and agronomic work. Laboratories engaged in rare-earth element work should have good flame emission equipment available. Although it is not widely realized at present, flame emission in its modern form can be quite useful, even, or perhaps especially, in the well-equipped laboratory.

**See also:** Atomic Absorption Spectrometry: Principles and Instrumentation; Interferences and Background Correction; Flame. Atomic Emission Spectrometry: Inductively Coupled Plasma. Quality Assurance: Internal Standards.

## Further Reading

Cresser M (1994) *Flame Spectrometry in Environmental Chemical Analysis: A Practical Guide*. Analytical



- Spectroscopy Monograph Series. Cambridge: Royal Society of Chemistry.
- Dean JA and Rains TC (eds.) (1969) *Flame Emission and Atomic Absorption Spectrometry*, vol. 1. Theory. New York: Dekker.
- Dean JA and Rains TC (eds.) (1971) *Flame Emission and Atomic Absorption Spectrometry*, vol. 2. Instruments and Techniques. New York: Dekker.
- Dean JA and Rains TC (eds.) (1975) *Flame Emission and Atomic Absorption Spectrometry*, vol. 3. Elements and Matrices. New York: Dekker.
- Ingle JD, Jr. and Crouch SR (1988) *Spectrochemical Analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Jackson KW and Chen G (1996) Atomic-absorption, atomic emission, and flame emission spectrometry. *Analytical Chemistry* 68(12): 231R–256R.
- Pickett EE and Hawkins JL (1981) Determination of lithium in small animal tissues at physiological levels by flame-emission photometry. *Analytical Biochemistry* 112: 213–218.
- Pickett EE and Koirtzmann SR (1969) Emission flame photometry – a new look at an old method. *Analytical Chemistry* 41: 28A–42A.

## Inductively Coupled Plasma

**J M Mermet**, Universite Claude Bernard, Villeurbanne, France

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

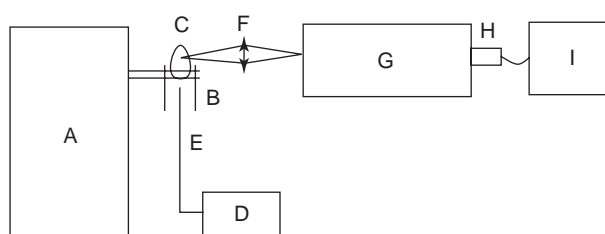
Early in the 1960s, a number of plasmas were described for the purposes of chemical analysis. This led to the availability of commercial systems for elemental analysis during the mid-1970s. Among them, inductively coupled plasma (ICP) has gained general acceptance as an atomization and a radiation source in atomic emission spectrometry (AES). This article will review the properties of such a plasma and the instrumental design required for its use in AES. It should be noted that a plasma may also be used as an ionization source.

### Basic Principle of Signal Generation

In AES, a source has two roles through its available energy: volatilization and atomization of the sample to obtain free atoms, and excitation (and ionization) of the atoms. The subsequent radiative de-excitation of the excited species is used to obtain the specific spectra of the elements present in the sample. A dispersive system is used to isolate the analytical lines (Figure 1). Among the various sources, plasmas were found to be the most suitable because of their properties. A plasma is an ionized gas that is macroscopically neutral. If a gas X is used, a plasma can be described by the following equilibrium:

$$X = \sum_{n=1}^q X^{n+} + \sum_{n=1}^q ne$$

where  $X^{n+}$  is an ion with  $n$  charges and  $e$  is the electron. Some of the properties of ideal gases, such



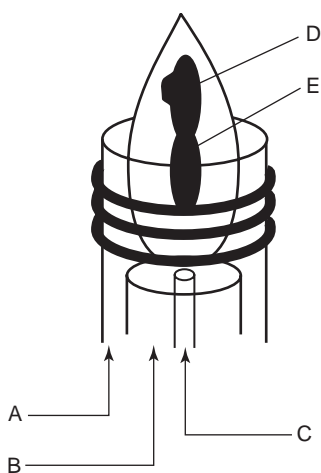
**Figure 1** Schematic diagram of an ICP system. Key: (A) HF generator; (B) torch; (C) plasma; (D) sample introduction system; (E) injector; (F) collimating system; (G) dispersive system; (H) photon detector; and (I) data acquisition, processing, and editing.

as the relationship between pressure and volume, still apply. However, the presence of charged particles (ions, electrons) leads to plasmas exhibiting different properties from those of ideal gases in terms of viscosity and thermal conductivity.

To create a plasma it is necessary to supply external energy to the electrons in order to ionize the gas. In this, a plasma differs from a flame, which derives its energy internally. An external energy source must be available that represents a significant constraint. However, it is possible, through the external source, to maintain a certain degree of control over the properties of the plasma. In practice, the plasma acts as an energy reservoir.

### Instrument Design

One of the most common ways to sustain a plasma is to use a high-frequency electrical field produced by an electrical generator. Kinetic energy is provided to the species in the plasmas to facilitate ionization of argon. The electrical field is produced through an induction coil, so there are no contaminating electrodes. This type of plasma is therefore called an ICP. Several types of oscillators are used to generate the high-frequency current: the free-running



**Figure 2** Schematic diagram of an ICP torch. Key: (A) plasma (outer gas); (B) auxiliary (intermediate) gas; (C) aerosol carrier gas; (D) emission zone for ionic lines; and (E) emission zone for atomic lines.

oscillator and the crystal-controlled oscillator. The design of a free-running oscillator is based on inductive heating technology, whereas that of the crystal-controlled oscillator is based on broadcasting technology. To be useful, the generator must exhibit enough flexibility to sustain a plasma over a large variation in impedance, corresponding to different working conditions such as ignition and the presence of water and organic solvents. Most generators are currently driven by a computer.

The plasma is created in a special device called a torch (Figure 2), which serves three purposes: it electrically insulates the plasma from the induction coil, stabilizes the plasma, and injects the sample to be analyzed. Because of the skin effect resulting from the high-frequency field, the energy is mostly deposited in the external layer of the plasma. A channel is therefore obtained along the axis of the plasma where the sample can be injected. The sample is consequently confined to the axis of the plasma. The torch materials used must be both insulating and refractory. Current torch designs employ three concentric tubes. The intermediate tube serves to accelerate the gas passing between the outer and the intermediate tubes, which reduces the gas consumption. The inner tube, or injector, serves to introduce the sample into the plasma. The tubes are usually made of silica, with the possibility of using a different material for the injector, such as alumina. Typical outer torch diameters are  $\sim 20$  mm with the inner diameter of the injector in the range 1.5–3 mm. Traditionally, optical observation is performed perpendicularly to the axis of the torch (radial viewing), with the observation height defined with reference to the highest point of the coil. Currently, most

commercially available systems make use of an observation along the axis of the torch, the so-called axial viewing.

Although it might be possible to use any gas to create a plasma, a noble gas is chosen because of its high ionization energy, the absence of stable combination with other elements, and the simple spectrum observed in emission due to the monoatomic nature of the vapor. Argon is usually selected as a compromise between its ionization energy (15.76 eV) and its availability and acceptable cost. ICPs are produced at atmospheric pressure to facilitate sample introduction and to obtain a high kinetic temperature. The kinetic temperature of heavy particles (atoms and ions) is in the range 5000–7000 K. The sample will reach the same temperature if the energy exchange is efficient. In this instance, the atomization processes are complete. The efficiency is related to the residence time of the sample in the plasma. Although laboratory plasmas are slightly ionized (0.1%), the electron number density is of the order of  $10^{20}$ – $10^{21}$  m $^{-3}$  at atmospheric pressure. This value is significantly higher than in a chemical flame. Consequently, interferences due to the presence of easily ionized elements are minimized. However, the spectral background in the ultraviolet (UV) and visible domains, which is due to radiative recombination between the electrons and the argon ions, is high, as its intensity is proportional to the square of the electron number density. It reaches a maximum at  $\sim 450$  nm, which explains the blue color of an argon ICP.

The viscosity of an argon plasma increases with temperature in the temperature range observed with an ICP. It is therefore difficult to introduce a sample carried by a cold gas, as it will encounter a medium whose viscosity is much greater. A major advantage of the ICP is that it facilitates the injection of a sample along its axis, where there is a zone of lower viscosity.

The significant operating parameters of an ICP are the following: generator power (600–2000 W) serving essentially to provide energy for sample dissociation; plasma gas flow rate (8–15 l min $^{-1}$ ) serving to obtain a stable plasma; aerosol carrier gas flow rate (0.5–1.0 l min $^{-1}$ ); and observation height with radial viewing (2–20 mm), which determines a zone within the plasma where the best excitation conditions for atomic and ionic lines are obtained. On most ICP systems, these parameters are computer controlled. They can be optimized and stored in an analytical method. Another parameter is generator frequency (27.12 or 40.68 MHz), which has an influence on plasma configuration and on the required power, as well as on the analytical properties. International regulations are available for high-frequency applications. They imply that both the

excursion range and the intensity of the field should have limited values. However, commercially available ICP systems are sufficiently shielded so that the influence of the high-frequency field on other equipment is negligible.

Although a noble gas plasma would appear to provide a simple medium, many species exist and are competing with one another. Many mechanisms of excitation and ionization have been described to explain the observed spectra in terms of the roles played by electrons and argon ions. The use of argon makes it possible to ionize a large number of the elements of the periodic table. The exceptions are helium, neon, and fluorine. However, it may be difficult to supply enough energy to excite certain ions to higher energy levels, which precludes the observation of such ions. The alkali metals are examples of elements whose ions have too high excitation energies. Most elements will exhibit both atomic and ionic lines, in which case the ionic lines usually provide the best sensitivities. For some elements (e.g., antimony, arsenic, boron, gallium, and phosphorus), only the atomic lines are used for analytical purposes. Moreover, for some elements such as chlorine, bromine, nitrogen, and oxygen, the most sensitive atomic lines are located in the UV at wavelengths shorter than 150 nm. This requires a vacuum (or nitrogen purge) between the plasma and the optical detector, and optical components made of  $\text{MgF}_2$ . Systems that allow a wavelength range down to 120 nm are now commercially available. In practice,  $\sim 70$  elements of the periodic table can be determined using an ICP in AES.

## Sample Introduction Techniques

Although an ICP exhibits a high kinetic temperature, the residence time during which the sample remains in the plasma is limited to a few milliseconds. Consequently, it is necessary to introduce the sample in the form of small liquid or solid particles of the order of a few micrometers to ensure complete volatilization and atomization. This is a serious constraint in the design of the sample introduction system.

In general, solids are introduced in the form of fine particles, either in suspension in a liquid (slurries) or in a gas stream. Spark ablation and laser ablation can also be used for this purpose. Direct introduction of solids, placed in a refractory vessel at the base of the plasma, has also been described.

The most common way for a sample to be introduced into an ICP is a solution. In order to produce fine droplets from a liquid, a nebulizing system is required. To date, the most widely used type is the pneumatic nebulizer, which has three

variants: the concentric, the cross-flow, and the V-type. V-nebulizers are less subject to blocking. In using a nebulizer, the aim is to achieve efficient nebulization at a low carrier gas flow rate ( $< 1 \text{ l min}^{-1}$ ). However, the nebulizers in service today provide droplet distributions centered on  $\sim 20 \mu\text{m}$ . It is therefore necessary to add a chamber, called a spray chamber, in order to filter out droplets larger than a few micrometers in diameter. The overall efficiency of the nebulizer associated with the spray chamber is very low – only  $\sim 2\text{--}5\%$  for liquid consumption in the range  $1\text{--}2 \text{ ml min}^{-1}$ . Currently, micronebulizers, i.e., micronebulizers efficient below  $0.2 \text{ ml min}^{-1}$  are commercially available, with efficiency higher than 50%. A pneumatic nebulizer is usually fed via a peristaltic pump to ensure a constant uptake rate irrespective of the viscosity of the liquid.

Ultrasonic nebulizers are free of these difficulties as aerosol production is independent of gas flow, and droplet size is related to ultrasonic frequency. Droplets with diameters of  $\sim 1 \mu\text{m}$  can be obtained with frequencies above 1 MHz. The efficiency is high enough to require the addition of a desolvation system to eliminate water. The desolvation system consists of an oven associated with a cooling system to trap the water vapor. Detection limits are usually improved by a factor of 10 in comparison to a pneumatic nebulizer. Although ultrasonic nebulizers are attractive in this respect, their use is as yet not widespread, due to their prohibitive cost.

Gases can be directly introduced into the plasma. The only limitation is the change in resistivity of the plasma, which can degrade the coupling efficiency between the generator and the plasma. The tolerance of the plasma to the amount of foreign gases is typically below  $0.1 \text{ l min}^{-1}$ . An application of gas introduction is the use of volatile hydrides (As, Bi, Se, Pb, etc.) or the addition of molecular gases ( $\text{N}_2$ ,  $\text{H}_2$ ) to improve the thermal conductivity of the plasma.

In order to avoid solid dissolution, which may be tedious and time consuming, direct solid analysis may be performed by using ablation, i.e., removal of material from a target. Spark ablation may be used for a conductive material, while laser ablation may be used for any type of material. In the case of solids in the form of fine powders, suspensions in water (slurries) may be an alternative to sample dissolution.

A current trend is the coupling of another technique with AES, such as liquid chromatography, ion chromatography, or flow injection analysis, to enhance the capability of the plasma, particularly in the field of speciation (the so-called ‘hyphenated’ techniques). This coupling can be considered as a highly sophisticated sample introduction system. In

general, the interface between the two hyphenated techniques makes use of the standard nebulizing system, although some adaptation may be required.

## Dispersive Systems

The task of a dispersive system is to measure the net line intensity of an analytical line. This includes the measurement of both the gross line intensity and the background. Off-peak background measurements are usually carried out with an interpolation of the background values at the analytical line wavelength. In AES, work is performed essentially in the range 165–800 nm, with Al II 167 nm and K I 770 nm being the lowest and highest wavelengths, respectively. As noted previously, there is a trend to work down to 120 nm for lines such as Cl I 134 nm. Dispersive systems with diffraction gratings are used. Two categories of dispersive systems are used: simultaneous polychromators and sequential monochromators, depending (1) on the type of grating, i.e., plane or concave grating with high line number or plane echelle grating with low groove number, and (2) on the type of detector, either photomultiplier tube or multichannel detector. Polychromators allow several wavelengths to be measured simultaneously. Conventional polychromators make use of a concave grating, and are generally of the Paschen–Runge type, with the entrance slit and the exit slits located on a so-called Rowland circle. A photomultiplier tube is set up behind each exit slit. The wavelength selection is in principle difficult to change. Background correction can be carried out through a small discrete lateral displacement of the entrance slit. Alternatively, multichannel detection can be used for polychromators. With a concave grating, a polychromator is obtained with a linear association of array detectors along the so-called Rowland circle. With an echelle grating, cross-dispersion based on the use of a prism is designed to avoid order overlaps due to the use of a low ruling density but high orders. Consequently, a two-dimensional spectrum is obtained, which is highly suitable for a two-dimensional multichannel detector. The main advantage of a simultaneous system is the speed of measurement. Monochromators work sequentially, by displacing one of the optical components in the dispersive system. They make use either of a plane grating (with an Ebert–Fastié or a Czerny–Turner mount) or a concave one (with a Paschen–Runge mount). In the first instance, wavelength selection is obtained by rotating the grating and in the second instance by displacing the exit slit and detector. An echelle grating monochromator is based on the motion of either of two optical components, grating and prism,

or the exit slit and the detector. Drastic improvements have been made in the accuracy and speed of positioning for wavelength. The positioning is controlled directly by a computer with periodic wavelength recalibration. Two measurement modes can be used: peak search mode and direct peaking mode. The former mode is usually for systems making use of grating rotation, whereas the latter is for systems based on slit and detector displacement. The peak search mode uses a scan over the line profile. The peak intensity is computed with the use of curve fitting. The main advantage of the peak search mode is the knowledge of the shape of the line profile, which can be displayed or printed. However, this mode is time consuming. The main advantage of the direct peaking mode is the speed of measurement but no information is obtained on line profile. When using multichannel detection, this problem is no longer a cause for concern, because in a single shot both the line profile and the adjacent background are measured.

The use of gratings with a large number of lines per millimeter (up to 4200) and that can work in the second order implies that the practical resolution in monochromators is limited essentially by the band pass and the optical aberrations. Practical resolution is in the range 5–20 pm with the highest resolution in the UV range. Similar resolutions are obtained with the use of echelle gratings.

## Spectral and Matrix Effects

Interference phenomena can be classified in two categories: interference that modifies the signal because of a change in the aerosol transport and filtration and a change in atomization, excitation, or ionization conditions due to the matrix or other elements, and interference that disturbs observation of the signal by partially or totally overlapping the analytical line. While the consequences of matrix or interelement effects have been widely described, their origins are still unexplained. The effects usually correspond to a change in the analyte line intensities and a shift in the spatial distribution along the axis of the plasma. However, these effects can be minimized by increasing the generator power and the residence time of the sample, which is obtained by decreasing the carrier gas flow rate and increasing the inner diameter of the injector.

Spectral interferences are much more critical and can result from the spectra of argon, OH, bands and elements injected into the plasma. Argon emits lines only above 300 nm or in the far-UV. Therefore, no unknown lines can be assigned to argon in the range 165–300 nm. Although water is almost entirely

dissociated, the OH species exists below a temperature of 3000 K and its emission is significant  $\sim 300$  nm, which hampers the measurement of a few analytical lines. In the case of elements, spectral interferences occur because of the complexity of their spectra, and the broadening of the spectral lines. This complexity can be explained by excitation phenomena within the plasma, as well as by improvements in the detection systems. None of the wavelength tables that are currently available cover all the lines emitted by an element. Line-broadening is mainly due to the Doppler effect resulting from the high kinetic temperature. Typical full-widths at half-maximum are in the range 1–6 pm. A few elements exhibit hyperfine structure. Elements such as cobalt, iron, molybdenum, niobium, tantalum, tungsten, and uranium are the most interfering. This is why software written for sequential dispersive systems and multichannel detector based dispersive systems allows the selection of an analytical line free of spectral interferences by displaying the spectra of the analyte and the other elements. For each new matrix, it may be necessary to make such a line search.

## Analytical Performance

The AES limits of detection in a solution obtained with an ICP fall between 0.01 and  $50 \mu\text{g l}^{-1}$ . For a very sensitive element such as Be, the best limits of detection are below  $0.05 \mu\text{g l}^{-1}$ , while for a difficult element such as Pb, they are below  $4 \mu\text{g l}^{-1}$  (Table 1). The principal limitation is the high background, which produces significant background noise. As a salt concentration of up to  $50 \text{ g l}^{-1}$  can be used, corresponding limits of detection less than  $1 \text{ mg per kg}$  can be obtained in a solid before dissolution.

ICP-AES is not an absolute method for concentration measurement. A calibration using known standards must be first carried out for each element, and then the concentration of an element in an

unknown solution is determined via the calibration graph. One important feature of the ICP is the large linear response range of over four to six orders of magnitude. Curvatures can be observed at high concentrations because of the self-absorption phenomenon. However, calibration graphs are generally used within a limited concentration range.

The accuracy of an analytical method, i.e., the agreement between the average of the results and a 'true' value, can be estimated by using reference materials. In spite of the large range of materials currently available, it may be difficult to obtain adequate standards for work at trace levels. Use of an internal standard can improve accuracy. An accuracy of 1% can be expected.

The precision refers to two concepts: repeatability, or the dispersion of the results obtained during an experiment where it is assumed that the operating parameters have not changed; and reproducibility, or the dispersion of the results among several experiments. Precision is usually estimated by using either the standard deviation or the relative standard deviation (RSD) of the fluctuations observed for several replicates of the measurements. Typically, the RSD values fall between 0.4% and 2% without an internal standard and can reach 0.1% with the use of an internal standard. The main limitations in the precision results from shot noise and flicker noise. The former noise is related to the random nature of the photoelectrons emitted at the detector level. Its RSD is inversely proportional to the square root of the signal. Flicker noise is mostly related to the sample introduction system, i.e., sample formation and transport. Its RSD is independent of the signal over a large range of concentration.

Long-term stability is an important factor: absence of drift allows the ICP user to avoid periodic, time-consuming recalibration. The main causes of drift are due to changes in the generator power and the sample introduction system efficiency. The latter is usually caused by a partial blocking of the nebulizer or a variation in the spray chamber temperature. Currently, an RSD of 1% can be expected over a period of 4 h. As mentioned previously for precision, use of an adequate internal standard can improve the long-term stability.

## Limitations

One limitation of the ICP is that it is based primarily on the use of solutions. The number of elements soluble in the same solution is limited. Sample preparation is time consuming and there is a risk of loss or contamination at the trace amount level.

**Table 1** Typical limits of detection ( $\mu\text{g l}^{-1}$ ) obtained with an axial viewing-based ICP-AES system

Ag	0.6	Fe	0.2	Sb	1.5
Al	0.2	Hg	1.3	Sc	0.1
As	1.2	I	10	Se	1.5
Au	0.6	K	1.5	Si	1.5
B	0.3	Li	0.5	B	1.3
Ba	0.04	Mg	0.03	Sr	0.03
Be	0.04	Mn	0.05	Ti	0.15
Bi	2.6	Mo	0.2	Tl	1
Ca	0.03	Na	0.6	V	0.2
Cd	0.1	Ni	0.3	W	2
Co	0.2	P	1.5	Y	0.2
Cr	0.15	Pb	1.5	Zn	0.1
Cu	0.18	S	3	Zr	0.3

Obviously, this is not a limitation for samples already in the form of a solution. The use of a pneumatic nebulizer is another limitation of the technique. Besides risk of blocking, most of the instabilities arise from the nebulizer. Nevertheless, the main limitation of ICP-AES is the possibility of spectral interferences as mentioned above. Although the most sensitive lines can be used most of the time, it may be necessary to verify that they are free from spectral interferences when a new matrix is used.

## Applications

Virtually any type of inorganic and organic sample can be analyzed in ICP-AES using the multielement capability of the technique: biological samples, metals, alloys, electronic materials, glass, ceramics, environmental samples, oils, geological samples, and so on. This versatility combined with adequate analytical performance at a reasonable cost justify the current acceptance of this technique, which is illustrated by the number of instrument companies involved in this field.

**See also:** **Atomic Emission Spectrometry:** Principles and Instrumentation; Interferences and Background Correction. **Atomic Mass Spectrometry:** Inductively Coupled Plasma.

## Further Reading

- Boumans PWJM (1987) *Inductively Coupled Plasma Emission Spectrometry. Part I: Methodology, Instrumentation and Performance. Part II: Applications and Fundamentals*. New York: Wiley.
- Broekaert JAC (2002) *Analytical Atomic Spectrometry with Flames and Plasmas*. Weinheim: Wiley-VCH.
- Hill SJ (1999) *Inductively Coupled Plasma Spectrometry and its Application*. Sheffield: Sheffield Academic Press.
- Montaser A and Golightly DW (1987) *Inductively Coupled Plasmas in Analytical Atomic Spectrometry*. New York: VCH Publishers.
- Moore GL (1989) *Introduction to Inductively Coupled Plasma Atomic Emission Spectrometry*. Amsterdam: Elsevier.
- Sneddon J (1990) *Sample Introduction in Atomic Spectroscopy*. Amsterdam: Elsevier.
- Thompson M and Walsh JN (1989) *Handbook of Inductively Coupled Plasma Spectrometry*. Glasgow: Blackie.

## Microwave-Induced Plasma

**G M Greenway**, University of Hull, Hull, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Microwave-induced plasmas (MIPs) can be used as excitation sources for atomic emission spectrometry (AES). The main advantage of the MIP over the more commonly used inductively coupled plasma (ICP) is that it can be operated with helium as the support gas, although argon has also been widely used. With helium plasmas halogens and other nonmetals with high excitation energies can be determined with much greater efficiency due to the presence of highly energetic species in the plasma. The microwave plasma is also less expensive to operate than an ICP as it typically runs at lower power and needs less cooling, so that there is lower gas consumption. These factors mean, however, that the plasma is physically small and does not afford a high enough plasma energy density to vaporize or evaporate solid or liquid samples. In practice, this means that the plasma is easily quenched.

## Basic Principles of Signal Generation

To produce the type of MIP used most commonly, a microwave generator operating at 2450 MHz is

connected by a waveguide to a resonant cavity. When switched on, a standing wave is set up in the cavity and at the center of the cavity where the electrical field is greatest there is a discharge tube through which the plasma gas flows. The plasma is initiated by seeding the plasma gas with electrons, usually derived from a Tesla coil. These electrons oscillate with the microwave field, initially being in phase with the applied field, but as the field is reversed the electrons are not able to change direction rapidly enough, and hence a phase lag is initiated between electron motion and field. The accelerating electrons collide with gaseous atoms, causing frequent direction changes. The process continues with the electrons gaining energy from the field and losing energy through collisions until they gain enough energy to excite or ionize an atom. Stabilization of the MIP occurs through repetition of this process and successive collisions between electrons released by ionization and neutral gas atoms until equilibrium is obtained.

The MIP formed is not in local thermodynamic equilibrium, that is, it cannot be characterized by one temperature, but needs defining in terms of the translational, rotational, vibrational, and electron temperatures as well as the more commonly understood surface temperature. The electron temperature is

highest as the energy is first transferred to the chemical species via electron collisions. The surface temperature as measured by a thermocouple or pyrometer has been found to be between 500 and 1000 K for a low-power helium plasma. The excitation mechanism is complex and involves both high- and low-energy electrons, ions, and metastable atoms and molecules. It is thought to be the highly energetic metastable helium species that give the helium MIP its high excitation capacity.

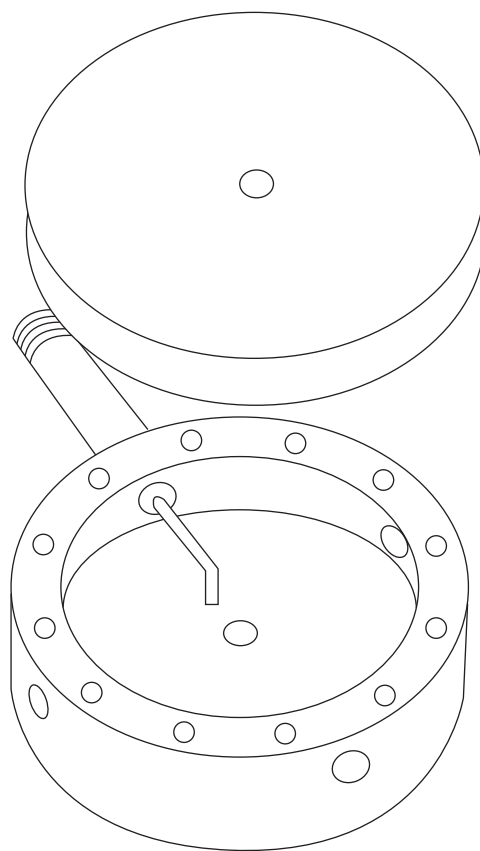
## Instrument Design

The components required to produce a low-power atmospheric helium plasma include a generator, a resonant cavity, and a discharge torch. In the past, the microwave generator was usually from a medical diathermy unit that supplied 100–200 W at 2450 MHz. More recently magnetron tubes manufactured for microwave ovens have been used. These are capable of supplying higher power levels (500 W), giving greater stability to the plasma, which facilitates sample introduction.

The microwave cavity must be designed to transfer power efficiently from the generator to the plasma gas. Usually, a coupling device is used to match the impedance of the cavity and plasma to that of the rest of the system, thus keeping reflected power to a minimum. To account for changes caused by igniting the plasma and different plasma conditions the cavity is often provided with tuning and impedance-matching devices. Performance of the cavity depends both on the material from which it is constructed and the mode in which it operates.

Fields and currents produced in microwave cavities decrease exponentially with penetration into the conductive material. The degree of penetration is called the skin depth and the cavity should therefore be constructed with a material of low skin depth, which is of high conductivity. However, as it is only the surface of the cavity that is important, cavities are usually made from a brass or aluminum substrate with a coating of high-conductivity metal such as silver or copper.

The cavities most often used in earlier work were either the shortened 3/4-wave or 1/4-wave coaxial types, but these were unable to sustain an atmospheric helium plasma. Fehsenfeld introduced a shortened 1/4-wave radial cavity that was improved by Beenakker and with certain minor modifications this proved to be a successful design for sustaining an atmospheric helium plasma due to its strong electrical field. The cavity, shown in **Figure 1**, was designed to have its resonant frequency at 2450 MHz, with a minimal cavity volume so that a high-energy density



**Figure 1** A  $TM_{010}$  type resonant cavity. The position of the coupling loop and the viewing and cooling ports are shown. The discharge tube is centered in the holes at the top and bottom faces of the cavity. (Matousek JP, Orr BJ, and Selby M (1984) Microwave-induced plasmas: Implementation and application. *Reviews in Analytical Atomic Spectroscopy* 7: 275–314.)

at a given input power was attained. The discharge tube was located at the position where the electrical field was a maximum. It was operated in the  $TM_{010}$  mode, where T and M stand for transverse and magnetic, respectively. The subscripts refer to the number of full wavelength patterns around the circumference, the number of half-wavelength patterns across the radius, and the number of half-wavelength patterns across the axis. The resultant cavity had an axially directed electrical field with a maximum at the center. The power was transferred to the cavity by a coupling loop that was mounted perpendicular to the circularly directed magnetic field. The tuning and coupling of the cavity were not ideal and several workers made modifications to improve this design. The designers of the commercially available atomic emission detector (AED) required a cavity that did not need constant retuning and they chose to use a re-entrant cavity with a special coupling loop design.

With conventional MIP discharge tubes the helium plasma forms a bolus in contact with the torch



walls and with no central channel. One of the main practical consequences of this is the short lifetime of the discharge tubes. One approach to improve torch lifetime has been to water-cool the discharge tube by modifying the cavity slightly. This has been found to be successful in reducing erosion of the torch walls as well as in reducing background levels and noise in the detection of silicon and oxygen. Reduction in noise and interference for nitrogen and oxygen detection can also be achieved by adding a window at the exit of the discharge tube to prevent back-diffusion of air. A useful addition to the plasma setup is a gas flow system to allow for solvent venting and the addition of make-up and reagent gases, all of which can greatly improve the plasma performance.

Another disadvantage of the conventional MIP discharge tubes is that the sample may pass around the cooler edges of the plasma and therefore not be excited efficiently. Part of the success of the ICP is that it has an elongated doughnut shape with a central channel into which the sample can easily be introduced. Different torch designs have therefore been investigated where toroidal or turbulent flow of the plasma and support gases have been designed to keep the plasma off the torch walls and to introduce the sample into the center of the plasma.

The optical detection systems used in MIPs are the same as those used for other atomic spectrometers and can be either single or multichannel. Fourier transform-based spectrometers have also been used. Conventional optical systems are best designed if the plasma is viewed from the exit of the discharge tube, as is possible with the  $TM_{010}$  type cavity, rather than through the walls of the discharge tube, which become etched. The commercially available AED uses a computer-controlled silicon photodiode array detector which has multielement detection capability over segments of spectra. In recent years, MIP sources have also been investigated as ion sources for mass spectrometry.

### Alternative Microwave Discharges

The focus of alternative microwave designs has been to produce a plasma that could cope more easily with the introduction of liquid aerosols. Often, higher powers (up to kilowatts) are used to give increased power density and plasma size. One of the original alternative designs was the capacitively coupled plasma, which is generated by transmitting microwaves through a rectangular waveguide to an electrode. This operates at high power and is therefore more robust and can easily cope with liquid sample introduction but tends to be less precise and have a higher background than the MIP.

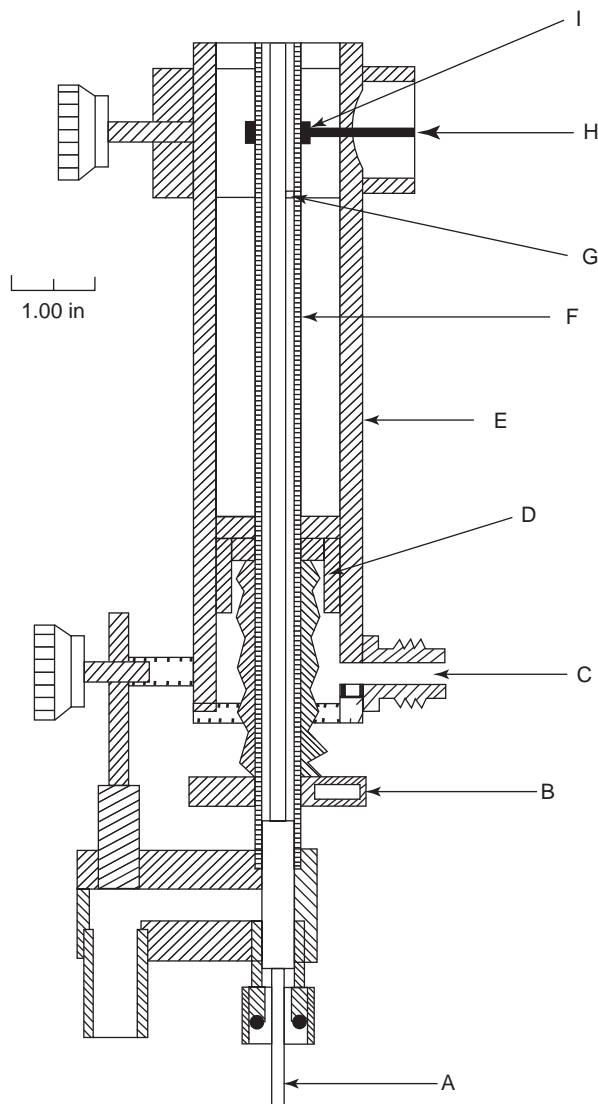
The surfatron or surface wave device has also been investigated, especially for liquid sample introduction. Instead of using a resonant cavity to sustain this type of plasma, a length of coaxial transmission line with a capacitive gap at one end and a short circuit at the other is used to launch surface waves along the length of the device. The plasma is sustained in a silica tube that is held in a cavity that shapes the axial electric field. Depending on the power, flow rates, and discharge tube geometry, a helium plasma can be formed that consists of an annulus of plasma surrounding a darker central hole. This central plasma channel appears to allow greater sample penetration into the plasma and the plasma is also stable over a wide working range, making it particularly suitable for sample introduction of liquids by different nebulization techniques if operated at high power.

A more radically different approach for liquid sample introduction is the microwave plasma torch (MPT) (Figure 2). In a recent design three concentric tubes were used where the outer two brass tubes acted as a coaxial waveguide for the microwaves eliminating the need for a cavity. The inner quartz channel was nonconducting and the helium plasma that was formed (at power levels 70–200 W) was similar to the ICP plasma, having a central channel that was more efficient for sample introduction as described previously. A sheath gas was used to stabilize the MPT and reduce air entrainment.

Different plasma gases have been used especially argon. A microwave-induced nitrogen discharge at atmospheric pressure has been developed with higher thermal energy which could accept aerosol samples and cope with larger sample sizes (3  $\mu$ g). The plasma was run at powers of  $\sim 250$  W and had a tail flame that reached out of the resonant cavity. A nitrogen MIP based on a  $TE_{011}$  (transversal electric) mode cavity (2.45 GHz, 1.5 kW) was compared with an argon ICP for liquid aerosol sample introduction atomic emission detection but was found to have poorer detection limits by one to two orders of magnitude. Recently, MIP-boosted glow discharge AES has also been investigated. A microsecond-pulse glow discharge source was boosted by an MIP to give a tandem glow discharge source for solid sample and surface analysis applications.

Perhaps the most interesting recent developments, however, have been in the area of miniaturization. A miniaturized MIP gas chromatography (GC) detector has been developed using a single 1.5 mm sapphire wafer (Figure 3). The system uses a microstrip to transmit the microwaves, having top and bottom planar conductors sandwiching an insulator through which the microwaves travel. This design allows high



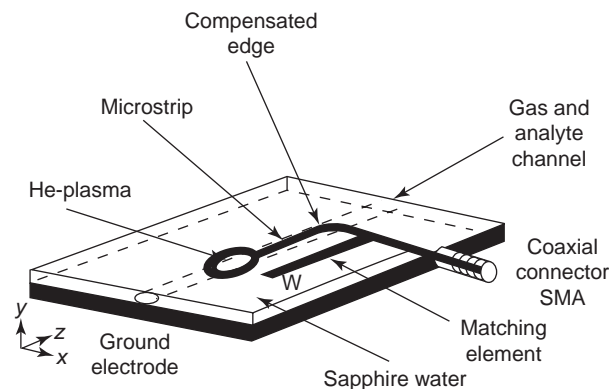


**Figure 2** Schematic diagram of the microwave plasma torch. A, quartz central channel; B, tuning assembly; C, introduction of sheath gas; D, plunger to change cavity length for tuning; E, outer conductor; F, inner copper conductor; G, copper frit; H, microwave coupler; I, microwave coupling collar. (Reprinted with permission from Pack and Hieftje (1997) An improved microwave plasma .... *Spectrochimica Acta B: Atomic Spectrometry* 52: 2163–2168; © Elsevier.)

field strengths in the plasma channel but can be operated using the small, low-cost integrated microwave power supplies used in mobile communication devices. A self-igniting helium MIP can be sustained with a power input of 5–30 W and a gas flow of 50–10 000 ml min<sup>-1</sup>.

## Spectral and Matrix Effects

The background emission spectrum due to the plasma is low for both helium and argon and the spectra of both these gases are well characterized so that



**Figure 3** Microstrip plasma source for helium (He-MSP). The square sapphire wafer is 30 mm long and has a height of 1.5 mm. It has a grown-in channel of 0.9 mm in diameter. The microstrip line has a width,  $w$ , of 0.8 mm. (Bilgiç *et al.*, (2000) *Journal of Analytical Atomic Spectrometry* 15: 579–580; reproduced by permission of The Royal Society of Chemistry.)

specific spectral lines may be avoided. If a particularly low background is needed, wavelength modulation may be used. Other spectral interferences are due to impurities from the support gases and the atmosphere such as carbon dioxide, nitrogen gas, and water vapor, or impurities from the sample introduction system and discharge tube. Memory effects are also a problem due to analyte atoms collecting in etched areas of the discharge tubes.

The matrix effect of major concern in the MIP is the effect of sample introduction on the stability of the plasma due to the fact that the plasma is not in local thermodynamic equilibrium. The size of sample that can be introduced can be very limiting, depending of course on the operating conditions used. The MIP has also been found to be affected by the presence of refractory compounds in the sample matrix, which may promote collisional deactivation of metastable species, thus reducing the excitation ability of the MIP. This can sometimes be overcome by increasing the microwave power. This problem is in addition to the more common chemical interferences that can also occur where the analyte forms involatile oxides. In low-pressure plasmas, signal enhancement has been seen in the presence of alkali and alkaline earth chlorides. This is probably due to the formation of volatile chloride salts that enhance atomization, although it may be that they affect the excitation mechanisms in some way. The lack of local thermodynamic equilibrium in the MIP also means that ionization interferences can be significant because changes in the electron density of the plasma will affect its stability and excitation mechanisms. The presence of easily ionized elements tends to lead to strong enhancement effects (although suppression is sometimes seen).

The use of doping gases to overcome matrix effects, especially when the MIP is being used as a gas chromatographic detector, is widespread. The gases act as scavengers and are added to prevent carbon from depositing on the wall of the discharge tube, because if this occurs severely distorted chromatographic peaks may be obtained. Different gases are used for the determination of different groups of elements. For the simultaneous determination of carbon, hydrogen, chlorine, and bromine, oxygen is the chosen gas; but if elements that form refractory compounds are present, hydrogen would be needed.

## Sample Introduction Techniques

Correct sample introduction is the key to the successful use of the MIP, as unless a modified design is used, it can only cope with such small amounts of sample and solvents tend to quench it. Its main application has been as a detector for GC as gases are less likely to affect the stability of the plasmas, but a considerable amount of effort has been spent developing suitable nebulization systems for solutions. The various techniques are discussed below.

### Solution Nebulization

As described earlier, probably the best approach to introducing solutions into the MIP is to use a different plasma design such as the MPT; however, pneumatic nebulizers can be used with argon plasmas if the sample is desolvated first. Ultrasonic nebulizers have also been used, usually followed by desolvation. A low-power argon MIP was investigated with wet aerosol using an integrated TE<sub>101</sub> rectangular cavity with an ultrasonic aerosol plasma cooling system. It was found to perform reasonably well for environmental samples but the detection limits were not low.

### Electrothermal Atomization

This approach is more successful for liquid and solid samples as the electrothermal atomizer will evaporate the solvent, and vaporize and atomize the analyte before it enters the MIP. This means the plasma has to excite only the sample atoms and the problems associated with the low thermal temperature are overcome, only leaving a restriction on sample size. The sample is passed from the atomizer to the MIP in a stream of the plasma support gas, the flow of which can easily be adjusted to match the requirements of the plasma.

Carbon furnaces and cups have been utilized as electrothermal atomizers but these tend to be rather complex and most work has been with metal strips or filaments of platinum, tantalum, or tungsten.

A liquid sample is applied to the filament, which may be placed directly in the support gas stream. In this simplest of configurations great care must be taken with the evaporation of the solvent to ensure that the stability of the plasma is retained. Alternatively, a valve system can be incorporated to vent the solvent whilst the flow of plasma support gas is maintained to the plasma. For all these systems the emission profiles tend to be very sensitive to any changes in conditions and different conditions are needed for different elements.

### Laser Ablation

Laser ablation has been investigated for MIP sample introduction with some success. The advantage of this technique is that, as with electrothermal atomization, a two-step process is used. The laser is used to volatilize and atomize the sample before it is introduced into the plasma for excitation, thus overcoming any problems with low thermal temperatures in the MIP. The sample is transferred from the laser ablation cell to the plasma via a carrier gas that is usually the support gas. The main problem with laser ablation is lack of precision due to the shot-to-shot variation in laser power and its nonlinear effect on the ablation process. To overcome this several different normalization techniques have been investigated.

A major advantage of this technique is that it can be used for solid samples.

### Chemically Generated Vapor and Gas Introduction

The MIP can be used for direct analysis of gas samples either by generating the plasma in the sample, for example, when looking for impurities in argon, or by injecting the gaseous sample into the plasma. Alternatively, an analyte in a liquid sample can be converted to a vapor by chemical reaction and the vapor can then be flushed into the plasma support gas stream via drying tubes. Perhaps the most widely used of these techniques is hydride generation, which can be used for the determination of several elements including arsenic, antimony, and selenium. The other well-established technique used is the cold-vapor mercury method in which mercury compounds are reduced by SnCl<sub>2</sub>. Generators have also been designed to produce volatile chlorides such as bismuth chloride and volatile carbonyl compounds such as nickel carbonyl and halogens.

### Gas Chromatography Detection

The main application of the MIP atomic emission spectrometer has been as a detector for GC, where it is more commonly known as an AED.

A commercially available GC–AED system has been available since 1991, this was originally produced by Hewlett Packard (now Agilent) but in 2002 the product was transferred to a partner company, Joint Analytical Systems GmbH (JAS). Initial attempts at combining GC with MIP were made difficult because only packed separation columns were available and the flow rates and sample sizes for these chromatographic systems were too great for the MIP so that complex venting systems were necessary. The introduction of the capillary column much simplified the GC–MIP interface. The capillary columns used much slower flow rates and smaller samples so that the interface could be achieved by placing the end of the column into the discharge tube only a few millimeters away from the plasma. Loss of sample on transfer was therefore unlikely, although the torch design did need modifying so that the plasma support gas supply was not interrupted. In the new commercial systems a solvent venting system has been retained, as this still tends to give better peak shape and allows the injection of larger sample sizes up to 100  $\mu\text{l}$ .

The advantage of the MIP as a detector for GC is that it is sensitive and can be used in two modes. It can be used as a selective detector for organic compounds containing heteroatoms such as halogens or phosphorus in pesticides. Alternatively, if the system has multielement detection capabilities several elements can be monitored simultaneously and empirical formulae can then be calculated. To a certain extent it is possible to have compound-independent calibrations with an MIP because the plasma breaks down most compounds into their constituent elements. This is particularly useful if calibration standards are not available because it allows the quantification of a whole series of compounds with a specific heteroatom based on a calibration for one analyte with the selected heteroatom. However, there are problems with this technique because the elemental response has been shown to alter in chemically different compounds unless great care is taken. This is probably due to difficulties in the atomization process. In recent years there has been a great reduction in the cost of GC–MS (mass spectrometry) and due to its high sensitivity this technique has tended to take over from the GC–AED. The ability of the GC–AED to selectively detect specific atoms is however still of great use and some workers have used hybrid systems with both MS and AED for identification purposes.

#### Detection for Other Separation Techniques

Because the MIP has been such a successful detector for GC, it was inevitable that workers have tried to

extend its use. For liquid chromatography the problem of introducing a liquid into the plasma has been addressed. One way to overcome this is to use a moving wheel. The column effluent is directed as a mist onto a continuous moving wheel interface. The solvent is then evaporated with a flow of hot nitrogen. This leaves the dry analyte residue that is carried by the wheel into the plasma for volatilization, atomization, and excitation. For supercritical fluid chromatography, a surfatron has been used with carbon dioxide and dinitrogen oxide as the chromatographic carrier gas. The conditions were optimized and evaluated. Limited work has also been published on coupling the MIP with capillary electrophoresis using an ion-exchange interfacing capillary. A plasma has also been created directly onto a thin-layer chromatography plate by connecting the inner conductor of a coaxial cable to a stainless steel capillary tube through which helium plasma gas could flow. Further developments in these areas are likely to continue.

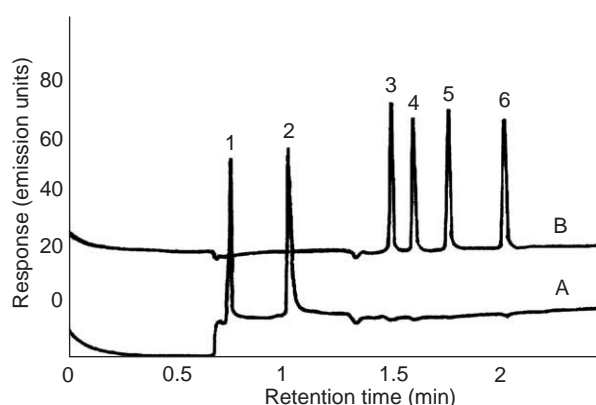
## Applications

Despite the problems described above the MIP has been used for many applications. The limits of detection obtained for the system very much depend on the sample introduction technique and are best either for vapor introduction or preatomization systems. The major application of the MIP has been as a detector for GC, either in the selective mode or for determining empirical formulae with multielement detection systems. Table 1 shows the limits of

**Table 1** Analytical characteristics of AED detection for selected elements

Element	Wavelength (nm)	LOD ( $\text{pg s}^{-1}$ )	Selectivity over carbon ( $\times 10^{-3}$ )
N	174.2	15–50	2–5
S	180.7	1–2	5–20
C	193.1	0.2–1	–
P	178.1	1–3	5–8
C	495.8	15	–
H	486.1	1–4	–
Cl	479.5	25–40	3–10
Br	478.6	30–60	2–6
F	685.6	60–80	20–50
O	777.2	50–120	10–30
Si	251.6	1–7	30
Hg	253.7	0.1–0.5	250
Pb	261	0.2–1	300
Sn	271	1	300

Reprinted with permission from Leo *et al.* (2002) Gas chromatography with atomic emission detection: A Powerful technique. *Trends in Analytical Chemistry* 21(9 & 10): 618–626; © Elsevier.



**Figure 4** GC-MIP AES chromatogram for the simultaneous determination of mercury ( $50 \text{ ng l}^{-1}$  as Hg) and Sn ( $15 \text{ ng ml}^{-1}$  as Sn) compounds of ethylated species (direct injection): 1,  $\text{MeHg}^+$ ; 2,  $\text{Hg}^{2+}$ ; 3, monobutyltin; 4, tripropyltin chloride; 5, dibutyltin; 6, tributyltin. A, 253.65 nm Hg emission line; B, 270.65 nm tin emission line. (Botana *et al.* (2002) *Journal of Analytical Atomic Spectrometry* 17: 904–907; reproduced by permission of The Royal Society of Chemistry.)

detection, selectivity, and dynamic range obtainable with an AED.

In the selective mode this type of system is particularly useful for speciation studies and has been used to determine different organolead, organotin, and organomercury compounds in environmental samples such as fish, water, and air. **Figure 4** shows chromatograms obtained for the rapid simultaneous determination of methyl mercury, inorganic mercury, and mono-, di-, and tributyltin species in water samples. The analytes were derivatized for analysis and concentrated using headspace solid-phase micro-extraction before injection into the GC. By using a

multicapillary column with the commercial AED system, mercury and tin compounds were detected at low ppt-levels in 5 min timescales. The detector has also been widely used for the determination of both halogen- and phosphorus-containing pesticides in biological samples and for industrial applications where heteroatoms such as sulfur and halogens need to be determined in oils, polymer additives, and pharmaceutical compounds.

*See also:* **Atomic Absorption Spectrometry:** Interferences and Background Correction. **Laser-Based Techniques. Liquid Chromatography:** Principles; Instrumentation. **Mass Spectrometry:** Overview. **Pesticides. Phosphorus. Sulfur.**

## Further Reading

- Croslyn AE, Smith BW, and Winefordner JD (1997) A review of microwave plasma sources in atomic emission spectrometry: Literature from 1985 to the present. *CRC Critical Reviews in Analytical Chemistry* 27(3): 199–255.
- Risby TH and Talmi Y (1983) Microwave induced electrical discharge detectors for gas chromatography. *CRC Critical Reviews in Analytical Chemistry* 14: 231–265.
- Rosenkranz B and Bettmer J (2000) Microwave-induced plasma-optical emission spectrometry-fundamental aspects and applications in metal speciation. *Trends in Analytical Chemistry* 19: 138–156.
- Van Stee LLP, Brinkman UATH, and Bagheri H (2002) Gas chromatography with atomic emission detection: A powerful technique. *Trends in Analytical Chemistry* 21: 618–626.
- Zander AT and Heftje GM (1981) Microwave supported discharges. *Applied Spectroscopy* 35: 357–371.

# ATOMIC FLUORESCENCE SPECTROMETRY

**D J Butcher**, Western Carolina University, Cullowhee, NC, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Atomic fluorescence spectrometry (AFS) is an analytical method used to determine the concentration of elements in samples. The sample is converted to gaseous atoms, and the element of interest is excited to a higher electronic energy level by a light source. Following excitation, the atoms are deactivated by the emission of a photon. The measured fluorescence is this emission process. Instrumentation for AFS

includes a light source to excite the atoms radiatively; an atom cell to convert the sample into gaseous atoms; and a detection system to collect the fluorescence radiation. The combination of a tunable laser system as the light source and an electrothermal atomizer (ETA), or graphite furnace, as the atom cell provides an instrument that can determine femtogram ( $10^{-15} \text{ g}$ ) quantities of many elements. Commercial AFS instrumentation employs vapor-generation techniques that can provide sensitivity for several elements to the parts per trillion (ppt) level. This article briefly reviews basic principles and instrumentation of AFS, and discusses the technique's analytical figures of merit, including its detection limits and linear dynamic range of calibration curves.

In addition, physical phenomena that hinder AFS analysis, including background signals, are discussed, along with methods to minimize their effects. Representative applications of AFS are described to demonstrate the ability of the technique to determine elements in samples.

## Basic Principles of Atomic Fluorescence

AFS is a method of elemental analysis that involves the use of a light source to excite gaseous atoms radiatively to a higher energy level, followed by a deactivation process that involves emission of a photon. This emission process provides the measured fluorescence signal. AFS can be distinguished from the related atomic spectrometric techniques of atomic absorption spectrometry (AAS) and atomic emission spectrometry (AES) because it involves both radiative excitation and deexcitation.

### AFS Transitions and Spectra

AFS excitation and deexcitation processes involve changes in the energy of valence electrons and are called electronic transitions. Analytical applications of AFS employ transitions in the ultraviolet (UV)–visible region of the electromagnetic spectrum (between 200 and 800 nm). AFS transitions may involve a combination of absorption, fluorescence, and non-radiative processes.

Resonance fluorescence involves radiative excitation of an atom from the ground state 0 to an excited state 1, followed by the emission of a photon (fluorescence) to deactivate the atom (Figure 1A). Resonance fluorescence is frequently the most intense transition for a given element, and is usually

employed when conventional (nonlaser) light sources are employed.

In a nonresonance fluorescence transition, the photons involved in absorption and fluorescence processes have different wavelengths (Figure 1B). The particular transition shown in Figure 1B is called Stokes direct-line fluorescence, which is frequently used for AFS with laser excitation. Nonresonance transitions have the advantage that a wavelength selection device can be used to distinguish between fluorescence and scattered source radiation.

Atomic fluorescence spectra are composed of a number of narrow lines whose width at half-maximum is typically 0.005–0.01 nm. The narrow width of AFS lines compared to molecular fluorescence spectra obtained in solution (whose width is frequently 10–100 nm or more) is due to the absence of vibrational and rotational transitions in atoms and the presence of the atoms in the gas phase.

### AFS Radiance Expressions

The expressions for the fluorescence radiance depend upon the intensity of the light source. For a relatively low-intensity (conventional) light source, the fluorescence radiance ( $B_F$ ) for a resonance transition (Figure 1A) is given by

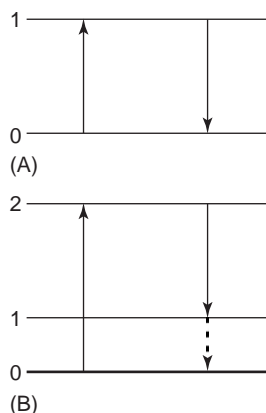
$$B_F = GYE(\lambda_0)\sigma(\lambda)n_T \quad [1]$$

where  $G$  considers the geometry over which fluorescence is collected;  $Y$  is the fluorescence yield, which describes the fraction of atoms that are not quenched by collisions, and is typically 0–0.5;  $E(\lambda_0)$  is the irradiance of the light source at the central frequency of the absorption transition;  $\sigma(\lambda)$  is the absorption cross-section, which describes the ability of atoms of an element to absorb light; and  $n_T$  is the total number of atoms present in the atom cell.

Several conclusions can be drawn from eqn [1]. First, quenching reduces the fluorescence signal when low-intensity light sources are employed. Second, under these conditions, the fluorescence intensity is directly proportional to source irradiance (Figure 2). Under these conditions, the use of a more intense light source increases the fluorescence signal size. Third, the fluorescence signal is proportional to the number of atoms, indicating that AFS can be used for quantitative analysis.

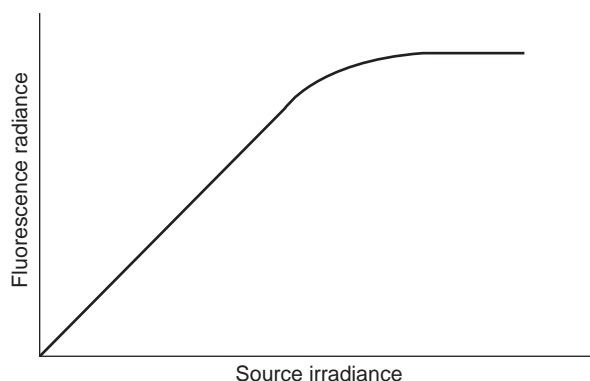
A different expression for the fluorescence radiance is employed by high-intensity sources (lasers):

$$B_F = GFAn_T \quad [2]$$



**Figure 1** AFS transitions: (A) resonance fluorescence and (B) nonresonance fluorescence. Radiative transitions are shown by solid lines; nonradiative transitions by dashed lines.

where  $F$  is a constant that includes Planck's constant and the central frequency of the absorption



**Figure 2** AFS saturation curve.

transition, and  $A$  is the Einstein coefficient for fluorescence, which describes the rate of emission from the excited state. At high source intensities, the fluorescence signal is independent of the fluorescence signal yield and the source intensity. The saturation curve (Figure 2) shows that the fluorescence intensity increases with source irradiance until a maximum signal is obtained. This condition is called saturation. The high sensitivity of laser-excited AFS (LEAFS) is due to the ability of a laser to achieve saturation and therefore provide the maximum possible fluorescence signal.

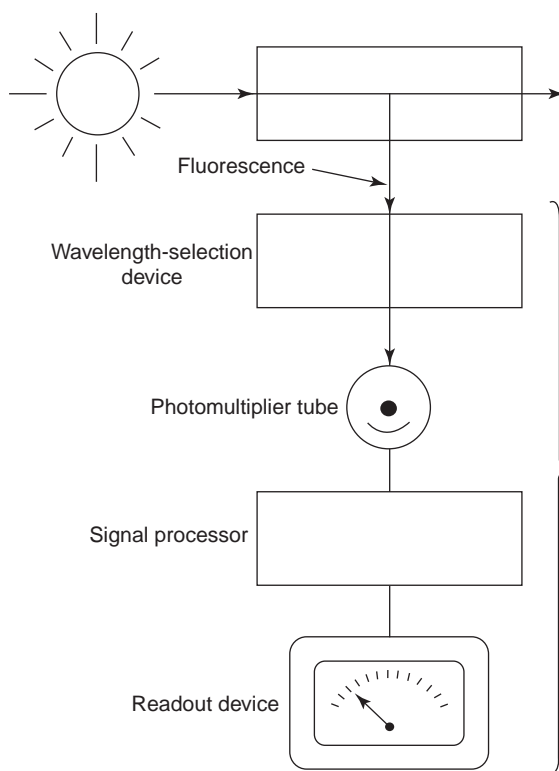
## Instrumentation

AFS was developed as an analytical technique in the 1960s and 1970s following the commercial development of AES and AAS instrumentation. For most elements, AFS detection limits (minimal detectable signal) were about the same as those of AES and AAS, and hence instrument companies did not have a compelling reason to develop AFS instrumentation. Hence, commercial development of AFS instrumentation has been relatively limited.

The principal components of AFS instrumentation are shown in Figure 3. These include a light source, to excite the atoms radiatively; an atom cell, to convert the sample into gaseous atoms; and a detection system, which consists of a wavelength selector, a photodetector, a signal processor, and a readout device. The detection system is usually orientated at  $90^\circ$  or  $180^\circ$  with respect to the light source to minimize the amount of stray source light reaching the photodetector.

### Light Sources

The light source is employed to excite atoms radiatively in the atom cell. From the discussion in the section on AFS radiance expressions, it is clearly advantageous to use an intense light source to obtain



**Figure 3** Instrumentation for AFS.

the maximum AFS signal. Light sources for AFS can be classified into two primary types: conventional (nonlaser) sources and laser sources.

**Conventional light sources** A significant amount of research has been done to develop more intense conventional sources to increase the fluorescence signal size. Conventional light sources that have been used for AFS include hollow cathode lamps (HCLs), electrodeless discharge lamps (EDLs), and continuum sources.

**Hollow cathode lamps** A HCL is composed of a silica envelope that contains 1–5 Torr of argon or neon and two metal electrodes. HCLs are almost ideal line sources for AAS because of their high stability and narrow linewidth (0.002 nm), but their relatively low intensity is a disadvantage for AFS. High-intensity hollow cathode lamps (HI-HCLs) provide increased intensity by use of an additional electrode to separate the atomization and excitation processes. The irradiance of the HI-HCLs is a factor of 20–100 times greater than that of conventional HCLs, and provides better sensitivity for AFS.

**Electrodeless discharge lamps** EDLs employ either microwave energy (microwave-excited EDLs) or radiofrequency energy (radiofrequency-excited EDLs)

to atomize and excite analyte atoms in a sealed silica tube containing an inert gas at low pressure. Microwave-excited EDLs have been commonly employed for AFS because of their high spectral irradiance, which for many elements is a factor of 10 higher than for HI-HCLs. In spite of their high intensity, the use of microwave EDLs is limited due to their commercial unavailability.

**Continuum sources** Continuum sources, unlike the line sources discussed above, offer the potential for multielemental analysis with a single excitation source. High-pressure xenon-arc lamps have been most widely used as continuum sources for AFS because of their high intensity. Each lamp consists of a silica envelope that contains 10–30 atm of xenon and two electrodes to excite the xenon. Improved sensitivity is obtained by the use of a parabolic mirror to focus the emission into an intense beam.

**Lasers** Lasers provide sufficient intensity to saturate atomic transitions, and hence provide the maximum fluorescence signal. For LEAFS, the laser must be capable of generating wavelengths throughout the UV and visible regions in order to excite as many elements as possible. These requirements necessitate the use of a laser system composed of three components: a pump laser, a dye laser, and a frequency doubler.

A pump laser emits light of a single wavelength that is obtained in the form of pulses whose duration ranges from 3 to 200 ns. Commonly used pump lasers include nitrogen, excimer, and neodymium–yttrium aluminum garnet (Nd:YAG) lasers. The Nd:YAG laser requires a frequency doubling system to convert its infrared radiation into visible light to pump a dye laser. Excimer lasers have been commonly used for LEAFS because of their relatively high pulse energies (10–1000 mJ) and their relatively high repetition rates (up to 1000 pulses per second). A high repetition rate is desirable because the AFS detection limit is inversely proportional to the square root of the repetition rate.

A dye laser employs a solution of a dye that absorbs the pump laser radiation and emits fluorescence at wavelengths that are characteristic of that dye. A wavelength-selective device, such as a grating, is incorporated into the laser cavity. The laser radiation is tuned by moving this device. By appropriate choice of the dye, laser radiation at wavelengths between 320 and 900 nm is obtained. Since the mid-1990s, tunable solid-state lasers have been developed as alternative to dye lasers. Optical parametric oscillator lasers have advantages of broad tuning range and ease of operation compared to dye lasers.

A frequency doubler employs a nonlinear crystal to convert the visible output into UV radiation between 190 and 320 nm by a process called second harmonic generation. Generally, only three or four crystals are needed to cover completely this range of wavelengths. The combination of these components provides a light source with sufficient power to saturate AFS transitions throughout the UV–visible region. The principal drawbacks of the laser systems are difficulty of use, high capital and maintenance costs, and poor reliability.

### Atom Cells

The atom cell converts the sample into gaseous atoms. Generally, it is necessary to convert all standards and samples into solution form by the use of acids or other reagents before introduction into the atom cell. Commonly used atom cells for AFS include cold vapor (CV) cells, flames, and ETAs (graphite furnaces).

**Cold vapor cells** Mercury is commonly determined using a CV quartz cell. Chemical reagents, typically tin(II) chloride, are employed to convert mercury dissolved in aqueous solution to gaseous elemental mercury. A flow of argon is employed to transport the mercury to the quartz cell. CV-AFS is a very simple and sensitive technique, with detection limits as low as 0.1 ppt for commercial instrumentation (Table 1). AFS also has the advantage of a long linear range of its calibration curves, typically four to five orders of magnitude.

**Flames** The most commonly used atom cell for AFS is the flame. A flame is formed by burning two gases, one that serves as the oxidant (e.g., air) and the other as the fuel (e.g., acetylene). A capillary is used to suck the sample solution into the flame by the passage of air past one end of the capillary (the Venturi effect). The solution is converted into an aerosol that is introduced into the flame. Flame detection limits are relatively poor because conversion of a solution into

**Table 1** Cold vapor and hydride generation AFS detection limits and linear dynamic range

Element	Technique	Detection limits (ppt)	Linear dynamic range (orders of magnitude)
Hg	Cold vapor	0.1	4–5
As	Hydride generation	10	4–5
Bi	Hydride generation	10	4–5
Sb	Hydride generation	10	4–5
Se	Hydride generation	2	4–5
Te	Hydride generation	10	4–5



an aerosol results in only about 1–10% of the sample actually entering the flame.

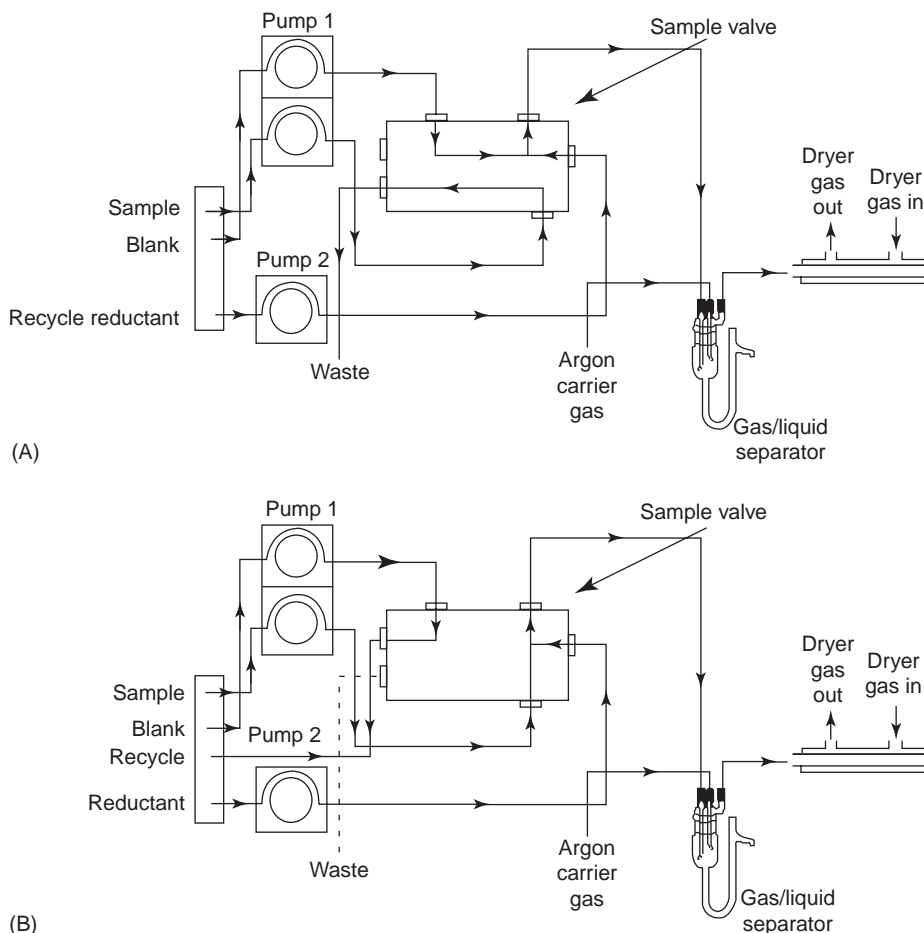
Advantages of flames include their low cost and ease of use. Disadvantages of flames for AFS include relatively poor sensitivity, relatively low temperatures, and high chemical reactivity, which may allow the analyte to form molecules and causes a reduction in the fluorescence signal (chemical interferences).

An alternative method of sample introduction for the hydride-forming elements (antimony, bismuth, arsenic, selenium, tellurium, tin) is called hydride generation (HG). Aqueous samples or standards are treated with sodium borohydride and hydrochloric acid to form the volatile hydrides (e.g.,  $\text{SbH}_3$ ), which are transported to an argon–hydrogen diffusion flame. A schematic diagram of a commercial HG system is shown in **Figure 4**. The flame decomposes the hydrides into gaseous atoms. HG provides much lower detection limits than conventional flame techniques because of the higher atomization efficiency and the low spectral background from the relatively cool diffusion flame. Detection limits between 2 and

10 ppt are available using commercially available AFS instrumentation (**Table 1**).

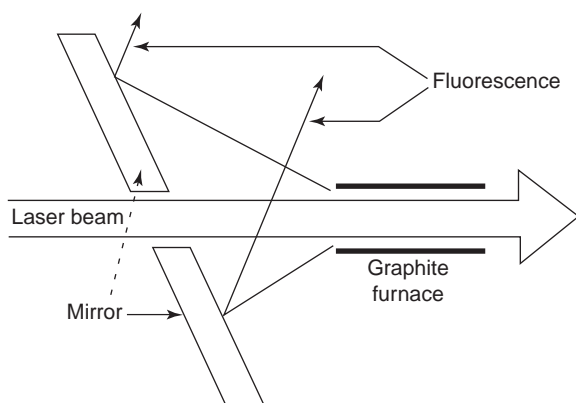
**Electrothermal atomizers (graphite furnaces)** ETAs, also called graphite furnaces, are widely used atom cells for AAS. The atom cell consists of a graphite tube (25–30 mm long by 4–6 mm ID with a 1 mm wall thickness), which is mounted between two electrodes. The temperature of the graphite tube is varied from room temperature up to 2700°C by the application of up to 12 V and several hundred amperes across the furnace by a power supply. A volume of between 5 and 100  $\mu\text{L}$  of sample is introduced into the tube through a sample port. An inert gas such as argon surrounds the tube to prevent decomposition in the presence of oxygen.

Three heating steps are required to obtain a furnace signal. First, the furnace is heated to between 100°C and 200°C to remove the solvent in the drying step. The temperature of the furnace is then raised to 300–1500°C to try to remove as much nonanalyte material in the sample as possible without the loss of



**Figure 4** Sample preparation module of a commercially available AFS instrument. (Reprinted with permission from <http://www.psanalytical.com>.)





**Figure 5** Front-surface illumination for graphite furnace LEAFS.

analyte. This is called the pyrolysis step. The furnace is then heated to a sufficiently high temperature to produce analyte atoms in the atomization step. A transient signal is obtained whose temporal width is typically 3–7 s.

The principal advantage of the ETA compared to a flame is higher sensitivity. ETA-AAS detection limits are one to three orders of magnitude lower than flame AAS detection limits for two reasons. First, the entire sample is introduced into the furnace, without the low efficiency of a flame. Second, the residence time (period of time that atoms remain in the atom cell) of atoms in a furnace is 3–7 s, while atoms remain in a flame for only a few milliseconds.

A considerable body of work has been performed using ETAs with LEAFS. Fluorescence is collected from a graphite tube using a collection scheme called front-surface illumination (Figure 5), which is the collection of fluorescence at  $180^\circ$  to the direction of the laser beam. A mirror, through which a hole is drilled to allow passage of the laser radiation, is used to collect fluorescence along the bore of the tube. Front-surface illumination allows the use of unmodified graphite tube furnaces for LEAFS.

Table 2 compares the detection limits of ETA-LEAFS with those of ETA-AAS. LEAFS detection limits are typically one to five orders of magnitude lower than those of ETA-AAS. In addition, calibration graphs for ETA-LEAFS are linear for four to seven orders of magnitude, which is far superior to the two to three orders of magnitude obtained by ETA-AAS.

## Detection System

### Wavelength-Selection Devices

A wavelength-selection device serves to discriminate against light of all wavelengths except for the

**Table 2** Comparison of ETA-LEAFS and ETA-AAS detection limits

Element	ETA-LEAFS absolute limit of detection (fg)	ETA-AAS absolute limit of detection (fg)
Ag	10	500
Al	100	4 000
Au	10	10 000
Cd	0.5	300
P	8000	5 500 000
Pb	0.2	5 000
Tl	0.1	10 000

fluorescence wavelength of the analyte. Monochromators have been frequently employed as wavelength selectors for AFS because of the ease of wavelength selection. A high-resolution monochromator is not required for line-source excited AFS because resolution is determined by the width of the light source rather than the detection system.

Bandpass filters have also been used extensively in AFS instruments. For example, they are employed in the commercially available CV- and HG-AFS instruments. For graphite furnace LEAFS, careful studies have been made that compare the use of monochromators and filters. Detection limits for lead and thallium were improved by two to three times by the use of very narrow (1 nm) bandpass filters.

### Photodetectors

The photomultiplier tube (PMT) has been widely used as the detector for AFS because of its sensitivity and long linear dynamic range. The operation of the PMT is described elsewhere in this encyclopedia. Commercial AFS instrumentation employs a solar-blind PMT, which is highly sensitive for wavelengths in the UV and vacuum UV wavelengths.

## Sources of Background Emission

A background signal is caused by light originating from sources other than analyte fluorescence reaching the photodetector. Although a wavelength-selection device is used to distinguish against background that is not of the fluorescence wavelength, it cannot remove background at the fluorescence wavelength. Important sources of background for AFS include scattered source radiation, atom cell emission, spectral line, and spectral band interferences.

Under many experimental conditions, the largest source of background for AFS is scattered source radiation. Scatter may be caused by source radiation striking particles in the atom cell, by reflections off the atom cell (e.g., in a furnace), or by components of

the sample matrix. Scatter is minimized by the use of a narrow line source to reduce the amount of unabsorbable radiation, and a nonresonance transition to allow the wavelength-selection device to discriminate against it.

A second source of background signal is atom cell emission, which is light produced by the atom cell. Atom cell emission generally increases with the fluorescence wavelength and with the temperature of the atom cell. Hence, for AFS, the temperature of the atom cell should be as low as possible to minimize atom cell emission, but high enough to atomize the analyte without chemical interferences.

Other sources of background include spectral line (nonanalyte atomic fluorescence) and spectral band (molecular fluorescence) interferences. Spectral line interferences are caused by the presence of another element that can absorb source radiation and emit fluorescence sufficiently close to the analyte wavelength to be collected by the detection system. Spectral band interferences involve the absorption of source light by a molecule whose fluorescence is collected by the detection system. Nonanalyte atomic fluorescence and molecular fluorescence are minimized by the use of a narrow line source and a nonresonance transition. This is in contrast to AES, where spectral interferences are sufficiently severe that a high-resolution monochromator is required.

## Methods of Background Correction

Background correction is widely used in AAS to distinguish between the absorption of light by the analyte and other phenomena that reduce the intensity of the light transmitted through the atom cell, such as molecular absorption and scattering of light by particles. AFS methods of background correction are based upon techniques that were originally developed for AAS. Background correction for AFS involves the measurement of analyte fluorescence plus background, followed by a measurement of background only. The subtraction of these two measurements gives a background-corrected signal. The types of background correction used for AFS have depended upon the type of light source employed. Here, three methods of background correction for AFS are described: two-line background correction, wavelength modulation, and Zeeman background correction.

### Two-Line Background Correction

Two-line background correction, which is employed with conventional line sources, uses a second wavelength that is emitted by the light source (or

alternatively, by a second light source) at which background is measured. Background correction is achieved by subtraction of the measurement made at the second wavelength from the measurement made at the analytical wavelength. In order for the technique to work, the second line may not correspond to a fluorescing line in the sample. It is also assumed that the background is the same at both wavelengths, and hence the wavelengths should be close together. Although two-line background correction works reasonably well for dilute samples, it is generally ineffective for samples that contain high concentrations of scattering materials.

### Wavelength Modulation

Wavelength modulation, which has been employed with continuum source and laser excitation, involves alternatively measuring signal plus background at the analytical wavelength and background 0.001–0.1 nm away from this wavelength. Generally, between 10 and 500 measurements per second are made at and away from the analytical wavelength. The instrumentation used to carry out wavelength modulation depends on the excitation source. Wavelength modulation is relatively effective at discriminating against scatter and atom cell emission because the background measurement is made very close to the analytical wavelength, but it cannot correct for background whose size varies near the analytical wavelength (structured background).

### Zeeman Background Correction

Zeeman background correction, which is widely used in AAS, commonly uses an AC electromagnet placed around a graphite furnace to split atomic energy levels of the analyte. For AFS, when the magnet is off, fluorescence plus background are measured. When the magnet is on, the analyte energy levels are split away from the analytical wavelength, so that absorption (and hence fluorescence) cannot occur, and only background is measured. Subtraction of the two measurements provides a background-corrected signal. Zeeman background correction has been used for AFS with EDLs and lasers as the excitation sources. The primary advantage of Zeeman compared to wavelength modulation or two-line background correction is that the correction is made at the analytical wavelength, and hence Zeeman accurately corrects for structured background.

## Applications

With the availability of commercial AFS instrumentation, hundreds of applications have been reported

since the mid-1990s. One of the most important applications of AFS is chemical speciation, which involves quantitative analysis of each of the chemical forms of an element in a sample. Speciation information is necessary in environmental and toxicological applications, because of differences in toxicity between chemical compounds.

For example, although inorganic mercury compounds are toxic, methylated mercury compounds are of special concern because of their ability to penetrate biological membranes and their efficient accumulation. Moreover, inorganic mercury is converted to methylmercury (MMHg) in aquatic sediments by sulfate-reducing bacteria. Because MMHg is retained in the fatty tissues of animals, it accumulates through the aquatic food chain. Contaminated fish pose a potential route of exposure for humans to MMHg.

AFS is well suited to speciation because of its simple instrumentation and excellent sensitivity. Instrumentation typically consists of either gas chromatography (GC) or high-performance liquid chromatography, to separate the various chemical species, with the AFS instrument serving as the detector. GC has the advantages of higher sensitivity and faster analysis times, but it is limited to volatile and thermally stable compounds.

In addition to mercury, AFS has been employed for the speciation of antimony, arsenic, selenium, and tellurium. Samples analyzed include environmental samples (sediments, water, etc.), food (fish, rice, etc.), and biological fluids (urine, blood, liver, etc.). It is anticipated that the applications of AFS will continue to increase because of the technique's favorable analytical characteristics.

**See also:** **Atomic Absorption Spectrometry:** Principles and Instrumentation; Interferences and Background Correction; Flame; Electrothermal. **Atomic Emission Spectrometry:** Principles and Instrumentation; Flame Photometry. **Elemental Speciation:** Practicalities and Instrumentation. **Laser-Based Techniques.** **Optical Spectroscopy:** Radiation Sources; Detection Devices.

## Further Reading

- Butcher DJ, Dougherty JP, McCaffrey JT, *et al.* (1987) Conventional source excited atomic fluorescence spectrometry. *Progress in Analytical Spectroscopy* 10: 359–506.
- Butcher DJ, Dougherty JP, Preli FR, *et al.* (1988) Laser excited atomic fluorescence spectrometry in flames, plasmas and electrothermal atomizers. *Journal of Analytical Spectrometry* 3: 1059–1078.
- Butcher DJ (1995) Laser-excited atomic and molecular fluorescence in a graphite furnace. In: Sneddon J (ed.) *Advances in Atomic Spectroscopy*, vol. 2. San Diego: JAI Press.
- Caballo-López A and Luque de Castro MD (2003) Slurry sampling-microwave assisted leaching prior to hydride generation-pervaporation-atomic fluorescence detection for the determination of extractable arsenic in soil. *Analytical Chemistry* 75: 2011–2017.
- Greenfield S, Hieftje GM, Omenetto N, Scheeline A, and Slavin W (1986) Twenty-five years of analytical atomic spectroscopy. *Analytica Chimica Acta* 180: 69–98.
- <http://www.psanalytical.com/>
- Michel RG (1987) Atomic fluorescence spectrometry. In: Meyers RA (ed.) *Encyclopedia of Physical Science and Technology*, vol. 2. New York: Academic Press.
- Ortiz AIC, Albarrán YM, and Rica CC (2002) Evaluation of different sample pre-treatment and extraction procedures for mercury speciation in fish samples. *Journal of Analytical Atomic Spectrometry* 17: 1595–1601.
- Smith BW, Glick MR, Spears KN, and Winefordner JD (1989) A comprehensive table of atomic fluorescence detection limits and experimental conditions. *Applied Spectroscopy* 43: 376–414.
- Sneddon J, Thiem TL, and Lee Y-I (eds.) (1997) *Lasers in Analytical Atomic Spectroscopy*. New York: VCH.
- Tao H and Nakahara T (2002) New developments in hydride generation – atomic spectrometry. In: Sneddon J (ed.) *Advances in Atomic Spectroscopy*, vol. 7. Amsterdam: Elsevier.
- Winefordner JD and Omenetto N (1979) Atomic fluorescence spectrometry: basic principles and applications. *Progress in Analytical Spectroscopy* 2: 1–183.
- Yan X-P, Yin X-B, Jiang D-Q, and He X-W (2003) Speciation of mercury by hydrostatically modified electroosmotic flow capillary electrophoresis coupled with volatile species generation atomic fluorescence spectrometry. *Analytical Chemistry* 75: 1726–1732.

# ATOMIC FORCE MICROSCOPY

**See** MICROSCOPY TECHNIQUES: Atomic Force and Scanning Tunneling Microscopy

# ATOMIC MASS SPECTROMETRY

Contents

**Inductively Coupled Plasma**

**Laser Microprobe**

## Inductively Coupled Plasma

**E H Evans**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Inductively coupled plasma mass spectrometry (ICP-MS) is a combination of two established techniques, namely the inductively coupled plasma (ICP) and mass spectrometry (MS). The ICP is an extremely suitable ion source for inorganic MS because:

- the high temperature of the ICP ensures almost complete decomposition of the sample into its constituent atoms and
- conditions within the ICP result in highly efficient ionization of most elements in the periodic table and, importantly, these ions are almost exclusively singly charged.

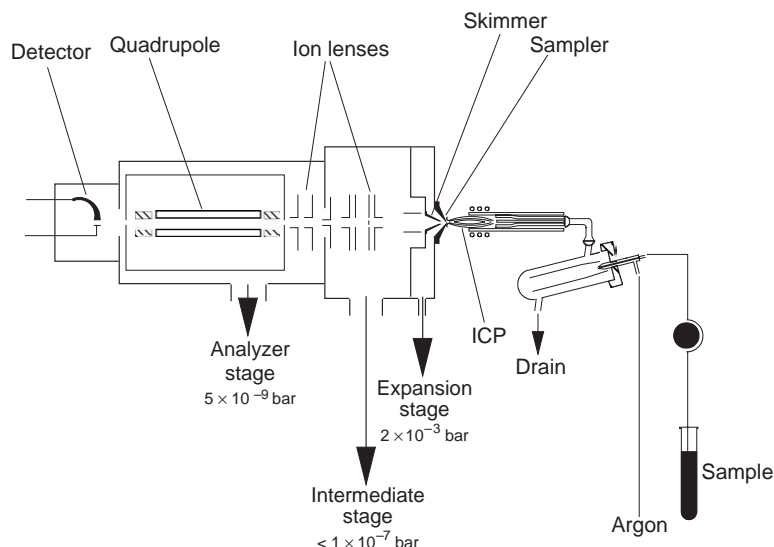
A schematic diagram of an ICP-MS instrument is shown in **Figure 1**. The ICP component is very similar to the ICP used for atomic emission spectrometry

and, in the recent past, sample introduction systems, radiofrequency (RF) generators, and tuning networks were often the same for ICP-MS and ICP-atomic emission spectrometry systems, though more specialization has recently been introduced.

### Instrumentation

#### The Inductively Coupled Plasma

The ICP is generated by coupling the energy from a RF generator into a suitable gas via a magnetic field that is induced through a two or three turn, water-cooled copper coil. The RF energy is normally supplied at a frequency of 27.12 MHz, delivering forward power at between 500 and 2000 W. Two gas flows, usually argon, flow in a tangential manner through the outer tubes of a concentric, three-tube quartz torch that is placed axially in the copper coil. The outer and intermediate gases flow tangentially (i.e., they swirl around as they pass through the torch), so the plasma is continually revolving and has a 'weak spot' at the center of its base, through which the inner gas flow, containing the sample, can be introduced. A spark is used to seed the gas with electrons, which then accelerate in the magnetic field and reach energies sufficient to ionize gaseous atoms in the



**Figure 1** Schematic diagram of a commercial inductively coupled plasma mass spectrometer.

field. Subsequent collisions with other gaseous atoms cause further ionization, and so on to form a self-sustaining plasma. This occurs almost instantaneously. The magnetic field causes the ions and electrons to flow in the horizontal plane of the coil, thereby heating neutral argon by collisional energy exchange, and a hot fireball is produced. The hottest part of the ICP has a temperature between 8000 and 10 000 K, which is the temperature of the surface of the Sun, though the analytically useful region is in the tail-flame with a temperature between 5000 and 6000 K.

In the absence of analyte atoms, water, and sample matrix components, the predominant species in an argon ICP will be Ar, Ar<sup>+</sup>, and e<sup>-</sup>, though other species can be important when considering analyte ionization mechanisms. The RF energy used to sustain the plasma is only coupled into the outer region of the plasma, so these species are primarily formed in this region and are thermally transferred to the center and circulate between the outer and central regions. Additionally, when the sample is introduced through the axial channel, the center will contain a flow of cooler gas containing the analyte and species derived from the sample matrix and water.

### Sample Introduction

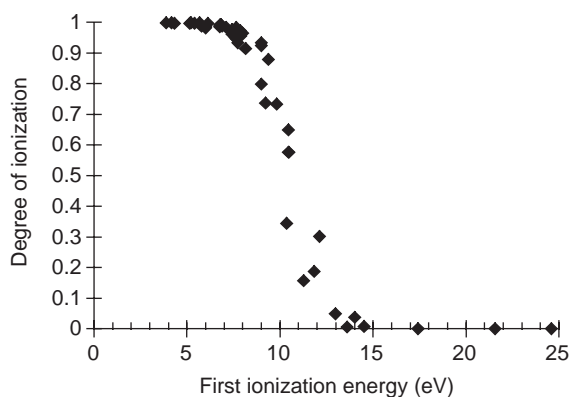
The commonest form of sample introduction is by means of an aerosol generated using a pneumatic nebulizer. The most commonly used nebulizer is a glass concentric type; however other types, such as cross-flow, ultrasonic, and v-groove, have also been used successfully. Aerosols generated by nebulization are directed through a spray chamber, which is usually constructed from glass, quartz, or an inert polymer. The spray chamber prevents larger aerosol droplets from reaching the plasma, which would otherwise cause flicker and consequent imprecision, and generally the bulk of droplets are of the order of 2–5 µm in diameter when the aerosol exits the spray chamber. One consequence of this is that the sample transport system is inefficient, of the order of 1–15% efficiency, depending on the type of nebulizer used. On exiting the spray chamber the gas stream containing the aerosol is directed through the injector tube, which forms part of the quartz torch, and thence into the base of the ICP. The injector tube has an internal diameter of 2 mm or so, so the 0.6–1.0 l min<sup>-1</sup> of gas flow exiting it has sufficient velocity to punch a hole through the center of the ICP, thereby forming an annular or doughnut shaped plasma. For ICP-MS, the sample must usually contain less than 0.1% dissolved solids to prevent salt build-up on the sampler and skimmer cones.

Once in the ICP, the aerosol droplets are successively desolvated, decomposed, and ionized by the

high temperature of the plasma. If an electron absorbs sufficient energy, equal to its first ionization energy, it escapes the atomic nucleus and an ion is formed. In the ICP the major mechanism by which ionization occurs is thermal ionization. When a system is in thermal equilibrium, the degree of ionization of an atom is given by the Saha equation:

$$\frac{n_i n_e}{n_a} = 2 \frac{Z_i}{Z_a} \left( \frac{2\pi m k T}{h^2} \right)^{3/2} \exp\left(-\frac{E_i}{kT}\right) \quad [1]$$

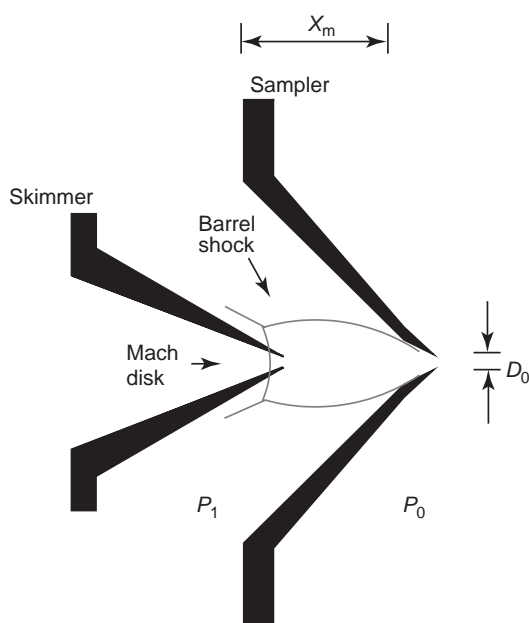
where  $n_i$ ,  $n_e$ , and  $n_a$  are the number densities of the ions, free electrons, and atoms, respectively;  $Z_i$  and  $Z_a$  are the ionic and atomic partition functions, respectively;  $m$  is the electron mass;  $k$  is the Boltzmann constant;  $T$  is the temperature;  $h$  is Planck's constant, and  $E_i$  is the first ionization energy. In this case, ionization is effected by ion-atom and atom-atom collisions, where the energy required for ionization is derived from thermal agitation of the particles. The degree of ionization is dependent on the electron number density, the temperature, and the ionization energy of the element in question. Taking the average electron number density for an argon ICP to be  $4 \times 10^{15} \text{ cm}^{-3}$  and the ionization temperature to be 8730 K, the degree of ionization as a function of first ionization energy, predicted by eqn [1], is shown in Figure 2. Most of the elements in the periodic table have first ionization energies of less than 9 eV and are over 80% ionized in the ICP. The remaining third are ionized to a lesser extent depending on their first ionization energy, with the most poorly ionized elements being: He, Ne, F, O, N <1% ionized; Kr, Cl <10%; C, Br, Xe, S <30%; P, I, Hg, As, Au, Pt <80%. Such thermal ionization is probably the dominant mechanism of ionization in the ICP, and predominantly forms singly charged ions, which are ideal for analysis by MS.



**Figure 2** Degree of ionization as a function of first ionization energy calculated using the Saha equation.

## Ion Sampling

In order to perform MS on the ions formed in the plasma they must be extracted into the mass spectrometer, which is required to be at extremely low pressure, so the sampling interface must be in direct contact with the plasma. The problem of extracting ions from an extremely hot plasma at atmospheric pressure into a mass spectrometer at  $\sim 10^{-9}$  atm is overcome by making use of a series of differentially pumped vacuum chambers held at consecutively lower pressures. A schematic diagram of the ICP-MS sampling interface is shown in Figure 3. The ICP is aligned so that the central channel is axial with the tip of a water-cooled, sampling cone, typically made of nickel or copper, which has an orifice of  $\sim 1$  mm in diameter. The pressure behind the sampling cone is reduced, by means of a vacuum pump, to  $\sim 2 \times 10^{-3}$  atm so the plasma gases, together with the analyte ions, expand through the sampling orifice to form a shock-wave structure as shown in Figure 3. This expansion is isentropic (i.e., no change in the total entropy) and adiabatic (i.e., there is no transfer of energy as heat), resulting in a supersonic expansion accompanied by a fall in temperature. This supersonic expansion takes the form of a cone with a shock-wave structure at its base called a Mach disk. The region within the expansion cone is called the 'zone of silence', which is representative of the ion species to be found in the ICP, i.e., the ionization conditions have been 'frozen'.



**Figure 3** Schematic of the ICP-MS interface showing the supersonic expansion formed in the expansion chamber, the barrel shock, and position of the Mach disk.

The skimmer cone is another metal cone, the tip of which has an orifice of  $\sim 0.7$  mm in diameter, which protrudes into the zone of silence, and is axially in line with the sampling orifice as shown in Figure 3. The ions from the zone of silence pass through the orifice in the skimmer cone, into a second intermediate vacuum chamber held at  $< 10^{-7}$  atm, as an ion beam. The ion beam can then be focused by means of a series of ion lenses, which deflect the ions along a narrow path and focus them onto the entrance to the mass analyzer.

## Mass Analysis

Mass analysis is simply a method of separating ions of different mass-to-charge ratio ( $m/z$ ). However, since the ions of interest are almost exclusively singly charged the  $m/z$  is equivalent to mass for most practical purposes. There are three types of mass analyzer used for ICP-MS, quadrupole, time-of-flight (TOF), and magnetic sector.

**Quadrupole ICP-MS** Quadrupoles are typically operated in the mass range from 2 to 260  $m/z$  by scanning through the mass range sequentially. Either the whole mass range or selected masses can be scanned; and the speed of the scan (100 ms to scan from 2 to 260  $m/z$ ) makes it seem almost like simultaneous mass analysis. The quadrupole mass analyzer has the advantage of being cheap, reliable, and compact, with single mass resolution which is sufficient for most applications, and is therefore the most commonly used mass analyzer. However, if an extremely high degree of resolution or true simultaneous mass analysis is required, then other types of mass spectrometer must be used.

**Time-of-flight ICP-MS** In quadrupole and sector field mass analyzers, the ion signal is a continuous beam; however, in TOF-MS the ion beam is pulsed so that the ions are either formed or introduced into the analyzer in 'packets'. These ion packets are introduced into the field-free region of a flight tube 30–100 cm long. The principle behind TOF analysis is that, if all ions are accelerated to the same kinetic energy, each ion will acquire a characteristic velocity dependent on its  $m/z$  ratio. The ion beam reaches its drift energy (2700 eV) in less than 2 cm. The ions are then accelerated down the TOF tube with whatever velocity they have acquired. Because all ions have essentially the same energy at this point, their velocities are inversely proportional to the square roots of their masses. As a result, ions of different mass travel down the flight tube at different speeds, thereby separating spatially along the flight tube with



lighter, faster, ions reaching the detector before the heavier ions.

While simple in theory, the TOF analyzer caused numerous problems when coupled with a plasma source such as an ICP. The spread in ion kinetic energies caused by the ion-sampling process results in the ions entering the field-free region of the flight tube at different angles. Another difficulty with ICP-TOF-MS is that ions have to be introduced in 'packets'. This can be achieved, for example, by using an orthogonal interface with a pulsed repeller plate. Background noise can be reduced by using a combination of quadrupole ion optics and an energy discriminator before the detector.

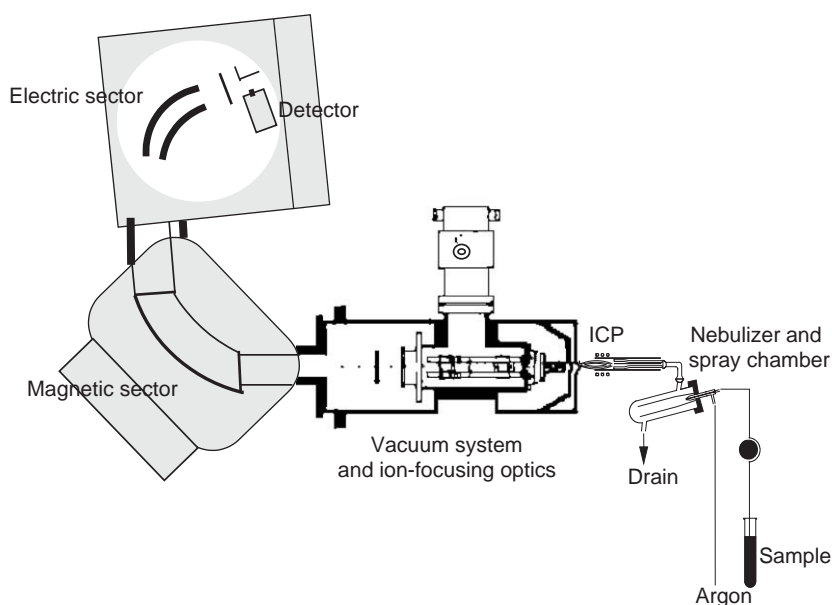
**Magnetic sector ICP-MS** Magnetic sector mass analyzers rely on the fact that ions are deflected by a magnetic field, with ions of greater mass or charge being deflected to a greater extent. However, ions produced in the ICP will have varying energies, 10–20 eV, depending on their point of formation. Magnetic sector mass analyzers require a high-energy (3–8 kV) ion beam for effective ion transmission and resolution, so the ions must be accelerated in a high-voltage field which will accentuate any differences in ion energy. If a magnetic sector is used on its own (single focusing) the difference in ion energy leads to peak broadening and low resolution. In order to overcome this, a combination of an electric and magnetic sector (double focusing), or a hexapole and magnetic sector, can be used.

Double focusing sector field (SF) instruments use electric and magnetic fields to disperse ions according to their momentum and translational energy. The

electric sector is usually (though not exclusively) placed before the magnetic sector, and acts as an energy focusing device, which transmits a narrow band of ion energies, which are then separated by the magnetic sector. A schematic of a double-focusing SF-ICP-MS instrument is shown in Figure 4.

The problem of variable ion energies can also be solved by placing a hexapole collision cell in front of the magnetic sector, into which a small amount of argon gas is injected. Collisions with the argon gas thermalizes the ion beam, reducing the ion energy spread to <1 V. This means that the simpler, single focusing magnetic sector geometry can be used. The ions are then accelerated to 6 kV before entering the magnetic sector.

The main advantage of magnetic sector analyzers is their vastly superior resolution compared with quadrupole instruments. The resolution of a mass analyzer can be expressed as ( $R = M/\Delta M$ ), where  $R$  is the resolution,  $M$  is the mass of the isotope of interest, and  $\Delta M$  is the peak width of the isotope at 5% peak height. The increased resolution is advantageous because polyatomic ion interferences can be resolved from analyte isotopes of interest. Table 1 shows a number of polyatomic ion interferences and the resolution required to separate them from the elemental isotope of interest. Most quadrupole and TOF mass analyzers operate with an upper resolution of 400, which enables unit mass resolution, while magnetic sector instruments for ICP-MS have been operated up to a resolution of 10 000 enabling peaks of fractions of a mass unit to be resolved.



**Figure 4** Schematic diagram of a double-focusing SF-ICP-MS instrument.

**Table 1** Common polyatomic ion interferences with the mass analyzer resolution necessary to resolve them from the analyte of interest

Analyte ion		Interfering ion		
Nominal m/z	Accurate m/z	Nominal m/z	Accurate m/z	Resolution required
<sup>24</sup> Mg	23.9850	<sup>12</sup> C <sub>2</sub>	24.0000	1599
<sup>28</sup> Si	27.9769	<sup>14</sup> N <sub>2</sub>	28.0060	962
<sup>28</sup> Si	27.9769	<sup>12</sup> C, <sup>16</sup> O	27.9949	1555
<sup>31</sup> P	30.9737	<sup>14</sup> N, <sup>16</sup> O, <sup>1</sup> H	31.0057	968
<sup>31</sup> P	30.9737	<sup>15</sup> N, <sup>16</sup> O	30.9950	1455
<sup>32</sup> S	31.9721	<sup>16</sup> O <sub>2</sub>	31.9898	1807
<sup>44</sup> Ca	43.9555	<sup>12</sup> C, <sup>16</sup> O <sub>2</sub>	43.9898	1282
<sup>48</sup> Ti	47.9479	<sup>32</sup> S, <sup>16</sup> O	47.9670	2511
<sup>51</sup> V	50.9440	<sup>35</sup> Cl, <sup>16</sup> O	50.9637	2586
<sup>52</sup> Cr	51.9405	<sup>35</sup> Cl, <sup>16</sup> O, <sup>1</sup> H	51.9715	1676
<sup>52</sup> Cr	51.9405	<sup>40</sup> Ar, <sup>12</sup> C	51.9623	2383
<sup>54</sup> Fe	53.9396	<sup>40</sup> Ar, <sup>14</sup> N	53.9653	2099
<sup>56</sup> Fe	55.9349	<sup>40</sup> Ar, <sup>16</sup> O	55.9572	2509
<sup>63</sup> Cu	62.9296	<sup>40</sup> Ar, <sup>23</sup> Na	62.9521	2797
<sup>64</sup> Zn	63.9291	<sup>32</sup> S, <sup>16</sup> O <sub>2</sub>	63.9619	1950
<sup>64</sup> Zn	63.9291	<sup>32</sup> S <sub>2</sub>	63.9442	4234
<sup>75</sup> As	74.9216	<sup>40</sup> Ar, <sup>35</sup> Cl	74.9311	7887
<sup>80</sup> Se	79.9165	<sup>40</sup> Ar <sub>2</sub>	79.9246	9867

Another advantage associated with focusing instruments is that ions of different mass are spatially separated on exiting the analyzer. This fact has been exploited by a number of manufacturers who have developed 'multidetector' instruments. An array of detectors is placed downstream of the analyzer, allowing up to 10 ions to be detected simultaneously. This has the advantage of reducing imprecision caused by temporal fluctuations in the ion beam, thereby improving the precision of isotope ratio measurements to better than 0.01% RSD.

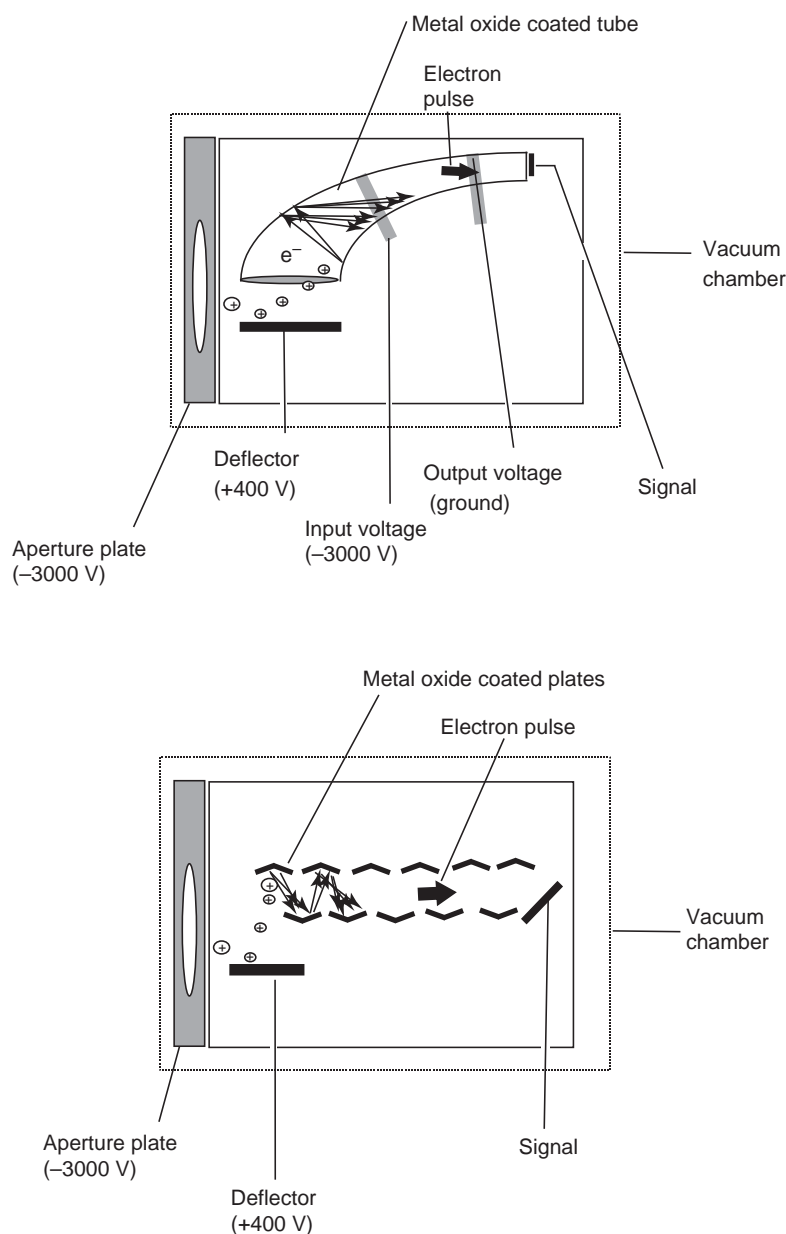
### Ion Detection and Signal Handling

**Electron multiplier** This is the commonest type of detector, and is used to detect ion currents of less than  $10^{-15}$  A. At higher currents a Faraday cup detector is normally used (see below). When the ion beam exits the mass analyzer it strikes a conversion plate that converts ions into electrons. The ions are drawn toward the plate by a strong voltage applied to the conversion plate. On striking the conversion plate the ions stimulate the ejection of electrons, which are accelerated by the voltage applied to the plate. The electrons are multiplied using either continuous or discrete dynodes. A continuous dynode channel electron multiplier is shown in **Figure 5**. This consists of a curved glass tube of  $\sim 1$  mm in internal diameter with an inner resistive coating and a flared end. The multiplier can be operated in one of two modes. In pulse counting mode – the most

sensitive mode of operation – a high voltage of between  $-2600$  and  $-3500$  V is applied to the multiplier, which attracts ions into the funnel opening. When a positive ion strikes the inner coating the collision results in the ejection of one or more secondary electrons from the surface, which are accelerated down the tube by the potential gradient and collide with the wall, resulting in further electron ejection. Hence, an exponential cascade of electrons rapidly builds up along the length of the tube, eventually reaching saturation toward the end of the tube, resulting in a large electron pulse and a consequent gain of  $10^7$  to  $10^8$  over the original ion collision. The electron pulses are read at the base of the multiplier and are approximately 50–100 mV and 10 ns in duration. Alternatively, the multiplier can be operated in analog mode with a gain of only  $10^3$ – $10^4$  so that the multiplier does not become saturated and the pulses vary greatly in size. In this mode the applied voltage is between  $-500$  and  $-1500$  V and the electron pulses are read at the collector electrode where they are amplified and averaged over a short time interval to allow rapid data acquisition. The greatest sensitivity is achieved with the detector in pulse counting mode, but the detector will become saturated at counting rates above  $10^6$  Hz, which are encountered when the analyte is at a high concentration in the sample. If the detector is switched into analog mode it is less sensitive, but can be used for analyte concentrations that are much higher, typically up to three orders of magnitude higher than for pulse counting. Such dual-mode operation results in an extremely large linear dynamic range of up to nine orders of magnitude. A discrete dynode detector consists of an array of discrete dynode multipliers, usually containing 15–18 dynodes, coated with a metal oxide that has high secondary electron emission properties. The dynodes are placed in one of two configurations, either Venetian blind or box and grid fashion. Secondary electrons emitted by the metal oxide are forced to follow a circular path by a magnetic field, so they strike successive dynodes, thereby multiplying the signal.

**Faraday cup** The Faraday cup detector is used for the detection of much higher ion fluxes than the pulse counting detector. It consists of a collector electrode that is surrounded by a cage. The electrode is positioned at an angle in respect to the ion beam, so that ions exiting the analyzer strike the electrode but secondary emissions are reflected away from the detector entrance. The function of the cage is to prevent detected ions and secondary electrons escaping from the detector. The electrode is connected to the ground via a resistor. The ion current striking the electrode is





**Figure 5** Schematic diagram of an electron multiplier showing: (top) continuous; (bottom) discrete dynodes.

neutralized by electron flow from the ground through the resistor. This causes a potential drop which is amplified to create a signal. Currents as low as  $10^{-15}$  A have been successfully measured in this way.

## Performance

Performance characteristics for quadrupole and magnetic sector instruments are shown in Table 2 and Figure 6. The main advantages that ICP-MS has over other techniques are: low detection limits, in the  $0.1\text{--}100\text{ pg ml}^{-1}$  range for quadrupole instruments (Figure 6), with magnetic sector instruments a factor of  $10\text{--}100$  lower; large linear dynamic

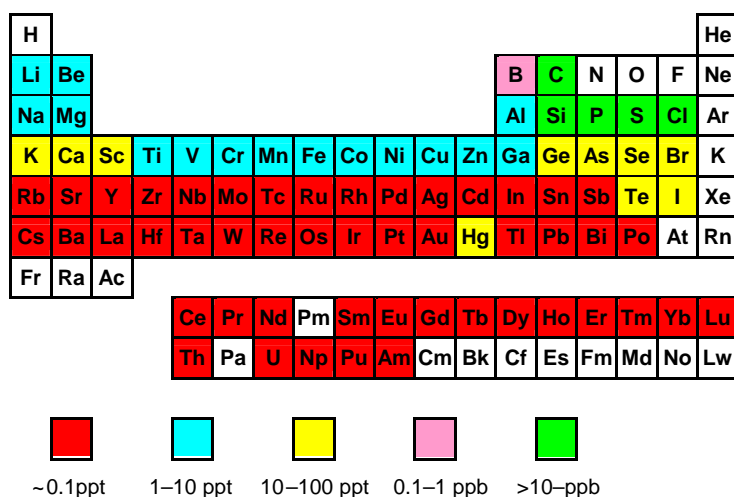
**Table 2** Performance characteristics for ICP-MS

	Quadrupole	Magnetic sector
Detection limit ( $\text{pg ml}^{-1}$ )	0.1–100	0.001–0.1 <sup>a</sup> 0.1–100 <sup>b</sup>
Linear dynamic range	$10^6$ $10^9$ dual mode	$10^6$ $10^{10}$ dual mode
Precision (% RSD)	1–2	1–2
Mass resolution	400	400–10 000

<sup>a</sup>Resolution 400.

<sup>b</sup>Resolution 3500.

range ( $10^6\text{--}10^{10}$ ); and rapid multielement capability. However, ICP-MS also suffers from a number of interferences.



**Figure 6** Typical detection limits for quadrupole ICP-MS. Those achievable with magnetic sector ICP-MS are between 10 and 100 times lower.

## Interferences

ICP-MS suffers from two main types of interference, spectroscopic and nonspectroscopic. Spectroscopic interferences arise when an interfering species has the same nominal  $m/z$  as the analyte of interest. The interfering species can be either an isotope of another element (which are well documented and hence easily accounted for), or a molecular ion formed between elements in the sample matrix, plasma gas, water, and entrained atmospheric gases. The molecular ions are less easy to correct for since they will vary depending on the nature of the sample matrix. Some common molecular ion interferences are shown in **Table 1**. Many of these interferences can be overcome by choosing an alternative, interference-free isotope of the analyte, though a sacrifice in sensitivity may result. If a 'clean' isotope is not available then one solution is to separate the analyte from the matrix before the analysis using chemical extraction and chromatography, or use a magnetic sector instrument that is capable of resolving the interfering species from the analyte. Many of these interferences are thought to be formed in the interface due to a secondary discharge. This discharge can be eliminated by operating the plasma at low power, typically 600 W, and modifying the torch by inserting a grounded shield between it and the coil, or by using a center-tapped RF coil. Under these so-called 'cool plasma' conditions interferences due to  $\text{ArO}^+$ , for example, can be eliminated, though the sensitivity for refractory elements and those with high ionization potentials may be reduced. An alternative method of reducing molecular ions is to use a reaction cell prior to the quadrupole. A gas, such as helium, ammonia, or water vapor, is introduced into the cell where it

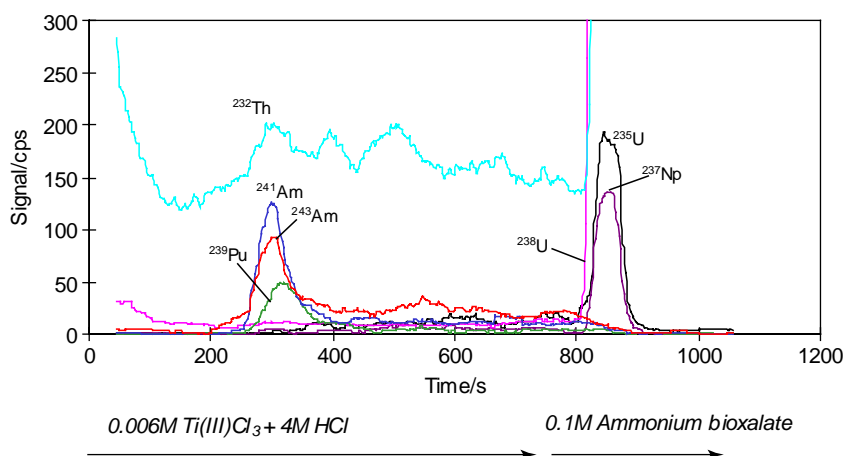
either reacts with, collisionally dissociates, or neutralizes the polyatomic species or precursors.

Nonspectroscopic interferences are caused by the sample matrix, and are manifest as an apparent enhancement or suppression in the analyte signal in the presence of a concentrated sample matrix. Such effects are thought to be caused primarily by space charge in the ion beam, whereby positive analyte ions are repelled from the ion beam by the high positive charge of the matrix ions, with low mass ions being relatively more affected than high mass ions. Such interferences are usually compensated for by using an internal standard, or by separating the matrix from the analyte before analysis. ICP-MS has very low detection limits so it is sometimes possible to dilute the sample to such an extent that the interference becomes negligible.

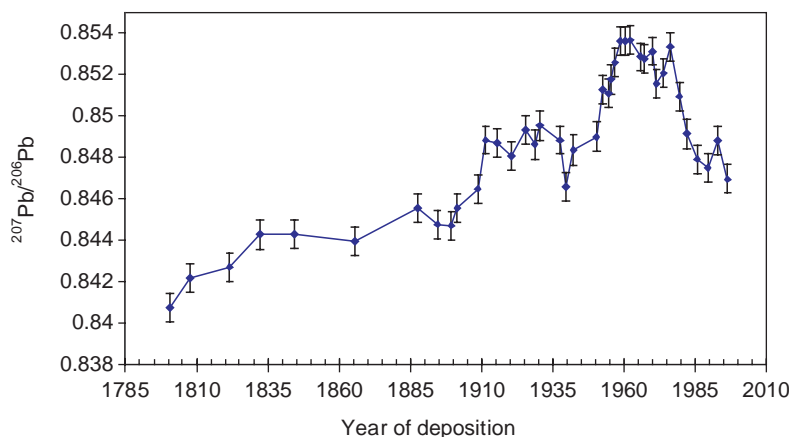
## Applications

Applications of ICP-MS cover a wide range of sample types which often make use of its excellent sensitivity and isotope ratio capabilities, such as the determination of ultra-low levels of impurities in semiconductors, long-lived radionuclides in the environment, and geochronology. ICP-MS is well suited to the determination of the lanthanide series of elements in many geological applications. Sample preparation methods are similar to those generally used for trace metals analysis; however, nitric acid is favored for sample digestion because other mineral acids contain elements which cause spectroscopic interferences.

The capability for rapid multielement analysis makes ICP-MS particularly suited to sample



**Figure 7** Sequential elution of  $\sim 100$  fg of each of the actinides using SF-ICP-MS detection. (Adapted from Truscott J, Jones P, Fairman B, and Evans EH (2001) *Analytica Chimica Acta* 433: 245–253 with permission.)



**Figure 8** Variation in  $^{207}\text{Pb}/^{206}\text{Pb}$  isotope ratio with depth in a salt-marsh core, measured using multicollector SF-ICP-MS. The error bars represent the expanded uncertainty ( $K=2$ ) of the isotope ratio measurements. (Adapted from Gehrels WR, Newnham RM, Kirby JR, *et al.* (2001) High-resolution reconstruction of sea-level change during the past 300 years: Geological Society of America Abstracts with Programs, vol. 33, p. A40.)

introduction methods that give rise to transient signals. This means that electrothermal vaporization, flow injection, and chromatographic methods can be interfaced, and many elements monitored in a single run. An example of this, combined with the extremely low detection limits obtainable with a sector field instrument, is shown in **Figure 7**. The actinide elements have been separated from the sample matrix, and partially separated from each other, using TRU-Spec<sup>TM</sup> resin and the isotopes  $^{232}\text{Th}$ ,  $^{237}\text{Np}$ ,  $^{238}\text{U}$ ,  $^{239}\text{Pu}$ ,  $^{240}\text{Pu}$ ,  $^{241}\text{Am}$ ,  $^{243}\text{Am}$  detected in a single run. Detection limits for the transuranic elements were of the order of 2 fg per g in solution.

A major attraction is the ability to perform isotope ratio measurements, e.g., in many geological applications to determine the age of rocks, biological and geological fractionation of elements, anthropogenic origin, stable isotope tracer studies, and isotope dilution analysis. In this respect, magnetic sector

ICP-MS equipped with simultaneous detection is capable of extremely good isotope ratio precision, of the order of 0.01% RSD, and the technique is supplanting thermal ionization MS as the method of choice in many applications because of its much higher throughput. **Figure 8** shows the variation in the  $^{207}\text{Pb}/^{206}\text{Pb}$  ratio in a salt-marsh core, taken at Chezzetcook, Nova Scotia, measured using multicollector sector-field ICP-MS. The ratio varies between 0.8407 and 0.8536 reflecting the changing influences of the underlying mineralogy and anthropogenic inputs. Note the peak in the early part of the twentieth century due to increased industrialization, and between 1955 and 1975, corresponding to increased automobile usage and leaded gasoline, until it was phased out when the ratio fell again.

See also: **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Atomic Spectrometry:** Overview.

**Isotope Dilution Analysis. Isotope Ratio Measurements. Mass Spectrometry:** Overview.

## Further Reading

Ebdon L, Evans EH, Fisher A, and Hill S (eds.) (1998) *An Introduction to Analytical Atomic Spectrometry*. Chichester: Wiley.

Evans EH, Giglio JJ, Castellano T, and Caruso JA (1995) *Inductively Coupled and Microwave Induced Plasma*

*Sources for Mass Spectrometry*. Cambridge: Royal Society of Chemistry.

Hill SJ (ed.) (1998) *ICP Spectrometry and its Applications*. Sheffield, UK: Sheffield Academic Press.

Holland G and Tanner SD (eds.) (2001) *Plasma Source Mass Spectrometry: The New Millennium*. Cambridge: Royal Society of Chemistry.

Montaser A (ed.) (1998) *Inductively Coupled Plasma Mass Spectrometry*. Hoboken, NJ: Wiley-VCH.

## Laser Microprobe

**L Van Vaeck**, University of Antwerp, Antwerp, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The defining attribute of laser microprobe mass spectrometry (LMMS) is the use of a pulsed ultraviolet (UV) laser focused on a  $<5\ \mu\text{m}$  diameter spot for one-step desorption-ionization (DI) of components of a local microvolume in a solid. Subsequent separation of the ions according to their mass-to-charge ratio ( $m/z$ ) is achieved using a time-of-flight (TOF) or Fourier transform (FT) mass analyzer, which provides a low or high mass resolution, respectively. The alternative acronyms used are laser probe microanalysis (LPM or LPMA), laser ablation microprobe mass spectrometry (LAMMS), laser ionization mass analysis (LIMA), and laser microprobe mass analysis (LAMMA). A recent development is 'aerosol TOF-mass spectrometry' (aTOF-MS), dedicated to analysis of suspended single particles in the micrometer-size range.

Material properties often depend on the constituents of inclusions as small as a few micrometers large or surface layers with a thickness of 0.1–100 nm. From the 1970s onward, microanalysis has become a major challenge. Initial methods such as electron probe X-ray microanalysis (EPXMA) or dynamic secondary ion mass spectrometry (SIMS) determine elements at the microscopic scale, but charge build-up and beam-induced sample damage hamper their application to nonconducting organic materials.

The use of photons conceptually overcame these problems, but the development of LMMS had to wait for reliable high-power pulsed UV lasers and sufficiently fast electronics, required for panoramic registration of ions from a single laser shot in TOF-MS. Although TOF-LMMS originally aimed at elemental analysis in dielectric materials, its major

strength soon turned out to be the generation of molecular (adduct) ions for molecular weight (MW) determination, together with structural fragments for characterization of functional groups in organic as well as inorganic analytes.

The successful application of TOF-LMMS to a variety of material science problems has provided evidence of the potential of laser microbeam irradiation, in particular when a high mass resolution is available. Fundamental research on DI processes has led to their successful coupling with FT-MS. The resulting FT-LMMS methodology remains unique in combining a microanalytical detection sensitivity with the ultimate identification power and specificity of a mass resolution value greater than 100 000 and a mass accuracy within 1 ppm. Under favorable conditions, monolayer detection is feasible. Although a laser shot creates craters as deep as 0.1–1  $\mu\text{m}$ , the detected ions come from the upper 10–50 nm surface layer. Unlike static secondary ion mass spectrometry (S-SIMS), LMMS can be used to probe the composition of the upper subsurface in the case of accidental contamination. In addition, its applicability to nonconducting thick samples facilitates material research.

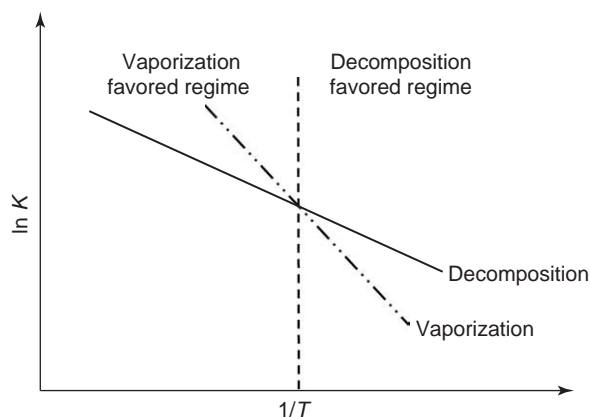
## Fundamentals

Despite the proliferating use of lasers in materials processing, the physics of DI process has not been fully understood yet. Analytical use of the methodology requires chemists to develop practical concepts rationalizing the way that detected ions are formed and reflecting the molecular composition of the sample.

Various models attempt to explain the apparent contradiction between the application of destructive power density conditions to thermolabile compounds and the detection of intact molecular (adduct) ions. Most of the concepts therein share

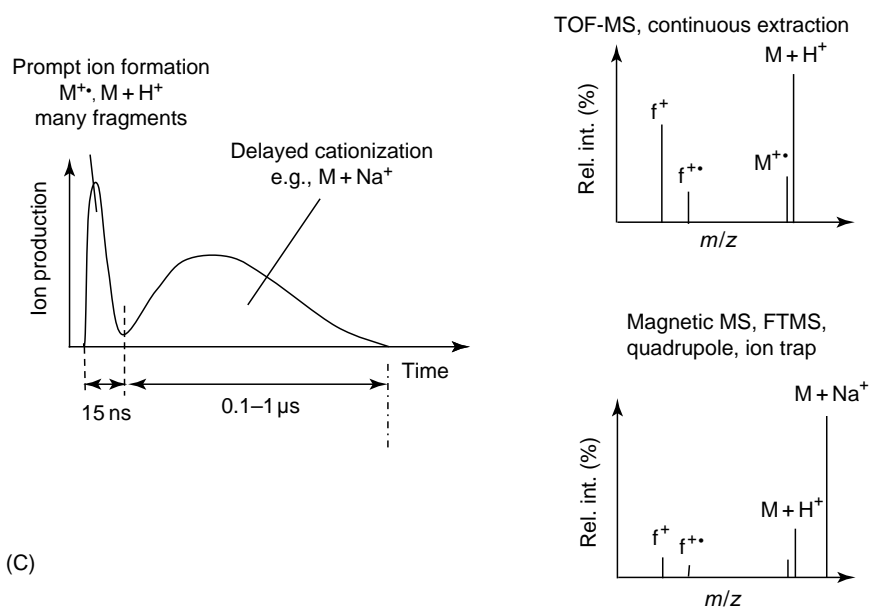
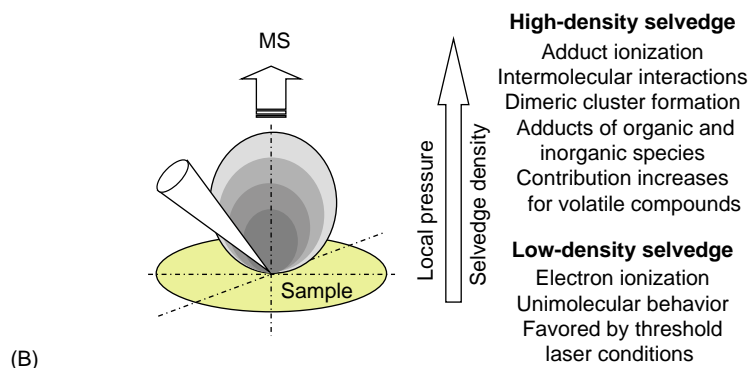
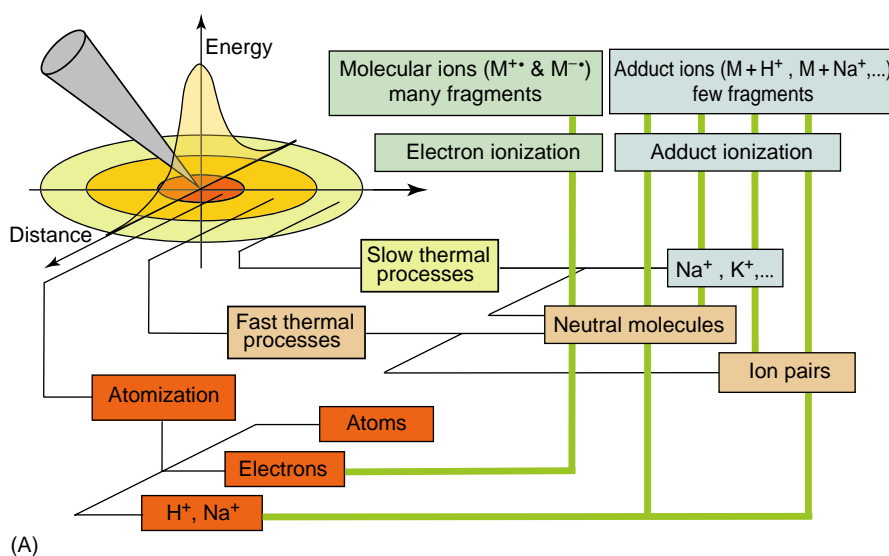
the assumption of a nonthermal process involving mechanisms such as direct ejection of (ionized) species through shock waves, ablation resulting from a nonlinear volume absorption of laser energy by the subsurface, 'solid-state chemical ionization', and photoionization of species already desorbed in the seldge (i.e., dense gas phase above the sample). However, it is often overlooked that thermally driven processes can indeed transfer labile compounds from solids to the gas phase without decomposition. **Figure 1** depicts Arrhenius plots for the vaporization of a thermolabile analyte without and with thermal decomposition. Unless the activation energy of the two routes is identical, the different slopes of the corresponding lines cause their intersection at a given temperature. Hence, there is a temperature range where vaporization prevails over decomposition. Conventionally, a compound is called thermolabile when conductive heating cannot attain this critical temperature before complete degradation of all available analytes occurs. However, laser irradiation of solids leads to a heating rate up to  $10^{10} \text{ K s}^{-1}$ , which brings the sample rapidly into the regime where vaporization prevails over decomposition.

A systematic study of inorganic and organic polyfunctional molecules combined with research on ion formation has allowed us to elaborate a tentative DI model, permitting the formation of detected ions to be rationalized and even predict those of a given analyte structure. The approach illustrated in **Figure 2** basically considers the effects to be expected from the three prime parameters in MS, namely energy, pressure, and time.



**Figure 1** Arrhenius plots describing the temperature dependence of vaporization and thermal decomposition of a thermolabile compound. (Adapted from Daves G (1979) Mass spectrometry of involatile and thermally unstable molecules. *Accounts of Chemical Research* 12: 359–365; American Chemical Society.)

The detection of atomic ions together with molecular (adduct) ions from fragile analytes is linked to the energy gradient created along the surface (**Figure 2A**). A laser impact leads to a combination of processes such as atomization, generation of free electrons, destructive pyrolysis, desorption of intact molecules or ion pairs, and thermionic emission of e.g., alkali ions. Unlike the often-advocated direct ejection of ions, the DI model explicitly considers the initial generation of neutrals, in agreement with the principles of molecular physics. Ionization occurs in the seldge by electron ionization (EI) or adduct ionization (AI). The former process, driven by laser-ejected electrons, generates radical molecular ions with sufficient internal stress to produce numerous fragments. In contrast, ion–molecule interactions between thermal(ized) seldge species form adduct ions. Their fragmentation is restricted to elimination of small molecules. Unlike other approaches that only consider the generation of even-electron parents, the DI model logically explains the detection of radical ions in LMMS. Also the  $[\text{M} - \text{CH}_2]^+$  signals from quaternary methyl ammonium salts are readily linked to the sequence of thermal degradation and AI. The pressure gradient in the seldge is assumed to govern the relative importance of EI and AI (**Figure 2B**). As the density of the neutrals determines the probability of thermalization and ion–molecule interactions, analytes that are readily desorbed with a low laser power density tend to form dimeric and trimeric adducts (e.g.,  $[\text{2M} + \text{Na}]^+$ ), while 'volatile' compounds only show cationization or protonation of single molecules. Finally, the match between the time domains of ionization and mass analysis is considered (**Figure 2C**). This key feature of the DI model makes it applicable to organic and inorganic mass spectra in LMMS, laser desorption with spot diameters of 0.1–1 mm, matrix assisted laser desorption, and S-SIMS. Basically, a given analyte can yield different mass spectra under identical irradiation, depending on the use of a magnetic sector, quadrupole, or FT- or TOF-analyzer. The reason is the bimodal ion production. Specifically, a first ion bunch is generated within 15–25 ns (essentially during the laser pulse), but seldge ionization continues for microseconds. On the other hand, a TOF-analyzer separates only ions if they enter the drift within 10–25 ns. Hence, TOF-LMMS is limited to the first 'prompt' ion contribution, while an FT-analyzer accepts the entire ion production. The existence of a fast and slow ionization process is experimentally supported by kinetic energy ( $E_{\text{kin}}$ ) and emission angle data as well as delayed ion extraction experiments in FT-LMMS performed with external ion sources.



**Figure 2** Rationalization of the detected ions using the DI model in LMMS by the effects of the (A) energy gradient created along the surface, (B) the pressure gradient in the selfedge, and (C) the time domain of ion formation and mass analysis. (Adapted from Van Vaeck L, Struyf H, Van Roy W, and Adams F (1994) Organic and inorganic analysis with laser microprobe mass spectrometry. Part I: instrumentation and methodology. *Mass Spectrometry Reviews* 13: 189–208; Wiley.)

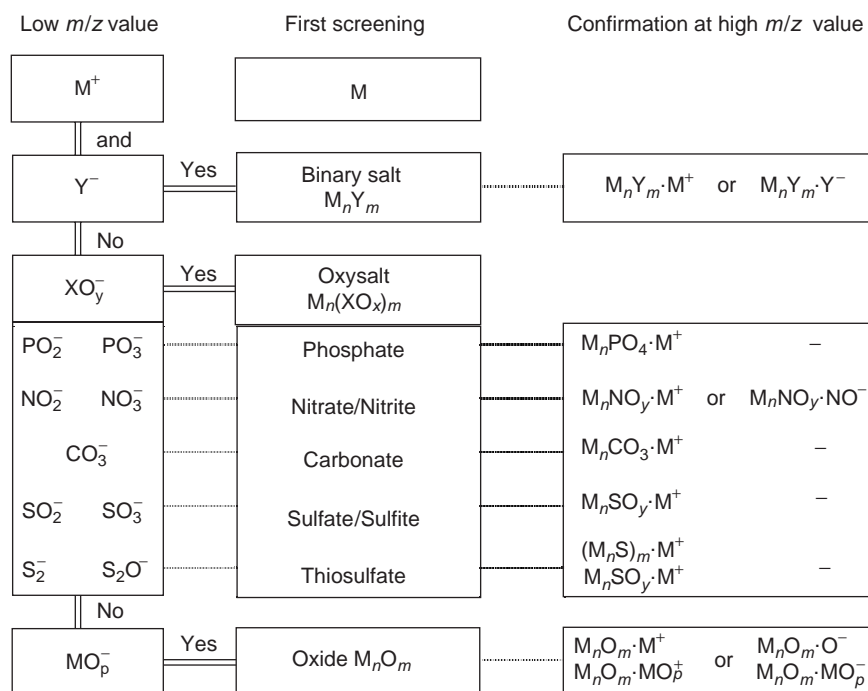
## Diagnostic Use of Mass Spectra

Analytical applications are often based on fingerprinting, i.e., comparison with reference spectra without detailed structural assignment of the ions. This approach fails to exploit the real strength of LMMS, namely the identification of unknown or unexpected compounds. The high mass resolution and accuracy of FT-LMMS are often essential for verifying the elemental composition of detected ions. Deductive identification is a major advantage in e.g., industrial problem solving because many steps, (technical grade) reagents, and various contaminants can be involved. The feasibility of pinpointing one or a few likely causes for the anomaly simply by looking at the mass spectrum readily pays back the cost of an LMMS analysis.

Nowadays, solid-state speciation or molecular identification of inorganic analytes is a hot topic in analytical chemistry. Molecular speciation allows individual components to be identified in mixtures, an impossible task for element or functional group detection methods. In fact, LMMS has been a forerunner in this field. **Figure 3** illustrates the speciation scheme used in FT-LMMS. Basically, the  $m/z$  difference between prominent signals from atomic ions and their adducts, the isotope patterns, and structural fragments such as oxide (adduct) ions in the case of oxysalts readily allow the molecular composition

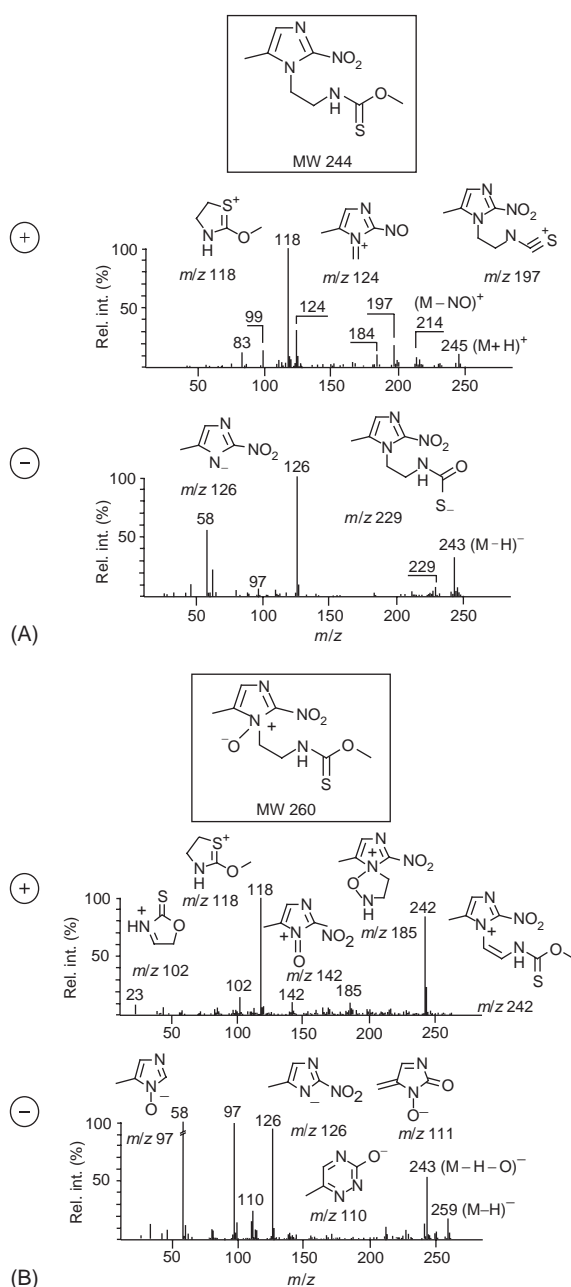
of the analyte to be determined, even at a low mass resolution level. However, FT-LMMS remains of interest for resolving isobars in multicomponent systems, while sampling of the entire ion production increases the relative importance of AI. As a result, the analytical specificity is higher than in TOF-LMMS.

The identification power of LMMS is particularly impressive for organic compounds, not only because of the local scale but also because of the applicability to labile molecules. **Figure 4** shows the mass spectra recorded from the residue of a single peak in analytical high-pressure liquid chromatography. A sample quantity of 1–10 pg is sufficient, while conventional MS often requires material from several elutions to be combined. Even then, positive identification of important metabolites such as N-oxides remains troublesome because classical methods yield identical spectra for the drug and the metabolite. In contrast, LMMS readily distinguishes between both compounds. The structural assignment of the ions follows the well-known mechanisms of gas phase EI and AI. The parent peak in the positive ion mass spectrum of the metabolite refers to the favored loss of  $\text{H}_2\text{O}$  from  $\text{M}^+$ . The detection of odd-electron fragments shows the occurrence of EI, even though the  $\text{M}^+$  ions themselves are not seen here. Furthermore, this example also illustrates how positive and negative ions yield strikingly complementary information on the analyte.



**Figure 3** Scheme for molecular speciation of inorganic compounds using FT-LMMS. (Reprinted with permission from Struyf H, Van Vaeck L, Poels K, and Van Grieken R (1998) Fourier transform laser microprobe mass spectrometry for the molecular identification of inorganic compounds. *Journal of the American Society for Mass Spectrometry* 9: 482–497; © Elsevier.)





**Figure 4** Comparison of the positive and negative ion mass spectra recorded using TOF-LMMS from carnidazole (top) and the corresponding N-oxide (bottom). (Reprinted from Van Vaeck L, Van Espen P, Gijbels R, and Lauwers W (1988) Structural characterisation of drugs and oxygenated metabolites by laser microprobe mass spectrometry (LAMMA). *Biomedical and Environmental Mass Spectrometry* 16: 121–130; Wiley.)

## Instrumentation

### Time-of-Flight Laser Microprobe Mass Spectrometry

Commercially made instruments are available from Leybold-Heraeus (Köln, Germany) and Kratos

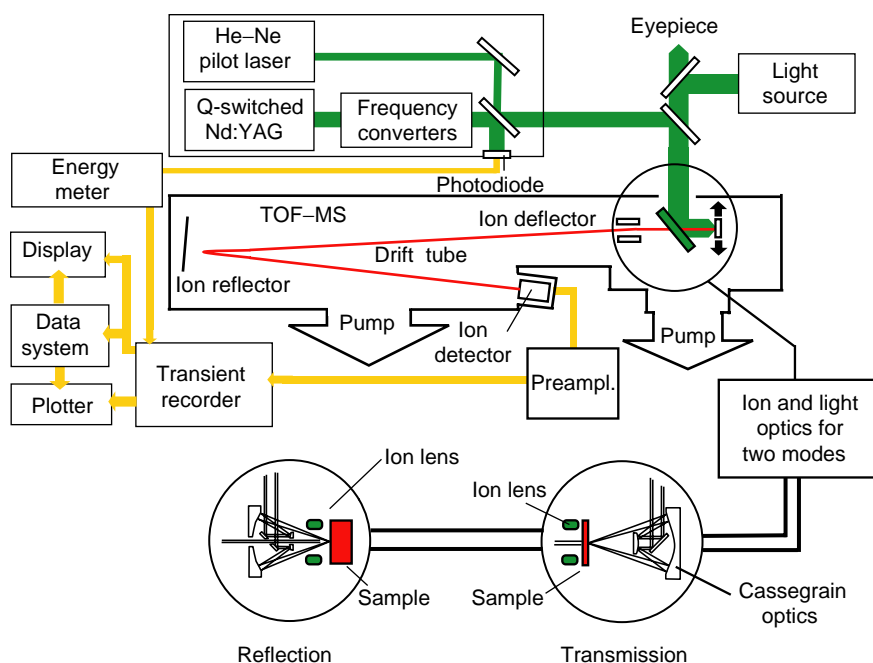
(Manchester, UK) under the names LAMMA<sup>®</sup> and LIMA<sup>®</sup>, respectively. Figure 5 shows a functional diagram of the latter instrument. The ingenious design of interchangeable sample holder and laser optics allows analysis to be performed in reflection or transmission (laser irradiation and ion extraction at the same side or opposite sides of the sample, respectively). The former mode is suitable for surface analysis of samples as thick as 1 cm, while transmission experiments are limited to thin sections (0.25–1  $\mu\text{m}$ ) or micrometer-sized particles on a polymer film. The sample is mounted on micropositioning devices in the high vacuum of the MS. The viewing microscope permits the operator to move the region of interest in the sample under the visible spot of the He–Ne pilot laser, collinearly aligned with the UV beam. Ionization is achieved by a single pulse of 15 ns duration from a frequency-quadrupled Q-switched Nd:YAG laser that delivers up to  $\sim 2$  mJ at  $\lambda = 266$  nm. Absorption filters or polarizers are used to tune the power density on the sample during irradiation between  $10^7$  and  $10^{11} \text{ W cm}^{-2}$ . Although a diffraction-limited spot of 0.5  $\mu\text{m}$  can be achieved in transmission-type instruments, a spot of 1–3  $\mu\text{m}$  is more workable.

The initial assumption that ion formation would be confined to the laser pulse duration has motivated the use of continuous ion extraction. After acceleration to  $\sim 3$  keV, ions travel through the field-free region of 1.5–2 m with a velocity that is inversely proportional to  $\sqrt{m/z}$ . A current electron multiplier, coupled to a 100 MHz memory scope, records the sequential arrival of ions as a function of time. The repulsing field of the ion reflector allows ions with higher  $E_{\text{kin}}$  values to penetrate deeper, follow a longer path, and arrive together with those having lower  $E_{\text{kin}}$  values. The resulting reduction of the peak width improves the mass resolution. The TOF-LMMS design combines simple construction and operation with excellent transmission and panoramic registration for each ion bunch over an, in principle, unlimited  $m/z$  range. However, the mass resolution is lower than expected because nominal mass separation (10% valley criterion) is only feasible up to  $m/z$  values of 500. The reason is the fundamental weakness of coupling laser microbeam DI with TOF-MS due to the specific time domain of ion generation (cf. *supra*, Figure 2C). First, the time spread on the ion bunch limits the mass resolution. Moreover, the slow component of the laser-generated ion population is only detected as an increased baseline.

### Aerosol Time-of-Flight Mass Spectrometry

The successful application of TOF-LMMS to single particle analysis has motivated the development of



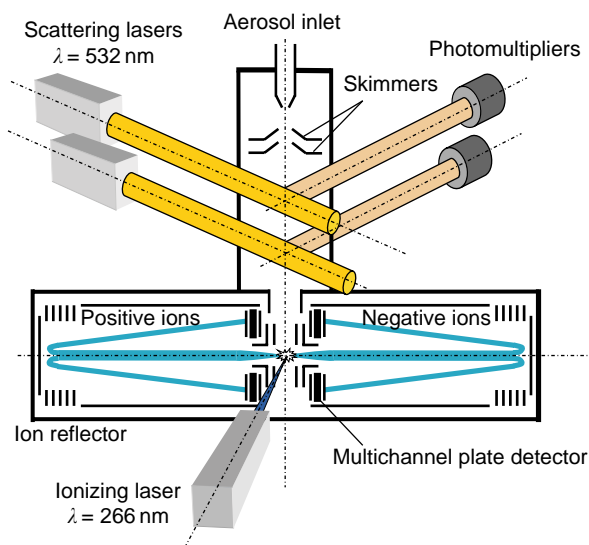


**Figure 5** Schematic diagram of LIMA<sup>®</sup> 2A, a commercial TOF-LMMS. (Reprinted from technical documentation, Kratos Analytical.)

dedicated instruments for the individual ionization of suspended particles. Different design options have been tried out and have recently yielded a compact and rugged instrument for field studies. **Figure 6** illustrates the aTOF-MS developed at University of California in Riverside and commercialized later on (TSI, Shoreview, MN). Ambient air is introduced through a nozzle and subsequently mounted skimmers to overcome the pressure difference due to the vacuum in the source. The presence of a particle in a predefined size range is probed through the time difference between the scattering events when it passes through continuous laser beams at a given distance. Once a suitable particle is detected, the Nd:YAG laser ( $\lambda = 266$  nm) fires when the aerosol moves through the waist of the ionizing beam. A brilliant idea that has never been exploited before in LMMS is the use of two TOF analyzers to record the positive and negative ions from the same particle.

#### Fourier Transform Laser Microprobe Mass Spectrometry

Basically an FT-MS is an ion storage analyzer that traps ions in an electrostatic potential valley of  $\sim 2$  eV along the axis of a strong magnetic field  $B$  (4–7 T). The latter forces the ions on circular orbits with a frequency,  $\nu_{\text{ion}}$ , inversely proportional to their  $m/z$  value. Initially, ions orbit with random phase on trajectories with a diameter than 1 mm. A radio-frequency field with frequency  $\nu_{\text{ion}}$  between the



**Figure 6** Schematic diagram of aTOF-LMMS developed at the University of California, Riverside. (Adapted from Gard E, Mayer JE, Morrical BD, *et al.* (1997) Real-time analysis of individual atmospheric aerosol particles: design and performances of a portable aTOF-MS. *Analytical Chemistry* 69: 4083–4091; American Chemical Society.)

transmitter plates allows the ions to take up energy and widen their trajectory until they orbit close to the receiver plates. The excitation also converts the original random phase of ions with a given  $m/z$  value into phase-coherent orbiting, allowing detectable image currents to be induced in the receiver plates. Hence, a sinusoidal signal with frequency  $\nu_{\text{ion}}$

and amplitude corresponding to the  $m/z$  value and the number of ions, respectively, is generated. When the trap contains ions of different  $m/z$  values, FT resolves the superposition of individual sinusoidal components into the individual  $\nu$  components ( $m/z$ ) and amplitudes.

The strong  $B$  field gives rise to a detectable  $\Delta\nu$  for ions with a  $\Delta m/z$  in the millimass range. This gives FT-MS an inherent potential for ultrahigh mass resolution and mass accuracy if the signal can be sampled over sufficient periods. However, collisions between the ions and neutrals from the residual vacuum, imperfect trapping fields, and sometimes space charge effects (too many ions in the cell) cause the phase coherence to vanish with time. Consequently, the signal becomes exponentially damped, and its decay rate determines the mass peak width and resolution. Hence, the resolution is simply increased by pumping the cell, optimizing of the trap voltages ( $V_{\text{trap}}$ ), or reducing the number of ions. The same actions also improve the sensitivity by optimizing the phase coherence and spread on the ion orbit radius after excitation. Unlike all other forms of MS, better mass resolution means better sensitivity in FT-MS.

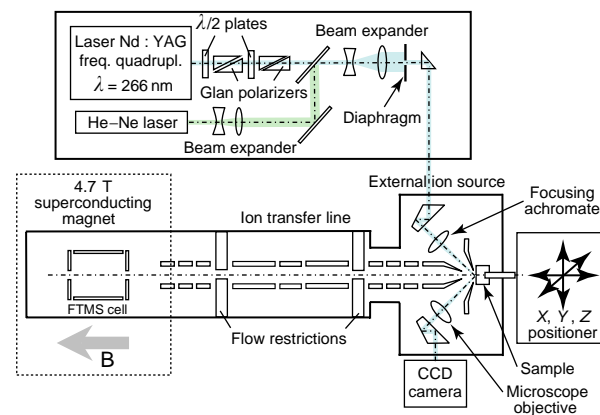
The duration of the ionization step is not critical and can be in the millisecond range as long as no ions are added when excitation starts. Hence, unlike TOF-LMMS, FT-analyzers detect the prompt and postlaser component of the DI process. Ionization can be performed directly inside the cell or in an external ion source, in which case the ions are injected through an orifice in the temporarily grounded trapping plate. Differential pumping between the external ion source and the cell allows the sample chamber to be kept at a relatively high pressure without compromising the vacuum in the analyzer cell. Alternatively, a dual cell can be used in which the orifice of the common trap plate serves as the conductance limit for differential pumping. Transfer of ions from the high- to the low-pressure compartment occurs by grounding the middle trap plate, allowing the ion bunch to be equally distributed between the two cells.

All basic FT-MS configurations have been tried out for LMMS. The instrument at the IBM laboratories uses internal ionization in a single cell, with the sample behind an orifice in the trapping plate. The laser beam passes between the receiver and transmitter plates and impinges on the sample at an angle of less than  $45^\circ$  with a spot of  $5\text{--}8\text{ }\mu\text{m}$  diameter. The presence of the sample inside the high-vacuum analyzer cell excludes specimens containing volatile components, while the pump-down time after sample exchange is significant. Therefore, the FT-LMMS developed at the University of Metz (France) uses a dual cell and internal ionization. The laser beam is

directed through the two cells and strikes the sample perpendicularly to improve energy deposition. Ingenious Cassegrain optics around a refractive lens allow excimer and/or tunable dye beams to be used for LMMS experiments with and without laser postionization. Unfortunately, the orifice in the middle trap plate must be relatively large to allow the laser beam to pass. As a result, the pressure in the two cells is different by less than a factor of 100.

Apart from the pressure problem, internal ionization implies that the sample introduction and positioning system and the optics for laser focusing, sample viewing, and illumination are inside the narrow bore of the magnet. All devices must be remote controlled and made from materials compatible with the high  $B$  field. Hence, significant compromises on the specifications are unavoidable. Another fundamental limitation is regarding trapping. A cell with a  $V_{\text{trap}}$  value of 2 V only stores ions with a total energy ( $E_{\text{tot}}$ )  $< 2\text{ eV}$ .  $E_{\text{tot}}$  is the sum of the ion's  $E_{\text{kin}}$  (up to several electronvolts in laser DI) and the potential energy ( $E_{\text{pot}}$ ), defined by the electrostatic potential at the point of ionization. Hence, ions produced from a sample with a  $V_{\text{trap}}$  value of 2 V impact on the opposite trapping plate. Lowering the sample potential to keep  $E_{\text{pot}} + E_{\text{kin}} < 2\text{ eV}$  cause ions to undergo one reflection before they strike back on the sample. Trapping requires sufficient cooling with collisions with residual neutrals to occur within one back-and-forth trajectory. In practice, this implies a high cell pressure, detrimental to FT-MS performances.

Therefore, a FT-LMMS with a single cell and external ion source, shown in Figure 7, has been built at the University of Antwerp. The ionization chamber is far away from the magnet and freely accessible,



**Figure 7** Schematic diagram of the FT-LMMS with an external ion source developed at the University of Antwerp. (Adapted from Van Vaecck L, Van Roy W, Struyf H, Adams F, and Caravatti P (1993) Development of a laser microprobe Fourier transform mass spectrometer with external ion source. *Rapid Communications in Mass Spectrometry* 7: 323–331; Wiley.)

allowing adequate devices to be mounted for laser focusing to a spot size  $<5\text{ }\mu\text{m}$ , sample positioning, and viewing. Differential pumping of the transfer line allows the source and cell to be kept at  $10^{-6}$  and  $10^{-10}$  Torr, respectively. Dedicated ion optics producing static electrical fields guide ions with an initial emission angle of up to  $35^\circ$  to the cell. Furthermore, the potential of the sample and the selvedge can be freely tuned to keep  $E_{\text{ion}} = E_{\text{kin}} + E_{\text{pot}} < 2\text{ eV}$  for optimal trapping. As a result, the limit of detection (LOD) is as low as  $10^6$ – $10^7$  molecules, the mass resolution routinely exceeds 4 000 000 and 150 000 at  $m/z$  values of 56 and 1000, respectively, and the mass accuracy is better than 1 ppm.

## Analytical Figures of Merit

Local analysis of solids excludes the use of prior separation or enrichment steps. Hence, the specificity becomes as important as the sensitivity. The LOD essentially determines the minimal sample volume in which an analyte with a given concentration can be traced back. Specificity refers to a combination of information levels (elements, bonds, functional groups, MW, and/or structure), resolving power to overcome interferences, lateral resolution, and information depth. The last two parameters determine the scale on which compositional discontinuities can be observed.

The characterization of a thin layer or coating is the easiest case when its thickness exceeds the information depth. Additionally, minimal spectrometric resolution is required, while the analysed area can be enlarged to match the LOD. In contrast, analysis of microobjects on a substrate requires adequate sensitivity and sufficient specificity to distinguish between the analyte and substrate signals when the phase of interest is smaller than the lateral resolution and information depth. However, proper substrate selection makes this case less demanding than local inclusions in a complex heterogeneous matrix. Finally, buried analytes are the most difficult task. Sufficient information depth is required while the matrix contribution to the detected signals increases with the distance between the analyte phase and the surface.

Looking at the specificity in terms of the information level, LMMS and S-SIMS inherently rank high because full molecular information (i.e., MW + structural fragments) is obtained. In contrast, EPX-MA, X-ray and synchrotron radiation based methods are confined to elemental analysis, while the bond-specific information of X-ray photoelectron spectroscopy,  $\mu$ -infrared (IR), or Raman spectroscopy is insufficient for characterizing unknown mixtures.

The lateral resolution of LMMS is typically 1 and  $5\text{ }\mu\text{m}$  for (a)TOF- and FT-LMMS, respectively. The diffraction-limited spot size is 400 nm at  $\lambda = 200\text{ nm}$ . In S-SIMS, a full mass spectrum is typically recorded from an area between  $25 \times 25$  and  $250 \times 250\text{ }\mu\text{m}^2$ , while  $1$ – $5\text{ }\mu\text{m}$  is the minimal size of the features to be imaged with structural ions. Although a single laser shot erodes material to a depth of  $0.1$ – $1\text{ }\mu\text{m}$ , ions are only generated from the upper  $10$ – $50\text{ nm}$ . In contrast, the ion beam damage of the subsurface restricts the information depth of S-SIMS to one monolayer ( $0.1$ – $1\text{ nm}$ ). Both S-SIMS and LMMS have sufficient detection sensitivity to trace back the major components within one or a few monolayers. This brings the methods within the reach of nanoscale applications. The multilayer information depth in LMMS is an advantage in comparison with S-SIMS when accidental contaminants cover the surface of interest.

Only  $10^6$ – $10^7$  molecules are needed for detection of inorganic and organic analytes with average ionization yield in FT-LMMS with an external source. Under favorable conditions, an organic monolayer can be detected. Thanks to the sampling of the post-laser ionization, the LODs in high mass resolution FT-LMMS compare favorably with those of low mass resolution TOF-LMMS. Registration of a full mass spectrum from conversion layers on aluminum requires similar sample consumption in S-SIMS and FT-LMMS, yielding low and high mass resolution data, respectively. There are no limitations with respect to sample conductivity in LMMS, while charge build-up hampers the application of S-SIMS to dielectric coatings even as thin as a few micrometers.

So far, FT-LMMS is the only methodology for microanalysis with the specificity of high mass resolution. This makes the method superior by far to S-SIMS when it comes to the identification of unknowns. It also gives FT-LMMS a significant potential to deal with nanoscale analyte phases smaller than the spot.

Sampling the full ion production in FT-LMMS improves reproducibility in comparison with TOF-LMMS. Because the kinetics of the DI process does not affect the mass analysis any longer, FT-LMMS is applicable to analytes with higher MWs and more polar groups than TOF-LMMS. On the other hand, S-SIMS excels in generating structural ions at high  $m/z$  values from polymer materials, while LMMS tends to lead to pyrolysis.

Ion imaging is one of the attractive features of S-SIMS. Because the ionization in LMMS depends critically on exact focusing of the ionizing beam on the sample surface, mapping experiments are only feasible on very flat samples. The quantitative

capabilities of both methods are similar and depend on the availability of suitable reference samples for calibrating the signal as a function of the local composition. Cryofreezing of solutions and ultraslow lyophilization in vacuum yield adequate reference samples for determinations within typically 10%.

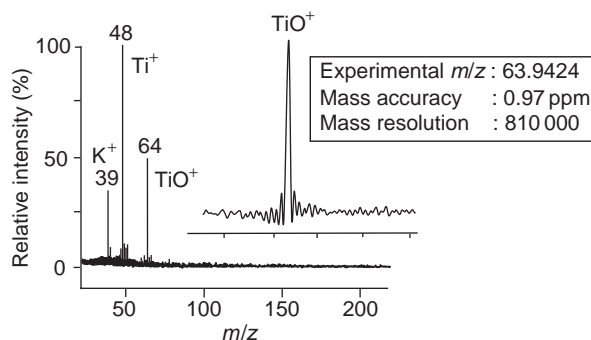
A major limitation is the inaccessibility of analytes buried deep under the surface. In this respect, non-destructive methods going from electron beam to spectroscopic techniques such as  $\mu$ -IR or  $\mu$ -Raman are more adequate but lack sensitivity for nanolayer detection.

## Selected Examples of Applications

### Biomedical and Biological Applications

The diffraction-limited spot of  $0.5\ \mu\text{m}$  allows TOF-LMMS in the transmission geometry to localize elements at the subcellular level. Tissue sections can be prepared using the common methods used for optical and electron microscopy. Cryotechniques are preferred because embedding of tissues in resin causes redistribution of mobile elements. Adequate calibration samples for heterogeneous tissue are a problem for quantitative work. Successful studies involve the localization of aluminum in bones of haemodialysis patients, accumulation of lead in kidneys due to chronic intoxication, incorporation of heavy metals from dental alloys in amalgam tattoos of the oral mucosa. Examples of speciation include the identification of hydroxy-apatite in the spheruliths of the Bowman's membrane from patients treated with high doses of cyclosporin. Also, the pathogenesis of aseptic loosening of joint prostheses has been related to the presence of zirconium oxide in granular foreign bodies of the surrounding tissue.

It has been anticipated that FT-LMMS with an external source would be inadequate for tissue research because of its spot size of  $5\ \mu\text{m}$  and reflection geometry hampering the visualization of sections. However, the composition and heterogeneity of tissues makes the analytical specificity of high mass resolution an invaluable asset. For instance, local analysis of foreign bodies in the inflamed tissue surrounding a titanium implant has allowed relatively simple mass spectra with prominent peaks at  $m/z$  values of 48 and 64 to be recorded. The inset in Figure 8 illustrates the detection of  $\text{TiO}^+$  at an  $m/z$  value of 64 with a mass resolution of 800 000, clearly sufficient to eliminate all isobaric interferences. The mass accuracy within 1 ppm ensures that the experimental  $m/z$  value is significant up to the fourth decimal place and excludes an alternative elemental composition. This leads to the unambiguous



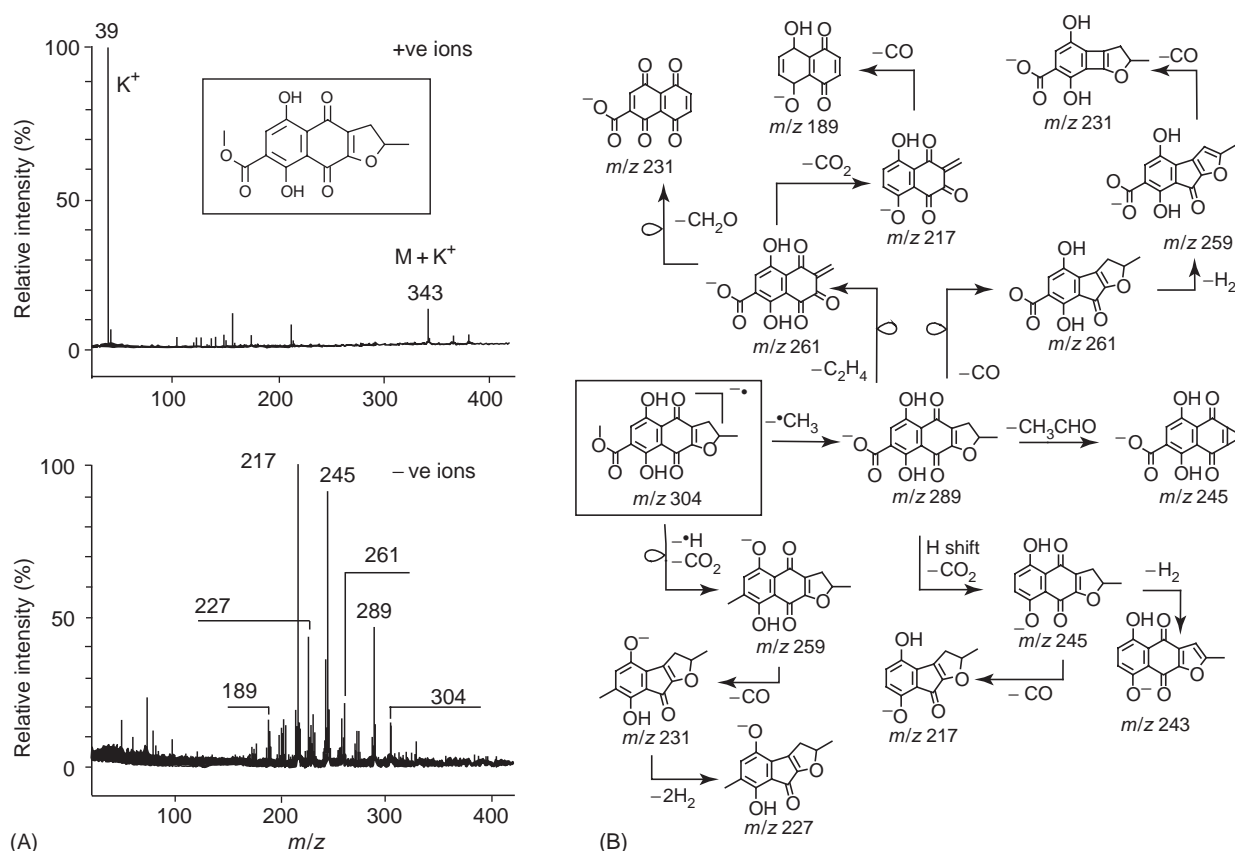
**Figure 8** Identification of foreign body cells in inflammatory tissue around a metal implant using FT-LMMS. The inset shows the high mass resolution trace of the  $^{48}\text{TiO}^+$  ions, offering sufficient separation and mass accuracy for unambiguous identification. (Reprinted from De Nollin S, Poels K, Van Vaeck L, *et al.* (1997) Molecular identification of foreign inclusions in inflammatory tissue surrounding metal implants by Fourier transform laser microprobe mass spectrometry. *Pathology Research and Practice* 193: 313–318; Gustav Fischer Verlag.)

identification of titanium oxide wear particles from the prosthesis.

Amongst biological applications using TOF-LMMS, the uptake of aluminum by fish in aquatic ecosystems subjected to acid rain has been studied. Use of the mass spectra of bacterial cultures as fingerprints has yielded screening methods for checking chemotherapy. The analytical specificity of high mass resolution FT-LMMS facilitates identification of organic substances in biological matrices. Figure 9 illustrates the results of an *in situ* analysis of the pigments responsible for sunlight interception and energy transfer in the apothecia of the microlichen *Haematomma ventosum*. The mass spectra from local spots with a diameter of  $5\ \mu\text{m}$  readily reveal the MW through the  $[\text{M}+\text{K}]^+$  and  $\text{M}^-$  ions at  $m/z$  values of 343 and 304, respectively. The detection of  $\text{M}^-$  confirms the occurrence of EI in LMMS. The high mass accuracy data in Table 1 are essential for the structural assignment of the numerous negative ion fragments and determination of the functional groups. As a result, FT-LMMS allows the time-consuming steps of isolation and subsequent high resolution MS to be performed while information on the local distribution is being obtained. Sample preparation only involves cutting raw biological material to dimensions not larger than  $1 \times 1 \times 1\ \text{cm}^3$ .

### Environmental Applications

The composition of single aerosol particles in the micrometer range is important for assessing their environmental health hazard and for studying the atmospheric chemistry of particle formation and transformation during atmospheric transport. Both



**Figure 9** *In situ* analysis of the pigment haemovantosin in the apothecia of a microlichen *Haematomma ventosum* using FT-LMMS. The structural assignment of the anions of major diagnostic interest is based on the high accuracy  $m/z$  data in Table 1. (Reprinted from Van Roy W, Mathey A, and Van Vaecck L (1996) *In-situ* analysis of lichen pigments by Fourier transform laser microprobe mass spectrometry with external ion source. *Rapid Communications in Mass Spectrometry* 10: 562–572; Wiley.)

issues make a distinction between the surface and bulk of individual particles desirable. In fact, TOF-LMMS is the first microprobe with sufficient sensitivity and specificity for molecular identification of inorganic and organic components in single aerosol particles collected through cascade impactor sampling. Successive recording of mass spectra at low laser power densities can be used to provide evidence of the presence of a nitrate shell covering a core of sodium chloride in a sea-salt aerosol. The detected adduct ions (cf. scheme in Figure 3) allow the chemical composition to be deduced with superior specificity in comparison with element detection using e.g., EPXMA. Furthermore, the information depth of  $\sim 1\mu m$  prevents the latter method from distinguishing between the composition of the core and shell. In contrast, EPXMA is superior to LMMS with respect to quantification and automated characterization of large particle populations.

The advent of aTOF-MS instruments that are capable of in-field experiments of online sampling and analysis has virtually annihilated the use of TOF-LMMS for aerosol research. Additionally, aTOF-MS

provides a major advantage by detecting both positive and negative ions from the same particle. Figure 10 shows the mass spectra of a single wood smoke particle that has yielded intense signals from structural fragments of (oxygenated) hydrocarbons. The ambient aerosol generates peaks from combustion related vanadium oxide and characteristic ions from ammonium nitrate. This information would not be available from one ion detection mode. Unlike TOF-LMMS, aTOF-MS can be automated for survey studies with high time resolution. The resulting huge data sets are processed using (multivariate) statistical procedures. On the other hand, aTOF-MS gives no information on the particle's morphology, which often includes interesting hints about its origin and formation. Also, the limited possibilities of fine-tuning the laser power density according to the particle type prevent distinction between the core and shell composition and cause many particles to generate insufficient spectral information. Ionization is improved by filling the source with a gaseous UV absorber, which coats the particle before irradiation.



**Table 1** Mass measurement of diagnostic ions detected from haemaventosin as shown in Figure 9

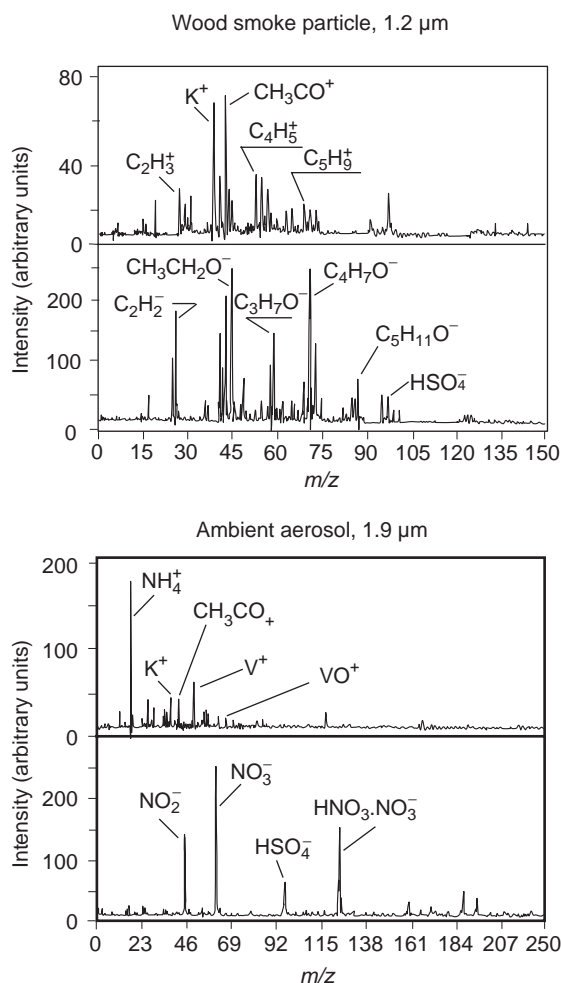
Experimental $m/z$	Elemental composition	Accuracy of $m/z$ (ppm)
<b>Positive ions</b>		
343.022	$C_{15}H_{12}O_7K^+$	1.3
<b>Negative ions</b>		
305.030	$C_{14}H_9O_8^-$	0.4
304.059	$C_{15}H_{12}O_7^-$	0.1
289.035	$C_{14}H_9O_7^-$	0.4
261.041	$C_{13}H_9O_6^-$	0.5
261.004	$C_{12}H_5O_7^-$	1.0
259.061	$C_{14}H_{11}O_5^-$	0.5
259.025	$C_{13}H_7O_6^-$	0.6
245.046	$C_{13}H_9O_5^-$	0.1
245.010	$C_{12}H_5O_6^-$	1.2
243.030	$C_{13}H_7O_5^-$	0.9
231.067	$C_{13}H_{11}O_4^-$	1.1
231.030	$C_{12}H_7O_5^-$	1.7
230.994	$C_{11}H_3O_6^-$	0.2
227.035	$C_{13}H_7O_4^-$	0.1
217.051	$C_{12}H_9O_4^-$	0.2
217.014	$C_{11}H_5O_5^-$	0.7
189.020	$C_{10}H_5O_4^-$	1.1

Reprinted from Van Roy W, Mathey A, and Van Vaeck L (1996) *In-situ* analysis of lichen pigments by Fourier transform laser microprobe mass spectrometry with external ion source. *Rapid Communications in Mass Spectrometry* 10: 562–572; Wiley.

### Industrial Materials and Troubleshooting Applications

The advantages of LMMS when compared with S-SIMS are the easy analysis of nonconducting samples, the increased diagnostic information due to the softer energy regime, and the capability of probing the molecular composition in the near-subsurface, i.e., under the upper monolayer, which often consists of accidental contaminants. As a result, LMMS is often applied to industrial processing faults.

Advanced metallurgical research has used TOF-LMMS to study casting or welding of new light alloys. The formation of borosilicates in the joints of oxide-dispersion-strengthened (ODS) superalloys has been related to metallic glass foils used for brazing. The presence of such heterogeneous phases in the joints of e.g., turbine engines can be detrimental under high-stress conditions. The discoloration of steel can be due to the deposition of unwanted species or to light interference by a thin homogeneous oxide layer. Unlike Auger electron spectroscopy, TOF-LMMS distinguishes between the two causes. Other examples involve the identification of surface impurities that prevent subsequent hardening or anticorrosion coatings from adhering. Paint defects become apparent only when the different lacquer layers have been already applied. In that case, successive laser shots



**Figure 10** Positive and negative ion mass spectra detected using aTOF-MS from a single wood smoke particle (top) and atmospheric aerosol sampled in Riverside, CA. (Reprinted from Gard E, Mayer JE, Morrical BD, *et al.* (1997) Real-time analysis of individual atmospheric aerosol particles: design and performances of a portable aTOF-MS. *Analytical Chemistry* 69: 4083–4091; American Chemical Society.)

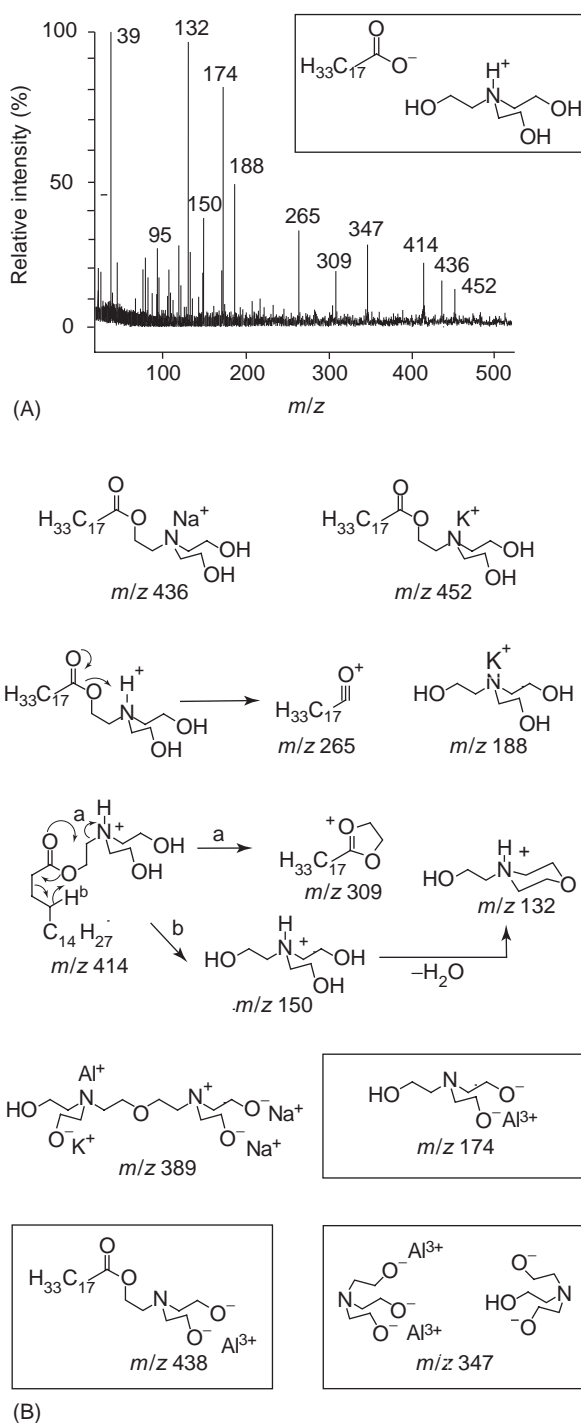
allow buried contamination to be exposed for analysis. Unlike S-SIMS, laser erosion is nondestructive to molecular information, and structural ions have been recorded from fragments of polyethyleneterephthalate (PET) gloves left by coachbuilders on bare metal.

Although LMMS is limited in the characterization of polymers, it excels in the study of local inclusions compromising material performances. For instance, poorly dispersed accelerators and local accumulation of metals from corroded reactor or extruder walls have been traced back using TOF-LMMS. The dispersion of magnetic material inside the PET matrix of faulty floppy disks has been studied using external source FT-LMMS. Simultaneous detection of the inorganic and organic components has allowed disk failure to be correlated to compositional anomalies.

Easy speciation and isotope detection have been identified as playing important roles in optimizing the production of superconducting  $\text{Nb}_3\text{Sn}$  wires by heating a composite of bronze and niobium filaments. Application of TOF-LMMS has provided evidence of the out-diffusion and accumulation of phosphorous, inhibiting  $\text{Nb}_3\text{Sn}$  formation when tin diffuses inward. The preferential incorporation of the lower tin isotopes in the  $\text{Nb}_3\text{Sn}$  layer has confirmed the process to be driven by diffusion. The semiconductor industry has used TOF-LMMS to verify the lateral diffusion of dopants, e.g., boron in tantalum and cobalt silicide runners only  $3\text{ }\mu\text{m}$  wide. Another application is the identification of microscopic residues left on integrated circuits after removal of the photoresist.

The increased identification power due to high mass resolution in FT-LMMS significantly enlarges the range of troubleshooting applications in comparison with TOF-LMMS. In particular, anomalies in the extremely well-controlled production of pharmaceuticals and fine chemicals require completely unexpected compounds to be identified. Quality control procedures nowadays require the reason for only a few microscopic particles in complex formulations to be traced back. Application of FT-LMMS to a few micrometer-size particles has allowed specific causes to be pinpointed, such as components from the piston and ceramic seal pointing to a faulty high pressure liquid chromatography pump, polyimide fragments indicating wear of the belt driving a stirrer, and an organic salt pointing to an incomplete washing step.

Figure 11 shows mass spectra taken from the organic additive on the surface of a rolled aluminum plate, using FT-LMMS with an external ion source. The mass resolution easily exceeds 100 000, and the mass accuracy (Table 2) is within 0.1–1 ppm. The specific problem studied here relates to the deposition and interaction of organic additives from the lubricating emulsion used in the aluminum hot rolling process. Irreversible binding of organic additives to the aluminum at high temperatures and pressures causes microscopic defects in the subsequently applied anodization layer or coating. The positive ion mass spectra contain several signals not found with the pure additive (framed structures in Figure 11). The elemental composition derived from high mass accuracy data indicates the presence of  $\text{Al}^{3+}$  together with two anionic functionalities. These ions must originate from ion pairs that exist already in the solid state before ionization since the laser only generates  $\text{Al}^+$  to form adducts in the selvage. In this example, FT-LMMS has been used to achieve one of the most difficult tasks in MS, namely the detection



**Figure 11** Identification of the interaction products between the triethanolamine oleate additive of a lubricating emulsion and rolled aluminum. Top: positive ion mass spectrum recorded using FT-LMMS with an external ion source. Bottom: structural assignment of the ions of major diagnostic interest. The corresponding high accuracy  $m/z$  data are listed in Table 2. The framed structures are indicative of the binding of the additive to aluminum. (Reprinted from Poels K, Van Vaeck L, Van Espen P, Terryn H, and Adams F (1996) Feasibility of Fourier transform laser microprobe mass spectrometry for the analysis of lubricating emulsions on rolled aluminum. *Rapid Communications in Mass Spectrometry* 10: 1351–1360; Wiley.)

**Table 2** Mass measurement of diagnostic ions detected from triethanolamine oleate on aluminum as shown in **Figure 11**

Experimental <i>m/z</i>	Elemental composition	Accuracy of <i>m/z</i> (ppm)
452.314	C <sub>24</sub> H <sub>47</sub> NO <sub>4</sub> K <sup>+</sup>	0.4
438.316	C <sub>24</sub> H <sub>45</sub> NO <sub>4</sub> Al <sup>+</sup>	0.6
436.340	C <sub>24</sub> H <sub>47</sub> NO <sub>4</sub> Na <sup>+</sup>	0.9
414.358	C <sub>24</sub> H <sub>48</sub> NO <sub>4</sub> <sup>+</sup>	0.1
389.101	C <sub>12</sub> H <sub>25</sub> N <sub>2</sub> O <sub>5</sub> KNa <sub>2</sub> Al <sup>+</sup>	0.4
347.134	C <sub>12</sub> H <sub>25</sub> N <sub>2</sub> O <sub>6</sub> Al <sub>2</sub> <sup>+</sup>	0.6
309.279	C <sub>20</sub> H <sub>37</sub> O <sub>2</sub> <sup>+</sup>	0.6
265.253	C <sub>18</sub> H <sub>33</sub> O <sup>+</sup>	0.2
188.068	C <sub>6</sub> H <sub>15</sub> NO <sub>3</sub> K <sup>+</sup>	0.1
174.071	C <sub>6</sub> H <sub>13</sub> NO <sub>3</sub> Al <sup>+</sup>	0.4
150.113	C <sub>6</sub> H <sub>16</sub> NO <sub>3</sub> <sup>+</sup>	0.2
132.102	C <sub>6</sub> H <sub>14</sub> NO <sub>2</sub> <sup>+</sup>	0.3
109.101	C <sub>6</sub> H <sub>13</sub> <sup>+</sup>	0.1
97.1011	C <sub>7</sub> H <sub>13</sub> <sup>+</sup>	0.9
95.0855	C <sub>7</sub> H <sub>11</sub> <sup>+</sup>	0.8

Reprinted from Poels K, Van Vaeck L, Van Espen P, Terry H, and Adams F (1996) Feasibility of Fourier transform laser microprobe mass spectrometry for the analysis of lubricating emulsions on rolled aluminum. *Rapid Communications in Mass Spectrometry* 10: 1351–1360; Wiley.

of interaction products between two distinct phases, which requires the intermolecular bonds with each phase to be broken without destruction of the intermolecular binding across the interface.

See also: **Fourier Transform Techniques. Mass Spectrometry:** Time-of-Flight. **Surface Analysis:** Auger Electron Spectroscopy; Laser Ionization.

## Further Reading

- Asamoto B (ed.) (1991) *FT-ICR/MS: Analytical Applications of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry*. New York: VCH Publishers Inc.
- Bruynseels F and Van Grieken R (1985) Detection of sulfate and nitrate layers on sampled marine aerosols by laser microprobe mass analysis. *Atmospheric Environment* 19: 1969–1970.
- Noble CA and Prather KA (2000) Real-time single particle mass spectrometry: a historical review of a quarter century of the chemical analysis of aerosols. *Mass Spectrometry Reviews* 19: 248–274.
- Pelletier M, Krier G, Muller JF, Weil D, and Johnston M (1988) Laser microprobe Fourier transform mass spectrometry. Preliminary results for feasibility and evaluation. *Rapid Communications in Mass Spectrometry* 2: 146–150.
- Van Vaeck L, Struyf H, Van Roy W, and Adams F (1994) Organic and inorganic analysis with laser microprobe mass spectrometry. Part II: applications. *Mass Spectrometry Reviews* 13: 209–232.

# ATOMIC SPECTROMETRY

## Overview

**J A Holcombe**, University of Texas at Austin, Austin, TX, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Traditionally, analytical atomic spectroscopy implied the use of electromagnetic radiation in the ultraviolet (~200–350 nm) and visible (~350–800 nm) region of the spectra for qualitative and quantitative analysis. With the use of some of the same sources to produce ions for detection using mass spectrometry, the term often encompasses the area of elemental mass spectroscopy. In this section the focus will remain on optical techniques.

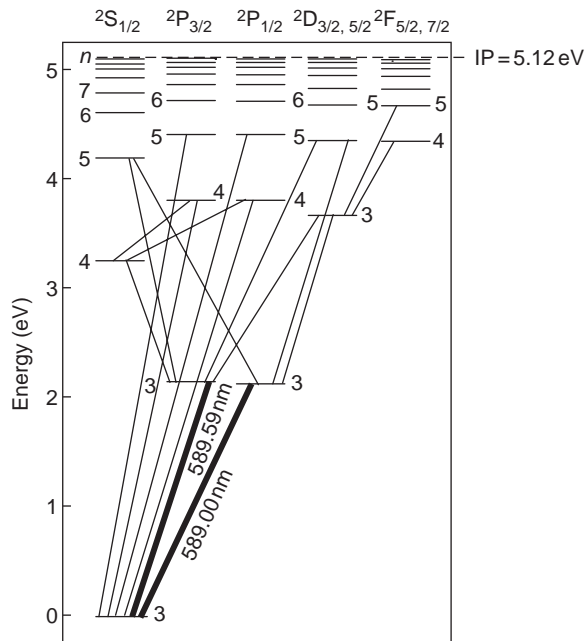
Atomic spectroscopists are concerned with the energy states of the atom or ion under study. These energy states can be seen diagrammatically in a term

(or Grotian) diagram, as shown in **Figure 1** for Na. These are the energy states for the isolated atom in the gas state. Molecules such as NaCl<sub>(g)</sub> have their own unique set of allowable energy states but have little relationship to the atomic spectral states. As a consequence, in order to use atomic spectroscopy to conduct elemental analysis on a sample, the species of interest (the analyte) must be converted to gaseous atoms using a source (flame, plasma, etc.). It should also be noted that the energy states are entirely different for the free atom and its ion. As a result, if we are to determine the atom concentration we need to minimize the extent of ionization if monitoring a neutral atom line, maximize ionization if monitoring an ion line or, at the very least, keep the degree of ionization constant from one sample to the next. This way the number of atoms or ions detected in our source remains proportional to the analyte concentration in the sample.

Three general types of experiment can be used in atomic spectroscopy: emission, absorption, and fluorescence (**Figure 2**). The 'atom cell' is where free gaseous atoms are generated (e.g., a flame). The



intensity of the radiation detected ( $I$ ) is measured and used to determine ultimately the analyte concentration in the sample, which is generally a solution.



**Figure 1** Term diagram for sodium showing location of various atomic states of the neutral atom. The vertically oriented lines denote some of the allowable transitions that could occur that give rise to spectral lines. The two transitions that are annotated with the emission (or absorption) wavelengths of 589.59 and 589.00 nm are the resonance lines that appear in the yellow region of the spectra and are often termed as the Na D lines.

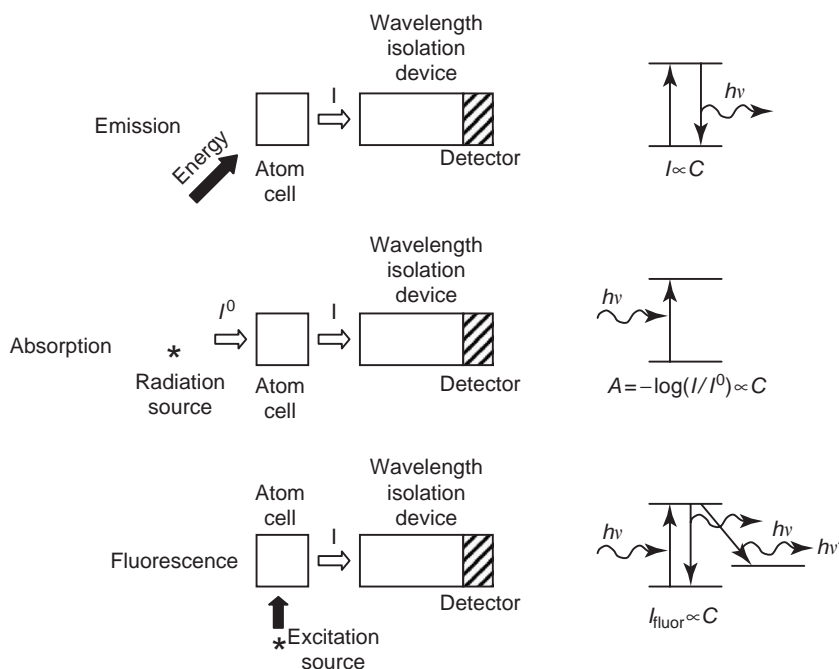
## Fundamentals of Atomic Spectrometry

### Electronic Transitions

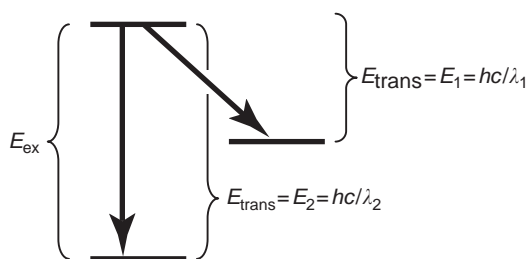
Figure 1 suggests that a number of transitions to various energy states could take place in the atom. In emission spectroscopy, the excited states are generally populated by collisional processes (e.g., heat); and when they relax ('fall down' to a lower energy state), they can lose the gain in internal energy either radiatively or nonradiatively. In spectroscopy, we are interested in radiative relaxation processes in which the wavelength of the emitted radiation is related to the transition energy,  $E_{\text{trans}}$  (Figure 3),

$$E_{\text{trans}} = hc/\lambda \quad [1]$$

where  $h$  is Planck's constant ( $6.626 \times 10^{-34}$  J s),  $c$  is the velocity of light ( $2.99 \times 10^8$  m s $^{-1}$ ), and  $\lambda$  is the wavelength in meters. Thus, a number of spectral lines can exist for any given element, depending on the extent of population of the various excited states. However, not all transitions are quantum mechanically allowed, thus reducing some of the spectral complexity. For most atomic spectroscopy applications, the primary transitions that provide the maximum sensitivity involve the ground state and are referred to as resonance transitions, and the resulting spectral lines referred to as resonance lines. In most sources, these are the primary lines used for analysis.



**Figure 2** Schematic diagram of three dominant processes (emission, absorption, and fluorescence) that can be used for spectrochemical analysis.



**Figure 3** Radiative relaxation processes that could occur from population of an excited state that has an excitation level of  $E_{\text{ex}}$  above the ground state. The transition energies,  $E_{\text{trans}}$ , of  $E_1$  and  $E_2$  give rise to emission lines at  $\lambda_1$  and  $\lambda_2$ , respectively.

While Figures 1 and 3 suggest the lines are of a single wavelength, each line has a finite width. The natural line width is as narrow as a spectral line can be and is governed by Heisenberg's uncertainty principle,  $\Delta E = h/(2\pi\tau)$ , where  $\Delta E$  is the uncertainty in the exact location of the excited state and  $\tau$  is the excited state lifetime, which is  $\sim 1$  ns for many electronic transitions of atoms. For a 300 nm line and  $\tau \sim 1$  ns, this predicts a natural line width of  $\sim 0.05$  pm. Line widths of most sources are broader than this because of collisional (i.e., pressure) or Doppler broadening mechanisms. For sources at 1 atm pressure and a few thousand degrees, both these broadening mechanisms produce line widths in the range of a few picometres or about a hundred times larger than the natural line width. The very narrow lines (even with broadening) permit unique wavelength assignments to specific elements and provide a degree of selectivity found in few other analytical techniques.

### Population of States

Most thermal sources used in atomic spectrometry provide temperatures in the range 1500–7000°C. Even at extreme temperatures, a majority of the neutral atoms exist in the ground state. Table 1 provides some indication of the temperature-dependent population of the first excited state for Na, which is relatively easily populated. The table also shows the population of the first excited state of Zn, whose transition energy, in contrast, is more than twice that of Na. At the higher temperatures, ionization can become appreciable and a majority of the analyte may exist as the ionized species, although, those sodium species still present as neutral atoms will still exhibit the relative populations shown in Table 1.

The fractional population of excited states in a system under local thermodynamic equilibrium (LTE) can be determined using Boltzmann's equation,

$$\frac{N^*}{N_T} = \frac{g^* e^{-E^*/kT}}{Z(T)} = \frac{g^* e^{-E^*/kT}}{\int g_i e^{-E_i/kT}} \quad [2]$$

**Table 1** Relative population of the first excited states of sodium and zinc at various temperatures

Temperature (K)	Na (589.6 nm)	Zn (213.9 nm)
1000	$2.6 \times 10^{-11}$	$2.3 \times 10^{-29}$
1500	$8.9 \times 10^{-8}$	$1.2 \times 10^{-19}$
2000	$5.1 \times 10^{-6}$	$8.3 \times 10^{-15}$
2500	$5.9 \times 10^{-5}$	$6.8 \times 10^{-12}$
3000	$3.0 \times 10^{-4}$	$5.9 \times 10^{-10}$
3500	$9.5 \times 10^{-4}$	$1.4 \times 10^{-8}$
4000	$2.3 \times 10^{-3}$	$1.6 \times 10^{-7}$
4500	$4.4 \times 10^{-3}$	$1.0 \times 10^{-6}$
5000	$7.5 \times 10^{-3}$	$4.5 \times 10^{-6}$
5500	$1.2 \times 10^{-2}$	$1.5 \times 10^{-5}$
6000	$1.6 \times 10^{-2}$	$4.2 \times 10^{-5}$
6500	$2.2 \times 10^{-2}$	$9.9 \times 10^{-5}$
7000	$2.8 \times 10^{-2}$	$2.1 \times 10^{-4}$

where  $N^*$  and  $N_T$  are the number densities (e.g., atoms  $\text{cm}^{-3}$ ) of atoms in a given excited state and all the atoms, respectively;  $g$  is the statistical weight;  $E^*$  is the excitation energy (energy above the ground state – see also Figure 3);  $k$  is Boltzmann's constant ( $1.38 \times 10^{-23} \text{ J K}^{-1}$ ); and  $T$  is the temperature in Kelvin.  $Z(T)$  is the partition function, which basically represents the sum of the fractional populations of all other states in the atom. The numerical value of  $Z(T)$  can be found in tables. Since most of the atoms are located in the ground state, eqn [2] can be approximated by

$$\frac{N^*}{N_T} \approx \frac{N^*}{N^0} = \frac{g^* e^{-E^*/kT}}{g^0} \quad [3]$$

where  $N^0$  is the population of the ground state and  $g^0$  is the statistical weight of the ground state.

### Ionization

The extent of ionization depends on the element, the temperature, and the free electron density. It varies from source to source. Ionization can be written as an equilibrium process:



Given no other dominant source of electrons, the degree of ionization will increase as the number density (or partial pressure) of the analyte in the source decreases. The extent of ionization also increases as the ionization potential of the element decreases. The degree of ionization can be calculated using Saha's equation.

Excessive ionization can reduce the atomic spectroscopy signal from the neutral atom, although accurate quantitation can still be conducted as long as the degree of ionization remains constant for all samples and standards. In some high-temperature sources, ionization is sufficiently large for ions to be

the dominant species, and the ion resonance line is used for analysis.

### Interferences

The term 'interference' is applied to any source or event that causes the signal for a given amount of analyte to be bigger or smaller than would be predicted from the calibration curve. These are often subdivided into chemical, ionization, and spectral interferences. In brief, there is an expectation that the atom density in an atomic spectroscopic source (and the relative population in an excited state for emission techniques) remains proportional to the elemental concentration in the sample. Anything in the source or sample that invalidates this assumption causes an interference.

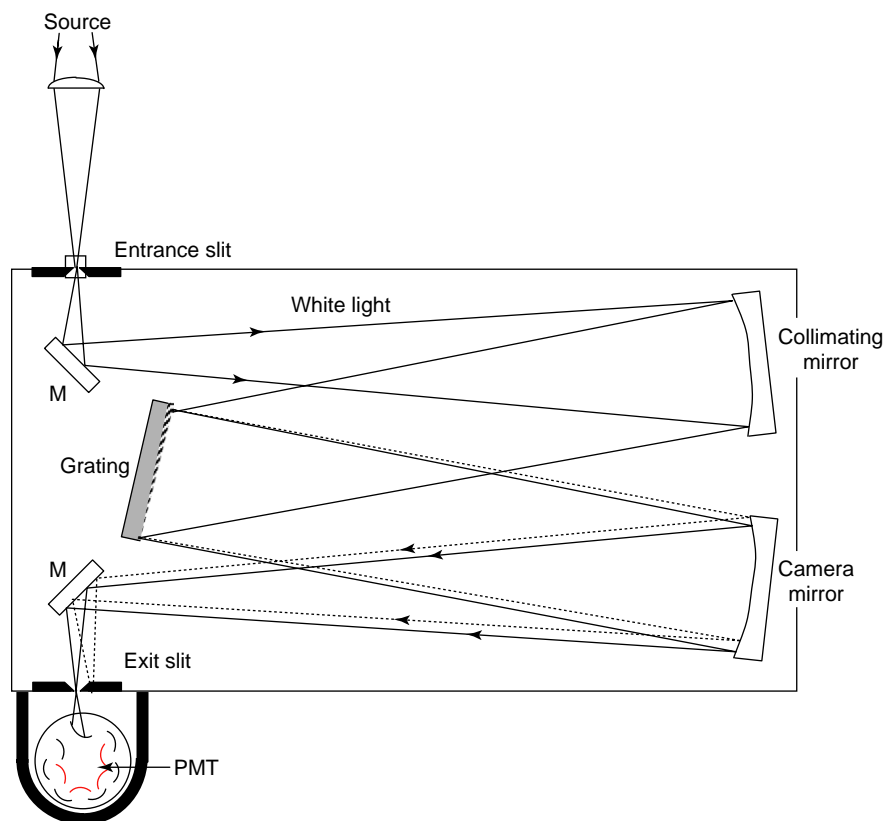
## Basic Instrumentation

### Spectrometers

Since a wavelength characteristic of the element(s) of interest must be isolated, a dispersive device must be included. Generally gratings are used in

spectrometers to accomplish this function. The 'size' (i.e., focal length of the mirrors) combined with the density of rulings on the grating or mounting configuration of the optical components govern the resolution available. A higher resolution is needed when the analyte line must be isolated from a very large number of lines or when multiple detectors are used and there is subsequent need for physical space to place the detectors in the focal plane of the spectrometer. **Figure 4** illustrates a simple dispersive device based on a diffraction grating. The term 'spectrometer' is generic and applies to all types of dispersive-type instruments. 'Spectrophotometer' generally refers to an instrument where comparison measurements are made, such as in absorption spectroscopy. When a single exit slit exists and the grating is rotated to select the wavelength, the device is called a monochromator. With multiple exit slits and multiple detectors, the term 'direct reader' or 'polychromator' is often used.

In some instances where the spectral isolation demands are minimal, simple interference filters can be used in place of a dispersive-type instrument. In these cases, the term 'filter photometer' rather than 'spectrometer' is employed.



**Figure 4** Schematic diagram of one configuration ('mount') for a reflection grating monochromator. This particular arrangement of mirrors and grating is relatively common and referred to as a Czerny–Turner mount. The dashed and solid lines leaving the grating represent shorter and longer wavelengths, respectively, that are angularly separated after diffraction by the grating has occurred. Simple front-surfaced plane mirrors (M) serve only to 'bend the beam'. The detector shown in this diagram is a photomultiplier (PMT).

## Detectors

While photographic film was traditionally used as a detector, various types of photoelectric detectors have replaced it. The photomultiplier (PMT) is a commonly used, sensitive detector where photons striking the photocathode cause the ejection of an electron if the photon energy can overcome the work function of the material used in fabricating the photocathode. The probability of a photon producing a photoelectron from the cathode is termed the quantum efficiency, which is dependent on the cathode material and wavelength. A quantum efficiency of 10% is typical. This photoelectron is electrically directed and accelerated to the first dynode where multiple (typically two to six) secondary electrons are ejected and accelerated to the next dynode. At each successive dynode a cascade of electrons strikes the next dynode. A typical PMT has approximately 9–16 dynodes. Ultimately, a large number of electrons are produced and are collected at the anode. Thus, a measurable current can be detected from a single photon. Gains of  $10^5$ – $10^8$  (electrons per photon) are not uncommon.

Arrays of solid-state detectors are becoming increasingly popular. These include linear photodiode arrays and various ‘panoramic’ or two-dimensional detector arrays such as charge coupled devices (CCDs) and charge injection devices (CIDs). These detectors can monitor multiple wavelengths simultaneously and provide varying degrees of time resolution when needed. While they have minimal gain compared with the PMT, they do have a higher quantum efficiency, i.e., the number of electron/hole pairs produced per incident photon in the solid-state devices is higher than the number of photoelectrons produced per photon incident on the photocathode of a PMT.

## Sample Introduction

For many of the techniques, a nebulizer is the primary means of producing very small droplets of a solution sample or standard that is transported to a high-temperature source where it undergoes desolvation, vaporization, and atomization. While a range of droplet sizes is generated, some means of size discrimination is present to insure that larger droplets (greater than  $\sim 2.5\ \mu\text{m}$ ) are kept from the thermal source. Nebulizers come in a variety of different designs, often serving slightly different functions. Some require very small volume uptake rates (e.g., microconcentric and direct injection nebulizers), others are more efficient at producing small droplets (e.g., ultrasonic nebulizers), and still others are capable of handling viscous and

particle-laden solutions (e.g., Babington and cross-flow nebulizers).

A variety of other approaches are also used including discrete micropipetting of solutions, production of gaseous hydrides, ablation of solid materials using a laser or electrical spark, and even reduction and direct use of metal vapor in the headspace above a liquid such as in the case of Hg determinations. In addition to ablation, which samples a relatively large mass of material from a solid ( $\sim 1\ \mu\text{g}$  per pulse), sputtering processes (e.g., ion beams and various low-pressure discharge processes) have also been used. In general, any approach that permits the introduction of a vapor or small aerosol particles into a source from a solid, liquid, or solution could be considered as a possibility for conducting quantitative analytical atomic spectrometry. Analysis of gaseous samples by direct introduction, of course, can also be done.

## Atomic Emission

### Fundamentals

The primary objective is to have the analyte emit radiation in proportion to the concentration in the sample being analyzed. Resonance lines are typically monitored for maximum sensitivity. For thermal sources in LTE, Boltzmann’s equation (eqn [2]) governs the relative population of the excited states. The intensity of the emission from radiative relaxation from this state is proportional to this population, the atom density in the source being monitored, and a factor that indicates the strength of the emission. This last factor is often referred to as the oscillator strength (which is also proportional to Einstein’s A coefficient for spontaneous emission) and gives an indication of the probability of radiative relaxation from an excited state. The oscillator strength is small for ‘disallowed’ transitions, such as those where an apparent change of spin state in the atom occurs. As a consequence, a low-lying excited state whose population appears to be high may emit a weak spectral line because of a small oscillator strength.

Sources tend to deviate from LTE as the rate of fluctuation in energy input into the source increases, such as in an electrical discharge with a high-frequency electric field used for excitation or when the source is at a reduced pressure and the collision frequency is too low to distribute the energy amongst the various energy modes, e.g., electronic excitation energies versus translational energy distribution. A source at LTE will have all its ‘temperatures’ at the same value. A combustion flame is an example of an

LTE source, while a low-pressure discharge lamp (such as a hollow cathode lamp (HCL)) is not. Similarly, there are some who argue that the inductively coupled plasma (ICP) is not at LTE, possibly as a consequence of the rapid oscillatory energy delivery to the plasma in the megahertz frequency range.

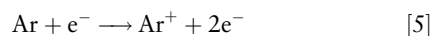
## Sources

The classic source is a chemical combustion flame such as an acetylene–air flame. **Table 2** shows flame temperatures for various fuels and oxidants. In today's instruments, acetylene–air and acetylene–nitrous oxide are the most common fuel–oxidant mixtures. In general, flames are considered relatively low-temperature sources (2000–3000°C) and have their greatest utility in determination of alkali metals and alkaline earth metals. (The cyanogen/oxygen flame shown in the table is considered an 'exotic flame' and is rarely used in analysis both because of the toxicity of the fuel as well as the high production rate of CO as a combustion byproduct.)

Electrical discharges were among the first moderate- to high-temperature sources and included arcs (DC currents of 1–20 A between solid electrodes) and sparks (repetitive, oscillating AC, or DC discharges resulting from discharging a capacitor charged to several thousand volts between two electrodes). Direct current plasmas (DCPs) were also popular for a while. Simply stated, a DCP is an arc where the metal electrodes are positioned at a 45° angle to each other and aspirated solutions are injected into the 'V' formed by the plasma and the upward streaming inert sheath gas. While sparks still entertain an enthusiastic audience in the metallurgical industry, arcs, DCPs, and sparks are generally absent from most analytical labs today.

The 'high-temperature', analytical workhorse for emission spectroscopy is the inductively coupled plasma (ICP). Inducing a current in a flowing stream of ionized Ar gas produces this electrodeless discharge. The rapid movement of electrons caused by

the megahertz frequency in the radio frequency (RF) generator sustains the plasma through secondary ionization of the Ar gas via electron collisions:



The typical power usage of these excitation sources is 1–3 kW. Although they are well sealed to insure no RF leakage, they have nonetheless been assigned two bands (27 and 40 MHz) by the Federal Communication Commission in the US and similar regulatory agencies in other countries. This is done since they employ power levels that one might expect from a medium-sized radio station! ICP temperatures are generally reported to be in the 6000–8000°C range, with temperatures as high as 10 000°C having been reported. While lower-temperature sources (e.g., flames) contain a number of molecular species, many of which emit characteristic bands that block regions of the spectra, the ICP temperatures are sufficient to atomize and ionize nearly all material within the discharge. The spectral features are line spectra of atoms and ions.

In glow discharges a milliamper discharge is established between one electrode and the conducting sample in a low-pressure, inert atmosphere. This gentle discharge produces line-rich spectra with a high degree of excitation and ionization but very low kinetic temperatures (a non-LTE source). Other types of discharges include microwave-induced plasmas (MIPs) that have been operated both at reduced and atmospheric pressures. Generally, an MIP is operated with only a few hundred watts of power and, as a consequence, often does not have the energy needed to vaporize large, condensed-phase particles (e.g., liquid aerosol particles from nebulizers), although some literature reports on success with such samples in addition to direct introduction of gaseous, analyte-containing molecules. The degree of molecular dissociation to form atoms and ions is again high in this type of source, and the spectrum is line rich.

## Spectrometers and Detectors

Simple, low-dispersion monochromators or even interference filters are used for most flame emission applications since few *atomic line* spectral interferences are expected as a result of the limited population of the higher-lying excited states. For high-temperature sources such as ICPs, higher-dispersion spectrometers are typically used. Instruments set up to do simultaneous multielemental analysis can use direct readers with PMT detection. However, most modern detections systems for this type of source for simultaneous multielemental analysis employ a high-dispersion echelle grating spectrometer and an array detector such as a CCD or CID.

**Table 2** Flame temperatures for various fuel/oxidant combinations<sup>a</sup>

Fuel	Oxidant	Temperature (K)
Propane	Air	2300
Hydrogen	Air	2400
Acetylene	Air	2700
Hydrogen	Oxygen	2900
Acetylene	Nitrous oxide	3000
Cyanogen	Oxygen	5000

<sup>a</sup> Temperatures can vary slightly from the values given depending on the fuel:oxidant ratio used. Acetylene-based flames are the most commonly used for analytical spectroscopy.

Moderate- to high-resolution monochromators equipped with computer-controlled scanning and PMT detection is another way of obtaining multi-elemental analysis. In these cases the intensity at the first wavelength of interest is measured, followed by a rapid slewing to the next wavelength region and a slow scan over the second spectral line, slewing to the third line, scanning the third line, etc. The instrumentation is often referred to as a slew-scan monochromator.

## Analysis

As noted earlier, flame emission has its biggest niche in analysis of the alkali metals and alkaline earth metals. As a result, they are often found in such areas as clinical laboratories, where such elements as Na and K are of routine interest. The technique is quite mature, and many proven methods for a variety of analyte-matrix combinations are available.

ICP optical emission spectrometry (ICP-OES) is also well established, and numerous procedures exist for a host of complex sample types. It is not uncommon to employ internal standards for ICP-OES. This is particularly useful to compensate partially for variations in sample delivery rate and, to a lesser extent, for small fluctuations in the plasma. The technique is relatively free of chemical interferences, and most spectral interferences can be corrected for. In addition to providing parts per billion to parts per million limits of detection for most of the periodic table with precisions of a few per cent or less, ICP-OES can also provide five to six orders of magnitude in analytical dynamic range.

Standard additions in lieu of simple aqueous standards and a calibration curve may be used in instances of very complex matrices. When the approach is used by adding a small volume of a standard to the unknown solution in the course of analysis to validate the legitimacy of the slope of the calibration curve (i.e., the sensitivity), the term 'spike' is applied.

## Atomic Absorption

### Fundamentals

Atomic absorption (AA) is based on 'stimulated' absorption since the number of photons absorbed is dependent on the photon flux of the incident light. However, the fraction of the light absorbed is relatively independent of source intensity until the incident intensity reaches extremely high levels and begins to populate substantially the excited state. At that point the probability of absorbing a photon

is no longer significant relative to having a photon re-mitted by stimulated emission from the high-lying excited state. Under the normal, lower light level operating mode, the relationship between absorbance and concentration is similar to the Beer-Lambert law used for solution absorption, with minor modification of terms:

$$A = k_{\text{abs}} \times C \quad [6]$$

where  $k_{\text{abs}}$  is the absorption coefficient and  $C$  is the analyte concentration; as in solution spectrophotometry, the linear relationship holds as long as a number of assumptions hold. In particular, it must be assumed that  $k_{\text{abs}}$  is singularly valued over the wavelength interval being monitored, i.e., the width of the incident line is much narrower than the width of the absorbing wavelength. When this is not the case, curvature of the calibration curve ( $A$  versus  $C$ ) results and is most pronounced at larger absorbance values. While the line widths of light sources typically used in AA (e.g., discharge lamps or HCLs) are narrow, so is the absorbing line width, and linearity is not assured over a large range of concentrations. Similar nonlinear responses can be observed if the spectrometer bandpass encompasses more than one spectral line from the incident light source or if stray light in the spectrometer is significant.

Since the typical discharge light source usually derives its spectrum from the metal that is used in fabricating the source, AA provides a high degree of elemental selectivity. This selectivity and general absence of spectral interferences from other atomic species is further enhanced by the requirement that the absorbing state must be populated, which means that resonance transitions originating in the ground state are the only absorbing lines that show any measurable absorbance. Similarly, Boltzmann's distribution shows that for most elements and most AA 'atomic cells', 99 + % of the atoms are in the ground electronic state, and this condition does not change with relatively wide variations in temperature. This places much less stringent requirements than for emission spectroscopy on the source temperature. In general, the primary role of the energy in an AA atom cell is to insure efficient – or at least consistent – free atom production.

### Sources

AA requires two 'sources': one in which free atoms are produced or the sample is 'atomized' (e.g., flame) and one that produces the radiation that is to be absorbed by these atoms. To avoid confusion, these will be referred to as 'atomizer' and 'radiation source' in the following discussion.

**Atomizer** As noted above, the primary function of the atomizer is to make isolated atoms in the gas phase efficiently. There is no need for excitation, and it would be undesirable, generally, to ionize the analyte.

In the case where the sample is introduced as a solution aerosol, the atomization source must evaporate the solvent, vaporize the resulting salt particles, and dissociate any analyte-containing molecules – tasks that are similarly required when wet aerosols are used in emission spectroscopy. Since collisions with high-temperature gases are the most efficient means of accomplishing the first two tasks, thermal sources such as flames are typically used (e.g., flame AA).

Small tube furnaces also accomplish these tasks efficiently by drying out a discretely deposited solution sample ( $\sim 10\text{--}20\mu\text{l}$ ) that is placed on the surface of the tube. For some samples where the analyte is not prematurely vaporized, some of the desolvated salt can be thermally decomposed (and sometimes vaporized) during a ‘char’ or ‘thermal pretreatment’ cycle in the furnace. This effectively eliminates some of the matrix prior to the ‘atomization’ heating cycle when the analyte vapor is introduced into the optical path. These atomizers are often referred to as graphite furnaces or more generally as electrothermal atomizers (ETAs). They exhibit an advantage over flames because the atoms have a significantly longer residence time in the absorbing volume. This  $\sim 100$ -fold increase yields an  $\sim 100$ -fold improvement in sensitivity. Additionally, the entire sample dosed into the ETA has the potential of being introduced into the observation volume. In contrast, the nebulizer and aerosol delivery system in a flame are only  $\sim 5\%$  efficient. Thus, even at a  $5\text{ ml min}^{-1}$  aspiration rate, the flame receives the sample at only  $\sim 5\mu\text{s}^{-1}$ , in contrast to the ETA, where  $20\mu\text{l}$  is introduced, albeit as a transient pulse.

Another common atomizer is a simple heated quartz tube into which a gaseous metal hydride is introduced, generally produced by reaction of the sample with a strong reducing agent such as sodium borohydride ( $\text{Na}_2\text{BH}_4$ ). These volatile metal hydrides are relatively labile and can be dissociated at relatively low temperatures to yield the free metal. In addition to a quartz tube, the hydride generation technique can also be used by introducing the hydride directly into a flame. While applicable to a number of metals, hydride generation finds its most common use with the semimetals (e.g., Se, As, Te, In, Bi) and a few others (e.g., Pb, Sn, Sb, Cd).

Mercury is unique since its elemental form has a moderately high vapor pressure even at room temperature. Thus, by adding a reducing agent to a

sample and purging the solution with a stream of gas (Ar or  $\text{N}_2$ ), the free  $\text{Hg}_{(\text{g})}$  can be pumped through a quartz tube with optical windows and absorbance measurements made. This is often referred to as a cold vapor technique and is almost exclusively applied to Hg determinations where parts per trillion detection limits have been reported.

**Radiation sources** The most common radiation source used with AA is a HCL, which consists of a tubular-shaped cathode made of the metal of interest and a simple anode. A milliamp DC discharge is established between these electrodes in a low-pressure noble gas environment. The discharge results in a very line-rich spectrum of the cathode material. Electrodeless discharge lamps (EDLs) are brighter and require a microwave power supply. For elements whose HCL lines are weak, EDLs are often the lamps of choice.

There are two other sources worth noting, although they are currently used in a very small fraction of the instruments employed. Continuum sources can be used if their intensity is sufficient to minimize noise levels and if the spectrometer has sufficient dispersion to make the spectral bandpass comparable with the absorbing line width. While feasibility has been demonstrated in research laboratories, there currently is no commercial instrument available. At the other extreme, using a very bright, stable source with a narrow line width has produced viable absorbance readings that are two to three orders of magnitude below those available with HCLs and EDLs. The source that provides this detection enhancement is a tunable diode laser.

### Spectrometers and Detectors

The spectrometer is generally a monochromator of moderate to low resolution since it is only necessary, in most cases, to isolate the resonance line of interest from the other atomic lines in an HCL or EDL. The PMT is typically used for detection.

The optics are generally moderately complex and of a double-beam design using modulated lamps with a chopper to measure all signals needed to calculate absorbances while accounting for lamp drift, emission from the atomizer, background absorbance/scatter, and any drift or dark current in the detector.

### Analysis

AA is useful for most of metals and semimetals, with flames and ETAs providing parts per million and low parts per billion limits of detection, respectively. Both are relatively mature techniques, with a variety of methods for handling various analyte–matrix

combinations. Generally, the ETA has to accommodate complex background signals that can arise from the matrix. As a result, background correction techniques are available on all commercial instruments.

AA is generally considered a 'single element technique' where one analyte is determined at a time. However, high degrees of automation make it relatively simple to deal with a large number of samples, standards, and analyte elements with minimal to no operator attention. There are also some instruments now available that permit a limited number of elements (four to eight) to be determined simultaneously, and a system using a continuum source and high-dispersion spectrometer with array detection can provide true simultaneous multielement capabilities.

## Atomic Fluorescence

### Fundamentals

Similar to AA in its requirements for operation, atomic fluorescence (AF) requires free gaseous atoms in the ground state. Population of an excited state is accomplished by absorption at a wavelength corresponding to a resonance transition. Once in the excited state, it can radiatively relax – like atomic emission – and the intensity of this fluorescence is used for quantitation.

Both resonance and nonresonance fluorescence can be used. The emitted resonance fluorescence is shown in **Figure 2** as  $h\nu$  and the nonresonance fluorescence as  $h\nu'$ . Nonresonance fluorescence is often preferred since the detected wavelength is different from the excitation wavelength, and the scatter of the exciting beam in the atomizer can be more easily excluded by the monochromator. Unlike AA, where a small decrease in intensity (which has fluctuations or noise inherent to it) is the key to determining the limits of detection (LODs), AF searches for the appearance of photons against a 'dark background'. This has the potential of yielding significantly improved LODs.

In contrast to AA, a brighter light source will yield better sensitivities since the signal is proportional to the concentration of the excited state atoms. If the intensity of the source is sufficient (e.g., laser radiation), the population of the excited state can be brought to saturation. In addition to maximizing the fluorescence intensity, this also makes the fluorescent intensity insensitive to small fluctuations in the excitation source intensity.

As with emission spectroscopy, anything present in the sample that alters the degree of radiative relaxation will impact the signal and result in 'interference effects'. In AF, the 'term quantum yield' is used to

describe the fraction of the excited state that relaxes radiatively. A good absorber with a high quantum yield should provide optimal sensitivity.

### Sources

**Atomizers** The same requirements that exist for AA also exist for AF. Consequently, similar sources have been used. However, since the long path length needed with AA to maximize the sensitivity is not needed with AF, ICPs and more circular flames have been used in place of the traditional chemical combustion flame with a slot burner. ETAs have also been used to enhance the sensitivity for AF as has been done for AA, and the resulting LODs are some of the best for the atomic spectroscopic suite of techniques when combining an ETA with laser excitation.

**Radiation sources** As in AA, line sources are typically used, although high-intensity sources are much more critical in AF. Boosted HCLs and EDLs have been employed. However, tunable lasers certainly provide the optimal sensitivity, i.e., laser excited atomic fluorescence spectrometry.

### Spectrometers and Detectors

Since the fluorescence spectrum should be relatively simple because of the selective radiative excitation of only the analyte, monochromators with only limited resolution are needed. Detectors that can be used in a photon counting mode (i.e., can detect individual photons) optimize the LODs. For this reason, PMTs are commonplace in AF.

### Analysis

Although research papers related to AF continue to appear, practical applications are less common than for the other techniques. As mentioned previously, when combined with an efficient atom cell (e.g., ETA) and a bright light source (e.g., laser), AF can provide exceedingly low detection limits.

This overview on analytical atomic spectrometry touches on the basics of three dominant methods of conducting optical spectroscopy for the purposes of qualitative and quantitative elemental analysis. There are a number of variations in sources, atom cells, dispersive devices, etc. that have not been discussed. As an example, laser-induced breakdown spectroscopy employs a high-intensity laser to ablate samples where the extreme radiant energy also produces a plasma that ultimately produces electronic excitation of the ablated material. Similarly, there are a number of nonoptical approaches that represent variations of some of these schemes that have



produced success. For example, laser-enhanced ionization uses photon absorption combined with thermal excitation to produce enhanced ionization, which ultimately causes a change in the current flowing between a pair of electrodes inserted within the excitation source such as a flame.

While the general approaches to elemental analysis using optical spectroscopy that have been discussed in the text are currently dominant in analytical laboratories, the wealth of options of producing atoms/ions and detecting their presence in order to identify the elemental composition of a sample keeps the field vibrant and foretells the likelihood of new, improved methodologies. The ultimate objective of analytical developments in this area is to devise a means of conducting elemental analysis faster, cheaper, and

more accurately on a wider variety of sample types. Improved precision, sensitivity, and elemental coverage round out the objectives sought for the ideal instrument.

*See also:* **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Principles and Instrumentation. **Atomic Fluorescence Spectrometry.**

### **Further Reading**

Ingle JD, Jr. and Crouch SR (1988) *Spectrochemical Analysis* 588p. Eaglewood NJ: Printice Hall.

Skoog DA, Holler FJ, and Nieman TA (1998) *Principles of Instrumental analysis*, 5th edn. Saunders College Publishing.

## **AUGER ELECTRON SPECTROSCOPY**

*See* **SURFACE ANALYSIS: Auger Electron Spectroscopy**

# B

## BACKGROUND CORRECTION

See **ATOMIC ABSORPTION SPECTROMETRY: Interferences and Background Correction. ATOMIC EMISSION SPECTROMETRY: Interferences and Background Correction**

## BIOASSAYS

Contents

**Overview**

**Microbial Tests**

**Bioautography**

### Overview

**J Chayen and L Bitensky**, University of Surrey, Guildford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 341–347, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

### What Constitutes a Bioassay?

As Gaddum in 1953 has pointed out, the term bioassay 'may be applied to any experiment in which the potency of a drug is measured by effect on living organisms or tissues'. (In this definition, 'hormone' can be inserted in place of 'drug'.) Gaddum's review of the various types of bioassay and of the mathematical evaluation of such bioassays is still cardinal to this type of investigation. The two more recent fundamental reviews on the mathematical handling of such assays are those of Emmens, 1969 and that found in the European Pharmacopoeia, 1971.

Before progressing further, it is essential to define the term 'bioassay'. In agreement with Gaddum, the term covers those assays that involve either whole animals or isolated tissues and is extended to include assays done on isolated cells. Assays performed on fragments of cells, such as isolated membranes or other cell fragments, are not included, since these

deal with the immediate binding of the hormone, not the intracellular effect of the hormone. Assays performed on isolated cell receptors are also excluded: the same receptor will bind specific inhibitors of the hormone and it will be difficult to differentiate this binding from that of the hormone. The essence of a bioassay is as follows: the hormone (or drug) binds to its specific site, or receptor; when sufficient concentration of the hormone is bound, it stimulates a response. The immediate response might be an activation of some 'second messenger' system, such as the formation of cyclic adenosine monophosphate (cAMP). But this 'messenger' is produced by many stimuli: it is not specific to this hormone, whereas the final intracellular response is indeed specific to the hormone. Therefore, a true bioassay must involve at least whole cells, preferably in intact tissue, responding to a specific intracellular response.

### Validation of Bioassays

A bioassay involves measuring how well a preparation of the hormone, or a sample of plasma, activates a biological function in its target tissue; this is compared with how well graded concentrations of a standard preparation of that hormone perform that function.

For example, intravenous injection of graded concentrations of the adrenocorticotrophic hormone (ACTH) into hypophysectomized rats caused a graded loss of ascorbate from the adrenal glands.

Improvements to this technique, with the retrograde injection of the test, and standard, material into the adrenal gland improved the sensitivity of this assay to  $100 \text{ pg ml}^{-1}$ , which was just sufficiently sensitive to detect pathologically elevated circulating levels of this hormone. Modification of the procedure used on isolated cells improved the detection level to  $1 \text{ pg ml}^{-1}$ .

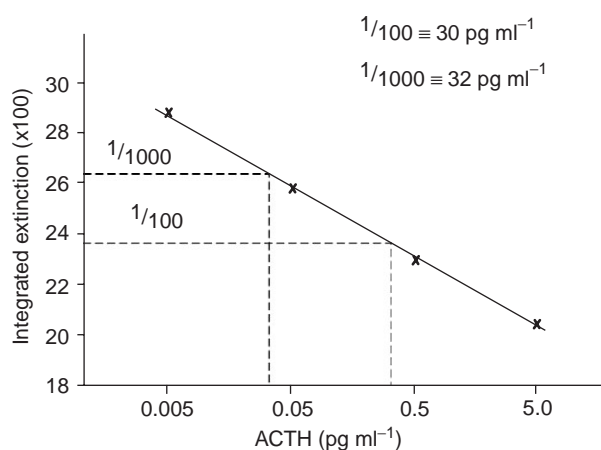
The cytochemical bioassay (discussed later) involves six segments of the adrenal glands of a guinea pig. Four segments are used to produce a calibration graph of the effect of logarithmically graded concentrations of a standard preparation of the hormone and the other two are exposed to different concentrations of the plasma (or other type of sample) to be tested; the dilutions of the plasma are normally 1:100 and 1:1000. The limit of detection by this bioassay is  $0.005 \text{ pg ml}^{-1}$ . A typical cytochemical bioassay of ACTH in a sample of human plasma is shown in Figure 1.

### Statistics and Probability

The basic criterion of a bioassay is the index of precision ( $\lambda$ ). This depends on the slope of the logarithmic dose-response graph ( $b$ ) and the standard deviation ( $s$ ) of the points from the line (obtained by subtracting each value of  $y$  from its recorded value  $y_c$ ):

$$s = \frac{\sum (y - y_c)^2}{n - 1} \quad [1]$$

where  $n$  is the number of readings. From this value one can calculate  $\lambda = s/b$ . (It may be noted that, in some publications, the term  $s_{yx}$  is used instead of  $s$ ).



**Figure 1** Result of a typical cytochemical bioassay, the assay of ACTH in human plasma. Four segments of the adrenal glands of one guinea pig are exposed to graded concentrations (crosses) of a standard preparation of the hormone in T8 medium. Two segments are exposed to one of two concentrations of the plasma in T8 medium (broken lines); when corrected for dilution, the values are consistent.

Generally, when  $\lambda = 0.3$ , the log dose estimated from a single observation has a standard error of 0.3. Consequently, within the limits of a single standard deviation (SD: 68% confidence limits) this single observation would be in error to plus/minus the antilog of 0.3, namely two times or one-half the value (50% to 200%) produced by the assay. Surprisingly, this is often considered the upper limit of an adequate assay. However, most bioassayists would aim towards a value of  $\lambda$  of 0.1, which implies that a single observation would have a 68% probability (i.e. one standard deviation) of being from 79% to 126% of the correct value. If one wants more precision, namely two standard deviations (95% confidence limits), a  $\lambda$  of 0.1 would give 63% to 158% of the true value.

A more exact indicator of the precision of a bioassay is given by the fiducial limits. To calculate these generally requires at least two dilutions of the test material and two or three of the standard reference compounds (as in Figure 1). These calculations take into account the degree of parallelism of the response of the test as against that of the standard (also see below). The equations for calculating these are given by Gaddum (1953) and by the European Pharmacopoeia (1971).

### Other Factors Involved in Assessing Bioassays

**Accuracy** In terms of bioassay, the accuracy of an assay is given by the percentage recovery of the pure substance that was added to the sample before the assay was done. In normal practice the hormonal content of an aliquot of a sample of plasma is first determined. Then a known amount of a standard preparation of the hormone (e.g.  $100 \text{ mU ml}^{-1}$ ) is added to the sample and a second aliquot is assayed. If, for example, the first sample was assayed at  $10 \text{ mU ml}^{-1}$  and the second at  $105 \text{ mU ml}^{-1}$ , the percentage recovery was 95%.

**Sensitivity** Sensitivity is defined by the smallest change in concentration of the hormone, or its lowest concentration, that can be measured significantly different from 'background' by the assay system.

**Variability** Variability essentially tells you how much reliance can be placed on the bioassay system. Ideally a single sample should be assayed on a number of occasions: to what extent do the results vary? It should also be remembered that some specimens may deteriorate with time of storage: this is also useful information.

**Specificity and Parallelism** The results given by two or three different concentrations of the test material

are plotted against those obtained from similar concentrations of the standard preparation (as in **Figure 1**). Generally, whether the responses are parallel can be determined by simple inspection. For greater precision mathematical methods can be used. It is only when the graphs are parallel that it can be assumed that the test material is likely to be the same as that in the standard preparation. This is a much more rigorous criterion than is used in immunoassays.

## Do We Need Bioassays?

All biologically active molecules are characterized by their biological activity as measured by a 'functional assay'. In contrast, there has been a tendency to rely on 'analytical assays', including saturation assays and physicochemical analytical procedures such as high-pressure liquid chromatography. The argument is that since hormones are specific chemical moieties, they should be analyzed by the highly sophisticated analytical procedures that have been developed recently. These usually are rapid and automated, and many can be done by relatively unskilled operators, whereas bioassays are more cumbersome and much more time-consuming.

The surprising fact is that bioassays are chemically more discriminating than are the new analytical procedures. This arises from the fact that quite minor changes in the hormone molecule can produce major physiological effects. For example, the terminal 1–34 part of the parathyroid hormone molecule (1–34 PTH) is fully active on bone and on kidney but desamino 1–34 PTH has only 1% of this activity in the renal adenylcyclase assay and 50% of the 1–34 PTH activity in the *in vivo* bone hypercalcemia assay. Further modification of the PTH molecule (3–34 PTH) removes all biological activity. It might be argued that such a change in the molecule would be detected by analytical methods. However even slight alteration in what is normally regarded as the 'biologically inert' region of the PTH molecule, changing aspartate in place of asparagine at position 76, can cause almost complete loss of PTH-like activity.

This highlights a worry about the current widespread use of immunoassays in that such assays do not monitor each part of the whole molecule. Even with immunoassays that have antibodies directed to several regions of the hormone, it is likely that a very small change, such as the asparagine residue to an aspartate residue, would not be noted, yet the striking loss of biological activity that this has been shown to entail would certainly be noted by bioassay. In fact there was a fairly recent, widespread example of this type. There was a great deal of

concern about a bone peptide, osteocalcin, and whether it had a role in ossification. The evidence, from radioimmunoassay of circulating levels of this peptide, was entirely contradictory. However, it was then realized that the important role of this peptide depended on the carboxylation of three glutamate residues to  $\gamma$ -glutamate (Gla): it was this carboxy-glutamate that was involved in binding calcium, not the normal glutamate residues, but the radioimmunoassays at that time were incapable of distinguishing between the biologically inactive and the biologically active (Gla-form) of osteocalcin.

There are still several entities that have considerable biological activity but for which no immunoassay is available. The most commonly found example of this is the long-acting thyroid-stimulating immunoglobulin of Graves disease, which can be assayed by the same cytochemical bioassay as is used for measuring circulating levels of thyroid-stimulating hormone except that the response is very much delayed. Equally, the levels of thyroid growth-stimulating immunoglobulins, associated with goiter, and the immunoglobulins that are related to the blockade of thyroid function in myxoedema have to be measured by suitable bioassay. Other examples include the assay of the anti-parietal cell immunoglobulins of pernicious anemia and some, as yet uncharacterized, natriuretic factors.

## Function of Bioassays

### Testing the Functional Potency of Hormone Preparations

The World Health Organization (WHO) Expert Committee on Biological Standardization is responsible for establishing international standard and reference preparations of hormones. The activity of other preparations of the hormone, or of the hormone in the circulation, is referred to this standard. That committee recognized that 'a limitation on the use of immunoassays for evaluating hormonal bioactivity is that the methods measure a composite of antigenic activity, which is not necessarily related to the bioactivity of the hormone'. Several excellent preparations of hormones have been obtained by the WHO only for it to be found that, while they gave high values by immunoassay, they had little effect in bioassays. Such preparations become labeled 'for immunoassay only'. This problem became more severe when polypeptide hormones began to be prepared from human tissue. There was too little of the final, purified hormone for it to be assayed by conventional bioassay and the WHO Expert Committee called for the development of new, very micro

bioassays to cope with this problem. That was the basis of the development of the cytochemical bioassays (discussed later), which required only 10 pg for each assay.

### **Discrepancy between Clinical Assessment and Immunoassay**

If a clinician suspects a hormonal disorder, the circulating level of that hormone in the patient needs to be measured. This can be done either by a functional procedure (i.e. bioassay) or by an analytical method such as immunoassay. Since a hormone is defined and recognized by its biological function it would be reasonable to measure its presence by bioassay, which depends on its function. This is measured in terms of units of activity relative to that of a standard preparation of that hormone. On the other hand, bioassays are laborious and time-consuming relative to immunoassays, which measure how much of that hormone molecule is present. In general, bioassays and immunoassays have produced similar results, so that it is obviously expedient to use the latter. However, there are many instances in which the two results have proved very discrepant. There are examples where immunoassay has given misleadingly high results: in some instances it has included in its measurements molecules such as 'big gastrin' or 'big ACTH', which are relatively inactive biologically; in others, immunoassay measured biologically inactive fragments of the hormone, especially where these have a long half-life in the circulation, as occurred with assays of parathyroid hormone in secondary hyperparathyroidism.

An unusual example concerns the presence of a PTH-like factor in the circulation of cancer patients. The clinical condition indicated a considerable concentration of PTH-like activity, associated with loss of bone and hypercalcemia (the hypercalcemia of malignancy). Immunoassay detected no elevated PTH levels; only the cytochemical bioassay was able to show high concentrations of PTH-like activity, although such activity did not show true parallelism. The material causing this effect has now been isolated and purified: it is known as the PTH-related peptide.

### **The Study of Hormonal Effects**

Often the name of a hormone is related to the first effect that the hormone has been shown to influence. For example, although prolactin may indeed have effects on the lactating breast, it has far-reaching influences other than this. Consequently, it is often valuable to study the cellular mechanisms influenced by a hormone other than that activity by which it is

generally named. This type of investigation necessarily depends on some form of bioassay; the cytochemical bioassay system is particularly useful because of the range of activities that can be monitored by it. A particular use of these methods has been the analysis of the peculiarities involved in the condition known as pseudohypoparathyroidism type I. (It may also be related to nutritional vitamin D deficiency). In this the often huge discrepancy between immunoassayable PTH-like material and very low biologically assayable PTH has been shown to be due to material that is detected by immunoassay but which, in fact, inhibits true PTH activity. The inhibiting material has now been isolated.

### **The Information That can be Obtained by Bioassay**

Some critics of bioassays have been concerned that the bioactivity of a hormone may vary depending on how the hormone is administered to the whole animal, for example intravenously, intraperitoneally, or intramuscularly. Another criticism is that *in vitro* bioassays do not always distinguish between sialated and desialated forms of the hormone. In the authors' view, these criticisms show a fundamental ignorance of the purpose of bioassays. Thus, a bioassay should measure the concentration of the hormone that, acting on the relatively intact target tissue (or cells), produces a measured response. In practice, this is related to the response produced by a standard preparation of that hormone acting under identical conditions. The fact that a hormonal preparation becomes modified in the circulation should be immaterial to the bioassay of that hormone: that is a question for a physiological study.

### **Whole-Organism Bioassays**

The most widely used whole-animal bioassay consists of variants of the original McKenzie assay for the thyroid-stimulating hormone (TSH), which provides a type-example of bioassays in intact animals.

The general procedure is as follows. Female mice, of about 15 g, are fed a low-iodine diet for a week. They are then injected intraperitoneally with 5–8  $\mu\text{Ci}$  of radioactive iodine ( $^{131}\text{I}$ ). Thyroxine (10–15  $\mu\text{g}$ ) is injected 4, 24, and 48 h after this injection. Twenty-four hours after the last injection of thyroxine, the samples to be analyzed, namely concentrations of the unknown and of a standard preparation of TSH, are injected into the tail vein of the mice. (In some modifications, thyroxine is added to the feed, or to the drinking water). At the time the TSH was injected,



and 2 h later, blood samples are taken for determining the radioactivity present. The percentage increase in radioactivity, with increasing concentrations of the preparations, is plotted on a log-dose scale. It should be linear over a useful range. The method detects 0.25 mU of TSH activity (reported to range from 0.17 to 0.29). However, as with most 'whole-organism' bioassays, the method is not sufficiently sensitive to measure plasma levels within the normal range.

By determining the radioactivity in blood samples 9–24 h after the material to be assayed has been injected, it is also possible to assay the long-acting thyroid-stimulating immunoglobulin (LATS).

Both  $2 \times 2$  and  $2 \times 3$  assay designs have been used, with 6–12 mice for each dosage point.

Other *in vivo* bioassays, including those for the follicle-stimulating hormone (FSH), for the luteinizing hormone (LH) and for the growth-stimulating hormone (GH), have been described but are not used much nowadays.

## Isolated-Cell Bioassays

Cells have been isolated from many organs, including the pituitary, adrenal and thyroid glands, the testes and the corpus luteum. They have been used, without being maintained *in vitro*, for assaying the relevant hormones. Possibly the most used of such bioassays are those for LH and the structurally similar human chorionic gonadotrophin. The isolated cell bioassay of LH is used here as an example. This is an unusual isolated cell bioassay because it does assay the plasma directly; in most other assays the isolated cells have become modified so that plasma is cytotoxic to them.

Small pieces of adult mouse testis are placed in Eagle's medium with 2% calf serum (6 testes/50 ml) and stirred for 10 min to liberate the cells. The medium is then filtered to remove pieces of tissue. The filtrate, containing the isolated cells, is preincubated (1 h at 37°C in an atmosphere of 93.5% O<sub>2</sub>–6.5% CO<sub>2</sub>) in a shaker at low speed. The cell dispersion is then placed in ice and centrifuged for 5 min at 4°C at low speed. The supernatant is discarded and the cells are resuspended in Eagle's medium containing 2% calf serum.

Cell suspension (0.1 ml, containing  $6 \times 10^4$  cells) is added to tubes, in ice, that contain 0.1 ml of the appropriate amount of the gonadotrophic hormone in Eagle's medium containing 2% calf serum. The samples (test and controls) are incubated at 34°C for 3 h, at 80 rpm in the O<sub>2</sub>:CO<sub>2</sub> atmosphere. The amount of testosterone produced is measured by

radioimmunoassay; the response is rendered linear by suitable mathematical adjustment.

The method appears to be highly specific, with an index of precision of 0.044 (19 assays). It is also sufficiently sensitive for assaying circulating levels of the hormone. In samples taken over the entire menstrual cycle, the biological activity measured by this assay was consistently ~5.5 times that recorded by immunoassay.

## Isolated-Tissue Bioassays

### Background

The main isolated tissue bioassays are the cytochemical bioassays (CBA). These are  $\sim 10^3$  times more sensitive than the equivalent immunoassays, so that they readily measure normal, and subnormal, circulating levels of the hormones. Because of their sensitivity, very little plasma is required; even a heel-prick from a neonate is sufficient. They are very specific in that they depend first on the specific recognition of the hormone at the surface of the target cells, and then on the provocation of the relevant 'second messenger' to transmit the message to the specific intracellular metabolic system that will give rise to the biochemical mechanism by which that hormone is normally recognized. These bioassays measure the altered metabolic system. The considerable sensitivity of these bioassays arises from a number of factors: (1) they are 'within-animal' assays, so obviating the variability imposed by using several animals for each point; (2) the hormone is not diluted into the large volume of circulating blood, as in whole-animal assays, but is applied directly to the target organ; and (3) the metabolic activity within the target tissue is allowed to recover from such previous stimulation as may have occurred inside the test animal before the hormone (or the plasma) is allowed to act on the target cells.

### Procedure

Small segments (e.g., one-third of a lobe of a thyroid gland for the TSH assay) of the target tissue of a suitable animal (normally a guinea-pig or a rat) are placed on lens tissue on a metal grid table in a vitreosil dish in a culture vessel (**Figure 2**). Trowell's T8 medium is added to the vitreosil dish until it reaches the lens tissue, but does not overflow it. This medium contains amino acids and a balanced mixture of salts and is suited to these nonproliferative organ-maintenance cultures. The culture vessels are gassed with a mixture of 95% O<sub>2</sub>–5% CO<sub>2</sub> and sealed. They are left for 5 h to allow the tissue to recover from the trauma of excision and from the effects of



**Figure 2** The vessel for organ culture. The method is described in the text.

previously circulating hormones. This allows the metabolic activity of the cells to revert to basal level. Then the culture medium is replaced by a fresh sample of T8 containing one of four dilutions (e.g.,  $5 \text{ fg ml}^{-1}$  to  $5 \text{ pg ml}^{-1}$ ) of the standard preparation of the hormone or of one of two dilutions of the plasma, normally at 1:100 and 1:1000 concentrations in T8 medium (as in **Figure 1**). The use of unextracted plasma obviates any degradation of polypeptide hormones that could occur during the separation of the serum and also the possible release of active moieties such as polyamines from the blood cells. The duration of the exposure of the segments to the concentrations of the standards or of plasma to achieve the first, rapid response to the hormone is normally very short: for example 4 min in the ACTH assay or 8 min in the assay of parathyroid hormone. The tissue is then chilled to  $-70^{\circ}\text{C}$  for 1 min and transferred to cold, dry tubes at this temperature for storage, which may be a matter of a day or two.

Sections are cut in a cryostat fitted with an automatic cutting device to ensure sections of constant thickness (to within  $\pm 5\%$ ) with the cabinet temperature at  $-25^{\circ}\text{C}$  and with the knife cooled further with solid carbon dioxide. The sections are flash-dried on to glass slides and reacted by quantitative histochemical methods for the biochemical activity that is typical of the effect of the hormone on its target cells. These methods have been adjusted to ensure no loss of material. The colored reaction

product is then measured, solely in the target cells, by scanning and integrating microdensitometry.

### Section Bioassays

Part of the strength of the cytochemical bioassays is that they are 'within-animal' assays. But normally only six, or at the most eight, segments can be obtained from one animal, four being required for the standard graph and the rest for the two dilutions of the plasma from one, or at the most two, subjects. To increase the 'throughput' of some of these bioassays, it was found possible to use sections, instead of segments.

In these section bioassays, relatively large segments of the target-organ are maintained for 5 h, as for the segment bioassays. The segments are then chilled to  $-70^{\circ}\text{C}$  and relatively thick sections that enclose whole cells are cut (e.g., at  $20 \mu\text{m}$ ). These sections, under suitable stabilizing conditions, are then exposed to the various concentrations of the standard preparation of the hormone or to one of two concentrations of samples of plasma from a number of subjects. The response now occurs in a matter of tens of seconds rather than of minutes. The cytochemical reactions and measurement, the response and the sensitivity, are the same as for the segment bioassays.

### Trends in Bioassay

Routinely, when hormonal assays are done to support a diagnosis, immunoassays are ideal because they are fast, require relatively little skill, and can be done on a large scale. When the results of such assays are at variance with the clinical picture, bioassay becomes essential. A particular recent example of this concerns a new, highly specific immunoassay of luteinizing hormone (LH). In some individuals this gave zero values even though there were indications that these values should have been high. This was then found to be caused by the presence of a molecular variant of LH that assayed normally by bioassay, so emphasizing the importance of bioassay in exceptional cases and in research. Bioassay becomes important also with respect to immunoglobulins that exert endocrine influence, such as those that affect the thyroid gland. It is also important in detecting inhibitory influences that can readily be detected by bioassay. This is done by first doing a straightforward bioassay on the plasma and then adding two known concentrations of a standard preparation of the hormone. The amount of this added hormone that is 'recovered' in the bioassay will indicate whether there is inhibitory material in the plasma.

It may well be that this detection of such material will be a major use of bioassays.

A particular advantage of bioassay is that it can assay material, such as the various thyroid-stimulating or inhibitory immunoglobulins that have eluded other forms of assay. The cytochemical bioassays have practical benefit in that the same expertise and equipment can be used for assaying biologically active moieties, whether they be polypeptide hormones or endocrinologically significant immunoglobulins or factors that inhibit endocrine function.

## Acknowledgments

The authors are grateful to Dr Alaghband Zadeh for his advice and help with this article, and to Miss A. O'Farrell for her help with it.

*See also:* **Blood and Plasma. Hormones:** Steroids. **Immunoassays, Applications:** Clinical. **Quality Assurance:** Reference Materials.

## Further Reading

Chayen J and Bitensky L (1988) Bioassay. In: Williams DL and Marks V (eds.) *Principles of Clinical Biochemistry*, 2nd edn., pp. 440–447. Oxford: Heinemann Medical.

Cohen SM, Robinson D, and MacDonald J (2001) Alternative models for carcinogenicity testing. *Toxicological Sciences* 64: 14–19.

European Pharmacopoeia (1971) *Statistical Analysis of Results of Biological Assays and Tests*, vol. II, pp. 441–498. Paris: Council of Europe, Maisonneuve S.A.

Goltzman D, Stewart AF, and Broadus AE (1981) Malignancy-associated hypercalcemia: Evaluation with a cytochemical bioassay for parathyroid hormone. *Journal of Clinical Endocrinology and Metabolism* 53: 899–904.

Hoffman DJ, Rattner BA, Burton GA, and Cairns J (eds.) (2003) *Handbook of Ecotoxicology*. Boca Raton: Lewis Publishers.

Keddy CJ, Greene JC, and Bonnell MA (1995) Review of whole-organism bioassays. *Ecotoxicology and Environmental Safety* 30: 221–251.

Loveridge N, Fischer JA, Nagant de Deuxchaisnes C, *et al.* (1982) Inhibition of cytochemical bioactivity of parathyroid hormone by plasma in pseudohypoparathyroidism type I. *Journal of Clinical Endocrinology and Metabolism* 54: 1274–1275.

Recio L and Everitt J (2001) Use of genetically modified mouse models for evaluation of carcinogenic risk: Consideration for the laboratory animal scientist. *Comparative Medicine* 51: 399–405.

Robertson JL (1992) *Pesticide Bioassays with Arthropods*. Boca Raton: CRC Press.

## Microbial Tests

**N Christofi**, Napier University, Edinburgh, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Bioassays are methods that utilize living materials to detect substances and/or determine the potential toxicity of chemicals or contaminated matrices. They are widely used to screen for potential hazardous chemicals in contaminated soils, potable and wastewater, foods, and other materials. Bioassays are important in toxicology, which is the 'study of toxic substances and their effect on organisms and the environment', and are used in assessing acute and chronic toxicity. The general purpose of toxicity testing is to obtain information for use in the management of the release of toxic substances into the environment. Whether a substance is toxic or not depends largely on two factors: the test system being used and the concentration of the chemical. This implies that most substances are potentially toxic if

they are present in the environment at sufficiently high concentrations and it is known that many essential micronutrients are also toxic at high doses. A large number of toxicity bioassays have been used in ecotoxicology. These include rapid screening tests involving microorganisms to multispecies testing utilizing microcosms and mesocosms embracing bacteria, algae, invertebrates, and vertebrates. This article deals with microbiological toxicity tests.

## Analyses of Toxicants

Toxic compounds in natural samples can be detected either through chemical analyses or through bioassay. A disadvantage of the first detection method is that it requires sophisticated instrumentation to determine contaminants in a sensitive and accurate way although often in complex effluents, for example, contaminants are not known and it is therefore not easy to carry out adequate analysis. Some bioassays provide quantitative information on polluting substances but most determine the toxicity of a given



**Table 1** Microbial bioassay methods

Method	Technique
Growth rate, biomass, numbers	Turbidometric, potentiometric, spectrophotometric, biochemical testing, electronic cell counters (e.g., flow cytometer), viable cell counting (MPN), direct observation
Respiration	Respirometric (manometric, electrolytic, DO)
Nitrification inhibition	NH <sub>4</sub> <sup>+</sup> utilization, NO <sub>3</sub> <sup>+</sup> production, Minntox, Amtox
Enzyme assays	Esterases, dehydrogenases, phosphatases
<i>Vibrio fischeri</i> bioluminescence	Biotox, Lumistox, Microtox
Chemiluminescence	Horseradish peroxidase, Eclox, Aquanox
Genetically modified bacteria	Reporter genes, e.g., <i>lux</i> , <i>luc</i> , <i>gfp</i> marked microorganisms
Biosensors for cytotoxicity, genotoxicity	<i>Escherichia coli</i> , <i>Pseudomonas putida</i> , <i>Salmonella typhimurium</i>

sample. A major advantage of a bioassay over chemical analysis is that the former monitors bioavailability.

Toxicity testing involves recording the response of biological material (whole organisms, enzymes, etc.) to a toxicant over a range of concentrations. The concentration, which affects a predetermined percentage of the biological material over a certain time period for a given response, is then calculated. Acute and chronic toxicity assays exist for organism testing. The endpoint of acute toxicity tests has traditionally been death of an organism, which provides information in terms of a lethal concentration (LC). The concentration at which 50% of the test organisms are killed (LC<sub>50</sub>) is the standard assessment in acute toxicity testing. Chronic toxicity testing involves monitoring sublethal effects such as growth, reproduction, or any other activity that will affect the long-term survival of the organism. Results are given in terms of an effective concentration (EC) or inhibitory concentration (IC), depending on the parameter tested.

The use of the term EC and IC should be clarified. Environment Canada advocated that 'inhibitory concentration should be used for any toxicological test that measures a change in the rate in a continuous response. The term effective concentration should be limited to other toxicity tests in which quantal measurements are taken'. Quantal in this context means that at a given exposure concentration a certain percentage of the test organism will show an effect whilst the remaining percentage will not show the effect.

It is impossible to test the toxicity of a chemical to every species and normally the species chosen for testing should be the most sensitive to the chemical in question or environmentally the most relevant. There has recently been a move toward the use of a battery of tests consisting of species from each trophic level in a community, instead of just a single surrogate species, for environmental risk assessment. Traditionally, fish have been the bioassay of choice;

however, the *Daphnia magna* test has also become a standard.

An ideal bioassay should be reliable and reproducible; economic; able to yield statistically robust data; relevant; practicable; readily understood by the layman; able to utilize test organisms continually from reliable stock; simple to emulate; regularly intercalibrated; have a clearly defined endpoint, and sensitive to a wide range of pollutants.

In addition to these qualities the species physiology and ecology of the organisms used should also be well documented.

The problem with most of the traditional tests is that they are long, expensive, and require time-consuming propagation of test organisms from higher trophic levels. The logistics of screening hundreds of chemicals and effluents is therefore not feasible, and there are ethical concerns associated with many traditional tests. Microorganisms and microbial constituents can be cost-effective alternatives to higher organism testing and a number of rapid methods utilizing these have been proposed (Table 1). Most of these tests could be used when microbial parameters are under investigation, and enable screening of large numbers of samples as a preliminary indication of toxicity.

## Bioassays

### Bioassays Using Changes in Growth Rate, Biomass, Numbers

Bioassays utilizing changes in growth rate have been employed to test the effect of chemicals on microorganisms. Such tests tend to be lengthy and labor intensive as they involve the use of techniques monitoring changes in microbial populations and biomass indicators that are not easy to automate. Microorganisms used in tests are either present as single, axenic, cultures or as mixed populations, and are subjected to varying concentrations of toxicants. Techniques used to monitor growth rates or biomass,

including changes in cell numbers, are biochemical analyses, turbidometry, spectrophotometry, potentiometry, electronic counting using, e.g., flow cytometry, viable and total cell counting using cultivation (plate counting or the most probable number (MPN)) or microscopic techniques. With microscopic methodology, dyes are used which can stain all cells, dead or alive, or which specifically target dead or living cells. Staining techniques utilize visible and ultraviolet light for detection. For the latter, fluorescent dyes such as DAPI (4',6-diamidino-2-phenylindole), CTC (5-cyano-2,3-ditolyl tetrazolium chloride), INT (2-(*p*-iodophenyl)-3-(*p*-nitrophenyl)-5-phenyltetrazolium chloride), FITC (fluorescein isothiocyanate), SYTOX Green, LIVE/DEAD. BacLight Bacterial Viability Kit (Molecular Probes) and others can be used. In addition, changes in specific microorganisms can be followed using fluorescent antibodies. Data on temporal changes in biomass are used to calculate, e.g., LC<sub>50</sub> concentrations for specific toxicants.

Toxicity bioassays are available which rely on measuring ATP changes. ATP, a high-energy compound present in all living organisms, is rapidly destroyed when cells die. An ATP luminescence bioassay determines the changes in ATP content of a given inoculum in the presence of toxic substances. Such a method of ATP detection is based on the light producing properties of the luciferase enzyme derived from the firefly. Light is emitted during the reaction of the luciferase enzyme, luciferin (substrate), magnesium ions, and ATP.

The overall reaction may be summarized as



The total light output is directly proportional to the amount of ATP present in the sample.

There are a number of techniques available to extract ATP from cells. Some use trichloroacetic acid (TCA), which also aids in stabilizing the molecule. TCA has a good extraction capacity, is relatively quick, has a reduced potential for human error, and does not discriminate between the different microbiological fractions of activated sludge. Luminescence produced in the luciferin–luciferase reaction can be measured using scintillation counting equipment or a range of luminometers currently on the market.

### Respirometry

Respirometry involves the measurement and interpretation of the biological oxygen consumption. Oxygen consumption is directly associated with both microbial growth and substrate removal.

A heterogeneous microbial community present in a sample allows the system to be flexible, even when challenged with considerable fluctuations of toxicant. Respiration rate is usually measured with respirometers.

All respirometers are based on techniques that measure the rate at which cells take up oxygen. This can be done directly by measuring the rate at which biomass takes up dissolved oxygen (DO) from the liquid or indirectly by measuring gaseous oxygen. Units measuring gaseous oxygen are more sophisticated as they do not need to estimate the oxygen consumption of the samples before the start of an experiment. They include reoxygenation of the culture as oxygen uptake proceeds. Where reoxygenation is achieved by electrolysis, the apparatus is known as an electrolytic respirometer.

Electrolytic respirometers rely on the electrolysis of water to provide the oxygen necessary for the growth of microorganisms in the sample. The inoculated medium is stirred in each respirometric unit, which consists of a closed vessel and an electrolytic cell for oxygen supply. As oxygen is consumed by the biological activity in the reactor vessel a slight vacuum lowers the electrolyte level in the electrolysis cell triggering oxygen production. Oxygen is produced at the positive electrode and added to the reaction vessel until the original pressure is restored. An electrical control unit monitors the amount of oxygen required to equalize the pressure, thus a continuous cumulative oxygen demand readout is obtained and expressed as a concentration. Measurements can be taken for extended periods of time without interruption or reduction of the oxygen content in the reaction vessel.

In ISO 8192, 'Test for inhibition of oxygen consumption by activated sludge', the oxygen uptake of a well-aerated sample of actively respiring sludge is measured using a DO electrode. The decrease in DO concentration is recorded as a function of time. An IC of test substance decreases the oxygen uptake by the microorganisms, and the IC<sub>50</sub> is calculated. The percentage inhibition of the microbial respiration is calculated by comparison with a control mixture without test substances.

Respirometric techniques have been extensively used for the determination of BOD and biokinetic parameters. A recent development is the use of this technique for the respiration inhibition kinetics analysis to quantify the toxic (or inhibitory) effect of xenobiotic compounds on the biogenic-carbon removal in biological wastewater treatment systems.

A number of respirometers are available for bioassay purposes. A common type is the Arthur Respirometer used in toxicity testing and also for

the assessment of treatability of specific sources of wastewater entering a wastewater treatment plant.

### **Nitrification Inhibition**

Nitrification, the conversion of nitrogen from a reduced to a more oxidized state, is carried out by a diverse group of microorganisms including chemoheterotrophic and chemolithotrophic nitrifying bacteria. Nitrogen removal is an important aspect of modern wastewater treatment processes as it prevents eutrophication of inland and coastal waters. Treatment plants employing biological nitrogen removal encounter problems arising from the presence of toxic compounds in the influent to the treatment works. It is therefore important to develop methods for investigating the inhibitory effects of such waters. Methods using pure cultures of nitrifying bacteria or samples of nitrifying activated sludge are adequate for determining the inhibition of nitrification.

There are two principle methods of determining the inhibition of nitrifying bacteria. One of them is based on the changes in the oxygen uptake rate (OUR) of the nitrifying bacteria when an inhibitor has been added. The OUR of the activated sludge is determined initially and then, when a specific inhibitor of the nitrifying bacteria has been added (i.e. allylthiourea (ATU)), the OUR is again determined. The change in activity of the nitrifying bacteria is calculated from the difference in these two rates. As the proportion of nitrifying microorganisms in the activated sludge is quite low, the greatest amount of oxygen consumption is due to the heterotrophic bacteria and a small variation in oxygen measurements may affect oxygen consumption by nitrifiers in nitrogen oxidation. Also, these methods require addition of selective inhibitors and are time consuming. The other type of test is based on the direct measurement of the nitrifying activity by means of the ammonia consumption rate or the total oxidized nitrogen production rate. The measurement of nitrifying activity from the ammonia consumption rate is not as accurate a reflection of activity as the production of total oxidized nitrogen because ammonia can be used in other processes apart from nitrification. Ammonia can be used by heterotrophs as a nitrogen source to yield new biomass and also it can be turned into  $\text{NH}_3$  gas that can be lost from the medium. The measurement of the nitrifying activity by means of the production of total oxidized nitrogen is both a direct method and is a better estimation of the nitrifying activity.

A great deal of work has been published on the OUR of nitrifying bacteria and toxicity methods

have been developed using a biological electrolytic respirometer. As only the total amount of oxygen consumed is measured, ATU is used to distinguish nitrification. Nitrification is evident when the sample without ATU uses more oxygen and at a faster rate than the sample with ATU. Nitrification toxicity tests are performed by inoculating wastewater samples with active nitrifying seed sludge and comparing resulting activities with activities measured for a reference wastewater that is known to support nitrification. The long response time and the laborious procedure make this method unsuitable for toxicity testing.

A proposed mini-nitrification test for toxicity screening is called MINNTOX. The test is carried out in 15 ml capped test tubes placed on a rotating shaker making it easier to screen large numbers of samples. The test is performed using 6 ml of liquid in a 15 ml capped test tube and when the tube is rotated, the liquid is aerated efficiently due to the 9 ml headspace. The gas phase contains a sufficient amount of oxygen ( $\sim 2.5 \text{ mg O}_2$ ) to satisfy the demand for nitrification and sludge respiration during the 2-hour test. A limitation of the test is encountered under extreme conditions with very high chemical oxygen demand concentrations. This method is suitable for volatile compounds; however, the compounds will be distributed between the liquid and gas phases, and therefore the bacteria are exposed to a lower concentration than the nominal concentration added. During the test, equilibrium is quickly attained, and the actual liquid concentration can be easily calculated.

Various modifications of the MINNTOX method have been made to ensure, for example, a nonlimiting oxygen concentration during the performance of the test. The test can be carried out in 30 ml capped test tubes and 10 ml of liquid are aerated during the 2-h test. The method is suitable for the screening of inhibitory wastewaters with high organic loads because it provides a modification with pure oxygen added to the gas phase to ensure nonlimiting oxygen concentrations. In Sweden, this method is now being used for regulatory purposes, including a toxicity limit based on this test.

The method using the production of total oxidized nitrogen to measure inhibition is easy to carry out, makes it possible to screen a large number of samples without any specific equipment, and directly measures nitrification inhibition.

### **Enzyme Bioassays**

Enzymes are known to play a crucial role in the metabolism of all organisms. In living cells they catalyze

**Table 2** Microbial enzyme activities – pathways, enzymes, application, and photometrical and fluorogenic substrates

Metabolic pathway	Enzymes	Photometric substrates	Fluorogenic substrates
Nonspecific hydrolases	Esterases Phosphatases	Fluorescein diacetate Nitrophenylphosphate	Fluorescein diacetate MUF-phosphate
Carbohydrate metabolism	Amylases Cellulases Glucosidases	Amylopectin azure Carboxymethyl-cellulose Nitrophenyl- $\alpha/\beta$ -glucosides	MUF-cellobiopyranoside MUF- $\alpha/\beta$ -glucoside
Protein metabolism	Proteases Alanine-peptidases  Leucine-peptidases	Hide powder azure L-Anilinenitroanilide- hydrochloride Leucine- <i>p</i> -nitroanilide	Casein resorufin L-Alanine-4-Methoxy- $\beta$ -naphthylamide L-Leucine-4-methyl-7- coumarinyl- amidehydrochloride
Electron transport system	Dehydrogenases	Iodonitrotetrazoliumchloride	

the chemical reactions within anabolism, catabolism, or energy transfer. The high specificity of a single enzyme to transform a single substrate to the product is fundamental for analyzing specific substances. Enzyme tests are usually carried out in food analysis and clinical chemistry. The activity of a specific enzyme (substrate turnover per hour) can be used to quantify a substance or to characterize the physiological conditions of organs.

Two approaches to measuring enzymatic activity exist. *In vitro* tests use commercially available isolated enzymes that have been purified and characterized. In medical laboratories *in vitro* enzyme tests are used as diagnostic aids to determine blood constituents. In food chemistry, carbohydrates, organic acids, and alcohols are quantified by using isolated enzymes. The second method is to determine the activity of enzymes *in vivo*. A synthetic substrate labeled with a chromophore is added to the sample. The enzyme in the sample catalyses the cleavage of the substrate chromophore bond and the chromophore can be detected photometrically or fluorometrically. The amount of dye released per hour corresponds to the enzyme activity as substrate turnover per hour.

Organic and inorganic materials of natural or anthropogenic origin are mainly degraded by microorganisms. Degradation of these substances provides a continued source of nutrients and energy. Microorganisms catabolize organic and inorganic extracellular materials using exo-enzymes. These are mostly bound to the cell membrane but some are also released into the surrounding soil or water. Both organic and inorganic materials may be transported through the cell membrane into the lumen of the cell where it is degraded by endoenzymes. The

determination of enzymatic activities *in vivo*, therefore, is an indicator of microbial degradation in soil, sediment, and water. Table 2 shows a list of existing tests to determine enzymatic activity *in vivo*. The methods to determine enzymatic processes in aquatic habitats are well established and one method used to determine alanine aminopeptidase fluorimetrically is now a DIN (Deutsches Institut für Normung) standard. The use of *in vivo* enzymatic tests for toxicity testing has been used and compared to other microbiological testing techniques.

### Bioluminescence/Chemiluminescence

Chemiluminescence is the emission of photons (electromagnetic radiation as light) when chemically excited molecules decay to ground state following a chemical reaction. In bioluminescence, light emission involves reactions in living organisms. The marine luminescent bacterium *Vibrio fischeri* produces light in a luciferin–luciferase system, linked to the energy transfer in the cell and to respiration. This light can be measured in a suitable luminometer. *Vibrio fischeri* uses the fluorescent pigment luciferin as a terminal electron acceptor. It is therefore an index of cellular metabolism. Toxicity is defined as causing a disturbance to the normal metabolism of the bacterium. The degree of toxicity is proportional to the measured light loss.

In the *V. fischeri* bioassay, the reduction in light output from a suspension of the bacteria on exposure to a toxic sample is monitored. A control is used to detect natural light decrease over time. In the acute test, measurements can be made at intervals of usually 5, 15, or 30 min. The result is often given in terms of an EC<sub>50</sub> (EC that results in a 50% reduction



in bacterial light output). Quality controlled *V. fischeri* bacteria are freeze-dried to maintain their physiological state during storage and are reconstituted on use. Various manufacturers produce versions of this test and testing bacteria (including AZUR; BioOrbit; Dr. Lange; Merck). A chronic test has also been devised. The *V. fischeri* organisms have also been immobilized for toxicity testing.

Chemiluminescent-based toxicity bioassays rely on the reaction of luminol with an oxidant to form light that can be monitored using a sensitive luminometer. Horseradish peroxidase, an enzyme that catalyses the reaction between the oxidant and an enhancer, can be used to generate an enhancer radical reacting with luminol to produce a luminol radical emitting light. In this bioassay, chemicals that scavenge free radicals (e.g., antioxidants) and enzyme inhibitors can reduce the light output of the reaction and this is usually dependent on the concentration of the toxicant. A number of proprietary chemiluminescence tests are available on the market (e.g., Eclox, Aquanox).

### **Reporter Genes/Whole Microbial Cell Biosensors**

The use of DNA in diagnostics has increased in recent years and is likely to become more important following the now completed Human Genome Project and the rapid elucidation of more and more microbial genomes. Reporter gene technology and biosensors can make use of DNA in the development of toxicity bioassays that are rapid, sensitive, and cost-effective.

Microorganisms can be genetically modified to harbor genes that generate a 'reporter' signal in response to a specific recognition event. This can be achieved by fusing together two genetic elements. The first element is the promoter that is activated by an analyte or toxic substance, and, this in turn activates the second element, a reporter gene. Reporter genes code for RNA and proteins that are different from the normal intracellular and extracellular molecules of a particular microorganism. They can be used in bioanalytical methods to produce signals indicating the presence of a target analyte. Reporter genes are often used to detect transcriptional activity within cells. Reporter proteins include alkaline phosphatase, beta( $\beta$ )-galactosidase, bacterial luciferase, and various fluorescent proteins such as green fluorescent protein. In addition to not being induced endogenously, the reporter gene should respond to stimulus in a rapid and quantifiable manner. Constitutive promoters are also used in stress responses where toxicity is monitored by a reduction in microbial metabolism affected by chemicals. For example, recombinant bioluminescent bacteria exhibit

light reduction in the presence of toxic or lethal chemicals concentrations.

Bioluminescence genes (*lux*, *luc*, *Aequorin*) are most commonly used in biosensors for monitoring toxicity and measuring pollutants. In response to analytes, including pollutants, microbial pathways involved in metabolism or stress may be activated. This results in the expression of one or more genes. If the expressed gene has a bioluminescent reporter gene fused within its promoter sequence, transcription of the reporter mRNA and protein synthesis will take place, indicating activation of the pathway. In this case there is rapid detection of activation manifesting in light emission by microbial cells measured using a luminometer. The use of reporters, which do not give an optical signal, would require additional assays to detect mRNA, the reporter protein or its enzymatic activity.

Bioassays utilizing reporter gene technology are available for a wide range of substances and genetically modified microbial cells can be used to detect organic and inorganic chemicals in natural samples. Numerous organic constituents in samples can be detected including naphthalene, toluene, phenolics, and inorganic elements such as Hg, As, Cu, and Pb. Assays can be carried out using microorganisms in suspension or immobilized onto surfaces or in specific matrices. Suspension bioassays have used microtiter plates and microplate luminometers such as Anthos Lucy 1, Dynatech, Berthold, where an optical signal is produced.

Generically, a biosensor is a device that incorporates a biological sensing element in close contact with a transducer that can convert a biological change in the sensing element to a measurable physical response. Microbiological sensing components include whole cells, nucleic acids, enzymes or enzyme components, ion channels, receptor molecules, or antibodies. The transduction technique may be potentiometric, amperometric, conductimetric, impedimetric, optical, calorimetric, or acoustic and depends on the signal emitted by the biosensor.

An optical toxicity biosensor device can utilize light-emitting bacteria immobilized directly onto a surface or incorporated in a polymer that is directly deposited onto glass or other surfaces. Toxicants added onto the immobilized bacteria cause an increase or a decrease in light output that is measured by a sensitive photo diode situated below the immobilized bacterial film.

Environmental chemicals, both natural and synthetic, play an important role in the cause of human cancer because they can react with DNA (genotoxic). This led to the development of a reverse mutation test, the Ames test, using auxotrophic bacterial

mutants, such as strains of *Salmonella typhimurium* requiring histidine (*his*) for growth. The test involving *his* mutants of *S. typhimurium* relies on point mutations of one or more DNA base pairs reversing the initial mutation in the test organism enabling it to become prototrophic for the essential amino acid (non-histidine-requiring). Point mutations are often the cause of many genetic diseases in humans. The Ames test is a relatively complex procedure requiring numerous manipulations and more than one day to perform. Genotoxicity can be detected using reporter gene technology where the reporter gene is associated with genes involved in DNA repair (SOS response involving > 30 genes such as *Rec*, *Lex*, *uvrA*, etc.). Thus, in the case of bioluminescence genes, DNA damage will activate the stress promoter which will lead to light emission.

See also: **Bioluminescence.** **Chemiluminescence:** Overview. **Enzymes:** Enzyme-Based Assays. **Microbiological Techniques.** **Water Analysis:** Biochemical Oxygen Demand.

## Further Reading

- Dalzell DJB, Alte S, Aspichueta E, *et al.* (2002) A comparison of five rapid direct toxicity assessment methods to determine toxicity of pollutants to activated sludge. *Chemosphere* 47: 535–545.
- D'Souza SF (2001) Review: microbial biosensors. *Biosensors and Bioelectronics* 16: 337–353.
- Hastings JW, Potrikus CJ, Gupta SC, Kurfurst M, and Makemson JC (1985) Biochemistry and physiology of bioluminescent bacteria. In: Rose AH and Tempest DW (eds.) *Advances in Microbial Physiology*, vol. 26, pp. 235–291. London: Academic Press.
- ISO 8192 (1986) *Water Quality Test for Inhibition of O<sub>2</sub> Consumption by Activated Sludge*, pp. 1–9. Geneva: International Standards Organization.
- King EF and Dutka BJ (1996) Respirometric techniques. In: Dutka BJ and Bitton G (eds.) *Toxicity Testing Using Microorganisms*. Florida: CRC Press.
- Landis WG and Ming-Ho Y (1995) *Introduction to Environmental Toxicology*. Boca Raton: Lewis Publishers.
- Wells P, Lee K, and Blaise C (eds.) (1998) *Microscale Testing in Aquatic Toxicology: Advances, Techniques and Practice*. Boca Raton: CRC Press.

## Bioautography

**L Botz, B Kocsis, and S Nagy**, University of Pécs, Pécs, Hungary

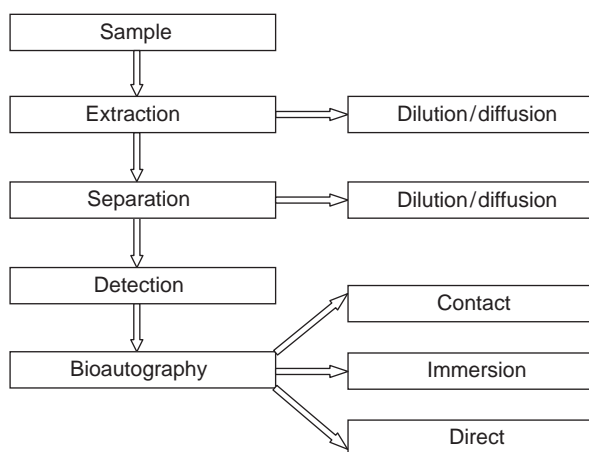
© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Bioautography is a postchromatographic detection method that is widely used for the bioassay of antimicrobial effects. This technique is based on paper or thin-layer chromatography (TLC) separations. Bioautography detects the growth-inhibiting or growth-promoting biological effects of the applied substances. Various bioautographic assays may be used to detect antibacterial, antifungal, antiprotozoal, and cytotoxic substances. The bioactivity-based bioautographic analysis is the most effective method to detect antimicrobial compounds because it detects activity even in complex sample material. Paper chromatography or TLC combined with other analytical techniques is the best separation techniques for bioautography. The combined chromatographic and microbiological assays must be carried out under controlled conditions as the experimental circumstances significantly influence the results. The combination of microbial bioassay with extraction and separation methods is summarized in Figure 1.

The potential applications of bioautography are the following:

1. Additional detection methods in the screening of samples.
2. Targeted or bioactivity-driven isolation processes.
3. Detection of antibiotic residues in human or animal food.



**Figure 1** Flowchart illustrating the combination of the microbial bioassay with the separation techniques.

4. Tool to investigate the activity of the products of pharmaceutical preparations, combinatorial chemistry, and microbial extracts.
5. Part of combined methods for standardization of extracts from biologically active materials (e.g., medicinal plant extracts).
6. Special and selective substance-specific postchromatographic detection processes.

## Theory of Detection Assay

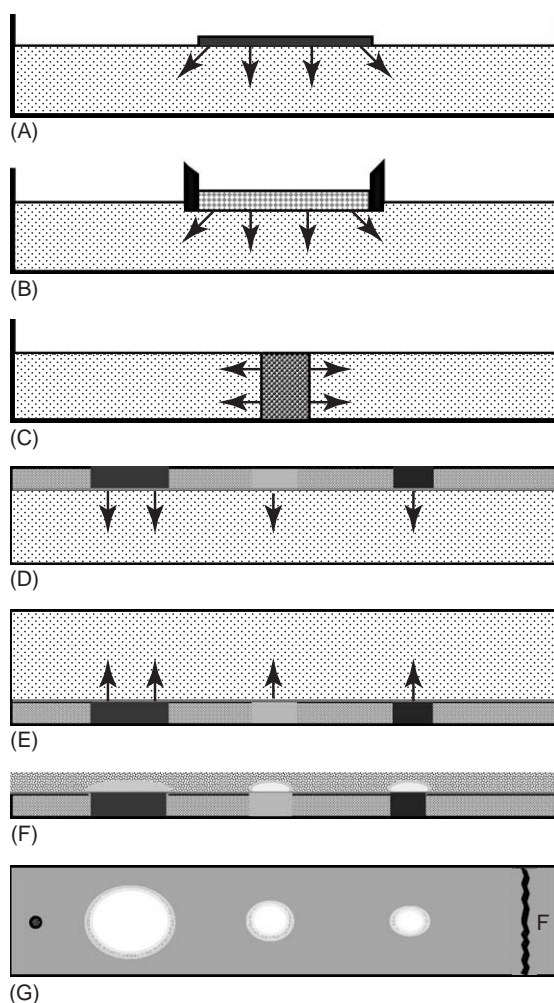
The methods for detection of antimicrobial activity may be classified into three groups: (1) dilution, (2) diffusion, and (3) bioautography.

The dilution method mixes the samples with a suitable medium inoculated with the test organism. After incubation the growth of the microorganism can be determined by direct visual or turbidimetric comparison between the test culture and the control culture. The diffusion method is an agar-overlay method that uses disk, cylinder, or hole as a reservoir for the microorganism to be tested. In the disk method, a filter-paper disk treated with the sample substance is placed on the surface of the agar block inoculated with the microorganism. In the hole-plate method, the samples are placed in a vertical hole in the solidified agar layer. The sample diffuses into the agar where it inhibits or stimulates the growth of the microorganisms around the hole (Figure 2). The cylinder method is similar to the hole-plate method. The sterile cylinders containing the sample are usually placed on the agar surface. After incubation the cylinders are removed and the average diameter of each zone of growth is measured.

There is no relationship between the diffusive power and the antimicrobial activity. The diffusion method cannot be employed when the sample contains lipophilic compounds or with complex samples prepared without separation.

The well-known AMES test is a similar technique in principle. This widely used, simple, and rapid bacterial mutagenicity test detects carcinogenic effects of chemicals. It is based on two assumptions. First, cancer is commonly caused by damage to the cell DNA resulting in somatic mutations. Second, chemicals that damage bacterial DNA and induce mutations are also likely to induce mutations in mammalian cells.

The bacteria-reversed mutation test is another method to investigate the mutagenic and teratogenic effect of chemicals. This uses numerous amino acid-dependent strains of *Salmonella enterica* (formerly *S. typhimurium*). The histidine-dependent strain of *S. enterica* has artificially induced point mutations. The test is carried out in a histidine-deficient bacterial medium. The histidine-independent bacterial



**Figure 2** Types of the migration of separated substances from the chromatoplates to the culture medium. (A) Filter paper disk, (B) cylinder, (C) hole plate, (D) contact or agar-diffusion, (E) immersion or agar-overlay, (F) direct, and (G) upper view of the developed and detected sorbent layer. The white spots show the inhibition zones (F: eluent front).

colonies may arise from spontaneous reversions. When a chemical, or a potential mutagen, is added to the plate, the number of revertant colonies per plate increases in proportion with the applied dosage. The mutagenicity of a substance is determined by comparing the control with the treated culture.

In bioautography, the compounds are separated by chromatography and extracted from the sorbent, then diffused into the inoculated agar. There are three variations of bioautography: (1) contact bioautography, where the antimicrobial compounds are transferred from TLC plate to the inoculated agar plate by direct contact, (2) immersion or agar-overlay method, where the developed chromatoplate is covered with the inoculated agar medium, and (3) direct

bioautography, where the microorganisms can be grown directly on the chromatoplate.

The dilution and diffusion methods mentioned above are all well-known classical microbial assays. Only the bioautographic methods are suitable for direct combination of the detection techniques with any kind of chromatographic separation methods. This technique opens up new avenues to detect the effects of samples on microbes after separation of the compounds by opened sorbent-bed (e.g., TLC) chromatography.

First, it is essential to ensure the appropriate transport (diffusion) of the compounds from the sample layer to the appropriate medium where the microbes can grow, or to create the suitable conditions for the microbial growth directly on the sorbent material. Second, the selection of microbes suitable for detecting the inhibitory effects of the sample substance is also a crucial step.

## Microbe Selection

The selection of the appropriate test organism depends on the purpose of the assay. In a general-purpose test the selected microorganisms must be as diverse as possible. They must represent all important groups of bacteria with group-characteristic physical and chemical compositions and resistance patterns. It is important to take into account that many crude extracts from medicinal plants have specific inhibitory activity against *Staphylococcus aureus*. This is the result of the synergistic effects of various plant constituents. The conditions required for bacterial growth also affect the bacteria selected. It is general practice that pathogenic and apathogenic aerobic microbes, which can be easily cultivated and sensitive to antimicrobial substances, are chosen. These are the Gram-positive *Bacillus subtilis*, *Bacillus cereus*, *Corynebacterium xerosis*, *Micrococcus luteus*, *Staphylococcus aureus*, *Sarcina lutea*, *Streptococcus aureus*, *Clostridium perfringens*, and the Gram-negative *Escherichia coli*, *Erwinia carotovora*, *Erwinia atroseptica*, *Erwinia herbicola*, *Pseudomonas aeruginosa*, *Pseudomonas syringae* bacteria. From amongst fungi *Cladosporium cucumerinum*, *Candida albicans*, *Fusarium* species, *Penicillium expansum*, *Saccharomyces cerevisiae* are usually chosen.

Sterilization is expensive, time consuming, and requires suitable microbial laboratory conditions. Sometimes it is sufficient to protect samples from contaminating bacteria by using ethanol as an extraction solvent. The TLC plates should be sterile, but generally this is not essential, because the large population of test bacteria will overgrow the few contaminant species. Irradiation of the chromatographic

plate with ultraviolet light is an acceptable method to reduce contamination.

## Cultivation of Test Bacteria

Most bacteria and yeasts can be cultivated on standard Müller–Hinton broth. This broth is a general-purpose liquid medium for the cultivation of large amount of bacteria and may be used to test the antimicrobial susceptibility. Standard test microbes should be obtained from the American Type Culture Collection (ATCC) or from other recognized culture collections, otherwise results cannot be comparable. Sometimes it is necessary to use clinical isolates, when multiresistant microorganisms are selected for the screening studies. Anaerobe bacteria, because of the exposure to oxygen, and slowly growing microbes (e.g., *Mycobacterium tuberculosis*) are difficult or impossible to test by bioautography.

The standard procedure describes the growth conditions to be applied for the strain, age, and growth of the tested organism. When microbes are used the inoculum must be taken from cultures in the exponential logarithmic growth phase. This is the time when multiplication occurs and the culture is the most homogeneous. To evaluate the assay the inhibition effect has to be visualized, and the live, dead, or impeded microbes have to be distinguished.

## Visualization of Inhibition of Separated Compounds

The zones or spots of separated compounds are generally visualized by detecting dehydrogenase activity with tetrazolium salt-based reagents. The metabolically active bacteria convert the tetrazolium salt into red formazan dye (2,3,5-triphenyl-2H-tetrazolium chloride, tetrazolium red). A number of tetrazolium salts were evaluated as potential substrates for the enzymatic reaction, but *p*-iodonitrotetrazolium violet, tetranitro blue tetrazolium, and MTT were found to be suitable substrates. The equations below show the principle of the reaction with tetrazolium salt, 3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyltetrazolium bromide (MTT, tetrazolium dye).



NADH/NAD<sup>+</sup> = nicotinamide adenine dinucleotide, pyruvate = pyruvic acid, formazan = tetrazolium dye, LDH = lactate dehydrogenase.



The MTT is reduced to formazan by the NADH, which is generated in the NAD-dependent lactate dehydrogenase reaction. The presence of detergents such as Triton X-100 surfactant (alkylaryl polyether alcohol) accelerates the reduction of the tetrazolium salt manifold. After an additional incubation inhibition zones become visible. The antibacterial compounds appear as clear spots on the colored formazan background.

The obligate aerobic bacterium like *Xanthomonas pruni* does not reduce tetrazolium dyes. The visualization can still be carried out by modification of the assay. The developed plate is pressed onto the agar plate that contains phytopathogenic microbes. These microbes hydrolyze the incorporated gelatin and starch in the nutrient agar, and produces acid from several sugars. The result can be detected by incorporation of bromocresol purple into the media. After transferring the chromatograms onto the agar medium the visualization of the separated cytotoxic antibiotics can be carried out with 2,6-dichlorophenolindophenol (Tillman's reagent, phenolindo-2,6-dichlorophenol). The viable test cells reduce the dye while the dead cells do not.

## **Contact or Agar Diffusion Bioautography**

The typical bioautography method is based on the migration of the antibacterial compound from the chromatographic plate to the inoculated agar plate by diffusion. The chromatographic plates are placed on the surface of agar plates inoculated with microorganisms. After 15–30 min the chromatographic plates are removed and the substances that already diffused into the agar inhibit the growth of the test microbes. Inhibition zones can be easily visualized by the use of suitable microbe strains and mainly dehydrogenase-activity-detecting reagents. The advantage of this technique is that before the antimicrobial test the heterogeneous samples are separated by TLC. The antibacterially active spots can be located. Agar-diffusion assays are particularly applicable to polar and moderately polar compounds. It could be further improved for antitumor antibiotics by applying an appropriate filter paper placed between the inoculated agar and the plate.

## **Immersion or Agar-Overlay Bioautography**

In agar-overlay or immersion bioautography, the developed plate is covered with agar medium. After solidification the plates are inoculated with the

microorganism and growth or inhibition bands can be visualized. Active compounds are transferred from the stationary phase to the agar layer by diffusion. Agar medium is essential in bioautography and provides optimum conditions for bacterial growth. The use of agar gel as a support medium has, however, considerable disadvantages including slow diffusion and bad contrast. In order to achieve longer incubation time, potassium nitrate is incorporated into the basal and seeded agar layers to overcome the problem of limited oxygen availability. After incubation, the plate is sprayed with a tetrazolium salt (e.g., MTT), which is converted to a formazan dye by the microorganism. This technique is a hybrid of the contact and direct bioautographies.

Generally, a less stringent mobile phase (appropriately altered solvent composition of the eluent) is recommended during chromatography in order to achieve increased sensitivity and selectivity of the microorganism in the agar plate or nutrient broth.

It must be noted that there are no sharp division lines between the different techniques. By the use of bioautographic methods great variety of assays can be carried out.

Preliminary screening of anthracycline antitumor antibiotics and the antifungal activity of polyphenols have been tested by the use of bioassay-driven fractionation. Sometimes the reduction of the thickness of the agar layer is beneficial. This is often the case when antifungal compounds are probed and activity-driven fractionated with indicator microorganism *C. albicans*. The essence of this modification is that the inoculated medium is poured rapidly onto the developed chromatogram. The agar overlay method requires the transfer of the active compounds from the stationary phase into the agar layer by diffusion. The thin agar films facilitate this process.

## **Direct Bioautography**

In direct bioautography, the microorganism cultures grow directly on the TLC plate. Each step of the method is performed on the plate. Originally, this method was developed for spore-producing fungi like *Aspergillus*, *Penicillium* and for bacteria. The suspension of the microbial liquid culture is poured over the developed TLC plate. Incubation in a humid atmosphere enables growth of bacteria, and dehydrogenases from the living microorganisms convert tetrazolium salt into the corresponding red formazan. The locations of antibacterial activity can be easily visualized as the antibacterial compounds appear as clear spots on the red background. The sensitivity of the assay enables determination of the minimum active concentration or the minimum

amount of compound. The quantitative determination of compounds by bioautographic analysis can be also implemented by regression analysis of the sizes of the zones or bands. In a simple procedure, *Glomerella cingulata* is directly sprayed onto TLC plates and fungitoxic compounds detected. The detection of inhibitors of *Erwinia carotovora* and *E. herbicola* on TLC plates is also possible. In this case suspensions of the bacteria are sprayed onto the TLC plates by the use of a fine spray. Even the phototoxic furocoumarin compounds can be determined by using the fungus *Penicillium expansum* as the test organism, spraying the TLC plate with a suspension of conidia and irradiated with UV light. With the modification of this technique it is possible to detect the pollen germination inhibitors on TLC plates.

A ready-to-use kit (Merck, Germany), called Chrom Biodip Antibiotika, has recently become available. This was developed for special direct bioautographic detection. This kit contains a premixed culture medium, *Bacillus subtilis* spore suspension, nutrient medium, and MTT detection reagent. Generally, these methods require only heater and therefore it is possible to use this detection in a conventional chemical laboratory. The total determination time is ~21 h. The main advantages of this test kit are: (1) the test needs little laboratory time and cost, (2) the user does not need to maintain strains of bacteria for the test, and (3) the spore suspension can be used at any time.

Overpressured-layer chromatography, also called forced-flow planar chromatography, is an instrumental version of planar liquid chromatography. It develops TLC under controlled conditions. The advantages of this technique are the optimized flow of the mobile phase, the short separation time, the high resolution, the good reproducibility, the high sample throughput, the simple precleaning process, and, finally, the reduced spot or band broadening.

In direct bioautography there is usually no need for diffusion from the adsorbent to the agar medium. Both water-soluble and lipid-soluble substances are in direct contact with the bacteria and the growth medium on the surface of the adsorbent particles. There is no dilution and diffusive loss of the separated substances, and this results in low detection limits.

However, the planar chromatographic adsorbent material is, not always suitable for microbial studies. If the test bacterium requires special conditions (e.g., extremely long incubation time) the advantages of the method may not be exploited. The main factors influencing the evaluation of direct bioautography will be discussed below.

The direct bioautography process can be divided into three steps: (1) growing test bacteria cultures for

dipping, (2) separation of substances for assay by TLC, and (3) microbial assay on the TLC plate. A flowchart for the direct bioautography process is shown in Figure 3.

## Factors Influencing the Microbial Assay

The bioautography screening tests are influenced by many chromatographic and microbial factors.

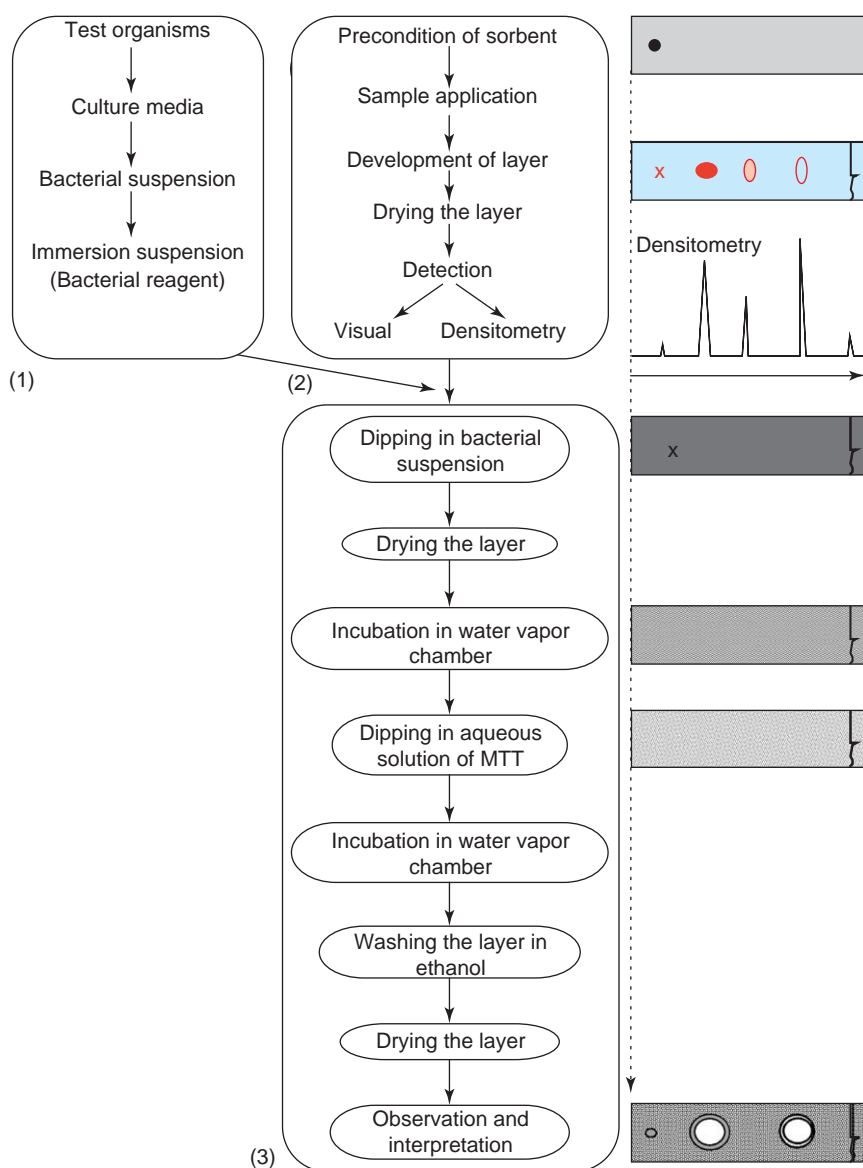
The chromatographic part of the assay is affected by solvents, additives, eluents, sorbents, spot volume, sample application, development mode, spot or band broadening, resolution of the separated compounds, the analytical detection, and the derivatization.

The microbial part of the assay depends greatly on the origin, contamination, pH of the samples, the composition of the culture medium, the chemical character of the analyzed material, the immersion suspension, the incubation temperature, the incubation time, and the visualization of the inhibition zones.

The components of the elution or extraction solutions are essential. Solvents can contain impurities and/or stabilizers, which might inhibit bacterial or fungal growths. Also, solvents or their residues in the adsorbents can alter results (e.g., traces of acetic acid remaining in the adsorbent inhibit bacterial growth). Tetrahydrofuran is not a suitable chromatographic solvent because even in a minute amounts it inhibits bacterial growth. It is beneficial if the mobile phase contains highly volatile solvents, so they can be completely removed before the microbial assay. This is not the case with most of the acids and bases. The good eluents used in direct bioautography does not disturb the microbial detection.

The selection of the appropriate adsorbent is not simple. The adsorbent has to fulfill the requirements of the chromatographic and the microbial processes. Normally, precoated silica gel plates are used as it contains sufficient binding power. The TLC technique mainly uses silica gel. It is also very promising to use a water-resistant glass plate. In special circumstances other adsorbents (e.g., alumina for gentamicin) are more appropriate. The use of polyamide and cellulose layers for direct bioautography is limited. Precleaning of the chromatographic plates with appropriate solvent is vital.

The insoluble compounds in the bands can crystallize after drying. This reduces the sensitivity of the assay considerably. Drying of compounds on silica gel plates can be also detrimental for some fungicides, probably because of their inherent chemical reactivity. Very concentrated sample loads, in terms of



**Figure 3** Implementation of the direct bioautography assay.

material per unit area, are required for the greatest sensitivity after visual detection. In the case of natural substances the sensitivity of direct bioautographic assay was most successful on silica gel layers. The most efficient detection system uses high-performance TLC layers with *Bacillus subtilis*. The detection limit can be determined by measuring of the diameter of the clear zone (inhibition ring). The values of the minimum detectable dose can be determined similarly to conventional TLC.

### Optimization of Direct Bioautography

The key factors in the optimization of the assay are the quality of the bacterial suspension and the conditions of the TLC incubation.

Frequent practice is the use of the overnight cultures, like in the case of *B. subtilis*, or 6-h cultures, like in the case of *E. coli*. The optical density (OD) (absorbance) of the bacteria suspension is varied. It is recommended to be between 0.40 and 0.60, which represents  $\sim 4 \times 10^7$  cells  $\text{ml}^{-1}$ .

It is important to ensure the maximum metabolic activity of the microbe suspension. The ATP (adenosine triphosphate) content reflects the metabolic activity of cells. Thus, the measurement of ATP content is a good indicator of the viability of bacteria. The measured ATP content must be expressed on the basis of one of the cellular attributes (protein content or cell number). The bacterial suspensions are the most suitable for coating the TLC plates when their growth is in the exponential log phase, and their OD

is between 0.2 and 0.5. When the OD is higher than this, the cell's activity, which is mirrored by the ATP content, is already in decline.

The surface of the sorbent material does not provide good living conditions for microbes. The growth medium adhered to the silica particles in TLC acts a source of nutrients for the test bacteria.

The relationship between the ATP and the protein contents is crucial. An optimum ratio between these two parameters is required for the successful microbial test of the plate. The best stage of growth shows high ATP content, without protein degradation. Only the metabolically active bacteria can convert the tetrazolium salt.

As a rule, higher sensitivity of the assay may be obtained by shorter incubation times for chromatoplates. The optimum incubation time (when the culture is in the exponential log phase) for Gram-positive sporeforming test bacteria is between 8 and 12 h while for Gram-negative enterobacterial bacteria is between 3 and 6 h.

In summary, the principles of optimization are the following:

1. The plate must be coated with a bacterial culture at its maximum ATP activity.
2. The immersion suspension must contain a large number of vigorous microbes.
3. The higher OD value does not necessarily guarantee better detection.
4. By mixing bacterial cultures in different stages of growth in order to achieve the optimum OD value for the suspension is not acceptable.

See also: **Bioassays:** Microbial Tests. **Food and Nutritional Analysis:** Overview; Contaminants. **Pharmaceutical Analysis:** Plant Extracts. **Thin-Layer**

**Chromatography:** Principles; Plate Technology; Method Development. **Water Analysis:** Microbiological.

## Further Reading

- Botz L, Nagy S, and Kocsis B (2001) Detection of micro-biologically active compounds. In: Nyiredy Sz (ed.) *Planar Chromatography. A Retrospective View for the Third Millennium*, pp. 489–516. Budapest: Springer Scientific Publisher.
- Colf MD (1994) Key antifungal antibacterial and anti-insect assays – a critical review. *Biochemical Systematics and Ecology* 22: 837–856.
- Hostettmann K, Terreaux C, Marston A, and Potterat O (1997) The role of planar chromatography in the rapid screening and isolation of bioactive compounds from medicinal plants. *Journal of Planar Chromatography* 10: 251–257.
- MacDonald A, Chen G, Duke P, *et al.* (1979) Bioautography. In: Touchstone JC and Sherma J (eds.) *Densitometry in Thin Layer Chromatography*, pp. 201–221. New York: Wiley.
- Mortelmans K and Zeiger E (2000) The Ames *Salmonella* microsome mutagenicity assay. *Mutation Research* 455: 29–60.
- Paxton JD (1991) Assays for antifungal activity. In: Dey PM and Harborne JB (series eds.) *Methods in Plant Biochemistry*. Hostettmann K (ed.) *Assays for Bioactivity*, vol. 6, pp. 33–46. London: Academic Press.
- Rios JL, Recio MC, and Villar A (1988) Screening methods for natural products with antimicrobial activity: A review of the literature. *Journal of Ethnopharmacology* 23: 127–149.
- Vanden Berghe DA and Vlietinck AJ (1991) Screening methods for antibacterial and antiviral agents from higher plants. In: Dey PM Harborne JB (series eds.), *Methods in Plant Biochemistry*. Hostettmann K (ed.) *Assays for Bioactivity*, vol. 6, pp. 47–66. London: Academic Press.

# BIOCHEMICAL OXYGEN DEMAND

See **WATER ANALYSIS: Biochemical Oxygen Demand**

# BIOLUMINESCENCE

**L J Kricka**, University of Pennsylvania, Philadelphia, PA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Bioluminescence ('living light') is the name given to the visible light emission from living organisms. It is

widely distributed in nature (~666 genera from 13 phyla) and some representative examples of bioluminescent organisms are shown in **Table 1**. In the marine environment, it has been estimated that in the dimly lit mid-ocean between 200 and 1200 m ~95% of fish and 86% of shrimps and squid are bioluminescent. In surface water <10%, and in the abyssal

**Table 1** Examples of bioluminescent organisms

Organism	Common name	Wavelength of light emission (range or max) (nm)
<i>Aequorea aequorea</i>	Jelly fish	500–523
<i>Agyroplecus affinis</i>	Hatchet fish	~480
<i>Arachnocampa luminosa</i>	New Zealand glowworm	–
<i>Diplocardia longa</i>	Earthworm	500
<i>Gonyaulax polyedra</i>	Dinoflagellate	479
<i>Lampyrus noctiluca</i>	Glowworm	–
<i>Malacosteus niger</i>	Stomioid fish	469–702
<i>Mnemiopsis leidyi</i>	Sea comb	485
<i>Obelia geniculata</i>	Sea fir	475
<i>Phrixothrix tiemani</i>	Railroad worm	~560 (A) <sup>a</sup> 625 (H) <sup>a</sup>
<i>Pholas dactylus</i>	Piddock	490
<i>Photinus pyralis</i>	Firefly	530–590
<i>Photoblepharon palpebratus</i>	Flashlight fish	490
<i>Pleurotus japonicus</i>	Moon night mushroom	524
<i>Plutonaster</i> spp	Starfish	–
<i>Quantula striata</i>	Land snail	–
<i>Renilla reniformis</i>	Sea pansy	480
<i>Vargula hilgendorffii</i>	Sea firefly	460
<i>Vibrio fischeri</i>	Marine bacterium	489
<i>Watasenia scintillans</i>	Firefly squid	–

<sup>a</sup> A = abdominal organ; H = head organ.

Visible spectrum: blue 400–500 nm, green–yellow 500–575 nm, orange–red 575–700 nm.

depths <25% of organisms are bioluminescent. The phenomenon of bioluminescence has been known since ancient times. Aristotle (384–322 BC) described the luminescence of fungi and dead fish in *De Anima*, and in 1668 Robert Boyle established the requirement for oxygen in bioluminescent reactions. Dubois in the late 1880s performed experiments with hot and cold water extracts of fire beetle and extracts of clam, and showed that light emission resulted when the hot and cold water extracts were mixed together in the presence of oxygen. He named the heat-labile (cold water) extract luciferase, and the heat-stable (hot water) extract luciferin. These generic names have come to be used for the enzymes (luciferase) and substrates (luciferin) in bioluminescent reactions. However, it must be noted that luciferases and luciferins from different organisms differ widely in chemical structure and composition (Table 2); for example, luciferins include aldehydes, imidazopyrazines, benzothiazoles, tetrapyrroles, and flavins (Figure 1). The function of bioluminescence is diverse and includes luring prey (Midshipman fish, *Porichthys*), camouflage (Hatchet fish), schooling

**Table 2** Characteristics and molecular biology of luciferases and photoproteins

Organism	Protein	<i>M<sub>r</sub></i> (subunits, <i>M<sub>r</sub></i> ) (kDa)	Gene
<i>Aequorea victoria</i>	Apoaequorin	~20	<i>aeq</i>
<i>Gonyaulax polyedra</i>	Luciferase	420 <sup>a</sup>	<i>luc</i>
<i>Obelia geniculata</i>	Apoobelin	~20	<i>ApoObi</i>
<i>Photinus pyralis</i>	Luciferase	100 (2, 62) <sup>b</sup>	<i>Luc</i>
<i>Pyrophorus plagiophthalmus</i>	Luciferase	62	<i>Luc</i>
<i>Renilla reniformis</i>	Luciferase	35	<i>Ruc</i>
<i>Vargula hilgendorffii</i>	Luciferase	~60 (6, 10)	<i>Vuc</i>
<i>Vibrio fischeri</i>	Luciferase	77 (2, 40 + 37)	<i>luxA</i> , <i>luxB</i>

<sup>a</sup> 140-kDa monomer, also an active 35 kDa proteolytic fragment.

<sup>b</sup> Sodium dodecyl sulfate electrophoresis gives one band, 62 kDa.

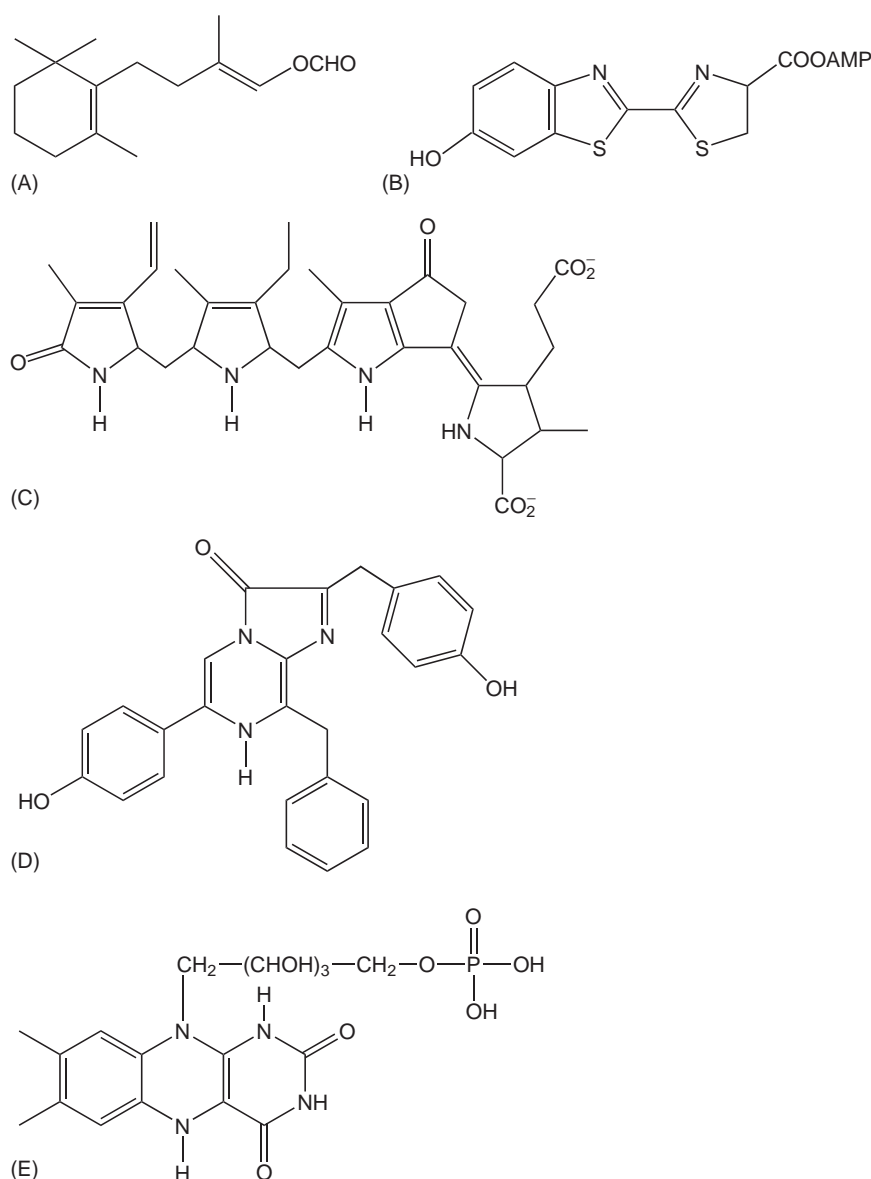
(euphausiid shrimp), vision (Flashlight fish), mating (firefly), and symbiosis (marine bacteria).

## Principles

In a bioluminescent reaction, a luciferase or a photoprotein catalyzes a reaction that produces chemically excited intermediates that decay to the electronic ground state and release energy in the form of visible light (390–750 nm). The efficiency of the overall process is characterized by a quantum yield for bioluminescence ( $\phi_{BL}$ ). This is the product of the fraction of the molecules reacting  $\times$  fraction of molecules entering the bioluminescence pathway that become electronically excited  $\times$  fluorescence quantum yield of the excited state product. Most bioluminescent reactions are relatively inefficient (e.g., for the *Renilla* luciferin–luciferase reaction *in vitro*,  $\phi_{BL} = 0.05$ ); however, the firefly reaction is unique in having a quantum yield for bioluminescence close to unity ( $\phi_{BL} = 0.88$ ). The intensity of light emission is  $> 2 \times 10^8$  photons  $s^{-1} cm^{-2}$  in the Flashlight fish (*Photoblepharon*) and  $> 2 \times 10^9$  photons  $s^{-1} cm^{-2}$  in the dinoflagellate *Gonyaulax*. Studies of bioluminescent spectra reveal that the maximum light emission of most deep-sea species is in the range 450–490 nm (blue), that that of coastal species is in the range 490–520 nm (green), and that terrestrial and freshwater species emit in either the yellow (550–580 nm) or green (510–540 nm) part of the spectrum.

A common feature of bioluminescent reactions is that they are oxidation reactions. Only oxidation reactions can provide the energy needed to produce high-energy states that will emit photons when they decay to the ground state (e.g., the energy required to produce blue light at 450 nm is 265.4 kJ mol<sup>-1</sup>). One





**Figure 1** Structures of luciferins: (A) limpet luciferin (e.g., *Latia*); (B) firefly luciferin (e.g., *Photinus*); (C) dinoflagellate luciferin (e.g., *Gonyaulax*); (D) coelenterate luciferin (e.g., *Renilla*, *Aequorea*); (E) bacterial luciferin (e.g., *Vibrio*, *Photobacterium*).

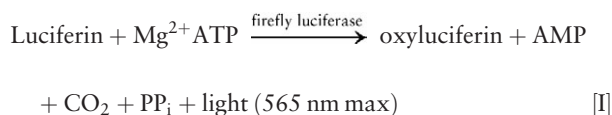
of the most important recent developments has been the elucidation of the molecular biology of bioluminescence and the cloning of many of the genes that encode luciferases and photoproteins (Table 2). These genes, particularly the firefly gene (*luc*), have been used as alternatives to the chloramphenicol acetyl-transferase gene (CAT) as reporter genes for the study of gene regulation and expression. Also, pairs of bioluminescent genes (e.g., firefly and *Renilla* luciferase) have been employed for two-color monitoring of the expression of two different genes. Genes for bioluminescent proteins have been transferred to and expressed in different bacteria, yeast, mammalian cells, transgenic mice, and plants. They

have also been fused with genes encoding other proteins and the fused gene expressed to produce a bioluminescent fusion protein (e.g., apoaequorin + IgG heavy chain, bacterial luciferase + interferon). The fusion proteins have found use as conjugates in immunoassay and have replaced conjugates prepared by conventional covalent coupling techniques.

## Firefly

The North American firefly, *Photinus pyralis* (order Coleoptera, family Lampyridae), has light organs on

the ventral surface of the sixth and seventh abdominal segments. The light emission arises from the firefly luciferase (EC 1.13.12.7) catalyzed ATP-dependent oxidative decarboxylation of firefly luciferin (reaction [I]). The reaction proceeds in two stages; first an AMP–luciferin complex is formed and then this is oxidized by oxygen to form the excited-state oxyluciferin via a dioxetanone intermediate:



A yellow–green light emission (565 nm max) is produced at neutral pH, but at low pH and in the presence of doubly charged ions such as  $\text{Pb}^{2+}$ ,  $\text{Hg}^{2+}$ ,  $\text{Zn}^{2+}$ , and  $\text{Cd}^{2+}$ , the light emission is shifted to longer wavelengths (610–615 nm).

Firefly luciferase is a dimer comprising two identical 62 kDa subunits (one active site per dimer, pH optimum 7.75). It has a binding site for  $\text{MgATP}^{2-}$ , and the  $K_m$  for luciferin is  $1\text{--}10 \mu\text{mol l}^{-1}$  and for  $\text{MgATP}^{2-}$  it is  $0.25 \text{ mmol l}^{-1}$ . The product of the reaction, oxyluciferin, is an inhibitor ( $K_i = 0.23 \mu\text{mol l}^{-1}$ ), as are pyrophosphate, free ATP, and various anions ( $\text{SCN}^- > \text{I}^- > \text{ClO}_4^- > \text{acetate}$ ).

*In vitro* the light emission is in the form of a rapid flash ( $\sim 300$  ms to reach peak). Detergents (e.g., Triton X-100 at concentrations above the critical micelle concentration) and polymers (e.g., poly(ethylene glycol) 6000) modify the kinetics to a protracted glow and this has been interpreted as due to selective removal of the inhibitory oxyluciferin product by micelles. Firefly luciferase can be covalently immobilized to solid supports such as Sepharose and nylon and still retain enzymatic activity. Likewise, it can be covalently coupled to small molecules such as biotin and methotrexate. However, attempts to link the enzyme to large molecules, e.g., immunoglobulin G (IgG), to produce conjugates that retain firefly luciferase activity have been unsuccessful. Firefly luciferase is one of the most hydrophobic proteins known, and this may explain the difficulties encountered in preserving activity during chemical modifications. Enzyme activity is inhibited by anesthetics (e.g., ether, chloroform) and this is attributed to a similarity between the hydrophobic binding site on luciferase and the protein-binding site for general anesthetics.

## Click Beetle

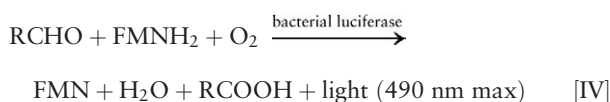
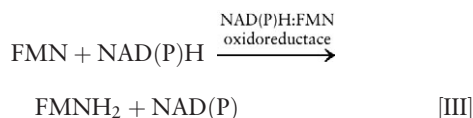
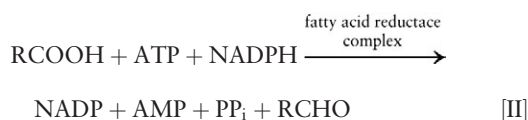
The Jamaican click beetle, *Pyrophorus plagiophthalmus* ('kitty boo'), is capable of emitting different colors of light from the pair of light organs on the surface of the head (green, yellow–green), and the

single organ on the ventral surface of the abdomen (yellow–green to orange). The luciferin in this species is exactly the same as the luciferin from the firefly. The different colors of bioluminescence arise from different luciferases. Four have been characterized (green, 546 nm; yellow–green, 560 nm; yellow, 578 nm; and orange, 593 nm). The genes encoding these luciferases have been cloned and the amino acid sequences of the four luciferases show 94–99% identity, but <50% identity with firefly luciferase.

## Bacteria

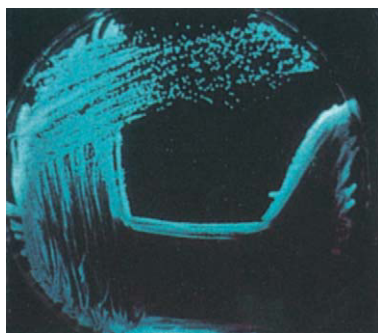
Luminous bacteria (Gram-negative motile rods) are widely distributed and the most abundant light-emitting organism. Luminous bacteria are classified into three genera, *Vibrio*, *Photobacterium*, and *Xenorhabdus*. They are found principally in a marine environment, but freshwater (e.g., *Vibrio albenis*) and terrestrial species (e.g., *Xenorhabdus luminescens*) are also known. In the marine habitat, luminous bacteria are free-living, or as symbionts in fish (e.g., teleost fish + *Photobacterium leiognathi*).

The reactions leading to light emission involve a fatty acid reductase complex (reaction [II]), NADH:FMN oxidoreductase (EC 1.6.99.3) or NADPH:FMN oxidoreductase (EC 1.6.99.1) (reaction [III]), and bacterial luciferase (EC 1.14.14.3) (reaction [IV]). The  $K_m$  values for NADH and NADPH with the oxidoreductases of the appropriate selectivity are 23 and  $22 \mu\text{mol l}^{-1}$ , respectively. For luciferase the  $K_m$  for  $\text{FMNH}_2$  is  $0.3 \mu\text{mol l}^{-1}$ . In the luciferase reaction enzyme-bound  $\text{FMNH}_2$  is oxidized to a 4a-peroxy flavin intermediate and this reacts with tetradecanal to form a stable intermediate that decays slowly with the emission of light. The aldehyde (tetradecanal) is oxidized to tetradecanoic acid:



The genes (*lux* genes) encoding luciferase and the other proteins involved in the bioluminescent reaction have been cloned. Luciferase subunits are coded





**Figure 2** Bioluminescence marine bacteria *Beneckea harveyi*. (Reproduced with permission of LJ Kricka, Hospital of the University of Pennsylvania, Philadelphia, USA.)

by the *luxA* and *luxB* genes, and the fatty acid reductase complex is coded by the *luxCDE* genes.

A feature of the growth of bioluminescent marine bacteria is that light emission lags behind cell growth **Figure 2**. One reason for this lag is the requirement for an autoinducer to accumulate in the medium. The autoinducers in *Vibrio fischeri* and *Vibrio harveyi* have been identified as  $\beta$ -caproylhomoserine lactone and  $\beta$ -hydroxybutyrylhomoserine lactone, respectively.

Marine bacterial luciferase can be immobilized on to a variety of solid supports (e.g., Sepharose, nylon) and it has also been coimmobilized with NAD(P)H:FMN oxidoreductase. The close proximity of the two enzymes leads to efficient channeling of FMNH<sub>2</sub> produced by the oxidoreductase to the luciferase. More extensive coupled coimmobilized systems have been prepared in which all of the enzymes in the enzymatic conversion of glucose to alcohol were coimmobilized with bacterial luciferase and the oxidoreductase, and the bioluminescent enzymes were used to monitor different stages in the biotransformation of glucose.

## Other Systems

### Aequorea

The hydrozoan jellyfish, *Aequorea forskalea*, contains the Ca<sup>2+</sup>-activated photoprotein aequorin (as a family of isoaequorins). Aequorin is composed of apoaequorin and a prosthetic group (coelenterazine). *In vitro* this can be prepared by incubating apoaequorin with coelenterazine in the presence of oxygen. Light emission (460 nm max, rapid flash) is triggered by simply adding a source of Ca<sup>2+</sup>; there is no requirement for oxygen because the luciferin has already reacted with oxygen to form a stable peroxide ('precharged'). Aequorin can also be triggered by

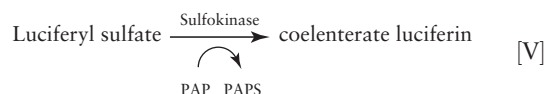
other metal ions (Sr<sup>+</sup>, Ba<sup>2+</sup>, La<sup>3+</sup>, Yb<sup>3+</sup>). The main application of this protein has been as an intracellular calcium indicator.

### Gonyaulax

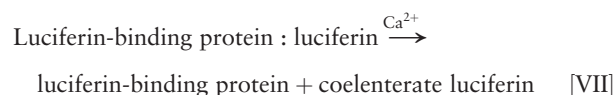
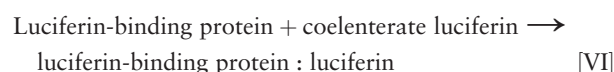
The dinoflagellate *Gonyaulax polyedra* produces brief flashes of light (<0.1 s) that originate from numerous (~400) small (0.5  $\mu$ m) organelles caused scintillons. The scintillons contain an open tetrapyrrole-type luciferin bound to a luciferin-binding protein (dimer, 72 kDa subunits). Change in the pH from 8 to 5.7 releases the luciferin, which then reacts with luciferase (420 kDa) to produce a flash of light at 479 nm. The genes encoding both the luciferase and the luciferin-binding protein have been isolated and cloned. An interesting feature of this bioluminescent organism is that light emission exhibits a circadian rhythm in which cells emit light at night but not during the daytime. This is achieved by translational control of luciferin and luciferin-binding protein synthesis (mRNA levels for these proteins are constant).

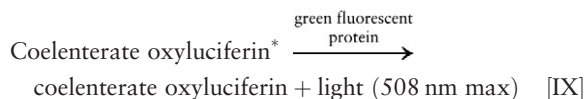
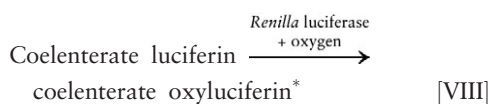
### Renilla

The sea pansy, *Renilla reniformis*, contains a luciferase ( $K_m$  luciferin = 30 nmol l<sup>-1</sup>, pH optimum ~7.6), a luciferin-binding protein (18.5 kDa, two Ca<sup>2+</sup> binding sites) that sequesters and then releases luciferin, and a green fluorescent protein (dimer, 34 + 53 kDa,  $\epsilon = 4.3 \times 10^4$  mol l<sup>-1</sup> cm<sup>-1</sup>, fluorescence quantum yield = 0.7–0.9). Bioluminescence arises from the sequence of reactions shown in reactions [V]–[IX], and involves a charge transfer between green fluorescent protein and the excited-state coelenterate oxyluciferin. The gene for *Renilla* luciferase has been cloned and used to prepare a recombinant luciferase for investigational use. In addition, the gene for green fluorescent protein (*gfp*), isolated from the bioluminescent *Renilla*, has become widely used for gene expression, both *in vivo* and *in vitro*:



(PAP = adenosine 3',5'-diphosphate; PAPS = adenosine 3'-phosphate 5'-phosphosulphate)





## Instrumentation

Instruments for measuring light emission (luminometers) utilize a range of detectors, including film (photographic or X-ray), photomultiplier tubes (in current or single-photon-counting mode), silicon photodiodes, and charge-coupled device (CCD) cameras. The most sensitive detection of light emission is achieved using single-photon-counting techniques. Luminometers vary in size and complexity from simple portable luminometers for use in the field to automated high-throughput instruments capable of measuring light emission from many hundreds of tubes. The kinetics of light emission from some bioluminescent reactions is in the form of a flash of light and thus some luminometers are equipped with injectors to initiate the reaction directly in front of the detector. For specialized applications, such as immunoassay, automatic analyzers that incorporate a luminometer have been produced. These instruments automate sampling, reagent addition, washing, incubation, and aspiration steps in an assay, and finally the initiation and measurement of the light emission.

## Applications

The sensitivity and versatility of bioluminescent reactions has led to a wide range of applications (Table 3). Three bioluminescent reactions, firefly luciferase, marine bacterial luciferase, and the aequorin reaction, dominate and account for more than 90% of the applications. The versatility of bioluminescent reactions stems from the dependence of the reactions on key substances such as ATP (firefly reaction), NAD/NADH (marine bacterial reaction), and metal ions such as calcium (aequorin), and the fact that ATP and NAD can in turn be coupled to kinases and dehydrogenases to measure either the enzyme or its substrate.

### Clinical

There is a diverse range of clinical applications for bioluminescence; however, few have found widespread use in routine clinical laboratories. Detection of ATP as an indicator of cell mass is the most important clinical use of bioluminescence (reaction

[X]). The firefly assay detects 80 fmol ATP, and it is now possible to detect the ATP from fewer than 10 bacterial cells. The presence of abnormally increased numbers of bacteria in urine (bacteriuria) can be detected rapidly by this technique. The bioluminescent detection of ATP using the firefly reaction is also useful in drug susceptibility testing and this is now being used to assess the chemosensitivity of tumor tissue to cancer drugs (a susceptible tumor will not grow in the presence of the appropriate drug and thus the number of cells and hence the ATP

**Table 3** Applications of bioluminescent proteins, photoproteins, and genes

#### Clinical

Assays for enzymes (adenylate kinase, ATPase, creatine kinase, creatine kinase isoenzyme MB, lactate dehydrogenase, proteases)

Assays for substrates (glucose, glycerol, inositol phosphate, malate, oxalate, triglycerides)

ATP assays (rapid microbiology)

Ionized calcium

Lipid peroxidation

Red-cell and platelet viability

Sperm viability

Steroids (androsterone, bile acids, testosterone)

Tumor chemosensitivity

#### Immunoassay and nucleic acid probe assays

*Labels:* recApoaequorin, firefly luciferase, marine bacterial luciferase, *Vargula* luciferase

*Detection reactions:* alkaline phosphatase label (luciferin-O-phosphate/firefly luciferase), glucose 6-phosphate dehydrogenase label (marine bacterial luciferase/NADH:FMN oxidoreductase reaction)

#### Reporter genes (*lux*, *luc*, *ruc*, *vuc*)

Assessing promoter activity

Membrane permeability

Microbial invasion of plants

Monitoring gene expression

#### Food

Bacteria in raw milk

Beverage testing (e.g., microbial contamination of fruit juices)

Brewery hygiene

Meat testing

Microbial contamination of packaging materials

*Salmonella* testing

Sterilization effectiveness

#### Environmental

Aquatic toxicity testing

Microbial adhesion

Pollutants in sewage

Water quality monitoring

#### Industrial

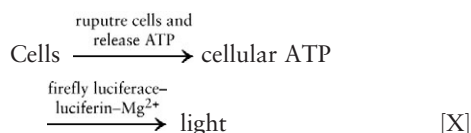
Biocide efficacy tests

Fossil fuel process water testing

Hygiene monitoring in production areas

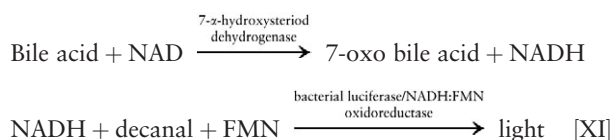
Microbial contamination testing of textiles

concentration will be low):



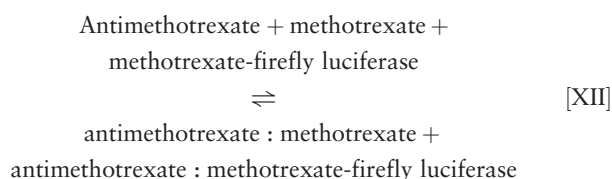
The marine bacterial bioluminescent system when coupled to 7- $\alpha$ -hydroxysteroid dehydrogenase provides a very sensitive (0.5 pmol) and precise method for serum bile acids according to the analytical scheme shown in reaction [XI]. This assay is typical of bioluminescent coupled enzyme assays for either a dehydrogenase enzyme or its substrate.

In order to improve the efficiency of the coupled enzymes, they are used as a coimmobilized preparation (coimmobilized on Sepharose beads). The assay is very simple to perform; a serum sample containing bile acids is mixed with a suspension of the coimmobilized enzymes and the coreactants (NAD, decanal, FMN). Light emission commences almost immediately and the maximum intensity is measured and related to the concentration of bile acid in the sample. Similar reactions can be performed using soluble enzymes; for example, the distribution of metabolites in tissue sections can be determined by incubating the section with a mixture of enzymes and cofactors (e.g., lactate is detected using a mixture of lactate dehydrogenase, bacterial luciferase/oxidoreductase, decanal, FMN, and NAD) and imaging the light emission directly with a CCD or similar type of photon camera:



### Immunoassay and Nucleic Acid Probe Assays

Firefly, marine bacterial, *Renilla* and *Vargula* luciferase, and the photoprotein apoaequorin have all been employed as labels in immunoassays (Table 3). An example of the experimental design for a competitive bioluminescent immunoassay for the drug methotrexate based on a firefly luciferase label is shown in reaction [XII]:



The other major application for bioluminescent reactions is the detection of conventional enzyme

labels. Alkaline phosphatase can be detected down to 10 zmol ( $10 \times 10^{-21}$  mol) using a firefly D-luciferin-O-phosphate substrate. Released luciferin is detected with a mixture of firefly luciferase, ATP, and magnesium ions. Glucose 6-phosphate dehydrogenase has been used as a label in both immunoassay and DNA probe assays. It is detected by quantifying the NADH formed in its reaction with NAD and glucose 6-phosphate using the marine bacterial luciferase/NADH:FMN oxidoreductase reaction.

### Reporter Genes

The gene for firefly luciferase (*luc*) has become one of the most popular reporter genes because the gene product, firefly luciferase, can be detected quickly and sensitively (detection limit for luciferase  $< 10^{-20}$  mol). In many applications it is replacing the more conventional chloramphenicol acetyltransferase gene (CAT) in reporter gene studies. A construct comprising the entire set of marine bacterial genes required for light emission, *luxCDEAB*, can also be assembled and used as a reporter cassette. However, it is the gene for green fluorescent protein (*gfp*) that has become one of the most important reporter genes and is now in widespread use. An advantage of this reporter gene is that it can be visualized using excitation light of the appropriate wavelength (470 nm) without destruction of the cells. Various green fluorescent protein mutants with different light emission maxima in the blue, green, and yellow-green regions of the spectrum provide gene combinations for multicolor analysis.

### Food, Environmental, and Industrial

Most applications for bioluminescence in these three areas are related to the detection of microbes via ATP content using the firefly reaction. Rapid results are very important in many industrial settings, e.g., hygiene monitoring in food production areas, and the rapidity of bioluminescent assays is an attractive and advantageous feature. There is also an increasing number of tests for toxins based on cells containing the *lux* and *luc* reporter genes.

*See also:* **Blood and Plasma. Chemiluminescence:** Overview. **Enzymes:** Immobilized Enzymes; Enzymes in Physiological Samples. **Immunoassays, Techniques:** Luminescence Immunoassays.

### Further Reading

- Bronstein IB, Fortin J, Stanley PE, Stewart GSAB, and Kricka LJ (1994) Chemiluminescent and bioluminescent reporter gene assays. *Analytical Biochemistry* 219: 169–181.  
Campbell AK (1988) *Chemiluminescence*. Chichester: Ellis Horwood.

- Case JF, Herring PJ, Robison BH *et al.* (eds.) (2001) *Bioluminescence and Chemiluminescence*. Singapore: World Scientific.
- DeLuca MA and McElroy WD (eds.) (1986) *Bioluminescence and Chemiluminescence: Methods in Enzymology*, vol. 133. Orlando: Academic Press.
- Harvey EN (1957) *A History of Luminescence – From Earliest Times Until 1900*. Philadelphia: American Philosophical Society.
- Hastings JW and Morin JG (1991) Bioluminescence. In: Prosser CL (ed.) *Neural and Integrative Animal Physiology*, 4th edn. New York: Wiley-Liss.
- Herring PJ (1987) Systematic distribution of bioluminescence in living organisms. *Journal of Bioluminescence and Chemiluminescence* 3: 147–163.
- Hill P, Stewart GSAB, and Stanley PE (1993) Bioluminescence and chemiluminescence literature – luciferase reporter genes – *lux* and *luc*. *Journal of Bioluminescence and Chemiluminescence* 8: 267–291.
- Kricka LJ (1988) Clinical and biochemical applications of luciferases and luciferins. *Analytical Biochemistry* 175: 114–121.
- Meighen EA (1991) Molecular biology of bacterial bioluminescence. *Microbiological Reviews* 55: 123–142.
- Pazzagli M, Cadenas E, Kricka LJ, Roda A, and Stanley PE (1989) *Bioluminescence and Chemiluminescence: Studies and Applications in Biology and Medicine*. Chichester: Wiley.
- Roda A, Pazzagli M, Kricka LJ, and Stanley PE (1999) *Bioluminescence and Chemiluminescence: Perspectives for the 21st Century*. Chichester: Wiley.
- Stanley PE (1992) A survey of more than 90 commercially available luminometers and imaging devices for low-light measurements of chemiluminescence and bioluminescence, including instruments for manual automatic and specialized operation, for HPLC, LC, GLC and microtitre plates. Part I: Descriptions. *Journal of Bioluminescence and Chemiluminescence* 7: 77–108.
- Stanley PE and Kricka LJ (eds.) (2002) *Bioluminescence and Chemiluminescence: Progress and Current Applications*. Singapore: World Scientific.
- Ziegler MM and Baldwin TO (2000) *Bioluminescence and Chemiluminescence. Methods in Enzymology*, vol. 305, Part C. San Diego: Academic Press.

## BIOSENSORS

See **SENSORS: Overview**

## BLEACHES AND STERILANTS

**M A Busch and K W Busch**, Baylor University, Waco, TX, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Bleaching agents remove or prevent unwanted discoloration of a substrate. Chemical sterilants and disinfectants destroy undesirable microorganisms on contact. This article surveys the more important chemicals that are employed for these purposes and discusses the most important analytical methods that are used for their determination. Because the residual levels of disinfectants following municipal water treatment are subject to governmental regulation, this article gives particular emphasis to the standard methods for determining chlorine residuals in drinking water.

### Major Areas of Use and Application

Bleaching agents are formulations that whiten or lighten a substrate by solubilizing color-producing

substances or by altering their light-absorbing properties. Bleaching agents are used extensively in the pulp and paper industry, the textile industry, and commercial and household laundering.

The chemistry of bleaching agents is predominantly that of oxidizing agents: chlorine ( $\text{Cl}_2$ ) and some of its compounds or peroxygen species such as hydrogen peroxide ( $\text{H}_2\text{O}_2$ ), ozone ( $\text{O}_3$ ), and sodium perborate ( $\text{NaBO}_3$ ). The decolorizing reaction generally involves the removal of chromophoric sites in which electron delocalization over conjugated double bonds has made the substrate capable of absorbing visible light. The bleaching agent will typically react by irreversibly cleaving or adding across these double bonds.

A few bleaching compounds act by chemical reduction; these include sulfur dioxide ( $\text{SO}_2$ ), sulfurous acid ( $\text{H}_2\text{SO}_3$ ), hydrogensulfite ( $\text{HSO}_3^-$ ), sulfite ( $\text{SO}_3^{2-}$ ), and dithionite ( $\text{S}_2\text{O}_4^{2-}$ ), as well as sodium tetrahydroborate (borohydride) ( $\text{NaBH}_4$ ). Their application is primarily in pulp and textile manufacturing, where the bleaching action is thought to occur by reduction of a chromophoric carbonyl group.



Other applications include the bleaching of glues, gelatin, soap, and food products.

Chemical bleaching of textiles (as opposed to bleaching by sunlight, a process known as crofting) had its advent soon after the discovery of the element chlorine by Scheele in 1774. Finding that aqueous solutions of chlorine gas weakened the fiber of textiles, Bertholet experimented with solutions of potassium hypochlorite ( $\text{KOCl}$ ), made by dissolving chlorine gas in a solution of caustic potash ( $\text{KOH}$ ). Labarraque made hypochlorite solutions industrially economic by replacing the more expensive caustic potash with caustic soda ( $\text{NaOH}$ ). The liquid household bleach normally available today is a 5% solution of sodium hypochlorite ( $\text{NaOCl}$ ).

Sterilants are strictly defined by the US Environmental Protection Agency as substances that totally destroy all forms of life, including viruses, bacteria, fungi, and their spores on inanimate surfaces, in water, or in the air. Many substances that render objects microbiologically safe for certain applications (i.e., reduce the level of living microorganisms below some predetermined level) are also commonly referred to as sterilants.

Sterilization can be accomplished by treatment with chemicals, as well as by heat, cold, and radiation. Chemical sterilants are also typically oxidizing agents, and their effectiveness is related to their ability to oxidize the cell wall of the microorganism and to diffuse through the cell wall and disrupt cellular activity. Many of the same substances that act as bleaching agents (Table 1) are also used extensively as sterilants.

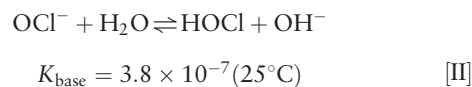
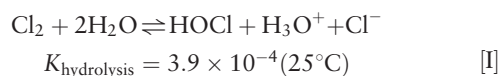
Chemical disinfection with chlorine gas was recommended as early as 1801 and was widely installed in British and European hospitals by 1823. Following the discovery of the microbial basis for contagious diseases, Robert Koch (1881) demonstrated that hypochlorites were effective in killing bacteria. In 1905, following an outbreak of typhoid fever in

London, continuous chlorination of the public water supply was inaugurated. Chemical disinfection of a public water supply was first practiced in the United States in 1908, and since that time has been a routine part of municipal water treatment. Chemical disinfectants are also used extensively in the production of bottled water and the manufacture of containerized beverages of all types, in the treatment of waste water, and in the treatment of water in swimming pools and spas.

## Chemicals Used as Bleaches and Sterilants

A comprehensive discussion of the analytical chemistry of bleaches and sterilants is complicated by the diversity of chemicals that are employed. The following is a brief survey of some of the more important chemicals that are used for these purposes.

Modern chlorine-based bleaching agents include elemental chlorine ( $\text{Cl}_2$  gas), a variety of hypochlorites ( $\text{OCl}^-$ ), certain *N*-chlorinated organic compounds, and chlorine dioxide ( $\text{ClO}_2$ ). The first three bleaching agents all hydrolyze to produce hypochlorous acid in aqueous solution, according to the equilibria in reactions [I]–[III]



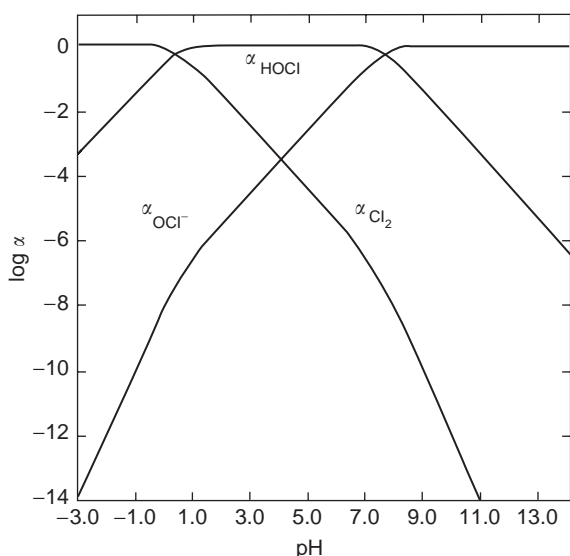
where R and R' represent a variety of substituents.

Figure 1 shows the effect of pH on the chemical form of dissolved chlorine in aqueous solution for a fixed chloride ion concentration of  $1 \times 10^{-3} \text{ mol l}^{-1}$ . The neutral hypochlorous acid molecule, which is thought to be more effective in penetrating cell walls than the charged hypochlorite ion, has a disinfecting ability that is greater than that of  $\text{OCl}^-$  by 2 orders of magnitude. As a result, an aqueous solution of chlorine will have its maximum disinfecting ability over a pH range from  $\sim 2$  to 7.

Solutions of hypochlorites and hypochlorous acid are subject to gradual deterioration, forming chlorite ( $\text{ClO}_2^-$ ) and sometimes chloride ( $\text{Cl}^-$ ) and oxygen ( $\text{O}_2$ ) over a period of time. In general, the lower the pH, the less stable the solution. (Solid preparations of hypochlorite salts are quite stable if kept dry and cool.)  $\text{HOCl}$  and  $\text{ClO}^-$  react strongly with organic

**Table 1** Oxidizing strengths of some typical bleaching agents and disinfectants as represented by their standard reduction potentials

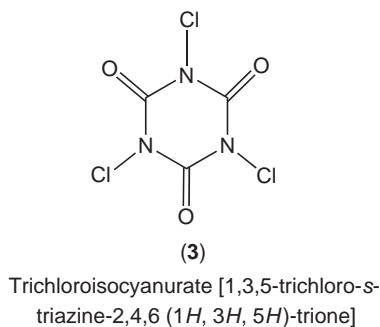
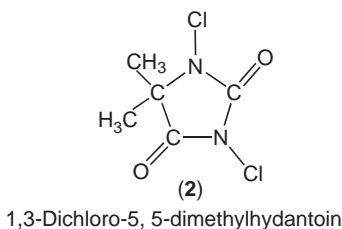
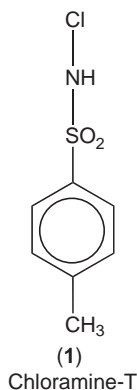
Compound	Standard reduction potential (V)	Half-cell
$\text{Cl}_2$	+1.3595	$\text{Cl}_2 + 2\text{e}^- \rightarrow 2\text{Cl}^-$
$\text{Br}_2$	+1.0652	$\text{Br}_2 + 2\text{e}^- \rightarrow 2\text{Br}^-$
$\text{I}_2$	+0.5355	$\text{I}_2 + 2\text{e}^- \rightarrow 2\text{I}^-$
$\text{O}_3$	+2.07	$\text{O}_3 + 2\text{H}_3\text{O}^+ + 2\text{e}^- \rightarrow \text{O}_2 + 3\text{H}_2\text{O}$
$\text{ClO}_2$	+1.91	$\text{ClO}_2 + 2\text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{Cl}^- + 4\text{OH}^-$
	+0.95	$\text{ClO}_2 + \text{e}^- \rightarrow \text{ClO}_2^-$
$\text{H}_2\text{O}_2$	+1.776	$\text{H}_2\text{O}_2 + 2\text{H}_3\text{O}^+ + 2\text{e}^- \rightarrow 4\text{H}_2\text{O}$



**Figure 1** Distribution diagram for aqueous chlorine in a closed system at 25°C with  $[\text{Cl}^-] = 1 \times 10^{-3} \text{ mol l}^{-1}$ . The fraction present in a given form ( $\alpha$ ) is equal to the concentration of the given species divided by  $C_T$  where  $C_T = [\text{Cl}_2] + [\text{HOCl}] + [\text{OCl}^-]$ .

matter and, for maximum effectiveness, organic contamination must be removed prior to use.

Examples of *N*-chlorinated organic compounds that gradually hydrolyze to HOCl include chloramine-T (*N*-chloro-*p*-toluenesulphonamide) (1), *N*-chlorohydantoin (2), and various chlorinated isocyanurates (3). They are mainly employed as disinfectants, because their low solubility and/or hydrolysis in water affords poor bleaching ability. The one exception is sodium dichloroisocyanurate dihydrate, which is the most water soluble, the fastest to dissolve, and the least hazardous.



Chlorine dioxide is fast replacing aqueous  $\text{Cl}_2$ , particularly in pulp and paper manufacture, because the reaction of  $\text{ClO}_2$  with organic materials does not appear to form carcinogenic trihalomethanes (THMs) as side-products and because  $\text{ClO}_2$  is 10

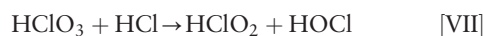
times more effective as a disinfectant under high chlorine demand. Chlorine dioxide is an explosive gas and is usually prepared onsite by the hydrolysis of an alkali-stabilized solution of sodium chlorite ( $\text{NaClO}_2$ ) (see reactions [IV] and [V]):



Somewhat higher yields are afforded by reacting sodium chlorite with chlorine gas (reaction [VI]):



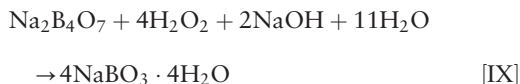
Excess chlorine must be removed from the mixture prior to use to prevent THM formation. Somewhat larger amounts of chlorine dioxide, such as needed in pulp processing and textile bleaching, are generally prepared by reduction of chlorate with hydrochloric acid (reactions [VII] and [VIII]):



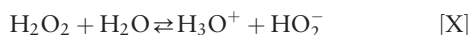
Hydrogen peroxide and other peroxy compounds derived from hydrogen peroxide are also used as bleaching agents. These compounds all contain the peroxide linkage ( $-\text{O}-\text{O}-$ ). Liquid solutions of hydrogen peroxide are commercially available (30–35%, 50%, or 65–70% by weight); however, stable solid peroxy compounds that hydrolyze in solution to give hydrogen peroxide offer improved

stability and convenience in handling. Typical among this class of solid bleaching agents (so-called safety bleaches) is sodium perborate, which is formed by reacting borax (sodium metaborate) with hydrogen peroxide in the presence of base

(reaction [IX]):



Hydrogen peroxide is a very weak acid and is relatively stable in acidic solution. Bleaching is performed in alkaline solution, where part of the hydrogen peroxide is converted to the perhydroxyl anion ( $\text{HO}_2^-$ ), believed to be the active bleaching agent (reaction [X]):



Disproportionation to water and oxygen is markedly accelerated by the presence of heavy metals, and metal chelating agents are often added to solutions of  $\text{H}_2\text{O}_2$  as stabilizers.

Concern over the adverse health effects of trace amounts of THMs produced during the chlorination of drinking water has led to the use of nonchlorinated disinfection agents, such as ozone ( $\text{O}_3$ ), in the treatment of domestic water, particularly in Europe. Ozone has not been extensively employed in the United States because it is difficult to maintain an active residual once the water leaves the treatment plant. Ozone gas is also used in hospitals to sterilize instruments and surgical dressings.

Ozone decomposes rapidly to oxygen ( $\text{O}_2$ ) and must be generated on demand, in close proximity to the utilization site. Ozone is generated by passing air or oxygen through a high-voltage corona discharge. Aqueous solutions are prepared by bubbling the gas through a column of water.

Sulfur dioxide ( $\text{SO}_2$ ) is a gas, formed by burning sulfur in air. When dissolved in water,  $\text{SO}_2$  forms sulfurous acid, a complex mixture of  $\text{SO}_2$ ,  $\text{H}_3\text{O}^+$ ,  $\text{S}_2\text{O}_3^{2-}$ , and  $\text{HSO}_3^-$ . All of these species are good reducing agents, making the bleaching effect independent of pH over the range 3–10.

## Survey of Analytical Methods

In industrial operations like pulp and paper manufacturing and textile bleaching, the selection of analytical methodology is largely dictated by process analytical and quality control requirements. By contrast, the methods used for the determination of residual levels of disinfectants in drinking water are subject to regulation by national and local governments. As a result, 'standard methods' for these procedures, which are recognized by courts in any legal action, are continuously being developed and revised. Up-to-date, detailed procedures for the determination of residual levels of disinfectants in

drinking water can be found in the latest publications and reports from the sources listed below:

- 'Standard Methods for the Examination of Water and Wastewater', published jointly by the American Public Health Association, the American Water Works Association, and the Water Pollution Control Federation, Washington, DC, USA.
- 'US Environmental Protection Agency', Analytical Reference Service Reports, Municipal Environmental Laboratory, Cincinnati, Ohio, USA.
- 'Standing Committee of Analysts', Department of the Environment, London, UK.
- 'American Water Works Association', Research Foundation Reports, Denver, Colorado, USA.
- 'Guidelines for Canadian Drinking Water Quality', Health and Welfare Canada, Ottawa, Ontario, Canada.

Since these methods can generally be adapted for the analysis of bleaching agents in industrial processes, this article will focus on the standard analytical procedures developed for water analysis.

## Determination of Chlorine Residuals in Drinking Water

Because chlorine-based disinfectants are widely used in water treatment, much effort has been invested in the development of analytical methods for their determination in municipal water samples over the concentration range from 0.001 to  $10 \text{ mg l}^{-1}$  (expressed as  $\text{Cl}_2$ ). The overriding challenge in the development of these procedures has been the need to distinguish between the free and combined forms of residual chlorine. This terminology is highly specific and, in order to clarify the discussion that follows, the definitions of these and other relevant terms are listed below.

- Chlorine demand: free available chlorine (FAC) required to react with all oxidizable species present in the water sample.
- Combined (available) chlorine: total available chlorine in the form of undissociated *N*-chloro compounds that can hydrolyze according to reaction [III].
- Free (available) chlorine: total available chlorine in the form of dissolved  $\text{Cl}_2$  gas, hypochlorous acid ( $\text{HOCl}$ ), and hypochlorite ion ( $\text{ClO}^-$ ).
- Residual chlorine: total available chlorine remaining after reaction of FAC with oxidizable species present in the water sample.
- Total (available) chlorine: total concentration of chlorine-based oxidants in the form of dissolved



$\text{Cl}_2$  gas, hypochlorous acid ( $\text{HOCl}$ ), hypochlorite ion ( $\text{ClO}^-$ ), and  $N$ -chlorinated organic compounds that can hydrolyze according to reaction [III].

When chlorine is added to a surface water or groundwater, a number of species present in the water can consume chlorine, and such reactions are said to constitute the chlorine demand of the water. Reactions with inorganic species such as iron, manganese, and sulfides produce chloride ion ( $\text{Cl}^-$ ), which has no residual oxidizing power. However, nitrogenous bases, such as ammonia ( $\text{NH}_3$ ), urea, and amino acids, react to form chloramines and other organically bound chlorine compounds. These combined forms of chlorine (called combined available chlorine) typically retain some disinfectant properties, but are generally weaker than solutions containing hypochlorous acid and hypochlorites in an uncombined state (FAC). The combined as well as free forms of available chlorine that remain after the chlorine demand of the water has been satisfied constitute the chlorine residual.

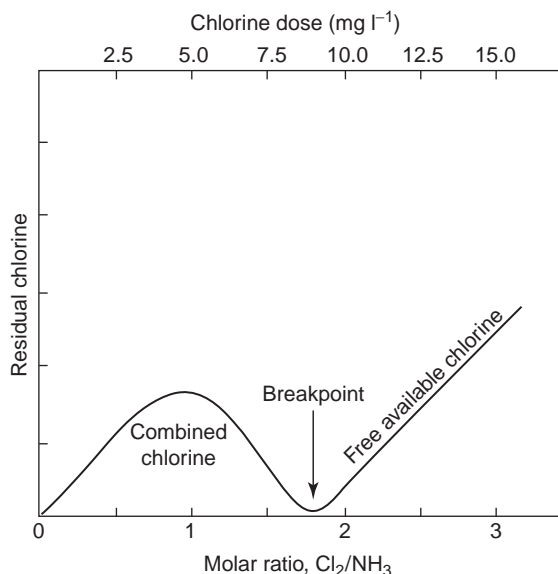
An FAC residual is regarded by the US Environmental Protection Agency as a sign of adequate disinfection. The problem encountered in ensuring that water leaving the treatment plant has been treated with enough chlorine to leave an FAC residual is illustrated by breakpoint chlorination. In breakpoint chlorination, dissolved chlorine is added to the water in a stepwise manner to determine the chlorine demand and to allow for the formation of chloramines.

Figure 2 is an idealized plot of residual chlorine as a function of the chlorine dose added to the water sample that contains only ammonia. Initially, addition of chlorine causes the rapid formation of monochloramine ( $\text{NH}_2\text{Cl}$ ), and the chlorine residual rises. A maximum in the curve occurs at a  $\text{Cl}_2/\text{NH}_3$  molar ratio near 1, where residual chlorine is present only as monochloramine and some unstable dichloramine ( $\text{NHCl}_2$ ). At a higher chlorine dose, nitrogen gas ( $\text{N}_2$ ), nitrate ( $\text{NO}_3^-$ ), and chloride are formed (reactions [XI] and [XII]),



causing the residual chlorine to decline. Finally, when all of the ammonia has been consumed, a minimum (breakpoint) is reached. At this breakpoint, the chlorine demand of the water has been completely satisfied, and further addition of chlorine causes the chlorine residual (now present as FAC) to increase linearly.

Since governmental regulations specify an FAC residual for adequate disinfection, accurate analytical

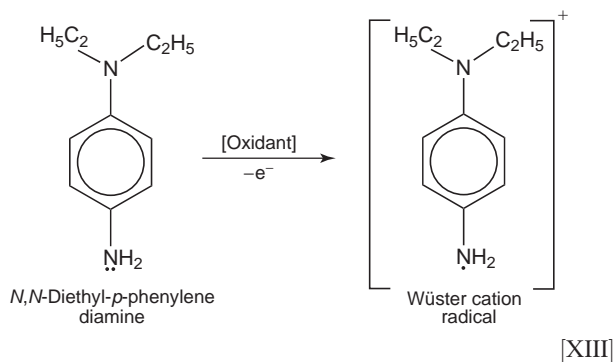


**Figure 2** Idealized breakpoint chlorination curve showing chlorine residual as a function of chlorine dose for a water sample at pH 7 that contains only ammonia at a concentration of 1 mg nitrogen (as  $\text{NH}_3$ ) per liter.

methods are required for the determination of FAC at levels less than  $10 \text{ mg l}^{-1}$  in the possible presence of chloramines. If combined chlorine is measured along with free chlorine, the true FAC will be overestimated, and the disinfectant level may be inadequate.

Although no ideal method exists, at present, that can distinguish between free and combined chlorine without interference, analytical methods based on amperometry and the use of  $N,N$ -diethyl- $p$ -phenylene diamine (DPD) are the most frequently employed in water analysis. With both of these methods, FAC is first determined in the absence of iodide ( $\text{I}^-$ ). Iodide can then be added to the solution, and the iodine (present as triiodide ion,  $\text{I}_3^-$ ), formed from the oxidation of iodide by combined chlorine, is determined.

**Methods using DPD** DPD is reversibly oxidized to a magenta-colored cation radical known as a Wüster cation [XIII]:



This reagent has replaced neutral orthotoluidine as the indicator in the standard titrimetric determination of FAC by ammonium iron(II) sulfate ((NH<sub>4</sub>)<sub>2</sub>Fe(SO<sub>4</sub>)<sub>2</sub>) (reaction [XIV]):



In the DPD procedure for FAC, the titration is continued until the red color of the indicator disappears. Because mono- and dichloramines hydrolyze slowly in solution to release hypochlorous acid (reaction [III]), titrations must be carried out rapidly to reduce the interference of combined chlorine.

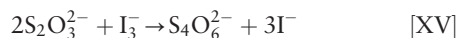
The concentration of the Wüster cation produced by reaction with FAC can also be determined spectrophotometrically at 515 nm. The absorbance should be measured within 1 min following the addition of DPD reagent to avoid interference of combined chlorine. Calibration standards for the spectrophotometer can be prepared from previously standardized chlorine solutions (by titration with ammonium iron(II) sulfate) or by using standardized potassium permanganate (KMnO<sub>4</sub>) solutions to develop the DPD color.

Both turbidity and colored water samples can produce an interference in the spectrophotometric procedure. Oxidizing agents such as bromine, iodine, chlorine dioxide, ozone, and oxidized forms of manganese will also interfere with either the titrimetric or the spectrophotometric DPD procedure.

In aqueous solution, the DPD indicator is catalytically oxidized by oxygen when trace metals are present. In the standard method, the DPD solution is mixed with a strong phosphate solution, buffered at pH 6.2, and ethylenediaminetetraacetate (EDTA) is added to complex trace metals. Thioacetamide (CH<sub>3</sub>CSNH<sub>2</sub>) can be added immediately following the addition of DPD to decrease the interference from monochloramine. Addition of mercury(II) chloride (HgCl<sub>2</sub>) to the phosphate buffer has also been found to reduce the interference from monochloramine, presumably by complexation of trace iodide by Hg<sup>2+</sup>.

When both free and combined chlorine are to be measured, the FAC is determined first using an excess of DPD. Potassium iodide is then added to the solution and reacts readily with chloramines to produce I<sub>2</sub>, which, in turn, reacts with part of the excess DPD. If only a small amount of KI is added, monochloramine will be the primary reactant. If an excess of KI is added, both mono- and dichloramine can be made to produce I<sub>2</sub>. Thus, sequential addition of KI to the sample can, in principle, allow the various chlorine fractions to be determined.

**Iodometric methods** Iodometric titrations can be used as reference methods for the determination of total available chlorine at concentrations greater than 1 mg l<sup>-1</sup> (as Cl<sub>2</sub>). In this method, an excess of KI is added to the sample. All oxidizing agents having *E*<sup>o</sup> values much greater than +0.54 V will react to form iodine (present as I<sub>3</sub><sup>-</sup>), which is titrated with a standard solution of sodium thiosulfate (Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub>) (reaction [XV]):



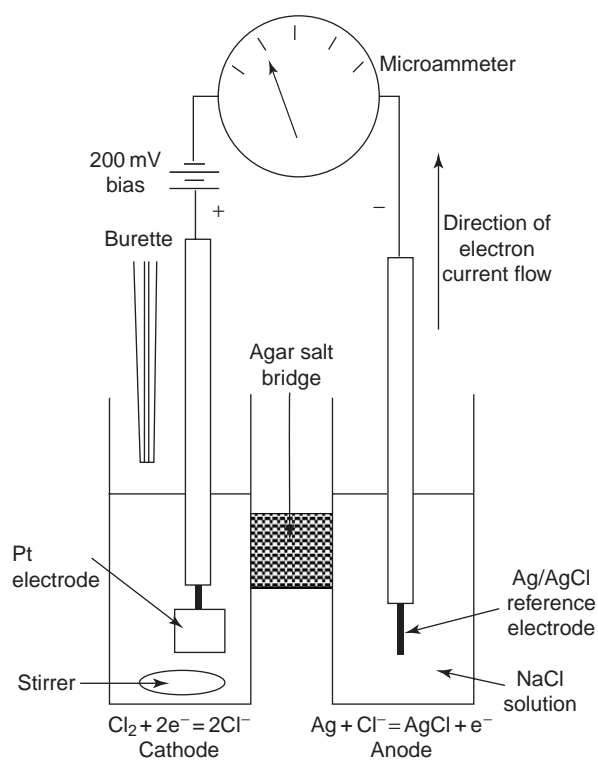
The equivalence point is taken as the volume of titrant required to react with all of the I<sub>3</sub><sup>-</sup> formed. The endpoint can be determined using a starch indicator (i.e., at the disappearance of the blue starch-iodine complex) or by amperometry.

For polluted waters or chlorinated waste waters where residual chlorine is present mainly in the combined form, analytical methods do not need to distinguish between FAC and combined available chlorine. Iodometric titration is not recommended because these water samples typically contain small amounts of reducing agents that can react with, and thereby reduce, the liberated iodine over the course of the titration. In this case, an iodimetric procedure is recommended. Arsenic(III) (as arsenic trioxide, As<sub>2</sub>O<sub>3</sub>) is added to the sample in a known amount that is in excess of the combined chlorine, and unreacted As(III) is titrated with a standard solution of I<sub>3</sub><sup>-</sup> to a starch endpoint (formation of a blue color).

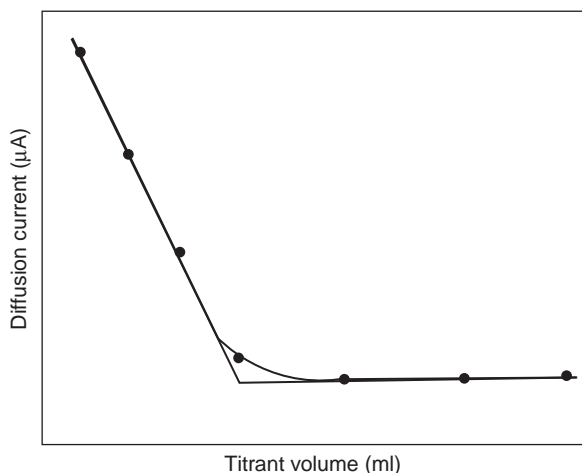
**Amperometric methods** Amperometric methods are based on the measurement of current produced in an electrochemical cell at an appropriate applied voltage. **Figure 3** shows a typical cell employed for the amperometric determination of dissolved chlorine, which consists of a reference electrode, whose potential is fixed, and a working electrode, whose potential can be adjusted. If a reducible substance is present in the cell, and the cell potential is sufficiently high, the diffusion current will be proportional to the concentration of reducible species.

In amperometric titrations, the diffusion current is plotted versus the volume of titrant added to give a titration curve consisting of two straight-line segments. **Figure 4** corresponds to the titration of a reducible analyte with a titrant that is not reducible. The titrant volume corresponding to the intersection of the two straight-line segments is taken as the endpoint.

In the amperometric determination of dissolved chlorine (**Figure 3**), an Ag/AgCl reference electrode is

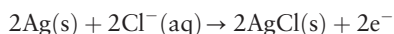
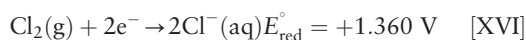


**Figure 3** Amperometric titration apparatus for the determination of dissolved chlorine.

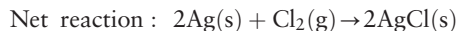


**Figure 4** Amperometric titration curve showing diffusion current versus titrant volume for a reducible analyte and a titrant that is not reducible.

employed, and a spontaneous cell is formed between the platinum electrode (reaction [XVI]) and the reference electrode (reaction [XVII]):



$$E_{\text{ox}}^\circ = -0.222 \text{ V} \quad [\text{XVII}]$$

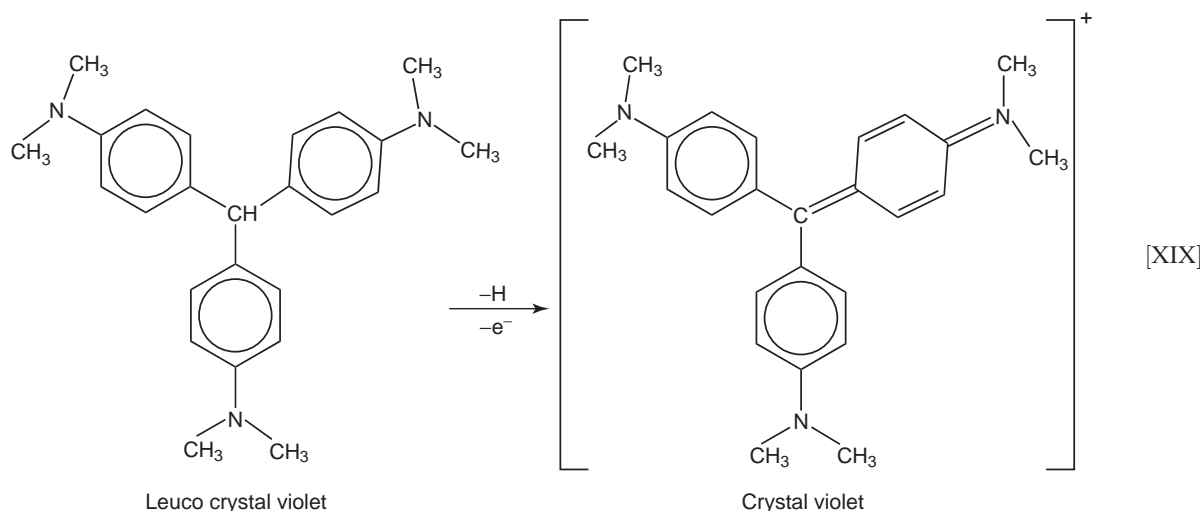


$$E_{\text{cell}}^\circ = +1.138 \text{ V} \quad [\text{XVIII}]$$

In these equations, the reactions have been represented in terms of standard-state conditions. The true voltage will be somewhat less than that shown in reaction [XVIII] because the  $\text{Cl}_2$  is present in solution and its vapor pressure will be much less than 1 atm, as implied for standard-state conditions. Nevertheless, it is apparent that the voltage produced by this spontaneous cell is more than sufficient to provide limiting current conditions at the platinum electrode. However, most standard methods suggest that +200 mV be applied to the platinum electrode to improve sensitivity (Figure 3). Amperometric titration is the standard method for comparison of accuracy in determining residual chlorine in drinking water. FAC is determined by titration with a standard solution of As(III) (in the form of phenylarsine oxide,  $\text{C}_6\text{H}_5\text{AsO}$ ) at a pH between 6.5 and 7.5. In this pH range, combined chlorine reacts slowly. Total chlorine may be determined by addition of KI to the sample, followed by amperometric titration of the iodine (liberated by the free and combined forms of chlorine) using a standard solution of phenylarsine oxide.

By comparison with spectrophotometric procedures, amperometric methods require somewhat greater skill. The solution must be well mixed and contain sufficient inert supporting electrolyte to reduce the ohmic potential gradients that interfere with the development of a uniform diffusion layer through which the electroactive species must diffuse in order to reach the working electrode. For accurate work, however, care must also be taken to avoid volatilization losses of chlorine from solution by mechanical stirring.

**Other spectrophotometric methods** Other methods for the determination of dissolved chlorine include the use of leuco crystal violet and syringaldazine as color-developing reagents. In the leuco crystal violet method, a colorless precursor (leuco crystal violet) is oxidized to crystal violet (reaction [XIX]), which has an absorption maximum at 588 nm. Interference from combined available chlorine can be avoided if the test is carried out within 5 min of indicator addition. Total chlorine determination can be accomplished by the addition of iodide ion to produce hypiodous acid (HIO), which reacts instantly with the indicator. Combined available chlorine is determined by difference.

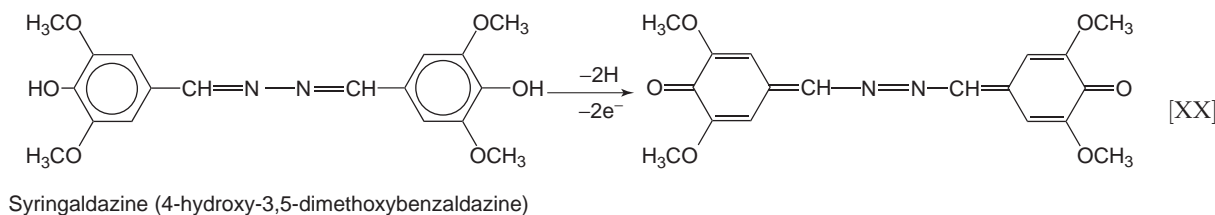


In the free available chlorine test (FACTS), syringaldazine (formed by reacting syringaldehyde with hydrazine dihydrochloride in basic solution) is dissolved in 2-propanol. At a pH between 6 and 7, an aqueous solution of this reagent has an intense yellow color, but upon oxidation by FAC forms a violet, quinone-type product that has an absorption maximum at 530 nm (reaction [XX]). The pH range between 6 and 7 represents a compromise condition that produces rapid color development but slow color fading. Monochloramine (up to  $18 \text{ mg l}^{-1}$ ), dichloramine (up to  $10 \text{ mg l}^{-1}$ ), and oxidized forms of manganese (up to  $1 \text{ mg l}^{-1}$ ) do not interfere with the determination of FAC with syringaldazine. The primary difficulty with the FACTS method is the relative insolubility of the syringaldazine in 2-propanol or water, which causes problems in reagent preparation and color stability.

upon being oxidized, may provide increased sensitivity.

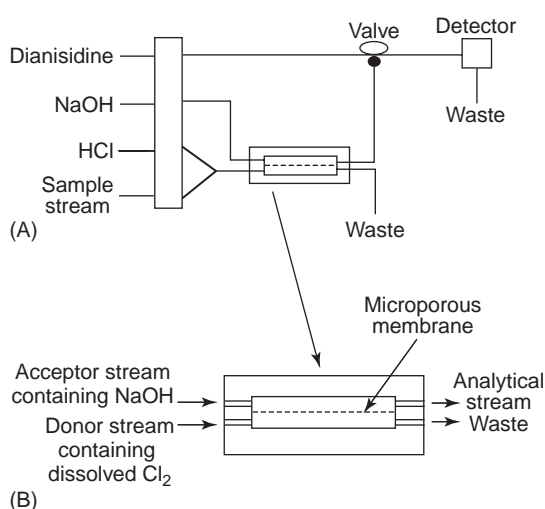
Bromine in aqueous solution has been determined fluorometrically with rhodamine B. The method can be adapted for the determination of dissolved chlorine by using the dissolved chlorine to form bromine in the presence of an excess of bromide ion.

Flame infrared emission (FIRE) spectrometry is a new technique that is useful in determining FAC in liquid bleach. In the FIRE method, solutions of sodium hypochlorite are acidified to produce aqueous  $\text{Cl}_2$  (reactions [I] and [II] and **Figure 1**). Dissolved  $\text{Cl}_2$  is liberated from solution in a purge tube and converted to vibrationally excited HCl molecules in a hydrogen-air flame. The intensity of the P-branch of the HCl stretching vibration at  $3.8 \mu\text{m}$  is monitored with a simple filter infrared photometer that employs a lead selenide detector.



**Other spectral methods** Red chemiluminescence is observed at 635 nm when hydrogen peroxide reacts with hypochlorite ion in alkaline solution. Unfortunately, the reaction is not sensitive enough for application to potable waters; however, substances such as luminol or lophine, which give chemiluminescence

By contrast with other methods for the determination of available chlorine, which are all based on oxidation-reduction chemistry and subject to potential interference from other oxidants present in the sample, FIRE actually measures a signal that is solely related to the amount of dissolved  $\text{Cl}_2$ .



**Figure 5** (A) Flow-injection assembly. (B) Detail of gas diffusion unit.

The sample is purged within seconds following addition of acid, and is not affected by the presence of chlorinated compounds that hydrolyze slowly in acid solution. A flow injection procedure has recently been developed that makes use of a continuous purge cell.

**Flow-injection methods** Flow-injection techniques have been developed and studied for the determination of free chlorine in industrial formulations and water samples. Figure 5 shows a schematic diagram of a flow-injection analysis (FIA) assembly that employs a gas-permeable membrane (0.5  $\mu\text{m}$  Fluoropore, Millipore) as a gas diffusion unit to separate dissolved chlorine from other potential matrix interferences.

With this apparatus, the sample (donor stream) is first mixed with hydrochloric acid to produce dissolved chlorine gas,

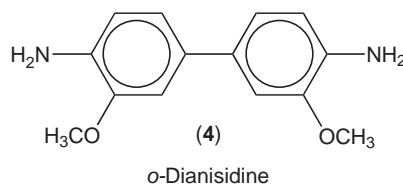


which subsequently diffuses across the gas-permeable membrane. Upon diffusing across the membrane, the  $\text{Cl}_2$  is converted back to hypochlorite ion by mixing with a NaOH acceptor stream,



Use of a basic acceptor stream facilitates transport of  $\text{Cl}_2$  across the membrane while, at the same time, permits preconcentration of the sample by stopping the flow of NaOH solution for a period of time. Following the preconcentration period, the flow of the acceptor stream is resumed and the sample is then mixed with *o*-dianisidine (4), producing a colored product that can be monitored at 445 nm. With a preconcentration period of a minute, limits of detection for chlorine

on the order of  $0.04 \text{ mg l}^{-1}$  can be achieved:



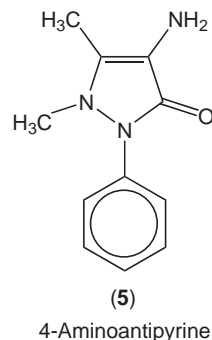
Gas diffusion FIA has also been used for the online determination of chlorine dioxide in potable water. In this example, chlorophenol red was used as the chromogenic reagent and a detection limit of  $0.02 \text{ mg l}^{-1}$  was achieved.

**Chromatographic methods** Chlorine dioxide is being used with increasing frequency in kraft mills and for water treatment because it does not form potentially carcinogenic THMs the way chlorine does. Since iodometric methods for the determination of chlorine dioxide are questionable, they should be avoided. While chlorine dioxide can be determined by amperometric titration and FIA with absorption spectrophotometric detection, chromatographic methods are becoming increasingly popular. Both high-performance liquid chromatography and ion chromatography (IC) have been used to determine inorganic chlorine species in aqueous samples. Since chlorine dioxide itself is not ionized in solution, it must be converted to  $\text{ClO}_2^-$  for IC analysis. To accomplish this conversion, samples are prepared in a basic solution containing hydrogen peroxide,



The  $\text{ClO}_2$  present is determined by difference between the chlorite found in a treated sample and that found in a separate aliquot that has not been dissolved in basic hydrogen peroxide.

Chlorine dioxide and hypochlorite have also been determined simultaneously in aqueous samples by high-performance liquid chromatography. The analytes are first separated on an anion-exchange column. After separation, a postcolumn reaction with 4-aminoantipyrine (5), and phenol is carried out, forming a chromogenic substance that absorbs at 503 nm through an oxidation-condensation reaction.



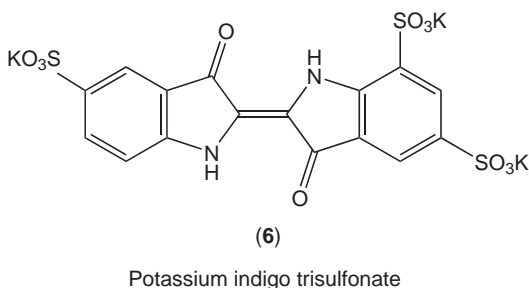


### Determination of Other Bleaching Agents and Disinfectants

**Oxidizing agents** Other oxidizing agents like peroxides and ozone are typically determined using variations of the methods used to determine FAC in solution. For example, peroxides may be determined by iodometric titration, by chemiluminescence, or by means of a colorimetric method based on syringaldazine.

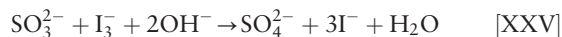
Determination of ozone in aqueous solution is perhaps the most problematic for a variety of reasons: (1) ozone is unstable; (2) ozone is volatile and easily lost from solution; and (3) ozone reacts with many organic compounds to form products such as ozonides and hydrogen peroxide that are also good oxidants. Careful study of the use of iodometric methods for the determination of ozone in aqueous solution has revealed that the stoichiometric ratio of ozone reacted with iodine produced in the reaction varies from 0.65 to 1.5, depending on pH, buffer composition and concentration, iodide ion concentration, and other reaction conditions. As a result, iodometric methods are not recommended. Ozone can be determined iodimetrically by addition of an excess of a standard solution of As(III), followed by titration of the excess As(III) with a standard solution of iodine to a starch endpoint. Methods using DPD, syringaldazine, and amperometric titrations have also been developed.

In comparison to standard procedures involving iodide/iodine, colorimetric determination of ozone with indigo trisulfonate (**6**) has fewer interferences. In this procedure, indigo trisulfonate is oxidized by ozone to a leuco (colorless) form, and the decrease in absorbance is monitored at 591 nm. Ozone decomposition products and ozonolysis products formed from organic compounds do not interfere. Moreover, chlorite ( $\text{ClO}_2^-$ ), chlorate ( $\text{ClO}_3^-$ ), perchlorate ( $\text{ClO}_4^-$ ), and hydrogen peroxide do not decolorize the indigo reagent.

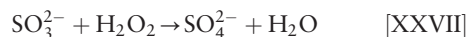


**Reducing agents** Methods for the determination of dissolved sulfite and dithionite generally involve reaction with oxidants. For example, aqueous solutions of sulfite can be determined by iodometric

titration to a blue, starch endpoint (reaction [XXV]):

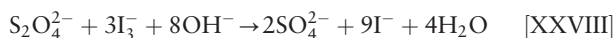


Hydrogen sulfite can be determined in the presence of sulfite by oxidation with hydrogen peroxide to form hydrogen sulfate and sulfate (reactions [XXVI] and [XXVII])



followed by titration of the acidic hydrogensulfate ion with sodium hydroxide.

Aqueous solutions of dithionite ion can also be determined by iodometric titration to a blue starch endpoint (reaction [XXVIII]):



Dithionite solutions can be determined without interference from sulfite by the addition of formaldehyde. Formaldehyde reacts with sulfite ion to form an adduct that is not oxidized by iodine. Spectrophotometric methods for dithionite solutions are based on the reductive bleaching of dyes.

### Sampling

Aqueous solutions of chlorine are not stable and decrease in strength with time. Exposure to strong light accelerates the decomposition. Strong agitation should be avoided to prevent loss of chlorine gas by volatilization. Samples should be analyzed as soon as possible after collection and not stored. Most analytical methods require 100–200 ml of sample.

Because of the relative instability of chlorinated water samples, determination in the field is often required. Colorimetric test kits are available commercially for this purpose.

### Precautions in Handling

Bleaching agents and disinfectants are all strong oxidants that in concentrated form can react explosively with reductants, including organic matter. Disinfectants are effective because they are toxic to microorganisms, and in sufficient concentration will produce adverse health effects in humans. Chlorine gas, which is commercially available in steel cylinders as a gas over liquid, is a powerful irritant that can cause fatal pulmonary edema. Inadvertent addition of acid to liquid bleach ( $\text{NaOCl}$ ) causes the release of poisonous  $\text{Cl}_2$ . While chlorine dioxide dissolved in water is stable (if kept cool and away from light), the gas will detonate at pressures above

300 Torr. To avoid formation of explosive concentrations of  $\text{ClO}_2$  above the solution, the concentration of chlorine dioxide must be kept below 5 g  $\text{ClO}_2$  per liter of  $\text{H}_2\text{O}$ . Concentrated solutions of hydrogen peroxide also react explosively with organic matter.

Preparation of standard solutions of iodine and the amperometric determination of chlorine in water samples make use of arsenic trioxide ( $\text{As}_2\text{O}_3$ ) or phenylarsine oxide ( $\text{C}_6\text{H}_5\text{AsO}$ ), respectively, as reagents. These titrations generate toxic waste that must be disposed of in an environmentally safe manner. The cost of collecting and disposing of these reagents must be included in the cost of performing such determinations. Crystal violet is a suspected carcinogen and solutions formed from the leuco crystal violet method should also be treated as toxic waste.

**See also:** **Amperometry.** **Atomic Emission Spectrometry:** Flame Photometry. **Chemiluminescence:** Overview; Liquid-Phase. **Flow Injection Analysis:** Principles. **Fluorescence:** Quantitative Analysis. **Ion Exchange:** Ion Chromatography Instrumentation. **Liquid Chromatography:** Overview. **Ozone.** **Sampling:** Theory. **Sulfur.** **Textiles:** Natural; Synthetic.

## Further Reading

Connell GF (1996) *The Chlorination/Chloramination Handbook*. Denver: American Water Works Association.  
Farr JP, Smith WL, and Steichen DS (1992) Bleaching agents. In: Kroschwitz JI and Howe-Grant M (eds.) *Kirk-Othmer Encyclopedia of Chemical Technology*, 4th edn., vol. 4, pp. 271–300. New York: Wiley.

Gordon G, Cooper WJ, Rice RG, and Pacey GE (1992) *Disinfectant Residual Measurement Methods*, 2nd edn. Denver: American Water Works Association Research Foundation.

Gordon G, Cooper WJ, Rice RG, and Pacey GE (1988) Methods of measuring disinfectant residuals. *Journal of the American Water Works Association* 80(9): 94–108.

Greenberg AE, Clerceri LS, and Eaton AD (1992) *Standard Methods for the Examination of Water and Wastewater*, 18th edn. Washington, DC: American Public Health Association, American Water Works Association, Water Pollution Control Federation.

Jensen JN and Johnson JD (1989) Specificity of the DPD and amperometric titration methods for free available chlorine: A review. *Journal of the American Water Works Association* 81(12): 59–64.

Kubala SW, Tilotta DC, Busch MA, and Busch KW (1989) Determination of chloride and available chlorine in aqueous samples by flame infrared emission. *Analytical Chemistry* 61: 2785–2791.

Madec C, Quantel F, Courtot-Coupez J, and Dore M (1987) Principales methodes analytiques applicables au dosage de traces de chlore, de chlorite et de chlorate lors du traitement des eaux douces par le dioxyde de chlore. *Analisis* 15(2): 69–76.

Masschelein WJ (1982) *Ozonization Manual for Water and Wastewater Treatment*. New York: Wiley.

Palin AT (1986) Current DPD methods for disinfectant residual measurement. *Journal of the Institution of Water Engineers and Scientists* 40: 501–510.

Scounce JS (1962) *Chlorine, Its Manufacture, Properties, and Uses*. New York: Reinhold.

White GC (1986) *The Handbook of Chlorination*, 2nd edn. New York: Van Nostrand Reinhold.

# BLOOD AND PLASMA

**E A Stein and J Pum**, Medical Research Laboratories International, Highland Heights, KY, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by P S Bachorik, pp. 384–388, © 1995, Elsevier Ltd.

## Introduction

Blood consists of a complex mixture of molecular and cellular components, some of which have yet to be identified. Quantitative and qualitative measurements in blood are indispensable for medical practice and biological investigation, but interpretation of the measurements requires knowledge not only of the analytical methods themselves, but of a variety of

other factors, some physiological and some artifactual, that can influence the results.

Measurements are made in blood for a variety of reasons, most commonly for diagnostic purposes and to assess the health status of the patient. The interpretation of such measurements requires the recognition of when the concentration of a particular component is in the abnormal reference range, which requires the knowledge of the normal reference range. This, of course, assumes that the normal concentration range is known. The normal concentration range is perhaps best established by measuring the component of interest in the general (i.e., free-living) population. This establishes the prevailing concentration range for the target population and can be considered as a first step in recognizing



abnormal values. In the absence of any additional information, the normal range is generally and somewhat arbitrarily defined as values that fall between the 5th and 95th percentiles of the prevailing distribution for components that are normally distributed in the population; values that fall outside this range are considered to be sufficiently unlikely that they can be considered abnormal. The prevailing concentration, however, may not always be the 'healthy' concentration. For example, the 95th percentile for serum cholesterol concentration in the USA is  $\sim 300 \text{ mg dl}^{-1}$ , but it is now well recognized that cholesterol concentrations exceeding  $200 \text{ mg dl}^{-1}$  in adults increase the risk for development of coronary artery disease. Thus, it should be recognized that the 'normal reference range' is not necessarily synonymous with 'healthy reference range'.

The second step in recognizing abnormal, or more properly, undesirable, levels of a particular component is to establish the relationship between its concentration and some manifestation of disease or risk for disease. Such relationships can be considerably more difficult to determine than ranges, because it is necessary to establish the relationship between concentration and disease or risk for disease without regard to arbitrarily defined limits, and ultimately, to establish in certain disorders whether the link between concentration and disease may be causal.

In many cases, blood analysis requires the qualitative determination of whether a particular component is present at all, for example, when assessing the presence of an antibody to establish whether the patient has been exposed to a particular pathogen. Finally, measurements in blood are also made in metabolic studies designed to elucidate normal metabolic pathways, test the safety and efficacy of medications, and determine the concentrations of various substances under defined metabolic conditions. Regardless of where or why such measurements are made, the ultimate aim of the laboratory is to provide accurate and precise measurements in individual specimens. In practice, this ideal can only be approximated. The extent to which it can be obtained depends on a number of factors including the nature of the methods employed, the availability of reference materials with which methodological error can be judged, and the circumstances under which samples are collected and handled before analysis (preanalytical factors). These factors can vary widely and can have serious medical, scientific, and economic consequences if their influence is not recognized. This article discusses some of the general principles that apply regardless of what is being measured. Specific details such as the sampling conditions, analytical methods, and quality control

procedures that apply to individual analytes are discussed in the entries dealing with particular measurements.

## Blood Composition

Blood is a complex non-Newtonian fluid that consists of several kinds of cells carried through the circulation in an isotonic aqueous medium, the plasma. Plasma contains a large number of proteins, salts, lipids, and various other nutrients and components. The cellular components normally account for  $\sim 45\%$  of the total blood volume. They include erythrocytes, leukocytes, and platelets. Erythrocytes constitute most of the total cell volume;  $\sim 1\%$  is accounted for by leukocytes, which include several types of cells. The major blood cell types and their concentrations are shown in **Table 1**. About 52% of total blood volume is water and the remaining 3% is contributed by dissolved solids, including protein and nonprotein components. The normal pH of plasma is  $\sim 7.4$ .

On average, the total protein concentration of the plasma, exclusive of the cellular elements, is  $\sim 78 \text{ g l}^{-1}$ . Approximately 58% of the plasma protein is albumin, and another 15% is IgG. The plasma lipoproteins include very-low-density lipoprotein, low-density lipoprotein (LDL), and high-density lipoprotein (HDL), which carry essentially all of the circulating lipids. Collectively they account for  $\sim 8\%$  of the total circulating plasma protein. Most plasma proteins are negatively charged at physiological pH. The major plasma proteins are shown in **Table 2**.

**Table 1** Normal cell composition of blood

Cell type	Mean cell concentration <sup>a</sup> (cells per liter)
Erythrocytes	
Males	$5.2 \times 10^{12}$
Females	$4.6 \times 10^{12}$
Leukocytes <sup>b</sup>	$7.4 \times 10^9$
Neutrophils (total)	$4.4 \times 10^9$
Band neutrophils	$0.2 \times 10^9$
Segmented neutrophils	$4.2 \times 10^9$
Eosinophils	$0.2 \times 10^9$
Basophils	$0.04 \times 10^9$
Lymphocytes	$2.5 \times 10^9$
Monocytes	$0.3 \times 10^9$
Platelets	$310 \times 10^9$

<sup>a</sup>Data from Nelson DA and Morris MW (1991) Basic examination of blood. In: Henry JB (ed.) *Clinical Diagnosis and Management by Laboratory Methods*, 18th edn., pp. 553–603. Philadelphia: WB Saunders.

<sup>b</sup>Leukocytes include a number of cell types broadly characterized as neutrophils, band cells, lymphocytes, monocytes, eosinophils, and basophils.

**Table 2** Major plasma proteins

Protein	Concentration <sup>a</sup> (g l <sup>-1</sup> )
Albumin	45.0
IgG	11.0
$\beta$ -Lipoprotein (low-density lipoprotein)	5.0
$\alpha_1$ -Antitrypsin	3.0
Transferrin	3.0
A <sub>2</sub> -Macroglobulin	2.5
Fibrinogen	2.5
IgA	2.0
Haptoglobin	1.7
IgM	1.1
C3	1.0
Prealbumin	0.3
IgD	0.3
TgE	0.003

<sup>a</sup> Data from McPherson RA (1991) Specific proteins. In: Henry JB (ed.) *Clinical Diagnosis and Management by Laboratory Methods*, 18th edn., pp. 215–228. Philadelphia: WB Saunders.

In addition to proteins, the plasma contains a large number of other anions, the most predominant of which are chloride and hydrogencarbonate. Smaller amounts of phosphate, sulfate, and organic acids are also present. Plasma also contains a number of cations, most predominantly sodium, and lesser quantities of potassium, calcium, and magnesium. The total cation concentration is  $\sim 150 \text{ mmol l}^{-1}$ , and since electrical neutrality must be maintained, the total anion concentration is similar. However, when subtracting the concentration of the major anions ( $\text{Cl}^-$  and  $\text{HCO}_3^-$ ) from the major cations ( $\text{Na}^+$  and  $\text{K}^+$ ), the result is not zero, but usually between 10 and  $20 \text{ mmol l}^{-1}$ . This is known as the ‘anion gap’ and represents the unmeasured anions (e.g.,  $\text{PO}_4^{2-}$ ,  $\text{SO}_4^{2-}$ , and organic acids).

These components represent diverse functions including the transport of nutrients to the tissues and of waste products from the tissues to be reused or excreted, the maintenance of blood pH within physiological limits and the movement of effectors such as hormones from their sites of synthesis to their respective target tissues. Some of these components are actually dissolved in the plasma, and some, such as lipids that are insoluble in an aqueous environment, are transported in the form of lipid–protein complexes, the lipoproteins. In other cases, for example, calcium and hormones, some may exist free in the plasma, and some may be protein bound. In these cases, only the free fraction is biologically active.

## Blood Sampling

When making a measurement in a blood sample obtained from a patient on a given occasion, the

measurement should accurately represent the concentration of the measured component at the time the sample is obtained. The accuracy of the measurement can be affected by artifacts arising from improper blood collection techniques (such as posture, hemoconcentration due to prolonged venous occlusion, and destruction of cells (hemolysis)), the conditions under which serum or plasma is prepared, stored, and transported to the laboratory, and from analytical errors inherent in making the measurement itself. Reliable measurements are more likely to be obtained when the samples have been collected atraumatically and handled properly prior to the analysis. While an unusual value in an individual patient may be commonly ascribed to ‘laboratory error’, in many cases blood samples are drawn in locations remote from the laboratory and under conditions over which the laboratory has no control. The sample may suffer changes such that the measured concentration no longer represents the concentration of the component at the time the patient was sampled. The conditions for blood sampling, preparation, and storage vary according to the measurements required, and it is not possible here to discuss the requirement for individual tests. Nonetheless, certain general principles can be outlined that will apply in most circumstances, although procedures may have to be modified for particular tests.

## Serum and Plasma

Most clinical tests are performed in venous samples of serum or plasma. Serum is prepared by allowing the blood specimen to coagulate. This is best accomplished by collecting the specimen into a glass tube and allowing it to stand at room temperature for at least 45 min. The clot is then sedimented by centrifuging at  $\sim 2500g$  (relative centrifugal force ( $g$ ) =  $1.118 \times 10^{-5} \times r \times n^2$  ( $r$  is the radius in centimeters from center of rotation to the bottom of the tube in the rotor bucket,  $n$  the rotor speed in revolutions per minute (rpm))) for 15–30 min, and the supernatant serum is removed to a clean storage container. Serum separator tubes contain a gel, which facilitates the separation of the serum from the cellular components based on the difference in density. Using such tubes for blood collection obviates the need to transfer the serum to a separate container as the gel forms an effective barrier. The serum contains all of the original plasma components except the cellular elements and protein components that constitute the clot. The concentrations of serum components are taken to represent their concentrations in the circulation.

For some components, it is necessary to use whole blood or plasma. This requires the addition of an anticoagulant to the blood sample as it is being collected or immediately thereafter. Some commonly used anticoagulants include heparin, sodium, potassium, or lithium ethylenediaminetetraacetic acid (EDTA), citrate, and oxalate salts. EDTA, oxalate, and citrate prevent clotting by complexing  $\text{Ca}^{2+}$ , which is required for coagulation; heparin interferes with the formation of fibrin. The particular anticoagulant used varies according to the test. Some anticoagulants, such as citrate and oxalate, can, in the concentrations used to prevent clotting, change the osmolarity of the plasma and induce a shift of water from red cells to the plasma. This dilutes the plasma and can reduce the concentrations of nondiffusible plasma components such as proteins and lipoproteins by 10% or more. Heparin, because it is used in much lower concentrations to prevent coagulation, exerts almost no osmotic effect, whereas EDTA can cause a less severe, but still noticeable (3–5%) dilution. Depending on the component being measured and the anticoagulant used, it may be necessary when interpreting the measurements to account for these effects, particularly for some components that are usually measured either in serum or plasma.

When plasma is to be used, the anticoagulant must be completely mixed with the atraumatically obtained blood sample. This can be accomplished using a blood mixing apparatus, or by gently inverting the collection tube 5–10 times. The cellular components are sedimented by centrifuging at 4°C for 15–30 min at ~2500g. Under these conditions, virtually all of the cellular elements are removed; the few remaining platelets can be removed by centrifuging the plasma again, but in most cases this is not necessary. The plasma is removed and stored in a clean storage container. The containers should be sealed to prevent evaporation and contamination during storage and transport to the laboratory.

Regardless of whether serum or plasma is used, and with some exceptions, the cellular components should be removed as soon as possible, generally within 1 h. In most cases, the serum or plasma should be stored at 2–4°C before analysis. The maximum period of storage varies with the analyte, but in most cases, periods of up to several days are acceptable. For longer periods, samples can generally be stored at –20°C for moderate periods (up to several months), or at temperatures of –70°C to –80°C or lower for longer periods (years). Self-defrosting freezers should not be used because their temperature can fluctuate between ~–2°C and ~–20°C during the daily defrost cycle. Such wide temperature variations can reduce sample stability. In all cases, the

samples should be handled as aseptically as possible to prevent bacterial growth. Samples drawn in one location are frequently shipped to the laboratory at room temperature as a matter of convenience. Depending on where and how the sample is shipped, however, the climate at the time of shipment, and other factors such as ‘room temperature’ can vary considerably and may have deleterious effects on the sample. For some components, room temperature shipment is not acceptable because of compositional changes that may occur due to enzymatic or other alterations. In such circumstances samples are normally shipped at 2–4°C or frozen and shipped on dry ice, which maintains a temperature of ~–40°C.

### Posture

The patient is usually in the sitting position when blood is drawn. In some instances, for example, when the patient is confined to bed, it may be necessary to sample the patient in the supine position. The concentrations of some components vary in response to postural change, however, and if it is necessary to use the supine position, the same position should be used each time that patient is sampled. Posture-related changes in concentration occur in part because of the redistribution of water between the blood and the tissues. When a standing subject sits or reclines, water tends to move from the tissues to the circulation and dilutes large components such as proteins and lipoproteins, or small molecules that are associated with the macromolecules, which are not as readily diffusible as water itself. Other factors not presently understood can also contribute to these changes. The magnitude of postural effects thus varies depending on the component. The changes occur rapidly, are maximal within 20–40 mm, and are reversible upon returning to the standing position. For example, plasma triglycerides, which are associated with lipoproteins, can decrease by ~20% when a standing subject reclines. Plasma cholesterol, also carried on lipoproteins, decreases by ~10% under these conditions. The changes are about half as great when a standing subject sits. On the other hand, the concentration changes for components such as sodium, potassium, creatinine, or glucose are insignificant. Table 3 illustrates posture-related changes in the concentrations of some commonly measured blood components. As can be seen, the changes range from insignificant to ~20%.

### Venous versus Capillary Blood

Venous blood is most commonly used, but capillary samples can also be used, depending on the analyte

**Table 3** Postural changes in blood component concentrations

Component	Percent change
<i>Supine compared to standing</i>	
Cholesterol <sup>a</sup>	-10
Triglycerides <sup>a</sup>	-18
HDL-cholesterol <sup>a</sup>	-8
Protein <sup>b</sup>	-10
Albumin <sup>b</sup>	-10
Alkaline phosphatase <sup>b</sup>	-10
Bilirubin <sup>b</sup>	-15
Calcium <sup>b</sup>	-5
Hematocrit <sup>c</sup>	-10
Sodium <sup>d</sup>	NS
Potassium <sup>d</sup>	NS
Inorganic phosphate <sup>d</sup>	NS
Uric acid <sup>c</sup>	NS
Glucose <sup>c</sup>	NS
Creatine <sup>c</sup>	NS
<i>Sitting compare to standing</i>	
Cholesterol <sup>a</sup>	-5
Triglycerides <sup>a</sup>	-10
HDL-cholesterol <sup>a</sup>	-7

<sup>a</sup>Data from Miller MM, Bachorik PS, and Cloey TC (1992) Normal variation of plasma lipoproteins: postural effects on plasma lipid, lipoprotein and apolipoprotein concentration. *Clinical Chemistry* 38: 569-574.

<sup>b</sup>Data from Dixon M and Paterson CR (1978) Posture and the composition of plasma. *Clinical Chemistry* 24: 824-826.

<sup>c</sup>Data from Hagen RD, Upton SJ, Avakian EV, and Grundy S (1986) Increases in serum lipid and lipoprotein levels with movement from the supine to standing position in adult men and women. *Preventive Medicine* 15: 18-27.

<sup>d</sup>Data from Renoe BW, McDonald JM, and Ladenson JH (1979) Influence of posture on free calcium and related variables. *Clinical Chemistry* 25: 1766-1769.

NS, not significant.

and method to be used for the measurement. Capillary blood is usually obtained by fingerstick, heel-stick (commonly used for infants), or from an earlobe. It more closely resembles arterial than venous blood, but in many instances values obtained from capillary and venous blood are similar. It is important to realize, however, that the concentrations of some analytes will more closely resemble those found in whole blood than in venous serum or plasma (capillary blood glucose concentration for example may be 10-15% lower when measured in a point of care instrument compared to plasma or serum concentration).

Venous samples are usually taken from an antecubital vein. A tourniquet is usually used for this procedure, but prolonged application of the tourniquet can artifactually increase the concentrations of non-diffusible blood components. For lipids, these effects are observed when the tourniquet is applied for longer than ~2 min. Capillary samples are more easily obtained than venous samples, and are

convenient when the analyses must be done rapidly, require only a small amount of specimen, and when methods are available to perform the measurements at the sampling site rather than in the laboratory. Measurements in capillary samples tend to be more variable than in venous samples, however. The reasons for this are not completely understood, but factors such as the possible contamination of blood with tissue fluid during sample collection, the puncture procedure used and variations in the ease with which blood flows from the puncture site, and the characteristics of the analytical methods themselves undoubtedly play a role.

## Normal Physiological Variation

The concentration of a particular component is generally not fixed, but fluctuates around a person's homeostatic set point during the normal course of daily activity. Measurements made in a sample obtained from a given individual on one occasion will usually differ from the values measured in samples obtained on other occasions. These variations occur even though the individual may be in a steady state, i.e., consuming a usual diet, not ill, not gaining or losing weight, or changing medications or dosages, and who is maintaining normal patterns of activity. Such normal variation can occur as a consequence of food intake, postural change, routine daily activity, diurnal variations, and other factors. Since the concentrations of blood components vary somewhat in individuals over time, a measurement made on a single occasion may not reflect the average or usual concentration of a particular component in the individual, even in the absence of analytical error. Physiological variations can usually be reduced to some extent, but they cannot be eliminated entirely. Fasting samples can be used for components whose concentrations may change postprandially. Blood drawing can be standardized to a particular posture, and the patient can be requested to remain in this posture for a specified period before blood is drawn. Even under such conditions, however, there will be concentration differences among serial specimens, and the results in several samples may have to be averaged to determine the patient's usual concentration. This is because the mean of two or more measurements varies less than the individual measurements themselves. The number of serial specimens required depends on a number of factors including the purpose of the measurement (e.g., whether it will be used as a screening tool; for diagnosis or follow-up; for a population survey), the extent to which the concentration is expected to vary normally, the



possible medical consequences or medical decisions that may arise from an unusually high or low value in a particular patient, and the reproducibility of the analytical procedure used for the measurement.

The physiological variation that occurs in a particular individual when in the steady state can be assessed using two or more samples drawn on different occasions. The magnitude of the variations observed in such measurements depends on two factors, normal physiological or biological fluctuations and the reproducibility of the measurement method itself. The latter is referred to as 'analytical variation' and is generally determined by the laboratory through the repetitive measurement of one or more serum pools of known concentration. Such pools are analyzed with every batch of patient samples and are used to monitor the accuracy and reproducibility of the measurements. The mean (and standard deviation, SD) measured value for a particular pool is calculated and the reproducibility of the measurement is usually expressed as a percentage in terms of the coefficient of analytical variation ( $CV_A$ ):

$$CV_A = \frac{SD}{Mean} \times 100$$

Physiological variation is established as follows. The mean (and SD) for measurements in serial samples from the same individual is calculated. The variation can also be expressed in terms of a coefficient of total variation,  $CV_T$ .  $CV_T$  includes the contributions of both physiological and analytical variations. Physiological variation ( $CV_P$ ) is estimated by adjusting the coefficient of total variation for the contribution of analytical variation. If analytical variation is small compared to physiological variation,  $CV_T$  will

approximate  $CV_P$ . The following formula is used to assess whether two serial samples, taken from the same individual, are significantly different:

$$SD (\%) = 2.8 \times \sqrt{CV_A^2 + CV_P^2}$$

If the two results differ by more than the significant difference there is a greater than 95% probability that this difference is not due to the analytical or physiological variation. Knowledge of the analytical and physiological variations also allows the calculation of the number of specimens required to ensure that the mean result is within 5% of the individual's homeostatic set point. This is given by the following

**Table 5**  $CV_T$  for mean total cholesterol concentrations in serial samples from individuals

Number of specimens <sup>a</sup>	$CV_A$ (%)	$CV_T^b$ (%)	
		$CV_P = 5.0\%$	$CV_P = 9.0\%$
1	2.0	5.4	9.2
2		4.1	6.7
3		3.5	5.6
4		3.2	4.9
5		3.0	4.5
1	5.0	7.1	10.3
2		6.1	8.1
3		5.8	7.2
4		5.6	6.7
5		5.5	6.2

<sup>a</sup>Specimens should be taken at least 12 weeks apart.

<sup>b</sup> $CV_P$  data taken from Kafonek SD, Derby CA, and Bachorik PS (1992) Biological variability of lipoproteins and apolipoproteins in patients referred to a lipid clinic. *Clinical Chemistry* 38: 864–872.

**Table 4** Coefficients of physiological variation for several plasma constituents

Component	Average (%)	50th percentile (%)	75th percentile (%)	95th percentile (%)
Cholesterol <sup>a,c</sup>		5	9	14
Triglyceride <sup>a,c</sup>		18	26	44
HDL-cholesterol <sup>a,c</sup>		7	14	25
LDL-cholesterol <sup>a,c</sup>		8	12	20
Uric acid <sup>b</sup>	10			17
Urea nitrogen <sup>b</sup>	12			19
Lactate dehydrogenase <sup>b</sup>	9			22
Phosphate <sup>b</sup>	8			13
Albumin <sup>b</sup>	4			7
Total protein <sup>b</sup>	3			5
Potassium <sup>b</sup>	5			9
Glucose <sup>b</sup>	6			9

<sup>a</sup>Data from Kafonek SD, Derby CA, and Bachorik PS (1992) Biological variability of lipoproteins and apolipoproteins in patients referred to a lipid clinic. *Clinical Chemistry* 38: 864–872.

<sup>b</sup>Data from Harris EK, Kanofsky P, Shakarji G, and Cottlove E (1970) Biological and analytical variation in long-term studies of serum constituents in normal subjects. II. Estimating biological components of variation. *Clinical Chemistry* 16: 1022–1027.

<sup>c</sup>Bachorik *et al.*

formula:

$$n = z^2 \times \left( \frac{CV_A^2 + CV_P^2}{d^2} \right)$$

where  $n$  is the number of specimens,  $z$  is the number of standard deviations required for a stated probability under the normal curve (for  $p < 0.05$ ,  $z = 1.96$ ), and  $d$  is the desired percentage closeness to the homeostatic setpoint.

The magnitude of intraindividual variation differs in different individuals. An example of this is provided in Table 4, which lists the average and selected percentiles for  $CV_P$  for several plasma components. As can be seen, the physiological fluctuations for the components shown vary from 3% to ~20% on average, but can be considerably higher in some individuals.

The  $CV_T$  observed for the mean level in one or more serial specimens from the same individual depends on the physiological variation in that individual, the number of serial specimens collected, and the  $CV_A$  for the measurement method itself. This is illustrated in Table 5 for total cholesterol. The table shows the expected  $CV_T$  for the mean of up to five serial specimens for individuals with  $CV_P$ s at the 50th and 75th percentiles when the measurements are performed with  $CV_A$ s of 2.0% or 5.0%. The

lower value reflects reasonably well the  $CV_A$ s that would occur in a well-controlled clinical laboratory using modern automated equipment. The higher  $CV_A$  is in the range observed with many physicians' desktop analyzers. For an individual with a median  $CV_P$  (5%) whose samples were analyzed with a  $CV_A$  of 2%, the coefficient of total variation would be below 5% if using two serial specimens. To achieve a  $CV_T$  below 5% would require at least four serial specimens if the subject had a  $CV_P$  at the 75th percentile (9%). Even five serial samples would not be sufficient to achieve a  $CV_T$  of less than 5% if the  $CV_A$  was 5%. Because of the interplay of these factors, it is necessary.

*See also:* **Clinical Analysis:** Sample Handling. **Lipids:** Determination in Biological Fluids. **Proteins:** Physiological Samples.

### Further Reading

Devilin TM (ed.) (1992) *Textbook of Biochemistry with Clinical Correlations*. New York: Wiley-Liss.

The Working Group on Lipoprotein Measurement (Bachorik PS, Myers GL, Ross JW, Stein EA, and Warnick GR) (1995) *Recommendations on Lipoprotein Measurement*. National Institutes of Health – National Heart, Lung, and Blood Institute. NIH Publication No. 95-3044. September.

## BOD

*See* **WATER ANALYSIS: Biochemical Oxygen Demand**

## BUFFER SOLUTIONS

**A Hulanicki and S Głab**, University of Warsaw, Warsaw, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

A buffer solution, or simply a buffer, is a chemical system added to keep constant, or at least minimize, the variation of a particular property. Initially, buffer solutions were used to stabilize the pH of the reaction medium. Subsequently, metal buffers were introduced to keep the free metal ion concentration constant. Redox buffers are used to stabilize the

redox potential. The term buffer is also used in other situations encountered in analytical chemistry. In this article, the simple theory of the buffering mechanism is given along with some examples of different buffers used in the laboratory practice.

### Various Concepts of Buffers

A buffer solution is a system that has the property of being able to eliminate or diminish the influence of external conditions on a chemical system. This term is most commonly used for acid–base systems, and they are named pH buffers. These buffers are added to solutions to prevent a change in their pH value

occurring on addition of acidic or basic solutions, when the solution is diluted with a solvent, or when in a reaction an undesirable pH is expected to occur. The discussion of such buffers will occupy the main part of this article.

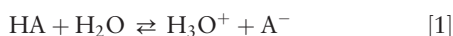
The term buffer is also occasionally applied to other systems. Metal buffers are used when it is necessary to maintain free metal concentration at a constant level. Redox buffers are designed to keep the redox potential of a solution constant. They contain a mixture of oxidized and reduced substances. In measurements by ion-selective electrodes, the term ionic strength buffer is used to denote a solution that is added to keep the ionic strength, and in consequence the activity coefficients, of a given solution constant. Such systems may be called buffers only when they are able to keep the buffered parameter constant and when in the absence of the buffer the given parameter alters significantly. For example, the total ionic strength adjustment buffer, used in measurements with ion-selective electrodes, exhibits a buffering capacity only because it has a relatively high content of the ionic-strength-adjusting electrolyte. However, at the same time they may play another role in the measurements, e.g., complexing some interferents.

In another field, the term buffer (but not buffer solution) is used for substances that are added to the samples introduced to the excitation source in atomic spectroscopy to prevent any change in the excitation conditions due to accompanying elements. Mainly this refers to maintaining constant temperature or electron density.

## Theory of pH Buffers

A pH buffer owes its buffering action to the fact that after the addition of hydrogen ions or hydroxyl ions the position of the equilibrium of weakly dissociated electrolytes (acids or bases) is shifted in such a way that the added ions are consumed. This is mainly applicable to aqueous solutions, but the same concept is valid for any amphoteric solvent.

The weak acid, HA, reacts in solution according to eqn [1]:



This equilibrium is characterized quantitatively by the acid dissociation constant given in eqn [2]:

$$K_a = [\text{H}_3\text{O}^+][\text{A}^-]/[\text{HA}] \quad [2]$$

Here, the expressions in square brackets denote molar concentrations.  $K_a$  is a concentration constant,

but more correct would be to express it in terms of activity. In this case, the constant is termed a thermodynamic constant. Depending upon which convention is used, the appropriate constant must be used.

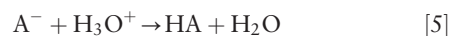
From eqn [2], the hydrogen ion concentration can be calculated as shown in eqn [3]:

$$\begin{aligned} [\text{H}_3\text{O}^+] &= K_a[\text{HA}]/[\text{A}^-] \quad \text{or} \\ \text{pH} &= \log([\text{A}^-]/[\text{HA}]) + \text{p}K_a \end{aligned} \quad [3]$$

This equation shows that in order to maintain the pH value constant the logarithmic term containing the concentration ratio should be allowed to vary as little as possible. When the solution contains a mixture of the acid, HA (Brønsted acid), and the weak base,  $\text{A}^-$  (Brønsted base), the addition of a base results in further dissociation of the acid according to eqn [4]:



The concentration of free hydrogen ions therefore decreases. In the reverse situation, when hydrogen ions are added or produced they combine with the base according to eqn [5]:



When the buffer solution is diluted the ratio of concentrations does not change at all and the pH value, in principle, should remain constant.

The buffering property of a solution is preserved as long as the concentrations of the components in the buffer are greater than the amounts of added hydroxyl ions (base) or hydrogen ions (acid). If this is not the case then the buffer solution does not function according to expectations. It is then said that the buffer capacity is too small.

The buffer capacity depends on the absolute concentrations of the two components: buffer capacity toward addition of acid depends on the concentration of the base, while buffer capacity toward addition of base depends on the concentration of the acid. The definition of the buffer capacity can be expressed as

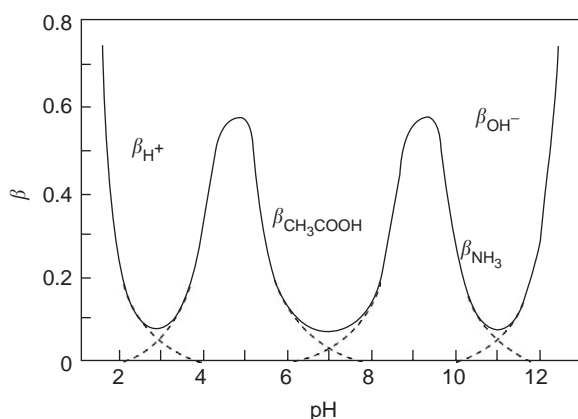
$$\beta = dC/d\text{pH} \quad [6]$$

Its exact value can be calculated from a much more complicated expression that takes into account the presence of several buffering systems:

$$\beta = \left( \frac{K_w}{[\text{H}_3\text{O}^+]} + [\text{H}_3\text{O}^+] + \sum \frac{C_i K_a [\text{H}_3\text{O}^+]}{(K_a + [\text{H}_3\text{O}^+])^2} \right) \quad [7]$$

In this equation,  $K_a$  refers to the acid dissociation constant (reciprocal of the protonation constant) of





**Figure 1** The buffer capacity of a mixture of two systems,  $0.10 \text{ mol l}^{-1}$  acetic acid and  $0.10 \text{ mol l}^{-1}$  ammonia, as a function of pH.

the buffer component with total concentration,  $C_i$ , and  $K_w$  the autoprotolysis constant of water. The  $\Sigma$  symbol indicates that the equation can take into account several buffering systems present in the solution. The first term in the brackets is equivalent to the concentration of the base ( $\text{OH}^-$ ), which in combination with water present as solvent also forms a buffer system. Therefore, it is often said that the solution of a strong base has buffering properties. The same can be said about the second term, which refers to the acid ( $\text{H}_3\text{O}^+$ ) and also buffers in the acidic range (Figure 1).

## Preparation of Buffer Solutions

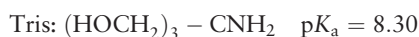
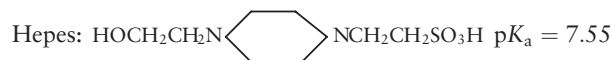
In principle, any acid–base system can be used to prepare a buffer solution. In practice, several conditions must be fulfilled. First, the logarithm of the protonation constant should not differ by more than one unit from the required pH value of the buffer. This will provide sufficient buffer capacity for the buffer solution. The other conditions are that the components in the buffer must be available at high purity grade or can be easily purified; they must be stable in time in both pure state and in solution; and they should not undergo reactions such as oxidation by air oxygen. In a number of buffer solutions, in particular when they contain organic species, mold growth may occur in the pH range 3–11. This obviously disqualifies the use of the buffer. To prevent this, a preservative (e.g., thymol) may be added.

A standard set of such acid–base systems has been recommended for establishing the pH scale, i.e., for calibrating pH meters. The pH value of standard buffer solution corresponds to the activity of the hydrogen ion. The pH value varies with temperature as

the activity coefficients and the dissociation constants are temperature dependent. Generally, the  $\text{dpH/dT}$  coefficient is negative, which means that with increasing temperature the pH value decreases. Among the standard buffers the change is largest for the carbonate buffer and corresponds to  $0.0096 \text{ pH unit per degree Celsius}$ .

When pH buffers are used in analytical procedures to provide approximately constant pH value, they are prepared by mixing the acid with the base in given proportions, by adding a strong base (e.g., NaOH, KOH) to the solution of a weak acid until the desired pH value is attained, or by adding a strong acid (e.g., HCl,  $\text{HNO}_3$ ) to a weak base solution. Acetate, phosphate, or ammonia buffers are commonly used in analytical practice. It is important to check that the buffer components will not react in an uncontrolled way with the substrates or products of the reaction being studied or applied analytically. In this respect, phosphate buffers are less convenient as they may react with many of the metal ions studied.

The buffers used in biochemical studies should have a pH value not far from neutral and should not complex or form precipitates with calcium. Among them are the so-called ‘Good buffers’ that contain zwitterionic amino acids such as *N*-substituted taurines or glycines. Examples of such buffers are Hepes (4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid), Tris (tris(hydroxymethyl) aminomethane), and Tricine (*N*-[Tris(hydroxymethyl)methyl]glycine):



For special applications additional requirements are imposed on buffer solutions. For spectrophotometry, the buffer used should not absorb radiation in the region where the species that are being studied exhibit absorption. Buffer solutions find important application in electrophoretic separations of biomolecules in the electric field, which, depending on the pH value of the medium, acquire a positive or negative net charge. The components of the buffer have similar properties as the buffers used in biochemical investigation, but, in addition, have a polymerizable part, may be immobilized on a polymer matrix, forming a stable pH gradient. Buffers used in voltammetric studies should not undergo, in the used potential range, electrochemical reduction or oxidation. The buffers used in extraction, spectrophotometry, or voltammetry often possess complexing properties that may enhance the selectivity of

determination. For biochemical and medical applications, the buffers used should be isotonic with respect to the fluids under investigation. Such buffers usually contain a mixture of mono- and disodium hydrogenphosphate, the ionic strength being adjusted with sodium chloride.

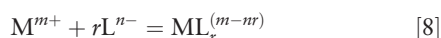
For routine work commercially available mixtures are useful. These contain several components and the addition of specified amounts of a strong base is all that is required for the preparation of the buffer solution. Among these is Britton–Robinson buffer (Table 1), which contains acetic, phosphoric, and boric acids. Universal buffers for spectrophotometry may contain such components as chloroacetic, formic, acetic, phosphoric, succinic, citric, boric acids, tris(hydroxymethyl)aminomethane, and butylamine. These buffers are transparent at wavelengths at least down to 240 nm.

Many naturally occurring systems are able to maintain a constant pH value by making use of buffer systems. In blood the hydrogencarbonate and protein systems maintain a pH  $\sim 7.4$ . In open ocean water the pH is kept within the range 7.9–8.3 by a multicomponent buffer that includes aluminosilicates and carbonates.

The composition and pH values of a range of commonly used pH buffers, including Good buffers, is given in the appendix section of the encyclopedia.

## Metal Buffers

Metal buffers are solutions that contain a given concentration of free metal ions, usually very small, which is kept constant by the addition of a suitable ligand. The basic theory is similar to that for the pH buffers. If a metal ion, M, reacts with a ligand, L, according to the equation:



for which the equilibrium (stability) constant can be written as

$$\beta_r = [ML_r^{(m-nr)}] / [M^{m+}][L^{n-}]^r \quad [9]$$

then

$$pM = \log([L^{n-}] / [ML_r^{(m-nr)}]) + \log \beta_r \quad [10]$$

where  $m$  and  $n$  are charge numbers for M and L, respectively, and  $r$  is the stoichiometric coefficient for the M–L complex.

If the ligand is present in excess and the complexes are relatively strong, then the ratio  $[ML_r^{(m-nr)}] / [M^{m+}]$  is practically constant and pM, being the negative logarithm of metal ion concentration, is maintained at a constant level. These considerations

**Table 1** Composition of Britton Robinson buffers. To 100 ml of a solution containing  $0.04 \text{ mol l}^{-1}$  acetic acid,  $0.04 \text{ mol l}^{-1}$  phosphoric acid, and  $0.04 \text{ mol l}^{-1}$  boric acid, X ml of  $0.2 \text{ mol l}^{-1}$  sodium hydroxide solution is added to obtain the required pH at  $18^\circ\text{C}$

X	pH	X	pH	X	pH
0.0	1.81	35.0	5.02	70.0	9.15
2.5	1.89	37.5	5.33	72.5	9.37
5.0	1.98	40.0	5.72	75.0	9.62
7.5	2.09	42.5	6.09	77.5	9.91
10.0	2.21	45.0	6.37	80.0	10.38
12.5	2.36	47.5	6.59	82.5	10.88
15.0	2.56	50.0	6.80	85.0	11.20
17.5	2.87	52.5	7.00	87.5	11.40
20.0	3.29	55.0	7.24	90.0	11.58
22.5	3.78	57.5	7.54	92.5	11.70
25.0	4.10	60.0	7.96	95.0	11.82
27.5	4.35	62.5	8.36	97.5	11.92
30.0	4.56	65.0	8.69	100.0	11.98
32.5	4.78	67.5	8.95		

are true in the exceptional case where the ligand has no protolytic properties. Usually, the system shows strong pH dependence and an additional pH buffer is needed. When polyprotic species are used as ligands they may simultaneously act as pH buffers. The ligands used for metal ion buffers are strong chelating agents such as polyaminopolycarboxylic acids (e.g., nitrilotriacetic acid; ethylenediaminetetraacetic acid (EDTA)) and their salts, macrocyclic polyamines (e.g., 1,4,7,10-tetraazacyclododecane), or aliphatic polyamines (e.g., ethylene diamine). As an example, a solution containing  $1 \times 10^{-3} \text{ mol l}^{-1} \text{ Ca}^{2+}$  and  $5 \times 10^{-3} \text{ mol l}^{-1}$  nitrilotriacetic acid at pH 8.04 in the presence of  $0.1 \text{ mol l}^{-1} \text{ NaCl}$  acts as a calcium buffer (pCa buffer) for constant ionic strength. Such a buffer keeps the calcium ion concentration at the chosen pCa level in the range from four to seven.

There are two types of one-phase metal buffers: one containing the metal ion of interest and a complexing ligand, and the other containing, in addition, another metal ion, which is in excess compared to the two other components but is less strongly complexed by the ligand. The latter type of system is less influenced by pH changes of the solution. This is because the pH changes affect the stability of the second complex in the same way as that of the first metal ion. In the system containing  $10^{-3} \text{ mol l}^{-1} \text{ Pb}^{2+}$ ,  $0.1 \text{ mol l}^{-1} \text{ Mg}^{2+}$ , and  $10^{-2} \text{ mol l}^{-1}$  EDTA the value of pPb equals 11.25, 11.29, and 11.29 at pH 5.0, 7.0, and 9.0, respectively. However, dilution affects the pM values of these metal buffers.

Metal buffers are used for the calibration of ion-selective electrodes at low ion levels (below  $10^{-5} \text{ mol l}^{-1}$  where the preparation of standard

solutions by simple dilution may give rise to errors due to losses (adsorption, side reactions) or contamination. Another application is to provide a medium of fixed ion concentration for selective precipitations. An important application of metal buffers is for the investigation of biochemical reactions influenced by the ion concentration of a particular metal (e.g., calcium), as in the study of calcium-sensitive enzymes such as ATP phosphohydrolase.

**See also:** **Electrophoresis:** Isoelectric Focusing. **Ion-Selective Electrodes:** Glass. **pH. Quality Assurance:** Instrument Calibration.

### Further Reading

- Bates RG (1973) *Determination of pH, Theory and Practice*. New York: Wiley.
- Craggs A, Moody GJ, and Thomas JDR (1979) Calcium ion-selective electrode measurements in the presence of complexing ligands. *The Analyst* 104: 961–972.

- Good NE, Winget GD, Winter W, *et al.* (1966) Hydrogen ion buffers for biological research. *Biochemistry* 5: 467–471.
- Hulanicki A (1987) *Reactions of Acids and Bases in Analytical Chemistry*. Chichester: Horwood.
- Hulanicki A, Ingman F, and Wanninen E (1991) Metal buffers in chemical analysis. Part II. Practical considerations. *Pure and Applied Chemistry* 63: 639–642.
- IUPAC (1998) In: Inczedy J, Lengyel T, and Ure AM (eds.) *Compendium of Analytical Nomenclature*. Oxford: Blackwell Science.
- Nicholson E and Duff EJ (1981) Fluoride determination in water – an optimum buffer system for use with fluoride-selective electrode. *Analytical Letters* 14: 887–912.
- Perrin DD and Boyd D (1981) *Buffers for pH and Metal Ion Control*. London: Chapman and Hall.
- Scharff O (1979) Comparison between measured and calculated concentrations of calcium ions in buffers. *Analytica Chimica Acta* 109: 291–305.
- Wanninen EV and Ingman F (1991) Metal buffers. Part I. Theoretical considerations. *Pure and Applied Chemistry* 59: 1681–1692.

## BUILDING MATERIALS

**L W Cheriton and J P Gupta**, Fosroc International Ltd., Birmingham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 407–417, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

An extremely wide range of materials can be used for building. These can be either natural or man-made. Natural materials include aggregates, bitumen, clays, rubber, stone, and wood; man-made materials include brick, inorganic cements, glass, plaster, metals and their alloys, synthetic polymers, and wood preservatives.

The purpose of analyzing any of these materials can vary from product failure, through quality control to product development. From an analytical point of view the materials are best classified as organic or inorganic. The inorganic group can be further subdivided into metallic and nonmetallic.

Modern instrumental analytical techniques are considered where appropriate, as they usually provide the most cost-effective, efficient, and most accurate results. Frequently British Standard (BS) and American Society for Testing and Materials (ASTM) methods are quoted where appropriate. Even though these methods can be quite similar, small differences in results can arise. It therefore

follows that if testing is required to a certain standard, then the actual standard must be used. At the time of writing, work is underway to incorporate many British Standards into European Standards (ENS). These have not been referred to at this time.

### Organic Materials

Two important organic materials used in the building industry are wood- and oil-based products.

#### Wood

Wood consists of 40–50% cellulose, 20–30% hemicelluloses, 21–23% lignin, 0.2–0.5% protein, and 1–5% mineral matter, which is left behind as ash when the wood is burned. Besides its straightforward use as timber, much wood is pulped for use in paper-making. Lignin is extracted in the process by either sulfonation or chlorination. Sulfonation produces lignosulfonates, which are widely used as dispersing agents, for example as chemical admixtures for concrete to disperse the cement particles (see the section on Cement and Concrete Admixtures, below).

Timber itself is usually treated with chemical preservatives (fungicides/biocides) to prevent attack by fungi or insects. Such materials include organotin salts, tri-*t*-butyltin (TBT), TBT oxide (TBTO), TBT

phosphate (TBTP), TBT naphthanate (TBTN); as well as 2-mercaptobenzothiazole (MBT), copper(II) oleate, organozinc and copper salts, pentachlorophenol (PCP), creosote, and phenyl mercury(II) acetate. These will require analysis, for example, in a survey to ascertain whether the timber used in a structure has been treated.

**Analysis** The chemical analysis of wood is carried out by ASTM (American Society for Testing and Materials) and TAPPI (Technical Association of the Pulp and Paper Industry) standards. These methods can be summarized as in Table 1 following milling to  $-40$  mesh ( $400\ \mu\text{m}$ ) and air drying.

Modern analytical such as infrared (IR) spectroscopy, liquid chromatography (LC), and gas-liquid chromatography (GLC) are in use for the identification of organic components, whilst the mineral constituents can be estimated by using X-ray fluorescence (XRF) spectrometry, X-ray diffraction (XRD), atomic absorption spectrometry (AAS), or inductively coupled plasma (ICP) atomic emission spectrometry (AES).

The above methods are also employed for the analysis of wood preservatives. The metallic component is determined by AAS or ICP-AES after oxidation of the organic matter in a mixture of concentrated sulfuric and perchloric acids. Nuclear magnetic resonance (NMR) spectroscopic techniques have also been mentioned in the literature for the identification of components from a complex formulated wood preservative.

## Oil

Oil is the principal raw material for the plastics industry and is used to produce a vast range of products used in the building industry.

**Polymers** Polymers in long chains are known as thermoplastic, e.g., polypropylene, which can be used for hot-water pipes.

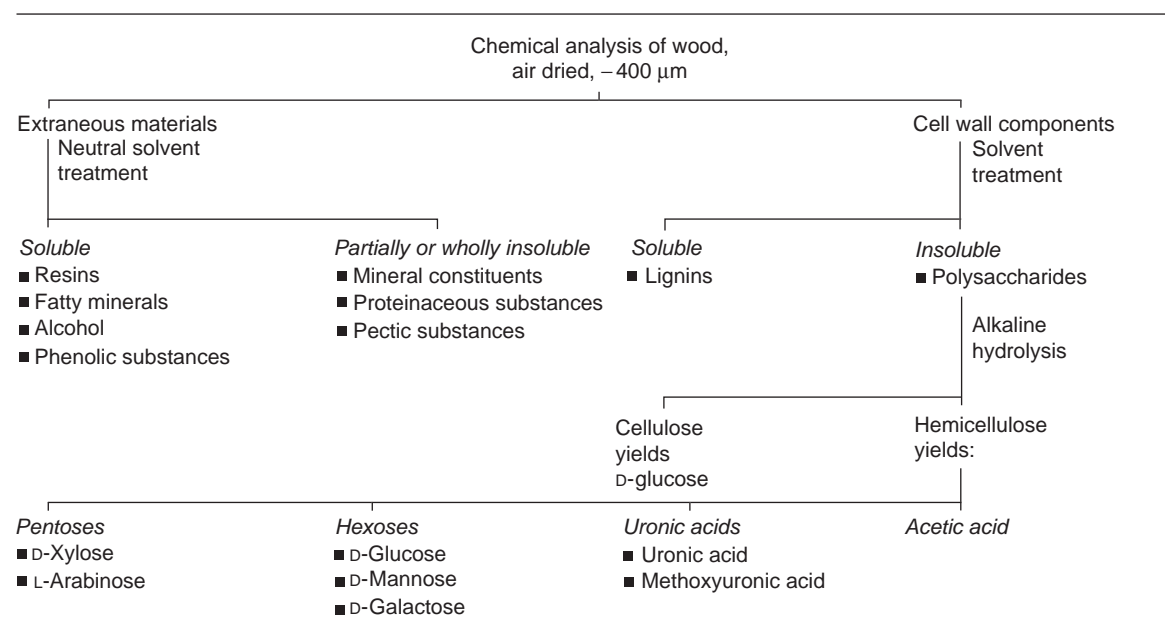
Cross-linked polymers are known as thermosetting. Polyesters have wide applications in building, e.g., in glass-reinforced plastic (GRP) and chemical anchoring.

Polymers in the form of spongy chains are known as elastomers. Polyurethane is an example of an elastomer. Elastomers are used to fill moving joints in construction, e.g., between concrete slabs on a road or runway.

Polymers are also used to form compounds that, when added to building materials, modify their properties. For example dibutylphthalate (DBP) is often used as a plasticizer to improve the flexibility of a plastic formulation.

Melamine formaldehyde sulfonate and naphthalene sulfonates are added to concrete to 'plasticize' the mix. A 'plasticizer' in concrete technology is a cement-dispersing agent. Other polymers can be added to cement mortars/concretes to improve flexibility and reduce permeability to water, e.g., styrene butadiene rubber (SBR) emulsions; to give air entrainment for freeze-thaw resistance, e.g., alkyl sulfonates; or to thixotropize a mix, e.g., polyacrylamide.

**Table 1** Analytical methods for wood



Polymers are also used extensively in the paint industry. The trend is away from solvent-based systems to the more environmentally friendly water-based emulsions.

It is important to be able to determine the small amounts of polymers present in concrete, for example, in the event of failure. This will be discussed further later.

**Analysis** The International Standards Organization (ISO) Technical Committee 61 on plastics has developed and promulgated more than 100 standards (ISO, Geneva, Switzerland), many of which describe test methods for the analysis and evaluation of plastic materials.

The development of instruments for the analysis and characterization of plastics has greatly facilitated the study of their composition, structure, molecular parameters, and performance. **Table 2** summarizes these methods and parameters for which they are suitable.

Paint is also analyzed using the methods given in **Table 2** as it is a mixture of polymers, pigments, extenders, carrier solvents, and additives. More appropriately, its testing almost always involves determining performance properties rather than chemical composition. For example, hiding power is not calculated from refractive index and particle size measurements, but by measuring and comparing the reflectance of the paint over black and white substrates.

**Table 2** Polymer characterization

Method	Determination
Chromatography	
GPC	Molecular mass/distribution, composition
LC	of additives such as plasticizers,
GC	antioxidants, stabilizers
Light scattering	Molecular mass, shape and size
Ultracentrifugation	Molecular mass/distribution
Thermal analysis	
DTA/DSC	$T_g$ , $T_m$ , degree of crystallinity and purity
TG	Weight loss, degradation mechanism,
	reaction kinetics, activation energies,
	thermal stability, and organic filler
	composition
TMA/DMA	Mechanical properties
NMR ( $^1\text{H}$ and $^{13}\text{C}$ )	Structure, configuration, molecular
	dynamics
MS	Polymer composition and structure,
	degradation products
IR	Structure, composition
X-ray spectroscopy	
XRF	Filler composition
XRD	Filler identification, degree of crystallinity

GPC = gel permeation chromatography; MS = mass spectroscopy;  $T_g$  = glass transition temperature;  $T_m$  = melting point; TMA = thermomechanical analysis.

**Bitumens** Another important class of oil-based products is bitumens. Bitumen is a generic term defined by ASTM as a class of black or dark-colored (solid, semisolid, or viscous) cementitious substances, natural or manufactured. Bitumens are composed principally of high-molecular-mass hydrocarbons of which asphalt, tars, pitches, and asphaltite are typical. Commercially these are produced from the destructive distillation of coal, crude oils, and other organic matter. Asphalt occurs naturally either in rock or a lake. In the United States the terms bitumen and asphalt are interchangeable, whereas in the United Kingdom 'asphalt' is reserved for the naturally occurring product and bitumen is the residue from crude oil distillation. It is important to note that the compatibility of various classes of bitumen with other raw materials can vary widely. This can lead, for example, to marked differences in chemical and solvent resistance of the end product.

The typical composition of bitumen/asphalt used in roadbuilding, etc., is

Bitumen content	4–20%
Softening point	28–38°C
Penetration of 100 g at 25°C over 5 s (mm/10)	Too soft–360
Remainder	Graded mineral fillers (sand, sandstone or limestone)

**Analysis** *Elemental*: C, S, H, O, N (typically 79–88% C; 7–13% H; up to 8% S; 2–8% O; up to 3% N). *Trace metals*: Fe, Ni, V, Ca, Ti, Mg, Na, Co, Cu, Sn, Zn. *Molecular mass*: typically  $M_r = 500$ –2500. *Acid number*: typically 0.1–2.8 mg KOH per g. *Distillation range*: ASTM D3279. *Composition*: bitumen insoluble in paraffin naphtha (AASHTO T46 or ASTM D3279); bitumen soluble in carbon disulfide (ASTM D4). *Purity*: solubility, ash, water content (ASTM D95). *Softening point*: ASTM D36. *Flash point*: ASTM D92.

## Nonmetallic Inorganic Materials

Some important nonmetallic inorganic materials used in construction are cements; pozzolanic materials and slags; clays; calcium sulfates; and rocks, aggregates, sand, and glass.

### Cements

The Pocket Oxford Dictionary gives the first meaning of cement as 'Substance made by calcining lime and clay, applied as a paste and hardening into stony consistence ...'. This describes the principal cement



produced world-wide in hundreds of millions of tons, which is more precisely termed ordinary Portland cement.

Analysis is required right from the initial geological survey of a proposed raw material quarry through to the finished product. Ordinary Portland cement is essentially made from 4 parts of a calcareous material such as limestones and 1 part of an argillaceous material (e.g., clay or shale) and fired in a rotary kiln at  $\sim 1400^\circ\text{C}$ . The resulting clinker is rapidly cooled and then ground with  $\sim 5\%$  of gypsum to a specific surface area (measured by air permeability) of  $\sim 350\text{ m}^2\text{ kg}^{-1}$ . Gypsum is added to prevent a rapid ‘flash’ set when the cement is mixed with water.

The end-product consists of  $\sim 50\%$  tricalcium silicate (the main contributor to compressive strength development up to 28 days),  $25\%$  dicalcium silicate (contributes to long-term strength),  $10\%$  tricalcium aluminate (involved in setting, little contribution to strength),  $10\%$  tetracalcium aluminoferrite (little contribution to strength), and  $5\%$  gypsum. The relative proportions of the various phases principally govern cement performance.

When Portland cement is added to water, reactions take place to form calcium silicate hydrates and release calcium hydroxide. These hydrates act as the glue in mortars and concretes to hold the matrix together.

Chemicals can be added to the system to modify the rate of cement hydration; for example, sugars retard it and salts such as calcium chloride accelerate it (see Cement and Concrete Admixtures, below).

Of the non-Portland cements used in building, the most important is high-alumina cement (Ciment Fondu).

### High-Alumina Cement

High-alumina cement (HAC) is made from a mixture of limestone and bauxite melted together at  $\sim 1600^\circ\text{C}$  to produce a dark clinker. The clinker is ground to a similar specific surface area to Portland cement of  $\sim 350\text{ m}^2\text{ kg}^{-1}$ . The principal clinker phases are  $\sim 40\text{--}50\%$  calcium aluminate ( $\text{CaO} \cdot \text{Al}_2\text{O}_3$ ),  $5\text{--}10\%$  mayenite,  $[(\text{CaO})_{12}(\text{Al}_2\text{O}_3)_7]$ ,  $20\text{--}40\%$  brownmillerite  $[(\text{CaO})_4\text{Al}_2\text{O}_3 \cdot \text{Fe}_2\text{O}_3]$ , and  $5\text{--}10\%$  gehlenite  $[(\text{CaO})_2\text{Al}_2\text{O}_3 \cdot \text{SiO}_2]$ . Calcium aluminate is the principal strength-giving constituent, whereas mayenite reacts very rapidly and primarily governs the setting time. Gehlenite is undesirable and reacts only slowly with water. A small increase in the silica content of the raw materials can lead to a large increase in gehlenite content and silica must therefore be carefully controlled.

When HAC is added to water, calcium aluminate hydrates are formed very rapidly that can give Portland cement type 28-day strengths in one day. Unfortunately, in damp, warm, conditions these hydrates can ‘convert’ to a denser form, increasing the porosity of the matrix and thereby reducing strength. However, provided water:cement ratios of 0.4 or less are used, HAC-based concretes and mortars will be sufficiently strong even after conversion has taken place. The degree of conversion can be ascertained by differential thermal analysis/differential scanning calorimetry (DTA/DSC) or XRD.

In a manner similar to Portland cement, the hydration of HAC can be accelerated or retarded by the use of suitable admixtures. The most common accelerators are lithium salts, which can make HAC mortars set in a few minutes and the temperature rise to the boiling point of water. Hydroxycarboxylic acids are the most commonly used retarders.

HAC is used where rapid strength gain is required, in refractory applications and, with calcium sulfate, to form ettringite-based products either for shrinkage compensation or for high water:cement ratio grouts (e.g., 5:1). It is important to control carefully the relative proportions of HAC and calcium sulfate in ettringite-based systems to ensure optimum properties.

### Calcium Sulfoaluminate Cement

A promising ‘new’ cement that was discovered in the 1950s by Klein is based on calcium sulfoaluminate (CSA). It is made in a rotary kiln in a manner similar to Portland cement, but the raw materials used are bauxite, limestone, and calcium sulfate.

CSA cements can be ground with calcium sulfate and used in construction, where they are reported to be more durable than Portland cement, particularly in aggressive sulfate or chloride environments. With more calcium sulfate they can be used to produce ettringite-based products in a manner similar to HAC.

Like HAC, CSA-based cements react much faster with water than Portland cement and are therefore very useful for cold-weather working, for example down to an ambient temperature of  $-15^\circ\text{C}$ . However, CSA hydrates are not vulnerable to the HAC conversion problems.

### Chemical Cements

Even faster setting and strength gain can be achieved by direct acid–base reactions. For example, magnesium chloride and magnesium oxide (Sorel cement) react very rapidly in water to form

$\text{MgO} \cdot \text{MgCl}_2 \cdot 11\text{H}_2\text{O}$  and  $\text{Mg}(\text{OCl})_2$ . However, Sorel cement is attacked by water and it attacks steel.

The reaction of phosphoric acid (usually in the form of ammonium phosphate) with magnesium oxide in the presence of water gives magnesium phosphate cements. These are used for rapid repairs, for example of pot holes in busy roads.

### Pozzolans

Both natural and artificial pozzolans can be used to make cements by reaction with lime in the presence of water.

Natural pozzolans are usually volcanic in origin and the most widely used artificial pozzolan is the pulverized flyash (PFA) obtained from coal-fired power stations. Their composition varies widely but they are all silico-aluminate based.

They can be used alone with lime for low-cost bulk filling operations such as backfilling in old mine workings or to extend and improve the properties of Portland cement. Their use reduces permeability and heat of hydration, leading to improved durability. The downside is that strength development is slower. Analysis is important to discover the likely reactivity of the pozzolan and to determine whether it contains anything likely to be deleterious to its intended use.

Another material used to extend and improve cement systems is granulated ground blast-furnace slag (GGBFS). Approximately one ton of slag is produced per ton of iron and provided the quality is right it can be sold for about 2/3 of the price of Portland cement. The major components are 40–50%  $\text{CaO}$ , 35%  $\text{SiO}_2$ , 10–15%  $\text{Al}_2\text{O}_3$ , and 7–12%  $\text{MgO}$ .

For optimum reactivity with Portland cement it is important that the slag has the right chemical composition and a high glass content (>85%). The hydraulic index (eqn [1]) should be >1.0, preferably >1.5, where

$$\text{Hydraulic index} = \frac{\% \text{CaO} + \% \text{MgO} + \% \text{Al}_2\text{O}_3}{\% \text{SiO}_2} \quad [1]$$

The major oxides are determined by X-ray fluorescence (XRF) spectrometry and the glass content by XRD (BS 6699, 1992).

GGBFS is often used at up to 70% of a cement system, where its low heat of hydration, low permeability, and high chemical resistance make it particularly suitable for large concrete structures in aggressive environments, such as the Bahrain Causeway.

Blast-furnace slag can also be used with ~10% calcium sulfate and 5% Portland cement to produce supersulfated cement. However, it is much more commonly used to extend and improve Portland cement.

Another use of pozzolans and blast furnace slag is to reduce the available alkalinity of cement systems to prevent the occurrence of alkali aggregate reaction (see later).

**Analysis** Basically the analysis of cements and related materials is carried out for elemental composition, mineral morphology, and particle size distribution. In the past wet classical methods were used but these were largely superseded by modern instrumental techniques in the early 1970s owing to their higher efficiency, better accuracy, and greater cost-effectiveness. Such techniques include AAS, ICP-AES, XRF, XRD, DTA, DSC, scanning electron microscopy/energy-dispersive X-ray analysis (SEM/EDX) for elemental composition and laser light scattering, sedimentation, or the Coulter method for particle size.

XRF or AAS/ICP-AES is used to obtain an elemental composition such as percent  $\text{SiO}_2$ ,  $\text{CaO}$ ,  $\text{MgO}$ ,  $\text{Al}_2\text{O}_3$ ,  $\text{Fe}_2\text{O}_3$ ,  $\text{SO}_3$ ,  $\text{Na}_2\text{O}$ , and  $\text{K}_2\text{O}$ . For XRF analysis either a homogeneous glass bead is obtained by fluxing the milled sample with lithium tetraborate at 1100°C or a briquette is obtained by pressing. Glass beads overcome particle size effects and are most commonly used in R&D laboratories for nonroutine work. Briquettes are commonly used at production plants for routine analysis. Prior to AAS/ICP-AES analysis the sample is turned into the liquid state by a suitable method, e.g., fusion or reaction with hydrofluoric or mineral acids in a bomb calorimeter at 150°C.

Loss on ignition measurements and levels of calcium hydroxide, calcium carbonate, basic carbonates, and organic material (e.g. from grinding aids) are determined by thermogravimetric analysis. This technique shows weight loss at characteristic temperatures that can be related to the amount of various phases present. Free (uncombined) lime ( $\text{CaO}$ ) is determined by extraction with ethylene glycol and titration with  $0.1 \text{ mol l}^{-1}$  hydrochloric acid.

Qualitative and semiquantitative analysis of mineral phases and the estimation of glass contents are carried out by XRD. The mineralogical composition of ordinary Portland cement can also be estimated from the oxide analysis by the Bogue equation (ASTM C 150 (1986)). In addition, optical microscopy on polished sections has been very successfully employed over the years both to estimate the phases present and to provide insight into such things as kiln burning conditions.

The particle size distribution is usually measured by BS 410 sieves down to 45  $\mu\text{m}$  and by laser light scattering, sedimentation, or electrical (Coulter) techniques for particles finer than this. The cement must be suspended in a nonaqueous solvent such as dried isopropyl or butyl alcohols. SEM in conjunction with



EDX has also been used for the characterization of cementitious particles (e.g., OPC/PFA blends).

### Cement and Concrete Admixtures

Admixtures are chemicals that are added to cement and concretes to impart desirable characteristics such as retardation, acceleration, water reduction/workability, air entrainment, and impermeability. Admixtures can be classified as shown in Table 3.

Cement admixture companies are usually formulators rather than base chemical producers. Commercial admixtures are usually a synergistic blend of the above chemical types to impart a desired range of properties to the cement/concrete.

**Analysis** Admixtures are usually tested in the as-supplied condition in accordance with BS 5075 and/or ASTM C 494. The requirements of these standards include relative density, pH, solids content, chloride ion, ash, total alkali content, and occasionally a standard IR scan. For formulation purposes,

additional testing is carried out for composition and molecular mass by modern analytical techniques such as ultraviolet, Fourier-transform infrared (FTIR) spectrometry, LC, GLC, DTA/DSC, thermogravimetry (TG), and microscopy (for air entraining admixtures.) This very specialist work is unique to the author's laboratory.

For the analysis of hardened concrete the determination of admixtures is a difficult and seldom performed task. The above techniques have been used in the authors' laboratory. Water-reducing and retarding admixtures in hardened concrete have been successfully determined. The problem with the other admixtures is not the method of analysis but the difficulty of their extraction from the hardened concrete matrix owing to their tendency to form complexes. Sodium carbonate extraction followed by neutralization with an ion exchange resin is generally the preferred technique. This is usually followed up by modern analytical techniques. For air-entraining admixtures, microscopy is used to check that a good bubble size distribution has been achieved. SEM is being used for the characterization of mineral admixtures.

**Table 3** Cement and concrete admixtures

<i>Principal effect</i>	<i>Chemical types</i>
Retardation	Hydroxycarboxylic acids and their salts, e.g., citric, tartaric, gluconic, glucoheptonic, maleic, salicylic, tannic Carbohydrates, e.g., glucose, sucrose, hydroxylated polymers (corn starch syrup) Chelating agents, e.g., EDTA, NTA, borates Inorganic salts, e.g., zinc, lead, phosphates
Retardation/water reduction	Lignosulfonates, corn starch syrup
Water reduction	Salts of sulfonated naphthalene and melamine formaldehyde condensates, lignosulfonates
Acceleration	Calcium chloride, calcium formate, sodium nitrite, sodium carbonate, triethanolamine, calcium thiocyanate, Ca/Na thiosulfate, sodium aluminate, Na/K silicates
Air-entrainment	Lignosulfonate, alkyl aryl sulfonates, proteins, vinsol resin, ethoxylates, fatty acid derivatives
Air-detrainment	Organophosphates such as tributylphosphate (TBP), silicones
Permeability reduction	Silica fume, granulated ground blast furnace slag (GGBFS), pulverized flyash or fuelash (PFA), polymers such as styrene butadiene rubber (SBR), polymers and copolymers of ethylene and vinyl acetate
Water retention	Cellulose ethers, polyethylene oxide, vinyl polymers, and copolymers

### Clays

Clays were formed many millions of years ago by the weathering of rocks such as granite and gneiss. They are all silicoaluminates and their use depends on chemical composition and crystal structure.

Kaolin (38%  $\text{Al}_2\text{O}_3$ , 45%  $\text{SiO}_2$ , and 14% bound water) is predominantly used for chinaware and papermaking. It can be burned to  $\sim 900^\circ\text{C}$  to form meta-kaolin, which is another useful pozzolan for addition to Portland cement. Further burning to  $\sim 1100^\circ\text{C}$  produces mullite ( $3\text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2$ ), which is used as a refractory.

Terracotta clays have a higher iron content than kaolin and are fired to make bricks, tiles, and pipes. In general, the higher the firing temperature, the lower the porosity of the final product and the higher the strength. Temperatures used range from  $970$ – $1150^\circ\text{C}$  for common bricks to  $1200$ – $1300^\circ\text{C}$  for engineering bricks. Bricks tend to be fairly porous to water; chemicals such as silicones can be used to improve water-repellency.

Montmorillonite clays are used in many construction grouts to impart thixotropic and water displacement properties. Another useful group of clay-like minerals are the illite micas e.g., muscovite,  $(\text{OH})_4\text{K}_2(\text{Si}_6\text{Al}_2)\text{Al}_4\text{O}_{20}$ , biotite,  $(\text{OH})_4\text{K}_2(\text{Si}_6\text{Al}_4)(\text{MgFe}_6)\text{O}_{20}$ , and phlogopite,  $(\text{OH})_4\text{K}_2(\text{Si}_6\text{Al}_2)\text{Mg}_6\text{O}_{20}$ . These can be used for making fireproof board, lamp chimneys, and electrical insulating materials.

**Analysis** The analysis of aluminate silicate clays is carried out to determine elemental composition, interlayer and bound water, particle size, and crystal structure.

Like cements, the composition of clay is established by either XRF or AAS/ICP-AES. However, it is necessary to modify the method of preparation of the fused bead by using a mixture of lithium tetraborate and lithium metaborate as the flux rather than lithium tetraborate alone.

Thermogravimetric (TG) analysis is used to determine the water content of clays. The weight loss at 110°C measures the interlayer water, whereas the bound water is lost between 400°C and 1000°C. The temperature at which bound water is lost is used in the identification of clays. For example, kaolin decomposes at 500–650°C to lose its bound water, whilst illite shows staggered weight loss between 400°C and 1000°C. Identification of clays is also carried out by IR spectroscopy in conjunction with XRD. The crystal structure is usually determined by XRD. Refractive index is also a very useful indicator of clay types.

### Calcium Sulfate

Calcium sulfate has many uses in the building industry, from the cement systems already mentioned to gypsum plasters and boards. It can occur naturally as anhydrite ( $\text{CaSO}_4$ ) or gypsum ( $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$ ). Large amounts are also available as by-products from various industrial processes, for example, the manufacture of hydrogen fluoride and phosphoric acid, and the desulfurization of flue gases in coal-fired power stations.

When gypsum is heated to about 160°C it forms calcium sulfate hemihydrate ( $\text{CaSO}_4 \cdot \frac{1}{2}\text{H}_2\text{O}$  – plaster of Paris). Further heating to about 190°C leads to the formation of soluble  $\gamma$ -anhydrite ( $\text{CaSO}_4$ ). At 380°C phase transformation takes place and  $\beta$ - $\text{CaSO}_4$  is formed.

When calcium sulfate hemihydrate is mixed with water it sets almost immediately to reform gypsum. Retarders such as tartaric acid or keratin can be added to slow down the reaction.

Anhydrite reacts much more slower than plaster, particularly if it contains an appreciable proportion of free lime. Accelerators such as potassium or aluminum sulfate can then be added. Anhydrite is often used in floor screed compositions.

**Analysis** Calcium sulfates are analyzed for particle size, type (crystalline form), pH, moisture content, and impurities such as free silica, lime (calcium oxide/hydroxide) and calcium fluoride. TGA and DTA

are complementary for the identification of calcium sulfate type. Moisture can only be measured by Karl Fischer titration. Using the elemental composition, thermal analysis data, and an XRD investigation, a probable composition for a calcium sulfate can be established.

Like cements, the elemental composition is determined by XRF or AAS techniques. The XRF bead is made using lithium tetraborate at 1050°C. Sulfide content cannot be determined by XRF. Sulfite,  $\text{SO}_3^{2-}$ , and sulfate,  $\text{SO}_4^{2-}$ , are safely analyzed by XRF.  $\text{Na}_2\text{CO}_3 + \text{K}_2\text{CO}_3$  fusion is carried out for Ca, Mg, Fe, and Al analysis by AAS. Lanthanum chloride is used as a sulfate interference suppressant. Gravimetric sulfate determinations are also carried out by precipitation as barium sulfate. The Leco Carbon–Sulfur Analyzer can also be used for quality control purposes. The fluoride is determined by XRF or a pyrohydrolysis method. The measurement of particle size distribution is carried out in a manner similar to that for cements and clays.

### Rocks, Aggregates, and Sand

Quarried stone has been used for many thousands of years as a building material. All was well until the industrial revolution started to put acidic gases such as sulfur dioxide into the air. The ensuing acid rain has severely damaged many limestone buildings. Analysis is therefore required of flue gases from industrial processes to ensure emission targets are being met. Old limestone buildings can also require protection.

Aggregates and sand are bound together with cement to make concrete. It is important that the aggregates and sands used are compatible with the cement. Modern dry-process Portland cements contain more of the alkali metals present in the raw materials because much less is lost with the flue gases. These alkali metals tend to be concentrated in the pore water in the concrete, leading to a pH of  $\sim 13.5$ . Such a high pH can attack certain siliceous aggregates and sands to form alkali metal silicate. These silicate gels swell in moist conditions by the process of osmosis, which then cracks the concrete. In severe cases this may lead to a structure having to be demolished. This process is known as the alkali silica reaction (ASR) and it is important that aggregates and sands are tested prior to use and the alkali content of the cement is known. Current building codes of practice limit the amount of alkali expressed as  $\text{Na}_2\text{O}$  equivalent to  $3 \text{ kg m}^{-3}$  of concrete.

**Analysis** A very important property of aggregates and sands for use in concrete is particle size. For a

concrete mix to have the right degree of cohesiveness and workability it needs a controlled aggregate/sand grading. The usual method of size analysis for sands and aggregates is by means of BS 410 test sieves. XRF, XRD, and AAS can be used for the elemental composition. Impurities such as clays, calcium carbonate, iron salts,  $\text{AlO}(\text{OH})$ , and illites can be estimated using instrumental methods. The shapes, sizes, and refractive indices of particles can also be studied by microscopy.

### Glass

Glass is manufactured by heating a mixture of silica sand and various metallic oxides/carbonates together in a furnace to a smooth and bubble-free melt at  $\sim 1500^\circ\text{C}$ . The usual glass used in the construction industry is soda glass. Formulations are very variable but typically might be:

$\text{SiO}_2$	70–75%
$\text{Na}_2\text{O}$	14–16%
$\text{CaO}$	8–13%
$\text{Al}_2\text{O}_3$	$1\frac{1}{2}$ –2%
$\text{MgO}$	2–5%

Fiber glass has a higher  $\text{Al}_2\text{O}_3$  content of up to 14% with an equivalent reduction in  $\text{SiO}_2$ . Increasing the  $\text{Na}_2\text{O}$  content reduces the melting point. Acid treatment of glass increases the  $\text{SiO}_2$  level to 99% by the removal of alkalis and other materials. This increases the melting point to  $1100^\circ\text{C}$ . Acid-treated glass is used to produce a cost-effective coarse fiber.

All glasses when they cool from the melting temperature pass through a range of temperatures at which crystalline compounds are likely to form. This process is known as devitrification. Once the glass has cooled to room temperature its viscosity is so high ( $10^{15}$  poise) that spontaneous devitrification is extremely difficult.

Glass is highly resistant to attack by water atmospheric pollution. However, persistent exposure to polluted atmospheres dulls the surface, increases the risk of devitrification, and leads to a loss in strength.

**Analysis** XRF, AAS, ICP-AES, and wet-chemical methods are used to determine the elements present in glass. Major elements are  $\text{SiO}_2$ ,  $\text{Al}_2\text{O}_3$ ,  $\text{SrO}$ ,  $\text{BaO}$ ,  $\text{ZrO}_2$ ,  $\text{Na}_2\text{O}$ ,  $\text{K}_2\text{O}$ ,  $\text{Li}_2\text{O}$ ,  $\text{MgO}$ ,  $\text{PbO}$ , and  $\text{CaO}$ . Minor elements include  $\text{Cu}$ ,  $\text{Co}$ ,  $\text{Cr}$ ,  $\text{Fe}$ ,  $\text{Mn}$ , and  $\text{B}$ .

XRF is carried out on pressed briquettes or fused beads. Lithium and boron cannot be determined by XRF. Fusion with sodium carbonate is carried out before the aqueous leachate is subjected to AAS,

ICP-AES, or flame emission photometry. Wet-chemical analysis can also be carried out on the above leachate.

## Metallic Inorganic Materials

Metals frequently used in construction are steels, alloy steels, aluminum, copper, lead, and zinc. The prevention of metal corrosion is extremely important.

### Steels

The first stage in the production of steel is the manufacture of iron from iron ore in a blast furnace. Molten iron is drawn off from the bottom of the furnace. Impurities are removed with the slag, which has already been mentioned in connection with cements.

Oxygen is now blown through the molten iron to oxidize waste constituents. The resulting pure iron is termed 'wrought iron', which is used for gates, fencing, and the like.

Carbon, manganese, chromium, nickel, titanium, molybdenum, vanadium, and tungsten can be added to the molten pure iron to form various steels.

Steels with a very low carbon content ( $< 0.2\%$ ) are softish and ductile. Mild steel accounts for 90% of the output of the steel industry. In the building industry it finds uses in reinforcement bars and mesh for concrete owing to similar thermal coefficients of linear expansion, and in steel pipes, steel sheeting, etc.

Structural steels for use in beams, girders, etc., are generally made with a carbon content of 0.15–0.25%. They also include a number of other constituents, such as up to 1.5% manganese and up to 0.50% chromium. Sulfur and phosphorus must be less than 0.05% each. High-carbon steels (0.5–2.0% C) are used for hammers, drills, cold chisels, etc.

### Alloy Steels

Elements such as chromium, manganese, nickel, tungsten, vanadium, silicon, and molybdenum are added to steels to obtain alloys with improved properties, for example strength, elasticity, hardness, abrasion resistance, rust resistance, and chemical resistance.

Manganese increases tensile strength and hardness at the expense of ductility. Chromium and nickel improve strength and toughness and their alloys are used for high-tensile steel bolts, for the manufacture of steel wire and cable for prestressing and post-tensioning, and for the construction of springs. Tungsten, vanadium, molybdenum and cobalt are used for

cutting steels that maintain their hardness almost to red heat.

Chromium and nickel are added to form stainless steels. When chromium is exposed to a corrosive atmosphere, an adherent film of chromium oxide ( $\text{Cr}_2\text{O}_3$ ) is formed. Nickel serves to improve the stability of the oxide film. Martensitic stainless steels contain  $\sim 0.1\text{--}0.4\%$  carbon and  $\sim 14\%$  chromium. They are very hard and can be tempered. Their use is in stainless-steel cutting tools.

Austenitic stainless steels are by far the most common and contain  $\sim 18\%$  chromium and  $8\%$  nickel, hence the term 18/8 stainless steel. The austenitic structure, which is the solid solution of iron carbide and all the metal additives in iron, is stabilized at all temperatures, producing a soft and flexible product. It is used for fume hoods, decorative cladding, and architectural features.

Ferritic stainless steels contain  $0.1\%$  carbon and  $15\text{--}20\%$  chromium and have a high elongation figure. They are used almost exclusively for pressed and deep-drawn articles such as kitchen sinks.

## Aluminum

Aluminum is the third most abundant element in the earth's crust. Unfortunately, it is mostly present as complex silicates and clay from which extraction is uneconomic.

Aluminum production is therefore based on bauxite, with an alumina content of at least  $55\%$ . Bauxite is powdered, calcined, milled with strong caustic soda (sodium hydroxide) and treated with steam at 5 bar. The sodium aluminate formed is filtered and diluted and a small quantity of alumina is added for seeding purposes. The pure aluminum hydroxide that precipitates is filtered off, roasted, and then electrolyzed in an electric furnace with cryolite ( $\text{Na}_3\text{AlF}_6$ ) used as a flux at  $1000^\circ\text{C}$ . Carbon anodes are used and are rapidly consumed. The whole process is very energy intensive.

The electrical resistance of aluminum is 1.5 times that of copper; however, its density is only  $2700\text{ kg m}^{-3}$  compared with  $8900\text{ kg m}^{-3}$  for copper. Hence the conductivity of aluminum is twice that of copper per unit weight. It is therefore used widely for power transmission cables. Aluminum is in general more resistant to corrosion than iron and steel owing to the formation of a well-adhering coating of aluminum oxide.

Aluminum is readily attacked by alkalis. The rate of attack by wet cement is rapid, with hydrogen liberated during the reaction. This reaction is made use of in light-weight cement block production and

gas-forming grouts that use the gaseous expansion to overcome plastic settlement.

The primary aluminum alloys used in building are as follows.

1. Manganese alloys containing  $1.2\%$  Mn. These are stronger than pure aluminum and more corrosion resistant, with no loss of workability. They are widely used in cladding and roofing materials.
2. Magnesium alloys are strong, ductile, and extremely corrosion-resistant. They are used for tubular scaffolding and the like.
3. Magnesium silicide alloys contain up to  $1.5\%$   $\text{Mg}_2\text{Si}$ . They are used for extrusions such as window sections.

## Copper

Copper usually occurs in sulfide ores intimately mixed with iron sulfides and many other compounds. The ores are ground and the copper is concentrated by froth flotation. The concentrate is heated in a reverberatory furnace where a high-density molten copper 'matte' and a low-density molten slag are formed. The molten matte is separated from the slag and transferred to a converter where air is blown through, allowing the oxidation of copper sulfide ( $\text{Cu}_2\text{S}$ ) and other compounds to copper. Copper can then be either 'fire refined' for use in alloys or 'electrolytically refined' for high-purity grades.

Copper is extremely resistant to atmospheric corrosion as it rapidly forms a complex green hydrated copper(I) oxide/carbonate film that prevents further corrosion. Copper is very suitable for underground services as it is extremely resistant to attack. Corrosion troubles occur when copper and steel pipes are joined together, because copper induces corrosion in steel objects joined to it by electrochemical action.

Copper has good chemical resistance, but is attacked by concentrated sulfuric acid, nitric acid and ammonia.

Copper is mainly used in the building industry for water and gas pipes. Its corrosion resistance and ease of bending and joining with good frost resistance make it eminently suitable. Copper sheet is used in roof coverings and in hot water tanks and domestic boilers. The high electrical conductivity of copper is used to good effect in most electrical cables.

**Brasses** Brasses are alloys of copper and zinc. The most popular is 60/40 used for the manufacture of a wide range of fittings.

**Bronzes** Bronzes are alloys of copper and tin, usually with some other elements such as phosphorus added. Normal phosphor bronze contains  $5\text{--}6\%$  tin



and traces of phosphorus. It is used for woven wire filters and screens. Bronze containing 12% tin is used for ornamental doors and bronze fittings. Copper–nickel–silica alloys are used for architectural features such as shop windows and counters.

### Lead

Lead occurs as galena ( $\text{PbS}$ ), cerisite ( $\text{PbCO}_3$ ), and anglesite ( $\text{PbSO}_4$ ). Galena is the most common. Lead ores are usually roasted and then reduced in a blast furnace with coke.

Purification is carried out by adding zinc, which removes silver (Parkes process), or by electrolysis in a solution of lead hexafluorosilicate and hexafluorosilicic acid (Betts process).

One of the most useful properties of lead is its ability to absorb nuclear and X-ray radiation. Lead has good chemical resistance but is attacked by nitric and organic acids. Lead is extremely resistant to atmospheric corrosion as it rapidly becomes covered with an impervious film of oxides, carbonates, and sulfates.

Lead was used for water pipes and other plumbing components. However, it is no longer installed for the transportation of drinking water. In soft-water areas there are programs to replace existing water authority lead pipes with plastic owing to possible permanent adverse health effects amounting from even low levels of lead in drinking water.

Lead is better than copper for roofing on account of its durability, but it is expensive and relatively heavy. Lead is also widely used for flashings and trimmings, anti-vibration mats, and sound proofing.

### Zinc

The most common ore is zinc blende ( $\text{ZnS}$ ), which is found in association with lead, iron, and other sulfides.

Before smelting, zinc sulfide is usually concentrated by a float and sink method. The zinc sulfide is then roasted to form zinc oxide, followed by reduction with carbon and/or carbon monoxide. Zinc metal is distilled off.

About half the zinc produced is used as a coating metal. The most common method of coating is hot-dip galvanizing. Iron or steel objects are degreased, followed by immersion in a bath of molten zinc at  $455^\circ\text{C}$ . Zinc protects the base metal from corrosion by acting as a sacrificial anode. Other methods are molten zinc spraying, electroplating, sheradizing (objects are heated in powdered zinc just below its melting point at  $\sim 380^\circ\text{C}$ ), and painting with zinc-rich paints.

### Corrosion of Reinforcing Steel

Corrosion of metals is a problem that costs many millions of US\$1£ per annum. In building it is therefore essential to minimize the likelihood of corrosion at the design stage in a cost-effective manner.

Corrosion is an electrolytic process that requires an anode, a cathode, and an electrolyte. In reinforced concrete the reinforcing steel becomes covered in a passivating layer of  $\gamma$ -iron oxide formed in the high-pH environment ( $\sim \text{pH } 13$ ) provided by the Portland cement. Unfortunately, the high pH can be neutralized by acidic gases such as carbon dioxide in the atmosphere. Once the pH in the concrete drops below  $\sim 10$ , generalized corrosion of the steel will take place in the low-pH area.

This phenomenon can be overcome by ensuring good concrete cover over the steel, using high-quality low-permeability, well-compacted concrete, and coating the concrete with a surface coating that prevents the ingress of carbon dioxide while allowing the passage of water vapor.

The chemical test for this type of corrosion is very straightforward. A lump of concrete is freshly broken out from the corroding area and sprayed with a 1% phenolphthalein solution in isopropyl alcohol. If the phenolphthalein remains colorless, then the alkalinity of the concrete has been neutralized. If the phenolphthalein goes purple (i.e., the concrete is sufficiently alkaline) *but* reinforcement corrosion is observed, then the far more serious problem of chloride attack is likely to be occurring. An explanation of the mechanism of chloride attack follows.

When chlorides reach the steel the passivating layer is broken down and a small anode is set up. This positive anode attracts further chloride ions, leading to local concentration and a lowering of pH. The steel away from the area of attack acts as a large cathode. This type of corrosion is more serious than the generalized corrosion caused by carbonation of the concrete because pits are formed and steel cross-section is rapidly lost, which can lead to structural failure.

Chlorides can be present in concrete from a number of sources. Calcium chloride is the best accelerator for Portland cement and was frequently added in precast concrete in the 1960s to maximize plant throughput. Chlorides will be present if unwashed marine dredged aggregates are used or seawater is used for mixing. A small amount of chloride can also be present in the Portland cement. Even after care was taken to minimize chlorides present in concrete, structures still started to show corrosion damage within a few years of completion. This is caused by the ingress of chlorides into the concrete

**Table 4** Major and minor constituents of various metals and alloys

Metals/alloys	Constituents	
	Major	Minor
Iron and steel	Fe, C, S, Si, Mn, P, Cr, Ni, Co, Cu, Nb, W, V, Mo	Sn, Sb, Zr, Pb, As, Bi, Ti, Mg, Te, Ce, La, V, Al
Copper	Cu	Sn, Pb, Zn, Fe, Ni, Al, Si, Mn, As, Bi, Sb
Aluminum/Al alloys	Al, Si, Mg, Pb, Zn, Cu, Ni, H	Fe, Mn, Cr, Ti, Sr, Zn, Na, Bi, V, Ca, P, Be, Li
Brass	Cu, Zn	Sn, Pb, Fe, Ni, Al, Si, As, Mn, Bi, Sb, P, S, Cr, Co, Mg
Zinc	Zn	Pb, Mg, Al, Cd, Fe, Sn, Cu, Ni, Mn
Lead	Pb	Sb, Sn, Bi, Cu, As, Ag, Zn, Cd, Ni, Te, Ca

from the use of deicing salts or from marine environments. The permeability of concrete to chlorides can be minimized by using a penetrating siloxane sealer, low water/cement ratios, and a PFA- or slag-modified cementitious phase.

Analysis is required to ensure all materials used in the manufacture of concrete contain less than the specified amounts of chloride, to assess the depth of chloride penetration into concrete from external sources, to determine the water/cement ratio used, and to describe the type of cement employed.

#### Analysis of Metals

The direct-reading spectrograph is still a widely used instrument for analysis for metals and most low-level residual impurities using the ASTM E414-7 method. The technique consists of briquetting the sample in the form of chips, drillings or powder, followed by excitation in a d.c. arc opposite a high-purity metal rod.

Since the 1960s new techniques such as AAS, ICP-AES, and XRF have been used for plant control and environmental measurements. Iodometric titrations are still used in many industrial laboratories, but the newer methods are more rapid and convenient. Other techniques including colorimetry, electrodeposition,

X-ray microprobe, neutron activation, and electro-analytical methods have been mentioned in the literature. For irons and steels the carbon and sulfur level is determined by a Leco carbon/sulfur analyzer or similar in addition to the elemental analysis. Similarly, the hydrogen content is measured in aluminum by a hydrogen analyzer. Table 4 summarizes the elements determined for various metals and alloys.

*See also:* **Cement.** **Forensic Sciences:** Glass. **Glasses.** **Paints:** Water-Based; Organic Solvent-Based. **Pesticides.**

#### Further Reading

- ASTM (2002) *Annual Book of ASTM Standards*. Conshohocken, PA: ASTM.
- Bianchina P (1993) *Building Materials and Techniques*. New York: Wiley.
- Cardarelli F (2000) *Materials Handbook*. London: Springer.
- Harper CA (2000) *Modern Plastics Handbook*. New York: McGraw-Hill.
- Salamore JC (1999) *Concise Polymeric Materials Encyclopedia*. Boca Raton, FL: CRC Press.
- Ullmann's Encyclopedia of Industrial Chemistry* (2001). Weinheim: VCH.

# C

## CADMIUM

**A Rodríguez-Cea, M R Fernández de la Campa, and A Sanz-Medel,** University of Oviedo, Oviedo, Spain

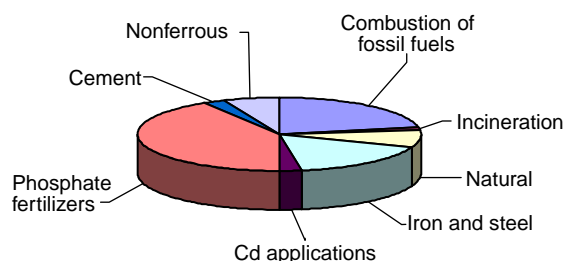
© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Although cadmium emissions into the environment have generally decreased in recent years, they still constitute a severe environmental problem in many areas, putting human health at a risk. Since Cd mobility and toxicity depends on its chemical form, the speciation analysis of Cd seems to be the more realistic way for the understanding, diagnosis, and eventual treatments for such potential health problems. The importance of Cd in living organisms relates not only to its bioaccumulation but also to its active role as an inducer of the biosynthesis of metallothionein. For this reason, the interest of researchers has moved recently from environmental samples toward the biochemical speciation analysis of Cd in biological fluids and tissues. Today, Cd speciation analysis in bioinorganic analytical chemistry includes determinations of the bioinorganic moiety (e.g., Cd at very low levels) and the bioorganic ligand (e.g., proteins, polypeptides, oligopeptides, etc.). Mass spectrometry (MS) techniques are critical, as we will see later on, in such studies that eventually aim to understand the role that Cd and its species play in living organisms. Also, the application of isotope dilution analysis to the precise quantification of such Cd species will be finally addressed in this article.

### Cadmium Sources and Toxicity

Apart from natural emissions, Cd is usually found in the environment because of several anthropogenic activities. Due to its chemical and physical properties, Cd is widely used in special alloys (e.g., copper-, tin-, lead-, or zinc-based alloys), pigments (as Cd sulfide and selenide), stabilizers (e.g., Cd incorporated into poly(vinyl chloride)), NiCd batteries and coatings (e.g., steel, aluminum and other non-ferrous metals). Of course, the manufacturing,



**Figure 1** Relative importance of the various cadmium sources to human exposure.

increasing use, and disposal of products containing Cd will eventually release it into the environment. However, major Cd emissions into the environment arise from products to which Cd was not intentionally added (i.e., it was an impurity). For instance, the combustion of fossil fuels, iron and steel production, and the use of phosphate fertilizers in agriculture constitute the main anthropogenic sources of Cd emissions to air, water, and soil. The relative importance of the various Cd sources to human exposure is illustrated in Figure 1.

The ingestion of foods, e.g., plants grown in soils containing Cd or meat from animals that ingested those plants, and inhalation, e.g., cigarette smoking or from Cd occupational exposure, are considered the main routes of human exposure to Cd, the maximum tolerable daily intake being 1 µg per kg of body weight (World Health Organization).

Concerning its toxicity, Cd is known to accumulate in the human kidney for a relatively long time and it causes serious renal dysfunction. Moreover, breathing high levels of Cd causes severe damages to lungs and several studies have pointed out its likely role in human carcinogenesis.

Therefore, constant Cd emissions due to anthropogenic sources has led to severe environmental pollution problems and increased interest on studying the mechanisms of this metal toxicity.

### Cadmium Speciation, What and Why?

In general, identification and quantitation of elemental species provide very important information about



its impact on the environment, its bioaccumulation, and toxicological effects. This fact has led to the development of the so-called 'speciation analysis'. As for other toxic elements, Cd toxicity depends upon its physicochemical form; hence, the determination of its different chemical forms is paramount for a correct evaluation of the environmental and health risks posed by Cd. For example, inorganic species of Cd are more toxic than organic ones (and, of course, among the inorganic ones,  $\text{Cd}^{2+}$  is most toxic).

### Cadmium Speciation in Environmental Samples

Most Cd speciation studies available have been carried out in biological systems, while comparatively little attention has been paid to the speciation of this metal in environmental samples and matrices (e.g., water samples, soils, sediments, or fly ash). Moreover, most of such studies devoted to Cd species in the environment are closer to fractionation than to real chemical speciation analysis. Fractionation, defined by International Union of Pure and Applied Chemistry (IUPAC) as "the process of classification of an analyte or a group of analytes from a certain sample according to physical or chemical properties", has been used as a separative step ('sequential extraction' procedures and rough classification of the total metal content into fractions of different toxicities) in soils, fly ashes, and sediments evaluated from the point of view of their environmental or toxicological risk. Typical sequential extraction procedure usually allows classifying Cd in at least four fractions: exchangeable, reducible, oxidizable, and total residual fraction. In soils and sediments, several studies pointed out that Cd shows a strong binding to organic matter (e.g., to humic and fulvic acids) and a relative sorption to clay, which limited its bioavailability.

In water samples, Cd speciation studies most often comprise the physical separation of  $\text{Cd}^{2+}$  from its inorganic or organic Cd complexes, along with the analysis in the sample of other elements, including Cu, Co, Fe, Mn, Ni, Pb, and Zn. The determination of  $\text{Cd}^{2+}$  in environmental samples usually involves a preconcentration step, which also serves to remove the sample matrix. Thus, this step is commonly accomplished for speciation in waters by passing the water sample through a column packed with an appropriate chelating cation exchanger. Final determination of Cd, released with a convenient reagent, can be achieved by atomic absorption spectrometry (AAS), electrothermal atomic absorption spectrometry, inductively coupled plasma-optical emission spectrometry (ICP-OES), or ICP-MS. Recently, flow injection analysis systems have been recommended

for carrying out the preconcentration step and the final release of the metal for transport to the specific (atomic) detector.

The preconcentration step is in particular useful in a saline matrix (e.g., seawater Cd speciation). For instance, ICP-MS is susceptible to polyatomic ion interferences from seawater and its performance is much better after such a matrix removal step.

Finally, electrochemical techniques including voltammetric techniques and ion-selective electrodes (ISE) for Cd have also been reported in the literature for 'free' Cd determinations in waters, allowing the differentiation between labile Cd species from strong Cd complexes (e.g., bound to humic and fulvic acids and colloids). Mercury is still the electrode material of choice for detection of Cd due to its large hydrogen overvoltage and its remarkable reproducibility.

The development of routine and easy handling procedures for continuous and real-time speciation of trace metals in waters has led, in the last years, to the development of microsensors coupled to voltammetric techniques. Microelectrodes offer several advantages for speciation measurements in real-world samples, including their application in low ionic strength media (e.g., freshwaters), reproducibility, and sensitivity. Some Cd speciation studies carried out in river waters, heavily loaded with suspended material, using microelectrodes demonstrated that most of Cd was associated with colloidal material. In addition, this technique also enables the determination of the corresponding complexation stability constants for Cd and protons.

### Speciation of Cadmium in Biological Material

Speciation analysis of Cd (as of other metals, e.g., Cu and Zn) in biological materials has become a real challenge, in trying to understand the role of the metal in living organisms. In this area, most investigations have focused so far on Cd speciation in metal-proteins or metal-peptides complexes (e.g., metallothioneins and phytochelatins) in biological tissues. Cd speciation in biological fluids (e.g., serum, urine, and milk) has received lesser attention.

Metallothioneins (MTs) are a group of small cysteine-rich proteins that exhibit high affinity for some metal ions such as  $\text{Cd}^{2+}$ ,  $\text{Zn}^{2+}$ , and  $\text{Cu}^{+}$ , forming characteristic metal-thiolate clusters. Apart from homeostatic functions, MTs play a role in the detoxification of toxic metals (e.g., Cd, Hg, and Ag). Thus, MTs can chelate  $\text{Cd}^{2+}$  and accumulate it in the form of nontoxic Cd-MT complexes. The fact that multiple forms of MT exist in many biological systems has lead to a great interest in studying

individual forms of MTs (the so-called MT isoforms) in order to clarify their actual biological role. In this sense, Cd speciation studies in MTs are generally concerned with the complexation of Cd with these individual forms.

MTs were first isolated from equine kidney, but they have widely been found throughout the animal kingdom. In typical mammals, there are two main isoforms, MT-1 and MT-2, named in the order of their elution from an anion-exchange column. Several experiments under Cd exposure, as well as under natural conditions, have demonstrated that Cd binds to both MT isoforms, in combination with copper and zinc. The relative affinity of such metal ions for MTs is in the following order:  $\text{Zn}^{2+} < \text{Cd}^{2+} < \text{Cu}^{+}$ .

MTs in invertebrates have also been named MT-like proteins, MLPs, because they differ significantly from those of vertebrates. In spite of these differences, both vertebrate and invertebrate MTs have been isolated for Cd speciation analysis using similar analytical procedures. A typical preparative procedure to extract MTs from animal tissues is given in Scheme 1.

On the other hand, it is known that Cd accumulation in soil and water now poses a major environmental problem due to various industrial and agricultural activities (as reported in previous sections). In fact, the first papers on Cd speciation were mainly focused on such environmental issues.

The use of metal-accumulating plants to remove this toxic metal from soils and water streams has been proposed as a possible remediation of this problem. This type of environmental restoration is termed phytoremediation. In plants, Cd appears to accumulate preferentially in sulfhydryl-rich peptides named phytochelatins (PCs). PCs have the general formula  $(\text{GluCys})_n\text{Gly}$  where  $n$  is between 2 and 11.

It is known that these peptides are induced in plants exposed to  $\text{Cd}^{2+}$ , which in turn may form PC–Cd complexes removing toxic-free  $\text{Cd}^{2+}$  from the solution.

As MTs, PCs are synthesized in plants in response to other metals also (e.g., Cu, Ag, Pb, Zn, or Hg), but the study of PCs in connection with Cd is most important today because this metal is the most efficient inducer of PCs in plants.

In conclusion, the need to speciate Cd contents in both MTs and PCs is of great scientific and analytical interest nowadays.

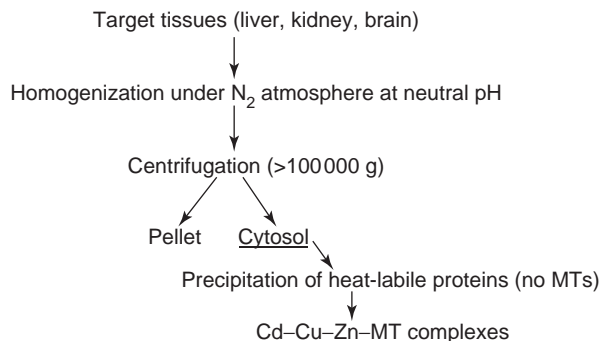
### Cadmium Speciation in Metallothioneins and Related Molecules by Electrochemical Techniques

Electrochemical techniques were extensively used to measure ‘free’ Cd in waters. In the case of MTs and related molecules, the amino acid chain is electroactive due to the presence of thiol groups. Moreover, the reduction of  $\text{Cd}^{2+}$  is electrochemically reversible at the mercury electrode and, therefore, it is possible to obtain two different responses simultaneously for a given sample. In addition, the electrochemical response depends on the chemical form of the element and allows one to easily monitor changes of the different species in solution. At present, many studies on the electrochemical behavior of Cd–thioneins have been reported. Most of them used mercury as the working electrode and so the polarograms of MTs exhibit one peak that is attributed to the oxidation of the mercury electrode in the presence of the Cd–thionein complexes.

Some advantages of electrochemical techniques are that the redox potentials of MTs at the electrode surface or the electrode solution interface can be accurately determined, allowing to probe the electroactive groups that undergo the electron transfer reaction. On the other hand, the precise control of the electrode potential allows examining Cd release from the MT or Cd uptake by the MT molecules.

Recently, electroanalysis of Cd-complexation with different mammalian MTs (e.g., rabbit liver, horse kidney, and human kidney) has become more frequent. The study of cysteine-containing peptides such as glutathione and MTs peptidic fragments (mainly  $\alpha$ -domain of mammalian MTs), in the absence and in the presence of  $\text{Cd}^{2+}$ , has been carried out by means of several electrochemical techniques such as differential pulse polarography (DPP), linear sweep voltammetry, cyclic voltammetry, direct current

#### Sample treatment scheme



**Scheme 1** Typical sample procedure to obtain the cytosolic fraction that contains Cd–MT complexes. Oxidation of metal–MTs complexes is avoided with N<sub>2</sub> atmosphere and the inclusion of reducing agents in the buffers. Cytosol can suffer a heat step in order to termocoagulate non-MTs proteins.

polarography, square wave voltammetry (SWV), and anodic stripping voltammetry. For instance, DPP is a good technique for providing information about complexation properties. This technique has been applied to *in vitro* studies of synthetic MTs in order to understand the properties of Cd–MT complexation. Additional advantages of DPP and SWV are that these techniques allow very low metal concentrations to be tested, where potentiometry with ISE are not reliable.

In real samples, the voltammetric responses of differential pulse anodic stripping voltammetry have been applied to investigate the Cd-binding properties in MLPs from mussels. The existence of two different Cd–MLP complexes was thus demonstrated.

## Cadmium Speciation Analysis Using Hyphenated Techniques

As for other elements, the analytical techniques developed to tackle real-life Cd speciation problems are mainly hybrid techniques today. Hybridization is achieved by coupling a powerful separation technique to an element-specific detector allowing the online determination of Cd in the separated species. For instance, the coupling of liquid chromatography (LC) with AAS or with ICP-OES turned out to be especially useful for the Cd-speciation in proteins in contrast to more classical offline methods. Nevertheless, ICP-MS has taken over as the most powerful atomic technique for the simultaneous determination of trace metals, due to its extreme sensitivity, capability of giving isotopic information, and a relative lack of interferences. Due to such characteristics, liquid chromatography LC–ICP-MS is today the preferred hybrid approach in speciation of Cd, although capillary electrophoresis (CE)–ICP-MS has also been proposed. Of course, ‘organic’ MS should be used to identify and characterize the particular bioligand (e.g., protein) binding Cd. In this latter vein, matrix-assisted laser desorption ionization–time-of-flight and electrospray ionization–tandem mass spectrometry (ESI–MS/MS) are increasingly used in speciation work. All those ‘hyphenated techniques’ have been successfully employed for the separation and characterization of Cd-containing metalloproteins and metal–peptides complexes (such as MTs and PCs, respectively).

### LC Coupled to ICP-MS

Cadmium species in metalloproteins can be separated in a first step by size exclusion chromatography (SEC) according to their size (and, to a lesser extent, their shape) enabling the separation of higher

molecular and lower molecular weight Cd-containing proteins. Although the online coupling of (SE)–LC with ICP-MS allows real-time element-specific detection of various Cd-containing molecular weight fractions, the SEC resolution is insufficient for Cd–MT isoforms’ discrimination. Further separation studies of a given MT fraction, using complementary chromatographic separation techniques (e.g., anionic exchange, reverse phase, or capillary zone electrophoresis, (CZE)), are needed.

On the other hand, (SE)–LC coupled to ICP-MS, and also coupled to UV and AAS detection, have been used to investigate PCs and PC–Cd complexes. Unfortunately, SEC does not provide enough resolution to separate all individual PC–Cd complexes. Thus, SEC has become a sort of preparative step to separate the MT and PC pools from the matrix of the sample.

Although few studies have been carried out for Cd speciation in biological fluids (e.g., human serum, human milk, cow milk, and formula milk), in cow and formula milk Cd seems to be bound to caseins while in human serum Cd appears to be complexed by ceruloplasmin, a well-known glycoprotein for metal storage.

On the other hand, anion exchange (AE)–LC has been used for the separation of the two main MT isoforms (MT-1 and MT-2) based on their different negative charge at neutral pH. This technique, especially the so-called anionic exchange fast protein liquid chromatography (AE–FPLC), offers good resolution. The use of FPLC with a strong anion exchanger allows the fast isolation of proteins, peptides, and amino acids with high efficiency. The coupling of AE–FPLC to ICP-MS has been successfully applied for the isolation and characterization of Cd–metalloproteins in both invertebrates (MLPs in mussel tissue) and in vertebrates (fish liver and kidney MTs), indicating the urgent need of Cd speciation studies in environmental issues.

Once a fraction is separated (e.g., MT-1), further separation of MT isoforms is possible using reverse phase (RP)–LC, which is the preferred technique in terms of resolution because the packing material is usually free of ligands for metals. The hybrid technique RP–LC–ICP-MS has been used, in fact, for such speciation purposes in biological samples as liver, kidney, and brain MTs, the main limitation being the loss of sensitivity in the ICP-MS due to the presence of the organic modifiers.

### CZE as a Tool for Cadmium Speciation

CZE is a well-established separation technique for metalloproteins and metal-binding peptides owing to

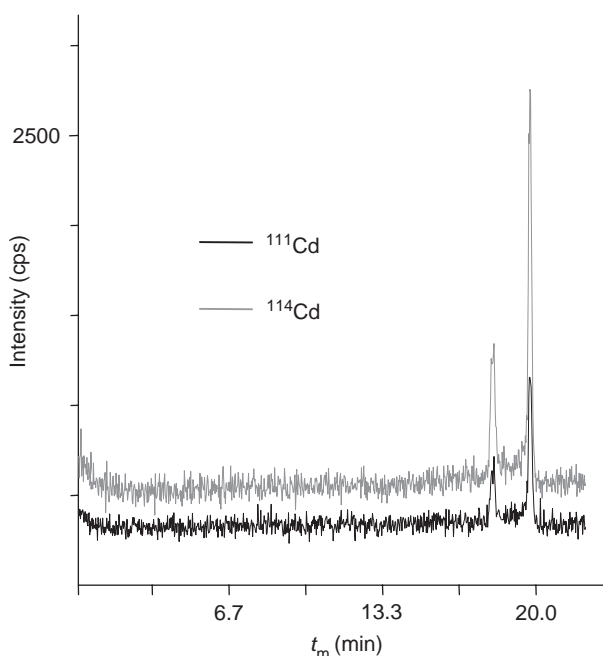
its high separation efficiency, the small sample requirement (several nanoliters), short analysis times, versatility in modes of operation, and the absence of packing susceptible to interactions with metals affecting the complexation equilibria. Many authors have shown the analytical potential of CZE as a high-resolution technique for the speciation of Cd-MTs. Cd-MTs can be well resolved using uncoated fused-silica capillaries under neutral or slightly alkaline pH. Under these conditions, MTs (pI values in the range of 3.9–4.6) have negative charge and move counter-electroosmotically toward the inlet end in normal polarity conditions. However, they can be detected due to the electroosmotic flow which is higher in magnitude than the electrostatic force they experiment toward the anode. CE-ICP-MS separations are particularly difficult because the different MT isoforms and subisoforms appear very close to each other in the timescale. Thus, any suction effect in the nebulizer would degrade the separation achieved. Moreover, the low flows emerging from a CE capillary do not match those required in typical ICP-MS nebulization devices. For this reason, the design of the interface between CE and ICP-MS is critical.

Due to nanoliter per minute flows of CE the concentration-based detection limits are seriously degraded in comparison with those obtained by LC, limiting the application of CE in Cd speciation analysis of real samples. In any case, there are CE strategies (e.g., large volume sample stacking) that are able to preconcentrate the analytes and have been successfully applied to real samples in order to improve the detectability of Cd species in MTs by CE-ICP-MS. **Figure 2** shows typical electrophoretic results by a hybrid technique for speciation of Cd-MTs in fish samples of environmental monitoring interest.

The use of CE coupled to high-resolution ICP-MS instruments has also been reported in the literature. It has been shown that Cd-MT complexes can be separated by CE and different metallic elements and sulfur can be simultaneously detected by using a high-resolution sector field (SF) ICP-MS detector operating at medium resolution to avoid polyatomic interferences that affect the determination of sulfur.

#### Gel Electrophoresis with Laser Ablation Applied to Cadmium Speciation in Proteins

Gel electrophoresis is a well-known separation technique for complex media such as proteins. However, classical modes of detection (including dye staining, immunoreaction with antisera, and autoradiography) do not allow the detection of metal-protein complexes.



**Figure 2** Electrophoretic separation of fish Cd-MTs complexes (real sample) by CE-ICP(Q)MS. Signals for  $^{111}\text{Cd}$  and  $^{114}\text{Cd}$  were monitored showing that Cd binds two main MT isoforms. (Alvarez-Llamas G, Fernández de la Campa MR, and Sanz-Medel (2003) Sample stacking capillary electrophoresis with ICP-(Q)MS detection for Cd, Cu, and Zn speciation in fish liver metallothioneins. *Journal of Analytical Atomic Spectrometry* 18: 460–466; reproduced from The Royal Society of Chemistry.)

In the context of chemical speciation, a new hyphenated technique, laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS), in combination with gel electrophoresis, appears as an emerging and powerful tool for metal complexation studies of proteins, giving multielement information about metals bound to the previously separated metalloproteins.

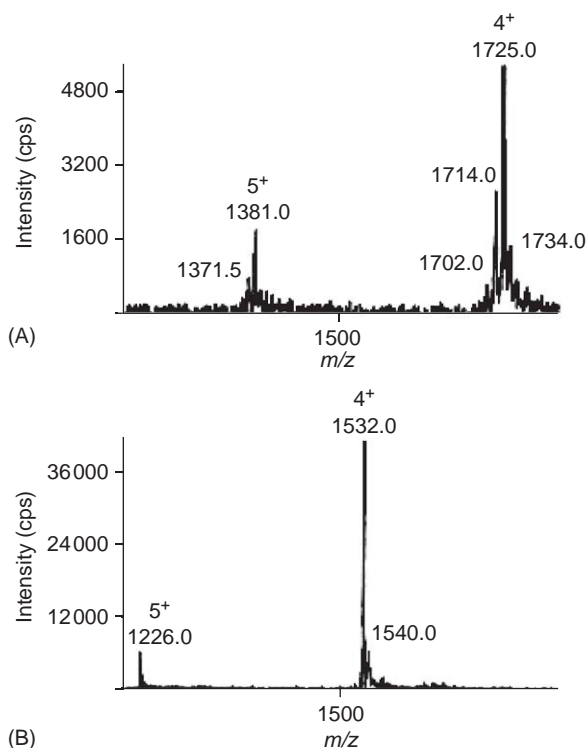
After the separation of the protein has taken place by one- or two-dimensional electrophoresis, gels are dried and subjected directly to LA-ICP-MS in order to detect, map, and quantify metal distribution in individual stains (compounds). In brief, a spot of the sample is ablated by the laser and the ablated plume is brought to the plasma by a continuous gas flow, generally argon. Then, ICP-MS gives the multielemental composition of the protein present in the ablated site. This method has been already reported for Cd speciation in MTs and in other Cd-binding proteins from bacterial extracts. Metal protein patterns in gels coming from cells grown under Cd-stress conditions can be compared very quickly to nonstressed cultures to investigate induction of MTs by metals.

## Methods for Bioligand Identification and Determination

ESI is a soft ionization technique (in contrast to ICP) and can be used to produce primarily intact protonated molecular ions of peptides and proteins. This technique, in combination with MS or tandem mass spectrometry (MS/MS), has been investigated in the last years for the ionization of metal-containing species (including Cd species) to obtain precise information of molecular weights and structural characterization of such bioligands at trace levels in complex matrices.

In ESI, peptides and proteins molecules are protonated, generating from single to multiple charged ions (e.g.,  $[M + H]^+$ ,  $[M + 2H]^{2+}$ , ...,  $[M + nH]^{n+}$ ). The number of protons attached to the molecules depends on their molecular mass, but generally, metal-containing species with mass up to 1000 Da are singly protonated, while polypeptides, proteins, and their metal complexes became multicharged ions. Thus, Cd-containing proteins as MTs (6–7 kDa) usually acquire charges of +4 and +5 in the ESI. The ions generated are sampled into the high-vacuum region of the mass analyzer, most often a quadrupole, where they are detected according to their  $m/z$  ratio. In this way, the resolution even in the simple quadrupole allows to distinguish  $\pm 1$  Da of molecular mass, enabling to differentiate very similar compounds (e.g., charged protein isoforms and their metallated complexes), which allows the confirmation of the presence of known compounds and their metal complexes. On the other hand, when identifying and characterizing unknown chemical species, including Cd-containing peptides or proteins, ESI-MS/MS (e.g., a triple-quadrupole mass spectrometer system) is needed.

The application of LC online with ESI-MS is especially attractive for the speciation analysis of Cd and Zn complexes. Metallated complexes such as MTs may suffer from several artifacts in their chromatographic separation due to the presence of those metals. For instance, the presence of other metals may modify the retention of the complex Cd–Zn–MTs on a RP column. As an alternative, ESI-MS with pH control can help in the analysis of metals in native and reconstituted metalloproteins, in order to exactly elucidate which biocompound binds Cd, how many atoms of Cd per bioligand molecule are bound, and what is the combination with other metals (e.g., Zn and Cu) in such molecules. In this line, samples of rabbit liver and horse kidney MTs (previously isolated by SEC and anionic-exchange chromatography) were analyzed under pH selected conditions. In moderately acidic (pH 4.0) and neutral (pH 6–7) media, the complexes of MTs with Cd and Zn retain



**Figure 3** ESI mass spectra taken in the vicinity of the apexes of the RP-LC elution of rabbit liver MT-2. (A) Spectrum obtained without postcolumn acidification.  $M_{r1} = 6804$ ,  $Cd_5Zn_2$ -MT-2a;  $M_{r2} = 6852$ ,  $Cd_6Zn$ -MT-2a;  $M_{r3} = 6898$ ,  $Cd_7$ -MT-2a;  $M_{r4} = 6932$ ,  $Cd_7$ -MT-2c. (B) Spectrum obtained after postcolumn acidification.  $M_{r1} = 6125$ , apo-MT-2a;  $M_{r2} = 6156$ , apo-MT-2c. (Reprinted with permission from Chassaigne H and Lobinski R (1998) Characterization of metallothionein isoforms by reversed-phase high-performance liquid chromatography with online postcolumn acidification and electrospray mass spectrometric detection. *Journal of Chromatography A* 829: 127–136; © Elsevier.)

their conformation in the gas phase and produce ions of intact complexes at the ESI source. In contrast, strong acidic conditions (pH 1.9) cause the loss of Cd and Zn and so the corresponding apo-forms of the protein are obtained. Therefore, when ESI is applied to the sample under neutral pH, the mass spectra of the metallo-subisoforms are obtained. Conversely, under acidic pH the mass spectra of the apo-forms (free ligand) are obtained. As a result, the difference between both mass spectra allows the determination of the stoichiometry of the separated analytes (metallo-subisoforms). In this way, Cd and Zn were speciated in rabbit liver MT-2 (see Figure 3), showing that Cd was mainly bound to MT-2a subisoform ( $M_r = 6125$ ) and MT-2c ( $M_r = 6156$ ). Both subisoforms bind Cd and Zn forming several metal complexes (e.g.,  $Cd_7$ -MT,  $Cd_6$ -Zn-MT,  $Cd_4$ -Zn<sub>3</sub>-MT, etc.).

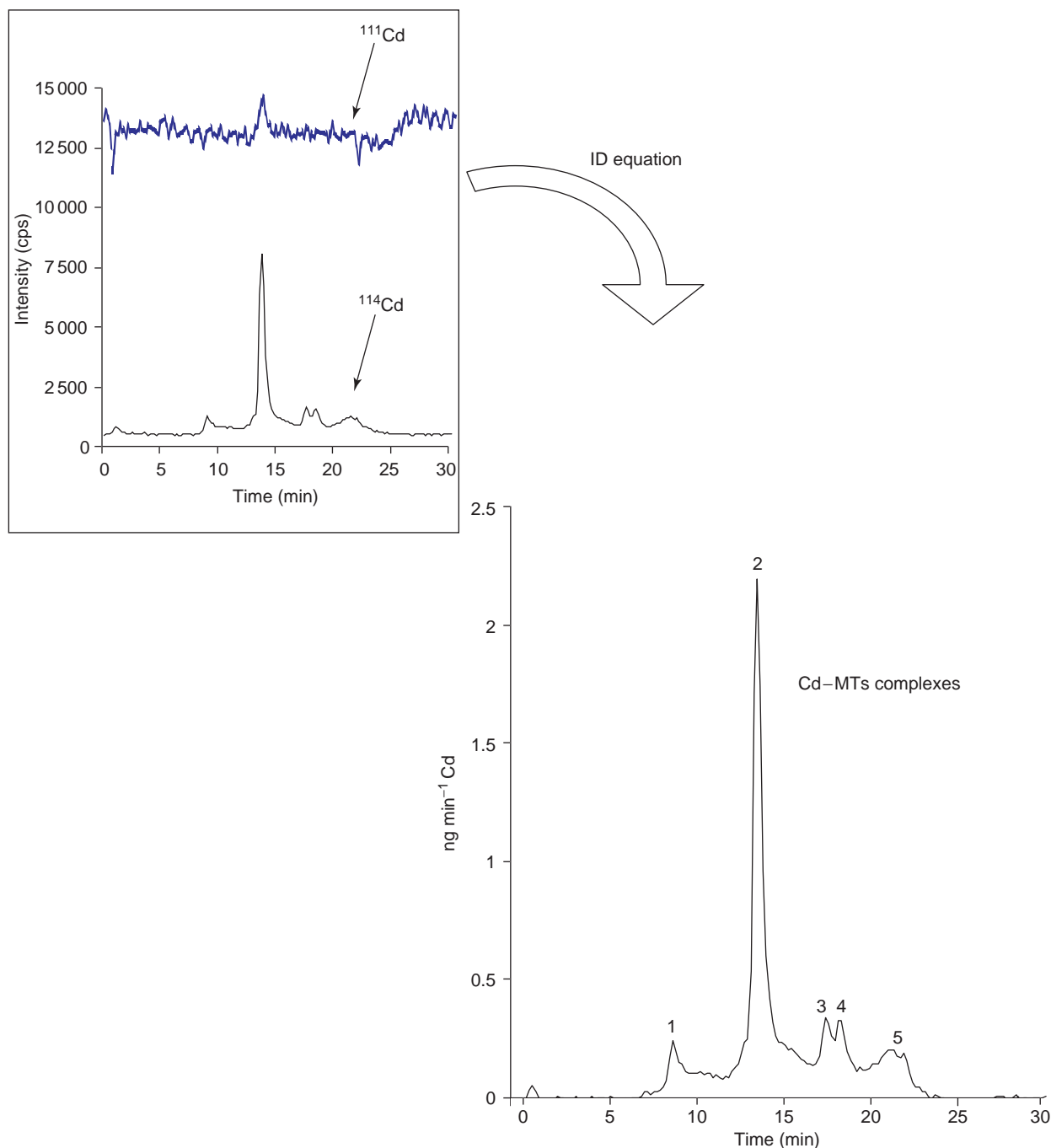
The direct analysis of PC–Cd complexes by ESI-MS/MS from plants extracts is relatively simple and

PC–Cd complexes produced in plants, e.g., in response to metal stress, can be studied.

Hyphenation of CE to ESI-MS seems to be difficult because it requires special coupling interfaces. However, such technology is available at present and, in

fact, some Cd speciation studies have been performed recently.

To conclude this section, it should be pointed out that despite the advantages of using ‘organic’ MS (ESI-MS) for the speciation of Cd in metallated



**Figure 4** Quantification of separated Cd species in fish MTs (real sample) by AE-FPLC–ID-ICP-MS.  $^{111}\text{Cd}$  was used as spike. Inset shows the transformation of intensity signals into mass flow chromatogram (by means of ID equation). The quantification of Cd bound to each metalloproteins (numbered from 1 to 5) could be achieved by this technique. (Rodríguez-Cea A, Fernández de la Campa MR, Blanco González E, Andón B, and Sanz-Medel (2003) Metal speciation analysis in eel (*Anguilla anguilla*) metallothioneins by anionic exchange FPLC-isotope dilution-ICP-MS. *Journal of Analytical Atomic Spectrometry* 18: 1357–1364; reproduced from The Royal Society of Chemistry.)

biocomplexes, there is 100–1000 times lower sensitivity for this detection in comparison to ICP-MS for real sample analysis. This renders molecular MS analysis of MTs or PCs at basal (not induced) levels in real samples yet a rather difficult task.

## Quantification of Cadmium Species by Isotope Dilution ICP-MS

Once a given Cd-bioligand has been identified and its chemical nature established, an accurate and precise determination of such species should be aimed at. The risk of species alteration during the analytical procedures and the relatively few papers published on validation of quantitative speciation results explain that validation of species determinations is still a challenge in real-life situations in speciation analysis. In this sense, isotope dilution (ID) analysis has been accepted as a definitive method for precise and accurate analysis in organic MS for years. This ID concept was extended to trace metals speciation by ICP-MS in environmental samples almost a decade ago. Then, it has been applied to accurate determinations of organotin, organolead, and organomercury compounds. The prerequisite of ID is that the analyte of interest must have at least two stable isotopes. Thus, it is not surprising that this method has also been applied for the quantification of Cd and its species, mainly to quantify the complexes formed between  $\text{Cd}^{2+}$  and MTs.

In ID, the natural isotopic abundance ratio of Cd is altered in the sample by spiking it with an exact and known amount of Cd-enriched isotope (the so-called 'spike', with a different isotopic abundance ratio than natural cadmium). The reference isotope is usually the isotope of highest natural abundance ( $^{114}\text{Cd}$ ), while the spike isotope is one of the lesser abundant natural isotopes (normally  $^{106}\text{Cd}$ ,  $^{116}\text{Cd}$ , or  $^{111}\text{Cd}$ ). As a result of the spiking process, the measurement by ICP-MS of the new isotope ratio (e.g.,  $^{114}\text{Cd}/^{111}\text{Cd}$ ) and its comparison with the natural isotope ratio offers the original Cd concentration in the sample. If the isotope dilution is performed 'online' in an LC- or CZE-ICP-MS experiment, quantification of Cd in each of the isolated species can be accurately achieved by integration of each chromatographic/electrophoretic peak after transformation of the data into 'mass flow' by means of the ID equation.

The quantification of Cd and other metals in MTs offers a great environmental interest in order to understand the role of metals in proteins and the potential use of such metals–MT concentrations as biomarkers of metal pollution. Usually, these kinds of environmental studies involve the analysis of Cd,

Cu, and Zn bound to MTs (such metal–MTs concentrations are interrelated). In this sense, LC-ID-ICP-MS has been applied to the speciation of Cd, Cu, and Zn in liver and kidney of fish under Cd stress. In **Figure 4**, the Cd speciation profile obtained by ID-ICP-MS in fish liver after online separation by anionic-exchange chromatography is illustrated. The ID quantification of Cd in each chromatographic peak provides the metal amount (nanogram) bound to different proteins. As can be seen, the relative amount of Cd (peak area) bound to each protein is considerable, pointing out the importance of Cd speciation studies in investigating bioindicators of environmental contamination.

On the other hand, combination of CE with ID-ICP-MS has also been successfully applied to Cd speciation, but using a double focusing ICP-MS detector. The determination of Cd, Cu, and Zn molar ratios in MTs, using such instruments, which enables the determination of  $^{32}\text{S}$  and  $^{34}\text{S}$  without interferences, has already been described.

**See also:** **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Capillary Electrophoresis:** Overview. **Elemental Speciation:** Overview; Waters, Sediments, and Soils. **Ion-Selective Electrodes:** Water Applications. **Isotope Dilution Analysis.** **Liquid Chromatography:** Size-Exclusion; Liquid Chromatography–Mass Spectrometry; **Mass Spectrometry:** Peptides and Proteins. **Voltammetry:** Overview.

## Further Reading

- Chassaigne H (2003) Electrospray methods for elemental speciation. In: Cornelis R, Caruso J, Crews H, and Heumann K (eds.) *Handbook of Elemental Speciation: Techniques and Methodology*, pp. 357–377. Chichester: Wiley.
- Das AK, de la Guardia M, and Cervera ML (2001) Literature survey of on-line elemental speciation in aqueous solutions. *Talanta* 55: 1–28.
- De Beer D (2000) Potentiometric microsensors for in situ measurements in aquatic environments. In: Buffle J and Horvai G (eds.) *In Situ Monitoring of Aquatic Systems. Chemical Analysis and Speciation*, pp. 161–194. Chichester: Wiley.
- Heumann KG (1990) Elemental species analyses with isotope dilution mass spectrometry. In: Broekaert JAC, Gücer S, and Adams F (eds.) *Metal Speciation in the Environment*, pp. 155–167. Berlin: Springer.
- Lobinski R, Chassaigne H, and Szpunar J (1998) Analysis for metallothioneins using coupled techniques. *Talanta* 46: 271–289.



Prange A and Schaumlöffel D (2002) Hyphenated techniques for the characterization and quantification of metallothionein isoforms. *Analytical and Bioanalytical Chemistry* 373: 441–453.

Sanz-Medel A, Montes-Bayón M, and Fernández Sánchez ML (2003) Trace element speciation by ICP-MS in large

biomolecules and its potential for proteomics. A review. *Analytical and Bioanalytical Chemistry* 377: 236–247.

Suzuki KT (1991) Metallobiochemistry Part B, metallothionein and related molecules. In: Riordan JF and Vallee BL (eds.) *Methods of Enzymology*, vol. 205, pp. 252–262. San Diego: Academic Press.

## CALORIMETRIC SENSORS

See **SENSORS: Calorimetric/Enthalpimetric**

## CAPILLARY ELECTROCHROMATOGRAPHY

**M Macka and P R Haddad**, University of Tasmania, Hobart, TAS, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Electrochromatography was first introduced in 1939. From its discovery it took a further 50 years to attain its modern form, capillary electrochromatography (CEC). This may sound astounding, but it must be realized that not only was it necessary for the theoretical framework to be developed, but also essential technologies underpinning the very existence of CEC needed to emerge. The first of these technologies was the advent of the fused silica capillary in 1979, spurring the development of capillary gas chromatography, and later of capillary electrophoresis (CE) from the 1980s, followed by CEC in the 1990s. Some of the important developments in CEC are summarized in **Table 1**. An evaluation of published literature in CE and CEC (**Figure 1**) reveals that CEC is a relatively recent analytical separation method and is applied much less frequently than CE.

CEC is a combination of liquid chromatography (LC) and CE, combining the advantages of the presence of a chromatographic stationary phase (as occurs in LC) with the separation efficiency achieved in CE as a result of using an applied voltage to drive the separation. Because CEC is a hybrid method, it is prudent to provide an exact definition of what falls under the category of CEC. CEC can be defined as any separation method performed in a liquid phase in the presence of a chromatographic stationary phase in a capillary column, under application of

voltage longitudinally along the column. This means that capillary methods in which both pressure and voltage are applied can be categorized as CEC.

It should be noted that electrophoretic separations can be performed by utilizing interactions with a pseudostationary phase contained within the electrolyte. This pseudostationary phase is so named because it consists of soluble species (or a suspension of small particles), which actually move with the flowing electrolyte. That is, this phase is not stationary. Typical pseudostationary phases include soluble polymers, colloid, micelles, or even particles of a heterogeneous solid chromatographic stationary phase dispersed in the electrolyte. Separations using pseudostationary phases are sometimes categorized as electrochromatography, based on the argument that chromatographic partitioning in the pseudostationary phase is involved. To make a clear distinction, we will here classify as CEC only those separations involving a true stationary chromatographic phase, while separation systems with pseudostationary phases will be categorized as electrokinetic chromatography and will be covered elsewhere. A further point regarding the nomenclature to be used in this article is that the terms electrolyte (rather than mobile phase) and migration time (rather than retention or elution time) will be used, based on the reasoning that electrochromatography always involves the use of separation voltage and therefore is an electromigration separation method.

### Instrumentation

At its current stage of development, where commercial instrumentation from the major manufacturers is

Prange A and Schaumlöffel D (2002) Hyphenated techniques for the characterization and quantification of metallothionein isoforms. *Analytical and Bioanalytical Chemistry* 373: 441–453.

Sanz-Medel A, Montes-Bayón M, and Fernández Sánchez ML (2003) Trace element speciation by ICP-MS in large

biomolecules and its potential for proteomics. A review. *Analytical and Bioanalytical Chemistry* 377: 236–247.

Suzuki KT (1991) Metallobiochemistry Part B, metallothionein and related molecules. In: Riordan JF and Vallee BL (eds.) *Methods of Enzymology*, vol. 205, pp. 252–262. San Diego: Academic Press.

## CALORIMETRIC SENSORS

See **SENSORS: Calorimetric/Enthalpimetric**

## CAPILLARY ELECTROCHROMATOGRAPHY

**M Macka and P R Haddad**, University of Tasmania, Hobart, TAS, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Electrochromatography was first introduced in 1939. From its discovery it took a further 50 years to attain its modern form, capillary electrochromatography (CEC). This may sound astounding, but it must be realized that not only was it necessary for the theoretical framework to be developed, but also essential technologies underpinning the very existence of CEC needed to emerge. The first of these technologies was the advent of the fused silica capillary in 1979, spurring the development of capillary gas chromatography, and later of capillary electrophoresis (CE) from the 1980s, followed by CEC in the 1990s. Some of the important developments in CEC are summarized in **Table 1**. An evaluation of published literature in CE and CEC (**Figure 1**) reveals that CEC is a relatively recent analytical separation method and is applied much less frequently than CE.

CEC is a combination of liquid chromatography (LC) and CE, combining the advantages of the presence of a chromatographic stationary phase (as occurs in LC) with the separation efficiency achieved in CE as a result of using an applied voltage to drive the separation. Because CEC is a hybrid method, it is prudent to provide an exact definition of what falls under the category of CEC. CEC can be defined as any separation method performed in a liquid phase in the presence of a chromatographic stationary phase in a capillary column, under application of

voltage longitudinally along the column. This means that capillary methods in which both pressure and voltage are applied can be categorized as CEC.

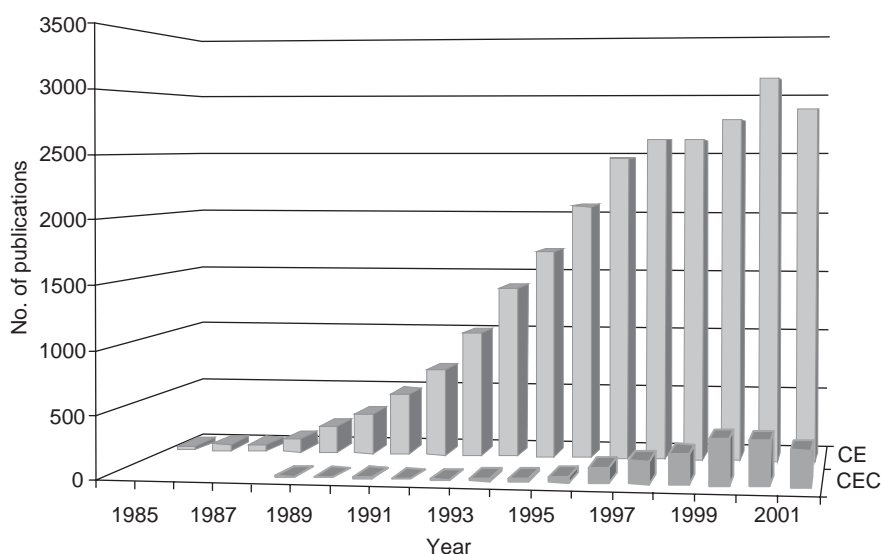
It should be noted that electrophoretic separations can be performed by utilizing interactions with a pseudostationary phase contained within the electrolyte. This pseudostationary phase is so named because it consists of soluble species (or a suspension of small particles), which actually move with the flowing electrolyte. That is, this phase is not stationary. Typical pseudostationary phases include soluble polymers, colloid, micelles, or even particles of a heterogeneous solid chromatographic stationary phase dispersed in the electrolyte. Separations using pseudostationary phases are sometimes categorized as electrochromatography, based on the argument that chromatographic partitioning in the pseudostationary phase is involved. To make a clear distinction, we will here classify as CEC only those separations involving a true stationary chromatographic phase, while separation systems with pseudostationary phases will be categorized as electrokinetic chromatography and will be covered elsewhere. A further point regarding the nomenclature to be used in this article is that the terms electrolyte (rather than mobile phase) and migration time (rather than retention or elution time) will be used, based on the reasoning that electrochromatography always involves the use of separation voltage and therefore is an electromigration separation method.

### Instrumentation

At its current stage of development, where commercial instrumentation from the major manufacturers is

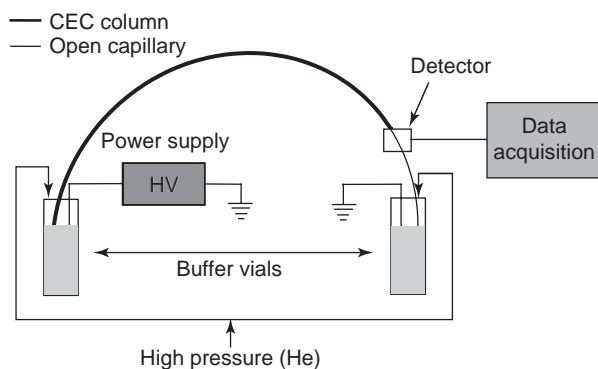
**Table 1** Some important developments in CEC

Time	Development
1939	Electric field used in LC (Strain HH (1939) Combination of electrophoretic and chromatographic adsorption methods. <i>Journal of American Chemical Society</i> 61: 1292–1293)
1943	Term 'electrochromatography' introduced when discussing paper electrophoresis (Berraz G (1943) Electrocapillary analysis. <i>An. Assoc. Quim. Argent</i> 31: 96–97)
1949	Electric field used in transport of compounds through a gel (Shepard CC and Tiselius A (1949) The chromatography of proteins. The effect of salt concentration and pH on the adsorption of proteins on silica gel. <i>Discuss. Faraday Soc</i> 7: 275–285)
1952	Electric field used in TLC (Mould DL and Synge RLM (1952) Electrokinetic ultrafiltration analysis of polysaccharides. A new approach to the chromatography of large molecules. <i>Analyst</i> 77: 964–970)
1974	Electrochromatography in a packed column of 1 mm ID (Pretorius V, Hopkins BJ, and Schieke JD (1974) Electroosmosis. New concept for high speed liquid chromatography. <i>Journal of Chromatography</i> 99: 23–30)
1979	Fused silica capillary technology (Dandenau RD and Zerenner EH (1979) An investigation of glasses for capillary chromatography. <i>HRC&amp;CC</i> 1: 351–356)
1981	CEC in 170 $\mu$ m ID capillary packed with 10 $\mu$ m particles (Jorgenson JW and Lukacs KD (1981) High-resolution separations based on electrophoresis and electroosmosis. <i>Journal of Chromatography</i> 218: 209–216)
1987	CEC in open-tubular and packed columns (Tsuda T (1987) Electrochromatography using high applied voltage. <i>Analytical Chemistry</i> 59: 521–523)
1987, 1991	CEC with small particles, detailed studies (Knox JH and Grant IH (1987) Miniaturization in pressure and electroendosmotically driven liquid chromatography: some theoretical considerations. <i>Chromatographia</i> 24: 135–143; Knox JH and Grant IH (1991) Electrochromatography in packed tubes using 1.5 to 50 $\mu$ m silica gels and ODS bonded silica gels. <i>Chromatographia</i> 32: 317–328)
1991	Hyphenation with MS (pressurised CEC–CF–FAB–MS) (Verheij ER, Tjaden UR, Niessen WMA, and van der Greef J (1991) Pseudo-electrochromatography–mass spectrometry: a new alternative. <i>Journal of Chromatography</i> 554: 339–349)
1995	Rigid polyacrylamide gel as a CEC stationary phase (Hjertén S, Eaker D, Elenbrink K <i>et al.</i> (1995) <i>Japanese Journal of Electrophoresis</i> 39: 105–???)
1997	Rigid macroporous polybutylmetacrylate as a monolithic CEC stationary phase (Peters EC, Petro M, Svec F, and Fréchet JMJ (1997) <i>Analytical Chemistry</i> 69: 3646–3649)

**Figure 1** Numbers of publications for CEC and CE from 1985 to 2002. Source: Chemical Abstracts database accessed through SciFinder Scholar, searched for 'capillary electrophoresis – concept', and 'capillary electrochromatography – concept'.

marketed as hybrid CE/CEC instruments, CEC has very strong similarities to CE in its instrumental requirements. However, there are two main differences between the instrumental implementation of CEC and CE. The first is the presence of some form of

stationary phase inside the separation capillary column (the capillaries used for the CEC columns are usually the same fused silica capillaries as used in CE) and the second is the frequent requirement for overpressure to be applied onto the electrolyte



**Figure 2** Schematic representation of a CEC instrument.

reservoirs (at both injection and detection ends of the capillary) during the separation to help prevent bubble formation. A schematic representation of such a CEC instrument is shown in **Figure 2**. It should be noted that the current situation with commercial CE instruments being used for CEC does not access the full potential of CEC, especially the use of gradient elution analysis. A  $\mu$ -LC/nano-LC pump capable of pumping the electrolyte through the column at defined flow rates would bring significant advantages for usage of pressure-assisted CEC, flushing, and conditioning of CEC columns, or for loading of sample solution onto the CEC column. Therefore, the need for a hybrid gradient  $\mu$ -LC/nano-LC-CEC-CE instrument seems to be clearly evident. Such instruments have been constructed by several research groups but have not been introduced by any of the major commercial CE/CEC instrumentation manufacturers.

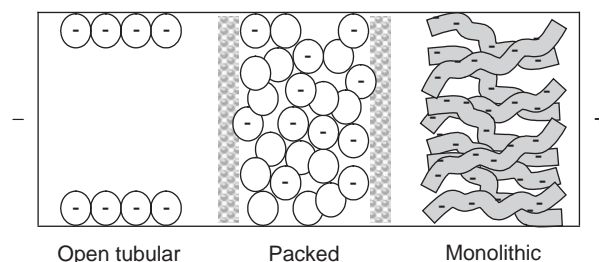
### Sample Introduction

As in CE, hydrodynamic or electrokinetic (electromigration) sampling is used, i.e., by applied pressure or voltage.

Electrokinetic sample introduction is known to be matrix dependent; therefore, hydrodynamic sample introduction is normally preferred. If electromigration injection is to be employed, matrix effects should be investigated and standard addition calibration rather than external standard calibration should be used.

### CEC Columns

The CEC columns have similarities to capillary columns used in  $\mu$ -LC/nano-LC, but there are several important differences. First, in CEC, the separation current resulting from the applied separation potential produces heat that must be dissipated effectively, otherwise bubble formation or even drying out of



**Figure 3** Modes of CEC with respect to the column type. From left to right illustrated schematically are: open-tubular columns, packed columns, and monolithic columns.

portions of the CEC column bed can result. Therefore, the inner diameter of CEC fused silica capillary columns usually does not exceed 100  $\mu$ m and an efficient capillary cooling system is essential. Additionally, electrolyte vials are often overpressurized (using helium or nitrogen at  $\sim 10$  bar) to further prevent bubble formation. Second, the nature of the electroosmotic flow (EOF) driving the flow of electrolyte through the CEC column allows the use of very small packing particles (even with diameters  $< 1 \mu$ m). There are three main types of CEC columns, namely open-tubular (OT) columns, packed columns, and monolithic columns, as illustrated schematically in **Figure 3**. These modes of CEC are described in more detail in the section 'Modes of CEC'.

### Detection

The most common detection method used in CEC is direct photometric detection, which offers the same advantages and disadvantages as its use in CE. Its strengths include simplicity, robustness, and widespread applicability to most analytes, while its lack of structural information for identification of analytes and moderate detection sensitivity are its drawbacks. Using direct photometric detection, a direct comparison of detection sensitivity between CEC and high-performance liquid chromatography (HPLC) for the same set of analytes yielded a 100-fold poorer method sensitivity for CEC, which is not surprising given that a typical path length in CEC using on-capillary detection (50–100  $\mu$ m) is about two orders of magnitude smaller than a typical detector cell path length in HPLC (5–10 mm).

Other detection methods can be applied to CEC. Indirect spectrophotometric detection methods are commonly used for ultraviolet (UV)-transparent ionic analytes (such as inorganic ions) in CE, but indirect detection in CEC tends to suffer from unstable baselines, caused by adsorption of the indirect detection probe onto the CEC ion-exchange stationary phases (SP). Fluorimetric detection is a very sensitive and robust method, but is limited to a relatively small

number of analytes. Electrochemical detection methods, such as conductometry, amperometry, and potentiometry, can be utilized in a manner similar to CE, and the same principles, strengths, and weaknesses apply. Amperometric detection would be the detection method of choice for analytes that can be easily reduced or oxidized, but the range of suitable analytes is rather limited. On-capillary contactless conductivity detection can be used for ionic analytes exhibiting different mobility to that of the electrolyte coion, but again its performance in CEC tends to be inferior compared to CE.

The lack of reliably performing indirect detection in CEC, and in fact of any reliable, sensitive, and universal detection method, makes the need for an alternative universal detection method even more obvious than in CE. Various mass spectrometric (MS) detection methods hyphenated with the CEC separation column can fulfill this task. In the future, MS detection methods are likely to further grow in popularity, especially with continuing improvements in the MS instrumentation and its affordability.

## Principles and Practical Considerations

### Conductivity of CEC Columns

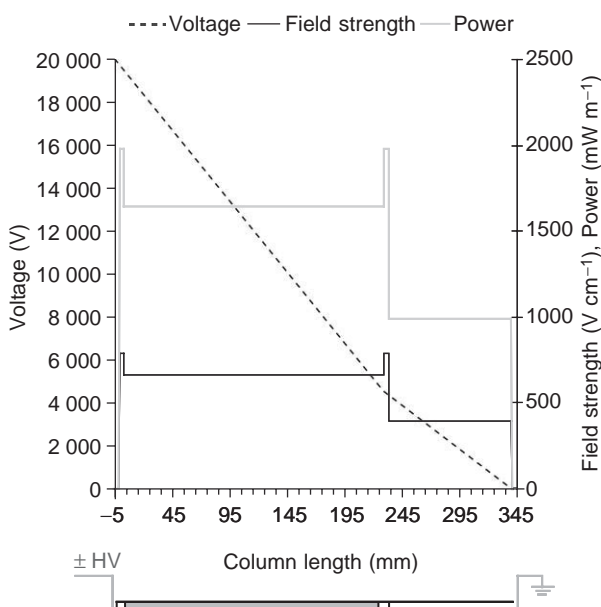
The conductivity of the stationary phases in CEC is considered negligible, as is the contribution of EOF to the overall conductivity through a CEC column. Therefore, under typical conditions it is solely the electrolyte in the pores of the CEC column that is responsible for conducting the current. Unlike in CE, where the conductivity through the capillary filled with electrolyte is simply given by the conductivity of the electrolyte filling the whole cross-section of the capillary, the conductivity of a CEC column will be considerably smaller due to the presence of nonconductive packing in the CEC column. The 'conductivity ratio',  $\Phi$ , can be defined as the ratio of the conductivities of a CEC capillary column and of an open capillary of same dimensions. Using the Boyack–Giddings equation,  $\Phi$  can be related to the interstitial porosity of the packing and its morphological characteristics:

$$\Phi = \frac{\sigma_{\text{packed}}}{\sigma_{\text{open}}} = \varepsilon_i O T^{-2} \quad [1]$$

where  $\sigma_{\text{packed}}$  is the conductivity of the CEC column,  $\sigma_{\text{open}}$  the conductivity of the CE open capillary,  $\varepsilon_i$  the interstitial porosity,  $O$  the constrictive factor, and  $T$  the tortuosity. The last two experimental parameters are associated with inconsistent channel cross-sectional area and with the fact that the channels and

the electric field vectors generally do not run parallel. It should be noted that interstitial porosity in LC is termed 'total porosity' and is simply the fraction of the void volume relative to the volume of the empty column, thus corresponding to the volume between the particles of the packing plus the volume of all accessible pores. Typical values in HPLC fall in the range 0.25–0.4. The values of conductivity ratios in CEC fall within the range of  $\sim 0.3$ –0.7, with CEC columns packed with pellicular chromatographic particles being at the lower end of the range, followed by typical 300 Å silica packings, and packings of macroporous particles and monolithic packings giving conductivity ratios at the upper end of the range.

The conductivity of the CEC column is also important when considering the profile of heat generation and dissipation that is to be expected along the CEC column. Unlike in CE, where the whole capillary is filled with a homogeneous electrolyte exhibiting a uniform conductivity, such that the electric field and the generated heat will also be uniform along the capillary length, in CEC the conductivity will be different for the open capillary section, the frits, and the column bed (and also can vary along the column bed if it is inhomogeneous). This situation is illustrated in Figure 4 for a hypothetical CEC packed capillary column held between two frits and



**Figure 4** Simulated heat-generation profile calculated along a typical packed CEC capillary column with the packing held between two frits, with an empty capillary section from the detection side frit. Parameters – total capillary length: 0.340 m; CEC column length including frits: 0.237 m; length of inlet and outlet frits: 4 mm; porosity of frits: 0.5; porosity of stationary phase: 0.6; voltage: 20 kV; current: 25  $\mu$ A.

with an open capillary section after the frit on the detection end of the capillary. In this example, the porosity of the frits was somewhat lower than the packed bed (which is a realistic scenario for frits prepared by sintering the same material as in the packed section). The graph clearly shows that the calculated profile of the electric field and the generated heat is nonuniform, and the system is likely to exhibit 'hot-spots' at the frits. This can cause bubble formation with deleterious consequences for the CEC analysis. Apart from overheating, other factors attributed to bubble formation are nonuniform EOF resulting in pressure differences at the interfaces between the different sections of the CEC capillary column and nonuniform surface tension along the CEC capillary column.

### Electroosmotic Flow

The principles of EOF and its role in CEC separations are the same as for CE.

The presence of EOF as the source of flow of the electrolyte through the CEC column has two very profound effects. First, by the nature of the EOF, the pressure difference across the whole CEC column bed is zero. This means that much smaller particle sizes can be used, down to  $\sim 1 \mu\text{m}$  particle diameter, with 3–5  $\mu\text{m}$  particles being used routinely. Second, the plug-like profile of the EOF is responsible for the typical higher separation efficiencies achieved in CEC compared to LC.

EOF arises due to the presence of charges on the inner surface of the capillary wall, and in CEC also on the surface of the CEC stationary phase, which give rise to the presence of a charged double-layer of a fixed (Stern) layer and a mobile, diffuse (Gouy–Chapman) layer. In the presence of an electric field directed longitudinally relative to the charged surfaces (i.e., where high voltage is applied at the ends of the CEC capillary column), the charges in the diffuse layer will move in the electric field along a shear plane between the fixed and diffuse layers, dragging the bulk liquid. Locally at the packing surface, the electroosmotic mobility is expressed by the Smoluchowski equation:

$$\mu_{\text{CEC,local}} = -\frac{\varepsilon\varepsilon_0\zeta}{\eta} \quad [2]$$

where  $\zeta$  is the zeta potential of the surface of the CEC stationary phase,  $\eta$  the viscosity of the electrolyte,  $\varepsilon$  the dielectric constant, and  $\varepsilon_0$  the dielectric constant *in vacuo*.

Alternatively, when expressing the zeta potential as a function of the charge density at the shear surface and of the double-layer thickness, and substituting

for the double-layer thickness from the Debye–Hückel approximation, one arrives at the Hunter equation:

$$\mu_{\text{CEC,local}} = \frac{\sigma[\varepsilon\varepsilon_0RT/2cF^2]^{1/2}}{\eta} \quad [3]$$

where  $\sigma$  is the charge density at the shear surface,  $R$  the gas constant,  $T$  the temperature,  $c$  the electrolyte concentration, and  $F$  the Faraday constant. Importantly, this equation shows that the EOF decreases with electrolyte concentration (in the absence of double-layer overlap which occurs in very narrow channels and/or at low electrolyte concentrations).

An overall (average) electroosmotic mobility over the whole CEC column bed takes into account the conductivity ratio from eqn [1]:

$$\mu_{\text{CEC}} = \mu_{\text{CEC,local}} \left( \frac{\sigma_{\text{packed}}}{\sigma_{\text{open}}} \right) \quad [4]$$

This shows that the EOF in CEC increases with increasing porosity of the stationary phase bed.

Another important conclusion can be drawn from the fact that packing particle diameter does not appear in the equations for EOF in CEC. Unlike in LC, in CEC the packing particle diameter does not exert any direct major influence on the flow of the electrolyte through the CEC column bed. Therefore, smaller packing particles can be used, often below 1  $\mu\text{m}$ . The minimum applicable particle size is limited by double-layer overlap occurring typically in channels below 1  $\mu\text{m}$ . This effect is also important in perfusive packings, which are macroporous packings with pores large enough (usually approaching 1000 Å or larger) to support perfusive (intraparticulate) EOF.

Not only is the EOF a fairly complex phenomenon in itself, but another complexity arises from the fact that surfaces inside the CEC column can exhibit substantially different EOF (at the capillary wall surface, frit surface, SP surface, or at different points of the stationary phase, such as in different pores), and that the surfaces inside the CEC column follow a complex morphology. These factors can lead to very complex flow patterns inside the CEC column. The main contributions to the bulk flow through the column are the EOF-induced flow through the chromatographic bed (and the frits, if present), and the EOF of the capillary wall in the open section. The EOF through the chromatographic bed can act as an 'EOF pump', inducing pressure-driven flow in the open capillary section, since the flow resulting from the CEC bed will be generally different to the EOF in the open section.

### Efficiency

The smaller particle size used in CEC compared to LC, together with the beneficial effects of EOF on the 'A' and 'C' terms of the van Deemter equation, result in separation efficiencies for CEC which exceed those of LC by a factor of  $\sim 2$ – $5$ . Efficiencies of  $\sim 200\,000$ – $500\,000$  theoretical plates per meter are typical for CEC. The two main beneficial effects of EOF, namely a more uniform flow profile across the column bed, and an enhanced mass transfer, are explained in more detail in the paragraphs below.

In the CEC bed, the linear flow velocity of the EOF through channels of different diameters will be independent of the channel diameter, provided the channel is wide enough to avoid double-layer overlap (under typical experimental conditions the critical channel diameter is  $\sim 0.1\,\mu\text{m}$ ). This significantly reduces (usually by a factor of  $2$ – $5$  times) the 'A' term in the van Deemter equation, also referred to as 'Eddy diffusion', and is primarily responsible for the significantly higher separation efficiencies achieved on a capillary column operated in the CEC mode compared to the same column operated in LC mode (pressure-driven).

Another effect contributing to higher separation efficiencies in CEC compared to LC is thought to arise from improved mass transfer (the 'C' term in the van Deemter equation, which is usually decreased by a factor of  $1.5$ – $5$  times). This results from mixing effects of vortices created by nonuniform EOF of the pore wall and physical obstructions in the flow path such as narrowing of the pore or even the presence of dead-end pores.

### Retention and Selectivity

The overall speed with which an analyte moves through the CEC column bed is given by a combination of both electrophoretic (CE) and chromatographic (LC) factors. Rathore and Horvath defined electrochromatographic retention factor as

$$k_{\text{CEC}} = k_{\text{LC}} + k_{\text{LC}}k_{\text{CE}} + k_{\text{CE}} \quad [5]$$

where  $k_{\text{LC}}$  is chromatographic retention factor and  $k_{\text{CE}}$  the electrophoretic velocity factor defined as a ratio of electrophoretic mobility of an analyte to electroosmotic mobility. It should be noted that for a neutral solute,  $k_{\text{CE}} = 0$  and  $k_{\text{CEC}}$  will revert to an expression for chromatographic retention factor. The disadvantage of this approach is that the electrochromatographic retention factor  $k_{\text{CEC}}$  cannot be determined from CEC data alone and an independent measurement of the CE contribution is needed. Therefore, for practical purposes, the position of

the analyte peak in the CEC electrochromatogram is characterized using a dimensionless retention factor, often expressed as

$$k'_{\text{CEC}} = \frac{t_{\text{m}} - t_0}{t_0} \quad [6]$$

where  $t_{\text{m}}$  and  $t_0$  are the migration times of the analyte and an inert neutral marker in a CEC column. It must be stressed that this retention factor is devoid of any insight into the mechanism of retention. Its disadvantage is that for cationic analytes migrating faster than the EOF it is possible that  $t_{\text{m}} < t_0$  and in that case  $k'_{\text{CEC}} < 0$  which might appear to be irrational by comparison with the chromatographic retention factor.

As described by eqn [5], the overall separation selectivity in CEC will be composed by contributions from both the LC and the CE mechanisms. Especially when developing a CEC method based on an existing LC method, it is important to realize that the CEC separation selectivity will always differ to some degree from that of the parent LC method. For uncharged analytes (usually separated in the reversed-phase CEC mode), the CEC separation selectivity should in principle closely resemble that of the LC separation using the same packing. However, in practice CEC retention factors tend to be larger by  $\sim 20\%$  compared to  $\mu$ -LC under the same conditions, and this discrepancy is attributed to possible influences of the electric field in CEC on the chromatographic stationary phase and the distribution constants of the analytes. Even more importantly, for any charged analyte, the overall CEC separation selectivity will also include a CE component based on the migration of the charged analytes. Thus, for instance, in a CEC separation on a reversed-phase stationary phase conducted in the usual arrangement with positive voltage polarity at the injection end and cathodic EOF driving the electrolyte toward the detector side, in the absence of ion-exchange interactions between the analyte and the stationary phase, the migration time of any cationic analyte will be smaller compared to the LC mode, and that of any anionic analyte will be larger.

### Modes of CEC

CEC columns can be categorized with respect to the type of stationary phase as OT columns, packed columns, and monolithic columns, as illustrated in Figure 3. It should be noted that a particular CEC stationary phase type can be realized in a fused silica capillary, or in a channel of a microfluidic chip.



## Open-Tubular Columns

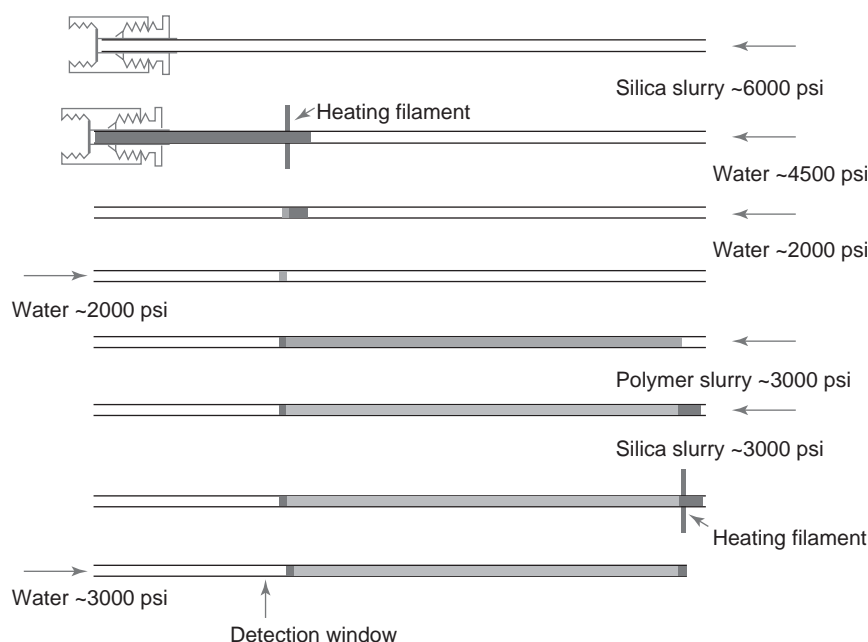
OT columns in CEC are analogous to OT columns in gas chromatography. A significant practical advantage of OT columns, especially compared to packed column CEC, is the ease with which the electrolyte can be flushed through the OT column using a moderate applied pressure.

OT CEC columns are relatively easy to prepare, with the stationary phase being bound onto the capillary wall using well-established procedures. For example, reversed-phase OT stationary phases can be formed by attaching hydrophobic C18 groups using the same chemical reactions as applied in the synthesis of reversed-phase LC stationary phases. The principal problem of OT columns is the relatively low ratio of stationary to mobile phase, resulting in low retention and capacity of the OT columns. This can be partially addressed by increasing the amount of stationary phase attached onto the capillary wall, usually by etching the capillary wall before attaching the stationary phase. The limit of this approach is the danger of decreasing separation efficiency by increasing the resistance to mass transfer (the 'C' term in the van Deemter equation). Another problem is the necessarily small capillary diameter required because of the low diffusion coefficients of analytes in liquids. These factors cause the capillary diameters of the OT columns to be generally smaller compared to packed and monolithic CEC columns, with capillary diameters of 25  $\mu\text{m}$  ID or less being typical. As a result, this places demands on detection methods, with the

sensitivity of photometric detection being typically unsatisfactory for these small ID columns.

## Packed Columns

Packed column CEC is a logical extension of previous research in  $\mu\text{-LC}$ /nano-LC systems conducted through the 1970s and the 1980s. The most common procedure for preparation of packed CEC columns is slurry packing under high pressure. A typical example of this procedure is shown in **Figure 5**, which illustrates the steps required to pack a bed of polymeric stationary phase, retained by two frits composed of silica. In the top frame, an aqueous slurry of silica particles is pressure-packed into the end-section of the column where the detection end frit is to be made. The frit is then made by sintering a short portion of the silica packing with a heating filament. The unsintered silica packing remaining on both sides of the formed frit is then flushed out of the capillary. A slurry of the desired chromatographic packing (in this example a polymer-based packing material) is then packed from the injection end of the column until only a few millimeters of unpacked capillary remain. This portion of the capillary is then packed with more of the silica packing and the injection end frit is made using the same sintering procedure as for the detection end frit. Finally the capillary is cut such that the injection end frit forms the front (injection end) of the CEC column. Other CEC column packing procedures have been developed for special cases, such as electrokinetic packing



**Figure 5** Packing procedure for CEC columns with sintered silica frits and a polymer particulate packing placed between the frits. See the text for explanation. (Reproduced with permission from Hilder EF (2000) Ph.D. thesis, University of Tasmania.)

for packing CEC sections of a channel on a micro-chip.

Probably the greatest advantage of packed CEC columns is the fact that a wealth of existing LC stationary phases can be used, with reversed-phase materials being the most commonly applied, followed by ion-exchange stationary phases. It could be argued that the very well-defined properties and known selectivity of these stationary phases should make method development easier in CEC. However, two factors that complicate the translation of an LC method into the CEC format must be kept in mind. First, as discussed earlier, the CEC separation selectivity for both uncharged and even more for charged analytes will differ to that of  $\mu$ -LC/nano-LC under identical conditions. Second, specialty CEC stationary phases which differ from typical LC stationary phases are often used, for instance those containing sulfonic acid groups which remain fully ionized over the available pH range and will therefore support a strong and pH-independent EOF through the CEC column bed. An example of this type of stationary phase is a reversed-phase material derivatized with alkylsulfonate chains, or with a mixture of alkyl and alkylsulfonate chains.

The miniaturization of the packed columns and especially of the frits makes the preparation of packed CEC columns highly technologically demanding. The columns are difficult to reproduce in terms of parameters such as the EOF, and they are prone to breakage, frit failure, blockage, and the formation of voids. Despite concentrated efforts of many research teams into various alternative frit making and column packing procedures, these problems have not been addressed satisfactorily, leaving an acute demand for an alternative column technology. This demand seems to have been satisfied by the advent of monolithic capillary columns, which are described in the following section. From this perspective, packed columns have been largely replaced by the monolithic capillary columns. Some limited use of packed columns in the future will probably be justified in cases where a specific existing chromatographic stationary phase packing has to be used because of its particular chromatographic properties.

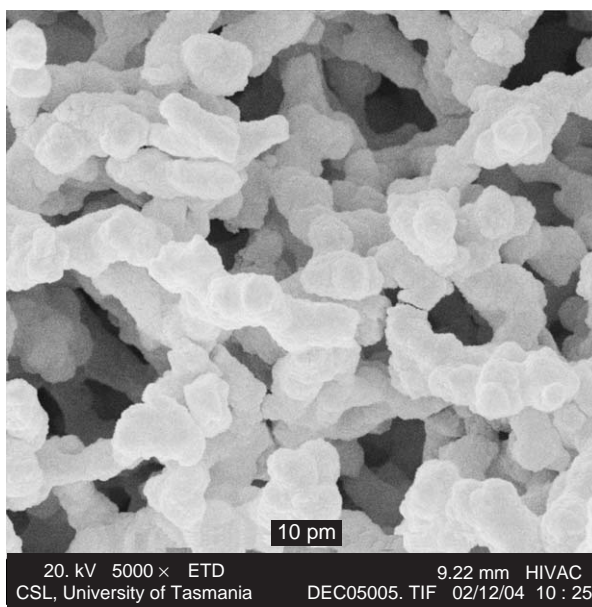
### Monolithic Columns

Monolithic, or continuous bed, columns for LC and CEC are relatively new, but the combination of high capacity, very low pressure resistance, simple *in situ* preparation procedures, and no requirement for frits has made them a very popular choice both in LC and in CEC.

Monolithic CEC columns can be formed from organic porous polymeric monoliths or porous silica sol-gel monoliths. Other monolithic CEC column types, such as immobilized particles, have also been demonstrated but have not shown the potential of the first two types. The monolithic stationary phases are created *in situ* by polymerization reactions under controlled conditions, such that a porous bed is created.

Silica monoliths are prepared by sol-gel technology using hydrolysis and polycondensation reactions. Functionalities for hydrophobic interactions with analytes can be introduced in a subsequent treatment of the bed with alkylsilanes, or by introducing hydrophobic functionalities in the monomer mixture (Figure 6).

For organic polymers, a mixture of monomers and an organic solvent acting as a porogen is polymerized using free radical UV-light or thermal initiation. Clearly, these procedures have to be optimized carefully, but once that is done, the process is claimed to be reproducible. The chromatographic selectivity of the resultant monolith depends on the monomers used and can be altered by attachment of additional functionalities, such as alkyl chains for hydrophobic interactions or charged groups for ion-exchange interactions. Typical polymers used are acrylate, methacrylate, polystyrene, and various copolymers. EOF is provided by charged groups, usually sulfonate



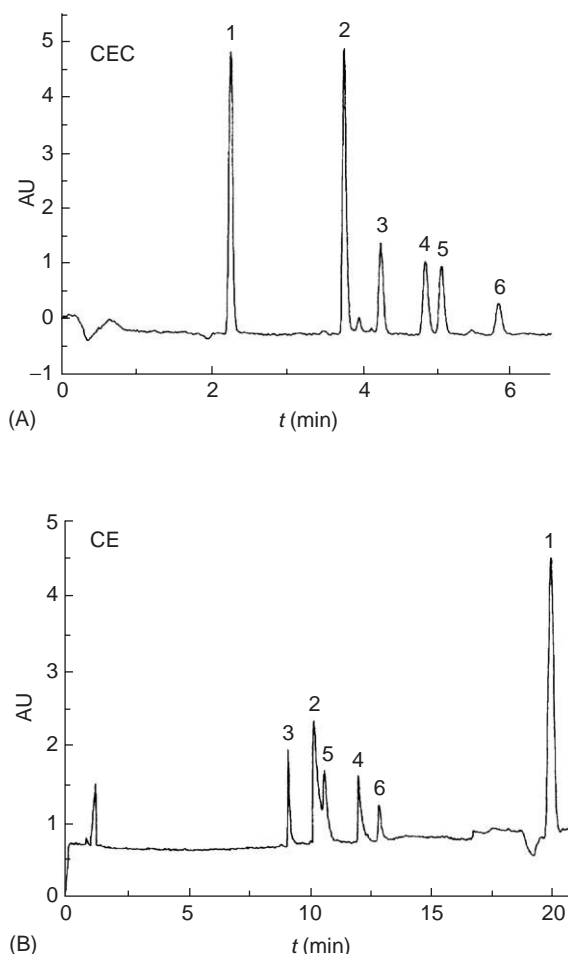
**Figure 6** SEM of silica monolithic CEC column inside a fused silica capillary. Conditions – fused silica ID: 75  $\mu$ m; the monolith was prepared by the procedure described by Ishizuka N, Kobayashi H, Minakuchi H, *et al.* (2002) *Journal of Chromatography A* 960: 85–96. (Courtesy of Mr. J. Hutchinson.)

groups for cathodic EOF, or quaternary ammonium groups for anodic EOF. A special category of stationary phases is molecular imprinted polymers used for enantioselective separations, where one isomer is 'imprinted' into the stationary phase during its polymerization, thus providing molecular-size cavities where this isomer will be specifically retained. A remarkably powerful technique when using reactions initiated by UV-light is photopatterning, where different sections of the capillary are covered with a mask, and the uncovered section is exposed to UV-light to induce photopolymerization reactions. Different sections of the CEC capillary column can therefore be filled with monolithic materials having different functionalities. This can provide powerful separation tools especially in the microchip format.

## Application Potential

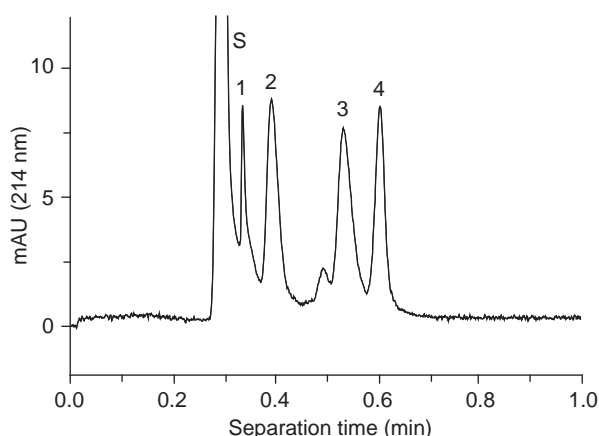
During the past decade many sample types have been investigated using CEC in order to assess its analytical potential. Because CEC carries an additional degree of difficulty compared especially to CE (namely the presence of a stationary phase), it has been suggested that there should be a clear advantage of CEC, such as higher separation efficiency or shorter analysis time, to justify why CEC should be used in preference to CE. The number of cases where such an advantage could be demonstrated clearly would be small. However, taking into account all the advantages offered by the presence of a chromatographic stationary phase (chromatographic gradient separation, sample preconcentration from volumes beyond one capillary volume, and chromatographic peak focusing), it should be recognized that CEC has very significant analytical application potential. It is to be expected that monolithic column technology will prove effective in providing CEC methods, which are competitive with the main alternatives – CE and  $\mu$ -LC/nano-LC.

The examples of CEC separations shown in Figures 7–9 arise from the above considerations: most importantly, in all cases they are efficient and rapid separations – Figure 7 in 6 min and Figures 8 and 9 in less than 1 min. Figure 7 shows a separation of several peptides on a strong cation-exchange stationary phase in the CEC mode (A), which is superior in both separation efficiency and separation time to CE (B). It would be fair to point out that use of a coated capillary containing sulfonate groups in the CE separation would provide a stronger EOF at the low pH used (3.0) and the separation would be faster, but peaks 2 and 5 then would not be separated.



**Figure 7** Separation of peptides by ion-exchange CEC (A) and by CE (B). Conditions: (A) 75  $\mu$ m ID capillary column, total length 0.310 m, column bed 0.010 m packed with Spherisorb-SCX 5  $\mu$ m; temperature, 20°C; voltage, 15 kV; electrolyte, 60% acetonitrile in 30 mmol L<sup>-1</sup> potassium phosphate buffer (pH 3.0); overpressure in both vials, 6.9 bar; detection, photometric at 200 nm; injection, electrokinetic, 5 kV for 10 s; (B) 50  $\mu$ m ID capillary, total length 0.502 m, length to detector 0.400 m; voltage, 25 kV; injection, hydrodynamic, 0.035 bar for 5 s. Peaks: 1 = benzylalcohol; 2 = Gly-Thr; 3 = Gly-Ala-Gly; 4 = Glu-Glu; 5 = Gly-Gly-Asn-Ala; 6 = Glu-Glu-Glu. (Reproduced with permission from Ye M, Zou H, Liu Z, and Ni J (2000) Separation of peptides by strong cation-exchange capillary electrochromatography. *Journal of Chromatography A* 869: 385–394.)

Figure 8 again shows a separation of several peptides on a cation-exchange stationary phase, in this case a monolithic material. A short column bed, a high separation voltage, and also probably a high EOF contribute to a short separation time of less than 1 min. Figure 9 shows a reversed-phase separation of labeled peptides on a monolith used in a microchip format, with separation time again less than 1 min. The last two examples illustrate the potential of monolithic CEC columns and of CEC in the microchip format.

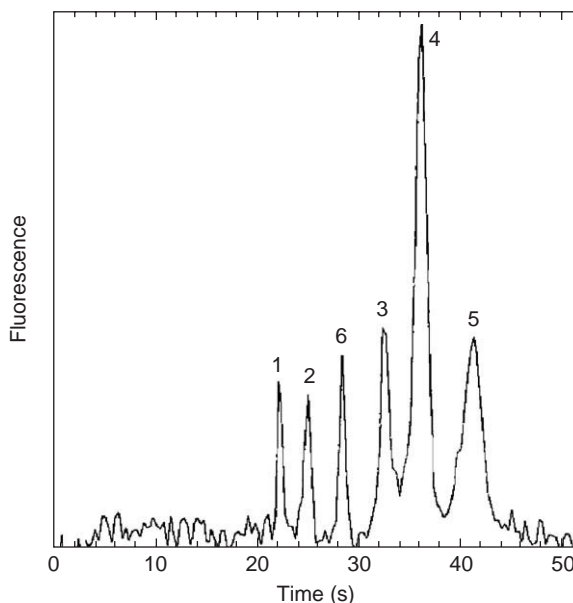


**Figure 8** Separation of peptides using monolithic capillary grafted with 2-acrylamido-2-methyl-1-propanesulfonic acid. Conditions: 100  $\mu\text{m}$  ID capillary column, total length 0.345 m, monolith 0.085 m, 30 s grafting; temperature, 60°C; voltage, 15 kV; electrolyte, 100  $\text{mmol l}^{-1}$  NaCl solution in 10  $\text{mmol l}^{-1}$  phosphate buffer, pH 6.0; overpressure in both vials, 0.8 MPa; detection, photometric at 214 nm; injection, pressure-driven, 0.8 MPa for 0.05 min; sample, 0.1  $\text{mg ml}^{-1}$ . Peaks: S = system peak; 1 = Gly-Tyr; 2 = Val-Tyr-Val; 3 = methionine enkephalin; 4 = Ileucine enkephalin. (Reproduced with permission from Hilder EF, Svec F, and Frechet MJM (2002) Polymeric monolithic stationary phases for capillary electrochromatography. *Electrophoresis* 23: 3934–3953.)

### Separation Strategies and Experimental Variables

In a typical CEC arrangement, a positive separation voltage is applied to the inlet of a CEC column exhibiting a cathodic EOF, which means that the EOF flows toward the detector. Any unretained cation (exhibiting no interaction with the stationary phase) will migrate before the EOF, and any unretained anion will migrate after the EOF. Because of the experimental similarity of CEC to CE, the major experimental parameters that can be optimized are the electrolyte composition, separation voltage, and temperature. Further parameters, which are specific to CEC are the stationary phase and dimensions of the column.

In a major difference to LC, where the column temperature can be varied over a wide range from below room temperature to close to the boiling point of the most volatile component of the mobile phase, in CEC the column temperature has to be kept low (close to room temperature or below) to minimize the problems associated with heat dissipation (bubble formation). As the EOF exhibits a linear dependence on the separation voltage, the separation voltage is normally kept at the maximum value determined by the avoidance of column overheating. Typically, this maximum voltage available is usually  $\pm 30\,000\text{ V}$ .



**Figure 9** Separation of NDA labeled peptides by reversed-phase CEC in a microchip with negatively charged lauryl acrylate monolithic stationary phase. Conditions: glass microchip, channels 40  $\mu\text{m}$  deep and 120  $\mu\text{m}$  wide, offset double-T design, main channel length 70 mm, length from the junction to waste 60 mm; stationary phase, porous acrylate monolith containing sulfonic groups, all the channels were filled with the monolith to avoid EOF mismatch; voltage, 5 kV (field strength 770  $\text{V cm}^{-1}$ ); electrolyte, 30:70, acetonitrile/25  $\text{mmol l}^{-1}$  borate, pH 8.2, containing 10  $\text{mmol l}^{-1}$  octane sulfonate; detection, fluorimetric (LIF) using 413 nm line of Kr ion laser, at  $\sim 50\text{ mm}$  from the junction; sample, peptides were labeled with naphthalene-2,3-dicarboxaldehyde (NDA). Peaks: 1 = papain inhibitor, 2 = proctolin, 3 = Opioid peptide ( $\alpha$ -casein fragment 90–95), 4 = Ileangiotensin III, 5 = angiotensin III, and 6 = GGG. (Reproduced with permission from Throckmorton DJ, Shepodd TJ, and Singh AK (2002) Electrochromatography in microchips: reversed-phase separation of peptides and amino acids using photopatterned rigid polymer monoliths. *Analytical Chemistry* 74: 784–789.)

Once the column type and stationary phase have been chosen, and the column temperature and separation voltage determined as explained above, the remaining operational parameters that have to be selected relate to electrolyte composition and include electrolyte concentration (ionic strength), pH, organic solvent type and content, and various additives. As in LC and CE, control of the composition of the electrolyte is the most powerful and straightforward tool for governing separation selectivity. As for CE, increasing ionic strength will decrease the EOF. However, unlike in CE, very small channels (less than 1  $\mu\text{m}$ ) are often present in a CEC column, so that double-layer overlap may come into effect especially in low ionic strength electrolytes (see the section ‘Electroosmotic flow’). In such cases, an opposite dependence of EOF on the ionic strength

can be seen. Therefore, the best overall performance of a CEC column in terms of both maximal EOF and separation efficiency can usually be achieved at moderate electrolyte concentrations, usually in the range between 10 and 50 mol l<sup>-1</sup>.

The choice of pH will be dependent primarily on the acid–base properties of the analytes and the charge they acquire at a given pH. Usage of CEC stationary phases containing permanently charged functional groups is helpful in optimizing the pH because the EOF will be relatively independent of pH, so the electrolyte pH can be varied in order to achieve the desired charges on the analytes. The choice of organic solvent again relates to principles well known from LC, for example, in reversed-phase separations where the organic solvent content governs the chromatographic interaction with the stationary phase. The content of organic solvent in the electrolyte will also influence the EOF, the heat generated in the column, and the magnitude of the sample focusing effect upon sample injection. However, these effects can be fairly complex and are beyond the scope of this article. This also applies to the optimization of any other parameters, such as the use of other electrolyte additives, for example, surfactants, enantioselective ligands, etc.

## CEC Used for Sample Preconcentration

In terms of the analytical potential of CEC for real samples, one of the biggest potential advantage of CEC is the presence of a chromatographic stationary phase. This can be highly beneficial for applying sample preconcentration and cleanup and peak focusing on the stationary phase. It has to be noted that the sample preconcentration and cleanup step can be performed in the LC mode since pressure-driven flow allows these operations to be performed at greater speed than flow driven by EOF. As a result of the fact that CEC is a relatively young method, sample preconcentration methods have yet to be fully applied and developed. Monolithic columns again seem to be ideally suited for this purpose as they combine a high sample capacity with high porosity, low pressure resistance, and low resistance to mass transfer, consequently allowing rapid sample preconcentration.

## CEC in Microfluidic Chip Format

The concept of micrototal analytical systems ( $\mu$ TAS) in the late 1980s and the development of the microfluidic chip (also called a microchip) from the early 1990s presented one of the most dynamic and exciting

developments in analytical chemistry. Separation methods realized in the chip format have so far concentrated mainly on CE, as voltage-driven microchip devices are instrumentally more simple than pressure-driven devices containing pumps, valves, etc. Therefore, microchip CEC is preferable to microchip LC. However, the development of CEC on microchips has been hampered by the practical difficulties encountered in packing a CEC column in a microchip. The development of monolithic CEC column technology has provided a viable alternative to conventional packing methods and it can be expected that CEC on microchips using monolithic columns will develop into a powerful separation technique.

## Emerging Instrumentation

There is a need for commercial hybrid capillary-LC/CEC/CE instrument equipped with pumps and gradient solvent delivery. Such an instrument could provide essential sample cleanup and preconcentration capabilities by loading samples under pressure onto a monolithic adsorbent. The CEC mode could then be used for focusing and separation of the analytes and the combination of these processes would provide greater analytical potential than LC or CEC used alone. It has been already demonstrated that a monolithic capillary column can be quantitatively loaded with sample and flushed using extremely high flow rates translating to a linear flow velocity of  $\sim 50 \text{ mm s}^{-1}$ .

Monolithic technology is the best approach for CEC columns, and in fact is the only workable technology to date for CEC on a microchip. Therefore, it seems likely that monolithic stationary phases will be used widely in CEC in general, and more or less exclusively in CEC on microchips. The reason for the success of the monolithic technology is not only that the monoliths are relatively easy to prepare, but especially their low pressure resistance and fast mass transfer make them well-suited for rapid separations. The use of photopolymerization and patterning of different sections of the channel with masks, leading to sections with different stationary phase chemistries, hold great promise for real-world applications to the analysis of difficult samples, where sample preconcentration, cleanup, and separation steps are performed sequentially in two or more different sections of a CEC microchip. Microchip technology also has great potential as a platform for the development of multidimensional separations, combining, for example, LC–CEC, CEC–CE, etc.

Detection techniques are probably the weakest point of the current CEC technology and further

developments in this area are necessary, especially aiming at increased detection sensitivity. Hyphenation of CEC with mass spectrometry may become routine when simpler, smaller and more affordable mass spectrometers become available. Similarly, in inorganic analysis and element speciation, routine use of hyphenation with inductively coupled plasma mass spectrometers would be logical.

**See also:** **Capillary Electrophoresis:** Overview. **Mass Spectrometry:** Overview; Principles. **Micellar Electrokinetic Chromatography.** **Micro Total Analytical Systems.**

## Further Reading

- Bartle K and Myers P (eds.) (2001) *Capillary Electrochromatography*. Cambridge: Royal Society of Chemistry.
- Debovski JK (2002) Selected applications of capillary electrochromatography in the pharmaceutical industry: to buy or not to buy? *Journal of Liquid Chromatography and Related Technologies* 25: 1875–1917.
- Deyl Z and Svec F (eds.) (2001) *Capillary Electrochromatography*. *Journal of the Chromatographic Library*, vol. 62. Amsterdam: Elsevier.
- Fintschenko Y, Kirby BJ, Hasselbrink EF, Singh AK, and Shepodd TJ (2003) Miniature and microchip technologies. In: Svec F, Tennikova TB, and Deyl Z (eds.) *Monolithic Materials*. *Journal of the Chromatographic Library*, vol. 67, pp. 659–685. Amsterdam: Elsevier.
- Hilder EF, Svec F, and Frechet JMJ (2002) Polymeric monolithic stationary phases for capillary electrochromatography. *Electrophoresis* 23: 3934–3953.
- Kang J, Wistuba D, and Schurig V (2002) Polymeric monolithic stationary phases for capillary electrochromatography. *Electrophoresis* 23: 4005–4021.
- Krull IS, Stevenson RL, Misty K, and Swartz ME (2000) *Capillary Electro-Chromatography and Pressurized Flow Capillary Electro-Chromatography*. New York: HNB Publishing.
- Lämmerhofer M and Lindner W (2003) Capillary electrochromatography. In: Svec F, Tennikova TB, and Deyl Z (eds.) *Monolithic Materials*. *Journal of the Chromatographic Library*, vol. 67, pp. 489–559. Amsterdam: Elsevier.
- Malik A (2002) Advances in sol–gel based columns for capillary electrochromatography: sol–gel open-tubular columns. *Electrophoresis* 23: 3973–3992.
- Mikšik I and Deyl Z (2003) Survey of chromatographic and electromigration separations. In: Svec F, Tennikova TB, and Deyl Z (eds.) *Monolithic Materials*. *Journal of the Chromatographic Library*, vol. 67, pp. 623–658. Amsterdam: Elsevier.
- Mistry K, Krull I, and Grinberg N (2002) Capillary electrochromatography: an alternative to HPLC and CE. *Journal of Separation Science* 25: 935–958.

# CAPILLARY ELECTROPHORESIS

Contents

**Overview**

**Pharmaceutical Applications**

**Low-Molecular-Weight Ions**

**Environmental Applications**

**Food Chemistry Applications**

**Clinical Applications**

## Overview

**M Bowser**, University of Minnesota, Minneapolis, MN, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Capillary electrophoresis (CE) is an example of a technique that was conceptualized before its time. The technology needed to fully realize the potential

of this technique was not available when Hjerten first demonstrated inorganic ion and protein separations in 1–3 mm quartz tubes in 1967. Hjerten was able to demonstrate that separations were possible in narrow tubes, but there was no reliable source of the submillimeter inner diameter (ID) fused silica capillary that was necessary to realize the full advantages of the technique. Even if there was, detector technology had not evolved to the point where the miniscule amounts of analyte present in nanoliter detection volumes could be measured. It was not until the early 1980s that conditions were right for CE to become a mainstream analytical technique.



developments in this area are necessary, especially aiming at increased detection sensitivity. Hyphenation of CEC with mass spectrometry may become routine when simpler, smaller and more affordable mass spectrometers become available. Similarly, in inorganic analysis and element speciation, routine use of hyphenation with inductively coupled plasma mass spectrometers would be logical.

**See also:** **Capillary Electrophoresis:** Overview. **Mass Spectrometry:** Overview; Principles. **Micellar Electrokinetic Chromatography.** **Micro Total Analytical Systems.**

## Further Reading

- Bartle K and Myers P (eds.) (2001) *Capillary Electrochromatography*. Cambridge: Royal Society of Chemistry.
- Debovski JK (2002) Selected applications of capillary electrochromatography in the pharmaceutical industry: to buy or not to buy? *Journal of Liquid Chromatography and Related Technologies* 25: 1875–1917.
- Deyl Z and Svec F (eds.) (2001) *Capillary Electrochromatography*. *Journal of the Chromatographic Library*, vol. 62. Amsterdam: Elsevier.
- Fintschenko Y, Kirby BJ, Hasselbrink EF, Singh AK, and Shepodd TJ (2003) Miniature and microchip technologies. In: Svec F, Tennikova TB, and Deyl Z (eds.) *Monolithic Materials*. *Journal of the Chromatographic Library*, vol. 67, pp. 659–685. Amsterdam: Elsevier.
- Hilder EF, Svec F, and Frechet JMJ (2002) Polymeric monolithic stationary phases for capillary electrochromatography. *Electrophoresis* 23: 3934–3953.
- Kang J, Wistuba D, and Schurig V (2002) Polymeric monolithic stationary phases for capillary electrochromatography. *Electrophoresis* 23: 4005–4021.
- Krull IS, Stevenson RL, Misty K, and Swartz ME (2000) *Capillary Electro-Chromatography and Pressurized Flow Capillary Electro-Chromatography*. New York: HNB Publishing.
- Lämmerhofer M and Lindner W (2003) Capillary electrochromatography. In: Svec F, Tennikova TB, and Deyl Z (eds.) *Monolithic Materials*. *Journal of the Chromatographic Library*, vol. 67, pp. 489–559. Amsterdam: Elsevier.
- Malik A (2002) Advances in sol–gel based columns for capillary electrochromatography: sol–gel open-tubular columns. *Electrophoresis* 23: 3973–3992.
- Mikšik I and Deyl Z (2003) Survey of chromatographic and electromigration separations. In: Svec F, Tennikova TB, and Deyl Z (eds.) *Monolithic Materials*. *Journal of the Chromatographic Library*, vol. 67, pp. 623–658. Amsterdam: Elsevier.
- Mistry K, Krull I, and Grinberg N (2002) Capillary electrochromatography: an alternative to HPLC and CE. *Journal of Separation Science* 25: 935–958.

# CAPILLARY ELECTROPHORESIS

Contents

**Overview**

**Pharmaceutical Applications**

**Low-Molecular-Weight Ions**

**Environmental Applications**

**Food Chemistry Applications**

**Clinical Applications**

## Overview

**M Bowser**, University of Minnesota, Minneapolis, MN, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Capillary electrophoresis (CE) is an example of a technique that was conceptualized before its time. The technology needed to fully realize the potential

of this technique was not available when Hjerten first demonstrated inorganic ion and protein separations in 1–3 mm quartz tubes in 1967. Hjerten was able to demonstrate that separations were possible in narrow tubes, but there was no reliable source of the submillimeter inner diameter (ID) fused silica capillary that was necessary to realize the full advantages of the technique. Even if there was, detector technology had not evolved to the point where the miniscule amounts of analyte present in nanoliter detection volumes could be measured. It was not until the early 1980s that conditions were right for CE to become a mainstream analytical technique.



Procedures for the production of large quantities of fused silica capillaries were developed to fill the growing demand for capillary gas chromatography columns. The advent of laser induced fluorescence (LIF) and electrochemical (EC) detectors provided the high mass sensitivity necessary for detection in nanoliter volumes. Jorgenson and Lucas published several key papers on modern CE in 1981. Since then CE has experienced exponential growth to the point where there are well over 2000 articles published annually involving CE.

## Instrumental Overview

Figure 1 shows a schematic of a typical CE instrument. A length of fused silica capillary (typically 20–50 cm) filled with separation buffer extends between two reservoirs. To perform a separation the inlet of the separation capillary is moved to the sample for injection. The inlet is returned to the separation buffer reservoir after a volume ( $\sim 20$ – $50$  nL) of sample has been introduced. An electric potential ( $\sim 10$ – $30$  kV) is applied between the two buffer reservoirs. Analytes migrate through the capillary according to their electrophoretic mobility and the electroosmotic flow (EOF). Analytes are detected either on or off column, depending on the particular detector. Separations with efficiencies of 100 000–150 000 plates and analysis times of 10–20 min are typical.

## The Capillary

The capillary is the central component of the CE instrument. Many of the technical advantages as well

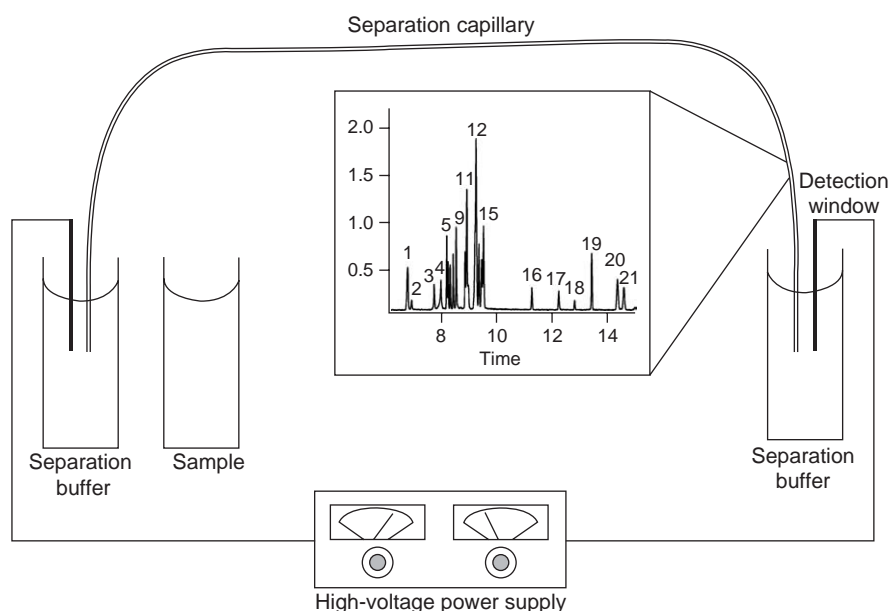
as limitations of CE are due to the unique aspects of performing electrophoresis in narrow tubes. Fused silica capillaries dominate modern capillary electrophoresis. Typical dimensions are 5–150  $\mu\text{m}$  ID and 150–360  $\mu\text{m}$  outer diameter (OD), with 50–100  $\mu\text{m}$  ID and 360  $\mu\text{m}$  OD being most common.

There are two major advantages to performing electrophoresis in capillaries. The first is heat generation (referred to as Joule heating) and dissipation. The electric field in slab gel electrophoresis is typically limited to  $\sim 15$ – $40$   $\text{V cm}^{-1}$ . At higher fields, enough heat may be generated to melt the gel. Although gels are not necessary in CE, excess heat generation leads to temperature, and therefore viscosity, gradients across the diameter of the capillary. The analyte near the center of the capillary experiences a lower viscosity and travels faster than the analyte near the capillary walls. This mobility gradient across the diameter of the capillary contributes to peak broadening, degrading the separation.

The heat generated during electrophoresis is similar to any common resistive element and can be described according to

$$\frac{\partial H}{\partial H} = \frac{IV}{LA} \quad [1]$$

where  $I$  is the current,  $V$  the applied voltage,  $L$  the length of the capillary, and  $A$  the cross-sectional area of the capillary. The heat dissipation is proportional to the surface area, not the volume of the capillary. If the diameter of the capillary is increased, the heat generated increases faster than the ability of the capillary to dissipate this heat, giving rise to a



**Figure 1** Schematic of a CE instrument.

**Table 1** Effect of capillary diameter on Joule heating<sup>a</sup>

Capillary diameter ( $\mu\text{m}$ )	Cross-sectional area ( $\text{m}^2$ )	Current ( $\mu\text{A}$ )	$\Delta T(^{\circ}\text{C})$
10	$7.85 \times 10^{-11}$	3	0.00054
25	$4.91 \times 10^{-10}$	19	0.021
50	$1.96 \times 10^{-9}$	75	0.34
75	$4.42 \times 10^{-9}$	169	1.7
100	$7.85 \times 10^{-9}$	300	5.4
150	$1.77 \times 10^{-8}$	675	27

<sup>a</sup>Voltage = 30 kV,  $K = 1 \times 10^{-9} \text{ m}^2 \text{ J}^{\circ}\text{C}$ .

temperature gradient across the capillary. The difference in temperature between the center and inner wall of the capillary ( $\Delta T$ ) is given by

$$\Delta T = 0.24 \frac{Wr^2}{4K} \quad [2]$$

where  $W$  is the power,  $r$  is the radius of the capillary, and  $K$  is the thermal conductivity of the buffer, capillary wall, and polyimide coating.

As shown in **Table 1**, the effect of Joule heating decreases drastically as capillary diameter is decreased, allowing much higher electric fields to be applied than in slab gel electrophoresis. A value of  $800 \text{ V cm}^{-1}$  is common, but much higher fields are starting to be used in high-speed CE. It is these higher electric fields that give rise to the high efficiency, increased separation power, and short analysis times characteristic of CE.

The second significant advantage of performing electrophoresis in capillaries is the elimination of convective flow. In electrophoresis chambers with large cross-sectional areas, thermal gradients created during electrophoresis induce convective mixing. This convection disrupts analyte bands, degrading separation efficiency. Gels were originally introduced into slab gel electrophoresis systems to prevent convection, not to improve the separation as is commonly thought. The capillary dimensions typically used in CE are too small for these convection currents to develop, allowing separations to be performed in free solution. The elimination of convective currents as a source of band broadening further contributes to the high efficiencies typically observed in CE.

## Capillary Chemistry

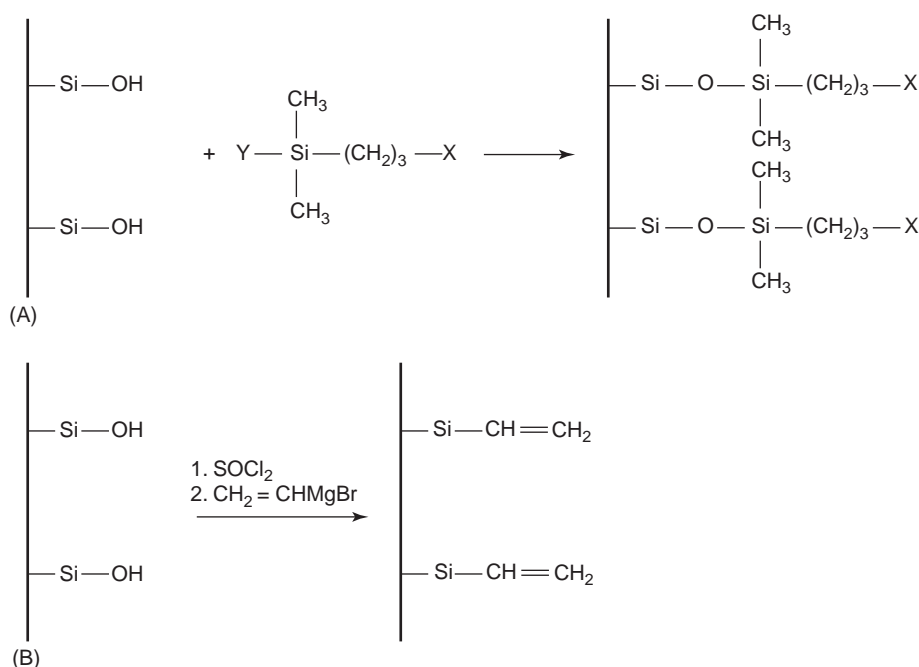
Fused silica capillaries are used almost exclusively in CE. Fused silica offers good mechanical stability and transparency and is relatively cheap and easy to obtain. The inner surface chemistry of these capillaries is determined by silanol functionalities. These silanols are weak acids that dissociate over a pH range

of 3–9. Therefore, under most separation conditions the capillary wall bears a negative charge, giving rise to EOF in the direction of the cathode. Often this is an advantage, allowing both cations and anions to be analyzed simultaneously. Many analytes, especially small neutral or negative molecules, show little interaction with the capillary wall, allowing the high efficiencies typically observed in CE.

There are many analytes that do interact with the negative silanols of the capillary wall though. Interactions with the wall introduce a mass transfer term into the peak broadening equation, often giving rise to peak tailing. Large molecules with cationic functional groups are most problematic. There are several approaches to modify the capillary surface chemistry for minimizing wall interactions. In some cases lowering the pH value below 3 will decrease interactions with cations. Most silanols are neutral at this pH, decreasing ionic interactions. There are many cases though where this solution is not practicable. Other options include permanent derivatization or dynamic coating of the capillary surface.

## Covalent Derivatization

Early wall coatings took advantage of silane coupling reactions that had been well characterized during the development of chromatographic stationary phases. Halo or alkoxyl silanes can be coupled to the capillary surface through the formation of siloxane bonds (see **Figure 2**). After this initial coupling, glycidoxyl functional groups are converted to the glyceryl derivative or substituted with another hydrophilic functionality such as polyacrylamide. These coatings were successful in diminishing EOF and minimizing protein–wall interactions. Unfortunately, siloxane bonds are easily hydrolyzed in moderately alkaline solutions, limiting their use to pH values below 8. More recently Grignard chemistry has been used to attach a vinyl functionality to the capillary surface through the formation of Si–C bonds. Free radical polymerization is initiated at the vinyl sites to complete the coating procedure. Linear polyacrylamides are most common, but polymers such as epoxides, diols, polyethylene glycol, polyvinyl alcohol (PVA), cellulose derivatives, dextran, and polyvinylpyrrolodone (PVP) have seen increasing use. The Si–C bond is much more stable to hydrolysis than the siloxane bond, allowing capillaries coated using this chemistry to be used over a wider pH range (2–10.5). Cationic (e.g., polyvinylamine) or anionic (e.g., 2-acrylamido-2-methylpropanesulfonic acid (NaAMPS)) polymers can be used in the final polymerization step as well. These charged coatings offer more reproducible EOF than uncoated



**Figure 2** Derivatization strategies for permanently coating the capillary surface. Si–O bonds are formed through silanization (A). Si–C bonds are formed using a Grignard reaction (B).

capillaries. In addition, cationic coatings can be used to reverse the direction of the EOF, which may be useful in optimizing certain separations.

## Dynamic Coatings

Noncovalent interactions can also be used to coat the capillary surface. Dynamic coatings can be generated by adding compounds to the separation buffer that interact with the capillary surface. Cationic surfactants such as cetyltrimethylammonium bromide (CTAB) form double layers over the capillary silanols when added to the buffer at concentrations higher than the critical micelle concentration, effectively coating the wall with a cationic surface. Similarly, neutral surfactants, such as Brij 35, can be used to coat the wall with a neutral surface. A potential drawback to this approach is that the introduction of micelles to the separation buffer can have dramatic effects on the separation. Other small polyamines (e.g., spermine and hexamethonium bromide) have been successfully used as dynamic coatings. Polyamines demonstrate a higher affinity for the capillary surface than do surfactants, allowing them to be effective at lower concentrations. In some cases it is possible to remove the additive from the separation buffer if the surface is regenerated by rinsing the capillary with the polyamine after several separations.

Hydrophilic polymers can be adsorbed onto the capillary surface to generate a neutral semipermanent coating. Examples include polysaccharides,

PVA, PVP, polyethylene oxide, and polyalkylene glycols. The properties of these polymers must be carefully balanced. More hydrophilic polymers tend to be most successful in decreasing interactions with proteins. Conversely, the stability of the coating increases with hydrophobicity.

## Capillary Gels

The development of capillary gels, analogous to slab gels, has largely been driven by the genomics boom. Separation of DNA fragments of varying length is not feasible in free solution CE. Initially capillaries were cast with permanent gels. Cross-linkers and extenders were added during the polymerization step described of the permanent polyacrylamide coatings described above. Gels of varying density can be generated by varying the concentration of the cross-linker. Although these gels made high-speed DNA sequencing possible, they suffered from short lifetimes and were relatively difficult to make. More recently polymeric solutions have been used as replaceable gels. High-pressure rinses are used to load capillaries with the polymer solutions. The polymer forms a cross-linked network if the concentration is above the entanglement concentration. This approach allows the gel to be replaced after every separation if necessary. Polymer concentration and size can be easily changed, giving increased options when optimizing the separation. Polyalkylene oxides

and polyalkylene glycols are commonly used as replaceable gels.

## Injection

As a general rule, less than 2–5% of the length of the capillary should be loaded with a sample to prevent the injection volume from significantly contributing to the peak width. For a 50 cm long, 50  $\mu\text{m}$  ID capillary, this corresponds to 20–50 nL, obviously much too small to be loaded using technologies developed for chromatography.

Electrokinetic injections are performed by placing the inlet of the capillary in the sample and applying a voltage for a set period of time. The quantity of analyte ( $Q$ ) loaded onto the capillary is described by

$$Q = \mu\pi r^2 E C t \quad [3]$$

where  $\mu$  is the analyte mobility,  $r$  is the capillary radius,  $E$  is the electric field,  $C$  is the analyte concentration, and  $t$  is the injection time. As shown in eqn [3], the amount of analyte injected depends in part on the mobility of the analyte. Analytes with higher mobility are preferentially injected over analytes with lower mobilities. This phenomenon is referred to as electrophoretic or electrokinetic bias. An electrokinetic bias can give rise to differing limits of detection (LODs) and peak widths for analytes, depending on their mobility. A less obvious drawback of electrokinetic injections is their high susceptibility to matrix effects. The electric field in the sample, and therefore the mobility of the analyte in the sample, is dependent on the sample conductivity. The amount of analyte injected is therefore highly dependent on the ionic strength of the sample.

Hydrodynamic injections are performed by inserting the inlet of the separation capillary into the sample and applying pressure at the inlet or vacuum at the outlet. The volume of sample injected ( $V$ ) is determined by

$$V = \frac{\Delta P D^4 \pi}{128 \eta L} \quad [4]$$

where  $\Delta P$  is the pressure difference between the inlet and outlet,  $D$  is the ID of the capillary,  $\eta$  is the viscosity of the separation buffer, and  $L$  is the length of the capillary. Siphoning can be used to perform a gravimetric injection. In this case  $\Delta P$  in eqn [4] is determined by

$$\Delta P = \rho g \Delta h \quad [5]$$

where  $\rho$  is the density of the sample,  $g$  is the gravitational constant, and  $\Delta h$  is the height difference

between the top of the inlet solution and the outlet solution. Hydrodynamic injections do not give rise to an injection bias. Additionally, matrix effects often have less of an effect on hydrodynamic injections. For these reasons hydrodynamic injections have become more common than electrokinetic injections, the exception being capillary gel electrophoresis, where the application of pressure disrupts the capillary gel. Both pressure and electrokinetic injections are usually available on commercial instruments. Research instruments built in-house are typically limited to electrokinetic or gravimetric injection.

## Capillary Cooling

As discussed above, Joule heating can become a significant contributor to band broadening at high electric fields, especially if high ionic strength buffers or large-diameter capillaries are used. Temperature fluctuations can also give rise to variability in migration times since the mobility is inversely proportional to the viscosity. A 10°C rise in separation buffer temperature can decrease the migration time by as much as 20%. Temperature control systems have become common components of commercially built instruments to help address these issues. Most systems pass air or liquid coolants over the separation capillary to dissipate Joule heat. Air cooling systems typically employ high flow rates (10  $\text{m s}^{-1}$ ) to improve heat dissipation. Liquid cooling systems have lower flow rates, but this is balanced by the higher heat capacity of the liquid coolant. Both systems work well to improve run-to-run migration time reproducibility. Typically capillaries can be thermostatted anywhere from 10°C below ambient temperature to  $\sim 60^\circ\text{C}$ .

## Power Supply

A high-voltage power supply is necessary for applying the electric field across the capillary. Power supplies with maximum outputs of  $\pm 30$  kV and 300  $\mu\text{A}$  are typical. Constant voltage, current, and power modes are often available, although most CE separations are performed in the constant voltage mode. Current monitoring is very useful in troubleshooting. When operating in the constant voltage mode the current should remain relatively constant throughout the separation. Spikes in current suggest partial blockage or breakage of the capillary. Similarly, no current across the capillary suggests complete blockage or breakage. Dual polarity power supplies are useful for cases where the analyte migration direction is reversed (e.g., a DNA separation in a neutral coated

capillary). By convention, normal polarity refers to the case where the anode is at the inlet of the capillary and the cathode is at the outlet. Reverse polarity refers to the case where the inlet potential is negative with respect to the outlet potential. The output potential should be relatively precise since variations in electric field contribute to irreproducibility in migration time. Precise timing is also necessary if electrokinetic injections are used. Platinum electrodes are used to make electrical contact with the inlet and outlet buffers. Platinum is chosen because of its inertness and long-term stability.

Performing separations at the 30 kV limit is not uncommon since separations often improve as the electric field is increased, assuming Joule heating does not become significant. Special precautions are necessary when applying higher voltages since the fields generated above 30 kV begin to approach the electric breakdown potential of air. The maximum current limit of the power supply is usually not a concern since Joule heating often becomes unacceptable at currents well below 300  $\mu\text{A}$ .

## Detectors

The progress of CE has been closely linked to the development of detector technology. Although the mass limits of detection (MLODs) for CE are often better than those for high-performance liquid chromatography (HPLC), the concentration limits of detection (CLODs) are usually orders of magnitude higher. The excellent MLOD is a result of the high peak efficiency typically observed in CE. Very little dilution of the analyte occurs during the separation in comparison with HPLC. Conversely, the dimensions of the capillary severely limit the CLOD. In an on-column detector, the detection volume significantly contributes to the peak width if the length of the detection region is greater than 100–200  $\mu\text{m}$ . In a 50  $\mu\text{m}$  ID capillary, this corresponds to a detection volume of only 100–200 pL, 10 000 times smaller than that of a typical HPLC detector. Even if the mass detection is excellent, the fact that it is in such a small volume limits the CLOD, forcing researchers in the CE field to push continually for further improvements in detector sensitivity.

## Absorbance Detectors

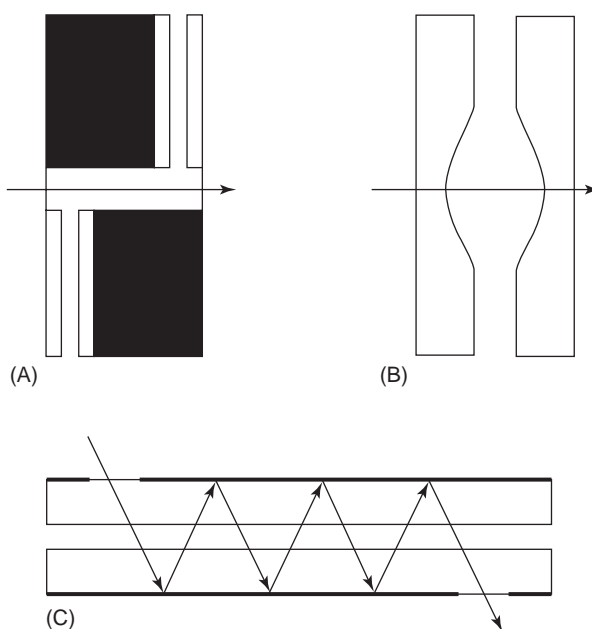
Ultraviolet (UV) and visible (Vis) absorbance are commonly used for detection in CE. The wavelengths available are limited by the UV-cutoff of the solvent and the fused silica capillary, allowing wavelengths as low as 190 nm to be used in aqueous solvents.

Many analytes contain aromatic or extended pi systems that absorb in this range, making UV–Vis somewhat universal detectors. Detection is performed on column. A narrow region of polyimide coating is removed from the separation capillary to make a detection window. Deuterium or tungsten lamps are common excitation sources in the UV and Vis wavelengths, respectively. Transmitted light is measured using photodiodes or photomultiplier tubes. Photodiode arrays can be used for multiple wavelength or full spectrum detection.

The on-column approach to CE absorbance detectors limits the path length to the ID of the CE capillary. The path length for a 50  $\mu\text{m}$  ID capillary will be 200 times shorter than that of a common 1 cm cuvette. The short path length significantly limits the sensitivity of CE absorbance detectors (see Table 2). A number of absorbance cells have been designed to increase the detection path length (see Figure 3). The

**Table 2** Representative LODs for CE detectors

Detector	Concentration LOD ( $\text{mol l}^{-1}$ )	Mass LOD (mol)
UV–Vis absorbance	$10^{-6}$ – $10^{-4}$	$10^{-15}$ – $10^{-13}$
Fluorescence	$10^{-16}$ – $10^{-6}$	$10^{-21}$ – $10^{-11}$
Amperometry	$10^{-8}$ – $10^{-6}$	$10^{-19}$ – $10^{-17}$
Conductivity	$10^{-8}$ – $10^{-7}$	$10^{-17}$ – $10^{-16}$
Potentiometry	$10^{-8}$ – $10^{-7}$	$10^{-17}$ – $10^{-16}$
MS	$10^{-9}$ – $10^{-5}$	$10^{-17}$ – $10^{-15}$



**Figure 3** High-sensitivity detection cells for UV–Vis absorbance CE detectors: (A) Z-cell; (B) bubble cell; and (C) multiple-reflection cell.

Z-cell introduces a flow through chamber parallel to the excitation beam, extending the path length to  $\sim 1$  mm. The ID of the bubble cell is widened at the detection point. The surface of the capillary can be coated with a reflective coating to make a detection cell where the excitation beam crosses the capillary multiple times. All these approaches offer a modest improvement in sensitivity at the expense of an increased detection volume, which can increase the observed width of the analyte peaks.

The optical properties of an on-column detection cell are far from ideal. During CE's early development, HPLC absorbance detectors were used. Unfortunately the dimensions of the capillary are much smaller than the typical HPLC absorbance cell. Much of the excitation light did not pass through the capillary and scatter was high. The circular cross-section of the capillary gives rise to varying path lengths across the excitation beam. The combination of high scatter and variation in excitation beam path length limited the linear range of early CE detectors. More recent absorbance detectors incorporate a ball lens to focus light through the center of the capillary, decreasing both scatter and variation in path length. This optical arrangement decreases detector noise 100-fold and increases the linear range to over three orders of magnitude.

## Fluorescence

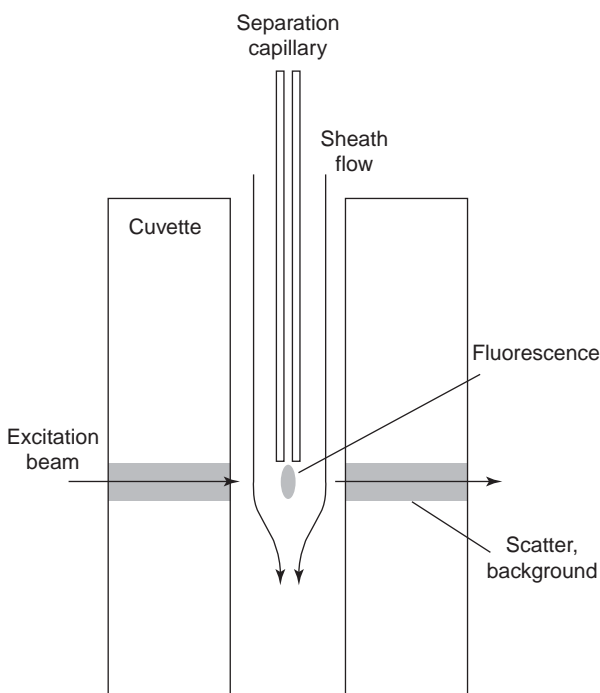
Of the CE detectors available, fluorescence is by far the most sensitive. Mass detection limits have been reported as low as a single molecule. CLOD values routinely approach  $<10^{-13} \text{ mol l}^{-1}$ . Unfortunately, most analytes are not natively fluorescent, making sample derivatization necessary. Labeling protocols for many functional groups, including amines, carboxylic acids, and thiols, have been developed.

Lasers are almost exclusively used as the excitation source in CE detectors. The low output divergence of lasers allows the excitation beam to be easily focused to a spot on the size scale of the ID of the capillary, giving rise to high excitation intensity and less scatter. Laser technology is constantly improving, making gains in reliability and affordability. Lasers are available with emission lines across the visible region of the spectrum. Development of lasers that emit deep into the UV has made native fluorescence detection of proteins and DNA increasingly feasible.

Fluorescence detection is usually performed on column, using an approach similar to that described for absorbance detectors. Excitation is collected through an objective at  $90^\circ$  to the excitation beam. Filters and field stops are used to block scatter and

pass fluorescence. Care must be used when choosing excitation and emission wavelengths to minimize the background signal generated by Raman scattering from the solvent. Fluorescence is usually detected using high-sensitivity photomultiplier tubes, although recent advances have made charge coupled device detection more common.

As discussed with absorbance detectors, the circular shape of the capillary makes for a nonideal optical arrangement when performing on-column fluorescence detection. The excitation beam is scattered over  $\sim 360^\circ$  at the air-fused silica and to a lesser extent the fused silica–buffer boundaries. Impurities in the fused silica capillary also contribute to the background fluorescence. Development of high-sensitivity fluorescence detection cells for CE has been an area of active research. The most successful design has been the sheath-flow cuvette (see **Figure 4**). The outlet of the capillary is positioned in a quartz cuvette ( $2 \text{ mm} \times 2 \text{ mm}$  square with a  $200 \mu\text{m} \times 200 \mu\text{m}$  square bore, optical quality). The separation buffer is flowed around the outlet of the separation capillary, pulling the analyte off the capillary in a laminar flow profile. LIF is performed off column in this laminar flow region, spatially isolating the fluorescence signal from the background fluorescence and the scatter that occur at the cuvette walls. The sheath-flow cuvette typically improves the S/N ratio by  $>100$ -fold when compared with on-column fluorescence detectors.

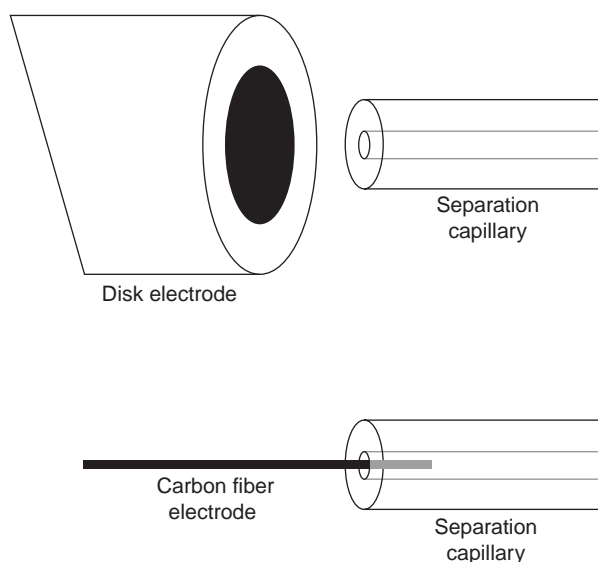


**Figure 4** Schematic of a sheath-flow, high-sensitivity, off-column LIF detection cell.

## EC Detection

EC detection has become increasingly popular over the past several years. EC detectors offer excellent sensitivity and in some cases selectivity. A wide range of analytes can be detected using EC. There are four modes of EC detection that have found use in CE: amperometry, conductivity, voltammetry, and potentiometry.

Amperometric detection has seen the most use to date. In this approach an electrode is placed near or in the outlet of the separation capillary. The electrode is held at a fixed potential. The current is measured as the analyte undergoes redox chemistry at the electrode. Amperometric detection is usually performed in the oxidative mode since dissolved oxygen in the separation buffer interferes with detection in the reductive mode. Electrodes are usually made of carbon fibers or glassy carbon. Several electrode designs have been implemented (see Figure 5). The most common approach is to place a disk electrode at the outlet of the capillary. Conversely, a carbon fiber can be inserted into the end of the capillary. Inserting a carbon fiber electrode into the end of the capillary decreases the detection volume, improving the peak shape. The oxidation efficiency, and therefore the sensitivity, is also improved since the analyte must migrate past the electrode surface to exit the capillary. The oxidation efficiency, and therefore the signal, is dependent on reproducibly positioning the electrode, which may have limited the spread of EC CE detectors. Another concern is isolating the detection electrode from the electric field generated by the separation voltage. Off-column disk electrodes are



**Figure 5** Disk and carbon fiber electrodes used in amperometric and voltammetric CE detectors.

only minimally affected by the separation potential if the separation current is kept low and the electrode is placed a reasonable distance from the capillary outlet. When on-column carbon fibers are used, the separation voltage must be grounded before the amperometry electrode. The capillary can be fractured or etched to provide a route for closing the circuit. The gap created to complete the CE circuit must be small enough to prevent a significant amount of analyte from escaping from the capillary. Palladium decouplers and electrically isolated potentiostats have also been employed for isolating the amperometry electrode.

Conductivity detection is performed by measuring the current between two electrodes. Most analytes analyzed using CE will be ionic and therefore should give rise to a change in conductivity. An AC voltage is used in the conductivity cell to prevent faradic reactions. Unfortunately, the LODs for conductivity detectors are higher than those for amperometry because small changes in buffer composition give rise to noise at the detector.

Voltammetric detection is performed by measuring the redox current at an electrode at the end of the capillary. The current is related to the analyte concentration, and the correlation of the onset potential and the peak current gives information about the identity of the analyte. The same electrode arrangements used in amperometric detection can be used in voltammetric detectors. The LODs using voltammetry are much poorer than with amperometric detectors due to capacitive charging and nonfaradic current generated when the potential is cycled.

Ion-selective electrodes are used in potentiometric detectors. Early designs inserted micropipette electrodes into the outlet of the separation capillary. The capillary outlet was widened using etching to make placement easier and decrease the electric field at the electrode. As such it was found that electrical decoupling was unnecessary. More recent designs have used coated wire electrodes where a solid wire is coated with a polyvinyl chloride membrane. The electrode is placed  $\sim 50\ \mu\text{m}$  from the outlet of the capillary. Coated wire electrodes are much less fragile and easier to position than micropipette electrodes.

## Mass Spectrometry

Mass spectrometry (MS) detection in CE has grown enormously over the past decade. MS detection is very sensitive and is well suited to the small sample volumes typical of CE. MS also provides structural information about the analytes, often allowing



unambiguous peak identification, a feature not available with other detection schemes. Most analytes can be analyzed using MS, making this a nearly universal detection approach.

The major difficulty in the initial development of MS detectors was the coupling of CE, a liquid-based technique, with MS, a gas phase analysis. Electrospray ionization (ESI) interfaces have found the most use to date. In contrast with HPLC, where flow rates need to be attenuated, the flow rates in CE are generally too small ( $\sim 1\text{--}100\text{ nL min}^{-1}$ ) to support an electrospray in a standard interface. A sheath flow was introduced to add additional volume to the CE buffer. The voltage applied to the sheath flow both generates the electrospray and closes the CE circuit (see Figure 6).

More recently sheathless-flow ESI interfaces, also referred to as nanospray interfaces, have been developed for CE. The electrospray is generated directly at the tip of the capillary. The capillary outlet is pulled to a tapered tip to reduce the flow rate and maximize the electric field, thereby stabilizing the electrospray at flow rates as low as several nanoliters per minute. The S/N ratio is greatly improved over sheath-flow ESI interfaces since there is no dilution of the analytes as they exit the capillary. Several approaches have been developed for applying the electrospray voltage (and completing the CE circuit) at the capillary outlet. Early designs coated the tip of the capillary outlet with gold, allowing an electrical connection to be made. Although successful, these

gold-coated tips had relatively short lifetimes due to EC and electrical degradation. Nanospray tips have been coupled to the outlet of the CE capillary in low-dead volume stainless steel unions or Nafion tubing. In this arrangement the electrical connection is made through the gap between the CE outlet and the nanospray tip. A downside of this approach is the relatively large dead volume of the interface, which can lead to peak broadening. A wire can be inserted into the CE capillary outlet or a hole near the outlet to complete the ESI and CE circuits. Again, while this approach is successful, it is not without drawbacks. The CE capillary must have an ID  $> 25\text{ }\mu\text{m}$  to admit the wire. This is unfortunate since the S/N ratio increases dramatically as the diameter of the nanospray tip is decreased.

Sheath-flow and nanospray ESI interfaces have successfully coupled CE to almost every mode of MS including the use of quadrupole, time-of-flight, ion-trap, and magnetic sector instruments. CE-MS has been employed in the analysis of a wide range of analytes but has been especially successful in biological assays. The mass selectivity of MS detection is especially useful in complex DNA, protein, and carbohydrate separations.

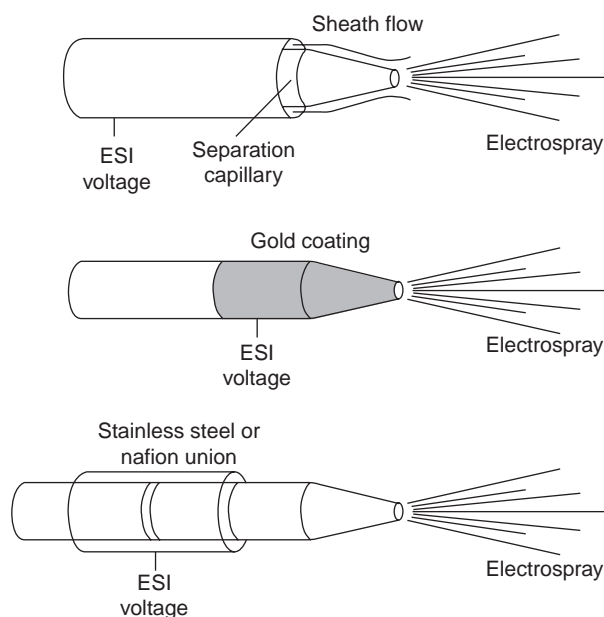
## Safety

Researchers should be reminded that CE makes use of relatively high voltages (up to 30 kV). Although the current is limited to 300  $\mu\text{A}$  in many CE power supplies, the potential for injury should be respected. CE instruments are often housed in Plexiglas boxes to prevent accidental exposure to high voltages. An interlock is used to disengage the high voltage automatically if the box is opened without first turning off the CE power supply. Similar safety features are present on commercially built CE instruments.

See also: **Electrophoresis:** Overview; Principles. **Fluorescence:** Overview; Instrumentation. **Mass Spectrometry:** Overview; Atmospheric Pressure Ionization Techniques; Electrospray.

## Further Reading

- Holland LA and Leigh AM (2002) Amperometric and voltammetric detection for capillary electrophoresis. *Electrophoresis* 23: 3649–3658.
- Horvath J and Dolnik V (2001) Polymer wall coatings for capillary electrophoresis. *Electrophoresis* 22: 644–655.
- Landers JP (ed.) (1994) *Handbook of Capillary Electrophoresis*. Boca Raton, FL: CRC Press.
- Liu CY (2001) Stationary phases for capillary electrophoresis and capillary chromatography. *Electrophoresis* 22: 612–628.



**Figure 6** CE-MS interfaces: (A) sheath-flow ESI; (B) gold-coated nanospray tip; and (C) fractured nanospray tip grounded through a stainless steel or Nafion junction.

- Mehdi M (2002) Capillary electrophoresis mass spectrometry and its application to the analysis of biological mixtures. *Analytical and Bioanalytical Chemistry* 373: 466–480.
- Righetti PG, Gelfi C, Verzola B, and Castelletti L (2001) The state of the art of dynamic coatings. *Electrophoresis* 22: 603–611.
- Swinney K and Bornhop DJ (2000) Detection in capillary electrophoresis. *Electrophoresis* 21: 1239–1250.
- Tanyanyiwa J, Leuthardt S, and Hauser PC (2002) Conductimetric and potentiometric detection in conventional and microchip capillary electrophoresis. *Electrophoresis* 23: 3659–3666.
- von Brocke A, Nicholson G, and Bayer E (2001) Recent advances in capillary electrophoresis/electrospray–mass spectrometry. *Electrophoresis* 22: 1251–1266.
- Weinberger R (ed.) (2000) *Practical Capillary Electrophoresis*, 2nd edn. London: Academic Press.

## Pharmaceutical Applications

**G K E Scriba**, University of Jena, Jena, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Although chromatographic techniques such as gas chromatography (GC) and (primarily) high-performance liquid chromatography (HPLC) are still used in most industrial laboratories for the analysis of drugs and excipients, capillary electrophoresis (CE) has been increasingly applied to analyze pharmaceuticals in recent years. CE can be operated at a similar performance and level of automation as HPLC and has in many instances advantages compared to HPLC in terms of rapid method development and lower operating costs due to reduced consumption of chemicals and samples. However, the major strength of CE is the fact that the separation principle is different from chromatographic techniques so that CE and HPLC in fact form a powerful combination for the analysis of complex molecules. Generally, the scope of applications of CE in pharmaceutical analysis is identical to that of HPLC. Therefore, often a choice between the two techniques has to be made.

In recent years, an increasing number of pharmaceutical companies have included CE methods in early drug discovery testing and routine quality control as well as in documents for regulatory submission. CE methods are accepted by the regulatory authorities such as the US Food and Drug Administration and the European Agency for the Evaluation of Medicinal Products and the technique has been implemented as analytical method by the United States Pharmacopeia and the European Pharmacopoeia.

Based on the number of publications, drugs are actually the preferred analytes in CE. While they served as model compounds for the investigation of specific aspects in some studies, CE has been used to solve ‘real’ pharmaceutical problems in the majority

of applications. Besides small molecules, inorganic ions, peptides, and proteins, oligonucleotide pharmaceuticals have also been analyzed. In addition to pharmaceutical applications, CE is also utilized as an analytical technique by the chemical, cosmetic, and food industries as well as in environmental and forensic analysis.

### Modes of Operation

With respect to small molecules, capillary zone electrophoresis (CZE) for charged molecules and micellar electrokinetic chromatography (MEKC) for the analysis of uncharged compounds has been applied. However, the MEKC mode may also be applied to charged analytes in order to enhance the separation selectivity of a given separation. Enantiomer separations generally require the presence of a chiral selector. For the analysis of large biomolecules such as nucleic acids and proteins, capillary gel electrophoresis (CGE) and capillary isoelectric focusing (CIEF) have also been applied. Isotachophoretic techniques are primarily used for sample concentration.

### Method Development, Validation, and System Suitability

The aim of method development in any analytical separation technique is to obtain an assay that allows the successful separation of the analytes of interest in a short analysis time, with high reproducibility and ruggedness. In CE, factors such as buffer pH, molarity and type of the background electrolyte, applied voltage, temperature of the capillary, and buffer additives such as surfactants, organic solvents, ion-pairing reagents, complexing agents influence a separation. In recent years, chemometrics have been employed to minimize the number of experiments upon definition of the dominant variables of a given separation.

- Mehdi M (2002) Capillary electrophoresis mass spectrometry and its application to the analysis of biological mixtures. *Analytical and Bioanalytical Chemistry* 373: 466–480.
- Righetti PG, Gelfi C, Verzola B, and Castelletti L (2001) The state of the art of dynamic coatings. *Electrophoresis* 22: 603–611.
- Swinney K and Bornhop DJ (2000) Detection in capillary electrophoresis. *Electrophoresis* 21: 1239–1250.
- Tanyanyiwa J, Leuthardt S, and Hauser PC (2002) Conductimetric and potentiometric detection in conventional and microchip capillary electrophoresis. *Electrophoresis* 23: 3659–3666.
- von Brocke A, Nicholson G, and Bayer E (2001) Recent advances in capillary electrophoresis/electrospray–mass spectrometry. *Electrophoresis* 22: 1251–1266.
- Weinberger R (ed.) (2000) *Practical Capillary Electrophoresis*, 2nd edn. London: Academic Press.

## Pharmaceutical Applications

**G K E Scriba**, University of Jena, Jena, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Although chromatographic techniques such as gas chromatography (GC) and (primarily) high-performance liquid chromatography (HPLC) are still used in most industrial laboratories for the analysis of drugs and excipients, capillary electrophoresis (CE) has been increasingly applied to analyze pharmaceuticals in recent years. CE can be operated at a similar performance and level of automation as HPLC and has in many instances advantages compared to HPLC in terms of rapid method development and lower operating costs due to reduced consumption of chemicals and samples. However, the major strength of CE is the fact that the separation principle is different from chromatographic techniques so that CE and HPLC in fact form a powerful combination for the analysis of complex molecules. Generally, the scope of applications of CE in pharmaceutical analysis is identical to that of HPLC. Therefore, often a choice between the two techniques has to be made.

In recent years, an increasing number of pharmaceutical companies have included CE methods in early drug discovery testing and routine quality control as well as in documents for regulatory submission. CE methods are accepted by the regulatory authorities such as the US Food and Drug Administration and the European Agency for the Evaluation of Medicinal Products and the technique has been implemented as analytical method by the United States Pharmacopeia and the European Pharmacopoeia.

Based on the number of publications, drugs are actually the preferred analytes in CE. While they served as model compounds for the investigation of specific aspects in some studies, CE has been used to solve ‘real’ pharmaceutical problems in the majority

of applications. Besides small molecules, inorganic ions, peptides, and proteins, oligonucleotide pharmaceuticals have also been analyzed. In addition to pharmaceutical applications, CE is also utilized as an analytical technique by the chemical, cosmetic, and food industries as well as in environmental and forensic analysis.

### Modes of Operation

With respect to small molecules, capillary zone electrophoresis (CZE) for charged molecules and micellar electrokinetic chromatography (MEKC) for the analysis of uncharged compounds has been applied. However, the MEKC mode may also be applied to charged analytes in order to enhance the separation selectivity of a given separation. Enantiomer separations generally require the presence of a chiral selector. For the analysis of large biomolecules such as nucleic acids and proteins, capillary gel electrophoresis (CGE) and capillary isoelectric focusing (CIEF) have also been applied. Isotachophoretic techniques are primarily used for sample concentration.

### Method Development, Validation, and System Suitability

The aim of method development in any analytical separation technique is to obtain an assay that allows the successful separation of the analytes of interest in a short analysis time, with high reproducibility and ruggedness. In CE, factors such as buffer pH, molarity and type of the background electrolyte, applied voltage, temperature of the capillary, and buffer additives such as surfactants, organic solvents, ion-pairing reagents, complexing agents influence a separation. In recent years, chemometrics have been employed to minimize the number of experiments upon definition of the dominant variables of a given separation.

Generally, water-soluble charged compounds are analyzed by CZE. Typical method development starts with the selection of an appropriate buffer pH followed by investigation of the other variables. Standard conditions using a pH 2.5 phosphate buffer have been developed for the analysis of basic drugs. A general method for the separation of acidic drugs uses a pH 9.5 borate buffer. However, it may also be suitable to investigate the pH range close to the  $pK_a$  values of compounds especially in the case of structurally closely related compounds and diastereomers. Suppression of the electroosmotic flow (EOF) can be obtained by dynamic or permanent coating of the capillary wall. Many reagents and methods for this purpose have been described. When selectivity requirements exceed mobility differences obtainable by CZE then MEKC conditions can be applied. Sodium dodecyl sulfate (SDS) is often employed but many other surfactants are available. Type and concentration of the surfactant are important factors that need to be considered. Alternatively, the use of complexing agents and chiral selectors may increase the selectivity of a given separation system. MEKC is also suitable for the simultaneous analysis of charged and neutral compounds. Furthermore, MEKC is the method of choice for neutral or water-insoluble compounds, for example, steroids or lipid-soluble vitamins. The addition of organic solvents such as methanol, ethanol, 2-propanol, acetonitrile to the background electrolyte further increases the solubility of the analytes in the background electrolyte. Alternatively, nonaqueous conditions can be employed for the analysis of water-insoluble drugs. The separation of enantiomers can be achieved by the indirect method upon derivatization with stereochemically pure reagents or by direct separation of the enantiomers, which generally requires the addition of a chiral selector to the background electrolyte in CZE as well

as MEKC. Numerous types of chiral selectors are commercially available (see below). Type and concentration of a chiral selector need to be evaluated during method development. Combinations of chiral selectors may apply.

As in other analytical techniques, careful and comprehensive method validation is also required in CE in order to obtain a reproducible and rugged method that is suitable for routine analysis and may be accepted by the regulatory authorities. Generally, the same principles apply for CE as for chromatographic techniques depending on the scope of the analysis but some specifics apply (Table 1).

Selectivity (sometimes also termed specificity) is the ability of a method to discriminate an analyte from potentially interfering substances including the matrix. As the selectivity will be particularly altered by a buffer pH that is close to the  $pK_a$  of the analytes and electrolysis of the background electrolyte will inevitably occur, buffer capacity, volume of the electrolyte reservoir, applied current, and the period of time of buffer usage have to be considered. In addition, when using derivatized chiral selectors such as cyclodextrins for enantiomer separations, it should be confirmed that different batches of the selectors from different suppliers give similar performance because the degree of substitution, polydispersity, or purity of the selectors may alter the selectivity. Commercially available randomly substituted cyclodextrins are a mixture of positional isomers with different numbers of substituents. Peak areas in CE depend on the velocity by which analytes travel through the detector cell. Thus, for obtaining accurate results corrected peak areas, i.e., peak area divided by the respective migration time, have to be used for correcting the different residence times of the compounds in the detector cell. In addition, ionization of the analytes or complexation by buffer

**Table 1** Validation requirements in capillary electrophoresis

Characteristic	Identification	Achiral or chiral impurities		Main component assay
		Limit only	Quantitative	
General requirements				
Specificity	+	+	+	+
Accuracy	–	–	+	+
Linearity	–	–	+	+
Range	–	–	+	+
Limit of detection	–	+	+	–
Limit of quantitation	–	–	+	–
Precision	–	–	+	+
CE specific requirements				
Background electrolyte stability	+	+	+	+
Purity of chemicals and chiral selectors	+	+	+	+
Capillary pretreatment, storage, and rinsing sequences	+	+	+	+

additives may alter the response factors in ultraviolet (UV) detection compared to HPLC.

The linearity as a function of the detection signal and the analyte concentration should be evaluated depending on the type of the assay. Thus, a method for the determination of the main component may be calibrated in the range of 50–150% or 80–120% while the linearity of an impurity should be determined in the presence of the main constituent around the maximally tolerated level of the impurity, for example, from the limit of quantitation to 200% of the tolerated level. In bioanalysis, a wider range has to be explored covering two to three orders of magnitude in concentration. The linear range of UV detectors in CE is more restricted than in HPLC due to the circular geometry of the capillary. Moreover, for indirect UV detection the linear range is more restricted compared to direct UV detection.

Precision as the variability of the individual measurements is generally lower compared to HPLC due to the small injection volumes that may vary with altered viscosity or temperature of the test solutions. The use of an internal standard is recommended to correct for injection errors. In addition, the influence of the capillary has to be evaluated by testing capillaries from different batches and suppliers and by carefully validating capillary pretreatment, rinsing, and storage procedures. In routine analysis, capillaries should be dedicated to only one specific application with defined experimental conditions. This is especially true for MEKC methods. The purity of the electrolytes and chiral selectors also has to be considered. As in HPLC the stability of test and reference solutions should be assessed, but in contrast to HPLC, where large volumes of the mobile phase are consumed, only a small volume of electrolyte solution is used in CE. Typically, 500–1000 ml of the background electrolyte is prepared at a time but the shelf life of the solutions may vary depending on the composition.

Robustness as a measure of the method to remain unaffected by small but deliberate variations of the experimental conditions is not listed in the ICH guidelines of method validation but is an important characteristic that should be part of any validation procedure. The most relevant factors such as buffer pH and composition, concentration of a chiral selector, temperature, rinse conditions and times are varied around the values of a method and evaluated. In addition, experimental design has been used for robustness testing allowing assessing the interaction of different factors. Response surface plots give an impression of the variations that can be expected. Robustness testing is especially easy in CE due to the

short equilibration time when changing the composition of the electrolyte.

Method transfer between instruments and laboratories may require some revalidation in CE due to differences in the construction of the instruments, especially the detector and injection systems. Therefore, it is preferable to specify an injection volume that is independent of a specific instrument. To assess the performance of a method in routine analysis system suitability tests comparable to HPLC such as selectivity, resolution, or system precision are recommended. Peak symmetry is not considered as the injection of high concentration often leads to peak distortion in CE.

## Main Component Analysis

A large number of reports have documented that precision and accuracy of CE methods for quality control of pharmaceuticals are comparable to HPLC analyses. RSDs  $\leq 1\%$  can be achieved by carefully controlling the operational parameters. A single set of experimental conditions often allows the analysis of a range of structurally diverse compounds. The use of an internal standard is generally recommended to correct for injection errors. The standard should be selected in such a way that it is well separated from the main component and related substances. Sample preparation steps consist of simple filtration in the case of the analysis of solid dosage forms; solutions can be injected directly upon suitable dilution if necessary. Thus, CE methods have been employed for the analysis of drug substances and pharmaceutical formulations (Table 2) and have been included in regulatory submission files.

Synthetic as well as herbal drugs have been determined by CE. In addition, starting materials of drug synthesis and excipients such as carbohydrates, sweeteners, preservatives, surfactants, dyes, lecithin, fatty acids, solubilizers can be analyzed. CE methods can also be effectively applied to determine the composition of drugs derived from natural or genetically modified organisms that are a mixture of closely related compounds such as the antibiotics erythromycin or gentamicin (Figure 1). In addition to small molecules, CE is increasingly used for the characterization of recombinant protein pharmaceuticals. Examples include insulin, human growth hormone, human platelet-derived growth factor, human epidermal growth factor, interferons, therapeutic monoclonal antibodies, a soluble tumor necrosis factor- $\alpha$  receptor, somatotropin, cytokines, and immunoglobulins. The CE methods employed include CZE, CIEF, and CGE. CE-SDS, which is analogous to SDS-polyacrylamide

**Table 2** Examples of validated capillary electrophoresis methods for main component analysis of pharmaceuticals

<i>Drug</i>	<i>Sample</i>	<i>Mode</i>	<i>Conditions</i>
Amoxicillin and clavulanic acid	Injection	MEKC	Sodium borate-phosphate, pH 8.66, 1.44% SDS
Ampicillin and sulbactam	Injection	MEKC	Sodium borate-phosphate, pH 8.66, 1.44% SDS
Benzalkonium chloride	Ophthalmic solutions	CZE	Sodium phosphate, pH 4.0, 40% acetonitrile
Cephalexin	Oral suspension	MEKC	Sodium tetraborate, 20 mmol l <sup>-1</sup> SDS, 0.1% laurylpolyoxyethylenic ether
Clodronate	Liposomal formulation	CZE	1-Nitroso-2-naphthol-3,6-disulfonic acid disodium salt, pH 8.0
Diclofenac sodium	Tablet	CZE	Sodium tetraborate, pH 9.23
Enoxacin	Tablets	CZE	Sodium tetraborate, pH 8.6
Fluoxetine	Capsules	CZE	Sodium phosphate, pH 2.5
Formoterol	Inhaler capsules	CZE	Sodium phosphate, pH 2.5, 20% acetonitrile
Lansoprazole	Capsules	CZE	Sodium tetraborate, pH 8.7
Meloxicam	Tablets	CZE	Sodium tetraborate, pH 8.5, 5% methanol
Mirtazapine	Tablets	CZE	Sodium phosphate, pH 7.0
Nimesulide	Tablets	CZE	Sodium tetraborate, pH 8.1, 10% ethanol
Reboxetine	Tablets	CZE	Sodium phosphate, pH 2.5
Rufloxacin hydrochloride	Drug substance, tablets	CZE	Sodium borate, pH 8.8
Salbutamol	Tablets, syrup	CZE	Sodium acetate, pH 5.0, 13 mg ml <sup>-1</sup> CM- $\beta$ -CD
Tobramycin	Drug substance	CZE	Sodium tetraborate, pH 10.2, 25% acetonitrile
Ursodeoxycholic acid	Tablets	CZE	Sodium <i>p</i> -hydroxybenzoic acid, pH 8.0
Ximelagatran	Drug substance, tablets	CZE	Sodium phosphate, pH 1.9, 10% acetonitrile, 11 mmol l <sup>-1</sup> HP- $\beta$ -CD
Recombinant human epidermal growth factor	Injection	MEKC	Sodium tetraborate, pH 8.5, 12.5 mmol l <sup>-1</sup> SDS

SDS, sodium dodecylsulfate; HP- $\beta$ -CD, hydroxypropyl- $\beta$ -cyclodextrin; CM- $\beta$ -CD, carboxymethyl- $\beta$ -cyclodextrin.

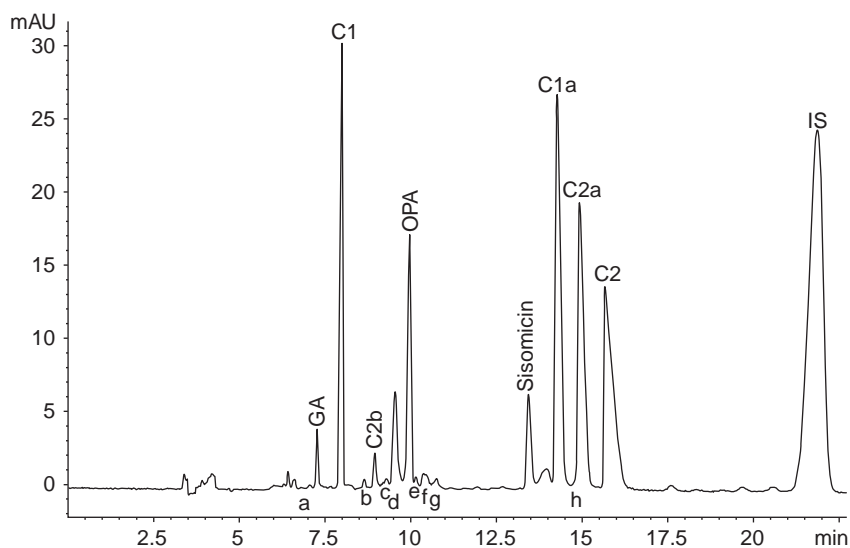
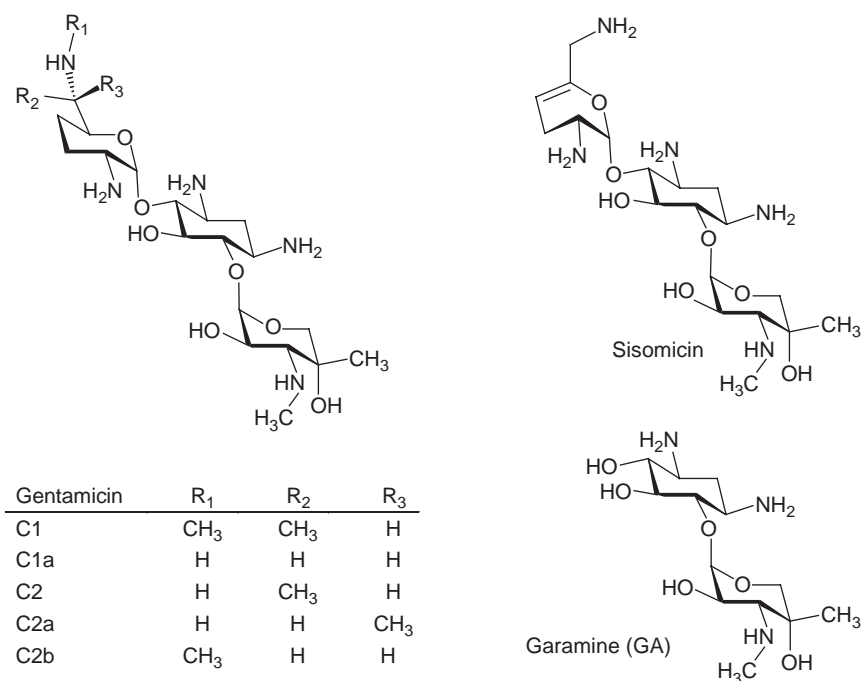
gel electrophoresis, can provide a size-based fingerprint of a product with respect to size variants from the desired product such as aggregates and fragments. Compared to classical gel electrophoresis CE provides greater automation. CE allows the detection of different isoforms and glycoforms of proteins. One example is the determination of the glycoforms of recombinant erythropoietin as one of the tests of the European Pharmacopoeia (Figure 2). In addition to quantitation data, CE can provide information on the degradation profile of peptides and proteins such as aggregation, cleavage, deamidation, or oxidation. A problem that may be encountered in protein analysis is the adsorption of proteins to the capillary wall that will greatly reduce the reproducibility and ruggedness of a method. However, several strategies have been developed to minimize wall adsorption such as operating at extreme pH values, the use of buffer additives (ionic compounds, polyamines, surfactants, etc.), and dynamic or permanent coating of the capillary surface.

## Determination of Drug-Related Impurities

In pharmaceutical analysis, the demonstration of the purity of a drug as a substance or in a formulation is

essential. Besides impurities that can be explained as reaction by-products or degradation products often unknown impurities may be present. As a high-resolution technique CE is suitable for analyzing related substances in drugs as demonstrated by a large number of sensitive, validated methods (Table 3). CZE as well as MEKC assays have been elaborated and CE methods were included in regulatory submission files. Predominantly small synthetic molecules have been analyzed but there are also examples for large compounds such as proteins. Often, identical operational parameters suitable for main component analysis can be applied to the determination of the impurities. In addition to the analysis of the purity of pharmaceuticals, CE may also be used for the profiling of illicit drugs in forensic sciences. For the determination of the chiral purity of drugs see section on chiral analysis.

Currently, regulatory agencies demand the identification and quantitation of impurities at the 0.1% level. The revised ICH guideline Q3A that came into operation in August 2002 states that impurities have to be reported if they are present above 0.05% (reporting threshold), identified if above 0.1% (identification threshold), and qualified if above 0.15% (qualification threshold). These limits apply to drugs with a maximum daily dose of 2 g per day and lower limits apply for a drug with a higher daily



**Figure 1** Electropherogram of a commercial sample of gentamicin sulfate following derivatization with  $\alpha$ -phthalaldehyde (OPA) and thioglycolic acid; IS internal standard (picric acid), a–h unknown impurities. Experimental conditions: 24.5/33 cm fused-silica capillary, 50  $\mu\text{m}$ , 100  $\text{mmol L}^{-1}$  sodium tetraborate buffer, pH 10.0, 20  $\text{mmol L}^{-1}$  sodium deoxycholate, 15  $\text{mmol L}^{-1}$   $\beta$ -cyclodextrin, 12 kV, UV detection at 340 nm. (Reproduced with permission from Wiene F and Holzgrabe U (2003) A new micellar electrokinetic capillary chromatography method for separation of the components of aminoglycoside antibiotics. *Electrophoresis* 24: 2948–2957; © Wiley-VCH.)

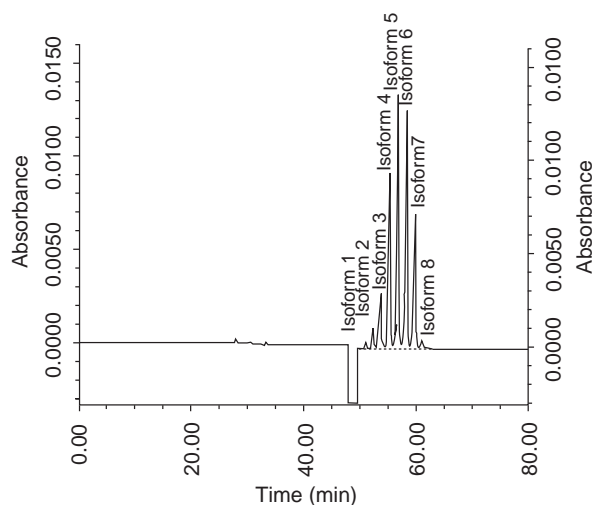
intake. **Figure 3** shows the analysis of a sample of levofolinic acid spiked with 0.1% of related substances including the diastereomer (6*R*,2'*S*)-folinic acid. CE assays with a limit of detection as low as 0.001% of the main component have been reported (**Table 3**). The identification or confirmation of unknown impurities can be performed by coupling of

CE to mass spectrometry (MS). In addition, CE and HPLC are complementary techniques due to the different respective separation mechanisms. Therefore, CE and HPLC represent a powerful combination especially for the purity determination of drugs. CE can be used to verify HPLC impurity data and vice versa.



## Determination of Inorganic Ions

Inorganic ions as counterions of acidic and basic drugs and as constituents in parenteral solutions can be analyzed by CE using indirect UV detection.



**Figure 2** Reference electropherogram of erythropoietin of the European Pharmacopoeia. (Reproduced with permission from the *European Pharmacopoeia*, 4th edn. (2002). Strasbourg, France: European Directorate for the Quality of Medicines.)

Typically, electrolytes containing imidazole or benzylamine derivatives have been employed for the determination of metal ions such as potassium, sodium, calcium, or magnesium. Complexing agents including hydroxyisobutyric acid, lactic acid, citric acid, or tartaric acid may be added as the differences in the electrophoretic mobility of metal cations in the co-electroosmotic mode are not always sufficient for a separation. The analysis may also be performed in nonaqueous electrolytes consisting of acetic acid and imidazole in methanol. Inorganic anions such as chloride, sulfate, or nitrate have been analyzed using chromate as UV-absorbing electrolyte co-ion. Tetradecyltrimethylammonium bromide or other quaternary ammonium compounds have been added in order to reverse the EOF. The organic anionic counterions acetate, succinate, maleate, citrate, or tartrate were determined by indirect UV detection using phthalate-containing buffers.

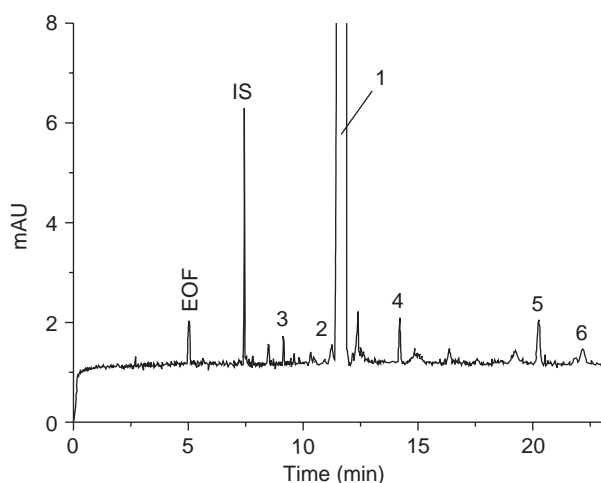
## Chiral Analysis

Chiral analysis is probably the premier application of CE due to the generally high resolution of the technique. Enantioseparations can be performed by the

**Table 3** Examples of validated capillary electrophoresis methods for the determination of related substances in pharmaceuticals

Drug	Sample	Conditions	LOD/LOQ of impurities
Alcuronium chloride	Drug substance, injection	Potassium phosphate, pH 5.5, 12 mmol l <sup>-1</sup> DM- $\beta$ -CD	LOD 0.1%, LOQ 0.2%
Calcium levofolinate	Drug substance	Sodium tetraborate, pH 9.9, 20 mg ml <sup>-1</sup> DM- $\beta$ -CD	LOD 0.025-0.05%, LOQ 0.05-0.1%
Cefalexin	Drug substance	Sodium acetate, pH 5.25, 50 mmol l <sup>-1</sup> SDS	LOD 0.05%, LOQ 0.1%
3,4-Diaminopyridine	Drug substance	Sodium phosphate, pH 2.5	LOD 0.025%, LOQ, 0.05%
Glutathione	Drug substance	Sodium phosphate, pH 1.8	LOD 0.002-0.008%, LOQ 0.005-0.02%
Kanamycin	Drug substance	Sodium tetraborate, pH 10.0, 16.0% methanol Derivatization with 1,2-phthalic dicarboxaldehyde/mercaptoacetic acid	LOQ 0.14%
Loratadine	Drug substance	Sodium phosphate, pH 2.5, 10% methanol	LOQ 0.05%
Metacycline	Drug substance	Sodium carbonate, pH 10.35, 1 mmol l <sup>-1</sup> EDTA, 13% methanol	LOD 0.024%, LOQ 0.06%
Mirtazapine	Drug substance	Sodium phosphate, pH 2.0, 25% methanol	LOD 0.02-0.04%, LOQ 0.06-0.13%
Protegrin IB-367	Drug substance	Sodium phosphate, pH 2.6	LOD 0.05%, LOQ 0.5%
Ranitidine hydrochloride	Drug substance, injection	Sodium citrate, pH 2.6	LOD 0.05%, LOQ 0.1%
Spiramycine	Drug substance	Sodium phosphate, pH 7.5, 12 mmol l <sup>-1</sup> CTAB, 20 mmol l <sup>-1</sup> sodium cholate	LOD 0.025%, LOQ 0.08%
Ximelagatran	Drug substance, tablets	Sodium phosphate, pH 1.9, 10% acetonitrile, 11 mmol l <sup>-1</sup> HP- $\beta$ -CD	LOQ 0.05%

CTAB, cetyltrimethylammonium bromide; EDTA, ethylenediaminetetraacetic acid; LOQ, limit of quantitation; LOD, limit of detection; DM- $\beta$ -CD, dimethyl- $\beta$ -cyclodextrin; HP- $\beta$ -CD, hydroxypropyl- $\beta$ -cyclodextrin.

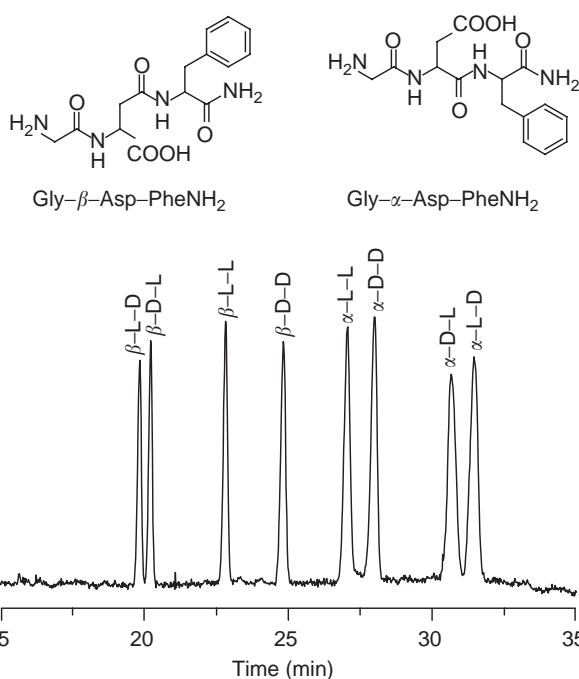


**Figure 3** Separation of levofolinic acid and related substances at the 0.1% level; 1, levofolinic acid; 2, (6*R*,2'*S*)-folinic acid; 3, dihydrofolinic acid; 4, folic acid; 5, *N*-(4-aminobenzoyl)-L-glutamic acid; 6, 10-formylc acid; IS, internal standard (methotrexate). Experimental conditions: 40/50 cm fused-silica capillary, 50  $\mu\text{m}$ , 40 mmol  $\text{L}^{-1}$  sodium tetraborate buffer, pH 9.9, 20 mg  $\text{mL}^{-1}$  heptakis-(2,6-di-*O*-methyl)- $\beta$ -cyclodextrin, 21 kV, UV detection at 214 nm. (Reproduced with permission from Süß F, Harang V, Sängers-van de Griend CE, and Scriba GKE (2004) Development and validation of a robust capillary electrophoresis method for impurity profiling of calcium levofolinate including the (6*R*,2'*S*)-diastereomer using statistical experimental design. *Electrophoresis* 25: 766–777; © Wiley-VCH.)

indirect method upon derivatization with a stereochemically pure agent to form diastereomers that are subsequently separated in an achiral system. The direct enantioseparation, which is by far the most popular technique in chiral CE, is based on the formation of transient diastereomeric complexes between the analyte enantiomers and an optically pure chiral selector added to the background electrolyte.

While the migration principle, i.e., the driving forces moving the analytes through the separation capillary, is based on electrophoretic mechanisms the chiral separation is based on enantioselective interactions between the analyte enantiomers and a chiral selector and is, therefore, a chromatographic separation principle. The fact that the selector is in the same phase as the analytes in CE and not part of a stationary phase that is immiscible with the mobile phase as found in chromatography does not represent a conceptional difference between both techniques. The chiral selector in CE is also called pseudophase as it is not a physically different phase and may also possess an electrophoretic mobility. Enantioseparations in CE have also been termed 'capillary electrokinetic chromatography'.

The chiral selectors applied to CE include native cyclodextrins as well as neutral and charged derivatives, oligo- and polysaccharides, chiral crown ethers,



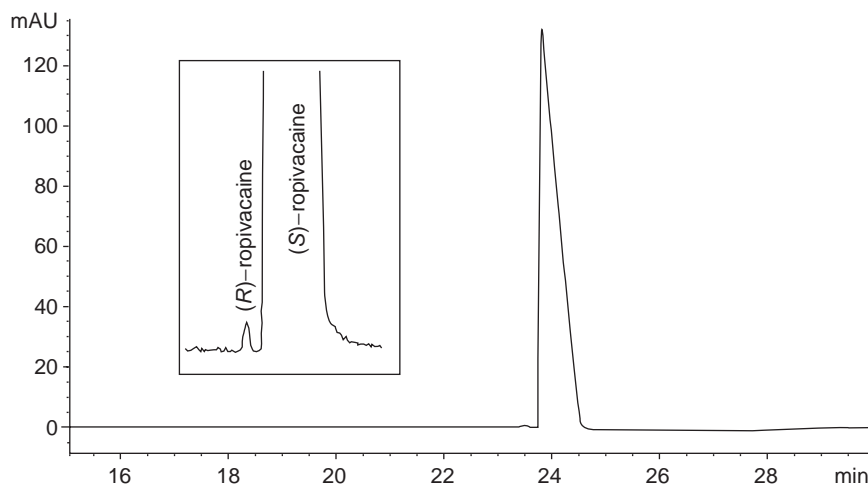
**Figure 4** Simultaneous chiral separation of the isomeric tripeptides Gly- $\alpha$ -Asp-PheNH<sub>2</sub> and Gly- $\beta$ -Asp-PheNH<sub>2</sub>. Experimental conditions: 40/47 cm polyacrylamide-coated capillary, 50  $\mu\text{m}$ , 50 mmol  $\text{L}^{-1}$  sodium phosphate buffer, pH 5.25, 60 mg  $\text{mL}^{-1}$  carboxymethyl- $\beta$ -cyclodextrin, –20 kV, UV detection at 215 nm. (Reprinted with permission from Sabah S and Scriba GKE (1998) Electrophoretic stereoisomer separation of aspartyl dipeptides and tripeptides in untreated fused-silica and polyacrylamide-coated capillaries using charged cyclodextrins. *Journal of Chromatography A* 822: 137–145; © Elsevier.)

macrocytic glycopeptide antibiotics, proteins and synthetic cyclopeptides, calixarenes, as well as chiral surfactants derived from steroids, amino acids, tartaric acid, or glycosides. In addition, ligand exchange and chiral ion-pairing reagents have been used for enantioseparations. Among these compounds, cyclodextrins are by far the most common chiral selectors applied to enantioseparations. Numerous neutral and charged cyclodextrin derivatives are commercially available. Charged cyclodextrins also allow the analysis of neutral compounds. The resolving power of chiral CE is demonstrated by the simultaneous separation of the enantiomers of the isomeric tripeptides Gly- $\alpha$ -Asp-PheNH<sub>2</sub> and Gly- $\beta$ -Asp-PheNH<sub>2</sub> (Figure 4).

For method optimization of chiral separations the type and concentration of the chiral selector has to be specifically considered in addition to factors also important in achiral CE such as pH, molarity and type of the background electrolyte, and buffer additives. When using randomly substituted cyclodextrins as chiral selectors testing of different batches of the cyclodextrin and samples from different suppliers is

recommended to ensure sufficient robustness of the method. In addition to aqueous background electrolytes, the use of nonaqueous solvents is becoming increasingly popular. Nonaqueous solvents may offer the advantage of increased solubility of chiral selectors and samples, lower Joule heating, or reduced solute-wall interactions. Moreover, an increase in selectivity can often be observed in nonaqueous media. The modes of chiral CE separations include partial filling and counter-current techniques, carrier-mode separations, mobility counterbalanced separations, etc.

Chiral CE has been applied to the determination of the chiral purity of drugs and racemate composition. Drug substances and pharmaceutical formulations have been analyzed and chiral CE methods have been included in regulatory submission files. As an example, the determination of the chiral purity of (*S*)-ropivacaine is shown in Figure 5. Further examples of chiral CE drug analysis using cyclodextrins as chiral selectors listed in Table 4 underline the performance capability of CE methods. ICH guideline Q3A on impurities in new drug substances explicitly



**Figure 5** Limit of quantitation of (*R*)-ropivacaine in a ropivacaine hydrochloride injection. Experimental conditions: 72.0/80.5 cm fused silica capillary, 50  $\mu\text{m}$ , 0.1  $\text{mol l}^{-1}$  phosphoric acid adjusted to pH 3.0 with triethanolamine, 10  $\text{mmol l}^{-1}$  heptakis-(2,6-di-*O*-methyl)- $\beta$ -cyclodextrin, 30 kV, UV detection at 206 nm. (Reprinted with permission from Sanger-van de Griend CE, Wahlstrom H, Groningsson K, and Widahl-Nasman M (1997) A chiral capillary electrophoresis method for ropivacaine hydrochloride in pharmaceutical formulations: validation and comparison with chiral liquid chromatography. *Journal of Pharmaceutical and Biomedical Analysis* 15: 1051–1061;  Elsevier.)

**Table 4** Examples of the validated capillary electrophoresis methods for the determination of the chiral purity of drugs

Drug	Selector	Background electrolyte	LOD/comment
DOPA, carbidopa	S- $\beta$ -CD	Sodium phosphate, pH 2.45	0.02% D-enantiomer
Fentoconazole	TM- $\beta$ -CD	Sodium phosphate, pH 3.0	0.2% minor enantiomer, separation of related substances
Galantamine	$\alpha$ -CD	Sodium phosphate, pH 3.0	0.04% minor enantiomer, separation of related substances
L-Hyoscyamine	S- $\beta$ -CD	Sodium phosphate, pH 8.8	0.25% D-hyoscyamine
Ketoprofen	TM- $\beta$ -CD	Sodium acetate, pH 5.0	0.04% ( <i>R</i> )-ketoprofen, stability study of oral solution
Melagatran	DM- $\beta$ -CD	Sodium phosphate, pH 1.8, 10% methanol	0.013% ( <i>S,R</i> )-enantiomer
Methyl dopa	SB- $\beta$ -CD	Sodium phosphate, pH 2.4	0.05% ( <i>R</i> )-methyl dopa, separation of related substances
Propranolol	CM- $\beta$ -CD/DM- $\beta$ -CD	Tris-phosphate, pH 3.0	0.1% ( <i>R</i> )-propranolol
Ropivacaine	DM- $\beta$ -CD	Triethanolamine-phosphate, pH 3.0	0.05% ( <i>R</i> )-ropivacaine, analysis of injection solution
Selegiline	$\beta$ -CD polymer	Tris-phosphate, pH 2.5	0.25% ( <i>S</i> )-enantiomer
Terbutaline	HP- $\beta$ -CD	Sodium phosphate, pH 2.5	0.1% (+)-enantiomer
E-6232	SB- $\beta$ -CD/DM- $\beta$ -CD	Sodium tetraborate, pH 9.2, 10% methanol	0.03% ( <i>R</i> )-enantiomer, separation of related substances

LOD, limit of detection;  $\alpha$ -CD,  $\alpha$ -cyclodextrin,  $\beta$ -CD,  $\beta$ -cyclodextrin; CM- $\beta$ -CD, carboxymethyl- $\beta$ -cyclodextrin; DM- $\beta$ -CD, dimethyl- $\beta$ -cyclodextrin; HP- $\beta$ -CD, hydroxypropyl- $\beta$ -cyclodextrin; TM- $\beta$ -CD, trimethyl- $\beta$ -cyclodextrin, S- $\beta$ -CD, sulfated  $\beta$ -cyclodextrin, SB- $\beta$ -CD, sulfobutyl- $\beta$ -cyclodextrin.

excludes enantiomeric impurities from the scope but guideline Q6A on test procedures and acceptance criteria for new drug substances states that for chiral drugs which are developed as a single enantiomer control of the other enantiomer should be considered in the same manner as for other impurities. In addition to enantiomer analysis, the methods often allow the simultaneous determination of chiral and achiral related substances.

## Bioanalysis

The analysis of drugs, their metabolites, and other exogenous compounds in body fluids and tissue samples is another important aspect of pharmaceutical analysis. Drug metabolism and pharmacokinetics contribute to the optimization of drug therapy in addition to providing fundamental data for drug discovery, development, and regulatory aspects. Because of its high efficiency CE is especially suited in the case of a complex matrix or to address separation problems that cannot be solved by chromatography. Target applications include stereoisomer separations, analysis in the case of very small sample volumes, assays that lack specificity, or methods requiring large amounts of (organic) solvents. In addition, CE has been applied to screening of body fluids and hair for licit or illicit drugs in clinical and forensic samples.

Sample preparation may be absent or include simple liquid handling operations (filtration, centrifugation), removal of endogenous constituents of the matrix (protein precipitation, ultrafiltration), and for selectivity and/or sensitivity enhancement extraction techniques (liquid–liquid extraction, solid-phase extraction) as well as derivatization. Microdialysis samples and urine can be injected onto the capillary in the case of sufficiently high drug concentrations in these samples but the inherent high ionic strength of urine can cause problems and require dilution prior to the injection. MEKC employing SDS often allows the direct injection of protein-containing fluids such as plasma and serum. SDS solubilizes the proteins and prevents their adsorption to the capillary wall, which causes deterioration of the analytical system due to an irreproducible EOF, drifting migration times, and peak asymmetry. A protocol for the direct injection of plasma with subsequent removal of adsorbed protein by rinsing with a SDS solution between runs has been reported. To avoid protein-associated problems protein precipitation by acetonitrile or trifluoroacetic acid can be performed followed by injection of the supernatant upon centrifugation. This operation simplifies the matrix and liberates protein-bound drugs but dilution of the sample occurs. When acetonitrile is used the inherent

electrokinetic solute concentration effect counteracts sample dilution allowing detection limits of the drugs comparable to the direct injection of the sample. Liquid–liquid extraction and solid-phase extraction simultaneously concentrate the solutes by one to two orders of magnitude and increase selectivity by simplification of the sample matrix.

In contrast to the determination of the main component or impurities in drug substances or pharmaceutical formulations where sample concentration is not a major issue, the limited sensitivity of CE may hamper its application to bioanalysis. The lowest detectable amount of compounds using UV detection as the most common detection mode in CE is in the range of  $1\text{--}10\ \mu\text{mol l}^{-1}$ , which is one to two orders of magnitude lower compared to HPLC with UV detection. Increasing the optical path using bubble or Z-shaped detection cells results in a 10-fold improvement, but at some loss of separation efficiency. Employing laser-induced fluorescence detection enhances assay sensitivity by a factor of 100–1000 compared to UV detection and also adds specificity to the assay as not all compounds show fluorescence. However, the approach is limited to analytes that can be excited at the wavelength of the commercially available He–Cd laser (325 nm) or argon-ion laser (488 nm). Hyphenation of CE and MS (CE–MS) yields structural information on drugs and their metabolites. However, while HPLC–MS is more sensitive compared to HPLC with UV detection, CE–MS is less sensitive than UV due to the construction of the currently used CE–MS interfaces employing sheath liquids to increase the CE efflux, which results in a dilution of the sample.

Sample concentration is another way to increase the sensitivity of a method. Besides liquid–liquid and solid-phase extraction protocols, which result in more concentrated solutions in addition to sample cleanup, several electrokinetic concentration techniques can be applied including sample stacking and field-amplified sample injection/stacking. Over 1000-fold sensitivity enhancement compared to hydrodynamic sample injection without stacking has been reported. When acetonitrile has been used for the removal of proteins from plasma, sample injections of up to 50% of the capillary volume become possible. Stacking techniques are also available for MEKC.

Numerous validated achiral and chiral assays of drugs in biological matrices have been reported (Table 5). The assay of lamotrigine was adapted to routine analysis with multilevel internal calibration on different commercial instruments. Evaluation of the calibration and control data from 4 years as well as cross-validation with data from different laboratories clearly demonstrated the suitability of

**Table 5** Examples of the validated capillary electrophoresis methods for the determination of drugs in biological matrices

Drug	Matrix	Extraction	Conditions	Application
Albendazole sulfoxide	Plasma	LLE	Tris-phosphate, pH 7.0, 3% S- $\beta$ -CD	Patient samples
Amphetamine, ecstasy, and analogs	Urine	LLE	Potassium phosphate, pH 2.5, 8.3 mmol l <sup>-1</sup> HP- $\beta$ -CD	Urinary excretion, patients and drug abusers
Cilazapril and metabolites	Urine	SPE	Sodium tetraborate, pH 9.5	Patient samples
Ciprofibrate, ciprofibrate glucuronide	Urine	SPE	Sodium phosphate, pH 6.0, 7.5 mmol l <sup>-1</sup> $\gamma$ -CD	Urinary excretion, stereoselective metabolism
Doxorubicin and metabolite, daunorubicin and metabolite	Serum	LLE/acetonitrile protein precipitation	Sodium phosphate, pH 5.0, 60 $\mu$ mol l <sup>-1</sup> spermine, 70% acetonitrile	Patient samples
Ibuprofen	Plasma	LLE	Triethanolamine-phosphate, pH 5.0, 2% S- $\beta$ -CD	Stereoselective pharmacokinetics
Ketoprofen	Serum	LLE	Triethanolamine-phosphate, pH 5.0, 50 mmol l <sup>-1</sup> TM- $\beta$ -CD	Stereoselective pharmacokinetics
Lamivudine/didanosine/saquinavir	Serum	SPE	N,N-Dimethyloctylamine-phosphoric acid, pH 2.5	Patient samples
Lamotrigine	Serum	Acetonitrile protein precipitation	Sodium acetate, pH 4.5	Patient samples
Mianserin	Plasma	LPME	Triethylamine-phosphate, pH 5.0, 2 mmol l <sup>-1</sup> HP- $\beta$ -CD	Patient samples
Ofloxacin and metabolites	Urine	Direct injection	Triethylamine-phosphate, pH 2.0, 0.3 mmol l <sup>-1</sup> SB- $\beta$ -CD	Urinary excretion, stereoselective metabolism
Quinapril and metabolites	Urine	SPE	Sodium tetraborate, pH 9.5	Patient samples
Tramadol and metabolites	Urine	LLE	Sodium tetraborate, pH 10.1, 30 mg ml <sup>-1</sup> CM- $\beta$ -CD	Urinary excretion, stereoselective metabolism
Trimebutine maleate	Serum, tissue	Acetonitrile protein precipitation	Sodium phosphate, pH 6.0	Pharmacokinetics and tissue distribution
Zaleplon and metabolites	Urine	SPE	Tris-borate, pH 9.4, 50 mmol l <sup>-1</sup> CM- $\beta$ -CD	Urinary excretion

LLE, liquid–liquid extraction; SPE, solid-phase extraction; LPME, liquid-phase microextraction; S- $\beta$ -CD, sulfated  $\beta$ -cyclodextrin; HP- $\beta$ -CD, hydroxypropyl- $\beta$ -cyclodextrin; SB- $\beta$ -CD, sulfobutyl- $\beta$ -cyclodextrin, CM- $\beta$ -CD, carboxymethyl- $\beta$ -cyclodextrin; TM- $\beta$ -CD, trimethyl- $\beta$ -cyclodextrin.

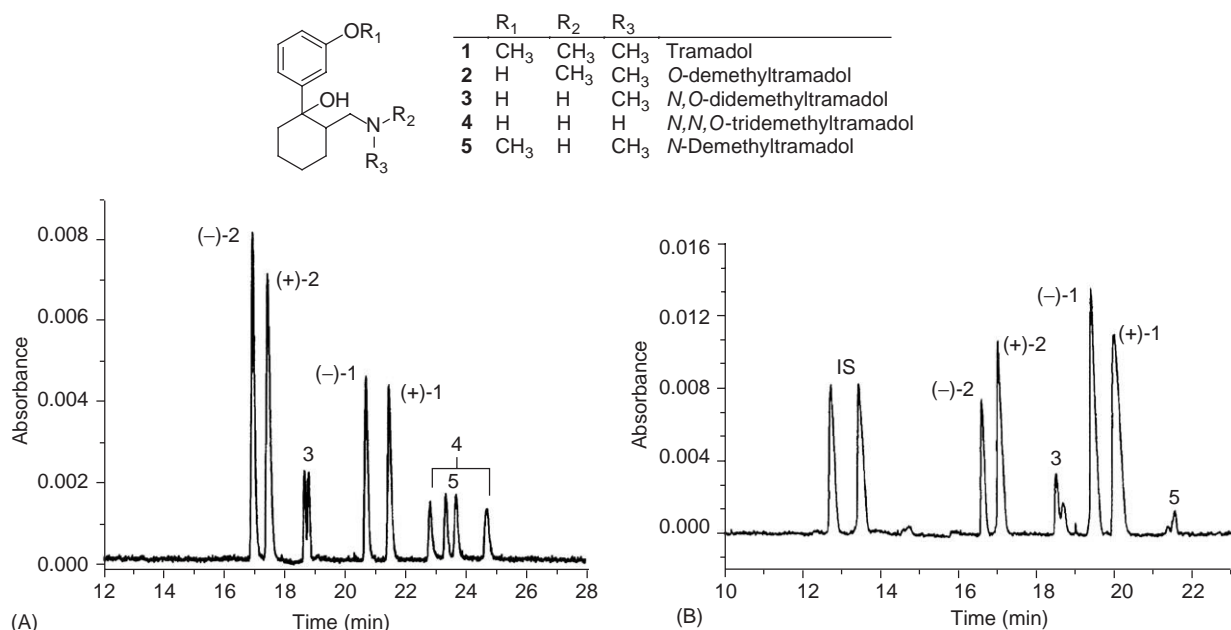
CE-based assays for therapeutic drug monitoring. Despite the fact that HPLC is currently the most widely applied method for drug bioanalysis, CE possesses a clear advantage over HPLC for the chiral analysis of drugs as well as the determination of drugs and their metabolites (Figure 6). The simultaneous analysis of drugs and their phase-II glucuronide and sulfate metabolites is hardly possible by HPLC. Furthermore, CE may be the method of choice when only small sample volumes are available such as blood samples from children and infants or *in vivo* microdialysates.

## Miscellaneous Applications

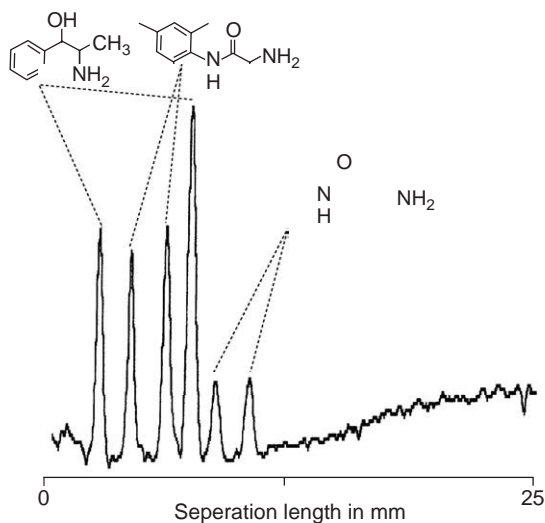
In addition to the applications described above, CE has been applied to reaction monitoring, drug stability studies, the determination of content release testing from tablets, as well as the analysis of drugs in

cell culture-based transport studies. Furthermore, CE can be employed for the determination of physicochemical properties of compounds such as  $pK_a$  values and lipophilicity. The advantage compared to classical methods is the low consumption of sample and the fact that the compounds do not have to be pure and mixtures of compounds can be utilized. CE is also well suited to study the interactions between ligands and biomolecules such as antibody–antigen interactions or the affinity of drugs to proteins, excipients, or liposomes including stereospecific aspects. Affinity CE, frontal analysis, the Hummel–Dreyer method, and vacancy peak analysis have been applied for this purpose. These analyses do not require pure compounds either and compound libraries can be screened. Moreover, complexation constants have been determined by CE methods.

Current trends in pharmaceutical applications of CE include the use of microchips as well as capillary



**Figure 6** Chiral capillary electrophoresis analysis of tramadol and its metabolites in urine. (A) Separation of reference compounds, (B) analysis of a urine sample collected 6–8 h after oral administration of 100 mg tramadol. The internal standard (IS) was a chiral analog that is also resolved. Experimental conditions: 50/57 cm fused silica capillary, 50  $\mu$ m, 50 mmol l<sup>-1</sup> sodium borate buffer, pH 10.1, 30 mg ml<sup>-1</sup> carboxymethyl- $\beta$ -cyclodextrin, 20 kV, UV detection at 214 nm. (Adapted with permission from Kurth B and Blaschke G (1999) Achiral and chiral determination of tramadol and its metabolites in urine by capillary electrophoresis. *Electrophoresis* 20: 555–563; © Wiley-VCH.)



**Figure 7** Simultaneous chiral separation of three basic drugs. Only about half of the total length of 25 mm of the microchip separation channel was needed to obtain complete resolution in 11 s. Experimental conditions: 25 mmol l<sup>-1</sup> triethylammonium phosphate buffer, pH 2.5, 5% highly sulfated  $\gamma$ -cyclodextrin, UV detection at 200 nm. (Reproduced with permission from Ludwig M, Kohler F, and Belder D (2003) High-speed chiral separations on a microchip with UV detection. *Electrophoresis* 24: 3233–3238; © Wiley-VCH.)

electrochromatography (CEC). CE on microchip devices is a rapidly emerging technology having the potential of analyzing hundreds of samples within minutes. Commercial instruments and different

detection modes as well as the coupling to a mass spectrometer are available. While the premier application of microchip CE still is nucleic acid and protein analysis the determination of peptides and pharmaceutical drugs including bioanalysis and chiral separations by this technique have been reported. **Figure 7** shows the simultaneous chiral separation of three basic drugs on a microchip in 11 s.

CEC is considered a hybrid technique between HPLC and CE combining the high peak efficiency of CE with the high separation selectivity of multivariate stationary phases available in HPLC. The general applicability of CEC to the analysis of drugs including chiral separations has been demonstrated, but it has not yet been implemented in the ‘arsenal’ of pharmaceutical analytical techniques.

See also: **Capillary Electrophoresis:** Overview; Clinical Applications. **Pharmaceutical Analysis:** Overview; Drug Purity Determination.

## Further Reading

- Altria KD (1998) *Analysis of Pharmaceuticals by Capillary Electrophoresis*. Wiesbaden, Germany: Vieweg.
- Altria KD, Chen AB, and Clohs L (2001) Capillary electrophoresis as a routine analytical tool in pharmaceutical analysis. *LCGC Europe* 19: 972–985.

- Blaschke G and Chankvetadze B (2000) Enantiomer separation of drugs by capillary electrophoresis. *Journal of Chromatography A* 875: 3–25.
- Blomberg LG and Wan H (2000) Determination of enantiomeric excess by capillary electrophoresis. *Electrophoresis* 21: 1940–1952.
- Chen A, Nashabeh W, and Wehr T (2001) *CE in Biotechnology: Practical Applications for Protein and Peptide Analyses*. Wiesbaden, Germany: Vieweg.
- Dolnik V, Liu S, and Javanovich S (2000) Capillary electrophoresis on microchip. *Electrophoresis* 21: 41–54.
- Heegaard NHH and Kennedy RT (1999) Identification, quantitation, and characterization of biomolecules by capillary electrophoretic analysis of binding interactions. *Electrophoresis* 20: 3122–3133.
- Hilhorst MJ, Somsen GW, and De Jong GJ (2001) Capillary electrokinetic separation techniques for profiling of drugs and related products. *Electrophoresis* 22: 2542–2564.
- Nishi H (1999) Capillary electrophoresis of drugs: current status in the analysis of pharmaceuticals. *Electrophoresis* 20: 3237–3258.
- Scriba GKE (2002) Selected fundamental aspects of chiral electromigration techniques and their application to pharmaceutical and biomedical analysis. *Journal of Pharmaceutical and Biomedical Analysis* 27: 373–399.
- Shaw CJ and Guzman NA (2002) Applications of capillary electrophoresis technology in the pharmaceutical industry. In: Ohannesian L and Streeter AJ (eds.) *Handbook of Pharmaceutical Analysis*, pp. 313–386. New York: Dekker.
- Thormann W (2002) Progress of capillary electrophoresis in therapeutic drug monitoring and clinical and forensic toxicology. *Therapeutic Drug Monitoring* 24: 222–231.
- Thormann W, Way AB, Lurie IS, *et al.* (1999) Capillary electrophoresis in clinical and forensic analysis: Recent advances and breakthrough to routine applications. *Electrophoresis* 20: 3203–3236.
- Wätzig H, Degenhardt M, and Kunkel A (1998) Strategies for capillary electrophoresis. Method development and validation for pharmaceutical and biological applications. *Electrophoresis* 19: 2695–2752.

## Low-Molecular-Weight Ions

**V Pacáková and K Štulík**, Charles University, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Low-molecular-weight (LMW) ions include both inorganic and organic compounds. The main application area of capillary electrophoresis lies in analyses for inorganic and organic anions and cations in environmental samples, in clinical chemistry, pulp and paper industry, process control, industrial applications, explosive residue analysis, biological samples, or drugs and drug intermediates. LMW organic acids are important intermediates or final metabolites of many biochemical pathways in living organisms, as well as degradative metabolites of amino acids, fats, and carbohydrates. LMW organic cations, e.g., quaternary ammonium compounds, are widely used in industry as antiseptic, antistatic, and antimicrobial agents.

In contrast to determinations of organic analytes, where the use of high-performance separations is almost inevitable in complex matrices, inorganic analysis has powerful tools in highly selective and sensitive spectroscopic methods and thus a separation step is often unnecessary. This is especially true for determination of inorganic cations. However, the

selection of methods for determination of inorganic anions is much more limited.

The traditional approach to high-performance separations of LMW ions, ion chromatography (IC), is partially being replaced by capillary electrophoresis (CE) owing to the following main advantages of the CE: a higher separation efficiency caused by a flat local velocity profile in CE running buffers compared to parabolic profiles in IC; a higher speed of analysis caused by the absence of a partition process between the mobile and stationary phases (except for micellar electrokinetic chromatography (MEKC) where the analytes are distributed between a pseudostationary phase and the electrolyte, and capillary electrochromatography employing capillaries packed with stationary phase particles); very small amounts of sample required permit, e.g., an analysis for cations in a single small rat or mice eye lens or analysis for cations and anions present in a single rain drop; very low mass detection limits; low consumption of chemicals (low cost per analysis); relatively simple instrumentation, easy automation, and good tolerance to sample matrix, e.g., to high pH values.

The disadvantages of CE compared to IC include lower sensitivity, higher concentration detection limits, and somewhat poorer reproducibility of qualitative and quantitative data owing to instability of the electroosmotic flow (EOF). IC has so far been developed more extensively. Validated procedures and computer optimization approaches are available



- Blaschke G and Chankvetadze B (2000) Enantiomer separation of drugs by capillary electrophoresis. *Journal of Chromatography A* 875: 3–25.
- Blomberg LG and Wan H (2000) Determination of enantiomeric excess by capillary electrophoresis. *Electrophoresis* 21: 1940–1952.
- Chen A, Nashabeh W, and Wehr T (2001) *CE in Biotechnology: Practical Applications for Protein and Peptide Analyses*. Wiesbaden, Germany: Vieweg.
- Dolnik V, Liu S, and Javanovich S (2000) Capillary electrophoresis on microchip. *Electrophoresis* 21: 41–54.
- Heegaard NHH and Kennedy RT (1999) Identification, quantitation, and characterization of biomolecules by capillary electrophoretic analysis of binding interactions. *Electrophoresis* 20: 3122–3133.
- Hilhorst MJ, Somsen GW, and De Jong GJ (2001) Capillary electrokinetic separation techniques for profiling of drugs and related products. *Electrophoresis* 22: 2542–2564.
- Nishi H (1999) Capillary electrophoresis of drugs: current status in the analysis of pharmaceuticals. *Electrophoresis* 20: 3237–3258.
- Scriba GKE (2002) Selected fundamental aspects of chiral electromigration techniques and their application to pharmaceutical and biomedical analysis. *Journal of Pharmaceutical and Biomedical Analysis* 27: 373–399.
- Shaw CJ and Guzman NA (2002) Applications of capillary electrophoresis technology in the pharmaceutical industry. In: Ohannesian L and Streeter AJ (eds.) *Handbook of Pharmaceutical Analysis*, pp. 313–386. New York: Dekker.
- Thormann W (2002) Progress of capillary electrophoresis in therapeutic drug monitoring and clinical and forensic toxicology. *Therapeutic Drug Monitoring* 24: 222–231.
- Thormann W, Way AB, Lurie IS, *et al.* (1999) Capillary electrophoresis in clinical and forensic analysis: Recent advances and breakthrough to routine applications. *Electrophoresis* 20: 3203–3236.
- Wätzig H, Degenhardt M, and Kunkel A (1998) Strategies for capillary electrophoresis. Method development and validation for pharmaceutical and biological applications. *Electrophoresis* 19: 2695–2752.

## Low-Molecular-Weight Ions

**V Pacáková and K Štulík**, Charles University, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Low-molecular-weight (LMW) ions include both inorganic and organic compounds. The main application area of capillary electrophoresis lies in analyses for inorganic and organic anions and cations in environmental samples, in clinical chemistry, pulp and paper industry, process control, industrial applications, explosive residue analysis, biological samples, or drugs and drug intermediates. LMW organic acids are important intermediates or final metabolites of many biochemical pathways in living organisms, as well as degradative metabolites of amino acids, fats, and carbohydrates. LMW organic cations, e.g., quaternary ammonium compounds, are widely used in industry as antiseptic, antistatic, and antimicrobial agents.

In contrast to determinations of organic analytes, where the use of high-performance separations is almost inevitable in complex matrices, inorganic analysis has powerful tools in highly selective and sensitive spectroscopic methods and thus a separation step is often unnecessary. This is especially true for determination of inorganic cations. However, the

selection of methods for determination of inorganic anions is much more limited.

The traditional approach to high-performance separations of LMW ions, ion chromatography (IC), is partially being replaced by capillary electrophoresis (CE) owing to the following main advantages of the CE: a higher separation efficiency caused by a flat local velocity profile in CE running buffers compared to parabolic profiles in IC; a higher speed of analysis caused by the absence of a partition process between the mobile and stationary phases (except for micellar electrokinetic chromatography (MEKC) where the analytes are distributed between a pseudostationary phase and the electrolyte, and capillary electrochromatography employing capillaries packed with stationary phase particles); very small amounts of sample required permit, e.g., an analysis for cations in a single small rat or mice eye lens or analysis for cations and anions present in a single rain drop; very low mass detection limits; low consumption of chemicals (low cost per analysis); relatively simple instrumentation, easy automation, and good tolerance to sample matrix, e.g., to high pH values.

The disadvantages of CE compared to IC include lower sensitivity, higher concentration detection limits, and somewhat poorer reproducibility of qualitative and quantitative data owing to instability of the electroosmotic flow (EOF). IC has so far been developed more extensively. Validated procedures and computer optimization approaches are available

**Table 1** Equivalent ionic conductivities ( $\Lambda_{eq}$ ) and ionic mobilities ( $m_i$ ) of selected inorganic cations and anions

Analyte	$\Lambda_{eq} (cm^2 \Omega^{-1} mol^{-1})$	$m_i (10^{-4} cm^2 V^{-1} s^{-1})^a$	$m_i (10^{-4} cm^2 V^{-1} s^{-1})$
$NH_4^+$	73.5	7.62	7.62
$K^+$	73.5	7.62	7.62
$Ca^{2+}$	59.5	6.17	6.17
$Mg^{2+}$	53.0	5.49	5.50
$Na^+$	50.1	5.19	5.19
$SO_4^{2-}$	80.0	8.29	8.29
$Cl^-$	76.3	7.91	7.91
$NO_2^-$	71.8	7.44	7.46
$NO_3^-$	71.4	7.40	7.41
$PO_4^{3-}$	69.0	7.15	7.15
$F^-$	55.4	5.74	5.74

<sup>a</sup> Calculated from the equivalent ionic conductivities.

in IC, while routine applications of CE are still less common. The selectivity range in CE is limited, as the selectivity can only be manipulated by the electrolyte composition. CE and IC are complementary rather than competing techniques and exhibit different selectivity; CE is used for separation of those mixtures of anions and cations that are difficult to separate by IC and vice versa.

## Theoretical Consideration

The electrophoretic mobility of an ion,  $\mu_{ep(ion)}$ , can be related to the limiting ionic equivalent conductivity,  $\lambda_{eq}$ , by

$$\mu_{ep(ion)} = \lambda_{eq}/F = q_i/6\pi\eta r_i \quad [1]$$

where  $F$  is the Faraday constant ( $F = 9.6487 \times 10^4 \text{ A s mol}^{-1}$ );  $\lambda_{eq} (cm^2 mol^{-1} \Omega^{-1})$  is related, by the Stokes law, to the charge of the hydrated ion,  $q_i$ , to the dynamic viscosity of the electrolyte,  $\eta (g cm^2 s^{-1})$ , and to the radius of the hydrated ion,  $r_i (cm)$ .

The  $\mu_{ep(ion)}$  values can be calculated from the experimental data, the apparent mobility of the ion,  $\mu_{app(ion)}$ , and the mobility of the EOF,  $\mu_{eo}$ , according to

$$\begin{aligned} \mu_{ep(ion)} &= \mu_{app(ion)} - \mu_{eo} \\ &= (1/t_{m(ion)} - 1/t_{m(eo)})(L_T L_D/V) \end{aligned} \quad [2]$$

where  $t_{m(ion)}$  and  $t_{m(eo)}$  (both in seconds) are the migration times of the ion, and of an EOF marker (an uncharged solute), respectively;  $L_T$  and  $L_D$  (both in centimeters) are the overall capillary length and the length of the capillary to the detector, respectively;  $V$  is the voltage (in volts).

The limiting ionic conductivities and the experimental and calculated electrophoretic mobilities of some anions and cations are compared in Table 1.

The electrophoretic mobilities derived from the limiting ionic equivalent conductivity differ somewhat from the experimentally measured values that are dependent on the composition of the background electrolyte and its pH. Differences in the ionic mobilities as small as  $0.1 \times 10^{-9} m^2 V^{-1} s^{-1}$  are sufficient for the separation of ionic species, provided that the separation is highly efficient. As follows from Table 1, all the ions can be separated except for  $K^+$  and  $NH_4^+$  whose mobilities are identical.

## Optimization of Separation

The running buffer composition is of primary importance to CE selectivity optimization. The buffer contains at least one anion and one cation, and one of these ions should have an adequate buffering capacity. If possible, the EOF should have the same direction as the migration of the analytes to shorten the analysis. The co-ion should have a similar mobility as the analyte to ensure a good peak shape.

The separation is optimized by changing the composition and concentration of the running buffer and by adjusting its pH. As follows from eqn [1], the electrophoretic mobility of ions depends on their charge-to-mass ratio. For weak acids and bases this ratio can be changed by changing the pH in the vicinity of the analyte  $pK_a$ . The effect of the pH on the separation of weak acids can be demonstrated on an example of separation of a high concentration of phosphate (more than  $800 \mu g l^{-1}$ ) from a low concentration of fluoride ( $1 \mu g l^{-1}$ ). Protonation of hydrogenphosphate at a pH of 7 results in its slower migration and leads to an improved separation from fluoride. Another example is a very fast separation of nitrate and nitrite within 10 s at pH 2.5.

An addition of organic solvent to the background electrolyte changes the separation selectivity. This

can be explained by changes in the relative hydration of ions. The organic solvent destroys the hydrated layer and thus changes the effective mass of the ions. Ions such as iodide and chloride, which are difficult to separate in an electrolyte consisting of pyromellitic acid and hexamethonium hydroxide, can be separated after addition of methanol. Organic solvents added to the running electrolyte decrease the EOF; they increase the viscosity and decrease the  $pK_a$  of the silanol groups on the capillary wall and improve the reproducibility of the migration times, with a less noisy baseline. The temperature also affects the selectivity, as it influences both the mobilities and the EOF through a change in the solution viscosity.

### Cations – Direct Analysis

Inorganic cations are smaller and thus have higher charge densities ( $q_i/r_i$  ratios) than most organic ions; therefore, as follows from eqn [1], their electrophoretic mobilities are higher. The problems connected with CE analysis of inorganic cations are caused by small differences in their migration rates (see Table 1), and, similar to the CE analysis of inorganic anions, by their low absorption of ultraviolet (UV) radiation, which complicates detection.

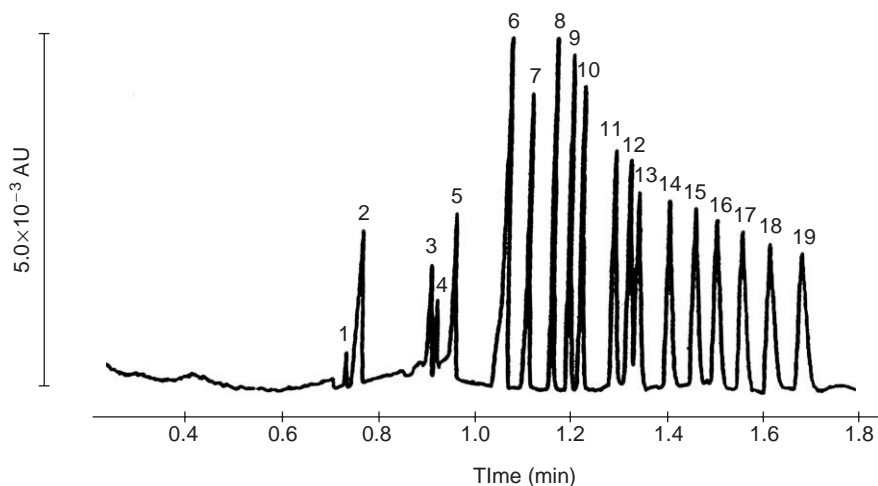
Only the alkali metal ions exhibit large differences in their mobilities and thus can easily be separated within very short times (less than 2 min), if the directions of the EOF and of the cation migration are

the same, toward the cathode. Exceptional is the separation of  $K^+$  and  $NH_4^+$  ions whose mobilities are identical at a slightly acidic pH. Their separation can be attained in alkaline buffers as the ammonium ions are less protonated and their mobility decreases while the  $K^+$  ion mobility is not affected by the pH. Quaternary ammonium ions can be separated in co-electroosmotic mode. However, to have a sufficiently large separation window, the EOF has to be decreased, e.g., by addition of organic solvents.

### Cations – Separation of Complexes

Complexation with auxiliary ligands is used to promote cation separation. Different degrees of complexation lead to different migration times. The migration can thus be influenced by the type and the concentration of the complexing agent, pH, ionic strength, viscosity, etc.

Several complexing agents have been proposed and successfully tested in analyses of cations. Weak complexing agents, such as lactate, phthalate, tartrate, or hydroxyisobutyric acid (for separation of lanthanoids), have been used. For example, 19 alkali, alkaline earth, and rare earth metal ions can be separated in very short time (less than 2 min), using hydroxyisobutyric acid as a complexing agent and an indirect UV detection method (Figure 1). Other complexing agents, such as ethylenediaminetetraacetic acid (EDTA), diaminocyclohexanetetraacetic acid, and 18-crown-6, have also been used.



**Figure 1** Electropherogram of the separation of alkali, alkaline earth, and lanthanide metal ions. Capillary, 36.5 cm, 75  $\mu$ m ID, fused silica; running electrolyte, 10 mmol  $l^{-1}$  UV-Cat-1, 4.0 mmol  $l^{-1}$  HIBA, pH 4.4; separation voltage, 30 kV; detection, indirect UV at 214 nm. Peaks ( $mg\ l^{-1}$ ): 1, Rb (2); 2,  $K^+$  (5); 3,  $Ca^{2+}$  (2); 4,  $Na^+$  (1); 5,  $Mg^{2+}$  (1); 6,  $Li^+$  (1); 7,  $La^{3+}$  (5); 8,  $Ce^{3+}$  (5); 9,  $Pr^{3+}$  (5); 10,  $Nd^{3+}$  (5); 11,  $Sm^{3+}$  (5); 12,  $Eu^{3+}$  (5); 13,  $Gd^{3+}$  (5); 14,  $Tb^{3+}$  (5); 15,  $Dy^{3+}$  (5); 16,  $Ho^{3+}$  (5); 17,  $Er^{3+}$  (5); 18,  $Tm^{3+}$  (5); 19,  $Yb^{3+}$  (5). (Reproduced with permission from Weston A, Brown PR, Jandik P, Jones WR, and Heckenberg AL (1992) *Journal of Chromatography* 593: 289–295.)

In principle, two experimental approaches are taken in the CE analysis of cations, an offline preparation of complexes, prior to the CE analysis, and online complexation in the separation capillary. Their application depends on the stability of the complexes formed.

If weak complexes are rapidly formed, on-capillary partial complexation can be used. A ligand is added to the running electrolyte and a rapid equilibrium between the free metal ions and their complexes is established, with most of the ions present in the free form. Owing to different complexation degrees with various charges on the complexes, the ions have different migration rates. The capillary zone electrophoresis (CZE) mode and an indirect UV photometric detection method are usually employed in this case, as only a small fraction of the cations is complexed.

If the complexes of metal ions with ligands are sufficiently stable under the CE conditions, then off-capillary complexation is preferred. An excess of a strongly complexing agent is added to the sample prior to the CE analysis. On-column UV photometric detection is possible as the fraction of the complexed ions is large. If there is a danger of the dissociation of the complex during the CE analysis, the complexation agent is added to the running buffer in a high concentration. Poor peak shapes can be caused by slow attainment of complexation equilibria in the capillary.

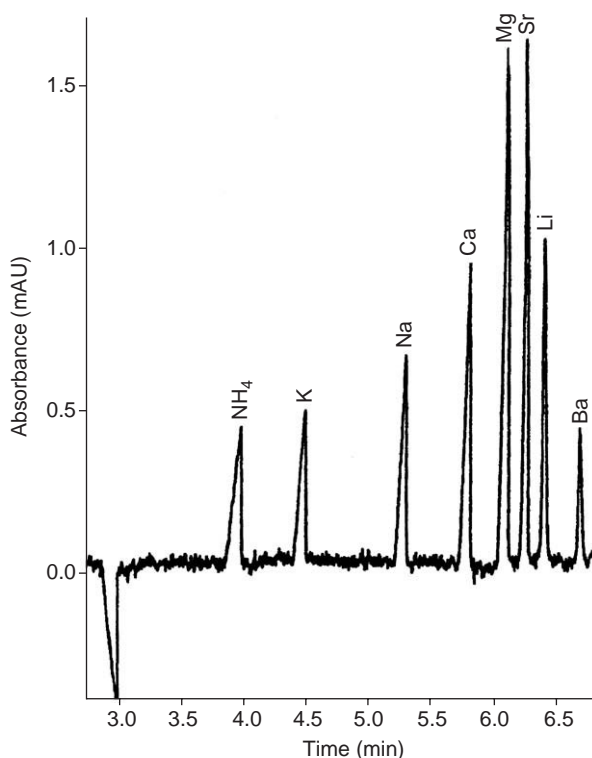
The equilibria of complexation reactions with weakly acidic or basic ligands are influenced by the pH of the running buffer and by the concentration of the complexing agent. The optimum pH for the separation is around the  $pK_a$  of the monoprotic acid. With di- and triprotic acids, the pH must lie between  $pK_{a,1}$  and  $pK_{a,2}$  ( $pK_{a,3}$ ).

The complexing agent must contain suitable binding groups (e.g., carboxyl, hydroxyl), its dissociating sites should not interfere with the complexation equilibrium (they should be located far from the binding groups), and the agent should absorb radiation at wavelengths different from that of the complex.

Cationic, anionic, or neutral metal complexes can be formed offline. In dependence on their structure and charge, they can be separated by CZE or MEKC. An advantage of cationic complexes lies in rapid analyses in CZE due to coelectroosmotic migration. However, the number of metals that can be separated in this mode is limited by a narrow migration window and/or by their similar mobilities. When using MEKC, the polarity can be reversed by an addition of a cationic surfactant to the electrolyte and thus the separation can be improved. Anionic metal complexes, such as those with cyanide, move, during CZE in

uncoated capillaries, to the cathode against the EOF and are detected within an acceptable time due to their rapid migration caused by their small size and/or high charge. Complexes with large ligands move very slowly or not at all. It is then necessary to suppress or reverse the EOF by adding a cationic surfactant or by a suitable coating of the capillary. Neutral complexes can be separated by MEKC.

Separation of cations can be influenced by their interaction with crown ethers, which depends on the sizes of the cation and the crown ether cavity. The concentration of a crown ether in the running buffer also plays a role. The best results have been obtained with 18-crown-6-ether where the selectivity changes were largest. The use of crown ethers makes it possible to separate, e.g., potassium from ammonium. Electrolyte containing  $4 \text{ mmol l}^{-1}$  18-crown-6,  $4 \text{ mmol l}^{-1}$  copper sulfate, and  $4 \text{ mmol l}^{-1}$  formic acid was successfully applied to complete separation of all alkali and alkaline earth cations including ammonium (Figure 2).



**Figure 2** Electropherogram of a mixture of alkali and alkaline earth metal ions. Capillary, 58.5 cm, 75  $\mu\text{m}$  ID, fused silica; running electrolyte,  $4 \text{ mmol l}^{-1}$  18-crown-6,  $4 \text{ mmol l}^{-1}$  copper sulfate, and  $4 \text{ mmol l}^{-1}$  formic acid, pH 4.4; separation voltage, 20 kV; detection, indirect UV at 215 nm. Peaks ( $\text{mg l}^{-1}$ ):  $\text{NH}_4^+$  (20);  $\text{K}^+$  (19.5);  $\text{Na}^+$  (11.5);  $\text{Ca}^{2+}$  (20);  $\text{Mg}^{2+}$  (12.2);  $\text{Sr}^{2+}$  (43.8);  $\text{Li}^+$  (3.5);  $\text{Ba}^{2+}$  (68.7). (Reproduced with permission from Havel J, Janoš P, and Jandík P (1996) *Journal of Chromatography A* 745: 127–134.)

## Anions

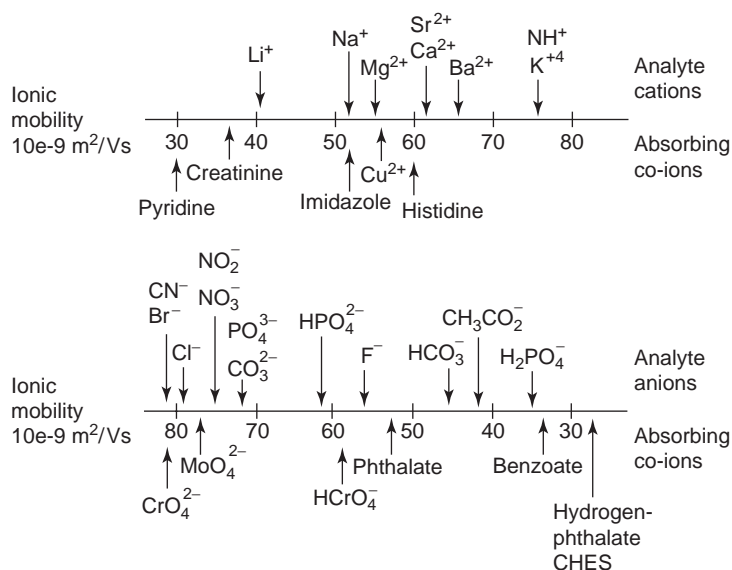
While cations can be analyzed directly in CE, the elution of anions in bare silica capillaries requires reversion of the EOF. The EOF can be changed by addition of cationic additives to the running electrolyte, by changing the concentration of the running electrolyte, or by a chemical modification of the capillary walls. Cationic surfactants are mostly added to reverse the EOF in analyses of anions. At concentrations below the critical micelle concentration, hemicelles are formed at the capillary wall that reverses the EOF. If anions interact with a monomeric surfactant present in the electrolyte, then their electrophoretic mobilities are influenced through ion association. The most common cationic surfactants used are, e.g., tetrabutylammonium, dodecyltrimethylammonium, tetradecyltrimethylammonium, and cetyltrimethylammonium (CTA) bromides or hydroxides, hexadimethrine, and hexamethonium hydroxides. Higher concentrations of additives lead to longer migration times owing to the formation of equilibrium ion-pairs. Differences in the selectivity allow complex samples to be analyzed by properly selecting the EOF modifier concentration and thus increasing the weak acid anion migration times.

Another possibility for EOF modification is coating of the capillary walls with cationic-soluble polymers. A net positive charge is then formed at the capillary walls. Reducing the EOF, e.g., by coating the capillary walls with silane, is sufficient for attaining sufficiently short analysis times for some anions. The EOF can also be modified by using capillaries made of materials other than fused silica, e.g., of polypropylene.

## Detection Modes

Detection in CE takes place directly on the separation column. The UV/Vis photometric detection is most common in capillary electrophoresis, as it is simple and reliable. The problems connected with the detection of LMW ions are caused by their low absorption in the UV region. Therefore, direct UV detection is only applicable to a few inorganic anions, e.g., to nitrate, sulfide, nitrite, iodide, bromide, and thiocyanate (for example, a detection limit of  $10 \mu\text{g l}^{-1}$  has been attained for sulfide in waste water using direct UV detection at 229 nm). LMW carboxylic acids can be detected at low wavelengths (200 nm and below).

Direct UV/Vis or fluorescence detection is often employed in analyses of cations after their complexation, provided that the ligand contains a chromophore or a fluorophore. Several metal ions, e.g.,  $\text{Fe}^{2+}$ ,  $\text{Fe}^{3+}$ ,  $\text{Co}^{2+}$ ,  $\text{Ni}^{2+}$ ,  $\text{Cu}^{2+}$ ,  $\text{Cd}^{2+}$ , form complexes with cyanide. Analogously,  $\text{Fe}^{2+}$  and  $\text{Fe}^{3+}$  form complexes with *o*-phenanthroline in the presence of EDTA and  $\text{Au}^{3+}$  with chloride and these complexes can be detected spectrophotometrically. 8-Hydroxyquinoline-5-sulfonic acid forms fluorescing complexes with metals. Dithizone sulfonate complexes have been used in determination of traces of inorganic mercury. 4-2-(Pyridylazo)resorcinol (PAR) forms colored complexes with  $\text{Cu}^{2+}$ ,  $\text{Pb}^{2+}$ ,  $\text{Ni}^{2+}$ ,  $\text{Cd}^{2+}$ ,  $\text{Fe}^{2+}$ , and  $\text{Fe}^{3+}$ , arsenazo I forms complexes with  $\text{Ba}^{2+}$ ,  $\text{Sr}^{2+}$ ,  $\text{Ca}^{2+}$ , and  $\text{Mg}^{2+}$ , arsenazo III has been recommended for separation of the lanthanoids and  $\text{U}^{6+}$ , sulfonoazo III for the determination of  $\text{Ba}^{2+}$  and  $\text{Sr}^{2+}$ . Detection limits of  $10^{-7} \text{ mol l}^{-1}$  have been



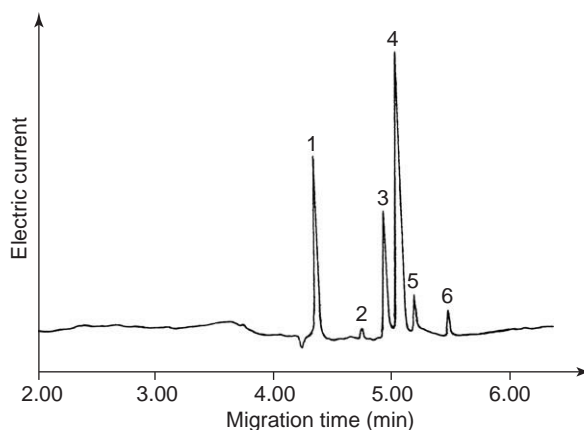
**Figure 3** Ionic mobilities of analytes and absorbing co-ions.

attained, using PAR complexes with the transition metals.

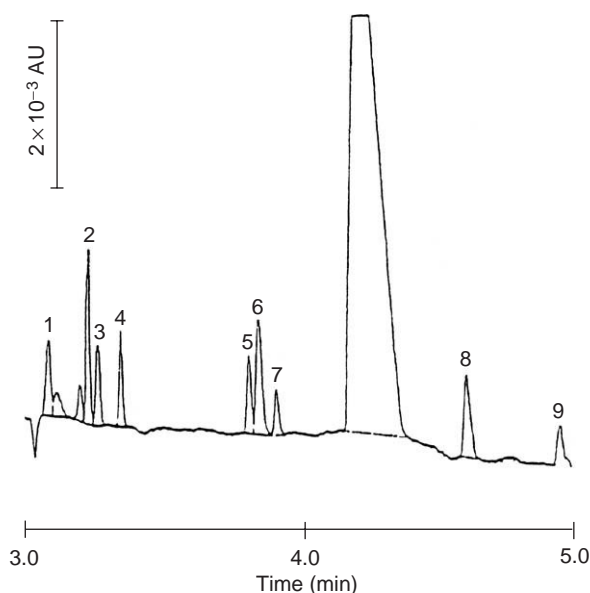
Indirect UV or fluorescence detection, based on charge displacement of an absorbing co-ion, is nearly universal. A disadvantage of indirect detection is a high background absorbance and thus a high noise and a limited linear dynamic range. UV absorbing anions or cations are added to the running electrolyte. For indirect UV detection, the co-ion should strongly absorb in the UV region and should have a mobility similar to those of the analytes to ensure optimum separation and peak shape. The EOF should have the same direction as the analyte migration to improve the speed of analyses. Cupric sulfate, imidazole, histidine, 4-methylbenzylamine, or creatinine are employed in analyses of cations while chromate, pyromellitic acid, phthalate, *p*-aminobenzoate, or molybdate in analyses of anions. **Figure 3** helps in the selection of a suitable UV-absorbing ion for indirect analysis of cations and anions. The chromate anion matches the mobilities of small anions while phthalate or benzoate is suitable for large anions. Molybdate has been shown to be a better visualization agent than chromate, yielding improved peak shapes because the molar absorptivity of molybdate is higher than that of chromate. Detection limits in ppb region have been obtained using a running electrolyte containing chromate, with a time of analysis of 3 min. *p*-Aminobenzoate has been found useful for simultaneous determination of low mobility organic and high mobility inorganic anions. The separation has been facilitated by an addition of a barium salt. So far the best separation of anions with indirect UV detection has been attained in the IonPhor PMA electrolyte buffer consisting of  $2.5 \text{ mmol l}^{-1}$  pyromellitic acid,  $6.5 \text{ mmol l}^{-1}$  NaOH,  $0.75 \text{ mmol l}^{-1}$  hexamethonium hydroxide, and  $1.6 \text{ mmol l}^{-1}$  triethanolamine, with a pH of 7.7. Anionic chromophores (benzoate, anisate) and cationic buffers (Tris, ethanolamine) have been tested for simultaneous detection of nonabsorbing anions and cations.

A CE method with indirect UV detection has been validated for eight anions and two electrolyte systems: pyromellitic acid + hexamethonium hydroxide and chromate + TTAB. The detection limits are between 1 and  $3 \text{ mg l}^{-1}$ , the repeatability and reproducibility of the measurement differ for different compounds and amounts to 5%, except for fluoride and phosphate. Linear calibration curves have been obtained within a concentration range between 1 and  $10 \text{ mg l}^{-1}$ .

Conductivity detection (CD) is a nearly universal bulk property detection mode for small ions and, similar to detection in IC, both nonsuppressed and



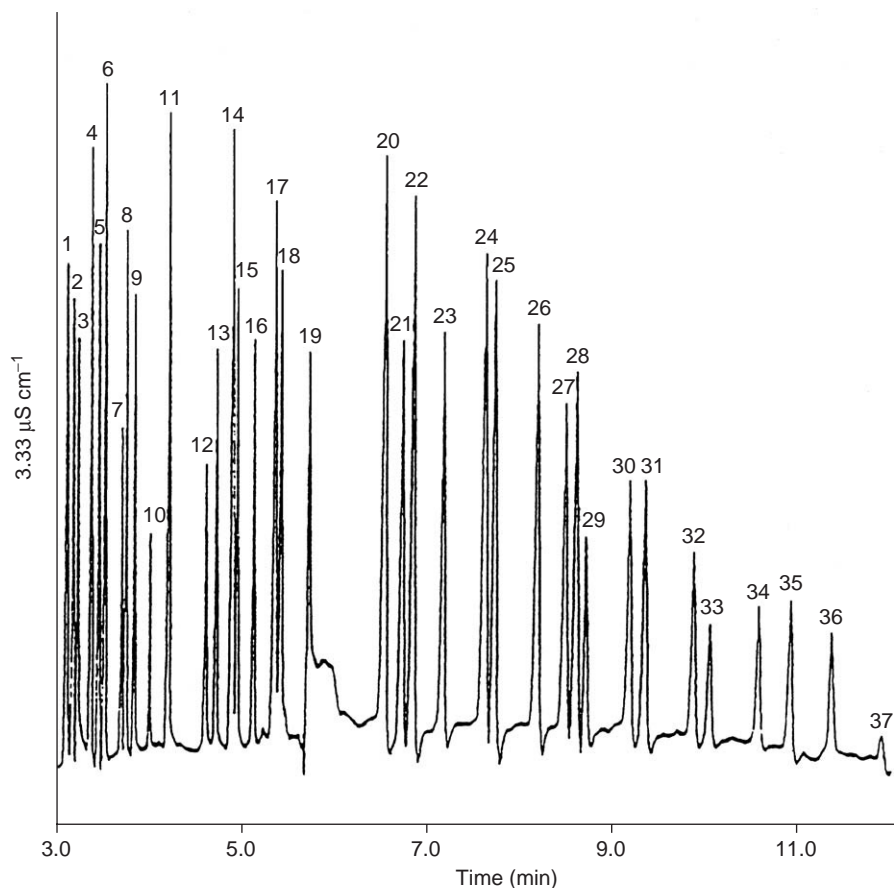
**Figure 4** Electropherogram of the analysis of cationic impurities in drug Carbetocin with conductivity detector. Capillary, 70.0 cm,  $50 \mu\text{m}$ , fused silica; running electrolyte,  $30 \text{ mmol l}^{-1}$  histidine,  $30 \text{ mmol l}^{-1}$  2-morpholinoethan-sulfonic acid; separation voltage, 30 kV; injection, hydrodynamic, 20 mbar for 6 s; detection, conductometric. Peaks: 1,  $\text{K}^+$ ; 2,  $\text{Ba}^{2+}$ ; 3,  $\text{Ca}^{2+}$ ; 4,  $\text{Mg}^{2+}$ ; 5,  $\text{Na}^+$ ; 6,  $\text{Li}^+$ . (Kindly provided by Dr. I. Jelínek from our Department.)



**Figure 5** Trace determination of inorganic and organic anions in pure water, after an electrophoretic enrichment at 5 kV for 45 s with an addition of  $75 \mu\text{mol l}^{-1}$  octanesulfonate to the sample. Capillary, 60 cm,  $75 \mu\text{m}$ , fused silica; running electrolyte,  $10 \text{ mmol l}^{-1}$  sodium chromate and  $0.5 \text{ mmol l}^{-1}$ , pH 8, UV detection at 254 nm. Anions (concentration in  $\text{mg l}^{-1}$ ): 1, chloride (3.5); 2, sulfate (4.8); 3, nitrate (6.2); 4, oxalate (5); 5, fluoride (1.9); 6, formate (5); 7, phosphate (3.2); 8, acetate (5); 9, propionate (5). (Reproduced with permission from Bondoux G, Jandik P, and Jones WR (1992) *Journal of Chromatography* 602: 79–88.)

suppressed CDs are used. There are more options for the selection of the running electrolyte in combination with CD. The co-ion must have a substantially different conductivity. In nonsuppressed





**Figure 6** Determination of inorganic and organic anions with direct conductivity detection. Capillary, 60 cm, 50  $\mu\text{m}$ , fused silica; running electrolyte, 50  $\text{mmol l}^{-1}$  2-*N*-cyclohexylamino-ethane-sulfonate, 20  $\text{mmol l}^{-1}$  LiOH, 0.03% Triton X-100; separation voltage, 25 kV; injection, hydrodynamic, 25 mbar for 12 s; the EOF was modified by preflushing the capillary with a 1  $\text{mmol l}^{-1}$  CTAB. Anions (concentration in  $\text{mg l}^{-1}$ ): 1, bromide (4); 2, chloride (2); 3, hexacyanoferrate (7); 4, nitrite (4); 5, nitrate (4); 6, sulfate (4); 7, azide (2); 8, oxalate (3); 9, molybdate (5); 10, tungstate (6); 11, 1,2,4,5-tetracarboxylic acid (7); 12, fluoride (1); 13, tartrate (5); 14, selenite (10); 15, phosphate (4); 16, citraconate (5); 17, glutarate (10); 18, phthalate (10); 19, carbonate (4); 20, acetate (10); 21, chloroacetate (10); 22, ethanesulfonate (20); 23, dichloroacetate (15); 24, propionate (15); 25, propanesulfonate (20); 26, crotonate (15); 27, butanesulfonate (20); 28, butyrate (15); 29, toluenesulfonate (15); 30, penatenesulfonate (20); 31, valerate (15); 32, hexanesulfonate (20); 33, caproate (15); 34, heptanesulfonate (20); 35, morpholineethanesulfonate (35); 36, octanesulfonate (20); 37, D-gluconate (40).

CD, low mobility buffers with higher ionic strengths provide an extended linearity and improve preconcentration by sample stacking.

In comparison with indirect UV detection, the sensitivity of CD is  $\sim 10$  times greater. The linear dynamic range extends over three concentration decades and the reproducibility of the migration times, peak area, and height is very good. A borate buffer (2  $\text{mmol l}^{-1}$ , pH 9.2) combined with suppressed conductivity detection provides good peak shapes owing to a close match of the borate mobility with those of the separated anions and meeting the principal condition of suppressed CD, i.e., the suppression leads to a weakly conducting species. Additives, such as barium ions, decrease the EOF and the migration velocity of high mobility anions, so that

they can be analyzed simultaneously with organic anions. Detection limits within a range of 1–10 ppb have been reported with suppressed conductivity detection.

Similar to CE analysis of anions, on- or end-column CD can be used for cations, with a sensitivity  $\sim 10$  time greater than that of the indirect UV detection. CD is a nearly universal bulk property detection mode for small ions. Inorganic or organic buffers with low conductivities, e.g., borate or MES-histidine, and higher ionic strengths are used when employing conductivity detection. **Figure 4** depicts a determination of cationic impurities ( $\text{K}^+$ ,  $\text{Ba}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Na}^+$ , and  $\text{Li}^+$  ions) in the drug Carbetocin using a MES-histidine buffer with conductivity detection. Contactless conductivity detection has



been suggested for analysis of  $\text{Rb}^+$ ,  $\text{K}^+$ ,  $\text{Na}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Cd}^{2+}$ , and  $\text{Li}^+$  cations, with a detection limit of  $\sim 1 \times 10^{-4} \text{ mol l}^{-1}$ . On-capillary cells combining contactless conductivity and UV/Vis detection improve the identification potential in separations of LMW ions.

CE-MS has been used in analyses of anions and cations as a method simultaneously providing positive identification and quantitation. CE-ESI MS is particularly well suited for analyses of quaternary ammonium salts. Speciation analysis of As can be carried out using a CE-ICP-MS. Problems in interfacing the ICP-MS detection to CE are associated with low flow rates and small samples analyzed. Detection limits in ppb region can be attained, using postcapillary hybridization prior to ICP-MS.

## Preconcentration Techniques

Sample preconcentration is usually necessary for analyses of highly dilute solutions, e.g., for separations of anions in deionized water or their determination in the presence of a large excess of a matrix component.

Isotachophoretic enrichment by electrostacking at the sample-buffer interface can be used. The sample matrix can assist in the stacking process by functioning as the leading or terminating electrolyte. The co-ion of the running electrolyte has to be chosen so that the analyte mobilities are between those of the ions of the electrolyte and the matrix. Limits of detection lower than  $50 \text{ nmol l}^{-1}$  have been attained in the simultaneous analysis of inorganic and organic anions in rain water when enriching by sample stacking with a dynamic injection. This preconcentration method permits determination of inorganic anions in the presence of a fluoride matrix up to an analyte-matrix ratio of  $1:6 \times 10^6$ .

Preconcentration with the electrokinetic injection can be used for nonionic matrices. With long injection times, the ionic components are preconcentrated at the expense of the nonionic interferents. The EOF has the direction opposite to the migration of the analytes. The matrix effects caused by ionic components can be decreased by suppressing their dissociation by a pH change, thus enriching the analytes by up to two orders of magnitude. The choice of the amount injected is influenced by the analyte mobilities, the magnitude of the EOF, and the sample and buffer ionic strengths. The reproducibility of the electrokinetic injection is poorer than that of the dynamic pressure injection and strongly depends on the running electrolyte ionic strength. Internal standards are usually added to improve the accuracy

and precision. A trace determination of some inorganic and organic anions in deionized water after electrokinetic enrichment with indirect UV detection is shown in Figure 5. A number of applications can be found in the references under Further Reading. An electropherogram of inorganic and organic anions with direct conductivity detection is given in Figure 6.

**See also:** Ion Exchange: Principles; Ion Chromatography Applications.

## Further Reading

- Fritz JS (1997) Determination of inorganic anions and metal cations. In: Camilleri P (ed.) *Capillary Electrophoresis Theory and Practice*, 2nd edn. Boca Raton, FL: CRC Press.
- Haddad RP (ed.) (2003) Inorganic species and small organic ions (Paper Symposium). *Electrophoresis* 12-13.
- Haddad RP, Macka M, Hilder EF, and Bogan DP (1997) Separation of metal ions and metal-containing species by micellar electrokinetic capillary chromatography, including utilization of metal ions in separations of other species. *Journal of Chromatography A* 780: 329-341.
- Haddad PR (1997) Comparison of ion chromatography and capillary electrophoresis for the determination of inorganic ions. *Journal of Chromatography A* 770: 281-290.
- Jandik P and Bonn B (1993) *Capillary Electrophoresis of Small Molecules and Ions*. New York: VCH.
- Jones WR (1994) Electrophoretic capillary ion analysis. In: Landers JP (ed.) *Handbook of Capillary Electrophoresis*, pp. 209-232. Boca Raton, FL: CRC Press.
- Macka M and Haddad PR (1997) Determination of metal ions by capillary electrophoresis. *Electrophoresis* 18: 2482-2501.
- Okada T (1997) Micellar electrochromatography of inorganic compounds. *Journal of Chromatography A* 780: 343-360.
- Pacáková V and Štulík K (1997) Capillary electrophoresis of inorganic anions and its comparison with ion chromatography. *Journal of Chromatography A* 789: 169-180.
- Pacáková V, Coufal P, and Štulík K (1999) Capillary electrophoresis of inorganic cations. *Journal of Chromatography A* 834: 257-275.
- Pacáková V, Coufal P, Štulík K, and Gaš B (2003) The importance of capillary electrophoresis, capillary electrochromatography and in chromatography in separations of inorganic ions. *Electrophoresis* 24: 1883-1891.
- Timerbaev AR (1997) Strategies for selectivity control in capillary electrophoresis of metal species. *Journal of Chromatography A* 792: 495-518.

## Environmental Applications

Y Picó, University of Valencia, Valencia, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Capillary electrophoresis (CE) is now emerging in environmental analysis as an advantageous tool because of its features such as higher separation efficiency, shorter analysis times, simplicity with regard to instrumentation, less consumption of expensive reagents and toxic solvents, field-screening capabilities, and microminiaturized format, e.g., CE on a chip. The originally reported CE limitation of inadequate sensitivity for environmental analysis, as a result of the small sample volumes typically injected ( $\sim 1$ – $10$  nl), has been overcome by off- and on-column trace enrichment schemes and improvements in sensitivity of detectors. CE can offer to the environmental chemist separation of compounds with a gas chromatography (GC)-like efficiency, and the ability of liquid chromatography (LC) to determine components that are thermally degradable or nonvolatile, without the derivatization of the analyte. Current CE methods are robust enough to provide valuable contributions to environmental assessment and their different selectivity make them complementary to LC and GC. In this article, the role of electrophoresis in environmental chemistry is introduced, followed by an examination of how CE is used in those environmental applications (sample treatment, on-column preconcentration procedures, CE modes of operation, and detectors).

### Types of Environmental Applications

The growing importance of this technique in the field of environmental analysis is emphasized by the appearance of first CE methods that are applicable to routine problems such as the determination of polar volatiles, most semivolatiles, nonvolatiles (e.g., herbicides), inorganic cations, inorganic anions, and natural organic matter (NOM). Most of the compounds determined by CE in different environmental matrices are shown in Table 1.

CE has been applied to anion and cation analysis during the last 15 years. This technique constitutes a viable alternative to ion chromatography. The United States Environmental Protection Agency (US EPA) has already approved a CE method for determining hexavalent chromium (in Region VII) and is currently

considering this technique to be included in the methods compendium SW-846 as an anion analysis method. This technique is appropriate for the analysis of dissolved inorganic anions in drinking water, ground water, surface water, and wastewater. CE has the impressive capability of separating inorganic ions. As an example, several metal cations, including 13 lanthanides, can be separated in only 6 min using lactate to partially complex the metal ions (Figure 1).

In environmental chemistry, the identification and quantification of element species are becoming increasingly important. It has been noted that the distribution, bioavailability, accumulation, and toxicological properties of heavy metals are strongly dependent on the chemical binding forms in which they occur in natural compartments. Since separation in CE is mainly governed by differences in charge-to-size ratio of the analyte, the technique is extremely powerful in discriminating the speciation pattern of redox-sensitive elements and organometallics. This has been proved for various environmentally significant elements such as mercury, arsenic, selenium, tellurium, antimony, lead, iron, chlorine, sulfur, and nitrogen (Figure 2).

The CE analysis of organic contaminants has caught the attention of analytical chemists because there are hundreds of these compounds that are released in the environment as a consequence of human activity. Many of them are considered, owing to their toxicity, as priority pollutants by the US EPA and the European Union (EU). The groups of compounds that have received most attention are pesticides, phenols, nitroaromatics, and other chemical warfare-related compounds, amines, aromatic sulfonic acids, carboxylic acids, endocrine disrupting compounds, phthalate esters, carbonyl, and dyes. Because of analyte preconcentration and stacking injection techniques, CE has become competitive in trace analysis and it is now being applied to environmental real matrices (Figure 3).

The chiral resolution of environmental pollutants by CE is a very interesting feature, since one of the chiral isomers may be more toxic than the other. In addition, biological transformation of the enantiomers is many times stereoselective, and, therefore, their uptake, metabolism, and excretion can be different. Besides, CE has also been utilized to separate the structural isomers of various toxic pollutants such as phenols, polyaromatic hydrocarbons, etc.

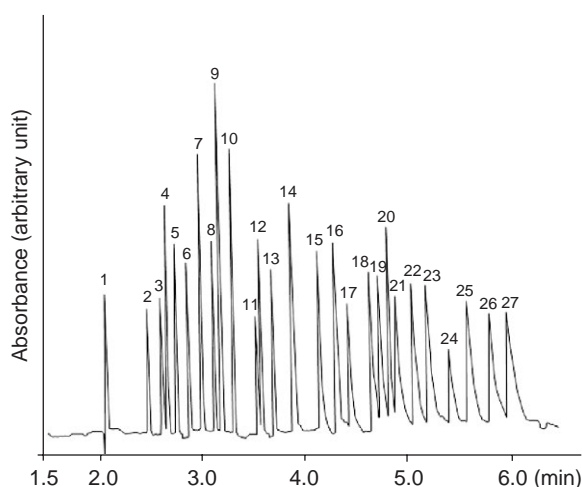
The term soil organic matter represents the organic part of soil, which includes high-molecular-mass organic material (proteins, etc.), small molecules

**Table 1** Environmental applications of electrophoresis

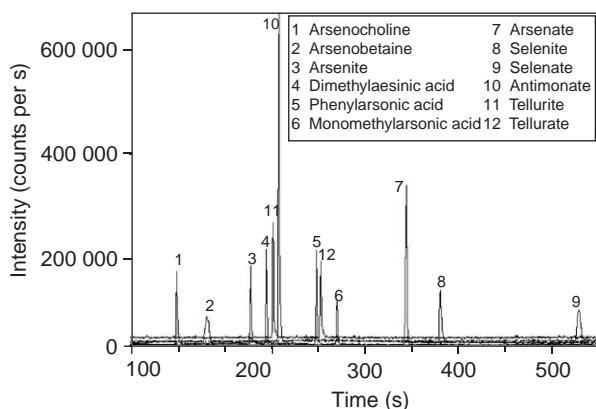
Analytes	Matrices
<b>Inorganic species</b>	
$F^-$ , $Cl^-$ , $Br^-$ , $BrO_3^-$ , $I^-$ , $IO_3^-$ , $ClO_2^-$ , $ClO_3^-$ , $ClO_4^-$ , $O_2$ , $SO_4^{2-}$ , $S_2O_3^{2-}$ , $S_4O_6^{2-}$ , $S^{2-}$ , $NO_2^-$ , $NO_3^-$ , $PO_4^{3-}$ , $HPO_4^{2-}$ , $CO_3^{2-}$ , $SCN^-$ , $SeO_3^{2-}$ , $SeO_4^{2-}$ , $AsO_2^-$ , $AsO_4^{3-}$ , $HCO_3^-$	Atmospheric samples: aerosols, air particulate matters, rain depositions
$Li^+$ , $Na^+$ , $K^+$ , $Rb^+$ , $Cs^+$ , $Ca^{2+}$ , $Mg^{2+}$ , $Sr^{2+}$ , $Ba^{2+}$ , $NH_4^+$ , $Mn^{2+}$ , $Cd^{2+}$ , $Cu^{2+}$ , $Cr^{3+}$ , $Zn^{2+}$ , $Fe^{2+}$ , $Fe^{3+}$ , $Co^{2+}$ , $Al^{3+}$ , $Ni^{2+}$ , lanthanides, $[PtCl_4]^-$ , $[PtCl_6]^-$ , Pt (metallic), $MoO_4^{2-}$ , $Cr^{III}$ , $Cr^{VI}$	Aquatic samples: drinking, mineral, and tap water, surface water, seawater, wastewater, soil solutions
$CH_3Hg^+$ , $(CH_3)_3Pb^+$ , $(C_2H_5)_3Pb^+$ , $(CH_3)_3Sn^+$ , $(C_4H_9)_3Sn^+$ , $(C_4H_9)_2Sn^{2+}$	Soil, sediment, and particulate matters
$As^{III}$ , $As^V$ , dimethylarsenic acid species: $HAsO_4^{2-}$ , $HAsO_3S^{2-}$ , $SO_2$ , $O_2$	Biological materials
<b>Organic contaminants</b>	
<b>Surfactants</b>	
Linear alkylbenzene sulfonates (LAS) and carboxylic degradation products, cationic: alkylammonium and alkylbenzyltrimethylammonium surfactants	Aquatic samples (drinking, surface, and wastewater)
Polycyclic aromatic hydrocarbons (PAHs)	Aquatic samples (drinking, surface, and wastewater), soil
<b>Dyes</b>	Aquatic samples (groundwater), soil, biological materials
Sulfonated, sulfonated azo, and fluorescent	
<b>Amines</b>	
Aliphatic, aromatic, heterocyclic aromatic, and biogenic	Aquatic samples (natural water, river water, and aqueous solutions), biological materials
<b>Aromatic acids and aromatic sulfonic acids</b>	
Benzenesulfonates, naphthalenesulfonates	Aquatic samples (tap and river water)
<b>Pesticides</b>	
Triazine and metabolites, phenoxy acid, sulfonylureas, phenylureas, urea-derived herbicides, carbamates (including dithiocarbamates), benzoimidazoles, bipyridilium compounds, pyrethrins, triazole fungicides, and other miscellaneous groups	Atmospheric samples (air particulate matters) Aquatic samples (drinking, mineral, and tap water, surface water, seawater, wastewater, soil solutions); soil, sediment, and particulate matters; biological materials
<b>Phenols</b>	
Chloro and nitrophenols	Aquatic samples (drinking, river, industrial, and waste waters)
<b>Endocrine disrupting compounds</b>	
Alkylphenols, alkylphenol ethoxylates, steroids, and bisphenol A	Aquatic samples (wastewater influents and effluents)
<b>Haloacetic acids</b>	
Trichloroacetic acid, dichloroacetic acid, dibromoacetic	Aquatic samples (tap water)
<b>Cyanide</b>	
Aquatic samples (leaching solutions from a gold mine)	
<b>Complexing agents</b>	
EDTA	Aquatic samples (wastewater, surface water)
<b>Warfare agents</b>	
Alkylphosphonic acids and their degradation products	Aquatic samples (surface water and soil solutions)
<b>Natural organic matter</b>	
Humic acids	Aquatic samples (river water), soil (Antarctic soil)

(amino acids, sugars, etc.), and humic substances (HS) (humic acids, fulvic acids, and humin). HS are complex, multicomponent mixtures of natural products, which result from the decay of plant and animal residues. They are ubiquitous in soil, water, and sediment, comprising ~40–60% or more of the dissolved organic carbon in natural water systems. HS play an essential role in determining the fate of environmental contaminants because of their ability to bind, sequester, and transport a wide range of organic compounds, heavy metals, and radionuclides.

The overriding feature of HS is their intrinsic heterogeneity of structure and conformation. Researchers have attempted to solve the structural puzzle of these materials by using mass spectrometry (MS) and nuclear magnetic resonance, but have simply shown that HS are complex mixtures of compounds with no defined structure or molecular mass. Resolution of HS has been equally challenging. Their charged nature renders them separable using CE, although they tend to migrate together in a single, broad peak (Figure 4).

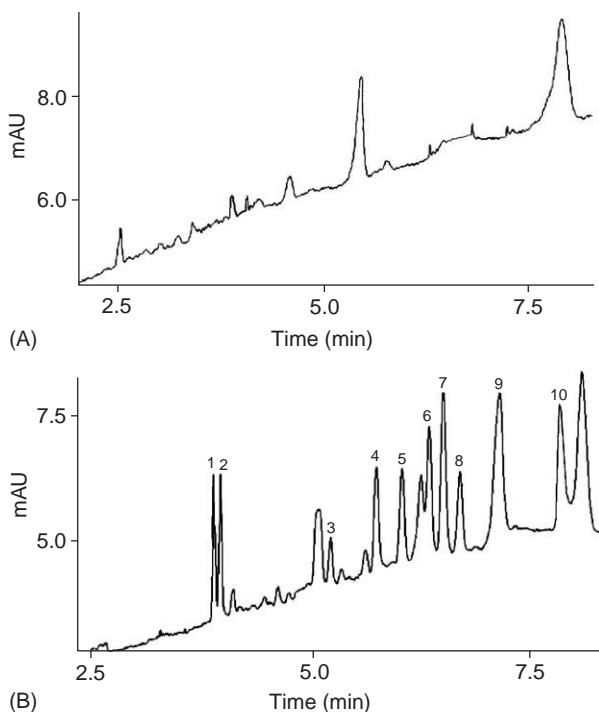


**Figure 1** Separation of 27 alkali, alkaline earth, transition, and rare earth metal ions in a single run using lactate. Electrolyte: 15 mmol L<sup>-1</sup> lactic acid, 8 mmol L<sup>-1</sup> 4-methylbenzylamine, 5% methanol, pH 4.25. Applied voltage: 30 kV. Peaks: 1=K<sup>+</sup>; 2=Ba<sup>2+</sup>; 3=Sr<sup>2+</sup>; 4=Na<sup>+</sup>; 5=Ca<sup>2+</sup>; 6=Mg<sup>2+</sup>; 7=Mn<sup>2+</sup>; 8=Cd<sup>2+</sup>; 9=Li<sup>+</sup>; 10=Co<sup>2+</sup>; 11=Pb<sup>2+</sup>; 12=Ni<sup>2+</sup>; 13=Zn<sup>2+</sup>; 14=La<sup>3+</sup>; 15=Ce<sup>3+</sup>; 16=Pr<sup>3+</sup>; 17=Nd<sup>3+</sup>; 18=Sm<sup>3+</sup>; 19=Gd<sup>3+</sup>; 20=Cu<sup>2+</sup>; 21=Tb<sup>3+</sup>; 22=Dy<sup>3+</sup>; 23=Ho<sup>3+</sup>; 24=Er<sup>3+</sup>; 25=Tm<sup>2+</sup>; 26=Yb<sup>3+</sup>; 27=Lu<sup>3+</sup>. (Reprinted with permission from Shi Y and Fritz JS (1993) Separation of metal ions by capillary electrophoresis with a complexing electrolyte. *Journal of Chromatography A* 640(102): 473–479; © Elsevier.)



**Figure 2** Simultaneous separation of 12 species of four elements. Concentration of the elements: As, Sb, Te 100 µg L<sup>-1</sup> each, Se 1000 µg L<sup>-1</sup>. (Prange A and Schaumlöfel D (1999) Determination of element species at trace levels using capillary electrophoresis-inductively coupled plasma sector field mass spectrometry. *Journal of Analytical and Atomic Spectrometry* 14: 1329–1332; reproduced by permission of The Royal Society of Chemistry.)

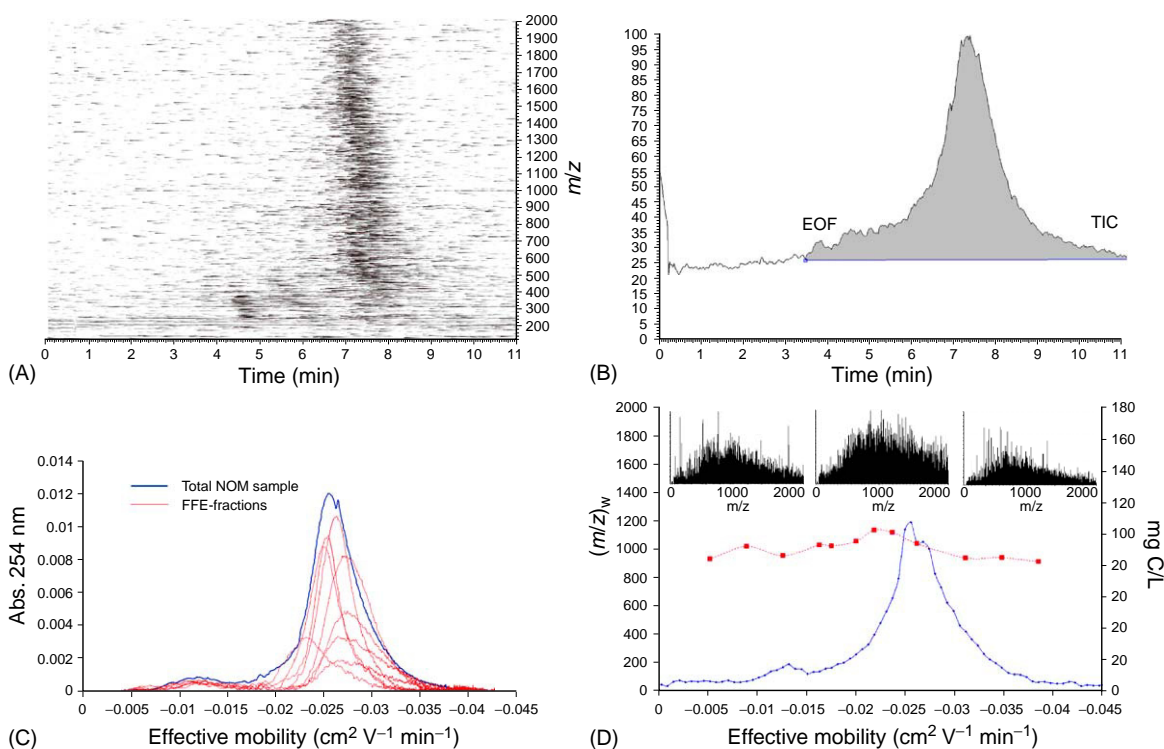
A variety of environmental matrices have been analyzed for the characterization of inorganic compounds, organic contaminants, and NOM. The measurement of inorganic cations and anions in various airborne samples (ambient air, aerosols, and dust) and wet deposition (rain and snow) continues



**Figure 3** Single wavelength (240 nm) electropherogram of (A) an extract tomato blank and (B) extract of fortified tomato blank with 10 compounds at the 0.5 mg per kg level. Running buffer: 4 mmol L<sup>-1</sup> borate (pH 9.2) containing 35 mmol L<sup>-1</sup> SDS. Peak identification: 1, triasulfuron; 2, chloresulfuron; 3, monuron; 4, flumeturon; 5, metobromuron; 6, chlorotoluron; 7, isoproturon; 8, diuron; 9, methabenzthiazuron; 10, flufenoxuron (Reproduced with permission from Rodríguez R, Picó Y, Font G, and Mañes J (2001) Determination of urea-derived pesticides in fruits and vegetables by solid-phase preconcentration and capillary electrophoresis. *Electrophoresis* 22: 2010–2016; © Wiley-VCH.)

being the subject of many studies. Much of the work on atmospheric samples involves the determination of common inorganic anions (Cl<sup>-</sup>, NO<sub>3</sub><sup>-</sup>, and SO<sub>4</sub><sup>2-</sup>), and cations of interest (alkali and alkaline earth metal and ammonium cations). Atmospheric samples that are characterized by low ionic strength provide very low detection limits under stacking conditions. Innovative applications of CE techniques in atmospheric science are the direct measurement of ionogenic atmospheric gases, and the determination of soluble gases and organic vapors (phenolic compounds and pesticides).

Recent applications of CE have been focused on natural aquatic samples. Water constitutes a clean environmental substratum suitable for direct injection in CE. The most intensively analyzed samples have been drinking, mineral, surface, ground, and sea waters. A number of methods have been described that yield detection limits for a variety of analytes below the concentrations at which they are present in real samples: inorganics, organometallics, surfactants, polycyclic aromatic hydrocarbons



**Figure 4** (A) CZE-ESI-MS (negative ionization modus) intensity plot of Suwannee River NOW (ammonium carbonate, pH 10); (B) TIC in timescale; (C) CZE-UV (254 nm) electropherogram in effective mobility scale of Suwannee River NOM and superimposed electropherograms of selected FFE fractions (measured in CE immediately after FFE separation). (D) FFE-TOC of Suwannee river NOM in effective mobility scale with superimposed weighted  $(m/z)_w$  values as obtained from offline ESI-MS (positive ionization modus) of selected FFE fractions. (Adapted with permission from Schmitt-Kopplin P and Kettrup A (2003) Capillary electrophoresis-electrospray ionization-mass spectrometry for the characterization of natural organic matter: an evaluation with free flow electrophoresis-off-line flow injection electrospray ionization-mass spectrometry. *Electrophoresis* 24: 3057–3066; © Wiley-VCH.)

(PAHs), dyes, amines, aromatic and aromaticsulfonic acids, pesticides, phenols, endocrine disrupting compounds, and NOM.

A difference between seawater and most other natural aquatic sources is its high ionic strength. This often complicates direct CE analysis. Another serious problem is due to large differences in concentrations between matrix components, such as sodium or chloride, and minor cationic and anionic constituents. One way to solve seawater matrix effects on separation and detection is sample dilution, but compounds present at lower concentrations can be diluted below the detection limits.

From an environmental point of view, analysis of waste waters is of great importance for characterizing their potential danger, and, in the case of emission to the environment, for detecting their spread and/or their origin. As for seawater, the matrix of waste water is complex and usually requires a treatment prior to analysis. Numerous examples of CE applications to different typologies of waste waters are reported in the current literature. CE was relevant to determine inorganics in samples of many different industrial processes (leather industry, pulp

and paper industry, extraction and precipitation of alumina from bauxite, and metal processing plants) and to the simultaneous identification of organic and inorganic anions in dumping sites. CE was also used in the control of relatively pure waste waters, such as the water from power or water purification plants, which requires control of inorganics, surfactants, PAHs, pesticides, phenols, and endocrine disrupting compounds at very low levels.

Determination of contaminants in sediments and soils by CE has received relatively little attention compared to the profusion of data concerning water analysis. This is probably related to the complexity of the sample matrices, and to the analytical difficulties involved in bringing the analytes into solution. CE is applied to anion and cation analysis in soils including sulfate analysis – widely used as an indicator of soil fertility – and separation of lithium, potassium, magnesium, barium, zinc, lead, lanthanum, samarium, europium, and dysprosium cations. This technique is also promising for determining organic contaminants associated with soil, such as PAHs, dyes, organo-tin compounds, and pesticides. Metal speciation is an unsettled issue in soil analysis

to provide relevant information regarding contamination and bioavailability of toxic metals. CE has been used to separate platinum species (many of them are potent sensitizers, even in very small doses), alkyl-lead and alkyl-tin compounds (widely used in industry, known to be potential causes of environmental pollution) and inorganic and organic mercury. Finally, the most important challenge to the pedologists is the characterization of the organic matter in soil and to understand better the humification process. CE seems to be an essential tool.

CE procedures for the analysis of samples of biological origin have also been reported. Although the number of published methods remains small, a representative range of pesticides in various fruits, vegetables, and grains; dyes (used as fruit fly toxicants) in coffee cherries and green roasted beans; heterocyclic aromatic amines in fried beefsteak, meat

extract, and baked salmon; biogenic amines in soy sauce, and fishery products; and methyl mercury in fish, can be determined by CE.

## CE Modes in Environmental Analysis

### CE Separation

A variety of CE modes are suitable to separate inorganic species, organic contaminants, and NOM, depending on the particular problem to be solved. Table 2 provides a comparison of the different CE modes and approaches applied to solve most environmental analytical problems.

The basic CE relies on the mobility of charged species in an electric field. The uncoated fused-silica capillary is filled with some type of electrolyte solution, known as the running buffer or background

**Table 2** Applications of different CE approaches to solve the separation problems in environmental applications

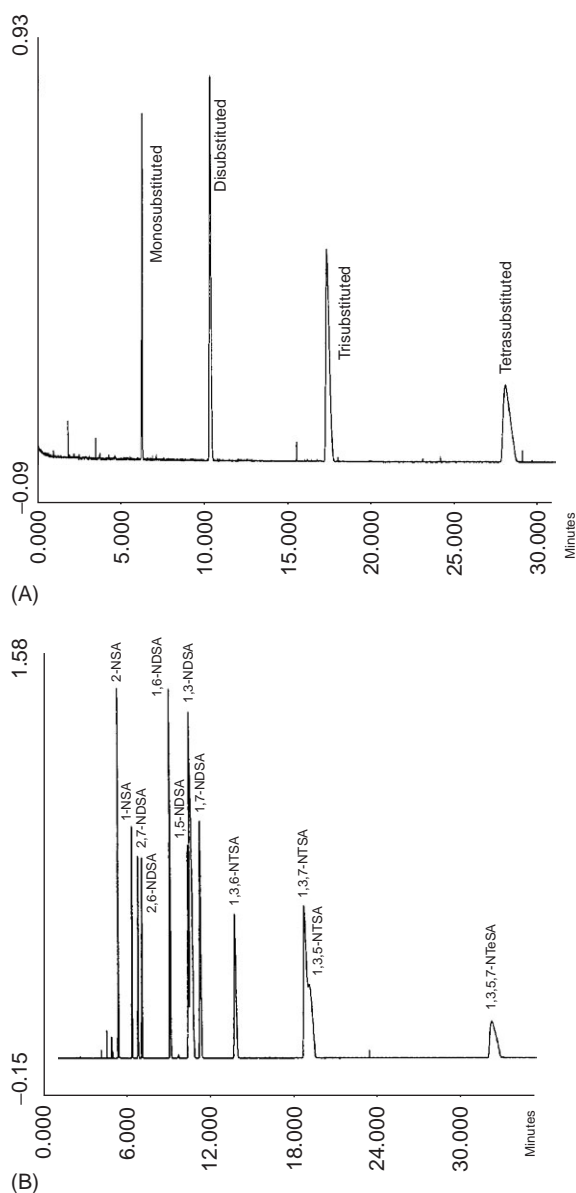
<i>Application</i>	<i>Problem</i>	<i>Solution</i>
Ionizable species	Lack of resolution	Improvement of separation by adjusting the buffer pH, alteration of the buffer ionic strength, and addition of organic solvents as methanol, 2-propanol, or acetonitrile
Inorganic cations	Differences in mobilities between cations of similar $m/z$ are often not sufficient for separation	Addition of weak complexing reagent to the sample and/or to the BGE gives complexed species. Complexed ligands are lactate, HIBA, crown esters, EDTA
Positively charged species	Strong adsorption of these substances at the inner surface of the silica capillary	Addition of organic modifiers in large amounts, the use of coated capillaries, or separation in absolute NACE
Negatively charged species	EM of the anionic species is close to or greater than EOF mobility	Reversal EOF mode by acidic pH or by addition of cationic surfactants at concentrations below the CMC that form a positively charged layer on the inner wall of the capillary
Cations and anions simultaneously	Simultaneous determination of high mobility inorganic anions and cations in a single run is difficult, especially with low EOF because they migrate in opposite direction	Transform cations into negatively charged complexes with EDTA, NTA, or PDC Use of one capillary, the sample is injected in both ends, and the detector is placed in the middle of the capillary
Ionizable species	Species that have very small charge-to-mass differences	NACE
Enantiomers	Chiral complexing reagents are required	Addition of CD to the BGE
Organic contaminants	Differences in mobilities between compounds of similar $m/z$ are often not sufficient for the separation	NACE
Polar and neutral solutes	Solubilization and mobility of neutral compounds	Formation of pseudostationary phase by MEKC or by MEECC Utilization of a stationary phase by CEC
Very hydrophobic analytes	Low solubility in the aqueous phase. They are totally incorporated into the micelle and cannot be separated	Addition of CDs to achiral surfactant solutions, which can include hydrophobic compounds changing the apparent distribution coefficients between the micelle and the non-micellar phases
Trace-level hydrophobic analytes	Unacceptable detection limits without preconcentration	Concentration by RMMEKC characterized by anionic micelles moving faster than the EOF

electrolyte (BGE). An electric field is applied to the capillary, and then cations go to the cathode whereas anions migrate to the anode. However, in the most conventional conditions, there is an electroosmotic flow (EOF) of the BGE induced by the negative ionization of the silica groups on the inner surface of the capillary. The excess cations in the diffuse double layer migrate toward the cathode, with their hydrogen shell, thus generating a net movement of liquid toward the negative electrode. This waterflow has a flat profile. As a consequence, the EOF provided high separation efficiencies compared to pressure-driven flow. It is the dynamic force of CE that will move all analytes toward the negative electrode (cathode). This basic mode has been used for determining inorganic ions, surfactants, PAHs, photoactive dyes, aromatic and heterocyclic aromatic amines, aromatic acids, naphthalene sulfonates, phenols, haloacetic acids, metallo-cyanide complexes, ethylenediaminetetraacetic acid (EDTA), alkylphosphonic acids, humic acids, and several classes of pesticides such as triazines, chlorophenoxyacids, sulfonylureas, and organo-tin compounds. BGE solutions can be either aqueous or nonaqueous, and can also contain additives such as cyclodextrin (CD) or polymers.

Nonaqueous capillary electrophoresis (NACE) involves the separation of analytes in a medium composed of organic solvents. The viscosity and dielectric constants of organic solvents affect both sample ion mobility and the level of EOF. The changes in separation selectivity in nonaqueous conditions contribute to a better separation of some substances that have very small charge-to-mass differences in aqueous phases (linear alkylbenzene sulfonates, or triazines). Further adsorption on the capillary wall and/or ion interactions, which cause solute precipitation (e.g., anionic surfactants with cations), can be avoided using NACE.

CE offers unsurpassed efficiency in chiral separations. As enantiomers have identical electrophoretic mobilities, some chiral complexing reagents must be added to the separation buffer to form diastereomeric complexes in dynamic equilibrium. One of the most popular procedures is the addition of CDs to the separation buffers. The outstanding chiral separation ability of the CE systems is accredited for phenoxy acid herbicides and triazole fungicides, which possess at least one chiral center. Addition of CDs can also be used for separating positional isomers as complex mixtures of unsubstituted naphthalene mono- to tetrasulfonic acids by CE (Figure 5).

Electrokinetic capillary chromatography (ECC) is another CE mode based on a combination of electrophoresis and interactions of the analytes with



**Figure 5** Separation of a mixture of naphthalenesulfonic acids (A) without CD and (B) with mixed  $\beta$ - and  $\gamma$ -CDs at a concentration of  $0.005 \text{ mol l}^{-1}$  each. Capillary, 75 cm (60 cm to detector)  $\times$   $50 \mu\text{m}$  ID uncoated fused-silica. Borate buffer,  $0.025 \text{ mol l}^{-1}$  (pH 9); voltage  $+25 \text{ kV}$ ; injection at 25 mbar, 0.1 min. Detection UV, 230 nm; capillary temperature  $35^\circ\text{C}$ ; overpressure  $2500 \text{ Pa}$  applied across the capillary (Reprinted with permission from Fisher J, Jandera P, and Stanek V (1997) Effects of the working electrolyte (cyclodextrin type and pH) on the separation of aromatic sulfonic acids by capillary zone electrophoresis. *Journal of Chromatography A* 772: 385–386; © Elsevier.)

additives (e.g., surfactants), which form a secondary phase moving at different velocities.

Micellar electrokinetic chromatography (MEKC) is a special and widely used case of ECC methods based on the differences between interactions of analytes with micelles present in the separation buffer,



which can easily separate charged and neutral solutes with either hydrophobic or hydrophilic properties. Micelles are formed by adding a surfactant, at a concentration above its critical micelle concentration (CMC), to the resolution buffer. The partition of analytes in and out of the micelles mimics reversed-phase LC conditions. The most striking feature is the resolution of MEKC, which is more likely to separate complex mixtures than CE. The most widely used surfactant has been SDS, but bile salts including sodium cholate, sodium deoxycholate, cetyltrimethylammonium bromide (CTAB), tetradecyltrimethyl ammonium bromide (TTAB), and others are also used frequently.

Most recently, a special technique of ECC, microemulsion electrokinetic chromatography, where a microemulsion is employed as dispersed phase, has been used for determining endocrine disrupting compounds.

Capillary electrochromatography (CEC) is a rapidly evolving hybrid technique between LC and CE. In essence, a voltage is applied across CE capillaries filled with an LC packing that generates two effects: differential partitioning and electrophoretic migration of the solutes during their transportation toward the detector. Generally, carrier electrolytes employed contain high levels (40–80%) of organic solvents such as methanol or acetonitrile. Exceptional resolution of water-insoluble and neutral solutes is readily achieved in CEC whilst these separations are more difficult to achieve by CE or LC. CEC has much latent potential to be exploited in the area of environmental analysis, especially those concerning the use of new materials, as molecularly imprinted polymers and immunosorbents, as stationary phases.

Future trends in the separation area will include translation of all these methods to microchip format, which promises to lead the next revolution in chemical analysis. MEKC and isotachopheresis, a CE separation technique in a discontinuous buffer system, have already been adapted to microchips and applied to assay herbicides, biogenic amines, and ions. Microchannels on a chip-like structure are likely to be exploited more frequently in CE after further development of nanotechnology because it results in extremely rapid separations that consume only picoliter sample volumes and introduce the possibility of merging sample preparation and analysis in a single device.

### Sample Enrichment Procedures

**Offline preconcentration procedures** The analysis of organic and inorganic compounds in environmental matrices presents several difficulties that are

common to all analytical separation techniques. These problems include the presence of a large number of compounds in the matrix, which leads to difficulties in resolving the analytes of interest. The occurrence of substances such as proteins, organic matter, or surfactants can modify the capillary characteristics, and, together with the low analyte concentrations, complicate the detection. Therefore, quite often a sample pretreatment is required. Typical sample preparations include solid-phase extraction (SPE) or liquid–liquid extraction (LLE). A further concentration is achieved by evaporating the organic solvent and reconstituting the sample in a smaller volume.

SPE and LLE can be used to extract the analyte from a matrix with some degree of specificity. Selectivity is obtained by adjusting the nature of the solid phase (in SPE) or the composition of extracting or eluting solvents. As a consequence of this selectivity, there is not one general extraction method for all the compounds determined in environmental matrices. Information on the different extraction procedures applied is given in Table 3. SPE or LLE present the advantage of preconcentrating the analytes, often 10–100 times.

Inorganic elements can often be released from solid matrices by water extraction, occasionally at elevated temperatures, or by homogenization in water followed by centrifugation, both procedures guaranteeing a high recovery without denaturation. In some cases, however, complete sample dissolution is required. The system of election is the closed-vessel microwave assisted acid digestion because this type of procedure has much lower risk of contamination and enhances the rate of decomposition. The extraction of organometallic compounds requires a partition step with an organic solvent immiscible with water.

The common practice to extract NOM from soils is to treat the sample with a strong base (sodium hydroxide) that solubilizes the humic materials.

Although LLE is effective in extracting many organic contaminants, it has been replaced by SPE, which has become the most powerful technique available for rapid and selective extraction of organic contaminants. SPE combines extraction and preconcentration of analytes from a liquid phase by adsorption on solid material, followed by desorption with a small quantity of an organic solvent providing extraction, cleanup, trace enrichment, and exchange of analyte environment for subsequent analysis in a unique step. Most of the SPE applications are based on the extraction using disposable cartridges or disks packed with C<sub>18</sub>-bonded silicas, porous graphitic carbon, or polymers (polystyrene–divinylbenzene; poly(divinylbenzene-co-*N*-vinylpyrrolidone) Oasis<sup>®</sup>

**Table 3** Sample preparation methods for capillary electrophoresis in environmental analysis

<i>Sample preparation</i>	<i>Analytes</i>	<i>Matrices</i>	<i>Remarks</i>
Homogenization in water and further filtration or centrifugation	Inorganic analytes and pesticides (dithiocarbamates, glyphosate, and metabolites)	Soil, sediment, and particulate matter; and biological material	Incomplete sample dissolution
Microwave-assisted acid digestion	Inorganic analytes	Soil, sediment, and particulate matter	Losses and decomposition of some analytes
Extraction with organic solvent and partitioning	Organometallic compounds and pesticides	Soil, sediment, and particulate matter; and biological material	High consumption of expensive and toxic organic solvents
Solid-phase extraction (SPE)	Endocrine disrupting compounds; surfactants; dyes; aliphatic and aromatic amines; pesticides; phenols; haloacetic acids; cyanide; EDTA; and alkylphosphonic acids	Aquatic samples (drinking, mineral, and tap water, surface water, seawater, wastewater, soil solutions); soil, sediment, and particulate matter; and biological material	Before SPE can be used with solid matrices (e.g., biological materials or soil), homogenization with a hydroalcoholic solvent and further filtration or centrifugation are required
Solid-phase microextraction (SPME)	PAHs and acidic pesticides	Aquatic samples (drinking water); and biological materials	Before SPME can be used with solid matrices, homogenization with a hydroalcoholic solvent and further filtration or centrifugation are required

HLB), along or together with strong anion-exchanger resins to improve the extraction of most polar analytes.

Solid-phase microextraction (SPME) is a miniaturized variation of the SPE, based on the partitioning of the analytes between the sample matrix and the stationary phase, which is coated on a fused-silica fiber. SPME with CE has been applied in environmental analysis to determine PAHs in water, and pesticides in vegetables. Trapped analytes can be desorbed by an organic solvent or directly into the CE electrolyte stream, via an adapter.

Examples of basic studies of offline extraction and preconcentration of pesticide residues using other techniques, such as online dialysis, steam distillation, supercritical fluid extraction, pressurized liquid extraction, cloud point extraction, or liquid–liquid membranes, have been reported. The large amounts of matrix coextractives and the need for clean extracts in CE/ultraviolet (UV) analysis are the main reasons for their scarce application.

**On-column preconcentration procedures** The word ‘stacking’ defines any on-capillary mode of concentration or focusing analytes based on changes of electrophoretic velocity due to the electric field across concentration boundaries. Sample stacking can be performed in both hydrodynamic (e.g., gravity or pressure) and electrokinetic (e.g., voltages) injection modes. The sample solution is sandwiched between two portions of the CE separation buffer. When high

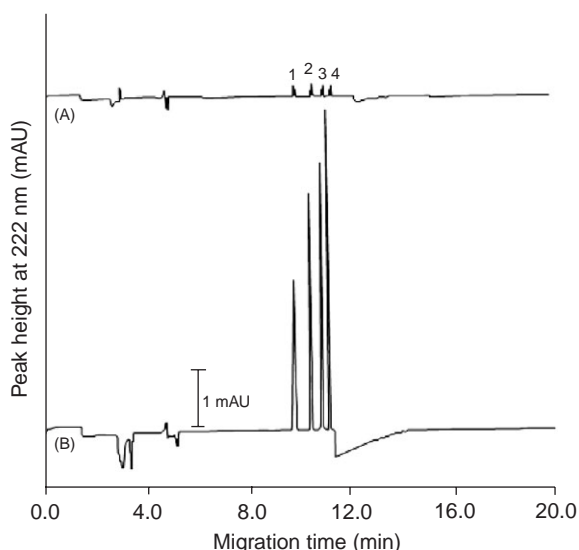
voltage is applied across the capillary, the analytes move quickly through the sample until they reach the buffer zone where they slow down because of the different compositions of the separation buffer. Neutral compounds, which are determined using MEKC, do not take part in a normal stacking procedure because they are not accelerated in the high electric field that is developed across the injection zone. The stacking technique has several variations including those designed for neutral compounds, as summarized in **Table 4**.

Among the strategies developed for neutral compounds, sweeping and stacking with reverse migration micelles, with and without the insertion of a plug of water before sample injection, are well established for different pesticide classes (triazine, benzimidazole, ura, and carbamate), providing detection limits of the order of  $2\text{--}46\ \mu\text{g l}^{-1}$ . Evidently, by the application of a stacking technique, the detection sensitivity of analytes can be greatly enhanced (**Figure 6**).

An on-column trace enrichment method for electrochromatography of dilute samples, similar to stacking methods, involves online preconcentration by frontal electrochromatography under conditions of strong solute binding to the stationary phase, followed by a step-gradient elution electrochromatography with a mobile phase of high eluting strength. The effectiveness of this on-column trace enrichment procedure is largely influenced by the affinity of the individual analytes for the stationary phase.

**Table 4** Preconcentration in CE for environmental analysis

<i>CE mode</i>	<i>Principle</i>	<i>Remarks</i>	<i>Enhancement (<math>\times</math> fold)</i>
<b>CE</b>			
Stacking	Ionic analytes are focused in a sharp sample band, at the interface, between a low conductivity matrix and a high conductivity separation buffer	Broadening of the stacked zones injection not longer than 30 s without loss of the separation efficiency	3–30
Stacking with matrix removal	The sample buffer is removed after stacking reversing the voltage polarity (EOF is directed toward the inlet expelling the excess of sample buffer). The polarity is reversed again before the analytes are also removed from the capillary by EOF	The capillary can be filled with an extremely large sample volume retaining high resolution	100–1000
Isotachoforesis	Sample zones between the BGE of higher (leading electrolyte) and lower (terminating electrolyte) electrophoretic mobilities	Application to the microchip format	100–10 000
<b>MEKC</b>			
Stacking and stacking with matrix removal	Addition of SDS to the injection solution causes neutral analytes to behave like anions because they become adsorbed on the anionic SDS micelles	The two methods gave similar stacking efficiencies up to an injection volume of 80 nl	85
Normal neutral analyte stacking	Neutral analytes in a low conductivity sample without surfactant. Micelles from the BGE reservoir race across the sample zone toward the capillary inlet incorporating analytes. Once the micelles reach the boundary between the sample zone and the BGE they are stacked into narrow bands	The enhancement in detection limits is poor	10
Reverse neutral analyte stacking	Neutral analytes in a low conductivity sample without surfactant. Stacking is performed by applying a negative voltage at the inlet. As the sample solution is backed out the capillary by the EOF, micelles from the inlet buffer reservoir migrate to the detector. The polarity is reversed and the separation performed in normal mode	Improvement in the stacking efficiency The reproducibility of the method is limited by the analyst's ability to determine when to switch the polarity	20
Stacking with reverse migration micelles (SRMM)	Preparing the sample in a solution of lower conductance than that of the BGE, and without the surfactant. The potential at the inlet is negative and the EOF pushes the sample plug out of the capillary. However, the high concentration of positive charges reduces the EOF, producing the micelles' net migration toward the detector	The advantage of SRMM is that the focusing process and the removal of the sample occur upon the application of voltage at negative polarity	50
Stacking using reverse migration micelles plug (SRW)	SRW is performed unlike the previous procedure, injecting a water plug prior to long hydrodynamic and water injection	The advantage is that it provides a second enhancement field zone	100
Sweeping	Samples are injected in a buffer solution with a similar conductivity as that of the BGE but without surfactant. Separation is performed in negative polarity and with EOF suppression. As the micelles migrate from the BGE reservoir toward the detector they sweep the neutral analytes along	The effectiveness of this sample concentration technique has been shown to be dependent on the analytes' affinity for the micelle	80–5000



**Figure 6** Detection sensitivity of analytes measured under two different separation conditions: (A) without sample stacking (2.5 s injection with an injection pressure of 0.4 psi; sample concentration  $1.0 \mu\text{g ml}^{-1}$ ); (B) with sweeping-stacking (30 s injection with an injection pressure of 1 psi; sample concentration,  $1.0 \mu\text{g ml}^{-1}$ . Separation buffer,  $40 \text{ mmol l}^{-1}$  TTAB in  $40 \text{ mmol l}^{-1}$  phosphate buffer at pH 6; capillary,  $70 \text{ cm} \times 50 \mu\text{m}$  ID; applied voltage,  $-20 \text{ kV}$ ; detection wavelength,  $22 \text{ nm}$ ; temperature,  $25^\circ\text{C}$ ; sample concentration,  $10 \mu\text{g ml}^{-1}$ , sample dissolved in a sample matrix containing 4% of acetonitrile solution. Peak identification: 1, simetryn; 2, ametryn; 3, prometryn; 4, terbutryn. (Reprinted with permission from Lin CE, Liu YC, Yang TY, Wang TZ, and Yang CC (2001) Online concentration of s-triazine herbicides in micellar electrokinetic chromatography using a cationic surfactant. *Journal of Chromatography A* 916: 239–245; © Elsevier.)

Another way of enhancing the sensitivity of CE in environmental analysis is by on-column SPE, using a capillary column with a plug of reverse-phase material or a material bounded to the inner wall of the capillary. These approaches seem very promising but, to our knowledge, no applications to the environmental analysis have been reported till date.

### Detection Systems

The detection system is the other cornerstone to improve the concentration limits of detection. Although CE is more easily interfaced with optical detection methods based on UV–visible absorption and laser induced fluorescence (LIF), a variety of detection systems are suitable for the identification and quantification of inorganic compounds, organic contaminants, and NOM in the environment. Table 5 lists the most common detectors used in CE environmental applications.

The majority of CE instruments commercially available today are equipped with UV–visible absorbance detectors because of their wide range of

applicability. The most popular is the photodiode array (DAD) because it displays the UV spectra of eluting compounds. In these detectors, the capillary itself serves as the cylindrical detection cell constraining them to operate within the minuscule dimensions of the capillary. Although UV detection is very mass sensitive, concentration limits of detection are substantially higher because of the small injection volume used.

Inorganic ions are mainly UV-transparent, and most CE applications utilize indirect UV detection to quantify them. The key to this approach is the displacement of a highly absorbing electrolyte co-ion by the sample ions. Direct UV of these compounds requires a derivatization step using suitable chromophores prior to separation.

Organic compounds absorb appreciably at wavelengths below  $250 \text{ nm}$ , but the problem with the UV detection is that many reagents and matrix-derived compounds absorb in the same spectral region. Figure 3B illustrates the presence of endogenous peaks in an electropherogram of a tomato sample. For this reason, CE–UV analysis is more popular in very clean environmental substrates such as water.

LIF detectors are more sensitive and selective than the UV–visible absorption ones, but only a few different laser sources ( $488 \text{ nm}$  Ar ion,  $442 \text{ nm}$  He–Cd, and  $324 \text{ nm}$  He–Ca) are available. Direct detection of native fluorescence compounds separated by CE has been demonstrated for some fluorescent dyes. Fluorescence detection of other inorganic and organic contaminants is also achieved by indirect methods – either direct fluorescence by the formation of complexes or derivatives, or by incorporating a fluorophore into the BGE.

Table 6 summarizes the variety of chromophores, fluorophores, and derivatizing agents used in the indirect format of UV and LIF detection.

Recently, electrochemical detection methods, namely, conductimetry, amperometry, and potentiometry, have also become accessible. All three variants of electrochemical detection are intrinsically simpler than the optical methods, and their success depends highly on the electrode materials and designs used. Conductivity detection relies on measurement of the differences between the conductivities of the analyte and the separation electrolyte; this provides a direct relationship between migration times and response factor, and makes this detector universal. On the contrary, amperometric detection is restricted to electroactive species and potentiometric detection is not possible for certain small ions with multiple charges. Conductimetric detection works better for inorganic compounds since the higher mobility of

**Table 5** Different detection modes for capillary electrophoresis in environmental analysis

<i>Detection mode</i>	<i>Analyte types</i>	<i>Detection limits (mol)</i>	<i>Advantages</i>	<i>Disadvantages</i>	<i>Solutions</i>
UV-Vis absorption	All analytes	$10^{-13}$ – $10^{-16}$	Universal; easily interfaced; and spectral information using DAD	Low sensitivity; and substances UV-transparent	Work at wavelengths less than 200 nm; wide-bore capillaries; bubble or z-shaped cells; sample preconcentration techniques; and indirect by BGE or complexation
Fluorescence	All analytes	$10^{-15}$ – $10^{-21}$	Easily interfaced; sensitive; and selective	Few laser sources commercially available; few substance naturally fluorescents	Indirect by BGE, complexation, or derivatization
Electrochemical conductivity	Ionic	$10^{-18}$ – $10^{-20}$	Universal; relatively inexpensive; and easy to adapt	Low sensitivity; requires special electronic and capillary modifications	Use low conductivity buffers
Potentiometry	Ionic	$10^{-13}$ – $10^{-15}$	Universal	Not suitable for small multicharged ions; requires special modifications	Adequate selection of the ionophore
Amperometry	Electroactive compounds	$10^{-18}$ – $10^{-20}$	Sensitive and selective	Restricted to electroactive analytes	Indirect by BGE, complexation, or derivatization
ES-MS	All analytes	$10^{-16}$ – $10^{-17}$	Sensitive, selective; and structural information	Restriction in buffer choice	Partial filling and anodically migrating micelles procedures
ICP-MS	Cations	$10^{-13}$ – $10^{-16}$	Selective and enable determination of different valences	Limited number of detectable elements	

the small ions directly affects the detection signal. However, impressive results have also been achieved with this technique for many organic contaminants with weakly acidic or basic character. Amperometric detection is an electroanalytical technique most used to determine organic contaminants and is also a valuable method for most transition metals and some electroactive anions (composite anions such as nitrogen- and sulfur-containing species). In contrast, alkali and alkali-earth metals are not accessible to amperometric detection. Potentiometric detection methods have been reported mainly for inorganic

species, as this is the usual domain of ion-selective electrodes, but, as has been recently shown, it can also be used for organic species. A disadvantage in the application to organic ions is the poor selectivity for most multicharge ions.

Presently, the coupling of CE to MS is already an attractive approach, which is generating a growing interest because it facilitates the analytes' identification, overcomes a large number of interfering substances, and improves detection sensitivity. Electrospray (ES) is the ionization technique most successfully combined with CE by means of the

**Table 6** Indirect methods for environmental analysis by UV–Vis absorbance and CE–LIF

<i>Reactives</i>	<i>Analytes</i>	<i>Remarks</i>
UV–Vis absorbing electrolytes Creatinine, imidazole, ephedrine, pyridine, copper sulfate, 4-aminopyridine, 4-methylbenzylamine, benzimidazole, <i>p</i> -toluidine, <i>N,N'</i> -dimethylamine, malachite green, <i>o</i> -aniline, 1-(4-pyridyl)pyridinium	All analytes	Most universal and can be applied to many non-UV absorbing analytes if a suitable buffer system is selected
UV absorbing complexes Cyanide	Inorganic cations	Involves the formation of metal–ligand complexes The resultant negatively charged complexes require an appropriate electrolyte and an EOF modifier
Fluorescence electrolytes Cerium <sup>3+</sup> , fluorescein, 2,5-dihydroxybenzoic acid, quinine sulfate	All analytes	Restriction in buffer choice
Fluorescence complexes 8-Hydroxyquinoline-5-sulfonic acid (HSQ)	Inorganics	Most applicable to inorganics
Precolumn fluorescent derivatization 5-Aminonaphthalene-1-sulfonic acid (ANSA) 7-Aminonaphthalene-1,3-disulfonic acid (ANDSA) 8-Aminonaphthalene-1,3,6-trisulfonic acid (ANTS) Fluorescein isothiocyanate (FITC) <sup>a</sup> 9-fluorenylmethyl chloroformate <sup>a</sup>	All the analytes	ANSA enhances the enantio-separation
On-capillary derivatization Quaternary ammonium surfactants (CTAB) catalyzing the thermal decomposition of <i>N</i> -methyl carbamates to liberate methylamine Separation buffer contains the derivatizing components <i>o</i> -phthalaldehyde/ 2-mercaptoethanol	<i>N</i> -methylcarbamates	<i>o</i> -Phthalaldehyde is the most popular fluorescence reactant An intact capillary can serve consecutively as chamber for separation, decomposition, derivatization, and detection

<sup>a</sup> The reagent itself is fluorescent and must be removed or well separated from derivatized peaks of interest.

sheath flow interfaces. It holds the most prominent position in the field of environmental analysis. Forthcoming developments in CE/MS will focus on ES ionization with triple quadrupoles, double focusing instruments, ion traps, and time-of-flight mass spectrometers.

However, the application of CE–ES–MS to various environmental problems is still limited by the use of nonvolatile buffer components and organic additives, such as surfactants and chiral compounds, for achieving a wide variety of CE separations. Anodically migrating micelles and partial-filling (PFMEKC) techniques have been proposed to overcome these problems. In the first approach, the micellar velocity is directly manipulated by the adjustment of the EOF against the electrophoretic velocity of the micelle by changing the solution pH in MEKC. The elimination of MEKC surfactant introduction into ES–MS is

achieved with an anodically migrating micelle, moving away from the ES interface.

The PFMEKC consists in filling the capillary first with the background buffer (without surfactant) and then with the micellar buffer. After the analytes reach the detector, and before the SDS molecules elute from the capillary, the separation is stopped.

The hyphenation of CE with element-selective inductively coupled-plasma mass spectrometry (ICP–MS) has been proven to be an ideal technique to make elemental speciation measurements and to characterize NOM. However, ensuring accuracy is still a challenge and many factors, which may potentially change the form of the analyte when establishing sample preparation and storage procedures, should be taken into account. The same precautions necessary for accurate CE and/or ICP–MS measurements must be considered.

Other new detection methods, including radioactivity, X-ray, or flame photometric detection in the P-selective mode, or the coupling of CE to biosensor detection have been studied for the determination of environmental contaminants showing certain promising features, including high sensitivity and good selectivity, but these approaches have not received acceptance since they are not easily available yet.

**See also:** **Elemental Speciation:** Practicalities and Instrumentation. **Endocrine Disrupting Chemicals. Environmental Analysis. Geochemistry:** Soil, Major Inorganic Components; Soil, Minor Inorganic Components; Soil, Organic, Components. **Herbicides. Humic and Fulvic Compounds. Polycyclic Aromatic Hydrocarbons:** Determination. **Surfactants and Detergents. Water Analysis:** Organic Compounds.

## Further Reading

- Ali L, Gupta VK, and Aboul-Enein HY (2003) Chiral resolution of some environmental pollutants by capillary electrophoresis. *Electrophoresis* 24: 1360–1374.
- Cai J and Henion J (1995) Capillary electrophoresis–mass spectrometry. *Journal of Chromatography A* 703: 667–692.
- Dabek-Zlotorzynska E, Aranda-Rodríguez R, and Keppel-Jones K (2001) Recent advances in capillary electrophoresis and capillary electrochromatography of pollutants. *Electrophoresis* 22: 4262–4280.
- Fukushi K, Takeda S, Chayama K, and Wakida S (1999) Application of capillary electrophoresis to the analysis of inorganic ions in environmental samples. *Journal of Chromatography A* 834: 349–362.
- Heinig K and Vogt C (1999) Determination of surfactants by capillary electrophoresis. *Electrophoresis* 20: 3311–3328.
- Johns C, Macka M, and Haddad PR (2003) Enhancement of detection sensitivity for indirect photometric detection of anions and cations in capillary electrophoresis. *Electrophoresis* 24: 2150–2167.
- Li SFY (1996) *Capillary electrophoresis. Principles, Practice and Applications*. Amsterdam: Elsevier.
- Liu YM and Cheng JK (2003) Elemental speciation analysis in capillary electrophoresis. *Electrophoresis* 24: 1993–2012.
- Martinez D, Cugat MJ, Borrull F, and Calull M (2000) Solid-phase extraction coupling to capillary electrophoresis with emphasis on environmental analysis. *Journal of Chromatography A* 902: 65–89.
- Paull B and King M (2003) Quantitative capillary zone electrophoresis of inorganic anions. *Electrophoresis* 24: 1892–1934.
- Picó Y, Rodríguez R, and Mañes J (2003) Capillary electrophoresis for the determination of pesticide residues. *Trends in Analytical Chemistry* 22: 133–151.
- Song LG, Xu ZH, Kang JW, and Cheng JK (1997) Analysis of environmental pollutants by capillary electrophoresis with emphasis on micellar electrokinetic chromatography. *Journal of Chromatography A* 780: 297–328.
- Sovocool GW, Brumley WC, and Donnelly JR (1999) Capillary electrophoresis and capillary electrochromatography of organic pollutants. *Electrophoresis* 20: 3297–3310.
- Timerbaev AR, Dabek-Zlotorzynska E, and van der Hoop MAGT (1999) Inorganic environmental analysis by capillary electrophoresis. *Analyst* 124: 811–826.
- Willauer HD and Collins GE (2003) Analysis of inorganic and small organic ions with the capillary electrophoresis. *Electrophoresis* 24: 2193–2207.

## Food Chemistry Applications

**R A Frazier**, University of Reading, Reading, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Capillary electrophoresis (CE) has become an established technique in chemical analysis, including many applications in food chemistry. CE is an automated technique that bridges the gap between high-performance liquid chromatography (HPLC) and traditional electrophoresis techniques, offering separation modes whose basis ranges from zone electrophoresis through to chromatography. The most common

modes of separation employed for food applications are capillary zone electrophoresis (CZE) and micellar electrokinetic chromatography (MEKC). These two modes are complementary not only to each other, but also to the separation modes of HPLC. CE importantly offers the ability to control selectivity through straightforward changes to the run buffer composition while usually requiring no changes to the capillary. This is a considerable advantage in terms of the development of new methods, both in time and cost savings.

CZE separates sample components based on differences in their electrophoretic mobility, which is determined by their charge-to-size ratio. CZE is therefore only applicable to the separation of charged



Other new detection methods, including radioactivity, X-ray, or flame photometric detection in the P-selective mode, or the coupling of CE to biosensor detection have been studied for the determination of environmental contaminants showing certain promising features, including high sensitivity and good selectivity, but these approaches have not received acceptance since they are not easily available yet.

**See also:** **Elemental Speciation:** Practicalities and Instrumentation. **Endocrine Disrupting Chemicals. Environmental Analysis. Geochemistry:** Soil, Major Inorganic Components; Soil, Minor Inorganic Components; Soil, Organic, Components. **Herbicides. Humic and Fulvic Compounds. Polycyclic Aromatic Hydrocarbons:** Determination. **Surfactants and Detergents. Water Analysis:** Organic Compounds.

## Further Reading

- Ali L, Gupta VK, and Aboul-Enein HY (2003) Chiral resolution of some environmental pollutants by capillary electrophoresis. *Electrophoresis* 24: 1360–1374.
- Cai J and Henion J (1995) Capillary electrophoresis–mass spectrometry. *Journal of Chromatography A* 703: 667–692.
- Dabek-Zlotorzynska E, Aranda-Rodríguez R, and Keppel-Jones K (2001) Recent advances in capillary electrophoresis and capillary electrochromatography of pollutants. *Electrophoresis* 22: 4262–4280.
- Fukushi K, Takeda S, Chayama K, and Wakida S (1999) Application of capillary electrophoresis to the analysis of inorganic ions in environmental samples. *Journal of Chromatography A* 834: 349–362.
- Heinig K and Vogt C (1999) Determination of surfactants by capillary electrophoresis. *Electrophoresis* 20: 3311–3328.
- Johns C, Macka M, and Haddad PR (2003) Enhancement of detection sensitivity for indirect photometric detection of anions and cations in capillary electrophoresis. *Electrophoresis* 24: 2150–2167.
- Li SFY (1996) *Capillary electrophoresis. Principles, Practice and Applications*. Amsterdam: Elsevier.
- Liu YM and Cheng JK (2003) Elemental speciation analysis in capillary electrophoresis. *Electrophoresis* 24: 1993–2012.
- Martinez D, Cugat MJ, Borrull F, and Calull M (2000) Solid-phase extraction coupling to capillary electrophoresis with emphasis on environmental analysis. *Journal of Chromatography A* 902: 65–89.
- Paull B and King M (2003) Quantitative capillary zone electrophoresis of inorganic anions. *Electrophoresis* 24: 1892–1934.
- Picó Y, Rodríguez R, and Mañes J (2003) Capillary electrophoresis for the determination of pesticide residues. *Trends in Analytical Chemistry* 22: 133–151.
- Song LG, Xu ZH, Kang JW, and Cheng JK (1997) Analysis of environmental pollutants by capillary electrophoresis with emphasis on micellar electrokinetic chromatography. *Journal of Chromatography A* 780: 297–328.
- Sovocool GW, Brumley WC, and Donnelly JR (1999) Capillary electrophoresis and capillary electrochromatography of organic pollutants. *Electrophoresis* 20: 3297–3310.
- Timerbaev AR, Dabek-Zlotorzynska E, and van der Hoop MAGT (1999) Inorganic environmental analysis by capillary electrophoresis. *Analyst* 124: 811–826.
- Willauer HD and Collins GE (2003) Analysis of inorganic and small organic ions with the capillary electrophoresis. *Electrophoresis* 24: 2193–2207.

## Food Chemistry Applications

**R A Frazier**, University of Reading, Reading, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Capillary electrophoresis (CE) has become an established technique in chemical analysis, including many applications in food chemistry. CE is an automated technique that bridges the gap between high-performance liquid chromatography (HPLC) and traditional electrophoresis techniques, offering separation modes whose basis ranges from zone electrophoresis through to chromatography. The most common

modes of separation employed for food applications are capillary zone electrophoresis (CZE) and micellar electrokinetic chromatography (MEKC). These two modes are complementary not only to each other, but also to the separation modes of HPLC. CE importantly offers the ability to control selectivity through straightforward changes to the run buffer composition while usually requiring no changes to the capillary. This is a considerable advantage in terms of the development of new methods, both in time and cost savings.

CZE separates sample components based on differences in their electrophoretic mobility, which is determined by their charge-to-size ratio. CZE is therefore only applicable to the separation of charged

molecules. Neutral molecules are not resolved from one another and migrate through the capillary under the influence of the electroosmotic flow (EOF).

MEKC can be used to separate both charged and neutral molecules. In MEKC, a surfactant (e.g., sodium dodecyl sulfate, SDS) is added to the run buffer at a concentration in excess of its critical micelle concentration. The surfactant micelles that are formed act as a pseudostationary phase, enabling separation upon the basis of partitioning between the hydrophobic core of the micelles and the aqueous buffer. This partitioning mechanism acts to enable the resolution of neutral molecules, whose migration with respect to the EOF will be influenced by the degree of partition into the micellar phase. Charged sample components are separated on the basis of charge-to-size ratio, although migration times and order may be influenced additionally by the synergistic partitioning mechanism.

CE applications in the field of food chemistry are becoming increasingly widespread, especially in the wake of several key developments in instrument technology that have brought significant improvements in terms of sensitivity, reproducibility, and flexibility. CE is applicable to as extensive a range of food components as it is possible to analyze by HPLC, but has the benefit of greater resolving power leading to shorter analytical run times. This article presents an overview of the current applications of CE to food chemistry, which are grouped in terms of the various food component types. The aim has been to show the diversity in these applications rather than to present an exhaustive review; therefore, the Further Reading list includes review articles that will provide more in-depth coverage.

## Proteins

Food proteins are vital nutrients and possess functional properties that can be exploited to modify and stabilize processed food structure. Protein content is therefore a significant factor in determining the commercial value of major food commodities such as cereal grains and milk. As a result there is intense interest in the development of improved methods for food protein analysis and the separation mechanisms provided by CE have brought the technique to the forefront of recent developments. The majority of methods employ straightforward CZE and MEKC based methodology, but in some cases capillary gel electrophoresis (CGE) or SDS-CGE (available as commercial kits) are employed. These latter separation modes employ a gel-filled capillary that allows

separation based on a molecular sieving mechanism. The major applications of CE to food protein analysis are in commercially important areas including the determination of protein quality, product authenticity, and processing and storage effects.

### Milk Proteins

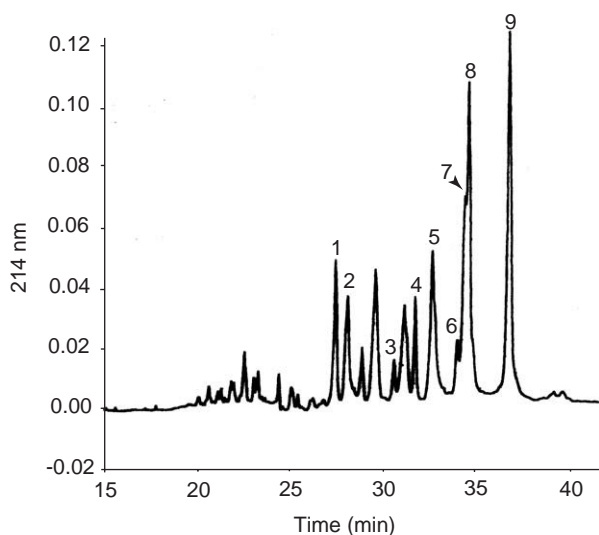
Milk proteins are the most-studied food proteins and consequently the application of CE to their analysis is widespread. Milk proteins can be separated using CE according to genetic polymorphism, which enables the application of CE methods to establish genetic origins of milks used to produce various dairy products. Proteins and peptides are also key markers for studying the development of biochemical and organoleptic properties during processing of dairy products, such as heat treatment and drying. Such procedures can lead to protein denaturation and undesirable Maillard reactions, which may affect nutritional and flavor properties. Furthermore, proteolysis is an integral feature of the production of many dairy products, primarily cheese, leading to the formation of biochemical breakdown products that can have important implications for product quality (e.g., flavor). CE is a valuable tool for the separation and characterization of proteins and peptides following technological or proteolytic treatments.

Milk proteins consist of casein and whey protein fractions. Caseins are the major protein in milk and are sub-grouped into  $\alpha$ -,  $\beta$ -, and  $\kappa$ -caseins. Caseins form micelles that constitute the colloidal phase of milk, and the formation of these aggregates during CE analysis is prevented by adding  $6 \text{ mol l}^{-1}$  urea to the sample and separation buffers. Using a hydrophilic coated capillary and a low pH buffer (pH 3.0) containing polymeric additives (e.g., methylhydroxyethylcellulose) enables the separation of  $\beta$ -casein into its genetic variants according to differences in the number of basic amino acid residues. The use of these conditions is primarily in order to avoid protein adsorption to the capillary walls, which can lead to peak shape distortion and poor resolution.

As mentioned earlier, proteins can be used as markers for establishing genetic origin, which is a major issue concerning dairy product quality. A prime example is the adulteration of caprine and ovine milk products with small amounts of bovine milk. This occurs because of the high market price of caprine and ovine milk and seasonal fluctuations in their production. CE offers a simple and rapid approach to the detection of adulteration because it is able to resolve milk proteins according to

genetic polymorphism. Figure 1 shows the analysis of the casein fraction of a milk mixture of one-third of each species from which it is possible using multivariate regression analysis to identify and quantify the presence of caseins from the three species.

Whey proteins, which are a group of globular proteins that include  $\beta$ -lactoglobulin and  $\alpha$ -lactalbumin ( $\alpha$ -La), are readily separated by a variety of CE approaches. The method used is normally dependent on the application, and whether resolution of whey proteins and caseins is required within a single analysis. The US Customs Service use CE as a rapid method (5 min run-time) for the analysis of  $\alpha$ -La in milk protein powders to aid the enforcement of import duties on milk protein concentrates, whose duty classification is partly defined by their content of  $\alpha$ -La.  $\alpha$ -La is present at low levels in milk (<5% of total protein) and is subject to degradation/denaturation during processing of milk powders. Its analysis by CZE is simple, using a borate run buffer (pH 8.0) for separation after extraction of the protein into an acetic acid medium.



**Figure 1** Electropherograms of the casein fraction of a mixture of one-third each of bovine, ovine, and caprine milk. Peak identification: 1, bovine  $\alpha_{S1}$ -casein; 2, ovine  $\alpha_{S1}$ -casein; 3, bovine  $\kappa$ -casein; 4, ovine  $\kappa$ -casein; 5, bovine  $\beta$ -casein A<sup>1</sup>; 6, caprine  $\kappa$ -casein; 7, bovine  $\beta$ -casein A<sup>2</sup>; 8, ovine  $\beta_2$ -casein and caprine  $\beta_2$ -casein; 9, ovine  $\beta_1$ -casein and caprine  $\beta_1$ -casein. Conditions: hydrophilic coated fused-silica capillary 50  $\mu$ m i.d.  $\times$  57 cm; separation buffer, 6 mol l<sup>-1</sup> urea, 0.32 mol l<sup>-1</sup> citric acid, 20 mmol l<sup>-1</sup> sodium citrate, pH 3.0, containing 0.5 g l<sup>-1</sup> methylhydroxyethyl-cellulose; applied voltage, 25 kV; injection, 15 s; temperature, 45°C. (Reprinted with permission from Molina E, Martin-Alvarez PJ, and Ramos M (1999) Analysis of cows', ewes' and goats' milk mixtures by capillary electrophoresis: Quantification by multivariate regression analysis. *International Dairy Journal* 9: 99–105; © Elsevier.)

## Meat Proteins

Muscle proteins are an important component of meat and can be classified according to solubility as sarcoplasmic (water soluble), myofibrillar (salt soluble), or stromal (insoluble) proteins. The application of CE to the analysis of meat proteins has been predominantly for separation of sarcoplasmic proteins in aqueous extracts from fish, bovine, and chicken muscle. The sarcoplasmic proteins that are present are mainly metabolic enzymes and therefore their separation profiles are useful for the purpose of species identification. Some reports also exist of the simultaneous separation of sarcoplasmic and myofibrillar meat proteins using SDS-CGE.

## Cereal Proteins

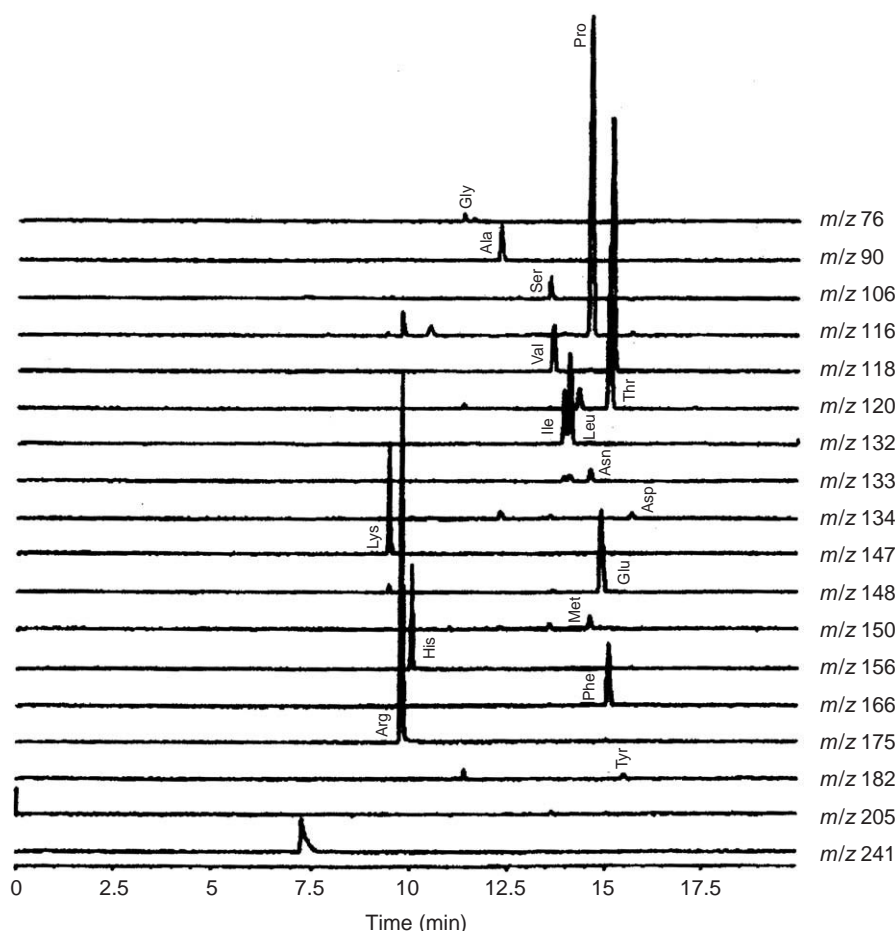
Cereal proteins play an important role in food functional properties and have been traditionally separated and characterized using slab-gel electrophoresis techniques. CE is emerging as a valuable new technique for the analysis of cereal proteins, particularly to bring improvements in resolution and speed of analysis. Methods based on CZE and SDS-CGE can be used for a variety of applications, including cultivar differentiation and purity screening.

## Amino Acids

Amino acids are important nutrients and can be used as indicators of protein composition. However, they can pose difficulties for analysis because the majority of amino acids do not possess a chromophore. For this reason the major focus for method development in CE is not to achieve separation, which is relatively straightforward, but to achieve sensitive and selective detection. Two routes are possible to meet this end, which are either amino acid derivatization followed by laser-induced fluorescence (LIF) or electrochemical detection, or detection of underivatized amino acids by conductivity or electrospray ionization mass spectrometry (ESI-MS). An advantage of ESI-MS is that it gives greater accuracy in the assignment of electropherogram peaks, which can be vital for complex sample matrices. Figure 2 shows an example of the application of ESI-MS to the detection of amino acids in soy sauce. In this example, free amino acids are analyzed simultaneously using a low acidic pH to confer positive charge on the analytes.

## Maillard Reaction Products

The Maillard reaction is of critical importance in food chemistry and is the origin of many colors and



**Figure 2** CE-ESI-MS selected ion electropherograms for amino acids in soy sauce. Experimental conditions: fused-silica capillary 50  $\mu\text{m}$  i.d.  $\times$  100 cm; electrolyte, 1  $\text{mol l}^{-1}$  formic acid; applied voltage, 30 kV; injection 3 s at 50 mbar; temperature, 20  $^{\circ}\text{C}$ ; sheath liquid, 10  $\mu\text{l min}^{-1}$  of 5  $\text{mol l}^{-1}$  ammonium acetate in 50% (v/v) methanol–water. (Reprinted with permission from Soga T and Heiger DN (2000) Amino acid analysis by capillary electrophoresis electrospray ionization mass spectrometry. *Analytical Chemistry* 72: 1236–1241; © American Chemical Society.)

flavors, including some that are undesirable. The Maillard reaction is the nonenzymic browning reaction of reducing sugars with amines and typically involves amino acids, proteins, and peptides. The chemistry underlying the Maillard reaction is complex and it is therefore a major challenge to effectively separate and characterize the complete range of its products. These products often have closely related chemical or biochemical structures, thus the resolving power of CE has been widely applied to tackle their separation. CE is attractive for these applications because of its simplicity and high resolving power combined with minimal requirements in terms of sample preparation.

## Vitamins

Vitamins are a structurally heterogeneous group of compounds that are essential in the diet for the

maintenance of healthy growth and development. Vitamins can be classified into two main groups: the water-soluble vitamins (B-group and C) and the fat-soluble vitamins (A, D, E, and K). Water-soluble vitamins are readily separated by CZE and MEKC methods, and several approaches have been developed for their determination in foods. There are fewer applications that have been developed for the determination of fat-soluble vitamins in foods. However, it should be noted that it is possible to separate mixtures of fat- and water-soluble vitamins using MEKC or by methods that employ microemulsion droplets as the pseudostationary phase. Separations of such mixtures have not been realized by other separation techniques.

### Vitamin A

Vitamin A is present in foods as retinol and retinyl esters (usually retinyl acetate or retinyl palmitate). Retinoids are often employed as food additives

because of their beneficial antioxidant effects. To resolve mixtures of retinoids, which are fat soluble, the use of MEKC methods is required. The choice of micellar phase is important, because common MEKC micellar phases such as SDS or the bile salts sodium cholate and sodium deoxycholate all exhibit a strong solubilization effect on retinoids. They are therefore unsuitable since poor resolution is obtained. However, by using a mixed micellar phase comprising sodium deoxycholate and polyoxyethylene(23) dodecanol (Brij 35) it is possible to achieve separation of, for example, retinol and the retinyl esters.

### Vitamin B<sub>1</sub>

Thiamine (vitamin B<sub>1</sub>) occurs in foods in free and bound forms, the free form predominates in cereals and plants, whereas the pyrophosphate ester is the main form in animal products. Acid hydrolysis is required to release thiamine from the food matrix. Enzymatic hydrolysis is then needed to convert phosphate esters to thiamine. Prior to CE analysis it is necessary to clean up samples by using ethanol to precipitate protein and by passing through an ion-exchange resin. Thiamine has been determined in meat and milk samples using MEKC with ultraviolet (UV) detection at 254 nm, obtaining comparable sensitivity to that achieved by HPLC using an ion-pair reversed-phase column with postcolumn derivatization and fluorescence detection.

### Vitamin B<sub>2</sub>

The analysis of riboflavin (vitamin B<sub>2</sub>) and related flavin cofactors (flavin adenine dinucleotide and flavin mononucleotides) can take advantage of their native fluorescence, which enables LIF detection to be used. LIF detection is extremely sensitive and imparts a high degree of selectivity for the determination of flavins from food samples. Taking advantage of this selectivity, riboflavin can be separated using a straightforward CZE approach with a basic running buffer (e.g., phosphate buffer at pH 9.8). This has been applied for analysis of riboflavin in wines, vegetables, wheat flour, and tomatoes.

### Vitamin C

The principal compound with vitamin C activity that is found naturally is L-ascorbic acid. L-Ascorbic acid is readily oxidized in solution to L-dehydroascorbic acid, which means that methods for the determination of total vitamin C content require quantification of both forms. Acids are normally used for extraction to protect vitamers from degradation and to precipitate

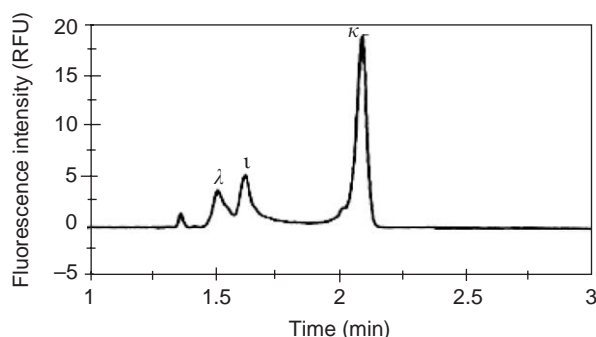
associated proteins. Depending on the complexity of the sample matrix, ascorbic acid is separated using either CZE or MEKC and detected by UV at 254 nm.

### Niacin

Niacin is found in foods as nicotinic acid and nicotinamide, which are the free forms, and as the co-enzymes nicotinamide adenine dinucleotide and nicotinamide adenine dinucleotide phosphate. Niacin is stable and can be extracted into acids or alkalis. Niacin has been determined by CE in a range of food matrices, such as meat, fish, fruit, and vegetables, and CE is generally found to be faster and more cost-effective in comparison to established HPLC methods.

## Carbohydrates

A major issue for the determination of carbohydrates is the lack of a chromophore suitable for UV detection. In addition, most of the carbohydrate sugars are neutral and therefore cannot be separated by CZE. Both issues can be addressed by the use of an alkaline borate sample and running buffer, since sugar-borate complexes are negatively charged and can be detected by UV at 195 nm. However, more sensitive detection can be achieved by derivatization of carbohydrates with a fluorescent label via reductive amination of the reducing end. **Figure 3** shows the separation of 9-aminopyrene-1,4,6-trisulfonic acid (APTS) labeled  $\iota$ -,  $\kappa$ -, and  $\lambda$ -carrageenans by a CZE method using LIF detection. Carrageenans are sulfated linear polysaccharides that are used as



**Figure 3** Separation of mixture of APTS labelled  $\iota$ -,  $\kappa$ -, and  $\lambda$ -carrageenans, each 0.06 mg ml<sup>-1</sup>. Experimental conditions: fused-silica capillary 50  $\mu$ m i.d.  $\times$  47 cm; electrolyte, 25 mol l<sup>-1</sup> ammonium acetate, pH 8.0; field strength,  $6.4 \times 10^4$  V m<sup>-1</sup>; temperature, 50°C. (Reproduced with permission from Mangin CM, Goodall DM, and Roberts MA (2001) Separation of  $\iota$ -,  $\kappa$ - and  $\lambda$ -carrageenans by capillary electrophoresis. *Electrophoresis* 22: 1460–1467; © Wiley-VCH.)

gelling agents ( $\iota$ - and  $\kappa$ -carrageenans) or thickeners ( $\lambda$ -carrageenan) in food systems. The quantitative separation of carrageenan subtypes is possible by anion exchange or gas chromatography, but requires hydrolysis or methanolysis prior to analysis, which could lead to loss of molecular information contained in the intact structure. The analysis by CE is extremely rapid and uses an electrolyte system that would be compatible for use with ESI-MS detection.

## Food Additives

Food additives can be any substance added to a food to perform a particular function. The term is most commonly applied to preservatives, colors, and flavorings. Many food additives are synthetic molecules and therefore legislative requirements are in place to control their use.

### Preservatives

Several simple and rapid methods have been reported for the analysis of preservatives, which are additives used to prevent biological degradation of foods. Common preservatives include benzoic acid, sorbic acid, sulfite, and *p*-hydroxybenzoic acid esters. The resolving power of CZE allows the facile separation of complex mixtures of preservatives whose separation by HPLC is less straightforward. For HPLC, the wide variation in polarity of the acid preservatives in comparison to the esters makes simultaneous determination difficult without the use of ion-pairing reagents. However, in CZE the charge-to-size ratio of the analytes is sufficient even to separate the methyl, ethyl, propyl, and butyl esters of *p*-hydroxybenzoic acid, which all have similar  $pK_a$  values but different molecular weights.

Anionic preservatives, such as sulfite (as sulfate), nitrite, and nitrate can be determined by reversing the polarity of the applied voltage and employing a cationic polymer capillary coating (e.g., polyethyleneimine) or a cationic surfactant buffer additive (e.g., hexadecyltrimethylammonium bromide, CTAB) to reverse the direction of EOF. This strategy, also commonly employed for the analysis of inorganic ions, allows faster separation since the anions migrate in the same direction as the EOF (anions conventionally migrate in the opposite direction to the EOF). Sulfate does not absorb in the UV region, so indirect detection is used with chromate as the absorbing ion. Advantages of CE over classical colorimetry and titrimetry methods include shorter analysis time, automation, and lower sample consumption.

### Antioxidants

Antioxidants occur naturally in many foods, but they are also important additives in fat-rich foods in order to prevent rancidity. Natural and synthetic antioxidants are commonly polyphenols, although some other natural antioxidants include ascorbic acid and tocopherols. Synthetic polyphenolic antioxidants are commonly separated in CZE or MEKC modes using an alkaline borate buffer. CE methodologies for the quantitative determination of natural polyphenolic antioxidants include the use of electrochemical detection, microemulsion electrokinetic chromatography (MEEKC) and nonaqueous CE (NACE). Electrochemical detection has been advocated for its high sensitivity and selectivity for electroactive analytes such as phenolic compounds. Indeed, the selectivity of electrochemical detection may be a useful approach to the determination of phenolic compounds in complex matrices, since it would minimize the extent of sample cleanup required. MEEKC is an interesting alternative to MEKC methods for phenolic compounds since it gives different selectivity, which can be valuable to separate structurally similar compounds such as the green tea catechins. The MEEKC mode of separation for neutral compounds is based on analytes partitioning between mobile charged oil droplets and an aqueous buffer phase. The microemulsion droplets are normally stabilized by the addition of a surfactant. NACE similarly allows improved selectivity and separation efficiency over aqueous-based CE methods for the separation of theaflavins.

### Food Colors

The majority of synthetic food colors are ideal candidates for separation by CE. This is because they commonly contain sulfonic acid or carboxylic acid functional groups that form negatively charged colored ions at alkaline pH. CE has also been applied to the analysis of natural colors such as caramels, which occur as four distinct classes according to the reactants used during sugar caramelization. CE analysis can be used to identify and quantitate the class of caramel present in a sample.

### Artificial Sweeteners

Sweeteners such as aspartame, saccharin, and acesulfame-K are added as sugar substitutes to reduced-calorie food and beverage products. All three of these commonly used sweeteners can be resolved in samples from reduced-calorie products using a mixed phosphate/borate buffer system at alkaline pH and including sodium deoxycholate as a micellar phase.



### Simultaneous Determination of Additive Mixtures

Some of the most elegant examples of the resolving power of CE have been applications for the separation of mixtures of food additives. There is a real need for such methods since mixtures of additives are often present in foods and beverages. This especially evident in reduced-calorie soft drinks, which are usually colored and include artificial preservatives to counter the loss of sugar as a natural preservative. **Figure 4** shows an example of a simple MEKC method that is able to resolve 13 components in a mixture of preservatives, sweeteners, and colors within a runtime of just 14 min. The use of MEKC is necessary to resolve benzoic acid and saccharin, which are not baseline resolved under similar CZE conditions. The mixture can be separated using CZE with a longer capillary, but this leads to long analysis times of the order of 30 min.

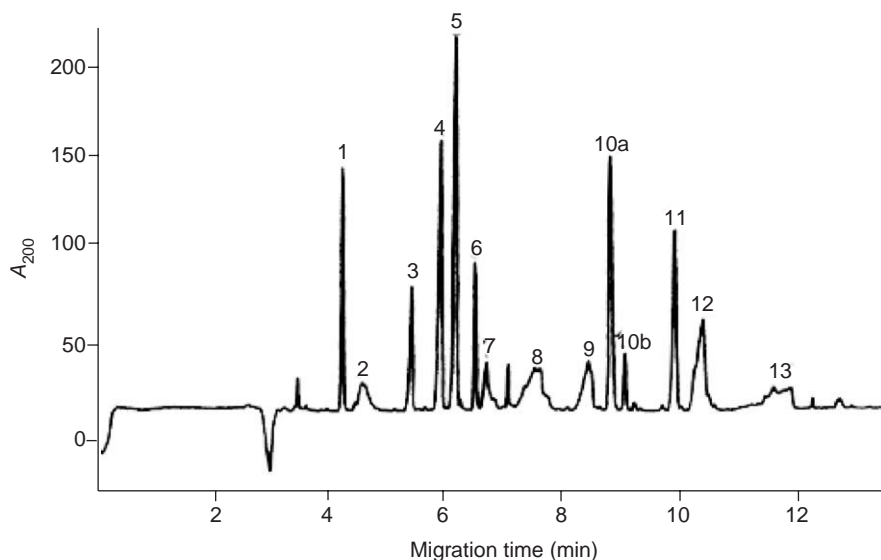
### Organic Acids

CE is widely accepted as a rapid and cheap approach for the determination of organic acids, which occur naturally in many foods and beverages, including beer, dairy products, and tomatoes. Organic acids are important contributors to the organoleptic properties of these and other foods. The rapid analysis of organic acids can be performed by CZE using reversed polarity and buffer additives such as CTAB or hexadimethrine bromide to reverse the EOF direction. This approach has been used for the analysis of tartaric

acid, which is an important by-product that is produced in large quantities during wine production and cannot be rejected into the environment. Wineries precipitate tartaric acid using calcium hydroxide to produce a solid residue containing calcium tartrate that can be later purified to recover tartaric acid. The analytical methodologies used to determine the tartaric acid concentration in these solid wine residues are often long and tedious and have poor reproducibility. However, CE can be used to analyze for tartaric acid very rapidly (2 min runtime) and reproducibly, giving a viable improvement to existing methodologies.

Organic acids are also added to products fraudulently in order to mask poor quality or to disguise origin. CZE has been validated for measuring citric, isocitric, malic, and tartaric acids as authenticity markers in orange juices. These acids can be separated after simple sample preparation involving only dilution and filtration. This approach is applicable to authenticity determination based on the unique organic acid profiles of different fruit juices. In beer, organic hop acids, specifically iso- $\alpha$ -acids and their reduced derivatives, are sources of bitter flavor and can be readily separated by MEKC to generate characteristic quality profiles.

A recent development for the analysis of organic acids has been the development of a novel device based on the principles of CZE that enables the fast detection of anionic components, including organic acids and inorganic ions, in sugar and wine samples. The device uses a shortened fused silica capillary to



**Figure 4** Separation of a standard sample containing 13 food additives. Peak identification: 1, caffeine; 2, aspartame; 3, sorbic acid; 4, benzoic acid; 5, saccharin; 6, green S; 7, acesulfame K; 8, sunset yellow FCF; 9, quinoline yellow; 10, brilliant blue FCF; 11, carmoisine; 12, ponceau 4R; 13, black PN. Conditions: fused-silica capillary 50  $\mu\text{m}$  i.d.  $\times$  48.5 cm with  $\times$  3 extended path length detection cell; separation buffer, 20  $\text{mmol l}^{-1}$  carbonate, pH 9.5, containing 65  $\text{mmol l}^{-1}$  SDS; applied voltage, 20 kV; temperature, 25  $^{\circ}\text{C}$ .



rapidly analyze samples in a runtime of <1 min with acceptable sensitivity that would suggest potential application as a near-real-time method for monitoring processing steps in sugar and wine production.

## Agricultural and Veterinary Residues

CE is viewed as a potentially important technique for the determination of pesticide residues in environmental and food matrices. This application needs simultaneous separation of multicomponent mixtures with low limit of detection. The application of CE for the determination of pesticide residues has been aided recently by the development of improved methods of sample enrichment and detection. These methods can overcome the sensitivity limitations presented by the small sample volumes that are normally analyzed. It is somewhat perverse that this limitation should also be one of the technique's advantages in terms of reagent consumption.

Various veterinary residues can be found in food, particularly antibacterial agents used as curative or prophylactic treatments in livestock. A concern with the presence of residual levels of these antibacterial drugs in foods is the increase in antimicrobial resistance. CE methods have been applied to the determination of drug residues in fish and chicken muscle. The quantitative analysis of oxolinic acid fish muscle can be achieved using CZE with a basic phosphate buffer (pH 9) after solid-phase extraction. Enrofloxacin and its metabolite ciprofloxacin are detectable in chicken muscle using LIF detection after separation by CZE in an acidic phosphate buffer (pH 2.2). Oxolinic acid and flumequine can be simultaneously determined in chicken by using a basic phosphate buffer (pH 8.02) and UV-visible diode array detection.

See also: **Capillary Electrophoresis:** Overview. **Carbohydrates:** Overview. **Fluorescence:** Derivatization. **Food and Nutritional Analysis:** Overview. **Liquid**

**Chromatography:** Amino Acids; Food Applications. **Mass Spectrometry:** Electrospray. **Micellar Electrokinetic Chromatography.** **Pesticides.** **Proteins:** Foods. **Sweeteners.** **Vitamins:** Overview.

## Further Reading

- Bean SR and Lookhart GL (2001) Recent developments in high-performance capillary electrophoresis of cereal proteins. *Electrophoresis* 22: 1503–1509.
- Boyce MC (2001) Determination of additives in food by capillary electrophoresis. *Electrophoresis* 22: 1447–1459.
- Eash DT and Bushway RJ (2000) Herbicide and plant growth regulator analysis by capillary electrophoresis. *Journal of Chromatography A* 880: 281–294.
- Frazier RA (2001) Recent advances in capillary electrophoresis methods for food analysis. *Electrophoresis* 22: 4197–4206.
- Frazier RA and Papadopoulou A (2003) Recent advances in the application of capillary electrophoresis for food analysis. *Electrophoresis* 24: 4095–4105.
- Frazier RA, Ames JM, and Nursten HE (1999) The development and application of capillary electrophoresis methods for food analysis. *Electrophoresis* 20: 3156–3180.
- Frazier RA, Ames JM, and Nursten HE (2000) *Capillary Electrophoresis for Food Analysis: Method Development*. Cambridge: Royal Society of Chemistry.
- Gutierrez JEN and Jakobovits L (2003) Capillary electrophoresis of  $\alpha$ -lactalbumin in milk powders. *Journal of Agricultural and Food Chemistry* 51: 3280–3286.
- O'Flaherty B, Yang W-P, Sengupta S, and Cholli AL (2001) Fast detection of anionic components in sugar and wine samples using a novel device based on capillary zone electrophoresis. *Food Chemistry* 74: 111–118.
- Pico Y, Rodriguez R, and Manes J (2003) Capillary electrophoresis for the determination of pesticide residues. *Trends in Analytical Chemistry* 22: 133–151.
- Recio I, Ramos M, and Lopez-Fandino R (2001) Capillary electrophoresis for the analysis of food proteins of animal origin. *Electrophoresis* 22: 1489–1502.
- Trenerry VC (2001) The application of capillary electrophoresis to the analysis of vitamins in food and beverages. *Electrophoresis* 22: 1468–1478.

## Clinical Applications

**Z K Shihabi**, Wake Forest University, Winston-Salem, NC, USA

## Introduction

Capillary electrophoresis (CE) is a general, relatively new analytical technique for separation and quantification of a wide variety of molecules including

rapidly analyze samples in a runtime of <1 min with acceptable sensitivity that would suggest potential application as a near-real-time method for monitoring processing steps in sugar and wine production.

## Agricultural and Veterinary Residues

CE is viewed as a potentially important technique for the determination of pesticide residues in environmental and food matrices. This application needs simultaneous separation of multicomponent mixtures with low limit of detection. The application of CE for the determination of pesticide residues has been aided recently by the development of improved methods of sample enrichment and detection. These methods can overcome the sensitivity limitations presented by the small sample volumes that are normally analyzed. It is somewhat perverse that this limitation should also be one of the technique's advantages in terms of reagent consumption.

Various veterinary residues can be found in food, particularly antibacterial agents used as curative or prophylactic treatments in livestock. A concern with the presence of residual levels of these antibacterial drugs in foods is the increase in antimicrobial resistance. CE methods have been applied to the determination of drug residues in fish and chicken muscle. The quantitative analysis of oxolinic acid fish muscle can be achieved using CZE with a basic phosphate buffer (pH 9) after solid-phase extraction. Enrofloxacin and its metabolite ciprofloxacin are detectable in chicken muscle using LIF detection after separation by CZE in an acidic phosphate buffer (pH 2.2). Oxolinic acid and flumequine can be simultaneously determined in chicken by using a basic phosphate buffer (pH 8.02) and UV-visible diode array detection.

See also: **Capillary Electrophoresis:** Overview. **Carbohydrates:** Overview. **Fluorescence:** Derivatization. **Food and Nutritional Analysis:** Overview. **Liquid**

**Chromatography:** Amino Acids; Food Applications. **Mass Spectrometry:** Electrospray. **Micellar Electrokinetic Chromatography.** **Pesticides.** **Proteins:** Foods. **Sweeteners.** **Vitamins:** Overview.

## Further Reading

- Bean SR and Lookhart GL (2001) Recent developments in high-performance capillary electrophoresis of cereal proteins. *Electrophoresis* 22: 1503–1509.
- Boyce MC (2001) Determination of additives in food by capillary electrophoresis. *Electrophoresis* 22: 1447–1459.
- Eash DT and Bushway RJ (2000) Herbicide and plant growth regulator analysis by capillary electrophoresis. *Journal of Chromatography A* 880: 281–294.
- Frazier RA (2001) Recent advances in capillary electrophoresis methods for food analysis. *Electrophoresis* 22: 4197–4206.
- Frazier RA and Papadopoulou A (2003) Recent advances in the application of capillary electrophoresis for food analysis. *Electrophoresis* 24: 4095–4105.
- Frazier RA, Ames JM, and Nursten HE (1999) The development and application of capillary electrophoresis methods for food analysis. *Electrophoresis* 20: 3156–3180.
- Frazier RA, Ames JM, and Nursten HE (2000) *Capillary Electrophoresis for Food Analysis: Method Development*. Cambridge: Royal Society of Chemistry.
- Gutierrez JEN and Jakobovits L (2003) Capillary electrophoresis of  $\alpha$ -lactalbumin in milk powders. *Journal of Agricultural and Food Chemistry* 51: 3280–3286.
- O'Flaherty B, Yang W-P, Sengupta S, and Cholli AL (2001) Fast detection of anionic components in sugar and wine samples using a novel device based on capillary zone electrophoresis. *Food Chemistry* 74: 111–118.
- Pico Y, Rodriguez R, and Manes J (2003) Capillary electrophoresis for the determination of pesticide residues. *Trends in Analytical Chemistry* 22: 133–151.
- Recio I, Ramos M, and Lopez-Fandino R (2001) Capillary electrophoresis for the analysis of food proteins of animal origin. *Electrophoresis* 22: 1489–1502.
- Trenerry VC (2001) The application of capillary electrophoresis to the analysis of vitamins in food and beverages. *Electrophoresis* 22: 1468–1478.

## Clinical Applications

**Z K Shihabi**, Wake Forest University, Winston-Salem, NC, USA

## Introduction

Capillary electrophoresis (CE) is a general, relatively new analytical technique for separation and quantification of a wide variety of molecules including

those of clinical interest utilizing narrow bore capillaries under high voltage. The flexibility of this technique stems from the ability to incorporate different additives easily in the separation buffer, which can interact with some of the analytes relative to others to alter their velocity in order to achieve the desired separation. Most of the clinical tests can be adapted to CE; however, in practice, some are better suited than others for analysis by this method. In general, large molecules, such as proteins and DNA and those that carry a charge are more suited to be analyzed by CE. Here, we discuss the most important applications.

## Nucleic Acids Analysis

The analysis of DNA by CE represents one of the best examples for application of this technique in clinical analysis where the low cost, the full automation, and the speed of analysis compared to the slab gels accelerated the implementation of this technique in sequencing the human genome. Polymerase chain reaction (PCR), the Genome project, viral detection, blood transfusion safety, and forensic identification are all pushing the widespread use of DNA analysis especially by CE further than most investigators have expected.

DNA is a complex linear biopolymer composed of long repeat of nucleotides, which carry the genetic information through their sequence in the form of different genes. It is present in a complementary base pair (bp) as a double-stranded helix. Because the DNA is present in very huge long strands, it needs to be cut first into smaller fragments by special enzymes (nucleases) before analysis. These fragments can be studied for the number of the bases (size), their sequences, or for the presence of mutations. Since these fragments are present in low concentration, below the detection of the instruments, the next step is amplification through PCR in the presence of a primer, enzyme, and nucleotides.

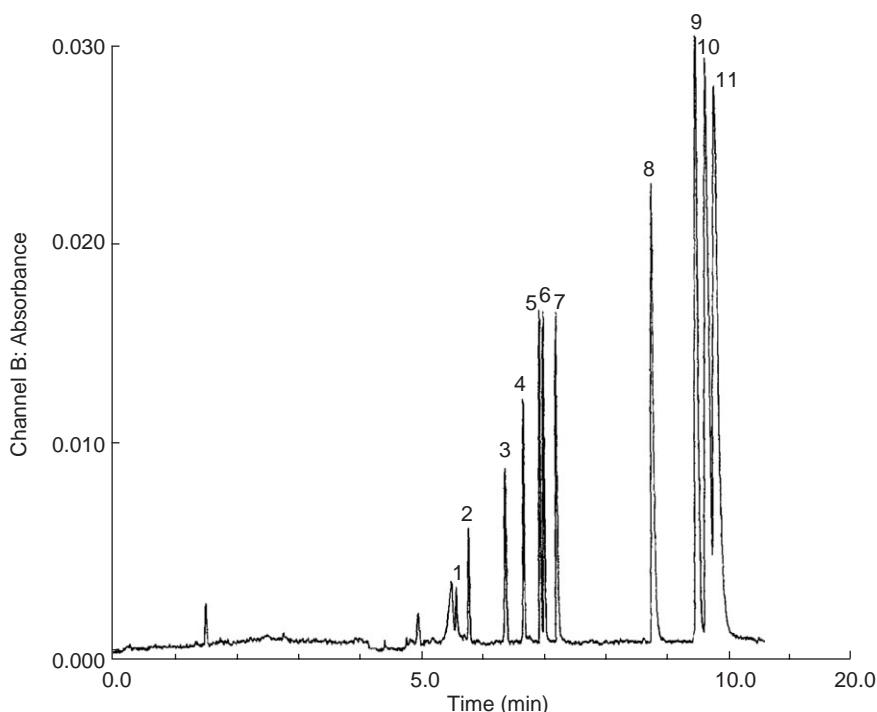
DNA molecules all carry essentially the same charge; thus, they cannot be separated by electrophoresis simply based on charge alone. However, the separation can be achieved much better based on size by molecular sieving. Gels with different pores or long polymers are added to the buffer to retard the migration of large fragments relative to the smaller ones. Traditionally, agarose gels are used to separate double-stranded DNA (dsDNA) fragments after digestion for mapping; while slab polyacrylamide gel (SPG) is used to separate single-stranded DNA (ssDNA) fragments for sequencing.

Initially, CE methods for DNA separations attempted to mimic those of the SPG by using the same gels. Capillaries of 20–80 cm length and 50–75  $\mu\text{m}$  diameter were filled with cross-linked polyacrylamide gels similar to those used in SPG. DNA samples are injected hydrodynamically or by electroinjection. Small DNA fragments migrate rapidly and emerge first while the larger ones migrate slowly and emerge later. Thus, the migration time is used to characterize these fragments. These gel-filled capillaries gave excellent separations. They were able to resolve a single base pair (1 bp) with good resolution up to 450 bp; however, they had several practical problems such as filling the capillary without introducing air bubbles, short lifetime, and high cost. However, later on, solutions of uncrossed polymer network have been introduced. They were easier to work with, less expensive, and gave better precision. These polymers no doubt have spurred the interest in employing CE for DNA analysis. The polymers include uncrossed polyacrylamide, many derivatives of cellulose, poly(vinylpyrrolidone), poly(ethylene oxide), and dextrans, with some of these being proprietary. These polymers are replaceable gels each time the capillary is filled. They separate fragments of DNA <100 bp to >2 kbp, and in some instances up to 12 kbp with resolution of  $\sim 10$  bp (Figure 1).

The dsDNA fragment separation in CE is accomplished as in the SPG based on molecular sieving after the PCR step of the DNA fragment, while the ssDNA requires also a denaturing substance such as urea or formamide. Some workers found cellulose polymers are more suited for large fragments while the linear polyacrylamide is more suited for the smaller fragments. However, the optimum concentration and the molecular weight are more important than the type of the polymer.

For DNA sequencing, enzymatic synthesis of the purified template is performed in the presence of deoxynucleotides, and specific labeled (fluorescent) dideoxynucleotides (analogues) that terminate the DNA strand at specific nucleotides. Thus, fragments with different lengths and with different labels complementary to the original are synthesized. These are resolved for size and with detection for the fluorescent terminal end by CE. A single base-pair separation is important with the ability to sequence large fragments of DNA on the order of  $\sim 1$  kbp. This can be accomplished in the CE by  $\sim 80$  min. The addition of intercalator dyes such as ethidium bromide or YO-PRO stiffens the structure of the DNA and thus improves the separation.

CE has been used extensively for fragment size analysis up to several kilobases (Figure 1) and for DNA sequencing. The dsDNAs after PCR amplification



**Figure 1** Separation of 11 DNA fragments in the untreated capillary by electrokinetic injection for 90 s at 5 kV (fragments: 1, 72; 2, 118; 3, 194; 4, 234; 5, 271; 6, 281; 7, 301; 8, 603; 9, 872; 10, 1078; 11, 1353 bp). (Reprinted with permission from (1999) *Journal of Chromatography A* 853: 349–354; © Elsevier.)

have been utilized to identify bacteria and viruses. The sequence is used often to monitor the changes in the DNA of the virus (e.g., HIV) to match the drug treatment with the genetic makeup of the infectious agent.

The detection of DNA can be performed by ultraviolet (UV) detection. However, this has poor sensitivity and subject to much interference. The addition of special dyes such as YO-PRO1, BODIPY, and thiazole orange to the separation buffer in CE enables the use of laser-induced fluorescence (LIF), which improves the detection by a factor of ~100 times. LIF eliminates the interferences from the nucleotides in the reaction mixture, which absorb light in the UV region. Also, because the sample is diluted, matrix effects become negligible. This enables introducing the sample by electrokinetic injection with stacking (concentration on the capillary) too.

Genetic diagnosis of an inherited disease or cancer often involves analysis for unknown point mutations, as well as known mutations in several genes. There is great demand for detecting mutation due to single nucleotide polymorphisms (SNPs) by fast and simple methods without the lengthy steps of sequencing. Among several techniques denaturing gradient gel electrophoresis, heteroduplex analysis (HA), and single-strand conformation polymorphism (SSCP) have been used frequently in slab gels and have been

applied later on to CE. SSCP is one of the most frequently used methods for detecting unknown mutations. In this method, the strand is amplified through PCR and denatured through heat and formamide. The separated strands adopt special folded structures determined by their nucleotide sequences. A single base alteration is detected by SSCP when the folding of the single strand changes (conformational) sufficiently to alter its electrophoretic mobility. Change between the amplified DNA of the wild type and mutation is detected by CE. The gel concentration and the presence of additives are important for the detection, with the temperature effect on mobility being a key element. A short-chain polyacrylamide is helpful in this technique.

HA for detecting SNP has also been applied to CE. In HA, the PCR-amplified DNA of allelic fragments are denatured and re-annealed to give a mixture of four duplexes, two homoduplexes and two heteroduplexes, in the heterozygote samples. Heteroduplexes have an aberrant, distorted structure with bubbles or bulges at the sites of mismatched bases, and generally move more slowly in the gel than homoduplexes. Thermal-profile SSCP is unique to CE. In this method, a temperature program is used to follow the conformational changes. All these methods have been applied to the detection of SNPs of many genes, for example, all *BRCA1* and *BRCA2*

mutations, P 53, and C 677T methylene tetrahydrofolate reductase mutation. CE has the advantage of speed, low cost, and full automation over SPG. Several machines have been designed specifically for DNA analysis. Some of these instruments use multicapillaries (8–300) to speed up the analysis.

### DNA Chip

In order to speed up further the analysis and conserve the expensive reagents, the chip, a small piece of glass, silica, or plastic, has been introduced for chemical analysis with two distinct types of chips: binding and separation. The binding chip contains on its surface certain molecules that recognize and bind specific analytes such as DNA or protein. The DNA binding chip (DNA microarray chip) contains on its surface hundreds of certain oligonucleotides that bind complementary genes and thus hundreds or even thousands of genes can be analyzed simultaneously on a very small surface. The separation chip is etched on its surface with multiple microgrooves (typically about 5–20  $\mu\text{m}$  deep and  $\sim 5\text{ cm}$  long) in place of the capillaries or columns to serve for the separation.

The main advantages of the chip include very rapid separation (in terms of seconds) while the reagent consumption is minimal, with no moving parts to wear. This is achieved because of the narrow and short dimensions of the etched grooves. In addition to that, many channels can be etched, so several reactions can be carried out simultaneously in the same chip. Several analytical steps can also be performed on the chip such as reagent movement, mixing, separation, and detection. These chips and the instrument to detect the reaction are commercially available now. Some of these chips are aimed for use at the point of care (physician's office, or hospital bed).

### Proteins, Peptides, Polypeptides, and Amino Acids

Proteins and peptides are composed of amino acids. They are zwitterions carrying both positive and negative charges. The peptides and polypeptides are similar to proteins in structure but smaller in size. While DNA carries the genetic information, proteins perform catalytic, hormonal, and structural functions. In many of the proteins of clinical interest, such as the urinary protein uromodulin, their functions are not well understood; nevertheless, they remain to be important and vital. Proteins are important for their function as well as for their diagnostic significance. CE is useful for quantitation, purity

check, isoforms, and microheterogeneity detection of proteins.

The interest in protein separation for clinical diagnosis by electrophoresis dates back to the work of Arne Tiselius (around 1937). Because proteins have different amino acid compositions with different isoelectric points they tend to behave differently in capillary zone electrophoresis (CZE). For example, basic proteins tend to bind to the capillary wall and give distorted peak shape. To improve their separation by CE, different additives, high salts, or coated capillaries are used to decrease the binding to the walls.

In CE, after the sample is injected and separated it is detected by monitoring the absorbance of the peptide bond at 214 nm. Thus, the advantages of CE over agarose gel (AG) electrophoresis for analysis of proteins are the speed, automation, small sample volume, and avoidance of staining–destaining steps. This led some companies to design instruments dedicated only to protein analysis by CE. Furthermore, other manufacturers designed special CE instruments to perform capillary isoelectric focusing (CIEF) with absorption imaging detectors in order to focus, concentrate, and separate better the different proteins. These instruments can detect protein microheterogeneity better than the common CZE instruments. Presented below are some clinical applications of proteins and peptides measurement by CE.

### Serum Proteins

Serum proteins comprise more than a few hundred different proteins. However, for clinical diagnosis, they are separated into 5–12 bands by AG electrophoresis, which is a time-consuming method. They are analyzed routinely on a daily basis in most large hospitals to detect several disorders such as renal failure, infections, and most importantly monoclonal gammopathies.

In CZE, serum proteins can be separated also into 5–12 zones using different buffers such as Tris, borate, and tricine, with a pH of 8–11. Serum protein separation can be completed by CE in  $\sim 2$ –10 min in contrast to 1–2 h for AG (Figure 2). The correlation coefficient between CE and AG for the separated bands is good. Transferrin isoforms, which are important as marker of alcoholism, have been separated by CZE. Some commercial instruments use multicapillaries of narrow diameter (25  $\mu\text{m}$ ) to increase the throughput of the analysis. The narrow capillaries produce better resolution than the wider capillaries with a shorter migration time.

Serum immunoglobulins are composed of heavy and light chains and classified based on their reaction

with specific antibodies into the classes of: IgG, IgA, IgM, IgD, and IgE. Multiple myeloma, Waldenström's disease, and light chain disease can cause increased levels of these proteins (paraproteins). These paraproteins are detected with immunofixation, a laborious procedure performed in gel electrophoresis. CZE has been adapted to perform the immunofixation method based on reacting serum proteins with specific antibodies bound to a solid matrix. The sample is assayed before and after binding. The difference between the two 'immunofixations' represents the specific type of the

monoclonal abnormal serum protein. This method has been shown to be reliable.

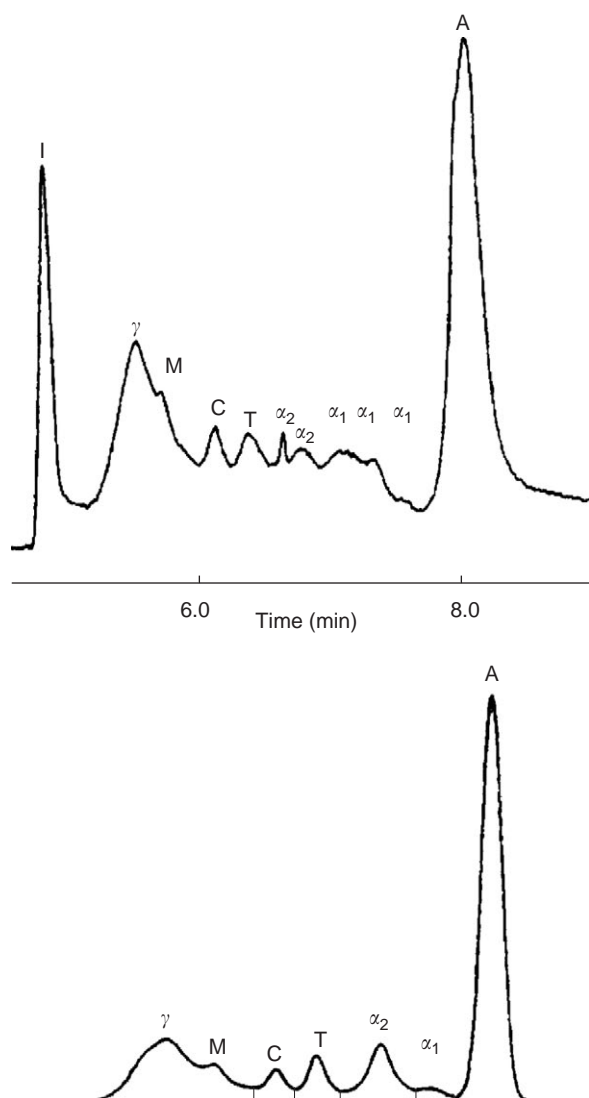
### Cryoglobulins

Cryoglobulins are special immunoglobulins that reversibly precipitate from serum at cold temperatures. Cryoglobulins can be classified as monoclonal globulins (type I), mix of polyclonal-monoclonal (type II), or mix of polyclonal-polyclonal (type III) immunoglobulins. Cryoglobulins can precipitate in different tissues of the body such as the kidney and the extremities. They are associated with several immune-type disorders, viral infection, glomerulonephritis, peripheral neuropathy, and diffuse vasculitis.

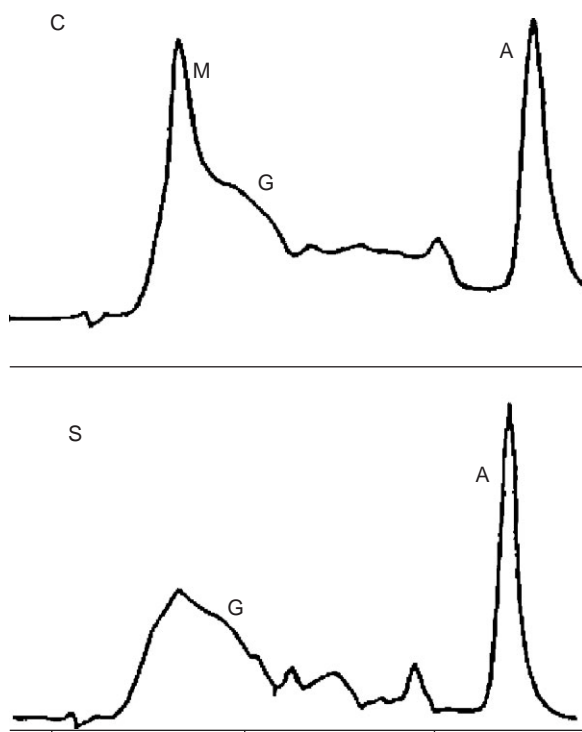
Cryoglobulins are detected by precipitating an aliquot of the serum at 4°C, centrifuging and dissolving the precipitate in a buffer followed by electrophoresis using the same conditions as those for serum proteins (**Figure 3**). Cryoglobulins are well suited for analysis by CE. The main advantages are the higher sensitivity, the use of small volumes of serum, speed, and improved quantification compared to the AG method.

### Urinary Proteins

Usually urine proteins are present at ~10–100 times lower concentrations as compared to serum. In



**Figure 2** Serum electrophoresis of a patient with a small monoclonal band: (top) CE; (bottom) agarose electrophoresis (I = internal standard; A = albumin;  $\alpha_1$  =  $\alpha_1$ -globulin;  $\alpha_2$  =  $\alpha_2$ -globulin; T = transferrin; C = complement; M = monoclonal band;  $\gamma$  =  $\gamma$ -globulins; X = marker). CE conditions: capillary 30 cm  $\times$  50  $\mu$ m (ID), 9 kV, detection at 214 nm. Separation buffer: 7 g boric acid, 7 g sodium carbonate, and 5 g of polyethylene glycol 8000 in 1000 ml water.



**Figure 3** Type II cryoglobulin of a patient by CE: top (C) is cryoprecipitate, and bottom (S) is serum (A = albumin, G = globulin, M = monoclonal peak). CE conditions as in **Figure 2**.

addition to that urine contains numerous interfering UV-absorbing compounds. This renders urinary proteins more difficult to measure by CE when compared to serum. The urine contains several proteins of clinical interest, especially Bence–Jones proteins. Some urine samples can be analyzed directly without any preparation. However, the majority require concentration before the analysis using membrane concentrators. The same buffers and conditions for serum proteins are basically used for analysis of urine protein.

### Cerebrospinal Fluid Proteins

The main clinical significance of cerebrospinal fluid (CSF) protein electrophoresis is for the detection of the oligoclonal bands, which are present in multiple sclerosis in the gamma region. Because proteins in CSF occur at much lower concentrations than in serum (100 times less), a 10–20-fold sample concentration is needed. CSF protein separation can be accomplished in less than 10 min with CE versus 2 h for AG with the ability to detect oligoclonal banding by this technique.

### Enzymes

Since enzymes are essentially proteins they can be measured in CE by direct UV absorbency or by their enzymatic activity. The enzyme, the substrate, or the

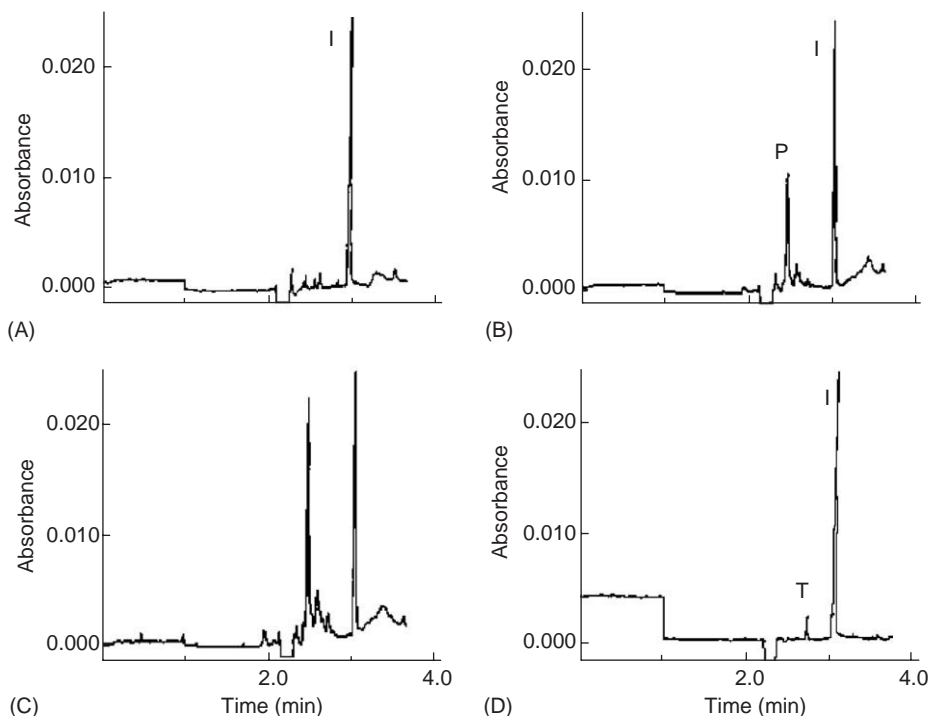
products can all be measured in CE. Catalytic activity is more suited for enzymes with low activity because the reaction product can be amplified easily several fold. On the other hand, enzymes, which are present in high concentration, can be measured directly by their light absorbency.

In CE, catalytic activity can be measured in several ways: (1) incubation in the capillary, (2) online, postcapillary reaction, and (3) incubation outside the capillary. If a long incubation step is needed then it is more convenient to perform the incubation outside the instrument. Proteolytic enzymes with low activity are well suited for analysis by CE in this manner. Several enzymes have been analyzed by CE such as chloramphenicol acetyl transferase, glutathione peroxidase, ornithine transcarbamylase, angiotension converting enzyme, and Cathepsin D (Figure 4).

The advantages of CE for analysis of enzymes are the use of small volumes, versatility, and ability to avoid the extra steps of indicator reactions. In practice, kinetic spectrophotometric methods remain to be most widely used for routine work while CE is reserved for difficult and specialized tests.

### Hemoglobin Variants, Hemoglobin A1c, and Globin Chains

Hemoglobin (Hb) carries the oxygen. Its level is used clinically as indicator of the presence of different



**Figure 4** Enzymatic activity of breast tumor homogenate activity ( $106 \text{ pmol mg}^{-1} \text{ protein}$ ) at different periods of incubation: A=0 min; B=10 min; C=20 min; and D= at 20 min in the presence of pepstatin (P=split peptide, I=iothalamic acid (internal standard); and T=pepstatin). (Reprinted with permission from (1996) *Journal of Chromatography B* 683: 125; © Elsevier.)



types of anemia. Hb is present as two  $\alpha$ -chains and two  $\beta$ -chains. The  $\beta$ -chain is more susceptible for amino acid substitution (mutations), which results in specific variants that occur often in special populations. Some of these variants are harmless; however, others are associated with severe anemia, decreased capacity to carry oxygen, and altered red blood cell shape. The most encountered variants of Hb are A, F, S, and C. Because of the small charge difference of the isoelectric point (pI), Hb variants do not separate well by CZE. For good separation of Hb variants by CZE, a high buffer concentration, a narrow capillary (20–30  $\mu\text{m}$  ID), minimum volume of sample, and low voltages are required. Tris, tricine, and arginine buffers at pH 8–8.4 give a good separation. The separation by CZE resembles very closely that of the alkaline separation by AG.

Although CE instruments are not well designed for CIEF, many variants can be separated better by this technique. In addition to the common variants, HBA<sub>1c</sub> (a good measure for long-term hyperglycemia in diabetes), G Philadelphia, A2, and Bart's can all be separated by CIEF. The separation by CIEF compares well to high-performance liquid chromatography (HPLC) and to gel isoelectric focusing. The variants have also been analyzed by both CE and CIEF equipped with special absorption imaging detectors. These types of detection devices eliminate the extra steps needed to move the peaks, after the focusing step, to the detector and can simultaneously detect several capillaries with better precision and faster results than CE instruments. The globin chains, which are useful for investigating the thalassemias, have been analyzed by CZE in phosphate buffer either at pH 11.8, or at pH 2.5–4.5 after acetone precipitation. Furthermore, the tryptic digests of the globin chains were analyzed by CZE.

### Peptides and Polypeptides

Peptides can arise from the digestion of purified proteins. In this case, they are present in high concentrations and can be analyzed using the same or similar conditions for proteins. There is a great interest in peptide analysis as means to identify those proteins coded by the different genes discovered recently. After hydrolysis of the protein the different peptides are separated and analyzed by mass spectra. Online protein digestion using the bacterial peptidase followed by fluorescence detection, or mass spectra coupled to CE, has been described. On the other hand, peptides can present naturally in different biological fluids such as in serum or spinal fluid. In this case, they usually are present in low concentration among several interfering substances. Thus,

the analysis becomes more difficult. For example, glutathione in blood has been analyzed by CE. Peptide analysis by CE can be used for quality control or purity check in the pharmaceutical industry. In this respect, CE is well suited for this purpose. For example, endorphins, insulin peptides, aprotinin, and substance P all have been analyzed by CE (Figure 5).

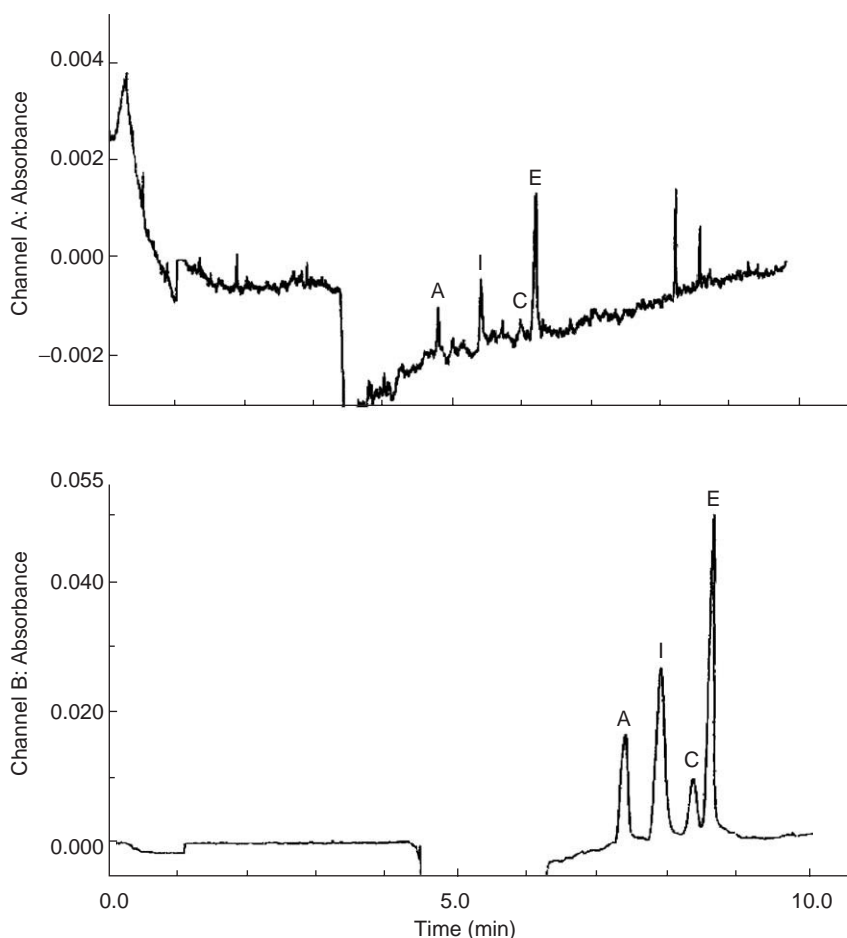
### Amino Acids

Amino acid analysis is used in three distinct areas:

1. Detection of certain inborn errors of metabolism, such as phenylketonuria or maple syrup urine disease, where analysis of a single or very few amino acids present in high concentration is sufficient. This task is a relatively simple one. Several amino acids have been determined for this purpose by CE such as tyrosine, proline, and phenylalanine.
2. Protein hydrolysate, and amino acid mixtures: This analysis is important in determination of the structure of protein or its nutritional value. Here, the analysis of ~20 amino acids is sufficient. This type of analysis is relatively more difficult than that for a single amino acid. Most of the CE work is aimed at this type of analysis.
3. Free amino acids in physiological fluids, where the separation is much more difficult because of the need to separate also the other interfering natural substances (such as small peptides) in addition to the amino acids.

Analysis of amino acids poses several problems. The majority of amino acids lack a strong chromophore. Thus, they need extra steps to be derivatized pre- or postseparation. Since most of the amino acids are very similar in structure they are difficult to separate. Physiological fluids like serum contain many interfering compounds such as peptides and the uncommon amino acids.

Many workers have attempted analysis of amino acids by CE using different pre- and postreactions to enhance their detection. Both free CZE and micellar electrokinetic capillary chromatography (MEKC) have been used with different degrees of success similar to that of amino acids detection by HPLC. The separation of amino acids is better achieved in coated capillaries and also better with MEKC. Derivatization, especially with fluorescent agents, offers much better sensitivity. Several additive agents such as urea, cyclodextrin, and tetrabutyl ammonium salts improve the separation of amino acids. The analysis of amino acids from biological fluids by CE without interference and with good reproducibility remains a challenge. A dedicated CE instrument for amino



**Figure 5** Separation and stacking of some natural peptides (A = angiotensin, I = insulin B chain, C = impurity in the insulin B chain, and E = Leu-enkephalin). Top at 1.5% loading of the capillary and bottom at 30%. (Reprinted with permission from (1996) *Journal of Chromatography A* 744: 231; © Elsevier.)

acids separation is an attractive idea and might someday be commercially available.

## Drug Analysis

Drugs are analyzed clinically for several purposes: metabolism, pharmaceutical, forensic, and therapeutic drug monitoring (TDM). The majority of drugs present in serum can be analyzed successfully by any of several techniques including immunoassay, CE, gas chromatography (GC), and HPLC. Because of the ease, convenience, and automation, immunoassays remain to be the favorite choice in routine laboratories. However, new drugs usually do not have commercial immunoassays available and the real choice for these drugs is between HPLC and CE. These two methods are complementary to each other. However, CE, when compared to HPLC for TDM, is faster and easier with better resolution especially for the polar compounds with less operating cost. The capillary is less expensive than the HPLC column and

can tolerate extreme pH and direct protein injection. The use of large volumes of expensive organic solvents in HPLC is not needed in CE. As the environmental protection and safety rules are getting tighter, with regard to organic solvents disposal, the merits of CE become more evident. Methods development for drugs in general is faster by CE and the sample size is very small in CE.

Most of the analysis for drugs by CE utilizes either CZE in free solutions or MEKC in the presence of micelles. The two methods are complementary to each other. Drugs that are not sufficiently water soluble can be analyzed by nonaqueous CE (NACE). The separation in this method occurs in organic solvents in place of aqueous buffers. NACE offers several advantages for the analysis of such drugs, e.g., better solubility, different selectivity, and the ability to use higher voltages.

Since a main problem in CE is the poor detection limit, several strategies to concentrate the drugs from the biological samples by stacking (sample

concentration on the capillary) and by extraction are employed. Many new extraction methods are emerging as more suitable for CE, such as solid-phase microextraction compared to the traditional method of liquid-liquid extraction.

Several clever ideas have been described to obtain basic information on drugs utilizing CE; e.g., fast  $pK_a$  determination or drug binding to proteins and cell membranes. Commercial CE-mass spectrometry instruments are now available for drug metabolism and for drug confirmation. Most of the drugs have been determined by CE such as pentobarbital, theophylline, iohexol, keppra, and lamotrigine. As an example, analysis of the drug fenofibric acid in serum by CE with stacking is shown in **Figure 6**. The sharp peaks (high plate number) and fast analysis is evident in this figure.

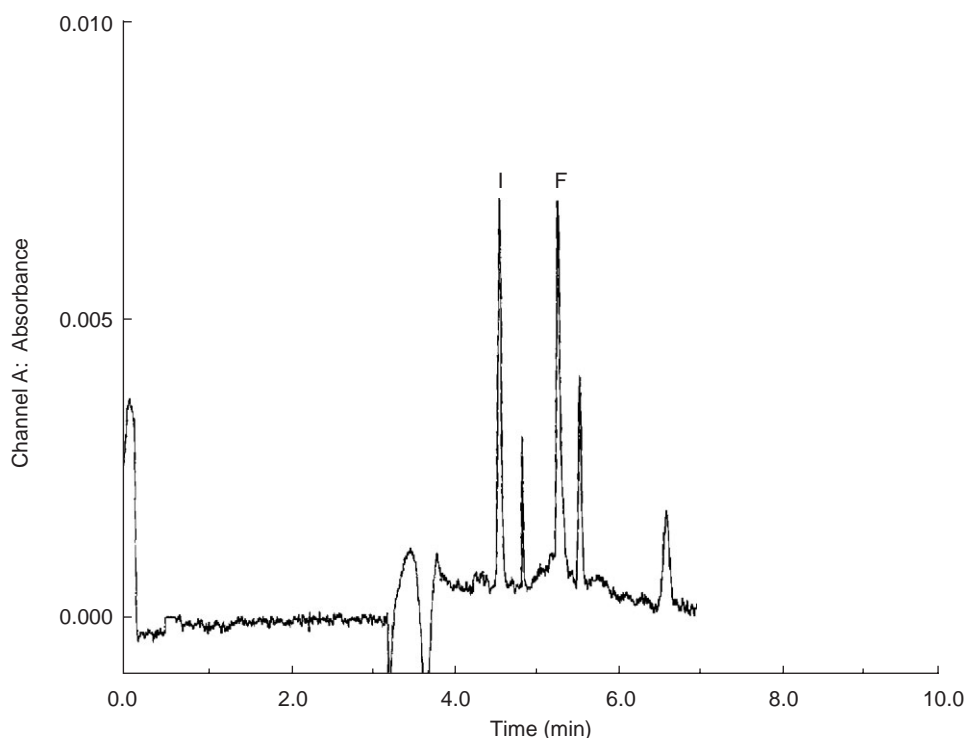
### Chiral Separations

About 40% of the drugs contain at least one chiral center. Isomers have very close chemical structures; however, in many cases they exhibit different biological effects, bind to proteins, or are metabolized differently. Thus, chiral separation is very important. The principles of this technique in CE depend on the addition to the buffer an additive compound with a special cavity such as cyclodextrins, heparin, or certain antibiotics, which

selectively interact with the migration of one isomer. Chiral selectivity results from inclusion of a hydrophobic portion of the solute in the cavity and also from hydrogen bonding to the chiral hydroxyl groups. Because the amount used in CE is very small, the expensive chiral additives are usually added to the buffer and the separation is performed by MEKC. CE and capillary electrochromatography (CEC) are becoming more utilized for this purpose. CEC utilizes columns similar to those of HPLC but the buffer movement performed is driven by voltage rather than by a high-pressure pump. The advantage of CEC over HPLC is the higher plate number. In CEC, the chiral separations can be performed by columns similar to those for HPLC or by adding the chiral selector directly to the mobile phase. Rapid methods for chiral separation based on using the short end of the capillary with highly sulfated cyclodextrins have been described. These stereoselective methods are being applied to the analysis of drugs not only in tablets but also to those in serum samples.

### Drug Screening

Drug screening for forensic and emergency room testing is performed in many labs mainly by immunoassays. Unfortunately, the migration time in CE is not reproducible enough for identification of the numerous



**Figure 6** Analysis of fenofibric acid (F,  $42 \text{ mg l}^{-1}$ ) in serum by CE, with stacking. Iohexol (I) is added as internal standard. Conditions of analysis: capillary of  $25 \text{ cm} \times 50 \mu\text{m}$  (ID), voltage 7 kV, wavelength 280 nm, sample size 13% of the capillary volume. Separation buffer as in **Figure 2**.

drugs. On the other hand, the mobility data are much more reproducible and can be used for this purpose. CE has the potential of being a simple, economical, and powerful method of separation.

CE offers two powerful modes of separation for the forensic drugs, which can be complementary to each other, CZE and MEKC. Early work has shown the promise of CE for drug analysis. Drug screening by CE is a very enormous task, which requires large effort for building up a computerized database for all the controlled substances with their mobility and spectral data.

## Ion Analysis

Ions, organic and inorganic, are difficult to measure by most methods, especially when present at low concentrations. In HPLC, they require special expensive columns. CE is quite suited for analysis of these charged particles. Because of their relative charge to the small mass they tend to migrate rapidly giving fast separation with very high plate numbers and at low cost per test. Both cations and anions can be analyzed in the same run. The separation can be based on simple free solution CE or based on suitable chelating additive. Cations in general are measured in a low pH electrolyte containing a UV active species (imidazole or benzylamine), while the anions are measured after reversing the electroosmotic flow and also after adding a high-mobility UV active species.

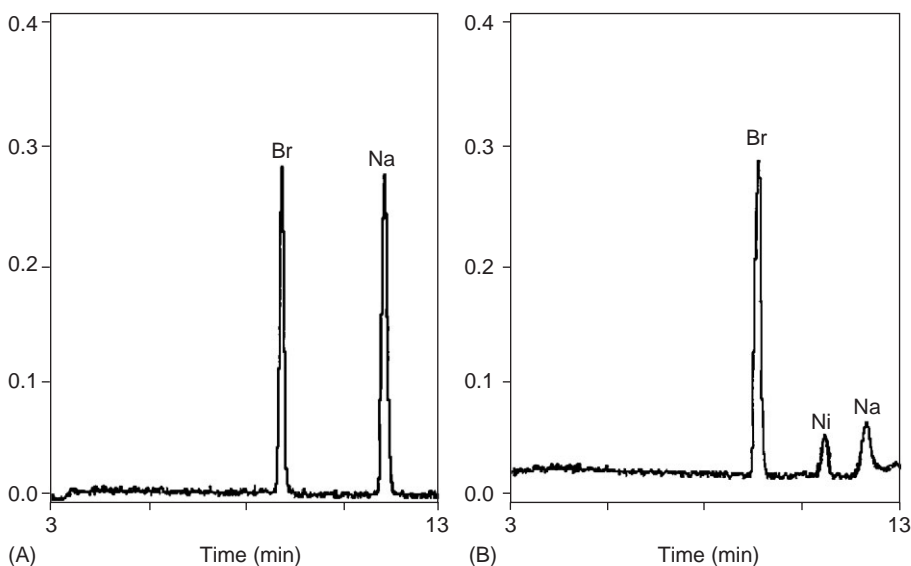
Many of the inorganic and small organic ions in the biological fluids have clinical and diagnostic importance. Inorganic ions in the serum are important for maintaining the osmotic pressure and the pH of

the cells constant to prevent muscle, renal, and heart malfunctions. These are measured routinely in clinical labs and they represent a large volume of the pathology lab work. In practice, most of the common inorganic ions such as Na, K, Ca, can be measured more conveniently by ion-selective electrodes in the clinical labs. However, some of the uncommon ions such as nitrite and nitrate can be measured better with CE (Figure 7).

On the other hand, organic acids are more difficult to measure in the labs. Usually, these are measured by GC or HPLC. However, CE offers speed, precision, and specificity over other methods. Organic acids are important in inborn errors of metabolism, in infection, and different metabolic disorders. Many of these compounds have been measured by CE directly, or by indirect UV absorbency after addition of a UV absorbing compound such as benzoate, naphthalene sulfonate, imidazole, or benzylamine. For example, oxalate and citrate, which are important in stone formation, have been measured after urine dilution by both direct and indirect detections. Lactate, pyruvate, ascorbate, and oxalate were measured by CE in the CSF of patients in ~10 min. Methylmalonic acid, which is a sensitive measure of vitamin B<sub>12</sub> deficiency preceding any clinical symptoms or changes in the serum, has been determined in urine by CE after sample extraction and concentration.

## Miscellaneous Tests

There are numerous clinical tests that are less frequently analyzed by CE. Nevertheless, they remain to



**Figure 7** CE analysis of nitrite and nitrate in: (A) urine sample, (B) CSF sample. (Reprinted with permission from (1997) *Journal of Chromatography A* 781: 491; © Elsevier.)

be clinically important. Some examples are given below.

### Carbohydrates and Glycoproteins

Simple carbohydrates are important in diabetes and in inborn errors of metabolism. Carbohydrates bound to proteins (glycoproteins) are important for cell recognition, receptor interaction, and in immunity. Many of these compounds are also of pharmaceutical interest. The separations of transferrin isoforms, which are different in their sialic acid content, were among the first glycoproteins to be separated by CE. Other compounds such as ribonuclease, ovalbumin, tissue plasminogen, and human chorionic gonadotropin isoforms were separated later by CE. Borate buffers as well as the presence of different additives that affect the surface charge of the capillary (e.g., alkylammonium salts, diaminobutane, formic acid) are important for the separation.

Carbohydrates analysis by CE poses several problems. These compounds have poor UV absorbency. Their indirect detection depends on using buffers with high chromophoric ions such as sorbate, sulfosalicylic acid, tricarboxylbenzoic acid, and tryptophane. Simple sugars are not ionized unless the pH is above 12. They can be separated by CZE using buffers of NaOH or LiOH with added riboflavin for indirect UV detection. The use of small capillaries and efficient heat dissipation is necessary for a low baseline noise. Carbohydrates can also be derivatized with different fluorescent reagents or complexed with borate ions. As the chain length of the carbohydrate increases the analysis becomes more difficult.

### Lipids and Lipoproteins

Lipids and lipoproteins are important in atherosclerosis and coronary vascular disease. Some of the lipoproteins are atherogenic such as low-density lipoprotein (LDL) while others such as high-density lipoprotein (HDL) are protective. The different fractions of lipoproteins (e.g., LDL, HDL, Lp(a)) have been separated by MEKC. However, the separation by CE remains difficult and thus is not common.

### Immunoassays

In CE, analysis by immunoassays depends on the principles that the antigen and antibody migrate differently when they are bound compared to when they are free. One of these two compounds (mostly the antigen) is labeled with a fluorescent tag. The unknown sample is mixed with labeled antigen for competitive binding assay and the mixture is separated by CE. The label in the bound fraction (or

the ratio of bound to free) can be calculated and the unknown measured from a standard curve. These methods have been applied for the analysis of insulin, cortisol, and a few drugs. Unfortunately, this approach is not as simple or convenient as those commercially automated immunoassay instruments dedicated for this purpose.

### Single Cell Analysis

In many instances, such as in cancer, it is important to analyze a single or very few cells rather than homogenizing the whole tissue. Small capillaries of  $\sim 10\mu\text{m}$  in diameter have been used to study the level of some metabolites in single cells, e.g., catecholamines, 5-hydroxyindole acetic acid including the level of the enzyme lactate dehydrogenase. However, cell introduction into the capillary and lysis is not an easy task. Furthermore, this analysis requires the use of very sensitive detection.

### Stacking

Most of the clinical tests are below the direct detection of the CE instrument. In order to improve the detection limits, sample concentration either through extraction or on the capillary is utilized. Ionized compounds, in general, migrate more rapidly in dilute compared to concentrated buffers. In practice, stacking is a very simple technique to achieve. It depends on injecting a large sample volume (5–30% of the capillary volume) and performing the electrophoresis step under discontinuous buffer conditions: different pH, different field strength, or different buffers in such a way that the two edges of the sample migrate at different rates leading to narrowing of the sample plug leading to sample concentration (Figure 5). This step is done often simultaneously with the separation step.

**See also:** Capillary Electrophoresis: Overview. Electrophoresis: Overview; Principles; Isoelectric Focusing; Polyacrylamide Gels; Affinity Techniques; Clinical Applications; Nucleic Acids; Proteins.

### Further Reading

- Altria KD (1996) *Capillary Electrophoresis Guidebook: Principles, Operation and Applications*. Totowa, NJ: Humana Press.
- Altria KD (1999) Overview of capillary electrophoresis and capillary electrochromatography. *Journal Chromatography A* 856: 443–463.
- Bocsek P, Vespalec R, and Geise RW (2000) Selectivity in CE. *Analytical Chemistry* 72: 587A–695A.
- Dovich NJ (1997) Capillary gel electrophoresis for large DNA sequencing. In: Landers JP (ed.) *Handbook of*

- Electrophoresis*, pp. 545–566. Boca Raton, FL: CRC Press.
- Guttman A and Ulfeder KJ (1998) Separation of DNA by capillary electrophoresis. *Advances in Chromatography* 38: 301–340.
- Jolliff CR (2001) Serum and urine paraprotein capillary electrophoresis. In: Peterson JR and Mohammad AA (eds.) *Clinical and Forensics Applications of Capillary Electrophoresis*, pp. 93–104. Totowa, NJ: Humana Press.
- Jones WR (1997) Capillary ion chromatography. In: Landers JP (ed.) *Handbook of Electrophoresis*, pp. 155–188. Boca Raton, FL: CRC Press.
- Landers JP (1997) *Handbook of Electrophoresis*, 2nd edn. Boca Raton: CRC Press.
- Li SFY (1992) *Capillary Electrophoresis. Journal Chromatography Library*, 52. Amsterdam: Elsevier.
- Michelson KR and Cheng J (2001) *Capillary Electrophoresis of Nucleic Acids*. Methods in Molecular Biology, vol. 162. Totowa, NJ: Humana Press.
- Oda RP, Roche MA, Landers JP, and Shihabi ZK (1997) Capillary electrophoresis for the analysis of drugs in biological fluids. In: Landers JP (ed.) *Handbook of Electrophoresis*, pp. 457–566. Boca Raton, FL: CRC Press.
- Palfrey S (1999) *Clinical Applications of Capillary Electrophoresis*. Totowa, NJ: Humana Press.
- Peterson JR and Mohammad AA (2001) *Clinical and Forensics Applications of Capillary Electrophoresis*. Totowa, NJ: Humana Press.
- Shihabi ZK (2003) Separation science in therapeutic drug monitoring. In: Abuol-Enien HY (ed.) *Separation Techniques in Clinical Chemistry*, pp. 103–138. New York: Dekker.
- Shintani H and Polonsky J (1997) *Handbook of Capillary Electrophoresis Applications*. London: Chapman and Hall.

# CARBOHYDRATES

## Contents

### Overview

#### Sugars – Spectrophotometric Methods

#### Sugars – Chromatographic Methods

#### Sugar Alcohols

#### Starch

#### Dietary Fiber Measured as Nonstarch Polysaccharides in Plant Foods

## Overview

**A M Stephen**, Fish Hoek, Cape Town, South Africa  
**E H Merrifield**, Zymogenetics, East Seattle, WA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Carbohydrates abound in Nature and are implicated in every form of living activity including commercial manipulation. The underlying reason for their analysis in matrices or for assessment of purity is to establish their identity and functions, and to control and utilize responsibly the resources they supply. Methods of carbohydrate analysis span a century, many of the early colorimetric, titrimetric, and gravimetric methods remaining indispensable, but being continually adapted to rapid and highly sensitive, accurate, automated procedures. A major breakthrough came 50 years ago with the development of chromatography, and of electrophoresis, as tools

for the separation of components in mixtures and subsequent application of a wide range of spectroscopic and electrochemical methods of detection, and of molecular structure determination. Modern analytical trends lie in the direction of ultramicroassay, based upon flow-injection analysis with voltammetric/amperometric detection, the automation of high-performance liquid chromatography (HPLC) in its various forms using fluorimetry and light scattering, ion chromatography with pulsed amperometric detection, and the use of biosensors with immobilized enzymes built into specific electrodes. Classical methods of planar and column chromatography based on adsorption and partition have largely given way to rapid, high-performance methodology including high-performance thin-layer chromatography (HPTLC), and HPLC on robust columns of graphitized carbon, or of modified silica for reversed-phase and hydrophobic interaction, or hydrophilic interaction chromatography, ion-exchange resins, immobilized lectins, and gels or their silica counterparts for noninteractive, size-exclusion chromatography (SEC). Capillary electrophoresis (CE) in different forms, using a

- Electrophoresis*, pp. 545–566. Boca Raton, FL: CRC Press.
- Guttman A and Ulfeder KJ (1998) Separation of DNA by capillary electrophoresis. *Advances in Chromatography* 38: 301–340.
- Jolliff CR (2001) Serum and urine paraprotein capillary electrophoresis. In: Peterson JR and Mohammad AA (eds.) *Clinical and Forensics Applications of Capillary Electrophoresis*, pp. 93–104. Totowa, NJ: Humana Press.
- Jones WR (1997) Capillary ion chromatography. In: Landers JP (ed.) *Handbook of Electrophoresis*, pp. 155–188. Boca Raton, FL: CRC Press.
- Landers JP (1997) *Handbook of Electrophoresis*, 2nd edn. Boca Raton: CRC Press.
- Li SFY (1992) *Capillary Electrophoresis. Journal Chromatography Library*, 52. Amsterdam: Elsevier.
- Michelson KR and Cheng J (2001) *Capillary Electrophoresis of Nucleic Acids*. Methods in Molecular Biology, vol. 162. Totowa, NJ: Humana Press.
- Oda RP, Roche MA, Landers JP, and Shihabi ZK (1997) Capillary electrophoresis for the analysis of drugs in biological fluids. In: Landers JP (ed.) *Handbook of Electrophoresis*, pp. 457–566. Boca Raton, FL: CRC Press.
- Palfrey S (1999) *Clinical Applications of Capillary Electrophoresis*. Totowa, NJ: Humana Press.
- Peterson JR and Mohammad AA (2001) *Clinical and Forensics Applications of Capillary Electrophoresis*. Totowa, NJ: Humana Press.
- Shihabi ZK (2003) Separation science in therapeutic drug monitoring. In: Abuol-Enien HY (ed.) *Separation Techniques in Clinical Chemistry*, pp. 103–138. New York: Dekker.
- Shintani H and Polonsky J (1997) *Handbook of Capillary Electrophoresis Applications*. London: Chapman and Hall.

# CARBOHYDRATES

## Contents

### Overview

#### Sugars – Spectrophotometric Methods

#### Sugars – Chromatographic Methods

#### Sugar Alcohols

#### Starch

#### Dietary Fiber Measured as Nonstarch Polysaccharides in Plant Foods

## Overview

**A M Stephen**, Fish Hoek, Cape Town, South Africa  
**E H Merrifield**, Zymogenetics, East Seattle, WA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Carbohydrates abound in Nature and are implicated in every form of living activity including commercial manipulation. The underlying reason for their analysis in matrices or for assessment of purity is to establish their identity and functions, and to control and utilize responsibly the resources they supply. Methods of carbohydrate analysis span a century, many of the early colorimetric, titrimetric, and gravimetric methods remaining indispensable, but being continually adapted to rapid and highly sensitive, accurate, automated procedures. A major breakthrough came 50 years ago with the development of chromatography, and of electrophoresis, as tools

for the separation of components in mixtures and subsequent application of a wide range of spectroscopic and electrochemical methods of detection, and of molecular structure determination. Modern analytical trends lie in the direction of ultramicroassay, based upon flow-injection analysis with voltammetric/amperometric detection, the automation of high-performance liquid chromatography (HPLC) in its various forms using fluorimetry and light scattering, ion chromatography with pulsed amperometric detection, and the use of biosensors with immobilized enzymes built into specific electrodes. Classical methods of planar and column chromatography based on adsorption and partition have largely given way to rapid, high-performance methodology including high-performance thin-layer chromatography (HPTLC), and HPLC on robust columns of graphitized carbon, or of modified silica for reversed-phase and hydrophobic interaction, or hydrophilic interaction chromatography, ion-exchange resins, immobilized lectins, and gels or their silica counterparts for noninteractive, size-exclusion chromatography (SEC). Capillary electrophoresis (CE) in different forms, using a



variety of sensitive detection methods, has undergone equally dramatic advances. These and other techniques are described in the sections that follow.

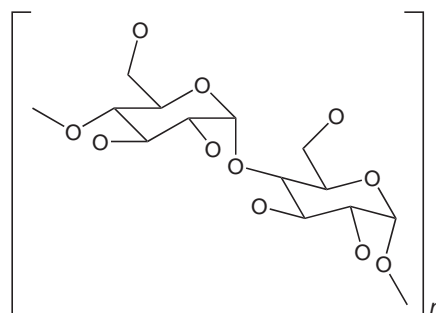
Impinging on the nucleic acid field, the technology of genetic manipulation has developed dramatically, restriction endonucleases being employed on a vast scale to characterize genes and the enzymes they encode, with a view to modifying products, such as starch and other polysaccharide components from plants, that could exhibit better functional uses. The medical field is set to benefit likewise through the targeting at molecular level the sites of infection or aberrant behavior, nuclear magnetic resonance (NMR) and X-ray technology being adapted to locating carbohydrate and other types of molecule at the sites of interest. As organic compounds carbohydrates and their derivatives are routinely analyzed for elemental composition, spectroscopic properties (infrared (IR), ultraviolet (UV), NMR, mass spectrometry (MS), ORD, and circular dichroism (CD)), and molecular mass (distribution) according to standard protocols, as an essential means of establishing identity and purity. These data are required particularly in the course of synthetic work. Selected information is needed for labeling products for marketing. In recent decades, dramatic advances have resulted from the use of computerized instrumentation and data handling.

## Structures

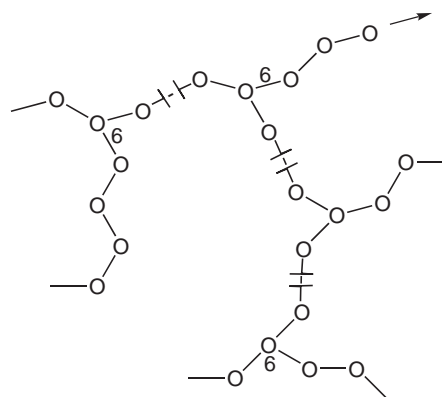
Carbohydrates in general are classified as (1) monosaccharides, which are polyhydroxy aldehydes (aldoses) or ketones (ketoses) with a continuous linear skeleton of carbon atoms (usually five or six), and (2) di-, oligo-, and polysaccharides, which comprise two or many more monosaccharide units joined as O-glycosides. Carbohydrates may be bonded covalently to amino acid constituents of proteins (as glycoproteins), to lipid (as glycolipids of varying complexity), and to lignin, tannin, flavonoids, terpenoids, or steroids by covalent bonds or as ill-defined association complexes. In all carbohydrate structures hydrogen bonding plays a highly prominent role, mediated in many instances by water molecules. Microorganisms, plants, and higher animals contain carbohydrate in the form of metabolites of short life, or oligo- and polysaccharides for energy storage, and as structural (cell wall) material.

The most abundant carbohydrate is D-glucose, which occurs usually in glycosidically bound form, though it is metabolized as phosphate ester and is found as the free 'sugar' in honey, plant nectars, and fruit juices. The predominant storage carbohydrate in plants is starch, a complex mixture of linear

((1); amylose) and branched (2); amylopectin)  $\alpha$ -D-glucans of which (2) usually predominates. It should be noted that isolates of (1), except in unusual circumstances, are accompanied by some branched material; the biosynthesis of starch is a complicated but well-understood process during which the linear molecules are progressively changed under the action of branching and debranching enzymes to form granular (2). The animal counterpart is glycogen, which resembles (2). Cellulose (3), a  $\beta$ 1 $\rightarrow$ 4-linked glucan, contributes hugely to biomass. The nitrogen-containing sugar derivative N-acetyl-D-glucosamine (2-acetamido-2-deoxy-D-glucose), units of which are linked  $\beta$ 1 $\rightarrow$ 4 as in cellulose, constitutes chitin (4), the shell coating of myriads of insects and crustaceans (lobsters) and a component of fungi, and is a vital component of glycoconjugates. The amphoteric compound muramic acid (3-D-lactyl ether of D-glucosamine) is ubiquitous as the N-Ac derivative in bacterial cell walls.

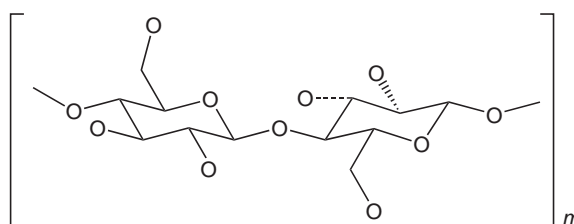


(1)

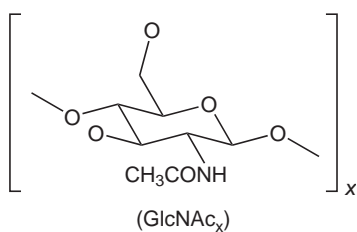


O=1,4-Linked  $\alpha$ -D-Glc

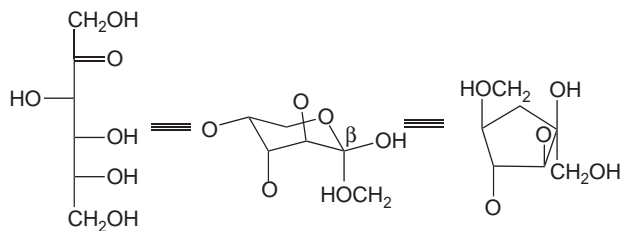
(2)



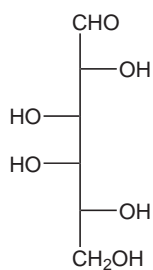
(3)

(GlcNAc<sub>x</sub>)

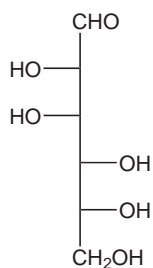
(4)



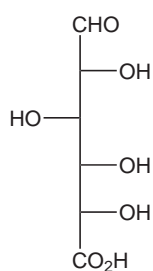
(5)



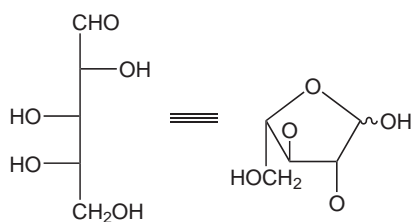
(6)



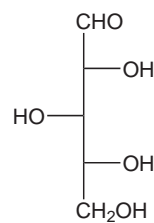
(7)



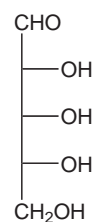
(8)



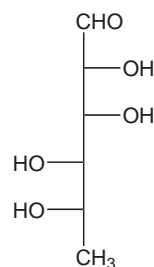
(9)



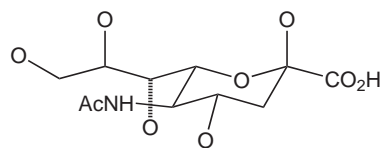
(10)



(11)

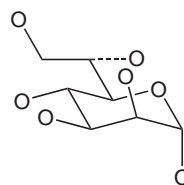


(12)



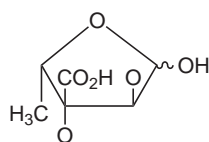
5-Acetamido-3,5-dideoxy-D-glycero-D-galacto-nonulosonic acid  
N-acetylneuraminic acid  
(NeuAc)

(13)

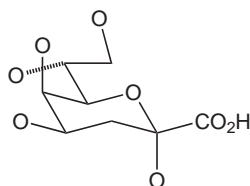


L-Glycero-D-manno-heptose (LD-Hep)

(14)



(15)



3-Deoxy-D-manno-octulosonic acid (Kdo)

(16)

Isomers of glucose in approximate order of natural abundance are the ketosugar D-fructose (5), which when combined with glucose forms common sugar (sucrose), and the stereoisomers of glucose D-galactose (6) and D-mannose (7). The monosaccharides (all  $C_6H_{12}O_6$ ) are accompanied in Nature by the C6-oxidized derivatives D-glucuronic (8), D-galacturonic (often as Me ester), and D-mannuronic acids. Five-carbon sugars L-arabinose (9; compare the stereochemistry of D-galactose) and D-xylose (10) (related to D-glucose) are abundant in plant tissues (pectins and hemicelluloses), while D-ribose (11) and 2-deoxy-D-ribose are constituents of RNA and DNA, respectively. The 6-deoxy sugars L-rhamnose (12; stereochemically related to L-mannose) and L-fucose (related to L-galactose) are components of pectins, xyloglucans, mucilages, and glycoconjugates; critically important in its biological activity is the mannose derivative *N*-acetylneuraminic acid, one of an extensive family of sialic acids (13). These examples by no means exhaust the variations in structure found naturally. Rare but important sugar derivatives have seven-carbon chains (14). Antibiotics and steroidal glycosides contain sugar units modified by several changes of functional group, and a key unit in all pectins is the chain-branched aceric acid (15). Well over a hundred modified sugars (including many Me ethers, and the ketoacid Kdo (16)) are known to be constituent units in bacterial polysaccharides.

Many polysaccharide types are highly complex in structure, containing perhaps six or seven different sugar units in the molecule. Each of these units may be linked in several ways. The number of combinations actually found is, however, remarkably small considering the many possibilities that exist for the linkage of any two sugar units to each other, though variations in fine structure (microheterogeneity) are the norm in plant products. The capsular polysaccharides of bacteria, however, contain repeating units consisting of several sugars in well-defined sequences. Any

experimental work aimed at determining the molecular structure of a polysaccharide commences with a decision, often difficult, as to when the isolation procedure has reached a stage at which the polysaccharide sample might be regarded as sufficiently homogeneous to warrant further examination. Modern spectroscopic methods sometimes permit structural features to be observed in unmodified natural specimens. MS affords an accurate measure of molecular weights in solid samples: polysaccharides in solution usually show complicated molecular weight distribution patterns (in the range  $10^4$ – $10^7$ ).

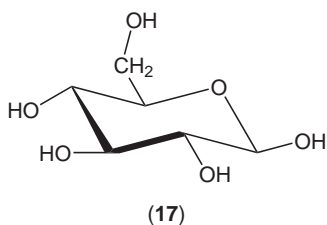
The type of analysis to be performed varies according to the nature of the carbohydrate sample. If a mixture of monosaccharides is presented, chromatographic and spectroscopic procedures of many types are available to identify and quantify the sugars. Optical rotation measurements are invaluable in defining the purity of carbohydrate specimens, and in analyzing the composition of mixtures if the identity of the components is known. Differentiation between D and L enantiomers is best achieved on a small scale by conversion of derivatized sugars to glycosides using a chiral reagent, as the diastereoisomers formed have different thermodynamic properties and hence chromatographic retention times and NMR spectra. Alternatively, the sugars might be chromatographed on a chiral stationary phase. Polysaccharides present analytical problems at different levels according to their complexity; if the molecular structure is known, quantification is usually straightforward. The determination of the modes of linkage and sequences of the constituent monosaccharide units in complex polysaccharides, however, is a lengthy procedure requiring the application of numerous chemical, enzymatic, and spectroscopic techniques. Identification of the conformation, degree of crystallinity, and organization of the polysaccharide in its biochemical and physiological environment, or in foodstuffs, represents a further extension of the analytical problem.

## Main Reactions and Methodology

### Carbohydrate Reactions

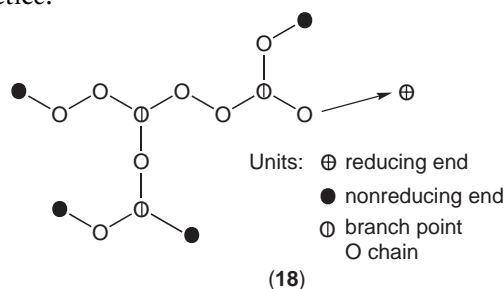
The key functional group in carbohydrates is the carbonyl ( $>C=O$ ) group, which in sugars is masked by internal addition of an OH group in the same molecule to form five- or six-membered hemiacetals (named furanoses and pyranoses, respectively). Diastereoisomeric forms of each sugar, known as anomers, designated  $\alpha$  and  $\beta$ , are formed upon cyclization; this with the family configuration, determined as D or L by the highest-numbered chiral carbon atom,

denotes a particular molecular form of the sugar. Thus, for common D-glucose there are four forms ( $\alpha$  and  $\beta$ , furanose and pyranose). In solution these molecules equilibrate through the intermediary aldehyde, and for thermodynamic reasons  $\beta$ -D-glucopyranose (17) predominates. This form adopts a chair conformation, in which all the substituent groups on ring carbons are equatorial. The exceptional stability of this fundamental structure is manifested in the linear molecule of cellulose, the most abundant of all carbohydrates, in which the regular repeating unit is the dimer (cf. (3)). In cellulose the chains of D-glucose units ( $\sim 500$ – $10\,000$  in number) are united by extensive hydrogen bonding, and approximately one-half of the isolated material shows crystallinity. Starch in the form of amylose (essentially a linear  $(1 \rightarrow 4)$ - $\alpha$ -D-glucan) and amylopectin (having branches at O-6) presents many subtle variations in molecular structure and morphology (seen increasingly in genetically modified materials), and has properties and functions vastly different from those of cellulose, to which it is second in natural abundance.



The inter-sugar linkage in all di-, oligo-, and polysaccharides is the glycosidic bond, which is stable in water except when the carbohydrate is heated under acidic conditions or incubated with specific, hydrolytic enzymes (hydrolases). Mechanisms of glycosidic fission differ from those of the spontaneous ring opening and closing of the hemiacetal (ring) forms of monosaccharides. Consequently, the chemistry of the potential aldehydic groups in sugars (glycoses) is markedly different from that of the acetal linkages present in glycosides (see below). Every combination of sugars in carbohydrates comprising two or more units is characterized by one reducing end, nonreducing end groups in proportion to the number ( $n$ ) of chains, a corresponding number ( $n-1$ ) of branch points, and, in the case of polysaccharides, large numbers of chain units (18); these categories are subtly different in their chemical and biological properties. Fructans are essentially linear polymers of  $\beta$ -D-fructose, linked  $(2 \rightarrow 1)$  and  $(2 \rightarrow 6)$ ; the former, inulins, are unique because none of the ring atoms save the anomeric C-2 are included in the glycosidic, polymeric chains. As their name implies, the ends of chain units in cyclodextrins are joined, forming cages that are of

great importance in analytical separations based on stereochemical differences, and in pharmaceutical practice.

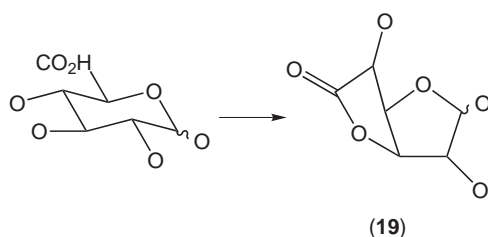


The carbonyl group in most carbohydrates is therefore masked (1) by hemiacetal formation in free sugars and in reducing-end sugar units, or (2) additionally by glycoside formation. The practical difference between (1) and (2) is that in solution the carbonyl function of sugars reacts as aldehyde toward oxidation (to yield carboxylic acid), and as aldehyde or ketone toward reducing agents (yielding alditols) and nucleophiles (forming derivatives). Complicated rearrangements follow enolization of sugars in aqueous alkaline solutions, so that they are intrinsically unstable at a raised pH, and lactone formation usually results after oxidation to the carboxylic acid. Reduction stabilizes the sugar molecule in these respects, yielding from an aldose a single product, and is accordingly widely exploited in analytical procedures. Only after hydrolysis (solvolysis) are sugars, released from being bound as glycosides, able to react as hemiacetals/aldehydes. The glycosidic bond being stable toward hydrolysis under neutral and alkaline conditions, the anomeric C atom in glycosides, disaccharides, oligosaccharides, and polysaccharides is not readily oxidized or reduced and is inert toward nucleophiles.

The hydroxyl groups attached to most carbon atoms in sugars and glycosides behave as do those in polyols, undergoing normal processes of dehydration, etherification, esterification, displacement, oxidation, and, in pairs, complex formation. They act as nucleophiles, displacing suitably activated anomeric hydroxyls of other sugar molecules to form glycosidic bonds. Hydroxyl groups are polar and hydrophilic, and form hydrogen bonds. In the special case of the primary alcohol functionality, oxidation proceeds to the carboxyl level, yielding, for example, the important D-glucuronic acid (8) and its lactone (19); Scheme 1.

### Analytical Aspects

If not soluble in water, carbohydrates are difficult to disperse otherwise. Anhydrous dimethyl sulfoxide, formamide, 1-methylmorpholine-1-oxide, 1-methyl-2-pyrrolidinone, and other aprotic solvents can sometimes be used, and sonication may be employed



Scheme 1

at risk of causing depolymerization of macromolecules; this procedure is a favored one for homogenizing plant tissues or foodstuff preparations prior to analysis. Enzymes may be needed to release polysaccharide components from their association with other substances in cellular tissue in order to bring them into solution in water, or to sever the glycan bond to peptide in a glycoconjugate. In view of the universality of HPLC methods of analysis it is advantageous to extract not with ethanol–water but with acetonitrile–water mixtures, which can be injected directly into the analytical system.

Analytical examination of solid samples includes observing effects of heat (melting point measurement, differential thermal analysis, and scanning calorimetry), and solid-state spectroscopy (IR, NMR, MS, X-ray). Microscopic observations play a part in identifying raw starch and its gelatinization. Investigation of aqueous solutions involves the measurement of chirality (optical activity for the sodium D-line; or ORD), viscosity and gel strength for polymers, and  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopic parameters; the power and applicability of NMR and MS in particular has made an enormous impact in recent years. Carbohydrate extracts are invariably submitted to one or more of a wide range of chromatographic separations, which provide a means of characterizing sugars (or oligosaccharides) and their chemical derivatives, and of establishing the degree of homogeneity or purity of all carbohydrates, including polysaccharides and glycoconjugates.

Carbohydrates char on being heated and melting points (except those of derivatives) are not always reliable. Qualitative identification is performed by applying the time-honored purple-ring Molisch test (mixing 1-naphthol in aqueous ethanol with carbohydrate in water, and pouring concentrated sulfuric acid cautiously down the side of the tube containing the mixture). The test, of which there are many variants, is highly sensitive. Other phenols (orcinol, *o*-cresol, resorcinol), aromatic amines (aniline, *p*-anisidine, and carbazole), or anthrone produce colorations that distinguish one class of sugar from another, and uronic acids from neutral aldoses and ketoses; a

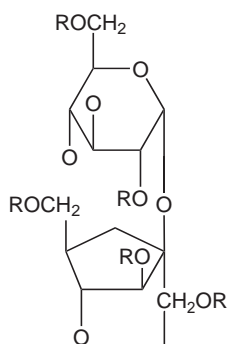
strong mineral acid is used to cause dehydration of the sugar, and the resulting furfural derivative combines with a phenol or amine to generate the chromophore. The standard Smith assay quantifies the reaction of phenol and concentrated sulfuric acid with carbohydrates generally, the optical absorption at 490 nm being compared with that of suitable calibrants. Carbohydrates separated on planar chromatograms are revealed when sprayed with such reagents. Derivatization of carbohydrates takes many forms, rendering them volatile (for GC), or changing polarity and providing easily detectable functional groups (for HPLC).

Structural analysis of oligo- and polysaccharides is most frequently accomplished by the methylation and hydrolysis procedure, modified if need be by reductive cleavage. Hydroxyl groups are converted quantitatively to methyl ethers of component sugar units, from which, after release and analysis, modes of linkage are inferred. Paper chromatography and TLC are simple and valuable tools for the separation of mixtures of methylated sugars; the extent of methylation and the positions of O-methyl groups are deduced from the mobility and color response of these derivatives to reagents such as *p*-anisidine hydrochloride. Methylated pentoses, hexoses, deoxyhexoses, and uronic acids exhibit different ranges of color. GC coupled with MS of suitable derivatives (methyl glycosides, acetylated aldononitriles, and alditols) is, however, the standard quantitative analytical procedure. Difficulties arise when uronic acid or acetamido sugar units are present in the carbohydrate; the glycosidic bonds resist acid hydrolysis and judicious derivatization is required to produce volatile substances for GC–MS identification. The contrary applies to polysaccharides containing acid-labile units such as D-fructofuranose (in inulins) or 3,6-anhydrogalactose (in agars and carrageenans) where such devices as protection by reduction with borohydride must be employed. Methanolysis may be an alternative to acid hydrolysis, and reductive cleavage is useful, but it is in the field of glycoconjugate analysis that this procedure becomes obligatory. The solubility of the analyte may favor methanolysis, particularly in working with glycolipids; and where such sensitive units as the sialic acids or certain sugar components of bacterial cell walls are present, methanolysis is essential; trimethylsilyl ether derivatives are most suitable for subsequent GC analysis. Separations of the mixtures of monosaccharides released on hydrolysis, if liquid chromatography (LC) is employed, do not necessarily require precolumn derivatization. Enzymes are used for many purposes, e.g., to estimate D-glucose or D-galactose, to establish anomeric linkages, and to effect hydrolysis or other



chemical change. Periodate oxidation of carbohydrates and analysis of the products is a technique applied extensively, and reaction procedures (degradations catalyzed by base) have been designed to localize and identify sugar units in proximity to uronic acid residues. Of prime importance is the application of NMR techniques to define the environments of H and C atoms and other bioelements in sugars or even in mixtures; intact oligosaccharides of great complexity may now be characterized spectroscopically. MS is informative at all levels of molecular size; however, the dramatic advances made in X-ray crystallography have afforded the most precise information as to conformational detail, with respect to helical structures in linear polysaccharides for example.

Sucrose (20) is a bulk chemical of exceptional purity, easily extracted and refined using relatively few steps. Some other sugars are also obtainable in crystalline form, but are generally hygroscopic. Polysaccharides retain moisture (5–10%) when air dried, the exact percentage being difficult to determine by thermogravimetry using vacuum-oven heating (microwave technology is feasible), or by Karl Fischer titration, and apart from cotton cellulose, some starches, bacterial exopolysaccharides, and plant exudates, very few can be isolated in a pure state without extensive chemical and physical manipulations. Many polysaccharides even defy exact definition, amyloids, pectins, hemicelluloses, and dietary fiber among them.



Family (R = e.g.,  $\alpha$ -D-Galp) of  
sucrose (R = H) derivatives

(20)

## Occurrence

Carbohydrates occur in all forms of life, and survive in plant and animal tissues for varying periods thereafter. Cellulose and the hemicellulosic components of dead wood decompose eventually under the action of enzymes secreted by fungi, whereas soluble carbohydrates disappear rapidly from plants and animals following well-established metabolic pathways. Industrially important carbohydrates, available in

bulk, include cellulose and mixtures of substituted xylans and glucomannans, which occur associated with the polyphenolic substance lignin in hardwoods and softwoods. In addition to the major nutritive components of cereal grains and vegetables (starches,  $\beta$ -(1 $\rightarrow$ 3) (1 $\rightarrow$ 4)-D-glucans and galactomannans), soluble and insoluble dietary fibers (xylans, pectins, and xyloglucans) are recognized as fulfilling an essential role. Natural sweeteners (sucrose, glucose, and fructose) are obtained from cane or beet, or from hydrolyzed corn (glucose, maltose, and related oligosaccharides). In food processing a variety of polysaccharides are employed to produce thickening and gelling; many such are of bacterial origin (xanthan, gellan, and scleroglucan), others are extracted from seaweeds (agars, carrageenans, and alginates), and some are obtained as exudates from trees and shrubs (gum arabic, gum tragacanth).

## Uses: In Foods, Heavy Industry, Biology, and Medicine

### Foods

The dietetic requirement for carbohydrate in humans and animals is well documented and has been widely discussed. A prime consideration is the ingestion of carbohydrate as an energy source, which is a staple need for all, especially in time of famine, and required in particular by manual workers and sportsmen among humans, and for the purposes of fattening livestock and caring for pet animals. Starch and commercial sugar are by far the most abundant carbohydrates consumed by humans (cellulose by ruminants) but for health reasons polysaccharides, which are not easily metabolized (dietary fiber), assume vital roles. Taste, texture, and palatability (mouth feel) are prime considerations in the supply of acceptable and adequate foodstuffs, and a range of different types of polysaccharide (hydrocolloids) are recognized food additives. Cooking normally being required, there are attendant chemical changes in the composition and properties of carbohydrates, and in their interaction with proteins and other constituents (cf. the Maillard reaction). Restrictions upon the sale, marketing, and distribution for food use of specific carbohydrates are imposed by legislation in most countries, and stringent attention to detail is required in the labeling of products; for this purpose and for assessing the nutritional value of foodstuffs, the importance of analysis is self-evident. Spoiling and the effects of ageing need to be assessed and monitored, together with the presence of contaminants or adulterants; these may include colorants, toxic substances, or compounds that contribute

adversely to taste, and usually require application of a wide range of analytical techniques. Low-calorie sweeteners are in demand, some of which (sorbitol = D-glucitol; maltitol; chlorine-containing sucrose derivatives) are related to sugars; others are bitter. It is not surprising, therefore, that some 5000 different analytical procedures are applied within a single large company (Eurofins) engaged in food analysis; the entries on food analysis in *Chemical Abstracts* exceed  $10^5$  in number.

### Food Additives (Hydrocolloids)

Because many persons in developed countries consume low-calorie and low-fat foods and high-intensity sweeteners, the analysis of these products is of great importance in meeting the requirements of health regulations. Starches (foods in their own right) and starch-derived polysaccharides, maltodextrins, pectins, cereal  $\beta$ -D-glucans, modified celluloses, bacterial gums, seed and algal polysaccharides, plant gum exudates, inulins, and semisynthetic polyglucose may all be added to prepared foods in varying amounts, either singly or as mixtures. These polysaccharides

(hydrocolloids) cause thickening of solutions and stabilizing of emulsions, promote gelling and affect numerous related improvements in functionality. The quantities incorporated in food preparations may be very large or as low as fractions of 1%.

Starch hydrolysates (dextrins, syrups) are extensively used in the food industry, the degree of polymerization (DP) of these oligomers of D-glucose being an important analytical characteristic. The DP may be estimated by determining the reducing power relative to glucose, there being but one reducing end-group unit in each molecule; alternatively, SEC using gels or silica-based packings in columns and calibrating with substances of similar structure and known molecular weight may be used as a direct measure of molecular size and distribution.

Table 1 lists some of these food hydrocolloids, and their utilization, together with a general description of their respective molecular structures.

### Other Applications of Hydrocolloids

The industrial uses of these polysaccharides are by no means limited to foodstuffs, and although the

**Table 1** Molecular structures and uses of food hydrocolloids

Hydrocolloid	Application <sup>a</sup>	Structure
Starch	Bakery, cereals, soups, meats, desserts, etc.	Essentially linear, (1 → 4)- $\alpha$ -D-glucan (amylose) Ditto, branches at O-6 (amylopectin), average chain length 25
Modified starches	Meats, dairy, soups, confectionery, starch pastes	Starch phosphates, acetates, adipates
Maltodextrins	Dairy, desserts, beverages, preserves, bakery, confectionery	Linear, (1 → 4)- $\alpha$ -D-glucans, mol. wt. <4000
Microcrystalline cellulose	Sauces, dairy, desserts, bakery, meats	Linear, (1 → 4)- $\beta$ -D-glucan, fibers <1–15 $\mu$ m diameter
Cellulose derivatives	Dairy, bakery, meats, batters, sauces	Ditto, $-\text{OCH}_3$ , $-\text{OCH}_2\text{CHOHCH}_3$ , or $-\text{OCH}_2\text{CO}_2\text{H}$ at C-6
$\beta$ -D-Glucans	Cereal products, bakery	Linear, (1 → 3) (1 → 4)- $\beta$ -D-glucans
Synthetic polydextrose	Confectionery, beverages, soft foods (laxative)	Branched D-glucan, random linkages
Seed xyloglucans	Jams, confectionery	Linear, (1 → 4)- $\beta$ -D-glucan, $\alpha$ -D-Xyl at O6, $\beta$ -D-Gal at O2 of Xyl
Konjac mannan	Low-calorie processed foods	Linear, (1 → 4)- $\beta$ -D-glucan and $\beta$ -D-mannan; $-\text{OAc}$
Seed galactomannans	Dairy, bakery, sauces, pet foods	Linear, (1 → 4)- $\beta$ -D-mannan, $\alpha$ -D-Gal- at O-6
Inulins	Dairy, beverages, confectionery, bakery	Linear, (2 → 1)- $\beta$ -D-fructofuranan
Xanthan gum	Dairy, sauces, meats, bakery, beverages	Cellulosic chains, substituted by $\beta$ -D-GlcA, T and $\rightarrow$ 2-D-Manp; $-\text{OAc}$ , pyruvate
Gellan gum	Confectionery, dairy	Linear, $\beta$ -D-Glc (2), $\beta$ -D-GlcA, L-Rha; $-\text{OAc}$ , $-\text{O-L-glycerate}$
Pectins	Preserves, jellies, dairy, health foods	Linear and branched, partly Me-esterified and O-acetylated (1 → 4)- $\alpha$ -D-galacturonan, $\rightarrow$ 2-L-Rhap $\rightarrow$ ; D-Gal, L-Ara, etc.
Alginates	Bakery, sauces, beverages, dairy	Linear, (1 → 4)- $\beta$ -D-mannuronan and $\alpha$ -L-guluronan
Agars and carrageenans	Bakery, dairy, confectionery, meats, sauces	Sulfated, linear galactans, including 3,6-anhydrogalactose, etc.
Gum arabic	Confectionery, beverages, sauces	Highly branched, (1 → 3) and (1 → 6)- $\beta$ -D-galactan, substituted by L-Rhap, D-GlcA, L-Ara; some glycoprotein
Gum karaya	Bakery, confectionery, dairy, meats	Modified pectin, partially acetylated
Gum tragacanth	Bakery, confectionery, dairy, sauces	Modified, acidic arabinogalactan, and modified pectin

<sup>a</sup> Food hydrocolloids are often used in combinations, so that an analyte (e.g., a fat replacer) can be expected to contain two or more different polysaccharides.



stringent demands imposed on edible products are not necessarily applied, their characterization by standard analytical procedures is normally required. The usefulness of hydrocolloids is based largely on their high capacity to attract water, producing thickening and gelling of aqueous solutions. Other functions, such as emulsification, suspension, flocculation, binding, and film formation, find outlets in the textile, adhesive, paint, paper, and mining industries, and the manufacture of pharmaceuticals and agricultural chemicals, as well as in large numbers of other trade applications.

### Heavy Industry

Cellulose is a well-publicized renewable and biodegradable energy source, at present in abundant supply ( $>10^8$  tons per annum), but subject to enormous demands from the paper and textile industries, for building purposes, and for fuel. The bulk comes from forests, particularly gymnosperms, for which the operations of milling and grinding prior to lignin removal are less complicated, while the most refined form is derived from cotton. The insolubility and inertness of cellulose, which contribute to the rigidity of plant tissues, form the basis of its chemical analysis; organic solvent extraction of the test material is followed by chemical treatment for the removal of reactive adhering substances, and the residual  $\alpha$ -cellulose is estimated gravimetrically. Cellulose may be assayed without isolation in a similar manner by degradation and extraction of the more-easily hydrolyzed components of the cell wall. Cellulose derivatives (esters, ethers) employed in the manufacture of fibers or prepared foods are identified and analyzed on the basis of the percentages of substituent groups present. Carbohydrate polymers generally constitute an important group of substances requiring special analytical approaches, including pyrolysis and GC-MS.

The starch content of agricultural products and foods is sometimes reported by difference after determining moisture, ash, organic solubles (lipid), and protein. The depth of blue-black colorations formed on interaction with iodine/iodide gives a direct estimate of I, while a host of methods are available for isolating and characterizing the size (1–100  $\mu\text{m}$  diameter), shape, and properties of starch granules (organized, hydrogen-bonded structures containing bound water), which reflect the plant source. Not all starch is soluble even in hot water, and recourse may have to be made to acid hydrolysis and estimation of the glucose released, with appropriate account being taken of other glucans present in the test sample. Amylose and amylopectin (as well as intermediate)

fractions can be separated using methods based on the formation of complexes with 1-butanol or thymol, combined with SEC, or by lectin affinity-chromatography. Commercial test-kits for measuring amylose to amylopectin ratios are available. Analyses based on enzymatic hydrolysis are specific and highly informative as to the molecular structures of starch components. Genetically modified plant sources give rise to starches that vary considerably in amylopectin structure.

### Biology and Medicine

Glucose (dextrose), fructose (laevulose), sucrose, lactose, dextrans, starch, and gums form the bulk of the carbohydrates in pharmaceutical use. Of the particular analyses performed upon biological fluids, automated procedures for glucose based upon reducing power, or color reactions linked with the use of insolubilized enzymes, outnumber all others; nevertheless, the range of substances analyzed that are derived from natural sources is vast, embracing carbohydrate molecules small and large. Glycoconjugates, glycoproteins, and glycolipids particularly are of immense importance in all living systems, and present numerous unique analytical challenges. The clot-dissolving drug tissue plasminogen activator (t-PA), the primary regulator of red blood cell formation in mammals, erythropoietin, and the blood anticoagulant heparin are examples of glycoproteins that are used extensively in the medical field. Not the least important use of carbohydrates is their conversion on an immense scale to fermentation products, from ethanol to antibiotics.

## Overview of Methods of Analysis and Their Merits

### Colorimetric and Spectrometric Methods

While there are many general reagents (anthrone, triphenyltetrazolium chloride), it is fortunate that the various classes of sugars (pentoses, hexoses, amino sugars, deoxysugars, uronic acids) respond differently to color tests, as these are generally easy to carry out with the aid of standard chemical reagents and simple photometric apparatus, and are readily adapted to automation. Well over 50 combinations of reagents were in use for these purposes as much as a decade ago. Absorbance is measured and the concentration of carbohydrate determined after calibration using purified, known substances. The method is applicable to monosaccharides and their methyl ethers, and is usually based on heating with mineral acid in the presence of one of a great variety of amines, amino acids, aldehydes, phenols, and many

other classes of organic compound. A quantitative method based on the reduction of 3,5-dinitrosalicylic acid has long been established.

Infrared spectroscopy (IR, near-infrared (NIR), and Fourier transform infrared (FTIR)) is universally applied to the characterization of carbohydrates along with organic and inorganic substances generally, and modern statistical analysis of data (chemometrics) has enlarged its usefulness. At higher levels of sophistication are other, modern spectroscopic methods of analysis, predominant among which are many forms of NMR spectroscopy and MS, the latter being normally coupled with chromatographic separations, GC and HPLC predominating. It is possible by NMR procedures to identify and quantify the common sugars in fruit tissues and to define the microenvironments of polysaccharide components in biological samples. Laser-assisted fluorescence spectroscopy and voltammetry coupled with electrophoretic and enzymatic techniques have brought the limits of detection of carbohydrates to picomole levels. MS in all its forms is now of universal application in carbohydrate analysis. Where crystalline material is available, the ultimate probe of molecular structure is X-ray diffraction technology. Such an approach is indispensable in the characterization of biopolymers, starch for instance.

### Enzymatic Methods

Although in a sense enzymatic techniques are ancillary to the use of chemical, chromatographic, and spectroscopic methods of analysis, they are in fact indispensable for many diagnostic and other purposes. Thermolabile and generally costly, especially when suitably purified (and this may be very difficult), enzymes are also subject in their action to end-product inhibition. On the other hand, enzymes are used only in catalytic amounts, are adaptable to microassay, are often rapid and uniquely specific in their mode of action, and are applicable both on a massive scale, as in degradation of cellulose and fermentation of starch, and, when immobilized and coupled with co-factors, in the detection and assay of sugars at medium to very low levels. The development of biosensors using enzymes is expanding rapidly. Enzymatic methods are ideally suited to the analysis of biological fluids and chromatographic eluates. Without enzymes it would be impossible to solubilize cell walls in order to study their components; carbohydrases bring about the reduction in size of polysaccharides and permit their release from encrusting material, and proteolytic enzymes or lipases effect the removal of unwanted substances from the carbohydrate to be isolated and examined. This is

specifically the case in the examination of dietary fiber, largely carbohydrate, which by definition is not decomposed by digestive enzymes. The molecular study and assay of all polymeric carbohydrates is virtually dependent on the use of hydrolyases and lyases for their degradation in a rational manner, exo and endo (e.g., with  $\alpha$ - and  $\beta$ -amylases). The resultant mono- and oligosaccharides are then conveniently analyzed with the aid of specific oxidases and dehydrogenases. A plethora of restriction endonucleases isolated from microorganisms is available for controlled DNA fragmentation and subsequent genetic manipulation of plants, for example, with accompanying enzyme changes and modification of the polysaccharides, such as starch, xyloglucan, galactomannan, or pectin, that they produce.

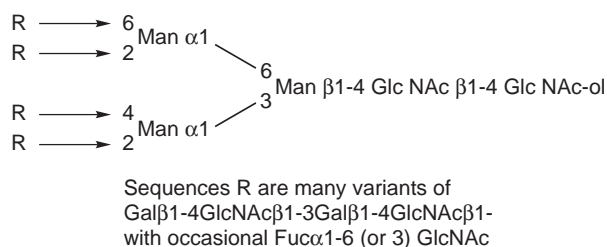
### Chromatographic Procedures

The degree of purity of a sugar sample can be ascertained or the identification and assay of mixtures of sugars can normally be achieved by any one or a combination of chromatographic methods, of which GC of volatile derivatives or HPLC with electrochemical detection of the sugars or derivatives prepared in order to confer suitable spectroscopic properties, are comparable in their efficacy. Supercritical fluid chromatography (SFC) is also a flexible, rapid, and efficient technique comparable with HPLC but by no means as widely used.

For sugar mixtures, GC on packed or capillary columns of varying polarity, with temperature programming, is carried out on alditol or aldonoitrile acetates, trimethylsilylated methyl glycosides, acetylated or trimethylsilylated oximes, N- or O-trifluoroacetyl derivatives, or any other suitably volatile derivatives. Retention times in comparison with standard derivatives provide a sufficient identification in most cases, but for more precise work MS with selected ion monitoring of fragments from the separated derivatives is advisable. MS is an invaluable approach to positioning the methyl substituents in sugar methyl ethers formed on hydrolysis of permethylated oligo- and polysaccharides, for example.

HPLC methods are extremely sensitive toward sugars and sugar derivatives, an important advance over detection by means of refractive index measurement lying in the use of ion chromatography with pulsed amperometric detection at copper, platinum, or gold electrodes. Analytical processes may be fully automated using various postcolumn derivatization procedures as required.

SEC enables carbohydrates to be differentiated, broadly on the basis of an average molecular weight. Thus, degradation products of many types of polysaccharide,

**Scheme 2**

mixtures of reduced, complex oligosaccharides (Scheme 2) from glycoproteins, and asparagine-linked glycopeptides have been analyzed by SEC.

DEAE-Sephadex is a column packing with both ion-exchange and size-exclusion properties, and chiral stationary phases have widespread uses. Affinity (or biospecific interaction) chromatography finds wide application in the purification of carbohydrates and glycoproteins. The technique of hydrophobic-interaction chromatography has been applied successfully to lignin-carbohydrate complexes and glycoproteins, the noncarbohydrate moieties being responsible for binding to hydrophobic groups in the column packings. Glycosaminoglycuronans (GAGs) have also been chromatographed on hydrophobic media. Ion-pair chromatography on reversed-phase columns is a process whereby such anionic carbohydrates as sugar phosphates, pectic oligomers, and fragments from hyaluronic acid, paired with the tetrabutylammonium ion, have been fractionated using aqueous eluents.

Modern methods employ CE with electrochemical detectors.

### Methods Used in Di- and Oligosaccharide Analysis

Sucrose and lactose occur naturally and are marketed on a huge scale, so that numerous standard procedures of analysis exist. Being furanosidic, sucrose requires special care in its extraction and handling, contact with aqueous acid to be avoided, and precautions taken against fermentation. Standard tables for specific gravity, refractive index, and specific rotation measurements (saccharimetry) have long been used, but modern techniques are based upon acid- or enzyme-catalyzed hydrolysis and assay of the reducing, glucose-fructose mixture ('invert sugar') produced. For this purpose, cuprimetric methods (Somogyi-Nelson; use of Munson-Walker tables) are well established and suitable for routine analyses. There are also biochemical techniques employing phosphorylases in coupled reactions that do not require prior hydrolysis. These approaches, to which numerous others including the use of IR reflectance and elaborate biosensors might be added for the analysis of

sucrose and related oligosaccharides, are general for di- and oligosaccharides, most of which are obtained in practice by cleavage of polysaccharides and glycoconjugates, often by the use of specific hydrolases. The usual approach for the analysis of disaccharides is HPLC of hydrolyzed disaccharides. NMR and MS are indispensable in modern structural analysis of carbohydrates.

### Methods Used in Polysaccharide Analysis

The requirements of polysaccharide analysis are essentially similar to those of oligosaccharides. Thus, many types of analytical operation are applied that the relevant techniques will be discussed on the basis of a structural classification of the polysaccharides themselves. Apart from cellulose and its derivatives, which comprise the major components of paper, textiles, and adhesive tapes, most industrial polysaccharides are incorporated in one form or another into processed foods or used in fermentation processes. Starch predominates among these but pectins, galactomannans, galactans, and alginates are also valuable ingredients. Most of the methods employed in the analysis or characterization of polysaccharides are based upon inspection, their solubility relationships, molecular weight distribution, optical activity, and electrophoretic behavior, with subsequent determination of the monosaccharide constituents and their modes of linkage. Fundamentally important criteria of polysaccharides are their immunological behavior and response towards enzymes.

**Starch** While there are specific methods, such as the measurement of the depth of the blue iodine coloration with lipid-free (1), and degradation by enzymes (amylases) followed by sugar estimation, which make the analysis of starch a relatively straightforward procedure, the determination of starch is complicated by problems of extraction from the natural source or by the fact that, in processed foods, dextrans and other products of partial hydrolysis of starch may also be present. As all glucans are hydrolyzed to D-glucose, assay of this sugar gives a measure not only of starch but also of accompanying compounds containing bound glucose unless the hydrolysis had been accomplished by specific enzymes, the most popular method today. Similarly, methods based upon measurement of optical rotation of starch ( $[\alpha]_D \sim +200^\circ$ ) solutions are entirely reliant upon the known presence or absence of other chiral substances, and upon the degree to which the natural starch granules have been brought into solution. After moisture and ash determination, hot water affects a rough separation of the more soluble amylose and

dextrins from amylopectin. 'Starch' is extracted from cereals in calcium chloride solution, 52% aqueous perchloric acid, or moist dimethyl sulfoxide. For various other classes of foodstuff, types of extraction have been evolved to suit the nature of the material. Selective acid hydrolysis of starch, or incubation at pH 4.5–5.5 with amyloglucosidases assisted by  $\alpha$ -amylase following prior gelatinization in hot water, precedes measurement of the glucose liberated. As detailed above, reducing power, response to color reagents, or enzymatic analysis specific for D-glucose constitute methods of estimating starch when appropriate conversion factors are applied. Biosensors are now used. It is important to observe that the portion of starch that is undegraded by enzymes is regarded as a component of dietary fiber.

Whole or degraded starch is currently analyzed by SEC, by the related technique of field flow fractionation (FFF), HPLC in different forms, and by a range of physical and chemical techniques including microscopy (light and SEM), X-ray diffraction, differential scanning calorimetry, NIR and FTIR spectrometry, solid-state  $^{13}\text{C}$  NMR, and most importantly by the use of such enzymes as the  $\alpha$ - and  $\beta$ -amylases and amyloglucosidase.

**Dextrins, glucose syrups, and modified starches** Dextrins, formed on heating starch, are assayed together with the unmodified polysaccharide. The dextrin-iodine coloration is reddish brown. Glucose syrups (from starch) are extremely soluble in water, and the reducing power ('dextrose equivalent') of a sample whose moisture and ash content is known affords a measure of the length of the chain of glucose residues. Free glucose is measured by the glucose oxidase method. The molecular weight distribution, which is an important property affecting viscosity, is best measured by SEC, or by an HPLC procedure. If modification of the starch by oxidation, etherification, or esterification (e.g., phosphate formation) has been carried out, methods appropriate to the specific analyses required must be adopted.

**Water-soluble, nonstarch polysaccharides** These components accompany starch in extraction processes, and are removed if possible by bringing the solution to 80% with respect to ethanol, proteins being precipitated as well. Acid hydrolysis of a test portion of the polysaccharide indicates from a chromatographic analysis of the neutral sugars released whether glucans (usually  $\beta 1 \rightarrow 3$  and  $\beta 1 \rightarrow 4$ -linked), galactomannans, or arabinoxylans are present, while the uronic acids (galacturonic acid for pectins or gums, glucuronic for gums or pectins, mannuronic and guluronic for alginates) are useful markers for different types of acidic polysaccharide. Neutral and

acidic polysaccharides in solution are differentiated by addition of cetyltrimethylammonium bromide (Cetavlon), which complexes with acidic polysaccharides, causing flocculation. The separated precipitate is then stirred in a minimum quantity of  $1 \text{ mol l}^{-1}$  sodium chloride solution, to bring about dissociation, and the polysaccharide is regenerated by pouring into ethanol (containing acetone and ether if necessary). More often than not a range of monosaccharide and uronic acid components are detected on hydrolysis of samples of the extracted polysaccharides, in which even a program of separation procedures (selective precipitation, dialysis, or chromatographic approaches) is adopted in order to isolate reasonably homogeneous fractions for detailed study, after their initial assay by weighing.

**Insoluble, noncellulosic polysaccharides** There is a gradation in the ease of extraction of polysaccharides from cell wall or endosperm from plant sources, the harshest methods generally yielding residual, colorless cellulose that can be estimated gravimetrically. After water extraction of the sample previously washed with organic solvents, reagents that sequester calcium ions (solutions of oxalate, citrate, bicarbonate, ethylenediaminetetraacetic acid, and related amphoteric compounds) are effective in bringing pectic substances into solution. These are acidic polysaccharides and the counterion will be dictated by the cation of the salt used. Salts are then conveniently removed by dialysis against distilled water, and the retentate on being freeze-dried (lyophilized) yields the pectic material in (usually) a colorless, fibrous form that is suitable for structural analysis or utilization. The gelling properties of such products (typically from fruits, leaves, or plant stems) are normally of greatest interest. Continued extraction of plant tissue, now with dilute aqueous alkali under an atmosphere of nitrogen to prevent oxidative degeneration, brings into solution various ill-defined polysaccharide mixtures for which the traditional name of hemicellulose is used. Neutralization of the solution with acetic acid yields hemicellulose A as a precipitate, and addition of ethanol to the supernatant solution gives hemicellulose B. A third soluble fraction, hemicellulose C, results on treatment with 17.5% NaOH. In any serious study of the hemicelluloses, further separations of hemicellulosic preparations are required, such as may be effected by column chromatography, but the tendency of the extracted polysaccharides to revert to insoluble material constitutes a serious barrier to effective isolation of 'pure' products, such as acidic, arabinose-substituted  $\beta$ -D-xylans and glucomannans, both of which have a cellulose-like core.

After these successive extractions of plant material, the residual cellulose is accompanied by dark brown polyphenolic material (lignin) that (unlike the cellulose) is insoluble in 72% (m/m) sulfuric acid–water. In the course of the solvent separation scheme, often before alkali is used, the standard method of removing lignin is to bring it into solution by means of sodium chlorite solution at an adjusted pH; the delignification process opens up plant tissues, permitting more thorough extraction of carbohydrate, but also causes some degradation and must be employed with due caution. Hydrolysis of the cellulosic residue normally reveals sugars, other than glucose, such as xylose or mannose, thus providing a measure of the accompanying xylan or glucomannan.

Any systematic study of cell walls requires an  $\alpha$ -D-galacturonanase to release pectic and other matrix components. Molecular structural studies have so far progressed that elaborate models showing the various polymeric constituents in juxtaposition have been generated. After enzymatic elimination of starch and protein, hydrolysis of the residual polysaccharide and estimation of the uronic acids and monosaccharides released furnish considerable information on the composition of agricultural samples. Partial depolymerization affords the complex, well-studied rhamnogalacturonans I and II (RG-I and RG-II).

**Dietary fiber** The analysis of soluble polysaccharides that resist enzyme action in the mammalian digestive system is important as the significance of this material, termed ‘dietary fiber’, is now fully recognized.

As with most analyses of polysaccharides, definition of the substances involved is the key to the methodology adopted. Dietary fiber comprises pectic and hemicellulosic components, and excludes cellulose, which is insoluble, and most of the ingested starch. Included are the natural plant gums (complex, branched, acidic heteroglycans), seed galactomannans, and industrially produced bacterial polysaccharides that are employed as food additives to impart desirable rheological properties and to limit caloric intake. Formidable as the problem is to obtain analytical results that have meaning in a physiological context, a large body of information on the dietary fiber content of foodstuffs is available. Common to the different methods of analysis is the removal of starch and protein, by enzymes that simulate human digestion, or by the use of solvent extraction with aqueous neutral detergents (yielding NDF); residual fiber is then weighed or analyzed after hydrolysis and estimation of the monosaccharides released. Future development in nutritional studies is

largely dependent upon  $^{13}\text{C}$  labeled polysaccharides becoming available.

**Carbohydrate hydrocolloids** As hydrocolloids constitute the majority of polysaccharides encountered in industry and research, the methods of analysis that apply variously to components of this group, including starch that is strictly in a category of its own on account of the overwhelming quantities used, embrace virtually all aspects of polysaccharide analysis. Glycoconjugates require certain additional, special procedures that are considered in the section that follows this account.

Hydrocolloids are dissolved in water with warming and agitation or sonication, and precipitated by addition of alcohols (MeOH, EtOH, or propan-2-ol), acetone, or ammonium sulfate. Acidic polysaccharides form insoluble complexes with Cetavlon, the precipitate being redissolved in aqueous salt solution, and the polysaccharide recovered as the sodium salt after dialysis and freeze-drying. Freeze-dried specimens of polysaccharides generally are analyzed for C, H, N, mineral ash, sulfate, and phosphate, and tested by color reactions or spectroscopy (IR, NMR) for uronic acid, amino- and acetamidodeoxy sugar, pentose, hexose, deoxyhexose, and anhydrosugar components. Purification is performed where necessary by column chromatographic methods (size-exclusion, partition, ion-exchange, affinity). Optical rotation measurements afford an index of purity and identity, and acid hydrolysis of the sample gives a good indication of the class or classes of polysaccharide present. The composition of the polysaccharide hydrolysate is normally determined by GC, HPLC, or SFC, with derivatization where necessary.

There are more or less specific color tests for starch (iodine blue) and carrageenans (using methylene blue), and in some instances the action of enzymes is diagnostic and used in reagent kits. A relatively new approach to the assay of food gums and stabilizers, one that shows many advantages in speed and specificity, is enzyme-linked immunosorbent assay, which has a number of variants.

Classification of the polysaccharide type is achieved fundamentally by hydrolysis to identifiable sugars, and by methylation analysis, in which the cleavage products from the permethylated polysaccharide are identified and assayed, after conversion to alditol acetates, acetylated aldononitriles, or other derivatives, by GC and GC–MS. Methanolysis of the permethylated polysaccharide and conversion to TMS derivatives is an alternative. The proportions of monosaccharides and their linkage modes are deduced from the molar quantities of sugar methyl ethers derived from the processed polysaccharide

sample. The category to which the polysaccharide belongs is thus determined (Table 1).

Spectroscopic methods of analysis are invaluable, particularly  $^1\text{H}$  and  $^{13}\text{C}$  NMR, which in one- and two-dimensional and various other modes permits the assignment of most of the signals obtained from polysaccharide solutions (in water,  $\text{D}_2\text{O}$ , or DMSO) to the protons and carbon atoms involved, irrespective of the molecular weight of the polysaccharide. The units and even sequences of the sugar moieties in bacterial exopolysaccharides, which comprise regular repeating structures of from three to seven monosaccharide units, may be deduced with additional information being obtained by MS. This study is aided by preliminary depolymerization using viral lyases (bacteriophages), and in an analogous manner all other polysaccharides that can be subjected to attack by endohydrolases can be readily characterized by NMR spectroscopic analysis of the polysaccharides and their products. Anomeric proton and carbon signals are easily identified, as are non-carbohydrate moieties (pyruvate ketals, methyl ethers, acetyl, or sulfate derivatives). Though not a sensitive method, low-resolution proton spectroscopy is used as a probe for examining physical and chemical properties that underlie the palatability of foodstuffs. Relaxation times ( $T_1$  and  $T_2$ ) are used to characterize the decay rates of proton signals, which, being sensitive to molecular mobility, can differentiate solid from liquid and define the quantity, mobility, and microenvironment of water molecules/droplets in complicated matrices. Solid-state NMR using the technique of cross-polarization, magic angle spinning permits the assignment of chemical shifts (values) in crystalline amyloses and in complex food structures, selected regions of which are examined by NMR imaging.  $^{13}\text{C}$  NMR enables alginates, galactomannans, starch, agar, carrageenans, and cellulose derivatives to be identified.

Measurements of IR absorption of thin films are useful for identification with authentic samples, particularly if in addition pyrolysis and GC/IR analysis of the decomposition products is carried out. The use of specific enzymes is important in aiding the isolation of polysaccharide components from natural sources, and in producing identifiable oligosaccharides that facilitate molecular structural analysis.

**General considerations** If the purpose of analysis is to determine the molecular structure of a polysaccharide (presumed to be homogeneous) in all its detail, standard procedures require the quantitative determination of all sugar units and of their respective linkage modes, and the establishment of sequences. X-ray diffraction measurements can in many cases

supply three-dimensional information, e.g., for oriented fibers or polysaccharide gels. CD supported by NMR is applied to monitor the behavior of polysaccharides in solution.

Methods involving partial breakdown of the polysaccharide sample are of great value in elucidating sugar sequences within limited regions of the macromolecule. These methods normally involve attack at the glycosidic intersugar linkages where these are weak either because they are furanosidic or have been rendered labile by selective chemical modification of sugar residues. Graded acid or enzymatic hydrolyses yield mixtures of oligosaccharides that require separation on a sufficient scale to permit structural identification of the products. Modern equipment couples chromatographic and spectrometric instrumentation so that isolation and structure determination are accomplished in one operation. Acetolysis, methanolysis, and mercaptolysis prior to aqueous acid hydrolysis may be required. Hexose/hexose and glucuronic acid/hexose bonds are normally resistant to fission; on the other hand, those positions in a polysaccharide to which labile groups are bonded may be ascertained by comparing the analytical properties of the initial and the partially hydrolyzed ('autohydrolyzed') polysaccharides.

A widely used procedure is the Smith degradation. The polysaccharide (or glycoconjugate) is oxidized with periodate, which selectively attacks all sugar units with exposed 1,2-diol groups, then reduced with borohydride to convert the resulting aldehyde to primary alcohol groups, and given a mild acid hydrolytic treatment that must be so controlled as to leave glycosidic bonds unattached but to effect complete cleavage of the glycolaldehyde acetals (at the sites of all units that have been periodate-oxidized and reduced). All sugar units protected by, e.g., (1 $\rightarrow$ 3)- or (1 $\rightarrow$ 2,4)-linkages remain intact, and joined together if contiguous in the parent polysaccharide. Isolated sugar units are recovered as polyol (ethan-1,2-diol, or glycerol) glycosides. In this manner (1 $\rightarrow$ 3)-linked  $\beta$ -D-galactan moieties have been obtained from arabinogalactans, and erythritol glucosides from (1 $\rightarrow$ 3) (1 $\rightarrow$ 4)- $\beta$ -D-glucans. Separation of the products of Smith degradation is performed by solvent extraction, SEC, or HPLC. Other forms of selective degradation include the reductive cleavage method of Gray, whereby adjustment of the catalyst controls the extent of glycosidic breakdown by triethylsilane, and a series of related base-catalyzed degradations ( $\beta$ -eliminations) that cause the fission of substituent sugars from 0 to 4 of glycuronate esters and concomitant detachment of the degraded acidic unit from its interior sugar unit. HPLC is the usual method adopted in isolating meaningful, structural

fragments from such degradations, which, together with several other approaches to targeting uronic acid constituents, have been developed by Aspinall in the study of pectins and plant gums.

## Glycoconjugates

Glycoconjugates contain protein, lipid, or polyphenolic moieties joined to carbohydrate, and though determinations of the proportion of total carbohydrate and of the monosaccharide units, and the modes of linkage involved follow standard procedures as outlined, the overall analysis has certain other requirements. Isolation methods are frequently attended by the risk of degradation. In general, the quantities available for analysis are not large, so that the development of micromethods has been of first importance. As natural compounds glycoconjugates respond to enzymatic treatment and are frequently assayed by immunological techniques.

## Glycoproteins

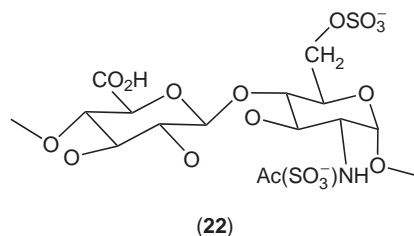
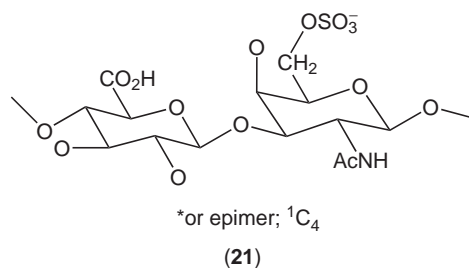
Glycoproteins result from the covalent association of glycans with proteins, and may be classified according to whether the glycosyl linkage is to O or N (Table 2). N-glycosylprotein glycans are divided into three groups according to monosaccharide composition and oligosaccharide structure as high mannose (containing only Man and GlcNAc), complex (galactose, sialic acid, and Fuc in addition), and hybrid (oligomannosidic, with peripheral N-Ac lactosamine; Scheme 2) types. A given oligosaccharide located at a specific amino acid in a glycoprotein may be polymorphic due to partial substitution of sugar residues on a similar core structure, leading to a type of diversity called microheterogeneity that is common to almost all glycoproteins. Glycan structures associated with glycoproteins range from short linear chains to more highly ramified oligosaccharides with multiple (usually two to five) outer chains (Scheme 2), termed antennae.

**Table 2** The amino acid and sugar linkages engaged in carbohydrate–peptide linkages<sup>a</sup> in glycoproteins

Amino acid	Linkage type	Corresponding sugar
L-Asparagine	N	2-Acetamido-2-deoxy-D-glucose
L-Threonine	O	2-Acetamido-2-deoxy-D-galactose
L-Serine	O	2-Acetamido-2-deoxy-D-galactose or 2-acetamido 2-deoxy-D-xylose
5-Hydroxy-L-lysine	O	D-Galactose

<sup>a</sup> A glycoprotein may exhibit both O and N linked chains.

Glycoproteins are widely distributed in animals, plants, microorganisms, and viruses, in which the glycans perform a range of important biological roles particularly in respect of intercellular recognition and adhesion. The glycan structure of the cell membrane glycoproteins (and glycolipids) is profoundly altered in cancer cells leading to the appearance of cell surface neoantigens that could be a factor in cancer induction and metastatic diffusion. Proteoglycans are a large subgroup of glycoproteins that consist of a central core protein to which are attached a number of long, highly charged glycosaminoglycan (GAG) chains, which confer exceptional water-holding properties. Proteoglycans are high molecular weight, unbranched heteropolymeric molecules consisting of repeating disaccharides (4-linked uronic acid to 3- or 4-linked acetamidodeoxyhexose; (21) and (22)) substituted with sulfate ester groups, and are distributed ubiquitously throughout the extracellular matrices of connective tissues. A range of plant proteoglycans and glycoproteins occurs in plant cell walls, many of them cross-linked in the matrix, so that treatment with cold, anhydrous HF is invaluable in effecting their isolation. The arabinogalactan-proteins (AG-Ps) are, however, water-soluble, and many enzymes, lectins, and the cell wall component extension may be extracted directly into aqueous NaCl.



Glycoproteins are purified using methods based on precipitation with ethanol or various salts, and by chromatographic (SEC, ion-exchange, hydrophobic interaction, HPLC) and electrophoretic techniques. The two latter techniques have been extensively exploited in mapping N-linked glycans and glycopeptides by two- and even three-dimensional techniques, the unique significance of sugar sequences in biological recognition events being generally understood. Lectin-affinity chromatography is used extensively



to separate glycoproteins according to their glycan composition. Prior to analysis of the oligomers in glycoproteins it is essential that they be split off from the protein core. This is achieved enzymatically with endo-glycosidases, or chemically with trifluoromethane sulfonic acid, alkaline sodium borohydride, hydrazine, or anhydrous HF. O- but not N-glycosidic bonds are cleaved at room temperature with the last-named reagent, and by operating at temperatures down to  $-23^{\circ}\text{C}$ , selective fission of glycosidic bonds may be achieved.

Kits for analyzing glycoproteins are now available, utilizing recombinant endo-enzymes together with electrophoresis and fluorescent labeling with ANTS (aminonaphthalene-1,3,6-trisulfonic acid); PMP (1-phenyl-3-methyl-5-pyrazolone), or other, similar type of reagent can be used. MS plays a crucial role in glycopeptide characterization where techniques such as electrospray LC-MS and MALDI-TOF MS are utilized.

The analytical procedures applied to glycoconjugates are essentially those described for pure glycans. In most instances, however, only minute quantities of carbohydrate are available, and the enhancement of detection sensitivity (to the picogram level) by incorporation of a radioactive or fluorogenic label is essential. HPLC with fluorimetric or pulsed amperometric detection is used extensively, and modern developments in MS (e.g., electrospray ionization MS-MS) are especially applicable. NMR using computerized databanks are of outstanding importance, and as an example of the exceptional power of the method, high-resolution magic-angle spinning NMR can detect and characterize O-specific polysaccharide components of the cell wall on the surface of 'living' bacteria. Given larger amounts of material such as cartilage or skin, the unique structural feature of alternating 3-linked and 4-linked sugar units of the type present in some GAGs (21) enables the processes of Smith degradation (cf. cereal  $\beta$ -D glucans) and  $\beta$ -elimination (22) using mild alkali to be used to good effect. Lyases, or the use of mild alkali, also cause fission of the 4-O-glycosyl substituent from the uronic acid chain units of (21) and (22), producing double bonds that are easily detectable.

## Glycolipids

Glycolipids are distributed throughout animal, plant, and microbial cells. The glyco-components, linked to ceramide or other glycerol derivatives, are mono- or oligosaccharide chains that are frequently branched, and may be substituted with acetyl or sulfate groups. The precise carbohydrate structures, including, e.g., the presence or absence of sialic acid, confer

biological specificity on glycolipids. The extraction of glycolipids is generally achieved by Folch procedure, in which fresh tissue is homogenized with chloroform/methanol and the insoluble material is removed by filtration or centrifugation. The crude glycolipid extract may then be purified using a number of chromatographic techniques including ion exchange, silica gel, TLC, HPLC, and SFC methods.

A considerable amount of information regarding the glycan structures of glycolipids may be obtained by a combination of methylation and mass spectral analysis of the intact molecule. Sialic acids are easily removed, and oligosaccharide chains may be cleaved from the lipid portion either chemically or enzymatically with a specific endo-glycoceramidase.

*See also:* **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Chiroptical Analysis.** **Chromatography:** Overview; Principles. **Clinical Analysis:** Glucose. **Enzymes:** Enzyme-Based Electrodes. **Food and Nutritional Analysis:** Overview. **Infrared Spectroscopy:** Overview. **Mass Spectrometry:** Overview. **Nuclear Magnetic Resonance Spectroscopy:** Overview. **Nuclear Magnetic Resonance Spectroscopy Applications:** Food. **Optical Spectroscopy:** Detection Devices. **Sampling:** Theory. **Spectrophotometry:** Overview. **Sweeteners.** **X-Ray Absorption and Diffraction:** Overview.

## Further Reading

- Anumula KR and Dhume ST (1998) High resolution and high sensitivity methods for oligosaccharide mapping and characterization by normal phase high performance liquid chromatography following derivatization with highly fluorescent anthranilic acid. *Glycobiology* 8: 685–694.
- Anumula KR and Du P (1999) Characterization of carbohydrates using highly fluorescent 2-aminobenzoic acid tag following gel electrophoresis of glycoproteins. *Analytical Biochemistry* 275: 236–242.
- Bush CA, Martin-Pastor M, and Imberty A (1999) Structure and conformation of complex carbohydrates of glycoproteins, glycolipids, and bacterial polysaccharides. *Annual Review of Biophysics and Biomolecular Structure* 28: 269–293.
- Chaplin MF and Kennedy JF (eds.) (1994) *Carbohydrate Analysis – A Practical Approach*. Oxford: IRL Press.
- Charlwood J, Bryant D, Skehel JM, and Camilleri P (2001) Analysis of N-linked oligosaccharides: Progress towards the characterisation of glycoprotein-linked carbohydrates. *Biomolecular Engineering* 18: 229–240.
- Charlwood J, Hanrahan S, Tyldesley R, Dwek M, and Camilleri P (2002) Use of proteomic methodology for the characterisation of milk fat globular membrane proteins. *Analytical Biochemistry* 301: 314–324.

- Collins PM and Ferrier RJ (1995) *Monosaccharides*. Chichester: Wiley.
- Doco T, O'Neill MA, and Pellerin P (2001) Determination of the neutral and acidic glycosyl-residue compositions of plant polysaccharides by GC–EI–MS analysis of the trimethylsilyl methyl glycoside derivatives. *Carbohydrate Polymers* 46: 249–259.
- Duus JO, Gotfredsen H, and Bock K (2000) Carbohydrate structural determination by NMR spectroscopy: Modern methods and limitations. *Chemical Reviews* 100: 4589–4614.
- El Rassi Z (ed.) (2002) *Carbohydrate Analysis by Modern Chromatography and Electrophoresis*. Amsterdam: Elsevier.
- Kawasaki N, Ohta M, Hyuga S, Hyuga M, and Hayakawa T (2000) Application of liquid chromatography/mass spectrometry and liquid chromatography with tandem mass spectrometry to the analysis of the site-specific carbohydrate heterogeneity in erythropoietin. *Analytical Biochemistry* 285: 82–91.
- Linhardt RJ and Bazin HG (2001) Separation and purification of carbohydrates. *Glycoscience* 1: 63–74.
- Sato Y, Suzuki M, Nirasawa T, Suzuki A, and Endo T (2000) Microsequencing of glycans using 2-aminobenzamide and MALDI-TOF mass spectrometry: Occurrence of unique linkage-dependent fragmentation. *Analytical Chemistry* 72: 1207–1216.
- Southgate DAT (1991) *Determination of Food Carbohydrates*, 2nd edn. London: Elsevier Applied Science.
- Stephen AM (ed.) (1995) *Food Polysaccharides*. New York: Dekker.
- Whistler RL and BeMiller JR (1997) *Carbohydrate Chemistry for Food Scientists*. St Paul, MN: Eagan Press.

## Sugars – Spectrophotometric Methods

**S C Churms**, University of Cape Town, Rondebosch, South Africa

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Analysis of sugars is a crucial aspect of carbohydrate chemistry, not only because of the economic importance of certain sugars, such as glucose, fructose, and sucrose in, for example, the food industry, but also because structural studies of polysaccharides and glycoconjugates require identification and quantitation of the constituent monosaccharides. Mixtures of sugars are usually analyzed by chromatographic methods. However, spectrophotometric methods also play an important role, as pre- or postcolumn derivatization is often employed to enhance sensitivity in qualitative and quantitative analysis of sugars by liquid chromatography (LC). In this article, emphasis is laid on the use of ultraviolet (UV)–visible spectrophotometry and fluorimetry for that purpose. The standard methods for detection and determination of carbohydrates are described elsewhere.

### Precolumn Derivatization for LC of Sugars

#### For UV Detection

Hydroxyl groups of sugars are points of substitution for chromophoric aryl groups, the chromophores being introduced by formation of ester or ether

linkages. Benzoylation has been widely employed for precolumn derivatization of sugars to be analyzed by LC on unmodified silica or in reversed-phase mode. Benzoyl chloride in pyridine was initially used as the derivatizing agent, but subsequently the preferred reagent has been benzoic anhydride, catalyzed by 4-dimethylaminopyridine; this is especially the case where amino sugars are present in the mixture to be analyzed, since some *N*-benzoylation occurs with benzoyl chloride. To obviate problems arising from the production of multiple peaks by anomeric pairs and pyranose/furanose isomers, sugars are often reduced to the corresponding alditols (by treatment with sodium borohydride) before benzoylation. Benzoylates of sugars are usually detected and determined after LC by UV photometry at 230 or 254 nm, although 275 nm has been the wavelength generally used in detection of benzoylated alditols.

4-Nitrobenzoylation has been used as a derivatization method by some in preference to benzoylation, since the incorporation of the additional chromophoric group produces a 10-fold increase in the sensitivity of UV detection. The derivatizing agent is 4-nitrobenzoyl chloride in pyridine, and the derivatives are detected by UV photometry at 260 nm.

Conversion of reducing sugars to hydrazone derivatives is also an effective method of introducing chromophoric groups. Precolumn derivatization to 2,4-dinitrophenylhydrazones has been recommended for LC analysis of the neutral sugar components of glycoproteins, as the reagent is selective for these, with no interference from amino sugars, uronic acids,

- Collins PM and Ferrier RJ (1995) *Monosaccharides*. Chichester: Wiley.
- Doco T, O'Neill MA, and Pellerin P (2001) Determination of the neutral and acidic glycosyl-residue compositions of plant polysaccharides by GC–EI–MS analysis of the trimethylsilyl methyl glycoside derivatives. *Carbohydrate Polymers* 46: 249–259.
- Duus JO, Gotfredsen H, and Bock K (2000) Carbohydrate structural determination by NMR spectroscopy: Modern methods and limitations. *Chemical Reviews* 100: 4589–4614.
- El Rassi Z (ed.) (2002) *Carbohydrate Analysis by Modern Chromatography and Electrophoresis*. Amsterdam: Elsevier.
- Kawasaki N, Ohta M, Hyuga S, Hyuga M, and Hayakawa T (2000) Application of liquid chromatography/mass spectrometry and liquid chromatography with tandem mass spectrometry to the analysis of the site-specific carbohydrate heterogeneity in erythropoietin. *Analytical Biochemistry* 285: 82–91.
- Linhardt RJ and Bazin HG (2001) Separation and purification of carbohydrates. *Glycoscience* 1: 63–74.
- Sato Y, Suzuki M, Nirasawa T, Suzuki A, and Endo T (2000) Microsequencing of glycans using 2-aminobenzamide and MALDI-TOF mass spectrometry: Occurrence of unique linkage-dependent fragmentation. *Analytical Chemistry* 72: 1207–1216.
- Southgate DAT (1991) *Determination of Food Carbohydrates*, 2nd edn. London: Elsevier Applied Science.
- Stephen AM (ed.) (1995) *Food Polysaccharides*. New York: Dekker.
- Whistler RL and BeMiller JR (1997) *Carbohydrate Chemistry for Food Scientists*. St Paul, MN: Eagan Press.

## Sugars – Spectrophotometric Methods

**S C Churms**, University of Cape Town, Rondebosch, South Africa

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Analysis of sugars is a crucial aspect of carbohydrate chemistry, not only because of the economic importance of certain sugars, such as glucose, fructose, and sucrose in, for example, the food industry, but also because structural studies of polysaccharides and glycoconjugates require identification and quantitation of the constituent monosaccharides. Mixtures of sugars are usually analyzed by chromatographic methods. However, spectrophotometric methods also play an important role, as pre- or postcolumn derivatization is often employed to enhance sensitivity in qualitative and quantitative analysis of sugars by liquid chromatography (LC). In this article, emphasis is laid on the use of ultraviolet (UV)–visible spectrophotometry and fluorimetry for that purpose. The standard methods for detection and determination of carbohydrates are described elsewhere.

### Precolumn Derivatization for LC of Sugars

#### For UV Detection

Hydroxyl groups of sugars are points of substitution for chromophoric aryl groups, the chromophores being introduced by formation of ester or ether

linkages. Benzoylation has been widely employed for precolumn derivatization of sugars to be analyzed by LC on unmodified silica or in reversed-phase mode. Benzoyl chloride in pyridine was initially used as the derivatizing agent, but subsequently the preferred reagent has been benzoic anhydride, catalyzed by 4-dimethylaminopyridine; this is especially the case where amino sugars are present in the mixture to be analyzed, since some *N*-benzoylation occurs with benzoyl chloride. To obviate problems arising from the production of multiple peaks by anomeric pairs and pyranose/furanose isomers, sugars are often reduced to the corresponding alditols (by treatment with sodium borohydride) before benzoylation. Benzoates of sugars are usually detected and determined after LC by UV photometry at 230 or 254 nm, although 275 nm has been the wavelength generally used in detection of benzoylated alditols.

4-Nitrobenzoylation has been used as a derivatization method by some in preference to benzoylation, since the incorporation of the additional chromophoric group produces a 10-fold increase in the sensitivity of UV detection. The derivatizing agent is 4-nitrobenzoyl chloride in pyridine, and the derivatives are detected by UV photometry at 260 nm.

Conversion of reducing sugars to hydrazone derivatives is also an effective method of introducing chromophoric groups. Precolumn derivatization to 2,4-dinitrophenylhydrazones has been recommended for LC analysis of the neutral sugar components of glycoproteins, as the reagent is selective for these, with no interference from amino sugars, uronic acids,

alditols, or amino acids. The recommended reagent is a solution of 2,4-dinitrophenylhydrazine (1.5%, m/v) in 1,2-dimethoxyethane. The reaction is acid-catalyzed and therefore a 2% (v/v) solution of trifluoroacetic acid in methanol is also added to the sugar mixture. The mixture is heated at 65°C for 90 min, cooled, and then excess 2,4-dinitrophenylhydrazine is precipitated by addition of acetone. The sugar derivatives are extracted into 1,2-dimethoxyethane for LC analysis, the peaks being recorded at 352 nm.

Reductive amination is widely used as a means of introducing chromophoric or fluorescent groups into sugars. Aldose sugars are converted to acyclic imines by reaction with the amine in the presence of an acid catalyst, and the imines produced are reduced to secondary amines by treatment with sodium cyanoborohydride. The formation of multiple peaks is prevented by this method of derivatization. Reductive amination with 4-aminobenzoic acid ethyl ester (ABEE) is very effective in increasing the sensitivity of detection of sugars in reversed-phase high-performance liquid chromatography (HPLC). The UV absorption maximum for such derivatives is at 229 nm, but the 254 nm of a standard UV photometer is often adequate for their detection. There is another absorbance peak at 304 nm. The derivatization reagent is usually prepared by mixing ABEE ( $470 \mu\text{g} \mu\text{l}^{-1}$ ), sodium cyanoborohydride ( $100 \mu\text{g} \mu\text{l}^{-1}$ ), and acetic acid (11.7%, v/v) in warm methanol. For analysis of sugars at picomolar level this reagent ( $40 \mu\text{l}$ ) is heated with the sample at 80°C for 1 h in a sealed tube. The reaction mixture is then cooled to room temperature and vortexed with a 1:1 mixture of water and chloroform (0.4 ml) to remove excess reagent. The upper, mainly aqueous, layer is used in HPLC.

Reductive amination of aldoses with chiral L-(–)- $\alpha$ -methylbenzylamine in the presence of sodium cyanoborohydride, carried out overnight at room temperature, followed by acetylation with acetic anhydride in pyridine (100°C, 1 h) gives diastereoisomeric 1-(N-acetyl- $\alpha$ -methylbenzylamino)-1-deoxyalditol acetates. LC of these derivatives on silica gel permits resolution of enantiomeric pairs, which are detected by UV photometry at 230 nm.

### For Fluorimetric Detection

Sensitivity of detection in LC of sugars is greatly enhanced by the introduction of a fluorescent group, thus permitting the use of fluorimetric detection. A reagent frequently employed for this purpose is 5-dimethylaminonaphthalene-1-sulfonyl hydrazine (dansyl hydrazine). The reagent solution (1%, m/v, in

ethanol or 5% in acetonitrile) is added to the sugars in the presence of trichloroacetic acid as a catalyst (0.5%, m/v, in ethanol, 10% in acetonitrile) and the mixture is heated at 50°C for 90 min. The derivatized sugars are detectable at picomolar levels by fluorimetry (excitation wavelength 350 nm, emission 500 nm). More recently, detection at subpicomolar levels has been reported in LC of sugars following reduction to the alditols and reaction with 2-naphthoylimidazole, with 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) as a catalyst. The pernapthoates are detected down to levels of  $\sim 0.1 \text{ pmol}$  by fluorimetry (excitation wavelength 234 nm, emission 374 nm).

Reductive amination is probably the method most widely used today for the introduction of fluorescent groups into sugars prior to analysis by chromatographic or electrophoretic methods. The fluorescent derivatives thus produced can also be detected by UV absorption, and this is sometimes preferred where the analytes are present at higher concentrations as, in contrast to fluorimetric detectors, the response of UV detectors remains linear over a wide range. However, if only trace amounts are to be analyzed, fluorimetry is the method of choice, as in most cases the derivatives are detectable at picomolar or even femtomolar levels under these conditions.

The reagent that has been most used for fluorogenic labeling of sugars and, especially, the complex oligosaccharides produced in degradative studies of glycoconjugates is 2-aminopyridine. Reductive amination with this reagent can be applied to amino sugars and sialic acids as well as neutral sugars, and this permits simultaneous analysis of, for example, the monosaccharide components of glycoproteins by reversed-phase HPLC with fluorimetric detection (excitation wavelength 320 nm, emission 400 nm) in the concentration range 10 pmol to 10 nmol. Glycoconjugates have been analyzed by this method using samples of only 100–200 pmol (glycoproteins) or 1–2  $\mu\text{g}$  (glycolipids). The reagent recommended by Hase, Takemoto, and co-workers is an aqueous solution (pH 6.2) of 2-aminopyridine (4.5%, m/v) containing hydrochloric acid (3.6%, v/v), which remains stable for at least a year if stored at  $-20^\circ\text{C}$ . The reagent ( $5 \mu\text{l}$  for analysis at picomolar level) is added to the dry sample and the mixture is heated at 100°C for 15 min in a sealed tube. The reducing agent, a freshly prepared aqueous solution of sodium cyanoborohydride ( $20 \text{ mg ml}^{-1}$ ,  $2 \mu\text{l}$ ) is then added, the tube is resealed, and heating is continued at 90°C for 8 h. Before HPLC analysis the excess reagents and by-products are removed by size-exclusion chromatography on a column of low porosity, with a solution containing a volatile salt

(such as ammonium acetate) as eluent. The 2-amino-pyridyl derivatives can also be used in nuclear magnetic resonance and mass spectrometry (MS) studies.

In recent years, several new fluorogenic reagents have been introduced for labeling of sugars and other carbohydrates by reductive amination prior to reversed-phase HPLC, often coupled to MS, since these derivatives give highly characteristic mass spectra. The derivatization process has been simplified by reaction of the analytes with the amine and the reducing agent (the borane–dimethylamine complex is now sometimes used instead of sodium cyanoborohydride) in a single step, as in the case of the reaction with ABEE mentioned before. For example, the reagent recommended by Anumula for reductive amination with 2-aminobenzoic acid (anthranilic acid) consists of a mixture of anthranilic acid (30 mg ml<sup>-1</sup>) and sodium cyanoborohydride (20 mg ml<sup>-1</sup>) in methanol containing acetic acid (2%, v/v). The sample is heated with this reagent at 80°C for 1 h and the products are purified by vortexing with the (mainly aqueous) mobile phase used in HPLC. Fluorimetric detection (excitation wavelength 230 nm, emission 425 nm) permits detection of the sugar derivatives (both neutral and amino sugars in glycoprotein hydrolysates) at femtomolar levels, so that analyses can be performed with glycoprotein samples of less than 1 µg.

Other reagents used for this purpose include 2-aminoacridine (AMAC), which gives acridone derivatives that are uncharged at the pH used in HPLC and are therefore suitable for the separation of both neutral and acidic oligosaccharides. Fluorimetry (excitation wavelength 428 nm, emission 525 nm) makes possible detection at subpicomolar levels. Reductive amination with 8-aminonaphthalene-1,3,6-trisulfonic acid (ANTS) has also been widely adopted recently as a technique for derivatization of sugars before reversed-phase HPLC or capillary electrophoresis. These derivatives can be detected by UV at 300 nm but fluorimetry (excitation wavelength 370 nm, emission 515 nm) allows detection to levels of ~0.2 pmol. The technique has been applied in the analysis of small amounts of sugar chains from glycoproteins.

The specificity and sensitivity of some of the detection methods discussed in this section are summarized in Table 1.

## Postcolumn Derivatization

The rapid development of HPLC/MS, for which the derivatives mentioned in the previous section are very suitable, has led to decreased use of postcolumn derivatization in LC of carbohydrates. Nevertheless, such methods do have certain advantages, apart from

**Table 1** Specificity and sensitivity of precolumn derivatization methods for sugars

<i>Method</i>	<i>Specificity</i>	<i>Detection limit</i>
Benzoylation	All carbohydrates	1–10 nmol
4-Nitrobenzoylation	All carbohydrates	0.1–1 nmol
Formation of 2,4-dinitro-phenylhydrazones	Reducing sugars	50 pmol
Formation of dansyl-hydrazones	Reducing sugars	3–5 pmol <sup>a</sup>
As alditol pernapthoates	Reducing sugars	0.1 pmol <sup>a</sup>
Reductive amination	Reducing sugars	
With ABEE		0.5 nmol
With 2-aminopyridine		10 pmol <sup>a</sup>
With anthranilic acid		100 fmol <sup>a</sup>
With ANTS or AMAC		0.2 pmol <sup>a</sup>

<sup>a</sup>With fluorimetric detection.

the obvious one of direct injection of sugars into the chromatograph: it has been noted that, with careful standardization, the formation of artifacts is of minor significance and the derivatization reactions do not necessarily have to be completed or fully defined. Thus, postcolumn derivatization methods with spectrophotometric or fluorimetric detection remain useful in carbohydrate analysis.

The various reactions applied for this purpose can be classified into the same main types as the chromogenic reagents used to visualize sugars in thin-layer chromatography. In addition, there are certain reactions of other types that have been adapted for LC detection systems. The different kinds of reaction used are described and exemplified below.

## Production of Furfural Derivatives

The classical methods for quantitative analysis of carbohydrates were based mainly on the formation of colored products on treatment of the analytes with strong mineral acids and condensation of the resulting 2-furaldehyde (furfural) derivatives with a suitable chromogenic reagent (usually a phenol or aromatic amine). Examples include the well-known procedures involving the use of sulfuric acid with phenol, orcinol, cysteine, or anthrone.

The phenol–sulfuric acid method involves treatment of the solution containing the carbohydrate analyte (5–40 µg ml<sup>-1</sup>) with an aqueous solution of phenol (5%, m/v) and then concentrated sulfuric acid: a characteristic yellow color is produced, with an absorption maximum at 490 nm for hexoses, 480 nm for pentoses, deoxy sugars, and uronic acids. Only amino sugars do not react. Oligo- and polysaccharides are hydrolyzed to their constituent monosaccharides by the addition of concentrated

sulfuric acid, which generates a great deal of heat, and therefore they can also be analyzed in this way. This long-established method has proved very effective in manual analyses of carbohydrates. However, adaptation of the procedure to automation in LC systems poses problems due to the excessive heat and pulsing produced when the acid is mixed with the aqueous effluent from the column.

The orcinol–sulfuric acid method, in which the recommended reagent is a solution of orcinol (0.1%, m/v) in diluted sulfuric acid (70%, v/v), was the first to be adapted for use in an automated carbohydrate analyzer. After mixing with the reagent, the column effluent is heated at 95°C for 15 min. Absorbance is usually measured at 420–425 nm; however, measurement at two wavelengths (420 and 510 nm) has been recommended where deoxy sugars or uronic acids are present.

The cysteine–sulfuric acid method, with a reagent consisting of a solution of cysteine (0.07%, m/v) in 86% (v/v) sulfuric acid, has also been successfully automated. The reagent is mixed with the column effluent in a ratio of 5:1 (v/v) and this is followed by heating at 97°C for 3 min. The absorption maximum of the colored derivative ranges from 390 nm (pentoses and uronic acids) through 400 nm (6-deoxyhexoses) to 412–414 nm (hexoses) and therefore the measurement of absorbance at more than one wavelength in this range has been recommended.

Anthrone (9,10-dihydro-9-oxo-anthracene) reacts with most carbohydrates in concentrated sulfuric acid to produce a characteristic blue color, with an absorption maximum at 625 nm. This reaction has been automated and was once widely used in analysis of dextran fractions emerging from size exclusion columns. However, as in the case of the phenol–sulfuric acid method, the presence of concentrated acid necessitates special precautions (such as the inclusion of a pulse suppressor) and for this reason, and also because of the instability of the reagent, the method is seldom used today except in manual analysis.

In addition to the reactions mentioned above, there are others of this type that are used in detecting and determining specific classes of carbohydrates. The carbazole–sulfuric acid method has been much used in analysis of uronic acids and acidic polysaccharides, which give a weaker response than neutral sugars in the other methods. The presence of borate ions decreases interference by proteins and salts, increases color yield, and reduces the time required for color development. In a recommended procedure for manual analysis, the sample solution (containing 0.02–0.2  $\mu\text{mol}$  of uronic acid) is mixed (1:5, v/v) with a solution of sodium tetraborate (25  $\text{mmol l}^{-1}$ ) in concentrated sulfuric acid and the mixture is heated at 100°C for 10 min. After rapid

cooling to 4°C the reaction mixture is mixed (30:1, v/v) with a solution of carbazole (0.125%, m/v) in ethanol. Heating at 100°C is continued for a further 15 min and, after cooling to room temperature, the absorbance at 530 nm is measured. The procedure has been successfully automated.

As there is some interference from hexoses and, to a lesser extent, pentoses in this analytical method, it has been recommended that for analysis of uronic acids in the presence of neutral sugars in appreciable proportion the carbazole should be replaced by 3-hydroxydiphenyl. In this case the sample (containing 0.5–20  $\mu\text{g}$  of uronic acid) is mixed (1:6, v/v) with a solution of sodium tetraborate (12.5  $\text{mmol l}^{-1}$ ) in concentrated sulfuric acid and, after heating at 100°C for 5 min and cooling to 4°C, 20  $\mu\text{l}$  of the reagent (3-hydroxydiphenyl, 0.15%, m/v, in 0.5% NaOH) is added. Absorbance at 520 nm should be measured within 5 min of mixing. As neutral sugars at high concentration produce a pale pink color with  $\text{Na}_2\text{B}_4\text{O}_7\text{--H}_2\text{SO}_4$  at 100°C, the absorbance of a blank, in which the reagent is replaced by 20  $\mu\text{l}$  of 0.5% NaOH, should be subtracted from the reading. This method is more sensitive and specific than the carbazole method. The molar absorptivity of the chromophore produced by mannuronic acid is  $\sim 50\%$  of that of the product from glucuronic acid, but the other uronic acids commonly encountered, i.e., galacturonic and iduronic acids, give absorbances only slightly below that given by glucuronic acid.

The resorcinol–hydrochloric acid method is specific for ketoses or other carbohydrates containing keto groups, such as sialic acids. In an automated method that has proved useful in chromatographic analysis of commercial syrups containing fructose in the presence of glucose and other aldoses, the column effluent is mixed 1:5 (v/v) with the reagent (resorcinol, 0.05%, m/v, in concentrated HCl) and the mixture 20:1 (v/v) with an aqueous solution of 1,1-diethoxyethane (0.05%, m/v) and heated at 95°C for 3 min. The absorbance of the product is measured at 550 nm.

The necessity for corrosive acids in the analytical system is a serious drawback of all methods of this type when utilized in automated LC systems. Nevertheless, postcolumn derivatization based on these principles remains important because it allows the simultaneous analysis of reducing and nonreducing carbohydrates. Most of the noncorrosive reagents that are used under neutral or mildly alkaline conditions (discussed next) are applicable only to reducing sugars.

### Reactions Involving Reduction by Sugars

Some metallic cations, such as iron(III) and copper(II), are easily converted to lower oxidation states

in the presence of reducing sugars. Therefore, if the system contains ligands that are incapable of interacting with the cations in the higher oxidation state but can bind to the reduced metal ions to form chelates having absorption maxima in the UV or visible region, spectrophotometric determination of the reducing sugars is possible. A prime example is the copper–bicinchoninate method, which is based on the formation of a deep lavender complex between Cu(I) and 2,2′-bicinchoninate. In this method an aqueous solution containing  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$  (0.1%, m/v), with aspartic acid (0.37%, m/v) as a masking agent, is mixed 1:1 (v/v) with a solution of disodium 2,2′-bicinchoninate (0.2%, m/v) in 0.38% (m/v)  $\text{Na}_2\text{CO}_3 \cdot 10\text{H}_2\text{O}$  ( $0.133 \text{ mol l}^{-1}$ ) to produce the reagent. In an automated system the column effluent is mixed (2:1, v/v) with the reagent and, after heating at  $100^\circ\text{C}$  for 5 min, the absorbance of the colored product is measured at 562 nm. The detection limits for most sugars, including amino sugars and uronic acids as well as neutral sugars, are in the range of 100–200 pmol.

Another method based on the reducing power of sugars is that in which 4-anisyltetrazolium chloride (tetrazolium blue) is reduced to diformazan in alkaline solution. This is not applicable to systems using aqueous eluents, in which the product precipitates, but can be used with eluents consisting largely of ethanol or acetonitrile, where it is freely soluble. An automated system has been described in which the reagent, a solution of tetrazolium blue (0.2%, m/v) in NaOH ( $0.18 \text{ mol l}^{-1}$ ), is mixed (2:5, v/v) with the column effluent and heated at  $80^\circ\text{C}$  for 1 min. The absorbance of the product is measured at 520 nm. The method is highly sensitive, detection limits being of the order of 100 pmol for neutral sugars.

### Periodate Oxidation

Carbohydrates containing vicinal diol systems are oxidized by periodate, and there are analytical methods that are based on reaction of the products with chromogenic reagents. The classic example is the Hantsch reaction, in which the formaldehyde produced on oxidation with sodium metaperiodate is reacted with 2,4-pentanedione in the presence of ammonia to give a pyridine derivative with an absorption maximum at 412 nm. However, this is applicable mainly to alditols; aldoses give a much lower response. A generally more useful method involves monitoring the elution of carbohydrate in aqueous HPLC systems by continuous measurement of the periodate consumption of the effluent, on the basis of the decrease in UV absorbance that accompanies reduction of periodate to iodate. A reagent consisting of  $\text{NaIO}_4$  ( $1 \text{ mmol l}^{-1}$ ) in borate buffer ( $0.5 \text{ mol l}^{-1}$ ,

pH 8.6) has been recommended, with heating at  $90^\circ\text{C}$  for 5 min. The absorbance maximum for iodate is at 223 nm, but the loss in sensitivity at the standard 254 nm of simple photometers is marginal. The method is applicable not only to alditols but also to sugars, both reducing and nonreducing, with detection limits in the nanomolar range.

### Condensation with Hydrazino Compounds

Reactions of reducing carbohydrates with hydrazides of benzoic acid derivatives in strong alkali give a yellow color, believed to be due to the formation of the anionic forms of carbohydrate hydrazones. This affords a sensitive analytical method, which can also be applied to nonreducing oligosaccharides if they are first hydrolyzed to reducing sugars by passage of the analyte solution at elevated temperatures through a cation-exchange resin in the hydrogen form. In a detection method for sugars in aqueous media a reagent prepared by mixing a solution of 4-hydroxybenzoic acid hydrazide (5%, m/v) in hydrochloric acid ( $0.5 \text{ mol l}^{-1}$ ) 10:57 (v/v) with NaOH ( $0.75 \text{ mol l}^{-1}$ ) is mixed 1:2 (v/v) with the column effluent and, after heating at  $95^\circ\text{C}$  for 40 s, the absorbance of the product is measured at 410 nm. The amino analog of this reagent, 4-aminobenzoic acid hydrazide (ABH), has proved to be even more sensitive to sugars. The reagent, prepared by mixing a stock solution, of the same concentration and HCl content as that given above for the hydroxy compound, with  $2.4 \text{ mol l}^{-1}$  NaOH in a ratio of 1:2 (v/v), must be clarified by ultrafiltration and sonication before use. The other analytical conditions are the same. Detection limits for reducing sugars, and for sucrose if the catalytic resin column is used, are of the order of 40 pmol. This has been shown to be one of the most sensitive of all the postcolumn derivatization methods for sugar analysis.

### Specific Methods for Amino Sugars

Amino sugars give lower responses or do not react at all in many of the analytical methods that have been described. Analyses for these sugars are generally based on the reaction of the amino groups with ninhydrin, the reagent used in the standard amino acid analyzer. This type of system may therefore be applied to the quantitative determination of amino sugars, with detection at 570 nm. Greater sensitivity is possible, however, if the ninhydrin reagent is replaced by 2-phthalaldehyde, which is used in alkaline solution in the presence of ethanethiol or 2-mercaptoethanol. This commercially available reagent produces fluorescent 1-alkylthio-2-alkyl-substituted isoindoles with amino sugars, which permits fluorimetric detection



(excitation wavelength 340 nm, emission 455 nm), with detection at levels down to 3 nmol. Amino acids interfere, but are readily separated from the sugar derivatives by cation-exchange chromatography.

### Reactions with Fluorogenic Reagents

The use of fluorogenic reagents in postcolumn derivatization of sugars can result in detection at picomolar levels. Among the reagents that have been successfully used for this purpose are simple aliphatic amines. In a weakly alkaline medium at elevated temperatures reducing sugars react with ethylenediamine to give fluorescent products resulting from isomerization of the sugars (Lobry de Bruyn–van Ekenstein reaction) and reaction of the intermediates with excess amine. This was first suggested as a sensitive detection method for sugars analyzed by borate anion-exchange chromatography, ethylenediamine ( $7.5 \text{ mmol l}^{-1}$ ) being added postcolumn to the mobile phase (borate buffer,  $0.7 \text{ mol l}^{-1}$ , pH 8.6) used in this chromatographic method. The optimal

reaction temperatures were found to be 110–120°C for pentoses, 120–130°C for hexoses (including hexuronic acids), and 140–145°C for heptoses and simple oligosaccharides (cleaved to monosaccharides by alkaline hydrolysis under the conditions used). Fluorimetry (excitation wavelength 360 nm, emission 455 nm) gave detection limits in the range 100–400 pmol for most reducing sugars. The ethylenediamine reagent can be used with any aqueous eluent under the appropriate pH conditions (which can be adjusted postcolumn) but is not satisfactory with organic eluents. As the products of derivatization are oxidizable, the reagent can also be used with electrochemical detection at a glassy carbon electrode, under which conditions even higher sensitivity is possible.

Ethanolamine is another alkylamine that reacts with reducing sugars to give fluorescent products. An aqueous solution containing boric acid and ethanolamine (both 2%, m/v) is used as the reagent and is mixed 1:3 (v/v) with the column effluent, at a temperature of 150°C, with fluorimetric detection of the products (excitation wavelength 357 nm,

**Table 2** Specificity and sensitivity of postcolumn derivatization methods for sugars

Method	Specificity	Detection limit <sup>a</sup>
<i>Production of furfural derivatives</i>		
Anthrone–H <sub>2</sub> SO <sub>4</sub>	Neutral sugars	100 nmol
Carbazole–H <sub>2</sub> SO <sub>4</sub>	Uronic acids	25 nmol
Cysteine–H <sub>2</sub> SO <sub>4</sub>	Neutral sugars, uronic acids	25–50 nmol
Orcinol–H <sub>2</sub> SO <sub>4</sub>	Neutral sugars, uronic acids	20–40 nmol
Phenol–H <sub>2</sub> SO <sub>4</sub>	Neutral sugars, uronic acids	20–40 nmol
Resorcinol–HCl	Ketoses	40–50 nmol
<i>Use of reducing power</i>		
Cu <sup>2+</sup> -2,2'-bicinchoninate	Reducing sugars	100–200 pmol
Tetrazolium blue	Neutral reducing sugars	50–200 pmol
<i>Periodate oxidation</i>		
NaIO <sub>4</sub>	Sugars with vicinal diol systems	1–10 nmol
<i>Condensation with hydrazides</i>		
4-Hydroxybenzoic acid hydrazide	Neutral reducing sugars	20 nmol
4-Aminobenzoic acid hydrazide	Neutral reducing sugars	40 pmol
<i>Specific for amino sugars</i>		
Ninhydrin	Sugars with primary amino groups	25 nmol
2-Phthaldehyde 3 nmol (fluorimetric detection)		3 nmol
<i>With fluorogenic reagents</i>		
Ethanolamine–boric acid (fluorimetric detection)	Neutral reducing sugars	1–5 nmol
Ethylenediamine	Neutral reducing sugars	
Fluorimetric detection		100–400 pmol
Amperometric detection		1 pmol
2-Cyanoacetamide	All reducing aldoses	Neutral sugars
UV detection		0.1–1 nmol; uronic acids 0.2–0.8 nmol; aminodeoxy sugars 2–3 nmol; acetamidodeoxy-hexoses 400 pmol
Fluorimetric detection		0.25 nmol; uronic acids 50–120 pmol; aminodeoxy sugars 2–3 nmol

<sup>a</sup>UV–visible spectrophotometric detection, except where otherwise indicated.

emission 436 nm). This reagent has been used successfully with both aqueous and acetonitrile-rich eluents. Detection limits are below 5 nmol for monosaccharides, but higher for oligosaccharides, the response decreasing sharply with increasing degree of polymerization and hence reducing power. Sensitivity for oligosaccharides (nonreducing as well as reducing) can be improved by online postcolumn hydrolysis with 4-toluenesulfonic acid prior to the reaction.

The most versatile of the fluorogenic reagents that have been used in sugar analysis is 2-cyanoacetamide, which reacts with reducing aldoses in weakly alkaline solutions at elevated temperatures to give intensely fluorescent compounds (excitation wavelength 331 nm, emission 383 nm). Like ethylenediamine, this reagent was originally employed in borate anion-exchange chromatography of sugars, where an aqueous solution (10%, m/v) was mixed 1:1 (v/v) with borate buffer (0.5–0.6 mol l<sup>-1</sup>, pH 9.0) before being added to the column effluent and heated at 100°C. The reagent may also be applied with acetonitrile-rich eluents, in which case a concentration of 5% (m/v) in potassium borate (0.1 mol l<sup>-1</sup>, pH 10.4) has been recommended, with a higher reaction temperature (135°C) to prevent the precipitation of borate salts on mixing with the organic eluent. The sensitivity is greater under these conditions (detection limit ~250 pmol for all neutral aldoses). Detection with 2-cyanoacetamide has been successfully applied not only to neutral sugars but also to uronic acids, amino sugars, and the monosaccharide constituents of glycoproteins (including acetamido-deoxyhexoses and sialic acids), as well as to alditols. The fluorescent products of the reaction of 2-cyanoacetamide with reducing aldoses are believed to be 3-cyano-2-pyridone and 3-cyano-2-pyrrolidone derivatives, and there is also a product, containing a conjugated diene system in the molecular structure, which absorbs strongly at 270 nm, so that UV detection is also sensitive. As all of these products are readily oxidizable at a glassy carbon electrode, electrochemical detection is another option with this useful analytical reagent.

The specificity and sensitivity of the reagents discussed in this section are summarized in Table 2.

**See also:** Carbohydrates: Overview; Sugars – Chromatographic Methods. **Derivatization of Analytes.** **Fluorescence:** Fluorescence Labeling. **Liquid Chromatography:** Size-Exclusion. **Spectrophotometry:** Overview; Derivative Techniques.

## Further Reading

- Anumula KR (1994) Quantitative determination of monosaccharides in glycoproteins by high-performance liquid chromatography with highly sensitive fluorescence detection. *Analytical Biochemistry* 220: 275–283.
- Churms SC (ed.) (1982, 1991) *CRC Handbook of Chromatography: Carbohydrates*, vols. I and II. Boca Raton, FL: CRC Press.
- El Rassi Z (ed.) (2002) *Carbohydrate Analysis by Modern Chromatography and Electrophoresis. Journal of Chromatography Library*, vol. 66. Amsterdam: Elsevier.
- Hase S (1996) Precolumn derivatization for chromatographic and electrophoretic analyses of carbohydrates. *Journal of Chromatography A* 720: 173–182.
- Hase S (2002) Pre- and postcolumn detection-oriented derivatization techniques in HPLC of carbohydrates. In: El Rassi Z (ed.) *Carbohydrate Analysis by Modern Chromatography and Electrophoresis. Journal of Chromatography Library*, vol. 66, pp. 1043–1069. Amsterdam: Elsevier.
- Honda S (1996) Postcolumn derivatization for chromatographic analysis of carbohydrates. *Journal of Chromatography A* 720: 183–199.
- Honda S, Kakehi K, Fujikawa K, Oka Y, and Takahashi M (1988) Mechanism of the reaction of reducing carbohydrates with 2-cyanoacetamide, used for postcolumn labelling in high performance liquid chromatography for photometric, fluorimetric, and electrochemical detection. *Carbohydrate Research* 183: 59–69.
- Kwon H and Kim J (1993) Determination of monosaccharides in glycoproteins by reverse-phase high-performance liquid chromatography. *Analytical Biochemistry* 215: 243–252.
- Takemoto H, Hase S, and Ikenaka T (1985) Microquantitative analysis of neutral and amino sugars as fluorescent pyridylamino derivatives by high-performance liquid chromatography. *Analytical Biochemistry* 145: 245–250.

## Sugars – Chromatographic Methods

**S C Churms**, University of Cape Town, Rondebosch, South Africa

## Introduction

Chromatographic analysis of samples containing mixtures of sugars is an essential operation in any laboratory concerned with carbohydrates, whether

emission 436 nm). This reagent has been used successfully with both aqueous and acetonitrile-rich eluents. Detection limits are below 5 nmol for monosaccharides, but higher for oligosaccharides, the response decreasing sharply with increasing degree of polymerization and hence reducing power. Sensitivity for oligosaccharides (nonreducing as well as reducing) can be improved by online postcolumn hydrolysis with 4-toluenesulfonic acid prior to the reaction.

The most versatile of the fluorogenic reagents that have been used in sugar analysis is 2-cyanoacetamide, which reacts with reducing aldoses in weakly alkaline solutions at elevated temperatures to give intensely fluorescent compounds (excitation wavelength 331 nm, emission 383 nm). Like ethylenediamine, this reagent was originally employed in borate anion-exchange chromatography of sugars, where an aqueous solution (10%, m/v) was mixed 1:1 (v/v) with borate buffer (0.5–0.6 mol l<sup>-1</sup>, pH 9.0) before being added to the column effluent and heated at 100°C. The reagent may also be applied with acetonitrile-rich eluents, in which case a concentration of 5% (m/v) in potassium borate (0.1 mol l<sup>-1</sup>, pH 10.4) has been recommended, with a higher reaction temperature (135°C) to prevent the precipitation of borate salts on mixing with the organic eluent. The sensitivity is greater under these conditions (detection limit ~250 pmol for all neutral aldoses). Detection with 2-cyanoacetamide has been successfully applied not only to neutral sugars but also to uronic acids, amino sugars, and the monosaccharide constituents of glycoproteins (including acetamido-deoxyhexoses and sialic acids), as well as to alditols. The fluorescent products of the reaction of 2-cyanoacetamide with reducing aldoses are believed to be 3-cyano-2-pyridone and 3-cyano-2-pyrrolidone derivatives, and there is also a product, containing a conjugated diene system in the molecular structure, which absorbs strongly at 270 nm, so that UV detection is also sensitive. As all of these products are readily oxidizable at a glassy carbon electrode, electrochemical detection is another option with this useful analytical reagent.

The specificity and sensitivity of the reagents discussed in this section are summarized in Table 2.

**See also:** Carbohydrates: Overview; Sugars – Chromatographic Methods. **Derivatization of Analytes.** **Fluorescence:** Fluorescence Labeling. **Liquid Chromatography:** Size-Exclusion. **Spectrophotometry:** Overview; Derivative Techniques.

## Further Reading

- Anumula KR (1994) Quantitative determination of monosaccharides in glycoproteins by high-performance liquid chromatography with highly sensitive fluorescence detection. *Analytical Biochemistry* 220: 275–283.
- Churms SC (ed.) (1982, 1991) *CRC Handbook of Chromatography: Carbohydrates*, vols. I and II. Boca Raton, FL: CRC Press.
- El Rassi Z (ed.) (2002) *Carbohydrate Analysis by Modern Chromatography and Electrophoresis. Journal of Chromatography Library*, vol. 66. Amsterdam: Elsevier.
- Hase S (1996) Precolumn derivatization for chromatographic and electrophoretic analyses of carbohydrates. *Journal of Chromatography A* 720: 173–182.
- Hase S (2002) Pre- and postcolumn detection-oriented derivatization techniques in HPLC of carbohydrates. In: El Rassi Z (ed.) *Carbohydrate Analysis by Modern Chromatography and Electrophoresis. Journal of Chromatography Library*, vol. 66, pp. 1043–1069. Amsterdam: Elsevier.
- Honda S (1996) Postcolumn derivatization for chromatographic analysis of carbohydrates. *Journal of Chromatography A* 720: 183–199.
- Honda S, Kakehi K, Fujikawa K, Oka Y, and Takahashi M (1988) Mechanism of the reaction of reducing carbohydrates with 2-cyanoacetamide, used for postcolumn labelling in high performance liquid chromatography for photometric, fluorimetric, and electrochemical detection. *Carbohydrate Research* 183: 59–69.
- Kwon H and Kim J (1993) Determination of monosaccharides in glycoproteins by reverse-phase high-performance liquid chromatography. *Analytical Biochemistry* 215: 243–252.
- Takemoto H, Hase S, and Ikenaka T (1985) Microquantitative analysis of neutral and amino sugars as fluorescent pyridylamino derivatives by high-performance liquid chromatography. *Analytical Biochemistry* 145: 245–250.

## Sugars – Chromatographic Methods

**S C Churms**, University of Cape Town, Rondebosch, South Africa

## Introduction

Chromatographic analysis of samples containing mixtures of sugars is an essential operation in any laboratory concerned with carbohydrates, whether

sugars as such or carbohydrates of higher molar mass (polysaccharides or glycoconjugates and their depolymerization products). Characterization of the latter requires, *inter alia*, complete degradation to the constituent monosaccharides in order to determine their nature and proportions, for which purpose chromatographic methods are crucial. Both gas chromatography (GC) and liquid chromatography (LC) are widely used in sugar analysis, but during the past decade LC has been increasingly preferred over GC. This is because derivatization prior to analysis is not usually required in LC, unless it is necessary to enhance sensitivity of detection by introduction of a chromophoric or fluorescent group. The major area of growth in chromatographic analysis of sugars in recent years has undoubtedly been in the development of what has become known as high-performance anion-exchange chromatography (HPAEC), a form of ion chromatography that is applied specifically to analysis of sugars and other carbohydrates. This is now the method of choice in many laboratories worldwide.

This article will first give an overview of the GC methods most frequently applied to sugar analysis, with emphasis on significant recent developments, such as the automation of derivatization of sugars to alditol acetates, and the potential for enhanced sensitivity in GC coupled to mass spectrometry (MS) afforded by the introduction of tandem MS (GC-MS/MS) into this field. In view of the growing preference for LC, however, the focus will be mainly on this form of chromatography, and especially HPAEC. In planar chromatography, the more rapid thin-layer chromatography (TLC) has now entirely superseded the classical technique of paper chromatography in qualitative analysis of mixtures of sugars, all the more since the widespread adoption of high-performance TLC (HPTLC) plates. A brief overview of the most effective TLC systems for this purpose, and some useful detection reagents, will be given in the final section of this article.

## Gas Chromatography

Sugars, being nonvolatile, must be converted to volatile derivatives before GC analysis is possible. During the 1970s and 1980s, many different methods of derivatization were explored, but most of the novel procedures were not pursued further. The derivatization methods that have been generally adopted as standard procedures for GC analysis of sugars of various types include conversion to acetylated alditols or aldononitriles, oximation followed by acetylation, trimethylsilylation or trifluoroacetylation,

and trimethylsilylation of the mixtures of methyl glycosides obtained on methanolysis of carbohydrates of higher molar mass. These procedures, together with some promising new techniques, are briefly discussed below.

Unless the anomeric centers of sugars are eliminated prior to GC, the chromatogram will be complicated by the production of multiple peaks. Therefore, for analysis of complex mixtures, the sugars are often reduced to the corresponding alditols by treatment with sodium borohydride (or sometimes borodeuteride). After acidification with acetic acid to destroy excess borohydride, and removal of the resulting borate by repeated evaporation with methanol, the residue is acetylated by reaction with acetic anhydride, the acetate remaining in the mixture serving as a basic catalyst. This procedure, together with the steps involved in removal of reagents and recovery of products, is relatively simple but time consuming, which is probably one of the main reasons for the shift in preference from GC to high-performance liquid chromatography (HPLC) analysis of sugars. Therefore, the recent development of a fully automated system for the preparation of alditol acetate derivatives can be considered to be an important contribution to the field. The computer-controlled instrument described by Alvin Fox and co-workers enables the multistage process to be performed sequentially without manual intervention, and also makes possible the simultaneous processing of several samples in a manifold of glass/polytetrafluoroethylene (PTFE) reaction chambers, seated in a movable heating block the temperature of which is adjusted automatically. Electrically driven solenoid valves are connected online to the sample manifold, a set of solvent valves controlling the input of solvent, reagent, and/or nitrogen gas to each chamber, and a set of gas valves controlling output to atmosphere or vacuum. Closure of all valves allows the samples to be sealed into closed chambers. Use of this apparatus permits the reduction and acetylation of samples to be performed overnight, and only the final post-derivatization cleanup (removal of excess acetic anhydride and extraction of the derivatized samples into chloroform or dichloromethane) requires manual operations. The advent of automated systems of this type, which remove the tedium from derivatization, may well lead to renewed interest in GC analysis of sugars.

An alternative method of eliminating the anomeric center in a sugar is conversion to the aldononitrile or ketoxime, and this reaction, followed by acetylation of the remaining hydroxyl groups, has also been much used as a derivatization procedure for GC analysis. The standard method, which is less time

consuming than the classical procedure for conversion of sugars to alditol acetates, involves heating the sample at 70°C with hydroxylamine hydrochloride in the presence of pyridine for 20 min, then addition of acetic anhydride and heating for a further 20 min under the same conditions. The products are extracted into chloroform for GC analysis. This derivatization method is applicable to both aldoses and ketoses; the latter are analyzed as acetylated ketoxime derivatives while, for the aldoses, the oximes formed initially are dehydrated to aldonitriles during the acetylation step. Not only neutral sugars but also aminodeoxy- and acetamidodeoxyhexoses can be derivatized in this way, but longer reaction times are required for the amino sugars. Each aldose gives a single, unique, linear derivative, but each ketose yields two products, corresponding to the geometrical isomers that differ in the configuration of the O-acetyl oxime group.

Where amino sugars are present together with neutral sugars in the mixture to be analyzed (as in hydrolysates from glycoproteins), derivatization to O-methyloximes prior to acetylation is generally preferred, as the time required for oximation is only 20–25 min for both neutral and amino sugars. In this case, the reagent used is O-methylhydroxylamine, with 1-dimethylamino-2-propanol or 4-(dimethylamino)-pyridine as a catalyst. The O-methyloxime acetates derived from the amino sugars are well separated from the neutral sugar derivatives on GC, in run times of 45 min or less. Double peaks are produced in all cases.

In recent years, trimethylsilylation has been increasingly preferred over acetylation following oximation of sugars in derivatization for GC analysis. The trimethylsilyl (TMS) oximes are more volatile, and therefore the method is applicable to sugars of all types, including acidic sugars. This GC technique has proved useful in the analysis of complex mixtures of neutral and acidic sugars (for example, those produced on hydrolysis of industrial gums) and also permits simultaneous analysis of neutral sugars, polyols, acidic sugars, and other components, such as carboxylic acids and amino acids, in natural matrices (e.g., fruit).

There has been some interest in the use of trifluoroacetylated oximes as volatile derivatives for GC analysis of sugars, as analysis can be performed at relatively low temperatures and the sensitive electron capture detector can be used. However, these derivatives are unstable in the presence of moisture, and it is therefore difficult to remove the derivatization reagent (trifluoroacetic acid anhydride); compounds other than sugars may thus be converted to halogenated derivatives, resulting in increased

background. More recently, introduction of halogen atoms by derivatization of sugars to O-pentafluorobenzyl oximes, followed by acetylation for GC analysis, has been recommended. These derivatives are stable and give highly characteristic mass spectra. They are therefore very useful in GC–MS analysis, particularly with MS in negative-ion chemical ionization mode.

Capillary GC of the TMS derivatives of the methyl glycosides and methyl glycoside methyl ethers produced on methanolysis of polysaccharides or glycoconjugates is a highly sensitive method of analysis, allowing the determination of individual components at picomolar levels. The derivatives from all the methyl glycosides in a complex mixture, including those from neutral, amino, and acidic sugars, can be separated in a single run, with relatively short analysis times. The method has thus proved particularly useful in analysis of the constituents of glycoconjugates and of acidic polysaccharides such as industrial gums. An interesting application has been the use of GC–MS of the TMS methyl glycosides from methanolysates in characterizing the plant gums and other vegetable substances present in objects of archaeological interest. Multiple peaks, corresponding to anomeric pairs and pyranose or furanose ring forms, are given by each sugar derivative, but the characteristic patterns can aid in identification, which in recent years has been greatly facilitated by the accumulation of a large body of GC–MS data for TMS methyl glycosides.

Recently, a novel method for rapid, sensitive analysis of the sugar constituents of glycoconjugates, by GC analysis of the heptafluorobutyrate derivatives of the methyl glycosides produced on methanolysis, was described by Zanetta and co-workers. Acylation with heptafluorobutyric anhydride derivatizes in a single step not only hydroxyl but also amino groups, forming products that do not interact strongly with the methyl siloxane stationary phases generally used in GC of apolar derivatives. Thus, amino sugars can be analyzed together with other sugar components within a short time. This applies also to acidic sugars, and therefore the method can be used in analysis of glycosaminoglycans.

Recommended conditions for GC analysis of the various sugar derivatives that have been discussed in this section are given in **Table 1**.

### Advances in GC–MS

GC–MS analysis of sugars has been greatly facilitated by the availability of databases containing *m/z* values and relative intensities of selected ions among those produced by the various derivatives used in

**Table 1** Conditions recommended for GC analysis of sugars

<i>Derivative</i>	<i>Column type</i>	<i>Phase</i>	<i>Temperature (°C)</i>	<i>Gas; flow rate (ml min<sup>-1</sup>)</i>
Alditols, acetylated	Packed	OV-225, 3% on Chromosorb W-HP (80-100 mesh)	210	He; 40
	Glass capillary	Silar 10C	190 (4 min); 190→230 at 4°C min <sup>-1</sup> ; 230 (8 min)	H <sub>2</sub> ; 9
	Glass capillary	SP-2340	185 (10 min); 185→220 at 3°C min <sup>-1</sup> ; 220 (5 min); 220→235 at 15°C min <sup>-1</sup> ; 235 (30 min); for separation of amino and neutral sugar derivatives	He; 5.5
	Fused-silica capillary	OV-1701 (bonded phase)	140 (0.5 min); 140→195 at 25°C min <sup>-1</sup> ; 195→250 at 6°C min <sup>-1</sup> ; 250 (8 min); for separation of amino and neutral sugar derivatives	He; 0.6
Aldononitriles and ketoximes, acetylated	Packed	OV-17, 2% on Chromosorb W-HP (80–100 mesh)	130→300 at 5°C min <sup>-1</sup>	N <sub>2</sub> ; 22
	Packed	Poly(neopentyl glycol succinate) 3% on Chromosorb W (60–80 mesh)	140→250 at 3°C min <sup>-1</sup>	N <sub>2</sub> ; 32
	Fused-silica capillary	OV-1	175 (4 min); 175→260 at 4°C min <sup>-1</sup> ; 260 (5 min); to separate amino from neutral	He; 0.5
O-Methyl oximes, acetylated	Fused-silica capillary	OV-1	As for acetylated aldononitriles; see above	He; 0.5
O-Methyl oximes, trimethylsilylated	Fused-silica capillary	SP-2100	180	He; 1
O-Pentafluorobenzyl oximes, acetylated	Fused-silica capillary	CP-Sil 88	140 (4 min); 140→240 at 10°C min <sup>-1</sup> ; 240 (10 min); separates amino from neutral	He; 3
Oximes, trimethylsilylated	Fused-silica capillary	DB-5 (bonded phase)	60 (2 min); 60→155 at 13°C min <sup>-1</sup> ; 155 (10 min); 155→250 at 13°C min <sup>-1</sup> ; 250 (12 min); to separate mono-, di-, and trisaccharides	NA
	Fused-silica capillary	DB-5	60 (2 min); 60→120 at 20°C min <sup>-1</sup> ; 120→155 at 6°C min <sup>-1</sup> ; 155 (10 min); 155→250 at 13°C min <sup>-1</sup> ; 250 (12 min); 250→330 at 20°C min <sup>-1</sup> ; 330 (10 min); to separate sugars, alditols, sugar acids, and amino acids	NA
Methyl glycosides, trimethylsilylated	Fused-silica capillary	CP-Sil 5	140 (2 min); 140→260 at 8°C min <sup>-1</sup> ; separates neutral, amino, and acidic derivatives	He; 1
	Fused-silica capillary	DB-5	150→220 at 2°C min <sup>-1</sup> ; separates neutral and acidic derivatives	N <sub>2</sub> ; 1
	Fused-silica capillary	DB-1 (bonded phase)	120→145 at 1°C min <sup>-1</sup> ; 145→180 at 0.9°C min <sup>-1</sup> ; 180→230 at 50°C min <sup>-1</sup> ; separates derivatives from all components of plant cell-wall polysaccharides, including 2-keto-3-deoxy sugars and branched sugars	H <sub>2</sub> ; 2
Heptafluorobutyrate of methanolysis products from glycoconjugates	Fused-silica capillary	CP-Sil 5 CB (low bleed)	90 (3 min); 90→260 at 5°C min <sup>-1</sup> ; 260 (10 min); separates neutral, amino, and acidic derivatives	He; 0.8 bar

NA = not available.

such analyses. A vast body of data has been collected for identification of sugars as their alditol acetate derivatives, and diagnostic MS data for characterization of the other derivatives mentioned above is now rapidly accumulating.

During the 1990s, tandem mass spectrometry attracted wide interest, and this technique is being used to an increasing extent in conjunction with GC for detection and identification of carbohydrates present in trace quantities in complex matrices (such as bacterial cell walls), an application in which optimal specificity and sensitivity are required. Both triple-quadrupole and ion-trap MS/MS have been used successfully in trace analysis of sugars (for example, muramic acid) that serve as bacterial markers in samples of environmental or clinical interest. Such sugars are usually analyzed as their alditol acetate derivatives. Quantitation is more accurate when a triple-quadrupole mass spectrometer is used, but sensitivity is higher with the ion-trap instrument. Specificity in detecting trace amounts of carbohydrate markers in complex matrices is greatest when MS/MS is used in multiple reaction monitoring mode. The sensitivity of GC–MS/MS in trace analysis of sugars is currently unmatched by that of MS/MS used in conjunction with any form of LC.

### Enantioselective GC of Sugars

Enantiomeric pairs can be resolved by capillary GC of sugar derivatives formed by reaction with chiral compounds: acetylated (+)-2-octyl glycosides and trimethylsilylated (–)-2-butyl glycosides are examples of derivatives that have been used for this purpose. In recent years, however, the focus has shifted toward the use of enantioselective stationary phases with sugar derivatives that are more readily available. König and co-workers, who are recognized as the leaders in this field, have produced very effective enantioselective stationary phases for carbohydrate analysis by modification of cyclodextrins through the introduction of hydrophobic groups. The objective of this modification has been both to lower the melting points of the cyclodextrins, to permit their use as liquid phases in GC, and to increase thermal stability. Tri-*O*-pentylated derivatives of  $\alpha$ - and  $\beta$ -cyclodextrins [hexakis(2,3,6-tri-*O*-pentyl) cyclomaltohexaose and heptakis(2,3,6-tri-*O*-pentyl) cyclomaltoheptaose] have been found to exhibit a high degree of enantioselectivity toward trifluoroacetylated derivatives of aldoses, alditols, and methyl glycosides, baseline resolution of enantiomers being achieved within 5–10 min on glass capillary columns coated with these phases (marketed as Lipodex A and C, respectively). These separations can

be achieved at relatively low column temperatures (80–120°C).

More recently, other modified cyclodextrins, such as heptakis(2,6-di-*O*-methyl-3-*O*-pentyl)cyclomaltoheptaose or octakis(2,6-di-*O*-methyl-3-*O*-pentyl)cyclomaltooctaose, have been used as stationary phases in enantioselective GC of permethylated derivatives of methyl glycosides and anhydroalditols. The acetylated derivatives of the mixture of 1,5-anhydro-*O*-methylgalactitols obtained on reductive cleavage of the permethylated galactan from a snail (*Helix pomatia*) were well resolved by capillary GC on a column coated with the modified cycloheptaose, and in this way the occurrence of terminal residues of L-galactose as well as the more usual D-galactose was clearly established. This is a striking demonstration of the potential of enantioselective GC in structural studies of polysaccharides.

## Liquid Chromatography

Of the several different modes of LC that have been developed over the years there are three that have been applied in analysis of sugars. These are: chromatography on polar stationary phases with a less polar mobile phase (formerly known as ‘normal-phase chromatography’ but now termed ‘hydrophilic interaction chromatography’, with the acronym HILIC), reversed-phase chromatography (RPC) (which generally requires precolumn derivatization of the polar sugars), and HPAEC. The use of each of these modes of chromatography in this field is briefly reviewed below.

### Hydrophilic Interaction Chromatography

This term encompasses chromatography on strongly polar, unmodified silica gel or on stationary phases bearing bonded polar groups, on a silica or a polymeric matrix. Chromatography on microparticulate silica as such can be governed by mechanisms involving either adsorption or partition, or both, while partition is generally the predominant mechanism where bonded phases are the active groups.

Unmodified silica sorbents have been used mainly in HPLC of carbohydrate derivatives, and have long been regarded as applicable to analysis of sugars only after derivatization of the analytes, since with the solvent systems generally used sugars themselves are too strongly bound to the highly polar silica to permit effective separations. The necessity for precolumn derivatization is not really a disadvantage, since the introduction of chromophoric or fluorescent groups by this means makes possible the use of detectors, such as the ultraviolet (UV) photometer or



fluorimeter, that are more sensitive than the differential refractometer generally used in HILIC of sugars. For this purpose, sugars have been derivatized to benzoates or 4-nitrobenzoates (often after reduction to the alditols to eliminate the production of multiple peaks), also to 2,4-dinitrophenylhydrazones, or fluorescent dansylhydrazones (produced by reaction with 5-dimethylaminonaphthalene-1-sulfonyl hydrazine, known as dansyl hydrazine) or pernapthoates. Of special interest is the use of reductive amination with chiral 1-( $\alpha$ -methylbenzylamine, followed by acetylation, to achieve resolution of enantiomeric pairs of aldose sugars by HPLC on silica.

In 1992, Herbreteau and co-workers reported that unmodified silica stationary phases did exhibit high selectivity for sugars and alditols if solvent systems consisting of dichloromethane–methanol mixtures, with only a trace (0.2%, v/v) of water, were used instead of the usual acetonitrile–water eluents. Good resolution of the common monosaccharides was obtained by chromatography on 5- $\mu$ m silica with such solvent systems containing 20% of methanol, while increasing the proportion of methanol to 28% permitted the separation of several disaccharides and the trisaccharide raffinose. The presence of water in only trace amounts in these solvent systems prevents the very strong binding of sugars that is exhibited by hydrated silica, allowing chromatographic separation by a partition process. Such chromatographic systems provide an attractive alternative to the use of aminopropyl silica columns (see below) for HILIC of sugars.

Of the HPLC sorbents carrying bonded polar phases, aminopropyl silica types were for many years the most widely used in analysis of sugars. However, these column packings have a short lifetime in this application, due to interaction between reducing sugars and the bonded amine groups to form glycosylamines. This reaction results not only in deactivation of the column but also losses of the sugar analytes, and is thus one of the main reasons for the widespread adoption of other forms of HPLC for sugar analysis that has been such a marked trend during the past decade.

Several alternative column packings carrying different polar bonded phases have been introduced with the objective of overcoming the problem of the instability of alkylamino groups in analyses involving reducing sugars. These include polyamine, aminocyano, and carbamoyl amide phases. However, although some successes have been achieved in HILIC of linear oligosaccharides and cyclodextrins, such stationary phases have not proved very effective in separations of monosaccharides. The most successful

for this purpose has been the hydroxylic type, in which the microparticulate silica matrix carries bonded diol groups. In the earliest applications of stationary phases of this type to the separation of sugars broad peaks were obtained, due to partial resolution of anomers, with the usual acetonitrile–water solvent systems. More recently, greatly improved resolution, with sharper peaks, was observed by Herbreteau and co-workers on replacing these solvent systems with dichloromethane–methanol systems similar to those that had proved effective in chromatography of sugars on unmodified silica. Use of the evaporative light-scattering detector under these conditions made possible gradient elution, with resulting improvements in selectivity, and the detection of the sugars at nanogram levels.

Novel stationary phases that have some advantages (especially with regard to efficiency and stability) over many of the others that have been applied to HILIC of sugars are those in which  $\alpha$ - or  $\beta$ -cyclodextrin is bonded to 5- $\mu$ m silica. Retention data published by Armstrong and Jin for a wide variety of carbohydrates, including monosaccharides from triose to heptose, deoxy sugars, di-, tri-, and tetrasaccharides, on these column packings, with eluents consisting of aqueous acetonitrile (80–85%) or acetone (85–90%) demonstrate the feasibility of separations of many such compounds that are not easily resolved by other LC methods.

Polymeric supports for HILIC are used to advantage under certain conditions where silica-based packings would be unstable; for example, at elevated column temperatures or with mobile phases containing relatively high proportions (over 20%) of water. The polymer must be in microparticulate bead form (5–10  $\mu$ m in diameter) and rigid enough to withstand high pressures and flow rates; furthermore, it should not shrink or swell to any great extent with changes of solvent.

Column packings of this type that are suitable for HILIC of sugars include highly crosslinked polystyrenesulfonate ion-exchange resins, which function as supports for partition chromatography rather than as ion-exchangers in mobile phases containing organic solvents mixed with water or buffer solutions. In this case, the separation mechanism depends on the hydrophilicity of the anionic groups on the resin and their counter-ions. For example, a resin cross-linked with 55% of divinylbenzene has been successfully used, in the hydrogen form, with an aqueous acetonitrile mobile phase to separate all the monosaccharide constituents of glycoproteins, neutral and amino sugars, in a single run. The same resin, in the sodium or calcium form, is capable of resolving the  $\alpha$ - and  $\beta$ -anomers of most of the aldopyranoses if the column is operated at low temperature (4°C).

Polymers developed specifically for HPLC include the macroporous crosslinked vinylpyridinium type, which has proved very effective in HILIC of sugars in aqueous acetonitrile. These polymers can be used at column temperatures up to 70°C, under which conditions, with the resin in the phosphate or sulfate form, baseline resolution of mixtures of the common monosaccharides and some disaccharides, such as maltose and lactose, can be achieved. Stationary phases in which a polymer replaces silica as the support for the aminopropyl bonded phase much used in HILIC of carbohydrates have also been successfully applied in separations of mono- and disaccharides; in this case precolumn derivatization with 4-amino-benzoic acid ethyl ether (ABEE) has been recommended to overcome the problem of glycosylamine formation that occurs with underivatized sugars.

### Reversed-Phase Chromatography

Chromatography on C<sub>18</sub>-bonded silica has been applied to underivatized sugars, with water as the mobile phase, but with limited success as resolution is poor and is complicated by the formation of double peaks due to anomer separation. The mechanism in this case is not really reversed-phase partition but rather a form of hydrophobic interaction chromatography. True RPC of sugars requires the introduction of nonpolar groups by precolumn derivatization. As has been mentioned, this can be advantageous in making possible the use of sensitive UV or fluorimetric detection, and the derivatives are readily identifiable by MS. Furthermore, the solvent systems used in RPC are more compatible with on-line mass spectrometry than are, for example, the acetonitrile-rich eluents generally used in HILIC. For these reasons, there has been increased use of RPC for analysis of sugars in recent years.

A highly effective method of precolumn derivatization of sugars for RPC is reductive amination, which has the effect of destroying the anomeric center, thus preventing the formation of multiple peaks, and simultaneously labeling the reducing end of each sugar with a group that is amenable to UV, fluorimetric, and MS detection. Reagents widely used for this purpose are 2-aminopyridine and, more recently, ABEE. Others that are finding application in this way include 2-aminoacridine, 8-aminonaphthalene-1,3,6-trisulfonic acid, and 2-aminobenzoic acid (anthranilic acid).

Derivatization for RPC may also be effected by introduction of chromophores through reaction with the hydroxyl groups of the sugars, but in this case multiple peaks are obtained. Nevertheless, perbenzoylation followed by RPC has been used as an

alternative to GC of the trimethylsilylated product for the analysis of methanolysates from polysaccharides and glycoconjugates. With microbore RPC columns, analysis of the benzoylated methyl glycosides is possible at the picomolar level. RPC has also proved effective in analysis of acetylated carbohydrates: for example, the method can be used instead of GC in analysis of sugars as the peracetylated aldono-nitrile or ketoxime derivatives.

As in the case of HILIC, the use of polymeric supports for RPC can be advantageous. A vinyl alcohol co-polymer carrying bonded C<sub>18</sub> chains has been successfully used in chromatography of sugars and oligosaccharides in aqueous media, the formation of double peaks due to anomerization being prevented by addition of sodium hydroxide to the mobile phase (to pH 11). Silica-based supports are not stable under such conditions.

An important development during the past decade has been the introduction of graphitized carbon stationary phases for HPLC of underivatized carbohydrates. Monosaccharides are weakly retained on these columns and, in the aqueous mobile phases used to separate these sugars from oligosaccharides, double peaks are produced on chromatography. However, this system gives sharp resolution of disaccharides if sodium hydroxide is added to the eluent: for example, nine gluco-disaccharides have been resolved in 25 min on a graphitized carbon column by use of an acetonitrile gradient (1.5–5.0%) in dilute sodium hydroxide (1 mmol l<sup>-1</sup>).

### High-Performance Anion-Exchange Chromatography

Neutral sugars, being very weak electrolytes with pK<sub>a</sub> values of the order of 12–13, do not interact with anion-exchangers to any appreciable extent in neutral aqueous media. They can be converted to anionic complexes in the presence of borate, which then permits analysis by anion-exchange chromatography on a microparticulate resin. This was formerly a widely used technique in sugar analysis. However, run times are long and have remained so (up to 6 h) despite the introduction of more efficient resin columns during the 1980s. Therefore, ion-exchange chromatography of sugars is now performed almost exclusively at high pH, under which conditions the hydroxyl groups on the sugars are dissociated, so that the sugars become anionic.

Chromatography of sugars at high pH presents some problems, which had to be overcome before the use of this technique became feasible. First, carbohydrates are susceptible to base-catalyzed reactions, such as isomerization of reducing sugars or degradation

of oligosaccharides by  $\beta$ -elimination, on prolonged exposure to strong alkali. For this reason, chromatographic supports that are capable of faster interaction than the usual silica or polymeric types are required, to enable separations to be finished rapidly enough to minimize these structural changes during chromatography. High efficiency is also necessary, as the carbohydrate anions are weakly retained in comparison with the common inorganic anions. In 1983, the Dionex Corporation (Sunnyvale, CA, USA) introduced stationary phases that had been developed for this specific purpose: these are pellicular supports consisting of nonporous latex beads, of particle diameter 5 or 10  $\mu\text{m}$ , coated with a thin film of strongly basic anion-exchanger (now marketed under the trade name CarboPac).

The second problem was that the detectors generally used in HPLC were not suited to the conditions to be employed here. This was overcome by the simultaneous development of the pulsed amperometric detector (PAD). The carbohydrates eluted from the column are oxidized at the surface of a gold electrode, a selected potential being applied between this and a silver/silver chloride reference electrode. Pulsed-potential operation is necessary to avoid rapid deactivation of the electrode surface due to adsorption of intermediates produced during carbohydrate oxidation. The potential required for oxidation of the analytes is applied to the working electrode for a very short time (100 ms or less), and then it is increased to oxidize fully any material adsorbed on the electrode surface, after which it is reversed to a strongly reducing potential to convert the oxide layer back to the metal. These detectors require highly alkaline conditions for their operation, and are thus ideal for use in chromatography of carbohydrates at high pH. The sensitivity of the detector can be varied over a wide range, as can the pulse width. Detection of carbohydrates at picomolar levels is possible.

The technique of HPAEC–PAD has now become indispensable in the analysis of sugars and other carbohydrates. The power of the method lies in its superior efficiency, high sensitivity, and simplicity of sample handling (derivatization is not required). It is applicable not only to sugars but also to carbohydrates of relatively high molar mass, including the complex oligosaccharides released on degradation of glycoconjugates.

For analysis of monosaccharides the eluent that has been most widely used is NaOH (150–160  $\text{mmol l}^{-1}$ ), although various modifications have been introduced for specific purposes. For example, xylose and mannose, which occur together in hydrolysates from glycoconjugates of plant origin,

coelute under these conditions, and therefore the use of a more dilute alkaline eluent (1  $\text{mmol l}^{-1}$  NaOH) containing sodium acetate (0.03  $\text{mmol l}^{-1}$ ) has been recommended for analysis of such samples. With this solvent system and others containing the strong alkali at low concentrations postcolumn addition of NaOH is necessary for effective functioning of the PAD. This applies in particular to analyses involving amino sugars, which are best resolved with eluents containing 10–15  $\text{mmol l}^{-1}$  NaOH, but are detected satisfactorily only after addition of 0.3  $\text{mol l}^{-1}$  NaOH to the column effluent.

It has been shown that the presence in the mobile phase of monovalent ions other than  $\text{Na}^+$  does not significantly change the selectivity of the HPAEC system for sugars. However, recent studies have demonstrated that the introduction of alkaline earth cations in small proportion (1–2  $\text{mmol l}^{-1}$ ) into dilute alkaline eluents (5–20  $\text{mmol l}^{-1}$  NaOH) results in sharper peaks and improves resolution, especially in analysis of mixtures of monosaccharides and alditols.  $\text{Ba}^{2+}$  and  $\text{Sr}^{2+}$  ions appear to be more effective in this respect than  $\text{Ca}^{2+}$ . For example, with an eluent consisting of dilute NaOH (5  $\text{mmol l}^{-1}$ ) containing 1  $\text{mmol l}^{-1}$  barium acetate, a mixture of alditols and sugars, the former eluting earlier than the latter, can be completely resolved on the standard CarboPac PA-1 column (Dionex) within 15 min. Use of a higher concentration of alkali (20  $\text{mmol l}^{-1}$  NaOH) with the same concentration of barium acetate permits sharp resolution of the mixtures of neutral and amino sugars in hydrolysates from glycoconjugates, also in a short run time (less than 12 min). The presence of  $\text{Ba}^{2+}$  or  $\text{Sr}^{2+}$  ions in the mobile phase has also been shown to increase PAD response, for neutral, amino, and acidic sugars. These effects of alkaline earth cations have been ascribed to the formation of soluble complexes between these cations and sugars or alditols, as well as removal from the eluent of carbonate ions formed by dissolution of atmospheric  $\text{CO}_2$ .

The presence of acetate ions in the mobile phase can also be advantageous in HPAEC, especially in analysis of oligosaccharides, which are eluted earlier under such conditions. These anions have a slightly higher affinity for the strongly basic anion-exchange groups on the stationary phase than have  $\text{OH}^-$  ions, and therefore strongly bound solutes, such as oligosaccharides (including those of high degree of polymerization) and acidic saccharides, are more easily displaced in the presence of acetate. The use of an acetate gradient has proved very effective in such analyses (see Table 2).

The PAD has remained the most frequently used detector for HPAEC, but other methods are occasionally

**Table 2** Conditions recommended for LC analysis of sugars

Stationary phase	Mobile phase	Derivatization	Temperature <sup>a</sup> (°C)	Detection <sup>b</sup>
Unmodified silica	Dichloromethane–methanol–water (80:19.8:0.2)	None	RT	RI or evaporative light-scattering
	Dichloromethane–methanol–water (72:27.8:0.2); separates mono-, di-, and some trisaccharides	None	RT	As above
	<i>n</i> -Hexane–dioxane–dichloromethane, linear gradient, 22:2:1 → 4:2:1 (80 min)	As benzoylated alditols	RT	UV (275 nm)
	<i>n</i> -Hexane–chloroform–acetonitrile (5:2:1)	As 4-nitrobenzoyl alditols	RT	UV (260 nm)
	Chloroform–methanol–water (91.7:7.6:0.7)	As 2,4-dinitro-phenylhydrazones	RT	UV (352 nm)
	Chloroform–ethanol–water, 86:12:2 (4 min); 78:19.5:2.5 (20 min)	As dansylhydrazones	27	Fluorimetry (excitation 350 nm, emission 500 nm)
	(A) 2-Propanol (0.025%) in dichloromethane; (B) Acetonitrile (4%) in A; A + B, 85:15 (15 min); 85:15 → 20:80 (15 min); 20:80 (10 min)	As pernapthoates of derived alditols	RT	Fluorimetry (excitation 234 nm, emission 374 nm)
	<i>n</i> -Hexane–ethanol (19:1)	As 1-( <i>N</i> -acetyl- $\alpha$ -methylbenzyl-amino)-1-deoxy-alditol acetates; enantiomers resolved	RT	UV (230 nm)
Aminopropyl silica	Acetonitrile–water (4:1 for monosaccharides, 3:1 for di- and trisaccharides)	None	RT	RI
Diol-bonded silica	Dichloromethane–methanol (84:16)	None	RT	RI or evaporative light-scattering
$\alpha$ -Cyclodextrin bonded to silica	Acetonitrile–water (4:1) or acetone–water (17:3)	None	RT	RI
$\beta$ -Cyclodextrin bonded to silica	Acetonitrile–water (17:3) or acetone–water (9:1)	None	RT	RI
Polystyrenesulfonate cation-exchange resin, 55% DVB, H <sup>+</sup> -form	Acetonitrile–water (23:2); separates neutral and amino sugars in glycoprotein hydrolysates	None	30	Postcolumn reaction with 2-cyanoacetamide; UV (280 nm)
Resin as above, Na <sup>+</sup> - or Ca <sup>2+</sup> -form; to resolve anomeric pairs	Acetonitrile–water (4:1)	None	4	As above
<i>N</i> -Methyl-4-vinyl-pyridinium polymer, phosphate form	Acetonitrile–water (4:1); separates mono- and disaccharides	None	70	RI
Amine-bonded vinyl alcohol copolymer	Acetonitrile–water (9:1 for monosaccharides, 85:15 for disaccharides)	Reductive amination with ABEE	25	UV (304 nm)
C <sub>18</sub> -silica	Acetonitrile–water, linear gradient, 35 → 90% acetonitrile (65 min); 90% acetonitrile (10 min)	As benzoylated methyl glycosides	RT	UV (230 nm)
	Water–tetrahydrofuran–methanol (13:2:5) or methanol–water, linear gradient, 40 → 5% methanol (10 min)	As peracetylated aldonitriles and ketoximes	RT	UV (195, 207, and 215 nm)
	Acetic acid (80 mmol l <sup>-1</sup> )–acetonitrile (79:21 for monosaccharides, 76:24 for mixtures of mono- and disaccharides)	As dansylhydrazones	RT	Fluorimetry (excitation 370 nm, emission 540 nm)

Table 2 Continued

Stationary phase	Mobile phase	Derivatization	Temperature <sup>a</sup> (°C)	Detection <sup>b</sup>
Pellicular anion-exchanger (CarboPac PA-1)	Sodium citrate buffer (0.25 mol l <sup>-1</sup> , pH 4.0) containing acetonitrile (1.0%)	Reductive amination with 2-amino-pyridine	25	Fluorimetry (excitation 320 nm, emission 400 nm)
	(A) Sodium acetate buffer (50 mmol l <sup>-1</sup> , pH 4.5); (B) Solvent A–acetonitrile– methanol (2:2:1); A and B mixed 3:1	Reductive amination with ABEE	45	UV (254 and 304 nm)
	(A) 1-Butylamine (0.2–0.3%), H <sub>3</sub> PO <sub>4</sub> (0.5%) and tetrahydrofuran (1%) in water; (B) Solvent A–acetonitrile (1:1); A + B (95:5) for 25 min; 5 → 15% B (50 min)	Reductive amination with anthranilic acid	RT	Fluorimetry (excitation 230 nm, emission 425 nm)
	NaOH, 150–160 mmol l <sup>-1</sup> for most neutral sugars, 10– 15 mmol l <sup>-1</sup> for amino sugars	None	RT	Pulsed amperometric detector (PAD)
	NaOH, 5 mol l <sup>-1</sup> , containing barium acetate, 1 mmol l <sup>-1</sup> ; separates alditols from sugars	None	RT	PAD
	NaOH, 20 mmol l <sup>-1</sup> , containing barium acetate, 1 mol l <sup>-1</sup> ; separates neutral from amino sugars	None	RT	PAD
	NaOH, 15 mmol l <sup>-1</sup> (15 min); linear gradient, 15 → 100 mmol l <sup>-1</sup> NaOH, 0 → 150 mmol l <sup>-1</sup> sodium acetate (40 min); separates acidic sugars from neutral and amino	None	RT	PAD

<sup>a</sup> RT = room temperature.<sup>b</sup> RI = refractive index detector; UV = ultraviolet.

employed where sugars are present in trace amounts in complex biological samples such as cell lysates. In this case, radiochemical labeling or derivatization with fluorogenic reagents is used to increase the sensitivity of detection. Direct coupling of HPAEC to MS presents problems due to the high salt content of the mobile phases; however, recent improvements in desalting methods (for example, the use of micro-membrane suppressors or online microdialysis) are rapidly overcoming these difficulties. Since matrix-assisted laser desorption/ionization/time-of-flight mass spectrometry shares with HPAEC the capability of analyzing carbohydrates in their native forms with high efficiency and sensitivity, there has been an increasing tendency to combine the two techniques in seeking to optimize such analyses. Coupling of HPAEC systems to electrospray ionization MS is also becoming more successful with the development of online desalting systems, with disaccharides such as sucrose, maltose, and trehalose now detectable at picogram levels.

Some recommended conditions for analysis of sugars by HPAEC and the other LC systems discussed in this section are given in Table 2.

### Thin-Layer Chromatography

Although TLC on cellulose plates is effective in separating mixtures of sugars, including amino and acidic sugars, multiple development or two-dimensional TLC with different solvent systems is usually required, and the method is seldom used today. TLC on unmodified silica gel plates is suitable only for carbohydrate derivatives of low polarity, such as methyl ethers, as polar sugars are too strongly adsorbed by the highly polar stationary phase (see also section on Liquid Chromatography). Prior impregnation of the silica gel layer with an inorganic salt that interacts with carbohydrates is necessary for effective TLC of mixtures of closely related sugars and alditols. Borate or phosphate buffers have been widely

used for this purpose. Inclusion of boric acid or lactic acid in the solvent system has also proved effective. A recent study has shown that both the cation and the anion of the impregnating salt can influence the selectivity of silica gel plates for sugars; for example, transition metal cations have a greater effect than alkali metals.

The introduction of HPTLC plates, coated with silica gel of small particle size, in thinner, denser, and more uniform layers than those on conventional TLC plates, has resulted in vast improvements in efficiency and resolution for TLC of sugars. Separations of analytes in nanogram and picogram amounts are possible with sensitive detection, and run times have been reduced (often to less than 10 min). As in the case of conventional TLC impregnation of the plates with phosphate or borate buffers or addition of boric acid to the solvent system is usually necessary to improve resolution. With HPTLC plates carrying a bonded aminopropyl phase, which tends to react covalently with the hydroxyl groups of the sugars, prior impregnation of the plate with sodium dihydrogen phosphate is essential to prevent such interaction.

Some TLC systems that have proved effective in separations of sugars are listed in Table 3.

### Detection Reagents for TLC of Sugars

The chromogenic spray reagents used to visualize sugars on TLC plates can be classified into four main types, according to the kind of reaction responsible for color formation:

1. Reduction by sugars.
2. Reaction of the sugars with an acid to produce 2-furaldehyde (furfural) derivatives, which then condense with phenols or aromatic amines.
3. Glycol cleavage with periodic acid or sodium metaperiodate, followed by reaction of the products to yield visible or fluorescent spots.
4. Reactions involving specific structural features, such as keto, amino, or carboxylic acid groups.

Detection limits with such visualization reagents are generally at microgram levels. Greater sensitivity is possible by use of reagents that are fluorogenic or produce a UV-absorbing chromophore in derivatizing sugars. Detection at nanogram levels is possible when fluorimetry or techniques such as UV diffuse-reflectance densitometry are used in scanning the plates. As in LC prederivatization of sugars is sometimes advantageous in TLC. The most successful reagent for this purpose has been dansyl hydrazine, which gives intensely fluorescent products with

**Table 3** TLC systems useful in separation of sugars

<i>Stationary phase</i>	<i>Mobile phase<sup>a</sup></i>	<i>Special features</i>
Silica gel 60	2-Butanone–acetic acid–saturated aqueous solution of boric acid (9:1:1)	Good resolution of monosaccharides
	2-Propanol–acetone–1 mol l <sup>-1</sup> lactic acid (2:2:1)	Separates mono-, di-, and trisaccharides
	2-Butanone–2-propanol–acetonitrile–0.5 mol l <sup>-1</sup> boric acid + 0.25 mol l <sup>-1</sup> 2-propylamine (4:3:1:2)	Separates mono-, di-, and trisaccharides
Silica gel 60, impregnated with boric acid (30 mmol l <sup>-1</sup> )	1-Butanol–ethyl acetate–2-propanol–water (35:10:6:3)	Good resolution of aldoses and ketoses
Silica gel 60, impregnated with	Acetone–water (9:1)	Separates mono-, di-, and trisaccharides
NaH <sub>2</sub> PO <sub>4</sub> (0.2 mol l <sup>-1</sup> )	1-Butanol–2-propanol–water (3:5:2)	Separates mono- and disaccharides
NaH <sub>2</sub> PO <sub>4</sub> (0.3 mol l <sup>-1</sup> )	Acetone–1-butanol–water (8:1:1)	Good resolution of aldoses and ketoses
NaH <sub>2</sub> PO <sub>4</sub> (0.5 mol l <sup>-1</sup> )	2-Propanol–acetone–0.1 mol l <sup>-1</sup> lactic acid (2:2:1)	Separates mono-, di-, and trisaccharides
Silica gel 60, HPTLC plates	2-Propanol–1-butanol–80 mmol l <sup>-1</sup> boric acid (5:3:1)	Separates alditols and monosaccharides
	Ethyl acetate–pyridine–water–acetic acid–propionic acid (10:10:2:1:1)	Separates mono-, di-, and trisaccharides
Silica gel 60, HPTLC plates, impregnated with phosphate buffer, pH 8.0	2-Butanone–acetic acid–80 mmol l <sup>-1</sup> boric acid (9:1:1)	Good resolution of pentoses
Aminopropyl silica, HPTLC plates	Acetonitrile–water (7:3)	Separates di-, tri-, and tetrasaccharides; aldoses react with amino groups
As above, impregnated with NaH <sub>2</sub> PO <sub>4</sub> (0.2 mol l <sup>-1</sup> )	Acetonitrile–water (7:3)	Separates mono-, di-, tri-, and tetrasaccharides

<sup>a</sup>All proportions are by volume.

**Table 4** Detection reagents for TLC of sugars

<i>Reagent</i>	<i>Procedure after spraying</i>	<i>Color produced</i>	<i>Detection limits</i>
<b>Chromogenic reagents</b>			
Aniline phosphate: aniline ( $2 \text{ mol l}^{-1}$ ) in water-saturated 1-butanol, mixed 1:2 (v/v) with $\text{H}_3\text{PO}_4$ ( $0.67 \text{ mol l}^{-1}$ ) in 1-butanol	Heated at $105^\circ\text{C}$ for 10 min	Aldopentoses – red-brown; aldohexoses – yellow-brown; sorbose – yellow-brown; other ketoses react only at high concentration	2–4 $\mu\text{g}$ for most aldoses
Aniline phthalate: aniline hydrogen phthalate (0.93%, m/v) mixed with phthalic acid (1.66%, m/v) in water-saturated 1-butanol	Heated at $105\text{--}120^\circ\text{C}$ for 5 min	Aldopentoses – red; aldohexoses, deoxy, and amino sugars – brown; uronic acids – orange-brown; very weak reaction with ketoses	1–2 $\mu\text{g}$ for most aldoses
Diphenylamine–aniline–phosphoric acid: diphenylamine (2%, m/v) and aniline (2%, v/v) mixed 10:1 (v/v) with 85% aqueous $\text{H}_3\text{PO}_4$	Air-dried, then heated at $100^\circ\text{C}$ for 10 min	Aldopentoses – olive green; aldohexoses – blue-grey; deoxypentoses – pink to violet; deoxyhexoses – yellow to green; ketoses – yellow to red; for disaccharides colors depend on both composition and linkage	Monosaccharides 1–5 $\mu\text{g}$ ; disaccharides 5–10 $\mu\text{g}$
1-Naphthol–phosphoric acid: 1-naphthol (1-hydroxynaphthalene) (0.5%, m/v) in aqueous ethanol (50%, v/v), mixed 10:1 (v/v) with $\text{H}_3\text{PO}_4$	Heated at $90^\circ\text{C}$ for 10–15 min	Aldopentoses – purple; deoxy sugars – orange; pentuloses – blue; other ketoses – violet	Aldopentoses and 6-deoxyhexoses 0.5 $\mu\text{g}$ ; 2-deoxyhexoses 2 $\mu\text{g}$ ; ribulose 0.5 $\mu\text{g}$ ; other ketoses 1–2 $\mu\text{g}$
Naphthoresorcinol: naphthoresorcinol (1,3-dihydroxynaphthalene) (0.2%, m/v) in ethanol, mixed 25:1 (v/v) with $\text{H}_2\text{SO}_4$	Heated at $105^\circ\text{C}$ for 5 min	Xylose – brown; other aldopentoses – light blue; aldohexoses – blue to violet; 2-deoxyribose – green; 2-deoxyhexoses – light blue; 6-deoxyhexoses – pink; ribulose – green; xylulose – orange; other ketoses – purple-red; uronic acids – blue	Aldopentoses 1–2 $\mu\text{g}$ ; aldohexoses 2–4 $\mu\text{g}$ ; deoxy sugars 0.5 $\mu\text{g}$ ; ribulose 0.5 $\mu\text{g}$ ; xylulose 0.2 $\mu\text{g}$ ; hexuloses 0.1–0.5 $\mu\text{g}$ ; heptuloses 0.5–4 $\mu\text{g}$ ; uronic acids 1–5 $\mu\text{g}$
Periodate–permanganate: aqueous solution of $\text{NaIO}_4$ (2%, m/v) mixed 2:1 (v/v) with solution of $\text{KMnO}_4$ (1%, m/v) in aqueous $\text{Na}_2\text{CO}_3$ (2%, m/v)	Air-dried	Yellow spots produced by most sugars after 20–30 min	Aldopentoses 4 $\mu\text{g}$ ; other aldoses 1–2 $\mu\text{g}$
Triphenyltetrazolium chloride: 2,3,5-triphenyltetrazolium chloride (4%, m/v) in methanol, mixed 1:1 (v/v) with $\text{NaOH}$ ( $1 \text{ mol l}^{-1}$ ) in methanol	Heated at $100^\circ\text{C}$ for 5–10 min	Reducing sugars give pink to red spots	1 $\mu\text{g}$ for most reducing sugars
<b>Specific reagents</b>			
<b>For amino sugars</b>			
Ninhydrin: ninhydrin (0.1%, m/v) in 1-butanol	Heated at $105\text{--}110^\circ\text{C}$ for 10 min	Amino sugars and amino acids give purple spots; better distinguished in presence of cupric ions (see below)	0.5 $\mu\text{g}$
Ninhydrin–Cu(II): (A) ninhydrin (2%, m/v) in ethanol; acetic acid (20%, v/v) and collidine (4%, v/v) added; (B) $\text{Cu}(\text{NO}_3)_2 \cdot 3\text{H}_2\text{O}$ (1%, m/v) in ethanol; A and B are mixed 50:3 (v/v) just before use	Heated at $105^\circ\text{C}$ for 2–3 min	Aminodeoxyhexoses – intense yellow-brown spots; aminodideoxyhexoses (e.g. fucosamine) – orange-yellow; amino acids – grey, red, or purple	0.1–0.2 $\mu\text{g}$ for most amino sugars

Continued



Table 4 Continued

Reagent	Procedure after spraying	Color produced	Detection limits
For 6-deoxyhexoses Sulfosalicylic acid: sulfosalicylic acid (2%, m/v) in H <sub>2</sub> SO <sub>4</sub> (0.5 mol l <sup>-1</sup> )	Heated at 110°C for 15 min	6-Deoxyhexoses – yellow; all other sugars – gray to gray- brown	Rhamnose 0.3 µg; fucose 0.5 µg; other sugars 1–2 µg
For ketoses 2,4-Dinitrophenyl- hydrazine: 2,4-dinitro- phenylhydrazine (0.4%, m/v) in HCl (2 mol l <sup>-1</sup> )	Heated at 105°C for 5 min	Ketoses or ketals – orange	1–2 µg for most keto compounds
Resorcinol–hydrochloric acid: resorcinol (0.2%, m/v) in 1-butanol, mixed 1:1 (v/v) with HCl (0.25 mol l <sup>-1</sup> ) just before use	Heated at 105°C for 5 min	Pentuloses – blue; hexuloses – red	1–2 µg for most ketoses
Thiobarbituric acid– H <sub>3</sub> PO <sub>4</sub> : thiobarbituric acid (0.5%, m/v) in ethanol, mixed 50:1 (v/v) with H <sub>3</sub> PO <sub>4</sub> (85%)	Heated at 100°C for 10 min	Ketoses and oligosaccharides containing ketoses – yellow to orange	Pentuloses 4 µg; hexuloses 0.5 µg; heptuloses 0.5–2 µg
Urea–sulfuric acid: urea (25%, m/v) in H <sub>2</sub> SO <sub>4</sub> (2 mol l <sup>-1</sup> ), mixed 1:5 (v/v) with ethanol	Heated at 100°C for 20–30 min	Hexuloses – blue; other ketoses – yellow-brown	Pentuloses and hexuloses 0.5 µg; heptuloses 1 µg
For uronic acids Mixed indicators: thymol blue (0.005%, m/v), methyl red (0.025%, m/v), and bromothymol blue (0.06%, m/v) in ethanol (95%); pH adjusted with NaOH (1 mol l <sup>-1</sup> ) until blue-green color reached	Observed at room temperature	Uronic acids and oligomers (e.g. oligogalacturonic acids) give red spots on dark green background	About 5–10 µg
<i>Fluorogenic reagents</i>			
Fluorescamine: (A) triethylamine (10%, v/v) in dichloromethane; (B) fluorescamine (4-phenyl- spiro[furan-2-(3H),1'- phthalan]-3,3'-dione) (0.05%, m/v) in acetone	Plate sprayed with A, air- dried, sprayed with B, air-dried, then sprayed again with A; spraying with base is necessary to stabilize and intensify fluorescence	Specific for amino- and acetamidodeoxy sugars; fluorimetric scanning (excitation 390 nm, emission 475 nm)	50–100 pmol
Lead tetraacetate–2,7- dichlorofluorescein: (A) lead tetraacetate (saturated solution) in glacial acetic acid; (B) 2,7- dichlorofluorescein in water (1%, m/v, or lower, down to 0.2%, if back- ground interferes with detection); A and B are mixed 1:1 (v/v), then toluene (19:1, v/v) is added	Plate is dipped in reagent for ~10 s, then heated at 100°C for 3 min	All carbohydrates with vicinal diol groups, oxidized by lead tetraacetate, are detected; fluorimetric scanning (excitation 313 nm, emission 366 nm)	50–100 pmol
Malonamide: malonamide (1%, m/v) in sodium carbonate buffer (1 mol l <sup>-1</sup> , pH 9.2)	Heated at 120°C for 20 min	Fluorimetric scanning (excitation 328–382 nm, emission 383– 425 nm)	0.25 nmol for most reducing sugars; deoxyhexoses 0.5 nmol

reducing aldoses, and the corresponding chloride, which reacts only with amino compounds (including amino sugars).

Some useful chromogenic and fluorogenic reagents for detection of sugars in TLC are listed in Table 4.

*See also:* **Carbohydrates:** Overview; Sugars – Spectrophotometric Methods. **Chromatography:** Overview. **Gas Chromatography:** Overview; Mass Spectrometry; Chiral Separations. **Ion Exchange:** Overview; Ion Chromatography Instrumentation; Ion Chromatography Applications. **Liquid Chromatography:** Overview; Normal Phase; Reversed Phase; Clinical Applications; Food Applications. **Mass Spectrometry:** Overview; Archaeological Applications; Clinical Applications; Food Applications. **Thin-Layer Chromatography:** Overview.

## Further Reading

- Armstrong DW and Jin HL (1989) Evaluation of the liquid chromatographic separation of monosaccharides, disaccharides, trisaccharides, tetrasaccharides and sugar alcohols with stable cyclodextrin bonded phase columns. *Journal of Chromatography* 462: 219–232.
- Bleton J, Mejanelle P, Sansoulet J, Goursaud S, and Tchaplal A (1996) Characterization of neutral sugars and uronic acids after methanolysis and trimethylsilylation for recognition of plant gums. *Journal of Chromatography A* 720: 27–49.
- Fox A (2002) A current perspective on analysis of sugar monomers using GC–MS and GC–MS/MS. In: El Rassi Z (ed.) *Carbohydrate Analysis by Modern Chromatography and Electrophoresis. Journal of Chromatography Library*, vol. 66, pp. 829–843. Amsterdam: Elsevier.
- Heinrich J, König WA, Bretting H, and Mischnick P (1997) Enantiomer separation of permethylated monosaccharides and 1,5-anhydroalditols and simultaneous determination of linkage positions and absolute configuration in the galactan of *Helix pomatia*. *Carbohydrate Research* 299: 1–6.
- Herbreteau B (1992) Review and state of sugar analysis by high performance liquid chromatography. *Analysis* 20: 355–374.
- Katona ZF, Sass P, and Molnar-Perl I (1999) Simultaneous determination of sugars, sugar alcohols, acids and amino acids in apricots by gas chromatography–mass spectrometry. *Journal of Chromatography A* 847: 91–102.
- Klaus R and Fischer W (1988) Quantitative thin-layer chromatography of sugars, sugar acids and polyalcohols. *Methods in Enzymology* 160: 159–175.
- König WA (1989) Enantioselective gas chromatography with modified cyclomalto-oligosaccharides as chiral stationary phases. *Carbohydrate Research* 192: 51–60.
- Seymour FR (1993) Identification and characterization of saccharides by GLC separation and MS analysis of their peracetylated aldonitrile (PAAN) and ketoxime (PAKO) derivatives. In: BeMiller JN, Whistler RL, and Shaw DH (eds.) *Methods in Carbohydrate Chemistry*, vol. IX, pp. 59–85. New York: Wiley-Interscience.
- Zanetta J-P, Timmerman P, and Leroy Y (1999a) Gas-liquid chromatography of heptafluorobutyrate derivatives of the O-methyl glycosides on capillary columns: a method for the quantitative determination of the monosaccharide composition of glycoproteins and glycolipids. *Glycobiology* 9: 255–266.
- Zanetta J-P, Timmerman P, and Leroy Y (1999b) Determination of constituents of sulphated proteoglycans, using a methanolysis procedure and gas chromatography/mass spectrometry of heptafluorobutyrate derivatives. *Glycoconjugate Journal* 16: 617–627.
- Zhang Y and Lee YC (2002) High-performance anion-exchange chromatography of carbohydrates on pellicular resin columns. In: El Rassi Z (ed.) *Carbohydrate Analysis by Modern Chromatography and Electrophoresis. Journal of Chromatography Library*, vol. 66, pp. 207–250. Amsterdam: Elsevier.

## Sugar Alcohols

**M E Legaz and C Vicente**, Universidad Complutense de Madrid, Madrid, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Sugars and sugar alcohols are one of the most abundant classes of organic molecules in various

environments and have biologically significant roles. Recent advances in biochemistry have stimulated the demand for the analytical determination of these compounds. In addition, these molecules are widespread in food and beverages; hence, their determination can be considered an important step because sugars and alditols contribute to the nutritional value, flavor, and organoleptic characteristics of foods. The development of high-performance liquid chromatography (HPLC) techniques, based on

reducing aldoses, and the corresponding chloride, which reacts only with amino compounds (including amino sugars).

Some useful chromogenic and fluorogenic reagents for detection of sugars in TLC are listed in Table 4.

*See also:* **Carbohydrates:** Overview; Sugars – Spectrophotometric Methods. **Chromatography:** Overview. **Gas Chromatography:** Overview; Mass Spectrometry; Chiral Separations. **Ion Exchange:** Overview; Ion Chromatography Instrumentation; Ion Chromatography Applications. **Liquid Chromatography:** Overview; Normal Phase; Reversed Phase; Clinical Applications; Food Applications. **Mass Spectrometry:** Overview; Archaeological Applications; Clinical Applications; Food Applications. **Thin-Layer Chromatography:** Overview.

## Further Reading

- Armstrong DW and Jin HL (1989) Evaluation of the liquid chromatographic separation of monosaccharides, disaccharides, trisaccharides, tetrasaccharides and sugar alcohols with stable cyclodextrin bonded phase columns. *Journal of Chromatography* 462: 219–232.
- Bleton J, Mejanelle P, Sansoulet J, Goursaud S, and Tchaplal A (1996) Characterization of neutral sugars and uronic acids after methanolysis and trimethylsilylation for recognition of plant gums. *Journal of Chromatography A* 720: 27–49.
- Fox A (2002) A current perspective on analysis of sugar monomers using GC–MS and GC–MS/MS. In: El Rassi Z (ed.) *Carbohydrate Analysis by Modern Chromatography and Electrophoresis. Journal of Chromatography Library*, vol. 66, pp. 829–843. Amsterdam: Elsevier.
- Heinrich J, König WA, Bretting H, and Mischnick P (1997) Enantiomer separation of permethylated monosaccharides and 1,5-anhydroalditols and simultaneous determination of linkage positions and absolute configuration in the galactan of *Helix pomatia*. *Carbohydrate Research* 299: 1–6.
- Herbretreau B (1992) Review and state of sugar analysis by high performance liquid chromatography. *Analysis* 20: 355–374.
- Katona ZF, Sass P, and Molnar-Perl I (1999) Simultaneous determination of sugars, sugar alcohols, acids and amino acids in apricots by gas chromatography–mass spectrometry. *Journal of Chromatography A* 847: 91–102.
- Klaus R and Fischer W (1988) Quantitative thin-layer chromatography of sugars, sugar acids and polyalcohols. *Methods in Enzymology* 160: 159–175.
- König WA (1989) Enantioselective gas chromatography with modified cyclomalto-oligosaccharides as chiral stationary phases. *Carbohydrate Research* 192: 51–60.
- Seymour FR (1993) Identification and characterization of saccharides by GLC separation and MS analysis of their peracetylated aldonitrile (PAAN) and ketoxime (PAKO) derivatives. In: BeMiller JN, Whistler RL, and Shaw DH (eds.) *Methods in Carbohydrate Chemistry*, vol. IX, pp. 59–85. New York: Wiley-Interscience.
- Zanetta J-P, Timmerman P, and Leroy Y (1999a) Gas-liquid chromatography of heptafluorobutyrate derivatives of the O-methyl glycosides on capillary columns: a method for the quantitative determination of the monosaccharide composition of glycoproteins and glycolipids. *Glycobiology* 9: 255–266.
- Zanetta J-P, Timmerman P, and Leroy Y (1999b) Determination of constituents of sulphated proteoglycans, using a methanolysis procedure and gas chromatography/mass spectrometry of heptafluorobutyrate derivatives. *Glycoconjugate Journal* 16: 617–627.
- Zhang Y and Lee YC (2002) High-performance anion-exchange chromatography of carbohydrates on pellicular resin columns. In: El Rassi Z (ed.) *Carbohydrate Analysis by Modern Chromatography and Electrophoresis. Journal of Chromatography Library*, vol. 66, pp. 207–250. Amsterdam: Elsevier.

## Sugar Alcohols

**M E Legaz and C Vicente**, Universidad Complutense de Madrid, Madrid, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Sugars and sugar alcohols are one of the most abundant classes of organic molecules in various

environments and have biologically significant roles. Recent advances in biochemistry have stimulated the demand for the analytical determination of these compounds. In addition, these molecules are widespread in food and beverages; hence, their determination can be considered an important step because sugars and alditols contribute to the nutritional value, flavor, and organoleptic characteristics of foods. The development of high-performance liquid chromatography (HPLC) techniques, based on

hydrophilic interactions, size exclusion, or ion exclusion, has provided effective methods for separating sugar alcohols. Very recently, capillary electrophoresis (CE) has also emerged as a powerful separation tool in these analyses. A major problem with carbohydrate detection in liquid chromatography (LC) or CE is that this class of molecules, which frequently lack chromophores or fluorophores, limits the use of ultraviolet (UV) and fluorescence detectors, unless these molecules were previously derivatized. The most popular detectors are the differential refractive index and evaporative light scattering detectors, both of which have poor detection limits. Also reported is the use of preadditives in the background electrolyte (BGE) in CE for on-column derivatization, which convert these molecules in high-absorbing moieties. On the other hand, direct electrochemical detection without a prior derivatization procedure has gained prominence. For identification, LC-mass spectrometry (MS) and CE-MS are preferred.

## Types of Compounds and Matrices

Sugar alcohols are sugar derivatives in which the aldo or keto group has been reduced to the corresponding hydroxyl group. Polyols are usually classified into two different groups: (1) glycitols, or acyclic polyols, consisting of linear chains of three to seven carbon atoms (or more when they are branched), and (2) cyclitols, or cyclic polyols, such as inositol and inositol derivatives.

The most common acyclic polyols are shown in **Figure 1**. If the various possible branch-chained, methylated, and longer-chain polyols are considered, the number of possible sugar alcohols increases enormously. Some of these compounds are found in nature, such as D-volemitol, D-glycero-D-manno-D-thaloheptitol, and D-perseitol. They are widespread in the animal and plant kingdoms as free compounds or forms conjugated to lipids, polysaccharides, and plant growth regulators. Bacteria also produce a large number of acyclic polyols, some of them related to phosphorylation-dephosphorylation mechanisms of hexose transport through plasmalemma.

Inositols are cyclohexanehexols that naturally occur as free, methylated, or phosphorylated forms. Of the nine possible isomers (**Figure 2A**), three are synthetic (*epi*-, *allo*-, and *cis*-inositol), seven are optically inactive as *meso* forms, and two are enantiomorphs (D- and L-*chiro* inositol). However, phosphorylation or methylation of the inositol isomer most widely distributed in the nature, *myo*-inositol, at one of the hydroxyl groups 1, 3, 4, or 6 lead to chiral compounds, as shown in **Figure 2B**. Inositol

methylethers are found only in plants whereas *scyllo*-inositol also occurs in animals, presumably derived from plant foods.

Phosphorylated inositols are the most important form of cyclitols in plant and animal metabolisms. Phosphate residues can individually be bound to hydroxyl groups of the inositol as a phosphomonoester, forming an inositol-(poly)-orthophosphate (see **Figure 3**). Alternatively, a phosphate residue can esterify two neighboring hydroxyl groups to form a phosphodiester. Orthophosphate can be replaced by pyrophosphate (diphosphate) in a linearly or cyclically linked ester or diester, respectively, although phosphodiesters of inositol rarely occur. Depending on the number of phosphate residues, inositol phosphates are called as mono-, bis-, tris-, tetrakis-, pentakis-, or hexakisphosphate.

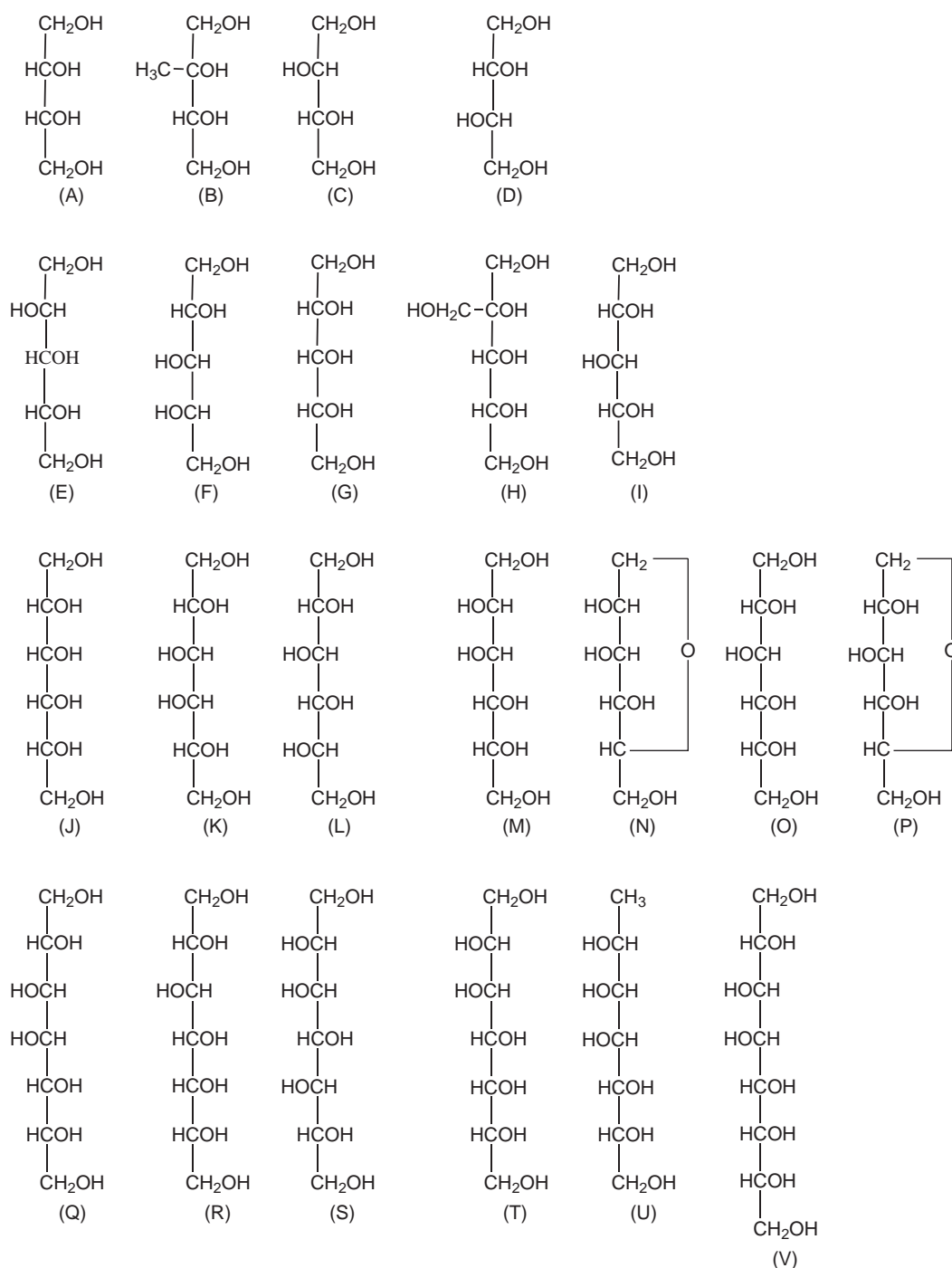
## Special Considerations in Sample Handling and Storage of Samples

Sugar alcohols should be extracted immediately from fresh material to avoid their enzymatic degradation. Air-dried plant material and even material stored at deep freeze temperatures ( $-25^{\circ}\text{C}$ ) over a long period should be avoided for the same reason. Plant material must be free of microbial infection and contamination.

Fractionation of the material is recommended because certain substances interfere with some detection methods. Percolation with petroleum ether removes lipophilic substances. Proteins and mucilages can be removed by precipitation with metal ions or 5% (w/v) trichloroacetic acid (TCA). Discoloration of biological materials is achieved with charcoal. Owing to the insolubility of benzyldine derivatives of sugar alcohols in water, they can easily be crystallized and purified from adhering matter by washing with water.

Human plasma and cerebrospinal fluid (CSF) samples must be frozen immediately after they are collected, and must be stored at  $-70^{\circ}\text{C}$  until analyzed. For plasma and CSF, borosilicate glass test tubes must be silanized with 5% (w/v) trimethylchlorosilane in hexane for 30 min at  $75^{\circ}\text{C}$  and unreacted reagent is removed by rinsing with methanol.

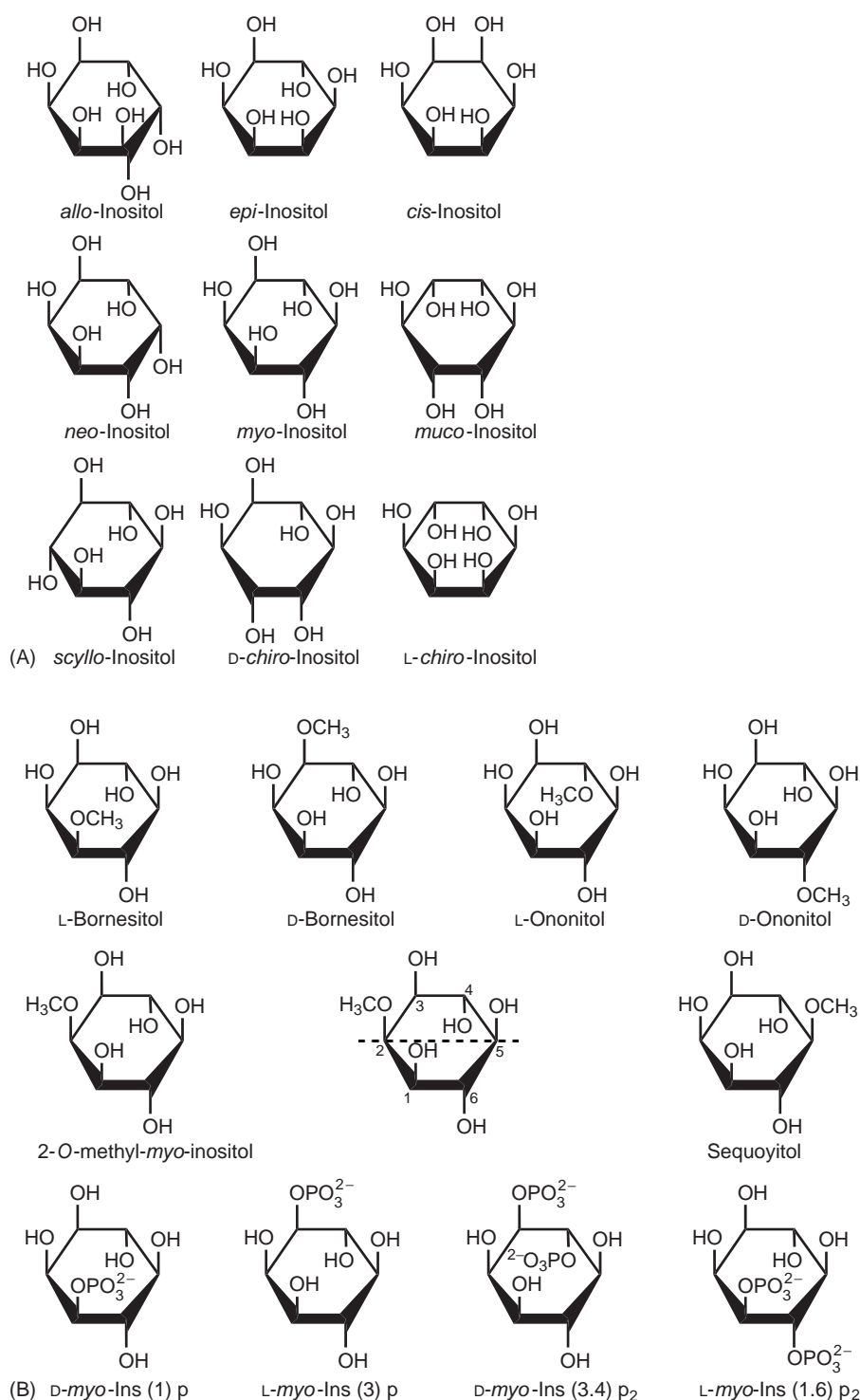
Glucose, sucrose, and polyhydric alcohols, glycerol, sorbitol, and mannitol, display nucleophilic reactivity with simple activated esters in aqueous solution buffered at neutral to alkaline pH. This nucleophilic reactivity is attributed to the anion resulting from ionization of a hydroxyl group. These polyhydric alcohols have been shown to be catalytically active in the hydrolysis of cephalosporins in



**Figure 1** Structural formulas of acyclic polyols: A = *meso*-erythritol; B = 2-C-methylerythritol; C = D-Threitol; D = L-threitol; E = D-aribitol; F = L-arabitol; G = *meso*-ribitol; H = D-hamamelitol; I = *meso*-xylitol; J = *meso*-allitol; K = *meso*-galactitol; L = L-Iditol; M = D-mannitol; N = 1,5-anhydro-D-mannitol; O = D-glucitol; P = 1,5-anhydro-D-glucitol; Q = D-glycero-D-galacto-heptitol; R = D-glycero-D-glucoheptitol; S = D-glycero-L-guloheptitol; T = D-glycerol-D-mannoheptitol; U = 7-deoxy-D-glycero-D-mannoheptitol; V = D-erythro-D-galacto-octitol.

aqueous solution. The reaction mechanism involves opening of the  $\beta$ -lactam moiety by an alkoxide ion derived from proton ionization of one of the hydroxyl groups generating an intermediate ester which undergoes further hydrolysis.

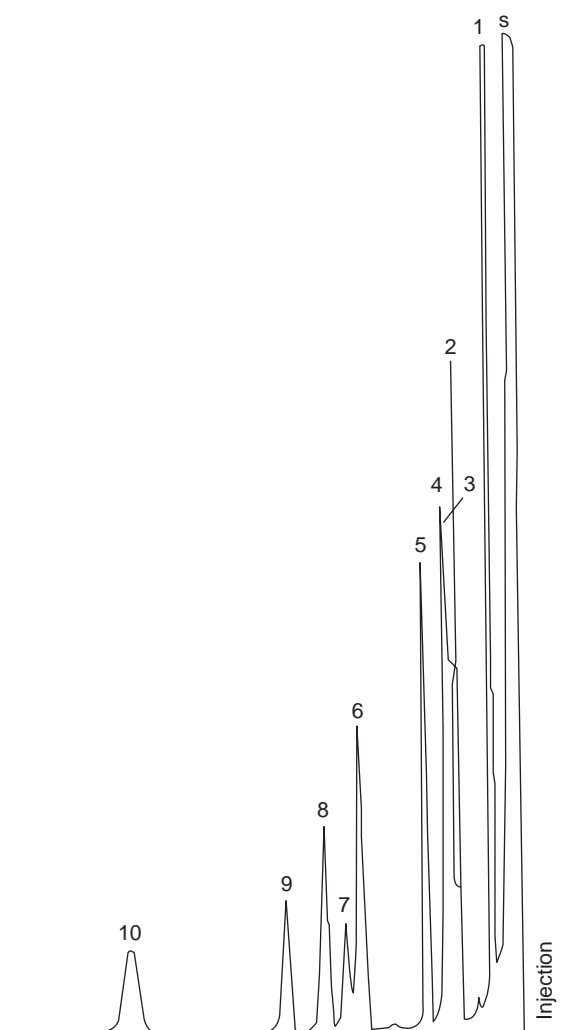
Glucose may interfere with sorbitol and galactitol estimation by gas-liquid chromatography (GLC) and therefore it must be removed prior to analysis by incubating samples with glucose oxidase. It is also recommended that monosaccharides be



**Figure 2** Structural formulas of cyclitols and the appearance of chiral forms after methylation or phosphorylation.

converted to the corresponding methyloxime derivatives prior to acetylation in order to avoid interferences in galactitol and mannitol estimation when using GLC as separation method. When GLC is chosen as analytical procedure, pyridine

tends to mask polyol peaks with short retention time values. In such cases, the reacted sample may be dried down and redissolved in a more volatile solvent such as chloroform, hexane, or heptane.



**Figure 3** Chromatogram in GC of a standard solution of ten sugar alcohols. Peaks are: s=solvent front; 1=erythritol; 2=pentaerythritol; 3=ribitol; 4=arabitol; 5=xylitol; 6=mannitol; 7=galactitol; 8=sorbitol; 9=inositol; 10=perseitol. The retention time of pentaerythritol was 4 min.

## Methods of Determination

### Extraction Procedures

Extraction of sugar alcohols is usually performed with boiling ethanol containing different percentages in water (e.g., 70%, 80%, 96%, v/v), although some modifications have been described. These are:

- extracting in boiling 80% ethanol overnight and the following morning a further extraction for 5 min with boiling 20% ethanol;
- three extractions for 5 min with 80% boiling ethanol;
- extraction in hot 80% methanol and then extraction by boiling in 80% methanol at 60°C; and

- extraction in cold 80% ethanol stored at  $-13^{\circ}\text{C}$  for 14 h to precipitate insoluble material.

After centrifugation, the supernatant is heated at  $60^{\circ}\text{C}$  for 20 min and then dried. To the dry residue, cold 80% ethanol is added, and then heated again to dryness. The procedure is repeated three times under the same conditions.

For extraction of mannitol, sorbitol, and inositol from fruit pulps, homogenization and extraction with 80% methanol ensures the best recovery; 50% ethanol provides higher yield than 89% ethanol. After 15 min refluxing, the supernatant can be filtered. Extraction of pinnitol, mannitol, sorbitol, and inositol from Australian plants is performed with 5% perchloric acid at  $4^{\circ}\text{C}$  and then adjusted to pH 3–3.5 with  $1.0\text{ mol l}^{-1}$  potassium carbonate. Samples are purified by using Sep-Pak  $\text{C}_{18}$  cartridges preconditioned with methanol and Milli-Q purified water; the loaded cartridges are washed with deionized water and then 30% (v/v) acetonitrile in  $1.0\text{ mol l}^{-1}$  hydrochloric acid is passed through the cartridge and collected. In order to avoid acid hydrolysis of sucrose in fructose and glucose, samples are dissolved in 50% (v/v) ethanol together with  $0.1\text{ mol l}^{-1}$  imidazole buffer, pH 7.0.

Another procedure for extraction of polyols consists of grinding the tissue for 1 min at 20 000 rpm. The homogenate is centrifuged at 20 000g at  $4^{\circ}\text{C}$  for 10 min and the supernatant solution passed through a column ( $1 \times 2\text{ cm}$ ) of a styrene-divinylbenzene resin, strongly basic cation exchanger with active sulfonic groups, 200–400 mesh ( $\text{H}^{+}$ ), and after this another column ( $1 \times 2\text{ cm}$ ) of the same resin, 200–400 mesh ( $\text{Cl}^{-}$ ). The residue obtained by evaporation of the eluate is dissolved in a solvent suitable for chromatography.

For extraction of inositol phosphates from nonradioactively labeled cell or tissue specimens, cell suspensions are directly mixed with ice-cold perchloric acid to give a final acid concentration of  $0.5\text{ mol l}^{-1}$ . This extract should be free of sulfate and low in phosphate, ethyleneglycol bis(2-aminoethyl ether)- $\text{N,N,N',N'}$ -tetraacetic acid and other multicharge anions. Tissue specimens are either freeze-clamped or focused-microwave-irradiated. Frozen specimens are powdered in a liquid- $\text{N}_2$ -cooled steel-ball mill and the frozen powder cooled on ice. They are homogenized in 2 ml of ice-cold  $0.5\text{ mol l}^{-1}$  perchloric acid for 20 s. After removal of the precipitate formed after 20 min on ice by centrifugation, the 5000g supernatant is adjusted to pH 5 by adding potassium hydroxide, and the potassium perchlorate precipitate formed after 20 min on ice is removed by centrifugation. Nucleotides are removed by charcoal treatment.



If samples contain many salts, these are diluted with water and applied to a column containing 0.5 ml of Q-Sepharose (adjusted to the chloride form). After washing twice with 4 ml of  $2.5 \text{ mmol l}^{-1}$  hydrochloric acid, inositol phosphates are eluted with  $2 \times 2.5 \text{ ml}$  of  $0.6 \text{ mol l}^{-1}$  hydrochloric acid. The eluate is frozen and freeze-dried to remove the acid. Dried sample is dissolved in  $2.2 \text{ ml}$  of  $5 \text{ mmol l}^{-1}$  sodium acetate, pH 5.0. For subsequent  $^1\text{H}$  nuclear magnetic resonance (NMR) and  $^{13}\text{C}$  NMR spectra, dried extracts containing polyols are dissolved in  $0.7 \text{ ml}$  of deuterium oxide (99.996% deuteration) and the pH value adjusted to either 6.0 or 9.0 by adding deuterated formic acid and deuterated ammonia.

### Separation Methods

Chromatographic and electrophoretic methods are used for the isolation of sugar alcohols.

**Paper chromatography (PC)** Owing to the similar  $R_f$  values of monosaccharides and sugar alcohols, the distinction between these groups and between the stereoisomeric sugar alcohols may be very difficult. Chromatographic separation of the stereoisomeric sugar alcohols is achieved using solvents allowing the formation of ionic complexes of sugar alcohols with borate. For paper chromatograms, special filter paper (Whatman) is used. Solvent systems are listed in Table 1. Isopropanol–pyridine–water–acetic acid solvent has the useful property of migrating sugars, polyols, uronic acids, purine, and pyrimidine ribosides and their phosphate esters all within the  $R_f$  range of 0.1–0.9.

**Paper electrophoresis (PE)** Prior to PE separation of sugar alcohols into tetrityls, pentitols, etc., separation by paper or thin-layer chromatography (TLC) is recommended. Basic lead acetate ( $0.178 \text{ mol l}^{-1}$ ) is the most useful electrolyte; the run takes 3 h at 800 V. Other systems used are  $0.05 \text{ mol l}^{-1}$  borate, run for 1.5 h at 500 V, or  $0.2 \text{ mol l}^{-1}$  calcium acetate in  $0.2 \text{ mol l}^{-1}$  acetic acid run for 2 h at 670 V.

**Thin-layer chromatography** The velocity of separation in TLC is superior to PC and the quality of separation is similar in both techniques. Resolution of isomers has not been very satisfactory with solvent systems such as *n*-butanol–acetic acid–water saturated with boric acid (9:6:3:1, v/v/v/v) employing silica gel-coated plates. Separation is improved when silica gel plates are sprayed with  $0.5 \text{ mol l}^{-1}$  sodium dihydrogen phosphate in 50% ethanol, allowed to dry for 15 min, and then activated by heating to  $110^\circ\text{C}$  for

**Table 1** Solvent systems used in the separation of sugar alcohols by PC

<i>Solvent system</i>	<i>Proportion (v/v)</i>
Ethyl methyl ketone–acetic acid–water saturated with boric acid	9:1:1
Ethyl acetate–pyridine–water saturated with boric acid	60:25:10
1-Butanol– $0.75 \text{ mol l}^{-1}$ boric acid	85:10
<i>n</i> -Butanol–acetic acid–water	4:1:5 or 5:1:2
<i>n</i> -Butanol–ethanol–water–concentrated $\text{NH}_4\text{OH}$	45:5:49:1
Ethyl acetate–acetic acid–water	14:3:3
<i>n</i> -Butanol–ethanol–water	4:1.1:1.9
Isopropyl alcohol–pyridine–water–acetic acid	8:8:4:1

60 min. The solvent system consists of isopropanol–acetone– $0.2 \text{ mol l}^{-1}$  lactic acid (60:30:10, v/v/v), which must be made up freshly.

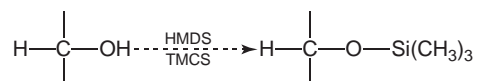
**‘Column’ chromatography** The technique involves elution of sugar alcohols from a column packed with charcoal, celite, cellulose, and other adsorbents. Solvent systems employed are similar to those described for PC.

Cation-exchange resins in the calcium or barium form have also been employed. A column of styrene–divinylbenzene resin, 8% cross-linked (220–400 mesh,  $\text{Ba}^{2+}$  form) can separate mannitol, sorbitol, and xylitol. Borate buffers are used as eluents from 5 to  $40 \text{ mmol l}^{-1}$ . Sugar alcohols can also be separated in Dowex 450W  $\times 2$  (200–400 mesh,  $\text{Ca}^{2+}$  form) columns by using only water as an eluent.

Recently, the use of a quite small column, namely  $14.0 \times 2.2 \text{ cm}$  of Dowex 50W  $\times 4$  in the  $\text{Nd}^{3+}$  form eluted with water at a rate of  $50 \text{ ml h}^{-1}$ , has been described for the separation of D-mannitol, D-glucitol, and D-iditol. The separation is thought to be caused by the different strength of complex formation between different polyols and metal cations that depends on the relative positions of the complexing hydroxyl groups.

**Gas–liquid chromatography** This technique allows the quantitative and qualitative estimation of sugar alcohols. It is limited by the volatility and stability of derivatives and by the accuracy with which quantitative conversion of polyols to their volatile derivatives can be achieved. The volatile derivatives of sugar alcohols commonly used are trimethylsilyl (TMS) ethers. To prepare TMS derivatives, samples are dried, dissolved in pyridine, and reacted by sequential addition of two volumes of hexamethyldisilazane (HMDS) and one volume of trimethylchlorosilane (TMCS). Other

silyl donor reagents can be used. All free hydroxyl groups are converted to the TMS ether as follows:



The reaction is carried out for 1 h at 70°C or for 24 h at room temperature. Since all solvents with free hydroxyl groups will react with HMDS and TMCS, it is necessary to exclude water at all stages of derivative preparation (i.e., by using anhydrous magnesium sulfate). The nonpolar phases give the best results for the separation of TMS derivatives of sugar alcohols. Stationary phases used include 2–5% silicon SE 30 or SE 52 (poly(dimethyl siloxane)), QF-1 (poly(methyltrifluoropropyl siloxane)), and EGS (poly(ethylene glycol succinate)) on uncoated adsorbent from diatomite (an acid washed and silanized diatomaceous earth support), acid-washed and dimethylchlorosilane-treated (AW DMCS). Examples of stationary phases are given in Appendix A.

Acetate derivatives of polyols show certain advantages for GC compared with TMS ethers: they are more stable, unaffected by water and other solvents containing free hydroxyl groups, and are freely soluble in the pyridine used for their preparation. Acetate derivatives are generally prepared by the action of acetic anhydride in the presence of a suitable catalyst. Reaction is carried out for 1–4 h at 70°C. Free hydroxyl groups are converted to the acetyl esters, so that substitution occurs on all carbon atoms of the polyols, and all but one of the aldoses. In the most widely used method, the sample is dried and treated with equal volumes of acetic anhydride and pyridine, which can be substituted with sulfuric acid and sodium acetate.

The best separations of polyols have been achieved on mixed phase packings, which combine the high resolving capacity of the polar phase with the stability of the nonpolar phase. Reported stationary phases are 2–10% EGSS-X (ethylene glycol succinate copolymerized with dimethyl polysiloxane) or ECNSS-M (ethylene glycol succinate copolymerized with cyanopropyl polysiloxane) on uncoated adsorbent from diatomite, acid washed, and dimethylchlorosilane-treated or uncoated adsorbent from diatomite acid-washed and silanized, 3% XE-60 (poly(methyl cyanopropyl siloxane)) on uncoated adsorbent from diatomite acid-washed and silanized, GP 3% SP-2340 (poly(dicyanopropyl siloxane)) on diatomite support and many other combinations.

Mannitol, sorbitol, and inositol separation is performed on a DB-5 or DB-1 (Durabon) fused silica capillary column of 30 m × 0.24 mm,  $d_f = 0.25 \mu\text{m}$ , with programmed temperature. These polyols can also be separated on fused silica columns coated with

CP-Sil-5CB (poly(dimethyl siloxane)), film thickness 0.12  $\mu\text{m}$ , and a temperature program without resorting to methoximation of sugars. Capillary gas chromatography with a phenyl (50%) dimethylpolysiloxane phase has been validated for the simultaneous determination of D-pinitol, *myo*-inositol, and their derivatives, as trimethylsilyl-derivatives.

Methyloxime-acetyl derivatives of sugar alcohols that avoid interference due to monosaccharides and resolve isomeric hexitols have been described. The procedure is as follows: 0.5 ml of methoxylamine hydrochloride in pyridine (10 mg ml<sup>-1</sup>) is added to dried samples containing polyols, and then incubated at 70°C for 30 min. The specimens are acetylated by adding 0.5 ml of acetic anhydride for a further 10 min at 70°C. Two drops of methanol are added and solvents evaporated in a water bath at 30°C under a stream of air. The residues are desiccated for at least 1 h and then dissolved in 50  $\mu\text{l}$  of methanol to be chromatographed. The column was filled with 3% XE-60 (80–100 mesh) and the chromatograph operated at a column temperature of 230°C with nitrogen as carrier gas at a flow of 50 ml min<sup>-1</sup>. **Figure 3** shows a chromatogram obtained using an aqueous solution containing 10 sugar alcohols.

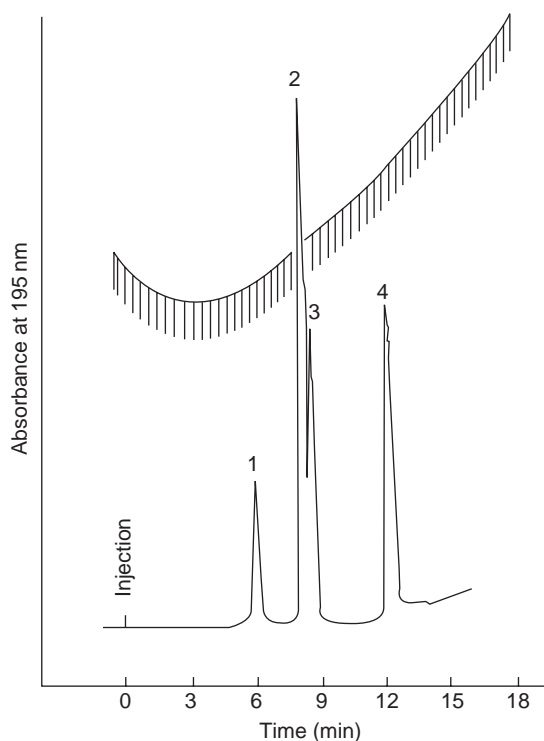
**Liquid chromatography** LC provides a rapid separation of sugar alcohols for analytical purposes without the need for sample derivatization. Extracts cannot be used directly in LC since substances such as proteins, lipids, and salts rapidly may inactivate the column packing; in any case, purification prior to chromatography must be carried out. Solutions can be deproteinized, e.g., by adding potassium hexacyanoferrate(III) (15%, w/v) and zinc acetate (30%, w/v) solutions. Alternatively, TCA can be used. The excess acid is subsequently removed by partitioning with six volumes of diethyl ether. The protein-free extracts are passed through a 0.22  $\mu\text{m}$  pore filter before chromatography.

Separation may be accomplished using adsorption chromatography, straight or reversed-phase partition chromatography, and ion-exchange (cation and anion) chromatography. Most useful for the separation of sugar alcohols are cation-exchange resins, using water or borate buffer as mobile phase. The optimum temperature range depends on the column, normally from 50°C to 80°C. **Figure 4** shows an LC separation of a standard mixture of three polyols.

The combined use of parabens as preservatives and the polyols as sweeteners or stabilizers is very extensive in the pharmaceutical, cosmetic, and food industries. However, it has been proved that these mixtures are not stable, because a slow *trans*-esterification reaction takes place, yielding different

degradation products, such as *p*-hydroxybenzoic acid, with lower preservative properties. Thus, the appearance of these degradation products should be controlled to guarantee the preservative capability of the paraben formulation. Sorbitol, which is one of these frequently used polyols, reacts with methylparaben (MPB) to form the ester sorbitol-paraben (SPB). MPB and SPB can also hydrolyze to form *p*-hydroxybenzoic acid and both reactions reduce the preservative capability of the formulation. An HPLC method by using  $C_{18}$  column with mobile phase of methanol–water (30:70, v/v) has been developed to separate the mixture of SPB.

From 1996 to the present, many stationary phases have been used. For the separation of sorbitol from other mono- and disaccharides, stationary phases prepared by the reaction of porous particles of chloromethylated styrene–divinylbenzene copolymer with various amines are preferred. Because of subtle differences in the  $pK_a$  values of hydroxyl groups, the oxyanions of carbohydrates generated under alkaline conditions interact with the positively charged stationary phase and facilitating their separation by the anion-exchange mechanism.



**Figure 4** HPLC separation of a standard mixture of 10 µg each of D-ribose, ribitol, arabinol, and mannitol on a MicroPak SP  $NH_2$ -5 (4 mm ID  $\times$  150 mm) column. Solvent, acetonitrile–water (80:20, v/v), isocratically. Flow rate, 1.3 ml min<sup>-1</sup>. Peaks: 1 = D-ribose as internal standard; 2 = ribitol; 3 = arabinol; 4 = mannitol. The line drawn across the chromatogram indicates baseline correction.

Widely utilized is the Dionex™ systems with an anion-exchange column, CarboPac PA1 and CarboPac MA1 (4  $\times$  250 mm) with a guard column (4  $\times$  50 mm). By using CarboPac MA1 and 16 mmol l<sup>-1</sup> NaOH as eluent, 30 different carbohydrates and alcohols can be separated: their migration times for alditols ranging from 10.7 min of erythritol to 27.7 min for maltitol. This column is able to separate a more wide variety of alcohols, sugar alcohols, and carbohydrates than the CarboPac PA1 optim, although retention times in the latter column are much lower.

Sorbitol is separated on a cationic exchange column (Sugar-Pak I), using 0.01 mol l<sup>-1</sup> potassium dihydrogen phosphate, pH 2.6, in isocratic mode, as mobile phase. Maltitol, *iso*-maltitol, and lactitol are separated by high-pH anion-exchange chromatography using a CarboPac PA100 with guard column using 40 mmol l<sup>-1</sup> sodium hydroxide and 1.0 mmol l<sup>-1</sup> barium acetate as the mobile phase (see Table 2). A new sulfonated monodisperse resin-based column (PL Hi-Plex) in protonated form has been developed and it is especially useful for profiling both carbohydrates and organic acids by combining mechanisms of ion exclusion and partition. Mannitol, sorbitol, maltitol, maltotriitol, lactitol, xylitol, and maltotetraitol have been separated on this column.

**Capillary electrophoresis** CE is a powerful separation technique that can provide high-resolution efficiency. One methodological difference between CE and LC is that CE utilizes an open tubular capillary. In CE, even if the samples contain a matrix, it can be injected with minimal sample preparation. In CE, ionic species are separated based on their charge and size; therefore, the electrolyte pH can be an important factor in optimizing resolution. Since  $pK_a$  values of neutral sugars and sugar alcohols are high (from 12 to almost 14), strongly alkaline conditions must be used to ensure ionization. Borate buffer from 20 to 75 mmol l<sup>-1</sup>, at pH values ranging from 10 to 13 has been used as BGE. Mannitol, sorbitol, galactitol, xylitol, and inositol are separated from other acidic, neutral, and amino sugars by CE using 2,6-pyridine dicarboxylic acid (PDC) and cetyltrimethyl ammonium bromide (CTAB) at pH 12.1 as electrolyte to reverse the direction of electroosmotic flow. Cetyltrimethyl ammonium hydroxide (CTAH) can also be used instead of CTAB.

**Other separation methods** Sorbitol and their paraben derivatives have been separated by micellar electrokinetic chromatography by using 25 mmol l<sup>-1</sup> phosphate buffer, pH 7.0, containing 100 mmol l<sup>-1</sup>

**Table 2** Mobile phase composition and elution mode of some sugar alcohols analyzed by anion-exchange chromatography-pulsed amperometric detection

Sugar compounds	Mobile phase composition and elution mode
Monosaccharides, alditols	8.0 mmol l <sup>-1</sup> Ba(OH) <sub>2</sub> and 0.5 mmol l <sup>-1</sup> HOAc, and 1.0 mmol l <sup>-1</sup> Ba(OH) <sub>2</sub> and 0.125 mmol l <sup>-1</sup> HOAc
<i>myo</i> -Inositol	1.0 mol l <sup>-1</sup> NaOH isocratic
Mannitol and glucose	480 mmol l <sup>-1</sup> NaOH
Mono- and disaccharides, alditols, aminosugars	5, 10, or 20 mmol l <sup>-1</sup> NaOH and 1–2 mmol l <sup>-1</sup> Ba(OAc) <sub>2</sub>
Alditols, mono-, and disaccharides	Water and postcolumn base addition
Alditols	0.5 mol l <sup>-1</sup> NaOH
Ribitol	650 mmol l <sup>-1</sup> NaOH
Alditols and sugars	0.6 or 0.45 mol l <sup>-1</sup> NaOH
Alditols and sugars	0.5 mol l <sup>-1</sup> NaOH and 1.0 mmol l <sup>-1</sup> Sr(OAc) <sub>2</sub> , Ba(OAc) <sub>2</sub> , or Ca(OAc) <sub>2</sub>
Alditols and sugars	0.58 mol l <sup>-1</sup> NaOH and 2.0 mmol l <sup>-1</sup> Ba(OAc) <sub>2</sub>
Alditols of mono- and disaccharides	0.5–0.6 mol l <sup>-1</sup> NaOH; 40 mmol l <sup>-1</sup> NaOH, and Ba(OAc) <sub>2</sub> ; 75 mmol l <sup>-1</sup> NaOH and Ba(OAc) <sub>2</sub>
Monosaccharides and alditols	NaOH gradient
Alditols, mono-, and oligosaccharides	0.1 mol l <sup>-1</sup> NaOH and NaOH and acetate gradient oligosaccharides

sodium dodecyl sulfate, over the critical micellar concentration, as BGE.

Capillary isotachopheresis (ITP) method with conductimetric detection has been proposed for separating polyols, such as mannitol, sorbitol, dulcitol, and xylitol. To facilitate the electromigration of the polyols, the strategy based on their conversion to anionic species by complex formation with B(III) as the central ion was adopted. The hydroxyl groups of the polyols react rapidly with borate to give anionic complexes that could presumably be separated by ITP. The complex-forming agent, i.e., boric acid, serves as the terminator and thus, *in situ* conversion of the neutral analytes to ionic species took place only during ITP analysis. The operational electrode system consisted of 10 mol l<sup>-1</sup> hydrochloric acid and 20 mmol l<sup>-1</sup> imidazole, pH 7.0, as the leading electrolyte, and 20 mmol l<sup>-1</sup> boric acid, pH 8.0, as the terminating electrolyte.

Subcritical fluid chromatography (SubFC) analysis of nine monosaccharides and three polyols, *meso*-erythritol, xylitol, and mannitol, on silica phases has been developed. Mobile phase composition was CO<sub>2</sub> modifier (80:20, v/v), the modifier being MeOH–water–triethylamine (91.5:8.0:0.5, v/v/v). In SubFC, the addition of triethylamine slightly increase the retention since interactions with the solutes and/or a strong interaction with the silanols and the residual silanols can occur.

## Detection Methods

**Spectrophotometric detection** Polyols can be determined by estimating the formaldehyde produced after mild periodate oxidation using arsenic(III) oxide

to destroy the excess periodate before addition of the chromogenic solution. Periodate oxidation is carried out at pH 4.5 for exactly 1 min, the periodate consumption being measured spectrophotometrically.

**Detection of spots on paper and thin-layer chromatograms** The similar *R<sub>f</sub>* values of monosaccharides and sugar alcohols in PC and TLC necessitate the use of specific detection reagents for sugar alcohols, i.e., benzidine–sodium periodate or reaction of monosaccharides with triphenyltetrazolium chloride prior to chromatography. Polyols can be identified with alkaline silver nitrate after composing the boric acid–sugar complex (if necessary) with hydrofluoric acid. Another alternative is the identification of polyols with *p*-anisidine phosphate, a characteristic reagent for sugars. Polyols appear colorless while spots of sugars are colored if brown cellulose paper previously dyed by the action of the acidic reagent is used. Mobilities and color of spots of some polyols are listed in Table 3. For PE, the paper strips are dried at 100°C and the color of sugar alcohols developed with 2.5% (w/v) potassium permanganate–chromium trioxide reagent.

Thin-layer plates may be developed by spraying with aniline–diphenylamine–acetone–80% phosphoric acid (4 ml:4 g:200 ml:30 ml) followed by heating at 105°C for 30 min. The combination of the <sup>14</sup>C isotope technique with PC is still one of the best methods, not only for the detection of sugar alcohols but also for the determination of their turnover and elucidation of their biosynthesis. Radioactive spots are eluted from chromatograms, evaporated on to planchettes, and the radioactivity measured with a counter.

**Table 3** Movement and detection of polyols in PC with Whatman No. 4 paper after 2 h at 30°C with isopropanol–pyridine–water–acetic acid (8:8:4:1, v/v/v/v) as mobile phase and separation by TLC in layers of phosphate-impregnated SilicaGel 60

Polyol	Reagent solution		$R_f^a$	$R_f$ in TLC
	$5 \text{ mol l}^{-1}$ periodic acid followed by $10 \text{ mol l}^{-1}$ benzidine	$0.01 \text{ mol l}^{-1}$ potassium permanganate		
meso-Inositol	W <sup>b</sup>	Y <sup>b</sup>	0.31	–
Galactitol	Y	Y	0.51	0.25
Mannitol	Y	Y	0.60	0.30
D-Sorbitol	W	Y	0.62	0.21
Pinitol	Y	Y	0.67	–
Arabitol	Y	Y	0.70	0.47
Ribitol	Y	W	0.70	0.53
Glycerol	W	Y	0.79	0.73
Erythritol	–	–	–	0.64
Threitol	–	–	–	0.61
Xylitol	–	–	–	0.38
Volemitol	–	–	–	0.18
Perseitol	–	–	–	0.13

<sup>a</sup>PC results, modified from Gordon HT, Thornburg W, and Werum LN (1956) Rapid paper chromatography of carbohydrates and related compounds. *Analytical Chemistry* 5: 849–855. TLC results, modified from Richardson (1985).

<sup>b</sup>Colors of spots: W = white; Y = yellow.

**Detection in GC** The universal detector for TMS and acetate derivatives of sugar alcohols is the flame ionization detector (FID). Linear response of the FID has been established for such an analysis. Detector temperatures in the gas chromatograph vary from 250°C to 300°C.

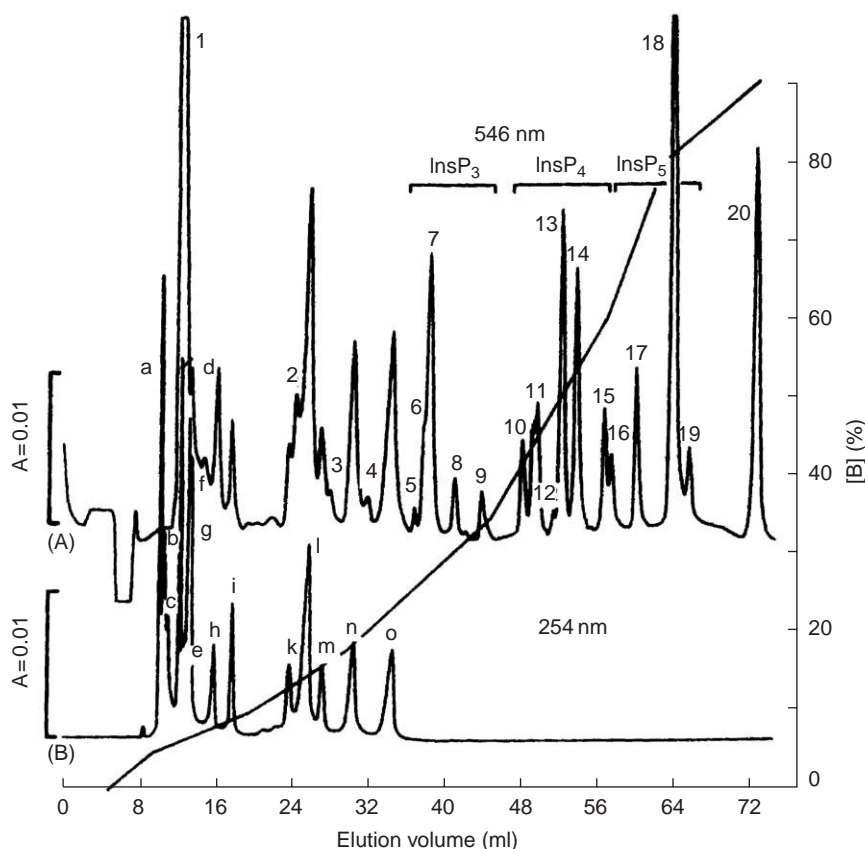
GC–MS methods for the simultaneous quantitation of sugars, sugar alcohols, and acids, measures as their TMS-oxyme ether/ester derivatives, has been developed. The reproducible determination of analytes was performed both on the basis of total ion current and selected fragment ions values.

**Detection in LC** Carbohydrates, as well as carboxylic acids, are difficult to detect by conventional spectrophotometric methods because they lack suitable chromophores or fluorescent groups. Sugar alcohols can be detected by refractive index change or by postcolumn reaction. Refractivity measurement is not specific for sugar alcohols and detectors are incompatible with gradient programs of the mobile phase. Detection of polyols can be achieved using an UV absorbance detector at 195 nm, which improves the sensitivity of polyol determination. In this case, it is absolutely necessary to use the purest mobile phases that are possible. Precolumn derivatization of sorbitol and galactitol with phenylisothiocyanate gives UV-absorbing derivatives with an absorbance maximum at 240 nm. A nonradiometric LC technique, employing dephosphorylation in an enzyme-loaded postcolumn reactor and subsequent orthophosphate

determination has been described for inositol polyphosphate isomers, based on molybdophosphate formation but requires relatively large amounts of tissue. A new dye-based ternary-complexometric technique not requiring dephosphorylation is based on the finding that transition-metal(III) ions bind with very high affinity to both 4-(2-pyridylazo)-resorcinol (PAR) and polyanions like inositol phosphates. In this two-ligand one-metal system, the dye functions as a reporter substance, indicating optically the presence of competitive ligands. This technique has been termed ‘metal-dye detection’. **Figure 5** shows a separation of a standard mixture of nucleotides and inositol polyphosphates by LC with metal-dye detection.

The tendency of polyols to form chelate complexes with some central ions, such as B(III), Ge(IV), or Mo(VI) is of analytical importance since the complex formation is accompanied by the liberation of protons from the polyol molecule that are normally not split off in aqueous medium. The pH changes occurring due to complex formation has been utilized in the potentiometric or spectrophotometric determination.

Electrochemical detection (potentiometry and amperometry) has been recognized as a useful method following separation by LC. A number of methods based on electrochemical oxidation using various metallic electrodes has been reported. Foiling of gold and platinum electrodes may be overcome by pulsed amperometric detection, which combines amperometric detection with alternate anodic and cathodic



**Figure 5** HPLC-m.d.d. analysis of a standard mixture of nucleotides and  $\text{InsP}_x$ . Separation was performed on a  $15\text{ cm} \times 0.5\text{ cm}$  Mono Q column. Eluent A and B contained  $9$  and  $14\text{ }\mu\text{mol l}^{-1}$   $\text{YCl}_3$ , respectively, and reagent C contained  $200\text{ }\mu\text{mol l}^{-1}$  PAR. The flow rate was  $1.2\text{ ml min}^{-1}$  for A/B, and  $0.6\text{ ml min}^{-1}$  for C. The gradient applied is depicted. The upper monitor tracing (A) was obtained by m.d.d. at  $546\text{ nm}$ , the lower one (B) by UV detection. The components of the nucleotide mixture are given in the lower tracing. The mixture of  $\text{InsP}_x$  was from partly hydrolyzed  $\text{InsP}_6$ . An equivalent of  $10\text{ nmol}$  of  $\text{InsP}_x$  was injected. Assignments are in part based on pure standard available. Assignments of isomers to individual peaks thus possible are indicated by an asterisk:  $1 = \text{P}_i + \text{InsP}$ ;  $2 = \text{Ins}(1,2)\text{P}_2^* + \text{Ins}(1,6)\text{P}_2^*$ ;  $3 = \text{PP}_i$ ;  $4 = \text{unidentified}$ ;  $5 = \text{Ins}(1,3,5)\text{P}_3^* + \text{Ins}(2,4,6)\text{P}_3^*$ ;  $6 = \text{Ins}(1,3,4)\text{P}_3^*$ ;  $7 = \text{Ins}(1,2,3)\text{P}_3^* + \text{Ins}(1,2,6)\text{P}_3^* + \text{Ins}(1,4,5)\text{P}_3^*$  preceding  $\text{Ins}(2,4,5)\text{P}_3^*$ ;  $8 = \text{Ins}(1,5,6)\text{P}_3^*$ ;  $9 = \text{Ins}(4,5,6)\text{P}_3^*$ ;  $10 = \text{Ins}(1,2,3,5)\text{P}_4^* + \text{Ins}(1,2,4,6)\text{P}_4^*$ ;  $11 = \text{Ins}(1,2,3,4)\text{P}_4^* + \text{Ins}(1,3,4,6)\text{P}_4^*$ ;  $12 = \text{Ins}(1,3,4,5)\text{P}_4^*$ ;  $13 = \text{Ins}(1,2,5,6)\text{P}_4^*$ ;  $14 = \text{Ins}(2,4,5,6)\text{P}_4^*$ ;  $15 = \text{Ins}(1,4,5,6)\text{P}_4^*$ ;  $16 = \text{Ins}(1,2,3,4,6)\text{P}_5^*$ ;  $17 = \text{Ins}(1,2,3,4,5)\text{P}_5^*$ ;  $18 = \text{Ins}(1,2,4,5,6)\text{P}_5^*$ ;  $19 = \text{Ins}(1,3,4,5,6)\text{P}_5^*$ ;  $20 = \text{InsP}_6$ . a = AMP + CMP + NAD; b = cyclic AMP; c = GMP; d = NADH; e = UMP; f = NADP; g = ADP + ADP-ribose; h = ATP + CTP; i = GDP; k = IDP; l = GTP; m = UDP; n = ITP; o = UTP. Elution volumes of bisphosphates in the  $23\text{--}33\text{ ml}$  range (not all included in this chromatogram) were (in milliliter):  $\text{Ins}(1,3)\text{P}_2^* = 23.28$ ;  $\text{Ins}(1,6)\text{P}_2^* = 23.45$ ;  $\text{Ins}(1,4)\text{P}_2^* = 23.95$ ;  $\text{Ins}(1,5)\text{P}_2^* = 23.95$ ;  $\text{Ins}(1,2)\text{P}_2^* = 23.97$ ; sedoheptulose-1,7-bisphosphate =  $24.00$ ; glucose 1,6-bisphosphate =  $24.07$ ;  $\text{Ins}(4,5)\text{P}_2 = 24.34$ ;  $\text{Ins}(2,4)\text{P}_2 = 24.35$ ; fructose 1,6-bisphosphate =  $25.33$ ;  $\text{PP}_i = 27.87$ ; 2,3-BPG =  $32.29$ .

polarization to clean and reactivate the electrode surface. However, constant potential amperometric is preferred because of its instrumental simplicity. Metals such as Cu, Ni, and Ag, as well as chemical modified electrodes, have been developed for the electrocatalytic oxidation of carbohydrates in alkaline media. Mobile phases consist of high concentration of sodium hydroxide ( $0.1\text{--}0.7\text{ mol l}^{-1}$ ), isocratically or in gradient with water containing or not containing the same modifiers. Eluent modification with barium ions enhances both the amperometric response and the chromatographic data reliability without column regeneration and postcolumn addition of strong bases. The cobalt-modified

electrode, prepared by anodic electrodeposition of cobalt oxyhydroxide on the polished glassy carbon surface by voltage cycling, has been used to detect inositol, xylitol, sorbitol, and mannitol from complex mixtures of carbohydrates. It is known that in alkaline medium the high valence states of the cobalt oxyhydroxides show interesting electrochemical activity toward the electrooxidation of several polyhydric organic compounds and, thus, catalytic oxidation of alditols occurs at potentials higher than  $0.3\text{ V}$  versus SCE.

Cooper wire electrodes, used for amperometric and potentiometric detection in HPLC, have been developed for the simultaneous detection of sugars,

polyols, and carboxylic acids. Diluted solutions of acids on water have been used as the eluent for the separation of carbohydrates and carboxylic acids in ion-exclusion chromatography and as the carrier for the potentiometric detection, with postcolumn addition of sodium hydroxide as the electrolyte for the amperometric detector.

Precolumn derivatization of sugar alcohols with a chromophoric reagent, such as phenylisocyanate, benzoate, and 2,4-dinitrobenzoate allows the use of inexpensive UV detectors. *p*-Nitrobenzoyl chloride was found to be a rapid and quantitative derivatizing agent for sugar alcohols since strong UV-absorbing derivatives at 260 nm are produced.

**Detection in CE** Since carbohydrates lack both a charge and a strong UV chromophore, several derivatization techniques have been described. Very recently, a new method based on precapillary derivatization with luminol (3-amino-phthalhydrazide) for carbohydrate analysis has been proposed with online chemiluminescence detection. While this method leads to improved sensitivity and resolution, the complexity of derivatization limits its use. Alternatively, methods for the analysis of underivatized carbohydrates have been developed. These methods include the use of high alkaline electrolyte to ionize carbohydrates and make them suitable for indirect UV detection. Mannitol, sorbitol, galactitol, xylitol, and inositol have been detected with indirect UV detection using PDC and CTAB/CTAH as electrolytes.

### Identification Methods

In the analytical procedures described above, spot or peak identification is based on a comparison of  $R_f$  or retention time of unknown spots or peaks with those of authentic compounds, followed by cochromatography to obtain coincident spots or peaks on the chromatogram. Further analyses using different stationary phases and derivatives provide unique combinations of retention values based on various physical and chemical characteristics. MS offers a good potential when coupled to GC.

The structure of some universal polyols can also be confirmed by NMR techniques, involving  $^{13}\text{C}$  and  $^{31}\text{P}$  NMR. *myo*-Inositol-1,3,4,5,6-pentakisphosphate has been identified by these techniques in avian erythrocytes after separation of inositol polyphosphates by Dowex 1  $\times$  2 chromatography. Arabitol, mannitol, and ribitol have also been identified using natural abundance criteria by  $^{13}\text{C}$  NMR spectroscopy in a number of lichen species.

### Areas of Special Interest

Acyclic sugar alcohols are considered as the most important agents for osmosis in plants, related to the maintenance of the water potential in many lower and higher plants living in arid conditions. Many lichens accumulate erythritol, mannitol, arabitol, or sorbitol as the main system to maintain a convenient, residual amount of water to survive. However, a balanced [polyol]/[hexose] ratio is required to avoid catabolite repression of many enzymes related to amino acids and phenolics metabolism.

In plants and animals, phosphoinositides are involved in the permeation mechanism for calcium ions. This mechanism is the basis of membrane potential changes that produce signal transduction and transmission of electrical stimuli. General physiological functions, such as neural transmission, memory, and vision in animals, or phytochrome-regulated morphogenesis in plants, are started by a cascade process involving phosphoinositide metabolism. For example, hydrolysis of phosphatidylinositol-(4,5)-bisphosphate (PTDIns(4,5) $\text{P}_2$ ) leads to inositol-(1,4,5)-triphosphate (Ins(1,4,5)- $\text{P}_3$ ) and diacylglycerol, the latter well known as an activator of protein kinases. Also calcium ions, released from their intracellular stores after the binding of Ins(1,4,5) $\text{P}_3$  to its endoplasmic reticulum receptor, activates protein kinases.

Sugar alcohols are widespread in food and beverages because they contribute to the nutrition value, flavor, and organoleptic characteristics of foods. Polyols have reduced cariogenic properties. The low hygroscopicity and the high stability of polyols have provoked a growing interest in food industry. Polyols are often claimed to be useful sweeteners, specially for diabetics. However, this is not true; polyols affect the blood glucose level (through slower than sucrose). Besides polyols contain calories; they have energy values of  $10\text{ kJ g}^{-1}$  on average, compared to  $17\text{ kJ g}^{-1}$  of sucrose and, therefore, must be accounted for the meal planning of diabetics. The polyols are often used in confectionery products, such as chewing gum pastilles, chocolate, candy, sugar-free candy, and cakes. Xylitol, sorbitol, mannitol, isomalt, and maltitol are the main polyols used in food products. Mannitol, sorbitol, and xylitol are pharmaceutically important nonionic osmotic diuretics.

### Sources of Error

Analysis of complex mixtures containing several polyols requires some attention to the possibility of analyte loss. Polyols have different solubilities in solvent currently used in extraction and



chromatography. Most acyclic polyols are freely soluble in water whereas the water solubility of inositol and inositol derivatives is limited. For example, a saturated solution of sorbitol in water contains ~89% (w/v) of this polyol whereas the limit of inositol solubility in water at 25°C is 140 g l<sup>-1</sup>. By contrast, inositol and inositol monophosphate are practically insoluble in pure ethanol. However, phytic acid is slightly soluble in pure ethanol and methanol whereas sorbitol is quite soluble in hot ethanol and sparingly soluble in cold alcohol.

All of polyols are very stable in the cold even when they are mixed with diluted acids, alkalis, or mild oxidizing agents. Inositol-1-phosphate is remarkably resistant to hydrolysis by boiling with strong alkali, although it can be hydrolyzed by boiling with 6 mol l<sup>-1</sup> hydrochloric acid for 14 h. Phytic acid decomposes slightly on heating.

**See also:** **Carbohydrates:** Overview. **Derivatization of Analytes.** **Extraction:** Solvent Extraction Principles. **Gas Chromatography:** Column Technology; Detectors; Mass Spectrometry. **Ion Exchange:** Ion Chromatography Instrumentation. **Liquid Chromatography:** Overview; Column Technology. **Nuclear Magnetic Resonance Spectroscopy – Applicable Elements:** Phosphorus-31. **Phosphorus.** **Sample Handling:** Comminution of Samples. **Thin-Layer Chromatography:** Overview.

## Further Reading

- Beck E and Hopf H (1990) Branched-chain sugars and sugar alcohols. In: Dey PM (ed.) *Methods in Plant Biochemistry. Vol. 2. Carbohydrates*, pp. 235–289. London: Academic Press.
- Holligan PM and Drew EA (1971) Routine analysis by gas-liquid chromatography of soluble carbohydrates in

extracts of plant tissues. II. Quantitative analysis of standard carbohydrates and the separation and estimation of soluble sugars and polyols from a variety of plant tissues. *New Phytologist* 70: 271–297.

Lamari FN, Kuhn R, and Karamanos NK (2003) Derivatization of carbohydrates for chromatographic, electrophoretic and mass spectrometric structure analysis. *Journal of Chromatography B* 793: 15–36.

Legaz ME, Pedrosa MM, and Vicente C (1998) Carbohydrates, polymeric sugars and their constituents. In: Deyl Z, Mikšik I, Tagliaro F, and Tesařová E (eds.) *Advanced Chromatographic and Electromigration Methods in Biosciences*, pp. 257–314. Amsterdam: Elsevier.

Loewus FA (1990) Cyclitols. In: Dey PM (ed.) *Methods in Plant Biochemistry. Vol. 2. Carbohydrates*, pp. 219–233. London: Academic Press.

Mayr GW (1988) A novel metal-dye detection system permits picomolar-range h.p.l.c. analysis of inositol polyphosphates from non-radioactively labelled cell or tissue specimens. *Biochemical Journal* 254: 585–591.

Mount JN and Laker MF (1981) Estimation of sugar alcohols by gas-liquid chromatography using a modified acetylation procedure. *Journal of Chromatography* 226: 191–197.

Que AH and Novotny MV (2002) Separation of neutral saccharide mixtures with capillary electrochromatography using hydrophilic monolithic columns. *Analytical Chemistry* 74: 5184–5191.

Rotzsche H (1991) *Stationary Phases in Chromatography*. Amsterdam: Elsevier.

Tommaso R, Cataldi I, Campa C, and De Benedetto GE (2000) Carbohydrate analysis by high-performance anion-exchange chromatography with pulsed amperometric detection: the potential is still growing. *Fresenius Journal of Analytical Chemistry* 368: 739–758.

Vicente C, Mateos JL, Pedrosa MM, and Legaz ME (1991) High-performance liquid chromatographic determination of sugars and polyols in extracts of lichens and sugarcane juice. *Journal of Chromatography* 553: 271–283.

## Appendix A

	Trivial name	Chemical name
2–5% silicon SE 30 or SE 52	Methyl silicone	Poly(dimethyl siloxane)
QF-1	Methyl fluorosilicone	Poly(methyltrifluoropropyl siloxane)
EGS	Ethylene glycol succinate	Poly(ethylene glycol succinate)
2–10% EGSS-X	Ethylene glycol succinate methyl-siloxane copolymer-type X	Ethylene glycol succinate copolymerised with dimethyl polysiloxane
2–10% ECNSS-M	Ethylene glycol succinate cyano-siloxane copolymer-type M	Ethylene glycol succinate copolymerised with cyanoethyl methyl polysiloxane
3% XE-60	Methyl cyanopropyl silicone	Poly(methyl cyanopropyl siloxane)
GP 3% SP-2340	Cyanopropyl silicone	Poly(dicyanopropyl siloxane)
CP-Sil-5CB	Methyl silicone	Poly(dimethyl siloxane)

## Starch

**G Mitchell**, Vilvoorde Research and Development Centre, Vilvoorde, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 511–518, © 1995, Elsevier Ltd.

### Introduction

Starch is a polymer of glucose, existing in two basic configurations. Amylose is predominantly a linear chain polymer, composed mostly of  $\alpha$ -1,4 linkages. Amylopectin is a branched polymer containing  $\alpha$ -1,4 and  $\alpha$ -1,6 bonds. The amylose fraction also has some degree of branching. Starch is found widely distributed in nature and is present in high concentration in cereal grains, in potatoes, and in roots such as manioc. It is the major complex carbohydrate in the diet of humans. Processes have been developed on an industrial scale to separate starch from the starch-rich sources. Starch is the starting material for the manufacture of glucose syrups and modified starches, and this family of products finds wide application in food and industrial use.

In every area of its use and manufacture, a need exists for measurement of starch. The industrial preparation of starch is based on wet-milling, and the efficiency of the process determines the purity of the final product. Analysis of the starting raw material, the starch, all by-products, and the factory wastewater, is necessary to assess the separation procedure. Starchy foods are favored by nutritionists, and labeling of foods with compositional data becomes the norm.

### Polarimetric Methods

Dissolution of starch can be achieved by various means, but there is usually associated hydrolysis. Once a solution is obtained, then the optical activity of the starch can be utilized for quantification. Two long-established methods are in common use.

#### Ewers Method

The first is based on controlled partial hydrolysis of the starch with hydrochloric acid, and was proposed in 1908 by Ewers. First, 2.5 g of sample is added to a 100 ml volumetric flask and dispersed in 50 ml of 1.128%, m/v, hydrochloric acid. The flask is plunged into a vigorously boiling water bath, with agitation,

and after 15 min precisely, the contents are diluted, and the flask is cooled in an ice bath. Before making-up, protein is precipitated with Carrez reagent (zinc acetate + potassium hexacyanoferrate(II)). Following filtration, the optical rotation is measured. To account for 'nonstarch' carbohydrate present, the sample is extracted with 40%, v/v, ethanol. This solution is hydrolyzed with acid, the optical rotation is measured, and its contribution is subtracted from the total analysis figure. Starches from different sources hydrolyze in acid at different rates, and Ewers had experimentally determined a factor for specific starch types.

The starch content of the sample is calculated according to eqn [1]:

$$\% \text{ Starch} = \frac{2000(P - P')}{[\alpha]_{\text{D}}^{20}} \quad [1]$$

with cell path length = 200 mm, sample weight = 2.500 g, and where  $P$  is the total optical rotation in degrees;  $P'$  is the optical rotation, in degrees, of substances soluble in 40% ethanol;  $[\alpha]_{\text{D}}^{20}$  is the specific optical rotation of the pure starch; and 2000 represents the diminution of the mathematical terms for the conditions of the analysis.

Conventionally accepted factors for different starch types are:

- Maize starch + 184.6°
- Wheat starch + 182.7°
- Potato starch + 185.4°
- Rice starch + 185.9°
- Barley starch + 181.5°
- Oat starch + 181.3°
- Other types of starch and mixtures + 184.0°

This method is widely used in Europe, and is proposed by many legislative bodies including the European Commission. It is an empirical method and therefore has the inherent obligation of strict adherence to the protocol, and the possibility of operator dependence.

#### Calcium Chloride Dissolution

An alternative method depends on the ability of concentrated solutions of electrolytes to solubilize starch. Clendenning in 1945 developed a method using a 40%, w/v, solution of calcium chloride, adjusted to pH 2.0 with acetic acid. Two grams of sample is added to an Erlenmeyer flask, dispersed in

10 ml of water, followed by addition of 60 ml of the salt solution. A few glass beads are included, and some drops of octanol are added to control foaming. A reflux condenser is fitted and the flask is heated strongly so as to reach boiling within 5 min. After 30 min of reflux, the flask is removed and cooled to room temperature. The whole content is transferred carefully to a 100 ml volumetric flask using the calcium chloride solution for all transfer and washing. Carrez reagent is added to precipitate protein, prior to final volume adjustment. Work done by Clendenning showed that a universal factor of  $203^\circ$  optical rotation could be attributed to all starches, regardless of source. Starch content is calculated accordingly to eqn [2]:

$$\% \text{ Starch} = \frac{P \times 100}{203 \times 2 \times m/100} \quad [2]$$

where  $P$  is the optical rotation of the sample, 2 is the cell path length (in dm), and  $m$  is the sample weight in grams.

While hydrolysis is less than in the Ewers method, it is still an empirical method, and some restrictions apply to both approaches. Wheat fractions with high gluten content cause excessive foaming in the calcium chloride method, and this is often impossible to diminish. In addition, both methods are suited to samples containing only granular native starch, and not gelatinized starch. Pentosans cause errors, having a negative influence on the calcium chloride method, and a positive effect with acid hydrolysis. This means that the methods are unsuitable for fractions containing wheat bran, which contains a high pentosan level. International collaborative exercises have found a tolerance of  $\sim 1.5\%$  on these two methods. Despite the drawbacks of these methods, starch manufacturers successfully use them for calculating mass balances.

## Hydrolysis Methods

### Acid Hydrolysis

Starch can be fully hydrolyzed to dextrose, which can be measured, with subsequent calculation of the original starch content. While the use of enzymes is probably the best means, acid hydrolysis is a simple alternative. Importantly, dilute acid should be used to avoid condensation of the dextrose. When 1 g of starch is refluxed for 3 h in  $0.25 \text{ mol l}^{-1}$  hydrochloric acid, the hydrolysis is complete; this has been verified by chromatography. An official method based on this approach uses the Lane and Eynon titration based on Fehling's solution to determine the dextrose. However, alternative methods for dextrose

quantification abound, and there is further discussion within this article. Some modified starches resist total hydrolysis, but the approach offers a simple fingerprint technique if liquid chromatography is applied to the hydrolysate.

### Enzyme Hydrolysis

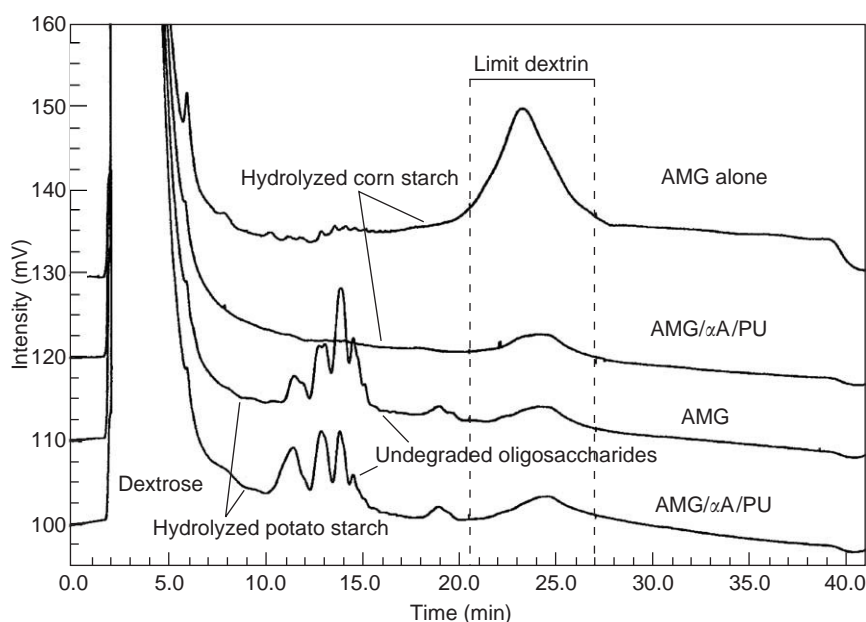
When using enzymes, there are three important stages in the method: (1) dissolution of the starch; (2) complete hydrolysis of the starch to dextrose (glucose); and (3) accurate measurement of the dextrose.

Satisfactory dissolution of the starch at atmospheric pressure is difficult to achieve and a thermostable  $\alpha$ -amylase is often included. Pressure cooking with an autoclave at  $130^\circ\text{C}$  is better, but such equipment is uncommon in laboratories. Dimethyl sulfoxide, with 10% water included, is an excellent solvent for starch, but this water content is critical, and the organic solvent is present during the subsequent enzyme hydrolysis. However, commercial kits are available that employ dimethyl sulfoxide as solvent. Acid may also be used for dissolution but, if a large amount is used, care must be taken that the starch does not 'salt-out' on neutralization. A most convenient solvent has proved to be cold dilute sodium hydroxide solution. The starch, 250 mg, is slurried in 25 ml of water, an equal volume of  $1 \text{ mol l}^{-1}$  sodium hydroxide is added and stirring is continued for 5 min. Before addition of enzymes the solution is adjusted to pH 4.5–5.0 with acetic acid.

To ensure full conversion to dextrose, the ideal conditions for hydrolysis must be chosen. This implies excess of enzyme, low starch concentration to avoid reversion of the dextrose formed, and optimal pH. Enzymes should be of the highest purity available.

Some methods available offer the use of amyloglucosidase alone. Work done with corn and potato starches illustrated that even under ideal conditions amyloglucosidase does not fully convert starch to dextrose, although the shortfall is small. The limit dextrin was more noticeable in corn starch hydrolysis. Proof was obtained by studying the reaction kinetics, and analyzing the hydrolysates by ion chromatography using pulsed amperometric detection. There remained always a small amount of a limit dextrin, and in the case of potato starch some other low molecular mass residues.

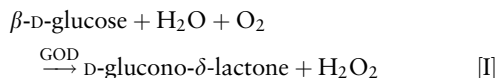
A study done within an International Standards Organization (ISO) work group has shown that the inclusion of pullulanase effectively eliminates the limit dextrin, as shown in **Figure 1**. Hydrolysis also appears accelerated by the presence of  $\alpha$ -amylase, although the pH (4.8–5.0) is not optimal. A mixture



**Figure 1** Expanded chromatograms from Dionex ion chromatograph of enzyme hydrolysates of corn and potato starch, the use of a single enzyme, amyloglucosidase (AMG), and a mixture of AMG,  $\alpha$ -amylase ( $\alpha$ -A), and pullulanase (PU), to show the effect on hydrolysis of limit dextrin. Anion exchange column AS6, with pulsed amperometric detection. Gradient flow solvents (1)  $150 \text{ mmol l}^{-1}$  NaOH and (2)  $150 \text{ mmol l}^{-1}$  NaOH +  $500 \text{ mmol l}^{-1}$  NaOOCCH<sub>3</sub>. Postcolumn addition of  $0.3 \text{ mmol l}^{-1}$  NaOH.

consisting of amyloglucosidase (500 U), pullulanase (0.3 U), and  $\alpha$ -amylase (1440 U) proved an adequate system for complete hydrolysis of the starch. The maximum amount of starch present should be 250 mg, and the enzymes are added immediately after neutralization of the alkali. This avoids the possibility of retrogradation of the starch. However, this combination did not fully hydrolyze potato starch, as up to 2% phosphate-containing oligosaccharides can remain in the hydrolysate.

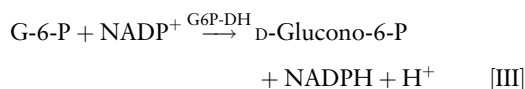
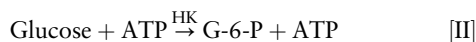
Various ways of measuring glucose exist. Probably the most often employed principle is the oxidation of glucose with glucose oxidase (GOD), to produce hydrogen peroxide and D-glucono- $\delta$ -lactone (reaction [I]):



Many methods are available in which the hydrogen peroxide, usually in the presence of peroxidase, takes part in reactions to form colored products. Typical reagents are a combination of phenol with either *p*-hydroxybenzoic acid or 4-aminopyrrole, or 2,2'-azino-bis-(3-ethyl-benzthiazoline-6-sulfonate) (ABTS) alone. Interference can be a problem; for example, ABTS is sensitive to the presence of proteins. Another drawback of colorimetric methods is that calibration with glucose is obligatory.

Hydrogen peroxide can be quantified electrochemically and a commercial instrument for glucose measurement is available, which consists of a cell containing GOD immobilized in a membrane. The hydrogen peroxide formed as the dextrose passes through the membrane diffuses to an anode where it is oxidized, and the resulting current produced is measured at the equilibrium point of the oxidation.

The most specific measurement of glucose is by a totally enzymatic procedure. D-Glucose is phosphorylated to glucose 6-phosphate (G-6-P) by adenosine 5'-triphosphate (ATP) in the presence of hexokinase (HK) (reaction [II]). Subsequently, G-6-P is oxidized by nicotinamide adenine dinucleotide phosphate (NADP) to gluconate 6-phosphate in the presence of glucose-6-phosphate dehydrogenase (G6P-DH). NADPH is formed in a stoichiometric amount with regard to the glucose present (reaction [III]). It can be quantified by absorbance at 340 nm:



All additions and reactions occur within the cuvette of the spectrophotometer, so the method is essentially semimicro. As sample size may be only

100  $\mu$ l, all micropipettes must be calibrated precisely. Better repeatability has been found by weighing all additions into the cuvette. While this procedure is theoretically superior to other approaches, its application needs the greatest care. Although the suppliers of kits quote a tolerance of  $\pm 1\%$ , this figure is optimistic, and is not derived from a collaborative study effected under ideal conditions for statistical evaluation.

With the advent of highly sensitive differential refractometers, liquid chromatography can be applied to the measurement of the glucose. Success has been achieved by acidifying the enzyme hydrolysate with sulfuric acid to inactivate the enzymes, and then chromatographing this solution directly on a cation-exchange column in the hydrogen form. An internal standard, *meso*-erythritol, should be included from the start of the analysis; baseline resolution of all components is found. This method was developed in an ISO work group, and a collaborative exercise performed on native starches yielded 98% recovery, with a tolerance of  $\pm 2\%$ .

## Choice of Method

Whereas the ideal situation is that a single method will suffice for all starch-bearing matrices, this is unlikely from a practical standpoint. For example, a customs laboratory needing daily to confirm the integrity of large series of native starches relies on official methods based on polarimetry; speed and reproducibility are the requirement. On samples that contain pregelatinized native starch, or other substances that are optically active, an enzymatic method must be used; specificity is the requirement.

The starch content of a natural product or of a man-made product may be required. The latter products might contain modified starches, and a limited number are permitted for food purposes; acetate, adipate, succinate, oxidized, hydroxypropyl. For nonfood applications cationic and carboxymethyl starches are in common use. Polarimetric methods, when applied to such modified starches, reveal that calcium chloride dissolution is better than the Ewers method. Even with cross-bonded starches, which are difficult to solubilize, complete dissolution is achieved within 30 min with calcium chloride solution. Recovery, based on a universal optical rotation of  $203^\circ$ , is over 95%. While this application of the method needs further verification, several laboratories in an ISO work group report similar findings.

The enzymatic methods are specific for native starch; modification usually hinders the enzyme action. However, the esters acetate, adipate, and

**Table 1** Summary of suitable methods for different starch types

Starch type	Polarimetric		Enzymatic
	Ewers	CaCl <sub>2</sub>	
<i>Granular</i>			
Native	+	—	+
Acid degraded	—	—	+
Modified:			
Acetate	—	—*	+**
Adipate			
Succinate			
Cationic	—	—*	—
Carboxymethyl			
Hydroxypropyl			
Oxidized			
Phosphate			
<i>Pregelatinized</i>			
Native	—	—	+
Acid degraded	—	—	+
Modified:			
Acetate	—	—	+**
Adipate			
Succinate			
Cationic	—	—	—
Carboxymethyl			
Hydroxypropyl			
Oxidized			
Phosphate			

+, suitable; —, not suitable; \*, better suited than Ewers, 95% recovery; \*\*, ester to be initially hydrolyzed.

succinate are very labile, and if the substituent is first hydrolyzed under alkaline conditions the starch content can be quantified enzymatically. A summary of the applicability of different methodology to the range of starches likely to be encountered is given in **Table 1**. Evidently, for certain pregelatinized modified starches, no method presently exists for quantification of the starch content.

An enzymatic approach is used to characterize dietary starch, which is observed as falling into different categories according to the speed of digestion in humans: rapidly digestible, slowly digestible, and resistant starch. Total starch is determined by enzyme hydrolysis after dissolution of the starch in potassium hydroxide. The digestible starch is determined by incubation with pancreatic enzymes at  $37^\circ\text{C}$ . Rapidly digested starch is calculated from glucose released after 20 min and slowly digested starch from glucose released after 2 h. Resistant starch is that which is not hydrolyzed within the 2 h.

## Starch Purity

The determination of starch purity, for certain economic reasons, became an important point of discussion in Europe. Whatever the source, starch

**Table 2** Typical analysis of starches from different sources<sup>a</sup>

Source	Moisture <sup>b</sup> (%)	Ash <sup>c</sup> (%)	Fat <sup>d</sup> (%)	Protein <sup>e</sup> (%)	Starch (by difference) (%)
Regular maize	13.13	0.06	0.68	0.37	98.9
Waxy maize	12.65	0.04	0.08	0.22	99.7
Potato	18.85	0.21	0.04	0.09	99.7
Wheat	12.24	0.15	0.56	0.30	99.0
Pea	8.06	0.03	0.09	0.27	99.6
Rice	11.74	0.25	0.71	0.52	98.5

<sup>a</sup>All values are % on dry substance.<sup>b</sup>Moisture determined by drying for 1.5 h at 130°C, atmospheric pressure.<sup>c</sup>Ash determined by pyrolysis at 550°C for 2 h.<sup>d</sup>Fat determined by hydrolysis with acid, and extraction of the residue with petroleum ether.<sup>e</sup>Protein determined by Kjeldahl method.

produced economically from an industrial process will always contain minor constituents, namely protein, fat, and ash. Fiber may also be present but at minute levels. According to the starch type, these will additively represent a maximum of 1.5% of the native starch. As examples, potato starch has little fat or protein and is probably 99.5% pure, while maize starch contains up to 0.6% fat and 0.4% protein, so is less than 99% pure. As all methods for determining starch have optimistically a best tolerance of  $\pm 1\%$ , true starch content is better determined by measuring the minor constituents and subtracting these from 100. Measurement of protein is done by standard methods such as the Kjeldahl or Dumas techniques. It is important, however, that the correct method for fat is used, as a simple solvent extraction is ineffective. To obtain the correct results, the starch must first be hydrolyzed with hot hydrochloric acid, and the resulting residue is collected prior to extraction with petroleum ether. Ash content is usually  $\sim 0.1\text{--}0.2\%$  and is determined by direct combustion in a furnace. Typical composition of six different starches is given in Table 2.

## Qualification of Search

There is no universally accepted definition of starch with regard to minimum chain length or relative molecular mass. A common way of defining starch is by stipulating a certain concentration of ethanol in water solution in which the 'starch' fraction is insoluble. The popular concept of the controlling medium for starch qualification is a 40%, v/v, ethanol solution.

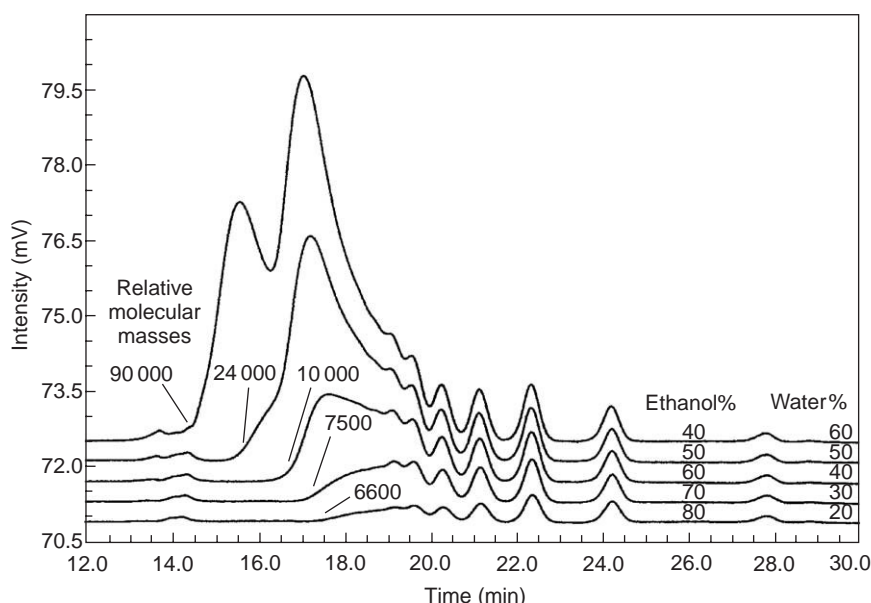
To examine the solubility of starch, a maltodextrin with a low dextrose equivalent of 5 was dispersed in a series of alcohol solutions at different concentrations. After filtration, the solubilized fractions were injected onto size-exclusion columns to determine maximum relative molecular mass  $M_r$ . The results in

Figure 2 show that the recommended concentration of 40% dissolves species of  $M_r$  up to 90 000. The question arises whether the alcohol concentration is too low, as such high  $M_r$  species cannot be classified as oligosaccharides. As expected, the higher alcohol contents show a correspondingly lower limiting solubility. The findings apply only if the original matrix has been fully pregelatinized. With granular starch, only minor amounts of very low  $M_r$  substances are removed. If labeling of foods increases in importance, there may be a need for an international definition of starch.

## Relative Molecular Mass Distribution

Native starches are physically or chemically modified to increase application possibilities, and the relative molecular mass distribution gives an excellent indication of how a starch will perform. The starch might be used as an adhesive, surface size, hydrocolloid, or bulking agent. Size-exclusion chromatography (SEC) is much improved with respect to columns, calibrants, and detectors. First, the starch has to be solubilized without molecular degradation, which eliminates the use of acid and probably alkali. The latter has been proposed but hydrolysis is difficult to avoid. If water is the basic eluent of the chromatography, then a solvent consisting of 90% dimethyl sulfoxide and 10% water is most effective; this ratio is critical. Starch concentration should be as low as possible, and a 0.5% solution gives a concentration adequate for the chromatography. Usually, dissolution is complete after 10 h at 60°C. Cross-bonded starches (e.g., phosphorylated, adipates) cannot be dissolved.

A large number of columns are available for SEC, but this part of the system needs the most careful selection. All packing materials so far evaluated show different degrees of starch adsorption. This can



**Figure 2** Size-exclusion chromatograms of lightly hydrolyzed starch extracted with mixtures of ethanol–water at different ratios. Determination of limiting solubility. Columns, Shodex S803/S801. Solvent, water at 60°C, 1 ml min<sup>-1</sup>. Detection, refractometer. Calibrants, Shodex kit, Pullulan, P802.

be diminished by inclusion of organic solvent, salts, surfactant, or alkali, according to the nature of the packing materials, but doubts may still exist concerning complete elution of the starch. Two packing materials on which only minor adsorption occurs are polystyrene cross-linked with divinylbenzene, sulfonated and in the sodium form, and a copolymer of allyldextran and *N,N*-methylene bis-acrylamide.

The choice of detector depends on what information is required. If SEC is performed as routine evaluation of samples, requiring an approximately correct and repeatable result, then a modern differential refractometer calibrated with commercially available carbohydrate standards will suffice. When more detailed analysis is required, with higher accuracy, no calibration, and information such as degree of branching and radius of gyration, then a light-scattering detector should be used. Instruments are available based on either low-angle laser light scattering (LALLS) or multiple-angle laser light scattering (MALLS). LALLS eliminates the need for complicated mathematics required for higher angles, but is least suited for water-soluble materials. The reverse situation holds for MALLS. With both detectors the amount of attention required will be much greater than with a refractometer. However, further advances are anticipated, and such detectors may become as easy to handle as refractometers. Typical SEC profiles for a corn starch and an acid-thinned corn starch are given in Figure 3.

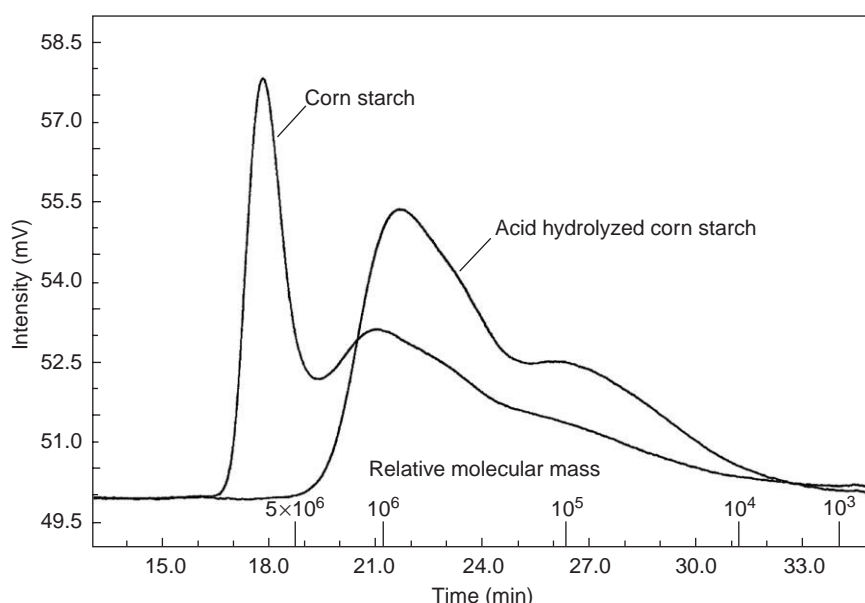
## Amylose/Amylopectin Content

The adsorption of iodine by amylose chains to give a deep blue complex is universally used to qualify starch, and to quantify amylose. For some industrial uses, it is imperative that starch hydrolysates contain no starch, and this can be confirmed by cooling the matrix to 5–10°C, adding 0.02 mol l<sup>-1</sup> iodine dropwise, and observing whether a blue color appears; the detection limit is 50 mg per kg.

Much has been published on ways of measuring iodine affinity, and care must be taken in applying the methodology. The simplest test is measurement of the 'blue value'. This entails dissolving 100 mg of starch in 10 ml of 1 mol l<sup>-1</sup> sodium hydroxide, making up to 100 ml, taking an aliquot, neutralizing, and developing the complex with iodine. Calibration is effected with amylose, and the intensity of the blue complex is measured at 620 nm. In all methods where calibration is effected with an amylose fraction, it should be representative of the amylose being analyzed. Iodine affinity is affected by chain length, degree of branching, and the nature of the amylopectin present. Fat also interferes, and the sample should preferably be defatted by refluxing for 24 h with 85% methanol solution. A more precise method is to measure the iodine affinity by potentiometric titration, using an amylose fraction separated from the native starch as calibrant.

Amylose/amylopectin ratio can be determined by measuring, respectively, the responses of the iodine





**Figure 3** Relative molecular mass distribution of corn starch, and an acid degraded corn starch. Samples dissolved in 90/10 dimethyl sulfoxide–water. Columns, Shodex S805/S804. Eluent,  $0.05 \text{ mol l}^{-1}$  NaOH. Detection, refractometer. Calibration, Shodex Pullulan kit P802.

complex at 620 and 550 nm. Another possibility is applying SEC analysis to a solution of the starch in dimethyl sulfoxide, adding iodine solution postcolumn, and measuring the absorption at the critical wavelengths. Either of these methods will require knowledge of the specific absorption of each fraction to realize quantification. An alternative is to debranch the amylopectin with isoamylase, subject the hydrolysate to SEC analysis, and calculate the carbohydrate profiles from the peak areas of the linear residues.

## Rheology

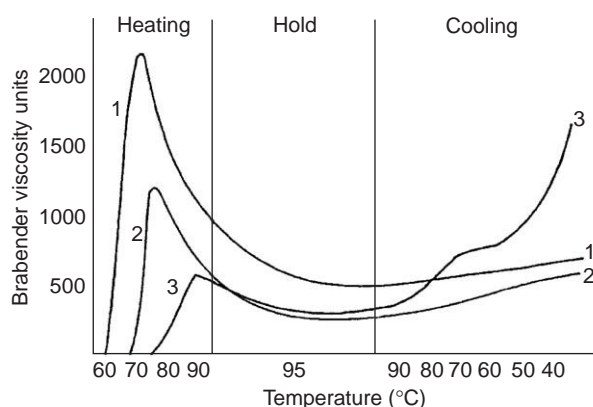
Probably the most important single physical property for application purposes is the viscosity profile of starch during its paste preparation. There are various types of viscometer available for evaluating its behavior but, whatever instrument is used, the conditions of paste preparation must be standardized and adhered to strictly, to obtain repeatability. When a dispersion of starch is heated, the granules absorb water and swell to many times their original volume before bursting. The peak viscosity during cooking is encountered at the maximum swollen state of the hydrated granules. At any moment during stirring and heating, there will be dissolved starch present that has leached out from broken granules. When all the swollen granules have burst, and the starch is dispersed, viscosity decreases as the temperature is raised.

Several types of viscometer are available that follow viscosity during paste preparation. Industry has widely adopted the Brabender viscometer. This has a rotational bowl and the sample exerts a force on fixed sensing elements that are dynamically balanced by a calibrated torsion spring, this causes an angular deflection of the sensing element shaft, which is continuously recorded. After the cooking, the temperature can be held at  $95^{\circ}\text{C}$ , cooled at a predetermined rate, then held at, say,  $50^{\circ}\text{C}$  to evaluate stability of the paste. The instrument is calibrated in arbitrary Brabender Units of viscosity. Typical viscosity curves of some native starches are shown in **Figure 4**.

As a relatively fast product quality and process control, the Scott Cup viscometer is often used. The starch paste is prepared according to a defined protocol and maintained at  $100^{\circ}\text{C}$ , then a fixed volume is timed as it passes through an orifice of defined dimensions. Sample weight is adjusted so that a constant flow time is obtained. Viscosity is then defined as weight Scott in grams. Of the universally used instruments, the Brookfield viscometer is extensively employed. Again, this technique is only of value if sample preparation is standardized.

## Modified Starches

Starch is modified to enhance its properties so as to extend application possibilities. The usual modification is to change the viscosity profile, although an ionic charge can be introduced, or emulsifying



**Figure 4** Brabender viscosity profiles of native starches (8%, w/v, dispersions): 1, potato; 2, waxy maize; 3, regular maize.

properties improved. There are strict regulations on the type that may be used in food, and it is imperative to have reliable methodology to determine type and degree of chemical substitution. The simplest modification is hydrolytic degradation with acid, which decreases the peak viscosity but increases the paste viscosity on storage. For such a product a viscosity profile is the important evaluation. Starch treated with hypochlorite under alkaline conditions exhibits a lower peak viscosity and has a clear, stable, low-viscosity paste. Carboxyl groups are introduced, and these can be quantified by titration with alkali, after conversion into the hydrogen form. Its anionic nature can be confirmed by staining the starch granules with methylene blue.

Of the esters, starch phosphate is produced by reaction with phosphorus oxychloride, polyphosphates, or metaphosphates; a cross-bonded product results. Total degree of substitution is determined by measuring the phosphorus content, and the mono- to disubstitution ratio can be calculated by potentiometric titration. Allowance is made for the natural phosphorus content of the starch. Treatment of starch with acetic anhydride produces starch acetate, which has improved paste stability over native starch. The acetyl group is very labile, and hydrolyses readily under mild alkaline conditions. When a known amount of alkali is used, the excess can be titrated and the ester function measured. This is not specific, however, and a method based on an enzymatic measurement of the acetate has been developed in an ISO work group. The modified starch is hydrolyzed under acidic conditions, which releases acetic acid and permits filtration of the resulting solution. Acetic acid is then measured by a commercially available enzyme test kit. Both bound and free acetyl groups can be measured, and the method is applicable

to granular and pregelatinized starches. The ISO collaborative exercise gave a reproducibility of 8% on 2% acetyl content.

Starch adipate, a cross-bonded starch for food use, is made by reaction with adipic anhydride, which is formed from adipic acid in the presence of excess acetic anhydride. It is also a labile ester, and after hydrolysis with alkali, followed by acidification, is extracted with ethyl acetate and silylated. Gas chromatographic analysis is performed on a capillary column of fused silica coated with dimethyl siloxane, film thickness 5  $\mu\text{m}$ . Pimelic acid is the internal standard. The method cannot differentiate between mono- and disubstitution. Alkyl succinate substitution can be determined using the same procedure as for adipate.

Starch ethers present a more difficult analytical problem because the bond is very resistant to attack. Carboxy-methyl starch substitution is determined by the titration method as used for carboxyl function. However, many of the commercial products are water soluble, and must first be precipitated with ethyl alcohol prior to analysis. Hydroxyalkyl starches can be analyzed using the Zeisel procedure based on decomposition with boiling hydroiodic acid. The alkyl iodide and corresponding olefin are formed, which can be measured by trapping in silver nitrate and bromine water, respectively. An alternative approach is to use proton nuclear magnetic resonance (NMR) spectroscopy, as the spectra from the methyl protons of the alkyl substitution permits quantification of the alkyl moiety. Cationic starches are commonly made from reaction with glycidyltrimethylammonium chloride, and substitution is usually determined from the nitrogen content; cationic nature is confirmed by staining with yellow-green SF dye. Proton NMR can also be used for this substituent.

**See also:** Chiroptical Analysis. **Enzymes:** Overview. **Food and Nutritional Analysis:** Oils and Fats. **Laser-Based Techniques.** **Liquid Chromatography:** Principles; Size-Exclusion; Food Applications. **Proteins:** Foods. **Quality Assurance:** Internal Standards.

## Further Reading

- Bergmeyer HU (ed.) (1983) *Methods of Enzymatic Analysis*, 3rd edn. Weinheim: Verlag Chemie.
- Mitchell GA (1990) Methods of starch analysis. *Starch/Stärke* 4: 131–134.
- Radley JA (1953) In: Tripp EH (ed.) *Starch and Its Derivatives*, 3rd edn. London: Chapman and Hall.
- Whistler RL (1964) *Methods in Carbohydrate Chemistry*. London: Academic Press.
- Whistler RL, BeMiller JN, and Paschall EF (1984) *Starch: Chemistry and Technology*, 2nd edn. London: Academic Press.

## Dietary Fiber Measured as Nonstarch Polysaccharides in Plant Foods

**H N Englyst, G J Hudson, and M E Quigley**, Englyst Carbohydrates Research and Services Ltd., Southampton, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The original hypothesis on 'dietary fiber' was that a diet rich in unrefined plant foods was linked to a low incidence of 'Western diseases' (including diabetes). The hypothesis has been largely confirmed, and this is the background for the recommendations in national dietary guidelines for the consumption of a diet rich in fruit, vegetables, and whole-grain cereals. Unrefined plant foods are naturally rich in plant cell-wall material; the major constituents (~90%) of this material are non-starch polysaccharides (NSP). The naturally occurring plant cell-wall NSP content of foods is therefore a good marker for the unrefined plant foods embodied in the dietary fiber hypothesis and recommended in current dietary guidelines. In line with this, the Joint WHO/FAO Expert Consultation on Diet, Nutrition and the Prevention of Chronic Diseases (2000) specifies the recommended increased intake of fiber in the prevention of obesity, diabetes, and coronary heart diseases in terms of NSP.

The encapsulation of starch and sugars within the walls of the intact plant cell restricts the rate at which they are digested and absorbed in the small intestine, resulting in low glycemic index (GI) values and low rapidly available glucose (RAG) values (see Table 1). This ability to influence the rate and extent of digestion and absorption, and thus the physiological effects of other nutrients, is an important property of the cell wall in unrefined plant foods. This encapsulation effect is largely destroyed by excessive food processing, and cannot be reintroduced by the addition of 'fiber supplements'. It should be noted that some fiber supplements, when taken in large amounts, do lower the glycemic response via their effect on gastrointestinal function. This is especially true for the soluble materials, such as gums, which, due to their high viscosity, slow gastric emptying and restrict access of digestive enzymes to the food bolus. However, their effect is not related to their origin as plant material or even to their chemical identity (i.e., NSP); it is due solely to their physical properties (solubility and viscosity). Accordingly, the use of fiber supplements to augment the fiber-poor diets

typical of modern Western societies is not promoted by dietary recommendations, as these supplements are not part of the original hypothesis and do not impart the same overall beneficial properties as a diet rich in naturally occurring plant cell-wall material.

NSP, by definition, includes all the plant polysaccharides other than starch. For the reasons given above, NSP are divided into two broad classes in the classification and measurement scheme described here (Table 1): (1) the cell-wall NSP, which impart rigidity, and encapsulate and control the release of other nutrients, and (2) other NSP, including gums and refined preparations of cell-wall material, which occur in foods mainly as additives. There is no endogenous human enzyme for the hydrolysis of NSP, which are therefore nonglycemic, and all become available for fermentation in the large intestine.

The Englyst procedure for the measurement of NSP as a marker for 'dietary fiber' has evolved from the principles laid down by McCance and Widdowson, and later by Southgate. The procedure involves: (1) removal of starch by enzymatic hydrolysis, (2) collection of NSP by precipitation in ethanol, (3) release of the constituent sugars of NSP by acid hydrolysis, and (4) quantification of the NSP by measurement of the constituent sugars by gas chromatography (GC), high-performance liquid chromatography (HPLC), or colorimetry. Values for total, soluble, and insoluble NSP may be obtained.

### Applications of the Analytical Values

The detailed information obtained from the chromatographic methods, which identify and measure the individual constituent sugars, is particularly useful in studies of the relation between intakes of NSP and health. The results of NSP analysis, with values for the individual constituent sugars, are shown in Table 2 for a range of plant foods.

The spectrum of the constituent sugars is characteristic for various types of plant NSP and may indicate the origin of the NSP measured. The values for wholemeal wheat products are characterized by high levels of insoluble NSP in the form of cellulose (measured as insoluble NSP glucose) and arabinoxylans. Wholemeal wheat NSP are fermented slowly and incompletely, and exert a considerable effect on fecal bulk. White bread contains only 30% as much NSP as wholemeal bread, and this NSP is

**Table 1** Classification of dietary carbohydrates

	Main components	Comments
<i>Free sugars</i>		(Soluble in 80% ethanol; $\leq 2$ sugar units)
Mono- and disaccharides	Glucose, fructose, sucrose maltose, lactose	Glucose, maltose, and sucrose digested rapidly Fructose and lactose may, in part, escape digestion and absorption in the small intestine Physiological response depends on identity Free glucose + glucose from sucrose = free sugar glucose (FSG)
<i>Sugar alcohols</i>		(Soluble in 80% ethanol; $\leq 2$ sugar units)
Mono- and disaccharides	Sorbitol, inositol, mannitol galactitol, maltitol	Poorly absorbed in the small intestine. May reach the large intestine
<i>Short-chain carbohydrates</i>		(Soluble in 80% ethanol; $> 2$ sugar units)
Maltodextrins	$\alpha$ -Glucans	Partly hydrolyzed starch. Normally included in the measurement of starch
Resistant short-chain carbohydrates (nondigestible oligosaccharides)	Fructo-oligosaccharides, galacto-oligosaccharides, pyrodextrins, polydextrose	Escape digestion in the small intestine and are fermented to different extents. Some may stimulate growth of bifidobacteria. Physiological effect largely unknown
<i>Polysaccharides</i>		(Insoluble in 80% ethanol)
Starch	$\alpha$ -Glucans	The most abundant dietary carbohydrates
Rapidly digestible starch (RDS)	Rapidly released glucose	Rapidly digested in the small intestine RDS + rapidly released FSG = RAG
Slowly digestible starch (SDS)	Slowly released glucose	Slowly digested in the small intestine SDS + slowly released FSG = SAG
Resistant starch (RS)	RS <sub>1</sub> (physically inaccessible); RS <sub>2</sub> (resistant granules); RS <sub>3</sub> (retrograded starch)	The three types of RS escape digestion in the small intestine and are fermented to different extents. Physiological effect largely unknown
Non-starch polysaccharides (NSP)	Many different types of polysaccharides	Escape digestion in the small intestine and are fermented to different extents
Plant cell-wall NSP	Main constituents: arabinose, xylose, mannose, galactose, glucose, uronic acids	Encapsulate and slow absorption of nutrients. Good marker for naturally fiber-rich diets for which health benefits have been shown
Other NSP	Many types of constituents	Food additives. Minor components of the human diet. The amounts added to foods are known and regulated

FSG, free-sugar glucose; RAG, rapidly available glucose; SAG, slowly available glucose.

largely soluble and expected to have only a moderate effect on fecal bulk. Oats and rye contain a greater proportion of soluble NSP compared with wheat products, and the main fraction is a  $\beta$ -glucan, which is measured as soluble NSP glucose. This is associated with a greater effect on cholesterol metabolism, which is in agreement with the claims for oat products. Corn flakes provide an example of highly fiber-depleted breakfast cereal. Fruit and vegetables have high levels of soluble fiber, and the main fraction in these foods is pectin, which is measured as soluble NSP uronic acids. In general, cereal products contain more xylose than arabinose, while fruits and vegetables contain less xylose than arabinose, which is measured

mainly in the soluble fraction. The presence of the minor constituents, rhamnose and fucose, and the high values for uronic acids indicate a diet rich in fruits and vegetables. **Figure 1** illustrates the differences in NSP constituent sugars seen for cereals, fruit, and vegetables. When detailed information on the constituent sugars is not required, values for total, soluble, and insoluble dietary fiber may be obtained by the more rapid colorimetric endpoint procedure.

The detailed NSP analysis, which yields scientifically defensible values for physically and chemically identified constituents, is necessary for two major lines of approach to understanding the links between

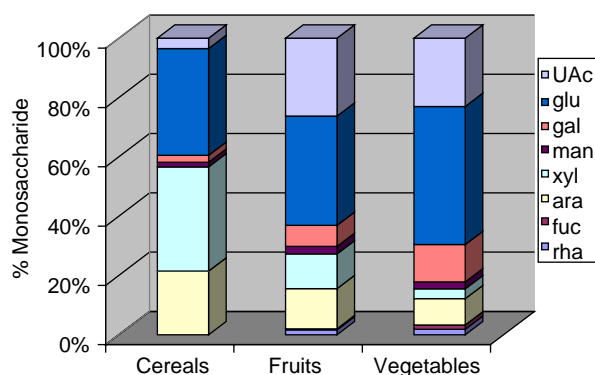
**Table 2** NSP in (g per 100 g dry matter) for a range of plant foods

<i>Plant foods</i>		<i>NSP constituents</i>								
		<i>Total</i>	<i>Rha</i>	<i>Fuc</i>	<i>Ara</i>	<i>Xyl</i>	<i>Man</i>	<i>Gal</i>	<i>Glc</i>	<i>UAc</i>
<i>Cereals</i>										
Bread, wholemeal	Soluble	2.3	t	t	0.7	0.8	0.1	0.2	0.4	0.1
	Insoluble	6.9	t	t	1.8	2.7	0.1	0.1	2.0	0.2
	Total	9.2	t	t	2.5	3.5	0.2	0.3	2.4	0.3
Bread, rye	Soluble	6.7	t	t	1.7	3.2	t	0.2	1.6	t
	Insoluble	6.6	t	t	1.8	2.6	0.1	0.1	1.9	0.1
	Total	13.3	t	t	3.5	5.8	0.1	0.3	3.5	0.1
Bread, white	Soluble	1.6	t	t	0.5	0.8	t	0.1	0.2	t
	Insoluble	1.1	t	t	0.3	0.4	0.1	t	0.3	t
	Total	2.7	t	t	0.8	1.2	0.1	0.1	0.5	t
Corn flakes	Soluble	0.4	t	t	t	0.2	t	t	0.1	0.1
	Insoluble	0.5	t	t	0.1	0.1	t	t	0.3	t
	Total	0.9	t	t	0.1	0.3	t	t	0.4	0.1
Quaker oats	Soluble	5.0	t	t	0.3	0.3	t	0.1	4.3	t
	Insoluble	3.5	t	t	0.8	1.1	0.1	0.1	1.2	0.2
	Total	8.5	t	t	1.1	1.4	0.1	0.2	5.5	0.2
<i>Fruits</i>										
Apple	Soluble	5.8	0.2	0.1	1.2	0.1	t	0.3	0.1	3.8
	Insoluble	7.5	0.1	0.1	0.9	0.7	0.3	0.6	4.5	0.3
	Total	13.3	0.3	0.2	2.1	0.8	0.3	0.9	4.6	4.1
Orange	Soluble	9.8	0.3	t	1.9	0.1	0.1	1.4	0.1	5.9
	Insoluble	5.2	t	t	0.3	0.5	0.3	0.4	3.4	0.3
	Total	15.0	0.3	t	2.2	0.6	0.4	1.8	3.5	6.2
Peach	Soluble	7.1	0.2	t	1.9	t	0.2	0.9	t	3.9
	Insoluble	6.4	t	t	0.6	0.8	0.2	0.4	4.2	0.2
	Total	13.5	0.2	t	2.5	0.8	0.4	1.3	4.2	4.1
Pineapple	Soluble	0.8	t	t	0.1	t	0.3	0.1	t	0.3
	Insoluble	8.3	0.1	t	1.1	2.1	t	0.6	4.0	0.4
	Total	9.1	0.1	t	1.2	2.1	0.3	0.7	4.0	0.7
Strawberry	Soluble	5.1	0.2	t	0.6	t	t	0.3	t	4.0
	Insoluble	6.8	t	t	0.2	1.4	0.2	0.2	4.5	0.3
	Total	11.9	0.2	t	0.8	1.4	0.2	0.5	4.5	4.3
<i>Vegetables</i>										
Cabbage	Soluble	16.6	0.7	t	4.4	0.2	0.3	2.3	0.2	8.5
	Insoluble	20.8	t	0.1	1.3	1.8	0.8	1.3	14.7	0.8
	Total	37.4	0.7	0.1	5.7	2	1.1	3.6	14.9	9.3
Carrot	Soluble	14.9	0.8	t	2.4	t	t	4	t	7.7
	Insoluble	11.1	t	t	0.4	0.4	0.5	0.6	8.9	0.3
	Total	26	0.8	t	2.8	0.4	0.5	4.6	8.9	8.0
Pea	Soluble	5.9	0.2	t	1.9	0.3	0.1	0.6	t	2.8
	Insoluble	15	0.1	t	1	0.5	t	0.2	12.6	0.6
	Total	20.9	0.3	t	2.9	0.8	0.1	0.8	12.6	3.4
Potato	Soluble	3.5	0.1	t	0.4	t	t	1.5	0.4	1.1
	Insoluble	3.2	t	t	0.1	0.1	t	0.2	2.7	0.1
	Total	6.7	0.1	0.5	0.1	0.1	t	1.7	3.1	1.2
Tomato	Soluble	7.4	0.2	t	0.5	0.1	t	1.0	0.2	5.4
	Insoluble	11.4	0.1	t	0.4	0.9	1.3	0.7	11.6	0.3
	Total	18.8	0.3	t	0.9	1.0	1.3	1.7	11.8	5.7

Rha, rhamnose; Fuc, fucose; Ara, arabinose; Xyl, xylose; Man, mannose; Gal, galactose; Glc, glucose; UAc, uronic acids.

dietary carbohydrates and health: (1) the correlation between dietary carbohydrate and disease as revealed from the analysis of epidemiological studies, and (2) the elucidation of the mechanisms that underlie the physiological importance of dietary carbohydrates for health. When used in food databanks and for food labeling, these data represent a reliable basis for

selecting the type of diet beneficial to health. The Englyst procedure for the measurement of NSP as a marker for dietary fiber in plant foods has been tested thoroughly in large international collaborative trials. Values for dietary fiber measured as NSP by this technique are used in the McCance and Widdowson UK food tables.



**Figure 1** NSP in different food types expressed as percentage monosaccharide present.

The purpose of labeling food with composition data is to provide the consumer with information on food quality. Which individual components appear on the label (e.g., energy, protein, fat, carbohydrate, sugar, starch, NSP, fiber) and how the values are obtained (by analysis by proscribed methods or by calculation) is decided by the legislators in each country or group of countries. There is no worldwide consensus on what material should be included in the term 'fiber' (or dietary fiber) and, hence, no worldwide agreement on a method for the measurement of fiber. Nutrition studies concerned with the qualitative and quantitative aspects of foodstuffs rely on specific definitions of food components and the availability of reliable methods of measurement. Identification of the NSP and the measurement procedure described here, with values for NSP constituent sugars, satisfy both requirements. The use of undefined terms, such as dietary fiber, and the use of nonspecific methodology satisfy neither of the criteria, and leads to the collection of uninterpretable and/or misleading data.

The argument is largely beyond the scope of this article but the reader is guided to relevant sources of information in the Further Reading and especially to the two WHO/FAO documents. If the information on the food label is to be meaningful for the consumer, it is important that the compositional values are accurate and, in a world of international trade, transcend geographical and language barriers. The Joint FAO/WHO Expert Consultations on Carbohydrates in Human Nutrition in 1998, concerning the analysis of dietary carbohydrates stated:

That the analysis and labelling of dietary carbohydrates, for whatever purpose, be based on the chemical divisions recommended. Additional groupings such as polyols, resistant starch, non-digestible oligosaccharides and dietary fibre can be used, provided the included components are clearly defined.

The measurement of plant cell-wall NSP as described here is in line with these recommendations, and provides values that can appear on the food label in many countries as fiber or dietary fiber. Any of the endpoint measures that can be used in the Englyst NSP procedure, GC, HPLC, or colorimetry, is suitable for food labeling and for quality control.

Studies designed (1) to examine the mechanisms underlying the links between diet and disease, and (2) to be the basis for preventive measures for the improvement and maintenance of public health, have an absolute requirement for specific definitions of food components and the availability of reliable methods of measurement. The approach used in the Englyst procedure for the measurement of the constituent sugars of the plant cell-wall NSP provides detailed analytical data that allow for sophisticated hypotheses to be erected and tested.

## Methods and Principles of Measurement

The Englyst procedure uses enzymatic and chemical methods to measure NSP. (For detailed working protocols, see Further Reading.) All starch is hydrolyzed enzymatically and NSP are measured as the sum of the constituent sugars released by acid hydrolysis. The sugars may be measured by GC or by HPLC to obtain values for individual monosaccharides, or a single value for total sugars may be obtained by colorimetry. Values may be obtained for total, soluble, and insoluble NSP, and a small modification allows cellulose to be measured separately.

The procedure as described here provides the following options:

1. GC procedure: measures NSP as the sum of neutral sugars obtained by GC and uronic acids measured separately.
2. HPLC procedure: measures NSP as the sum of neutral sugars and uronic acids.
3. Colorimetric procedure: measures NSP as reducing sugars.

The main procedural steps are

1. Dry/defat sample if necessary.
2. Disperse and hydrolyze starch enzymatically.
3. Precipitate NSP in acidified aqueous ethanol.
4. Disperse and hydrolyze NSP with sulfuric acid.
5. Measure released constituent sugars by colorimetry, GC, or HPLC.

Sample preparation includes freeze-drying and milling so that representative samples may be taken.

High-fat (>10%) samples are defatted with acetone. In order to disperse the starch completely, the sample is treated with dimethylsulfoxide, then buffered and the starch hydrolyzed with a mixture of pancreatic  $\alpha$ -amylase and pullulanase. After precipitation of the NSP with acidified ethanol, the supernatant is removed and discarded, or it may be used for the measurement of resistant short-chain carbohydrates (see **Table 1** and Further Reading). The precipitated NSP are washed with 80% ethanol to remove any free sugars and then dried with acetone. The dried, starch-free residue is treated with 72% sulfuric acid to disperse cellulose, followed by 2 mol l<sup>-1</sup> sulfuric acid for 1 h at 100°C to hydrolyze all the cell-wall polysaccharides to their constituent monosaccharides. When detailed information is required, the NSP sugars may be measured by GC or by HPLC, giving values for individual monosaccharides, or a single value may be obtained by colorimetry. In the GC procedure, the sugars are reduced to their alditols with alkaline sodium borohydride and acetylated with acetic anhydride in the presence of methylimidazole as catalyst. Conventional GC is used to measure the resulting alditol acetate derivatives of the neutral sugars. The uronic acid-containing polysaccharides are more difficult to hydrolyze and require treatment with concentrated acid at high temperature; for this reason, they are measured separately by colorimetry. In the HPLC assay, the hydrolysate is diluted, an internal standard is added, and the neutral sugars and uronic acids are measured directly by electrochemical detection.

Two kits are available from Englyst Carbohydrate to help ensure that accurate analytical results are obtained. The kit for the colorimetric procedure contains color reagent and a second kit contains a solution of allose as internal standard for the GC procedure. Both kits contain the required enzymes, sugar solutions, and reference materials and are rigorously tested in-house.

The four reference materials included in the kits may be used as part of a complete quality control procedure:

- Reference sample 1 (white flour). This sample is used to check the efficiency of the starch hydrolysis and washing steps when there is a small amount of NSP in the presence of a large amount of starch.
- Reference sample 2 (haricot bean). This sample contains all the constituent sugars of NSP and is used to check the efficiency of the starch hydrolysis and washing steps for samples that may cause handling problems, such as aggregation.

- Reference sample 3 (high-amylose starch resistant to  $\alpha$ -amylase). This sample is taken through the entire procedure for total NSP and is used to check the efficiency of the starch hydrolysis and washing steps.
- Reference sample 4 (cellulose). This sample is subjected to direct acid hydrolysis only and is used to check the acid hydrolysis steps.

## Method Validation

### Comparison of the Colorimetric, GC, and HPLC Assays

Three alternative endpoint techniques may be used to obtain values for NSP. The chromatographic procedures yield detailed information about the individual NSP constituent sugars. We have demonstrated that these chromatographic procedures give identical values for both neutral sugars and uronic acids. The colorimetric procedure has been shown in interlaboratory studies to give values for total, soluble, and insoluble NSP that are identical with the values obtained by the chromatographic procedures. The colorimetric assay is ideally suited for food labeling purposes, where a single value is required.

### International Collaborative Trials

The Englyst procedure has been the subject of a series of international collaborative trials organized by the UK Ministry of Agriculture, Fisheries and Food (MAFF). In the MAFF IV study, 37 laboratories from 11 countries compared the accuracy and precision of the Englyst GC and colorimetric procedures. This has culminated in the publication of the GC and the colorimetry techniques as MAFF Approved Methods.

## Certification of Reference Materials

As the result of a large international trial of methodology, following rigorous study of stability of the test materials, five EU Community Bureau of Reference certificated reference materials (CRMs) are available for use with the Englyst GC and colorimetry NSP procedures:

1. dried haricot bean powder, CRM 514;
2. dried carrot powder, CRM 515;
3. dried apple powder, CRM 516;
4. full-fat soya flour, CRM 517; and
5. dried powdered bran breakfast cereal, CRM 518.



These CRMs can be used to check the performance of the analytical method and as quality control of analytical measurements for nutritional labeling.

**See also:** **Carbohydrates:** Overview; Sugars – Spectrophotometric Methods; Sugars – Chromatographic Methods. **Liquid Chromatography:** Food Applications.

## Further Reading

- Burkitt DP and Trowell H (1975) *Refined Carbohydrate Foods and Disease. Some Implications of Dietary Fibre*. New York: Academic Press.
- Englyst HN, Wiggins HS, and Cummings JH (1982) Determination of the non-starch polysaccharides in plant foods by gas-liquid chromatography of constituent sugars as alditol acetates. *Analyst* 107: 307–318.
- Englyst HN, Trowell H, Southgate DAT, and Cummings JH (1987) Dietary fiber and resistant starch. *American Journal of Clinical Nutrition* 46: 873–874.
- Englyst HN, Quigley ME, and Hudson GJ (1994) Determination of dietary fibre as non-starch polysaccharides with gas-liquid chromatographic, high-performance liquid chromatographic or spectrophotometric measurement of constituent sugars. *Analyst* 119: 1497–1509.
- Englyst HN, Quigley ME, Englyst KN, Bravo L, and Hudson GJ (1996) Dietary fibre. Measurement by the Englyst NSP procedure. Measurement by the AOAC procedure. Explanation of the differences. *Journal of the Association of Public Analysts* 32: 1–52.
- Englyst KN, Englyst HN, Hudson GJ, Cole TJ, and Cummings JH (1999) Determination of rapidly available glucose in foods. A measurement that reflects the glycemic response. *American Journal of Clinical Nutrition* 69: 448–454.
- FAO/WHO (1998) *Carbohydrates in Human Nutrition*. Report of a Joint FAO/WHO Expert Consultation, Rome 14–18 Apr. 1997, FAO Food and Nutrition paper 66, pp. 1–140. Rome: FAO.
- Holland B, Welch AA, Unwin ID, *et al.* (1991) *Mccance & Widdowson's The Composition of Foods*, 5th edn. Cambridge and London: Ministry of Agriculture Fisheries & Food and Royal Society of Chemistry.
- Quigley ME and Englyst HN (1992) Determination of neutral sugars and hexosamines by high-performance liquid chromatography with pulsed amperometric detection. *Analyst* 117: 1715–1718.
- Trowell H (1985) Dietary fibre: A paradigm. In: Trowell HC, Burkitt D, and Heaton KW (eds.) *Dietary Fibre, Fibre-Depleted Foods and Disease*, pp. 1–20. London: Academic Press.
- WHO/FAO (2002) *Joint WHO/FAO Expert Consultation on Diet, Nutrition and the Prevention of Chronic Diseases*. Geneva, Switzerland, 28 Jan–1 Feb, 2002; <http://www.who.int/hpr/nutrition/26Aprildraftrev1.pdf>
- Wood R, Englyst HN, Southgate DAT, and Cummings JH (1993) Determination of dietary fibre in foods – collaborative trials IV. Comparison of Englyst GC and colorimetric measurement with the Prosky procedure. *Journal of the Association of Public Analysts* 29: 57–141.

# CARBON

**M Nimmo**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Carbon, as inorganic or organic species, is ubiquitous in all environmental regimes. Global carbon reservoirs of significance include land biota, soils, surface marine sediments, and the oceans. Of these, quantitatively the most important carbon reservoir at the earth's surface is the dissolved inorganic carbon (DIC) pool in the ocean, estimated to hold  $\sim 38\,000\text{ Gt C}$  ( $\text{Gt} = 10^{15}\text{ g}$ ). In contrast, the atmospheric pool of C as  $\text{CO}_2$  is estimated to hold  $\sim 750\text{ Gt}$ , compared to that estimated  $\sim 600\text{ Gt C}$  in preindustrial times. The changes in the chemical form (inorganic – see below – and organic C both in the dissolved and particulate phases) and environmental

reservoir quantities and fluxes between the reservoirs of C are linked to physical (atmospheric transport, diffusion, and solubility), chemical (acid/base equilibria), biological (respiration and photosynthesis), as well as anthropic processes (fossil fuel burning, forest destruction). Currently,  $\sim 4\%$  of the  $\text{CO}_2$  emitted into the atmosphere is of anthropogenic origin, this historic increase in  $\text{CO}_2$  inputs has led to the well-documented increase of  $\text{CO}_2$  atmospheric levels and the potential impacts on the global climate. The ocean is considered a significant sink for anthropogenically emitted atmospheric  $\text{CO}_2$ , taking up  $\sim 2.0 \pm 0.6\text{ Gt-C year}^{-1}$ ; the dynamics of this are mainly determined by the rate of transport of surface water that is rich in  $\text{CO}_2$  to depth. Therefore, to better quantitatively define the global C cycle and reservoirs, accurate and precise monitoring of the C concentrations in atmospheric, terrestrial, and marine samples is imperative.

These CRMs can be used to check the performance of the analytical method and as quality control of analytical measurements for nutritional labeling.

**See also:** **Carbohydrates:** Overview; Sugars – Spectrophotometric Methods; Sugars – Chromatographic Methods. **Liquid Chromatography:** Food Applications.

## Further Reading

- Burkitt DP and Trowell H (1975) *Refined Carbohydrate Foods and Disease. Some Implications of Dietary Fibre*. New York: Academic Press.
- Englyst HN, Wiggins HS, and Cummings JH (1982) Determination of the non-starch polysaccharides in plant foods by gas-liquid chromatography of constituent sugars as alditol acetates. *Analyst* 107: 307–318.
- Englyst HN, Trowell H, Southgate DAT, and Cummings JH (1987) Dietary fiber and resistant starch. *American Journal of Clinical Nutrition* 46: 873–874.
- Englyst HN, Quigley ME, and Hudson GJ (1994) Determination of dietary fibre as non-starch polysaccharides with gas-liquid chromatographic, high-performance liquid chromatographic or spectrophotometric measurement of constituent sugars. *Analyst* 119: 1497–1509.
- Englyst HN, Quigley ME, Englyst KN, Bravo L, and Hudson GJ (1996) Dietary fibre. Measurement by the Englyst NSP procedure. Measurement by the AOAC procedure. Explanation of the differences. *Journal of the Association of Public Analysts* 32: 1–52.
- Englyst KN, Englyst HN, Hudson GJ, Cole TJ, and Cummings JH (1999) Determination of rapidly available glucose in foods. A measurement that reflects the glycemic response. *American Journal of Clinical Nutrition* 69: 448–454.
- FAO/WHO (1998) *Carbohydrates in Human Nutrition*. Report of a Joint FAO/WHO Expert Consultation, Rome 14–18 Apr. 1997, FAO Food and Nutrition paper 66, pp. 1–140. Rome: FAO.
- Holland B, Welch AA, Unwin ID, *et al.* (1991) *Mccance & Widdowson's The Composition of Foods*, 5th edn. Cambridge and London: Ministry of Agriculture Fisheries & Food and Royal Society of Chemistry.
- Quigley ME and Englyst HN (1992) Determination of neutral sugars and hexosamines by high-performance liquid chromatography with pulsed amperometric detection. *Analyst* 117: 1715–1718.
- Trowell H (1985) Dietary fibre: A paradigm. In: Trowell HC, Burkitt D, and Heaton KW (eds.) *Dietary Fibre, Fibre-Depleted Foods and Disease*, pp. 1–20. London: Academic Press.
- WHO/FAO (2002) *Joint WHO/FAO Expert Consultation on Diet, Nutrition and the Prevention of Chronic Diseases*. Geneva, Switzerland, 28 Jan–1 Feb, 2002; <http://www.who.int/hpr/nutrition/26Aprildraftrev1.pdf>
- Wood R, Englyst HN, Southgate DAT, and Cummings JH (1993) Determination of dietary fibre in foods – collaborative trials IV. Comparison of Englyst GC and colorimetric measurement with the Prosky procedure. *Journal of the Association of Public Analysts* 29: 57–141.

# CARBON

**M Nimmo**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

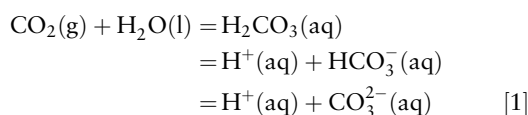
Carbon, as inorganic or organic species, is ubiquitous in all environmental regimes. Global carbon reservoirs of significance include land biota, soils, surface marine sediments, and the oceans. Of these, quantitatively the most important carbon reservoir at the earth's surface is the dissolved inorganic carbon (DIC) pool in the ocean, estimated to hold  $\sim 38\,000\text{ Gt C}$  ( $\text{Gt} = 10^{15}\text{ g}$ ). In contrast, the atmospheric pool of C as  $\text{CO}_2$  is estimated to hold  $\sim 750\text{ Gt}$ , compared to that estimated  $\sim 600\text{ Gt C}$  in preindustrial times. The changes in the chemical form (inorganic – see below – and organic C both in the dissolved and particulate phases) and environmental

reservoir quantities and fluxes between the reservoirs of C are linked to physical (atmospheric transport, diffusion, and solubility), chemical (acid/base equilibria), biological (respiration and photosynthesis), as well as anthropic processes (fossil fuel burning, forest destruction). Currently,  $\sim 4\%$  of the  $\text{CO}_2$  emitted into the atmosphere is of anthropogenic origin, this historic increase in  $\text{CO}_2$  inputs has led to the well-documented increase of  $\text{CO}_2$  atmospheric levels and the potential impacts on the global climate. The ocean is considered a significant sink for anthropogenically emitted atmospheric  $\text{CO}_2$ , taking up  $\sim 2.0 \pm 0.6\text{ Gt-C year}^{-1}$ ; the dynamics of this are mainly determined by the rate of transport of surface water that is rich in  $\text{CO}_2$  to depth. Therefore, to better quantitatively define the global C cycle and reservoirs, accurate and precise monitoring of the C concentrations in atmospheric, terrestrial, and marine samples is imperative.

Additionally, water, soil, urban and enclosed air quality may be defined by their inorganic C content (e.g., alkalinity, CO in air, carbonate content in soils), also requiring accurate/precise analytical techniques. This article will therefore critically discuss both direct and indirect analytical methods currently available for inorganic C analysis.

## Aquatic Systems

Inorganic C may exist in natural water systems in a number of different chemical forms. These are carbonate ( $\text{CO}_3^{2-}$ ), hydrogencarbonate ( $\text{HCO}_3^-$ ), and dissolved  $\text{CO}_2$  (free and hydrated form). In natural waters, the inorganic forms of carbon are derived from the chemical weathering of carbonate rocks, the decomposition of biotic material (respiration), and solvation of atmospheric  $\text{CO}_2$ . The dissolved inorganic C species distribution is a function of pH and is described by equations given below (i.e., the carbonate system):



The total inorganic carbon (TIC) in aqueous solution is the sum of the concentrations of all the species, such that:

$$\text{TIC} = C_T = [\text{CO}_3^{2-}] + [\text{HCO}_3^-] + [\text{H}_2\text{CO}_3] + [\text{CO}_2] \quad [2]$$

The relative concentrations (to TIC) of the different inorganic C species (defined as  $\alpha_0$ ,  $\alpha_1$ , and  $\alpha_2$ ) will be dependent upon the solution pH and acid dissociation constants, such that

$$\begin{aligned}[\text{H}_2\text{CO}_3^*]/C_T &= [\text{H}_3\text{O}^+]^2 / [\text{H}_3\text{O}^+]^2 + K_1[\text{H}_3\text{O}^+] + K_1K_2 \\ &= \alpha_0\end{aligned}\quad [3]$$

$$\begin{aligned}[\text{HCO}_3^-]/C_T &= K_1[\text{H}_3\text{O}^+] / [\text{H}_3\text{O}^+]^2 + K_1[\text{H}_3\text{O}^+] + K_1K_2 \\ &= \alpha_1\end{aligned}\quad [4]$$

$$\begin{aligned}[\text{CO}_3^{2-}]/C_T &= K_1K_2 / [\text{H}_3\text{O}^+]^2 + K_1[\text{H}_3\text{O}^+] + K_1K_2 \\ &= \alpha_2\end{aligned}\quad [5]$$

where

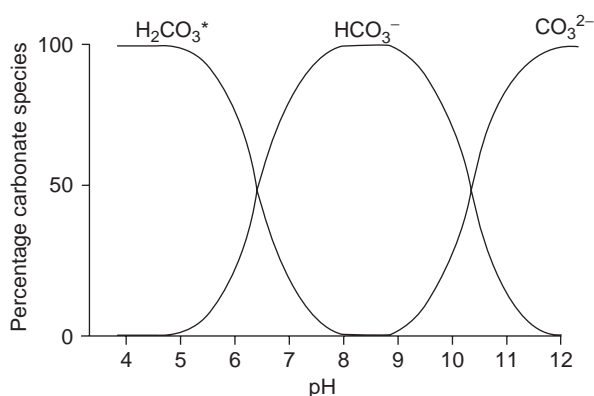
$$K_1 = [\text{H}_3\text{O}^+][\text{HCO}_3^-] / [\text{H}_2\text{CO}_3] \quad [6]$$

and

$$K_2 = [\text{H}_3\text{O}^+][\text{CO}_3^{2-}] / [\text{HCO}_3^-] \quad [7]$$

$$pK_1 = 6.3; \quad pK_2 = 10.3$$

The distribution of the various C species against pH is presented in **Figure 1**. It is clear from **Figure 1** that



**Figure 1** The distribution of dissolved inorganic carbonate species in aqueous solution as a function of pH.

in the pH range of the majority of natural waters systems (6.5–9), the  $\text{HCO}_3^-$  chemical form will be the most predominant. At around pH 10.5,  $\text{CO}_3^{2-}$  becomes the predominant form.

The carbonate system plays an important role in moderation of the chemistry of natural aquatic systems (e.g., acts a pH buffer in seawater), which, in turn, influences biotic and chemical activities. Inorganic C is also a major component of the global C cycle; its subsequent atmospheric/oceanic interaction plays a major role in atmospheric  $\text{CO}_2$  levels and hence global warming processes (see above). For example, an oceanic uptake of 40% will only correspond to an average change in DIC of  $1 \mu\text{mol l}^{-1}$ ; therefore, its accurate and precise determination and distribution is essential.

## Determination of Inorganic Carbon in Aquatic Systems

Samples collected for the subsequent analysis of TIC would typically require the use of polyethylene/borosilicate glass stoppered bottles (rinsed at least twice with the sample) to which is added a biocide (such as chloroform, mercuric chloride) and stored (avoiding a headspace) in the dark at  $4^\circ\text{C}$  (if immediate sample analysis is not possible). Airtight seals on the sample bottle should be ensured so that there is no loss or gain of  $\text{CO}_2$ . If TIC is to be determined by alkalinity titration, the pH of the sample should be noted upon sample collection.

Several approaches have been adopted to determine the TIC. Perhaps the most commonly applied is the indirect approach by determination of the total alkalinity via titrimetry, and then computation of the various inorganic C species assuming attainment of

chemical equilibria. The total alkalinity is defined as the number of moles of  $\text{H}^+$  equivalents to the excess of proton acceptors (bases formed from weak acids with a dissociation constant  $K < 10^{-4.5}$  at  $23^\circ\text{C}$ ) over proton donors (acids with  $K > 10^{-4.5}$ ) species in one kilogram of sample. Whereas the carbonate alkalinity (presented in eqn [3]) represents the contribution made to the alkalinity of inorganic C species only:

$$\begin{aligned} \text{Carbonate alkalinity} \\ = 2[\text{CO}_3^{2-}] + [\text{HCO}_3^-] + [\text{OH}^-] - [\text{H}^+] \quad [8] \end{aligned}$$

Generally, in freshwaters total alkalinity determinations are carried out using additions of strong acids (either  $\text{H}_2\text{SO}_4$  or  $\text{HCl}$  in the concentration ranges  $0.02\text{--}0.1\text{ mol l}^{-1}$ ) to an end point pH of  $\sim 4.2\text{--}5.1$ , using an appropriate indicator or potentiometer. For accurate analysis determination a two-point or multi-point Gran plot method may be employed.

The amount of acid added to reach the end point corresponds to the concentration of  $\text{CO}_3^{2-}$  and  $\text{HCO}_3^-$  in the sample. The TIC may, therefore, be calculated from alkalinity and pH using

$$[\text{Alkalinity}] = C_T(\alpha_1 + \alpha_2) + k_w/[\text{H}_3\text{O}^+] - [\text{H}_3\text{O}^+] \quad [9]$$

( $k_w$  being the ionization constant for water and  $[\text{H}_3\text{O}^+]$  is the proton concentration in water sample). Using this approach, good precision may be obtained ( $< 5\%$ ), assuming all the usual precautions are taken when carrying out the titration.

The titrimetry approach is subject to a number of possible interferences. For example, in aqueous samples that are enriched in particulate material (riverine, estuarine, sewage) filtration is recommended (typically through a  $0.45\text{ }\mu\text{m}$  membrane filter) to remove any particulate surfaces that might contribute to the solution alkalinity. In addition, aqueous samples that contain high concentrations of surfactants (e.g., industrial, sewage) may need longer pH electrode equilibration times during the alkalinity titrations.

It should, however, be remembered that there are aquatic systems where alkalinity is not solely due to the presence of carbonate species, e.g., seawater/aqueous solution derived from anthropic activities. For seawater therefore the alkalinity would be more accurately represented by eqn [10]:

$$\begin{aligned} \text{Seawater total alkalinity} \\ = 2[\text{CO}_3^{2-}] + [\text{HCO}_3^-] + [\text{B}(\text{OH})_4^-] + [\text{HPO}_4^-] \\ + [\text{H}_3\text{SiO}_4] + [\text{MgOH}] + [\text{OH}^-] - [\text{H}^+] \quad [10] \end{aligned}$$

To determine the total alkalinity in seawater, a similar method to that for freshwater samples may be employed, i.e., the sample could be titrated with  $\text{HCl}$  (containing background electrolytes maintaining a comparable ionic strength of seawater). The titration would be monitored using a glass electrode. Total alkalinity may then be calculated from the titrant volume plotted against the electrode potential using a modified Gran plot approach. However, TIC can only be related to alkalinity when the other contributing species concentrations are known. Therefore, a direct method is more preferable. Hence, for the direct determination of TIC in seawater, a  $\text{CO}_2$  coulometer is used. The technique requires the removal of  $\text{CO}_2$  from an acidified (phosphoric acid) seawater sample and the subsequent absorption of  $\text{CO}_2$  in a solution of ethanolamine. The weak acid is then titrated by a strong base, using thymolphthalein as the indicator. The equivalence point is evaluated photometrically ( $610\text{ nm}$ ). Accuracy of the technique is assessed using a sodium carbonate standard solution. The adopted technique performs with a high degree of precision ( $\pm 0.1\%$ ) and accuracy ( $\pm \mu\text{mol kg}^{-1}$ ). This approach was recently adopted for the determination of DIC in seawater for the JGOFS (Joint Global Ocean Flux Study) program.

Aqueous samples may in addition be determined for TIC directly by nondispersive infrared (NDIR) absorption spectrometry providing very good selectivity and sensitivity. Typically during a sample analysis, the dissolved inorganic C is converted to  $\text{CO}_2$  by sample acidification and purged to the sample cell of an NDIR spectrometer using an inert IR carrier gas. Calibration is achieved by using a standard sodium carbonate solution. Interferences may arise from gaseous species evolved into the carrier gas from the sample having overlapping absorption bands. This for natural water samples is most likely to arise from water vapor, which is therefore removed prior to  $\text{CO}_2$  detection.

Following detection, alkalinity may be expressed in a number of units including equivalent concentrations of calcium carbonate in units of  $\text{mg CaCO}_3$ ; micro/milliequivalents, or molarity. Freshwater alkalinity may range from less than  $-200$  to  $500\text{ meq l}^{-1}$ , whereas in seawater a more consistent concentration of  $\sim 2.35\text{ meq l}^{-1}$  is generally observed.

## Inorganic Carbon Species in Soils

TIC in soils predominantly derives from carbonate minerals such as calcite ( $\text{CaCO}_3$ ) and dolomite ( $\text{MgCO}_3$ ). TIC will also be impacted upon by soil's

biotic respiration processes, which in turn may alter the soil pH and hence nutrient availability and uptake. Collection and storage methods of soil samples are not as critical as those required for water and gaseous samples. Samples should be stored dried between 35°C and 38°C. TIC is essentially determined by the addition of acid to the soil sample, converting the inorganic C into gaseous CO<sub>2</sub> which may then be determined by an infrared gas analyzer or a carbon dioxide coulometer (see above; analysis of seawater). Alternatively, CO<sub>2</sub> may be absorbed and then precipitated as BaCO<sub>3</sub>, followed by back titration. Inorganic C in soils has also been achieved by analyzing the CO<sub>2</sub> by a modified pressure calcimeter method (with a detection limit of 0.17 g in organic C kg<sup>-1</sup>). In addition, TIC in soils may be determined using the Van Slyke method. In this method, the volume of CO<sub>2</sub> and air in a reaction vessel, containing the soil sample that has been acidified, is measured. The volume of air and CO<sub>2</sub> is also determined after being passed through a solution of NaOH, effectively removing the CO<sub>2</sub>. Measuring the residual air (at specific temperature and pressures) will allow the determination of the volume of CO<sub>2</sub> evolved from the acidified soil sample, and hence the TIC. Detection limits are ~0.01% (w/w) using this method.

## Determination of Inorganic Carbon in Gaseous Media

Inorganic C in the atmosphere may be present either associated with aerosol material (mainly crustal material, derived from physical weathering of soils and rocks) or as carbon monoxide/dioxide. In terms of human and environmental impacts, gaseous CO and CO<sub>2</sub> are of most importance. Both species have anthropogenic and natural sources. Natural sources include principally biotic processes (i.e., respiration, e.g., bacterial decomposition of plant material) whereas anthropogenic sources include combustion of fossil fuels via the internal combustion engine, power generation, and waste incineration. More recently, deforestation has contributed directly and indirectly to the atmospheric CO<sub>2</sub> levels. Atmospheric concentrations have increased from preindustrial revolution concentrations of ~280 ppmv (parts per million by volume) to ~370 ppmv (current day concentrations). The rapid enhancement of atmospheric CO<sub>2</sub> levels has, of course, led to the potential enhancement of the rate of global warming, leading to potential future climatic change. As a result of climate change, sea level rise, changes in the global oceanic circulation, and rainfall patterns are just a

number of potential environmental impacts. As a result, originating from the Kyoto agreement, nations have agreed to limit their CO<sub>2</sub> emissions. Therefore, there is a need to have accurate analytical systems to monitor ambient atmospheric CO<sub>2</sub> levels, emissions from industrial processes to ensure the efficiency of process control, as well as the effectiveness of environmental legislation and define trend in atmospheric CO<sub>2</sub> levels, to enable the better prediction of future climate change.

The impacts of CO<sub>2</sub> and CO on human health are also a consideration. CO exposure of levels down to 220 ppm may cause physiological effects such as the impairment of mental activity, headache, and irritability with enhanced levels in excess of 800 ppm causing loss of consciousness and even death if exposure is prolonged.

## Sample Collection

Any samples collected away from the analyzer are done so using a discrete sampler with exit and input ports. Air samples may be actively collected with a vacuum system – with a volume of air passed through the system typically being 10 times greater than that of the collection system. Alternatively, samples may be collected in an evacuated container. Continuous monitoring (long-term temporal environmental sampling or industrial process monitoring) should consist of an inert sampling probe located at minimal distance from the analyzer (to provide real-time data) and where high-temperature processes are being monitored and a cooling and dehumidifying system should be incorporated. Particle/corrosive gas removal systems may also be required for monitoring certain industrial processes. These are required to minimize spectral interferences by light scattering and degradation of the transport system and analytical instrumentation, respectively. In addition to sampling discrete air samples or continuously online, passive samplers may be employed.

## Detection of CO<sub>2</sub> and CO in Air

Detection of both gases may simply be carried out by injection of the sample into a GC (gas chromatograph) integrated with a detection system that might be one of the following: thermal conductivity detector, flame ionization detector (FID), or helium glow discharge ionization detector (DID). Precision of such measurements may range in the order of 2% (manual injection) down to 0.5% (direct gas

sampling). Using a TCD, detection limits are typically  $\sim 10 \text{ ng ml}^{-1}$ . Modifying the chromatographic parameters (such as the stationary phase chemical characteristics) may separate any interfering gases. The sensitivity of the detection system for CO may be enhanced (possibly lowering of the detection limits by an order of magnitude) by reacting hydrogen with CO in the presence of a catalyst (heated in Ni) to produce methane before FID detection. However, the methanization will add an extra stage in the analytical process and contamination of hydrogen with methane is often a problem.

Carbon dioxide may be further determined using an NDIR analyzer. The NDIR measures the absorption of IR radiation ( $4.26 \mu\text{m}$ ) due to the presence of  $\text{CO}_2$  traveling through the optical path of the detection system. Measurements are based on a comparison of ambient air- $\text{CO}_2$  mixing ratios in tanks of compressed reference gases. The analyzers operate in differential mode, with a 'zero' reference gas of  $\text{CO}_2$  mixing ratios 20–30 ppm below ambient  $\text{CO}_2$  levels. However, the NDIR is subject to interference by other gaseous species (having overlapping absorption bands) including water vapor, methane, and ethane. The application of the NDIR is the preferred approach for the determination of atmospheric  $\text{CO}_2$  in the atmosphere at remote regions. To lower water vapor interference sampled air is generally dried prior to the introduction into the gas analyzer. Accuracy of such systems are typically  $\pm 0.1 \text{ ppm}$ .

Carbon monoxide may be determined in the work place using direct reading monitors (sampling times of up to 8 h) and detection limits down to 1.6 ppmv (upper measurement limit 2000 ppm) and a precision of  $\sim 10\%$ . The system consists of a three-electrode diffusional electrochemical sensor incorporating a CO filter. A more sensitive approach is to incorporate a GC with a DID detection system. Helium is generally used as the carrier gas/ionized species. Passing the sample through a molecular sieve eliminates  $\text{CO}_2$  interference. The system may achieve detection limits of  $\sim 0.4 \text{ ppm}$ .

*See also:* **Titrimetry:** Potentiometric. **Water Analysis:** Seawater – Dissolved Organic Carbon.

### Further Reading

- Clesceri LS, Greenberg AE, and Eaton AD (eds.) (1998) *Standard Methods for the Examination of Water and Wastewaters*, 20th edn. Washington, DC: American Public Health Association.
- Kramer JR (1982) In: Minear RA and Keith LH (eds.) *Water Analysis, Vol. 1, Inorganic Species*, part 1, ch. 3. New York: Academic Press.
- Langmuir D (1997) Carbonate chemistry, ch. 6. In: *Aqueous Environmental Geochemistry*, pp. 193–230. Englewood Cliffs, NJ: Prentice-Hall.
- Robinson C and Williams PJLeB (1991) Development and assessment of an analytical system for the accurate and continual measurement of total inorganic carbon. *Marine Chemistry* 34: 157–175.

## CATALYTIC TECHNIQUES

*See* **KINETIC METHODS:** Catalytic Techniques

## CATHODIC STRIPPING VOLTAMMETRY

*See* **VOLTAMMETRY:** Cathodic Stripping

# CEMENT

**P W Hurley**, Huntingdon, UK

**R G Pritchard**, Philips Analytical X-Ray b.v., Almelo, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 567–572, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

## Production of Cement

Portland cement is a complex mixture of compounds formed from the oxides of calcium (CaO), silicon (SiO<sub>2</sub>), aluminium (Al<sub>2</sub>O<sub>3</sub>), and iron (Fe<sub>2</sub>O<sub>3</sub>). In addition to these four main constituents, it also contains smaller amounts of magnesium oxide (MgO) and oxides of the alkali metals potassium (K<sub>2</sub>O) and sodium (Na<sub>2</sub>O). It is produced by heating together naturally occurring raw materials containing the required oxides in a kiln at ~1400–1500°C, which results in a product called clinker. The clinker is then ground together with gypsum (which controls setting) to a fine powder. Additional constituents such as granulated blast furnace slag, pulverized fuel ash (PFA), or limestone filler may be used. The grinding process and the addition of these constituents contribute to the different setting times and strength growth patterns of the range of cements needed to meet market requirements.

The source of CaO in the clinker is calcium carbonate either as chalk or limestone, and the SiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>, and Fe<sub>2</sub>O<sub>3</sub> are normally obtained from either clay or shale. The raw materials are quarried, crushed, and blended in the required proportion and the resulting powdered material (raw meal) is then fed into a rotating kiln. The raw meal passes through a burning zone in which the high temperature is generated by burning finely ground coal inside the kiln. The compounds formed can break down on cooling, and it is necessary to cool the resulting clinker rapidly to preserve the cementitious compounds that have been formed. As a final step in the process, the clinker is mixed and ground with a proportion of calcium sulfate to provide a product with the required setting and strength properties. The calcium sulfate can be naturally occurring gypsum, or come from an industrial by-product such as power station desulfurization waste. Other materials may be added either as ‘fillers’, e.g. raw meal, or as ‘extenders’, when substances that have some inherent cementitious characteristics are used, e.g. PFA or

**Table 1** Typical compositions of Portland cements

Oxide	OPC (wt. %)	WPC (wt. %)	SRPC (wt. %)
SiO <sub>2</sub>	20.7	22.5	20.5
Al <sub>2</sub> O <sub>3</sub>	5.75	4.50	3.75
Fe <sub>2</sub> O <sub>3</sub>	2.50	0.30	5.50
Mn <sub>2</sub> O <sub>3</sub>	0.05	0.03	0.05
TiO <sub>2</sub>	0.30	0.33	0.30
P <sub>2</sub> O <sub>5</sub>	0.15	0.17	0.15
CaO	64.0	67.5	64.5
MgO	1.00	0.35	0.75
SO <sub>3</sub>	2.75	2.50	2.20
K <sub>2</sub> O	0.60	0.10	0.40
Na <sub>2</sub> O	0.20	0.12	0.20
Free lime	2.0	2.5	1.0

Loss on ignition as CO<sub>2</sub> and H<sub>2</sub>O is 0.5–3.0 wt. %.

blast furnace slag. It is more common for these extenders to be added by the concrete makers rather than at the cement works.

The most commonly produced cement is Portland cement to British Standard BS 12, which represents ~90% of the market, which includes products such as rapid hardening and coarse ground cements, and white Portland cement (WPC). Other cements covered by other British Standards include sulfate resisting Portland cement (SRPC), Portland blast furnace cement, high slag blast furnace cement, and Portland PFA cement.

WPC is so-called because of its color. It is made from the purest chalk and kaolinite, and contains much less Fe<sub>2</sub>O<sub>3</sub> than the normal gray Portland cement. SRPC is produced for its resistance to sulfate attack, and has a higher Fe<sub>2</sub>O<sub>3</sub> content, often added to the raw meal as iron oxide. The other types named above contain a major (>5 wt. %) constituent. They are used for a variety of reasons, sometimes simply cost reduction, but often to exploit their particular properties. Standards and Codes of Practice regulate their use by specifying appropriate applications and minimum strength, and/or minimum cement content, of a concrete.

Typical compositions of three types of Portland cement are shown in Table 1. Although other oxides are present, it is the contents of CaO, SiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub> and Fe<sub>2</sub>O<sub>3</sub>, and their state of combination, that are of prime importance in the production of high quality Portland cement.

## Cementitious Compounds

The firing together of the calcareous and siliceous materials at 1400–1500°C produces a mixture of



four crystalline compounds. They are tricalcium silicate ( $3\text{CaO} \cdot \text{SiO}_2$ ), dicalcium silicate ( $2\text{CaO} \cdot \text{SiO}_2$ ), tricalcium aluminate ( $3\text{CaO} \cdot \text{Al}_2\text{O}_3$ ), and an aluminoferrite phase which approximates to tetracalcium aluminoferrite ( $4\text{CaO} \cdot \text{Al}_2\text{O}_3 \cdot \text{Fe}_2\text{O}_3$ ). In practice, small quantities of other elements are also present in solid solution in these compounds, and in the cement industry they are known by the shortened names  $\text{C}_3\text{S}$  or alite,  $\text{C}_2\text{S}$  or belite,  $\text{C}_3\text{A}$  and  $\text{C}_4\text{AF}$ , respectively. Provided that the compositions of the raw materials are accurately known, and the correct amounts are blended together to make the raw meal, then all the  $\text{CaO}$  present will potentially react with one of the other oxides to form the above compounds. In practice, the reactions are never entirely complete and some unreacted 'free lime' will be present in the clinker. A high free lime content, for instance by incorrect formulation of the raw meal, can give rise to expansion and cracking in concrete. If an excess of siliceous material is present, this adversely affects the  $\text{C}_3\text{S}$  content of cement, and consequently decreases strength development at early ages.

## Control and Quality Parameters

The production of high-quality cement is a combination of control of the composition of the blended raw materials that are fed to the kiln, and quality checks on the composition of the resulting clinker and cement. Perhaps the single most important control parameter is the chemical composition of the raw meal, which in turn relies on a knowledge of the composition of the raw materials. If these two parameters are precisely controlled, and the kiln is operated in the required manner, then the composition of the clinker should be correct. In practice of course, its quality and that of the final product are regularly checked.

Although cement making employs and produces what is essentially a mixture of oxides, control of the process is monitored through the determination of a series of moduli which are derived from the concentrations of the oxides.

These moduli provide information which is more easily interpreted than the individual oxide contents, and yield key information leading to predictions on the chemical species in the final product.

The main modulus used to monitor the blending of the raw materials is called the lime saturation factor (LSF). This is a measure of the chemical balance between the calcareous and siliceous components. When there is complete balance the LSF has a value of 100%, and this theoretically leads to the

maximum possible proportion of the silica being combined as  $\text{C}_3\text{S}$ . The LSF is given by the expression (where compounds are expressed as wt.%)

$$\text{LSF} = \frac{[(\text{CaO}) - 0.7(\text{SO}_3)] \times 100\%}{[2.8(\text{SiO}_2) + 1.2(\text{Al}_2\text{O}_3) + 0.65(\text{Fe}_2\text{O}_3)]}$$

The  $\text{SO}_3$  content may not be appropriate for raw meal.

Since it is known that there is always some uncombined  $\text{CaO}$  in the clinker, another modulus is used which takes account of this 'free lime'. It is called the lime combination factor (LCF) and is given by the expression

$$\text{LCF} = \frac{[(\text{CaO}) - 0.7(\text{SO}_3) - (\text{free CaO})] \times 100\%}{[2.8(\text{SiO}_2) + 1.2(\text{Al}_2\text{O}_3) + 0.65(\text{Fe}_2\text{O}_3)]}$$

Two other moduli that are used allow the proportions of  $\text{C}_3\text{S}$ ,  $\text{C}_2\text{S}$ ,  $\text{C}_3\text{A}$ , and  $\text{C}_4\text{AF}$  to be predicted. These are the silica ratio and the alumina-iron ratio. They are expressed (as wt.%) as follows:

$$\text{silica ratio} = \frac{\text{SiO}_2}{\text{Al}_2\text{O}_3 + \text{Fe}_2\text{O}_3}$$

$$\text{alumina-iron ratio} = \frac{\text{Al}_2\text{O}_3}{\text{Fe}_2\text{O}_3}$$

## Bogue Calculations

Whilst the above moduli are used for online control of the process, the ultimate properties of the cement, such as setting time and strength, depend on the contents of the compounds  $\text{C}_3\text{S}$ ,  $\text{C}_2\text{S}$ ,  $\text{C}_3\text{A}$ , and  $\text{C}_4\text{AF}$  present in the clinker. These may either be arrived at by direct analytical means or alternatively, and more commonly, by using what are known as the Bogue calculations. These calculations use the contents of the four major oxides,  $\text{CaO}$ ,  $\text{SiO}_2$ ,  $\text{Al}_2\text{O}_3$  and  $\text{Fe}_2\text{O}_3$ , to predict, with stated assumptions, the contents of the above compounds.

First, it is assumed that all the  $\text{Fe}_2\text{O}_3$  is present as  $\text{C}_4\text{AF}$ . Then from the  $\text{Fe}_2\text{O}_3$  content, and using stoichiometry, the  $\text{C}_4\text{AF}$  content is calculated. This calculation also yields the amounts of  $\text{CaO}$  and  $\text{Al}_2\text{O}_3$  required, and therefore by difference the amounts remaining for further compound formation.

The next assumption is that all the remaining  $\text{Al}_2\text{O}_3$  is present as  $\text{C}_3\text{A}$ . Again using stoichiometry the amount of  $\text{C}_3\text{A}$  can be calculated.

Some of the remaining  $\text{CaO}$  is present as 'free lime', and some is also present in the gypsum which is added. So assuming that all the  $\text{SO}_3$  is present as

$\text{CaSO}_4$ , then the  $\text{CaO}$  combined with the  $\text{SO}_3$  can be calculated.

Ignoring for the sake of simplicity the possible role of other minor constituents, the  $\text{CaO}$  remaining, after subtracting from the total  $\text{CaO}$  the amounts present as  $\text{C}_3\text{A}$ ,  $\text{C}_4\text{AF}$ , free lime, and gypsum, is combined with  $\text{SiO}_2$  as  $\text{C}_3\text{S}$  and  $\text{C}_2\text{S}$ .

To arrive at values for these two compounds, the amount of  $\text{CaO}$  required to convert the whole of the silica to  $\text{C}_2\text{S}$  is calculated, and then the remaining  $\text{CaO}$  is used to convert its equivalent of  $\text{C}_2\text{S}$  to  $\text{C}_3\text{S}$ . Thus the contents of the four compounds can be calculated from concentrations of  $\text{Al}_2\text{O}_3$ ,  $\text{Fe}_2\text{O}_3$ ,  $\text{CaO}$ ,  $\text{SiO}_2$ ,  $\text{SO}_3$ , and the free lime.

These calculations can be condensed into four equations (where compounds are expressed as wt.%):

$$\text{C}_3\text{S} = 4.07 \times (\text{CaO} - \text{free lime}) - 7.60 \times \text{SiO}_2 \\ - 6.72 \times \text{Al}_2\text{O}_3 - 1.43 \times \text{Fe}_2\text{O}_3 - 2.85 \times \text{SO}_3$$

$$\text{C}_2\text{S} = 2.87 \times \text{SiO}_2 - 0.7544 \times \text{C}_3\text{S}$$

$$\text{C}_3\text{A} = 2.65 \times \text{Al}_2\text{O}_3 - 1.69 \times \text{Fe}_2\text{O}_3$$

$$\text{C}_4\text{AF} = 3.04 \times \text{Fe}_2\text{O}_3$$

## Analytical Requirements

Cement production is a continuous process, with the blended raw meal entering the kiln at one end, and clinker emerging at the other to be mixed and ground with gypsum and optionally other fillers to produce the final product. This process continues 365 days a year, and close continuous analytical control must be maintained to ensure product quality. Since a large kiln is capable of producing 100–200 tons per hour, analysis must be rapid, whilst at the same time being both precise and accurate.

Table 2 shows the typical range for Portland cement, although in practice the range for an individual works would be much narrower.

## Analytical Methods

### Determination of $\text{CaO}$ , $\text{SiO}_2$ , $\text{Al}_2\text{O}_3$ , $\text{Fe}_2\text{O}_3$ , $\text{MgO}$ , $\text{SO}_3$ , $\text{K}_2\text{O}$ , and $\text{Na}_2\text{O}$ in Raw Materials, Raw Meal, Clinker, and Cement

Although chemical methods exist and are published in BS 4550 and EN 196, X-ray fluorescence analysis has become the standard industrial method for determining these eight oxides in raw materials, raw meal, clinker, and cement. It fulfills all the requirements for speed, precision, and accuracy and is also simple to use, such that control analysis can be

**Table 2** Composition ranges for Portland cement

Oxide	Concentration range (wt.%)
$\text{SiO}_2$	18.0–24.0
$\text{Al}_2\text{O}_3$	4.0–8.0
$\text{Fe}_2\text{O}_3$	1.5–4.5
$\text{Mn}_2\text{O}_3$	0.03–0.5
$\text{TiO}_2$	0.2–0.4
$\text{P}_2\text{O}_5$	0.05–0.3
$\text{CaO}$	62.0–66.0
$\text{MgO}$	0.7–4.0
$\text{SO}_3$	1.5–3.5
$\text{K}_2\text{O}$	0.1–1.5
$\text{Na}_2\text{O}$	0.1–0.9
Free lime	0.5–3.0
Loss on ignition	up to 3.0
Insoluble residue	up to 1.5

carried out by relatively inexperienced shift chemists on a 24 h per day basis.

Three types of X-ray fluorescence (XRF) spectrometers are available:

1. Wavelength-dispersive scanning XRF spectrophotometers.
2. Wavelength-dispersive multichannel XRF spectrophotometers.
3. Energy-dispersive XRF spectrophotometers.

All three types can be found in use in cement works control laboratories, but the preferred system is the multichannel type fitted with channels for the simultaneous determination of the eight elements Fe, Ca, K, S, Si, Al, Mg, and Na. Depending on the local geology, a channel for the determination of fluorine may be fitted if the limestone deposit is close to a source of fluorspar ( $\text{CaF}_2$ ). (The determination of fluorine is beyond the scope of an energy-dispersive XRF spectrometer.) At some works a channel is added to monitor the chlorine content of the clinker, which can be introduced from the fuel used to fire the kiln.

A suitable sample for an X-ray spectrometer is normally circular with a diameter of 30–40 mm and thickness about  $\sim 3$ –5 mm. For oxide materials a choice must be made between either direct pelletization of the powdered sample, or fusion of the sample with a suitable flux to form a glass bead of the appropriate dimensions. This choice rests between speed and accuracy, with direct pelletization being faster, and fusion potentially more accurate. As a general rule pelletization may be used for raw material and raw meal analysis, when the seams are reasonably pure, and when only two materials are employed (i.e., chalk/limestone and clay/shale).

When raw material composition is very variable, or when more than two are used as in SRPC, when iron ore and sand may be added, then fusion is the preferred technique.

### Sample Preparation by Direct Pelletization

For acceptable accuracy the particle size of the powdered sample for X-ray fluorescence should be less than the minimum path length of the X-radiation being measured in the analysis. For the elements present in cement this means a particle size of less than  $\sim 10\mu\text{m}$ . Since the particle size of the samples received for analysis is normally much greater than this, it is necessary to reduce it using a laboratory grinding mill. Those employed are usually of the 'swing mill' type, with a tungsten carbide barrel and grinding tools. Empirical grinding trials should be carried out to determine the optimum mill loading and grinding time to achieve a constant spectral line intensity for all the elements to be determined.

When suitable grinding parameters have been found, an aliquot of the ground powder is pressed into a pellet, normally of 40 mm diameter, up to 40 tons per square inch ( $6.2 \times 10^5$  kPa). The pellet may either be self-supporting or be made in an aluminum cup or metal ring to provide support. In order to produce a robust pellet, it may be necessary to use a chemical such as stearic acid or boric acid. Stearic acid can be introduced into the grinding barrel with the sample to be ground. It not only acts as a binder for the pellet, but also as a grinding aid to prevent agglomeration of the particles. Alternatively boric acid may be introduced into the bottom of the aluminum cup before pressing, with the ground powder placed on top. There are many different chemicals that can be used in this way, and for further information, reference should be made to texts on X-ray fluorescence. It should be said that failure to produce a robust pellet can lead to dust ingress to the spectrometer which can adversely affect its performance.

Manual, semiautomatic and fully automatic grinding mills and pelletizing presses are available, and there are machines that combine all these operations including cleaning the system between samples.

### Sample Preparation by Fusion

Elimination of the so-called particle size and mineralogical effects in X-ray spectrometry is the key to accuracy in the analysis of powdered materials. As previously stated, there are circumstances when acceptable accuracy can be achieved using pressed powder samples, but in many cases it is necessary to dissolve the sample in a suitable medium to destroy

the particles and render the sample homogeneous. Many cement works have now converted their XRF sample preparation to fusion having previously used pressed pellets, and the move in this direction seems to be inevitable as product quality requirements become more stringent, and sources of pure raw materials become less readily available.

The fusion medium in most common use is either sodium or lithium tetraborate ( $\text{Na}_2\text{B}_4\text{O}_7$  or  $\text{Li}_4\text{B}_2\text{O}_7$ ). Either may be used, but the sodium salt, whilst cheaper precludes the possibility of analyzing for sodium. It is also slightly deliquescent, so calibration standards must kept in a vacuum desiccator. Lithium tetraborate suffers from the disadvantage that its softening point is in the region of  $1200^\circ\text{C}$ , a temperature beyond the reach of some laboratory muffle furnaces. A mixture of lithium metaborate ( $\text{LiBO}_2$ ) and tetraborate in a ratio of 4:1 reduces softening point considerably, and is also reported to be more reactive with siliceous materials. This mixed flux is commercially available.

The fusion method involves dissolving an accurately known weight of the sample in an accurately known weight of the flux. The dissolution is usually carried out in a platinum alloy crucible, with the melt being cast into a bead in a preheated casting dish of the same alloy. A bead of suitable size for the spectrometer should weigh  $\sim 9$ – $10$  g, and flux to sample ratios in common use are 10:1 and 5:1. This means either 1 g or 1.5 g of sample and 10 g or 7.5 g of flux, respectively. The ratio 5:1 is preferable since not only does it dilute the sample to a lesser extent, thus achieving greater sensitivity of the X-ray measurement, but it is also less expensive on flux. Fusion can be carried out either in a laboratory muffle furnace, over a gas burner, or in one of the purpose-built commercially available fusion machines. Some of these machines are fully automatic, and include automatic weighing, multisample handling, and crucible-cleaning devices.

### Calibration

Calibration of the spectrometer is carried out by measuring samples of known composition to derive the parameters for the calculation of the concentration of each element in an unknown sample from the measured X-ray intensities. For the pressed powder method calibration samples covering a range of composition should be obtained from the normal production process. Since the aim of the process is to produce a constant composition, this can be difficult to achieve. The temptation to produce calibration samples by 'spiking' production samples to produce a range should be avoided, since this practice can lead

to gross inaccuracy. The powder method will only be successful when the calibration is carried out with samples that are from production and not synthetically produced. Powdered calibration samples can either be analyzed using wet chemical techniques, or by X-ray fluorescence spectrometry using the fusion method.

Calibration for the fusion method is achieved by the dissolution of carefully weighed aliquots of pure chemicals in the flux to provide a set of synthetic calibration beads to cover the required analytical range for each element. Where the pure oxide of any element is not available, then a compound which will convert to the oxide during fusion should be used. For example  $K_2CO_3$  can be used for  $K_2O$ . Normally five to ten beads should be sufficient, but care must be taken in the design of the set. Each individual bead should contain the equivalent of 100% (i.e., the sample weight for unknown samples, e.g., 1.5 g), and with the concentration range for each element well covered to avoid any possibility of extrapolation. Also the concentration series for one element should not follow the values for any other, either in an increasing or decreasing progression. This is important should the calculation of interelement correction factors be necessary.

The chosen calibration samples should be measured such that the counting statistical error for each element is small compared with other errors contained within the samples. These other errors are those associated with the chemical analysis of powdered samples and the weighing and other manipulations concerned with glass bead making. It is common practice to measure calibration samples for longer times than those used for the measurement of unknown samples. This helps to eliminate any unnecessary error in the calibration and therefore in the final analysis of unknowns.

Since the composition of the sample is variable, and therefore so is its mass absorption coefficient, it is sometimes necessary to carry out interelement corrections. The values of the correction factors can either be calculated from fundamental theory using commercially available software, or by multilinear regression analysis using the calibration samples. The former is preferred since the latter can force a mathematical fit to the data which bears no relation to X-ray theory.

**Determination of free lime** The determination of the uncombined CaO can be tackled in one of two ways. Historically this has been done using an acid-base titration after reaction of a weighed sample of clinker with ethylene glycol. The method is rapid and reliable and is still in common use. It is published in BS 4550 Part II.

More recently the industry has been investigating the use of an X-ray diffraction method, and this is now becoming accepted as an alternative to the chemical method. The advantages of instrumental techniques are clear in that they are much less prone to human error and by and large are simpler and less time-consuming. However, their accuracy must be proven, and the problems of preferred orientation and particle size effects have hindered such proof for the determination of free lime by X-ray diffraction. It should be borne in mind that CaO will react with moist air to form the hydroxide, the carbonate and the hydrogencarbonate. Analysis should be carried out rapidly once the sample has been taken, since the formation of these other chemical species will seriously affect the accuracy of the free lime determination, in particular by X-ray diffraction.

**Determination of the cementitious compounds  $C_3S$ ,  $C_2S$ ,  $C_3A$ , and  $C_4AF$**  As stated above, the common practice in cement works is to determine the contents of these compounds by using Bogue calculations together with the necessary assumptions. It should be stressed that the quantities of these compounds that are formed are a function, not only of the chemical composition, but also of the process itself, and the theoretical and actual quantities have been shown to be quite different in some cases. These inherent dangers are well recognized within the industry, and alternative methods are seen to exhibit inaccuracies of the same or greater magnitude.

Microscopy with point-counting can be used and, although it can be time-consuming, it is quite attractive in terms of its low capital cost. It also produces information on kiln performance not available from other methods.

The only other method that has gained any credence is X-ray diffraction. Whilst it is true that a full understanding of the reactions that take place, and the compounds that are formed in a cement kiln has resulted largely from X-ray diffraction studies, accurate quantitative analysis of cement clinker for the cementitious compounds has remained a research and development tool rather than a technique for industrial analysis. The accuracy of results achieved by quantitative X-ray diffraction analysis (QXDA) is limited by modifications to the diffracted beam intensities caused by substitution and solid solution. While the accuracy of Bogue calculations is limited by the soundness of the assumptions made, comparisons of results obtained using Bogue calculations and X-ray diffraction exhibit large discrepancies at times, but the decision as to which set of results is the more accurate remains difficult to take. Since the majority of cement works operate successfully using

Bogue calculations, it must be assumed that they are meaningful.

### Sample Preparation for QXDA

As in X-ray spectrometry, there is a need to reduce the particle size to below the level at which it will adversely affect the diffracted intensities. In practice, this means to about the 5  $\mu\text{m}$  level. Since the action of grinding the sample can at times induce phase changes, it is important that the grinding action is not too vigorous. This can be achieved by slow grinding in an agate ball mill with an inert liquid medium such as cyclohexane. This helps to keep the temperature down and also acts as a grinding aid.

Calibration samples and samples for analysis should be treated in the same way. The ground sample should be introduced into the sample holder of the diffractometer using the 'back-loading' method which reduces the effects of preferred orientation.

### Calibration

Calibration can be achieved by the use of synthetically prepared cement minerals, although these may be difficult to obtain. Since the composition of the sample is variable, and therefore so is its mass absorption coefficient, it is advisable to introduce an internal standard to both the calibration samples and samples for analysis. The elimination of the effect of variable absorption is achieved simply by taking all measured lines as a proportion of a line from the internal standard.

**Determination of loss on ignition and insoluble residue** Loss on ignition is determined by heating a weighed quantity of sample in a vitreous crucible at 950–1000°C in a muffle furnace.

Insoluble residue is determined by reacting the sample with hydrochloric acid, followed by solution in sodium carbonate and weighing the ignited insoluble residue.

Both methods are published in full in (BS) EN 196.

### Automation

X-ray fluorescence spectrometry has been established as the prime analytical technique for cement works control since the early 1970s. During the 20 years that have followed, X-ray spectrometers have been incorporated into complete control systems, which include sample transport from the sampling points to the works laboratory, sample preparation and transport into the spectrometer, analysis, calculation of control moduli, and generation and feedback of control signals to the plant to modify the process when necessary.

More recently, systems are being introduced that include an X-ray diffraction capability for the determination free lime content, and in some cases the  $\text{C}_3\text{S}$  (alite) content, of the clinker.

*See also:* **Building Materials. Ceramics. Sample Handling:** Comminution of Samples; Automated Sample Preparation. **X-Ray Fluorescence and Emission:** Wavelength Dispersive X-Ray Fluorescence; Energy Dispersive X-Ray Fluorescence.

### Further Reading

American Institute of Concrete (2001) *ACI Manual of Concrete Practice*. Farmington Hills: ACI.  
ASTM International (2002) *Annual Book of ASTM Standards*. Conshohocken: ASTM International.  
Lea FM and Hewlett PC (1998) *Lea's Chemistry of Cement and Concrete*. London: Arnold.

# CENTRIFUGATION

Contents

**Analytical Ultracentrifugation**

**Preparative**

## Analytical Ultracentrifugation

**A Furst**, Beckman Coulter, Inc., Palo Alto, CA, USA  
**D N Taulbee**, University of Kentucky, Lexington, KY, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Analytical centrifugation is a method of studying the hydrodynamic properties of a molecule as it moves through a fluid medium. Hydrodynamic behavior depends on mass, density, and shape, so one can study these properties of a molecule by accelerating

Bogue calculations, it must be assumed that they are meaningful.

### Sample Preparation for QXDA

As in X-ray spectrometry, there is a need to reduce the particle size to below the level at which it will adversely affect the diffracted intensities. In practice, this means to about the 5  $\mu\text{m}$  level. Since the action of grinding the sample can at times induce phase changes, it is important that the grinding action is not too vigorous. This can be achieved by slow grinding in an agate ball mill with an inert liquid medium such as cyclohexane. This helps to keep the temperature down and also acts as a grinding aid.

Calibration samples and samples for analysis should be treated in the same way. The ground sample should be introduced into the sample holder of the diffractometer using the 'back-loading' method which reduces the effects of preferred orientation.

### Calibration

Calibration can be achieved by the use of synthetically prepared cement minerals, although these may be difficult to obtain. Since the composition of the sample is variable, and therefore so is its mass absorption coefficient, it is advisable to introduce an internal standard to both the calibration samples and samples for analysis. The elimination of the effect of variable absorption is achieved simply by taking all measured lines as a proportion of a line from the internal standard.

**Determination of loss on ignition and insoluble residue** Loss on ignition is determined by heating a weighed quantity of sample in a vitreous crucible at 950–1000°C in a muffle furnace.

Insoluble residue is determined by reacting the sample with hydrochloric acid, followed by solution in sodium carbonate and weighing the ignited insoluble residue.

Both methods are published in full in (BS) EN 196.

### Automation

X-ray fluorescence spectrometry has been established as the prime analytical technique for cement works control since the early 1970s. During the 20 years that have followed, X-ray spectrometers have been incorporated into complete control systems, which include sample transport from the sampling points to the works laboratory, sample preparation and transport into the spectrometer, analysis, calculation of control moduli, and generation and feedback of control signals to the plant to modify the process when necessary.

More recently, systems are being introduced that include an X-ray diffraction capability for the determination free lime content, and in some cases the  $\text{C}_3\text{S}$  (alite) content, of the clinker.

*See also:* **Building Materials. Ceramics. Sample Handling:** Comminution of Samples; Automated Sample Preparation. **X-Ray Fluorescence and Emission:** Wavelength Dispersive X-Ray Fluorescence; Energy Dispersive X-Ray Fluorescence.

### Further Reading

American Institute of Concrete (2001) *ACI Manual of Concrete Practice*. Farmington Hills: ACI.  
ASTM International (2002) *Annual Book of ASTM Standards*. Conshohocken: ASTM International.  
Lea FM and Hewlett PC (1998) *Lea's Chemistry of Cement and Concrete*. London: Arnold.

# CENTRIFUGATION

Contents

**Analytical Ultracentrifugation**

**Preparative**

## Analytical Ultracentrifugation

**A Furst**, Beckman Coulter, Inc., Palo Alto, CA, USA  
**D N Taulbee**, University of Kentucky, Lexington, KY, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Analytical centrifugation is a method of studying the hydrodynamic properties of a molecule as it moves through a fluid medium. Hydrodynamic behavior depends on mass, density, and shape, so one can study these properties of a molecule by accelerating

a sample preparation in a centrifugal field. The analytical ultracentrifuge (AUC) contains a built-in optical system allowing one to observe the movement of a sample as it is spun in a centrifuge rotor. The centrifuge permits molecules to be studied in their native state, in solution, and has been useful in characterizing how proteins and other biological macromolecules bind to one another to form higher ordered structures. It has many other applications as well, including studies of polymers and colloids.

A great advantage of centrifugal studies is that the movement of particles in a centrifugal field can be accurately described by the laws of physics. Measurements can be made without comparisons to molecular standards. This makes it particularly useful for the study of proteins with properties that may differ from those of typically globular protein standards, such as glycoproteins, highly asymmetric molecules, and systems where association is reversible and concentration dependent. Analytical ultracentrifugation has also been referred to as solution interaction analysis to reflect its utility in the study of interacting systems.

The first AUC was built by T Svedberg, J B Nichols, and co-workers at Upsala in the 1920s. Svedberg began to develop the method, and used it to characterize proteins and other molecules in solution, demonstrating the individual character of biological macromolecules. His work led to the 1926 Nobel prize in Chemistry. The technology was commercialized by the Specialized Instrument Company (Spinco) in 1946, in the form of the Model E analytical ultracentrifuge. Now a part of Beckman Coulter, Inc., it is the sole manufacturer of AUC equipment today. Modern versions of AUC make use of a variety of optical systems and use an external computer for facilitated data acquisition and analysis.

## Theory

### Sedimentation

When a particle in a fluid medium is subjected to a centrifugal field it experiences three forces: sedimentation, buoyancy, and friction. The sedimentation force,  $F_s$ , is proportional to the particle mass,  $m$ , and to the centrifugal acceleration:

$$F_s = m\omega^2 r \quad [1]$$

where  $\omega$  is the rotational velocity and  $r$  is the distance from the center of rotation. The sedimentation force is opposed by a buoyant force,  $F_b$ , which is

proportional to the mass of the medium displaced by the particle:

$$F_b = -m\bar{v}\rho\omega^2 r \quad [2]$$

where  $\rho$  is the solvent density and  $\bar{v}$ , the partial specific volume, is the volume of fluid displaced per unit mass of solute. If the particle is less dense than the medium, the buoyant force will be greater than the sedimentation force, and the particle will float. Finally, the particle experiences a frictional resistance to its motion

$$F_f = -f\nu \quad [3]$$

where  $\nu$  is the particle velocity and  $f$  a frictional coefficient that depends upon the size and shape of the particle as well as the viscosity of the medium.

During centrifugation, these forces come quickly into balance and the particle reaches a terminal velocity,  $\nu$ , described by the Svedberg equation:

$$s = \frac{\nu}{\omega^2 r} = m(1 - \bar{v}\rho)/f \quad [4]$$

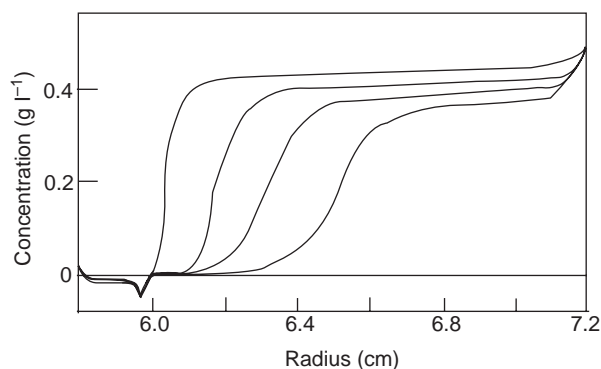
This equation defines a sedimentation coefficient,  $s$ , which may be thought of as the intrinsic speed of the particle and which depends on the molar mass and on the frictional coefficient. The sedimentation coefficient is expressed in Svedberg units, which equal  $10^{-13}$  s. Since it varies with temperature and solvent viscosity, it is often expressed as  $s_{20,w}$ , that is, the equivalent in 20°C water. The quantity  $m(1 - \bar{v}\rho)$  is called the buoyant molecular weight, and corresponds to the mass of the particle less than that of the displaced medium.

### Dynamic Distribution

In the case of macromolecules in a medium of lower density, centrifugation will cause the macromolecules to sediment, that is, to move through the medium and eventually to pellet on the outer wall of the sample compartment. At high centrifugal speeds, a boundary is observed at the top of the sample column, corresponding to the trailing edge of the solute distribution. In a sedimentation velocity experiment, a series of optical scans made during the pelleting of a sample is analyzed to determine its rate of movement and its shape (Figure 1).

By measuring the rate at which the boundary moves, an average sedimentation coefficient for the sample may be obtained. If multiple components are present, the boundary will broaden and may even resolve into discrete steps. However, diffusion will also cause the boundary to spread. Numerical or graphical methods are usually required to distinguish





**Figure 1** Formation and movement of the boundary in a sedimentation velocity experiment.

diffusional broadening from the presence of multiple species.

The evolution of the boundary is described by a partial differential equation, the Lamm equation

$$\frac{dc}{dt} = \frac{1}{r} \frac{d}{dr} \left[ rD \frac{dc}{dr} - s\omega^2 r^2 c \right] \quad [5]$$

This describes how the rate of change of a concentration distribution  $c(r, t)$  will be related to  $s$  and to the particle's diffusion coefficient,  $D$ .

Analysis of the sedimentation boundary and its displacement is used to study sample purity and molecular shape, and to describe systems such as polymers and associating macromolecules.

### Equilibrium Distribution

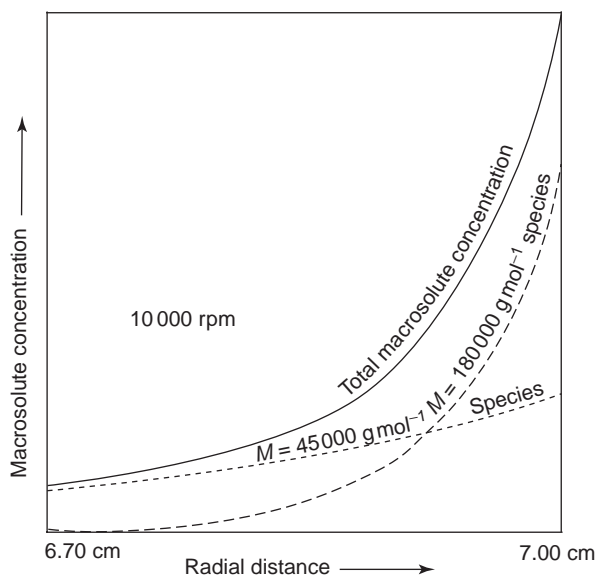
The boundary is a concentration gradient, and solutes will tend to diffuse against the gradient. At low centrifugal speeds, diffusional spreading will oppose sedimentation until eventually an equilibrium is achieved. Sedimentation equilibrium experiments lead to static exponential concentration distributions like the one shown in **Figure 2**.

At equilibrium, by definition, the concentration distribution no longer changes with time, and the Lamm equation is equal to zero. Then, an exponential solution may be found, of the form

$$c = c_k \exp \left[ \frac{\omega^2}{2RT} M(1 - \bar{v}\rho) (r^2 - r_k^2) \right] \quad [6]$$

where  $c_k$  is the concentration at a reference radial position  $k$ .

Both equilibrium and velocity studies are used to characterize a sample. Velocity runs are quick and provide an indication of the number of species present. Equilibrium runs give a more precise measurement of the masses of macromolecules and can be used to monitor their interactions. Together,



**Figure 2** Equilibrium distributions for a reversibly associating monomer–tetramer system. Individual species give rise to the lower traces. The upper trace is the sum of these, and is the one actually observed. (Courtesy of Beckman Coulter, Inc., used by permission.)

these complementary methods can be used to characterize the stoichiometry and kinetics of formation of large macromolecular complexes. Equilibrium distributions also provide a tool for determining the molecular weights of proteins with bound sugar or lipid moieties that are not amenable to measurement by techniques that rely on a comparison to protein standards.

### Nonideality

Although the theory of sedimentation as described holds for the near-ideal conditions encountered in most dilute biological preparations, refinements of the theory are required to deal with nonideal effects. Deviations from ideal behavior are caused by charge effects, hydration, or steric interference of molecules in concentrated solutions. Nonideal effects may result in an apparent concentration dependence of the sedimentation coefficient and an artificial sharpening of the boundary.

### Instrumentation

AUC instrumentation consists of a centrifugal drive unit, a rotor in which the sample is held, one or more optical systems to observe the sample, and a computer to acquire and to analyze the data.

Relative to preparative centrifugation, hardware requirements are particularly stringent with respect to rotor stability and temperature control. Rotor precession can lead to uncertainty in the measurement

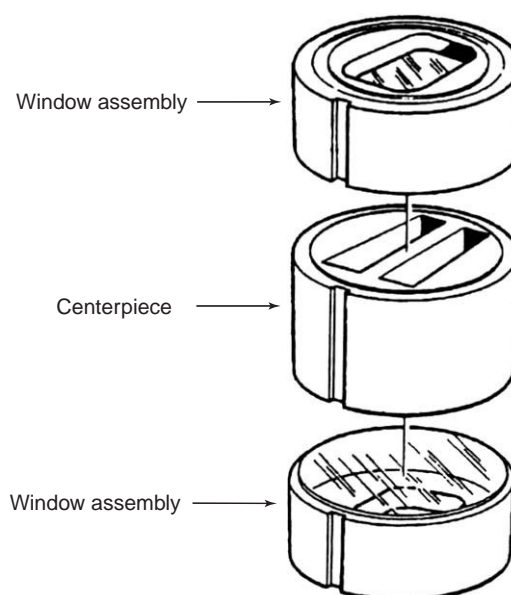
of concentration and of radial position and must be held to an absolute minimum. Precise temperature determination is required for all data analysis calculations, since the sedimentation coefficient depends upon temperature. In addition, temperature instability may cause convective mixing of the sample. For this reason, the rotor is spun in a vacuum to eliminate frictional heating.

### Centrifuge and Rotor Systems

AUC systems are available commercially with several optical systems. The ProteomeLab XL-I<sup>TM</sup> model (Beckman Coulter, Inc.) is equipped with two optical detection systems. A Rayleigh interference system, driven by a 660 nm laser, measures concentration by monitoring refractivity, and a 190–800 nm xenon-lamp system measures sample absorbance. An XLA version, equipped with absorbance-optics only, is also available.

A 60 000 rpm four-position rotor and a 50 000 rpm eight-position rotor are supplied for these instruments. These rotors are made of titanium in order to withstand forces that may reach ~300 000 times the force of gravity. Normally, one sample position in the rotor is occupied by a counterbalance with ports for radial calibration, which is required to correct for the significant stretch exhibited by rotors at high speed. The other positions are occupied by cells in which the sample is held in a centerpiece between two windows. In the centerpiece are one or more sets of paired sample and reference cavities. For velocity experiments, a single pair of sector-shaped cavities is used. Each cavity wall coincides with a radius of the rotor – like slices of a pie – in order to minimize collision effects during sedimentation. Centerpieces for equilibrium experiments may have many sets of small cavities for multiple samples. Smaller cavities allow shorter runtimes, as a steady state is achieved more rapidly. These interchangeable centerpieces provide capacity to simultaneously process as many as 56 samples for equilibrium experiments, or seven samples for velocity runs. Sample volumes between 50 and a few hundred microliters are required, and the samples may be recovered after the run. A typical two-sector centerpiece and window assembly is shown in **Figure 3**.

There are also centerpiece designs that allow the sample to be added to a column of buffer dynamically, i.e., when the rotor is turning. In a band-forming centerpiece a small volume of sample is introduced to a buffer column through a capillary and migrates centrifugally in a narrow band. A synthetic boundary centerpiece allows two fluids to be layered while the rotor spins. This provides a sharp



**Figure 3** Sample cell, consisting of a centerpiece held between quartz windows (exploded view). The assembled cell is inserted into holes in the rotor and allows light to pass through the sample while the rotor spins. (Courtesy of Beckman Coulter, Inc., used by permission.)

interface in a strong centrifugal force field and is used for studies of diffusion.

### Optical Systems

The centerpiece is sealed by pressure between two windows made of quartz or of sapphire. Quartz windows are good for transmitting light in the far-ultraviolet (UV). Sapphire, which is less susceptible to stress deformation in high force fields, is preferred with interference optics.

Absorbance optics are useful for samples that absorb light in the available frequency range, and are frequently employed in the UV for studies of proteins and nucleic acids. The sample is illuminated by flashes from a wide-spectrum xenon lamp. Timing pulses derived from magnets embedded in the rotor are used to ensure that the lamp flashes when either the sample or reference sector is directly over a detector located beneath the rotor on the centrifuge chamber floor. A moving slit provides radial resolution by masking portions of the solution column. The measured absorbance is proportional to centerpiece thickness, i.e., path length, and to sample concentration. Three millimeter and 12 mm path length centerpieces are commercially available. These systems are capable of examining proteins at concentrations down to a few micrograms per milliliter or, in the case of

self-associating systems, for the determination of equilibrium constants as low as  $\sim 10^{-8} \text{ mol l}^{-1}$ .

In an interference optical system, coherent light from a laser is split into two beams. One is passed through the sample, and the other through the reference sector. The beams are then recombined and a pattern of fringes is created by constructive or destructive interference of the light waves. If one beam is slowed by higher refractivity of the sample, the fringe pattern will be shifted from a central reference position. This displacement is related to the sample concentration by

$$\Delta Y = c \cdot \frac{dn}{dc} \cdot L / \lambda \quad [7]$$

where  $\Delta Y$  is the pattern shift measured in fringes,  $dn/dc$  is the refractive increment of the macrosolute,  $L$  the path length (determined by the centerpiece thickness), and  $\lambda$  the wavelength. Interference systems are less sensitive than absorbance optics but are very precise, and cover a broad concentration range. Since all macromolecules cause refractive index changes, they may be used with any solute. Polysaccharides, for example, which do not absorb UV radiation, are amenable to study with this system. Interference optics are typically useful for protein concentrations in the range  $0.1\text{--}10 \text{ mg ml}^{-1}$ .

At the time of writing this article, a fluorescence optical system is under development by Aviv Biomedical Inc., and will be available as an accessory to the XL-I and XL-A instruments. Fluorescence optics provide very high sensitivity for studies of labeled or naturally fluorescent compounds. Test systems have proved useful with concentrations as low as  $800 \text{ pmol}$ , allowing determination of protein equilibrium constants below  $10^{-9} \text{ mol l}^{-1}$ . Studies with fluorescent tracers may also be used to characterize macromolecules at higher concentrations.

## Data Analysis

The analysis of AUC data has undergone a revolution in recent years because of the availability of powerful personal computers. Although graphical methods, which were the norm in the early days of the technique, may still be useful, the personal computer has fueled the development of several new and powerful analysis algorithms.

Desktop computational power has supported two significant trends in data analysis. Global analysis, the simultaneous processing of data from multiple experiments, has allowed characterization of molecular systems over broad ranges of concentration. And the use of iterative methods to arrive at approximate

solutions to the Lamm equation has increased precision significantly. In general, there will be several ways to analyze data from a given experiment, all of which offer comparable results.

## Equilibrium Experiment Analyses

Sedimentation equilibrium experiments may be used to determine the molecular weight of a solute. For molecules that self-associate or bind, the stoichiometry of the complex and a measure of the kinetics of association may also be determined.

Molecular weights are determined by solving the equilibrium form of the Lamm equation, generally by nonlinear least-squares fitting of the data to a presumptive model. For a monomeric solute (single ideal species), eqn [6] is a typical model. For associating systems, eqn [6] is extended by terms for each species present in the model. A monomer- $n$ -mer system is described by the model

$$c_{\text{total}} = c_{\text{k,monomer}} \exp \left[ \frac{w_2}{2RT} M(1 - \bar{v}\rho) (r^2 - r_k^2) \right] + c_{\text{k,nmer}} \exp \left[ \frac{w_2}{2RT} n M(1 - \bar{v}\rho) (r^2 - r_k^2) \right] \quad [8]$$

Stoichiometry can be determined from the mass of the largest species in a model that best fits the data. To study kinetics, an association constant,  $k_{\text{assoc}}$ , is defined such that

$$k_{\text{assoc}} = c_{n\text{-mer}} / (c_{\text{monomer}})^n \quad [9]$$

where  $c_{n\text{-mer}}$  is the concentration of a specific oligomer and  $c_{\text{monomer}}$  that of the monomer. Since all species are in thermodynamic equilibrium throughout the range of concentrations in the gradient,  $k_{\text{assoc}}$  may also be derived by solving the extended Lamm equation. Numerous software packages are available to perform this function. Foremost among these is a fitting algorithm called Nonlin, developed by Yphantis and colleagues at the University of Connecticut.

## Velocity Experiment Analyses

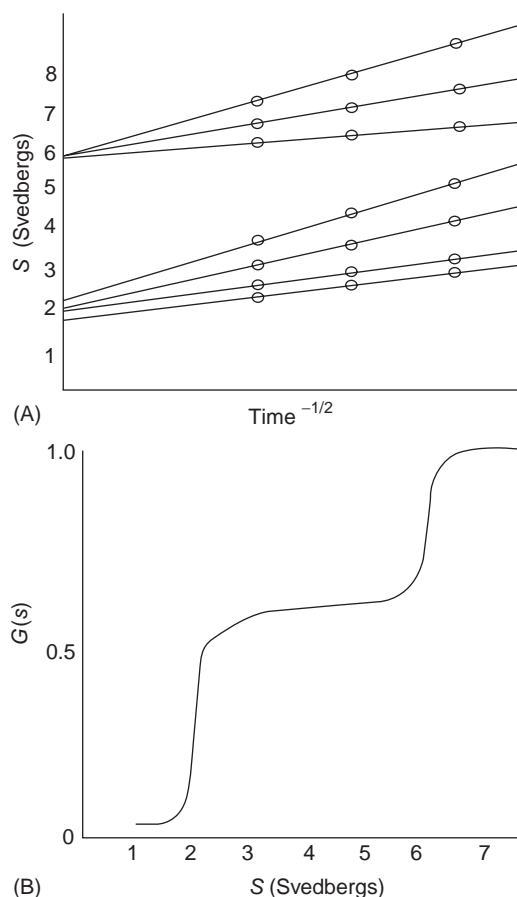
The primary goal of sedimentation velocity experiments is the determination of the sedimentation coefficient, which provides information about the mass and shape of sedimenting species. To this end, the analysis of velocity data seeks to determine (1) the rate of movement of the boundary and (2) whether the shape of the boundary indicates the presence of multiple species. Sample heterogeneity, or the presence of multiple species, will result in a broadening

of the boundary. A similar effect is produced by diffusion, so that it is necessary to distinguish the two.

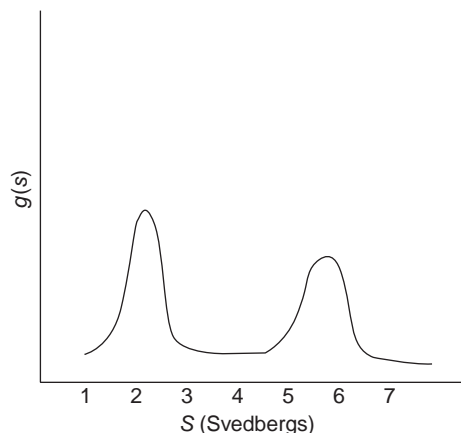
Classical methods, such as the transport and second moment methods, provided only average sedimentation coefficients. Modern computer methods are more capable of distinguishing individual sample components, and generally return a sedimentation coefficient distribution that describes the range of particle sizes in the sample. Two forms of this distribution are common. In an integral sedimentation profile,  $G(s)$ , the percentage of material in the sample with sedimentation coefficient equal to or less than a particular value of  $s$  is plotted against  $s$ . In the absence of significant sample nonideality, these plots increase systematically from left to right, to a value of 100%. Individual species with specific sedimentation values will produce vertical limbs in these plots (Figure 4B). The differential sedimentation profile,  $g(s)$ , is simply the derivative of the integral distribution. Individual species appear as peaks in these plots, which are similar in appearance to chromatograms (see Figure 5).

**Extrapolation** A powerful approach making use of the integral distribution is the method of Van Holde and Weischet, a graphical method that evaluates the boundary to determine if all portions of it are moving at the same rate. The presence of multiple species will cause the upper portion of the boundary to appear to move more rapidly until, ideally, a step is eventually formed. This method takes advantage of the fact that diffusion is a much slower process than sedimentation by extrapolating the data to a time point at infinity in order to eliminate the effects of the former. The boundary is subdivided into a number of vertical segments and a sedimentation coefficient corresponding to the radial position of each segment is calculated. This process is repeated for boundaries observed at different times during the run, and the resulting  $s$  values are plotted against the inverse square root of the time. A straight line is fitted to the data for each segment and extrapolated to the axis, which represents infinite time (Figure 4A). A sample containing a single sedimenting species will result in a single intercept, whereas multiple species result in multiple intercepts. The data may be presented in the form of the extrapolation plot or else transformed into a plot of  $G(s)$  versus  $s$ .

**Time-derivative methods** Another approach to analyzing the sedimentation coefficient distribution is to generate the differential  $g(s)$  profile from the time derivative of the concentration profile,  $dc/dt$ . The derivative is approximated as the difference between paired scans acquired very rapidly. In practice,



**Figure 4** Van Holde–Weischet analysis for a sample mixture of two species. (A) Extrapolation plot. Each vertical array of points corresponds to a boundary obtained at the same time during the run, and straight lines are fitted to the points derived from similar portions of the boundary at different times. Extrapolation to the axis, or infinite time, eliminates the effects of diffusion. (B) Integral sedimentation coefficient, or  $G(s)$ , distribution representing the same data.



**Figure 5** Time derivative analysis for a sample mixture of two species. The rate of change of the concentration profile is approximated from the difference between scans, and transformed into a plot showing the distribution of sedimentation coefficients in the sample.

a series of scans is taken over a short interval and subtracted in pairs. The differences are then averaged to approximate  $dc/dt$ . Since faster-moving particles move a greater distance than slower ones during the finite time interval studied, a final transformation is required to convert  $dc/dt$  into  $g(s)$  (Figure 5). An advantage of this technique is that time invariant experimental noise, such as window scratches, is eliminated by the subtraction.

For molecules of similar shape, this approach may be further refined by correcting for diffusional spreading. A differential sedimentation coefficient distribution that incorporates a relationship between  $s$  and  $D$ , termed  $c(s)$ , is generated by a complex calculation. Exceptional sensitivity can be obtained in this manner.

**Direct boundary fitting** Sedimentation velocity data may also be analyzed by least-squares fitting to models based on the Lamm equation. Software packages exist which provide approximate analytical solutions to the equation and which run on desktop computers. Other software packages offer more precise finite element solutions to the equation but require substantial computational power.

### Websites for Major Analysis Software

Some websites that are useful resources for the prominent analysis software are:

- Sedit (P Schuck), includes  $c(s)$  method: <http://analyticalultracentrifugation.com>
- Beckman Coulter – Origin, a general package: <http://www.beckmancoulter.com>
- Ultrascan (B Demeler), a general package: <http://www.ultrascan.uthscscsa.edu>

- Svedberg (J Philo), velocity methods: <http://www.jphilo.mailway.com/svedberg.htm>
- SEDNTERP (J Philo, T Laue), calculation of hydrodynamic parameters: <http://www.rasmb.bbri.org>

See also: **Centrifugation:** Preparative.

### Further Reading

- Hansen JC, Lebowitz J, and Demeler B (1994) Analytical ultracentrifugation of complex macromolecular systems. *Biochemistry* 33: 13155–13163.
- James LC and Jeffery CH (1999) Analytical ultracentrifugation as a contemporary biomolecular research tool. *Journal of Biomolecular Techniques* 10(4): 163–176.
- Johnson ML, Correia JJ, Yphantis DA, and Halvorson HR (1981) Analysis of data from the analytical ultracentrifuge by nonlinear least squares techniques. *Biophysical Journal* 36: 575–588.
- Lebowitz J, Lewis MS, and Schuck P (2002) Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Science* 11: 2067–2079.
- Schuck P and Demeler B (1999) Direct sedimentation boundary analysis of interference optical data in analytical ultracentrifugation. *Biophysical Journal* 76: 2288–2296.
- Schuck P (2000) Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophysical Journal* 78: 1606–1619.
- Stafford WF III (1992) Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. *Analytical Biochemistry* 203: 295–301.
- Van Holde KE and Weischet WO (1978) Boundary analysis of sedimentation velocity experiments with monodisperse and paucidisperse solutes. *Biopolymers* 17: 1387–1403.

## Preparative

**D N Taulbee**, University of Kentucky, Lexington, KY, USA

**A Furst**, Beckman-Coulter, Inc., Palo Alto, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Centrifugation is a mechanical process that utilizes a spinning medium to separate one or more components of a sample according to density or size. While gaseous or immiscible liquids can be separated by

centrifugation, the majority of applications involve sedimentation of solid particles in a liquid medium. Centrifugal separations may be classified as either analytical or preparative. In ‘analytical centrifugation’, the objective is to monitor particle sedimentation behavior in order to characterize particle properties, e.g., molecular weight, shape, and association. In ‘preparative centrifugation’, the objective is to separate and recover one or more components from a sample mix. Preparative centrifugation encompasses the vast majority of centrifugal applications.

a series of scans is taken over a short interval and subtracted in pairs. The differences are then averaged to approximate  $dc/dt$ . Since faster-moving particles move a greater distance than slower ones during the finite time interval studied, a final transformation is required to convert  $dc/dt$  into  $g(s)$  (Figure 5). An advantage of this technique is that time invariant experimental noise, such as window scratches, is eliminated by the subtraction.

For molecules of similar shape, this approach may be further refined by correcting for diffusional spreading. A differential sedimentation coefficient distribution that incorporates a relationship between  $s$  and  $D$ , termed  $c(s)$ , is generated by a complex calculation. Exceptional sensitivity can be obtained in this manner.

**Direct boundary fitting** Sedimentation velocity data may also be analyzed by least-squares fitting to models based on the Lamm equation. Software packages exist which provide approximate analytical solutions to the equation and which run on desktop computers. Other software packages offer more precise finite element solutions to the equation but require substantial computational power.

### Websites for Major Analysis Software

Some websites that are useful resources for the prominent analysis software are:

- Sedit (P Schuck), includes  $c(s)$  method: <http://analyticalultracentrifugation.com>
- Beckman Coulter – Origin, a general package: <http://www.beckmancoulter.com>
- Ultrascan (B Demeler), a general package: <http://www.ultrascan.uthscscsa.edu>

- Svedberg (J Philo), velocity methods: <http://www.jphilo.mailway.com/svedberg.htm>
- SEDNTERP (J Philo, T Laue), calculation of hydrodynamic parameters: <http://www.rasmb.bbri.org>

See also: **Centrifugation:** Preparative.

### Further Reading

- Hansen JC, Lebowitz J, and Demeler B (1994) Analytical ultracentrifugation of complex macromolecular systems. *Biochemistry* 33: 13155–13163.
- James LC and Jeffery CH (1999) Analytical ultracentrifugation as a contemporary biomolecular research tool. *Journal of Biomolecular Techniques* 10(4): 163–176.
- Johnson ML, Correia JJ, Yphantis DA, and Halvorson HR (1981) Analysis of data from the analytical ultracentrifuge by nonlinear least squares techniques. *Biophysical Journal* 36: 575–588.
- Lebowitz J, Lewis MS, and Schuck P (2002) Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Science* 11: 2067–2079.
- Schuck P and Demeler B (1999) Direct sedimentation boundary analysis of interference optical data in analytical ultracentrifugation. *Biophysical Journal* 76: 2288–2296.
- Schuck P (2000) Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophysical Journal* 78: 1606–1619.
- Stafford WF III (1992) Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. *Analytical Biochemistry* 203: 295–301.
- Van Holde KE and Weischet WO (1978) Boundary analysis of sedimentation velocity experiments with monodisperse and paucidisperse solutes. *Biopolymers* 17: 1387–1403.

## Preparative

**D N Taulbee**, University of Kentucky, Lexington, KY, USA

**A Furst**, Beckman-Coulter, Inc., Palo Alto, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Centrifugation is a mechanical process that utilizes a spinning medium to separate one or more components of a sample according to density or size. While gaseous or immiscible liquids can be separated by

centrifugation, the majority of applications involve sedimentation of solid particles in a liquid medium. Centrifugal separations may be classified as either analytical or preparative. In ‘analytical centrifugation’, the objective is to monitor particle sedimentation behavior in order to characterize particle properties, e.g., molecular weight, shape, and association. In ‘preparative centrifugation’, the objective is to separate and recover one or more components from a sample mix. Preparative centrifugation encompasses the vast majority of centrifugal applications.

While centrifugal techniques have been in use for at least a thousand years, the first recorded scientific study using a centrifuge did not appear until 1806, when Thomas Knight reported differences in the orientation of roots and stems of seedlings when placed in a rotating wheel. Some of the more significant developments since that time have included: the first commercial centrifuge, a hand-cranked cream separator, introduced in 1878 by the Swedish inventor, DeLaval; introduction of the analytical centrifuge for viewing particle sedimentation in 1923 by Theodor Svedberg; the isolation of subcellular components by centrifugal techniques in the 1940s; and the first use of a density-gradient medium by Edward Pickels in 1943. More recent advances have been characterized by significant improvements in materials and equipment and a broadening of applications.

Today, centrifuges are routinely used in a variety of disciplines including the medical, pharmaceutical, mineral, chemical, dairy, food, and agricultural industries. Available centrifuge designs and configurations seem almost as numerous as the applications themselves. An in-depth description of centrifuge designs and applications is beyond the scope of this treatise. Instead, this article will present the reader with an introduction to the theory of centrifugation, an overview of the various types of preparative centrifugal separations, and a description of some of the more common centrifuge and rotor designs along with their more common applications.

## Theory

### Sedimentation by Gravity

A particle suspended in a liquid medium of lesser density tends to sediment downward due to the 'gravity force',  $F_g$ :

$$F_g = mg = m \times 980 \text{ cm s}^{-2} \quad [1]$$

where  $m$  is the mass of the object and  $g$  is the gravitational-force constant.

For an object settling in a liquid or gaseous medium, there are two forces which oppose the gravitational force, the 'buoyancy force',  $F_b$  and the 'frictional force',  $F_f$ .

**Buoyancy force ( $F_b$ )** Archimedes showed that a particle suspended in a fluid experiences an upward force equivalent to the weight of the fluid being displaced:

$$F_b = m_M g = V_p \Delta_M g \quad [2]$$

where  $m_M$  is the mass of the fluid medium,  $V_p$  the volume of the particle (= volume of the displaced fluid), and  $\Delta_M$  the density of the displaced fluid.

**Frictional force ( $F_f$ )** In addition to the buoyancy force, the movement of a particle through a gaseous or fluid medium is hindered by the viscosity of the medium,  $\eta$  as described for a spherical particle by the Stokes' equation:

$$F_f = 6\pi\eta r_p (dx/dt) \quad [3]$$

where  $\eta$  is the medium viscosity in poise,  $P$  ( $\text{g cm}^{-1} \text{s}^{-1}$ ),  $r_p$  the radius of a spherical particle (cm), and  $(dx/dt)$  the velocity of the moving particle.

The frictional force increases as a function of particle velocity, eventually combining with the buoyancy force to precisely oppose the gravitational force. This condition is known as the limiting or 'terminal velocity'.

**Diffusion** Diffusion stems from random Brownian motion and results in the net movement of solute or suspended particles from regions of higher to regions of lower concentration. Thus, diffusion works in opposition to sedimentation which tends to concentrate particles.

While the precise impact of diffusion can be difficult to calculate for complex systems, it often suffices to know that the rate of diffusion is generally more pronounced for smaller particles, it increases with temperature, and its effects are lessened by higher centrifugal-force fields.

### Sedimentation in a Centrifugal Field

A particle moving in a circular path continuously experiences a 'centrifugal force',  $F_c$ , which may be expressed as

$$F_c = ma = m\omega^2 r \quad [4]$$

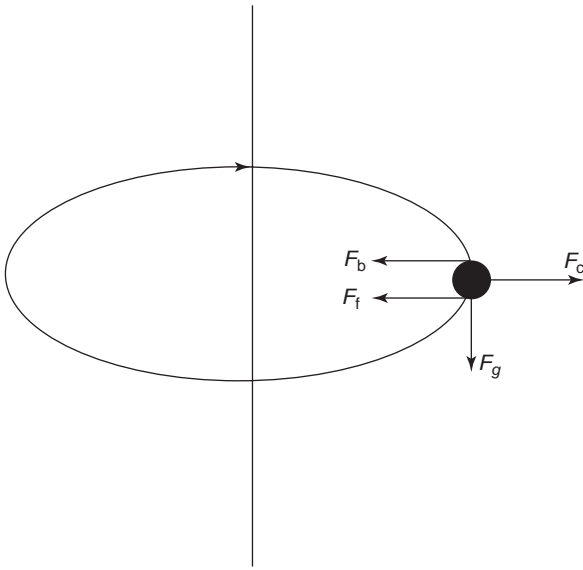
where  $m$  is the particle mass (g),  $a$  the acceleration ( $\text{cm s}^{-2}$ ),  $\omega$  the angular velocity ( $\text{radians s}^{-1} = 2\pi \text{ rpm}/60$ ), and  $r$  the radial distance from the axis of rotation to the particle (cm).

The centrifugal force acts to move particles away from the axis of rotation, while the buoyancy and frictional forces oppose this movement (**Figure 1**).

Centrifugal force can be related to the gravitational force,  $g$ , by the 'relative centrifugal force', (RCF), more commonly referred to as the  $g$  force:

$$\begin{aligned} \text{RCF} &= F_c/F_g = (m\omega^2 r)/(mg) = (\omega^2 r)/g \\ &= 1.19 \times 10^{-5} (\text{rpm})^2 r \end{aligned} \quad [5]$$





**Figure 1** Forces acting on a particle in a centrifugal field ( $F_b$ , buoyancy;  $F_f$ , frictional;  $F_c$ , centrifugal; and  $F_g$ , gravitational).

While RCF is unitless, it is frequently expressed in units of  $g$  to indicate the number of times that the force of the applied centrifugal field is greater than the force of gravity.

Analogous to gravity settling, a particle will attain a limiting or terminal velocity when the sum of the buoyancy and frictional forces equals the centrifugal force:

$$F_c = F_b + F_f \quad [6]$$

Assuming a spherical particle ( $4/3\pi r^3$ ), particle velocity can be calculated as

$$v = (d^2(\Delta_P - \Delta_M)\omega^2 r) / 18\eta \quad [7]$$

Equation [7] can be integrated to determine the time required for a particle to traverse a given radial distance from  $r_0$  to  $r_1$ :

$$t = [18\eta / (d^2(\Delta_P - \Delta_M)\omega^2)] \ln(r_1/r_0) \quad [8]$$

where  $r_0$  is the initial position of the particle and  $r_1$  the final position of the particle.

Equations [7] and [8] describe the impact of the more significant and controllable parameters that govern particle movement in a centrifuge. They reveal that, for a given particle density, the rate of sedimentation:

- Increases as a square function of rotor speed, e.g., doubling the speed will lessen the run time by a factor of four.

- Increases proportionally with distance from the axis of rotation.
- Is inversely related to the viscosity of the carrier medium.

### Sedimentation Coefficient

As shown, the sedimentation velocity,  $v$ , is proportional to  $\omega^2 r$ . This proportionality can be expressed in terms of the sedimentation coefficient,  $s$ , which is a measure of the sedimentation velocity per unit of centrifugal force. For a spherical particle,

$$s = (dx/dt) / (\omega^2 r) = 2r_p^2(\Delta_P - \Delta_M) / 9\eta \quad [9]$$

The sedimentation coefficient has the dimensions of time and is expressed in Svedberg units,  $S$ , equal to  $10^{-13}s$ . Its value is dependent on the mass and shape of the particle and on the sedimentation medium.

### Rotor Efficiency

The time required for a particle to traverse a rotor may be determined from the clearing- or  $k$ -factor, also known as the pelleting efficiency, with lower  $k$ -factors equating to shorter pelleting times. The  $k$ -factor is calculated at the maximum rated rotor speed and is a constant for a given rotor.  $k$ -Factors provide a convenient means of determining the minimum time required to pellet a particle and are thus, useful for comparing different rotors:

$$\begin{aligned} k &= \ln(r_{\max} - r_{\min}) \times 10^{13} / (3600\omega^2) \\ &= 2.53 \times 10^{11} \times \ln(r_{\max} - r_{\min}) / (\text{rpm})^2 \end{aligned} \quad [10]$$

where  $r_{\max}$  and  $r_{\min}$  are the maximum and minimum distances from the centrifugal axis, respectively.

When  $k$  is known (normally provided by manufacturer), the minimum run time required for particle pelleting may be calculated from the relation

$$t = k/s \quad [11]$$

where  $t$  is the time in hours required for pelleting and  $s$  is the sedimentation coefficient in Svedberg units.

### Deviation from Ideal Behavior

There are other, more difficult to characterize effects that lead to significant deviations from the settling velocity predicted by eqn [7]. The most common such effect occurs when the particles are nonspherical. Another example is for biological particles which interact with the medium via hydration, the extreme case being for osmotic particles which can result in drastic changes in particle density. Interparticle forces, e.g., charge or hydrophobic effects, may also impact sedimentation behavior.

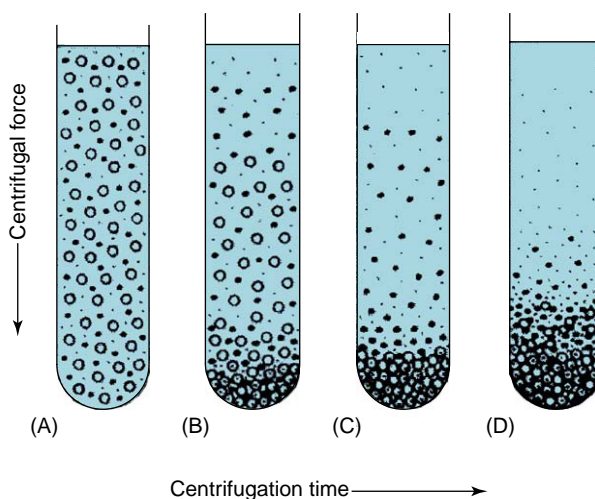
## Types of Separations

Preparative centrifugal separations are often classified according to the phases of the media and the material to be purified, e.g., gas–gas, liquid–liquid, or liquid–solid. Gas-phase separations are very important in certain applications, e.g., uranium-isotope enrichment, but are highly specialized and not widely used. Liquid–liquid or even liquid–liquid–solid separations, on the other hand, are much more common. However, the majority of preparative separations involve the sedimentation of solid particles in a liquid medium.

Another way of classifying preparative separations is according to the method by which purified components are recovered. Three modes are used: (1) batch-mode, in which a sample mix is loaded, processed, and then recovered at the conclusion of the run by decanting the supernatant and scraping the pellet from the rotor wall; (2) semibatch mode, in which the sample mix is continuously fed to a spinning rotor as the supernatant is continuously discharged and the pellet is permitted to accumulate for postrun removal; and (3) continuous mode, in which the sample mixture is fed continuously, the supernatant is continuously discharged, and denser liquid or solid materials are either intermittently or continuously discharged during the run.

### Differential Sedimentation

As discussed, larger and/or denser particles will sediment more rapidly in a centrifugal-force field and thus, pellet onto the rotor wall faster than smaller or lighter particles. Most applications are based on this difference in behavior, referred to as differential sedimentation or ‘pelleting’. In a simple batch-mode pelleting separation, a sample mixture termed the homogenate (immiscible liquids or solid suspensions), is separated into two fractions as depicted in **Figure 2**. The unsedimented material is termed the ‘supernatant’ and the sedimented material is the ‘pellet’. This approach works well when the objective is to pellet all the solid particles or to clarify the liquid. However, obtaining high-purity separations by this approach can be difficult since the centrifugal field required to pellet larger or denser particles that are initially nearer the axis of rotation is capable of pelleting smaller or lighter particles that are initially closer to the outer wall (**Figure 2**). A more efficient one-step approach for isolating smaller or lighter particles is to layer the sample mixture on top of a preloaded cushion of dense medium then stopping the run before the lighter or smaller particles reach the rotor wall.



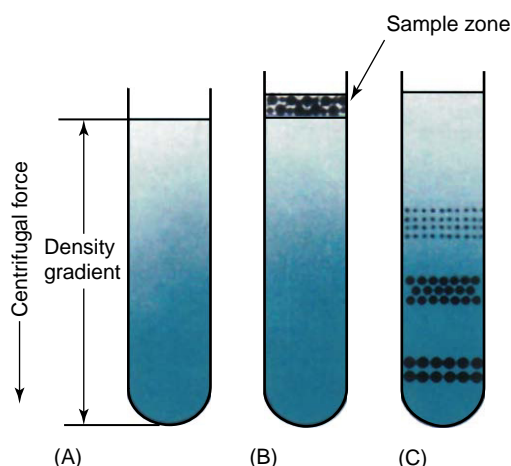
**Figure 2** Differential sedimentation or pelleting. (Courtesy of Beckman Coulter, Inc.)

### Density Gradient Centrifugation

A density gradient is a liquid medium that increases in density from the layers nearest the axis of rotation to those farthest away. This is achieved through variation in the concentration of an aqueous solute, or other gradient material, across the rotor. Density gradient centrifugation (DGC) compensates for some of the problems with separation efficiency encountered when using a homogeneous medium while permitting the simultaneous separation of multiple components from a single sample. DGC may be conducted as either rate-zonal or isopycnic separations.

**Rate-zonal separations** It can be used to separate particles of similar density according to size or to separate particles of different density and size as a function of their sedimentation coefficient,  $s$ . In its simplest form, the sample mixture is layered in a narrow band on top of a preloaded, homogeneous medium as shown in **Figure 3**.

The use of a density gradient instead of a homogeneous medium offers several advantages for rate-zonal separations. The steep gradient beneath the layered sample suppresses premature sedimentation as well as convection currents in the liquid column. Additionally, the continuous increase in density, and often accompanying increase in viscosity across the rotor, serves to slow the faster moving particles and provide better resolution in the sample component ‘bands’. Increasing-viscosity gradients also lessen diffusional effects, though this advantage may be offset by an increase in run time. Rate-zonal separations are well suited for mixtures of



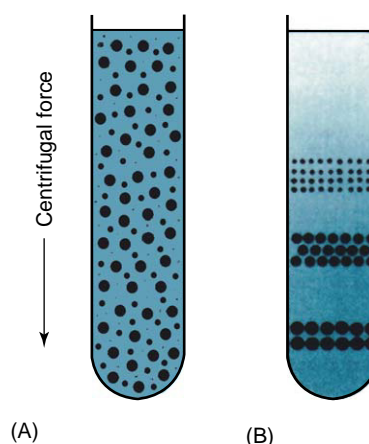
**Figure 3** Rate-zonal separations in a swinging-bucket rotor. (A) Centrifuge tube filled with density gradient solution, (B) sample applied to top of gradient, and (C) under centrifugal force, particles move at different rates depending upon their mass. (Courtesy of Beckman Coulter, Inc.)

similar-density particles that exhibit two or more well-defined modes of size distribution.

**Isopycnic separations** They rely on differences in particle densities and are conducted in a density gradient. The density range of the gradient often spans the full range of particle densities so that particles never reach the rotor wall regardless of run time. Instead, particles sediment until they reach a position in which the density of the medium is equivalent to their own (Figure 4). Differences in particle size only affect their rate of movement, though this may ultimately dictate the required run time. When the range of particle densities exceeds the range of the density gradient, then both pelleting and isopycnic separation will occur as some particles fully traverse the rotor and pellet while others attain their isopycnic position and remain suspended.

**Gradient materials** The selection of an appropriate gradient material is an important consideration. As summarized by Griffith and Ridge, the ideal gradient material should:

- Span a density range sufficient to permit separation of the particles of interest.
- Be stable in solution.
- Be inert toward the fractionated materials.
- Exert the minimum osmotic effect, ionic strength, and pH.
- Be removable from the product.
- Be either inexpensive or easily recycled.
- Be sterilizable.



**Figure 4** Isopycnic separation with a self-generating gradient. (A) Uniform mixture of sample and gradient and (B) under centrifugal force, gradient redistributes and sample particles band at their isopycnic positions. (Courtesy of Beckman Coulter, Inc.)

On the other hand, the selected media should not generate a prohibitively high viscosity, interfere with the assay technique, be corrosive, or generate flammable or toxic aerosols.

Unfortunately, no single ideal gradient material exists, as each separation imposes its own set of requirements. The list of materials that have been used for gradient formation is extensive with examples of some commonly used materials listed in Table 1.

With respect to biological inertness and low viscosity, the most ideal aqueous gradient material is deuterium oxide ( $D_2O$ ). However,  $D_2O$  is expensive and has a relatively low maximum density ( $1.11 \text{ g cm}^{-3}$ ). Sucrose is a commonly used material due to its price, transparency, availability, and nontoxic nature. Densities to  $1.33 \text{ g cm}^{-3}$  can be achieved which is sufficient for separating most cells and intracellular organelles. More recently, Iodixanol (Optiprep<sup>TM</sup>) has gained in popularity. It has a density range similar to sucrose but a significantly lower viscosity and is iso-osmotic.

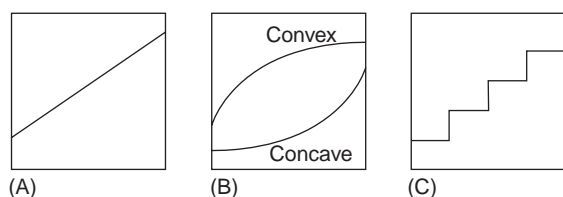
Salts are used to generate higher-density aqueous solutions. Cesium chloride ( $CsCl$ ) solutions, the most widely used of this class, can reach densities of  $\sim 1.9 \text{ g cm}^{-3}$  at saturation while providing low viscosity at lower concentrations.  $CsCl$  can be readily recovered and purified, but tends to be highly osmotic, corrosive, and expensive. The practitioner must be cautious not to overstress the rotor when using high-density salt solutions or when using salts that may precipitate, leading to catastrophic rotor failure.

For nonaqueous gradients, organic liquids such as toluene, methanol, or kerosene may be blended to

**Table 1** Physical properties of gradient materials in aqueous solutions at 20°C

	Maximum solution			20% w/w solution	
	Concentration (% w/w)	Density ( $\text{g cm}^{-3}$ )	Viscosity (cP)	Density ( $\text{g cm}^{-3}$ )	Viscosity (cP)
Sucrose	65	1.33	182	1.08	2
Ficoll	43	1.17	600	1.07	27
Ludox-SM	—	1.40	—	1.13	2
Percoll	23	1.13	10	1.11	8
Metrizamide	56	1.44	58	1.12	2
Cesium chloride	65	1.91	1.3	1.17	0.9
Lithium silicotungstate	85	2.89	14	1.20	—
Sodium polytungstate	85	2.89	26	1.20	2

Reproduced with permission from sheeler. *Centrifugation in Biology and Medical Science*.

**Figure 5** Gradient shapes: (A) linear, (B) exponential, and (C) isokinetic.

attain gradient densities lower than  $1.0 \text{ g cm}^{-3}$ . Of these, methanol presents an additional advantage of being water soluble, thereby allowing gradients to be formed from a combination of the two. Halogenated liquids such as diiodomethane, bromoform, and tetrabromoethane can be used to prepare very dense solutions ( $>2.8 \text{ g cm}^{-3}$ ). Problems associated with flammability, toxicity, and deterioration of transfer lines and seals must be considered when using these materials.

**Gradient shape and formation** Gradient shape refers to the density profile across the tube or rotor as a function of gradient volume (Figure 5). Its choice is important as it governs the sedimentation rate as well as the terminal position in isopycnic runs.

Gradients may be classified as 'step' or 'continuous'. Step (discontinuous) gradients, are prepared by the step-wise addition of solutions of successively higher density to the rotor or tube. Step gradients have the advantages that they may be formed without the need for a gradient generator and may be easily 'tailored' to provide larger volumes of medium in the ranges that correspond to the density profile of the particles to be separated. For continuous gradients, the medium density varies in a continuous manner across the rotor or tube and may be further classified as linear, exponential, or isokinetic (Figures 5A–5C). 'Isokinetic' gradients are designed to produce a uniform sedimentation velocity throughout the

gradient by counterbalancing the increase in centrifugal force the particles experience as they traverse the gradient with an increase in medium density and viscosity. Simple linear sucrose gradients provide a near isokinetic gradient.

Numerous methods are used to form gradients. The simplest approach is to form the gradient *in situ*, i.e., self-generating, at high centrifugal speeds. However, higher rotor speeds and longer run times are often required for self-generating gradients and not all media or centrifugal equipment are amenable to this approach. Step gradients are also easily formed by simply pumping targeted volumes of successively denser solutions to the rotor. Inexpensive peristaltic pumps provide a convenient means to form step gradients. Gradient pumps are available to generate the targeted gradient curve shape including both mechanical and programmable electronic pumps that mix variable amounts of low- and high-density solutions.

Several approaches are used to analyze and/or fractionate the rotor effluent. The simplest is to split the gradient into fractions according to volume. Alternatively, the effluent may be routed through one or more in-line flow cells to monitor a selected gradient property, e.g., density, absorbance, refractive index, fluorescence, etc. Automated fractionators that switch collection vessels according to effluent volume or to feedback from an in-line detector are available.

### Continuous Centrifugation

In continuous separations, a sample mixture is continuously introduced to a spinning rotor as the clarified supernatant continuously exits. Continuous-feed centrifuges may be used for rate, pelleting, filtration, or isopycnic separations. They are best suited for applications in which large volumes and/or low-concentration samples must be processed, the particle sedimentation coefficient is high  $\geq 50 \text{ s}$ , or long acceleration/deceleration times are required.

In continuous centrifugation, the centrifugal force and flow rate must be controlled to provide sufficient time for solid or denser liquids to sediment before being carried out with the supernatant but not so long as to underutilize the rotor-throughput capacity. With information on liquid volume within the rotor and assuming laminar flow, the maximum flow rate can be determined from eqn [8]. Alternatively, if the rotor  $k$ -factor and the particle sedimentation coefficient are known, the minimum residence time required for pelleting can be calculated from eqn [11].

### Ultracentrifuges

Ultracentrifugation is an ill-defined term originally applied to analytical centrifuges and subsequently to units with rated speeds greater than  $\sim 25\,000$  rpm, regardless of the media or rotor design. Some manufacturers now reserve this term for centrifuges that operate at sufficient speeds to require a significant vacuum to reduce frictional drag and/or rotor heating.

### Filtration

Filtration is a mechanical means of separating solids from a liquid suspension via a porous medium or screen that permits the liquid to pass while retaining the solids. Centrifugal filtration is driven by the pressure exerted by a liquid medium and is opposed by the combined resistance of the porous filter and filter cake.

Centrifugal filtration is a complex process that is dependent on a number of parameters including liquid viscosity, cake thickness, centrifugal force, screen area, and importantly, the size and packing characteristics of the particles themselves. This technique is generally not amenable to broad generalizations and is, therefore, best approached on a case by case basis.

## Centrifugal Equipment

Centrifuges and rotors are commercially available in literally hundreds of shapes, sizes, and configurations. They range from small laboratory-scale units that can spin at speeds in excess of  $100\,000$  rpm delivering forces in excess of  $1\,000\,000g$ , to industrial-scale decanters that may continuously process up to  $300\,000\text{ l h}^{-1}$ . The primary rotor or centrifuge selection criteria must focus on the objective for conducting the separation. Parameters such as batch versus continuous; the required centrifugal force, purity; throughput; the number of components to be recovered; sample toxicity/corrosiveness; time; cost;

**Table 2** Strength data for commonly used rotor construction materials

Material	Density ( $\text{g cm}^{-3}$ )	Ultimate strength ( $\text{g cm}^{-3}$ )	Strength–density ratio
Aluminum	2.79	2159	774
Titanium	4.84	6088	1258
Steel	7.99	7915	991

available space; and noise tolerances should be considered.

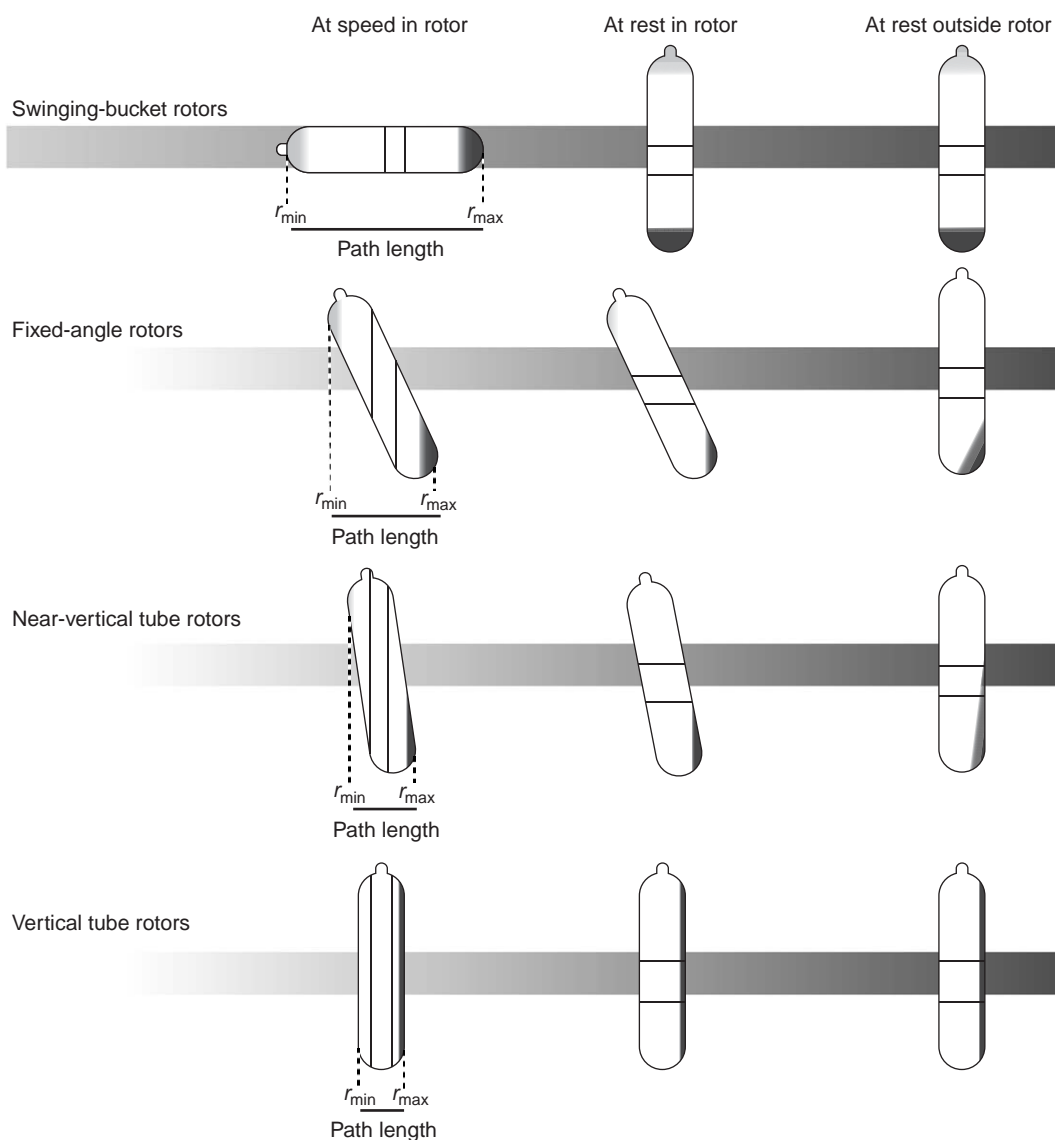
Early rotors were often manufactured of steel or brass, but now, they are more commonly constructed of aluminum and titanium (Table 2). Newer composite materials have made tremendous gains in popularity, with plastics for small-scale applications, and stainless steel for industrial-scale units in common use. Though somewhat more expensive, titanium is particularly suitable as it has both a higher strength–density ratio and a high resistance to corrosion and erosion.

Centrifuge bottles and tubes are also constructed from a variety of materials including glass, stainless steel, and plastics, with polycarbonate being one of the more popular materials due to its transparency and strength. The choice of material is generally dictated by the required  $g$ -force and chemical compatibility with the sample and medium.

### Bottle Centrifuges

Bottle centrifuges consist of a motor-driven vertical spindle to which a horizontal rotor, machined with a number of sample positions is attached. Such units are normally equipped with a timer, tachometer, refrigeration system, and manual or automatic braking. Bottle centrifuges are usually smaller, bench-top units but are also available as larger, free-standing units.

Bottle-centrifuge rotors are classified as swinging bucket, fixed-angle, and vertical (Figure 6). In the ‘swinging-bucket’ design, the bottles are in a vertical position at rest but swing outward to a horizontal orientation as the rotor speed increases. The centrifugal force is applied along the length of the tube providing the best resolution for rate-zonal separations. ‘Fixed-angle’ rotors are loaded and operated with the tube remaining at a fixed angle both at rest and while spinning. This angle is typically  $20$ – $45^\circ$  from vertical, though ‘near-vertical’ rotors are less than  $10^\circ$  from vertical. The fixed-angle design provides a shorter path length with a corresponding reduction in run time. During the run, particles aggregate on the outer wall of the tube and slide down the tube wall to form a pellet in the bottom. Fixed-angle rotors provide the highest available force



**Figure 6** Particle separation in swinging bucket, fixed angle, and vertical tube rotors. (Courtesy of Beckman Coulter, Inc.)

fields. ‘Vertical’ rotors can be considered as an extension of fixed-angle rotors in which the angle of repose is  $0^\circ$  from vertical. In this design, the maximum path length is equal to tube diameter, thereby providing the lowest  $k$ -factors for a given tube. Vertical tube rotors are often used for isopycnic banding where short run times are desired. Density gradients will reorient during the acceleration and deceleration of a fixed-angle or vertical rotor, such that the gradient is vertical (aligned with gravity) while the rotor is at rest, and horizontal (aligned with the centrifugal field) while at speed.

### Zonal Rotors

Zonal rotors are bowls or cylindrical cavities equipped with a central core and attached vanes

or septa that divide the rotor into four or more sector-shaped compartments. Zonal rotors provide larger internal volumes for a given radius, minimal wall effects, and maximum particle and gradient resolution.

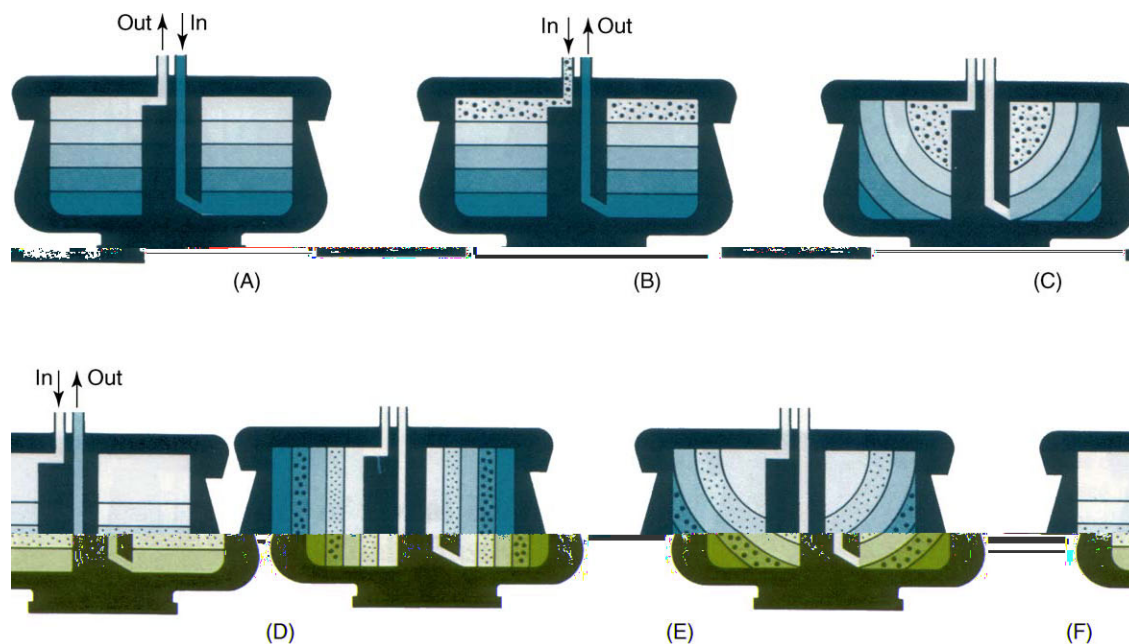
Zonal centrifuges can be operated in batch, semibatch, or continuous modes and may be loaded and unloaded with the rotor stopped (‘static or reorienting’) or spinning (‘dynamic or rotating seal’). In static loading, the gradient is loaded with the rotor at rest, the rotor slowly accelerated to permit the gradient to reorient from a horizontal to a vertical configuration, and then slowly decelerated back to rest for postrun unloading (Figure 7). The advantages of the static-loading technique are simplicity and the avoidance of rotating seals that may leak or fail during loading/unloading. The major disadvantage is



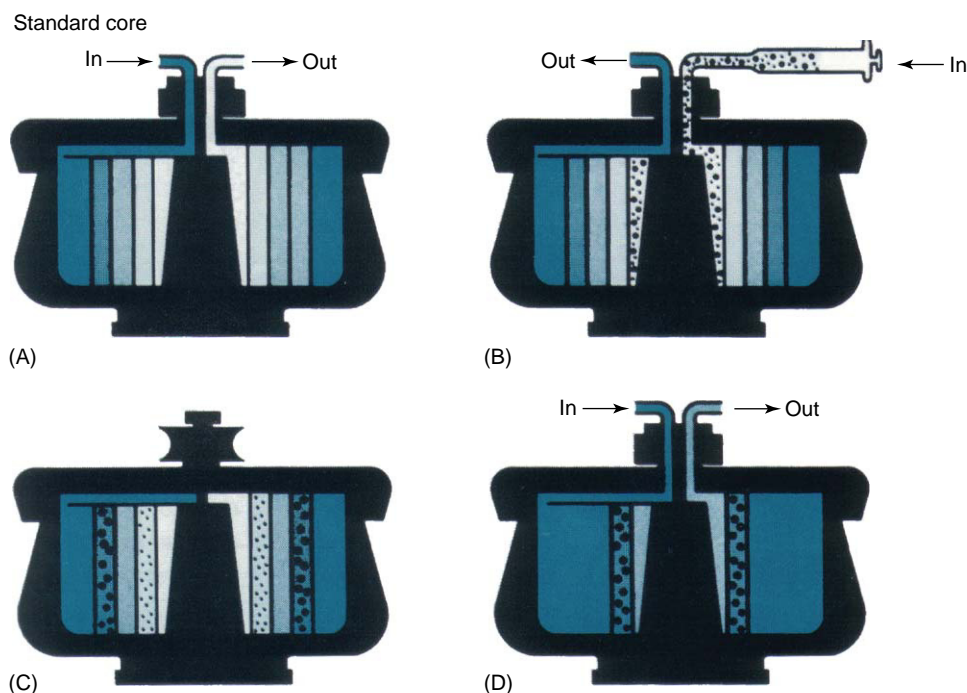
the tendency of the gradient to swirl as it reorients leading to a loss in resolution.

Dynamic loading and unloading is conducted as the rotor spins (Figure 8). The gradient is pumped

through a rotating seal in the center of the rotor lid to the outer wall of the rotor via passages machined into the rotor core. Lighter-density solutions are loaded first, forming a vertical layer which is



**Figure 7** Static loading and unloading of a zonal rotor with a reorienting gradient core. (A) Gradient loaded, light end first, with rotor at rest; (B) sample solution layered on top of gradient; (C) Rotor accelerated, layers reoriented under centrifugal force; (D) layers vertical, particles separated with rotor at speed; (E) rotor decelerated, layers reoriented; (F) static unloading, contents displaced with air pressure, heavy end first. (Courtesy of Beckman Coulter, Inc.)



**Figure 8** Dynamic loading and unloading of a zonal rotor. (A) Gradient loaded with rotor spinning at 2000 rpm, (B) sample injected at 2000 rpm, followed by injection of overlay, (C) particles separated with rotor at speed, and (D) contents unloaded by introducing dense solution at rotor edge, displacing fractions at center. (Courtesy of Beckman Coulter, Inc.)



displaced inward by the ensuing denser solutions. The rotor is unloaded by routing a high-density immiscible liquid to the outer wall forcing the gradient from the rotor, lighter fractions first (center unloading) or by pumping a light liquid to the center, displacing the heavier fractions first (edge unloading). While more cumbersome, dynamic loading generally provides better resolution than static loading/unloading.

### Continuous Centrifuges

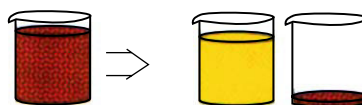
**Figure 9** shows the major industrial applications of continuous centrifuges.

**Disk centrifuges** Disk centrifuges operate on the principle of differential sedimentation and are used for two-phase (liquid–solid or liquid–liquid) and three-phase (liquid–liquid–solid) separations. Disk centrifuges are essentially a rotating bowl equipped with an internal set of conical settling plates which

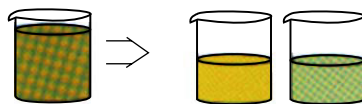
serve to decrease the sedimentation path length and increase the sedimentation surface area, i.e., capacity factor (**Figure 10**). Denser materials sediment onto and slide across the plate surfaces before accumulating on the bowl wall. These are highly efficient units with some industrial-scale units generating forces of 10 000g and capable of pelleting particles as small as 0.1  $\mu\text{m}$  diameter.

Three variations of disk centrifuges, distinguished by their solids-handling capability, are commonly used: solids-retaining, intermittent solids-ejecting, and continuous solids-ejecting (**Figure 10**). Solids-retaining designs (**Figure 10A**) are appropriate for liquid–solid or liquid–liquid separations where the solids content is less than  $\sim 1\%$  by volume. Common applications of solids-retaining disk centrifuges include separation of cream from milk, organic waste from water, purification of lubricating oils, or removal of water and solids from jet fuel.

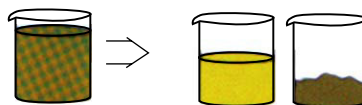
Clarification: Separate suspended particles from a liquid



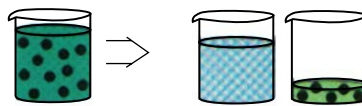
Purification: Separate immiscible liquids (even with solids present)



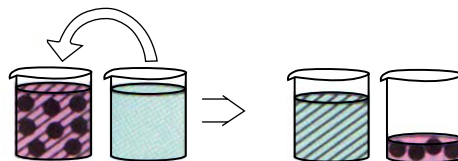
Dewatering: Concentrate a slurry



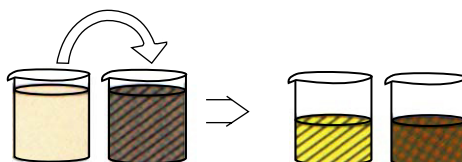
Classification: To split a suspension into two streams with different particle size distributions



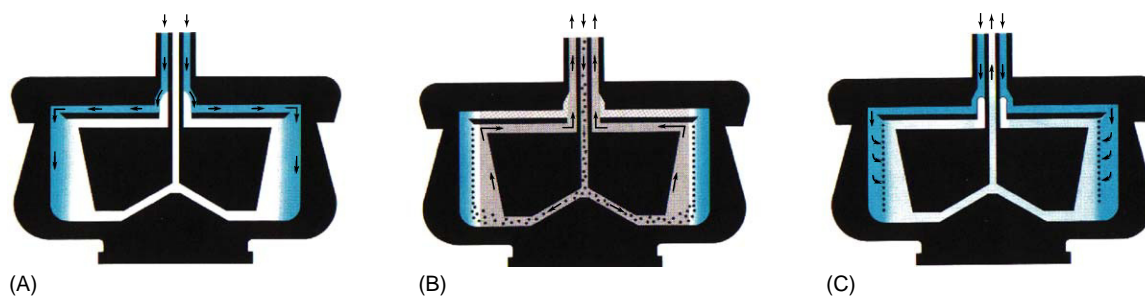
Washing: Countercurrent washing or dissolving of impurities in suspended, crystallized or amorphous solids



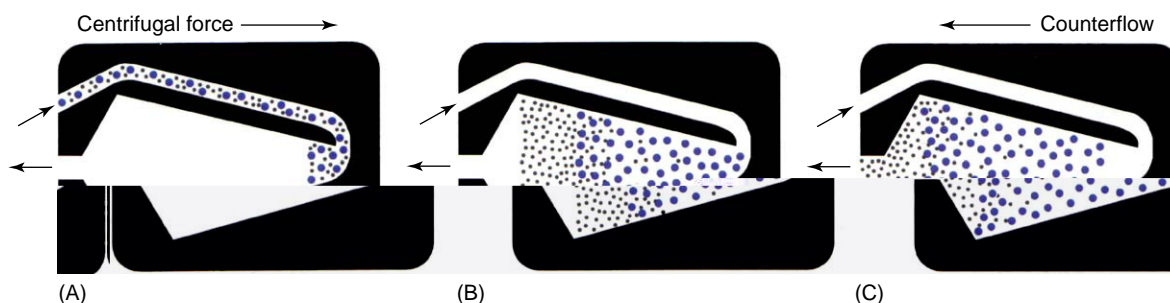
Extraction: To mix a liquid containing a mineral or extract with a liquid agent, and then separate out the agent, which contains the extract or mineral, from the original liquid



**Figure 9** Major industrial applications for continuous centrifuges. (Courtesy of Alfa Laval, Inc., used by permission.)



**Figure 12** Flow regimes in a continuous-flow zonal rotor. (A) Flow during rotor loading at 2000 rpm, (B) sample flow at operating speed, and (C) unloading after a banding or isopycnic experiment. (Courtesy of Beckman Coulter, Inc.)



**Figure 13** The elutriation process. (A) Sample suspended in medium enters chamber; (B) sedimentation tendency of particles balanced by counterflow; (C) flow increased, slow sedimenting particles elutriated from chamber. (Courtesy of Beckman Coulter, Inc.)

design to process feed streams with low solids content. One configuration, designed for recovery of two immiscible liquids and a solids product, is shown in **Figure 11**. Due to their high speed and short settling path, tubular centrifuges are well suited for the pelleting of ultrafine particles, liquid clarification, and difficult to separate immiscible liquids, particularly when the objective is efficient recovery of high-value products. Industrial models are available with throughput rates to  $250 \text{ m}^3 \text{ h}^{-1}$  and forces ranging to  $20\,000g$  while laboratory models are available with throughput rates to  $150 \text{ l h}^{-1}$  and centrifugal forces ranging to  $62\,000g$ . Typical applications include recovery of *Escherichia coli* cells and flu viruses, blood fractionation, and deinking.

**Continuous zonal rotors** These rotors are similar to those designed for batch separation but with a larger diameter core providing a different flow pattern as illustrated in **Figure 12**. These rotors are best suited for low-concentration, high-volume samples. Applications include purification of viruses from tissue-culture media, harvesting bacteria, or separating fine-clay particles from water.

**Elutriation rotors** Elutriation or counterstreaming is a batch technique used to separate particles with

differing sedimentation rates (rate separation). Rotors with conical- or funnel-shaped cavities are used with the small end positioned farthest from the axis of rotation (**Figure 13**). The sample mixture is introduced at a constant rate to the small end of a pre-filled, spinning cavity. As the particles migrate inward toward the larger-diameter end of the cavity, they experience a decreasing centrifugal force due to the shorter radius of rotation and a decreasing frictional drag due to the decrease in the velocity of the counterflowing medium. As these forces are balanced, the particles will classify according to size. Particles of increasing size may then be collected by increasing the flow rate or by decreasing the rotor speed. Elutriation rotors provide high cell viability by operating at low centrifugal fields and permitting the separation to be conducted in any compatible medium. A common application is the isolation of specific mammalian cell types.

### Centrifugal Filtration Equipment

Filtration centrifuges are used for the separation of solids from liquid slurries, chiefly in industrial applications, and are usually characterized in terms of the final moisture content, drainage time, and centrifugal force. They are most useful for high-volume

processing of fast draining solids. Three of the more common designs are batch/semibatch basket centrifuges, and continuous push-type, and continuous conical centrifuges. Applications for continuous centrifugal filters include dewatering of crystalline solids, extraction of solids from fruit and vegetable pulps, and dewatering of coal fines.

See also: **Centrifugation**: Analytical Ultracentrifugation.

## Further Reading

- Brakke MK (1952) Density gradient centrifugation: a new separation technique. *Journal of the American Chemical Society* 73: 1847–1848.
- Coulson JM, Richardson JF, Backhurst JR, and Harker JH (1978) *Chemical Engineering, Volume 2: Unit Operations*, 3rd edn., 807pp. Oxford: Pergamon.
- Griffith OM (1986) *Techniques of Preparative, Zonal, and Continuous Flow Ultracentrifugation*; DS-468H, p. 50. Palo Alto, CA: Spinco Division of Beckman Instruments, Inc.
- Hsu HW (1981) In: Perry ES (ed.) *Techniques of Chemistry, Vol. XVI, Separations by Centrifugal Phenomenon*, 466pp. New York: Wiley.
- Lavanchy AC and Keith EW (1979) Centrifugal separation. In: Grayson M and Eckroth D (eds.) *Encyclopedia of Chem. Techn.*, 3rd edn., vol. 5, pp. 194–233. New York: Wiley.
- Price CA (1982) *Centrifugation in Density Gradients*, 430pp. New York: Academic Press.
- Ridge D (1978) In: Birnie GD and Rickwood D (eds.) *Centrifugal Separation in Molecular and Cell Biology*, ch. 3. London: Butterworths.
- Sheeler P (1981) *Centrifugation in Biology and Medical Science*, 269pp. New York: Wiley.
- Taulbee DN and Maroto-Valer MM (2000) Centrifugation. In: Cooke M and Poole C (eds.) *Encyclopedia of Separation Science*, pp. 17–40. London: Academic Press.
- [www.beckmancoulter.com/resourcecenter/labresources/centrifuges/rotorcalc.asp](http://www.beckmancoulter.com/resourcecenter/labresources/centrifuges/rotorcalc.asp).
- <http://researchlink.labvelocity.com/tools/conversionTools/CentrifugationTool.jhtml>.
- [www.centrifuges.bz/page2.htm](http://www.centrifuges.bz/page2.htm).

# CERAMICS

**G J Oliver and R N White**, Ceram Research Limited, Stoke-on-Trent, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The term ‘ceramic’ covers a wide range of materials, which makes it an interesting area of activity for the inorganic analyst. The range of matrices is described with reference to problems in preparation for analysis. Various methods of analysis are described and compared with respect to cost, general applicability, sensitivity, and accuracy. As might be expected, this article is primarily about bulk analysis, but speciation, surface analysis, and microstructural analysis are becoming more critical these days.

## Types of Ceramic Material

Ceramic materials are produced by heat treatment of minerals or mixtures of minerals, and the matrices have much in common with geological samples. For this reason, methods established for the analysis

of ceramic samples may also be applied to rocks, soils, ores and minerals as well as refractories, white-ware, glass, enamels, and cement. Ceramic materials may be divided into four industrial types as follows:

1. *Whitewares* (tiles, tableware, sanitaryware). These are traditional fired and unfired bodies, glazes, raw materials such as china clay, talc, bone or feldspar, and intermediates such as frits, slips, glazes, and colors.
2. *Building materials*. Bricks, breezeblock, cement, glass, concrete, tiles, clays, limestone, and additives such as stains and barium carbonate.
3. *Refractories*. Aluminosilicates, silica and aluminous material, basic materials such as magnesite, chrome-bearing materials such as magnesia-chrome bricks, zirconia-bearing materials, and silicon carbide. There are also raw materials such as fireclay, ferrosilicon, and graphite.
4. *Industrial or engineering ceramics*. This class includes a wide range of finished products and raw materials for biological, military, electronic, automotive, aerospace, and high-temperature uses. Typical materials include alumina, zirconia, titanate, silicon carbide and nitride, SiALON, and mixtures of rare earths.

processing of fast draining solids. Three of the more common designs are batch/semibatch basket centrifuges, and continuous push-type, and continuous conical centrifuges. Applications for continuous centrifugal filters include dewatering of crystalline solids, extraction of solids from fruit and vegetable pulps, and dewatering of coal fines.

See also: **Centrifugation**: Analytical Ultracentrifugation.

## Further Reading

- Brakke MK (1952) Density gradient centrifugation: a new separation technique. *Journal of the American Chemical Society* 73: 1847–1848.
- Coulson JM, Richardson JF, Backhurst JR, and Harker JH (1978) *Chemical Engineering, Volume 2: Unit Operations*, 3rd edn., 807pp. Oxford: Pergamon.
- Griffith OM (1986) *Techniques of Preparative, Zonal, and Continuous Flow Ultracentrifugation*; DS-468H, p. 50. Palo Alto, CA: Spinco Division of Beckman Instruments, Inc.
- Hsu HW (1981) In: Perry ES (ed.) *Techniques of Chemistry, Vol. XVI, Separations by Centrifugal Phenomenon*, 466pp. New York: Wiley.
- Lavanchy AC and Keith EW (1979) Centrifugal separation. In: Grayson M and Eckroth D (eds.) *Encyclopedia of Chem. Techn.*, 3rd edn., vol. 5, pp. 194–233. New York: Wiley.
- Price CA (1982) *Centrifugation in Density Gradients*, 430pp. New York: Academic Press.
- Ridge D (1978) In: Birnie GD and Rickwood D (eds.) *Centrifugal Separation in Molecular and Cell Biology*, ch. 3. London: Butterworths.
- Sheeler P (1981) *Centrifugation in Biology and Medical Science*, 269pp. New York: Wiley.
- Taulbee DN and Maroto-Valer MM (2000) Centrifugation. In: Cooke M and Poole C (eds.) *Encyclopedia of Separation Science*, pp. 17–40. London: Academic Press.
- [www.beckmancoulter.com/resourcecenter/labresources/centrifuges/rotorcalc.asp](http://www.beckmancoulter.com/resourcecenter/labresources/centrifuges/rotorcalc.asp).
- <http://researchlink.labvelocity.com/tools/conversionTools/CentrifugationTool.jhtml>.
- [www.centrifuges.bz/page2.htm](http://www.centrifuges.bz/page2.htm).

# CERAMICS

**G J Oliver and R N White**, Ceram Research Limited, Stoke-on-Trent, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The term ‘ceramic’ covers a wide range of materials, which makes it an interesting area of activity for the inorganic analyst. The range of matrices is described with reference to problems in preparation for analysis. Various methods of analysis are described and compared with respect to cost, general applicability, sensitivity, and accuracy. As might be expected, this article is primarily about bulk analysis, but speciation, surface analysis, and microstructural analysis are becoming more critical these days.

## Types of Ceramic Material

Ceramic materials are produced by heat treatment of minerals or mixtures of minerals, and the matrices have much in common with geological samples. For this reason, methods established for the analysis

of ceramic samples may also be applied to rocks, soils, ores and minerals as well as refractories, white-ware, glass, enamels, and cement. Ceramic materials may be divided into four industrial types as follows:

1. *Whitewares* (tiles, tableware, sanitaryware). These are traditional fired and unfired bodies, glazes, raw materials such as china clay, talc, bone or feldspar, and intermediates such as frits, slips, glazes, and colors.
2. *Building materials*. Bricks, breezeblock, cement, glass, concrete, tiles, clays, limestone, and additives such as stains and barium carbonate.
3. *Refractories*. Aluminosilicates, silica and aluminous material, basic materials such as magnesite, chrome-bearing materials such as magnesia-chrome bricks, zirconia-bearing materials, and silicon carbide. There are also raw materials such as fireclay, ferrosilicon, and graphite.
4. *Industrial or engineering ceramics*. This class includes a wide range of finished products and raw materials for biological, military, electronic, automotive, aerospace, and high-temperature uses. Typical materials include alumina, zirconia, titanate, silicon carbide and nitride, SiALON, and mixtures of rare earths.

The initial problem with ceramics lies in the fact that they have been designed to be resistant to wear, heat, and chemical attack, so that there are major difficulties in breaking them down or bringing them into solution. The other problem is that quite often there are two or more major constituents to determine. Materials discussed in this section can be regrouped in terms of difficulty of sample preparation.

(a) *High silica*. Samples are amenable to hydrofluoric acid attack, followed by analysis of the fused residue.

(b) *Silica/alumina*. Substances that contain silica and/or alumina as the major constituent. They generally require fusion for total decomposition.

(c) *Silicates*. These are effectively a variant of silica/alumina with another major element, and include talc, calcium silicates, cement, bone china, bone ash, and zircon. They are treated as silica/alumina, with lesser or greater difficulty in fusion and dissolution.

(d) *Basic*. Magnesite, limestone, and dolomite, which are decomposed by acid, usually with a fusion of the residue. Strontium and barium oxide also fall into this class but precipitation of sulfates is a problem.

(e) *Oxide*. Zirconia, titania, and other metal oxides that are usually difficult to decompose, whereas transition metal oxides (e.g.,  $\text{Fe}_2\text{O}_3$  and  $\text{MnO}$ ) are easier to decompose.

(f) *Reduced materials*. These include silicon carbide and nitride, SiAlON, silicon, graphite, and ferroalloys that require special methods for dissolution and speciation.

(g) *Complex materials*. These include some special ceramics, chrome-bearing refractories, colors, and enamels, where the presence of many minor and major constituents, especially rare earths, presents problems in analytical procedures. Decomposition procedures can fall into any of the six categories above. Ceramic materials also require analysis when added to composite matrices together with paint, plastic, rubber, metal, and cement.

## Sample Reduction and Communication

Obtaining a representative sample, which is not significantly contaminated with any of the elements to be determined, is an essential and vital part of ceramic analysis. However sophisticated the analytical procedures used, their results are worthless if sampling is not carried out properly. Many samples submitted for ceramic analysis are relatively small quantities of powder or small pieces that require

minimal grinding. Other samples, however, may be submitted in quantities of several kilograms in such forms as large bricks, broken rock, or wet clay.

Wet samples first require drying, possibly at temperatures as low as  $30^\circ\text{C}$ . Bulk samples are jaw crushed to  $\sim 4\text{ mm}$  and then subdivided by coning and quartering or by use of a riffle or mechanical sample divider to obtain a representative sample of 50–200 g for fine grinding.

As most analysis is by X-ray fluorescence (XRF) spectrometry and other instrumental methods, fine grinding is usually carried out in a tungsten carbide vial on a swing mill. The grinding parameters are set to produce a sample that would pass a  $125\text{ }\mu\text{m}$  sieve. Unless specified in a standard method, the sample is not actually sieved. For wet-chemical analysis, an agate pestle and mortar is to be preferred, but for hard materials of low silica content two samples are prepared: one in an iron percussion mortar, and another in an alumina mortar for subsequent iron determination. Tungsten carbide is unsuitable for 'wet' analysis as tungsten interferes with many of the determinations. For trace analysis, agate or boron carbide mortars are preferable but have to be used only for trace analysis to avoid surface contamination.

## Loss on Ignition

The determination of loss on ignition is a standard and is generally carried out at  $1025^\circ\text{C}$  on the dried  $110^\circ\text{C}$  sample.

Some samples require special procedures. Clays high in carbon such as brick or ball clay need a slower more careful pre-ignition to avoid silicon carbide (SiC) formation. Other materials require different temperatures. Bone ash requires ignition at  $1200^\circ\text{C}$ , fluorspar  $500^\circ\text{C}$ , glasses  $550^\circ\text{C}$ , plaster (gypsum)  $600^\circ\text{C}$  on an undried sample, and high-cobalt samples  $700^\circ\text{C}$ . Ferroalloys are not normally ignited, while glazes are heated over a burner at softening point. Silicon carbide, silicon nitride, silicon oxynitride, and SiAlON are ignited using a programmable furnace that allows slow burnout of carbon followed by a dwell at  $750^\circ\text{C}$ . In a proper full analysis, differential thermal analysis and thermogravimetric analysis (DTA/TGA) should be run both in air and argon atmospheres.

## Conventional Wet-Chemical Analysis

Wet-chemical analysis has largely been superseded by XRF analysis (with ICP for  $\text{Li}_2\text{O}$  and  $\text{B}_2\text{O}_3$ ) in all major ceramic laboratories. Some small laboratories cannot afford expensive capital equipment and continue to use wet methods, sometimes combined with

atomic absorption spectroscopy (AAS). Currently these methods are being revised within ISO rather than CEN as most European laboratories will not use them. Within ISO, wet methods are currently being brought up to date for silica–aluminosilicates–alumina (NP21587), limestone and dolomite (ISO10058), zircon and zirconia (based on JIS2012), and chrome-bearing samples (based on JIS2212). These procedures also allow the use of inductively coupled plasma (ICP) for some determinations.

The majority of wet-chemical analysis is for the silica–aluminosilicate–alumina range, and this is used to illustrate the methods used. The oxides determined are the normal eight (those of silicon, titanium, aluminum, iron, calcium, magnesium, potassium, and sodium). In ceramics, it is traditional to refer to the elements as their oxides. For the analyst, this has the advantage of providing an analytical total to cross-check the accuracy of the analysis.

The procedure for the determination of the alkalis ( $\text{Na}_2\text{O}$ ,  $\text{K}_2\text{O}$ , and  $\text{Li}_2\text{O}$ ) is effectively identical for all classes of material in that ~0.25 g of sample is decomposed with hydrofluoric acid together with dilute nitric and sulfuric acids in a platinum dish on a sand bath. The residue is dissolved in dilute nitric acid and alkalis determined directly on the solution by flame photometry or flame atomic absorption spectroscopy (FAAS) in emission mode. Cesium and aluminum sulfate buffers are added to aliquots for the flame photometric determination of sodium and potassium.

For classes (a)–(c) and for SiC, the main analysis solution is prepared via a low flux/sample ratio fusion or, in the case of alumina, a sinter at 1200°C. The flux used is a mixture of fusion mixture ( $\text{NaKCO}_3$ ) and boric acid; a lower flux/sample ratio is used for aluminous materials, and a larger flux/sample ratio is used for zircons. The melt is dissolved and the silica is dehydrated in a mixture of hydrochloric and sulfuric acids with the assistance of polyethylene oxide. The main silica is then determined gravimetrically. Other constituents, including residual silica, are determined on the filtrate combined with the residue remaining after the gravimetric silica determination.

- *Residual silica.* Colorimetric with ammonium molybdate reduced with tin(II) chloride (800 nm) or a mixture of tartaric and ascorbic acids to give the blue reduced heteropolymolybdate (650 nm).
- *Total silica.* Add the gravimetric and residual silica figures.
- *Iron oxide.* Colorimetrically with 1,10-phenanthroline (510 nm). At high levels in chrome-bearing samples, iron from the stock solution is eluted from a cation resin exchange column with

hydrofluoric acid; the iron is then complexed with CyDTA (*trans*-1,2-cyclohexanediamine-*N,N,N,N'* tetraacetic acid) and back-titrated with Zn using xylenol orange indicator. For chrome-bearing samples, a cleanup through an ion-exchange resin is required.

- *Titania.* Colorimetrically with hydrogen peroxide in phosphoric acid (398 nm) or di-antipyrimethane (DAM) at 390 nm. If  $\text{ZrO}_2$  is present, the solution is treated with CyDTA solution first.

- *Lime (calcium oxide).* Titrimetrically with EDTA or EGTA in triethanolamine and potassium hydroxide with a screened calcein indicator. Alternatively with AAS, particularly for  $\text{ZrO}_2$ -containing materials and limestone.

- *Lime plus magnesia.* Titrimetrically with EDTA in triethanolamine and ammonia using methylthymol blue indicator. For chrome-bearing samples, a cleanup through an ion-exchange resin is required first followed by EDTA titrimetry with an Eriochrome Black T indicator.

- *Magnesia (magnesium oxide).* Derived by the difference of the above and lime or directly by AAS, especially if  $\text{ZrO}_2$  is present.

- *Alumina high levels.* Cleaning up into chloroform with cupferron followed by adding excess 1,2-diaminocyclohexanetetraacetic acid (DCTA) or CyDTA and back-titrating with zinc. For chrome-bearing samples, a cleanup through an ion-exchange resin is required.

- *Alumina low levels.* By ICP.

- *Phosphorus (V) oxide.* If  $\text{ZrO}_2$  is present, mask with CyDTA. Then adjust pH and determine colorimetrically with ammonium molybdate reduced in ascorbic acid at 830 nm.

- *Zirconia + hafnia in zircon and zirconia.* Gravimetrically by DL-mandelic acid.

- *Manganese oxide ( $\text{MnO}$ ).* Colorimetrically by oxidation to permanganate with potassium periodate at 524 nm. For chrome-bearing samples, a cleanup through an ion-exchange resin is required.

- *Chromium oxide ( $\text{Cr}_2\text{O}_3$ ).* Up to 0.1% colorimetrically with diphenylcarbazide at 540 nm. Above 0.1% but as a minor constituent, colorimetrically with EDTA at 550 nm. As a major constituent by oxidation to dichromate by peroxodisulfuric acid using a silver nitrate catalyst, destruction of permanganate with HCl and titration against ferrous ammonium sulfate using diphenylamine-4-sulfonate indicator.

## X-Ray Fluorescence Spectrometry

In many ways, modern ceramic analysis is synonymous with XRF, usually combined with the fused,

cast-bead method. This combination is so close to being a universal method that other techniques are regarded as only gap fillers or a cheap alternative. This method of sample preparation and the XRF technique combine ease of preparation, primary calibrations, and better accuracy than can be achieved by 'wet' methods. This has revolutionized ceramic analysis and allowed faster and cheaper analysis for routine samples and, for others (e.g., glazes and colors), has made their analysis economically viable. It also means that very complex materials can finally be analyzed. In 2003, the definitive method ISO12677 was published, which is a standard for the analysis of refractories and refractory raw materials by the XRF fused cast-bead method. The standard features two methods of calibration. One is the synthetic calibration route, used in much of Europe, and the other employs 'Series Reference Materials', an off-the-peg approach used successfully in Japan, which are mixtures of oxides and compounds for short-range calibrations that are checked chemically. Although designed for refractories, ISO12677 will be applied to all ceramics as well as geological materials. It includes uncertainty data that are required under ISO17025: 2000, the Test House Quality Standard.

### Sample Preparation

As implied above, the most popular sample preparation for ceramic analysis by XRF is the fused cast-bead technique whereby the ground but heterogeneous sample is converted into a homogeneous glass by fusing in a flux. Because lithia (lithium oxide) cannot be determined by XRF, and boric oxide is difficult to determine, most XRF fluxes have been based on combinations of lithia and boric oxide. Lithium tetraborate is applied to most basic materials such as magnesite or limestone as well as to oxides such as titania and iron oxide categories ((d) and (e) above). Typical sample flux ratios are: 3:1 for trace analysis on limestone; 5:1 for limestone and dolomite; 10:1 for magnesite, magnesite spinel, magnesite-chrome, iron oxide, and manganese oxide; and 12:1 for zirconia, titania, barium carbonate, and barium titanate.

Lithium tetraborate is not ideal for siliceous materials (categories (a), (b), (c), and much of (g) above), because it is too acidic for an acidic siliceous matrix. A near-eutectic mixture of one part by weight of lithium tetraborate to four of lithium metaborate used at a ratio of 5:1 (w/w) with the sample allows easy fusion of high-silica materials. The lower melting point also permits lower fusion temperatures, aiding the retention of volatile components such as sulfur, and reducible metal oxides such as PbO,

provided oxidizing conditions are maintained. The same flux is equally applicable to silica, alumina, clays, glazes, bone, plaster, dish-washing detergents, talc, and zircon. The fusion temperature is usually 1200°C except where volatiles need to be retained, when 1050°C is used. When one-off samples are analyzed, it is often simpler to analyze them by fusing in this flux as two dilutions, one in Al<sub>2</sub>O<sub>3</sub> and the other in SiO<sub>2</sub>.

Chrome-bearing refractories are particularly difficult to fuse, and, although not entirely satisfactory, a flux mixture of 10 parts of lithium metaborate and 12.5 parts of lithium tetraborate to one part of sample is generally used for chrome–magnesite, chrome–magnesia–zirconia and chrome ore samples, but not metallurgical chrome ore, which requires further dilution in MgO.

The reduced materials (category (f)) require special fusion conditions because dissolution produces an exothermic reaction that can destroy the platinum alloy fusion vessels. For example, the flux used for siliceous materials is reconstituted as lithium tetraborate and lithium carbonate. A SiC or other reduced sample is mixed with lithium carbonate and sintered on top of a protective layer of lithium tetraborate that has been fused and spread over the dish. The weight of the reduced sample therefore needs to be adjusted to maintain the flux to a (oxidized) sample ratio at 5:1. If samples lie in category (c), lithium tetraborate is replaced by boric oxide, which together with the lithium carbonate will ultimately give the appropriate lithium tetraborate/sample ratio (ignited basis).

Having selected the appropriate flux, the preparation of a sample is as follows. The ignited sample is mixed with the appropriate weight of ignited (500°C) flux (or a weight of flux corrected for this ignition) in a platinum/5% gold alloy dish. The mix is fused over a burner or in a 1050°C muffle for reducible samples until the sample is dissolved. The next part of the procedure is that used for all samples where the homogeneous melt is then poured (cast) from a 1200°C furnace into a preheated platinum/5% gold alloy mold, and the subsequent bead, in its mold, is annealed over an air-jet. This generates the fused cast bead. Either surface may be presented to the spectrometer for analysis, provided the surface is used consistently throughout.

There are automatic fusion devices for preparing samples for XRF, but most laboratories preparing a mixture of sample types still prefer a manual process. Because of various new features of XRF, the pressed disc technique, of mixing the ground sample with a binder and pressing in a die, is finding more application. Its use is discussed below.



## Analysis by XRF Spectrometry

Analysis is best exemplified by the most usual class of samples (category (a)). The constituents determined include the eight normally determined by wet-chemical analysis. However, XRF is much more selective than 'wet' methods, so that to obtain a reasonably complete analysis, some or all of the following oxides may need to be determined:  $\text{SiO}_2$ ,  $\text{TiO}_2$ ,  $\text{Al}_2\text{O}_3$ ,  $\text{Fe}_2\text{O}_3$ ,  $\text{CaO}$ ,  $\text{MgO}$ ,  $\text{K}_2\text{O}$ ,  $\text{Na}_2\text{O}$ ,  $\text{P}_2\text{O}_5$ ,  $\text{Cr}_2\text{O}_3$ ,  $\text{Mn}_3\text{O}_4$ ,  $\text{ZrO}_2$ ,  $\text{HfO}_2$ ,  $\text{BaO}$ ,  $\text{SrO}$ ,  $\text{ZnO}$ ,  $\text{PbO}$ , and  $\text{SO}_3$  (retained after fusion). In addition,  $\text{WO}_3$  must always be determined to allow for the correction for the effects of contamination with tungsten carbide on grinding, both in terms of dilution and the loss on ignition figure. For zircon and zirconia, it is also necessary to determine  $\text{U}_3\text{O}_8$ ,  $\text{ThO}_2$ ,  $\text{SnO}_2$ , and  $\text{CuO}$ . Rare earths are often added to a zirconia matrix, so  $\text{La}_2\text{O}_3$ ,  $\text{CeO}_2$ ,  $\text{Nd}_2\text{O}_3$ , and  $\text{Pr}_6\text{O}_{11}$  are often found as major or minor constituents and the others as impurities.  $\text{Y}_2\text{O}_3$  is added as an alternative to  $\text{CaO}$ ,  $\text{MgO}$ , or  $\text{CeO}_2$  as a stabilizer for zirconia.

With the fused cast-bead technique, synthetic calibration is used to generate a primary calibration needed for referee quality work. Calibration derived from a series of binary standards with a silica base. These same binaries, plus others in alumina, allow quantification of line interference of background effects of one element on another. The third part of the calibration is the determination of mass absorption or  $\alpha$  correction effects (eqn [1])

$$I_T = I_M(1 + \alpha_{ij}C_j) \quad [1]$$

where  $I_T$  is the true intensity of the analyte,  $I_M$  is the measured intensity of the analyte,  $\alpha_{ij}$  is the absorption or enhancement effect on the analyte per percentage of interfering oxide  $j$ , and  $C_j$  is the concentration of that oxide  $j$ . For more than one interfering oxide,  $\alpha_{ij}C_j$  is replaced with  $\sum \alpha_{ij}C_j$ .

These corrections are derived from the binary synthetic beads above, together with appropriate tertiary synthetic beads. With the advent of generally available computer software, there is a move towards theoretically derived corrections either from an Excel spreadsheet or supplied through the instrument manufacturer's software. The calibration for each element involves up to  $(n-1)^2$  line interference effects and  $(n-1)^2$  corrections for each element, plus drift correction.

XRF instrument parameters for ceramics differ from those used for some other materials because of the light element determinations. Rhodium, or less frequently scandium, X-ray tubes are used at voltages in the relatively low kilovolt range. Thallium acid phthalate or layered crystals are used for

soda and magnesia and indium antimonide for silica. Another difference is that in ceramic analysis zirconia, strontia, and lead oxide are frequently present at high concentrations, necessitating the use of longer wavelength lines on a Ge crystal, e.g.,  $\text{ZrL}\alpha$  not  $\text{ZrK}\alpha$ ,  $\text{PbM}\alpha$  not  $\text{PbL}\beta_{1,3}$ . Zirconia causes many line overlaps, so a flow counter is used in preference to a sealed Kr counter in many instances. High-order overlaps are reduced using 30 kV on the X-ray tube rather than 60 kV, so Hf will be determined on the  $\text{L}\beta$  line using a flow counter, narrow collimator on at 30 kV. Lithium fluoride (220) crystals are used in preference to the normal lithium fluoride (220) to separate Ba from Ti peaks.

## Other Uses of XRF

Modern instruments are now being sold with excellent semiquantitative software, light element facilities, and the ability to carry out some speciation. The latter will be dealt with in a separate section below.

There are at least two good semiquantitative packages on the market. These can be used with powders, metals, lumps, pressed discs, liquids and slurries (with He atmosphere), and metal samples. These are very useful for identification or fault finding (e.g., identification of metallic impurities in a slip). With additional information like density, dimension, and elements known to be present or not present, accuracy can be greatly improved. Newer versions of software can be used as an alternative approach to calibration rather than type calibration. Powder samples can be pressed into discs with a wax or  $\text{H}_3\text{BO}_3$  powder or put directly onto a Mylar film and analyzed like liquids.

The major instrument manufacturers can now supply crystals and collimators to determine O, F, C, and B. One manufacturer also claims that Be is possible. Samples in this case are usually ground down or micronized to a particle size similar to that used for XRD below. Wax and  $\text{H}_3\text{BO}_3$  are the best binders. If B and C are required, then two discs are required in the two binders unless the sample is self-binding. Glass and glaze can be analyzed directly, provided a reasonably flat surface is available. Light element analysis is currently of semi- to quantitative quality but work is currently in hand to improve this, as there is a large incentive to avoid the lengthy sample preparations, as discussed above, for a few elements.

A further use of XRF is in depth profiling Pb in glazes. The  $\text{M}\alpha$  line measures the surface and the  $\text{L}\beta$  is used to give a bulk Pb content. This shows Pb redeposition on the surface, which leads to high metal release results.

## Techniques Complementary to XRF

Alternative methods are used where XRF fails, i.e., for elements below sodium in the periodic table, for volatile elements and precious metals not amenable to fusion, for liquids, and for some trace analyses. In the case of trace analysis (in ceramics  $<100\text{ }\mu\text{g per g}$ ), XRF meets the requirements for many elements even with a 5:1 flux/sample ratio, particularly for U, Th, Y, La, V, Rb, Cs, Ga, Ge, Ce, Nd, Pr, Sc, and Ni.

The matrices for trace analysis are generally restricted to silica, alumina, aluminosilicates, cobalt aluminate, limestone, zircon, barium titanate, and zirconia. Sample preparation for the techniques below is similar to those for full analysis by wet methods above, but using purer reagents.

### Atomic Absorption Spectrometry

In ceramics, unlike many industries, AAS has not been greatly used, partly because of chemical and other matrix problems, and partly because XRF appeared on the scene shortly after AAS came into use. Its main application in the ceramic industry has been in the determination of metal release from ceramic ware. This is a class of tests designed to establish the likelihood of lead or cadmium leaching from ceramic ware and involves a 24 h extraction at  $22^{\circ}\text{C}$  with 4% (v/v) acetic acid and subsequent determination of lead and cadmium by flame AAS. Current legislation in the USA is driving limits to a level where atom trap AAS, ICP, or graphite furnace AAS is needed.

Within the greater ceramic community, there is a demand for the use of AAS as part of a wet-chemical analysis. As this is not of interest in Europe, standards are being developed at ISO rather than CEN. Currently, there are methods within ISOTC33 for silica/aluminosilicate/aluminous materials, magnesite and dolomite (ISO10058 revision), zircon/zirconia, and chrome-bearing samples.

For trace analysis, the main ceramic elements of interest are Zn, Pb, Cu, Bi, Sb, Sn, Ag, As, Mn, Cr, Se, and Hg. Many of these are environmentally important. In certain cases the detection limits of flame AAS are inadequate, so that hydride generation for antimony, selenium, arsenic and bismuth, cold vapor for mercury, and graphite furnace AAS for lead and cadmium are required. A variation of AAS is atomic fluorescence, and this is used to achieve the detection limits needed for Hg and Se in environmental samples. Microwave digestion techniques for sample preparation are becoming more common, where, unlike fusion, there is no risk of loss of volatile elements from unfired samples and fewer reagents are

required. This technique scores for batches of samples of similar type.

### Inductively Coupled Plasma Atomic Emission Spectrometry

Inductively coupled plasma atomic emission spectrometry (ICP-AES) is the next most important ceramic instrumental method, and is competitive with XRF at minor determination levels and better at trace levels. Its main use in conventional analysis is in the determination of lithium and boron in glass, glazes, borates, petalite, and magnesite. Sample preparation is identical to that for AAS. Another major use is for the analysis of liquid samples such as acid- or water-soluble salts from clay or bricks, slip liquors, and ceramic waste-product leachates (for landfill, packaging waste, and special waste regulations). A major application of ICP is in the analysis of composition of mortars (BS4551) and concrete (BS1881). Once the separations are complete, ICP gives a very fast series of determinations and greater element coverage. The former means that mortars can be checked during construction and the latter allows a better elucidation of historic mortars. As discussed under AAS, at ISOTC33, ICP is cited as a method of determination of minor constituents as part of a full analysis.

ICP-AES is normally preferred for trace analysis because of its generally better detection limits, freedom from chemical matrix effects, and multi-element capability. When available, it is used instead of AAS. Like AAS, hydride and mercury vapor attachments can be used to enhance the detection limits of the elements listed above. As an alternative to normal sample preparation or microwave procedures, laser ablation is creeping in as an alternative approach. The same pressed discs used for XRF are ablated by an IR or UV laser in a chamber that replaces the spray chamber used normally. The sample preparation takes minutes rather than days or hours, eliminates reagents (even pure reagents are less pure than ceramic powders), and is less susceptible to external contamination.

### Ion Chromatography

Ion chromatography (IC) is capable of determinations impossible by XRF and ICP and can carry out speciation. It is applied mainly to the determination of anions such as  $\text{Cl}^-$ ,  $\text{F}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{SO}_3^{2-}$ ,  $\text{NO}_3^-$  and  $\text{NO}_2^-$ . The determinations are made after sodium carbonate fusion or on water- or acid-soluble extracts. Fluorine is usually separated by pyrohydrolysis, where IC is used as a more sensitive and selective method of determination than the use of ion-selective electrodes. These days it is used for environmental

purposes under IPPC together with additional columns for ammonia, cyanide, and sulfide.

### Single-Element Methods

Even the four techniques discussed above are inadequate for completing all ceramic analyses, and additional methods are required for specific elements:

- *Boric oxide.* A cheaper alternative to ICP–AES is a photometric determination by carminic acid.
- *Carbonate* ( $\text{CO}_3^{2-}$ ). By reaction with acid followed by absorption and weighing of the evolved carbon dioxide.
- *Total carbon, oxygen, and nitrogen.* Usually pyrometrically by commercially available apparatus originally developed for the steel industry.
- *Total sulfur.* Commercially available apparatus employing pyrolysis.

### Speciation

Speciation analysis is particularly demanded for SiC- and nitrogen-containing ceramics, and because species like  $\text{Al}_4\text{C}_3$ , BN,  $\text{B}_4\text{C}_3$ , free aluminum, and free magnesium are being added to ceramics. Another driver is environmental legislation. X-ray diffraction (below) is particularly useful for crystalline species where standards are available. DTA/TGA used in air and an inert atmosphere can be used for the determination of free and SiC carbon in complex silicon carbide materials, especially those containing reactive graphite. XRF can also be used in some instances. Normally the analyst carefully chooses the  $K\alpha$  line for lighter elements or the  $L\alpha$  line for elements for the determination of elements of atomic number 38 (Sr) or above. This is because these transitions are not affected by electronic environment. If the  $L\beta$  line is used for the lighter elements or an L line for the heavier elements, the lines show the effects of chemical environment. These can be manifest in one, two, or three ways – peak shift, a satellite peak, or a shoulder on the peak. A workable method has been developed for  $\text{CrVI}/\text{Cr}^{3+}$  in refractories. Free silica has a peak shift away from both Si and SiC. The generation of a gas into a nitrometer is another useful tool. Some methods being used at present are the following:

- *Free carbon in SiC, etc.* By extended temperature-controlled loss on ignition, using a special apparatus specified by DIN 51075 part 2 (1984) or using DTA/TGA.
- *Silicon carbide.* The elegant way is by the subtraction of free carbon from total carbon and

multiplying by 3.3389. A cheaper alternative is to multiply the analytical total from XRF, including loss on ignition by 2.0060.

- *Free silica.* There are two simple ways, depending on the constituents present. The first is to determine total silica and subtract the Si in free silicon, SiC,  $\text{Si}_3\text{N}_4$  and multiply by 2.1392 or a similar approach using total oxygen, subtracting the O in other species and multiplying by 1.8778. The more elegant method applicable to nitride-bonded silicon carbides is by distillation of  $\text{SiF}_4$  from the sample with HF followed by ICP determination of Si in the distillate and multiplying by 2.1392. The use of the XRF  $K\beta$  line may have potential in this determination.
- *$\alpha$ - $\text{Si}_3\text{N}_4$ ,  $\beta$ - $\text{Si}_3\text{N}_4$ , and  $\text{Si}_2\text{ON}_2$ .* Determine the individual species by XRD and normalize them to the total nitrogen.
- *Sialon* ( $\text{Si}_{(6-z)}\text{Al}_z\text{O}_z\text{N}_{(8-z)}$ ). This determination is similar to the one above and relies on sialon giving an XRD intensity similar to  $\text{Si}_3\text{N}_4$  and calculation from the total nitrogen figure less than that in other nitrogen species. The  $z$  value and hence the formula is calculated by resolving the crystal spacing.
- *Free silicon.* By displacement of silver from silver fluoride or hydrogen into a nitrometer from sodium hydroxide after first washing out free iron and aluminum with acid. If FeSi or FeSi<sub>2</sub> are present, then these need to be determined by XRD and corrections made. XRD also produces accurate silicon results.
- *Free aluminum.* If free silicon and ferrosilicon is low, it can be determined by the free silicon method with a suitable factor. Alternatively, it can be determined by the volume of gas evolved with (1 + 1) HCl and corrected for the free iron content.
- *Free iron.* By bromine/methanol or copper sulfate extraction, followed by ICP determination of Fe.
- *Ferrous iron.* By hydrofluoric acid attack excluding air followed by redox titration with dichromate.
- *Hexavalent chromium.* The method BS1902-9.3, in which the ground sample is extracted with a solution of 2% NaOH/3%  $\text{Na}_2\text{CO}_3$  and the extract is measured spectrophotometrically using diphenyl carbazide at 540 nm.
- *Other inorganic species.* Methods are currently being developed at CEN TC187 WG4.
- *Organic constituents.* Organic constituents are being increasingly used in a ceramic context and their determination may be required in solid samples, organic liquids, gaseous emissions, or aqueous effluents; determination is by gas chromatography (GC) or better by gas chromatography–mass spectrometry (GC–MS) after thermal desorption, solvent extraction, or purge and trap. FTIR and HPLC and colorimetric methods are used for other species.

## Microstructural Analysis

In addition to chemical data on ceramic materials, information on the mineralogical constituents by speciation can be vital to a complete understanding of the material (see above). Both X-ray diffraction (XRD) and microscopic examination can provide such mineralogical information. Additional structural knowledge must be obtained using microscopic techniques.

### X-Ray Diffraction

XRD analysis provides a means by which different crystalline phases are characterized and identified. Samples are normally prepared as finely ground material and then presented to the X-ray beam in such a way that individual crystallites are randomly orientated. Comparison of diffraction traces with standard reference profiles enables the identification of phases to be made. Reference patterns are published by the International Centre for Diffraction Data. In some cases with a careful use of reference materials, internal standards and mass absorption corrections, some quantitative results are obtainable using a variety of mathematical formulas on the peak heights or areas.

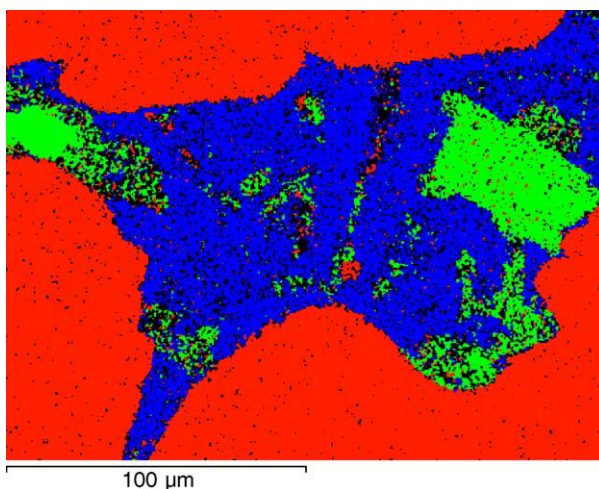
XRD may be applied to a wide variety of ceramic problems such as simple phase identifications, crystallite size measurements, and determination of crystal lattice parameters. The mineral assemblages produced during firing or in different service environments can be readily studied. Techniques are now available for micro XRD, in which an intense X-ray beam is collimated or focused onto a specific feature in the sample in order to characterize the crystal structure.

### Microscopy

Microscopic examination can be used in combination with both chemical and XRD results so that interrelationships between phases, grain shapes, and sizes can be examined to give a more complete evaluation. Ceramic materials are generally examined using reflective light microscopy on highly polished sections. Selective etching techniques are employed to enhance features such as grain boundaries and contrasts between mineral phases.

Optical microscopy can be applied to examine porosity, grain size, grain intergrowths, degree and type of ceramic bonding, fracture systems, and identification of contaminants and causes of failure. It can also be used as a tool to predict the likely behavior of materials in a variety of service environments.

Scanning electron microscopy (SEM) can be used in a similar manner, although examination at much



**Figure 1** Spatial distribution of components in a ceramic material (brown fused alumina) as shown by overlaid elemental distribution using energy dispersive analysis.

higher magnification is possible. SEM examination is not constrained to polished flat surfaces and can be applied to rough surfaces such as fractures and powders. Microprobe analysis is often conducted in conjunction with SEM examination and provides quantitative chemical data from areas of samples as small as a few micrometers in size. The high spatial resolution enables analyses to be made on specific crystals, phases, and positions within diffusion profiles. By integrating the control of the electron microscope with a microanalysis system, it is possible to collect spatial chemical information or chemical maps of a sample. These can be false colored to visually differentiate component parts of the sample (Figure 1). Another feature that can be used for electron microscopy is integrating electron diffraction with the microscope so that information on crystal orientation can be obtained. Combining electron diffraction with control of the microscope, grain boundaries can be delineated as points of change of crystallographic orientation.

*See also:* **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Building Materials. Cement. Gas Chromatography:** Mass Spectrometry. **Ion Exchange:** Ion Chromatography Instrumentation; Ion Chromatography Applications. **Sampling:** Theory. **Sulfur.**

## Further Reading

Bennett H (1977) Analysis of silicates – ceramic and geochemical. A review. *Analyst* 102: 153–179.

- Bennett H and Oliver GJ (1992) *XRF Analysis of Ceramics, Minerals and Allied Materials*. Chichester: Wiley.
- Bennett H and Reed RA (1971) *Chemical Methods of Silicate Analysis*. London: Academic Press.
- Drake SA and Tyson JF (1993) Interaction of laser radiation with solid materials and its significance to analytical spectrometry. *Journal of Analytical Atomic Spectrometry* 8: 145–200.
- Hendry GL and Thomas JE (1988) *Trace Analysis Data for Five BCS Certified Reference Material*. Special

Publication 118. Stoke-on-Trent: British Ceramic Research Ltd.

Julietti RJ and Reeve BCE (1991) A new approach to the analysis of silicon carbide materials. *Transactions and Journal of the Institute of Ceramics* 90: 85.

Oliver GJ, Salt PD, Cooper JJ, Hodson PTA, Thompson PD, and Bennett H (1984) *Proceedings of the Seventh Ceramic Chemists' Conference*, Special Publication 110. Stoke-on-Trent: British Ceramic Research Ltd.

For BS, ISO and EN standards, search [www.bsonline.co.uk](http://www.bsonline.co.uk)

## CEREBROSPINAL FLUID

**E J Thompson**, The Institute of Neurology, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

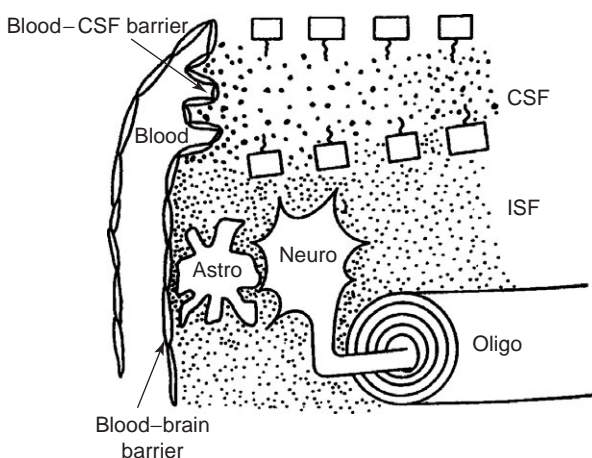
Cerebrospinal fluid (CSF) contains a myriad of analytes. Rather than attempting to give a complete list of all the constituents, the emphasis in this article will be on giving some of the guiding principles for interpretation of the results of analysis. There will also be a critical approach to the methods of collection of CSF, since this can also have a profound effect on the final results of analysis. The interpretation of results also depends primarily on the nature of the analyte.

### What is CSF?

CSF is not a homogeneous fluid, but is derived from various sources. It is therefore relevant to consider, if only in a brief fashion, the source of the various brain fluids, each of which will contribute relatively different amounts of analytes. Studies on the origin of CSF are conveniently described under the four classical headings of anatomy, physiology, biochemistry, and pathology.

#### Anatomy

About two-thirds of CSF is derived from the choroid plexi (blood–CSF barrier) located in the lateral ventricles and eventually finds its way back into the venous blood through the superior sagittal sinus. The other one-third of the fluid is derived from the brain parenchyma, in part from filtration across the blood–brain barrier (see **Figure 1**) and in part from



**Figure 1** Two normal brain barriers that proteins transgress to reach the CSF.

endogenous synthesis within the brain, where water is also produced as an end product of metabolism. As well as flowing over the brain, CSF also goes down the spinal sac to envelop the spinal cord. The most common position in which the needle is inserted for a lumbar puncture to obtain a sample of CSF is at the bottom of the spinal cul-de-sac. However, one can also perform a cisternal puncture through the foramen magnum, as well as a ventricular puncture through a burr hole. The concentration of total proteins is 300% higher in lumbar CSF than in the ventricular fluid. A characteristic protein of ventricular fluid is transthyretin (prealbumin), which is synthesized locally by the choroid plexi. In many species there is significant drainage of the CSF along the cranial nerves, with egress into the lymph nodes of the head and neck. However, this route is of only minor significance in humans. The fluid is propelled by the cardiac pulsations and in humans the total volume is ~150 ml. With a production rate of some

- Bennett H and Oliver GJ (1992) *XRF Analysis of Ceramics, Minerals and Allied Materials*. Chichester: Wiley.
- Bennett H and Reed RA (1971) *Chemical Methods of Silicate Analysis*. London: Academic Press.
- Drake SA and Tyson JF (1993) Interaction of laser radiation with solid materials and its significance to analytical spectrometry. *Journal of Analytical Atomic Spectrometry* 8: 145–200.
- Hendry GL and Thomas JE (1988) *Trace Analysis Data for Five BCS Certified Reference Material*. Special

Publication 118. Stoke-on-Trent: British Ceramic Research Ltd.

Julietti RJ and Reeve BCE (1991) A new approach to the analysis of silicon carbide materials. *Transactions and Journal of the Institute of Ceramics* 90: 85.

Oliver GJ, Salt PD, Cooper JJ, Hodson PTA, Thompson PD, and Bennett H (1984) *Proceedings of the Seventh Ceramic Chemists' Conference*, Special Publication 110. Stoke-on-Trent: British Ceramic Research Ltd.

For BS, ISO and EN standards, search [www.bsonline.co.uk](http://www.bsonline.co.uk)

## CEREBROSPINAL FLUID

**E J Thompson**, The Institute of Neurology, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

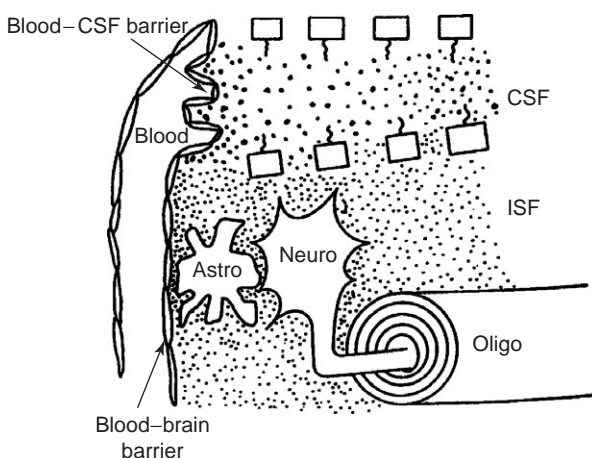
Cerebrospinal fluid (CSF) contains a myriad of analytes. Rather than attempting to give a complete list of all the constituents, the emphasis in this article will be on giving some of the guiding principles for interpretation of the results of analysis. There will also be a critical approach to the methods of collection of CSF, since this can also have a profound effect on the final results of analysis. The interpretation of results also depends primarily on the nature of the analyte.

### What is CSF?

CSF is not a homogeneous fluid, but is derived from various sources. It is therefore relevant to consider, if only in a brief fashion, the source of the various brain fluids, each of which will contribute relatively different amounts of analytes. Studies on the origin of CSF are conveniently described under the four classical headings of anatomy, physiology, biochemistry, and pathology.

#### Anatomy

About two-thirds of CSF is derived from the choroid plexi (blood–CSF barrier) located in the lateral ventricles and eventually finds its way back into the venous blood through the superior sagittal sinus. The other one-third of the fluid is derived from the brain parenchyma, in part from filtration across the blood–brain barrier (see **Figure 1**) and in part from



**Figure 1** Two normal brain barriers that proteins transgress to reach the CSF.

endogenous synthesis within the brain, where water is also produced as an end product of metabolism. As well as flowing over the brain, CSF also goes down the spinal sac to envelop the spinal cord. The most common position in which the needle is inserted for a lumbar puncture to obtain a sample of CSF is at the bottom of the spinal cul-de-sac. However, one can also perform a cisternal puncture through the foramen magnum, as well as a ventricular puncture through a burr hole. The concentration of total proteins is 300% higher in lumbar CSF than in the ventricular fluid. A characteristic protein of ventricular fluid is transthyretin (prealbumin), which is synthesized locally by the choroid plexi. In many species there is significant drainage of the CSF along the cranial nerves, with egress into the lymph nodes of the head and neck. However, this route is of only minor significance in humans. The fluid is propelled by the cardiac pulsations and in humans the total volume is ~150 ml. With a production rate of some



500 ml day<sup>-1</sup>, this means that the fluid is renewed about once every 6 h. The flow rate is slower around the spinal cord because this is in the cul-de-sac.

Turning to the cellular anatomy of the CSF, the normal cell count is up to 4 µl<sup>-1</sup>. This is constituted of approximately two-thirds lymphocytes and one-third macrophages. This comprises an ideal mixture for the production of antibodies. Since the brain has very few lymphatic ducts, it has been argued that one of the main functions of CSF is to act as a kind of lymph for the brain tissue. Under pathological conditions this function becomes especially prominent.

### Physiology

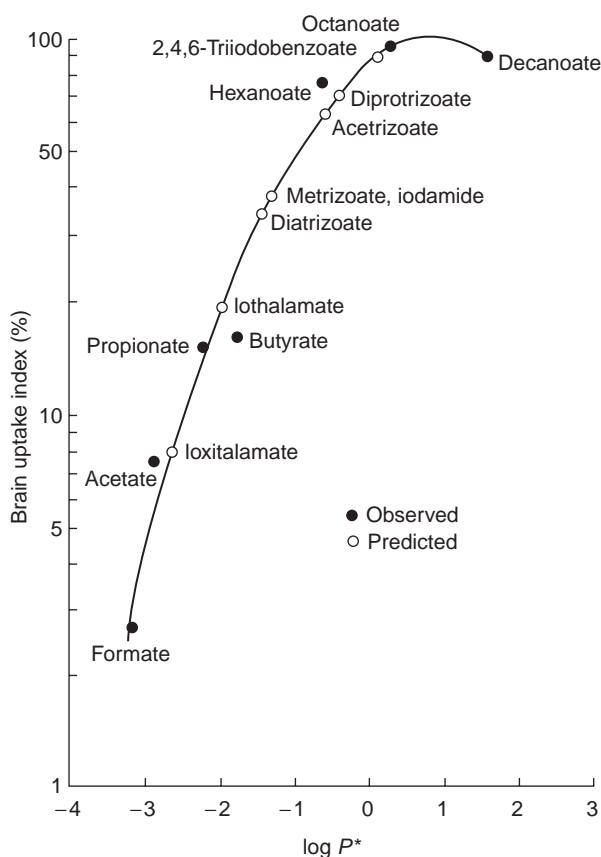
Part of the fluid is derived from the blood as an ultrafiltrate of plasma and part is derived from the brain due to local synthesis of water by glucose metabolism. The brain parenchyma contributes not just water but many analytes. Separate and apart from this, the cells that normally migrate within the CSF also produce analytes. In the blood the main cells are the polymorphonuclear leucocytes, which are, however, rarely found in normal CSF. They may be present in CSF as a result of a traumatic puncture that causes the abnormal contamination of CSF with blood, or as a consequence of various pathological conditions.

Analytes that are derived from the blood plasma can be conveniently divided into those that are actively transported and typically maintained at homeostatic levels, as opposed to those that diffuse down concentration gradients. In between is carrier-mediated transport, which is rate limited typically by a saturable enzymatic transporter system. The passive diffusion of macromolecules such as proteins is influenced by the molecular sieving effect of various barriers that are constituted by basement membranes with varying degrees of effective 'pore size'.

Depending upon where the fluid is sampled along the rostro-caudal neuraxis (ventricular, cisternal, or spinal puncture), there may easily be differences in concentration for particular analytes; for instance, various neurotransmitters synthesized within the thalamic nuclei will have a much higher concentration in the cistern than that found in the spinal cul-de-sac.

### Biochemistry

Perhaps the most important molecular characteristic is that of hydrated size or relative molecular mass ( $M_r$ ), whereas the role of net charge is much less important. The other important characteristic of molecules, especially those of small molecular size, is their degree of lipophilia. For molecules of  $M_r$  less than ~500, the oil-water partition coefficient is the



**Figure 2** Correlation between higher percentage brain uptake (ordinate axis) and increasing lipid solubility (the octanol-water coefficient, modified by the dissociation constant) consequent upon increasing chain length from formate to octanoate (abscissa).

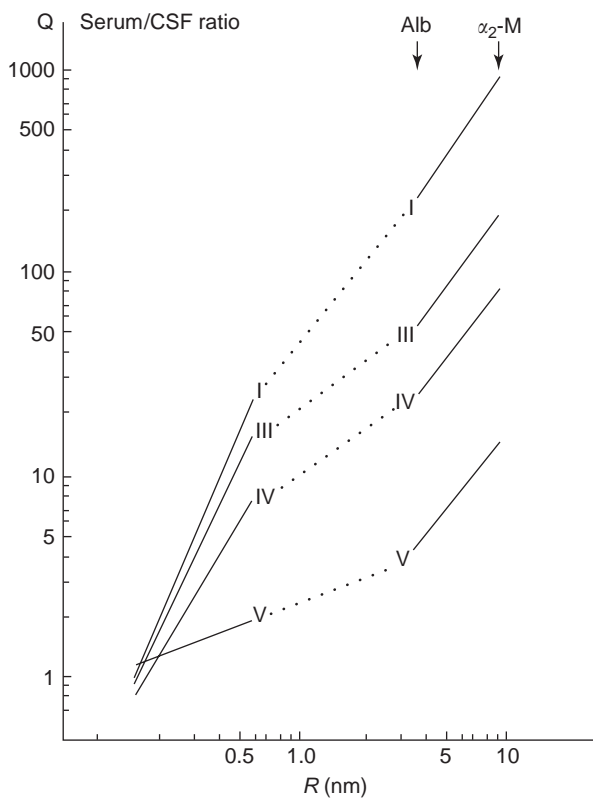
main determinant for entry into the CSF from the serum (Figure 2). In an attempt to bridge the gap between low  $M_r$  and high  $M_r$ , it can be seen that the slopes are not derived from an obvious continuous line of unchanging slope (Figure 3).

Proteins can be conveniently classified according to a kind of quantum rule, in which increasing orders of magnitude for percentage transfer (Table 1) seem to conveniently divide proteins from the two extremes of those synthesized within the brain versus those filtered from the plasma, whilst in between, some proteins are synthesized by the brain to the same extent as other tissues (as opposed to being a 'unique' synthesizer). The remaining proteins seem to be synthesized by the cells that normally inhabit the CSF, the only exception being prealbumin, which shows partial intrathecal synthesis. A more complete list of proteins is given in Table 2.

### Pathology

The presence of significant numbers of polymorphs within the CSF can be iatrogenic, due to traumatic





**Figure 3** Serum/CSF ratio plotted against Stokes hydrated molecular radius,  $R$ . Comparison of low  $M_r$  (left) vs. high  $M_r$  (right) analytes in five stages (I–V) of increasing barrier breakdown. The most dramatic change of slope is in stage V of the low  $M_r$  analytes. The dotted lines also show that it is not a continuous linear function between the two analytes (low vs. high  $M_r$ ). (Reproduced from Felgenhauer K, Liappis N, and Nekic M (1982) Low molecular solutes and the blood cerebrospinal fluid barrier. *Klinische Wochenschrift* 60: 1385–1392; © Springer-Verlag.)

puncture, but is also found in inflammatory disease of the brain. They can synthesize characteristic proteins, such as lactoferrin, lysozyme, and eosinophil cationic protein. There are two other barriers that normally contribute only a minor amount of fluid to the normal CSF, but under various pathological conditions, they can become major contributors. These are the dorsal root ganglia, which can leak large amounts of protein as in Guillain-Barré polyneuropathy. The other source is the meninges, which in various infections, or with carcinomatous deposits, can also give rise to massive increases in the level of CSF proteins.

## Why Analyze It?

The two main reasons why people have traditionally analyzed CSF are: (1) because it represents a kind of

**Table 1** Proteins grouped in order of 'percentage transfer' (or concentration in CSF divided by concentration in serum)

1	500% (approx.)
	Primary intrathecal synthesis
	$\beta$ Trace
	$\gamma$ Trace
	$\tau$ Protein
	Myelin basic protein
	Glial fibrillary acid protein
2	100% (approx.)
	Intrathecal synthesis similar to systemic levels
	$\beta_2$ -Microglobulin
	Enolase ( $\gamma$ )
	D-2 antigen (N-CAM protein, a neural cell adhesion molecule)
	Fibrinogen degradation products
3	10–1%
	Partial intrathecal synthesis
	Prealbumin
	Lysozyme
	Eosinophil cationic protein
	Lactoferrin
	Ferritin
4	<1%
	Mainly filtration from plasma – rest of proteins as in

**Table 2**

From Thompson EJ (1988) *The CSF Proteins: A Biochemical Approach*. Amsterdam: Elsevier; © Elsevier.

liquid biopsy of the brain and is, by definition, the 'milieu interieur'; and (2) because it is a kind of brain lymph.

Brain proteins are easily recognized since they typically have at least a fivefold higher concentration in CSF than in serum, or the percentage transfer given in **Table 1** is greater than 500%. Although several laboratories have sought marker proteins that might be specific for individual diseases, this has so far not proved a simple solution with 'one protein—one disease'. What is clear, however, is that many brain proteins can be used as indicators of prognosis in response to acute insults such as stroke or physical trauma. Basically, the higher the level of the brain protein in CSF coupled with the longer it persists at high levels, the worse the prognosis.

By far the most common analyses have been conducted through the use of CSF as lymph fluid. Historically, there has been a great deal of work on antibodies in CSF. With continuing improvements in technology, there has been increasing emphasis placed on the analysis of specific antigens. This formation of antibodies can take some 7–10 days. Analysis for antigens is therefore particularly helpful in making the initial diagnosis. In addition, they can also be monitored sequentially to ascertain the effectiveness of therapy for the antigen in question.

**Table 2** The roster of the CSF proteins

Protein	mg l <sup>-1</sup>	% Transfer	% Total
1 Albumin	200	0.5	67
2 $\beta$ -Trace	26	500	9
3 IgG	22	0.2	7
4 Prealbumin	17	6	6
5 Transferrin ( $\tau \sim 1/3$ )	14	0.6	5
6 $\alpha_1$ -Antitrypsin	8	0.4	3
7 Apo-A-lipoprotein	6	0.4	2
8 $\gamma$ trace	6	500	2
9 Orosomucoid	3.6	0.6	<1
10 Hemopexin	30	0.3	
11 Group components	2.5		
12 Haptoglobin	2.1	0.1	
13 Antichymotrypsin	2.1		
14 $\alpha_2$ -Macroglobulin	2.0	0.2	
15 $\alpha_2$ -Hermann Schultz	1.7	0.3	
16 Complement C'3	1.5	0.3	
17 Complement C'9	1.5	2	
18 Fibrinogen degradation products	1.4	50	
19 IgA	1.3	0.1	
20 $\beta_2$ -Microglobulin	1.0	33	
21 Ceruloplasmin	1.0	0.5	
22 Complement C'4	1.0	0.3	
23 Lysozyme	1.0	7	
24 $\beta_2$ -Glycoprotein I	1.0	0.5	
25 Fibrinogen	0.65	0.02	
26 D-2 antigen (NCAM)	0.64	40	
27 $\beta$ -Lipoprotein	0.59	0.01	
28 Zinc $\alpha_2$ -glycoprotein	0.27		
29 Plasminogen	0.25	0.2	
30 IgM	0.15	0.02	
31 Aldolase C4	0.085		
32 $\alpha_1$ -Microglobulin	0.035	0.1	
33 Enolase ( $\gamma$ )	0.010	100	
34 Lactoferrin	0.0066	2	
35 Glial fibrillary acidic protein	0.0033	500	
36 Ferritin	0.0023	2	
37 S-100	0.0020	500	
38 Creatine kinase-BB	0.0009	500	
39 Eosinophil cationic protein	0.0009	5	
40 Myelin basic protein	0.0006	500	

From Thompson EJ (1988) *The CSF Proteins: A Biochemical Approach*. Amsterdam: Elsevier; © Elsevier.

## Sampling – Conditions for the Collection of CSF

Because CSF is not a homogeneous mixture, being derived from multiple sources, it is particularly important to keep precise track of the volume of CSF that is removed. Given the various gradients in concentration of some analytes, which are known to exist along the neuraxis, one should compare the same 'locus' in specimens from different individuals. In humans, for instance, the first 20 ml of spinal fluid comes from the spinal cord, whereas, say, the 15th

milliliter is derived from near the fourth ventricle. In other words, the latter would be CSF similar in composition to that obtained from a cisternal puncture.

When blood is collected, various anticoagulants can be used, not only to prevent clot formation, but also to stop the cascade of various proteolytic enzymes such as the complement series. If these analytes are to be measured in CSF, then it should also be collected in the presence of anticoagulant.

Cells will normally sediment rapidly from CSF to the bottom of the tube, unlike blood plasma that is much more viscous. Nevertheless, it is important to remember that CSF should be centrifuged if the cells may contain the analyte rather than just the supernatant CSF above the cell pellet. Fragments of brain are found in normal spinal fluid and these should be removed by centrifugation. It may thus be of interest to analyze the contents of the pellet as well as that which remains in solution in CSF. It is particularly important to remove cells from CSF before freezing, since lysis of the cells will result. Another source of difficulty with proteins is the use of techniques that involve concentration of CSF prior to analysis, since the proteins will commonly become denatured; also they may not go back into solution when reconstituted following freeze-drying. Due to the very low protein content of CSF, it may be necessary to add inhibitors of proteolysis. This is necessary, for instance, in the case of cytokines, which are not normally destroyed in the blood plasma due to the high concentrations of protein that provide a kind of buffer against hydrolysis by enzymes. In CSF, by contrast, it is essential to use a protease inhibitor such as aprotinin.

The age of the animal can also have a dramatic effect. The CSF total protein is very high in humans in the first few months of life but then decreases. At about age 60 it begins to increase again. Mobility is important, as the circulation of CSF is aided by movement and thus prolonged immobility, for instance, due to coma, can lead to high levels of total protein in the spinal sac. It is therefore important to keep note of any drugs that can reduce the mobility of the patient, likewise any diseases, including psychological effects such as catatonic schizophrenia, which also give rise to less movement and can thus elevate the spinal total protein level. The concentrations of other analytes that depend upon CSF circulation can also be altered.

## Methods of Analysis

The amounts of protein present (see Table 2) can vary over several orders of magnitude. Thus various

**Table 3** Methods of protein detection in order of increasing sensitivity

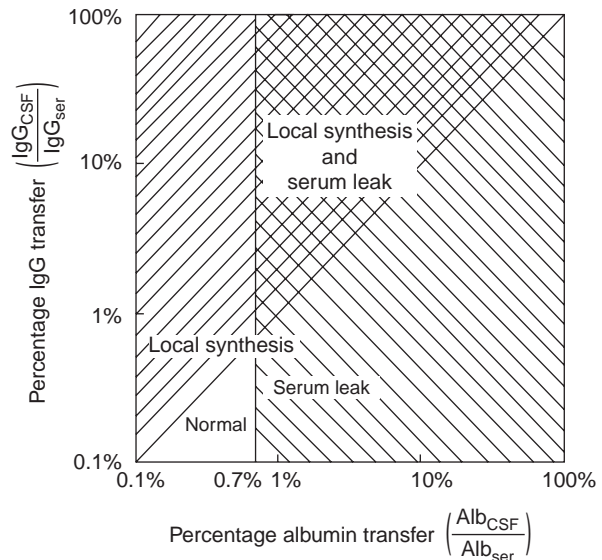
<i>Sensitivity of solid-phase assays</i>	
Stain:	
Naphthalene black	3 µg per band
Coomassie	1 µg per band
Nigrosine	10 ng per band
Silver	10 ng per band
Enzymatic (HRP)	1 ng per band
<i>Sensitivity of immunoassays in solution</i>	
Radial (Mancini)	10 mg l <sup>-1</sup>
Electro (Laurell)	5 mg l <sup>-1</sup>
Nephelometry	1 mg l <sup>-1</sup>
Particle counting	1 µg l <sup>-1</sup>
Enzymatic (ELISA)	1 µg l <sup>-1</sup>
Chemiluminometry	10 pg l <sup>-1</sup>
Radioimmunoassay	10 pg l <sup>-1</sup>

HRP, horseradish peroxidase; ELISA, enzyme-linked immunosorbant assay.

From Thompson EJ (1988) *The CSF Proteins: A Biochemical Approach*. Amsterdam: Elsevier; © Elsevier.

different methods are relevant, depending upon the concentration of the protein in question. The assay can be performed in solution on a total amount of protein, regardless of its physical state, e.g., native versus fragments versus aggregates, all of which are typically found in the CSF of normal as well as in altered proportions in abnormal CSF. Additional information is thus available by fractionation with attendant separation of the different physical forms either on the basis of molecular size (e.g., sodium dodecyl sulfate polyacrylamide gel electrophoresis, SDS-PAGE), molecular charge (e.g., isoelectric focusing), or both (e.g., non-SDS-PAGE). Depending on the concentration and/or amount of CSF available for analysis, one would choose the appropriate method to visualize/estimate (Table 3). The cells of the CSF, which typically amount to less than 5 µl<sup>-1</sup>, must not be neglected. If an abnormality is suspected, a cytocentrifuge is an important investigation. A Gram stain is also essential if bacterial infection is suspected. If the level of total protein is normal, bacterial infection is unlikely. CSF lactate is the most discriminating test for bacterial meningitis, having supplanted the less satisfactory determination of glucose.

Determination of CSF albumin and IgG are typically done in parallel with the serum levels of albumin and IgG. This can indicate whether the various blood–CSF barriers are intact and/or whether there is local synthesis of antibody within the central nervous system. It is recommended that the results should be plotted in a nonlinear fashion (Figure 4), especially for measurements of IgA and IgM, since they are

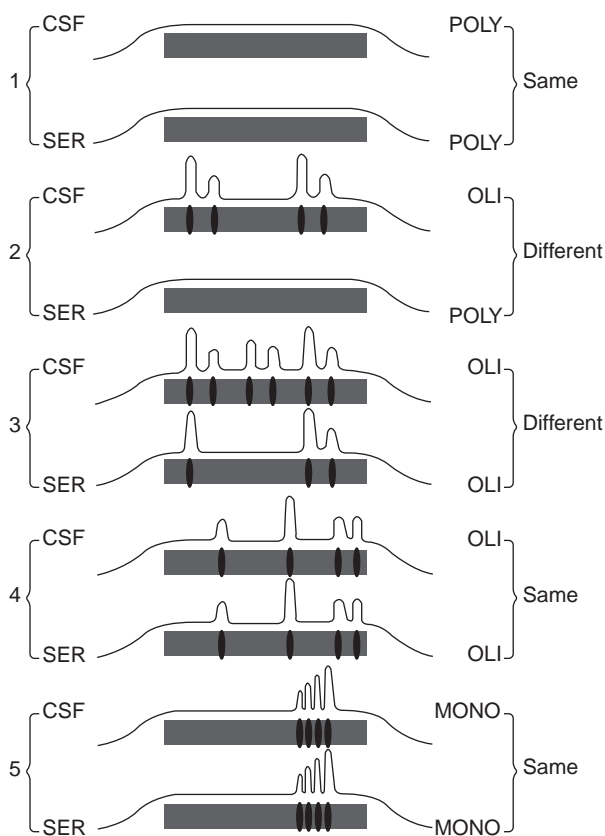
**Figure 4** Use of logarithmic plot to distinguish IgG barrier abnormalities (serum leak) from local synthesis.

much larger proteins than IgG and will therefore show disproportionate leakage with greater destruction of the barriers for the CSF. The most discriminating method (which is also the most widely accepted) for detection of local synthesis of antibodies within the central nervous system is the technique of isoelectric focusing, again of CSF in parallel with serum (Figure 5). Although this is a qualitative test, it is more sensitive than quantitative tests for intrathecal synthesis of IgG, regardless of which mathematical formulation one adopts.

## Applications

Perhaps the most important reason for analyzing CSF is as part of the differential diagnosis of brain inflammation. This inflammation can be either acute or chronic. Probably the most common cause of chronic brain inflammation is multiple sclerosis. About 90–95% of patients with multiple sclerosis will have a typical pattern of immunoglobulin bands called ‘oligoclonal’ in their CSF, which are not found in the corresponding serum when analyzed by isoelectric focusing. As there are various treatments for multiple sclerosis, these can be monitored by longitudinal lumbar puncture to ascertain whether the brain lymphocytes have been suppressed in their production of IgG.

In acute inflammatory diseases, such as bacterial meningitis, the prognosis is inversely correlated with the time required to make the diagnosis. Bacterial meningitis is thus an acute medical emergency.



**Figure 5** Representation of the five patterns of CSF IgG following isoelectric focusing. Densitometry scans show relative proportions of different bands in CSF and parallel serum specimens. Different CSF/SER patterns denote local IgG synthesis. CSF, cerebrospinal fluid; SER, serum; POLY, polyclonal; OLI, oligoclonal; MONO, monoclonal. Cathode is on the right.

The physician should hold the fluid up to the daylight to see if it is turbid (Tyndall effect). If so, antibiotic therapy should be immediately considered. The CSF should then be examined within a short time by microscopy to look for the presence of polymorphs. A Gram stain for bacteria should also be done. Apart from bacteria, acute inflammation can also be due to a number of different viruses that can cause meningitis or encephalitis. Herpes simplex is a DNA virus and herpes simplex infection is treatable using Acyclovir, a drug that acts by blocking replication of the virus. The diagnosis of herpes simplex infection can be made using the polymerase chain reaction. Human immunodeficiency virus (HIV) is a neurotropic virus that primarily enters the microglial cells of the brain. Since HIV infection is associated with immunodeficiency, there are a number of intercurrent infections that tend to appear in the brains of these patients and these can be detected by looking for the appropriate antigens within their CSF. Since the HIV is well known to cause immunoparesis,

traditional diagnostic tests based upon antibody titers may be less useful than antigen tests.

There are many more applications that involve the analysis of CSF but, because of limitations of space, only a selected few of the more pressing applications have been presented. There are entire books devoted to the cells, proteins, and individual brain diseases. For the most recent studies and new markers of disease, see the two texts written in 1999 by Felgenhauer and by Thompson.

## Final Interpretation

The first decision to make is whether the analyte in question is actively pumped or whether it diffuses passively. There is always the caveat of any possible rostro-caudal gradients, in which case the volume removed is important and it is always prudent to measure the levels of analyte in blood plasma (or serum). For proteins with passive diffusion one must correct simultaneously for both the serum level and for an independent measure of barrier function. The quotient of CSF analyte/serum analyte divided by CSF albumin/serum albumin can be expressed as an 'index' value, but if there are alterations in any of the barriers there may be a nonlinear relationship that is particularly characteristic of large molecular weight proteins. This relationship is thus best expressed by a nonlinear function. The golden rule must be to compare like with like, and make appropriate corrections for known differences due to inhomogeneous samples.

**See also:** **Blood and Plasma. Clinical Analysis:** Sample Handling. **Electrophoresis:** Isoelectric Focusing. **Proteins:** Physiological Samples.

## Further Reading

- Felegenhauer K and Beuche W (1999) *Labordiagnostik neurologischer Erkrankungen*. Stuttgart: George Thieme Verlag.
- Fishman RA (1992) *Cerebrospinal Fluid in Diseases of the Nervous System*. Philadelphia: W B Saunders Co.
- Reiber HO and Lange P (1991) Quantification of virus-specific antibodies in cerebrospinal fluid and serum: sensitive and specific detection of antibody synthesis in brain. *Clinical Chemistry* 37: 1153–1160.
- Thompson EJ (1999) Neurochemistry. In: Baynes F and Dominiczak MH (eds.) *Medical Biochemistry*, ch. 37, pp. 487–495. London: Mosby.
- Thompson EJ and Keir G (1990) Laboratory investigation of cerebrospinal fluid proteins. *Annals of Clinical Biochemistry* 27: 425–435.

## CERTIFIED REFERENCE MATERIALS

See **QUALITY ASSURANCE: Reference Materials; Production of Reference Materials**

## CHARGE COUPLED DEVICES

See **OPTICAL SPECTROSCOPY: Detection Devices**

## CHEMICAL OXYGEN DEMAND

See **WATER ANALYSIS: Chemical Oxygen Demand**

## CHEMICAL WARFARE AGENTS

**P A D'Agostino**, Defence R&D Canada Suffield, Medicine Hat, AB, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Chemical warfare agents are a group of toxic chemicals that have been defined in the Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and their Destruction (commonly referred to as the Chemical Weapons Convention or CWC) as 'any chemical which through its chemical effect on life processes can cause death, temporary incapacitation or permanent harm to humans or animals...'. Poisonous or toxic compounds have been utilized in an effort to gain military superiority throughout history but it is only during the past century that chemical warfare agents have been produced and used on a large scale. Tear gas grenades were used in 1914 by the French at the outbreak of the First World War, but it was not until the Germans first used chlorine near Ypres in 1915 that the world entered the modern era of chemical warfare. Other chemical warfare agents such as phosgene and mustard were weaponized during the First World War and were used by both sides throughout the conflict.

The use and development of chemical warfare agents continued following the First World War

despite the signing of the 1925 Geneva Protocol, which bans the first use of chemical weapons. Mustard was used by the Italians against the Abyssinians (Ethiopia) during the 1936–37 war and just prior to the Second World War, the Germans discovered and produced the first nerve agent, tabun. Tabun was weaponized by the Germans but neither side made use of their chemical weapons stocks. More effective nerve agents, such as VX, were developed in the 1950s, mustard was used in the Yemen War (1963–67) and allegations of chemical warfare agent use were reported in South East Asian conflicts. Nerve and mustard agents were used by Iraq in the 1980s war between Iran and Iraq, and were considered a real threat to United Nations armed forces during their action against Iraq (1990–91). More recently, sarin and mustard were collected in 1992 from a site where chemical weapons were thought to have been previously used against the population of a Kurdish village. Most recently, sarin was released by the Aum Shinrikyo cult in the Tokyo underground transit system (1995) resulting in thousands seeking medical attention and 12 deaths.

After considerable effort the CWC was opened to signature in 1993, with the treaty coming into force on April 29, 1997. More than 160 States Parties have ratified the CWC and agreed not to develop, produce, stockpile, transfer, or use chemical weapons and agreed to destroy their own chemical weapons

and production facilities. A strong, compliance monitoring regime involving site inspections was built into the CWC to ensure that the treaty remains verifiable. The Organisation for the Prohibition of Chemical Weapons, or OPCW, based in The Hague, has responsibility for implementation of the treaty. Routine OPCW inspections have or will take place at declared sites, including small-scale production, storage and destruction sites, and challenge inspections will take place at sites suspected of noncompliance. Proliferation of chemical weapons and their use will hopefully decrease over the coming years as the CWC proceeds toward its goal of worldwide chemical weapons destruction.

Concerns over possible terrorist use, continued interest by the defense community and the requirements of a verifiable CWC, have driven the development and application of analytical methods for the detection, characterization, and confirmation of chemical warfare agents. Analytical techniques play an important role in this process as sampling and analysis will be conducted to ensure treaty compliance, to investigate allegations of use, and to verify the use of these weapons for forensic purposes.

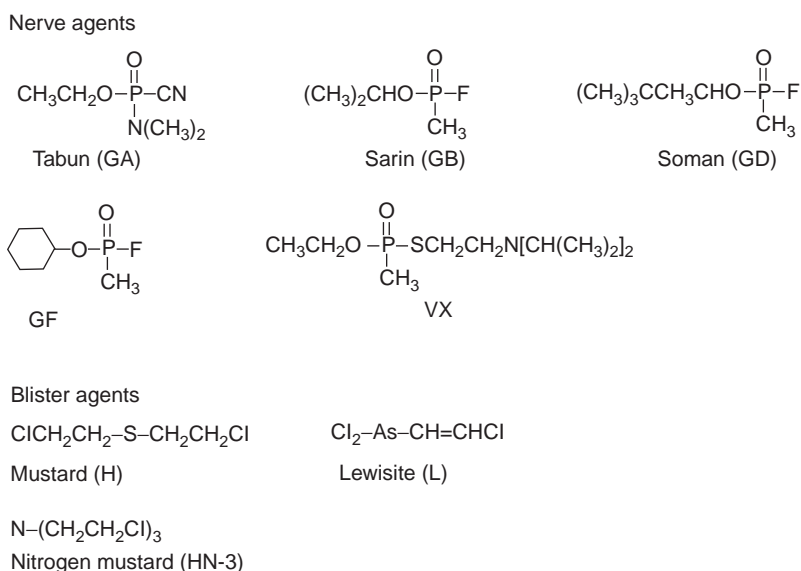
## Chemical Warfare Agent Categories

Chemical warfare agents have been classified into nerve, blister, choking, vomiting, blood, tear, and incapacitating agent categories based on their effect on humans. The most significant chemical warfare agents in terms of military capacity and past use are the nerve and blister agents. For these reasons the analysis of these compounds will be emphasized over the other groups. The choking, blood, and vomiting agents are for the most part obsolete chemical agents that were employed during the First World War. The tear agents were used during the Vietnam War but their primary use, because of their inability to produce high casualties, remains in riot control and for the training of military personnel in chemical defense. Incapacitating agents have been included in the CWC as the United States did develop an agent in this category.

The compounds listed in Table 1 represent the most common chemical warfare agents, by category with their Chemical Abstracts registry numbers, and is not intended to be exhaustive. It has been estimated that more than 10 000 compounds are controlled

**Table 1** Common chemical warfare agents

Full name (trivial name(s))	CA no.
<i>Nerve (react irreversibly with cholinesterase which results in acetylcholine accumulation, continual stimulation of the body's nervous system, and eventual death)</i>	
1-Methylethyl methylphosphonofluoridate (sarin, GB)	107-44-8
1,2,2-Trimethylpropyl methylphosphonofluoridate (soman, GD)	96-64-0
Cyclohexyl methylphosphonofluoridate (GF)	329-99-7
Ethyl dimethylphosphoramidocyanidate (tabun, GA)	77-81-6
O-Ethyl S-(2-diisopropylaminoethyl) methylphosphonothiolate (VX)	50782-69-9
<i>Blister (affect the lungs, eyes, and produces skin blistering)</i>	
Bis(2-chloroethyl)sulfide (mustard, H)	505-60-2
Bis(2-chloroethylthio)ethane (sesquimustard, Q)	3563-36-8
Bis(2-chloroethylthioethyl)ether (T)	63918-89-8
Tris(2-chloroethyl)amine (HN-3)	555-77-1
(2-Chloroethenyl)arsonous dichloride (lewisite, L)	541-25-3
<i>Choking (effects respiratory tract and lungs)</i>	
Chlorine	7782-50-5
Carbonic dichloride (phosgene, CG)	75-44-5
<i>Vomiting (causes acute pain, nausea, and vomiting in victims)</i>	
Diphenylarsinous chloride (DA)	712-48-1
10-Chloro-5,10-dihydrophenarsazine (adamsite, DM)	578-94-9
Diphenylarsinous cyanide (DC)	23525-22-6
<i>Blood (prevents transfer of oxygen to the body's tissues)</i>	
Hydrogen cyanide (HCN, AC)	74-90-8
<i>Tear (causes tearing and irritation of the skin)</i>	
[(2-Chlorophenyl)methylene]propanedinitrile (CS)	2698-41-1
2-Chloro-1-phenylethanone (CN)	532-27-4
Dibenz[b,f][1,4]oxazepin (CR)	257-07-8
<i>Incapacitating (prevents normal activity by producing mental or physiological effects)</i>	
3-Quinuclidinyl benzilate (BZ)	6581-06-2



**Figure 1** Structures of common chemical warfare agents.

under the CWC, although in practical terms the actual number of chemical warfare agents, precursors, and degradation products that are contained in the OPCW database is in the hundreds. The structures of some common nerve and blister chemical warfare agents are illustrated in **Figure 1**.

## Sample Handling

Samples contaminated with chemical warfare agents generally fall into one of the following general categories: (1) munitions or munition fragments (e.g., neat liquid or artillery shell casing) or (2) environmental (e.g., soil, water, vegetation, or air samples), man-made materials (e.g., painted surfaces or rubber), and biological media (e.g., blood or urine). The ease of analysis depends on the amount of sample preparation required to obtain a suitable sample or extract for chromatographic analysis. In the simplest case where neat liquids can be obtained the sample typically only requires dilution with a suitable solvent prior to analysis. Aqueous samples may be analyzed directly or following solvent extraction with an organic solvent, while biological fluids typically require extensive sample handling and/or derivatization prior to analysis. Soil and other solid samples generally require, at a minimum, solvent extraction and concentration prior to analysis.

Soil may be collected following chemical warfare agent contamination and is one of the more commonly analyzed media. Extraction of chemical warfare agents from soil samples may be accomplished using a number of solvents (e.g., hexane,

dichloromethane, or water), with dichloromethane being the most commonly employed extraction solvent for chemical warfare agent determinations. A small portion of soil sample (e.g., 2 g) would typically be ultrasonically extracted with dichloromethane (e.g., 4 ml) in screw-capped glass vial or culture tube for ~10 min. The fines may be removed from the dichloromethane extract by centrifugation or filtration prior to concentration (if necessary) and analysis by gas chromatography–mass spectrometry (GC–MS) or another analytical technique. The sample handling procedures developed for contaminated soils can usually be extended to other solid samples with minor modifications to extraction solvent, volume, or vessel.

## Identification Methods

Chemical warfare agents have often been referred to as warfare gases and, in the military, the phrase ‘gas, gas, gas’ has become synonymous with attack by chemical warfare agents. In fact, many chemical warfare agents exist as liquids at ambient temperatures but have varying degrees of volatility and pose a significant vapor hazard as well as a liquid contact hazard. This physical characteristic has made the analysis of chemical warfare agents amenable to the analytical techniques commonly employed for most environmental analyses, namely GC with a variety of detectors including MS. Synthetic or relatively pure samples are typically characterized by nuclear magnetic resonance (NMR) or Fourier transform infrared (FTIR) spectroscopy. Liquid chromatography–mass spectrometry (LC–MS) has been



used with increasing regularity for the detection and confirmation of chemical warfare agents and their nonvolatile degradation products in aqueous samples and extracts. Thin-layer chromatography (TLC) methods have been thoroughly investigated but have been largely superseded by GC and LC.

The OPCW inspectorate, an important end user of analytical techniques for chemical warfare agents, requires the use of two or more spectrometric techniques and the availability of authentic reference standards for the unambiguous identification of these controlled compounds. For this reason, the combined use of GC-FTIR has received increased attention as newer technologies have led to detection limits approaching those routinely reported during GC-MS analysis. For analyses involving low levels of chemical warfare agents in the presence of high levels of interfering chemical background, tandem mass spectrometry (MS/MS) is often employed.

### Chromatography

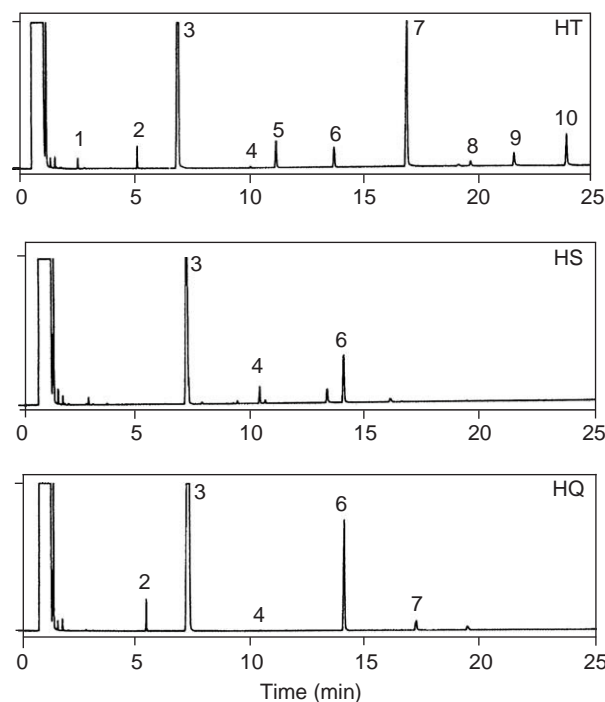
Samples contaminated with chemical warfare agents typically contain multiple components that are best characterized following chromatographic separation. TLC was routinely employed for the detection of chemical warfare agents but with the advent of GC this technology has been used less frequently for analytical applications. Work continues in the area of two-dimensional overpressured TLC with applications being reported for nerve agents. TLC methods have also been proposed for rapid field analyses, but at present this technique sees most application in support of synthetic programs for the isolation of pure materials or as a quick screening procedure.

Capillary column GC remains the most frequently employed analytical separation method for the screening of samples contaminated with chemical warfare agents. Separation of chemical warfare agents may be achieved with many of the commercially available fused silica columns coated with polysiloxane or other films and retention index data relative to *n*-alkanes and alkyl bis(trifluoromethyl)phosphine sulfides (M-series) have been reported for many chemical warfare agents and related compounds. In general, the best separations have been achieved with a moderately polar film such as (86%)-dimethyl-(14%)-cyanopropylphenyl-polysiloxane. Chiral stationary phases have been developed for the resolution of stereoisomers of several chiral nerve agents, most notably soman. The use of multiple columns of differing polarity during one analysis has also been successfully employed during chemical warfare agent analysis and the term

'retention spectrometry' was coined to describe this technique.

Most of the GC detectors commonly applied to pesticide residue analysis have also been applied to the screening of samples for chemical warfare agents with detection limits typically being in the nanogram to picogram range. Flame ionization detection (FID) is routinely used for preliminary analyses as this technique provides a good indication of the complexity of a sample extract. **Figure 2** illustrates typical GC-FID chromatographic separations obtained for three different munitions-grade mustard formulations, HT, HS, and HQ. Mustard comprised 54%, 74%, and 82% of the volatile organic content in HT, HS, and HQ, respectively, based on peak area measurements. The longer chain blister agents, sesquimustard (Q) and bis[(2-chloroethylthio)ethyl]ether (T) were significant in all three samples along with a number of other related compounds that may provide synthetic procedure or source information.

The need for higher specificity and sensitivity has led to the application of element-specific detectors



**Figure 2** Capillary column GC-FID chromatograms of three munitions-grade mustard samples: HT (top), HS (middle), and HQ (bottom). Identified compounds include: (1) 1,4-thioxane, (2) 1,4-dithiane, (3) mustard (H), (4) bis(2-chloroethyl)disulfide, (5) 2-chloroethyl (2-chloroethoxy)ethyl sulfide, (6) sesquimustard (Q), (7) bis(2-chloroethylthioethyl)ether (T), (8) 1,14-dichloro-3,9-dithia-6,12-dioxatetradecane, (9) 1,14-dichloro-3,6,12-trithia-9-oxatetradecane, and (10) 1,16-dichloro-3,9,15-trithia-6,12-dioxaheptadecane. (GC conditions: 15 m  $\times$  0.32 mm ID J&W DB-1; 50°C (2 min) 10°C min<sup>-1</sup> 280°C (5 min).)

such as flame photometric detection, thermionic detection, atomic emission, and electron capture detection. The simultaneous use of FID with one or more element specific detectors has also been demonstrated during dual- or tri-channel GC analysis using conventional and thermal desorption sample introduction. While data obtained with these detectors may provide strong collaborative evidence for the presence of chemical warfare agents, they cannot be used for full confirmation. Use of GC with one or more spectrometric technique such as MS is required to confirm the presence of chemical warfare agents.

LC-ESI-MS is being used increasingly, as electrospray mass spectrometric data may be used to directly identify chemical warfare agents, degradation products, and related compounds in collected aqueous samples or extracts. Both the nerve and blister agents undergo hydrolysis in the environment and methods are required under the CWC for retrospective detection and confirmation of these compounds. These compounds are significant as they would not be routinely detected in environmental samples and their identification strongly suggests the prior presence of chemical warfare agents. The degradation products of the chemical warfare agents, in particular the nerve agents, are nonvolatile hydrolysis products that must be derivatized prior to GC analysis. A variety of derivatization reagents, leading to the formation of pentafluorobenzyl, methyl, *tert*-butyldimethylsilyl, and trimethylsilyl ethers (or esters), have been investigated to allow GC analysis of, in particular, the organophosphorus acids related to the nerve agents (e.g., alkyl methylphosphonic acids and methylphosphonic acid).

### Mass Spectrometry

Mass spectrometry is the method of choice for the detection and characterization of chemical warfare agents, their precursors, degradation products, and related compounds. Extensive use has been made of GC-MS and the mass spectra of numerous chemical warfare agents and related compounds have been published, with the most common chemical warfare agent mass spectra being available in the OPCW, commercial, or defense community databases.

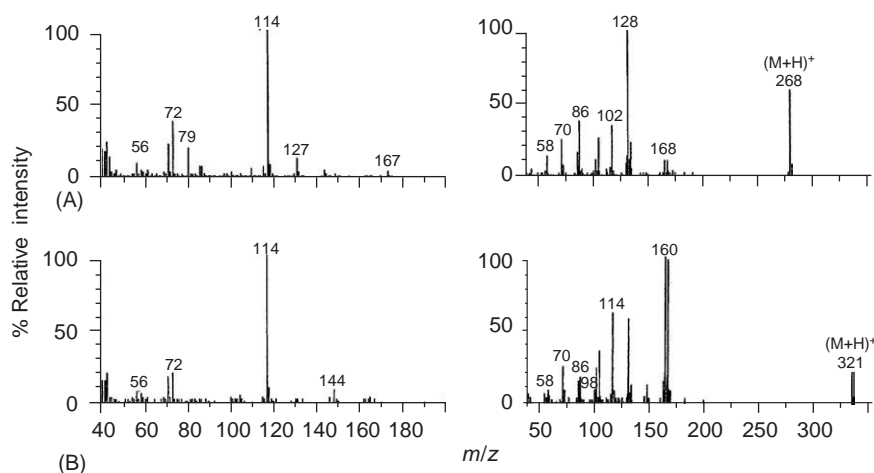
Most of these data were obtained under electron impact (EI) ionization conditions. However, many of the chemical warfare agents, in particular the organophosphorus nerve agents and the longer chain blister agents related to mustard, such as T, do not provide molecular ion information under EI-MS. This hinders confirmation of these chemical warfare agents and makes identification of novel chemical warfare agents or related impurities difficult.

Considerable effort has been devoted to the use of chemical ionization (CI) as a complementary ionization technique. This milder form of ionization generally affords molecular ion information for the chemical warfare agents and has been used extensively for the identification of related compounds or impurities in chemical warfare agent munitions samples and environmental sample extracts. The identity of these related compounds is important because the origin of samples, synthetic process information, or degree of degradation (weathering) information may aid OPCW or other investigations.

Isobutane, ethylene, and methane gases were initially demonstrated as suitable CI gases for the acquisition of organophosphorus nerve agent CI-MS data. The efficacy of ammonia CI-MS for organophosphorus nerve agents and related compounds has been demonstrated and many laboratories now employ this complementary confirmation technique. Ammonia CI not only offers abundant molecular ion data but also affords a high degree of specificity as less basic sample components are not ionized by the ammonium ion. Additional data may be obtained through the use of deuterated ammonia CI, as this technique provides useful hydrogen/deuterium exchange data that indicates the presence of exchangeable hydrogen(s) in CI fragmentation ions. Finally, for full confirmation, the acquired EI and CI mass spectrometric data should be compared to authentic reference data obtained under identical experimental conditions.

Figure 3 illustrates EI and ammonia CI data obtained for VX and a significant VX degradation product, bis[2-(diisopropylamino)ethyl] disulfide. The acquired EI data for both compounds, as well as other VX related compounds, are remarkably similar. Both compounds lack a molecular ion and contain a base ion at  $m/z$  114 due to  $(\text{CH}_2\text{N}(\text{iPr})_2)^+$  and additional ions related to the  $-\text{SC}_2\text{H}_4\text{N}(\text{iPr})_2$  substituent. Under ammonia CI conditions, mass spectra containing pseudomolecular and CI fragmentation ions were acquired, with these data being used to confirm molecular mass and differentiate VX related compounds that exhibit similar EI data.

Capillary column GC-MS/MS offers the analyst the potential for highly specific, sensitive detection of chemical warfare agents as this technique significantly reduces the chemical noise associated with complex biological or environmental sample extracts. The specificity of product scanning with moderate sector resolution, as well as the specificity of ammonia CI, was demonstrated with a hybrid tandem mass spectrometer during analysis of painted panel samples circulated during an international round robin verification exercise.



**Figure 3** EI (left) and ammonia CI (right) mass spectrometric data obtained for (A) VX and (B) bis[2-(diisopropylamino)ethyl] disulfide.

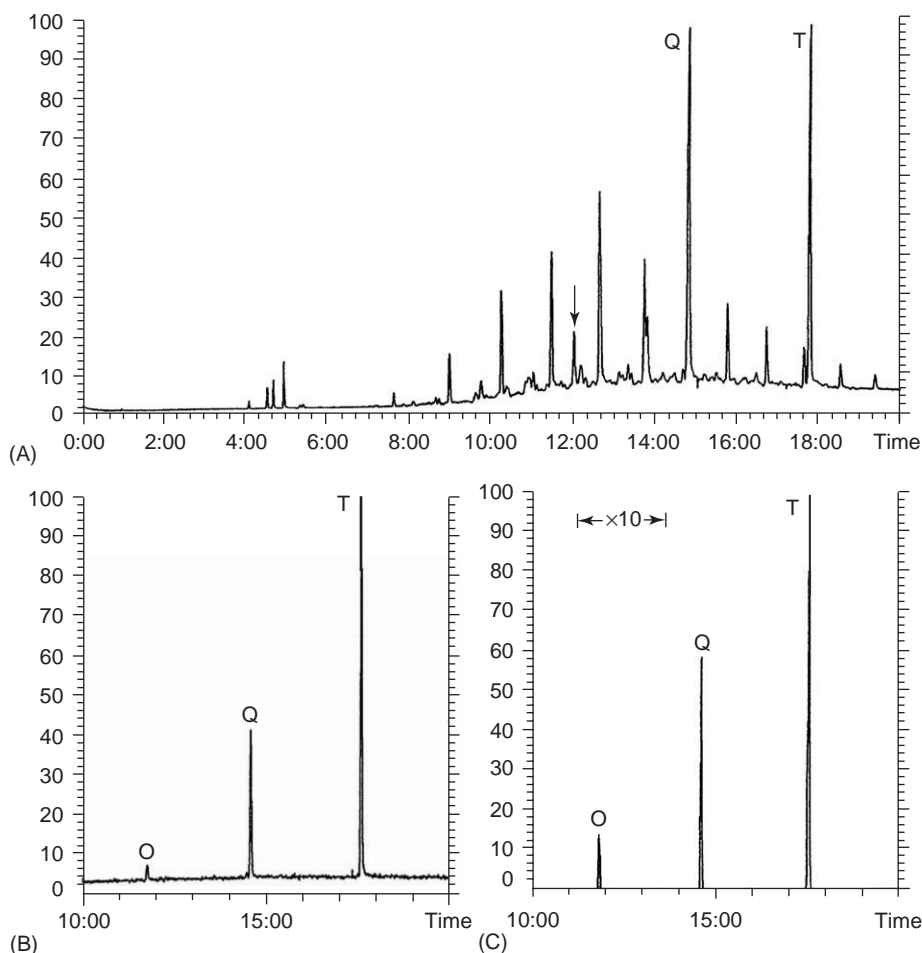
The painted panel extract was contaminated with numerous hydrocarbons and only two of the three longer chain blister agents, sesquimustard (Q) and bis(2-chloroethylthioethyl)ether (T), could be identified during capillary column GC–MS (EI) analysis (Figure 4A). The arrow indicates the chromatographic retention time of the third blister agent, 2-chloroethyl (2-chloroethoxy)ethyl sulfide (O). The specificity of ammonia CI (Figure 4B) was clearly demonstrated during this analysis. All three longer chain blister agents were determined in the presence of high levels of interfering hydrocarbons, as the hydrocarbons were not sufficiently basic to ionize. Similarly, it was possible to use the resolution of hybrid MS/MS to discriminate between ions at  $m/z$  123 arising from the longer chain blister agents from those ions at  $m/z$  123 arising from the hydrocarbon background. The resultant GC–MS/MS chromatogram (Figure 4C), where only  $m/z$  123 ions due to the blister agents were transmitted into the collisional activated dissociation cell, was virtually free of chemical noise and all three components were detected. The three longer chain blister agents were well resolved with the J&W DB-1701 capillary column, with all three components exhibiting similar product spectra during GC–MS/MS analysis.

Both the nerve and blister agents undergo hydrolysis in the environment and methods are required for retrospective detection and confirmation of these hydrolysis products. Hydrolysis products are significant as they are generally compounds that would not be routinely detected in environmental samples and their presence strongly suggests the prior presence of chemical warfare agents. The degradation products of the chemical warfare agents, in particular the nerve agents, are nonvolatile hydrolysis products that

must be derivatized prior to GC analysis. Alternatively, aqueous samples or extracts may be analyzed by LC–MS, negating the need for additional sample handling steps and derivatization.

Use of thermospray MS and more recently the atmospheric pressure ionization (API) (e.g., electrospray (ESI), ionspray, and atmospheric pressure CI) techniques have enabled the direct mass spectrometric analysis of the hydrolysis products of chemical warfare agents. Both techniques may be interfaced to liquid chromatography for component separation, with thermospray having been largely superseded by API for most LC–MS applications. LC–ESI–MS methods have been used for the direct analysis of chemical warfare agent hydrolysis products in a number of studies and have recently been demonstrated for the analysis of nerve agents. These new methods complement existing GC–MS methods for the analysis of chemical warfare agents and their hydrolysis products and LC–ESI–MS methods will replace some GC–MS methods used for the analysis of contaminated aqueous samples or extracts.

Mustard and longer chain blister agents hydrolyze to their corresponding diols, with thiodiglycol being the product formed following hydrolysis of mustard. Figure 5A illustrates a typical LC–ESI–MS chromatogram obtained for the aqueous extract of a soil sample taken from a former mustard storage site. The soil sample extract contained thiodiglycol (Figure 5B) and 6-oxa-3,9-dithia-1,11-undecanediol (Figure 5C), the hydrolysis products of blister agents mustard and bis(2-chloroethylthioethyl)ether, respectively. ESI–MS data for both compounds contained protonated molecular ions that could be used to confirm molecular mass and characteristic lower mass product ions.

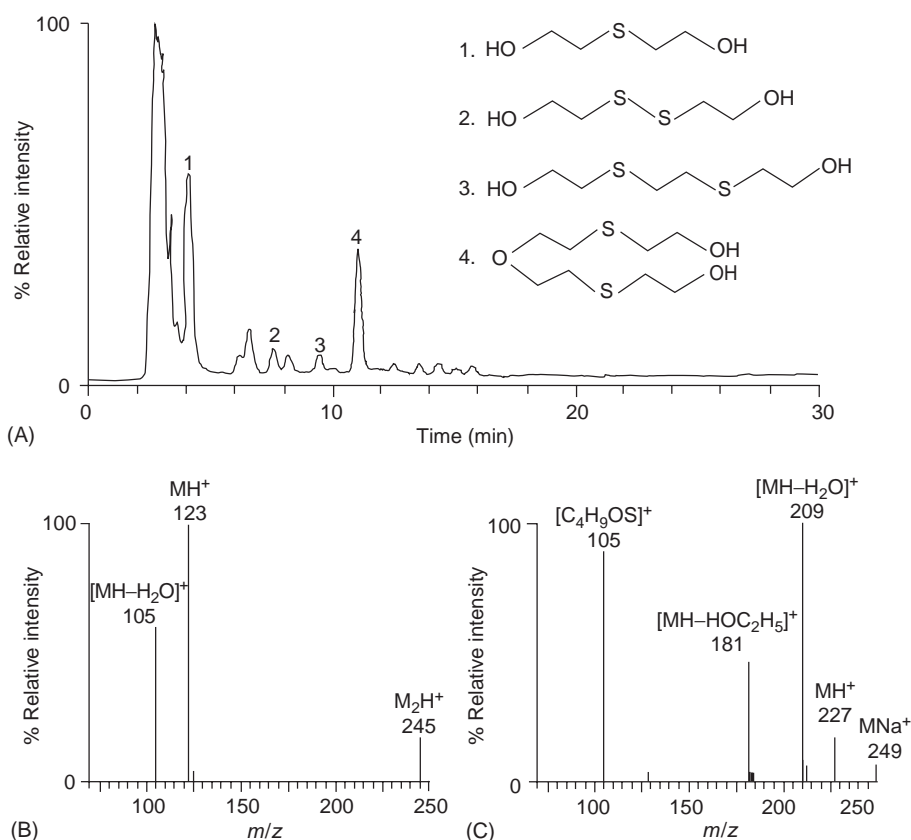


**Figure 4** Capillary column (A) GC–MS (EI); (B) GC–MS (ammonia CI); and (C) GC–MS/MS (EI) chromatograms obtained during analysis of international round robin painted panel extracts. Sequimustard (Q) and bis(2-chloroethylthioethyl)ether (T) were detected during EI analysis. The downward arrow in (A) indicates the retention time of 2-chloroethyl (2-chloroethoxy)ethyl sulfide (O). This compound was masked by the sample matrix during EI analysis and was only detected following (B) ammonia CI and (C) MS/MS analysis. (GC conditions: 15 m  $\times$  0.32 mm ID J&W DB-1701, 40°C (2 min) 10°C min<sup>-1</sup> 280°C (5 min), x-axis: time (min).)

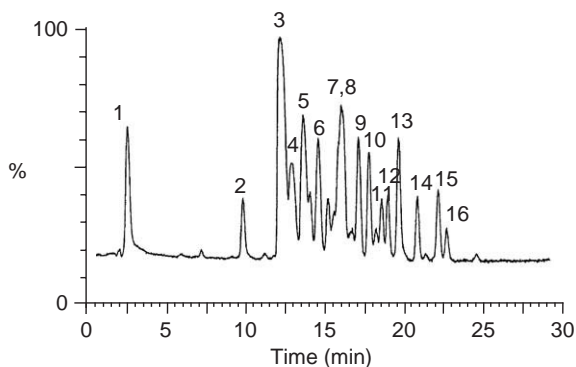
**Figure 6** illustrates the LC–ESI–MS chromatogram for a complex munitions-grade tabun sample. Tabun and a number of related compounds were identified based on their acquired ESI–MS data. The mass spectra contained  $(M + H)^+$ ,  $(M + H + ACN)^+$  ions and/or protonated dimers that could be used to confirm the molecular mass of each compound. Structural information was provided by inducing product ion formation in either the ESI interface or the quadrupole collisional cell of a MS/MS instrument. Product ions due to alkene loss from the alkoxy substituents, and the acetonitrile adduct associated with these product ions, were generally observed. **Figure 7** illustrates typical ESI–MS data obtained for tabun and three other nerve agents.

Considerable effort has been expended on the development of field portable MS and GC–MS

instruments, as this technique holds the greatest promise for the confirmation of chemical warfare agents under field situations. The OPCW has available field portable GC–MS instrumentation that may be taken onsite to confirm the presence of chemical warfare agents. An atmospheric pressure MS/MS has also been developed and evaluated for real-time detection of nerve agents in air. Alternatively, air samples may be collected on solid-phase microextraction fibers or on Tenax tubes that may be thermally desorbed into an onsite GC–MS instrument. Secondary ion mass spectrometry has been used for the detection of chemical warfare agents and their hydrolysis products on leaves, soil, and concrete, offering a new option for the detection of these compounds on adsorptive surfaces. Finally, rapid separation and detection of chemical warfare agents has recently been



**Figure 5** (A) Packed capillary LC-ESI-MS chromatogram obtained for the water extract of a soil sample obtained from a former mustard site. ESI-MS data obtained for (B) thiodiglycol (sampling cone voltage: 20 V) and (C) 6-oxa-3,9-dithia-1,11-undecanediol (sampling cone voltage: 30 V). (LC conditions: 150 mm  $\times$  0.32 mm ID  $C_{18}$ , acetonitrile/water gradient.).

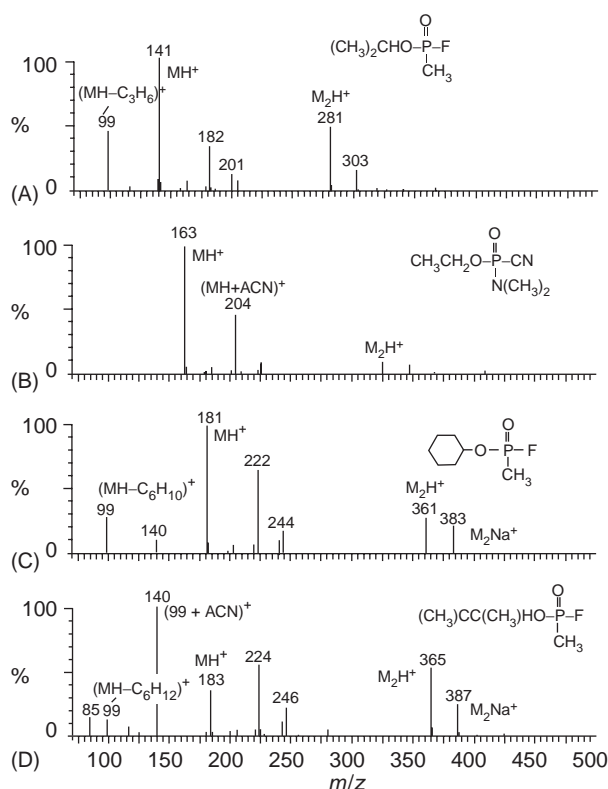


**Figure 6** Packed capillary LC-ESI-MS chromatogram obtained for 0.1 mg  $ml^{-1}$  munitions-grade tabun sample. Tabun (peak number 4) and 15 related organophosphorus compounds were identified by ESI-MS. (LC conditions: 150 mm  $\times$  0.32 mm ID  $C_{18}$ , acetonitrile/water gradient.)

demonstrated with ESI-ion mobility spectrometry (IMS)-MS. IMS is commonly employed in military devices for rapid field detection and this approach could lead to the development of instrumentation for the analysis of aqueous samples.

### Other Techniques

NMR is an important technique for the structural analysis and characterization of chemical warfare agents, particularly for the authentication of reference materials or unknown chemical warfare agents and related compounds. The presence of heteronuclei such as  $^{31}P$  and  $^{19}F$  in the nerve agents leads to diagnostic splitting patterns and coupling constants due to  $^1H$ - $^{31}P$  and  $^1H$ - $^{19}F$  spin-spin coupling. The utility of NMR for analysis of complex sample mixtures or for trace analysis is somewhat limited. Specific heteronuclear experiments such as  $^{31}P$  NMR may be used to identify organophosphorus nerve agents in complex matrices. Characteristic chemical shifts of compounds containing a phosphorus-carbon bond and splittings due to phosphorus-fluorine spin-spin coupling can be used to screen for the presence of nerve agents (Table 2). However,  $^{31}P$  chemical shifts are sensitive to temperature, concentration, and solvent and the identification must be supported with additional spectrometric data such as MS. Two-dimensional correlation experiments have been used to help in structural elucidation of



**Figure 7** ESI-MS data obtained for (A) sarin (GB); (B) tabun (GA); (C) cyclohexyl methylphosphonofluoridate (GF), and (D) soman (GD) with a sampling cone voltage of 20 V.

**Table 2** Phosphorus-31 chemical shifts of nerve agents in deuteriochloroform

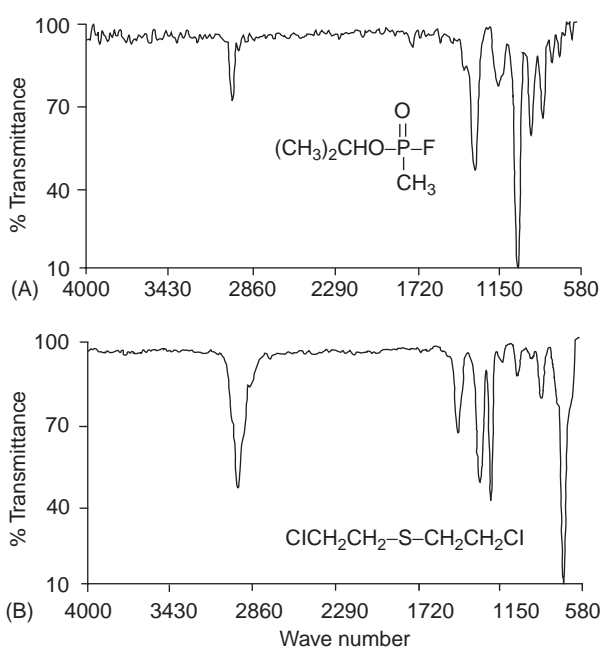
Nerve agent	$\delta^{31}\text{P}$ (ppm) <sup>a</sup>	$J_{\text{PF}}$ (Hz) <sup>b</sup>
VX	53.9	—
Soman (two isomers)	29.1	1047
	28.1	1047
GF	28.5	1047
Sarin	28.4	1046
Tabun	−9.8	—

<sup>a</sup>Chemical shift relative to an internal reference standard of triethyl phosphate, −1 ppm.

<sup>b</sup>Spin-spin ( $^{31}\text{P}$ – $^{19}\text{F}$ ) coupling constants.

unknowns in contaminated samples, making NMR a valuable technique to be used alongside other spectrometric techniques.

Condensed phase infrared (IR) data exist for many chemical warfare agents and related compounds as this technique was routinely used prior to the advent of GC–MS. Capillary column GC–FTIR offers considerably more promise for the identification and characterization of chemical warfare agents in multiple component sample extracts and has been utilized as a complementary confirmation technique.



**Figure 8** Vapor phase FTIR spectra obtained for (A) sarin (GB) and (B) mustard (H) during capillary column GC–FTIR analysis.

Sensitivity is generally poorer than that obtained by MS but may be improved by using large volume (e.g., 50  $\mu\text{l}$ ) injections with peak compression onto an uncoated precolumn with lightpipe technology or through the use of cryodeposition.

Figures 8A and 8B illustrates the FTIR vapor phase data obtained for sarin and mustard, respectively. Sarin exhibits characteristic absorption bands at 2991, 1313, 1018, 924, and 843  $\text{cm}^{-1}$  due to C–H, P–F or P=O, P–O–C, P–CH<sub>3</sub>, and P–F, respectively. Mustard contains bands at 2969 and 717  $\text{cm}^{-1}$  that can be assigned to C–H and C–Cl, respectively. The spectra for other chemical warfare agents differ sufficiently such that library searching may be routinely employed for tentative identification.

## Military Detection

A variety of detection devices and other chemical warfare agent defense equipment have been developed for specific military applications. Most of the effort in this area resulted from the perceived threat during the Cold War era and although this threat has decreased dramatically, interest in chemical detection equipment persists because of worldwide chemical weapons proliferation. During the 1990–91 Iraq War chemical detection equipment was deployed into the

Persian Gulf and similar equipment has been used to support the United Nations Special Commission during the destruction of Iraqi chemical weapons. Equipment of this type has been used by the OPCW and could potentially be utilized again by the United

Nations in peacekeeping or intervention roles where the threat of chemical weapons use exists. **Table 3** lists examples of chemical detection equipment by country and indicates the principle of detection and capabilities of each system.

**Table 3** Selected military chemical warfare agent detection devices

<i>Country</i>	<i>Device name and capabilities</i>
Canada	Chemical Agent Detection System (CADS II) – Early warning system that controls a network of Chemical Agent Monitors (see UK) for the real time detection of nerve and blister agents
China	Chemical Warfare Agent Identification Kit, M-75 – Wet chemistry detection of nerve, blister, choking, vomiting, and blood agents
Denmark	INNOVA 1312 Multi-Gas Monitor – Photoacoustic detection of nerve, blister, choking, and blood agents
Finland	Chemical Agent Detection System, M90 – Alarm for the ion mobility spectrometric detection of nerve and blister agents
France	PROENGIN Portable Chemical Contamination Monitor AP2C – Hand-held flame photometric detection of nerve and blister agents – Also designs for fixed sites (AP2C-V and ADLIF)
Germany	MM-1 Mobile Mass Spectrometer – Quadrupole mass spectrometric detection of chemical warfare agents Rapid Alarm and Identification Device – 1 (RAID-1) – Ion mobility spectrometric detection of nerve and blister agents
Hungary	Chemical Agent Sensor GVJ-2 – Ion mobility spectrometric detection of nerve and blister agents Remote Chemical Agent Sensor VTB-1 – Field deployable laser radar for the detection of chemical warfare agents
Romania	Nerve Agent Alarm ASTN-2 – Nerve agent detector based on optical and acoustic signals
Switzerland	IMS 2000 CW Agent Detector – Ion mobility spectrometric detection of nerve and blister agents
CIS (formerly USSR)	Automatic Nerve Agent Detector Alarm, Model GSP-11 – Enzyme inhibition for the detection of nerve agents
UK	Chemical Agent Monitor (CAM), GID-2/GID-3 Detectors – Ion mobility spectrometry based monitor for the detection of nerve and blister agents NAIAD – Nerve agent immobilized enzyme detector and alarm
USA	ICAD Miniature Chemical Agent Detector – Personal detector based on electrochemical principals for the detection of nerve, blister, blood, and choking agents MINICAMS – Gas chromatographic detection of nerve and blister agents M21 Remote Sensing Chemical Agent Alarm (RSCAAL) – Passive infrared detection of chemical warfare agents Chemical Agent Detection Kit, M256A1 – Wet chemistry detection of nerve, blister, choking, and blood agents SAW MINICAD MK II – Surface acoustic wave detection of nerve and blister agents



## Safety and Disposal

Chemical warfare agents are extremely hazardous and lethal compounds. They can only be used in designated laboratories by personnel trained in safe-handling and decontamination procedures and with immediate access to medical support. Safety and standard operating procedures must be developed and approved before any chemical warfare agents are handled. Chemical warfare agents can only be used in laboratory chemical hoods with a minimum face velocity of 100 linear feet per minute equipped with emission control devices that limit exhaust concentration to below  $0.0001 \text{ mg m}^{-3}$ . Personnel handling chemical warfare agents should wear rubber gloves, lab coats, and full faceshields and keep a respirator (gas mask) within easy reach. Sufficient decontaminant to destroy all chemical warfare agents being handled must be on hand before commencing operations.

Blister and nerve agents can be destroyed using saturated methanolic solutions of sodium or potassium hydroxide. Decontaminated chemical warfare agents must be disposed of in an environmentally approved method according to local legislation.

See also: **Gas Chromatography**: Overview; Fourier Transform Infrared Spectroscopy. **Liquid Chromatography**: Overview. **Mass Spectrometry**: Overview; Ionization Methods Overview; Gas Analysis. **Nuclear Magnetic Resonance Spectroscopy**: Overview.

## Further Reading

- Compton JAF (1988) *Military Chemical and Biological Agents*. Caldwell NJ: The Telford Press.
- D'Agostino PA, Provost LR, and Hancock JR (1998) Analysis of mustard hydrolysis products by packed capillary liquid chromatography – electrospray mass spectrometry. *Journal of Chromatography A* 808: 177–184.
- D'Agostino PA, Hancock JR, and Provost LR (1999) Packed capillary liquid chromatography – electrospray mass spectrometry analysis of organophosphorus chemical warfare agents. *Journal of Chromatography A* 840: 289–294.
- D'Agostino PA, Chenier CL, and Hancock JR (2002) Packed capillary liquid chromatography – electrospray mass spectrometry of snow contaminated with sarin. *Journal of Chromatography A* 950: 149–156.
- Eldridge J (2001) *Jane's Nuclear, Biological, and Chemical Defence*. Coudsdon, UK: Jane's Information Group Limited.
- Ellison DH (2000) *Handbook of Chemical and Biological Warfare Agents*. Washington: CRC Press.
- Kientz ChE (1998) Chromatography and mass spectrometry of chemical warfare agents, toxins and related compounds: State of the art and future prospects. *Journal of Chromatography A* 814: 1–23.
- Ivarsson U, Nilsson H, and Santesson J (1992) *A Briefing Book on Chemical Weapons*. Sweden: Ljungforeytagen Oregro.
- Methodology and Instrumentation for Sampling and Analysis in the Verification of Chemical Disarmament* (1977–1994). Helsinki: The Ministry of Foreign Affairs of Finland.
- Read RW and Black RM (1999) Rapid screening procedures for the hydrolysis products of chemical warfare agents using positive and negative ion liquid chromatography – mass spectrometry and atmospheric pressure chemical ionization. *Journal of Chromatography A* 862: 169–177.
- Soderstrom MT and Ketola RA (1994) Identification of nerve agents and their homologues and dialkyl methylphosphonates by gas chromatography/Fourier transform infrared spectrometry (GC–FTIR). *Fresenius Journal of Analytical Chemistry* 350: 162–167.
- Somani SM (1992) *Chemical Warfare Agents*. New York: Academic Press.

# CHEMICALLY MODIFIED ELECTRODES

See **SENSORS: Chemically Modified Electrodes**

# CHEMILUMINESCENCE

Contents

**Overview****Liquid-Phase****Gas-Phase****Electrogenerated**

## Overview

**N W Barnett and P S Francis**, Deakin University,  
Geelong, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by T A Nieman, pp. 608–613, © 1995, Elsevier Ltd.

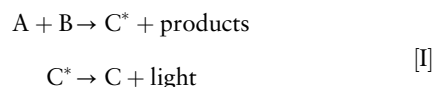
## Introduction

All chemical reactions are accompanied by energy changes, with any excess usually dissipated as heat. However, certain redox reactions yield light (chemiluminescence) at wavelengths from the near-ultraviolet to the near-infrared. Chemiluminescence has been observed in living systems since antiquity and from synthetic compounds since the late nineteenth century. Some commonly observed chemiluminescence emanates from fireflies and the various commercially available 'glow sticks'. These reactions also exhibit analytical utility because the emission intensities are a function of the concentrations of chemical species involved. Routine application of chemiluminescence as an analytical tool did not emerge until the 1970s for gas-phase reactions and the 1980s for liquid-phase reactions. This article will cover the basic principles common to all chemiluminescent reactions and discuss those characteristics that are important for analytical chemistry. The articles that follow in this section will elaborate on the mechanism and analytical applications of specific chemiluminescence reactions. Chemiluminescence in living systems or with reagents originating in living systems is called bioluminescence and is covered elsewhere in this encyclopedia.

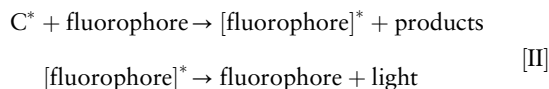
## Principles

In a chemiluminescence reaction between species A and B, some fraction of the product species, C, are formed in an electronically excited state, C\*, which

can subsequently relax to the ground state with emission of a photon (reaction [I]):



In some cases, the excited intermediate (C\*) transfers energy to a suitable fluorophore, which may then exhibit its characteristic fluorescence emission (reaction [II]). This phenomenon is referred to as indirect or sensitized chemiluminescence:



Chemiluminescence involves both a luminescence process and a chemical reaction. Consequently, the observed intensity depends upon the rate of the chemical reaction, the number of species excited, and their light emission efficiency (eqn [1]):

$$I_{\text{CL}} = \Phi_{\text{CL}} \frac{dP}{dt} = \Phi_{\text{EX}} \Phi_{\text{EM}} \frac{dP}{dt} \quad [1]$$

where  $I_{\text{CL}}$  is the chemiluminescence emission intensity (photons emitted per second),  $dP/dt$  the rate of the chemical reaction (molecules reacting per second),  $\Phi_{\text{CL}}$  is the chemiluminescence quantum yield (photons emitted per molecule reacted),  $\Phi_{\text{EX}}$  is the excitation quantum yield (excited states produced per molecule reacted), and  $\Phi_{\text{EM}}$  is the emission (luminescence) quantum yield (photons emitted per excited state).

The excitation quantum yield ( $\Phi_{\text{EX}}$ ) is the product of the efficiencies of (1) the chemical reaction, (2) the conversion of chemical potential into electronic excitation energy and in the case of sensitized chemiluminescence, and (3) the energy transfer. As a consequence, most chemiluminescent reactions have relatively low quantum yields compared to those of photoluminescence; the exception being the enzymatically mediated bioluminescent processes. In spite of this low quantum efficiency, chemiluminescence remains an attractive option for chemical analysis. This stems from three factors: (1) improved

signal-to-background and signal-to-noise ratios due to the absence of an excitation source; (2) inexpensive and robust instrumentation; and (3) increased selectivity due to the limited number of available reactions.

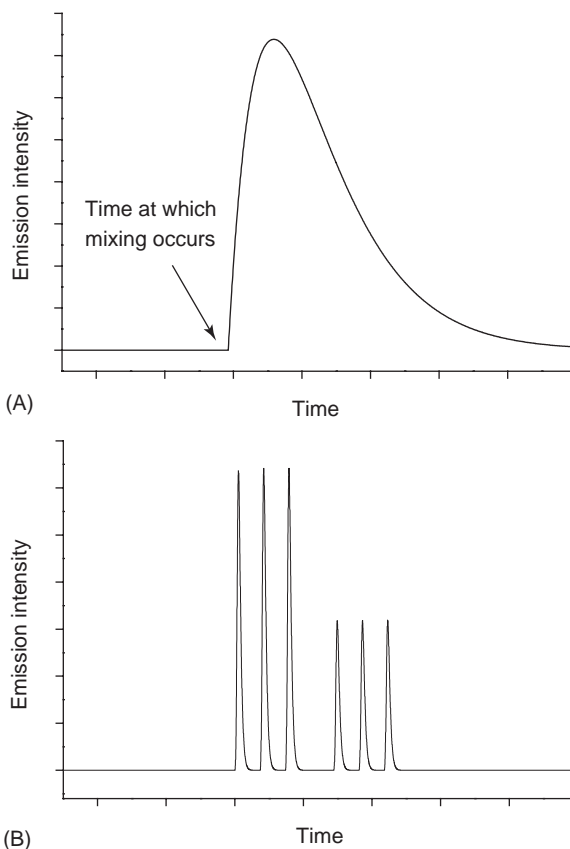
The energy for the creation of the electronically excited state comes from the chemical reaction. For emission in the region from the near-infrared to the near-ultraviolet, a reaction liberating between 130 and  $340 \text{ kJ mol}^{-1}$  is required. This amount of energy usually comes from bond cleavage or electron transfer. In systems involving bond cleavage (e.g., luminol or peroxyoxalates), the reagent can be used only once. However, some electron transfer reactions (including those of the radical ions of rubrene and *p*-benzoquinone or tris(2,2'-bipyridyl)ruthenium(III)) elicit emission without bond cleavage or rearrangement and as such the reagents can be recycled.

## Reaction Kinetics and the Observed Signal

Unlike photoluminescence, where the rate of emission reflects the lifetime of the excited state, the rate of generation of light from a chemical reaction is dependent on reaction kinetics. Consequently, the chemiluminescence emission intensity is inherently transient as illustrated in **Figure 1A**, which shows the variation of intensity with time after mixing the reactants. The overall timescale can vary from a short flash lasting less than a second to a pseudocontinuous glow lasting minutes to hours, depending upon the reaction involved and the concentration of reactants.

There are two basic approaches to employing chemiluminescence for analytical detection: static solutions and flow analysis. With the former, discrete portions of the reagent(s) and analyte are mixed in front of a detector and the chemiluminescence intensity versus time profile is monitored (**Figure 1A**). Generally, the height or area of the peak is measured and correlated with analyte concentration, but the temporal distribution of this signal can also be exploited. Analytes that react at different rates will produce disparate intensity-time profiles. In cases where the chemiluminescence kinetics are reproducibly distinct for two or more compounds, the possibility exists for mathematical deconvolution of signals resulting from mixtures of analytes and hence enhanced selectivity without a separation step.

With respect to flow analysis methodologies (including flow injection, sequential injection, pulsed flow, and chromatography), the analyte and reagent(s) are propelled in streams and merged to initiate the reaction close to or within a detection cell. As a consequence, the observed signal is dependent



**Figure 1** (A) Intensity versus time profile recorded from a static solution in a batch luminometer after a triggering reagent has been added. (B) Two sets of triplicate sample injections using a flow analysis system.

upon the reaction kinetics and the physical parameters of the flow system. These include the volume between the confluence and observation points, the flow cell geometry, and the flow rate. In order to achieve the desired sensitivity, these parameters require optimization. In the case of particularly fast reactions, it is crucial that the mixing occurs as close as practicable to the point of detection. Some typical flow analysis response profiles are shown in **Figure 1B**. If the reagent and analyte are continuously combined a steady-state signal will arise and this, coupled with a spectrofluorimeter, allows the collection of chemiluminescence spectra.

## Chemiluminescence versus Fluorescence Detection

Chemiluminescence and fluorescence have the potential for low detection limits and wide linear calibration ranges. Both are sensitive to environmental conditions (solvent type, the presence of quenchers, pH, ionic strength, and temperature) that may alter the quantum efficiency of emission. In addition, these

factors significantly affect the rate of the chemiluminescence reaction and the chemi-excitation efficiency and as such require control to ensure analytical precision. Fluorescence instrumentation is more complex than that employed for chemiluminescence, as the former requires an excitation source and two monochromators. Consequently, chemiluminescence measurements do not suffer from source fluctuation noise or light scattering. As a result, chemiluminescence detection can sometimes afford superior detection limits compared to those achievable with fluorescence. On the other hand, the spectrofluorimeter provides measurement versatility not possible with chemiluminescence: selection of emission and excitation wavelengths, source intensity and excitation beam size, and location. Many chemiluminescence reactions exhibit a measurable chemical background (or blank) signal, which can degrade detectability. In principle, a fluorescent molecule can be excited more than once in a given measurement, whereas very few chemiluminescence agents can be regenerated (a notable exception is tris(2,2'-bipyridyl)ruthenium(III)). Nevertheless, the use of intense excitation sources in fluorescence to maximize the signal is limited by problems of photodecomposition.

## Selectivity

Unlike photoluminescent techniques, where some degree of selectivity is often derived from the intrinsic excitation and emission wavelengths of the analyte, the inherent selectivity of chemiluminescence detection arises from the limited number of chemical reactions that produce significant amounts of light. Furthermore, wavelength discrimination usually offers no advantage to the chemiluminescence detection, as different analytes often lead to the same emitting species, and should be avoided due to the detrimental effect on sensitivity.

Selectivity of chemiluminescence reactions can be a concern. Some reactions are essentially compound-specific. An example is tetrakis(dimethylamino)ethylene, which undergoes chemiluminescence reaction only with  $O_2$ . Although such specificity provides freedom from measurement interference, that chemiluminescence reaction then lacks universal application. In other cases, several species could yield emission with a given reagent. These situations then require coupling of the chemiluminescence detection with some sort of highly selective physical or chemical step (such as chromatography, immunoassay, enzyme reactions) to achieve an interference-free measurement. An

example of this type of chemiluminescence reaction would be the metal-ion catalyzed oxidation of luminol by transition metals. Many transition-metal ions and organometallics will yield intense chemiluminescence emission upon reaction with an alkaline solution of luminol and hydrogen peroxide. The emission emanates from the oxidation product of luminol, thus there is no spectroscopic way to identify which metal-ion catalyst was responsible for the emission.

## Corrected Emission Spectra

Chemiluminescence spectra provide possible insights into the nature of the emitting species and consequently should be mandatory for any speculation concerning the light-producing pathway. Clearly, this is not the case with indirect chemiluminescence as the emission is identical to the photoluminescence of the added fluorophore. As with photoluminescence, chemiluminescence emission spectra should be corrected for the wavelength dependence of the detector response and monochromator transmission. A correction factor can be created with a lamp of known spectral irradiance or with fluorescence emission standards. Correction is particularly important for broadly distributed bands and those extending into the near-infrared region where the sensitivity of most instruments is poor. For example, the uncorrected emission spectrum from the oxidation of urea with alkaline hypobromite appeared to be broadly distributed across the visible region with a maximum at 510 nm. After correction, the maximum was at  $\sim 710$  nm.

## Gas-Phase Reactions

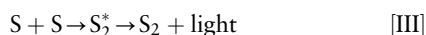
Numerous gas-phase chemiluminescence reactions have been studied and several of them commercialized for atmospheric monitoring and gas chromatography (GC) detection. **Table 1** and the discussion that follows highlight a few of the most important reactions. More details on gas-phase chemiluminescence is contained in a subsequent chapter. These reactions can be grouped into two categories: those occurring at room temperature and those occurring in flames between high-energy molecular or atomic species.

The most well-known flame chemiluminescence reactions are for sulfur and phosphorus detection. These form the basis of flame photometric detectors used in gas chromatography. When sulfur compounds are pyrolyzed in a hydrogen-rich flame, excited diatomic sulfur is formed as illustrated in

**Table 1** Selected gas-phase chemiluminescence reactions

Analyte	Reagent(s)	Emitter	Wavelengths
Ozone	Ethylene or other alkenes	Numerous	Visible and near-IR
Nitric oxide (NO)	Ozone	NO <sub>2</sub>	Visible and near-IR
Nitrogen compounds	Ozone after conversion to NO	NO <sub>2</sub>	Visible and near-IR
Phosphorus compounds	Hydrogen-rich flame	PHO	526 nm
Sulfur compounds	Hydrogen-rich flame	S <sub>2</sub>	384 nm
Sulfur compounds	H <sub>2</sub> flame followed by ozone	SO <sub>2</sub>	360 nm

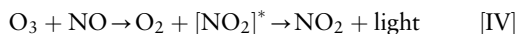
reaction [III]:



The emission from S<sub>2</sub> is an intense blue and is monitored through a filter at ~400 nm, thus eliminating interference from the flame. The calibration function has a squared dependence on the analyte concentration, as the chemiluminescence reaction involves two sulfur atoms. This detector is especially good for sulfur dioxide, carbon disulfide, methyl mercaptan, and hydrogen sulfide. Similarly, burning of phosphorus compounds yields emission from PHO\*, which is monitored at 526 nm. The phosphorus flame photometric detector yields linear working curves for over four decades and is useful for detection of phosphorus-containing pesticides.

Analytically useful chemiluminescence can be generated from room-temperature gas-phase reactions with ozone for the determination of olefins, nitric oxide, metal carbonyls, nonmetal hydrides, and other species. The reaction of ozone with ethylene (or other olefins) yields emission in the range from 300 to 600 nm. Using a constant flow of ethylene as the reagent, this reaction can be used for monitoring atmospheric ozone over a four-decade concentration range with a detection limit of three parts in 10<sup>9</sup> by volume air. This reaction is also the basis of a GC detector for hydrocarbons with detection limits in the nanogram range. Elevated reaction temperatures (>200°C) are required to detect either aromatic or saturated hydrocarbons.

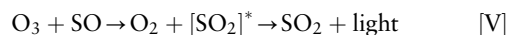
The reaction of ozone with nitric oxide (reaction [IV]) has been utilized for several analytical applications, including the determination of atmospheric nitric oxide, which has a linear response over six orders of magnitude and a detection limit of about one part in 10<sup>12</sup> by volume. The wavelength distribution of this chemiluminescence ranges from the visible to the near-infrared:



This reaction can also be applied to a variety of analytes following their conversion to nitric oxide. For example, *N*-nitrosoamines can be detected at

picogram levels after pyrolytic decomposition at temperatures between 300°C and 350°C as utilized in the GC detector erroneously termed a thermal energy analyzer. Additionally, catalytic oxidation with molecular oxygen at 800–1000°C will yield nitric oxide from nitrogen-containing compounds, which can be used instead of the Kjeldahl determination for total nitrogen. Nitric oxide can also be produced by reacting organic compounds with nitrogen dioxide at a heated gold catalyst; thereby extending this type of chemiluminescence detection to a variety of analytes that do not contain nitrogen. Interestingly, saturated and chlorinated hydrocarbons give no response under these conditions, which provides a limited degree of selectivity.

The reaction between ozone and SO (see reaction [V]) elicits chemiluminescence between 260 and 480 nm and is the foundation of an alternative sulfur detector that is commercially available for use with gas, supercritical fluid, and liquid chromatography. The SO is created from sulfur-containing analytes by burning the chromatographic effluent in a hydrogen flame:



The detection limits are ~20 times better than the flame photometric detector and the response is linear with concentration over five orders of magnitude.

## Liquid-Phase Reactions

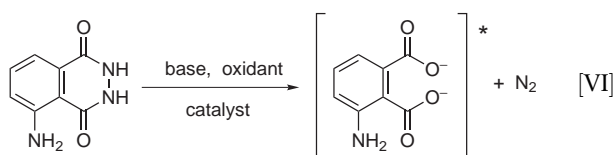
Liquid-phase chemiluminescence reactions have been extensively studied and applied to flow analysis, high-performance liquid chromatography, capillary electrophoresis, and batch luminometers. Some of the more important chemistries are listed in Table 2. The discussion that follows will highlight three reactions, which have each been selected to illustrate certain unique attributes. Much more detail on liquid-phase chemiluminescence can be found in a subsequent chapter.

Luminol (5-amino-2,3-dihydro-1,4-phthalazinedione) is one of the most commonly used liquid-phase chemiluminescence reagents. In aqueous alkaline

**Table 2** Selected liquid-phase chemiluminescence reactions

Reagent(s)	Analytes	Emission max.
Luminol (alkaline)	Transition metal ions, H <sub>2</sub> O <sub>2</sub> , peroxidase, reactive oxygen species	425 nm
Lucigenin (alkaline)	Transition metal ions, reactive oxygen species	440 nm
Tris(2,2'-bipyridyl)ruthenium(III) (acidic)	Amines, amino acids, oxalate, NADH, some alkaloids	610 nm
Potassium permanganate (acidic with polyphosphates)	Catechols, catecholamines, indoles, ascorbic acid	690 nm
Peroxyoxalates (alkaline)	Fluorescent compounds and analytes derivatized with suitable fluorophores	Dependent on fluorophore

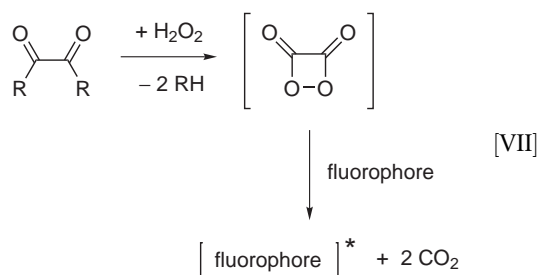
solution, luminol is oxidized to yield excited-state 3-aminophthalate, which emits at  $\sim 425$  nm (reaction [VI]):



Various oxidants (including potassium permanganate, sodium hypochlorite, and iodine) can be used but hydrogen peroxide is the most common. The use of hydrogen peroxide requires a catalyst and these include transition metal cations (such as Co(II), Cu(II), and Fe(III)), potassium hexacyanoferrate(III), or certain metallo-complexes (hemin, hemoglobin, peroxidases). Within limits, the chemiluminescence emission intensity is directly proportional to the concentration of luminol, hydrogen peroxide, or catalyst; consequently, this reaction has been used to determine all of these species. Many applications have resulted from the luminol-based determination of hydrogen peroxide that is generated by reactions of oxidase enzymes with appropriate substrates; in this way analytes such as glucose, cholesterol, uric acid, sucrose, and glucoside metabolites can be quantified. Approximately 20 catalysts can also be determined using this chemistry; these include horseradish peroxidase, microperoxidase, Co(II), Cr(III), and Cu(II) at sub-nanomolar levels. Peroxidases are common labels for immunoassays, and as such luminol chemiluminescence is a useful method for their sensitive determination. The structurally related compound iso-luminol finds application as a chemiluminescent label in both immunoassay and for pre-column chromatographic derivatization.

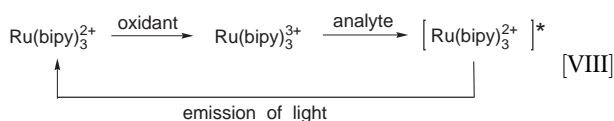
The light generated from the oxidation of substituted diaryloxalate esters with hydrogen peroxide, in the presence of a fluorophore, is known as peroxyoxalate chemiluminescence (see reaction [VII]). This group of reactions is highly sensitive to base

catalysis by potassium hydroxide or benzyltrimethylammonium hydroxide, and exhibit the highest non-biological chemiluminescence quantum yields (up to 30%):



In these systems, a high-energy intermediate excites a suitable fluorophore, which then emits its characteristic fluorescence spectrum; consequently, they are termed indirect or sensitized chemiluminescence. The most common analytical application has been as a postcolumn reaction detector for liquid chromatography. Various fluorescent analytes (polycyclic aromatic hydrocarbons and polycyclic aromatic amines) and compounds derivatized using dansyl chloride, fluorescamine, or *o*-phthalaldehyde have been determined with sub-femtomole detection limits.

In the tris(2,2'-bipyridyl)ruthenium(II) [Ru(bipy)<sub>3</sub><sup>2+</sup>] system, an orange emission centered at 610 nm results from the excited state [Ru(bipy)<sub>3</sub><sup>2+</sup>]\*. Central to the utilization of this type of chemiluminescence is the generation the reactive oxidant tris(2,2'-bipyridyl)ruthenium(III) [Ru(bipy)<sub>3</sub><sup>3+</sup>] and its subsequent reaction with a reductive analyte, as shown in reaction [VIII]:



In the vast majority of applications the reagent has been electrochemically generated, but there are also

examples where simple redox chemistries have been employed with equal efficacy. Notwithstanding the method of reagent production, this chemistry has been employed to sensitively and selectively detect a variety of analytes including sodium oxalate, NADH, amino acids, antibiotics, opiates, and pharmaceuticals.  $\text{Ru}(\text{bipy})_3^{2+}$  can be used as a label for immunoassay or DNA probes. Importantly, as electron transfer rather than bond cleavage or rearrangement leads to the excited state and the reaction results in regeneration of the  $\text{Ru}(\text{bipy})_3^{2+}$ , it is possible to perform analyses with no net consumption of the reagent.

## Solid-Phase Reactions

In contrast to the large number and extensive application of chemiluminescence reactions in the gas and liquid phases, the use of solid-phase chemiluminescence reactions is limited. The oxidation of many organic compounds is accompanied by weak chemiluminescence. Measurement of this emission can be used to characterize oxidative changes in materials (such as polymer degradation due to exposure to heat or ionizing radiation, or flavor alteration in foods) and to evaluate stabilizers that are used to retard these processes. The oxidation of polymers often results in the formation of peroxides, cross-linking, and chain cleavage. Polyolefins, polyamides, rubber, epoxies, lubricating oils, and edible oils are examples of some materials characterized by chemiluminescence. Measurements on solids involve heating at a controlled temperature (25–250°C) in front of the detector and monitoring the intensity of chemiluminescence emission over time. The shape of the chemiluminescence intensity versus time profile is characteristic of particular materials and can help determine their composition or histories. In the presence of oxygen, the chemiluminescence is

proportional to the rate of oxidation. In the absence of oxygen, the emission is due to decomposition of previously formed peroxide groups. Nevertheless, in both cases the mechanism involves formation of excited-state aldehydes or ketones, and the emission intensity can be enhanced by the addition of highly fluorescent molecules (such as substituted anthracenes).

## Acknowledgment

The authors would like to acknowledge the contribution of the late Professor Timothy A. Nieman, who wrote this chapter for the first edition, as we have employed a similar format and some of his material for this article.

*See also:* **Bioluminescence.** **Chemiluminescence:** Liquid-Phase; Gas-Phase. **Luminescence:** Solid Phase. **Phosphorus.** **Sulfur.**

## Further Reading

- Birks JW (ed.) (1989) *Chemiluminescence and Photochemical Reaction Detection in Chromatography*. New York: VCH.
- Burr JG (1985) *Chemi- and Bioluminescence*. New York: Dekker.
- Campbell AK (1988) *Chemiluminescence: Principles and Applications in Biology and Medicine*. New York: VCH.
- García-Campaña AN and Baeyens WRG (eds.) (2001) *Chemiluminescence in Analytical Chemistry*. New York: Dekker.
- Mendenhall GD (1990) Chemiluminescence techniques for the characterization of materials. *Angewandte Chemie International Edition in English* 29: 362–373.
- VanDyke K (ed.) (1985) *Bioluminescence and Chemiluminescence: Instruments and Applications*. Boca Raton: CRC Press.

## Liquid-Phase

**N W Barnett and P S Francis**, Deakin University, Geelong, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Although many liquid-phase chemiluminescence reactions have been observed and investigated, only a

limited number of reagents are responsible for the majority of analytical applications; these include acidic potassium permanganate, acridinium esters, diaryloxalates, dioxetanes, hypohalites, luminol, and tris(2,2'-bipyridyl)ruthenium(III). The elucidation of reaction mechanisms for chemiluminescent processes is highly desirable to facilitate improvement of the overall quantum efficiency and hence the analytical figures of merit. Knowledge of the mechanism



examples where simple redox chemistries have been employed with equal efficacy. Notwithstanding the method of reagent production, this chemistry has been employed to sensitively and selectively detect a variety of analytes including sodium oxalate, NADH, amino acids, antibiotics, opiates, and pharmaceuticals.  $\text{Ru}(\text{bipy})_3^{2+}$  can be used as a label for immunoassay or DNA probes. Importantly, as electron transfer rather than bond cleavage or rearrangement leads to the excited state and the reaction results in regeneration of the  $\text{Ru}(\text{bipy})_3^{2+}$ , it is possible to perform analyses with no net consumption of the reagent.

## Solid-Phase Reactions

In contrast to the large number and extensive application of chemiluminescence reactions in the gas and liquid phases, the use of solid-phase chemiluminescence reactions is limited. The oxidation of many organic compounds is accompanied by weak chemiluminescence. Measurement of this emission can be used to characterize oxidative changes in materials (such as polymer degradation due to exposure to heat or ionizing radiation, or flavor alteration in foods) and to evaluate stabilizers that are used to retard these processes. The oxidation of polymers often results in the formation of peroxides, cross-linking, and chain cleavage. Polyolefins, polyamides, rubber, epoxies, lubricating oils, and edible oils are examples of some materials characterized by chemiluminescence. Measurements on solids involve heating at a controlled temperature (25–250°C) in front of the detector and monitoring the intensity of chemiluminescence emission over time. The shape of the chemiluminescence intensity versus time profile is characteristic of particular materials and can help determine their composition or histories. In the presence of oxygen, the chemiluminescence is

proportional to the rate of oxidation. In the absence of oxygen, the emission is due to decomposition of previously formed peroxide groups. Nevertheless, in both cases the mechanism involves formation of excited-state aldehydes or ketones, and the emission intensity can be enhanced by the addition of highly fluorescent molecules (such as substituted anthracenes).

## Acknowledgment

The authors would like to acknowledge the contribution of the late Professor Timothy A. Nieman, who wrote this chapter for the first edition, as we have employed a similar format and some of his material for this article.

*See also:* **Bioluminescence.** **Chemiluminescence:** Liquid-Phase; Gas-Phase. **Luminescence:** Solid Phase. **Phosphorus.** **Sulfur.**

## Further Reading

- Birks JW (ed.) (1989) *Chemiluminescence and Photochemical Reaction Detection in Chromatography*. New York: VCH.
- Burr JG (1985) *Chemi- and Bioluminescence*. New York: Dekker.
- Campbell AK (1988) *Chemiluminescence: Principles and Applications in Biology and Medicine*. New York: VCH.
- García-Campaña AN and Baeyens WRG (eds.) (2001) *Chemiluminescence in Analytical Chemistry*. New York: Dekker.
- Mendenhall GD (1990) Chemiluminescence techniques for the characterization of materials. *Angewandte Chemie International Edition in English* 29: 362–373.
- VanDyke K (ed.) (1985) *Bioluminescence and Chemiluminescence: Instruments and Applications*. Boca Raton: CRC Press.

## Liquid-Phase

**N W Barnett and P S Francis**, Deakin University, Geelong, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Although many liquid-phase chemiluminescence reactions have been observed and investigated, only a

limited number of reagents are responsible for the majority of analytical applications; these include acidic potassium permanganate, acridinium esters, diaryloxalates, dioxetanes, hypohalites, luminol, and tris(2,2'-bipyridyl)ruthenium(III). The elucidation of reaction mechanisms for chemiluminescent processes is highly desirable to facilitate improvement of the overall quantum efficiency and hence the analytical figures of merit. Knowledge of the mechanism

involved enables structural and environmental changes to be made on a fundamental rather than an empirical basis. The major stumbling block to the understanding of many chemiluminescent reaction mechanisms is the isolation and characterization of the key intermediates of the reaction pathways. In spite of these difficulties, a significant body of work exists on this area and some of the proposed mechanisms for the more analytically useful chemiluminescent systems are summarized in the following sections. This is complemented with a discussion on the application of these detection chemistries with flow analysis, high-performance liquid chromatography (HPLC), capillary electrophoresis, immunoassay, and DNA assays.

### Acidic Potassium Permanganate

The earliest use of acidic potassium permanganate as a reagent for chemiluminescence is most probably attributable to A.A. Grinberg, who in 1920 observed emission of light during the oxidation of pyrogallol. Since that time, this reagent has been employed under a variety of conditions to elude chemiluminescence from a wide range of analytes. The nature of the emitting species in these reactions has been the subject of considerable conjecture over the past three decades; candidates have included fluorescent oxidation products of the analyte, a triplet dimer of carbon dioxide, singlet oxygen, sulfur dioxide, molecular nitrogen, nitric oxide, and manganese(II). However, it was not until 2002 that the necessary spectroscopic and chemical evidence was obtained to confirm several previous postulations that the emitter was an excited manganese(II) species. In common with tris(2,2'-bipyridyl)ruthenium(II) chemiluminescence, the emission from acidic potassium permanganate reactions is an example of solution-phase chemically induced phosphorescence.

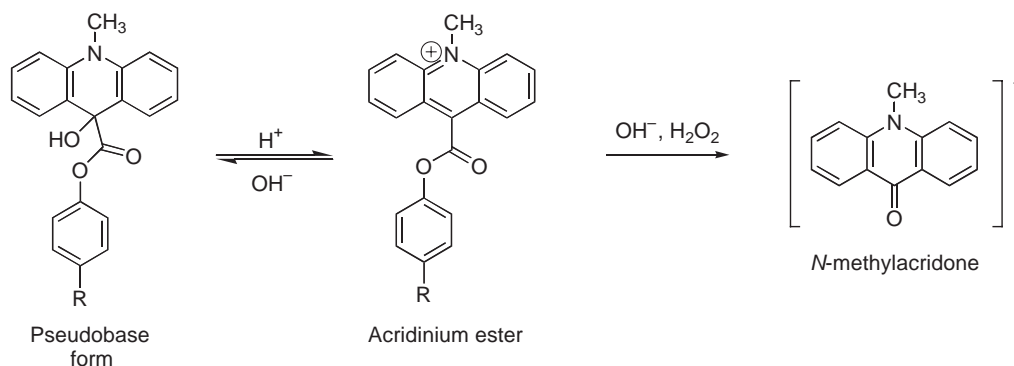
With respect to the analytical applications, ~100 papers have been published since the first report by Stauff and Jaeschke appeared in 1975. The applications can be broadly divided into two classes: inorganic (e.g.,

sulfur dioxide, sulfite, hydrogen sulfide, manganese(II), hydrazine, hydrogen peroxide, and iron(II)) and organic (e.g., opiate and strychnine alkaloids, catechols, catecholamines, indoles, ascorbic acid, and a variety of pharmaceuticals). The majority of these analyses were undertaken using flow analysis or HPLC; the efficacy of this detection chemistry has also been demonstrated with capillary electrophoresis. The utilization of this chemiluminescence has not been as extensive as some other systems. Nevertheless, acidic potassium permanganate can sensitively detect molecules containing phenolic and/or amine moieties and therefore it has considerable potential for the determination of a wide range of important analytes.

### Acridinium Esters

Acridinium phenyl esters react with alkaline hydrogen peroxide to yield excited state *N*-methylacridone, which then emits at ~40 nm (**Scheme 1**). Acridinium esters are structurally related to lucigenin (bis-*N*-acridinium nitrate), which is one of the classic organic chemiluminescence reagents. Both luminol and lucigenin react with alkaline hydrogen peroxide and are catalyzed by transition-metal ions, but the chemiluminescence reaction between acridinium esters and hydrogen peroxide does not require catalysis.

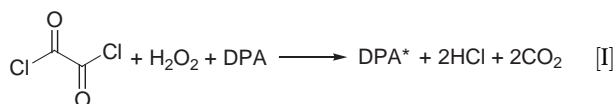
Although acridinium esters can be used to determine hydrogen peroxide, they are more commonly used as labels for immunoassay or DNA probes. As no catalyst is required, acridinium ester chemiluminescence has a lower chemical blank than the oxidation of luminol and therefore has superior detection limits. A strongly alkaline solution (pH 12–13) is required, but under these conditions acridinium esters undergo reversible conversion to the non-chemiluminescent pseudobase form (**Scheme 1**), which slowly decomposes. As a consequence, acidic hydrogen peroxide is added to reconvert the pseudobase to the acridinium ester before the chemiluminescence reaction is initiated with a sodium hydroxide solution.



**Scheme 1**

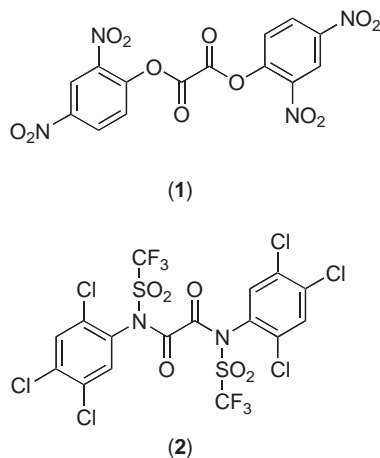
## Diaryl Oxalates and Oxamides

Indirect (sensitized) chemiluminescence was first reported by E.A. Chandross in 1963 as a result of the reaction between oxalyl chloride and hydrogen peroxide in the presence of 9,10-diphenylanthracene (DPA), as shown in reaction [I]. The observed transient blue emission corresponds to the fluorescence of the aromatic hydrocarbon and is generated via energy transfer from an excited-state reaction intermediate:



A few years later it was discovered that certain substituted esters and amides of oxalic acid could also induce chemiluminescence from suitable fluorophores in the presence of hydrogen peroxide. This class of chemiluminescent reactions has become known trivially as ‘peroxyoxalate’.

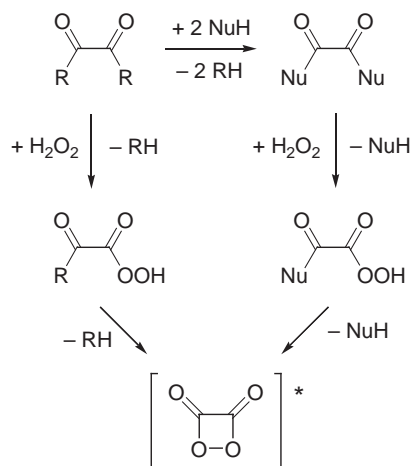
In stark contrast to most other nonbiological chemiluminescence reactions, the overall quantum yield for certain peroxyoxalate systems under optimum conditions can range from 0.13 to 0.34. Unlike luminol, the properties of reactivity, energy conversion/transfer, and luminescence are not the province of a single molecule, and theoretically a fluorophore with a quantum yield tending toward unity could be coupled with an oxalate or oxamide designed to give the largest possible quantity of the key high-energy reaction intermediate. The best chemiluminescence quantum yields have been obtained with certain bisphenyloxalates and oxamides that have electron-withdrawing substituents. For example, by reacting compound (1) with hydrogen peroxide in dimethylphthalate and employing rubrene as a fluorophore, an overall quantum yield of 0.23 has been achieved. In analogous fashion, compound (2) initiated chemiluminescence from 1-chloro-9,10-bis(phenylethynyl)anthracene with a quantum efficiency of 0.34.



The electron withdrawing power of the trifluoromethylsulfonyl (triflyl) group attached to the nitrogen activates the oxamide toward the reaction with hydrogen peroxide. However, the correlation of the electron withdrawing power of the substituents with efficiency of light production is limited; the 2,4,6-trichloro isomer of compound (2) is less than half as efficient under identical reaction conditions. Peroxyoxalate chemiluminescence is quite versatile with several reagents showing efficacy with an array of fluorophores that have first excited singlet state energies in the range from 200 to 440 kJ mol<sup>-1</sup> (600–270 nm). Generally, analyte molecules with low first excited singlet state energies and high fluorescence quantum yields exhibit the best overall efficiency. However, there are many exceptions to this simple model and a quantitative relationship between chemiluminescent quantum yield and singlet excitation energy does not exist.

The mechanism of peroxyoxalate chemiluminescence centers on the nature of the postulated ‘key intermediate’ and its mode of interaction with, and excitation of, the fluorophore. An early proposal for this key intermediate, the highly strained 1,2-dioxetanedione, was confirmed more than three decades later using low-temperature <sup>13</sup>C nuclear magnetic resonance spectroscopy in combination with *ab initio* calculations. As shown in Scheme 2, the formation of the key intermediate is subject to both nucleophilic and general-base catalysis by concurrent mechanisms.

Spectroscopic analysis of peroxyoxalate chemiluminescence also revealed several other, as yet unknown, transient entities. The next step in the proposed mechanism was the formation of a charged transfer complex (between 1,2-dioxetanedione and the fluorophore) that decomposes to yield the



Scheme 2

fluorophore in the excited state. Evidence obtained from mechanistic studies on the thermally induced chemiluminescent reactions of dioxetanes and dioxetanones led to a modification of this excitation step. After oxidation of the fluorophore, the charge transfer complex undergoes loss of one molecule of carbon dioxide to produce a radical ion pair. The subsequent annihilation of the two radical ions affords the fluorophore in the excited state. This is commonly termed chemically initiated electron exchange luminescence (CIEEL).

A detailed kinetic investigation of the reaction between bis(pentachlorophenyl)oxalate and hydrogen peroxide with sodium salicylate as a catalyst and 9,10-diphenylanthracene as the fluorophore in chlorobenzene solvent supported the CIEEL approach. This study predicted a linear dependence of light intensity on initial hydrogen peroxide concentration together with a direct proportionality between overall quantum yield and the oxidation potential of the fluorophore. Both these predictions have been experimentally verified. Studies employing various oxalate esters in solvent systems containing small amounts of water have revealed an even more complex mechanism. This is manifested by the observation of two maxima in the intensity–time profile of the chemiluminescent reaction. The result has been a proposed mechanism with five or more reaction intermediates of which only a certain species may transfer energy to the fluorophore. As the water content of the solvent system increases to that commonly used in HPLC (20–50%, v/v) the chemiluminescence intensity–time profile exhibits only one maximum.

A simplified kinetic model for a comparable system has been described based around the concept of three pools of substances: (A) oxalate, hydroxide, and catalyst; (B) intermediate substances; and (C) products. Pool A reacts with a pseudo-first-order rate constant  $r$  to produce a pool of intermediates (B), which are subsequently converted at a rate constant  $f$  to pool C, as shown in reaction [II]:



The chemiluminescent intensity–time function ( $I_t$ ) can therefore be derived as (eqn [1])

$$I_t = \frac{d(h\nu)}{dt} = \frac{Mr}{f-r}(e^{-rt} - e^{-ft}) \quad [1]$$

where  $M$  is the theoretical maximum intensity for quantitative conversion of reactants into emitting species. This model has been used to determine overall quantum yields of partially aqueous systems.

In spite of these investigations, there is still considerable ambiguity regarding the mechanisms involved. Such studies are hampered by the very nature of the reaction; subtle changes in the oxalate structure, chemical environment, and type of fluorophore can result in large variations in overall quantum yield and emission intensity versus time profiles.

Peroxyoxalate chemiluminescence reactions are analytically important (particularly in HPLC). The most commonly used reagents are bis-(2,4,6-trichlorophenyl)oxalate (TCPO) and bis-(2,4-dinitrophenyl)oxalate (DNPO). Fluorescent compounds (including anthracene, perylene, aminoanthracenes, and aminopyrenes) and suitably derivatized analytes (such as amines, steroids with dansyl chloride; thiols with *N*-[4-(6-dimethylamino-2-benzofuranyl)-phenyl]maleimide and catecholamines with fluorescamine) can be sensitively detected.

Peroxyoxalate chemistry can be used over a wide pH range; TCPO from pH 5 to 9 and DNPO at more acidic values. The more rapid reaction kinetics of DNPO can result in slightly more sensitive assays; however, it is not as stable as TCPO. A major concern with diaryloxalates is the necessity for an organic solvent system owing to their limited aqueous solubility and stability. Common solvent choices have been ethyl acetate/ethanol/water or acetonitrile/water. The requirement to include an organic solvent is a limitation for the determination of hydrogen peroxide generated from oxidase enzymes. Although water-soluble oxamides have been reported with comparable performance to that of TCPO, this class of reagents is not fully evaluated or readily available. The most common application of peroxyoxalate chemiluminescence has been with HPLC for the detection of analytes either exhibiting native fluorescence or those suitably derivatized. Because the fluorophore is not involved in the chemical reaction, it may go through multiple chemiluminescence reaction/excitation/emission cycles to generate several photons per analyte molecule. The lack of source scatter and noise can sometimes give chemiluminescence a sensitivity advantage over fluorescence for the same analyte. Fluorophores such as dansyl derivatives and rubrene can be detected at sub-femtomolar levels with peroxyoxalate chemiluminescence.

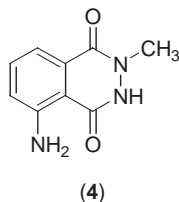
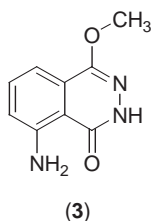
## Dioxetanes

Dioxetanes are four-membered cyclic peroxides and their relative stability depends on the types of substituent groups present. Certain 1,2-dioxetanes are stable at room temperature but can be chemically triggered to produce chemiluminescence, these have an adamantyl group on one side of the ring and a

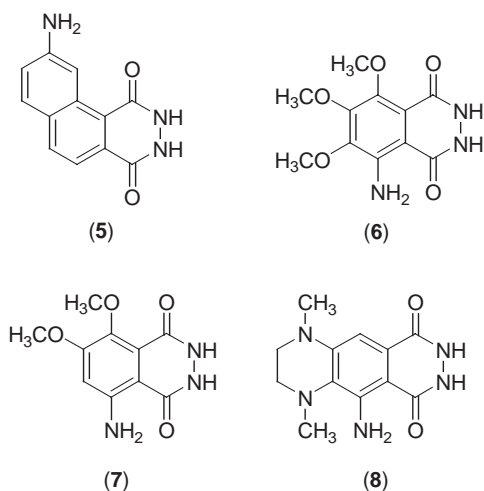


The tautomerism is driven by the increased basic strength of 3-aminophthalate in the aprotic solvent with the resultant structural change decreasing the energy of lowest excited singlet state. The kinetics of the reaction in dimethylsulfoxide are first order with respect to the concentrations of hydroxide, oxygen, and luminol. An important analytical feature is the catalytic effect of certain transition metal cations ( $M^{n+}$ ) upon the reaction shown in **Scheme 4**. The luminescence intensity is proportional to the concentration of species including cobalt(II), copper(II), chromium(III), iron(II), and nickel(II) at levels down to  $\sim 10^{-11} \text{ mol l}^{-1}$ .

Various studies have been conducted on compounds that are structurally similar to luminol with a view to improving the overall quantum yield ( $\phi_{cl}$ ). It was shown that even minor alterations to the heterocyclic ring destroyed the molecules' chemiluminescent property. Thus, compounds such as (3) and (4) showed no chemiluminescence under the reaction conditions outlined in **Scheme 4**.

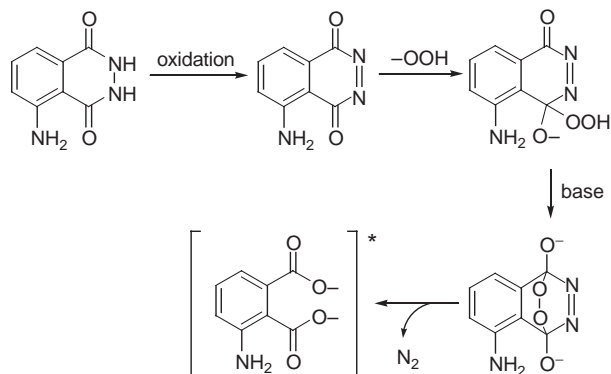


For the quantum yield to be useful the aromatic ring requires substitution, with 5-position isomers exhibiting significantly higher efficiency than those at the 6-position. The electron-withdrawing ability of the substituent is roughly inversely proportional to the efficiency of light production. This correlation arises from the consequent lowering of the fluorescence quantum yield of the emitting species with the increased electron-withdrawing power of the substituents on the aromatic ring. For example, 5-nitro-2,3-dihydro-1,4-phthalazinedione is  $\sim 10\,000$  times less chemiluminescent than luminol. The knowledge gained from the investigations into the structural aspects of luminol analogs paved the way for the synthesis of compounds such as (5)–(7), which are more chemiluminescent than luminol. However, the relationship between molecular structure and chemiluminescent efficiency is not completely understood, since the bis-dialkylamino derivative (8) is approximately one-third as efficient as luminol.



Despite the numerous and detailed studies concerned with the luminol reaction mechanism, the exact nature of the pathways and intermediates involved are somewhat speculative. However, the most likely starting point appears to be the oxidation of the cyclic diacyl hydrazine moiety to give an azaquinone (**Scheme 6**). In the presence of basic hydrogen peroxide, nucleophilic attack by the hydroperoxide ion would seem most likely; this idea is supported by the luminescence intensity dependence upon hydrogen peroxide concentration. Although there are several possible outcomes from the reaction of the azaquinone with hydroperoxide ion, we shall (for reasons of simplicity) concentrate on that shown in **Scheme 6**.

The mechanism is also consistent with the observation that cyclic diacyl hydrazines are far more chemiluminescent than similar acyclic compounds. The bicyclic peroxide intermediate shown in **Scheme 6** may be precluded by rupture of one C–N bond after the addition of the hydrogen peroxide anion, but before the formation of an anti-aromatic endoperoxide that decomposes to form the emitter. Obtaining hard evidence for the existence of these



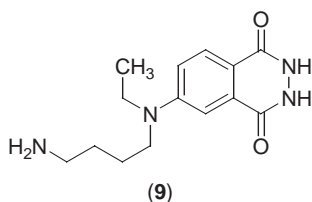
**Scheme 6**



intermediate species, whilst problematic, may be feasible with low-temperature nuclear magnetic resonance spectroscopy, since the chemiluminescence from luminol is achievable at temperatures down to  $-50^{\circ}\text{C}$ . This serves to underline the difficulties in relation to the elucidation of chemiluminescent reaction mechanisms.

Luminol is one of the most commonly used liquid-phase chemiluminescence reagents. Oxidants such as permanganate, hypochlorite, or iodine can be used, but hydrogen peroxide is the most common. As previously noted, a catalyst is required with this chemistry and these include transition-metal ions, hexacyanoferrate(III), hemin, and heme proteins (hemoglobin, peroxidases, catalase, and cytochromes). The optimum reaction pH varies between 8 and 11, depending upon the catalyst. This chemistry can be used to sensitively determine the oxidant, catalyst, or species derivatized with luminol or related compounds. It is possible to electrochemically initiate the chemiluminescence reaction of hydrogen peroxide and luminol at an electrode that is held at about  $+0.5\text{ V}$  (versus  $\text{Ag}/\text{AgCl}$ ). The electrode takes the place of the conventional dissolved catalyst. The chemiluminescence reaction is fast enough to confine emission close to the electrode surface. For determination of species labeled with luminol, two electrodes can be used, one at  $-1.0\text{ V}$  to generate the necessary hydrogen peroxide and one at about  $+0.5\text{ V}$  to initiate the chemiluminescence reaction. Detection limits are the same with electrogenerated chemiluminescence as with a conventional dissolved catalyst.

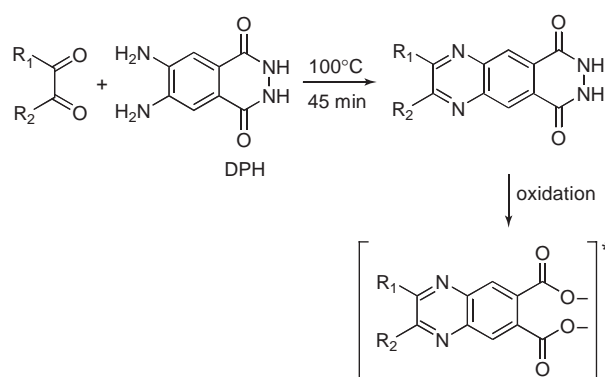
Luminol can be used to derivatize analytes (for immunoassay or HPLC) through substitution at the primary amine but this results in a 10- to 100-fold decrease in chemiluminescence efficiency. Isoluminol (6-amino-2,3-dihydro-1,4-phthalazinedione) is somewhat less efficient than luminol; however, it does not suffer a decrease in efficiency upon binding. Aminobutylethylisoluminol (ABEI, **9**) is a useful precolumn labeling reagent for amines and carboxylic acids because the hydrocarbon spacer isolates the chemiluminophore from the analyte.



Similarly, 4,5-diaminophthalhydrazide (DPH; 6,7-diamino-2,3-dihydro-1,4-phthalazinedione) has been

**Table 1** Comparison of metal ion detection limits with different chemiluminescence systems. All ions have detection limits of  $10\ \mu\text{mol l}^{-1}$  or lower. Ions in bold type have detection limits of  $10\ \text{nmol l}^{-1}$  or lower. Only first-row transition-metal species are considered

Luminol	Lucigenin	Gallic acid	Lophine
Ti(IV)			
V(II)			
<b>Cr(III)</b>	Cr(III)		Cr(III)
<b>Mn(II)</b>		Mn(II)	
<b>Fe(II)</b>	Fe(II)		
<b>Fe(III)</b>	Fe(III)		
<b>Co(II)</b>	<b>Co(II)</b>	Co(II)	Co(II)
<b>Ni(II)</b>	Ni(II)		
<b>Cu(II)</b>	Cu(II)		Cu(II)



**Scheme 7**

used to derivatize aromatic aldehydes,  $\alpha$ -keto acids, and  $\alpha$ -dicarbonyl compounds (Scheme 7).

Certain species significantly increase the chemiluminescence intensity and duration, particularly when horseradish peroxidase is employed as the catalyst. For example, *p*-iodophenol enhances the intensity by more than three orders of magnitude. Enhanced luminol chemiluminescence is advantageously employed for determination of horseradish peroxidase in immunoassay.

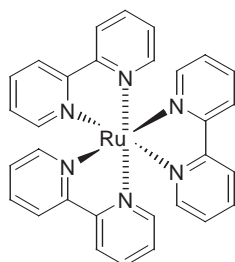
Similar to luminol and lucigenin, both lophine and gallic acid also react with alkaline hydrogen peroxide to yield chemiluminescence. In all of these reactions, the emission intensity is proportional to transition metal ion catalyst concentration over finite ranges. Table 1 provides an abbreviated comparison of the selectivity and sensitivity available with these systems.

### Tris(2,2'-bipyridyl)ruthenium(II)

The photoluminescent properties of selected ruthenium(II) *N,N'*-chelates were reported in 1959 and



although the chemiluminescence from tris(2,2'-bipyridyl)ruthenium(II)  $[\text{Ru}(\text{bipy})_3]^{2+}$  (**10**) was observed in 1962, it was not published until 1966. Much of the subsequent research on this compound has focused on the photochemical oxidation of water as a means of solar energy storage and, consequently, knowledge regarding the chemiluminescent reaction mechanisms is limited.



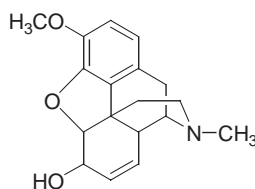
(10)

Chemiluminescence requires the production of tris(2,2'-bipyridyl)ruthenium(III) from the oxidation (chemical or electrochemical) of tris(2,2'-bipyridyl)ruthenium(II). Subsequent reaction with a suitable reducing agent (analyte) elicits light from an excited state of tris(2,2'-bipyridyl)ruthenium(II) (see Scheme 8). This charge transfer luminescence is, in fact, chemically induced phosphorescence (in simple solution) originating from a short-lived  $d\pi^*$  triplet with a lifetime of  $\sim 6 \times 10^{-7}$  s in water and can be regenerated. Excitation is afforded by the promotion of an electron from the  $t_{2g}d^6$  orbital on the ruthenium to the  $\pi^*$  antibonding orbital on the ligand. Intersystem crossing to the triplet state is enhanced by the heavy, paramagnetic ruthenium(II) and the excited state is considered to be a  $d^5 \text{Ru(III)} \text{L}_3^{\bullet-}$ .

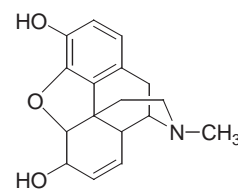
The ruthenium(III) chelate is only moderately stable in acidic aqueous solutions and as such it is normally produced immediately prior to reaction. This can be achieved using oxidants such as chlorine, cerium(IV), or lead dioxide. Lead dioxide is a convenient reagent as no excess oxidant remains in solution. The deep green ruthenium(III) complex can also be produced electrochemically, often with the chelate immobilized on a suitable polymer membrane coated onto a working electrode maintained at  $\sim 1.25$  V. In this configuration, the reagent is easily regenerated. The ruthenium(II) complex has been modified and covalently bound onto silica beads to produce a

chemically regenerable form of the reagent. Alternatively, the anhydrous perchlorate salt of tris(2,2'-bipyridyl)ruthenium(III) is temporally stable as a solid or dissolved in dry acetonitrile.

Although a large variety of compounds can reduce tris(2,2'-bipyridyl)ruthenium(III), only certain species (e.g., aliphatic amines, amino acids, NADH, some alkaloids, aminoglycoside or tetracycline antibiotics, and the oxalate ion) will produce the characteristic orange luminescence with this reagent. Subtle differences in chemical structure can have a dramatic effect on chemiluminescence intensity. This is exemplified by the determination of the papaver alkaloid codeine (**11**) compared to structurally similar morphine (**12**). At pH 6.8, codeine can be determined down to a concentration of  $10^{-11} \text{ mol l}^{-1}$ , whereas morphine produces a chemiluminescent response equivalent to that of the blank. In many applications this degree of selectivity is most desirable.

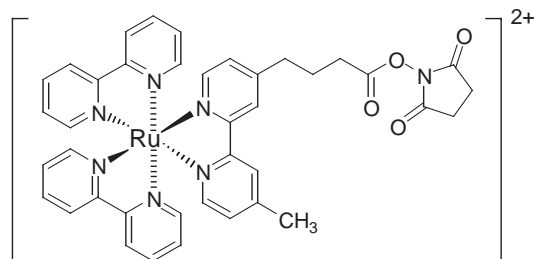


(11)

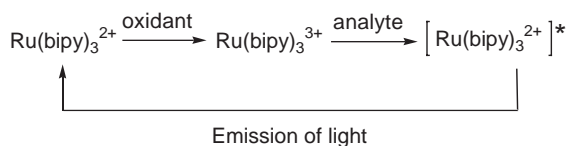


(12)

The chemiluminescence intensity from the reaction of amines with tris(2,2'-bipyridyl)ruthenium(III) is generally in the order tertiary > secondary > primary, but no definitive mechanisms have been elucidated. Derivatives of tris(2,2'-bipyridyl)ruthenium(II) have been introduced as labels for immunoassay or DNA probes (e.g., (**13**)). The use of chemiluminescent labels is clearly advantageous due to their nonradioactive nature and they can be quantified at sub-picomolar levels via oxidation in the presence of tripropylamine.



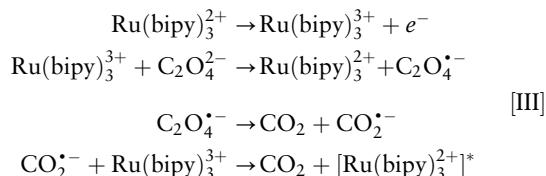
(13)



Scheme 8

The reaction with the oxalate anion ( $\text{C}_2\text{O}_4^{2-}$ ) has been extensively studied; most investigations were electrochemical with the reagent and analyte both present in the cell and chemiluminescence observed at a particular oxidative potential. A feasible reaction

pathway has resulted from this considerable body of work. Initially, the ruthenium chelate is oxidized at a platinum or carbon electrode (reaction [III]); subsequent steps are thought to occur in the diffusion layer near the electrode. It is actually the highly reducing radical anion of carbon dioxide that reacts with the Ru(III) form of the reagent to produce the excited state:



## Practical Considerations

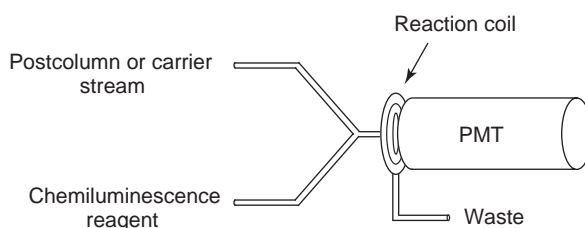
Numerous applications involve coupling liquid-phase chemiluminescence detection to physical or chemical separation processes. Conversely, adequate selectivity can also be achieved for particular analytes in a range of sample matrices through a judicious selection of reagent and reaction conditions. Successful detection strategies have been employed for HPLC, flow analysis, electrophoresis, immunoassay labels, DNA probes, and enzyme reactions.

### HPLC and Flow Analysis

Chemiluminescence detectors for HPLC or flow analysis are often custom built and generally incorporate a flat reaction coil situated flush against the window of a suitable photomultiplier tube, as shown in Figure 1.

The distance (and hence volume) between the confluence point of the sample and reagent streams and the detection cell needs to be optimized for the kinetics of the reaction used. The flow cell must be transparent to the wavelength of the chemiluminescence emission and inert to the chemical reaction or solvent system; glass, quartz, and Teflon tubing are commonly employed. Instrumentation for electrogenerated chemiluminescence requires suitable electrodes and a potentiostat to facilitate and control the electrochemistry. Gold, platinum, or carbon working electrodes are placed in the observation cell with counter and reference electrodes situated downstream.

For HPLC, the compatibility between the chemical parameters necessary for efficient separation and sensitive detection is of paramount importance. A compromise between two optima is often required. For example, peroxyoxalate reactions need alkaline conditions but reverse-phase separations (using silica-based stationary phases) require an eluent pH



**Figure 1** Chemiluminescence detection for HPLC or flow analysis using a photomultiplier tube (PMT).

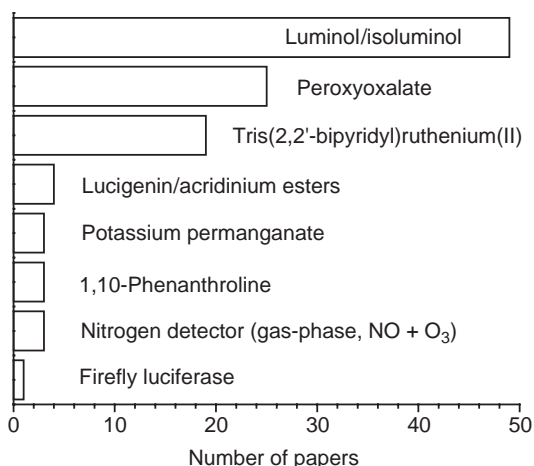
below 6.5. Differences in solvent polarity can also be a problem. Large amounts of water can quench peroxyoxalate emission through reagent insolubility and decomposition.

There has been some investigation into the consumption of immobilized chemiluminescence reagents in flow-through reactors. In flow injection and HPLC, the peroxyoxalate reagent, TCPO, has been used as a packed bed of solid reagent for hydrogen peroxide and fluorophore determinations; the solvent slowly dissolves TCPO from the reactor. Luminol has been immobilized on small particles (silica, nylon, and carbon) contained in flow-through reactors for use in hydrogen peroxide determinations. In all of these schemes luminol is released as required. The TCPO and luminol reactors are eventually exhausted but hundreds of samples can be assayed without the normal requirement of pumping and mixing a separate reagent solution.

Some work has been done with fiber-optic-based chemiluminescence sensors. Luminol has been used with various enzymes immobilized on the end of an optical fiber. Sensors based on electrogenerated chemiluminescence with luminol or tris(2,2'-bipyridyl)ruthenium(II) have appeared; the latter is particularly attractive due the regenerable nature of the reagent. Importantly, for these applications of liquid-phase chemiluminescence, the reactions are occurring at a reactive surface rather than in a homogeneous solution and as such mass transfer probably controls the overall response.

### Capillary Electrophoresis

Since the first publication by Hara and co-workers in 1991, most of the analytically important chemiluminescent reactions have been employed for detection with capillary electrophoresis, as shown in Figure 2. Although chemiluminescence is an attractive mode of detection for capillary electrophoresis, interfacing the detection system with the electrophoretic separation is more complex than with HPLC. Nevertheless, a number of chemiluminescence detectors have been developed and successfully employed for the



**Figure 2** Number of papers that have been published (by mid-2003) on various chemiluminescence reactions used for detection in capillary electrophoresis.

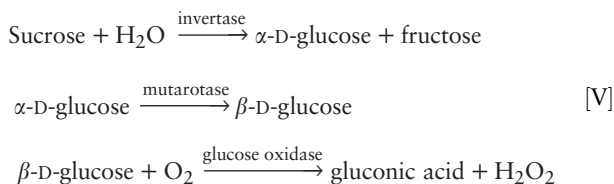
determination of compounds such as amino acids, catecholamines, metal ions, and proteins. The interface design can be categorized as merging flow, coaxial flow, and reservoir mixing, based on the way the chemiluminescence reaction is initiated, or as off-, on-, and end-column based on the site of detection and if it is isolated from the electrophoresis high-voltage supply. Chemiluminescence has also been investigated as a method of detection for microchip-based capillary electrophoresis.

### Enzyme Reaction Products

Coupling chemiluminescence reactions with immobilized enzymes enables the detection of a variety of substrates. A useful example is oxidase enzymes that generate hydrogen peroxide, which can be quantified with several chemiluminescence systems. Substrates including glucose, cholesterol, choline, amino acids, aldehydes, lactate, and uric acid (reaction [IV]) can be determined at the nanomolar level:



This approach can be extended with sequential enzyme steps to ultimately convert an analyte to a desired reagent. In this way, sucrose, maltose, lactose, fructose, glucosides, cholesterol esters, creatinine, and acetylcholine have been measured (reaction [V]):



The most commonly employed system for the determination of hydrogen peroxide under these conditions

is luminol with a peroxidase catalyst. Nevertheless, there are limitations: a high blank signal, interference by other oxidants (O<sub>2</sub>, OCl<sup>-</sup>, O<sub>2</sub><sup>-</sup>), and quenchers present in biological samples. If the solvent and pH requirements of the enzyme are incompatible with reaction conditions, a change after the enzyme reaction will be necessary. There are other enzyme systems (such as  $\beta$ -glucuronidase with glucuronide metabolites, which forms glucuronic acid) that can be detected with lucigenin. Likewise, glucose dehydrogenase reacts with glucose (and NAD<sup>+</sup>) to form NADH, which can be detected with tris(2,2'-bipyridyl)ruthenium(III). After separation by ion chromatography, acetylcholine and choline react with immobilized enzymes to produce hydrogen peroxide, which can be subsequently detected with TCPO and perylene.

### Immunoassay and DNA Assays

A chemiluminescence application that has become much more prevalent and important in recent years is the detection of proteins and oligonucleotides in immunoassay and DNA probe assays (DNA fingerprinting, DNA sequencing, and detection of DNA following electrophoresis and blotting). Nonradioactive labels are of great interest in these areas, and chemiluminescence is an attractive option, owing to advantages of sensitivity (equal to or better than radioactive labels like <sup>125</sup>I), dynamic range, simple instrumentation, long shelf-life, and low cost. In chronological order, the chemiluminescence systems that have found application are luminol, acridinium esters, dioxetanes, and tris(2,2'-bipyridyl)ruthenium(II). The assay can involve either direct labeling with the chemiluminescence species or indirect labeling with a species that catalyzes a chemiluminescence reaction. After incubation (to achieve binding) and then separation of the bound and unbound materials (if necessary), suitable reagents are added to initiate the chemiluminescence reaction. Reagents for a variety of analytes are commercially available.

Direct labels include luminol or aminobutylethylisoluminol, acridinium esters, and tris(2,2'-bipyridyl)ruthenium(II). Luminometers designed for these measurements add the appropriate trigger solution directly in the sample compartment, as the chemiluminescence emission is a short flash lasting from 1 to 5 s. Although one can monitor peak intensity, it is also common to integrate the entire light output. In general, each label molecule reacts only once and will produce only one photon (actually fewer, given the quantum efficiency). An exception to this rule is tris(2,2'-bipyridyl)ruthenium(II), as it can be continuously recycled and re-excited. With acridinium ester labels, the chemiluminescence reaction

results in cleavage of the acridinium ring portion prior to emission. As a result, the labeled material shows the same reaction kinetics and emission properties as the free label.

Catalysts for chemiluminescence reactions can be used as indirect labels, each of which catalyzes the reaction of many substrate molecules and leads to the emission of many photons. For enzyme immunoassay, the most common labels are horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, and for DNA detection the most common nonisotopic label is alkaline phosphatase. Chemiluminescence systems are available for all these: enhanced luminol for horseradish peroxidase, and dioxetanes for alkaline phosphatase and  $\beta$ -galactosidase. If a sufficiently large concentration of the chemiluminescence reagent is added so that its concentration remains essentially constant during measurement, the rate of the chemiluminescence reaction is limited by the enzyme catalyst concentration. Emission is then in the form of a 'continuous' glow that lasts many minutes to hours and it is no longer necessary for instrumentation or procedures to provide reaction initiation directly in front of detector. Consequently, multiple recordings of emission intensity can be made on microtiter plates or electrophoresis gels. Quantification with enzyme labels generally involves monitoring peak intensity or integration over a fixed time rather than over the entire emission intensity-time profile. As the concentration of the catalyst changes, the rate of reaction changes but the equilibrium position remains constant.

**See also:** **Enzymes:** Immobilized Enzymes. **Flow Injection Analysis:** Detection Techniques. **Forensic Sciences:** DNA Profiling. **Immunoassays, Techniques:** Luminescence Immunoassays. **Liquid Chromatography:** Column Technology; Instrumentation.

## Further Reading

- Blum LJ and Gautier SM (1991) Bioluminescence- and chemiluminescence-based fibre optic sensors. In: Blum LJ and Coulet PR (eds.) *Biosensor Principles and Applications*, pp. 213–247. New York: Dekker.
- Campbell AK (1988) *Chemiluminescence; Principles and Applications in Biology and Medicine*. New York: VCH.
- De Jong GJ and Kwakman PJM (1989) Chemiluminescence detection for high-performance liquid chromatography of biomedical samples. *Journal of Chromatography* 492: 319–343.
- García-Campaña AN and Baeyens WRG (eds.) (2001) *Chemiluminescence in Analytical Chemistry*. New York: Dekker.
- Gerardi RD, Barnett NW, and Lewis SW (1999) Analytical applications of tris(2,2'-bipyridyl)ruthenium(III) as a chemiluminescent reagent. *Analytica Chimica Acta* 378: 1–41.
- Hindson BJ and Barnett NW (2001) Analytical applications of acidic potassium permanganate as a chemiluminescent reagent. *Analytica Chimica Acta* 445: 1–19.
- Huang XJ and Fang ZL (2000) Chemiluminescence detection in capillary electrophoresis. *Analytica Chimica Acta* 414: 1–14.
- Kricka LJ and Carter TJN (eds.) (1982) *Clinical and Biochemical Luminescence*. New York: Dekker.
- Nieman TA (1991) Analytical chemiluminescence in flowing streams; FIA and HPLC. In: Baeyens WRG, DeKeukeleire D, and Korkidas K (eds.) *Luminescence Techniques in Chemical and Biochemical Analysis*, pp. 523–565. New York: Dekker.
- VanDyke K and VanDyke R (eds.) (1990) *Luminescence Immunoassay and Molecular Applications*. Boca Raton: CRC Press.
- Weeks I (1992) *Chemiluminescence Immunoassay*. Amsterdam: Elsevier.
- Yamaguchi M, Yoshida H, and Nohta H (2002) Luminol-type chemiluminescence derivatization reagents for liquid chromatography and capillary electrophoresis. *Journal of Chromatography A* 950: 1–19.

## Gas-Phase

**N W Barnett and P S Francis**, Deakin University, Geelong, VIC, Australia  
**J S Lancaster**, BP Chemicals, Hull, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The earliest recorded example of chemiluminescence occurring in the gas phase is the oxidation of

phosphorus vapor just above the solid surface by molecular oxygen. Hennig Brandt observed the green emission from this reaction in the seventeenth century. The reaction mechanism is not fully characterized, but the emitting species have been identified as  $(\text{PO})_2$  and  $\text{HPO}$ . Gas-phase chemiluminescence reactions are now known to occur naturally in the upper atmosphere where ultraviolet (UV) radiation produces radicals and ions. Recombination of these reactive oxygen and nitrogen species yields one or

results in cleavage of the acridinium ring portion prior to emission. As a result, the labeled material shows the same reaction kinetics and emission properties as the free label.

Catalysts for chemiluminescence reactions can be used as indirect labels, each of which catalyzes the reaction of many substrate molecules and leads to the emission of many photons. For enzyme immunoassay, the most common labels are horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, and for DNA detection the most common nonisotopic label is alkaline phosphatase. Chemiluminescence systems are available for all these: enhanced luminol for horseradish peroxidase, and dioxetanes for alkaline phosphatase and  $\beta$ -galactosidase. If a sufficiently large concentration of the chemiluminescence reagent is added so that its concentration remains essentially constant during measurement, the rate of the chemiluminescence reaction is limited by the enzyme catalyst concentration. Emission is then in the form of a 'continuous' glow that lasts many minutes to hours and it is no longer necessary for instrumentation or procedures to provide reaction initiation directly in front of detector. Consequently, multiple recordings of emission intensity can be made on microtiter plates or electrophoresis gels. Quantification with enzyme labels generally involves monitoring peak intensity or integration over a fixed time rather than over the entire emission intensity-time profile. As the concentration of the catalyst changes, the rate of reaction changes but the equilibrium position remains constant.

**See also:** **Enzymes:** Immobilized Enzymes. **Flow Injection Analysis:** Detection Techniques. **Forensic Sciences:** DNA Profiling. **Immunoassays, Techniques:** Luminescence Immunoassays. **Liquid Chromatography:** Column Technology; Instrumentation.

## Further Reading

- Blum LJ and Gautier SM (1991) Bioluminescence- and chemiluminescence-based fibre optic sensors. In: Blum LJ and Coulet PR (eds.) *Biosensor Principles and Applications*, pp. 213–247. New York: Dekker.
- Campbell AK (1988) *Chemiluminescence; Principles and Applications in Biology and Medicine*. New York: VCH.
- De Jong GJ and Kwakman PJM (1989) Chemiluminescence detection for high-performance liquid chromatography of biomedical samples. *Journal of Chromatography* 492: 319–343.
- García-Campaña AN and Baeyens WRG (eds.) (2001) *Chemiluminescence in Analytical Chemistry*. New York: Dekker.
- Gerardi RD, Barnett NW, and Lewis SW (1999) Analytical applications of tris(2,2'-bipyridyl)ruthenium(III) as a chemiluminescent reagent. *Analytica Chimica Acta* 378: 1–41.
- Hindson BJ and Barnett NW (2001) Analytical applications of acidic potassium permanganate as a chemiluminescent reagent. *Analytica Chimica Acta* 445: 1–19.
- Huang XJ and Fang ZL (2000) Chemiluminescence detection in capillary electrophoresis. *Analytica Chimica Acta* 414: 1–14.
- Kricka LJ and Carter TJN (eds.) (1982) *Clinical and Biochemical Luminescence*. New York: Dekker.
- Nieman TA (1991) Analytical chemiluminescence in flowing streams; FIA and HPLC. In: Baeyens WRG, DeKeukeleire D, and Korkidas K (eds.) *Luminescence Techniques in Chemical and Biochemical Analysis*, pp. 523–565. New York: Dekker.
- VanDyke K and VanDyke R (eds.) (1990) *Luminescence Immunoassay and Molecular Applications*. Boca Raton: CRC Press.
- Weeks I (1992) *Chemiluminescence Immunoassay*. Amsterdam: Elsevier.
- Yamaguchi M, Yoshida H, and Nohta H (2002) Luminol-type chemiluminescence derivatization reagents for liquid chromatography and capillary electrophoresis. *Journal of Chromatography A* 950: 1–19.

## Gas-Phase

**N W Barnett and P S Francis**, Deakin University, Geelong, VIC, Australia  
**J S Lancaster**, BP Chemicals, Hull, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The earliest recorded example of chemiluminescence occurring in the gas phase is the oxidation of

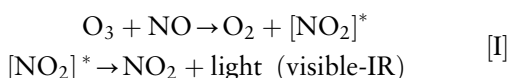
phosphorus vapor just above the solid surface by molecular oxygen. Hennig Brandt observed the green emission from this reaction in the seventeenth century. The reaction mechanism is not fully characterized, but the emitting species have been identified as  $(\text{PO})_2$  and  $\text{HPO}$ . Gas-phase chemiluminescence reactions are now known to occur naturally in the upper atmosphere where ultraviolet (UV) radiation produces radicals and ions. Recombination of these reactive oxygen and nitrogen species yields one or

more products in an electronically excited state. Relaxation of the excited state by radiative emission (fluorescence or phosphorescence) completes the sequence of reactions leading to chemiluminescence.

Chemiluminescence reactions generally offer three major advantages in analytical applications: (1) wide dynamic range, (2) low detection limits, and (3) high selectivity. The nitric oxide–ozone reaction, for example, fulfills all these criteria with linearity over six orders of magnitude, detection limits down to low parts per billion ( $\mu\text{g per kg}$ ) or less, and little interference from other species. Gas-phase chemiluminescence reactions have several applications, particularly in the field of atmospheric analysis for nitrogen- and sulfur-containing pollutants, and as a highly sensitive means of detection in gas chromatography (GC). One of the most widely used chemiluminescence reagent gases is ozone ( $\text{O}_3$ ) and we shall begin this review with a discussion on the reaction of ozone with nitrogen compounds, sulfur compounds, and hydrocarbons.

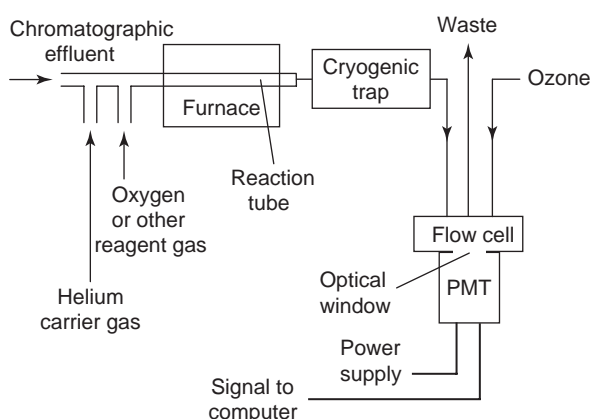
## Chemiluminescence of Nitrogen Compounds

Nitric oxide gives a chemiluminescence emission when reacted with ozone to yield electronically excited nitrogen dioxide (reaction [I]):



This reaction is one of the most well-known and widely used gas-phase chemiluminescence reactions. The emission, which is broadly distributed from 600 to 3000 nm with a maximum intensity at 1200 nm, appears to originate from a combination of the  $^2\text{B}_1$  and  $^2\text{B}_2$  states of nitrogen dioxide. Commercial ozone-mediated nitrogen-specific detectors are fitted with a red filter that allows the emission from excited nitrogen dioxide to pass to the photomultiplier tube while absorbing light from interfering species. To minimize the nonradiative decay of excited nitrogen dioxide molecules, the reaction chamber is operated under reduced pressure, typically 1–100 mmHg.

Detection of compounds other than nitric oxide requires conversion of the analytes (either directly or indirectly) into nitric oxide. The so-called ‘ $\text{NO}_x$  box’ that is frequently used in atmospheric field studies to determine both NO and  $\text{NO}_2$ , requires measurement with and without conversion of  $\text{NO}_2$  to NO by photolysis. Similarly, the ‘total reactive oxides of nitrogen’ can be converted to NO by passing the

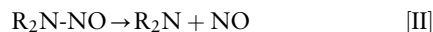


**Figure 1** Ozone-mediated chemiluminescence detector. The chromatographic effluent reacts with oxygen or a reagent gas in the furnace. The excited-state species is generated by reaction with ozone, and the emission is detected by a photomultiplier tube (PMT).

sample through a heated molybdenum tube with a catalytically active oxide surface, before quantification with a  $\text{NO}_x$  box. For GC, the conversion of analytes involves a postcolumn reaction. The resulting stream may then be passed through a cryogenic trap to remove less-volatile species, before mixing with ozone in the reduced-pressure reactor. Three commercially available detectors that are based on the nitric oxide–ozone reaction are discussed below. **Figure 1** shows the components of the basic detector.

## Thermal Energy Analyzer

The thermal energy analyzer (TEA) was designed to selectively detect *N*-nitrosoamines, based on their decomposition to nitric oxide (reaction [II]). In this method, the term ‘thermal’ refers to the heat required for decomposition, rather than the mode of detection. A catalyst is often employed to detect a wider variety of compounds at lower temperatures:



The nitric oxide is then determined by its reaction with ozone (reaction [I]). This reaction is specific and can be used for screening purposes without any kind of separation. Alternatively, it can be used as a detector for gas or liquid chromatography where the effluent passes through a pyrolysis chamber and the *N*-nitrosoamines decompose to form nitric oxide. The pyrolysis chamber consists of a quartz or ceramic tube with the catalyst coated onto the walls, or packed into the tube. The most effective catalyst has been found to be a mixture of tungsten oxides.



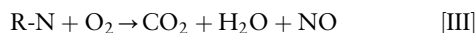
The temperature of the pyrolysis tube affects the selectivity and must therefore be reproducibly controlled. The nitric oxide radical is cleaved from *N*-nitrosoamines at relatively low temperatures owing to the weak N–NO bond. Cold traps and packed columns of a solid phase such as Tenax are placed between the pyrolysis tube and the detector to remove potentially interfering compounds, while allowing nitric oxide to pass.

*N*-nitrosoamines can undergo pyrolytic decomposition in the absence of a catalyst, to yield nitric oxide, though the required temperature is higher. This can result in a loss of selectivity in the presence of nitro-compounds. The noncatalyzed system is used because of its simplicity in cases where selectivity is less important, for example, in food testing. The TEA is now the detector of choice for the determination of *N*-nitrosoamines in the environment, foods, and tobacco products.

### Nitrogen-Selective Detector

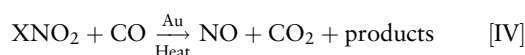
Most nitrogen-containing compounds can be converted to nitric oxide at temperatures between 800°C and 1100°C in the presence of a catalyst. This type of detector, known as a nitrogen-selective detector (NSD) or chemiluminescent nitrogen detector (CLND), is similar to the TEA, except that a stream of oxygen is added to the column effluent before it enters the pyrolysis tube (reaction [III]). Compounds with higher oxidation states of nitrogen are formed in addition to nitric oxide, but at the temperatures employed they thermally decompose to nitric oxide. The nitric oxide is then determined by its reaction with ozone (reaction [I]). Detection limits for the TEA and NSD are  $\sim 1 \times 10^{-12} \text{ mol l}^{-1}$  of carbon-bound nitrogen and the detector response is linear over at least four orders

of magnitude:



### Redox Chemiluminescence Detection

The redox chemiluminescence detector (RCD) is also based on the nitric oxide–ozone reaction. However, unlike the two detectors previously described, the analytes do not contain any oxidizable nitrogen. In this case a nitrogen-containing reagent such as nitrogen dioxide or nitric acid is reduced to nitric oxide by the analyte at a heated gold surface (reaction [IV]):



This detector can therefore be used for any reducing species that can react with nitrogen dioxide, or several oxidized nitrogen species that react with carbon monoxide. Conversion of  $\text{XNO}_2$  compounds occurs rapidly and quantitatively where X is an alkyl, alkoxy, or hydroxyl group and where carbon monoxide is in excess. The formation of nitric oxide is rate limiting and the degree of conversion can be altered by selecting the appropriate catalysts and temperatures. The detector can therefore be made to respond selectively to a particular compound or class of compounds.

A typical gas chromatograph-RCD consists of a nitrogen dioxide inlet, a carbon monoxide inlet, a heated redox reaction zone with a suitable catalyst, and a nitric oxide/ozone reactor for chemiluminescence generation. The RCD, however, has not gained wide acceptance. This may be partly due to the tendency of the gold catalyst to become poisoned and require conditioning with oxygen at high temperature. Table 1 lists some applications of the three

**Table 1** Some applications of nitrogen oxide–ozone chemiluminescence

Analyte	Method	Reference
Total nitrosamines in food extracts	TEA	<i>Journal of AOAC International</i> , 78 (1995) 1435
Nitrosamines in tobacco and smoke	GC-TEA	<i>Cancer Letters</i> , Shannon, Ireland, 97 (1995) 1
Nitrate in foods	GC-TEA	<i>Journal-Association of Official Analytical Chemists</i> , 68 (1985) 41
Nitroaromatics in explosives	SGC-TEA	<i>Journal of Chromatography A</i> , 902 (2000) 413
Involatile nitrosamines	HPLC-TEA	<i>Journal of Chromatography</i> , 328 (1985) 362
Nitro- and nitroso-compounds	SFC-TEA	<i>Journal of Microcolumn Separations</i> , 6 (1994) 395
Amines in air	GC-NSD	<i>Journal of Chromatography</i> , 239 (1982) 617
Nitrogen pyrolysis products from soil	GC-NSD	<i>Biology and Fertility of Soils</i> , 20 (1995) 174
Thiotriazine compounds in urine	GC-NSD	<i>Analytica Chimica Acta</i> , 424 (2000) 7
Nucleotides and nucleosides in foods	HPLC-NSD	<i>Developmental Food Science</i> , 37A (1995) 379
Ammonia, nitrite and nitrate	FIA-NSD	<i>Analytica Chimica Acta</i> , 349 (1997) 11
Nitric oxide from nitrovasodilator drugs	RCD	<i>Journal of Pharmacological Methods</i> , 25 (1991) 19
Carbon monoxide	GC-RCD	<i>Journal of Chromatography</i> , 395 (1987) 9
Reducing sugars	HPLC-RCD	<i>Journal of Chromatography</i> , 441 (1988) 125

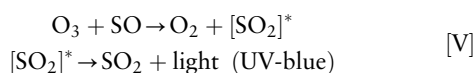
TEA, thermal energy analyzer; NSD, nitrogen selective detector; RCD, redox chemiluminescence detector.



types of chemiluminescence detectors for nitrogen compounds.

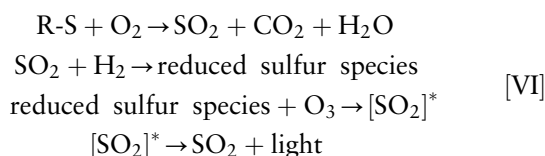
## Chemiluminescence of Sulfur Compounds

The oxidation of sulfur monoxide with ozone (reaction [V]) is similar to that of nitric oxide (reaction [I]), but it is more exothermic. The chemiluminescence emission spectrum accompanying this reaction extends from about 260 to 480 nm. At least four excited states of sulfur dioxide are thought to be involved;  $^3B_1$ ,  $^1A_2$ ,  $^1B_1$ , and  $^1B_2$ , which have radiative lifetimes of 8 ms, 30  $\mu$ s, 600  $\mu$ s, and 30 ns, respectively:



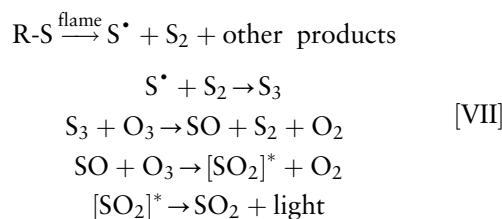
This reaction enables the detection of sulfur monoxide and other reduced sulfur compounds that can react with ozone to initially form sulfur monoxide. The chemiluminescence intensity depends on the type of compound, with thiols giving the largest response followed by alkyl sulfides, hydrogen sulfide, and thiophenes. Detection limits in the parts per billion range can be achieved for thiols. Alkenes, which also react with ozone to give emission centered at 354 nm, potentially interfere but selectivity can be achieved with a suitable optical filter.

Universal sulfur-selective detectors for GC or supercritical fluid chromatography (SFC), that incorporate a cool hydrogen-rich flame or closed hydrogen/air burner have been developed and are now commercially available. Combustion products are transferred via a connecting line to a low-pressure reaction cell, where they are mixed with ozone. Although sulfur monoxide is believed to be the common intermediate that reacts with ozone to form the excited sulfur dioxide emitter, it is not necessarily the product of the first conversion step, which has caused debate between instrument manufacturers. One proposed mechanism proceeds as shown in reaction [VI] and it was suggested that the reduced sulfur species could be  $H_2S$ :



Another group ruled out hydrogen sulfide and many other sulfur species as intermediates and tentatively proposed that  $S_3$  forms in the transfer line from

the reaction of sulfur atoms with diatomic sulfur molecules (reaction [VII]):



Sulfur chemiluminescence detectors (SCD) are highly selective; a signal seven orders of magnitude greater than potential interfering species has been reported. In addition, an equimolar response from all sulfur compounds is observed, as the emitter contains only one sulfur atom. These detectors are rapidly becoming the method of choice for the analysis of many sulfur-containing compounds. Furthermore, as nitric oxide is also formed in the SCD and it survives the hydrogen reduction step (reaction [VI]), sulfur and nitrogen can be determined simultaneously. The column effluent can also be split and directed toward an FID, but compromises between the three systems can lead to a reduction in performance.

## Chemiluminescence of Hydrocarbons

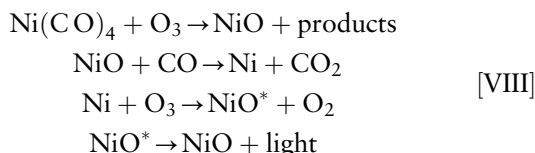
Ozone reacts rapidly with alkenes to give a chemiluminescence emission from several excited-state species. Excited species identified as emitters in the ozone oxidation of ethene include formaldehyde at 350–520 nm ( $^1A_2$ ) and hydroxyl radicals at 700–1100 nm ( $X^2\Pi$ ,  $v \leq 9$ ) and 306 nm ( $A^2\Sigma^+$ ). In addition, phosphorescence is observed from glyoxal and methyl glyoxal ( $^3A_u \rightarrow ^1A_g$ ) when substituted alkenes are oxidized.

Alkanes and aromatic compounds are less readily oxidized than alkenes, but can be made to react with ozone at higher temperatures to give a chemiluminescence emission. This difference in reactivity can be exploited to increase selectivity. For example, a postcolumn reactor at 100°C allows only the detection of alkenes, but when the temperature is increased to ~150°C aromatic compounds will also elicit chemiluminescence and at 250°C all other hydrocarbons will give rise to emission. Aromatic compounds are likely to react in a similar way to alkenes, by forming an ozonide intermediate. The reaction pathway for the ozone-mediated chemiluminescence of alkanes is unknown and results in much lower chemiluminescence intensity. The detector response is linear when ozone is present in excess and absolute

detection limits for alkenes are reported to be in the nanogram range.

## Other Reactions with Ozone

The concentration of nickel carbonyl in air can be determined using the chemiluminescence reaction with ozone and purified carbon monoxide (reaction [VIII]). The emission intensity is measured at  $\sim 500$  nm and the detection limit is 2 ppbv:

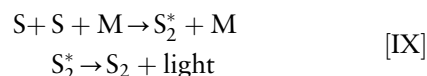


The reaction between ozone and arsine leads to an emission continuum in the visible region and a discrete emission in the UV. The hydrides of arsenic, antimony, and tin can be determined down to sub-ppb levels. The chemiluminescent reaction between ozone and phosphine has been employed to determine phosphate and other phosphorus-containing compounds after reduction. The chemiluminescent oxidation of silane was used to determine silicate in natural waters. A variety of applications involving ozone as a reagent gas have been included in Table 2.

## Flame Photometric Detector

Chemiluminescence emission may occur in cool flames at temperatures below that required for thermal excitation. This is the basis of the flame

photometric detector (FPD). An example of this is the molecular emission of sulfur in cool ( $<1000^\circ\text{C}$ ) hydrogen-rich flames. Sulfur atoms are formed in the flame and recombine to form electronically excited  $\text{S}_2$  ( $\text{B}^3\Sigma_u^-$ ), which emits a photon to return to the ground state ( $\text{X}^3\Sigma_g^-$ ), as shown in reaction [IX], where M is another atom or molecule. The reaction forming the emitter is second order with respect to sulfur atoms and therefore the signal has a quadratic dependence on analyte concentration:



Organophosphorus compounds can also be detected; they react to give excited HPO. The same detector can be made selective to phosphorus and sulfur-containing species by using the appropriate optical filters with dual photomultipliers. Emission from the excited  $\text{S}_2$  molecule is monitored at 394 nm and that from HPO at 526 nm. This detector is subject to quenching and analyte dependency. Two flames can be used to separate the region of sample decomposition and excitation. The first flame breaks down large molecules in the column effluent into smaller fragments that reportedly do not quench the emission. The excited-state sulfur and phosphorus species are formed in the second flame. The dual flame design results in greatly improved selectivity with respect to hydrocarbons. Selectivity ratios reported for phosphorus and sulfur are of  $10^{-5}$  and up to  $10^{-6}$ , respectively.

Although determinations of sulfur and phosphorus compounds constitute the main applications of the FPD, there are  $\sim 20$  emitters that produce analytically useful signals. Nitrogen-containing

**Table 2** Some applications of other chemiluminescence detectors

Analyte	Detector	Reference
Sulfur compounds in coal tar	SCD (ozone)	<i>Chromatographia</i> , 33 (1992) 507
Volatile sulfur compounds in wines	SCD (ozone)	<i>Analisis</i> , 26 (1998) 142
Aromatic sulfur compounds in gas oils	SCD (ozone)	<i>Analytical Chemistry</i> , 74 (2002) 3849
Sulfur compounds in irradiated foods	PFPD	<i>Journal of Agricultural and Food Chemistry</i> , 50 (2002) 4257
Sulfur and phosphorus in chemical weapons	PFPD	<i>Field Analytical Chemistry and Chemical Technology</i> , 4 (2000) 170
Phosphorus and chlorine compounds in fats	PFPD	<i>Journal of Separation Science</i> , 25 (2002) 527
Organotin compounds	PFPD	<i>Journal of Chromatography A</i> , 999 (2003) 123
Potassium content of soil	FPD	<i>Environment and Ecology</i> , 20 (2002) 987
Nitrous oxide ( $\text{N}_2\text{O}$ )	FPD	<i>Journal of Chromatography A</i> , 598 (1992) 313
Polyhalogenated hydrocarbons	Sodium vapor	<i>Journal of Chromatography</i> , 238 (1982) 347
Trimethyl aluminum	Active nitrogen	<i>Analytical Chemistry</i> , 51 (1979) 1399
Phosphates (as phosphines)	Ozone	<i>Analytical Chemistry</i> , 61 (1989) 2699
Silicate in natural waters (as silane)	Ozone	<i>Analytical Chemistry</i> , 65 (1993) 1814
Hydrocarbons	Ozone	<i>Journal of Chromatography</i> , 112 (1975) 253
Atmospheric hydrocarbons	Ozone	<i>Atmospheric Environment</i> , 32 (1998) 1435
Halogenated hydrocarbons	Ozone	<i>Analytica Chimica Acta</i> , 451 (2002) 189

SCD, sulfur chemiluminescence detector; PFPD, pulsed flame photometric detector; FPD, flame photometric detector.

organic compounds react in an analogous manner to phosphorus, to form excited HNO, which emits at  $\sim 690$  nm. Compounds containing manganese or ruthenium exhibit photon yields (from  $\text{Mn}^*$  and an unknown emitter) that are comparable with those of sulfur and phosphorus. Iron compounds produce a rich spectrum of atomic lines from  $\text{Fe}^*$  at a temperature of only  $346^\circ\text{C}$ . Selenium and tellurium compounds, like those of sulfur, produce diatomic emitters ( $\text{Se}_2^*$  and  $\text{Te}_2^*$ ) and therefore a nonlinear relationship between concentration and intensity is observed.

The pulsed FPD (PFPD) is an improved FPD that was developed during the past decade. The PFPD uses a pulsed flame and time-resolved emission detection, which in addition to providing greater selectivity, has substantially improved the overall performance. This sensitivity of this multianalyte detector is similar to those specifically designed for nitrogen/sulfur chemiluminescence, but the selectivity for nitrogen and sulfur over hydrocarbons is still much lower. This approach has been extensively applied to organotin and organophosphorus compounds. GC-PFPD can be coupled with mass spectrometry to provide even greater information. Some applications of PFPD are shown in Table 2.

## Other Reagent Gases

### Molecular Fluorine

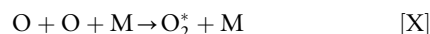
Fluorine reacts with certain organosulfur compounds to generate excited-state species via an unknown pathway. The reaction is highly exothermic owing to the fission of the very weak F–F bond, and the formation of strong H–F and C–F bonds. This is the basis for a highly selective chemiluminescence detector for reduced sulfur compounds. The chemiluminescence emission in the red and near-infrared is due to the formation of vibrationally excited hydrogen fluoride. The fluorine chemiluminescence detector (FCLD) has been successfully interfaced with GC, high-performance liquid chromatography (HPLC), and SFC, and shown to be analytically useful for thiols, sulfides, disulfides, and thiophenes, with detection limits in the low picogram range. Thiophenes are rather less reactive owing to the stability imparted by their aromaticity, and their detection limits are an order of magnitude poorer. The detector response is linear in most cases over more than three orders of magnitude. This detector has also been used for organotellurium and organoselenium compounds, phosphines, alkyl phosphines, and phosphinate esters.

### Atomic Fluorine

Atomic fluorine can be generated by microwave discharges in  $\text{F}_2$ ,  $\text{CF}_4$ , or  $\text{SF}_6$ . Chemiluminescence from the reaction of fluorine atoms with hydrocarbons is almost universal and results from the production of vibrationally excited HF via hydrogen abstraction. Although the HF overtone band at 880 nm can be used to detect many species that contain hydrogen, this reaction also generates chemiluminescence from  $\text{C}_2$  and CH at 470 and 431 nm, respectively, which provides selective detection of hydrocarbons. Other classes of compounds can be selectively monitored. For example, iodo compounds react to produce excited IF, which has been monitored at 580 nm, with reported detection limits of  $\sim 1$   $\mu\text{g}$ .

### Atomic Oxygen

Oxygen atoms are more powerful oxidants than ozone, and produce chemiluminescence with a wide range of analytes. Nevertheless, the analytical application of this reagent is very limited. One reason for this is that the five different excited states of molecular oxygen formed in the recombination of oxygen atoms (reaction [X]) produce a background emission across the entire visible region that increases quadratically with oxygen atom concentration:



The reaction between oxygen atoms and nitric oxide produces a continuum between 400 and 1400 nm from excited nitrogen dioxide. These are significantly lower wavelengths than those of the previously discussed reaction between nitric oxide and ozone. This reaction has been used to determine oxygen atoms in kinetic experiments. As with the oxidation of sulfur monoxide with ozone, oxidation with oxygen atoms produces sulfur dioxide in electronically excited states. In this case, the emission is distributed from 240 to 400 nm with a maximum at  $\sim 270$  nm.

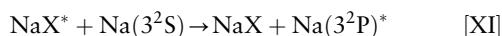
### Active Nitrogen

Passing molecular nitrogen ( $\text{X}^1\Sigma_g^+$ ) through a microwave or electrical discharge produces ‘active nitrogen’, which is a cool plasma of atoms ( $^4\text{S}$ ) and excited molecules ( $\text{A}^3\Sigma_u^+$  and  $\text{B}^3\Pi_g$ ). These nitrogen species undergo collisional energy transfer with saturated hydrocarbons, leading to excited states of these species. In addition, the active nitrogen can undergo a variety of abstraction and addition reactions with unsaturated molecules, resulting in excited-state cyanogen radicals. This is the basis

for a universal chemiluminescence response to hydrocarbons. Detection limits in the nanogram range have been reported for saturated hydrocarbons, whilst for unsaturated analytes detection limits approaching 100 pg have been reported. Compounds other than hydrocarbons (including organometallics, halogens, and compounds containing oxygen, phosphorus, or sulfur) can be detected by monitoring different emitters. Organometallic compounds have been shown to produce excited metal species in addition to cyanogen radicals, and by monitoring the characteristic atomic emission line of the metal of interest some selectivity can be achieved.

### Atomic Sodium

Atomic sodium, generated by heating the metal to 400°C, has been used for the specific detection of halocarbons containing more than one halogen atom in a low-pressure postcolumn reactor. The reaction pathway leading to chemiluminescence is thought to involve abstraction of the halogen atoms to form excited sodium halide and an alkene. Collision between the excited sodium halide (NaX) and another sodium atom produces a doublet sodium atom (reaction [XI]), which can emit at 589 nm (i.e., the sodium D-line). Detection limits in the low picogram to high femtogram range have been reported for chlorinated and brominated compounds, but fluorinated compounds are much less sensitive owing to the strong C–F bond:



### Chlorine Dioxide

Chlorine dioxide is a highly selective chemiluminescence reagent for two sulfur compounds: hydrogen sulfide and mercaptoethanol. The exact reaction pathway is unknown, although it is thought that sulfur atoms produced in the reaction recombine to form excited S<sub>2</sub> molecules, as in the FPD. A detection limit of 3 µg l<sup>-1</sup> hydrogen sulfide has been quoted.

### Official Methods

Several American and British official analytical methods (ASTM, BSI) and Environmental Protection Agency (EPA) recommended procedures for monitoring industrial and environmental samples utilize

gas-phase chemiluminescence detection. These methods include the trace analysis of nitrogen-containing species in liquid petroleum hydrocarbons, the atmospheric monitoring of ozone by measurement of the chemiluminescence emission due to the ozone oxidation of ethene, and a standard assay for chemically bound nitrogen in natural waters. Most process analyses for the measurement of nitrogen oxides in gaseous effluent streams utilize gas-phase chemiluminescence detection.

**See also:** **Air Analysis:** Outdoor Air. **Food and Nutritional Analysis:** Contaminants. **Gas Chromatography:** Detectors. **Luminescence:** Overview. **Nitrosamines. Ozone. Phosphorus. Polycyclic Aromatic Hydrocarbons:** Determination. Environmental Applications. **Sulfur. Supercritical Fluid Chromatography:** Overview; Applications.

### Further Reading

- Aue WA and Singh H (2001) Chemiluminescent photon yields measured in the flame photometric detector on chromatographic peaks containing sulfur, phosphorus, manganese, ruthenium, iron or selenium. *Spectrochimica Acta Part B* 56: 517–525.
- Belcher R, Bogdanski SL, and Townshend A (1973) Molecular emission cavity analysis. A new flame analytical technique. *Analytica Chimica Acta* 67: 1–16.
- Birks JW (ed.) (1991) *Chemiluminescence and Photochemical Reaction Detection in Chromatography*, pp. 39–95. New York: VCH.
- Boulter JE and Birks JW (2001) Gas-phase chemiluminescence detection. In: García-Campaña AM and Baeyens WRG (eds.) *Chemiluminescence in Analytical Chemistry*, pp. 349–391. New York: Dekker.
- Campbell AK (1988) *Chemiluminescence*. Chichester: Ellis Horwood.
- Mihalatos AM and Calokerinos AC (1995) Ozone chemiluminescence in environmental analysis. *Analytica Chimica Acta* 303: 127–135.
- Verdin A (1973) *Gas Analysis Instrumentation*. New York: Macmillan.
- Yan X (1999) Detection by ozone-induced chemiluminescence in chromatography. *Journal of Chromatography A* 842: 267–308.
- Yan X (2002) Sulfur and nitrogen chemiluminescence detection in gas chromatographic analysis. *Journal of Chromatography A* 976: 3–10.
- Zainullin RF and Berezkin VG (1991) Flame photometric detectors in chromatography: A review. *Critical Reviews in Analytical Chemistry* 22: 183–199.

## Electrogenerated

**J Kankare**, University of Turku, Turku, Finland  
**S Kulmala**, Helsinki University of Technology, Espoo, Finland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

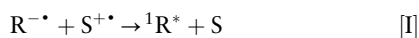
Although the application of chemiluminescent phenomena to various analytical problems is well established, electrogenerated chemiluminescence as an analytical tool has been developed relatively recently. In electrogenerated chemiluminescence (ECL) one or more of the reagents is generated *in situ* in an electrolytic process. ECL shares many analytical advantages with chemiluminescence, the most important one being the low detection limit owing to the low background emission. This article will briefly review the working principles outlined in the most important publications on analytical assays based on ECL and their application areas.

### Electrochemiluminescent Systems

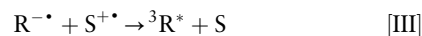
As in conventional chemiluminescence, the change in free energy of the reaction leading to the emitting species needs to be of the order of  $200 \text{ kJ mol}^{-1}$  in order to produce emission in the visible range. This, together with the structural demands on the emitting species makes chemiluminescent reactions rather rare. Methods based on ECL have the advantage, however, that electrolytic processes are capable of producing highly energetic, although unstable, reactants not easily produced through any other chemical means, so that the high energy demands of chemiluminescence may be met. In most cases at least one of the reactants of any chemiluminescent system can be generated electrochemically, and hence ECL is actually more generally applicable than chemiluminescence.

#### Aromatic Radical Ion Annihilation

Reactions between oppositely charged aromatic or heteroaromatic radicals are among the most thoroughly studied electrochemiluminescent reactions. These reactions, which are carried out in aprotic solvents, can be generally written as in reactions [I–V],



or



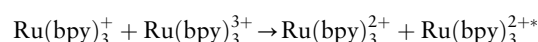
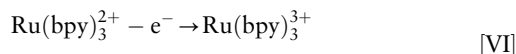
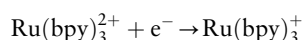
Here singlet and triplet states are denoted by the left superscripts 1 and 3, respectively. In the first case the reaction is energy sufficient, i.e., the change in free energy of the reaction is high enough to generate the excited singlet state directly in one step. In the latter case the reaction is energy deficient and the first step [III] generates the lowest triplet state. The emitter is then generated by triplet–triplet annihilation [IV]. The radical cations and anions are generated either at a single electrode by applying alternate positive and negative pulses or at two separate closely spaced electrodes. The species R and S may be the same or different. Typical examples are anthracene, 9,10-diphenylanthracene, *N,N,N',N'*-tetramethyl-*p*-phenylenediamine, rubrene, and pyrene.

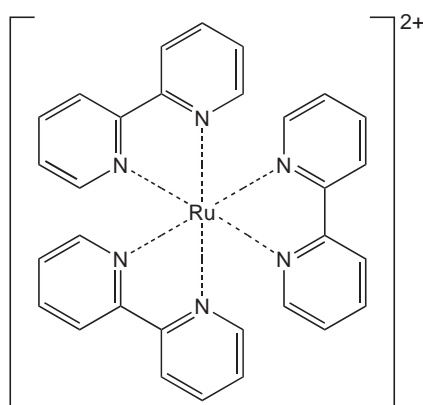
From the analytical point of view the luminescent annihilation of aromatic radicals suffers from the detrimental effect of oxygen and protic solvents, especially water. In fairly well-controlled conditions the system has been used as a basis for a selective liquid chromatographic detector, although the advantages compared with, for example, fluorometric detection are not apparent.

#### Platinum Metal Chelates

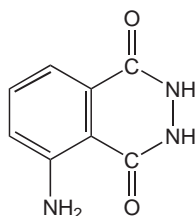
Some platinum metals, especially ruthenium but also rhenium, osmium, and iridium, form luminescent complexes with nitrogen-containing ligands, e.g., 2,2'-bipyridine (bpy). These complexes may participate in redox reactions with concomitant chemiluminescence in three different ways.

(a) Annihilation reactions in aprotic media, e.g., for tris(2,2'-bipyridyl)ruthenium(II),  $\text{Ru}(\text{bpy})_3^{2+}$  ((1), Figure 1); see reaction [VI]:

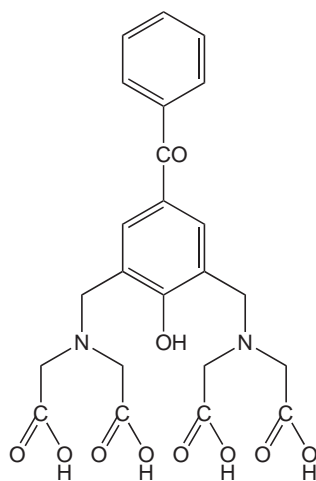




(1)



(2)



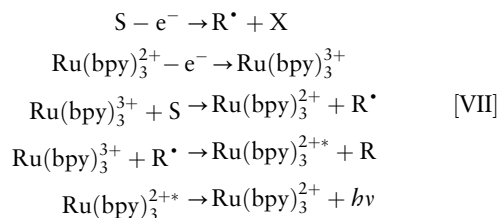
(3)

**Figure 1** Three typical electrochemiluminescent compounds: (1) tris(2,2'-bipyridyl)ruthenium(II),  $\text{Ru}(\text{bpy})_3^{2+}$ ; (2) luminol; and (3) 2,6-bis(*N,N*-bis(carboxymethyl)aminomethyl)-4-benzoylphenol.

These reactions are analogous to the annihilation reactions of aromatic compounds and suffer from the same drawback, namely the sensitivity to water.

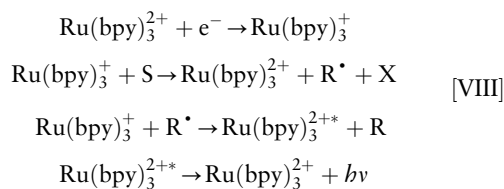
(b) Oxidation–reduction mechanisms (ox–red) where a substrate (co-reactant) *S* is initially oxidized directly at the electrode or through a reaction with  $\text{Ru}(\text{bpy})_3^{3+}$ , forming a strong reductant *R* that reacts with the ruthenium complex, generating light

(reaction [VII]):



A typical substrate *S* is oxalate, where *X* is then  $\text{CO}_2$  and  $\text{R}^\bullet$  is the anion radical  $\text{CO}_2^{\bullet-}$ . Even more efficient than oxalate are tertiary amines. It has been recently shown that the cation radical produced by tripropylamine by one-electron oxidation is a stronger oxidant than  $\text{Ru}(\text{bpy})_3^{3+}$  and has a rather long lifetime in aqueous solution at an appropriate pH value prior to deprotonation-induced redox inversion to a strongly reducing radical. Thus this cation radical can serve as an oxidizing mediator reaching  $\text{Ru}(\text{bpy})_3^{2+}$  species at quite long distances from the electrode surfaces. This is very useful when  $\text{Ru}(\text{bpy})_3^{3+}$  derivatives are used as electrochemiluminescent labels in binding assays carried out on the surfaces of magnetic latex beads that are finally collected by applying magnetic field on the working electrode during the detection step of the assay.

(c) Reduction–oxidation mechanisms (red–ox) where a substrate *S* is first reduced, forming a strong oxidant  $\text{R}^\bullet$  (reaction [VIII]). Normally, a substrate cannot undergo a one-electron reduction at an active metal electrode, and therefore, only the reduction by  $\text{Ru}(\text{bpy})_3^+$  is typically significant:



A typical substrate is the peroxodisulfate ion, with *X* being the sulfate ion and  $\text{R}^\bullet$  the sulfate radical  $\text{SO}_4^{\bullet-}$ .

In contrast with (a), the reactions in (b) and (c) also take place in protic media. The quantum yield of these ECL processes is rather high, and consequently the analytical methods based on these reactions have a very low detection limit. The most promising applications of these compounds are as labeling compounds in the ECL binding assays, i.e., immunoassays and DNA probe assays.

### ECL of Conventional Chemiluminescent Compounds

Many chemiluminescent reactions are based on hydrogen peroxide as a co-reactant. One of

these reactions is the oxidation of luminol (3-aminophthalhydrazide (2), **Figure 1**). In this reaction luminol reacts with hydrogen peroxide in the presence of catalysts, generating light. There are a large number of analytical methods for either hydrogen peroxide or catalysts based on this reaction. Hydrogen peroxide can also be generated electrolytically, and in fact, the oldest reported ECL reaction is based on this.

Some analytical methods, for example, for detection of trace metals, have been devised based on this reaction. Luminol has also been suggested as a labeling compound for the ECL immunoassay. In addition to luminol there are a number of analogous chemiluminescent compounds that require hydrogen peroxide in their luminescent reactions. Among these compounds the acridine derivatives lucigenin and acridinium esters have been used in ECL methods.

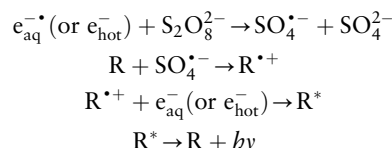
### ECL at Thin Insulating Film-Coated Electrodes

Emission of light from valve metal or semimetal electrodes (e.g., aluminum, magnesium, zirconium, tantalum, silicon) is observed when current is flowing in either direction through the interface between the electrode and the electrolyte solution during AC or pulsed excitation. The intensity and spectrum of the cathodic emission depends on the composition of the contacting electrolyte solution. Cathodic ECL is induced by certain metal cations, either complexed or uncomplexed, and organic fluorescent compounds. In most cases the presence of a co-reactant, a compound that is electrolytically reduced to a strongly oxidizing agent, is needed, e.g., peroxodisulfate ion or oxygen. The cathodic ECL at thin insulating film-coated electrodes can be obtained using cathodic potential pulses either with precoated electrodes (e.g., silicon electrodes coated with high-quality thermal oxide films) or using anodically oxidizable valve metal electrodes that are oxidized *in situ* during the ECL measurement through anodic potential pulses, yielding insulating surface films. Those anodic metal oxide films that behave as n-type semiconductors (e.g., tantalum oxide) cannot be utilized this way in most applications, and only those wide band gap oxide films having their Fermi level in the midgap region are generally usable. However, semiconducting oxide films can also offer an interesting basis for analysis when they are doped with luminescent metal ions and are used to probe hydrogen peroxide or clinical analytes, which are enzymatically coupled with their production.

The primary step of cathodic ECL at thin insulating-film coated electrodes is hot electron injection into the electrolyte solution using tunnel emission,

which is also called direct field assisted tunneling. The insulating film at the electrode must be ultrathin, i.e., its thickness must be of the order of 4 nm.

In contrast to traditional electrochemistry at active metal electrodes, one-electron reductions in aqueous solutions can be easily carried out at thin insulating film-coated cathodes and in the reduction potential range not obtainable at active electrodes. For instance, even toluene (R) can be excited in aqueous solution/emulsion through hot electron injection at thin insulating film-coated electrodes by the following ox-red mechanism:



A pulsed voltage waveform is normally used for the excitation. A large number of organic fluorescent compounds produce ECL at thin insulating film-coated electrodes in aqueous solutions, providing an alternative excitation method to photoluminescence. Some weak background emission is also observed without any luminescent additives during cathodic pulses due to high electric field-induced solid state electroluminescence from the insulating oxide films. By using excitation with short cathodic pulses and compounds displaying long-lived luminescence, the effect of this background can be largely eliminated.

### Instrumentation

The apparatus for making ECL measurements comprises the following main parts (**Figure 2**).

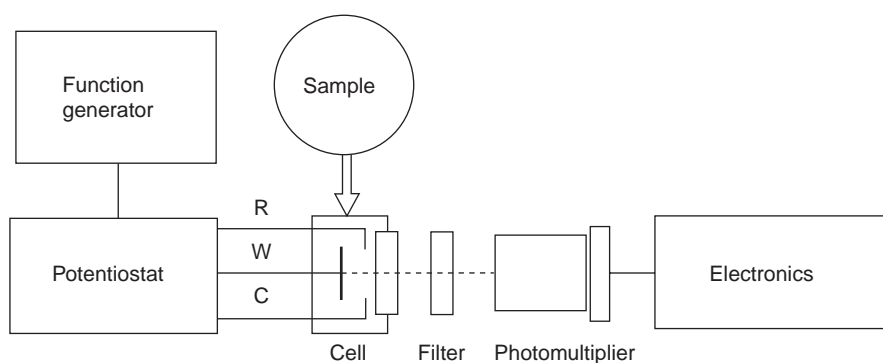
#### Function Generator

In most cases the excitation waveform in ECL is a periodic signal, often consisting of rectangular pulses with a variable duty cycle, or a sawtooth signal (**Figure 3**). For research purposes the function generator should preferably be programmable to allow easy generation of different voltage waveforms.

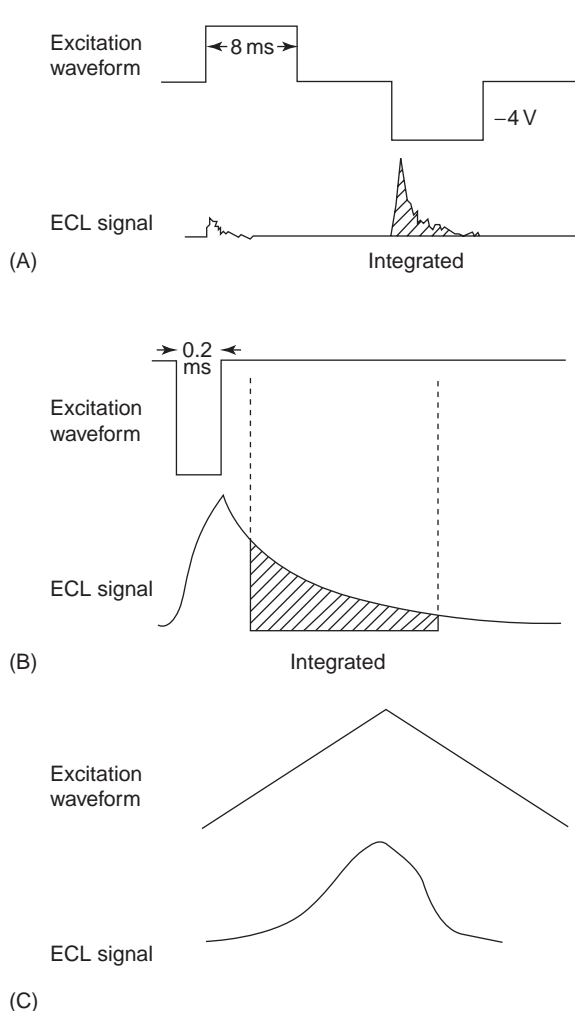
#### Potentiostat

The potential of the working electrode is controlled by means of a potentiostat. In some applications, a real reference electrode is not necessary and it is sufficient to use a pseudo-reference electrode or even to use the counter-electrode as a pseudo-reference electrode. In the latter case, only the voltage between the working electrode and counter-electrode is controlled, i.e., the real working electrode potential is reproducible only if the properties of the





**Figure 2** Schematic diagram of ECL detection system. W, working electrode; R, reference electrode; C, counter-electrode.



**Figure 3** Examples of voltage excitation waveforms and resulting luminescence signals in ECL measurements. (A) Symmetrical double-step waveform used in, e.g., ECL measurements with valve metal electrodes. (B) The sharp cathodic pulse and resulting signal in time-resolved ECL measurements with e.g., Tb(III) complexes. The shadowed region of the signals in (A) and (B) is integrated on a gated integrator. (C) Sawtooth waveform and resulting signal used, e.g., in ECL measurements with ruthenium complexes at active metal electrodes.

electrochemical cell remain unchanged between separate measurements.

### Electrochemical Cell and Electrodes

The design of the electrochemical cell depends on the method of application. In general, the cell contains two or three electrodes and an optical window. Platinum is the most commonly used material for the working electrode, although glassy carbon and gold have also been used. In some cases the electroluminescent material is immobilized in a Nafion perfluorinated ion exchange polymer film coated on the surface of a platinum or carbon electrode. In methods using oxide-coated electrodes, aluminum, silicon, zirconium, and tantalum are the materials mostly used. Often a rotating disk electrode is used as a working electrode, especially when the mechanism of ECL is being studied.

### Sample Introduction

The ECL measurement is carried out as a simple batch measurement or alternatively in a liquid stream. In the first case a spike of the analyte is added to the appropriate electrolyte solution, the solution is transferred to the cell, and the ECL signal is recorded. In the latter case, the sample is injected into the flowing liquid, which transfers it to the ECL cell either with (e.g., liquid chromatography (LC)) or without (flow injection analysis) a separation step.

### Detector

The intensity of ECL is in most cases so low that the most sensitive light detector, a photomultiplier tube (PMT), is necessary. Also avalanche photodiodes have been developed for photon counting purposes, and they can be expected to compete successfully with PMTs in the future, especially in miniaturized analytical systems. Depending on the application, the

detecting device may be preceded by an optical filter but in analytical applications hardly ever by a monochromator.

### Electronics

The electronic circuitry used depends on the method of measurement. At low light levels photon counting allows better discrimination against the noise than does analog recording. Both these methods have been used, although one can anticipate that with the continuous decrease in the cost of electronics photon counting will eventually win. The only instrumental advantage of analog recording of the photomultiplier anode current is better linearity at high light levels owing to the pulse pile-up in photon counting. In the time-resolved measurements the detection circuit is synchronized with the excitation pulses, allowing a fixed or variable time delay between the pulse and the recording window (Figure 3B).

### Analytical Applications

Electrochemiluminescence offers an alternative method of luminescence excitation. Its advantages as an analytical method compared with photoluminescence are partly the same as with chemiluminescence:

1. No expensive excitation optics is needed. Electronic excitation requires comparatively simple and inexpensive instrumentation.
2. Electrochemiluminescent reactions are quite rare. Although this may be a drawback if general methods are sought for, it is clearly an advantage if the electroluminescent compound is used for labeling, because of the smaller likelihood of interference.
3. ECL is often confined to the surface of the electrode or its close vicinity. Reactive intermediates generated by the electrode processes are often very short lived, and the action distance from the surface toward the bulk solution is small. This feature can be exploited, for example, in developing methods for homogeneous immunoassay.

One disadvantage of ECL methods is the frequent fouling of the electrodes. This effect can sometimes be prevented by regular electrochemical cleaning of the electrodes. In the case of an inexpensive electrode material such as aluminum, or even silicon manufactured in large quantities, the electrode is disposable.

### Inorganic Ions

Ruthenium can be determined down to subpicomolar level by complexing with suitable ligands, e.g., bpy, and applying ECL measurement. For example, various concentrations of  $\text{Ru}(\text{bpy})_3^{2+}$  are added to a solution of  $18 \text{ mmol l}^{-1} \text{ Na}_2\text{S}_2\text{O}_8$  and  $0.1 \text{ mol l}^{-1} \text{ Bu}_4\text{NBF}_4$  in acetonitrile–water (1:1, v/v) and the ECL emission is measured after deaeration. The logarithm of the light intensity is linear with the logarithm of concentration over six orders of magnitude from  $10^{-13}$  to  $10^{-7} \text{ mol l}^{-1}$ . Using tripropylamine instead of peroxodisulfate, linearity extending up to seven orders of magnitude can be achieved. The method may equally be used for the determination of peroxodisulfate and oxalate. For instance, oxalate can be determined by its chemiluminescent reaction with  $\text{Ru}(\text{bpy})_3^{3+}$  in the concentration range  $10^{-6}$ – $10^{-4} \text{ mol l}^{-1}$ . This happens to encompass the concentrations found in normal human blood and urine. Luminol-based ECL has been used for the determination of trace concentrations of cobalt(II) and copper(II), based on the fact that these metal ions effectively catalyze the luminescent reaction between luminol and electrogenerated hydrogen peroxide.

Certain metal ions in aqueous solution in the presence of hydrogen peroxide can be determined using ECL with a rotating aluminum disk electrode. With the uncomplexed metal ions the lowest detection limit ( $10^{-10} \text{ mol l}^{-1}$ ) is achieved with thallium(I), but mercury(II), copper(II), lead(II), and silver(I) induce relatively intensive ECL allowing detection in trace quantities. From the analytical point of view the most important metal ions determined at an oxide-covered aluminum electrode are europium(III) and terbium(III). Terbium, when complexed with 2,6-bis(*N,N*-bis(carboxymethyl)aminomethyl)-4-benzoylphenol ((3), Figure 1) or other ligands of analogous structure can be determined down to subpicomolar levels at stationary oxide-coated aluminum and silicon electrodes in the presence of potassium peroxodisulfate. The linear range of the log–log plot of emission intensity versus concentration of terbium is in these cases from  $10^{-12}$  to  $10^{-6} \text{ mol l}^{-1}$ . It should be noted that the luminescence lifetime of this terbium complex is  $\sim 2 \text{ ms}$ . The lifetime of the background solid state cathodic electroluminescence originating from the oxide film is much shorter, which allows time-resolved resolution of the analyte signal from the background. Determination of these metal cations is not important *per se* but is important in connection with the development of labeling compounds for binding assays (see below). Aromatic lanthanide(III) chelates ( $\text{Ln(III)}$ ) can be electrochemically excited in aqueous solutions only at thin

**Table 1** Examples of organic compounds determined using ECL

Compound	Method	Compound	Method
Luminol	a	Salicylic acid	c
Lucigenin	a	7-Hydroxy-4-methylcoumarin	c
Fluorescein	a, c	8-Quinololinol	c
Naphthalene	b	Eosin	c
Anthracene	b	5-Dimethylamino-1-naphthalenesulfonic acid	c
Phenanthrene	b, c	Rhodamine B	c
Rubrene	b	Morin	c
9,10-Diphenylanthracene	b, c	Tripropylamine	d
Dibenzofuran	b	Proline	d
Brucine	b	NADH	d
Perylene	b, c	Streptomycin	d
<i>N,N,N',N'</i> -tetramethyl- <i>p</i> -phenylene-diamine	b	Gentamicin	d

a, conventional chemiluminescence with electrogenerated  $\text{H}_2\text{O}_2$ ; b, radical ion annihilation; c, ECL on an oxide-covered aluminum electrode in the presence of  $\text{K}_2\text{S}_2\text{O}_8$ ; d, as a co-reactant in the ECL of a ruthenium–bipyridine complex.

insulating film-coated electrodes through hot electron chemistry. The excitation occurs through the so-called ligand-sensitized mechanism, in which the ligand is first excited by the above-mentioned ox–red or red–ox excitation pathway and then energy is transferred from the ligand intramolecularly to the central Ln(III) ion, which finally emits through its typical sharp-peaked f-shell radiative transitions, having relatively long luminescence lifetimes (up to  $\sim 2.2$  ms).

### Organic Compounds

The organic analytes that can be determined using ECL belong to two groups: (1) compounds or their reaction products that are luminescent; (2) compounds that participate in the chemiluminescent reaction without being luminescent *per se*. A representative list of organic compounds belonging to each group along with the method of measurement is shown in Table 1.

Polycyclic aromatic hydrocarbons (PAHs) belong to the group of compounds that in principle can be determined directly using ECL with the radical annihilation method. One practical drawback making the quantitative assay difficult is the sensitivity of the method to water and oxygen. However, ECL has been used as a detection system for reversed-phase LC with a mobile phase containing 10–20% water. ECL with oxide-coated electrodes has been used for the PAH assay in a micellar aqueous phase.

If an organic compound is fluorescent and reasonably soluble in water (or in a micellar aqueous phase), it most probably produces ECL at thin insulating film-coated electrodes in the presence of peroxodisulfate or oxygen. For instance, salicylic acid can be determined with a linear range of  $10^{-8}$  to  $10^{-4} \text{ mol l}^{-1}$  at a rotating oxide-covered aluminum

electrode, and e.g., aminonaphthalene sulfonates, fluorescein, eosine, and some coumarins have detection limits of  $\sim 10^{-10}$ – $10^{-9} \text{ mol l}^{-1}$  at oxide-covered stationary aluminum electrodes.

Ruthenium complexes need a strong reductant or oxidant for chemiluminescent reactions of type (b) and (c) above. These reactive compounds are usually radicals derived from simple organic compounds such as amines, amino acids, the reduced form of  $\beta$ -nicotinamide adenine dinucleotide (NADH), and some antibiotics. Analytical procedures have been developed for some of these compounds on this basis, e.g., for tripropylamine, which gives a detection limit of  $\sim 2 \times 10^{-8} \text{ mol l}^{-1}$ .

### Hydrogen Peroxide

Hydrogen peroxide is produced in various enzymatic reactions, and its determination is the basis for a number of assay methods. A direct ECL method is based on the use of tantalum or zirconium electrodes covered with a terbium(III)-doped oxide layer. Light with the typical emission spectrum of terbium(III) is emitted from the surface of the electrode in the presence of hydrogen peroxide.

In indirect ECL methods, hydrogen peroxide is generated electrolytically at a negatively biased glassy carbon or gold electrode and detected through the chemiluminescence of, e.g., luminol. Hydrogen peroxide is transported by a liquid flow toward the chemiluminescent reagent and during this time is partially decomposed by the sample molecules, which act as a catalyst. The sample could be, for instance, heme components, which effectively catalyze the decomposition of hydrogen peroxide.

### ECL Labels

Labeling techniques are used extensively in immunoassays and DNA probe assays. The label molecules

have a certain distinct feature, e.g., a radioactive atom or fluorescence, that allows detection and quantification at a very low concentration level. The use of labels based on detection using ECL has advantages in certain applications. In ECL the emission source is a very narrow zone in the close vicinity of the electrode. If in the binding assay one of the reactants is immobilized on the electrode, then complexation with the labeled molecule brings the ECL label close to the electrode, where it is effectively excited. This makes the nonseparation or 'homogeneous' binding assays possible. Another factor favoring ECL compared with fluorescence is its simpler and less expensive excitation technique. Immunoassays based on ECL of luminol and ruthenium complexes as well as ECL at oxide-covered electrodes have been demonstrated. For example, assays for the important thyroid-stimulating hormone (TSH) have been developed based both on ruthenium labels on a gold electrode and terbium labels on disposable oxide-coated aluminum and silicon electrodes. Both methods give a reasonably linear calibration line for TSH in the clinically important concentration range. One of the benefits of hot electron-induced cathodic ECL at thin insulating film-coated electrodes is that luminophores having very different optical and redox properties can be simultaneously excited. Thus, e.g., Tb(III) chelates, Ru(bpy)<sub>3</sub><sup>2+</sup>, metalloporphyrins, luminol, fluorescein, and many other fluorescent labeling compounds can be excited simultaneously in multiparametric assays. In addition, the signals from different types of label compounds emitting in the ultraviolet, visual, and near-infrared ranges can be separated from each other either by wavelength discrimination or by time discrimination or even by their combination. Some phosphorescent metalloporphyrin labels can be alternatively used instead of Ln(III) chelate labels when long-lived ECL displaying labels are required.

**See also:** **Elemental Speciation:** Overview. **Immunoassays, Techniques:** Luminescence Immunoassays. **Polycyclic Aromatic Hydrocarbons:** Determination; Environmental Applications.

## Further Reading

- Armstrong NA, Wightman RM, and Gross EM (2001) Light-emitting electrochemical processes. *Annual Review of Physical Chemistry* 52: 391–422.
- Blackburn GF, Shah HP, Kenten JH, *et al.* (1991) Electrochemiluminescence detection for development of immunoassays and DNA probe assays for clinical diagnostics. *Clinical Chemistry* 37: 1534–1539.
- Downey TM and Nieman TA (1992) Chemiluminescence detection using regenerable tris(2,2'-bipyridyl)ruthenium(II) immobilized in Nafion. *Analytical Chemistry* 64: 261–268.
- Fahnrich KA, Pravda M, and Guilbault GG (2001) Recent applications of electrogenerated chemiluminescence in chemical analysis. *Talanta* 54: 531–559.
- Faulkner LR and Bard AJ (1977) Techniques of electrogenerated chemiluminescence. In: Bard AJ (ed.) *Electroanalytical Chemistry*, vol. 10, pp. 1–95. New York: Dekker.
- Greenway GM (1990) Analytical applications of electrogenerated chemiluminescence. *Trends in Analytical Chemistry* 9: 200–203.
- Haapakka K, Kankare J, and Puhakka O (1988) Fluorophore-enhanced cathodic electroluminescence at an oxide-covered aluminum electrode. *Analytica Chimica Acta* 207: 195–210.
- Haapakka K, Kankare J, and Kulmala S (1985) Feasibility of low-voltage cathodic electroluminescence at oxide-covered aluminum electrodes for trace metal determinations in aqueous solutions. *Analytica Chimica Acta* 171: 259–267.
- Helin M, Vare L, Håkansson M, *et al.* (2002) Electrochemiluminoimmunoassay of hTSH at disposable oxide-coated n-silicon electrodes. *Journal of Electroanalytical Chemistry* 524–525: 176–183.
- Kankare J, Fälden K, Kulmala S, and Haapakka K (1992) Cathodically induced time-resolved lanthanide(III) electroluminescence at stationary aluminum disc electrodes. *Analytica Chimica Acta* 25: 17–28.
- Kankare J, Haapakka K, Kulmala S, *et al.* (1992) Immunoassay by time-resolved electrogenerated luminescence. *Analytica Chimica Acta* 266: 205–212.
- Knight AW (1999) A review of recent trends in analytical applications of electrogenerated chemiluminescence. *Trends in Analytical Chemistry* 18: 47–62.
- Kulmala S, Ala-Kleme T, Heikkilä L, and Väre L (1997) Energetic electrochemiluminescence of (9-fluorenyl) methanol induced by injection of hot electrons into aqueous electrolyte solution. *Journal of the Chemical Society, Faraday Transactions* 93: 3107–3113.
- Kulmala S, Håkansson M, Spehar AM, *et al.* (2002) Heterogeneous and homogeneous electrochemiluminoimmunoassays of hTSH at disposable oxide-covered aluminum electrodes. *Analytica Chimica Acta* 458: 271–280.
- Miao W, Choi J-P, and Bard AJ (2002) Electrogenerated chemiluminescence 69: The tris(2,2'-bipyridine)ruthenium(II), (Ru(bpy)<sub>3</sub><sup>2+</sup>)/Tri-n-propylamine (TPrA) system revisited – a new route involving TPrA<sup>•+</sup> cation radicals. *Journal of the American Chemical Society* 124: 14478–14485.
- Richter MM (2004) Electrochemiluminescence (EL). *Chemical Reviews* 104: 3003–3036.

# CHEMOMETRICS AND STATISTICS

## Contents

### Statistical Techniques

#### Experimental Design

#### Optimization Strategies

#### Multivariate Classification Techniques

#### Multivariate Calibration Techniques

#### Expert Systems

#### Multicriteria Decision Making

#### Signal Processing

#### Spectral Deconvolution and Filtering

## Statistical Techniques

**J N Miller**, Loughborough University, Loughborough, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

In the analytical sciences it is usual to make a fairly small number of measurements on a given material. This small dataset in statistical terms constitutes a sample. Because of random errors the replicate values within the sample will not be identical, but the analyst must provide an estimate of these errors – without such an estimate the quantitative results presented would be meaningless. Statistics provides the means to do this by estimating the results that would have been obtained if a hypothetically infinite population of measurements had been made. This distinction between samples and populations is fundamental: results relating to samples are given English symbols (e.g.,  $\bar{x}$  for a sample mean), whereas those describing populations are given Greek symbols (e.g.,  $\mu$  for a population mean). The link between samples and populations is generally made by assuming that the data come from a population with a specific frequency distribution.

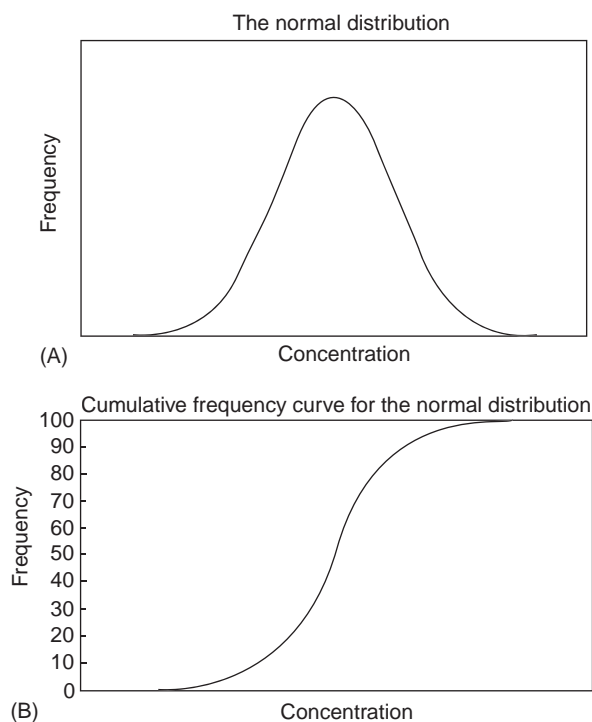
## The Normal Distribution

When replicate analyses of the same material are made in the same experimental conditions it is usual to assume that the frequency distribution of the results is a normal or Gaussian one. This distribution

has the mathematical form:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad [1]$$

where  $y$  is the frequency with which a given value of the measurement  $x$  occurs. The normal distribution has the form shown in **Figure 1A**. The exact shape of the curve is defined by the parameters  $\mu$  and  $\sigma$ , which are the mean and standard deviation of the population, respectively, and describe the position of the curve (the  $x$ -value at which the frequency is maximal – this is the population mean,  $\mu$ , because the curve is



**Figure 1** The normal (Gaussian) distribution represented as (A) a frequency curve: and (B) a cumulative frequency curve.

Simulated normal distribution with mean 0.500, s.d. 0.03  
Nitrate level in water, ppm

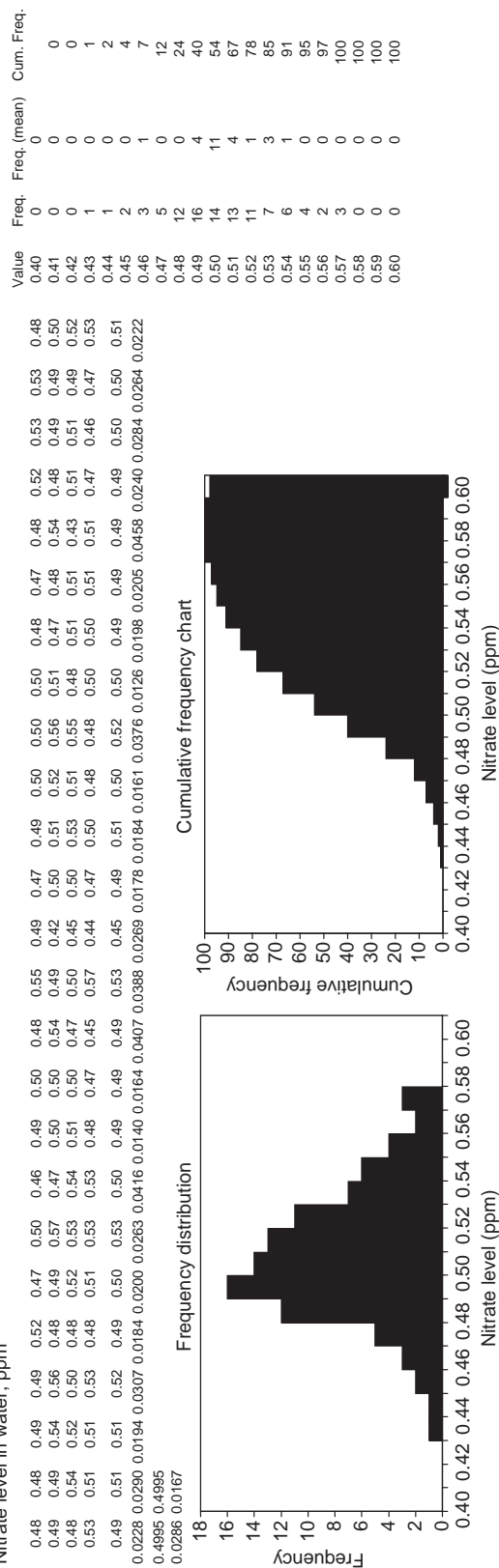


Figure 2 Microsoft Excel® simulation of 100 measurements of the nitrate ion concentration in water.

symmetrical) and its width (see below). An alternative representation of the normal distribution is the cumulative frequency curve, shown in **Figure 1B**. In this case, the y-axis gives the frequency, on a scale of 0–1% or 0–100%, of any value of  $x$  and all the lower values of  $x$ . The cumulative frequency reaches 0.5% or 50% at the mean value of  $x$ , i.e., the maximum of the frequency plot in **Figure 1A**.

**Figure 2** shows the outcome of a Microsoft Excel® simulation of 100 values of the nitrate ion concentration in a single water sample (ppm). It is most unlikely that any material would actually be analyzed 100 times, but the data are used to demonstrate methods by which such results are presented. Each result is given to two places of decimals, all the results lying in the range 0.43–0.57 ppm. The ‘Frequency’ column in the spreadsheet shows the number of times that each value 0.43, 0.44, 0.45, ..., 0.57 occurs, and this column along with its neighbor headed ‘Value’ comprises a frequency table. The data in this table can be plotted as a bar chart, as shown in **Figure 2**. The shape of this bar chart is approximately the same as the smooth ideal curve for the normal distribution shown in **Figure 1A**. An additional column shows the cumulative frequency values, and when these are plotted as a bar chart the curve obtained looks similar to that in **Figure 1B**. The bar chart, frequency table, and cumulative frequencies are just three of several ways in which replicate experimental data can be presented.

Using eqn [1] it can be shown that ~68% of the measurements in a normal population lie within  $1\sigma$  of the mean value,  $\mu$ , i.e., between  $\mu - \sigma$  and  $\mu + \sigma$ . Ninety-five percent of the measurements fall in the range  $\mu \pm 1.96\sigma$  often approximated as  $\mu \pm 2\sigma$ , and 99.7% of the measurements in the range  $\mu \pm 3\sigma$ . This is equivalent to saying that 68% of the area under the curve lies between  $\mu - \sigma$  and  $\mu + \sigma$  and so on). These multipliers of  $\sigma$ , i.e., 1 for 68% of the results and 1.96 (for 95%), are examples of the statistic  $z$ , which can be used to express any value of  $x$  in terms of its deviation from the mean in standard deviation units, i.e.,

$$z = \frac{x - \mu}{\sigma} \quad [2]$$

These properties of the normal distribution are of great importance, as they are used to calculate confidence intervals and limits (see below).

In practice, when a sample of measurements is studied, it is necessary to estimate the value of  $\mu$  using the sample mean,  $\bar{x}$ :

$$\bar{x} = \sum_{i=1}^n x_i / n \quad [3]$$

the  $x_i$  values being the  $n$  individual measurements. An alternative measurement of central tendency or

location is the median, obtained by arranging the results in numerical order and identifying the middle one if  $n$  is odd, and the average of the two middle values if  $n$  is even. The median has the advantage that it is unaffected by extreme values and avoids consideration of possible outliers (see below). The standard deviation of the population,  $\sigma$ , is given by

$$\sigma = \left\{ \frac{\sum_{i=1}^n (x_i - \mu)^2}{n} \right\}^{0.5} \quad [4]$$

In practice, since  $\mu$  is estimated using  $\bar{x}$  it is necessary to estimate  $\sigma$  using the sample standard deviation,  $s$ :

$$s = \left\{ \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \right\}^{0.5} \quad [5]$$

Equation [5] includes the term  $(n - 1)$  in its denominator rather than  $n$  to ensure that  $s$  is an unbiased estimator of  $\sigma$ . The number  $(n - 1)$  is known as the number of degrees of freedom of the  $s$  value, i.e., the number of independent deviations  $(x_i - \bar{x})$  used to calculate  $s$ . If  $(n - 1)$  such deviations are known, the last deviation can be deduced from the obvious result  $\sum_i (x_i - \bar{x}) = 0$ . The concept of degrees of freedom is much used in significance testing (see below).

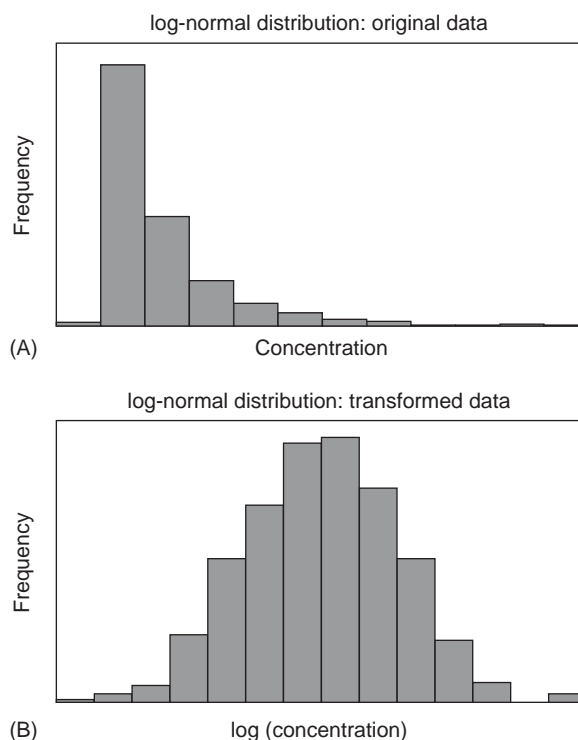
The square of the standard deviation is called the variance (and has many important properties and uses), while the standard deviation expressed as a percentage of the mean, i.e.,  $100s/\bar{x}$ , is called the relative standard deviation (r.s.d.) or the coefficient of variation (c.v.).

## Other Frequency Distributions

The normal distribution is very important in experimental science but other frequency distributions do occur in practice. One important situation quite common in analytical data arises when an otherwise normal distribution has heavy tails, i.e., has more values than expected that are much higher or lower than the mean. This situation may arise when a normal random distribution of errors is supplemented by additional gross errors, i.e., possible outliers (see below). Another cause of heavy-tailed distributions is the superimposition of two or more normal distributions with the same mean but different standard deviations, for example, when the same analysis is done by two or more different people or instruments and the results are combined. Distributions of this kind are now commonly studied using robust statistical methods, which seek to down-weight measurements that are much higher or lower than the mean. (This is regarded as a more logical approach than the use of statistics such as the median (see above), which ignores extreme results completely.) A simple robust

statistic is the 10% trimmed mean, i.e., the mean value calculated after the highest and lowest 10% of the measured values are neglected. This statistic has the advantage that possible outliers are ignored. If there are no such outliers, then the mean obtained is closely similar to the overall mean, so no harm is done. Considering the 100 measurements in Figure 2, their overall mean is 0.4995, and the 10% trimmed mean (i.e., the mean of the 80 values obtained when the highest 10 and the lowest 10 are omitted) has the very similar value 0.5038. Robust standard deviations can also be calculated, and virtually all conventional univariate and multivariate statistical methods can be similarly modified ('robustified') to handle situations where outliers may occur. Robust statistics are considered in detail elsewhere in this encyclopedia.

If a series of measurements, using the same technique in every case, is carried out on a number of different sample materials, another frequency distribution that often occurs is the log-normal distribution. This situation often arises in the natural world (e.g., the concentrations of antibody in the blood sera of different individuals) and in environmental science (e.g., the levels of nitrate in different water samples taken over successive days or weeks). The log-normal distribution curve has a long tail at its high end (Figure 3A), but can be converted to look like a



**Figure 3** The log-normal distribution with the horizontal axis showing (A) measured values and (B) logarithms of the measured values.



normal distribution by plotting the logarithms of the measured values on the horizontal axis (Figure 3B). After transformation in this way, the logarithmic data have the properties expected of a normal distribution. If the mean of the logarithmic values is calculated and its antilogarithm found, it equals the geometric mean of the original values, given by

$$\bar{x}_g = \sqrt[n]{x_1 x_2, \dots, x_n} \quad [6]$$

where  $x_1$ ,  $x_2$ , etc., are the individual measurements. Confidence limits (see below) can be calculated for log-normal data in the same way as for normally distributed results, but again they are the confidence limits of the geometric mean.

Sometimes it is required to test whether a set of measurements could have come from a population with a particular distribution (most commonly the normal distribution). A method for carrying out such tests is described in a later section.

### The Sampling Distribution of the Mean and the Central Limit Theorem

When samples of size  $n$  are repeatedly taken from a population of data, and the sample means –  $\bar{x}$  values – are plotted as a frequency diagram, the result is called the sampling distribution of the mean. A very important statistical principle called the central limit theorem shows that the shape of the curve obtained in such cases will be close to that for a normal distribution, even if the parent population is not normally distributed. This approximation improves as the sample size  $n$  increases. The standard deviation of the  $\bar{x}$  values, generally (and perhaps misleadingly) called the standard error of the mean (s.e.m.), is given by  $\sigma/\sqrt{n}$ . (Inevitably it is necessary in practice to use the estimated value,  $s/\sqrt{n}$ ). This result is (roughly) exemplified by the two numbers in the bottom left-hand corner of Figure 2. The standard deviation of the original 100 measurements is 0.0286, quite close to the simulated value of 0.03. But the standard deviation of the means (s.e.m.) of 25 samples with  $n=4$ , i.e., the 25 column means, is only 0.0167.

Figure 4 summarizes a simulation of the central limit theorem using Microsoft Excel® which gives 250 values of the weight of a vial from a uniform distribution, i.e., all the weights in the range covered (99.6–100.5 mg) should be equally likely to occur. This is approximately true, all the 10 bars in the darker histogram representing the original 250 values having sizes between 20 and 29 results. The paler bars show the effect of taking 50 samples of size five from the same 250 measurements and plotting the mean values of these samples. These bars have a

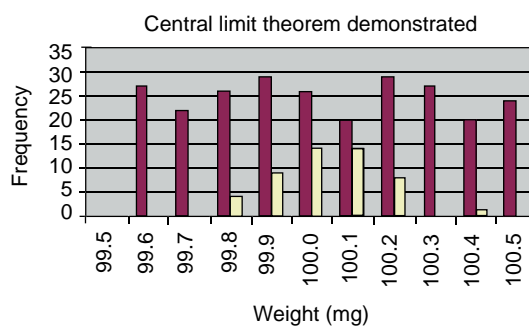


Figure 4 Microsoft Excel® simulation of the central limit theorem. For details see text.

shape similar to that of the normal distribution, and cover a narrower range than the original data. The importance of the Central Limit Theorem lies in the fact that it allows us to use the properties of the normal distribution to calculate confidence intervals and limits from the relatively small numbers of measurements (small sample sizes) often obtained in analytical work.

### Confidence Limits and Intervals

When a small number of replicate analyses are performed their mean value,  $\bar{x}$  is used to estimate the population mean,  $\mu$ . In the absence of systematic errors (i.e., with unbiased measurements),  $\bar{x}$  would be the true value of the analyte concentration. But because random errors occur  $\bar{x}$  will not be exactly equal to  $\mu$ . We thus wish to define an interval within which  $\mu$  lies with a given degree of probability, and thanks to the central limit theorem we can apply normal distribution properties to the sampling distribution of the mean to do this. Remembering that the standard error of  $\bar{x}$  is  $\sigma/\sqrt{n}$  (see above), the normal distribution shows that 95% of the values of  $\bar{x}$  will lie within  $1.96\sigma/\sqrt{n}$  of the mean. That is, the 95% confidence limits of the mean are given by

$$\mu - 1.96\sigma/\sqrt{n} < \bar{x} < \mu + 1.96\sigma/\sqrt{n} \quad [7]$$

which can be rewritten as

$$\bar{x} - 1.96\sigma/\sqrt{n} < \mu < \bar{x} + 1.96\sigma/\sqrt{n} \quad [8]$$

Equation [8] is more appropriate than [7], as we usually have a single experimental mean  $\bar{x}$  and wish to use it to provide a range for  $\mu$ , the true value. The range  $\mu = \bar{x} \pm 1.96\sigma/\sqrt{n}$  is known as the confidence interval for  $\mu$ . Similarly, 99.7% confidence interval and limits are obtained from  $\mu = \bar{x} \pm 3\sigma/\sqrt{n}$ .

When dealing with small samples, however, it is necessary to adjust the methods just described in two

ways. First, as already noted, we must in practice use  $s$  to estimate  $\sigma$ . A second problem is that as  $n$  diminishes,  $s$  becomes increasingly less reliable as an estimator of  $\sigma$ . This is exemplified in Figure 2. The 25 individual columns of data, i.e., 25 samples with  $n=4$ , have  $s$  values as low as 0.0126 and as high as 0.0458, yet each of these might be taken as an estimate of  $\sigma$ . This problem was famously studied almost a century ago by W.S. Gosset, writing under the pseudonym 'Student'. He showed that confidence intervals derived for small samples are given by

$$\mu = x \pm t_{n-1}s/\sqrt{n} \quad [9]$$

In [9],  $s$  has replaced  $\sigma$  as expected, and the numbers 1.96 or 3,  $z$ -values derived from the properties of the normal distribution, are replaced by the statistic  $t$ . This varies with the sample size  $n$ , becoming larger as  $n$  decreases to take account of the greater unreliability of  $s$  as an estimator of  $\sigma$ . The value of  $t$ , readily obtained from statistical tables using the data for  $(n-1)$  degrees of freedom, also depends on the confidence level required (95%, 99.7%, etc.).

## Uncertainty and the Propagation of Errors

The discussion in the previous section assumed that systematic errors were absent from an analytical experiment. In reality, this assumption can never be justified, and experience (e.g., in the conduct of proficiency testing schemes and method comparison studies) shows that quite large systematic errors can occur in many analyses. In recent years, considerable pressure has developed for all analytical results to be accompanied by an estimate of their uncertainty. The uncertainty of a result provides a range within which the true value of the analyte level is predicted to lie, taking into account all sources of error, and with a given degree of probability, typically 95%. There are two possible approaches to estimating uncertainty. The top-down approach utilizes the results of proficiency testing schemes, in which carefully prepared sample materials are circulated to a number of participating laboratories, each of which estimates the analyte content using its own conventional method. The range of results obtained by different laboratories (evaluated and published independently by a controlling body) then provides, it is argued, an estimate of the overall uncertainty of an analysis for that particular analyte. This approach may underestimate sampling errors (as the samples are specially prepared before dispatch to the participants): equally

it may overestimate the spread of results obtained by a single laboratory. Its main advantage is that it avoids the tedious study of errors occurring in the numerous individual steps in a particular analysis.

If appropriate proficiency scheme results are not available this detailed examination of each step in the analytical process, the so-called bottom-up approach to uncertainty estimates, must be used. In doing so, it is necessary to use a number of equations that allow the calculation of the overall random error arising in a process containing two or more steps. This is technically referred to as the propagation of errors. The basic problem in such cases is that, if the two or more experimental steps have random error sources that are independent of each other, such errors will partly, but not wholly, cancel each other out. In the following equations the final result of the analysis is called  $X$ , and the individual analytical steps contributing to  $X$  are  $A$ ,  $B$ , etc. The random errors in  $X$ ,  $A$ ,  $B$ , etc. (normally standard deviations) are given the symbols  $\delta X$  (which we wish to determine),  $\delta A$ ,  $\delta B$ , etc. Then,

$$\begin{aligned} \text{If } X &= A \pm B \pm C \dots, \\ \text{then } \delta X^2 &= \delta A^2 + \delta B^2 + \delta C^2 + \dots \end{aligned} \quad [10]$$

$$\begin{aligned} \text{If } X &= A/B \text{ or } A \times B \dots, \\ \text{then } (\delta X/X)^2 &= (\delta A/A)^2 + (\delta B/B)^2 + \dots \end{aligned} \quad [11]$$

$$\text{If } X = A^q, \text{ then } (\delta X/X) = q(\delta A/A) \quad [12]$$

Using equations of types [10]–[12] singly or jointly we can combine two or more random error contributions to provide an overall random error estimate. (Note that the independence of the sources of error is assumed throughout. For example, if  $X = A^2$ , then  $(\delta X/X) = 2(\delta A/A)$  from [12]: using eqn [11], giving  $(\delta X/X) = \sqrt{2}(\delta A/A)$ , would be incorrect.) To determine uncertainties we must also include estimates of systematic errors in mathematical forms (i.e., hypothetical population distributions) that allow them to be combined algebraically with the random ones. Guidelines for doing this are provided in the Further Reading section. Although the bottom-up method is potentially tedious, as even the simplest analytical procedures are found on close inspection to involve numerous error-prone steps, it can be very valuable. It reveals the individual steps that contribute most to the overall error, and which thus need to be improved if the range of values given by the uncertainty estimate is to be reduced. Equally, it will reveal steps that contribute little to the overall error, and which are thus not worth trying to improve.

## Significance Testing Principles

The commonest applications of univariate statistics in the analytical sciences are probably in significance testing. We may wish, for example, to decide whether one of the results in a sample of measurements is a possible outlier; to decide whether a data sample could have come from a normal (or some other) population; or to compare two or more means or standard deviations. Although the methods used in these cases differ in detail, the same underlying principles are used in every case. The starting point for such a test is always a null hypothesis ( $H_0$ ), i.e., a hypothesis assuming no difference between the data being compared. So in the examples cited above the null hypotheses would be that the suspect value was not an outlier; that the data *do* fit a normal or other specified distribution; or that the means or standard deviations being compared *do not* differ significantly from one another. The next step is to calculate the probability of obtaining the actual experimental results, assuming the null hypothesis to be correct. If this probability is less than a user-determined value, often  $p = 0.05$  or 5%, sometimes  $p = 0.01$  (1%), etc., then it is concluded that the null hypothesis should after all be rejected. If the probability of getting the experimental data exceeds, e.g., 5%, then the null hypothesis is retained. (The results of such tests should always be expressed in exactly this ' $H_0$  rejected' or ' $H_0$  retained' form, along with a record of the probability level used and the sample size.)

The probability of obtaining the experimental data was until recently determined by converting them into a suitable test statistic (see below) and comparing the result with critical values for that statistic ( $t$ ,  $F$ , etc.). These values are available in standard sets of tables for different numbers of degrees of freedom and different probability levels. Nowadays, it is commoner to use software that provides the probabilities directly, a simpler procedure that moreover gives more detailed information.

There is inevitably a chance that any significance test may lead to an erroneous conclusion. If it is decided to reject a null hypothesis at the  $p = 0.05$  level, there must be a 5% chance that it will be rejected when it should be retained. Such an error is called a type I error. Similarly, it is possible to retain a null hypothesis that should be rejected – a type II error. It is natural to consider that a good test procedure should reject an erroneous null hypothesis as often as possible, so  $(1 - \text{the probability of a type II error})$  is called the power of a test. (For a given test in specified conditions the power is a calculable number, not merely a vague concept.) It is clearly possible to reduce the chance of a type I error occurring by

performing a test at, e.g., the  $p = 0.01$  level rather than at the  $p = 0.05$  level. Unfortunately, for any given sample size this inevitably increases the chance of a type II error occurring. The only way to reduce the chance of both types of error in a given test is to increase  $n$ , i.e., to take more measurements.

Significance tests can be performed as one-tailed (sometimes called one-sided) or two-tailed (two-sided) tests. If two (or more) sets of results are compared with no *a priori* reason to believe that, e.g., one of the mean values should be higher or lower than the other, then the comparison is a two-tailed one: we test to see whether the two results differ or not. This is the commoner situation. In some cases, however, we are only interested in one possibility. For example, in testing the efficacy of a catalyst we only wish to study whether one reaction rate is specifically higher than another: this would require a one-tailed test. The decision on which of the two types of test is required should be taken in advance, but the procedure is very similar whether a one-tailed or two-tailed test is used. The only difference arises in the critical values taken from the statistical tables for the given test, or in the choice made from a software menu.

## Common Significance Tests

Hundreds of significance tests are available for different purposes in the handling of experimental data. Here we outline only those most commonly used in analytical work. One extremely common requirement is to test whether a set of data might come from a normally distributed (or other defined) population. For relatively small samples, the most efficient method is the Kolmogorov test (later adapted by Smirnov, and sometimes known as the Kolmogorov–Smirnov test). The procedure involves comparing the cumulative frequency plot of the data to be tested (this will look like a series of steps, each step corresponding to one measurement) with that of the hypothesized distribution (a smooth curve, often available in graphical form in sets of statistical tables). The maximum vertical distance between the two plots is the test statistic, which can be compared with critical values as usual. In testing for the normal distribution it is necessary to start by transforming the data to the standard normal variable,  $z$ , using eqn [2].

A very common problem in analytical work is the occurrence of outliers, or suspect results, i.e., one (or more) results in a series of replicate measurements that appears to be out of line with the rest. Should the suspect value be rejected before calculation of the mean, standard deviation, etc.? The treatment of

outliers is controversial, as samples drawn from, e.g., a normal distribution may legitimately include some measurements very different from the mean (Figure 1). Robust methods (see above) provide one approach to the problem, but the use of significance tests for outliers is still common. ISO now recommends the Grubbs test for this purpose. The test statistic here,  $G$ , is given by

$$G = |\text{suspect value} - x|/s \quad [13]$$

with  $s$  calculated with the suspect value included. Values of  $G$  above the critical value suggest that the suspect value could be rejected. Like most outlier tests, this method has the twin disadvantages that it is hard to apply when two or more outliers occur, and that it assumes that the population distribution is normal. This could be dangerous: a measurement that appears to be an outlier with the assumption of a normal distribution may not be an outlier if the distribution is, e.g., log-normal.

The important  $F$ -test allows us to compare two standard deviations,  $s_1$  and  $s_2$ , of samples with sizes  $n_1$  and  $n_2$ . We calculate

$$F = s_1^2/s_2^2 \quad \text{or} \quad s_2^2/s_1^2 \quad [14]$$

so that  $F > 1$ . Since the squares of the standard deviations are the variances of the samples, this is sometimes known as the ratio of variances test. Critical values of  $F$  depend on the number of degrees of freedom of each sample,  $(n_1 - 1)$  and  $(n_2 - 1)$ , and on the probability level chosen ( $p = 0.05$ , etc.). Experimental  $F$ -values greater than the critical value allow the rejection of the null hypothesis that the standard deviations are not significantly different. When  $n_1$  and  $n_2$  are small, the critical values of  $F$  are large, i.e., the two standard deviations must be very different before such a divergence becomes statistically significant. If the two standard deviations are found not to be significantly different, they can for some purposes be *pooled* to give a single value of  $s$ , using the equation

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad [15]$$

This equation indirectly demonstrates the important additivity properties of variances. The numerator contains the two variances, each multiplied by the respective number of degrees of freedom, and the denominator gives the total number of degrees of freedom for the two samples. So the pooled variance,  $s^2$ , is the average of the two individual variances, taking degrees of freedom into account.

A series of important significance tests is provided by the  $t$ -statistic (see above). Its simplest use is in the comparison of an experimental mean,  $x$  with a true or standard value,  $\mu$ , for example, in testing for systematic errors. In this case the test statistic is given by

$$t = |x - \mu|/\sqrt{n}/s \quad [16]$$

The numerical value of  $t$  is then compared with the critical values at the chosen probability level and  $(n - 1)$  degrees of freedom. Experimental  $t$  values that exceed the critical value lead to a rejection of the null hypothesis that  $\mu$  and  $x$  are not significantly different.

A similar method can be used to compare two mean values,  $x_1$  and  $x_2$  from samples with sizes  $n_1$  and  $n_2$ , respectively. The null hypothesis that these two means are not significantly different is equivalent to the statement that  $(x_1 - x_2)$  is not significantly different from zero. So eqn [16] can be adapted to give

$$t = \frac{|x_1 - x_2|}{s\sqrt{(1/n_1) + (1/n_2)}} \quad [17]$$

where  $s$  is the pooled standard deviation calculated from eqn [15]. The value of  $t$  is again compared with critical values at the desired probability level, and with  $(n_1 + n_2 - 2)$  degrees of freedom. The null hypothesis of equal means can be rejected if the experimental  $t$ -value exceeds the critical one. In the (unlikely) event that the two standard deviations cannot be pooled modified equations are necessary to calculate both  $t$  and the appropriate number of degrees of freedom. These equations are given below.

A final and useful application of the  $t$ -statistic is in the paired  $t$ -test. This is applicable when, e.g., several different materials containing different levels of analyte are studied using two different methods. Any differences between the results of the two methods might be masked by differences between the analyte levels in the samples if a conventional  $t$ -test were used. So it is necessary to determine for each material the difference between the results of the two methods, which would average zero if the null hypothesis, i.e., that the methods give indistinguishable results, is correct. The value of  $t$  is then obtained from

$$t = d\sqrt{n}/s_d \quad [18]$$

where  $d$  and  $s_d$  are the mean and standard deviation, respectively, of the differences between the pairs of results. Again the null hypothesis is rejected if the experimental value of  $t$  exceeds the critical value at the chosen probability level. The number of degrees

of freedom in this case is  $(n - 1)$  where  $n$  is the number of pairs of data.

Comparisons between means (and standard deviations) can be extended to the study of three or more sets of data. Such comparisons require the very important statistical method called analysis of variance (ANOVA).

## Application of Statistical Techniques

Many of the significance tests and other procedures summarized in this article (and many others) are very readily performed with the aid of Microsoft Excel<sup>®</sup>, Minitab<sup>®</sup>, and other widely available programs. Such software also gives instant access to the most important descriptive statistics (mean, median, standard deviation, s.e.m., confidence limits, etc.). In practice, the major problems are therefore (1) accurate entry of the experimental data into the program (this problem may not arise if an analytical instrument is directly interfaced to a PC); and (2) choice of the appropriate test once data entry has been successfully completed. Guidance on the latter

crucial issue is provided by the resources listed in the Further Reading section.

**See also: Chemometrics and Statistics:** Experimental Design; Optimization Strategies; Multivariate Classification Techniques; Multivariate Calibration Techniques; Expert Systems; Multicriteria Decision Making; Signal Processing; Spectral Deconvolution and Filtering.

## Further Reading

- De Levie R (2001) *How to Use Excel<sup>®</sup> in Analytical Chemistry*. Cambridge: Cambridge University Press.
- ISO (1993) *Guide to the Expression of Uncertainty in Measurement*. Geneva: International Organisation for Standardisation.
- Meier PC and Zund RE (2000) *Statistical Methods in Analytical Chemistry*, 2nd edn. New York: Wiley.
- Miller JN and Miller JC (2000) *Statistics and Chemometrics for Analytical Chemistry*, 4th edn. Harlow: Pearson Education.
- Neave HR (1981) *Elementary Statistics Tables*. Routledge: London.

## Experimental Design

**G Hanrahan, J Zhu, S Gibani, and D G Patil,**  
California State University, Los Angeles, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by J Goupy, pp. 659–666, © 1995, Elsevier Ltd.

## Introduction

Experimental design methods allow the experimenter to understand better and evaluate the factors that influence a particular system by means of statistical approaches. Such approaches combine theoretical knowledge of experimental designs and a working knowledge of the particular factors to be studied. Although the choice of an experimental design ultimately depends on the objectives of the experiment and the number of factors to be investigated, initial experimental planning (as shown in Figure 1) is essential.

The relationship between the various factors and response within a given system can be shown

mathematically as follows:

$$y = f(x_1, x_2, x_3, \dots, x_k) \quad [1]$$

where  $y$  is the response of interest in the system and  $x$  are the factors that affect the response when their values change. In general, the following types of factors can be distinguished: (1) continuous, e.g., temperature; and (2) discrete, e.g., experimenters. Factors are considered to be independent if there is no relationship between them and dependent if a relationship exists. The values or settings attributed for each factor are called levels. Each experimental run in an experimental design study requires that one or more treatments (stimulus applied to one or more factors) be applied to the system and the response measured. The experimenter then employs statistical design methods to determine if the treatment of interest or combination of treatments was significant in influencing the response of the system under study. Calculation of the treatment effects can then be used to identify which variables lead to an optimal response.



of freedom in this case is  $(n - 1)$  where  $n$  is the number of pairs of data.

Comparisons between means (and standard deviations) can be extended to the study of three or more sets of data. Such comparisons require the very important statistical method called analysis of variance (ANOVA).

## Application of Statistical Techniques

Many of the significance tests and other procedures summarized in this article (and many others) are very readily performed with the aid of Microsoft Excel<sup>®</sup>, Minitab<sup>®</sup>, and other widely available programs. Such software also gives instant access to the most important descriptive statistics (mean, median, standard deviation, s.e.m., confidence limits, etc.). In practice, the major problems are therefore (1) accurate entry of the experimental data into the program (this problem may not arise if an analytical instrument is directly interfaced to a PC); and (2) choice of the appropriate test once data entry has been successfully completed. Guidance on the latter

crucial issue is provided by the resources listed in the Further Reading section.

**See also: Chemometrics and Statistics:** Experimental Design; Optimization Strategies; Multivariate Classification Techniques; Multivariate Calibration Techniques; Expert Systems; Multicriteria Decision Making; Signal Processing; Spectral Deconvolution and Filtering.

## Further Reading

- De Levie R (2001) *How to Use Excel<sup>®</sup> in Analytical Chemistry*. Cambridge: Cambridge University Press.
- ISO (1993) *Guide to the Expression of Uncertainty in Measurement*. Geneva: International Organisation for Standardisation.
- Meier PC and Zund RE (2000) *Statistical Methods in Analytical Chemistry*, 2nd edn. New York: Wiley.
- Miller JN and Miller JC (2000) *Statistics and Chemometrics for Analytical Chemistry*, 4th edn. Harlow: Pearson Education.
- Neave HR (1981) *Elementary Statistics Tables*. Routledge: London.

## Experimental Design

**G Hanrahan, J Zhu, S Gibani, and D G Patil,**  
California State University, Los Angeles, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by J Goupy, pp. 659–666, © 1995, Elsevier Ltd.

## Introduction

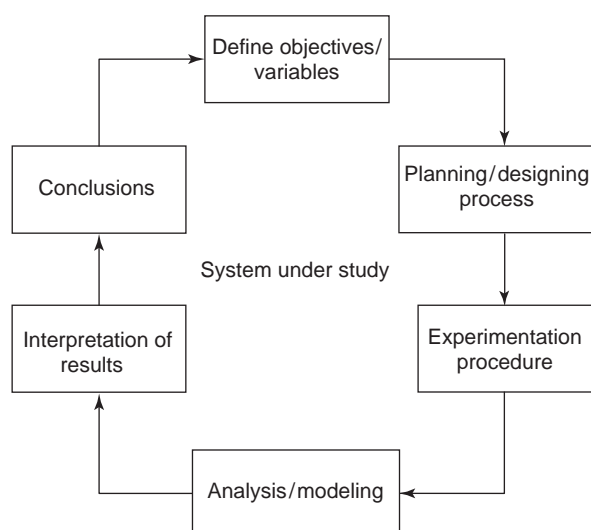
Experimental design methods allow the experimenter to understand better and evaluate the factors that influence a particular system by means of statistical approaches. Such approaches combine theoretical knowledge of experimental designs and a working knowledge of the particular factors to be studied. Although the choice of an experimental design ultimately depends on the objectives of the experiment and the number of factors to be investigated, initial experimental planning (as shown in Figure 1) is essential.

The relationship between the various factors and response within a given system can be shown

mathematically as follows:

$$y = f(x_1, x_2, x_3, \dots, x_k) \quad [1]$$

where  $y$  is the response of interest in the system and  $x$  are the factors that affect the response when their values change. In general, the following types of factors can be distinguished: (1) continuous, e.g., temperature; and (2) discrete, e.g., experimenters. Factors are considered to be independent if there is no relationship between them and dependent if a relationship exists. The values or settings attributed for each factor are called levels. Each experimental run in an experimental design study requires that one or more treatments (stimulus applied to one or more factors) be applied to the system and the response measured. The experimenter then employs statistical design methods to determine if the treatment of interest or combination of treatments was significant in influencing the response of the system under study. Calculation of the treatment effects can then be used to identify which variables lead to an optimal response.



**Figure 1** Essential criteria during early experimental planning.

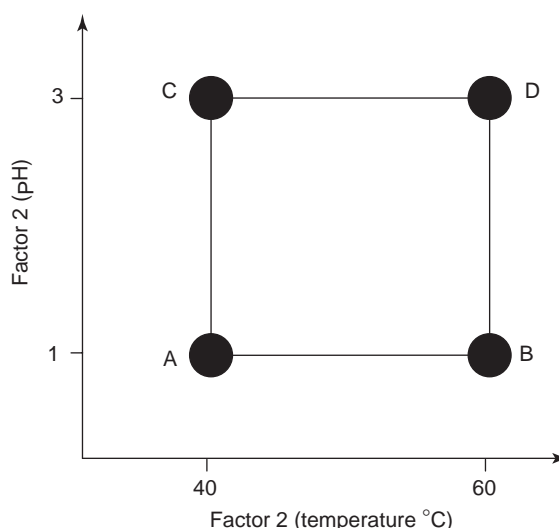
The two main applications of experimental design are screening, in which the factors that influence the experiment are identified, and optimization, in which the optimal settings or conditions for an experiment are found. A screening experiment is a systematic approach to identifying the key input parameters of a process or product that affect the output performance. The usual approach is to start with a screening design including all controllable factors that may possibly influence the experiment, identify the most important ones, and then proceed with an experimental optimization design.

## Experimental Designs

### Full Factorial Designs (Two Levels per Factor)

The most general two-level design is a full factorial design and described as  $2^k$ -designs where the base 2 stands for the number of factor levels and  $k$  the number of factors each with a high and low value. In a full factorial design, the levels of the factors are chosen in such a way that they span the complete factor space. Often, only a lower and upper level is chosen. With two factors, this defines a square in the factor space, and with three factors, this defines a cube. The lower level is usually indicated with a '−' sign; the higher level with a '+' sign.

The method can be graphically illustrated in a simplified example: the effects of reaction temperature and pH in determining the spectrophotometric response (absorbance) of a standard analyte solution. **Figure 2** shows a graphical definition of the experimental domain, with the reaction temperature varying from 40°C (low level) to 60°C (high level)



**Figure 2** Graphical definition of the effects of reaction temperature and pH in determining the spectrophotometric response of a standard analyte solution.

**Table 1** Experimental matrix: spectrophotometric response of a standard analyte solution

Experiment number	Temperature	pH	Response
A	− 1	− 1	$y_1$
B	+ 1	− 1	$y_2$
C	− 1	+ 1	$y_3$
D	+ 1	+ 1	$y_4$
Factor levels			
(−)	40°C	pH 1	
(+)	60°C	pH 3	

and the reaction pH varying from 1 (low level) to 3 (high level). The best experimental points in the domain are located in the corners A, B, C, and D as follows: A (40°C, pH 1); B (60°C, pH 1); C (40°C, pH 3); D (60°C, pH 3).

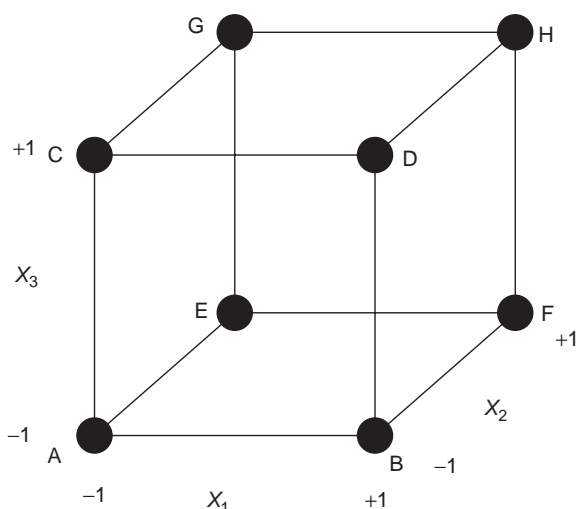
The four trials of experimental matrix used in this experiment are shown in **Table 1**, with the results of each experiment indicated in the response column and the factor levels in the rows below the experimental matrix. Note that −1 is used for the low level of each factor and +1 for the high level.

If we introduce another variable (e.g., reagent concentration) in the experiment, it is then possible to represent the factors as faces on one or more cubes with the responses at the points. The distribution of experimental points within this type of experimental domain ( $2^3$  design) is shown schematically in **Figure 3**.

### Fractional Factorial Design

Fractional factorial designs are arguably the most widely used designs in experimental investigations,





**Figure 3** Full factorial design at two levels,  $2^3$  design.

and mainly used for the screening portion of experiments. Such designs are good alternatives to a full factorial design, especially in the initial stage of a project, and considered a carefully prescribed and representative subset of a full factorial design. In fractional factorial designs, the number of experiments is reduced by a number  $p$  according to a  $2^{k-p}$  design. In the most commonly employed fractional design, the half-fraction design ( $p=1$ ), exactly one-half of the experiments of a full design are performed.

Suppose a situation occurs in which three factors, each at two levels are of interest, but the experimenter does not want to run all eight treatment combinations ( $2^3 = 8$ ). A design with four treatment combinations can then be performed when considering the one-half fraction of the  $2^3$  design ( $2^{3-1} = 4$ ).

The fractional factorial design is based on an algebraic method of calculating the contributions of factors to the total variance with less than a full factorial number of experiments. Such designs are useful when the numbers of potential factors are relatively large because they reduce the total number of runs required for the overall experiment. However, by reducing the number of runs, a fractional factorial design will not be able to evaluate the impact of some of the factors independently.

### Latin Squares

A Latin square is a block design with the arrangement of  $v$  Latin letters into a  $v \times v$  array (a table with  $v$  rows and  $v$  columns). Latin square designs are often used in experiments where subjects are allocated treatments over a given time period where time is

thought to have a major effect on the experimental response. Suppose the treatments are labeled A, B, and C. In this particular situation, the design would be

Day 1	A B C
Day 2	C A B
Day 3	B C A

This type of design allows the separation of an additional factor from an equal number of blocks and treatments. If there are more than three blocks and treatments, then a number of Latin square designs are possible. It can be noted that Latin square designs are equivalent to specific fractional factorial designs (e.g., the  $4 \times 4$  Latin square design is equivalent to a  $4^{3-1}$  fractional factorial design).

### Greco-Latin Squares

The Greco-Latin square design involves two Latin squares that are superimposed on each other. It contains two treatment factors instead of one and contains four factors overall instead of three. An example design would look as follows:

	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>
B <sub>1</sub>	C <sub>1</sub> D <sub>3</sub>	C <sub>2</sub> D <sub>4</sub>	C <sub>3</sub> D <sub>1</sub>	C <sub>4</sub> D <sub>2</sub>
B <sub>2</sub>	C <sub>4</sub> D <sub>2</sub>	C <sub>1</sub> D <sub>1</sub>	C <sub>2</sub> D <sub>3</sub>	C <sub>3</sub> D <sub>4</sub>
B <sub>3</sub>	C <sub>3</sub> D <sub>1</sub>	C <sub>4</sub> D <sub>2</sub>	C <sub>1</sub> D <sub>3</sub>	C <sub>2</sub> D <sub>4</sub>
B <sub>4</sub>	C <sub>2</sub> D <sub>4</sub>	C <sub>1</sub> D <sub>3</sub>	C <sub>3</sub> D <sub>2</sub>	C <sub>4</sub> D <sub>1</sub>

The analysis for the Greco-Latin square design is similar to that of a Latin square design. However, one noticeable difference is that two treatment sum of squares have to be computed (factors C and D) by listing two sets of means outside the design table. As an additional note, Greco-Latin squares are most effective if replicated and are subject to the same randomization rules as for the Latin squares.

### Response Surface Designs (More than Two Levels for One or More Factors)

Response surface methodology is designed to allow experimenters to estimate interactions, therefore giving them an idea of the shape of the response surface they are investigating. This approach is often used when simple linear and interaction models are not adequate, e.g., experimentation far from the region of optimum conditions. Here, the experimenter can expect curvature to be more prevalent and will need a mathematical model, which can represent the curvature. The simplest such model has the quadratic

form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 \quad [2]$$

which contains linear terms for all factors, squared terms for all factors, and products of all pairs of factors. The two most common designs generally used in response surface modeling are central composite designs and Box–Behnken designs. In these designs the inputs take on three or five distinct levels, but not all combinations of these values appear in the design.

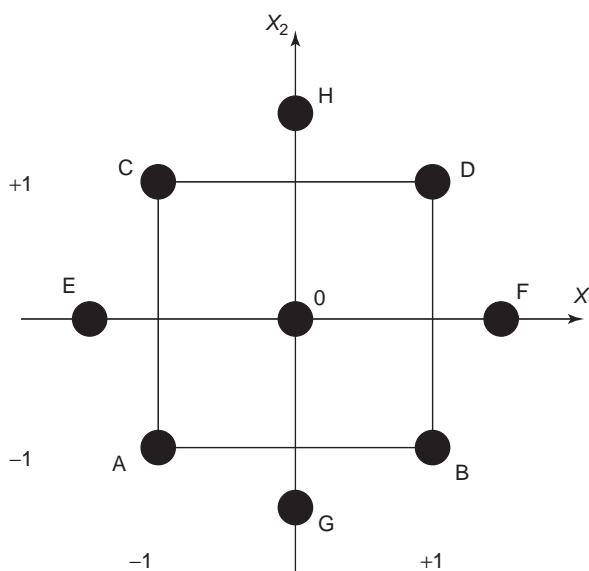
Central composite designs contain imbedded factorial or fractional factorial design with center points that is augmented with a group of axial (star) points that allow estimation of curvature (Figure 4). One central composite design consists of cube points at the corners of a unit cube that is the product of the intervals  $[-1, 1]$ , star points along the axes at or outside the cube, and center points at the origin. Points A, B, C, and D are the points of the initial factorial design with points E, F, G, and H being the star points at the central 0.

A central composite design always contains twice as many star points as there are factors in the design. The star points represent new extreme values (low and high) for each factor in the design. There are three types of central composite designs, ultimately depending on where the star points are placed. Circumscribed central composite designs are the original forms of the central composite design. The star points are at some distance  $\alpha$  from the center based on the properties desired for the design and the number of factors in the design. The  $\alpha$  values depend on the number of factors in the factorial part of the design.

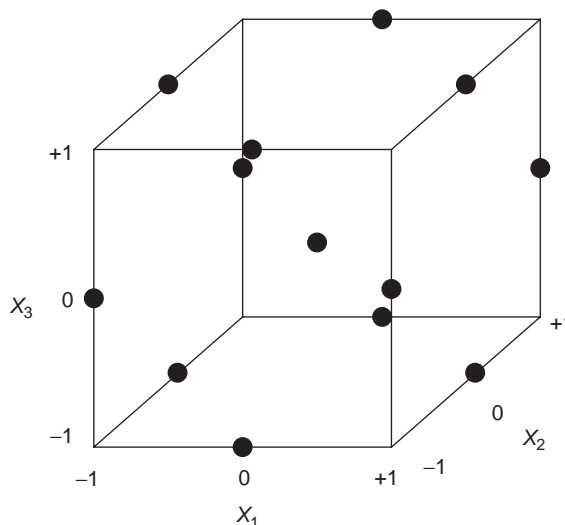
Inscribed central composite designs are a scaled down version of the circumscribed version in that they use the factor settings as the star points and create factorial or fractional factorial design within those limits. In other words, inscribed designs are those where each factor of the circumscribed version is divided by  $\alpha$  for their generation. Face centered central composite designs occur when the star points are at the center of each face of the factorial space, where  $\alpha = \pm 1$ .

### Box–Behnken Designs

The Box–Behnken is considered an efficient option in response surface methodology and an ideal alternative to central composite designs. It has three levels per factor, but avoids the corners of the space, and fills in the combinations of center and extreme levels (Figure 5). It combines a fractional factorial with incomplete block designs in such a way as to avoid the



**Figure 4** Central composite design consisting of a full factorial two-level and star design.



**Figure 5** Box–Behnken design with three levels per factor.

extreme vertices and to present an approximately rotatable design with only three levels per factor. A design is rotatable if the variance of the predicted response at any point  $x$  depends only on the distance of  $x$  from the design center point. It must be noted, however, that Box–Behnken designs should be confined to uses where the experimenter is not interested in predicting response at extremes (corners of the cube).

### Mixture Designs

In a mixture design experiment, the independent factors are proportions of different components of a

blend and often measured by their portions, which sum to 100% or normalized to 1, i.e.,

$$\sum_{i=1}^N X_i = 1 \quad \text{for } x_i \geq 0 \quad [3]$$

As shown, mixture components are subject to the constraint that they must equal to the sum of one. In this case, standard mixture designs for fitting standard models such as simplex-lattice and simplex-centroid designs are employed. When mixtures are subject to additional constraints, constrained mixture designs (extreme-vertices) are then appropriate. Like the factorial experiments discussed above, mixture experimental errors are independent and identically distributed with zero mean and common variance. In addition, the true response surface is considered continuous over the region being studied. Overall, the measured response is assumed to depend only on the relative proportions of the components in the mixture and not on the amount.

**See also:** **Chemometrics and Statistics:** Optimization Strategies; Multivariate Calibration Techniques.

## Further Reading

- Box GEP, Hunter WG, and Hunter JS (1987) *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*. New York: Wiley.
- Cornell JA (1990) *Experiments with Mixtures*. New York: Wiley.
- Dean A and Voss D (1999) *Design and Analysis of Experiments*. New York: Springer.
- Goupy J (1993) *Methods for Experimental Design. Principles and Applications for Physicists and Chemists*. Amsterdam: Elsevier.
- Kemphorne O (1976) *Design and Analysis of Experiments*. New York: Wiley.
- Martens H and Martens M (2001) *Multivariate Analysis of Quality*. Chichester: Wiley.
- Massart DL, Vandeginste BGM, Buydens LMC, et al. (1997) *Handbook of Chemometrics and Qualimetrics, Part A*. Amsterdam: Elsevier.
- Miller JN and Miller JC (2000) *Statistics and Chemometrics for Analytical Chemistry*, 4th edn. New York: Prentice-Hall.
- Morgan E (1991) *Chemometrics: Experimental Design*. Chichester: Wiley.
- Myers RH and Montgomery DC (2002) *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. New York: Wiley.
- Otto M (1999) *Chemometrics: Statistics and Computer Application in Analytical Chemistry*. Chichester: Wiley-VCH.
- Sharaf MA, Illman DL, Kowalski BR (1986) *Chemometrics*. Chemical Analysis Series, vol. 82. New York: Wiley.
- Walters FH, Parker LR, Morgan SL, and Deming SN (1991) *Sequential Simplex Optimization*. Boca Raton, FL: CRC Press.

## Glossary

- Blocking** procedure by which experimental units are grouped into homogeneous clusters in an attempt to improve the comparison of treatments by randomly allocating the treatments within each cluster or 'block'.
- Control** A control is a treatment, which is included to provide a reference set of data which can be compared with data obtained from the experimental treatments.
- Experiment** An investigation in which the investigator applies some treatment(s) to experimental units to be observed and evaluated by measuring one or more response variables.
- Experimental unit** A physical entity or subject subjected to the treatment independently of other units.
- Factor** A categorical explanatory variable studied in an experiment, e.g., pH, flow rate.
- Factorial designs** A factorial design is used to evaluate two or more factors simultaneously. The treatments are combinations of levels of the factors. The advantages of factorial designs over one-factor-at-a-time experiments are that they are more efficient and allow interactions to be detected.
- Levels** The different values assigned to a factor.
- Randomization** A random assignment of experimental material to treatments prior to the start of the experiment. Randomization is vital in the experimental design process and provides: (1) the basis for a valid interpretation of the experimental outcomes in terms of a test of statistical significance, and (2) the basis for computing a valid estimate of experimental error by justifying the assumption of independence of responses over experimental units.
- Replication** When a given combination of factors is present in a system, replication can be used to: (1) demonstrate that results are reproducible, (2) provide a degree of assurance against erroneous results due to unforeseen reasons, (3) provide the means to estimate experimental error, and (4) provide the capacity to increase

	the precision for proper estimates of treatment means.		to address the objectives of a particular experiment.
Response variable	A characteristic of an experimental unit measured after treatment and analyzed	Treatments	The set of circumstances created for an experiment.

## Optimization Strategies

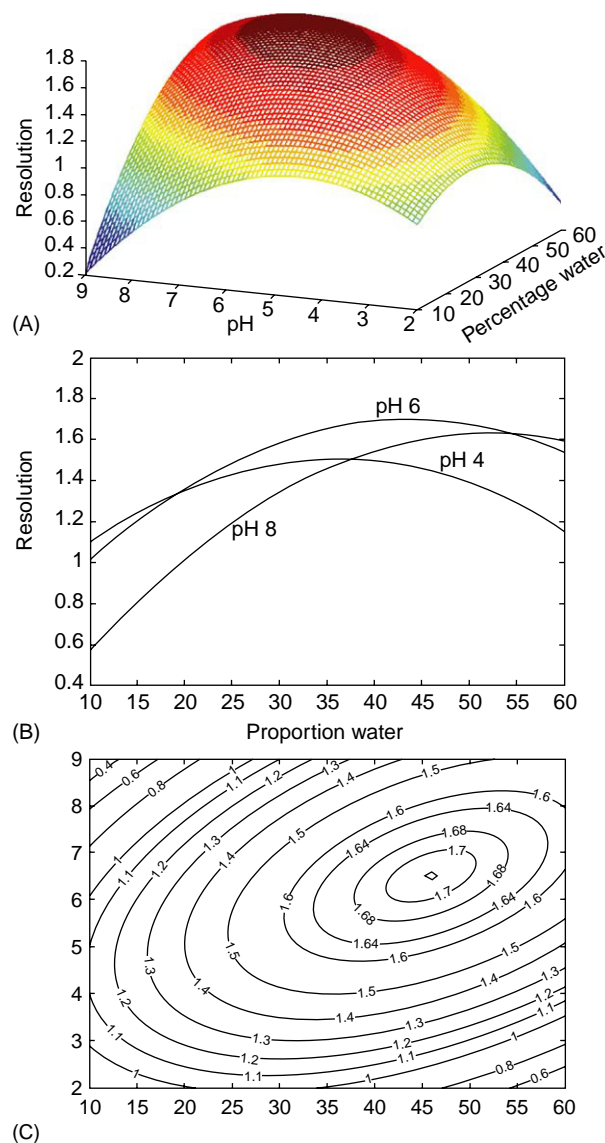
R G Brereton, University of Bristol, Bristol, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Optimization has a significant role in analytical science. There are many reasons for finding an optimum. For example, it may be important to maximize the extraction efficiency of a compound from a matrix; there may be a large number of factors involved in the extraction procedure. Other examples involve improving chromatographic separations and optimizing the factors that influence signal intensity in atomic spectroscopy.

Traditional methods involve studying the influence of one factor at a time. For example, we might want to look at chromatographic resolution of two compounds using isocratic conditions, as a function of proportion of water (using a mobile phase of water and acetonitrile) and acidity (as controlled by the nature and amount of buffer). We do not know how the resolution varies as a function of proportion of water and pH prior to the experiment. **Figure 1A** illustrates a possible underlying response surface. We might try to set pH at a constant level, then vary the percentage water, until we reach an optimum, then set this percentage constant, varying the pH until a fresh optimum is chosen, which we use as the best conditions. The problem with this strategy is illustrated in **Figure 1B**. The change in resolution as a function of proportion of water differs according to pH, so a different optimum proportion of water is found according to the pH that is used. Hence, if we performed an optimization in which the initial experiments were performed at pH 4, we would obtain a different result for the best conditions to an optimization at pH 8 and so, in both cases a false optimum unless we happen by accident to have hit on the correct pH in the original experiments. The reason for this is that the influence of pH and mobile phase composition are said to interact and so cannot be considered independent factors. If there are several different factors involved in an optimization



**Figure 1** (A) Response surface for a typical chromatographic resolution as pH and proportion of water (in a water–acetonitrile mixture) is altered. (B) Resolution as a function of proportion of water at three different values of pH. (C) Representation as a contour plot.

it can be extremely hard to find true optima using traditional methods.

In order to overcome these difficulties, it is necessary to employ systematic approaches for determination

	the precision for proper estimates of treatment means.		to address the objectives of a particular experiment.
Response variable	A characteristic of an experimental unit measured after treatment and analyzed	Treatments	The set of circumstances created for an experiment.

## Optimization Strategies

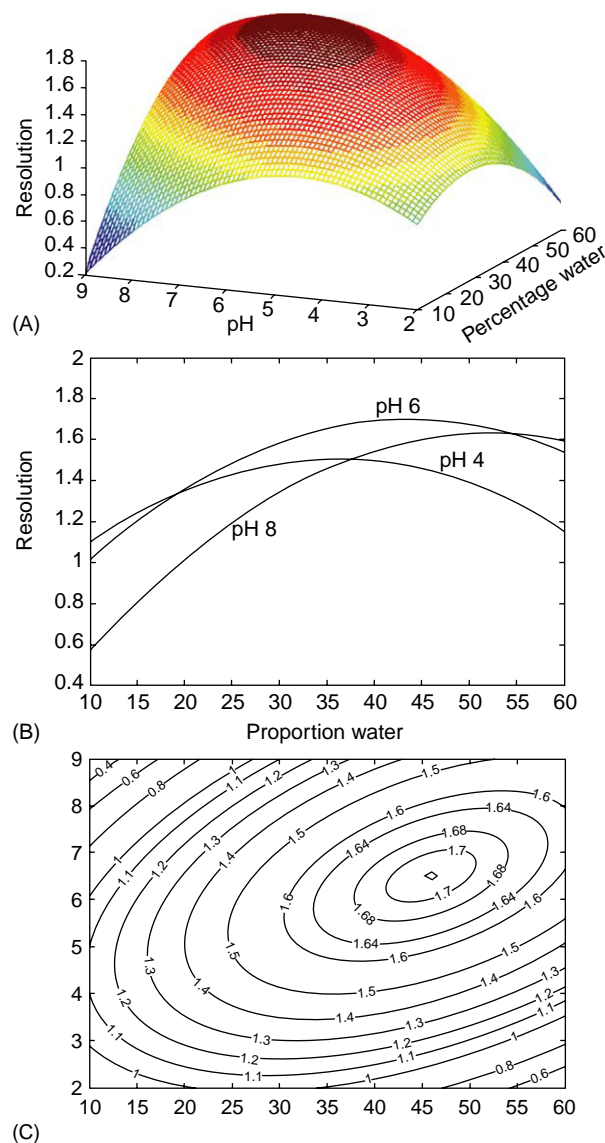
R G Brereton, University of Bristol, Bristol, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Optimization has a significant role in analytical science. There are many reasons for finding an optimum. For example, it may be important to maximize the extraction efficiency of a compound from a matrix; there may be a large number of factors involved in the extraction procedure. Other examples involve improving chromatographic separations and optimizing the factors that influence signal intensity in atomic spectroscopy.

Traditional methods involve studying the influence of one factor at a time. For example, we might want to look at chromatographic resolution of two compounds using isocratic conditions, as a function of proportion of water (using a mobile phase of water and acetonitrile) and acidity (as controlled by the nature and amount of buffer). We do not know how the resolution varies as a function of proportion of water and pH prior to the experiment. **Figure 1A** illustrates a possible underlying response surface. We might try to set pH at a constant level, then vary the percentage water, until we reach an optimum, then set this percentage constant, varying the pH until a fresh optimum is chosen, which we use as the best conditions. The problem with this strategy is illustrated in **Figure 1B**. The change in resolution as a function of proportion of water differs according to pH, so a different optimum proportion of water is found according to the pH that is used. Hence, if we performed an optimization in which the initial experiments were performed at pH 4, we would obtain a different result for the best conditions to an optimization at pH 8 and so, in both cases a false optimum unless we happen by accident to have hit on the correct pH in the original experiments. The reason for this is that the influence of pH and mobile phase composition are said to interact and so cannot be considered independent factors. If there are several different factors involved in an optimization



**Figure 1** (A) Response surface for a typical chromatographic resolution as pH and proportion of water (in a water–acetonitrile mixture) is altered. (B) Resolution as a function of proportion of water at three different values of pH. (C) Representation as a contour plot.

it can be extremely hard to find true optima using traditional methods.

In order to overcome these difficulties, it is necessary to employ systematic approaches for determination



of the optimum. The simplest method for overcoming the problem could be a grid design. This would involve performing the experiment at a range of different pHs and proportion of water (e.g., 10 in each case), there would be 100 possible combinations. In every case, the separation is performed and the resolution measured. This would prevent the experimenter from missing key information that could happen by accident using traditional approaches. Once it appears that an optimum has been found, another series of experiments could be performed around the best points in the grid, using a finer step-size, should it be desired to obtain an even more accurate optimum. Although this method is likely to yield an appropriate result, there are several weaknesses. The major one is that a large number of experiments need to be performed. This wastes time and if there are several factors involved is impracticable. A second one is that experiments are often irreproducible, and if the optimum is flat, it can often be hard to find an exact position. The third is that it ignores observations that could easily be made during the process of the optimization. **Figure 1C** is the contour representation of the response. It is clear that the optimum will not be found in the top left, so performing a large number of experiments in that region is a waste of time. The experimenter only needs to perform a few to notice that this is far from the optimum.

Therefore, it is possible to refine the simple grid search to help find an optimum more quickly. There are statistical rules that allow optimization to be performed rapidly and safely, which can be used formally by the analytical scientist. This article describes some of the main approaches.

## Screening

The first step in many designs involves screening. Often there may be a large number of possible factors that influence a response. For example, we may be interested in determining the extraction efficiency of a process. There may be many different factors, such as oven temperature, pH, extraction time, nature of the filter, enzymes, or reagents added (where appropriate), and so on. The first step an experimenter should do is to list these factors. Some will be irrelevant and so it is not worth spending time on studying their influence, whereas others may have a significant effect. The first step in many optimizations is to eliminate those factors that are not interesting, and this is called screening. In order to do this it is normal to perform a series of experiments at different levels of the factors.

## Full Factorial Designs

The principles of how these designs are set up are described in more detail elsewhere in this encyclopedia.

Although it is not difficult to perform a screening experiment for two factors, in realistic cases it is more likely that there will be many more factors; for example, if we want to screen for six factors,  $2^6$  or 64 experiments are required. The extra experiments are mainly used to study what are called the interaction terms. For example, a two-factor full factorial design at two levels requires  $2^2$  or four experiments. These experiments can be used to obtain a model between the response and the level of the factors often expressed in the form

$$y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2$$

where  $y$  is the response, the  $x$ s represent the coded values of the factors and the  $b$ s are the coefficients. The term  $b_{12}$  is called an 'interaction' term; if four experiments are performed, four terms are studied, one of which is an interaction factor. For a six-factor, 64 experiment, design, most of the experiments are used to study interaction terms including five- or six-term interactions between the factors. Whereas these terms may be interesting, especially the lower order interactions, for the purpose of screening we are primarily interested in the significance of the single-factor terms, and it is only necessary to perform  $f+1$  experiments for an  $f$  factor design (the extra experiment being used to determine the intercept or average), or seven experiments in the case of six factors. Hence,  $64 - 7$  or 57 of the original experiments are unnecessary for the purpose of screening, wasting time. If the number of factors is small (e.g., three), it is still worth doing a full factorial design, but when the number of factors becomes large it is prohibitive to perform a full factorial, and tricks are required to reduce the number of experiments.

## Fractional Factorial Designs

Consider a three-factor, two-level design. Eight experiments are listed in **Table 1**, the conditions being coded as usual. How can we reduce the number of experiments safely and systematically? Two-level fractional factorial designs are used to reduce the number of experiments by 1/2, 1/4, 1/8, and so on. Can we halve the number of experiments? At first glance, a simple approach might be to take the first four experiments of **Table 1**. However, these would leave the level of the first factor at  $+1$  throughout. A problem is that we now no longer study the variation of this factor, so we do not obtain any information on how factor 1 influences the response.

Can a subset of four experiments be selected that allows us to study all three factors? Rules have been developed to produce these fractional factorial designs obtained by taking the correct subset of the original experiments. **Table 2** illustrates a possible fractional factorial design that enables all factors to be studied.

The matrix of effects can be calculated as presented in **Table 2** and is also interesting. A full model for three factors is given by

$$y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3$$

Note that there are eight terms, corresponding to eight experiments. The last four terms are not interesting at this stage, but the values of  $x$  can be calculated, simply by multiplying the relevant columns. Whereas the first four columns of this matrix are all different, the last four each correspond to one of the first four columns. For example the  $x_1x_2$  column exactly equals the  $x_3$  column. What does this imply? As the number of experiments is reduced, the amount of information is correspondingly reduced. Since only four experiments are now performed, it is only possible to measure four unique factors. The interaction between factors 1 and 2 is said to be 'confounded' with factor 3. This might mean, for example, that, using this design the interaction between factors 1 and 2 is indistinguishable from the influence of factor 3 alone. However, not all interactions will be

significant, and the purpose of a preliminary experiment is often simply to sort out which main factors should be studied in detail later.

Note that two-level fractional factorial designs only exist when the number of experiments equals a power of 2. In order to determine the minimum number of experiments, the following procedure is adopted:

1. Determine how many terms are interesting.
2. Then construct a design whose size is the next greatest power of 2.

### Plackett–Burman Designs

Where the number of factors is quite large, the constraint that the number of experiments must equal a power of 2 can be rather restrictive. Since the number of experiments must always exceed the number of factors by at least one, this would mean that 32 experiments are required for the study of 19 factors, and 64 experiments for the study of 43 factors. In order to overcome this problem and reduce the number of experiments further, other approaches are needed.

Plackett and Burman proposed a number of two-level factorial designs, where the number of experiments is a multiple of 4 as opposed to a power of 2. Hence, designs exist for 4, 8, 12, 16, 20, 24, etc., experiments. The number of experiments exceeds the number of factors,  $f$ , by 1.

One such design is given in **Table 3** for 11 factors and 12 experiments and has various features:

1. In the first row, all factors are at the same level.
2. The first column from rows 2 to  $f$  is called a 'generator'. The key to the design is that there are only certain allowed generators that can be obtained from tables. Note that the number of factors will always be an odd number equal to  $f = 4m - 1$  (or 11 in this case), where  $m$  is any integer. If the first row consists of '–', the generator will consist of  $2m$  ( $= 6$  in this case) experiments at the '+' level and  $2m - 1$  ( $= 5$  in this case) at the '–' level, the reverse being

**Table 1** Three-factor full factorial design with a '+' indicating a high level and a '–' indicating a low level of each factor

Experiments	Factor 1	Factor 2	Factor 3
1	+	+	+
2	+	+	–
3	+	–	+
4	+	–	–
5	–	+	+
6	–	+	–
7	–	–	+
8	–	–	–

**Table 2** Developing a three-factor fractional factorial design

Experiments			Matrix of effects							
Factor 1	Factor 2	Factor 3	$x_0$	$x_1$	$x_2$	$x_3$	$x_1x_2$	$x_1x_3$	$x_2x_3$	$x_1x_2x_3$
+	+	+	+	+	+	+	+	+	+	+
+	–	–	+	+	–	–	–	–	+	+
–	–	+	+	–	–	+	+	–	–	+
–	+	–	+	–	+	–	–	+	–	+



**Table 3** Plackett–Burman design for 11 factors

Experiments	Factors										
	1	2	3	4	5	6	7	8	9	10	11
1	–	–	–	–	–	–	–	–	–	–	–
2	+	–	+	–	–	–	+	+	+	–	+
3	+	+	–	+	–	–	–	+	+	+	–
4	–	+	+	–	+	–	–	–	+	+	+
5	+	–	+	+	–	+	–	–	–	+	+
6	+	+	–	+	+	–	+	–	–	–	+
7	+	+	+	–	+	+	–	+	–	–	–
8	–	+	+	+	–	+	+	–	+	–	–
9	–	–	+	+	+	–	+	+	–	+	–
10	–	–	–	+	+	+	–	+	+	–	+
11	+	–	–	–	+	+	+	–	+	+	–
12	–	+	–	–	–	+	+	+	–	+	+

true if the first row is at the ‘+’ level. In Table 3, the generator is + + – + + + + – – – + –. 3. The next  $4m - 2$  ( $= 10$  in this case) columns are generated from the first column simply by shifting the down cells by one row. This is indicated by diagonal arrows in the table. Notice that experiment 1 is not included in this procedure.

4. The level of factor  $j$  in experiment (or row) 2 equals to the level of this factor in the row  $f$  for factor  $j - 1$ . For example, the level of factor 2 in experiment 2 equals the level of factor 1 in experiment 12.

There are as many high as low levels of each factor over the 12 experiments, as would be expected. The most important property of the design, however, is called ‘orthogonality’. Consider the relationship between factors 1 and 2. There are six instances in which factor 1 is at a high level, and six at a low level. For each of the six instances at which factor 1 is at a high level, in three cases factor 2 is at a high level, and in the other three cases it is at a low level. A similar relationship exists where factor 1 is at a low level. This implies that the factors are orthogonal or uncorrelated, an important condition for a good design. Any combination of two factors is related in a similar way. Only certain generators possess all these properties, so it is important to use only known generators for the designs.

Standard Plackett–Burman designs exist for 7, 11, 15, 19, and 23 factors; generators are given in Table 4. Note that for 7 and 15 factors it is also possible to use conventional fractional factorial designs, and it can be shown that the two approaches are equivalent, except that the columns and rows are swapped around; so, in such situations, it makes no difference as to which approach is adopted.

**Table 4** Generators for Plackett–Burman designs

Factors	Generator
7	+ + + – + – –
11	+ + – + + + – – – + –
15	+ + + + – + – + + – – + – – –
19	+ + – + + + + – + – + – – – – + +
	–
23	+ + + + + – + – + + – – + + – – +
	– + – – – –

If the number of experimental factors is less than that of a standard design (a multiple of  $4 - 1$ ), the final factors are ‘dummy’ ones. Hence, if there are only 10 real factors, use an 11 factor design, the final factor being a dummy one: this may be a variable that has no effect on the experiment, such as the technician who handed out the glassware, or the color of laboratory furniture.

## Designs for Optimization Requiring Modeling

For a typical analytical chemical problem, preliminary considerations, if necessary using screening designs, will reduce the number of factors to a manageable amount, typically between two and five. The next phase is to study these in detail, including their interactions. The most common approaches require a series of experiments, the results of which are then analyzed using a mathematical model, which results in an equation that expresses the response as a function of the coded values of the factors. The optimum of the response surfaces (e.g., the best resolution or highest extraction efficiency) can be determined in one of three ways. The first is by using a computational optimization package, often available as part of commercial package for the analysis data obtained from experimental designs. The second is by visualizing a response surface and determining the optimum graphically; this is feasible mainly if the number of factors is limited. The third is algebraically, using calculus, solving differential equations.

Note that although a model may already have been obtained using a screening design, such a model is rarely useful for optimization studies, because usually only linear terms are estimated. Most screening designs do not involve recording replicates, which is important in order to know how well the model is obeyed, nor do they provide information on squared terms, for example, the dependence of the response on  $x^2$ , which are important for optimization; some such as Plackett–Burman and highly fractional factorials do not even provide details of interactions.

Two of the most common classes of design are described below, although there are many other potential methods used throughout analytical chemistry.

### Central Composite or Response Surface Designs

Many designs for optimization are based on the central composite design (sometimes called a response surface design). Table 5 illustrates a typical design used to optimize a procedure with three factors, temperature, pH, and time.

A major advantage of response surface designs is that in addition to linear and interaction terms they can include squared terms, because the factors are at several levels. For many situations in optimization, these terms are important to provide curvature in the resultant response surface. Hence, after using screening designs that mainly indicate which factors are important, it is possible to then study in detail how the factors influence the optimum, and once a model is obtained to predict the experimental optimum conditions.

### Mixture Designs

A separate set of designs are important in analytical chemistry, where the factors add up to a constant total, an example being the composition of a mobile

phase in chromatography. Special designs called mixture designs are used in such situations.

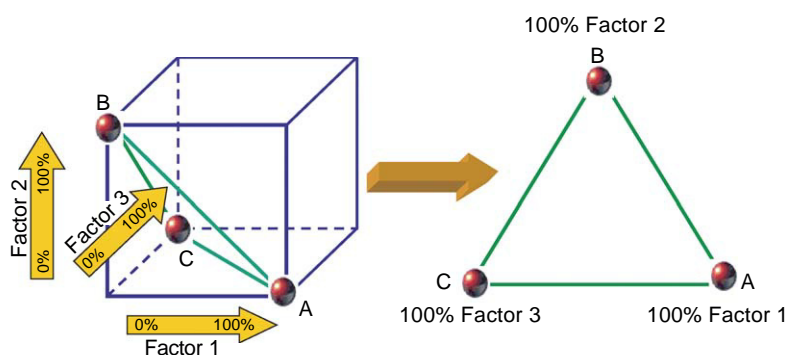
Most chemists represent their experimental conditions in mixture space, which corresponds to all possible allowed proportions of components that add up to 100%. A three component mixture can be represented by a triangle (Figure 2), which is a two-dimensional cross-section of a three-dimensional space showing the allowed region in which the proportions of the three components add up to 100%. Points within this triangle or mixture space represent possible mixtures or blends. As the number of components increases so does the dimensionality of the mixture space. Physically meaningful mixtures can be represented as points in this space, for two components the mixture space is simply a straight line, for three components a triangle, for four components a tetrahedron.

There are a number of common designs that can be envisaged as ways of determining a sensible number and arrangement of points within the simplex.

The most widespread are the simplex centroid designs. For  $f$  factors they involve performing  $2^k - 1$  experiments, i.e., for four factors, 15 experiments are performed. It involves all possible combinations of the proportions 1, 1/2, to 1/ $f$ . A three-factor design consists of three single-factor combinations, three

**Table 5** A three-factor central composite design in both coded and real factors

<i>Coded</i>			<i>Real</i>		
<i>Factor 1</i>	<i>Factor 2</i>	<i>Factor 3</i>	<i>pH</i>	<i>Temperature (°C)</i>	<i>Time (h)</i>
1	1	1	9	50	3
1	1	-1	9	50	1
1	-1	1	9	30	3
1	-1	-1	9	30	1
-1	1	1	5	50	3
-1	1	-1	5	50	1
-1	-1	1	5	30	3
-1	-1	-1	5	30	1
0	0	1.682	7	40	3.68
0	0	-1.682	7	40	0.32
0	1.682	0	7	56.82	2
0	-1.682	0	7	23.18	2
1.682	0	0	10.36	40	2
-1.682	0	0	3.64	40	2
0	0	0	7	40	2
0	0	0	7	40	2
0	0	0	7	40	2
0	0	0	7	40	2
0	0	0	7	40	2
0	0	0	7	40	2
0	0	0	7	40	2
<i>Coding</i>					
	pH	Temperature	Time		
-1	5	30	1		
1	9	50	3		



**Figure 2** Representation of three-factor mixture space.

**Table 6** Three-factor simplex centroid design

Experiment	Factor 1	Factor 2	Factor 3
<i>Single factor</i>			
1	1	0	0
2	0	1	0
3	0	0	1
<i>Binary</i>			
4	1/2	1/2	0
5	1/2	0	1/2
6	0	1/2	1/2
<i>Ternary</i>			
7	1/3	1/3	1/3

**Table 7** Three-factor simplex lattice design

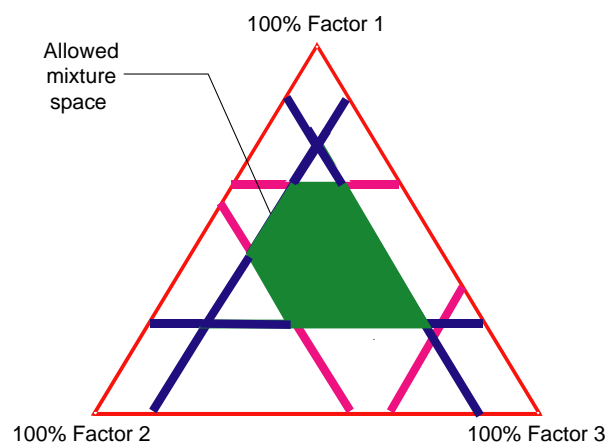
Experiment	Factor 1	Factor 2	Factor 3
<i>Single factor</i>			
1	1	0	0
2	0	1	0
3	0	0	1
<i>Binary</i>			
4	1/3	2/3	0
5	2/3	1/3	0
6	1/3	0	2/3
7	2/3	0	1/3
8	0	2/3	1/3
9	0	1/3	2/3
<i>Ternary</i>			
10	1/3	1/3	1/3

binary combinations, and one ternary combination, as presented in **Table 6**.

Another class of designs called ‘simplex lattice’ designs have been developed and are often preferable to the reduced simplex centroid design when it is required to reduce the number of interactions. They span the mixture space more evenly. A  $\{f, m\}$  simplex lattice design consists of all possible combinations of  $0, 1/m, 2/m, \dots, m/m$ , or a total of  $N = (f + m - 1)! / [(f - 1)!m!]$  experiments where there are  $k$  factors. A  $\{3, 3\}$  simplex lattice design can be set up analogous to the  $\{3, 3\}$  simplex centroid design given in **Table 7**.

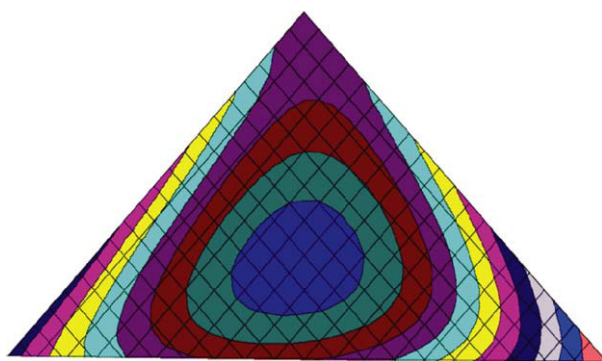
Special designs are required when there are constraints on the factors. This is quite common in analytical chemistry. For example, if we want to study chromatographic separations using a three-solvent mixture, it is often not useful to study the separations using pure solvents. There are several types of constrained mixture designs, which may be represented graphically by a region of the original mixture space. For three components a constrained mixture design is illustrated in **Figure 3**.

Once the design is performed, it is normal to fit a mathematical model to the response (e.g., signal intensity, resolution, extraction efficiency) according



**Figure 3** Constrained mixture design.

to the values of the factors, and this allows the optimal conditions to be calculated. If there are two or three factors, this can be visualized as in **Figure 4**.



**Figure 4** Visualizing an optimum in a three-factor mixture space.

## Sequential Methods for Optimization

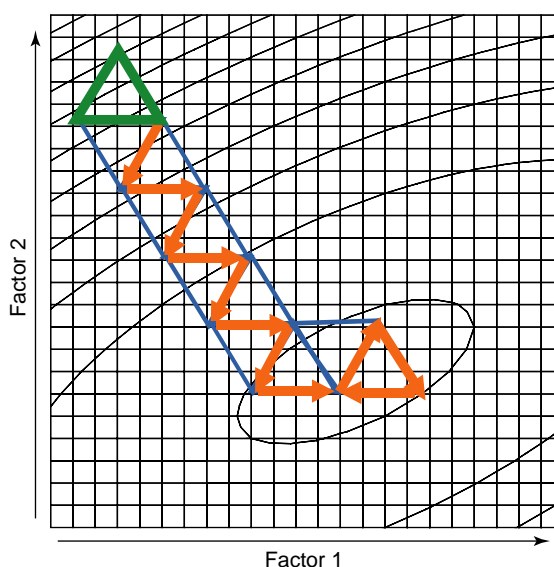
The best known sequential method for optimization is the simplex method. This requires the experimenter to perform a series of experiments until he or she reaches an optimum. The response surface is not mathematically modeled, but the experimenter follows a series of rules until he or she cannot improve. A simplex is the simplest possible object in  $N$  dimensional space, e.g., a line in one dimension, and a triangle in two dimensions. Simplex optimization means that a series of experiments are performed on the corners of such a figure.

The most common, and easiest to understand, method of simplex optimization is called the fixed sized simplex.

The main steps are as follows, exemplified by a two-factor experiment:

1. Define how many factors of interest, which we will call  $k$ .
2. Perform  $k + 1$  ( $= 3$  for two factors) experiments on the vertices of a simplex (or triangle for two factors) in factor space. The conditions for these experiments depend on the step-size. This defines the final 'resolution' of the optimum. The smaller the step-size the better the optimum can be defined, but more experiments are necessary. It is important to establish sensible initial conditions, especially the spacing between the experiments.
3. Rank the response (e.g., the chromatographic resolution) from 1 (worst) to  $k + 1$  (best). In vector form the conditions for the  $n$ th response are given by  $\mathbf{x}_n$  where each element in the vector corresponds to the value of a factor, e.g., pH or eluotropic strength in a chromatographic separation.
4. Establish new conditions for the next experiment as follows:

$$\mathbf{x}_{\text{new}} = \mathbf{c} + \mathbf{c} - \mathbf{x}_1$$



**Figure 5** Progress of a typical fixed size simplex.

where  $\mathbf{c}$  is the centroid of the responses 2 to  $k + 1$  (excluding the worst response), defined by the average of these responses represented in vector form.

5. Continue as in steps 3 and 4 unless the new conditions result in a response that is worst than the remaining  $k$  ( $= 2$ ) conditions, i.e.,  $y_{\text{new}} < y_2$  where  $y$  is the corresponding response and the aim is maximization. In this case return to the previous conditions, and calculate

$$\mathbf{x}_{\text{new}} = \mathbf{c} + \mathbf{c} - \mathbf{x}_2$$

where  $\mathbf{c}$  is the centroid of the responses 1 and 3 to  $k + 1$  (excluding the second worst response) and can also be expressed by  $\mathbf{x}_{\text{new}} = \mathbf{x}_1 + \mathbf{x}_3 - \mathbf{x}_2$ , for two factors. Keep these new conditions together with the worst and the  $k - 1$  best responses from the previous simplex. The second worst response from the previous simplex is rejected.

6. Check for convergence. When the simplex is at an optimum it normally oscillates around in a triangle or hexagon. If the same conditions reappear, stop. There are a variety of stopping rules, but it should generally be obvious when optimization has been achieved.

The progress of a fixed sized simplex is illustrated in **Figure 5**. Many elaborations have been developed over the years. One of the most important is the  $k + 1$  rule. If a vertex has remained part of the simplex for  $k + 1$  steps, perform the experiment again. The reason for this is that response surfaces may be noisy, so

an unduly optimistic response could have been obtained because of experimental error. This is especially important when the response surface is flat near the optimum. Another important problem relates to boundary conditions. Sometimes there are physical reasons why a condition cannot cross a boundary, an obvious case being a negative concentration. It is not always easy to deal with such situations, but it is possible to use step 5 rather than step 4 above under such circumstances. If the simplex constantly tries to cross a boundary either the constraints are a little unrealistic and so should be changed, or the behavior near the boundary needs further investigation. Starting a new simplex near the boundary with a small step-size may solve the problem.

A weakness with the standard method for simplex optimization is a dependence on the initial step-size. Another method is called the modified simplex algorithm and allows the step size to be altered, reduced as the optimum is reached, or increased far from the optimum.

For the modified simplex, step 4 of the fixed sized simplex is changed as follows. A new response at point  $\mathbf{x}_{\text{test}}$  is determined, where the conditions are obtained as for fixed sized simplex:

1. If the response is better than all the other responses in the previous simplex, i.e.,  $y_{\text{test}} > y_{k+1}$  then 'expand' the simplex, so that

$$\mathbf{x}_{\text{new}} = \mathbf{c} + \alpha(\mathbf{c} - \mathbf{x}_1)$$

where  $\alpha$  is a number greater than 1, typically equal to 2.

2. If the response is better than the worst of the other responses in the previous simplex, but worse than the second worst, i.e.,  $y_1 < y_{\text{test}} < y_2$ , then 'contract' the simplex but in the direction of this new response

$$\mathbf{x}_{\text{new}} = \mathbf{c} + \beta(\mathbf{c} - \mathbf{x}_1)$$

where  $\beta$  is a number less than 1, typically equal to 0.5.

3. If the response is worst of the other responses, i.e.,  $y_{\text{test}} < y_1$ , then 'contract' the simplex but in the opposite direction of this new response

$$\mathbf{x}_{\text{new}} = \mathbf{c} - \beta(\mathbf{c} - \mathbf{x}_1)$$

where  $\beta$  is a number less than 1, typically equal to 0.5.

4. In all other cases simply calculate

$$\mathbf{x}_{\text{new}} = \mathbf{x}_{\text{test}} = \mathbf{c} + \mathbf{c} - \mathbf{x}_1$$

as in the normal (fixed-sized) simplex.

Then perform another experiment at  $\mathbf{x}_{\text{new}}$  and keep this new experiment plus the  $k$  ( $=2$ ) best experiments from the previous simplex to give a new simplex. Rule 5 of the fixed sized simplex still applies, if the value of the response at the new vertex is less than that of the remaining  $k$  responses, return to the original simplex and reject the second best response, repeating the calculation as above.

There are yet further sophistications such as the supermodified simplex, which allows mathematical modeling of the shape of the response surface to provide guidelines as to the choice of the next simplex. Simplex optimization is only one of several computational approaches to optimization, including evolutionary optimization, and steepest ascent methods, however, it is the most commonly used sequential method in analytical chemistry, with diverse applications ranging from autosimming of instruments to chromatographic optimizations, and can easily be automated.

See also: **Chemometrics and Statistics: Experimental Design.**

## Further Reading

- Bayne CK and Rubin IB (1986) *Practical Experimental Designs and Optimisation Methods for Chemists*. Deerfield Beach: VCH.
- Brereton RG (2003) *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, ch. 2. Chichester: Wiley.
- Cornell JA (1990) *Experiments with Mixtures: Design, Models, and the Analysis of Mixture Data*, 2nd edn. New York: Wiley.
- Deming SN and Morgan SL (1994) *Experimental Design: A Chemometric Approach*, 2nd edn. Amsterdam: Elsevier.
- Massart DL, Vandeginste BGM, Buydens LMC, *et al.* (1997) *Handbook of Chemometrics and Qualimetrics Part A*, chs. 21–26. Amsterdam: Elsevier.
- Otto M (1998) *Chemometrics: Statistics and Computer Applications in Analytical Chemistry*, ch. 4. Weinheim: Wiley-VCH.



## Multivariate Classification Techniques

M J Adams, RMIT University, Melbourne, VIC, Australia

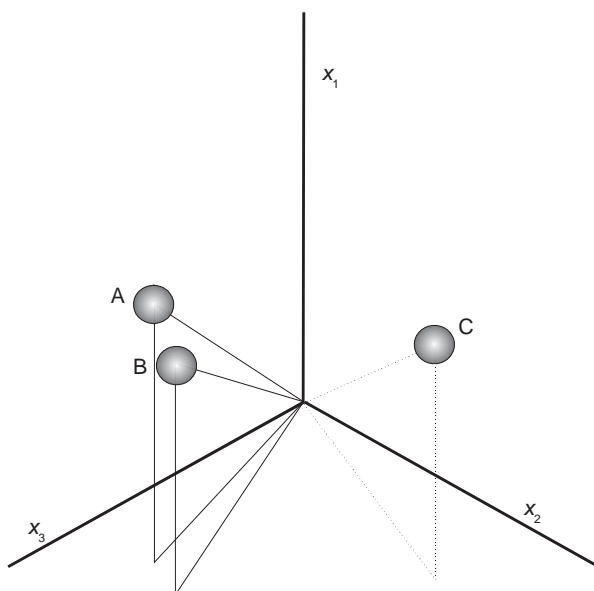
© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

As instrumental chemical analysis techniques have become more sophisticated with increasing levels of automation, the number of samples routinely analyzed has grown and the amount of data per sample has increased and can appear overwhelming.

Multivariate classification techniques attempt to make sense of these data by identifying inherent patterns that may provide insight into the structure and form of the data and, hence, of the samples themselves. This general area of study is referred to as pattern recognition; a field rich in applications in analytical science. For example, the identification of the origin of goods and foodstuffs is an important task for the analyst. Given a limited set of possibilities then the problem at hand is one of classification. From a training set of samples of known origin a classification scheme is developed that can classify unknown, test samples. In the biochemical field, nuclear magnetic resonance (NMR) spectroscopy of biofluids and cells provides a unique insight into changes in metabolism caused by drugs and toxins. Although biological NMR spectra are extremely complex the essential diagnostic parameters are carried in the overall patterns of the spectra. Metabonomics employs pattern recognition methods to interrogate the databases of proton NMR spectra. This approach allows a mathematical classification of toxicity based on disparate types of multidimensional metabolic data so giving new insights into the modes and biochemical mechanisms of toxicity. This work is of value in the prediction of toxicity in drug development studies.

The aims of pattern recognition are to determine summarizing structure within analytical data, using, for example, exploratory data analysis techniques and cluster analysis, and to identify such patterns (classification) according to their correspondence with previously characterized examples. Many numerical techniques are available to the analytical scientist wishing to interrogate their data, but all have the same starting point. The data are expressed in a matrix form in which each sample, or object, is described by a vector of measurements. This representation leads logically to the concept of a pattern space with as many dimensions as the number of



**Figure 1** Three objects, A, B, and C, displayed in the three-dimensional pattern space defined by the variables  $x_1$ ,  $x_2$ , and  $x_3$ . Objects A and B are closer to each other, and therefore considered more similar, than to object C.

variables in the vector, and an object occupying a single point in that multidimensional space. Furthermore, it is assumed that points representing similar patterns (i.e., similar samples) will tend to cluster in the pattern space. Conversely, samples of dissimilar patterns will lie in different regions of space (Figure 1).

### Data Reduction

It is often the case, particularly in spectrochemical analysis, that the number of variables far exceeds the number of samples. This is not surprising given that a single infrared spectrum, for example, can comprise absorption measurements at several thousand wavelengths. Although there are many statistical techniques available for identifying the major variables (features) responsible for defining the pattern space occupied by a sample, by far the most common technique employed in chemometric analysis of analytical data is the method of principal components analysis (PCA). The method is an important tool for analysts in exploratory data analysis.

PCA involves rotating and transforming the original axes representing the original variables into new axes, so that the new axes lie along the directions of maximum variance of the data. These new axes are orthogonal, i.e., the new variables are uncorrelated. Because of the high correlation that frequently exists between analytically measured

variables, it is generally found that the number of new variables needed to describe most of the sample data variance is significantly less than the number of original variables. Thus, PCA provides a means to reduce the dimensionality of the parameter space. In addition, PCA can reveal those variables, or combinations of variables, that describe some inherent structure in the data and these may be interpreted in chemical or physicochemical terms.

Principal components are linear combinations of original variables, and the linear combination with the largest variance is the first principal component (Figure 2). Once this is determined, then the search proceeds to find a second normalized linear combination that has most of the remaining variance and is uncorrelated with the first principal component. The procedure is continued, usually until all the principal components have been calculated and a subset of the

principal components is then selected for further analysis and for interpretation. The scheme for performing PCA is illustrated in Figure 3.

PCA can often be so effective and efficient in reducing the dimensionality of analytical data that it can provide immediate visual indication of patterns within data and is a commonly employed exploratory technique.

## Unsupervised Pattern Recognition

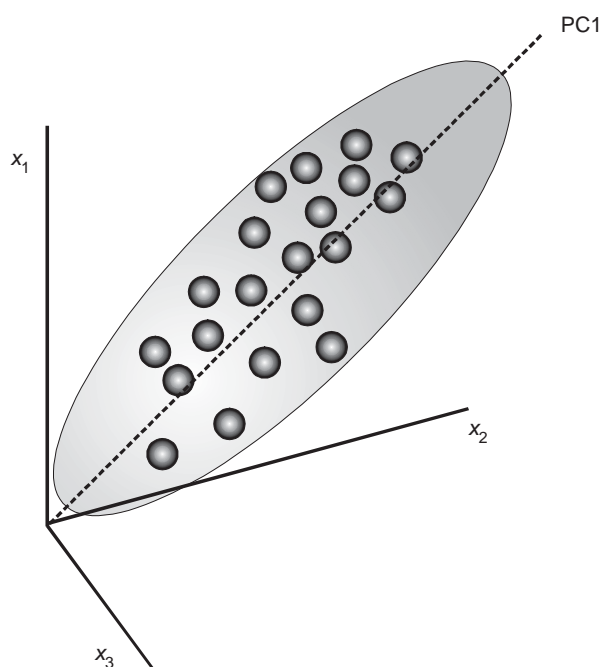
In many initial, exploratory studies information concerning the presence of groups of similar samples, or the pattern design of such groups as may be present, is not known. This may be due to a lack of knowledge of the pattern generating process. Indeed, this could be the principal aim of the study – to determine the inherent designs of patterns existing within the analytical data. The purpose of unsupervised pattern recognition is to identify groups, or clusters, of similar objects characterized by a multivariate feature set. No *a priori* information about the structure of possible groups present is required.

The general aim is to attempt to partition a given dataset into homogeneous subsets (clusters) according to similarities of the points in each subset and their relationship to the elements of other subsets. This approach is referred to as cluster analysis. Quantification of the degree of similarity of objects expressed in pattern space is a key process in cluster analysis.

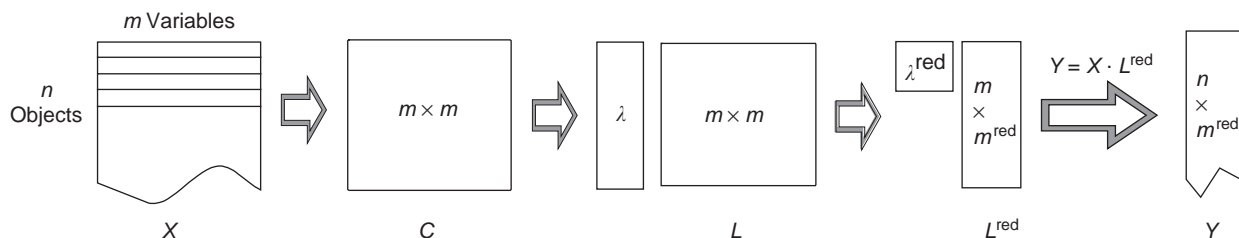
It is generally accepted without proof that similarity and distance are complementary; objects close together in multidimensional space are more alike than those further apart. Most of the similarity measures used in practice are based on some distance function, and whilst many such functions are referenced in the literature the most common is the simple Euclidean distance metric.

In multidimensional space the Euclidean distance,  $d_{(1,2)}$ , between two objects is given by

$$d_{(1,2)} = \sqrt{\sum_{j=1}^m (X_{1,j} - X_{2,j})^2} \quad [1]$$

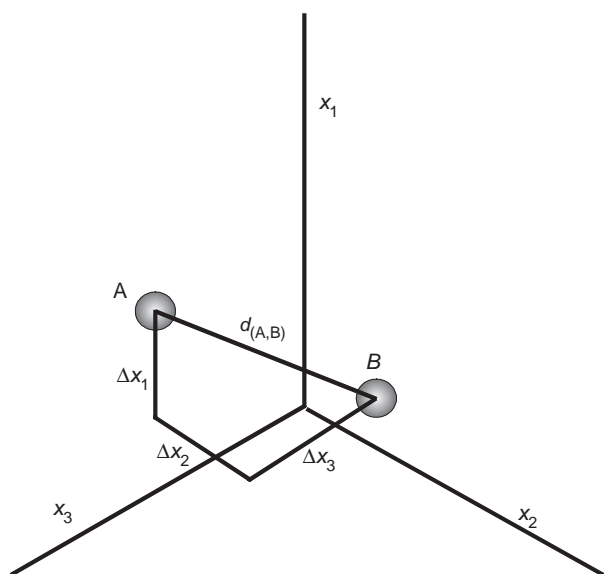


**Figure 2** The first principal component, PC1, is a new axis representing the combination of original variables providing the greatest variance in the data.



**Figure 3** The original data,  $X$ , comprising  $n$  objects or samples described by  $m$  variables, is converted to a dispersion (covariance or correlation) matrix  $C$ . The eigenvalues,  $\lambda$ , and eigenvectors,  $L$ , are extracted from  $C$ . A reduced set of eigenvectors,  $L^{\text{red}}$ , i.e., selected and the original data projected into this new, lower-dimensional pattern space  $Y$ .





**Figure 4** In multidimensional pattern space the Euclidean distance,  $d_{(A,B)}$ , between two objects A and B is provided by the square root of the sum of the squares of the differences between the values of the defining variables,  $\Delta x_j$ .

where  $x_1$  and  $x_2$  are the feature vectors of objects 1 and 2 and  $m$  is the number of variables (Figure 4). The Euclidean distance can be calculated for all pairs of objects and the data matrix transformed into a square, symmetric distance matrix.

There are basically two approaches to data clustering, dynamic methods and hierarchical techniques.

### Dynamic Clustering

Dynamic clustering employs an iterative algorithm to optimize some clustering criterion function such as the average affinity of points to a cluster's mean value – the  $c$ -means algorithm. During each iteration, a data point is assigned to a cluster on the basis of its closeness to the cluster's center. The cluster centers are recalculated to reflect changes brought about by data point assignments, and the new cluster models are used in the next iteration to reclassify the data. The process is continued until a stable partition is obtained, i.e., until no object is reclassified. The number of expected clusters or groups must be specified before commencing the analysis.

With the  $c$ -means algorithm a mean vector is taken as representing a cluster and for asymmetric shaped clusters this will be inadequate. The algorithm can readily be generalized to more sophisticated models so that a cluster is represented not by a single point (such as described by a mean vector) but rather by a function describing some attribute of the cluster's shape in pattern space.

### Hierarchical Cluster Analysis

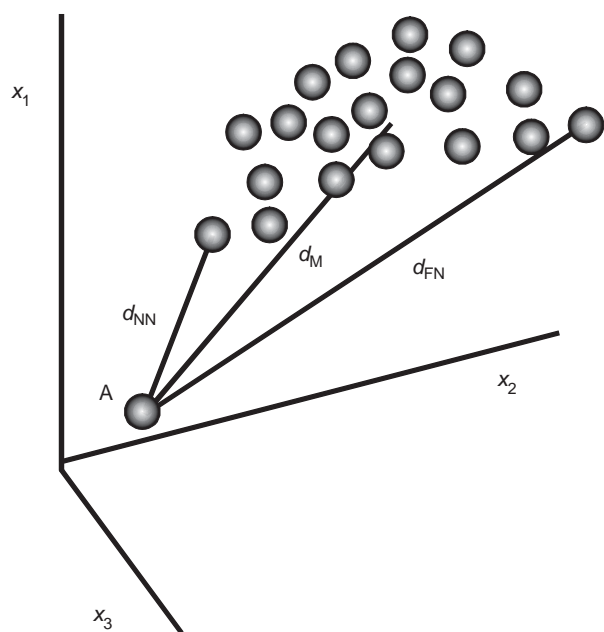
Hierarchical cluster analysis is noniterative and various implementation methods are commonly encountered in analytical science. Initially, each object in the dataset is considered as a separate cluster. At each subsequent stage of the algorithm the two clusters that are most similar are merged to create a new cluster. The algorithm terminates when all clusters are combined. The number of clusters in the dataset does not need to be known or provided *a priori*. Using the original data matrix, a suitable matrix of similarity measures between objects is first constructed (e.g., the Euclidean distance matrix). From this similarity matrix, the most similar objects are combined to produce a new 'grouped' object and the process repeated until all objects have been included in a single cluster. The choice of an appropriate similarity metric and the manner in which objects are grouped (or clustered) gives rise to many combinations of potential methods. Popular interpoint distances used for clustering are the nearest neighbor, furthest neighbor, and the mean (Figure 5).

The choice of measure may greatly influence the result of clustering. The nearest neighbor metric will link together quite distinct clusters if there exists a path of closely located points connecting the two clusters. The furthest neighbor algorithm does not suffer this problem but can be very sensitive to outliers and will not detect elongated clusters. The hierarchical structure of clusters provided by the algorithm is represented graphically by a dendrogram (Figure 6).

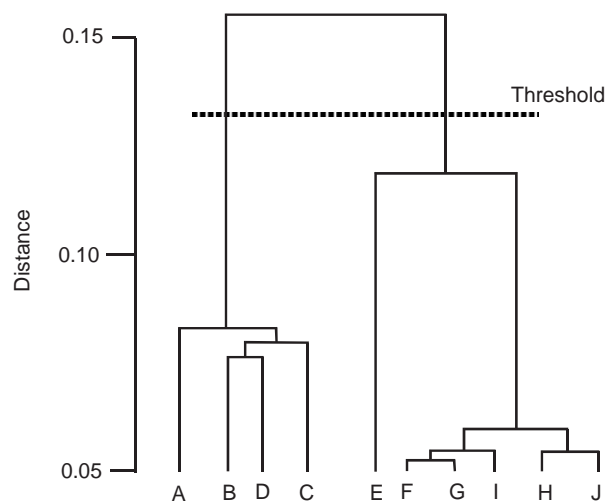
The dendrogram illustrates the cluster merging sequence and the corresponding values of the similarity measure employed. A threshold similarity value, selected by the user, splits the data into the perceived correct number of clusters. Generally, the threshold level should be chosen so that intercluster distances are considerably greater than the intracluster distances.

The application of unsupervised pattern recognition methods should be undertaken with caution. Many factors will influence the results of the analysis, including variable scaling, metric used, similarity measure, clustering criterion, number of data points, etc. Dynamic clustering methods can be computationally efficient, particularly with large data sets, but unrealistic groupings can easily be achieved and analysis with several different cluster representative functions is recommended. If the number of samples is small then describing some cluster representation function is meaningless and in such situations hierarchical methods are more useful.

Cluster analysis is not a statistical based operation and should not be employed to prove the existence of groups. Rather, cluster analysis is best employed as



**Figure 5** The distance between object A and the existing cluster can be defined by its proximity to the cluster's nearest neighbor,  $d_{NN}$ , its center,  $d_M$ , or its furthest neighbor,  $d_{FN}$ .



**Figure 6** A dendrogram provides a two-dimensional representation of the similarity between objects according to the distance between an object and a cluster. A threshold level, selected by the user, defines the number of distinct groups in the data.

part of the toolkit for exploratory data analysis. The evidence for certain groups and clusters, and the cause of structure found should be investigated by other techniques.

## Supervised Pattern Recognition

Whereas cluster analysis neither needs nor assumes *a priori* information about cluster properties or pattern design, supervised pattern recognition uses a training

set of known group members to define patterns and develop partition functions. The aim is to identify and quantify relationships between groups, and assign unclassified objects to one of the groups.

A wide range of modeling algorithms is available to perform this type of classification. Widely used techniques include the  $k$ -nearest neighbor ( $k$ -NN) method and linear discriminant analysis (LDA).

### Nearest Neighbor Classification

For successful operation the nearest neighbor rules for classification rely on knowing a large number of correctly, previously classified patterns. The basic principles are that samples that are close together in pattern space are likely to belong to the same class or have similar distributions of the classes. The first idea gives rise to formulation of the single nearest neighbor rule, 1-NN, and the second provides for extension of the rule to  $k$ -NNs.

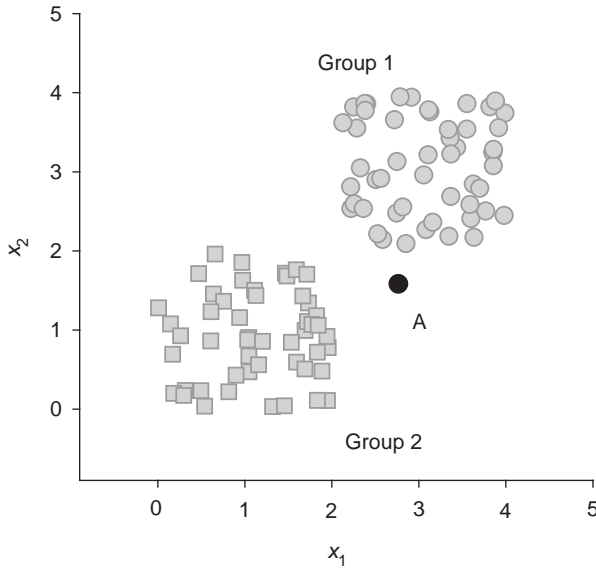
A test object to be classified is assigned to that class containing its nearest neighbor as measured by some distance metric. The Euclidean distance is easily calculated and is often used. This 1-NN scheme can readily be extended to more ( $k$ ) neighbors with the class assignment decided according to a majority vote procedure, i.e., assignment to the class most represented in the set of  $k$ -NN classified objects. Common values for  $k$  are 3 or 5 (Figure 7).

The sample-based decision boundary between groups is defined by the relatively small number of samples belonging to the outer envelopes of the clusters. Those samples deeply imbedded within clusters do not contribute to defining the boundary and may be discarded for classification of new objects. This concept can be exploited for increasing the computational efficiency of nearest neighbor classification since the distance from a new, unclassified object to every classified sample need not be calculated. A variety of so-called condensing algorithms is available and aims to provide a subset of the complete training dataset such that 1-NN classification of any new pattern with the subset is identical to 1-NN classification with the complete set.

Since nearest neighbor methods are based on similarity measured by some distance metric then variable scaling and the units used to characterize the data can influence results. Variables with the largest amount of scatter (greatest variance) will contribute most strongly to the Euclidean distance and in practice it may be advisable to standardize variables before performing classification analysis.

### Discriminant Function Analysis

In LDA, or discriminant function analysis, we are seeking to create new synthetic features (variables)



**Figure 7** With nearest neighbor classification a cluster is defined by the elements in the boundary layer and an object is classified as belonging to that group containing its nearest neighbor. Object A will be assigned to group 1 rather than group 2.

that are linear combinations of the original variables and that best indicate the differences between the known groups in contrast to the variable variances within the groups.

The process of performing LDA aims to derive and construct a boundary between the known classes of the training objects using statistical parameters. This boundary is derived using a discriminant function that provides a value or score when applied to a test object that indicates the group to which the new object should be assigned.

If  $f(x_i, k)$  is some measure of likelihood of object  $x_i$  belonging to group or class  $k$ , then the discriminant score  $D_i$ , for assigning  $x_i$  to one of two classes is given by

$$D_i = f(x_i, k_1) - f(x_i, k_2) \quad [2]$$

which may be interpreted as saying we classify test object  $x_i$  into class 1 if  $D_i$  is positive, otherwise  $x_i$  is considered as belonging to class 2.

The value of the discriminant score is calculated from a linear combination of the recorded values of the variables describing the objects, each suitably weighted to provide optimum discriminatory power. For two variables

$$D_i = w_0 + w_1x_{i,1} + w_2x_{i,2} \quad [3]$$

The weights, or variables coefficients used in calculating  $D$  (eqn [3]), are determined by

$$w = (\bar{x}_j(1) - \bar{x}_j(2))S^{-1} \quad [4]$$

$$w_0 = -\frac{1}{2}[(\bar{x}_j(1) - \bar{x}_j(2))S^{-1}(\bar{x}_j(1) + \bar{x}_j(2))]$$

This represents the ratio of the separation of the means of the two groups to the within-group variance for the groups as given by the pooled covariance matrix,  $S$ .

$\bar{x}_j(1)$  and  $\bar{x}_j(2)$  are the vectors of the mean values for variable  $j$  for groups (1) and (2), respectively, and easily obtained from

$$\bar{x}_j(k) = \frac{\sum_{i=1}^{n(k)} x_{i,j}(k)}{n(k)} \quad [5]$$

where  $n(k)$  is the number of objects in group ( $k$ ), and  $x_{i,j}(k)$  is the value for object  $i$  of variable  $j$  in group ( $k$ ).

The pooled covariance matrix,  $S$ , of two training classes is given by

$$S = \frac{S_{(1)} + S_{(2)}}{n(1) + n(2) - 2} \quad [6]$$

$S_{(k)}$  represents the covariance matrix of group ( $k$ ).

The vector of weights coefficients can be thus be calculated from the classified training data and the discriminant score computed for each new, unclassified sample.

The discriminant function is linear; all the terms are added together to give a single number, the discriminant score.

For higher-dimensional pattern space the boundary is a hyperplane of  $m - 1$  dimensionality, where  $m$  is the number of variables. The partition boundary between two classes is defined at  $D_i = 0$  and in the two-dimensional case it is given by

$$w_0 + w_1x_{i,1} + w_2x_{i,2} = 0 \quad [7]$$

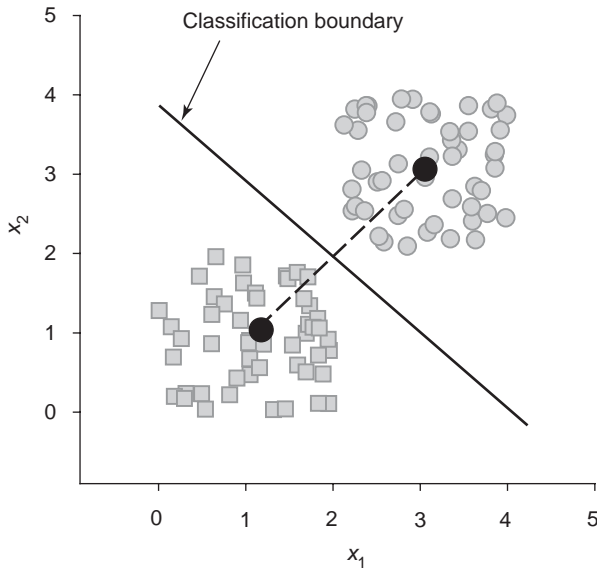
The classification boundary bisects a line between the centroids of the two clusters (Figure 8).

A useful result of performing LDA is the production of what is termed a discriminant plot, where every data point (from the training set or the new test sample set) is projected onto the discriminant function displayed as one-dimensional axis (Figure 9).

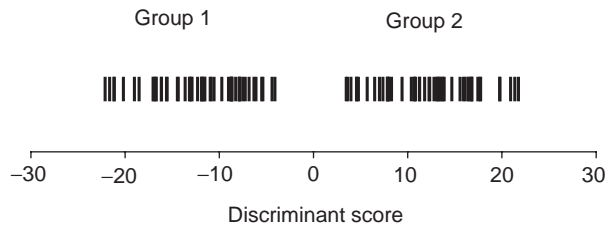
The concept of a linear discriminant axis reduces the multidimensional classification problem to a single dimension, with the projection achieved so that discrimination between classes is preserved as well as possible.

## Bayes Classification

The application of discriminant analysis can be extended to include probabilities of class membership and, assuming a multivariate normal distribution of data, confidence intervals for class boundaries can be calculated. The Bayes rule for classification simply states that 'an object should be assigned to that



**Figure 8** Linear discriminant analysis provides a linear partition boundary between the two known groups, bisecting the line between the centroids of the two groups.



**Figure 9** A discriminant plot projects the data on to a single axis (defined by the discriminant function).

group having the highest conditional probability'. If there are two possible classes, then a sample is assigned to group 1 if

$$P_{(1|x)} \geq P_{(2|x)} \quad [8]$$

where  $P_{(1|x)}$  is the conditional probability for group 1 given the pattern vector  $x$ . This conditional probability can be estimated from

$$P_{(1|x)} = \frac{P_{(x|1)} \cdot P_{(1)}}{P_{(x|1)} \cdot P_{(1)} + P_{(x|2)} \cdot P_{(2)}} \quad [9]$$

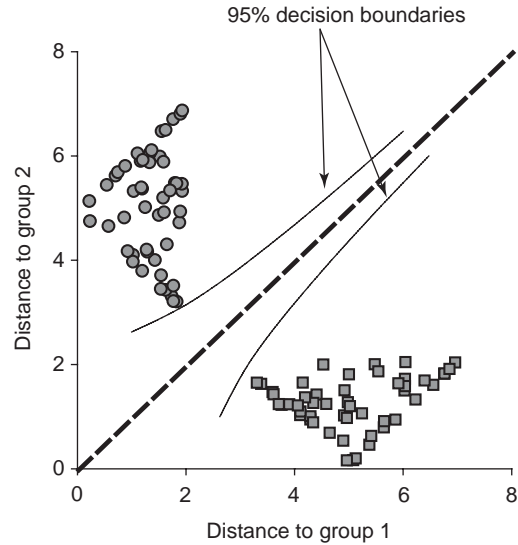
where  $P_{(1)}$  and  $P_{(2)}$  are the probabilities of the sample belonging to group 1 or group 2 in the absence of analytical data.

$P_{(x|1)}$  express the conditional probability of the pattern  $x$  arising from a member of group 1.

Thus, a sample is assigned to group 1 if

$$P_{(x|1)} \cdot P_{(1)} > P_{(x|2)} \cdot P_{(2)} \quad [10]$$

and the partition boundary where a sample has equal likelihood of belonging to group 1 or group 2, is



**Figure 10** A Coomans plot provides a visual representation of classification results and probability boundaries can also be displayed ( $P_1 = 0.95$  and  $P_2 = 0.95$  levels are illustrated here).

given by

$$P_{(x|1)} \cdot P_{(1)} = P_{(x|2)} \cdot P_{(2)} \quad [11]$$

If the data are assumed to come from a population having a multivariate normal distribution and it is furthermore assumed that the covariance matrix of each group is similar, then the conditional probability values can be calculated from the multidimensional normal distribution

$$P_{(x|k)} = \frac{1}{2\pi|S|^{1/2}} \exp[-0.5(x - \bar{x}_k)^T S^{-1}(x - \bar{x}_k)] \quad [12]$$

where  $S$  is the covariance matrix and  $\bar{x}_k$  the vector of variable means for group  $k$ .

The term  $(x - \bar{x}_k)^T S^{-1}(x - \bar{x}_k)$  is a squared metric (the Mahalanobis distance),  $dm_k^2$ , representing the distance of each object from a group center taking into account correlation within the data.

Using the vectors of means for each group, then values for  $dm_1$  and  $dm_2$  can be calculated for every sample. A plot of  $dm_1$  against  $dm_2$  is referred to as a Coomans plot and displays the results of classification (Figure 10).

On the same diagram probability boundaries (confidence levels) can also be displayed. By substitution into eqn [12], the partition boundary is given by

$$P_{(1)} \exp(-0.5 dm_1^2) = P_{(2)} \exp(-0.5 dm_2^2) \quad [13]$$

and since  $P_{(2)} = 1 - P_{(1)}$ , then this equation can be rearranged to

$$dm_1 = \sqrt{dm_2^2 + 2 \ln\left(\frac{P_{(1)}}{1 - P_{(1)}}\right)} \quad [14]$$

Selecting an appropriate value for  $P_{(1)}$  then  $dm_1$  values can be obtained for a range of  $dm_2$  values, and similarly  $dm_2$  values for a set of given  $dm_1$  figures.

Classification and discriminant analysis algorithms are available with all multivariate statistical software packages. New or modified procedures are regularly being introduced and the application of such techniques and methods in analytical science is growing.

See also: **Chemometrics and Statistics:** Statistical Techniques; Expert Systems; Multicriteria Decision Making. **Computer Modeling. Nuclear Magnetic Resonance Spectroscopy Techniques:** Multidimensional Proton.

## Further Reading

Adams MJ (2004) *Chemometrics in Analytical Spectroscopy*, 2nd edn. Cambridge: Royal Society of Chemistry.

Brereton RG (2003) *Chemometrics: Data Analysis for the Laboratory and the Chemical Plant*. Chichester: Wiley.

Coomans D and Massart DL (1992) Hard modelling in supervised pattern recognition. In: Brereton RG (ed.) *Multivariate Pattern Recognition in Chemometrics*, pp. 249–288. Amsterdam: Elsevier.

Devijver PA and Kittler J (1982) *Pattern Recognition: A Statistical Approach*. London: Prentice-Hall International.

Einax JW, Zwaninger HW, and Geiss S (1997) *Chemometrics in Environmental Data Analysis*. Heidelberg: VCH.

Everitt B (1980) *Cluster Analysis*, 2nd edn. London: Heinemann Educational.

Malinowski E (1991) *Factor Analysis in Chemistry*. New York: Wiley.

Nicholson JK, Connelly J, Lindon JC, and Holmes E (2002) Metabonomics: A platform for studying drug toxicity and gene function. *Nature Reviews (Drug Discovery Volume 1)*: 153–161.

Varmuza K (1980) *Pattern Recognition in Chemistry*. New York: Wiley.

## Multivariate Calibration Techniques

**G Hanrahan, F Udeh, and D G Patil**, California State University, Los Angeles, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Calibration is the process of measuring the instrument response ( $y$ ) of an analytical method to known concentrations of analytes ( $x$ ) using model building and validation procedures. These measurements, along with the predetermined analyte levels, encompass a calibration set. This set is then used to develop a mathematical model that relates the amount of sample to the measurements by the instrument. In some cases, the construction of the model is simple due to relationships such as Beer's Law in the application of ultraviolet spectroscopy.

Traditional univariate calibration techniques involve the use of a single instrumental measurement to determine a single analyte. In an ideal chemical measurement using high-precision instrumentation, an experimenter may obtain selective measurements linearly related to analyte concentration (**Figure 1A**). However, univariate techniques are very sensitive to the presence of outlier points in the data used to fit a

particular model under normal experimental conditions. Often, even only one or two outliers can seriously skew the results of a least squares analysis. The problems of selectivity and interferences (chemical and physical) also limit the effectiveness of univariate calibration methods causing some degree of nonlinearity (**Figure 1B**). In addition, such calibration techniques are not well suited to the multitude of data collected from the sensitive, high-throughput instrumentation currently being used in the analytical sciences. These datasets often contain large amounts of information, but in order to fully extract and correctly interpret this information, analytical methods incorporating a multivariate approach are needed.

In multivariate calibration, experimenters use many measured variables ( $x_1, x_2, \dots, x_k$ ) simultaneously for quantifying the target variable (a variable whose value is to be modeled and predicted by others (i.e., the variable on the left of the equal sign in linear regression)). In order to effectively use multivariate techniques, proper experimental design is essential. Experimental design allows the proper assessment of systematic variability (e.g., interferences) and helps in minimizing, for example, the effects of random noise. Experimental design is often limited by problems associated with the estimation of experimental



Selecting an appropriate value for  $P_{(1)}$  then  $dm_1$  values can be obtained for a range of  $dm_2$  values, and similarly  $dm_2$  values for a set of given  $dm_1$  figures.

Classification and discriminant analysis algorithms are available with all multivariate statistical software packages. New or modified procedures are regularly being introduced and the application of such techniques and methods in analytical science is growing.

See also: **Chemometrics and Statistics:** Statistical Techniques; Expert Systems; Multicriteria Decision Making. **Computer Modeling. Nuclear Magnetic Resonance Spectroscopy Techniques:** Multidimensional Proton.

## Further Reading

Adams MJ (2004) *Chemometrics in Analytical Spectroscopy*, 2nd edn. Cambridge: Royal Society of Chemistry.

Brereton RG (2003) *Chemometrics: Data Analysis for the Laboratory and the Chemical Plant*. Chichester: Wiley.

Coomans D and Massart DL (1992) Hard modelling in supervised pattern recognition. In: Brereton RG (ed.) *Multivariate Pattern Recognition in Chemometrics*, pp. 249–288. Amsterdam: Elsevier.

Devijver PA and Kittler J (1982) *Pattern Recognition: A Statistical Approach*. London: Prentice-Hall International.

Einax JW, Zwaninger HW, and Geiss S (1997) *Chemometrics in Environmental Data Analysis*. Heidelberg: VCH.

Everitt B (1980) *Cluster Analysis*, 2nd edn. London: Heinemann Educational.

Malinowski E (1991) *Factor Analysis in Chemistry*. New York: Wiley.

Nicholson JK, Connelly J, Lindon JC, and Holmes E (2002) Metabonomics: A platform for studying drug toxicity and gene function. *Nature Reviews (Drug Discovery Volume 1)*: 153–161.

Varmuza K (1980) *Pattern Recognition in Chemistry*. New York: Wiley.

## Multivariate Calibration Techniques

**G Hanrahan, F Udeh, and D G Patil**, California State University, Los Angeles, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

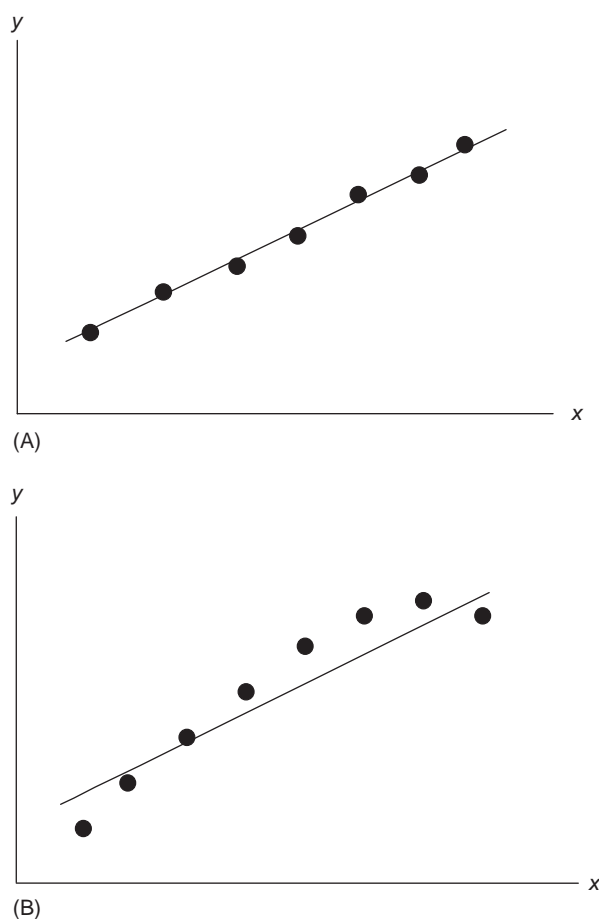
## Introduction

Calibration is the process of measuring the instrument response ( $y$ ) of an analytical method to known concentrations of analytes ( $x$ ) using model building and validation procedures. These measurements, along with the predetermined analyte levels, encompass a calibration set. This set is then used to develop a mathematical model that relates the amount of sample to the measurements by the instrument. In some cases, the construction of the model is simple due to relationships such as Beer's Law in the application of ultraviolet spectroscopy.

Traditional univariate calibration techniques involve the use of a single instrumental measurement to determine a single analyte. In an ideal chemical measurement using high-precision instrumentation, an experimenter may obtain selective measurements linearly related to analyte concentration (**Figure 1A**). However, univariate techniques are very sensitive to the presence of outlier points in the data used to fit a

particular model under normal experimental conditions. Often, even only one or two outliers can seriously skew the results of a least squares analysis. The problems of selectivity and interferences (chemical and physical) also limit the effectiveness of univariate calibration methods causing some degree of nonlinearity (**Figure 1B**). In addition, such calibration techniques are not well suited to the multitude of data collected from the sensitive, high-throughput instrumentation currently being used in the analytical sciences. These datasets often contain large amounts of information, but in order to fully extract and correctly interpret this information, analytical methods incorporating a multivariate approach are needed.

In multivariate calibration, experimenters use many measured variables ( $x_1, x_2, \dots, x_k$ ) simultaneously for quantifying the target variable (a variable whose value is to be modeled and predicted by others (i.e., the variable on the left of the equal sign in linear regression)). In order to effectively use multivariate techniques, proper experimental design is essential. Experimental design allows the proper assessment of systematic variability (e.g., interferences) and helps in minimizing, for example, the effects of random noise. Experimental design is often limited by problems associated with the estimation of experimental



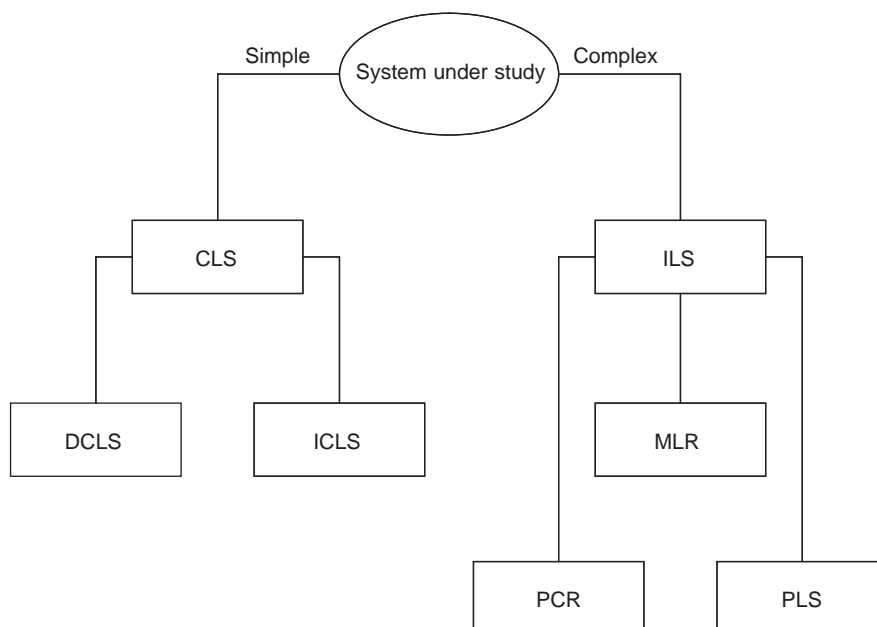
**Figure 1** (A) An ideal chemical measurement where ( $y$ ) is linearly related to ( $x$ ). (B) How selectivity and interference problems limit the effectiveness of univariate calibration methods causing nonlinearity.

parameters, and in some cases the data do not contain sufficient information to obtain proper estimates of the parameters. This can be especially true in latent variable models, where extremely large training sets are often needed to provide enough information for efficient parameter estimation. It is then paramount to search for optimum experimental designs to maximize the information. Forming a decision tree (Figure 2) can be a useful design tool in determining which multivariate calibration technique to use in a given problem.

Both classical least squares (CLS) and inverse least squares (ILS) approaches should be considered. The main advantages and disadvantages of both approaches are listed in Table 1, with detailed descriptions of the various techniques within both approaches described below.

### Data Preprocessing

Data preprocessing is a very important step in many chemometric techniques, which can be used separately (before a method is applied), or as a self-adjusting procedure that forms part of the chemometric methods. Ideally, data preprocessing can be used to remove known interference(s) from data to improve selectivity and enhance more important information to improve robustness. Techniques such as principal components (PCs; see 'Principal components regression' section below) are scale dependent. If one variate has a much higher variance than the others, it is necessary to scale the original variates before

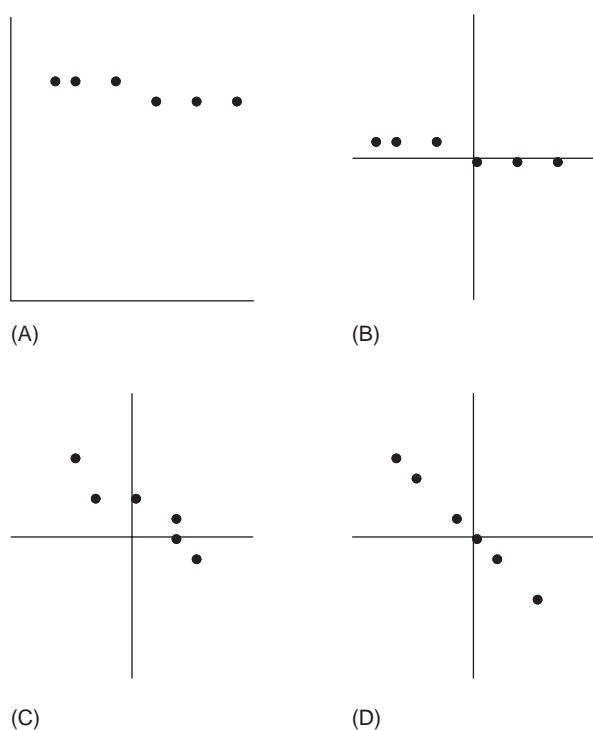


**Figure 2** Multivariate calibration decision tree.



**Table 1** Main advantages and disadvantages of CLS and ILS multivariate calibration approaches

Approach	Advantages	Disadvantages
CLS	Used in estimating multivariate limits of detection, often based directly on Beer's Law Used for moderately complex mixtures  Wavelength selection is not necessarily required for calibration Averaging effects make it less susceptible to noise	Not useful for mixtures with components that interact Requires knowledge of all components in calibration mixture Susceptible to baseline effects
ILS	Used in estimating multivariate limits of detection, often based directly on Beer's law Allows calibration of very complex mixtures  Calculations are relatively fast	Interferences must be included in model Wavelength selection can be difficult and time consuming Accurate calibration often requires large numbers of samples Number of wavelengths used in the model limited by the number of calibration samples

**Figure 3** Scaling technique effects: (A) original data, (B) column centered, (C) autoscaled, and (D) autoscaled profiles.

calculating PCs. Common preprocessing scaling techniques include (but not limited to) column centering, autoscaling, column standardization, and autoscaled profiles (Figure 3). Other data preprocessing techniques used in chemometrics include variance normalization, baseline corrections (first and second derivative, subtraction), and path length corrections (MSC, SNV, thickness).

### Classical Least Squares

The CLS (also known as K-Matrix) approach is best applied to systems where the concentration of every

analyte in the sample is known and obeys a linear relationship with measurement vectors. For a single wavelength and a single analyte, this relationship (Beer's Law) can be explained mathematically by the following equation:

$$A_{\lambda} = \varepsilon_{\lambda} b c \quad [1]$$

where  $A_{\lambda}$  is the absorbance at wavelength  $\lambda$ ,  $\varepsilon_{\lambda}$  the molar absorption coefficient at wavelength  $\lambda$  ( $\text{l mol}^{-1} \text{cm}^{-1}$ ),  $b$  the cell path length (cm), and  $c$  the concentration of the analyte ( $\text{mol l}^{-1}$ ). The absorbances can be normalized to a constant cell path length and results in the simplified relationship:

$$A_{\lambda} = k_{\lambda} c \quad [2]$$

where  $k_{\lambda}$  is the single constant combination of the absorptivity coefficient and path length. This equation can be easily solved by measuring the absorbance of a single sample of known concentration and using these values to solve for  $k_{\lambda}$ . Multivariate systems (e.g., analytes A and B) are more complex and the mixture of analytes is equal to the sum of the instrument responses to the component responses (Figure 4).

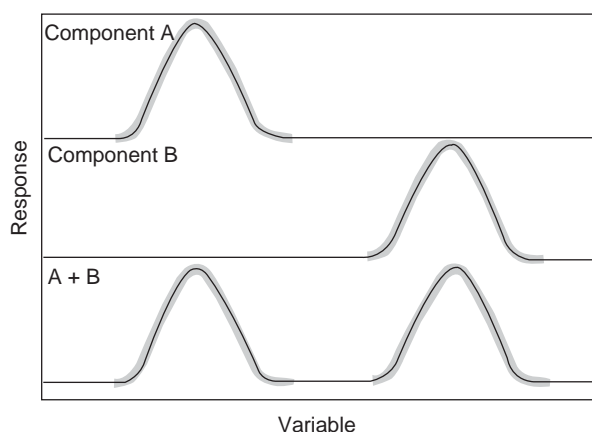
The final response of the instrument to this mixture can be represented as

$$A_{\lambda} = K c \quad [3]$$

As shown in the decision tree (Figure 1), two CLS methods are used, direct (DCLS) and indirect (ICLS). In the DCLS method, pure spectra are measured directly. In the ICLS method, pure spectra are estimated using mixtures.

### Inverse Least Squares

In the ILS approach, the dependent variable (concentration) is solved by calculating a solution from multiple independent variables (responses at



**Figure 4** Mixture of analytes in multivariate systems.

the selected wavelengths). This is different from CLS where absorbance at a single wavelength is calculated as an additive function of the component concentrations. In ILS, we can combine the absorptivity coefficient ( $\varepsilon_i$ ) and cell path length ( $b$ ) from Beer's Law to form a single constant relationship (matrix notation) with concentration:

$$c = PA_i \quad [4]$$

where  $P$  is the matrix of coefficients. The ILS approach is used in complex mixtures where it is not necessary to obtain the spectra of pure analytes present. The three inverse techniques of multiple linear regression (MLR), principal components regression (PCR), and partial least squares (PLS) are discussed in detail below.

### Multiple Linear Regression

MLR is a method used to estimate the size and statistical significance of the relationship between a dependent variable ( $y$ ) and one independent or predictor variable, ( $x_1$ ), after adjustment for confounders ( $x_2, \dots$ ). As discussed earlier, models constructed from spectroscopy are relatively simple due to linear combinations of the instrumental measurements. Models for a broader range of conditions (i.e., measurements from several wavelengths) have been constructed in order to separate overlapping peaks elicited from the analyte plus other unknown components or conditions. These multiple linear methods for separating outliers are based upon the following equation:

$$x_i = b_0 + b_1y_{i1} + b_2y_{i2} + \dots + b_qy_{iq} + e_i \quad [5]$$

where  $x_i$  is the analyte level of the  $i$ th specimen;  $y_{ij}$  the  $j$ th instrumental measurement with the  $i$ th

specimen;  $b$  the model parameters, and  $e_i$  the error associated with  $x_i$ . Simply, it finds a set of partial regression coefficients such that  $y$  could be approximated as well as possible by a linear combination of the independent variables. The value of the partial coefficients can be found using ordinary least squares (OLR) and the MLR equation can be expressed conveniently in matrix notation. MLR is useful in situations where the number of variables is small, not significantly collinear, and has a strong relationship to the response of the system. When one or more of these conditions is sacrificed, MLR can be inefficient. When the number of variables is reduced, for example, poorer error detection and less precise estimates can occur.

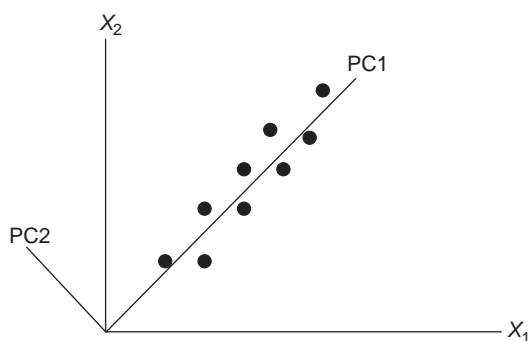
### Principal Components Regression

PCR is a two-step multivariate calibration method involving compression of the data ( $x$ -) matrix into latent variables by principal components analysis (PCA), followed by MLR. PCA (also known as Karhunen–Loève expansion or Eigen- $xy$  analysis) mathematically transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called eigenvectors (or PCs). Essentially, PCA is the breakdown of the original data matrix ( $X$ ) to a product of scores matrix ( $T$ ) and a loadings matrix ( $L$ ). The loading matrix describes the direction of the PC. These relationships can be represented by the equation:

$$X = TL^T$$

where the superscript  $T$  is the transpose of a matrix. The PCs of PCA decomposition represent the spectral variations that are common to all of the spectroscopic calibration data. Therefore, using that information to calculate a regression equation will produce a robust model for predicting concentrations of the desired constituents in very complex samples. The latent variable which best describes the relative distances between objects is given by the direction of maximum variance. **Figure 5** illustrates a two PC situation for two variables,  $x_1$  and  $x_2$ . The first PC accounts for as much of the variability in the data as possible and termed the first PC (PC1). The second PC (PC2) accounts for as much of the remaining variability as possible and is orthogonal (at right angles) to PC1.

Further PCs can be determined by the continuation of this process. However, most of the variation is often found with the combination of PC1 and PC2, thus resulting in a two-dimensional representation of



**Figure 5** PCA with two variables,  $x_1$  and  $x_2$ .

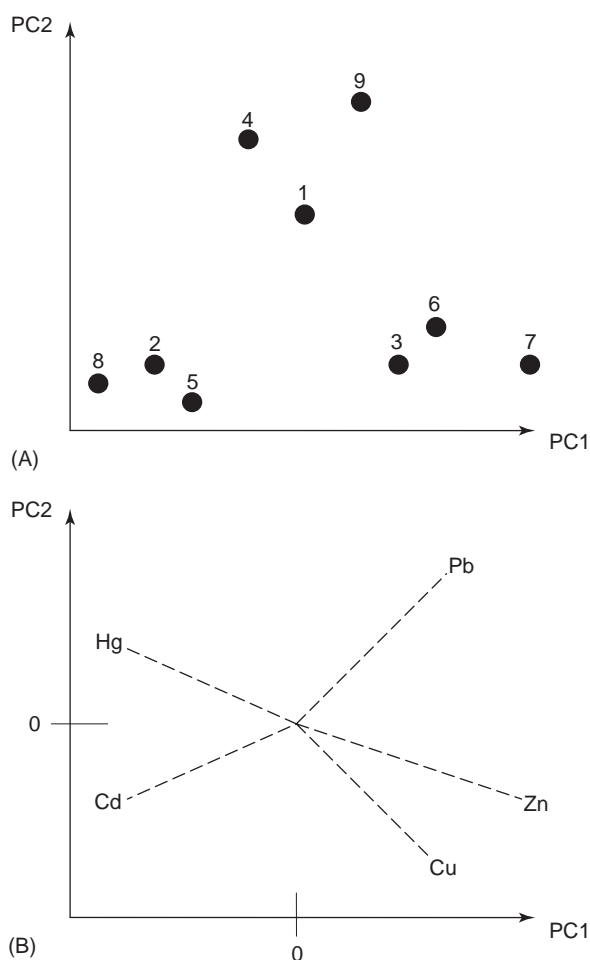
**Table 2** PCA applied to the determination of heavy metals in various sites of a coastal waterway

Site	Cd ( $\mu\text{g l}^{-1}$ )	Hg ( $\mu\text{g l}^{-1}$ )	Zn ( $\mu\text{g l}^{-1}$ )	Cu ( $\mu\text{g l}^{-1}$ )	Pb ( $\mu\text{g l}^{-1}$ )
1	2.5	0.28	42.1	2.9	9.9
2	8.6	0.19	69.8	7.5	12.0
3	3.0	0.26	84.0	4.6	8.9
4	3.9	0.25	64.6	7.9	9.8
5	10.2	0.20	75.8	10.2	11.5
6	4.0	0.19	89.2	13.0	7.6
7	5.3	0.29	99.7	11.9	13.3
8	3.9	0.13	48.9	3.9	8.2
9	4.2	0.18	50.6	5.2	20.3

the data. The main issue in employing PCA for data modeling is that of choosing the appropriate number of retained components. When estimating the number of PCs (latent variables) to use, criteria such as the percentage of explained variance, Scree-test, and cross validation are considered. Overall, PCA works extremely well for the detection of experimental patterns, trends, and groups among multivariate datasets as well as for outlier detection. Let us consider an example of PCA applied to the determination of heavy metals in various locations along a coastal waterway. Table 2 lists the nine sites analyzed and the elements detected.

Figures 6A and 6B show the data analyzed using a simple PCA approach with two PCs (PC1 and PC2). Figure 6A shows the relevant score matrix of the site locations. The score matrix tells the experimenter where the points lie along the new coordinate axis. Figure 6B depicts the loadings and how much the original variables contribute to the principle component.

As mentioned, PCR is a two-step process; the PC scores and the PC loadings are calculated with the PC scores regressed against the analyte concentrations using a regression method. In PCR, the PC scores are chosen to describe as much of the variation in the predictors as possible. The total variance of the experimental dataset and the sum of the eigenvalues



**Figure 6** (A) PCA score matrix of heavy metal determinations from various coastal waterway locations. (B) PCA loadings and how much the original variables contribute to the principle component.

both equal the number of variables. Note that the PCA eigenvectors and scores can be calculated independently of any knowledge of these concentrations and merely represent the largest common variations among all the spectra in the dataset. Most likely, these variations will be related to changes in the analyte concentrations. In applications with collinear  $x$ -variables, PC scores with smaller eigenvalues will likely be more influenced by  $x$ -noise than PC scores with larger eigenvalues. Such conditions exist, for example, in near-infrared analysis where spectral data are highly collinear. The PCR method represents a considerable improvement over MLR and CLS. By using latent variables (scores), it is possible to use a large number of variables (frequencies), but without having knowledge of all possible interferences.

### Partial Least Squares Regression

PLS was originally developed in the 1960s as an ecometric method, but is now employed by many

chemometricians. There are actually two versions of the PLS algorithm (PLS-1 and PLS-2). In an experimental context, PLS can be presented as a kind of simultaneous PCA and regression. The practical implication of this approach is that the experimenter may create new explanatory variables while carrying the same information as the original explanatory variables. Like PCR, PLS is a spectral decomposition technique highly used in multivariate calibration. However, PLS is a one-step process with no regression step involved. Instead, PLS performs the decomposition on both the spectral and concentration data simultaneously.

In an original PCR analysis, for example, only the spectra are taken into account for the determination of the PC scores. For PLS the analyte concentrations of the calibration samples are also incorporated. The factors are presented in such a way that the variation of the content substances can be clarified, as well as possible, by the PCs which are to be determined. Calibration by means of PLS therefore requires, under normal circumstances, fewer factors than a PCR calibration. However, PLS does have its limitations. For example, this approach is often times not appropriate for screening factors that have a minimum effect on the response. Calculations in PLS are generally slower than most classical methods, especially PLS-1. PLS-2 calibrates for all constituents simultaneously, thus allowing for possible sacrificing of accuracy in the predictions of the constituent concentrations. This is especially true for mixtures. In addition, a large number of samples are normally required for accurate calibration.

**See also:** **Chemometrics and Statistics:** Experimental Design; Optimization Strategies.

## Further Reading

- Anderson TW (2003) *An Introduction to Multivariate Statistical Analysis*, 3rd edn. New York: Wiley-Interscience.
- Beebe KR, Pell RJ, and Seasholtz MB (1998) *Chemometrics: A Practical Guide*. New York: Wiley.
- Brereton RG (2003) *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. Chichester: Wiley.
- Brown PJ (1993) *Measurement, Regression and Calibration*. Oxford: Clarendon Press.
- Frank IE and Todeschni R (1994) *The Data Analysis Handbook*. Amsterdam: Elsevier.
- Henrion R and Henrion G (1994) *Multivariate Data Analysis*. Berlin: Springer.
- Kramer R (1998) *Chemometric Techniques for Quantitative Analysis*. New York: Decker.
- Malinowski ER (1991) *Factor Analysis in Chemistry*. New York: Wiley.

- Martens H and Martens M (2001) *Multivariate Analysis of Quality*. Chichester: Wiley.
- Martens H and Naes T (1989) *Multivariate Calibration*. New York: Wiley.
- McLachlan GJ (1992) *Discriminant Analysis and Statistical Pattern Recognition*. Chichester: Wiley.
- Miller JN and Miller JC (2000) *Statistics and Chemometrics for Analytical Chemistry*, 4th edn. New York: Prentice-Hall.
- Naes T, Isaksson T, Fearn T, and Davies T (2002) *A User Friendly Guide to Multivariate Calibration and Classification*. Chichester: NIR Publications.
- Otto M (1999) *Chemometrics: Statistics and Computer Application in Analytical Chemistry*. Chichester: Wiley-VCH.
- Vandeginste BGM, Massart DL, Buydens, LMC *et al.* (1998) *Handbook of Chemometrics and Qualimetrics: Part B*. Amsterdam: Elsevier.

## Glossary

- |                        |   |
|------------------------|---|
| Collinearity           | Approximate linear dependence among experimental variables.   |
| Cross-validation       | Concept where every object of the X-matrix is removed from the dataset once, and a model is computed with the data that remains.                            |
| Eigenvalue             | A measure of the magnitude of variation of a derived variable within a multivariate analysis technique.   |
| Latent variable        | Unobserved variables which do not necessarily have to be orthogonal.  |
| Multivariate analysis  | Statistical, mathematical, or graphical techniques that consider multiple variables simultaneously.   |
| Outlier                | A sample that does not follow the same model as the rest of the experimental dataset.   |
| Principle component    | A set of variables that encompasses the maximum amount of variation in a dataset and is orthogonal to the previous principle component of the same dataset. |
| Scree-test             | Based on the idea that residual variance levels off when the proper number of principal components is obtained.   |
| Univariate calibration | Construction of a relation between variables $x$ and $y$ , such that $x$ can be used to predict $y$ .   |
| Variance scaling       | Variance scaling is used to emphasize small variations in the data by giving all values equal weighting.  |

## Expert Systems

M J Adams, RMIT University, Melbourne, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Expert systems constitute a branch of artificial intelligence and the name is often considered as synonymous with the term knowledge-based systems. They are computer programs; software products that attempt to replicate the knowledge and skills of a human expert in a specific domain in order to solve problems.

Early computerized intelligent systems, belonging to the so-called first generation of intelligent programs, concentrated on attempts to code general problem-solving strategies. Knowledge about any specific problem being solved was integrated, often buried, within the program code. Results obtained required tedious analysis and interpretation and generally little progress was made in applying such systems to real-world problems.

Second-generation programs represented a significant advance in user friendliness. This was largely achieved by separating the procedural, reasoning part of the code from a database containing data and information relevant to the specific problem. Such programs are typified by DENDRAL (used to infer plausible molecular structures from mass spectra and nuclear magnetic resonance data) and are interactive and flexible in terms of user communication.

Subsequent development and progress in expert system technology built on this scheme of keeping data/information and processing functions separate. Other improvements included enhanced flexible reasoning models and greatly improved interactivity and user communications with regard to the programs explaining their line of reasoning and being capable of interrogation about processes undertaken.

Once it was realized and accepted that expert systems worked best on problems limited to a specific domain, then a considerable number of applications of such systems in analytical science rapidly developed. This is particularly the case in automating the interpretation of spectral data. The apparently obvious correlation or correspondence between a spectral pattern and its parent molecular structure has proved to be an attractive and fruitful field for researchers developing expert system technologies and applications. Initially mass spectrometry, and later infrared (IR),  $^1\text{H}$ , and  $^{13}\text{C}$  nuclear magnetic

resonance (NMR) spectroscopies have all been domains in which expert systems have been developed. The scientific literature contains many hundreds of papers detailing their development and application. In addition to interpreting single-source data, many attempts have been reported that seek to combine sample spectral data from multiple sources. The early work on expert systems in analytical science led to much debate on the usefulness of expert systems in chemistry and today most reported applications are more modest in scope and tend to be focused on a specific application.

### Structure of Knowledge Base Systems

A general characteristic of expert systems is the separation of the domain knowledge, i.e., knowledge specific to a problem domain (e.g., IR spectra or proton NMR spectra), from the program's operational processes that are contained in an inference engine (Figure 1).

#### The Knowledge Base

The knowledge base holds facts, which may comprise short-term and long-term information, and rules, which can be considered as long-term information, on how to generate new facts. Within the knowledge base information can be represented in a number of ways, including predicate calculus, frames, semantic networks, and production rules.

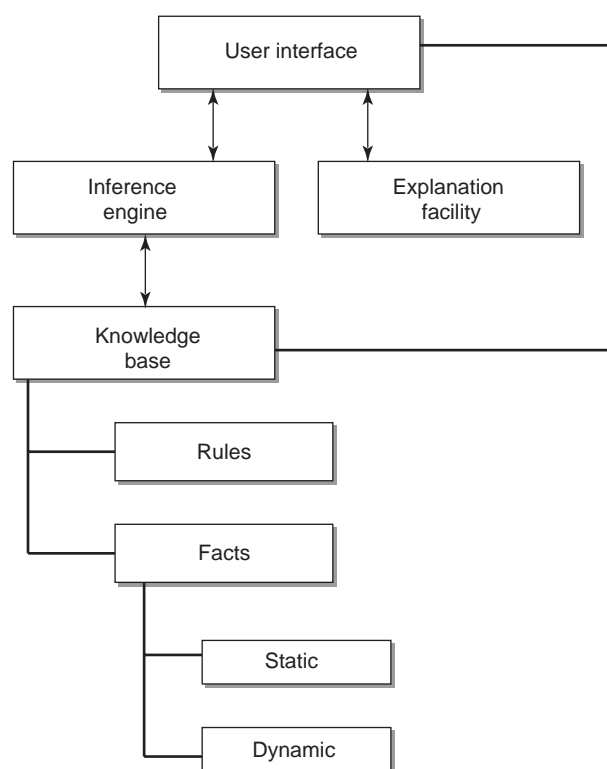
Predicate calculus is derived from propositional logic, in which propositions or statements can result in one of two values, e.g., true or false. Several logical operators exist to enable propositions to be developed. Some of the most important are:

- logical AND ( $\wedge$ ),
- logical OR ( $\vee$ ),
- negation NOT ( $\sim$ ),
- implication ( $\rightarrow$ ), and
- equivalence ( $\leftrightarrow$ ).

Statements can be developed using operators such as 'for all' ( $\forall$ ) and there exists ( $\exists$ ) and combining such operators can provide a modular, precise, and abstract representation of interpretation strategies. For example, the interpretation of a X-ray fluorescence (XRF) spectrum could include the statement,

$$(\forall x, y) \text{PeakAt}(x, 57.5) \wedge \text{PeakAt}(y, 51.7) \\ \rightarrow (\exists z) \text{Iron}(z) \wedge \text{Present}(z) \quad [1]$$



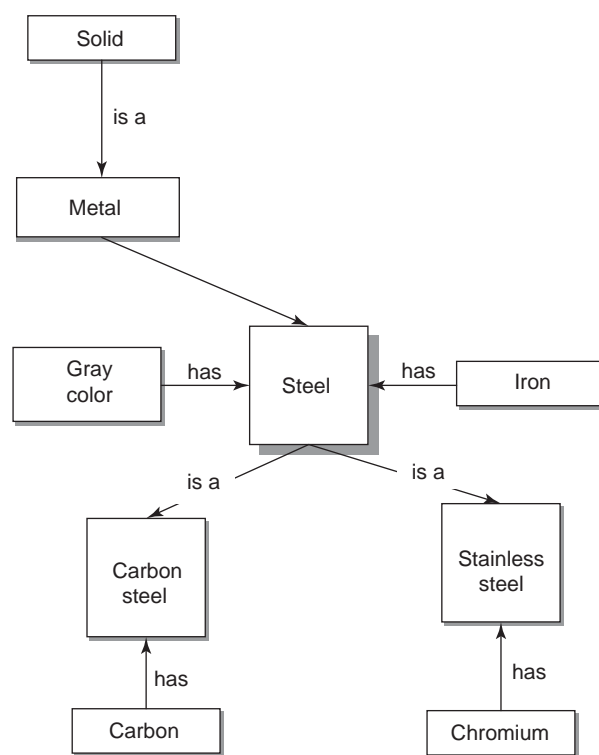


**Figure 1** The modular structure of an expert system separates the reasoning process (the inference engine) from the specific domain information (the knowledge base) comprising both rules and data. The explanation facility provides a means of tracing the reasoning scheme during or after an analysis.

This can be interpreted as stating that a pair of peaks occurring at angles of  $57.5^\circ$  and  $51.7^\circ$  in the dispersion spectrum implies iron is present in the sample giving rise to the spectrum.

Semantic networks and frames are similar structures for storing information. Networks use nodes, generally containing a single piece of information, linked to other nodes by relationship arcs. Semantic networks have a less formal and rigid structure than frames that are used to describe hierarchical data. A frame may contain several slots, each holding data or initiating an action if the value in the slot changes. An important feature and characteristic of these structures is the property of inheritance; objects or frames automatically assume the characteristics associated with their parent frame in addition to the specific information contained in their own slots (Figures 2 and 3).

Production rules, also known as 'if-then' pairs or condition-action pairs, form the basis of rule-based expert systems. They resemble implication statements in predicate calculus and can not only generate new facts, but can also run other programs, activate or deactivate other rule-sets, and even change the



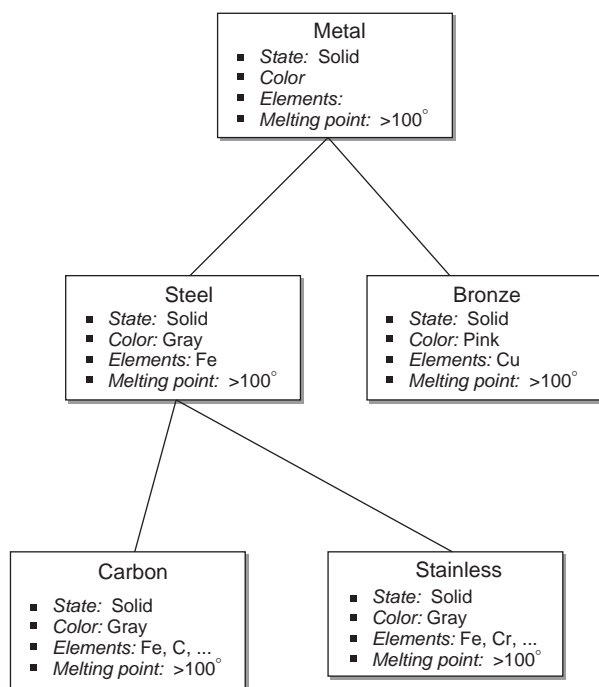
**Figure 2** A semantic network is a collection of objects and relationships and is used to represent knowledge. Whilst more flexible than frame structures (Figure 3), their poor formal representation can complicate the operation of the inference engine.

context of inference. Rules are usually independent of each other, making construction and editing of new rules simple, and their use imposes a uniform structure on the knowledge base. They are often a natural way of expressing certain knowledge, but in a complex system they can be difficult and confusing to trace (Figure 4).

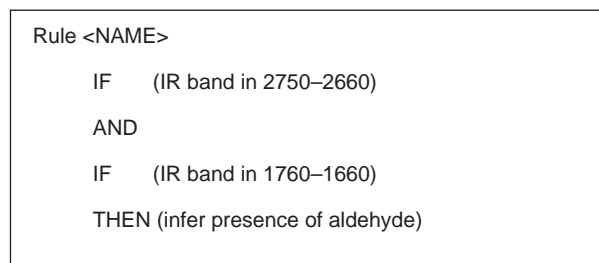
### The Inference Engine

The data and information in the knowledge base are processed by the inference engine, the part of the expert system containing the general problem-solving strategy. The inference engine uses production rules and facts held in the knowledge base to infer new knowledge. The scheme can be backward driven, forward driven, or a combination of both.

Forward- or data-driven inferencing starts with known facts and using rules derives new facts that are themselves stored in the knowledge base. The inference engine iterates this operation until no new facts can be inferred or some specified goal is achieved. Forward chaining will infer everything that can be inferred from the data provided, and this process can result in unfocused questioning



**Figure 3** Frame-based representation of knowledge provides a hierarchical format. Like semantic networks the nodes, or frames, can contain multiple pieces of information in slots, and are linked according to defined relationships between two or more objects. An important property of the frame structure is the ability for frames to inherit slot values. Thus, *Steel* inherits its *state* value from the *Metal* frame.



**Figure 4** Production rules are the most common formal form of knowledge representation. Rules are easily understood with each rule representing a single piece of knowledge, and they can be easily added to, removed from, or modified in a rule base.

when further information is requested from a user (Figure 5).

Backward-driven inferencing, or goal-driven inferencing as it is sometimes called, involves starting with some goal or specific fact (e.g., 'is lead present in the sample' or 'is a carbonyl functional group indicated by an IR spectrum') and determines which rules will satisfy this goal. This backward-chaining process is repeated until all necessary subgoals are resolved and a value for the primary goal derived.

Backward-driven inferencing is very efficient in determining a specific answer, but not as good or effective at general reasoning (Figure 6).

Since both inference strategies have advantages and disadvantages, most successful systems, and commercially available shells, generally incorporate both schemes.

## User Interface

In addition to the knowledge base and inference engine, an expert system should include an explanatory interface. Its purpose, as its name suggests, is to provide the developer or user with an account of the reasoning strategy employed. Important reasons for a system having an efficient and effective explanation facility are

- to demonstrate to the user that the results presented are reasonable;
- to instruct or train new users about the knowledge in the system; and
- to aid development and debugging.

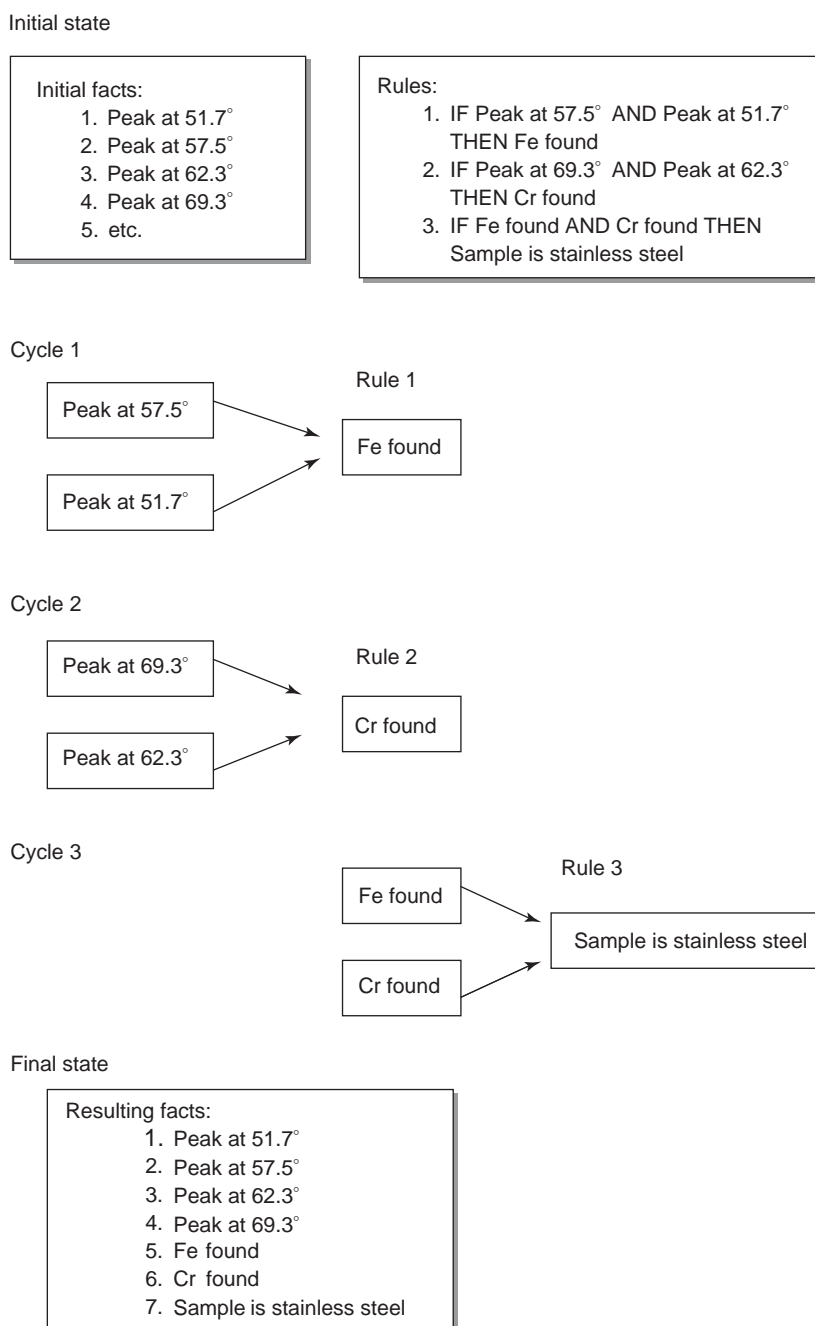
It is generally accepted that most current expert systems described in the literature or available commercially are relatively primitive in their ability to explain the reasoning involved in reaching a conclusion.

The interface with the user is an important and significant part of an expert system. An expert system's task is not necessarily complete when a question is answered or a problem solved; the system must be capable of explaining its reasoning and display how conclusions are attained. The importance of this explanation facility should not be underestimated. A user may be a novice in the domain in which the system operates, and it can have, and manipulate, knowledge outside of the competence of the user. For a user to trust and confidently act on the decisions and results proposed by an expert system its operations and strategies must be transparent.

## Expert System Implementation

Developing and implementing an expert system requires the use of a computer language or development tool. The programming languages used for expert system applications are generally either problem-orientated languages such as BASIC, FORTRAN, PASCAL, or C, or symbol-manipulation languages such as LISP or PROLOG. Although symbolic programming languages were developed specifically for applications in artificial intelligence, the





**Figure 5** The forward-chaining scheme initiates the search for a goal by sequentially firing all rules until either a goal is found or no rule fires. In this much-simplified example, the system is presented with a peak list from an XRF spectrum and using simple production rules can infer that the sample is a stainless steel.

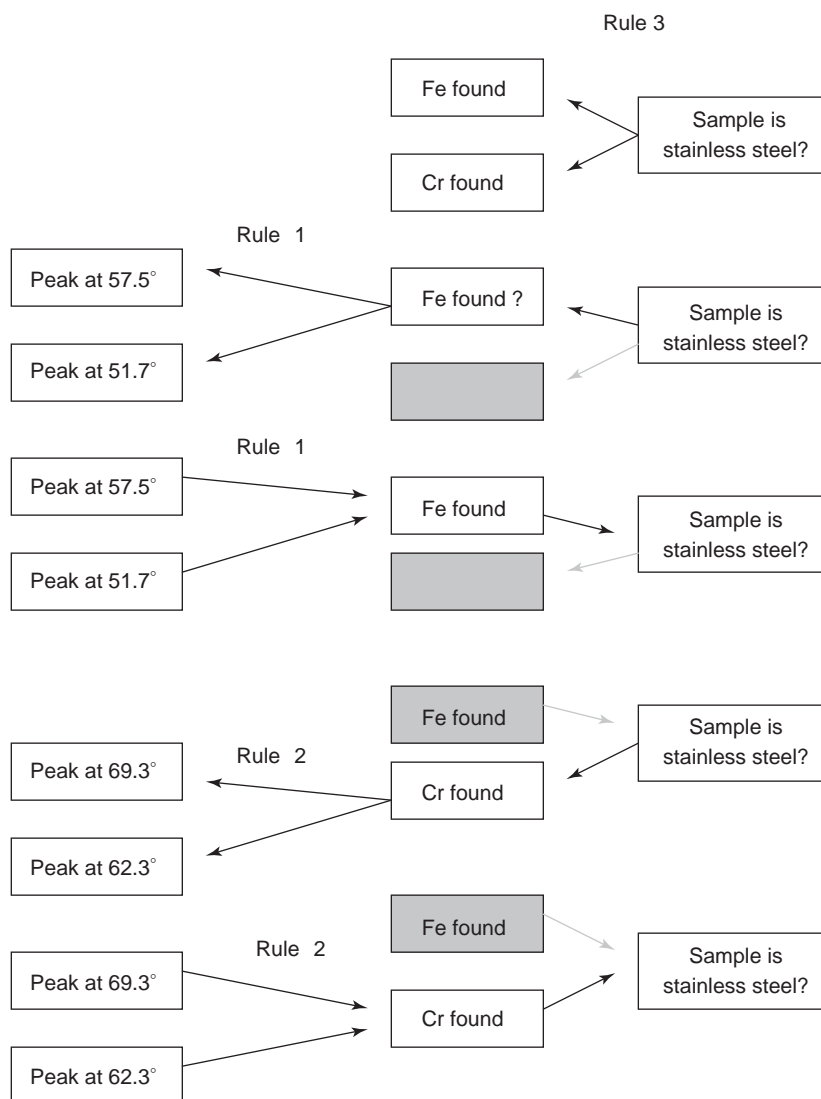
vast majority of expert systems are implemented in the more traditional problem-orientated languages.

It is not uncommon for developers of expert systems to use programming tools, knowledge engineering languages referred to as shells, specifically designed for this purpose. Shells include a number of features common to all expert systems, including an inference engine, a predefined rule language, a user interface, and a development interface. A number of

such shells are available commercially and their use can considerably aid the development and application of expert systems.

## Uncertainty in Expert Systems

Many applications of expert systems in analytical science have to cope with and manage situations



**Figure 6** Backward chaining starts with the goals and initiates the firing of rules according to the question asked. It is a commonly employed technique where the volume of data is large. Here the question 'is the sample a stainless steel' is answered by seeking confirmation for the presence of both iron and chromium in the XRF peak list.

involving uncertainty that may arise from imperfect domain knowledge or imperfect data. Although it is generally agreed and accepted that uncertainty is an important aspect of expert system reasoning strategies, there is little general agreement as to how it is best implemented. Some common and well-documented methods include probability theory, certainty factors, and fuzzy logic.

Conditional probability is based on Bayes' rule that may be simply represented as

$$P_{(H|E)} = \frac{P_{(E|H)} \cdot P_{(H)}}{P_{(E)}} \quad [2]$$

$P_{(H|E)}$  is read as the probability of hypothesis  $H$  given that evidence  $E$  exists.  $P_{(E|H)}$  is the probability

of evidence  $E$  being observed given hypothesis  $H$ ,  $P_{(H)}$  is the probability of  $H$  in the absence of evidence  $E$ , and  $P_{(E)}$  is the probability associated with evidence  $E$ . Although eqn [2] is adequate for a single hypothesis, when a number of competing hypotheses exist each has to be considered in the calculation of  $P_{(H|E)}$ . A more generalized form of Bayes' rule is

$$P_{(H_i|E)} = \frac{P_{(E|H_i)} \cdot P_{(H_i)}}{\sum [P_{(E|H_i)} \cdot P_{(H_i)}]} \quad [3]$$

The need to include all hypotheses in the calculation is the main weakness of this approach in handling uncertainty.

Some of the problems associated with using Bayesian probability are overcome by the implementation

of certainty factors (CF). A certainty factor can assume a value in the range  $-1$  to  $+1$ , representing the conditions of certainly false to certainly true. The system can link evidence to hypotheses using production rules of the form,

$$\begin{aligned} &\text{IF evidence } E_1 \wedge E_2 \wedge \dots E_n \text{ THEN hypothesis} \\ &H_i \text{ is true, certainty } C_i \end{aligned} \quad [4]$$

The value of  $C_i$  is derived from CF values of the evidence weighted by the CF factor associated with the rule:

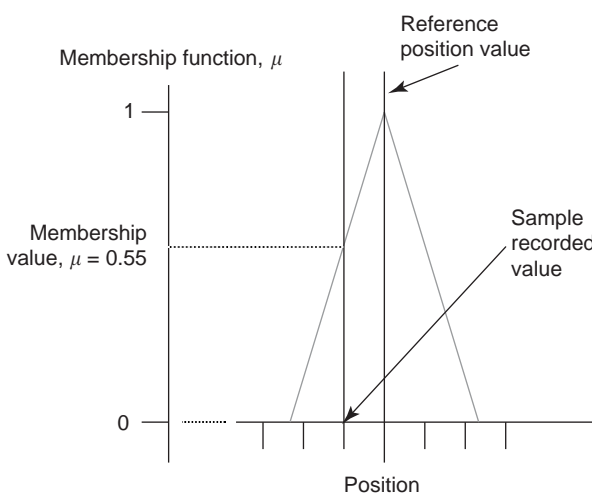
$$\begin{aligned} &CF(H_i, E_1 \wedge E_2 \wedge \dots E_n) = CF(\text{rule}) \\ &\times \min[CF(E_1), CF(E_2), \dots, CF(E_n)] \end{aligned} \quad [5]$$

This scheme can model uncertainty or errors associated with both imperfect data ( $CF(E_1) \dots CF(E_n)$ ) and imperfect domain knowledge ( $CF(\text{rule})$ ). Adoption of the scheme assumes that the conclusion of one rule is independent of the conclusion of other rules.

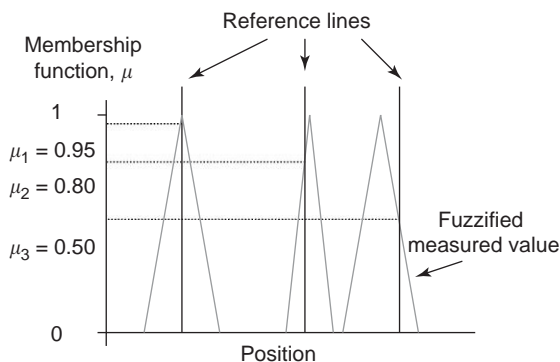
Fuzzy logic is derived from fuzzy set theory. It has received considerable attention in the scientific literature and has been associated with a number of expert systems in analytical science. Traditional set theory declares an object as either a member of a class or not, there is no ambiguity. Data interpretation, however, is often ambiguous and in practice a class may not have a clearly defined boundary. For example, inspection of an IR spectrum may unambiguously indicate the presence of a carbonyl group in a sample, but characterizing the band as originating from a ketone or ester will be less certain. Fuzzy set theory allows objects to have a possibility of membership to a class. The value of this membership ( $\mu$ ) can range from 0 (no membership) to 1 (definitely a member). Consider the apparently simple task of identifying a single peak in a spectrum. A common approach is to compare the position (expressed as wavelength, frequency, angle, etc.) of the recorded peak to a reference table to find a suitable match. The problem is defining an acceptable match. Is a difference in position of one unit acceptable or must the difference be zero? The problem can be overcome by replacing the fixed, crisp value of the peak position by a fuzzy value defining the likelihood of acceptance (Figure 7).

The three basic operators for fuzzy sets are complement, union, and intersection:

$$\begin{aligned} \mu_{\bar{A}}(x) &= 1 - \mu_A(x) \\ \mu_{A \cup B}(x) &= \max[\mu_A(x), \mu_B(x)] \\ \mu_{A \cap B}(x) &= \min[\mu_A(x), \mu_B(x)] \end{aligned} \quad [6]$$



**Figure 7** A triangular function centered about a spectral line reference value provides for uncertainty and error when matching a measured value to the reference value. The further the measured value is from the reference value, the less the match strength as given by its membership value,  $\mu$ .



**Figure 8** Multiple recorded spectral lines can be matched to reference values by recording the intersection of each with the membership function associated with each reference value. An overall match value can be gained by, for example, averaging the membership function values for each line.

where  $\mu_A(x)$  is the membership function (possibility) value for object  $x$  being in class  $A$ .

An important aggregation operator is defined by the arithmetic mean ( $h$ ),

$$h(\mu_1, \mu_2, \dots, \mu_n) = \frac{\mu_1 + \mu_2 + \dots + \mu_n}{n} \quad [7]$$

Application of such aggregation operators permits, for example, the matching of multiple recorded lines to reference lines by fuzzifying the observed peak positions (Figure 8).

## Knowledge Acquisition

Acquiring and eliciting the data and knowledge required by an expert system and structuring such

knowledge into a useable form is a primary bottleneck in expert system development. Rarely can the process be automated and few tools exist to aid the process.

Knowledge for an expert system can originate from many sources, including text books, databases, case studies, personal experience, and empirical data. Usually, the main source of knowledge is the domain expert and information is obtained through direct interaction with the human expert, at a prolonged series of intense, systematic interviews. The so-called knowledge engineer will present the expert with both model and realistic problems to solve typical of the problems the expert system will be expected to deal with. Domain experts often have great difficulty expressing their problem-solving strategies in a structured manner. Where a machine can be seen to solve problems in a piecewise manner, building on simple information in a stepwise manner to reach a more complex judgment, a human expert will seldom be seen to operate at a basic level. The expert will often make complex judgments rapidly. Pieces of basic knowledge are assumed and combined so quickly that it may be difficult for the expert to recognize or describe the process. Information may appear to be neglected and a complex reasoning process simply passed off as intuition. It is what to consider basic and relevant that makes such a person an expert.

Psychologists and management scientists have studied experts and their problem-solving techniques, using both observational and intuitive methods to measure performance and disclose expertise. Whatever methods are used, and most applications employ a wide range of techniques for acquiring the required knowledge, the process is likely to be the most time consuming and expensive part of developing the expert system.

## Expert Systems in Analytical Science

Expert systems have been extensively applied in many branches of analytical science, and in a number of noteworthy cases (generally involving molecular structure elucidation from spectroscopic data) such applications have led to the development of the technology. In addition to organic, molecular spectroscopy automated spectral interpretation systems have also been developed for X-ray diffraction, X-ray fluorescence, and, as advisors for instrument optimization, for atomic absorption spectrometry.

In chromatographic and separation science considerable investment has been made in advisory

systems for selecting the best column and conditions for a specific analysis. An interesting development of some of these systems is that they can have the ability to search online scientific literature for appropriate information.

Expert systems can be embedded within analytical instruments, to monitor equipment performance, report defects and errors, and in some cases provide maintenance and repair help to users.

The advent and implementation of laboratory information management systems has provided many opportunities for expert system technologies. Prioritizing samples and instruments and scheduling workloads can be undertaken using expert systems, in addition to summarizing and reporting on the analytical results obtained.

A well-developed explanation facility and a good user interface provides expert systems with a valuable role for training and instructing novice users of the analytical technique.

In addition to providing advice and recommendations, expert systems can be extended to act on their own results by controlling, for example, instrumentation or process plant. Since such real-time operations require decisions within a fixed, and often limited, timescale, restrictions may be imposed on the way such expert systems function. Intelligent control of instruments in industrial plants is growing. The speed and quantity of data produced by modern instrumentation outstrips the rate at which humans can process and respond to it. An advisory expert system may contain many thousands of pieces of information but the majority of this will be static and change only slowly. A control expert system, in contrast, can contain a considerable amount of both fixed and variable information about its environment. Data may change substantially within a very short period of time and the expert system must be capable of not only accessing these data but also responding to them.

The value of an expert system depends on the quality (both accuracy and completeness) of the data it contains, as well as the sophistication of the shell or inference engine. As computer power grows and instrumentation becomes more automated, the development of larger and better knowledge bases will continue and the application of expert systems will become more evident.

*See also:* **Chemometrics and Statistics:** Statistical Techniques; Multivariate Classification Techniques; Multicriteria Decision Making.

## Further Reading

- Beynon-Davies P (1991) *Expert Database Systems*. London: McGraw-Hill.
- Buchanan BG and Shortcliffe EH (1985) *Rule-Based Expert Systems*. Massachusetts: Addison-Wesley.
- Cartwright HM (1993) *Applications of Artificial Intelligence in Chemistry*. Oxford: Oxford University Press.
- Gaines BR and Boose JH (eds.) (1990) *Machine Learning and Uncertain Reasoning*. Toronto: Academic Press.
- Gray NA (1986) *Computer-Assisted Structure Elucidation*. New York: Wiley.
- Hippe Z (1991) *Artificial Intelligence in Chemistry*. Warsaw: Elsevier.
- Hohne BA and Pierce TH (2001) *Expert System Applications in Chemistry*. Oxford: Oxford University Press.
- Jackson P (1998) *Introduction to Expert Systems*, 3rd edn. Wokingham: Addison-Wesley.
- Kandel A (1991) *Fuzzy Expert Systems*. Boca Raton: CRC Press.
- Waterman DA (1985) *A Guide to Expert Systems*. California: Addison-Wesley.

## Multicriteria Decision Making

**S Kokot and G A Ayoko**, Queensland University of Technology, Brisbane, QLD, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The name 'chemometrics' was first used by Svante Wold in a Swedish journal. In general, it refers to a chemical discipline that focuses on maximizing the extraction of information from data and experimental measurements with the aid of mathematical, computational, and logic methods. The data or information collected are submitted for analysis by one or more methods of chemometrics typically associated with pattern recognition, classification, or prediction.

A detailed discussion of these three common chemometrics methods may be found elsewhere in this Encyclopedia or in compendia such as that by D.L. Massart *et al.* In general, they provide a backdrop against which the concepts and contributions of the multicriteria decision-making methods (MCDM) may be compared and understood.

Raw data are normally arranged in matrix form with rows representing objects, such as samples, spectra, chromatograms, and columns representing variables, such as different chemical, physical, or biological characteristics of objects, spectral frequencies or wavelengths, and time. Thus, chemometrics facilitates the interpretation of multivariate systems generally collected in two-dimensional data matrix format although multiway methods of data presentation are also available. Clearly, chemometrics relies on fast computational processes, and as a discipline did not seriously appear on the agenda of analytical

chemistry until the 1970s when sufficiently fast desktop computers became more common. Since that time, thousands of scientific investigations and practical industrial or laboratory applications have been reported and utilized. These range, for example, from the classical rapid near infrared spectroscopic determination of properties of wheat or petroleum products to the more recent simultaneous spectrophotometric determination of several metal ions present in real-world electroplating solutions or the simultaneous kinetic-spectrophotometric determination of carbamate pesticides with the aid of three-way data unfolding.

There are significant advantages in using the multivariate approach:

- Chemometrics methodology is often able to model quite complex responses consisting of many overlapping signals from individual analytes.
- The ability to predict simultaneously, concentrations or property values of several analytes from one measured response obtained from a mixture of the analytes.
- Requirement for sophisticated and costly instrumentation can sometimes be replaced by quite inexpensive methods, the performance of which is enhanced with respect to data interpretation and calibration modeling by chemometrics.

In the context of analytical chemistry, these advantages are well illustrated in recent studies such as the simultaneous determination of five metal ions found in electroplating solutions, with the use of visible spectrophotometry and chemometrics and the simultaneous kinetic-spectrophotometric determination of carbamate pesticides.

## Further Reading

- Beynon-Davies P (1991) *Expert Database Systems*. London: McGraw-Hill.
- Buchanan BG and Shortcliffe EH (1985) *Rule-Based Expert Systems*. Massachusetts: Addison-Wesley.
- Cartwright HM (1993) *Applications of Artificial Intelligence in Chemistry*. Oxford: Oxford University Press.
- Gaines BR and Boose JH (eds.) (1990) *Machine Learning and Uncertain Reasoning*. Toronto: Academic Press.
- Gray NA (1986) *Computer-Assisted Structure Elucidation*. New York: Wiley.
- Hippe Z (1991) *Artificial Intelligence in Chemistry*. Warsaw: Elsevier.
- Hohne BA and Pierce TH (2001) *Expert System Applications in Chemistry*. Oxford: Oxford University Press.
- Jackson P (1998) *Introduction to Expert Systems*, 3rd edn. Wokingham: Addison-Wesley.
- Kandel A (1991) *Fuzzy Expert Systems*. Boca Raton: CRC Press.
- Waterman DA (1985) *A Guide to Expert Systems*. California: Addison-Wesley.

## Multicriteria Decision Making

**S Kokot and G A Ayoko**, Queensland University of Technology, Brisbane, QLD, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The name 'chemometrics' was first used by Svante Wold in a Swedish journal. In general, it refers to a chemical discipline that focuses on maximizing the extraction of information from data and experimental measurements with the aid of mathematical, computational, and logic methods. The data or information collected are submitted for analysis by one or more methods of chemometrics typically associated with pattern recognition, classification, or prediction.

A detailed discussion of these three common chemometrics methods may be found elsewhere in this Encyclopedia or in compendia such as that by D.L. Massart *et al.* In general, they provide a backdrop against which the concepts and contributions of the multicriteria decision-making methods (MCDM) may be compared and understood.

Raw data are normally arranged in matrix form with rows representing objects, such as samples, spectra, chromatograms, and columns representing variables, such as different chemical, physical, or biological characteristics of objects, spectral frequencies or wavelengths, and time. Thus, chemometrics facilitates the interpretation of multivariate systems generally collected in two-dimensional data matrix format although multiway methods of data presentation are also available. Clearly, chemometrics relies on fast computational processes, and as a discipline did not seriously appear on the agenda of analytical

chemistry until the 1970s when sufficiently fast desktop computers became more common. Since that time, thousands of scientific investigations and practical industrial or laboratory applications have been reported and utilized. These range, for example, from the classical rapid near infrared spectroscopic determination of properties of wheat or petroleum products to the more recent simultaneous spectrophotometric determination of several metal ions present in real-world electroplating solutions or the simultaneous kinetic-spectrophotometric determination of carbamate pesticides with the aid of three-way data unfolding.

There are significant advantages in using the multivariate approach:

- Chemometrics methodology is often able to model quite complex responses consisting of many overlapping signals from individual analytes.
- The ability to predict simultaneously, concentrations or property values of several analytes from one measured response obtained from a mixture of the analytes.
- Requirement for sophisticated and costly instrumentation can sometimes be replaced by quite inexpensive methods, the performance of which is enhanced with respect to data interpretation and calibration modeling by chemometrics.

In the context of analytical chemistry, these advantages are well illustrated in recent studies such as the simultaneous determination of five metal ions found in electroplating solutions, with the use of visible spectrophotometry and chemometrics and the simultaneous kinetic-spectrophotometric determination of carbamate pesticides.



Alternatively, MCDM methods are principally concerned with selection, optimization, and decision-making. The bringing together of chemometrics and MCDM methods so as to maximize information from a set of data has been an interesting, useful, and expanding endeavor in chemistry.

## Role of Decisions

Making decisions intuitively, subjectively, or objectively is the experience of every human being. We make decisions when we have fun, when we are at work, at leisure, or when threatened, to give but a few examples. We ask such questions as: am I comfortable? Is this correct? Should I spend this money? And if I do what happens to some other scenarios contingent on finances? We make many of our decisions qualitatively. Recently, Brans discussed this problem in the context of the Natural Real World, and it was noted that the human species developed an approach for making decisions, which is based on measurement and modeling. However, mathematical models are developed in virtual space and are usually approximations of reality. Consequently, there is always a role for the decision-maker. This implies that a degree of subjectivity is used to modify the mathematical model, which is regarded as the rational solution. Interestingly, three elements of thought are incorporated into the decision-making process – rationality, subjectivity, and ethics. It is in this context that the MCDM models are broadly considered today and applied to a huge variety of problems, e.g., comparison of performance of baseball teams, development of negotiation support systems, selecting landmine detection strategies, etc. Of particular interest is a study that investigated the contribution of the MCDM methods for decision support systems. Over 20 studies involving a variety of problems and methods were investigated, and the outcomes of each study were formally analyzed with respect to the contribution of MCDM analysis.

All of the above illustrations are outside the realm of analytical chemistry and instrumental analysis, which are vital for qualitative and quantitative definition or description of scenarios in environment, agriculture, civil engineering, medicine, and similar fields of scientific and technological endeavor. Since chemical data derived by analytical methods are critical for decision-making in these fields, many examples of the applications of MCDM can be found in such fields. One important common element in decision-making in these fields is that it is becoming more and more multidisciplinary, i.e., several disparate stakeholders, such as administrators, scientists,

technologists, doctors, contribute to the outcome of the decision-making process, and, thus, the solution sought is a compromise. Some MCDM methods facilitate the analysis of such multidisciplinary scenarios by providing valuable guidance to the final decision, in which the decision makers ultimately have the final choice.

## Comparison of Some MCDM Methods

MCDM methods commonly offer partial preordering as well as net full ordering or ranking of objects. The latter ordering method is well known. The objects are simply lined up according to some index either top-down (largest index value is preferred; sometimes referred to as maximized ranking) or bottom-up (the smallest index value is preferred, i.e., minimized ranking). Partial preordering, on the other hand, is less common. It is concerned with the situation where objects may perform equally well but on different variables, i.e., they cannot be compared and one object cannot be preferred to others. When this occurs, such objects are alternatives. The simple matrix in Table 1 illustrates these concepts.

From Table 1, it can be clearly seen that O1 is the preferred object (highest row sum) while O4 is the worst performing one (lowest row sum); O2 and O3 have the same row sums, which are derived from different variables. Hence, these two objects cannot be compared, and are therefore alternatives. The partial preorder of objects 1–4 may be represented as O1, (O2/O3), O4 ('maximized' ranking). If the criterion was reversed, i.e., the lowest row sum was preferred ('minimized ranking'), then the rank order would also be reversed, O4, (O2/O3), O1.

Some common MCDM methods include PARETO, elimination et choix traduisant la réalité (ELECTRE), simple multiattribute ranking technique (SMART), ORESTE, NAIADE, preference ranking organization method for enrichment evaluation (PROMETHEE), Hasse diagram technique (HDT), multiple attribute utility theory (MAUT), and analytical hierarchy process (AHP). For details of their algorithms, differences, and refinements the reader is

**Table 1** Matrix illustrating ranking concepts (preference criterion: highest row sum preferred)

Objects	Variables				Row sum
	V1	V2	V3	V4	
O1	1	1	1	1	4
O2	1	0	1	1	3
O3	1	1	0	1	3
O4	0	0	0	0	0



referred to the primary literature, a start to which may be made with the references from the 'Further Reading' section. The central aspect of these methods is that they are designed to provide a method of comparison in terms of performance or preference of one object to another. The methods provide this information in different ways, e.g., ELECTRE and PROMETHEE are outranking methods, AHP gives a priority of alternatives. Some methods provide a full outranking order of objects, and others also give their partial preorder, which reflects their incomparability. A visual display of the rankings with the aid of line diagrams is sometimes available to indicate the rank order as illustrated in **Figure 1**. This mode of presentation is particularly useful because many people from different backgrounds can easily relate to pictorial information relative to equivalent numerical presentation. This is important because quite often decisions have to be explained to participants or decision makers with different educational backgrounds.

There are no universal methods for comparing the performance of MCDM methods. However, there have been many studies concerned with the selection of the most appropriate MCDM method. In one substantial study, Al-Shemmeri *et al.* evaluated the performance of 16 MCDM methods to solve a multicriteria water resources problem. The methods were rated on the basis of 24 criteria spread more or less evenly over four categories such as characteristics describing (1) the problem, (2) the decision maker, (3) the techniques, and (4) the solution obtained. PROMETHEE was declared the best performing method. However, in another study concerned with the ranking of chemical substances, and focusing on the relatively recent MCDM approach, the partial preorder HDT, Lerche *et al.* found that this method was to be preferred over others including PROMETHEE. The latter method, nevertheless, ranked well ahead of methods such as NAIADE, ORESTE, and AHP. The important criteria for

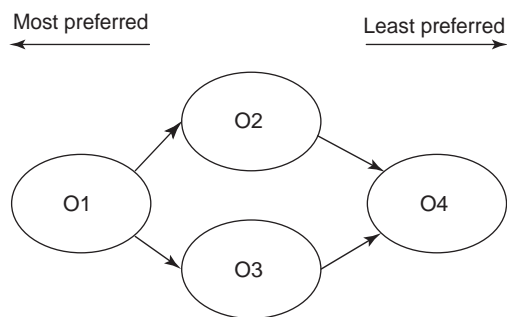
method comparison in this case were the minimization of subjectivity and facilitation of method transparency, i.e., the role of the decision maker was minimized. Others have argued that the real world is about risks and balances and strongly involves the decision maker. Therefore, provision for the inclusion of subjectivity, ethics, as well as the rational contributions from science and technology is important. Three MCDM methods regarded by Salminen *et al.* to be particularly suitable for application to environmental problems, SMART, ELECTRE III, and PROMETHEE, were compared, and it was found that there was little difference between PROMETHEE and SMART, but it would seem that ELECTRE III had some extra functionality. Interestingly, others regard PROMETHEE to be more refined than ELECTRE in that the former method quantifies the degree of preference by estimating a global preference index of an object compared with another for each criterion. The consistent general conclusions from such method comparison studies suggest that:

- there is no unique best MCDM method;
- the decision maker is always an important contributor to the MCDM process; and
- it is useful to apply more than one MCDM method to find the compromise solution to the problem.

## MCDM Methods and Chemometrics

As indicated in the previous sections, the scope of application of MCDM methods is generally huge. This is in contrast to their applications in chemically related fields such as analytical chemistry and similar fields where chemical criteria play a significant role. In general, the qualitative exploratory principal component analysis (PCA) and the quantitative prediction of analytes have dominated analytical chemistry by multivariate calibration methods. Other chemical fields have followed more or less similarly although MCDM methods have found significant application in the environmental chemistry area. This is probably so because there is a substantial need to involve multidisciplinary teams and criteria in the resolution of environmental problems. In addition, there are usually many stakeholders and decision makers that provide inputs to find the compromise solution in such problems. Therefore, decision-making support methodologies that have the capacity and flexibility to offer diverse modeling options and include the subjectivity of the decision makers are required.

Keller *et al.* have been credited with the introduction of MCDM methods to chemometrics in 1991.



**Figure 1** Diagram for the partial preorder of the example shown in **Table 1**.

They explained the decision-making concepts with the use of data matrix containing information on the development of a chemical formulation for the treatment of textile products. PARETO Optimality, ELECTRE, and PROMETHEE partnered by GAIA MCDM methods were applied and compared. The first of these, PARETO Optimality, was found to be useful in simple bicriteria cases, while the remaining methods illustrated the more substantial multivariate approaches. ELECTRE and PROMETHEE provided similar ranking and selection information regarding the formulation. Both facilitate graphical presentation of partial and net outranking flows. However, when linked to GAIA, which is essentially a form of a PCA biplot (PC1 versus PC2), very useful additional information is provided about the rank order of the objects and particularly about the criteria, which are significant in influencing the ranking. Interestingly, the data matrix for GAIA is generated from the PROMETHEE net outranking flows, and anecdotal observations suggest that the PCA clustering from this matrix is rather crisper than that achieved from the conventionally standardized or normalized raw data matrix. It is also interesting to note that in their 1997 definitive text on chemometrics, Massart *et al.* relied heavily on the above work by Keller *et al.* in their section on MCDM. Apart from the three methods above, arguably only one other substantial method, Utility Functions, was discussed. Following Keller *et al.*'s study, PROMETHEE and GAIA have been applied to a number of chemically related areas, including the problem of selecting the most suitable microwave digestion method for the dissolution of soil samples and comparison of the quality of Australian and Vietnamese rice grains. In an in-depth study by Kokot and Phuong, which appeared in 1999, PROMETHEE models were tested on a combined matrix consisting of the chemical (elemental content), biochemical (proteins), and physiobiological (length, chalkiness) characteristics of various types of Australian and Vietnamese rice grains. More recent examples of the application of MCDM in chemometrics are illustrated by the PROMETHEE and GAIA analysis of the prediction performance of various multivariate calibration models derived from different methods of chemometrics, e.g., PLS, BP-ANN, RBF-ANN, PARAFAC, and others. Chemistry and microbiology were brought together in a further PROMETHEE and GAIA study of ranking of some organotin(IV) compounds in relation to their fungicidal properties, while several recent investigations, reported at symposia, applied the same MCDM methodology to air quality problems in homes and in relation to performance of truck engines. These MCDM methods have also been successfully used

to resolve life cycle assessment problems and to plan the development of river alluvial plains. While these examples are by no means exhaustive, they serve to illustrate the range of more recent applications of MCDM methodology in chemometrics. In addition, they demonstrate how these two versatile and well-established MCDM methods may be set up to interpret a problem and illustrate the powerful visual presentation of the outcomes.

## The PROMETHEE Model

PROMETHEE is a nonparametric method applied in Euclidian space to rank objects. Each variable in the raw data matrix is set to maximize or minimize; then the data array is converted to a difference,  $d$ , matrix. This is achieved for each criterion by comparing all values pairwise by subtraction in all possible combinations. Preference indices are computed for each  $d$  value for each object with the use of a mathematical preference function selected independently for each variable, e.g., linear function or a Gaussian. When commercial software is used, such functions are simply chosen from the selection provided but when the algorithm is written in-house then any function deemed suitable might be included. From the individual preference indices, the global preference can be computed for each object and the positive and negative preference flows,  $\Phi^+$  and  $\Phi^-$ , are calculated from these indices; the former indicates how each object outranks all others and the latter how each object is outperformed by all others. These outranking flows are compared and produce a partial pre-order according to three rules:

- i. one object is preferred to another,
- ii. there is no difference between the objects,
- iii. objects cannot be compared.

It is rule (iii) that gives rise to the alternative objects of the same rank illustrated in **Figure 1**. This type of ranking is called PROMETHEE I. If rule (iii) is removed, the net flows,  $\Phi$ , are obtained and the procedure is known as PROMETHEE II. This latter scale is intuitively more satisfying but in fact tends to be less reliable than that derived from PROMETHEE I. However, particularly when large matrices are used the PROMETHEE I diagrams become very complex and challenging to interpret, and PROMETHEE II net flows may be preferred.

## PROMETHEE and GAIA: An Example

PROMETHEE and GAIA can be applied to the matrix shown in **Table 2** in order to choose the best

**Table 2** Performance matrix for some materials

Material	Property P1	Property P2	Property P3	Property P4	Property P5	Ease of use	Cost (in \$)
O1	2.0	1.5	3.0	3.0	4.0	2.0	50.0
O2	4.0	3.0	6.0	4.0	8.0	4.0	50.0
O3	3.0	1.5	3.0	3.0	4.0	2.0	50.0
O4	2.0	1.5	3.0	2.5	3.5	3.0	30.0
O5	2.0	1.5	3.0	2.5	3.0	1.0	30.0
O6	3.0	3.5	6.0	2.5	6.0	2.0	50.0
O7	1.5	1.0	2.0	1.0	3.0	3.0	5.0
O8	1.0	0.5	0.5	0.5	2.0	2.0	5.0
O9	2.0	1.5	2.0	1.5	1.0	1.0	30.0
O10	2.0	1.5	2.0	3.0	3.0	4.0	40.0

material for a special function; for example, removal of stains from carpets.

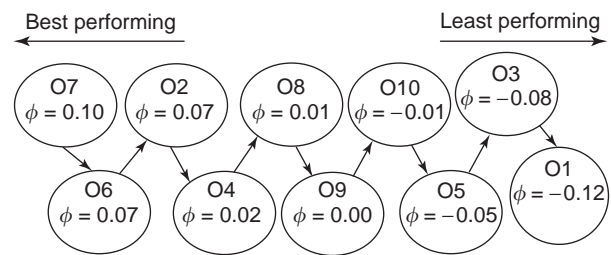
A typical decision maker would minimally prefer materials that possess the desired properties, and are easy to use while not overtly expensive. Since no one single material obviously gives the best outcome for all of the variables, PROMETHEE can assist the decision maker in arriving at the 'best compromise' solution in this multicriteria problem.

In Table 2, 'ease of use' is rated on a five-point scale, with 5 as the best and 1 as the worst while the properties are rated on 0–10 scales. To apply PROMETHEE and GAIA to the matrix, for each variable, it must be determined whether higher or lower values are preferred, i.e., whether the variable should be maximized or minimized. Next, the appropriate preference function for each variable must be chosen from the list of six preference functions available in the commercial software. In this example, based on the premise that the best material should be, efficient, easy to use, and cost-effective, the maximized linear preference function could be selected for properties P1, P2, and P3 as well as the 'ease of use', while the minimized linear preference function is used for cost and properties P4 and P5. One of the advantages of PROMETHEE and GAIA over other comparable ranking procedures is their amenability to sensitivity analysis. Thus, weights, which indicate the priority the decision maker attaches to each variable, can be set.

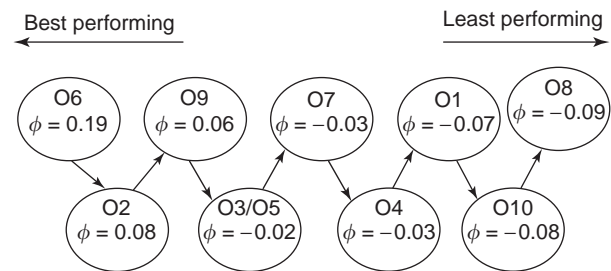
In this example (Table 2), if the decision maker attaches equal priority to all variables and they are weighted equally, the complete ranking result shown in Figure 2 would be obtained.

Thus, material O1 is the worst performing while O7 is the best performing. As illustrated by Figures 3 and 4, the ranking is sensitive to the variables considered.

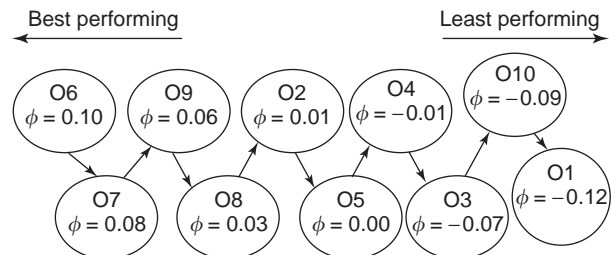
It is also noteworthy that the position of the decision axis is sensitive to the weighting set for the variables (see Figures 5 and 6).



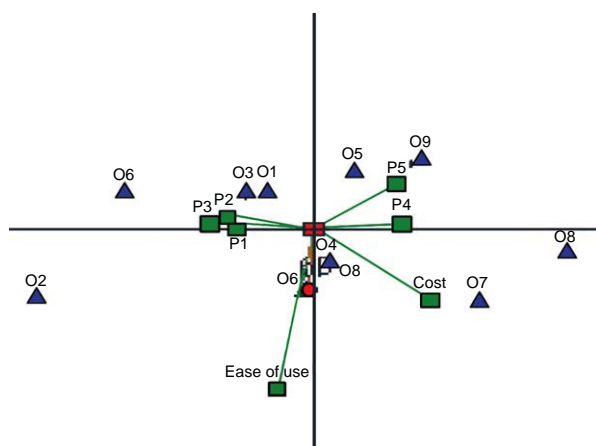
**Figure 2** Complete ranking of materials based on their properties, ease of use, and cost.  $\phi$  denotes the net outranking flow for each material: the higher its value the higher the preference of the material.



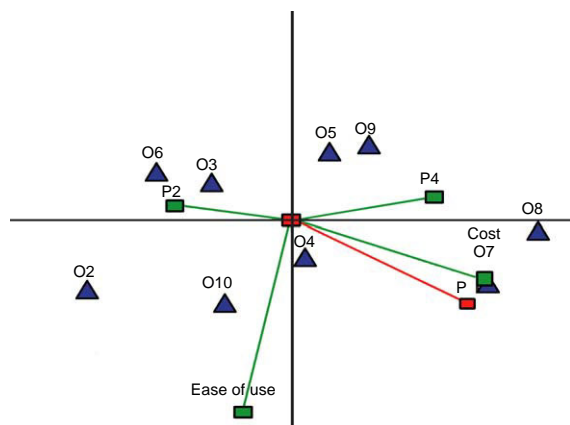
**Figure 3** Complete ranking of materials based on their properties, P, only. O8 is the worst-performing material; O6 the best-performing material, and O7 is ranked sixth.



**Figure 4** Complete ranking of materials based on their properties and cost. O1 is the worst-performing material and O6 is the best-performing material. Although the least-performing material is similar to that in Figure 3, the rank orders of the other materials varied as did the net outranking flows.



**Figure 5** GAIA plane obtained for the matrix shown in Table 2 when all variables are equally weighted; the  $\pi$  (pi) decision axis, which represents the condensed weighting of the variables, is close to the vectors for 'ease of use', suggesting that it is the most important factor influencing the ranking of the materials. A total of 87% of the data variance is accounted for by the first two principal components.



**Figure 6** GAIA biplot obtained when 'cost' is given a weighting of 5 and other variables are given unit weighting. The decision axis visibly moves toward the 'cost' axis. Vectors for properties P1, P2, and P3 are oriented in the same direction in Figure 5, suggesting that they represent equivalent information. Consequently, all three properties may be represented by P2 in this figure. Similarly, properties P4 and P5 are oriented in the same direction in Figure 5. Therefore, they may be represented in this figure by P4.

MCDM methods are powerful tools that can provide guidance for many practical problems where there are many stakeholders, and they offer extensive opportunities to extract compromise solutions, which take into considerations not only the rational findings of science and technology but also ethics and subjectivity of the decision maker.

See also: **Chemometrics and Statistics:** Statistical Techniques; Optimization Strategies; Multivariate Classification Techniques; Multivariate Calibration Techniques.

## Further Reading

- Al-Shemmeri T, Al-Kloub B, and Pearman A (1997) Model choice in multi-criteria decision aid. *European Journal of Operational Research* 97: 550–560.
- Asker N and Kokot S (1991) The application of NIR spectroscopy for the prediction of properties of Australian refined reformat. *Applied Spectroscopy* 45: 1153–1157.
- Ayoko GA, Bonire JJ, Abdulkadir SS, *et al.* (2003) A multi-criteria ranking of organotin(IV) compounds with fungicidal properties. *Applied Organometallic Chemistry* 17: 749–758.
- Brans JP (2002) Ethics and decisions. *European Journal of Operational Research* 136: 340–352.
- Eom SB and Min H (1999) The contributions of multi-criteria decision making to the development of decision support systems subspecialties: An empirical investigation. *Journal of Multi-criteria Decision Analysis* 8: 239–255.
- Janssen R (2001) On the use of multi-criteria analysis in environmental impact assessment in the Netherlands. *Journal of Multi-criteria Decision Analysis* 10: 101–109.
- Keller HR, Massart DL, and Brans JP (1991) Multi-criteria decision making: A case study. *Chemometrics and Intelligent Laboratory Systems* 11: 175–189.
- Kokot S and Phuong TD (1999) Elemental content of Vietnamese rice. Part 2. Multivariate data analysis. *Analyst* 136: 561–569.
- Lerche D, Brüeggemann R, Soerensen P, Carlsen L, and Nielsen OJ (2002) A comparison of partial order technique with three methods of multi-criteria analysis for ranking of chemical substances. *Journal of Chemical Information and Computational Sciences* 41: 1086–1098.
- Ni Y, Chen S, and Kokot S (2002) Spectrophotometric determination of metal ions in electroplating solution in the presence of ethylenediaminetetraacetic acid by multivariate calibration and artificial neural networks. *Analytica Chimica Acta* 463: 305–316.
- Massart DL, Vandeginste BGM, Buydens LMC, *et al.* (1997) *Handbook of Chemometrics and Qualimetrics: Part A*. Amsterdam: Elsevier.
- Panayiotou H and Kokot S (1999) Matching and discrimination of single human – scalp hairs by FT-IR microspectroscopy and chemometrics. *Analytica Chimica Acta* 392: 223–235.
- Salminen P, Hokkanen J, and Lahdelma (1998) Comparing multi-criteria methods in the context of environmental problems. *European Journal of Operational Research* 104: 485–496.



## Signal Processing

R G Brereton, University of Bristol, Bristol, UK

© 2005, Elsevier Ltd. All Rights Reserved.

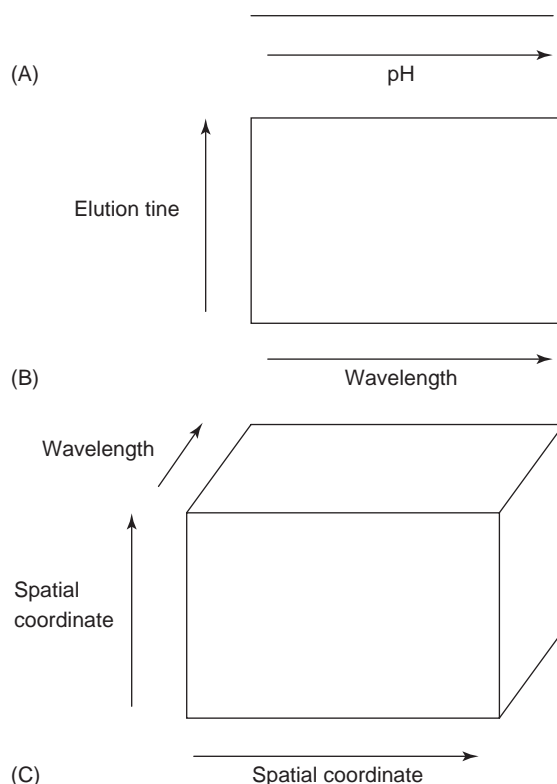
### Introduction

Most analytical information is obtained in computerized form in both spectroscopy and chromatography. Signal processing is required to convert this raw digital information into graphs and numbers that are interpretable to the analytical scientist. Aims of signal processing include data reduction, integration, determining the numbers, positions, and shapes of peaks, and detection of weak signals.

The majority of analytical signals are univariate where a single response is measured normally as a function of time, although other variables such as spectral frequency or voltage can be employed. Examples include the intensity of a signal in gas chromatography obtained using a flame ionization detector as a function of time, the magnetization of a nuclear magnetic resonance spectrum as a function of frequency, and the intensity of absorbance at a specific wavelength in ultraviolet visible spectroscopy as a function of pH. Multivariate signals occur when the response is measured as a function of more than one variable, commonly two, such as diode array high-performance liquid chromatography where the intensity of absorbance of a spectrum is measured as a function of both elution time and wavelength. Higher-order signals where the response is a function of more than two variables are possible, an example is chemical microscopy where a spectrum is recorded of an image; the response is absorbance, and one of the three variables is spectral wavelength and the other two are spatial. These types of signals are illustrated in Figure 1.

### Digital Sampling

Scientific instrumentation is at the heart of the modern analytical laboratory, and the majority of information obtained is in the form of digitized signals. Conventionally, many analytical textbooks deal with analog signals, for example, such as that coming from a chart recorder, but this type of signal is rare in the modern laboratory. The output from instruments is primarily in terms of information consisting of a series of numbers, normally recorded sequentially, although some instruments, such as those containing a diode array, may collect these numbers simultaneously.



**Figure 1** Types of signals (A) univariate, e.g., the absorbance at a single spectroscopic wavelength as a function of pH. (B) Multivariate as a function of two variables, e.g., the intensity of absorbance in diode array high-performance liquid chromatography as a function of elution time and wavelength. (C) Multivariate as a function of three variables, e.g., intensity of absorbance in chemical microscopy as a function of wavelength and two spatial coordinates.

Signal processing is required to convert this raw digital information into something that is readily interpretable, normally graphically, such as a spectrum or a chromatogram. It is important to understand the digital nature of the data.

### Sampling Interval

Normally signals are sampled at regular intervals, for example, a ultraviolet/visible (UV/vis) spectrum may be sampled every nm, or a Fourier transform nuclear magnetic resonance (FT-NMR) spectrum every ms, usually the interval is regularly spaced. However, it is important always to check the units of the raw data, although, for example, mid-infrared (IR) spectra could be presented either in units of  $\text{cm}^{-1}$  or nm, generally the signals are evenly acquired in intervals based on  $\text{cm}^{-1}$ .

For multivariate signals, two or more variables are sampled; for example, in diode array high-performance liquid chromatography (HPLC) these are wavelength and time. If a spectrum is sampled every 2 nm, between 200 and 398 nm, and the chromatogram is acquired at 1 s intervals between 5 min and 6 min 59 s, then a total of  $100 (\text{wavelengths}) \times 120 (\text{points in time})$  or 12 000 datapoints are acquired.

## Sampling Window

In many forms of spectroscopy each data point is acquired over a window, rather than at a precise point. For example, in UV/vis spectroscopy a typical sampling window may be 7 nm, even if the sampling rate is smaller. This means, for example, that the spectral intensity measured at 220 nm is in fact the average intensity between 217 nm and 223 nm. This is illustrated in **Figure 2**. Many spectrometers employ windows that are large relative to sampling rates because this increases the signal-to-noise ratio, but comes at the cost of signal blurring. However, in certain other spectroscopies, such as FT-NMR, the window is by necessity small relative to sampling rate.

## Analog-to-Digital Conversion

Signals are not only digitized by recording a series of numbers usually obtained in time, but also the intensity is digitized, because the raw data have to be acquired by a computer. Computers acquire information in the form of bits, which are binary representations of integers. For example, the number 6 has a binary representation of 110 ( $= 4 + 2 + 0$ ). Any integer has a binary equivalent. However, the range of integers that can be acquired depends on the number of bits in an analog-to-digital converter (ADC): a 6 bit ADC measures a maximum range of numbers of 111 111 or  $2^7 - 1$  or 127. In many cases half the integers are reserved for negative numbers,

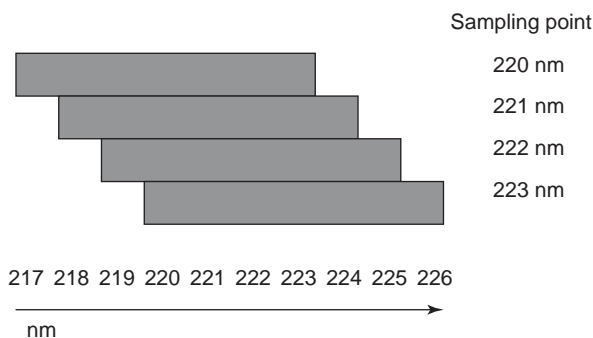
so a 6 bit ADC could be used to measure a range of numbers between  $-63$  and  $+63$ . Typically, ADCs consist of much larger number of bits, for example, 20 bits, which allows a range of about a million numbers. Raw readings, often in the form of voltages, are converted to these numbers by the ADC. A procedure is required to determine the relationship between raw measurement and integer value, this is usually via a receiver gain. In many instruments the size of the receiver gain is not under the control of the operator, because sensible measurements will be within a specific range; for example, in the case of UV/vis measurements, a typical range of 0.01–2 AU may be encountered in practice, outside this range spectroscopic measurements have little quantitative meaning. In other types of spectroscopies, such as NMR, there may be a wider range of measurements (e.g., dependent on concentrations of compounds) and the operator can set this parameter manually. Correct use of instruments maximizes the usefulness of the entire digital range. An ADC that is too low can result in low-intensity signals that are poorly digitized in the intensity direction, whereas one that is too high can result in overloaded signals that are cut off.

## Signals and Noise

A typical spectrum or chromatogram consists of a series of signals, usually of chemical origin, imposed upon noise. The main aim of signal processing is to obtain information about these underlying signals and remove the noise. The signals are often in the form of a series of peaks, which may correspond, for example, to the chromatographic elution profiles of a series of compounds or to electronic spectroscopic transitions. The analytical scientist usually wants to interpret these peaks both qualitatively (e.g., to identify compounds) and quantitatively (e.g., to determine the concentrations of chemical species).

## Peakshapes

Each peak is characterized by several parameters. The most usual are (1) the position of the center (which may, for example, correspond to an elution time or an  $m/z$  value), (2) the area (which may relate to concentration), and (3) the width – generally at half-height (which may provide important diagnostic information, for example, how well a chromatographic column is performing). This is illustrated in **Figure 3**. In some applications, such as chromatography, peakwidths at the base are measured, by fitting a triangle to the peak and seeing the size of the base of the triangle.



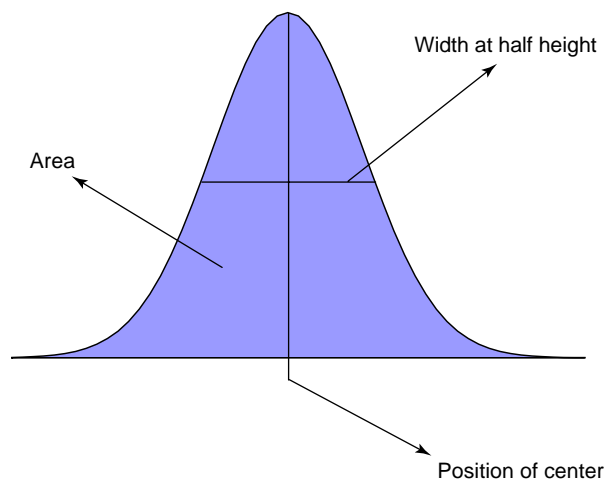
**Figure 2** Illustration of windows, for four successive samples of a spectrum centered on 220, 221, 222, and 223 nm.

Peaks can be of a variety of shapes. However, pure peaks are usually unimodal, and in most cases symmetric. The two most common shapes are Gaussian (as occurs in most types of spectroscopy and chromatography) and Lorentzian (specifically encountered in NMR). Lorentzian peaks tail more than Gaussians as illustrated in **Figure 4**. Occasionally, asymmetric peak shapes are encountered, for example, tailing chromatographic peaks as illustrated in **Figure 5**. Usually an symmetric peak shape is modeled empirically, using a Lorentzian to model the tailing half and a Gaussian the sharper half.

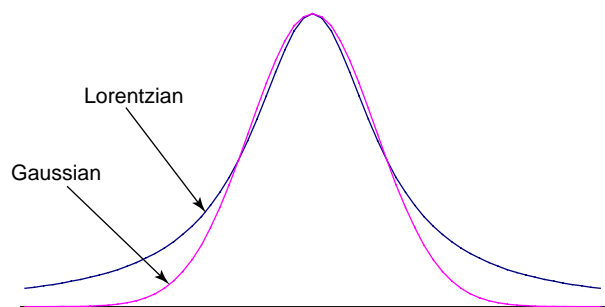
Typical analytical data can be represented as a sum of several such peaks. For example, a chromatogram arising from seven compounds could be modeled by a sum of seven Gaussians, each characterized by a position, width, and area, and so represented by 21 parameters in total.

### Definition of Resolution

Resolution relates to how well peaks can be characterized, and depends on the separation and widths of peaks. **Figure 6** shows two pairs of peaks both separated by an identical amount, but the top pair is less



**Figure 3** Main parameters that characterize a peak.



**Figure 4** Difference between Gaussian and Lorentzian peak-shapes.

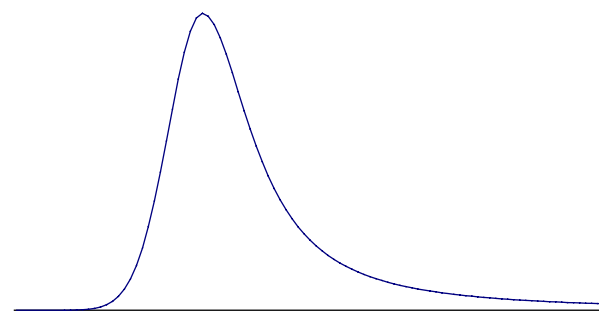
well resolved than the bottom. There are various methods for quantifying resolution; in chromatography a popular one is to divide the difference between the peak centers by the average peak width at their base. A value of 1.5 implies excellent resolution, and one of 0.5 implies the peaks are hardly resolved.

### Noise

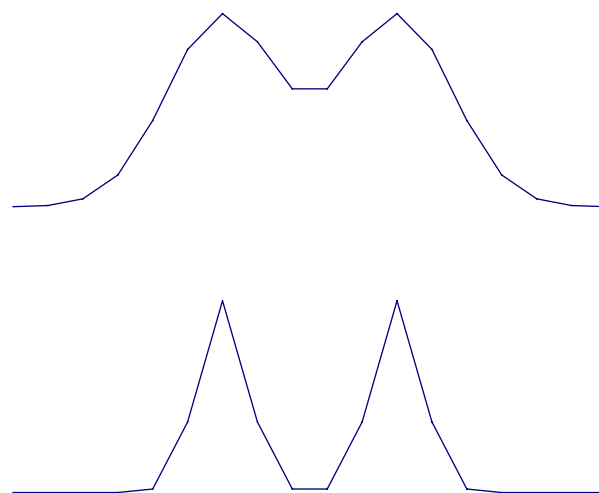
Superimposed on signals is noise. There are several common types of noise.

If the noise at each successive point (normally in time) does not depend on the level of noise at the previous point we call it stationary noise. There are two major types.

1. *Homoscedastic noise*. This is the simplest to envisage. The features of the noise, normally the mean and standard deviation, remain constant over the entire data series. The most common type of noise is given by a normal distribution. In most real-world situations, there are several sources of instrumental noise, but a combination of different symmetric noise distributions often tends toward a



**Figure 5** A tailing peak.



**Figure 6** Two pairs of peaks, both of identical separation but different resolution.



normal distribution. Hence, this is a good approximation in the absence of more detailed knowledge of a system.

2. *Heteroscedastic noise*. This type of noise is dependent on signal intensity, often proportional to intensity. The noise may still be represented by a normal distribution, but the standard deviation of that distribution is proportional to intensity.

Sometimes raw signals are transformed mathematically, for example, in optical spectroscopy it is usual to convert transmittance to absorbance data using a logarithmic transformation. This changes the noise characteristics, often to a log-normal distribution, although the origins of the instrumental noise are still homoscedastic.

A second type of noise is nonstationary. As a series is sampled, the level of noise in each point depends on that of the previous point. This is quite common in process monitoring. For example, there may be problems in one aspect of the manufacturing procedure, an example being the proportion of an ingredient. If the proportion is in error by 0.5% at 2 p.m., does this provide an indication of the error at 2.30 p.m.? Many such sources of noise cannot be modeled in great detail, but a generalized approach is that of autoregressive moving average (ARMA) noise. The moving average component relates the noise at the current to the values at previous times. The autoregressive component relates the noise to the observed value of the signal at one or more previous times. ARMA processes are well described in texts on time series.

### Signal-to-Noise Ratio

It is often important to measure the signal-to-noise ratio. The higher this is the easier a signal is to detect. There are a variety of methods for determining this ratio, but a common one is as follows. First, a region of the data is selected where there appear to be no signals. Then either the root mean square or the standard deviation of the data in this region is determined as an indicator of the noise level. Finally, the height of a signal is divided by the noise level.

In addition to providing a quantitative measure of the quality of spectroscopy or chromatography, this can also be used as an indicator as to whether a peak can be detected or not. If the signal-to-noise ratio exceeds 3 it is assumed that the peak is real. This is because, using a normal distribution, a value exceeding 3 standard deviations from the mean is expected in  $\sim 0.1\%$  of cases, so there is a chance of more than 99.9% that the peak is a real one, rather than originating from an artifact arising from the noise.

Signal-to-noise criteria are commonly employed as part of peak picking algorithms.

## Aims of Signal Processing

There are several important aims.

### Detection

Detection of weak signals is of significant importance. There are two reasons why a signal may be hard to detect. The first is because the noise level is high or because the amount of analyte is small, resulting in a low signal-to-noise ratio. A simple approach to overcome this limitation, possible in most types of spectroscopy where the positions of signals are quite reproducible, is time averaging. The signal-to-noise ratio increases by  $\sqrt{N}$  where  $N$  is the number of signals averaged; however, this can be impracticable in some cases, for example, if it takes 5 min to scan a spectrum it will take 500 min to increase the ratio ten-fold. Alternatives involve using smoothing functions as discussed in this encyclopedia elsewhere, or to use a faster technique for obtaining data, e.g., Fourier transform spectroscopy.

A second reason is that a minor peak may be buried within a larger one. This is common, for example, in chromatographic purity monitoring. Improving the physical measurement method (e.g., by better chromatography) is one approach, by using computational resolution enhancement functions as discussed in this encyclopedia elsewhere is also possible. If there is another dimension (e.g., coupled chromatography), it is possible to employ multivariate methods for peak purity assessment.

### Resolution

This involves determining the number and characteristic parameters of peaks in a cluster. Computational approaches are complementary to physical ones such as shimming spectrometers or changing columns. In some cases, such as NMR, the number and relative intensities of peaks in a cluster may have a precise physical meaning (e.g., a triplet implies a specific coupling mechanism), whereas in chromatography, each new component corresponds to another compound and the shapes cannot easily be predicted mathematically.

### Integration

This is one of the hardest problems in signal processing. It relates to the areas of peaks, which, in turn, can be calibrated to the concentrations and/or relative amounts of each component. The commonest

approach is to sum the intensities at each sampling point.

Many common methods for determining relative intensities result in unreliable estimates of concentrations. For example, one of the commonest difficulties relates to using signal heights rather than areas to measure intensities, because it is easier to develop signal processing software to measure heights. Only if peaks are well resolved and have identical shapes can heights be employed as a reliable approach for determining relative concentrations.

## Handling the Signal

There are several common procedures for handling digital signals as described below.

### Baseline Correction

This is a common requirement. The first step is to determine regions of baseline, as indicated in **Figure 7**; this is often done graphically, but there are also automated computational algorithms available. Next a mathematical model is fitted to these regions, which could be simple, such as an average, or more complex, such as a polynomial. This model is then subtracted from the entire dataset. In areas such as coupled chromatography it is common to fit a baseline model to the chromatogram at each wavelength or  $m/z$  value separately.

### Interpolation

Sometimes it is necessary to obtain regularly spaced data. The raw data may not be sampled regularly, for example, when studying the evolution of a process with time. In other cases it may be important to align two different types of information; an example is when using two different detectors in coupled chromatography, each of which is sampled at a slightly different rate. There are a number of methods but a simple one is as follows:

1. Establish a desired sampling interval, which is often approximately equal to the average sampling interval for the overall dataset.

2. For each interpolated sampling time, see if a real sample is obtained at exactly that time, and if so, keep it.

3. If not, take the real samples immediately before and after the desired interpolated sampling time and obtain a weighted average of the signal, the closer it is to the desired interpolated sampling time the higher the weight, so if we have signals at time 8 and 11, the interpolated signal at time 10 is two-thirds of that at time 11 and one-third of that at time 8.

### Variable Reduction

There are a large number of methods for variable reduction. In many cases the regions between peaks are of no interest. Retaining this information means storing large files; it also means that noise can dominate the analysis. Often we search for parts of the spectrum or chromatogram where there appear to be signals and reject the remainder, a common example in mid-IR spectroscopy where there are only certain discrete peaks that are interesting.

### Derivatives

Derivative spectroscopy is quite common. This involves calculating mathematical derivatives of the spectrum. The advantage is that close peaks can be resolved out. Derivatives rely on the assumption that pure peaks have only one maximum. The summation of two peaks results in a 'turning point' that can be resolved out. However, the problem is that noise is enhanced, so derivatives are often combined with smoothing functions. These methods are described in more detail in this encyclopedia.

### Resolution and Signal-to-Noise Enhancement

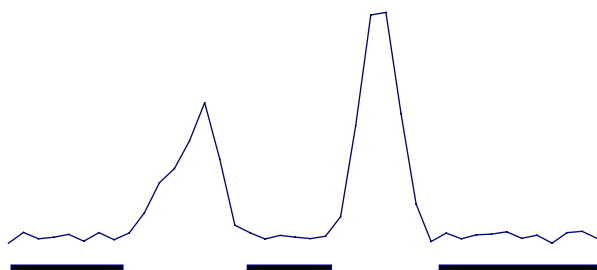
These techniques are described elsewhere in this encyclopedia.

## Types of Signal

There are several different types of signals commonly encountered.

### Fourier Transform Data

Such data may arise from either pulsed (e.g., NMR) or interferometric (e.g., IR) instrumentation. Data are recorded directly in the time domain. Normally, the data consist of a sum of decaying sine waves, superimposed on noise. Fourier transformation converts these into a frequency domain spectrum. For each component, the frequency of the sine wave corresponds to the position in the spectrum, the initial



**Figure 7** Typical signals with three baseline regions indicated.

intensity to the integral, the decay pattern to the lineshape (commonly an exponential decay corresponds to a Lorentzian and a Gaussian decay to a Gaussian peakshape), and the decay rate to the peak width (the faster the decay the broader the peak).

The faster a time series is sampled, the wider the spectral width in the resultant transform. The longer the time series is sampled, the greater the digital resolution in the spectrum. These parameters must be taken into account when considering the accuracy of integration. For example, a typical peak may be only a few data points wide in FT-NMR, there can be substantial distortions in peak area when a peak is characterized by only two or three data points at half-height.

The advantage of using Fourier transform spectroscopy is that data are acquired very much faster than via conventional, continuous wave, methods. This means a better signal-to-noise ratio can be achieved for an identical amount of compound in the same period of time. The increase in speed is  $\sim 50$ -fold.

#### Hadamard Transform Data

Hadamard spectroscopy is rare, but is an alternative to Fourier transform spectroscopy for improving signal-to-noise ratios. In a Hadamard spectrum, different frequencies of the raw spectrum are blocked by a 'mask'. The overall sum of intensities at the frequencies that are not blocked is then recorded. The mask is systematically changed, until several such experiments have been performed. It is possible to derive the individual spectrum from a series of experiments. The advantage is that the signal-to-noise ratio of each experiment is higher than for any individual wavelength. However, this approach has not been widely used, primarily because of the difficulty of constructing rapidly changing masks.

#### Correlation Spectroscopy and Chromatography

These approaches are quite rare. Methods for time series analysis are required to analyze the data. A signal, consisting of a time series, can be correlated with another. The resultant correlogram consists of the correlation coefficient as one series is lagged

against the other in time. Each successive lag position uses one less data point for calculation of correlation coefficients. There is a large literature on time series analysis, and the choice of number of lag positions in the correlogram is of substantial interest. The more the lag positions, the lower the signal-to-noise ratio but the greater the digital resolution.

Once the correlogram is computed, it is usually Fourier transformed. It is also possible to filter the signal at various different stages in the analysis.

#### Direct Recording

The most common way of recording spectra or chromatograms is by directly changing a sequential parameter such as wavelength or time. The result is directly interpretable without further transformation. UV/vis spectroscopy and chromatography are common examples. Considerations should, however, be given to sampling frequency and slit width as discussed above.

See also: **Chemometrics and Statistics:** Spectral Deconvolution and Filtering.

#### Further Reading

- Brereton RG (2003) *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, chapter 3. Chichester: Wiley.
- Chatfield C (1989) *Analysis of Time Series: An Introduction*. London: Chapman and Hall.
- Gans P (1992) *Data Fitting in the Chemical Sciences: By the Method of Least Squares*. Chichester: Wiley.
- Lynn PA and Fuerst W (1998) *Introductory Digital Signal Processing with Computer Applications*, 2nd edn. Chichester: Wiley.
- Otto M (1998) *Chemometrics: Statistics and Computer Applications in Analytical Chemistry*, chapter 3. Weinheim: Wiley-VCH.
- Vandeginste BGM, Massart DL, Buydens LMC, *et al.* (1998) *Handbook of Chemometrics and Qualimetrics Part B*, chapter 40. Amsterdam: Elsevier.
- Wegscheider W (1998). In: Kellner R, Mermet J-M, Otto M, and Widmer HM (eds.) *Analytical Chemistry*, section 12.3. Weinheim: Wiley-VCH.

## Spectral Deconvolution and Filtering

R G Brereton, University of Bristol, Bristol, UK

### Introduction

Two main classes of method, univariate and multivariate, are commonly employed. Univariate methods

intensity to the integral, the decay pattern to the lineshape (commonly an exponential decay corresponds to a Lorentzian and a Gaussian decay to a Gaussian peakshape), and the decay rate to the peak width (the faster the decay the broader the peak).

The faster a time series is sampled, the wider the spectral width in the resultant transform. The longer the time series is sampled, the greater the digital resolution in the spectrum. These parameters must be taken into account when considering the accuracy of integration. For example, a typical peak may be only a few data points wide in FT-NMR, there can be substantial distortions in peak area when a peak is characterized by only two or three data points at half-height.

The advantage of using Fourier transform spectroscopy is that data are acquired very much faster than via conventional, continuous wave, methods. This means a better signal-to-noise ratio can be achieved for an identical amount of compound in the same period of time. The increase in speed is  $\sim 50$ -fold.

#### Hadamard Transform Data

Hadamard spectroscopy is rare, but is an alternative to Fourier transform spectroscopy for improving signal-to-noise ratios. In a Hadamard spectrum, different frequencies of the raw spectrum are blocked by a 'mask'. The overall sum of intensities at the frequencies that are not blocked is then recorded. The mask is systematically changed, until several such experiments have been performed. It is possible to derive the individual spectrum from a series of experiments. The advantage is that the signal-to-noise ratio of each experiment is higher than for any individual wavelength. However, this approach has not been widely used, primarily because of the difficulty of constructing rapidly changing masks.

#### Correlation Spectroscopy and Chromatography

These approaches are quite rare. Methods for time series analysis are required to analyze the data. A signal, consisting of a time series, can be correlated with another. The resultant correlogram consists of the correlation coefficient as one series is lagged

against the other in time. Each successive lag position uses one less data point for calculation of correlation coefficients. There is a large literature on time series analysis, and the choice of number of lag positions in the correlogram is of substantial interest. The more the lag positions, the lower the signal-to-noise ratio but the greater the digital resolution.

Once the correlogram is computed, it is usually Fourier transformed. It is also possible to filter the signal at various different stages in the analysis.

#### Direct Recording

The most common way of recording spectra or chromatograms is by directly changing a sequential parameter such as wavelength or time. The result is directly interpretable without further transformation. UV/vis spectroscopy and chromatography are common examples. Considerations should, however, be given to sampling frequency and slit width as discussed above.

See also: **Chemometrics and Statistics:** Spectral Deconvolution and Filtering.

#### Further Reading

- Brereton RG (2003) *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, chapter 3. Chichester: Wiley.
- Chatfield C (1989) *Analysis of Time Series: An Introduction*. London: Chapman and Hall.
- Gans P (1992) *Data Fitting in the Chemical Sciences: By the Method of Least Squares*. Chichester: Wiley.
- Lynn PA and Fuerst W (1998) *Introductory Digital Signal Processing with Computer Applications*, 2nd edn. Chichester: Wiley.
- Otto M (1998) *Chemometrics: Statistics and Computer Applications in Analytical Chemistry*, chapter 3. Weinheim: Wiley-VCH.
- Vandeginste BGM, Massart DL, Buydens LMC, *et al.* (1998) *Handbook of Chemometrics and Qualimetrics Part B*, chapter 40. Amsterdam: Elsevier.
- Wegscheider W (1998). In: Kellner R, Mermet J-M, Otto M, and Widmer HM (eds.) *Analytical Chemistry*, section 12.3. Weinheim: Wiley-VCH.

## Spectral Deconvolution and Filtering

R G Brereton, University of Bristol, Bristol, UK

### Introduction

Two main classes of method, univariate and multivariate, are commonly employed. Univariate methods

are as follows. Curve fitting is used when quite a good model of the individual peaks is known. Moving averages, Savitsky–Golay filters, and derivatives are simple approaches for enhancing signal-to-noise ratios and resolution of spectra. In the Fourier (time) domain, most filters involve multiplying the data by a function that can be related to moving averages via the convolution theorem. Specialized windows such as correlograms are used for time series. Finally, nonlinear approaches such as maximum entropy show great promise. Multivariate data occur in hyphenated techniques or when a series of spectra is recorded on a number of samples. Multivariate methods (sometimes called factor analysis) are useful when there is more than one measurement, e.g., hyphenated chromatography or fluorescence excitation-emission spectroscopy.

There are numerous aims of deconvolution and filtering. The simplest is to determine the number of components in a mixture. In nuclear magnetic resonance (NMR), determining how many peaks are in a cluster can tell us about the coupling pattern, and gives us diagnosis about the structure of the underlying molecular species. In pyrolysis mass spectrometry (MS), how many significant components are in a series of mixture spectra? In some areas of chromatography, detection of small peaks, e.g., in impurity monitoring, is important. Most of these questions require answers in the form of graphs rather than numerical values.

Once this qualitative information is obtained, there are a number of more quantitative questions. One of the simplest is to determine the centers or positions of each peak. In molecular spectroscopy, the position may be highly diagnostic of a specific compound. A more sophisticated problem is to pull out the spectra of each component, and in multivariate methods, to determine the chromatographic elution time, pH, or concentration profile of each component of a mixture. The individual spectra can then be compared to a library or manually interpreted. In pH equilibrium titrations, the profiles may allow determination of  $pK_a$ s.

The most difficult problems relate to quantification, normally involving integrating individual peaks or spectra arising from specific compounds. Some methods for deconvolution distort these, for example, many methods for Fourier transform resolution enhancement are dependent on peakwidths of individual components. If a spectrum consists of peaks of different widths, each peak will be enhanced to a different extent, making relative integrals meaningless. Other approaches, such as in chemometric multivariate resolution, should preserve relative integrals.

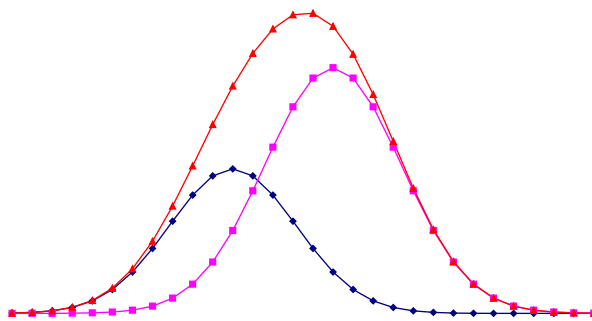
The analytical scientist has a very large battery of methods to choose from.

## Univariate Methods

### Curve Fitting

One of the conceptually simplest methods is curve fitting. The number and parameters of peaks that make up a cluster can be determined by fitting a model. Normally, this consists of a sum of well-defined peakshapes such as Gaussians and Lorentzians as illustrated in **Figure 1**. Normally, each peak is characterized by a center, a width, and a height. If there are three peaks in a cluster, then nine parameters are required to fit the data. Most algorithms are based on least squares fitting, but some newer algorithms involve imposing constraints, such as nonnegativity and unimodality.

A problem with these approaches is that the nature of the underlying data should normally be known in advance. There are many difficulties with nonlinear (e.g., Gaussian) curve fitting routines and it is sometimes hard to find a global minimum. Most computational algorithms require initial parameters, often guesses of the peakshapes, and some information about the method for iteration. In many forms of molecular spectroscopy such as NMR, where peakshapes are known with a high degree of confidence from the underlying spin physics, curve fitting routines are extremely effective. In cases such as chromatography where peakshapes are far less well established, curve fitting can be dangerous. Sometimes if only one or two parameters are varied (constrained minimization) such routines can be useful: for example, if it is desired to follow the rate of a two-component reaction by ultraviolet/visible spectroscopy, and the spectra of the pure components are well known, it is only necessary to vary a single parameter, corresponding to the proportion of each component in a mixture.



**Figure 1** Curve fitting is employed to fit the main peak into two components.



Curve fitting usually aims to represent the profile of each component separately, whereas most other methods represent the profiles in the same graph, but aims to distinguish different peaks both from each other and from the noise.

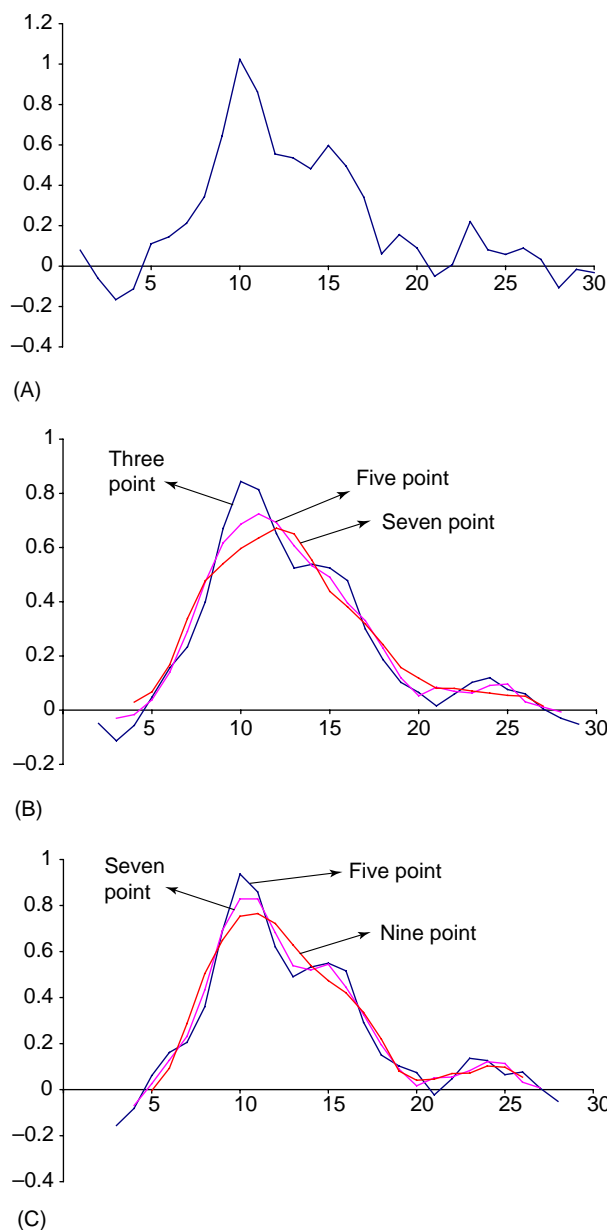
### Direct Methods: Linear Filters

**Moving averages** There are many direct methods that are used to enhance the resolution and increase the signal-to-noise ratio of peaks. One of the simplest is a moving average. A three-point moving average replaces a reading at a given point in time by the average of itself and the previous and successive points or

$$x_{i,\text{new}} = \sum_{j=-1}^1 x_{i+j} / 3$$

where  $x_i$  is the signal at point  $i$ . Noise tends to vary randomly, whereas signals vary more systematically, and are broader. Hence, the noise is smoothed away resulting in an increased signal-to-noise ratio. The number of points in the moving average filter can be increased, but is generally an odd number, often called the 'window' size. The optimum size depends primarily on the signal width. The broader the signal, the bigger the optimum size; however, as the size becomes comparable to the peakwidth in data points, peaks start to be smoothed away and so resolution and intensity is lost. **Figure 2A** illustrates a cluster of peaks and **Figure 2B** the result of three-, five-, and seven-point moving average filters. Above three points, it is evident that the data are smoothed away quite considerably. Note that it is often possible to record a spectrum or chromatogram using higher digital resolution and this may allow more powerful filters to be applied.

**Higher-order filters including Savitsky–Golay filters** It is not necessary to restrict filters to simple local averages. Often, signals are better approximated by polynomials, such as quadratics or cubics. This is especially important in the center of peaks which will inevitably be smoothed away using simple averages. Polynomial filters work by fitting a local polynomial, for example, a five-point quadratic filter will take five successive points, fit these to a quadratic model, and replace the central point by the estimated value from the quadratic model. This filter is moved along a spectrum or chromatogram. For example, the first window might involve recalculating the value of point 3 in the spectrum using points 1–5, the second window recalculates the value of point 4 using points 2–6, and so on. The extreme points of the spectrum or chromatogram are removed.



**Figure 2** (A) Raw data together with the results of (B) moving average and (C) Savitsky–Golay filters.

Because it is computationally intense to calculate fit polynomials to each point, Savitsky and Golay proposed an alternative and simpler computational method in which a new data point is expressed as

$$x_{i,\text{new}} = \sum_{j=-p}^p c_j x_{i+j}$$

where the size of the filter window is  $2p + 1$  and coefficient  $c$  can be obtained from standard tables. Different coefficients relate to window size and the order of the model (e.g., quadratic, cubic). Most

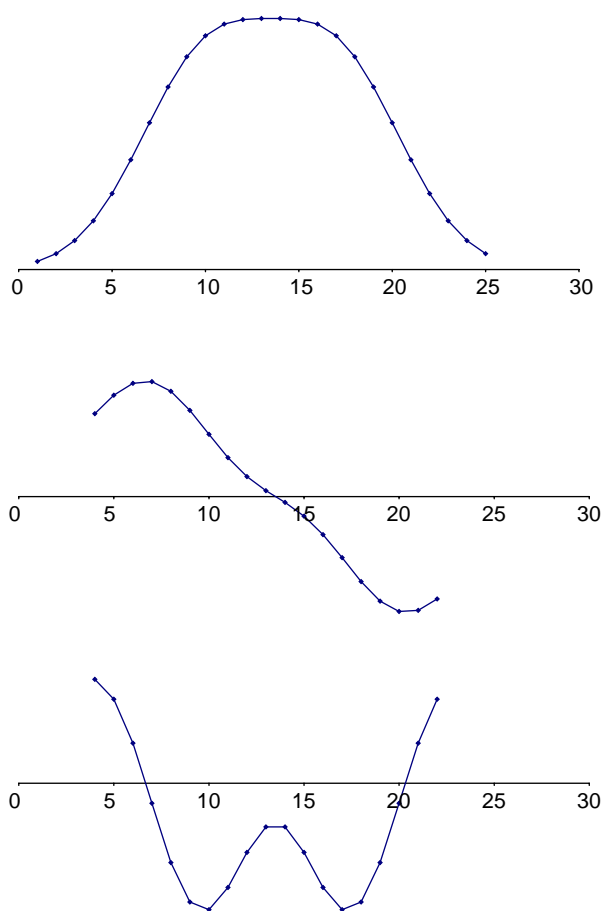
**Table 1** Savitsky–Golay filters for smoothing

Window size, $j$	Quadratic/cubic model			Quartic/quintic model	
	5	7	9	7	9
–4			–21		15
–3		–2	14	5	–55
–2	–3	3	39	–30	30
–1	12	6	54	75	135
0	17	7	59	131	179
1	12	6	54	75	135
2	–3	3	39	–30	30
3		–2	14	5	–55
4			–21		15
Normalization constant	35	21	231	231	429

filters can, in fact, be expressed in these terms and so are often called linear filters, because the new data point is expressed as a linear combination of the original data, even if the order of polynomial is greater than 1. For moving averages the value of  $c$  is  $1/(2p + 1)$ , for Savitsky–Golay smoothing functions the values are presented in **Table 1**. To calculate the coefficients, simply divide the values in the table by the normalization constant. The results of five-, seven-, and nine-point quadratic/cubic filters on the data of **Figure 2A** are presented in **Figure 2C**; the seven-point filter is about optimal, reducing noise but still preserving resolution and making it fairly obvious that this is a two-peak cluster. A nine-point filter is too extreme. Often, it is necessary to experiment using several different functions before a satisfactory answer is obtained.

There are several other linear filters such as Hanning and Hamming windows which can be found in specialist spectroscopic texts. Often, for historic reasons, users of specific spectroscopic techniques have employed particular named filters.

**Derivatives** Derivatives are frequently used to enhance resolution. The principle is that pure peaks are generally unimodal, so contain only one maximum. Peak clusters arising from several underlying components, if not completely resolved, contain inflexion points. Derivatives resolve these out, as illustrated in **Figure 3**. These also allow peak maxima and widths to be measured more easily. For example, the center of a peak has a first derivative of 0, so crosses the baseline in the derivative spectrum. It is often easier to measure a peak crossing the baseline than estimate the exact center of a flat peak. In some forms of spectroscopy such as electron spin resonance, spectra are conventionally recorded directly in the derivative mode, which allows good resolution of hyperfine

**Figure 3** Two closely overlapping peaks (top) together with their first (middle) and second (bottom) derivatives.

coupling constants. Alternatively derivatives may be calculated computationally.

There are various problems with derivatives, a major one is that noise is amplified. In order to overcome this, it is usual to smooth derivative



**Table 2** Savitsky–Golay filters for calculation of derivatives

Window size, <i>j</i>	5	7	9	5	7	9
<i>First derivatives</i>	Quadratic model			Cubic/quartic model		
–4			–4			86
–3		–3	–3		22	–142
–2	–2	–2	–2	1	–67	–193
–1	–1	–1	–1	–8	–58	–126
0	0	0	0	0	0	0
1	1	1	1	8	58	126
2	2	2	2	1	67	193
3		3	3		–22	142
4			4			–86
Normalization constant	10	28	60	12	252	1 188
<i>Second derivatives</i>	Quadratic/cubic model			Quartic /quintic model		
–4			28			–4 158
–3		5	7		–117	12 243
–2	2	0	–8	–3	603	4 983
–1	–1	–3	–17	48	–171	–6 963
0	–2	–4	–20	–90	–630	–12 210
1	–1	–3	–17	48	–171	–6 963
2	2	0	–8	–3	603	4 983
3		5	7		–117	12 243
4			28			–4 158
Normalization constant	7	42	462	3	99	4 719

spectra. This can be performed by using linear coefficients as proposed by Savitsky and Golay in a similar manner to the direct smoothing. These coefficients are presented in Table 2.

### Time Domain (Fourier) Methods

Sometimes, it is useful to perform deconvolution in the time domain. In this article, time domain is used to refer to a time series and frequency domain to the directly interpretable spectrum/chromatogram. Fourier transformation is employed to transform data from the time (Fourier) domain to the frequency (spectral) domain. It is important to note that all spectra have a time domain associated with them even if directly recorded in the frequency domain.

It can be shown, for example, that a damped oscillator described by the equation

$$f(t) = A \cos(\omega t) e^{-\lambda t}$$

transforms into a Lorentzian peakshape. The value of  $\lambda$  is a time constant and influences the peakwidth. The greater its value the faster the decay rate and the broader the peak. Hence, multiplying the decay curve by a positive exponential decreases the decay rate, and so sharpens peaks improving resolution. The problem is that this process also increases the noise. Therefore, it is more usual to employ a filter containing two adjustable parameters, one to increase resolution and the other to decrease noise, a

common one being a double exponential as

$$g(t) = e^{+\kappa t - \nu t^2}$$

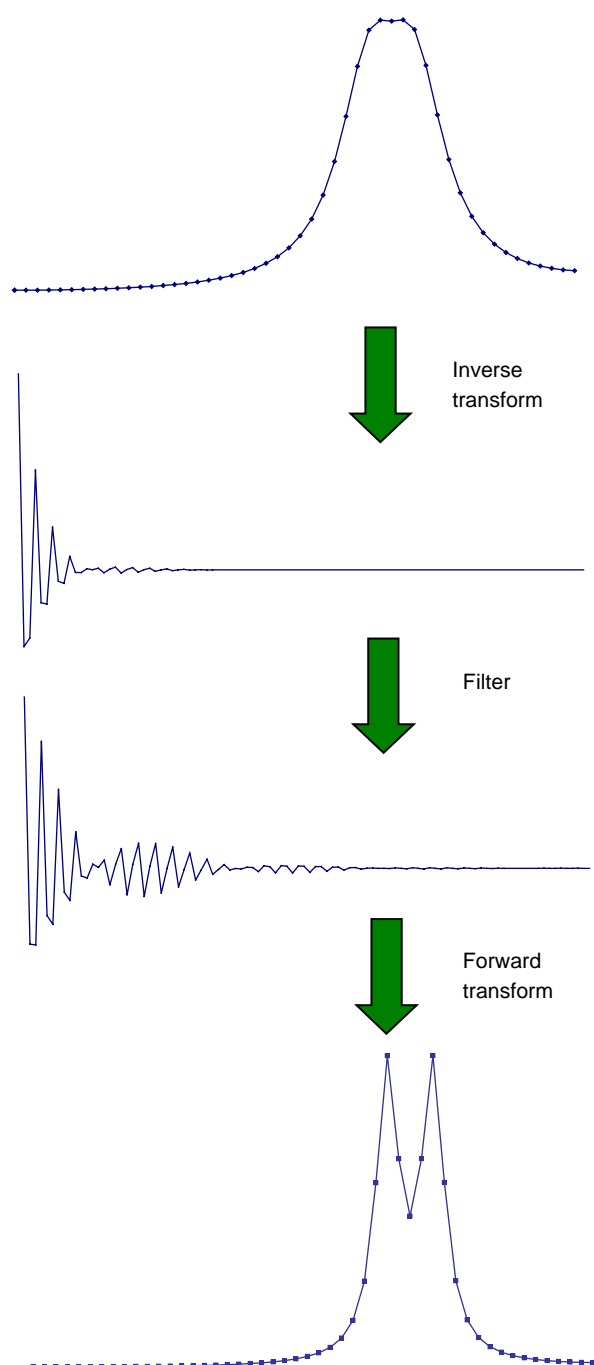
which is used to multiply the original time domain signal. The values of  $\kappa$  and  $\nu$  have to be adjusted and depend on noise levels and peakwidths. There are a wide variety of alternative filter functions in the literature but most depend on two adjustable parameters.

After applying filter functions, it is normal to Fourier transform the data. Sometimes for data recorded directly in the frequency domain, it is possible to perform Fourier self-deconvolution. This involves inverse transforming the spectrum or chromatogram back into a time domain, applying the filters, and then forward transforming to the frequency domain as illustrated in Figure 4.

### Convolution Theorem

Both methods for time domain (Fourier) and frequency domain (direct) filters are equivalent and are related by the convolution theorem, and both have similar aims, to improve the quality of spectroscopic or chromatographic or time series data. Two functions,  $f$  and  $g$ , are said to be convoluted to give  $h$ , if

$$h_i = \sum_{j=-p}^{i=p} f_j g_{i+j}$$



**Figure 4** Principles of Fourier self-deconvolution.

Convolution involves moving a window or digital filter function (such as a Savitsky–Golay or moving average) along a series of data such as a spectrum, multiplying the data by that function at each successive data point. A three-point moving average involves multiplying each set of three points in a spectrum by a function containing the values (1/3, 1/3, 1/3), and the spectrum is said to be convoluted by the moving average filter function. Filtering a time

series, using Fourier time domain filters, however, involves multiplying the entire time series by a single function, so that

$$H_i = F_i \cdot G_i$$

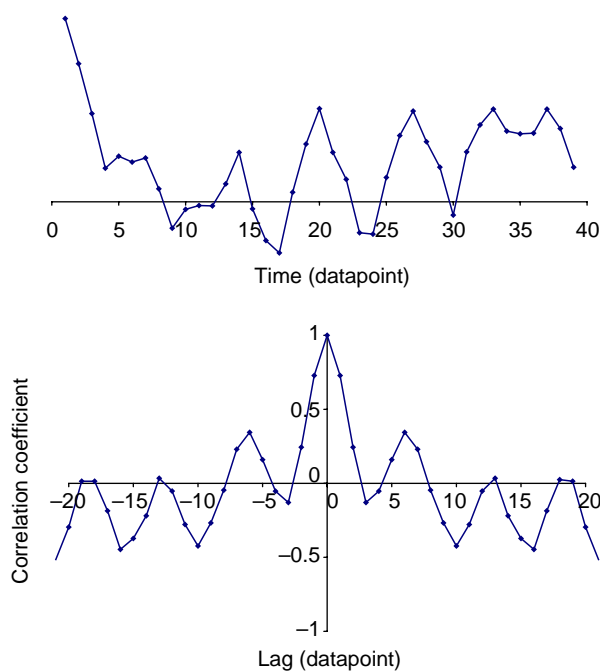
The convolution theorem states that  $f$ ,  $g$ , and  $h$  are Fourier transforms of  $F$ ,  $G$ , and  $H$ . Hence, linear filters as applied directly to spectroscopic data have their equivalence as Fourier filters in the time domain. In other words, convolution in one domain is equivalent to multiplication in the other, which approach is dependent largely on computational complexity and convenience. For example, both moving averages and exponential Fourier filters are easy to apply, one applied direct to the frequency spectrum and the other to the raw time series. Convoluting a spectrum with the Fourier transform of an exponential decay is a difficult procedure and so the choice of domain is made according to how easy the calculations are.

### Time Series Methods

In a small number of areas such as correlation spectroscopy, time series analysis is necessary. These methods have been used for many years, especially in economics, geology, and engineering, where noisy time series are recorded. This involves computing what is called a correlogram. The way this is performed is by taking a time series, and shifting it against itself by an amount called a lag, then the correlation coefficient is calculated between the original time series and the lagged series. For example, if a time series is recorded over 50 data points and lagged by 5 data points, then points 1–45 of the original time series is compared to points 6–50 of the lagged series (the last five points are lost). Lags can be both positive and negative.

There is a limit to the number of lag positions, because if the lag is too much, there are not many points remaining, a good guide is to employ around half the data points from the original time series. A graph of the correlation coefficient versus lag is called an autocorrelogram and is symmetric around a lag of 0. This is illustrated in **Figure 5** for a typical time series. The correlogram can be further manipulated by Fourier transformation, smoothing, and filtering if necessary. The aim is primarily to emphasize cyclical signals from others. The technique is especially useful when there is correlated noise.

It is also possible to compare two different time series, a similar procedure is used to compute a cross-correlogram where each time series is lagged against one another. This correlogram is no longer symmetric. It may be useful if, for example, we record a



**Figure 5** A time series (top) together with its autocorrelogram.

reference signal from a particular compound and want to see whether it is present in an experimental dataset. The time series from the pure component is cross-correlated with the experimental one.

### Nonlinear Deconvolution, Including Maximum Entropy

All the methods for filtering described above are linear; that is, the filtered data can be expressed as linear sums of the original data, multiplied by coefficients. However, one of the problems with linear methods is that there are no restrictions to physically meaningless solutions. In many cases it may be known, for example, that negative signals are meaningless. It may be useful to include these in the algorithm, meaning that linear methods are no longer adequate.

Maximum entropy is probably the most well-known nonlinear method. Consider tossing two unbiased coins. What is the most likely distribution? Is it two tails, one head and one tail, or two heads? The middle distribution is twice as likely and is said to have twice the degeneracy of the other distributions. The problem of the toss of two coins is one in which there are two states (heads and tails) and three levels (0, 1, or 2 coins in each state). The number of coins expected in each of the three levels can be converted to a probability, so the probability of level 0 is 0.25, level 1 is 0.5, and level 2 is 0.25.

A spectrum can also be considered a probability distribution with  $I$  states where each represents a

sampling point (e.g., wavenumber) and an infinite number of levels (subject to a sufficiently high analog to digital conversion). It can be shown that for a sufficiently large number of states relative degeneracy is proportional to entropy, defined by

$$S = - \sum_{i=1}^I p_i \log(p_i)$$

where  $p_i$  is the probability of absorption at point  $i$  in the spectrum. In many cases the probability is defined as the spectral intensity at a given point divided by the sum of all other intensities in the spectrum. Note that entropy requires positive values. For any possible spectrum, the greater the entropy the more likely the spectrum. In the absence of other information, a flat spectrum is one with maximum entropy.

Maximum entropy deconvolution, however, is normally iteratively combined with other methods. For example, least squares curve fitting attempts to create structure in a spectrum. Maximum entropy picks solutions with least structure: the reason is that in the absence of external evidence, there is no reason to assume structure in the reconstructed spectrum: the 'flat map' is the most likely answer. Most algorithms try to decrease mean square errors iteratively. A target residual is selected, according to the desired speed of computational convergence, and the solution with maximum entropy is chosen. Then, the residual is decreased and, again, a maximum entropy solution is found, until convergence.

Maximum entropy has been applied to a large number of techniques such as NMR, Raman, and MS. The method can be very powerful, but on the whole is best suited when there is poor knowledge of the data, or the dataset is somewhat incomplete, e.g., noisy data. If a large amount of information is already known, such as exact lineshapes and there are high signal-to-noise ratios, other more targeted methods are probably more effective.

There are several other nonlinear methods applied in different spectroscopies, the book edited by Jansson provides a broad review of nonlinear deconvolution.

### Multivariate Methods

There has been especially large growth in methods for multiway resolution over the past decade. These relate to signals where there are two or more dimensions, such as in the study of pH equilibria by optical spectroscopy, where one dimension relates to pH and the other wavelength, or diode array high-performance liquid chromatography (HPLC) where

one dimension relates to elution time and the other wavelength. Normally, one dimension (e.g., pH or elution time) is studied in sequence and the aim is to determine the profiles of individual chemical species as a function of this dimension together with their spectral characteristics.

### Principal Component and Factor Analysis

Consider the case of coupled chromatography, such as HPLC–DAD. For a simple chromatogram, the underlying dataset is the sum of responses for each significant compound in the data, which are characterized by (1) an elution profile and (2) a spectrum. This is often written in matrix terms as

$$X = C \cdot S + E$$

where  $X$  is the original data matrix or coupled chromatogram,  $C$  is a matrix consisting of the elution profiles of each compound,  $S$  is a matrix consisting of the spectra of each compound, and  $E$  is an error matrix. An important aim may be to predict  $C$  and  $S$  from  $X$ , allowing the resolution of the original multivariate data matrix into each constituent. Ideally, the predicted spectra and chromatographic elution profiles are close to the true ones, but it is important to realize that we can never directly or perfectly observe the underlying data. There will always be measurement error even in practical spectroscopy. Chromatographic peaks may be partially overlapping or even embedded, meaning that chemometric methods will help resolve the chromatogram into individual components.

One aim of chemometrics is to obtain these predictions after first treating the chromatogram as a multivariate data matrix, and then performing principal component analysis (PCA). Each compound in the mixture is a ‘chemical’ factor with its associated spectra and elution profile, which can be related to principal components, or ‘abstract’ factors, by a mathematical transformation.

PCA results in an abstract mathematical transformation of the original data matrix, which takes the form

$$X = T \cdot P + E$$

where  $T$  represents the scores, and has as many rows as the original data matrix,  $P$  the loadings and have as many columns as the original data matrix, and the number of columns in the matrix  $T$  equals the number of rows in the matrix  $P$ . It is possible to calculate scores and loadings matrices as large as desired, provided the ‘common’ dimension is no larger than the smallest dimension of the original data matrix, and corresponds to the number of

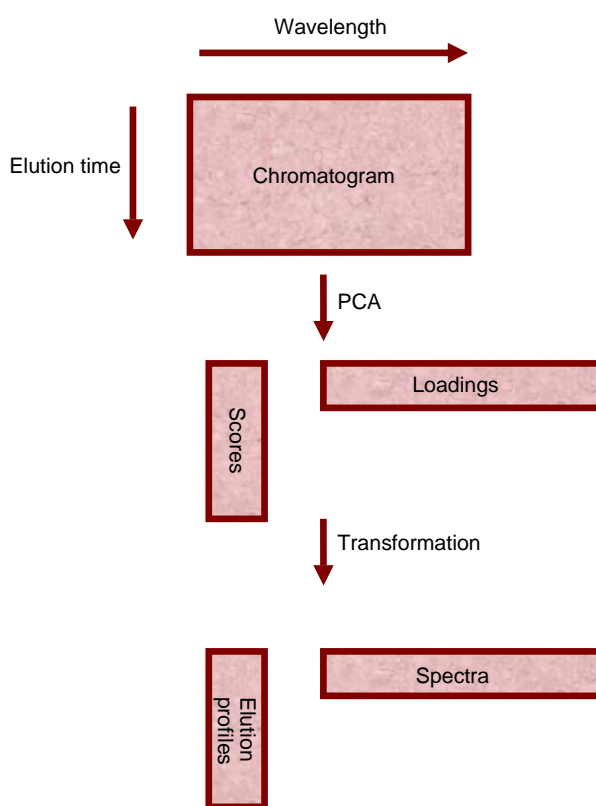


Figure 6 Difference between PCA and factor analysis.

principal components that are calculated. However, ideally the number of columns in  $P$  equals the number of significant components or compounds in the chromatogram.

Principal components regression relates these abstract factors to chemically meaningful factors by finding a matrix  $R$  to give

$$X = C \cdot S + E = T \cdot R \cdot R^{-1} \cdot P + E$$

This procedure is called transformation, rotation, or factor analysis depending on the author and is illustrated in Figure 6. The majority of methods for multivariate resolution differ in how  $R$  is determined.

### Target Factor Analysis

Target factor analysis (TFA) tries to see how closely the predicted pure data relate to targets, for example, a library of known spectra. Can we find a transformation matrix so that the predicted spectra fit closely to known spectra (the targets)? Another popular method is iterative transform factor analysis where the targets are defined iteratively. For example, the original targets may be guesses of the pure component spectra obtained by choosing regions of the data that are expected to most represent pure

components. This model is refined iteratively until a good fit is obtained.

### Eigenvalue-Based Window Methods

A separate group of methods involve calculating the eigenvalues of a multivariate matrix within a window. Each principal component has a size, which is often given by an eigenvalue; the larger, the more it is significant. So if a section of a multiway dataset is subjected to PCA, and there are three compounds in this section, there are likely to be three significant eigenvalues. The size of the eigenvalues can be calculated as the window is changed in size or position.

There are two main alternatives, using a fixed sized window, which is moved along the data, or using an expanding window. For the fixed sized window, the number of significant eigenvalues increases and decreases according to the number of components in the data, whereas for the expanding window it will always increase as the window evolves. In the latter case it is normal to calculate eigenvalues for both forward and backward windows, the latter starting at the end of the data. The two types of method are illustrated in **Figure 7**. The resultant eigenvalues are generally plotted on a logarithmic scale against time.

The result is used to determine regions of maximum purity. These approximate to the elution profiles or spectra (in coupled chromatography) of each component in a mixture that can then be employed in factor analysis as some information on each pure component is known. Sometimes there are embedded peaks, for which there is no pure (or selective or composition 1) region. The eigenvalue plots can still provide valuable information as to where each component elutes but sometimes it is hard to obtain unique mathematical solutions to the determination of information on each compound.

However, using constraints such as non-negativity and unimodality can help.

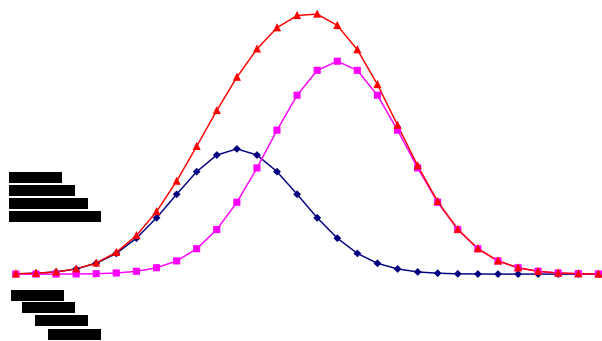
### Purity-Based Approaches

A separate set of approaches relates to determine selective or pure variables. These are variables that most characterize each individual component in a mixture. For example, in the GC-MS of mixtures there is likely to be a specific  $m/z$  value most characteristic of each component. In HPLC-DAD, we look for selective elution times for each compound.

There are a variety of named approaches such as SIMPLISMA and the Orthogonal Projection Approach. In addition, it is possible to determine derivatives between successive scaled spectra, regions of high purity will result in low values, and so minima in the scaled derivatives can be employed to pinpoint pure regions. Correlation coefficients are another effective alternative, the higher the correlation between two successive spectra in a chromatogram, the more likely two successive points correspond to the same spectrum and so correspond to one pure compound. In regions of coelution, the derivatives will be high and the correlation coefficient low, because the nature of the spectrum will be changing.

Once pure variables have been determined, then it is possible to perform factor analysis as described above, using these pure values either as guesses of spectra (e.g., selective regions of a chromatogram) that can be used to determine elution profiles in overlapping regions, or as guesses of profiles (e.g.,  $m/z$  values of characteristic ions in LC-MS) that can be used to determine spectra of each component. Sometimes constraints such as non-negativity are included in the calculations.

See also: **Chemometrics and Statistics:** Signal Processing.



**Figure 7** Illustration of expanding (top) and fixed (bottom) windows.

### Further Reading

- Brereton RG (2003) *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, chapters 3 and 6. Chichester: Wiley.
- Chatfield C (1989) *Analysis of Time Series: An Introduction*. London: Chapman and Hall.
- Gans P (1992) *Data Fitting in the Chemical Sciences: By the Method of Least Squares*. Chichester: Wiley.
- Jansson PA (ed.) (1984) *Deconvolution: With Applications in Spectroscopy*. New York: Academic Press.
- Lynn PA and Fuerst W (1998) *Introductory Digital Signal Processing with Computer Applications*, 2nd edn. Chichester: Wiley.



- Malinowski ER (2002) *Factor Analysis in Chemistry*, 3rd edn. Chichester: Wiley.
- Otto M (1998) *Chemometrics: Statistics and Computer Applications in Analytical Chemistry*, chapter 3. Weinheim: Wiley-VCH.
- Savitsky A and Golay MJE (1964) Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* 36: 1627–1639.
- Sibisi S, Skilling J, Brereton RG, Laue ED, and Staunton J (1984) Maximum entropy signal processing in practical NMR spectroscopy. *Nature* 309: 801–802.
- Vandeginste BGM, Massart DL, Buydens LMC, *et al.* (1998) *Handbook of Chemometrics and Qualimetrics Part B*, chapters 34 and 40. Amsterdam: Elsevier.
- Wegscheider W (1998) In: Kellner R, Mermet J-M, Otto M, and Widmer HM (eds.) *Analytical Chemistry*, section 12.3. Weinheim: Wiley-VCH.

## CHIROPTICAL ANALYSIS

**H-G Kuball**, University of Kaiserslautern, Kaiserslautern, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

‘Chiroptical analysis’ is the application of chiroptical spectroscopy – including absorption, emission, and optical refraction measurements – in order to obtain information about chiral molecules and phases. Lord Kelvin has defined chiral objects in his famous ‘Baltimore Lectures 1889’ by: “I call any geometrical figure or group of points ‘chiral’, and say it has ‘chirality’, if its image in a plane mirror, ideally realized, cannot be brought – by rotation and translation of the geometrical figure or the group of points – into coincidence with itself.” Examples include molecules and chiral anisotropic phases with their long-range orientational and long-range positional order (for a classification by a length scale, four levels of chirality can be introduced. First level: atoms (weak interaction); second level: molecules (geometry); third level: suprastructural chirality (long-range orientational and positional order of atoms/molecules); fourth level: macroscopic objects like helical pillars, staircases, etc.) without symmetry of second kind, i.e., no rotation–reflection symmetry (rotation–reflection axes  $S_n$ ,  $n = 1, 2, \dots$ ). These chiral objects do exist always in two different forms, one being the mirror image of the other (enantiomers and enantiomorphous solids). The introduction of a reference system, like that of Cahn, Ingold, and Prelog (CIP), allows dividing the world of chiral molecules into two classes, the ‘S or R (M or P) world’. For chiral anisotropic phases no universally valid reference systems exist. Exceptions are the left (M) or right

(P) handed helical long-range orientational order of cholesteric and smectic  $C^*$  phase and the left (M) or right (P) handed screw axes of crystals.

In the last decades chiroptical spectroscopy has been mostly applied to chiral ‘isotropic solutions’, the so-called ‘dissymmetric solutions’. Nowadays, chiroptical spectroscopy has been proven to be a prospective tool to analyze chiral anisotropic phases. Here, chirality measurements yield information about the anisotropy of chiral properties of molecules and the chiral long-range positional and orientational order of phases. Especially, the differences of optical constants will be measured possessing for two enantiomers or enantiomorphous phases the same absolute value but opposite signs due to their chirality. Because of these differences in optical constants, light can only propagate in a chiral sample with chiral eigenstates of polarization, i.e., as two orthogonal elliptically or circularly polarized light beams. These eigenstates, determined by the symmetry of the sample (Table 1), penetrate the material without changing their states of polarization. Their existence can be proven by the splitting of unpolarized light penetrating a wedge-shaped sample (Figure 1).

Chiroptical spectroscopy, for which only the circular and elliptical eigenstates are of interest, is usually performed in experimental arrangements with two light beams propagating parallel to each other. They superpose after leaving the sample (Figure 2) to an elliptically polarized light that differs in intensity and polarization from that of the incident light. In a dissymmetric solution with its two refractive indices  $n_L$  and  $n_R$  (Figure 2), the plane of polarization of an incident linearly polarized light is rotated with an angle  $\varphi$  (optical rotation):

$$\varphi = \pi \tilde{v} d (n_L - n_R) \quad (\text{rad}) \quad [1]$$

- Malinowski ER (2002) *Factor Analysis in Chemistry*, 3rd edn. Chichester: Wiley.
- Otto M (1998) *Chemometrics: Statistics and Computer Applications in Analytical Chemistry*, chapter 3. Weinheim: Wiley-VCH.
- Savitsky A and Golay MJE (1964) Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* 36: 1627–1639.
- Sibisi S, Skilling J, Brereton RG, Laue ED, and Staunton J (1984) Maximum entropy signal processing in practical NMR spectroscopy. *Nature* 309: 801–802.
- Vandeginste BGM, Massart DL, Buydens LMC, *et al.* (1998) *Handbook of Chemometrics and Qualimetrics Part B*, chapters 34 and 40. Amsterdam: Elsevier.
- Wegscheider W (1998) In: Kellner R, Mermet J-M, Otto M, and Widmer HM (eds.) *Analytical Chemistry*, section 12.3. Weinheim: Wiley-VCH.

## CHIROPTICAL ANALYSIS

**H-G Kuball**, University of Kaiserslautern, Kaiserslautern, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

‘Chiroptical analysis’ is the application of chiroptical spectroscopy – including absorption, emission, and optical refraction measurements – in order to obtain information about chiral molecules and phases. Lord Kelvin has defined chiral objects in his famous ‘Baltimore Lectures 1889’ by: “I call any geometrical figure or group of points ‘chiral’, and say it has ‘chirality’, if its image in a plane mirror, ideally realized, cannot be brought – by rotation and translation of the geometrical figure or the group of points – into coincidence with itself.” Examples include molecules and chiral anisotropic phases with their long-range orientational and long-range positional order (for a classification by a length scale, four levels of chirality can be introduced. First level: atoms (weak interaction); second level: molecules (geometry); third level: suprastructural chirality (long-range orientational and positional order of atoms/molecules); fourth level: macroscopic objects like helical pillars, staircases, etc.) without symmetry of second kind, i.e., no rotation–reflection symmetry (rotation–reflection axes  $S_n$ ,  $n = 1, 2, \dots$ ). These chiral objects do exist always in two different forms, one being the mirror image of the other (enantiomers and enantiomorphous solids). The introduction of a reference system, like that of Cahn, Ingold, and Prelog (CIP), allows dividing the world of chiral molecules into two classes, the ‘S or R (M or P) world’. For chiral anisotropic phases no universally valid reference systems exist. Exceptions are the left (M) or right

(P) handed helical long-range orientational order of cholesteric and smectic C\* phase and the left (M) or right (P) handed screw axes of crystals.

In the last decades chiroptical spectroscopy has been mostly applied to chiral ‘isotropic solutions’, the so-called ‘dissymmetric solutions’. Nowadays, chiroptical spectroscopy has been proven to be a prospective tool to analyze chiral anisotropic phases. Here, chirality measurements yield information about the anisotropy of chiral properties of molecules and the chiral long-range positional and orientational order of phases. Especially, the differences of optical constants will be measured possessing for two enantiomers or enantiomorphous phases the same absolute value but opposite signs due to their chirality. Because of these differences in optical constants, light can only propagate in a chiral sample with chiral eigenstates of polarization, i.e., as two orthogonal elliptically or circularly polarized light beams. These eigenstates, determined by the symmetry of the sample (Table 1), penetrate the material without changing their states of polarization. Their existence can be proven by the splitting of unpolarized light penetrating a wedge-shaped sample (Figure 1).

Chiroptical spectroscopy, for which only the circular and elliptical eigenstates are of interest, is usually performed in experimental arrangements with two light beams propagating parallel to each other. They superpose after leaving the sample (Figure 2) to an elliptically polarized light that differs in intensity and polarization from that of the incident light. In a dissymmetric solution with its two refractive indices  $n_L$  and  $n_R$  (Figure 2), the plane of polarization of an incident linearly polarized light is rotated with an angle  $\varphi$  (optical rotation):

$$\varphi = \pi \tilde{v} d (n_L - n_R) \quad (\text{rad}) \quad [1]$$

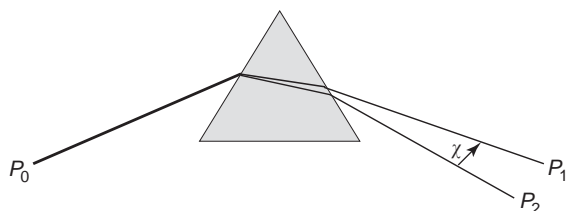


**Table 1** Eigenstates of light

<i>Achiral isotropic material</i>	<i>All states of polarization</i>
Chiral isotropic material (dissymmetric material)	Left and right circularly polarized light
Achiral anisotropic material	Parallely and perpendicularly <sup>a</sup> linear polarized light <sup>b</sup>
Chiral anisotropic material	Left and right elliptically polarized light

<sup>a</sup>With respect to the optical axis/axes of the material.

<sup>b</sup>The 'plane of polarization' is defined by the plane built up by the electric field vector and the propagation direction of light.



**Figure 1** Splitting of an unpolarized light beam  $P_0$  by a dissymmetric solution in a wedge-shaped cell or a wedge-shaped chiral anisotropic solid into the two orthogonal eigenstates  $P_1$  and  $P_2$ .  $\chi \neq 0$  except in the case of isotropic material where all states of polarized light, including the unpolarized light, can propagate ( $P_0 = P_1 = P_2$ ).

where  $d$  is the path length of the sample in centimeters,  $\bar{\nu} = 1/\lambda$  the wavenumber of the incident light, and  $\lambda$  the wavelength in centimeters. The specific rotation  $[\alpha]$  for a gaseous, a fluid, or a solid compound is then defined by:

$$[\alpha] = \frac{\alpha}{\rho l} = \frac{\alpha}{ql} \left[ \frac{\text{deg cm}^3}{\text{mol dm}} \right] \quad [2]$$

where  $\rho$  is the density of a pure compound in grams per centimeter cube and  $q$  is the concentration of a dissolved compound in grams per (100 cm<sup>3</sup> solution).  $\alpha$  is given in degree and  $l$  is the length of the cell in decimeters. The molar rotation  $[M]$  can be expressed by

$$[M] = \frac{[\alpha]M}{100} = \frac{100\alpha}{cd} \left[ \frac{\text{deg cm}^3}{\text{mol dm}} \right] \quad [3]$$

where  $M$  is the molecular mass,  $c$  the concentration in moles per liter, and  $d$  the path length in centimeters. Here, one has to have in mind that the conversion of units is made under the condition that  $[M]$  and  $[\alpha]$  maintains their historical units. No IUPAC regulation for SI units does exist, nowadays.

By the absorption of a dissymmetric solution an incident linearly polarized light changes to an

elliptically polarized light with the ellipticity  $tg\theta$

$$\begin{aligned} tg\theta &= \frac{10^{-\varepsilon_L cd/2} - 10^{-\varepsilon_R cd/2}}{10^{-\varepsilon_L cd/2} + 10^{-\varepsilon_R cd/2}} \cong \theta \\ &= \frac{\ln 10}{4} (\varepsilon_L - \varepsilon_R) cd = 0.5757 \Delta \varepsilon cd \end{aligned} \quad [4]$$

$\varepsilon_L$  and  $\varepsilon_R$  are the molar decadic absorption coefficients of left and right circularly polarized light, respectively.  $\Delta \varepsilon = \varepsilon_L - \varepsilon_R$  (l mol<sup>-1</sup> cm<sup>-1</sup>) is the circular dichroism (CD). Often the molar ellipticity  $[\theta]$

$$[\theta] = \frac{100\theta}{cd} = 3300 \Delta \varepsilon \quad [5]$$

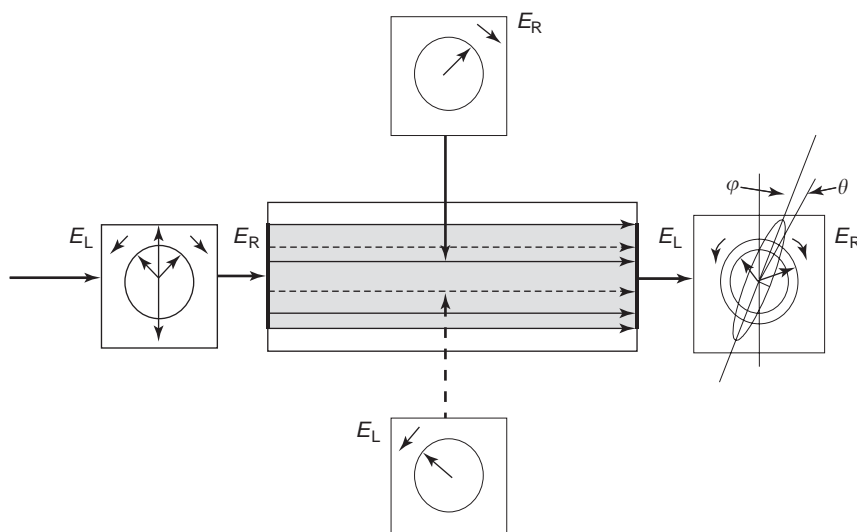
is used instead of  $\Delta \varepsilon$ . Because both eigenstates are independent (Figure 2), the Lambert–Beer law does not hold

$$\begin{aligned} I &= I_0 10^{-\bar{\varepsilon} cd} \cosh \left[ \frac{2.303}{2} (\varepsilon_L - \varepsilon_R) cd \right] \\ &\cong I_0 10^{-\bar{\varepsilon} cd} \end{aligned} \quad [6]$$

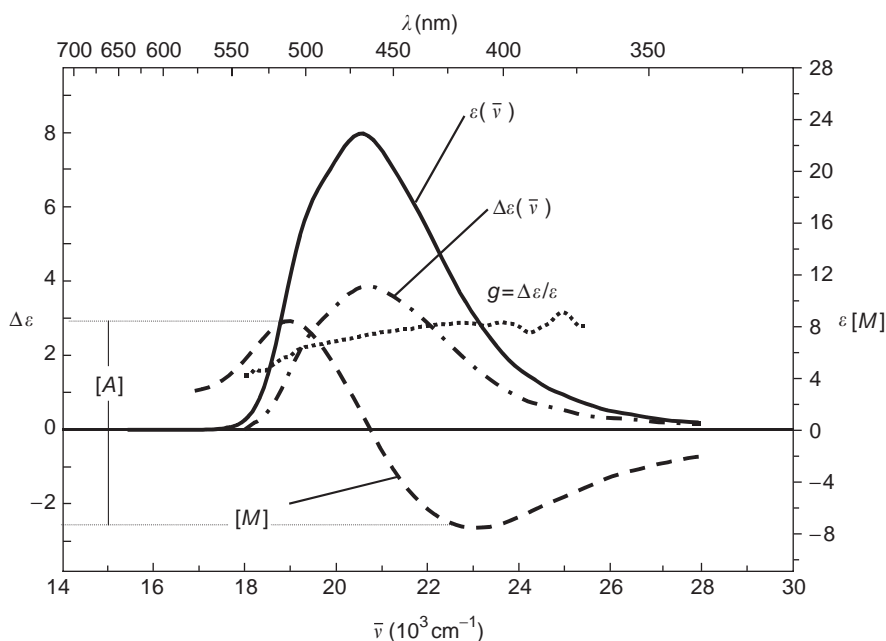
In most cases the deviation from the Lambert–Beer law is negligible.  $\bar{\varepsilon}$  is the average absorption coefficient of the sample:

$$\bar{\varepsilon} = \frac{1}{2} (\varepsilon_L + \varepsilon_R) \quad [7]$$

The CD and optical rotatory dispersion (ORD) can be given either as a function of the wavenumber  $\bar{\nu}$  (cm<sup>-1</sup>) or the wavelength  $\lambda$  (nm) (Figure 3). The wavenumber is proportional to the energy  $\bar{\nu} = (1/hc)E$  ( $E$  energy). From the physical point of view the presentation as a function of  $\bar{\nu}$  should be preferred but the  $\lambda$  dependence is often taken for historical reasons because most spectroscopic instruments use a  $\lambda$  scale. Especially, the recognition and the analysis of vibrational progressions of an electronic absorption band is easier to handle in the  $\bar{\nu}$  representation than in the nonlinear  $\lambda$  representation. The ORD curve is of a simple sigmoid form for a symmetric and structureless CD band. This ORD



**Figure 2** Propagation of the two orthogonal circularly polarized eigenstates within a dissymmetric sample and the superposition of both eigenstates after leaving the sample.  $\varphi$  is the optical rotation and  $tg\theta = b/a$  the ellipticity of the light ( $a$  is the long axis,  $b$  is the short axis of the ellipse).  $\varphi$  is positive for a clockwise rotation of the plane of polarization. The elliptically polarized light is left handed for  $tg\theta < 0$  and  $\Delta\varepsilon < 0$  and right handed for  $tg\theta > 0$  and  $\Delta\varepsilon > 0$ . The electric field vector  $E$  of the circularly polarized light is given for selected positions (the four quadrants) within the optical path in planes perpendicular to the propagation direction of the light. The state of polarization of the eigenstates does not change while propagating through the sample whereas the absorption changes their intensity. For the sign convention: the observer is facing the propagation direction of the light.



**Figure 3** Cotton effect (CE): CD (.....), ORD (-----), dissymmetry factor  $g$  (.....) and the UV (—) band. The CE is positive if  $\Delta\varepsilon > 0$ . Then the positive lobe of the sigmoid ORD curve lies on the long-wavelength/smaller wavenumber side of the absorption band.  $\Delta\varepsilon$  and  $[M]$  change signs (mirrored at the abscissa) for the enantiomer (negative CE). The long-wavelength region between  $\sim 525$  and  $600$  nm ( $17\,000$  and  $19\,000\text{ cm}^{-1}$ ) is a 'plain positive rotation' or 'normal rotatory dispersion'. The sigmoidal form below  $525$  nm is an 'anomalous rotatory dispersion'. The positive maximum of the ORD curve is a so-called 'peak' whereas the negative minimum is a 'trough'.  $[A] = ([M(\lambda_1)] - [M(\lambda_2)]) / 100$  ( $\lambda_1 > \lambda_2$ ) is the amplitude of the rotatory dispersion curve.

curve can be characterized by its amplitude

$$[A] = \frac{[M(\bar{\nu}_1)] - [M(\bar{\nu}_2)]}{100}$$

The amplitude is positive (negative) if  $[A] > 0$  ( $[A] < 0$ ) for  $\bar{\nu}_1 < \bar{\nu}_2 \{ \lambda_1 > \lambda_2 \}$  ( $\bar{\nu}_1 > \bar{\nu}_2 \{ \lambda_1 < \lambda_2 \}$ ). **Figure 3** presents the so-called positive Cotton effect

(CE) where  $\Delta\varepsilon > 0$  and  $[A] > 0$ . For the corresponding enantiomer the negative CE will be obtained:  $\Delta\varepsilon < 0$  and  $[A] < 0$ .

CD and ORD are correlated by Kramers–Kronig transformations

$$\varphi(\omega) = \frac{2}{\pi} \oint \frac{\omega \theta(\bar{\omega})}{\bar{\omega}^2 - \omega^2} d\bar{\omega} \quad [9]$$

and

$$\theta(\omega) = -\frac{2}{\pi} \oint \frac{\bar{\omega} \varphi(\bar{\omega})}{\bar{\omega}^2 - \omega^2} d\bar{\omega}. \quad [10]$$

$\omega = 2\pi\nu = 2\pi\bar{\nu}c$  with  $\nu$  the frequency and  $c$  the vacuum velocity of light. The Kramers–Kronig transforms can be used to check an experimental result in critical situations and assure ORD measurements by CD measurements and vice versa by calculating the ORD from the CD and vice versa by eqns [9] and [10].

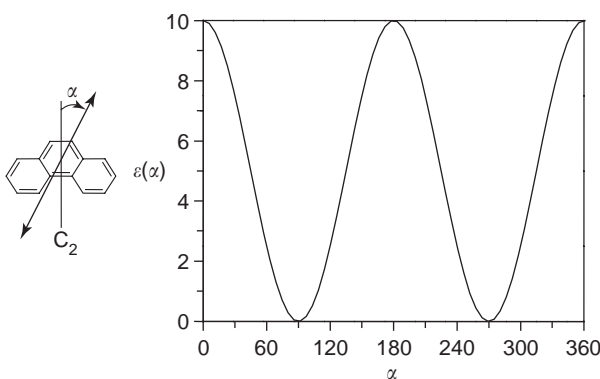
## Theory

For molecules the ‘optical chirality measurements CD and ORD’ can be performed between vacuum ultraviolet (UV) (ECD: CD of electronic transitions) and infrared (IR) (VCD: CD of vibrational transitions) spectral regions in which  $\lambda$  is large in comparison with the diameter of a molecule. Suprastructural chirality can be observed in a spectral region in which  $\lambda$  is in the order of the periodicity of the chiral structures as can be established by the ‘selective reflection’ of a cholesteric phase in the IR spectral region.

The simplest approach to understand the interaction of light with chiral molecules is the assumption of a diastereomeric interaction between circularly polarized light and chiral molecules. This leads to different absorption coefficients and refractive indices for left ( $\varepsilon_L$ ,  $n_L$ ) and right ( $\varepsilon_R$ ,  $n_R$ ) handed circularly polarized light with left handed and right handed molecules, respectively. The interaction of light with a molecule can be quantitatively described by the introduction of an electric dipole transition moment  $\langle \underline{\mu} \rangle_{ga}$  and a magnetic dipole transition moment  $\langle \underline{m} \rangle_{ag}$ . The scalar products of these vectors  $\langle \underline{\mu} \rangle_{ga}$  and  $\langle \underline{m} \rangle_{ag}$  (eqns [13] and [17]) are proportional to the probability by which the molecule in the state ‘g’, characterized by the wavefunction  $\psi_g$  and the energy  $E_g$ , is excited to the state ‘a’, given by the wavefunction  $\psi_a$  with the energy  $E_a$ :

$$\langle \underline{\mu} \rangle_{ga} = \langle \psi_g | \underline{\mu} | \psi_a \rangle \quad [11]$$

$$\langle \underline{m} \rangle_{ag} = \langle \psi_a | \underline{m} | \psi_g \rangle \quad [12]$$



**Figure 4** Intensity of absorption as a function of the angle between the electric field of the light and the electric dipole transition moment direction  $\langle \underline{\mu} \rangle_{ga}$ :  $\varepsilon(\alpha) = \langle \underline{\mu} \rangle_{ga}^2 \cos^2 \alpha$ .

where  $\underline{\mu}$  [Debye] and  $\underline{m}$  [BM] are the electric and magnetic dipole operators, respectively. The directions of the vectors  $\langle \underline{\mu} \rangle_{ga}$  and  $\text{Im} \{ \langle \underline{m} \rangle_{ag} \}$  determine the directions along which the light has to be polarized in order to have the maximum of absorption ( $\alpha = 0^\circ$  and  $180^\circ$  in **Figure 4**). The magnetic dipole transition moment is a complex quantity. Thus,  $\text{Im} \{ \}$  means the imaginary part of  $\langle \underline{m} \rangle_{ag}$ . Light that is polarized perpendicular to  $\langle \underline{\mu} \rangle_{ga}$  and  $\text{Im} \{ \langle \underline{m} \rangle_{ag} \}$ , respectively, penetrates the sample without being absorbed ( $\alpha = 90^\circ$  and  $270^\circ$  in **Figure 4**).

For a transition between the vibronic ground state  $g \equiv |Nn\rangle$  and the excited state  $a \equiv |Kk\rangle$  follows for the integrated intensity of the absorption:

$$D^{NnKk} = \langle \underline{\mu} \rangle_{NnKk}^2 + \langle \underline{m} \rangle_{KkNn}^2 \cong \langle \underline{\mu} \rangle_{NnKk}^2 \quad [13]$$

$D^{NnKk}$  is the dipole strength of the transition  $|Nn\rangle \rightarrow |Kk\rangle$ .  $N$  and  $K$  characterize the electronic ground and excited state and the quantum numbers  $n$  and  $k$  do correspond to their vibrational states, respectively. The contribution of the magnetic transition in eqn [13] can be neglected because of  $|\langle \underline{\mu} \rangle_{NnKk}|^2 \gg |\langle \underline{m} \rangle_{KkNn}|^2$ . The corresponding integrated intensity of the CD band, the rotational strength, is given by the scalar product

$$\begin{aligned} R^{NnKk} &= \sum_{i=1}^3 \text{Im} \{ \langle \underline{\mu}_i \rangle_{NnKk} \langle \underline{m}_i \rangle_{KkNn} \} \\ &= \text{Im} \{ \langle \underline{\mu} \rangle_{NnKk} \cdot \langle \underline{m} \rangle_{KkNn} \} \\ &= |\langle \underline{\mu} \rangle_{NnKk}| |\langle \underline{m} \rangle_{KkNn}| \cos \vartheta \end{aligned} \quad [14]$$

$\text{Im} \{ \}$  means the imaginary part of the scalar product.  $\vartheta$  is the angle between the transition moments  $\langle \underline{\mu} \rangle_{NnKk}$  and  $\langle \underline{m} \rangle_{KkNn}$ . The quantities  $\langle \underline{\mu}_i \rangle_{NnKk}$  and  $\langle \underline{m}_i \rangle_{KkNn}$  ( $i = 1, 2, 3$ ) are the three components (coordinates) of the electric and magnetic dipole

transition moments with respect to a chosen molecule-fixed coordinate system.

With the line shape  $F^{NnKk}(\bar{\nu})$  of the transition  $|Nn\rangle \rightarrow |Kk\rangle$ , which is equal for most of the experimental conditions for the absorption and the CD bands, the absorption coefficient  $\varepsilon$  is given by

$$\varepsilon(\bar{\nu}) = \frac{B\bar{\nu}}{12} \sum_n \sum_{Kk} D^{NnKk} F^{NnKk}(\bar{\nu}) = \sum_K \varepsilon^{NK}(\bar{\nu}) \quad [15]$$

$$B = \frac{32\pi^3 N_A}{10^3 h c_0 \ln 10} = 7.653 \times 10^{40} \quad \text{cgs} \quad [16]$$

$N_A$  is Avogadro's number,  $c_0$  the velocity of light, and  $h$  is Planck's constant. Summation over all vibrational contributions then leads to the dipole strength of the electronic transition  $|Nn\rangle \rightarrow |Kk\rangle$ :

$$D^{NK} = \langle \underline{\mu} \rangle_{NK}^2 \quad [17]$$

For CD  $\Delta\varepsilon$  follows

$$\Delta\varepsilon(\bar{\nu}) = \frac{B\bar{\nu}}{3} \sum_n \sum_{Kk} R^{NnKk} F^{NnKk}(\bar{\nu}) = \sum_K \Delta\varepsilon^{NK}(\bar{\nu}) \quad [18]$$

Summation over all vibrational contributions leads to the rotational strength of the electronic transition  $|Nn\rangle \rightarrow |Kk\rangle$ . The dipole strength and the rotational strength of an electronic transition can be experimentally determined by integration over the corresponding UV and CD band by

$$D^{NK} = \frac{12}{B} \int_{\text{band}} \frac{\varepsilon^{NK}(\bar{\nu})}{\bar{\nu}} d\bar{\nu} \quad [19]$$

and

$$R^{NK} = \frac{3}{B} \int_{\text{band}} \frac{\Delta\varepsilon^{NK}(\bar{\nu})}{\bar{\nu}} d\bar{\nu} \quad [20]$$

With the Kramers–Kronig transforms (eqn [9]), the frequency dependence of the molar optical rotation (ORD) can be obtained as follows:

$$\begin{aligned} [M(\bar{\nu})] &= \frac{288\pi N_A \bar{\nu}}{3hc} \sum_n \sum_{Kk} R^{NnKk} J_2^{NnKk}(\bar{\nu}) \\ &= \sum_K [M^{NK}(\bar{\nu})] \end{aligned} \quad [21]$$

$$J_2^{NnKk}(\bar{\nu}) = \oint \frac{\bar{\nu}' F^{NnKk}(\bar{\nu}') d\bar{\nu}'}{\bar{\nu}'^2 - \bar{\nu}^2}. \quad [22]$$

$J_2^{NnKk}(\bar{\nu})$  (eqn [22]) is the principal value of the integral over the singularity at the wavenumber of the corresponding excitation ( $\bar{\nu}' = \bar{\nu}$ ). The dispersion

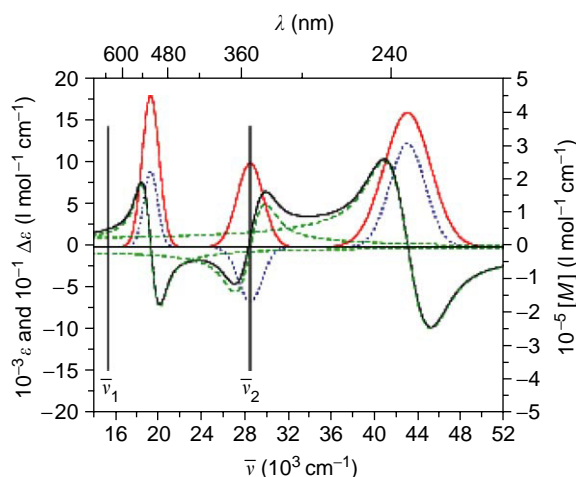
of the well-known Rosenfeld equation for the ORD for  $\lambda \gg \lambda_{NK}$  is obtained from eqn [22] if vibrational states  $n$  and  $k$  do not contribute:

$$J_2^{NK}(\bar{\nu}) = \frac{\bar{\nu}^2}{\bar{\nu}_{NK}^2 - \bar{\nu}^2} = \frac{\lambda_{NK}^2}{\lambda^2 - \lambda_{NK}^2} \quad [23]$$

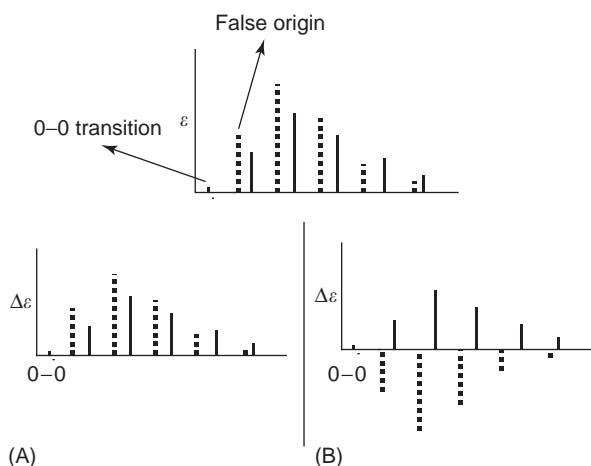
$\bar{\nu}_{NK}$  and  $\lambda_{NK}$  are the spectral positions of the corresponding transition.

To every absorption band  $\varepsilon^{NK}(\bar{\nu})$  belongs a CD band  $\Delta\varepsilon^{NK}(\bar{\nu})$  and a corresponding ORD curve  $[M^{NK}(\bar{\nu})]$  (Figure 3). The right-hand side of eqns [18] and [21] are the sum over contributions of all electronic transitions. Thus, the CD and ORD spectra can be interpreted as a sum of Cotton effects. It is evident from Figures 3 and 5 that the ORD curve possesses a larger half-bandwidth than  $\Delta\varepsilon^{NK}(\bar{\nu})$ . As a consequence the molar optical rotation  $[M(\bar{\nu})]$  is, even in spectral regions far away from absorption bands, always a sum of contributions of many transitions. The experimental CD bands in a given wavelength region are often determined by only one or at least by a few transitions.

Completely unstructured CD and UV bands are the exception, not the rule. With any electronic excitation a large number of rotational and vibrational states are involved in the absorption process. Neglecting rotational states, because they do not appear in normally measured CD spectra, every vibronic electronic state  $|Nn\rangle$  or  $|Kk\rangle$  can be described within the Born–Oppenheimer approximation by the product of vibrational states  $n(N)$  and  $k(K)$  and the electronic states  $N$  and  $K$ , respectively. Often, formally forbidden – mostly weak – absorption bands gain their



**Figure 5** CD, ORD, and UV spectra as a sum of contributions of different transitions, i.e., different chromophores  $|N\rangle \rightarrow |K\rangle$ .  $\bar{\nu}$  is a wavenumber in the region of a 'normal rotatory dispersion' and  $\bar{\nu}$  is a wavenumber at the center of a CE.



**Figure 6** Vibrational progressions contribute to a CD spectrum of an electronic transition either with equal (A) or different (B) sign. For (A) the CD and UV bands are quite similar whereas for (B) the CD band possesses a larger vibrational structure than the absorption band (pseudovibrational structure) because of the superposition of CD curves of different signs.

intensity by vibronic coupling with intensive allowed transitions via totally symmetric ( $\bar{\nu}_1$ ) and nontotally symmetric ( $\bar{\nu}_X$ ) vibrations (intensity borrowing). Then, these vibrations contribute via vibrational progressions (eqns [24] and [25]) to the CE with equal or different sign (Figure 6).

$$\bar{\nu} = \bar{\nu}_{00} + n\bar{\nu}_1 \quad [24]$$

$$\bar{\nu} = \bar{\nu}_{00} + \bar{\nu}_X + n\bar{\nu}_1 \quad [25]$$

$n = 1, 2, 3, \dots$  and  $\bar{\nu}_{00}$  is the wavenumber of the 0–0 transition, i.e., the vibrational free electronic transition.

Absorption bands can be classified by the symmetry of the involved vibrational and electronic states. This symmetry determines also the transition moment directions, i.e., the orientation of the transition moment vector  $\langle \mu \rangle_{NnKk}$  within the molecule. On the other hand, an experimentally determined transition moment direction allows assigning an absorption band. The transition moment directions can be evaluated from the linear dichroism described in polarized spectroscopy in terms of the degree of anisotropy  $R$ , given here for a uniaxial phase:

$$R = \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + 2\varepsilon_2} = \frac{\varepsilon_1 - \varepsilon_2}{3\varepsilon} \quad [26]$$

$\varepsilon_1$  and  $\varepsilon_2$  are the absorption coefficients for linear polarized light polarized parallel or perpendicular to the optical axis of the uniaxial phase. The following

equations hold:

$$\varepsilon_1 = \varepsilon + (\varepsilon_{33}^* - \varepsilon)S^* + \frac{1}{\sqrt{3}}(\varepsilon_{11}^* - \varepsilon_{22}^*)D^* \quad [27]$$

with

$$\varepsilon = \frac{1}{3}(\varepsilon_1 + 2\varepsilon_2). \quad [28]$$

$S^*$  and  $D^*$  are the Saupe order parameters that describe the order of the molecules in the absorbing sample.  $\varepsilon_{ii}^*$  ( $i = 1, 2, 3$ ) are the diagonal elements of the transition moment tensor. They are proportional to the squares of the components of the transition moments ( $\langle \mu_i^* \rangle$  ( $i = 1, 2, 3$ )) given with respect to a molecule fixed-coordinate system ( $x_i^*$ ):  $\varepsilon_{ii}^* \propto (\langle \mu_i^* \rangle)^2$  (see section on Circular dichroism of chiral anisotropic phases without suprastructural chirality).

A further classification of the CE of a band is obtained with the dissymmetry factor as a function of wavenumbers by

$$g(\bar{\nu}) = \frac{\Delta\varepsilon^{NK}(\bar{\nu})}{\varepsilon^{NK}(\bar{\nu})} \quad [29]$$

or for the electronic transition  $|N\rangle \rightarrow |K\rangle$  by

$$g = \frac{4R^{NK}}{D^{NK}} \quad [30]$$

The size of  $g$  also indicates whether the CD of an absorption band is measurable or not.

## Chiroptical Methods

### Methods, Techniques, and Problems

ECD, VCD, Raman optical activity (ROA), and ORD are at disposal in a spectral range between the IR and the vacuum UV. ACD and AORD, the CD and ORD of chiral anisotropic phases are only available in the UV/vis spectral range. The specific rotation, as a standard, is measured with the sodium D-line. For special applications other lines, e.g., mercury lines, have been taken. Indirect methods for chiroptical analyses are the nuclear magnetic resonance (NMR) spectroscopy of diastereomeric compounds and the chiral induction of cholesteric phases (helical twisting power (HTP)) combined with ACD/CD and the corresponding selective reflection.

There are three main fields for chiroptical analyses:

#### (A) Structural information

1. the absolute configuration,
2. the structure and diversity of conformations,
3. the characterization of compounds,
4. the recognition of time dependent structural variation

- (B) Analytical measurements
5. the concentrations of chiral compounds, e.g., saccharose (saccharimetry)
  6. the enantiomeric excess via optical purity
  7. the enantiomeric separation during a chromatographic run in liquid chromatography
- (C) Phase properties
8. pretransitional order near a phase transition
  9. the suprastructural chirality in crystals, films, membranes, and liquid crystal phases.

All techniques have been more or less successfully applied to solve problems in different scientific fields. Table 2 gives a survey of the available techniques and their main field of application.

### CD Contra ORD

In principle, CD and ORD yield identical information because they are Kramers–Kronig transforms (eqns [9] and [10]). Thus, it seems to be sufficient to measure either CD or ORD. This is only true from the theoretical point of view because of some practical reasons:

1. CD and ORD can only be measured in confined parts of the spectrum.
2. It is more convenient to analyze CD than ORD spectra because of its smaller bandwidth. The large bandwidth of  $[M^{NK}(\bar{\nu})]$  leads to a large background effect in the spectral region of all other chromophores. This background effect is difficult to correct.
3. The half-bandwidth of an absorption band ( $\epsilon^{NK}(\bar{\nu})$ ) and the corresponding CD band ( $\Delta\epsilon^{NK}(\bar{\nu})$ ) is approximately the same. This allows an easier assignment of the CD band to the corresponding absorption band.
4. The analysis of vibrational progressions can be easier performed with CD than ORD because vibrational progressions often contribute with bands

of different signs to CD and thus an ‘apparent’ vibrational structure of the CD band is obtained in contrast to the less structured band shape of the absorption band. The vibrational structure of an ORD curve is very complex and nearly not separable into contributions of different vibrational progressions.

5. The ECD spectra can be theoretically calculated with sufficient accuracy for a direct comparison of a calculated and an experimental spectrum.

6. Last but not least, the calculation of an ORD spectrum from the CD spectrum with the Kramers–Kronig relation (eqn [9]) is easier to perform than the calculation of the CD spectrum from the ORD curve.

The very first systematic chiroptical analyses have been obtained with ORD measurements in the 1950s and 1960s. The dominance of the ORD analyses has been broken by commercial CD instruments in the 1960s when CD and ORD were measurable with an equivalent accuracy. Since that time ORD is just of interest for selected problems. It should be mentioned here that a new development based on the invention of new numerical quantum mechanical methods, which allows calculating  $[\alpha]_D^T$  with a sufficient accuracy, may lead to a situation in which even CD, VCD, and ORD measurements can be avoided. Here, in order to avoid experimental artifacts by association, etc.,  $[\alpha]_D^T$  has to be extrapolated with  $c \rightarrow 0$ :  $[\alpha]_{D,c \rightarrow 0}^T$  (see section on Quantum mechanical computation for electronic circular dichroism and vibrational optical activity).

ORD measurements in the IR region never have played any role. The ORD method has to be applied in the UV/vis spectral region if a compound does not possess a suitable chromophore. But in order to avoid an ORD analysis, ‘chromophore-free compounds’ will often be substituted with suitable absorbing groups (see section on Electronic circular dichroism for compounds without a chromophore).

### The Chromophore as a Probe for Chirality – Assignment of Transitions

The concept of the ‘chromophore’ is of eminent importance for probing chiral structures of molecules and their surroundings. But what is a chromophore? In chemical textbooks, a chromophore is a molecule or a group in a molecule that is the ‘carrier of color’ in the visible spectral region. More generally, one chromophore does belong to every absorption band and vice versa. But besides this, the notion is only intuitively defined. From the spectroscopic point of view a chromophore can be defined as an area of a molecule in which some properties have changed after the molecule became excited. In this context measurable and computable properties are the

**Table 2** The main applications of chiroptical methods

Numbering given in the list above	Technique/method						
	ECD	VCE	ROA	ORD	$[\alpha]$	ACD	HTP
A-1	+	+	+		∇	+	∇
A-2	+	+	+	∇			
A-3	∇	∇	∇	∇	+	+	+
A-4	∇	∇					
B-5	∇	∇		∇	+		
B-6	+				+		
B-7	+				+		
C-8				∇	+		
C-9	+					∇	+

+, main application; ∇, possible application.



geometry, charge distribution, polarizability, etc. From a chemist's view a chromophore can be identified by, e.g., a substitution of an auxiliary group that leads to a large or at least a distinct change of the absorption if the area around the position of the substitution belongs to the chromophore. In most cases a chromophore is located in a small restricted part of a molecule, bound to functional groups. Sometimes the whole molecule can be a chromophore as, e.g., with naphthalene with its  $\pi\pi^*$  transitions. The geometrical areas of different chromophores do often overlap, especially those that belong to absorption bands in very different spectral regions. A new chromophore comes into being if the excitation energy of, e.g., two chromophores, is of similar size, a phenomenon well known from the exciton coupling. The location of a chromophore in a molecule is of importance for the applicability of sector and helicity rules for adjacent chromophores. Adjacent chromophores should possess similar or even identical rules and, by this, should give similar or identical results about the surroundings of the chromophore.

In the context discussed above, 'characterizing a chromophore' means an assignment of the absorption band to its corresponding quantum mechanical states. The electric transition moment direction of the absorption band belonging to the chromophore that can be experimentally determined from the degree of anisotropy  $R$  (eqn [26]) is an important information for an unequivocal assignment. Furthermore, the dissymmetry factor  $g$  (eqn [30]) rewritten by eqns [14] and [17] into

$$g = \frac{4|\langle m \rangle_{NK}|\cos \vartheta}{|\langle \underline{\mu} \rangle_{KN}|} \quad [31]$$

allows to identify four types of chromophores. Three types of chromophores possess a local symmetry and so are called inherent symmetric. Their chirality effects are induced through a perturbation by atoms or groups in their surroundings. This leads to transitions that are:

1. magnetically allowed and electrically forbidden;
2. magnetically forbidden and electrically allowed; and
3. magnetically and electrically forbidden.

Dipole transitions magnetically allowed and electrically forbidden are, e.g., the  $n\pi^*$  transitions of the carbonyl group in aldehydes or ketones or the d-d transition of transition metal complexes. For an allowed magnetic dipole transition,  $|\langle m^{NK} \rangle|$  is  $\sim 10^{-21}$  (cgs). For a forbidden electric dipole

transition,  $|\langle \mu^{NK} \rangle|$  is  $\sim 10^{-19}$  to  $10^{-20}$  ( $\epsilon \sim 10$ – $100$ ). Thus,  $g$  is  $\sim 10^{-2}$  to  $10^{-1}$  and the CD is easily measurable. The absorption of the  $n\pi^*$  transitions often gains intensity from the effect of vibronic coupling with allowed  $\pi\pi^*$  transitions via nonsymmetric vibrations induced by the perturbation of the symmetry of the chromophore by its surroundings. For an electrically allowed and magnetically forbidden dipole transition,  $g$  is  $\sim 10^{-3}$  to  $10^{-6}$  and thus its corresponding CD is difficult to measure because of its strong absorption ( $\epsilon$  between  $10^4$  and  $10^6$ ) as found, e.g., for  $\pi\pi^*$  and  $\sigma\sigma^*$  transitions. For the third type of chromophore no common rule can be given. They are often transitions between  $\pi$  and  $\sigma^*$  or  $\sigma$  and  $\pi^*$  orbitals.

The fourth, the so-called inherent dissymmetric chromophore, does not possess local symmetry. Therefore, the transitions belonging to these chromophores are magnetically and electrically allowed. Inherent dissymmetric chromophores are often found with so-called 'form chiral' molecules for which atropisomers like binaphthols are typical representatives. Further examples are chromophores that come into being by exciton coupling. In both cases the dissymmetry factor  $g$  is in the order of  $10^{-2}$  to  $10^{-1}$  and the CD is easily measurable in spite of the fact that the absorption coefficients of these compounds are often very high ( $\epsilon$  between  $10^4$  and  $10^6$ ).

## Absolute Configuration – Stereochemical Application

In spite of the fact that nowadays well established and widely applicable X-ray techniques to determine the absolute configuration are available, these methods cannot displace chiroptical analyses, not only because the X-ray analyses needs good crystalline material. With ECD, VCD, and ROA, besides configurational chirality the conformational chirality of the diverse conformations of nonrigid molecules can also be analyzed.

## Experimental Information

**Electronic and vibrational circular dichroism spectroscopy** With commercially available instruments the accessible spectral region of ECD and VCD spectroscopies lies nowadays between  $\sim 800$  ( $1.25 \mu\text{m}$ ) and  $62\,500 \text{ cm}^{-1}$  ( $160 \text{ nm}$ ). From the spectroscopic point of view the chosen solvents should be free of absorption and in order to have only a small solvent/solute interaction they should be as nonpolar as possible. Acyclic or cyclic hydrocarbons are a good choice for the UV/vis absorption region if sufficient solubility is guaranteed. As a compromise, dioxane,



methanol, and acetonitrile are often used for the UV/vis spectral region. For the IR region besides neat liquids any IR solvent can be used, especially H<sub>2</sub>O, D<sub>2</sub>O, DMSO-d<sub>6</sub>, etc. Measurements in the gaseous state are difficult to perform because the low density of the gas requires long path lengths – often in the order of 100 cm and more. The concentration range for the CD in the UV/vis region has to be chosen between  $\sim 10^{-5}$  and  $10^{-1}$  mol l<sup>-1</sup>, depending on the type of chromophore and the optical quality of the CD instruments. For the IR measurements the concentration is in the range of  $\sim 0.01$  and  $0.6$  mol l<sup>-1</sup> and path length lies between 5  $\mu$ m and 1 cm.

**Fluorescence detected circular dichroism** Often  $\Delta\epsilon$  of a compound is too small or the amount of available material is not sufficient to perform a CD or ORD measurement. The fluorescence detected circular dichroism (FDCD) allows improving the sensitivity and selectivity of the analysis by observing the difference of the fluorescence intensities of molecules excited with left ( $I_L$ ) and right ( $I_R$ ) circularly polarized light. This difference is caused by a different population of the excited states of an enantiomer of a compound when excited with left or right circularly polarized light. Therefore, the information obtained corresponds to the ground state like that measured in absorption. In spite of the fact that FDCD is a long known technique its application was limited because of a poor signal-to-noise ratio and the often existing serious artifacts. For measurements using viscous solvents also the anisotropy induced by photoselection can falsify the result. Recently gained experimental improvements have changed this situation.

**Chiroptical luminescence** Circular polarized luminescence (CPL) allows obtaining information about excited state structures. Corresponding achiral information can be obtained from electrochromic phenomena of molecules. Serious experimental problems, which often lead to more artifacts than acceptable, have prohibited a broad application of this method. Information can be obtained about S $\rightarrow$ T transitions of enones and ketones. For metal complexes kinetic studies about stereochemical dynamics are available. New experimental developments seem necessary for a broad application of the CPL.

### Assignment of the Absolute Configuration

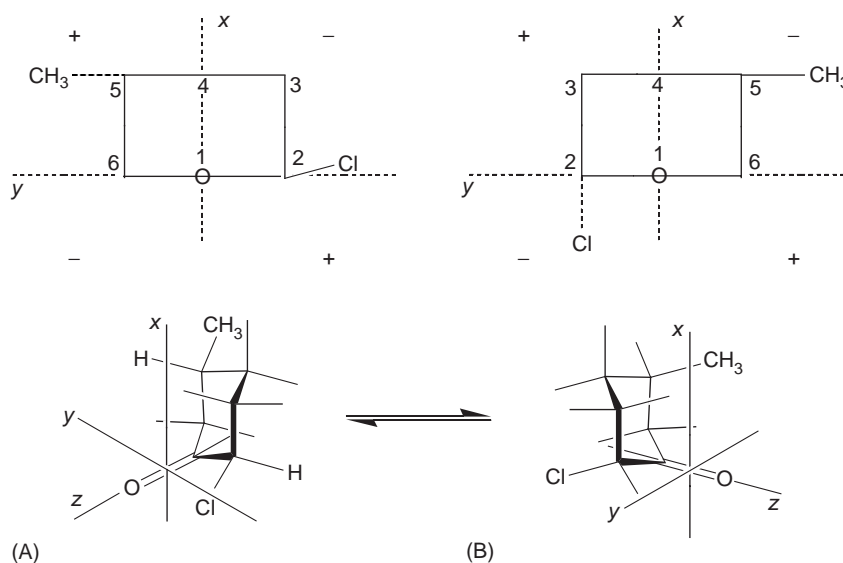
The basis for an assignment of the absolute configuration is the knowledge of the CE, the rotational strength  $R$  (eqn [14]), or the dissymmetry factor  $g$

(eqns [29] and [30]) of one or more absorption bands. In addition, methods are needed that correlate the sign of CE,  $R$ , or  $g$  with the absolute configuration. In principle, there are two different possibilities:

1. The assignment by semiempirical and empirical rules known as sector and helicity rules. Sector rules can be applied to inherent symmetric chromophores. Helicity rules have been deduced for inherent dissymmetric chromophores.
2. The comparison of the computed rotational strength or the corresponding CD spectra with experimentally determined CD spectra.

Whereas sector rules and helicity rules have been very successfully applied for the CD of electronic transitions (ECD), their applicability in VCD spectroscopy is very limited.

**Sector rules for electronic circular dichroism** The first known rule, the octant rule for the  $n\pi^*$  transition of cyclic ketones, can be depicted using the local  $C_{2v}$  symmetry of the carbonyl group: C=O. Their two symmetry planes decompose the space outside the chromophore into four quadrants (Figure 7). The simplest chiral system is obtained by taking the C=O chromophore and one atom positioned out of the symmetry planes in one of the quadrants. Because this simple system is chiral, the atom induces a CE in the  $n\pi^*$  band that is assumed for the following discussion to be of positive sign. The same atom in a mirror image position, i.e., in one of the two neighboring quadrants, leads to the enantiomer of the original system and, thus, the induced CE is of opposite sign. Moving the atom on a circle parallel to the  $x, y$  plane through the four quadrants the sign pattern for the induced CD has to be ‘plus-minus-plus-minus’ because no chiral zero is allowed on the circle except the atom lies in the  $x, z$  or  $y, z$  symmetry plane. Qualitative summation of the effect of all atoms of the molecule in the different quadrants leads to the CE of the  $n\pi^*$  transition. Identical atoms in mirror image position compensate each other. Atoms positioned in the symmetry planes do not contribute to the CE at all. Atoms or groups with a higher polarizability yield a higher contribution except, e.g., in the case of fluorine. Experimental results have shown that a quadrant rule does not hold for the C=O group. An additional plane is needed for a correct sign pattern. This additional plane, which decompose the space into eight octants, has been found to be the nodal plane of the  $\pi^*$  orbital of the carbonyl group. In a modern presentation of the octant rule, instead of the flat nodal plane of the  $\pi^*$  orbital, a curved plane has been introduced. By convention, a



**Figure 7** Octant rule depicted with the equilibrium of diequatorial (A) and diaxial (B) *trans*-2(*R*)-chloro-5(*R*)-methylcyclohexanone and their octant projection diagrams. The cyclohexanone ring lies in the four rear octants. The oxygen of the carbonyl group is directed along the positive *z*-axis (viewing direction is against the positive *z*-axis). The numbers in the octant projection diagrams indicate the position of carbon atoms of the cyclohexanone ring. The chlorine atom in (A) is positioned nearer to the symmetry plane of the carbonyl chromophore ( $\sigma_{yz}$ ). The sign pattern for the CE contributions of atoms is given for the four rear octants.

compound is positioned in the octant as shown in **Figure 7**. The main part of the molecule then lies in the four rear octants. The induced sign of the CD by a perturbing atom in the *x*, *y*, *z* position is then given by the product of its coordinates, i.e.,  $\text{sign}(x \cdot y \cdot z)$  (**Figure 7**). An attempt to quantify the octant rule has failed because of a number of theoretical restrictions of this simple perturbation model. The rule can also be derived from E. Ruch's and A. Schönhofer's theory of chirality functions using only symmetry arguments. Furthermore, the  $n\pi^*$  transition in a  $C_{2v}$  chromophore is a forbidden electric dipole transitions that gains its absorption intensity from vibronic coupling with neighbored  $\pi\pi^*$  transitions. Hereby different vibrational progressions yield contributions of equal or opposite sign to the CD band. The rule fails when the ratio of the CD intensities, belonging to the positive and negative progressions, become interchanged in comparison to the ratio of the reference compounds from which the rule have been derived. There is another problem worthy of mentioning: How large can a perturbation be before the perturbing atom or group has to be included into the originally defined chromophore in order to create a new chromophore. The new chromophore possibly possesses another symmetry or is even converted to an inherently dissymmetric chromophore. In any case, then other sector or helicity rules hold.

For many inherent symmetric chromophores sector rules have been derived. With an increasing

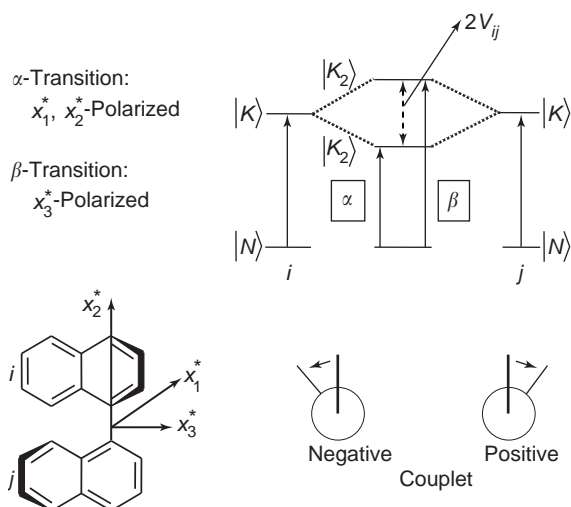
number of symmetry elements for a chromophore the spatial areas in which atoms induce a contribution to the CD of the same sign will get smaller and smaller. Therefore, such rules should be used only with care.

**Helicity rules for electronic circular dichroism** For inherent dissymmetric chromophores all transitions are magnetically and electrically allowed and the dissymmetry factor *g* is, in principle, larger than for inherent symmetric chromophores. Furthermore, the CD is only determined by the geometry of the chromophore without any influence from its surroundings. An unequivocal correlation between a helical sense of a chromophore and the sign of the CE can be given as, e.g., for polymers or biopolymers (e.g., DNA) if atoms or groups in the chromophore can be exactly ordered on a helical line.

For small molecules the selected atoms for defining a helical structure are often only approximately on a helical line. Thus, the helical sense for the following helicity rule of the cisoid enone is uncertain, in principle: The helical sense of the cisoid enone chromophore, e.g., is left handed (right handed) if the CD of the  $n\pi^*$  transition  $\sim 340$  nm is negative (positive). For the  $\pi\pi^*$  transition  $\sim 220$  nm  $\Delta\epsilon > 0$  ( $\Delta\epsilon < 0$ ) for a left handed (right handed) helix. An equivalent rule can be applied to the long-wavelength  $\pi\pi^*$  transition of a diene chromophore. In the given example the positive (negative) helicity is defined by a positive (negative) dihedral angle about the central C–C bond

of the enone chromophore. In this example, the helical sense is only unequivocally defined because the orientation of the helix axis within the molecule, the central C–C bond, is given, i.e., when choosing atoms to approximate a helix, at first a direction for the helix axis has to be chosen because a given helical sense is not unequivocal without the knowledge of orientation of this axis within the molecule. For the twisted diene C=C–C=C, e.g., two different axes with helices of opposite sense can be approximated: one lying parallel to the C<sub>2</sub> symmetry axis, the other parallel to the C–C single bond. One has to be aware of this problem especially when also different atoms have been chosen to define a helical order within a molecule. It is interesting to see that for well chosen axes the sign change of the helix sense can be correlated to a measurable anisotropy of the CD, experimentally found in ACD spectroscopy with, e.g., binaphthols.

The helicity rule of the exciton chirality method is one of the most applied rules for the UV/vis spectral region. This rule possesses a very high reliability for inherent dissymmetric molecules with two identical or similar very strongly absorbing groups. The interaction of two identical chromophores *i* and *j* with their degenerate excited state leads to two non-degenerate states *K*<sub>1</sub> and *K*<sub>2</sub> (Figure 8). The resulting partially overlapping absorption bands  $|N\rangle \rightarrow |K_1\rangle$  and  $|N\rangle \rightarrow |K_2\rangle$  are polarized approximately perpendicular to each other as a consequence of a



**Figure 8** The energy scheme of the exciton chirality method presented for binaphthyl.  $\alpha$  and  $\beta$  indicate the exciton transitions  $|N\rangle \rightarrow |K_1\rangle$  and  $|N\rangle \rightarrow |K_2\rangle$ , respectively.  $2V_{ij}$  is the energy splitting induced by the interaction between both chromophores *i*, *j*. The Newman projections depict the rule for the assignment of the absolute configuration by correlating the sign of the couplet and the dihedral angle.

symmetric and an antisymmetric coupling of the electric dipole transitions in the monomeric units. In the simplest approximation the dipole strength of the absorption of the ‘dimer’ is twice the absorption of the monomer unit. The CD of the new transitions are opposite in sign and their rotational strengths is given by

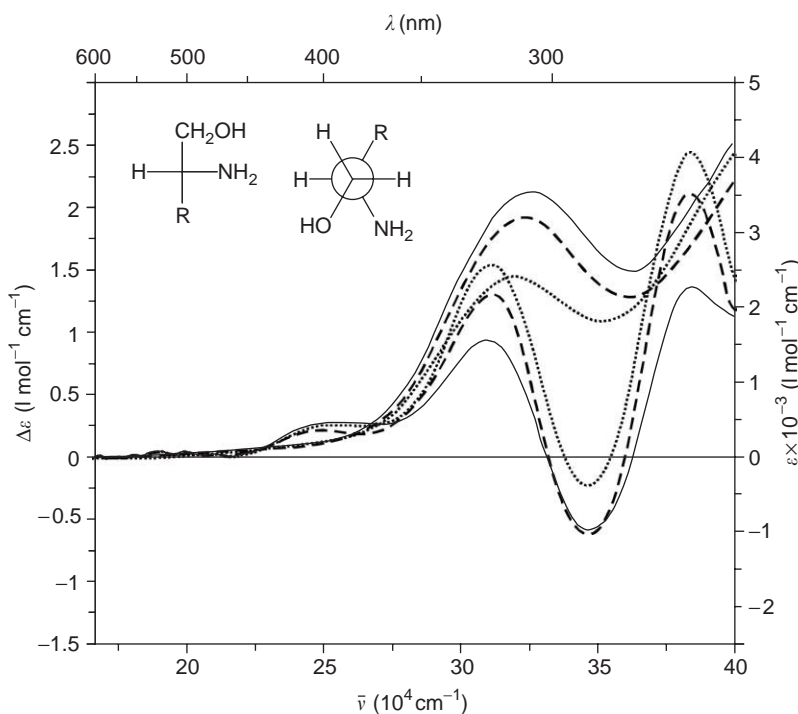
$$R^{NK_1} = -R^{NK_2} = \underline{R}_{ij}(\langle \underline{\mu} \rangle_{NK_1}^i \cdot \langle \underline{\mu} \rangle_{NK_2}^j) V_{ij} \quad [32]$$

where  $2V_{ij}$  is the Davidov splitting and  $\langle \underline{\mu} \rangle_{NK_1}^i$  and  $\langle \underline{\mu} \rangle_{NK_2}^j$  are the transition moments in the unperturbed monomeric units.  $\underline{R}_{ij}$  is the interchromophoric distance vector between the monomeric units.

Often the energy splitting  $2V_{ij}$  is small. Then the splitting of the absorption bands is also small and the second band only appears as a shoulder whereas in the CD spectrum both bands of opposite sign are easily detected. The superposition of the CD curves of both transitions leads to a sigmoidal CD curve, a couplet, which can be positive ( $\Delta\epsilon(\bar{\nu}_1) > \Delta\epsilon(\bar{\nu}_2)$ ;  $\bar{\nu}_1 < \bar{\nu}_2$ ) or negative ( $\Delta\epsilon(\bar{\nu}_1) < \Delta\epsilon(\bar{\nu}_2)$ ;  $\bar{\nu}_1 < \bar{\nu}_2$ ). The couplet changes its sign when the angle between the transition moments of the monomers crosses the value of  $\sim 110^\circ$ . The amplitude of the couplet for the angle of the crossing point is zero. This results from an accidental degeneration of the excited states  $|K_1\rangle$  and  $|K_2\rangle$  by which both exciton CD bands compensate each other (crossing of the energies of both states). Determining the direction of the transition moments of the exciton transitions by polarized spectroscopy allows a reliable assignment of the transitions, in both cases.

For a quantitative calculation with eqn [32], a suitable interchromophoric distance has to be determined. Theoretically this distance is determined within the point dipole approximation where it has been assumed that  $|\underline{R}_{ij}|$  is large in comparison to the diameter of the monomeric units. In spite of the fact that the latter assumption is mostly not fulfilled, the exciton model has been applied with large success.

**Electronic circular dichroism for compounds without a chromophore** To determine the absolute configuration of compounds without a chromophore in the accessible spectral region one or more proper groups can be derivatized to obtain an appropriate chromophoric system. The substitution of dihydroxysteroids with two *p*-methoxybenzoates can be mentioned as an example. The substitution leads to an inherent dissymmetric chromophore for which the exciton chirality method can be applied. For chiral diols and amino alcohols another interesting possibility does exist. Complexing of these diols and amino alcohols with a transition metal



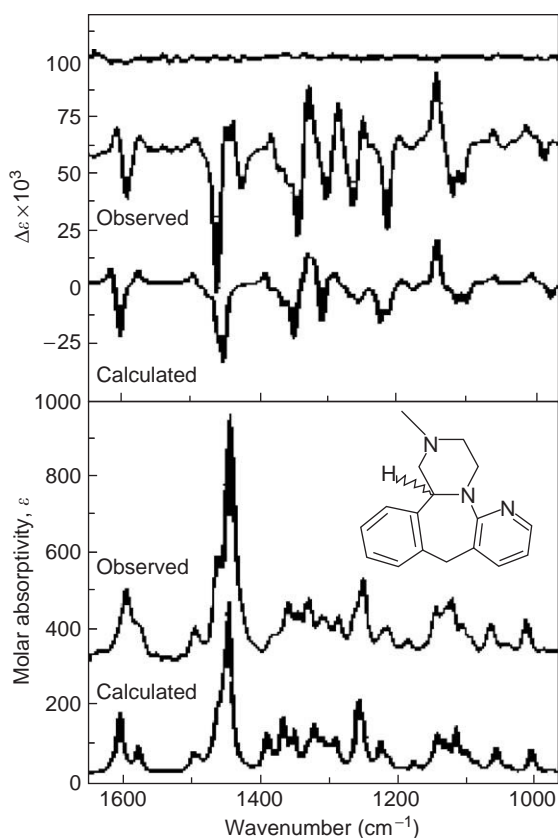
**Figure 9** The absolute configuration of *vic*-amino alcohols. The CD and UV spectra of *R*-alaninol (---), *R*-leucinol (.....), and *R*-phenylalaninol (—) in the presence of  $[\text{Mo}_2(\text{OAc})_4]$  referred to the total concentration of the *vic*-amino alcohols in DMSO at room temperature. The positive (negative) sign of the CE,  $\sim 330\text{ nm}$ , is related to the negative (positive) torsional angle of the N–C–C–O subunit. (Courtesy of Frelek J and Ruśkowska P, unpublished results.)

like dimolybdenum or dirhodium tetraacetate ( $\text{Me}_2(\text{O}_2\text{CCH}_3)_4$ ) leads to a chiral complex with a number of d–d and charge transfer transitions. The CD of the d–d transitions is determined by the absolute configuration of the diol and aminoalcohol units. The weakly absorbing chromophore is probably inherent dissymmetric with an allowed magnetic dipole transition (Figure 9) and, thus, easy to measure. One of the advantages of this method is that the complexes need not be isolated because only the signs of one or more of the CEs are needed. The disadvantages are the restriction of the rule to special classes of compounds and the often unknown assignment of the d–d transitions in the used spectral region.

**Quantum mechanical computation for electronic circular dichroism and vibrational optical activity** With the exception of the exciton chirality method for the development of sector and helicity rules, reference compounds are needed. Their absolute configuration has been usually obtained by X-ray analyses. The progress in numerical techniques of quantum mechanics nowadays allows performing computations of the rotational strength and its sign and, thus, the CD spectra with high reliability for CD

and also VCD for many classes of compounds. Here, the calculation of the rotational strength of more than one transition increases the reliability because not for all types of transitions the sign of the rotational strength can be obtained with the same quality. A large number of atoms of a molecule and the often existing large number of conformers also restrict the method. By a further improvement of the numerical techniques the application of empirical and semiempirical sector and helicity rules may hopefully decrease in their importance in the next years. For VCD spectroscopy, the empirical and semiempirical sector and helicity rules never have had the same importance as for electronic transitions. In Figure 10, an example for a computation is given in order to demonstrate the quality of calculated spectra.

The simplest form to characterize a chiral molecule was and is the specific optical rotation for one wavelength  $[\alpha]_D$ , e.g., in spite of the fact that  $[\alpha]_D$  is very sensitive against external influences because it is determined by a sum of many ORD sigmoid curves (eqn [21]; Figure 5), which belong to absorption bands lying far away, also in the UV spectral region. Until now it seemed that it is impossible to calculate  $[\alpha]_D^T$  with a sufficient reliability in order to use the specific rotation to more than a pure characterization



**Figure 10** The absolute configuration of mirtazapine from the comparison of the VCD and IR spectra of (–) mirtazapine and the calculated spectra (Gaussian 98, DFT level) of the *R*-enantiomer. The uppermost trace is the VCD noise. (Reproduced with permission from Freedman TB, Dukor RK, van Hoof PJCM, Kellenbach ER, and Nafie LA (2002) Determination of the absolute configuration of (–) mirtazapine using vibrational circular Dichroism. *Helvetica Chimica Acta* 85: 1160–1165.)

of a compound. But in the last years the quality of the computation of  $[\alpha]_D^T$  has much improved so that for some classes of compounds the absolute configuration was determinable. If this can be corroborated for diverse classes of compounds, the determination of the absolute configuration would be extremely simplified (see section on CD contra ORD).

#### Vibrational Circular Dichroism/Raman Optical Activity and Electronic Circular Dichroism – Partners or Opponents

The VCD, ROA, and ECD measure the same phenomenon. The ECD is well established since years whereas VCD and ROA are new and developing methods, especially also in their instrumentation. Therefore, the question arises whether the ECD can be superseded by VCD and RAO.

For the ECD only a small number of absorption bands are accessible in the UV/vis spectral region. Applying empirical rules and the results of the

theoretical computations with care allows determining the absolute configuration with high reliability. On looking closely one has to say that the local absolute configuration of the surroundings of the chromophore or of the chromophore itself is obtained for inherent symmetric and dissymmetric chromophores, respectively. Noncoupled or only weakly coupled overlapping chromophores probe an equal or similar surroundings. But as found for the  $n\pi^*$  and the  $\pi\pi^*$  transitions of the enone chromophore it happens that contradictory results are obtained from two chromophores. This shows the different sensitivities of different chromophores against small structural variations in their surroundings.

For the VCD and ROA a relatively large number of IR/Raman bands are accessible in the available spectral region. Normal vibrations are only very weakly coupled and thus, the information of the CD is localized to the surroundings of the normal coordinate. Because good empirical rules are not available, the quantum mechanical numerical computations are not an additional but, in general, the only information for the assignment of the absolute configuration. Restriction of the method lies within the scope of the quality of the calculation of the VCD/ROA spectra.

Comparing the information from ECD and VCD/ROA spectra is equivalent to a comparison of the different meanings of the VCD/ROA and the ECD chromophores as probes. The normal vibration is often the smaller chromophore. Thus, the information is more localized. But in contrast to the ECD, equal or similar VCD/ROA chromophores – functional groups in which a normal vibration is located – are often distributed over the skeleton of the whole molecule. In these cases the vibrational chromophores act as probes for a ‘mean’ geometry of the molecule whereas the ECD chromophore is a probe for itself or for its own surroundings. The correlation to the absolute configuration of the total molecule is unequivocal for ECD and VCD/ROA in most cases but for a number of questions their answers and the quality of answers are different. One obvious example is the different information of the analysis of vibrational progressions with the ECD and VCD. Whereas the ECD measures the effect of vibrations of the excited state, involved in the absorption process, the VCD yields information about vibrations of the electronic ground state. Thus, suitable chosen ECD and VCD are partners not opponents.

#### Conformational Analysis

The equilibria between chiral and achiral conformers of nonrigid molecules changes with temperature,



solvents, and by phase transitions because their energy difference  $\Delta E$  are in the order of  $k_B T$  ( $k_B$ , Boltzmann constant). Even compounds that are 'achiral' (racemic achiral), e.g., binaphthyl or chloroethane, possess chiral, or chiral and achiral conformers, respectively. ECD, VCD, and ROA have been proven to be very powerful techniques to determine these equilibria. *trans*-2-Chloro-5-methylcyclohexanone was the first example (Figure 7; originally via ORD spectra) where an equilibrium shift from the equatorial toward the axial conformation in a less polar solvent has been analyzed. The temperature dependence of  $\Delta\epsilon(T)$  has been used to evaluate the energies of conformers by a nonlinear fit of  $\Delta\epsilon(T)$  as a function of the Boltzmann factor  $\exp\{-\Delta E/k_B T\}$ . Nowadays an often applied method to handle the influence of conformations is the quantum mechanical calculation of the ECD, VCD, and ROA spectra taking the Boltzmann weighted distribution of conformers into consideration. Here, VCD seems to be superior to the ECD analysis. ECD and VOA are superior to analyses of absorption spectra, especially when the CD spectra of conformers have opposite signs by which the sensitivity of the method is increased.

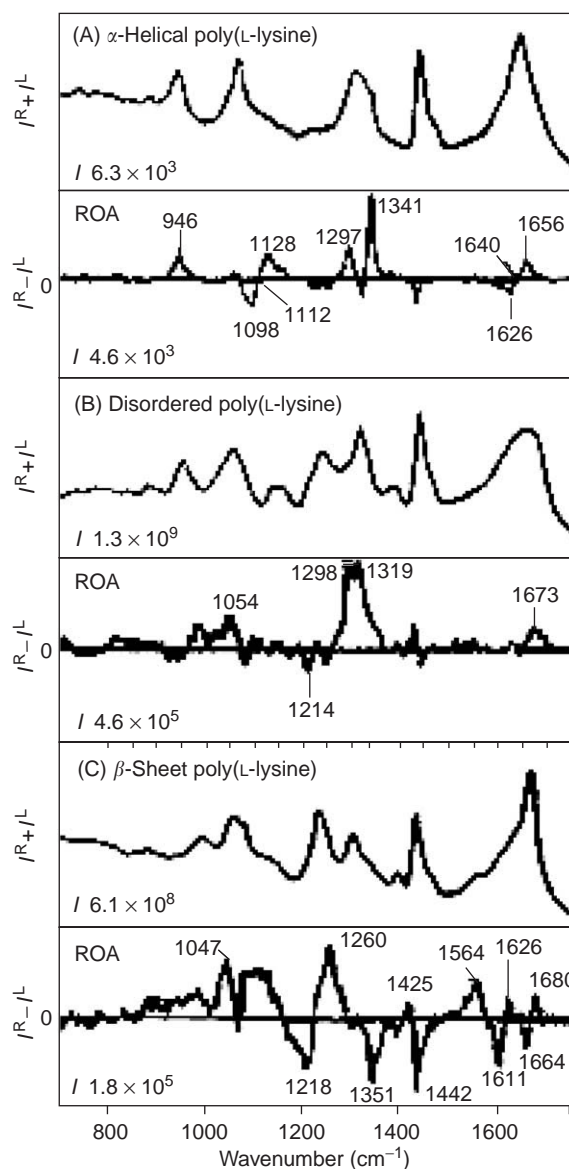
### Association and Aggregation

Association of molecules changes their  $\Delta\epsilon$  as has been proven for metal complexes, charge transfer complexes, Lewis acids, etc. In any case, complex formation can be followed by the variation of the CD. Molecular association is the first step to an aggregation and, furthermore, is the presupposition for supramolecular structures. Compounds with intermolecular exciton interaction, e.g., carotenoids or cyanine dyes, have been successfully analyzed. The spontaneous association to chiral associates from achiral compounds, e.g., cyanines, have been proven and analyzed with the help of ECD spectroscopy. In this context, the successful CD analyses with Langmuir–Blodgett films as well as the analyses of membranes have to be mentioned where suprastructural chirality can also emerge from achiral monomers.

### Macromolecules

Conformations and structures of polymers and biopolymers are accessible by ECD, VCD, and ROA spectroscopy. Suitable chromophores for the ECD are aromatic amino acids, the bases of nucleic acids, and the peptide bonds in the spectral region  $\sim 270$  nm and from 230 down to 160 nm. The determination of the  $\alpha$ -helix content of proteins was

one of the first successful applications. In addition to other features, VCD and ROA provide information about the secondary and tertiary structure of polypeptides or the base stacking of nucleic acids. As an example, ROA and Raman spectra of poly(L-lysine) in its  $\alpha$ -helical, disordered, and in the  $\beta$ -sheet structure are given in Figure 11. Even viral nucleic acids and proteins of intact viruses have been successfully identified in complex mixtures.



**Figure 11** ROA ( $I^R - I^L$ ) and backscattered Raman ( $I^R + I^L$ ) spectra of the  $\alpha$ -helical (A), disordered (B), and  $\beta$ -sheet (C) poly(L-lysine) at 20°C. (Reproduced with permission from Barron LD, Blanch EW, McColl IH, *et al.* (2003) Structure and behavior of proteins, nucleic acids and viruses from vibrational Raman optical activity. *Spectroscopy* 17: 101–126.)

### Pharmaceutical Applications

It is well known that right and left handed molecules of chiral drugs can differ dramatically in their pharmaceutical impact. Even a small racemic fraction in a chiral drug can be a dangerous impurity. Therefore, the determination of the optical purity is one of the most important applications of the chiroptical analysis. Specific techniques are needed in order to detect a very small percentage of a chiral compound in a mixture of other chiral compounds (see section on Polarimetry and CD detectors in liquid chromatography). A second analytical problem concerns the conversion of pure enantiomers into the other, often unacceptable form in a biological milieu. The high potential of VCD for determining the optical purity and the absolute configuration of newly synthesized chiral molecules intended for medical application has attracted increasing attention of this newer technique in the chemical and pharmaceutical industries (see section on Macromolecules).

### Time-Resolved Circular Dichroism

CD instruments have not primarily been developed for chiroptical kinetic analyses. In this context, racemization kinetics, complexing of metal ions, and also folding processes, e.g., of proteins, are an essential field of chiroptical analyses. The modulation technique as well as the scanning operations of the spectra restrict kinetic analyses to times not faster than some 10–100 ms. New developments, using diode array technologies, do not seem to be available at the moment. Instruments aimed for the nanosecond time regime are individual constructions generally incapable of and unavailable for systematic studies, nowadays, in spite of the fact that very interesting results can be expected by this technique.

### Indirect Methods of Chiroptical Analysis

The assignment of the absolute configuration with the help of sector and helicity rules is often seriously restricted. The assignment by quantum mechanical methods is now, as before, time consuming and not always available. Therefore, a way forward may have to be sought through new methods.

### NMR Measurements of Diastereomers with Substituents of Known Absolute Configuration

The NMR spectroscopy of diastereomeric compounds with unknown absolute configuration, derivatized with suitable substituents of known absolute configuration, has been introduced as a new

‘indirect’ chiroptical method that also allows determining the optical purity.

### Chiral Induction of Cholesteric Phases – Circular Dichroism and Selective Reflection

A chiral compound, dissolved in a nematic liquid crystal phase, transforms this phase into a chiral phase that is very often a chiral nematic – cholesteric – phase. Under the same condition of concentration and temperature two enantiomers induce helical structures with the same pitch but of opposite sign. The helical pitch  $p$  is for low concentrations of the dopant a linear function of mole fraction  $x$ . The molecular measure for the chiral induction is the helical twisting power (HTP):

$$\text{HTP} = px^{-1} \quad [33]$$

The pitch can be obtained by the CANO method or from the wavelength  $\lambda_0$  of the maximum of the selective reflection of circularly polarized light by

$$\lambda_0 = \bar{n}p. \quad [34]$$

$\bar{n} = (n_e + n_o)/2$  is an average of the refractive indices  $n_e$  and  $n_o$  of the ordinary and extraordinary rays. In spectral regions far away from the selective reflection a CD can be observed within the absorption bands of dopants in the cholesteric phase. Even achiral molecules, dissolved in a cholesteric phase, give rise to a CD. In spite of the complex origin of the HTP, correlations between the absolute configuration and the sign of the HTP have been found experimentally, especially for ‘form chiral’ molecules (inherent dissymmetric molecules). These phenomena may be an important field for prospective chiroptical analyses. Because of the broad research field about the chiral induction it is impossible to go into details, here.

## Analytical Applications

### Polarimetry

The unambiguous determination of enantiomeric excess and optical purity is an important task for the characterization and application of nonracemic chiral compounds. There are two different principles available: the determination of chiroptical properties such as CD, ORD, or the specific rotation  $[\alpha]_\lambda^T$  or to analyze the diastereomeric interaction with other chiral environments. The enantiomeric excess

$$\%ee = \frac{c_R - c_S}{c_R + c_S} 100 \quad [35]$$



can be obtained via the optical purity

$$P = \frac{[\alpha]^T}{[\alpha]_{\lambda, \max}^T} \quad [36]$$

if there is no diastereomeric interaction between the enantiomers in the enantiomeric mixture (e.g., complexing).  $c_R$  and  $c_S$  are the concentrations of the *R* and *S* enantiomer, respectively.  $[\alpha]^T$  is the specific rotation for the enantiomeric mixture and  $[\alpha]_{\lambda, \max}^T$  the specific rotation of the pure enantiomer.  $[\alpha]$  is a function of temperature, concentration (density of a neat liquid), and the solvent. The optical rotation is mostly measured far away from absorption bands – standard is the sodium D-line – in spite of the fact that there the optical rotation is small and often not very characteristic.

There are a number of requirements to the accuracy of the method because the two borderline cases, the racemic mixture (50:50) and the pure enantiomer (100%ee), have to be determined with high accuracy. The accuracy of a polarimetric measurement is  $\sim 10^{-3} - 10^{-4}$  degree. For small values of  $[\alpha]^T$  a suitable concentration and path length have to be chosen. For large values of  $[\alpha]_{\lambda, \max}^T$  any measured angle can be  $\alpha \pm 180n^\circ$ . To check  $n$ , a dilution experiment has to be performed. Furthermore, an independently proven absolute specific rotation  $[\alpha]_{\lambda, \max}^T$  has to be available. Also the purity of a compound is of importance. An enantiomeric excess counts as an impurity. The optical purity is only equal to the enantiomeric excess if and only if the conditions given above are fulfilled.

For a characterization within a series of compounds often polarimetry supersedes CD spectroscopy because the chosen concentration for its measurement is not restricted by a too high absorption. The broad acceptance of the polarimetric standardization of chiral compounds always has been a motivation for improvements of polarimeters. Especially for the daily routine work there is a requirement for instruments with high convenience (see section on Polarimetry and CD detectors in liquid chromatography).

### Polarimetry and Circular Dichroism Detectors in Liquid Chromatography

A polarimetric equipment is an important detector in liquid chromatography. The combination of an  $[\alpha]_{\lambda, \max}^T$  measurement combined with a detection of the UV absorption allows to determine the optical purity even in the case of an incomplete chromatographic enantiomeric separation. The use of CD as a chromatographic detector is limited in liquid

chromatography. One of the most serious experimental limitations is the very small CD signal because the concentrations must be kept very low in a chromatographic process and the path lengths, viewed across the exit tubes, is short. Furthermore, the scanning time for a spectrum has to be short in comparison to the retention time if the analysis needs a complete CD band in order to gain more information. In spite of these limiting facts the CD detector has been often applied with success. The measurement of the dispersion of  $\Delta\epsilon cd$ , or with less information the optical rotation, and the absorbency  $\epsilon cd$  have been used for optical purity online analyses by deconvolution of the overlapping CD peaks of enantiomers. Furthermore, enantiomerization during the chromatographic process can be analyzed. In order to overcome the limitation of the time-consuming registration of a spectrum a simultaneous measurements of a CD spectrum could be of interest (see section on Time-resolved circular dichroism).

### Properties of Chiral Phases

New and promising fields are the chiroptical analyses of chiral crystals, chiral liquid crystal phases, chiral membranes, Langmuir–Blodgett films, chiral surfaces, etc. Their chirality is caused by the molecular chirality and the suprastructural chirality of the phase. Because the spatial extension of the structures of the suprastructural chirality is often about three orders of magnitude larger than molecular diameters, the spectral regions of chiroptical phenomena for both levels of chirality are different.

From our daily life it is well known that chiral objects like spiral staircases or flowers look different from different directions. Therefore, it seems evident that chiral molecules and chiral suprastructural phases are anisotropic, too. Until now, chiroptical methods have been only very sparsely applied to chiral anisotropic systems because of serious experimental problems. At first one has to draw attention to artifacts induced by the always existent linear dichroism and birefringence of anisotropic systems (elliptical dichroism and birefringence). Secondly, objects without symmetry do not allow to measure directly ‘chiral and achiral anisotropies’ without additional requirements. That is, new techniques and unequivocal definitions are needed to decompose the results of measurements with chiral anisotropic phases into ‘chiral’ and ‘achiral’ components. A minimum of symmetry is needed to adapt suitable situations where the CD and ORD as chirality measurements of anisotropic systems can be observed directly.

### Pretransitional Phase Effects

In the isotropic phase near the phase transition to a chiral liquid crystal phase a strong increase of the optical rotation has often been observed. The so-called pretransitional phase effect announces a structural change by approaching the phase transition temperature. Corresponding CD measurements do not exist because the extreme increase of the optical rotation has been measured always with neat compounds where large absorption prohibits CD measurements. For induced chiral phases with low concentrations of chiral dopants pretransitional phase effects have not been observed.

### Circular Dichroism of Chiral Anisotropic Phases without Suprastructural Chirality

The chiral anisotropy of a phase is the consequence of the anisotropy of the results of chirality measurements with oriented molecules, i.e., information about chiral molecules when viewed from different directions. This is of interest, e.g., when two molecules, oriented with respect to each other, interact in an asymmetric synthesis.

A sufficient amount of oriented chiral molecules can be obtained in an induced cholesteric liquid crystal phase if the induced helical structure has been untwisted by an electric field. In the following description tensors are needed for the sake of simplicity (At least there are three tensors required: the transition moment tensor  $D_{ij}$  (absorption tensor  $\varepsilon_{ij}$ ), the rotational strength tensor  $R_{ij}$  (circular dichroism tensor  $\Delta\varepsilon_{ij}$ ), and the order tensor  $g_{ij33}$  ( $i, j = 1, 2, 3$ ). If the molecules do not possess any symmetry, the principal axes of all of these tensors are differently oriented with respect to the molecular frame (the coordinate system in which only the three diagonal elements of a tensor are different from zero.) The only tensorial property, needed here explicitly, is the existence of three coordinates (components) of a tensor with respect to three specially chosen mutually perpendicular axes. This means that three information instead of one information about a molecule are needed: instead of one CD value, namely  $\Delta\varepsilon$ , three CD values, namely  $\Delta\varepsilon_{ii}$  ( $i = 1, 2, 3$ ), have to be introduced.  $\Delta\varepsilon$  is then one-third of a sum of the three so-called tensor coordinates of the CD tensor:

$$\Delta\varepsilon = \frac{1}{3}(\Delta\varepsilon_{11} + \Delta\varepsilon_{22} + \Delta\varepsilon_{33}) \quad [37]$$

A coordinate  $\Delta\varepsilon_{ii}$  for  $i = 1, 2$ , or  $3$  is the result of a CD measurements with a light beam propagating along the axis  $i$  of the molecule. In addition, all molecules in each measurement have to be parallelly oriented with respect to the axis  $i$  and the ensemble

of the molecules have to be rotationally symmetrically distributed about the chosen axis  $i$ . The latter condition is fulfilled with a light beam propagating parallel to the optical axis of a uniaxial phase. It is not fulfilled for a light beam propagating perpendicular to its optical axis. Here the result is of ‘mixed symmetry’, i.e., a mixture of a scalar (achiral) and a pseudoscalar (chiral) information.

The CD of a uniaxial phase with a light beam propagating parallel to the optical axis of the phase, the ACD, is then given by

$$\Delta\varepsilon^A = g_{1133}^* \Delta\varepsilon_{11}^* + g_{2233}^* \Delta\varepsilon_{22}^* + g_{3333}^* \Delta\varepsilon_{33}^* \quad [38]$$

where the orientational distribution coefficients  $g_{ii33}^*$  are measures of the order of the molecules in a uniaxial phase of partially oriented molecules (eqns [39] and [40]). The orientational distribution coefficients are correlated with the earlier introduced Saupe order parameters by

$$S^* = \frac{1}{2}(3g_{3333}^* - 1) \quad \text{and} \quad D^* = \frac{\sqrt{3}}{2}(2g_{2233}^* + g_{3333}^* - 1) \quad [39]$$

In addition the following relations hold:

$$0 \leq g_{ii33} \leq 1 \quad \text{for } i = 1, 2 \text{ or } 3 \quad \text{and} \quad \sum_{i=1}^3 g_{ii33}^* = 1 \quad [40]$$

The convention  $g_{3333}^* \geq g_{2233}^* \geq g_{1133}^*$  allows to introduce the ‘orientation axis ( $x_3^*$ )’ of a molecule, i.e., ‘the best ordered molecular axis in an anisotropic phase’. The  $g_{ii33}^*$  ( $i = 1, 2, 3$ ) are given in their principal axes, otherwise nondiagonal elements of  $\Delta\varepsilon_{ij}$  ( $i = 1, 2, 3$ ) are required in eqn [38]. The  $\Delta\varepsilon_{ij}^*$  are complicated functions of the involved electric and magnetic dipole and the electric quadrupole transition moments. The contribution of the electric dipole, magnetic dipole, and the electric quadrupole transition moments can be obtained for the transition  $|0\rangle \rightarrow |n\rangle$  directly by

$$\Delta_1 = \frac{1}{2}(\Delta\varepsilon_{11}^* - \Delta\varepsilon_{22}^* - \Delta\varepsilon_{33}^*) = -\frac{1}{2}B\bar{v} \sum_n \text{Im}\{\langle \mu_1 \rangle_{0n} \times \langle m_1 \rangle_{n0} \} F^{0n}(\bar{v}) + \Delta\varepsilon_{11}^{\mu Q} \quad [41]$$

$$\Delta_2 = \frac{1}{2}(\Delta\varepsilon_{22}^* - \Delta\varepsilon_{11}^* - \Delta\varepsilon_{33}^*) = -\frac{1}{2}B\bar{v} \sum_n \text{Im}\{\langle \mu_2 \rangle_{0n} \times \langle m_2 \rangle_{n0} \} F^{0n}(\bar{v}) + \Delta\varepsilon_{22}^{\mu Q} \quad [42]$$

$$\Delta_3 = \frac{1}{2}(\Delta\varepsilon_{33}^* - \Delta\varepsilon_{11}^* - \Delta\varepsilon_{22}^*) = -\frac{1}{2}B\bar{v} \sum_n \text{Im}\{\langle \mu_3 \rangle_{0n} \times \langle m_3 \rangle_{n0} \} F^{0n}(\bar{v}) + \Delta\varepsilon_{33}^{\mu Q} = -\frac{3}{2}\Delta\varepsilon - \Delta_1 - \Delta_2 \quad [43]$$

The three diagonal elements  $\Delta\epsilon_{ii}^{\mu Q}(\bar{\nu})$  ( $i = 1, 2, 3$ ) are proportional to products of electric dipole times electric quadrupole transition moments. They do not contribute to the isotropic CD because the sum over the three coordinates  $\Delta\epsilon_{ii}^{\mu Q}(\bar{\nu})$  ( $i = 1, 2, 3$ ) is zero.  $\Delta\epsilon_{ii}$ , measured for oriented guest molecules in ordered liquid crystal phases, yield spectroscopic and structural information and, has been used, especially for the check of sector and helicity rules. First numerical quantum mechanical calculations of the CD tensor coordinates  $\Delta\epsilon_{ii}$  have been published recently.

### Chiral Crystals

Solid-state CD spectra provide information on chiral conformations, suprastructural chirality of solid phases, packing effect in crystals, and absolute configuration of the molecular and the phase chirality. Especially the anisotropy of chirality measurements – the pseudoscalar information – of small molecules and also biopolymers has been obtained, which are completions to the results of ACD measurements of noncrystalline phases with oriented molecules (see section on Circular dichroism of chiral anisotropic phases without suprastructural chirality). For the general case of crystals without or with low symmetry the measured CD is a mixture of achiral and chiral information, in principle – a function of a pseudoscalar and a scalar contribution – because the eigenstates of light are elliptically polarized. In this context one should remember that also for achiral crystals ( $m$ ,  $mm2$ ,  $\bar{4}$ ,  $\bar{4}2m$ ) or molecules with a special long-range order a CD can be obtained, which is by no means caused by chirality. In order to avoid mixed information there are at least three possibilities:

1. To measure a crystal – or general any phase – along the optical axis of a uniaxial phase where the eigenstates are circularly polarized.
2. To measure a powdered crystalline material with isotropically distributed crystallites, e.g., pressed in KBr or as a suspension in Nujol.
3. To measure along an arbitrary direction apart from the optical axis/axes. Here a technique is needed which allows to decompose the result into its pseudoscalar (chiral) and scalar (achiral) contribution.

**Isotropically distributed powdered crystals** Powdered crystalline material pressed in KBr or as a suspension in Nujol with isotropically distributed chiral crystals should yield the CD of isotropically distributed molecules and the CD of the isotropically distributed ‘crystals with their suprastructural

chirality’. Until now, a decomposition into these two effects has not been performed. One of the most difficult problems for such CD measurements is an always existing residue of anisotropy by incompletely distributed powdered material or by scattering of crystallite surfaces, which leads to an azimuth dependent CD of the sample. However, for many compounds the CD has been successfully analyzed with this technique. In spite of the success of this method there are a number of unresolved problems. The measured ‘isotropic CD’ of these crystal powders has been corrected with equations derived with the Jones Calculus for homogeneous material and for coherent light waves. Furthermore, no contribution of a tensorial contribution of the anisotropy of the CD of crystallites has been taken into account. Here, artifacts originated by the artificial anisotropy of grounded material and their inhomogeneities need further considerations.

**Crystals** Measurements with enantiomorphic crystals along an optical axis have often been reported in biographies of crystal optics. Measuring the ORD and CD for uniaxial and biaxial crystals in any direction apart from the optical axis (axes) are up to now one of the most difficult experiments (see section on Experimental equipments). J. Kobayashi’s extended HAUP method is time consuming but very successful. CD, CB, linear dichroism (LD), and linear birefringence (LB) can be obtained in one and the same experiment. Most of the information has been obtained only for a single wavelength because no suitable equipment is available to determine the dispersion automatically. Besides optically active inorganic crystals, crystals of small chiral molecules and especially biopolymers have been analyzed as, e.g., crystalline glyoxylamide,  $\beta$ -lactams, aspartic acid, glutamic acids, poly-L-lactic acid, and lysozyme. The effect of freezing of conformations in the crystalline state and the anisotropy of the chirality measurement of an oriented helix strand have been proven.

### Circular Dichroism of Chiral Anisotropic Phases with Suprastructural Chirality

**Amorphous phases** A new field of applications of chiroptical analyses is the search of chiral structures in Langmuir–Blodgett films or thin films of polymers, membranes, and ‘chiral surfaces’, surfaces on which a few chiral molecules are adsorbed. The recently published technique of reflectivity CD may allow systematic analyses of chiral metal surfaces. This field is too new to be summarized and, therefore, some selected papers have been cited in the Further reading section.

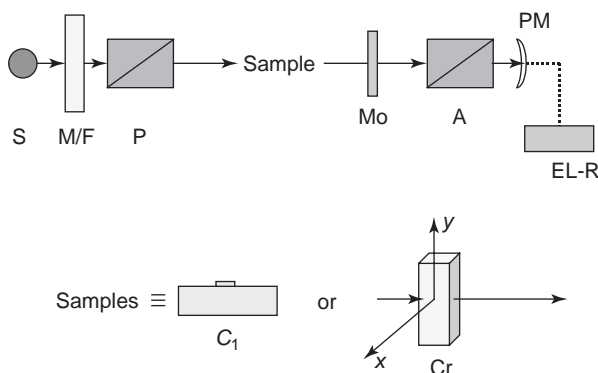
**Liquid crystals** There is a broad field for chiroptical analyses with chiral liquid crystal phases. Because it is beyond the scope of this article to discuss all types of chiral liquid crystals, only two chiral phases with a simple suprastructural chirality will be mentioned here, which are of special interest also for display technology. First, there is the chiral nematic  $N^*$ -cholesteric phase mostly used in twisted nematic (TN) displays. Here, the long-range orientational order does possess a helical structure of chiral and in an induced phase of chiral and achiral molecules. Their longest axes, their orientation axes, are oriented in the mean perpendicular to the helix axis. The helix is then obtained by the periodic rotation of the orientation axes of the molecules about the helix axis. Here again the two levels of chirality, the molecular and the suprastructural phase chirality, are involved. In the second important phase, the smectic  $C^*$  phase, a helical order is produced by molecules that are tilted with their long axis (orientation axis) against the normal of the smectic layers. The chiral structure of the phase is obtained by a periodical rotation of the tilt angle in the smectic layers about the optical axis – the helix axis – of the phase.

Until now, only a few chiroptical analyses have been performed with liquid crystal phases. This has again its origin in the experimental problems of CD measurements with anisotropic phases. Besides all those problems that appears in context with ORD and CD measurements of crystals with low symmetry (see section on HAUP method) one has to overcome the falsification of results by the intense scattering of the liquid crystal phases. To some extent this scattering can be subdued by measuring thin films with a thickness in the order of micrometers.

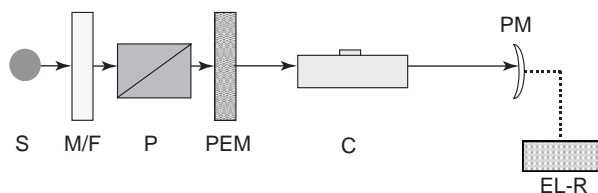
## Experimental Equipment

Either the difference of two absorption coefficients ( $\Delta\epsilon$ ) and the difference of two refractive indices ( $\Delta n$ ) of the circularly polarized eigenstates or the state of polarization – the ellipticity and the optical rotation – of the light beam leaving the sample have to be measured. Nowadays, experimentally preferred are the determinations of  $\Delta\epsilon$  and the optical rotation.

The optical rotation can be obtained from the change of the azimuth angle of an analyzer (Figure 12) in an originally crossed polarizer–analyzer system. The half-shade device of historical instruments is substituted by a modulator Mo for the state of polarization, which is transformed into an intensity modulation by a following analyzer. The accuracy with common available polarimeters and spectral polarimeters lies between  $10^{-4}$  and  $10^{-3}$  degrees.



**Figure 12** Block diagram of a spectral polarimeter (a), a polarimeter (b), and the HAUP spectrometer (c). (S = source of radiation: xenon lamp (a); sodium or mercury lamp (b), laser (c); M/F = monochromator (a) or filter (b, c); P = polarizer (a, b, c); c = sample cell C<sub>1</sub> (a, b) or crystal plate Cr (c); Mo = modulator: half-shadow equipment (a, b); A = analyzer (a, b, c); PM = photomultiplier (a, b, c); EL-R = electronic device/computer/recorder (a, b, c).)



**Figure 13** Block diagram of a CD spectrometer with a periodic variation of the state of polarization and by this an intensity modulation by the absorption difference of the sample for the UV/vis (a) and the IR (b) spectral region if the optical elements are suitable chosen (S = source of radiation; M/F = monochromator (a, b), or Fourier transform interferometer (b); P = polarizer; PEM = photoelastic modulator; C = sample cell; PM = photomultiplier (a, b) or another detector (b); EL-R = electronic equipment/computer/recorder).

ECD and VCD spectrometers modulate (Mo) the state of polarization of the incident light. Through CD of the sample the polarization modulation is transformed into an intensity modulation (Figure 13). The amplitude of the modulated intensity is proportional to the CD. Nowadays, ECD spectrometers possess a fully developed technology. Therefore, only about a few instrumental developments have been reported recently. There are efforts to subdue the sensitivity to linear dichroism (LD) and birefringence (LB) in order to measure the CD of chiral anisotropic crystals (ACD). But the use of a selected artifact-free optical modulator (Mo) and a rotation of the sample were until now only of limited success. In order to gain intensity in an ECD spectrometer, both orthogonally polarized light beams leaving the polarizing system P (Figure 13) have been used to measure the CD signal. The symmetry of both beams can be used to suppress artifacts induced



by anisotropic samples. The dissymmetry factor  $g$  (eqns [29] and [30]), a reliable measure for an estimate of the sensitivity of CD instruments, lies for the UV/vis region between  $10^{-7}$  and  $10^{-6}$  whereas for the IR region a value  $\sim 10^{-6}$  can be obtained.

To increase the sensitivity of FD CD fluorescence spectrometer a new equipment for the excitation of the sample with left ( $I_L$ ) and right ( $I_R$ ) circularly polarized light (eqn [44]) has been developed by the use of an ellipsoidal mirror for the exciting light and a polarizing equipment for the emitted light. The contribution of 'wrongly' polarized light has been strongly reduced. And, thus the results essentially improved:

$$\frac{I_L - I_R}{I_L + I_R} \quad [44]$$

VCD instruments, based also on the principles depicted in **Figure 13**, are in comparison to the ECD instruments, relatively new products on the market. Thus, there have been further developments. Two of them should be mentioned, namely, the use of the Fourier transform technology and the use of a second PEM with and without a second polarizing element. This 'dual polarization modulation (DPM)' method increases the quality of the VCD spectra. Also, the Raman spectrometers for ROA, with which the difference of scattered left and right circularly polarized Raman light is measured, have been developed to a standard that allows systematic chiroptical analyses, nowadays.

The determination of the optical constants CD, CB, LD, and LB of crystals of low symmetry in any direction apart from the optical axis/axes allows the high-accuracy universal polarimeter (HAUP) consisting of a polarizer, an analyzer (**Figure 12**) and as for any polarimeter an 'electronic half-shade' device as a modulator to increase the signal to noise ratio. A plane-parallel plate of a crystal is oriented with its linear azimuth, symmetry azimuth, or minimum azimuth (Using the nomenclature of G. Szivessy and Cl. Münster. These three azimuths converge to the azimuth of the eigenpolarization of, e.g., a uniaxial crystal.) approximately parallel to the almost crossed polarizer and analyzer. The fit of the intensity as a function of two angles, the angles between the polarizer and the analyzer and, e.g., the minimum azimuth of the crystal, respectively, allows determining all four optical constants. Artifacts induced by the experimental equipment can be measured independently without the crystal plate. The time consuming technique impeded the applicability of the HAUP instrument to measure the wavelength dependence which is often needed for an interpretation of molecular systems.

**See also:** **Infrared Spectroscopy:** Overview. **Liquid Chromatography:** Overview. **Raman Spectroscopy:** Instrumentation.

## Further Reading

- Agranat I, Caner H, and Caldwell JP (2002) Chirality to work: the strategy of chiral switches. *Nature Reviews Drug Discovery* 1: 753–768.
- Asahi T and Kobayashi J (2003) Polarimeter for anisotropic optically active materials. In: Weiglhofer WS and Lakhtakia A (eds.) *Introduction to Complex Mediums for Optics and Electromagnetics*, pp. 645–676. Bellingham, WA: SPIE Press.
- Berova N, Nakanishi K, and Woody RW (eds.) (2000) *Circular Dichroism: Principles and Applications*. New York: Wiley.
- Chalmers JM and Griffiths PR (eds.) (2002) *Handbook of Vibrational Spectroscopy*, vol. 1. Chichester: Wiley.
- Demus D, Goodbey JW, Gray GW, Spiess HW, and Vill V (eds.) (1998) *Handbook of Liquid Crystals, Four Volume Set*. Chichester: Wiley.
- Hicks JM (ed.) (2002) *Chirality: Physical Chemistry*, ACS Symposium Series 810. Washington DC: American Chemical Society.
- Kiesswetter R, Brandl F, Kastner-Pustet N, and Mannschreck A (2003) Chiroptical detection during liquid chromatography: deconvolution of overlapping peaks of enantiomers and its applications. *Chirality* 15: 40–49.
- Kitzerow H-S and Bahr C (eds.) (2001) *Chirality in Liquid Crystals*. Berlin: Springer.
- Kuball H-G and Höfer T (1999) Chiroptical spectroscopy oriented molecules and anisotropic systems. In: Lindon JC, Tranter GE, and Holmes JL (eds.) *Encyclopedia of Spectroscopy and Spectrometry*, pp. 267–281. London: Academic Press.
- Lightner DA and Gurst JE (2000) *Organic Conformational Analysis and Stereochemistry from Circular Dichroism Spectroscopy*. New York: Wiley-VCH.
- Michl J and Thulstrup EW (1986) *Spectroscopy with Polarized Light – Solute Alignment by Photoselection, in Liquid Crystal, Polymers, and Membranes*. New York: VCH Publishers.
- Nehira T, Parish CA, Jockusch S, et al. (1999) Fluorescence-detected exciton-coupled circular dichroism: scope and limitation in structural studies of organic molecules. *Journal of the American Chemical Society* 121: 8681–8691.
- Polavarapu PL (2002; 2003) Optical rotation: recent advances in determining the absolute configuration. *Chirality* 14: 768–781; 15: 284–285.
- Snatzke G (1991) Helicity of molecules different definitions and application to circular dichroism. In: Janoscheck R (ed.) *Chirality – From Weak Bosons to the  $\alpha$ -Helix*, pp. 59–85. Berlin: Springer.
- Zsila F, Bikadi Z, Deli J, and Simonyi M (2001) Chiral detection of carotenoid assemblies. *Chirality* 13: 446–453.

# CHLOROFLUOROCARBONS AND OTHER HALOCARBONS

**D E Oram**, University of East Anglia, Norwich, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A wide variety of C1–C2 halocarbons have been identified in the Earth's atmosphere at concentrations ranging from a few to several hundred ppt (parts per trillion, or parts in  $10^{12}$ ). Many of the species identified are of natural origin, being produced in the oceans (e.g.,  $\text{CH}_3\text{Cl}$ ,  $\text{CH}_3\text{Br}$ ,  $\text{CH}_3\text{I}$ ,  $\text{CHBr}_3$ ), from the natural burning of biomass ( $\text{CH}_3\text{Cl}$ ,  $\text{CH}_3\text{Br}$ ), and even from geochemical processes ( $\text{CF}_4$ ). With the exception of  $\text{CH}_3\text{Cl}$ , which is present at  $\sim 500$  ppt, the natural background mixing ratios of these species are quite low, typically less than 10 ppt, and consequently their contributions to global processes such as stratospheric ozone depletion and global warming are relatively small. However, it has become apparent that the halocarbon content of the lower atmosphere has been significantly perturbed during the twentieth century as a result of expanding production and emission of an ever-increasing number of synthetic compounds, most notably the chlorofluorocarbons or CFCs. Commercial production of CFCs began in the 1930s and they have subsequently found widespread use in refrigeration, foam blowing, and aerosol propellant applications. Other halocarbons have been used as industrial solvents, chemical feedstocks, and as cleaning/degreasing agents ( $\text{CH}_3\text{CCl}_3$ ,  $\text{CCl}_4$ ,  $\text{CH}_2\text{Cl}_2$ ,  $\text{C}_2\text{Cl}_4$ ), for fire-fighting ( $\text{CBrClF}_2$ ,  $\text{CBrF}_3$ ), and as fumigants in the agricultural industry ( $\text{CH}_3\text{Br}$ ). Several halocarbons are also emitted as by-products of industrial processes ( $\text{CF}_4$ ,  $\text{C}_2\text{F}_6$ ,  $\text{CHF}_3$ ). This article will focus on these synthetic halocarbons, discussing atmospheric sampling methods, analytical techniques, and their effect on stratospheric ozone and climate.

## Environmental Concerns

Concern over the build-up of halocarbons in the atmosphere began in the early 1970s when it was proposed that long-lived CFCs could be photodissociated in the stratosphere, releasing chlorine atoms capable of ozone destruction. Up until that time, being inert, nonflammable, and nontoxic, CFCs

had been considered benign and their production and range of applications had grown considerably since their introduction in the 1930s. By 1974, the combined production of the two major CFCs,  $\text{CCl}_2\text{F}_2$  (CFC-12) and  $\text{CCl}_3\text{F}$  (CFC-11), had reached over 800 000 metric tons per year. However, despite extensive research driven by considerable public and political concern, it was not until the mid-1980s that the first evidence of significant ozone depletion was found in the unexpected form of the Antarctic ozone hole. The ozone hole was inextricably linked to reactions involving chlorine and bromine atoms, and this led to the signing of the Montreal Protocol in 1987. This treaty, and subsequent amendments, imposed restrictions on the consumption of substances thought to contribute to ozone depletion, most notably the CFCs, halons, and the chlorinated solvents,  $\text{CH}_3\text{CCl}_3$  and  $\text{CCl}_4$ .

Following the Montreal Protocol, industry was forced to look for suitable alternatives to the banned compounds and amongst them the prime candidates were hydrochlorofluorocarbons (HCFCs) and hydrofluorocarbons (HFCs). Both types of compound contain at least one C–H bond that renders them susceptible to attack by hydroxyl radicals (OH) in the Earth's troposphere, thereby shortening their atmospheric lifetimes and reducing their potential threat to stratospheric ozone. Although their ozone depletion potentials (ODPs) are generally much smaller than those of the CFCs, HCFCs still contain chlorine and are themselves regulated under the Montreal Protocol, with phase-out scheduled for 2030. HFCs are generally regarded as more acceptable replacements for CFCs as they contain no chlorine. Under the auspices of AFEAS (Alternative Fluorocarbon Environmental Acceptability Study) and PAFT (Programme for Alternative Fluorocarbon Toxicity Testing), the environmental impact of HCFCs, HFCs, and their degradation products was thoroughly investigated before significant commercialization began (see Table 1).

Another group of chemicals regulated by the Montreal Protocol is the halons. Halons are fully halogenated, bromine-containing, CFC analogs, used as fire extinguishing agents. Despite their relatively low abundance, halons are an important source of stratospheric bromine. Atom for atom, bromine is 40–100 times more efficient than chlorine at depleting ozone, and is involved in 25–50 and 10% of

**Table 1** Atmospheric lifetime, ozone depletion and global warming potentials of the major atmospheric halocarbons

Formula	Common name	Lifetime (years)	ODP <sup>a</sup>	GWP <sup>b</sup>
CCl <sub>3</sub> F	CFC-11	45	1.0	4 680
CCl <sub>2</sub> F <sub>2</sub>	CFC-12	100	1.0	10 720
CClF <sub>2</sub> CCl <sub>2</sub> F	CFC-113	85	1.0	6 030
CBrF <sub>3</sub>	Halon-1301	65	12	7 030
CBrClF <sub>2</sub>	Halon-1211	16	6.0	1 860
CHClF <sub>2</sub>	HCFC-22	12.0	0.05	1 780
CH <sub>3</sub> CClF <sub>2</sub>	HCFC-142b	17.9	0.07	2 270
CH <sub>3</sub> CCl <sub>2</sub> F	HCFC-141b	9.3	0.12	713
CHF <sub>3</sub>	HFC-23	270	<0.0004	12 240
CF <sub>3</sub> CH <sub>2</sub> F	HFC-134a	14.0	<0.00002	1 320
CF <sub>4</sub>	FC-14	50 000	–0	5 820
C <sub>2</sub> F <sub>6</sub>	FC-116	10 000	–0	12 010
SF <sub>6</sub>	Sulfur hexafluoride	3 200	–0	22 450
CH <sub>3</sub> Cl	Methyl chloride	1.3	0.02	17
CH <sub>3</sub> Br	Methyl bromide	0.7	0.38	5
CH <sub>3</sub> CCl <sub>3</sub>	Methyl chloroform	5.0	0.12	144
CCl <sub>4</sub>	Carbon tetrachloride	26	0.73	1 380

<sup>a</sup> Ozone depletion potential relative to CFC-11.<sup>b</sup> Global warming potential relative to CO<sub>2</sub> for a 100-year time frame.Adapted from WMO (World Meteorological Organisation) (2003) *Scientific Assessment of Ozone Depletion: 2002*. Global Ozone Research and Monitoring Project, Report No. 47, Geneva.

ozone loss in polar regions and mid-latitudes, respectively. Halons account for 30–40% of bromine in the stratosphere. The other major source of stratospheric bromine (~50%) is methyl bromide (CH<sub>3</sub>Br), a chemical used as an agricultural fumigant. Despite also having significant natural sources and a relatively short lifetime (<1 year), CH<sub>3</sub>Br was added to the Montreal Protocol in 1992.

A second concern regarding the accumulation of halocarbons in the atmosphere is their potential effect on climate. Many halocarbons are efficient absorbers of infrared radiation, particularly in the atmospheric ‘window’ region (800–1200 cm<sup>–1</sup>) where naturally occurring species such as CO<sub>2</sub> and water vapor do not absorb strongly. In the atmosphere these compounds trap radiation emitted from the Earth’s surface, thereby contributing to the warming of the lower atmosphere, the so-called greenhouse effect. With their long atmospheric lifetimes certain halocarbons, most notably the perfluorocarbons (PFCs) and SF<sub>6</sub>, are among the most potent greenhouse gases known, with global warming potentials (GWPs) several orders of magnitude higher than that of CO<sub>2</sub>. PFCs, HFCs, and SF<sub>6</sub> do not participate in ozone depletion so are not controlled by the Montreal Protocol, but they have recently been included in the Kyoto Protocol, which aims to reduce emissions of greenhouse gases by 2008–2012.

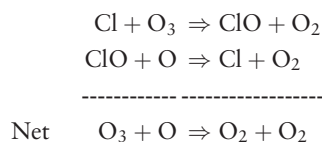
### Stratospheric Ozone Depletion

In the troposphere, CFCs are unreactive and are not susceptible to the normal degradation mechanisms

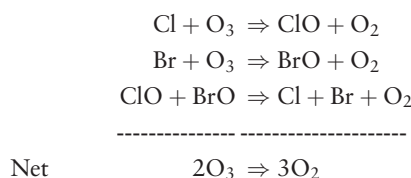
such as reaction with OH radicals, rainout, or surface deposition. Furthermore, the short wavelength ultraviolet radiation required to break C–Cl bonds does not penetrate into the troposphere, being absorbed at higher altitudes by ozone and molecular oxygen. Consequently, CFCs have long atmospheric lifetimes and are slowly transported to the stratosphere where photolysis can occur, for example:



Cl atoms can participate in catalytic cycles leading to ozone destruction. Examples include:



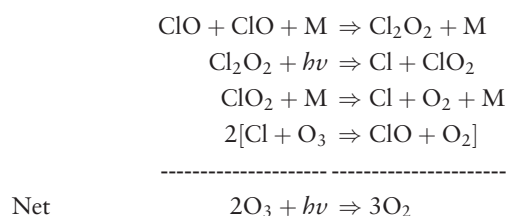
and



Observed ozone losses are highly variable in both space and time. The global average total column ozone in the late 1990s was ~3% below pre-1980



values. However, much greater losses are seen in polar regions, particularly in Antarctica, where losses of up to 70% have been recorded. During austral winter, a vortex of cold, sinking air forms over Antarctica, which persists for several months. Temperatures inside the vortex become sufficiently cold for polar stratospheric clouds to form and reactions on the surface of these clouds convert unreactive chlorine reservoirs (HCl, ClONO<sub>2</sub>) into photolytically labile species (Cl<sub>2</sub>, HOCl). Although stable in the dark winter conditions, Cl<sub>2</sub> and HOCl are rapidly photolyzed in spring releasing Cl atoms. The most important catalytic cycle at polar sunrise, responsible for ~70% of ozone loss, involves the Cl<sub>2</sub>O<sub>2</sub> dimer:



## Halocarbon Production, Applications, and Sources

CFCs were developed in the late 1920s by scientists at General Motors/Frigidaire as a safe alternative to traditional refrigerants such as sulfur dioxide and ammonia. Being nontoxic, nonflammable, noncorrosive, energy efficient, and inexpensive, CFCs proved to be ideal and soon captured much of the global refrigeration and air conditioning market. They have also found widespread use as foam blowing agents, aerosol propellants, and chemical solvents (see Table 2). Global production of CFCs reached a maximum in the late 1980s but has declined rapidly in response to the Montreal Protocol. In the developed world, production of the three most widely used CFCs (11, 12, 113) fell from over 1000 Gg in 1987 to just 30 Gg by 2001.

HCFC-22 (CHClF<sub>2</sub>) has been widely used since the 1930s for refrigeration, air conditioning, and as a chemical feedstock. Annual production continued to rise throughout the latter part of the twentieth century reaching over 250 Gg by the mid-1990s. Production of other HCFCs, most notably CH<sub>3</sub>CClF<sub>2</sub> (HCFC-142b) and CH<sub>3</sub>CCl<sub>2</sub>F (HCFC-141b), began in earnest in the 1990s coinciding with the phase out of CFCs. Global production of HCFC-141b, for example, rose from ~0.1 Gg in 1990 to 135 Gg in 2000. HCFCs 142b and 141b are primarily used as foam blowing agents (see Figure 1).

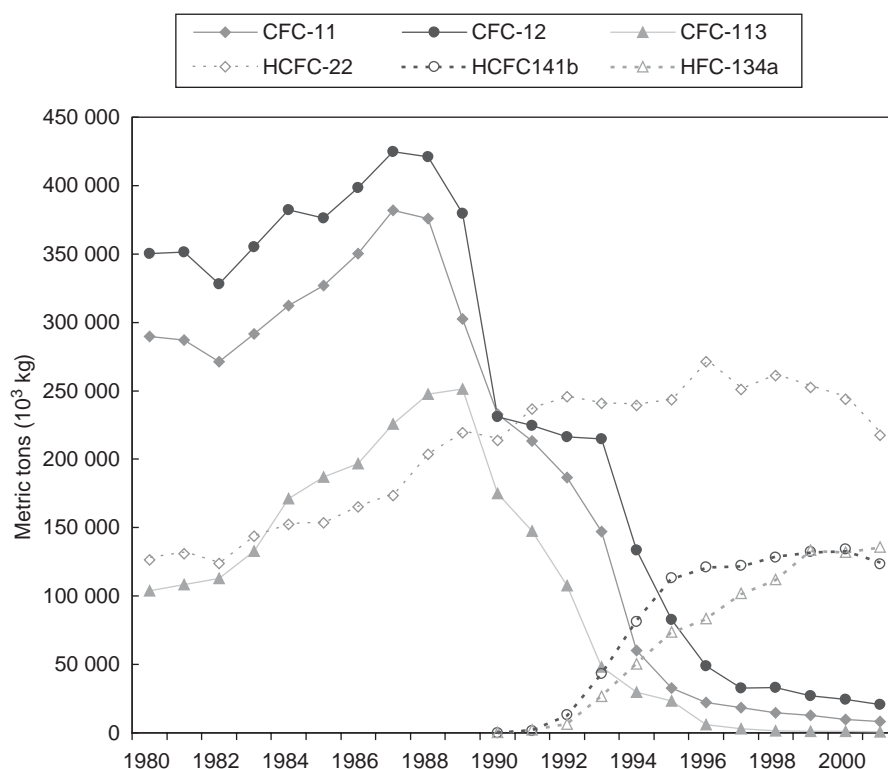
**Table 2** Principle applications and major industrial sources of halocarbons

Chemical	Application/source
CFC-11	Foam blowing, aerosol propellant
CFC-12	Refrigeration, aerosol propellant
CFC-113	Solvent, cleaning agent
CH <sub>3</sub> CCl <sub>3</sub>	Cleaning/degreasing agent
CCl <sub>4</sub>	Chemical feedstock/solvent
Halon-1211	Fire extinguishing
Halon-1301	Fire extinguishing
CH <sub>3</sub> Br	Fumigation, leaded petrol
HCFC-22	Refrigeration, air conditioning, feedstock
HCFC-141b	Foam blowing
HCFC-142b	Foam blowing
HFC-23	HCFC-22 production
HFC-134a	Refrigeration, air conditioning
HFC-152a	Refrigeration, chemical feedstock
HFC-143a	Refrigeration
HFC-125	Refrigeration
CF <sub>4</sub>	Aluminum production, electronics
C <sub>2</sub> F <sub>6</sub>	Aluminum production, electronics
SF <sub>6</sub>	Electrical insulation, magnesium smelting

In common with most HCFCs, large-scale production of HFCs did not begin until the 1990s. The most widely used is CF<sub>3</sub>CH<sub>2</sub>F (HFC-134a), which is the favored replacement refrigerant for CFC-12. Global production rose from practically zero in 1989 to 135 Gg in 2001. Another important HFC (CHF<sub>3</sub>, HFC-23) is released as a by-product during the manufacture of HCFC-22. HFC-23 has a particularly long atmospheric lifetime (270 years) and, on a per molecule basis, is one of the most potent greenhouse gases detected in the atmosphere to date.

Halons 1211 (CBrClF<sub>2</sub>) and 1301 (CBrF<sub>3</sub>) were developed in the 1940s and found widespread use in fixed (1301) and portable (1211) fire extinguishing systems. A third halon, CBrF<sub>2</sub>CBrF<sub>2</sub> (H-2402), was produced in the former Soviet Union, although in much smaller quantities. Global production of the major halons peaked in the late 1980s and had been phased out in all but a handful of developing countries by 1994. CH<sub>3</sub>Br is produced by both natural and anthropogenic processes. Major sources to the atmosphere include agricultural fumigation, biomass burning, the oceans, and a number of terrestrial ecosystems. The Montreal Protocol specifies that production of CH<sub>3</sub>Br in developed countries be reduced by 50% from 1991 levels by 2001, and recent estimates of production are consistent with this reduction.

The major source of PFCs is the aluminum industry, CF<sub>4</sub> and C<sub>2</sub>F<sub>6</sub> being released as by-products during the electrolytic conversion of aluminum oxide to aluminum. Annual CF<sub>4</sub> emissions



**Figure 1** Annual production of selected CFCs, HCFCs, and HFCs by companies reporting to AFEAS (Alternative Fluorocarbons Environmental Acceptability Study) over the period 1980–2001. The AFEAS data typically account for over 90% of total global production. (Data available at <http://www.afeas.org/>.)

of  $\sim 11 \text{ Gg year}^{-1}$  were estimated for the period 1992–98, 40% lower than the equivalent estimates for 1978–90. PFCs are also used in the semiconductor industry for processes such as plasma etching. A natural source of  $\text{CF}_4$  from fluorite minerals has also been identified, which may account for  $\sim 50\%$  of the current atmospheric burden.  $\text{SF}_6$  is mostly used as an insulating fluid in electrical switchgear and as a blanket gas in magnesium production.

## Atmospheric Loss Processes and Lifetimes

The primary atmospheric removal processes for halocarbons are photolysis and reaction with tropospheric hydroxyl radicals (OH). For the fully halogenated CFCs and halons, photolysis is the only important sink and their atmospheric lifetimes are dependent on their absorption cross-sections, the solar flux, and the surface to stratosphere transport time. As a general rule, the greater the number of Cl, Br, or I atoms on any one carbon atom, the larger the cross-section and the shorter the lifetime. For example, the lifetimes of  $\text{CClF}_3$  (CFC-113),  $\text{CCl}_2\text{F}_2$

(CFC-12), and  $\text{CCl}_3\text{F}$  (CFC-11) are 640, 100, and 45 years, respectively. The type of halogen atom is also important, as this influences not only the atmospheric lifetime but also the region in the atmosphere where photolysis can occur. For example, bromine-containing halons such as  $\text{CBrClF}_2$  (H-1211) typically absorb at longer wavelengths than their CFC analogs ( $\text{CCl}_2\text{F}_2$ ) and are, therefore, photolyzed at lower altitudes. This has important consequences for ozone chemistry, as bromine source gases tend to release their Br atoms in the lower stratosphere where ozone concentrations are greatest. Iodocarbons are readily photolyzed at wavelengths that penetrate to the surface and, consequently, their lifetimes are comparatively short, ranging from a few hours ( $\text{CH}_2\text{ClI}$ ) to several days ( $\text{CH}_3\text{I}$ ).

For halocarbons that contain a carbon–hydrogen bond, reaction with tropospheric OH becomes important and their atmospheric lifetimes become dependent on the relative rates of the OH–halocarbon reaction and the global concentration and distribution of OH. Compounds for which OH reaction is the predominant sink include HCFCs, HFCs,  $\text{CH}_3\text{CCl}_3$ ,  $\text{CH}_3\text{Cl}$ , and  $\text{CH}_3\text{Br}$ .

PFCs and  $\text{SF}_6$  are highly resistant to normal atmospheric removal processes and will persist in

the atmosphere for many thousands of years. The lifetime of  $\text{CF}_4$ , for example, has been estimated at 50 000 years. Minor loss process may include high-temperature combustion in power plants and incinerators and reactions with free electrons or  $\text{O}^+$  in the upper atmosphere.

## Analytical Methods

### Atmospheric Sampling

Measurements of halocarbons in the atmosphere can either be made *in situ*, where an analytical instrument is operated in the field or, alternatively, air samples can be collected and returned to the laboratory for subsequent analysis. Advantages of *in situ* measurements include more frequent data collection, which can lead to a greater understanding of daily and seasonal variations, and problems associated with sample storage are avoided. On the other hand, sample collection enables measurements to be made from a much wider range of location, giving a broader picture of spatial distribution. Collected samples can also be analyzed using different instruments, increasing the number and type of compound measured, and can also be stored for future analysis.

Stainless-steel canisters are the preferred choice for air sample collection, although glass vessels, polytetrafluoroethylene bags, and chemical absorbents have also been used, depending on the type and concentration range of the target compound. The internal surfaces of stainless-steel canisters are normally treated to reduce the number of active adsorption sites, thereby increasing the stability of compounds during storage. Electropolishing using techniques such as the SUMMA process, in which a pure chrome–nickel oxide layer is coated on the inner metal surface, is most common although canisters with a fused silica coating have recently become available, which helps to improve the stability of certain species. Canisters are normally fitted with a noncontaminating stainless steel valve to allow for filling or venting of air, and are available in a variety of sizes ranging from 0.5 to 30 l. Before use they should be cleaned by evacuation and repeat flushing with zero air or nitrogen, preferably at elevated temperatures (100–150°C). Humidification of the cleaned cylinder can help the stability of certain compounds.

Canisters can be filled in a variety of ways, the simplest being to fill an evacuated canister to ambient pressure by opening the valve at the sampling location. Alternatively, they can be filled using a small, noncontaminating pump. The pump should be

clean, oil and grease free, and preferably constructed so that sampled air does not come into contact with plastic or rubber seals, a potential source of contamination. Metal bellows type pumps are often used which are capable of pumping to pressures of 3–4 bar, giving ~10 l of air in a 3 l canister. Air samples can also be collected cryogenically, by partially immersing an evacuated canister in liquid nitrogen and allowing air to be drawn in and condense. Cryogenic sampling is clean, as no pump is required, and is particularly useful for collecting large volumes of air, e.g., for use as long-term calibration standards.

The suitability of stainless-steel canisters for halocarbon measurements at sub-ppb levels is well documented. Although stability over extended periods of time has been shown for many species, including CFCs, halons, HCFCs, HFCs, and PFCs, it is normally advisable to analyze samples as soon as possible after collection to minimize the risk of sample degradation. Specific problems relating to the storage of certain compounds, notably  $\text{CCl}_4$ , have been reported and the storage of more reactive halocarbons such as  $\text{CH}_3\text{Br}$  and  $\text{CH}_3\text{I}$  is not always reliable. As a general rule, sample stability is improved when samples are collected wet (or the containers are prehumidified) and when the air is stored at higher pressures.

Aluminum cylinders have also been used for air sampling, and are particularly useful as long-term standards since they can be filled to much higher pressures. For halocarbon measurements the internal surface can be passivated to allow for optimum compound stability. Long-term storage (several years) of CFCs, HCFCs, and HFCs at ppt levels has been demonstrated in Aculife-treated cylinders, although the stability of more reactive halocarbons, such as  $\text{CH}_3\text{Cl}$  and  $\text{CH}_3\text{Br}$ , remains problematical. Unlike stainless-steel canisters, aluminum cylinders must be filled with dry air as water can damage the passivation treatment.

### Chromatographic Methods

Historically, atmospheric halocarbon measurements have been made primarily using gas chromatography with electron capture detection (GC-ECD). The ECD was invented during the 1960s and is particularly sensitive to halocarbons with multiple chlorine, bromine, and/or iodine atoms. In general, the response of an ECD increases with the number and type of halogen atoms present (i.e.,  $\text{F} < \text{Cl} < \text{Br} < \text{I}$ ). For example, an ECD is more sensitive to  $\text{CH}_3\text{I}$  than  $\text{CH}_3\text{Br}$ , and more sensitive to  $\text{CHCl}_3$  than  $\text{CH}_2\text{Cl}_2$ . The CFCs, which contain multiple chlorine atoms, are particularly suited to electron capture even at low

ppt levels. GC-ECD has also been used for chlorinated solvents, including  $\text{CH}_3\text{CCl}_3$  and  $\text{CCl}_4$ , various bromo- and iodo-carbons, and  $\text{SF}_6$ .

HCFCs and HFCs are more difficult to measure by conventional GC-ECD. HCFCs typically contain less chlorine than their CFC counterparts, whilst HFCs, being composed of only carbon, hydrogen, and fluorine, are largely insensitive to the ECD. It is possible to increase the sensitivity of an ECD to certain halocarbons by doping the nitrogen make-up gas with a small amount of oxygen, typically  $\sim 0.2\%$ . The increased response is thought to be due to oxygen ( $\text{O}_2^-$ ) serving as a catalyst for electron capture. Response factors can be further improved by careful choice of operating conditions, including the use of lower detector temperatures (typically  $250^\circ\text{C}$ ) and lower ECD standing currents. Compounds that show an enhanced response to oxygen-doped ECD include HCFC-22,  $\text{CH}_3\text{Cl}$ ,  $\text{CH}_3\text{Br}$ , and  $\text{CH}_2\text{Cl}_2$ . Disadvantages of oxygen doping include increased background noise and greater potential for coelutions.

A more popular, and sometimes necessary, alternative for the analysis of CFC replacements and other ECD-weak halocarbons is gas chromatography/mass spectrometry (GC-MS). The most common GC-MS systems utilize a quadrupole analyzer, although magnetic sector instruments have also been used, which are capable of operating at higher mass resolution and have lower detection limits (parts in  $10^{15}$ ). Both are scanning analyzers, where ions are detected sequentially with time. The advantages of a quadrupole include cost, ease of use, rapid scanning speeds, and the linear variation of ion count and quadrupole field voltage. However, they are only able to operate at low resolution and are typically several orders of magnitude less sensitive than the best sector instruments.

For rapid scanning of GC output and to maximize sensitivity, mass analyzers are normally operated in selected ion mode, where only a few pre-selected ions are monitored at any one time. Electron ionization at 70 eV is normal although increased response has been observed in some systems using helium carrier gas when operated at lower ionization energies (10–20 eV). There is a growing interest in the use of negative ion chemical ionization as this technique is very sensitive to certain halocarbons, notably those containing bromine or iodine.

GC-MS has a number of advantages over other GC techniques. The ability to identify a compound by its mass spectrum or by the coelution of specific ion fragments is preferable to relying on retention time analysis alone. MS detection can help in the identification of unknown compounds and also in the

detection and identification of coeluting species. Furthermore, where two compounds are shown to coelute, either or both can still be analyzed quantitatively with the use of single ion monitoring, provided they produce nonidentical ions. Another advantage of GC-MS is the wider variety of compounds that can be measured with just one instrument. For atmospheric halocarbon measurements, GC-MS enables the detection of many more compounds, including HCFCs, HFCs, and PFCs. Finally, in the case of magnetic sector instruments, the ultimate detection limit of parts in  $10^{15}$  or better is probably two orders of magnitude better than an ECD.

Due to their superior resolving power and suitability for use with GC-MS (lower flow rates, reduced chemical bleed), capillary columns are gradually replacing packed columns as the column of choice for halocarbon analysis. Low polarity methyl and methyl/phenyl polysiloxane columns are widely used and suitable for all but the most volatile species. Column lengths typically range from 30 to 100 m and internal diameters from 0.25 to 0.53 mm ( $\leq 0.32$  mm for GC-MS). Alumina-PLOT columns offer excellent separation of the fully halogenated CFCs, PFCs, and halons, as well as HFCs, but are very sensitive to water and cannot be used for many hydrogenated compounds containing chlorine, bromine, or iodine as these can decompose on the column. Other columns widely used include various other PLOT columns (Poraplot-Q, GS-Q) and the 624-type columns, designed for the analysis of volatile pollutants using US-EPA methods.

With packed column GC-ECD it is possible to measure the major CFCs (11, 12, 113) as well as several chlorinated solvents ( $\text{CH}_3\text{CCl}_3$ ,  $\text{CCl}_4$ ) and  $\text{SF}_6$  in the background atmosphere by a simple loop injection of 1–10 ml of air. However, to measure a greater number of compounds and to be able to work with capillary columns and GC-MS it is necessary to preconcentrate samples before analysis. A variety of techniques have been developed to accomplish this and there are several systems available commercially. The general principal involves passing air through a trap, often held at subambient temperatures, so that the halocarbons and other compounds of interest are trapped, whilst the bulk of the air (mainly  $\text{N}_2$  and  $\text{O}_2$ ) passes through unretained. Typical volumes processed range from 100 to 2000 ml, actual volumes being determined by mass flow or pressure difference. Traps are normally made of stainless steel or glass and may be empty or packed with glass beads or chemical adsorbents such as Tenax or Carboxen. The most sophisticated traps may contain multiple adsorbents to allow for analysis of the widest possible range of compounds. To increase trapping



efficiency, traps are normally cooled to subambient temperatures using cryogenic fluids, recirculating refrigerants, or peltier-type coolers, the required temperature being dependent on the efficiency of the trapping material. Injection of the trapped sample onto the GC column is facilitated by rapid heating of the sample loop. When preconcentrating large volumes of air it is usually necessary to remove water vapor from the sample stream. Water can not only cause a blockage in the trap but can also affect chromatographic performance and detector response. If the dimensions of the trap are sufficiently large and care is taken with the desorption temperature, it is possible to retain water on the trap during desorption, particularly if there is a second prefocusing trap before injection onto the column. Otherwise water vapor can be removed using Nafion-type, semipermeable membranes or chemical dryers such as magnesium perchlorate or phosphorus pentoxide. Another potential interferent is CO<sub>2</sub>, particularly when analyzing very volatile halocarbons or trapping larger volumes of air.

### Calibration and Units

The most accurate gas-phase calibration standards are those made gravimetrically. Gravimetric standards are normally prepared in dry aluminum or humidified stainless-steel cylinders and typically consist of a small amount of pure material, gas or liquid, diluted with real air or synthetic air/nitrogen at high pressure. High-precision balances are necessary to determine the weight of raw material and sometimes that of the diluent gas. Commercial gravimetric standards for some halocarbons are available, although these are normally in the high ppb (parts per billion) range and have to be diluted further before analysis, either by static dilution into known volumes or by dynamic dilution with zero air or nitrogen.

Other calibration techniques include mixing a known volume of gas with nitrogen or air in a fixed volume container and the use of liquid standards, where analytes are purchased or prepared in a liquid matrix, such as methanol, and then diluted into the gas phase. These methods usually involve several dilution steps to generate standards in the ppt range. Permeation devices have also found limited use.

For atmospheric analysis, secondary or working standards are normally analyzed alongside real samples and are used to monitor daily or longer-term changes in instrument response. These standards are typically large-volume, high-pressure air samples, which are calibrated against primary standards before and after use. As both primary

and secondary standards are often used for many months or years, care has to be taken to monitor compound stability and to ensure continuity between different standards.

Atmospheric concentrations are normally expressed as mixing ratios, the preferred unit being dry air mole fraction, normally parts per trillion (10<sup>12</sup>), ppt, or pmol mol<sup>-1</sup>. For standards prepared volumetrically, corrections for nonideal behavior should be applied.

### Spectroscopic Methods

Halocarbon measurements in the atmosphere have also been made by remote sensing techniques, which negate the requirement for sample collection or preconcentration. Since many halocarbons have strong absorption bands in the 8–15 μm region, infrared spectroscopy has proved to be a valuable technique for the measurement of certain molecules including CFCs 11, 12, and 113, HCFC-22, CCl<sub>4</sub>, CF<sub>4</sub> and SF<sub>6</sub>. Most spectrometers in current use are rapid-scanning, high-resolution, Fourier-transform interferometers that use sunlight as the radiation source, i.e., the instrument records the difference between the true solar spectrum and the solar spectrum after passage through a portion of the atmosphere. Surface-based spectrometers are capable of quantifying total vertical column amounts, whilst balloon- or aircraft-borne instruments provide better spatial and vertical resolution. The ATMOS spectrometer (Atmospheric Trace Molecule Spectroscopy Experiment) has been flown several times on the space shuttle generating high-resolution (0.01 cm<sup>-1</sup>) absorption spectra of the atmosphere from space with a vertical resolution of ~2 km. Measurements have also been made from satellites.

### Global Halocarbon Monitoring

Halocarbon measurements have been made from a variety of platforms including ground-based sites, ships, aircraft, and balloons. There are several well-established global networks making regular, long-term measurements of certain species using *in situ*, canister, and remote sensing techniques. Canister samples are regularly collected from dedicated research aircraft, from commercial airliners, and even from balloon-borne samplers reaching altitudes of ~30 km. Lightweight, fast-response GC-ECD systems have also been flown on research aircraft and balloons, and the first flying GC-MS systems are being developed. As fast response times are desirable, airborne GCs are normally designed to analyze a limited number of compounds, typically two to three

compounds every 1–2 min. Chromatographic conditions such as column selection and pressure control are critical and systems normally use precolumns and/or back-flushing to prevent unwanted components from reaching the detector.

Several studies have reported halocarbon measurements in air trapped in polar firn (unconsolidated snow). Air pumped out of the firn at different depths is representative of a particular period of time, depending on diffusion rates within the snow. Firn can reach depths in excess of 100 m, and the oldest derived atmospheric records date back to the early twentieth century. These confirm that natural sources of CFCs, HCFCs, HFCs, halons, chlorinated solvents, and SF<sub>6</sub> are negligible or nonexistent. Halocarbons with significant preindustrial abundance include CF<sub>4</sub>, CH<sub>3</sub>Cl, and CH<sub>3</sub>Br.

## Atmospheric Abundance and Temporal Trends

Routine monitoring of the major CFCs, CH<sub>3</sub>CCl<sub>3</sub> and CCl<sub>4</sub> began in the late 1970s. Atmospheric mixing ratios increased steadily until the Montreal Protocol took effect in the early 1990s and growth rates began to slow. The global mean surface concentrations of CFC-11 and CFC-113 reached maxima of 275 and 85 ppt around 1993 and 1996, respectively, and have been declining slowly since. By 2000, the concentration of CFC-12 had reached 540 ppt, although its growth rate had fallen to just 2 ppt year<sup>-1</sup>, compared with 20 ppt year<sup>-1</sup> during the 1980s. With lifetimes ranging from 50 to 100 years, it will be several decades before CFC

concentrations decline substantially. In contrast, the atmospheric abundance of CH<sub>3</sub>CCl<sub>3</sub>, which has a relatively short lifetime (5 years), decreased from 130 ppt in 1992 to 45 ppt in 2000, a decline of over 60% (see Table 3).

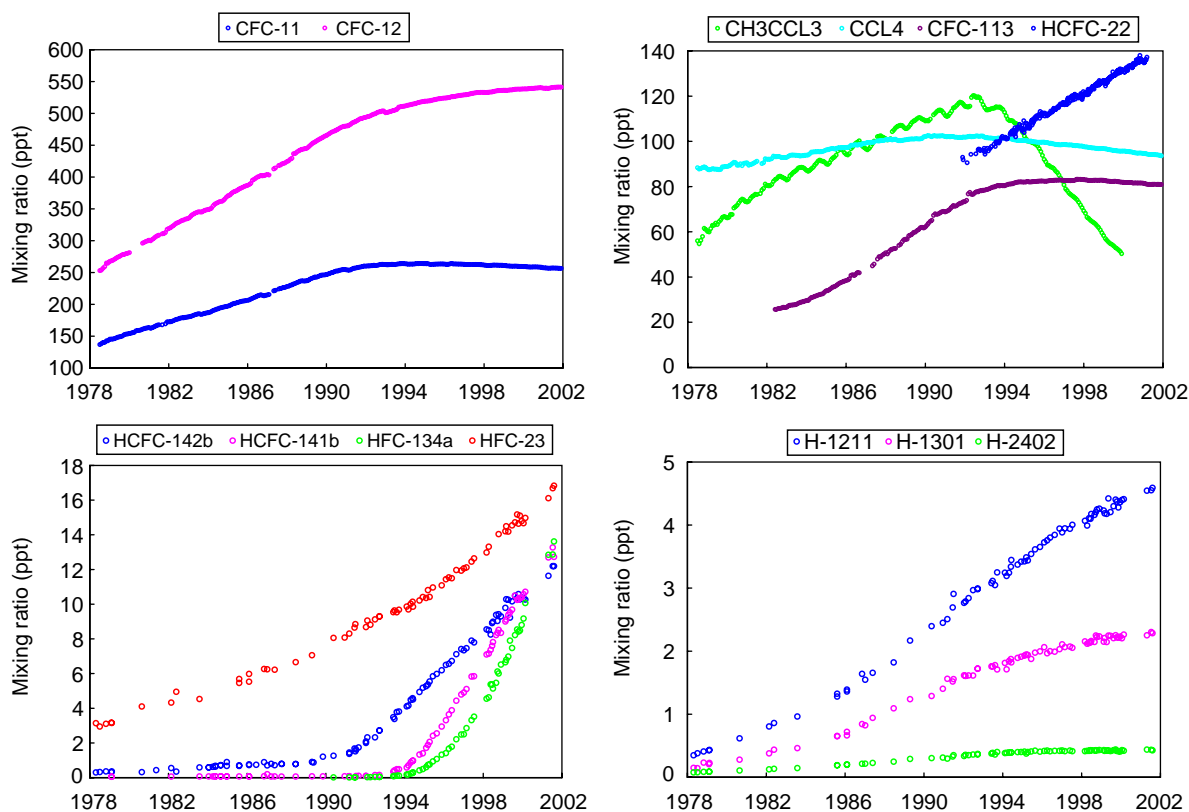
Global atmospheric mixing ratios of HCFCs and HFCs continue to rise. Several compounds (HCFC-141b, HCFC-142b, HFC-134a) have increased rapidly from zero or near-zero levels in the late 1980s as they have found use as CFC replacements. Others, like HCFC-22 and HFC-23, have grown more steadily. However, with the exception of HCFC-22, which had reached 140 ppt by 2000, the concentrations of other HCFCs and HFCs were all below 20 ppt, still significantly lower than the major CFCs. A number of other CFC replacements have also been identified in the atmosphere, including HCFC-123, HCFC-124, HFC-125, HFC-143a, and HFC-152a. Atmospheric levels of these compounds are currently a few ppt or less, but may be expected to increase in coming years (see Figure 2).

The behavior of the CFCs has generally followed a predictable pattern with atmospheric growth rates declining as industrial production and emissions have decreased. CFC-12 is taking longer to stop as much of its use involved slow release applications such as refrigeration. The behavior of halons is different and atmospheric levels of H-1211 and H-1301 are still increasing despite having one of the earliest phase-out dates. The reason for this lies partly in the way halons were used. Compared to CFCs and CH<sub>3</sub>CCl<sub>3</sub>, a far greater proportion of total halon production is stored in existing equipment and forms part of a bank of unreleased material. Two other halons to be

**Table 3** 2000 surface mixing ratios and growth rates of the major atmospheric halocarbons

Formula	Common name	Mixing ratio 2000 (ppt)	Growth rate 2000 (ppt year <sup>-1</sup> )
CCl <sub>3</sub> F	CFC-11	260	-1.5
CCl <sub>2</sub> F <sub>2</sub>	CFC-12	540	2.0
CClF <sub>2</sub> CCl <sub>2</sub> F	CFC-113	82	-0.4
CBrF <sub>3</sub>	Halon-1301	2.5	0.07
CBrClF <sub>2</sub>	Halon-1211	4	0.12
CHClF <sub>2</sub>	HCFC-22	140	5.2
CH <sub>3</sub> CClF <sub>2</sub>	HCFC-142b	12	1.0
CH <sub>3</sub> CCl <sub>2</sub> F	HCFC-141b	13	1.8
CHF <sub>3</sub>	HFC-23	16	0.9
CF <sub>3</sub> CH <sub>2</sub> F	HFC-134a	14	3.5
CF <sub>4</sub>	FC-14	80	
C <sub>2</sub> F <sub>6</sub>	FC-116	2.5	
SF <sub>6</sub>	Sulfur hexafluoride	4.7	0.2
CH <sub>3</sub> Cl	Methyl chloride	540	0
CH <sub>3</sub> Br	Methyl bromide	8	0
CH <sub>3</sub> CCl <sub>3</sub>	Methyl chloroform	45	-10
CCl <sub>4</sub>	Carbon tetrachloride	100	-1.0

Adapted from WMO (World Meteorological Organisation) (2003) *Scientific Assessment of Ozone Depletion: 2002*. Global Ozone Research and Monitoring Project, Report No. 47, Geneva.



**Figure 2** Southern hemispheric surface mixing ratios of selected halocarbons over the period 1978–2002. CFC,  $\text{CH}_3\text{CCl}_3$ , and  $\text{CCl}_4$  data are from the ALE/GAGE/AGAGE network (Prinn *et al.*, 2000, *J. Geophys. Res.*, 115: 17751–17792. <http://cdiac.ornl.gov/>). HCFC-22 data are from NOAA-CMDL (Montzka *et al.*, 1996, <http://www.cmdl.noaa.gov/hats/index.html>). Other HCFC, HFC, and halon data are updated from Oram *et al.* (1995, 1996, 1998) and Fraser *et al.* (1999).

detected in the background atmosphere are H-2402 and H-1202. Although global surface mixing ratios of individual halons were less than 5 ppt in 2000, total halon-derived bromine had risen to 8 ppt and was still increasing at  $0.2 \text{ ppt Br year}^{-1}$ . Firn studies suggest that  $\text{CH}_3\text{Br}$  was present at 5–6 ppt at the beginning of the twentieth century (prior to industrial use) and had risen to 10 ppt by the 1990s. However, recent measurements indicate that the abundance of  $\text{CH}_3\text{Br}$  may now be in decline.

A number of perfluorinated carbon and sulfur compounds have been detected in the background atmosphere, including  $\text{CF}_4$ ,  $\text{C}_2\text{F}_6$ ,  $\text{C}_3\text{F}_8$ ,  $\text{c-C}_4\text{F}_8$ ,  $\text{SF}_6$ , and  $\text{SF}_5\text{CF}_3$ . With the exception of  $\text{CF}_4$ , which is present at  $\sim 80 \text{ ppt}$ , atmospheric mixing ratios for individual compounds are currently less than 5 ppt, although all are increasing and could become important contributors to radiative forcing in the future.

See also: **Air Analysis:** Sampling; Outdoor Air. **Environmental Analysis. Gas Chromatography:** Column Technology; Instrumentation; Detectors; Mass Spectrometry; Environmental Applications. **Mass Spectrometry:**

Environmental Applications. **Ozone. Water Analysis:** Seawater – Organic Compounds.

## Further Reading

- AFEAS, Alternative Fluorocarbon Environmental Acceptability Study (<http://www.afeas.org/>).
- Andersen SO and Sarma KM (2002). In Sinclair L (ed.) *Protecting the Ozone Layer: the United Nations History*. London: Earthscan.
- Aragón P, Atienza J, and Climent MD (2000) Analysis of organic compounds in air: a review. *Critical Reviews in Analytical Chemistry* 30(2&3): 121–151.
- Brasseur GP, Orlando JJ, and Tyndall GS (eds.) (1999) *Atmospheric Chemistry and Global Change*. New York: Oxford University Press.
- Fabian P and Singh ON (eds.) (1999) *The Handbook of Environmental Chemistry*, Vol. 4, Part E, *Reactive Halogen Compounds in the Atmosphere*. Berlin: Springer-Verlag.
- Fraser PJ, Oram DE, Reeves CE, Penkett SA, and McCulloch A (1999) Southern hemispheric halon trends (1978–1998) and global halon emissions. *Journal of Geophysical Research* 104: 15985–15999.



- Helmig D (1999) Review: air analysis by gas chromatography. *Journal of Chromatography*, A 843: 129–146.
- Krska R and Kellner R (1985) Chlorofluorohydrocarbons. In Townshend A (ed.), *Encyclopedia of Analytical Science*, 1st ed. London: Academic Press.
- Molina MJ and Rowland FS (1974) Stratospheric sink for chlorofluoromethanes: chlorine atom catalysed destruction of ozone. *Nature* 249: 810–812.
- Montzka SA, Butler JH, Myers RC, Thompson TM, Swanson TH, Clarke AD, Lock LT, and Elkins JW (1996) Decline in the tropospheric abundance of halogen from halocarbons: implications for stratospheric ozone depletion. *Science* 272: 1318–1322.
- Oram *et al.* (1995) *Geophysics Research Letters* 22: 2741–2744.
- Oram *et al.* (1996) *Geophysics Research Letters* 23: 1949–1952.
- Oram *et al.* (1998) *Geophysics Research Letters* 25: 35–38.
- Prinn RG, Weiss RF, Fraser PJ, Simmonds PG, Cunnold DM, Alyea FN, O'Doherty S, Salameh P, Miller BR, Huang J, Wang RHJ, Hartley DE, Harth C, Steele LP, Sturrock G, Midgley PM, and McCulloch A (2000) A history of chemically and radiatively important gases in air deduced from ALE/GAGE/AGAGE. *Journal of Geophysical Research* 105(D14): 17751–17792.
- Singh HB (1995) Halogens in the atmospheric environment, Chapter 7. In Singh HB (ed.) *Composition, Chemistry and Climate of the Atmosphere*. New York: Van Nostrand Reinhold.
- WMO (World Meteorological Organisation) (1999) *Scientific Assessment of Ozone Depletion: 1998*. Global Ozone Research and Monitoring Project, Report No. 44, Geneva.

# CHROMATOGRAPHY

Contents

**Overview**

**Principles**

**Multidimensional Techniques**

## Overview

**V R Meyer**, EMPA St Gallen, St Gallen, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 720–729,  
© 1995, Elsevier Ltd., with revisions made by the Author.

## Introduction

Chromatography can be used to solve a very broad range of analytical problems. This versatility is reflected in the large number of chromatographic techniques that are successfully applied today. They can be classified according to a number of criteria, the most important of which is the type of mobile phase used. Subsequently the shape of the chromatographic bed and the properties of the stationary phase expand the possibilities offered by chromatography. Once a particular method is chosen, it is possible to influence the separation using programmed elution. Finally special techniques can be used to perform difficult analyses or to obtain short separation times. The individual methods are discussed in detail in

their respective articles, so here only a short introduction is given.

## Classification of Chromatographic Techniques

### The Type of Mobile Phase

Chromatographic techniques can be classified on the basis of a number of different criteria. Nevertheless, a logical hierarchy is given by first specifying the type of mobile phase, second the shape of the chromatographic bed and third the type of stationary phase. In **Table 1** the chromatographic methods are listed according to this scheme. The type of mobile phase is the most important criterion because it determines the class of samples that can be analyzed with one of the associated techniques. (From a theoretical point of view, each type of mobile phase also governs a certain range of diffusion coefficients, which determines and limits the speed of analysis and the efficiency of the method.) Therefore a distinction can be made between gas, supercritical fluid, and liquid chromatography (GC, SFC, and LC, respectively).

- Helmig D (1999) Review: air analysis by gas chromatography. *Journal of Chromatography*, A 843: 129–146.
- Krska R and Kellner R (1985) Chlorofluorohydrocarbons. In Townshend A (ed.), *Encyclopedia of Analytical Science*, 1st ed. London: Academic Press.
- Molina MJ and Rowland FS (1974) Stratospheric sink for chlorofluoromethanes: chlorine atom catalysed destruction of ozone. *Nature* 249: 810–812.
- Montzka SA, Butler JH, Myers RC, Thompson TM, Swanson TH, Clarke AD, Lock LT, and Elkins JW (1996) Decline in the tropospheric abundance of halogen from halocarbons: implications for stratospheric ozone depletion. *Science* 272: 1318–1322.
- Oram *et al.* (1995) *Geophysics Research Letters* 22: 2741–2744.
- Oram *et al.* (1996) *Geophysics Research Letters* 23: 1949–1952.
- Oram *et al.* (1998) *Geophysics Research Letters* 25: 35–38.
- Prinn RG, Weiss RF, Fraser PJ, Simmonds PG, Cunnold DM, Alyea FN, O'Doherty S, Salameh P, Miller BR, Huang J, Wang RHJ, Hartley DE, Harth C, Steele LP, Sturrock G, Midgley PM, and McCulloch A (2000) A history of chemically and radiatively important gases in air deduced from ALE/GAGE/AGAGE. *Journal of Geophysical Research* 105(D14): 17751–17792.
- Singh HB (1995) Halogens in the atmospheric environment, Chapter 7. In Singh HB (ed.) *Composition, Chemistry and Climate of the Atmosphere*. New York: Van Nostrand Reinhold.
- WMO (World Meteorological Organisation) (1999) *Scientific Assessment of Ozone Depletion: 1998*. Global Ozone Research and Monitoring Project, Report No. 44, Geneva.

# CHROMATOGRAPHY

Contents

**Overview**

**Principles**

**Multidimensional Techniques**

## Overview

**V R Meyer**, EMPA St Gallen, St Gallen, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 720–729,  
© 1995, Elsevier Ltd., with revisions made by the Author.

## Introduction

Chromatography can be used to solve a very broad range of analytical problems. This versatility is reflected in the large number of chromatographic techniques that are successfully applied today. They can be classified according to a number of criteria, the most important of which is the type of mobile phase used. Subsequently the shape of the chromatographic bed and the properties of the stationary phase expand the possibilities offered by chromatography. Once a particular method is chosen, it is possible to influence the separation using programmed elution. Finally special techniques can be used to perform difficult analyses or to obtain short separation times. The individual methods are discussed in detail in

their respective articles, so here only a short introduction is given.

## Classification of Chromatographic Techniques

### The Type of Mobile Phase

Chromatographic techniques can be classified on the basis of a number of different criteria. Nevertheless, a logical hierarchy is given by first specifying the type of mobile phase, second the shape of the chromatographic bed and third the type of stationary phase. In **Table 1** the chromatographic methods are listed according to this scheme. The type of mobile phase is the most important criterion because it determines the class of samples that can be analyzed with one of the associated techniques. (From a theoretical point of view, each type of mobile phase also governs a certain range of diffusion coefficients, which determines and limits the speed of analysis and the efficiency of the method.) Therefore a distinction can be made between gas, supercritical fluid, and liquid chromatography (GC, SFC, and LC, respectively).

**Table 1** Classification of chromatographic methods

Type of mobile phase	Gas	Supercritical fluid	Liquid
Shape of chromatographic bed	Column	Column	Column (Open tubular) Packed
Type of stationary phase	Open tubular: Liquid, cross-linked liquid Packed: Liquid, Solid	Open tubular: Cross-linked liquid Packed: Solid	Column (Packed): Solid Paper: Liquid Thin-layer: Solid
Method	Gas-liquid chromatography Gas-solid chromatography	Capillary supercritical fluid chromatography Packed column supercritical fluid chromatography	Column liquid chromatography Paper chromatography Thin-layer chromatography
Abbreviation	GC, GLC GC, GSC	SFC	LC HPLC <sup>a</sup> PC TLC HPTLC <sup>b</sup>
Type of method	Adsorption, molecular sieve, porous polymer	Adsorption, bonded phase	Adsorption, reversed phase, bonded phase, ion exchange, affinity, size exclusion

<sup>a</sup>High-performance liquid chromatography.

<sup>b</sup>High-performance thin-layer chromatography.

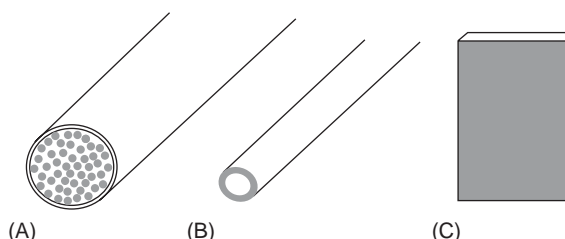
The high-performance methods use stationary phases of very small particle diameter.

SFC is less important than GC or LC. In many cases, it is not difficult to choose between the latter two methods: the prerequisite for GC is that the analyte is volatile and thermally stable (although derivatization in order to obtain these properties is possible in many cases).

### The Shape of the Chromatographic Bed

If the mobile phase is a gas or a supercritical fluid, it is necessary to let it flow through a tube, a so-called column, that contains the stationary phase. In the case of liquid chromatography one can choose between a column or planar geometry because the mobile phase can move through a sheet of paper or a thin layer by capillary action. If a column is used, the mobile phase is forced through it by pressure generated by a pump or by a gas stored in a pressurized cylinder. (As a preparative laboratory technique, liquid chromatography is also performed in columns packed with coarse stationary phases; in this case simple hydrostatic pressure may be sufficient.)

The column can be an open capillary or a packed tube. In the first case the mobile phase is coated as a thin film on the inner wall of the capillary. If the mobile phase has a certain solvating power, as in SFC, it is necessary to cross-link this liquid film, whereas in GC linear polymers are used in many cases because the usual carrier gases, helium and hydrogen, cannot dissolve any stationary phase. (Owing to problems with manufacturing and



**Figure 1** Three possibilities for creating a chromatographic bed: (A) packed column, (B) open capillary, and (C) plane. The third technique can only be used with a liquid mobile phase; here a thin-layer plate is drawn but the plane can also consist of a sheet of paper. In all three cases the stationary phase is shown in gray. The particles of the packed column can be round, as in the figure, or irregular.

instrumentation, liquid chromatography with open capillaries is only of theoretical interest.) If the column contains packing, many possibilities are offered by contemporary technology. The stationary phase can be an inorganic adsorbent, a cross-linked and thereby rigid organic polymer, an inorganic or organic material with chemically modified surface, or even a liquid film coated on a granular carrier material.

Figure 1 shows the three possibilities: packed column, open capillary, and plane.

### Terminology of the Methods

Taking the type of chromatographic bed and stationary phase into account, GC, SFC, and LC can

now be subdivided, although the usual terms do not follow the same criteria in all cases. **Table 1** uses the expressions gas–liquid, gas–solid, capillary supercritical fluid, packed-column supercritical fluid, column liquid, paper, and thin-layer chromatography. The following abbreviations are used: GLC, GSC, but in most cases the type of stationary phase is omitted and both techniques are termed GC; SFC; LC (usually only used for column techniques, although thin-layer and paper chromatography are also ‘LC’); PC (sometimes also used for ‘preparative chromatography’); and TLC. For LC and TLC, which both use a granular stationary phase, a special term was introduced to distinguish the more recent instrumental methods based on very fine stationary phases from the classical ones: HPLC and HPTLC where HP is for ‘high performance’. Here the particle diameter is not larger than  $\sim 10\ \mu\text{m}$ , which is the key to obtaining high plate numbers per unit length.

Especially in LC, and also in other fields, it is usual to distinguish in more detail between very different types of method. This will be discussed below.

### Comparison of the Methods

**Table 2** lists some characteristic features of GC, SFC, and LC. In most cases GC is used as open-tubular GLC, and LC is performed in packed columns. As can be seen from the physical parameters of density, viscosity, and diffusion coefficient, SFC lies between GC and LC and it is no surprise that it can be used equally well with open capillaries and packed columns. The values in the table are to some extent arbitrary but are typical. The values of the three basic physical parameters are not only of theoretical interest but are linked directly to some of the main

properties of the methods: the density governs the solvating power of the mobile phase and thereby determines whether the separation can be influenced by a particular choice of eluent; the viscosity influences the pressure needed to force the mobile phase through the chromatographic bed and sets the upper limit of the solute diffusion coefficient. This latter should be high because low diffusivity means slow mass transfer and therefore slow chromatography. If the sample diffusion coefficient is low, it is necessary to keep low the characteristic chromatographic dimension, i.e., the capillary or particle diameter. Therefore, LC with  $5\ \mu\text{m}$  particles is more efficient than with a  $100\ \mu\text{m}$  packing.

Practical aspects of the three methods are listed under Variables and Sample Prerequisites at the end of **Table 2**.

### Gas Chromatography

If the mobile phase is a gas, the sample needs to be volatile. Its boiling point at atmospheric pressure should not be higher than  $\sim 360^\circ\text{C}$ . If the temperature of the GC column or capillary is adequate, the sample molecules will be transported by the gas owing to their volatility. Retention is governed by both vapor pressure and affinity to the stationary phase of a given compound. The gaseous mobile phase has no direct influence on the separation. GC can be a simple and rapid technique and is the method of choice for the investigation of volatile and even very complex samples. An example is given in **Figure 2**.

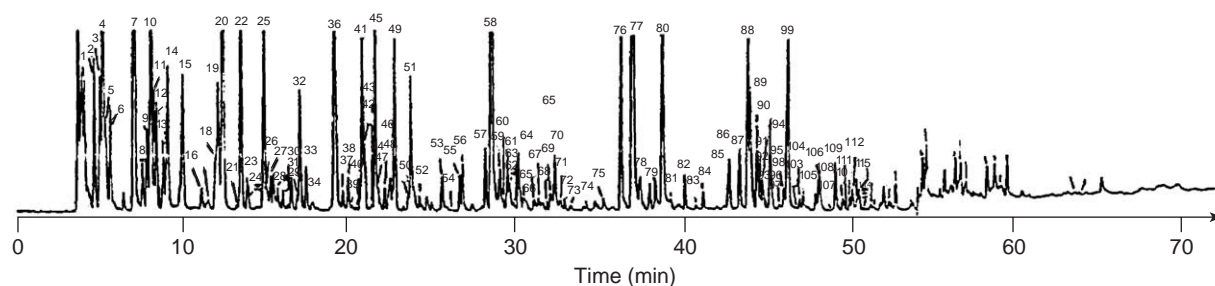
The most frequently used mobile phases for GC are hydrogen and helium. The lower the molecular

**Table 2** Comparison of analytical column-type chromatographic methods

	GC (open-tubular GLC)	SFC	LC
Density of mobile phase ( $\text{g ml}^{-1}$ )	$10^{-3}$	0.5	1
Viscosity of mobile phase (poise)	$10^{-4}$	$10^{-3}$	$10^{-2}$
Diffusion coefficient of solute in mobile phase ( $\text{m}^2\text{s}^{-1}$ )	$10^{-5}$	$10^{-7}$	$10^{-9}$
Diameter of capillary ( $\mu\text{m}$ )	320	100	
Diameter of packing ( $\mu\text{m}$ )		5	5
Length of column (m)	25	25, 0.25 <sup>a</sup>	0.1
Number of theoretical plates ( $\text{m}^{-1}$ )	3 000	3 000, 50 000 <sup>a</sup>	50 000
Number of theoretical plates per column	75 000	75 000, 12 000 <sup>a</sup>	5 000
Pressure drop (bar)	1	Variable <sup>b</sup>	100
Variables	Stationary phase Temperature	Stationary phase Mobile phase Temperature Pressure	Stationary phase Mobile phase (Temperature)
Sample prerequisites	Volatility Thermal stability	Solubility (Thermal stability)	Solubility

<sup>a</sup>For open tubular and packed columns, respectively.

<sup>b</sup>In SFC the pressure drop over the column can be chosen.



**Figure 2** Gas chromatographic separation of hydrocarbons found in an urban air sample. Open capillary, 0.32 mm i.d.  $\times$  60 m length; stationary phase, DB-1 (dimethyl polysiloxane); film thickness, 0.25  $\mu\text{m}$ ; carrier gas, helium; temperature programme, 5°C isothermal for 3 min, 5–50°C at a rate of 3°C min<sup>-1</sup>, 50–220°C at a rate of 5°C min<sup>-1</sup>; detector, flame ionization. With this method, a total of 142 hydrocarbons could be separated and identified; 128 of them were found in the urban air sample. (After Ciccioli P, Cecinato A, Brancaleoni E, Frattoni M, and Liberti A (1992) Use of carbon adsorption traps combined with high resolution GC–MS for the analysis of polar and nonpolar C4–C14 hydrocarbons involved in photochemical smog formation. *Journal of High Resolution Chromatography* 15: 75.)

mass of a gas, the lower its own diffusivity as well as the diffusivity of the sample molecules and the faster the chromatography. Therefore, hydrogen would be the favored carrier gas but it is often barred on safety grounds. Sometimes nitrogen is used because it is cheap but this can only be recommended for simple analytical problems because the separation performance is poorer than with gases of low molecular mass. The fact that some detectors demand the use of a certain gas must also be taken into consideration.

A typical stationary phase for GC is a viscous liquid with low vapor pressure (at the temperature required for a given range of application). The two most important types of stationary phases are silicones and polyglycols; their structures are given in Table 3. The silicones especially can be substantially chemically modified in order to obtain a wide range of polarities and specialized functionalities (including chiral groups). The stationary phase is coated as a thin film (typically 0.25  $\mu\text{m}$ ) on the inner wall of the open capillary or on the surface of a granular, porous, inert packing material, in this case called a solid support. For special types of analyses the stationary phase is not a liquid but a porous–solid packing. Adsorbents (silica), molecular sieves and porous polymers are used for the GSC of highly volatile samples such as mixtures of permanent gases or low-molecular-mass hydrocarbons.

In GC, the eluted compounds are most often detected with a flame-type detector that generates ions, the so-called flame ionization detector, FID; for special purposes nitrogen- and phosphorus-sensitive FIDs, electron capture, or thermal conductive detectors and mass spectrometers are used.

If a sample is not volatile, several derivatization techniques are known that allow reduction in the boiling points of certain classes of compounds.

Alcohols, amines, amino acids, carboxylic acids, carbohydrates, and steroids can be trimethylsilylated; amines, phenols, carbohydrates, and steroids can be acylated with trifluoroacetic acid or a higher homolog; carbonic acids and phenols can be alkylated.

## Liquid Chromatography

Liquid chromatography has a number of different configurations with regard to technical (instrumental) as well as separation modes. Paper, thin-layer, and classical column techniques all belong to liquid chromatography and the ‘high performance’ technique especially (though to a lesser extent the other methods also) offers a great variety of separation principles.

### Paper Chromatography

The simplest and cheapest technique is paper chromatography, where the chromatographic bed consists of a sheet of paper, i.e., cellulose. The stationary phase consists of water adsorbed to the cellulose as well as of the polymer itself, although ion exchange and complexation processes may play an important role. The sample solution is applied as a spot near one end of the paper. A few centimeters of the sheet are dipped into the mobile phase which then ascends (or descends, as descending mode is also possible) into the stationary phase. When the mobile phase has almost reached the other end of the sheet the paper is removed from the developing tank and dried. If the analytes are not visible because they are not colored, the sheet is treated with a reagent to visualize the spots.

**Table 3** Important stationary phases for GLC and (HP)LC

GC		
Silicones	$\left( \begin{array}{c} \text{R} \\   \\ \text{—Si—O—} \\   \\ \text{R}' \end{array} \right)_n$	With the proper choice of R and R' a wide range of polarities and special functionalities is available
Polyglycols	$\text{—(O—CH}_2\text{—CH}_2\text{)}_n\text{—OH}$	Polar stationary phase: $n$ ranges from 4 to 800
LC		
Silica	$\begin{array}{c}   \\ \text{(SiO}_2\text{)}_n\text{—Si—OH} \\   \end{array}$	Three-dimensional network
Octadecyl silica	$\begin{array}{c}   \\ \text{(SiO}_2\text{)}_n\text{—Si—C}_{18}\text{H}_{37} \\   \end{array}$	} Reversed phases
Octyl silica	$\begin{array}{c}   \\ \text{(SiO}_2\text{)}_n\text{—Si—C}_8\text{H}_{17} \\   \end{array}$	
Diol silica	$\begin{array}{c}   \\ \text{—(SiO}_2\text{)}_n\text{—Si—CH}_2\text{—CHOH—CH}_2\text{OH} \\   \end{array}$	} Polar bonded phases
Nitrile silica	$\begin{array}{c}   \\ \text{—(SiO}_2\text{)}_n\text{—Si—CH}_2\text{—CH}_2\text{—CN} \\   \end{array}$	
Amino silica	$\begin{array}{c}   \\ \text{—(SiO}_2\text{)}_n\text{—Si—CH}_2\text{—CH}_2\text{—NH}_2 \\   \end{array}$	
Polystyrene	$\begin{array}{c} \text{—(CH—CH}_2\text{)}_n\text{—} \\   \\ \text{C}_6\text{H}_5 \end{array}$	Three-dimensional network due to cross-linking with divinylbenzene
Strong cation exchanger	• $\text{—SO}_3^-\text{H}^+$	} • Can be silica or polystyrene
Weak cation exchanger	• $\text{—COO}^-\text{H}^+$	
Strong anion exchanger	• $\text{—NR}_3^+\text{OH}^-$	
Weak anion exchanger	• $\text{—NH}_3^+\text{OH}^-$	

### Thin-Layer Chromatography

Thin-layer chromatography is more versatile than paper chromatography since a number of different stationary phases are available such as silica, derivatized silica, or cellulose (the analogue to paper); also, developing times are much shorter. An immense number of spray reagents have been published that allow detection of any type of analyte. HPTLC is the 'high-performance' version of TLC and uses 10  $\mu\text{m}$  or 5  $\mu\text{m}$  stationary phase particles. The separation performance of these plates is higher, but to take full advantage it is necessary to use instrumentation for sample application, development, and detection. As an example of TLC, Figure 3 presents the separation of ten rare earths.

### Liquid Chromatography

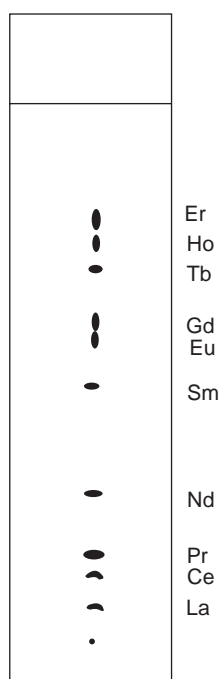
Whereas chromatography in open columns is mainly used for preparative purposes, the analytical technique

is LC using microparticulate packings. Under these circumstances it is necessary to use a pump for mobile phase transport and a detector for the observation of the fractions (usually the concentration of the analytes in the eluate is low), e.g. UV absorbance, fluorescence, refractive index, or electrochemical detectors according to the properties of the analytes. It is also possible to derivatize the sample prior to or after the separation. Precolumn derivatization can be performed offline or online; postcolumn derivatization is usually carried out online. An example of LC is given in Figure 4, which shows the separation of the three stereoisomers of mivacurium, a neuromuscular blocking agent, in a plasma extract.

### Liquid Chromatographic Separation Principles

Liquid chromatography can be performed in a variety of modes; the most important ones are





**Figure 3** High-performance thin-layer chromatographic separation of ten rare earths (as nitrates). Sample, 1  $\mu\text{g}$  each of rare earth; layer, silica, impregnated with ammonium nitrate prior to the separation; mobile phase, 4-methyl-2-pentanone/tetrahydrofuran/nitric acid/2-ethylhexylphosphonic acid mono-2-ethyl hexylester 3:1.5:0.46:0.46; developing distance, 5 cm; detection reagent, (1) spray of saturated alizarin solution in ethanol, (2) ammonia vapour, (3) gentle heating. (After Wang QS and Fan DP (1991) *Journal of Chromatography* 587: 359.)

presented briefly. Schematic drawings are shown in **Figure 5**, and **Table 3** also lists some stationary phases used in LC.

**Adsorption chromatography** The stationary phase is a polar adsorbent, in most cases silica. The mobile phase is nonpolar (usually a solvent with polarity within the range from hexane to esters). It competes with the sample molecules for adsorption at the active sites of the stationary phase. Nonpolar compounds are eluted first, followed by solutes of increasing polarity. Steric properties of the sample compounds can play an important role and therefore adsorption chromatography is the method of choice for the separation of many classes of isomers.

**Reversed-phase chromatography** The stationary phase here is nonpolar; in most cases it is derivatized silica that carries  $\text{C}_{18}$  (i.e.,  $\text{C}_{18}\text{H}_{37}$ ) or  $\text{C}_8$  (i.e.,  $\text{C}_8\text{H}_{17}$ ) groups. The mobile phase is polar, in most cases a mixture of water (or buffer solution) with methanol, acetonitrile, or tetrahydrofuran. Such an eluent cannot wet the surface of the stationary phase and the solutes are retained owing to an energy gain

that comes from the decrease in contact area between the two phases as long as the sample molecules adhere to the hydrocarbon chains. Polar analytes are eluted first and homologs will be retained more strongly the longer their chain length. Ionic compounds can be separated on reversed phases if a neutral ion-pair is formed by the addition of a counter-ion to the eluent. This was carried out in the separation shown in **Figure 4**: mivacurium is a quaternary amine, and therefore a sulfonic acid was added as an agent to mask its charge.

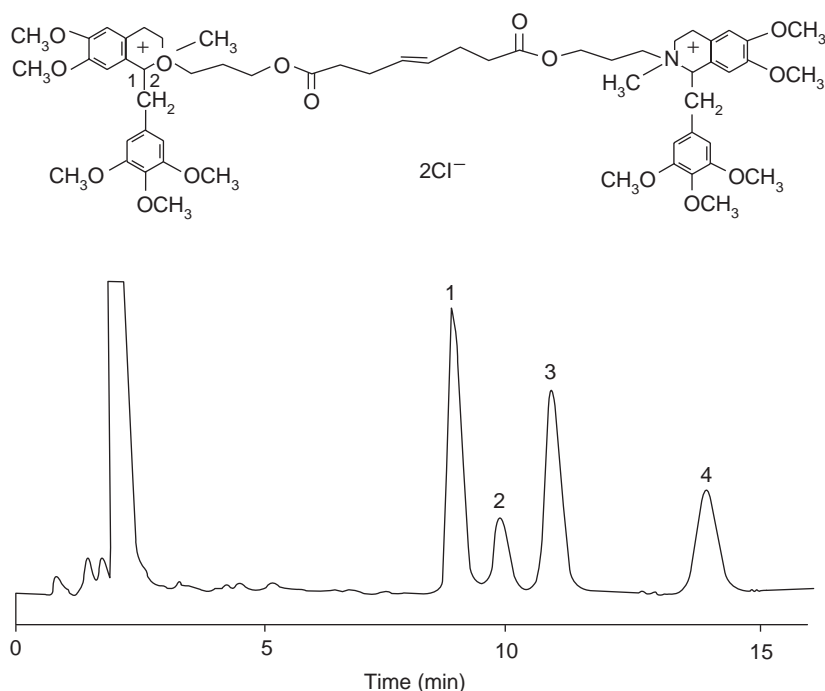
**Other bonded phases on silica (not illustrated in Figure 5)** Besides the nonpolar hydrocarbons, other functional groups can also be bonded to silica. Important stationary phases are diol, nitrile, amino (see **Table 3**), and a great number of special functionalities including chiral groups. The retention mechanisms are as variable as the stationary phases and are not known in some cases.

**Ion exchange chromatography** Ion exchange groups can be bonded to silica or to polystyrene. ‘Classical’ ion exchange is based on ionic equilibria between solute, buffer and stationary phase ions and counter-ions. Besides this ion exclusion mechanisms can also be utilized and special types of ion exchangers have been developed for the separation of the ions of strong acids and bases.

**Size exclusion chromatography** If the mobile phase has a good affinity for both the sample molecules and the stationary phase and if the latter has a well-defined pore structure, such a chromatographic system will separate the solutes according to their size. They will not be retained by the column packing but will enter the pores where the mobile phase is stagnant. Large molecules can utilize a smaller fraction of the pore volume than small ones and will be eluted earlier. Molecules that are too large to enter the pores are excluded and will appear as the first fraction at the column end.

**Affinity chromatography** The stationary-phase matrix can be loaded with chemically bonded, biologically active groups such as enzymes or antibodies. If a complex sample is injected into an affinity column, only those molecules will be retained that bind to the ligands; in the cases mentioned above these will be substrates or antigens. All other compounds will be swept away by the mobile phase. Afterwards the retained molecules can be eluted by switching to a specially designed mobile phase (e.g., change of pH or ionic strength). Affinity chromatography is a





**Figure 4** Liquid chromatographic separation of mivacurium stereoisomers in human plasma extract. The drug is a mixture of three isomers; the structure is drawn without stereochemical preference. Column, 4.6 mm i.d.  $\times$  12.5 cm length; stationary phase, LiChrospher 60 RP (reversed-phase) select B, 5  $\mu$ m; mobile phase, acetonitrile/water 40:60 with 0.005 mol l<sup>-1</sup> octanesulfonic acid (as ion-pair reagent), 1 ml min<sup>-1</sup>; detector, fluorescence 202/320 nm. Peaks: (1) is the *trans-trans* isomer (1*R*, 1'*R*, 2*S*, 2'*S*); (2) is the *cis-trans* isomer (1*R*, 1'*R*, 2*R*, 2'*S*); (3) is the *cis-cis* isomer (1*R*, 1'*R*, 2*R*, 2'*R*); (4) is the internal standard, the *trans-trans* analog of mivacurium with a benzene ring instead of the double bond in the middle of the molecule. (After Brown AR, James CD, Welch RM, and Harrelson JC (1992) Stereoselective HPLC assay with fluorometric detection for the isomers of mivacurium in human plasma. *Journal of Chromatography* 578: 302.)

highly selective method and works by an 'on-off switching' mechanism.

## Supercritical Fluid Chromatography

The phase diagram of a pure compound shows not only regions of the solid, liquid, and gaseous states, the equilibrium lines and the triple point, but also the critical point. If pressure and temperature exceed the critical values, the compound will be neither a liquid nor a gas, nor will the two phases coexist, but a supercritical fluid exists. This phase is denser and more viscous than a gas without attaining the properties of a liquid as shown in Table 2. The advantage of SFC over LC lies in the higher diffusion coefficient, which allows faster separations; in comparison to GC the mobile phase has a large solvating power and thus influences selectivity.

In most SFC separations carbon dioxide is used as the mobile phase; often a modifier (of polarity) such as methanol, other alcohols or water is added. The critical data for CO<sub>2</sub> are 31.3°C and 72.9 bar, values that can easily be handled by instrumental chromatography. To keep the column outlet under critical conditions, a restrictor (a device with a high

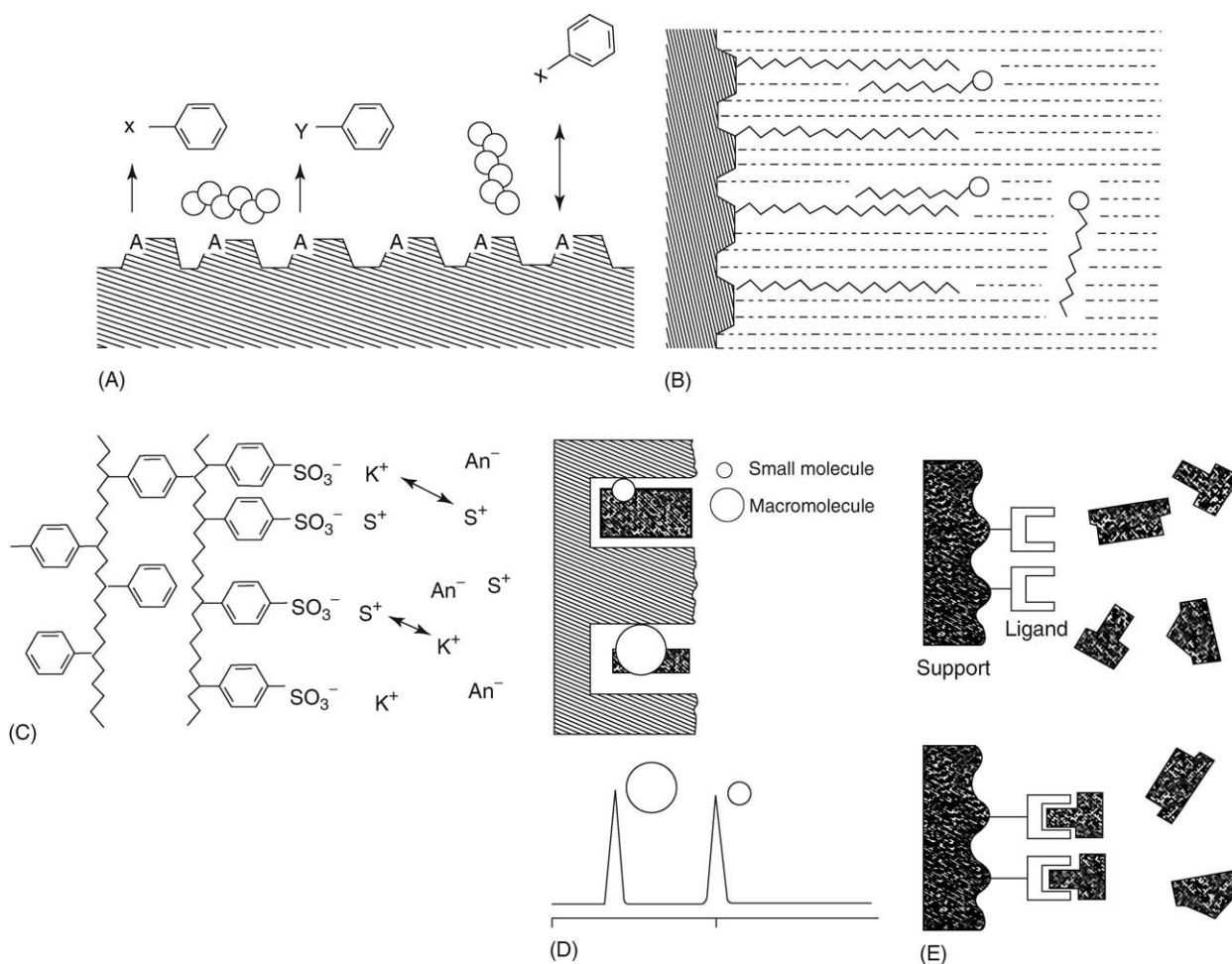
resistance to the eluent flow) needs to be installed after or at the outlet of the column.

Owing to its intermediate position between GC and LC, SFC can be performed equally well in open capillaries and packed columns. The separation can be influenced by the type of stationary phase and of modifier, by pressure, pressure drop, and temperature. In contrast to GC, SFC can also be used for the separation of nonvolatile or thermally labile compounds (although some temperature compatibility is necessary). The separation of enantiomers on chiral stationary phases can be very attractive because the temperature is lower than in GC, which increases the separation factors. SFC is an alternative to normal-phase LC because it is fast and carbon dioxide is ecologically sound. An example of an SFC separation can be found in the previous article, Principles, where Figure 2 shows the separation of orange oil components.

## Special Chromatographic Techniques

### Preparative Methods

Chromatography can equally well be used for analytical and preparative purposes. 'Preparative' is not



**Figure 5** Separation principles of liquid chromatography. (A) Adsorption chromatography: the adsorptive sites of the stationary phase are symbolized by A; the solute molecules interact with their polar groups X or Y; the mobile phase drawn is hexane, which can also interact weakly with A. (B) Reversed-phase chromatography: the solute molecules interact via their nonpolar groups with the nonpolar stationary phase. (C) Ion-exchange chromatography: a styrene-divinylbenzene type cation exchanger is shown; sample ions  $S^+$  and buffer cations  $K^+$  compete for interaction with the exchange sites. (D) Size exclusion chromatography: sample molecules can occupy the pore volume according to their size, therefore the macromolecule will spend less time in the pores and elute first. (E) Affinity chromatography: only certain sample molecules can fit to the ligands of the stationary phase, the others are washed out.

reserved to the fractionation of large samples, but indicates that the separated compounds are collected and used for a subsequent purpose: identification or structure elucidation, chemical modification by synthetic methods, use as a reference material, determination of chemical or biological properties, or for sale. If only small amounts of material are needed, the only difference from analytical chromatography lies in the use of a fraction collector; for routine separations it should be computer controlled.

If the sample size is increased, the shape of the peaks changes to rectangular (in the case of volume overload) or triangular (with mass overload); mixed forms and distorted peak shapes are also observed. Displacement effects can occur where a compound is 'pushed' and concentrated by a following one that has a stronger affinity to the stationary phase.

If large samples need to be separated, the diameter, and also often the length, of the column are increased. (Obviously, open capillaries cannot be used for this purpose.) Preparative GC is an attractive approach (though the fraction collector needs to be cooled) but few commercial instruments are available. Preparative LC is the most important technique in organic synthesis, biochemical research, downstream processing in biotechnology, and for the commercial preparation of certain chemicals or drugs.

### Programmed Elution

In a complex sample the individual analytes often have very different retention factors in a given chromatographic system. It is therefore not possible to

separate and elute them efficiently without changing the properties of the system, i.e. under so-called isothermal (GC) or isocratic (LC) conditions. In this case a GC separation is started at relatively low temperature, an LC separation at low eluting power of the mobile phase. Subsequently the temperature or mobile phase strength is increased in order to elute compounds that were strongly retained under the initial conditions. In GC this technique is called a temperature program (see **Figure 2**); the corresponding LC term is gradient elution. Note that in normal-phase LC the polarity of the mobile phase needs to be increased (however, gradient elution on silica is almost never performed because steep gradients are not possible and it takes a long time to re-equilibrate the column after the separation), whereas in reversed-phase LC the eluent polarity is decreased. A gradient from 10 to 100% acetonitrile in water can separate a very broad range of compounds on a reversed-phase column; pH or ionic strength gradients are also possible. In SFC, mobile phase, pressure, and temperature gradients are of equal importance.

### Column Switching

An alternative to programmed elution can be the coupling of two (or even more) columns with different stationary phases. This technique is known as multidimensional chromatography. The first column, for example, will separate the sample according to polarity groups. Then selected fractions are switched online to the second column, where the fine separation into chemically pure compounds takes place.

It is even possible to couple LC and GC; here, LC plays the role of a sample preparation technique that eliminates compounds that would affect the gas chromatographic separation. Because GC cannot tolerate high volumes of liquid, it is necessary to use narrow-bore LC columns, to split the eluate, or to use a special interface that eliminates most of the liquid.

### Special GC Techniques

**Headspace analysis** For the investigation of the volatile ingredients of complex mixtures, e.g., of olfactory principles, the sample is stored in a closed vial and perhaps gently heated. A portion of the vapor that fills the space over the solid or liquid sample is collected by a syringe and injected into the gas chromatograph. To obtain reproducible results it is necessary to control storage temperature and time strictly.

**Thermal desorption** Volatile compounds in gases such as pollutants in air can be trapped in a small adsorption tube, either by pumping the gas through or by passive diffusion. The packing in the trap can be chosen from a wide variety of adsorbents (molecular sieves, graphitized carbon blacks, organic polymers). After sample collection the adsorption tube is rapidly heated in a stream of purge gas which transports the released analytes to the GC column where the separation runs.

**Pyrolysis chromatography** For the GC analysis of high-molecular-mass samples such as plastics or wood, the sample can be pyrolyzed (heated until breakdown into smaller molecules occurs) online prior to injection. A 'fingerprint' of the material is obtained that can be used for quality control or identification purposes.

*See also:* **Chromatography:** Principles. **Gas Chromatography:** Column Technology; Pyrolysis; Detectors. **Headspace Analysis:** Static; Purge and Trap. **Ion Exchange:** Overview. **Liquid Chromatography:** Overview; Ion Pair; Size-Exclusion. **Supercritical Fluid Chromatography:** Overview; Applications. **Thin-Layer Chromatography:** Overview.

### Further Reading

- Anton K and Berger C (eds.) (1997) *Supercritical Fluid Chromatography with Packed Columns: Techniques and Applications*. New York and Basel: Dekker.
- Fowles IA (1995) *Gas Chromatography*. Chichester: Wiley.
- Fried B and Sherma J (1999) *Thin-Layer Chromatography*. New York and Basel: Dekker.
- Grant DW (1996) *Capillary Gas Chromatography*. Chichester: Wiley.
- Grob RL and Barry EF (1995) *Modern Practice of Gas Chromatography*, 4th edn. Chichester: Wiley.
- Guiochon G and Guillemin GL (1988) *Quantitative Gas Chromatography*. Amsterdam: Elsevier.
- Hahn-Deinstrop E (2000) *Applied Thin-Layer Chromatography: Best Practice and Avoidance of Mistakes*. Weinheim: Wiley-VCH.
- Jennings W, Mittlefehldt E, and Stremple P (1997) *Analytical Gas Chromatography*, 2nd edn. New York: Academic Press.
- Katz E, Eksteen R, Schoenmakers P, and Miller N (eds.) (1998) *Handbook of HPLC*. New York and Basel: Dekker.
- Lindsay S and Barnes J (eds.) (1992) *High Performance Liquid Chromatography*, 2nd edn. Chichester: Wiley.
- Lough WJ and Wainer IW (eds.) (1995) *High Performance Liquid Chromatography. Fundamental Principles and Practice*. London: Blackie.

- McMaster MC (1994) *HPLC – A Practical Users Guide*. Chichester: Wiley.
- McNair HM and Miller JM (1997) *Basic Gas Chromatography*. Chichester: Wiley.
- Meyer VR (1999) *Practical High-Performance Liquid Chromatography*, 3rd edn. Chichester: Wiley.

- Smith RM (1999) Supercritical fluids in separation science – the dreams, the reality and the future. *Journal of Chromatography A* 856: 83–115.
- Wells PS, Zhou S, and Parcher JF (2003) Unified chromatography with CO<sub>2</sub>-based binary mobile phases. *Analytical Chemistry* 75: 18A–24A.

## Principles

V R Meyer, EMPA St Gallen, St Gallen, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

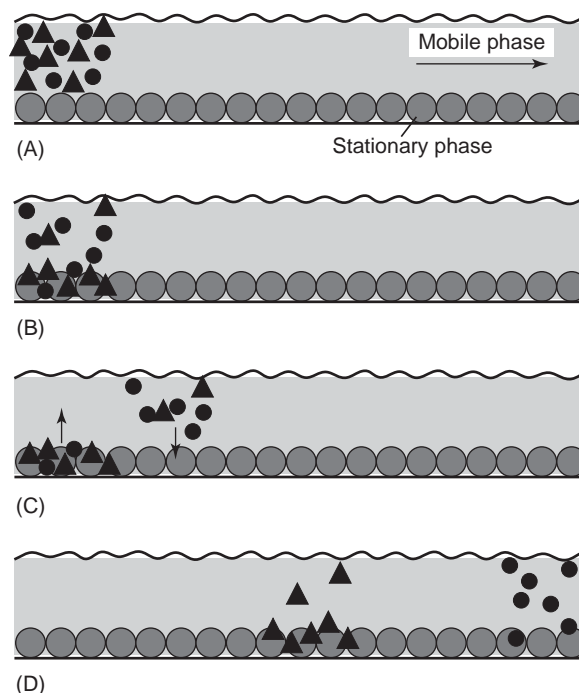
Chromatography is one of the most important analytical techniques. It allows the separation and subsequently the qualitative and quantitative analysis of complex mixtures, as long as the samples are volatile or soluble in a suitable solvent. Since chromatography is based on the partition of the sample components between two phases, one stationary and one moving, it is necessary to distinguish between gas, liquid, and supercritical fluid chromatography, according to the type of mobile phase used. Chromatography is versatile and can be highly efficient; full automation is possible. Some basic principles of its theory are presented here as knowledge of the underlying phenomena is necessary to take real advantage of all the possibilities offered by chromatographic techniques.

### The Chromatographic Process

In order to obtain a chromatographic separation, two phases are needed – a moving or mobile phase and a fixed or stationary phase. The stationary phase can be either a solid or a liquid, the mobile phase is a liquid, a gas, or a supercritical fluid. Both phases must be able to interact physically or chemically with the sample molecules; chromatography is based on transport, solvation, and ‘adsorption’ (in a very broad sense) phenomena. When the mobile phase is flowing through or over the stationary phase the analytes in the sample mixture undergo characteristic partition between the two phases. The mobile phase transports, the stationary phase retains. A mixture can be separated if its compounds are retained to varying degrees.

Figure 1 is a simple representation of the process. In (A) a mixture of seven molecules each of ● and ▲ is introduced into the chromatographic system. In (B) they are distributed between the upper mobile and lower stationary phase, and in (C) the mobile phase has transported the dissolved molecules over a small distance: new equilibria between the phases are established. When this process has been repeated many times, as in (D), the compounds ● and ▲ are separated because their preference for one of the two phases differs strongly.

In practice, chromatography takes place on a plane or in a tube. The plane can be a sheet of paper (paper



**Figure 1** Schematic representation of the process of a chromatographic separation. (A) Sample injection; (B) partition between the two phases; (C) progression of the mobile phase and new equilibrium; and (D) separation of the two compounds after a number of partition processes.

- McMaster MC (1994) *HPLC – A Practical Users Guide*. Chichester: Wiley.
- McNair HM and Miller JM (1997) *Basic Gas Chromatography*. Chichester: Wiley.
- Meyer VR (1999) *Practical High-Performance Liquid Chromatography*, 3rd edn. Chichester: Wiley.

- Smith RM (1999) Supercritical fluids in separation science – the dreams, the reality and the future. *Journal of Chromatography A* 856: 83–115.
- Wells PS, Zhou S, and Parcher JF (2003) Unified chromatography with CO<sub>2</sub>-based binary mobile phases. *Analytical Chemistry* 75: 18A–24A.

## Principles

V R Meyer, EMPA St Gallen, St Gallen, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

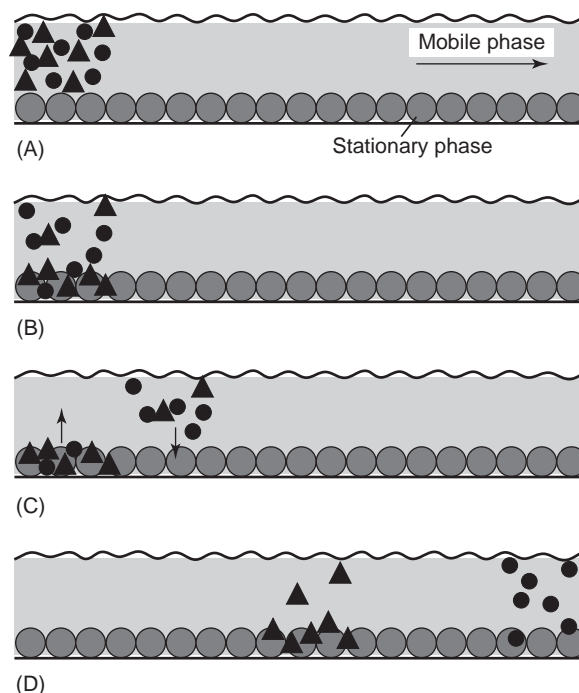
Chromatography is one of the most important analytical techniques. It allows the separation and subsequently the qualitative and quantitative analysis of complex mixtures, as long as the samples are volatile or soluble in a suitable solvent. Since chromatography is based on the partition of the sample components between two phases, one stationary and one moving, it is necessary to distinguish between gas, liquid, and supercritical fluid chromatography, according to the type of mobile phase used. Chromatography is versatile and can be highly efficient; full automation is possible. Some basic principles of its theory are presented here as knowledge of the underlying phenomena is necessary to take real advantage of all the possibilities offered by chromatographic techniques.

### The Chromatographic Process

In order to obtain a chromatographic separation, two phases are needed – a moving or mobile phase and a fixed or stationary phase. The stationary phase can be either a solid or a liquid, the mobile phase is a liquid, a gas, or a supercritical fluid. Both phases must be able to interact physically or chemically with the sample molecules; chromatography is based on transport, solvation, and ‘adsorption’ (in a very broad sense) phenomena. When the mobile phase is flowing through or over the stationary phase the analytes in the sample mixture undergo characteristic partition between the two phases. The mobile phase transports, the stationary phase retains. A mixture can be separated if its compounds are retained to varying degrees.

Figure 1 is a simple representation of the process. In (A) a mixture of seven molecules each of ● and ▲ is introduced into the chromatographic system. In (B) they are distributed between the upper mobile and lower stationary phase, and in (C) the mobile phase has transported the dissolved molecules over a small distance: new equilibria between the phases are established. When this process has been repeated many times, as in (D), the compounds ● and ▲ are separated because their preference for one of the two phases differs strongly.

In practice, chromatography takes place on a plane or in a tube. The plane can be a sheet of paper (paper



**Figure 1** Schematic representation of the process of a chromatographic separation. (A) Sample injection; (B) partition between the two phases; (C) progression of the mobile phase and new equilibrium; and (D) separation of the two compounds after a number of partition processes.



chromatography), or an adsorbent that is fixed onto a sheet of aluminum or glass (thin-layer chromatography); the tube can be an open capillary with the stationary phase coated on the inner wall (open-tubular chromatography) or a packed tube (packed-column chromatography). This is explained in more detail in the following article.

Two important aspects of **Figure 1** should be noted:

1. The degree of preference of compound  $x$  for one of the phases is expressed by the partition coefficient  $K$ , also called distribution constant.

$$K_x = \frac{c_{\text{stat}}}{c_{\text{mob}}} \quad [1]$$

where  $c$  is the concentration (strictly speaking, activity) of the compound in the mobile and stationary phase, respectively, at equilibrium.

In **Figure 1**,  $K_{\blacktriangle}$  is  $5/2 = 2.5$ ,  $K_{\bullet}$  is  $2/5 = 0.4$ . If  $K = \infty$  the compound is totally adsorbed by the stationary phase and will not be transported through the chromatographic bed. If  $K = 0$  no retention takes place and the residence time in the column will be minimal.

2. The chromatographic bed seems to consist of small, discrete units that correspond, in this case, to approximately three particle diameters of the stationary phase, and the mobile phase is moving discontinuously. Although in practice it is unrealistic to assume such distinct zones, it is a helpful concept. The length of the chromatographic bed over which equilibrium is reached is called a 'theoretical plate'.

If the components of a mixture have different partition coefficients in a given chromatographic system, they can be separated if the number of plates is high enough. This process is impeded by the fact that the compound bands become broader as they travel through the layer or column. The effect of band broadening must be less than that of separation, otherwise complete resolution cannot be obtained.

Chromatography is based on the interaction of molecules with molecules (not with a field) and it is inherently governed by time-dominated processes. The  $x$ -axis of a chromatogram is time (and not energy as in spectroscopy). Chromatography and spectroscopy are totally orthogonal techniques, and consequently their online combination (spectroscopic investigation of the separated sample components) is a most powerful approach.

As an example of how efficient chromatography can be, **Figure 2** shows the separation of polymethoxylated flavone isomers found in orange oil by supercritical fluid chromatography with a packed column.

## The Chromatogram

If the compounds are eluted from the chromatographic bed, as is always the case in all types of tubular or column chromatography, usually an instrument, called a detector, is coupled to the outlet that responds to the eluting bands. According to the type of chromatography and to the compounds present in the sample mixture, detectors are highly sophisticated devices but in any case the chromatograms look similar. A simple chromatogram is shown in **Figure 3**.

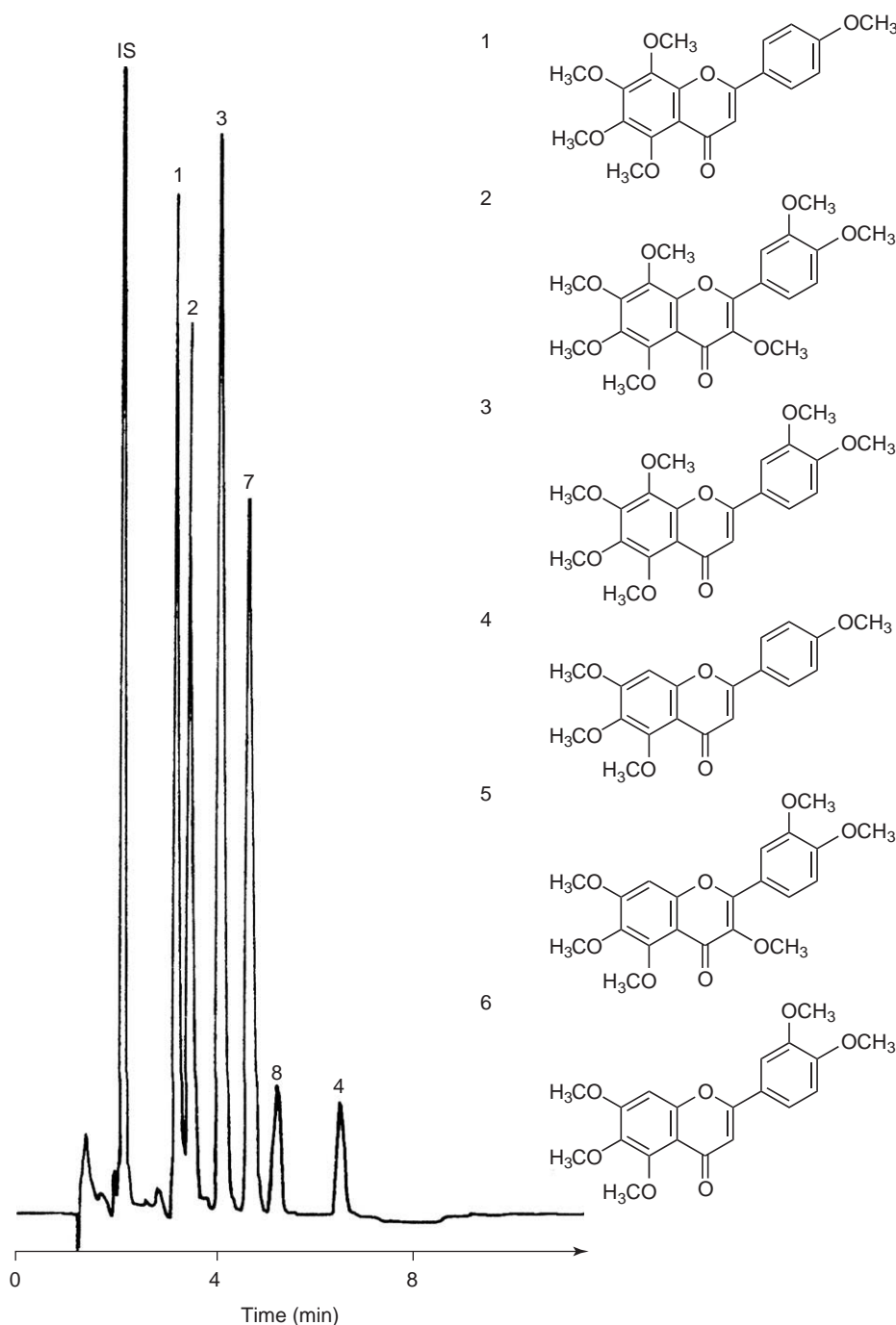
Owing to the band broadening processes, the bands, now called peaks, are of Gaussian (normal distribution) shape, as in **Figure 3**. In reality, the peaks often show a degree of asymmetry in that the trailing end is broader than the leading one. This phenomenon is known as tailing and for its mathematical description it is possible, for example, to overlay an exponential function.

The nomenclature in **Figure 3** are defined as follows:  $t_0$  is the breakthrough time or dead time (sometimes also called  $t_M$ ); time of a nonretained solute or of the mobile phase itself to travel through the chromatographic bed. The linear velocity of the mobile phase  $u$  (say, in  $\text{mm s}^{-1}$ ) can be calculated from  $t_0$  if the length of the column  $L_c$  is known:

$$u = \frac{L_c}{t_0} \quad [2]$$

$t_R$  is the retention time; time elapsed from sample injection to the maximum of a particular peak.  $t'_R$  is the net retention time,  $t'_R = t_R - t_0$ . For any solute,  $t_0$  represents the time spent in the mobile phase and  $t'_R$  the time spent in the stationary phase. To achieve a separation, the individual compounds of a mixture need to show differing net retention times in a given chromatographic system since on average all molecules spend the same time in the mobile phase.  $w$  is the peak width at the baseline.  $w$  is defined as the intersection of the inflection tangents with the baseline and is four times the standard deviation  $\sigma$  of the Gaussian peak:  $w = 4\sigma$ .

If the chromatographic separation is performed in a plane (paper or thin-layer chromatography), the separated bands are not eluted but remain in the layer. Usually, the development of the chromatogram is stopped before the front of the mobile phase reaches the opposite end of the plane. The separated bands then appear as spots although their concentration profile is also described by a (perhaps distorted) Gaussian function. This becomes obvious if the spots are scanned by a densitometer: a chromatogram as shown in **Figure 3** is obtained. **Figure 4**



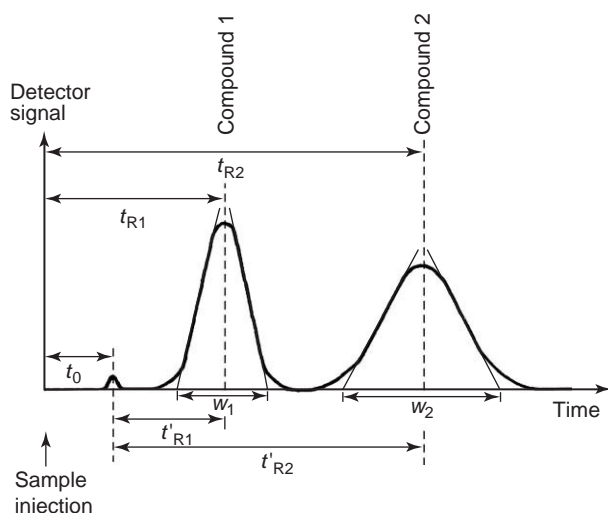
**Figure 2** Chromatographic separation of polymethoxylated flavones from orange oil. Method, packed column chromatography with a supercritical mobile phase; column, 4.6 mm i.d.  $\times$  25 cm length; stationary phase, silica 5  $\mu$ m; mobile phase, carbon dioxide–methanol 9:1; column inlet pressure, 220 bar; pressure drop over the column, 20 bar; temperature, 40 °C; detector, UV 313 nm. IS is the peak of coumarin used as internal standard. (Reprinted with permission from Morin Ph, Gallois A, Richard H, and Gaydou E (1991) Fast separation of polymethoxylated flavones by CO<sub>2</sub> SFC. *Journal of Chromatography* 586: 171.)

presents a schematic planar chromatogram with its characteristic definitions.  $z_m$  is the distance from the starting point to the front of the mobile phase and  $z_x$  is the distance from the starting point to the spot center of compound x.

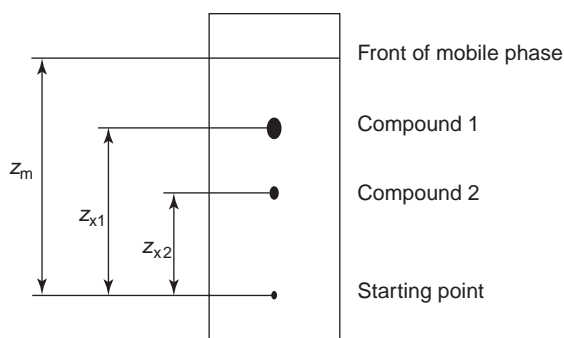
## Concepts Used in Chromatography

The chromatogram allows calculation of the basic parameters of the separation. These are highly useful tools in understanding and optimizing chromatographic





**Figure 3** A schematic elution chromatogram.



**Figure 4** A schematic planar chromatogram.

separations and, most importantly, utilizing the method for chemical analysis, although in some cases their basic scientific justification is weak; this is especially true for the plate number.

### Retention Factor

The scientific description of the appearance of a peak in an elution chromatogram is not given by its retention time but by its retention factor  $k$  (formerly called capacity factor  $k'$ ).

$$k = \frac{t'_R}{t_0} = \frac{t_R - t_0}{t_0} \quad [3]$$

where  $k$  is the mole ratio of a solute compound at equilibrium and is directly related to the partition coefficient  $K$  defined in eqn [1] by the phase ratio (as long as no overloading occurs, i.e. for low-injected sample mass):

$$k_x = \frac{n_{\text{stat}}}{n_{\text{mob}}} = K_x \frac{V_{\text{stat}}}{V_{\text{mob}}} \quad [4]$$

where  $n_{\text{stat}}$  is the number of moles of compound  $x$  in the stationary phase at equilibrium;  $n_{\text{mob}}$  is the number of moles of compound  $x$  in the mobile phase at equilibrium;  $V_{\text{stat}}$  is the volume of the stationary phase in the chromatographic system; and  $V_{\text{mob}}$  is the volume of the mobile phase in the chromatographic system.

If the stationary phase is an adsorbent, its characteristic property is not the volume but its specific surface, usually expressed in  $\text{m}^2 \text{g}^{-1}$  (because the sample molecules do not penetrate into the stationary phase but are retained on its surface only).

In a planar chromatogram the position of a spot is defined by its  $R_F$  value (sometimes called retardation factor) which is given by

$$R_F = \frac{z_x}{z_m} \quad [5]$$

The theoretical relation between  $R_F$  and  $k$  is given by

$$R_F = \frac{1}{1 + k} \quad [6]$$

In practice, deviations from this equation are found owing to nonideal behavior of planar chromatographic separation (e.g., depletion of solvent at the front of the mobile phase).

### Relative Retention, Separation Factor

Two compounds can only be separated if they differ in their respective  $k$  values. This is expressed by  $\alpha$ , the relative retention or separation factor:

$$\alpha = \frac{k_2}{k_1} = \frac{t_{R2} - t_0}{t_{R1} - t_0} = \frac{K_2}{K_1} \quad \text{with } k_2 \geq k_1 \quad [7]$$

A separation is only possible if  $\alpha > 1$ . The higher  $\alpha$  the easier is the separation and the fewer theoretical plates are necessary to resolve two neighboring bands; see eqn [11]. The relative retention is a measure of the inherent ability of a given chromatographic system to separate a mixture of interest, i.e. of its specificity. It has nothing to do with its quality, which is defined by its plate number.

### Theoretical Plate Number and Plate Height

A visual idea of a theoretical plate is presented in **Figure 1**. The mathematical expression of the number of theoretical plates  $N$  in a chromatographic system is obtained from the width of a peak in relation to its retention:

$$N = 16 \left( \frac{t_R}{w} \right)^2 = 5.54 \left( \frac{t_R}{w_{1/2}} \right)^2 = 2\pi \left( \frac{h_p t_R}{A_p} \right)^2 \quad [8]$$

where  $w_{1/2}$  is the peak width at half height of the peak;  $h_p$  is the peak height; and  $A_p$  is the peak area.

The greater the plate number of a chromatographic system, the more difficult the separation problems that can be solved. In principle the number of plates can be increased by using a longer column. Practical problems, mainly the pressure drop across the column and difficulties in manufacturing long packings or capillaries, set a limit at  $\sim 10^5$  plates in all chromatographic techniques.

To describe the quality of a chromatographic system the height equivalent to a theoretical plate, HETP or  $H$ , is more useful than  $N$ :

$$H = \frac{L}{N} \quad [9]$$

where  $L$  is the length of the chromatographic bed.

### Resolution

The separation factor  $\alpha$  describes the chromatographic system and its interaction with a sample from a thermodynamic point of view, but it says nothing about the actual resolution of two peaks. This is done by using the resolution  $R$  which compares the difference between the retention times with the peak widths:

$$R = 2 \frac{t_{R2} - t_{R1}}{w_1 + w_2} = 1.18 \frac{t_{R2} - t_{R1}}{w_{1/2(1)} + w_{1/2(2)}} \quad [10]$$

Two neighboring peaks have  $R = 1$  if their tangents of inflection touch each other at the baseline; their areas have a low degree of overlap. In Figure 2 this is approximately the case for peaks 1 and 2. There is negligible area overlap if  $R > 1.5$  (for peaks of similar size).

Resolution can be described as the interplay of retention factor  $k$ , separation factor  $\alpha$ , and plate number  $N$ :

$$R = \frac{1}{4}(\alpha - 1)\sqrt{N} \frac{k_1}{1 + \bar{k}} = \frac{1}{4} \frac{\alpha - 1}{\alpha} \sqrt{N} \frac{k_2}{1 + \bar{k}} \quad [11]$$

$\bar{k} = (k_1 + k_2)/2$  and it is assumed that  $N_1 = N_2$ . (Other representations of this resolution equation are also used.)

Equation [11] is key to understanding how a chromatographic separation can be influenced. The most important parameter is  $\alpha$  but it is often the most difficult one to change. The influence of  $k$  is rather weak, especially at higher retention factors.  $N$  acts only through its square root, i.e. doubling the length of a column increases the resolution only by a factor of 1.4. For mixtures with a small number of components it is most important to increase  $\alpha$  in order to

be able to perform rapid analyses: a special case is the separation of a pair of enantiomers. If complex mixtures need to be separated it is essential to choose systems with high  $N$  because it is impossible to optimize  $\alpha$  for all components present.

### Peak Capacity

The peak capacity  $n$  is a measure of the performance of a chromatographic system (although analysis time is not directly involved) because it describes how many peaks can be separated with  $R = 1$  within a given 'window' of  $k$  values:

$$n = 1 + \left( \frac{\sqrt{N}}{4} \right) \ln(1 + k_{\max}) \quad [12]$$

where  $k_{\max}$  is the retention factor of the last eluted peak.

Even excellent chromatographic columns have rather low peak capacities (e.g.,  $n = 86$  for  $N = 20\,000$  and  $k_{\max} = 10$ ). The number of adequately resolved peaks in a real chromatogram is in fact markedly lower because their  $k$  values are distributed in an unpredictable manner and because the peaks differ in size (see Statistical Peak Overlap below). A way to overcome this limitation is programmed elution, i.e., running temperature programs or solvent gradients.

### Separation Number, Trennzahl

In gas chromatography a common measure for characterizing the quality of a capillary is the separation number, SN (also called Trennzahl, TZ). This gives the number of peaks that can be placed with resolution  $R = 1.18$  between two members  $n$  and  $n + 1$  of a series of homologs:

$$SN = \frac{t_{R(n+1)} - t_{R(n)}}{w_{1/2(n+1)} + w_{1/2(n)}} - 1 \quad [13]$$

Whereas the plate number can only be determined in non-programmed (isothermal or isocratic) chromatography, SN is always obtained from a gas chromatogram with temperature programming.

### Statistical Peak Overlap

In a chromatographic system the peaks are eluted in accordance with their partition coefficients, i.e. as they are driven or retained by their physicochemical properties with regard to the phase system. For the observer the elution order of complex mixtures resembles an unpredictable pattern and in a first approximation a chromatogram must be looked at as a random mixture of peaks. (Exceptions are homologs, e.g.  $n$ -alkanes in mineral oil samples.)

Partial or total peak overlap is a common phenomenon that has to be taken into account.

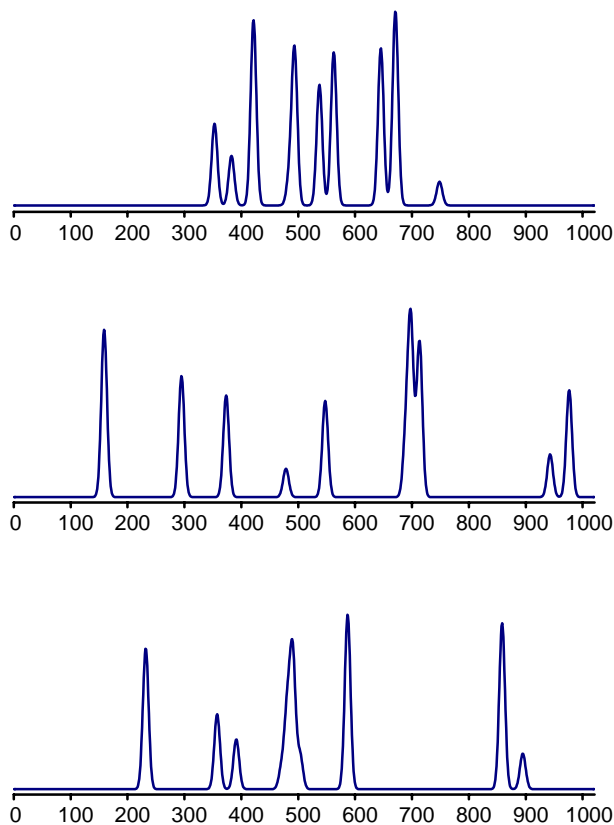
If  $m$  detectable analytes are present in the sample and if the chromatographic system has a peak capacity of  $n$ , then the average number  $p_x$  of  $x$ -tets (fused peaks that are composed of  $x$  single-component peaks) is

$$p_x = m \cdot e^{-2m/n} (1 - e^{-m/n})^{n-1} \quad [14]$$

In other words, the probability  $P$  that a given peak is single-standing with resolution  $R \geq 1.0$  is

$$P = e^{-2m/n} \quad [15]$$

In a chromatogram of 10 compounds, to be separated on a system with peak capacity 45, one can expect to find 6 or 7 singlets ( $p_1 = 6.4$ ), at least one doublet ( $p_2 = 1.3$ ), and perhaps even a triplet ( $p_3 = 0.25$ ). The probability that a peak of interest will be resolved is  $P = 0.64$  or almost two-thirds. This situation is shown in Figure 5.



**Figure 5** Artificial chromatograms generated with random numbers. Programmed separations are simulated with  $t_0 = 100$  s, end of chromatogram at 1000 s, and a peak capacity of 45. The width of each peak is 20 s. Peak sizes can vary between 1 and 100 area units. Ten compounds are present but none of the three chromatograms shows them all.

## Reduced Parameters

From an engineer's point of view it is highly desirable to describe the chromatographic system by reduced parameters. This helps to understand some very basic principles of chromatography and to compare very different systems. Reduced parameters are dimensionless. Instead of the plate height  $H$  and mobile phase linear velocity  $u$ , one defines the reduced plate height  $h$  and the reduced velocity  $v$  by the following two equations:

$$h = \frac{H}{d} \quad [16]$$

where  $d$  is the characteristic length of the chromatographic system; for packed columns this is the particle diameter, for open capillaries this is the diameter of the capillary:

$$v = \frac{ud}{D} \quad [17]$$

where  $D$  is the diffusion coefficient of the solute in the mobile phase.

If the quality of chromatographic columns is to be compared,  $h$  is a good measure because it is independent of their physical dimensions. In chemical engineering  $v$  is known as the Peclet number; it relates the carrier flow velocity to the diffusion coefficient, which is the parameter that limits the speed of chromatography. To obtain optimum results from a given chromatographic system,  $v$  should be kept within a well-defined range.

For all chromatographic techniques the time needed for an analysis can be predicted from plate number  $N$ , reduced plate height  $h$ , characteristic length  $d$ , reduced velocity  $v$ , diffusion coefficient  $D$ , and retention factor of the last eluted peak  $k_{\text{last}}$ :

$$t_R = \frac{Nhd^2}{vD} (1 + k_{\text{last}}) \quad [18]$$

## Band Broadening and the van Deemter Curve

As mentioned in the discussion of Figure 1, the bands become broader during their passage through the chromatographic bed. Several very different reasons are responsible for this unwanted behavior.

1. Eddy-diffusion and flow distribution, which occur in packed beds. Not all sample molecules travel along the same paths and these paths are not straight, therefore their residence times in the mobile phase differ. Moreover, the flow velocity between the particles of the stationary phase is not uniform but

exhibits a parabolic distribution, which again allows some molecules to travel faster than others. This is the so-called *A*-term of the van Deemter equation (see below). It is nonexistent for open capillaries. In packed columns it can be kept low by using particles of as uniform a size as possible.

2. Longitudinal diffusion. The sample molecules are not only transported passively by the mobile phase, they also move back and forth through their own diffusion. This effect cannot be influenced because it depends on the diffusion coefficient, but it is not (or should not be) a problem since it only manifests itself if the chromatographic process is too slow. However, then, it affects the separation significantly. This is the *B*-term.

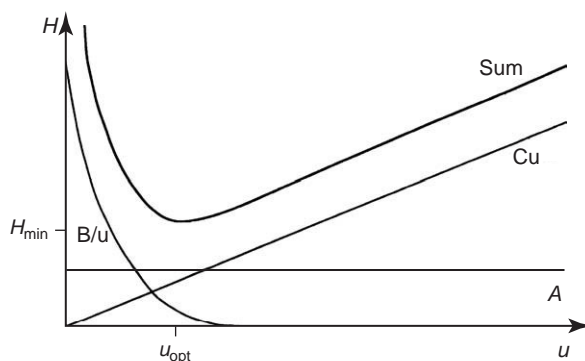
3. Resistance to mass transfer between the mobile and stationary phase. It takes the solute molecules some time to reach the stationary phase, to enter it, to remain there for a duration that is governed by statistics, and to leave the stationary phase and enter the mobile phase again. Molecules that in the same period are not engaged in mass transfer will be well ahead, along the chromatographic bed, thereby broadening the band. This is the *C*-term. It can be kept low by using low-viscosity mobile phases, packed columns with very fine particles, open capillaries of small inner diameter, and thin films of the stationary phase if this is a liquid.

Although a detailed theory of band broadening is complicated, it can be presented in a simple form through the van Deemter equation, which expresses the plate height *H* or the reduced plate height *h* as a function of *A*, *B*, *C*, and the linear flow velocity of the mobile phase:

$$H = A + \frac{B}{u} + Cu \quad [19a]$$

$$h = Av^{1/3} + \frac{B}{v} + Cv \quad [19b]$$

The equation  $H = f(u)$ , the van Deemter equation, is shown in Figure 6. One sees that *H* has a minimum at an optimum flow velocity that can be determined empirically or by knowledge of *A*, *B*, and *C* (which in many cases is not difficult). It is not recommended to work in the region left of the minimum because here the *B*-term becomes strongly dominant. If the *A*-term is missing, as in open capillaries, the minimum plate height will be decreased. Most important is the *C*-term because it limits the speed of chromatography: it is possible to work in the region to the right of the minimum, but only with a certain sacrifice of separation quality.



**Figure 6** The van Deemter curve, which describes the dependence of the height of a theoretical plate *H* on the linear flow velocity of the mobile phase *u*. Values are typically *H* in  $\mu\text{m}$ , *u* in  $\text{mm s}^{-1}$ .

## Chromatographic Analysis

Chromatography can be used for both qualitative and quantitative analysis.

### Qualitative Analysis

The retention time (as an easily determined value, especially when an electronic registration device is used), or more rigorously, the retention factor, and in planar chromatography the  $R_F$  value, are physico-chemical characteristics of a pure compound in a given chromatographic system. However, this system needs to be thoroughly defined. This means that the type of the stationary phase (and even the brand and batch), the composition of the mobile phase and its flow rate, the temperature and pressure, the column dimensions, and the detector properties all need to be known and kept more or less constant for a given analysis. Even the composition of the sample (e.g. the solvent used for dilution) can have a distinct influence on the separation. Owing to the limited peak capacity of all chromatographic methods, a *k* value determined in one system is not real proof of identity. It is also necessary to perform the analysis on a very different system (a good example would be adsorption and partition chromatography) and to compare again with a pure standard. Alternatively, positive identification can be obtained by coupling to a spectroscopic technique that provides additional qualitative information, especially mass spectrometry.

### Quantitative Analysis

It is a convenient fact that both the peak area and peak height of a well-defined (not necessarily Gaussian) peak are proportional to the mass of compound registered by the detector. This means that quantitative analysis can easily be performed using common techniques such as external or internal

standardization or standard addition. However, good resolution from all neighboring peaks is a prerequisite; poor resolution is prone to producing erroneous results and should be avoided whenever possible. The signal-to-noise ratio of the peak should be higher than 10. Even when these basic principles are observed, numerous pitfalls await the analyst. They can only be avoided by experience, by knowledge of the relevant literature, and by always being alert and aware of what one is doing. As a result chromatography can be considered a most powerful, rapid, and versatile technique.

*See also:* **Chromatography:** Overview; Multidimensional Techniques. **Gas Chromatography:** Detectors; Mass Spectrometry. **Liquid Chromatography:** Liquid Chromatography–Mass Spectrometry. **Thin-Layer Chromatography:** Overview; Principles.

## Further Reading

- Beesley TE, Buglio B, and Scott RPW (2000) *Quantitative Chromatographic Analysis*. New York and Basel: Dekker.
- Dyson N (1998) *Chromatographic Integration Methods*, 2nd edn. London: Royal Society of Chemistry.
- Felinger A and Pietrogrande MC (2001) Decoding complex multicomponent chromatograms. *Analytical Chemistry* 73: 619A–626A.
- Giddings JC (1991) *Unified Separation Science*. New York: Wiley-Interscience.
- Heftmann E (ed.) (1992) *Chromatography*, 5th edn. Amsterdam and New York: Elsevier.
- Miller JN (1988) *Chromatography – Concepts and Contrasts*. Chichester: Wiley.
- Poole CF (2002) *The Essence of Chromatography*. Amsterdam: Elsevier.
- Ravindranath B (1988) *Principles and Practice of Chromatography*. Chichester: Wiley.

## Multidimensional Techniques

**P J Marriott and D Ryan**, RMIT University, Melbourne, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Limitations of One-Dimensional Chromatography

Chromatography currently serves as the premier technique for the separation and analysis of an extensive range of samples and sample types, since intrinsic properties of the sample analytes can be accommodated by selecting the appropriate type of chromatography. For example, volatile compounds will be usually analyzed by gas chromatography (GC). One-dimensional chromatography can be defined as the use of one particular chromatographic method that utilizes only one basic mechanism (and one type of stationary and mobile phase) for component separation. The resolving power of a particular or discrete chromatographic separation can be evaluated in terms of its ‘peak capacity’, a term that was introduced in 1967 by Giddings to represent the number of compounds that can be placed side by side in a separation channel with a given minimum resolution performance for neighboring components. This concept assumes that compounds are evenly positioned in the separation dimension.

Unfortunately, the composition of complex samples when subjected to chromatographic analysis will ensure a more ‘random’ displacement of components in the separation space, with multiple occurrences of peak overlap. This is the main hindrance to one-dimensional chromatography for adequate sample analysis. In order to reduce the amount of peak overlap, one must significantly inflate the peak capacity of the chromatographic system, such that the peak capacity is excessively high compared to the actual number of components in the sample. This may not always be possible (depending upon the chromatographic system used), or can be achieved only at the expense of dramatically increasing the analysis time to obtain reasonable analyte resolution. Alternatively, multidimensional chromatography, which offers superior peak capacity compared to one-dimensional methods, can be used for the successful analysis and separation of complex samples.

### Multidimensionality in Chromatography

Multidimensional chromatography can be defined as the coupling of two or more chromatographic displacement systems for enhanced sample resolution. **Figure 1** is representative of the key features of a two-dimensional separation system. In respect of



standardization or standard addition. However, good resolution from all neighboring peaks is a prerequisite; poor resolution is prone to producing erroneous results and should be avoided whenever possible. The signal-to-noise ratio of the peak should be higher than 10. Even when these basic principles are observed, numerous pitfalls await the analyst. They can only be avoided by experience, by knowledge of the relevant literature, and by always being alert and aware of what one is doing. As a result chromatography can be considered a most powerful, rapid, and versatile technique.

*See also:* **Chromatography:** Overview; Multidimensional Techniques. **Gas Chromatography:** Detectors; Mass Spectrometry. **Liquid Chromatography:** Liquid Chromatography–Mass Spectrometry. **Thin-Layer Chromatography:** Overview; Principles.

## Further Reading

- Beesley TE, Buglio B, and Scott RPW (2000) *Quantitative Chromatographic Analysis*. New York and Basel: Dekker.
- Dyson N (1998) *Chromatographic Integration Methods*, 2nd edn. London: Royal Society of Chemistry.
- Felinger A and Pietrogrande MC (2001) Decoding complex multicomponent chromatograms. *Analytical Chemistry* 73: 619A–626A.
- Giddings JC (1991) *Unified Separation Science*. New York: Wiley-Interscience.
- Heftmann E (ed.) (1992) *Chromatography*, 5th edn. Amsterdam and New York: Elsevier.
- Miller JN (1988) *Chromatography – Concepts and Contrasts*. Chichester: Wiley.
- Poole CF (2002) *The Essence of Chromatography*. Amsterdam: Elsevier.
- Ravindranath B (1988) *Principles and Practice of Chromatography*. Chichester: Wiley.

## Multidimensional Techniques

**P J Marriott and D Ryan**, RMIT University, Melbourne, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

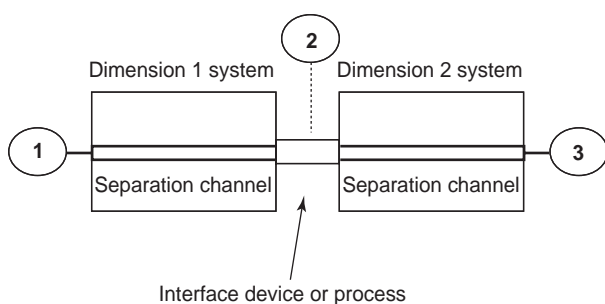
### Limitations of One-Dimensional Chromatography

Chromatography currently serves as the premier technique for the separation and analysis of an extensive range of samples and sample types, since intrinsic properties of the sample analytes can be accommodated by selecting the appropriate type of chromatography. For example, volatile compounds will be usually analyzed by gas chromatography (GC). One-dimensional chromatography can be defined as the use of one particular chromatographic method that utilizes only one basic mechanism (and one type of stationary and mobile phase) for component separation. The resolving power of a particular or discrete chromatographic separation can be evaluated in terms of its 'peak capacity', a term that was introduced in 1967 by Giddings to represent the number of compounds that can be placed side by side in a separation channel with a given minimum resolution performance for neighboring components. This concept assumes that compounds are evenly positioned in the separation dimension.

Unfortunately, the composition of complex samples when subjected to chromatographic analysis will ensure a more 'random' displacement of components in the separation space, with multiple occurrences of peak overlap. This is the main hindrance to one-dimensional chromatography for adequate sample analysis. In order to reduce the amount of peak overlap, one must significantly inflate the peak capacity of the chromatographic system, such that the peak capacity is excessively high compared to the actual number of components in the sample. This may not always be possible (depending upon the chromatographic system used), or can be achieved only at the expense of dramatically increasing the analysis time to obtain reasonable analyte resolution. Alternatively, multidimensional chromatography, which offers superior peak capacity compared to one-dimensional methods, can be used for the successful analysis and separation of complex samples.

### Multidimensionality in Chromatography

Multidimensional chromatography can be defined as the coupling of two or more chromatographic displacement systems for enhanced sample resolution. **Figure 1** is representative of the key features of a two-dimensional separation system. In respect of



**Figure 1** Schematic diagram of a multidimensional chromatography system. The two systems correspond to the methods employed, e.g., the GC or LC system. The separation channel may be an elution column or separation plane. Between the two systems is the process for transferring solute from the first to second dimensions. It may be a heart-cut process (GC–GC), a fraction collection step (LC–GC), a modulation process (GC  $\times$  GC), or a plate positional change for planar TLC. Some systems will allow or require a detection step between the two systems, such as a monitor detector in MDGC. In the figure, step 1 refers to the sample application/injection; step 2 to the interface or intermediate sample-processing device; and step 3 to the elution of the separated sample into a detection system.

separations, about two dimensions is the practical limit of such a system, although a higher-dimension analysis system can be devised if a third dimension comprising a spectroscopic detection step (e.g., mass spectrometry) is added. Multidimensional separation methods are particularly amenable to the analysis of complex samples since different separative mechanisms are exploited in each of the separate dimensions for the ultimate resolution of sample components. Successful separations are easily achieved, due to the expanded peak capacity of the multidimensional system. The extent of peak capacity enhancement will depend upon the type of multidimensional system used, and how eluate fractions from the first dimension are transferred to the subsequent dimension(s). This is an important aspect, which needs further elaboration.

Multidimensional techniques may be classified as hyphenated or comprehensive, depending upon the method of effluent transfer between coupled dimensions. The historical origins of multidimensional chromatography derive from simple heart-cutting techniques whereby only a specific retention region(s) is isolated and directed to the second dimension for further separation. This type of multidimensional separation represents a ‘hyphenated’ method, and is denoted using the hyphen symbol, for example, liquid chromatography–gas chromatography (LC–GC). In this example, selected effluent fractions from the LC are isolated and further separated by GC. The peak capacity of such hyphenated systems is the sum of the individual peak

capacities of the systems that are coupled. In comparison, comprehensive methods ensure that all solutes in the sample are fully subjected to each separation dimension of the multidimensional system in an independent or orthogonal manner. This can be achieved by directing all of the effluent from the first dimension to the subsequent dimension, or otherwise, taking a representative fraction of the effluent and directing it to the coupled dimension. What is fundamental in the latter instance is that the second dimension faithfully samples all peaks eluting from the first dimension. In comprehensive systems, the resulting peak capacity is the product of the individual peak capacities, and the multiplex sign is used to indicate a comprehensive multidimensional chromatography system (e.g., LC  $\times$  GC). It is important to recognize that simply joining two columns together and allowing solutes to pass directly through the two-column set without hindrance does not satisfy the definition of a comprehensive two-dimensional separation. It is clear that the most substantial gains in peak capacity are achieved using comprehensive multidimensional chromatography systems as opposed to hyphenated systems. This can be expressed mathematically using the following equations where total peak capacity ( $n_{\text{tot}}$ ) is expressed as

$$n_{\text{tot}} = \sum_{i=1}^m n_m$$

for a hyphenated system with  $m$  dimensions, each of capacity  $n_m$ ; and

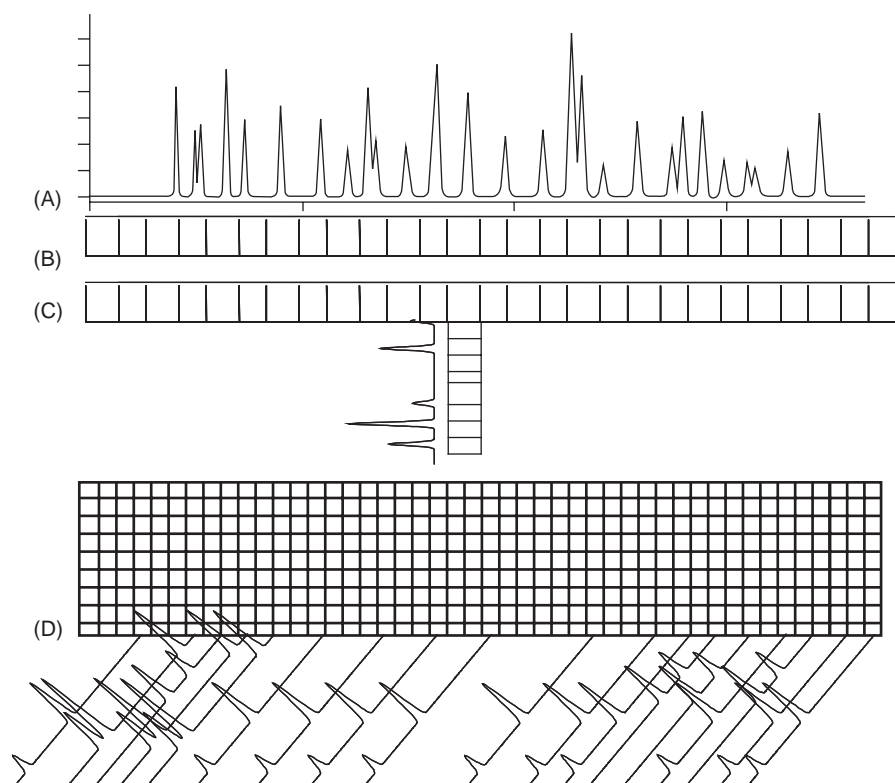
$$n_{\text{tot}} = \prod_{i=1}^n n_m$$

for a comprehensive system with  $m$  dimensions.

**Figure 2** further illustrates the concept of hyphenated and comprehensive multidimensional chromatography with respect to total peak capacity. In the figure, the single channel separation in ‘A’ is represented as an equivalent capacity diagram in ‘B’. The conventional multidimensional method traditionally selects a given region to apply to the second dimension, which then yields greater capacity for that selected region, as shown in ‘C’. The total capacity is expanded by the additional capacity of each selected region. However, only a few regions can usually be subjected to the second dimension. In the comprehensive method, by employing a suitable interface and a higher sampling frequency from first to second dimension, it is possible to apply the multidimensional method to the whole first dimension displacement. ‘D’ illustrates this concept.

Giddings defined two main parameters, which are fundamental to multidimensional separations. The





**Figure 2** A schematic illustration comparing the peak capacities of chromatographic systems with different dimensionalities. (A) shows the one-dimensional chromatographic trace, and (B) represents the equivalent one-dimensional peak capacity. (C) and (D) show the comparative gains in peak capacity achieved using a hyphenated two-dimensional system and a comprehensive two-dimensional system, respectively.

first simply states that components of a mixture must be subjected to two or more separation steps or mechanisms in which their displacements depend on different factors. The second, however, is critical, and states that when two or more components are substantially separated in a single step, the integrity of the separation must be maintained until the completion of the separative operation. Adherence to the latter parameter may therefore rule out single tandem (directly coupled) column arrangements whereby components separated by the first column may remerge on the second column.

The beauty of multidimensional chromatography rests in the fact that adequate separation of two components along any separation axis is sufficient for their ultimate resolution. This means that there can still be serious component overlap along any one axis, so long as physical deconvolution occurs along another axis. Superior multidimensional systems, however, depend upon the coupling of powerful one-dimensional systems whereby the final resolution achieved is dependent upon differences between the two (or more) dimensions. The highest resolution is gained when there is no correlation between the

separation mechanisms, in which case we say that the mechanisms are 'orthogonal'. This means that the mechanisms used for separation are very much different. A simple example of an orthogonal multidimensional system would be LC–GC (or LC  $\times$  GC) since in the LC case, the mobile phase is a liquid, and separation arises due to (for example) compound hydrophobicity, whereas in the GC system, the mobile phase is a gas, and separation is determined by compound boiling point or volatility. There is no correlation between the two separation methods; they are orthogonal, and hence the resulting dimensionality of the system is 2.

Dimensionality is an important term, which can be used to describe the nature of the sample, or of the multidimensional system. System dimensionality simply refers to the number of coupled separation stages or dimensions used in the multidimensional system. For example, a single column liquid chromatograph would have a dimensionality of 1, whilst a LC–GC system described above has a dimensionality of 2. On the other hand, sample dimensionality, introduced by Giddings in 1995, is used to describe the intrinsic nature of the sample, and can be defined

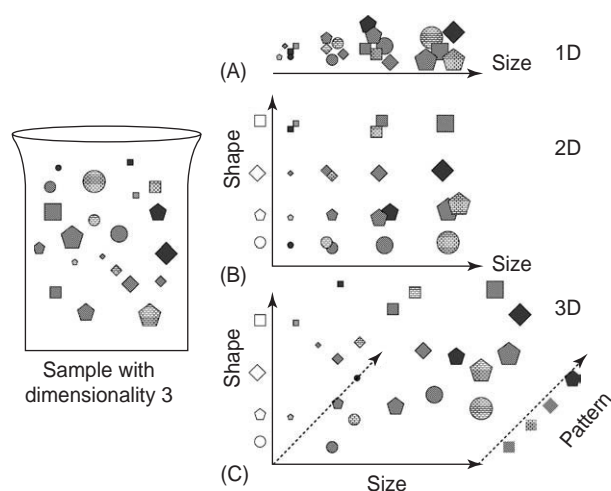
as the number of individual variables that must be specified to identify components in a sample. For example, in a mix containing only straight chain alkanes, the sample dimensionality would be based exclusively upon carbon number, thus yielding a dimensionality of 1. Naturally, as the sample becomes more complex, composed of multiple chemical classes (e.g., aromatics), more parameters or mechanisms of separation would be needed to define the sample, therefore, increasing its dimensionality. It is also important to consider whether we have available separation tools that can be employed to resolve or differentiate the identified sample dimensionality. If not, then these sample dimensions would collapse into a single dimension.

Theoretically, optimum separation is achieved when the sample dimensionality and system dimensionality are equivalent, resulting in an ordered separation (Figure 3). In the above example, only one separation dimension would be required to analyze the alkane sample. If, however, the dimensionality of the sample exceeds that of the system, sample components will not be resolved in an orderly fashion, but rather a disordered or chaotic separation will result. In Figure 3, three descriptors are required to define the sample – shape, pattern, and size. In a chemical sense, these might be molar mass, polarity, and molecular shape. As more dimensions of separation are applied, greater definition of the mixture components is achieved. Unfortunately, for very complex samples, the sample dimensionality will be

significantly greater than the system dimensionality, since the development of a multidimensional system with a system dimensionality equivalent to that of the sample cannot be achieved. Alternatively, a multidimensional system with a system dimensionality greater than the sample dimensionality is merely wasted, in that the increased peak capacity of the system will not yield any significant improvement in separation of the sample. For example, if a sample is comprised of straight chain and branched alkanes (sample dimensionality 1), then a system of dimensionality greater than 1 will probably not provide any further separation advantage than a single separation dimension. In contrast, a sample comprising alkanes and aromatics (sample dimensionality 2) will require a system dimensionality of 2 to provide full class separation. In this case, a nonpolar GC column in the first dimension, and a second column with a stationary phase of high percentage phenyl content (which retains aromatic compounds compared to aliphatics) should provide separation of alkanes and aromatics that had coeluted on column 1. Clearly, knowledge of sample dimensionality should therefore be exploited, when possible, to enable the most appropriate multidimensional technique to be chosen to suit the sample type.

## History of Multidimensional Chromatography

The origins of multidimensional chromatography derive from paper chromatography, with the realization that different eluents could be applied to the paper stationary phase in different directions. This procedure soon progressed with the development of two-dimensional thin-layer chromatography (TLC) by Kirchner and his team in 1951. Development of multidimensional column chromatography techniques came later, and in 1958, the first multidimensional GC analysis (GC–GC) was conducted by Simmons and Snyder for the analysis of C<sub>5</sub>–C<sub>9</sub> petroleum fractions. This system, referred to as ‘two-stage gas–liquid chromatography’, utilized a valve for the online transfer of heart-cut fractions from the first (nonpolar) column to the second (polar) column. The development of multidimensional LC, however, proceeded at a much slower pace, and it was not until 1978, 20 years after the development of multidimensional GC, that an online multidimensional LC system (LC–LC), utilizing orthogonal columns was developed by Erni and Frei for the analysis of complex plant extracts. The first dimension was composed of a gel permeation chromatographic column, whilst a reversed-phase column was used as the



**Figure 3** Schematic illustration showing the successive improved resolution of a three-dimensional sample mixture comprising dimensions ‘shape’, ‘size’, and ‘pattern’. (A) The 1D separates by size, but this leads to poor resolution of pattern and shape; (B) adding a shape selective dimension leaves some overlapping pattern components; and (C) only the three-dimensional system permits full resolution of all components.

second dimension. Heart-cut fractions from the first column were collected and transferred to the second column using an eight-port dual loop injection valve.

It is beyond the scope of this article to trace the development of all hyphenated multidimensional chromatographic systems, since this would also encompass the development of hyphenated chromatographic and spectroscopy techniques (such as LC-MS) and also chromatography coupled to electrophoretic techniques such as liquid chromatography-capillary electrophoresis (LC-CE). What is important to realize is that multidimensional chromatography was initially conducted using hyphenated systems, since hyphenated couplings could be achieved simply, and enabled the better characterization of specific regions of retention complexity via heart-cutting. Peak capacity gains were significant, but only amounted to the addition of the individual component peak capacities for each time a given dimension is employed. Subsequently, these methods then progressed to comprehensive techniques capable of producing multiplicative gains in peak capacity.

The first report of comprehensive two-dimensional chromatography was made by Bushey and Jorgenson in 1990 using LC  $\times$  LC. A microbore cation exchange column was used as the first dimension, and operated under gradient elution conditions. Effluent from this column was collected using an eight-port valve, and forced onto a second, size exclusion, column using a second LC pump. The entire system was automated and operated under computer control, and in this system, the entire first column effluent was separated by the second column, yielding a comprehensive analysis. It is noteworthy that LC  $\times$  CE was also developed by Bushey and Jorgenson, but in this instance, only 5% of the LC effluent was applied to and separated by the CE step. Still, this analysis can be considered as comprehensive since all LC peaks were representatively sampled and separated on both dimensions. In the following year, comprehensive GC  $\times$  GC was introduced by Liu and Phillips. In this work, a two-stage elevated temperature (thermal) modulator was used to interface two serially coupled and somewhat orthogonal GC columns. The modulator used ensured the comprehensive analysis of the entire solute composition of a hydrocarbon standard mixture and a coal liquids sample. The development of the principles of comprehensive separation systems in these landmark studies has served as the foundation for the development of many more comprehensive chromatography technologies, which are currently in use today.

The most obvious benefit of comprehensive analyses rests in the multiplicative gains in peak capacities that are attainable when orthogonal dimensions

are coupled in such a way as to yield the comprehensive result. Furthermore, the need for timed heart-cuts, as mandated in multidimensional analyses for selection of target regions, are redundant in comprehensive systems, thereby simplifying the potential for automation. Critical to the success of comprehensive multidimensional systems is that the interface between the coupled systems should be reliable and ensures the faithful transferring of solutes from one dimension to the next. At present, the focus of much research is the development of improved interfaces, for example, in GC  $\times$  GC, the development of improved modulators.

## Two-Dimensional Planar Chromatography and Coupled-Column Chromatography

There are two distinct approaches used when undertaking multidimensional chromatography. The first is known as two-dimensional planar separation, or discrete separation, which is achieved by using the two right angle dimensions of a continuous surface or thin film bed for separation. This technique derives from paper chromatography, and has progressed to such methods as multidimensional thin-layer chromatography (MD-TLC). Discrete separation is achieved since an isolated sample is applied to the two-dimensional bed, generally at a corner position, and development proceeds along the two planar axes. Spots, corresponding to individual compounds, appearing at discrete positions on the planar surface reflect the resulting separation.

The second type of multidimensional chromatography is coupled column or continuous separation. As the name suggests, this technique relies on serially coupled chromatography columns for sample resolution. Partially resolved effluent fractions from the first column are sequentially directed to a second column, or series of columns, with different separation capabilities, for subsequent separation. It is important in this multidimensional method that the columns are carefully chosen to maximize their orthogonality to ensure optimum resolution is achieved. Furthermore, for the success of this method, it is important that the effluent cuts taken from the first column are sufficiently small so as to minimize the number of components in each cut, and therefore increase the probability of their ultimate separation. Typically, cuts are taken at about the duration of peak standard deviation timescale. This ensures the first dimension separation is not significantly degraded.

Choosing between a two-dimensional planar separation and a coupled column separation will largely depend upon the sample type and its complexity. Contrasting the two techniques shows that planar separations give a very broad characterization of the sample. These techniques are generally quite simple to incorporate, and the two-dimensional separation is clearly illustrated and easily interpreted. Quantification of the isolated components, however, is complex. The main advantage of two-dimensional planar chromatography, however, is that if sufficiently good resolution is achieved along one axis, the separation cannot be destroyed by any relative displacement along the second axis. This is not always the case in coupled column separations whereby separation in the first dimension is not always maintained in the subsequent dimension(s), and components can remerge, thus ruining the initial resolution. Nevertheless, coupled column chromatography systems are advantageous because they can be used to target particular areas of complexity in the elution space through the collection of specific eluate fractions. These fractions can then be further separated as necessary. Furthermore, additional separation mechanisms can be easily added to the existing coupled column arrangement, giving the coupled column arrangement significant flexibility. The use of conventional detectors simplifies quantification.

## Considerations for Multidimensional Couplings

In order to maximize the resolving power of multidimensional systems, very different, and uncorrelated, one-dimensional systems should be coupled. When high correlation occurs, the two-dimensional (or  $n$ -dimensional) space effectively reduces back to a

one-dimensional (or lower-dimensional) space with little improvement in separation power. Ideally, one-dimensional systems with high peak capacities should be coupled, since according to the statistical model of overlap theory, a sample should have fewer components than 37% of the column peak capacity in order to be relatively certain that acceptable peak resolution in a one-dimensional system can be achieved for a purely random analysis.

However, for various reasons coupling of particular one-dimensional systems will not always produce effective multidimensional separations, and therefore couplings should be carefully considered. First, consideration should be given to the sample type and properties, and whether it is amenable to analysis by a certain chromatographic method. It would be illogical, for example, to analyze uncharged species using ion chromatography (IC). Secondly, coupling of chromatographic methods must be physically compatible with respect to mobile phase composition (Table 1); whilst multidimensional GC is relatively straightforward, direct coupling of particular LC methods, such as normal-phase (NP) with reversed-phase (RP) chromatography, are generally problematic due to the incompatibility of the two mobile phases. Mobile phase incompatibility can sometimes be accommodated, but this will depend on the separative displacement processes used by the coupled chromatographic system. Likewise, a GC step can follow a suitable liquid-phase separation such as NP-LC, but LC will not be appropriate as a downstream separation procedure after a GC first dimension – the gas phase separated solutes cannot be effectively re-introduced into a liquid phase.

Giddings introduced the notion of a sequential or simultaneous displacement process to characterize the type of displacement used in a multidimensional arrangement. Thus, a sequential displacement is

**Table 1** Compatibility table: Possible couplings of selected two-dimensional column chromatographic systems based upon dimension compatibility. Note that the coupling order is important. Entries do not imply that the method has been implemented in practice. The coupling may be online direct coupling (e.g., GC–GC) or may involve special interfaces that allow phase isolation in the coupled method (e.g., vaporization of liquid phase in discrete sampling injection in NP-LC–GC)

First dimension	Second dimension						
	GC	NP-LC	RP-LC	SFC	IC	CE	SEC
GC	✓						
NP-LC	✓	✓				✓	✓
RP-LC	✓		✓			✓	✓
SFC	✓	✓		✓			✓
IC					✓	✓	
CE	✓					✓	✓
SEC	✓	✓	✓		✓	✓	✓

GC, gas chromatography; NP-LC, normal-phase liquid chromatography; RP-LC, reversed-phase liquid chromatography; SFC, supercritical fluid chromatography; IC, ion chromatography; CE, capillary electrophoresis; SEC, size exclusion chromatography.

essentially achieved using two (or more) separation processes, which occur in different media, under different conditions. The initial separation achieved in the first dimension is transferred in linear zones to the subsequent dimension of differing composition, for further displacement. Sequential multidimensional separations therefore offer significant flexibility to the analyst, and can be used to overcome mobile phase incompatibilities between the systems. Thus, using the example described in the previous paragraph, fractions eluting from a normal-phase separation could be collected, evaporated to dryness, and reconstituted in a polar solvent, which would be compatible for subsequent separation using a reversed-phase system.

Conversely, simultaneous displacements are much more rigid, in that the displacements take place in the same media, under the same conditions, such as in the same or similar solvent, at the same temperature and pressure. Therefore, chromatographic techniques using different states of mobile phases (i.e., gas, liquid, or solid) cannot be described as simultaneous. In 1984, Giddings described the coupling of chromatographic methods for simultaneous displacements as redundant, and predicted that no combination would enhance the separation gained under one-dimensional analysis; enhanced separation would only occur by coupling a chromatographic method with a separative method based upon field-induced displacements, such as electrophoresis. However, it cannot be disputed that GC  $\times$  GC using coupled, 'somewhat orthogonal' columns inside a single GC oven constitutes a powerful multidimensional system achieved using simultaneous displacement.

Sequential multidimensional chromatographic methods can be conducted either online or offline. A typical offline LC–GC method will be the collection of polarity-defined fractions of an oil separated by NP-LC into saturates, cyclics, aromatic, heteroatomic fractions, etc., with subsequent separation of each fraction by high-resolution GC. Similarly, NP-LC can be successfully coupled online with GC. In the online set up, fractions from the liquid chromatograph are transferred to the gas chromatograph using one of two techniques. The first is known as 'concurrent eluent evaporation'. In this method, as the name suggests, the eluate from the liquid chromatograph is completely evaporated during its introduction into the gas chromatograph injector. This method is simple and enables relatively large volumes of LC effluent to be transferred; however, it is restricted to the analysis of solutes with intermediate-to-high elution temperatures. The alternative is known as the retention gap method, whereby a valve

is positioned between the LC and GC interface, and an on-column injector and uncoated precolumn and retaining column are positioned prior to the analytical GC column. Sample is introduced into the GC at a temperature below that of the LC eluent boiling point, resulting in solvent trapping effects. The retention gap method is best used for the analysis of highly volatile solutes since the solutes of interest must elute at least 50°C above the boiling point of the mobile phase. In fact, the retention gap method is undermined by the introduction of nonvolatile materials, which may build up in the precolumn causing peak broadening, and is not compatible with water-containing LC eluents since water reacts with the precolumn and can cause active sites within the retention gap.

Online techniques have the advantage of being more rapid than offline techniques. Online processes can be automated and ensure higher sample throughputs as opposed to offline processes. Generally, the mobile phases of the coupled methods need to be compatible, or else the equipment must be capable of performing a suitable solvent changeover prior to the solutes entering the next column. Often, online multidimensional systems will require additional equipment such as valves and pumps, which will require increased technical maintenance, and vigilance to ensure devices such as valves perform consistently for chromatographic reproducibility. On the other hand, offline techniques are generally quite simple to incorporate and do not require any additional sophisticated equipment. However, these techniques have the disadvantage of requiring more time, in that solvent evaporation may be required, or modification of the first dimension fractions must be undertaken prior to subsequent chromatography. The need for increased sample handling of the fractions can be detrimental to solute recovery, and can introduce the risk of sample contamination.

## Status of Multidimensional Chromatography

It is only in recent decades that multidimensional chromatography has been given adequate attention, and the capabilities of multidimensional chromatographic systems realized. Still, the general practice of multidimensional chromatography, as opposed to conventional one-dimension analysis, is limited, as is the awareness and/or appreciation of these techniques. Fortunately, once a chromatographer learns about multidimensional techniques that can be applied to their own field of work, their fascination and curiosity surrounding such techniques is



usually aroused. Work which may have taken extensive periods of time and required tedious sample preparation using one-dimensional chromatography can be replaced with new and faster methods offering unsurpassed and almost unimaginable possibilities for solute resolution. Sample preparation prior to instrumental analysis can be reduced, and multidimensional methods can be used for sample enrichment and selective resolution. Thus, there are ample opportunities available to chromatographers, and attendant advantages from, application of multidimensional separations for sample analysis.

There is often a misconception that multidimensional systems are difficult and costly to set up and implement as a routine tool for analysis. It is true that in some instances additional equipment such as pumps and switching valves for multidimensional LC methods will need to be purchased, but once correctly installed, the costs and additional system maintenance required is insignificant when the improvements in resolution are considered, especially when most online multidimensional techniques proceed in the same time it would take to achieve a typical one-dimensional separation. Currently, some comprehensive techniques, such as GC  $\times$  GC, are overshadowed by a lack of computer software for integration purposes and data reduction. However, it is only a matter of time before user-friendly multidimensional data presentation packages are developed and are commercially available.

At present, most multidimensional chromatography methods are two-dimensional, composed of two one-dimensional systems coupled together. It is foreseeable that these two-dimensional technologies will proceed to three-dimensional systems and beyond. As the number of dimensions increases, and peak capacity further expands, more specific and unambiguous information is gained from each dimension, resulting in superior sample characterization possibilities. It can be predicted that in the future one-dimensional chromatographic analyses of complex samples will be redundant, replaced with faster and more efficient multidimensional methods.

In this article, the role of the detector has not been discussed. At the minimum the detector operates and is chosen exactly as that in a one-dimensional chromatography experiment. Whilst it was alluded

that spectroscopic detection constitutes a second dimension in a two-dimensional (hyphenated) analytical system, it is apparent that multidimensional separations with spectroscopic detection may be termed a three-dimensional system. The obvious and most powerful implementation of this is the use of mass spectrometry. Discussion of chromatography–mass spectrometry is beyond the scope of this article, and the reader is referred to appropriate articles in this text.

*See also:* **Chromatography:** Overview; Principles. **Gas Chromatography:** Overview; Principles; Multidimensional Techniques; Online Coupled LC–GC; Mass Spectrometry. **Liquid Chromatography:** Multidimensional.

## Further Reading

- Cortes HJ (ed.) (1990) *Multidimensional Chromatography*. New York: Dekker.
- Davis JM and Giddings JC (1983) Statistical theory of component overlap in multicomponent chromatograms. *Analytical Chemistry* 55: 418–424.
- Giddings CJ (1984) Two-dimensional separations: concept and promise. *Analytical Chemistry* 56: 1258A–1270A.
- Giddings JC (1962) Theory of minimum time operation in gas chromatography. *Analytical Chemistry* 34: 314–319.
- Giddings JC (1987) Concepts and comparisons in multidimensional separation. *Journal of High Resolution Chromatography* 10: 319–323.
- Giddings JC (1995) Sample dimensionality: a predictor of order–disorder in component peak distribution in multidimensional separation. *Journal of Chromatography A* 703: 3–15.
- Liu Z and Lee ML (2000) Comprehensive two-dimensional separations using microcolumns. *Journal of Microcolumn Separation* 12: 241–254.
- Majors RE (1980) Multidimensional high performance liquid chromatography. *Journal of Chromatographic Science* 18: 571–579.
- Marriott PJ and Kinghorn RM (2001) Multidimensional and comprehensive multidimensional gas chromatography. In: Handley AJ and Adlard ER (eds.) *Gas Chromatographic Techniques and Applications*. Sheffield: Sheffield Academic Press.
- Mondello L, Lewis AC, and Bartle KD (2001) *Multidimensional Chromatography*. Chichester: Wiley.

# CHROMIUM

**M Sperling**, University of Muenster, Muenster, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The special interest that exists for chromium arises from the fact that chromium in its trivalent oxidation state is considered to be essential for human and animal nutrition, while the hexavalent species have been identified as being toxic, causing cancer and DNA damage. On the other hand, chromium in its various forms is of commercial importance in a wide range of products meeting not only the perceived needs of the society but also leading to a significant emission of chromium to the environment. Since both the chemistry and the biological activity of chromium compounds depend on the valence state of chromium, it is evident that the determination of total concentrations for answering questions about risks for man and its environment is not only insufficient but irrelevant. Speciation analysis can seldom be more significant for the understanding of the behavior of a metal in the natural or occupational environment than it is in the case of chromium. The determination of the distribution of chromium between the two mainly occurring oxidation states, however, is a challenging analytical task, since the analytical approach has to assure that possible transformations between the two forms during sampling, sample preparation, separation, and detection do not occur or at least do not corrupt the analytical result.

## Origin and Nature

Chromium is the seventh most abundant element on earth, most chromium residing in the core and mantle. In the earth's crust chromium ranks 21st, the average concentration being estimated to be  $185 \text{ mg kg}^{-1}$ , much lower than that for the earth as a whole, estimated to be  $3700 \text{ mg kg}^{-1}$ . In its crustal distribution, it shows a particular preference for ultrabasic and basic rocks and specifically eschews the feldspar minerals. Apart from the rare mineral crocoite ( $\text{PbCrO}_4$ ) all naturally occurring chromium is found in the trivalent state. The most important and only commercial source of chromium is chromite ore, in which the chromic oxide content falls typically within the range 15–65%. The main deposits

**Table 1** Some physical properties of chromium

Atomic number	24
Relative atomic mass	51.996
Electronic configuration	$[\text{Ar}]3d^54s^1$
Melting point ( $^{\circ}\text{C}$ )	1857
Boiling point ( $^{\circ}\text{C}$ )	2672
Relative density ( $20^{\circ}\text{C}$ )	7.19

are in southern Africa with 96% of the known reserves.

## Chemical and Physical Properties

Chromium, a member of Group VIB of the periodic table, exists in four stable isotopes with the following abundance:  $^{50}\text{Cr}$  (4.31%),  $^{52}\text{Cr}$  (83.76%),  $^{53}\text{Cr}$  (9.55%), and  $^{54}\text{Cr}$  (2.38%). It has five short-living radioactive isotopes (half-life in brackets),  $^{48}\text{Cr}$  (23 h),  $^{49}\text{Cr}$  (41.9 min),  $^{51}\text{Cr}$  (27.8 days),  $^{55}\text{Cr}$  (3.5 min), and  $^{56}\text{Cr}$  (5.9 min), but only  $^{51}\text{Cr}$  is commonly used for tracer studies. Some of the basic physical properties of chromium are summarized in Table 1.

Chromium is a steel-gray, lustrous, hard, brittle metal that takes a high polish. It dissolves readily in nonoxidizing mineral acids but not in cold aqua regia or nitric acid because of passivation. Because of this resistivity against oxidizing attack, chromium is widely used as a protective corrosion inhibitor. The inorganic chemistry of chromium is not only rich in its variety of colors (which was the basis for the naming of the element), but also in its many oxidation states and the geometry of its compounds. Being a typical transition metal, chromium can occur in any of the oxidation states from  $-II$  to  $+VI$ , but it is not commonly found in oxidation states other than 0, III, and VI. Chromium can be present in the environment as  $\text{Cr(VI)}$  or  $\text{Cr(III)}$ , the latter can be present either in soluble form or in particulate form.

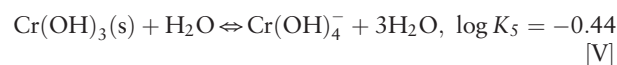
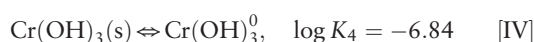
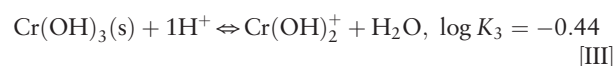
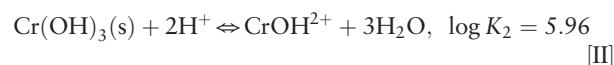
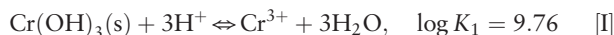
## Types of Compounds

$\text{Cr(III)}$  exhibits a  $d^3$ -electron configuration and forms many compounds including numerous complexes with oxygen and nitrogen being the preferred coordination partners (Table 2). Kinetically, these compounds are inert in keeping with the half-filled  $t_{2g}$  level of the  $d^3$ -configuration in octahedral geometry. The violet hexaquo species is inert to ligand exchange, with a ligand exchange rate of  $k = 3.5 \times 10^{-6} \text{ s}^{-1}$ , i.e., a half-life of a few days.

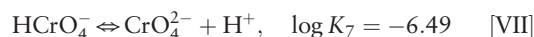


Cr(III) hydroxo complexes are expected to be the dominant species in natural waters, including  $\text{CrOH}^{2+}$ ,  $\text{Cr(OH)}_2^+$ ,  $\text{Cr(OH)}_3^0$ ,  $\text{Cr(OH)}_4^-$ ,  $\text{Cr}_2(\text{OH})_2^{4+}$ , and  $\text{Cr}_3(\text{OH})_4^{5+}$ . Polynuclear complexes

of Cr(III) have been reported in the literature, but these species are of no significance in natural systems, especially in solutions of low total chromium concentrations. Under the assumption of thermodynamic equilibrium, the distribution of these complexes can be calculated from thermodynamic data shown in reactions [I]–[V]:



Cr(VI) exhibits  $d^0$ -electron configuration and forms complexes mainly with oxo- or hydroxo-ligands with a tetrahedron configuration. Cr(VI) may be present in aqueous solutions as chromate ( $\text{CrO}_4^{2-}$ ), dichromate ( $\text{Cr}_2\text{O}_7^{2-}$ ), hydrogen chromate ( $\text{HCrO}_4^-$ ), dihydrogen chromate (chromic acid,  $\text{H}_2\text{CrO}_4$ ), hydrogen dichromate ( $\text{HCr}_2\text{O}_7^-$ ), trichromate ( $\text{Cr}_3\text{O}_{10}^{3-}$ ), and tetrachromate ( $\text{Cr}_4\text{O}_{13}^{2-}$ ) (Figure 1). The last three ions have been detected only in solutions of  $\text{pHCrpmol l}^{-1}$ . Thermodynamically, the aqueous equilibria are as in reactions [VI]–[VIII]:

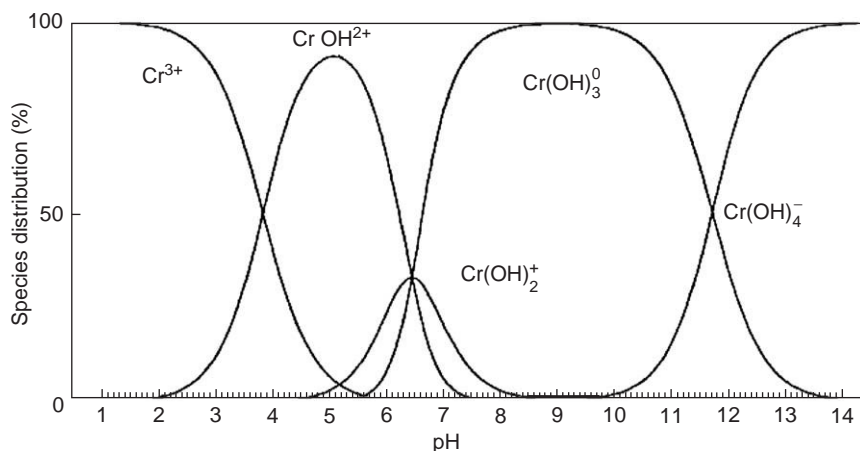


**Table 2** Representative chromium compounds and their formal oxidation states and geometry

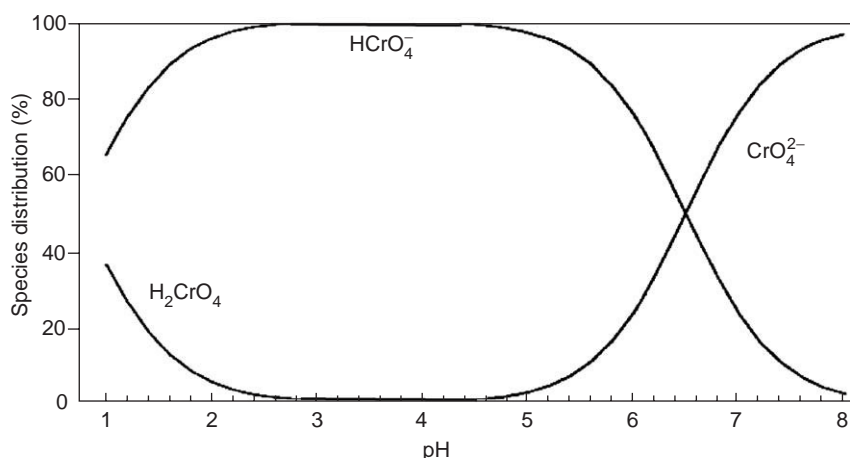
Oxidation state	Coordination number	Geometry	Compound
– II	5	Trigonal bipyramidal	$[\text{Cr}(\text{CO})_5]^{2-}$
– I	6	Octahedron	$[\text{Cr}_2(\text{CO})_{10}]^{2-}$
0	6	Octahedron	$[\text{Cr}(\text{CO})_6]$
+ I	6	Octahedron	$[\text{Cr}(\text{SCN})_6]^+$
+ II	4	Twisted tetrahedron	$\text{CrCl}_2(\text{CH}_3\text{CN})_2$
	5	Trigonal bipyramidal	$[\text{Cr}(\text{Me}_6\text{tren})\text{Br}]^+$
	6	Distorted Octahedron	$\text{CrCl}_2$
	7	Capped trigonal prism	$[\text{Cr}(\text{CO})_2(\text{diars})_2\text{X}]^+$
+ III	3	Planar	$\text{Cr}(\text{NPr}_2)_3$
	4	Twisted tetrahedron	
	5	Trigonal bipyramidal	$[\text{Cr}(\text{CH}_2\text{SiMe}_3)_4]^-$
	6 <sup>a</sup>	Octahedron	$[\text{Cr}(\text{H}_2\text{O})_6]^{3+}$
+ IV	4	Tetrahedron	$\text{Cr}(\text{OC}_4\text{H}_9)_4$
	6	Octahedron	$[\text{CrF}_6]^{2-}$
+ V	4	Tetrahedron	$\text{CrO}_4^{3-}$
	5	Square pyramid	$[\text{CrOCl}_4]$
	6	Octahedron	$[\text{CrOCl}_5]^{2-}$
	8	Quasi-dodecahedron	$[\text{CrO}_8]^{3-}$
+ VI	4	Tetrahedron	$[\text{CrO}_4]^{2-}$

<sup>a</sup>Most frequent geometry.

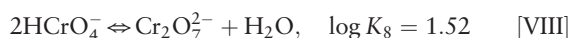
R, alkyl or aryl; Me, methyl; Pr, propyl; diars, *o*-phenylene-bis-dimethylarsine; tren, tris(2-aminoethyl)amine.



**Figure 1** Calculated distribution of inorganic Cr(III) species as a function of pH (solution in equilibrium with  $\text{Cr(OH)}_3$  precipitate).



**Figure 2** Calculated distribution of inorganic Cr(VI) species as a function of pH for a total concentration of  $10^{-6} \text{ mol l}^{-1}$ .



The distribution of these species with pH is shown in Figure 2.

Three pH regions may be distinguished for Cr(VI) species:

1. pH 0, where  $\text{H}_2\text{CrO}_4$  is a significant species;
2. pH 2–6, where  $\text{HCrO}_4^-$  and  $\text{Cr}_2\text{O}_7^{2-}$  occur together; and
3. pH > 9, where  $\text{CrO}_4^{2-}$  predominates.

In studies with various cell systems, starting with chromate, Cr(VI) has been shown to be present as an intermediate. It was found also in the presence of naturally occurring organic matter such as humus. The concentration of dichromate ( $\text{Cr}_2\text{O}_7^{2-}$ ) is negligible for total Cr(VI) concentrations below  $10^{-3} \text{ mol l}^{-1}$  over the whole pH range.

The interest in determining the concentration of the specific chemical forms of chromium rather than only its total concentration is the fact that the chemical forms have a profound effect on its mobility, bioavailability, and toxicity. Of the two oxidation states found in nature, trivalent and hexavalent, the former is relatively benign and the latter toxic.

Most organochromium compounds are not very stable and undergo hydrolysis and/or protolysis in the presence of water or even decompose in the presence of air. Because of their instability these species do not accumulate in the environment; however, they may play a role in some transport phenomena and metabolic pathways as transient species. The determination of such species in environmental or biological samples is hampered by their extremely low concentrations and the problem of their

differentiation from chromium physically sorbed onto organic species.

## Areas of Special Interest

In view of the different behavior of chromium species with respect to mobility, bioavailability, and toxicity, and in order to follow the pathways for interconversion both in the environment and in biological systems, it is increasingly important to monitor the concentration of the individual chemical species as well as the total concentration of chromium.

## Health Considerations and Occupational Monitoring

Cr(III) is held to be essential for human and animal nutrition, necessary for the maintenance of glucose, lipid, and protein metabolism, and is therefore used as a dietary supplement mostly in the form of its picolinate or nicotinate. Trivalent chromium is poorly absorbed principally because, unless heavily complexed, it will precipitate under most physiological pH conditions. Whilst Cr(III) compounds appear to be able to produce genetic effects with purified nucleic acids or cell nuclei, there is generally no such activity in intact cellular systems due to the relatively poor ability of Cr(III) to cross cell membranes.

In contrast, hexavalent chromium species, predominantly chromate at physiological pH, behave very differently and are toxic for bacteria, plants, and animals. Primary effects are related to the oxidative nature of Cr(VI), which manifests itself in irritation of the skin and the mucous membranes and allergic effects on lung and bronchia. Cr(VI) that survives reduction by body fluids is rapidly taken up by the

erythrocytes and lymphocytes penetrating the cell membrane via a general nonselective anion transport channel. In the cell nucleus, chromate can oxidatively damage genetically important components leading to DNA damage, reverse mutation, forward mutation, sister chromatid exchange, chromosomal aberrations, cell transformation, and alterations in mitotic cell cycle. Evidence is gathered that the ability of body fluids and long-lived nontarget cells to reduce Cr(VI) greatly reduces its potential toxicity.

The working group of the International Agency for Research on Cancer (IARC, France) concluded that there is sufficient evidence in humans for the carcinogenicity of Cr(VI) compounds as encountered in the chromate production, chromium pigment production, and chromium plating industries, and that there is inadequate evidence in humans for the carcinogenicity of metallic chromium and of Cr(III) compounds. Therefore, the major concern is in reducing and controlling the exposure to Cr(VI) for workers in industries handling chromium compounds as well as those being exposed by secondary products, e.g., welding fumes. However, there is no scope for chromium speciation in biological monitoring as all Cr(VI) in the body is converted to Cr(III) before it is excreted.

### Biological Monitoring

The effects of chromium on glucose metabolism have been studied extensively in both animals and man since the initial suggestion that a chromium compound in brewer's yeast influences glucose control. However, most of the findings were based on concentrations greatly influenced by contamination or lack of selectivity (e.g., atomic absorption spectrometry (AAS) with continuum source background correction; results obtained with such techniques should be questioned). Interpretation of the data on the effects of chromium on lipid and glucose metabolism is therefore difficult and the organic compound assumed to be the biologically active form of chromium has not yet been characterized and consequently is not available for supplementation studies. In addition, there is increasing evidence for the suggestion that the existence of the link between chromium and the glucose tolerance factor (Cr-GTF) is doubtful, and that earlier findings indicating the existence of the Cr-GTF complex have been heavily influenced by contamination.

The concentration of chromium in serum of unexposed healthy individuals is  $\sim 0.2 \mu\text{g l}^{-1}$ , i.e., 5000 times lower than that of zinc and copper. The interest for speciation of chromium in serum was initiated by the finding of extremely high chromium

levels (25–50 times that of healthy individuals) in the serum of dialysis patients as compared to healthy persons due to irreversible take-up from impurities of the dialysate. In order to follow biochemical processes involving the element, it is necessary to know the chromium-containing species. Because the time that is needed for classical ion-exchange procedures is so long, such that denaturation of the proteins occurs, it is necessary to perform these separations on a fast protein liquid chromatography system. The tedious workload to determine the very low concentrations in the various fractions and the immense hazards of contaminating with exogenous chromium can be partly overcome by the use of *in vitro* and *in vivo* labeled  $^{51}\text{Cr}$  plasma reducing the detection of the element to a simple radioactivity measurement. With the help of such methods chromium was found to be mainly bound to transferrin (85%) and to a lesser extent to albumin (8%) and other components (6%).

### Environmental Monitoring

Chromium enters the environment as a result of effluent discharge from steel works, electroplating, tanning industries, timber treatment, oxidative dyeing, chemical industries, and cooling water towers. The metal may also enter drinking water supply systems from the corrosion inhibitors used in water pipes and containers or by contamination of the underground water from sanitary land fill leaching. In view of the differences between the chromium species, chromium speciation rather than chromium total concentration has to be considered in pollution control and environmental risk assessment studies. Special attention has to be paid to industrial effluents, waste disposal, and sewage sludge deposition.

Tannery wastes usually contain high chromium concentrations (1–5%). Historical references to the antiquated, and virtually obsolete, two-bath chrome-tanning process, which uses Cr(VI), and the possibility of chromate pigments in finishing, have frequently resulted in confusion and exaggerated fears of tannery waste being hazardous from Cr(VI) contamination. Cr(VI) is readily reduced to Cr(III) in low-pH, organic-matter environments. Furthermore, Cr(VI) is not a tanning agent or otherwise useful material in modern tanning processes. On the other hand, due to the progressive exhaustion of landfills, land disposal of such sludge is increasingly attractive for economic as well as logistic reasons. However, chromium could affect metabolism and/or accumulate in living organisms. Furthermore, unavoidable oxidation of Cr(III) to more soluble and toxic species could cause groundwater pollution. Analytical interest in this field is devoted both to the control of such

waste disposals and to the development of less polluting processes, including recycling of the chromium.

### Aquatic and Marine Environment

Chromium is present in natural water systems in two thermodynamically stable oxidation states, Cr(III) and Cr(VI), depending on the redox conditions. Cr(VI) is often mistakenly referred to as a powerful oxidant in seawater, when, in fact, at pH 8.1 it has a redox potential lower than that of seawater. While thermodynamic calculations indicated that in fresh water chromium should exist almost exclusively as Cr(VI), experimental results differ significantly from prediction. Although the chemistry of chromium is understood in principle, it is strongly influenced by reactions that are kinetically slow, such as the oxidation of Cr(III) in seawater by oxygen. The Cr(III) ion is readily hydrolyzed and binds strongly to particles and organic material. Because Cr(III) is rapidly scavenged by particles, it is mainly in particulate form.

Adsorption on manganese oxide, followed by oxidation at the surface, helps considerably to convert Cr(III) into the thermodynamically stable Cr(VI). However, because of the low concentration of suspended  $\text{MnO}_2$  in the oceans, it is not clear whether this catalyzed oxidation is quantitatively more important than the direct oxidation by dissolved oxygen. Thus, chemical speciation of chromium in seawater is still an important issue in marine chemistry.

The concentrations of dissolved trace metals in interstitial waters of marine sediments often exceed predicted concentrations based on the solubility products of the most likely mineral phase. Complexes with organic matter having the general characteristics of humic substances with intermediate polarity may explain the observed excess solubility. Between 23% and 55% of the total, Cr in the interstitial water was found to be bound to organic material. It is assumed that the driving force behind this complex formation is the large crystal field stabilization energy associated with the formation of organic coordination complexes. These complexes may be an important factor in increasing the solubility of chromium and thus allowing its transport to the overlying water column. The trivalent chromium-organic matter association is so stable that it can be found even in the oxidizing water column.

While complexation/dissociation of chromium and natural organics has been postulated very often to explain data otherwise not in agreement with thermodynamic model calculations, there is a lack of direct evidence of its significance. Advances in understanding the biogeochemistry of chromium,

therefore, depend on the accurate determination of chromium species and the kinetics of their inter-conversion.

In conclusion, it can be stated that the valence-state distribution of dissolved chromium in natural waters will depend on the following important factors:

1. the oxygen content and redox potential of the water;
2. the presence of dissolved or particulate organic matter; and
3. the presence of suspended inorganic matter.

Samples for speciation studies, therefore, should be characterized accordingly with additional measurements of pH,  $pE$ , suspended inorganic matter, organic material, and oxygen content.

### Soil

The biomethylation of metals and metalloids in soils and sediments appears to be a widespread phenomenon. Owing to their altered physical properties like volatility and lipid solubility, the methylated compounds are important in the mobilization, transport, and bioaccumulation of trace elements in the environment. The methylation of Cr(II) by methylcobalamine appears to proceed by a homolytic pathway involving the transfer of a methyl radical.  $\text{CH}_3\text{Cr}(\text{H}_2\text{O}_3)^{2+}$  is the product of this reaction, but is rapidly cleaved under acidic conditions to give methane and Cr(III).

Generally, interest in the speciation of chromium in soil is to gain insights into mechanisms responsible for the mobility and bioavailability of the element. Therefore, the parameters under investigation are chromium solubility and its oxidation state, which are related to the bioavailability and toxicity. Because of its low mobility, landfill disposals of Cr(III) would not present a pollution problem, but in case oxidation of Cr(III) to Cr(VI) occurs it can be more serious. The migration of chromium is determined by the competition between complexation, dissolution/precipitation, redox processes, and adsorption/desorption mechanisms. Cr(III) will migrate under acidic conditions and/or if present as dissolved organic matter complex. Hexavalent chromium generally migrates more rapidly but its mobility is inhibited in presence of Fe(II) and high concentrations of organic matter and when sorption processes are favored at low pH. Numerous separation schemes are in use, based on the leaching of soil with different solvent and buffer systems. Generally, these 'procedural speciation techniques' lack in standardization and comparability, and interpretation of results is

not straightforward. In an effort to improve this situation, the European Community Bureau of Reference (BCR) has promoted the evaluation, improvement, and standardization of a sequential three-stage extraction scheme that today has been characterized for a variety of sample matrices.

## Methods for the Determination of Chromium Species

The variety of methods for the determination of chromium species may be classified into two fundamental categories: species-selective, direct measurements and nonselective detection combined with species-selective separation methods. In the case of chromium, two strategies can be distinguished for the later approach: simultaneous separation and determination of both species or determination of one species and the total chromium concentration, obtaining the result for the second species by calculation 'through difference'. Since such an indirect approach involves some risk due to cumulative errors, it is nowadays avoided whenever possible. Such and other problems can often be prevented by techniques combining a selective separation technique online with a sensitive detection technique, creating a very powerful 'hyphenated technique'.

### Special Considerations in Sample Handling

The sampling, sample pretreatment, and storage of samples are the most critical stages in any procedure for trace metal analysis but even more so for speciation analysis. These critical steps should be performed either directly by, or under the supervision of, qualified personnel. The requirements for the determination of chromium speciation depend strongly on the type of sample and the concentration range. The main concerns are analyte loss during sample preparation, contamination, and species transformation. Generally, attention should be paid to the selection of tools and containers for sampling and sample pretreatment. Tools such as knives, mixers, homogenators, or mills made from stainless steel, most often containing chromium, should be avoided and replaced by tools made from polymeric materials, quartz, and titanium. Colored polymeric materials, such as screw-caps for bottles, dispensers, and pipette tips are suspect and should be controlled for chromium contamination.

Most stringent requirements on sampling, sample pretreatment, and sample storage have to be fulfilled in the analysis of biological materials, because physiological concentrations are in the low microgram per liter or submicrogram per liter level.

Disposable steel needles used widely in medical practice for drawing of blood samples cannot be used for sampling blood intended for trace metal analysis. Contamination of blood samples by such needles can exceed the actual level of chromium in the sample by a factor of 100–1000. In case liquid chromatography (LC) is used for separation of species, the system inherent contamination has to be taken in account, when using high-performance liquid chromatography (HPLC) equipment (pumps, valves, columns, capillaries) made from stainless steel.

The problems of contamination obviously extend to any sample pretreatment procedures, which, therefore, should be reduced to the absolute minimum for ultratrace determinations. If sample pretreatment is necessary or considered appropriate, great care must be taken to ensure that chromium from the reagents or containers used, or indeed from the environment, does not contaminate the sample and invalidate the results.

In the field of water analysis the interest is in differentiation between soluble and particulate chromium as well as between the two main oxidation states. The system Cr(III)/Cr(VI) is subject to redox reactions, especially in the presence of oxygen and/or organic materials. As soon as possible after its collection, water samples should be filtered if metal speciation is to be studied in order to avoid any remobilization of metals bound or sorbed to particulate matter. While pressure filtration offers advantages in terms of speed of filtration it may contribute to enhanced soluble organic matter by rupture of phytoplankton cells.

Collection of particulate matter from atmospheric air or aerosols being present in the workplace environment for subsequent species determination needs special care. Variable amounts of Cr(VI) can be reduced by filter materials like paper or cellulose or in the presence of materials co-collected on the filter. Poly(vinyl chloride) filters should be used and samples should be either stabilized by preconditioning of the filters or immediately after collection.

### Storage of Samples

For speciation studies it is very important that the analyses be performed as soon as possible after collection. If samples have to be stored, the risk of Cr(VI) reduction, especially in the presence of organic material under acidic conditions, has to be considered. Cr(VI) losses in the range of 15–20% have been observed during the first day after sampling. In contrast, Cr(III) is remarkably susceptible to oxidation in alkaline medium. Only in neutral solutions are the different chromium species comparatively resistant to redox reactions, partly because



of slow kinetics and also the oxidation potential of the Cr(III)/Cr(VI) couple in such solutions. The medium oxidation potential under neutral conditions relatively stabilizes the Cr(III)/Cr(VI) couple both against oxidation and reduction. Consequently, samples should be stored without acidification. Unfortunately, Cr(III) shows the highest instability at neutral pH with respect to losses by sorption on container walls, particularly in polyethylene containers, where losses up to 25% have been observed after 15 days of contact. If sample storage cannot be avoided, then the best stability of the species distribution is obtained in a buffer solution containing  $50 \text{ mmol l}^{-1}$   $\text{HCO}_3^-/\text{H}_2\text{CO}_3$  in polytetrafluoroethylene bottles if kept at  $5^\circ\text{C}$  under a  $\text{CO}_2$  blanket.

In order to reduce the possibilities for species transformation between sampling and analysis some researchers have proposed to perform the separation onsite by using short chromatographic columns operated under low-pressure conditions (flow injection, FI) as collection devices in the field. Rather than transporting the water samples back to the laboratory, the loaded columns will be used instead. It was demonstrated that the column variability is in the range of single-column replicates, allowing for the establishment of convenient, versatile, and easy to use methodology, with built-in preconcentration ready for automation.

### Species-Selective Direct Techniques

For the direct speciation analysis of chromium traces only a few techniques are available.

**UV-visible spectrophotometry** UV-visible spectrophotometry is a well-established technique for the selective determination of Cr(VI) with good detection power. The standard method for the selective determination of Cr(VI) is based on the formation of a red-violet colored complex with 1,5-diphenylcarbazide under acidic conditions, which can be detected spectrophotometrically at 540 nm. In order to achieve good reproducibility, several conditions such as temperature or amount of reagent and acids must be kept strictly constant. The molar absorption coefficient is from  $3.0 \times 10^4$  to  $8.0 \times 10^4$ , fairly high, allowing detection limits of  $\sim 5 \mu\text{g l}^{-1}$  under optimized conditions. The complex is very stable (less than 2% signal reduction over 90 min) and only a few ions like Mo(VI), Cu(II), Mn(II), Fe(III), and Hg interfere, but only at high concentrations. Some of these interfering ions can be masked either with phosphate or ethylenediamine-tetraacetic acid (EDTA) or kinetically differentiated. The main problem in its application is the acidic

conditions needed for complex formation with the potential risk of reducing Cr(VI) especially in the presence of organic materials, normally present in samples like natural or waste water. In general, spectrophotometric determinations in sample solutions that are originally colored is problematical.

In order to determine total chromium, Cr(III) has to be oxidized either during sample pretreatment or, more elegantly, online with Ce(IV) using a FI manifold. The methods of oxidizing chromium to the required state and of destroying the excess of oxidizing agent are of critical importance for the overall performance of the method. In the conventional analytical procedure an excess of permanganate has to be decomposed by reduction with azide or by precipitation as hydrous  $\text{MnO}_2$ , while an excess of peroxydisulfate can be decomposed by boiling the solution or by reduction with azide. Ce(IV) can be used for online oxidation; however, the blank introduced with this reagent seriously degrades the detection limit of this approach and the online method is less tolerant against interferents such as Fe(III), Mo(VI), Mn(II), or Cu(II). Other spectrophotometric methods based on complex formation with Cr(III) are less selective and sensitive and therefore are of not much importance.

**Chemiluminescence** Cr(III) can be selectively determined by chemiluminescence. The detection is based on the oxidation of luminol by  $\text{H}_2\text{O}_2$  catalyzed by the presence of small concentrations of Cr(III). The reaction is highly selective for Cr(III), Cr(VI) does not catalyze the luminol reaction. Interferences from other metal ions can be eliminated by the addition of a chelating compound, such as EDTA, which forms complexes with the interferents, thereby inactivating them as catalysts. Because of the kinetic inertness of Cr(III), chromium complexes do not form at room temperature during the analysis time. The whole procedure can be easily automated by the use of FI systems, allowing high sample throughput for routine analysis. The high detection power of this method allows the determination in environmental samples such as surface waters or seawater or even biological samples.

**Electrochemical methods (polarography and voltammetry)** The electrochemical behavior of Cr(III) and Cr(VI) is significantly different, allowing the species-selective determination by polarographic and voltammetric techniques. Cr(VI) is electrochemically active over the entire pH range; hence, a medium pH can be selected for the measurements, thus effectively protecting samples from undergoing redox reactions during the analytical procedure. Direct

differential pulse polarography provides detection limits of  $\sim 30 \mu\text{g l}^{-1}$ , this detection limit can be substantially improved by electrochemical preconcentration. Cathodic stripping voltammetry preceded by adsorptive collection of a complex of Cr(III) with, e.g., diethylenetriamine pentaacetic acid or Cr(VI) with diphenylcarbazide on a hanging Hg drop electrode provides the necessary sensitivity for natural water samples including seawater. Other techniques like amperometry or potentiometry are mainly applied to automated effluent monitoring and process control.

**High-resolution X-ray spectroscopy** High-resolution X-ray spectroscopy can be used for the direct speciation of chromium in solid samples, such as air particulates, welding dust, and soil. These techniques use the effect that the chemical environment of the target analyte chromium is producing a chemical shift that is different for Cr(III) and Cr(VI) and can be observed in both emission and absorption X-ray spectra. Applications have been described for using the shift on the Cr  $K_{\alpha 1}$  and  $K_{\alpha 2}$  emission lines (X-ray fluorescence, XRF), the shift on the L-transitions that can be excited by low-energy electrons ('soft X-ray') or the shift in the profile of the near-edge absorption spectra. Using such techniques in combination with chemometric data evaluation, the distribution of chromium species in solid samples can be measured quantitatively in cases where the total chromium concentration is sufficiently high.

### Speciation Techniques Based on Separation of Chromium Species

Speciation techniques based on the separation of different chromium species before detection have the advantage of making possible the combination of a highly selective technique with a highly sensitive detection technique. While often the detection power of the combined technique can even be increased by preconcentration, a serious drawback is the complex and time-consuming sample pretreatment, based on solvent extraction, co-precipitation, electrochemical separation, ion exchange, solid-phase extraction, or selective volatilization. It is also apparent that any separation-detection method must not alter the initial species profile in the sample undergoing analysis; that is, the results of the analytical method must accurately and correctly represent what was present in the initial sample before it was touched. Therefore, in general, more direct methods are in favor of more complex methods, where multiple steps are a potential risk with respect to contamination and alteration

of the original species distribution. If multiple steps cannot be avoided, their integration into a single combined 'hyphenated technique' has the advantage of improved reproducibility through the exact control of parameters and the speed of analysis, and the use of a 'closed system' reducing the risk of contamination.

The determination of chromium can be performed by a great number of physicochemical methods, with different detection powers, working range, and application field. Here, only those techniques that have found more general applications in different fields can be briefly discussed. The selection of a method for a special analytical task is not only dependent on the detection power and other performance characteristics of the technique, but is highly dependent on the sample constitution, the required sample pretreatment and means of sample introduction, and, last but not least, on its availability (Table 3).

Most of the mentioned analytical techniques can be coupled with separation procedures, either to enhance their detection limits by preconcentration, or to enhance their selectivity by separation from interferences or to allow for species selective determination. Especially for the last case, online coupling of separation techniques such as gas chromatography (GC), HPLC or FI techniques, with some of the detection techniques named in the above table are very attractive.

**Atomic absorption spectrometry** AAS is historically the most widely used technique for trace element determination in general and for the analysis of chromium in particular.

**Flame AAS (FAAS)** Chromium can be determined by FAAS using both air-acetylene flame and a nitrous oxide-acetylene flame, giving no flame background problems in the wavelength range used for determination. A large number of resonance lines of similar sensitivity are available, but for the majority of work the 357.9 nm line is used. The sensitivity and dynamic range of chromium determinations by FAAS are critically dependent on flame conditions and the observation zone. Using the air-acetylene flame, the reductive power of a fuel-rich flame is essential for effective atomization of chromium salts, giving a characteristic concentration of  $\sim 40\text{--}80 \mu\text{g l}^{-1}$ , depending on the aerosol transport efficiency of the nebulizer/burner system and the flame conditions, allowing precise determinations in the concentration range from 0.5 to  $\sim 10 \text{ mg l}^{-1}$ . Sensitivity for chromium using the air-acetylene flame is different for different chromium compounds with Cr(III) giving higher sensitivity than Cr(VI), the difference being



**Table 3** Various analytical techniques and their characteristics for the determination of chromium

Analytical method	Detection limit <sup>a</sup> ( $\mu\text{g l}^{-1}$ , 3 $\sigma$ )	Multi-element capability	Micromethod	Hyphenation	Sample types for direct sample introduction				Degree of interferences		
					Solution	Suspension	Solids		Mechanical, physical	Chemical	Spectral detection
NAA	5000	+	-	-	+	+	+		+	-	+
XRF	500	+	+	-	+	+	+		+	+	-
Flame-AAS	20-200	±	±	+	+	+	-		+	+	-
Polarography, DPP	10-50	+	-	+	+	+	-		+	+	-
Photometry	2-20	-	-	+	+	+	-		+	+	+
ICP-AES	1-5	+	-	+	+	+	LA, ETV, Arc		+	-	+
TXRF	0.2-1.0	+	+	-	+	+	-		+	+	+
Chemiluminescence	0.03-0.3	-	-	+	+	+	-		+	+	-
ET-AAS	0.05-0.15	±	+	+	+	+	+		+	+	+
ICP-MS	0.02-0.05	+	-	+	+	+	LA, ETV, Arc		+	-	+

<sup>a</sup> All detection limits are mean instrumental detection limits, which can be seriously degraded in the presence of a complex matrix.

most pronounced in fuel-rich air-acetylene flames and also at the observation height, giving the best sensitivity.

Iron and nickel cause severe depression of the atomic absorption signal from the air-acetylene flame, which can be attributed to the formation of refractory oxides. Several releasing agents such as ammonium chloride or fluoride, hydroxylamine hydrochloride, sodium sulfate, and others have been used to eliminate these interferences. Removal of interferences occurs in most instances through the introduction of a new dominating influence on the flame reactions controlling the atom formation process despite the presence of the original interferent. Many of these interferences decrease in hotter flames and in hotter regions of flames, and can be suppressed or even eliminated by the use of the nitrous oxide-acetylene flame but only at the expense of reduced sensitivity (about one-third) in comparison to the air-acetylene flame.

**Electrothermal atomization AAS (ETAAS)** ETAAS using the graphite furnace has been clearly the most common technique for the determination of chromium in the low trace and ultratrace range especially encountered with environmental and biological materials. Despite its broad application, the determination of ultratrace levels of chromium by ETAAS is not without problems. The accuracy of such determinations is determined, amongst other factors, by the precise correction of nonspecific absorption of light by concomitant compounds. Owing to the low intensity of the deuterium lamp at 357.9 nm, background absorption and emission produced by matrix components cannot be efficiently corrected when using this type of background correction. This situation improves when using Zeeman-effect background correction systems that is able to compensate for high absorbances, such as those produced by biological matrices, and is therefore the technique to be used in this field.

Standard procedures to overcome matrix interferences such as the use of platform atomization together with the use of matrix modifiers are only in part applicable to the ultratrace determination of chromium in biological samples. Matrix modifiers normally used such as palladium and magnesium are often too highly contaminated, and therefore cannot be used without degrading the detection limit, which is  $\sim 0.1 \mu\text{g l}^{-1}$ . Chromium interacts with carbon, forming carbides, a reaction that is responsible for chromium retention in the graphite tube, giving rise to tailing peaks and carry over. In general, high heating rates and atomization from a pyrolytic graphite coated surface will give the best performance. In

the case of transversely heated graphite tubes, such conditions can also be attained by using platform atomization.

The maximum charring temperature that can be used for chromium in aqueous solution is  $\sim 1200^{\circ}\text{C}$ . At this temperature, most if not the whole organic matrix is removed but a large proportion of the major inorganic salts present in biological samples remain even after prolonged charring times, giving rise to a substantial background signal during the atomization stage. Addition of magnesium nitrate as a matrix modifier extends the possible charring temperature to  $1600^{\circ}\text{C}$ , ensuring more complete removal of inorganic salts.

By using chelating compounds such as trifluoroacetylacetone as a modifier in combination with the graphite tube as a programmable thermal reactor, species-specific determination can be realized by ETAAS. The method is based on the formation of a volatile Cr(III) complex that is lost from the heated graphite tube prior to atomization, thus only detecting Cr(VI).

**Atomic emission spectrometry** The atomic emission spectrometry (AES) determination of chromium using the inductively coupled plasma (ICP) as the excitation source is a well-established method. ICP-AES in its conventional form, using a pneumatic nebulizer for sample introduction, has become the working horse for analyzing solutions especially in environmental analytical laboratories. Sample pretreatment is generally the same as for AAS, calling for leaching, dissolution/digestion as the main procedure for most solid samples while liquid samples, such as water or beverages, can be introduced directly in many instances.

Being a multielement technique, ICP-AES exhibits detection limits for chromium in the range of  $1\text{--}5\text{ }\mu\text{g l}^{-1}$ . Although chemical interferences are generally uncommon in ICP-AES, there are some restrictions in the total salt concentrations, which can be introduced into the plasma using conventional nebulization techniques, calling for further dilution. Moreover, detection limits can be seriously degraded by spectral interferences, much more common in ICP-AES than in AAS. Apart from metallurgical applications, sensitivity is satisfactory for the analysis of most environmental samples (dust, soil, sediments) and some biological materials (plants, foods), with the exception of more demanding samples like body fluids or tissues. Other sources for AES such as flame, arc, spark, glow-discharge source, direct-current plasma, microwave-induced plasma have been applied for the determination of chromium in special cases.

**Mass spectrometry** Mass spectrometry (MS) using the ICP as the ion source shares most of the advantages of ICP-AES, namely a stable ICP source with the additional feature of 10–100 times improved detection limits. For mass spectrometric detection of chromium most often the main isotope  $^{52}\text{Cr}$  is used, since  $^{53}\text{Cr}$  and  $^{54}\text{Cr}$  are overlapped by iron isotopes and  $^{50}\text{Cr}$  by a titanium isotope. Chromium isotopes are in an advantageous mass range in terms of sensitivity, allowing detection limits well below  $0.1\text{ }\mu\text{g l}^{-1}$  under optimized conditions, thereby closely reaching or even surpassing the detection power of GFAAS. The high detection power together with the multielement capability of ICP-MS makes this technique very suitable for environmental and biological samples. However, determinations below a few micrograms per liter are hampered by isobaric interference of the  $^{52}\text{Cr}$  by polyatomic ions between argon (the plasma gas) and nitrogen, oxygen and carbon, especially with samples rich in organic compounds or using organic solvents. These interferences can be overcome by either high-resolution ICP-MS using sector-field technology or by quadrupole ICP-MS in combination with chemical resolution provided by reaction/collision cell technology.

An especially attractive feature of ICP-MS is the possibility of using isotope dilution as a calibration strategy, enhancing the accuracy of measurements significantly. In its special form of speciated isotope dilution (SIDMS) this method can even correct for some species transformation during analysis. For this purpose, SIDMS is using the concept of spiking the sample with known amounts of enriched isotopes that have been chemically converted into the same forms of the species to be analyzed. The isotopic spike for each species has a unique isotopic enrichment. In the case of chromium two isotopic spikes, Cr(VI) enriched in  $^{53}\text{Cr}$  and Cr(III) enriched in  $^{50}\text{Cr}$ , are prepared for the simultaneous determination of Cr(III) and Cr(VI). Environmental samples containing Cr species are spiked with both  $^{53}\text{Cr(VI)}$  and  $^{50}\text{Cr(III)}$ . The species are separated at the end of the manipulation by means of chromatography and different isotope ratios are measured. In contrast to 'classical' chromatography, which provides a snapshot recording the state of affairs at the end of the manipulation, any interconversions that occur after spiking are traceable by SIDMS and can be quantitatively corrected by monitoring isotopes in each species. However, as with any isotope dilution method, accurate determinations can only be obtained when the spike behaves in every aspect like the species originally present and is 'equilibrated' with it before the determination step. This prerequisite is not easy to fulfill in some real samples, where the

freshly added Cr(III) spike behaves differently from the present Cr(III), which in a precipitated form is subjected to aging effects.

Online coupling of ICP-MS with LC is relatively straightforward. The separation power required to separate the two oxidation states of chromium is not very high, so that even some short columns operated under low- or medium-pressure conditions may be useful. Different chromatographic techniques such as suppressed ion chromatography, chelation ion chromatography, mixed anion/cation exchange chromatography, or reversed-phase chromatography of dithiocarbamates have been used for this purpose with success.

**X-ray fluorescence** XRF is one of the longest established techniques for trace elemental analysis. While XRF is not a very sensitive technique, its main advantages are the capability for direct solid sample analysis combined with multielement determinations. While sample pretreatment of solids can be substantially reduced or even omitted in some cases, perfect matching between standards and samples is required for accurate results, because of severe matrix effects. The main application field of XRF is, therefore, the analysis of solid materials, such as metallurgical and geological samples, where solid standards are readily available. Liquid samples can be analyzed either directly in special cells or by using preconcentration techniques with solid sorbents, which can be directly analyzed after sample loading. More modern methods, like total-reflection X-ray fluorescence, which is a multielement technique mainly for solutions, or particle-induced X-ray emission, which is a micromethod with some spatial resolution, have found limited application in some special areas. For speciation purposes, species separation has to be carried out in front in an offline mode.

### Radiochemical methods

**Neutron activation analysis** Thermal neutron activation of Cr leads to two radionuclides of which the short-lived radionuclide  $^{55}\text{Cr}$  has a low abundance compared to the parent isotope (2.36%), a relatively small cross-section for formation (0.36 barn), and, most significantly, it is almost a pure  $\beta$ -emitter and emits very few  $\gamma$ -rays (0.043%). The lower limit of detection using this radionuclide is very high and measurement of chromium by neutron activation analysis (NAA) is carried out using the long-lived radionuclide  $^{51}\text{Cr}$ . The half-life of  $^{51}\text{Cr}$  is within a range of 27.8 days, allowing sufficient time for sample manipulations.

In order to obtain high sensitivity the sample should be irradiated at high fluxes for long periods of

time (several days or at least 10–20 h). With these long irradiation times liquid samples will suffer considerable radiolysis, leading to the formation of large amounts of gases, which are likely to lead to explosion of the irradiation ampoule. Hence, a solid sample should be used and, rather than drying the liquid sample, it is preferable to preconcentrate the trace elements by co-precipitation. The best co-precipitants for NAA will be compounds that have small thermal neutron absorption cross-sections and which do not form  $\gamma$ -emitting radioisotopes on neutron absorption.

Sensitivity can further be increased by preconcentration following the radiation (and appropriate decay time) before detection with the advantage that contamination during these steps can be ignored. Whilst excellent results have been achieved with such procedures even for ultratrace determinations of chromium in biological material, the expensive instrumentation required precludes the use of this method for routine analysis, but it does play a vital role in cross-validation.

**Radioactive tracer techniques** Chromium-51 can be used as a radioactive tracer in isotope dilution analysis. After adding known amounts of the tracer to the sample, chromium is separated by a substoichiometric procedure. Under the assumption that the isotopic ratio for the spiked sample will not be changed by the separation procedure, the original chromium concentration can be calculated from the activity ratios measured. This assumption, however, is only valid if both the original chromium and the added chromium tracer have the same oxidation state and are homogeneously distributed. The whole sample preparation procedure has to be designed in order to fulfill these requirements.

### Separation Techniques

**Separation by extraction** Most of the methods of selective extraction are based on the extraction of Cr(VI) into organic solvents using ion pairs either with the anion of a mineral acid, or by some ion-pairing reagents such as trioctylphosphine oxide, trioctylamine, or Aliquat-336. Other methods are based on complex formation with chelating agents such as dithiocarbamates or liquid ion-exchangers such as Amberlite LA-2. The crucial point of this principle is the risk of co-extraction of Cr(III) complexes initially present in the sample. Another risk with this type of separation is the use of acidic media, which is often needed for the quantitative extraction of Cr(VI) with many reagents but favors reduction of Cr(VI) especially in the presence of organic material.

Preconcentration methods based on complex formation with Cr(III) have not been applied so widely, owing to the inert nature of hydrated Cr(III) species. Consequently, Cr(III) is normally obtained by difference after the determination of total chromium and Cr(VI). Especially in those cases, where Cr(III) is only a small part of the total chromium present in the sample, e.g., seawater, determination by difference may not be adequate, due to the large uncertainties introduced by such a calculation.

For the direct preconcentration and separation of Cr(III), complex formation has to be accelerated with most of the ligands available. The simplest method for the enhancement of the complexation rate is to heat the sample solution. Efforts have been made to decrease the reaction time and temperature required by the addition of large amounts of reagents that have catalytic effects on the ligand-exchange reactions. Reagents such as fluoride, benzoate, and dodecyl sulfate have been used for this purpose. Quinolin-8-ol was used as a complexation reagent, reacting more rapidly than others with Cr(III) and having the advantage of a good selectivity for Cr(III) and no reducing ability.

**Separation by co-precipitation** Co-precipitation as a method for species-selective separation offers a high preconcentration factor of  $10^2$ – $10^3$  by using only standard laboratory equipment. Working at pH levels between 6 and 9, the risk of influencing the redox equilibrium is comparatively small. The hydroxides of Fe(III), Al(III), Ti(III), or others have been used as carriers for the co-precipitation of Cr(III) in accordance with a suitable detection technique. Often, the high amount of carrier used must be removed prior to the determination of chromium to avoid serious interference problems. Partial adsorption of Cr(VI) onto the precipitate can reduce the selectivity of the approach.

**Separation by liquid chromatography, ion-exchange, sorption, and solid-sorbent extraction** The use of ion-exchange methods for chromium speciation is based on the assumption that Cr(III) definitely forms cations so that anions, collected on a column, are considered as Cr(VI) by definition. However, this premise does not hold for Cr(III) complexes, which may be present in an anionic form. Further, real samples, such as river water, may contain negatively charged colloids, including humic substances, kaolin, hydrated silica, and their mixtures with hydrated iron(III) oxide, which may sorb Cr(III) and be partly retained on the resins.

Problems inherent in manual sample manipulation have been overcome, at least, in part, by hyphenated

techniques such as LC or FI coupled to photometric, electrochemical, or spectrometric detection techniques. Reversed-phase  $C_{18}$  silica gel has been used for the chromatographic separation of chromium species, which often results in inadequate sensitivity for trace concentrations of chromium in real samples because of low sample loading. In order to achieve adequate detection limits for trace determinations some preconcentration, preferentially on-column, has to be performed. Preconcentration of Cr(III) is not as straightforward as for Cr(VI) because the inertness of Cr(III) hinders efficient complexation resulting in poor sensitivity, low sampling frequency, and incomplete recovery.

A number of advantages, in addition to automated sample handling, sample manipulation in a closed system, and miniaturization, emerge from the application of flow techniques such as FI or sequential injection (SI) to classical chemical procedures. A salient feature of these techniques is the well-known reproducibility of the processes involved, which allows for very short analysis times since it is not required to wait until equilibrium has been established. This last factor is particularly important because prolonged sample manipulation may affect the chromium species distribution significantly. When flow techniques are used for separation, the objective is to separate a single analyte or group of analytes from interfering sample components or matrices, often simultaneously achieving some degree of preconcentration and therefore gaining sensitivity at the expense of separation power. The FI/SI separation process is quite similar in this respect to batch filtration or solvent extraction procedures, and no chromatographic processes are involved, despite the fact that some chromatographic equipment like HPLC pumps and online columns might be used for this purpose. Different solid sorbents have been used for FI/SI preconcentration, such as chelating ion-exchangers, e.g., polydithiocarbamate, polyacrylamidoxime, polyhydroxamic acid, or iminodiacetate chelate resin, immobilized chelating reagents on solid supports such as quinoline-8-ol immobilized on controlled-pore glass, silica gel, or activated alumina. Activated alumina offers the principal possibility for preconcentration of both chromium species since it can function both as an anion exchanger and as a cation exchanger depending on the pH of the solution. Under acidic conditions alumina exhibits a high affinity for oxyanions whereas it strongly retains cations under basic conditions.

**Gas chromatography** Using trifluoroacetylacetone, Cr(III) can be converted into a volatile complex that can be separated from other similar metal complexes



by GC. Considerable sensitivity can be achieved by using the electron capture detector, microwave excitation, atomic absorption, or MS for its detection.

**Electrophoresis** Capillary electrophoresis (CE) has been successfully used to separate Cr(III) and Cr(VI) species in recent years. Detection was done by UV-photometry, chemiluminescence, or ICP-MS. The online coupling of CE and ICP-MS is often realized by using microflow nebulizers or by compensating the difference between the very low effluent flow rate from the CE capillary and the much higher uptake of conventional nebulizers with a make-up buffer flow. Flat bed electrophoresis can be used to separate high molecular weight compounds such as proteins. Element-selective detection can be achieved by scanning the chromatographic plates with laser ablation ICP-MS.

## Quality Assurance

Quality control and quality assurance have become a very important part of trace analysis, and analysts are requested not only to prove that their methods produce reliable results but they should also report the uncertainty coming with it. In general, two different approaches are available to discuss obtained results, either by comparing the results obtained by the method used against a reference method or by comparing the results for the analyzed samples against those for reference materials. Both approaches are extremely difficult to fulfill for speciation analysis, since both standard methods and certified reference materials (CRMs) are rare or even non-existing. Speciated isotope dilution analysis could in some cases play the role of a reference method, particularly if the already discussed prerequisite of a total equilibrium between spikes and originally present species can be guaranteed. Today, CRMs are only available for a very limited number of matrices, namely NIST 2108/2109 (aqueous solutions of pure Cr(III) and Cr(VI)), BCR 544 (both Cr(III) and Cr(VI) in lyophilized solution), BCR 545, which is certified for Cr(VI), and total leachable chromium in welding dust loaded on a filter, and DANREF Cement-1 and Cement-2, which are certified for their chromate content in cement.

*See also:* **Activation Analysis:** Neutron Activation. **Air Analysis:** Workplace Air. **Amperometry.** **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Principles and Instrumentation; Inductively Coupled Plasma. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Capillary Electrophoresis:** Overview. **Cement.** **Elemental**

**Speciation:** Overview; Waters, Sediments, and Soils. **Environmental Analysis.** **Extraction:** Solid-Phase Extraction. **Food and Nutritional Analysis:** Overview. **Geochemistry:** Soil, Minor Inorganic Components. **Ion Exchange:** Overview. **Isotope Dilution Analysis.** **Liquid Chromatography:** Overview. **Polarography:** Overview. **Sample Handling:** Sample Preservation. **Voltammetry:** Overview. **Water Analysis:** Industrial Effluents. **X-Ray Absorption and Diffraction:** X-Ray Absorption.

## Further Reading

- Aitio A and Tomatis L (1991) On the carcinogenicity of nickel and chromium and their compounds. In: Aitio A, Aro A, Järvisalo J, and Vainio H (eds.) *Trace Elements in Health and Disease*, pp. 159–168. Cambridge: Royal Society of Chemistry.
- Campanella L (1996) Problems of speciation of elements in natural waters: the case of chromium and selenium. In: Caroli S (ed.) *Element Speciation in Bioinorganic Chemistry*, pp. 419–443. New York: Wiley.
- Costa M (1997) Toxicity and carcinogenicity of Cr(VI) in animal models and humans. *CRC Critical Reviews in Toxicology* 27(5): 431–442.
- Cornelis R (1990) Speciation of trace elements in serum or plasma. In: Broekaert JAC, Güçer S, and Adams F (eds.) *Metal Speciation in the Environment*, NATO ASI Ser. G, vol. 23, pp. 169–194. Berlin: Springer.
- Darrie G (2001) The importance of chromium in occupational health. In: Ebdon L, Pitts L, Cornelis R, et al. (eds.) *Trace Element Speciation for Environment, Food and Health*, pp. 315–330. Cambridge: Royal Society of Chemistry.
- Dyg S, Cornelis R, Griepink B, and Verbeek P (1990) Stability study of Cr(III) and Cr(VI) in water for production of an aqueous chromium reference material. In: Broekaert JAC, Güçer S, and Adams F (eds.) *Metal Speciation in the Environment*, NATO ASI Ser. G, vol. 23, pp. 361–376. Berlin: Springer.
- Katz SA (1998) Chromium, environmental analytical chemistry. In: Meyers RA (ed.) *Encyclopedia of Environmental Analysis and Remediation*, pp. 1129–1141. New York: Wiley.
- Kimbrought DE, Cohen Y, Winer AM, Creelman L, and Mabuni C (1999) A critical assessment of chromium in the environment. *Critical Reviews in Environmental Science and Technology* 29(1): 1–46.
- Ma RL, Woods G, and McLeod CW (1999) Microcolumn field sampling and flow injection analysis: A strategy for enhanced trace analysis and element speciation. In: Sanz-Medel A (ed.) *Flow Analysis with Atomic Spectrometric Detectors*, pp. 439–458. Amsterdam: Elsevier.
- Marqués MJ, Salvador A, Morales-Rubio AE, and de la Guardia M (1998) Analytical methodologies for chromium speciation in solid matrices; a survey. *Fresenius' Journal of Analytical Chemistry* 362: 239–248.
- Marqués MJ, Salvador A, Morales-Rubio AE, and de la Guardia M (2000) Chromium speciation in liquid

- matrices: A survey of the literature. *Fresenius' Journal of Analytical Chemistry* 367: 601–613.
- Nriagu JO and Nieboer E (1988) *Chromium in the Natural and Human Environments*, vol. 20. New York: Wiley.
- Pavel J, Kliment J, Stoerk S, and Suter O (1987) Preservation of traces of chromium(VI) in water and waste water samples. *Fresenius' Zeitschrift für Analytische Chemie* 321: 587–591.
- Pedersen B, Thomsen E, and Stern RM (1987) Some problems in sampling, analysis and evaluation of welding fumes containing Cr(VI). *Annals of Occupational Hygiene* 31(3): 325–338.
- Petura JC, James BR, and Vitale RJ (1998) Chromium(VI) in soils. In: Meyers RA (ed.) *Encyclopedia of Environmental Analysis and Remediation*, pp. 1142–1158. New York: Wiley.
- Richard FC and Bourg ACM (1991) Aqueous geochemistry of chromium: A review. *Water Research* 25: 807–816.
- Vercoutere K and Cornelis R (1995) Chromium speciation in environmental and biological samples. In: Quevauviller Ph, Maier EA, and Griepink B (eds.) *Quality Assurance for Environmental Analysis*, pp. 195–213. Amsterdam: Elsevier.

## CLINICAL ANALYSIS

Contents

**Overview**

**Sample Handling**

**Electrolytes in Physiological Samples**

**Glucose**

**Sarcosine, Creatine, and Creatinine**

**Inborn Errors of Metabolism**

### Overview

**I D Watson**, University Hospital Aintree, Liverpool, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Overview of Aims and Scope – Philosophy of Clinical Analysis

Clinical laboratories have to be able to determine a wide range of analytes. These encompass a wide range: ions, small organic molecules, proteins, and lipids. Typically measurements are made on biological fluids. Most determinations are performed on plasma (or serum) the liquid fraction of blood, or urine; occasionally other fluids such as cerebrospinal fluid may be examined. While some of the measurements relate directly to pathology in the blood cells or plasma, many are markers of changes happening at the cellular level. Cell destruction releases intracellular contents, e.g., enzymes, tumors release products; cell dysfunction causes buildup or diminution of normal concentrations of molecules.

Analyses for some of the analytes are requested very frequently, often requiring a rapid turn-round

time (0.5–2 h). To meet this demand, high-capacity rapid-throughput analyzers have to be utilized. A different approach is the increasing availability of point of care testing, i.e., testing outside the laboratory.

Less frequent or more specialized assays are performed using one or other of the more usual armamentarium of analytical biochemistry such as chromatography and enzyme-based methods. Clinical laboratory testing is intended to enable a qualified individual to diagnose and monitor disease. A great number of different analytes are measured with this aim in mind. Some of the commonly determined analytes are relatively specific for a disease or are particularly useful for monitoring disease progression.

The clinical utility of a test is described by its sensitivity – the ability to detect the disease with no false negatives – and its specificity – the ability to avoid false positives in the nondiseased population. The ideal test detects all those with disease and does not include any of those without the disease. No test meets these criteria. Combinations of tests may yield better classification, termed efficiency, but rarely achieve 100% efficiency. Complicating this perspective is the fact that patients present at different stages of disease, also some clinical analytes may change in

- matrices: A survey of the literature. *Fresenius' Journal of Analytical Chemistry* 367: 601–613.
- Nriagu JO and Nieboer E (1988) *Chromium in the Natural and Human Environments*, vol. 20. New York: Wiley.
- Pavel J, Kliment J, Stoerk S, and Suter O (1987) Preservation of traces of chromium(VI) in water and waste water samples. *Fresenius' Zeitschrift für Analytische Chemie* 321: 587–591.
- Pedersen B, Thomsen E, and Stern RM (1987) Some problems in sampling, analysis and evaluation of welding fumes containing Cr(VI). *Annals of Occupational Hygiene* 31(3): 325–338.
- Petura JC, James BR, and Vitale RJ (1998) Chromium(VI) in soils. In: Meyers RA (ed.) *Encyclopedia of Environmental Analysis and Remediation*, pp. 1142–1158. New York: Wiley.
- Richard FC and Bourg ACM (1991) Aqueous geochemistry of chromium: A review. *Water Research* 25: 807–816.
- Vercoutere K and Cornelis R (1995) Chromium speciation in environmental and biological samples. In: Quevauviller Ph, Maier EA, and Griepink B (eds.) *Quality Assurance for Environmental Analysis*, pp. 195–213. Amsterdam: Elsevier.

## CLINICAL ANALYSIS

Contents

**Overview**

**Sample Handling**

**Electrolytes in Physiological Samples**

**Glucose**

**Sarcosine, Creatine, and Creatinine**

**Inborn Errors of Metabolism**

### Overview

**I D Watson**, University Hospital Aintree, Liverpool, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Overview of Aims and Scope – Philosophy of Clinical Analysis

Clinical laboratories have to be able to determine a wide range of analytes. These encompass a wide range: ions, small organic molecules, proteins, and lipids. Typically measurements are made on biological fluids. Most determinations are performed on plasma (or serum) the liquid fraction of blood, or urine; occasionally other fluids such as cerebrospinal fluid may be examined. While some of the measurements relate directly to pathology in the blood cells or plasma, many are markers of changes happening at the cellular level. Cell destruction releases intracellular contents, e.g., enzymes, tumors release products; cell dysfunction causes buildup or diminution of normal concentrations of molecules.

Analyses for some of the analytes are requested very frequently, often requiring a rapid turn-round

time (0.5–2 h). To meet this demand, high-capacity rapid-throughput analyzers have to be utilized. A different approach is the increasing availability of point of care testing, i.e., testing outside the laboratory.

Less frequent or more specialized assays are performed using one or other of the more usual armamentarium of analytical biochemistry such as chromatography and enzyme-based methods. Clinical laboratory testing is intended to enable a qualified individual to diagnose and monitor disease. A great number of different analytes are measured with this aim in mind. Some of the commonly determined analytes are relatively specific for a disease or are particularly useful for monitoring disease progression.

The clinical utility of a test is described by its sensitivity – the ability to detect the disease with no false negatives – and its specificity – the ability to avoid false positives in the nondiseased population. The ideal test detects all those with disease and does not include any of those without the disease. No test meets these criteria. Combinations of tests may yield better classification, termed efficiency, but rarely achieve 100% efficiency. Complicating this perspective is the fact that patients present at different stages of disease, also some clinical analytes may change in



**Table 1** Some common clinical analytes

Group name	Analytes (optional)	Matrix	Units	Useful in emergency	Indicative of
Urea and electrolytes	Urea	Serum or plasma	$\text{mmol l}^{-1}$	Yes	Kidney function
	(Creatinine)		$\mu\text{mol l}^{-1}$		Kidney function
	Sodium		$\text{mmol l}^{-1}$		Fluid and electrolyte balance
	Potassium		$\text{mmol l}^{-1}$		Potassium homeostasis
Liver function tests	(hydrogen carbonate)	Serum or plasma	$\text{mmol l}^{-1}$	No	Acid-base status
	(chloride)		$\text{mmol l}^{-1}$		Acid-base status
	Bilirubin		$\text{mmol l}^{-1}$		Liver obstruction
	Alanine transaminase		$\text{U l}^{-1}$		Liver damage
	Aspartate transaminase		$\text{U l}^{-1}$		
	Alkaline phosphatase ( $\gamma$ -glutamyl transpeptidase)		$\text{U l}^{-1}$ $\text{U l}^{-1}$		Liver obstruction
Bone chemistry	Calcium	Serum or plasma	$\text{mmol l}^{-1}$	Yes	Bone metabolism
	Albumin		$\text{g l}^{-1}$		Correct for $\text{Ca}^{2+}$ protein binding
	Phosphate		$\text{Mmol l}^{-1}$	No	Bone metabolism
	Alkaline phosphatase		$\text{U l}^{-1}$		Bone building cell activity
Cardiac markers	Troponin T (or I)	Serum or plasma	$\text{U l}^{-1}$	Possibly no	Indicators of heart attack after 12 h
	Creatine kinase				
Thyroid function test	Thyroxine	Serum or plasma	$\text{nmol l}^{-1}$	No	Thyroid function
	Triodo thyronine		$\text{nmol l}^{-1}$		
	Thyroid-stimulating hormone		$\text{U l}^{-1}$		
	Urate	Serum or plasma	$\text{mmol l}^{-1}$	No	Indicator of gout
	Creatinine clearance	Urine plus serum	$\text{ml min}^{-1}$	No	Indicator of kidney function
	Sodium	Urine	$\text{mmol}$	Yes	Indicator of circulating blood volume status
	Amino acids	Urine	$\text{mmol l}^{-1}$ and $\mu\text{mol l}^{-1}$	No	Inborn errors of metabolism
	Porphyrin	Serum, urine, or feces	$\mu\text{mol l}^{-1}$ or $\mu\text{mol day}^{-1}$	No	Inborn or acquired errors of metabolism

different diseases. This may be due to a primary change, i.e., directly due to the disease process or be secondary to a disease process, e.g., due to reduced metabolic clearance.

However, the use of suitable criteria enables decisions to be made as to the best analyte to measure or detect disease. For example, prostatic cancer may be detected by measuring prostate specific antigen, however, it may be raised in other conditions or be normal in the presence of cancer; the specificity and sensitivity while acceptable mandate the test be part

of a group of investigations. Lack of awareness of the performance characteristics of clinical analytes in detecting pathology is under-appreciated by professionals and patient knowledge is much worse. For the well-established tests (for example see **Table 1**) the specificity and sensitivity are typically poor but a particular test in combination with other tests and the clinical history is a useful indicator of disease. In the early 1970s this led to the concept of screening to detect occult disease. Experience showed this was not clinically useful or cost-effective and with the

modern selective multichannel analyzers the move is toward organ profiles, e.g., to detect bone diseases such as osteomalacia, bone malignancies, parathyroid adenoma, etc.

Clinical laboratories prioritize the turnaround of results. This is decided by clinical demand and analytical feasibility. Laboratories offer their full repertoire during the day and during the week and offer a more restricted service outside these times. However, the changes in healthcare are increasing demands on laboratories and there is more 24 h, 7 days a week availability of the test menu. Some tests reflect acute changes in metabolism, and are needed for immediate patient management and rapid results are required (<30 min). The concept of turn-round time is from the sample being taken to the requesting clinician assessing the significance of the result; so called 'vein to brain' time.

Many other tests are clinically useful if available within 3–4 h of receipt; others, of less immediate but possibly of as great clinical importance, greater complexity or greater cost, may require several days turn-round; they may be so rare, difficult or costly that they need to be sent to a specialist center.

The unique accession number given to a sample on receipt follows it through the analytical phase to the production of the final report. Machines with bar-code reading capability can read bar-coded laboratory accession numbers and link these with associated reports for these patients via computer interfacing. This allows the production of a cumulative report from which disease trends can be assessed. Bar-coded primary sample tubes (i.e., no aliquoting) can be handled on these analyzers.

Better integration of laboratory information management systems with hospital patient administration systems is resulting in tests being ordered at ward or primary care level with electronic result returns greatly improving the electronic patient record and minimizing delays. This is enhanced by positive patient identification with patients having bar-coded labels using palm computers, which enables production of requests and consideration of results at the bedside. This approach minimizes mixing of patients samples and enhances patient safety.

## Analytical Goals

As all analysts know, there is inherent imprecision in any analytical system. The question is at what point does the imprecision become acceptable for the purpose for which the analysis is going to be used. In clinical analysis this can be addressed by determining

what the biological analyte variability is in a population. It has been determined that analytical imprecision should be no greater than 25% of the population biological variation. The goal for bias is always zero, however, if the source and degree of interference are known this does not preclude the use of that analytical method. For example, recommendations on cholesterol screening advise a bias of zero and an imprecision of better than 3%. This enables significant changes to be detected. Using the Jaffe reaction to assay creatinine, a marker of renal function, imprecision is of the order of 5% and there is positive interference from ketones, which occur in diabetic coma and negative interference from bilirubin, the cause of the yellow color in jaundice.

Perhaps half of the methods in current use meet the current analytical goals. The clinical benefit that derives from meeting these goals has not been demonstrated to evidence-based medicine standards.

## Practical Aspects of Analysis

### Sample Handling

Clinical analysis is performed on a variety of matrices: whole blood (e.g. in hematology), serum, from clotted blood, or plasma from anticoagulated blood (both in clinical chemistry), urine, other fluids (e.g., amniotic fluid), feces, tissues, and occasionally on other matrices. Normally analytes are measured in serum, as it is easily accessible and is used to reflect target concentrations in tissue or receptors. Generally urine is examined to assess renal tract function and integrity.

Owing to the high workloads (often 2000 or more samples per day), samples are uniquely identified with a laboratory accession number. If serum or plasma is required, the blood sample is centrifuged (at ~3000 rpm) and the serum or plasma may be removed but primary tube sampling is now prevalent and a gel separator allows 'pouring off' of the supernatant if desired (see the section 'Large-capacity analyzers').

Sending samples to the correct path, aliquotting for different analytical workstations and tracking of samples is difficult and time-consuming. Given the volume of work in clinical laboratories and the potential significance of the results, these aspects are very important. The preanalytical requirements are now being automated. With bar-coded samples and host-query interfacing it is possible to correctly divide aliquot appropriate volumes and rack for analysis automatically. Following analysis automated sample handling enables X–Y–Z (column–row–rack)

storage and retrieval to be readily achieved, such systems range in price from £60 000 upward.

Traditional wet-chemical procedures required serum or plasma proteins to be removed prior to further analysis; this was achieved by precipitation, usually with acids such as trichloroacetic acid or metal ions, e.g., zinc sulfate. If organic acid extraction were to be used, this would also precipitate proteins. Such methods are scarcely used in routine practice. Automated procedures can readily be adapted to determine colorimetric endpoint or rates of reaction; the avoidance of protein interference is more difficult (see the section 'Large-capacity analyzers').

### Calibration

It is clearly recognized that analytes in aqueous solution do not perform in the same way as analytes in a biological matrix; it is therefore necessary to use, e.g., serum as diluent. Owing to the cost, and more recently health hazards, nonhuman, usually bovine, serum may be used. This can cause problems for some analytes, e.g., albumin.

Some analytes are difficult to standardize owing to the complexity of developing a primary standard, or analyte inhomogeneity, an International Standard is used to provide comparability. Such standards are based on the principle of transferability. A definitive method, e.g., isotope-dilution mass spectrometry (MS), is used to obtain the best possible estimate of the accurate concentration of the analyte. This value would be transferred to a reference method, which is usually a very carefully documented analytical procedure of known and impeccable performance, e.g., Abel-Kendal assay for cholesterol, and standards are compared in the procedure before being used in the filed.

### Quality Control

Internal, i.e., within department, quality control utilizing material of known concentration is used to assess the precision and bias of the assay. Inadequate analytical performance in internal schemes will result in rejection of the results from the analyzed samples. Performance criteria may simply be assessed as  $>2$  standard deviations from the mean or by more complex bias and imprecision estimates such as Westgard rules. Rejection of quality control may require analyses to be rejected, corrective action to be taken, and the analyses repeated.

External, i.e., outside-department, quality assurance schemes compare the accuracy of different methods and may derive targets from an all-laboratory trimmed mean (i.e., outliers discarded), on spiked values, or on a reference standard; usually

means and variance for method types are given. The consensual nature of such schemes validates the control material values and these can then be used by analytes, which correct poor method performance. The identities of the laboratories are unknown to all but the scheme organizers. In external schemes, in the event of difficulty there may be a communication followed by a visit from the panel of experts to help resolve a problem; in some countries continued poor performance will result in the laboratory losing its license.

### Units Used in Reporting

There is a diversity of units of concentration in use in analytical clinical biochemistry; however, clinical laboratories in Europe and elsewhere (except the United States) use the SI (System International) units. The base unit of volume is always the liter (written as/l), the concentration is expressed in moles or part thereof, e.g., micromole; numerical values should, ideally, be written between 1 and 999 per liter, e.g.,  $232 \text{ nmol l}^{-1}$ .

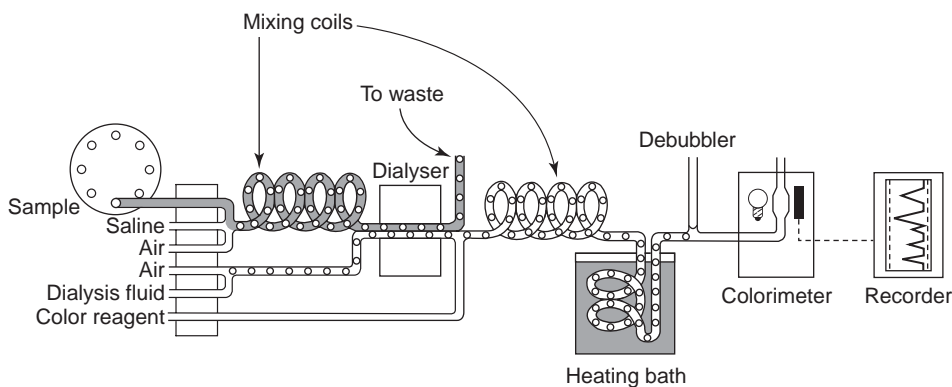
Archaic legal requirements, e.g., to report ethanol in milligram percent are sometimes used. To use molar units, the molecular weight of the species must be known as well as its degree of heterogeneity; for proteins, e.g., hemoglobin, the results are expressed as mass per liter, e.g.,  $\text{g l}^{-1}$  as there is high heterogeneity.

Worldwide agreement on the units to be used in analytical clinical biochemistry would greatly facilitate exchange of information and minimize unnecessary errors.

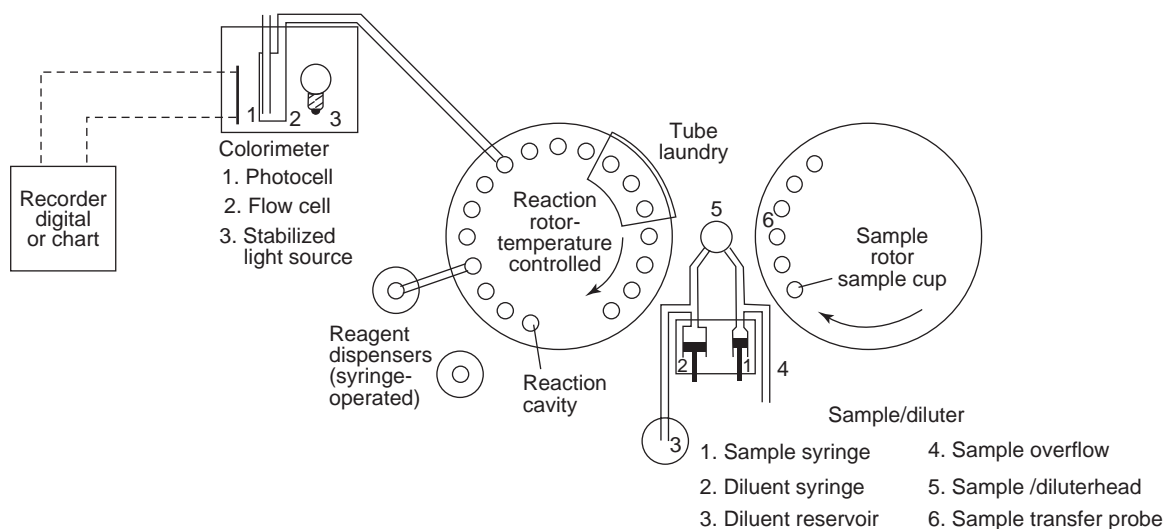
## Automation

### Large-Capacity Analyzers

Increasing workloads meant traditional manual wet-chemistry methods were too time-consuming and better diagnostics demanded better accuracy and precision. The first successful automation, in the early 1960s, was Skeggs continuous flow air-segmented systems in which a peristaltic pump utilizing tubing of different internal diameters proportionally delivered reagents, diluent, sample, and air for subsequent dialysis to remove protein and colorimetric or flame photometric analysis. Single analyte analyzers (Figure 1) were combined to produce the multichannel (multianalyte) analyzers. Analytical speeds of initially 20–60 samples per hour could be increased to 120 samples per hour for up to 24 different chemistries, but the inexorably increasing workloads exceeded the capabilities of this technology.



**Figure 1** Flow diagram of a typical single-channel, segmented flow Auto Analyser®.



**Figure 2** Flow diagram of a typical discrete analyser.

Discretionary analysis, as opposed to multianalyte panel screening developed from the realization that nonselective screening did not yield the expected health benefits. This approach was best addressed with discrete analyzers.

These modern discrete analyzers (Figure 2), typically utilize wet-chemistry with photometry of selective electrodes for the cations sodium and potassium. Dry-film technology using similar measurement principles are a significant minority of clinical analyzers. Each test is delivered as an individual slide, which minimizes infection risks. Either type allows different mixtures of tests within a defined menu to be done on different samples during the same analytical run (discretionary). The protein problem is circumvented by sample dilution and the use of surfactants.

Discrete analyzers can operate at up to 300 samples per hour, but may be dependent on the number

of tests requested per sample; speeds of 150 samples per hour for 16 tests or more are readily and commonly achievable. The other major advance was to move away from batch to random-access operation; this optimizes machine operations and uses bar-coding to enable random loading of the machine. Photometric detection is of either endpoint reactions or rate of reaction. Ion-selective electrodes are used to detect electrolytes. Many assays are based on enzymes specific for the substrate, with measurement of NADH/NADPH production or utilization or colorimetry of an end product. Users may be tied to the manufacturers' chemistries, but there is a trend to more 'open' systems where the analyst can use alternative chemistries. Combination of chemistry analyzers with immunoassay analyzers improves efficiency, these may be as an integrated analyzer or joined by a track system; they are further enhanced by linked preanalytical sample handling. Such

combinations are becoming increasingly popular despite the changes to building structures that they entail. Such analyzers can cost well over £200 000.

### Medium-Capacity Analyzers

Today such analyzers are discrete analyzers operating at high throughput on a wider menu of tests than their large cousins. They are usually suited to more sophisticated chemistries, such as those used in specific protein analysis or enzyme-mediated immunoassays for drug analysis.

### Specialized Analyzers

A number of specialized analyzers are found in clinical laboratories. They include machines to measure blood gases, serum and urine osmolality, and blood glucose.

### Point of Care Testing

Tests performed outside clinical laboratories are attracting increasing interest. Complex technology and chemistry are combined in easy-to-use machines; dry-film technology is the commonest technique used. The philosophy is that economically the immediacy of the result offsets the high consumables costs. For many analytes there is no need for such immediacy; however, patients can perform their own home glucose monitoring, GPs and others may offer cholesterol analysis as part of an assessment of cardiovascular disease, and intensive therapy units in hospitals invariably have blood gas machines to hand.

Clinical laboratories have a constructive approach to these developments and should offer training to operators and local quality control schemes to ensure that equipment is used correctly and is performing to specification. Problems in point of care testing include the use of inappropriate chemistries, operator-mediated analytical error, the use of inappropriate units of concentration; this results because the United States, unlike nearly every other country, does not use molar SI units for reporting clinical laboratory values (see the section 'Units used in reporting') and the failure to permanently record the results. These machines and their misuse have resulted in hazard warning notices from the Department of Health in the United Kingdom and the inclusion of such testing under the Clinical Laboratory Improvement Act 1988 in the United States. A significant issue is the lack of unequivocal integration of results in the electronic patient record.

## Other Techniques

Clinical laboratories also utilize fluorescence, cell culture techniques, and DNA analysis, depending on the demands made on them.

### Immunoassays

A significant proportion of many clinical laboratories' workload is dealt with by one or other form of immunoassay; drugs; proteins and hormones (steroid and protein). Radioisotopic assays (RIA) are now rarely used having been supplanted by safer non-isotopic automated analyzers. There has been rapid growth in the development of immunoassay analyzers with the ability to measure a wide range of hormones, tumor markers, and hematinics. These are high-capacity analyzers. Different antibody specificities and variation in calibrator material makes comparison of results between different machine types problematic. This is slowly being addressed. Combined general clinical chemistry analyzers with immunoassay analyzers are now available. Some suppliers provide 'track' systems enabling full automation of the combined sample reception and analytical process.

### Atomic Absorption Spectrometry

Determination of heavy metals and trace elements, e.g., lead, cadmium, copper, and zinc are best measured using atomic absorption. Electrothermal equipment has a wider application owing to its greater sensitivity. The workload, limited application of the tests, and the capital cost of equipment mean that spectrometers are located in larger laboratories. Inductively coupled plasma-mass spectrometry (ICP-MS) is used in some specialist centers, supplanting atomic absorption spectroscopy.

### UV-Visible Spectrophotometry

Spectrophotometry is the mainstay of the main chemical analyzers. However, stand-alone machines are occasionally used to measure colorimetric and endpoint assays. There has been a recent resurgence of interest in the use of visible spectrophotometry to detect bilirubin released following an intracranial bleed not detected by CT scan.

### Chromatography

Chromatography is used in a small proportion of routine clinical analysis. The sequential nature of sample introduction in column chromatography and the labor-intensive nature of thin-layer chromatography



means that chromatography is reserved for special assays where the ability to simultaneously determine metabolites and/or classes of compounds provides a distinct benefit e.g., amino acid analysis, anticonvulsant analysis.

Liquid chromatography (LC) is the most common of the chromatographic modes, as it has more general application in clinical laboratories than gas-liquid chromatography. Thin-layer chromatography is used for qualitative procedures, e.g., screening for drugs of abuse. Larger clinical laboratories will have a gas-liquid chromatograph and liquid chromatograph. The commonest use of LC is to measure hemoglobin A<sub>1c</sub> as a marker of diabetic control. Mass spectrometric detection linked to gas chromatography is used to confirm drugs of abuse by immunoassay screens. LC tandem MS have proven invaluable for screening for inborn errors of metabolism in pediatrics.

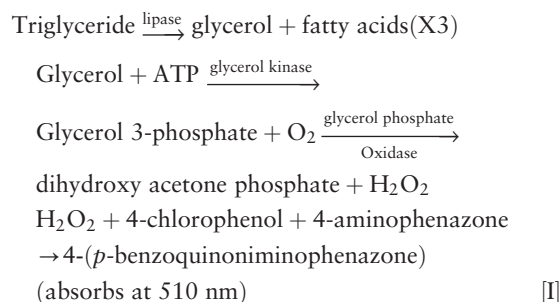
### Electrophoresis

Electrophoresis on cellulose acetate once widely used from separating serum proteins has been replaced by commercial high-resolution gel electrophoresis.

Isoelectric focusing for specific applications such as cerebrospinal fluid oligoclonal bands is of more specialist interest. Immunoelectrophoresis (i.e., electrophoresis followed by the application of specific antibodies) for typing monoclonal gammopathies (myelomas) is the procedure of choice. Capillary electrophoresis offers the opportunity to determine genotypes for homo or heterozygotes of, e.g., enzyme defects, following polymerase chain reaction.

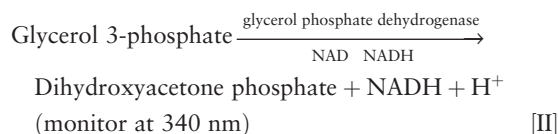
### Enzymes

Enzymes are detected in serum as markers of damage to cells or induction of synthesis. Enzymes are also used as reagents; commonly the product is linked to a color reaction with, for example, 4-aminophenzone (see eqn [I]).



Alternatively, the enzyme reaction is NAD- or NADP-linked; the change from or to the reduced

form can then be monitored (see eqn [II]).



The use of immobilized enzymes minimizes costs, but most analyzers still use liquid reagents.

### Future Directions

In clinical laboratories sample handling, including analysis, is becoming much more automated, requiring lower skill levels; point of care testing will further increase. There will be a need for central clinical laboratories to participate in point of care testing to ensure that high standards of analysis are set and maintained. Perhaps more routine testing out of centralized laboratories into the community: a disseminated laboratory.

New techniques will be introduced, e.g., proton nuclear magnetic resonance (NMR) spectroscopy of biological fluids or perhaps combined NMR imaging with spectroscopy allowing *in situ* analysis.

With the sequencing of the human genome DNA analysis will make a significant contribution to determining the individuals susceptibility to disease. Currently DNA analyzers main contribution is to the identifications of infective agents, in assessing tumor susceptibility and in the diagnosis of inborn errors of metabolism. To harness these advances, different, probably more sophisticated samples will be required. The next challenge is proteomics: exciting times lie ahead for clinical analysis.

**See also:** **Blood and Plasma. Chemometrics and Statistics:** Multivariate Calibration Techniques. **Chromatography:** Overview. **Electrophoresis:** Clinical Applications. **Enzymes:** Enzymes in Physiological Samples; Enzyme-Based Assays. **Quality Assurance:** Quality Control; Primary Standards; Laboratory Information Management Systems. **Sample Handling:** Automated Sample Preparation.

### Further Reading

- Gowenlock A (ed.) (1987) *Varley's Practical Clinical Biochemistry*. London: Heinemann.
- Kaplan LA and Presce AJ (1996) *Clinical Chemistry. Theory, Analysis and Correlation*, 3rd edn. St. Louis: C.V. Mosby.
- Noe DA (1985) *The Logic of Laboratory Medicine*. Baltimore: Urban and Schwarzenberg.
- Price CP and Hicks JM (eds.) (1999) *Point of Care Testing*. Washington: AACC Press.
- Price CP and Newman DJ (eds.) (1997) *Principles and Practice of Immunoassay*, 2nd edn. London: MacMillan.

## Sample Handling

**F Bressolle, O Nicolas, and M T Kelly**, University Montpellier I, Montpellier, France

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

During pharmacokinetic and metabolism studies, as well as during routine therapeutic drug monitoring, the levels of the drug and its metabolites are commonly monitored in biological matrices (blood, plasma, serum, saliva, urine, bile, feces, cerebrospinal fluid, tissues, and *in vitro* biological samples). However, there are many analytical issues that must be addressed including (1) low concentrations of drugs in the matrices, (2) the complexity of biological materials, (3) small sample volumes, and (4) the choice of the analytical method to accurately quantify the drug in the matrix. Among the different analytical methods available, gas chromatography (GC), and high-performance liquid chromatography (HPLC) are among the most widely used. However, as the majority of drugs are nonvolatile, their analysis by GC requires that they be derivatized to render them amenable to gas-phase analyses. Thus, HPLC based on traditional detection methods, such as ultraviolet and fluorescence is widely used. Indeed, this method is particularly well adapted to the analysis of drugs and their more polar metabolites. In the past few years, the use of liquid chromatography coupled with mass spectrometry (LC-MS) or with tandem mass spectrometry (LC-MS/MS) has been growing in importance especially with the introduction of the electrospray interface.

The aim of this article is to present the different methods used in sample preparation for biopharmaceutical analysis. Their advantages, disadvantages, and the possibility of automation will be presented.

Before proceeding to discuss the relevance and methodology of sample pretreatment in bioanalysis, it is worth noting that in a few limited instances, it is possible to inject untreated body fluids directly onto the analytical HPLC column. This approach is not possible in GC. Direct injection is only feasible for samples (urine or bile, for example) containing high drug concentrations of the analyte(s) and very low protein concentrations. In this case, the matrix is simply diluted in deionized water before injection. This approach is difficult for serum or plasma samples due to the problem of protein precipitation on

the analytical column. However, columns have been specially designed for direct injection of serum or plasma. These columns totally exclude proteins while retaining smaller organic molecules.

In most instances, a biological sample containing a compound of interest requires some kind of sample pretreatment. Such procedures are principally carried out to isolate the drug from interfering matrix substances, but they also serve to liberate the drug from protein-binding sites and to concentrate the drug for more sensitive analysis. Indeed, removal of endogenous components is particularly important where these interferents become irreversibly adsorbed onto the packing material of the column, as is the case with lipids, or precipitated in the chromatographic system, as is the case with proteins.

### Elimination of Interfering Compounds

#### Ultrafiltration

Since it is not the total, but rather the free drug concentration that correlates with the concentration at the site of action, the free drug concentration is often considered as the best estimate of the pharmacologically active drug concentration. A protein-free solution may be obtained by filtration through a size-selective semipermeable membrane under pressure or by centrifugation in a membrane cone. The separation of free drugs is now simplified by the availability of ultrafiltration devices. Among the most widely used commercially available kits are Centrifree and MPS micropartition devices from Amicon (molecular weight cutoff 30 000 Da), 'Emit' free level filter system from Syva, and Molcut II from Millipore (molecular weight cutoff 10 000 Da). Ultrafiltration membranes are microporous in their structure; all molecules greater than the largest pore diameter are retained and all the other molecules, smaller than the smallest pore pass completely through the membrane. The hydrostatic pressure applied during the ultrafiltration process varies between 1 and 10 atm. However, there are some problems associated with ultrafiltration such as (1) adsorption of the drug onto the inert membrane, particularly if the drug is present at trace levels, (2) leakage of bound drug through membrane, (3) stability of the binding equilibrium during the separation process, and (4) a major factor limiting the speed and effectiveness of the process is the buildup of a layer of proteins on the upstream surface of the membrane.



**Table 1** Reagents used for protein precipitation

Principle	Reagent
Modification of the ionic strength	Saturated solution of ammonium sulfate
Insoluble salt precipitation	
Acidic precipitants	10–20% trichloroacetic acid (v/v) 10–20% perchloric acid (v/v) 5% metaphosphoric acid (v/v)
Basic precipitants	Zinc sulfate–sodium hydroxide Zinc sulfate–barium hydroxide Copper sulfate–potassium hydroxide
Modification of dielectric constant	Acetonitrile Methanol Ethanol Acetone

### Protein Precipitation

Protein precipitation can be obtained by modification of (1) the ionic strength, (2) the pH, or (3) the dielectric constant (Table 1).

Proteins are positively charged in strongly acidic solutions and negatively charged in strongly basic solutions, owing to their zwitterionic nature. Acidic or anionic precipitants form insoluble protein salts with the positively charged amino groups of proteins at low pH. Trichloroacetic and perchloric acids are widely used to separate compounds from tissue and blood samples. After centrifugation, an aliquot of the supernatant may be injected into the analytical column. To avoid damaging the column, the analyst must use a column specifically designed for operation at low pH or use a mobile phase containing a high molar concentration of buffer. Acidic protein precipitants are obviously unsuitable for compounds which are prone to acid hydrolysis. Proteins may also be precipitated by using cationic precipitants; these reagents form insoluble protein salts with the carboxylate groups of proteins. This method is unsuitable for compounds that have a tendency to form metal complexes. Cationic and acid protein precipitants may be used in conjunction with organic precipitants. If precipitation is effected by the addition of inorganic salts, the supernatant liquid will contain other constituents besides proteins, and frequently the drug peak can be accompanied and complicated by several other peaks which adversely affect assay precision. Where dilution of the sample is not of major importance, organic precipitants should be considered as they are less aggressive than the ionic precipitants.

With the addition of a water-miscible organic solvent to a biological sample such as serum or plasma, the solubility of proteins is lowered and they

**Table 2** Efficiency of protein removal from plasma according to the proportion of water/organic solvent

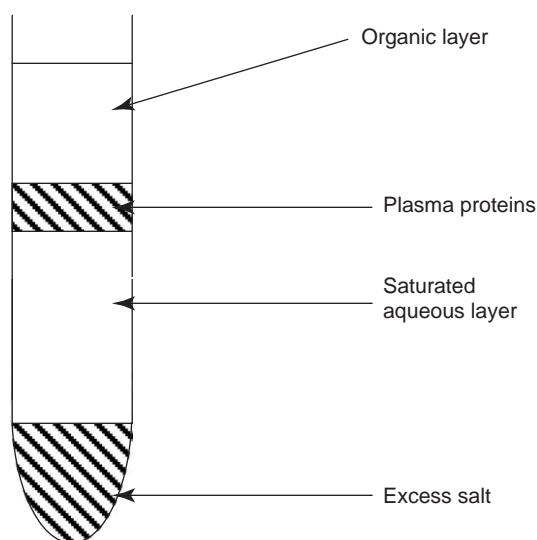
	Organic solvent/plasma (v/v)				
	0.2	0.6	0.8	1.0	2.0
Acetonitrile	13.4	45.8	88.1	97.2	99.7
Acetone	1.5	33.6	71.0	96.2	99.4
Methanol	17.6	32.2	49.3	73.4	98.7

From Blanchard J (1981) Evaluation of the relative efficacy of various techniques for deproteinizing plasma samples prior to HPLC analysis. *Journal of Chromatography Biomedical Applications* 226: 455–460.

precipitate out. The sample is centrifuged to produce a clear supernatant containing the compound of interest, and an aliquot is injected onto the column. It is important to use a protein precipitating solvent in which the analyte is highly soluble, otherwise it may adsorb onto, or co-precipitate with the protein. Diluting the plasma (or serum) up to threefold with the organic solvent will effect removal of 99% of the proteins (Table 2), but effectively decreases sensitivity of the method. This drawback can be counteracted to some extent by increasing the volume of injection, though this can adversely affect chromatographic efficiency and peakshape. Alternatively, the supernatant may be evaporated, but this measure also concentrates any remaining interfering compounds. Alternatively a water-immiscible organic solvent, such as dichloromethane, may be added to the mixture. In this case, the lower layer is a mixture of dichloromethane–acetonitrile (or methanol) and the supernatant is the aqueous phase. These two phases can be separated using phase separators such as silicone-treated filter paper Whatman 1 PS.

### Demixing (Salting-Out) Method

This method is particularly well adapted to drug-level monitoring. It has been used in the determination of antifungal agents and more recently it was employed in the quantitation of antiproteases in plasma. Its principle is as follows: a water-miscible organic solvent is added to the biological sample, followed by an excess of salt (sodium carbonate, potassium chloride, or sodium chloride). Polar water-miscible solvents are forced to form a separate layer using this salting out technique, resulting in four different phases (Figure 1). The analyte is generally present in the organic layer, which may be injected directly onto the column. The advantage of this technique includes its simplicity and rapidity, and dilution of the analyte may be circumvented by



**Figure 1** Demixing (salting-out) method.

evaporating the organic solvent, though this step considerably lengthens the procedure.

## Liquid-Liquid Extraction

### Classic Liquid-Liquid Extraction

Liquid-liquid extraction is a very widely used method for the preparation of biological samples for subsequent analysis. Convenience and ease of use have contributed to the large popularity of this technique. In addition, liquid-liquid extraction permits concentration of the analyte.

This technique is based on the extraction of a drug depending on its partition between an aqueous and an immiscible organic phase. The partition coefficient ( $K_i$ ) of a drug  $i$  is given by

$$K_i = \frac{[C_i]_{\text{org}}}{[C_i]_{\text{aq}}}$$

where  $[C]_{\text{org}}$  and  $[C]_{\text{aq}}$  are the concentrations of drug in the organic and aqueous phases, respectively. Hence, the degree of solvent extraction is dependent on  $K_i$ , which in turn is dependent on the type of organic solvent as well as the pH and ionic strength of the solution. It is well known that repeated extractions (two or more) with small portions of solvent can recover more analyte than a single batch extraction, but for practical purposes, a biological sample is extracted no more than twice in most applications.

The solvent is selected to provide maximum extraction efficiency and minimum contaminants.

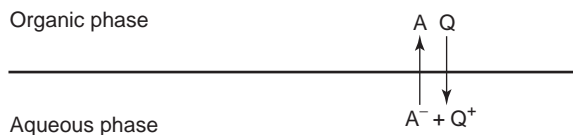
Polarity is usually the most important factor in the choice of the extraction solvent, and generally as this increases, the range of compounds extracted also increases. Hence, the solvent should be selected with minimum polarity consistent with high recovery of the drug. Drugs of high polarity are difficult to extract and require polar and consequently nonselective solvents. In some cases, to avoid losses by adsorption onto glass material, to increase the reproducibility of extraction, and to allow the extraction of more polar metabolites simultaneously with the parent drug, it is possible to select a mixture of solvents (1–5% of alcohol added to nonpolar organic solvents). For example, a mixture of methylene chloride-isopropyl alcohol (4:1, v/v) is used to extract a zwitterion, pefloxacin, from plasma and tissues.

Only the unionized form of a drug is extracted into the organic solvent. Therefore acidic drugs, which are unionized under acidic conditions are extracted from acidified matrices into organic solvents; basic drugs are likewise extracted from basified matrices. The optimal pH for acidic species is 1–2 pH units below their  $pK_a$  values, and for basic species it is 1–2 pH units above their  $pK_a$  values. Extraction can be difficult for compounds which are soluble in water at all pH values, for example water-soluble amphoteric and neutral drugs. In some cases, the addition of buffer salts to the aqueous solution increases its ionic strength and hence its polarity. This tends to decrease the affinity of polar compounds for the aqueous phase, and thus shifts the partition equilibrium in favor of extraction.

In order to increase the purity of extracts, interfering compounds may be preextracted into an organic phase which is then discarded. For example, urine contains many endogenous compounds, and a preliminary extraction from acidic urine improves the purity of a subsequent basic extract. Another useful method is back-extraction. By adjusting the pH of a new aqueous phase to that which will reionize the analyte, it can be back-extracted from the organic phase into the new aqueous layer. For back-extraction of basic drugs into an acidic phase, sulfuric and phosphoric acids are preferred to hydrochloric acid because many hydrochlorides are soluble in organic solvents. Back-extraction is more or less mandatory in GC since interfering substances in the initial extract tend to contaminate GC detectors, particularly nitrogen-phosphorus and electron capture detectors. This approach, when used prior to HPLC analysis, greatly improves the extraction selectivity, though overall recovery of the drug can be reduced and it adds considerably to analysis time.

### Ion-Pair Extraction

A useful method for dealing with a highly polar ionic drug is to convert it into a neutral ion-pair (IP) complex by the addition of an excess of suitable ions of opposite charge, followed by extraction of the complex into an organic solvent.



The extraction constant ( $K_{\text{ex}}$ ) is given by

$$K_{\text{ex}} = \frac{|\text{AQ}|_{\text{org}}}{|\text{A}^-|_{\text{aq}} \cdot |\text{Q}^+|_{\text{aq}}}$$

Table 3 lists common ion-pairing (IP) reagents. Formation of the complex depends on factors such as the pH of the aqueous phase, the type of organic solvent, and the nature and concentration of the counter ion. The IP extraction technique is useful for a variety of ionizable compounds which are difficult to extract in the unchanged form such as penicillins, amino acids, and conjugated metabolites. This method allows extraction of quaternary ammonium compounds such as tubocurarine which are ionized at all pH values.

**Table 3** Common ion-pairing reagents

<i>For basic analytes (most as sodium salt)</i>	<i>For acidic analytes</i>
Propanesulfonic acid	Triethylamine
Butanesulfonic acid	Tetramethylammonium bromide (or hydrogen sulfate)
1-Pentanesulfonic acid	Tetraethylammonium bromide (or hydrogen sulfate)
1-Hexanesulfonic acid	Tetrapropylammonium bromide (or hydrogen sulfate)
1-Octanesulfonic acid	Tetrabutylammonium bromide (or phosphate, iodide)
1-Nonanesulfonic acid	Tetrapentylammonium bromide
1-Decanesulfonic acid	Tetrahexylammonium bromide (or hydrogen sulfate)
1-Dodecanesulfonic acid	Tetraheptylammonium bromide
	Tetraoctylammonium bromide
Dodecylsulfate, sodium salt	Hexadecyltrimethyl ammonium hydroxide (or bromide, or hydrogen sulfate)
Diocylsulfosuccinate, sodium salt	Decamethylenebis (trimethylammonium bromide)
Trifluoroacetic acid	
Pentafluoropropionic acid	
Heptafluorobutyric acid	
Bis-2-ethylhexylphosphate	

In some cases in HPLC, the IP agent can be added to the mobile phase to reduce the retention times of the solutes and to obtain an acceptable asymmetry coefficient. For example, the retention times of enoxacin and its 4-oxo metabolite and of pefloxacin decrease as the concentration of tetrabutylammonium hydroxide in the mobile phase increases and the peak sharpens as a consequence of this decrease in retention time.

### Solvent Extraction Following Derivatization

Derivatization is commonly used to improve analyte detectability, although in some cases, derivatization is necessary to release analytes from their binding sites. For example, the antineoplastic agent cisplatin is chelated with diethyldithiocarbamate before HPLC–UV analysis.

### Problems with Liquid-Liquid Extraction

Solvent extraction is relatively time-consuming. It often requires the removal of solvents by evaporation. This lengthy step may also lead to decomposition of unstable compounds, particularly when heating is required. Losses can also occur at the redissolution stage. Another problem is the risk of adsorption of analytes onto glassware which can occur at the stage of extraction or solvent evaporation. Adsorption is much more noticeable when a drug is present at low concentrations. This problem may be partially overcome by (1) silanization of glassware, (2) the inclusion of 1–2% ethanol, isopropanol, or isoamyl alcohol into the extracting solvent, or (3) the use of polypropylene test tubes. These measures are, however, not always effective, and others may have to be devised.

### Offline Solid-Phase Extraction

Solid-phase extraction (SPE) has been available to analytical chemists for more than two decades. SPE often results in lower detection limits compared to other methods of sample preparation.

The principle of SPE is selective retention of the analytes on the sorbent by nonpolar, polar, and ionic interactions, and affinity and partition between the analytes, the sorbent, and/or the solvent.

Several SPE formats are commercially available and they may be divided in two broad categories. The first category includes hydrophilic packing materials such as inert particles of diatomaceous earth (kieselguhr). The sample and the water it contains are absorbed over a large surface area forming a thin aqueous film over the surface of each particle. A small volume of a water-immiscible organic solvent

such as dichloromethane is then passed through the column extracting the drug from the aqueous film of sample. Water and endogenous materials, such as pigments or other polar compounds are retained in the absorbed phase. The second category is based on the principles of chromatography. A six-step SPE protocol is the most common approach.

### Principle of the Six-Step Extraction Procedure

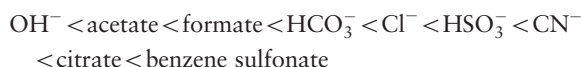
In each of the steps, conditions must be optimized for interaction between the analyte(s), matrix, and sorbent. These conditions include pH, ionic strength, solvent strength, solvent volume, and flow rates.

**Step 1: sample pretreatment** This step involves preparing the sample both physically and chemically for SPE extraction so that optimal conditions exist for analyte retention. The type of sample pretreatment is dependent on the analyte (particularly its stability), the type of matrix, the type of sorbent, and the nature of the analytes/sorbent interactions.

For aqueous samples containing analytes that are to be retained primarily by hydrophobic interactions, pH adjustment may be required to ensure that the surface and analyte (if ionizable) are not charged. It may be necessary to add a wetting agent (e.g., 1–2% methanol) to maintain an active sorbent surface. Where the primary interaction for analyte retention is ion exchange, the pH should be controlled to ensure complete ionization of the analyte and surface of the sorbent. The selectivity of the buffer cation (for cation exchange) or anion (for anion exchange) should be taken into account. Buffers that contain ions of lower affinity for the sorbent than the analyte itself facilitate analyte retention. The selectivity of some common cations is as follows (ions on the right will displace those on the left):



Selectivity of some common anions (ions on the right will displace those on the left):



Dilution of viscous samples may be necessary to reduce sample viscosity and to ensure a free-flowing sample.

**Step 2: SPE column conditioning** This step prepares SPE column for the extraction process by enabling the bonded phase to interact with the sample matrix. For aqueous samples the sorbent is wetted with an organic solvent such as methanol. It is important

when using nonpolar sorbents so that a residual amount of the solvent remains in order to keep the sorbent ‘wetted’. For nonaqueous samples, the sorbent is wetted with the matrix solvent. The volume of conditioning solvent is  $2 \times 1.0$  ml of solvent per 100 mg sorbent.

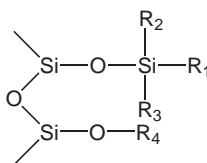
**Step 3: SPE column equilibration** This step removes excess conditioning solvent. To maximize analyte retention, the sorbent is treated with a ‘matrix-like’ solvent. For aqueous samples, the solvent should be similar to the sample matrix in terms of pH and ionic strength (if the retention mechanism is ion exchange, an ionic strength  $< 0.05$  M should be used). This step is not required for nonaqueous samples.

Proper conditioning is important for conventional reverse-phase supports, which require a ‘wetting’ step with a polar organic solvent followed by sorbent equilibration prior to sample application (steps 2 and 3) to ensure reproducible results. However, SPE columns that do not require preconditioning are now available. Such sorbents have been developed in response to the changing pace of drug discovery – high throughput screening LC–MS, tougher legislation on solvent use, and safety requirements.

**Step 4: sample loading** During this step, it is essential that the correct loading flow rate is used. In method development, it is important to start with a low flow rate to ensure that optimum recoveries are obtained. A good starting point is  $1 \text{ ml min}^{-1}$  for a 1 ml cartridge,  $3 \text{ ml min}^{-1}$  for a 3 ml cartridge and  $7 \text{ ml min}^{-1}$  for a 6 ml cartridge. Once the method chemistry has been established, the flow rate can be increased until some sample breakthrough (i.e., a drop in recovery) is observed. A flow rate which is slightly lower than the upper limit should be used.

**Step 5: interference elution step** This step selectively removes undesired compounds from the sorbent without eluting the analytes. Generally, the solvent used (1–2 ml for 100 mg of sorbent) is miscible with the sample matrix. For aqueous samples, the ionic strength and pH should be maintained to prevent analyte losses. The flow rate should be adjusted so that the solvent is in contact with the sorbent for 1–2 min. If the elution solvent is water-immiscible, drying of the column is required.

**Step 6: analyte elution** The solvent used must be one in which the analytes are soluble and that overcomes both primary and secondary sorption interactions. Small volumes of elution solvents should be used ( $250 \mu\text{l}$  per 100 mg of sorbent). The use of two small aliquots of solvent with a 1–4 min soak step

**Table 4** Properties and applications of the most widely used sorbents



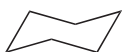
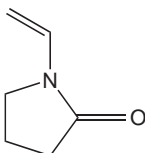
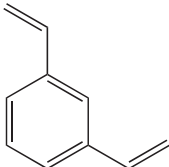
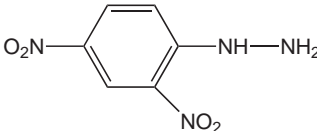
Sorbents	Retention mechanism	Applications
<b>Silica-based nonpolar</b>		
Monofunctional silane: C <sub>18</sub> , C <sub>8</sub> reversed-phase R <sub>1</sub> = (CH <sub>2</sub> ) <sub>17</sub> CH <sub>3</sub> , (CH <sub>2</sub> ) <sub>7</sub> CH <sub>3</sub> ; R <sub>2</sub> and R <sub>3</sub> = CH <sub>3</sub> ; R <sub>4</sub> = H (NC)*	Nonpolar, polar, cation exchange	Aqueous medium Wide polarity range *pH range, 3–7.5
Trifunctional silane: C <sub>18</sub> reversed-phase R <sub>1</sub> = (CH <sub>2</sub> ) <sub>17</sub> CH <sub>3</sub> ; R <sub>4</sub> = H (NC) R <sub>1</sub> = (CH <sub>2</sub> ) <sub>17</sub> CH <sub>3</sub> ; R <sub>4</sub> = Si(CH <sub>3</sub> ) <sub>3</sub> (EC)*	Nonpolar, polar, cation exchange Nonpolar	Aqueous medium Wide polarity range *pH range, 3–7.5
Trifunctional silane: C <sub>8</sub> ; C <sub>6</sub> ; C <sub>4</sub> ; C <sub>2</sub> reversed-phase R <sub>1</sub> = (CH <sub>2</sub> ) <sub>7</sub> CH <sub>3</sub> , (CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub> , (CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub> , CH <sub>2</sub> CH <sub>3</sub> ; R <sub>4</sub> = H (NC) R <sub>1</sub> = (CH <sub>2</sub> ) <sub>7</sub> CH <sub>3</sub> , CH <sub>2</sub> CH <sub>3</sub> ; R <sub>4</sub> = Si(CH <sub>3</sub> ) <sub>3</sub> (EC)	Nonpolar, polar, cation exchange Nonpolar	Aqueous medium; wide polarity range; less retentive than C <sub>18</sub> ; retention properties, C <sub>8</sub> > C <sub>6</sub> > C <sub>4</sub> > C <sub>2</sub> if retention based on nonpolar interactions alone
Trifunctional silane: cyanopropyl R <sub>1</sub> = (CH <sub>2</sub> ) <sub>3</sub> CN; R <sub>4</sub> = H (NC) R <sub>1</sub> = (CH <sub>2</sub> ) <sub>3</sub> CN; R <sub>4</sub> = Si(CH <sub>3</sub> ) <sub>3</sub> (EC)	Polar if nonpolar matrix Nonpolar, cation exchange (NC) if aqueous medium	Can be used as a less polar alternative to silica in normal-phase applications or as less hydrophobic alternative to C <sub>18</sub> or C <sub>8</sub> in reversed-phase applications
Trifunctional silane: phenyl		
R <sub>1</sub> =  ; R <sub>4</sub> = H (NC)	Nonpolar, polar, cation exchange	Aqueous medium; wide polarity range; less retentive than C <sub>18</sub> ; different selectivity than C <sub>8</sub> and C <sub>18</sub> for aromatic compounds
R <sub>1</sub> =  ; R <sub>4</sub> = Si(CH <sub>3</sub> ) <sub>3</sub> (EC)	Nonpolar	
Trifunctional silane: cyclohexyl (EC)		
R <sub>1</sub> =  ; R <sub>4</sub> = Si(CH <sub>3</sub> ) <sub>3</sub>	Nonpolar	Aqueous medium; wide polarity range; less retentive than C <sub>18</sub> ; different selectivity from C <sub>8</sub> and C <sub>18</sub> for phenolic compounds
<b>Silica-based polar</b>		
Trifunctional silane: aminopropyl R <sub>1</sub> = (CH <sub>2</sub> ) <sub>3</sub> NH <sub>2</sub> ; R <sub>4</sub> = H (NC)	Polar bonded phase with basic character if nonpolar matrix Anion exchange if aqueous medium	Can be used as a polar sorbent, like silica, with different selectivity for acidic/basic analytes or as a weak anion exchanger in aqueous medium
Trifunctional silane: diol R <sub>1</sub> = (CH <sub>2</sub> ) <sub>3</sub> OCH <sub>2</sub> CH(OH)CH <sub>2</sub> OH; R <sub>4</sub> = H (NC)	Polar bonded phase with neutral character if nonpolar matrix Nonpolar, semipolar if aqueous medium	Can be used as an alternative to silica in normal-phase applications where the acidic character of silica is undesirable or as very weakly interacting phase in aqueous applications



Table 4 Continued

Sorbents	Retention mechanism	Applications
Other polar		
Silica	Polar sorbent The binding mechanism can be hydrogen bonding or dipole–dipole interaction Silanol groups are ionizable and unbounded silica can be used as a weak cation exchanger	To adsorb analytes from nonpolar solvents
Alumina	Available in acidic, basic, and neutral High activity grades	Similar to silica
Florisil <sup>®</sup> (magnesia–silica gel)	Polar highly active, weak-basis sorbent for adsorption	Adsorption of low to moderate polarity compounds from non-aqueous solutions
Silica-based anion-exchanges		
R <sub>1</sub> = (CH <sub>2</sub> ) <sub>3</sub> NH <sub>2</sub> , (CH <sub>2</sub> ) <sub>3</sub> NH(CH <sub>2</sub> ) <sub>2</sub> NH <sub>2</sub> ; R <sub>4</sub> = H	Hydrophilic character Weak anion-exchangers (pK <sub>a</sub> = 9.8, 10.1 and 10.9) at pH ≤ 7.8, 8.1	Extraction of anionic analytes in aqueous and nonaqueous solutions
R <sub>1</sub> = (CH <sub>2</sub> ) <sub>3</sub> N <sup>+</sup> (CH <sub>3</sub> ) <sub>3</sub> Cl <sup>−</sup> , (CH <sub>2</sub> ) <sub>3</sub> N <sup>+</sup> (CH <sub>3</sub> ) <sub>3</sub> CH <sub>3</sub> COO <sup>−</sup> ; R <sub>4</sub> = H	Strong anion-exchangers	
Silica-based cation-exchanges		
R <sub>1</sub> = (CH <sub>2</sub> ) <sub>3</sub> COOH	Hydrophilic character Weak cation-exchanger, (pK <sub>a</sub> = 4.8) positive charge at pH ≥ 6.8	Extraction of cationic analytes in aqueous and nonaqueous solutions
R <sub>1</sub> = benzene sulfonic acid, (CH <sub>2</sub> ) <sub>3</sub> SO <sub>3</sub> <sup>−</sup> H <sup>+</sup> , ethylbenzene sulfonic acid R <sub>4</sub> = H	Strong cation-exchangers	
Others		
Poly(divinylbenzene–vinylpyrrolidone) resin	Hydrophobic character	An alternative to octadecyl-bonded silica for analytes that weakly adsorb to silica-based reversed-phase sorbents
Hydroxylated polystyrene–divinylbenzene co-polymer	Hyper cross-linked sorbent The sorbent does not need to be conditioned before use	Extraction of polar compounds; high sample loading flow rates can be used
Polystyrene–divinylbenzene co-polymer	Nonselective polymeric phase; Highly cross-linked sorbent	Extraction of polar compounds that are not adequately retained on C <sub>18</sub> or C <sub>8</sub> sorbents
Hydrophilic–lipophilic balance reversed-phase sorbent: co-polymer of	Water-wettable sorbent	Widely used in forensic analysis High capacity and retention of a wide variety of analytes Column use pH 1–14
<div><div></div><div></div></div>		
Others		
DNPH-silica	Acidified dinitrophenylhydrazine reagent coated on a silica sorbent	Analysis of formaldehyde and other aldehydes and ketones in air
		

\*NC, nonendcapped; EC, encapped.

between elution volumes is often more efficient than one large aliquot. Flow control is important to ensure reproducibility. For analytes that are retained by ion exchange, high ionic strength ( $>0.1 \text{ M}$ ) buffers

can be used for elution. For doubly charged analytes, buffers of  $>0.2 \text{ M}$  should be used. An organic component in the elution solvent may be necessary to overcome secondary hydrophobic interactions. Thus,

the elution of an ionic compound may be obtained by using a counter ion of the same charge as the analyte but which has a higher affinity for the sorbent. Alternatively, the pH of the solvent may be adjusted to 1 or 2 pH units above or below its  $pK_a$  (depending on whether the analyte is acidic or basic) in order to suppress its ionization, thus lowering its affinity for the ion-exchange sorbent.

### SPE Extraction Columns

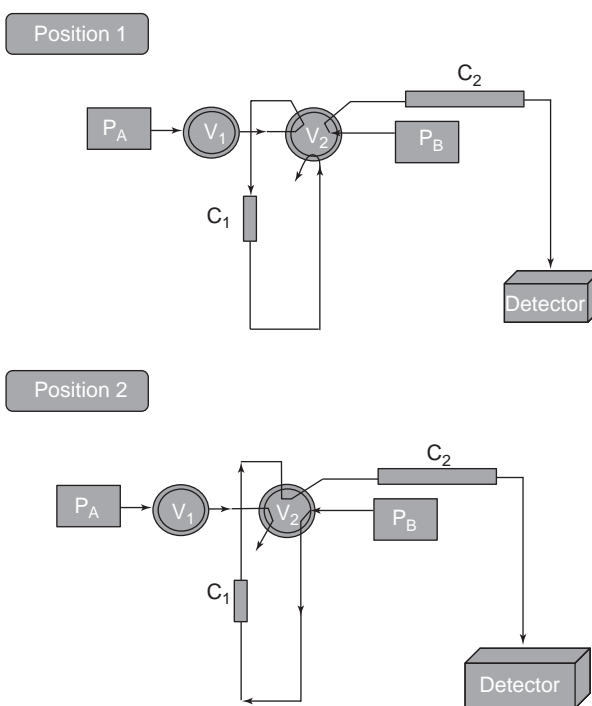
The most widely used sorbents with their characteristics are presented in **Table 4**. Mixed-mode columns containing mixtures of nonpolar ( $C_8$ ,  $C_{18}$ ,  $C_4$ , etc.) and strong anion or/and strong cation exchange functional groups are available for applications for both acidic and basic compounds. These columns are designed to capitalize on multiple interactions of the analyte or different types of interactions from more than one type of analyte. Mixed-mode columns allow the development of more robust procedures, which are less dependent on the matrix, making cleaner extracts possible.

IP SPE extraction is an additional method to the arsenal of techniques available to the analytical chemist. Cartridges containing  $C_{18}$ , phenyl, cyclohexyl, and polymeric media have been used in IP SPE. They are activated according to the manufacturer's recommendations followed by a solution of the IP reagent. IP reagent concentrations used range from 0.005 to 0.2 M. The same IP reagent is added to the samples before their application onto the conditioned cartridges. The cartridges may be washed with the aqueous IP reagent before elution with a stronger solvent, which may or may not contain the same IP reagent. Combining IP SPE with other SPE modes can yield substantial sample cleanup. One disadvantage of IP SPE is that occasionally there are problems with lot-to-lot differences of IP reagents that lead to variable recovery rates. IP SPE may also complicate and lengthen the method development process.

Recently, the use of 96-well SPE formats for extractions, automated using a robotic liquid-handling system, allows high throughput sample preparation.

### Online Solid-Phase Extraction

The typical scheme for an online sample cleanup procedure incorporating two pumps and a six-port switching valve is shown in **Figure 2**. Pump B is used to deliver the mobile phase and pump A is used to deliver the wash solvent, which is usually water or buffer, though a small percentage of organic solvent is sometimes added. In position 1, the wash solution



**Figure 2** Instrument arrangement for online solid-phase extraction with column switching.  $P_A$  and  $P_B$ , pumps;  $C_1$ , extraction column for the online SPE;  $C_2$ , analytical column;  $V_1$  and  $V_2$ , injection and switching valves; position 1, switching valve is positioned for online SPE cleanup and preconcentration; position 2, switching valve is positioned to transport sample to the analytical column.

is passed by pump A via the injector and the extraction column. Meanwhile, the mobile phase is pumped by pump B via the valve onto the analytical column, which is thus maintained in a state of constant equilibration. The sample was then loaded onto the extraction column, where cleanup and preconcentration take place. The polar matrix components are eluted to waste and the compounds of interest are selectively enriched on top of judiciously chosen sorbent. After a predetermined wash period, the valve is switched to position 2, the precolumn was then connected to the analytical column preferably in a back-flush configuration, where analytes were swept onto the analytical column by the HPLC mobile phase. The precolumn remains in line with the analytical column for predetermined period, and while the analytes are separated on the analytical column, the extraction column is re-equilibrated with the wash solution.

The column switching technique is particularly well adapted to biological fluids with low protein concentrations (for example, urine). For plasma samples, dilution or deproteinization may be required prior to injection into the extraction precolumn. It



should be noted that online SPE sorbents exist, known as 'internal surface reversed phase' that permit direct injection of protein-containing biological samples without prior cleanup.

**See also:** **Clinical Analysis:** Overview. **Extraction:** Solid-Phase Extraction.

## Further Reading

- Carson MC (2000) Ion-pair solid phase extraction. *Journal of Chromatography A* 885: 343–350.
- Gilar M, Bouvier ESP, and Compton BJ (2001) Advances in sample preparation in electromigration, chromatographic and mass spectrometric separation methods. *Journal of Chromatography A* 909: 111–135.
- Jemal M (2000) High-throughput quantitative bioanalysis by LC/MS/MS. *Biomedical Chromatography* 14: 422–429.
- Kataoka H (2003) New trends in sample preparation for clinical and pharmaceutical analysis. *Trends in Analytical Chemistry* 22: 232–244.
- Kelly MT (1992) Drug analysis in biological fluids. In: Smyth MR (ed.) *Chemical Analysis in Complex Matrices*. Chichester: Ellis Horwood.
- Maurer HH (1999) Systematic toxicological analysis procedures for acidic drugs and/or metabolites relevant to clinical and forensic toxicology and/or doping control. *Journal of Chromatography B: Biomedical Sciences and Applications* 733: 3–25.
- Mil'ánová C and Hutta M (2003) Role of biological matrices during the analysis of chiral drugs by liquid chromatography. *Journal of Chromatography B: Biomedical Applications* 797: 91–109.
- Oravcová J, Böhs B, and Lindner W (1996) Drug-protein studies. New trends in analytical and experimental methodology. *Journal of Chromatography B: Biomedical Applications* 677: 1–28.
- Riley CM, Lough WJ, and Wainer IW (1994) *Pharmaceutical and Biomedical Applications of High Performance Liquid Chromatography*. Amsterdam: Elsevier.
- Stevenson D (2000) Immuno-affinity solid-phase extraction. *Journal of Chromatography B: Biomedical Sciences and Applications* 745: 39–48.
- Thurman EM and Mills MS (1998) *Solid-Phase Extraction: Principles and Practice*. New York: Wiley.
- Wells D (2003) *High Throughput Bioanalytical Sample Preparation – Methods and Automation Strategies*. Amsterdam: Elsevier.

## Electrolytes in Physiological Samples

**W R Külpmann**, Medizinische Hochschule Hannover, Klinische Chemie, Hannover, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

In clinical chemistry, the term 'electrolytes' is confined to anions and cations, which are present in high concentrations in comparison to trace elements, such as iron, copper, or zinc. Thus, electrolytes comprise sodium, potassium, calcium, magnesium, chloride, phosphate, and hydrogen carbonate (bicarbonate). Electrolytes are determined frequently in patient care, because the clinical symptoms of electrolyte disturbances are often, at least at an early stage, not very evident. In addition, the clinical symptoms are frequently not clearly associated with electrolyte disturbances, but are ambiguous and may be caused by other diseases. The concentration of the electrolytes in serum is held constant biologically within narrow limits. During infusion therapy, rapid changes of electrolyte concentration may occur and need thorough

and rapid monitoring. Fast determinations of most electrolytes are now possible by the use of ion-selective electrodes (ISEs), which give results within minutes for blood; hence, there is no loss of time due to centrifugation to obtain plasma and no delay for clotting to obtain serum. Reliability of the results is as essential as speed. Reliability is warranted by concomitant procedures of quality assessment and plausibility control, which, however, can only be used to a very limited and inadequate extent to monitor appropriate preanalytical handling (sampling, transport, storage).

## Sodium

### Methods and Techniques

Sodium is determined in physiological samples using flame atomic emission spectrometry (FAES), UV-visible spectrophotometry, or potentiometry. All the methods can be used for serum (for convenience and brevity, 'serum' here includes plasma obtained from heparinized blood) or urine, whereas blood can only

should be noted that online SPE sorbents exist, known as 'internal surface reversed phase' that permit direct injection of protein-containing biological samples without prior cleanup.

**See also:** **Clinical Analysis:** Overview. **Extraction:** Solid-Phase Extraction.

## Further Reading

- Carson MC (2000) Ion-pair solid phase extraction. *Journal of Chromatography A* 885: 343–350.
- Gilar M, Bouvier ESP, and Compton BJ (2001) Advances in sample preparation in electromigration, chromatographic and mass spectrometric separation methods. *Journal of Chromatography A* 909: 111–135.
- Jemal M (2000) High-throughput quantitative bioanalysis by LC/MS/MS. *Biomedical Chromatography* 14: 422–429.
- Kataoka H (2003) New trends in sample preparation for clinical and pharmaceutical analysis. *Trends in Analytical Chemistry* 22: 232–244.
- Kelly MT (1992) Drug analysis in biological fluids. In: Smyth MR (ed.) *Chemical Analysis in Complex Matrices*. Chichester: Ellis Horwood.
- Maurer HH (1999) Systematic toxicological analysis procedures for acidic drugs and/or metabolites relevant to clinical and forensic toxicology and/or doping control. *Journal of Chromatography B: Biomedical Sciences and Applications* 733: 3–25.
- Mil'ánová C and Hutta M (2003) Role of biological matrices during the analysis of chiral drugs by liquid chromatography. *Journal of Chromatography B: Biomedical Applications* 797: 91–109.
- Oravcová J, Böhs B, and Lindner W (1996) Drug-protein studies. New trends in analytical and experimental methodology. *Journal of Chromatography B: Biomedical Applications* 677: 1–28.
- Riley CM, Lough WJ, and Wainer IW (1994) *Pharmaceutical and Biomedical Applications of High Performance Liquid Chromatography*. Amsterdam: Elsevier.
- Stevenson D (2000) Immuno-affinity solid-phase extraction. *Journal of Chromatography B: Biomedical Sciences and Applications* 745: 39–48.
- Thurman EM and Mills MS (1998) *Solid-Phase Extraction: Principles and Practice*. New York: Wiley.
- Wells D (2003) *High Throughput Bioanalytical Sample Preparation – Methods and Automation Strategies*. Amsterdam: Elsevier.

## Electrolytes in Physiological Samples

**W R Külpmann**, Medizinische Hochschule Hannover, Klinische Chemie, Hannover, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

In clinical chemistry, the term 'electrolytes' is confined to anions and cations, which are present in high concentrations in comparison to trace elements, such as iron, copper, or zinc. Thus, electrolytes comprise sodium, potassium, calcium, magnesium, chloride, phosphate, and hydrogen carbonate (bicarbonate). Electrolytes are determined frequently in patient care, because the clinical symptoms of electrolyte disturbances are often, at least at an early stage, not very evident. In addition, the clinical symptoms are frequently not clearly associated with electrolyte disturbances, but are ambiguous and may be caused by other diseases. The concentration of the electrolytes in serum is held constant biologically within narrow limits. During infusion therapy, rapid changes of electrolyte concentration may occur and need thorough

and rapid monitoring. Fast determinations of most electrolytes are now possible by the use of ion-selective electrodes (ISEs), which give results within minutes for blood; hence, there is no loss of time due to centrifugation to obtain plasma and no delay for clotting to obtain serum. Reliability of the results is as essential as speed. Reliability is warranted by concomitant procedures of quality assessment and plausibility control, which, however, can only be used to a very limited and inadequate extent to monitor appropriate preanalytical handling (sampling, transport, storage).

## Sodium

### Methods and Techniques

Sodium is determined in physiological samples using flame atomic emission spectrometry (FAES), UV-visible spectrophotometry, or potentiometry. All the methods can be used for serum (for convenience and brevity, 'serum' here includes plasma obtained from heparinized blood) or urine, whereas blood can only

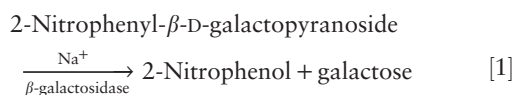
be analyzed for its sodium activity in the extracellular water phase by potentiometry.

**Flame atomic emission spectrometry** Basic information on FAES is presented elsewhere in this encyclopedia. Sodium measurements are performed at 590 nm with the use of a propane flame (1925°C). Physiological samples for sodium determination are highly diluted before measurement. The diluent and the calibrator solution contain the same concentration of lithium ions so as to balance flame instability by a concomitant measurement of lithium in the reference beam (the so-called lithium guideline). At the same time, lithium ions inhibit the ionization of sodium atoms. This procedure cannot be used in the case of therapy with lithium salts. That is why some authors prefer the concomitant measurement of caesium to that of lithium. Dilution adjusts the viscosity of the sample to that of the calibrator solution to produce identical aspiration rate and drop size on nebulization. As other electrolytes interfere with sodium measurement, their concentration in the calibrator solution must be similar to their concentration in the sample. For the measurement of sodium in urine, calibrator solutions different from those for serum measurement are needed as the electrolyte concentrations in urine samples are quite different from those in serum and their relations are very variable. As the concentration of the electrolytes in serum is rather constant, calibrator solutions for serum measurements can fulfill their function better than those for urine; in other words, urine determinations are usually less accurate. FAES proved to be sufficiently reliable to be used as the basic principle of the sodium reference measurement procedure. In routine use, however, FAES is less accurate. Its application is given up by most clinical laboratories in favor of potentiometric measurements

**UV-visible spectrophotometry** With the advent of FAES, the determination of sodium with magnesium uranylacetate became obsolete, as this method was insufficiently reliable for the determination of physiological samples.

In 1988 however, new methods were described for the measurement of sodium in physiological samples by UV-visible spectrophotometry.

1. *Enzymatic determination.* The activity of  $\beta$ -galactosidase is specifically dependent on the sodium concentration, and is determined according to



It is measured by following the absorption at 405 nm. As a linear relationship is found only at low sodium concentrations, a defined proportion of sodium ions is bound to a kryptand before the enzymatic reaction. The method is applicable to serum and urine.

2. *Kryptahemispherand determination.* The reagent consists of (1) an ionophore, with a cavity tailor-made for sodium ions, and (2) a chromophore, bound to the ionophore, with a characteristic absorption spectrum at a certain pH adjusted by buffer. On binding of sodium, the absorption spectrum is shifted proportionally. The method is interfered with by lipids and bilirubin.

FAES and ISEs need additional equipment, and FAES is hazardous (because of the use of propane). In contrast, sodium measurements by UV-visible spectrophotometry can be performed on an instrument that is always available in a clinical laboratory.

**Potentiometry** For sodium determinations by potentiometry, a glass electrode is usually used, but ion-carrier or ion-exchange membranes are also feasible for the ISE. A silver/silver chloride or a calomel electrode is used as the reference electrode.

According to the Nernst equation, the potential of the ISE is proportional to ion activity:

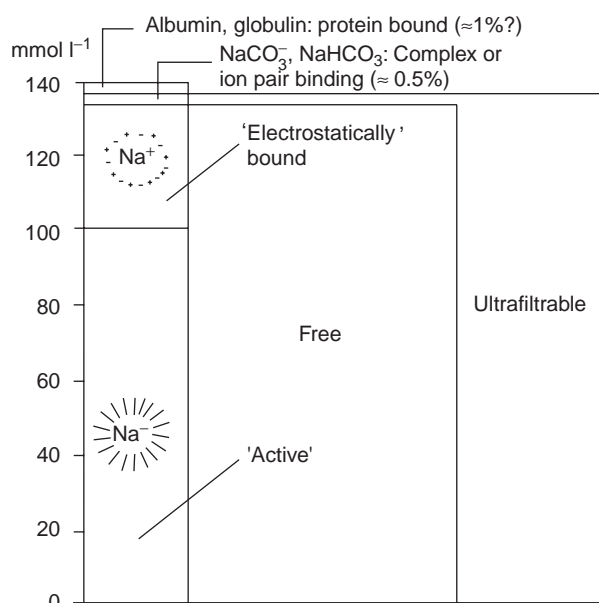
$$E = E_0 + (RT) (zF)^{-1} \ln \gamma c \quad [2]$$

where  $E$  is the ion-selective membrane potential measured,  $E_0$  a constant potential,  $R$  the gas constant,  $T$  the temperature (K),  $z$  the charge of the ion measured,  $F$  Faraday's constant,  $\gamma$  the molal activity coefficient, and  $c$  the concentration.

*Undiluted samples:* The activity coefficient of sodium in serum is nearly constant (0.747), as the ion concentrations do not vary much in this system. Sodium ions bound to hydrogen carbonate, protein, etc., escape measurement by ISEs and are not recorded. On the other hand, the measurement is independent of the protein and lipid concentration, i.e., of the variable size of the compartment of macromolecules, that does not contain electrolytes (Figure 1). The relationship between concentration and activity is given by

$$C_{\text{Na}} = (\tilde{m}_{\text{Na}} \gamma^{-1} + m_{\text{NaHCO}_3} + m_{\text{Na Prot}} + m_{\text{NaX}}) \rho_{\text{H}_2\text{O}} \quad [3]$$

where  $C_{\text{Na}}$  is the sodium concentration,  $\tilde{m}_{\text{Na}}$  the active molality of sodium,  $\gamma$  the molal activity coefficient,  $m_{\text{NaHCO}_3}$  the molal concentration of sodium bound to hydrogen carbonate (bicarbonate),  $m_{\text{Na Prot}}$  the molal concentration of sodium bound to



**Figure 1** Fractions of sodium in serum. (Modified from Maas AHJ, Siggaard-Andersen O, Weisberg HF, and Zijlstra WG (1985) Ion-selective electrodes for sodium and potassium: a new problem of what is measured and what should be reported. *Clinical Chemistry* 31: 482–485.)

protein,  $m_{\text{NaX}}$  the molal concentration of sodium bound to other anions, and  $\rho_{\text{H}_2\text{O}}$  the mass concentration of water in serum.

Usually measurements by ISEs are reported as concentrations, assuming identical activity coefficients for the sample and the corresponding calibrator. The values will differ from concentration measurements (e.g., by FAES), depending on the protein and lipid concentration (which is expressed by the water mass concentration  $\rho_{\text{H}_2\text{O}}$ ). In sera of 'normal' water concentration, the differences between ISE measurements and concentration measurements are less than expected from theory (+2.5% instead of 7%) because of factors such as sodium binding (see eqn [3]) and residual junction potential. Values obtained from ISE measurements in undiluted samples must be reported separately, as they reflect the chemical potential of the component in the system, i.e., its activity and not its amount of substance concentration.

**Diluted samples:** In highly diluted samples, the activity coefficients are identical for the calibration solution and the analytical portion of the sample. The influence of water concentration and sodium binding is negligible. Accordingly, the concentration of sodium in the sample can be calculated from the concentration of the calibration solution. Thus, measurements by ISEs in diluted samples give comparable estimates of the amount of sodium in the

**Table 1** Reference intervals ( $\text{mmol l}^{-1}$ ) for electrolytes in serum (adults)

Calcium, total	2.15–2.55
Calcium, free ('ionized') (pH 7.4)	1.17–1.29
Chloride, total	98–106
Hydrogen carbonate (standard)	21–26
Magnesium, total	0.75–1.10
Phosphate (inorganic), total	0.87–1.45
Potassium, total	3.5–5.1
Sodium, total	135–145

The intervals are not method dependent, but are only valid for normal water concentration of the serum.

system, and accurate results of sodium concentration can be expected.

**Solid-phase chemistry** Potentiometry is also used for the determination of sodium concentration in solid-phase chemistry (Vitros, Ortho). The results usually compare well with measurements by FAES or ISEs in diluted samples, but differ in the case of paraproteinemia.

**Analytical reliability** A (between-day) coefficient of variation below 1.5% should be achieved in routine laboratory work with a bias of less than 2% from the target value to meet clinical requirements. All the methods are adequately sensitive and specific for physiological samples; however, some ISEs are less robust and more susceptible to interference than FAES.

### Significance and Interpretation of Results

Owing to the relatively small reference interval of sodium in serum (Table 1), a change in the water concentration has a large influence on the interpretation of sodium concentration. Sodium concentration decreases proportionally to the increase in protein and/or lipid concentration ('pseudohyponatremia'). Sodium concentration increases proportionally to the decrease in proteins. Therefore, before interpreting the results for sodium concentration in serum, one must be sure that the water concentration of the sample is not grossly aberrant from normal. Otherwise, the concentration has to be readjusted to be comparable to the reference interval. However, if measurement is performed by ISEs in the undiluted sample, the result can be interpreted without respect to water concentration. It reflects the sodium activity in the water phase, which is a more meaningful quantity in biological systems than concentration. The amount of sodium ions in the body is adjusted by the renin–angiotensin–aldosterone system, which influences the tubular reabsorption of sodium.

**Table 2** Disorders of sodium balance

1	Hyponatremia combined with hypotonic dehydration
1.1	Cause: Sodium loss exceeds water loss
	Examples: Polyuric renal disease
	Disturbance of tubular sodium reabsorption (e.g., interstitial nephritis)
	Adrenal cortical hypofunction
	Vomiting
	Diarrhea
	Diuretics
1.2	Cause: Sodium substitution (still) more insufficient than water substitution
	Example: Inadequate therapy of isotonic dehydration
2	Hyponatremia combined with hypotonic hyperhydration
2.1	Cause: Redundant administration of water
	Examples: Intravenous application of solutions with low sodium concentration (water intoxication)
2.2	Cause: Decreased water excretion
	Examples: Heart failure
	Liver cirrhosis
	Renal insufficiency
	Nephrotic syndrome
	Syndrome of inappropriate ADH secretion
3	Hypernatremia combined with hypertonic dehydration
3.1	Cause: Insufficient water input
	Example: Thirst
3.2	Cause: Excessive water loss
	Examples: Diabetes insipidus
	Diabetes mellitus
	Diarrhea
	Excessive sweating
4	Hypernatremia with hypertonic hyperhydration
4.1	Cause: Sodium input inadequately high
	Example: Intravenous application of solutions with high sodium concentration
4.2	Cause: Sodium elimination decreased
	Example: Hyperaldosteronism

Modified from Külpmann WR, Stummvoll HK, and Lehmann P (1996) *Electrolytes: Clinical and Laboratory Aspects*. New York: Springer.

Sodium concentration depends mainly on the elimination and reabsorption of free water by the kidney; therefore, sodium disorders are often connected with water dysregulation or reflect its presence (Table 2).

In isotonic dehydration and isotonic hyperhydration, the sodium concentration is 'normal', and these conditions cannot be detected by sodium concentration measurement.

The osmolarity of serum ( $O_{\text{est}}$ ) can be estimated according to

$$O_{\text{est}} = 1.86[\text{Na}] + [\text{glucose}] + [\text{urea}] + 9 \quad [4]$$

where square brackets indicate the concentration of the analyte in  $\text{mmol l}^{-1}$ .

The osmotic gap is defined as the difference between osmolality, as determined by freezing point

**Table 3** Reference intervals ( $\text{mmol d}^{-1}$ ) for electrolytes in urine (adults)

Calcium	2.50–8.00
Chloride	85–170
Magnesium	2.50–8.50
Phosphate (inorganic)	11–32
Potassium	35–80
Sodium	30–300

depression, and calculated osmolarity ( $O_{\text{estim}}$ ). The osmotic gap is increased in pseudohyponatremia as well as in intoxication by ethanol or other compounds, which may be present in high concentration but are not considered in the formula. Thus, the osmotic gap can give an indication of a undetected intoxication. Sodium is filtered through the renal glomeruli ( $25\,000\text{ mmol d}^{-1}$ ), but 99% of the filtered sodium ions are reabsorbed in the tubular system. Therefore, fractional sodium excretion  $\text{FE}_{\text{Na}}$  is 1%. In disturbances of the tubular apparatus, sodium reabsorption is decreased and  $\text{FE}_{\text{Na}}$  is increased. In the case of renal hypoperfusion however, tubular reabsorption is increased and  $\text{FE}_{\text{Na}}$  is decreased. In intensive-care patients, measurement of urinary excretion of sodium (Table 3) may be necessary to monitor infusion therapy.

## Chloride

### Methods and Techniques

The favored method for the determination of chloride in serum and urine is potentiometry. The use of other techniques such as coulometry and UV-visible spectrophotometry is decreasing steadily. The determination of 'ionized' (free) chloride in serum or blood (extracellular water phase) becomes more popular by the introduction of combined blood gas/electrolyte analyzers.

**Coulometry** The basic information on the principle and practice of coulometry is presented elsewhere in this encyclopedia. In clinical chemistry, coulometry is exclusively used for the determination of chloride concentration. For this application coulometry is sufficiently specific, because the concentration of other halides is usually very low in comparison to the chloride concentration ( $\sim 100\text{ mmol l}^{-1}$  in serum). However, in case of long-lasting abuse of, e.g., hypnotics, which contain bromine (carbromal, bromisoval), bromide concentrations higher than  $5\text{ mmol l}^{-1}$  can be observed (bromide is eliminated very slowly (elimination half-life  $\sim 300\text{ h}$ ) and will accumulate). The bromide concentration will add



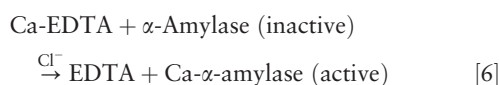
spuriously to the chloride concentration. Coulometry was considered to be adequately reliable for use as a reference method, albeit in a more elaborate procedure than that used in routine work. Unfortunately, because it is usually not part of a multichannel analyzer and highly mechanized coulometers are not available, coulometry can only be used for the measurement of a small number of samples.

**UV-visible spectrophotometry** (1) In this technique, the determination of chloride is performed mainly through the following reaction:



Thiocyanate ions are then detected by the addition of  $\text{Fe}^{3+}$ . The method can easily be mechanized for measurements on a large scale, but is subject to interference by bromide ions more than is coulometry. This method is no longer recommended, owing to the environmental hazards resulting from the mercury in the reagent.

(2) More recently, a method that uses the activation of  $\alpha$ -amylase by chloride has been proposed:



$\alpha$ -Amylase activity, which is proportional to chloride concentration, is determined by measuring the release of 2-chloro-4-nitrophenol from 2-chloro-4-nitrophenyl- $\beta$ ,D-maltoheptaoside. The method can be adapted to clinical chemistry analyzers, but is also susceptible to interference by bromide.

**Potentiometry** Silver chloride electrodes as well as membrane electrodes with ion exchange or ion carrier are used for chloride determinations. Measurements are usually performed with a highly diluted sample to obtain its chloride concentration. The measurement of chloride activity in undiluted samples is used more and more often, as it is easily performed by combined blood gas/electrolyte analyzers.

In solid-phase chemistry, chloride is also determined by potentiometry (Vitros, Ortho).

**Analytical reliability** The coefficient of variation (between days) should be less than 2% and the deviation from the target value should be less than 4% in a routine clinical laboratory to meet clinical requirements.

## Significance and Interpretation of Results

Chloride concentration measurement of serum (Table 1) is meaningful in the case of normal water concentration of the sample. In hyperlipemia or hyperproteinemia, 'pseudohypochloremia' occurs, because an increase of, e.g., 10% in the macromolecule-containing compartment decreases chloride concentration by 10%. A shift of 10% is especially important when the reference interval is relatively small. In the case of chloride, the upper limit of the reference interval differs from the lower limit by 11% (see Sodium above). The chloride concentration of serum therefore has to be readjusted (see Significance and Interpretation of Results above) when the water concentration is abnormal to be comparable with the reference interval.

The anion gap (AG) ( $\text{mmol l}^{-1}$ ) is calculated as

$$\text{AG} = [\text{Na}^+] - [\text{Cl}^-] - [\text{HCO}_3^-] \quad [7]$$

where  $[\text{Na}^+]$  is the sodium concentration in serum ( $\text{mmol l}^{-1}$ ),  $[\text{Cl}^-]$  is the chloride concentration in serum ( $\text{mmol l}^{-1}$ ), and  $[\text{HCO}_3^-]$  is the hydrogen carbonate concentration in serum ( $\text{mmol l}^{-1}$ ).

The reference interval of the anion gap is  $+8$  to  $+16 \text{ mmol l}^{-1}$ . The gap is due to the fact that the sum of the cations (e.g.,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ ), which is not considered, is lower than the sum of the anions, which are excluded from the calculation (proteins, organic acids). The anion gap is enlarged in the cases of diabetic ketoacidosis, lactacidosis, and certain intoxications with salicylate, methanol (formate), ethanol (lactate), and ethylene glycol (oxalate, glycolate, glyoxylate). The anion gap is decreased or even becomes negative in the case of high bromide concentrations, especially if chloride concentration is determined by a method that is more sensitive to bromide than to chloride.

Sodium concentration and chloride concentration generally change in the same direction and to the same extent, whereas chloride and hydrogen carbonate change reciprocally (Table 4).

Urinary excretion of chloride (Table 3) usually resembles sodium excretion. In the case of metabolic alkalosis due to extrarenal loss of chloride, the excretion of chloride is  $<10 \text{ mmol d}^{-1}$ . If metabolic alkalosis is caused by mineralocorticoid excess, chloride excretion is unaffected and  $>10 \text{ mmol d}^{-1}$ .

## Potassium

### Methods and Techniques

Potassium in serum or urine may be determined by FAES, absorption spectrometry, or potentiometry

**Table 4** Disorders of chloride balance

1	Hypochloremia
1.1	Cause: Insufficient intake of chloride
1.2	Cause: Loss of chloride
	Examples: Vomiting
	Gastric lavage
	Metabolic alkalosis
	Respiratory acidosis
	Diuretics
1.3	Cause: Hypotonic dehydration (see Sodium)
1.4	Cause: Hypotonic hyperhydration (see Sodium)
2	Hyperchloremia
2.1	Cause: Excessive input of chloride
	Examples: Infusion of high amounts of chloride
	Uretersigmoidostomie
2.2	Cause: Decreased elimination of chloride (due to hydrogen carbonate loss)
	Examples: Diarrhea
	Pancreatic fistula
	Renal tubular acidosis
	Hypoaldosteronism with acidosis
2.3	Cause: Hypertonic dehydration (see Sodium)
2.4	Cause: Hypertonic hyperhydration (see Sodium)

Modified from Külpmann WR, Stummvoll HK, and Lehmann P (1996) *Electrolytes: Clinical and Laboratory Aspects*. New York: Springer.

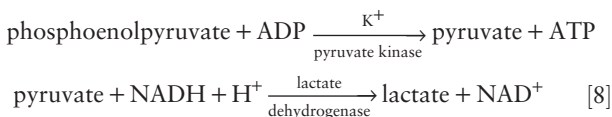
(ISEs), which is most often used nowadays. Blood itself can only be used for analysis with potentiometry.

**Flame atomic emission spectrometry** Potassium concentration is determined at a wavelength of 767 nm using a propane flame. Procedures that take advantage of the lithium guideline technique are preferred (see Flame Atomic Emission Spectrometry above).

Good precision, accuracy, and specificity of FAES provide the reason for its choice as the basic principle of the potassium reference method. In routine work however, the use of FAES is steadily decreasing.

### UV-visible spectrophotometry

1. *Enzymatic determination:* In 1989, an 'enzymatic' method for the determination of potassium was proposed:



Potassium ions activate pyruvate kinase; its activity therefore reflects the potassium concentration. Interference by sodium is eliminated by binding to a kryptand, and that from ammonium ions by the addition of glutamate dehydrogenase.

2. *Kryptahemispherand determination:* The trinitroaniline-kryptahemispherand is used for the

measurement of potassium. This consists of an ionophor with a cavity fitting to the potassium ion and an attached chromophor. The absorption spectrum shifts on potassium binding. However, the method is subject to interference by lipids and bilirubin, which can be compensated for by the use of a sample blank.

Both the above procedures can be adapted to chemical analyzers.

**Potentiometry** The favored ISE for potassium determination uses valinomycin, an ion carrier of natural origin, in its ion-selective membrane. In undiluted samples an estimate of the concentration in the water phase is made, whereas in highly diluted samples the results reflect the concentration of the total sample. In hyperlipemia and hyperproteinemia or in hypoproteinemia, the results from undiluted samples will differ from diluted samples to the same extent as described above for sodium. As the 'relative' reference interval of potassium is much larger than that of sodium (the upper limit of the reference interval of potassium differs from the lower limit by 50%, and in the case of sodium by 7%), the effects are clinically less meaningful.

**Solid-phase chemistry** Potentiometry is used for potassium measurement also in solid-phase chemistry. Potassium is bound to valinomycin, releasing a proton that causes a color change in an indicator dye. In paraproteinemia, deviations from FAES values will occur.

**Analytical reliability** In routine work, the between-days coefficient of variation should be lower than 2.7%, with the deviation from the target value no more than 3.7% to meet clinical requirements.

### Significance and Interpretation of Results

The concentration of potassium in erythrocytes is 25-fold greater than in plasma. The analysis of hemolyzed samples therefore does not yield clinically meaningful results. Generally, any process that allows potassium to leave erythrocytes or thrombocytes, such as blood storage with or without visible hemolysis (in a refrigerator, no haemolysis), will falsify the potassium concentration. Even clotting of blood will increase potassium concentration; hence, its concentration is  $\sim 0.4 \text{ mmol l}^{-1}$  higher in serum (Table 1) than in plasma. Thrombocytosis or chronic myelosis may cause 'pseudohyperkalemia' as well as *in vivo* haemolysis. For these reasons, heparinized plasma is preferable to serum, as it decreases potassium release from cells.



**Table 5** Disorders of potassium balance

1	Hypokalemia
1.1	Cause: Insufficient input of potassium Examples: Chronic alcoholism Anorexia nervosa Infusion of potassium-depleted solution
1.2	Cause: Increased loss of potassium Examples: <i>Hormonally induced:</i> Mineralocorticoid excess <i>Renal loss:</i> Renal tubular acidosis Diuretics <i>Gastrointestinal loss:</i> Vomiting Diarrhea Abuse of laxatives
1.3	Cause: Shift from extra- to intracellular compartment Examples: Alkalosis Insulin
2	Hyperkalemia
2.1	Cause: Abundant potassium input Example: Infusion of excessive amounts of potassium Blood transfusion Potassium-containing drugs (e.g., potassium-penicillin)
2.2	Cause: Decreased potassium excretion Examples: <i>Hormonally induced:</i> Adrenal cortical hypofunction <i>Disturbed renal function:</i> Acute renal failure (oliguria) Chronic renal insufficiency Diuretics
2.3	Cause: Shift from intra- to extracellular compartment Example: Acidosis Lysis of cells (rhabdomyolysis, burning, treatment with cytostatics) Hyperthermia

Modified from Külpmann WR, Stummvoll HK, and Lehmann P (1996) *Electrolytes: Clinical and Laboratory Aspects*. New York: Springer.

Potassium concentration in plasma is dependent on renal function and acid–base balance. Mineralocorticoids promote the renal excretion of potassium, and insulin favors cellular uptake (Table 5).

Potassium is filtered through the glomeruli and is almost totally reabsorbed in the proximal tubules and Henle's loop. Therefore, potassium in urine (Table 3) stems from the secretion of the distal tubules and collecting ducts.

In hypokalemia due to renal loss, the urinary excretion of potassium is elevated; if the condition is due to extrarenal loss, elimination of potassium by urine is decreased ( $<10 \text{ mmol d}^{-1}$ ).

## Calcium

### Methods and Techniques

Calcium is determined in physiological samples by flame atomic absorption spectrometry (FAAS), FAES,

UV–visible spectrophotometry, and potentiometry. The last is primarily used for the measurement of 'ionized' (free) calcium ions in the (extracellular) water phase of blood or serum. The other techniques are used for the determination of total calcium in serum and urine. Urine must be acidified (pH 1) to release calcium ions from stable salts in the sediment.

**Flame atomic absorption spectrometry** FAAS measurements are performed at the calcium resonance line at 422.7 nm using an air–acetylene flame.  $\text{SrCl}_2$  may be used as the internal standard. Before analysis, samples are diluted and the viscosity of the analytical portion and of the calibrator solution is adjusted. Phosphate ions form thermally stable calcium salts, which then escape atomization and measurement. For this reason,  $\text{LaCl}_3$  is added, which binds phosphate ions enabling accurate calcium measurements independent of the (varying) phosphate concentration. FAAS is considered to be the most reliable technique for total calcium determination, and therefore it has been chosen as the reference method.

**Flame atomic emission spectrometry** Although FAES is not as reliable as FAAS, it was widely used because it allows the simultaneous mechanized determination of sodium, potassium, and calcium with adequate analytical performance for routine laboratory work. As calcium is not as easily excitable as the alkali metals, an air–acetylene flame ( $2325^\circ\text{C}$ ) instead of an air–propane flame must be used, which is less appropriate for the measurement of sodium and potassium. The flame spectrum of calcium has emissions at 422.7, 554, and 622 nm, the first being an arc line and the others being molecular bands. Usually 622 nm ( $\text{CaO}$ ) is chosen for measurement. Errors from anionic interference due to phosphate salts, which are not easily volatilized, can be excluded by the addition of phosphate ions in excess. For further details, see Sodium above.

**UV–visible spectrophotometry** Calcium ions form a violet-colored complex with *o*-cresolphthalein. Its absorbance is proportional to the calcium concentration of the sample and can be measured at 546 nm. Protein-bound calcium is released by hydrochloric acid. The interference of magnesium ions is excluded by the addition of 8-hydroxyquinoline. The method is in widespread use on mechanized analysers for clinical chemistry.

**Potentiometry** Potentiometry by ISEs is primarily used for the measurement of 'ionized' (free) calcium ions and not of total calcium concentration. The ion-selective membrane of the electrode contains an ion

carrier, e.g., ETH 1001. Sodium chloride is added to the calcium chloride of the calibration solution to adjust the ionic strength to  $160 \text{ mmol kg}^{-1}$ , which is equivalent to the mean ionic strength of serum. This procedure allows one to minimize the error, which stems from measuring activity and reporting concentration. Any factor that has an impact on activity and is different in the analytical portion from that in the calibration solution will increase the difference between the reported and the true concentration of free ions.

In all, 40% of calcium ions are bound to protein at pH 7.4 (Figure 2).

In alkalosis the bound fraction increases, and in acidosis the bound fraction decreases. If the concentration of free calcium ions at the actual pH is to be determined, blood samples have to be processed anaerobically and measured within minutes. Results are reported for the actual pH and for pH 7.4 after recalculating according to a suitable algorithm. If the pH is considered to be 'normal', serum can be used for measurement after readjustment of pH to  $7.4 \pm 0.2$  and recalculating for pH 7.4. Reporting at pH 7.4 is performed to allow for comparison with the reference interval, which is only known for this pH value. Determination of 'ionized' calcium by ISEs is preferred to calculation of its concentration from the values of total calcium, albumin, and pH using an empirical nomogram.

**Solid-phase chemistry** After application of the sample to the slide, calcium is released from protein

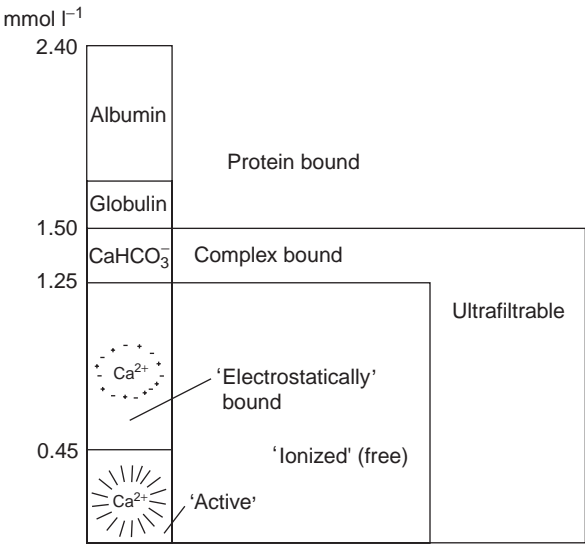
binding and penetrates the various layers to react finally with Arsenazo III, an indicator dye. The analytical performance of this solid-phase method is satisfactory, although systematic deviations from the reference method values have been observed with control sera (probably owing to special properties of their matrix).

**Analytical reliability** In routine clinical laboratories, imprecision between days for total calcium concentration measurements should be below a coefficient of variation of 3.0%, and deviation from the target value should not exceed 5% to meet clinical requirements.

**Significance and Interpretation of Results**

In all, 40% of calcium in serum is bound to protein and 10% to anions like citrate, hydrogen carbonate, etc.; 50% is 'ionized' (free) and only this fraction influences nervous excitability. Parathyroid hormone and calcitriol increase calcium concentration (Table 1); their antagonist is calcitonin. Input of calcium is by food, and output by incomplete tubular reabsorption of the glomerular filtrate and intestinal excretion (Table 6).

The determination of calcium in urine (Table 3) is usually of minor importance. In nephrolithiasis, the renal excretion of calcium is elevated in about one-third of all patients.



**Figure 2** Fractions of calcium in serum. (Modified from Siggaard-Andersen O, Thode J, and Fogh-Andersen N (1983) What is 'ionized calcium'? *Scandinavian Journal of Clinical Laboratory Investigation* 43: (Suppl. 165) 11–16.)

**Table 6** Disorders of calcium balance

1	Hypocalcemia
1.1	Cause: Insufficient input Examples: Calcium-depleted diet Reduced intestinal absorption (e.g., pancreatitis) Calcitriol deficiency
1.2	Cause: Excessive calcium loss Examples: Chronic renal failure Diuretics
1.3	Cause: Calcium binding exceeds calcium release Examples: Hypoparathyroidism Osteoblastic metastases
2	Hypercalcemia
2.1	Cause: Excessive calcium input Examples: Milk-Alkali syndrome Vitamin D intoxication
2.2	Cause: Insufficient calcium elimination Example: Diuretics
2.3	Cause: Calcium release exceeds calcium binding Examples: Hyperparathyroidism Osteolytic metastases Immobilization

Modified from Kùlpmann WR, Stummvoll HK, and Lehmann P (1996) *Electrolytes: Clinical and Laboratory Aspects*. New York: Springer.

## Magnesium

### Methods and Techniques

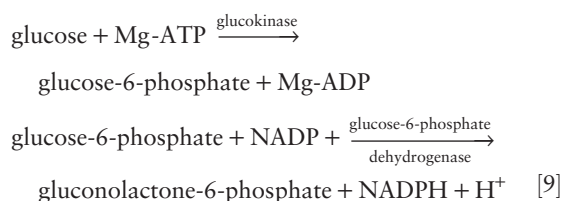
Magnesium concentration can be determined in physiological samples such as serum and urine by FAAS, FAES, and UV-visible spectrophotometry. Urine must be acidified (pH 1.0) by hydrochloric acid to release magnesium from salts of the sediment. Recently, potentiometry has become available for the measurement of 'ionized' (free) magnesium in serum and blood.

**Flame atomic absorption spectrometry** The determination of magnesium by FAAS is performed in diluted samples using the resonance line 285.2 nm and a stoichiometric air-acetylene flame. The interference due to phosphate is eliminated by the addition of  $\text{LaCl}_3$ . FAAS is used for the most reliable determination of total magnesium in physiological samples and has been chosen as the basis for the reference method.

**Flame atomic emission spectrometry** Magnesium has a flame spectrum with a band emission with peaks at 370 nm and 383 nm and an arc line at 285.2 nm. Wavelengths of 370 nm or 383 nm have been proposed for FAES, but the method is only rarely applied.

### UV-visible spectrophotometry

1. *Xylidyl blue (Magon)*: Magnesium forms a colored complex with sulfonated Magon in a strongly basic medium. The absorbance maximum of the complex is at 548 nm. Calcium interference is avoided by the addition of ethylene-bis-(oxyethylene-nitrilo)-tetraacetic acid (EGTA) and interference by heavy metals by using cyanide.
2. *Other methods*: Calmagite, methylthymol blue, and Titan yellow are also used for the measurement of magnesium.
3. *Enzymatic determination*: The activity of glucokinase is dependent on the concentration of the Mg-ATP complex:



This method is not subject to interference by calcium ions. It is not widely applied because the reagents are rather expensive.

**Potentiometry** ISEs for the determination of 'ionized' (free) magnesium are now available. They use membranes with neutral ion carriers (e.g., modified ETH 5220), which, however, lack specificity with respect to calcium. Therefore, magnesium and calcium must be measured simultaneously in the same analytical portion (with different electrodes) and magnesium concentration is calculated from this.

**Solid-phase chemistry** Magnesium is released from protein, and penetrates the layers of the slide to react with a formazan dye. Calcium is bound to a chelating agent to prevent interference. Accuracy was less satisfactory with control sera than with native sera.

**Analytical reliability** The coefficient of variation for imprecision between days should be below 4% in the clinical laboratory, and the deviation from the target value should be below 7% to meet clinical requirements.

### Significance and Interpretation of Results

About 20–35% of magnesium is bound to proteins and other anions at pH 7.4, less in acidosis and more in alkalosis as discussed for calcium. As a variable amount of magnesium is bound, total magnesium concentration (Table 1) is only a poor estimate of free magnesium concentration, which is considered to be the biologically active fraction.

Magnesium is ingested with food, although only 50% is absorbed by the intestine. In the kidney, magnesium is filtered through the glomeruli and 95% is reabsorbed by the tubular apparatus. Increased secretion of mineralocorticoids and glucocorticoids promotes urinary excretion, owing to reduced tubular absorption (Table 7).

It is assumed that in European countries, the magnesium concentration of 'healthy' persons (reference interval) is lower than it should be because people at risk of myocardial infarction and angiopathy benefit from magnesium-supplementation to their diet. This low reference interval of magnesium concentration in serum is said to stem from magnesium-deficient soil. It is taken that the magnesium concentration should exceed  $0.8 \text{ mmol l}^{-1}$ .

The determination of magnesium in urine (Table 3) is helpful in the investigation of hypomagnesemia. If the urinary excretion is increased, hypomagnesemia is probably due to renal insufficiency. However, if the excretion is decreased, an extrarenal cause

**Table 7** Disorders of magnesium balance

1	Hypomagnesemia
1.1	Cause: Insufficient intake
	Examples: Malabsorption
	Malabsorption
	Chronic alcoholism
	Magnesium-depleted infusions
1.2	Cause: Excessive magnesium loss
	Examples: Renal insufficiency with polyuria
	Hyperaldosteronism
	Diuretics
1.3	Cause: Shift from extra- to intracellular compartment
	Examples: Treatment of diabetic coma
	Hungry bone with increased magnesium binding
2	Hypermagnesemia
2.1	Cause: Excessive intake
	Examples: Infusion of excessive amounts of magnesium
	Administration of magnesium-containing antacids
2.2	Cause: Reduced renal excretion
	Examples: Renal insufficiency with oliguria
	Hypoaldosteronism
2.3	Cause: Shift from intra- to extracellular compartment
	Examples: Rhabdomyolysis
	Diabetic coma

Modified from Külpmann WR, Stummvoll KH, and Lehmann P (1996) *Electrolytes: Clinical and Laboratory Aspects*. New York: Springer.

must be assumed. Furthermore, a magnesium deficiency of the body, which at first may be accompanied by normomagnesemia, can be detected by a decreased excretion of magnesium in urine.

## Inorganic Phosphate

### Methods and Techniques

Inorganic phosphate concentration is determined, after an appropriate reaction, by UV–visible spectrophotometry. The methods are applicable to serum and urine, but not to blood.

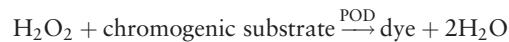
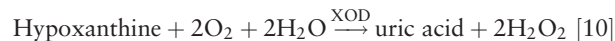
#### UV–visible spectrophotometry

1. *Molybdenum blue*: Phosphate ions react with molybdate to form ammonium phosphomolybdate, which is reduced to convert it to molybdenum blue. Tin(II) chloride and ammonium iron(II) sulfate are used as reducing agents.

2. *Phosphorus–molybdate complex*: On reaction of phosphate ions with molybdate, the phosphorus–molybdate complex is formed. Its concentration is determined directly (without reduction) at 340 nm.

3. *Molybdivanadophosphate*: Phosphate concentration can be determined by reaction of the phosphate ions with vanadium molybdenum heteropolyacid (without reduction).

4. *Enzymatic determination*: Inosine and phosphate form hypoxanthine and ribose-1-phosphate in the presence of purine nucleoside phosphorylase. The amount of hypoxanthine is proportional to the phosphate concentration of the sample:



where PNP is purine nucleoside phosphorylase, XOD is xanthine oxidase, and POD is peroxidase.

**Solid-phase chemistry** Phosphate determination is performed by the use of the molybdenum blue complex.

**Analytical reliability** In routine clinical chemistry, imprecision between days should be below 5% (coefficient of variation), and the deviation from the target value should be below 8% to meet clinical requirements. Results are very high with some non-enzymatic methods owing to hydrolysis of organic phosphate esters.

### Significance and Interpretation of Results

After intake, phosphate (Table 1) is incompletely absorbed in the intestine. It is filtered through the glomeruli and 85% of the filtrate is reabsorbed in the tubular apparatus. Furthermore, phosphate is secreted into the intestine and is excreted with the feces. Parathyroid hormone increases the renal elimination of phosphate, whereas calcitriol, somatotropin, and insulin inhibit urinary excretion by increasing tubular reabsorption (Table 8).

Urinary excretion (Table 3) of phosphate is dependent on (1) the amount that is filtered through the glomeruli and the fraction thereof that is reabsorbed in the (proximal) tubular system, (2) the intake of phosphate by food, and (3) the metabolism of bone.

This is why the determination of phosphate excretion in daily urine is clinically not very meaningful and should be supplemented by measurements of (1) phosphate clearance, (2) fractional tubular phosphate reabsorption, and (3) maximum tubular phosphate reabsorption.

## Hydrogen Carbonate or Total CO<sub>2</sub>

### Methods and Techniques

Total CO<sub>2</sub> or hydrogen carbonate (bicarbonate) is determined in serum by UV–visible spectrophotometry

**Table 8** Disorders of phosphate balance

1	Hypophosphatemia
1.1	Cause: Insufficient intake
	Examples: Starvation
	Phosphate-depleted food
	Malabsorption
	Calcitriol deficiency
	Phosphate-binding antacids
1.2	Cause: Increased renal excretion
	Examples: Tubular acidosis
	Fanconi's syndrome
	Hyperparathyroidism
	Diuretics
1.3	Cause: Shift from extra- to intracellular compartment
	Example: Burns
2	Hyperphosphatemia
2.1	Cause: Excessive intake
	Example: Milk-Alkali syndrome
2.2	Cause: Decreased excretion
	Examples: Renal insufficiency
	Hypoparathyroidism
	Vitamin D intoxication
2.3	Cause: Release from bony tissue
	Examples: Osteolytic metastases
	Extensive bone surgery

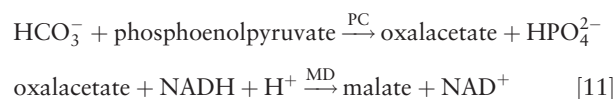
Modified from K lpmann WR, Stummvoll KH, and Lehmann P (1996) *Electrolytes: Clinical and Laboratory Aspects*. New York: Springer.

or potentiometry. In urine, these analytes are measured by titrimetry or manometrically. They can be calculated for the extracellular water phase of blood by using of quantities obtained in blood gas analysis.

### UV-visible spectrophotometry

1. In continuous-flow analysis, CO<sub>2</sub> gas is released by acidification of the analytical portion. It diffuses across a silicon-rubber membrane into an alkaline hydrogen carbonate buffer. The gas is converted into hydrogen carbonate and hydrogen ions; this causes a pH change, which is monitored by a pH indicating dye. The color change of the dye is measured by spectrophotometry and is proportional to total CO<sub>2</sub> of the sample.

2. In an enzymatic method, all CO<sub>2</sub> forms are converted to HCO<sub>3</sub><sup>-</sup> by the addition of alkali, which reacts with phosphoenolpyruvate:



where PC is phosphoenolpyruvate carboxylase and MD is malate dehydrogenase.

**Potentiometry** For the determination of hydrogen carbonate, lactic acid is added to release CO<sub>2</sub> from HCO<sub>3</sub><sup>-</sup>; CO<sub>2</sub> is measured by a pCO<sub>2</sub> electrode.

**Solid-phase chemistry** A potentiometric test is available for the determination of total CO<sub>2</sub> in serum by the use of ISEs sensitive to the carbonate ion.

**Calculation** Actual bicarbonate and total CO<sub>2</sub> are usually calculated in clinical chemistry from blood gas analysis.

*Actual bicarbonate:*

$$\log \text{HCO}_3^- = \text{pH} - \text{pK}' + \log p\text{CO}_2 + \log \alpha_{\text{CO}_2} \quad [12]$$

HCO<sub>3</sub><sup>-</sup>, hydrogen carbonate concentration in blood (mmol l<sup>-1</sup>); pK', apparent pK of CO<sub>2</sub> in blood (6.095); pCO<sub>2</sub>, partial pressure of CO<sub>2</sub>; α<sub>CO<sub>2</sub></sub>, molar solubility coefficient of CO<sub>2</sub>.

*Total CO<sub>2</sub>:*

$$\text{CO}_2 = \alpha_{\text{CO}_2} \cdot p\text{CO}_2 + [\text{HCO}_3^-] \quad [13]$$

α<sub>CO<sub>2</sub></sub>, Bunsen coefficient (0.0307); pCO<sub>2</sub>, partial pressure of carbon dioxide (mmHg); [HCO<sub>3</sub><sup>-</sup>], hydrogen carbonate concentration (mmol l<sup>-1</sup>).

**Analytical reliability** It has been proposed that deviations from the target value must be lower than the following: pH: ±0.03; pCO<sub>2</sub>: ±3 mmHg (or ±7.5%) (College of American Pathologists).

### Significance and Interpretation of Results

It must be ascertained that serum was obtained from blood that was processed anaerobically during sampling and centrifuging. Total carbon dioxide comprises dissolved CO<sub>2</sub> (3%), hydrogen carbonate (64%), and carbamino derivatives of plasma proteins (33%). Hydrogen carbonate is increased in metabolic alkalosis and decreased in acute and chronic acidosis. Hydrogen carbonate concentration (Table 1) is needed for the calculation of the anion gap (see Chloride above).

Hydrogen carbonate in urine is determined in addition to urinary titratable acid and ammonium for the estimation of the net excretion of hydrogen ions:

$$[\text{H}^+] = [\text{TA}] + [\text{NH}_4^+] - [\text{HCO}_3^-] \quad [14]$$

where TA is titratable acid.

The [H<sup>+</sup>] excretion is decreased in chronic renal insufficiency.



## Emerging Needs

Techniques that are currently available for the measurement of physiological samples are sufficiently reliable for routine patient care. However, they are not yet adequately precise to monitor intraindividual changes. For this purpose, improvement of accuracy must be achieved.

To alleviate medical interpretation, it is hoped that the measurement of the 'ionized' (i.e., free) fraction of the electrolytes by ISEs in undiluted samples will become the widespread technique. At the moment, the use of these methods is restricted because they are still less practical (with regard to sample volume and throughput) than traditional procedures, which measure total concentration. The improvement of practicability will anyhow be the focus of future developments to allow quick measurements at all times at low costs.

The *in vivo* monitoring of electrolytes in blood by ISEs may be achievable in the near future. Although this is an interesting new perspective for research, an urgent and widespread need for use in patient care is not yet evident.

The measurement of electrolytes in blood cells is under investigation, but on the whole the results are not too promising. There is evidence that the serum concentration of some electrolytes poorly reflects the electrolyte concentration in the tissues. Therefore, devices are that enable the measurement of electrolytes *in vivo* in muscle or bone are required.

*See also:* **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Principles and Instrumentation. **Ion-Selective**

**Electrodes:** Clinical Applications. **Quality Assurance:** Clinical Applications. **Spectrophotometry:** Overview.

## Further Reading

- Burnett RW, Covington AK, Fogh-Andersen N, *et al.* (2000) Recommendations for measurement of and conventions for reporting sodium and potassium by ion-selective electrodes in undiluted serum, plasma or whole blood. *Clinical Chemistry and Laboratory Medicine* 38: 1065–1071.
- Burnett RW, Covington AK, Fogh-Andersen N, *et al.* (2000) IFCC recommended reference method for the determination of the substance concentration of ionized calcium in undiluted serum, plasma or whole blood. *Clinical Chemistry and Laboratory Medicine* 38: 1301–1314.
- Burnett RW, Covington AK, Fogh-Andersen N, *et al.* (2001) IFCC reference measurement procedure for the substance concentration determination of total carbon dioxide in blood, plasma or serum. *Clinical Chemistry and Laboratory Medicine* 39: 283–289.
- Faulkner WF and Meites S (1982) Selected methods for the small clinical chemistry laboratory. *American Association of Clinical Chemistry*, vol. 9, p. 398.
- Henry RJ, Cannon DC, and Winkelman JW (1974) *Clinical Chemistry*, 2nd edn. Hagerstown: Harper & Row.
- Kaplan LA and Pesce AJ (1989) *Clinical Chemistry*. St. Louis: Mosby Company.
- Külpmann WR (1992) Determination of electrolytes in serum and serum water. *Wiener Klinische Wochenschrift* 104(Suppl. 192): 34–38.
- Waugh WH (1969) Utility of expressing serum sodium per unit of water in assessing hyponatremia. *Metabolism* 18: 706–712.

## Glucose

**H M Heise**, ISAS – Institute for Analytical Sciences, Dortmund, Germany

**P U Abel**, Forschungszentrum Sensorik Greifswald e.V., Greifswald, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

An important parameter in biochemistry and medical diagnostics is the concentration of glucose in biofluids, which is most frequently determined in the

clinical laboratory. For routine analysis the established methodology uses different enzymatic assays involving photometric or electrochemical detection. New trends strive for reagentless assays using biosensor technology. Another important research field concerns the development of self-monitoring devices for blood glucose, necessary for patients suffering from diabetes mellitus.

D-Glucose is by far the most abundant monosaccharide in physiological samples. It has two stereoisomers designated the  $\alpha$ - and  $\beta$ -anomeric forms. In aqueous solution each form changes slowly, by

## Emerging Needs

Techniques that are currently available for the measurement of physiological samples are sufficiently reliable for routine patient care. However, they are not yet adequately precise to monitor intraindividual changes. For this purpose, improvement of accuracy must be achieved.

To alleviate medical interpretation, it is hoped that the measurement of the 'ionized' (i.e., free) fraction of the electrolytes by ISEs in undiluted samples will become the widespread technique. At the moment, the use of these methods is restricted because they are still less practical (with regard to sample volume and throughput) than traditional procedures, which measure total concentration. The improvement of practicability will anyhow be the focus of future developments to allow quick measurements at all times at low costs.

The *in vivo* monitoring of electrolytes in blood by ISEs may be achievable in the near future. Although this is an interesting new perspective for research, an urgent and widespread need for use in patient care is not yet evident.

The measurement of electrolytes in blood cells is under investigation, but on the whole the results are not too promising. There is evidence that the serum concentration of some electrolytes poorly reflects the electrolyte concentration in the tissues. Therefore, devices are that enable the measurement of electrolytes *in vivo* in muscle or bone are required.

*See also:* **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Principles and Instrumentation. **Ion-Selective**

**Electrodes:** Clinical Applications. **Quality Assurance:** Clinical Applications. **Spectrophotometry:** Overview.

## Further Reading

- Burnett RW, Covington AK, Fogh-Andersen N, *et al.* (2000) Recommendations for measurement of and conventions for reporting sodium and potassium by ion-selective electrodes in undiluted serum, plasma or whole blood. *Clinical Chemistry and Laboratory Medicine* 38: 1065–1071.
- Burnett RW, Covington AK, Fogh-Andersen N, *et al.* (2000) IFCC recommended reference method for the determination of the substance concentration of ionized calcium in undiluted serum, plasma or whole blood. *Clinical Chemistry and Laboratory Medicine* 38: 1301–1314.
- Burnett RW, Covington AK, Fogh-Andersen N, *et al.* (2001) IFCC reference measurement procedure for the substance concentration determination of total carbon dioxide in blood, plasma or serum. *Clinical Chemistry and Laboratory Medicine* 39: 283–289.
- Faulkner WF and Meites S (1982) Selected methods for the small clinical chemistry laboratory. *American Association of Clinical Chemistry*, vol. 9, p. 398.
- Henry RJ, Cannon DC, and Winkelman JW (1974) *Clinical Chemistry*, 2nd edn. Hagerstown: Harper & Row.
- Kaplan LA and Pesce AJ (1989) *Clinical Chemistry*. St. Louis: Mosby Company.
- Külpmann WR (1992) Determination of electrolytes in serum and serum water. *Wiener Klinische Wochenschrift* 104(Suppl. 192): 34–38.
- Waugh WH (1969) Utility of expressing serum sodium per unit of water in assessing hyponatremia. *Metabolism* 18: 706–712.

## Glucose

**H M Heise**, ISAS – Institute for Analytical Sciences, Dortmund, Germany

**P U Abel**, Forschungszentrum Sensorik Greifswald e.V., Greifswald, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

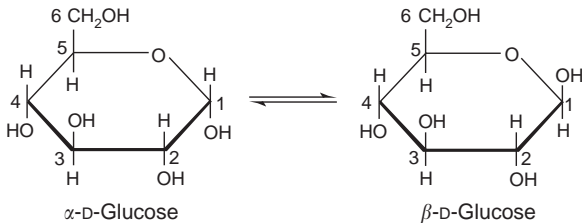
An important parameter in biochemistry and medical diagnostics is the concentration of glucose in biofluids, which is most frequently determined in the

clinical laboratory. For routine analysis the established methodology uses different enzymatic assays involving photometric or electrochemical detection. New trends strive for reagentless assays using biosensor technology. Another important research field concerns the development of self-monitoring devices for blood glucose, necessary for patients suffering from diabetes mellitus.

D-Glucose is by far the most abundant monosaccharide in physiological samples. It has two stereoisomers designated the  $\alpha$ - and  $\beta$ -anomeric forms. In aqueous solution each form changes slowly, by



means of the free aldehyde form, into an equilibrium mixture of approximately one-third  $\alpha$ -D- and two-thirds  $\beta$ -D-glucose (mutarotation). This is detectable as a change in the optical rotation due to the different optical activities of the two anomeric forms. These anomers may exhibit dissimilar properties in enzymatic reactions:



Since blood serves as the primary metabolic transport system between the body organs, its composition is the preferred indicator with regard to the pathophysiological condition of the patient. This is the reason why most determinations are concerned with whole blood or biofluids derived from it. Glucose in urine is also an index component of metabolic disorders in patients. Other biological fluids have been considered as an alternative assay for blood glucose.

The aldohexose glucose is a substance of high metabolic importance, since it can be considered as the main energy carrier in the human organism. Regulation of its concentration is extremely crucial because the metabolism of some organs, such as brain and kidney, and also erythrocytes, depend on it. Under physiological conditions the glucose concentration in mammalian blood varies as a result of nutritional and metabolic activities. There are, however, endocrine feedback mechanisms in the body to regulate the blood glucose concentration. An important role in lowering the glucose level is played by the insulin secreted by the  $\beta$ -cells of the Langerhans islet cells in the pancreas. In addition to insulin, the hormones glucagon, epinephrine, and cortisol take part in regulating the blood glucose concentration.

The quantification of glucose is a prerequisite for the diagnosis and therapy of patients suffering from disorders in their carbohydrate metabolism that are mainly caused by diabetes mellitus. In type I diabetes, the  $\beta$ -cells have been destroyed, leading to a severe insulin deficiency. In type II diabetes, enough insulin may be available, but there is an insulin resistance in the target organs.

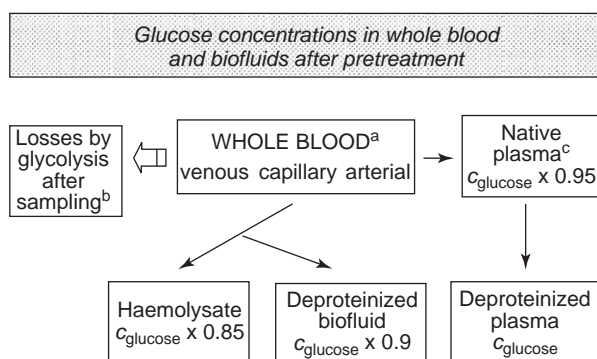
In therapy, the aim is always towards a blood glucose concentration comparable with that of metabolically healthy individuals (normoglycemia). For type I diabetics the external administration of insulin

is required, as insulin deficiency can lead to severe hyperglycemia, ketosis, coma, and death. However, the application of insulin and thus the regulation of blood sugar concentration are rather imperfect, although with intensified conventional insulin therapy supported by self-monitoring devices for blood glucose improvement can be gained. Patient kits with test strips have been developed based on dry chemistry. Urine testing was considered as the only way to obtain day-to-day control before blood-sugar tests were available. The presence of abnormal glucose concentrations in urine indicated a blood glucose level above the renal threshold, resulting in urinary excretion of glucose.

The technology currently available still delivers inadequate glycemic control, even for diabetic patients undergoing intensive insulin therapy. A close connection between late diagnosis or poor stabilization of diabetes mellitus and the development of late complications leading to large vessel and microvascular diseases is evident. The reasons for this are the long-term effects of pathological hyperglycemia. Severe hypoglycemia also has drastic consequences and can lead to unconsciousness. For effective control of blood glucose level through insulin administration, frequent blood glucose monitoring is mandatory. At present, research activities involve the development of reliable glucose biosensors that could be used for an artificial endocrine pancreas. Other applications concern, for example, *in vitro* analyzers and biomedical research.

## Methods and Techniques

Many different assays are available for glucose, and are distinguishable by their sample preparation and detection techniques. In routine clinical practice, analysis is usually of plasma or serum. Preparation of these samples requires centrifugation of the cellular blood components (hematocrit) for plasma and coagulation of proteins for serum. There can be specific problems, such as the stability of blood samples due to the glycolytic activity of the red and white blood cells, unless precautions are taken by using additives such as sodium fluoride, dilution in a hypotonic buffer, or immediate centrifugation. Hemolysed blood as a specimen is rather popular, since time-consuming deproteinization and centrifugation can be avoided. Another advantage is the stability of glucose in such a sample having been treated by the addition of a solution of digitonin and maleimide, by which the erythrocyte membranes are destroyed and glycolytic enzymes are deactivated. For some assays, however, deproteinization of the sample is essential.



**Figure 1** Variations in blood glucose assay results (100% value for deproteinized plasma by definition): <sup>a</sup>Arterio-venous differences depend on nutritional and physiological conditions. <sup>b</sup>Losses of plasma glucose at room temperature without preservatives by 0.2 (0.02) and 0.33 (0.06) mmol l<sup>-1</sup> per hour in blood from adults and newborns, respectively (values in brackets for blood on ice, see Lin YL, Smith CH, and Dietzler DN (1976) Stabilization of blood glucose by cooling with ice: An effective procedure for preservation of samples from adults and newborns. *Clinical Chemistry* 22: 2031–2033); effect of additives has been described by Landt M (2000) Glyceraldehyde preserves glucose concentrations in whole blood specimens. *Clinical Chemistry* 46: 1144–1149; <sup>c</sup>glucose concentrations in plasma are usually ~15% larger than in whole blood (see also Colagiuri S, Sandbæk A, Carstensen B, *et al.* (2003) Comparability of venous and capillary glucose measurements in blood. *Diabetic Medicine* 20: 953–956).

The actual concentration of glucose in blood differs from that in fluids that are derived from whole blood (see **Figure 1**). The concentrations for plasma are ~15% higher than for the original specimen due to a lower glucose concentration within the red blood cells. For deproteinized samples, the volume-displacement effect from separated proteins is relevant. Glucose concentrations in capillary blood samples tend to be slightly higher when compared with those from venous blood, especially after meals, with differences of up to 2 mmol l<sup>-1</sup>. For stationary conditions, as found for fasting subjects, the concentration difference is significantly reduced.

For tissue measurements, an appropriate pretreatment is necessary to allow for thorough cell disintegration and sample homogenization. Dilution of the sample fluid is advisable to reach the linear concentration range of the assay. When photometry is used for detection, sample preparation must result in a clear and colorless solution.

There are certain requirements for the analysis of glucose. The assays used should be rather selective for this monosaccharide and provide high sensitivity and linearity. For clinical laboratories, in particular, the methods require levels of accuracy and precision as set by the guidelines of professional organizations.

The assays should be preferably suitable for all kinds of physiological samples and measuring devices whether manual or mechanical, and applicable in routine analysis, emergency, or research laboratories. Further considerations are the cost factor and simplicity of the method.

Glucose methodology can be classified into several categories. Enzymatic procedures currently dominate clinical assays. However, before these were developed, classical chemical methods were widespread. In recent years, methods based on different spectroscopies have also been introduced.

### Chemical Methods

Although glucose possesses some aldehyde characteristics, it lacks others, e.g., due to the formation of the predominant cyclic hemiacetal in neutral or weakly acidic solutions, when forming two different monomethyl derivatives under conditions that normally convert an aldehyde to a dimethyl acetal. However, its reduction potential can be utilized. For many years qualitative testing of glucose in urine relied on this property using Tollens' reagent (Ag(I) → Ag), Nylander's reagent (Bi(III) → Bi), or freshly prepared Fehling's solution (Cu(II) → Cu(I)). Many modifications for quantitative work were carefully worked out based on the Fehling method.

In another popular chemical method, hexacyanoferrate(III) is reduced to hexacyanoferrate(II), and the unchanged Fe(III) can be quantitatively determined, e.g., by iodometric back-titration with sodium thiosulfate. Modifications have been proposed to allow direct photometric quantification. Other titrimetric methods have also been proposed, for example, using an iodometric method with vanadium(V) in perchloric acid as reagents. The simple reduction methods are regarded as nonspecific due to the interference of other similar compounds with reducing properties.

A useful chemical and rather selective method involves *o*-toluidine, leading to a colored reaction product, glycosylamine, which is photometrically determined at 630 nm; however, aldo-pentoses and aldo-hexoses, such as mannose and galactose, interfere. This method had been adapted for use in automatic analyzers. It had been favored due to its simplicity and reliability, but severe limitations occurred when the reagent used was classified as carcinogenic.

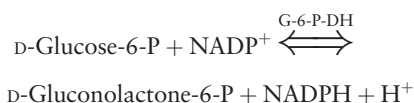
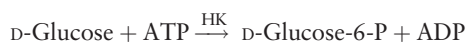
### Enzymatic Methods

In clinical chemistry, enzymatic methodology clearly dominates quantitative analysis, and most glucose biosensors are nowadays based on immobilized

enzyme technology. The reasons for this are the high selectivity and the simplicity of the assays, with possible implementation on analyzers. Different measurement techniques can be taken into account: in comparison with end-point methods, automated kinetic procedures with readings at fixed-time yield much shorter analysis times and, for some assays, a reduced sensitivity to interferences. A prerequisite is that the overall reaction follows pseudo-first-order kinetics with respect to glucose. For glucose determination three main enzyme systems are available, which are described in detail.

**Hexokinase (HK; EC 2.7.1.1) and glucose-6-phosphate dehydrogenase (G-6-P-DH; EC 1.1.1.49)** The determination of deproteinized samples is carried out according to the following reactions. For both enzymes, the presence of magnesium ions is required

for their catalytic activity:



In the first reaction, a phosphate group from ATP (adenosin triphosphate) is transferred to the substrate with the formation of ADP.

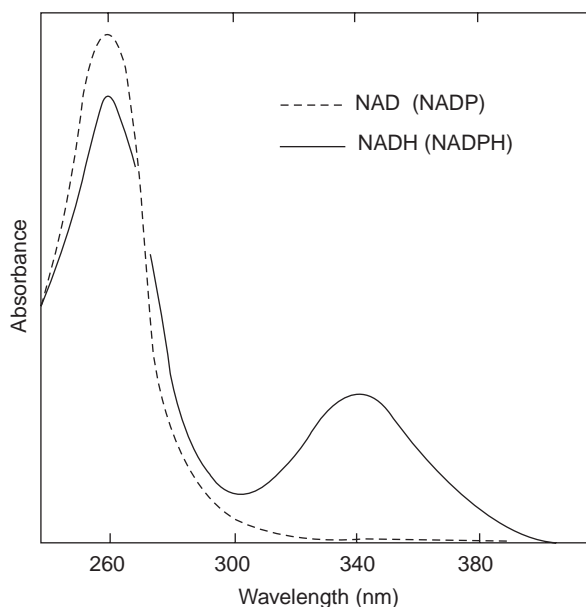
The second, so-called indicator reaction, provides the specificity of the method, since hexokinase also converts other hexoses, such as D-fructose and D-mannose, into their phosphorylated compounds. The formation of the coenzyme NADPH (the reduced form of nicotinamide adenine dinucleotide phosphate) can be measured photometrically, preferably at the absorption band maximum at 340 nm (see Figure 2). The hexokinase method has been internationally recommended as a reference method.

In addition, a fluorescence method has also been described, leading to a detection limit more than 10 times lower than that of the absorbance method. A further reduction, down to femtomolar concentrations, can be achieved by coupling the hexokinase procedure to a bioluminescent indicator reaction using firefly luciferase.

**Glucose dehydrogenase (G-DH; EC 1.1.1.47)**



An advantage here is the single reaction step, after which the lactone produced is spontaneously hydrolyzed to give gluconic acid. The addition of mutarotase accelerates the interconversion of the anomeric  $\alpha$ -form into the  $\beta$ -anomer. Quantification by photometry is the same as for the preceding method. Because only D-xylose and mannose, found at rather low concentration levels in physiological fluids, cause interference in clinical specimens (see also Table 1), this method is widespread and accepted for the determination of glucose due to its



**Figure 2** UV absorption spectra of NAD, NADP, and their reduced forms used for photometry in enzymatic glucose assays.

**Table 1** Reaction rates of different carbohydrates by glucose-dehydrogenase relative to  $\beta$ -D-glucose (100%)

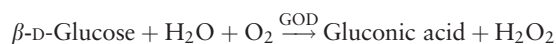
Substrate	Relative rate	Substrate	Relative rate
2-Deoxyglucose	125	Cellobiose	1
2-Amino-2-deoxy-D-glucose	31	D-Ribose	0.8
D-Xylose	15	Lactose	0.7
D-Mannose	8	2-Deoxyribose	0.1
6-Amino-6-deoxy-D-glucose	6	D-Galactose	0.0

Reproduced with permission from Vormbrock R (1988) UV-method with glucose dehydrogenase. In: Bergmeyer HU (ed.) *Methods of Enzymatic Analysis*, vol. VI, ch. 2.4.2. Weinheim: VCH-Verlag; © VCH-Verlag.

simplicity and speed. The reagents for this method are commercially available from several companies as an assay kit. Analysis of body fluids is possible either without or with prior deproteinization by means of, e.g., perchloric acid/perchlorate solution; even hemolysate from capillary blood can be used directly. Excellent correlations with the hexokinase method have been reported.

Alternatively, electrochemical detection by using an amperometric biosensor has been proposed using modified electrodes for the electrocatalytic oxidation of the reduced cofactors (NADH, NADPH). The oxidation current reflects the rate of glucose conversion. Additionally, covalent coupling of the coenzyme is a precondition of more advanced reagentless measuring devices. Further developments use an electron mediator such as ferrocyanide and PQQ/PQQH<sub>2</sub> (pyrroloquinoline quinone) as the cofactor pair.

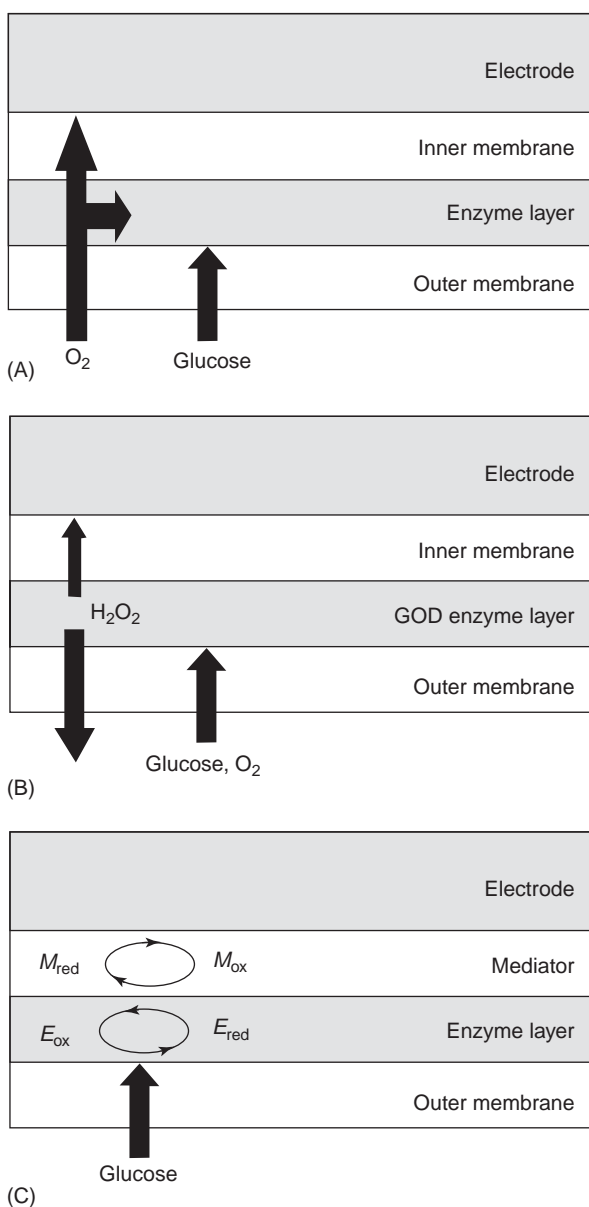
### Glucose oxidase (GOD; EC 1.1.3.4)



This enzyme reaction dominates the glucose assays. Usually, an indicator reaction is used which can be followed spectrophotometrically. In one such reaction the hydrogen peroxide produced reacts with phenol and 4-aminophenazone in the presence of peroxidase to give a colored dye. Other combinations have been proposed that have been adapted for several automatic analyzers.

Other detection procedures follow oxygen consumption by a so-called Clark amperometric oxygen electrode, which is thought to be the first biosensor combining electrochemistry and immobilized enzymes (see **Figure 3**). Alternatively, the formation of hydrogen peroxide can be monitored amperometrically. The removal of interferences from various reducing substances, e.g., ascorbic and uric acids, paracetamol, and others is a prerequisite for selectivity. One method of removing interferences uses a hydrogen peroxide-permeation selective membrane to screen out electrochemical interferences from biological fluids, so that the enzyme layer is sandwiched between permeable membranes. A recent approach employs a scavenging microreactor for preoxidation of interfering substances.

A significant extension of the utilization of amperometric probes could be obtained by electron transfer mediators. The idea is to replace oxygen by an alternative redox couple, allowing a rapid reaction of the reduced enzyme with the mediator, which shuttles electrons efficiently between electrode and



**Figure 3** General approaches for *in vivo* monitoring of glucose by different enzyme electrodes. (A) O<sub>2</sub> based, (B) H<sub>2</sub>O<sub>2</sub> based, and (C) mediator based: *E*<sub>ox</sub> and *E*<sub>red</sub> are the redox forms of the enzyme, *M*<sub>ox</sub> and *M*<sub>red</sub> the redox forms of the mediator needed for electron transfer. The outer layer represents a low-permeability diffusion barrier for glucose. (Reproduced with permission from Reach G and Wilson GS (1992) Can continuous glucose monitoring be used for the treatment of diabetes? *Analytical Chemistry* 64: 381A–386A.)

enzyme. A high selectivity by a pH-independent and low oxidation potential is desirable. One of the most successful mediator compounds is ferrocene (bis-(cyclopentadienyl) iron) and its derivatives. Based on this technology, credit-card and pen-sized devices have been introduced for patient self-monitoring of blood glucose and clinical emergency use (see **Table 2**). The instruments, with a response time down to



**Table 2** Instrumentation for self-monitoring of blood glucose

Product	Maker	Country	URL	Principle	Biosensor chip	Centesis	Sample volume ( $\mu$ l)	Measurement time (s)	Measurement range ( $\text{mmol l}^{-1}$ )
SureStep (GlucoTouch)	Lifescan	USA	www.lifescan.com	CM	GOD-POD Color coupler	Finger	10	30	0–27.8
Medisafe Reader	Terumo	Japan	www.terumo.co.jp/English	CM	GOD-POD Color coupler	Finger	4	18	1.1–33.3
Precision QID	Abbott Lab	USA	Abbottdiagnostics.com	EM	GOD Mediator	Finger	3.5	20	1.1–33.3
Ascensia CONTOUR	Bayer HealthCare	USA	www.bayercarediabetes.com	EM	GOD Mediator	Finger, etc.	0.6	15	0.5–33.3
Glucocard Diameter $\alpha$	Matsushita Kotobuki	Japan	Panasonic.co.jp/mke/en/	EM	GOD Mediator	Finger	2	15	1.1–33.3
Accu-Check Active	Roche Diagnostic	Germany	Diabetes.roche.com	CM	GDH Mediator	Arm, etc.	1	5	0.5–33.3
OneTouch Ultra	Lifescan	USA	www.lifescan.com	CM	Color coupler	Arm, etc.	1	5	1.1–33.3
Free Style	Therasense	USA	www.therasense.com/freestyle	EM	Enzyme	Arm, etc.	0.3	~15	1.1–27.8
Precision Sof-Tact	Abbott Lab	USA	Abbottdiagnostics.com	EM	GDH Mediator	Arm, etc.	3	20	1.7–25.0
Medisafe ez Voice	Terumo	Japan	www.terumo.co.jp/English	CM	GOD-POD Color coupler	Finger	4	18	1.1–33.3
GlucoWatch G2Biograph	Cygnus	USA	Www.glucowatch.com	EM	GOD	Arm	Minimal-invasive	–	2.2–22.2

CM, colorimetry; EM, electrochemical method; GOD, glucose oxidase; POD, peroxidase; GDH, glucose dehydrogenase.

In part from: Reproduced with permission from Nakamura H and Karube I (2003) Current research activity in biosensors. *Analytical and Bioanalytical Chemistry* 377: 446–468; © Springer-Verlag.

15 s, utilize small, disposable electrode strips. Several studies have been devoted to their response to different  $\text{O}_2$  partial pressures. For the reaction above, oxygen is needed to reoxidize the reduced enzyme-cofactor complex (glucose oxidase/FADH<sub>2</sub>) and for producing hydrogen peroxide. In the case of an electron mediator, oxygen is competing for reoxidation of FADH<sub>2</sub>.

The performance of the handheld meters matches the best test systems currently available, but still cannot compete with standard laboratory devices. It must be noted that the instruments should be handed out only to patients trained in the use of the meter to achieve the necessary testing quality. In the analytical instrument market, analyzers based on enzymatic electrochemical technology dominate. A summary of the performance of some commercially available devices with regard to linearity, coefficient of variation, sample frequency, etc., is presented in **Table 3**.

There are so many variants of enzymatic assays that they cannot be listed here completely, but fiber-optic biosensors must be mentioned. The so-called optodes employ immobilized indicators in combination with optical fibers and some kind of spectroscopic detection. **Figure 4** shows schematically an oxygen optode used for glucose sensing with fluorescence detection. Another category includes fiber-optic sensors based on bio- and chemiluminescence, which can give picomolar detection limits. A chemiluminescent reaction, for example, can be coupled to the production of hydrogen peroxide by using luminol. The ultrasensitivity of such assays has great attractions for the design of novel biosensors, in particular, as they allow miniaturization and simple instrumentation. As one of the latest developments, multiple-analyte biosensor chips have been designed based on electro-chemiluminescent measurements. For glucose, such biosensors utilize the reaction of the hydrogen peroxide produced by the GOD enzyme and electrochemically oxidized luminol, both immobilized on microbeads. The photons emitted are measured by a charge-coupled device (CCD) camera, leading to a detection range of  $20 \mu\text{mol l}^{-1}$  to  $2 \text{ mmol l}^{-1}$ .

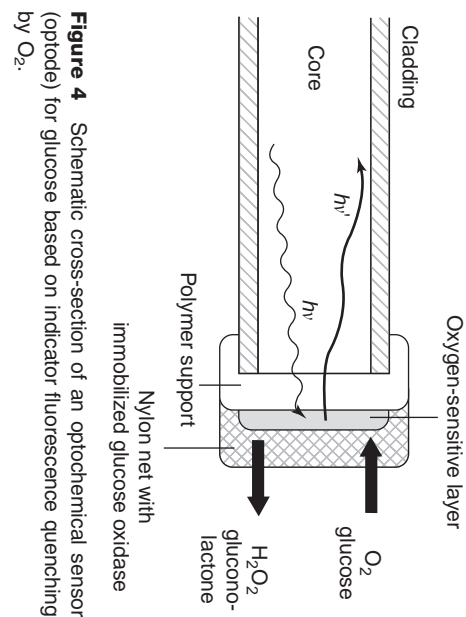
### Dry Chemistry

The first major upsurge in dry chemistry took place with the introduction of the Ames Clinistix (Miles, USA) in the 1950s for the testing of urinary glucose. The test strips utilized the oxidation of  $\beta$ -D-glucose as catalyzed by glucose oxidase with hydrogen peroxide formation and subsequent oxidation of a chromogen impregnated on a paper matrix to a colored compound. The approach has subsequently been

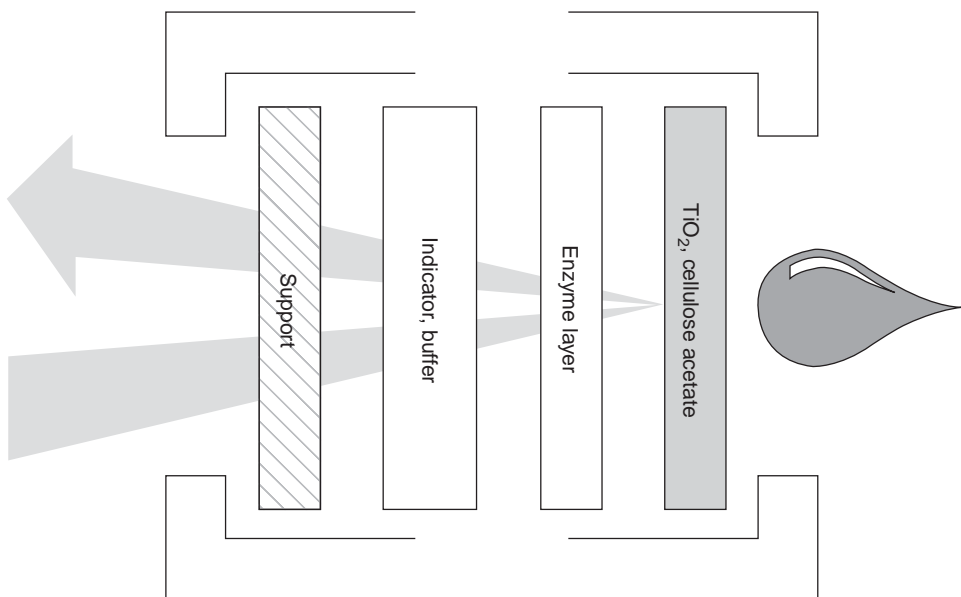
**Table 3** Selection of commercial clinical laboratory instruments for glucose determination

Manufacturer/URL	Instrument	Technology	Sample ( $\mu$ l)	Linear range	Frequency	Precision	Stability
A & T Corp., Japan www.aandt.co.jp/eng/	GA05	Immobilized GOD, oxygen electrode	35 serum/plasma 10 urine	0–280 mmol l <sup>-1</sup>	160–200 samples per hour	$\leq 0.3$ mmol l <sup>-1</sup>	Not reported
Eppendorf AG, Germany www.eppendorf.com	EBIO <sup>®</sup> plus	GOD, amperometric Thick film sensor	10–20 Whole blood, serum/plasma	0.6–50 mmol l <sup>-1</sup>	160 samples per hour	$\leq 2.5\%$ at 12 mmol l <sup>-1</sup>	7000 analyses or 60 days
YSI (Yellow Springs Instrument Co.), USA www.ysi.com	YSI 2300 STAT Plus Analyzer	Immobilized GOD, H <sub>2</sub> O <sub>2</sub> electrode	25	0–50 mmol l <sup>-1</sup>	36 samples per hour	2% of reading or 0.2 mmol l <sup>-1</sup>	21 days
i-STAT Corp./Abbott Lab. www.istat.com	i-STAT 1 Analyzer	GOD, amperometric with supply of cartridges for POC analysis	16–95 depending on cartridge type, whole blood, plasma	1.1–38.9 mmol l <sup>-1</sup>	Analysis time typically 130–200 s	1.6% at 2.3 mmol l <sup>-1</sup> 0.8% at 15.9 mmol l <sup>-1</sup>	At 20°C for 2 weeks; at 2–8°C until expiration date
HemoCue, United Kingdom www.hemocue.co.uk	HemoCue Glucose 201 +	Glucose dehydrogenase, photometry	5 whole blood	0–22.2 mmol l <sup>-1</sup>	60–90 samples per hour	1.5%	> 1000 assays

GOD, glucose oxidase; POC, point of care.



**Figure 4** Schematic cross-section of an optochemical sensor (optode) for glucose based on indicator fluorescence quenching by O<sub>2</sub>.



**Figure 5** Schematic diagram of a solid phase reagent system for glucose measurement (Ektachem System).

extended to other biofluids such as plasma and whole blood. The development was made possible by application of different technologies. The Ektachem system (Eastman Kodak Co, Rochester, NY, USA), for example, is shown schematically in Figure 5. All

reactions are carried out on a multilayer system similar to a photographic film. The serum or plasma droplet is placed on to a spreading layer. A multilayered reagent zone is below the reflective area filled with pigments such as  $\text{TiO}_2$  or  $\text{BaSO}_4$ . The chemical reaction products can be evaluated visually or by using a small device, which measures diffuse reflectance. The analyte concentration is mostly calculated relative to a reflectance standard using some signal linearization. Systems have also been developed to deal with whole blood that may be wiped off or remain on the test strip.

The availability of reagent strips and pocket-sized reflectance meters has had a great impact on the management of diabetes allowing for intensified therapy. There are many different devices on the market, which are clearly challenged by the novel electrochemical testing equipment described above. Dry chemistry has made significant contributions to clinical chemistry, in particular in the area of rapid diagnostics for point-of-care testing or self-monitoring of blood glucose by patients with diabetes, at the cost of slightly reduced reliability compared with standard methods. Instrumentation ranges in size from hand-held instruments, as for dedicated blood glucose analysis (see **Table 2**), to bench-top systems such as the Reflotron<sup>®</sup> Plus from Roche Diagnostics (Mannheim, Germany), which are capable of handling more than one analyte with a specific module for each test. Similar analysis slides are manufactured and sold by Johnson & Johnson for their Vitros analyzer (see also <http://www.vitros.com>).

### Automation in the Clinical Laboratory

For managing today's clinical testing workloads, automation is a prerequisite. Besides the general chemistry testing for glucose and further compounds, fully automated and computer controlled analyzers are capable of quantifying many additional parameters (electrolytes, therapeutic drugs, and others) in various biofluids such as serum, plasma, urine, and cerebrospinal fluid samples. However, some floor-model instrumentation may be restricted only to, for example, critical care clinical chemistry parameters. Usually, random access sample handling is included, and the frequency of tests can reach 800 and more per hour. A user interface enables loading of samples in racks or carousels, of controls, reagents, and wastes. The essential instrumental parts are concerned with efficient transport fluidics and robotics and further processing units as needed for agitation, mixing, measurement, etc. Carryover between samples is tightly controlled and achieved with small volumes of cleaning solutions. Photodiode-array

detectors with discrete wavelengths or interference filters have been mainly used for photometry. On the other hand, compact analytical devices incorporating one-parameter dedicated biosensors, able to produce either discrete or discontinuous electronic signals related to individual analytes, have also been designed for routine analysis as discussed above.

### In Vivo Monitoring

The basic components of an artificial pancreas are the blood glucose sensor and an automated and possibly implantable insulin pump with a 'closed loop' regulating process for insulin administration. For clinical applications, two developments – either based on optical methods or enzymatic–electrochemical approaches for glucose detection – must be noted. They led to extracorporeal monitoring devices that were connected through a double-lumen catheter to a peripheral vein for realizing a continuous, buffer diluted blood flow. Early attempts for a bedside artificial pancreas resulted in the Biostator (Miles Laboratories, Elkhart, IN, USA). The apparatus offered promising perspectives for clinical research programs, but the original therapeutic intention, i.e., optimization of insulin therapy in diabetic patients, could not be satisfactorily reached. Nevertheless, experience from the application of such instrumentation accelerated the efforts in the development of paracorporeal and intracorporeally applicable devices and implantable sensors for blood glucose determination under daily life conditions.

Nowadays, the intensive care discipline is especially interested in continuous monitoring or at least in frequent analysis of glycemia (point-of-care testing). Intensive insulin therapy in critically ill patients can thus be realized when it is necessary to 'clamp' their glucose concentration between 4.5 and  $6.4 \text{ mmol l}^{-1}$  for improving their long-term outcome.

**Implantable biosensors** In contrast to conventional laboratory analyzers and extracorporeal bedside devices, where any sample can be adapted to the needs of the sensing element by means of dilution, centrifugation, addition of chemicals, and other preparation steps, an intracorporeal sensor has to meet certain conditions to function under the natural environment of the surrounding biological medium. Furthermore, the sensing element and the ancillary electronics for signal amplification and processing must be miniaturized, and – at least for sensors with a limited lifetime – be mass-producible, easily applicable, and exchangeable, as well as affordable for the patient. Although several principles can be used to



realize glucose measurements intracorporally, most enzymatic biosensors employ electrochemical detection, but a few approaches using optical detection by infrared (IR) spectrometry or polarimetry also exist.

Many biosensors are now widely used for *in vitro* measurements, but implantable devices have been reported to possess only limited functionality due to inflammatory tissue reaction. Intravascular monitoring is particularly hampered by thrombotic deposition on the sensing element and further associated complications. From the alternative approaches, the direct implantation of a needle-type sensor into the subcutaneous tissue is often favored. Further requirements for the sensors are nontoxicity and sterilizability without loss of enzyme activity. An important prerequisite for the long-term application of such devices is the biocompatibility of the cover membranes. Special materials such as 2-methacryloyloxyethyl phosphorylcholine, polyethylene oxide-based hydrogels, or Nafion<sup>®</sup> have been shown to give satisfactory short-period results. Using the latter membrane, the interference from ascorbic acid can also be eliminated. Extensive research work has been carried out by many groups to develop membranes for long-term bio-stability.

An essential difference between any extracorporeal or laboratory measuring method and sensors in the body is the lack of cleaning of the measuring element. Thus, the continuous exposure of an intracorporally implanted sensor to the body tissue leads to protein build-up on its diffusion membranes and finally to a hypovascular foreign body capsule, resulting in unpredictable changes in the diffusion characteristics toward the biosensor surface. Experimental studies in animals as well as in humans have demonstrated that continuous glucose measurements over some days or weeks by means of subcutaneously implanted biosensors are at least possible.

**Microdialysis devices** Another interesting option to realize continuous or at least frequent glucose measurements is based on microdialysis. Comparable with a conventional double lumen catheter, the microdialysis cannula has a small dialysis membrane or a hollow fiber with a well-defined recovery of body fluid on the tip of the catheter. In most cases, microdialysis probes have been placed subcutaneously to allow compounds of low relative molecular mass such as glucose from the interstitial fluid to penetrate the dialysis membrane by diffusion. By means of a continuous or discontinuous flow of buffer solution, the harvested glucose, depending on the recovery through the membrane, will be diluted and transported to an extracorporeal measuring device. Compared with direct implantation, this extraction

technique significantly eases constraints that arise from sensor poisoning due to the elimination of proteins of molecular mass above the membrane cutoff. It allows pocket-sized instrumentation with the sensing device, usually an amperometric enzyme electrode, to be placed outside the body. Compared with bedside analyzers using diluted blood, the main advantage of the microdialysis sampling is the very small amount of interstitial fluid needed. This even allows application in children. Such systems have been used in the intensive care of neonates and after surgical operations for safeguarding patients against hypo- and hyperglycemia, especially after experiencing severe traumatic brain lesions and stroke.

However, a disadvantage, caused by the small sample volume, the needed length of the catheters, and the slow flow velocity of the buffer/sample mixture, is the time lag between sampling and detection. When adding up diffusion time, flow time, and response time of the measuring system, the lag time is usually ~15 min. To avoid long delay times, the need for small and low dead volume measuring devices is evident, and optimized devices with a total delay time of <3 min have been reported. Thus, microdialysis-based systems are suited for monitoring tasks and for measuring controlled therapeutic steps – even during rapid glucose dynamics.

Finally, there have been many investigations into the glucose concentration profiles in blood and subcutaneous tissue, or to be precise, within the interstitial fluid. Changes in the blood are transmitted to the subcutaneous tissue with a delay of 5–10 min, so that glucose homeostasis is possible by using, for example, a needle-type sensor, although the ‘gold standard’ is still set by the glucose concentration in capillary blood.

**Minimal-invasive technology** Other novel strategies such as minimal-invasive techniques, e.g., the so-called GlucoWatch biographer (Cygnus, Redwood City, CA), have been recently evaluated. The sensor is wearable like a watch, and an intermittent glucose analysis is carried out every 20 min by employing reverse iontophoresis through the skin for biofluid collection. The lifetime of the glucosensor is 12 h, and furthermore a skin preconditioning for 3 h is essential for quasi-continuous measurements, especially for hypoglycemia alert. With another approach, interstitial fluid is collected through an array of micropores created with a laser in the outer horny skin layer and measured in a patch containing a glucose sensor. For noninvasive techniques, refer to spectroscopic methods below.

## Spectroscopic Methods

**Polarimetry** With spectroscopic techniques the interaction of matter with electromagnetic radiation is evaluated. From the optical activity mentioned already, a rapid polarimetric method for glucose can be derived, which is of considerable technical importance, although it is subject to interference from other optically active compounds. The aqueous humor of the eye has been investigated for use in monitoring applications. Using the optical rotation of the *in vitro* humor, a correlation with blood glucose has been established. However, problems due to the time lag in the different concentration profiles and instrumentation for *in vivo* measurements remained unsolved (for a rabbit's eye used as a test model, a time lag of 5 min was determined when drastic changes in blood glucose concentration occurred). Micro-degree polarimetry using diode lasers with visible and near-IR radiation have recently received attention. For improving the method selectivity, multi-wavelength approaches have been tested in the past.

**Infrared spectroscopy** IR spectroscopy is a widely applied technique for obtaining quantitative results on a great variety of substances, but still rarely found in the clinical laboratory despite the recent progress observed for novel assay technologies. The technique is based on the wavelength- and concentration-dependent absorption of IR radiation by the molecules under study and can provide a simultaneous, continuous, and reagentless determination of many analytes. In principle, compounds participating in the enzymatic reactions of glucose can also be monitored. However, advances in direct spectrometry have yielded a significant contribution to analytical chemistry in the biomedical field. With the advent of multivariate calibrations with multi-wavelength information, the technique has become tremendously successful within the field of routine analysis and process control. The new challenge presented by biomedical applications is the low concentration range of compounds to be monitored against an extremely large and varying background.

The mid-IR spectral range contains much denser and more selective information compared to the near-IR, where overtone and combination bands of the fundamental molecular vibrations occur. As the intensity of bands in the mid-IR is higher, optical path lengths are on the micrometer scale. Such path lengths can be achieved with a special technique that uses attenuated total reflection, and which renders optical materials compatible with aqueous biofluids. Recently, dry films of biosamples of nanoliter volumes have been successfully applied for reagent-free

quantification. In the near-IR, cuvettes normally have millimeter path lengths due to smaller sample absorptivities.

Several extensive studies have investigated the prospects of clinical assays for glucose, as well as other substrates, in human plasma or whole blood. The intra-assay variation coefficients using large hospital sample populations for validation reached less than 5%. Since reference methods are needed to provide calibration standards, their imprecision can influence assay results significantly, as has been found for glucose. Using near-IR radiation, which is of great interest because of its potential for noninvasive glucose monitoring in tissue, quantitative results were slightly worse when studying plasma sample populations by multivariate near-IR spectrometry.

Different approaches have been proposed to achieve noninvasive *in vivo* monitoring of glucose, which is desirable for a variety of reasons (as already discussed under section: *In vivo* Monitoring). There are two alternatives: probing of subcutaneous tissue by using diffuse reflectance with radiation of wavelengths  $\sim 1.6 \mu\text{m}$ , where significant glucose absorption exists, or by using spectrometry within the so-called therapeutic window with wavelengths  $\sim 1.0 \mu\text{m}$ , which renders a feasible transmission measurement, e.g., of a finger tip, or exploits the glucose concentration-dependent scattering within skin tissue. It is hoped that acceptable equipment meeting the demands of patient self-monitoring blood glucose devices can be developed soon. Only a few results with calibrations against capillary blood glucose have been published so far. Reliability is achieved only for hyperglycemic concentrations, because average prediction errors for glucose have been found to be just below  $2 \text{ mmol l}^{-1}$ , which is still unacceptable for normal and hypoglycemic ranges.

**Raman spectroscopy** The Raman effect involves inelastic radiation scattering that provides molecular spectral fingerprints with similar, but often complementary spectral information to mid-IR spectroscopy. Therefore, also a simultaneous reagent-free measurement of multiple analytes is possible. An advantage is the weak water Raman spectrum, contrary to the intensive water absorption bands in the IR spectrum. Recently, glucose measurements in various biofluids such as plasma and even whole blood have been pursued, but interference from strong fluorescence of matrix substances or photo-decomposition of the sample by intensive laser radiation has hampered the studies. Therefore, most recent research has been carried out using near-IR excitation Raman spectroscopy, thus avoiding water and other biosample absorbers. However, tremendous trade-off

due to lower Raman intensities is found for longer wavelength excitation, compared with excitation with UV-visible lasers. Despite this, intensive excitation sources such as diode lasers and extremely sensitive detectors, e.g., silicon-based CCDs, are available for miniaturized spectrometer systems. More research will be needed for routine implementation of Raman spectroscopy based assays.

**Fluorescence spectroscopy** A promising technique for *in vitro* and *in vivo* assays of glucose involves the sensitive detection of fluorescence emitted by compounds that have absorbed radiation from an external monochromatic source. Because of their simplicity and sensitivity, glucose-specific fluorescence probes have been designed in the past, e.g., involving the binding of glucose to a protein such as concanavalin A. In a competitive assay, the fluorescence resonance energy transfer (FRET) is used between a fluorescent donor and an acceptor, each covalently linked to the protein or a dextran polysaccharide. In the absence of glucose, the binding between the protein and dextran leads to a high FRET efficiency, while the presence of glucose causes competitive binding, thus reducing the FRET effect. Preliminary *in vivo* tests with skin-implanted poly(ethylene glycol) microspheres, containing such a transcutaneously probed fluorescent compound pair, have been reported.

**NMR spectroscopy** Nuclear magnetic resonance (NMR) spectroscopy is an important technique for the study of biological fluids and intact cells. Compared with other analytical methods, it is nondestructive and noninvasive. For carbohydrates, only the proton ( $^1\text{H}$ ) NMR spectroscopy is relevant, although phosphorylated compounds have been measured using  $^{31}\text{P}$  NMR spectroscopy. So far, the measurement of biosamples is complicated by the water resonance that obscures a large portion of the spectrum, thus creating a dynamic range problem during data acquisition. With high-field NMR spectrometers, however, using special pulse sequences, high-resolution spectra can be obtained from physiological fluids such as blood plasma and, for example, red blood cells with concentrations of glucose at the  $\text{mmol l}^{-1}$  level. Detection limits reported are in the order of  $0.01\text{--}0.1\text{ mmol l}^{-1}$  for small molecules, allowing the study of intracellular glycolytic chemistry. For glucose in blood plasma a coefficient of variation of better than 4% has been reported and accuracy is generally within 5% of a clinical glucose oxidase method. The  $^1\text{H}$  NMR spectroscopic method cannot compete with, for example, enzymatic methods on a routine basis, due to spectrometer and

measurement complexity, but it shows its potential for applications needing simultaneous multicomponent analysis, especially for carbohydrates and their metabolites.  $^{13}\text{C}$  NMR can provide spectra of plasma glucose with isotopically enriched compounds, which are necessary because of the low NMR sensitivity and the low natural abundance of  $^{13}\text{C}$ .

**Positron emission tomography** With this technique, radioactive nuclides that emit positrons during decay are administered. The positrons interact with an antiparticle, the electron of a neighboring molecule. This event causes two  $\gamma$  photons to be emitted in opposite directions that can be detected by two separate detectors. This allows the assessment of three-dimensional images. For glucose to be doped, the  $^{11}\text{C}$  labeled compound must be synthesized. As this technique allows the investigation of biodynamic processes, for example, glycolytic rates, it has been successfully used for the measurement of regional cerebral or myocardial glucose metabolism in human subjects. These applications of a complex, but noninvasive method, provide insights into the functional physiology of the human brain unmatched by any other currently known method.

**Mass spectrometry (MS)** For *in vitro* glucose analysis, MS has been used in combination with gas chromatography (GC). As demonstrated in several studies, isotope dilution mass spectrometry can offer an analytical method of high accuracy and precision. In order to evaluate routine methods, reference methodology is needed. Great demands on analytical performance are required to get a 'definitive method' accepted. After adding  $^{13}\text{C}$  labeled glucose to the sample, glucose is separated from the matrix and derivatized to allow combined capillary GC-MS. The ratio of certain peaks areas, specific for normal and  $^{13}\text{C}$  labeled glucose (selected ion monitoring), are evaluated for quantification for sample and calibration standards. By use of a standard serum reference material, a relative accuracy of  $\sim 0.5\%$  has been established; values slightly below this can be achieved for the coefficient of variation as a precision parameter (within-day and between-day imprecision). Results achieved for determining glucose in whole blood were slightly larger (mean of 1.0% for within-run coefficients of variation).

Recently, matrix-assisted laser desorption/ionization (MALDI) MS of carbohydrates has found much interest, since it also enables the study of underivatized compounds. Besides methods for sample preparation and for extracting carbohydrates from biological media, various MALDI matrices were compared in a recent review.

An interesting field is the determination of the glucose turnover rate in human subjects. With the advances in MS instrumentation, rates can be measured using stable isotopes, and are a prerequisite for studying dynamic aspects of glucose metabolism. For this purpose, 6,6-dideuteroglucose is a suitable tracer, but recent results for studying glucose kinetics within intravenous glucose tolerance tests were also obtained using  $^{13}\text{C}$  labeled glucose. Glucose oxidation rates can be determined by administering uniformly  $^{13}\text{C}$  labeled glucose and analyzing the  $^{13}\text{C}$  enrichment of exhaled carbon dioxide. In this case, an isotope-ratio mass spectrometer is used. Other kinetic metabolic parameters, such as glucose carbon recycling and production rates, can be detected by mass isotopomer analysis of plasma  $^{13}\text{C}$  glucose, which overcomes the problems of measuring the low  $^{13}\text{C}$  isotopic abundance in plasma glucose. One limitation, however, is the rather high cost of the tracer.

### Separation Techniques

The use of GC was mentioned in the preceding section. It requires, prior to sample injection into the GC column, some sample cleanup and derivatization to produce volatile species. Several different derivatives, such as penta-trimethylsilyl derivative, aldono-nitrile acetate, diacetone glucose, and others, have been suggested allowing the separate determination of the two anomeric species. For mass spectrometric detection, the reaction of D-glucose with *n*-butylboronic acid (BBA) followed by acetic anhydride leading to a quantitative conversion to glucose-BBA can be utilized. A fused silica capillary column coated with SE-54 may be considered for separation of different aldohexoses.

The use of high-performance liquid chromatography (HPLC) for the analysis of carbohydrates has increased, as different column materials have become available. Water–acetonitrile mixtures are needed for partitioning on amine-bonded silica phases, whereas ion chromatography with anion or ligand exchange separation modes allows the use of water as the mobile phase. One advantage of HPLC is that the aqueous carbohydrate sample can be injected directly into the column and mono-, di-, and oligosaccharides can be separated in a single step. This method is important for the analysis of glycoproteins after hydrolysis. Detection can be via a refractive index or an optical rotation detector as referred to in the section ‘Polarimetry’. Since spectrophotometry offers only poor sensitivity for carbohydrates, postcolumn derivatization is required for fluorescence measurements. Other postcolumn reactions combine immobilized

enzyme reactors with amperometric or luminol chemiluminescence detection.

Pulsed amperometric detection (described below) delivers selective and sensitive measurements, also allowing gradient elution in HPLC. This detection technique has been a great breakthrough in carbohydrate analysis.

Capillary electrophoresis is another high-efficiency analytical technique that currently has significant impact as a tool for bioanalytical research. This separation technique is advantageous with respect to small sample volumes, rapid analysis, large resolution power, and low costs, thus making it ideal for the analysis of numerous endogenous substances in biofluids. Analysis of sugars at high pH needed for conversion to ionic species, with electrochemical detection, is a promising approach for separation. Recent developments have led to the use of microchip electrophoresis, with a total analysis time of 1 min, suited for clinical glucose assays. UV and fluorescence detection is also possible, but needs postcolumn derivatization for chromophore tag appendage. With amperometric detection, a linear response for glucose has been reported for the range of 10–1000  $\mu\text{mol l}^{-1}$ .

### Electrochemical Detection

Monosaccharides, for example, undergo direct electro-oxidation at modest positive potentials using noble metal electrodes, e.g., gold or platinum, but the electrode surface is poisoned by oxidation products, although the poisoning can be eliminated by oxidative desorption. The latter produces a metal oxide layer, which is reduced back by a negative potential to restore electrode activity. Detection limits with pulsed amperometric detection are  $\sim 10$ –50 pmol. A similar electrocatalytic sensor can be exploited for *in vivo* monitoring of glucose. Adjustments to the potential sequence at an appropriately membrane-covered electrode have been necessary to reduce interference from endogenous oxidizable substances such as urea, ascorbic acid, and others, or from ethanol and pharmaceutical compounds. Another technique is achieved by using constant potential amperometric detection with working electrodes of transition metals (Cu, Ni), which needs only simple instrumentation with detection limits of 20  $\text{nmol l}^{-1}$  reported for glucose.

### Significance and Interpretation of Results

In the clinical laboratory, the most frequently found samples are of blood or fluids derived from whole



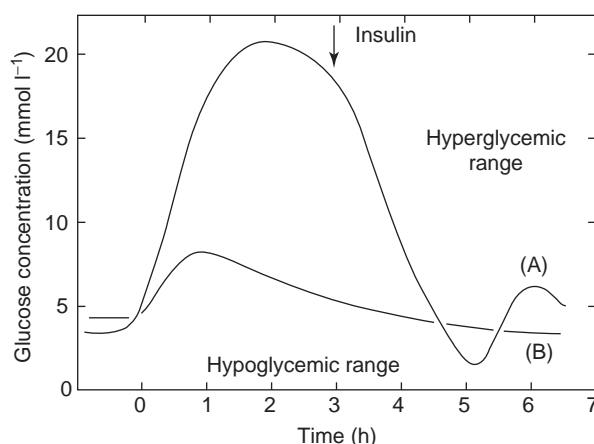
**Table 4** Reference fasting concentration intervals for glucose in capillary blood plasma samples (compiled from various references)

Subject	Range ( $\text{mmol l}^{-1}$ )
Adults	3.8–6.4
Children between 2 and 16 years	3.8–5.8
Children within 2nd year	3.3–5.5
Babies after 1st month, within 1st year	2.2–5.0
Newborns within 1st month	1.6–3.3

blood. Differences in glucose concentration between such specimens, which are specific for sample preparation and blood origin, have already been mentioned. Since the biologically significant glucose is found in the arterial vessels, often capillary blood samples are analyzed. For adults, physiological regulation mechanisms usually keep the blood glucose level between 2.5 and  $7.3 \text{ mmol l}^{-1}$ . Concentration values lying outside this range are grounds for further investigation. Reference intervals as statistically defined by using a population of healthy individuals are presented in Table 4. No differences between female and male subjects have been observed. For children, the ranges are generally shifted to lower values. Neonatal hypoglycemia has been suggested below  $1.1 \text{ mmol l}^{-1}$  for the newborn of low birth weight and, for normal newborns, below  $1.6 \text{ mmol l}^{-1}$  from birth until the fourth day.

The oral glucose tolerance test (see Figure 6) is the most commonly used test for diagnosis of diabetes mellitus and impaired glucose tolerance. There have been suggestions for standardization, especially with concern for the oral administration of the glucose load (aqueous solution of 75 g of glucose in 500 ml), and criteria for interpretation. The normal values for fasting plasma glucose are below  $6.1 \text{ mmol l}^{-1}$ , whereas the normal 2-h postload blood glucose concentration has been defined to be  $<7.8 \text{ mmol l}^{-1}$ . A glucose concentration level over  $11.1 \text{ mmol l}^{-1}$  after 2-h postadministration indicates the existence of diabetes, in particular with increased glucose concentration for the fasting subject above  $7.0 \text{ mmol l}^{-1}$  (all values for capillary whole blood). A fasting concentration level below the latter value, but between 7.8 and  $11.0 \text{ mmol l}^{-1}$  during the tolerance test, indicates an impaired glucose tolerance. Another type of metabolic disorder is gestational diabetes that can occur during pregnancy. Initial screening employs the following diagnostic criterion of a fasting plasma glucose  $>7.0 \text{ mmol l}^{-1}$ .

The normal renal threshold is between 9 and  $10 \text{ mmol l}^{-1}$ , but may show some individual scatter. At this level, glucose is spilled over into the urine, resulting in a more or less strong correlation of blood

**Figure 6** Glucose time dependence for an oral glucose tolerance test of a diabetic (trace A) and a healthy subject (trace B). Time is after glucose intake. The administration of insulin is usually not performed, but exemplifies the difficulties in control of blood glucose, since the hypoglycemic range is reached later. The central unshaded area shows schematically the normal and tolerated concentration ranges for subjects undergoing this test.

and urine glucose concentrations. As the concentration of several constituents varies in this body fluid, the excreted glucose values are usually stated on a per day basis (the average daily urine volume of adult subjects is  $\sim 1.5 \text{ l}$ , but can vary significantly). The reference range here has been defined up to  $2.8 \text{ mmol}$  per day. Concentrations above  $55 \text{ mmol l}^{-1}$  have been measured in urine samples of diabetic patients. Below the renal threshold, slightly larger urine glucose concentrations outside the physiological range can occur, indicating an impaired glucose tolerance as a sign for a developing diabetes or renal failure.

Another, but less frequently monitored body fluid is the cerebrospinal fluid. However, its concentration profile is important for neurodiagnostics. The glucose concentration is  $\sim 60\text{--}80\%$  of the blood level with concentration profiles delayed due to diffusion by  $\sim 4 \text{ h}$ . The reference interval is between 1.1 and  $4.4 \text{ mmol l}^{-1}$ .

Saliva has received much attention; because it is easily sampled, the correlation of the glucose concentration therein, to that in blood, has been studied in detail. A glucose level in parotid saliva for fasting subjects has been found to be between 5 and  $30 \mu\text{mol l}^{-1}$ . After consumption of carbohydrates, concentration increases by factors of two to four have been noticed. However, correlation between the concentrations of glucose in plasma and saliva is rather poor due to other factors influencing the salivary glucose concentration. Blood testing is still essential for biomedical diagnostics and, in

particular, for blood glucose self-monitoring by diabetic patients.

## Trends in Glucose Methodology

Biosensors are currently en vogue and their development rapid; and a new field of synthetic receptors has been established recently. Miniaturization of biosensors can be done on the basis of nano- and microfabrication technology, and mass production will make them inexpensive and disposable. Self-contained analytical devices of small size, which respond selectively and reversibly to the concentration of a compound in a biological sample, still attract many researchers. Nowadays, third-generation glucose meters are even equipped with fully developed quality control and patient data management. In particular, much interest is also shown in bio- and chemiluminescence detection in novel biochip technology or in combination with intracellularly applicable nanobeads due to their extremely low detection limits for glucose.

Innovative research has also led to interesting novel instrumentation, e.g., the development of needle-type sensors for the simultaneous monitoring of glucose and insulin. Other detection schemes have been realized, for example, with disposable biosensors employing a glucose-sensitive biochemomechanical gel, i.e., the pH-dependent gel shrinkage depends on glucose oxidase enzyme activity and glucose concentration. Assays with colorimetric glucose recognition have also been developed, for which glucose assembles special functional groups into a supramolecular complex (photonic crystal hydrogel with changes in the diffraction of visible light). A different detection technique has been used by Di-setronic Medical Systems' scientists. They exploit viscosity changes within a sensitive liquid, consisting of a macromolecular Dextran with terminal glucose residues and Concanavalin A, which is a glucose-specific lectin and links the Dextran into a highly viscous gel. Free glucose weakens the cross-linking with the result of a measurable viscosity reduction.

For the clinicians, other aspects are more important. Although for the last few years many predictions of a breakthrough in the field of biosensors have been made, a satisfactory system for long-term continuous *in vivo* glucose monitoring is not yet available. In this context, *in vivo* calibration of such sensor systems is still an issue. The two different approaches, either invasive methodology using electrochemical detection or noninvasive monitoring by spectroscopic techniques, may be successfully applied for diabetes therapy in the not too distant

future, although regular controls by means of standard assays for the determination of glucose in blood and urine will still be required.

**See also:** **Amperometry.** Blood and Plasma. **Cerebrospinal Fluid.** **Chiroptical Analysis.** **Clinical Analysis:** Sample Handling. **Enzymes:** Immobilized Enzymes; Enzyme-Based Assays. **Fluorescence:** Clinical and Drug Applications. **Gas Chromatography:** Mass Spectrometry. **Infrared Spectroscopy:** Near-Infrared. **Isotope Dilution Analysis.** **Liquid Chromatography:** Column Technology; Instrumentation. **Sensors:** Overview.

## Further Reading

- Abel PU and von Woedtke T (2002) Biosensors for *in vivo* glucose measurement: Can we cross the experimental stage? *Biosensors and Bioelectronics* 17: 1059–1070.
- American Diabetes Association (2003) Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* 26(Supplement 1): S5–S20.
- Baba JS, Cameron BD, Theru S, and Coté GL (2002) Effect of temperature, pH, and corneal birefringence on polarimetric glucose monitoring in the eye. *Journal of Biomedical Optics* 7(3): 321–328.
- Burtis CA, Ashwood ER, and Tietz NW (eds.) (1999) *Tietz Textbook of Clinical Chemistry*, 3rd edn. Philadelphia: WB Saunders.
- Fraser DM (ed.) (1997) *Biosensors in the Body: Continuous In Vivo Monitoring*. Chichester: Wiley.
- Hannestad U and Lundblad A (1997) Accurate and precise isotope dilution mass spectrometry for determining glucose in whole blood. *Clinical Chemistry* 43: 794–800.
- Harvey DJ (1999) Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrometry Reviews* 18: 349–450.
- Heise HM (2002) Glucose measurements by vibrational spectroscopy. In: Chalmers JM and Griffiths PR (eds.) *Handbook of Vibrational Spectroscopy*, vol. 5, pp. 3280–3294. Chichester: Wiley.
- Heller A (1999) Implanted electrochemical glucose sensors for the management of diabetes. *Annual Reviews on Biomedical Engineering* 1: 153–175.
- Kost GJ, Nguyen TH, and Tang Z (2000) Whole-blood glucose and lactate – trilayer biosensors, drug interference, metabolism, and practical guidelines. *Archives of Pathology and Laboratory Medicine* 124: 1128–1134.
- Lourido J, Ederoth P, Sundvall N, Ungerstedt U, and Nordström CH (2002) Correlation between blood glucose concentration and glucose concentration in subcutaneous adipose tissue evaluated with microdialysis during intensive care. *Scandinavian Journal of Clinical and Laboratory Investigation* 62: 285–292.
- Maran A, Crepaldi C, Tiengo A, et al. (2002) Continuous subcutaneous glucose monitoring in diabetic patients. *Diabetes Care* 25: 347–352.
- Marquette CA, Degiuli A, and Blum LJ (2003) Electrochemiluminescent biosensors array for the concomitant

detection of choline, glucose, glutamate, lactate, lysine and urate. *Biosensors and Bioelectronics* 19: 433–439.

Skeie S, Thue G, Nerhus K, and Sandberg S (2002) Instruments for self-monitoring of blood glucose: Comparisons

of testing quality achieved by patients and a technician. *Clinical Chemistry* 48(7): 994–1003.

Suzuki S and Honda S (2003) Miniaturization in carbohydrate analysis. *Electrophoresis* 24: 3577–3582.

## Sarcosine, Creatine, and Creatinine

**T Smith-Palmer**, St. Francis Xavier University, Antigonish, NS, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

#### Creatine and Creatinine

Creatine can be synthesized in the body from glycine, arginine, and methionine. Creatine can also be obtained from eating meat. There is ~120 g of creatine in the body of an average adult male, with a turnover rate of  $\sim 0.016 \text{ day}^{-1}$  for creatine. Vegetarians have actually been shown to have somewhat lower levels of creatine in their urine and serum than nonvegetarians. Over 90% of creatine in the body is found in the muscles, where it is converted to phosphocreatine in a reversible reaction with adenosine triphosphate (ATP), mediated by creatine kinase. When ATP is depleted during exercise, the ADP so formed is rephosphorylated in a reaction with the phosphocreatine. Phosphocreatine and creatine cyclize non-enzymically to form creatinine, which is then excreted by renal glomeruli and not reabsorbed by tubules. The amount of creatinine excreted in the urine is related to the amount of phosphocreatine present in the body, and is indicative of lean muscle mass. In a 24 h period, the amount of creatinine excreted is fairly constant for a particular subject. Thus, concentrations of other components of the urine may be normalized to the creatinine concentration. The clearance rate of creatinine is an important clinical diagnostic tool. For example, decreased excretion of creatinine is a symptom of muscle degeneration. Thus, the analysis of creatinine in urine is routine, and in a clinical setting is usually done on an automated system. Creatinuria can be an indication of a muscle-wasting disease such as muscular dystrophy, or be a result of creatine supplementation. Many athletes, body builders, and young males in general take creatine as a supplement to increase their muscle bulk and to increase their endurance during

resistance training. They typically start with a 5 day loading phase of ~20 g creatine per day. Much of this is excreted in the urine in the first 5 h after its ingestion. Creatine is found at low concentrations in the plasma and at higher concentrations in erythrocytes. Erythrocyte creatine concentrations decrease with the age of the cell and thus can be used to determine cell age. The role of the creatine in erythrocytes has not been defined, but the concentration can be increased by creatine supplementation.

Creatinine can be analyzed in automatic colorimetric analyzers using the Jaffe method, by gas chromatography–mass spectrometry (GC–MS) after derivatization, or simultaneously with creatine by high-performance liquid chromatography (HPLC) or capillary electrophoresis (CE). Enzyme-based methods for both creatine and creatinine are used for colorimetric analyses as well as in biosensors. Creatine can be converted to creatinine to be analyzed by the Jaffe reaction. Fluorescence analyses and methods involving partial least squares (PLS) with ultraviolet (UV), infrared (IR), or near-infrared (NIR) spectra can also be used.

#### Sarcosine

Sarcosine, or *N*-methyl-glycine, is an amino acid that forms as an intermediate in the metabolism of choline in the kidney and liver. It is formed by the methylation of glycine using *S*-adenosylmethionine and the enzyme *N*-methyltransferase. It is normally present in very low concentrations compared to other amino acids unless the subject is suffering from sarcosinemia. This is not a well-defined syndrome, but subjects often exhibit neurological problems. The cerebrospinal fluid of patients with Parkinson's disease and multiple sclerosis contains higher than normal levels of sarcosine.

Sarcosine is also formed in the hydrolysis of creatine mediated by creatine amidinohydrolase (creatinase). This reaction is used in a variety of sensors for creatine and creatinine.

Sarcosine can be analyzed by various amino acid analyses, although in some cases the detection limits



detection of choline, glucose, glutamate, lactate, lysine and urate. *Biosensors and Bioelectronics* 19: 433–439.

Skeie S, Thue G, Nerhus K, and Sandberg S (2002) Instruments for self-monitoring of blood glucose: Comparisons

of testing quality achieved by patients and a technician. *Clinical Chemistry* 48(7): 994–1003.

Suzuki S and Honda S (2003) Miniaturization in carbohydrate analysis. *Electrophoresis* 24: 3577–3582.

## Sarcosine, Creatine, and Creatinine

**T Smith-Palmer**, St. Francis Xavier University,  
Antigonish, NS, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

#### Creatine and Creatinine

Creatine can be synthesized in the body from glycine, arginine, and methionine. Creatine can also be obtained from eating meat. There is ~120 g of creatine in the body of an average adult male, with a turnover rate of  $\sim 0.016 \text{ day}^{-1}$  for creatine. Vegetarians have actually been shown to have somewhat lower levels of creatine in their urine and serum than nonvegetarians. Over 90% of creatine in the body is found in the muscles, where it is converted to phosphocreatine in a reversible reaction with adenosine triphosphate (ATP), mediated by creatine kinase. When ATP is depleted during exercise, the ADP so formed is rephosphorylated in a reaction with the phosphocreatine. Phosphocreatine and creatine cyclize non-enzymically to form creatinine, which is then excreted by renal glomeruli and not reabsorbed by tubules. The amount of creatinine excreted in the urine is related to the amount of phosphocreatine present in the body, and is indicative of lean muscle mass. In a 24 h period, the amount of creatinine excreted is fairly constant for a particular subject. Thus, concentrations of other components of the urine may be normalized to the creatinine concentration. The clearance rate of creatinine is an important clinical diagnostic tool. For example, decreased excretion of creatinine is a symptom of muscle degeneration. Thus, the analysis of creatinine in urine is routine, and in a clinical setting is usually done on an automated system. Creatinuria can be an indication of a muscle-wasting disease such as muscular dystrophy, or be a result of creatine supplementation. Many athletes, body builders, and young males in general take creatine as a supplement to increase their muscle bulk and to increase their endurance during

resistance training. They typically start with a 5 day loading phase of ~20 g creatine per day. Much of this is excreted in the urine in the first 5 h after its ingestion. Creatine is found at low concentrations in the plasma and at higher concentrations in erythrocytes. Erythrocyte creatine concentrations decrease with the age of the cell and thus can be used to determine cell age. The role of the creatine in erythrocytes has not been defined, but the concentration can be increased by creatine supplementation.

Creatinine can be analyzed in automatic colorimetric analyzers using the Jaffe method, by gas chromatography–mass spectrometry (GC–MS) after derivatization, or simultaneously with creatine by high-performance liquid chromatography (HPLC) or capillary electrophoresis (CE). Enzyme-based methods for both creatine and creatinine are used for colorimetric analyses as well as in biosensors. Creatine can be converted to creatinine to be analyzed by the Jaffe reaction. Fluorescence analyses and methods involving partial least squares (PLS) with ultraviolet (UV), infrared (IR), or near-infrared (NIR) spectra can also be used.

#### Sarcosine

Sarcosine, or *N*-methyl-glycine, is an amino acid that forms as an intermediate in the metabolism of choline in the kidney and liver. It is formed by the methylation of glycine using *S*-adenosylmethionine and the enzyme *N*-methyltransferase. It is normally present in very low concentrations compared to other amino acids unless the subject is suffering from sarcosinemia. This is not a well-defined syndrome, but subjects often exhibit neurological problems. The cerebrospinal fluid of patients with Parkinson's disease and multiple sclerosis contains higher than normal levels of sarcosine.

Sarcosine is also formed in the hydrolysis of creatine mediated by creatine amidinohydrolase (creatinase). This reaction is used in a variety of sensors for creatine and creatinine.

Sarcosine can be analyzed by various amino acid analyses, although in some cases the detection limits

only allow sarcosine to be detected in the serum of subjects with sarcosinemia. The best known amino acid analysis involves separation on a cation exchange column followed by derivatization with ninhydrin for detection by absorption. Sarcosine can also be derivatized and determined by liquid chromatography (LC) with absorption or fluorescence detection, or by GC with either flame ionization or mass spectrometric detectors. A biosensor has been made based on sarcosine oxidase immobilized in nafion, with amperometric detection.

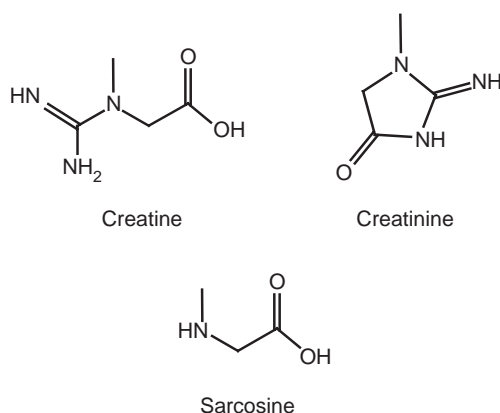
Creatine, creatinine, and sarcosine can be analyzed simultaneously by 500 MHz proton nuclear magnetic resonance (NMR).

### Properties of Creatine, Creatinine, and Sarcosine

The structures of creatinine, creatine, and sarcosine are shown in Figure 1. The  $pK_a$  values for creatinine are 4.8 and 9.2, for creatine are 2.6 and 14.3, and for sarcosine are 2.21 and 10.12.

Creatine and creatinine are in equilibrium with each other, but at room temperature at close to neutral pH the interconversion is slow. Conversion of creatine to creatinine is essentially complete if the solution is boiled in the presence of acid, while creatinine in basic conditions forms creatine. These conversions are not normally a problem during clinical analysis. In alkaline solutions of urine with high creatinine concentrations, conversion rates of  $\sim 5\%$  per hour may be found.

The absorbance of creatine is insignificant above 225 nm, but increases at lower wavelengths. It is generally monitored at 210 nm. Creatinine absorbs strongly at 210 nm, and slightly less strongly at 230 nm. Both wavelengths can be used for its determination. Uric acid and urea also absorb at these wavelengths. The absorptivity of urea is very low, although it may still cause slight interference because of its high concentrations.



**Figure 1** Structures of creatine, creatinine, and sarcosine.

### Reference Values

The normal range for serum creatinine is  $4\text{--}16\text{ mg l}^{-1}$  ( $35\text{--}140\text{ }\mu\text{mol l}^{-1}$ ). During kidney dysfunction this concentration may increase to  $1000\text{ }\mu\text{mol l}^{-1}$ . Values above  $140\text{ }\mu\text{mol l}^{-1}$  indicate that renal function needs further assessment. Values above  $530\text{ }\mu\text{mol l}^{-1}$  are indicative of severe renal problems. In urine,  $0.5\text{--}3.3\text{ g per day}$  is normally excreted.

Creatine concentrations in serum are generally less than  $5\text{ mg l}^{-1}$ . Increases are seen in muscle wasting, hyperthyroidism, and after supplementation.

The reference values for erythrocyte creatine are  $228\text{--}550\text{ }\mu\text{mol l}^{-1}$  for males and  $258\text{--}598\text{ }\mu\text{mol l}^{-1}$  for females. In urine, normal creatine concentrations range up to  $150\text{ mg per day}$ ; values as high as  $12\text{ g per day}$  are found when subjects are supplementing with creatine.

The reference values for sarcosine in urine are  $2\text{--}20\text{ }\mu\text{mol per day}$ , and in serum sarcosine concentrations of  $0.50\text{--}2.70\text{ }\mu\text{mol l}^{-1}$  are usual. In most analyses, sarcosine is just not detected in normal serum. In patients with sarcosinemia, the concentrations range up to  $603\text{ }\mu\text{mol l}^{-1}$  in serum and  $9.4\text{ }\mu\text{mol mg}^{-1}$  creatinine in urine.

### Sample Preparation

The presence of proteins in biological fluids constitutes an interference in many analyses and their removal is often the major part of sample preparation. Proteins are typically removed by precipitation with acetonitrile or sulfosalicylic acid. Ultrafiltration or ultracentrifugation can also be used.

The protein content of urine and cerebrospinal fluid is much lower than that of blood, and dilution is often the only sample treatment necessary. However, protein removal is necessary before reversed-phase column chromatography to protect the lifespan of the column. Sample dilution helps reduce the presence of proteins, but even running untreated urine samples diluted 50-fold can very quickly destroy a column.

## Absorption Methods

### Creatinine

In a clinical lab, creatinine is normally determined using a commercial analyzer utilizing an absorption method based on the Jaffe reaction. In the Jaffe reaction, creatinine is reacted with alkaline sodium picrate solution to form an orange creatinine-picrate complex. The absorbance at 510 nm can then be measured after a specified reaction time (endpoint method). There are, however, a variety of modifications to the basic method because of the positive

interference from species such as protein, glucose, ascorbic acid, and the ketone bodies (pyruvic, acetoacetic,  $\beta$ -hydroxybutyric, and acetone). The interference can be of the order of 20% in normal serum or plasma and higher if the subject is diabetic. The interference can also be higher in erythrocytes. Surfactants are sometimes incorporated in the reagent mix, or the reaction is monitored as a function of time (kinetic method). Cleanup of a sample on an ion-exchange column is effective but not practical for an automated procedure. One method that works well, and is used in some commercial kits, is acidification to pH 4 after the initial complex formation. The color due to complexed creatinine is destroyed, while that produced by interferences remains. The difference in the absorbance at 500 nm is proportional to the concentration of creatinine. The incorporation of dialysis in an automated flow method can also reduce interference.

*Note:* Picric acid is a strong oxidizing agent and is explosive when dry, so spills should be treated accordingly.

The second most common method used in automated analyzers is based on the use of enzymes to produce a compound that will react with a substance provided to form a colored product. For example, creatinine iminohydrolase (creatinine deiminase) catalyzes the conversion of creatinine to *N*-methylhydantoin and ammonia. This reaction can be carried out in a flow injection analysis system setup so that the ammonia diffuses into an acceptor stream containing a pH indicator; the color changes of which are measured by diffuse reflectance.

Creatinine amidohydrolase hydrolyzes creatinine to creatine, the creatine is hydrolyzed to sarcosine, urea by creatine amidohydrolase, and sarcosine oxidase catalyzes the breakdown of the sarcosine to glycine, formaldehyde, and peroxide. When horseradish peroxidase is included, the reduction of peroxide can be coupled with various compounds to produce absorbing species. One example of a suitable chromogen is 3,5-dichloro-2-hydroxybenzenesulfonic acid-4-aminophenazone, with the absorbance being measured at 510 nm.

Creatinine can also be determined by measuring the rate of disappearance of NADH after addition of creatinine amidohydrolase, creatine kinase, pyruvate kinase, and lactate dehydrogenase.

Although there are less interferences in the enzyme-based methods, the higher cost has prevented them from replacing the Jaffe method, although there are a number of commercial systems based on them. The limit of detection (LOD) is  $1 \text{ mg l}^{-1}$ .

Mid-IR and near-IR methods have been used in conjunction with PLS methods to determine

creatinine and other metabolites in blood and urine. The use of dry films solves the problem of water interference.

### Creatine

Traditionally, creatine has been determined by the Folin method whereby creatine is determined as creatinine by the Jaffe reaction after conversion in boiling acid for an hour, or autoclaving for 20 min at  $120^\circ\text{C}$ . Thus, one aliquot of a sample is analyzed for creatinine, and a second part is heated in acid prior to creatinine analysis. The difference between the two answers gives the concentration of creatine. Large errors are often reported. Also note that heating creatine for an hour in  $6 \text{ mol l}^{-1}$  HCl is reported to liberate sarcosine.

One reagent that has been used for the automated analysis of creatine is diacetyl-1-naphthol. However, some creatine is converted to creatinine during acid hydrolysis, and corrections must be made.

The preferred absorption-based methods are those based on enzyme-catalyzed reactions. The absorbance of NADH (340 nm) after addition of creatine amidohydrolase, sarcosine oxidase, formaldehyde dehydrogenase, and  $\text{NAD}^+$  to a creatine-containing solution is proportional to the concentration of creatine present. Creatinine amidohydrolase produces creatine from creatinine. Thus, any assays for creatinine which use creatinine amidohydrolase in the first step can be used for creatine by just leaving out that first step. Conversely, creatine can interfere in those analyses, although creatine is normally present in substantially lower concentrations than creatinine.

Creatine can be analyzed in diluted urine by a PLS method using UV absorption analysis, after calibration with a training set of urine samples analyzed using HPLC. The determination of creatinine in this way is not as reliable.

## Fluorescence

### Creatine

The fluorescence of a ninhydrin derivative of creatine can be used to determine creatine concentrations in blood and urine. Protein and arginine are precipitated by treatment with barium hydroxide and zinc sulfate, and after addition of ninhydrin the solution is made basic and the fluorescence excited at 390 nm is monitored at 495 nm. Creatinine and phosphocreatine do not form fluorescent derivatives, but some guanidine compounds can interfere, although they are not normally present in significant amounts.

Urine samples must be pretreated with anionic exchange resin to remove unidentified interfering species that are not a problem when analyzing blood.

### Sarcosine

Sarcosine is a secondary amine and reacts with fluorescamine at pH 12 to give a nonfluorescent aminoenone. Subsequent reaction with the primary amine (L-Leu-L-Ala) for 10 min at 70°C produces a pyrolinone that can be excited at 390 nm to fluoresce at 480 nm. Unreacted fluorescamine is rapidly hydrolyzed at pH 12 and thus does not interfere by reacting with the primary amine. The LOD is  $\sim 10 \mu\text{mol l}^{-1}$ .

### Biosensors

Sensors for creatine, creatinine, and sarcosine generally involve the use of enzymes (Table 1) and can be based on potentiometry or amperometry.

#### Creatinine

Because tests for creatinine are routine in a clinical setting, there has been much work done on devising sensors for creatinine in urine, serum/plasma, and hemodialysate. Such sensors would allow point of care (POC) testing which is becoming more common in places such as emergency departments. Some commercial POC analyzers are already available. A typical one such as an i-STAT<sup>®</sup> is handheld, with the blood sample being placed on a cartridge for insertion. The sensor (microfabricated thin film electrode) is calibrated automatically by means of an internal standard solution immediately before each analysis. The NOVA 16 benchtop analyzer incorporates determinations of  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Cl}^-$ , glucose, urea

nitrogen, and hematocrit along with the analysis of creatinine.

To create a potentiometric creatinine sensor, creatinine iminohydrolase, which catalyzes the production of ammonia from creatinine, can be immobilized on the surface of an ammonium ion selective electrode. There is no interference from creatine but some from the ammonium ions in blood and urine.

Another biosensor is based on an ion-sensitive field-effect transmitter. Creatinine iminohydrolase is immobilized on the gate by cross-linking with bovine serum albumin in glutaraldehyde vapor. Creatinine iminohydrolase catalyzes the formation of *N*-methylhydantoin (and ammonia) from creatinine. The differential signal between a reference and sensing chip is proportional to creatinine concentration.

#### Creatine and Creatinine

When the three-enzyme sequence based on creatinine amidohydrolase is used, any creatine present can interfere with the determination of creatinine, so two sensors are used: one to determine the total creatine plus creatinine and one to determine just creatine (by only using creatine amidinohydrolase and sarcosine oxidase). Creatinine is determined by difference. Amperometric sensors are generally based on this sequence and do not suffer from interferences. They are usually designed to respond to peroxide, though some have used oxygen electrodes. Typically, Pt electrodes are used. A sensor for just creatine only requires the creatine amidinohydrolase and sarcosine oxidase sequence.

One configuration that can be used to amperometrically detect the peroxide produced from the oxidation of sarcosine is a Pt working electrode, a Ag/AgCl reference electrode, and a carbon counter electrode screen-printed onto a polyester sheet. The enzymes can be immobilized on the working

**Table 1** Enzymes used for the analysis of creatine and creatinine, including the final species detected

<i>Enzyme 1</i>	<i>Enzyme 2</i>	<i>Enzyme 3</i>	<i>Enzyme 4</i>	<i>Species detected</i>
Creatinine iminohydrolase	<i>N</i> -Methylhydantoin amidohydrolase	Pyruvate kinase	Lactate dehydrogenase	$\text{NAD}^+$
Creatinine iminohydrolase	<i>N</i> -Methylhydantoin amidohydrolase	<i>N</i> -Carbamoylsarcosine amidohydrolase	Sarcosine oxidase	Red product of peroxidase on 4-aminoantipyrine and phenol
Creatinine iminohydrolase				Ammonia
Creatinine amidohydrolase	Creatine amidinohydrolase	Sarcosine oxidase	Horseradish peroxidase	3,5-Dichloro-2-hydroxybenzenesulfonic acid/4-aminophenazone
Creatinine amidohydrolase	Creatine kinase	Pyruvate kinase	Lactate dehydrogenase	NADH
Creatine amidinohydrolase	Sarcosine oxidase	Horseradish peroxidase		Methylene blue
Creatine amidinohydrolase	Sarcosine oxidase	Horseradish peroxidase		Benzoquinone-imine



electrode in a poly(carbamoyl sulfonate) hydrogel. The working electrode is first coated with a polymer such as nafion to exclude interfering substances. In a more elegant design, the enzymes are dissolved in phosphate buffer and incorporated into a paste of graphite powder and paraffin oil. The paste is inserted into plastic tips 3 mm in diameter and the exposed tip is polished. Silver wires are used for electrical contact. No membranes are needed to prevent fouling and the preparation of the electrodes is fast and straight forward. The response characteristics of a three-enzyme system (creatinine + creatine) were best at 650 mV, with an analytical range of 4–100 nmol l<sup>-1</sup>. With a two-enzyme system, the assay for creatine was optimal at 240 mV with a range of 4–200 pmol l<sup>-1</sup>. The actual concentration of creatinine is determined from the difference between the two readings.

Creatine and creatinine, along with *p*-aminohippuric acid and uric acid, can be determined using CE on a glass microchip with amperometric detection via screen-printed electrodes connected to an electrochemical analyzer. Creatinase, creatininase, and sarcosine oxidase are mixed with the sample, so that the actual separation is of peroxide, *p*-aminohippuric acid, and uric acid. This allows total (creatine plus creatinine) to be determined. The concentration of creatine can be determined in the absence of creatininase.

### Sarcosine

A biosensor for sarcosine can be constructed based on sarcosine oxidase entrapped in nafion, with amperometric detection using manganese dioxide-modified screen-printed electrodes. The LOD is 28 μmol l<sup>-1</sup>.

## Thin-Layer Chromatography

### Sarcosine

Sarcosine can be fluorescently labeled using 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole (NBD-Cl). This reaction facilitates quantitation of sarcosine after separation by thin-layer chromatography (TLC). TLC concentrates the species of interest into a limited area which is then scanned for fluorescence. This procedure gives extremely low detection limits (low picomoles), allowing quantitation of sarcosine even in serum samples from healthy subjects. However, ions that interfere in the derivatization step must be removed on an ion-exchange column prior to derivatization, and the sarcosine is best pre-separated from large amounts of other amino acids, which are present in sera. Thus, serum must be deproteinized using sulfosalicylic acid, then the sarcosine separated out using an amino acid analyzer connected to a

fraction collector. The sarcosine fraction is then deionized by ion-exchange chromatography before derivatization. Urine is deproteinized using ethanol before deionization and derivatization. Thus, considerable sample preparation is necessary, but the low LOD warrants this if more expensive equipment is not available.

## Liquid Column Chromatography

### Creatine and Creatinine

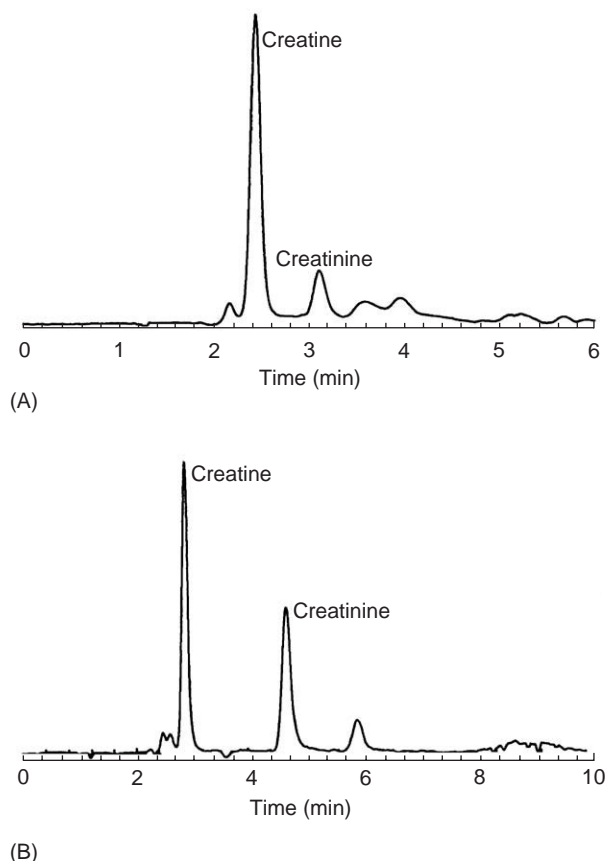
Chromatographic separations allow the simultaneous determination of several species of interest. Thus, there is a wide range of systems that can be used, and the method selection will be determined, to a large extent, by what other compounds need to be determined at the same time. This also dictates the detection method. The absorption of creatine is typically measured at 210 nm, while creatinine can be measured at both 210 and 230 nm and has a higher absorptivity. When amino acids or guanidino compounds are to be analyzed simultaneously with creatine and creatinine, postcolumn detection using ninhydrin can be used. Enzymatic methods can be used if coeluting compounds are a problem. HPLC methods can be automated, and are particularly useful when both creatine and creatinine need to be analyzed in urine, plasma, erythrocytes, or tissue (Figure 2).

The sample preparation varies from dilution to protein precipitation to use of a precolumn. Removal of protein is important for extending the lifetime of reversed-phase columns, although it does not affect the analysis *per se*. Urea is present in urine in high concentrations but does not absorb strongly, while uric acid and guanidinoacetate are present in relatively small amounts but absorb similarly to creatine and creatinine. These compounds can interfere with analyses of creatine and creatinine on reversed-phase columns, and their relative retention times are highly dependent on the specific makeup of the reversed-phase column being used, as well as on the eluent. Thus, a separation needs to be optimized for the particular reversed-phase column that is installed, so that these species are resolved from the species of interest.

Most eluents for the separation of creatine and creatinine on a reversed-phase column are based on a phosphate buffer (Table 2). The addition of a quaternary ammonium ion-pairing agent such as tetrabutyl ammonium sulfate increases the hold-up of interfering species to help separate them from creatine. This method is one of the most popular for simultaneous determination of creatine and creatinine. Porous graphitic columns can also be used.

## Sarcosine

Sarcosine can be analyzed by many of the standard amino acid analyzers. Typically in these, amino acids



**Figure 2** Chromatograms of plasma (A) and urine (B) from subjects supplementing with creatine. The separations were carried out on a Synergi 10 $\mu$  Hydro RP 80 250  $\times$  4.6 mm column using a 20 mmol l<sup>-1</sup> potassium phosphate buffer, pH 6.5, for the plasma and 14.7 mmol l<sup>-1</sup> potassium phosphate–2.3 mmol l<sup>-1</sup> tetrabutyl ammonium sulfate, pH 5.0, for the urine, with detection by absorption at 210 nm. For both samples, protein was precipitated out before injection using acetonitrile. The plasma and urine were diluted 5-fold and 87-fold, respectively, during sample preparation.

are separated by cation-exchange chromatography and must be derivatized for detection. There is little interference from contaminants, and minimal sample preparation is required. The most well-known derivatizing agent for amino acids is ninhydrin, which gives blue–purple derivatives (absorption measured at 570 nm) for the majority of amino acids, but for a few, including sarcosine, the derivative is purple–brown. Cation-exchange chromatography is favored because there are few matrix effects. Run times depend on the mix of amino acids and other compounds present and the particular eluent used. One common eluent system for the analysis of physiological samples is lithium citrate–chloride, which is used with a cation-exchange stationary phase regenerated with LiOH. Gradient elution is necessary for an efficient analysis. Elution generally begins at pH 2.8, and the pH and concentration of lithium citrate are increased. Buffers and columns are commercially available, with gradients customized for columns and the particular mix of ninhydrin-active species present in the samples. The initial portion of a chromatogram of a mix of amino acids including sarcosine is shown in Figure 3. Sarcosine of concentrations down to 5  $\mu$ mol l<sup>-1</sup> can be analyzed in urine or plasma. Accordingly, sarcosine does not show up in normal plasma when analyzed in this way. Although ninhydrin reagent is not very stable, there are various modifications commercially available that attempt to address this problem. For example, Trione<sup>®</sup> gives a more stable but less sensitive derivative.

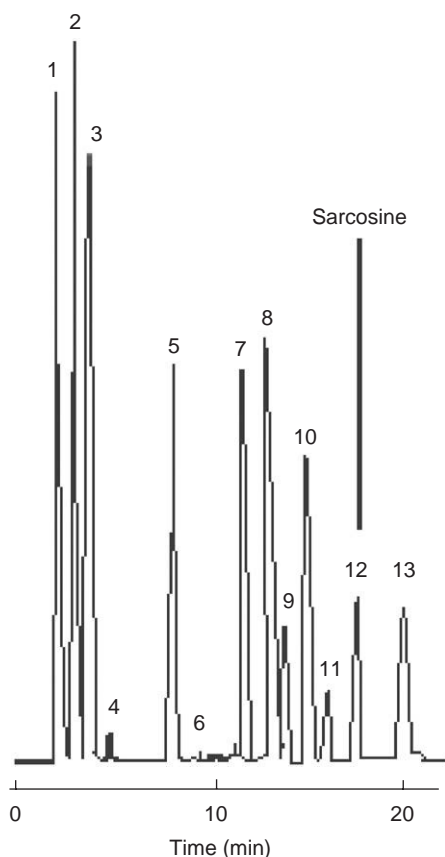
Pulsed amperometric detection offers a significant advantage in that it can be used after separation of the ions by ion chromatography without derivatization, and has detection limits  $\sim$ 50 times lower than those of ninhydrin.

Precolumn derivatization of amino acids causes them to become less polar and more amenable to separation by reversed-phase HPLC. Such analyses tend to give shorter run times and higher sensitivity

**Table 2** HPLC buffers used for the analysis of creatine and creatinine

Matrix	Column	Eluent	Other compounds analyzed
Tissue, urine	Reversed phase	Potassium phosphate (14.7 mmol l <sup>-1</sup> ), tetrabutylammonium sulfate (2.3 mmol l <sup>-1</sup> ), pH 5.0	Phosphocreatine
Urine	Reversed phase	Potassium phosphate (20 mmol l <sup>-1</sup> ), pH 6.5	Uric acid, hippuric acid
Serum	Reversed phase	(Sodium phosphate (100 mmol l <sup>-1</sup> ), SDS (30 mmol l <sup>-1</sup> )):acetonitrile, 3:1, pH 2.1	Guanidino compounds
Urine	Reversed phase coated with hexadecylsulfonate	EDTA (5 mmol l <sup>-1</sup> ), phosphoric acid (15 mmol l <sup>-1</sup> ), pH 2.5	Amino acids
Serum	Porous graphite	Sodium citrate (10 mmol l <sup>-1</sup> ), sodium octanesulfonate (5 mmol l <sup>-1</sup> )	Guanidino compounds
Serum, urine	Porous graphite	Trifluoroacetic acid (0.1%), acetonitrile (3%)	





**Figure 3** Chromatogram of standard amino acids. The separation was carried out on an  $80 \times 5.5$  mm accelerated lithium cation-exchange column using 100% Pickering™ Li280 for the first 8 min and then a gradient to 23% Pickering™ Li750 over the next 14 min. Detection was by absorption at 570 nm. The amino acids in the mix are 1, phosphoserine; 2, taurine; 3, phosphoethanolamine; 4, urea; 5, aspartic acid; 6, hydroxyproline; 7, threonine; 8, serine; 9, asparagine; 10, glutamic acid; 11, glutamine; 12, sarcosine; 13,  $\alpha$ -aminoadipic acid. (Adapted from website [www.pickeringlabs.com](http://www.pickeringlabs.com) for Pickering Labs.)

than methods involving postcolumn derivatization. One common precolumn derivatizing agent for amino acids is *o*-phthaldialdehyde, but this does not react with secondary amino acids such as sarcosine. On the other hand, phenylisothiocyanate (PITC), also known as Edman reagent, reacts with both primary and secondary amino acids, forming phenylthiocarbamyl (PTC) derivatives that are analyzed at 254 nm. Pico tag columns are recommended. Unreacted PTC must be removed under vacuum or by extraction, but the whole derivatization sequence can be automated. The separation of the PTC-aa derivatives can then be carried with gradient elution using  $0.14 \text{ mol l}^{-1}$  sodium acetate, pH 6.4, with increasing concentrations of acetonitrile. Triethylamine and ethylenediaminetetraacetic acid (EDTA) are sometimes added. The exact program required will

depend on the column particulars and the components present. Sarcosine elutes between asparagine and  $\beta$ -alanine. This analysis can be used for blood and urine samples, with linear ranges from 20 to 5000 pmol.

PITC derivatives can also be detected electrochemically using an operating potential of 1.1 V. This gives lower detection limits than absorbance, and the sample preparation can be shorter and simpler.

Derivatization of sarcosine with 9-fluorenylmethyl chloroformate (FMOC) facilitates detection via fluorescence, but unfortunately the sarcosine derivatives are among the least stable of the amino acid derivatives formed this way, and the instability can lead to peak broadening.

6-Aminoquinolyl-*N*-hydrosuccinimidyl carbamate reacts quickly with amino acids in a simple procedure to give stable derivatives that can be detected by fluorescence or absorption after separation on a reversed phase at 37°C. The reagents and buffers are commercially available (AccQ-TAG).

## Capillary Electrophoresis

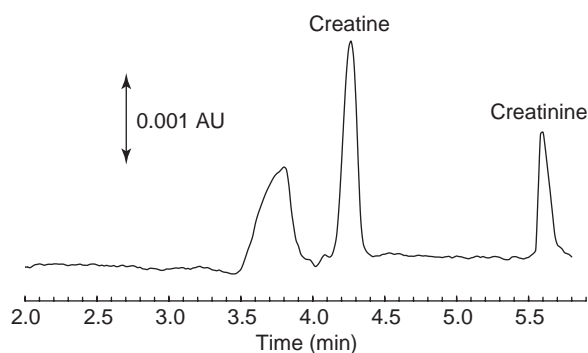
### Creatinine

Creatinine itself can be analyzed in 4 min using CE with a pH 6.4 phosphate buffer and detection by absorption at 235 nm. EDTA is added to prevent metal ions from interfering by binding with the creatinine. In a pH 4.05 buffer containing pyridine, tartaric acid, and 18-crown-6, cations can be analyzed along with the creatinine, using indirect photometric detection. To ensure reproducibility in CE when biological fluids are being analyzed, the capillary must be rinsed with base between runs (then water and buffer) to ensure that any proteins that may have adsorbed on the capillary wall are removed.

Addition of sodium dodecyl sulfate (SDS) in concentrations greater than that required for micelle formation ensures holdup of species that may interfere, and facilitates separation of other species in the sample that may be of interest. This is called micellar electrokinetic capillary chromatography (MECK). Apart from requiring a very small sample, an advantage of CE is that once the compound/s of interest has/have eluted, the remainder of the sample can be rapidly rinsed from the column using pressure or suction.

### Creatine and Creatinine

Baseline resolution of creatine, creatinine, and uric acid can be obtained using a pH 5.5,  $50 \text{ mmol l}^{-1}$  phosphate buffer. Best results are obtained when the



**Figure 4** Electropherogram of 100-fold dilution of urine from a subject supplementing with creatine. The separation was carried out at 17 kV with hydrostatic injection for 20 s, using a  $30 \text{ mmol l}^{-1}$  phosphate– $150 \text{ mmol l}^{-1}$  sodium dodecyl sulfate buffer, pH 6, and detection by absorbance at 214 nm. The unlabeled peak is a system peak.

sample is diluted with buffer containing 2% EDTA. MECK using  $150 \text{ mmol l}^{-1}$  SDS in a  $30 \text{ mmol l}^{-1}$  phosphate buffer, pH 6, can also be used for the analysis of creatine and creatinine (Figure 4). Migration times depend on the length of capillary and the applied voltage. The migration time for creatinine is particularly sensitive to pH, decreasing as the pH is raised. The LOD depends on the instrument used and may be as low as  $0.7 \mu\text{g ml}^{-1}$  for creatine and  $0.09 \mu\text{g ml}^{-1}$  for creatinine.

## Gas Chromatography

Creatine, creatinine, and sarcosine must all be derivatized before they can pass through a gas chromatograph.

### Creatine and Creatinine

Unfortunately, in most cases creatine and creatinine form the same derivatives, and so the two must be separated before derivatization. This can be accomplished using a weak cation-exchange column. Creatine elutes in water while creatinine elutes in  $1 \text{ mol l}^{-1}$  ammonia. Suitable derivatizing agents are shown in Table 3.

### Sarcosine

An <sup>TM</sup>EZfaast kit facilitates the cleanup and derivatization of mixtures of primary and secondary amino acids. The derivatives can then be analyzed by GC–FID or GC–MS. The latter method gives lower LODs.

Other derivatizing agents that can be used are shown in Table 4.

**Table 3** Derivatives of creatinine for analysis by GC–MS

Derivative	Reagent(s)
(2-Hydroxy, 2-methyl)ethylcreatinine di(trifluoroacetate)	1,2-Epoxypropane then trifluoroacetic acid
Ethyl ester of <i>N</i> -(4,6-dimethyl-2-pyrimidinyl)- <i>N</i> -methylglycine	2,4-Pentanedione, ethanol, acetic acid
Trimethylsilylcreatinine	<i>N</i> -Methyl- <i>N</i> -trimethylsilyltrifluoroacetic amide (MSTFA)
<i>O</i> -Trifluoroacetylcreatinine	Trifluoroacetic anhydride

**Table 4** Derivatives of sarcosine for analysis by GC–MS

Derivative	Reagent(s)
<i>N</i> ( <i>O</i> )- <i>tert</i> -Butyldimethylsilyl sarcosine	<i>N</i> -Methyl- <i>N</i> -( <i>tert</i> -butyldimethylsilyl)trifluoroacetamide and <i>N,N</i> -dimethylformamide
<i>N</i> -Pentafluoropropionyl hexafluoroisopropyl ester	Pentafluoropropionic anhydride and hexafluoroisopropanol (2:1)
Mono- and di-trimethylsilylsarcosine	<i>N,O</i> -bis(trimethylsilyl) trifluoroacetamide (BSTFA)

## Mass Spectrometry

GC–MS gives very low detection limits and positively identifies the presence of a species of interest. Internal standards are often used, typically an isotopically labeled species.

### Creatinine

Isotope dilution analyses give very accurate results but are not suitable for routine analysis. Standards can be labeled with  $^{15}\text{N}$ , deuterium,  $^{13}\text{C}$ , or a combination. Several isotope dilution-GC–MS methods for creatinine have been proposed as reference methods.

LC–APCI–MS can also be used to determine creatine and creatinine. However, the sample must first be cleaned up by ion-exchange chromatography, and the LOD is not as low as for some simpler methods. Urinary creatine, creatinine, and guanidinoacetate can be determined by HPLC–MS using 2% acetonitrile in  $100 \text{ mmol l}^{-1}$  ammonium acetate as eluent.

### Sarcosine

Sarcosine can be determined along with other amino acids by GC–MS after derivatization. Using single-ion monitoring, LODs of 50 fmol or lower can be obtained with sample sizes of a few hundred microliters. Derivatization by methylation with diazomethane is not suitable as glycine could be methylated to sarcosine, and produce a positive interference.

## Nuclear Magnetic Resonance

### Creatine, Creatinine, and Sarcosine

Proton NMR can be used to examine the composition of urine.

Using a 500 MHz instrument and solvent ( $\text{H}_2\text{O}$ ) suppression, the spectra are collected using 0.5 ml urine with  $50\ \mu\text{l}\ 20\ \text{mol l}^{-1}$  sodium 2,2,3,3-(trimethylsilyl)tetradeuteriopropionate in  $\text{D}_2\text{O}$ . Peaks are assigned by comparison to standards and by use of various pulse sequences. Peaks can be seen for creatine, creatinine, and sarcosine. Creatine is generally elevated in the urine of infants. Sarcosine concentrations in the urine of normal healthy patients are too low to be seen by this method.

$^{13}\text{C}$  NMR spectroscopy can be used to follow the metabolism of  $[2-^{13}\text{C}]$ creatine in tissue culture.

**See also:** **Amperometry.** Blood and Plasma. **Clinical Analysis:** Sample Handling; Inborn Errors of Metabolism. **Fluorescence:** Derivatization; Fluorescence Labeling. **Gas Chromatography:** Mass Spectrometry. **Isotope Dilution Analysis.** **Liquid Chromatography:** Clinical Applications.

### Further Reading

- Cuka S, Dvornik S, Drazenovic K, and Mihic J (2001) Evaluation of the Dade Behring Dimension RxL clinical chemistry analyzer. *Clinical Laboratory* 47: 35–40.
- Erlenkötter A, Fobker M, and Chemnitus G-C (2001) Biosensors and flow-through system for the determination of creatinine in hemodialysate. *Analytical Bioanalytical Chemistry* 372: 284–303.
- Grunau JA and Swiader JM (1992) Chromatography of 99 amino acids and other ninhydrin-reactive compounds in the pickering lithium gradient system. *Journal of Chromatography* 594: 165–171.
- Hill LM, Pink J, Kaiser C, Burke DG, and Smith-Palmer T (2003) Multivariate calibration for the determination of creatine in urine using UV absorption spectroscopy. *Canadian Journal of Analytical Sciences and Spectroscopy* 48: 157–161.
- Killard JK and Smyth MR (2000) Creatinine biosensors: principles and designs. *Tibtech* 18: 433–437.
- Kochansky CJ and Strein TJ (2000) Determination of uremic toxins in biofluids: Creatinine, creatine, uric acid and xanthines. *Journal of Chromatography B* 747: 217–227.
- Kotzian P, Beyene NW, Llano LF, et al. (2002) Amperometric determination of sarcosine with sarcosine oxidase entrapped with nafion on manganese dioxide-modified screen-printed electrodes. *Scientific Papers of the University of Pardubice, Series A: Faculty of Chemical Technology* 8: 93–101.
- Le Boucher J, Charret C, Coudray-Lucas C, Giboudeau J, and Cynober L (1997) Amino acid determination in biological fluids by automated ion-exchange chromatography: Performance of Hitachi L-8500A. *Clinical Chemistry* 43: 1421–1428.
- Molnár-Perl I (1994) Advances in the high-performance liquid chromatographic determination of phenylthiocarbamyl amino acids. *Journal of Chromatography A* 661: 43–50.
- Shaw RA, Low-Ying S, Leroux M, and Mantsch HH (2000) Toward reagent-free clinical analysis: quantitation of urine urea, creatinine, and total protein from the mid-infrared spectra of dried urine films. *Clinical Chemistry* 46: 1493–1495.
- Smith-Palmer T (2002) Separation methods applicable to urinary creatine and creatinine. *Journal of Chromatography B* 781: 93–106.
- Stefan R-I, Bokretson RG, van Staden JF, and Aboul-Enein HY (2003) Simultaneous determination of creatine and creatinine using amperometric biosensors. *Talanta* 60: 1223–1228.
- Walsh DA and Dempsey E (2002) Comparison of electrochemical, electrophoretic, and spectrophotometric methods for creatinine determination in biological fluids. *Analytical Chimica Acta* 459: 187–198.

## Inborn Errors of Metabolism

**D H Chace**, Pediatrix Screening, Bridgeville, PA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

There are ~1300 different types of metabolic diseases that have been found in children. Individually, these diseases are quite rare; combined, they are not. Statistics from a subset of inherited disorders of

metabolism (55 disorders analyzed by newborn screening) reveal a combined incidence of more than 1 in 1000 in a study population of 1.5 million. A vast array of methods has been used to measure diagnostic metabolites of these disorders in many biological fluids from newborns, infants, children, and adults. Many of these diseases share common symptoms including retarded growth and mental development, failure to thrive, abnormal electrolytes, etc., and produce severe disabilities, coma, or death if not

## Nuclear Magnetic Resonance

### Creatine, Creatinine, and Sarcosine

Proton NMR can be used to examine the composition of urine.

Using a 500 MHz instrument and solvent ( $\text{H}_2\text{O}$ ) suppression, the spectra are collected using 0.5 ml urine with  $50\ \mu\text{l}\ 20\ \text{mol l}^{-1}$  sodium 2,2,3,3-(trimethylsilyl)tetradeuteriopropionate in  $\text{D}_2\text{O}$ . Peaks are assigned by comparison to standards and by use of various pulse sequences. Peaks can be seen for creatine, creatinine, and sarcosine. Creatine is generally elevated in the urine of infants. Sarcosine concentrations in the urine of normal healthy patients are too low to be seen by this method.

$^{13}\text{C}$  NMR spectroscopy can be used to follow the metabolism of  $[2-^{13}\text{C}]$ creatine in tissue culture.

**See also:** **Amperometry.** Blood and Plasma. **Clinical Analysis:** Sample Handling; Inborn Errors of Metabolism. **Fluorescence:** Derivatization; Fluorescence Labeling. **Gas Chromatography:** Mass Spectrometry. **Isotope Dilution Analysis.** **Liquid Chromatography:** Clinical Applications.

### Further Reading

- Cuka S, Dvornik S, Drazenovic K, and Mihic J (2001) Evaluation of the Dade Behring Dimension RxL clinical chemistry analyzer. *Clinical Laboratory* 47: 35–40.
- Erlenkötter A, Fobker M, and Chemnitz G-C (2001) Biosensors and flow-through system for the determination of creatinine in hemodialysate. *Analytical Bioanalytical Chemistry* 372: 284–303.
- Grunau JA and Swiader JM (1992) Chromatography of 99 amino acids and other ninhydrin-reactive compounds in the pickering lithium gradient system. *Journal of Chromatography* 594: 165–171.
- Hill LM, Pink J, Kaiser C, Burke DG, and Smith-Palmer T (2003) Multivariate calibration for the determination of creatine in urine using UV absorption spectroscopy. *Canadian Journal of Analytical Sciences and Spectroscopy* 48: 157–161.
- Killard JK and Smyth MR (2000) Creatinine biosensors: principles and designs. *Tibtech* 18: 433–437.
- Kochansky CJ and Strein TJ (2000) Determination of uremic toxins in biofluids: Creatinine, creatine, uric acid and xanthines. *Journal of Chromatography B* 747: 217–227.
- Kotzian P, Beyene NW, Llano LF, et al. (2002) Amperometric determination of sarcosine with sarcosine oxidase entrapped with nafion on manganese dioxide-modified screen-printed electrodes. *Scientific Papers of the University of Pardubice, Series A: Faculty of Chemical Technology* 8: 93–101.
- Le Boucher J, Charret C, Coudray-Lucas C, Giboudeau J, and Cynober L (1997) Amino acid determination in biological fluids by automated ion-exchange chromatography: Performance of Hitachi L-8500A. *Clinical Chemistry* 43: 1421–1428.
- Molnár-Perl I (1994) Advances in the high-performance liquid chromatographic determination of phenylthiocarbamyl amino acids. *Journal of Chromatography A* 661: 43–50.
- Shaw RA, Low-Ying S, Leroux M, and Mantsch HH (2000) Toward reagent-free clinical analysis: quantitation of urine urea, creatinine, and total protein from the mid-infrared spectra of dried urine films. *Clinical Chemistry* 46: 1493–1495.
- Smith-Palmer T (2002) Separation methods applicable to urinary creatine and creatinine. *Journal of Chromatography B* 781: 93–106.
- Stefan R-I, Bokretson RG, van Staden JF, and Aboul-Enein HY (2003) Simultaneous determination of creatine and creatinine using amperometric biosensors. *Talanta* 60: 1223–1228.
- Walsh DA and Dempsey E (2002) Comparison of electrochemical, electrophoretic, and spectrophotometric methods for creatinine determination in biological fluids. *Analytical Chimica Acta* 459: 187–198.

## Inborn Errors of Metabolism

**D H Chace**, Pediatrix Screening, Bridgeville, PA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

There are ~1300 different types of metabolic diseases that have been found in children. Individually, these diseases are quite rare; combined, they are not. Statistics from a subset of inherited disorders of

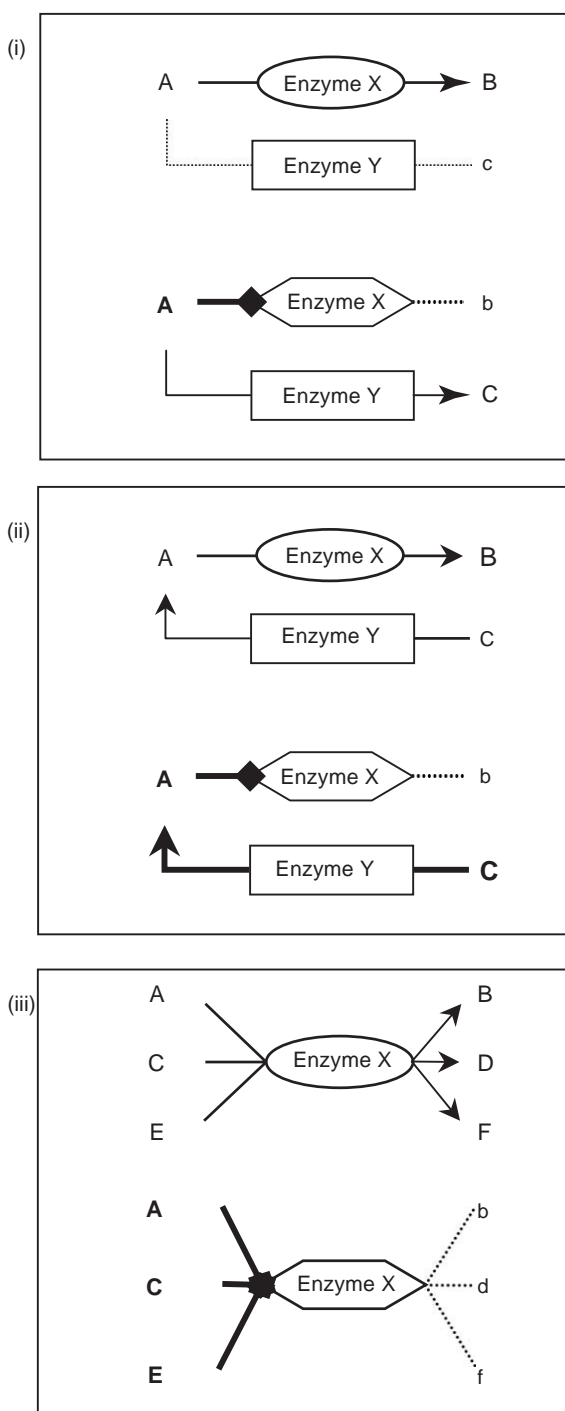
metabolism (55 disorders analyzed by newborn screening) reveal a combined incidence of more than 1 in 1000 in a study population of 1.5 million. A vast array of methods has been used to measure diagnostic metabolites of these disorders in many biological fluids from newborns, infants, children, and adults. Many of these diseases share common symptoms including retarded growth and mental development, failure to thrive, abnormal electrolytes, etc., and produce severe disabilities, coma, or death if not

treated early or appropriately. Historically, clinical methods were designed to detect disease in affected children with symptoms. The analysis performed was based on a suspected diagnosis and often was a single analyte test with a few notable exceptions. Recently, newer technology has enabled multianalyte testing for a variety of disorders in symptom-free neonates. This multiplex approach is cost-effective for the rare disorders that generally characterize inborn errors of metabolism (IEM). The primary focus of this article will be the accurate multianalyte techniques (e.g., mass spectrometry) rapidly replacing outdated older technology.

## Metabolites: Products and Substrates of Enzymes

An inborn error of metabolism presumes that a protein that comprises an enzyme or enzyme complex is abnormal. The kinetics/rates of metabolism of this altered protein/enzyme may be severely impaired or only mildly affected depending on how the protein is altered. Genetic mutations that directly alter its active site or interaction with co-factors would have deleterious consequences to enzyme activity as compared to minor alterations in protein structure or in less critical areas of protein structure and function. From an analytical perspective, alterations of these enzymes directly affect the concentration of their precursors and products. It might be deduced that enzymes with little activity would produce the most dramatic changes in the concentrations of metabolites. However, in these cases, the metabolites produced in high or low concentration have minimal effect on overall metabolism and may even have no consequence unless stressed by environmental or dietary circumstance. What is important to note is the simple fact that the concentration of metabolites is more closely correlated with a disease state than are abnormal enzymes and the genes that code them. The burden of diagnosing and confirming a disease requires the analysis of metabolites. The absolute concentration of a metabolite and its relative concentration to other metabolites will determine the likelihood or certainty of a metabolic disorder. The preliminary analytical results will subsequently influence the manner of further metabolic investigation and confirmation of the disease leading to ultimately, a final diagnosis.

A schematic illustration of the relationships of metabolites in various metabolic disorders is presented in Figure 1. In a simple metabolic system (Figure 1, panel i), metabolite A is converted to metabolite B by enzyme X. In presumably healthy states, the concentrations of A and B will be within a



**Figure 1** Illustrations showing substrates and products of enzymatic pathways involved in inborn errors of metabolism. Three panels (i, ii, iii) contain both normal and abnormal substrate (A) metabolism to product (B). Larger upper case bold fonts indicate higher concentrations than either normal (upper case fonts) or low (lower case, smaller fonts) concentrations. Panel (i) represents a simple scenario of metabolism of substrate A to B with alternate minor pathway c. Panel (ii) represents a scenario of substrate C metabolism to product A, which is a substrate for the enzyme X forming product B. Panel (iii) represents a scenario of substrates A, C, and E metabolized to products B, D, and F by the same enzyme X.



range considered 'normal'. The metabolite A may have alternative, minor metabolic pathways. This pathway involves enzyme Y that converts metabolite A to metabolite C. In a disease state where the activity of enzyme X is impaired, the concentration of A (the substrate) is increased, while the concentration of B (the product) is decreased. With a substantially elevated concentration of A, kinetics dictates that the alternate pathway will be utilized to a significant degree, producing increased concentrations of C. An example of this system may be observed in phenylketonuria where a deficiency of the enzyme phenylalanine hydroxylase produces a substantially increased concentration of phenylalanine while not producing the product tyrosine. From an analytical perspective, the analysis of a body fluid, blood or plasma, for example, would show an increased concentration of Phe and a possible decrease in Tyr. Further, very high levels of Phe are metabolized by alternative enzyme pathways to produce phenylpyruvic and phenylacetic acids. These metabolites are toxic in high concentration and produce brain damage.

A second metabolic system similar to the first is shown in **Figure 1**, panel ii. In this schematic, the emphasis is a metabolite one or two steps proximal to the primary enzyme deficiency. For example, in normal individuals, metabolite C is converted by enzyme Y to metabolite A, which is subsequently converted by enzyme X to metabolite B. As described above, in individuals with an inherited deficiency of enzyme X, metabolite A accumulates. In this particular scenario, compounds that are converted to metabolite A, namely metabolite C, will increase as enzyme Y is inhibited by basic kinetics. An example of this enzyme system is homocystinuria. In this disorder, the metabolism of homocysteine to cystathionine by cystathionine  $\beta$ -synthase is blocked. An increase in homocysteine causes an accumulation of S-adenosyl homocysteine and S-adenosyl methionine. Due to an increase in these metabolites, the metabolism of methionine to S-adenosyl methionine by methionine adenosyl transferase is decreased. Hence, methionine increases in the blood of individuals with homocystinuria. Note that in this example there were two enzymatic steps before the metabolism of homocysteine.

In the final example of metabolic disorders, **Figure 1**, panel iii, more than one metabolite are metabolized by a single enzyme system. In this instance, enzyme X converts three different compounds to three different substrates, A to B, C to D, and E to F. If a deficient enzyme X is present, all three substrates, A, C and E, are expected to increase while substrates B, D, and F are expected to decrease. An excellent

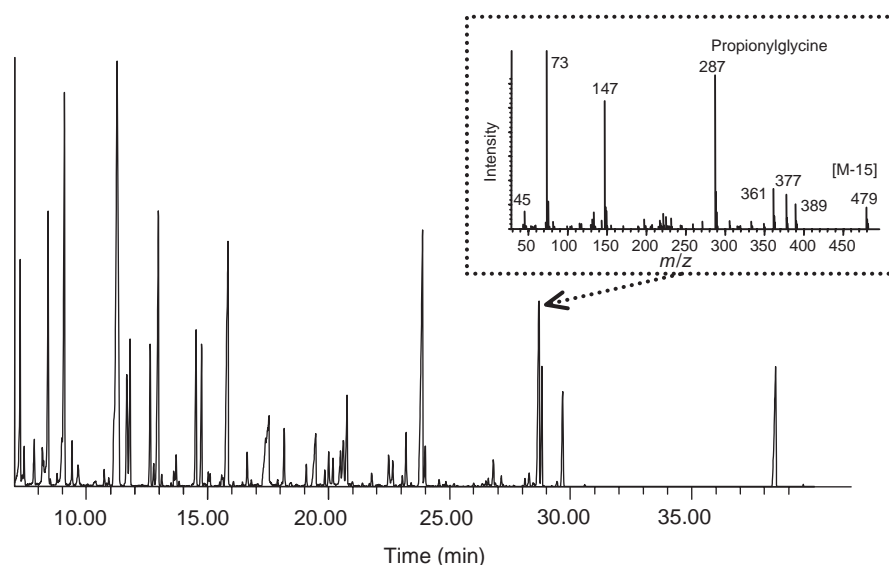
example of this scenario is medium chain acyl CoA dehydrogenase (MCAD) deficiency and their beta-oxidation disorders. The enzyme, MCAD, is specific for medium chain acyl CoAs. These include octanoyl CoA, hexanoyl CoA, and decanoyl CoA, among others. In MCAD deficiency, it would be expected that all three species would be elevated, and they are indeed.

These three schemes of metabolism offer a unique perspective for discussing the analysis of metabolites in clinical chemistry labs. Traditionally, many laboratories have utilized the one analyte, one disease approach to testing. Hence, metabolite A in each of the scenarios would be a single measurement made in a laboratory, often by a relatively nonspecific test such as an immunoassay or fluorometric analysis. There are often numerous circumstances that can produce an elevation of a particular metabolite in addition to an enzyme deficiency produced by a metabolic disease. The result is a false positive. Measurement of more than one metabolite that is affected by a particular enzyme, Phe and Tyr as in the case of phenylketonuria (PKU), reduces the likelihood of a false positive as these amino acids are 'linked' metabolically. The case is similar in scenarios 2 and 3 where the pattern of metabolites may indicate one or more disorders that share common metabolic pathways. One final point regards scenario 2. Measurements of metabolites 2 or more steps away from the primary metabolic block, theoretically and in practice, are somewhat less reliable indicators of disease than the primary substrate. Often, however, no alternative is presented due to the primary substrate being chemically unstable or difficult to measure analytically. Generally, these metabolites are easier to detect in older infants as they accumulate over time.

## The Metabolic Profile

Unlike many routine clinical chemistry tests, clinical analyses of IEM almost always involve a multiple metabolite analysis. The results form the basis of a metabolic profile in which both individual concentrations of metabolites and their relationship to each other can be viewed either in tabular form or in a graphical display. Perhaps the most comprehensive and historically significant test in IEM studies is gas chromatography/mass spectrometry (GC/MS) of a derivatized extract of urine. **Figure 2** is a chromatogram from an infant with propionic acidemia, an organic acid disorder of leucine metabolism. Hundreds of volatile compounds of carbohydrate, amino acid, fatty acid, and nucleic acid metabolism are separated in 40 min using capillary GC. Addition





**Figure 2** Capillary gas chromatogram of a derivatized urine extract from a patient with propionic acidemia. The mass spectrum of a peak at 28.7 min is shown in the inset and identified as propionylglycine, one of the metabolic indicators for propionic acidemia.

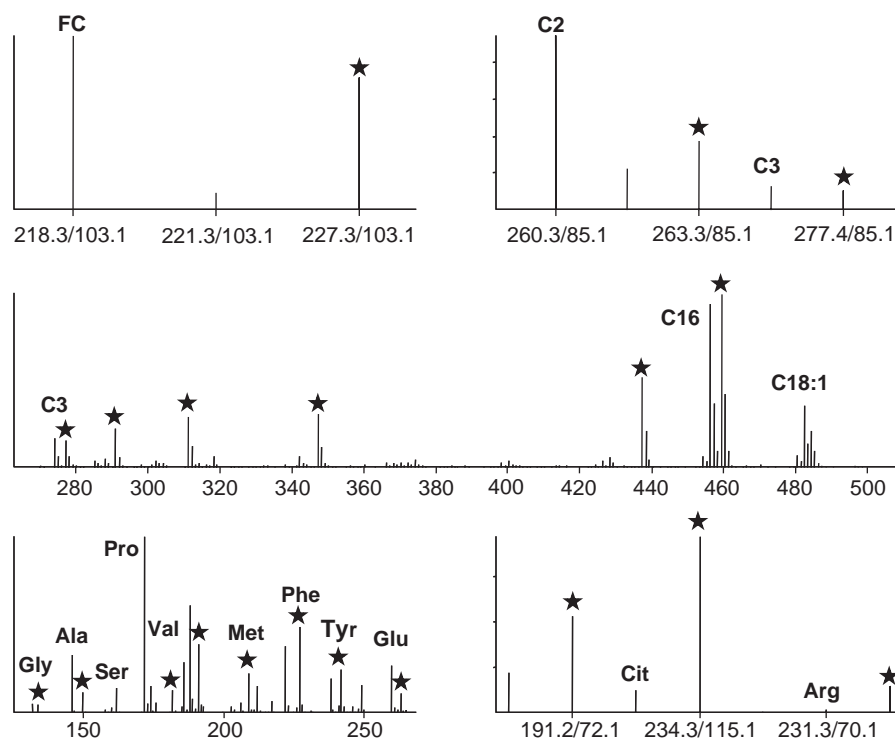
of a mass spectrometer as a detector improves identification of peaks at a particular retention time since separation of each component is often incomplete. An example of the mass spectrum of propionylglycine, a metabolite associated with propionic acidemia, is shown in the inset of **Figure 2**. Since other diseases can be detected in an organic acid analysis, it is perhaps still the most powerful tool in the diagnosis of metabolic disorders.

More recently, a complementary approach to clinical analysis of IEM is tandem mass spectrometry (MS/MS). This method is currently used to analyze numerous amino acids and acylcarnitines in a single analysis. One advantage of MS/MS (when compared to GC/MS) is that it does not require chromatography and hence lends itself to high volume clinical analysis of hundreds of samples per day. Another advantage of MS/MS is that it is a soft ionization, liquid chromatographic system that enables the analysis of ionic and polar compounds without extensive derivatization. An example of a typical MS/MS analysis of amino acids and acylcarnitines extracted from a dried newborn blood spot is shown in **Figure 3**. More than 65 separate compounds are detected in a single analysis. The concentrations of individual metabolites and, in some cases, their relationship to each other, are used to presumptively diagnose more than 35 disorders of amino acid, organic acid, and fatty acid metabolisms. Similar methods have been used to analyze blood, plasma, and urine from infants suspected of a metabolic disorder as well as to analyze postmortem specimens collected at autopsy from infants who have died of unknown cause.

Both GC/MS and MS/MS serve as confirmatory tests for each other. Modern clinical laboratories specializing in diagnosis of IEM utilize both techniques considering the quantity of information obtained using GC/MS and MS/MS. It is important to note that there are other methods in addition to GC/MS and MS/MS that are important in the diagnosis of disease and its confirmation. As the medical community places increased emphasis on reducing or controlling cost and clinical chemists are required to improve laboratory efficiency and become more regionalized and specialized, techniques such as GC/MS and MS/MS will remain the primary tools of choice for the foreseeable future.

## Specimen Characteristics

The type of biological specimen and form of collection is important in IEM. Although this statement appears obvious, there are a large number of considerations and implications of their use. Urine specimens are characterized primarily by hydrophilic metabolites that are often extensively modified from the original endogenous metabolite that accumulated in a metabolic disorder. As discussed previously, which compounds are present in the form of a metabolic profile is important in diagnosing a disorder but determination of concentration requires many assumptions. Because higher or lower urinary output affects the dilution of various metabolites, an indication is required to standardize urine concentration. Hence, creatinine is used as an indicator of urinary



**Figure 3** Tandem mass spectra from a typical analysis of the derivatized extract of a newborn blood spot. The five panels represent five types of analyses in a single 2 min run. These include (left to right, top to bottom), an analysis of free carnitine, short chain acylcarnitines, acylcarnitine profile, amino acid profile, basic amino acids. The stars represents internal standards for quantification.

output and most concentrations are reported as a concentration per milligram of creatinine. Few clinical labs analyze liquid blood directly for metabolites. Whole blood is either separated into plasma or serum and red cells. Traditionally, most clinical analysis is performed from plasma. Metabolites in plasma are often quite different from metabolites in urine, being of course less extensively modified. Plasma contains both hydrophilic and many lipophilic compounds with the latter requiring association with protein compounds for solubility. Blood can also be applied to an adsorbent filter paper specifically designed to have specific plasma volume equivalence per area of the sample at a specific hematocrit. This sample used in clinical labs, is the specimen of choice in newborn screening labs, and is increasingly used in metabolic specialty labs. Since blood can be applied to filter paper without precise application (a simple drop of blood from a heel stick), it is convenient from the screening perspective. However, it is quantitatively poorer since an individual's hematocrit and methods of application vary considerably. Nevertheless, its use is expanding with techniques such as MS/MS where emphasis is placed on a semiquantitative metabolic profile and relative ratios of metabolites to each other.

## Patient Characteristics

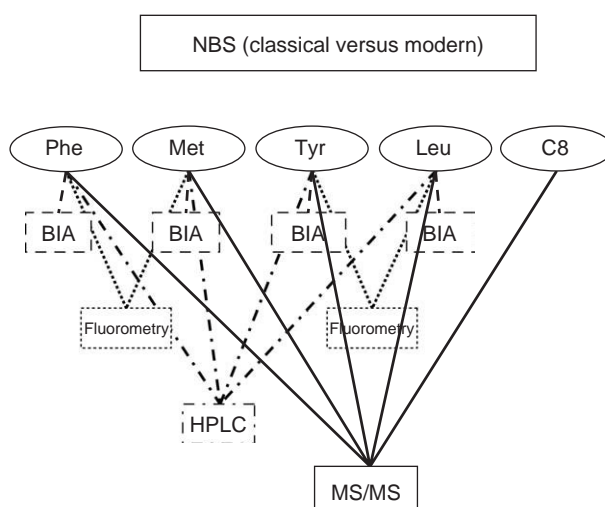
The age of the patient when a sample was collected is extremely important in a clinical analysis of IEM. Illustrations of this are plentiful in newborn screening where the first 7 days of life demonstrate many metabolic changes, the most important of which is at 'time zero' when the infant can no longer rely on maternal metabolism and nutrition. Generally, very early collection of blood, plasma, or urine specimens may be unreliable especially for diseases that are affected mostly by diet. The best example is phenylalanine and detection of PKU. The concentration of phenylalanine increases with time in affected individuals. This rate of accumulation depends upon many factors, but the most important being the degree of loss of enzymatic activity of phenylalanine hydroxylase and the influx of substrate, phenylalanine, from dietary protein sources. Historically, older, less accurate methods such as bacterial inhibition assays required a minimum of 48–72 h before an elevated Phe could be reliably used to detect PKU without substantial false negative and false positive rates. With newer technology as shown with MS/MS this can be reduced to 24 h without an increase in the false negative rate and with an important significant

decrease in the false positive rate. In fact, PKU and other amino acidopathies can be detected even earlier when necessary and as a safeguard to early discharge of patients. In other disorders such as  $\beta$ -oxidation defects that are analyzed via acylcarnitines, early detection (1–3 days) appears to be more reliable than an analysis of dried blood specimens more than 7 days of age due to substantially decreased concentration of carnitine and acylcarnitines. On the other hand, the reliability of organic acid analysis of this same disease improves with age. Other important characteristics are whether a patient is symptomatic, presymptomatic, deceased, prenatal, undergoing dietary treatment, supplementation, etc. Each has a different metabolic profile characteristic and is important in the analysis of a disease state and its detection.

## Analytical Characteristics

Choice of analytical methods in clinical chemistry is perhaps one of the most important decisions a clinical chemistry/laboratory director can make. Choosing a method requires the following considerations: types of specimens received, numbers of specimens per day, turnaround time, accuracy and specificity required, cost-effectiveness, cost-benefit ratio, expertise required, etc. Meaningfulness and interpretation of results are directly determined by which method is utilized. This is illustrated in two examples, methods used to diagnose PKU and MCAD deficiency. A flow chart (Figure 4) is utilized to illustrate the choices of methods for analyzing a modern panel of metabolites used in newborn screening showing older technology such as bacterial inhibition assay (BIA) and modern technology such as MS/MS.

In the earliest days of clinical chemistry, many methods relied on the biological responses of organisms to a particular stimulus. Robert Guthrie, in the early 1960s, developed a bacterial culture that would only grow in the presence of phenylalanine and would reverse the inhibition of a compound that inhibits its growth. This assay, known as BIA, was the stimulus for newborn screening of metabolic disorders. At the time, this method was inexpensive, high-volume capable, and easily implemented in public health laboratories. Further, samples could be collected on a piece of cotton-based filter paper from presymptomatic patients a few days after birth and mailed to a laboratory for analysis. The only other alternative at the time was a more expensive, high-performance liquid chromatography (HPLC) based analysis using liquid plasma from symptomatic patients. Problems with screening, however, became



**Figure 4** Flowchart showing classical and modern approaches to newborn screening. The top panel represents the desired analyte with bottom panels representing analytical approaches. The earliest technology, BIA, requires one assay per metabolite. A more recent classical approach is fluorometry, which may allow for two simultaneous analyses with one instrument. Still more recently is HPLC, which allows for an analysis of compounds in a panel with shared characteristics, e.g., amino acids. Finally, the most recent approach is MS/MS, which allows profile analysis of more than one compound type such as amino acids and acylcarnitines.

apparent by relatively high false results. Newer methods such as fluorometric assays later became integrated into laboratories. These methods improved reproducibility, no longer were biologically based assays dependent upon the response of an organism, and were still inexpensive. However, the false results were still high because like many immunoassays, the specificity of the fluorescent complexes with phenylalanine are fairly nonspecific for similar compounds. More recently, methods such as MS/MS overcame these limitations while presenting few shortcomings. First, MS/MS can be used in a highly selective analysis for Phe, it is versatile in that it can be used for both liquid and dried specimens, it is as accurate as either BIA or fluorometry (theoretically more accurate in liquid but not in filter paper specimens), it is inexpensive (for high-volume laboratories using MS/MS in conjunction with other assays), reliable, has a lower false positive rate that is at least 10–100 times better than other methods, and can improve diagnostic accuracy by measuring tyrosine in addition to phenylalanine (the ratio of Phe/Tyr is important in distinguishing a spurious elevation of Phe due to double spotting or hyperalimentation with amino acids). A limitation of the method is that it requires experience and expertise that is improving with progress in the ease of use of the

technology and increasing numbers of experts that are accessible. Further, due to its reliability and reduced false positive results, follow-up analysis and confirmation is more rapid, resulting in earlier treatment and medical care.

Most recently, a better understanding of the intermediary metabolism of fats and related disorders was realized primarily because analytical methodology improved to investigate the fundamental biochemistry. Interestingly, this same technology that led to improved understanding of  $\beta$ -oxidation resulted in an adaptation for the diagnosis of disorders of fat metabolism. Analysis of acylcarnitines was historically difficult using HPLC or GC due to the fact that these compounds are unusual in that they have both negative and positive side chains similar to amino acids in solution but with a fundamental difference in the esters they form with fatty acids on their aliphatic alcohol side chain. Depending upon the length of these side chains, compounds vary widely in their hydrophobicity. Hence, HPLC methods are excellent for measuring short chain acylcarnitines but poor for long chains. Radioimmunoassays and turbidimetric assays can only measure free carnitine. MS/MS, however, can measure both free and acylcarnitines providing both a measure of total carnitine and its fractions as well as individual metabolites. Measurement of individual metabolites in a profile is important in diagnosing numerous IEM. In fact, at this time, no other method can replace MS/MS as the primary analysis of numerous IEM in blood.

In both of these examples, PKU and fatty acid oxidation disorders, the methods utilized at this time are characterized by their comprehensiveness, cost-benefit ratios, and their ability to be used as part of a diagnosis of a particular disorder. In fact, these methods demonstrate that the future of clinical chemistry is multianalyte analysis that enables low cost even with relatively expensive instrumentation. Multianalyte analysis is not new to clinical chemists as the chemistry analyzers in hospitals and labs are used to analyze several dozens of compounds. A close look at these assays, however, will demonstrate that they are not true multiplexed analyses but rather a large robotic system performing 24 individual assays with 24 different chemistries and standards. MS/MS enables a single analysis, a single chemistry, and many results.

## Diagnostic Characteristics

It is important to emphasize that the purpose of a clinical analysis is to provide information regarding the metabolic status of a patient and whether a

particular analysis reflects a normal chemistry or that which deviates significantly from normal. To do this well, an analysis must fit the obvious criteria of high precision and accuracy resulting in few false results. There is, however, a challenge facing clinical chemists to provide more than a quantitative result and to provide information as to the probable diagnosis thus assisting a physician in making this determination. For example, consider analysis of a blood specimen for phenylalanine that reveals a concentration of  $240 \mu\text{mol l}^{-1}$  ( $4 \text{ mg dl}^{-1}$ ). From the physician's perspective, whether this patient has PKU or not is the question. A screening method that reports a concentration of 240 requires further testing and repeat analysis causing delay in treatment or unnecessary anxiety if subsequently normal. Consider a method that measures other amino acids, including tyrosine. Methods that provide both results enable a better estimation of whether or not this patient likely suffers from PKU. An elevated ratio of Phe/Tyr together with an elevated phenylalanine increases the probability of a truly positive case of PKU. Furthermore, a method that measures additional amino acids may suggest something different. For example, consider that leucine, methionine, and alanine were also elevated. Does this suggest a premature infant or perhaps an infant whose hyperalimentation has low concentrations of tyrosine (a possibility considering that some supplements have low concentrations of tyrosine due to its solubility characteristics)? Methods that provide more information improve the diagnostic characteristics of a test and help to determine whether repeat analysis is urgent, necessary, or should wait until medications are discontinued. As methods assist physicians in making a metabolic diagnosis, physicians can alter their response to analytical results.

It is also important to consider in any debate about whether a metabolic analysis is a diagnosis in and of itself: can a metabolic disease be diagnosed in an asymptomatic infant in the case of MCAD deficiency? The answer is no. If that is the case, is the analysis by MS the only method that can be used to diagnose a patient with MCAD deficiency in the newborn period? The answer is unclear in that no diagnosis would be made without an analysis but this analysis alone will not definitively diagnose MCAD deficiency without confirmatory testing such as DNA analysis, urine organic acids, or other studies. As is the case in forensic science, clinical science requires more evidence of a disease than a single test no matter how low the false positive or false negative rates are. Nevertheless, methods that improve the likelihood of an accurate diagnosis early enable a physician to spend more time treating a patient when the

maximum benefit is realized, before symptoms are present.

## Rare Disorders

Perhaps the single most important factor affecting clinical analysis in IEM is that these disorders are rare, relatively infrequent. This simple fact affects the who, what, why, and where for a clinical analysis of a metabolic disorder. Consider a disease that affects 1 in 15 000 individuals. The total number of patients expected to be affected per year in the US is ~250 based on the birth rate of 4 million. To what extent do we apply capital resources to developing methods to detect these disorders? What methods should be used? Is there benefit to early intervention and diagnosis? What is the cost of analysis? Are there metabolic experts knowledgeable in the diagnosis and treatment? Is the analysis accessible? Consider that a metabolic disease with a frequency of 1 in 15 000 is considered relatively frequent in the world of IEM as most diseases have frequency of 1 in 50 000 or less. IEM are therefore quite different than diseases affected by environmental influences or the aging process. A clinical analysis of inborn errors ideally would include methods that analyze more than one disease in a single test, perform many different analyses for multiple disorders in larger specialized labs rather than in individual hospitals, or biomedical labs that perform routine blood chemistries, and develop a network of metabolic experts to diagnose and treat these patients. Simply stated, consider a series of rare disorders screened with one technology as if they were one disease state, i.e., disorders of fat metabolism. This is best illustrated by MS/MS, which has changed the face of newborn screening, a change not realized since Dr. Guthrie first established the blood spot analysis of PKU. For example, PKU has a frequency of 1 in 12 000 whereas MCAD has a frequency of 1 in 14 000. Combining all the frequencies of disorders that should be screened regardless of individual frequencies without additional work produces a rate of more than 1 in 3000. Viewing metabolic diseases as a 'single group of disorders' provides the necessary financial and medical resources to support the clinical analysis of IEM.

See also: **Clinical Analysis:** Overview. **Gas Chromatography:** Mass Spectrometry; Forensic Applications. **Mass Spectrometry:** Overview; Principles; Electrospray;

Clinical Applications. **Quality Assurance:** Clinical Applications.

## Further Reading

- Bennett MJ, Rinaldo P, and Strauss AW (2000) Inborn errors of mitochondrial fatty acid oxidation. *Critical Reviews in Clinical Laboratory Sciences* 37: 1–44.
- Burtis CA, Ashwood ER, and Tietz NW (1999) *Tietz Textbook of Clinical Chemistry*. Philadelphia: W.B. Saunders.
- Chace DH (2001) Mass spectrometry in the clinical laboratory. *Chemical Reviews* 101: 445–477.
- Chace DH, DiPerna JC, Mitchell BL, *et al.* (2001) Electrospray tandem mass spectrometry for analysis of acylcarnitines in dried postmortem blood specimens collected at autopsy from infants with unexplained cause of death. *Clinical Chemistry* 47: 1166–1182.
- Chace DH, Hillman SL, Van Hove JL, and Naylor EW (1997) Rapid diagnosis of MCAD deficiency: quantitatively analysis of octanoylcarnitine and other acylcarnitines in newborn blood spots by tandem mass spectrometry. *Clinical Chemistry* 43: 2106–2113.
- Chace DH, Kalas TA, and Naylor EW (2002) The application of tandem mass spectrometry to neonatal screening for inherited disorders of intermediary metabolism. *Annual Review of Genomics and Human Genetics* 3: 17–45.
- Chace DH, Sherwin JE, and Hillman SL (1998) Use of phenylalanine-to-tyrosine ratio determined by tandem mass spectrometry to improve newborn screening for phenylketonuria of early discharge specimens collected in the first 24 hours. *Clinical Chemistry* 44: 2405–2409.
- Dougherty FE and Levy HL (1999) Present newborn screening for phenylketonuria. *Mental Retardation and Developmental Disabilities Research Reviews* 5: 144–149.
- Gerhards P (1999) *GC/MS in Clinical Chemistry*. Weinheim: Wiley-VCH.
- Hannon WH, Boyle J, Davin B, *et al.* (1997) Blood collection on filter paper for neonatal screening programs. NCCLS Document LA4-A3.
- Koch R, Williamson ML, and Donnell GN (1966) A cooperative study of two methods for phenylalanine determination: McCaman–Robins fluorimetric and microbiologic inhibition methods. *Journal of Pediatrics* 68: 905–911.
- Mei JV, Alexander JR, Adam BW, and Hannon WH (2001) Use of filter paper for the collection and analysis of human whole blood specimens. *Journal of Nutrition* 131: 1631S–1636S.
- Ozand PT and Gascon GG (1991) Organic acidurias: A review Part 1. *Journal of Child Neurology* 6: 196–219.
- Ozand PT and Gascon GG (1991) Organic acidurias: A review Part 2. *Journal of Child Neurology* 6: 288–303.
- Scriver CR (1995) *The Metabolic and Molecular Bases of Inherited Disease*. New York: McGraw-Hill.



# COAL AND COKE

**J A Pajares and M A Díez**, Instituto Nacional del Carbón (INCAR), Consejo Superior de Investigaciones Científicas, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Coal may be regarded as a combustible sedimentary rock. Whereas the majority of rocks are mostly inorganically derived, coal consists predominantly of organic matter, largely derived from a variety of plant remains (higher plants, ferns, fungi, algae) and different tissues (leaves, stalks, woody trunks, bark, pollen, spores, sclerotia, resins, etc.) with associated mineral constituents.

In the first stage of the evolution of coal from vegetable matter – diagenesis – changes in composition were due to biochemical processes, involving the degradation in a humid environment of the cellular structures and of the organic complex of the vegetable matter by activity of fungi and bacteria. Further alteration is a geochemical process – catagenesis – combination of the effects of pressure (due to overlying sediments) and mainly of heat (subsidence brings a rise of temperature) over a prolonged period. Evolution in chemical composition results in losses of water and gases ( $\text{CH}_4$ ,  $\text{CO}_2$ ) and, consequently, a reduction of hydrogen and oxygen contents, a decrease in volatile matter content, and a progressive enrichment in carbon content. Along these processes – coalification (Figure 1) – the original plant remains were progressively altered to

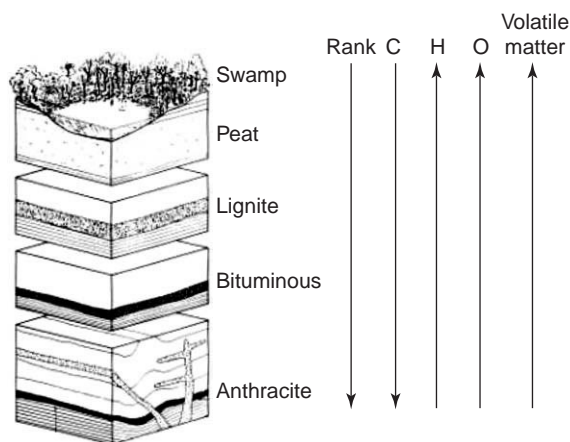
peat, lignite, subbituminous coal, bituminous coal, semianthracite, and anthracite.

The stage of coal alteration gives rise to the concept of rank – the greater the alteration, the higher the rank of the coal. Thus, lignite and subbituminous are described as low-rank coals, bituminous are termed medium-rank coals, whilst semianthracite and anthracite are termed high-rank coals. Coals of higher rank than lignites are often referred to as hard coals.

In addition to rank, two other parameters are important in coal classification, type and grade. Type of coal refers to the amount and sort of organic constituents called ‘macerals’ (see below), whereas grade refers to the relative purity or the amounts of ash-forming minerals found in coal.

Coal mines are spread over 100 countries. Mining may be surface or underground depending on whether the coal seam is exposed or at high depth. Known coal reserves are estimated over  $10^{13}$  tons and, at current production levels, workable seams are estimated to cover consumption beyond 225 years. About 90% of these reserves are in the Northern Hemisphere, with three countries – USA, China, and the former Soviet Union – accounting for about two-thirds of the total.

World hard coal production reached 3837 million tons in 2002, with a 50% rise in the last 25 years. Production of lignites was 897 Mt, Germany being the largest producer (~20% of the total). Data on the major producer countries are given in Table 1. Production of hard coal is increasing in Australia, South Africa, India, and in some new countries in the hard coal market, such as Indonesia, Colombia, and Venezuela. On the other hand, production is generally decreasing in those countries belonging to the European Union.



**Figure 1** Scheme of the formation of coal in terms of rank (coalification series).

**Table 1** Coal production. Major producers of hard coal (2002)

Country	Production (Mt)
China	1326
USA	917
India	334
Australia	276
South Africa	223
Russia	164
Poland	103
Indonesia	101
Ukraine	83
Kazakhstan	71

Data from World Coal Institute.



**Table 2** International hard coal trade. Major coal exporters (2002)

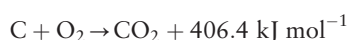
Country	Exportation (Mt)
Australia	198
China	86
Indonesia	73
South Africa	69
Russia	45
USA	36
Colombia	34
Canada	27
Poland	23

Data from World Coal Institute.

The recession in coal mining in Europe and the need of Southeast Asian countries for coal as energy source has determined a marked increase of the international coal trade from some producer countries to those two neuralgic points. Data for 2002 are given in **Table 2**. Total trade reached 625 Mt. USA's contribution is decreasing because coal companies are not interested due to weak coal market conditions. Australia is the first coal exporter, having actually a market share of 25% of the European coking coal market and practically 100% of the coking coal market in Japan.

Coal takes care of 25% of the world's primary energy, with forecasts for the next 15–20 years of 2% per year energy growth and a little higher growth, 2.2%, for coal production. Productivity depends on the grade of mechanization and other factors. Actually, the average mining costs in the USA are in the order of \$5–6 ton<sup>-1</sup> in surface and \$14–20 ton<sup>-1</sup> in underground mining. Productivity ranges from country to country from less than 1000 ton per person a year to 220 tons per person and for a shift of 8 h in the Powder River Basin surface mining. The price of a ton of coking coal (~\$50) is larger than the price of steam coal (\$40) in a global market of ~US\$125 billions.

Combustion, carbonization, gasification, and liquefaction are considered the four grand processes in the utilization of coal. In general terms, 92% of the coal production is used as fuel and ~8% is carbonized to produce metallurgical coke. Coal combustion is carried out in thermal utilities for electricity production, co-generation plants, and cement factories. Coal combines with oxygen from the air giving carbon dioxide and heat:



Conversion to electric energy reaches efficiencies from 15% to 20% for old installations to >45% for

new plants with pressurized fluid bed reactors or supercritical condition boilers. About 40% of the world's electricity production is based on coal; in countries such as China, Australia, or South Africa, more than 80% of electricity needs are met with using coal. Indeed, electricity from coal, with all restrictions derived from the Kyoto protocol, still meets 52% of electricity demands in countries such as Germany and USA.

The process of iron production in blast furnaces is the major application of metallurgical coke (about 90% of the coke produced worldwide). The whole coal consumption in the iron-making industry comes from two routes. On the one hand, coking coals are needed to produce metallurgical coke with strict quality specifications to feed the blast furnace. Currently, ~700–800 kg of coke are produced from 1 ton of coal, so to produce 340 Mt of coke for the blast furnace, ~450 Mt of coking coal needs to be carbonized in 250 coking plants around the world. Excellent coking coals are mined in Australia, Canada, China, USA, and Poland. On the other hand, coal is also consumed by the steel industry in blast furnaces operating with pulverized coal injection (PCI). As an example, blast furnaces with PCI require 700 kg of coal for each ton of hot metal produced, 525–600 kg of coking coals to make 350–400 kg of coke, and 100–200 kg of cheaper coal to be injected via the tuyeres.

Among the nonsteel blast furnace applications, foundry iron is the second most important application of coke produced from coking coals. An estimation of the global production is between 50 and 55 Mt per year. Coke for electrometallurgical reduction processes represents a minor consumption. In the nonsteel applications, coke is used as a reducing agent or as an energy source in industrial applications such as ferroalloys, lead–zinc smelting, silicon carbide production, and stone wool.

## Preparation

In most coal-producing and consuming countries, national standard methods are available for preparation, sampling, and analysis of coal. Through the International Organization for Standardization (ISO) considerable progress has been made toward the development of internationally acceptable procedures. Most of the ISO procedures are based on standard methods formulated by National Organizations such as the American Society for Testing Materials (ASTM), the British Standards Institution (BSI), the German Normenausschuss (DIN), and Poland's Standards Committee (PN). In this article,

specific ISO and ASTM standard procedures are referred to.

In surface mining, the overlying rock (overburden) is removed and the coal extracted. According to geological and topographical conditions, different methods (opencast, open pitch, etc.) and equipment (shovels, draglines) are used. Conventional room and pillar mining has been largely replaced by continuous methods – continuous miner and long-wall systems – in underground extraction. But, in all cases, the coal extracted by any of these techniques will have suffered contamination with incombustible rock and water.

Mineral matter is incorporated to the coal seam from different origins along the coalification period

1. as trace elements from decomposition of vegetable matter;
2. as detrital minerals from fluvial contribution during the diagenesis evolution;
3. as mineral inclusions (calcite, pyrite, kaolinite, etc.) from the neighboring rocks during tectonic folding in the subsidence period; and
4. through the mechanical contamination cited above.

All the contributions determine percentages in the inorganic matter of coal that range from less than 2 mass% (some Indonesian ultraclean bituminous coals) to more than 50 mass% in some run-of-mine material.

Depending on the intended end-use of coal, removal of this extraneous inorganic matter, to a greater or lesser extent, is necessary. Diminishing the ash and sulfur content is important for a better handling of boilers and to fight air pollution. Limits in emissions to concentrations of nitrogen and some minor elements, such as mercury, are already established or near to be established. Coking coal is required with 10 mass% ash and 1 mass% sulfur contents as upper limits. To achieve the derived end-products the run-of-mine mineral is treated in coal preparation plants.

In a preliminary treatment, the run-of-mine material is crushed to a size of <150 mm. This raw material, between 2 and 150 mm (named coarse), is treated through cleaning devices that use the difference in solid density between coal and mineral matter. Some heavy media – organic liquids, inorganic salts, and suspended solid particles of magnetite – are used to produce different specific gravities. Smaller mineral fractions (2–100 mm, small, and 0.5–1 mm, fines) are separated through hydraulic separation machines (jigs) based on the rate of sedimentation. Finally, solids <0.5 mm (ultrafine) are treated in

flotation cells. The cost is correspondingly higher for a higher proportion of fines. Dewatering is done by means of centrifuges and the different products from the preparation plant are stored separately in silos. By dosing and blending, ash and moisture contents can be adjusted automatically according to the market.

Although cleaning was driven by quality requirements of metallurgical coal, pollution control has increased the need to improve the quality of steam coal. Actually, the number of coal preparation plants is ~2200 and they treat ~1800 Mt of run-of-mine material with a 73% recovery rate. Although the average plant capacity is 190 tons h<sup>-1</sup> biased by the small capacity Chinese units (>1500 tons), an average of 1000 tons h<sup>-1</sup>, are more in line with those established in Western countries. In the USA, there are ~100 plants with a capacity of more than 900 tons h<sup>-1</sup>, the largest in the world being Grooteegelunk, Republic of South Africa (8200 tons h<sup>-1</sup>).

## Sampling

Coal sampling may be defined as the extraction of a small amount of material from a larger bulk of coal such that the sample extracted is representative, as far as possible, for all analytes of interest. Sample preparation, which may involve drying, crushing, and subsampling, is an integral part of the sampling process.

The compositional variability of the material is such that coal sampling becomes a complex and difficult operation. The various standardization organizations around the world have all issued detailed documents that specify the conditions and methods necessary to obtain representative coal samples for analysis.

The process of sampling, which may vary from a 2 ton domestic consignment to a 165 000 ton export shipment, involves obtaining a number of increments (spot samples, the product of a single action of the sampling device) that are combined to form one or several gross samples. Gross samples may then be crushed and subsampled (subdivided) to provide the analytical sample.

Coal samples may be obtained by either manual or mechanical sampling devices, the latter being more appropriate where large tonnages are experienced (e.g., at coal shipment terminals with throughputs up to 10 000 tons h<sup>-1</sup>) and where continuous operations or operator safety dictate their use.

Standard methods for obtaining coal samples specify minimum numbers of increments required to form a gross sample, based on a consignment

(unit) mass of 1000 tons. For consignments of greater mass, the number of increments may be increased to provide a larger gross sample, or the consignment may be considered as several subconsignments, each of 1000 tons. For each of them, a separate gross sample is collected, to be individually processed and analyzed.

As well as the mass of the consignment, the required minimum number of increments is influenced by

- the quality of coal – washed or cleaned coals are more homogeneous and therefore require fewer increments than blended or untreated coals and
- the sampling process to be used – stopped-belt or falling-stream sampling requires fewer increments to attain the required precision than sampling from barges or stockpiles.

Sample preparation involves a series of operations such as reduction of size, homogenization, and reduction of the mass of the gross sample to that suitable for analysis. For the determination of size distribution, no sample preparation other than air drying and reducing the mass of the gross sample is undertaken. For other analysis, the extent of sample preparation is dictated by the intended analysis. For example, when testing for Hardgrove Grindability Index (ISO 5074, ASTM D409 or equivalent), a subsample of ~1 kg coal prepared to a top-size of 4.75 mm is required, whereas for general analysis a final sample of 50–100 g crushed to less than 0.2 mm is sufficient.

Sample preparation procedures typically involve a fixed sequence of operations:

1. Extract a subsample for total moisture determination.
2. Air dry the gross sample (to ensure that the coal will flow smoothly through subsequent equipment). It should be noted that forced drying can have adverse effects, especially on coking properties; thus heating samples to more than 15°C above ambient temperatures should be avoided.
3. Reduce the particle size of the sample. The whole gross sample is crushed to some intermediate size (10, 5, and 3 mm) using a mechanical mill or crusher.
4. Mix the whole sample thoroughly to ensure homogeneity.
5. Reduce the mass of the gross sample (sample division) to a mass consistent with the present size of the coal, by using a mechanical sample divider.
6. Further crush the sample using a high-speed impact mill to attain the required particle size.

7. Finally, mix and divide to the mass required for the laboratory sample.

It should be noted that the coking properties of crushed coal samples deteriorate rapidly with time. Therefore, samples intended for such testing should be prepared immediately before analysis.

Coke sampling is marginally less problematic because the product from a single source derives from coal or blend of coals that have been prepared to a specification for ash, moisture, particle size distribution, etc. The final coke produced will be relatively homogeneous in all properties, with the exception of size distribution. Standard methods are available for coke sampling that reflects the somewhat less rigorous requirements for this material.

## Petrography and Chemical Composition

Coal is not a homogeneous rock. In coal seams, petrographers have distinguished with the naked eye more or less lustrous thin beds and regular accumulations that can be divided into four fairly well-defined classes called lithotypes – vitrain (brilliant layer), clarain (semibrilliant layer), durain (mat layer), and fusain (fibrous layer). When coal is examined through an optical microscope with reflected light under oil immersion, it is shown that coal is composed of discrete entities, called macerals, which result from the transformations of the original vegetal debris. According to their optical characteristics macerals can be classified into three groups:

- The vitrinite group originates from lignocellulosic tissues that are gelified by bacterial action. It is mainly composed of aromatic and hydroaromatic structures and is usually the most abundant maceral group. It often occurs as a matrix surrounding the other macerals and mineral constituents and has the property of swelling and agglomerating during the carbonization of medium-rank coals. Its density increases with rank from 1.2 to 1.7 g cm<sup>-3</sup>.
- The liptinite group is derived from organisms and organs that are relatively poor in oxygen such as algae, spores, pollens, cuticles, secretions. Chemically, it is characterized by a higher content in hydrogen and aliphatic structures. It is the lightest maceral with a density between 1.1 and 1.25 g cm<sup>-3</sup>. By thermal heating in an inert atmosphere, it is converted to volatile products, leaving little solid residue.
- The inertinite group often originates from vegetable matter that is partly burnt or has undergone lengthy aerobic oxidation before burial. Chemically,

it is mostly aromatic with a comparatively higher amount of carbon and lower amounts of hydrogen and volatile matter. It is the densest maceral (1.4–2.0 g cm<sup>-3</sup>). Most inertinite remains inert during the carbonization process, although the least reflective ones still retain plastic properties.

The maceral content defines the coal type: sapropelic, with >50% liptinite, or humic, more abundant, usually presenting a banded structure. On the other hand, based on the different optical properties of macerals, the reflectance of vitrinite is an essential characteristic used in coal identification and related to rank. A good analysis of maceral content provides knowledge about the chemical composition of a coal, their behavior in different conversion processes, and can also be used as a parameter of coal rank (see section on Petrographic analysis).

Thus, the chemical composition of the dominant organic part of a coal, mainly the amounts of carbon, hydrogen, and oxygen, which together comprise between 97% and 99% of the mass of pure coal, depends on the coal type, and has a clear evolution with coal rank, the carbon content increasing progressively from <65% for lignite to ~95% for anthracite. Hydrogen is a minor element, following an opposite trend. Oxygen decreases correspondingly from above 30% in lignites to <1% in the highest rank coals. No distinct trends are apparent for nitrogen (usually between 0.3 and 1.5%) and organic sulfur (0.3–2%). A progressive decrease in water and volatile matter contents (from >42 to <4%) goes parallel to an increasing rank. **Table 3** presents typical values for elemental analysis of coal precursors and coals of increasing rank, together with proximate analysis and calorific values.

The inorganic constituents of coal are the residues of plant components, minerals incorporated in coal seams, and rocks co-deposited or mixed with the coal during mining. As such, their chemical composition is enormously varied. The principal components found are quartz and minerals of clay, mica, and chlorite groups. Sulfide, carbonate, and chloride minerals, together with feldspars, barites, and titanium minerals, make up the remainder of the suite, along with lesser concentrations of minerals formed by weathering, such as limonite and sulfates of iron and calcium. Of particular interest is the content of sulfur minerals. Sulfur is an undesirable constituent of coal, and as such will attract a price penalty if present at greater than specified levels.

Virtually all other elements occur sporadically in coal. These 'trace' or 'minor' elements rarely constitute more than 1% of the mineral matter and frequently the proportion is much lower.

**Table 3** Composition of different rank coals

Coal	C (mass% dmmf)	H (mass% dmmf)	O (mass% dmmf)	N (mass% dmmf)	S (mass% dmmf)	Moisture (mass%)	Ash (mass% db)	Volatile matter (mass% db)	Calorific value (MJ kg <sup>-1</sup> )
Wood	50.0	6.0	43.0	0.5	0.5	15.0	1.0	70.0	16.61
Peat	57.0	6.0	34.5	1.5	1.0	13.0	3.0	59.0	18.56
Lignite	66.0	5.5	25.2	1.3	2.0	16.0	6.0	48.0	19.05
Bituminous									
hvb <sup>a</sup>	78.8	5.5	13.0	1.3	1.4	10.7	3.1	32.0	28.47
hvb <sup>a</sup>	84.7	5.4	7.3	1.8	0.8	4.7	5.0	35.5	31.40
hvb <sup>a</sup>	87.5	5.5	4.8	1.2	1.0	1.5	5.9	31.6	33.00
mvb <sup>a</sup>	89.7	4.9	3.3	1.4	0.7	0.8	4.8	24.5	34.42
lvb <sup>a</sup>	91.0	4.4	2.5	1.4	0.7	0.6	2.9	16.1	35.24
Semianthracite	92.4	4.0	1.3	1.5	0.8	1.3	3.0	10.9	34.84
Anthracite	94.0	2.9	0.8	1.5	0.8	2.9	2.3	5.1	33.96

<sup>a</sup>hvb, mvb, and lvb refer to high-, medium-, and low-volatile bituminous coals.

Data from *Encyclopedia of Analytical Science*, 1st edn, Academic Press.

## Analysis and Tests

The suite of analyses required for a particular coal mainly depends on its intended end-use. Steam coals are defined through parameters related to their calorific value and impurities content having a strong impact on environmental pollution; metallurgical, or coking coals are better classified through specific tests related to their properties to transform to a good-quality coke for a particular application. Testing and assessing of a coal starts with a consideration of its rank, washability, hardness, and its inherent moisture content and is followed by analyses of its chemical, physical, rheological, and microscopic properties, and, finally, pilot tests. As with coal sampling, national and international standard methods for the analysis of coal and coke have been developed in order to define the quality of the material through chemical, petrographic, and empirical tests. These enable the producer to monitor variations in the quality of the product and the purchaser to assess the suitability of marketed coal for a process.

Instrumental methods are now becoming accepted as alternative techniques, provided they are shown to give equivalent results to the conventional test. These methods can allow a greater throughput of samples, with less operator dependence than conventional techniques, and often generate useful data that were not previously available. In fact, there are some instrumental companies, i.e., LECO Corporation, devoted almost exclusively to the development and sale of apparatus for chemical analysis – immediate analysis, elemental analysis, sulfur determination – calorific value, ash fusibility, and other technical tests of coals.

Knowledge of the total moisture of a coal is essential in commercial activities. Proximate analysis – determination of residual moisture, ash, and volatile matter in a dried sample of coal – is made for all samples received for coal utilization. Elemental analysis (C, H, N, O) and analysis of sulfur are necessary in order to have a better knowledge of the quality of a coal. Petrographic analysis provides information on the rank, quality, and blending of coking coals. Friability tests are necessary for the behavior of steam coals, and plasticity and swelling tests for the behavior of coking samples. The analysis of mineral matter gives information on the possibility of pollution problems and tests of fusibility on the ash behavior of combustion coals.

Most analyses of coal are carried out on air-dried samples and results are normally reported on the dry basis (db), as percentage of the coal after the moisture has been removed. When samples are compared on the basis of certain properties of the pure coal,

results are given in a dry, ash-free basis (daf). If more complete analytical data are available they can be given in a dry, mineral matter-free basis (dmmf), results being a measure of only the organic component of coal.

Other bases may be required for the expression of analytical data. Moist, ash-free basis (maf) assumes that the sample is free of ash, but with moisture (ASTM D388, Standard Classification of Coal by Rank, requires calorific value to be expressed on a moist, mineral matter-free basis (mmmfm)). The various national standards organizations present different formulae for the calculation of mineral matter (which is not a generally determined value); reference to their publications is necessary to determine which calculation is appropriate in given circumstances.

An exhaustive discussion of individual methods is outside the scope of this article. Thus, the more commonly reported tests are commented upon below.

## Moisture

The water in coal is bound in different forms to its constituents. It can be divided into three types: (1) Free moisture, also referred to as external moisture, superficial moisture, or the primary moisture fraction, which is present in large cracks and capillaries. Water bound in this way retains its normal physical properties. (2) Inherent moisture, also referred to as internal moisture or the secondary moisture fraction, whose vapor pressure is lower, since it is absorbed within the pore structure of the coal. (3) Water of constitution, which is mainly combined with mineral matter normally present in coal. This water is generally driven off only at temperatures higher than those normally used for the determination of moisture content. Standard methods do not make use of these terms and define: (1) the total moisture content of a coal; and (2) the moisture content of the coal analysis sample. Total moisture determination must be made over the sample as received in the laboratory, in an air-proof recipient. The determination consists in drying in an oven at 105°C till constant weight. Its value is of huge interest both in international and domestic coal trade (ISO 589, ASTM D3173).

This determination must be differentiated from that of the moisture of the sample for analysis that corresponds to a sample of coal equilibrated at the laboratory conditions (usually, 24 h). The methodology is the same and the value is that to be used in the calculations at the different bases.

The equilibration of a coal sample in laboratory conditions is absolutely necessary in order to facilitate handling of the sample in the following



operations, such as sizing – humid coal samples do not answer to riffing – or the Hardgrove grindability test (see below).

### Sizing

Size analysis is useful in assessing problems in coal transportation and handling. For larger sizes coals, hand or mechanical sieving or screening is undertaken; fines (<1 mm) are better screened in humid medium; while for ultrafines (<0.06 mm) optical or electrical field effect techniques are more appropriate.

Size reduction is carried out on the samples for analysis, depending on the specific analysis or test (i.e., <0.212 mm for chemical analysis, <4.75 mm for Hardgrove; <1 mm for petrographic analysis).

### Hardgrove Grindability Index

This index gives a value for the grindability of coal and coke. The numerical value is related to the number of revolutions needed to reduce a given amount to a determined particle size, relative to a standard (see ISO 5074 and ASTM D409). Grindability is related with rank, Hardgrove index reaching a maximum for coals with 80–90% of carbon. High values of Hardgrove index (50–80) are indicative of easier to grind materials while values in the order of 20–30 indicate a higher hardness of a coal.

## Chemical Analyses

### Proximate Analysis

The analyses of the air-dried equilibrated sample for moisture, ash, and volatile matter are collectively termed the proximate analysis. Fixed carbon is, by definition, the difference between 100 and the sum of the analytes (moisture, ash, volatile matter). The proximate analysis gives information on the classification of coal by measuring the relative percentage of volatile and nonvolatile organic matter as those of moisture and noncombustible mineral matter.

Determination of moisture was discussed above (ISO 589, ASTM D3173). Volatile matter – parameter used in coal classification – can be defined as the percentage of gaseous components of the coal, except moisture, loss at a high temperature in an inert atmosphere. Its determination is carried out under prescribed conditions (i.e., 7 min at 900°C) starting from an amount of normalized material around a mass of 1 g (ISO 562, ASTM D3175).

Ash is the solid residue from inorganic material after the complete combustion of coal. Its composition and amount depends on the analytical conditions (ISO 1171, ASTM D3174) that must be

strictly maintained in order to obtain comparable results. The ash content differs from the mineral matter content, both in their composition and in their mass (always lower). The mineral matter content is often calculated following the empirical formula of Parr:

$$\text{Mineral matter (\%)} = 1.08\text{Ash} + 0.55S_{\text{pyritic}}$$

The difference between both contents – ash and mineral matter – can lead to considerable errors, larger when the contents are larger. For less than 30 mass%, the error is ~1.3%, which rises to 4% when the ash content is around 25% and to 28% for inorganic contents of 70%.

The whole analysis can be done in an automatic apparatus designed to reproduce the analytical conditions of the established procedures. The apparatus follows a program and can work simultaneously with 20 samples. The conditions for determination of moisture are well accomplished, but some details on the determination of volatile matter (time) and ashes (temperature, 815°C) are not exactly the same and the tolerance (2%) is higher than that of the standard method (1%). Anyhow they are largely used in electrical utilities that need to make hundreds of analyses from different suppliers each day.

### Calorific Value

The gross calorific value is the amount of heat released by burning with oxygen a coal sample in a calorimeter under controlled conditions. A correction for the heat absorbed by the remaining ashes must be calculated. If the correction includes the latent heat of vaporization, the net calorific value, important in the coal market, is determined.

The measurement can be done both in isothermal or adiabatic calorimeters, the latter being preferred. For isothermal measurement (see ASTM D3286), the temperature of the calorimeter jacket is held constant and a correction for heat transfer from the calorimeter is applied, while in the adiabatic measurement (see ISO 1928 and ASTM D2015), the temperature of the calorimeter jacket is continuously adjusted to approximate that of the calorimeter itself.

The calorific value can be correlated with the fixed carbon content of the coal. It is a good parameter for coal classification of specific coal types and as a price determination index (\$ per MBtu) for commercial steam coals.

### Ultimate or Elemental Analysis

Analysis for the elementary constituents of coal follows techniques similar to those employed in organic



chemistry. It comprises analyses for carbon, hydrogen, nitrogen, sulfur, and oxygen, and altogether give the composition of the organic matter of a coal. Corrections must be made because, apart from nitrogen, carbon, sulfur, and oxygen can be found in impurities (as carbonates, sulfides, sulfates, clays), and hydrogen and oxygen are present in the sample moisture.

Carbon and hydrogen are determined through complete combustion at 1200°C. The gases (CO<sub>2</sub> and H<sub>2</sub>O) are absorbed, respectively, on anhydrous and NaOH and measured gravimetrically (see ISO 609, ASTM D3178).

Nitrogen is determined by the Kjeldahl method. Nitrogen is transformed in ammonium sulfate by treatment with concentrated sulfuric acid at 900°C. Afterwards it is liberated as ammonia in a steam stream that passes through a diluted acid solution and a titration with alkali (see ISO 332, ASTM D3179).

Sulfur occurs in both organic and inorganic combinations. For certain purposes a knowledge of total sulfur content is adequate; however, for coal preparation and conversion processes, determination of the forms of sulfur (organic, pyritic, and sulfate) is valuable.

The total sulfur content may be determined by one of several methods that convert it to sulfate by wet chemical analysis. One of these, the Eschka method, involves combustion of coal at 800°C in the presence of alkaline/oxidant medium (e.g., two parts of calcined MgO and one part anhydrous sodium carbonate); all sulfur is converted to sulfate that by the addition of barium chloride precipitates as barium sulfate, which is calcined to BaO and measured gravimetrically (see ASTM D3177). This is a standard method in many countries. Another is the 'high-temperature method' where the coal is burned in oxygen at 1350°C, converting all sulfur present into SO<sub>2</sub>. The SO<sub>2</sub> is then converted to sulfuric acid for titrimetric determination.

Pyritic sulfur may be determined by the estimation of pyritic iron, which involves a pretreatment with hydrochloric acid to eliminate the nonpyritic iron and then dissolving the pyritic iron in nitric acid.

Sulfate sulfur is estimated by solution in hydrochloric acid followed by gravimetric estimation of the dissolved sulfates.

Procedures to determine the sulfur forms – sulfates, pyritic, and organic sulfur – are well described in ISO 157 and ASTM D2492.

Automatic systems have been developed and are largely used in the simultaneous analysis of carbon, hydrogen, and nitrogen. Samples (up to 20) of 50–100 mg are burned and the evolved gases, CO<sub>2</sub>,

H<sub>2</sub>O, and N<sub>2</sub>, monitored by a thermal conductivity detector.

The oxygen content of a coal is an important rank parameter, the younger coals being richer in this element than the more mature coals. Oxygen is traditionally calculated by difference mainly for high-rank coals from a knowledge of the amount of other chemical components (C, H, N, and S), moisture, and ash contents. Techniques involving direct determination of oxygen avoid the effect of cumulative error in the analysis by difference.

The direct determination of oxygen has been and continues to be a problem as it does not have an easy solution. Laboratory microanalyzers for C, H, N, S, and O have been developed. They work over smaller coal samples (~1 mg) and the oxygen determination, based on chemical methods available, is carried out in a supplementary device, where the sample is pyrolyzed at 1350°C in a helium stream and reducing atmosphere. The resulting gases are passed over activated carbon, which converts oxygenated products into carbon monoxide. The CO is converted catalytically to CO<sub>2</sub> and then quantified by means of an infrared detector. The method is increasingly used mainly in analysis of low-rank coals, in coal weathering research, and analysis of carbon materials.

### Other Chemical Analyses

The chemical analyses of ash for major elements and a test in ash fusibility temperature give guidelines on coal utilization, defining the suitability of coal for different combustion or conversion systems. Analysis for minor elements is gaining impetus as greater interest is taken in the environmental consequences of coal uses. Typical ranges of concentration for major, minor, and trace elements are summarized in **Table 4**.

Multielement analytical techniques – atomic absorption spectrometry, inductively coupled plasma mass spectrometry, X-ray fluorescence, neutron activation analysis, etc. – are used. The experimentation can be done directly on the mineral matter of the coal sample after the removal of the organic matter by a prolonged treatment of activation with oxygen plasma (low-temperature ashing). Neutron activation is also applied to online analyses of coal and fly-ashes on feeding-belts in order to provide information on a continuous basis.

Chlorine can also occur in coal as organic and inorganic compounds, but only the total chlorine is normally determined. Formulae for calculating mineral matter contents arbitrarily assume 50% of the chlorine to be inorganic.

**Table 4** Concentration ranges of major, minor, and trace elements in coal

<i>Major and minor constituents</i>		<i>Trace elements</i>	
<i>Ash analysis</i>	<i>Ash (mass%)</i>	<i>Element</i>	<i>Concentration (<math>\mu\text{g per g coal}</math>)</i>
SiO <sub>2</sub>	40–90	Be	0.1–15
Al <sub>2</sub> O <sub>3</sub>	20–60	Cr	0.5–60
Fe <sub>2</sub> O <sub>3</sub>	5–25	Mn	5–300
CaO	1–15	Co	0.5–30
MgO	0.5–4	Ni	0.5–50
Na <sub>2</sub> O	0.5–3	As	0.5–80
K <sub>2</sub> O	0.5–10	Se	0.2–10
P <sub>2</sub> O <sub>5</sub>	<1	Cd	0.1–3
TiO <sub>2</sub>	<2	Sb	0.05–10
		Hg	0.02–1
		Pb	2–80

## Petrographic Analysis

The most widely used petrographic analyses of coal are maceral analysis and vitrinite reflectance analysis. Both are performed on representative samples ground to <1 mm in size and embedded in resin. The polished surfaces are then examined under a white reflected light microscope.

### Maceral Analysis

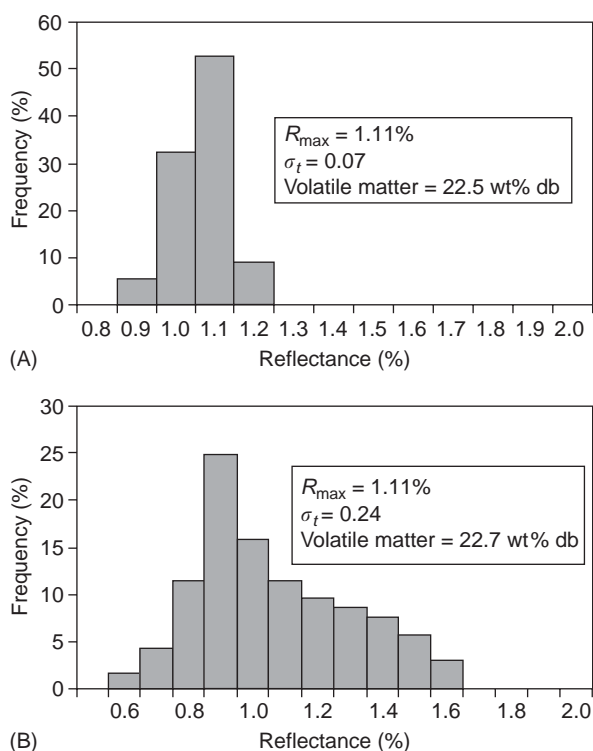
The relative proportion of petrographic components in a coal can be performed at a maceral level or maceral group level depending on the degree of detail desired. For most industrial applications the maceral group analysis is usually enough. For coals to be used in coking processes it might be desirable distinguishing between low- and high-reflecting inertinites since the former still retains plastic properties (see below). For a better discrimination of liptinite group macerals in low-rank coal the use of fluorescence light might be convenient. As with other coal analyses, several international and national standard procedures are available with minor differences among them and they describe with some degree of detail the procedure of pellet preparation, the total number of points counted, the magnification employed, the repeatability and reproducibility values, etc. (i.e., ISO 7404-3 and ASTM D2797). Oil immersion objectives are normally used to provide best identification conditions. For standard maceral analysis typically 500 points are recorded by point counting displacing the sample with a mechanical stage at regular intervals (0.5 mm). The results are expressed in volume percent based on the principle that the surface occupied by a given component in a randomly selected section is proportional to the volume of that component in the sample.

### Vitrinite Reflectance Analysis

The reflectance of vitrinite is a rank parameter that has well-established relationships with other chemical rank parameters such as volatile matter, C, and H contents. Among the advantages of the use of vitrinite reflectance as a rank parameter are that it follows a regular increase over the whole coalification scale and is not affected by maceral composition since it is recorded on individual particles. Reflectance is defined as the proportion of perpendicularly incident light reflected from a component compared to that reflected from a standard of known reflectance. Readings are taken with monochromatic light (546 nm) and oil immersion objectives (1.518 refraction index at 23°C) following the standard ISO 7404-5 procedure (see also ASTM D2798). Strict rank determination requires recording the readings on a single maceral from the vitrinite group – collotelinite – although for most industrial applications readings recorded on any maceral of vitrinite group might be sufficient.

Two types of readings can be taken on vitrinite particles: (1) random reflectance ( $R_r$ ) using nonpolarized light; and (2) maximum reflectance ( $R_{\max}$ ) using polarized light and turning the stage till the maximum reflectance is achieved. For low-rank coals,  $R_r$  and  $R_{\max}$  are equivalent. As coal rank increases vitrinite develops anisotropy yielding higher  $R_{\max}$  values than  $R_r$ . In both cases 100 readings selected using a regular frame are averaged to calculate mean reflectance values and to construct the reflectogram with distribution of vitrinite classes. Standard deviation must also be provided. Reflectance values typically approach to a Gaussian distribution for low-rank coals and standard deviation increases with the rank of the coal.  $R_{\max}$  is recommended for detailed studies on burial history of coal basins, whereas for most industrial applications random reflectance is preferred since its analysis is less time consuming.

Petrographic analysis is the only procedure to determine the various components of a coal blend, which is of foremost importance in coal utilization. In this case reflectance readings are taken on any vitrinite within the particle regardless the maceral on which the crosswire is landed. With this procedure the histogram plot reflects directly the proportion of the coals in the blend. The analysis yields the rank of the single-component coals and their amount. Coal discrimination is easy for blends of coals of rather different rank. Calculations may require histogram deconvolution for coals close in rank. A higher amount of readings is recommended for coal blend analysis.



**Figure 2** Histograms of the distribution of vitrinite in reflectance classes for a bituminous coal (A) and a complex coal blend (B).

As an example, **Figure 2** displays the histograms of the distribution of vitrinite in different reflectance classes for a bituminous coal and a complex coal blend with the same volatile matter content and mean vitrinite reflectance.

Both maceral and reflectance analysis may be performed by automated image analysis systems that process images acquired with a camera. They usually yield an average reflectogram of all components in the sample where thresholds must be established in a further step either manually or using more or less complex computer routines. The maceral group percentages are calculated as the area under the curve within two thresholds and the vitrinite reflectance value is based on the position of the peak assigned to vitrinite. Automated systems yield results similar to manual analysis for medium-rank vitrinite-rich coals, but problems remain with resin–mineral matter–lipinite thresholding for low-rank coals and inertinite–vitrinite thresholding for high-rank coals.

## Thermal and Rheological Properties

It is well established that only coals within a specific range of rank and type are suitable for coke production. Chemical analyses of the single coals, including ash and sulfur, are important parameters in

cokemaking, but criteria for selecting coals in coke manufacture are based on certain properties of coals that, when heated in an inert atmosphere, cause them to soften, become plastic, and coalesce into a coherent mass that swells and resolidifies to form a solid porous carbon material (coke). These particular types of coal are generally known as ‘caking coals’. The physical changes occurring in coal when heated in the temperature range of 350–550°C in the absence of oxygen are known as ‘plastic or thermoplastic properties’. The importance of such properties is related to the capacity of agglomeration between the reactive maceral constituents of the coal and the inert particles (some part of the inertinite maceral group and mineral constituents). On the other hand, the remaining coals are referred to as ‘noncaking’ coals and they produce a weakly coherent or noncoherent char. It is important to point out the difference between ‘caking coal’ and ‘coking coal’. The term ‘caking’ is reserved for coals that possess plastic (or agglutinating) properties in laboratory tests, while the use of ‘coking’ is related to caking coals in which plastic properties are so strongly expressed as to make them suitable for conversion in metallurgical and other industrial cokes. Caking coals are therefore strongly caking coals with a volatile matter content ranging between 19 and 32 wt%. Caking behavior can be assessed during rapid heating (100°C min<sup>-1</sup>), whereas coking behavior is assessed during slow controlled heating at a rate of 3–5°C min<sup>-1</sup>.

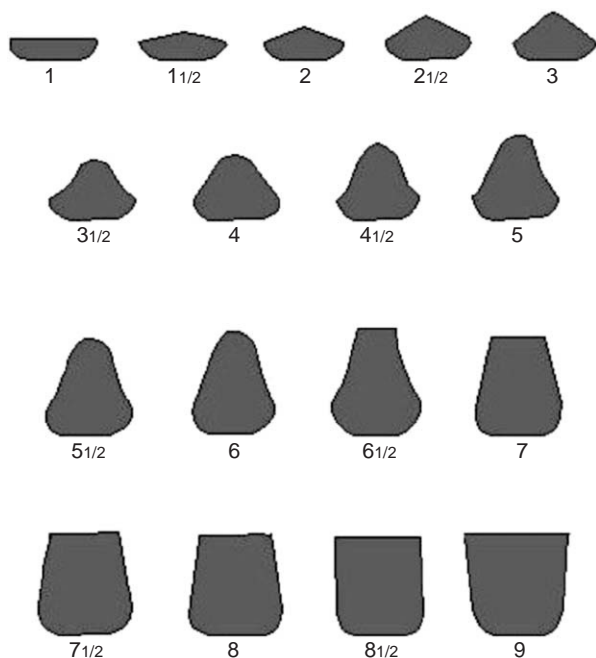
An appraisal of caking and related rheological properties is consequently not simple and usually requires measurement of several parameters. Distinct laboratory tests to assess these properties, swelling and caking, have been developed to simulate the coal behavior and the experimental conditions in coke ovens. Some of them have been standardized and are widely used to evaluate the suitability of a coal or coal blend for coke production. It has been reported that swelling and caking are therefore mutually reinforcing phenomena and that swelling is simultaneously the cause and the consequence of caking. It should be pointed out that coal blending has been adopted by the industry because of the limited availability and high cost of good coking coals and the continued demand for better quality coke for the blast furnace. Coal blends must be low in cost, produce a high-quality coke, and provide a safe oven pushing performance. They are composed of coals (four or more) differing in rank, rheological properties, geographical origin, and their proportion in the blend. As aids to coal selection and coke quality prediction, several mathematical models have been developed and extensively reported. Most of them

are based on parameters reflecting the rank, maceral behavior during carbonization (the concept of reactive and inert), and plastic properties of coal.

Some of the most common methods to evaluate the suitability of a coal or coal blend for coke production are the free-swelling index test, Gray–King assay, Roga assay, various dilatometer tests, and Gieseler plastometer test. In addition, research using larger amount of coal sample is also performed by means of a pilot coke oven.

### Free-Swelling Index

The free-swelling index (FSI), also called the crucible-swelling index, is the simplest test to evaluate whether a coal is of potential value for coke manufacture. It can be taken as a preliminary separation of coking and noncoking coals and it provides some measure of relative coking characteristics. It does not reflect, however, essential coking characteristics such as plasticity. This test involves rapid heating of a small sample of crushed coal in a standardized crucible to a temperature of  $\sim 800^{\circ}\text{C}$ . After heating, a small coke ‘button’ remains in the crucible. The cross-sectional profile of this coke ‘button’ is compared to a series of standard profiles numbered 1–9 with 1/2 increments – **Figure 3** (ISO 501, ASTM D720). The lower the profile of the coke ‘button’ formed, the less the free-swelling power and/or caking power of coal. An index of 0 is assigned to a noncoherent and pulverulent coke ‘button’ and indices of 1–3 are usually taken as the coal is only weakly caking. A high FSI (8–9) is also

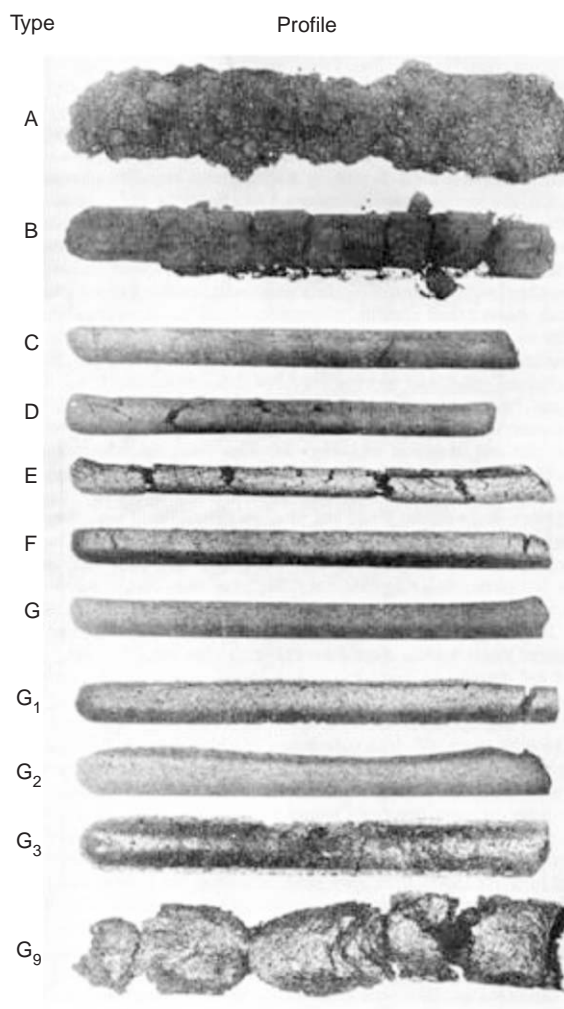


**Figure 3** Reference profiles of the free-swelling index test.

undesirable for cokemaking, because the coke is too weak with thin walls of its pores.

### Gray–King Assay

It is a predominantly visual assessment of agglomeration and swelling properties when an unconfined sample of coal (with or without inert diluent) is heated in a horizontal oven. The coal sample is slowly heated in a tube at  $5^{\circ}\text{C min}^{-1}$  up to  $600^{\circ}\text{C}$ . The appearance of the resulting coke is examined visually and classified by comparison with a series of reference cokes designated as A, B, C,..., G (ISO 502). The coal may remain in powder form, either granular or weakly coherent (coke types A, B, C), or it may be fused, but reduced in volume (coke types D, E, F) – **Figure 4**. Coal given a well-fused coke is classified as type G. If coal has a very high swelling power, it is tested with variable proportions of electrode carbon. The proportion of added carbon



**Figure 4** Gray–King coke types and profiles.



**Table 5** Comparison between free-swelling index, Gray–King coke type, and Roga index

Free-swelling index	Gray–King coke type
0–1/2	A–B
1–4	C–G2
4 1/2–6	F–G4
6 1/2–8	G3–G9
8 1/2–9	G7 and above

Free-swelling index	Roga index
0–1/2	0–5
1–2	5–20
2 1/2–4	20–45
> 4	> 45

Data taken from Ward CR (1984) *Coal Geology and Coal Technology*. Oxford: Blackwell Scientific.

characterizes the coal and the coke type is designated as G<sub>1</sub>–G<sub>8</sub>. Although it is not determined under the same conditions, the Gray–King coke type corresponds broadly to the FSI (Table 5).

### Roga Index

This test, originally developed in Poland, indicates the caking power of coal under standardized conditions (ISO 335). It is based on pressing a coal sample mixed with a standard anthracite under a small load at 850°C for 15 min in a standard crucible. The cohesion of the resultant coke is then tested in a drum tumbler, and the Roga index determined as the percentage of coarse material remaining. A general correspondence exists between the FSI and the Roga index. A Roga index of 45 indicates a coking coal with an FSI of  $\sim 3\frac{1}{2}$  to 4; the stronger coking coals have Roga indices above 45 (Table 5).

### Dilatometer Tests

Dilatometers allow the study of two types of phenomena: (1) the behavior of coal during fusion and softening; and (2) the contraction of semicoke or green coke beyond the temperature of resolidification. The dilatometry-based tests measure the volume changes when a compressed pencil of coal, confined by a piston, is heated in a vertical retort (ISO 349, ISO 8264, and ASTM D5515). This type of test differs from those described above in that swelling is not free, either because the coal has first been agglomerated by compression and because the piston is under load. It uses slow heating (usually, 3°C min<sup>−1</sup>) in an attempt to simulate the behavior of coal in a coke oven. The most commonly used dilatometer is that developed by Audibert–Arnu. A detailed description of the sample preparation,

Audibert–Arnu dilatometer, and the experimental conditions used are given in the International Standard ISO 349.

The Ruhr dilatometer test (ISO 8264), used in Germany and many other countries, is a modification of the Audibert–Arnu dilatometer test. By this test, the coking capacity G or G-factor, used for coke quality prediction, is calculated.

### Gieseler Plastometer Test

In coke production, crushed coal passes through a fluid or plastic state to become a fused porous solid carbon material (semicoke). The temperature over which coal exists in this fluid state and the extent of fluidity development are critical factors in coke manufacture, because they are responsible for good adhesion between reactive components of coal and inert particles. These characteristics of a coal or a coal blend (softening, reduction in viscosity, resolidification) are tested by using a constant-torque Gieseler plastometer (a specific type of viscosimeter) as described in the ASTM D2639 standard procedure. In this test, the coal sample (5 g with a particle size less than 0.425 mm) is heated from 300°C to 550°C at a rate of 3°C min<sup>−1</sup>. The parameters derived from this test are: (1) softening temperature, the temperature at which the coal starts to be fluid; (2) the temperature of maximum fluidity, the temperature at which the fluidity of the coal reaches a maximum; (3) resolidification temperature, the temperature at which the fluid mass resolidifies; (4) plastic range, which is defined as the difference between the resolidification and softening temperatures; and (5) maximum fluidity, obtained as dial divisions per minute (ddpm). Maximum fluidity may also be reported as the logarithm (base 10) of that value. The maximum fluidity is one of the most important parameters used to characterize coals for cokemaking. Parameters maximum fluidity and plastic range are of great relevance in the compatibility of coals for blend preparation.

### Coking Pressure Measurements

During carbonization, the volatile matters that evolve as a result of the thermal coal decomposition try to escape through the different layers formed in coke ovens. There is certain resistance to the passage of the volatile products and, consequently, a pressure is developed against the chamber walls, which is known as coking pressure. Certain coals generate high gas pressures in coke ovens that cause operational difficulties during coke pushing and coke-oven wall damage, thereby shortening coke oven life. This

type of coking coals, namely dangerous coals, are not carbonized individually, but are used as components in industrial coking blends in order to adjust the volatile matter and fluidity of the blend and to improve coke mechanical strength. It is therefore necessary to be able to predict and assess the danger of a coal, and keep the coking pressure below certain limits in order to prolong the life of the coke oven.

Two basic methods of approach can be undertaken to determine the suitability of a coal or coking blend: (1) indirect coking pressure measurements at laboratory scale and (2) direct coking pressure measurements by larger-scale tests, using a few hundreds of kilograms of coal while trying to reproduce industrial conditions.

Generally, the first type of tests involves heating of a coal sample while it is compressed by a piston acted by a standard load. As carbonization proceeds, the pressure generated in the coal charge causes movement of the piston, which is monitored. Depending on the expansion and/or contraction values observed, coals can be classified as very dangerous, dangerous, or nondangerous (safe). Among the laboratory-scale tests developed to monitor dangerous coals, the Koppers test and its variants and the sole-heated oven test are briefly described. The principal advantages of such tests are the shorter time, lower cost, and a smaller amount of coal required for testing.

### Koppers Test

The Koppers test uses an oven with unidirectional heating from one side. By this test, carbonization is performed with a sample of 80 g of coal that is subjected to a constant pressure of 10 kPa by means of a piston and the change in volume is measured. Modifications to the early Koppers test have been introduced by INCAR (Spanish Patent No. 524.258, 1983) based on Mott and Spooner's modifications. They include the heating system and experimental conditions such as pressure on the charge, bulk density, and the rate and duration of heating. As a result of the research conducted on several coals and coal blends, a criterion was established to classify coals. Coals giving a contraction greater than 10 mm and no expansion can be considered to be not dangerous during carbonization.

### Sole-Heated Oven Test

An ASTM procedure specifies the conditions under which the expansion or contraction of coal or coal blends during carbonization can be measured (ASTM D2014). This type of coking oven has unidirectional heating from one side (the sole).

### Movable Wall Oven

Direct measurement of coking pressure can be obtained in a movable wall oven, which is widely used to estimate coking pressure of a coal and a coking blend. A characteristic feature of these ovens is that one wall is mounted on runners such that it can move, or tend to move, away from the other oven wall. Its movement is nowadays restrained by a load cell that measures the force necessary to prevent the wall movement. By monitoring this force during a coking cycle, the pressure exerted by the charge on the wall can be calculated. There is no standard oven and procedure, although all movable wall ovens work on the same basic principles. These ovens with 250–400 kg capacity have similar width to commercial ovens, so that heating regimes can be accurately reproduced on the pilot scale. In addition to coking pressure measurements, the amount of coke produced allows to run full-scale coke tests in terms of physical and mechanical properties and reactivity to carbon dioxide.

### Coke

Metallurgical coke is a macroporous carbon material of high strength and relatively large lump size produced by the carbonization of coals with a specific rank or of coal blends at temperatures up to 1400 K. In conventional coke production, coal blends crushed to  $\sim 80\%$  less than 3 mm particle size are usually carbonized in batteries of coke ovens. The ovens are indirectly heated through the side walls at a temperature of  $\sim 1300^\circ\text{C}$  over a period of 18–20 h. About 90% of the coke produced from coal in the world is used to maintain the process of iron production in the blast furnace where it has three major roles: (1) as a fuel, it provides heat for the endothermic requirements of chemical reactions and the melting of slag and metal; (2) as a chemical reducing agent, it produces gases for the reduction of iron oxides; and (3) as a permeable support, it acts as the only solid material in the furnace that supports the iron-bearing burden and provides a permeable matrix necessary for slag and metal to pass down into the hearth and for hot gases to pass upwards into the stack.

Of these three roles, the first two can be substituted by oil, gas, plastics, and coal, which are injected at the tuyeres as generating energy and a carbon source. Such a substitution brings about a reduction in coke rates for the blast furnace. (Coke rate is the weight of coke required to produce 1 ton of iron). However, there is no other satisfactory material available, which can replace, fully or partially,



metallurgical coke as a permeable support of blast-furnace charge.

An assessment of the coke performance in the blast furnace operating with or without injection technology should include those properties of coke that reflect its resistance to degradation under the chemical and thermal environments of the blast furnace. Such properties are related to lump size, shape and size uniformity, chemical composition, mechanical strength, and thermal and chemical stabilities. Thus, coke for the blast furnace needs to be a successful compromise between structure and properties. To ensure good blast furnace performance, coke should be moderately large, with a narrow size range, and have a high mechanical strength and a high resistance to abrasion and thermal shock in the blast furnace. Because of the many unknown factors, it is not possible to establish universal quality indices common to all blast furnaces, although typical specifications for metallurgical coke quality are available. It could be said that each blast furnace depending on design and operation requires a tailored coke.

Impurities present in coke (moisture, volatile matter, ash, sulfur, phosphorous, and alkali contents) affect its performance in the blast furnace by decreasing its role as a fuel in terms of amounts of carbon available for direct and indirect reduction roles and also its role as a permeable support.

As in the case of coal, all determinations included in proximate analysis (moisture, ash, and volatile matter contents) are the subject of national and international standards (ISO 579 for the determination of total moisture content; ISO 687 and ASTM D3173 for the determination of moisture in the analysis sample; ISO 1171 and ASTM D3174 for the determination of ash; ISO 562 and ASTM D3175 for the determination of volatile matter content). **Table 6** summarizes typical coke chemical properties for some operating blast furnaces in Europe.

### Physical Tests

The importance of the physical properties of coke is linked to the need to support the ferrous burden and to give a permeable matrix through which reducing gases can flow and molten material can percolate in the lower blast-furnace region. These physical properties are related to its size (mean and distribution) and its resistance to breakage and abrasion. Coke size is mostly controlled by screening (ISO 728, ASTM D293). A large mean size with a narrow size distribution maintains adequate permeability. Most operators consider a mean optimum size to be in the range of 50–55 mm with a lower limit ~20–30 mm and an upper limit ~70–100 mm.

**Table 6** Required chemical properties of blast furnace coke

<i>Chemical property</i>	<i>European range</i>
Moisture (mass%)	1–6
Volatile matter (mass% db)	< 1.0
Ash (mass% db)	8–12
Sulfur (mass% db)	0.5–0.9
Phosphorous (mass% db)	0.02–0.06
Alkalies (mass% db)	< 0.3

Data taken from Leonard DC, Bonte L, Dufour A, Ferstl A, Raipala K, Scmole P, Schoone P, Verduras JL, and Willmers RR (1996) Coke quality requirements of European blast furnace engineers (joint EBFC-Paper). In: *Proceedings of the Third European Cokemaking Congress*, pp. 1–10. Gent, Belgium, CRM-VDEh.

Coke is a porous carbon material consisting of a network of pores of various dimensions and shapes, some of which are closed, but the majority of which are interconnected. Generally, porosity of industrial cokes is determined indirectly from the ratio of apparent and true relative densities, according to the equation

$$\text{porosity (\%)} = 100 \left( 1 - \frac{\text{apparent density}}{\text{true density}} \right)$$

A full description of the determination of apparent and true specific densities of lump coke is given in the ISO 1014 and ASTM D167 procedures.

Empirical mechanical strength tests, commonly used to measure resistance to size degradation, involve dynamic loading either in the form of shatter tests (ISO 6161, ASTM D3038), where breakage occurs by impact, or revolving drum tests such as ASTM Tumbler (ASTM D3402), MICUM, half- and extended-MICUM, IRSID, and JIS (JIS K2151) tests, where attrition takes place by a combination of breakage and abrasion. In Europe, the MICUM and IRSID (ISO 556 and ISO 1881) tests, which use the same equipment, are dominant. The JIS test is widely used in Japan and Australia, while the ASTM Tumbler is commonly used in North America. **Table 7** summarizes the cold mechanical strength methods for coke testing.

All these tests are based on the mechanical treatment of a specific amount of coke (10–50 kg) with a defined size (>60 to >20 mm) performed in a rotating drum under well-defined conditions (number of revolutions and rate). Afterwards, the coke is sieved and the different size fractions weighted. During mechanical treatments, coke fragmentation takes place by fissuring, cohesiveness, and abrasion. Two indices are normally derived from these tests: one referred to fissuring or cohesion and the other to abrasion.

**Table 7** Standard tests for assessing mechanical strength of coke

	<i>MICUM</i>	<i>Half-MICUM</i>	<i>IRSID</i>	<i>ASTM Tumbler</i>	<i>JIS</i>
<i>Standard procedure</i>	ISO 556	ISO 556	ISO 1881	ASTM D294	JIS K2151
<i>Coke characteristics</i>					
Weight (kg)	50	25	50	10	10
Particle size (mm)	>60	>60	>20	51–76	>50
Type of sieve	Rounded hole	Rounded hole	Rounded hole	Square hole	Square hole
<i>Drum characteristics</i>					
Length (m)	1	0.5	1	0.46	1.5
Diameter (m)	1	1	1	0.91	1.5
<i>Test characteristics</i>					
Drum (rpm)	25	25	25	24	15
Duration (min)	4	4	20	58	2 and 10
Total revolutions	100	100	500	1400	30 (2 min); 150 (10 min)
<i>Strength indices</i>					
Breakage	M <sub>40</sub> = mass% > 40 mm	As for MICUM	I <sub>40</sub> and I <sub>20</sub> = mass% > 40 and > 20 mm	Stability factor = mass% > 25 mm	DI30/15 = mass% > 15 mm; DI150/ 15 = mass% > 15 mm
Abrasion	M <sub>10</sub> = mass % < 10 mm	As for MICUM	I <sub>10</sub> = mass% < 10 mm	Hardness factor = mass% 6.3 mm	

Adapted from Loison R, Foch P, and Boyer A (1989) *Coke. Quality and Production*. London: Butterworth.

### High-Temperature Tests

The importance of high-temperature properties of coke was established by the Japanese Steel Industry from the dissection surveys of three blast furnaces that were quenching whilst operating. As a result the reactivity of coke to carbon dioxide was considered one of the most important characteristics to assess the quality of coke and the suitability of its use in the blast furnace. A high reactivity of cokes is disadvantageous because of a higher specific consumption of coke in the blast furnace. In addition, the reactive coke is mechanically less resistant, affecting the conditions of the gas flow in the blast furnace. Coke reactivity is commonly determined using a method developed by the Nippon Steel Corporation (NSC), which has been adopted as an ASTM standard procedure (ASTM D5341). The method measures the solution loss reaction of coke by carbon dioxide at 1100°C for 2 h on a dried coke sample of 200 g ( $20 \pm 1$  mm in size) under standardized gas flow rate conditions. The coke residue is cooled in nitrogen and the weight loss after reaction is defined as the coke reactivity index (CRI). The gasified coke is subjected to a mechanical strength test in a drum (600 revolutions at 20 rpm). The coke postreaction strength or coke strength after reaction index (CSR) is defined as the percentage of coke >9.5 mm or >10 mm remaining after the mechanical treatment. For cokes produced under similar coking conditions,

there is a high degree of correlation between the two indices, CRI and CSR both determined by NSC (Nippon Steel Corporation) method. It has been established that CRI is strongly influenced by the rank and type and quantity of mineral matter of the parent coal and CSR is also primarily controlled by coal properties and in general by coking conditions and oven design.

The range of optimum values in coke quality parameters is wide and strongly dependent upon the characteristics and operational conditions of the blast furnaces. As an example, typical coke size, mechanical strength before and after reaction with CO<sub>2</sub>, and reactivity index of cokes used in different blast furnaces worldwide are given in Table 8.

### Coke Petrography

Optical microscopy, using polarized reflected light, has been recognized as a very useful tool and has played a major role in studies of carbonization mechanisms and coke quality properties. Examination of a coke polished surface under a microscope reveals the presence of certain parts without optical activity (isotropic coke) and others with optical activity (anisotropic coke). The anisotropic carbon exhibits yellow, blue, and purple areas of various sizes, <1–200 μm, and shapes (optical texture) with an interchange of color on rotation of the specimen,

**Table 8** Required physical and high-temperature properties of blast furnace coke in current operation

	European range	Australian BHP Port Kembla	American range	Japan range
Mean size (mm)	47–70	50	50	45–60
M <sub>40</sub>	> 78–> 88	85	n.a.	n.a.
M <sub>10</sub>	< 5–< 8	6.5	n.a.	n.a.
I <sub>40</sub>	53–55	n.a.	n.a.	n.a.
I <sub>20</sub>	> 77.5	n.a.	n.a.	n.a.
DI150/15	n.a.	84.4	n.a.	83–85
ASTM stability	n.a.	63.6	60	n.a.
CSR	> 60	74.1	61	50–65
CRI	20–30	17.7	23	n.a.

n.a.: not available.

After Díez MA, Álvarez R, and Barriocanal C (2002) Coal for metallurgical coke production: Predictions of coke quality and future requirements for cokemaking. *International Journal of Coal Geology* 50: 389–412.

in association with the use of a retarder plate. Each color represents a given orientation of the aromatic lamellar molecules that constitutes the carbon matrix. To describe the different components of the optical texture of coke, different classifications and nomenclatures have been developed by research centers and industry. Most of them make a distinction between isotropic, mosaics of various sizes, flow-type anisotropy of various sizes and shapes, and inerts. Quantification of the optical textural components can be conducted by means of a point counting technique on a sample embedded in resin (e.g., ASTM D3997 and ASTM D5061).

By combining this microscopic technique with an image analyzer, the analysis program allows the determination of total porosity of coke, the number, the total perimeter, the average diameter, and the wall thickness of pores.

The use of optical microscopy for qualitative and quantitative analysis of the optical texture of coke is performed rather in the field of research on carbonization than in a full characterization of coke by the industry.

It is established that the rank and chemistry of the parent coal strongly influence the optical textures of cokes. Several studies show that the development of anisotropy (size, shape, and intensity) during carbonization varies mainly with (1) coal rank; (2) petrographic composition of the coals; (3) plasticity of the parent coal; (4) carbonization conditions such as rate of heating, soak time, and gas overpressure; as well as (5) the nature of additives used in the coal blends.

*See also:* **Activation Analysis:** Neutron Activation. **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Carbon. Geochemistry:** Soil, Organic Components. **Humic and Fulvic Compounds. Microscopy:**

**Overview. Microscopy Techniques:** Light Microscopy; Sample Preparation for Light Microscopy; X-Ray Microscopy. **Sample Handling:** Comminution of Samples. **Sampling:** Theory; Practice. **Sulfur. X-Ray Fluorescence and Emission:** Energy Dispersive X-Ray Fluorescence.

## Further Reading

- Alpern B and Lemos de Sousa MJ (2002) Documented international enquiry on solid sedimentary fossil fuels; Coal: definitions, classifications, reserves-resources and energy potential. *International Journal of Coal Geology* 50: 3–41.
- Bertkowicz N (1979) *An Introduction to Coal Technology*. London: Academic Press.
- Couch GR (2001) *Metallurgical Coke Production*. London: IEA Coal Research.
- Davidson RM and Clarke LB (1996) *Trace Elements in Coal*. London: IEA Coal Research.
- Elliot MA (1981) *Chemistry of Coal Utilization*, Second supplementary volume. New York: Wiley.
- Kural O (ed.) (1994) *Coal. Resources, Properties, Utilization, Pollution*. Turkey: Orhan Kural.
- Marsh H and Rodríguez-Reinoso F (eds.) (2000) *Sciences of Carbon Materials*. Spain: Secretariado de Publicaciones, The University of Alicante.
- Stach E, Mackowski MTh, Teichmüller M, *et al.* (1982) *Stach's Textbook of Coal Petrology*, 3rd edn. Berlin: Gebrüder Borntraeger.
- Scott AC (2002) Coal petrology and the origin of coal macerals: A way ahead. *International Journal of Coal Geology* 50: 119–134.
- Scott DA (1994) *Development Affecting Metallurgical Uses of Coal*. London: IEA Coal Research.
- Van Krevelen DW (1993) *Coal, Typology–Physics–Chemistry–Constitution*, 3rd edn. Amsterdam: Elsevier.
- Ward CR (1984) *Coal Geology and Coal Technology*. Oxford: Blackwell Scientific.
- Zimmerman RE (1979) *Evaluating and Testing of Coking Properties of Coal*. San Francisco: Miller Freeman Publication.

## COD

See WATER ANALYSIS: Chemical Oxygen Demand

## COFFEE

See FOOD AND NUTRITIONAL ANALYSIS: Coffee, Cocoa, and Tea

## COKE

See COAL AND COKE

## COLLOIDS

See FIELD-FLOW FRACTIONATION. WATER ANALYSIS: Particle Characterization

## COLOR MEASUREMENT

**J B Hutchings**, Bedford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Color is a perception. It is the human psychological response to electromagnetic radiation within the wavelength range of  $\sim 380\text{--}710\text{ nm}$ . In normal daylight the peak response occurs at  $550\text{ nm}$ . Color is an interdisciplinary subject in both its fundamental and applied levels. This has resulted in the use of a number of units to describe spectral colors. These colors do not occur at a single wavelength, energy, or frequency, but over bandwidths. The approximate equivalents are shown in Table 1. These are the colors on a neutral surrounding that are seen by a person of normal color vision.

The color of a substance is due to emission, if the substance is self-luminous, or, if it is not self-luminous,

to selective absorption during transmission or reflection of incident radiation. There are many different physical and chemical causes of color production; these are outlined below.

### Simple Excitations and Vibrations

External energy supplied to the substance is converted into thermal motion, and atoms are raised to an excited state by promotion of electrons to higher-energy orbitals.

After a short time each electron returns to its normal orbital, and the energy released may be manifested as radiation of a particular frequency (Planck's law). Incandescence is caused by heating, for example, in wire filaments and carbon arcs. Electrical stimulation is one cause of such excitation of specific atoms in vapors and gases. This occurs in sparks and mercury vapor lamps. In fluorescent tubes, ultraviolet (UV) emissions from mercury vapor lamps are

**Table 1** The visible spectrum and the approximate equivalence of units used to describe it

Energy (eV)	Color of light	Wavelength (nm)	Frequency ( $\times 10^{14}$ Hz)	Color of material that absorbs the light in this band
	UV			
3.1	Violet	420	7.5	Greenish yellow
2.7	Blue	450	7.0	Yellow
	Bluish green	490		Red
2.4	Green	520	5.7	Magenta
2.2	Yellow	580	5.2	Blue
2.0	Orange	600	4.9	Cyan
1.8	Red	650	4.3	Bluish green
	IR			

converted into lower-energy visible emissions by phosphor coatings. The coherent light produced by helium/neon gas lasers results from gas excitation amplified by optical feedback. Chemical excitation occurs in analytical flame tests (e.g., the yellow of sodium), and in high-energy displays of auroras.

Most molecular vibrations are of relatively low energy, and therefore interact in the infrared (IR) region. The presence of hydrogen bonding in water and ice raises vibrational energies and some absorption takes place at the red end of the visible spectrum. Thus, when pure and present in bulk, water and ice are perceived to be blue.

### Transitions Involving Ligand Field Effects

Compounds of the s- and p-block elements are almost always white, because the electron transition energies fall in the UV region. However, many compounds of transition elements are colored because they have unpaired electrons in d- or f-orbitals. Transitions taking place between these orbitals require less energy, and thus emissions occur in the visible range. Such colored emissions can occur in transition-metal compounds such as minerals and paint pigments, and when the transition-metal ion occurs as an impurity, as in many gems and glasses. The color depends on the energy difference between the two levels. Thus, in some complexes, the color depends on the type of ligand. For example,  $[\text{Ni}(\text{NH}_3)_6]^{2+}$  is blue,  $[\text{Ni}(\text{H}_2\text{O})_6]^{2+}$  is green, and  $[\text{Ni}(\text{NO}_2)_6]^{2+}$  is brown-red.

### Transitions between Molecular Orbitals

Highly colored organic molecules, such as those used as dyes or food colorants, are complex and unsaturated. Electrons are held in conjugated systems, and their excited states occur at comparatively low

energies, leading to the absorption of light of visible wavelengths. Longer conjugated chains require lower absorption energies for excitation. Color shifts are also achieved where there are electron donors, such as  $-\text{NH}_2$  or  $-\text{OH}$  groups, or electron acceptors, such as  $-\text{NO}_2$  or  $-\text{C}=\text{O}$  groups, which pump electrons into or out of the conjugated system, respectively. Re-emission may occur as fluorescence.

The colors of a number of inorganic species arise from charge transfer. For example, the d-levels of Cr(VI) and Mn(VII) are empty and their compounds ought to be colorless. However, charge transfer leads  $\text{CrO}_4^{2-}$  and  $\text{MnO}_4^-$  to be strongly colored. These transitions also cause the intense colors in blue sapphire and ultramarine.

### Transitions Involving Energy Bands

In certain types of materials, electrons are distributed in energy bands. The form of the band depends on the particular atomic orbitals and the atomic spacing and geometry. Electrons available for bonding occupy various levels within the band. In metals, light excites electrons within the band, and energy is strongly absorbed. The electromagnetic radiation induces electric currents within polished metal surfaces, and light is strongly re-emitted. The absorption and re-emission efficiencies depend on optical energy. If the efficiencies are equal at all energies in the visible range, polished metals will be silver when illuminated by white light. Colloidal metal particles dispersed in a medium also produce energy bands as in, for example, ruby glass.

Gaps may appear in energy bands to form a lower-energy (valence) band and a higher-energy (conduction) band. The lowest energy that can be absorbed is determined by the gap size. Large band gap materials such as diamond (5.4 eV) cannot absorb visible light and are colorless. Yellow cadmium sulfide (2.6 eV) absorbs somewhat in the blue, and red mercury(II) sulfide (vermillion) (2.0 eV) is able to



absorb blue, green, and yellow light. Pure semiconductors possessing lower-energy gaps (say 1.6 eV) are black.

Doped semiconductors possess donor or acceptor atoms within their band gaps. These effectively decrease the gap size and allow energy absorption. Diamond consists of carbon atoms, each having four valence electrons, while nitrogen atoms have five valence electrons. Where a small percentage (say 0.001%) of the carbon atoms has been replaced by nitrogen donor atoms, spare electrons entering a donor level within the band gap can be donated to the empty conduction band. Broadening of the donor level by heating, for example, leads to absorption at the blue end of the spectrum, allowing the diamond to appear yellow. Conversely, boron, which has three valence electrons, creates an acceptor hole that is filled by an electron from the full valence band. The energy needed for the transfer is very small, and leads to a lower-energy absorption and a red diamond. Other examples of these mechanisms can be found in light-emitting diodes, and some lasers and phosphors.

Transparent crystals and glasses often appear colored because they contain color centers. These encompass another form of band gap phenomenon. Sodium chloride is an ionic crystal consisting of a three-dimensional (3D) array of  $\text{Na}^+$  and  $\text{Cl}^-$  ions. When  $\text{Cl}^-$  is missing from the lattice, the presence of a free electron in its place can restore electrical neutrality. This is an F-centre (after the German *Farbe* = color), which forms a trapping energy level within the wide band gap of this white semiconductor crystal. An electron can be raised from the valence band into the trap by  $\gamma$ -, X-, or UV-radiation. The energy level (2.7 eV) present within the trap of an irradiated NaCl crystal containing such a defect (color center) can absorb blue light and leads to a yellow–brown color. Other examples of materials containing color centers are amethyst, smoky quartz, some fluorescing materials, and lasers.

### Geometrical and Physical Optics

The dispersion of light may be achieved by refraction through a prism. This dispersion occurs because the refractive index  $n$  of the prism material and the velocity of light passing through it vary with frequency of the light. Anomalous dispersion results when light absorption occurs in an otherwise transparent material. The complex refractive index ( $N$ ) is used to explain the behavior of such a material:  $N = n + ik$  where  $i$  is the square root of  $-1$  (an imaginary number), and  $k$  is an absorption coefficient. At

wavelengths where absorption occurs, the resonating frequency of the absorber interacts with the frequency of the light, and there are changes in both  $n$  and  $k$ . Such effects are also found in rainbows, and cause the flashes of fire in diamonds.

Scattering occurs when light is deflected in all directions, after striking fine particles or an irregular surface. The presence of scattering particles is not necessary because all materials scatter light to some extent. When light is scattered by particles that are very small compared with the wavelength of the light, the amount of scatter is related to the inverse of the fourth power of the wavelength. Scattering accounts for the blueness of the sky and watered-down skimmed milk, for redness at sunset, and for many blue and green bird-feathers. The Mie theory must be used to account for the scattering effects of larger particles. This theory accounts for phenomena in which scattering intensity and wavelength depend on the angle of view.

The interference of polarized white light produces the colors found in optically anisotropic materials. This phenomenon is made use of in the analysis of photoelastic stress patterns. When a single monochromatic light source is arranged so that it produces two overlapping beams, a series of interference fringes of alternate light and dark bands is obtained. The bands are the result of the alternate reinforcement and canceling of the coherent beam. Colors produced by interference without diffraction include those in nonstructured systems such as soap bubbles and oil slicks, and in structured systems such as mother of pearl and hummingbird feathers.

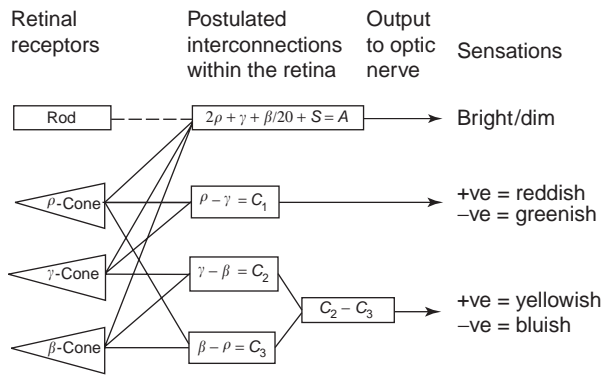
### Color Perception

The normal retina of the human eye has two types of detectors – rods, sensitive at low illumination levels, and cones, which operate under normal levels of daylight.

Their distribution is not uniform across the retina. There are three types of cone, each having different spectral sensitivities. The  $\beta$ -cones have peak sensitivities in the blue part of the spectrum, the  $\gamma$ -cones peak in the green, and the  $\rho$ -cones in the yellow–green. These different sensitivities to wavelength provide the basis for color vision.

There is a highly complex series of connections between the detectors and the fibers of the optic nerve, which takes the resulting signals to the visual cortex of the brain. There appear to be three types of nerve fiber, one of which is thought to carry the total achromatic or brightness information ( $A$ ), while the other two carry color information (**Figure 1**). The





**Figure 1** A postulated representation of the interconnections between receptors and the subsequent signals on which color sensations depend. (Reproduced with permission from Hunt RWG (1998) *Measuring Colour*, 3rd edn., p. 24. Kingston upon Thames: Fountain Press.)

brightness information consists of two parts, the rod response and all three cone responses weighted to compensate for the differing numbers of each type of cone. The color information from the retina is obtained from a series of signal differences and transmitted as two signals,  $C_1$  and  $(C_2 - C_3)$ .

These concepts can be used to indicate a visual basis of perceived color. Color can be thought of in terms of the brightness (by which an area appears to exhibit more or less light); hue (by which an area appears to be similar to one, or proportions of two, of the perceived colors red, yellow, green, and blue); and chroma, saturation, or colorfulness (by which an indication can be given of the amount or strength of hue present).

Good color vision is required for a number of tasks. These include judgments of color blending or matching, electrical and electronic circuit wiring, the setting up of color television sets, medical diagnosis, and medical/chemical analysis. Color vision standards exist for the armed services and public bodies concerned with transport.

Approximately 8% of men and 0.4% of women perceive colors in significantly different ways from the remainder of the population. Most of this color-deficient vision is inherited, but some can be acquired as a result of pathology, illness, or old age. The most common inherited causes relate to deficiencies or absence of the  $\rho$ - or  $\gamma$ -cones. The Ishihara charts may be used to detect the presence of such conditions. These charts consist of a number of plates made up of differently colored spots, which form numbers that may wrongly be described or not detected at all. This is a first screening for detection of a color vision deficiency.

Color vision probably evolved so that man could distinguish fruit from its surroundings. Our color

vision is not absolute but relative to other conditions in the environment, and this greatly assists in the recognition of the colors of objects in different viewing conditions. In consequence, failure to provide a consistent, rigorously observed set of conditions when the color of an object is being assessed or measured can lead to inconsistencies because of metamerism and other effects. Metamerism occurs when two colors match under one set of viewing conditions, but fail to match under a different set. Metamerism can be of four types: illuminant, observer, geometric, and field size.

## Measurement Principles and Instrumental Design

When viewing any object by reflection or transmission, three factors must be present for there to be a color: a light source, the object, and a viewing mechanism. If changes occur in any of these elements the color may change. Hence, any color measurement system must include a specification of these factors.

Color cannot be measured because it is perceived in the brain. However, it can be specified instrumentally using additive or subtractive mixing to produce a color that matches the object color. In the former, a match to most colors can be made by the additive mixing of light of the three primary colors. A measure of the intensity of the three primaries provides a specification of the color. Primary colors are defined such that none can be matched by mixtures of the other two. The principle of additive mixing is the basis of the common tristimulus and spectrophotometric methods of measurement.

Before these concepts could be developed to a generally acceptable measurement method, some agreement of the visual performance of a 'Standard Observer' was necessary. Observers with normal color vision obtained color matches for a series of wavelengths from 400 to 700 nm using additive mixing. The primaries were three monochromatic lights. The amount  $C$  of each color  $[C]$  was matched using amounts  $R$ ,  $G$ , and  $B$  of each particular stimulus  $[R]$ ,  $[G]$ , and  $[B]$ , where  $[R]$ ,  $[G]$ , and  $[B]$  represent red, green, and blue radiations, respectively. That is:

$$C[C] = R[R] + G[G] + B[B]$$

For most problems, it is convenient to separate the color quality, represented by the proportions of  $R$ ,  $G$ , and  $B$ , from the total intensity of light, represented by the absolute values of  $R$ ,  $G$ , and  $B$ . This is done by dividing  $R$ ,  $G$ , and  $B$  in turn by  $(R + G + B)$  to

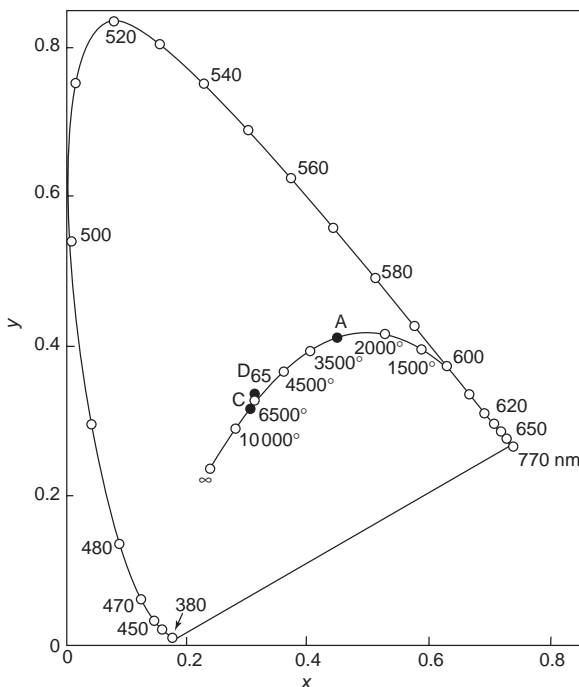
give:

$$1.0[C] = r[R] + g[G] + b[B]$$

where  $r = R/(R + G + B)$ ,  $g = G/(R + G + B)$ ,  $b = B/(R + G + B)$ , and  $r$ ,  $g$ , and  $b$  are chromaticity coordinates.

Before a match can be obtained at a number of wavelengths the spectral color has to be desaturated. That is, one of the three stimuli has to be added to the color being matched. For these colors one of the coordinates is negative. The use of three unreal primaries ( $X$ ,  $Y$ ,  $Z$ ) was proposed to overcome this disadvantage. These define, effectively, the three cone responses of an observer with normal color vision, using a standard subtended viewing angle. The resulting chromaticity coordinates ( $x$ ,  $y$ ,  $z$ ) are always positive. If two are specified the third can be calculated. Hence, a two-dimensional ( $x$ ,  $y$ ) chromaticity diagram can be plotted as in **Figure 2**. In this, spectral colors are shown as a spectrum locus. The line joining the two ends of the spectrum locus is the purple boundary, nonspectral colors that are mixtures of long and short wavelengths from the two ends of the spectrum. Two sets of unreal primaries have been specified corresponding to  $2^\circ$  and  $10^\circ$  subtended angles of viewing.

All colors can be located within the area enclosed in the ( $x$ ,  $y$ ) chromaticity diagram. The chromaticity



**Figure 2** The 1931 CIE ( $x$ ,  $y$ ) chromaticity diagram for a  $2^\circ$  observer, showing the Planckian locus and the chromaticities of standard illuminants A, C, and  $D_{65}$ .

of any color resulting from a mixture of two other colors falls on the line joining their chromaticities. Deep, spectrum-like colors have chromaticities near to the spectrum locus (or the line of purples). Paler, less chromatic colors fall nearer to the white point. The standard illuminant used in the measurement governs the position of the white point, which defines the position of the achromatic point. This can be defined in comparison with a Planckian radiator. As the temperature of such a body is increased, the color of its emission changes from red at  $\sim 1600$  K, to whiter at 4000 or 5000 K, to bluer at 8000–10 000 K. The locus of chromaticities of Planckian radiators is plotted in **Figure 2**. Although the Planckian radiator itself is impractical as a light source, it provides a color temperature reference against which more practical sources may be judged. The Commission Internationale de l'Eclairage (CIE) has defined a number of standard illuminants, and those most commonly reported are illuminants A, C, and  $D_{65}$ . Illuminant A is provided by a gas-filled tungsten lamp operating at a color temperature of 2856 K. Illuminants C and  $D_{65}$ , closer to ordinary daylight, have respective color temperatures of approximately 6774 and 6500 K, respectively.

A definition of a color stimulus includes the chromaticity and a relative intensity measure, the luminance factor. This is determined by the value of  $Y$ . This primary has a spectral form which is the same as the eye's overall sensitivity in daylight. Thus, a color can be specified by the three dimensions ( $x$ ,  $y$ ,  $Y$ ).

### Tristimulus and Spectrophotometric Measurement

The tristimulus colorimeter is the simplest means for instrumental color specification. A light source, filters, and a photodetector are combined such that together they yield a direct evaluation of the tristimulus values  $X$ ,  $Y$ , and  $Z$ . A spectrophotometer may also be used for color specification. That is, a spectral reflectance or transmittance curve ( $R$ ) can also be converted into  $X$ ,  $Y$ ,  $Z$  data. The curve ( $R$  versus wavelength  $\lambda$ ) is integrated over the visible range, with the spectral emission of an illuminant ( $E$ ), and the standard observer functions ( $\bar{x}$ ,  $\bar{y}$ ,  $\bar{z}$ ). The areas under the resulting curves yield the values of  $X$ ,  $Y$ , and  $Z$ ; that is:

$$X = k \int_{380}^{710} RE\bar{x} d\lambda, \quad Y = k \int_{380}^{710} RE\bar{y} d\lambda,$$

$$Z = k \int_{380}^{710} RE\bar{z} d\lambda$$

where  $k$  is a constant chosen so that  $Y=100$  for a perfect white. The spectral curve produced by a spectrophotometer may also yield useful information regarding possible chemical mechanisms or color-change kinetics of the system under investigation.

There are four instrument geometries recommended by the CIE. Quoting the illumination angle ( $^\circ$ ) first, and the viewing angle second, they are designated: 45/0, 0/45, and, where integrating spheres are used, diffuse/0 and near-0/diffuse. The 45/0 and 0/45 geometries by definition exclude the spectral component of the reflected light. Instruments using diffuse illumination or viewing can often be set up to include (SPINC) or exclude (SPEX) this component.

All measurements are made relative to a 'white standard', which the CIE nominate to be a perfect diffuser. Such perfection does not exist, but compensation for this can be made during measurement and calculation. A colored substandard is sometimes used to increase measurement reliability and sensitivity. This technique has been found particularly useful when interlaboratory color-quality specification procedures are being established. Instrumental performance may be checked with highly stable reflection or transmission standards.

The perfect sample for reflectance measurement is one that is flat, uniform in color, perfectly matt, and opaque. Results from nonperfect samples will be subject to error, and appropriate measurement techniques must be used. For example, the relationships between illumination, sample, and measuring areas are important to the successful color measurement of translucent materials.

Color measurements can be made of fluorescent materials, which absorb energy at one wavelength and re-emit it at another. A dual monochromator spectrophotometer (i.e., a spectrofluorimeter) can be used to illuminate the sample with monochromatic radiation, and record the energy reflected plus that re-emitted at each wavelength. In this way, a total radiance (reflected plus emitted energy) curve characteristic of the material can be built up. Values of  $X$ ,  $Y$ , and  $Z$  are calculated in the normal way. A simpler approach involves illumination of the sample with a white light source that emits the approximate amounts of near-UV wavelengths occurring in daylight. Tristimulus values of the resulting radiance can then be measured using tristimulus or spectral techniques.

A statement of the measurement geometry, standard illuminant, and standard observer ( $2^\circ$  or  $10^\circ$ ) must be a part of the color specification. When comparing results made on different occasions, or at different laboratories, or when comparing instrumental

measurements and visual assessments, great care must be taken to ensure that all measurement and calculation details correspond.

### Digital Analysis

The advent of digital analysis presents the opportunity for measurement of products and scenes in terms of total appearance analysis. That is, in terms not only of average color, achieved by conventional instrumentation, but also in terms of a detailed examination of the type of color variation that occurs across the surfaces of biological materials. Using color-calibrated digital technology measurements of gloss, surface texture (as well as visual structure) can also be attempted on irregular 3D materials, such as fruit and vegetables. The technology for specifying total appearance is based upon the digital camera that can capture images rapidly in digital format. These digital images can be easily processed, duplicated, modified, or transmitted via a network.

Unfortunately, camera  $R$ ,  $G$ ,  $B$  sensors do not have the same spectral sensitivities as the CIE standard observer. Hence, in order to measure an object in terms of device independent color from a digital camera, there is a need to correlate the camera  $RGB$  signals and CIE  $XYZ$  values. The most common technique for digital camera characterization consists of presenting the camera with a series of color patches in a standardized reference chart with known  $XYZ$  values and recording the averaged  $RGB$  signals for each patch. Polynomial fitting techniques are then applied to interpolate the data over the full range and to generate inverse transformations. An overall accuracy of  $\sim 0.5\text{--}2.0 \Delta E^*$  units is obtainable.

### Uniform Chromaticity Space and Uniform Color Scales

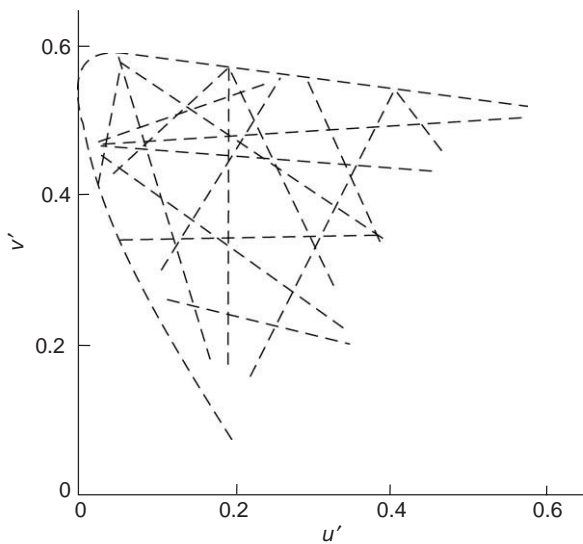
The  $(x, y)$  chromaticity diagram is not perceptually uniform. Colorant-using industries have long sought a color space in which a unit distance can be interpreted as equivalent to a unit of visual perception. The CIE-recommended space (CIE 1986) includes the 1976  $(u', v')$  uniform chromaticity diagram. The transformations are:

$$u' = 4X/(X + 15Y + 3Z) = 4x/(-2x + 12y + 3)$$

$$v' = 9Y/(X + 15Y + 3Z) = 9y/(-2x + 12y + 3)$$

Plotted in this space (Figure 3) are lines that are equivalent to equally perceivable differences. The ratio of the lengths of the longest and shortest lines, a measure of linearity, is  $\sim 4:1$ . The ratio is 20:1 for the untransformed  $(x, y)$  space.

The CIE has recommended two ways of combining luminance factor and chromaticity information.



**Figure 3** The CIE ( $u'$ ,  $v'$ ) chromaticity diagram showing lines of equally perceivable differences. (Reproduced by permission from Hunt RWG (1998) *Measuring Colour*, 3rd edn., p. 60. Kingston upon Thames: Fountain Press.)

These are the CIE 1976 ( $L^*$   $u^*$   $v^*$ ), or CIELUV, and the CIE ( $L^*$   $a^*$   $b^*$ ), or CIELAB spaces. The former was primarily intended for industries involved with additive color mixing, such as lighting and television. This space is obtained by plotting  $L^*$ ,  $u^*$ , and  $v^*$  at right angles to each other:

$$\begin{aligned} L^* &= 116(Y/Y_n)^{1/3} - 16 \quad \text{for } Y/Y_n > 0.008856 \\ L^* &= 903.3(Y/Y_n) \quad \text{for } Y/Y_n \leq 0.008856 \\ u^* &= 13L^*(u' - u'_n) \\ v^* &= 13L^*(v' - v'_n) \end{aligned}$$

where  $u'$  and  $v'$  are defined as described above, and where  $u'_n$ ,  $v'_n$  are the values of  $u'$ ,  $v'$  for the appropriate reference white.

The 1976 CIELAB space is obtained by plotting  $L^*$ ,  $a^*$ , and  $b^*$  at right angles to each other:

$$\begin{aligned} L^* &\text{ is defined as above} \\ a^* &= 500[(X/X_n)^{1/3} - (Y/Y_n)^{1/3}] \\ b^* &= 200[(Y/Y_n)^{1/3} - (Z/Z_n)^{1/3}] \end{aligned}$$

where  $X_n$ ,  $Y_n$ , and  $Z_n$  are values of  $X$ ,  $Y$ ,  $Z$  for the appropriate reference white.

Hue-angle ( $h_{uv}$ ), saturation ( $s_{uv}$ ), and chroma ( $C_{uv}^*$ ) may be calculated within CIELUV space. The CIELUV hue-angle is given by  $h_{uv} = \arctan(v^*/u^*)$ , where  $\arctan$  means 'the angle whose tangent is...'. Hue-angle is the angle from the  $u$ -axis moving anticlockwise.

The CIELUV chroma is given by

$$C_{uv}^* = (u^{*2} + v^{*2})^{1/2} = L^* s_{uv}$$

where  $s_{uv}$  = CIE 1976  $u$ ,  $v$  saturation =  $13[(u' - u'_n)^2 + (v' - v'_n)^2]^{1/2}$ .

Values of hue-angle and chroma can be similarly calculated for CIELAB space. Both the CIELUV and CIELAB spaces approximate to the Munsell system for the uniform spacing of colored samples of painted papers.

Many commercial colorimeters utilize variants of the Hunter 1958 (Lab) space. The coordinates are:

$$\begin{aligned} L &= 10Y^{1/2} \\ a &= \frac{17.5(1.02X - Y)}{Y^{1/2}} \\ b &= \frac{7.0(Y - 0.847Z)}{Y^{1/2}} \end{aligned}$$

where  $X$ ,  $Y$ ,  $Z$  are the 1931 CIE tristimulus values. Axes  $L$ ,  $a$ , and  $b$  are mutually perpendicular. Changes in  $a$  and  $b$  represent approximately the opponent nature of our visual response. Axis  $a$  represents the change from green ( $-a$ ) to red ( $+a$ ); axis  $b$  from blue ( $-b$ ) to yellow ( $+b$ ). An increasing value of  $L$  represents an increase in whiteness or lightness. Neutral colors are close to ( $a=b=0$ ). Similar associations apply to  $u'$ ,  $v'$ , and  $L^*$  in the CIELUV system and to  $a^*$ ,  $b^*$ , and  $L^*$  in the CIELAB system. Hue-angle ( $\arctan b/a$ ), and saturation index  $[(a^2 + b^2)^{1/2}]$ , give descriptive color information.

The fundamental units of color measurement are  $X$ ,  $Y$ ,  $Z$  and hence ( $x$ ,  $y$ ,  $Y$ ) space. All other color measurement spaces are derived from them.

### Subtractive Colorimetry

The methods described above are based on additive color mixing, but color can also be specified using subtractive colorimetry. Where the sample is transparent, colored glasses are used to subtract light from a standard source until the resulting color visually matches that transmitted by the sample.

Opaque colors can be similarly measured, by using the lamp to illuminate both the sample and the surface of a white diffuse reflector. The glass filters are then adjusted to produce a matching color on the diffuser.

The Tintometer system glasses are magenta (called red) to absorb green light; yellow to absorb blue; and cyan (called blue) to absorb red light. The glasses of each color are arranged in a numerical scale from practically colorless to highly saturated. The sample

color can be specified by the number of units of red, yellow, and blue required to make the match. In certain tintometers a visual match is made using two of the scales and a light intensity control. The results can be converted into  $X$ ,  $Y$ ,  $Z$  values.

### Color Difference Calculation

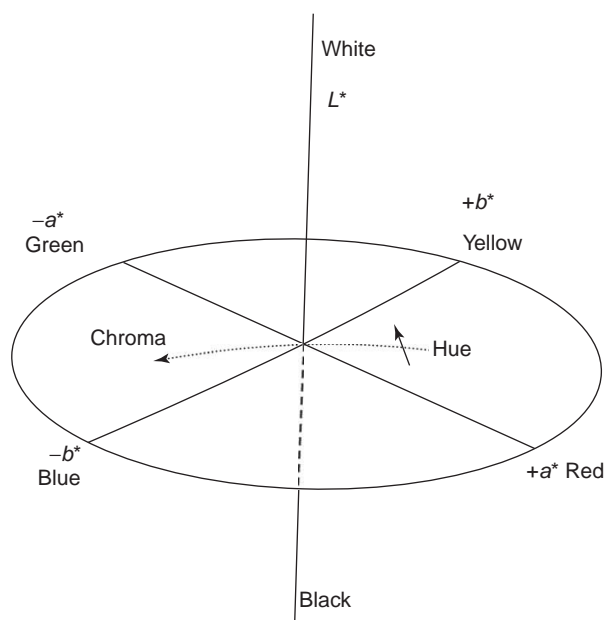
A number representing the difference between two colors can be calculated in any of the coordinate systems described above. The equation is based on the square root of the sum of the squares of the differences in each axis. For example:

$$\Delta E_{(\text{CIELAB})} = [(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]^{1/2}$$

The near-linear nature of (Lab), CIELUV, and CIELAB spaces make color difference calculation more meaningful through the whole of color space. More sophisticated equations yielding greater precision have been developed for specific industrial applications.

### Color Organization

Whatever the system of color measurement or color description ordering in 3D space is essential. In terms of the measurements described above relative values of  $L^*$ ,  $a^*$ ,  $b^*$  are indicated in **Figure 4** together with directions of increasing hue and chroma. However,



**Figure 4** The three dimensions of color measurements and color order systems.

color can also be organized, and hence specified, according to a color order system. Many such systems have been produced and those most well known and widely used are the Munsell System and the Swedish Natural Color System (NCS). Associated with these systems are atlases consisting of books of chips, which cover a wide color range, organized in logical sequences, also indicated by the scheme shown in **Figure 4**.

The Munsell color dimensions are hue ( $H$ ), value ( $V$ ), and chroma ( $C$ ). The Hue circle consists of 10 major hues, each divided into 10 visually equal steps. The central achromatic  $V$  (lightness) axis consists of 10 visually equal steps, extending from ideal black = 0 to ideal white = 10. The radial distance from this axis indicates an increase in  $C$ ; that is, an increase in hue content and departure from grey.  $C$  is zero at the achromatic axis, and increases in visually equal steps to 10, 12, 14 or more for particularly saturated colors. The Munsell atlas consists of pages of colored chips, one page for each hue. Chips are arranged so the vertical axis of the page represents an increase in  $V$ , the horizontal axis an increase in  $C$ . The Munsell description of a yellow-red color of hue 3YR, value 5 and chroma 6 is 3YR 5/6. Interpolation between whole units is possible, and single decimal places may be used where appropriate. Some product-control systems involve the use of two or more Munsell colors on a spinning disk. Their relative areas are changed until a color match is obtained. A large range of colors within the color saturation limits of the disks can be produced.

The Swedish NCS is based upon Hering's postulate that all colors may be placed in a system with reference to six elementary color sensations. These are: whiteness ( $V$ ), blackness ( $S$ ), yellowness ( $Y$ ), redness ( $R$ ), blueness ( $B$ ), and greenness ( $G$ ), again as indicated in **Figure 4**. Any one color may be specified in terms of the percentages of two chromatic and two achromatic attributes. Thus, a particular color can be described as 2030-Y90R, where 20 is the percentage blackness  $s$ , 30 is the percentage chromaticity  $c$  (where  $c$  is the total % of chromatic colors, say red and yellow), and Y90R is the hue of containing 10% yellowness and 90% reddishness. Note, the percentage whiteness normally omitted from the specification =  $(100 - 20 - 30)\%$ .

The conditions necessary for the use of color chips are relevant to all visual color-matching and assessment procedures. Such tasks must be done by individuals with normal color vision, using standard viewing and lighting conditions. The illumination should be white and diffuse, and there should be no interference from extraneous light sources, or reflections from the ceiling or other objects in the room.



The lamp often used for visual assessment is the artificial daylight fluorescent tube. This produces an emission spectrum intended to imitate approximately that of the standard D65 illumination specified in BS 950: 1967 part 1. To eliminate gloss effects, a  $0^\circ/45^\circ$  (or vice versa) relationship must be maintained between the illumination, the sample and the viewer. The sample should be placed on a neutral mid-grey to white background. The sample is assessed against one chip at a time, a grey mask being placed over other chips on the atlas page.

## Applications of Color Measurement

There are three main uses for color measurement. A major use is in systems in which pigments or dyes are manipulated to produce a product of specified or matching color. The second use is for the specification of a color when establishing a commodity standard, or for reference during, for example, the storage trial of labile material. The third use provides a means of understanding the human response to a product or visually perceived situation.

A number of interactions take place when light impinges upon a material. These determine the color and appearance of the material. The interactions take place (1) from the surface (specular and diffuse reflection, and refraction into the body of the material), (2) within the material (internal diffusion or scattering and absorption), and (3) through the material (regular and diffuse transmission).

The ability to handle, specify, and predict scattering and absorption properties is vital to the solving of many practical problems. An example is the prediction of colorant concentrations that will produce a match for an existing color. In some industries it is necessary to specify translucency. These problems can be approached using the Kubelka–Munk concept of a colorant layer of turbid material, such as a paint film. This postulates two light fluxes traveling in opposite directions across a thin layer within a strongly light-scattering material. The layer is thick compared with the size of pigment particles, but thin compared with the sample thickness. It has an infinite lateral dimension. As each light flux passes through the layer, it is affected by the absorption coefficient ( $K$ ) and the scattering coefficient ( $S$ ) of the material. The treatment is relevant to monochromatic radiation, and it does not allow for light lost through Fresnel reflection at the interface between the air and the medium. Also, the pigments are assumed to be randomly orientated, and the differential equations are only concerned with diffuse light traveling in two directions.

The upward flux ( $j$ ) is decreased by absorption  $-Kjdx$ , decreased by scattering  $-Sjdx$ , and increased by backscatter from the downward-proceeding flux  $+Si dx$ . Hence, the total change in the upward flux is:

$$dj = -(K + S)j dx + Si dx$$

Similarly, the downward-proceeding flux ( $i$ ) is changed by:

$$-di = -(K + S)i dx + Sj dx$$

The most quoted solutions to these equations, for a sample thickness  $X$ , concern the following relationships:

$$\frac{K}{S} = \frac{(1 - R_\infty)^2}{2 R_\infty}$$

$$a = \frac{1}{2}R + \frac{R_0 - R + R_g}{R_0 R_g} = \frac{(S + K)}{S}$$

$$R_\infty = a - b$$

$$T_i^2 = (a - R_0)^2 - b^2$$

$$SX = \frac{1}{b} \operatorname{arc} \coth \frac{1 - aR_0}{bR_0}$$

where the reflectance ( $R$ ) of a colorant layer is backed by a known reflectance ( $R_g$ );  $R_0$  is the reflectance of a layer with an ideal black background,  $R_g = 0$ ;  $R$  is the reflectance of an infinitely thick layer of the material;  $T_i$  is the internal transmittance;  $b = (a^2 - 1)^{1/2}$ ; and  $\operatorname{arc} \coth$  is an inverse hyperbolic function. Tables are widely available for calculating such hyperbolic functions. Two measurements of a thin layer of the sample, over backgrounds of white and black, enable  $K$  and  $S$  to be calculated.

Much work has been done in the field of colorant formulation. Where  $K$  and  $S$  behave linearly and additively with dye and scatterer concentrations, colorant formulation may be possible using the ratio  $K/S$ . Where this does not occur,  $K$  and  $S$  can be treated independently.

The Kubelka–Munk theory does not take into account the reflectance from the front surface of the sample, nor the diffuse light incident upon this surface from the inside. Part of the light approaching the surface from a particle within the sample is scattered back into the sample. If the ray approaches the surface at an angle greater than the critical angle it will be totally internally reflected. Light loss can be substantial. For example,  $\sim 50\%$  of the light attempting to emerge from a boundary of refractive index ratio of 1.33 (air and water) is back-reflected. A correction



for reflectance has been developed by Saunderson:

$$R_c = \frac{R_m - k_e}{(1 - k_e)(1 - k_i) + k_i R_m - k_e k_i}$$

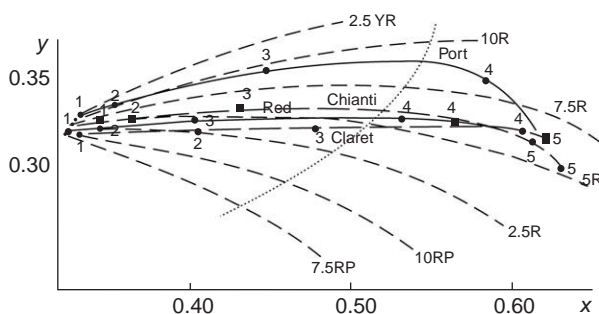
where  $R_c$  is the corrected reflectance,  $R_m$  the measured reflectance,  $k_e$  the fraction of incident light reflected externally, and  $k_i$  the fraction lost internally.

Methods are available for fine-tuning dye and pigment mixes by, for example, optimizing tristimulus values or using simplex methods. Such methods can be used to minimize metamerism and pigment costs. The many problems of dye and pigment formulation can be surmounted using specific equations for specific applications.

The second group of color measurement applications involves their use as standards or references where the original material might not be available for comparison. An example of the former includes standards for traffic light glasses (*Colors of Light Signals*, CIE publication 02.2). Measurements can be used as references during color change. Examples are weathering of paint and the comparison of the effect of treatment variables during food production trials. As for all color measurements, sample preparation must be appropriate to the form of the sample, and testing methodology must be rigorously consistent and shown to be reproducible.

The third application is directed toward the understanding of visual properties. For example, the effect of chemical changes on a sample color can lead to methods of defining the state of its chemistry. Tristimulus measurements have been used to monitor separately the progress of anthocyanin loss and browning increase occurring simultaneously in wines. Controlled oxidation and reduction reactions in tuna meat have been used to monitor the state of the pigment in the flesh, and to indicate damage inflicted during production.

An example of understanding in a psychophysical situation is provided by expert wine taster judgments. Tasters commented on the edge or rim color of four types of red wine. This property becomes apparent when the wine is swirled in the glass. Transreflectance spectrophotometric measurements were made on each wine using cells from 0.25 to 10 mm deep. Tristimulus properties were calculated from the spectra for the 2° observer and illuminant C. Using these measurements alone, it is difficult to visualize either the subtleties of color change with depth, or the color differences between the wines. However, Figure 5 shows that this situation can be remedied by superimposing Munsell constant hue and constant chroma lines onto the chromaticity locus of the wines. The change of color with cell depth



**Figure 5** Change of chromaticity with cell depth for four wines. From transreflectance measurements using a near 0°/diffuse instrument geometry, illuminant C, 2° observer. Cell 1, depth 0.25 mm; cell 2, depth 1 mm; cell 3, depth 2 mm; cell 4, depth 5 mm; cell 5, depth 10 mm. Lines of Munsell constant Hue and Chroma have been superimposed. (Reproduced with permission from Hutchings JB (1999) *Food Color and Appearance*, 2nd edn. Gaithersburg, MD: Aspen Publishers.)

is indicated by the number of constant hue lines crossed as cell depth is increased. The claret and the red wine change more than the port. Thin layers of claret and red wine are purpler than the chianti and port, because the depth-change chromaticity loci cross into the red–purple (RP) Munsell zone. Similarly, port is browner, because its locus enters the yellow–red (YR) zone. The chianti, noted as having slight browning on the rim, is not as purple in thin layer as the red wine or claret. This agrees with the observed order of purpleness of the thin layer when the wines are swirled in the glass – claret (most), red, chianti, port. Hence, tristimulus measurements can provide leads to psychophysical perception mechanisms in defined situations.

This approach, in which we are attempting to understand the color vocabulary of the expert, is being extended to the understanding of the vocabulary of total appearance. For example, consumer concepts such as ripeness of bananas and apparent flavor strength of orange juice are being approached using color calibrated digital analysis. Thus, the subject of soft metrology has been launched.

**See also:** **Optical Spectroscopy:** Wavelength Selection Devices; Detection Devices. **Spectrophotometry:** Instrumentation; Derivative Techniques; Biochemical Applications; Pharmaceutical Applications.

## Further Reading

- Berns RS (2000) *Billmeyer and Saltzman's Principles of Colour Technology*, 3rd edn. New York: Wiley.
- Chamberlin GJ and Chamberlin DG (1980) *Colour: its Measurement, Computation and Application*. London: William Clowes (Beccles).
- Fletcher R and Voke J (1985) *Defective Colour Vision*. Bristol: Adam Hilger.

- Hunter RS and Harold RW (1987) *The Measurement of Appearance*. New York: Wiley.
- Hutchings JB, Luo MR, and Ji W (2002) Calibrated colour imaging analysis of food. In: MacDougall DB (ed.) *Colour in Food, Improving Quality*, pp. 352–366. Cambridge: Woodhead Publishing.
- Judd DB and Wyszecki G (1975) *Color in Business, Science and Industry*, 3rd edn. New York: Wiley.
- McDonald R (ed.) (1987) *Colour Physics for Industry*. Bradford: Society of Dyers and Colorists.
- Minnaert M (1954) *The Nature of Light and Colour in the Open Air*. New York: Dover Publications.
- Nassau K (1983) *The Physics and Chemistry of Color*. New York: Wiley.
- Tilley R (2000) *Colour and the Optical Properties of Materials*. Chichester: Wiley.

## COMPLEXOMETRIC TITRATIONS

See **TITRIMETRY: Overview; Potentiometric; Photometric**

## COMPUTER MODELING

**J Mocak**, Slovak University of Technology, Bratislava, Slovakia

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by D B Hibbert, pp. 810–816, © 1995, Elsevier Ltd.

### Introduction

Modeling is a common human activity. From an early age people have tried to represent objects by a model. A model is a system of postulates, data, and inferences presented as a simplified description of an entity or state of a process. The very act of making or understanding a model allows us to appreciate what is involved in the thing that is being modeled. Mathematical models also predict the future behavior of the system being considered. Because models are necessarily simplifications, ignoring certain details, we cannot be sure that we have included all the important factors involved. Running a simulation through a model and comparing the results with what might be expected to happen can tell us whether our model fairly represents the modeled object. Using the results of our simulation, we can go back and alter our model as necessary. There exist several definitions of what is modeling, for example: (1) modeling is to design, develop, explore, and evaluate models of real or imaginary situations, (2) modeling is a representation of some part or aspect of an object or system, which can be based in reality or imagination, (3) modeling is simulating in a simplified and

numerical way a real-world situation, and (4) modeling is the search for an analytical function or a procedure (model) that will give a specified  $n$ -variable output for any  $m$ -variable input. The third and fourth definitions concern mathematical modeling, which substitutes a real experiment by mathematical simulation.

Modeling on a computer is similar to other kinds of modeling except that the approach is more abstract and utilizes mathematics and logic. Computer modeling is the implementation of algorithms coded in computer programs to generate new information about the studied aspect of reality. Applications in analytical science span a wide range of approaches and methods that depend on the system studied. Some forms of modeling are covered explicitly in this Encyclopedia, for example, expert systems, or implicitly by the end use as in optimization.

### Basic Principles

In terms of the definition given above there are three steps in creating a computer model;

1. *To understand the real system in terms of the properties to be modeled.* Thus, one needs to have some background understanding of the modeled objects or processes. Their main features and their interrelation are evaluated. First, it consists of the selection of the influencing factors (input variables) and the outcome (output variables). Then, a mathematical model is constructed at an appropriate level. For example, atoms and molecules can be modeled

- Hunter RS and Harold RW (1987) *The Measurement of Appearance*. New York: Wiley.
- Hutchings JB, Luo MR, and Ji W (2002) Calibrated colour imaging analysis of food. In: MacDougall DB (ed.) *Colour in Food, Improving Quality*, pp. 352–366. Cambridge: Woodhead Publishing.
- Judd DB and Wyszecki G (1975) *Color in Business, Science and Industry*, 3rd edn. New York: Wiley.
- McDonald R (ed.) (1987) *Colour Physics for Industry*. Bradford: Society of Dyers and Colorists.
- Minnaert M (1954) *The Nature of Light and Colour in the Open Air*. New York: Dover Publications.
- Nassau K (1983) *The Physics and Chemistry of Color*. New York: Wiley.
- Tilley R (2000) *Colour and the Optical Properties of Materials*. Chichester: Wiley.

## COMPLEXOMETRIC TITRATIONS

See **TITRIMETRY: Overview; Potentiometric; Photometric**

## COMPUTER MODELING

**J Mocak**, Slovak University of Technology, Bratislava, Slovakia

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by D B Hibbert, pp. 810–816, © 1995, Elsevier Ltd.

### Introduction

Modeling is a common human activity. From an early age people have tried to represent objects by a model. A model is a system of postulates, data, and inferences presented as a simplified description of an entity or state of a process. The very act of making or understanding a model allows us to appreciate what is involved in the thing that is being modeled. Mathematical models also predict the future behavior of the system being considered. Because models are necessarily simplifications, ignoring certain details, we cannot be sure that we have included all the important factors involved. Running a simulation through a model and comparing the results with what might be expected to happen can tell us whether our model fairly represents the modeled object. Using the results of our simulation, we can go back and alter our model as necessary. There exist several definitions of what is modeling, for example: (1) modeling is to design, develop, explore, and evaluate models of real or imaginary situations, (2) modeling is a representation of some part or aspect of an object or system, which can be based in reality or imagination, (3) modeling is simulating in a simplified and

numerical way a real-world situation, and (4) modeling is the search for an analytical function or a procedure (model) that will give a specified  $n$ -variable output for any  $m$ -variable input. The third and fourth definitions concern mathematical modeling, which substitutes a real experiment by mathematical simulation.

Modeling on a computer is similar to other kinds of modeling except that the approach is more abstract and utilizes mathematics and logic. Computer modeling is the implementation of algorithms coded in computer programs to generate new information about the studied aspect of reality. Applications in analytical science span a wide range of approaches and methods that depend on the system studied. Some forms of modeling are covered explicitly in this Encyclopedia, for example, expert systems, or implicitly by the end use as in optimization.

### Basic Principles

In terms of the definition given above there are three steps in creating a computer model;

1. *To understand the real system in terms of the properties to be modeled.* Thus, one needs to have some background understanding of the modeled objects or processes. Their main features and their interrelation are evaluated. First, it consists of the selection of the influencing factors (input variables) and the outcome (output variables). Then, a mathematical model is constructed at an appropriate level. For example, atoms and molecules can be modeled

by quantum mechanical or statistical mechanical equations. Equations such as Fick's laws of diffusion, the Navier–Stokes equation, rate equations, or equilibrium equations may be used to derive information about a system at a higher level. The model can only be as good as our understanding. The majority of failures of models stem from attempts to apply the resulting computer program to situations outside the range of applicability of the original model. While a theory has only the alternatives of being right or wrong, a model has a third possibility, namely that it may be right, but irrelevant.

2. *To develop an algorithm to describe the model.* The algorithm is defined as a detailed sequence of actions to accomplish some task. It is the procedure of implementing the model. This may include the solution of equations governing the system, compilation of rules in an expert system, and also the methods of input of data and output of results. It may be straightforward if the equation exists in a form from which the desired parameter can be explicitly calculated, or it may require the solution of nonlinear equations for a given variable, or the solution of differential equations (for example, Laplace's equation or rate equations). At the most basic level the model may simulate reality in a more direct way. For example, diffusion processes may be modeled by random walks. Here, particles move in a direction determined by the toss of a coin and in doing so implicitly solve Laplace's equation.

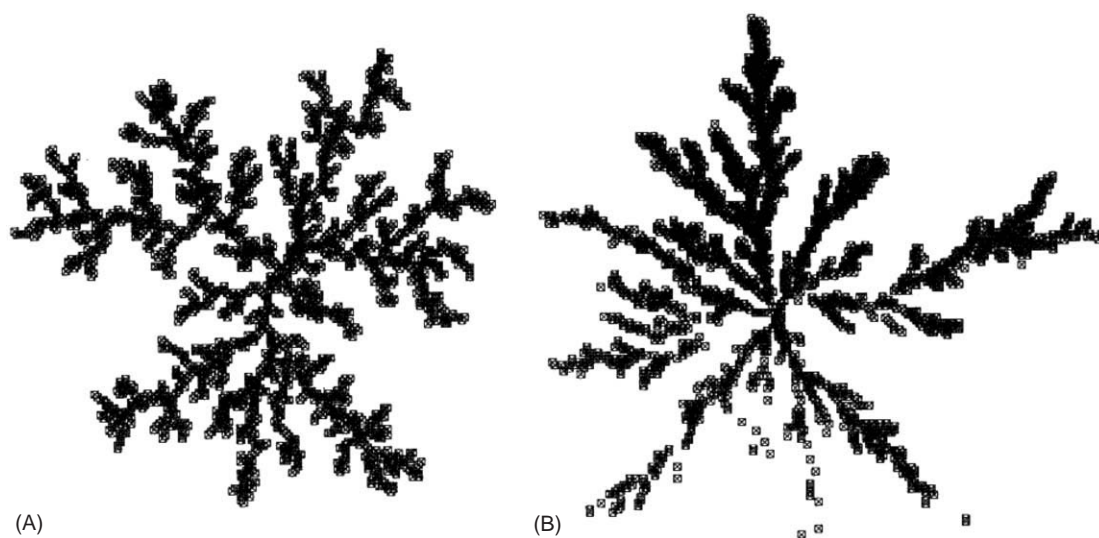
3. *To write a computer program to implement the algorithm.* In this step, the mathematical model is

converted into a computer model. This is usually a straightforward process but does require knowledge about the software used. It may be trivial, for example, to write a BASIC program to model the course of an acid–base titration, or it may result in programs that require hours of supercomputer time as in a molecular–dynamic simulation of liquid water. The required sophistication of the computer program depends on the complexity of the model and the constraints of the hardware and software. The code to perform the above-mentioned random walks calculation is simple and may be written in any high-level language. **Figure 1A** shows a computer model of a two-dimensional (2D) fractal created by allowing random walks (pixels on a computer screen) to move about until they hit and stick to the growth. A real growth of copper, electrodeposited from copper sulfate supported in filter paper, is shown in **Figure 1B**.

Computer modeling in analytical science takes a myriad of forms. **Table 1** brings together examples of the above principles for a range of models.

## Hardware and Software Considerations

When deciding to model a particular analytical problem thought must be given to the computer program and the platform on which it will run. While it is true that *ab initio* quantum mechanical packages exist to run on microcomputers, for serious work a



**Figure 1** Comparison between a simple random walk model of particle deposition and an electrochemically deposited copper fractal. (A) 2000 random walks on a square grid, (B) the digitized image ( $256 \times 256$  pixels) of copper electrodeposited from  $0.75 \text{ mol l}^{-1}$  copper sulfate and  $1 \text{ mol l}^{-1}$  sulfuric acid in an 11 cm Whatman 541 filter paper at 5 V. Both fractals are displayed using Lotus for Windows.



**Table 1** Computer modeling

<i>System</i>	<i>Model</i>	<i>Properties modeled</i>
Molecule	Quantum mechanical: <i>ab initio</i> , or empirical Force field model of interactions	Structure, electronic energy levels, thermodynamic data, analytical data, e.g., spectra, diffraction patterns, etc.
Clusters of molecules	Monte Carlo simulation Molecular dynamics Fluid mechanical equations	Structures of assemblies, dynamic evolution of structures, analytic data
Chemical reaction	Solution of rate equations Optimization for rate parameters	Concentration and related data (e.g., electrochemical current, temperature) with time
Industrial process	Solution of rate, diffusional, and hydrodynamic equations in space and time	Yields of product Optimal conditions to operate process (e.g., time, concentrations) Real-time process control
Analytical method	Expert system to model rules of configuration of method Optimization of method variables	Suggested conditions for given problem, e.g., mobile phase, column, and detector for ion chromatography
Analytical data	Multivariate analysis	Classification, calibration, pattern recognition
Instrument	Solution of equations governing operation of instrument given known inputs	Control of instrument Simulation of output of instrument Optimum design of new instruments

supercomputer running vectorized or parallel code is all but mandatory. The choice of hardware revolves about two factors, the amount of processing to be performed, which has a bearing on the speed of the machine, and the memory requirements of the task. The available computers range from microcomputers (personal computers, PCs) through workstations to minicomputers, mainframes, and supercomputers. Most computer modeling is performed on workstation computers. These have powerful processors combined with the flexibility and user-friendliness of microcomputer style user interfaces. The fast, high-quality graphics of workstations or the better microcomputers are of great use to the analytical scientist for presentation of results. Where speed and memory may be sacrificed for cheaper computing, microcomputers are useful, for example, in running instruments and in teaching. Desktop modeling of complex systems is changing the way we perform analytical science.

The advent of parallel computing also must be taken into account. It comprehends either using more than one computer, or a computer with more than one processor. For example, Dual Pentiums or Quad Pentiums may be used as a shared memory multiprocessor. Parallel and distributed computer systems support large-scale high-performance computing facilities.

If commercial software is available for a given simulation generally it should be purchased in preference to programs written in-house. At least the display of data, graphs, etc., and simple statistical manipulation should be performed by standard

routines. They can be found in large mathematical and statistical libraries, e.g., in the IMSL, NAG, or Netlib library, or the Numerical Recipes series of texts. The reasons for this advice are twofold. First, programmers employed by commercial houses are better than scientists at writing error-free, efficient, useable code. Second, the cost of writing commercial software is shared amongst all the purchasers.

Software written in a high-level language should be as portable as possible. Even if the code is intended for a single user, upgrading of a computer may render a program written in a dialect of a language specific to the earlier machine quite useless. Programming language is a formal language in which computer programs are written. The definition of a particular language consists of both syntax (how the various symbols of the language may be combined) and semantics (the meaning of the language constructs). Languages are classified as low level if they are close to machine code (the first and second generation) and high level if each language statement corresponds to many machine code instructions (starting from the third generation). There are five generations of computer languages: (1) First-generation languages (or 1GLs) are machine languages. (2) Second-generation languages (2GLs) generally consist of assembly languages. (3) Third-generation languages (3GLs) are high-level programming languages such as FORTRAN, BASIC, C and its successors, and Java. (4) Fourth-generation languages (4GLs) are closer to a human language since they use statements similar to it; they are commonly used in database programming, spreadsheets and scripts.

(5) Fifth-generation languages (5GLs) are used for artificial intelligence (AI) and contain visual tools to help develop a program. Good examples of the 5GLs are Visual Basic, Visual C++, and Delphi.

The 3GL of choice is C++, which is an object-oriented language, widely accepted, portable, and equipped by libraries of routines to extend the power of the basic language. It enables efficient and fast programming. However, in the chemistry- and analytical chemistry-oriented literature, many practical programs have been written in BASIC, FORTRAN, or PASCAL. The fourth generation includes macro languages written within applications such as databases and spreadsheets. Spreadsheet modeling is nowadays very popular even though it is useful only for relatively small systems. Lotus and MS Excel are the most popular spreadsheets, which include facilities to produce graphs, perform 'what if' calculations, solve equations, and to use simple statistical functions. In teaching analytical chemistry the student has a much better understanding of, for example, a pH titration if the form of the titration curve is displayed and the effects of changing concentrations and  $pK_a$ s may be viewed instantaneously. Spreadsheets are rather slow and cannot accommodate very large datasets, but within their limitations their use for chemical modeling is growing. Most spreadsheet applications are multidimensional, meaning that you can link one spreadsheet to another. A multidimensional spreadsheet, for example, is like a stack of spreadsheets all connected by formulas. A change made in one spreadsheet automatically affects other spreadsheets. Fifth-generation languages include declarative, list-based languages like LISP and PROLOG. The code of procedural 3GL informs the computer what to do with the data, e.g., 'loop round 10 times adding a number to a sum'. On the contrary, a declarative 5GL expresses what is true about a problem and then invites the compiler to engineer a solution that is logically correct. For example, a sorted list has the property that each succeeding number is greater than its predecessor, so that it can be used to implement a sort routine without DO loops in very few lines of code. Fifth-generation languages are at the heart of expert system shells because of their ability to handle logical data and the integration of code and data. Computers handle numbers and are optimized to manipulate numerical information. Among chemists and analytical chemists, the most popular 5GLs are MATLAB, MATHEMATICA, and MAPLE, as well as SAS (Statistical Analysis System) as a statistical language. The names of these languages are identical to the names of the software packages wherein they are used. MATLAB is both an intuitive language and a technical computing

environment. It provides core math and advanced graphic tools for data, analysis, visualization, and development of applications. MATHEMATICA is a list-based language, which combines techniques from procedural languages C/C++, and nontraditional languages. It is intelligent and ideal for AI applications enabling symbolic mathematical expressions much better than any human can, including integrals and derivatives. For example, by command  $D[\text{Log}[a-x]/4a^2, x]$  you can obtain the derivative of the first term in the square brackets according to  $x$  (the second term) with  $-a^2/4(a-x)$  at the output. If necessary, it is possible to simplify the result and/or make an evaluation for a given constant  $a$ . Lists are used mainly to represent vectors, matrices, tensors, sets, ranges of integrals and plots, as well as groups of arguments in functions. A common way to generate a list is with the Table command. An electro-analytical application of MATHEMATICA is described later.

## Model Architecture

The essential elements of any model are:

1. *Establishment of boundary conditions.* No model operates over all space and all time and so some limits must be specified. The integration of differential equations, whether explicitly analytical or numerical, requires limits to the integration (integration constants). The calculation of diffusion to an electrode requires the solution of Laplace's equation (Fick's laws). For this, the concentration of electroactive species at the electrode must be known in addition to how far into the solution the integration is to be made. In calculations involving time, conditions at zero time (and sometimes infinite time) are required. Optimization models generally require initial guesses and some knowledge of one state of the system (e.g., at zero time).
2. *The acquisition of input data.* Input data may be read from a file, generated in the model, or read from an experiment via analog-to-digital converters. It is these input data that distinguish one run of the model from another.
3. *Calculation of model output.* At the heart of a model simulation is the calculation itself. Usually this code is in a loop over time or distance. Considerable attention must be paid to the fitness of the model, both in terms of its underlying assumptions and its coding as a computer program. It must be understood that a model is not reality. Where the model deviates from reality must be known, and the computer realization of the model must not allow nonphysical behavior. For example, in fitting kinetic data rate constants cannot



be negative, even if the mathematical functions could be improved by allowing them to be so.

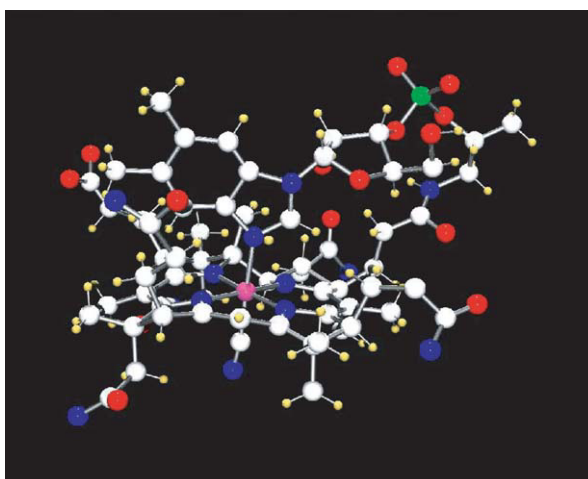
4. *Output of results.* The purpose of any model is to convey information about the real system. The results may be numbers output to file, printer, or screen; a visual presentation of the system such as the shape of a molecule, a graph of a property with time or space; or a voltage output via a digital-to-analog converter to control an instrument. The output of an expert system is advice on the problem.

## Examples of Modeling Techniques

### Molecular Structures and Processes

Molecular dynamics attempts to solve the dynamically evolving ensemble of molecules given the interactions between molecules. The form of the forces between molecules or atoms, the number of interactions (i.e., two- or three-body interactions), and the number of molecules that can be tackled by the program determine the success of the model. Molecular dynamics simulations can predict the internal energy, heat capacity, viscosity, and infrared spectrum of the studied compound and form an integral part in the determination and refinement of structures from X-ray crystallography or nuclear magnetic resonance (NMR) experiments.

Recent advances in commercially available software have concentrated on taking molecular modeling routines and embedding them in high-quality visualization packages with easy interfaces to libraries of molecules and programs that can predict the properties of the molecules created. The use of such packages in materials and drug design is relieving chemists of much wasted time in synthesis and analysis of compounds that are ultimately not suitable for the required task. Examples of modeling properties using computational instruments, that is, software that can predict the output of an analytical instrument applied to a compound, are X-ray diffraction (XRD), neutron and electron diffraction patterns, and high-resolution transmission electron microscope images. The chemist may make adjustments to a structure displayed on a computer screen and see the effect on, for example, the XRD of that compound in an adjacent window. Various three-dimensional (3D) graphics tools can be used for visualization of molecular structure. **Figure 2** shows a computer-generated 3D view of cyanocobalamin, vitamin B12. Databases of chemicals that contain 2D information of the structure of molecules in the form of connection tables may be used to generate 3D conformers. These structures may then be searched for specific properties such as ‘atoms of nitrogen and



**Figure 2** ORTEP 3 view of cyanocobalamin (vitamin B12) as created by POV-Ray 3D graphics tool. Crystal structure data were obtained from Cambridge Structural Database (CSD), code COVDEW01. Colors of atoms: carbon – white, hydrogen – yellow, nitrogen – blue, oxygen – red, phosphorus – green, cobalt – pink. (Courtesy of Moncol J and Koman M, Slovak University of Technology, Bratislava, Slovakia, with permission.)

oxygen separated by less than 0.5 nm each attached to an aromatic ring’. Candidate molecules are displayed or stored in a file for evaluation. In the pharmaceutical industry, the target molecular fragments are known as pharmacophores. If the interest is the interaction of a ligand at a receptor site and X-ray or NMR data are available for the site, then it is possible to model the interaction both in terms of the 3D shape and the forces between putative ligands and the site. Finally, the structure of a compound with the specified pharmacophore atoms at defined distances and angles is found and is used as the basis of a new drug.

### Dynamic Systems

Computer modeling of the kinetics of a reaction by solving rate equations is useful in the determination of mechanism and the estimation of rate parameters. Such analysis of kinetic data represents a higher-level approach to the problem of modeling than the molecular modeling discussed above. Here, we assume knowledge of the bulk properties of the system (the kinetic equations) and proceed to model the system comparing predictions to experimental measurements.

The problem is therefore to predict the concentrations of reactants and products with time given rate equations of the form:

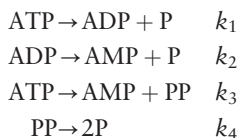
$$\frac{dc_i}{dt} = \sum_j \kappa_j \prod_l c_l^{n_{lj}}$$

where  $c_i$  is the concentration of species  $i$ ,  $\kappa_j$  is the rate coefficient of the  $j$ th reaction ( $\kappa_j$  is the product of the

rate constant  $k_j$  and the number of molecules consumed or formed in the given reaction, e.g.,  $-1$ ,  $0$ ,  $+1$ , etc.) and  $n_l$  is the order with respect to species  $l$ . If no analytical solution exists for the set of equations, which is the case for all but the simplest systems, then they must be integrated numerically. When concentration data are available the solution of the rate equations is done as part of an optimization for the rate coefficients. At each guess of the  $\kappa_j$  values, the set of equations are integrated and the calculated concentrations are compared with measured values. If temperature also varies, as in a temperature-programmed desorption experiment, the rate constant  $k_j$  must be expressed in terms of the Arrhenius equation ( $k = A \exp(-E_a/RT)$ ) and the constants  $A$  and  $E_a$  determined.

In modeling kinetic data, it is important not to include more than the minimal set of reactions (rate constants). A sensitivity analysis that determines  $d \ln c_i / d \ln k_j$  for all  $i$  and  $j$  will show if a particular compound is superfluous (too small  $c_i$ ) or if a reaction is not required (too small  $k_j$ ). Kinetic modeling also highlights the fact that a model and the reality are not mirror images. A correct model should reproduce the desired facet of reality. A model that appears to give numbers that are in accord with reality is not necessarily correct. A simple example may be used to illustrate these points.

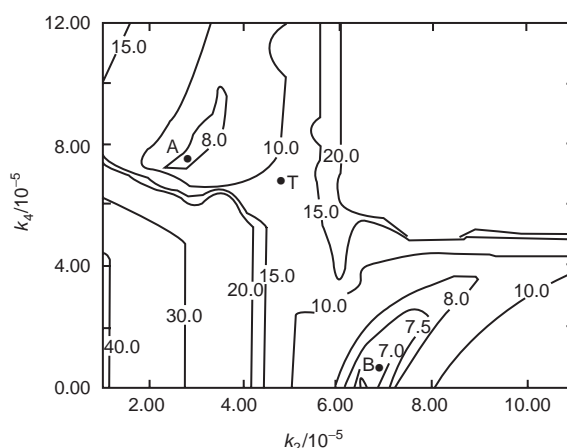
Adenosine 5'-triphosphate (ATP) hydrolyzes via the diphosphate (rate constant  $k_1$ ) and monophosphate ( $k_2$ ), or directly to the monophosphate with the pyrophosphate formation ( $k_3$ ). One mole of pyrophosphate then hydrolyzes to 2 mol of phosphate ( $k_4$ ):



Modeling of the two paths to fit phosphate concentration data by solving the rate equations in an optimization of the rate constants reveals that depending on the initial guess, two different, but apparently valid, results may be achieved. Figure 3 is the response surface of the optimization plotted as a function of the rate constants  $k_2$  and  $k_4$ . The response  $F(k)$  is defined as

$$F(k) = \left[ \sum_t (c_{t,\text{meas}} - c_{t,\text{calc}})^2 \right]^{1/2}$$

where  $k$  is the vector of rate constants,  $c_t$  is the concentration of phosphate at time  $t$ , and the subscripts 'meas' and 'calc' refer to the measured and calculated values, respectively.



**Figure 3** Response surface for the modeling of the hydrolysis of ATP as a function of two rate constants showing two minima A and B. T is the location of the true minimum. (Reproduced with permission from Hibbert DB and Sandall JPB (1989) Mechanism of the hydrolysis of adenosine 5'-triphosphate: A statistical evaluation of kinetic data. *Journal of Chemometrics* 3: 569–577; © John Wiley & Sons Ltd.)

Small values of the function (valleys) indicate good agreement between experiment and the model. The points labeled A and B are about equally good fits, but predict quite different mechanisms. To distinguish between the mechanisms pyrophosphate must also be measured. Visualization of data, in this case of the response surface in Figure 3, is a great aid in understanding the underlying processes when modeling.

Electrochemical systems lend themselves to modeling. In cyclic voltammetry, the peaks arise through a complex interplay of electrochemistry, solution chemistry, and diffusion, which influence the shapes of the voltammetric signals. Using microcomputers with sufficient power, the numerical solution of the relevant dynamic equations not only can confirm the mechanism, but good estimates of rate and diffusion coefficients can be obtained in real time. Several general simulation programs (e.g., DigiSim, ELSIM), libraries, and expert systems have been developed that bring the power of computer experiments to every electrochemist. One of the main tasks of computational electrochemistry is numerical modeling of electrochemical systems.

Many internal mathematical functions, available in mathematical software packages, are based on infinite series computations, optimized by a team of programmers, so that they are highly accurate for all practically used argument values. This has enabled to find, among numerous built-in functions of MATHEMATICA, those which can advantageously be used directly, or with adaptation, for calculations in linear

sweep voltammetry, thus creating an alternative way of computing the signal of this method. For example, the current–potential curve for a reversible electrode process can be calculated by means of the polylogarithm function  $\text{PolyLog}[n, z]$ , defined in MATHEMATICA as

$$\text{PolyLog}[n, z] = \sum_{k=1}^{\infty} z^k / k^n$$

Figure 4 shows the calculations of a single current output, a current–potential table (containing merely three output lines for the sake of space), as well as the corresponding plot, which are elegantly performable with only a few lines of instructions. The tables and plots for catalytic reaction with reversible charge transfer can be modeled with similar ease by

implementing the three-parameter transcendent ‘Lerch function’.

### Classification of Analytical Data

In linear multivariate data analysis (MDA) of a set of analytical data, the most common model is that linear relationships exist within the data that can be revealed, for example, in the principal components analysis (PCA) or linear discriminant analysis (LDA). Analytical data may represent the analyte concentrations of the multicomponent mixture, a spectrum of any kind, the results of clinical laboratory tests, the status of a technological process at a given time, etc. The classification of the investigated objects by the MDA techniques is an important task, which provides interpretation of complex problems, often

```

estar := 15                                     (* Part A *)

istar := Re [-PolyLog [-1/2, -Exp [estar]]]

SetAccuracy [istar, 16]

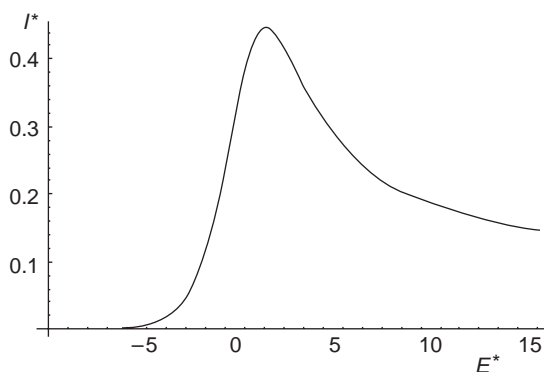
0.146 513 229 207 207

SetAccuracy [TableForm [Table[istar, {estar, -10, 10, 10}], 16] (* Part B *)

0.000 045 397 015 013
0.380 104 812 609 684
0.180 928 068 599 584

Plot [istar, {estar, -10, 15}, PlotPoints -> 251,          (* Part C *)
AxesLabel -> {"E*", "I*"}, AxesOrigin -> {-10.5, -0.01}]

```



– Graphics –

(\* Time of calculation: 15 s on PIII 500 MHz PC \*)

**Figure 4** Dimensionless current derived for a simple reversible charge transfer process using the polylogarithm function in MATHEMATICA. Part A: Calculation of the dimensionless current  $I^*$  at dimensionless potential  $E^* = 15$ . Part B: Calculation of the table of  $I^*$  values at  $E^*$  values  $-10, 0$ , and  $10$ . Part C: Calculated  $I^* - E^*$  graph over the potential range of  $-10$  to  $+15$  dimensionless units. Note: The command in line 2, part A, has to precede commands in subsequent parts. (Reprinted with permission from Mocak J and Bond AM (2004) Use of MATHEMATICA software for theoretical analysis of linear sweep voltammograms. *Journal of Electroanalytical Chemistry* 561: 191–202; © Elsevier.)

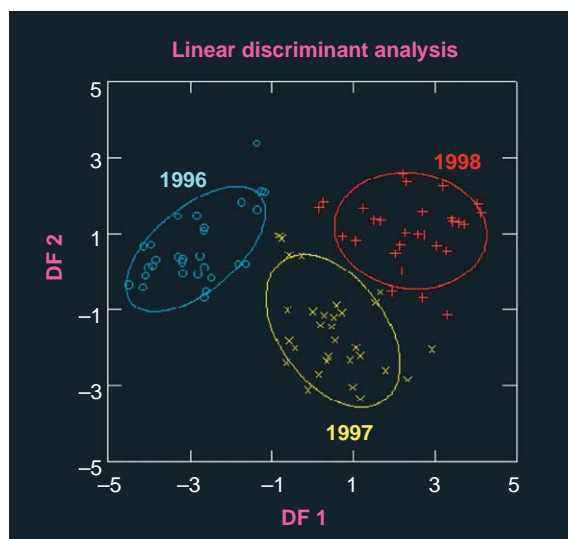
existing beyond the borders of Analytical Chemistry. However, it should be done by analytical chemists, not only due to its feedback to analytical measurements, but also to enhance their role from service to knowledge. For example, the models created by the MDA methods enable authentication of wines, their classification, and characterization according to the variety, vintage, or the producer. Moreover, based on sensory data from experts, wine samples may be classified into two or three classes according to sensorial quality. By analysis of the variables maximally influencing the discrimination of classes, which can be found that the chemical compounds have the largest impact upon the sensorial quality, i.e., the total taste, color, transparency, bouquet, and flavor. **Figure 5** shows the discrimination of white varietal wines (Gruener Veltliner, Mueller Thurgau, and Welsch Riesling) obtained from five producers during three consecutive vintages. Based on the gas chromatography–mass spectrometry analysis of volatile species influencing the wine aroma, the LDA enabled a very reasonable classification of wines by vintages. Nineteen variables (concentrations of volatile compounds) were finally used, among them six alcohols, five esters, seven terpenes, and one aliphatic acid. The variables

most influencing classification according to variety, vintage, producer, and sensorial quality were found. For solving different types of practical problems, the MDA methods use the same algorithms, processing the matrix of analytical data with objects in the rows and variables in the columns.

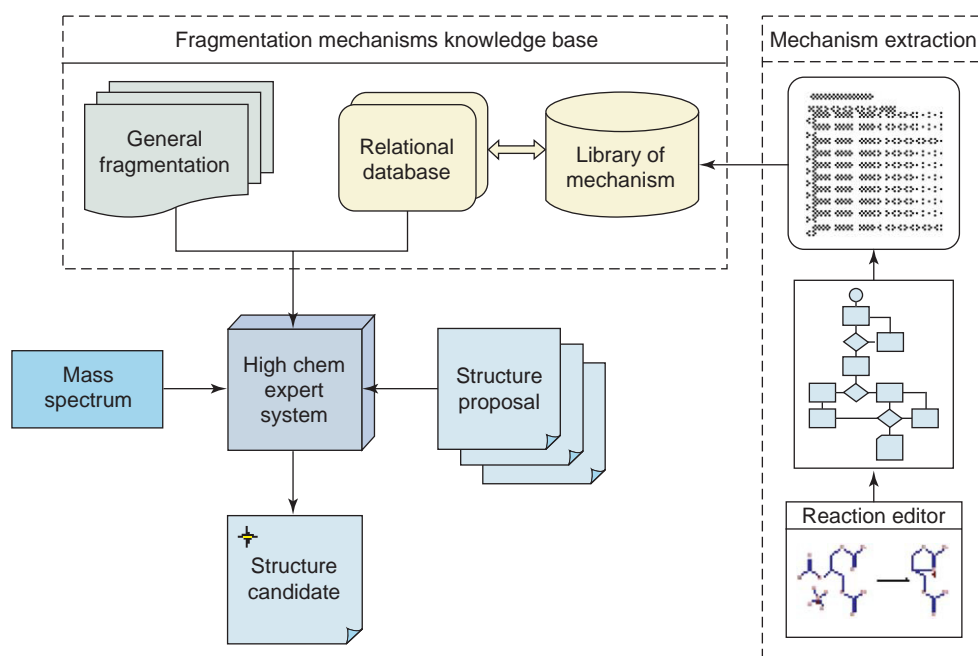
### Instruments

Many modern analytical instruments may be classed as virtual. They have no knobs or switches, but are run entirely from a microcomputer, the screen of which serves as the control panel of the instrument. The computer thus models the instrument and in doing so controls the real thing. In the virtual instrument approach the boundary between analog and digital information is pushed far towards the physical processes being measured. Lying at this boundary, analog-to-digital and digital-to-analog converters convert potentials to binary numbers, and vice versa, binary numbers from the computer are converted to potentials or currents that control the instrument's operation. The key advantage is that the specialized and costly 'analog processors' can be largely replaced by a powerful microcomputer with the specialized tasks being performed digitally under software control. It is now possible to build, rebuild, and modify virtual instruments in software without having to change the more expensive and less flexible hardware. The virtual approach also helps with complex analytical instruments in which many functions must be controlled simultaneously with a high degree of precision. Automation of repetitious tasks, as may be found in a clinical or forensic laboratory dealing with hundreds of samples, is made simpler by computer control. Use of a robot requires a model of the tasks to be developed in the controlling computer.

The process of modeling may be taken one step further in the development of new instrumentation. Modeling the output of a would-be instrument may save time and money in a trial and error attempt to refine parameters of its operation. An extension of the virtual instrument concept is the addition of artificial intelligence to the choice of parameters. **Figure 6** shows the flow diagram of the system for interpreting mass spectra based on automated prediction of fragmentation and rearrangement pathways (which is based on library mechanisms) combined with a module that uses general fragmentation rules. An intelligent fragmentation mechanism knowledge base, incorporated in a computer system, is used to predict unimolecular decomposition reactions. The knowledge base is encapsulated in a relational database that, in turn, is encapsulated in software with a graphical user interface. The software



**Figure 5** Classification of 88 white wine samples of three varieties from five producers according to *vintage* by linear discriminant analysis – plot in the coordinates of two main discriminant functions (DF2 versus DF1) composed of 19 original variables (concentrations of volatile, aroma creating compounds). Explanation of symbols: ○ denote the 1996 samples, × the 1997 samples, + the 1998 samples. Probability ellipses express the 95% probability level. (Reproduced with permission from Petka J, Mocak J, Farkas P, Balla B, and Kovac M (2001) Classification of Slovak varietal white wines by volatile compounds. *Journal of the Science of Food and Agriculture* 81: 1533–1539; © John Wiley & Sons Ltd.)



**Figure 6** The flow diagram of the system of interpreting mass spectra based on a combination of a fragmentation mechanism database and general fragmentation rules. A mechanistic rationale that accounts for the fragmentation is algorithmically extracted from the reaction drawing. (Courtesy of R. Mistrik, HighChem, Ltd., Bratislava, Slovakia, with permission.)

contains an expert system that automatically extracts the decomposition mechanism for each given fragmentation reaction and determines the compound class range that the mechanism can be applied to. This approach is highly selective, which assures that skeletal and charge-remote rearrangements, ring closures and expansions, as well as compound specific mechanisms are applied to an appropriate structure. It allows the application of database mechanisms to a user-supplied structure and automatically proposes fragmentation reactions for the given compound. Each proposed fragmentation step is connected to its template mechanism, which allows the source and original database entry used for the prediction to be reviewed.

The synergy of two applied techniques, based on different principles, provides a powerful method of interpretation of mass spectra even when complicated rearrangements occur in the overall dissociation process. This approach considerably increases the success rate when matching spectral peaks with generated fragments.

Contemporary instruments are increasingly equipped with software systems that combine computer control of the measurements with elements of the simulation and data analysis. We are not far from seeing instruments that will carry out scientific investigations, not just measurements.

**See also:** **Chemometrics and Statistics:** Multivariate Classification Techniques; Multivariate Calibration Techniques; Expert Systems. **Voltammetry:** Linear Sweep and Cyclic.

## Further Reading

- Bieniasz L (2002) Towards computational electrochemistry – a kineticist's perspective. In: Conway BE and White RE (eds.) *Modern Aspects of Electrochemistry*, No. 35, ch. 3. New York: Kluwer Academic/Plenum.
- Brandt J and Ugi IK (eds.) (1989) *Computer Applications in Chemical Research and Education*. Heidelberg: A. Hüthig Verlag.
- Britz D (1988) *Digital Simulation in Electrochemistry*, 2nd edn. Berlin: Springer.
- Buydens LMC and Schoenmakers PJ (eds.) (1993) *Intelligent Software for Chemical Analysis*. Amsterdam: Elsevier.
- Cohen NC (1996) *Molecular Modeling in Drug Design*. New York: Academic Press.
- Coveney P and Highfield R (1995) *Frontiers of Complexity, The Search for Order in a Chaotic World*. New York: Fawcett Columbine.
- Denning PJ and Metcalfe RM (eds.) (1997) *Beyond Calculation, The Next Fifty Years of Computing*. New York: Copernicus-Springer.
- Doucet JP and Weber J (1996) *Computer-Aided Molecular Design: Theory and Applications*. London: Academic Press.



- Frenkel D and Smit B (1996) *Understanding Molecular Simulations: From Algorithms to Applications*, 2nd edn. San Diego: Academic Press.
- Gould H and Tobochnik J (1996) *An Introduction to Computer Simulation Methods (Parts 1 & 2)*, 2nd edn. Reading: Addison-Wesley.
- Grant GH and Richards WG (1995) *Computational Chemistry*. Oxford: Oxford University Press.
- Jensen F (1999) *Introduction to Computational Chemistry*. New York: Wiley.
- Karjalainen EJ (ed.) (1990) *Scientific Computing and Automation (Europe)*. Amsterdam: Elsevier.
- Massart LM, Vandeginste BMG, Buydens LMC, De Jong, Lewi PJ, and Smeyers-Verbeke J (1997) *Handbook of Chemometrics and Qualimetrics: Part A*. Amsterdam: Elsevier.
- Press WH, Flannery SA, Teukolsky WT, and Vetterling BP (2002) *Numerical Recipes in C++ . The Art of Scientific Computing*. Cambridge: Cambridge University Press (available also in C and Fortran).
- Rapaport DC (1995) *The Art of Molecular Dynamics Simulation*. Cambridge: Cambridge University Press.
- Vandeginste BMG, Massart LM, Buydens LMC, De Jong, Lewi PJ, and Smeyers-Verbeke J (1998) *Handbook of Chemometrics and Qualimetrics: Part B*. Amsterdam: Elsevier.
- Zupan Z (1989) *Algorithms for Chemists*. New York: Wiley.

## CONDUCTIMETRY AND OSCILLOMETRY

**T S Light**, Lexington, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 834–842, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

### Introduction

Conductimetry (the measurement of conductivity) is a physical chemical measurement that provides information about the total ionic content of aqueous solutions. Using conductimetric techniques, electrolytic properties of ions such as diffusion, transport, mobility, and migration have been extensively studied and reported. Dissociation and dielectric constants of compounds have similarly been the subject of these physical chemistry studies in both aqueous and nonaqueous solutions.

Analytical chemistry has found great utility in conductimetric measurements in spite of its apparent nonspecificity. Rapid quantitative accuracy of a few tenths of a percent may be quickly accomplished by direct conductimetric determination of binary electrolytic solutions such as aqueous acids, bases, or salts. A nearly linear increase in conductivity is observed for solutions containing as much as 20% of solute. The concentration of strong solutions, such as the salinity of seawater, may be determined from conductance measurements; traces of electrolyte impurities, such as the impurity in ultrapure water, may be reported at the  $\mu\text{g l}^{-1}$  level. Conductimetric titrations may increase the accuracy of endpoint detection and permit titrimetric analysis of weak electrolytes, such as boric acid, which is not feasible by potentiometric or colorimetric

detection methods. Continuous monitoring of the progress of electrolytic chemical reactions and of industrial chemical streams may be done with cost-effectiveness, accuracy, and reliability exceeding those of many other 'online' analytical methods.

### Fundamentals

The electrical resistance of a solution is denoted by  $R$ ; its units are ohms which are commonly indicated by the Greek letter omega,  $\Omega$ . The electrical conductance is the reciprocal of the resistance; it may be given the symbol  $L$  and has the SI units of siemens (S) or the older units of reciprocal ohms (mhos) and the symbol  $\Omega^{-1}$ . Their relation is

$$L = 1/R \quad [1]$$

The resistance of a material is a property of its composition and geometry. For liquid samples, which are the subject of this article, a conductivity cell may be visualized as shown in **Figure 1**, where the area  $A$  of each of the parallel electrodes is  $1\text{ cm}^2$  and they are separated from each other by the distance  $d$  of 1 cm. The specific resistance, also called the resistivity, is characteristic of the solution material and concentration. It is denoted by the symbol  $\rho$ , has the units of ohm-centimeters and is defined by

$$\rho = R(A/d) = R/\theta \quad [2]$$

where  $\theta$  is the cell constant, useful for characterization of conductivity cells, with units of  $\text{cm}^{-1}$ :

$$\theta = d/A \quad [3]$$

The specific conductance, also called the conductivity, of the solution is denoted by the symbol



- Frenkel D and Smit B (1996) *Understanding Molecular Simulations: From Algorithms to Applications*, 2nd edn. San Diego: Academic Press.
- Gould H and Tobochnik J (1996) *An Introduction to Computer Simulation Methods (Parts 1 & 2)*, 2nd edn. Reading: Addison-Wesley.
- Grant GH and Richards WG (1995) *Computational Chemistry*. Oxford: Oxford University Press.
- Jensen F (1999) *Introduction to Computational Chemistry*. New York: Wiley.
- Karjalainen EJ (ed.) (1990) *Scientific Computing and Automation (Europe)*. Amsterdam: Elsevier.
- Massart LM, Vandeginste BMG, Buydens LMC, De Jong, Lewi PJ, and Smeyers-Verbeke J (1997) *Handbook of Chemometrics and Qualimetrics: Part A*. Amsterdam: Elsevier.
- Press WH, Flannery SA, Teukolsky WT, and Vetterling BP (2002) *Numerical Recipes in C++ . The Art of Scientific Computing*. Cambridge: Cambridge University Press (available also in C and Fortran).
- Rapaport DC (1995) *The Art of Molecular Dynamics Simulation*. Cambridge: Cambridge University Press.
- Vandeginste BMG, Massart LM, Buydens LMC, De Jong, Lewi PJ, and Smeyers-Verbeke J (1998) *Handbook of Chemometrics and Qualimetrics: Part B*. Amsterdam: Elsevier.
- Zupan Z (1989) *Algorithms for Chemists*. New York: Wiley.

## CONDUCTIMETRY AND OSCILLOMETRY

**T S Light**, Lexington, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 834–842, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

### Introduction

Conductimetry (the measurement of conductivity) is a physical chemical measurement that provides information about the total ionic content of aqueous solutions. Using conductimetric techniques, electrolytic properties of ions such as diffusion, transport, mobility, and migration have been extensively studied and reported. Dissociation and dielectric constants of compounds have similarly been the subject of these physical chemistry studies in both aqueous and nonaqueous solutions.

Analytical chemistry has found great utility in conductimetric measurements in spite of its apparent nonspecificity. Rapid quantitative accuracy of a few tenths of a percent may be quickly accomplished by direct conductimetric determination of binary electrolytic solutions such as aqueous acids, bases, or salts. A nearly linear increase in conductivity is observed for solutions containing as much as 20% of solute. The concentration of strong solutions, such as the salinity of seawater, may be determined from conductance measurements; traces of electrolyte impurities, such as the impurity in ultrapure water, may be reported at the  $\mu\text{g l}^{-1}$  level. Conductimetric titrations may increase the accuracy of endpoint detection and permit titrimetric analysis of weak electrolytes, such as boric acid, which is not feasible by potentiometric or colorimetric

detection methods. Continuous monitoring of the progress of electrolytic chemical reactions and of industrial chemical streams may be done with cost-effectiveness, accuracy, and reliability exceeding those of many other 'online' analytical methods.

### Fundamentals

The electrical resistance of a solution is denoted by  $R$ ; its units are ohms which are commonly indicated by the Greek letter omega,  $\Omega$ . The electrical conductance is the reciprocal of the resistance; it may be given the symbol  $L$  and has the SI units of siemens (S) or the older units of reciprocal ohms (mhos) and the symbol  $\Omega^{-1}$ . Their relation is

$$L = 1/R \quad [1]$$

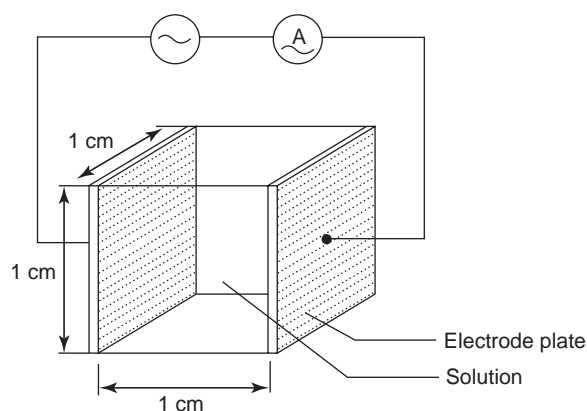
The resistance of a material is a property of its composition and geometry. For liquid samples, which are the subject of this article, a conductivity cell may be visualized as shown in **Figure 1**, where the area  $A$  of each of the parallel electrodes is  $1\text{ cm}^2$  and they are separated from each other by the distance  $d$  of 1 cm. The specific resistance, also called the resistivity, is characteristic of the solution material and concentration. It is denoted by the symbol  $\rho$ , has the units of ohm-centimeters and is defined by

$$\rho = R(A/d) = R/\theta \quad [2]$$

where  $\theta$  is the cell constant, useful for characterization of conductivity cells, with units of  $\text{cm}^{-1}$ :

$$\theta = d/A \quad [3]$$

The specific conductance, also called the conductivity, of the solution is denoted by the symbol



**Figure 1** Principle of conductivity cell with contacting electrodes and recommended cell constant of  $1 \text{ cm}^{-1}$ .

$\kappa$  and is related to the above quantities by

$$\kappa = 1/\rho = L\theta \quad [4]$$

Actual electrodes are usually not planar and parallel. Cell constants  $\theta$  are most accurately determined by use of eqn [4] after measurement in standard reference solutions of known conductivities rather than by dimensional measurements of electrodes. Table 1 lists some of these standard solutions.

## Electrolyte Theory

The conductivity of a pure solution is given by the summation

$$\kappa = 10^{-3} \sum_i z_i C_i \lambda_i \quad [5]$$

**Table 1** Reference solutions for calibration of cell constants

Approximate concentration ( $\text{mol l}^{-1}$ )	Method of preparation	Temperature ( $^{\circ}\text{C}$ )	$\kappa$ ( $\mu\text{S cm}^{-1}$ )
1.0	74.2460 g KCl per 1 l solution at $20^{\circ}\text{C}$	0	65 176
		18	97 838
		25	111 342
0.1	7.4365 g KCl per 1 l solution at $20^{\circ}\text{C}$	0	7 138
		18	11 167
		25	12 856
0.01	0.7440 g KCl per 1 l solution at $20^{\circ}\text{C}$	0	773.6
		18	1 220.5
		25	1 408.8
0.001	Dilute 100 ml of $0.01 \text{ mol l}^{-1}$ to 1 l at $20^{\circ}\text{C}$	25	146.93

From ASTM D1125-91 (1992) *Standard Test Methods for Electrical Conductivity and Resistivity of Water*. Philadelphia, PA: American Society for Testing and Materials.

**Table 2** Equivalent ionic conductivity of selected ions at infinite dilution at  $25^{\circ}\text{C}$  ( $\text{S cm}^2 \text{ mol}^{-1}$ )

Cations <sup>a</sup>	$\lambda_0$	Temperature coefficient <sup>b</sup>	Anions <sup>a</sup>	$\lambda_0$	Temperature coefficient <sup>b</sup>
$\text{H}^+$	349.8	0.0139	$\text{OH}^-$	198.6	0.018
$\text{Co}(\text{NH}_3)_6^{3+}$	102.3		$\text{Fe}(\text{CN})_6^{4-}$	110.5	0.02
$\text{K}^+$	73.5	0.0193	$\text{Fe}(\text{CN})_6^{3-}$	101.0	
$\text{NH}_4^+$	73.5	0.019	$\text{Co}(\text{CN})_6^{3-}$	98.9	
$\text{Pb}^{2+}$	69.46	0.02	$\text{SO}_4^{2-}$	80.0	0.022
$\text{La}^{3+}$	69.6	0.023	$\text{Br}^-$	78.14	0.0198
$\text{Fe}^{3+}$	68.0		$\text{I}^-$	76.8	0.0197
$\text{Ba}^{2+}$	63.64	0.023	$\text{Cl}^-$	76.4	0.0202
$\text{Ag}^+$	61.9	0.021	$\text{NO}_3^-$	71.42	0.020
$\text{Ca}^{2+}$	59.50	0.0230	$\text{CO}_3^{2-}$	69.3	0.02
$\text{Cu}^{2+}$	53.6	0.02	$\text{C}_2\text{O}_4^{2-}$	74.2	0.02
$\text{Fe}^{2+}$	54.0		$\text{ClO}_4^-$	67.3	0.020
$\text{Mg}^{2+}$	53.06	0.022	$\text{HCO}_3^-$	44.5	
$\text{Zn}^{2+}$	52.8	0.02	$\text{CH}_3\text{CO}_2^-$	40.9	0.022
$\text{Na}^+$	50.11	0.0220	$\text{HC}_2\text{O}_4^-$	40.2	
$\text{Li}^+$	38.69	0.0236	$\text{C}_6\text{H}_5\text{CO}_2^-$	32.4	0.023
$(n\text{-Bu})_4\text{N}^+$	19.5	0.02	Picrate <sup>-</sup>	30.4	0.025

<sup>a</sup>For ions of charge  $z$ , the figures given are on an equivalent basis, that is, they apply to the fraction  $(1/z)$  of a mole.

<sup>b</sup>The temperature coefficient, when known, is given as  $(1/\lambda_0)(d\lambda_0/dT)$ , with the units  $\text{K}^{-1}$ .

From Frankenthal RP (1963) In: Meites L (ed.) *Handbook of Analytical Chemistry*, pp. 5–30. New York: McGraw-Hill.

where  $z_i$  represents the charge of the  $i$ th ion,  $C_i$  its equivalent concentration, and  $\lambda_i$  its equivalent ionic conductivity. It is also useful to define the equivalent conductivity  $\Lambda$ , with the units  $\text{S cm}^2 \text{ mol}^{-1}$ :

$$\Lambda = \sum (\lambda_+ \lambda_-) = 1000 \kappa / C \quad [6]$$

The ionic conductance  $\lambda_i$  gives quantitative information about the contribution of each ionic species to the conductance of the solution. Its value is somewhat dependent on the total ionic concentration (denoted by the ionic strength  $\sum z_i C_i$ ) of the solution and increases with increasing dilution. It is convenient to tabulate numerical values of  $\lambda_0$ , the limiting value of  $\lambda$  as the concentration approaches zero (infinite dilution). Representative values of  $\lambda_0$  are given in Table 2. For many common acids, bases, and salts, values of the equivalent conductivity at infinite dilution and at real concentrations are given in Table 3.

Physical-chemical data concerning ionic equilibria can be obtained from conductimetric studies. Conductimetric measurement has led to the determination of acid and base dissociation constants, stability constants, and solubility product constants. Further details may be found in modern texts of physical chemistry.

## Instrumentation

Two types of conductance measurement systems are in use, employing either contacting or electrodeless principles. The most widely used method utilizes either two or four contacting electrodes immersed in the solution to be studied. A pair of contacting electrodes is commonly employed but their surfaces may become contaminated, adding spurious resistance to the circuit. Owing to the increased tolerance for fouling coatings, four-electrode conductivity is becoming available and popular.

The second type of conductance measurement utilizes noncontacting sensors and depends on inductive or capacitive effects to measure conductance.

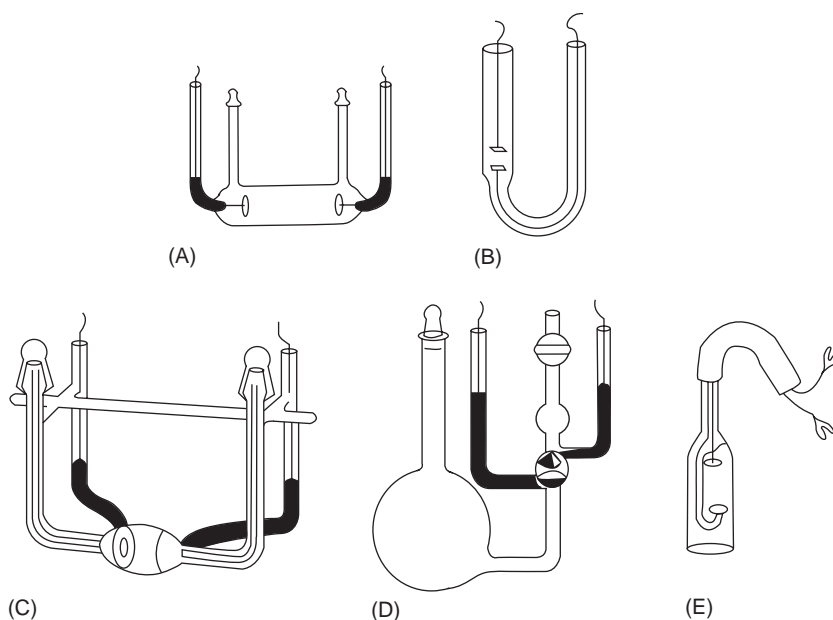
## Contacting Electrode Methodology

Figure 2 illustrates several cells with contacting electrodes. Cell designs are available for a variety of research, routine, and industrial applications. Routine measurements use conductivity cells of suitable size that may be directly immersed in beakers, flasks, etc. Water analysis and online industrial measurements may require immersion or flow-through cells for continuous monitoring. Precision measurements

**Table 3** Equivalent conductances of some electrolytes at 25°C

Electrolyte	Concentration ( $\text{eq l}^{-1}$ )							
	0.0000	0.0005	0.001	0.005	0.01	0.02	0.05	0.10
NaCl	126.45	124.50	123.74	120.65	118.51	115.76	111.06	106.74
KCl	149.86	147.81	146.95	143.55	141.27	138.34	133.37	128.96
LiCl	115.03	113.15	112.40	109.40	107.32	104.65	100.11	95.86
HCl	426.16	422.74	421.36	415.80	412.00	407.24	399.09	391.32
NH <sub>4</sub> Cl	149.7	—	—	—	141.28	138.33	133.29	128.75
KBr	151.9	—	—	146.09	143.43	140.48	135.68	131.39
KI	150.3	—	—	144.37	142.18	139.45	134.97	131.11
NaI	126.94	125.36	124.25	121.25	119.24	116.70	112.79	108.78
NaO <sub>2</sub> CCH <sub>3</sub>	91.0	89.2	88.5	85.72	83.76	81.24	76.92	72.80
NaO <sub>2</sub> CCH <sub>2</sub> CH <sub>3</sub>	85.92	84.24	83.54	80.90	79.05	76.63	—	—
NaO <sub>2</sub> C(CH <sub>2</sub> ) <sub>2</sub> CH <sub>3</sub>	82.70	81.04	80.31	77.58	75.76	73.39	69.32	65.27
KNO <sub>3</sub>	144.96	142.77	141.84	138.48	132.82	132.41	126.31	120.40
KHCO <sub>3</sub>	118.00	116.10	115.34	112.24	110.08	107.22	—	—
AgNO <sub>3</sub>	133.36	131.36	130.51	127.20	124.76	121.41	115.24	109.14
NaOH	248	246	245	240	237	233	227	221
$\frac{1}{2}$ CaCl <sub>2</sub>	135.84	131.93	130.36	124.25	120.36	115.65	108.47	102.46
$\frac{1}{2}$ BaCl <sub>2</sub>	139.98	135.96	134.34	128.02	123.94	119.09	111.48	105.19
$\frac{1}{2}$ SrCl <sub>2</sub>	135.80	131.90	130.33	124.24	120.29	115.54	108.25	102.19
$\frac{1}{2}$ MgCl <sub>2</sub>	129.40	125.61	124.11	118.31	114.55	110.04	103.08	97.10
$\frac{1}{2}$ Ca(OH) <sub>2</sub>	258	—	—	233	226	214	—	—
$\frac{1}{2}$ Na <sub>2</sub> SO <sub>4</sub>	129.9	125.74	124.15	117.15	112.44	106.78	97.75	89.98
$\frac{1}{2}$ NiSO <sub>4</sub>	—	118.7	113.1	93.2	82.7	72.3	59.2	50.8
$\frac{1}{2}$ LaCl <sub>3</sub>	145.9	139.6	137.0	127.5	121.8	115.3	106.2	99.1
$\frac{1}{4}$ K <sub>4</sub> Fe(CN) <sub>6</sub>	184	—	167.24	146.09	134.83	122.82	107.70	97.87

From MacInnes DA (1961) *The Principles of Electrochemistry*, p. 339. New York: Dover Publications.



**Figure 2** Types of contacting conductivity cells: (A) Jones and Bollinger; (B) Roseveare; (C) Shedlovsky; (D) flask type (Shedlovsky); (E) dipping cell. (From Light TS and Ewing GW (1990) *Measurement of electrolytic conductance*. In: Ewing GW (ed.) *Analytical Instrumentation Handbook*, pp. 641–658. New York: Dekker.)

**Table 4** Recommended cell constants for various conductance ranges

Conductance range ( $\mu\text{S cm}^{-1}$ )	Cell constant ( $\text{cm}^{-1}$ )
0.05–20	0.01
1–200	0.1
10–2000	1.0
100–20 000	10.0
1000–200 000	50.0

From ASTM D1125-91 (1992) *Standard Test Methods for Electrical Conductivity and Resistivity of Water*. Philadelphia, PA: American Society for Testing and Materials, with permission.

require that samples be contained within the conductivity cell, which may then be inserted in a constant-temperature bath, since the conductance of most solutions increases with the temperature at the rate of  $\sim 2\%$  per kelvin (cf. Table 2). Cells are available with built-in temperature sensors and associated instrumentation that computes automatic temperature compensation. Cells are made with various cell constants to accommodate solutions over wide ranges of conductances. Constants of 50, 10, 1.0, 0.1, and  $0.01\text{ cm}^{-1}$  are most widely used. Table 4 shows cell constants recommended for various ranges of conductance.

Electrodes are commonly made of platinum, although other noble or passive metals may be used including gold, silver, titanium, monel, stainless steel, and tungsten. The nonmetal graphite is also used because it is inexpensive, inert, and conductive.

Platinum electrodes may be coated with platinum black ('platinized') to increase the surface area and sensitivity. This finely divided form of platinum is prepared by electrolysis in a chloroplatinic acid solution containing a small amount of lead acetate. Platinization of platinum cells is not recommended on cells for measurements below  $10\text{ }\mu\text{S cm}^{-1}$ , although a trace of platinum black is sometimes suggested in these regions.

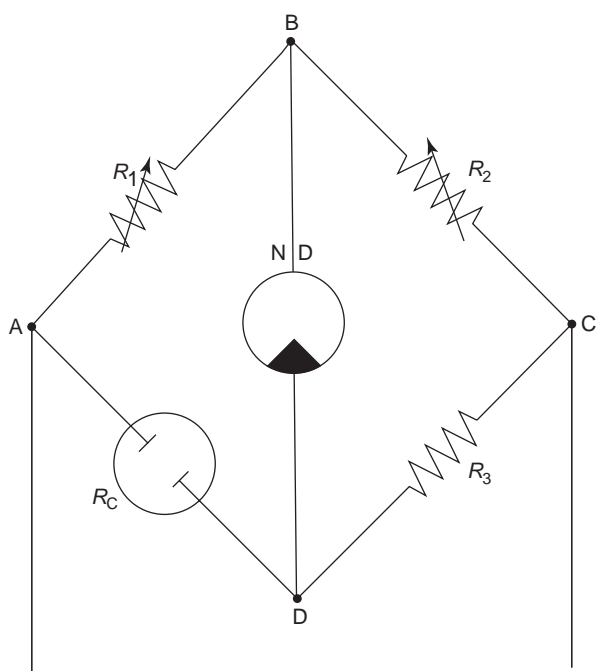
#### Circuit Considerations with Contacting Electrodes

Measurement of conductance, which is a measurement of the reciprocal of resistance, is done with a Wheatstone bridge such as that shown in Figure 3. In its traditional form, the bridge consists of four resistances, the unknown cell resistance  $R_C$  and three precision variable resistances,  $R_1$ ,  $R_2$ , and  $R_3$ , which are manipulated to show zero voltage drop across BD as observed by the null detector ND. Calculation of the unknown resistance is then derived from Ohm's law, where  $V$  is the voltage and  $I$  the electrical current across a resistance  $R$ :

$$V = IR \quad [7]$$

$$R_C = R_1(R_3/R_2) \quad [8]$$

If the voltage source used were direct current, chemical reactions might occur and deposits or gases evolving at the electrodes would change the electrical resistance. This effect, known as 'polarization' of the



**Figure 3** Wheatstone bridge for measuring conductance. (From Sawyer DT, Heineman WR, and Beebe JM (1984) *Chemistry Experiments for Instrumental Methods*, p. 61. New York: Wiley, with permission.)

electrodes, is eliminated by using alternating current as the voltage source to drive the bridge. However, the introduction of alternating current introduces a new set of measurement problems of its own. Instead of Ohm's law as in eqn [7], the voltage is governed by the alternating current impedance,  $Z$ :

$$V = IZ \quad [9]$$

The impedance now depends on three new variables: the frequency, inductance, and capacitance. The impedance,  $Z$ , in ohms, is given by equations such as

$$Z = (R_2 + X_2)^{1/2} \quad [10]$$

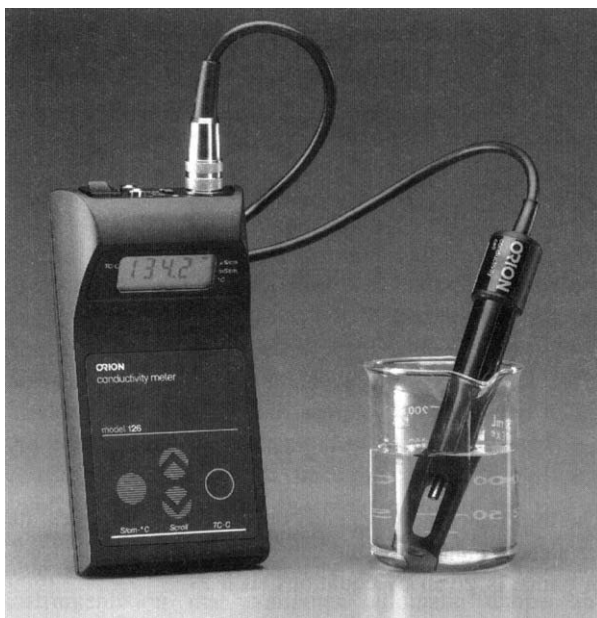
$$X = X_L - X_C \quad [11]$$

$$X_L = 2\pi fL \quad [12]$$

$$X_C = 1/(2\pi fC) \quad [13]$$

where  $R$  is the solution resistance in ohms as before,  $X_L$  is the inductive reactance in ohms,  $L$  is the inductance in henrys,  $X_C$  is the capacitive reactance in ohms,  $C$  is the capacitance in farads, and  $f$  is the frequency in hertz.

Driven by the alternating current, the conductivity cell now has capacitances between the electrodes, across the double layer presented by the solution-electrode interface and between the wires of the connecting cable. Additional resistances and inductances in the wire leads and connections further complicate



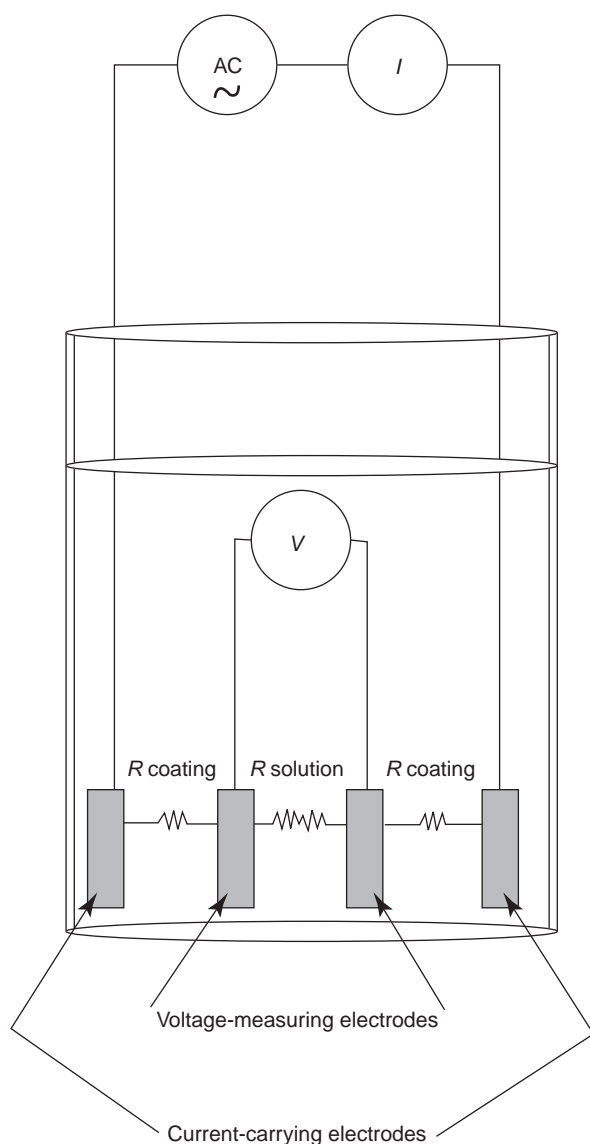
**Figure 4** Digital conductivity meter. (Courtesy of Thermo Electron Corporation, Boston, USA, used by permission.)

the construction of an accurate measuring circuit. Conductivity measuring circuits may have variable frequency sources ranging from 60 to 50 000 Hz. Low conductivity measurements made with small cell constants favor the lower frequency region. Modern conductivity instrumentation minimizes these sources of error by an appropriate combination of alternating current voltage and frequency, wave shape and sampling, and temperature compensation. A conductivity instrument with microprocessor, digital readout, and assorted conductivity cells is shown in **Figure 4**. Electronic instrumentation is available that automatically corrects for the cell constant and temperature and reads directly in terms of concentration of common electrolytes.

#### Four-Electrode Contacting Conductivity Cells

In many applications, electrodes become coated because of the characteristics of the measuring solution. A four-electrode technique that diminishes the effect of resistive electrode coatings is shown in **Figure 5**. Voltage from an alternating current source is applied to the two outer, current-carrying, electrodes in the same manner as the two-electrode measurement. The current,  $I$ , is a function of the sum of the solution and coating resistances. However, if two additional voltage measuring electrodes are introduced, such that the voltage across them is measured potentiometrically (with no current being withdrawn from the circuit by the measuring device), then the 'IR' drop of eqn [7] is measured only across the solution





**Figure 5** Principle of four-electrode conductivity measurement.

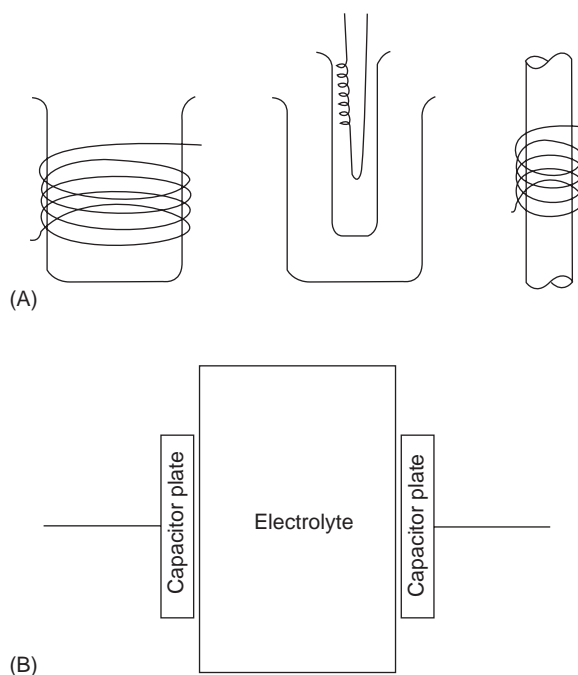
and is not affected by any resistive coating on the voltage measuring electrodes since no current is passing through this coating.

## Electrodeless (Noncontacting Methodology)

There are also two methods of conductance measurement that may be made without the electrodes in contact with solution. The first is a high-frequency method, frequently called oscillometry. The second is called 'electrodeless' or inductive conductivity.

### Oscillometry

Oscillometry employs high frequencies in the megahertz domain. Using a sample cell made of an



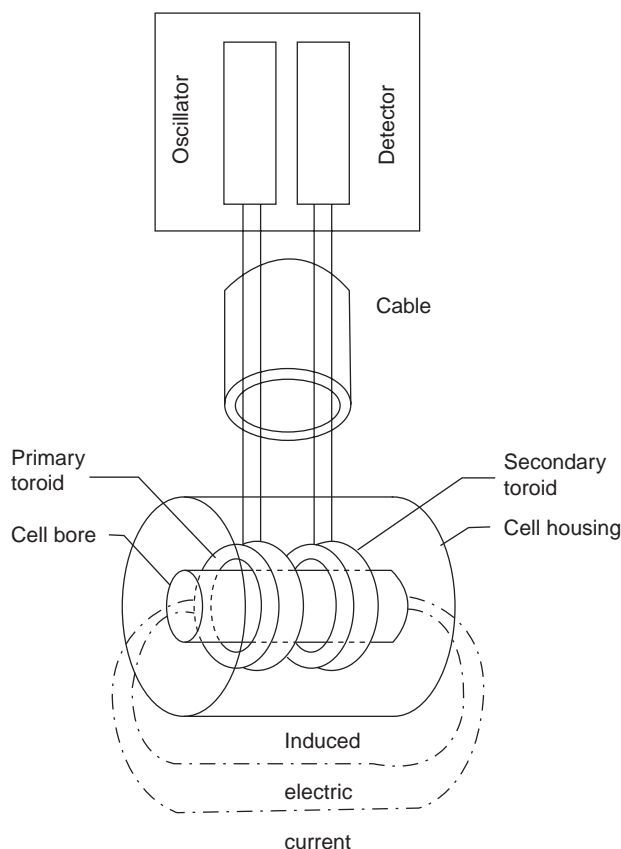
**Figure 6** (A) Inductance measuring oscillometric cell (From Pungor E (1965) *Oscillometry and Conductometry*. London: Pergamon, with permission). (B) Capacitance measuring oscillometric cell.

electrically insulating material such as glass or ceramic, the measuring cells can be classified as either capacitive or inductive. **Figure 6A** illustrates inductance high-frequency measuring cells and **Figure 6B** a capacitance cell. Special high-frequency circuits are associated with these measurements and extensive applications, paralleling those of classical conductimetry, have been reported using oscillometric techniques. Because of the complexity of interpretation, the observed data differ from those obtained with low-frequency methods. Measurements reported by oscillometry do not appear to have the exactitude of interpretation corresponding to conventional conductimetry. Commercial apparatus is of limited availability. An excellent and extensive monograph has been written by Pungor (see Further Reading), who has pioneered the work in this field.

### Electrodeless (Inductive) Conductivity

A method of measuring conductance without the use of contacting electrodes is applied for continuous and reliable measurement in the chemical process industries. A probe consisting of two encapsulated toroids in close proximity to each other is shown in **Figure 7**. This sensing unit, contained within non-conducting material such as glass, fluorocarbon, or other high-temperature resistant thermoplastic

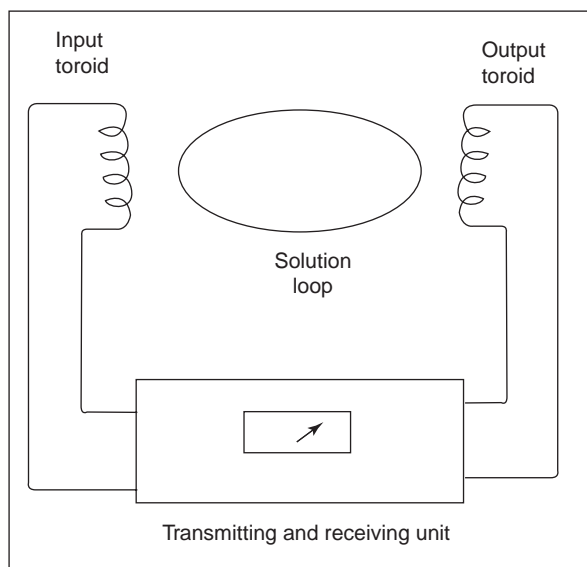




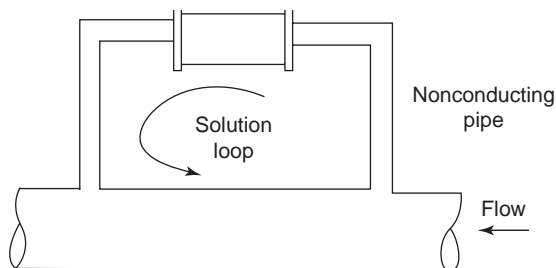
**Figure 7** Principle of the electrodeless conductivity cell and instrument. (From Light TS and Ewing GW (1990) *Measurement of electrolytic conductance*. In: Ewing GW (ed.) *Analytical Instrumentation Handbook*, pp. 641–658. New York: Dekker.)

material, is totally immersed in the electrolytic solution. Precipitates or coatings adhering to this probe have little or no effect on the measured conductance. The input toroid generates a low-frequency electric field. An alternating current is induced in the solution and sensed by the output toroid as shown in **Figure 8**. The equivalent electrical circuit may be compared to a transformer with the toroids forming the primary and secondary windings, and the core replaced by a loop formed by the conducting solution. A configuration in which the probe does not come in contact with the solution is also possible and is illustrated in **Figure 9**. Here, the toroidal unit is installed around a section of nonconductive pipe containing the test solution. A complete current loop must exist for this arrangement to function.

In either case, the generating toroid is energized from a source of a stable ultrasonic frequency, typically 20–50 kHz. The pickup toroid is connected to a receiving circuit that measures the current induced through the secondary winding. The current is then amplified and displayed on a meter or transmitted to a strip chart recorder or computer. The output is a



**Figure 8** Representation of an electrodeless conductivity measuring circuit. (From Light TS and Ewing GW (1990) *Measurement of electrolytic conductance*. In: Ewing GW (ed.) *Analytical Instrumentation Handbook*, pp. 641–658. New York: Dekker.)



**Figure 9** Electrodeless conductivity cell mounted externally to a nonconducting pipe. (From Light TS and Ewing GW (1990) *Measurement of electrolytic conductance*. In: Ewing GW (ed.) *Analytical Instrumentation Handbook*, pp. 641–658. New York: Dekker.)

direct function of the conductance of the solution loop, and is interpreted in a manner analogous to the traditional measurement with contacting electrodes.

Commercially available instruments, one of which is shown in **Figure 10**, have useful ranges extending from  $50 \mu\text{S cm}^{-1}$  to  $2 \text{ S cm}^{-1}$ , with relative accuracy of a few tenths of a percent of full-scale, after-temperature compensation. A temperature sensor is incorporated in the toroid probe, and a compensation circuit corrects readings to the standard reference temperature of  $25^\circ\text{C}$ .

The electrodeless conductivity technique is used mainly in the chemical process industries for continuous monitoring where its freedom from maintenance is superior to contacting techniques. The smallest electrodeless probe is  $\sim 3.6 \text{ cm}$  diameter,

has an equivalent cell constant of  $2.5 \text{ cm}^{-1}$ , and requires a minimum solution volume of several hundred milliliters to ensure a complete solution loop without wall effects that distort the apparent cell constant. For the lower conductivity ranges, which require a smaller cell constant (cf. Table 4), the diameter of the probe must increase, and accurate

measurement below  $\sim 10 \mu\text{S cm}^{-1}$  is not practical. The relatively large probe and sample size have hindered applications for laboratory use.

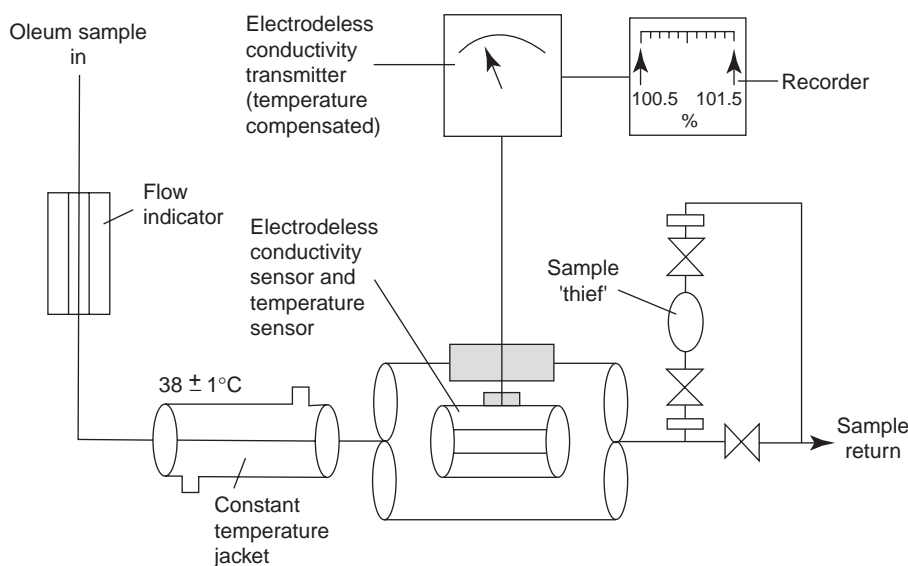
Numerous applications of electrodeless conductivity sensors have been published for the chemical, pulp and paper, aluminum, mining, and food industries. Similar instrumentation has been used for *in situ* measurements of the salinity of seawater. An instrument for continuous analysis of oleum in the range of 100–102% equivalent sulfuric acid, with an accuracy of 0.01%, is illustrated in Figure 11. A historical review has been published by Light (see Further Reading).



**Figure 10** Digital electrodeless conductivity instrument used for online process measurements. (Courtesy of The Foxboro Company, Foxborough, USA.)

## Applications

Specific conductance and resistance values for a number of solutions are shown in Figure 12. Water itself is a very poor conductor. Its specific conductance due to dissociation into  $\text{H}_3\text{O}^+$  and  $\text{OH}^-$  ions is  $0.0550 \mu\text{S cm}^{-1}$  (resistivity of  $18.18 \text{ M}\Omega \text{ cm}$ ) at  $25^\circ\text{C}$ . Water of this theoretical purity is produced using commercially available nuclear-grade ion exchange resins. It is used extensively in the semiconductor, power, and pharmaceutical industries, and has replaced distilled water for general laboratory use. The conductivity measurement is extremely sensitive to traces of ionic impurities. The presence of  $1 \mu\text{g l}^{-1}$  of sodium chloride will raise the conductivity to  $0.0571 \mu\text{S cm}^{-1}$  ( $17.5 \text{ M}\Omega \text{ cm}$ ) at  $25^\circ\text{C}$ . A good grade of distilled or deionized water in contact with air measures  $1 \mu\text{S cm}^{-1}$  or ( $1 \text{ M}\Omega \text{ cm}$ ) at  $25^\circ\text{C}$  and falls far short of the purity requirement of



**Figure 11** A continuous analyzer for oleum using an electrodeless conductivity sensor. (From Shaw R and Light TS (1982) Online analysis of oleum using electrodeless conductivity. ISA Transactions 21: 63.)

Resistivity ( $\Omega$ cm)	$10^8$	$10^7$	$10^6$	$10^5$	$10^4$	$10^3$	$10^2$	10	1
Conductivity ( $\mu$ S $\text{cm}^{-1}$ )	$10^{-2}$	$10^{-1}$	1	10	$10^2$	$10^3$	$10^4$	$10^5$	$10^6$
Ultrapure water									
Demineralized water									
Condensate									
Natural waters									
Cooling tower coolants									
Percent level of Acids, bases, and salt									
5% Salinity									
2% NaOH									
20% HCl									
Range of contacting cells									
Range of electrodeless									

**Figure 12** Conductivity spectrum. (From Light TS (1990) Conduct electricity without electrodes. *CHEM-TECH August*: 496–501.)

**Table 5** Conductivity and resistivity of pure water

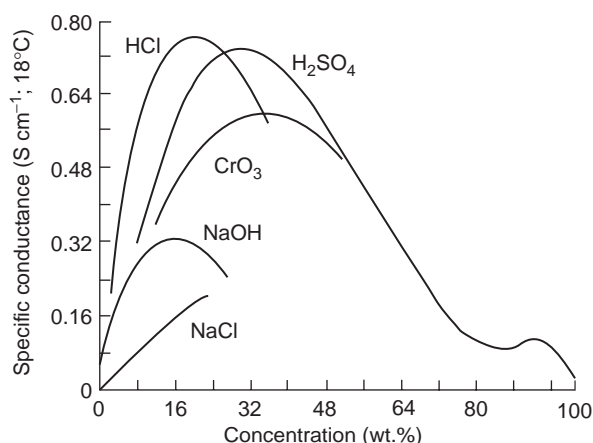
Temperature ( $^{\circ}$ C)	Conductivity ( $\mu$ S $\text{cm}^{-1}$ )	Resistivity ( $M\Omega$ cm)
0	0.011 65	85.84
10	0.023 10	43.30
20	0.041 94	23.84
25	0.055 00	18.18
30	0.071 01	14.08
40	0.113 5	8.810
50	0.172 7	5.791
60	0.251 4	3.978
70	0.351 6	2.844
80	0.474 4	2.108
90	0.620 5	1.612
100	0.793 0	1.261

From Thornton RD and Light TS (1989) A new approach to accurate resistivity measurement of high purity water. *Ultrapure Water* 6(5): 14–26, with permission.

many industries. The conductivity of water has a large temperature coefficient compared to salts, acids, and bases. **Table 5** shows the conductivity of water over the range of 0–100 $^{\circ}$ C.

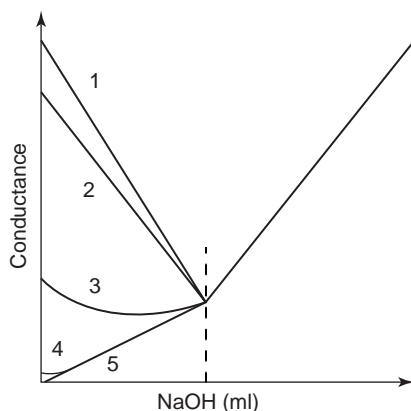
Solutions of strong electrolytes show a nearly linear increase of conductance with concentration up to  $\sim$ 10–20 wt.%. At higher concentrations the conductance decreases again due to such interactions as complexation reactions, formation of dimers, or higher polymers and increased viscosity. **Figure 13** shows the relation between conductivity and concentration for a few representative substances.

Conductometric titrimetry is widely applicable for titration reactions involving ions. **Figure 14**



**Figure 13** Conductivity–concentration curves for selected electrolytes. (From Ewing GW (1985) *Instrumental Methods of Chemical Analysis*, 5th edn., p. 337. New York: McGraw-Hill.)

demonstrates curves that result from titration of acids with sodium hydroxide. Graduations of acids, ranging from strong to very weak, are shown. Curve 1 represents a strong acid such as hydrochloric acid. As additions of sodium hydroxide are made, the conductance decreases until the equivalence point is reached. The point at which the slope changes defines the endpoint. For progressively weaker acids, such as dichloroacetic, monochloroacetic, or acetic acid, curves 2–4 represent the progress of titration. Unlike potentiometric or visual indicator methods, conductometric titrations do not require exact data in the vicinity of the endpoint. Extrapolation of the



**Figure 14** Conductometric titration curves of various acids by sodium hydroxide. Curve 1 represents a strong acid, and curve 5 an extremely weak one, while the others are intermediate. The acids are (1) hydrochloric acid, (2) dichloroacetic acid, (3) monochloroacetic acid, (4) acetic acid, and (5) boric acid. (From Ewing GW (1985) *Instrumental Methods of Chemical Analysis*, 5th edn., p. 337. New York: McGraw-Hill.)

data obtained before and after the slope change determines the endpoint. For titrations of very weak acids, such as curve 5 for boric acid, where the slope increases continuously, the endpoint cannot be detected by the usual potentiometric and visual methods. Using the slope change method of conductometric titrations, the endpoint may be determined by extrapolation.

Other types of titrations including those of mixtures of acids, precipitation titrations, such as silver with halides, and complex reactions are also adaptable to conductometric titration methods and frequently yield better analytical results than the

potentiometric or indicator methods. Analytical chemistry texts should be consulted for details of methodology.

Conductivity measurements are useful in following the reaction kinetics or diffusion processes when a change in ionic content or mobility is involved.

Ion chromatography using a conductivity detector has developed into one of the major instrumental tools. Ion chromatography fills a long-standing gap in analytical methods for anions by separating weak or strong ionic compounds on ion exchange columns followed by conductivity detection in a microcell.

**See also:** **Ion Exchange:** Ion Chromatography Instrumentation; Ion Chromatography Applications. **Water Analysis:** Overview; Seawater – Inorganic Compounds.

## Further Reading

International Electrotechnical Commission (1985) IEC Standard, Publication 746-3, *Expression of Performance of Electrochemical Analyzers, Part 3: Electrolytic Conductivity*. Geneva: Bureau Central de la Commission Electrotechnique Internationale.

Klug O and Lopatin BA (eds.) (1988) New developments in conductimetric and oscillometric analysis. In: *Wilson & Wilson's Comprehensive Analytical Chemistry*, vol. XXI. Amsterdam.

Light TS (1989) Electrodeless conductivity. In: Stock JT and Orna MV (eds.) *Electrochemistry, Past and Present*, pp. 429–441. ACS Symposium Series 390. Washington, DC: American Chemical Society.

Light TS, McHale EJ, and Fletcher KS (1989) Electrodeless conductivity. *Talanta* 36: 235–241.

# CONTINUOUS-FLOW ANALYSIS

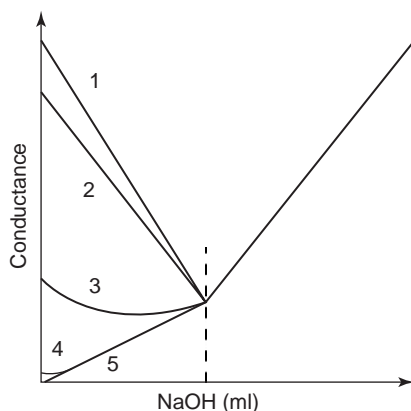
**See** **FLOW ANALYSIS:** Overview. **FLOW INJECTION ANALYSIS:** Principles; Instrumentation; Detection Techniques; Environmental and Agricultural Applications; Clinical and Pharmaceutical Applications; Industrial Applications. **SEGMENTED FLOW ANALYSIS.** **SEQUENTIAL INJECTION ANALYSIS**

# COSMETICS AND TOILETRIES

**A Salvador and M C Pascual-Martí**, University of Valencia, Valencia, Spain

## Introduction

The use of cosmetic products goes back thousands of years to antiquity; they were even found in some Egyptian graves. In the past, cosmetics were



**Figure 14** Conductometric titration curves of various acids by sodium hydroxide. Curve 1 represents a strong acid, and curve 5 an extremely weak one, while the others are intermediate. The acids are (1) hydrochloric acid, (2) dichloroacetic acid, (3) monochloroacetic acid, (4) acetic acid, and (5) boric acid. (From Ewing GW (1985) *Instrumental Methods of Chemical Analysis*, 5th edn., p. 337. New York: McGraw-Hill.)

data obtained before and after the slope change determines the endpoint. For titrations of very weak acids, such as curve 5 for boric acid, where the slope increases continuously, the endpoint cannot be detected by the usual potentiometric and visual methods. Using the slope change method of conductometric titrations, the endpoint may be determined by extrapolation.

Other types of titrations including those of mixtures of acids, precipitation titrations, such as silver with halides, and complex reactions are also adaptable to conductometric titration methods and frequently yield better analytical results than the

potentiometric or indicator methods. Analytical chemistry texts should be consulted for details of methodology.

Conductivity measurements are useful in following the reaction kinetics or diffusion processes when a change in ionic content or mobility is involved.

Ion chromatography using a conductivity detector has developed into one of the major instrumental tools. Ion chromatography fills a long-standing gap in analytical methods for anions by separating weak or strong ionic compounds on ion exchange columns followed by conductivity detection in a microcell.

**See also:** **Ion Exchange:** Ion Chromatography Instrumentation; Ion Chromatography Applications. **Water Analysis:** Overview; Seawater – Inorganic Compounds.

## Further Reading

International Electrotechnical Commission (1985) IEC Standard, Publication 746-3, *Expression of Performance of Electrochemical Analyzers, Part 3: Electrolytic Conductivity*. Geneva: Bureau Central de la Commission Electrotechnique Internationale.

Klug O and Lopatin BA (eds.) (1988) New developments in conductimetric and oscillometric analysis. In: *Wilson & Wilson's Comprehensive Analytical Chemistry*, vol. XXI. Amsterdam.

Light TS (1989) Electrodeless conductivity. In: Stock JT and Orna MV (eds.) *Electrochemistry, Past and Present*, pp. 429–441. ACS Symposium Series 390. Washington, DC: American Chemical Society.

Light TS, McHale EJ, and Fletcher KS (1989) Electrodeless conductivity. *Talanta* 36: 235–241.

# CONTINUOUS-FLOW ANALYSIS

**See** **FLOW ANALYSIS:** Overview. **FLOW INJECTION ANALYSIS:** Principles; Instrumentation; Detection Techniques; Environmental and Agricultural Applications; Clinical and Pharmaceutical Applications; Industrial Applications. **SEGMENTED FLOW ANALYSIS.** **SEQUENTIAL INJECTION ANALYSIS**

# COSMETICS AND TOILETRIES

**A Salvador and M C Pascual-Martí**, University of Valencia, Valencia, Spain

## Introduction

The use of cosmetic products goes back thousands of years to antiquity; they were even found in some Egyptian graves. In the past, cosmetics were



considered as beautifying substances or preparations (e.g., lipstick, eye-shadow) that when applied to the face or the body would make us more attractive, whereas the term toiletries was sometimes reserved for those articles used in grooming (e.g., shampoo, deodorant) for washing or caring for the appearance. Nowadays, our perception of cosmetics has become more extensive and toiletries, perfumes, and other products have been included within the broad term 'cosmetics'. The cosmetic industry generates a lot of jobs and is an important source of income in the industrialized countries, as well as an indicator of prosperity.

This article deals with the different types of cosmetic formulations (including toiletries), forms and functions, analytes of interest, pretreatment and sample preparation, and analytical techniques. A thorough description of the main ingredients of toiletries – namely surfactants, is given in another article and due to their particular relevance and special characteristics, perfumes are also considered elsewhere.

## Definitions and Nomenclature

A cosmetic, according to the current European Union (EU) legislation, is any substance or preparation intended to be placed in contact with the various external parts of the human body (epidermis, hair system, nails, lips, and external genital organs) or with the teeth and mucous membranes of the oral cavity with a view exclusively or mainly to cleansing them, perfuming them, changing their appearance, and/or correcting body odors and/or protecting them or keeping them in good condition.

The US Food and Drug Administration (FDA) defines cosmetics by their intended use, as articles intended to be rubbed, poured, sprinkled, or sprayed on, introduced into or, otherwise, applied to the human body for cleansing, beautifying, promoting attractiveness, or altering the appearance.

Different laws and regulations apply to cosmetics and drugs. However, some products meet both definitions. This may happen when a product has two intended uses. Such products must comply with the requirements for both cosmetics and drugs. For example, an antidandruff treatment shampoo is a drug because one of its intended uses is to treat dandruff, but it is also a cosmetic because its other use is to clean hair. The same happens with toothpastes containing fluoride. Other products such as sunscreens are considered by EU legislation as cosmetics (they protect the external part of the human body) whereas they are considered as drugs by US legislation (the

protection function is not included in the definition of the FDA). The cosmetics industry uses the word cosmeceuticals to refer to those cosmetic products that have medicinal or drug-like benefits, although the legislation does not recognize this term.

Generally, the cosmetic ingredients are designated either by their chemical name or, preferentially by the simplest International Nomenclature Cosmetic Ingredient (INCI) to take into account the need for a truly international approach enabling comprehensive and short names for use in labeling.

## Types of Cosmetics: Cosmetic Forms

Cosmetic formulations are composed of active ingredients and excipients. Active ingredients are compounds directly related with the efficacy of the cosmetic product. Excipients can have different functions such as: to facilitate the preparation of the formulation, to achieve the required physicochemical properties, to improve the efficacy or to provide stability to the finished product. On the other hand, active ingredients in some cosmetic formulations can act as auxiliary ingredients in other formulations.

Different reasons determine the physical form chosen by the manufacturers for a cosmetic formulation, these may include: the solubility of the active ingredients, the type of skin, the adequate level of skin penetration according to the function of the cosmetic, or some commercial reasons such as comfort or the easiness of use.

**Table 1** shows the different types of cosmetics and their more common cosmetic forms.

## Efficacy and Safety: Analytes of Interest in Cosmetics Analysis

Cosmetic samples are often very complex. They may contain numerous ingredients, in a very wide range of concentrations and for diverse applications. The complexity of these samples ranges from just a few to 20 or more ingredients, which can be found in any given manufactured cosmetic.

The active ingredients or excipients included in cosmetic formulations have different functions (e.g., antioxidants, surfactants or emulsifying agents, preservatives).

**Table 2** shows the most important functions of cosmetic products (according to the EU) and some examples of compounds which can provide these properties to the cosmetics.

The analytes of main interest in cosmetic formulations are those related to the efficacy and safety of the products.



**Table 1** Types of cosmetic products and their most common formulation forms

<i>Cosmetic products</i>	<i>Most common formulation forms</i>
<i>General toilet products</i>	
After-bath powders, hygienic powders	Powder
Toilet soaps, deodorant soaps	Solid soap
Toilet waters, eau de Cologne	Lotion, water, spray, paper toilet
Bath and shower preparations	Salt, foam, oil, gel
Products for care of the teeth and mouth	Paste, gel, lotion
Products for external intimate hygiene	Gel, lotion, paper towel
Deodorants and antiperspirants	Stick, spray
<i>Depilatories and shaving products</i>	
Depilatories	Lotion, wax
Shaving products	Cream, foam, lotion
<i>Sun bronzing and whitening products</i>	
Sunscreens products	Cream, lotion, emulsion, spray
Products for tanning without sun	Cream, lotion, emulsion, spray
Skin-whitening products	Cream, lotion, emulsion
<i>Skin care products</i>	
Products for skin care of hands, face, feet, lips, body, etc.	Cream, emulsion, lotion, gel, oil, solutions, suspensions
Antiwrinkle products	Cream, lotion, emulsion, gel
Face masks (with the exception of chemical peeling products)	Paste, cream, gel
<i>Make-up and removing products</i>	
Products for making-up the face and the eyes	Powder, cream, suspension, past, pencil
Products for removing make-up from the face and eyes	Cream, gel, lotion, paper towel
Products intended for application to the lips	Stick, cream
Products for nails	Lotion, suspension
<i>Tinted bases and hair care products</i>	
Products for dying, bleaching, waving, straightening, fixing, setting, cleansing, conditioning, hairdressing products	Cream, lotion, oil, powder, shampoo, spray

The efficacy of a cosmetic product can be defined as the capacity to provide the desired effect. The safety of cosmetic products is of great importance because these products are used without medical supervision, on sensitive areas of the body and/or for long periods of time. The safety relates to composition, packaging, and information and responsibility

for all these aspects falls totally on the producer or the importer, who is responsible for any marketing liability. A rigorous control of the cosmetic product on the market is a key requirement in ensuring the protection of the consumer; therefore, quality control of the analytes of interest in the finished product is necessary to assure both the efficacy and the safety (Figure 1). The cosmetic industry requires simple, accurate and, if possible, green analytical methods that can be routinely and safely used in the quality control laboratories.

### Efficacy of the Finished Product

Active ingredients provide the efficacy of the cosmetic formulation to achieve the desired result with the aid of appropriate excipients. For example, the function of deodorants is to inhibit the growth of microorganisms which cause unpleasant odors; zinc salts of the rinoic acid, or triclosan (2,4,4'-trichloro-2'-hydroxydiphenyl ether) can be used as active ingredients. Antiperspirants have to reduce the secretion of the sweat glands, aluminum or zirconium salts can do this. Organic coloring compounds are used in hair dye formulations. Organic compounds whose molecules contain  $\pi$  electrons which absorb UV radiation can be used as filters in sunscreens, thus imitating melanine which prevents the human body from absorbing solar radiation naturally. As the active ingredients are directly related to the desired efficacy of the cosmetic, the determination of their concentration at production level has to be controlled.

Some ingredients can degrade with time or storage conditions and they must be specially controlled in order to ensure the efficacy of the cosmetic. In this sense, the expiry data must be indicated with the warning: 'Use preferably before...'. If necessary, the storage conditions should also be indicated. Moreover, if the useful life of the product is less than 30 months, EU legislation obliges the manufacturer to indicate the expiry date after the product has been opened.

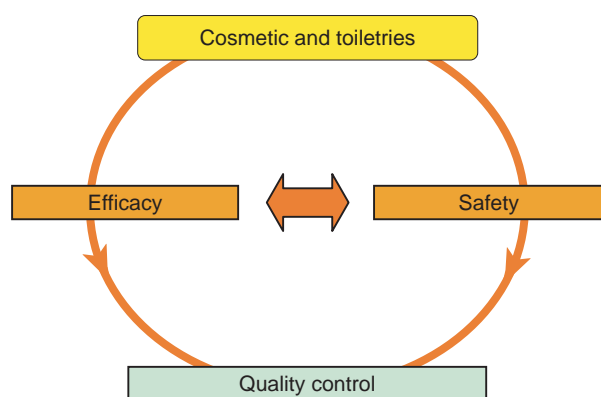
### Safety of the Finished Product

There are official statements in each country that legislate about finished cosmetic products, for example: the Scientific Committee on Cosmetic Products and Non-Food Products (SCCNFP) or the EU Directorate General Health and Consumer Protection or the Center for Food Safety and Applied Nutrition (CFSAN) in the US FDA.

A large number of compounds are considered as potentially harmful for human health and their use as cosmetic ingredients (e.g., cadmium and its

**Table 2** Representative compounds included in the cosmetic formulations to achieve different functions

Function	Examples of compounds used in the cosmetic formulations
Abrasives	Silica, alumina, pumice, sodium silicoaluminate, zeolites
Absorbents	Cellulose, bismuth subnitrate, talc, magnesium carbonate
Additives	Acetaldehyde, acetaminophen, gallic acids
Anticorrosive	Nitromethane, sodium nitrite
Antidandruff	Zinc pyrithione, ketoconazole, piroctone olamine
Antifoaming	Isopropyl alcohol, hexyl alcohol, silica silylate, dimethicone
Antimicrobials	Benzoic acid, potassium iodide, zinc acetate, zinc borate, <i>p</i> -hydroxybenzoates
Antioxidants	Butylhydroxytoluene, citric acid and its salts, hydroquinone, ascorbic acid, tocopherol
Antiperspirant	Aluminum chlorhydrate, aluminum chloride, aluminum sulfate
Antistatic	Alanine, oleamine, lanolin, pyridoxine, aspartic acid
Binders	Agar, alginic acid, starch, PTFE, synthetic wax
Biological additives	Amylase, ovum, oyster ( <i>Ostrea edulis</i> ), collagen, colostrums, folic acid
Bleaching	Ammonium persulfate, hydroquinone, strontium dioxide
Botanicals	<i>Aloe barbadensis</i> , <i>Avena sativa</i> , <i>Castanea sativa</i> , <i>Coffea arabica</i>
Buffering	Phosphoric acid, citric acid, lactic acid, ethanolamine
Chelating	Oxalic acid, bismuth citrate, diisopropyl oxalate, EDTA
Cosmetic coloring	Inorganic and organic natural pigments, synthetic pigments
Denaturants	Acetone, ammonia, eucalyptol
Deodorant	Zinc gluconate, zinc glutamate, zinc ricinoleate, triclosan
Depilatory	Thioglycolic acid, potassium thioglycolate, zinc sulfide
Emollients	Propylene glycol citrate, cholesterol, dihexyl adipate, vegetable oils
Emulsifying	Triethanolamine, stearic acid, stearyl alcohol, sodium lauryl sulfate
Emulsion stabilizers	Aluminum dilinoleate, magnesium hydroxide stearate
Film formers	Acetyltributyl citrate, polyacrylamide, polyisobutene, starch, triacetin
Hair dyes products	Henna, resorcinol, thymol, silver nitrate, benzenic derivatives
Humectants	Glycerin, propylene glycol, sorbitol, glucose, fructose, glutamic acid
Opacifiers	Silica, cellulose, zinc carbonate, alumina
Oral care	Calcium carbonate, allantoin, aluminum fluoride, sodium monofluorophosphate, sodium fluorosilicate
Oxidizing	Potassium chlorate, potassium bromate, potassium persulfate
Preservatives	Ammonium sulfite, formaldehyde, propionic acid, formic acid
Propellants	Carbon dioxide, nitrogen, ethane, propane
Reducing	Thioglycolic acid, isooctyl thioglycolate, sodium hydrosulfite (sodium dithionite)
Solvents	Water, alcohol, glycerin, acetone, pentane, isobutyl acetate
Surfactants	Tridecylbenzenesulfonic acid, potassium lauryl sulfate, potassium myristate, potassium laurate
UV absorbers	Titanium oxide, benzophenones, PABA and derivatives, methoxycinnamates, salicylates
Viscosity	Polyacrylic acid, methylcellulose, zinc laurate, tin oxide, gelatin, paraffin, agar, alginate, betaine

**Figure 1** Relationship between both efficacy and safety of the cosmetics and production quality control.

compounds, hexachlorophene) is prohibited by the different legislations.

Other substances are subject to restrictions in concentration and/or field of application (e.g., after

toxicological evaluations on a large number of dyes, the SCCNFP indicates the parts of the body where they can be safely applied, the cosmetics containing them, the maximum concentration levels authorized and some specific warnings).

Some substances have special labeling instructions (e.g., FDA suggests that the labeling of a cosmetic product that contains an alpha hydroxy acid (AHA) and that is intended for topical application to the skin or mucous membrane should carry instructions that convey the following information: 'Sunburn Alert: This product contains an alpha hydroxy acid (AHA) that may increase your skin's sensitivity to the sun and particularly the possibility of sunburn. Use a sunscreen and limit sun exposure while using this product and for a week afterwards.'

These substances, therefore, present particular concerns related to the safety of the finished product.

Furthermore, routine analysis and quality control often requires the analysis of raw materials. For

instance, even though methyleugenol is naturally present in essential oils used as components in cosmetic products, the SCCNFP recommends that methyleugenol should not be intentionally added as a cosmetic ingredient, and the maximum authorized level in the finished product is legislated. Possible contaminants present in the raw materials must also be checked.

A cosmetic product must not cause damage to human health when applied under normal or reasonably foreseeable conditions of use. These products must not be harmful, either immediately (for example, by causing allergic reactions) or in the long term (for example, leading to cancer or natal defects).

Different bioassays using animals were used in the past to evaluate the safety and efficacy of finished products but nowadays, authorities encourage researchers to develop alternative *in vitro* methods that can substitute all the animal assays. The aim is both to avoid or to reduce the number of these experiments at the present time and to completely forbid them in the near future.

Studies on safety and efficacy are continuously being developed by researchers. The advances in some of these studies result in nocive ingredients being forbidden and substituted by others. Other studies are devoted to search for new ingredients to improve the efficacy of the formulations.

Cosmetic products should remain where they are most effective, that is on the skin surface; they should not penetrate into tissue. However, different *in vivo* and *in vitro* studies show that some cosmetic ingredients are absorbed through the skin. Due to this, the concern about the selection of cosmetic ingredients now includes the analysis of samples taken from cosmetic users such as urine, blood serum, nails, hair, etc (e.g., determination of parabens on the skin of human fingers, determination of organic UV filters in urine). The development of selective and sensitive analytical methods, which are capable of controlling the bioaccumulation and excretion mechanisms of these compounds, is interesting from the viewpoint of health. These studies will inform us about ingredients which do not filter through the skin and can therefore be safely applied.

## Preparation of Samples for Analysis

Sample treatment is a vital and time-consuming part of the analytical procedure, which contributes decisively to the accuracy and precision of the results. For these reasons, the search for the most suitable sample pretreatment is of great interest in the overall analysis.

Sample pretreatment in cosmetic analysis depends on the cosmetic form of the finished products and the analytes to be determined, and the analytical techniques to be used.

### Preparation for the Determination of Elements

The majority of the analytical techniques used for determining elements in cosmetic samples require the dissolution and dilution of samples in an appropriate solvent, with some exceptions such as X-ray fluorescence spectrometry or neutron activation analysis (NAA) which allow, in the case of solid samples, direct measurement. For example, iron and zinc have been determined in compact eye-shadow, face-powder, and rouge by NAA without prior treatment of samples.

However, most cosmetic samples require a pretreatment like complete acid digestion or leaching of the analytes, which may or may not be relatively difficult depending on how efficiently the analytes are extracted from the matrix. Microwave energy permits rapid heating of samples, which considerably reduces pretreatment time. For example, heavy metals have been determined in some cosmetics (lipsticks, powders) using different atomic spectroscopic techniques after acid treatment assisted by microwave irradiation.

Liposoluble samples such as creams or oils can be directly emulsified in water, with the aid of a surfactant, and some metallic elements can be determined by atomic spectrometric techniques without any pretreatment and using aqueous standards. For example, sunscreen creams have been emulsified and zinc (present as zinc oxide) determined by atomic absorption spectrometry.

### Preparation for the Determination of Organic Compounds

Some cosmetic samples, such as creams, emulsions, lotions, lipsticks, etc. can be directly dissolved by the appropriate solvent (e.g., methanol, acetonitrile, dimethylformamide, tetrahydrofuran, ethanol, or mixtures) with the aid of manual or ultrasonic agitation and analyzed by chromatography or related techniques. Sometimes, however, the sample needs cleaning up before chromatographic determination, especially when gas chromatography is involved.

Techniques such as liquid-liquid extraction are usually employed to improve the selectivity of analytical techniques and, specially, for removing the matrix (e.g., antimicrobial ingredients have been extracted from deodorant by liquid-liquid partition with n-hexane/H<sub>2</sub>O before gas chromatographic determination).

Analytes which are present in some difficult to dissolve samples, such as make-up formulations, can be separated from the solid matrix by leaching using the appropriate solvent (e.g., different UV filters have been leached from lipstick samples by ethanol with the aid of ultrasonic irradiation).

Other separation techniques such as solid-phase extraction (SPE) enable the selective retention and elution of both matrix and analytes (e.g., benzophenone-4 and phenylbenzimidazole sulfonic acid have been determined in sunscreen sprays by UV spectrometry using an online SPE sequential injection system).

Supercritical fluid extraction (SFE) has a very good field of application in cosmetics analysis as separation treatment before different analytical techniques (e.g., waxes, UV filters, preservatives, or vitamins have been extracted from different cosmetic matrices).

## Analytical Techniques

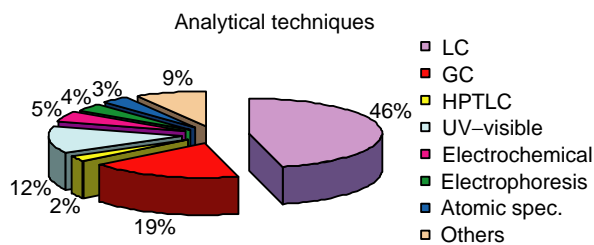
The number and varied chemical nature of the substances that make up cosmetics ingredients have lead to a great variety of methods for cosmetic analysis.

Classical noninstrumental methods are used in routine analysis due to their easily accessible laboratory equipment required for implementation. In addition, they are often the official methods proposed by the corresponding government commissions. Acid-base titrations are used for ingredients that present acid-base properties (e.g., potassium and/or sodium hydroxides), iodometric titrations for the determination of oxidizing agents (e.g., hydrogen peroxide in hair-care products); even gravimetric determinations are carried out (e.g., oxalic acid and/or its alkaline salts by precipitation with calcium or zinc by 8-hydroxyquinoline).

The use of modern instrumental techniques has increased over the last few years and a great number of scientific publications using different techniques (Figure 2) can be found in the cosmetic analysis literature. Chromatographic techniques are the most used because they allow the simultaneous determination of a number of components without the need for a laborious cleanup process.

### Chromatographic and Related Techniques

Classical thin-layer chromatography (TLC) allows identification and semiquantitative determination of multicomponents with very simple laboratory equipment (e.g., certain oxidation coloring in hair dyes by two-dimensional TLC, ionic surfactants in toiletries). This technique is often used as a previous separative step followed by an instrumental determination.



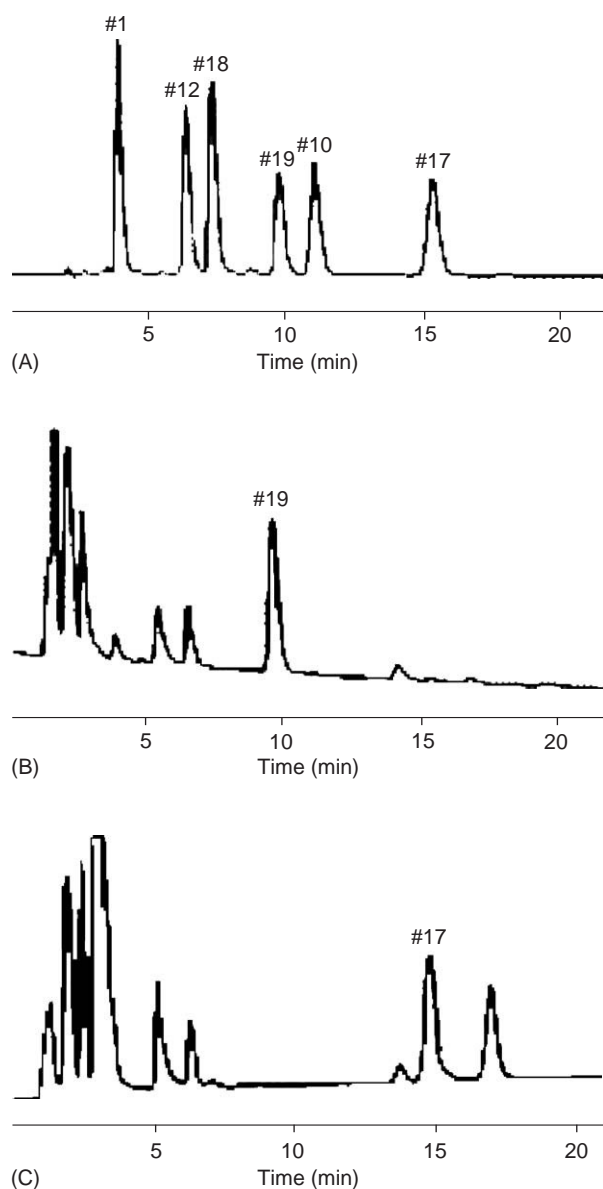
**Figure 2** Percentage distribution of analytical techniques used in cosmetics analysis. (Data obtained from the publications compiled in *Analytical Abstracts* before November 2003.)

High-performance thin-layer chromatography (HPTLC) presents better resolution, sensitivity, and precision than classical TLC, and allows determination of the analytes with sufficient accuracy (e.g., determination of preservatives in creams and lotions); besides this, sample preparation may be simple. However, due to the high versatility of other techniques, this is probably the less commonly used of the chromatographic techniques in this area.

Gas chromatography (GC) has a very good performance, but is limited to volatile and thermally stable analytes. Moreover, sample preparation is often more complex than required by other chromatographic techniques, laborious clean up and/or derivatization steps are generally necessary.

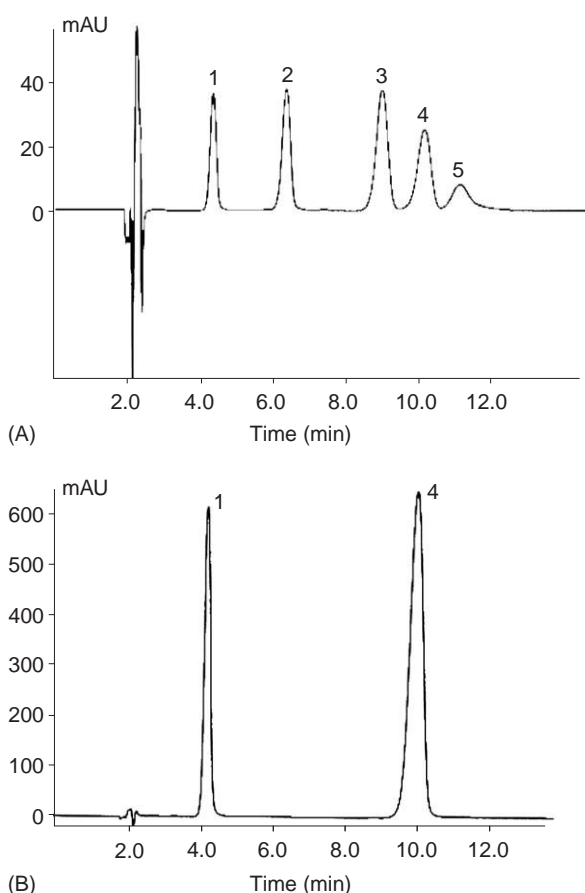
Liquid chromatography (LC) with UV detection is the most commonly used instrumental technique. Fluorescence and electrochemical detectors have also been used. The main advantage of this technique is its versatility. The great variety of separation mechanisms allows determination of any sort of analytes (e.g., anionic surfactants in toiletries by reverse phase ion-pair chromatography, ionic substances by ion-exchange chromatography, UV filters in sunscreens by reverse phase LC). Moreover, switching and multidimensional chromatographic techniques allow on-line sample pretreatment (cleanup, derivatization or trace enrichment) and increased resolution.

As mentioned before, pretreatment of cosmetic samples is often necessary before chromatographic determination. For instance, Figure 3 shows the chromatograms obtained from a solution containing some corticosteroids (forbidden by legislation) standards and from two extracts obtained from commercial cosmetic products. A previous screening of the samples using a TLC method to detect the presence of corticosteroids and a SPE cleanup pretreatment of samples was carried out before analysis. Figure 4 shows the chromatogram obtained in the determination of UV filters in a sunscreen lipstick purified by SFE.



**Figure 3** Chromatograms obtained from: (A) a standard solution containing hydrocortisone-21-acetate #1, betamethasone-17, 21-dipropionate #10, fluocinonide #12, betamethasone-17-valerate #18, and clobetasol propionate #19; (B) and (C) extracts from a shampoo and cream, respectively, obtained by solid-phase extraction. (Reproduced from Gagliardi L, de Orsi D, del Giudice MR, Gatta F, *et al.* (2002) *Analytica Chimica Acta* 457: 187–189.)

Electrophoretic techniques, mainly capillary zone electrophoresis (CZE) and also capillary isotachopheresis (ITP) or micellar electrokinetic chromatography (MEKC) have been used in cosmetic analysis (e.g., determination of cationic surfactants in toiletries, parabens in different cosmetics, fluoride, and polyphosphates in toothpaste, hair dyes, or acid preservatives in cosmetic lotions). However, their use is less extensive than LC, probably because the



**Figure 4** Chromatograms obtained from: (A) a standard mixture containing benzophenone-3 (2-hydroxy-4-methoxybenzophenone or oxybenzone) 1; octyl dimethyl PABA (2-ethylhexyl-*p*-dimethylaminobenzoate) 2; 4-methylbenzylidene camphor 3; octyl methoxycinnamate (2-ethylhexyl-*p*-methoxycinnamate) 4; butyl methoxydibenzoylmethane (4-*tert*-butyl-4'-methoxydibenzoylmethane) 5; (B) an extract from a lipstick, obtained by supercritical fluid extraction. (Reproduced from Scalia S (2000) *Journal of Chromatography A* 870: 199–205.)

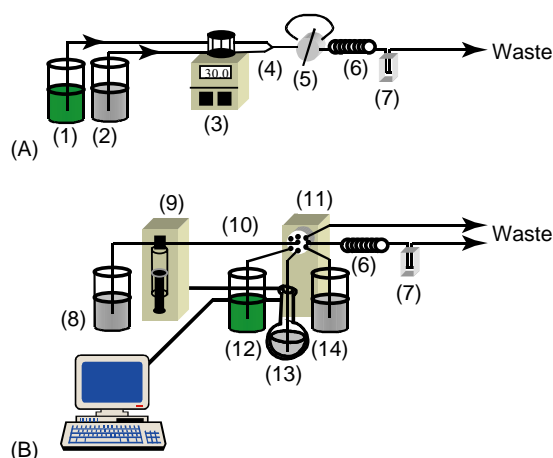
instrumentation is still not so general and moreover LC allows the determination of a great number of analytes whose concentration levels in cosmetic samples can be determined without difficulty.

Hyphenation between GC or LC and mass spectrometry (MS) has improved these techniques, however the high cost of the equipment limits its use in routine quality control. GC–MS has been used to determine different analytes in cosmetics (e.g., essential oils). LC–MS has also been used (e.g., ceramides, cationic and nonionic surfactants, preservatives).

### Molecular Spectrometry

Infrared spectroscopy (both IR and FTIR) has been used in some cases (e.g., anionic surfactants





**Figure 5** Flow injection systems used to determine benzophenone-3 (2-hydroxy-4-methoxybenzophenone or oxybenzone) in sunscreen creams: (A) flow injection analysis (FIA); (B) sequential injection analysis (SIA). Reagents (1), (2), (12), (14); peristaltic pump (3); merging point (4); injection valve (5); reaction coil (6); flow cell (7); ethanol (8); autoburette with syringe (9); holding coil (10); eight-channel selector valve (11); sample or standard solutions (13). (Reproduced from Chisvert A, Salvador A, Pascual-Martí MC, and March JG (2001) *Fresenius Journal of Analytical Chemistry* 369: 684–689.)

determination in shampoos, methanol and ethanol in different cosmetics, glycolic and lactic acids in creams). The low selectivity of UV absorption spectrophotometry is its main disadvantage in this field, which makes inclusion of a cleanup step mandatory most times. Some colorimetric reactions have been proposed to determine different analytes of interest in cosmetics. Luminescence techniques have scarcely been used (e.g., determination of antioxidants or potential nitrosating agents).

Dynamic methods have been proposed for carrying out online UV spectrometric procedures to determine active ingredients such as UV filters in sunscreens, based either on selective SPE of the analytes or on derivatization reactions to obtain colored or chemiluminescent products. Figure 5 shows two flow systems used to determine benzophenone-3 in sunscreens without matrix interferences by reaction with Ni(II) in ammoniacal medium.

### Atomic Spectrometry

Atomic spectrometry has been used to determine metallic components in toothpaste, sunscreen, shampoos, cosmetic pigments, etc. Elements can be found in cosmetics and toiletries as active ingredients (e.g., Se in shampoos) or as impurities from the raw materials. The elements most frequently determined are Pb, Hg, and As; but Se, Cd, Zn, Ti, Fe, Cr, Sr, Bi, Ba,

Mn, Cu, or S have also been determined. Flame atomic absorption spectrometry and electrothermal atomic absorption spectrometry have been the most used techniques. Other techniques used to a lesser extent, are inductively coupled plasma-atomic emission spectrometry (e.g., selenium in shampoos) or atomic fluorescence spectrometry (e.g., determination of mercury by cold vapor).

### Other Techniques

Different analytes are determined by using electrochemical techniques such as differential pulse voltammetry (e.g., metal ions and chlorhexidine in oral care products, glycolic acid in creams, dyes in lipsticks) or potentiometry (e.g., inorganic compounds and anionic and cationic surfactants in personal care products). Modified carbon electrodes and biosensors have been developed to determine some cosmetic ingredients by techniques such as voltammetry or potentiometry.

Other analytical techniques have less frequently been used: nuclear magnetic resonance spectroscopy (NMR) (e.g., tocopherols in toothpaste by hyphenated LC–NMR), energy dispersive X-ray fluorescence (ED-XRF) (e.g., heavy metals determination), surface enhanced Raman scattering (e.g., determination of 4-aminobenzoic acid or PABA, in sunscreens), neutron activation analysis (e.g., determination of iron and zinc), and thermometric analysis (e.g., fluoride in toothpaste).

**See also:** **Essential Oils. Food and Nutritional Analysis:** Antioxidants and Preservatives. **Perfumes. Quality Assurance:** Quality Control. **Surfactants and Detergents.**

### Further Reading

- Barel AO, Paye M, and Maibach HI (eds.) (2001) *Handbook of Cosmetic Science and Technology*. New York: Dekker.
- Bore P (ed.) (1985) *Cosmetic Analysis. Selective Methods and Techniques*. New York: Dekker.
- EU Directorate General Health and Consumer Protection Scientific, Committee on Cosmetic Products and Non-food Products intended for consumers (SCCPNP), [http://europa.eu.int/comm/health/ph\\_risk/committees/sccp/sccp\\_en.htm](http://europa.eu.int/comm/health/ph_risk/committees/sccp/sccp_en.htm)
- Mitsui T (1997) *New Cosmetic Science*. Amsterdam: Elsevier.
- Philippsborn HE (2000) *Elsevier's Dictionary of Cosmetic Science*. Amsterdam: Elsevier.
- US Food and Drug Administration, Center for Food Safety and Applied Nutrition (CFSAN), <http://vm.cfsan.fda.gov/~dms/cos-toc.html>



# COULOMETRY

P C Hauser, The University of Basel, Basel, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Coulometry is an electrochemical method in which the total charge (the number of coulombs) consumed in the redox conversion of an analyte at an electrode is measured. It is not to be confused with colorimetry, the spectroscopic method. Coulometry is distinguished from voltammetric and amperometric methods by not relying on mass transport current control to obtain a signal dependency on concentration. Coulometry is an absolute method, which means that calibration is generally not necessary as electrical charge can be measured with high accuracy. This is an advantage shared with gravimetry. A further advantage is its inherent simplicity and therefore limited expense. The method was very popular around the middle of the twentieth century but has been replaced in many applications by voltammetric methods such as differential pulse polarography or recent nonelectrochemical analytical methods such as high-performance liquid chromatography (HPLC). However, some applications are still current, most notably the important determination of water content by the Karl Fischer method. Interesting new developments with regard to miniaturization and sensitivity have also been reported.

## Fundamentals

The total charge passed between two electrodes in the redox reaction of an analyte is given by the Faraday equation:

$$Q(t) = \int_0^t i(t) dt = zFn = zFcV \quad [1]$$

where  $Q$  is the electrical charge in coulombs,  $i$  the current in amperes,  $t$  the time in seconds,  $z$  the number of electrons exchanged per analyte molecule,  $F$  the Faraday constant ( $96\,487\text{ C mol}^{-1}$ ),  $n$  the number of moles,  $c$  the molar concentration, and  $V$  the volume.

In coulometry, the analyte in the sample volume is exhausted completely. This distinguishes the method from amperometry or voltammetry where the level of current is measured, which is controlled by the concentration through its influence on the rate of

diffusion. This means that in coulometry, in contrast to most analytical methods, it is not the concentration that is determined but the total amount of analyte. From the knowledge of the volume it is, however, possible, of course, to derive the concentration. As electrical charge can be measured with high accuracy and precision, the method is an absolute method, i.e., calibration with standard solutions is not necessary. This is the aspect coulometry has in common with the classical methods of gravimetry and titrimetry and also with electrogravimetry. However, in contrast to electrogravimetry, the other electrochemical bulk method, the product of the redox reaction does not have to form a weighable solid deposit on an electrode. This makes coulometry a much more versatile option.

There are two options. Coulometry can be conducted in the constant current and the constant potential modes. The former is inherently simpler as in this case the total charge is obtained directly from the measured time to the completion of the reaction. However, it can only be used successfully if there is only one redox-active species present in the sample or if at least the redox potentials of species present are succinctly different. In constant current coulometry, the cell voltage needs to follow the depletion of the concentration of the analyte according to Nernst's law (for an oxidation):

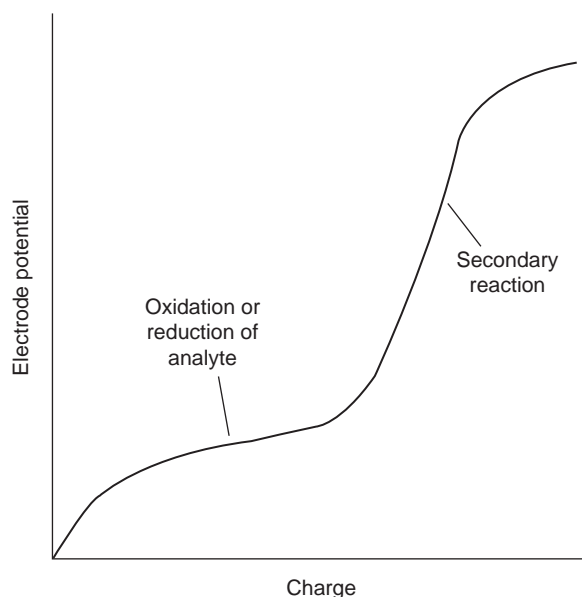
$$E = E_F^\circ + \frac{RT}{nF} \ln \frac{x}{1-x} \quad [2]$$

where  $E$  is the electrode potential,  $E_F^\circ$  the standard formal potential for the species being oxidized,  $R$  the universal gas constant,  $T$  the absolute temperature in Kelvin, and  $n$  the number of electrons exchanged for each molecule;  $x$  is the fraction of analyte oxidized.

From an examination of eqn [2] it follows that, for example, for 99.9% completion of the reaction, the applied voltage has to be  $\sim 180\text{ mV}$  more positive than at the start (for  $n = 1$ ).

In order to continue to pass the current through the cell, eventually an alternate reaction has to set in, such as the decomposition of the solvent or the electrode material. This leads to the voltage behavior shown in Figure 1.

Constant potential coulometry allows better control as electrode reactions are always governed by the electrode potential. Interfering reactions can be avoided as long as their redox potentials are sufficiently distinct. It may also be possible to



**Figure 1** Potential change during constant current coulometry. The secondary reaction is due to oxidation or reduction of a further electroactive species in the sample or decomposition of the electrode or solvent.

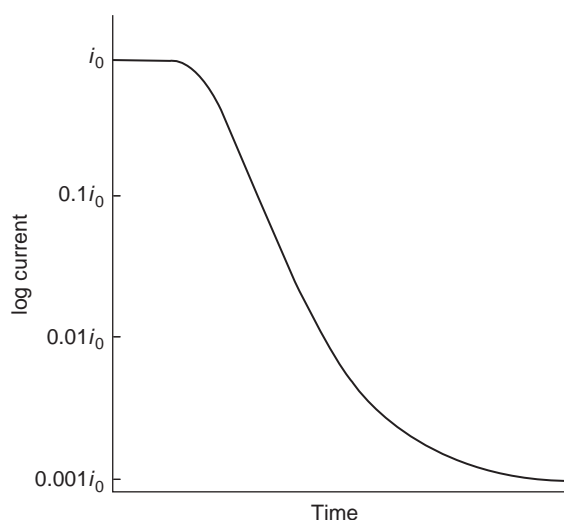
quantitatively remove an interfering species before the determination of the analyte, a process termed electroseparation, or to sequentially determine several analytes. In constant potential coulometry the current,  $i_t$ , decays with time according to eqn [3]:

$$i_t = i_0 \cdot 10^{-kt} \quad [3]$$

$i_0$  is the initial current,  $t$  the time, and  $k$  a constant dependent on such parameters as the electrode size and the diffusion coefficient of the species analyzed (see Figure 2).

It is therefore necessary to integrate the area under the curve to obtain the charge. Today this is readily achieved by electronic or computational methods and is not a hindrance, and precisions of better than 0.1% can be achieved. Such levels of precision are scarcely attainable with other instrumental methods.

A possible side reaction to be avoided is the decomposition of the solvent, which in most cases is water. Therefore, it has to be made sure that the pH-dependent potential stability boundaries of water are not exceeded in order to assure 100% current efficiency for the analytical process. For reductions it is often beneficial, as it is in other electrochemical methods, to use mercury electrodes because the reduction of water, i.e., the evolution of hydrogen, is kinetically hindered on this material. Therefore, the reduction of species such as metals with redox potentials more negative than that of hydrogen is enabled. For anodic processes, i.e., oxidations, the



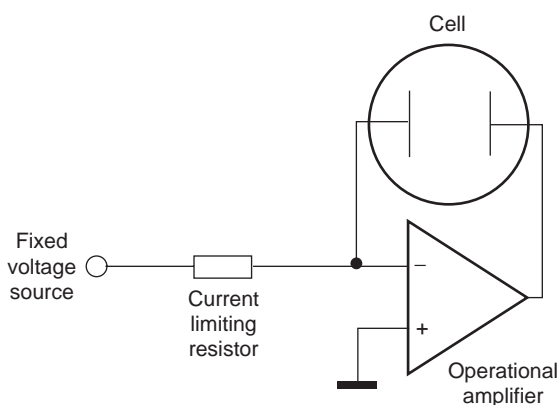
**Figure 2** Current decay on a logarithmic scale during constant potential coulometry.

dissolution of the electrode material is also a potential interfering reaction. Mercury is not suitable for such applications and glassy carbon should be used instead. For reductions, ambient oxygen dissolved in the sample solution will interfere by its reduction to hydrogen peroxide. Deoxygenation by passing a stream of nitrogen or other inert gas through the solution before analysis might therefore be necessary. Attention also has to be paid to potential secondary interfering reactions. It is, for example, possible to reduce  $\text{Cr}^{3+}$  to  $\text{Cr}^{2+}$ , but  $\text{Cr}^{2+}$  will react with protons in acidic solutions to yield again  $\text{Cr}^{3+}$  as product, thus lowering the Faradaic efficiency to below 100%. For very small analytical concentrations, non-Faradaic, capacitive charging currents at the working electrode will not be negligible and limit the lowest amount of charge measurable with adequate precision to  $\sim 10 \mu\text{C}$ . This corresponds to an ultimate mass detection limit of  $\sim 0.1 \text{ nmol}$  or  $\sim 1 \text{ ng}$  of analyte.

## Instrumentation

### Constant Current Coulometry

A constant current source can be approximated by simply using a voltage source of a relatively high level in series with a current control resistor. However, changes in cell voltage will cause an error in current in this arrangement (which is the smaller the higher the source voltage is) and therefore it is better to use a galvanostat. Stand-alone galvanostats are not widely available commercially these days, but some commercial potentiostats feature the galvanostatic mode as an option. Most commercial



**Figure 3** Standard galvanostat circuitry for constant current coulometry consisting of a voltage source, a resistor, and an operational amplifier.

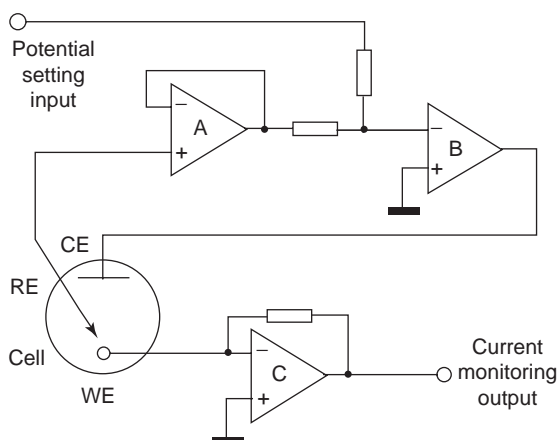
potentiostats (those using a standard internal electronic circuitry) without provision for galvanostatic experiments may be forced to work in the galvanostatic mode by inserting a resistor between the inputs for the working and reference electrodes and then connecting the working electrode to the reference input. A reference electrode is then not used. The current forced through the cell is determined by the potential set at the potentiostat and the resistor according to Ohm's law. It is also possible to construct a galvanostat that is perfectly adequate for the small currents used in analysis relatively easily and at low cost using a single operational amplifier as illustrated in Figure 3.

### Constant Potential Coulometry

Constant potential coulometry requires the presence of a reference electrode in solution, besides working and counter-electrodes, and the use of a potentiostat in order to accurately set the applied potential (see Figure 4).

## Methods

It is possible to distinguish between direct and indirect, or primary and secondary, coulometry, denoting methods in which either the analyte is oxidized or reduced directly at the working electrode, or in which a substance is generated at the working electrode which then reacts with the analyte. The latter procedure is also termed coulometric titration considering that the working electrode serves for the production of a titrant. Note that sometimes direct coulometry has also been termed a titration, a view in which electrons are seen as reagent. Coulometry may also be employed in a flow-through mode, which has found special applications.



**Figure 4** Basic potentiostat circuitry for constant potential coulometry. Amplifier A allows the current-free monitoring of the potential of the reference electrode RE, amplifier B regulates the potential at the working electrode WE, and amplifier C converts the current passing through the cell CE into a proportional voltage.

### Direct Coulometry

The analyte is directly oxidized or reduced at a working electrode and the charge required is determined. This may be carried out in either the constant current or the constant potential mode, but the latter is preferable because of the better control of side reactions. The completeness of the reaction is then indicated by the decay of the current toward zero. The reaction is stopped when the current has decayed to a predetermined fraction of its initial value such as 1 or 0.1%, assuming that this will give ~99 or 99.9% accuracy. However, it is often difficult to set a potential that assures complete reaction of the analyte whilst avoiding side reactions such as the decomposition of solvent. Stirring of the solution is required for enhancement of the transport of the analyte to the electrode but the current is still predominantly controlled by the diffusion through the stagnant non-stirred layer adhering to the electrode. For low concentrations the diffusion rate is low and therefore the current small (in the constant current mode this has to be set low). This means that the method tends to be slow. To overcome some of these difficulties, often a mediator is added that carries the current when the concentration of the analyte becomes low. This mediator then reacts in turn with the analyte so that the condition of 100% current efficiency is maintained. For example, iron may be determined by oxidation of  $\text{Fe}^{2+}$  to  $\text{Fe}^{3+}$  at a working electrode. However, as the concentration of  $\text{Fe}^{2+}$  decreases, potentials at which water is oxidized are required to complete the iron determination. This side reaction can be avoided if  $\text{Ce}^{3+}$  is added to the solution.

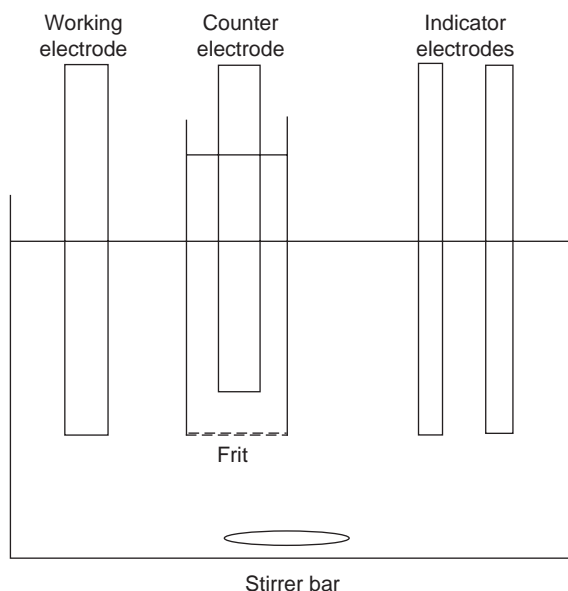
$\text{Ce}^{3+}$  is readily oxidized to  $\text{Ce}^{4+}$  at modest anodic potential where water oxidation is not yet taking place and this reaction will set in once  $\text{Fe}^{2+}$  has been depleted to some extent. However,  $\text{Ce}^{4+}$  readily oxidizes available  $\text{Fe}^{2+}$  to  $\text{Fe}^{3+}$  so that the remaining iron is determined indirectly. This approach should not be confused with coulometric titrations described below.

### Coulometric Titrations

In indirect coulometry, an intermediate reagent, or titrant, is produced by the electrode reaction and the analyte never reacts directly at the working electrode. Titrations are more widely used than direct methods because they overcome the disadvantages of the latter. The starting material for electrochemical generation of the reagent can be provided at sufficient concentration to sustain a high current so that the determination can be carried out more rapidly. This also minimizes the chance of side reactions and the constant current mode can be adopted. One disadvantage compared to direct coulometry is that both reactions, the generating electrode reaction and the subsequent reaction with the analyte, need to be 100% efficient. On the other hand, the method is much more flexible as it is possible to generate titrants by electrode reactions not only for redox titrations but also for the other three types of titration. In coulometric titrations, a method of endpoint detection is needed, just as in conventional volumetric titrations. Preferentially an instrumental method is adopted. In comparison to conventional titrations the coulometric mode has several advantages as long as a suitable generating reaction is available. Complete automation is simpler. The expense of a precise mechanical automatic burette is avoided. The use of unstable reagents is possible as these are produced *in situ* and species not normally used for conventional titrations such as  $\text{Cl}_2$ ,  $\text{Br}_2$ ,  $\text{Ag}^{2+}$ , or  $\text{Cu}^+$  are feasible. Calibration of a reagent solution is not necessary. The sensitivity is also higher than for classical titrations as smaller amounts of reagent can be delivered with accuracy than dispensed volumetrically. This is illustrated by the fact that a current of  $10\text{ }\mu\text{A}$  flowing over a period of 10 s corresponds to only about  $10^{-9}$  mol of analyte. Measuring currents and times of that order accurately is readily achieved with modern instrumentation.

A cell for coulometric titration is illustrated in Figure 5.

The counter-electrode reaction often interferes and for this reason this electrode is usually physically separated from the sample solution by using a diaphragm or frit to create a liquid junction. External



**Figure 5** Cell arrangement for coulometric titration in the constant current mode. The counter-electrode is isolated from the sample solution by a liquid junction formed, for example, with a glass frit. The indicator electrodes may belong to a potentiometric, amperometric, or conductometric detection system.

generation is also possible by employing a flow-through arrangement that delivers the desired product from the electrode to the sample.

All of the four types of titrations have been implemented coulometrically (i.e., acid-base, precipitation, complexometric, and redox titrations). Acid-base titrations are achieved by generating protons or hydroxide ions from the solvent water by electrolysis (the hydrogen and oxygen evolution reactions). A list of possible methods is given in Table 1.

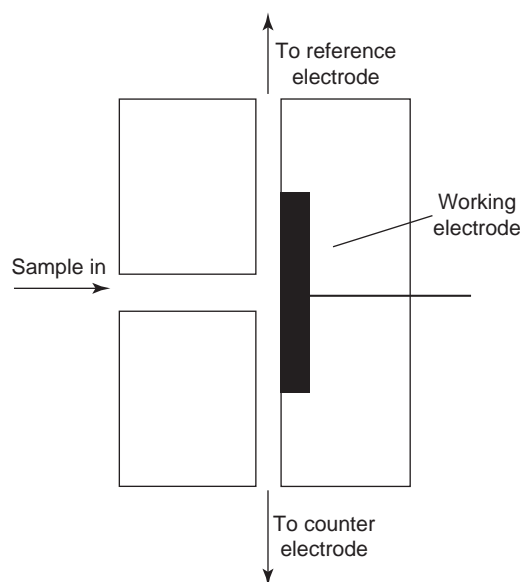
The means of detecting the endpoint will be dictated by the type of reaction employed. Acid-base titrations are most easily followed using a glass pH electrode while redox reactions lend themselves to amperometric detection (only a small fraction of the species detected is consumed at the indicator electrode). Other options are ion-selective electrodes and conductometric detection.

### Coulometric Detection in Flow Methods

Coulometry may be employed for detection in flow-injection analysis (FIA), an approach to automation of sample handling and processing, or in separation methods such as HPLC. Note that these methods are closely related to amperometric detection. The distinction is in the degree at which the analyte is converted. A high Faradaic efficiency is desirable in any case and 100% efficiency might actually be

**Table 1** Examples of coulometric titrations

Reagent	Generating reaction	Analytes determined
<i>Redox titrations</i>		
Cl <sub>2</sub>	$2\text{Cl}^- \rightarrow \text{Cl}_2 + 2\text{e}^-$	As <sup>3+</sup> , degree of unsaturation in fatty acids
Br <sub>2</sub>	$2\text{Br}^- \rightarrow \text{Br}_2 + 2\text{e}^-$	As <sup>3+</sup> , Sb <sup>3+</sup> , U <sup>4+</sup> , SCN <sup>-</sup> , NH <sub>3</sub> , phenols, alkenes
I <sub>2</sub>	$2\text{I}^- \rightarrow \text{I}_2 + 2\text{e}^-$	As <sup>3+</sup> , S <sub>2</sub> O <sub>3</sub> <sup>2-</sup> , H <sub>2</sub> S
Ce <sup>4+</sup>	$\text{Ce}^{3+} \rightarrow \text{Ce}^{4+} + \text{e}^-$	Fe <sup>2+</sup> , Ti <sup>3+</sup> , As <sup>3+</sup>
Cu <sup>+</sup>	$\text{Cu}^{2+} + \text{e}^- \rightarrow \text{Cu}^+$	Cr <sub>2</sub> O <sub>7</sub> <sup>2-</sup> , IO <sub>3</sub> <sup>-</sup>
<i>Acid–base titrations</i>		
H <sup>+</sup>	$2\text{H}_2\text{O} \rightarrow 4\text{H}^+ + \text{O}_2 + 4\text{e}^-$	Bases
OH <sup>-</sup>	$2\text{H}_2\text{O} + 2\text{e}^- \rightarrow 2\text{OH}^- + \text{H}_2$	Acids
<i>Precipitation titration</i>		
Ag <sup>+</sup>	$\text{Ag} \rightarrow \text{Ag}^+ + \text{e}^-$	Halides, CN <sup>-</sup> , SCN <sup>-</sup> , S <sup>2-</sup> , RSH
<i>Complexometric titration</i>		
EDTA	$\text{HgNH}_3(\text{EDTA})^{2-} + \text{NH}_4^+ + 2\text{e}^- \rightarrow \text{Hg} + 2\text{NH}_3 + \text{H}(\text{EDTA})^{3-}$	Ca <sup>2+</sup> , Zn <sup>2+</sup> , Pb <sup>2+</sup> , Cu <sup>2+</sup>

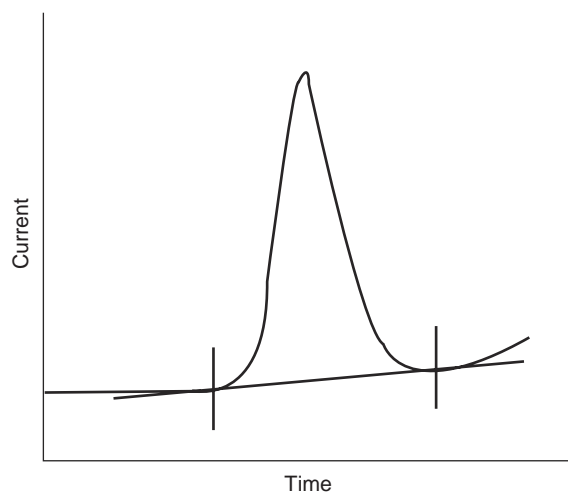
**Figure 6** Wall jet electrode for flow through coulometry.

achieved in methods nominally termed amperometric.

The wall jet cell arrangement commonly used in flow-through detection is illustrated in **Figure 6**.

Complete analyte reaction is assured by designing the cell in such a way that the ratio of electrode area to cell volume is large, so that each analyte molecule can reach the electrode during its residence time in the cell.

Applications are mostly based on oxidations, as the complete removal of oxygen needed for reduction reactions is difficult to achieve in flow-through systems. However, the flow-through approach elegantly allows discrimination against residual currents, due to electrode charging or Faradaic reactions of impurities, by enabling the measurement of transient

**Figure 7** Current peak obtained in flow through methods. The charge is obtained by integration of the peak area on top of a baseline current between the limits indicated.

coulometric peaks on top of baseline currents (see **Figure 7**).

FIA with coulometric detection is suitable when only one or few electroactive species are present and potential control assures selectivity. A typical application for an FIA system with coulometric detection is the determination of SO<sub>2</sub> in wine.

For more complex samples with a number of electroactive species to be determined, separation by HPLC, or other methods such as ion chromatography or capillary electrophoresis, followed by coulometric detection is better suited. Its applications in HPLC are usually to oxidizable organic species that cannot be determined by ultraviolet absorption, the standard detection technique in HPLC. Examples for such species include amines and phenols, catecholamines (such as the neurotransmitters adrenaline and



dopamine), vitamin A, carbohydrates, ketones, and nitrocompounds. The sensitivity of coulometric detection in HPLC is high; detection limits down to  $\sim 10^{-8} \text{ mol l}^{-1}$  have been achieved. For certain applications it is necessary to adopt pulsing methods in which the electrode is continuously cleaned electrochemically by applying an oxidizing voltage followed by a reducing voltage before applying an intermediate working potential. This is necessary if the oxidation products of the analyte lead to poisoning of the electrode. Indirect coulometric detection for HPLC, in which a reactive intermediate is formed at the working electrode, similarly to coulometric titration, has also been reported. Postcolumn reaction of analytes that are not natively electroactive before detection may also be adopted, for example, for the determination of amino acids.

Flow-through methods are also used for sample treatment prior to coulometric titration, for the purpose of cleanup or preconcentration on an ion-exchange column or for reduction to a desired oxidation state in reductor columns. Preadjustment of redox states or removal of interferents may also be carried out by an auxiliary electrochemical flow-through cell located ahead of the analytical cell.

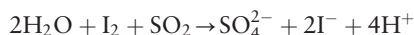
### Stripping Coulometry

Sensitive determination of metal ions is possible via coulometry by using preconcentration. Thereby the metals are first deposited by reduction onto a mercury drop or film or a porous carbon electrode. In a second step the metals are stripped by oxidation and the resulting charge is measured in the form of a peak allowing discrimination against background currents. Deoxygenation of the sample is generally not necessary in this procedure. Calibration-free determinations of heavy metals following this procedure at concentrations as low as 1 ppb have been reported.

### Determination of Water by the Karl Fischer Titration

A very common application of coulometry is in the determination of water by the Karl Fischer method.

Water reacts with iodine and sulfur dioxide according to the following reaction:



In order to achieve complete reaction, pyridine is added to bind the protons produced in the reaction. Pyridine also increases the solubility of iodine and sulfur dioxide by complex formation. The reagents are dissolved, and the reaction is carried out, in methanol or ethylene glycol monomethyl ether as

solvent to allow polar substances from the sample to be dissolved. Because of the toxicity of pyridine, this reagent is now often replaced with other bases, such as diethanolamine or imidazole, or salts of weak organic acids such as sodium salicylate.

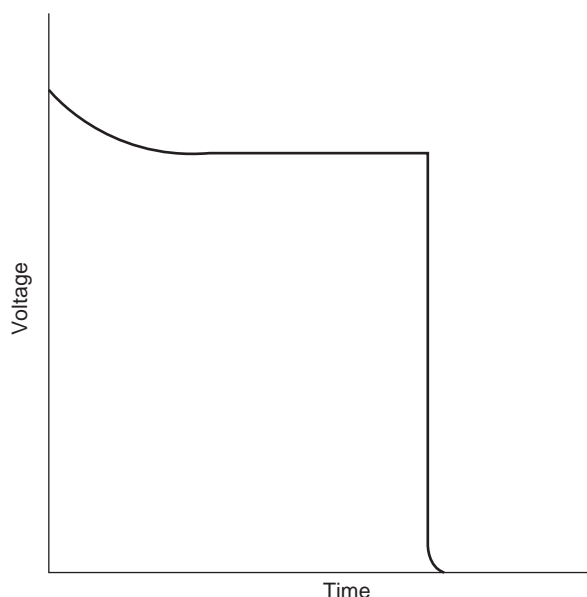
The Karl Fischer titration may be carried out volumetrically by using a mixture of the three compounds in a mole ratio of 1:3:10 (iodine, sulfur dioxide, pyridine), in which sulfur dioxide is present in excess. The reagent mixture is commercially available or readily made. In the coulometric Karl Fischer method, which has the usual advantages of coulometric titrations of easier automation and reagent handling as well as higher sensitivity, iodine is generated anodically from iodide in the reagent solution by the application of a fixed current of typically  $10 \mu\text{A}$  to two inert platinum electrodes in solution. Before the endpoint is reached, the current is maintained by oxidation of iodide produced in the titration reaction on the anode and the reduction of  $\text{H}^+$  on the cathode. The required voltage between the electrodes is  $\sim 300 \text{ mV}$  and corresponds roughly to the difference in standard potentials for the two redox reactions. After the endpoint, excess iodine is present in solution and the reduction of this species can carry the cathode current. As the iodine/iodide redox couple is well reversible, the cell voltage drops rapidly to a few millivolts only giving a very sharp indication of the endpoint. This mode of endpoint detection is also used in volumetric Karl Fischer titration. It is then usually termed biamperometric detection, although it still involves a constant current approach, because it is based on two electrode reactions of interest. It is then the oxidation of iodide produced in the titration reaction that carries the current (see Figure 8).

Karl Fischer titrations are frequently used for the analysis of water content in such samples as food materials, pharmaceuticals, and solvents used in industrial applications, or for the determination of hydration water in crystals. Dedicated units are commercially available and many automatic titrators or pH meters feature a connector for current biased voltage measurements labeled 'Karl-Fischer' or simply 'KF'.

### Layer Thickness

The layer of an easily oxidizable surface layer of metal may be dissolved anodically under potential control. If the area of the sample surface is known the layer thickness can be deduced from the total charge needed for complete stripping. A typical application of this procedure is the analysis of a layer of anticorrosive tin on steel.





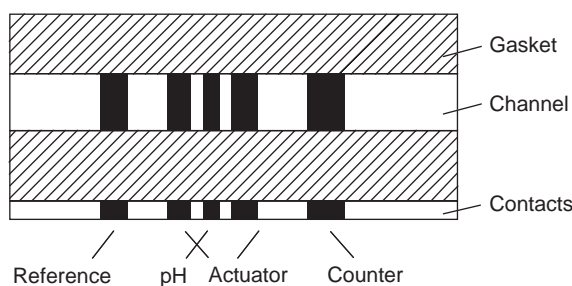
**Figure 8** Voltage drop obtained in coulometric Karl Fischer titrations indicating the endpoint.

### Determination of Gaseous Species

Coulometry in different variants is also used for the determination of gaseous species. Carbon dioxide or sulfur dioxide dissolved in beverages may be determined by sweeping these dissolved gases with nitrogen from a sample into an absorbing solution followed by coulometric titration in a batch or flow-through system. Samples may also be combusted with oxygen in a closed system and the resultant gases titrated. This approach is mainly used for the determination of total carbon content. Another application is the determination of the amount of chlorinated hydrocarbons present in a sample by determining hydrochloric acid liberated in the combustion.

### Coulometric Microtitrators

A relatively recent development has been the construction of microscale coulometric acid–base titrators using microlithographic techniques. One or more inert working electrodes (typically made from gold) are placed at a distance of typically 50  $\mu\text{m}$  from a solid-state pH sensor typically in the form of an ion-sensitive field effect transistor (ISFET). The working electrode, the actuator, is used to either generate protons or hydroxide ions by electrolysis of water at constant current. The thus created reagent diffuses away from the generator electrode in all directions but is consumed by the analyte. Therefore, a pH gradient moving outward results. This is defined by the rate of reagent generation but delayed by the buffering action of the analyte in dependence of its concentration. The pH-ISFET serves to sense this



**Figure 9** Possible arrangement of electrodes for a channel-based coulometric microtitrator. Two actuator electrodes at which protons or hydroxide ions are produced by electrolysis of water encompass a sensing solid-state pH electrode. The reference electrode belongs to the potentiometric pH electrode while the counter-electrode carries the actuator current and is removed from the working electrode to prevent interference.

gradient and the signal indicates the endpoint. The counter-electrode, where the complementary, but interfering, water electrolysis reaction is taking place is sufficiently far removed from the working electrode so that the titration is completed before the cloud of acid or base reaches the detection zone. The small dimensions of the entire assembly allow the placement of all electrodes at the tip of a probe the size of a conventional pH electrode. As the diffusion across the spacing between the working electrode and the pH sensor is very fast, the entire titration is completed in a few seconds. In contrast to conventional coulometric titration only a negligibly small fraction of the analyte is consumed. The analysis may be carried out repeatedly in the same sample. The sample is not stirred during the titration in order not to perturb the diffusion process. It is, however, necessary to carry out a calibration of the system. Coulometry carried out in this fashion is not an absolute method as diffusion is introduced as a variable parameter (see Figure 9).

See also: **Amperometry. Sensors: Overview. Titrimetry: Overview.**

### Further Reading

- Bard AJ and Faulkner LR (2001) *Electrochemical Methods: Fundamentals and Applications*, 2nd edn. New York: Wiley.
- Beinrohr E (2001) Flow-through coulometry as a calibrationless method in inorganic trace analysis. *Accredited Quality Assurance* 6: 321–324.
- Kissinger PT and Heineman WR (eds.) (1996) *Laboratory Techniques in Electroanalytical Chemistry*, 2nd edn. New York: Dekker.
- Olthuis W, Luo J, van der Schoot BH, Bomer JG, and Bergveld P (1990) Dynamic behaviour of ISFET-based sensor–actuator systems. *Sensors and Actuators B* 1: 416–420.

# COUNTERCURRENT CHROMATOGRAPHY

Contents

## Overview

### Solvent Extraction with a Helical Column

## Overview

**Y Ito**, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Countercurrent chromatography (CCC) is a type of liquid partition chromatography and has the distinct advantage of operation without the use of a solid support. Since no solid support is used, complications arising from the use of a solid support, such as adsorptive sample loss and denaturation, tailing of solute peaks, and contamination, can be avoided in CCC. Different from most types of liquid chromatography, CCC utilizes two immiscible solvent phases and the partition process takes place in an open column space where one phase (stationary phase) is retained and the other phase (mobile phase) continuously passes through it. In order to retain the optimum amount of the stationary phase in the column, the system uses appropriate combinations of various column configurations and the applied force field (gravitational or centrifugal). Hence, the CCC instruments display a variety of forms that are quite different from those used in conventional liquid chromatography.

All the existing CCC systems may be divided into two classes: hydrostatic equilibrium systems and hydrodynamic equilibrium systems. The hydrostatic system uses a stable force field to retain the stationary phase in the column while the mobile phase is flowing through the column. The CCC systems that belong to this category are helix CCC (toroidal coil CCC), droplet CCC, locular CCC, and centrifugal partition chromatography (CPC). On the other hand, the hydrodynamic system employs an Archimedean screw effect that facilitates constant mixing of the two phases while retaining one of the phases as a stationary phase. It includes high-speed CCC, nonsynchronous CCC, and slow rotary CCC.

## Hydrostatic CCC Systems

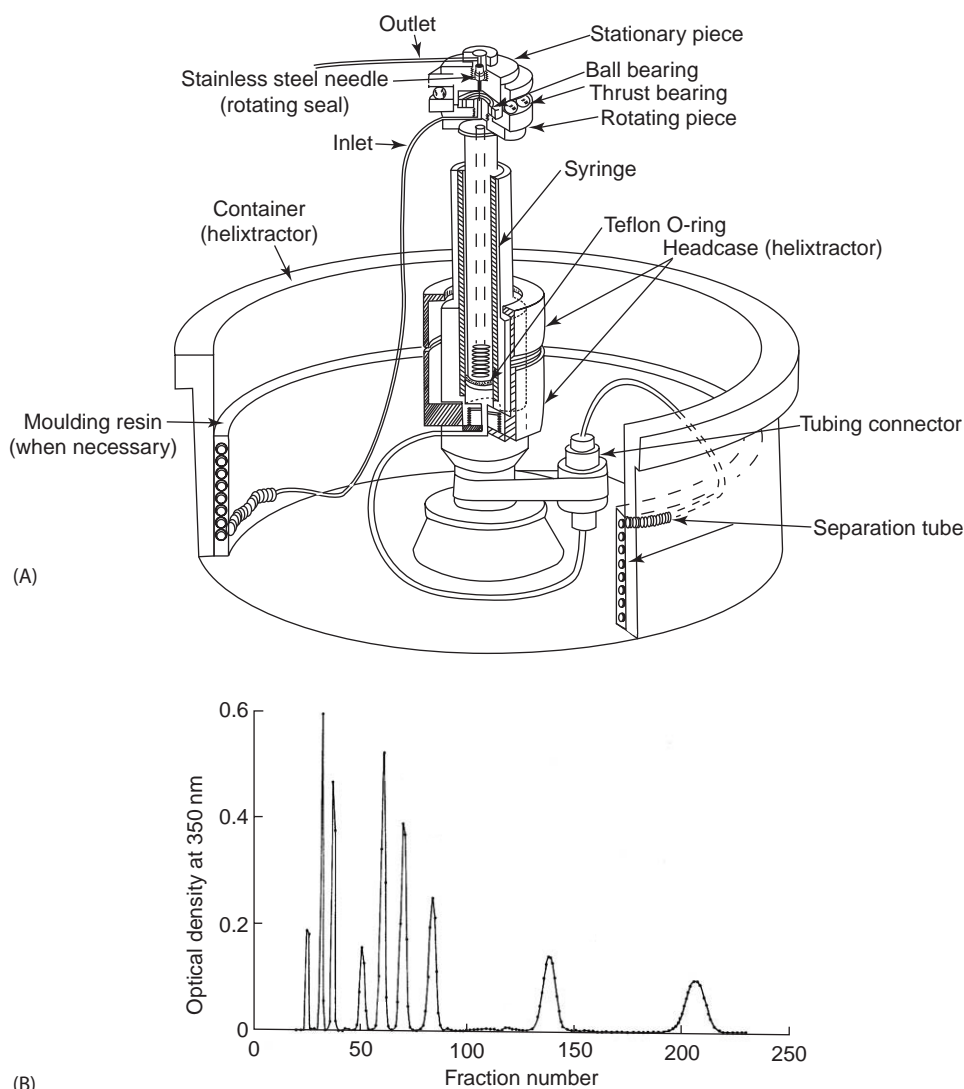
### Helix CCC (Toroidal Coil CCC)

This is the original CCC system that was given the name ‘countercurrent chromatography’ in 1970. It uses a long, fine coiled tube on a flexible core (6 mm OD) that is again coiled around the periphery of a centrifuge bowl (toroidal coil). Rotation of the bowl creates a strong centrifugal force field that retains the stationary phase, either lighter or heavier phase in each helical turn. The mobile phase is introduced into the inlet of the coiled column while the effluent is continuously monitored and collected into test tubes as in liquid chromatography. The original model (Figure 1A) was equipped with a rotating syringe to feed the mobile phase while the effluent from the outlet of the column was collected through a rotary seal. Later, this design was replaced by a seal-free flow-through system to eliminate the risk of leakage of the solvent and cross-contamination at the rotary seal. The system yields a high partition efficiency of several thousand theoretical plates as demonstrated by the separation of DNP (dinitrophenyl)-amino acid (Figure 1B). The partition efficiency of the column increases with the number of helical turns and/or when the core is made smaller and the inner diameters of the coiled tube are reduced. The limitation is the hydrostatic pressure accumulated in the column, which eventually breaks the tubing. This hydrostatic pressure ( $P$ ) is expressed by the following formula:

$$P = R\omega^2(\rho_1 - \rho_2)nd \quad [1]$$

where  $R$  denotes the distance between the center of rotation and the axis of the helix,  $\omega$  the angular velocity of rotation,  $n$  the number of coil units,  $d$  the distance between a pair of interfaces in each turn (less than the core OD), and  $\rho_1$  and  $\rho_2$ , the densities of heavier and lighter phases, respectively.

This analytical CCC method also requires careful optimization of the flow rate, which maximizes the partition efficiency and the retention of the stationary phase. A higher flow rate will result in higher column pressure and lower partition efficiency with the loss of the stationary phase from the column, whereas a lower flow rate requires a longer elution time.

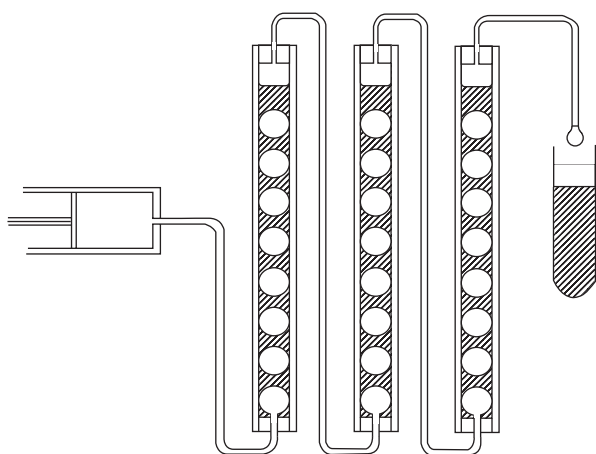


**Figure 1** Design and performance of the helix CCC apparatus: (A) Original design of the centrifuge head of helix CCC; and (B) DNP (dinitrophenyl) amino acid separation by helix CCC. Peaks identified in order of elution and their partition coefficient ( $C_{\text{upper phase}}/C_{\text{lower phase}}$ ) from left to right are: *N*-DNP- $\delta$ -L-ornithine (>100); *N*-DNP-L-aspartic acid (3.8); *N*-DNP-D,L-glutamic acid (1.9); *N,N*-di-DNP-L-cystine (0.94); *N*-DNP- $\beta$ -alanine (0.71); *N*-DNP-L-alanine (0.56); *N*-DNP-L-proline (0.45); *N*-DNP-L-valine (0.26); and *N*-DNP-L-leucine (0.18). Solvent system: chloroform/acetic acid/0.1 mol l<sup>-1</sup> HCl (2:2:1), upper aqueous phase mobile at 125  $\mu$ l h<sup>-1</sup>. The total elution time: ~40 h.

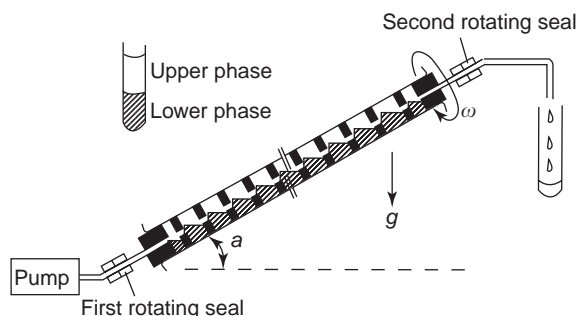
### Droplet CCC

This simple preparative CCC system grew out of the observation that the light phase with a low wall affinity formed discrete droplets that rose through the heavy phase with a visible evidence of very active interfacial motion. Under ideal conditions, each droplet could become a plate if kept more or less discrete throughout the system. The actual system was made of long vertical columns of narrow bore silanized-glass tubing with fine Teflon tubes to interconnect the wider bore glass tubes. Discrete droplets, formed at the tips of the finer tube inserted into the bottom of the long glass tube, were made to

follow one another with minimal space between them and at a diameter close to that of the internal bore of the column (Figure 2). These droplets divide the column into discrete segments that prevent longitudinal sample band spreading along the length of the column as they mix locally near the equilibrium. The fine Teflon tubing interconnecting the individual columns preserves the integrity of the partitioning with minimum longitudinal diffusion and helps in the formation of new droplets at the bottom of the next column. This regularity of droplet size and close spacing, a droplet size nearly filling the bore, and thin return tubing are important for best results.



**Figure 2** Droplet CCC system. Droplets formed at each column junction can travel through the column a rate of  $\sim 2 \text{ cm s}^{-1}$  without coalescence.

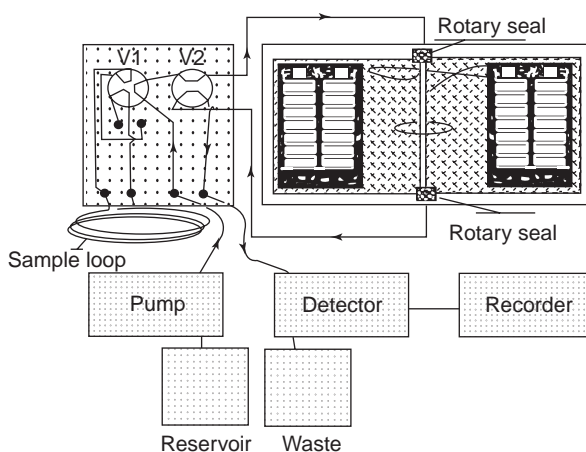


**Figure 3** Mechanism of rotation locular CCC apparatus.

The first prototype column assembly consisted of 300 glass tubes (60 cm long and 1.8 mm ID) and had a total capacity of 540 ml including  $\sim 16\%$  dead space in connections. Using a two-phase solvent system composed of chloroform, acetic acid, and 0.1 N HCl (2:2:1), several hundred milligrams of DNP-amino acids were separated within 80 h with an efficiency of 900 theoretical plates (TPs).

### Locular CCC

This CCC system was developed to compensate for the limitations of droplet CCC, which requires adequate droplet formation. **Figure 3** illustrates the column design and mechanism of rotation locular CCC. The column is made by inserting centrally perforated disks into the tubular column at regular intervals to form multiple compartments called 'locules'. Both retention of the stationary phase and interfacial area in each locule are optimized by inclination of the column while the mixing of the two phases is introduced by the rotation of the column.



**Figure 4** Schematic view of the centrifugal partition chromatograph.

The system permits universal application of the conventional two-phase solvent systems with either upper or lower phase as the mobile phase. Necessity of rotary seals at each terminal may be eliminated by gyration locular CCC in which the locular column is held vertical and gyrated to introduce circular motion of the two phases and their interface within each locule.

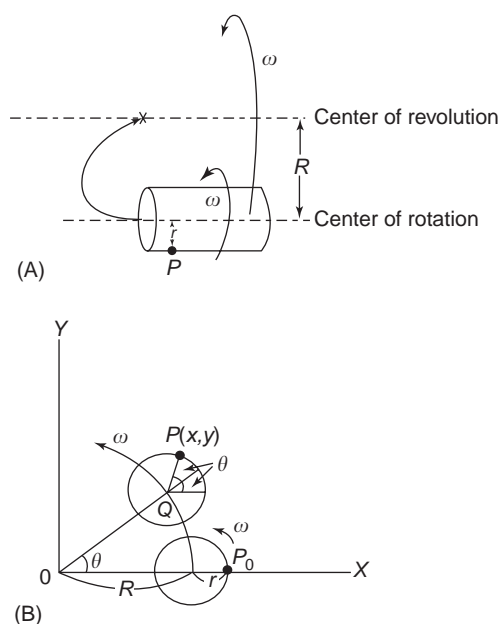
### Centrifugal Partition Chromatography

This CCC system is considered to be a hybrid of helix CCC and locular CCC. The doughnut-shaped separation column consists of a set of disks, each containing a number of small partition compartments interconnected with fine transfer ducts (**Figure 4**). Several units of these disks are stacked to form a column assembly. The whole column assembly is rotated with a flow-through centrifuge system equipped with a pair of rotary seals that can tolerate pressure up to  $60 \text{ kg cm}^{-2}$  ( $\sim 800 \text{ psi}$ ), and the whole system is computerized, as in liquid chromatography.

### Hydrodynamic CCC Systems

All hydrodynamic CCC systems utilize an Archimedean screw force that is created by coil rotation in the force field: some use low-speed rotation of the coil in unit gravity while others subject the coil to a high-speed planetary motion. An Archimedean force drives all objects in the coil toward one end of the coil, which is called the head; the other end is called the tail.

Since the low-speed rotary system is elsewhere in the encyclopedia, planetary centrifuge systems are reviewed here. All the centrifuge systems described



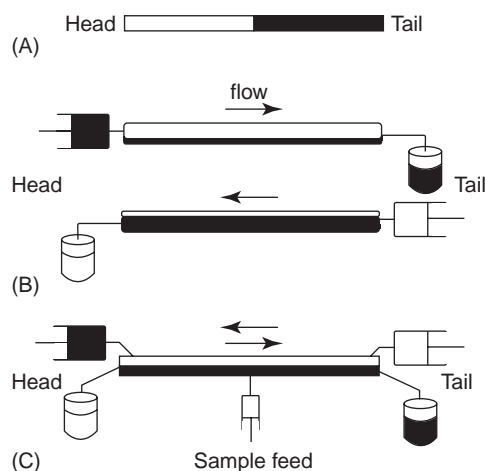
**Figure 5** Type-J synchronous planetary motion and resulting force field: (A) planetary motion; (B) coordinate system for computation of the acceleration field; and (C) centrifugal force distribution.

below use the seal-free flow-through mechanism to minimize the risk of leakage and cross-contamination.

### Type-J Multilayer Coil Planet Centrifuge

Figure 5 illustrates the type-J synchronous planetary motion of the coil holder and the resulting centrifugal force field. When the tubing is coaxially wound around the holder, the centrifugal force quickly separates the two-phase solvent system along the coil in such a way that one phase occupies the head side and the other the tail side.

This hydrodynamic behavior of the two phases can be effectively utilized for performing CCC as

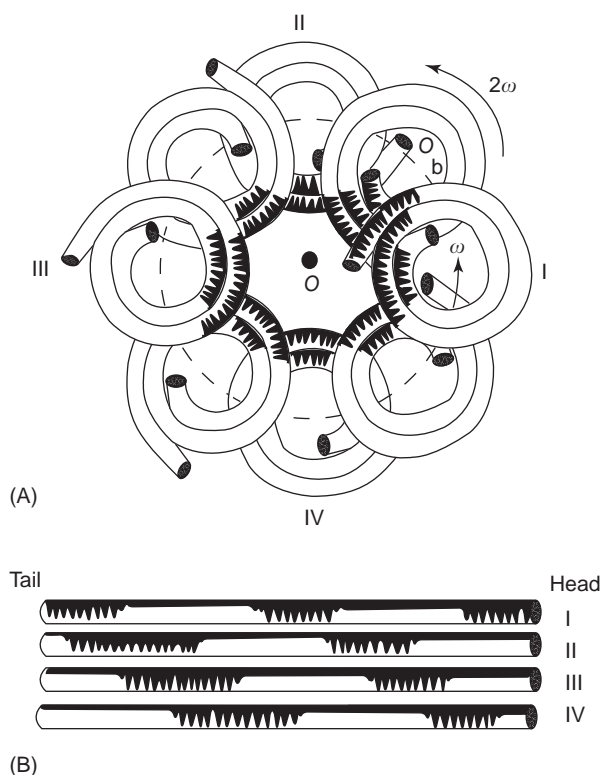


**Figure 6** Mechanism of high-speed CCC: (A) bilateral hydrodynamic distribution of the two phases in the coil, where the white phase occupies the head side and the black phase the tail side; (B) CCC operation utilizing the above hydrodynamic equilibrium condition; and (C) mechanism of dual CCC, where the two phases literally undergo countercurrent movement.

illustrated in Figure 6, where each coil is drawn as a straight tube. The coil at the top shows the bilateral hydrodynamic equilibrium where the white phase (head phase) occupies the head side and the black phase (tail phase) the tail side of the coil (Figure 6A). This equilibrium condition indicates that the white phase, if introduced at the tail, would move toward the head, and similarly the black phase introduced at the head would move toward the head. In Figure 6B, the upper coil is filled with the white phase and the black phase is introduced from the head end. The mobile phase (black) then travels rapidly through the coil, leaving a large volume of the stationary phase in the coil. Similarly, the lower coil is filled with the black phase and the white phase is introduced from the tail end. The mobile phase then travels through the coil, leaving a large volume of the stationary phase in the coil. In either case, solutes locally injected at the inlet of the coil are efficiently partitioned between the two phases and quickly eluted from the coil in the order of their partition coefficients, thus yielding high partition efficiency in a short separation time. The present system also permits simultaneous introduction of the two phases through the respective terminals, as illustrated in Figure 6C. This dual CCC operation requires an additional flow tube at each terminal to collect the effluent, and if desired a sample injection port is introduced at the middle portion of the coil. This system has been effectively applied to liquid-liquid dual CCC and foam CCC as described later.

The hydrodynamic distribution of the two solvent phases in the type-J planetary centrifuge has been

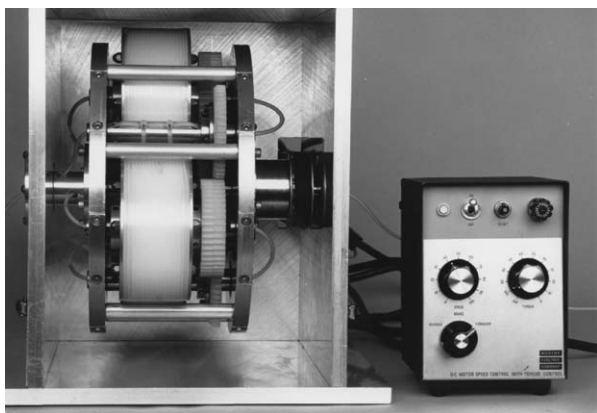




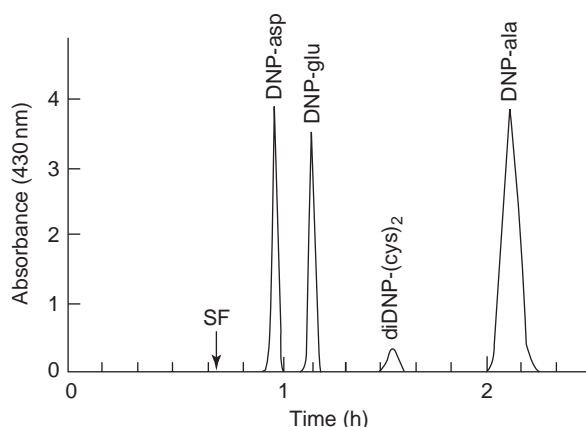
**Figure 7** Distribution of two immiscible solvent phases in the spiral column undergoing type-J synchronous planetary motion based on stroboscopic observation: (A) distribution of mixing and settling zones in the rotating spiral column and (B) motion of the mixing zones through the stretched spiral column.

observed under stroboscopic illumination. A spiral column was filled with the stationary phase and the colored mobile phase was eluted through the column in a proper elution mode. In the steady-state hydrodynamic equilibrium, the spiral column showed two distinct zones. As schematically illustrated in **Figure 7A**, vigorous mixing of the two solvent phases was observed in about one-fourth of the column area near the center of the centrifuge (mixing zone) while two phases are clearly separated into two layers in the rest of the area (settling zone). Because the location of the mixing zone is fixed to the centrifuge system, while the spiral column rotates about its own axis, each mixing zone is traveling through the spiral column at a rate of one round per revolution as shown in **Figure 7B**. This indicates an important fact that at any given portion of the column the two solvent phases are subjected to a repetitive partition cycle of mixing and settling at a high frequency of over 13 times per second at 800 rpm of column rotation. Because of its higher partition efficiency and speedy separation, the system was named 'high-speed CCC (HSCCC)'.

**Figure 8** shows the most advanced HSCCC centrifuge equipped with a set of three multilayer coils



**Figure 8** Photograph of HSCCC centrifuge equipped with a set of three multilayer coils connected in series.

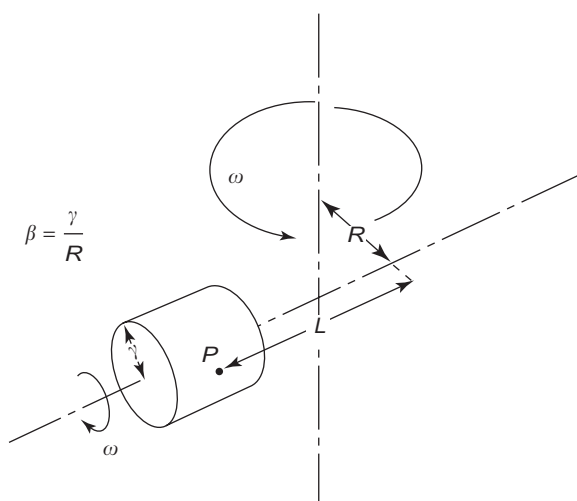


**Figure 9** Separation of DNP-amino acid by HSCCC. Experimental conditions: Apparatus: type-J HSCCC centrifuge with a set of three multilayer coil separation columns; column: semianalytical, 1.07 mm ID and 270 ml capacity; sample: DNP-amino acid, 10 mg; solvent system: chloroform/acetic acid/0.1 N HCl (2:2:1); mobile phase: upper aqueous phase; flow rate: 3 ml min<sup>-1</sup>; revolution: 1250 rpm; retention of stationary phase: 41.5% of the total column capacity.

connected in series on the rotary frame. It can produce highly efficient separations in a few hours. A typical separation obtained by HSCCC is illustrated in **Figure 9**, where four DNP-amino acids are resolved at a high efficiency ranging from 4200 (DNP-L-aspartic acid) to 3000 (DNP-L-alanine) TPs in a few hours. As in high-performance liquid chromatography (HPLC), HSCCC can also be interfaced with a mass spectrometer.

Although the type-J HSCCC has been successfully used for wide variety of natural and synthetic products, it fails to retain polymer phase systems that are useful for the separation and purification of various proteins. In order to improve the retention of the stationary phase for these extremely viscous solvent systems, a cross-axis coil planet centrifuge has been introduced as described below.





**Figure 10** Planetary motion of holder in cross-axis CPC.

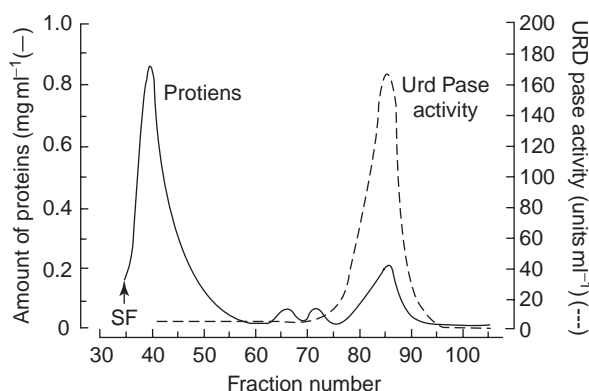
### Cross-Axis Coil Planet Centrifuge (Type-XL CPC)

This CCC system, which also belongs to a family of HSCCC, produces a characteristic planetary motion where the axis of the holder is perpendicular to the axis of the centrifuge (**Figure 10**). This modified planetary motion produces a complex pattern of the centrifugal force field fluctuating in a three-dimensional space, where one force vector acts across the diameter of the tube to enhance the separation of the two phases to improve the retention of the stationary phase. The ratio  $L/X$  (**Figure 10**) determines the retention stabilizing power, where  $L$  is the distance from the coil to the middle point of the rotary shaft, and  $X$  is the distance between the axis of the coil holder and the centrifuge axis. An apparatus such as the X-1.5L CPC has been successfully used for the separation of various recombinant enzymes with polymer phase systems as well as for large-scale preparative separation with polar solvent systems. Some examples are shown in **Figure 11** where recombinant uridine phosphorylase was purified by the X-1.5L CPC using a polymer phase system.

### Nonsynchronous Flow-Through Coil Planet Centrifuge

This most versatile coil planet centrifuge provides a freely adjustable coil rotation under a strong centrifugal force field produced by high-speed revolution, and yet the system does not require the use of a rotary seal. The apparatus is not yet commercially available probably due to its highly complex mechanical design (**Figure 12**).

This CCC system is ideal for the separation of biopolymers and bioparticles, such as proteins, nucleic



**Figure 11** Purification of recombinant uridine phosphorylase (Urd Pase) using polymer phase systems composed of 384 g of PEG (polyethylene glycol) 1000, 150 g each of  $\text{KH}_2\text{PO}_4$  and  $\text{K}_2\text{HPO}_4$  in 1716 g of water.

acids, mammalian cells and bacteria using polymer phase systems.

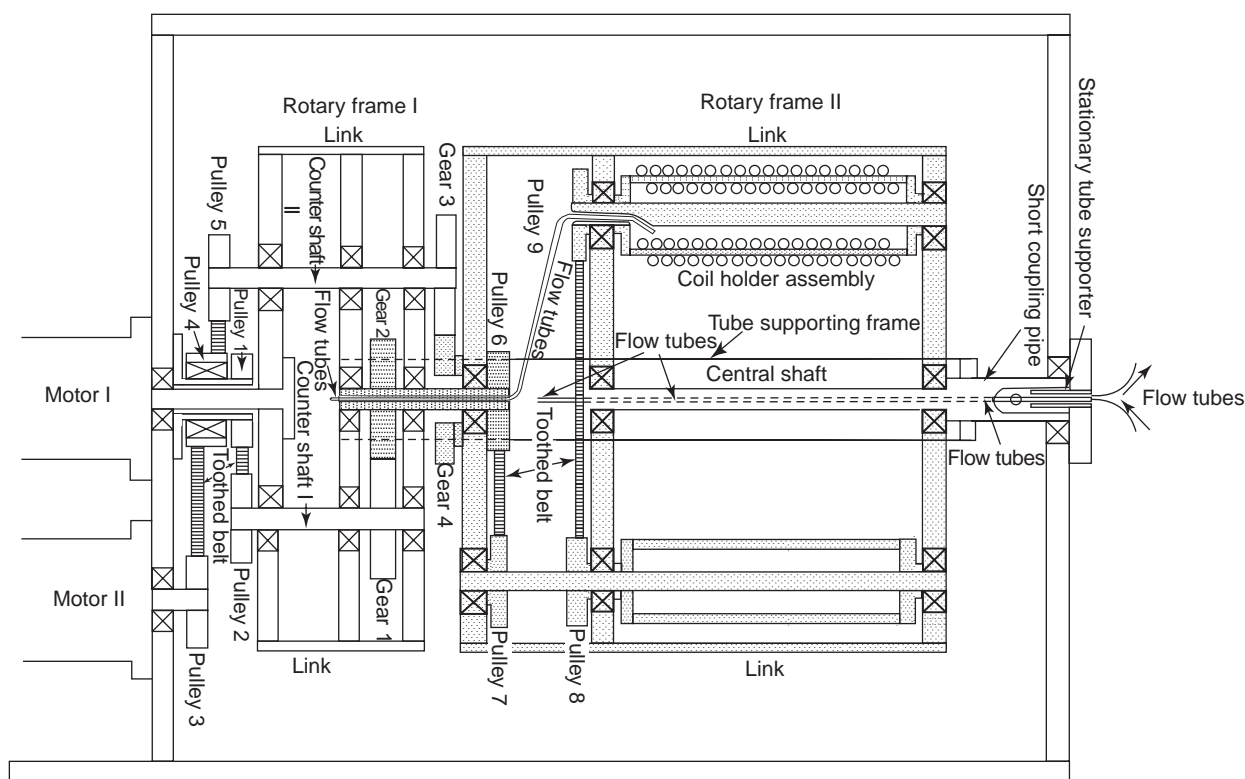
## Specific CCC Techniques

Recently two useful CCC techniques have been developed: pH-zone-refining CCC, which separates a large quantity of charged samples, and chiral CCC, which uses a ligand called chiral selector to separate enantiomers according to their affinity to the chiral selector.

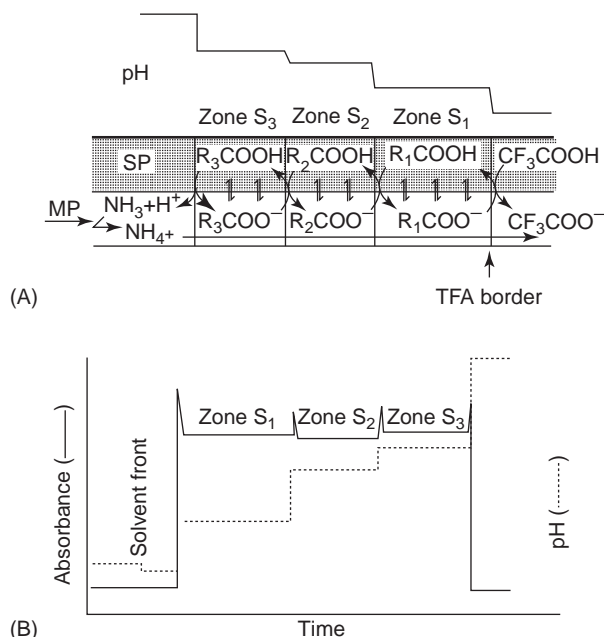
### pH-Zone-Refining CCC

This large-scale preparative CCC technique separates ionic compounds to produce a train of highly concentrated rectangular peaks fused together with minimum overlap, which is quite similar to that in displacement chromatography. The method uses a retainer in the stationary phase to retain the target compounds and an eluter in the mobile phase to elute them according to their hydrophobicity and  $\text{pK}_a$ . The most commonly used retainers are trifluoroacetic acid (TFA) for acidic samples and triethylamine for basic samples, while the eluter is usually ammonia for acidic samples and hydrochloric acid for basic samples.

Mechanism of pH-zone-refining CCC is shown in **Figure 13A**, which schematically illustrates a cross-sectional view of the separation column where three acidic compounds are eluted with a mobile phase containing ammonia and through a stationary phase containing TFA. Due to its nonlinear isotherm, the retainer (TFA) forms a sharp rear border that moves at a rate considerably lower than that of the mobile phase. Three analytes,  $S_1$ ,  $S_2$ , and  $S_3$ , competitively form solute zones behind the sharp TFA border according to their  $\text{pK}_a$  and hydrophobicity. Among



**Figure 12** Design of nonsynchronous flow-through CPC without rotary seals. Cross-sectional view through the central axis of the apparatus.



**Figure 13** Mechanism of pH-zone-refining CCC: (A) partition process within the column and (B) elution profile.

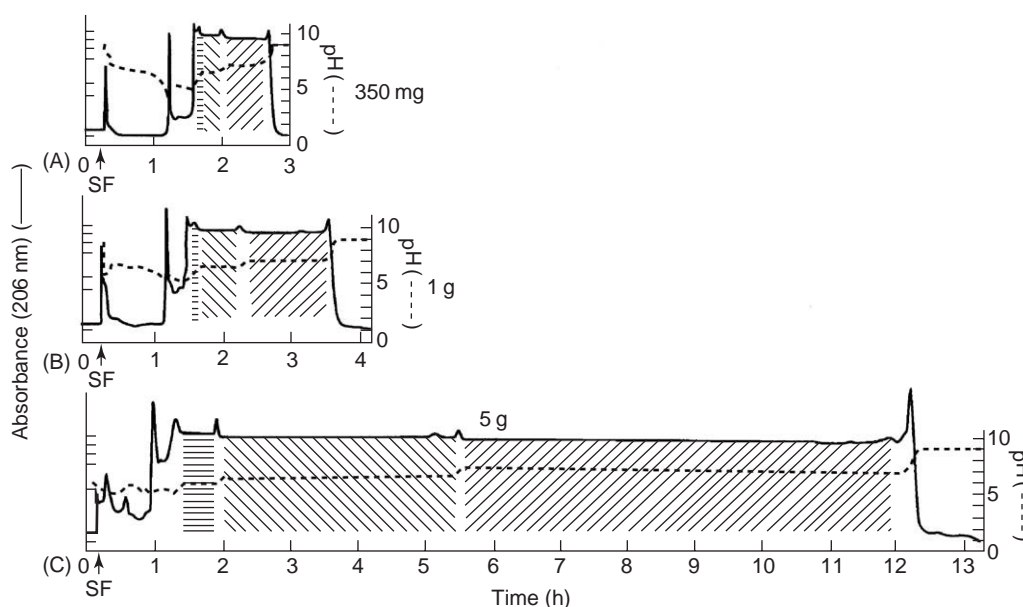
these,  $S_1$  with the lowest  $pK_a$  and hydrophobicity is located immediately behind the TFA border, while  $S_3$  with the highest  $pK_a$  and hydrophobicity is located at

the end of the solute zones where it forms a sharp trailing border. As indicated by curved arrows, proton transfer takes place at each zone boundary governed by the difference in pH between the neighboring zones. The loss of the solute from the mobile phase to the stationary phase at the zone front is compensated by its return at the back of each zone, while ammonium ion in the aqueous phase serves as counterion for all species. After equilibrium is reached three solute zones move at the same rate as that of the TFA border, while constantly maintaining their width and pH. Charged minor components present in each zone are efficiently eliminated either forward or backward according to their  $pK_a$  and hydrophobicity and eventually accumulate at the zone boundaries. Consequently, the three analytes elute as a train of rectangular peaks with sharp impurity peaks at their narrow boundaries, as illustrated in **Figure 13B**.

The zone pH ( $pH_{zone}$ ), which governs the order of the elution for a set of analytes, is expressed according to the following formula:

$$pH_{zone} = pK_a + \log \left\{ \left( \frac{K_D}{K_S} \right) - 1 \right\} \quad [2]$$

where  $K_D$  is the partition ratio of solute  $S$  ( $RCOOH$ ) and  $K_S$ , the partition constant.



**Figure 14** Separations of C&D Orange No.5 by pH-zone-refining CCC. Sample size: 350 mg (A); 1 g (B); and 5 g (C). Apparatus: HSCCC centrifuge with 10 cm revolution radius; column: semipreparative multilayer coil, mm  $\times$  1.6 mm ID, 325 ml capacity; solvent system: diethyl ether/acetonitrile/0.01 mol l<sup>-1</sup> aqueous ammonium acetate (pH 9 by ammonia) (4:1:5); Mobile phase: lower aqueous phase; retainer: TFA in the sample solution; flow rate: 3 ml min<sup>-1</sup>; revolution: 800 rpm; SF: solvent front.

Figure 14 illustrates the separation of D&C Orange No.5 by pH-zone-refining CCC where three chromatograms were obtained at different sample sizes of 350 mg, 1 g, and 5 g as indicated in the diagram. In each chromatogram, the shaded area represents fractions containing pure components. As clearly shown in this figure, increasing the sample size results in a proportional increase of the peak width, while the width of the mixing zone containing impurities remains the same. These results indicate one of the important advantages of the present technique, i.e., increasing the sample size gives a higher yield of pure fractions. A few other typical applications of pH-zone-refining CCC are illustrated in Figures 15A–15C, and its advantages over the standard CCC technique are summarized in Table 1.

### Chiral CCC

Analogous to liquid chromatography, CCC can be used for chiral separation by dissolving the chiral selector in the liquid stationary phase. The method has some advantages over liquid chromatography in that the amount of the chiral selector in the column can be much greater (since it is dissolved in the stationary phase) and the same column may be used for various kinds of chiral separations simply by changing the chiral selector in the stationary phase.

Figure 16 shows the separation of ( $\pm$ )DNB(di-nitrobenzoyl)-amino acids using a chiral selector,

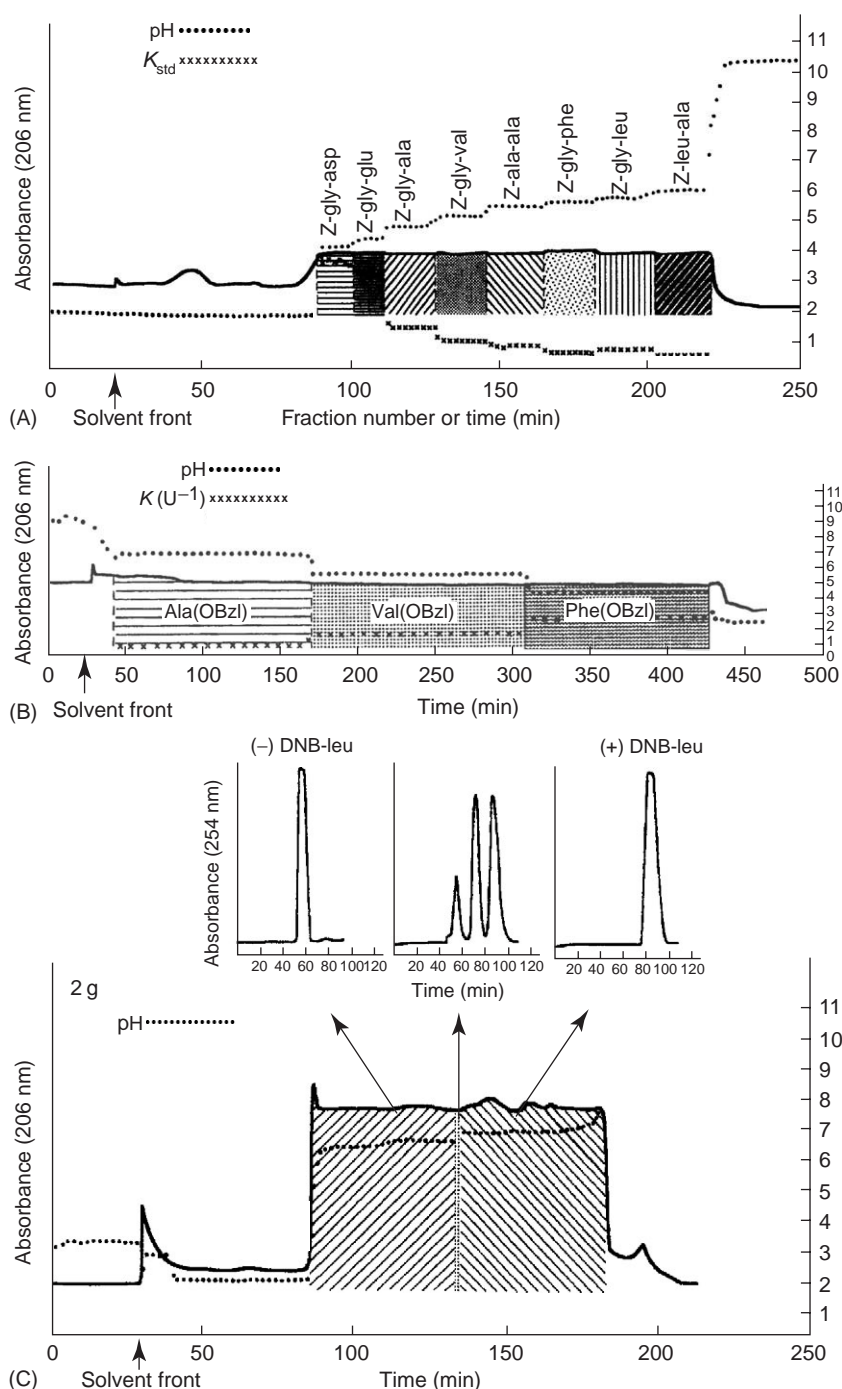
*N*-dodecanoyl-L-proline-3,5-dimethylanilide by HSCCC. The retention time of each enantiomer and its peak resolution are dependent on the  $K_F$  (formation constant) of chiral selector–amino acid complexes. Several ( $\pm$ )DNB-amino acids were resolved in a single run by adding the above chiral selector in the stationary phase.

A series of experiments on separation of ( $\pm$ )DNB-amino acids using the above chiral selector revealed that the peak resolution of the racemates are increased by (1) increasing the net amount or concentration of the chiral selector in the organic stationary phase and (2) adjusting the hydrophobicity of the solvent system so that the mean partition coefficient values for racemates fall between 0.6 and 0.8. The peak resolution will be further increased by the use of a longer and/or greater ID coiled column.

One of the advantages of chiral CCC over the conventional chiral chromatographic technique is that the method can be used to determine  $K_F$  (formation constant) of the chiral selector and target enantiomer.

### Dual CCC

As mentioned earlier, the type-J CPC can be used for dual CCC where two solvent phases literally countercurrent through a coiled tube (Figure 6C). In this system both polar and nonpolar components are eluted from the column in the respective outlet while the components with their partition coefficients at



**Figure 15** Examples of chromatograms obtained by pH-zone-refining CCC: (A) Separation of Z or CBZ(*N*-carbobenzoxy)-peptides. Solvent system: methyl-*tert*-butyl ether/acetonitrile/water (2:2:3), 16 mmol l<sup>-1</sup> TFA in organic stationary phase (pH 1.83) and 5.5 mmol l<sup>-1</sup> NH<sub>3</sub> in aqueous mobile phase (pH 10.62); sample: eight CBZ(*Z*)-dipeptides as indicated, each 100 mg dissolved in 50 ml solvent (25 ml each phase); flow rate: 3.3 ml min<sup>-1</sup>, head-to-tail elution mode; revolution: 800 rpm; (B) Separation of amino acid benzylesters. Solvent system: methyl-*tert*-butyl ether/water, 5 mmol l<sup>-1</sup> triethylamine in organic stationary phase and 20 mmol l<sup>-1</sup> HCl in aqueous mobile phase; sample: ala(OBzl), val(OBzl), and phe(OBzl) each 3.3 g in 100 ml of solvent; flow rate: 3 ml min<sup>-1</sup>; revolution: 800 rpm; and (C) Separation of (±)DNB(dinitro-benzoyl)-leucine using chiral selector. Solvent system: methyl-*tert*-butyl ether/water, TFA (40 mmol l<sup>-1</sup>) + chiral selector DPA (*N*-dodecanoyl-L-proline-3,5-dimethylanilide) (40 mmol l<sup>-1</sup>) in organic stationary phase and ammonia (20 mmol l<sup>-1</sup>) in aqueous mobile phase; sample: (±)-DNB-leucine 2 g; flow rate: 3 ml min<sup>-1</sup>; elution mode: head to tail; revolution: 800 rpm. (Analytical HSCCC was carried out using the same column by the standard CCC technique under the following conditions: solvent system: hexane/ethyl acetate/methanol/10 mmol l<sup>-1</sup> HCl (8:2:5:5), organic stationary phase containing DPA (20 mmol l<sup>-1</sup>); flow rate: 3 ml min<sup>-1</sup>; elution mode: head to tail; revolution: 800 rpm. All separations (A–C) were performed with the apparatus and column described in the **Figure 14** caption.)

near unity are retained in the column for a long period of time. The method is ideal for sample cleaning up for mass spectrometric analysis, since the repetitive sample injection is possible without loss of the stationary phase and minimum accumulation of impurities in the column.

The system also provides unique application to foam separation in which gas and liquid phases undergo a true countercurrent movement through a

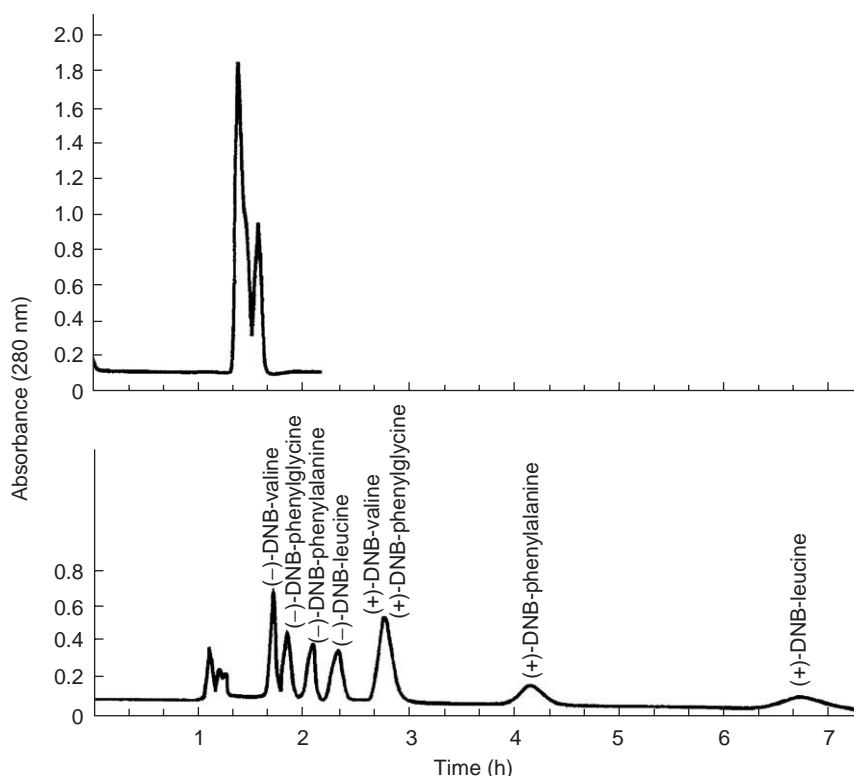
long narrow coiled tube. When the liquid phase contains a surfactant, the dual countercurrent process produces a foaming stream that moves toward the tail. The sample mixture introduced at the middle portion of the column is separated according to their foam affinity: foam-active components are quickly carried with the foaming stream toward the tail whereas the remainder is carried in the liquid stream in the opposite direction and collected at the head end of the coil. Also for samples with a strong foaming capacity such as proteins and peptides (bactracin), foam CCC can be performed without surfactant in the liquid phase.

**Table 1** Advantages of pH-zone-refining CCC over standard HSCCC

Sample size	Applicable sample size is increased over 10 times for a given column
Fraction	Eluted fractions are highly concentrated to near saturation
Yield	Higher the sample size, the greater the yield of pure fractions
Minor components	Charged minor components are concentrated and accumulated at zone boundaries
Detection	Solute with nonchromophore can be monitored with pH

## Selection of Two-Phase Solvent Systems

The procedure of CCC is somewhat similar to that for liquid chromatography except that CCC uses two solvent phases that should provide a suitable partition coefficient ( $K$ ) to the target compounds as well as satisfactory retention of the stationary phase in the column.



**Figure 16** Chiral separation by HS-CCC. One step separation of various ( $\pm$ ) DNB amino acids by HS-CCC with the chiral selector DPA (*N*-dodecanoyl-L-proline-3,5-dimethylanilide) in the stationary phase (lower chromatogram). Upper chromatogram was obtained without the chiral selector in the stationary phase under otherwise identical experimental conditions. Solvent system: Apparatus and column: analytical HS-CCC centrifuge with a set of three multilayer coils (0.85 mm ID Teflon tubing) connected in series with a total capacity of 60 ml; solvent system: hexane/ethyl acetate/methanol/10 mmol l<sup>-1</sup> HCl (8:2:5:5) without ligand (upper chromatogram) and with a ligand DPA 1.6 g in the stationary phase (lower chromatogram); flow rate: 1 ml min<sup>-1</sup>; elution mode: head to tail; sample: 10 mg each of ( $\pm$ ) DNB amino acids indicated in the chromatogram dissolved in 2 ml of solvent (1 ml each phase); revolution: 1000 rpm.



### Partition Coefficient ( $K$ )

Since CCC uses two-solvent phases, the solute partition coefficient can be easily determined by a test tube experiment prior to the separation. In each measurement the sample is first partitioned between the two equilibrated solvent phases in a test tube, an aliquot of each phase is diluted with a suitable solvent, and then the concentration of the solute in each solution is determined by a spectrophotometer. Also, other methods such as radioactivity, enzymatic activity can be used. The partition coefficient,  $K$ , is the ratio between these two measurements and is usually expressed as solute concentration in the stationary phase divided by that of the mobile phase. When the sample is a mixture of multiple components, the partition coefficient of each component can be obtained by HPLC, gas chromatography, or thin-layer chromatography of each phase in the test tube and by comparing the peak height or area between the corresponding peaks in the two chromatograms. In general, the most suitable range of  $K$  values is  $1 \leq K \leq 2$  for hydrostatic CCC systems and  $0.5 \leq K \leq 1$  for hydrodynamic CCC systems. Once the  $K$  value is determined, the retention volume of the solute can be computed from the following equation:

$$V_R = V_{SF} + K(V_C - V_{SF}) \quad [3]$$

where  $V_R$  is the retention volume of the solute,  $V_{SF}$  the retention volume of the solvent front (amount of the mobile phase in the column), and  $V_C$  the total column capacity.

### Retention of Stationary Phase

The retention of the stationary phase in the separation column is an important parameter to determine the resolution of solute peaks in CCC. Generally, the higher the retention of the stationary phase, the better the separation. In the hydrostatic systems, the retention of the stationary phase is improved simply by reducing the flow rate of the mobile phase.

In the hydrodynamic systems, which provide efficient mixing of the two phases, the retention of the stationary phase depends on the physical properties of the two-phase solvent system as well as the choice of the mobile phase and its elution mode. In HSCCC using the type-J planetary motion, the settling time of the two solvent phases in unit gravity provides a useful measure for the retention of the stationary phase and the elution mode. The test is performed as follows: the two phases are pre-equilibrated in a separating funnel and 2 ml of each phase is delivered into a 5 ml graduated cylinder equipped with a

stopper (an ordinary glass test tube of 13 mm OD and 10 cm length with a polyethylene cap can also be used). The contents are gently mixed by inverting the container five times and the time required to form two clear layers is measured. If this settling time is within 30 s, the solvent system can be used for separation by eluting the lower phase from the head toward the tail or the upper phase in the reversed mode. If the settling time exceeds 30 s, the above elution mode should be reversed while the retention of the stationary phase is usually considerably lower than an optimum range. However, this settling time test is not applied to the cross-axis coil planet centrifuges. These centrifuge systems provide excellent retention of the stationary phase for almost all two-phase solvent systems including viscous polymer phase systems used for partition of macromolecules and cell particles.

### CCC versus HPLC

CCC is a pure liquid-liquid partition chromatographic method using no solid support and has various advantages over HPLC. Although both methods are based on the same principle of solute partitioning between two phases, HPLC uses a bulk of solid support coated with the thin film of stationary phase, which limits the amount of stationary phase and its sample loading capacity. In contrast, CCC uses an empty column filled with a liquid stationary phase. When eluted with the mobile phase, the column can still retain over 50% of the stationary phase and therefore it has a large sample loading capacity. The solid support used in HPLC often irreversibly adsorbs the solutes and denatures the proteins. Solid support preparation is difficult to control, expensive, and often unstable. Most importantly, CCC separations are predictable. From the partition coefficient that is measured by a simple test tube experiment, one can predict the retention time of the target compounds.

*See also:* **Countercurrent Chromatography:** Solvent Extraction with a Helical Column. **Extraction:** Solvent Extraction: Multistage Countercurrent Distribution.

### Further Reading

- Berthod A (2002) *Countercurrent Chromatography*. Comprehensive Analytical Chemistry, vol. XXXVIII. Amsterdam: Elsevier.
- Conway WD and Petroski RJ (1993) *Modern Countercurrent Chromatography*, ACS symposium series, vol. 593. Washington, DC: American Chemical Society.



Conway WD (1990) *Countercurrent Chromatography: Apparatus, Theory and Applications*. New York: VCH.

Foucault A (1994) *Centrifugal Partition Chromatography*. New York: Dekker.

Ito Y (1981) Countercurrent chromatography. *Journal of Biochemical and Biophysical Methods* 5: 105–129.

Ito Y (1984) Development of high-speed countercurrent chromatography. In: Giddings JC, Grushka E, Cazes J, and Brown PR (eds.) *Advances in Chromatography*, vol. 24, pp. 181–226. New York: Dekker.

Ito Y (1986) High-speed countercurrent chromatography. *CRC Critical Review in Analytical Chemistry* 17: 65–143.

Ito Y (1991) Countercurrent chromatography in chromatography V. *Journal of Chromatography Library*, Part A, ch. 2. Amsterdam: Elsevier.

Ito Y and Conway WD (1996) *High-Speed Countercurrent Chromatography*. New York: Wiley-Interscience.

Ito Y and Ma Y (1996) pH-zone-refining countercurrent chromatography. *Journal of Chromatography A* 753: 1–36.

Mandava NB and Ito Y (1988) *Countercurrent Chromatography: Theory and Practice*. New York: Dekker.

Menet JM and Thiébaud D (1999) *Countercurrent Chromatography*. New York: Dekker.

## Solvent Extraction with a Helical Column

**Y Ito**, National Heart, Lung, and Blood Institute, Bethesda, MD, USA

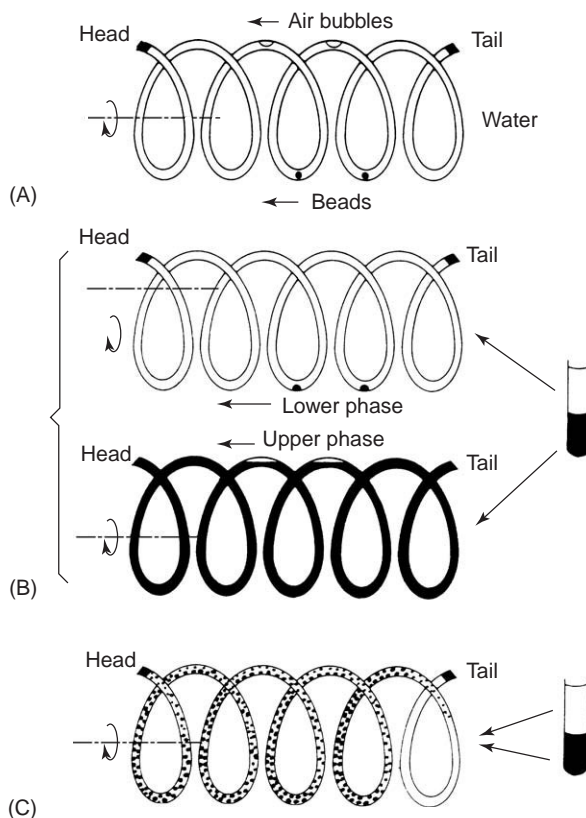
© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Three steps required for multistage solvent extraction, i.e., phase mixing, phase settling, and transfer of the mobile phase, are defined clearly in the discontinuous countercurrent distribution process using the Craig apparatus. These basic requirements are essentially fulfilled by the use of a coiled tube in a continuous fashion. Solvent extraction using a coiled column is most efficiently performed with a horizontally laid coil that rotates about its own axis. In this horizontal coil orientation, the rotation induces the well known Archimedean screw force, which can be utilized for performing countercurrent solvent extraction.

### Hydrodynamic Motion of Two Immiscible Solvent Phases in a Rotating Coil

This Archimedean screw effect on the hydrodynamic distribution of the solvent phases is illustrated in **Figure 1**, where each coil, consisting of five helical turns, is placed horizontally and rotated slowly around its axis. In **Figure 1A**, air bubbles and glass beads are introduced in the coil, previously filled with water, and both ends of the coil are sealed. Then, rotation of the coil in the indicated direction induces an Archimedean screw force that drives all the suspended objects (lighter or heavier than the



**Figure 1** Archimedean screw effect in a rotating coil. (A) Motion of air bubbles and glass beads suspended in water; (B) Motion of droplets of one phase of an equilibrated two-phase solvent system suspended in the other phase; (C) Hydrodynamic equilibrium of two immiscible solvent phases in a slowly rotating coil.

water) toward the end of the coil labeled 'head' (the other side of the coil is called the 'tail'). Under a slow coil rotation, the air bubbles (which are lighter than the water) always stay at the top of the coil, while the

Conway WD (1990) *Countercurrent Chromatography: Apparatus, Theory and Applications*. New York: VCH.

Foucault A (1994) *Centrifugal Partition Chromatography*. New York: Dekker.

Ito Y (1981) Countercurrent chromatography. *Journal of Biochemical and Biophysical Methods* 5: 105–129.

Ito Y (1984) Development of high-speed countercurrent chromatography. In: Giddings JC, Grushka E, Cazes J, and Brown PR (eds.) *Advances in Chromatography*, vol. 24, pp. 181–226. New York: Dekker.

Ito Y (1986) High-speed countercurrent chromatography. *CRC Critical Review in Analytical Chemistry* 17: 65–143.

Ito Y (1991) Countercurrent chromatography in chromatography V. *Journal of Chromatography Library*, Part A, ch. 2. Amsterdam: Elsevier.

Ito Y and Conway WD (1996) *High-Speed Countercurrent Chromatography*. New York: Wiley-Interscience.

Ito Y and Ma Y (1996) pH-zone-refining countercurrent chromatography. *Journal of Chromatography A* 753: 1–36.

Mandava NB and Ito Y (1988) *Countercurrent Chromatography: Theory and Practice*. New York: Dekker.

Menet JM and Thiébaud D (1999) *Countercurrent Chromatography*. New York: Dekker.

## Solvent Extraction with a Helical Column

**Y Ito**, National Heart, Lung, and Blood Institute, Bethesda, MD, USA

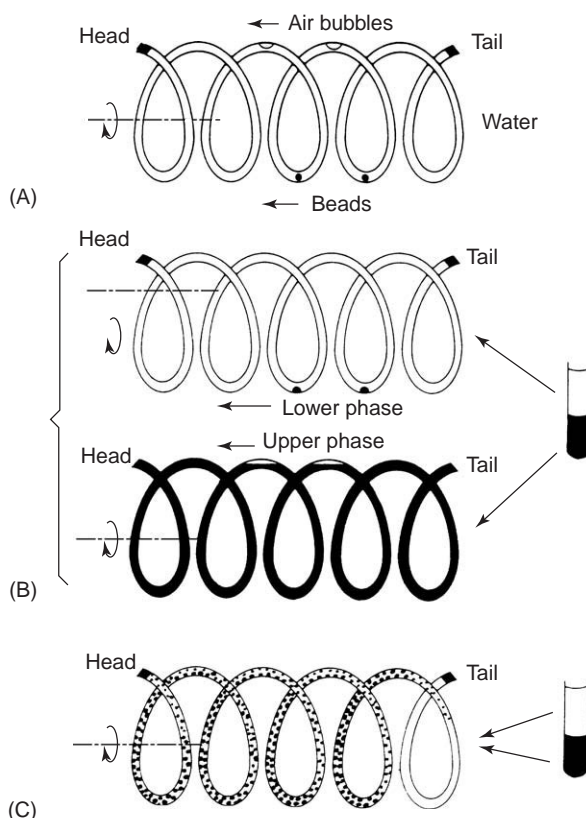
© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Three steps required for multistage solvent extraction, i.e., phase mixing, phase settling, and transfer of the mobile phase, are defined clearly in the discontinuous countercurrent distribution process using the Craig apparatus. These basic requirements are essentially fulfilled by the use of a coiled tube in a continuous fashion. Solvent extraction using a coiled column is most efficiently performed with a horizontally laid coil that rotates about its own axis. In this horizontal coil orientation, the rotation induces the well known Archimedean screw force, which can be utilized for performing countercurrent solvent extraction.

### Hydrodynamic Motion of Two Immiscible Solvent Phases in a Rotating Coil

This Archimedean screw effect on the hydrodynamic distribution of the solvent phases is illustrated in **Figure 1**, where each coil, consisting of five helical turns, is placed horizontally and rotated slowly around its axis. In **Figure 1A**, air bubbles and glass beads are introduced in the coil, previously filled with water, and both ends of the coil are sealed. Then, rotation of the coil in the indicated direction induces an Archimedean screw force that drives all the suspended objects (lighter or heavier than the



**Figure 1** Archimedean screw effect in a rotating coil. (A) Motion of air bubbles and glass beads suspended in water; (B) Motion of droplets of one phase of an equilibrated two-phase solvent system suspended in the other phase; (C) Hydrodynamic equilibrium of two immiscible solvent phases in a slowly rotating coil.

water) toward the end of the coil labeled 'head' (the other side of the coil is called the 'tail'). Under a slow coil rotation, the air bubbles (which are lighter than the water) always stay at the top of the coil, while the

glass beads (which are heavier than the water) stay at the bottom of the coil, both moving toward the head of the coil at a rate of one helical turn per rotation of the coil. Finally, both air bubbles and glass beads reach the head of the coil, where they remain by repeating a back and forth motion synchronous with the rotation of the coil.

In **Figure 1B**, similar experiments are performed with a two-phase solvent system. The first coil is filled with the lighter phase (white) of an equilibrated two-phase solvent system, and a small amount of the heavier phase (black) is added. Then, under a slow rotation of the coil, droplets of the heavier phase remain at the bottom of the coil, traveling through the coil at a rate equal to the coil rotation. Similarly, the second coil is filled with the heavier phase (black), and a small amount of the lighter phase (white) is added. Under a slow rotation, droplets of the lighter phase stay at the top of the coil, again traveling toward the head of the coil at a rate of one turn per rotation of the coil.

In **Figure 1C**, the coil is filled with nearly equal volumes of the two phases and rotated slowly about its axis. In this case, the lighter phase stays at the upper portion and the heavier phase at the lower portion of the coil, both competitively advancing toward the head of the coil. Sooner or later, the two phases establish a hydrodynamic equilibrium where each phase occupies about an equal space on the head side of the coil, and any excess of either phase remains at the tail end of the coil. Once this hydrodynamic equilibrium is formed, continued rotation of the coil mixes the two solvent phases vigorously, while the overall distribution of the two phases remains unaltered.

This hydrodynamic equilibrium can be used for performing solute extraction in the following manner: The coil is first filled entirely with the stationary phase, either the lighter or the heavier phase, and the sample solution dissolved in either phase is introduced at the head side of the coil. Then, the mobile phase is eluted through the coil from the head toward the tail while the coil is slowly rotated around its axis. As the mobile phase meets the stationary phase in the rotating coil, the two solvent phases establish a hydrodynamic equilibrium quickly: the two phases are vigorously mixed by rotation of the coil, while some amount of the stationary phase is permanently retained in the coil. This process continues in each helical turn of the coil. After the entire coil attains a hydrodynamic equilibrium state and the mobile phase begins to elute from the tail end of the coil, the mobile phase only displaces the same phase in the coil, leaving the retained stationary phase permanently in the coil. Consequently, the solutes present

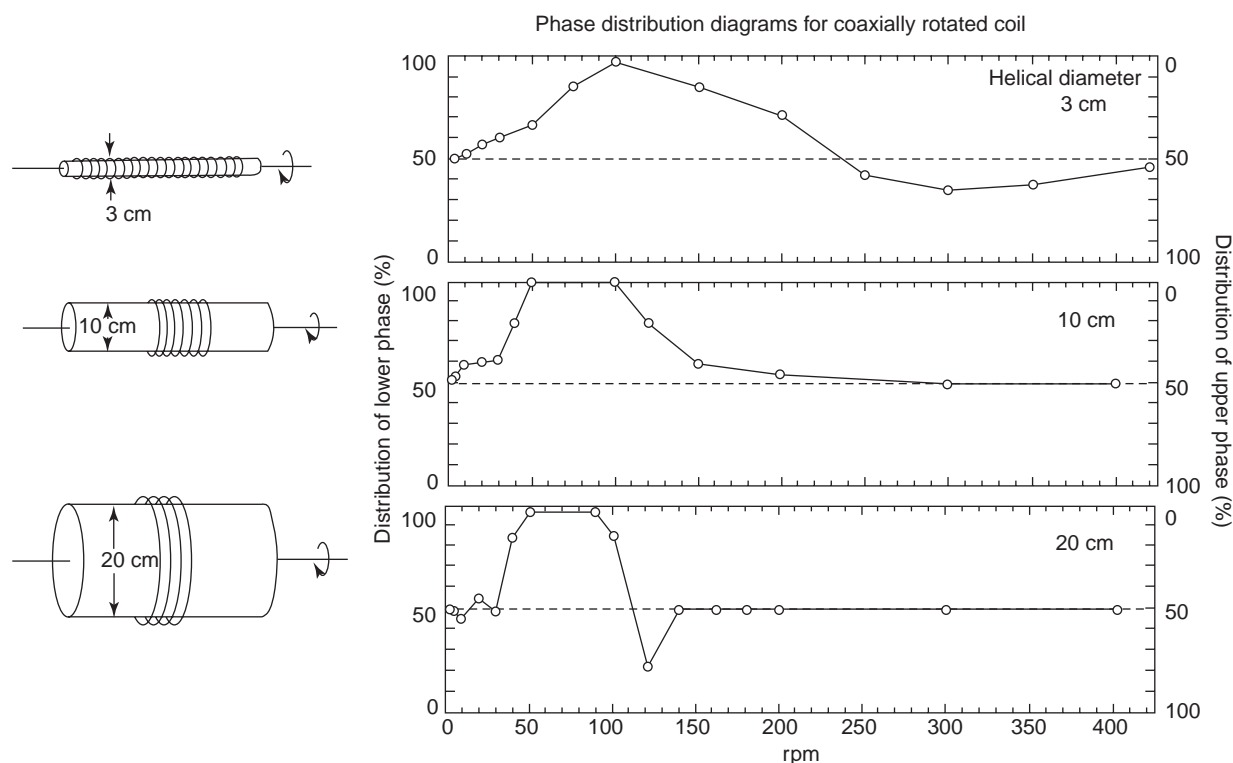
in the sample solution are subjected to an efficient partition process as in the multistage countercurrent distribution with the Craig apparatus but in a continuous manner.

The partition efficiency in this partition system is highly dependent upon the amount of the stationary phase retained in the column. Under a slow rotation of the coil as described above, the two solvent phases occupy competitively the head side of the coil where the elution of either phase from the head of the coil only permits retention of the stationary phase at a maximum level of 50% of the column capacity. It has been found, however, that the volume ratio of the two phases occupying the head side of the coil can be altered by increasing the rotation speed of the coil.

In **Figure 2**, three diagrams indicate the effect of rotation speed on the two-phase distribution on the head side of the coil. The experiments were performed using a set of coils with helical diameters of 3, 10, and 20 cm as indicated on the left. Each coil was first filled with about equal volumes of the lighter and heavier phases of chloroform/acetic acid/0.1 mol l<sup>-1</sup> hydrochloric acid (2:2:1, v/v), closed at both ends, and then rotated at the desired speed. After a hydrodynamic phase equilibrium was reached, the rotation was stopped and the volume of the two phases occupying head side of the coil was measured. The percentage volume of the heavier (lower) phase occupying the head of the coil was then plotted against the rotation speed of the coil.

The three diagrams obtained from different helical diameters show common features: In the slow rotation speed, between 0 rpm and 30 rpm, the two solvent phases distribute fairly evenly in the coil (stage I). When the rotation speed is increased, the heavier nonaqueous phase tends to occupy more space on the head side of the coil, and at a critical speed between 60 rpm and 100 rpm, the two phases are almost completely separated along the length of the coil, the heavier phase occupying the head side and the lighter phase the tail side of the coil (stage II). This particular two-phase distribution is called bilateral and most efficiently utilized for performing solvent extraction. After this critical speed range, the amount of heavier phase on the head side tends to decrease rather sharply, crossing below the 50% line (stage III). A further increase in the rotational speed again yields an even distribution of the two phases in the coil (stage IV). As the helical diameter increases, all these stages tend to shift toward the lower rpm range, apparently due to the enhanced centrifugal force field.

Series of similar studies have been carried out using various two-phase solvent systems with a broad spectrum of hydrophobicity. **Figure 3** illustrates a set



**Figure 2** Hydrodynamic distribution of a two-phase solvent system composed of chloroform/acetic acid/0.1 mol l<sup>-1</sup> hydrochloric acid (2:2:1, v/v) in rotating coils of three different helical diameters, shown on the left.

of phase distribution diagrams obtained from nine commonly used volatile solvent systems in glass coils of various sizes with or without a silicone coating. These diagrams are arranged from left to right in the order of hydrophobicity of the major organic solvents as labeled at the top of each column, whereas the inner diameter (ID) and core diameter of the coils are indicated on the left margin. In each diagram the solid curve was obtained from an untreated coil and the broken curve from the same coil after silicone coating. Thus, any difference between these two curves indicates the effects of the solvent-wall interaction. An absence of one or both distribution curves in the designated space indicates that the two solvent phases failed to move or displayed sluggish motion in the coil and, therefore, the measurement could not be completed. These data indicate that with only a few exceptions, various two-phase solvent systems establish a bilateral hydrodynamic distribution in 1–2 cm ID coils at rotation speeds ~100 rpm.

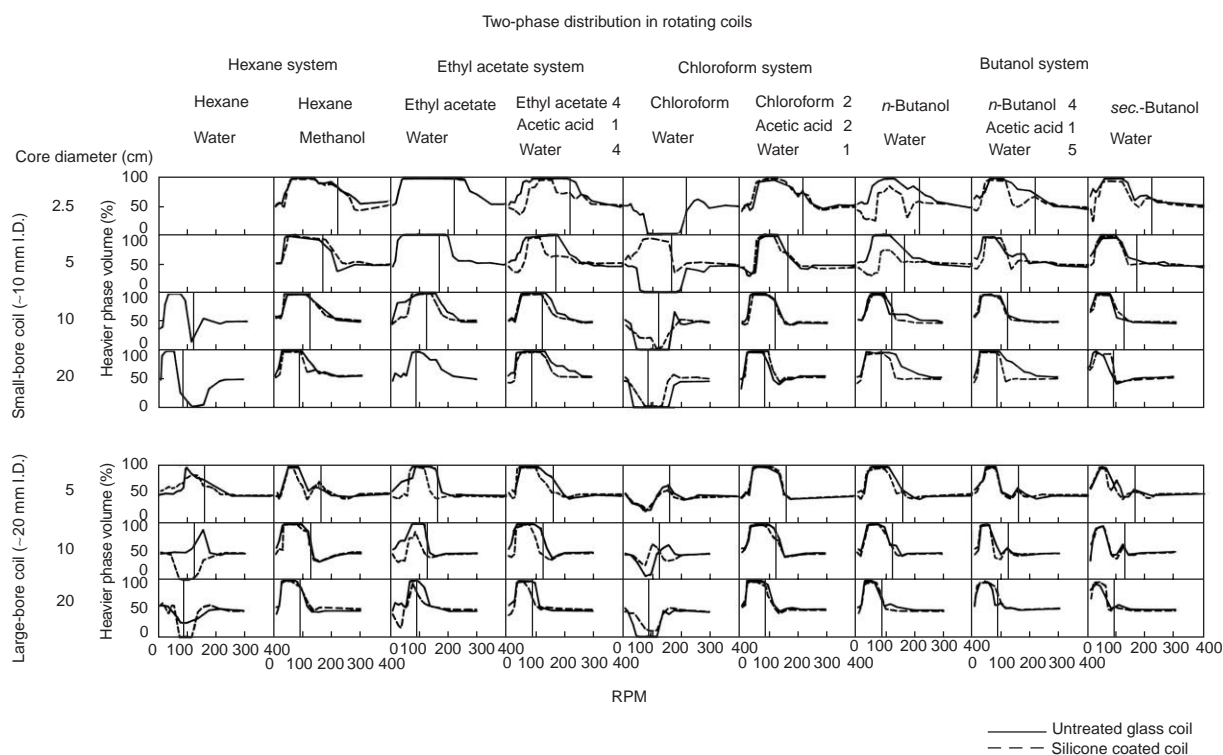
## Mechanism of Countercurrent Extraction

As mentioned above, the bilateral hydrodynamic distribution of the two phases can be utilized efficiently

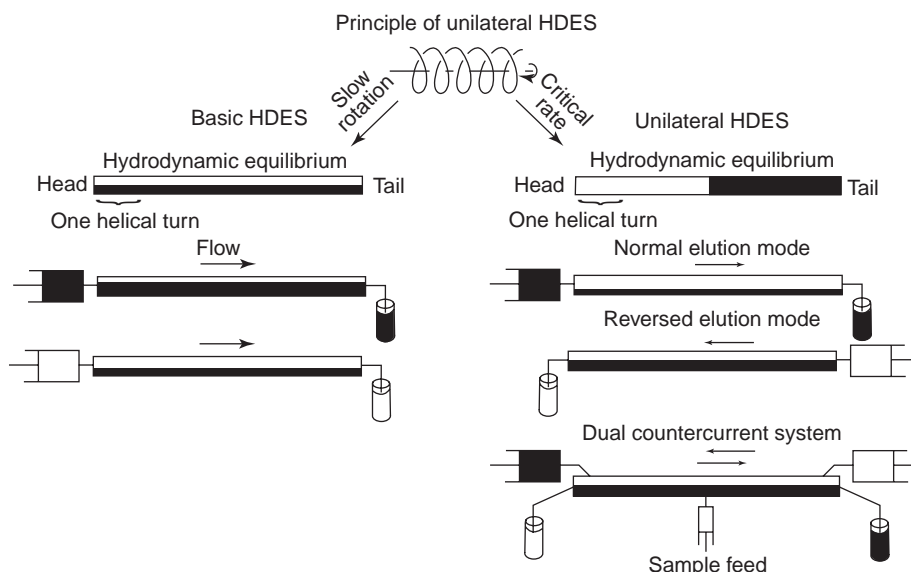
for performing solvent extraction. **Figure 4** illustrates schematically the hydrodynamic mechanisms in solvent extraction using the basic hydrodynamic distribution (stage I) under a slow coil rotation (left) and the bilateral hydrodynamic distribution (stage II) at the critical rotation speed (right). For simplicity, all rotating coils – except for one shown at the top – are drawn uncoiled to show the overall distribution of the two solvent phases along the length of the coil.

In the basic hydrodynamic equilibrium system (left), the slow rotation of the coil distributes the two phases evenly from the head of the coil (top). In order to obtain retention of the stationary phase in the coil under this hydrodynamic condition, the mobile phase, regardless of whether it is the heavier or lighter phase, should be introduced from the head of the coil. This operation results in a low level of stationary phase retention, usually much less than 50% of the total column capacity, as shown in the diagram. Elution of either phase from the tail of the coil would result in a total loss of the stationary phase from the coil.

In the bilateral hydrodynamic equilibrium system (right), the critical rotation speed distributes the two phases bilaterally along the length of the coil, the head phase (white) entirely occupying the head side and the tail phase (gray) the tail side of the coil as



**Figure 3** Phase distribution diagrams for nine volatile two-phase solvent systems obtained from glass coils with various dimensions, as indicated on the left. The solid curve indicates the data obtained from nontreated glass coils and the dotted curve, from silicone-treated glass coils. The thin vertical line in each diagram indicates the rpm value at which the centrifugal force field created by the rotation equals unit gravity. Note that most solvent systems exhibit the critical rpm value where one phase occupies 100% of the column space on the head side of the coil.



**Figure 4** Mechanism of solvent extraction with rotating coils. Left: basic hydrodynamic distribution produced by a slow coil rotation (stage I). Right: bilateral hydrodynamic distribution produced by the critical rotation speed (stage II).

shown in the first coil. In a rotating coil under unit gravity, the heavier phase usually becomes the head phase (see Figure 3). This hydrodynamic equilibrium condition indicates clearly that the tail phase, if

introduced at the head end of the coil, would travel through the head phase toward the tail and that the head phase, if introduced at the tail end of the coil, would travel through the tail phase toward the head.



This hydrodynamic trend is effectively utilized for performing the solvent extraction in two different manners: The tail phase is eluted from the head toward the tail of the coil, previously filled with the head phase. Alternatively, the head phase is eluted from the tail toward the head of the coil, previously filled with the tail phase. In either case, the mobile phase can travel quickly through the coil, leaving a large volume of the stationary phase in the coil.

This bilateral hydrodynamic system also permits simultaneous elution of the two solvent phases through the respective ends of the coil as shown at the bottom of the diagram. This dual countercurrent operation requires an additional flow tube at each end of the coil to collect the effluent and, if desired, a sample feed tube at the middle portion of the coil for continuous sample feeding.

Comparison of the above two hydrodynamic systems reveals that the bilateral hydrodynamic system (Figure 4, right) provides several advantages over the basic hydrodynamic system (Figure 4, left): the bilateral system gives a better retention of the stationary phase in the column and yields a higher partition efficiency in a given period of time due to more efficient phase mixing under a higher rotation speed of the coil. The system can also be applied to dual countercurrent operation, where the two solvent phases literally undergo countercurrent movement through a coiled column.

## Apparatus for Solvent Extraction

Three different types extraction instrument based on the bilateral hydrodynamic equilibrium are used: a slow rotary countercurrent apparatus (Figure 5) and two types of high-speed centrifuge systems, one for standard extraction (Figure 6) and the other for dual countercurrent extraction, with a spiral column

(Figure 7). All these instruments are free of rotary seals.

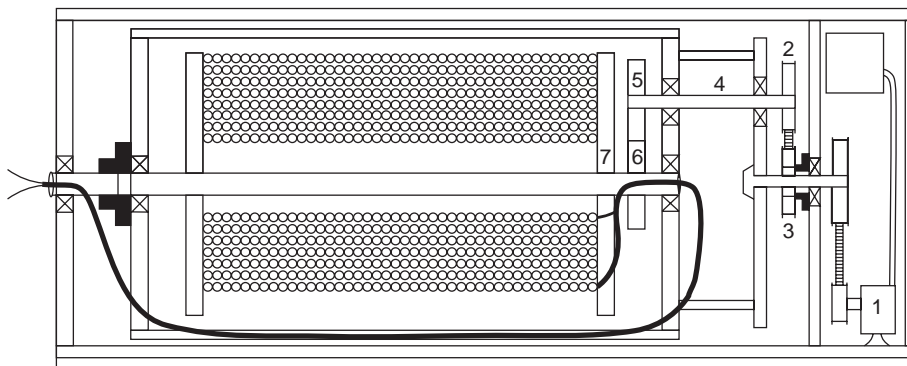
### Slow Rotary Countercurrent Chromatography Apparatus

The slow rotary countercurrent chromatography (CCC) apparatus holds a long column holder, which slowly rotates about its axis (Figure 5). The separation column is prepared by winding a long Teflon tube directly onto the holder hub, making multiple coiled layers. Two types of Teflon tube were used, standard tubing and convoluted tubing (similar to a miniature vacuum cleaner duct). A pair of flow tubes from each terminal is passed through the hole of the hollow central shaft and then making an arch supported by a lateral tube support and then exiting the centrifuge system through another hole on the central shaft.

The motor drives the rotary frame around the central axis of the centrifuge. Because the pulley at the right end of the countershaft is engaged to the stationary pulley with a toothed belt, the countershaft counter-rotates at the same rate on the rotating rotary frame. This motion is further conveyed to the column holder by a 1:1 gear coupling between the countershaft and the column holder. Consequently, the column holder rotates at the doubled speed with respect to the earth. This system allows the flow tubes to rotate without twisting, thus eliminating the need for rotary seals, and the high-speed centrifuge based on this rotary-seal-free system is widely used for apheresis at blood banks.

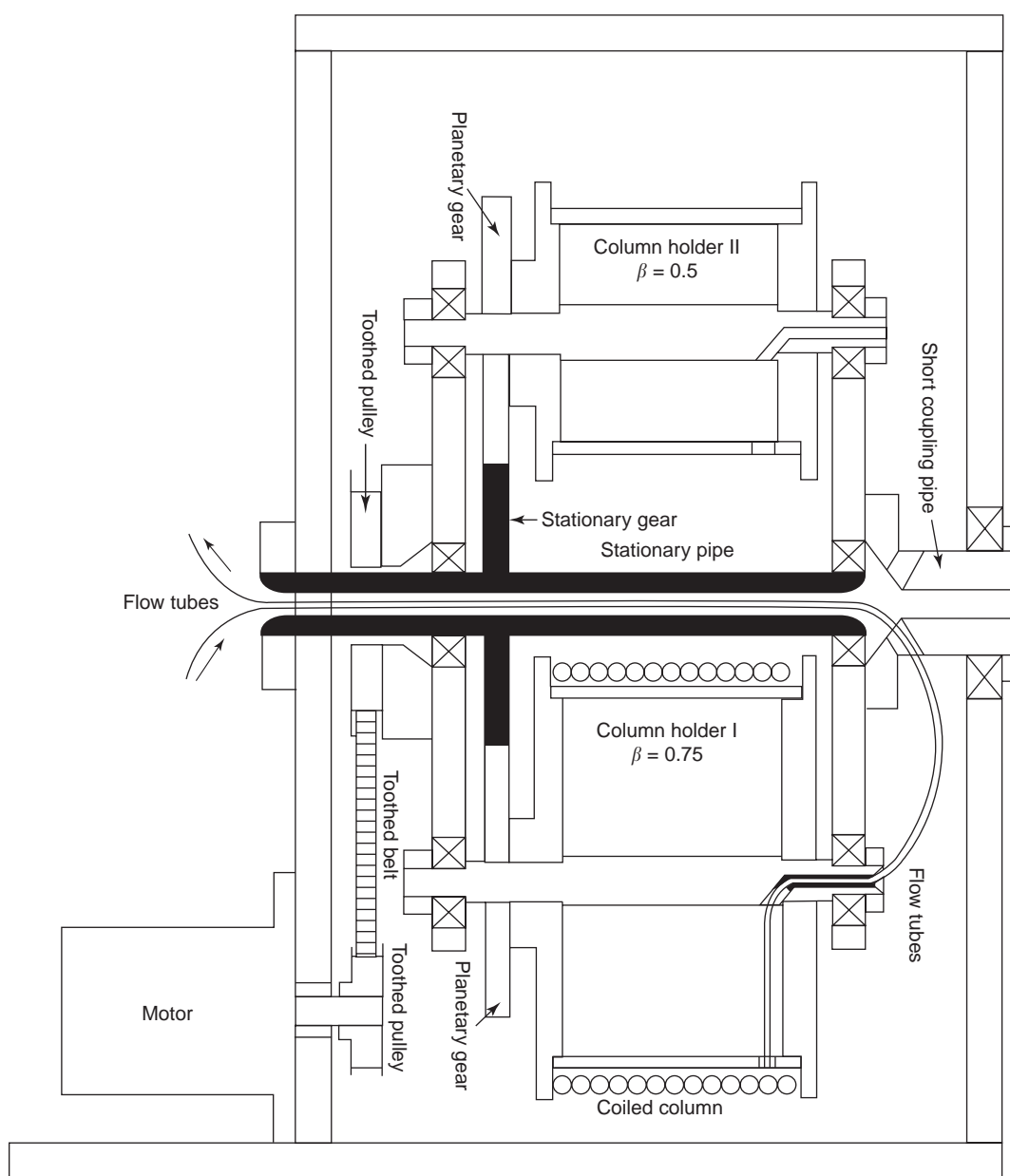
### Standard High-Speed CCC System

The multilayer coil assembly described above utilizes the Archimedean screw effect produced by the unit gravity. Thus, the relatively weak gravitational field limits the efficiency of the system. It has been found that the use of a centrifugal force field enhances the



**Figure 5** Cross-sectional view of seal-free slow rotary countercurrent chromatography (CCC) equipped with a large multilayer coil.





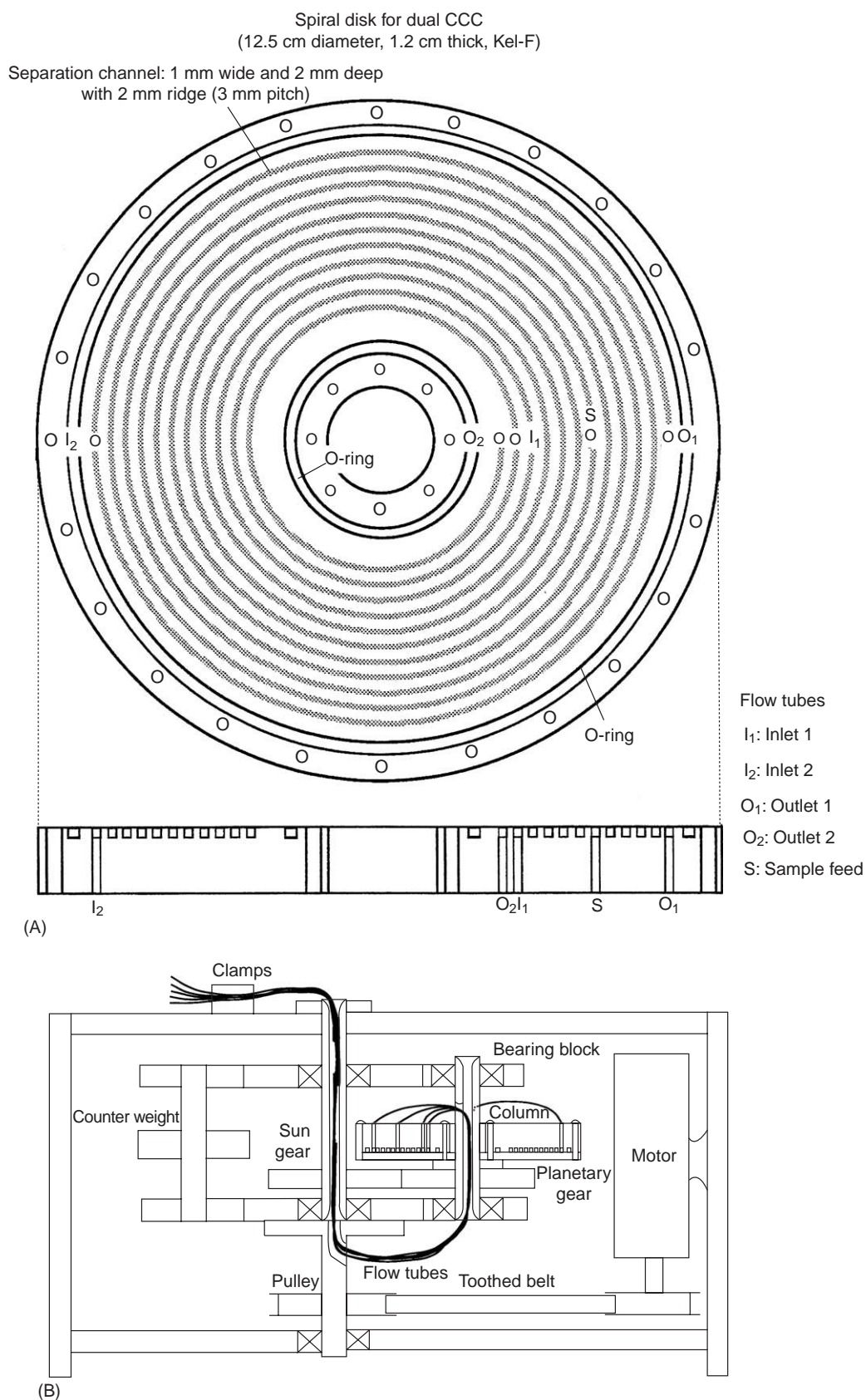
**Figure 6** Cross-sectional view of the coil planet centrifuge used for solvent extraction. One column holder (bottom) holds a coiled column, while the other (top) serves as a counterbalance.

partition efficiency in terms of both theoretical plate and the elution time.

Among various coil planet centrifuge systems developed in the 1980s and 1990s, the type-J synchronous system produces a particular mode of planetary motion that yields a bilateral hydrodynamic equilibrium of two solvent phases in a multilayer coil mounted coaxially around the holder. Consequently, the system is applied efficiently to both solvent extraction and CCC.

A cross-sectional view of the type-J coil planet centrifuge is illustrated in **Figure 6**. The motor (left,

bottom) drives the rotary frame via a pair of toothed pulleys and a toothed belt. The rotary frame holds a pair of column holders in symmetrical positions at a distance of 10 cm from the centrifuge axis. Each holder is equipped with a planetary gear that is engaged to an identical stationary sun gear (shaded) mounted rigidly around the central stationary pipe (shaded). This gear arrangement produces a planetary motion synchronous with the holder, i.e., one rotation about its own axis for each revolution around the central axis of the centrifuge in the same direction. A single-layer coiled column was mounted



**Figure 7** Instrumentation of dual countercurrent chromatograph. (A) Design of the spiral disk; (B) cross-sectional view of the apparatus.

around one column holder (lower), and the other holder (upper) served to counterbalance the centrifuge system. A pair of flow tubes from the separation coil first passes through the center hole on the holder shaft and then, by making an arch, reaches the side hole made on the short coupling pipe to enter the opening of the central stationary pipe. As mentioned elsewhere these tubes are not twisted during a centrifuge run.

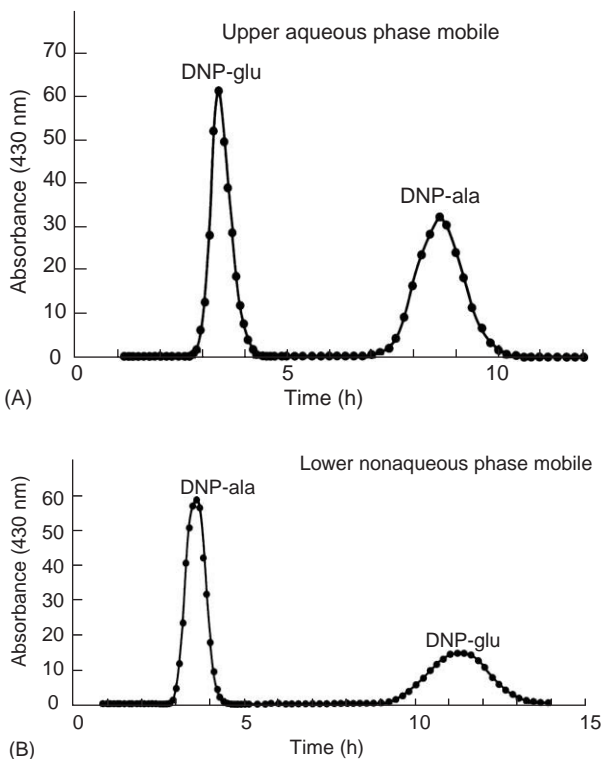
### Spiral Disk Dual Countercurrent Extraction Centrifuge

Figure 7 illustrates the new dual countercurrent extraction system (see Figure 4) using a spiral disk (A) mounted on the type-J synchronous planetary centrifuge system (B). The use of spiral channel configuration generates a radial centrifugal force gradient, enhancing the countercurrent movement of the two phases. This instrument is designed for separation of small amounts of pesticide in vegetable oil for mass spectrometric analysis. Because there is no stationary phase, multiple sample injections at regular intervals become possible without a risk of depleting the stationary phase or accumulating extracted oil, which would occur in the multilayer coil in the standard high-speed CCC technique.

### Solvent Extraction with a Rotating Coil in a Unit Gravity

Hydrodynamic studies (Figure 2) using the three coiled columns, with helical diameters ranging from 3 cm to 20 cm, demonstrated that the bilateral hydrodynamic distribution is established in all these coils  $\sim 100$  rpm. This result indicates that a long coiled column can be fabricated compactly by winding a single piece of plastic tubing around a spool-shaped rotary drum, making multiple coiled layers with dimensions as large as a 20 cm outside diameter (OD). The performance of this bilateral hydrodynamic extraction system in separating 2,4-dinitrophenyl (DNP) amino acids in a two-phase solvent system composed of chloroform, glacial acetic acid, and  $0.1 \text{ mol l}^{-1}$  hydrochloric acid at a volume ratio of 2:2:1 was examined. A large capacity multilayer coil was fabricated from a 30 m long, 5.5 mm ID fluorinated ethylene propylene (FEP) tube by winding it coaxially onto a 10 cm diameter, 25 cm wide spool support, making three layers of the coil with a total capacity of  $\sim 750$  ml.

Figure 8 shows the results of this preliminary separation, where both upper (A) and lower (B) phases were used as the mobile phase. In both separations two DNP amino acids are well resolved and eluted

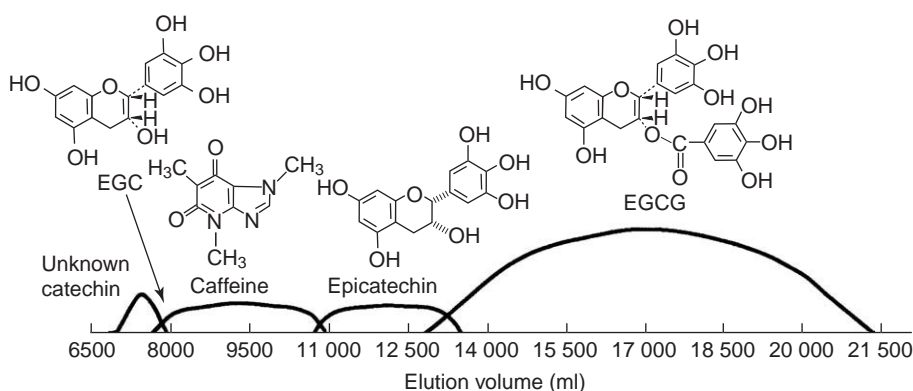


**Figure 8** Separation of DNP amino acids with a multilayer coil rotating in a unit gravity. Column: multilayer coil, 30 m long, 5.5 mm ID FEP tube with capacity 750 ml. Sample: DNP-glu and DNP-ala, 500 mg of each dissolved in 30 ml of solvent. Solvent system: chloroform/glacial acetic acid/ $0.1 \text{ mol l}^{-1}$  hydrochloric acid (2:2:1, v/v). Mobile phase: lighter aqueous phase (A) and heavier nonaqueous phase (B). Flow rate:  $516 \text{ ml h}^{-1}$ . Coil rotation: 80 rpm.

out as symmetrical peaks. The partition efficiency, expressed in terms of the theoretical plate number, is over 200, which is nearly equivalent to that obtained from 200 partition units in the Craig countercurrent distribution apparatus. This method can be applied to extraction or separation of various compounds using suitable two-phase solvent systems.

This slow rotary countercurrent extraction system has been applied for a large scale preparative separation of natural products using a multilayer coil prepared from convoluted PTFE tubing as shown in Figure 9. About 150 g amount of crude tea leaf extract was separated into four components in 72 h. In this separation, over 40 g of epigallocatechin gallate (EGCG) (fourth peak) was obtained at a high purity of 92.7%. The convoluted tubing provides some advantages over the standard tubing, such as ease of coil preparation and higher retention of the organic stationary phase.

This low-speed rotary coil extraction system has various desirable features for industrial applications: the sample loading capacity can be scaled up simply by increasing the diameter of the multilayer coil



**Figure 9** Separation of crude extract of tea leaves using low-speed countercurrent extraction apparatus equipped with a multilayer coil of convoluted tubing. Column: multilayer coil made of convoluted PTFE tubing, 200 m long, 8.5 mm average ID coiled  $\sim 9$  cm OD holder hub, forming seven layers, each consisting of 60 loops with a total capacity of 10 l (see **Figure 5**). Sample: 150 g of tea leaf extract dissolved in 1.2 l of solvent consisting of equal volumes of each phase. Solvent system: *n*-hexane/ethyl acetate/*n*-butanol/acetic acid/water (0.5:1:2:0.2:6, v/v). Mobile phase: lower aqueous phase. Elution mode: head to tail. Flow rate: 5 ml min<sup>-1</sup>. Column rotation: 21 rpm. Retention of the stationary phase: 33%.

and/or the width of the coil holder. The system also provides excellent safety features such as low rotation speed, low column pressure, and minimum risk of leakage of the solvent. Because of its simplicity, the system may be automated easily for long-term operation.

## Solvent Extraction with a Coil Planet Centrifuge

### Extraction of DNP Amino Acids

Continuous countercurrent extraction is efficiently performed using the high-speed coil planet centrifuge as shown in **Figure 6**. It enables extraction of a solute present in a large volume of the mobile phase into a small volume of the stationary phase retained in the coiled column. This requires a set of conditions such that the solute must favor partition to the stationary phase. With commonly used extraction media such as an ethyl acetate–aqueous system, partition coefficients of various biological materials can be adjusted conveniently by modifying the pH and/or ionic strength of the aqueous phase to meet the above requirement. For the model studies, a pair of DNP-amino acids, *N*-2,4-DNP-L-leucine (DNP-leu) and  $\delta$ -*N*-2,4-DNP-L-ornithine (DNP-orn), were selected as samples because they are readily observed through the column wall during the extraction process under stroboscopic illumination and also provide suitable partition coefficients.

A typical extraction procedure may be divided into three steps, i.e., extraction, cleaning, and collection. In each operation, the column was filled with the stationary phase and the mobile phase containing the

sample was eluted through the column in the proper direction while the apparatus was run at 600 rpm. The extraction process was continued until 400 ml of the mobile phase was eluted. Then the mobile phase was replaced by the same phase but free of solute to wash the column contents. This cleaning process was continued until the additional 100 ml of the mobile phase was eluted. This would elute out all impurities having a partition coefficient of 0.1 or greater. The sample extracted into the stationary phase in the coiled column was collected by eluting with the mobile phase in the opposite direction. The sample still remaining in the column was then washed out by eluting the column with the other phase originally used as the stationary phase. The degree of sample recovery was estimated by comparing the amount of the sample in the original mobile phase with that in the collected stationary phase.

The results of the experiments are summarized in **Table 1**. In experiments 1–3, DNP-leu dissolved in the aqueous mobile phase was extracted into  $\sim 10$  ml of the stationary nonaqueous phase. The sample recovery ranges from 94% to 100%. In experiments 4 and 5, DNP-orn dissolved in the nonaqueous mobile phase was extracted into the aqueous stationary phase. The sample recovery was in the range 97–100%. In practice, application of the method to aqueous crude extracts or physiological fluids requires a preliminary adjustment of the solvent composition for providing a suitable partition coefficient of the desired material for the applied pair of solvents. In this case, pre-equilibration of the two phases may not be essential. Experiment 6 shows an example of operation with such nonequilibrated solvents. The sample DNP-leu was first dissolved in

**Table 1** Experimental conditions and results of extraction

Exp. no.	Solvent system	Mobile phase	Stationary phase	Sample (P.C.) <sup>a</sup>	Sample conc. in mobile phase (mg%)	Extracted mobile phase volume (ml)	Flow rate (direction)	RPM	Collected stationary phase volume (ml)	Sample recovery (%)
1	Ethylacetate 1 0.5 mol l <sup>-1</sup> NaH <sub>2</sub> PO <sub>4</sub> 2	Aqueous	Nonaqueous	DNP-leu ( $<0.01$ )	4	400	516 ml h <sup>-1</sup> (Head-tail)	600	10.5	94
2	0.5 mol l <sup>-1</sup> NaH <sub>2</sub> PO <sub>4</sub> 2	Aqueous	Nonaqueous	( $<0.01$ )	0.4	400	516 ml h <sup>-1</sup> (Head-tail)	600	10.0	97
3	0.5 mol l <sup>-1</sup> NaH <sub>2</sub> PO <sub>4</sub> 2	Aqueous	Nonaqueous	( $<0.01$ )	0.04	400	516 ml h <sup>-1</sup> (Head-tail)	600	10.4	100
4	Ethylacetate 2 0.5 mol l <sup>-1</sup> NaH <sub>2</sub> PO <sub>4</sub> 1	Nonaqueous	Aqueous	DNP-orn ( $<0.01$ )	0.4	400	516 ml h <sup>-1</sup> (Tail-head)	600	11.8	97
5	0.5 mol l <sup>-1</sup> NaH <sub>2</sub> PO <sub>4</sub> 1	Nonaqueous	Aqueous	( $<0.01$ )	0.04	400	516 ml h <sup>-1</sup> (Tail-head)	600	11.8	100
6	Nonequilibrium system	5% Ethylacetate in 0.5 mol NaH <sub>2</sub> PO <sub>4</sub>	Ethylacetate	DNP-leu ( $<0.01$ )	0.4	400	516 ml h <sup>-1</sup> (Head-tail)	600	6.1	99

<sup>a</sup> Partition coefficient is defined as solute concentration in the mobile phase divided by that in the stationary phase.

400 ml of 0.5 mol l<sup>-1</sup> NaH<sub>2</sub>PO<sub>4</sub> aqueous solution containing ethyl acetate at 5%, which is slightly below the saturation level of ~7%. The column was filled with ethyl acetate followed by elution with the above sample solution. Both extraction and cleaning processes were performed as in other experiments. The sample solution collected from the column measured slightly over 6 ml. This depletion of the stationary phase resulted apparently from use of the nonequilibrated solvent pair but without any effect on the sample recovery.

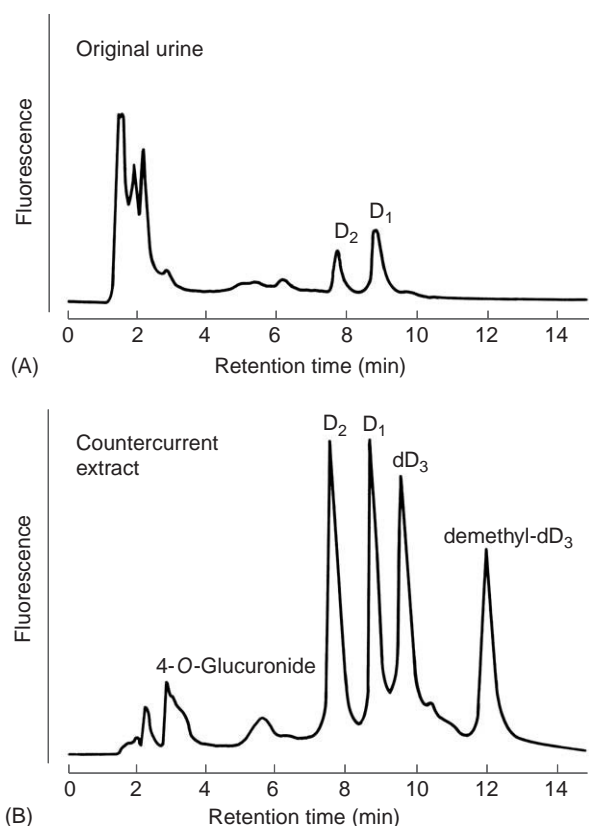
The overall results indicate the potential usefulness of the present method in processing large amounts of crude extracts or biological fluids in research laboratories. A small amount of the sample present in several hundred milliliters of the original solution can be enriched in 10 ml of the nonaqueous phase free of salt in 1 h, at a high recovery rate.

### Extraction of Urinary Drug Metabolites

The present method has been applied to extraction of urinary metabolites of daunorubicin, an anticancer drug. The extraction was performed with a two-phase solvent system composed of *n*-butanol/0.3 mol l<sup>-1</sup> Na<sub>2</sub>HPO<sub>4</sub>. Prior to the extraction, the urine sample was saturated with *n*-butanol, and then Na<sub>2</sub>HPO<sub>4</sub> was added at a concentration of 0.3 mol l<sup>-1</sup>. In each experiment, the column was first filled with *n*-butanol. Then, the apparatus was rotated at 650 rpm while aqueous Na<sub>2</sub>HPO<sub>4</sub> saturated with *n*-butanol was pumped into the column to equilibrate the stationary phase. The prepared urine sample, 1–2 l in volume, was then eluted through the column at a flow rate of 500–700 ml h<sup>-1</sup> with a metering pump. After all the sample solution was eluted, the column was cleaned by eluting with 100 ml of aqueous, *n*-butanol-saturated 0.3 mol l<sup>-1</sup> Na<sub>2</sub>HPO<sub>4</sub>. Then, the centrifuge run was terminated and the retained *n*-butanol phase was drained from the column by connecting the column inlet to an N<sub>2</sub> line under pressure. Several milliliters of *n*-butanol were flushed through the column to recover any remaining sample. The *n*-butanol extracts were combined and evaporated to dryness by flash evaporation.

The results of the experiments are illustrated in Figure 10, where the high-performance liquid chromatography (HPLC) analysis of the original urine sample and that of the countercurrent extract are compared. In Figure 10A, the chromatogram of the original urine sample shows a large amount of hydrophilic material at the solvent front and two metabolites peaks, D<sub>1</sub> and D<sub>2</sub>. As shown in Figure 10B, the chromatogram of the countercurrent extract with the coil planet centrifuge reveals enriched D<sub>1</sub> and D<sub>2</sub>





**Figure 10** Extraction of urinary metabolites of daunorubicin using the coil planet centrifuge. (A) HPLC analysis of the original urine; (B) HPLC analysis of the countercurrent extract.

peaks and three additional metabolite peaks, dD<sub>3</sub>, demethyl-dD<sub>3</sub>, and 4-O-glucuronide as indicated in the chromatogram.

See also: **Countercurrent Chromatography: Overview.**

## Further Reading

- Ito Y (1981) New continuous extraction method with a coil planet centrifuge. *Journal of Chromatography* 207: 161–169.
- Ito Y and Bhatnagar R (1981) Preparative countercurrent chromatography with a rotating coil assembly. *Journal of Chromatography* 207: 171–180.
- Nakazawa H, Riggs CE Jr., Egorin MJ, *et al.* (1984) Extraction of urinary anthracycline antitumor antibiotics with the coil planet centrifuge. *Journal of Chromatography* 307: 323–333.
- Ito Y (1988) Studies on hydrodynamic distribution of two immiscible solvent phases in rotating coils. *Journal of Liquid Chromatography* 11: 1–19.
- Ito Y and Conway WD (eds.) (1996) *High-Speed Countercurrent Chromatography*. New York: Wiley Interscience.
- Menet JM and Thibaut D (eds.) (1999) *Countercurrent Chromatography*. New York: Dekker.
- Du Q-Z, Wu P-D, and Ito Y (2000) Low-speed rotary countercurrent chromatography using a convoluted multi-layer helical tube for industrial separation. *Analytical Chemistry* 72: 3363–3365.
- Berthod (ed.) (2002) *Countercurrent Chromatography: The Support-Free Liquid Stationary Phase*. Amsterdam: Elsevier.
- Du Q-Z and Ito Y (2003) Slow rotary countercurrent chromatography. *Journal of Liquid Chromatography and Related Technologies* 26(11): 1827–1838.
- Du Q-Z, Winterhalter P, and Ito Y (2003) Large convoluted tubing for scale-up of slow rotary countercurrent chromatography. *Journal of Liquid Chromatography and Related Technologies* 26(12): 1991–2002.
- Ito Y (2003) *Spiral Disk for Dual Countercurrent Extraction*. US Patent pending.

## CRMs

See **QUALITY ASSURANCE: Reference Materials; Production of Reference Materials**

## CSV

See **VOLTAMMETRY: Cathodic Stripping**

## CYCLIC VOLTAMMETRY

See **VOLTAMMETRY: Linear Sweep and Cyclic**



# D

## DAIRY PRODUCTS

See **FOOD AND NUTRITIONAL ANALYSIS: Dairy Products**

## DERIVATIZATION OF ANALYTES

**S Görög**, Chemical Works of Gedeon Richter Ltd., Budapest, Hungary

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The derivatization of analytes is very important in several branches of analytical chemistry. It expands the fields of application of various spectroscopic techniques (ultraviolet-visible (UV-vis), fluorimetry, nuclear magnetic resonance (NMR), and mass spectroscopies), and in several cases increases also the selectivity and sensitivity of these techniques. Derivatization is also an inevitable tool in all chromatographic and electrophoretic techniques. In gas chromatography (GC), the main importance of derivatization is the improvement of the volatility/thermal stability of the analytes, and in all of the discussed separation techniques it has the potential of increasing the selectivity of the separation (including enantiomeric separations) and the sensitivity of the detection.

### Definition: Historical Background

In a broad sense, all chemical transformations of the analyte taking place in the course of an analytical procedure can be considered to be derivatization reactions. These methods can be classified as follows:

1. Practically all kinds of classical analytical procedures (acid-base, redox, precipitation, and complexometric titrations as well as gravimetry) fall into this category. An example from the field of titrations is the transformation of acetylsalicylic acid to sodium acetylsalicylate with sodium hydroxide as the titrant, or to sodium salicylate and sodium acetate if the method is based on the

hydrolysis of the ester group in the analyte followed by titration of excess sodium hydroxide with hydrochloric acid. Further examples are the reduction of iron(III) to iron(II) with iodide ions in the course of its iodometric determination, the precipitation reaction between chloride and silver ions taking place during the determination of chloride, or the transformation of a wide range of metal ions to the ethylenediamine tetraacetate complexes during their complexometric determination. Examples for gravimetric assays are the determination of sulfate after transformation to barium sulfate and determination of various aldehydes and ketones after their transformation to the 2,4-dinitrophenylhydrazones. The discussion of these classical reactions will not be the subject of this article.

2. After the introduction of instrumental analytical methods several possibilities became available where no chemical reactions were involved in the complex analytical procedure. Methods of this kind are various spectroscopic and spectrophotometric methods, GC with thermal conductivity or electron-capture detector, high-performance liquid chromatography (HPLC) and capillary electrophoresis (CE) with UV or fluorimetric detector, etc. In other cases chemical reaction takes place in the detector of the instruments as a step of the complex procedure leading to the analytical signal. Examples are the combustion of the analyte in the flame ionization detector of gas chromatographs, their oxidation, reduction, or other transformation in various electrochemical sensors and detectors, fragmentation of the molecules in mass spectrometers, etc. These reactions will also not be discussed in this article.
3. Derivatization in the up-to-date sense of the word is the transformation of the analyte by a chemical

reaction to a modified structure either outside or inside the analytical instrument prior to the formation of the analytical signal with the aim of expanding the application field of the analytical method to the analyte or increasing the selectivity or the sensitivity of the method. Chemical derivatization is not restricted to the formation of covalent bonds: the formation of ion pairs, complexes, and adducts that can greatly improve the selectivity of various analytical methods can also be considered to be derivatization reactions. These reactions – leading to covalently and noncovalently bound derivatives alike – will be the subject of this article. (Note: a widely accepted criterion of a derivatization reaction in analytical chemistry is that the molecule of the analyte is transformed to a larger, more complex molecule. This is, however, not applicable in all cases: reduction of ketones to hydroxyl derivatives, oxidation of the latter to ketones, acid-catalyzed elimination of water or carboxylic acids to form (conjugated) double bonds, hydrolysis of esters, etc., can also be considered to be (retro)-derivatization reactions.)

## Spectroscopic and Spectrophotometric Techniques

### UV-Visible Spectrophotometry

Although UV-vis spectrophotometers were already available in the 1910s, their use was limited to a few laboratories only. However, filter photometers operating only in the visible spectral region found wide-ranging applications. Thus, it was the precondition of their application to colorless compounds that the analytes were converted to colored derivatives prior to the measurement. Innumerable methods were developed in the following decades based on this principle with sometimes well established but in many cases obscure chemical background. After the spread of UV-vis spectrophotometers in the middle of the twentieth century the importance of these methods decreased considerably, but spectrophotometric/colorimetric methods based on chemical reactions have retained some importance among others in the field of metal and pharmaceutical analysis up to the present time. The reasons for this are that this is the only possibility for the spectrophotometric determination of UV-inactive materials in the form of UV-active or colored derivatives, the selectivity and sensitivity can often be increased by the formation of spectrophotometrically highly active derivatives ( $\epsilon \sim 10\,000\text{--}50\,000\text{ l mol}^{-1}\text{ cm}^{-1}$ ), and the transformation of analyte to colored derivatives is often the prerequisite of using automatic analyzers equipped

with inexpensive filter photometers mainly in the clinical analytical practice.

Some examples are the determination of iron(II) as the 2,2'-dipyridyl or 1,10-phenanthroline complex, iron(III) in the form of the thiocyanate complex, various metals (e.g., Cu(II), Ag(I), Pb(II), Zn(II), and Hg(II)) in the form of dithizone complexes, etc. (The formation of colored derivatives of metal ions with various reagents is useful in their determination by ion chromatography with a UV-vis detector.) Of the pharmaceutical applications that are still in use, typical examples are the determination of drugs with suitable functional groups after chelation reaction with various metal ions; e.g., the determination of salicylic acid impurity in acetylsalicylic acid as its complex with iron(III), determination of alkaloids and other basic materials after ion-pair formation with various acidic dyes, diazotization and azo coupling for the determination of aromatic amines and phenols, condensation of 4-ene-3-oxosteroids with isonicotinoyl hydrazine, indirect determination of corticosteroids with reducing  $\alpha$ -ketol-type side chain using tetrazolium reagents. Enzymatic reactions excel due to their high specificity. An example is the determination of paracetamol (4-acetaminophenol) in serum based on enzymatic hydrolysis to 4-aminophenol followed by oxidative coupling with *o*-cresol to form a colored indophenol derivative.

A more up-to-date application field of UV-vis spectrophotometry is flow injection analysis where the spectrophotometer is the generally used detector. In this technique, the sample and the reagent(s) are injected into a continuous, unsegmented flow produced by a pump. The reaction takes place in a temperature-controlled reaction coil, and the reaction product producing sharp peaks is measured in the detector.

### Fluorimetry

Compounds not having (sufficiently intense) native fluorescence can be transformed with suitable reagents to fluorescent derivatives enabling high sensitivity to be obtained in the course of their fluorimetric determination.

Two general types of these reactions can be mentioned:

1. The structure of the analyte is changed to a fluorescent derivative by inorganic reagents, mainly oxidizing agents. For example, morphine is oxidized by ferricyanide reagent to pseudomorphine, enabling its highly sensitive and selective determination in the presence of related derivatives such as codeine, dihydromorphine, diacetylmorphine, and apomorphine. Similarly, selective and sensitive determination is obtainable when vitamin B<sub>1</sub>

(thiamine) is oxidized with the same reagent to the highly fluorescent thiochrome derivative.

2. More general is the applicability of the other type of reactions where any kind of molecules with suitable functional groups can be transformed to highly fluorescent derivatives by means of a derivatization reaction with suitable fluorescent reagents. In the majority of cases the fluorescence of the excess reagent excludes the possibility of the direct determination of the derivatized analyte. These derivatization reactions are useful tools for the determination of various compounds by HPLC using fluorescence detection. This explains why at present much attention is devoted to the development of fluorogenic derivatizations in which the reagent is not fluorescent itself while the derivative exerts strong fluorescence. A classical example is fluorescamine, which is suitable for the very sensitive determination of primary amines (see Scheme 1).

Time-resolved fluorimetry, a modern variant of fluorimetry, excels with its high specificity. This technique is based on the derivatization reaction of various compounds with europium(III) to form highly fluorescent chelates with extremely long decay time. Using a suitable instrument the long decay time enables the long-lived fluorescence of the derivatized analyte to be resolved from the short-lived fluorescence of the background and interfering components in biological samples.

### NMR Spectroscopy

Hydrogen–deuterium exchange taking place between labile protons (e.g., OH–, NH–, COOH–, or acidic CH) of the analyte and deuterated solvents (D<sub>2</sub>O, CD<sub>3</sub>OD, CF<sub>3</sub>COOD, etc.) can be considered a simple derivatization reaction. This is a very useful tool in the assignment of signals in the respective groups and environments and is therefore part of the everyday routine both in <sup>1</sup>H and <sup>13</sup>C NMR spectroscopies. Further possibilities are also available for NMR spectroscopists, based on complex or adduct formation, to solve a wide variety of problems ranging

from signal assignment through the determination of absolute configuration of the analyte to the investigation of its molecular dynamic properties. In addition to the classical shift reagent Eu(acetylacetonate)<sub>3</sub> complex and others, chiral shift reagents (where the ligand of the Eu<sup>3+</sup> complex is an enantiomerically pure chiral compound) and especially cyclodextrins are extremely useful tools in estimating enantiomer excess or even enantiomeric purity of chiral compounds.

### Mass Spectrometry

See sections ‘Gas chromatography and GC–MS’ and ‘High-performance liquid chromatography and LC–MS’.

## Chromatographic and Electrophoretic Techniques

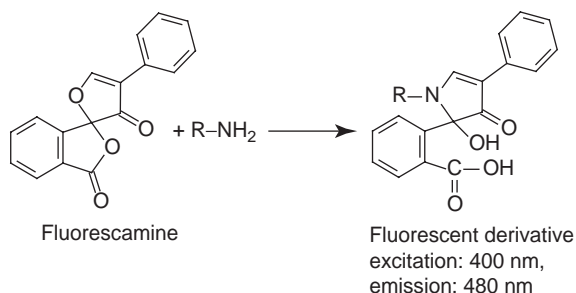
### Thin-Layer Chromatography

Although (*in situ*) prechromatographic derivatization of the analytes in thin-layer chromatography (TLC) has been described in several cases, much more important are the postchromatographic reactions with various reagents.

The aim of postchromatographic derivatization is to make the separated compounds visible at daylight or to transform them to fluorescent derivatives, thus enabling primarily the sensitivity and often also the selectivity of the detection and of the densitometric determination to be increased.

There are several modes to expose the separated compounds to the reagent. Of the vapor-phase reagents only iodine vapor is widely used. The incorporation of the reagent in the mobile or stationary phases has been reported only in a limited number of cases. The two main techniques are spraying the plate with relatively concentrated reagent solutions and dipping it into less concentrated solutions.

Some of the innumerable reagents used in TLC are based on reactions with more or less well-established mechanism. For example, sodium iodobismuthate (Dragendorff reagent) is widely used among others for alkaloids and quaternary ammonium compounds, 4-dimethylaminobenzaldehyde for primary amines and amino acids, 2,4-dinitrophenylhydrazine for aldehydes and ketones, ninhydrin for amino acids and some antibiotics, fluorescamine for primary and secondary amines, phosphomolybdic acid for lipids, various steroids, and other compounds, chlorine vapor followed by KI/starch for amines and amides. More complex is the mechanism of the reactions with some other reagents, containing high concentrations of sulfuric acid, vanillin/sulfuric acid, phosphoric acid, aluminum chloride, antimony(III)



Scheme 1

chloride, etc., which produce various colors and fluorescence upon heating the plate in the course of the analysis of various classes of organic compounds, e.g., steroids.

### Gas Chromatography and GC-MS

Although many organic compounds can be chromatographed in the gas phase without derivatization, it still plays an important role here also. The aim of derivatization prior to the GC-MS analysis is

- expanding the capabilities of GC by blocking the polar functional groups in the analytes, thus enabling compounds that are not sufficiently volatile to be analyzed;
- improving the peak shape of polar compounds;
- improving the sensitivity of the detection and quantitation; and
- improving the selectivity by enabling the separation of compounds that are not sufficiently separated in the underivatized form.

The most frequently derivatized functional groups are hydroxy, carboxy, and amino groups.

**Derivatization of hydroxy compounds** The most widely used derivatization reaction is their transformation to silyl (in the majority of cases trimethylsilyl) ether derivatives. Hexamethyldisilazane is the classical and still widely used silylating agent (usually mixed with trimethylchlorosilane, which is the catalyst generally used for various silylation reactions). For less reactive (sterically hindered) hydroxy groups more reactive reagents (*N,O*-bis(trimethylsilyl)acetamide, *N,O*-bis(trimethylsilyl)trifluoroacetamide, *N*-methyl-*N*-trimethylsilyltrifluoroacetamide, trimethylsilylimidazole, etc.) are used. In some cases (especially in quantitative analysis by GC-MS), it is useful to replace trimethylsilyl by *t*-butyldimethylsilyl group due to the easy formation of the abundant ( $M - 57$ )<sup>+</sup> peak.

Another important derivatization method of hydroxy compounds is their acylation with anhydrides of acetic, trifluoroacetic, heptafluorobutyric, etc. The latter enable very sensitive quantification to be performed using the electron-capture detector.

**Derivatization of amines** Any of the above described silylation and acylation reactions can be adopted for the derivatization of primary and secondary amines in the form of their *N*-silyl and *N*-acyl derivatives.

**Derivatization of carboxylic acids** Although the carboxyl group can also be transformed to the silyl

(ester) derivative, due to their hydrolytic instability the most widely used derivatization method here is the formation of methyl esters with methanol/hydrochloric acid at elevated temperature or with diazomethane. Double derivatization is often necessary, e.g., esterification of the carboxyl group and acylation of the amino group of amino acid and esterification of the carboxyl group and silylation of the hydroxy group(s) of bile acids.

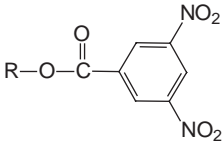
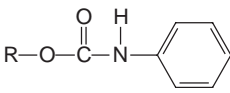
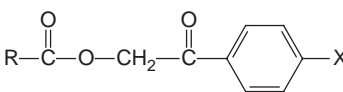

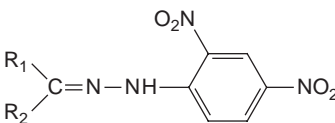
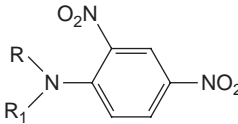
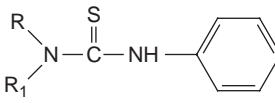
**Derivatization of inorganic materials** Various anions can be separated by GC after derivatization with, for example, pentafluorobenzyl *p*-toluenesulfonate. A variety of metals can be made volatile for GC separation by means of fluorinated  $\beta$ -diketones as the chelating agents.

### High-Performance Liquid Chromatography and LC-MS

**UV-vis derivatization** The main advantage of HPLC (introduced at the end of the 1960s) was claimed to be its much wider application field than that of GC: there are no limitations due to thermal instability, low volatility, high molecular weight, and very high polarity. This is why derivatization is much less important in HPLC than in GC. In spite of this, derivatization in HPLC is as old as the technique itself. The reason for this is that due to the low UV activity of several organic compounds the sensitivity of detection obtainable with the generally used UV detector is often not sufficient. Several reagents for the pre- or postcolumn derivatization of analytes to colored or highly UV-active derivatives have been described. Some of these are summarized in Table 1.

**Fluorimetric derivatization** Much more important is fluorimetric derivatization. The reason for this is that in biomedical analysis many of the native compounds, drugs, and drug metabolites are present at very low concentration (pg to ng ml<sup>-1</sup>) excluding the possibility of using the insensitive UV detector both for underivatized and derivatized analytes. The high sensitivity of fluorimetric detection enables the limit of detection to be decreased by several orders of magnitude. It should be noted, however, that for complex matrices as, for instance, dealt with in environmental samples the chemistry as such becomes the crucial step: the derivatization has to be successful for very low analyte concentrations in the presence of many possible matrix interferences. Nonetheless, HPLC with a fluorimetric detector is a very useful tool in biomedical analysis. Since only a limited number of analytes possess sufficiently strong native

**Table 1** Some derivatization reactions for HPLC/UV detection

Functional group	Reagent	Derivative
-OH	3,5-Dinitrobenzoyl chloride	Aroyl ester
		
-OH	Phenyl isocyanate	Phenyl carbamate
		
-COOH	Phenacyl bromide	Phenacyl ester
		
		X = H, Br, NO <sub>2</sub>
	2,4-Dinitrophenylhydrazine	Phenylhydrazone
		
		R <sub>1</sub> = alkyl, aryl; R <sub>2</sub> = H, alkyl, aryl
-NH <sub>2</sub> , -NHR <sub>1</sub>	1-Fluoro-2,4-dinitrobenzene	Dinitrophenylamine
		
		R <sub>1</sub> = H, alkyl
-NH <sub>2</sub> , -NHR <sub>1</sub>	Phenyl isothiocyanate	Phenyl thiourea
		
		R <sub>1</sub> = H, alkyl

fluorescence, derivatization with strongly fluorescent reagents is very important. Some characteristic examples are shown in **Table 2**. It has to be noted that the number of commercially available reagents of this type is over 100.

Laser-induced fluorescence detectors enable especially sensitive determinations to be carried out. In addition to the above-mentioned classical derivatizations, diode lasers with emission in the red or near-infrared (>630 nm) require special derivatizing agents (reactive dicarbocyanine-type dyes) for carboxylic acids, amines, and thiols.

**Chemiluminescence derivatization** Derivatization enables chromatographically separated organic compounds to be determined by a selective and sensitive chemiluminescence detector. The use of isoluminol as a chemiluminescence label is an

interesting combination of pre- and postcolumn derivatization. After the HPLC separation of the derivatives of various amines and carboxylic acids with *N*-(4-aminobutyl)-*N*-ethylisoluminol, the separated components are reacted in a postchromatographic reactor with hydrogen peroxide and hexacyanoferrate in alkaline solution to obtain the chemiluminescence. Lucigenin and peroxyoxalate systems have also been used for the analysis of amino acids, steroids, etc.

**MS derivatization** As a consequence of its extremely high specificity and sensitivity, HPLC combined with MS(MS) detection is the most widely used technique in all cases when the requirement is the detection, identification, and quantitation of minor organic components in complex matrices (biomedical,

**Table 2** Some derivatization reactions for HPLC/fluorescence detection

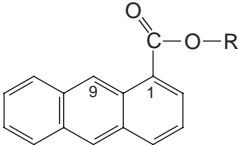
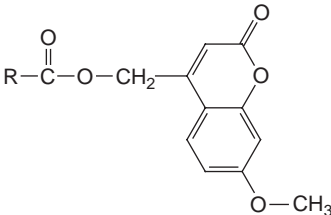

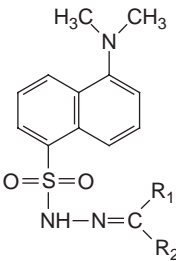
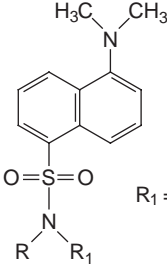
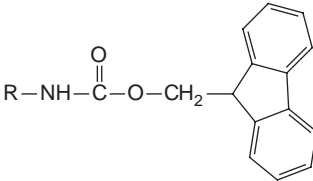
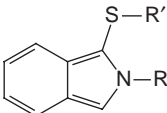
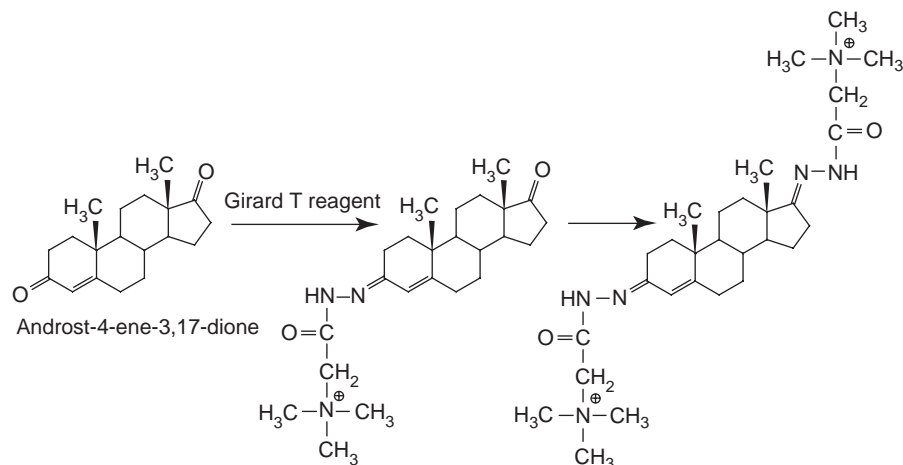
Functional group	Reagent	Derivative	
–OH	1- or 9-anthroynitrile	Anthroyl ester	
–COOH	4-Bromomethyl-7-methoxycoumarine	4-acyloxymethyl-7-methoxycoumarine	
	Dansyl hydrazine	Dansyl hydrazone	 <p><math>R_3 = \text{alkyl, aryl}; R_2 = \text{H, alkyl, aryl}</math></p>
–NH <sub>2</sub> , –NHR <sub>1</sub>	Dansyl chloride	Dansyl amide	 <p><math>R_1 = \text{H, alkyl}</math></p>
–NH <sub>2</sub> , –NHR <sub>1</sub>	Fluorescamine	See <b>Scheme 1</b>	
–NH <sub>2</sub> (amino acid)	9-Fluorenylmethyl chloroformate	Fluorenylmethyl carbamate	
–NH <sub>2</sub> (amino acid)	o-Phthaldialdehyde + thiol	Isoindole	



Table 2 Continued

Functional group	Reagent	Derivative
-SH	Bromobimanes	Bimane sulfide

$X = \text{H}, \text{Br}, \text{N}(\text{CH}_3)_3$



Scheme 2

environmental, food analyses). Although in the majority of cases no derivatization is necessary to obtain the desired selectivity and sensitivity, in many cases the ionization efficiency of some analytes in soft ionization MS is not sufficient, leading to poor sensitivity. The solution to the problem is derivatization introducing permanently charged or easily ionizable moieties for electrospray ionization MS as well as proton-affinitive and electron-affinitive groups for the positive and negative atmospheric pressure chemical ionization MS, respectively. Examples are derivatization of alcohols with 2-fluoro-1-methylpyridinium, *p*-toluenesulphonate, or ferrocenoyl azide and ketones with 2-nitro-4-trifluoromethylphenylhydrazine or Girard T or P reagents (carboxymethyltrimethylammonium chloride hydrazide and carboxymethylpyridinium chloride hydrazide). The reaction scheme of the reaction of a ketosteroid is depicted in Scheme 2.

**Electrochemical derivatization** The high sensitivity attainable in HPLC with the aid of electrochemical (usually amperometric) detectors prompted the transformation of nonelectroactive analytes into electroactive derivatives. Some of the reagents or

reagent types described for the electrochemical pre- or postcolumn derivatization have already been mentioned in the previous sections dealing with other HPLC derivatization methods. For example, the isoindole derivatives obtained when primary amines or amino acids are derivatized with the *o*-phthaldialdehyde + thiol reagent are active in the oxidative mode. The same applies to the thiocarbamide derivatives of amines (reagents: isothiocyanates) and the derivatives of alcohols and amines with various reagents containing the ferrocene moiety.

The most frequently used derivatizing agents introducing electroreducible groups contain aromatic nitro moiety. Examples are 2,4,6-trinitrobenzenesulfonic acid, 1-fluoro-2,4-dinitrobenzene, and 3,6-dinitrophthalic anhydride for the derivatization of amines, 4-nitro- or 2,4-dinitrophenylhydrazine reagents for aldehydes and ketones, 4-nitrobenzyl bromide and 4-nitrophenacyl bromide for carboxyl groups, etc.

### Supercritical Fluid Chromatography

The importance of supercritical fluid chromatography (SFC) is not comparable with that of the previously discussed chromatographic methods. The aim of

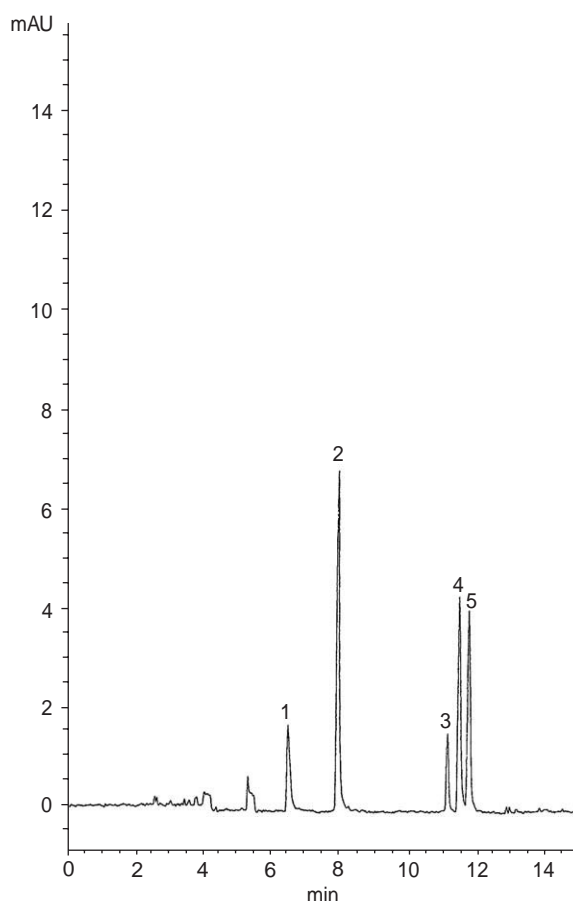
derivatization here is either to increase the solubility of highly polar analytes (e.g., carbohydrates) in CO<sub>2</sub> by permethylation, persilylation, peracetylation, or to improve the selectivity and mainly the detection limit using various detectors making use of the derivatization strategies described in the sections dealing with GC and HPLC.

### Capillary Electrophoresis and Related Techniques

Derivatization is often used in CE either to introduce electric charge to the uncharged analyte molecule or to improve the sensitivity of its detection and quantitation.

**Introduction of electric charge to the analyte molecule** In the course of the electrophoretic procedure, only electrically charged molecules migrate separately, toward the cathode or the anode; uncharged molecules move together toward the cathode driven by the electroosmotic flow. These molecules can be separated in the CE instrument by the application of micellar electrokinetic chromatography (MEKC). Another method (much less widely used but still remarkable) is to introduce electric charge to the molecule by means of suitable derivatization reactions prior to the CE run. Examples are reductive amination of carbohydrates using 2-aminopyridine as the reagent, which leads to charged molecules in acidic medium and the derivatization of ketosteroids with Girard P and T reagents (already mentioned in section “MS derivatization”). **Figure 1** shows the electropherogram of ethisterone and its synthetic precursors after derivatization with Girard T reagent. The two-step reaction of the intermediate diketo steroid androstenedione is shown in **Scheme 2**.

**Improving sensitivity** While the above discussed aspect of derivatization in CE aiming at introducing electric charge to the uncharged analytes is only seldom used, the other aspect, namely the improvement of the sensitivity by introducing chromophores and fluorophores, is of great importance. What is described about the potential of pre- or postcolumn derivatization in the section ‘High-performance liquid chromatography and LC–MS’ is also applicable to CE and related techniques (capillary electrochromatography (CEC) and MEKC): the same or similar derivatization reactions can be used. The result is extremely high sensitivity, enabling even the most delicate bioanalytical methods to be solved especially if the laser-induced fluorimetric detection with argon ion or helium–cadmium lasers is used. For example, as low a limit of detection as  $10^{-18}$  mol per injection was measured after the derivatization of peptides



**Figure 1** CE separation of ethisterone (17 $\alpha$ -ethynyl-17-hydroxy-androst-4-ene-3-one) and its synthetic precursors after derivatization with Girard T reagent. Key: 1, Girard T reagent; 2, androst-4-ene-3,17-dione-bis-Girard T derivative; 3, androst-4-ene-3,17-dione-3(mono)-Girard T derivative; 4, 3 $\beta$ -hydroxy-androst-5-ene-17-one Girard T derivative; 5, ethisterone Girard derivative. pH 4.8; voltage, 25 kV; temperature, 15°C; injection time, 3 s; UV detector: 250 nm. (Reprinted with permission from Görög S, Gazdag M, and Kemenes-Bakos P (1996). *Journal of Pharmaceutical and Biomedical Analysis* 14: 1115–1124; © Elsevier.)

with suitable dye reagents such as tetramethylrodamine 5-isocyanate isomer G.

### Immunoanalytical Methods

Immunoanalytical methods occupy an important position in bioanalytical chemistry. Chemical derivatization plays a fundamental role in this technique. Small analyte molecules (haptens) are covalently bound to proteins to raise the antibody, which is the basis for their highly selective and sensitive assay. Various labeled derivatives of the analyte are then prepared for competitive binding on the antibody (radiolabeling, labeling for enzyme immunoassay, fluorescence immunoassay, fluorescence polarization immunoassay, and luminescence immunoassay). These derivatization reactions are carried out by the

manufacturers of the various commercially available kits rather than the analysts themselves. However, in the case of enzyme immunoassays the basis for the measurement is usually chemical reactions, the products of which are measured spectrophotometrically or fluorimetrically. Two examples of the great variety of reactions used in the various assay methods are the reaction between hydrogen peroxide and 2-phenylenediamine catalyzed by peroxidase enzyme leading to a quinonediimine chromophore. If the reaction partner of hydrogen peroxide in this reaction is phenylacetic acid and the reaction product is further reacted with glycine, a fluorophore is formed enabling much higher selectivity and sensitivity to be attained especially if time-resolved fluorimetry (briefly introduced in the section 'Fluorimetry') is adopted.

## Chiral Derivatization

The simultaneous determination of enantiomers of drugs and other materials of biological importance and the estimation of enantiomeric purity of drugs administered as pure enantiomers are among the most important tasks in contemporary bioanalytical chemistry. Although a great variety of chiral stationary phases are commercially available for the direct separation of enantiomers by GC, TLC, SFC, and especially by HPLC, the classical approach, i.e., transformation of the enantiomers to a pair of diastereomers by derivatization with homochiral reagents followed by their separation using achiral chromatography or CE, is still a widely used general method. Two general methods are available for the formation of the diastereomers: formation of covalently bound derivatives with homochiral derivatizing agent prior to the chromatography and incorporation of the homochiral agent into the mobile phase to form a pair of diastereomeric adducts.

### Formation of Covalently Bound Diastereomeric Derivatives

In addition to the general requirements toward derivatization reagents and reactions it is important that the enantiomeric purity of the reagent is high (at least 99.9% if the aim of the analysis is the enantiomeric purity check of chiral drugs). The reagent should be enantiomerically stable (racemization must not occur), and kinetic resolution during the chromatographic run should be absent. An important advantage of this approach is that in possession of both enantiomers of the homochiral reagent, the elution order can be adjusted in such a way that the enantiomeric impurity is eluted first, thus enabling lower detection and quantitation limits to be obtained (see Figure 2).

Although the first separations after enantiomeric derivatization described in the literature were gas chromatographic procedures, at present only HPLC methods are widely used. Of course, in this case only precolumn derivatization is applicable. For the derivatization of the amino group the following main types of chiral derivatizing agents are used (the type of the reaction product in parentheses): acyl chlorides and anhydrides (carboxamides), chloroformates (carbamates), isocyanates (carbamide derivatives), isothiocyanates (thiocarbamide derivatives), acylamino-haloaryl derivatives (acylamino-amino-aryl derivatives), *o*-phthalaldehyde + chiral thiol (isoindoles). The effect of the choice of the derivatizing agent (d or l) is demonstrated in Figure 2, where the amino acids in a sample of plant origin were derivatized with *o*-phthalaldehyde + d- and l-N-isobutyrylcysteine, respectively.

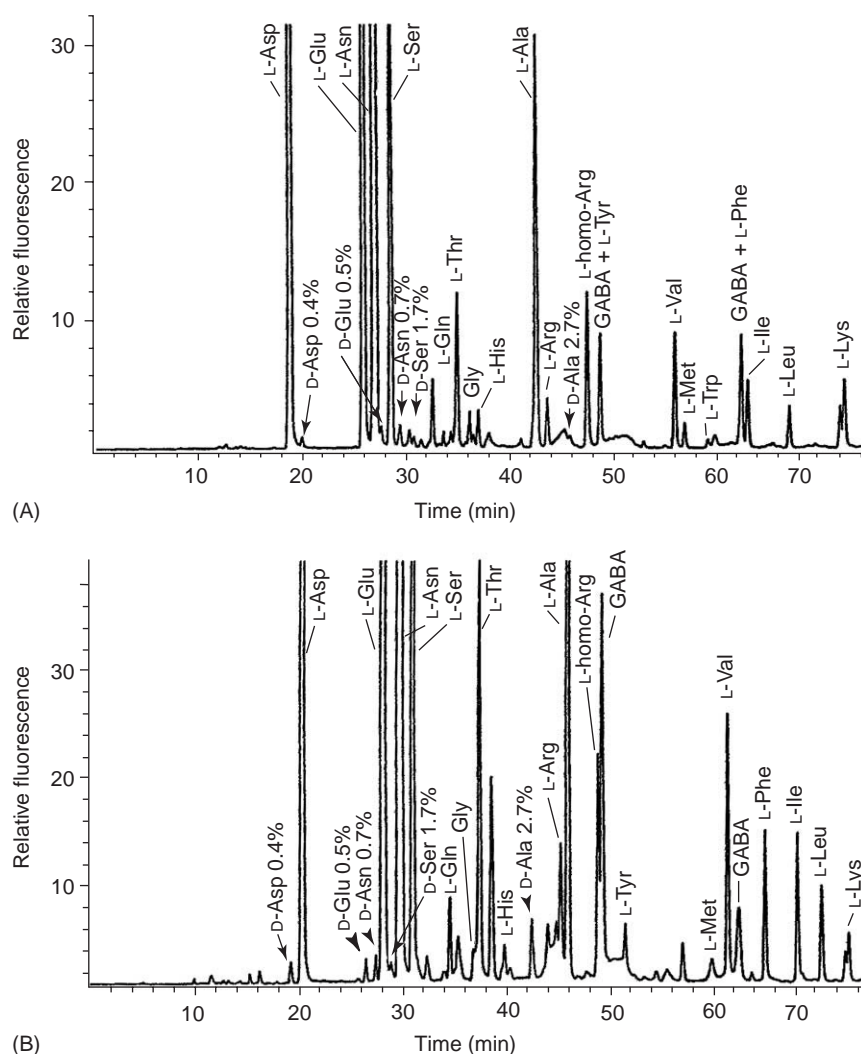
The main chiral derivatizing agents for alcohols and phenols are acyl chlorides and anhydrides (esters), isocyanates (urethanes), acyl nitriles (esters), while for carboxylic acids the most generally used derivatizing agents are chiral amines that can be applied after the activation of the carboxylic group in different ways.

### Formation of Diastereomeric Adducts

Enantiomeric separation based on dynamic formation of a pair of diastereomeric adducts (inclusion complexes, ion-pairs, chelates) is also a widely used method. In this case (mainly in HPLC but with an increasing rate also in CE, CEC, MEKC), the homochiral reagent is dissolved in the mobile phase and the main factor creating the basis for the separation of the enantiomers is the difference between the stability constants of the two diastereomers. The most frequently used homochiral reagents are cyclodextrins and their derivatives, but several applications have been described with chiral ion-pair forming agents such as 10-camphorsulfonic acid and quinine, chiral chelates of copper(II), which serve as the basis of ligand-exchange chromatography.

## General Aspects: Types of and Requirements toward Derivatization Reactions and Reagents

As for the timing and arrangement of derivatization reactions in the case of (mainly spectroscopic and spectrophotometric/fluorimetric) methods not involving chromatographic or electrophoretic separation steps, the question is whether the reaction is run online (automatic analyzers with segmented or non-segmented flow) or offline prior to the measurement.



**Figure 2** HPLC elution profile of L and D-amino acids from freshly prepared apple juice (Golden Delicious) derivatized as the isoindoles with (A) *o*-phthalaldehyde-*N*-isobutyl-L-cysteine; (B) *o*-phthalaldehyde-*N*-isobutyl-D-cysteine; Column: Hypersil ODS (250 × 4 mm; 5 μm; mobile phase, gradient elution; (A) = 23 mmol l<sup>-1</sup> sodium acetate (pH 5.95); (B) methanol-acetonitrile (600:50 v/v), linear gradient from 0% B to 53.5% B in 75 min; flow rate: 1 ml min<sup>-1</sup>; fluorescence detection (230 nm excitation, 445 nm emission). (Reprinted with permission from Brückner H, Haasmann S, Langer M, Westhauser T, and Wittner R (1994). *Journal of Chromatography A* 666: 259–273; © Elsevier.)

In the case of chromatographic or electrophoretic methods the derivatization reaction can take place before the separation (precolumn derivatizations) or between the column/capillary and the detector (postcolumn derivatizations). The arrangement of pre-column derivatizations can be offline (reaction separated from the chromatographic/electrophoretic procedure in time and space) or online (precolumn reactor before the injector). In the case of postcolumn derivatization, only online arrangement can exist: the reactor should be inserted between the column/capillary and the detector. The derivatization in the reactors takes place usually in the solution phase, but solid-phase reactors where the reagent is immobilized on the surface of a solid support

have also gained importance. An important application of postcolumn derivatization is in amino acid analyzers where amino acids are first separated by (usually ion-exchange) chromatography followed by transformation in a postcolumn reactor to colored derivatives with ninhydrin or fluorescent derivatives with an *o*-phthalaldehyde-thiol reagent for the spectrophotometric or fluorimetric measurement.

The requirements toward the derivatization reagent and reaction depend on the analytical problem to be solved. Generally, it is advantageous if the reaction conditions are as simple as possible. Side-reaction-free and rapid one-pot reactions, reaching completion within a few minutes at room temperature, are suitable for automation as ideal cases. Of course,

compromise can be made in many cases. For example, it is not an obstacle if heating is necessary to the completion of the reaction; moreover, 100% yield of the reaction is not by all means necessary (the yield should be, however, reproducible under the specified reaction conditions). The necessity of an extraction step is tolerable; moreover, this can also be automated.

The application of dual-purpose reagents is very advantageous. For example, of the derivatizing agents discussed in the preceding sections Girard T reagent is a useful tool to introduce electric charge to uncharged molecules, thus enabling their CE analysis to be performed; but at the same time it greatly improves the UV detectability of spectrophotometrically poorly active ketones. Moreover, it also improves the sensitivity of HPLC–MS detection. In the case of separation of spectrophotometrically poorly active enantiomers, it is advantageous if the homochiral derivatizing reagent transforms the analyte to spectrophotometrically or fluorimetrically active diastereomeric derivatives. Most of the reagent types shortly summarized in the section ‘Chiral derivatization’ to fulfill this requirement (see Figure 2).

*See also:* **Capillary Electrophoresis:** Overview. **Gas Chromatography:** Mass Spectrometry; Chiral Separations. **Immunoassays:** Overview. **Liquid Chromatography:** Liquid Chromatography–Mass Spectrometry. **Mass Spectrometry:** Overview. **Nuclear Magnetic Resonance Spectroscopy:** Overview. **Supercritical Fluid Chromatography:** Overview; Applications. **Thin-Layer Chromatography:** Overview.

## Further Reading

Blau K and Halket JM (1993) *Handbook of Derivatives for Chromatography*. Chichester: Wiley.

Chankvetadze B (1997) *Capillary Electrophoresis in Chiral Analysis*. Chichester: Wiley.

Cimpean G (2001) Pre- and post-chromatographic derivatization. In: Nyiredy Sz (ed.) *Planar Chromatography*, pp. 410–445. Budapest: Springer.

Drozd J (1981) *Chemical Derivatization in Gas Chromatography*. Amsterdam: Elsevier.

Gore MG (2000) *Spectrophotometry and Spectrofluorimetry: A Practical Approach*, 2nd edn. Oxford: Oxford University Press.

Görög S (1995) *Ultraviolet–Visible Spectrophotometry in Pharmaceutical Analysis*. Boca Raton: CRC Press.

Görög S (2000) Chiral derivatization. In: Wilson ID, Adlard TR, Poole CF, and Cook M (eds.) *Encyclopedia of Separation Sciences*, pp. 2310–2321. London: Academic Press.

Gosling JP (ed.) (2000) *Immunoassays: A Practical Approach*. Oxford: Oxford University Press.

Ichinose N, Schwedt G, Schnepel FM, and Adachi K (1991) *Fluorometric Analysis in Biomedical Chemistry. Trends and Techniques Including HPLC Applications*. Chichester: Wiley.

Krull IS, Deyl Z, and Lingeman H (1994) General strategies and selection of derivatization reactions for LC. *Journal of Chromatography B, Biomedical Applications* 659.

Lingeman H and Underberg WJM (1990) *Detection-Oriented Derivatization Techniques in Liquid Chromatography*. New York: Dekker.

Lobinski R and Marczenko Z (1996) *Spectrochemical Trace Analysis of Metals and Metalloids*. Amsterdam: Elsevier.

Lunn G and Hellwig LC (1998) *Handbook of Derivatization Reactions for HPLC*. New York: Wiley.

Pierce AE (1968) *Silylation of Organic Compounds: a Technique for Gas-Phase Analysis*. Rockford, IL: Pierce Chemical Company.

Toyo'oka T (1999) *Modern Derivatization Methods for Separation Sciences*. New York: Wiley.

# DETERGENTS

*See* SURFACTANTS AND DETERGENTS

# DIALYSIS

*See* MEMBRANE TECHNIQUES: Dialysis and Reverse Osmosis



# DIFFERENTIAL SCANNING CALORIMETRY

See **THERMAL ANALYSIS: Overview; Temperature Modulated Techniques; Coupled Techniques; Sample-Controlled Techniques; Nonbasic Techniques**

## DIODE ARRAY

See **SPECTROPHOTOMETRY: Diode Array**

## DIOXINS

**K Kannan and K M Aldous**, Wadsworth Center, Albany, NY, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

'Dioxins' is a generic term for the polychlorinated dibenzo-*p*-dioxins (PCDDs) and polychlorinated dibenzofurans (PCDFs), comprising 75 and 135 congeners, respectively. The generalized structures for these two groups of molecules are shown in Figure 1. The term dioxin is also used to refer specifically to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), the most toxic congener. In addition to the chlorinated dioxins, brominated analogs of PCDDs and PCDFs can also be formed when organic material containing bromides is incinerated. Throughout this article the term dioxins is used to refer to the family of chlorinated dioxins and furans, although the analytical principles described are generally applicable to the brominated congeners as well.

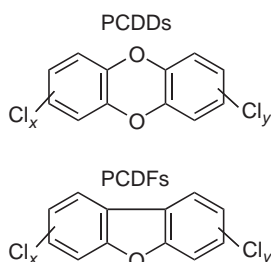
Dioxins occur as unintended by-products of incomplete combustion and certain chemical processes. The International Agency for Research

on Cancer and the United States Environmental Protection Agency (USEPA) have classified TCDD as a known human carcinogen. In view of the widespread distribution of dioxins in the environment, these compounds are being monitored routinely in air, water, soil, sediments, biota, and food-stuffs. Regulatory agencies have identified dioxins as priority environmental pollutants for routine monitoring, with the aim of reducing both emissions and exposures.

### Analytical Purpose and Requirements

The need for analysis of PCDDs and PCDFs arises from the apparent high toxicity of TCDD and 16 other 2,3,7,8-substituted congeners. This toxicity has been established from numerous studies on laboratory test animals. The true magnitude of their toxicity to humans probably cannot be defined, but precautionary guidelines for dietary intake limits have nonetheless been set in the low picogram per kilogram of body weight per day range (e.g., tolerable daily intake in the UK is 2 picogram per kilogram of body weight per day).

In general, the analysis of PCDDs and PCDFs is intended to determine their sources and pathways of human and wildlife exposures, and their environmental fates. The environmental levels and toxicological data are also used for risk assessment purposes. Several countries have promulgated compliance monitoring of wastes, in order to regulate the release of dioxins into the environment. In all cases, the analytical procedures used must be capable of measuring background environmental concentrations ranging from ng per g for soils and sewage sludge to fg l<sup>-1</sup> and fgm<sup>-3</sup> in water and air, respectively.



**Figure 1** Generalized structures of PCDDs and PCDFs.



Regulatory monitoring of emissions from industrial processes requires reliable determination of sub-nanogram per cubic meter in stack gases and a few parts per quadrillion ( $\text{pg l}^{-1}$ ) in water. From these examples, it is clear that a distinctive feature of analysis of PCDD/Fs is the need to achieve rather low detection limits – typically a few picograms in the samples analyzed in order to monitor and control the discharge of these compounds. PCDDs and PCDFs occur in environmental and biological matrices as complex mixtures with various other contaminants, such as polychlorinated biphenyls (PCBs) and biogenic compounds (lipids, proteins). Analyses of dioxins at picogram or femtogram levels are complicated by the presence of myriad of interfering compounds. Accurate determination of trace levels of dioxins in complex environmental matrices is a demanding analytical task involving extensive sample treatment and rigorous quality control during the instrumental analysis.

Another critical issue in the analysis of dioxins is the specificity at the individual congener level for use in risk assessments. As noted earlier, the congener with the highest toxicity is 2,3,7,8-TCDD. Other congeners whose substitution includes the same symmetrical arrangement of chlorines also exhibit high toxicities. There are a total of 17 such PCDD and PCDF congeners, whose relative toxic potencies vary over as much as four orders of magnitude. In order to express the toxic potential of a complex mixture of individual congeners as a single integrated parameter, the concept of the toxic equivalency (TEQ) value was developed. This requires the specific determination of the most toxic congeners (2,3,7,8-substituted), normalization of each of them to a 2,3,7,8-TCDD equivalent using a set of factors (called TCDD equivalency factors, or TEFs) that reflect their relative toxicities, and then summation. The most commonly used TEFs are those that were proposed by the World Health Organization in 1998. TEQs are a useful means of expressing toxicity in regulatory issues and risk assessment problems. Several legislative guidelines for PCDDs and PCDFs are based on TEQs rather than on concentrations of individual or all congeners. The legislative usage necessitates the determination of concentrations of individual congeners rather than total dioxins.

Considerable advances have been made in the analytical techniques used to measure dioxins over the last three decades. They range from rapid biological screening techniques, utilizing *in vitro* gene expression cell bioassays and immunoassays (referred to as bioanalytical techniques), to comprehensive two-dimensional gas chromatographic ( $\text{GC} \times \text{GC}$ ) separation and identification by time-of-flight mass

spectrometer (TOFMS). Dioxin analyses performed with combined high-resolution capillary GC/high-resolution MS (HRGC–HRMS) are still the method of choice, although time consuming and relatively expensive. While HRMS methods are continually becoming faster and less expensive, as the use of automated extraction and cleanup procedures increases, bioanalytical methods also offer advantages in speed, expense, and field portability. The bioanalytical methods can be used for the rapid and inexpensive detection and relative quantification of TEQs. Details regarding the use of cell bioassays and immunoassays for the detection of dioxins have been reviewed elsewhere. Quick, cost-effective field screening and monitoring techniques, such as the bioanalytical methods coupled with conventional HRMS testing, can increase the amount of information that is available for the determination of the presence and concentration of contaminants that could impact the environment and human health.

Compliance environmental monitoring of PCDDs and PCDFs requires the analysis of the 17 2,3,7,8-substituted congeners. Although non-2,3,7,8-substituted congeners are found in the environment, they are not monitored routinely, due to their low bioaccumulation potential and toxicity. However, analysis of non-2,3,7,8-substituted congeners can be useful in tracing the sources and fates of dioxins in the environment.

Current standard methods of dioxin analysis using HRGC–HRMS require rigorous quality assurance and quality control. The samples undergo extensive treatment and extraction prior to instrumental analysis. Each step in the sample handling process is critical in order to discriminate possible interferences from other substances that may be present and to avoid analyte losses; thus, the final qualitative and quantitative measurements can be made with the greatest possible precision and accuracy, according to the requirements of the method. The analytical process includes representative sampling, careful handling of samples, extraction, cleanup, GC separation, detection, and accurate quantification. The cleanup steps provide a suitable removal of the bulk matrix and some interfering compounds; HRGC allows appropriate separations among the various congeners; and HRMS affords a sensitive and selective method of detection. Finally, an isotopic dilution technique based on the use of  $^{13}\text{C}$ -labeled standards provides the reliable quantification needed to quantify the dioxins present.

The basic requirements that must be met by an analytical procedure for dioxin measurement are: high sensitivity, selectivity, and specificity, and high

accuracy and precision. In general, sample extraction, cleanup, chromatographic separation, and mass spectral identification and quantification methods together form a delicately matched set, each element of which is designed to meet the basic requirements.

## Sample Collection and Preparation

The quality and utility of the analytical data depend on the validity of the sample and the adequacy of the sampling program. The number of individual samples that need to be analyzed will depend on the kind of information required by the investigation. The collection of samples for dioxin analysis requires the application of standard practices, to ensure that representative samples are obtained that contain sufficient analyte to allow effective analysis. All of the equipment and containers used to collect and store samples should be precleaned with organic solvents, typically acetone and hexane, to reduce interferences and to ensure that they are free of dioxin contamination. They may additionally be heated in a muffle furnace at  $\sim 450^{\circ}\text{C}$  for 2 h if they are made of borosilicate glass. Samples are kept on ice or at  $< 4^{\circ}\text{C}$  from the time of collection until receipt at the analytical laboratory. Samples must be frozen upon receipt at the laboratory and maintained in the dark at  $< -10^{\circ}\text{C}$  until prepared for analysis. In the analysis of dioxins, there is no maximum holding time, as dioxins are stable and not subject to loss or degradation, when handled in accordance with standard operating procedures. Some regulatory agencies recommend a sample holding time of 7 days for aqueous samples and for up to 1 year for solid matrices, and an extract holding time of 40 days for water samples to up to 1 year for other matrices. Samples with high moisture content (e.g., vegetables) may dehydrate unless extracted immediately after collection. Freezing at  $< -10^{\circ}\text{C}$  may dehydrate such samples. Adequate precautions should be taken to account for the loss in moisture content while storing the samples in a freezer.

There are unique problems associated with the sampling of gaseous and aqueous matrices. Because of the low concentrations of dioxins in these media, high-volume sampling devices are used to collect several tens to hundreds of liters of water, or several hundreds to thousands of cubic meters of air. The sampling of dioxins in gases is complex and beyond the scope of this article, but it utilizes a variety of solid and liquid traps, tetrafluoroethane filters, resin beds, polyurethane foam plugs, reagent traps, and adsorbents in series, to collect both particulate and vapor-phase PCDDs and PCDFs. Aqueous samples

are preconcentrated on a glass fiber filter combined with absorbing resin cartridges. It is essential that these sampling media are precleaned with organic solvents and spiked with the appropriately labeled analogs, at the beginning of the sampling train. For biological fluids such as milk, urine, and blood, it is essential that an appropriate set of isotopically labeled analogs of the target congeners are introduced into the sample through spiking prior to processing or pretreatment.

Sample mass or volume has a profound influence on the detection limits for dioxins. For biological fluids such as blood plasma, the conventional HRGC–HRMS method requires a sample volume of 40–60 ml to attain a detection limit of  $\leq 1$  pg per g. Collection of 40–60 ml blood is laborious and problematic, particularly from infants and aged persons. Unless other methods such as solvent-cut, large-volume injection systems are used, the nominal sample size of 5–10 ml is not adequate to measure sub-pg per g levels of dioxin in blood.

Legislation has required the detection of dioxins in food and feed at fg per g (parts per quadrillion) levels. One way in which to achieve this detection level is by increasing the mass of the samples extracted. Nevertheless, it should be noted that the extraction of a large amount of sample can also increase the amounts of impurities and interference, which are coextracted, requiring additional cleanup steps. Optimization of the mass of samples used for extraction is necessary to determine which cleanup steps need to be used, and to attain the required level of detection.

The content of moisture or lipid in biological samples, and of organic carbon in soil or sludge, can account for variations in dioxin concentrations measured between samples. Therefore, normalization of concentrations on dry and/or lipid weight or organic carbon would improve the comparability of concentrations across samples.

Sample pretreatment involves modification of the physical form of the sample in such a way that the dioxins can be extracted efficiently. Some environmental matrices such as fly ash require pretreatment involving acid attack, usually with HCl, to destroy the structure of the matrix and to allow better recovery of the dioxins. For certain samples, denaturants are added so as to destroy any protein–analyte interactions and disrupt micellar formation, both of which can decrease recoveries of dioxins. For example, formic acid has been used as a denaturant prior to extraction of dioxins in plasma or serum. Sodium or potassium oxalate (20 mg per g milk) is added to milk samples to disrupt fat globules. Depending on the type of sample, grinding, homogenization, and blending of various sample phases

may be needed. For solid matrices, moisture in samples should be removed either by freeze-drying or by grinding with anhydrous sodium sulfate, when the sample is extracted with nonpolar organic solvents.

## Extraction

In general, extraction methods are standardized around liquid–solid extraction (e.g., Soxhlet), solid-phase extraction (SPE), and liquid–liquid extraction (LLE). The choice of procedures is dependent upon the amount and type of sample requiring extraction and the types of other compounds that may be present. Extraction techniques for food and other biological matrices are generally based on the lipophilic properties of dioxins, and therefore the extraction methods are based on the extraction of lipids from the sample matrix. Methods for the isolation of lipid fractions in food samples such as butter, oil, and animal fats involve direct dissolution with *n*-hexane or light petroleum to the desired fat concentration. Soxhlet extraction with, for example, toluene is the most common liquid–solid extraction procedure applied to solids. Although this technique achieves good extraction rates, it requires a long extraction time (24–48 h) and large volumes of solvent (300–400 ml). After extraction, the solvent must be concentrated, which results in air emissions or costly solvent disposal.

Saponification under alkaline conditions (in the presence of ethanolic KOH), followed by extraction with organic solvent is often employed for the analysis of large amounts of fats (several 10 s to 100 s of g). However, this method is known to degrade the more highly chlorinated dioxins, particularly octachlorodibenzofuran, thus leading to artifacts.

Over the past few years, efforts have been made to develop alternative extraction techniques that allow more efficient extraction, along with reduced solvent volumes, in shorter times, and incorporating high levels of automation. Methods such as accelerated solvent extraction (ASE) uses conventional liquid solvents at elevated pressures (1500–2000 psi) and temperatures (50–200°C) to extract solid samples rapidly, with much less solvent than is required by Soxhlet methods. The sample is placed in extraction cells, which are filled with an extraction solvent and heated for 5–10 min, with the expanding solvent vented to a collection vial. The entire procedure is completed within 10–20 min per sample, and uses only 15–20 ml of solvent. Similarly, microwave-assisted solvent extraction (MAE) requires a relatively short extraction time (<1 h) and small amount of solvent (~30 ml), and thus offers an attractive

alternative to the Soxhlet procedure. MAE uses microwave radiation that enables instantaneous and efficient heating in the presence of microwave-absorbing compounds. However, this technique requires further filtration to obtain the final extract. Both ASE and MAE require a relatively high initial investment. Supercritical fluid extraction (SFE) of solids offers a means of exercising close control over extraction conditions, and thus potentially improved repeatability. Carbon dioxide is most often used as the extraction solvent, because of its moderate critical temperature (31°C) and pressure (1070 psi). However, optimum SFE conditions vary from one sample matrix to another. Because of their advantages in speed, cost, and solvent quantities, ASE, MAE, and SFE are promising alternatives to the conventional Soxhlet extraction method.

High-capacity absorbent phases are used for the extraction of dioxins in aqueous samples. These phases trap the target molecules, which can be recovered later by reverse elution with a suitable solvent such as dichloromethane, hexane, or ethanol. SPE of dioxins, using a variety of commercially available sorbents, has been used for various biological fluids such as serum. LLE is used to extract lipid-phase containing dioxins from aqueous samples such as milk. LLE of milk involves mixing with sodium oxalate and ethanol or methanol, followed by repeated extractions using a combination of organic solvents such as acetone–pentane or diethyl ether–light petroleum. Solid-phase microextraction (SPME), which utilizes various coated adsorbents and/or absorbents on a fiber contained in a syringe-like device, has been tested for its utility in dioxin analysis of aqueous media. This method is based on selective absorption of dioxins to carbon SPME fibers and is being further validated to establish its applicability.

Because it is not possible to achieve complete extraction in all cases, all of the time, it is necessary to test the effects of varied extraction conditions (such as Soxhlet cycles, duration, solvent volume, and sample weight) and then establish a rigorous regimen by which optimum performance can be maintained. It should be noted that the addition and recovery of isotopically labeled analogs does not provide a reliable check on extraction efficiency, since the form of these analogs is different from that of the native analytes within the sample matrix. Inclusion of a quality check sample or standard reference material in each batch of samples analyzed does allow extraction performances to be monitored and controlled, to ensure reproducibility.

## Extract Purification

The isolation of dioxins as a group is aided by these compounds' resistance to oxidative and hydrolytic destruction. Thus, contact with strong acid and/or base can be used to eliminate much of the co-extracted material. Direct contact of concentrated extract with sulfuric acid is used to remove lipids, proteins, and certain oxidizable materials in the extracts. Gel permeation chromatography (GPC) has also been used to remove high molecular weight interferences including lipids. GPC presents several advantages, including robustness, versatility, and an ability to be used repetitively without the need for regeneration. Fat retainers and lipophilic gels have also been used to extract or cleanup lipids from sample matrices. The adsorbent columns do not allow the handling of more than 1 g of lipids, in order that they are able to produce clean extracts. Therefore, analyses of foods of animal origin, especially dairy products, require additional fat-removal steps. A multistep purification procedure is routinely followed to concentrate the final extract to appropriate volumes of solvent, thus allowing the detection of analytes at the ultratrace levels at which they usually occur. Successive cleanup steps using chromatographic adsorbents (silica, Florisil, alumina, and activated carbon) are most commonly used for cleanup and fractionation of dioxins from other interferences. Polypropylene tubes filled with sorbents are commercially available (as SPE cartridges). A recent technique, referred as multilayer silica gel column chromatography, involves the passage of concentrated extract through a column containing separate layers of acid- and base-modified silica, sandwiched between layers of nonmodified silica. Dioxin isomers pass through the column unaffected, while the layers retain polar contaminants and other extractables that can act as interferents during the GC analysis. The column can be connected to a vacuum manifold, to facilitate the extraction process and to decrease extraction time. Concentration of the eluate from such a column yields a solution containing not only PCDDs and PCDFs, but also other compounds such as PCBs, polychlorinated naphthalenes (PCNs), and polybrominated diphenyl ethers, all of which are similarly resistant to acid and/or base attack. The latter types of compounds are separated from the dioxins through a combination of liquid chromatography with a carbon column (porous graphitic carbon or pyrenyl ethyldimethyl silylated silica), making use of the relatively small differences between the polarities and sorptive characteristics of the dioxins and those of the PCBs and other chemically similar groups of molecules. Alumina, Florisil,

and/or activated carbon are adsorbents commonly used to separate PCBs from dioxins. A major drawback of carbon as a packing material is its high back-pressure. Therefore, activated carbon dispersed on a glass fiber, polyurethane foam, or silica gel support, or more recently on C<sub>18</sub> packing, is also used. This technique separates dioxins from PCB congeners based on  $\pi$ -cloud interaction with the graphitic carbon surface. Dioxins, with their rigid planar structures, are separated from PCBs because they are more strongly adsorbed onto the carbon than are the PCBs. However, PCNs and other planar compounds also co-elute with dioxins. Success at this stage depends on precise control over materials preparation and elution procedures, as well as on the choice of stationary and mobile phases. Hexane or a combination of hexane and dichloromethane is used frequently as the mobile phase. More complex mobile-phase profiles are required to separate dioxins from coplanar PCBs and PCNs, because of their similar adsorptive properties. These methods are time consuming and require manual sample preparation that may give rise to a decrease in precision and accuracy, in addition to the risks of human exposure. Lot-to-lot variations of adsorbents and solvents can lead to irreproducibility in the elution profiles of dioxins.

The development of extract purification techniques is directed, on the one hand, toward automation, and, on the other hand, toward the development of more selective adsorbents such as immunosorbents. Recently, automated cleanup systems have been developed based on the use of pressurized column chromatographic procedures. This is regarded as an alternative system that bypasses most of the disadvantages of the conventional cleanup methods, given its capacity for processing automatically different samples simultaneously in  $\sim 1$  h. An automated cleanup system that uses a classical set of multilayer silica, basic alumina, and carbon columns has been developed to increase the number of samples that can be treated simultaneously. Several modifications and upgrades of this system are currently examined for the analysis of dioxins in both abiotic and biotic samples. Immunoaffinity chromatography (IAC) has been investigated as a means to simplify cleanup of extracts. IAC columns are generated from antidioxin antibodies and are shown to bind dioxins selectively from samples. The IAC procedure is relatively fast and uses fewer organic solvents than conventional methods. However, it is not compatible with high-fat matrices and is not fully selective for all of the 17 toxic PCDD/F congeners.

Recoveries of dioxins through adsorbent columns must be checked and validated by spiking with known concentrations. Particular adsorbency



characteristics (size, weight, conditioning, etc.) and solvent polarity, elution rate, volume, and elution profiles need to be optimized to separate dioxins from other interferences. While the use of stable-isotope dilution calibration allows recovery losses to be corrected for, in calculation over an established range, actual recoveries should be monitored to ensure that procedures are under control.

## Chromatographic Separation

The GC separation of all target congeners is critical to the successful analysis of dioxins. The development of capillary GC columns has permitted congener-specific determination of a number of these mixtures. In principle, polar or nonpolar capillary columns (normally combined with splitless or on-column injection) may be used. The isomer-specific elution patterns of dioxins are well established for such widely used column coatings as 5% diphenyl-95% dimethyl-polysiloxane, 14% cyanopropylphenyl-86% methyl polysiloxane, and 90% biscyanopropyl-10% cyanopropylphenyl polysiloxane. However, chromatographic separation of the toxic isomers from all of the nontoxic isomers requires the use of at least two columns with differing compositions and polarities. To enhance the resolution, longer or narrower bore columns, or a combination of both, have been used. Nevertheless, several critical congeners co-elute with others. Through the use of nonpolar columns, a separation can be achieved between homolog groups. In addition, nonpolar columns separate the 2,3,7,8-substituted congeners from one another; however, they fail to resolve this group of congeners from the nontoxic congeners. With polar columns, the resolving power improves, but incomplete separation still remains for 2,3,7,8-TCDF, 1,2,3,7,8-PeCDF, and 1,2,3,4,7,8-HxCDF (Table 1). TCDF is the most difficult to resolve and requires medium polar columns. Nonpolar columns

are frequently used for the analysis of biological samples, as these samples generally do not contain congeners other than the 2,3,7,8-substituted ones. Polar columns are required for isomer-specific analyses of PCDDs and PCDFs in samples that may contain nonlaterally substituted isomers, such as environmental samples, fly ash, ambient air, and PCDD/F-contaminated products of various origins. Injection onto two different columns is recommended for an unambiguous determination. Several chromatographic columns are being developed for the purpose of overcoming the need for dual-column injections.

The chief concerns with single-column GC analysis are lack of resolution, lack of unambiguous identification, and lack of robustness. Multidimensional GC (MDGC) has been used as a complementary technique to resolve difficulties in the separation of co-eluting compounds. This technique provides an alternative to use of additional sample pretreatment steps, which could lead to a further loss of analytes. One of the MDGC methods, comprehensive GC  $\times$  GC, has been shown to be very useful for the separation of complex mixtures of environmental contaminants, including dioxins. This technique employs a pair of coupled columns of differing selectivities and it subjects the entire sample to a two-dimensional separation. Effluent gas from the first column is modulated to produce sharp chemical pulses, which are rapidly separated on the second column. A separation plane is produced by the orthogonal retention-time axes for the two columns. GC  $\times$  GC offers increased sensitivity and selectivity and improved reliability in the identification of target compounds.

## Spectroscopic Detection

Mass spectrometric identification is the most commonly used method in dioxin analysis. The coupling of GC and MS has been recognized as the most practical and widely used method of measuring dioxins. Although both high- (HRMS) and low-resolution mass spectrometry (LRMS) have been used in dioxin analysis, for analyses involving multiple congeners at the parts per trillion or parts per quadrillion level, an isotope dilution technique using HRMS is essential. LRMS is useful as a screening tool, but it is not a recommended technique for stack-emission data for PCDD/Fs; instead, the LRMS method was designed for samples that contain levels of dioxins in the parts per billion range.

For certain regulatory measurements, the use of HRMS is obligatory (e.g., USEPA method 1613, European Union method EN1948-1/2/3). HRMS is operated in the EI mode (electron energy 38 eV) at a

**Table 1** Common coelutants in the GC analysis of 2,3,7,8-substituted dioxins and furans using 5% diphenyl-based microbore column chromatography

<i>Toxic congener</i>	<i>Co-eluting nontoxic congener</i>
2,3,7,8-TCDF	1,2,4,9-; 2,3,4,6-; 2,3,4,8-; 1,2,7,9-; 2,3,4,7-TCDFs
2,3,7,8-TCDD	1,2,3,9-; 1,2,3,7-/1,2,3,8-TCDDs
1,2,3,7,8-PeCDF	1,2,3,4,8- PeCDF
2,3,4,7,8-PeCDF	1,2,6,7,9-; 1,2,3,6,9-; 1,2,4,8,9-; 2,3,4,8,9-PeCDFs
1,2,3,4,7,8-HxCDF	1,2,3,4,6,7-HxCDF
2,3,4,6,7,8-HxCDF	1,2,3,6,8,9-HxCDF
1,2,3,7,8,9-HxCDD	1,2,3,4,6,7-HxCDD

resolving power of 10 000. Under these conditions, various ions (including isotopically labeled ions) are monitored in the selected ion monitoring mode. Two characteristic ions of the molecular cluster are selected for each of the congener groups, from tetra- to octa-PCDD and PCDF. An important advantage of HRMS is its capability of allowing to test for suppression by interferent compounds, via lock mass monitoring.

Tandem MS (MS/MS) has been reported to be a valuable technique for improved selectivity in analysis of PCDD/Fs. Increased selectivity is obtained by two mass selection stages and collisionally induced dissociation (CID) of characteristic ions. PCDD/F molecular ions pass through the first mass analyzer to a collision cell, where CID occurs and specific daughter ions, usually the  $[M-COCl]^+$  ions, are then selected in the second mass analyzer. Although this technique provides high selectivity, the sensitivity is relatively low compared to the HRMS technique. Other disadvantages of GC-MS/MS include high initial cost and high maintenance costs. Recently, ion-trap tandem MS (ITMS-MS) has become an interesting alternative to HRMS because of its low cost. Under optimal ionization conditions, ITMS-MS performs reproducible quantification of dioxins. ITMS-MS's detection limit of TCDD is in the range of 25–50 fg.

As chromatographic run times are shortened through the use of microbore columns, conventional MS (quadrupole and magnetic sector) are unable to scan fast enough ( $\leq 1$  s per spectrum) that they can produce sufficient data points to accurately define a chromatographic peak. TOFMS can operate at scan speeds of more than 100 spectra per second. This speed provides a number of advantages over other mass spectrometers, such as the capability to deconvolute mass chromatograms for compounds with retention times that differ by more than  $\sim 150$  ms. A combination of GC  $\times$  GC with TOFMS provides both the sensitivity and selectivity required for the analysis of dioxins in environmental samples, and its application is being investigated.

The successful identification of dioxins is dependent on the following criteria: (1) retention times of chromatographic peaks must fall within the appropriate chromatographic windows; (2) simultaneous responses for the two masses monitored must be obtained; (3) signal-to-noise ratios must be  $> 3$ ; and (4) relative isotopic peak ratios must be within  $\pm 15\%$  of the theoretical values. Once these criteria are satisfied, assignment of congeners is performed by comparing the retention times with those of the corresponding labeled compounds added as internal standards. Quantification is carried out by an

isotopic dilution technique, based on the addition of labeled standards. Relative response factors for the individual isomers are obtained using five or more calibration standard solutions containing both native and labeled congeners at the concentration range in which the sample values fall. In addition to internal standards, which are spiked at an early stage of sample extraction, injection internal standards (which are usually non-2,3,7,8  $^{13}C_{12}$  labeled analogs) are added to the cleaned extract immediately prior to GC-MS analysis and provide a means to correct for variations in injection efficiencies; thus, actual recoveries of internal standards can be calculated. The recoveries of labeled standards are calculated using a mixture of two labeled PCDDs added before the HRGC-HRMS analysis.

**See also:** **Air Analysis:** Sampling. **Extraction:** Solvent Extraction; Multistage Countercurrent Distribution; Supercritical Fluid Extraction. **Gas Chromatography:** Multidimensional Techniques. **Mass Spectrometry:** Time-of-Flight. **Quality Assurance:** Quality Control; Reference Materials. **Water Analysis:** Overview.

## Further Reading

- Dallüge J, Beens J, and Brinkman UAT (2003) Comprehensive two-dimensional gas chromatography: A powerful and versatile analytical tool. *Journal of Chromatography A* 1000: 69–108.
- Eljarrat E and Barceló D (2002) Congener-specific determination of dioxins and related compounds by gas chromatography coupled to LRMS, HRMS, MS/MS and TOFMS. *Journal of Mass Spectrometry* 37: 1105–1117.
- Focant J-F, Eppe G, Pirard C, and De Pauw E (2001) Fast clean-up for polychlorinated dibenzo-*p*-dioxins, dibenzofurans and coplanar polychlorinated biphenyls analysis of high fat content biological samples. *Journal of Chromatography* 925: 207–221.
- Hilscherova K, Machala M, Kannan K, Blankenship AL, and Giesy JP (2000) Cell bioassays for the detection of aryl hydrocarbon (AhR) and estrogen receptor (ER) mediated activity in environmental samples. *Environmental Science and Pollution Research* 7: 159–171.
- Kibblewhite M and Fernandes A (1995) Dioxins. In: Townshend A (ed.) *Encyclopedia of Analytical Chemistry*, 1st edn, pp. 936–939. New York: Academic Press.
- Liem AKD (1999) Important developments in methods and techniques for the determination of dioxins and PCBs in foodstuffs and human tissues. *Trends in Analytical Chemistry* 18: 429–439.
- Rappe C (1996) Sources and environmental concentrations of dioxins and related compounds. *Pure & Applied Chemistry* 68: 1781–1789.
- Schomburg G (1995) Two dimensional gas chromatography: principles, instrumentation and methods. *Journal of Chromatography A* 703: 309–325.



USEPA (1994) *Method 1613: Tetra-through Octa-chlorinated Dioxins and Furans by Isotope Dilution HRGC/HRMS*. Washington DC: Office of Water, United States Environmental Protection Agency.

World Health Organization (1989) *Polychlorinated Dibenzo-para-dioxins and Dibenzofurans*. Environmental Health Criteria 88, p. 409. Geneva: World Health Organization.

## DISSOLUTION TESTING

See **PHARMACEUTICAL ANALYSIS: Dissolution Testing**

## DISTILLATION

**J D Green**, Beverley, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Distillation is a widely used technique in chemical analysis for characterizing materials by establishing an index of purity and for separating selected components from a complete matrix. The technique is even more widely used in preparative chemistry and throughout manufacturing industry as a means of purifying products and chemical intermediates. Distillation operations differ enormously in size and complexity from the semimicro scale to the 'thousands of tons per annum' production operations. For analytical purposes the scale employed is usually bench-level.

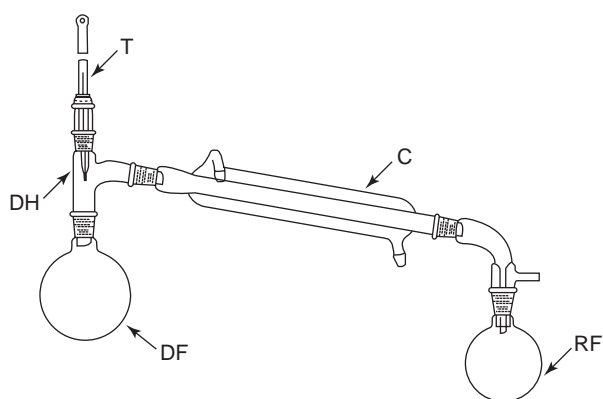
Numerous quoted standard specifications refer to distillation ranges as criteria of purity or suitability for use, or as indicators of performance. Published standards for analytical reagents in the AnalaR range and similar documentation by the American Chemical Society refer to distillation ranges as criteria of purity for appropriate materials.

Distillation is the process that occurs when a liquid sample is volatilized to produce a vapor that is subsequently condensed to a liquid richer in the more volatile components of the original sample. The volatilization process usually involves heating the liquid but it may also be achieved by reducing the pressure or by a combination of both. This can be demonstrated in a simple laboratory distillation apparatus comprising a flask, distillation head, condenser, and sample collector (**Figure 1**). A thermometer is included in the apparatus as shown to monitor the progress of the operation. In its simplest form this procedure results in a separation into a volatile

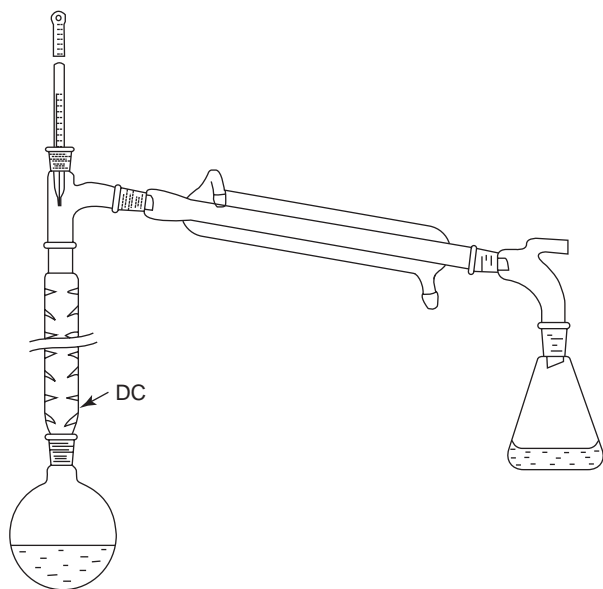
fraction collected in the receiver flask and a nonvolatile residue in the distillation flask. When a distillation column is incorporated in the equipment (**Figure 2**), the evaporation and condensation processes occur continuously. This results in a progressive fractionation of the volatiles as they pass up the column. The most volatile components emerge from the top of the column initially and the less volatile components emerge later. By changing the receivers throughout the course of the distillation a separation or fractionation is effected. Eventually, all the volatiles will have passed over into the sample collectors and any involatile residue present will remain in the distillation flask.

### Principles

The underlying principles are conveniently illustrated by reference to a vapor-liquid equilibrium diagram (**Figure 3**). The diagram relates to a binary mixture containing components P and Q. The lower curve gives the composition of the liquid boiling at various temperatures whilst the upper curve gives the composition of the vapor in equilibrium with the boiling liquid. Points x and y, therefore, give the boiling points of the individual components P and Q, respectively. For example, point A shows that at X degrees the vapor has a composition of approximately 90% P, whilst point B shows that the boiling liquid with which it is in equilibrium has a composition of approximately 80% P. In a continuous distillation process, such as occurs in a distillation column, liquid of composition C (90% Q, 10% P) vaporizes to vapor of composition D, which condenses to liquid of composition E. Subsequently, liquid E becomes vapor F and liquid G (composition: 50% Q, 50% P). This continuous process of vaporization and condensation occurs in the distillation column until a volatile fraction leaves the top of



**Figure 1** Simple distillation apparatus comprising distillation flask (DF), distillation head (DH), thermometer (T), condenser (C), and receiver (or collection) flask (RF). (Reproduced from Furniss BS, Hannaford AJ, Smith PWG, and Tatchell AR (1989) *Vogel's Textbook of Practical Organic Chemistry*, 5th edn., pp. 168–197. Harlow: Longman Scientific and Technical.)

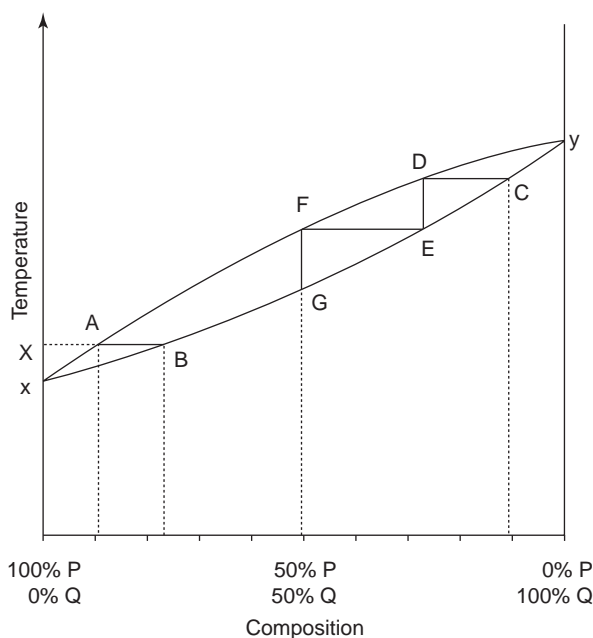


**Figure 2** Distillation apparatus including distillation column (DC). (Reproduced from Furniss BS, Hannaford AJ, Smith PWG, and Tatchell AR (1989) *Vogel's Textbook of Practical Organic Chemistry*, 5th edn., pp. 168–197. Harlow: Longman Scientific and Technical.)

the column and is removed from the process by being collected in the collection flask. At the same time the liquid in the distillation flask becomes progressively more concentrated in the involatile component.

Distillation techniques may be classified into several different types including:

- Distillation at atmospheric pressure
- Distillation under reduced pressure
- Steam distillation



**Figure 3** Vapor-liquid diagram for a binary mixture of components 'P' and 'Q', illustrating the principles of distillation (see text for details).

- Molecular distillation (short-path distillation)
- Azeotropic distillation
- Isopiestic distillation

Distillation at atmospheric or reduced pressure produces a separation according to the general principles discussed in the introduction.

Steam distillation is a means of distilling that part of a sample that is volatile in steam at a lower temperature than would otherwise be the case. This method is typically used for removing phenols from an aqueous sample. A means of introducing steam into the distillation flask must be provided.

Molecular distillation, sometimes termed short-path distillation, is used principally for compounds normally having high boiling points. In such cases, very low pressures are needed to achieve the desired low boiling points. The apparatus is constructed such that the condensing surface is located only a short distance from the distilling liquid and the pressure is reduced so that the process is governed to a large extent by the mean free path of the molecules involved. Hence the terms short-path distillation and molecular distillation.

Azeotropic distillation occurs when a mixture of two materials distils at constant composition. This technique is commonly used to remove water from samples. As an example, toluene may be added to a complex sample containing water, the distillation process results in the toluene-water

azeotrope distilling. The distillate can then be examined to determine the water content of the original sample.

Isopiestic distillation is a convenient way of producing metal-free aqueous samples of volatile acids. The 'crude' acid is placed in an open container, such as a beaker, in a desiccator containing also an open beaker of pure water. The acid vaporizes and subsequent condensation in the pure water produces an aqueous sample of the volatile acid without any of the involatile contaminants such as metals.

The alternative terms 'flash' distillation and 'fractional' distillation are sometimes used to describe some of the above procedures carried out in a particular way. Flash distillation effects a crude separation into volatiles and residue, whilst fractional distillation produces a series of 'cuts' of different volatility (or boiling point) ranges.

Additionally, there are other forms of sample purification and separation that are either a type of distillation or are related to a distillation process:

- Simultaneous distillation/extraction (see Applications section)
- Dean and Stark distillation (see Applications section)
- Simulated distillation (gas chromatographic technique)

Analytically, distillation is used for two principal purposes, firstly as a criterion of purity and secondly as a means of preparing a sample for analysis. Many specification tests include reference to a distillation range within the limits of which a stated percentage of the material of interest distills. Alternatively, distillation may be used to separate volatiles from a sample prior to a suitable analytical technique being employed on the distillate or on the residue. Standard tests are documented that involve distillation as a sample pretreatment method prior to titrimetry, potentiometry, and spectrophotometry.

It is, of course, essential if meaningful comparative results are to be obtained, that the design and use of the apparatus are standardized for such determinations.

## Apparatus

A wide variety of apparatus is available to satisfy the different distillation techniques. The appropriate design of apparatus depends upon the type of distillation to be performed, considering, for example, whether a vacuum is required or steam is needed. Descriptions of apparatus are to be found

in a number of different texts (see Further Reading). Standards referring to the design and use of distillation apparatus have been published by the British Standards Institute (BSI) and the American Society for Testing and Materials (ASTM). Simulated distillation, which is a gas chromatographic technique, is dealt with in a recent review by Robillard *et al.* and referred to in several standards.

Apparatus may be discussed in terms of the distillation flask, the distillation column, the condenser, and the collecting flask(s). By far the most effort has been expended in the design and operation of the distillation column, which is at the heart of the separation efficiency. The form of the column, its size, and the packing used are very influential upon the results that are achievable. A summary of some different types of columns is given in Table 1 and of packing in Table 2.

**Table 1** Types of distillation column

Column type	Description/comments
Dufton	An open tube into which a glass spiral fits closely
Hempel	A simple tube normally filled with a suitable packing (rings/helices) and having a side-arm near the top
Oldershaw	A column with fixed but perforated plates that maintains a fixed amount of liquid on each plate
Podbielniak	A simple tube with a wire packing to provide large contact area between liquid and vapor to effect high efficiencies
Spinning band	A tube fitted with a closely fitting spiral of PTFE or metal gauze that can be rotated at typical speeds of 600–3000 rpm as the vapor-liquid equilibrium is maintained in the column
Vigreux	A tube having pairs of indentations down its length that slope downwards and provide a large and designed surface area to enhance the liquid-vapor equilibrium

**Table 2** Distillation column packings

Packing	Description
Balls	Mostly made of glass. Columns have a tendency to flood easily
Helices	Made from metal or glass, although metal may be packed mechanically to produce a more uniform column
Rings	Usually made of glass of an appropriate size for the column but can be made of porcelain, stainless steel, aluminum, copper, or nickel. Depending upon design they can be termed Raschig, Lessing, or Dixon rings
Wire packings	Produced as 'Heli-Grid' and 'Heli-Pak' packings especially for use with Podbielniak columns

Once apparatus has been chosen carefully to compare with previously used apparatus or to conform to standards, the operation of the equipment must be considered. The following factors are among the most important to be controlled:

- The heating of the distillation flask must be carefully controlled.
- The distillation column must be operated so that it does not become flooded.
- The reflux ratio, that is, the ratio of material returning via reflux to the distillation column or the distillation flask compared to the amount presented to the condenser in unit time must be carefully controlled. The higher the reflux ratio, the purer the material collected from the distillation. Reflux ratios are controlled in simple distillation apparatus by adjustment of the heating rate and by maintaining stable thermal conditions throughout the apparatus.

## Applications

Documentation of analytical applications of distillation is widely dispersed. Distillation is used not only for the assessment of purity and characterization of volatile materials but also as a means of removing volatile compounds from an involatile matrix as part of an analytical procedure. Thus, distillation will be found in analytical procedures relating to volatile hydrocarbons, to polymers, and to environmental samples. Standards including distillation are concerned with specifying products, characterizing materials, preparing samples for analysis, and defining methods in terms of sampling, equipment, and procedures. Most advanced industrial countries have their own Standards Institutes with connections to International Organizations that promote harmonization of measurement and analysis. **Table 3** gives examples of the organizations whose standards are most frequently quoted in measurement science.

A search of the BSI and ASTM sites reveals several hundred standards involving distillation, including current and withdrawn versions, for material specifications and characterizations or for descriptions of sampling, equipment, and procedures of analytical methods involving distillation. As requirements change and analytical capabilities improve, new standards are introduced and old standards are modified or withdrawn. It is therefore advisable to consult the latest documentation prior to considering new testing methods.

**Table 3** Organizations involved with the development and promotion of standards and test methods

<i>Organization</i>	<i>Web address</i>
British Standards Institute	<a href="http://www.bsonline.techindex.co.uk">www.bsonline.techindex.co.uk</a>
American Society for Testing and Materials	<a href="http://www.astm.org">www.astm.org</a>
American National Standards Institute	<a href="http://www.ansi.org">www.ansi.org</a>
Institute of Petroleum	<a href="http://www.intertek-cb.com">www.intertek-cb.com</a>
International Standards Organization	<a href="http://www.iso.ch">www.iso.ch</a>
US Environmental Protection Agency	<a href="http://www.epa.gov">www.epa.gov</a>

Hydrocarbons derived from crude oil and other fossil deposits are separated into their various fractions by using distillation and so it is not surprising that distillation is used as an analytical tool to characterize them. **Table 4** list examples of applications of distillation to hydrocarbon characterization including bituminous materials, tars, fuels, and derived organic materials.

In a similar way to hydrocarbons the purity of organic chemicals is defined by the distillation range over which they distil under carefully controlled conditions. A range of organic compounds including alcohols, esters, and chlorinated compounds are specified in this way. Examples of the associated specification standards are listed in **Table 4**.

Moisture and water content are important parameters in the processing and sale of materials from foodstuffs to petroleum based fuels. Azeotropic distillation is commonly used to determine water content using apparatus developed by Dean and Stark and an immiscible liquid such as toluene as the codistillate.

Some aspects of water quality can also be determined by methods involving distillation; these include determination of 'phenol index', nitrate content, and ammonium content.

The determination of nitrogen by the Kjeldhal method includes a preliminary distillation of the sample. Protein content can be estimated using these methods as can the determination of ammoniacal and total nitrogen in various samples including fertilizers.

As trace analysis of residual compounds in consumables has become more important, methods for extracting these compounds have been developed. A method known as simultaneous distillation extraction developed from the original work of Likens and Nickerson has been particularly popular and effective for extracting the volatiles

**Table 4** Examples of standards involving application of distillation in analysis. This table is not intended to be comprehensive but is included to show the scope of different analyses that involve analysis. Compiled by selection from websites (March 2004)

Application	Standards
<i>Hydrocarbon characteristics and purity</i>	
Tars for road purposes	BS 76
Creosote for wood preservation	BS 144
Middle distillate contamination in petroleum	BS 2000-459.1
Tar, water gas, and coke yields from brown coals and lignite	ISO 647
<i>Organic liquids</i>	
Methanol specification for industrial use	BS 506-1
Butanol specification for industrial use	BS 508-1
Acetone specification for industrial use	BS 509-1
Ethyl acetate specification for industrial use	BS 553
Diethyl ether specification (technical)	BS 579
Propan-2-ol specification for industrial use	BS 1595-1
Methylene chloride specification	BS 1994
Acetic anhydride specification	BS 2068
<i>Standard methods</i>	
Methanol test methods	BS 506-2
Propan-2-ol test methods	BS 1595-2
Specification for apparatus for determination of distillation range	BS 658
Specification for Dean and Stark apparatus	BS 756
<i>Methods for determination of:</i>	
Water in petroleum and bituminous products	BS 2000-74; ISO3733
Distillation characteristics of cutback bitumen	BS 2000-27
Determination of asphaltenes in crude petroleum	BS 2000-143
<i>Nitrogen content</i>	
Chemical analysis of cheese (nitrogen)	BS 770
Nitrogen content of coal and coke	BS 1016-106.2
Liquid milk and cream (nitrogen)	BS 1741-5.2
Total nitrogen in ammonium nitrate	BS 4267-2
Ammoniacal nitrogen in fertilizers	BS 5551-4.1.2
<i>Moisture/water content</i>	
Methods of analysis for coal and coke (total moisture)	BS 1016
Water in formulated detergents	BS 3762
Moisture in condiments and spices	BS 4585-2
<i>Water quality determination</i>	
Ammonium determination	BS 6068-2.7; ISO 5664
Phenol index determination	BS 6068-2.12; ISO 6439
Inorganic total fluoride determination	BS 6068-2.48
<i>Miscellaneous methods involving distillation</i>	
Residue from essential oils after distillation	BS 5991
Sulfur dioxide in fruit and vegetable juices	BS EN 13196
Determination of boron in steel	ISO 13900

from foods and plant materials, and the herbicide and pesticide residues in agricultural products. The method involves steam distilling the compound of interest from an aqueous suspension of the crude sample while the condensed steam is continuously extracted with an immiscible organic solvent refluxing within the apparatus. The design of the apparatus allows the volatiles that are extracted from the condensed water to be flushed into the flask containing the organic solvent. After a previously determined time of extraction, the apparatus may be disassembled and the organic solvent removed by evaporation from the now concentrated extract. Further analytical techniques can be used to identify and quantify the components of the residue according to the particular requirements.

Distillation is an accepted method of material purity specification when carried out according to standard and agreed procedures. It is also a method of sample pretreatment prior to analysis by other methods including spectroscopy, titrimetry, or potentiometry.

## Further Reading

AnalaR Standards (1984) *AnalaR Standards for Laboratory Chemicals* (AnalaR is a registered trademark of Merck Ltd.).

*Annual Book of American Society for Testing and Materials*, ASTM International, 100 Barr Harbor Drive, West Conshohocken, PA, USA.

BSI Group, British Standards House, 389 Chiswick High Road, London W4 4AL, UK, [www.bsi-global.com](http://www.bsi-global.com)

Godefroot M, Sandra P, and Verzele MJ (1981) *Chromatography* 203: 325.

Institute of Petroleum (1998) *Standard Methods for Analysis and Testing of Petroleum and Related Products*. New York: Wiley.

Likens ST and Nickerson GB (1964) *American Society of Brewing Chemists, Proceedings*, 5.

Perrin DD and Armarego WL (eds.) (1988) *Purification of Laboratory Chemicals*, 3rd edn., pp. 5–12. Oxford: Pergamon Press.

Robillard MV, Spock PS, and Whitford JH (1991) *An Overview of Methodology and Column Technology for Simulated Distillation Analysis*. Bellefonte, PA: Supelco.

Stichlmair JG and Fair JR (1998) *Distillation: Principles and Practices*. New York: Wiley-VCH.



# DNA SEQUENCING

**K R Mitchelson**, Australian Genome Research Facility, Brisbane, QLD, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

DNA sequencing is very big business. Approximately US\$3 billion was spent in 2003 on sequencing reagents and enzymes, and on the analyzer equipment and software for automated sequence acquisition. The majority of this sequence output was determined using capillary electrophoresis (CE) technology, which has commensurately developed rapidly over the past 10 years. CE offers high resolution and high throughput, automatic operation, and data acquisition, with online detection of dyes bound to DNA extension products. Operational advances such as pulsed-field and graduated electric fields and automated thermal ramping programs as the run progresses result in higher base resolution and longer sequence reads. Advanced base-calling algorithms and DNA marker additives that utilize known fragment sizing landmarks can also help to improve fragment base-calling, increasing call accuracy and read lengths by 20–30%. Despite the high efficiency of CE sequencers, the complete delineation of the human genome and its implication for genome-wide analysis for personalized medicine is driving the development of devices and chemistries capable of massively increased sequence throughput, compared to the conventional CE sequencers. Miniaturization of CE onto chip-based devices provides all of the above facilities – a significant improvement in the speed and improved automation of analysis. New array-based sequencing devices also promise a quantum increase in efficiency. Each of these new devices provides an extremely high throughput, high-quality data, and low-process costs. This article also examines the automation and improvement of sequencing processes, DNA amplification processes, and alternative approaches to sequencing.

## High-Throughput Capillary-Array Sequencing

A CE instrument comprises two electrolyte chambers linked by a thin silica capillary 50–100  $\mu\text{m}$  in diameter, or as a fine microchannel on a silicon chip. The thin capillary rapidly dissipates heat generated by the large electric fields, stabilizing band resolution. An

online detector positioned close to the capillary outlet acquires data from the size-fractionated molecules. Typically, dyes attached to the DNA fragments are detected using laser-induced fluorescence (LIF). During the recent accumulation phase of genome research, electrophoresis-based capillary-array sequence analysis developed rapidly becoming the paradigm for DNA sequencing and providing simultaneous multiparallel analyses. Capillary-array electrophoresis is a multiplied version of conventional CE, with up to 100 parallel capillaries or channels on miniature CE chips that each simultaneously analyze individual samples (see **Figure 1**).

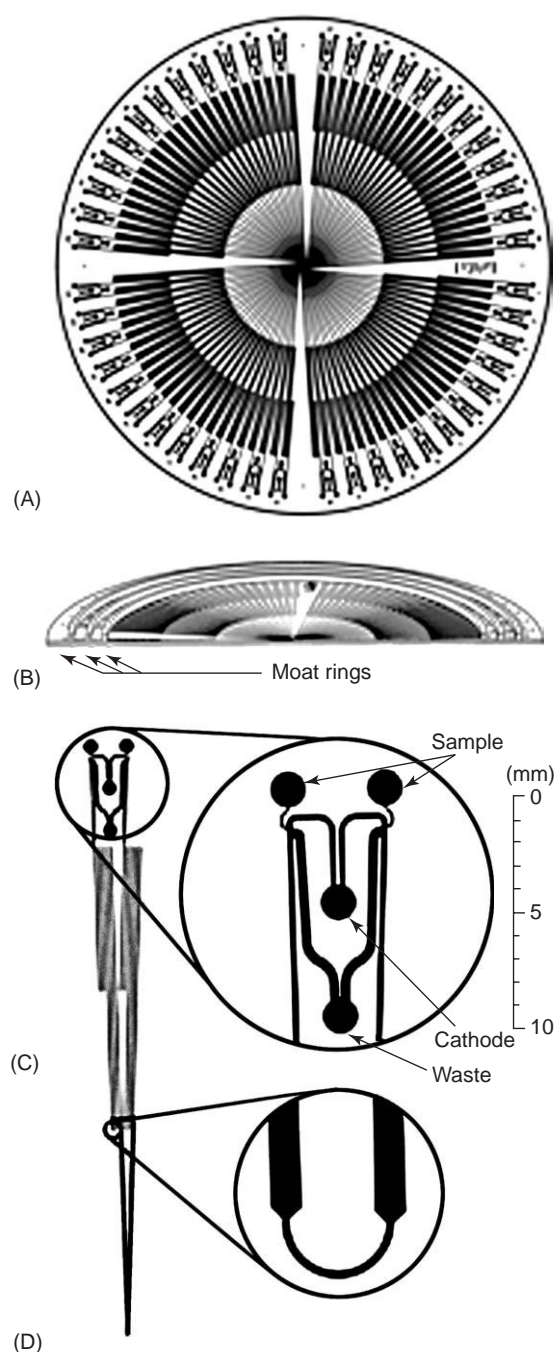
## Signal Detection Dyes

Energy-transfer dye-tagged nucleotides are efficiently incorporated into DNA by available enzyme systems and have higher sensitivity and better spectral discrimination than earlier fluorophore dyes. Two-dimensional direct-reading spectrographs can analyze fluorescence simultaneously from each of the collimated capillary dye streams passing across a charge-coupled device (CCD) camera as the stream emerges from the end of each capillary. Devices with 384 capillaries process sequence data from as many as 2500 samples per day, with each sequence presented as a stained-glass plot representing the dye-space electropherogram of each capillary. New sensitive photometric devices can now detect extremely low numbers of signal molecules, and even single molecules. Alternative detection methods such as time-resolved fluorescence decay, electrochemical detection, chemiluminescence, and near-infrared (NIR) detectors can now be incorporated in microdevices, and recent development of efficient incorporation processes will make widespread application possible. Dye-tagged nucleotides with longer linker lengths and charge matching improve the incorporation of these bulky molecules into nascent DNA (see **Figure 2**). The positively charged linker drives unincorporated dye-terminator nucleotides in the opposite direction to the sequencing products, also eliminating the need for reaction cleanup.

## Microchip Electrophoresis Sequencing

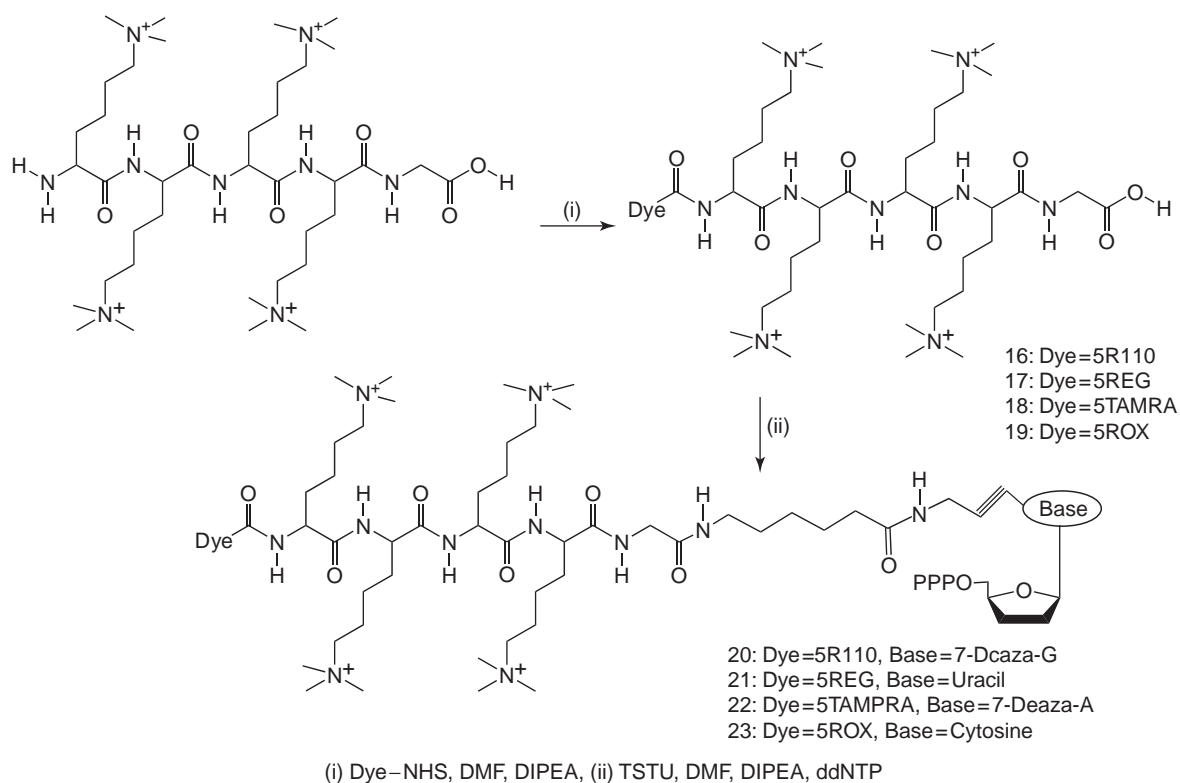
The microchip analysis format has significant advantages over conventional CE, being 10 times faster





**Figure 1** (A) Overall layout of the 96-lane DNA sequencing microchannel plate (MCP). (B) Cut-away of the MCP (vertical view). The concentric rings form two electrically isolated buffer moats that lie above the cathode and waste ports. (C) Expanded view of the injector. Each doublet has two sample reservoirs and common cathode and waste reservoirs. The sample to the separation channel arm is  $85\ \mu\text{m}$  wide, and the waste to separation channel arm is  $300\ \mu\text{m}$  wide. The separation channel is  $200\ \mu\text{m}$  wide. (D) Expanded view of the hyperturn region. The turns are symmetrically tapered with length  $100\ \mu\text{m}$ , turn channel width  $65\ \mu\text{m}$ , and a radius of curvature  $250\ \mu\text{m}$  (Reproduced with permission from Paegel BM, Blazej RG, Wedemayer GJ, *et al.* (2002) *Proceedings of the National Academy of Science USA* 99: 574–579; © National Academy of Sciences, USA.)

and using submicroliter volumes of analysis reagents. Strand-displacement amplification (SDA) and rolling-circle amplification are isothermal *in vitro* methods for amplification of a specific DNA sequence to concatomer lengths. Burns and colleagues elegantly demonstrated the extremely rapid DNA analysis using an integrated microchip system incorporating both DNA amplification and CE separation of SDA products. Significant advances in the amplification and detection of single DNA template molecules on integrated devices are providing unprecedented levels of sensitivity. New constriction-channel designs improve fragment resolution and increase the scope for longer pathlengths, permitting single base resolution over longer fragments. Revolutionary two-dimensional separation CE devices represent a new paradigm for sequence analysis, when coupled with two-dimensional array detectors. More conventionally, new low-voltage closed-loop CE devices offer the promise of handheld, or readily transportable analysis instruments. Microfabricated multireflection absorbance cells for microchip-based CE have also been built with 5- to 10-fold enhanced sensitivity over single-pass devices. Currently, these schemes are being built into devices with hundreds of capillaries to achieve high speed, extremely high throughput, and higher detection sensitivity. Microdevices that require no operator intervention and which integrate sample purification, sample amplification, amplicon product purification, and DNA sequencing by CE have been developed (see Figure 3). Unincorporated dye terminators are electrically separated from sequencing products under high voltage into a waste channel, prior to diverting the sample into a separation capillary for size resolution and sequence analysis. Micro-machined sheath-flow cuvettes that precisely control both capillary alignment and matrix flow. And confocal LIF systems with one lens for both excitation and detection optics, scanning continuously across a bundle of capillaries, can provide both longer reads and more accurate base-calling. Other nanoreactors, with serial electrodes that provide for high ‘sweeping field’ separation using low-voltage supplies suited to handheld devices, can achieve polymerase chain reaction (PCR) amplification in 15 min and CE analysis in 2 min. The nanoreactors can be interfaced either to microelectrophoresis chips or capillary gel tubes via micromachined capillary connectors or zero-dead volume unions, and signals are detected using a NIR fluorescence detector. Sequencing reactions from a small ( $\sim 60\ \text{nl}$ ) volume reactor could be coupled directly to a capillary gel column for separation with minimal loss in the efficiency of the separation process and capable of single-base resolution to  $>450$



**Figure 2** Synthesis of single dye-labeled oligo-trimethyllysine-modified 2',3'-dideoxynucleoside-5'-triphosphates. (Reprinted with permission from Finn PJ, Bull MG, Xiao H, *et al.* (2003) *Nucleic Acids Research* 31: 4769–4778; © Oxford University Press.)

bases. These devices permit complete CE analysis in an easily transportable format.

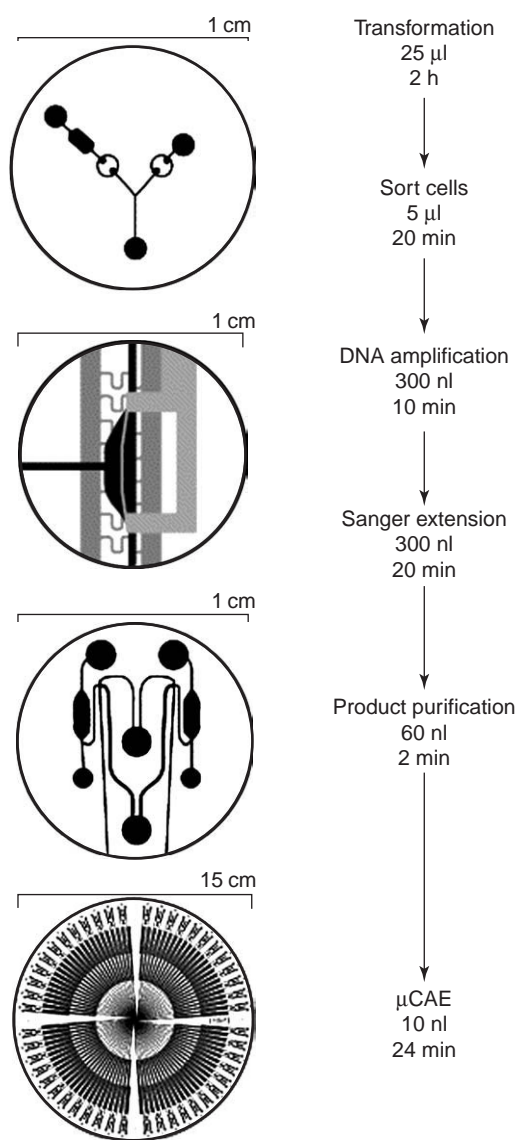
## Solid-Phase and Array-Sequencing Devices

The demand for ultrahigh throughput sequencing for personal medicine is driving the development of nonelectrophoretic methods for sequencing multiple single molecules in parallel. These devices use wide-ranging technologies such as array-based solid-phase techniques, sequencing by synthesis methods, ultra-sensitive optics, and mass spectrometry (see **Table 1**). If current process problems are solved, these new nanosequencing technologies could markedly increase in sequence capacity required by genome-sequencing projects and large genome resequence analysis programs.

## Single DNA Molecule Sequencing

The need for analysis of single DNA molecules has coincided with the development of technologies capable of single molecule sensitivity. The miniaturization

of chip CE systems with nanochannels  $< 1 \mu\text{m}$  allows analysis to be undertaken on limited numbers of molecules at pico- and nanomolar concentrations, with amplification and detection of signals from single template molecules. Single DNA molecule imaging can achieve simultaneous analysis of up to 100 000 distinct molecules every second. Several solid-state methods for single DNA molecule sequencing have been reported recently, again with promise of highly parallel, genome-scale efficiencies. 'Single-molecule sequencing with exonuclease' comprises the serial digestion of a single-DNA strand attached to a solid surface or microchannel surface. The fluorescent-tagged nucleotide subunits are sequentially released, then collected, and detected. The method demands highly efficient enzymatic incorporation of labeled analogs at each subunit nucleotide position. US Genomics offer another potential solution, with technologies that can directly analyze individual molecules of DNA at ultrahigh throughputs without amplification (e.g., PCR). The process of direct analysis begins with fluorescently tagging the sample genomic DNA at specific nucleotide sequence motifs. The sample is then interrogated using several lasers as it passes through a nanofluidic



Current opinion in Biotechnology

**Figure 3** A microfabricated 'sequencing factory-on-a-chip'. An individual subclone is sorted and directed to a PCR amplification chamber. Single-cell amplification is followed by template purification. The purified template is moved to a thermal cycling reactor for generation of sequencing extension fragments. The extension fragment products are isolated, concentrated, and then injected into an electrophoretic microchannel for length analysis. The total estimated process volume from cell to called bases is  $\sim 1 \mu$ l. (Reprinted with permission from Paegel BM, Blazej RG, and Mathies RA (2003) Microfluidic devices for DNA sequencing: Sample preparation and electrophoretic analysis. *Current Opinion in Biotechnology* 14: 42–50; © Elsevier.)

analyzer instrument, with thousands of molecules analyzed per minute. For DNA analysis, this represents a throughput of 10–30 million base pairs per second, with each molecule detected by the laser excitation of the fluorescent tags on the molecule.

## Nanofluidic Barrier Sequencing

Sequencing methods employing physical techniques and nanoscale fluidics combined with ultrasensitive optical systems are being developed to sequence single molecules, one molecule at a time. The nanodevices sort DNA molecules by using openings or pores or forests of pillars that are less than the 'radius of gyration' of the DNA fragments, just large enough for DNA molecules to run through in single file. When a DNA molecule is placed outside a 'forest' of tiny pillars arranged in a square grid and the molecules forced to move into the grid by an electric field, the DNA must uncoil to pass through. By varying the electric field pulse, the length of the DNA strands that are collected in the grid can be controlled and thus separated by size (length). When a mix of DNA fragments is driven electrically along a channel with several such barriers, fragments of different lengths will arrive at the far end in a series of bands not unlike those seen in conventional gel electrophoresis separations. Single-molecule sequencing (SMS) promises to radically improve DNA sequencing as it is potentially 10 000 times faster than current production systems that rely on single lanes. It can potentially start with genomic DNA reducing the need for sample preparation. SMS read lengths are also potentially significantly longer than those obtained from gel electrophoresis systems. Longer read-lengths will simplify sequence reconstruction and reduce the total number of runs required to get full coverage of the genome. Uniquely, SMS sequencing can directly detect haplotypes over several polymorphisms.

## Massively Parallel Signature Sequencing

Massively parallel signature sequencing combines techniques developed for expression analysis with data accumulated from representative but uncharacterized cDNA libraries. The system comprises microbeads with 32-mer tags comprising strings of eight 4-mer 'words' that hybridize amplified cDNA library targets. The beads each hybridize  $\sim 100\,000$  molecules of a unique cDNA, the beads are then sorted on an FACS analyzer, and distributed in a micromachined flow cell into a planar array. The sequence of each immobilized cDNA bead is then 'read' successively by the restriction enzyme *BbvI* after linking of encoded 'adapters' to the end of each exposed specific cDNA overhang and by a second set of fluorescently labeled 'decoder' oligonucleotides. The system has large capacity for sequence analysis, with some one million cDNA-coated beads being

**Table 1** Websites providing insights into different aspects of novel DNA sequencing technologies

Websites	Technologies
<a href="http://home.appliedbiosystems.com">http://home.appliedbiosystems.com</a> <a href="http://amershambiosciences.com">http://amershambiosciences.com</a>	Array capillary electrophoresis instrumentation and cyclic dye terminator chemistry.
<a href="http://www.sequenom.com">http://www.sequenom.com</a> <a href="http://www.methexis-genomics.com">http://www.methexis-genomics.com</a> <a href="http://www.nuvelo.com">http://www.nuvelo.com</a>	MALDI-TOF mass spectrographic sequencing. SNUPE and small oligomer fragmentation sequencing.
<a href="http://www.pyrosequencing.com">http://www.pyrosequencing.com</a>	'Sequencing by synthesis'. Sequential polymerization enzymology.
<a href="http://www.454.com">http://www.454.com</a>	Massively parallel 'sequencing by synthesis'.
<a href="http://www.solexa.co.uk">http://www.solexa.co.uk</a> <a href="http://www.genovoxx.de">http://www.genovoxx.de</a> <a href="http://www.visgen.com">http://www.visgen.com</a>	Array-based chemistry and advanced signal detection technologies. Advanced base-calling software.
<a href="http://www.nanofluidics.com">http://www.nanofluidics.com</a> <a href="http://www.Ionian-technologies.com">http://www.Ionian-technologies.com</a> <a href="http://www.usgenomics.com">http://www.usgenomics.com</a>	Nanofluidic barrier technology. Nanopore technologies. Sequencing of single DNA molecules.
<a href="http://www.molecularstaging.com">http://www.molecularstaging.com</a> <a href="http://www.fidelitysystems.com">http://www.fidelitysystems.com</a> <a href="http://www.nucleics.com">http://www.nucleics.com</a> <a href="http://www.stratagene.com">http://www.stratagene.com</a>	Alternative DNA sequencing tools. DNA-barrier breaking thermostable enzymes, sequencing enhancers, advanced base-calling software.
<a href="http://www.nimblegen.com">http://www.nimblegen.com</a> <a href="http://www.affymetrix.com">http://www.affymetrix.com</a> <a href="http://www.usgenomics.com">http://www.usgenomics.com</a>	DNA and oligonucleotide arraying technologies. Sequencing and re-sequencing by hybridization. Optical masking.
<a href="http://www.mpimg-berlin-dahlem.mpg.de/~capri">http://www.mpimg-berlin-dahlem.mpg.de/~capri</a> <a href="http://www.lab-on-a-chip.com/home">http://www.lab-on-a-chip.com/home</a> <a href="http://thebigone.caltech.edu/quake/research">http://thebigone.caltech.edu/quake/research</a>	Laboratory on a chip technologies, nano and microfluidic engineering. CE technologies.
<a href="http://www.cchem.berkeley.edu/~cab">http://www.cchem.berkeley.edu/~cab</a> <a href="http://depts.washington.edu/chemfac/dovich">http://depts.washington.edu/chemfac/dovich</a> <a href="http://www.chem.harvard.edu/faculty/whitesides.html">http://www.chem.harvard.edu/faculty/whitesides.html</a>	Fundamental advances in dye chemistries, CE equipment and micro/nanofabricated device engineering.

simultaneously analyzed in a single machine experiment within several days, with over  $\sim 20$  bases read per bead. This technology provides a 'signature' for each cDNA rather than full-length sequence analysis.

## Sequencing by Synthesis

'Sequencing by synthesis' is a method common to primer extension methods such as single nucleotide primer extension (SNUPE) and pyrosequencing, in which a unitary base addition chemistry that allows single nucleotide additions to growing chains to be monitored on each oligonucleotide feature, simultaneously with the addition of one of four differentially labeled, terminating nucleotides. Church and colleagues initiated the integration of solid-state DNA sequencing using polymerase colonies ('colonies') and cycles of fluorescent dNTP incorporation with high signal sensitivity, which allow multiple colonies to be sequenced in parallel (see **Figure 4**). Large-scale arrays

of discrete 'colonies' can be extended cyclically, bringing closer cost-effective genome-scale single array sequencing. 'Fluorescence *in situ* sequencing' involves the addition of one nucleotide; the extended fragments are then all detected simultaneously using CCD optics, the terminating moiety and fluor-tag are then removed chemically from each attached nucleotide ready for the addition of the next nucleotide to each chain in the following cycle. The series of base additions (e.g., C A T G C A T G, ...) is interrupted by scanning (for data acquisition) and chemical treatment of slides to remove signal prior to the next extension step. Ideally, the polymerase catalyzed strand extensions continue for 30 nucleotides or more.

The technology company 454 Life sciences has developed a solid-phase parallel microarray system of microfluidic wells in which  $10^8$  oligonucleotide molecules per square centimeter are used to capture complementary DNA fragments. Pyrosequencing reactions are simultaneously monitored on each

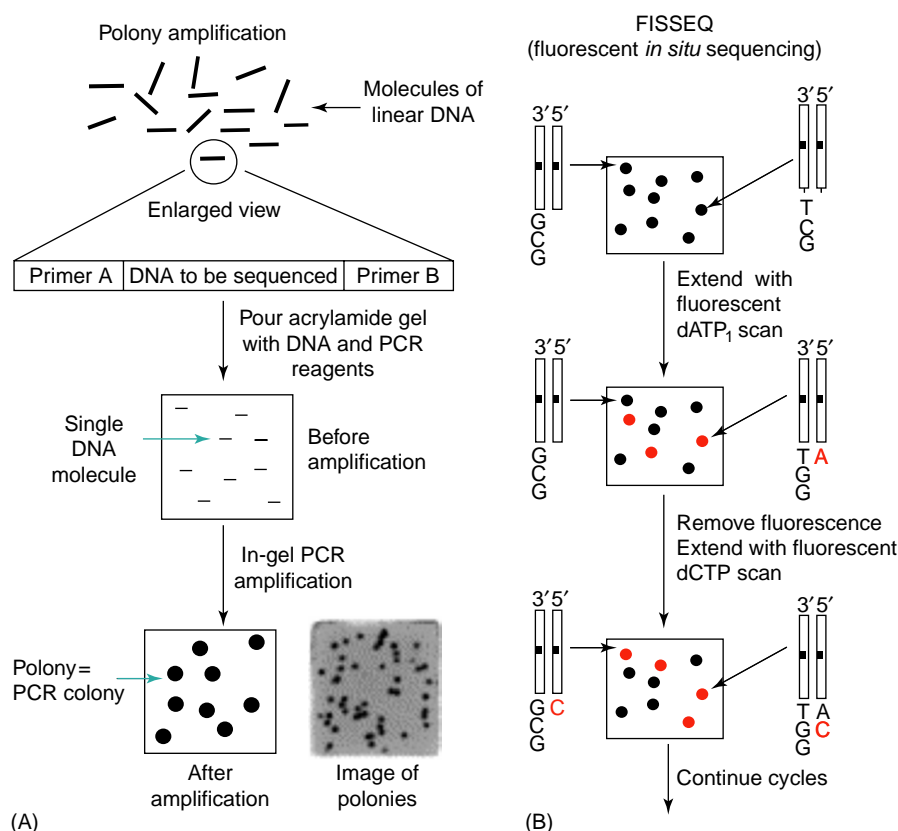


well-attached molecule, for each base addition. Solexy Pty Ltd. is another company, which sequences on arrays using a nanofeature format innovation such as zero-wave guide, a device confining optical excitation and detection to the few zeptoliters of fluid surrounding the polymerase, and the identification of nucleotides by fluorescent labels attached to the  $\gamma$ -phosphate leaving group of the dNTP. These two different arrayed analysers are expected to develop sequencing rates of  $10^8$  base-reads per day, the equivalent of a billion-lane sequencer that reads the sequence of each molecule at the speed of the addition reaction. Although currently the efficiency and uniformity of extension is poor, it is expected that if each molecule could be extended by an average 50 nucleotides, it will allow parallel discovery and detection of genetic variation on  $10^8$  molecules that can be aligned to known reference sequence (such as the human genome). 'Sequencing by synthesis' may permit *de novo* sequencing of entire genomes of low repetition, such as phage, virus, and bacterial genomes. In August 2003, 454 Life Sciences announced that 'sequencing by synthesis' technology

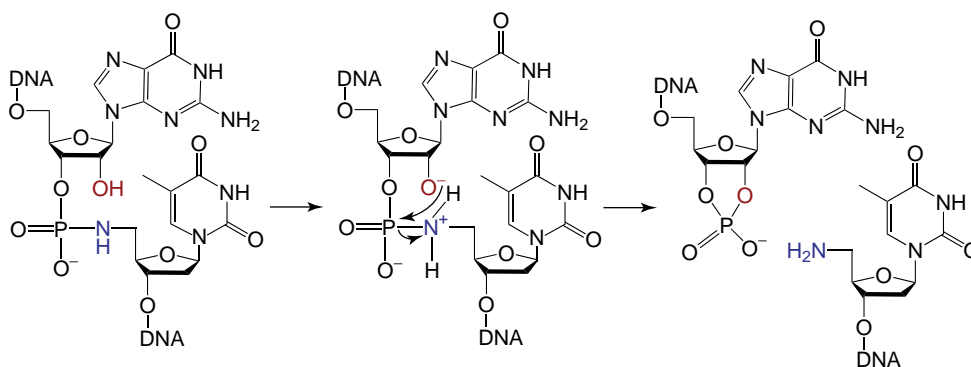
was used to sequence the entire 35 kb genome sequence of adenovirus, which due to its low complexity is amenable to this analysis.

## Sequencing by Mass Spectrometry

Mass spectrometry is used to determine the sequence of a polynucleotide by analysis of the atomic masses of a series of polynucleotide subfragments derived by the partial and uniform fragmentation, or by the uniform extension of the polynucleotide. The polynucleotide subfragments are released from a solid phase and analyzed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). The techniques include partial enzymatic cleavage, chemical cleavage, base-specific cleavage of PCR products, and primer extension Sanger sequencing SNUPE methodologies. Nucleotide analogs are used widely during DNA sequence analysis by mass spectrometry – to modify polynucleotide electrochemistry and stabilize the *N*-glycosidic linkages, to block base loss and subsequent random backbone cleavage, or to introduce



**Figure 4** (A) Polony amplification. A library of linear DNA molecules with universal priming sites is PCR amplified within a polyacrylamide gel. Each single template molecule gives rise to a polymerase colony or 'polony'. (B) Fluorescent *in situ* sequencing. Polonies are denatured, and a sequencing primer is annealed. Polonies are sequenced by serial additions of a single fluorescent nucleotide. (Reprinted with permission from Mitra RD and Church GM (2003) Fluorescent *in situ* sequencing on polymerase colonies. *Analytical Biochemistry* 320: 55–65; © Elsevier.)



**Figure 5** Proposed dinucleotide chemistry that permits cleavage at a two base sequence, such as at GT. (Reprinted with permission from Wolfe JL, Bing HW, Tomohiko K, and Vincent PS, Jr. (2003) Sequence-specific dinucleotide cleavage promoted by synergistic interactions between neighboring modified nucleotides in DNA. *Journal of the American Chemical Society* 125(35): 10500–10501; © American Chemical Society.)

site-specific backbone weakness for controlled fragmentation. Wolfe and others have employed analogs such as 7-deaza analogs, 5'-amino-5'-deoxy- and 5'-amino-2',5'-dideoxy-analogs, and stable mass isotope tags to substitute for particular nucleotides (see **Figure 5**). These nucleotide analogs introduce differential mass properties, or differential stability into the DNA subfragments that improve the mass separations. Boranophosphate-modified nucleotides incorporated randomly into DNA can reveal sequencing ladders directly by exonuclease digestion. Borano analogs form standard Watson–Crick base pairs, but are preferentially and uniformly cleaved compared to native nucleotides at their phosphoribose backbone. Similarly, acids and bases can cleave dideoxy and amine-modified analog backbones generating sequencing ladders, which may be analyzed by mass.

## Sequencing of Problematic DNA

Several new methods to overcome DNA motifs that are refractile to conventional sequencing have been developed. Transcriptional sequencing using RNA polymerases and dye-tagged ribonucleotides improve the sequencing of GC-rich regions, simple sequence repeats, hairpins (inverted repeats), tandem repeat DNA templates, and gap-closing in draft sequencing data. 'Thermofidase' is another highly effective sequencing system using engineered DNA polymerases and chimeric thermostable 'DNA polymerase' and 'topoisomerase V' in combination with cycle sequencing and modified oligonucleotides. These chimeric polymerases resist inhibitors and are also particularly effective for sequencing regions of GC-rich DNA and highly repeated centromeric DNAs. Another method for tackling of poly A- and AT-rich tracts uses the excision or truncation of the A/T-rich region. Poly A tails are shortened by type IIS restriction enzymes such as *GsuI* before cDNA cloning.

Whilst another method relies on the slippage of mung bean nuclease, which digests A/T-rich double-stranded DNA into a set of deletion fragments that can then be cloned and sequenced. 'Sequencing aided by mutation' is a radically different method that overcomes sequencing difficulties caused by 'problematic motifs' – regions where local sequence characteristics hinder existing sequencing technologies. Here, problematic motifs and regions are mutated in a random manner, sufficiently to overcome the obstruction to sequencing or cloning. The random locations of the mutations cause the copies of the region to have different sequences. The DNA sequences determined from a low number of the altered copies are then analyzed using Bayesian methods to reconstruct the original wild-type sequence, with cost efficiencies and accuracy similar to conventional sequencing.

**See also:** **Capillary Electrophoresis:** Clinical Applications. **Forensic Sciences:** DNA Profiling. **Lab-on-a-Chip Technologies.** **Mass Spectrometry:** Polymerase Chain Reaction Products. **Nucleic Acids:** Chromatographic and Electrophoretic Methods.

## Further Reading

- Brackmann S and Lobermann S (2002) A further step towards single-molecule sequencing: *Escherichia coli* exonuclease III degrades DNA that is fluorescently labelled at each base pair. *Angewandte Chemie: International Edition English* 41: 3215–3217.
- Braslavsky I, Hebert B, Kartalov E, *et al.* (2003) Sequence information can be obtained from single DNA molecules. *Proceedings of the National Academy of Science USA* 100: 3960–3964.
- Brenner S, Johnson M, Bridgham J, *et al.* (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology* 18: 630–634.



- Bruin GJ (2000) Recent developments in electrokinetically driven analysis on microfabricated devices. *Electrophoresis* 21: 3931–3951.
- Cochran DAE, Lala G, Keith JM *et al.* (2003) Sequencing by aligning mutated DNA fragments. In: Xing WL and Cheng J (eds.) *The Frontiers of Biochip Technologies*. Dordrecht: Kluwer Publishers.
- Deamer DW and Akeson M (2000) Nanopores and nucleic acids: Prospects for ultra-rapid sequencing. *Trends in Biotechnology* 18: 147–151.
- Hunt AR, Willey DL, and Quail MA (2003) Finishing genom sequence and dealing with problem sequences. In: Dunham I (ed.) *Genome Mapping and Sequencing*, pp. 315–355. Cambridge: Horizon Press.
- Krishnan M, Namasivayam V, Lin R, *et al.* (2001) Micro-fabricated reaction and separation systems. *Current Opinion in Biotechnology* 12: 92–98.
- Marziali A and Akeson M (2001) New DNA sequencing methods. *Annual Review of Biomedical Engineering* 3: 195–223.
- Mitchelson KR and Cheng J (eds.) (2001) *Capillary Electrophoresis of Nucleic Acids*, vols I and II. Methods in Molecular Biology, vols 162 and 163. Totowa: Humana Press.
- Murray KK (1996) DNA sequencing by mass spectrometry. *Journal of Mass Spectrometry* 31: 1203–1215.
- Sauer M, Angerer B, Ankenbauer W, *et al.* (2001) Single molecule DNA sequencing in submicrometer channels: State of the art and future prospects. *Journal of Biotechnology* 86: 181–201.
- Shendure J, Mitra R, Varma C, and Church GM (2004) Advanced sequencing technologies: Methods and goals. *Nature Reviews in Genetics* 5: 335–344.

## DRUG METABOLISM

Contents

### Overview

### Metabolite Isolation and Identification

### Isotope Studies

## Overview

**S A Roberts**, AstraZeneca R&D, Lund, Sweden

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A knowledge of the fate of a drug, its disposition (absorption, distribution, metabolism, and excretion; ADME) and pharmacokinetics (PK) (the mathematical description of these processes) is important throughout pharmaceutical research and development. The aim of a typical pharmaceutical company is to identify and develop good orally active candidate molecules that have suitable potency and selectivity against a target, sufficient oral bioavailability, metabolic stability, and duration of action. The importance of ADME/PK has been highlighted in surveys that have revealed that the main causes of failure during development are unacceptable clinical efficacy, toxicity, and poor PK. In fact, as much as 40% of all failures have been attributed to poor PK. *In silico* (computer-based), *in vitro* (cell-based), and *in vivo* (animal-based) techniques are

available for investigating the ADME/PK of new chemical entities (NCEs). Both the scientific concepts underpinning this discipline and the techniques and rationale employed will be detailed in this article.

## The Science of ADME/PK

All drugs have the potential to produce unwanted effects. In drug research and development it is essential to select chemicals that have a margin of safety between the dose that produces the desired (therapeutic) effects and the dose that produces undesired (toxic) effects.

The investigation of ADME/PK aids in the discovery and selection of new drugs, supports safety assessment, and helps define conditions for effective use in patients. The main objectives of ADME/PK studies can therefore be summarized as shown in Table 1.

### Drug Absorption

For a drug to work it must be a solution at the site of delivery. If orally administered it must dissolve in the lumen of the small intestine. Once in solution it must

**Table 1** Main objectives of ADME/PK studies

To characterize the basic pharmacokinetic parameters of a test substance
To understand the mechanism of action of a drug
To support toxicology studies
To validate the appropriateness of the animal species used in toxicology
To establish margins of safety between the therapeutic and toxic effects
To optimize dosage regime (size, formulation, frequency)
To establish target organs/modes of absorption and elimination
To establish nature of metabolism
To provide data for regulatory authorities

be absorbed through the walls of the intestine and subsequently cross the barriers between the blood stream and the site of action. Absorption can therefore be defined as the net transfer of a substance from its site of administration to the site of effect.

The entry of many orally administered drugs into the systemic circulation is dependent on their tendency to partition into lipid (lipophilicity). Lipophilic properties facilitate a compound's penetration of lipid membranes and transportation in blood lipoproteins but their lower water solubility reduces urinary excretion. In addition, lipophilic compounds are more likely to undergo extensive metabolism.

The degree of ionization of a drug can dictate how lipid soluble it is. Ionization, as measured by the dissociation constant ( $pK_a$ ), can vary with the pH of the environment. When the pH of a solution is equal with  $pK_a$  the concentration of the ionized species is the same as that of the unionized species. As many drugs are either weak acids or weak bases, the degree of ionization can impact on their absorption, distribution, and excretion characteristics as shown in Figure 1.

The solubility of a compound is a barrier that limits its concentration in a solution and therefore the amount of material available for absorption. A compound's crystal type (polymorph), its particle size, wettability, and the use of solubilizing agents in formulations can all influence solubility. Also, the formation of salts, esters, and complexes with a compound can improve its solubility. Solubility is a key parameter for determining the dissolution rate of compounds, a more informative measure that takes into account the form and particle size distribution in a particular medium.

A number of mechanisms of absorption exist but the most important of these, for the vast majority of compounds, is passive diffusion where the movement of compound is driven by the concentration gradient that can occur between cell junctions (paracellular)

or through cells (transcellular). Lipophilic compounds tend to cross the plasma membrane most readily and utilize the transcellular route, while hydrophilic compounds tend to utilize the paracellular route. In addition, some substances can be transported by an active system either with or against a diffusion gradient. An important transporter protein, P-glycoprotein (P-gp), is present in the plasma membrane of many cells and is responsible for the efflux of a wide range of substrates. The presence of a significant P-gp efflux capacity in the intestinal lumen reduces the overall absorption of many compounds. The main mechanisms of transport that can occur in the intestinal epithelium are shown schematically in Figure 2.

The small intestine is folded into villi and microvilli which increases surface area many times. Even poorly soluble compounds often undergo reasonable absorption because of the large surface area. Absorption can be reduced if blood flow in the gastrointestinal tract is insufficient to maintain 'sink conditions' or if transit times within the tract are too rapid.

### Drug Bioavailability

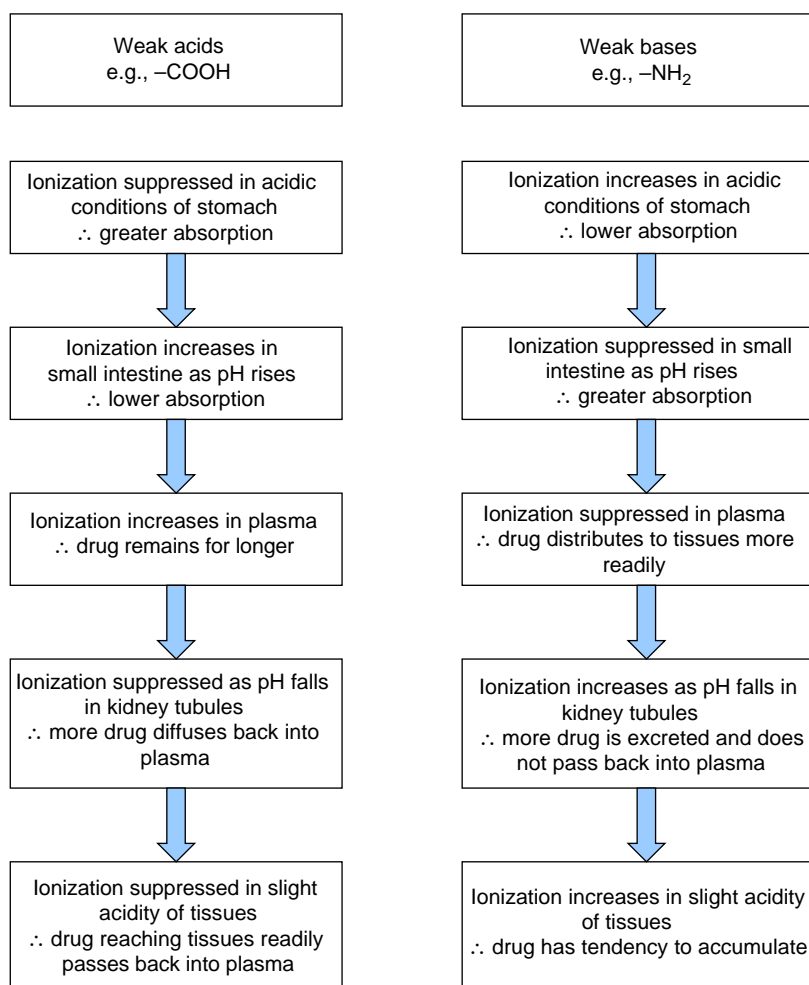
Following absorption, first-pass metabolism can reduce the total exposure of the body to drug. First-pass metabolism refers to any loss of the administered material by transmucosal or hepatic means after absorption and before reaching the systemic circulation, and this is shown schematically in Figure 3.

Bioavailability is a term used to describe the systemic availability of drug and can be defined as the rate and extent of appearance of unchanged drug in the systemic circulation following an extravascular (e.g., oral) dose. Experimentally, it is determined as the fraction of the maximal levels of drug present in the systemic circulation after an intravenous dose. It takes into account both absorption and metabolism and is dependent on area under the curve (AUC), peak concentration achieved ( $C_{max}$ ), and time to reach peak concentration ( $t_{max}$ ). It is generally expressed as a percentage and is represented by eqn [1]:

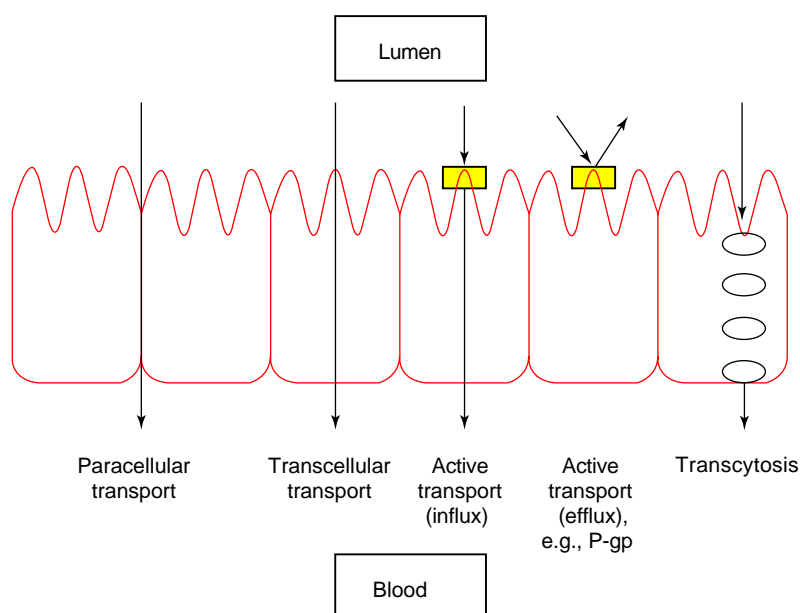
$$\% \text{Bioavailability} = \frac{\text{AUC}_1 \times \text{Dose}_2}{\text{AUC}_2 \times \text{Dose}_1} \times 100 \quad [1]$$

where 1 refers to an oral dose and 2 refers to an intravenous dose.

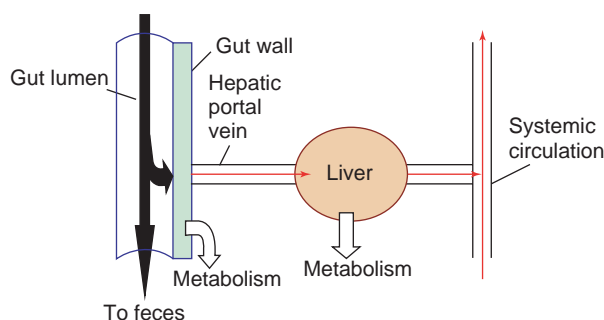
The assessment of many PK parameters depends upon the accurate measurement of AUC. Although software is used routinely to do this, the calculations are straightforward and are based on the sum of



**Figure 1** Consequences of ionization for weak acids or bases during their passage through the body.



**Figure 2** Multiple mechanisms of transport through the intestinal epithelium.



**Figure 3** Schematic diagram to show sites of first-pass metabolism.

trapezoid areas under the concentration versus time curve for each sampling interval (trapezoidal rule).

As pharmacological effect depends upon the quantity of unchanged drug reaching the target organ and its residence time, it is evident that changes in absorption and metabolism will lead to changes in bioavailability and therefore activity.

Relative bioavailability (bioequivalence) is the comparison of the rate and extent of drug appearance in the systemic circulation when one formulation is compared with another. When two or more formulations of the same drug produce statistically indistinguishable rates and extents of appearance they can be said to be bioequivalent.

### Drug Distribution

Drug distribution can be defined as the (reversible) transfer of drug around the body. This is usually very quick within the vascular system but variable into tissues, being dependent on a number of factors including blood perfusion rate, drug diffusion rate, plasma protein binding, and tissue binding.

The amount of drug distributed cannot normally be measured directly and so is inferred from PK analysis of the plasma concentration–time profile utilizing eqn [2]. The apparent volume of distribution ( $V_d$ ) can be defined as the volume that a drug would occupy were the drug present in all tissues at the same concentration as observed in the plasma:

$$V_d = \frac{\text{amount in body (dose)}}{\text{plasma concentration } (C_0)} \quad [2]$$

where  $C_0$  is the plasma concentration if drug had been instantaneously distributed into its final volume following an intravenous dose;  $V_d$  can vary from 0.05 to 800 l kg<sup>-1</sup>. The lower values are usually for drugs that remain in the plasma compartment, i.e., they are hydrophilic and/or tightly bound to plasma. The higher values are for drugs that have a high affinity for tissues.

When a membrane is not a substantial barrier to a drug, the rate of perfusion of the tissue becomes the limiting factor; this is commonly the case for small lipophilic molecules. Distribution occurs rapidly in highly perfused organs (e.g., lungs) and slowly in poorly perfused organs (e.g., fat). A diffusion rate limitation occurs when a tissue membrane is a substantial barrier to the movement of drug, e.g., the blood–brain barrier has poorly permeable capillary membranes that are effective in blocking the passage of many drugs into the brain. The rate of distribution is also dependent on total body mass; therefore, distribution equilibrium is likely to be achieved faster in smaller species.

As the most abundant plasma protein is albumin (which prefers to bind acidic drugs), the total binding for acidic drugs can be very high (greater than 99.9%) and therefore the volumes of distribution can be very low. Less abundant but still important is the plasma protein  $\alpha$ 1-acid glycoprotein that tends to bind basic drugs. The extent to which a drug binds is governed by an equilibrium (or affinity) constant, the concentration of the protein and of the drug. Therefore, as only unbound ‘free’ drug will pass through membranes, the dissociation rate can control the rate of distribution. In addition to plasma protein binding, drugs can bind to various sites within tissues. Lipophilic weak bases, for instance, have a high affinity for the acidic phospholipids, which make up biological membranes, and therefore they tend to have high volumes of distribution.

Drug distribution is related, as shown in eqn [3], to the plasma elimination half-life ( $t_{1/2}$ ), the time it takes for the plasma concentration or the amount of drug in the body to be reduced by 50%:

$$t_{1/2} = \frac{V_d}{Cl_p} \times 0.693 \quad [3]$$

where  $Cl_p$  refers to plasma clearance (see eqn [6]).

The  $t_{1/2}$  will therefore increase if  $V_d$  increases or  $Cl_p$  decreases, and if a drug has a high  $V_d$  and a long  $t_{1/2}$  this can indicate tissue accumulation.

### Drug Elimination

Drug elimination can be defined as the (irreversible) transfer of a drug from the site of measurement (usually plasma or blood) by either excretion (e.g., renal, biliary, pulmonary, sweat and milk excretion) or metabolism.

Renal excretion is particularly important. Glomerular filtration removes unbound low molecular weight compounds and active secretion and passive reabsorption processes occur throughout the kidney tubules depending on a drug’s ability to cross lipid

membranes and degree of ionization. The net rate of renal excretion can be estimated using eqn [4]:

$$\begin{aligned} \text{Net rate of renal excretion} = & \text{Rate of filtration} \\ & + \text{Rate of secretion} \\ & + \text{Rate of reabsorption} \quad [4] \end{aligned}$$

Biliary excretion results in the elimination of drugs in the bile and is determined by molecular weight. Little biliary excretion occurs for molecules with molecular weight less than 300 and very large molecules such as proteins. The approximate molecular weight thresholds above which drugs can be excreted directly in bile are higher in man and monkey (500) than rabbit (475), guinea-pig (400), and rat and dog (325). It is important to note that because the threshold for rat and dog, the usual toxicology species, is somewhat lower than for man, this can lead to species differences for compounds of intermediate molecular weight.

Enterohepatic recycling is where compounds are returned to the gut via the bile and then reabsorbed. This process can lead to very long  $t_{1/2}$  values, e.g.,  $t_{1/2}$  for digoxin in dogs is raised from 6 to 14 h by recycling.

Expired air is an important route of elimination for anesthetics and some volatile compounds are readily excreted in sweat. The elimination of compounds in mother's milk can lead to high exposures for the baby when it is their only source of nourishment. The mucosal secretion of drug back into the gut can also be an important route of elimination.

Clearance is an important property of a drug, and as already indicated in eqn [3], reducing clearance can extend the  $t_{1/2}$ . It can be defined as the volume of blood (or plasma) from which all drug is removed in unit time and depends upon the flow of blood through organs and their extraction ratio (ER). The ER can be calculated from drug concentrations as shown in eqn [5]:

$$ER = \frac{C_{in} - C_{out}}{C_{in}} \begin{cases} \text{If organ extracts nothing } ER = 0 \\ \text{If organ extracts everything } ER = 1 \end{cases} \quad [5]$$

where  $C_{in}$  and  $C_{out}$  are the concentrations of drug entering and leaving the organ, respectively. In the case of the liver,  $C_{in}$  is the concentration in the hepatic portal vein blood and  $C_{out}$  the concentration in the systemic circulation.

The total clearance is the sum of the clearances in all organs. If the clearance of drug from blood/plasma approaches the blood/plasma flow through the liver, this indicates that the drug is likely to undergo hepatic first-pass removal. An estimate of  $Cl_p$

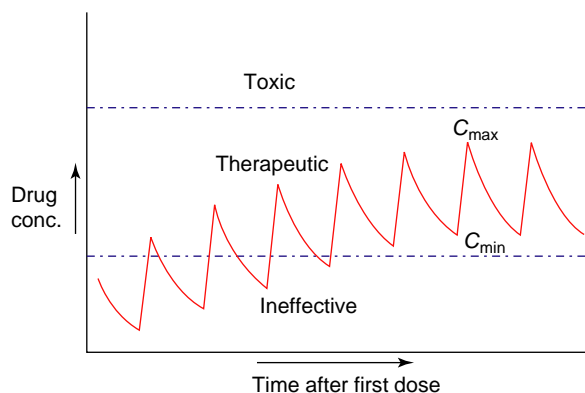
can be obtained from intravenous dosing data using eqn [6]:

$$Cl_p = \frac{\text{DOSE}}{\text{AUC}} \quad [6]$$

During a drug's development it is important to investigate the effect on exposure of changing the dose, and to prove that clearance is dose-independent. When  $Cl_p$ ,  $V_2$ , and  $t_{1/2}$  remain constant over a dose range and over time, a drug can be said to have linear PK. If an elimination process becomes saturated, AUC tends to increase with dose more than one would expect, if absorption or plasma protein binding are saturated, AUC tends not to increase with dose as much as one would predict.

Knowing the  $Cl_p$  and  $t_{1/2}$  of a drug one can modify the size and frequency of doses. Doses should be given at repeated intervals such that the 'rate in' equals the 'rate out'. If doses are too high and frequent, then toxic levels may be achieved. If doses are too low and infrequent, then the drug may prove ineffective. **Figure 4** shows the ideal situation where plasma concentrations reach a plateau and the same maximum ( $C_{max}$ ) and minimum ( $C_{min}$ ) concentrations are reproduced within the therapeutic window, i.e.,  $C_{min}$  is kept above the minimum effective concentration and  $C_{max}$  below the minimum toxic concentration.

While PK data indicate what the body does to the drug over time, pharmacodynamic (PD) data indicate what the drug does to the body over time. PK/PD modeling involves linking PK and PD data to assess the response-time relationship. The predictivity of models varies with their complexity and the type of data available. Often, response does not track directly with concentrations with the greatest effect being observed well after maximal concentrations have occurred. Generally, the most reliable models depend on the measurement of 'free' levels at the site



**Figure 4** Plot of plasma concentration versus time following the application of an 'ideal' dosing regimen.



of drug action and the accurate and precise measurement of drug effect.

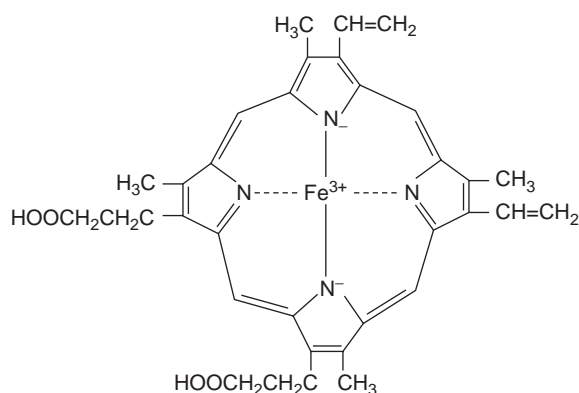
### Drug Metabolism

Drugs are metabolized if the body is unable to excrete them unchanged. The main site of metabolism is the liver. The principal reason for metabolism is to increase the water solubility and hence the excreatability of the compound. It is important to note that many of the enzymes involved in drug metabolism have evolved for the metabolism of naturally occurring compounds and only attempt to metabolize drugs because they resemble them in some way. The route of metabolism of a drug can determine whether it shows any pharmacological or toxicological activity. Usually, the metabolism of a drug reduces its toxicity but certain metabolites (usually short-lived, reactive intermediates) are able to produce toxic effects. Studying metabolism allows species comparison of metabolite patterns to validate toxicity studies, to help understand PK/PD, and to allow extrapolation to man. The PK of metabolites along with that of the parent compound may help to give a better overall understanding of the observed effect (desired or undesired).

The main routes of metabolism are the chemical reactions of oxidation, reduction, hydrolysis, and conjugation. These can be divided into two phases: phase I (or functionalization reactions) and phase II (or conjugative reactions).

**Phase I metabolism** Phase I reactions (mainly oxidation, reduction, and hydrolysis) act as a preparation of the drug for the phase II reactions, i.e., a chemically reactive group is produced or uncovered on which the phase II reactions can occur, e.g.,  $-\text{OH}$ ,  $-\text{NH}_2$ ,  $-\text{SH}$ ,  $-\text{COOH}$ . Most toxic metabolites are produced by phase I reactions. The P-450 isoenzymes (CYP enzymes), known collectively as the mixed function oxidase system, are found in the endoplasmic reticulum of many cells (notably those of liver, kidney, lung, and intestine) and perform many of these different functionalization reactions. The system requires the presence of molecular oxygen and co-factor nicotinamide adenine dinucleotide phosphate (NADPH) as well as cytochrome P450, NADPH-cytochrome P450 reductase, and lipid.

Cytochrome P450 is a heme-containing protein embedded in the endoplasmic reticulum, which exists in multiple forms and has a molecular weight of 45 000–55 000. The heme (Figure 5) is noncovalently bound to the apoprotein and serves as both the oxygen and substrate-binding site.



**Figure 5** Ferroporphyrin IX heme prosthetic group.

The oxidation reaction catalyzed by CYPs conforms to the stoichiometry shown in eqn [7]:



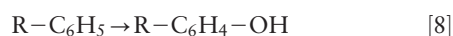
where RH represents an oxidizable drug substrate and ROH is the hydroxylated metabolite. The source of the reducing equivalent ( $2\text{H}^+$ ,  $2\text{e}^-$ ) is either NADPH or NADH mediated by a flavoprotein (FAD).

Three CYP families, CYP1, CYP2, and CYP3, account for ~70% of human hepatic microsome CYPs with CYP3 accounting for ~30%. These CYPs are the major ones responsible for the metabolism of commonly prescribed drugs. Specifically, the most important isoforms are CYP1A1, CYP2C9, CYP2C19, CYP2D6, and CYP3A4. CYP3A4 has the widest substrate specificity and is estimated to be involved in the metabolism of ~50% of drugs used in humans.

Oxidations can occur at both aliphatic and aromatic carbon atoms. This leads to increased water solubility because of an increase in hydrogen bonding and ionizability. Oxidations can also occur at nitrogen and sulfur atoms with the formation of nitrogen and sulfur oxides. Another common phase I reaction is dealkylation; the removal of an alkyl group from either an oxygen producing an alcohol, or a nitrogen atom producing an amine. Reduction reactions are generally inhibited by oxygen and occur mainly under anaerobic conditions, e.g., in the gastrointestinal tract. Hydrolysis is catalyzed chiefly by esterases and peptides. Some of the most common phase I oxidation, reduction, and hydrolysis reactions are summarized in eqns [8]–[20].

### Oxidations

#### *Aromatic hydroxylation*





For example, phenobarbital → *p*-hydroxyphenobarbital.

*Aliphatic hydroxylation*



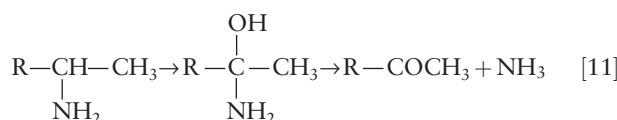
For example, pentobarbital → pentobarbital alcohol.

*Dealkylation*



For example, imipramine → desmethylinipramine (*N*-dealkylation) and codeine → morphine (*O*-dealkylation).

*Oxidative deamination*



For example, amphetamine → phenylacetone + NH<sub>3</sub>.

*N-oxidation*



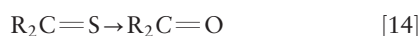
For example, chlorpromazine → chlorpromazine N-oxide.

*S-oxidation*



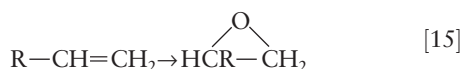
For example, chlorpromazine → chlorpromazine S-oxide.

*Desulfuration*



For example, thiopental → pentobarbital.

*Epoxidation*



For example, benzopyrene → benzopyrene epoxide.

*Alcohol oxidation*



For example, ethanol → ethanal.

**Reductions**

*Azo reduction*



For example, prontosil → sulfanilamide + triamino-benzol.

*Nitro reduction*



For example, *p*-nitrobenzoic acid → *p*-aminobenzoic acid.

**Hydrolysis**

*Ester hydrolysis*



For example, procaine → *p*-aminobenzoic acid + diethylaminoethanol and aspirin → salicylic acid + acetic acid.

*Amide hydrolysis*



For example, phenacetin → phenetidin + acetic acid.

**Phase II metabolism** Phase II conjugation reactions give rise to the bulk of the inactive, excreted products of a drug and involve the addition of an endogenous molecule to a compound. Generally, conjugation results in a large increase in water solubility and molecular weight, which encourages biliary excretion. Importantly, conjugation also leads to the inactivation of many reactive phase I products. The most common enzymes and functional groups involved are listed in Table 2.

Glucuronidation occurs at oxygen and nitrogen atoms and is one of the most common phase II reactions. It can occur with alcohols, phenols, hydroxylamines, carboxylic acids, amines, sulfonamides, and thiols and utilizes the energy-rich co-factor uridine diphosphoglucuronic acid (UDPGA).

Sulfation occurs at oxygen atoms. It can occur with phenols, alcohols, amines, and thiols and utilizes the energy-rich co-factor 3'-phosphoadenosine-5'-phosphosulfate (PAPS). Competition with glucuronidation for substrates can occur and in general sulfate conjugation predominates at low substrate concentrations and glucuronide conjugation at high concentrations, due to the kinetics of the two reactions and the more limited supply of PAPS compared to UDPGA.

Acetylation results in the addition of a CH<sub>3</sub>CO group to nitrogen atoms to give an amide and although this results in no great change in water solubility the nitrogen is 'inactivated'. It utilizes the co-factor acetyl (CoA).

Glutathione conjugation is recognized as a protective compound useful for the removal of potentially toxic electrophilic compounds. It can occur

with epoxides, haloalkanes, nitroalkanes, alkanes, and aromatic halo- and nitro-compounds. The levels of glutathione in the body are finite and depletion following an overdose with a compound that undergoes this form of metabolism can result in toxicity due to the presence of large amounts of unconjugated reactive electrophiles, e.g., paracetamol.

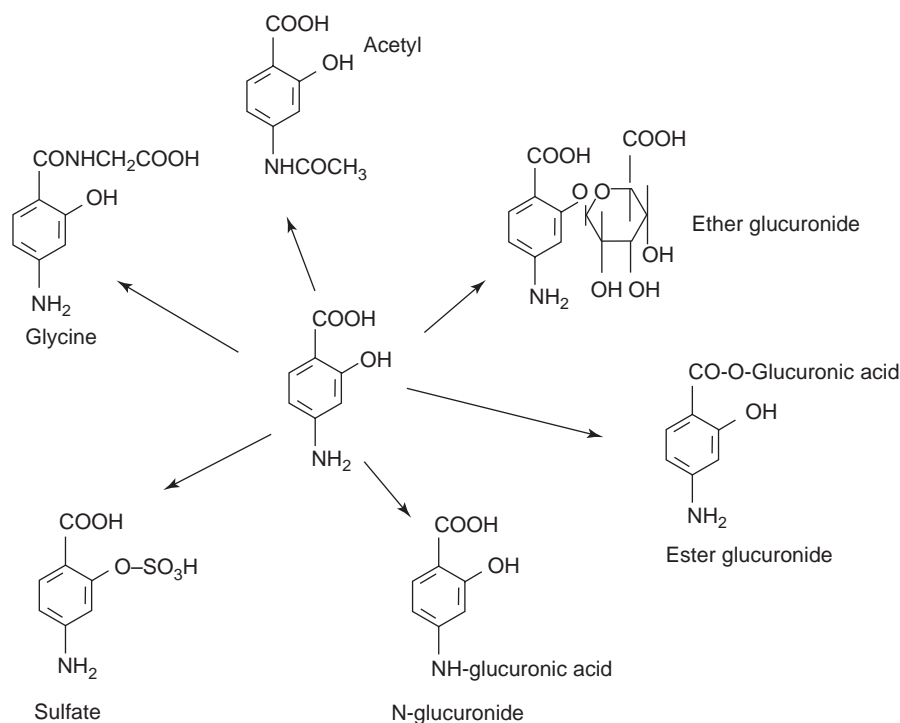
**Table 2** Principal conjugation reactions

Reaction	Enzyme	Functional group
Glucuronidation	UDP – glucuronyltransferase	–OH –COOH –NH <sub>2</sub> –SH
Glycosidation	UDP – glycosyltransferase	–OH –COOH –SH
Sulfation	Sulfotransferase	–NH <sub>2</sub> –SO <sub>2</sub> NH <sub>2</sub> –OH
Methylation	Methyltransferase	–OH
Acetylation	Acetyltransferase	–NH <sub>2</sub> –SO <sub>2</sub> NH <sub>2</sub> –OH
Glutathione conjugation	Glutathione-s-transferase	Epoxide Organic halide

A good example of a compound undergoing a number of key conjugation reactions is *p*-aminosalicylic acid (**Figure 6**).

All tissues have some metabolic capacity, but greatest activity is seen in the liver, gastrointestinal lumen plus microflora, kidney, and blood. Many factors can affect drug metabolism including the dose level (exposure) and route, species and genetic differences, age, sex, diet, environment, disease, and the inhibition and induction of enzymes. An important exposure-related effect sometimes seen at high doses is when a metabolizing enzyme becomes saturated causing a change in drug:metabolite ratios and an altered pharmacological and/or toxicological response. The route of dose can also change the levels of total circulating drug, in particular, dosing by routes other than the oral route (e.g., intravenous, subcutaneous, inhaled) will result in first-pass metabolism being avoided.

Differences in the metabolic route between species are common, e.g., the rat mainly hydroxylates amphetamine leading to conjugated products whereas the rabbit and guinea pig (and man) mainly deaminate amphetamine. Acetylation of aromatic amines occurs in humans but not dogs while glucuronic conjugation is very poor in the cat. A good example of a compound that has the same routes of metabolism in different species but different rates is caffeine, as shown in **Table 3**. Species differences can be



**Figure 6** *p*-Aminosalicylic acid metabolism.

**Table 3** Species differences in caffeine metabolism

Parameter	Man	Monkey	Rat	Rabbit
Total metabolism	322	235	160	137
Theobromine	28	13	15	19
Paraxanthine	193	11	20	42
Theophylline	16	190	20	30

Rates expressed as  $\mu$  moles of product per min per mg protein. Reproduced with permission from Berthou *et al.* (1992) *Xenobiotica* 22: 671–680; © Taylor & Francis Ltd.; <http://www.tandf.co.uk/journals>.

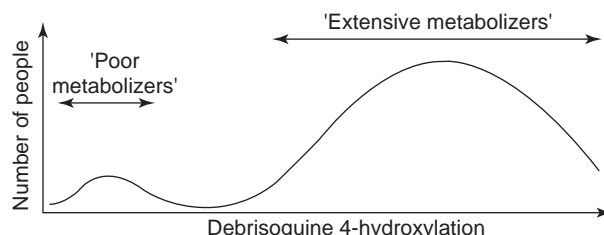
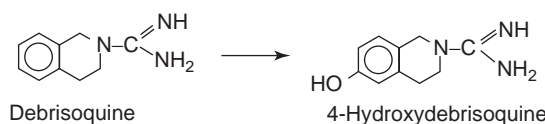
important and usually animals are selected for toxicology studies on the basis of whether their metabolism most closely mimics the expected metabolism in humans.

Genetic differences within species (genetic polymorphism) may result in absences or deficiencies in enzymes, which may predispose individuals to adverse reactions. A genetic deficiency in the P-450 isoenzyme family (lack of CYP2D6) in a proportion of the population is why the antihypertensive debrisoquine is poorly metabolized in certain individuals (Figure 7).

A genetic polymorphism that leads to individuals having differences in their ability to acetylate compounds means that those classed as 'slow-acetylators' can have greater exposures to unconjugated reactive phase I products as has been observed for the anti-tuberculosis drug isoniazid.

The rates and routes of metabolism can vary with age and the lower rates and reduced routes of metabolism usually seen in the young and the old can make them more susceptible to drug action. Differences in hormonal levels, in particular, can lead to differences in metabolism between sexes. Dietary and environmental factors such as the presence of alcohol and cigarette smoke can affect metabolism, as can certain disease conditions, e.g., liver disorders, diabetes. Such conditions can lead to decreased enzyme activity, altered hepatic blood flow, and changes in plasma protein levels which, in turn, can affect circulating 'active' drug levels.

**Drug–drug interactions** Many compounds are able to 'inhibit' metabolizing enzymes, i.e., decrease their activity (cause an increase in steady-state plasma concentrations and a decrease in clearance). Inhibition can take place through the destruction or direct inhibition of the enzyme or the formation of a complex with it. Drug–drug interactions can occur when inhibition of the metabolism of a drug by a co-administered one leads to higher than expected levels of a pharmacologically active moiety. Also, the inhibition of some conjugation reactions can allow



**Figure 7** Distribution and rate of metabolism of debrisoquine in the Caucasian population. (Reproduced with permission from Gibson G and Skett P (1994) *Introduction to Drug Metabolism*, 2nd edn., ch. 4, p. 113; © Nelson Thornes Ltd.)

chemically reactive metabolites to escape detoxification and cause a toxic effect. A well-known example of a drug–drug interaction was the case of the non-sedating antihistamine terfenadine (a substrate for CYP3A4), which when co-administered with the antifungal ketoconazole, a potent inhibitor of CYP3A4, led to higher than expected plasma levels and cardiotoxic side effects. Another example is that of the stomach ulcer drug, cimetidine, which, when co-administered with the anticlotting drug warfarin, inhibited its rate of metabolism and dangerously lengthened blood clotting times.

Many compounds are able to 'induce' metabolizing enzymes, i.e., increase their activity (cause a decrease in steady-state plasma concentrations and an increase in clearance). The induction of metabolizing enzymes results in either an increase in their own metabolism (auto-induction) or that of other drugs (drug–drug interactions). It can reduce the toxicity of the parent molecule but lead to the enhanced formation of a chemically reactive metabolite and thereby overwhelm the detoxification processes. Sometimes, these metabolites can prove to be carcinogenic, e.g., polycyclic aromatic hydrocarbons. However, because enzymes are often induced in animals at higher doses than those used in man, the overall effect may not be clinically significant. On a purely practical level, enzyme induction may mean that it will be difficult to maintain efficacious levels and during regulatory toxicity studies it may not be possible to achieve high enough exposures to see toxicity. A number of CYPs are now usually monitored to assess the inductive capacity of test drugs including: CYP1A1, 2C, 2E, and 3A isozymes. For the reasons stated it is better to avoid developing drugs that are potent enzyme inducers, but, nevertheless, many important marketed

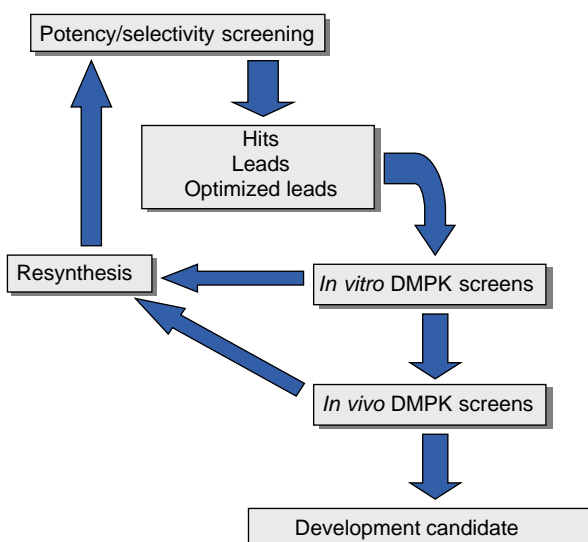
drugs are inducers, e.g., phenytoin, carbamazepine, phenobarbitone.

## Assessment of ADME/PK

ADME/PK has a major role in the research and development of new drugs. The process of drug discovery can be divided into a number of distinct phases: hit screening and lead generation/optimization. The goal of the hit screening phase is to screen compound libraries to find molecules with a specific biological response, usually an *in vitro* potency. At the lead generation/optimization step the main aim is to apply structure activity relationships to improve upon these hits to produce leads that in turn are amenable to further optimization to produce the development candidate. The ADME/PK evaluation of compounds occurs throughout these phases but is particularly useful during lead generation/optimization. A schematic diagram of the discovery process from an ADME/PK perspective is shown in Figure 8.

Many methods are available for investigating key ADME/PK issues. Absorption issues include poor membrane permeability, extensive P-glycoprotein efflux, and poor physicochemical properties. Clearance/metabolism issues include extensive gut metabolism, extensive hepatic metabolism, enzyme induction and inhibition, and transporter-mediated excretion. Distribution issues include extensive protein binding and inadequate tissue penetration. *In silico* (computer-based), *in vitro* (cell-based), and *in vivo* (animal-based) techniques are available for investigating these issues. Structure-based *in silico* models and physicochemical methods allow basic predictions, while *in vitro* models offer a greater level of complexity and most closely mimic the *in vivo* situation. The most promising compounds from an *in vitro* potency and ADME/PK perspective are tested in appropriate *in vivo* pharmacology models. The aim of a typical drug discovery project is to identify a candidate molecule with suitable potency and selectivity against a target, sufficient oral bioavailability, metabolic stability, and duration of action, and minimal developability issues, e.g., potential for drug–drug interactions, ease of synthesis, dose linearity, etc.

Once a compound is selected for development, further studies are performed in animals and man to support the safety assessment and ensure that optimum dosing regimens are selected. Definitive *in vivo* PK and toxicokinetic studies are performed but also more detailed *in vitro* studies, particularly for the assessment of drug–drug interactions. ADME/PK support is continued throughout development (pre-clinical and phases I–IV) and a summary list of the main activities is shown in Table 4.



**Figure 8** Schematic diagram illustrating the typical position of *in vitro* and *in vivo* DMPK screens in the lead generation/optimization phases of the drug discovery process. (Reproduced with permission from Roberts SA (2003): Drug metabolism and pharmacokinetics in drug discovery. *Current Opinion in Drug Discovery and Development* 6(1): 66–80; © Thomson Scientific.)

**Table 4** Main ADME/PK activities performed during drug development

Toxicity testing – monitoring of TK in two nonhuman species
Dose ranging – tolerability over a range of doses
Dose linearity – check of PK proportionality between doses
Definitive PK – single and multiple dose PK
Sex differences – influence of gender on PK
Food interactions – influence of food on PK
Tissue distribution – animals only (usually radiolabeled drug)
Excretion balance – recovery in excreta (usually radiolabeled drug)
Metabolite identification – major routes of metabolism
PK/PD and dose response – determination of clear dose response relationships
Genetic polymorphisms – influence of genetic differences on metabolism
Effect of disease – PK in target population
Dose formulations – PK of final and new formulations
Drug interactions – effect of co-administered compounds on PK

## In Silico Methods

Methods range from those based on molecular structure to those that use complex physiologically-based pharmacokinetic (PBPK) models requiring the input of experimentally determined data. Many physicochemical parameters can be generated automatically from chemical structure including H-bonding, molecular weight, log *P*, p*K*<sub>a</sub>, polar surface area (PSA), and solubility and various software are available to do this. Lipinski's 'rule-of-five' is used by many companies to flag molecules likely to have poor absorption, i.e., compounds that have a molecular weight of

$>500$ , a calculated  $\log P > 5$ , number of H-bond donors  $>5$ , and the number of H-bond acceptors  $>10$ . PSA and rotatable bonds have also proven useful parameters for predicting absorption. A number of more complex models exist, e.g., for plasma protein binding, gastrointestinal absorption, and blood–brain barrier (BBB) penetration, that utilize several physicochemical parameters (calculated and/or experimentally determined). For instance, an algorithm for predicting brain penetration,  $\log BB$  (the ratio of the steady-state concentrations of a compound in the brain and blood), designed using data on a series of brain penetrant compounds, requires PSA and calculated  $\log P$  values as shown in eqn [21]:

$$\log BB = -0.0148\text{PSA} + 0.152C \log P + 0.139 \quad [21]$$

However, one drawback of attempts to model intestinal absorption and BBB penetration is that the effect of transporters is not generally accounted for as the factors that make a compound a substrate or inhibitor are not well understood at present.

Metabolism has proven a difficult event to predict *in silico* but some rule-based expert systems exist and some quantitative structure activity relationship models have been developed. The development of CYP models has been hindered by the lack of three-dimensional crystal structures; however, some homology models have been constructed for key CYP isozymes, e.g., CYP2C9, 2C19, 3A4, and 2D6. In conjunction with homology modeling, site-directed mutagenesis has been shown to help in understanding the relationship between substrate specificity and CYP structures. Important factors for substrate binding and enzyme selectivity include molecular size and shape, number and disposition of hydrogen bond donor/acceptor atoms, and number of aromatic rings in the molecule.

### Physicochemical Methods

Measured physicochemical parameters are useful for the more accurate prediction of events such as drug absorption.  $\log P$  is typically obtained by using potentiometric and spectrophotometric methods to measure the partitioning of a drug between an octanol and a water phase. Solubility is usually measured directly by monitoring achievable drug concentrations in test solutions. For discovery purposes, it is common to measure ‘kinetic’ solubility in which a dimethylsulfoxide solution of the compound is added to aqueous buffer and turbidimetry, nephelometry, and direct ultraviolet methods are used for analysis. However, drug dissolution profiles over time or as a function of pH are generally a better indicator of drug absorption than just inherent

solubility. In addition, a number of artificial permeability assays now exist that attempt to mimic the lipophilic nature of cell membranes and give an indication of absorption potential.

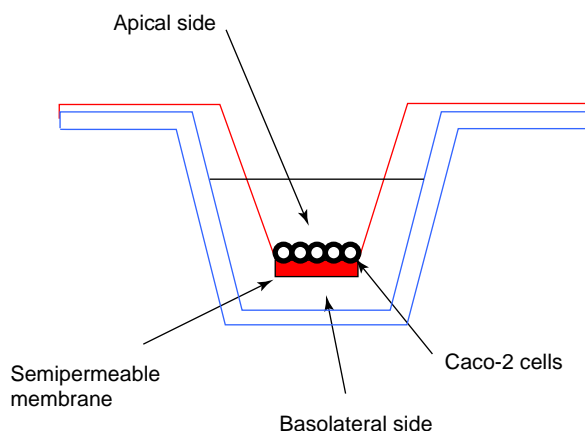
### In Vitro Methods

The most commonly used cell culture model today is the 21-day Caco-2 human colonic cell line derived from a human adenocarcinoma. They can be cultured in special transwell cell culture plates (Figure 9) that enable the investigation of passive diffusion (apical to basolateral side) and active transport (basolateral to apical side). Although the system has its limitations, for many compounds it can give a good indication of likely *in vivo* absorption. An alternative cell culture, which has been shown to correlate well with absorption *in vivo* and permeability in Caco-2 cultures, is the 3-day Madin-Darby canine kidney cell line. Also, the expression of transporter proteins in cell cultures has led to new screens being established for identifying transporter substrates.

A variety of techniques are available for protein binding measurement including equilibrium dialysis, ultrafiltration, spectrofluorometric, and microdialysis methods.

Various cell culture models are available for assessing BBB penetration and although they can give some indication of potential brain penetration, they generally do not form sufficiently tight cell junctions.

A number of *in vitro* hepatic systems are available for assessing metabolism including liver slices, isolated hepatocytes, subcellular fractions (e.g., S9 and microsomes), and purified enzymes. Predictions of *in vivo* clearance can be made using *in vitro* intrinsic clearance ( $Cl_{int}$ ) values derived from liver microsomes or hepatocytes. In addition to methods for assessing enzymic clearance, some exist for investigating drug efflux/uptake in the liver, e.g., isolated



**Figure 9** Schematic diagram of Caco-2 Transwell.



bile cannicular membranes and sandwich-cultured hepatocytes.

Mass spectrometry (MS) is now employed routinely for metabolite elucidation and in particular methods using quadrupole ion traps and time-of-flight liquid chromatography–MS (LC–MS) have proven very useful. Nuclear magnetic resonance (NMR) techniques are also being applied increasingly for more detailed structure elucidation. To aid in structure design, it is also useful to identify the P450 isoforms responsible for metabolism; cDNA-expressed enzymes and selective inhibitors can be used as well as isoform specific anti-P450 antibodies.

CYP inhibition assays include ones that utilize liver microsomes, isolated/cultured hepatocytes, and human cDNA-expressed enzymes. Typically, LC–MS/MS quantification is used for the microsome and hepatocyte methods while fluorimetric assays are used for the human cDNA-expressed enzymes. Typically, a weak inhibitor is defined as having a  $k_i > 20 \mu\text{mol l}^{-1}$ , whereas a potent inhibitor has a  $k_i < 1 \mu\text{mol l}^{-1}$ .

CYP induction assays are available that utilize liver slices, hepatocytes, and cell-based gene reporter constructs. *In vivo*, microsomes can be prepared from the livers of repeat-dosed test animals to examine if any key CYPs have been up-regulated. In discovery, the models utilizing hepatocyte cultures and cell-based gene reporter constructs have proven most useful. A number of analytical methodologies are available for supporting these various approaches including the LC–MS monitoring of probe compounds, mRNA analysis, and ribonuclease protection assays.

In addition to cell-based models, tissue-based models such as the Ussing chamber technique, the everted gut sac approach, and perfused isolated intestinal segments are also used, but only when it is important to understand the absorption processes in more detail. Unlike Caco-2, tissue-based models have the correct physiological levels of transporters and the presence of an apical mucus layer. Also, *in situ* and isolated organ perfusion methods exist for the gut, liver, lungs, kidneys, and brain and can provide data not directly obtainable *in vitro*. The isolated perfused liver is particularly useful since it allows an assessment of first-pass hepatic clearance, the quantitative distribution of metabolites in liver, blood, and bile, the effects of binding to plasma proteins and intracellular sites, and cellular uptake processes.

### In Vivo Methods

Key PK parameters such as bioavailability, plasma elimination half-life, clearance, and volume of

distribution can be generated from PK profiles following oral and intravenous dosing. The efficiency of PK data generation has improved in recent years with the introduction of more sensitive higher throughput methods of analysis. The use of venous cannulated animals and multiple compound (cassette) dosing have contributed to the greater use of *in vivo* PK in drug discovery. Additional information can also be obtained in animals from biliary excretion studies through the cannulation of the common bile duct or gall bladder. Also, transgenic animals lacking or possessing particular gene sequences responsible for important drug metabolizing enzymes and transporters can be used to investigate specific issues. *In vivo* studies provide definitive PK data during development, aid in the design of *in vivo* pharmacology and toxicology studies, and help confirm (validate) the predictivity of *in vitro* methodologies and *in silico* models throughout the drug discovery process.

Methods available for measuring the concentrations of drugs in tissues and body fluids include radioimmunoassay, LC with ultraviolet detection (LC–UV), LC with radioactivity detection, and LC–MS. LC–MS has proven the most attractive method with its good sensitivity and specificity for a wide range of compounds as well as having a minimal sample handling requirement. In addition, MS and NMR techniques are used extensively for metabolite elucidation.

Techniques for quantitative ADME typically employ radiolabeled drugs to look at tissue distribution either in homogenized samples or by whole body autoradiography. Whole body autoradiography involving the location and quantification of radiolabeled material in thin frozen sections of tissue can be particularly useful. Plasma protein binding and metabolism studies also benefit from the use of radiolabeled material.

**See also:** **Drug Metabolism:** Metabolite Isolation and Identification; Isotope Studies. **Pharmacokinetics:** Absorption, Distribution, and Elimination; Pharmacodynamics.

### Further Reading

- Benet LZ, Massoud N, and Gambertoglio JG (eds.) (1984) *Pharmacokinetic Basis for Drug Treatment*. New York: Raven Press.
- Borchardt RT, Smith PL, and Wilson G (eds.) (1996) *Models for Assessing Drug Absorption and Metabolism. Pharmaceutical Biotechnology*, vol. 8. New York: Plenum.
- Bourne DWA (2002) Pharmacokinetics. In: Banker GS and Rhodes CT (eds.) *Modern Pharmaceutics*, 4th edn., pp. 67–91. New York: Dekker.
- Gibaldi M and Perrier D (1982) *Pharmacokinetics*, 2nd edn. New York: Dekker.



- Gibson GG and Skett P (2001) *Introduction to Drug Metabolism*, 3rd edn. London: Nelson Thornes.
- Jang RJ, Harris RZ, and Lau DT (2001) Pharmacokinetics and its role in small molecule drug discovery research. *Medicinal Research Reviews* 21: 382–396.
- Poon GK (1997) Drug metabolism and pharmacokinetics. In: Cole RB (ed.) *Electrospray Ionisation Mass Spectrometry*, pp. 499–525. New York: Wiley.
- Ritschel WA and Kearns GL (1992) *Handbook of Basic Pharmacokinetics – Including Clinical Applications*, 5th edn. Washington: American Pharmaceutical Association.
- Rowland M and Tozer TN (1995) *Clinical Pharmacokinetics: Concepts and Applications*, 3rd edn. London: Williams and Wilkins.
- Smith DA and van de Waterbeemd H (2001) *Pharmacokinetics and Metabolism in Drug Design*. Chichester: Wiley-VCH.
- Testa B, van de Waterbeemd H, Folkers G, and Guy R (eds.) (2001) *Pharmacokinetic Optimization in Drug Research*. Cambridge: Wiley-VCH.
- Yacobi A, Skelly JP, Shah VP, and Benet LZ (eds.) (1993) *Integration of Pharmacokinetics, Pharmacodynamics, and Toxicokinetics in Rational Drug Development*. New York: Plenum.
- Yan Z and Caldwell GW (2001) Metabolism profiling, and cytochrome P450 inhibition and induction in drug discovery. *Current Topics in Medicinal Chemistry* 1: 403–425.

## Metabolite Isolation and Identification

I P Nnane and X Tao, Temple University, Philadelphia, PA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The study of the metabolism of new chemical entities is very important in selecting drug candidates, and increasingly metabolite isolation and identification studies are performed in the early phase of drug discovery. Drug metabolism studies provide vital information to facilitate the understanding of the pharmacological and toxicological action of drugs and satisfy regulatory requirements. The most common types of drug metabolism assays involve isolation and identification of metabolites generated *in vitro* and *in vivo*. The advent of several new technologies has facilitated the isolation and identification of metabolites, which are often formed in very small quantities in drug metabolism models. For example, the increased use of automation in bioanalysis and the availability of robust mass spectrometers have increased the rate at which data can be generated for elucidation of metabolic pathways. In this article, the current approaches and emerging technologies used for metabolite isolation and identification from various biological matrices are discussed.

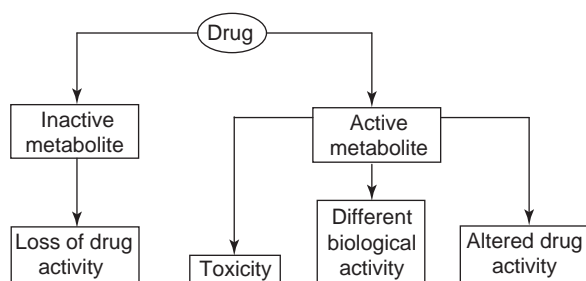
### The Importance of Drug Metabolism Studies

The fate of a drug after administration in animals and humans involves absorption, distribution, metabolism, and excretion (ADME). The body recognizes a

drug as a foreign compound and in most cases the drug is converted to a metabolite through multiple enzymatic pathways. Drug metabolism controls the levels of drugs in the body and may detoxify or activate a drug. In a nutshell, the goals of metabolite isolation and identification include (1) determination of routes of elimination, (2) estimation of rates of metabolic reactions, (3) assessment of metabolic stability, (4) testing of metabolizing enzyme inhibition, (5) structural characterization of formed metabolites, and (6) evaluation of the biological activity of metabolites. Morphine, for example, is converted to morphine-6-glucuronide, which is partially responsible for its pharmacological activity in humans. The various biological outcomes of drug metabolism are shown in **Figure 1**. In recent years, regulatory authorities have also demanded information on metabolite identity and biological activity before new drug applications are approved. In fact, the successful drug candidates must have acceptable ADME, pharmacological, and toxicological profiles. Thus, the role of metabolite isolation and identification studies in selecting better lead compounds with acceptable ADME properties *in vivo* is pivotal in drug discovery and regulatory filing for new drugs.

### Models for Drug Metabolism Studies

Both *in vitro* and *in vivo* models are used to optimize compounds for favorable metabolic properties during drug discovery and development. After administration of a drug to animals and humans, the drug may be transferred from the blood into the liver, the major site of drug metabolism in the body, where it is metabolized or excreted into the biliary system.



**Figure 1** Schematic representation of possible outcomes of drug metabolism.

In many cases, drug metabolism occurs in two stages: Phase I biotransformations, which introduce or unmask functional group(s) in a drug and Phase II biotransformations, which attach endogenous molecules to the parent drug or formed metabolites. Phase I reactions are mainly catalyzed by a family of enzymes known as the cytochrome P450s (CYPs). The common Phase I metabolic reactions include oxidation, reduction, hydrolysis, hydration, dealkylation, and isomerization. Generally, these preparatory reactions make a drug more water-soluble. The metabolites are commonly removed from the body through an excretory route such as the kidneys although other routes may be utilized. Phase II reactions, such as glucuronidation, sulfation, acetylation, amino acid conjugation, and glutathione conjugation, generate highly polar derivatives thereby facilitating drug elimination. Other factors often affect the *in vivo* metabolism of drugs. Hepatic blood flow, for instance, limits the metabolism of some rapidly metabolized compounds *in vivo*.

After administration of a drug to selected animal species, blood, urine, feces, and occasionally bile are collected at predetermined times for metabolite isolation and identification. These whole animal studies are expensive and the metabolic information generated may not be applicable to humans. In addition, the numbers of new drug candidates and the cost of screening them have increased significantly as a result of combinatorial chemistry in recent years. Consequently, *in vitro* drug metabolism methods are increasingly utilized in parallel with *in vivo* models for metabolic stability, metabolite generation, and metabolite profiling because they provide efficient and low-cost approaches in the screening of new drug candidates. The commonly used *in vitro* drug metabolism models include subcellular fractions, hepatocytes, liver slices, and cDNA expressed enzymes. As a result of increased availability and ease of controlling incubation conditions, *in vitro* models are very useful in elucidation of major metabolic pathways, characterization of metabolites, and large-scale synthesis of special metabolites. A major drawback of

*in vitro* models is that they are simplistic and may not accurately predict the metabolic fate of a drug in animals or humans. In addition, *in vitro* enzymatic activities tend to decrease during incubations. Thus, the choice of which *in vitro* model to use depends on several factors including availability and suitability of the model to predict the intrinsic clearance of drugs in humans.

### Subcellular Fractions

Hepatic microsomes remain the simplest and most popular subcellular fractions used to perform metabolic stability studies. Liver microsomal preparations contain numerous CYPs and the flavin-containing monooxygenases that reside in the smooth endoplasmic reticulum. CYP-mediated metabolism of drugs is responsible for many drug interactions. For example, the metabolism of tamoxifen and CYP inhibition by tamoxifen and its major metabolites has been examined in human liver microsomes. Four major metabolites are formed from CYP-mediated biotransformation of tamoxifen in human liver microsomes. Additionally, tamoxifen and its metabolites, *N*-desmethyltamoxifen, 4-hydroxytamoxifen, and 3-hydroxytamoxifen, reversibly inhibited midazolam 1'-hydroxylation, diltiazem *N*-demethylation, and testosterone 6 $\beta$ -hydroxylation with  $K_i$  ranging from 3 to 37  $\mu\text{mol l}^{-1}$  in human liver microsomes. Although liver microsomes are widely used for metabolite profiling, the information obtained from this model is not as complete as that obtained from cellular systems since cellular integrity is lost and some enzymes may require the incorporation of co-factors for optimal activity. Other subcellular fractions such as liver homogenates, which contain both Phase I and II enzymes, are available for *in vitro* drug metabolism studies.

### Isolated Hepatocytes and Liver Slices

The liver is the most commonly used tissue for drug metabolism experiments although other organs, such as the brain, heart, and kidney have also been utilized. Hepatocytes have increasingly become the preferred whole cell *in vitro* system for drug metabolism studies. They provide an environment where enzyme and co-factors are present in normal physiological concentrations and cellular integrity is maintained. Primary hepatocytes are isolated from fresh liver tissue and often possess a broad spectrum of enzyme activities, including reticular, cytosolic, and mitochondrial enzymes. A major disadvantage of primary hepatocytes is that the enzyme activity tends to decrease rapidly with time. However, with the introduction of hepatocyte technologies such as

sandwich culture, gel-immobilized, and cryopreserved hepatocytes, metabolic data are increasingly being generated with long-term hepatocyte cultures.

Liver slices retain intact cell-cell junctions, cellular integrity, and the complete array of drug metabolizing enzymes compared to other *in vitro* systems. Thus, liver slices are attracting increasing attention since complete information on drug metabolism reaction can be obtained in this system.

### Recombinant CYPs

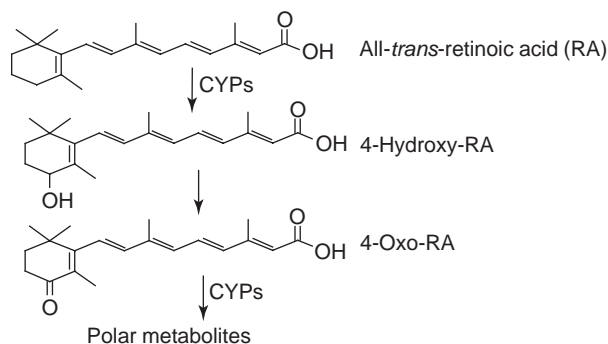
Isolated enzymes responsible for various metabolic pathways are now commercially available due to recent advances in molecular biology. Animal and human CYPs have been expressed from specific cDNA in yeast, bacterial, and mammalian cell lines. These recombinant enzymes are very useful tools for metabolite generation and isolation. The expression of CYPs has been extensively studied since they metabolize a wide variety of drugs. Recombinant human CYPs provide a consistent source of single CYP isoforms, allow incubation conditions to be readily controlled, generally exhibit similar kinetic profiles as demonstrated in human liver microsomes, and permit identification of specific enzymes involved in biotransformations. The metabolism of all-*trans*-retinoic acid (RA) to its 4-hydroxy derivative, for example, has been examined in cDNA-expression systems (Figure 2). These studies have identified CYP26A as the principal isoform that mediates the 4-hydroxylation of RA in animals. It is acknowledged that systems containing cDNA-expressed enzymes are artificial because the enzyme is not present in its native environment, is often over-expressed, and the activity of the expressed enzyme may differ from that of the native one. In order to overcome this limitation, well-characterized cDNA-expressed enzymes must be used and sufficient attempt made to

relate *in vitro* data to the level and activity of the enzyme in the native tissue.

### Metabolite Isolation

The sample types encountered in drug metabolism studies consist of enzymes, other endogenous compounds, and varying concentrations of target drugs and their metabolites. In many instances, the analysis of metabolites in these biological matrices requires a sample preparation step before the samples are introduced into an analytical instrument such as high-performance liquid chromatography (HPLC). The primary purpose of sample preparation protocols is sample purification, isolation, and concentration of the analytes prior to chromatographic analysis and structure elucidation of the drugs and their metabolites.

The overall efficiency of the analytical procedure and the accuracy of the results obtained are influenced by the effectiveness of the sample preparation method. Several technologies are used for metabolite isolation from various biological matrices during *in vitro* and *in vivo* drug metabolism studies. Appropriate isolation procedures account for the stability of the metabolites and allow for efficient separation of the metabolites from the matrix resulting in relatively clean fractions that make metabolite identification easier. The common techniques used for metabolite separation and isolation such as protein precipitation, solvent extraction, and solid-phase extraction (SPE) have been discussed at length previously. Isolation of metabolites that are bound to proteins may also be accomplished, although the commonly used procedures may change the structure of the metabolites. Protein-precipitating agents (e.g., trichloroacetic acid, acetonitrile, or methanol) are often used as the first step for denaturing the proteins that are present in the biological samples in order to get rid of interfering macromolecules and in the process liberate the bound drug and metabolites. Additional sample cleanup by solvent or SPE is usually conducted to isolate the unbound metabolites before analysis. The common methods for sample preparation are increasingly being transferred to automated platforms in order to support large-scale metabolite isolation studies. A significant degree of changes toward miniaturization and automation have been made in SPE in order to minimize the number of steps involved in sample preparation procedures, reduce potential sources of error and the time required to conduct routine assays, and generally to make the processes less labor-intensive. The use of multiwell plates and column-switching methodologies for SPE are making a major impact in sample preparation during the drug discovery phase.



**Figure 2** Metabolic pathway of all-*trans*-retinoic acid (RA) to its hydroxy- and oxo-derivatives by cytochrome P450 (CYPs).

SPE is carried out with many sorbent phases (e.g., C8, C18, silica, CN) in cartridge, disc, and 96-well plate formats for online sample preparation. For example, an online extraction column containing C18 has reportedly been used to study the CYP3A4-mediated hydroxylation of testosterone in liver microsomes. The separation of testosterone and its metabolites from microsomal proteins was based on reversed-phase partitioning and size exclusion; the C18 groups retained the drug and metabolite molecules in the pores in the inner surface of the column while the extraction mobile phase washed away the protein molecules. After the proteins had been washed away, the valve was switched to allow the elution mobile phase to go through the extraction column in order to elute the retained drug and metabolites onto the analytical column for chromatographic separation and analysis. This online sample preparation approach coupled to automated analysis using an autosampler is flexible to handle different assays.

## Chromatographic Methods for Metabolite Separation and Isolation

Chromatographic techniques still make up the vast majority of methods for separation, isolation, and purification of drugs and their metabolites from biological samples in drug metabolism studies. Although separate chromatographic methods are still required for each drug and its metabolites, significant improvements in the limits of detection of the assays have been realized, especially for hyphenated methods such as HPLC–mass spectrometry (MS) (*vide infra*). In addition, major improvements in software for controlling the analytical instruments and for data acquisition and management have resulted in significant reductions in the time required to conduct bioanalysis. The software is useful for accurate recording of the retention times, intensity of the chromatographic peaks, and for obtaining the spectra for analytes of interest, thereby allowing the bioanalyst sufficient time for data interpretation. These methods will continue to have a major impact in metabolite isolation and identification in the future.

### Metabolite Identification

The elucidation of metabolite structure is usually conducted in several steps that involve isolation of metabolites from appropriate biological matrices, metabolite detection, and quantification followed by structural characterization and comparison to synthesized standards. This approach is based on the

premise that it is possible to predict the common metabolic pathways such as oxidation and glucuronidation of new drugs and therefore a targeted search of potential metabolites is conducted. After metabolite isolation and separation is accomplished, specialized technologies are required for metabolite identification. The standard spectroscopic techniques, such as ultraviolet (UV), MS, infrared spectroscopy, and nuclear magnetic resonance (NMR) may be used offline after metabolite isolation or linked with gas or liquid chromatographic methods for metabolite identification. The use of radiolabeled and stable isotopes has also facilitated metabolite isolation and detection for some drug candidates, which lack strong chromophores or fluorophores in their structure. The isolation and identification of metabolites is tedious, time consuming, and requires careful data interpretation in order to unravel the chemical structures. In order to overcome this drawback and address the increasing number of compounds encountered during metabolic stability studies, automated hyphenated techniques utilizing chromatography with UV, fluorescence, radiometric, NMR, and MS detectors, and improved software for data acquisition are now in routine use. Of all the instruments available for metabolite identification, MS has undergone significant innovations and emerged as an ideal technique for the characterization of structurally diverse metabolites. Our discussion will now focus on innovations in some MS instruments that have impacted metabolite identification in recent years.

## Mass Spectrometry

New developments in MS technologies are useful in distinguishing minor differences in the mass ion spectra for different metabolites. The commonly used ion separation analyzers are quadrupole MS, ion-trap MS, time-of-flight (TOF) MS, and Tandem MS. The common modes of ionization in MS include (1) electron impact ionization (EI), (2) chemical ionization (CI), and (3) atmospheric pressure ionization (API). A detailed discussion of all the available MS instruments and their modes of operation is beyond the scope of this article. Briefly, the quadrupole MS is the most frequently used mass analyzer for small molecules and can scan a mass range of 80–700  $m/z$  in  $\sim 0.5$  s. Additionally, it can detect both positive and negative ions and perform product ion scanning, precursor ion scanning, neutral loss scanning, and selected reaction monitoring for high sensitivity. Ion-trap mass analyzers such as three-dimensional quadrupole ion-trap MS that have the potential to define the sites of metabolism more clearly are also



available. Modern ion-trap MS operates by storing ions in a trap and manipulating the ions using electric fields in a series of carefully timed events. This approach provides some unique capabilities to perform a variety of tandem MS experiments with very high sensitivity and resolution that are helpful in determining the structure of a metabolite.

TOF MS measures the mass-dependent time that it takes for ions of different masses to move from the ion source to the detector. For this to work well, the time at which the ions leave the ion source must be well defined. Therefore, ions are either formed by a pulsed ionization method or various kinds of rapid electric field switching may be used as a 'gate' to release the ions from the ion source in a very short time interval. TOF MS has the capability of acquiring spectra every 10–100  $\mu$ s and can efficiently define chromatographic peaks from very fast separations. Additionally, TOF MS is very sensitive with great mass accuracy.

Recent innovations in coupling MS instruments to chromatographic systems have made it possible to systematically isolate and elucidate the structures of a wide variety of metabolites. MS instruments can be readily coupled to GC; the gaseous mobile phases and volatile analytes make development of the interface between GC and MS relatively easy. GC–MS systems are very versatile in separating mixtures into their individual components, identifying structure, and providing quantitative and qualitative information for drugs and their metabolites. In contrast, the HPLC solvents are not an ideal interface for MS. The compounds in the liquid phase eluting from the HPLC system are at atmospheric pressure, while MS is only open to ions in a gas phase at high vacuum. Therefore, a suitable interface between HPLC and MS must have the capability to remove the solvents and ionize the analytes.

### Atmospheric Pressure Ionization Mass Spectrometry

The advent of API MS has revolutionized the analysis of drugs and metabolites in biological samples. The API source provides an ideal interface to link HPLC to MS utilizing two ionization methods: electrospray ionization (ESI), which ionizes the molecules in the mobile phase, before the ions arrive in the gas phase, and atmospheric pressure chemical ionization (APCI), which ionizes the molecules after they are already in the gas phase. In APCI mode, both solvent and solutes are rapidly evaporated by heat. After evaporation of neutral species, ions are formed in a gas phase at atmospheric pressure. All species in the gas phase undergo

significant collision with surrounding gases to form ions of sample molecules. APCI is a very efficient mode of ionization for compounds with high proton affinity and it is applicable for ionizable and polar molecules with molecular weight range of less than 2000  $m/z$ . The APCI MS using a quadrupole analyzer is a relatively small and low-cost system, offers good reproducibility, and has wide application in hyphenated systems such as HPLC–MS, GC–MS, and MS–MS systems. HPLC–APCI MS is suitable for identification of a wide range of metabolites and assay sensitivity can reach the femtomole levels. Recently, we developed an HPLC–MS for the simultaneous determination of all-*trans*-RA and its polar metabolites, 4-hydroxy-RA and 4-oxo-RA, in human liver microsomes (Figure 2). The method of detection was carried out using positive ion APCI and selected ion monitoring of the ions  $m/z$  299,  $m/z$  301, and  $m/z$  315 for quantitative analysis. The ion  $m/z$  299 represented the loss of the hydroxyl group from 4-hydroxy-RA, while the ions  $m/z$  301 and  $m/z$  315 represented the protonated RA and 4-oxo-RA molecules, respectively. The limit of quantification was determined to be in the picomole range with excellent accuracy and precision.

### Tandem Mass Spectrometry

A combination of different types of MS, with different ionization techniques, is also very valuable in identifying unknown metabolites. Controlled fragmentation, through tandem MS (also known as MS<sup>*n*</sup>, where  $10 \geq n \geq 2$ ) involving multiple stages of analysis with serially linked MS instruments, is useful for structural elucidation of novel drugs and metabolites. Tandem MS covers a variety of scanning methods and usually contains ion isolation and fragmentation stages in the same instrument. In the MS/MS mode, a first mass spectrometer (e.g., quadrupole MS) is tuned to allow only the metabolite ion of interest through based on its  $m/z$  ratio. The isolated ion is then passed into a collision cell where an inert gas such as argon is used to fragment the metabolite and the resulting ions are swept into a second mass spectrometer (e.g., ion-trap MS) for separation and detection. During MS<sup>*n*</sup> analysis, a series of fragment ion masses, which correspond to part of an unknown metabolite, can be searched sequentially, permitting the identification of different metabolites. The tandem MS approach can also be coupled with either GC or HPLC. The selectivity and sensitivity of GC–MS/MS and HPLC–MS/MS are extremely high and the structural information obtained from the technique is extensive. In addition,

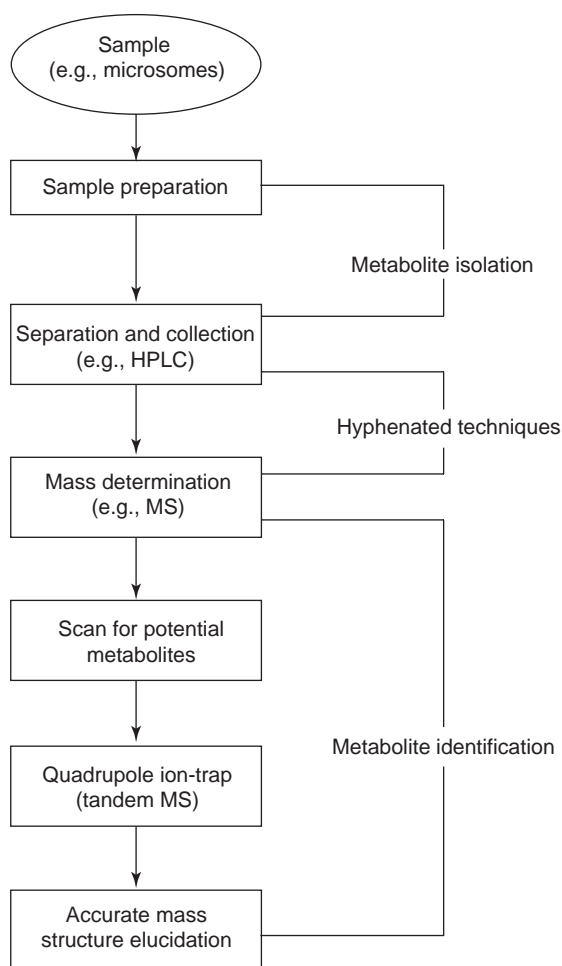
the GC–MS/MS and HPLC–MS/MS technique is robust and may be readily linked to online sample preparation methods for high throughput analysis. Furthermore, combination techniques such as GC–MS/MS and HPLC–MS/MS are the most powerful techniques in drug metabolism studies today by virtue of the fact that they allow the simultaneous separation, purification, quantification, and identification of drug and metabolites with high selectivity and sensitivity as depicted in Figure 3. Notwithstanding, the preferred approach and methods that a bioanalyst chooses depends on availability of a particular analytical instrument, its specificity, accuracy, and the limits of detection of the formed metabolites in a particular matrix.

Interestingly, the increased selectivity of MS instruments has made it possible to circumvent the process of HPLC method development and inject minimally prepared samples directly into the MS

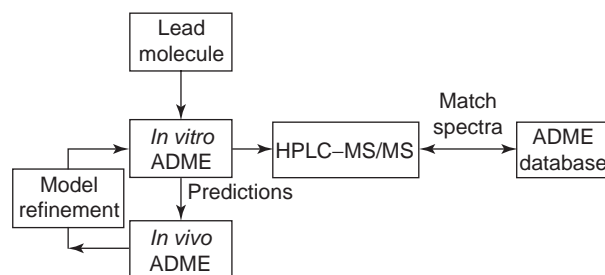
instrument. For example, the high sensitivity of ion-trap MS, by virtue of its capacity to accumulate specific ion populations in the trap, allows direct injection of some biological samples without pretreatment. Furthermore, in flow injection analysis combined with MS (FIA–MS) the sample is injected directly into a stream of mobile phase that is directed into either an ESI or API source, bypassing any HPLC columns. As a result, FIA–MS is the fastest approach for mass detection although it is the least able to determine purity. As improvements in direct-injection MS analysis continue to be made, its contribution to drug metabolism studies will inevitably grow.

## Computational Approaches

Computational approaches are being developed based on past *in vitro* and *in vivo* drug metabolism data in order to increase the efficiency of metabolite profiling and identification even further. In a typical scenario, the ion-trap MS may be used to monitor and characterize the metabolism of lead compounds *in vitro* and *in vivo*. The ions that are detected may be fragmented further by MS<sup>n</sup> methods for additional structural information. The mass spectral data are then transferred to networks for storage and retrieval and metabolite identification accomplished by comparing the fragmentation patterns obtained for specific metabolites to mass spectral databases of potential metabolites as depicted in Figure 4. A promising metabolite database, for providing more realistic predictions of metabolic pathways in various species, is also being developed based on metabolism studies reported in the literature. The importance of computational methods and efficient laboratory information management systems for metabolite identification will become even more apparent as the volume of data generated from drug metabolism studies increases.



**Figure 3** A schematic representation of sample preparation, metabolite isolation, and identification using modern analytical instruments.



**Figure 4** A computational approach for metabolite identification during drug metabolism studies.



See also: **Drug Metabolism: Isotope Studies.**

## Further Reading

- Clarke NJ, Rindgen D, Korfmacher WA, and Cox KA (2001) Systematic LC/MS metabolite identification in drug discovery. *Analytical Chemistry* 73(15): 430A–439A.
- Donato MT and Castell JV (2003) Strategies and molecular probes to investigate the role of cytochrome P450 in drug metabolism: Focus on *in vitro* studies. *Clinical Pharmacokinetics* 42(2): 153–178.
- Ekins S (2003) *In silico* approaches to predicting drug metabolism, toxicology and beyond. *Biochemical Society Transactions* 31(Pt 3): 611–614.
- Ekins S, Ring BJ, Grace J, McRobie-Belle DJ, and Wrighton SA (2000) Present and future *in vitro* approaches for drug metabolism. *Journal of Pharmaceutical Toxicological Methods* 44(1): 313–324.
- Gunaratna C (1998) Biotrap sample extraction column: A new approach for on-line sample preparation in drug metabolism studies. *Current Separations* 17(3): 83–86.
- Guttendorf RJ (1996) The emerging role of A.D.M.E. in optimizing drug discovery and design (<http://www.netsci.org/Science/Special/feature06.html>).
- Kostiainen R, Kotiaho T, Kuuranne T, and Auriola S (2003) Liquid chromatography/atmospheric pressure ionization-mass spectrometry in drug metabolism studies. *Journal of Mass Spectrometry* 38(4): 357–372.
- Nnane IP, Damani LA, and Hutt AJ (1995) Drug metabolism studies: Metabolite isolation and identification. In: Townsend A, Worsfold PJ, Haswell SJ, Macrae R, Werner HW, and Wilson ID (eds.) *Encyclopedia of Analytical Science*, vol. 2, pp. 944–951. London: Academic Press.
- Papac DI and Shahrokh Z (2001) Mass spectrometry innovations in drug discovery and development. *Pharmaceutical Research* 18(2): 131–145.
- Smith DA (1994) Design of drugs through a consideration of drug metabolism and pharmacokinetics. *European Journal of Drug Metabolism and Pharmacokinetics* 3: 193–199.
- Watt AP, Mortishire-Smith RJ, Gerhard U, and Thomas SR (2003) Metabolite identification in drug discovery. *Current Opinion in Drug Discovery & Development* 6(1): 57–65.

## Isotope Studies

**I P Nnane and X Tao**, Temple University, Philadelphia, PA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The evaluation of absorption, distribution, metabolism, and excretion (ADME) of new chemical entities using *in vitro* and *in vivo* models is a critical step in the discovery and development of potential new drugs. An essential part of drug metabolism studies involves isolation of metabolites from various biological matrices and identification using state-of-the-art technologies. In many cases, biotransformation studies involve detection and measurement of very low concentrations of the formed metabolites in biological matrices for structural characterization and elucidation of novel metabolic pathways. Specialized methods including the use of radioactive and stable isotopes play a key role in metabolite isolation, identification, and profiling. Tracer methods, which involve the use of drugs labeled with radioisotopes (e.g.,  $^{14}\text{C}$  or  $^3\text{H}$ ) or stable isotopes (e.g.,  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ), are widely used in drug metabolism studies because the tracers can be recognized and quantified by virtue of their decay profiles or mass differences. The application of radioactive and stable isotopes in

drug metabolism studies has a long history. Although stable isotopes such as deuterium were used in drug metabolism studies as early as the 1930s, they were replaced by radiotracer techniques with the advent of liquid scintillation counting. Resurgence in the application of stable isotopes to drug metabolism studies has occurred during the last several decades due to the increasing consideration for human health and improvements in analytical methods, such as high-performance liquid chromatography (HPLC), chemical reaction interface mass spectrometry (CRIMS), and gas chromatography–mass spectrometry (GC–MS). The use of isotopes in conjunction with chromatographic techniques linked to detectors such as ultraviolet, fluorescence, or mass spectrometry (MS) facilitates real-time monitoring of administered drugs and their metabolites in different parts of the body with high resolution, sensitivity, and specificity. In this article, we highlight how current and emerging analytical instrumentation for isotopic studies are used to obtain new insights into the biotransformation of drugs.

## Radioactive versus Stable Isotopes

In isotope-labeled compounds, one or more of the atoms of the molecules are replaced with a detectable quantity of radioactive or stable isotope(s) of the

elements present in the molecules by chemical, enzymatic, or other methods. An important consideration before using isotope-labeled compounds as tracers for obtaining metabolic and pharmacokinetic data is the choice of the label. Briefly, an ideal tracer should have the same physical, chemical, or biological properties of interest as the natural molecule but should also possess some characteristics that will facilitate its detection in the system where the natural molecule is present. In preparing an isotope tracer, it is also important to substitute one or more of the naturally occurring atoms in a specific position in the natural molecule, which is useful for determining the metabolism of the compound, with an isotope atom such that the label does not affect the ADME of the drug under investigation.

Radiolabeled isotopes are frequently used in drug metabolism studies by virtue of the fact that radioactive species do not occur naturally, can be readily traced, and provide a convenient method of measuring very low concentrations of unknown metabolites in complex biological matrices, with very high sensitivity and specificity. The most frequently used radioisotopes in drug metabolism studies are  $\beta$ -emitting nuclides such as carbon-14 ( $^{14}\text{C}$ ) and tritium ( $^3\text{H}$ ). Although it is easier to insert  $^3\text{H}$  into a drug molecule,  $^{14}\text{C}$  is usually preferred because it provides higher activity when inserted at a suitable site in a drug molecule and may offer far more accurate information on the fate of a drug in the body. Radioactive substrates decay by  $\beta$ -emission and the kinetic energy of the emitted particles poses a serious health problem. Other major limitations of using radioisotopes include (1) the environmental concerns of radioactive waste disposal, (2) the exposure to radiation may be too high to permit the use of radiolabeled isotopes for some drugs with relatively long half-lives, (3) the detection of some radiotracers with very short half-lives (Table 1) may be difficult to accomplish, and (4) radioactive isotope studies are not approved in women and children for obvious health concerns.

Stable isotopes offer considerable promise to overcome some of the drawbacks of radioactive isotopes. In stable isotope-labeled drugs, the naturally abundant isotopes of hydrogen, carbon, and nitrogen ( $^1\text{H}$ ,  $^{12}\text{C}$ , and  $^{14}\text{N}$ ), for example, are replaced with the respective rare stable isotopes of these elements ( $^2\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$ ). Although stable isotopes differ from the naturally abundant elements in mass, they do not decay and therefore are not radioactive. The extra mass found in stable isotopes can be exploited to distinguish between chemically identical molecules by precise measurement of their masses for structure determination. In fact, the addition of stable isotopes

**Table 1** Physical properties of some isotopes commonly used in metabolism studies

Element	Stable isotope		Radioactive isotope	
	Mass	Natural abundance (%)	Mass	Half-life
H	1	99.985	3	12.33 years
	2	0.015		
C	12	98.89	11	20 min
	13	1.11	14	5730 years
N	14	99.63	11	500 years
	15	0.37	13	10 min
I	127	100	121	2.12 h
			123	13.27 h
			125	59.4 days
			126	13.11 days
			128	24.99 min
			129	$1.57 \times 10^7$ years

at known concentrations to biological samples before analysis has enabled positive identification and quantification of several drugs and metabolites. Although stable isotopes differ from the natural isotope by one or two atomic mass units, they rarely generate additional toxicities beyond that of the natural molecule itself. Heavy stable isotopes such as  $^{13}\text{C}$ ,  $^{15}\text{N}$ , or  $^{18}\text{O}$ , for example, which differ marginally in mass from the predominant isotopes, do not show any adverse biological effects even at high enrichments. On the other hand, deuterium ( $^2\text{H}$ ), which differs markedly in mass from the predominant hydrogen isotope, may induce serious side effects at high concentrations in body fluids. It has also been shown that there is approximately a 15% difference in the excretion rates between deuterated and non-deuterated norephedrine in humans. In any case, there are several advantages of using stable isotopes over radioactive ones in drug metabolism studies: (1) use of stable isotopes is feasible in pregnant women and children; (2) stable isotopes allow simultaneous and repeated administration of several different tracers through different routes for evaluation of excretory pathways; (3) the substrate content and the isotopic enrichment could be measured simultaneously; (4) the biological specimen can be stored for long periods for subsequent analysis of the stable isotope; and (5) the isotopic effects associated with stable isotopes are smaller compared with radioactive isotopes.

The disadvantages of the use of stable isotope tracers result, to a large extent, from the fact that they also occur naturally, and therefore only changes in the natural background caused by administration of the tracer can be detected. Hence, the sensitivity

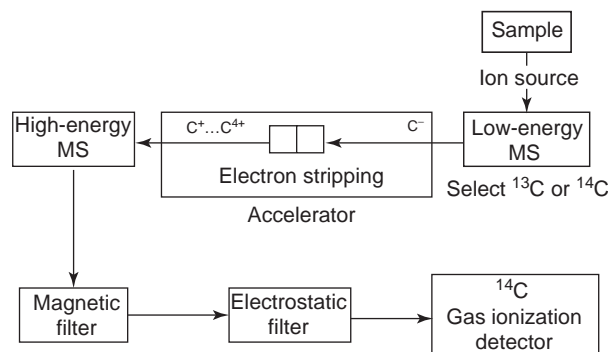
for stable isotopes is usually limited by their relatively high abundance in all natural materials. For example,  $^{13}\text{C}$ -labeled compounds are not distinguishable from the natural carbon compounds when diluted by about five orders of magnitude because  $^{13}\text{C}$  is 1% abundant in nature (Table 1). Since stable isotopes do not emit any radiation, measurements must be made directly on the substances, which in many cases require the isolation of the analyte from complex biological matrices. The measurement of stable isotopes also demands sophisticated analytical techniques and the studies are often quite expensive especially for high enrichment isotopes. Interestingly, new advances in MS have enabled the quantitative detection of stable isotopes with better selectivity, sensitivity, and resolution in drug metabolism studies. It is likely that stable isotopes will continue to play an important role in drug metabolism studies in the foreseeable future as a result of increasing availability of new methods of detection. Ultimately, the choice between radioactive or stable isotopes for drug metabolism studies should take into consideration the cost and safety of the isotopes and the availability of appropriate analytical instruments.

## Measurement of Radioactive Isotopes

In ADME studies, the administered drug and its metabolites must be isolated and characterized in order to obtain a clear picture of the metabolic pathways at play. Frequently, a radiolabeled drug is administered and measurement of radioactivity in the collected biological samples is conducted. Liquid scintillation counting (LSC) or HPLC coupled to a radioactivity detector are commonly used to detect and quantify the radioactivity in the collected tissues and body fluids. The application of LSC and HPLC techniques in drug metabolism studies was discussed in a previous account on this topic. Briefly, the LSC technique is based on decay properties and therefore, is not a particularly efficient way to measure radioisotopes that have long half-lives (Table 1). The counting of the decayed atoms of  $^{14}\text{C}$ , for example, is feasible since its half-life is relatively short (5730 years) although it is time consuming and requires a relatively large sample size. Other radioactive isotopes such as  $^{36}\text{Cl}$  and  $^{129}\text{I}$  have half-lives of up to  $10^7$  years, very slow decay rates, and require a large sample size and incredibly long counting times. Recent innovations in mass spectrometry, for example, accelerator mass spectrometry (AMS), are increasingly being applied to conduct isotope studies because they provide improved efficiency for measurement of radioisotopes.

## Accelerator Mass Spectrometry

AMS has quickly become the preferred method for radioactive isotope measurement because it is much faster and sensitive compared to the traditional methods. The AMS unit consists of several parts, all of which are controlled by a computer (Figure 1). However, a detailed discussion of the architecture and mode of operation of AMS is beyond the scope of this article. Briefly, the AMS procedure usually consists of several steps: a radiolabeled drug is administered to a subject and appropriate samples are collected. For the measurement of  $^{14}\text{C}$ -labeled compounds, for example, the samples are extracted, converted to  $^{14}\text{CO}_2$ , and subsequently to graphite. The AMS method is then used to measure the isotope ratio (e.g.,  $^{14}\text{C}/^{12}\text{C}$ ). AMS is a suitable method of detecting long-lived radioisotopes without regard for their decay products or half-life (Table 1). Since AMS counts individual atoms of radioisotopes, its sensitivity is extremely high and only milligram-sized samples are required. AMS has been reportedly applied to the detection of  $^{14}\text{C}$ -labeled urinary metabolites of the triazine herbicide, atrazine, and the analytical performance of AMS was directly compared to that of LSC. Ten human subjects were given a dermal dose of  $^{14}\text{C}$ -labeled atrazine, urine was collected, and analyzed by both AMS and LSC. AMS provided superior concentration ( $2.2$  versus  $27\text{ fmol ml}^{-1}$ ) and mass ( $5.5$  versus  $54\,000\text{ amol}$ ) detection limits relative to those of LSC in human urine samples. This level of sensitivity will positively impact drug metabolism studies especially during the drug discovery phase. Although AMS requires the use of radioactivity, the radiological dose required is less than  $10\text{ nCi}$  of  $^{14}\text{C}$  labeled drug. This amount of radioactivity is allowed in Phase I ADME studies since there is negligible radiation damage and no serious radioactive waste disposal problems are encountered. A major limitation to the application of



**Figure 1** Schematic representation of the accelerator mass spectrometer.

AMS technology to drug metabolism studies is low instrument availability as a result of its high cost and large size. It should also be pointed out that although several studies with AMS have been carried out with radioisotopes, the potential of this technology for the measurement of stable isotopes has not been fully exploited.

### Online Liquid Chromatography-Accurate Radioisotope Counting

Several analytical techniques such as GC and HPLC have been coupled with AMS and applied efficiently and satisfactorily to drug metabolism studies. However, direct combustion of the GC or HPLC effluents within a microwave-induced plasma chamber prior to AMS analysis is required since no suitable interface is currently available for introducing the HPLC effluent directly into the AMS online. In a recent development, however, an online detection method combining liquid chromatography-accurate radioisotope counting (LC-ARC) with a radioactivity detector and mass spectrometer has been reported. The interface of the LC-ARC system with the mass spectrometer permits effective online radioisotope measurement and acquisition of mass spectrometric data for metabolite isolation and identification. The LC-ARC system has been successfully used for online separation and identification of [ $^3\text{H}$ ]propranolol metabolites with limited sample preparation. Thus, hyphenated techniques involving a combination of LC-ARC with radioactivity detection and MS offer considerable promise as effective tools for metabolite isolation, identification, and profiling.

### Positron Emission Tomography

Whole-body autoradiography (WBA), a resourceful tool for the estimation of the tissue distribution of drugs in animal models of drug metabolism, was described previously. In any event, WBA is increasingly being used for lead optimization and tissue distribution studies in several animal species during early drug discovery. Additionally, lead compounds can be radiolabeled, administered, and whole-body sections imaged quickly to obtain information related to tissue pharmacokinetics, routes of elimination, cytochrome P450-mediated metabolism, and interspecies kinetics. Nowadays, new imaging techniques, such as positron emission tomography (PET), can also be applied to determine the metabolism and pharmacokinetics of labeled drugs *in vivo*. After administration of a labeled drug, PET may be used to measure the distribution of radioactivity in different tissues and obtain information on the metabolic stability of a drug *in vivo* with very high

sensitivity and specificity. PET employs drugs labeled with short-lived positron emitting radionuclides such as  $^{11}\text{C}$ ,  $^{13}\text{N}$ ,  $^{15}\text{O}$ , and  $^{18}\text{F}$ , and by measuring the destruction of radiation using a coincidence technique. A major advantage of PET is that human studies can be performed with the techniques since very low amounts of the radiolabeled drugs have to be administered. In most cases, the radionuclides for PET are usually produced in-house since they are short-lived. Carbon-11 is especially useful because almost every compound in nature contains carbon. Hydrogen, on the other hand, has no radioactive isotope that decays with emission of radiation, which can be detected outside of the human body. In order to circumvent this problem,  $^{18}\text{F}$  is often used as a replacement for a hydrogen atom in a drug, if such a substitution does not alter the properties of the drug in the body. Remarkably, PET is providing new opportunities for lead optimization during drug discovery and for conducting drug metabolism and pharmacokinetic studies *in vivo*.

### Measurement of Stable Isotopes

Mass spectrometric approaches are also very useful for the measurement of stable isotopes in drug metabolism studies. The application of MS to the quantitative measurement of stable isotope has been limited due to the high cost and sophistication of the instruments necessary for stable isotope enrichment studies. Nonetheless, recent improvements in instrument design and performance, as well as computer software for instrument control, data acquisition, and analysis, have increased the sensitivity and reliability of stable isotopic enrichment studies. These new MS instruments, including continuous-flow isotope ratio mass spectrometry (CF-IRMS) and HPLC-chemical reaction interface mass spectrometry (HPLC-CRIMS) are increasingly less expensive, easier to operate, and accessible for mass balance/metabolite identification studies with stable isotopes.

### Continuous Flow-Isotope Ratio Mass Spectrometry

In the CF-IRMS technique, a solid or liquid biological sample is introduced into a high-temperature combustion chamber of an elemental analyzer, where it is converted into a gas. The combustion gases (e.g.,  $^{13}\text{CO}_2$ ,  $\text{N}_2$ ,  $\text{H}_2\text{O}$ ) are carried by a continuous stream of helium, through a cleanup phase, to a gas chromatography-isotope ratio mass spectrometry device. The gases are separated by GC and sent to an ion source where an ion beam is generated for each combustion gas. In the magnetic sector of the mass spectrometer, the beam is split into multiple beams



depending on the molecular weights of the various isotopes contained in the gas. For example, the beam for  $N_2$  is separated into three ion beams and measured at preset masses ( $m/z$  28, 29, and 30) in the collectors. In a biological sample containing a  $^{15}N$  tracer, an increase in masses 29 and 30 above natural abundance will correlate with the concentration of the enriched tracer as atom percent excess. Some efforts have been made to combine HPLC and CF-IRMS for mass balance/metabolite identification studies. However, HPLC effluents must be collected first and then transferred onto CF-IRMS due to the lack of a suitable interface that can accommodate the HPLC effluent directly. In spite of its relatively modest sensitivity ( $\sim 1 \mu g ml^{-1}$ ), the CF-IRMS approach has been successfully applied to analyze  $^{15}N^{13}C_2$ -acetaminophen and to conduct mass balance studies in human samples.

#### HPLC-Chemical Reaction Interface Mass Spectrometry

The CRIMS is increasingly being utilized for the detection and quantification of stable isotope-labeled drugs and metabolites. The universal interface involves a microwave-induced helium plasma that dissociates all analytes to their elemental forms. These species go on to combine with atoms from a reactant gas to form small and stable gaseous molecules that flow into a mass spectrometer. The presence of  $^{13}C$  and  $^{15}N$  is monitored at  $m/z$  45 ( $^{13}CO_2$ ) and  $m/z$  31 ( $^{15}NO$ ), respectively. Since each of these isotopes is naturally abundant, simply detecting these two masses is not sufficient to recognize enriched materials. Nevertheless, if the natural abundance is fixed, the observed signal at the more abundance mass 44 ( $^{12}CO_2$ ) or 30 ( $^{14}NO$ ) is multiplied by the natural abundance of the chemical species being traced, and the product is subtracted from the observed signal for the heavier isotope, the responses from all unenriched material will disappear and only the enriched compound will be traced by the MS. In this approach, the same chemical species are being detected for all analytes; hence, the sensitivity for detection is equal for different analytes. Nowadays, CRIMS can be coupled with HPLC for online analysis of HPLC effluents. The hyphenated instrument, HPLC-CRIMS, constitutes the most powerful separation and detection technique for stable isotope detection in bioanalysis. The HPLC-CRIMS approach has been demonstrated to be effective in the study of the quantitative measurement of a wide range of drug molecules and their metabolites including biologics. Recently, uniformly  $^{13}C,^{15}N$ -labeled rat growth hormone was administered intravenously to rats and blood samples were collected and analyzed using HPLC/CRIMS. The absolute amounts of

$^{13}C,^{15}N$ -labeled rat growth hormone detected ranged from 825 to 66 pmol in a 20  $\mu l$  plasma sample. Interestingly, HPLC-CRIMS has shown excellent agreement to liquid scintillation counting of HPLC fractions.

#### Gas Chromatography-Mass Spectrometry

The application of GC-MS to the study of the metabolism of drugs is a well-developed field. The major advantages of this technique include its ability to rapidly determine many different compounds in complex biological matrices with high sensitivity. A high sample throughput is also realized because extensive sample preparation is often avoided. The present discussion will only highlight the usefulness of GC-MS methods for metabolite profiling of target drugs labeled with stable isotopes. In many cases, the natural and stable-isotope-labeled drug forms differ by one atomic mass unit. These two forms are not usually separated using analytical techniques such as HPLC or GC, since the change in molecular weight by itself does not cause a measurable change in retention times. However, the natural drug form can be distinguished from the stable-isotope-labeled form by GC-MS. In the GC-MS mode, the analytes are not combusted into gases such as  $CO_2$ ,  $N_2$ , and  $H_2O$ , which makes it more suitable for metabolite identification. Measurements of isotopic enrichment by GC-MS typically acquire data by selective ion monitoring (SIM) rather than by a full-scan acquisition. In GC-MS-SIM, ions of the natural and isotope-labeled fragment with specific  $m/z$  values of interest are monitored. The ratio of the tracer to the natural isotope is determined by correlation of the peak area ratio against those obtained for a set of standards of known isotopic enrichment. This technique minimizes potential problems due to label recycling and allows the use of multiple-labeled compounds simultaneously. However, the limited sensitivity of GC-MS-SIM sometimes requires that relatively large amounts of stable-isotope tracers be used.

Isotope dilution gas chromatography-mass spectrometry (ID/GC-MS) is also widely used for the quantitative determination of drugs and their metabolites due to its resolving power, sensitivity, specificity, and accuracy. The use of ID/GC-MS is based on adding a known amount of an isotope-labeled version of the analyte to the sample as an internal standard, equilibrating the labeled analyte with the endogenous analyte, processing the sample, followed by measurement of the ratio of unlabeled to labeled analyte using GC-MS. Assuming complete equilibration with the labeled analyte after spiking, less than complete recovery of the analyte does not affect the measured concentrations. The GC-MS approach using the stable-isotope dilution method has been

reported for the identification of the metabolites of dehydroepiandrosterone formed in human prostate homogenates. A mixture of natural and  $^2\text{H}$ -labeled dehydroepiandrosterone was incubated with human prostate tissue homogenates in the presence of appropriate co-factors. The metabolites were extracted and analyzed by GC–MS. Several metabolites including androst-5-ene-3 $\beta$ ,17 $\beta$ -diol, androst-4-ene-3,17-dione, testosterone, 5 $\alpha$ -dihydrotestosterone, androsterone, and 7 $\alpha$ -hydroxydehydroepiandrosterone were identified based on their chromatographic behavior and mass spectra. GC–MS analyses using stable isotopes are much more difficult and expensive than the usual HPLC and GC procedures. In spite of these limitations, GC–MS approaches using stable isotopes provide powerful tools for investigating drug metabolism in the body and will continue to do so in the near future.

**See also:** **Drug Metabolism:** Metabolite Isolation and Identification. **Liquid Chromatography:** Isotope Separations. **Mass Spectrometry:** Stable Isotope Ratio.

## Further Reading

- Abramson FP, Teffera Y, Kusmierz J, Steenwyk RC, and Pearson PG (1996) Replacing  $^{14}\text{C}$  with stable isotopes in drug metabolism studies. *Drug Metabolism and Disposition* 24(7): 697–701.
- Baba S, Kuroda Y, and Horie M (1986) Studies on drug metabolism by use of isotopes. *Biomedical & Environmental Mass Spectrometry* 13(3): 137–140.
- Baillie TA and Rettenmeier AM (1986) Recent advances in the use of stable isotopes in drug metabolism research. *Journal of Clinical Pharmacology* 26(6): 481–484.
- Bier DM (1987) The use of stable isotopes in metabolic investigation. *Bailliere's Clinical Endocrinology and Metabolism* 1(4): 817–836.
- Browne TR, Szabo GK, Ajami A, and Wagner D (1993) Performance of human mass balance/metabolite identification studies using stable isotope ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ) labeling and continuous-flow isotope-ratio mass spectrometry as an alternative to radioactive labeling methods. *Journal of Clinical Pharmacology* 133(3): 246–252.
- Gilman SD, Gee SJ, Hammock BD, *et al.* (1998) Analytical performance of accelerator mass spectrometry and liquid scintillation counting for detection of  $^{14}\text{C}$ -labeled atrazine metabolites in human urine. *Analytical Chemistry* 70(16): 3463–3469.
- Goldthwaite CA Jr., Hsieh FY, Womble SW, *et al.* (1996) Liquid chromatography/chemical reaction interface mass spectrometry as an alternative to radioisotopes for quantitative drug metabolism studies. *Analytical chemistry* 68(17): 2996–3001.
- Jackson GS, Weaver C, and Elmore D (2001) Use of accelerator mass spectrometry for studies in nutrition. *Nutrition Research Review* 14: 317–334.
- Nassar AEF, Bjorge SM, and Lee DY (2003) On-line liquid chromatography-accurate radioisotope counting coupled with a radioactivity detector and mass spectrometry for metabolite identification in drug discovery and development. *Analytical Chemistry* 75: 785–790.
- Nnane IP, Hutt AJ, and Damani LA (1995) Drug metabolism studies: Isotope studies and metabolite profiling. In: Townsend A, Worsfold PJ, Haswell SJ, *et al.* (eds.) *Encyclopedia of Analytical Science*, vol. 2, pp. 951–957. London: Academic Press.
- Osborn BL and Abramson FP (1998) Pharmacokinetic and metabolism studies using uniformly stable isotope labeled proteins with HPLC/CRIMS detection. *Biopharmaceuticals and Drug Disposition* 19(7): 439–444.
- Paans AMJ and Vaalburg (2003) Positron emission tomography in drug development and evaluation (<http://www.bentham.org/cpd-sample/paans/paans.htm>).
- Papac DI and Shahrokh Z (2001) Mass spectrometry innovations in drug discovery and development. *Pharmaceutical Research* 18(2): 131–145.
- Solon EG, Balani SK, and Lee FW (2002) Whole-body autoradiography in drug discovery. *Current Drug Metabolism* 3(5): 451–462.
- Wang H, Liu Y, and Sun H (2001) Application of  $^{14}\text{C}$ -AMS for carcinogen, pharmaceutical and nutrient studies. *Journal of Nuclear & Radiochemistry Sciences* 2(1–2): R9–R12.

# DRUG SCREENING IN SPORT

**See FORENSIC SCIENCES:** Drug Screening in Sport

## DRY ASHING

**See SAMPLE DISSOLUTION FOR ELEMENTAL ANALYSIS**



# ELECTROGRAVIMETRY

A R Hillman, University of Leicester, Leicester, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Electrodeposition of metals from solution is a well-established technology. The classical electrogravimetry technique uses the increase in electrode mass associated with exhaustive deposition to provide the quantity of metal species initially present in the solution. Recently, the term has been applied to a technique in which a quartz crystal resonator is used as a gravimetric probe; although not generally coupled with exhaustive deposition, this approach exploits the same processes and has the advantage of *in situ* applicability. In either case, control of conditions can facilitate selective deposition of individual elemental metals from mixtures. The technique can also be used via electrodeposition of compounds of known stoichiometry, notably oxides and halides. This article describes the principles involved, strategies for separation of mixtures, and experimental aspects of the technique.

## Overview

### Definition and Principle

Classical electrogravimetry involves the exhaustive electrolytic deposition of a target solution species onto an electrode as a product of well-defined stoichiometry. The increase in electrode mass then yields the amount of the target species originally present in the analyte. Except where otherwise specified, the text refers to this classical approach.

### Technique Characteristics

Electrogravimetry possesses moderate sensitivity, is a fairly rapid procedure, and has a typical precision of a few parts per thousand for a wide range of metals, present singly or in mixtures. Since the analyte concentration is derived from the quantity measured, via

deposit stoichiometry and relative atomic masses, calibration is not required. The measurand is the mass of deposited material, regardless of the charge required for effecting its deposition. The electron is a reagent that may be used in excess; a 100% current efficiency is not necessary, provided competing reactions do not result in codeposition of other materials or prevent deposition of the required species.

### Historical

The first application of electrogravimetry (described by Luckow and Gibbs in 1864) involved the electrodeposition of copper on platinum at an approximately constant current. The controlled potential method (described by Sand in 1906 and now implemented using a potentiostat) has greater selectivity, vital when determining mixtures of metals. Early applications focused on deposition of elemental metals on Pt cathodes and from 1920 to 1950 procedures were developed for separating and determining numerous metals and alloys. Other chemistries include deposition of lead (as  $\text{PbO}_2$ ) on Pt anodes and of  $\text{Cl}^-$  (as  $\text{AgCl}$ ) on Ag anodes.

### Deposition Strategies

The most common procedure involves the reduction of solution phase metal ions – hydrated or complexed – to elemental metal on the cathode. In the case of anodic deposition of metal as an oxide or halide, the stoichiometry must include any water present.

The simplest experimental configuration employs a two-electrode cell without control of the working electrode potential; this offers no control over the selectivity. If the solution contains more than one depositable species, selectivity–complete deposition of the target metal but no other – is required. This is accomplished in a three-electrode cell via independent control of the working electrode potential, and if necessary augmented by control over the pH and/or complexing agents.

### Relationship to Other Analytical Procedures

Generically, electrodeposition under potential control effects separation of the target species. In

electrogravimetry, subsequent determination is by mass. However, this selectivity can be used prior to determination using any other means. The most obvious example of this, within the same experiment but with the additional constraint of 100% current efficiency, is coulometry. An interesting complement to this is deposition and subsequent amalgamation of metals (Bi, Cd, Cr, Co, Cu, Ga, Ge, Au, In, Ir, Fe, Mg, Mo, Ni, Pd, Pt, Po, Rh, Ru, Ag, Te, Tl, Sn, and Zn) at a mercury cathode, in which the deposited metal may be determined coulometrically or voltammetrically during deposition by anodic stripping voltammetry or by decomposition and solution analysis of the amalgam. Electrodeposition is used for removing other metals from actinides and the separation of U, Np, and Am from higher actinides. Although the standard electrode potentials of the actinides prevent analytically useful electrodeposition, radiochemical samples can be prepared this way.

## Deposition Characteristics

### General Requirements

The primary chemical and physical requirements of the electrodeposited material are purity and adherence, respectively. The ideal deposit is dense, smooth, and stable in the presence of oxygen; this allows it to be washed, dried, and weighed without mechanical loss or reaction with the atmosphere. In the case of elemental metals, 'good' deposits are fine-grained and lustrous. 'Poor' deposits are spongy, powdery or flaky, and less adherent and may contain occlusions of foreign species.

### Deposition of Elemental Metals

A wide range of factors influence the deposit characteristics, and it is commonly necessary to change conditions for successive depositions from mixtures. Typical current densities lie in the range  $10^2$ – $10^4$  A m<sup>-2</sup>. Too high a current density yields a rough, less adherent deposit. If the current exceeds the mass transport limited value for the target ion, hydrogen evolution will occur.

Deposits from complexed metal ions are normally smoother and more adherent than those from uncomplexed metal ions. Complexes that are reduced very easily or with great difficulty tend to give poor deposits; complexes of intermediate stability are optimum. Commonly employed complexing agents are cyanide and ammonia. Large coordinating molecules, such as aromatics, give poor deposits. 'Brighteners' – organic additives used to improve the finish of metal plates – are also used. Some, but not all,

brighteners form metal complexes: for example, glycine, tartaric acid, citric acid, and metaphosphoric acid form complexes with silver and copper and act as brighteners for their electrodeposition, and gelatin improves the smoothness of lead and cadmium deposits. Care must be taken to avoid occlusion of these organics in the metal deposit.

Wherever possible, hydrogen evolution should be avoided: this may be accomplished by control of the pH or cathode potential or by addition of a depolarizer, such as nitrate in the constant current deposition of copper. If the metal ion reduction is accompanied by hydrogen evolution, the metal deposit may be brittle, porous, and nonadherent. The main underlying cause is believed to be the physical effect of bubble evolution rather than the formation and decomposition of metal hydrides.

Stirring the solution has several beneficial effects: it can improve the deposit by replenishing metal ions in the vicinity of the electrode, decrease hydrogen evolution, and increase the electrolysis rate, i.e., decrease analysis time. Temperature effects are less predictable. Temperatures in the range 60–90°C are sometimes used for increasing mass transport and decreasing electrolyte resistance, resulting in denser, more adherent deposits.

### Deposition of Oxides

Anodic deposition of metal oxides or other compounds can result in incorporation of some water. This is accounted for by empirical conversion factors (see Table 2).

## Experimental Aspects

### Solution Composition

Alloys are usually dissolved in acid mixtures; for example hydrochloric acid plus a little nitric acid is used for Cu-, Pb-, Sn-, Zn-, Al-, and Mg-based alloys. Alkaline or acid fusion is used for refractory metals. Complexing agents and then depolarizers are added, and finally the pH is adjusted.

Selective complexation can be used for separating metals of similar deposition potential or for reversing the order of deposition. For example, copper is deposited before cadmium in the presence of cyanide, and after in its absence.

Depolarizers prevent undesirable reactions by providing a facile reaction that prevents the cathode potential from becoming too negative or the anode potential from becoming too positive. Cathodic depolarizers are oxidizing agents that are reduced less readily than the depositing metal ion but more

readily than hydrogen ions and/or other metal ions; an example is nitrate, which is reduced to ammonium. Anodic depolarizers are reducing agents; an example is hydrazine, which is oxidized in an acid medium to nitrogen to prevent chlorine evolution or platinum anode dissolution. Hydroxylamine can act as a cathodic or anodic depolarizer. In determination of lead as the elemental metal, chloride is used to prevent PbO<sub>2</sub> deposition on the anode. When the determination is via PbO<sub>2</sub> deposition on the anode, nitrate is used to prevent elemental lead deposition on a copper-plated cathode.

### Instrumentation and Cell Design

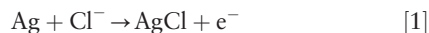
In a classical two-electrode cell the distribution of applied voltage is determined by the reactions at both electrodes, and so there is no independent control of the working electrode potential. Toward the end of electrolysis, the decrease in metal ion concentration allows the cathode potential to become more negative, resulting in codeposition of other metals. For mixtures it is preferable to use a three-electrode cell controlled by a potentiostat such that there is independent control of the working electrode potential with respect to a reference electrode passing no current. The cell resistance is minimized by large area electrodes, high salt concentrations, and (where required) high diaphragm porosity. The electrolysis time is reduced by maximizing the ratio of the electrode area to the cell volume (*V*) and by stirring to minimize the diffusion layer thickness ( $\delta$ ).

### Electrodes

Concentric cylindrical gauze electrodes ensure uniform potential and current distributions. When a reference electrode is employed – historically calomel or mercury/mercury sulfate, more recently (for occupational health reasons) using silver-based couples and occasionally a wire of the metal to be plated – placement of the tip of the salt bridge close to the working electrode minimizes the ohmic potential drop.

The most commonly used working electrode material is platinum, sometimes alloyed with a few per cent of iridium. Two difficulties can arise. First, in chloride media, the Pt anode may dissolve and codeposit with the intended metal on the cathode; platinum dissolution can be suppressed using a depolarizer such as hydrazine sulfate. Second, electrode cleaning is difficult when the deposited metal alloys with platinum. This arises with Bi, Cd, Ga, Hg, Pb, Sn, and Zn but can be prevented by precoating Pt with Cu or Ag.

Other solid electrodes include tantalum and silver (for copper) and nickel (for zinc). Silver anodes prevent evolution of oxygen (which could be reduced at the cathode), and silver cathodes prevent evolution of hydrogen, with associated local pH changes. Addition of chloride allows the anode reaction (eqn. [1])



to proceed at an approximately fixed potential, facilitating controlled potential deposition at the cathode in a two-electrode cell. Silver electrodes cannot be used in complexing media, for example, a cyanide solution or chloride solution sufficiently concentrated to form AgCl<sub>2</sub><sup>−</sup>. One can use a cathode of the metal to be deposited (with a Pt stem for preventing corrosion at the solution/air interface) or indeed any metal for which the deposition potential is more positive than for the metals in solution, provided one uses potential control and is alert to variations in deposition potential with complexing agents.

Before each determination, removal of previous deposits is essential. Anodic dissolution or chemical stripping is used.

A mercury cathode has the advantage of a very large overpotential for hydrogen evolution. The deposited (and amalgamated) metals are determined gravimetrically by distillation of the mercury. In dilute sulfuric acid, the mercury cathode can be used for depositing Fe, Cr, Zn, Cd, Ga, Ge, Cu, Sn, Mo, Pb, Bi, Ag, Au, the Pt metals (except Ru and Os), As, Se, Te, Tl, Re, and Hg. It can only partially deposit Mn. Al, Zr, Ti, P, V, U, Mg, and the alkaline and rare earth metals are left in solution. A hybrid approach is the Paweck electrode, an amalgamated brass gauze used for deposition of Ni, Co, Cd, and Zn.

### Gravimetric Measurement

The requirement is to avoid loss of deposited material. This is accomplished by removal of the electrode from the solution without breaking the circuit, with concomitant washing. With base metals, even prompt washing results in a slight (of the order of a milligram) loss of metal through air oxidation. The electrode is then dried (commonly assisted by washing with a volatile water-miscible organic solvent), heated (110°C for 2–3 min), air-cooled, and weighed.

### Microdeterminations

By using the electrode as the counterpoise of a balance, one can achieve 0.1% accuracy with 1–10 mg metal samples. The difficulties of using a three-electrode cell with small samples can be overcome by using a two-electrode cell with a large counterelectrode whose potential is maintained approximately

constant using a depolarizer. Small-scale separations of Cu from Pb, Sn, Ni, Zn, and Al (for analyses of brass, bronze, and gunmetal), of Bi and Pb, and of metals in solders have been described. Anodic deposition of Pb (as PbO<sub>2</sub>) can also be performed on a small scale. Extension of the conventional technique to submicroquantities presents difficulties associated with submonolayer deposits; the electrochemical quartz crystal microbalance (see below) can accurately quantify such small amounts of material.

## Separation Strategies

Most analytes contain more than one depositable metal, and so a strategy is required for selective or sequential deposition of individual metals; opportunities for this are now described.

### Thermodynamically Based Separation Strategies

Consider separation of metal ions of oxidation state  $z$ , deposited cathodically as the element. The final concentration of metal ions in solution,  $C_{\text{final}}$ , following exhaustive electrolysis of a solution of initial concentration  $C_{\text{initial}}$  at potential  $E_c$  is given by

$$E_c - E_{\text{dep}} = (2.303RT/zF) \log(C_{\text{final}}/C_{\text{initial}}) \quad [2]$$

where  $E_{\text{dep}}$  is the potential for the onset of deposition and the factor  $2.303RT/F$  is equal to 0.059 V at a temperature ( $T$ ) of 25°C. 'Quantitative' deposition requires  $(C_{\text{final}}/C_{\text{initial}}) < 10^{-3}$ , corresponding to  $(E_c - E_{\text{dep}}) < -(0.177/z) \text{ V}$ . Thus, metals with  $E_{\text{dep}}$  values separated by greater than  $0.177/z \text{ V}$  can be separated and determined by alternate deposition and weighing of the electrode. If the required value of  $E_c$  is more negative than  $E_{\text{dep}}$  for another metal, manipulating the potential alone cannot provide separation.

Controlled potential electrogravimetry, using a potentiostat, is commonly used. The applied cell potential,  $E_{\text{appl}}$ , is distributed between the equilibrium anode and cathode potentials ( $E_a$  and  $E_c$ ), the anodic and cathodic overpotentials ( $\eta_a$  and  $\eta_c$ ) and the solution ohmic drop ( $iR$ ):

$$E_{\text{appl}} = E_a + \eta_a - E_c + \eta_c + iR \quad [3]$$

In a three-electrode system, the anode contributions and  $iR$  drop are instrumentally removed. Usually  $\eta_c$  is negligible, and when the current tends to zero (at the end of deposition)  $\eta_c$  tends to zero. Thus, thermodynamic factors provide selectivity via the cathode potential.

When  $E_a$  and  $\eta_a$  remain constant, use of a constant  $E_{\text{appl}}$  approaches the controlled potential method.

This condition is met in a method called internal electrolysis (or spontaneous electrogravimetric analysis), first described by Ullgren in 1868, in which electrolysis occurs by spontaneous discharge of a galvanic cell. To illustrate the principle, consider two half-cells, comprising a zinc rod in a zinc sulfate solution and a copper rod in a copper sulfate solution. At open circuit, 25°C, the reversible cell potential is related to the two standard electrode potentials ( $E^0$ ):

$$E = E_{\text{Zn}}^0 + E_{\text{Cu}}^0 - (0.059/2) \log\{[\text{Zn}^{2+}]/[\text{Cu}^{2+}]\} \quad [4]$$

On connection of the two electrodes, the cell emf drives a short circuit current ( $i$ ) through the cell resistance ( $R$ ), according to

$$E = iR = E_{\text{Zn}}^0 + E_{\text{Cu}}^0 - (0.059/2) \log\{[\text{Zn}^{2+}]_0/[\text{Cu}^{2+}]_0\} \quad [5]$$

where the concentrations now refer to surface values. This is a displacement reaction but with separated half-cells, and so the metal deposits adherently on a stable (not dissolving) surface.

When the deposition potentials are so close that potential control alone is insufficient for effecting separation of the mixtures, it is common to manipulate the deposition potentials using complexing reagents. Complexed metal ions are normally reduced at more negative potentials than are the free ions. Exceptions, for kinetic reasons, are nickel/pyridine and  $\text{SnCl}_4^{2-}$  complexes. Complexation can also improve the deposit quality; for example, silver deposition is excellent from cyanide (as  $\text{Ag}(\text{CN})_2^-$ ), and copper yields much better deposits from pH 4–6 tartrate than from nitric or hydrochloric acids. The stability of many complexes is pH-dependent, and so combined complexation and pH control is used to separate multicomponent mixtures.

Since the pH defines the potential of hydrogen evolution, its variation may be used to extend negatively the potential window for metal ion discharge. One may also use hydrogen evolution for depolarizing or for preventing deposition of less noble metals, subject to the deleterious effects on the quality of the metal deposit (see above).

### Kinetically Based Separation Strategies

The primary exploitation of kinetics for effecting separation relates to hydrogen evolution, the kinetics of which varies significantly with the (metal) electrode material. Slow hydrogen evolution allows copper to be deposited from acid solutions and is the reason why such a wide range of metal ions can be reduced at mercury.

Kinetic separation of metals from each other is less common. A notable exception is the electrogravimetric determination of copper/bismuth mixtures. This relies on bismuth deposition (pH 5.2–6.0,  $-0.30$  V), being extremely slow on platinum or copper-plated platinum but rapid at a bismuth-plated cathode.

## Monitoring the Deposition Process

In controlled potential electrolysis under convective/diffusional control, the reactant concentration decays exponentially with time with an effective rate constant ( $k$  ( $s^{-1}$ )) given by

$$k = DA/V\delta \quad [6]$$

where  $D$  is the diffusion coefficient. The progress can be monitored via the current decay toward zero, at which point the extent of deposition is thermodynamically determined according to eqn [2]. Any  $iR$  drop in solution may extend the electrolysis time by failing to maintain a diffusion-limited current regime. Note that the rate of constant current deposition is subject to similar constraints since any current beyond the diffusionaly limited value will effect a different, possibly undesirable, reaction.

In internal electrolysis, since the cell emf is distributed across the cell as the  $iR$  drop, the deposition rate is inversely proportional to  $R$ : the maximal deposition rate is thus achieved by minimizing  $R$ . The progress of the reaction can be monitored via the cathode potential or current, although variation of  $R$  during the electrolysis distorts the simple exponential decay of the current. The determination itself can be based on spontaneous current measurement during internal electrolysis, although this is not normal practice. For example, determination of cyanide or fluoride in potable water can be based on empirical correlation of current and concentration.

## Applications and Separations

### Classification of Species

Conditions under which common elemental metals can be electrodeposited are shown in **Table 1**. The significance of the groupings is that metals from different groups can generally be separated using potential control alone: successive deposition and weighing sequences allow determination of such mixtures. Separation of metals within a group normally requires an additional control variable, usually

**Table 1** Grouping of metals deposited in the elemental state

Group	Metal	Solution requirement(s)	Comments
Precious metals group	Au	Acid or cyanide	Cyanide allows Au separation In halide media less Hg(II): require greater deposition potential In nitrate, tendency for oxide deposition on anode
	Pt	Acid (not cyanide)	
	Hg	HNO <sub>3</sub> , HCl, or NH <sub>3</sub>	
	Ag	Acid nitrate or citrate; neutral cyanide; ammonia	
Copper group	Cu	Acid sulfate, nitrate, or chloride; alkaline cyanide; ammonia	Cu <sub>2</sub> O formation at high pH
	Bi	HNO <sub>3</sub> , nitrate, chloride, sulfate, oxalate, tartrate	
	Sb	Hot HCl, require Sb(III) state	
	As	As(III) state	
Lead group	Pb	Chloride or tartrate	Alloys with Pt; can also be deposited as oxide on anode
	Sn	HCl	
Zinc group	Ni	Ammonia, oxalate	Redissolves in dilute acids; amalgamated brass cathodes useful (As Ni) (As Ni) (As Ni); alloys with Pt
	Co	Ammonia, oxalate	
	Cd	Alkaline, neutral, or acid media	
	Zn	Acid citrate (controlled pH; not strong acid)	
Mercury cathode group	As, Cr, Ga, Ge, In, Ir, Fe, Mo, Re, Se, Te, Tl	Alkaline media prevents H <sub>2</sub> evolution; complexing agents prevent hydroxide formation	High overpotential prevents H <sub>2</sub> evolution



**Table 2** Species determined as nonelemental electrodeposits

<i>Solution species</i>	<i>Deposited as</i>	<i>Anode</i>	<i>Required conditions</i>	<i>Conversion factor</i>
Pb <sup>2+</sup>	PbO <sub>2</sub>	Pt	Strong HNO <sub>3</sub>	0.8660–0.8605 <sup>a</sup>
Tl <sup>+</sup>	Tl <sub>2</sub> O <sub>3</sub>	Pt	Nitrate/ammonia (Mn, Pb, and Bi interfere)	0.895
Mn <sup>2+</sup>	MnO <sub>2</sub>	Pt	HCOOH/HCOONa	
Co <sup>2+</sup>	Co <sub>2</sub> O <sub>3</sub>	Ag	Acetate buffer	0.710
Br <sup>−</sup> , Cl <sup>−</sup>	AgX	Ag		

<sup>a</sup>Theoretical value: 0.8660.**Table 3** Deposition of metals by internal electrolysis

<i>Metal</i>	<i>Noninterfering metals</i>	<i>Solution</i>	<i>Anode half-cell</i>
Ag	Pb, Cu, Bi	HNO <sub>3</sub>	Cu/Cu(NO <sub>3</sub> ) <sub>2</sub>
Bi	—	HNO <sub>3</sub>	Zn/ZnCl <sub>2</sub>
Cd	Zn	Acetate/NH <sub>4</sub> Cl/N <sub>2</sub> H <sub>4</sub>	Zn/ZnCl <sub>2</sub>
Cu	Ni, Zn	HNO <sub>3</sub>	Zn/ZnCl <sub>2</sub>
Hg	Cu, Zn	H <sub>2</sub> SO <sub>4</sub>	Cu/CuSO <sub>4</sub>
Ni	—	NH <sub>3</sub> /NH <sub>4</sub> Cl/SO <sub>4</sub> <sup>2−</sup>	Zn/ZnCl <sub>2</sub>
Pb	Zn	HCl + 0.1% gelatin; acetate	Zn/Zn(NO <sub>3</sub> ) <sub>2</sub>
Zn	—	Acetate	Mg/NH <sub>4</sub> Cl, HCl

complexation and/or pH. Deposition of the more electronegative elements on a mercury cathode has the (surmountable) complication of mercury loss in transfer, as droplets or by volatilization.

An alternative to cathodic deposition of the elemental metal is anodic deposition of a higher oxide (Table 2). For lead, thallium, manganese, and cobalt, this allows separation from the vast majority of metals. This concept can be extended to determination of bromide and chloride, as the respective insoluble silver halides.

Where the metal can exist in two stable valence states in solution, there is a problem associated with redox cycling between the two electrodes. For example, in concentrated chloride media, reduction of Cu<sup>2+</sup> at the cathode generates CuCl<sub>3</sub><sup>−</sup>, which migrates to the anode and is oxidized back to Cu<sup>2+</sup>, and so Cu<sup>0</sup> is not deposited. In this case, the problem is avoided by keeping the chloride concentration low. An analogous problem can arise with Fe<sup>3+</sup>/Fe<sup>2+</sup> in tartrate media. Even if the cycled metal is not the object of deposition, the decrease in current efficiency can prevent successful analysis.

### Case Studies

Electrodepositions from solutions containing only one depositable metal ion are straightforward. Separations of two-component mixtures also present few problems. Examples include the separations of silver from copper (at −0.24 V in ammoniacal me-

dia), of cadmium from zinc (at −0.80 V in pH 4 acetate media), and of nickel from iron (at −1.10 V in an ammoniacal tartrate medium containing sodium sulfite). For more complex mixtures, the interplay of potential, deposited material, complexation, and pH one must use are illustrated by three case studies. Generally, one must retain close control of the pH in separations from citrate or tartrate media and be aware of the presence of one metal influencing the deposition of another, illustrated by the inhibition by aluminum of copper and bismuth deposition.

None of the examples below involves the internal electrolysis method. This is because, despite its simplicity, it suffers from slow deposition rates. This is less of a disadvantage when the amount of metal to be deposited is small (<50 mg), as in determinations of impurities in more base metals or alloys. Determinations of copper and bismuth in lead–tin alloys or of bismuth in pig-lead fall within this category; other examples are given in Table 3.

*Copper and lead in brasses and bronzes.* Nitric acid treatment of brass leads to chemical separation of tin as SnO<sub>2</sub> · xH<sub>2</sub>O; after filtration and an ignition procedure, the tin can be gravimetrically determined. Thereafter, the analysis is based upon cathodic deposition of Cu and anodic deposition of PbO<sub>2</sub>. If copper alone is to be determined, deposition from a pH 4 tartrate medium with hydrazine as the anodic depolarizer gives separation from most common metals. If lead is also to be determined, one uses a concentrated nitric acid medium, from which PbO<sub>2</sub> is deposited



quantitatively. Copper is not deposited quantitatively under these conditions, and so, following lead deposition, urea is added to lower the nitrate concentration and allow complete Cu deposition. Procedures exist for related samples, such as nickel bronzes and nickel silver (Cu, Ni, and Zn).

*Cu, Bi, Pb, Cd, Zn, and Sn mixtures.* In pH 5.2–5.9 tartrate media, Sn(IV) is complexed. Using gelatin as a brightener and hydrazine as a depolarizer, one can successively deposit and weigh Cu (at  $-0.30$  V), Bi (at  $-0.40$  V), and Pb (at  $-0.60$  V); Pt/Pb alloy formation is prevented by prior deposition of Cu. Addition of ammonia then allows sequential depositions of Cd (at  $-1.20$  V) and Zn (at  $-1.50$  V) since hydrogen evolution occurs at more negative potentials. Addition of HCl decomposes the tin–tartrate complex, allowing tin deposition (at  $-0.65$  V) on a fresh cathode to avoid Zn redissolution. The Pb, Cd, and Zn aspects of this procedure are relevant to the analysis of flue dust from zinc refineries.

*Alloys of Cu, Sb, Pb, and Sn.* This procedure for analysis of bearing metals relies on two successive depositions of pairs of metals. Each (mixed) deposit is dissolved, and then one of the components is selectively deposited (and determined) under different conditions and the other component determined by the difference. First, copper and antimony are codeposited from HCl, with hydrazine as a depolarizer. The (weighed) deposit is dissolved in nitric/hydrofluoric acid, which retains antimony in solution as a fluoride complex, allowing deposition of pure copper (at  $-0.40$  V). Tin and lead are codeposited (at  $-0.70$  V) from the initial residual solution. After redissolution of the (weighed) deposit in nitric/hydrofluoric acids, lead is deposited anodically as  $\text{PbO}_2$ . The Pb and Sn aspect of this procedure is useful for analysis of solders. An analogous procedure allows Ni and Co separation (via  $\text{Co}_2\text{O}_3$ ).

## Recent Developments

Two of the recognized disadvantages of the classical electrogravimetry technique are the need for removal of the electrode from solution for the gravimetric component of the procedure and the practicalities of direct mechanical weighing. The first of these places stringent demands on the deposit characteristics (see above) and makes the procedure slow. The second of these limits the sensitivity of the method. It would clearly be advantageous to have a means of determining the mass of electrodeposited films *in situ*.

In fact, measurements of inertial mass of thin films on solid substrates exposed to a gaseous (or vacuum) environment have long been made using

thickness shear mode (TSM) quartz crystal resonators. The concept is that the resonant frequency of a piezoelectric wafer (here, quartz) sandwiched between two exciting electrodes responds to changes in the acoustic load of the system. For a rigid film typified by the materials, one generally assays using electrogravimetry – this is synonymous with inertial mass. For AT-cut quartz crystal TSM resonators subject to such loading, a change in mass ( $m$  (g)) on an electrode (area  $A$  ( $\text{cm}^2$ )) can be determined from the accompanying resonant frequency change ( $\Delta f$  (Hz)) from its initial value ( $f_0$ ) via the Sauerbrey equation:

$$\Delta f = - \left( \frac{2}{\rho_q v_q} \right) f_0^2 \left( \frac{\Delta m}{A} \right) \quad [7]$$

where  $\rho_q$  and  $v_q$  are the density and shear wave velocity within the quartz respectively. Such a device is referred to as quartz crystal microbalance (QCM).

Immersion in a liquid (as required of an electrochemical experiment) attenuates the propagation of the shear wave, and it was long presumed that this was fatal to *in situ* operation of a QCM. However, it has recently been shown that the crystal continues to oscillate under such conditions. In the electrochemical context, where one uses a conducting solution, the electrochemical QCM (EQCM) technique can be used if only one electrode is exposed to the solution and forms the working electrode in a three-electrode cell. Typical instrumental performance allows frequency (mass) change measurements in response to electrochemically controlled processes on a timescale of less than 1 s. For typical resonators ( $f_0 \approx 5$ – $10$  MHz,  $A \approx 0.2$   $\text{cm}^2$ ), the lower level of measurable frequency changes is of the order of a few Hertz; this corresponds to mass changes of a few nanograms, i.e., a sensitivity of  $\sim 10$  pmol.

Although the EQCM is a relatively new technique and its full analytical potential has not been explored, the above characteristics make it an attractive sensing device. Despite the fact that the principle of mass measurement is different from the classical methodology, measurement of surface mass change using the technique is referred to as electrogravimetry. Interestingly, in addition to the types of system described earlier (metals, oxides, and salts), the *in situ* nature of the technique allows extension to different surface binding chemistries. These may be based on ion exchange, partition into a film, or chemical bonding. In the latter case, this can be anything from coordination chemistry of inorganic analytes to antibody–antigen interactions of biological analytes.

See also: **Coulometry. Gravimetry. Ion-Selective Electrodes:** Overview. **Kinetic Methods:** Principles and Instrumentation.

## Further Reading

- Buttry DA (1991) Applications of the quartz crystal microbalance to electrochemistry. In: Bard AJ (ed.) *Electroanalytical Chemistry*, vol. 17, pp. 1–85. New York: Dekker.
- Christian GD (1986) *Analytical Chemistry*, 4th edn., pp. 326–330. New York: Wiley.
- Curran DJ and Fletcher KS (1986) Coulometry and electrogravimetry. In: Christian GD and O'Reilly JE (eds.) *Instrumental Analysis*, 2nd edn. Boston: Allyn and Bacon.
- Harris DC (1999) *Quantitative Chemical Analysis*, 5th edn., pp. 34, 451–457, and 861. New York: W H Freeman.
- Hillman AR (2003) The electrochemical quartz crystal microbalance. In: Bard AJ and Stratmann M (eds.) *Encyclopedia of Electrochemistry*, vol. 3, pp. 230–289. New York: Wiley.

- Kolthoff IM, Sandell EB, Meehan EJ, and Bruckenstein S (1969) *Quantitative Chemical Analysis*, 4th edn., pp. 928–937 and 1040–1043. London: Macmillan.
- Lingane JJ (1958) *Electroanalytical Chemistry*, 2nd edn., chs. XIV–XVII, pp. 351–445. New York: Wiley.
- Page JA (1963) Electrogravimetry In: *Handbook of Analytical Chemistry*, pp. 5–170 to 5–186, New York: McGraw-Hill.
- Plambeck JA (1982) *Electroanalytical chemistry*. New York: Wiley.
- Skoog DA, West DM, and Holler FJ (1996) *Fundamentals of Analytical Chemistry*, 7th edn., pp. 440–446 and 852–853. New York: Saunders.
- Skoog DA, West DM, Holler FJ, and Crouch SR (2000) *Analytical Chemistry: An Introduction*, 7th edn., pp. 516–522 and 762–764. New York: Saunders.
- Tanaka N (1963) Electrodeposition. In: Kolthoff IM and Elving PJ (eds.) *Treatise on Analytical Chemistry*, part I, vol. 4, ch. 48, pp. 2417–2473. New York: Wiley.
- Wilson CL and Wilson DW (1964) *Comprehensive Analytical Chemistry*, vol. IIA, ch. II, pp. 7–64. Amsterdam: Elsevier.

# ELECTRON ENERGY LOSS SPECTROMETRY

**J Verbeeck, S Hens, P Potapov, and D Schryvers,**  
University of Antwerp, Antwerp, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction: Basic Principles

This article will focus on the use of electron energy loss spectrometry (EELS) in a transmission electron microscope (TEM) or a scanning transmission electron microscope (STEM). In a TEM or STEM, a beam of electrons is accelerated to energies typically between 100 keV and 1 MeV. The beam of electrons is transmitted through a sample that consists of a thin piece of material (typically less than 50 nm thickness). Interaction of the beam with the sample enables the operator to learn something about the sample, such as the chemical elements present, stoichiometry, energy levels, electronic structure, and more.

The EELS technique focuses on the inelastic interaction of the primary beam electrons with the electrons from the sample.

The loss of energy of the primary electrons that is commonly used in EELS is in the range of 0 eV–3 keV. This energy loss is measured by using the dispersive properties of a homogeneous magnetic

field on the charged electrons in an EEL spectrometer. When fast electrons travel through a region of homogeneous magnetic field  $B$  perpendicular to their momentum  $m_e v$ , they are bent in a circular path with radius  $r$  obtained by using the Lorentz force on the particles:

$$r = \frac{\gamma m_e v}{eB} \quad [1]$$

where  $e$  is the charge of the electrons and  $\gamma$  a relativistic correction factor. The radius of the path is thus dependent on the energy of the particle and can be used to discriminate particles with respect to their energy. A so-called EEL spectrum is recorded by counting the primary electrons sorted by energy. A typical spectrum is shown in **Figure 1**. The spectrum is obtained from a thin sample of  $\text{La}_{0.85}\text{Ca}_{0.15}\text{MnO}_3$  and can be divided into several regions that contain information about the sample.

The pronounced peak at 0 eV is labeled as the zero-loss peak. The zero-loss peak contains all primary electrons that did not undergo a noticeable energy loss during the transmission through the sample. The width of this zero-loss peak can mainly be attributed to the effect that the source of primary electrons, the electron gun, has its own intrinsic energy spread. This energy spread is dependent on the type of electron gun but is practically in the order of 0.3–1 eV.

See also: **Coulometry. Gravimetry. Ion-Selective Electrodes:** Overview. **Kinetic Methods:** Principles and Instrumentation.

## Further Reading

- Buttry DA (1991) Applications of the quartz crystal microbalance to electrochemistry. In: Bard AJ (ed.) *Electroanalytical Chemistry*, vol. 17, pp. 1–85. New York: Dekker.
- Christian GD (1986) *Analytical Chemistry*, 4th edn., pp. 326–330. New York: Wiley.
- Curran DJ and Fletcher KS (1986) Coulometry and electrogravimetry. In: Christian GD and O'Reilly JE (eds.) *Instrumental Analysis*, 2nd edn. Boston: Allyn and Bacon.
- Harris DC (1999) *Quantitative Chemical Analysis*, 5th edn., pp. 34, 451–457, and 861. New York: W H Freeman.
- Hillman AR (2003) The electrochemical quartz crystal microbalance. In: Bard AJ and Stratmann M (eds.) *Encyclopedia of Electrochemistry*, vol. 3, pp. 230–289. New York: Wiley.

- Kolthoff IM, Sandell EB, Meehan EJ, and Bruckenstein S (1969) *Quantitative Chemical Analysis*, 4th edn., pp. 928–937 and 1040–1043. London: Macmillan.
- Lingane JJ (1958) *Electroanalytical Chemistry*, 2nd edn., chs. XIV–XVII, pp. 351–445. New York: Wiley.
- Page JA (1963) Electrogravimetry In: *Handbook of Analytical Chemistry*, pp. 5–170 to 5–186, New York: McGraw-Hill.
- Plambeck JA (1982) *Electroanalytical chemistry*. New York: Wiley.
- Skoog DA, West DM, and Holler FJ (1996) *Fundamentals of Analytical Chemistry*, 7th edn., pp. 440–446 and 852–853. New York: Saunders.
- Skoog DA, West DM, Holler FJ, and Crouch SR (2000) *Analytical Chemistry: An Introduction*, 7th edn., pp. 516–522 and 762–764. New York: Saunders.
- Tanaka N (1963) Electrodeposition. In: Kolthoff IM and Elving PJ (eds.) *Treatise on Analytical Chemistry*, part I, vol. 4, ch. 48, pp. 2417–2473. New York: Wiley.
- Wilson CL and Wilson DW (1964) *Comprehensive Analytical Chemistry*, vol. IIA, ch. II, pp. 7–64. Amsterdam: Elsevier.

# ELECTRON ENERGY LOSS SPECTROMETRY

**J Verbeeck, S Hens, P Potapov, and D Schryvers,**  
University of Antwerp, Antwerp, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction: Basic Principles

This article will focus on the use of electron energy loss spectrometry (EELS) in a transmission electron microscope (TEM) or a scanning transmission electron microscope (STEM). In a TEM or STEM, a beam of electrons is accelerated to energies typically between 100 keV and 1 MeV. The beam of electrons is transmitted through a sample that consists of a thin piece of material (typically less than 50 nm thickness). Interaction of the beam with the sample enables the operator to learn something about the sample, such as the chemical elements present, stoichiometry, energy levels, electronic structure, and more.

The EELS technique focuses on the inelastic interaction of the primary beam electrons with the electrons from the sample.

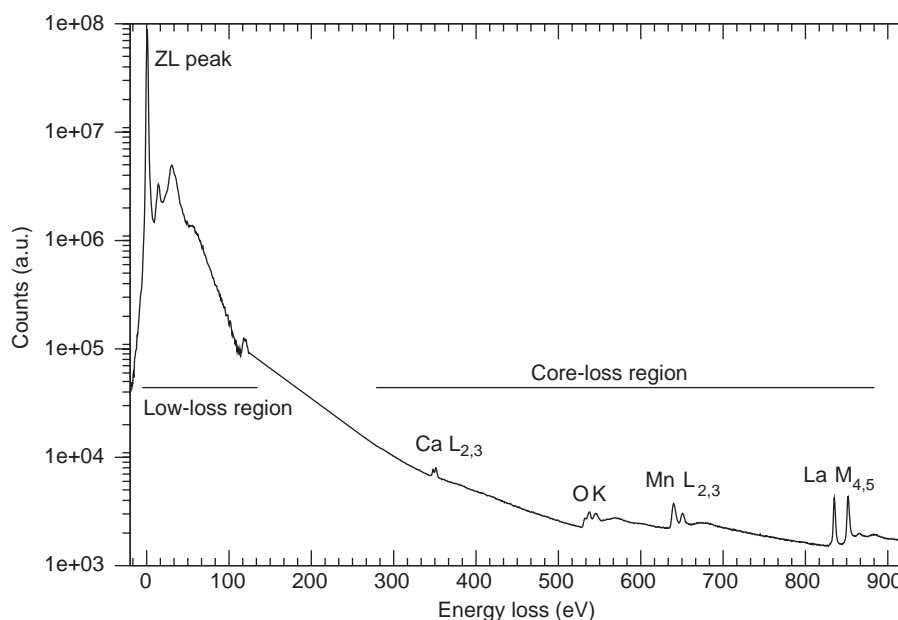
The loss of energy of the primary electrons that is commonly used in EELS is in the range of 0 eV–3 keV. This energy loss is measured by using the dispersive properties of a homogeneous magnetic

field on the charged electrons in an EEL spectrometer. When fast electrons travel through a region of homogeneous magnetic field  $B$  perpendicular to their momentum  $m_e v$ , they are bent in a circular path with radius  $r$  obtained by using the Lorentz force on the particles:

$$r = \frac{\gamma m_e v}{eB} \quad [1]$$

where  $e$  is the charge of the electrons and  $\gamma$  a relativistic correction factor. The radius of the path is thus dependent on the energy of the particle and can be used to discriminate particles with respect to their energy. A so-called EEL spectrum is recorded by counting the primary electrons sorted by energy. A typical spectrum is shown in **Figure 1**. The spectrum is obtained from a thin sample of  $\text{La}_{0.85}\text{Ca}_{0.15}\text{MnO}_3$  and can be divided into several regions that contain information about the sample.

The pronounced peak at 0 eV is labeled as the zero-loss peak. The zero-loss peak contains all primary electrons that did not undergo a noticeable energy loss during the transmission through the sample. The width of this zero-loss peak can mainly be attributed to the effect that the source of primary electrons, the electron gun, has its own intrinsic energy spread. This energy spread is dependent on the type of electron gun but is practically in the order of 0.3–1 eV.



**Figure 1** A typical EEL spectrum from an LCMO ( $\text{La}_{0.85}\text{Ca}_{0.15}\text{MnO}_3$ ) thin film. The spectrum is divided into low-loss and core-loss regions and the zero-loss peak and core excitation edges are labeled. Note the use of a logarithmic intensity scale to cover the wide dynamic range present in a typical EEL spectrum.

Smaller energy spreads are possible by making use of monochromators. The width of the zero-loss peak is the main limiting factor for the energy resolution of the EELS experiment since the experimental spectrum can be regarded as a convolution of an ideal spectrum with the energy distribution of the gun and spectrometer. It is customary to define the energy resolution in EELS experiments as the full-width at half-maximum of the zero-loss peak.

The energy range of  $\sim 0$ – $100$  eV is mostly dominated by the so-called plasmon peaks and is called the low-loss region. Typical for this range is a broad peak at  $E_p$  (typically  $10$ – $30$  eV) that is repeated with decreased intensity at  $2E_p$ ,  $3E_p$ , etc., until the peaks are no longer visible. The origin of these plasmon peaks lies in the excitation of the so-called plasmons (collective oscillations of the valence electrons) by the primary electrons. The position and width of these plasmon peaks are related to the electronic structure of the sample material and can be used to discriminate between different materials.

Apart from collective plasmon excitations, other excitations like valence state excitations and low lying core state excitations are present in the low-loss region and make this region generally difficult to use for chemical quantification. The high probability for excitations in the low-loss range, however, allows the acquisition of high-quality spectra in short exposure times, making it a good region when beam damage is

problematic. A disadvantage is the so-called delocalization effect that intrinsically limits the spatial resolution and can be understood in terms of a classical impact parameter.

Moreover, the dielectric response of the material can be obtained from the low-loss region via a Kramers–Kronig analysis if the EEL spectrum is recorded over a wide energy range. Although the resulting dielectric function has poorer energy resolution than that extracted from, e.g., light absorption spectroscopy, the energy range can be greater covering visible light, ultraviolet, and soft X-ray regions.

The region above  $100$  eV energy loss is generally called the core-loss region. This region is dominated by a smoothly decaying background and element-specific core excitations. These core excitations are caused by the excitation of an atomic core-state of an atom in the sample to an unoccupied higher lying state of the crystal or even to a free electron state. The energy position of these core excitations is determined by the difference in core-state energy and the energy of the first free-state and enables a sensitive discrimination between different elements. Core excitations are generally labeled according to the type of core state that was excited. The naming conventions are given in Table 1 and are used in Figure 1.

The methods of extracting chemical information from a core-loss EEL spectrum are analogous to that in energy-dispersive X-ray (EDX) spectroscopy. The

**Table 1** Naming convention for core excitations depending on core state quantum numbers  $n$ ,  $l$ , and  $j$ 

Label	Core state
K	1s <sub>1/2</sub>
L <sub>1</sub>	2s <sub>1/2</sub>
L <sub>2</sub>	2p <sub>1/2</sub>
L <sub>3</sub>	2p <sub>3/2</sub>
M <sub>1</sub>	3s <sub>1/2</sub>
M <sub>2</sub>	3p <sub>1/2</sub>
M <sub>3</sub>	3p <sub>3/2</sub>
M <sub>4</sub>	3d <sub>3/2</sub>
M <sub>5</sub>	3d <sub>5/2</sub>

probability of an inelastic core excitation is given by the cross-section  $\sigma_a$ . It relates the number of counts  $I_a$  in a chosen energy range of the core-excitation edge (disregarding the background) to the number of atoms  $N_a$  of a certain element in the interaction volume  $V$  probed by the primary electrons:

$$I_a = I_0 N_a \sigma_a \quad [2]$$

with  $I_0$  the total amount of electron counts in the spectrum.

The volume concentration  $n_a$  is then given by

$$n_a = \frac{I_a}{I_0 \sigma_a V} \quad [3]$$

This enables the operator to determine chemical concentrations from an EEL spectrum if all parameters are known. In general, accurate determination of all these parameters is difficult and mostly the element ratio technique is used to overcome this problem. In the element ratio technique, the ratio of two elements is determined, which cancels out all parameters if both parts of the excitation spectrum are taken at exactly the same experimental conditions:

$$\frac{n_a}{n_b} = \frac{\sigma_b[\Delta E]}{\sigma_a[\Delta E]} \frac{I_a[\Delta E]}{I_b[\Delta E]} = k_{ab}[\Delta E] \frac{I_a[\Delta E]}{I_b[\Delta E]} \quad [4]$$

where  $k_{ab}$  is the so-called  $k$ -factor that can be experimentally determined from calibration samples or can be approximately calculated from physical arguments. It is important to note that these  $k$ -factors are dependent on many parameters (like the energy width  $\Delta E$ ) and they should be considered as microscope dependent. This makes it difficult to publish tabulated  $k$ -factors in contrast to EDX spectroscopy. On the other hand, there is no need for backscattering and absorption corrections that make reliable EDX quantification so difficult.

Moreover, EELS is very sensitive to light elements that are problematic in EDX spectroscopy. This makes EELS an interesting technique to study, for instance, organic materials.

In order to obtain  $I_a$ , the counts in a certain core-excitation edge, one has to get rid of the nonspecific background that is always present in an EEL spectrum. The background is caused by the tails of all lower-lying excitations that add up to a background that is generally approximated by a power-law function  $B(E)$ :

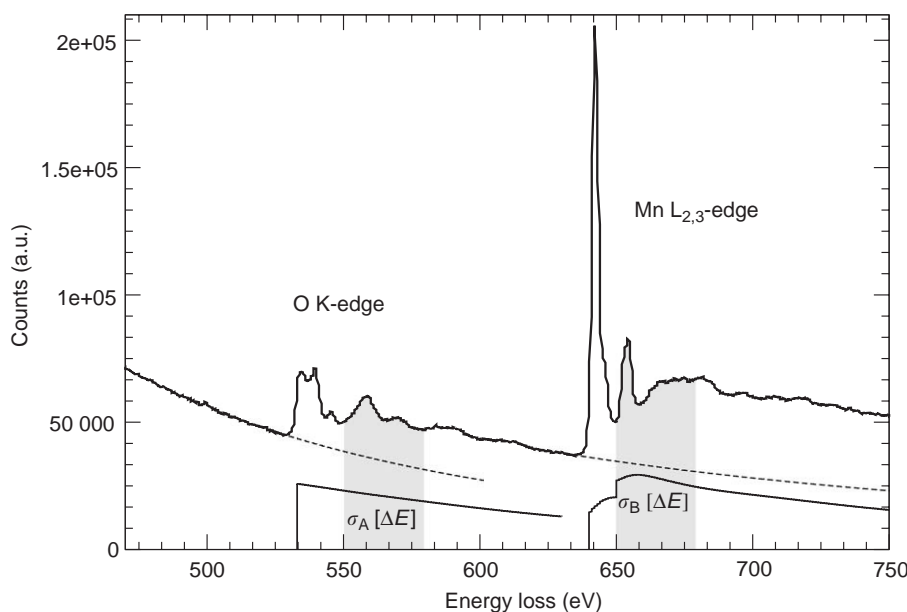
$$B(E) = AE^{-r} \quad [5]$$

where  $A$  and  $r$  are parameters that depend on the experiment and  $E$  the energy loss. The removal of this background is conventionally done by fitting the power-law function in a user-selected region preceding the excitation edge (the so-called fitting region) and then extrapolating it under the excitation edge. After removal of the background, the excitation edge can be integrated over a certain energy range to find the total amount of electrons  $I_a[\Delta E]$  that were scattered by this excitation in the given energy range  $\Delta E$ . These steps are schematically presented in Figure 2 for an MnO sample.

In many cases, the background removed excitation edge is further corrected for the so-called multiple scattering. Multiple scattering is the effect when an electron loses energy to more than one excitation process when interacting with the sample. This leads to a blurring of the experimental spectrum with redistribution of counts to higher energies. Removing this multiple scattering is mostly done by deconvoluting the core-loss spectrum with a low-loss spectrum. This is one of the reasons why in most cases both core-loss and low-loss spectra are recorded during an experiment, taking care that both are recorded under equivalent conditions.

Removal of multiple scattering is especially important if the shape of the excitation edge is to be studied, it removes the thickness-dependent blurring and redistribution and allows for a comparison between the shapes of excitation edges taken from different samples or from different regions in the same sample. The shape of an excitation edge close to the edge onset is related to the electronic structure and can be studied with energy loss near edge structure (ELNES). The shape of the edge at energies slightly above the edge onset can be used to obtain element-specific radial distribution functions with a technique called extended energy loss fine structure (EXELFS). For a detailed description of EXELFS see the textbooks listed in the Further Reading section.





**Figure 2** Extraction of the Mn/O elemental ratio from the spectrum of MnO. Both edges are analyzed within the energy window of the same width (marked in gray) and the characteristic intensity is separated from the extrapolated background. The EELS scattering cross-sections for O and Mn atoms are calculated and integrated over the same windows.

## Technical Details

EEL spectral quantification is very sensitive to the exact microscope setup used by the operator. A successful quantification requires knowledge of at least the following experimental parameters:

- acceleration voltage;
- convergence angle: the angle of the cone that is formed by the electron beam on the sample with respect to the optical axis; and
- collection angle: the angle with respect to the optical axis up to which the transmitted (inelastic) electrons are collected in the spectrometer.

Calibration of these parameters for a specific microscope is essential. The details on what determines these angles are dependent on whether the operator uses a TEM or STEM setup and on the type of spectrometer that is used. The details are treated at length in several of introductory textbooks on EELS mentioned in the Further Reading section.

## Spot Analysis

The simplest operation mode in EELS is focusing the incident electron beam at the place of interest and analyzing the spectral distribution of the transmitted beam. The probe size can be as small as 1 nm and can even be reduced down to atomic sizes in dedicated STEM instruments.

The most common information that can be obtained from EELS spot analysis is:

- the local chemical composition and
- local atomic bonding and valence characterization.

### Local Chemical Analysis

As discussed in the introduction, core excitation edges appear in the spectrum at element characteristic energies. Therefore, EEL spectra can directly be used for qualitative analysis of the local chemical composition. Solid-state effects might slightly shift the edge position depending on the specific electron structure of the material. However, this chemical shift does not generally exceed 5 eV, which, in the majority of cases, cannot cause wrong identification of the element. Quantification of the spectra then proceeds with removal of background and multiple scattering to yield element ratios.

### Local Bonding and Oxidation State

Near the threshold, ionization edges often show fine details, called ELNES, which contains information about the electronic structure of the probed material. In the event of ionization, electrons from a core shell are excited into unoccupied states above the Fermi level. Just near the threshold these states represent an unoccupied part of the conduction band, which is sensitive to changes of bonding and oxidation state in solids. Thus, it is possible to track bonding and



oxidation state trends at an extremely small scale by comparing nanoprobe EEL spectra with theoretical calculations or with experimental reference spectra. The latter is generally called fingerprinting and is widely used for identification of specific compounds. An example is given in **Figure 3** where the C K-edges from graphite and diamond can be easily distinguished by comparing the ELNES patterns. In graphite, a sharp peak associated with the  $\pi^*$  orbital is found a few electronvolts before the major group of peaks associated with  $\sigma^*$  orbitals. In many cases, however, the interpretation of the ELNES is not so straightforward and theoretical calculations are needed to relate the ELNES features with the bonding and oxidation state.

Following the inelastic scattering theory, the scattering cross-section can be expressed as

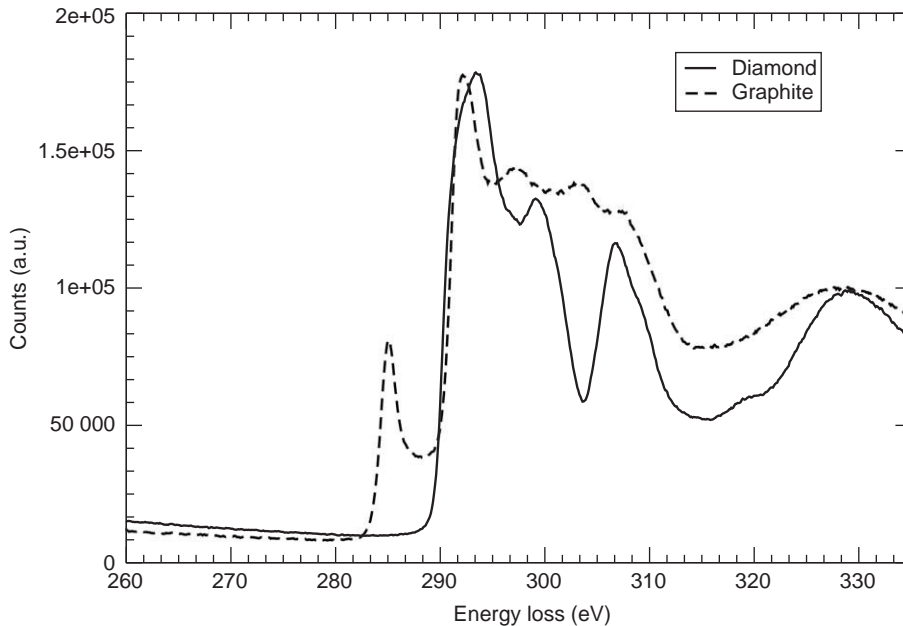
$$\sigma \propto \int \frac{d\vec{q}}{q^3} \sum_f |\langle f | e^{i\vec{q} \cdot \vec{r}} | i \rangle|^2 \quad [6]$$

where  $\vec{q}$  is the scattering vector,  $i$  and  $f$  are initial and final electron states, and the summation is performed over all possible unoccupied final states above the Fermi level. The term  $|\langle f | e^{i\vec{q} \cdot \vec{r}} | i \rangle|^2$  in the integral is significant only in an area where initial and core states spatially overlap. The initial core state is highly localized around a nucleus, thus the conduction band is probed selectively for each given atom position. The term  $e^{i\vec{q} \cdot \vec{r}}$  can be expanded as  $1 + \vec{q} \cdot \vec{r} + \frac{1}{2}(\vec{q} \cdot \vec{r})^2 + \dots$  and neglecting the terms higher than first

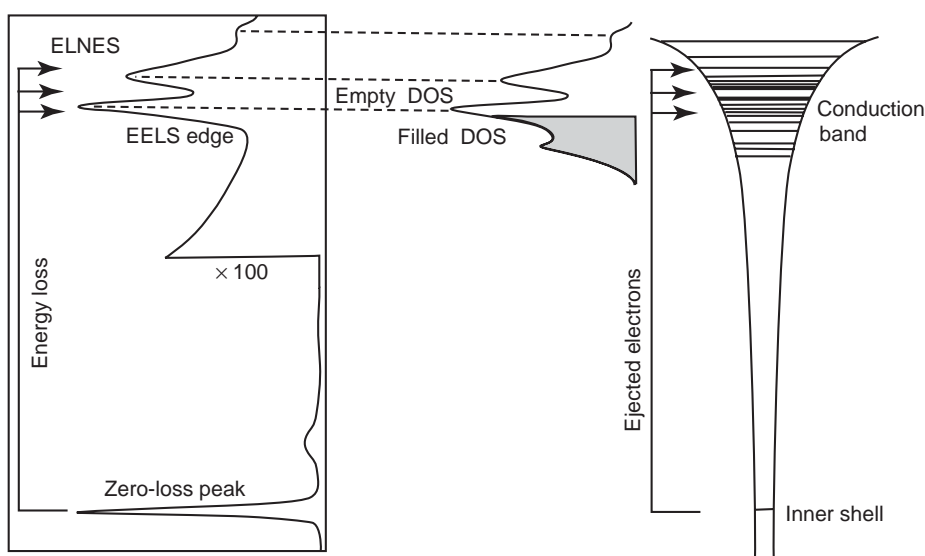
order results in significant simplification of eqn [6] called the dipole approximation. In the dipole approximation, the term  $|\langle f | \vec{r} | i \rangle|^2$  does not vanish only if the angular momentum  $l$  of the final state differs by one from the angular momentum of the core state. Thus, only final states with a selected angular momentum  $l$  are probed. A K-edge ( $l_0 = 0$ ) probes only unoccupied p-states with  $l = 1$  while  $L_{2,3}$ -edges ( $l_0 = 0$ ) probe both unoccupied d-states ( $l = 2$ ) and s-states ( $l = 0$ ) with a higher probability for going to d-states. The dipole approximation has been proved to be valid for small  $\vec{q}$  while for high scattering  $\vec{q}$  vectors dipole forbidden transitions can be observed. Furthermore, it can be shown that in the dipole approximation the energy differential cross-section is expressed as

$$\frac{\partial \sigma(E)}{\partial E} \propto N_{l_0+1}(E)A_{l_0+1}(E) + N_{l_0-1}(E)A_{l_0-1}(E) \quad [7]$$

where  $N_l(E)$  is the density of empty states with a given angular momentum  $l$  in the conduction band and  $A_l(E)$  is a term slightly dependent on energy. As mentioned above, in most cases, only the transition  $l_0 \rightarrow l_0 + 1$  needs to be considered. Equation [7] reveals that under certain conditions the energy differential EELS cross-section is roughly proportional to the angular momentum projected unoccupied local density of states  $N_l(E)$ . This is schematically sketched in **Figure 4**, in which ELNES peaks on the top of a given EELS ionization edge can be related



**Figure 3** ELNES fine structure for carbon and diamond. Although both edges are excitations from the 1s-state of carbon, the different unoccupied final states for both compounds create significant differences in the fine structure of the edges.



**Figure 4** Correspondence between the fine ELNES structure and the DOS in the empty part of the conduction band.

with the density of states (DOS) shape in the empty part of the conduction band.

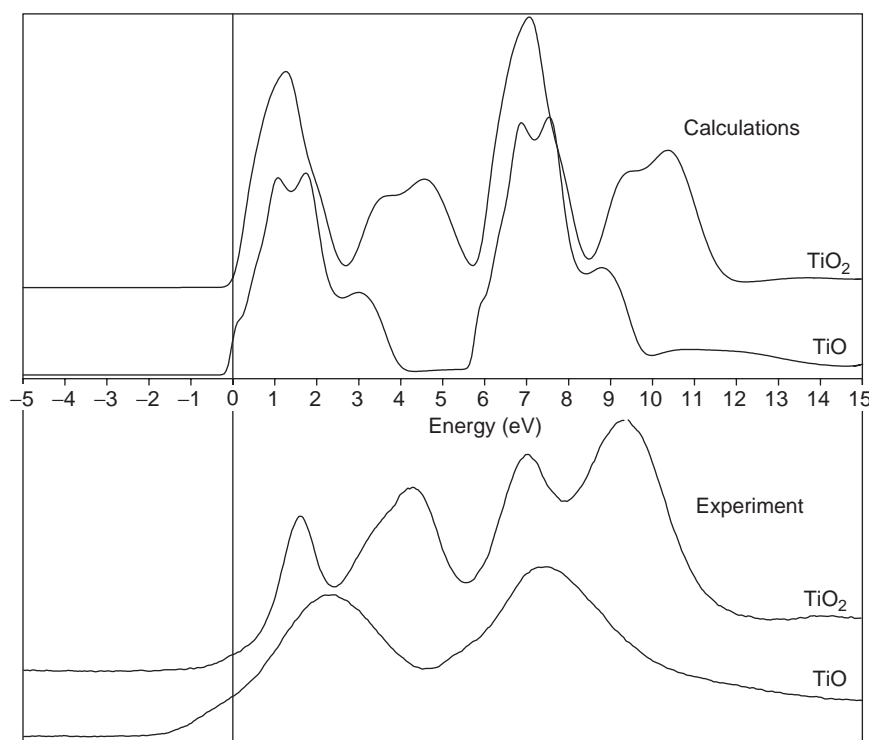
### Numerical Calculations

Several approaches were developed for numerical calculations of the scattering cross-section. In the simplest approximation of a free atom, the final states are assumed to be not affected by the other atoms in the crystal and can be presented as unoccupied bound atomic states and free electron states. Atomic calculations predict well the energy dependence of the EELS cross-sections far above the threshold and are widely used for quantitative chemical analysis (see above). As clearly seen from **Figure 2**, atomic cross-sections fail to reproduce the near-edge structure, which is strongly affected by solid-state effects. The latter can be accounted for by considering the interference of the outgoing ionized free electron wave with waves backscattered from the neighboring atoms. These calculations, called multiple-scattering calculations, reproduce successfully the features of the near-edge structure provided a sufficient number of neighboring atoms is considered. This method operates with atomic clusters in real space and allows one to treat disordered and aperiodic systems like alloys, interface, and surfaces. The main limitation arises from the rapid increase of the calculation time with growing cluster size.

Alternatively, standard electron band calculations like augmented plane waves, linearized augmented plane waves, or pseudopotential methods can be employed. Pseudopotential methods using plane waves and a smooth crystal potential are quick and sufficiently

precise in calculating the final states. These methods, however, cannot treat explicitly the core electrons. Fortunately, in the majority of cases, this is not required as the core electrons are not noticeably affected by crystal fields and are almost the same as those in free atoms. Another group of methods divides the crystal into muffin-tin (MT) spheres and interstitial space. The single particle electron functions are decomposed into the radial functions in the MT spheres and in plane waves in the interstitial region. In such a way the crystal potential produced by both core and outer electrons can be self-consistently calculated using the economic wave function basis yielding a reasonable computation time.

One of the most commonly used computer programs for all-electron calculations is the WIEN code. Except calculating the dipole-forbidden transitions, WIEN can also simulate the  $q$ -dependence of scattering and the effect of the orientation of crystalline material. An example of EEL spectra simulated by WIEN is shown in **Figure 5**. The major peaks in the ELNES of TiO and TiO<sub>2</sub> are well predicted by the *ab initio* calculations although there is still some disagreement in the exact peak positions and their relative intensities. The reason for this is that calculations describe the stationary ground state of an atom while the EEL spectra rather relate to the excited state where the initial unperturbed and the final ionized state are quantum-mechanically superposed. Taking into account the effect of a core hole remaining in the inner shell after excitation of an electron has been demonstrated to improve significantly the agreement with the experimental curves. However, the methods for accounting the core hole



**Figure 5** Theoretical profiles (ELNES) of the Ti L edge in the  $\text{TiO}_2$ -rutile and TiO compared with experimental ELNES from 100 nm particles. No core-hole effect is accounted for.

effect are still not well established and remain under discussion.

## Image Analysis

The technique of obtaining images of specific elements, the so-called elemental maps, is another important application in EELS analysis. Energy filtered TEM (EFTEM) images can be obtained with a so-called imaging spectrometer. However, the images are not directly interpretable as chemical maps because of the nonspecific background in the EELS spectrum.

Two techniques exist to solve this problem by taking more than one image: the three-window technique and the so-called jump-ratio technique. The first approach requires three images for elemental mapping while only two are needed in the second method. The method of the three windows allows one to make an image of which the intensity is proportional to the intensity in the EEL spectrum under a certain excitation edge. The method makes use of three images made up of electrons coming from three different energy regions in the spectrum. Two pre-edge images are used to estimate the non-specific background in each pixel of the image, by

estimating the parameters of the background model  $A \cdot E^{-r}$  for each pixel in the spectrum. The third image (post-edge image) is obtained from the energy region that contains the excitation edge of interest. The estimated background for each pixel is subtracted from the post-edge image and an excitation-specific image remains. The resulting image then contains an intensity per pixel, which is proportional to the number of electrons that have made the specific excitation. If the cross-section is known for the excitation in the specimen (for given experimental conditions), it is then possible to estimate the absolute concentration of a given element. Important parameters for the use of this three-window technique are:

- the position of the energy selecting window;
- the width of the energy window;
- the number of counts in the images (they should be as high as possible but too long exposure times and overexposure should be avoided);
- the spatial and energy drifts during the acquisition of the images (drift between the images can be corrected by using a cross-correlation technique);
- the type of excitation used (the greater the cross-section for an excitation the better the elemental mapping will work); and

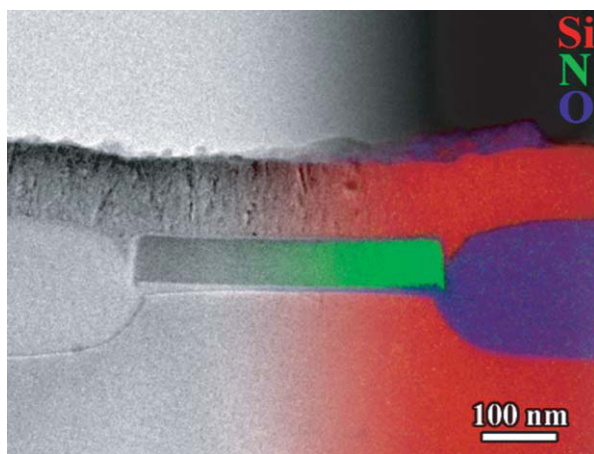
- the thickness of the specimen (comparing regions of comparable thickness is more accurate and easier to interpret).

The optimal settings for these parameters depend on the goal of the experiment.

An alternative to the three-window technique is the so-called jump-ratio technique. In this case, two EFTEM images are recorded: one pre-edge and one post-edge. Dividing the post-edge by the pre-edge image leads to a jump-ratio image. This image gives a qualitative image of the changes in the spectrum between the pre-edge and post-edge regions. No quantitative data can be extracted since the ratio of background plus edge is not directly related to the excitation edge strength but also depends on the background. The advantages of using this technique over the three-window method are:

- A better signal-to-noise ratio in the final image is obtained, since no extrapolation or parameter estimation is needed.
- The image is less sensitive to preserved elastic contrast in the EFTEM images. The ratio cancels out all elastic contrast information that is the same for both pre- and post-edge images.
- Drift problems are reduced since only two images are needed.

The disadvantage, of course, is the loss of quantitative information. Therefore, this technique is only used to discern real chemical changes from elastic contrast effects by comparing the result of the jump-ratio image to the elemental image. In general, it is useful to avoid zone-axis orientations in EFTEM work to limit elastic contrast effects.



**Figure 6** EFTEM image partially pasted on a TEM image of a semiconductor test structure showing the different elements present in the specimen as color coded.

A convenient way to present and compare elemental maps of different elements from the same sample are the so-called color coded maps. These images are prepared by overlaying up to three elemental maps using a different primary color for each. The total image is then a color coded image, with mixed colors where more than one element is present in the same region. The advantage is that the correlation between the different elemental maps is much more visible, compared to the presentation of the different elemental maps as separate gray-scale images (Figure 6).

The spatial resolution in EFTEM images can be as good as 1 nm if a careful selection is made of the apertures in the microscope and specimen drift is kept to a minimum. A good understanding of the imaging process is necessary to obtain high-resolution EFTEM images but treatment is beyond the scope of this article.

**See also:** **Microscopy Techniques:** Electron Microscopy; Scanning Electron Microscopy. **X-Ray Fluorescence and Emission:** X-Ray Fluorescence Theory; Wavelength Dispersive X-Ray Fluorescence; Energy Dispersive X-Ray Fluorescence.

## Further Reading

- Brydson R (2001) *Electron Energy Loss Spectroscopy*. New York: Springer.
- Colliex C, Imhoff D, Perez-Omil JA, *et al.* (1999) Electron energy-loss spectroscopy (EELS) on nanodimensional structures. *Journal of Electron Microscopy* 48(Supplement S): 995–1003.
- Egerton RF (1996) *Electron Energy-Loss Spectroscopy in the Electron Microscope*, 2nd edn. New York: Plenum.
- Jouffrey B, Schatschneider P, and Hebert C (2002) Ionization edges: some underlying physics and their use in electron microscopy. *Advances in Imaging and Electron Physics* 123: 413–450.
- Krivanek OL, Gubbens AJ, Dellby N, and Meyer CE (1992) Design and first applications of a post-column imaging filter. *Microscopy, Microanalysis and Microstructure* 3: 187–199.
- Reimer L (1995) *Energy-Filtering Transmission Electron Microscopy*, Springer Series in Optical Sciences, vol. 71. Berlin: Springer.
- Reimer L, Fromm I, and Rennekamp R (1988) Operation modes of electron spectroscopic imaging and electron energy-loss spectroscopy in a transmission electron microscope. *Ultramicroscopy* 24: 339–354.
- Rez P, Alvarez JR, and Pickard C (1999) Calculation of near edge structure. *Ultramicroscopy* 78: 75–183.
- Williams DB and Carter CB (1996) *Transmission Electron Microscopy: A Textbook for Materials Science*, vol. 4. New York: Plenum.

# ELECTRON PARAMAGNETIC RESONANCE SPECTROSCOPY

See ELECTRON SPIN RESONANCE SPECTROSCOPY: Principles and Instrumentation; Specialized Techniques; Biological Applications

## ELECTRON SPIN RESONANCE SPECTROSCOPY

Contents

**Principles and Instrumentation**

**Specialized Techniques**

**Biological Applications**

### Principles and Instrumentation

**C J Rhodes**, University of Reading, Reading, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by S A Fairhurst, pp. 1013–1017, © 1995, Elsevier Ltd.

Electron spin resonance (ESR) spectroscopy is also known as electron paramagnetic resonance (EPR) spectroscopy and less often, but more desirably, as electron magnetic resonance (EMR) spectroscopy. The last name highlights the similarities with nuclear magnetic resonance (NMR) spectroscopy. In fact, both ESR and NMR share the same basic theory. Whereas NMR deals with nuclei having magnetic moments, ESR refers to electrons, but these must be unpaired. At first sight this requirement appears to be a severe limitation on the usefulness of ESR, but the reality is that a wealth of information can be obtained from this branch of spectroscopy. A large number of transition metal compounds, both with odd and even numbers of electrons, are paramagnetic and thus can be studied by ESR. There are many chemical compounds that have odd numbers of electrons – namely, organic and inorganic free radicals. Other systems having unpaired electrons are molecules in electronic triplet states, impurities in semiconductors, electrons in unfilled conduction bands, and electrons trapped in radiation-damaged sites. Thus there are an enormous number of applications ranging from new materials to biochemical systems – some of these applications will be considered in this article.

### Theory

The fundamentals of the magnetic resonance phenomenon for nuclei and for electrons can be described both in quantum mechanical and in classical terms. The latter description is particularly useful when considering spin physics and pulse methods. Both descriptions lead to the resonance condition, for a particle having a spin value of  $1/2$ :

$$\Delta E = g\mu B \quad [1]$$

where  $\Delta E$  is the separation of energy levels produced by the application of an external magnetic field,  $B$ ,  $g$  is the  $g$ -factor (a tensor quantity), which has a value of 2.0023 for a free electron and 5.5856 for a proton. The magnetic moment  $\mu$ , is given by

$$\mu = eh/4\pi m \quad [2]$$

where  $e$  and  $h$  are fundamental constants and  $m$  is the mass of the particle. For the electron,  $\mu$  (the Bohr magneton) has a value of  $9.274 \times 10^{-24} \text{ J T}^{-1}$  while for the proton,  $\mu$  (the nuclear magneton) has a value of  $5.051 \times 10^{-27} \text{ J T}^{-1}$ . These values illustrate one of the main differences between ESR and NMR, that is, the sensitivity.  $\Delta E$  is  $\sim 600$  times larger for the electron, hence ESR is much more sensitive than NMR; for example, it is possible to record ESR spectra for  $10^{-7} \text{ mol l}^{-1}$  solutions of stable free radicals.

### $g$ -Factors

In a molecule, the unpaired electron may occupy an orbital that may be more or less localized on a single



atom, for example, in the methyl radical the unpaired electron is localized almost entirely in a p orbital at right angles to the plane of the molecule. The electron may also be virtually delocalized over the entire molecule, as in the case of the naphthalene negative ion. The  $g$ -factor can provide information on such delocalization since it is dependent on the spin-orbit coupling:

$$g = L + g_0S \quad [3]$$

where  $g_0 = 2.0023$ ,  $L$  is the total orbital angular momentum, and  $S$  is the total spin angular momentum: for large values of  $L$ ,  $g$ -factors can be as high as 9 or more. The value of  $L$  is particularly important for free radicals containing heavy atoms and for paramagnetic transition metal compounds. For the latter compounds the orbital contributions are very similar to those observed in measurements of magnetic moments of metal complexes. Since the  $g$ -tensor will clearly be anisotropic if  $L$  is not zero, linewidths of solution spectra can be affected and solid-state spectra can depend upon the orientation of the sample.

$g$ -Factors can also be affected by mixing of ground electronic states with excited states that have orbital angular momentum. The extent of the mixing is inversely related to the energy of the excited state, which may be represented by the equation:

$$g = 2.0023(1 - f\lambda/\Delta E) \quad [4]$$

where  $\lambda$  is the spin-orbit coupling constant,  $f$  is a composite numerical factor, and  $\Delta E$  is the separation of the ground and excited electronic states. Equation [4] is especially important for transition metal complexes: for atoms with fewer than five d electrons,  $\lambda$  is positive, it is zero for five electrons, and it is negative for more than five d electrons. The magnitude of  $\lambda$  depends upon the oxidation state of the atom, but it also increases rapidly with increase of atomic number, so the effect is large for the later 3d atoms and very large for the 4d and 5d atoms.

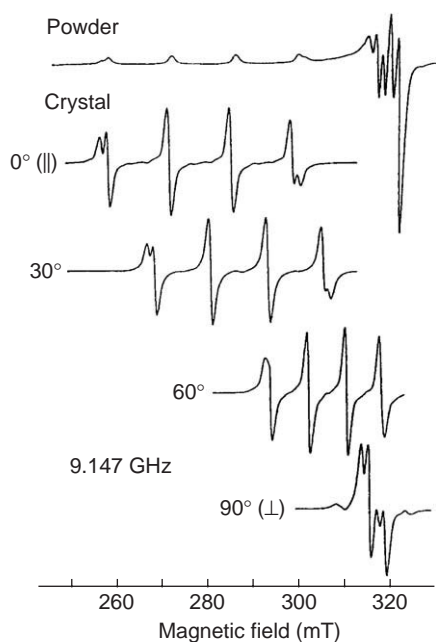
### Fine Structure

Transition-metal complexes can have more than one line in their spectra due to the fact that more than one unpaired electron may feature in the complex. The extra lines arise from zero-field splitting, which depends upon molecular structure, the extent of spin-orbit coupling and other factors. It is common for the zero-field splitting to be very large; in which case, either no resonance is observed or there is only one line in the spectrum.

A complex ESR spectrum is observed when there are hyperfine interactions due to the presence of

magnetic nuclei – the effect is analogous to the nuclear spin-spin interactions found in NMR spectra: in ESR spectroscopy the extra lines are referred to as hyperfine structure. A nucleus of spin value  $I$  gives  $2I + 1$  lines all of equal intensity and separated by the hyperfine coupling constant  $a$ . (It should be noted that extra lines and uneven separation of lines can occur when second-order effects are present: these effects come into play when the magnitude of  $a$  is large and approaches that of the applied magnetic field). The magnitude of  $a$  depends upon the unpaired electron spin density located at the magnetic nucleus in question. For nuclei having a positive magnetic moment, and under the influence of a large unpaired electron spin density,  $a$  is positive; however,  $a$  can be negative if the spin density is small since low spin densities can be negative. Spectra can be very complicated when there is interaction from a number of magnetic nuclei. There are three major factors which control the magnitude of  $a$ . The first is the Fermi contact term, which operates through the s character of the unpaired electron, there being a finite probability of s electrons being located at the nucleus. This term is important for transition metal complexes only when there is low symmetry because the unpaired electrons are formally d electrons and mixing of d and s electrons can only occur when they have the same symmetry. A second factor is the spin-polarization effect in which the unpaired spin density in valence orbitals polarizes the distribution of electrons in the core orbitals. The third factor operates in the solid state – it is the electron-nucleus dipolar coupling. This depends upon the distance between the unpaired electron and the magnetic nucleus and also upon their relative orientations. The orientation of the nuclear spin is determined by the applied magnetic field, while that of the unpaired electron depends upon its orbital character and the orientation of the molecule with respect to the applied magnetic field. It is obvious that dipolar coupling produces anisotropy in the hyperfine interaction leading to the three principal magnitudes  $A_{xx}$ ,  $A_{yy}$ , and  $A_{zz}$  (these are diagonal elements in a matrix): they can be positive or negative. Thus for a solid, the hyperfine interaction  $A$  is made up of two main terms, namely  $a + T$ , where  $a$  is the isotropic hyperfine interaction and  $T$  is the anisotropic hyperfine interaction. In a single crystal, both  $A$  and  $g$  can change with orientation. For powdered solids or frozen solutions, all orientations are represented in the spectrum but the principal values of  $A$  are associated with their corresponding  $g$ -factors. **Figure 1** illustrates the difference between a powder spectrum and a single crystal spectrum of a copper(II) complex diluted with the compound



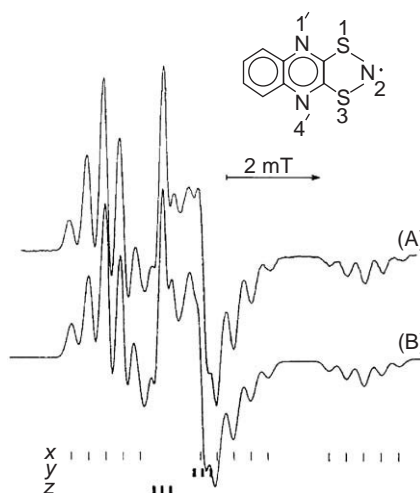


**Figure 1** First-derivative powder and single crystal (at four orientations) X-band ESR spectra of copper(II) in  $\text{CaCd}(\text{CH}_3\text{COO})_4 \cdot 6\text{H}_2\text{O}$ . The powder spectra show that parallel (||) features are absorption-like and they correspond to the  $\theta = 0^\circ$  orientation in the crystal spectra. (Reproduced with permission from Pilbrow JR (1990) *Transition Ion Electron Paramagnetic Resonance*. Oxford: Clarendon Press, Oxford University Press.)

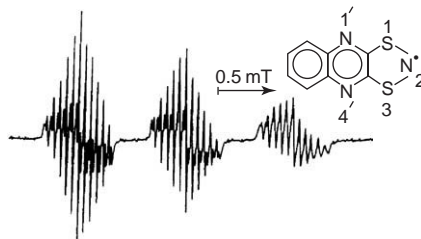
$\text{CaCd}(\text{CH}_3\text{COO})_4 \cdot 6\text{H}_2\text{O}$ : the fine structure comes from two copper nuclides both having a nuclear spin of  $3/2$  and similar magnetic moments. The full detailed structure in a spectrum may not always be resolved and a broadening of lines is observed. For solutions, the dipolar coupling is averaged to zero so that:

$$a_{\text{iso}} = [(a + T_{xx}) + (a + T_{yy}) + (a + T_{zz})]/3 \quad [5]$$

There may be small discrepancies between this average and the true value of  $a_{\text{iso}}$ . Evaluation of  $a$  and  $A$  allows estimations to be made, respectively, of  $s$  and  $p$  spin densities centered upon a particular magnetic nucleus. **Figures 2** and **3** illustrate the different ESR spectra obtained from a solution in the liquid and frozen states of the free radical 1,3,2-dithiazolo-[4,5-*b*]quinoxalin-2-yl (QDTA). The spectrum shown in **Figure 1** is characterized by one value of  $g_{\text{iso}}$ , a large isotropic hyperfine interaction from the nitrogen nucleus at position 2 that gives rise to the three groups of lines: the fine structure within each group comes from the two equivalent nitrogen nuclei at positions 1' and 4' and from the aromatic protons. Thus all the abundant magnetic nuclei in the molecule contribute to the spectrum. The frozen solution



**Figure 2** First-derivative X-band ESR spectrum of 1,3,2-dithiazolo-[4,5-*b*]quinoxalin-2-yl in toluene- $d_6$ :acetone- $d_6$  (4:1, v/v) at 128 K: (A) is the experimental spectrum and (B) is the computer-simulated spectrum.



**Figure 3** First-derivative X-band ESR spectrum of 1,3,2-dithiazolo-[4,5-*b*]quinoxalin-2-yl in toluene:acetone (4:1, v/v) at 222 K.

spectrum shown in **Figure 3** (both the experimental spectrum and the computer-simulated spectrum are given) is characterized by three anisotropic  $g$ -factors and by the corresponding anisotropic hyperfine interactions,  $A_{xx}$ ,  $A_{yy}$ , and  $A_{zz}$  for the nitrogen nucleus at position 2 and for the nitrogens at positions 1' and 4'. From the markers at the foot of **Figure 3** it may be seen that the latter pair of nitrogen nuclei have negligible hyperfine interactions in the  $y$ - and  $z$ -directions. In fact, the major interaction for all the nitrogens is in the  $x$ -direction, which is at right angles to the plane of the molecule, and this tells us that nitrogen  $p_x$  orbitals are responsible for the interaction. Because hydrogen nuclei have negligible anisotropic interactions they do not feature in the solid-state spectrum.

### Linewidths

The theoretical width of an ESR line is  $\sim 0.01$  mT and this may be approached in solutions of free radicals as shown in **Figure 2**. However, various relaxation

processes can shorten the lifetime of an excited state and cause line broadening. Electron spin–lattice and spin–spin relaxation times can be very short, which is why pulse methods are difficult to implement in ESR spectroscopy and why spectra of solid-state transition metal complexes are often broad and poorly resolved. It will be noted that the line-heights of the three main groups of lines in **Figure 2** are different or in other words, the lines are broader towards high fields. This is due to the rate of tumbling of the radical being insufficiently rapid to average out the anisotropic contributions. Other molecular dynamics processes can cause alternate lines in the spectrum to be broadened – this is known as the alternating linewidth effect. There are a wide variety of such processes which can affect the linewidths in these ways and they include electron spin exchange, electron transfer, proton exchange, ion association, and molecular inter-conversion. Though these effects contribute to the complexity of ESR spectra, analysis of the data they provide can yield very valuable structural and kinetic information.

## Instrumentation

The microwave radiation used in ESR is most commonly of frequency  $\sim 9\text{--}10\text{ GHz}$  [i.e.,  $(9\text{--}10) \times 10^9\text{ s}^{-1}$ ] – the so-named ‘X-band’ – but other frequency ranges are sometimes used. When a different choice of microwave frequency is made, a different strength of the external magnetic field is required to achieve the resonance condition. This follows from eqn [1]. Since the innate sensitivity of ESR depends on the excess of spins in the ground state, which relates to the Boltzmann factor, it is clear that measurements made at lower microwave frequencies will afford an intrinsically weaker signal. There are, however, very good reasons for undertaking so-called ‘multi-frequency’ experiments. For example, in the study of large biological or aqueous samples (including isolated organs, such as the liver, and whole, small animals), a frequency of  $\sim 1.5\text{ GHz}$  (L-band) is often beneficial. This is because the absorption of microwave radiation by water becomes increasingly attenuated on reduction of its frequency. It is also important to be aware of the ‘dimensions’ attendant to the experiment: namely the size of the resonant cavity. At X-band, the frequency ( $\nu$ ) is  $\sim 9.5 \times 10^9\text{ s}^{-1}$ , which corresponds to a wavelength ( $\lambda$ ) of 3.2 cm. In order to set up a resonant (standing) wave, the cavity must also be of this dimension. At L-band, since  $\nu = 1.5 \times 10^9\text{ s}^{-1}$ , the size of the cavity is now 20 cm (8 inches!), and so a far larger sample

might be accommodated. Allowing, overall, for the loss of sensitivity caused by the reduced Boltzmann factor at the lower frequency, and the compensation for which is possible with a much larger sample volume, a working rule is that an L-band experiment is of around a factor of ten lower in sensitivity than the one at X-band. Similar reasoning applies to measurements made at higher frequencies. For example, at W-band,  $\nu = 9.4 \times 10^{10}\text{ s}^{-1}$ . In this case,  $\lambda = 0.32\text{ cm}$  ( $\sim 3\text{ mm}$ ), and the sample is usually contained in a tube of overall diameter  $\sim 1\text{ mm}$ , with an internal diameter of  $\sim 0.6\text{ mm}$ . While it is true that the experiment is ideal for (highly expensive) biological materials, available only in small quantities, and that the greatly increased Boltzmann factor will largely offset the loss of sensitivity from such a small sample, the actual signal intensity is limited by problems of magnetic field inhomogeneity and component construction. The commonly encountered microwave frequencies, their wavelength equivalents, and representative attendant magnetic fields required for resonance, are all listed in **Table 1**, although we note that ESR experiments have been done at frequencies up to  $\sim 500\text{ GHz}$ .

Multi-frequency methods are generally resorted to in an effort to improve the resolution of a given ESR spectrum. In the solid state, it is sometimes unclear whether an apparent splitting arises from different g-components, or is due to hyperfine coupling. By increasing the frequency, the separation of g-components is increased, but a hyperfine splitting will be unaffected. Measurements made over a range of frequencies are also sometimes useful in achieving a more optimal separation of spectral features in transition metal complexes, and may also provide a useful range of ‘time-windows’ in studies of molecular reorientation (e.g., of spin-labels or spin-probes).

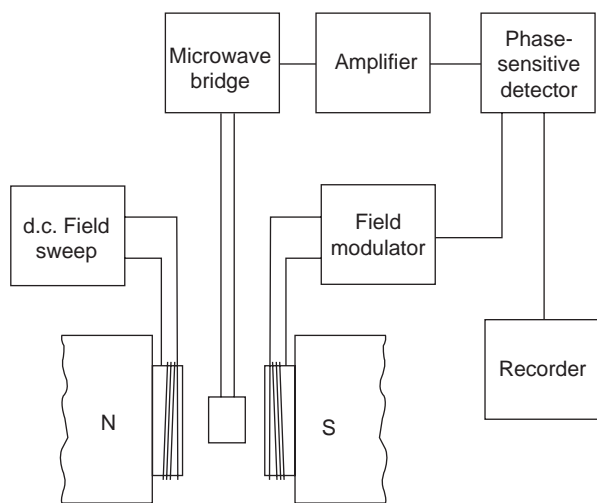
Pulse methods dominate over (CW) methods in the instrumentation used in NMR spectroscopy, while the reverse is true for ESR. The problem in adopting pulse methods for the latter arises for two main reasons: (1) a typical ESR spectrum of a free radical occupies  $\sim 90\text{ MHz}$ , and (2) spin–lattice relaxation

**Table 1** Frequency ranges, wavelengths, and representative resonant magnetic fields for different microwave bands

Band	$\nu$ (GHz)	$\lambda$ (cm)	$B$ (G)
L	1.5	20	500
S	3	10	1000
X	9.5	3.2	3400
K	24	1.3	8500
Q	35	0.8	12500
W	94	0.3	33500

times of free radicals are of the order of microseconds. Thus to obtain a complete ESR spectrum from a single pulse would require very large microwave powers and digitization of the spectrum in 1 ns. Some of the problems can be circumvented and commercial pulse ESR spectrometers are available but they are not as ubiquitously useful as their NMR counterparts.

A block diagram of a simple CW ESR spectrometer is shown in Figure 4. The basic components are a frequency source (klystron or Gunn diode), an electromagnet (having a field of  $\sim 330$  mT for X-band), and a resonant cavity (or loop gap resonator) in which to place the sample. The microwave source and the crystal detector are normally contained within one unit known as the microwave bridge. The microwaves are introduced into the sample cavity via a waveguide and the reflected radiation returns to the microwave bridge. The magnetic field is swept rapidly with a set of small coils mounted on the cavity for oscilloscope display of the signal, or the main field is swept slowly for pen recorder presentation. In order to improve the signal-to-noise ratio it is customary to modulate the microwave frequency at 100 kHz and detect at this frequency with a phase-sensitive detector. The result of doing this is to produce first-derivative absorption spectra as exemplified by Figures 1–3. Other components of an ESR spectrometer can include equipment for controlling the temperature of the sample from 4 to  $\sim 1000$  K, a field-frequency lock, NMR gaussmeter (for accurate magnetic field measurement), microwave frequency counter, data acquisition system, and computer control of the entire operation of the spectrometer.



**Figure 4** Block diagram of a simple continuous-wave ESR spectrometer.

For almost all ESR studies, high-purity silica is used to contain the sample and to fabricate the Dewar flask placed in the sample cavity for variable-temperature work. The major advantage of silica is that it is free from the paramagnetic iron and vanadium impurities found in borosilicate glasses: there is an additional advantage in that it helps to focus the microwave field in the sample cavity and thus the sensitivity is improved. High-purity silica does not readily generate paramagnetic defects when the sample is to be irradiated with ionizing radiation such as  $\gamma$ -rays. Flat cells or capillaries are used for lossy samples. Special cells are available for *in situ* electrochemical experiments and for rapid and stopped flow measurements. Most sample cavities have a 'window' to allow photochemical experiments to be carried out. The highest radical solution concentration, which is advisable to use, is  $10^{-4} \text{ mol l}^{-1}$  in order to avoid exchange broadening: concentrations for solids can be as high as  $10^{-2} \text{ mol l}^{-1}$ , which is low enough to prevent line broadening from dipole-dipole interactions. For single crystal work it is obvious that, as seen in Figure 1, a diamagnetic host crystal has to be used. For high-resolution studies of solutions, such as the example shown in Figure 2, it is essential to remove oxygen to reduce the effects of paramagnetic line broadening.

Most ESR spectroscopy is concerned with stable paramagnetic materials but, because of the high sensitivity of the technique, it can also be used to study short-lived species having half-lives as short as 1  $\mu\text{s}$ . Free radicals having half-lives in the millisecond region can be observed by flowing the reagents rapidly through a mixing chamber in the resonant cavity. Radicals which can be generated photochemically lend themselves to study with high-speed pulses from a laser.

## Applications

Several broad categories can be highlighted:

1. Transition metal complexes. These are studied for their intrinsic interest, and also in biological systems such as metalloproteins and in catalysts, for example when they are incorporated into zeolites.
2. Organic free radicals. There are many basic studies on these including structural investigations, their role as reaction intermediates, mechanistic studies, and studies of molecular dynamics. New free radicals are being synthesized for use in spin labeling and for use as spin probes. There is a tremendous amount of effort being put into using spin labels in biological work; they are used for

example to study the structure of membranes. Organic free radicals are produced when synthetic polymers undergo degradation.

3. Free radicals are formed when most materials are subjected to ionizing radiation, thus ESR can be used to monitor the process. In the food industry such monitoring is extremely important and ESR can assist in direct radical concentration measurement and also in providing a means of measuring the radiation dose. Radiation effects on rocks can be used to measure ages up to one million years.
4. Naturally occurring free radicals are found in all living systems and ESR provides a useful measure of these, for example the radical concentration in normal cells has been found to be higher than in cancer cells. Oxygen is responsible for the formation of some radicals in living systems, and ESR can be used to measure its concentration (oximetry). Spin traps are stable diamagnetic compounds which react rapidly with unstable free radicals to give a long-lived species that can be examined by ESR. The spin trapping technique is useful for studying a wide variety of chemical reactions as well as those of biological interest.

See also: **Electron Spin Resonance Spectroscopy: Specialized Techniques**; Biological Applications.

## Further Reading

- Berliner LJ and Reuben J (eds.) (1989) *Biological Magnetic Resonance: Spin Labelling Theory and Applications*, vol. 8. New York: Plenum.
- Bramley R and Strach SJ (1983) Electron paramagnetic resonance at zero magnetic field. *Chemical Review* 83: 49–82.
- Chung Y-L, Fairhurst SA, Gillies DG, *et al.* (1992) ESR spectroscopy of some dicyano-, pyrazino- and quinoxalino-1,3,2-dithiazolyl radicals. *Magnetic Resonance Chemistry* 30: 774–778.
- Hoff AJ (ed.) (1989) *Advanced EPR Applications in Biology and Biochemistry*. Amsterdam: Elsevier.
- Landolt-Börnstein *Numerical Data and Functional Relationships in Science and Technology*, new series, group II, vols. 1, 2, 8, 9a, 9b, 9c<sub>1</sub>, 9c<sub>2</sub>, 9d<sub>1</sub>, 9d<sub>2</sub>, 10, 11, and 12. Berlin: Springer.
- Weil JA, Bowman MK, Morton JR, and Preston KF (eds.) (1987) *Electronic Magnetic Resonance of the Solid State*. Ottawa: Canadian Society for Chemistry.
- Wertz JE and Bolton JR (1986) *Electron Spin Resonance: Elementary Theory and Practical Applications*. New York: Chapman and Hall.

## Specialized Techniques

**C J Rhodes**, Reading, UK

**H Morris**, Liverpool John Moores University, Liverpool, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Shirley Fairhurst, pp. 1017–1022, © 1995, Elsevier Ltd.

## Introduction

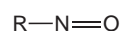
Most specialized electron spin resonance (ESR) techniques are largely instrumental but there are several techniques that are in widespread use that only require a standard ESR spectrometer for their application – two of these are spin trapping and spin labeling. Another important technique that is coming into use in biological work is oximetry. Details of the three techniques are given later.

Most of the specialized techniques involve the use of fairly sophisticated instrumentation; these can be divided roughly into two categories: continuous wave (CW) and pulse methods. Electron–nuclear double resonance (ENDOR) and ESR imaging will be discussed in the CW category, while a variety of

techniques (including pulse ENDOR) will be presented in the pulse section.

## Spin Trapping

A limitation to the application of ESR spectroscopy to the study of reactive intermediates in reactions is the difficulty of producing a sufficient quantity of free radicals for direct detection. The introduction of the technique of spin trapping has provided a simple and effective means of overcoming this difficulty. A diamagnetic compound (a spin trap) is introduced into the radical-producing system to give a relatively stable ESR-observable free radical (spin adduct). In favorable cases, the original free radical can be identified from the *g*-factor and hyperfine coupling constants of the spin adduct. The method allows studying free radicals that are being produced at low stationary-state concentrations (e.g., by  $\gamma$ -radiolysis) over a long period. Typical spin traps are nitroso compounds (1) or nitrones (2)



(1)

R is an alkyl or aryl group



example to study the structure of membranes. Organic free radicals are produced when synthetic polymers undergo degradation.

3. Free radicals are formed when most materials are subjected to ionizing radiation, thus ESR can be used to monitor the process. In the food industry such monitoring is extremely important and ESR can assist in direct radical concentration measurement and also in providing a means of measuring the radiation dose. Radiation effects on rocks can be used to measure ages up to one million years.
4. Naturally occurring free radicals are found in all living systems and ESR provides a useful measure of these, for example the radical concentration in normal cells has been found to be higher than in cancer cells. Oxygen is responsible for the formation of some radicals in living systems, and ESR can be used to measure its concentration (oximetry). Spin traps are stable diamagnetic compounds which react rapidly with unstable free radicals to give a long-lived species that can be examined by ESR. The spin trapping technique is useful for studying a wide variety of chemical reactions as well as those of biological interest.

See also: **Electron Spin Resonance Spectroscopy: Specialized Techniques**; Biological Applications.

## Further Reading

- Berliner LJ and Reuben J (eds.) (1989) *Biological Magnetic Resonance: Spin Labelling Theory and Applications*, vol. 8. New York: Plenum.
- Bramley R and Strach SJ (1983) Electron paramagnetic resonance at zero magnetic field. *Chemical Review* 83: 49–82.
- Chung Y-L, Fairhurst SA, Gillies DG, *et al.* (1992) ESR spectroscopy of some dicyano-, pyrazino- and quinoxalino-1,3,2-dithiazolyl radicals. *Magnetic Resonance Chemistry* 30: 774–778.
- Hoff AJ (ed.) (1989) *Advanced EPR Applications in Biology and Biochemistry*. Amsterdam: Elsevier.
- Landolt-Börnstein *Numerical Data and Functional Relationships in Science and Technology*, new series, group II, vols. 1, 2, 8, 9a, 9b, 9c<sub>1</sub>, 9c<sub>2</sub>, 9d<sub>1</sub>, 9d<sub>2</sub>, 10, 11, and 12. Berlin: Springer.
- Weil JA, Bowman MK, Morton JR, and Preston KF (eds.) (1987) *Electronic Magnetic Resonance of the Solid State*. Ottawa: Canadian Society for Chemistry.
- Wertz JE and Bolton JR (1986) *Electron Spin Resonance: Elementary Theory and Practical Applications*. New York: Chapman and Hall.

## Specialized Techniques

**C J Rhodes**, Reading, UK

**H Morris**, Liverpool John Moores University, Liverpool, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Shirley Fairhurst, pp. 1017–1022, © 1995, Elsevier Ltd.

## Introduction

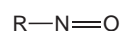
Most specialized electron spin resonance (ESR) techniques are largely instrumental but there are several techniques that are in widespread use that only require a standard ESR spectrometer for their application – two of these are spin trapping and spin labeling. Another important technique that is coming into use in biological work is oximetry. Details of the three techniques are given later.

Most of the specialized techniques involve the use of fairly sophisticated instrumentation; these can be divided roughly into two categories: continuous wave (CW) and pulse methods. Electron–nuclear double resonance (ENDOR) and ESR imaging will be discussed in the CW category, while a variety of

techniques (including pulse ENDOR) will be presented in the pulse section.

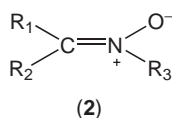
## Spin Trapping

A limitation to the application of ESR spectroscopy to the study of reactive intermediates in reactions is the difficulty of producing a sufficient quantity of free radicals for direct detection. The introduction of the technique of spin trapping has provided a simple and effective means of overcoming this difficulty. A diamagnetic compound (a spin trap) is introduced into the radical-producing system to give a relatively stable ESR-observable free radical (spin adduct). In favorable cases, the original free radical can be identified from the *g*-factor and hyperfine coupling constants of the spin adduct. The method allows studying free radicals that are being produced at low stationary-state concentrations (e.g., by  $\gamma$ -radiolysis) over a long period. Typical spin traps are nitroso compounds (1) or nitrones (2)



(1)

R is an alkyl or aryl group



$R_1$  and  $R_2$  can be an aryl group or a proton and  $R_3$  can be a *tert*-butyl group.

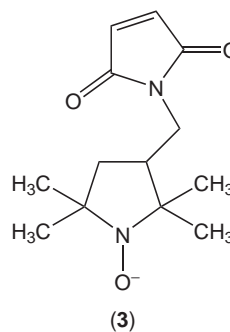
The reacting radical attaches itself to the nitrogen atom in a nitroso compound and to the unsaturated carbon in the nitron. The most commonly used nitroso compound is 2-methyl-2-nitrosopropane (MNP), while the most commonly used nitron is phenyl *N*-*tert*-butylnitron (PBN). When a methyl radical is trapped by MNP the resulting ESR spectrum is a 1:1:1 triplet of 1:3:3:1 quartets. However, if PBN is substituted for MNP the hyperfine structure is not indicative of the attached radical and its identity has to be deduced from hyperfine splittings originating from the nitrogen and lone proton already present in PBN. A nitron that is more sensitive to the structure of the incoming radical is 5,5-dimethyl-1-pyrroline-*N*-oxide. Nitrones have advantages over nitroso compounds in that they are thermally and photochemically more stable. Other disadvantages of nitroso compounds are (1) they have a tendency to form dimers that are inert toward radical trapping and (2) they are unreliable in applications involving oxygen-centered radicals. There is no such thing as a perfect spin trap; hence, it is necessary to be able to choose from a wide variety (more than 100 compounds have been proved to be suitable) for a particular application. One aid to the selection of a suitable trap is the availability of rate constants for many spin-trapping reactions.

Great care has to be taken in interpreting the results from spin-trapping experiments and they have to be augmented with thorough 'blank' experiments because it is possible to generate free radicals from spin traps in reactions in which free radicals are not involved.

## Spin Labeling

The principle of spin labeling is very simple – a macromolecule is attached to a stable free radical and the increase in molecular mass of the radical reduces its rate of tumbling, which can be easily monitored by ESR spectroscopy. The most suitable class of free radicals for the purpose are nitroxyls because of their stability and because the anisotropic  $g$  and  $A$  tensors are sufficiently different in the  $x$ -,  $y$ -, and  $z$ -directions to provide high sensitivity to a change in the rate of molecular tumbling. Rates of tumbling can be assessed, usually by computer simulation, by the change from a completely isotropic spectrum (rates

of  $\sim 10^{10} \text{ s}^{-1}$ ) to a near-powder spectrum (rates of  $\sim 10^6 \text{ s}^{-1}$ ). In order to attach itself easily to a macromolecule the nitroxyl radical has to be provided with a prosthetic group. Example (3) is a nitroxyl designed for attachment to a protein. Should a protein contain a thiol site with which the maleimide group can attach itself, then the ESR spectrum will change from a fast-tumbling to a slow-tumbling type. The method can be made quantitative so that a spin label solution can be 'titrated' against the protein solution. A great deal of skilled synthetic chemistry has been undertaken to provide specific spin labels.



A technique related to spin labeling uses spin probes to monitor molecular motion. In this technique, a radical is synthesized that mimics as closely as possible the structure of the molecule under investigation. Again nitroxyl radicals are the most suitable candidates. An example of their use is in the study of polymer melts: a spin probe is introduced into the polymer and then its rate of tumbling is measured as the polymer changes its viscosity and/or undergoes phase changes.

## Oximetry

In principle, the measurement of oxygen concentration in solution as small as  $10^{-6} \text{ mol l}^{-1}$  is possible. Since ESR linewidths are broadened by the presence of other paramagnetic substances such as oxygen, a radical possessing narrow lines can be used to measure the concentration of the latter. The ubiquitous nitroxyl radicals can be used in this application. Narrow lines can be achieved by keeping the number of magnetic nuclei to a minimum and by using deuterium instead of protium for those that cannot be avoided: the hyperfine splittings for deuterium are about one-sixth those of protium as a result of their different magnetogyric ratios.

Another approach is to use a solid that is permeable to oxygen and has a single very sharp line. This can be achieved extremely well with small crystals of



lithium phthalocyanine that, in the absence of oxygen, have a peak-to-peak linewidth less than 0.002 mT. This is so narrow that most commercial ESR spectrometers do not have sufficient stability to take full advantage. A less good, but much cheaper, material is a species of coal called fusinite.

## Quantitative ESR ('Q-ESR')

Obtaining reliable measurements of ESR signal intensities is extremely tricky, and it almost appears that all vital aspects of the ESR experiment act in conspiracy against managing this successfully! For a start, the sensitivity of the measurement depends acutely on the position of the sample in the microwave cavity, and is greatest at a single point at the cavity center. On moving the sample away from this point, in any direction, the ESR signal intensity is found to decrease. The sample tube can be constrained by means of a 'collet', so that movement only up and down in the cavity is possible. When the sample is a liquid (and therefore is homogeneous), potential problems in its positioning may be circumvented by using a completely filled sample tube that extends through the full length of the cavity. In this case, the exact position of the sample has far less influence on the signal intensity, provided that the sample tube is homogeneous in its diameter and wall thickness. For many solid materials, it is not possible to provide a 'homogeneous', cavity-length sample. Therefore, the sample position needs to be determined precisely. Various devices have been devised for this purpose; the simplest being merely to draw a calibration mark on the sample tube, so that its penetration depth through the collet and into the cavity is known and is consistent. For a series of samples, a set of matched tubes is required, each calibrated to the same depth, which ideally brings the sample to the cavity center, in order to maximize sensitivity.

For quantitative comparisons between samples, they should be as closely matched in composition as possible. This applies to the use of standard samples against which may be measured unpaired electron concentrations: these are more usually referred to as 'spin concentrations', and are quoted as the number of 'spins' (unpaired electrons) per gram of sample. So, at the very least, the sample tubes should all be filled to the same depth, and at best, the standard should consist of the same material as the sample. Since this is rarely possible, a more practical alternative strategy is to match the materials as closely as possible in their electromagnetic properties.

Finally, of course, the operating conditions of the ESR spectrometer, e.g., gain, modulation, and microwave power level, should be the same for both sample and standard, although the signal intensity response to the 'gain' is probably sufficiently linear to permit extrapolation from one sensitivity range to another, as is required when the two specimens have widely different spin concentrations. Strictly, the spin concentration should be determined against a standard, and, in principle, any paramagnetic material of known spin concentration might be used. In practice, 1,1-diphenylpicrylhydrazyl is widely used.

The spin concentration of the standard is determined from the quantity (number of molecules) present, and a corresponding 'area' under the absorption peak is obtained by double integration of the first-derivative spectrum. There are programs available for doing this if the spectra are first digitized. Then the sample is run, the spectrum is then also double integrated, and from the relative areas, the spin concentration of the sample may be deduced. This is not always necessary, and good results may be achieved on the basis of relative peak heights alone.

## ENDOR Spectroscopy

ENDOR spectroscopy offers enhanced resolution compared with conventional ESR; for example, isotropic hyperfine interactions as small as 0.004 mT can be measured. This enhanced resolution is achieved partly because the technique lies between ESR and nuclear magnetic resonance (NMR) and also because redundant lines are eliminated from the spectrum: essentially the ESR signal is monitored while sweeping through NMR frequencies. The simplification of the spectrum arises from the fact that each  $a$ -value produces only two lines in the spectrum, irrespective of how many nuclei contribute to that hyperfine coupling constant. Other advantages of the method are:

1. The hyperfine splittings of each nuclide are determined uniquely since the spectrometer is set at the appropriate NMR frequency.
2. Direct analysis of quadrupolar splittings can be made.
3. Spectra of individual radicals in a mixture can be disentangled.
4. Relative signs of hyperfine and of quadrupolar interactions can be determined.

ENDOR methods have been applied to the study of organic radicals in solution, transition-metal

complexes, and metalloproteins such as solids, glasses, frozen solutions, and in single crystals.

CW ENDOR is considered in this section; pulse ENDOR is dealt with below. The instrumentation is usually based around a computer-controlled CW ESR spectrometer and is commercially available. A radiofrequency coil, capable of handling up to 1 kW power, is used to introduce the NMR frequencies. The coil is contained in a special resonant cavity. Various modulation strategies are employed to improve the signal-to-noise ratio. In order to carry out the ENDOR experiment, the spectrometer is set for a given line in the ESR spectrum. The microwave power is increased to just beyond the saturation level, and then the selected NMR frequency is swept. Two NMR transitions are observed at the frequencies:

$$\nu = |\nu_n \pm a/2| \quad [1]$$

where  $\nu_n$  is the NMR frequency for the applied magnetic field of the ESR spectrometer and  $a$  is the ESR isotropic hyperfine coupling constant. Equation [1] applies when  $\nu_n > |a/2|$ ; however, when  $\nu_n < |a/2|$ , the two lines are centered at  $|a/2|$  and are separated by  $2\nu_n$ . **Figure 1A** shows an ESR spectrum characterized by just two  $a$ -values and **Figure 1B** shows the simplification brought about in the ENDOR spectrum. Other experiments can be carried out with an ENDOR spectrometer. In the special and general TRIPLE experiments (electron–nuclear–electron triple resonance), three frequencies are employed. In the special experiment, two NMR transitions belonging to one set of equivalent nuclei are irradiated, while in the general case different sets of nuclei are irradiated (**Figure 1C**). The former experiment gives improved spectra compared with ordinary ENDOR. The general TRIPLE experiment provides a means of determining the relative signs of hyperfine coupling constants (not possible from an ESR spectrum). In the example shown in **Figure 1D**, the hyperfine coupling constants have opposite signs. The ENDOR-induced ESR (EIE) technique is used to distinguish between radicals in a mixture. Here, the NMR frequency is positioned at the top of an ENDOR line and the intensity changes of this line are monitored while the external magnetic field sweeps over the range of the ESR spectrum. The EIE spectrum has the appearance of an absorption spectrum (see **Figure 1E**). With the technique it is even possible to distinguish between different isotopically substituted radicals.

## ESR Imaging

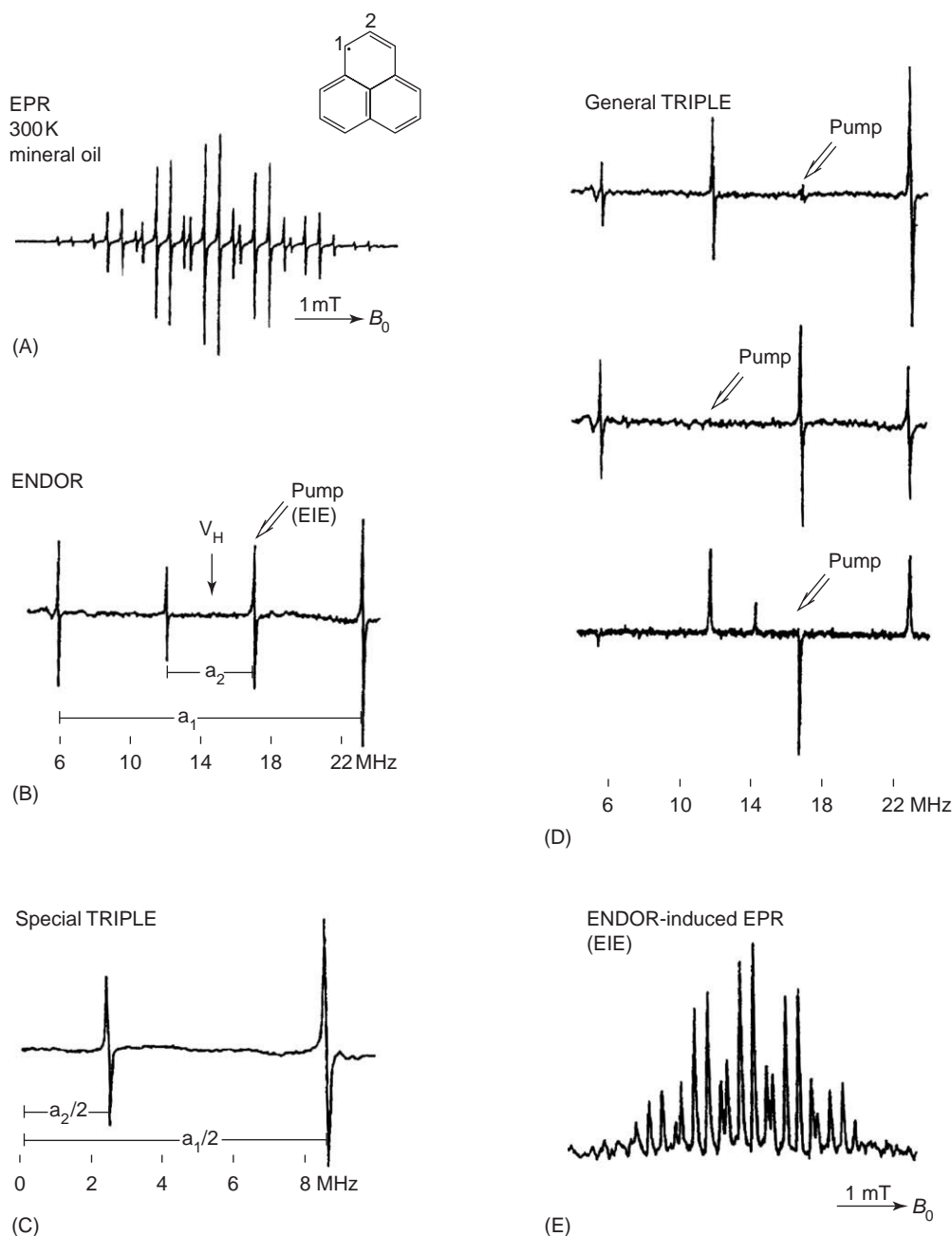
NMR imaging is now a well-established technique for both medical and materials research. The

principles are similar for NMR and ESR imaging – field gradients are used in the  $x$ -,  $y$ -, and  $z$ -directions to allow a volume element to be selected. Difficulties arise in ESR because the field gradients have to be much larger since an ESR spectrum occupies a much larger frequency range than does an NMR spectrum. Another problem arises due to the irradiating frequency used. As stated earlier, the most convenient frequency is  $\sim 9$  GHz, but this only allows a sample of maximum diameter 10 mm to be examined. This is satisfactory for ESR microscopy but is too small for the larger specimens encountered in biological research. For this purpose, frequencies as low as 200 MHz are being used, which permit samples as large as 100 mm to be investigated.

## Pulse Techniques

The basic principles of ESR and NMR are similar and thus one would expect pulse ESR to have a place in magnetic resonance. It does, but it cannot be expected to be as useful as pulse NMR because of practical difficulties. For example, the pulse power required to produce the frequency span of an ESR spectrum is  $\sim 20$  kW. The pulse would last for 1 ns and the signal would have to be digitized in this time. Methods have been devised to overcome these problems, the most common of which are based on spin echoes (echo techniques were developed in the early days of NMR). Echo-detected ESR spectroscopy can be very useful in detecting short-lived radicals generated, say, from a laser flash, where the time resolution is determined solely by the interval between the laser flash and the first spectrometer pulse. Using electron spin echo (ESE) spectroscopy, ENDOR spectra can be obtained from the Fourier transform of the echo-decay envelope. In contrast to CW ENDOR, the ESE envelopes always show deep modulation for the off-principal axis and little or no modulation for canonical orientations; hence, electron spin echo envelope modulation (ESEEM (also known as ESEM)) is a valuable supplement to CW ENDOR. **Figure 2A** shows the spectrum obtained from an ESEEM experiment on a single crystal of nickel(acacen) doped with cobalt(acacen) (acacen is  $N,N'$ -ethylene-bis(acetyl-acetate iminato)). The ENDOR spectrum that is obtained by carrying out a Fourier transform of these data is shown in **Figure 2B**.

The pulse ESR spectrometer is usually based on a standard CW spectrometer because a standard ESR spectrum is often required before carrying out the pulse experiment. The pulse part of the equipment is



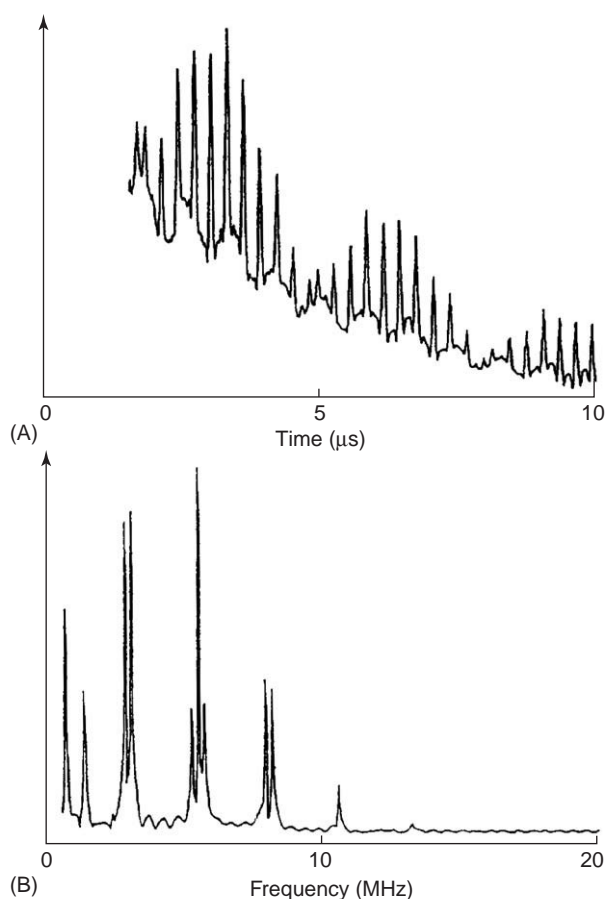
**Figure 1** First-derivative spectra obtained by continuous-wave techniques for the phenalenyl radical: (A) ESR, (B) ENDOR, (C) special TRIPLE resonance, (D) general TRIPLE resonance, and (E) ENDOR-induced ESR. The arrows indicate the positions of the RF pumping frequencies: there are several possibilities in (D). The RF pumping frequency in (E) is set in the same position as shown in (B). (Reprinted from Kurreck H, Kirste B, and Lubitz W (1988) *Electron Nuclear Double Resonance Spectroscopy of Radicals in Solution*. Weinheim: VCH Publishers.)

quite sophisticated in order to overcome some of the problems mentioned above. Some of the components, such as the traveling wave tube amplifier, are expensive; thus, the whole spectrometer can be costly, but no more costly than a medium-priced well-equipped NMR installation.

## Applications

### Spin Trapping

The main use of spin trapping is clearly in identifying radical intermediates in organic and inorganic



**Figure 2** Section of a three-pulse ESEEM for a single crystal of [Ni(acacen)] $0.5\text{H}_2\text{O}$  doped with [Co(acacen)] at 4.2 K. (A) The ESEEM spectrum; (B) the Fourier transform of (A) showing the transition frequencies of the nitrogens and some of the protons. (Reprinted from Schweiger A (1986) *Moderne methodische entwicklungen in der Electronenspinresonanz-Spektroskopie. Chimia* 40: 111–123.)

reactions. The reactions can take place in the gas phase or in solution. The reactions can be thermal, electrochemical, or radiation induced. Probably the most important application of spin trapping is to the study of radicals in biological systems. Some examples of the latter are the effects of ionizing radiation on aqueous solutions of peptides, amino acids, and nucleic acids, and the detection of  $\text{O}_2^{\bullet-}$  and  $\text{OH}^{\bullet}$  radicals in lipid peroxidation.

### Spin Labeling

The method provides a means of studying biochemical and biophysical characteristics of many lipids, proteins, and amino acids. Many investigations have been devoted to the synthesis of radicals having a suitable reactive functional group to couple as selectively as possible and with minimum perturbation of the biological system. The latter requirement

is particularly important when a drug moiety is converted to a spin label for use in detecting the drug–antigen complex. The technique has been developed to provide spin immunoassay and spin membrane immunoassay.

Carbohydrates have been labeled to help elucidate the structures of enzymes, lectins, and epimers, and they have also been used as spin probes for cell membrane structures.

Stable nitroxyl radicals have also been widely employed as spin probes for the study of relaxation and molecular dynamics of synthetic polymers.

### Oximetry

Probably the most important application of the technique is in the study of animals *in vivo*. For the nitroxyl type of oxygen sensor it is necessary to use an imaging technique and then an oxygen concentration map can be constructed. Imaging is not so important when solid sensors are used, as small particles can be injected into the organ of interest. When a crystal of lithium phthalocyanine is injected into a rat heart, the change of oxygen concentration with each heart-beat can be monitored.

### Quantitative ESR

Obtaining a reliable measure of the number of unpaired electrons ('spin concentration') in a sample is often extremely useful. Even reliable relative values measured across a series of samples can often provide useful information. There are various important applications, as may be illustrated by the following incomplete list: ESR dating, the determination of oxidized polycyclic aromatic hydrocarbons and of environmental carbon in samples of ambient air, the influence of air pollution (e.g.,  $\text{SO}_2$  and  $\text{NO}_2$ ) on plants and soils, quantification of  $\text{NO}_2^{\bullet}$ ,  $\text{RO}_2^{\bullet}$ , and  $\text{HO}_2^{\bullet}$  radicals in air samples, radiation dosimetry, redox activities of zeolite catalysts, and the metabolism of spin probes in cells and tissues.

### ENDOR Spectroscopy

Because of its close relationship with NMR spectroscopy, many of the applications involve a wide variety of magnetic nuclei. ENDOR spectroscopy not only yields information about static molecular properties, but it can provide an insight into molecular dynamics in a similar way to conventional ESR and NMR. ENDOR and ESR techniques have timescales of  $10^{-9}$ – $10^{-4}$  s, while that for NMR is  $10^{-5}$ – $10^{-1}$  s. Both ENDOR and ESR have been applied to studies of internal rotation, ring inversions, intramolecular electron transfer, or valence isomerizations.

These effects influence the linewidth if the timescale of the process is of the same order of magnitude as the timescale of the method.

Liquid crystals provide a useful approach to the investigation of quadrupolar nuclei since quadrupole interactions give rise to extra splittings in ENDOR spectroscopy.

ENDOR methods can be especially useful for investigating biological systems because powder-type spectra can yield isotropic and anisotropic coupling constants with a higher resolution and precision than can conventional ESR spectroscopy. Typical examples are the radical ions occurring as primary photoproducts in reaction centers of photosynthetic bacteria. Liquid-state ENDOR measurements have been made on biologically active organic species such as the vitamin quinones.

### ESR Imaging

ESR microscopy has been applied to a variety of problems including studies of defects in diamonds, diffusion of oxygen in model biological systems, radiometric dosimetry, and the distribution of radicals in a variety of solids. Examples include the *in situ* oxidation of coal, rock dating, and the investigation of the swelling of polymers using solvents containing spin probes.

Radiofrequency (and L-band) imagers are largely used for biological research to monitor the transport of spin-labeled reagents in small animals. Oximetry using solid particles also benefits from the spatial resolution derived from imaging techniques. These imagers clearly can be used for large specimens; for example, measuring the porosity of rocks using solutions of spin probes.

### Pulse Techniques

There are several obvious applications that are analogous to those encountered in NMR spectroscopy. Probably the most important of these is the measurement of relaxation times that are of the order of microseconds. In NMR, these times are of great structural value and lead to conclusions about molecular motion. No doubt these benefits will also apply to ESR as commercial pulse spectrometers become more widespread. If the relaxation times of two paramagnetic centers are sufficiently different, their spectra can be separated easily using the

echo-detected ESR method. When rapid reaction kinetics is being measured, information is also gained on transient phenomena such as chemically induced electron polarization.

ESEEM has been applied to the study of surface complexes, coordination of water in metmyoglobin, the structures of ligands in a copper protein, magnetic properties of electronic triplet states, and solitons in polyacetylene.

Pulsed ENDOR avoids some of the difficulties encountered with CW ENDOR in that the entire pulse sequence can usually be made short enough to exclude unwanted relaxation effects. This means that the pulsed ENDOR method can be used at any temperature, provided that an electron spin echo can be detected. There are a variety of pulse ENDOR techniques that have been applied to the investigation of free porphyrin bases and of nitrogen-14 nuclei in a disordered system.

**See also:** **Electron Spin Resonance Spectroscopy:** Principles and Instrumentation. **Geochemistry:** Inorganic; Soil, Organic Components. **Immunoassays:** Overview. **Nuclear Magnetic Resonance Spectroscopy:** Instrumentation.

### Further Reading

- Eaton GR, Eaton SA, and Ohno K (1991) *EPR Imaging and In Vivo EPR*. Boca Raton, FL: CRC Press.
- Fairhurst SA, Gillies DG, and Sutcliffe LH (1990) Electron spin resonance imaging. *Spectroscopy World* 2: 14–21.
- Hoff AJ (ed.) (1989) *Advanced EPR: Applications in Biology and Biochemistry*. Amsterdam: Elsevier.
- Schweiger A (1991) Pulsed electron spin resonance spectroscopy; basic principles, techniques and examples of applications. *Angewandte Chemie International Edition in English* 30: 265–286.
- Törmälä P (1979) Spin label and probe studies of polymeric solids and melts. *Journal of Macromolecular Science – Reviews in Macromolecular Chemistry* C17: 297–327.
- Yorandov ND, Mladenova B, and Petkov P (2002) Studies on the uncertainties in quantitative EPR estimations due to the construction of the cavities used. *Analytica Chimica Acta* 453: 155–162.
- Yordanov ND, Lubenova S, and Sokolova S (2001) On the possibility for separate determination of pyrolyzed products (soot and polycyclic aromatic hydrocarbons) in aerosols by EPR spectroscopy. *Atmospheric Environment* 35: 827–831.



## Biological Applications

**A F Vanin**, Semenov Institute of Chemical Physics,  
Moscow, Russia

**E E H van Faassen**, Universiteit Utrecht, Utrecht,  
The Netherlands

© 2005, Published by Elsevier Ltd.

### Introduction

The application of electron spin resonance (ESR) to biological studies started from the realization that free radicals play a very important role in enzymatic protein catalysis. In the 1950s Barry Commoner in the USA and Lev Blumenfeld in the USSR were the first to begin testing this hypothesis using ESR spectroscopy on tissue and cell preparations. Subsequent studies confirmed the crucial role of paramagnetic metal complexes for the enzymatic activity of biological materials. Two examples are afforded by the process of mitochondrial respiration and the kinetics of the photosynthetic electron transport chain in plants and bacteria. The paramagnetic centers involve flavin and semiquinone radicals, or short-lived cationic or anionic radical intermediates. In addition, Helmut Beinert in the USA discovered in the late 1950s the presence of stable paramagnetic iron–sulfur centers in frozen animal tissues and cell components. Geometrical information on tertiary protein structure can be obtained if ESR spectroscopy is used to measure the distances between two paramagnetic sites in a protein. Such sites may be native (iron–sulfur clusters or metal ions) or artificially incorporated by means of a paramagnetic label in a well-defined location of the protein. Driven by technological advances, the necessary labeling techniques and distance measurements using pulsed ESR and pulsed electron nuclear double resonance (ENDOR) have undergone very rapid progress in recent years.

In recent years, much interest has been focused on the detection of reactive oxygen species (ROS) and nitric oxide (NO) radicals in viable cell cultures and tissues. The spin trapping technique has proved to be very valuable, and its general characteristics have been described elsewhere in this encyclopedia. The complex biochemistry involved requires special techniques for spin trapping ROS and NO in biological materials. Such techniques form the content of the final sections.

Finally, spin labeling of proteins, ESR oximetry, and the use of spin probes for investigation of the structure of biomembranes are classic examples of biological ESR. These topics have been discussed elsewhere in this encyclopedia.

### Structure and Function of Metalloproteins

Enzymes can be considered as highly developed and selective catalysts that control much of the chemistry in living organisms. In nearly all cases, the crucial reaction steps involve a redox reaction mediated by a transition metal ion, for example at the heart of the enzyme. Such electron transfers affect the electronic configuration of the metal ion and the magnetic moment of the ion. Many, though not all, important enzymes have at least one paramagnetic charge state and may be observed in principle using ESR. The ESR spectrum of transition metal ions is remarkably sensitive to the so-called crystal field interactions with the atomic neighbors of the ion (spin–lattice interaction). Their outer electronic states consist of so called d-states, i.e., the five possible states characterized by an orbital angular momentum quantum number  $L=2$ . In an isolated atom, all five orbitals would have the same energy. This degeneracy is partially or completely lifted by the interaction with surrounding ligands and leads to distinct positions of the energy levels. In this way, tetrahedral, octahedral, tetragonal, or square planar surroundings may be easily distinguished. ESR spectra are particularly sensitive to violations of the structural symmetry, such as small distortions of the tetrahedral pyramid or loss of axial symmetry of the metal center. In addition, the ESR spectra may show hyperfine couplings between the electronic spins and the magnetic moments of nearby magnetic nuclei like nitrogen ( $I=1$ ), hydrogen ( $I=1/2$ ), or many transition metal ions like Cu ( $I=3/2$ ), Co ( $I=7/2$ ), or Mn ( $I=5/2$ ). Additional structural information may be obtained by isotopic substitution of natural nonmagnetic nuclei by magnetic isotopes like  $^{57}\text{Fe}$  ( $I=1/2$ , natural abundance 2.2%) or  $^{67}\text{Zn}$  ( $I=5/2$ , natural abundance 4%). A table of the most relevant paramagnetic metal ions may be found elsewhere in the encyclopedia.

### Distance Measurements in Biomolecules

In noncrystalline samples, ESR provides a convenient technique for estimating the distance between paramagnetic sites on a macromolecule. The spatial distance between the sites determines the strength of the spin–spin interaction, which may be experimentally determined using ESR. Such interactions may be



studied in a liquid or frozen solution, glasses, in porous matrices, and in amorphous or crystalline solids. At close distances, the spin–spin interactions are sufficiently strong to affect the shape of continuous wave (CW) ESR spectra. At larger separations these interactions are weak and better resolved using modern pulsed ESR spectroscopy methods like electron spin echo modulation (ESEEM) or pulsed electron–electron double resonance (ELDOR or PELDOR). ESR is very selective as it provides information on only those few sites that carry an electronic magnetic moment. It has proved to be very valuable for macromolecules that cannot be crystallized or that are too large for nuclear magnetic resonance investigations in solution. Although pulsed ESR requires cooling to low temperatures, it can be applied to fully functional proteins embedded in their natural matrix like a lipid cell membrane or attached to a coenzyme. It provides information on protein tertiary structure, conformational changes during enzymatic activity, and the dynamics of protein folding. The paramagnetic sites may be endogenous flavin or semiquinone radicals, cation or anion radical intermediates, or paramagnetic metal centers. Additionally, exogenous spins may be included by site-directed spin labeling. The 3-(methanesulfonylthiomethyl)-2,2,5,5-tetramethylpyrrolidin-1-yloxy spin label (also called methane–thiosulfonate) is a nitroxide that has proved to be particularly useful for selective labeling of cysteine residues. It attaches to the protein backbone via a rigid disulfide bridge, which makes the orientation and mobility of the label reflect those of the backbone itself.

The spin–spin interactions are caused by either dipolar magnetic coupling between the spins or electronic exchange interaction due to orbital overlap between the two unpaired electrons involved. Such electronic exchange may be direct or mediated through hydrogen bonds or bridging ligands between paramagnetic metal centers. In continuous wave ESR, the coupling strength is best determined via the intensity of the so-called half-field transition, which occurs near half the magnetic field for the usual single quantum transitions. This double quantum transition becomes weakly allowed due to the anisotropic part of the spin–spin interaction (dipolar interaction and anisotropic exchange). The practical experimental range extends to distances up to  $\sim 12$  Å for continuous wave ESR and up to 30 Å for pulsed excitation of this double quantum coherence (DQC). Far longer distances are accessible by specialized pulsed ESR methods. Particularly useful are double frequency resonance techniques where two different spin transitions are successively excited by pulse sequences with two microwave frequencies. The

**Table 1** Selection from ESR methods to determine the distance between paramagnetic sites. The acronyms are explained in the text

<i>Technique</i>		<i>Range (Å)</i>	<i>No. of frequencies used</i>
Spectral simulation	CW	4–20	1
Intensity of half-field transition	CW	4–12	1
Intensity of half-field transition	Pulsed	12–25	1
DQC	Pulsed	20–30	1
DEER	Pulsed	15–70	2
PELDOR	Pulsed	15–130	2

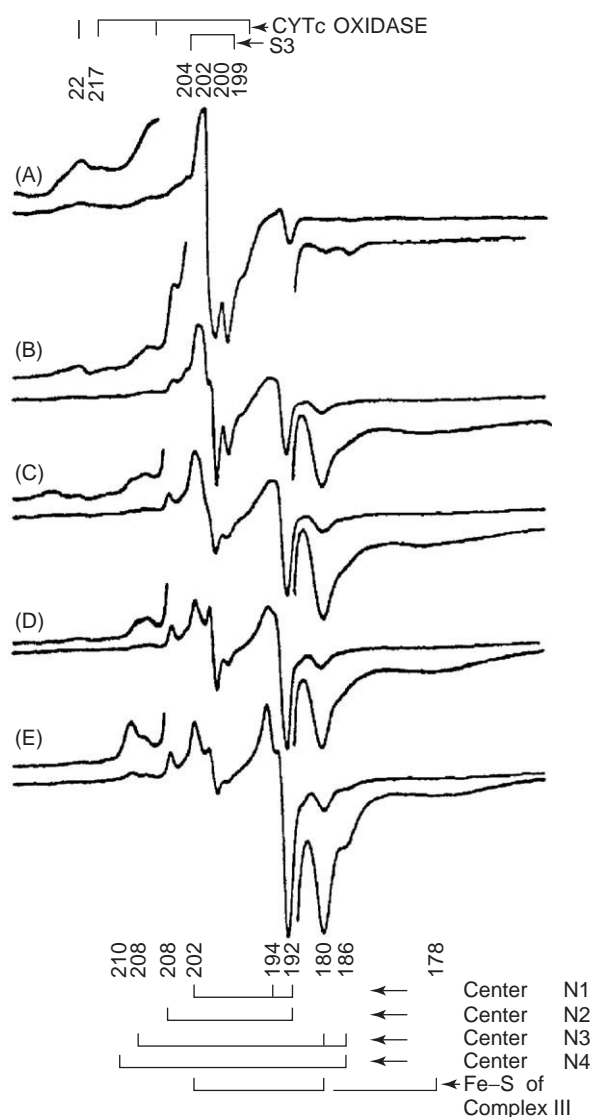
coherent excitation of two different microwave transitions provides a high selectivity that allows the determination of even very weak coupling between distant spins. The basic PELDOR experiment is based on a three-pulse sequence but suffers from experimental limitations like ‘blind spots’ and ‘dead time’. These problems may be reduced in the more sophisticated four-pulse double electron–electron resonance (DEER) spectroscopy.

The most commonly used ESR methods for distance measurement are listed in Table 1.

### ESR-Detectable Endogenous Paramagnetic Centers in Animal Tissues, Cells, and Bacteria

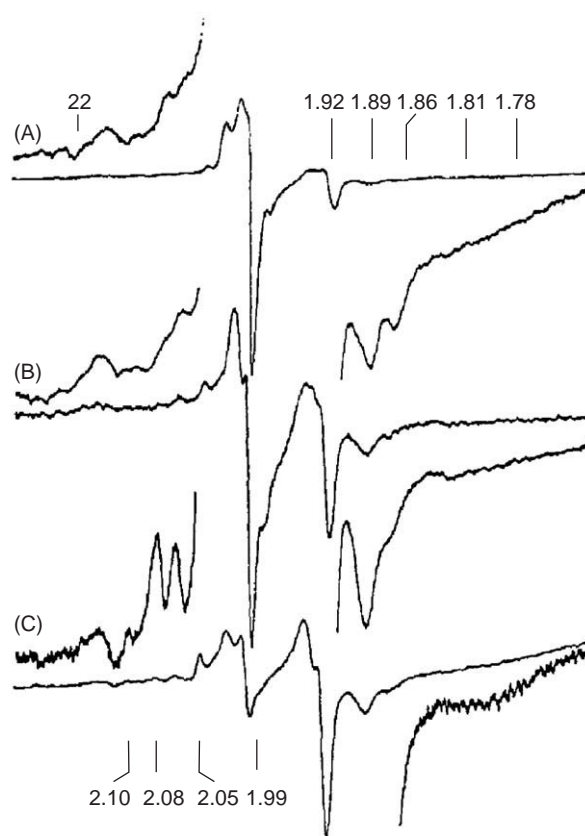
The majority of endogenous paramagnetic centers detected in animal cells and tissues arise in electron transport chains (ETCs) of mitochondria. They comprise heme or nonheme iron-containing proteins (cytochromes and iron–sulfur proteins, respectively),  $\text{Cu}^{2+}$ -containing proteins, protein-based radical centers (flavine mononucleotide or flavine adenine nucleotides) as well as certain radical intermediates of low molecular weight (ubiquinones). Due to intensive spin–lattice relaxation, the ESR spectra of iron-containing protein components of ETC can be recorded only at cryogenic temperatures in the range 4–80 K. Microwave power saturation of the ESR signals should be avoided at such low temperatures.

Figure 1 shows the complex ESR spectra from isolated cardiac mitochondria. They appear as a superposition of spectra from various paramagnetic components of the mitochondrial ETC. They are mainly iron–sulfur centers, denoted as N1, N2, N3 + 4 (located in complex I, NADH–ubiquinone oxidoreductase), S1 (complex II, succinate–ubiquinone oxidoreductase), and the Rieske iron–sulfur protein (complex III, ubihydroquinone–cytochrome C oxidoreductase). The positions of the components



**Figure 1** EPR spectra of electron transport particles (ETPs) from heart mitochondria from titration with NADH. (A) ETPs treated as other samples except that no reductant was added; (B)–(E) reduced with increased dose of NADH. (Reproduced with permission from Orme-Johnson N *et al.* (1974) Electron paramagnetic resonance-detectable electron acceptors in beef heart mitochondria. *Journal of Biological Chemistry* 249: 1922–1939.)

(i.e., the  $g$ -tensor values) are shown in **Figures 1** and **3**. Similar spectra are characteristic of various isolated animal tissues. As shown in **Figures 2** and **3**, the shape of the spectrum depends on the redox state of the paramagnetic centers. Full reduction of the mitochondrial ETC components is obtained upon depletion of oxygen in tissue. In the oxidized state, the ESR spectra appear as superpositions from cytochromes, the copper center in cytochrome oxidase, and a high-potential iron–sulfur center, S3 (complex II). An additional signal arises from the iron–sulfur

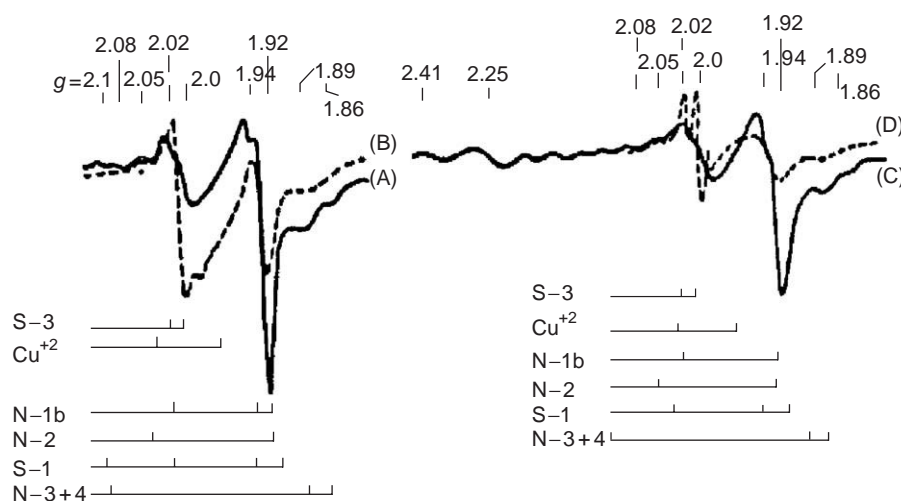


**Figure 2** EPR spectra of whole pigeon heart at different reductive states of the electron carriers (the extent of reduction increases from A to C). (Reproduced with permission from Orme-Johnson N *et al.* (1974) Electron paramagnetic resonance-detectable electron acceptors in beef heart mitochondria. *Journal of Biological Chemistry* 249: 1922–1939.)

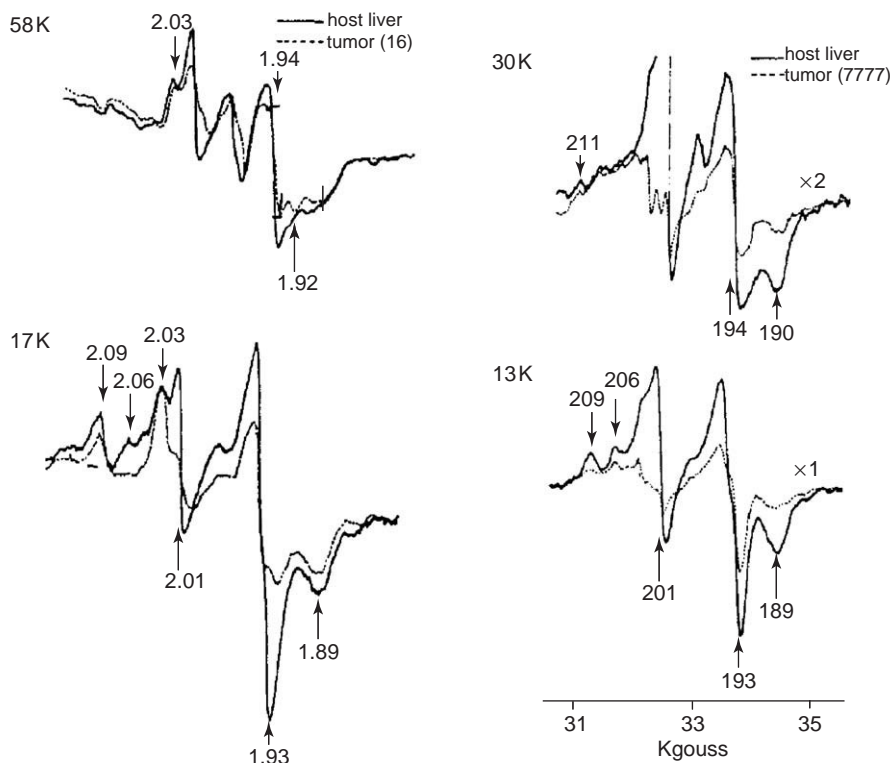
cluster of mitochondrial aconitase, the enzyme from the Krebs cycle.

The ESR spectra also give valuable information on pathologies like cancer. **Figure 4** shows that the iron–sulfur signals from hepatoma mitochondria in Morris rats are correlated with the mitochondrial respiratory activity. In the slow-growing hepatoma-16 tumors, the iron–sulfur signals from reduced nicotinamide adenine dinucleotide (NADH) dehydrogenase are smaller than from surrounding nonaffected liver tissue or liver from healthy rats. In the rapidly growing hepatoma-7777 cells, the ESR signals of iron–sulfur centers are even more diminished. The data show that the abnormal metabolic characteristics of some tumors are related to specific defects in the iron–sulfur components of the ETC.

Recent ESR data suggest that the final step of proton translocation in complex I involves an endoenergetic electronic disproportionation in the ubisemiquinone pair. If tightly coupled and functioning bovine heart submitochondrial particles are snap



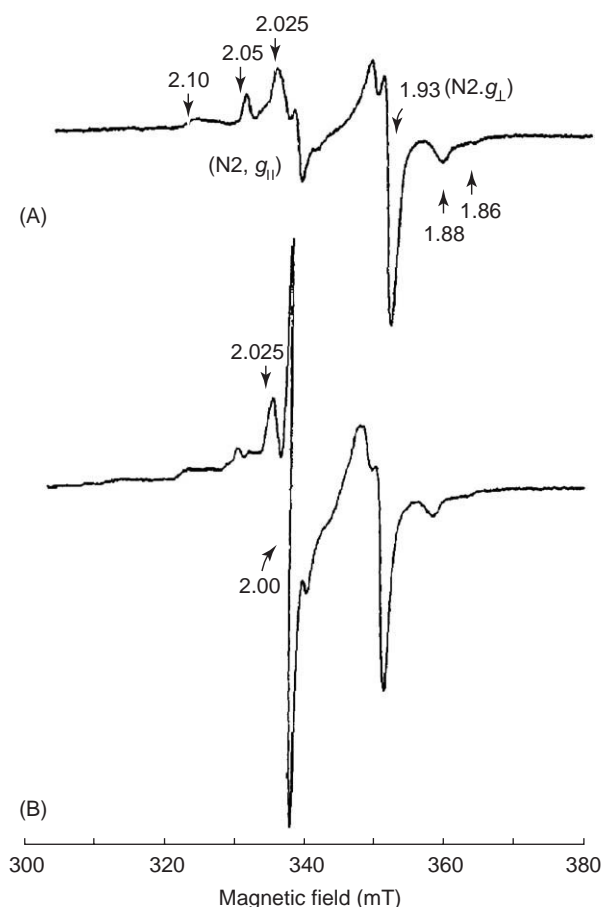
**Figure 3** EPR spectra of isolated tissues of mouse heart (A,B) and liver (C,D). (A,B) The tissues after isolation from the organism; (B,D) after low temperature oxidation (by maintaining isolated tissues at  $-12^{\circ}\text{C}$  for 5 h). (Reproduced with permission from Burbaev D *et al.* (1975) ESR spectra of animal tissues *in vitro*. *Biofizika* (Russian) 20: 1062–1067.)



**Figure 4** EPR spectra of reduced iron-sulfur centers in mitochondria prepared from slow growing Morris hepatoma-16 (left panel) or rapidly growing Morris hepatoma-7777 and host liver (right panel), measured at 58 and 17K, or 30 and 13K, respectively. Mitochondrial suspensions were brought to anaerobiosis by incubating with glutamate + malate for 10 min. (Reproduced with permission from Ohnishi T *et al.* (1973) Electron paramagnetic resonance studies of iron-sulfur centers in mitochondria prepared from three Morris hepatomas with different growth rates. *Biochemistry and Biophysics Research Communications* 55: 372–381.)

frozen in the presence of an NADH substrate, strong ESR signals at  $g=2.00$  show that one ubisemiquinone of this pair is in a radical state (spectra shown in Figure 5). It undergoes fast spin-spin relaxation

due to interaction with the iron-sulfur center, N2, located at a distance of only  $10 \text{ \AA}$ . This ubisemiquinone signal was not detected in uncoupled submitochondrial particles. The data suggest that energy

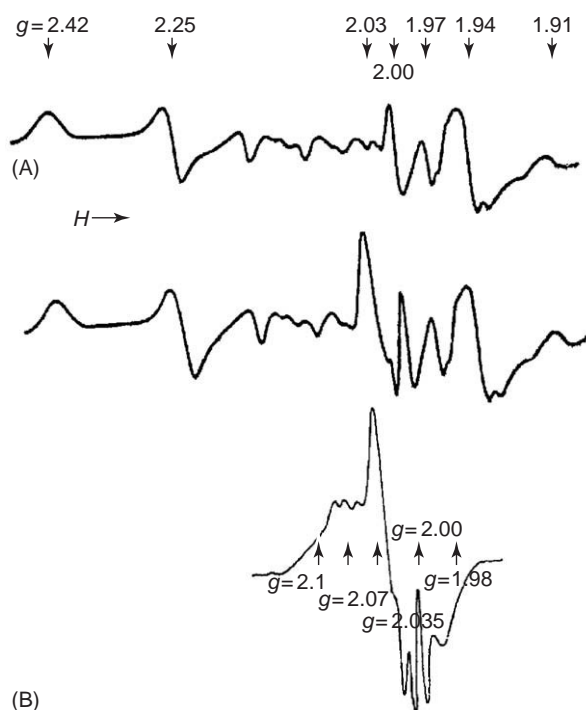


**Figure 5** EPR spectra of uncoupled (A) and tightly coupled (B) submitochondrial particles during steady-state NADH oxidation. Recordings were made with microwave power 2 mW and temperature 16 K. (Reproduced with permission from Vinogradov A *et al.* (1995) Energy-dependent Complex I-associated ubiquinones in submitochondrial particles. *FEBS Letters* 370: 83–87.)

transduction in mitochondria requires the participation of bound ubiquinones as well as the iron-sulfur center N2.

Analysis of these mitochondrial ESR spectra sheds light on the nature, redox state, and quantity of the ETC components. This has been very helpful in understanding the redox behavior of enzymatic centers and in deducing their orientation in mitochondrial membranes. It should be kept in mind that the ESR spectra reflect the ETC components in the frozen state. As such, the spin and redox states of the paramagnetic centers may differ from those of the active enzyme at a physiological temperature.

Oxidized cytochrome,  $P_{450}$ , is the terminal component of the microsomal ETC. Its  $g$ -values at  $g = 2.42$ ,  $2.25$ , and  $1.91$  are easily distinguishable from those of components in the mitochondrial ETC. Spectra from this enzyme are often detected in



**Figure 6** ESR spectra of liver tissue from mice at 77 K: (A) control animal on normal diet; (B) mouse on drinking water with 0.3% nitrite for 7 days. The nitrite consumption induces formation of dinitrosyl-iron (DNIC 0.03) and nitrosyl-heme complexes, at  $g = 1.98$ . (Reproduced with permission from Varich V (1979) Changes in amounts of dinitrosyl non-heme iron complexes in animal tissues depending on animal growth. *Biofizika* (Russian) 24: 344–347.)

isolated liver and kidney tissue at temperatures below 100 K (**Figure 6**). The redox processes in the microsomal ETC ensure detoxication of many xenobiotics as well as the biosynthesis and biodegradation of endogenous compounds like steroids or fatty acids. The detoxication defense of many bacteria, insects, fishes, yeasts, and plants involves a wide range of cytochromes. Like cytochrome  $P_{450}$ , these cytochromes have paramagnetic charge states. ESR spectroscopy has proved to be very useful for studying the catalytic cycle of this class of enzymes. A particular success was the detection of the so-called Compound 1, a highly redox-active intermediate porphyrin radical-cation- $\text{Fe(IV)}=\text{O}$  complex.

Paramagnetic centers are found in many specific enzymes and proteins that function in animal tissues and bacteria (oxygenases, sulfite or nitrite reductases, xanthine oxidase, nitrogenase, etc.) besides mitochondria and microsomes. The catalytic site can include Mo, Cu, Co, Ni, Mn, and other metal ions. However heme- and iron-sulfur centers constitute the majority of the paramagnetic centers found outside the mitochondria.

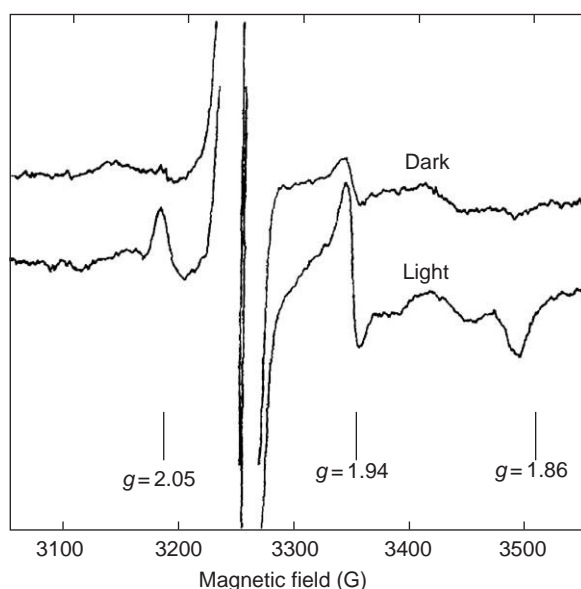
In recent years, the interest in the physiological role of nitric oxide radicals has greatly stimulated the investigation of iron–nitrosyl complexes. The generation of nitric oxide in mammalian tissues, cultured cells, or bacteria is usually accompanied by the formation of paramagnetic mononitrosyl–heme iron complexes and dinitrosyl–nonheme iron complexes (DNICs). The respective ESR signals of the complexes are shown in **Figure 6**. Nitric oxide is recognized now as a signaling and regulatory radical that influences diverse physiological and biochemical processes. In biosystems, it is generated either enzymatically by NO synthases or nonenzymatically (mainly from nitrite). A possible physiological role of the paramagnetic nitrosyl–iron complexes is now being hotly debated and investigated.

Various forms of DNIC may be distinguished according to the structure of the thiolate group that ligates to the central iron atom. The preliminary results suggest that DNICs may be implicated in the redox chemistry of *S*-nitrosothiols. The latter are known to have antioxidant as well as signaling functions in cells and tissues. Recently DNICs have been found in plant leaves upon supplementation of NO from an exogenous source or generated endogenously from nitrite. It is reasonable to suggest that the formation of DNICs from nonheme iron prevents iron precipitation in the form of hydroxide complexes. This would increase the iron availability for various intracellular components, particularly for chloroplasts.

### ESR-Detectable Endogenous Paramagnetic Centers in Photosystems of Plants

ESR spectroscopy has been very useful in studying the photosynthetic function in plants and bacterial photosystems. It had been known for a long time that illumination causes free radical ESR signals in photosynthetic systems, and subsequently the importance of nonheme iron and then manganese complexes was discovered. The photosynthetic activity of plants is based on two supramolecular assemblies, photosystems I and II, which act in tandem. Each photosystem consists of a light harvesting antenna structure containing chlorophyll dye molecules, and a cascade of electron acceptors that act in succession as an electron transport chain for charge separation.

The ESR signals from the photosystems may be observed at cryogenic temperatures and depend sensitively on the state of illumination. **Figure 7** shows the spectra of photosystem I from spinach chloroplasts. Under steady illumination, they appear as a superposition of an intense  $g = 2.00$  free radical line

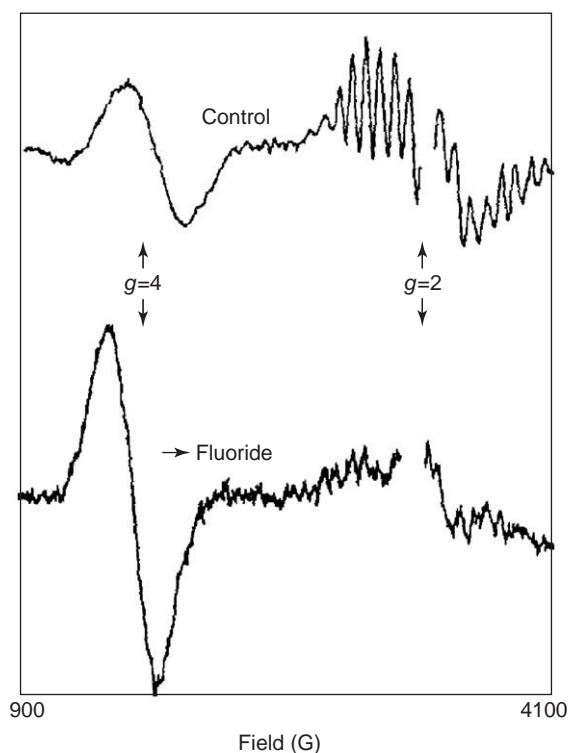


**Figure 7** Photoreduction of a bound Fe–S center in intact spinach chloroplasts after illumination at 77 K. Recordings were made at 10 K. (Reproduced with permission from Malkin R and Bearden A (1971) Primary reactions of photosystem: photoreduction of a bound ferredoxin at low temperature as detected by ESR spectroscopy. *Proceedings of the National Academy of Sciences of the USA* 68: 16–19.)

from oxidized pigment  $P700^+$  and an anisotropic ESR signal at  $g$ -values of 2.05, 1.94, and 1.86 assigned to a reduced Fe–S center. A quantitative comparison of the  $P700^+$  content and paramagnetic Fe–S center, as well as a comparison of their kinetics characteristics at temperatures 10–100 K, has shown that photosystem I has a good correlation between  $P700$  oxidation and the ESR intensity from the Fe–S center. This center turned out to be the secondary electron acceptor from the electron transport chain. The nature of the primary quinone–iron acceptor is revealed using ESR if the electronic pathway to the secondary acceptor is blocked. In the reduced state, this so-called X center has a broad spectrum with  $g = 1.78$ , 1.88, and 2.08.

Photoinduced oxidation of water into dioxygen is the central step in photosynthesis. It is carried out through a characteristic tetranuclear manganese cluster in photosystem II. This cluster is the active site for water oxidation by formation of the so-called oxygen-evolving complex (OEC). Photosynthetic oxidation of water is a highly complex mechanism involving a sequence of five intermediate states,  $S_i$ . It has been proved that the first intermediates,  $S_0$ – $S_3$ , of the Mn–OEC donor complex are sufficiently stable to be detected in frozen samples using ESR. For example, **Figure 8** shows the spectrum of the  $S_2$  state. The signal is mainly accepted to arise from a  $S = 1/2$

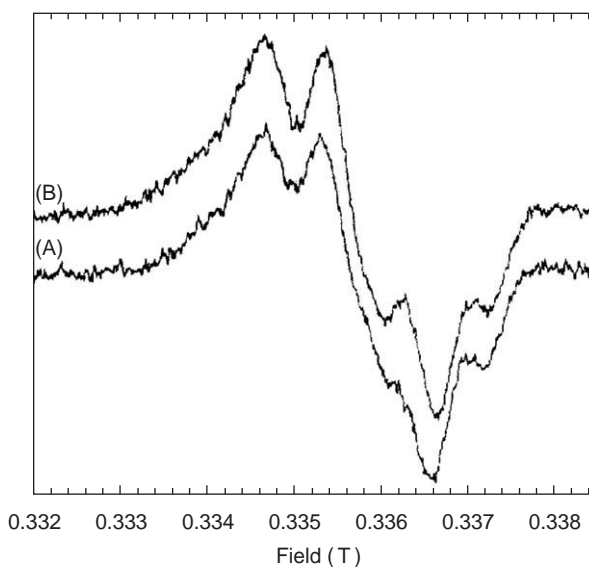




**Figure 8** EPR spectra associated with the oxygen-evolving center in photosystem II preparations: the multiline EPR signal and the signal at  $g = 4.1$  (top panel) from the preparation in the  $S_2$  state (illumination at 195 K); EPR signal (shown at the bottom) produced by continuous illumination of a sample at 200 K in the presence of  $20 \text{ mmol l}^{-1}$ , 20 mM fluoride. Recordings were made at 10 K. A narrow ESR line from a spurious free radical near  $g = 2.0$  was removed from the spectra. (Reproduced with permission from Yachandra V, Sauer K, and Klein M (1996) Manganese cluster in photosynthesis: Where plants oxidize water to dioxygen. *Chemical Reviews* 96: 2927–2950.)

antiferromagnetically exchange-coupled  $\text{Mn}_4(\text{III}, \text{IV}_3)$  cluster. A multiline structure is located near  $g = 2.00$  and shows well-resolved hyperfine couplings from the manganese nuclei. The additional low-field signal at  $g = 4.1$  is suggested to arise from some clusters existing in a high-spin  $S = 3/2$  or  $S = 5/2$  electron configuration. The  $g = 4.1$  signal can also be formed through infrared illumination of the  $S = 1/2$  state at 65 K. It is also observed in photosystem II upon exposure to  $\text{F}^-$  anions. Such samples lose the multiline structure near  $g = 2.00$  as well as the capacity to generate  $\text{O}_2$ .

Interestingly, the halide anion  $\text{Cl}^-$  is an essential cofactor for  $\text{O}_2$  production through photosystem II. The anion may be replaced by  $\text{Br}^-$  with full retention of activity. Moreover, either  $\text{Cl}^-$  or  $\text{Br}^-$  is required for the multiline structure near  $g = 2.00$ , which is characteristic of the  $S_2$  state. It strongly suggests that the halide anion is a ligand of the Mn atoms in the cluster, but up to now attempts to detect



**Figure 9** ESR spectra of free tyrosyl radical TyrD photogenerated in Mn-depleted photosystem II at 4 K (A) and warming to 200 K (B). The spectral change near 0.3362 T is attributed to a thermally activated deprotonation of the  $\text{Y}_D^+$  radical. (Reproduced with permission from Faller P, Goussias C, Rutherford AW, and Un S (2003) Resolving intermediates in biological proton-coupled electron transfer: A tyrosyl radical prior to proton movement. *Proceedings of the National Academy of Sciences of the USA* 100: 8732–8735.)

the hyperfine interaction from the halide ion have failed.

The Mn cluster is not directly photoactivated itself but via an indirect redox process involving a tyrosine intermediate. Upon photoactivation of the primary donor in photosystem II, the excited electron is injected into the electron transport chain and removed. The oxidized primary donor,  $\text{P680}^+$ , receives an electron from a nearby tyrosine  $\text{Y}_Z$ . The resulting  $\text{Y}_Z^+$  radical intermediate oxidizes the tetranuclear Mn–OEC to a high valence state that is capable of oxidizing water. This short-lived  $\text{Y}_Z^+$  radical intermediate has ms lifetime and has been studied with pulsed ESR and optical spectroscopy. In addition, illuminated photosystem II contains a long lived  $\text{Y}_D^+$  tyrosyl radical, which is stable at cryogenic temperatures. The ESR spectra reveal a thermal relaxation of the molecular structure around this  $\text{Y}_D^+$  radical at 80 K (see Figure 9). The spectral changes were attributed to a thermally induced deprotonation of the  $\text{Y}_D^+$  cation.

## Spin Trapping of Nitric Oxide

NO radicals are paramagnetic with a degenerate ground electronic state, which precludes direct ESR detection except at cryogenic temperatures in the



frozen state. When bound with various protein amino acid residues, it forms paramagnetic centers that may be observed even at room temperature. The ESR spectra resemble those from stable nitroxyl radicals like 2,2,6,6-tetramethyl piperidino oxy nitroxide (TEMPO). As these centers are susceptible to oxidation by superoxide, their concentrations usually remain below the detection threshold in biosystems. Spin trapping of NO molecules provides a good alternative. Various derivatives of dithiocarbamate ligands are known to enhance considerably the affinity of ferrous ions for NO molecules. The straightforward binding of NO to the Fe(II)–(dithiocarbamate)<sub>2</sub> complex is often referred to as a trapping reaction of NO. The adduct is a paramagnetic mononitrosyl-iron complex (MNIC) with dithiocarbamate ligands that can be detected and evaluated using ESR spectroscopy even at ambient temperature.

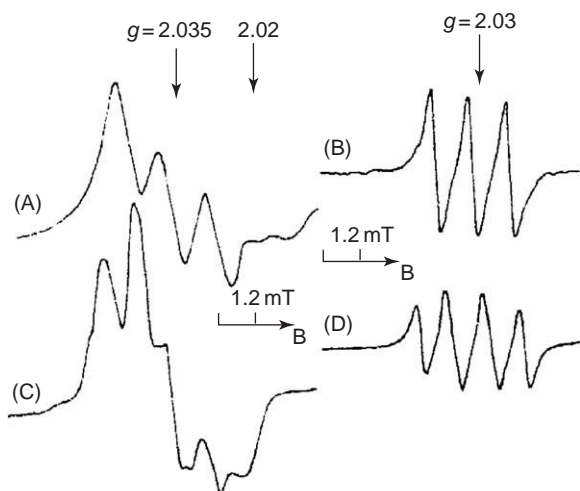
Figure 10 shows the spectra of MNICs in the frozen and liquid solution states, respectively. The triplet structure is caused by hyperfine interaction of the unpaired electron with the nitrogen nucleus ( $I=1$ ) of the NO ligand. The moderate magnitude of this hyperfine splitting demonstrates that the unpaired electron is largely transferred from the NO to the iron atom, which must have a formal monovalent

redox state of Fe(I). Depending on the hydrophobicity of the dithiocarbamate ligands, the MNIC complexes localize preferably in either the hydrophobic or the hydrophilic compartments of cells of tissues. Widely used dithiocarbamates include diethyldithiocarbamate (DETC) and *N*-methyl-D-glucamine dithiocarbamate (MGD), which give rise to hydrophobic and hydrophilic MNICs, respectively. After supplementation of free iron and dithiocarbamate ligands, such MNICs are detectable in many biosystems using ESR spectroscopy, from cultured viable mammalian cells and tissues to living plants (Figure 11). The NO trapping and MNIC formation proceeds *in vivo*, whereas the actual ESR detection and quantification of MNICs is usually carried out *ex vivo*. Recent advances in sensitivity have even allowed MNIC detection in small living mammals. In combination with ESR imaging, MNIC even allows NO tomography revealing the distribution of MNICs in the body of a small animal.

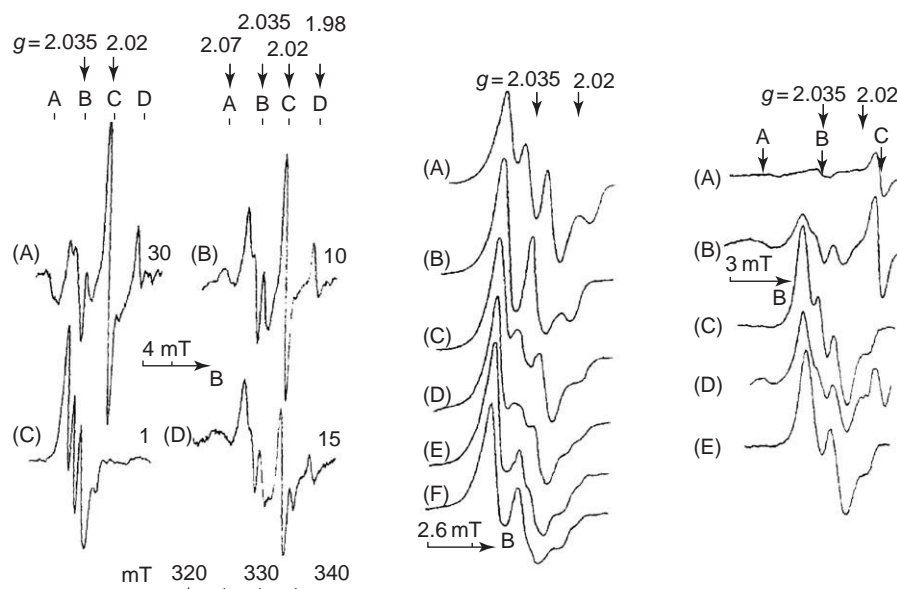
ESR spectroscopy has provided the first unequivocal proof that NO synthesis from L-arginine using NOS enzymes is the only relevant source of NO in mammals: MNIC formation can be prevented by infusion of NOS inhibitors. In addition, isotopic labeling of L-arginine with <sup>15</sup>N ( $I=1/2$ ) isotopes leads to the formation of MNICs with a doublet structure instead of the usual nitroxide triplets obtained with the natural <sup>14</sup>N isotope (Figure 11). In this context, it should be mentioned that modest quantities of NO may also be released from nitrite via nonenzymatic pathways under conditions of acidosis.

It is noteworthy that Fe–dithiocarbamate traps are normally used in the millimolar range. As such, MNIC formation will provide the dominant reaction pathway for endogenous NO molecules under normal physiological conditions. In particular, we can neglect the alternative pathway involving superoxide anions that is present in living systems at a concentration of not more than a few micromoles. Nevertheless, as will be discussed below, the presence of superoxide has a big effect on the MNIC lifetime. Experimental MNIC spectra obtained in tissues usually have to be corrected for the presence of an additional broad background caused by paramagnetic Cu(II)–DETC complexes. DETC is a sufficiently good chelator to scavenge spurious free copper and even extract Cu(II) ions from certain endogenous enzymes like superoxide dismutase (SOD).

The interpretation of NO trapping experiments using Fe–dithiocarbamate complexes is particularly complicated by the redox activity of both Fe–dithiocarbamate complexes and their nitrosyl adducts. The Fe(II)–dithiocarbamate complexes are easily



**Figure 10** EPR spectra of MNIC–MGD complexes in aqueous solution, including <sup>56</sup>Fe (A and B) or <sup>57</sup>Fe (C and D). Recordings were made at 77 K (A and C) or ambient temperature (B and D). Isotopic substitution of <sup>56</sup>Fe ( $I=0$ ) with <sup>57</sup>Fe ( $I=1/2$ ) results in changes in the signal shape that are due to the manifestation of an additional HFS induced by interaction of the unpaired electron with the <sup>57</sup>Fe nucleus ( $I=1/2$ ). EPR spectra from MNIC–DETC. (Reproduced with permission from Mikoyan *et al.* (1997) Complexes of Fe<sup>2+</sup> with diethyldithiocarbamate or *N*-methyl-D-glucamine dithiocarbamate as traps of nitric oxide in animal tissues: Comparative investigations. *Biochimica Biophysica Acta* 1336: 225–234.)



**Figure 11** Left panel: EPR spectra of MNIC-DETC complexes from a mouse liver preparation from control animal injected with DETC (A), an animal injected with bacterial lipopolysaccharide (LPS) (4 h) + DETC (B), an animal injected with LPS (4 h) + DETC and Fe-citrate complex (C). (D) EPR spectra of a mouse spleen preparation from an animal injected with LPS (4 h) + DETC and Fe-citrate complex. The MNIC-DETC complex gives an EP signal at  $g_{\perp}=2.035$ ,  $g_{\parallel}=2.02$  and triplet HFS at  $g_{\perp}$ . A,B,C, and D indicate the quartet HFS of the EPR signal due to the  $\text{Cu}^{2+}$ -DETC complex formed in the liver. The EPR signal with the components at  $g=2.07$  and  $1.98$  in (D) is due to a nitrosyl-heme iron complex. Recordings were made at 77 K. Middle panel: The shapes of the EPR signal of MNIC-DETC including  $^{14}\text{NO}$  (A) and  $^{15}\text{NO}$  (B) only, or at various ratios of these ligands:  $^{15}\text{NO}$ :  $^{14}\text{NO}$  = 3:7 (C), 1:1 (D), 2:1 (E), 4:1 (F). Recordings were made at 77 K. Right panel: EPR spectra of macrophages ( $10^7$  cells per sample) stimulated by LPS for 0 (A), 5 (B), and 11 h (C-E) and subsequently incubated with DETC,  $\text{FeSO}_4$ , and LPS for 2 h. ( $^{15}\text{N}^{\text{G}}$  L-arginine) was present in (D) and (E) for the last 5 h of incubation. Spectra were recorded at 77 K. (A-C) indicate the position of three low field HFS components from  $\text{Cu}^{2+}$  (DETC) $_2$  complexes.  $g=2.035$  and  $2.02$  indicate the position of  $g_{\perp}$  and  $g_{\parallel}$  of the  $\text{NOFe}^{2+}$  (DETC) $_2$  complex. (Reproduced with permission from Kubrina L *et al.* (1992) EPR evidence for nitric oxide production from guanidino nitrogen of L-arginine in animal tissues *in vivo*. *Biochimica Biophysica Acta* 1099: 223-237 and Vanin A *et al.* (1993) The relationship between L-arginine-dependent nitric oxide synthesis, nitrite release, and dinitrosyl-iron complex formation by activated macrophages. *Biochimica Biophysica Acta* 1177: 37-42.)

**Table 2** Spectroscopic properties of iron(II) and iron(III)-MGD complexes in aqueous solutions<sup>a</sup>

Iron-carbamate complex	Absorption wavelength (nm)	Extinction coefficient ( $\text{l mol}^{-1}$ )	Color
Fe(II)-MGD	—	—	Clear, diamagnetic
Fe(III)-MGD	340, 385, 520	20 000, 15 000, 3 000	Orange-brown, paramagnetic ( $S=3/2$ )
NO-Fe(II)-MGD	314, 368, 450	18 000, 12 500, 5 600	Deep green, paramagnetic ( $S=1/2$ )
NO-Fe(III)-MGD	No peaks	~20 000 at 350 nm	Yellow, diamagnetic

<sup>a</sup>Data adapted with permission from Vanin A *et al.* (2000) Redox properties of iron-dithiocarbamates and their nitrosyl derivatives: Implications for their use as traps of nitric oxide in biological systems. *Biochimica Biophysica Acta* 1474: 365-377.

oxidized to Fe(III)-dithiocarbamate complexes by dissolved oxygen. In an aqueous solution, the rate constant of the reaction for Fe(II)-MGD complexes was determined as  $5 \times 10^5 \text{ l mol}^{-1} \text{ s}^{-1}$  at ambient temperature. Superoxide radicals react considerably faster with a rate of  $3 \times 10^7 \text{ l mol}^{-1} \text{ s}^{-1}$ . The Fe(III)-(dithiocarbamate) $_3$  complexes can also bind NO molecules, which results in the formation of diamagnetic iron(III) nitrosyl derivatives. However, the latter are rather unstable and slowly transform into paramagnetic MNIC-dithiocarbamate under anaerobic

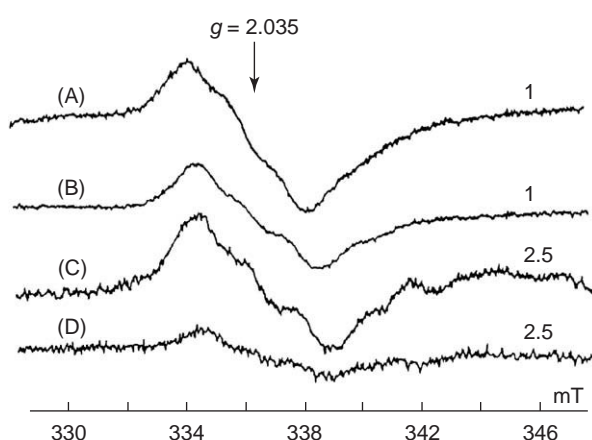
conditions. This transformation may be accelerated by addition of reducing agents like ascorbate. Interestingly, the presence of the NO ligand seems to facilitate this reduction step as the reduction of the iron from the Iron(III) to the Iron(II) state is considerably slower in the absence of NO. The redox conversion may be followed spectroscopically using ESR as well as optical absorption as the redox state and the presence of the NO ligand strongly affect the color of the iron complex. Table 2 lists the optical and ESR properties.

We stress that ESR spectroscopy only detects the paramagnetic versions of MNIC. As such, the ESR intensity may only reflect the NO levels in tissues or biomaterials if all the iron is reduced to the Iron (II) state. Three pathways for this transformation are known in biomaterials. The first is straightforward reduction of diamagnetic MNICs by endogenous reducing agents like glutathione and ascorbate. Second is a mechanism of reductive nitrosylation similar to that proposed for nitrosyl ferrihemoproteins, where the NO ligand reduces the iron to the ferrous state, followed by release of  $\text{NO}^+$  from the complex. Nevertheless, it should be kept in mind that these mechanisms do not ensure complete reduction of all complexes to the iron(II) state.

An additional problem concerns the stability of paramagnetic MNICs in biomaterials. It has recently been discovered that paramagnetic MNICs react rapidly with superoxide and peroxynitrite. The reaction products are ESR-silent nitroso complexes. In other words, the adducts are formed but do not survive. This complication can be partially redressed by the so-called ABC method. Its basic idea is to overwhelm these alternative pathways by bolus injection of a known quantity of exogenous MNIC to a cultured cells or animal tissue. It acts as a very efficient scavenger of superoxide anions and peroxynitrite, so that endogenously formed MNIC adducts survive. This experiment is repeated in the presence of a NO synthase inhibitor. The endogenous NO levels are estimated from the difference between these two experiments: first experiment =  $A + B - C$ , second experiment =  $A - C$ , where  $A$  is the signal from the exogenous MNIC bolus injection and  $B$  is the signal due to endogenous NO production. The quantity  $C$  accounts for the endogenous MNIC adducts transformed into the ESR-silent state by superoxide and/or peroxynitrite.

Figure 12 illustrates the advantage of the ABC method over the 'conventional' method in cultured endothelial BEND3 cells. The ABC method allows detection of a larger fraction of endogenously produced NO. As such, it gives a better representation of NO levels in biomaterials, tissues, and living systems. Still, we cannot exclude the possibility that an unknown fraction of NO ends up in ESR-silent nitroso compounds.

Organic spin traps are also used for NO detection in chemical and biological systems. For example, 2,5-dimethylhexadiene, when introduced into degassed organic solutions of NO, yields a three-line ESR signal, tentatively assigned to 2,2,5,5-tetramethyl-1-pyrroloxy. Paramagnetic nitronyl nitroxides change their ESR spectrum when they react with NO to form a paramagnetic imino nitroxide adduct.



**Figure 12** EPR spectra of MNIC-DETC complexes formed in suspensions of viable endothelial cells. (A) In the absence of NOS inhibitor, *N*-nitro-L-arginine (NLA); Endogenous and exogenous MNIC-DETC make contributions to the signal. (B) With  $1 \text{ mmol l}^{-1}$  NLA. Only exogenous MNIC-DETC makes a contribution to the signal. (C) Computed difference (A-B), showing the endogenous MNIC-DETC formed NO synthesis. (D) Reference spectrum showing MNIC-DETC complexes formed endogenously by conventional NO trapping. Spectra were recorded at 77 K. The relative gain settings are shown at the right side of the spectra ABC method. (Reproduced with permission from Vanin *et al.* (2001) Antioxidant capacity of mononitrosyl-iron-dithiocarbamate complexes: Implications for NO trapping. *Free Radical Biology and Medicine* 30: 813–824.)

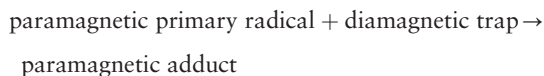
For example, PTIO or carboxy-(2-(*p*-carboxy)phenyl-4,4,5,5-tetramethylimidazoline 3-oxide-1-oxyl) reacts efficiently with NO to form paramagnetic carboxy-(2-(*p*-carboxy)phenyl-4,4,5,5-tetramethylimidazoline 3-oxide). However, in viable cells and tissues, such adducts are easily transformed into diamagnetic hydroxylamines. This property severely limits the ESR application of these traps in bioassays. Potentially valuable information may be obtained from the class of fluorescent cheletropic traps. These have weak red fluorescence and react with NO to form nonfluorescent paramagnetic adducts. The latter are rapidly transformed into a diamagnetic hydroxylamine with strong blue fluorescence, which is easily observable using fluorescence microscopes and spectrometers. Nevertheless, iron-dithiocarbamate complexes form the basis of the most valuable and reliable ESR technique for detecting NO in living tissues.

## Spin Trapping of Oxygen Radicals

Singlet oxygen, triplet oxygen, ozone, superoxide anions and the hydroxyl radicals are all fairly reactive oxygen species. Singlet oxygen is diamagnetic and unobservable using ESR, whereas the remaining

species are paramagnetic. They cannot be observed directly using ESR due to line broadening or the short lifetimes of the species. In biological materials, ozone does not occur in significant quantities and will not be considered further. Superoxide and hydroxyl radicals have been recognized as major and inseparable agents in the development of physiological pathologies. In biological research, these radicals always appear in the company of the nonradical hydrogen peroxide, to which they are linked via a complex series of chemical reactions (dismutation and iron-mediated Fenton chemistry are discussed elsewhere). Evolution has developed elaborate enzymatic defenses to reduce the *in vivo* levels of superoxide and hydrogen peroxide, in the form of superoxide dismutase and catalase, respectively. Nevertheless, minute quantities of oxygen radicals may escape the defenses and contribute to undesirable oxidative stress and oxygen toxicity.

The superoxide anion  $O_2^-$  is modestly reactive when compared with most other radicals but still is a stronger one-electron reductant and oxidant than  $O_2$  itself. This  $S = 1/2$  radical can be observed using ESR only in exceptional cases where the spin relaxation is reduced sufficiently. In biological systems, its concentration remains too low and the linewidth too broad for observation. In the late 1960s, nitroso and nitron compounds were introduced as spin trapping agents to detect superoxide radicals. The traps are diamagnetic, but their reaction with the primary radical leads to the formation of a more stable paramagnetic reaction product in the form of a nitroxide (adduct in spin trapping terminology):

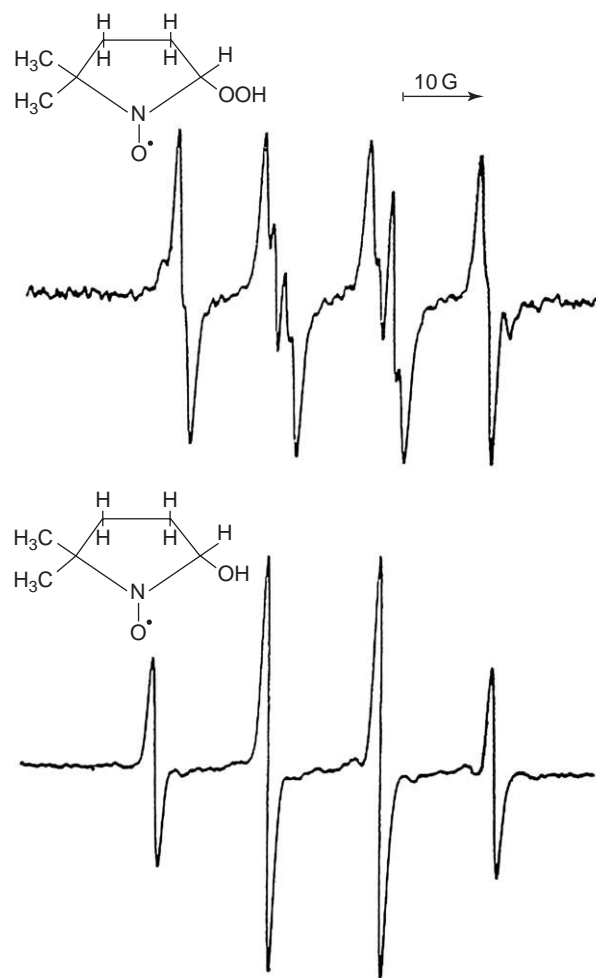


The ideal spin trap should be stable, with rigorously stable adducts. The adduct spectra should depend distinctively on the primary radical to allow unambiguous identification of the latter. In reality, the adducts are susceptible to hydroxylation into a diamagnetic hydroxylamine that is ESR silent. This hydroxylation is a serious complication for radical detection in biological systems, particularly in blood samples.

Compared with nitrones, nitroso compounds like 2-methyl-2-nitrosopropane (MNP) incorporate the superoxide radical closer to the nitroxide center. Therefore, the adduct spectra provide more information on the nature of the primary radical, making its identification easier. But nitroso traps and their adducts are more susceptible to thermal, photochemical, and hydroxylative degradation as well as -ene addition. They have limited solubility and tend to

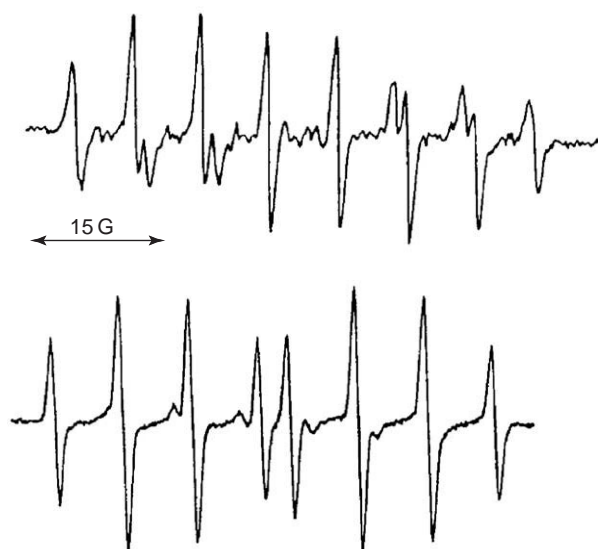
dimerize, leaving only a small concentration of the monomer, which alone is capable of the trapping reaction. Some traps like MNP are sufficiently volatile to be lost from solutions through degassing via nitrogen bubbling.

Wider applications include nitron spin traps, which are light insensitive, nontoxic, and lead to reasonably robust adducts. The adduct spectra do not distinguish between different primary radicals as these get inserted at the  $\beta$ -position further away from the nitroxide center. Phenyl-*tert*-butylnitron (PBN) and 5,5-dimethylpyrroline (DMPO) are most widely used for superoxide detection in hydrophobic and hydrophilic compartments of biological samples, respectively. DMPO is the standard of choice for application in cell cultures as it readily crosses cell membranes. The spectrum of its superoxide adduct, DMPO-OOH, is shown in Figure 13. The interpretation of the trapping experiments is complicated by the poor stability of DMPO-OOH,



**Figure 13** Room temperature spectra of DMPO-OOH adduct (top) and DMPO-OH adduct (bottom) in liquid solution.





**Figure 14** Room temperature spectra of DEPMPO-OOH adduct (top) and DEPMPO-OH adduct (bottom) in liquid solution.

which spontaneously dismutates into the very same DMPO-OH adduct that results from the trapping of the hydroxyl radical. Usually, additional experiments in the presence of effective superoxide scavengers like SOD are needed before the primary radical can be identified unambiguously. It should be noted that the rates of the trapping reaction depend sensitively on the pH. For DMPO, the reaction rate with protonated superoxide exceeds that of the superoxide anion by nearly three orders of magnitude.

The recently developed nitron 5-diethoxyphosphoryl-5-methyl-1-pyrroline-*N*-oxide (DEPMPO) leads to adducts with better stability. Additionally, its decomposition does not lead to potentially misleading DEPMPO-OH adducts. This important advantage has a price: the trap cannot penetrate cell membranes, and the ESR spectra appear as the spectral superposition of two isomers in varying concentrations (Figure 14).

Usually, spin traps exhibit fairly small reaction rates with superoxide, of the order of  $1\text{--}20\text{ l mol}^{-1}\text{ s}^{-1}$  at room temperature. Therefore, fairly long incubation times ( $\sim 1\text{--}30\text{ min}$ ) and high trap concentrations ( $10\text{--}50\text{ mmol l}^{-1}$ ) are required to accumulate sufficient adducts and to capture a substantial fraction of the superoxide in the assay. Recently, traps with higher reaction rates have become available that can be applied at high micromolar concentrations. For

blood, the highly reactive 1-hydroxy-3-carboxy-2,2,5,5-tetramethylpyrrolidine (CP-H) has been applied, in spite of a serious susceptibility to artificial adduct formation via secondary pathways (spurious metal ions, peroxyxynitrite, nonoxygen radicals).

In contrast to superoxide, the hydroxyl radical is one of the most reactive substances known. It reacts rapidly with nearly every substance at rates limited only by the time needed to approach its reagent diffusively. As such, it may become difficult to detect by spin trapping in biological materials, where many alternative reaction pathways compete with the trapping reaction. DMPO has proved to be useful in biological systems because the stability of the DMPO-OH adduct comfortably exceeds that of the adducts formed with PBN of  $\alpha$ -(4-pyridyl 1-oxide)-*N*-*tert*-butylnitron (4-POBN) spin traps.

**See also:** **Electron Spin Resonance Spectroscopy: Principles and Instrumentation; Specialized Techniques. Nitric Oxide.**

## Further Reading

- Beinert H and Albracht SPJ (1982) Iron-sulfur proteins. New insight and unresolved problems. *Biochimica Biophysica Acta Reviews on Bioenergetics* 683: 246–277.
- Berliner LJ (ed.) (1998) Spin labeling: The next millennium. In: *Biological Magnetic Resonance*, vol. 14. New York: Plenum.
- Berliner LJ, Eaton SS, and Eaton GR (eds.) (2000) Distance measurements in biological systems by EPR. *Biological Magnetic Resonance*, vol. 19. New York: Kluwer Academic.
- Berliner LJ and Swartz HM (eds.) (2000) *In vivo* EPR: Theory and application. In: *Biological Magnetic Resonance*, vol. 18. New York: Kluwer Academic.
- Henry Y, Guissani A, and Ducastel D (1996) *Nitric Oxide Research from Chemistry to Biology: EPR Spectroscopy of Nitrosylated Compounds*. Austin: Landes Bioscience.
- Palmer G (2000) Electron paramagnetic resonance of metalloproteins. In: Que L (ed.) *Physical Methods in Bioinorganic Chemistry*. Sausalito: University Science Books.
- Rosen GM, Britigan BE, Halpern HJ, and Pou S (1999) *Free Radicals: Detection and Spin Trapping*. New York: Oxford University Press.
- Schweiger A and Jeschke G (2001) *Principles of Pulse Electron Paramagnetic Resonance*. Oxford: Oxford University Press.
- Vanin A, Huisman A, and van Faassen E (2002) Iron di-thiocarbamate as spin trap for nitric oxide: Pitfalls and successes. *Methods in Enzymology* 359: 27–42.

# ELECTROPHORESIS

## Contents

### Overview

### Principles

### Isotachopheresis

### Isoelectric Focusing

### Electrophoresis in Ionic Liquids

### Polyacrylamide Gels

### Two-Dimensional Gels

### Affinity Techniques

### Blotting Techniques

### Carbohydrates

### Clinical Applications

### Nucleic Acids

### Proteins

## Overview

**R Stringer**, Royal Liverpool University Hospital, Liverpool, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Many substances exist in solution as electrically charged species. Once ionized, cations (+) or anions (−) can be separated in an electrical field where they will migrate toward the cathode or anode depending on their charge (Figure 1). This is the basis of electrophoresis (Gr. 'electro': electric, and 'phoresis': to carry).

The rate at which ions move toward the attracting electrode is dependent on the balance between the

impelling force of the electric field on the charged ion and the frictional and electrostatic retarding effects (i.e., the drag) between the sample and the surrounding medium. The degree of the drag depends on the size and shape of the molecule and on the viscosity of the matrix. An estimate of the velocity of a molecule in an electric field, in a solution, is given by the following equation:

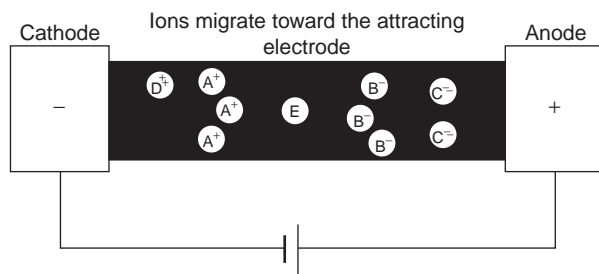
$$v = Eq/d6\pi r\eta$$

where  $v$  is the velocity at which the molecule is moving,  $E$  the electric field,  $q$  the net charge on the molecule,  $d$  the distance between the electrodes,  $r$  the radius of the molecule, and  $\eta$  the solution viscosity.

The movement of an ion is proportional to its charge and the electric field, and inversely proportional to its size and the viscosity of the surrounding medium.

## History

The first investigations with electrophoresis were carried out by Lodge in 1886, and were performed using a curved tube filled with jelly and an aqueous solution. Acidic and basic ions were added at the positive and negative ends of the tube and a current was applied, this caused the jelly/solution boundary to move giving an indication of the net ion migration in the system. Linder and Picton in 1896 replaced the jelly with a colored colloidal solution and observed the migration of inorganic compounds and hemoglobin. This method became known as moving boundary or 'free' electrophoresis. When



**Figure 1** General theory of electrophoresis. Many substances exist in solution as electrically charged species. Ions can be separated in an electrical field where the speed and direction of movement depends on their charge.



electrophoresis is conducted in a solution the minimal frictional resistance allows rapid migration of sample components. The technique was used mainly for separating components in biological fluids and also in water purification.

It was not until the work of Tiselius in the 1930s that the potential of electrophoresis as a biochemical tool was realized. Tiselius developed the use of electrophoresis for separating proteins in suspension on the basis of their charge. He optimized the geometry and temperature of the system in an apparatus known as the Tiselius 'moving boundary' system. Detection of the separated components was carried out by detecting concentration changes in optical refraction.

There were still limitations to this method as relatively large samples were required and complete separation was impossible due to boundary issues in the solution. These problems were solved with the use of stable media in which components of a sample migrate as distinct zones, which can then be detected with suitable equipment. This method is termed zone electrophoresis. In 1937, Konig developed a method using paper as the stable medium. Disadvantages associated with paper (dealt with below) led to the development of methods using other media, e.g., Smithies (1955) using starch gel and Kohn (1957) using cellulose acetate. In 1959, Ornstein and Davies and, independently, Raymond and Weintraub, and also Hjerten (1960), developed methods using polyacrylamide gel, where separation is based on both electrophoretic mobility and sieving effects.

Svensson (1961) and subsequently Vesterberg (1969) developed the important technique of isoelectric focusing (IEF). This method exploits the principles of moving boundary electrophoresis. Components are separated according to their pI by the use of carrier ampholytes in the supporting medium. An ampholyte is a compound that can have basic and acidic properties, such as amino acids. The migration of ampholytes is therefore pH dependent. This method has been combined with gel electrophoresis to form the powerful tool of two-dimensional gel electrophoresis pioneered by O'Farrell in 1975.

The next important development in electrophoresis occurred in the 1970s with the emergence of capillary electrophoresis (CE). This technique involves the application of high voltages across buffer-filled capillaries to achieve separations. Until this time electrophoresis equipment offered a low level of automation, long analysis times, with detection only possible by postseparation visualization. Conducting electrophoresis in capillaries overcame these problems with automated analytical equipment, fast analysis times, and online detection of separated

components. Jorgenson and Lukacs, in the 1980s, enhanced this technique and in recent times CE has accelerated the completion of the Human Genome Project.

## General Considerations in Electrophoresis

Several important parameters influence the behavior of a sample being electrophoresed. The electric field is obviously important. Characteristics of the charged molecule or ion itself can affect its migration rate. For example, the higher the net charge on a molecule the more quickly it will move. Whereas the larger the molecule the more slowly it tends to migrate due to increased frictional and electrostatic forces. The shape of a molecule can influence its migration rate since molecules of similar mass and net charge may migrate at different rates due to differential effects of frictional and electrostatic forces.

The choice of buffer is very important as it stabilizes the pH of the sample and the supporting medium. It should also act as an acceptable electrolyte. The buffer is generally chosen to ensure that the sample components do not bind to it altering their rate of migration. In some cases, however, this interaction is desirable. For example, borate buffers are used to separate carbohydrates as they form charged complexes with them. With high ionic strength buffers the proportion of the current carried by the buffer is high and the proportion carried by the sample is relatively lower, reducing the rate of migration. Heat production also increases. Conversely, at low ionic strengths the proportion of current carried by the sample is higher and the migration rate is faster. A low ionic strength buffer reduces the overall current and results in less heat production, but diffusion and resulting loss of resolution are higher. The pH of the buffer is very important when separating organic compounds as it influences the degree of ionization and therefore migration will be pH dependent. When separating fully ionized inorganic compounds the pH of the buffer will have little effect. The temperature of the system can affect the migration of ions, with an increase in temperature increasing velocity.

The support medium used to carry out the separation will have an effect on the migration of components due to adsorption, electroendosmosis, or molecular sieving. Adsorption occurs when the sample sticks to the support medium (such as to the cellulose hydroxyl groups of cellulose paper) giving a characteristic comet shape to the separation. Adsorption reduces both the rate and resolution of the

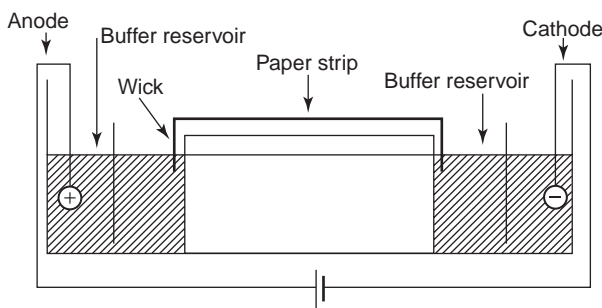
separation. Electroendosmosis is a phenomenon that occurs when the surface of the supporting medium becomes negatively charged in contact with water in the buffer solution. This can result in a change in mobility of charged molecules and is an important consideration when the isoelectric points of compounds are required. Allowance can be made for this effect by measuring the migration of neutral compounds in the same system. Molecular sieving occurs in gel media where the movement of large molecules is hindered by the gel matrix. Decreasing the pore size of the gel will decrease the mobility of the larger molecules. Consideration of all of these properties will determine the type of support medium to be used for particular samples to be separated.

## General Techniques and Applications

Separation of samples using electrophoresis can be useful both for analysis and as a preparative tool. The sample and purpose of the separation will determine which electrophoresis protocol will be the most useful. The following section is a brief overview of some of the techniques and media available and their applications. This is by no means a comprehensive list and readers are referred to the references at the end of this article for more detailed information.

### Paper Electrophoresis (Low Voltage)

Low-voltage paper electrophoresis is the simplest and cheapest form of electrophoresis. A strip of commercially available chromatography paper is soaked in buffer and placed with one end in each buffer reservoir (connecting wicks may be used). It is important that the paper is saturated with the buffer since it is the buffer that conducts the majority of the current. A spot of sample is placed in the center of the strip. When the current is applied ions separate out and migrate toward the attractive electrodes (Figure 2). Paper electrophoresis has several limitations. Only small charged molecules can be reliably separated since many macromolecules adsorb on to the paper, although it is possible to reduce the adsorption by using a buffer that is more alkaline than the isoelectric point of the sample. Paper systems are associated with high electrical resistance causing heating that dries out the paper. Although this method has been largely superseded by gel methods for many samples it is still useful for separation of small molecules such as amino acids, small peptides, nucleotides, and inorganic ions. An important limitation to this technique is that considerable diffusion of small molecules occurs at low voltage.



**Figure 2** Paper electrophoresis (low voltage). A paper strip is soaked in buffer (using connecting wicks). It is important that the paper remains wet. When a spot of sample is applied to the strip, and an electrical field applied, ions will migrate toward the attracting electrode.

### Paper Electrophoresis (High Voltage)

For the electrophoresis of small molecules it may be beneficial to use a high-voltage system. This results in better resolution and rapid separations (10–60 min). The voltage supplied can be up to 10 000 V, which produces a great deal of heat making it essential to have a direct cooling system. This method has been used to detect drug metabolites in urine, and for the separation of oligosaccharides, and peptides.

### Cellulose Acetate Electrophoresis

Commercially available cellulose acetate sheets have a homogeneous micropore structure. Unlike paper, very little adsorption occurs. Cellulose acetate is less hydrophilic than paper and so less buffer is held in the medium, leading to shorter separation time and better resolution than paper and less sample required than with other mediums. The cellulose acetate sheet must be saturated with buffer prior to electrophoresis. This is done by floating the strip on the surface in a shallow tank, since rapid immersion can trap air bubbles in the medium. The relatively low amount of buffer held by the sheet can lead to excessive heat production and drying out. Cellulose acetate electrophoresis plays an important role in clinical diagnostics, for example, in the analysis of hemoglobin. It is also commonly used in the separation of other blood proteins, enzymes, mucopolysaccharides, urine, and other bodily fluids.

### Thin-Layer Electrophoresis

Thin layers of silica, alumina, or cellulose can be prepared on glass plates. Electrophoresis is carried out in a horizontal unit as for the cellulose acetate method outlined above. The thin-layer media are saturated with buffer by capillary action from the reservoir by the connecting wicks. The method is

rapid and gives good resolution and sensitivity. Cellulose medium has been used in the separation of polysaccharides and peptides.

### Starch Gel Electrophoresis

Starch gel is made by boiling hydrolyzed potato starch in a buffer and pouring the molten starch into a preformed mold to cool. Gels of varying pore sizes can be made by altering the percentage of starch to buffer, although an exact pore size cannot be determined and batches can be inconsistent. Gels can be run either horizontally or vertically. A limitation to this gel type is that the starch possesses some charged groups that can cause adsorption of some macromolecules (e.g., proteins). An important application of starch gels is in allozyme electrophoresis (the separation of allelic variants of enzymes).

### Agar Gel Electrophoresis

Agar is a mixture of two galactose-based polymers, agarose and agarpectin. Agar will solubilize in aqueous buffers above 40°C and sets to form a gel at ~38°C. The gel has a large pore size and low frictional resistance enabling a rapid movement of ions, which assists in the separation of macromolecules. Agar gels can be used for separating nucleic acids at a lower cost than agarose gels (described next). A disadvantage to agar as a medium is that electroendosmosis is severe unless its sulfur content is removed prior to purification. Citrate agar electrophoresis is routinely used clinically to identify hemoglobins, for example, in diagnosis of sickle cell anemia.

### Agarose Gel Electrophoresis

Agarose is a polysaccharide derivative of agar. Gels are made by heating up agarose in an appropriate buffer. The gel contains microscopic pores that act as a molecular sieve. Certain molecules can also interact with agarose to varying degrees affecting their mobility. The higher the concentration of agarose the better the resolution. One per cent agarose is commonly used. The agarose gel is kept submerged under buffer in a horizontal apparatus, and for this reason it is sometimes called submarine electrophoresis, although it is possible to run agarose gels in a vertical apparatus. Nucleic acids are commonly separated by agarose gels and they are useful for DNA mapping and apoptosis ladder assays. As DNA and RNA have a constant charge to mass ratio (above 400 bp), separation is based on size and shape, not charge. Agarose gels are normally neutral, but in certain circumstances alkaline gels are used, e.g., for separation of nucleic acids as single strands. RNA preparations can be separated using a denaturing agarose gel. Low

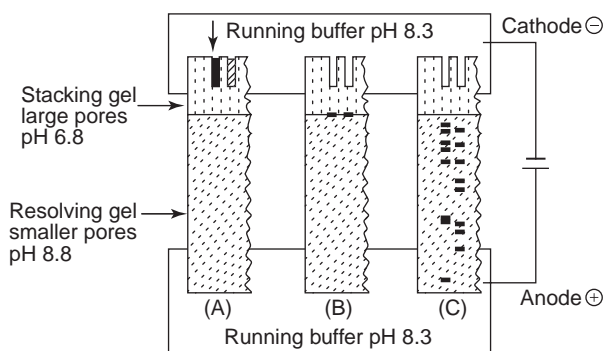
melting point (LMP) agarose is used for preparative gels. DNA fragments can be cut from the gel and the LMP agarose easily removed without causing strand separation due to its LMP. Very small fragments of DNA (less than 100 bp), are better resolved using the finer mesh of polyacrylamide gels.

### Polyacrylamide Gel Electrophoresis

Polyacrylamide gels are made by chemical polymerization of a mixture of acrylamide and bisacrylamide (a cross-linker) in the presence of a catalyst and an initiator of the polymerization reaction. The porosity of the gel is determined by the relative concentration of acrylamide to cross-linker and by the total percentage of monomers. The final concentration of acrylamide depends on the sample under study with high acrylamide concentrations giving better resolution of low molecular mass proteins, and vice versa. The gels can be prepared with a high degree of reproducibility. It is possible to finely tailor the gel pore size, for example, by preparing gradient gels where the acrylamide concentration can vary from 5% to 20% in the same gel, making it possible to separate a mixture of molecules with very varying molecular mass values. By varying gel and buffer parameters it is possible to separate samples on the basis of charge, size, or a combination of charge and size. Gels can be formed into either tubes or slabs and can be used for analysis or preparatively. Polyacrylamide gel electrophoresis (PAGE) is routinely used for protein analysis, and can also be used to separate nucleic acid fragments smaller than 100 bp. Nucleic acids are usually analyzed using a continuous buffer system where there is a constant buffer composition, pH, and pore size throughout the gel. Native proteins can be separated according to differences in their charge density, so long as the buffer in the gel is suitable for maintaining the protein in its native state. This enables enzyme preparations to be assessed for purity and also allows their activity to be assayed after the separation.

### Discontinuous PAGE

Discontinuous PAGE is a means of enhancing the high resolution of PAGE for analysis by focusing the sample into very sharp bands before entering the running gel, and is a common method used for proteins. It involves gels of two different pore sizes: the stacking gel has a larger pore size than the resolving gel below. The stacking gel offers little resistance to the passage of large molecules, which results in the concentration of the proteins at the interface between the two gels, thus giving a better resolution. There are also different buffer ions in the gel compared to



**Figure 3** Discontinuous polyacrylamide gel electrophoresis: (A) Sample is applied to the wells (solid arrow). (B) When current is applied, the sample concentrates at the gel interface. Proteins are stacked according to mass. (C) Molecular sieving occurs when the sample enters the resolving gel.

those in the electrode reservoir, hence the name discontinuous PAGE. These may differ in pH and/or buffer composition, and also assist in the focusing of samples before they enter the resolving gel (Figure 3). It is possible to use this system to get a good separation of a large sample volume.

### SDS-PAGE

Sodium dodecyl sulfate (SDS)-PAGE is a method used for separating and identifying proteins according to their molecular mass. Protein mixtures are boiled in a buffer containing SDS and a reducing agent. SDS is a detergent with a strong negative charge, which binds avidly to all proteins regardless of their natural charge. So, all proteins will carry a strong negative charge. The SDS binds to polypeptides with a constant weight ratio; therefore, the charge per unit mass is nearly constant and electrophoretic mobility becomes a function of molecular mass only (with a number of exceptions, e.g., glycoproteins, very basic proteins). The reducing agent breaks the disulfide bonds and the compact three-dimensional shape of the protein is changed to a rod-like structure. Separation is therefore not affected by the proteins' tertiary structure. The proteins are loaded onto a polyacrylamide gel, commonly a discontinuous gel system.

### Isotachopheresis

Isotachopheresis is a technique based on the principles of moving boundary electrophoresis. Two buffer systems are used: a leading electrolyte and a trailing electrolyte. The leading electrolyte has a higher mobility than the fastest sample component; likewise the trailing electrolyte has a slower mobility than the slowest component. When an electric field is applied

the leading electrolyte moves quickly toward the appropriate electrode with the sample ions following, creating zones in order of their mobilities. This method can only be used for either anions or cations, not both at the same time. Any charged substance can be separated by isotachopheresis. Isotachopheresis is usually carried out in capillaries. This method has the advantages of no sample preparation, speed, and is applicable to a wide sample range. Isotachopheresis has been used for analysis of organic acids in silage, anions in urine and serum, inorganic ions in water, proteins, and amino acids.

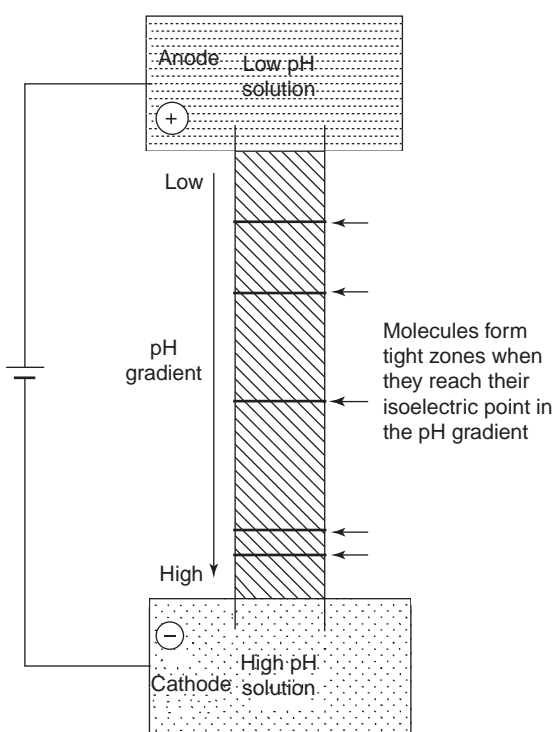
### Isoelectric Focusing

This is a method of separating molecules according to charge, and to a lesser extent tertiary structure. Amphoteric substances such as amino acids and peptides possess charges that vary from a net negative charge at low pH, through zero at some intermediate pH, through to a net positive charge at high pH. The isoelectric point is the pH at which the molecule has a net zero charge. Amphoteric molecules can be separated in an electric field across which there is both voltage and pH gradients. A pH gradient can be created by using ampholytes that are small organic molecules with different combinations of basic and acidic groups so that each one has a separate  $pK_a$ . By using a mixture of ampholytes with a low pH solution in the anode reservoir and a high pH solution in the cathode reservoir, each ampholyte will migrate to a pH equal to its pI and buffer the pH at that point. The sample molecules will migrate until they reach their isoelectric point and be focused into sharp bands (Figure 4). IEF can be carried out in various media, but agarose is frequently used as it can be gelled with a very large pore size, so reducing molecular sieving effects. A low percentage acrylamide gel is also used especially when carrying out two-dimensional gel electrophoresis (see below).

### Two-Dimensional Electrophoresis

Two-dimensional electrophoresis is a powerful technique that combines two different electrophoresis techniques, separating molecules by two different properties, resulting in a greater resolving power than each technique alone. For example, IEF in the first dimension and SDS-PAGE in the second dimension will separate proteins according to charge, followed by molecular mass. In a mixture of proteins from a cell lysate, thousands of different proteins can be resolved. This is a useful method for peptide mapping. The method can also be used as a preparative tool to obtain peptides for amino acid sequencing and also for the purification of antibodies. Thin-layer





**Figure 4** Isoelectric focusing. Molecules are separated according to charge in an electric field across which there is both voltage and pH gradient. The pH gradient is formed using ampholytes.

electrophoresis in one dimension and chromatography in another can also be used in a two-dimensional separation method for proteins and nucleic acids, and is frequently used in peptide mapping.

### Immuno-electrophoresis

Immuno-electrophoresis is defined as the separation and identification of proteins based on differences in electrical charge and reactivity with antibodies. The separating medium is usually an agarose gel. The gel has alternating wells and slots cut into it. The sample is placed in the wells and electrophoresis is carried out to separate the proteins. Antibody (or a mixture of antibodies) is added to the slots. The antibody and antigen migrate toward one another and form precipitin arcs at the region of optimal concentration. The method is mainly used clinically to determine the blood levels of immunoglobulins, and aids in the diagnosis and evaluation of the therapeutic response in many disease states affecting the immune system and also in the diagnosis of multiple myeloma. A similar technique is called rocket immuno-electrophoresis. In this case the gel contains specific antisera. The protein precipitates with the antisera in a rocket-shaped pattern. The height of the peak (or area under the

curve) is proportional to the concentration of the protein.

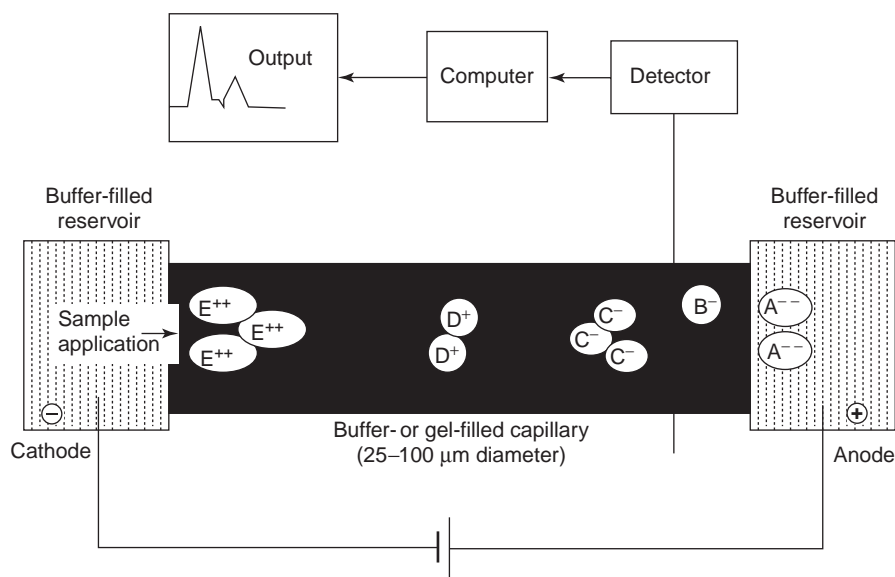
### Capillary Electrophoresis

CE separates samples by applying voltage across buffer-filled capillaries. The capillaries are usually made of fused silica with an external polyimide coating for increased mechanical strength. Capillaries are between 25 and 100 cm long with an internal diameter of 25–100  $\mu\text{m}$ . After passing through a detector the separated ions are seen as peaks (Figure 5). The area of each peak is proportional to the concentration of the molecule, which enables quantitative studies to be carried out. Advantages of CE are that analysis time is rapid: 1–30 min, and operating costs are low. Detection can be by a variety of methods depending on the sample. CE has become a range of techniques that include: capillary zone electrophoresis where separation is based on size and charge differences; capillary gel electrophoresis analogous to SDS-PAGE where separation is based solely on size; capillary isoelectric focusing, separation of neutral compounds using surfactant micelles (micellar electrokinetic capillary chromatography); and capillary isotachopheresis. Using these methods it is possible to use CE for separation of many different sample types, e.g., inorganic ions such as metal ions and anions (chloride, sulfate, and nitrate), DNA, serum proteins. CE is routinely used in a clinical setting, in the pharmaceutical industry, and also in forensic laboratories.

## General Operational Procedures

### Sample Application

Samples are generally applied using a micropipette or syringe. The position of application will depend on whether the sample contains components that will migrate to both electrodes. In this case the sample is usually applied to the center of the support medium as a spot or narrow streak and components will migrate in both directions. For samples where the components are all positively or negatively charged, sample application should be at one end of the medium farthest away from the attracting electrode. For a gel support medium the sample is usually applied to a well cast in the gel. For capillaries the sample is applied at one end furthest from the detector. This is usually achieved by dipping the capillary into the sample vial, the vial is then pressurized causing a volume to be forced into the capillary. Sample volumes used in electrophoresis can vary from  $1 \times 10^{-5} \text{ cm}^3$  (in CE) to  $5 \text{ cm}^3$  for some preparative separations.



**Figure 5** Capillary electrophoresis. As the ions migrate along the capillary, they pass through a detector. After integration, the separated ions are seen as peaks. Analysis is rapid: 1–30 min.

### Detection

Detection of unknown compounds or molecules after electrophoresis is usually *in situ*. Most biological molecules are colorless and need to be treated to produce stable colored compounds. Examples are the visualization of proteins in polyacrylamide gel with Coomassie Blue or silver stain. The medium can also be treated with compounds that bind to the sample causing them to fluoresce under ultraviolet (UV) light. Enzymes can be detected by histochemical methods where substrates are converted into insoluble colored products. It is possible to detect many biological compounds by immunological methods; for example, in western blotting the separated components are transferred from a polyacrylamide gel onto another suitable medium (e.g., nitrocellulose) that can then be treated with antisera and the conjugated product detected by various methods. If a sample contains radioactive components then the radioactive element can be detected by several different methods, e.g., phosphor imaging, fluorography. In CE, there are detectors built into the instrument to detect the peaks as they are resolved. The most frequently used is a UV absorbance detector. There are also fluorescence, conductivity, and indirect detectors available. It is possible to combine CE and mass spectrometry to obtain structural information about the resolved peaks, making this a very powerful technique.

A number of methods are used for recovering separated compounds from the support medium.

Cellulose acetate will dissolve in acetone leaving the separated compound in solution. Macromolecules can be recovered from starch and polyacrylamide gels by electrodialysis. In some forms of preparative electrophoresis the sample is run until the components ‘run off the end’ and can be collected.

**See also:** **Capillary Electrophoresis:** Overview. **Electrophoresis:** Principles; Isotachopheresis; Isoelectric Focusing; Polyacrylamide Gels; Two-Dimensional Gels; Blotting Techniques.

### Further Reading

- Andrews AT (1986) *Electrophoresis Theory, Techniques, and Biochemical and Clinical Applications*, 2nd edn. Oxford: Clarendon.
- Camilleri P (1998) *Capillary Electrophoresis Theory and Practice*, 2nd edn. Boca Raton: CRC Press.
- Catsimpoilas N (ed.) (1973) *Isoelectric Focusing and Isotachopheresis*. New York: New York Academy of Sciences.
- Chin HP (1970) *Cellulose Acetate Electrophoresis Techniques and Applications*. Ann Arbor Michigan: Ann Arbor Sci. Publ.
- Dunbar BS (1987) *Two Dimensional Electrophoresis and Immunological Techniques*. New York: Plenum.
- Dunn MJ (1993) *Gel Electrophoresis: Proteins (Introduction to Biotechniques)*. Oxford: BIOS Scientific Publishers Limited.
- Hames BD (ed.) (1998) *Gel Electrophoresis of Proteins: A Practical Approach*, 3rd edn. Oxford: Oxford University Press.



Hawcroft DM (1997) *Electrophoresis the Basics*. Oxford: Oxford University Press.

Rickwood D and Hames BD (eds.) (1990) *Gel Electrophoresis of Nucleic Acids: A Practical Approach*, 2nd edn. Oxford: Oxford University Press.

Righetti PG (1983) *Isoelectric Focusing: Theory, Methodology and Applications*. Amsterdam: Elsevier Science Publishers BV.

Righetti PG (1990) *Immobilized pH Gradients: Theory and Methodology*. Amsterdam: Elsevier Biomedical Press.

Simpson CF and Whittaker M (eds.) (1983) *Electrophoretic Techniques*. London: Academic Press Inc.

Westermeier R and Naven T (2002) *Proteomics in Practice: A Laboratory Manual of Proteome Analysis*. Weinheim: Wiley-VCH.

## Principles

**B Gas**, Charles University, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

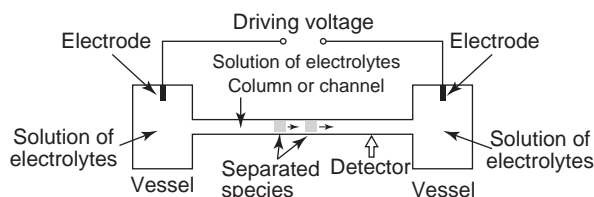
Electrophoresis is a separation method that is based on the migration of charged species in a supporting medium (a liquid or a hydrophilic gel) under the influence of an electric field. The ability of electrophoresis to separate charged species ranges from small inorganic or organic ions to charged biopolymers (like DNA or proteins), or even chromosomes, microorganisms, or whole cells. Even though electrophoresis was established more than 100 years ago, it is an important technique even today. In the form of gel electrophoresis it is almost the only technique used for solving complicated separation tasks in genomics and proteomics, where thousands and thousands of species must be separated. Furthermore, in the last two decades electrophoretic techniques in capillaries or microchips have had a rapid development. Due to the fast expansion of new electrophoretic methods, many of the older terminologies now have rather different meanings than before. In this article, the basic principles are explained and discussed.

### Basic Terms

The supporting medium where electrophoresis takes place is called the electrophoretic system or the separation system. The separation systems are solutions of electrolytes that are filled into a separation column or a separation channel (Figure 1). Both ends of the column are connected with vessels, which are also filled with the solution of electrolytes and serve as a stock of the solution. The electric voltage (500–30 000 V), which is called driving voltage, is provided by two electrodes in the two vessels and

raises the driving electric field. The analyzed species (analytes) are introduced through a narrow sample plug, which then migrate across the longitudinal axis of the column in the liquid, the negatively charged analytes (anions) in the direction of the anode and the positively charged ones (cations) in the direction of the cathode. As various species have different velocities in the electric field, they are spatially separated from each other and are detected by a suitable detector, which is located at a certain position of the column or, sometimes, at its end. The separation column is often a very thin tube, with an inner diameter of 10–500  $\mu\text{m}$ , which is called the capillary; the electrophoretic methods that utilize such a capillary are called the capillary electrophoretic methods. The separation column can also be formed by a channel in various materials (glass, plastic, silica, etc.), which is fabricated by means of technologies normally used for the production of electronic microchips. These approaches are called ‘lab-on-a-chip’ technologies and electrophoresis that uses this approach is called chip or microchip electrophoresis.

A hydrophilic cross-linked gel (mostly polyacrylamide or polysaccharide gel), which is soaked with a solution of electrolytes, or a viscous solution of a linear polymer with an electrolyte, can also serve as the separation medium. Electrophoresis using such gels or polymeric media is called gel electrophoresis. The polymeric chains of the gel serve in most cases not only as a supporting medium but also influence to a great extent the movement of the separated species. The gels or solutions of linear polymers can also be filled into columns (channels), but as the structure



**Figure 1** Schematic diagram of an electrophoretic instrument.

Hawcroft DM (1997) *Electrophoresis the Basics*. Oxford: Oxford University Press.

Rickwood D and Hames BD (eds.) (1990) *Gel Electrophoresis of Nucleic Acids: A Practical Approach*, 2nd edn. Oxford: Oxford University Press.

Righetti PG (1983) *Isoelectric Focusing: Theory, Methodology and Applications*. Amsterdam: Elsevier Science Publishers BV.

Righetti PG (1990) *Immobilized pH Gradients: Theory and Methodology*. Amsterdam: Elsevier Biomedical Press.

Simpson CF and Whittaker M (eds.) (1983) *Electrophoretic Techniques*. London: Academic Press Inc.

Westermeier R and Naven T (2002) *Proteomics in Practice: A Laboratory Manual of Proteome Analysis*. Weinheim: Wiley-VCH.

## Principles

**B Gas**, Charles University, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

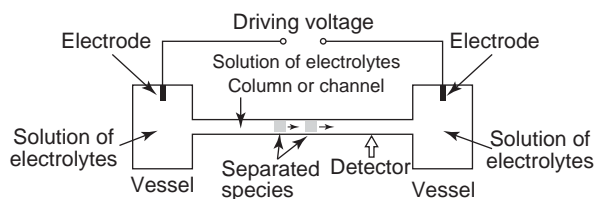
Electrophoresis is a separation method that is based on the migration of charged species in a supporting medium (a liquid or a hydrophilic gel) under the influence of an electric field. The ability of electrophoresis to separate charged species ranges from small inorganic or organic ions to charged biopolymers (like DNA or proteins), or even chromosomes, microorganisms, or whole cells. Even though electrophoresis was established more than 100 years ago, it is an important technique even today. In the form of gel electrophoresis it is almost the only technique used for solving complicated separation tasks in genomics and proteomics, where thousands and thousands of species must be separated. Furthermore, in the last two decades electrophoretic techniques in capillaries or microchips have had a rapid development. Due to the fast expansion of new electrophoretic methods, many of the older terminologies now have rather different meanings than before. In this article, the basic principles are explained and discussed.

### Basic Terms

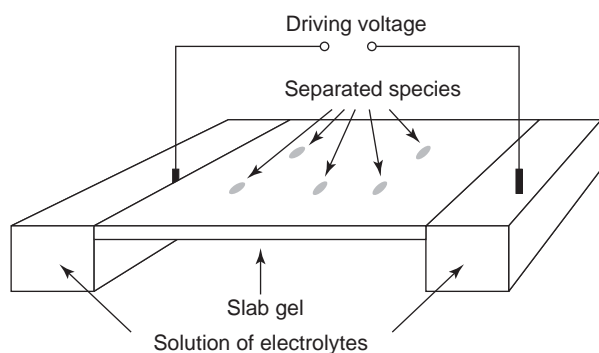
The supporting medium where electrophoresis takes place is called the electrophoretic system or the separation system. The separation systems are solutions of electrolytes that are filled into a separation column or a separation channel (Figure 1). Both ends of the column are connected with vessels, which are also filled with the solution of electrolytes and serve as a stock of the solution. The electric voltage (500–30 000 V), which is called driving voltage, is provided by two electrodes in the two vessels and

raises the driving electric field. The analyzed species (analytes) are introduced through a narrow sample plug, which then migrate across the longitudinal axis of the column in the liquid, the negatively charged analytes (anions) in the direction of the anode and the positively charged ones (cations) in the direction of the cathode. As various species have different velocities in the electric field, they are spatially separated from each other and are detected by a suitable detector, which is located at a certain position of the column or, sometimes, at its end. The separation column is often a very thin tube, with an inner diameter of 10–500  $\mu\text{m}$ , which is called the capillary; the electrophoretic methods that utilize such a capillary are called the capillary electrophoretic methods. The separation column can also be formed by a channel in various materials (glass, plastic, silica, etc.), which is fabricated by means of technologies normally used for the production of electronic microchips. These approaches are called ‘lab-on-a-chip’ technologies and electrophoresis that uses this approach is called chip or microchip electrophoresis.

A hydrophilic cross-linked gel (mostly polyacrylamide or polysaccharide gel), which is soaked with a solution of electrolytes, or a viscous solution of a linear polymer with an electrolyte, can also serve as the separation medium. Electrophoresis using such gels or polymeric media is called gel electrophoresis. The polymeric chains of the gel serve in most cases not only as a supporting medium but also influence to a great extent the movement of the separated species. The gels or solutions of linear polymers can also be filled into columns (channels), but as the structure



**Figure 1** Schematic diagram of an electrophoretic instrument.



**Figure 2** Electrophoresis in slab gel.

of gels is cross-linked they are mechanically stable even without being filled into columns. Therefore, they are used in the form of flat strips or layers that are sticking to a supporting plastic or glass plate and are called slab gels (Figure 2).

## Mobility and Conductivity

When a charged particle of an  $i$ th species with charge number  $z_i$  is placed in the electric field with an intensity  $E$  (in  $\text{V m}^{-1}$ ), the Coulombic force  $F_c = z_i e E$  acting on the particle causes its movement;  $e = 1.602 \times 10^{-19}$  C is the elementary charge. As in electrophoresis the particles to be separated move in liquid or gel form, they face obstacles during the movement – mostly molecules of the solvent or polymer chains of the gel. The collisions and/or interactions with them causes the friction force  $F_f$ , which is proportional to the velocity of the moving particle, and which acts against the Coulombic force. After a very short time ( $10^{-13}$ – $10^{-10}$  s) both forces are balanced,  $F_c = F_f$ , and a mean velocity of the particle attains a constant value. The net movement of charged particles in the electric field is called electromigration and the velocity of such movement is called electrophoretic velocity. Various species have, under given conditions, a different electrophoretic velocity. When moving for a certain time, they are spatially separated. This is a separation principle of the majority (but not all) electrophoretic separation techniques. The electrophoretic velocity of the  $i$ th species  $v_i$  is almost perfectly proportional to the intensity of the driving electric field,  $v_i \sim E$ , when other parameters (e.g., temperature) are constant. The coefficient of proportionality is a quantity called electrophoretic mobility,  $u_i$ , and is defined as

$$v_i = u_i E \quad [1]$$

The dimension of the mobility in SI units is  $\text{m}^2 \text{V}^{-1} \text{s}^{-1}$ . The electrophoretic mobility is unique for a given species and is therefore tabulated for a

**Table 1** Limiting ionic mobilities  $u_i^0$  of some ions at 25°C

Cations	$u_i^0$ , $10^{-9} \text{ m}^2 \text{V}^{-1} \text{s}^{-1}$	Anions	$u_i^0$ , $10^{-9} \text{ m}^2 \text{V}^{-1} \text{s}^{-1}$
Cs <sup>+</sup>	80.0 <sup>a</sup>	Br <sup>−</sup>	80.9 <sup>a</sup>
K <sup>+</sup>	76.2 <sup>a</sup>	Cl <sup>−</sup>	79.1 <sup>b</sup>
Na <sup>+</sup>	51.9 <sup>a</sup>	F <sup>−</sup>	57.0 <sup>b</sup>
Li <sup>+</sup>	40.1 <sup>a</sup>	SO <sub>4</sub> <sup>2−</sup>	82.9 <sup>b</sup>
Ammediol <sup>+</sup>	29.5 <sup>b</sup>	CH <sub>3</sub> COO <sup>−</sup>	42.4 <sup>a</sup>
β-Alanine <sup>+</sup>	36.7 <sup>b</sup>	β-Alanine <sup>−</sup>	30.8 <sup>c</sup>
H <sup>+</sup>	362.3 <sup>a</sup>	OH <sup>−</sup>	206.4 <sup>a</sup>

<sup>a</sup> Atkins PW (1994) *Physical Chemistry*. Oxford: Oxford University Press.

<sup>b</sup> Foret F, Křivánková L, and Boček P (1993) *Capillary Zone Electrophoresis*. Weinheim: VCH.

<sup>c</sup> Hirokawa T, Gojo T, and Kiso Y (1987) *Journal of Chromatography* 390: 201–223. Ammediol, 2-amino-2-methyl-1,3-propanediol.

given temperature and solvent for many ions. By convention, it is negative for negatively charged species (anions) and positive for positively charged species (cations). To a certain range it depends on the concentration of the electrolytes (or the ionic strength) in solution. Therefore, the usual data for tabulated electrophoretic mobilities of ions are the limiting or absolute ionic mobilities,  $u_i^0$ . They are the mobilities that are extrapolated for zero ionic strength assuming further that the substance would be completely in the given ionic form. Such conditions can be theoretically reached in infinitely diluted solutions. Table 1 gives limiting ionic mobilities  $u_i^0$  for several ions. The ‘real’ ionic mobilities under a given finite ionic strength and given conditions are called actual mobilities.

The electromigration of charged species in the separation column increases the electric current. The current density  $j$  (in  $\text{A m}^{-2}$ ) at a particular position of the column is given by

$$j = FE \sum_{i=1}^n c_i u_i z_i \quad [2]$$

where  $F = 96487 \text{ C mol}^{-1}$  is the Faraday constant,  $c_i$  is the concentration of the  $i$ th charged species (in  $\text{mol m}^{-3}$ ), and  $n$  is the total number of charged species present. Notice that the product  $u_i z_i$  is always positive, so the movement of positive species in one direction and negative species in the opposite direction contribute to the electric current density in the same way. According to Ohm’s law, the current density  $j$  is proportional to the intensity of the electric field  $E$ ,

$$j = \kappa E \quad [3]$$

The proportionality constant is the specific conductivity,  $\kappa$ , which is defined by the above relation.

The SI unit for specific conductivity is  $\text{S m}^{-1}$ . Obviously, the specific conductivity can be calculated if concentrations, charge numbers, and mobilities of all the charged species are known,

$$\kappa = F \sum_{i=1}^n c_i u_i z_i \quad [4]$$

## Mobility of Weak Electrolytes – Dependence on pH

Most substances behaving as electrolytes, especially organic acids, bases, and ampholytes, are weak electrolytes. Under given conditions their dissociation or protonation is not complete and reaches a chemical equilibrium. Then, significant parts of the substance can exist as ionic forms with various charge numbers or in the neutral form; and all these forms have different velocities. In most cases the acid–base reactions in solutions are very fast so the electromigration of the substance in the electric field can be regarded as the movement of one entity with all its ionic or neutral species moving with a certain mean velocity. Such a mean electrophoretic velocity  $v_A$  of electromigration of the weak electrolyte A is then

$$v_A = \frac{E}{\bar{c}_A} \sum_{k=1}^m c_{A,k} u_{A,k} \quad [5]$$

where  $\bar{c}_A = c_{A,0} + \sum_{k=1}^m c_{A,k}$  is the total (analytical) concentration,  $c_{A,k}$  and  $u_{A,k}$  are the concentration and actual mobility of the  $k$ th ionic species, respectively,  $c_{A,0}$  is the concentration of the neutral form, and  $m$  is the total number of all ionic species of the weak electrolyte A. The above equation allows a definition of the effective mobility  $\bar{u}_A$  of the weak electrolyte A:

$$\bar{u}_A = \frac{1}{\bar{c}_A} \sum_{k=1}^m c_{A,k} u_{A,k} = \sum_{k=1}^m x_{A,k} u_{A,k} \quad [6]$$

where  $x_{A,k}$  is the molar fraction of the ionic species  $k$ .

The molar fractions of the ionic species are dependent on pH, so the effective mobility of the weak electrolyte is dependent on the pH of the solution as well. This fact shows that changing pH is a very powerful tool that allows one to ‘tune’ the effective mobilities of analytes to obtain their best separation.

## Other Ways to Influence Electrophoretic Mobility

As the difference in electrophoretic mobilities of analytes is of key importance for obtaining good separation, additional ways are used to influence it.

Here, electrophoresis often employs principles used in chromatography – interaction of the separands with another phase, which is called the pseudostationary phase. The pseudostationary phase is a substance that is added to the separation system in the column, and can interact with the species to be separated. The substance can be neutral, then it does not have its own electrophoretic movement; or it can be charged, then it can move in the column with certain mobility. In both cases, the analytes with their own electrophoretic movements encounter on their way the molecules of the pseudostationary phase and interact with them by forming a temporary complex or by association. During the time when the separands are bound to the pseudostationary phase its mobility is different; however, when it is free it moves with its own mobility. As the rate constants of the interaction are mostly very high, in analogy with weak electrolytes, the analyte then moves with a certain mean mobility that lies somewhere between the bound and free mobilities. The mean mobility is in this way dependent on the interaction (complexation) constant. Many substances can be used for this purpose, such as 2-hydroxyisobutyric acid, which forms complexes with many ions, especially with lanthanides, and enables their electrophoretic separation when added to the separation systems.

## Enantioselective Electrophoresis

An important task that can be solved by electrophoresis is the separation of enantiomers of chiral compounds. It is performed by means of their different interactions with a chiral pseudostationary phase – another chiral substance, called chiral selector, which is present in the separation system. The enantiomers of a chiral analyte have the same physicochemical properties in the achiral environment, so their mobilities are also the same and cannot be separated when the separation system is achiral. However, when a chiral substance, which is able to interact (to form complexes) with the analyte, is added to the system, the constants of complexity are generally different and the mobilities of the complexes with both enantiomers may also be different. This enables their separation.

## Micellar Electrokinetic Chromatography

In micellar electrokinetic chromatography (MEKC), the added pseudostationary phase are micelles. The micelles are spherical aggregates of several molecules of a surface-active compound like sodium dodecyl sulfate (SDS). The core of the SDS micelles is composed of entangled aliphatic chains, while the outer shell is formed by charged sulfate groups.



The micelles are highly charged and move in the electric field. They can interact with molecules of analytes, even when they are neutral, and form transient associations. In this way the analytes obtain an electrophoretic movement and can be separated if their interaction constants are different.

### Mobility in Sieving Media: Gel Electrophoresis

The hydrophilic gel soaked with an electrolyte solution, which is often used in electrophoresis as a supporting medium has, additionally, an even more important function: it serves as a sieving medium for separation of high molecular mass biomolecules such as DNA or proteins according to their chain length or molecular size. For example, in sequencing of DNA by the Sanger method the resulting mixture contains a series of single-stranded DNA fragments with the chain lengths ranging from  $\sim 20$  to  $600\text{--}1000$  nucleotides differing only by a single nucleotide. As every nucleotide contains a phosphate group, which has a negative charge, the total charge of any fragment is exactly proportional to its length. Such a mixture of chains cannot be separated by electrophoresis in simple liquid solutions according to their length. The reason is that the chains are only slightly coiled and form the statistical free-draining coils – that is, the electrolyte solution can flow through the coils. Thus, both the Coulombic force and the friction force are proportional to the chain length in the same way, and hence the mobility of the fragments in free solution  $\mu^0$  is not dependent on the chain length. The free solution mobility of DNA is  $\sim 37 \times 10^{-9} \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$  at  $25^\circ\text{C}$ . The free solution mobility was independent of DNA molecular mass, from  $\sim 400$  bp to  $48.5$  kbp.

The diameter of the DNA coils can be characterized by the gyration diameter  $R_g$ , which is dependent on the length of the chain. When the separation system is formed by a neutral cross-linked hydrophilic gel, e.g., polyacrylamide, with a certain mean size of pores between polymer chains, the separated DNA fragments being rather spherical coils mechanically collide with the polymer chains and have to pass through the pores in some way. When the polymer chains are approximately regarded as straight long rods with radius  $r$ , such a type of movement of the DNA fragments is called the Ogston movement. Its effective mobility  $\bar{\mu}$  is given by

$$\bar{\mu} = \mu^0 \exp(-K_R \Phi) \quad [7]$$

where  $K_R \sim (R_g + r)^2$  is the retardation coefficient and  $\Phi$  is the concentration of the gel. As the gyration

radius is dependent on the chain length, the effective mobilities of the DNA fragments are also dependent on the chain length. Thus, DNA fragments can be separated by electrophoresis.

The DNA fragments and also other biopolymers (proteins, carbohydrates) can be separated by electrophoresis in sieving media even if their chains are much longer and their gyration radius is much greater than the radius of pores. Here, other more complex mechanisms of movement play a role: reptation model (the separated molecule creeps in chains of the sieving medium as a reptile in grass), biased reptation model, biased reptation with fluctuation, model of entropic barriers, and geometration. For very long double-stranded DNA fragments (longer than  $\sim 10$  kbp) pulsed field gel electrophoresis is used, a technique that adopts alternating driving electric field. Here, the mode of movement of the chains is even more complicated.

### Theory – Continuity Equation

A simple environment, in which the separation process in electrophoresis takes place, allows easy formulation of basic transport laws that describe electromigration with good exactness and enables the separation process to be understood well. For example, the approximate continuity equations that describe electromigration of  $n$  strong ions in free solution are

$$\frac{\partial c_i}{\partial t} = D_i \frac{\partial^2 c_i}{\partial x^2} + j \frac{\partial}{\partial x} \left( \frac{c_i u_i}{\kappa} \right), \quad i = 1, \dots, n \quad [8]$$

where  $D_i$  is the diffusion coefficient of the  $i$ th ion. These equations are able to provide information about the concentration of all  $n$  ions in the electric field in any time  $t$  and at any position  $x$  of the separation column, when the initial distribution of all ions is known. The first term on the right-hand side of eqn [8] describes diffusion and the second one electromigration. The division in the electromigration term is a nonlinear mathematical operation. This indicates that electromigration is intrinsically a nonlinear process. The nonlinearity is responsible for some phenomena in electrophoresis, such as self-sharpening of boundaries between zones of analytes or electromigration dispersion.

### Modes of Electrophoresis

The configuration of electrolytes in the separation system can substantially influence the electrophoretic separation. Some of the configurations or modes of electrophoresis showed to be well suited for

specific classes of analytical tasks and are now widely used. Here belong predominantly: zone electrophoresis, isotachopheresis (ITP), and isoelectric focusing (IEF).

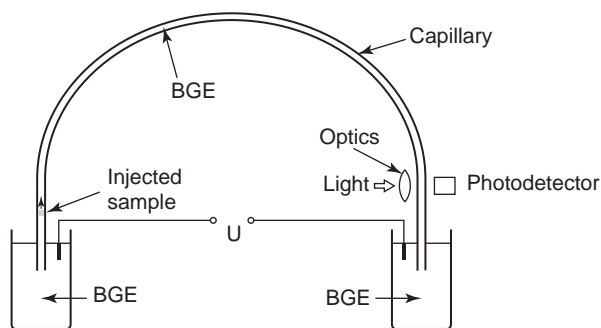
### Zone Electrophoresis

This mode is characterized by two features: (1) the whole separation system together with electrode vessels is filled with a homogeneous solution of electrolytes, which is called the background electrolyte (BGE), (2) the mixture of analyzed species to be separated is introduced or injected in a rather low amount as a short plug at one end of the separation column (Figure 3). The mode of zone electrophoresis is often used in capillary columns, when it is called capillary zone electrophoresis (CZE). The BGE is mostly formed by a suitable buffer to maintain a certain pH, which is optimal for the separation of a given mixture. This mode is also almost exclusively used for gel electrophoresis: here the whole separation system is formed by a buffer-soaked hydrophilic slab gel and the mixture to be separated is introduced in little amounts at one end. After applying the driving electric field the analytes migrate in the column or in the slab gel with mutually different velocities so that they are spatially separated. After some time they are detected with a suitable detector located near the second end of the separation column; in the case of the slab gel electrophoresis, the plate with the gel is removed from electricity and the separated analytes are visualized as spots by a suitable reagent. The conductivity of the BGE in the separation system is only slightly influenced by the presence of the separated analytes, if their amount is low enough. Then, the  $\kappa$  in the denominator of the second term in eqn [8] is practically constant in time and homogeneous across the column length so that the

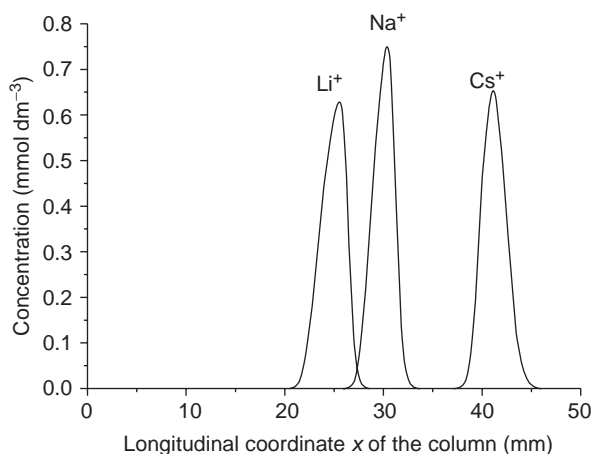
equations can be regarded as linear. The individual zones of analytes, although narrow at the beginning, become broader when moving in the column as diffusion causes their dispersion and their longitudinal profile attains a Gaussian shape. If, however, the amount of an analyte is rather high, it influences conductivity in its own zone, so it is no longer constant. Then, the nonlinearity arising in this way causes the electromigration dispersion – another type of broadening of the zones that further deforms zones to the triangular shape (Figure 4). The quantitative determination of a particular analyte is performed by integration of its detector signal.

### Isotachopheresis

ITP is used for separation of smaller species like inorganic or organic ions. The separation system, which serves for separation of either cations or anions, is composed of two different electrolyte solutions. The main features of ITP are as follows: (1) one electrode vessel and the separation column are filled with the leading electrolyte, whose cation (anion) has higher mobility (in absolute value) than any cation (anion) of the separated mixture; (2) the second electrode vessel is filled with the terminating electrolyte, whose cation (anion) has lower mobility than any cation (anion) of the separated mixture; (3) a plug of mixture of cations (anions) to be separated is introduced between the leading and terminating electrolytes (Figure 5). After applying the driving

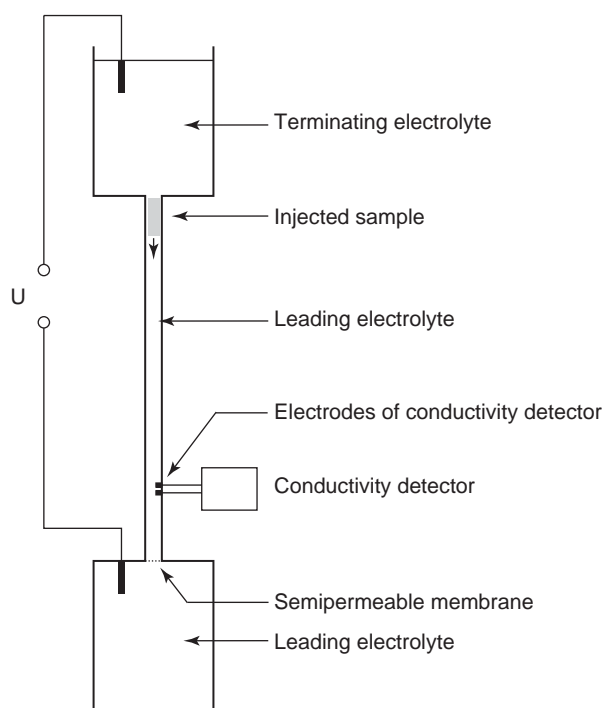


**Figure 3** Configuration of capillary zone electrophoresis. The arrangement shown is close to a real instrument. The capillary is often made of fused silica, whose typical length is 10–100 cm. It has an inner diameter of 20–100  $\mu\text{m}$  and its hydrodynamic resistance is rather high. The electroosmosis is allowed to flow. A photometric (UV) absorption detector is generally used.



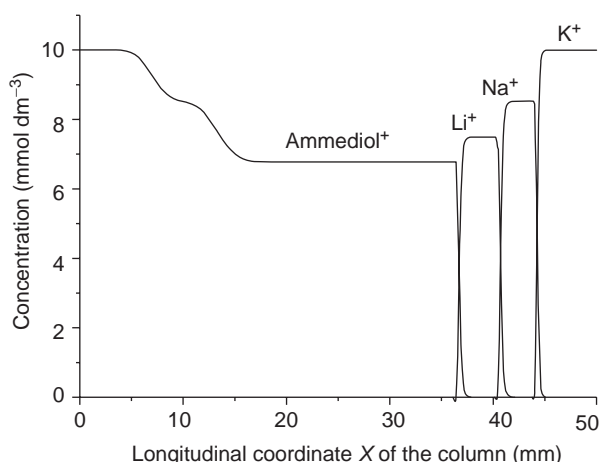
**Figure 4** Longitudinal profiles of analytes in CZE as calculated from eqn [8] by computer. Background electrolyte: 20  $\text{mmol l}^{-1}$  acetic acid/10  $\text{mmol l}^{-1}$  potassium. Sample introduction: 2 mm plug of mixture of 1  $\text{mmol l}^{-1}$   $\text{Cs}^+$ , 1  $\text{mmol l}^{-1}$   $\text{Na}^+$ , 1  $\text{mmol l}^{-1}$   $\text{Li}^+$  ions, injected at position of 7 mm. Current density  $j = 200 \text{ A m}^{-2}$ . The picture shows resulting profiles at  $t = 250 \text{ s}$ , when all ions are almost separated. A slight triangular deformation of analyte zones caused by electromigration dispersion is obvious in the figure.





**Figure 5** Configuration of capillary isotachopheresis. The arrangement shown is close to a real instrument. The capillary is often made of plastic. The semipermeable membrane prevents electroosmotic and hydrodynamic flows of the electrolyte but allows electromigration of ions. A conductivity detector is often used. It has miniature electrodes placed inside the capillary.

electric field the ion of the leading electrolyte leaves the separation column and ‘pulls’ the sample ions, while the terminating ion starts to fill the column and ‘pushes’ the ions. The plug of the separated ions moves ‘bracketed’ between the leading and terminating electrolytes and every ion forms its own individual zone after some time. This stage is called the stationary state. All the zones of the ions are then stacked one by one and move with the same velocity, which is given by the velocity of the leading ion. Unlike the mode of zone electrophoresis, here the specific conductivities  $\kappa$  of the zones of the separated analytes are different and no longer constant. The nonlinearity in ITP is responsible for a useful phenomenon: ‘sharpening’ of the boundaries between the individual zones. The boundaries are very steep and the longitudinal profile of the zones of the analytes is almost rectangular (**Figure 6**). When the zones pass a detector located near the second end of the separation column, the detector gives a rather stepwise signal. The concentration of the analytes in their zones is not the same, as it was in the original mixture: it attains a certain value. For singly charged univalent ions or strong multicharged ions it can be derived from a simple relation, which is called the Kohlrausch regulating function. As concentrations of



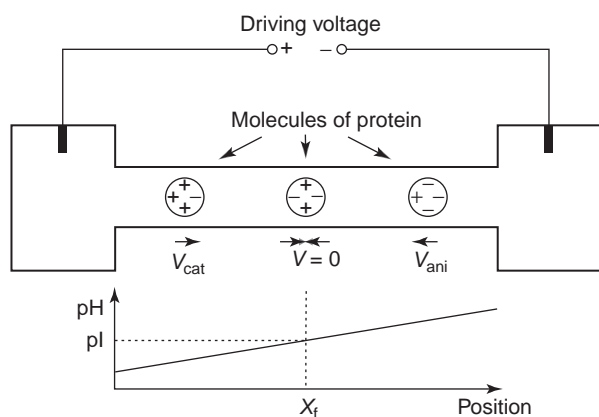
**Figure 6** Longitudinal profiles of analytes in capillary isotachopheresis as calculated from eqn [8] by a computer. Leading electrolyte:  $20 \text{ mmol l}^{-1}$  acetic acid/ $10 \text{ mmol l}^{-1}$  potassium. Terminating electrolyte:  $20 \text{ mmol l}^{-1}$  acetic acid/ $10 \text{ mmol l}^{-1}$  ammediol. Sample introduction:  $6 \text{ mm}$  plug of mixture of  $5 \text{ mmol l}^{-1}$   $\text{Na}^+$ ,  $5 \text{ mmol l}^{-1}$   $\text{Li}^+$  ions, injected at position of  $7 \text{ mm}$ . Current density  $j = 75 \text{ A m}^{-2}$ . The picture shows resulting profiles at  $t = 650 \text{ s}$ , when sodium and lithium are separated. They form isotachophoretic zones with virtually rectangular profiles stacked one by one.

the analytes in the ITP zones are fixed, the longitudinal length of the zone is directly proportional to their amount. The proportionality enables a precise quantitative determination of the analytes when measuring the length of their zones.

Interestingly, the principle of ITP is applied in the initial phase of disc electrophoresis, in the sample and stacking gels, for the purpose of sorting and concentrating (often by a factor of 10 000) the various protein zones. The ITP process is disrupted as the train of zones enters the running gel, where migration continues in the zonal mode.

### Isoelectric Focusing

IEF is an electrophoretic technique for separation of amphoteric species, mostly proteins. It is a technique based on differences in isoelectric points (pI). The species are separated in the pH gradient that is formed in the separation system. The pH gradient is generated in stabilizing matrices of special gels with proteolytic groups, mostly in slabs. Alternatively, it can be generated in a free solution without the gel by passing the electric current through the mixture of a series of ampholytes having isoelectric points in close proximity to each other. The analyte migrates in the separation system by electrophoretic movement according to its net charge. When it reaches a position where the pH is equal to its pI, the net charge becomes zero and the movement stops. The separated



**Figure 7** Principle of isoelectric focusing. The protein has cationic or anionic movement at pH which is below or above its pI, respectively. The velocity of the movement is zero, when protein reaches the position  $x_i$ , this is the site where it will be focused.

species are in this way focused at pH positions corresponding to their pI (Figure 7). When the slab gel is used, they are visualized by a suitable reagent; when focused in free solution in the column, the content of the column is pushed out by pressure or by changing electrolytes in the electrode vessels and detected by a detector located at the end.

## Electroosmotic Flow

The electroosmotic flow (EOF) is an electrokinetic phenomenon accompanying almost all electrophoretic methods. It arises due to the existence of the electric double layer at the inner surface of the separation channel. The electric double layer is generally formed at the interface between the liquid and solid phases. A possible specific adsorption of some ions from the liquid phase at the solid phase or dissociation of some ionogenic groups from the surface layer of the solid phase causes a layer of a surface charge, the Helmholtz layer, sticking to the channel inner wall. This is compensated by the opposite charge, which is distributed in the adjacent liquid solution of the BGE and which ranges to some distance from the inner wall. This part of the double layer is called the diffuse layer. The diffuse layer, which has an excess of volume charge, can be regarded as a very thin 'sleeve' of charged liquid adjacent to the inner wall. When the driving electric field is applied longitudinally to the channel axis, the Coulombic force acting at the charged 'sleeve' causes its movement. Due to friction forces the movement is, in very short time ( $10^{-4}$ – $10^{-2}$  s), spread over the whole volume of the channel and as a result the whole volume of the liquid flows in the channel. This

is the EOF. Unlike the electrophoretic movement where only charged species move in the separation system and neutral molecules of the solvent have no net movement, the EOF causes the liquid to move as a whole. In fact, it acts as a pump transporting a liquid from one electrode vessel to another. The EOF in itself cannot separate the analytes but it is vectorially superposed to its electrophoretic movement. The total electromigration velocity  $v_{EM,i}$  of the  $i$ th separated species is

$$v_{EM,i} = v_i + v_{EOF} \quad [9]$$

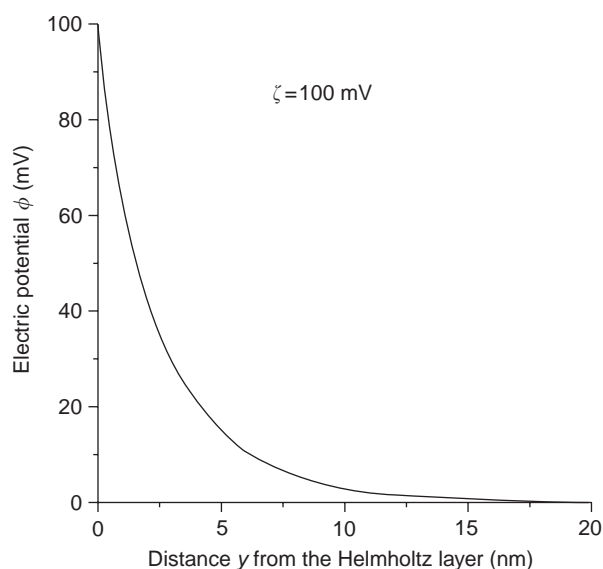
The EOF is especially important in CZE and MEKC. Here, the columns are often made of fused silica, and have negative charge at the inner wall due to dissociation of the silanol groups, which causes the EOF to flow in the direction of the cathode. Therefore, the electrophoretic movement of cationic species is speeded up, while the negative species move slower as the EOF flows against them. Sometimes the latter cannot reach the detector.

The distribution of the velocity of the EOF across the radius of the channel or column is an important issue. In the 'normal' laminar flow that is raised by a difference in pressure, the radial profile of the velocity vector has a parabolic shape: it is zero at the inner wall and maximum in the channel axis. The radial profile of the EOF is quite different: it is also zero at the wall but in a very short distance, which is approximately equal to the thickness of the diffusion layer, it attains a certain value, which is constant across the rest of radius. The radial profile of the EOF is 'plug-like'. This is a very beneficial feature of this type of flow, as it does not cause axial dispersion of the separand.

The excess of the volume charge in the diffuse layer causes the origin of the electric potential  $\phi(y)$  in liquid solution. It is dependent on the distance  $y$  from the Helmholtz layer (Figure 8). The potential  $\phi$  is conventionally set zero at a big distance from the wall. The value of the potential in the diffuse layer in the closest vicinity to the Helmholtz layer ( $y = 0$ ) is called the zeta potential,  $\phi(0) = \zeta$ . When longitudinal driving electric field  $E$  is applied, the velocity  $v_{EOF}$  of the plug-like EOF is related to the zeta potential by the Helmholtz–Smoluchowski equation:

$$v_{EOF} = -\frac{\varepsilon\zeta}{\eta} E \quad [10]$$

where  $\varepsilon$  is the permittivity and  $\eta$  is the dynamic viscosity, respectively, of the liquid solution. The zeta potential and, consequently, also the velocity of the EOF strongly depend on the quality of the surface. Various materials have different zeta potentials.



**Figure 8** Course of electric potential in the diffuse layer of  $10 \text{ mol dm}^{-3}$  NaCl. The zeta potential is supposed to be 100 mV. The curve is a solution of the Poisson–Boltzmann equation and is approximately exponential.

As pointed out above, fused silica used as a material for separation columns in capillary electrophoretic methods has normally a negative charge. Various ions, especially surfactants or big amphoteric ions, can be, however, adsorbed on the surface, which dramatically influences the zeta potential. For example, the addition of a small concentration of a suitable cationic surfactant like tetradecyltrimethylammonium bromide to the BGE causes its adsorption at the inner capillary wall and the reversal of the EOF in silica capillaries to the anodic side. Such surfactants are

often used in the separation of negatively charged species as the EOF in the anodic direction speeds up their movement. The EOF can be also utilized for pumping liquids in electrophoretic methods, especially in electrophoresis in chips. Further, it is used for pumping of the mobile phase in capillary electrochromatography. This method is a hybrid between electrophoresis and chromatography. Instead of using a mechanical pump for the transport of a liquid phase through the column it utilizes the EOF raised by the driving voltage.

**See also: Capillary Electrophoresis: Overview. Micellar Electrokinetic Chromatography.**

## Further Reading

- Auroux PA, Iossifidis D, Reyes DR, and Manz A (2002) Micro total analysis systems 1. Introduction, theory, and technology. *Analytical Chemistry* 74: 2637–2652.
- Chankvetadze B (1997) *Capillary Electrophoresis in Chiral Analysis*. Chichester: Wiley.
- Everaerts FM, Beckers JL, and Verheggen TPEM (1976) *Isotachophoresis. Theory, Instrumentation and Applications*. Amsterdam: Elsevier.
- Hames BD (1998) *Gel Electrophoresis of Proteins*. Oxford: Oxford University Press.
- Heller C (ed.) (1997) *Analysis of Nucleic Acids by Capillary Electrophoresis*. Wiesbaden: Friedrich Vieweg.
- Jones P and Rickwood D (1995) *Gel Electrophoresis: Nucleic Acids*. New York: Wiley.
- Righetti PG (1990) *Immobilized pH Gradients, Theory and Methodology*. Amsterdam: Elsevier.
- Righetti PG (ed.) (1996) *Capillary Electrophoresis in Analytical Biotechnology*. Boca Raton: CRC Press.

## Isotachophoresis

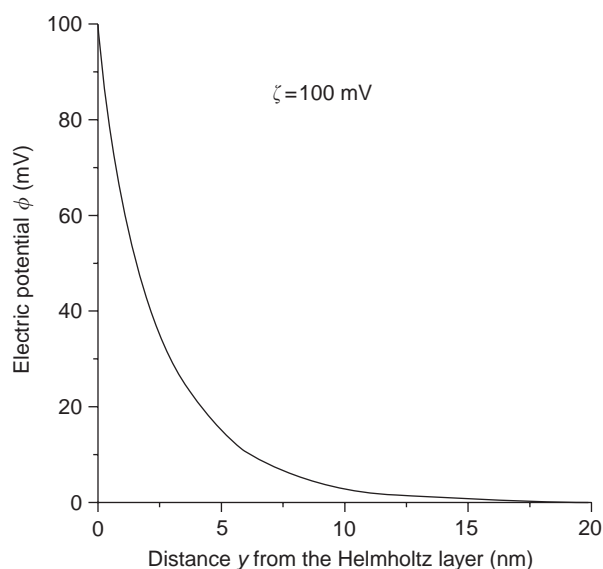
**L Křivánková and P Boček**, Institute of Analytical Chemistry, Brno, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Capillary isotachophoresis began to develop as a modern instrumental analytical method at the end of the 1960s and the beginning of the 1970s and is now a well-established method with a good theoretical background and available commercial instrumentation directed nowadays toward miniaturization. It is

highly useful for analysis of complex mixtures in solution, from simple inorganic ions up to proteins. Isotachophoresis (ITP) can be performed, however, also in any apparatus developed for capillary zone electrophoresis (CZE), and the advantages of this powerful analytical method can be employed in an effective online hyphenation with CZE or transiently within zone electrophoretic migration as well. Here, the basic theory, instrumentation, principles of qualitative and quantitative analyses, and examples of applications of single analytical ITP as well as combined techniques are described. The article is oriented to capillary ITP since capillaries provide high separation efficiency and speed of analysis; however,



**Figure 8** Course of electric potential in the diffuse layer of  $10 \text{ mol dm}^{-3}$  NaCl. The zeta potential is supposed to be 100 mV. The curve is a solution of the Poisson–Boltzmann equation and is approximately exponential.

As pointed out above, fused silica used as a material for separation columns in capillary electrophoretic methods has normally a negative charge. Various ions, especially surfactants or big amphoteric ions, can be, however, adsorbed on the surface, which dramatically influences the zeta potential. For example, the addition of a small concentration of a suitable cationic surfactant like tetradecyltrimethylammonium bromide to the BGE causes its adsorption at the inner capillary wall and the reversal of the EOF in silica capillaries to the anodic side. Such surfactants are

often used in the separation of negatively charged species as the EOF in the anodic direction speeds up their movement. The EOF can be also utilized for pumping liquids in electrophoretic methods, especially in electrophoresis in chips. Further, it is used for pumping of the mobile phase in capillary electrochromatography. This method is a hybrid between electrophoresis and chromatography. Instead of using a mechanical pump for the transport of a liquid phase through the column it utilizes the EOF raised by the driving voltage.

**See also:** **Capillary Electrophoresis: Overview. Micellar Electrokinetic Chromatography.**

## Further Reading

- Auroux PA, Iossifidis D, Reyes DR, and Manz A (2002) Micro total analysis systems 1. Introduction, theory, and technology. *Analytical Chemistry* 74: 2637–2652.
- Chankvetadze B (1997) *Capillary Electrophoresis in Chiral Analysis*. Chichester: Wiley.
- Everaerts FM, Beckers JL, and Verheggen TPEM (1976) *Isotachophoresis. Theory, Instrumentation and Applications*. Amsterdam: Elsevier.
- Hames BD (1998) *Gel Electrophoresis of Proteins*. Oxford: Oxford University Press.
- Heller C (ed.) (1997) *Analysis of Nucleic Acids by Capillary Electrophoresis*. Wiesbaden: Friedrich Vieweg.
- Jones P and Rickwood D (1995) *Gel Electrophoresis: Nucleic Acids*. New York: Wiley.
- Righetti PG (1990) *Immobilized pH Gradients, Theory and Methodology*. Amsterdam: Elsevier.
- Righetti PG (ed.) (1996) *Capillary Electrophoresis in Analytical Biotechnology*. Boca Raton: CRC Press.

## Isotachophoresis

**L Křivánková and P Boček**, Institute of Analytical Chemistry, Brno, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Capillary isotachophoresis began to develop as a modern instrumental analytical method at the end of the 1960s and the beginning of the 1970s and is now a well-established method with a good theoretical background and available commercial instrumentation directed nowadays toward miniaturization. It is

highly useful for analysis of complex mixtures in solution, from simple inorganic ions up to proteins. Isotachophoresis (ITP) can be performed, however, also in any apparatus developed for capillary zone electrophoresis (CZE), and the advantages of this powerful analytical method can be employed in an effective online hyphenation with CZE or transiently within zone electrophoretic migration as well. Here, the basic theory, instrumentation, principles of qualitative and quantitative analyses, and examples of applications of single analytical ITP as well as combined techniques are described. The article is oriented to capillary ITP since capillaries provide high separation efficiency and speed of analysis; however,

the principles described here hold also for the performance in large bore tubes or flat beds.

## Principles

In ITP, the sample is injected between two electrolytes differing in the mobility of the ion that migrates in the same direction as the analytes do and are named the leading and terminating ions. The ion that migrates in the opposite direction is common for both electrolytes and is named the counterion. The common name for such an arrangement of electrolytes is the discontinuous electrolyte system. In a simplified case, i.e., if only strong electrolytes are considered, the leading electrolyte contains the leading ion with the highest mobility, whereas the terminating electrolyte contains the terminating ion with the lowest mobility in the isotachopheretic system. The mobilities of the sample ions should lie between those of the leading and the terminating ions. For weak electrolytes, similar criteria and actual effective mobilities, see below, must be applied.

Figure 1 shows a scheme of an isotachopheretic separation of a mixture of anions. The initial conditions, where the sample containing ions X and Y, e.g., periodate and iodate, respectively, is injected between the leading and terminating electrolytes, are depicted in Figure 1A. The leading electrolyte contains a leading ion L and a counterion, R, e.g., chloride and histidine, when a buffering system of

hydrochloric acid and histidine is used as the leading electrolyte. The terminating electrolyte contains a terminating ion, T, e.g., aspartate, which is suitable for the mentioned separation of periodate and iodate. After the current is switched on, the counterion R of the leading electrolyte (protonated histidine) moves toward the cathode and becomes the common counterion for the whole system. The sample, i.e., the mixed zone of anions periodate and iodate, moves toward the anode and separation occurs. The original mixed zone becomes shorter and shorter (Figure 1B) and then disappears. Thus, complete separation is reached as well as the steady state where all the anions move toward the anode in separate zones (Figure 1C).

The fundamental properties of an isotachopheretic system can be described as follows:

1. The zone of any ion of a substance X, which migrates in an isotachopheretic way behind the zone of L, migrates with the same velocity as the leading zone L,

$$v_{\text{iso}} = v_T = v_X = v_L \quad [1]$$

$$v_{\text{iso}} = \bar{u}_T E_T = \bar{u}_X E_X = \bar{u}_L E_L \quad [2]$$

where  $v_{\text{iso}}$  is the isotachopheretic migration velocity,  $v_X$  is the migration velocity of a zone X,  $\bar{u}_X$  is effective mobility of a substance X, and  $E_X$  is intensity of electric field in a zone X. The effective mobility of a given substance is of key importance here and is defined as follows. A substance X present in a solution in the form of species ions or neutral molecules  $X_0, X_1, X_2, \dots, X_n$  with ionic mobilities  $u_0, u_1, u_2, \dots, u_n$ , migrates in the electric field as a single substance with the effective mobility  $\bar{u}_X$  given by

$$\bar{u}_X = \frac{1}{\bar{c}_X} \sum_{i=0}^n c_i u_i \quad [3]$$

where  $\bar{c}_X$  is the total (analytical) concentration of the substance,

$$\bar{c}_X = \sum_{i=0}^n c_i \quad [4]$$

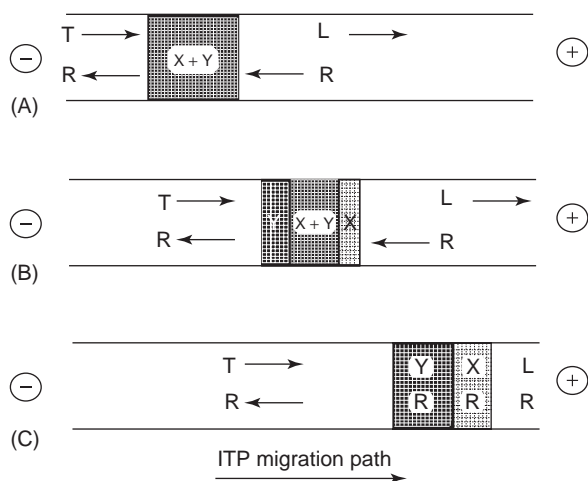
The equation defining  $\bar{u}_X$  can be rewritten in the form

$$\bar{u}_X = \sum_{i=0}^n \alpha_i u_i \quad [5]$$

where  $\alpha_i$  is the mole fraction of species  $X_i$ .

2. ITP proceeds in a channel of a constant cross-section  $S$ . Thus, the driving current density is constant along the whole path,

$$i = \frac{I}{S} = E_T \kappa_T = E_X \kappa_X = E_L \kappa_L \quad [6]$$



**Figure 1** Scheme of the separation process in anionic isotachopheresis of a mixture of anions X and Y. (A) initial conditions, the sample containing anions X and Y is injected between the leading electrolyte composed of the leading ion L and the counterion R, and the terminating electrolyte, containing the terminating ion T and the counterion R. (B) The formation of pure X and Y individual zones. (C) Final steady-state ready for detection of pure X and Y zones containing the entire quantity of sample anions.

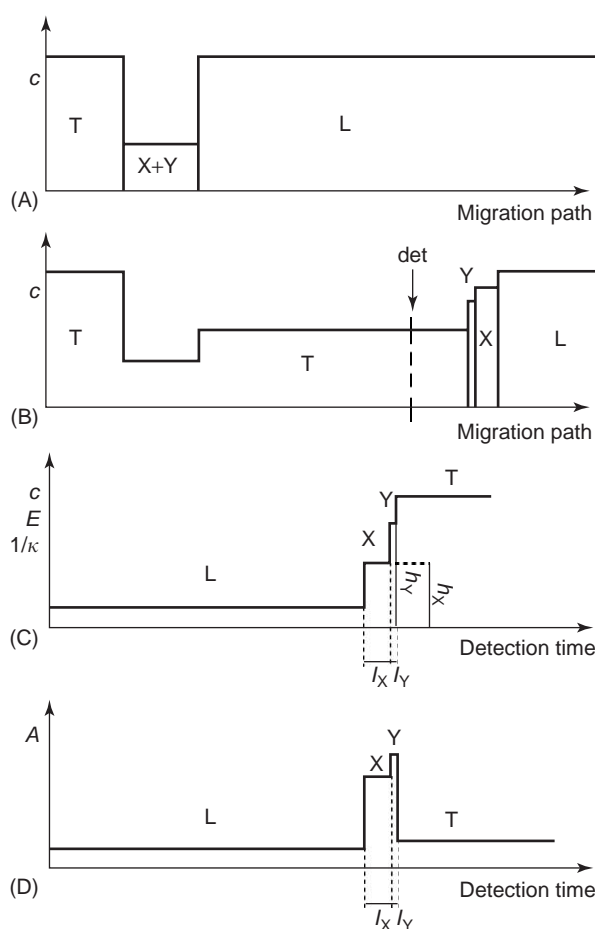


where  $I$  is current and  $\kappa_X$  is the electrical conductivity in a zone  $X$ .

3. The boundaries between individual isotachopheretic zones are sharp – they exhibit self-sharpening effect. For any moving boundary  $2 \rightarrow 1$ , for which  $|\bar{u}_2| < |\bar{u}_1|$ , it holds also  $\kappa_2 < \kappa_1$  and from the relationship given in eqn [6] it follows that  $E_2 > E_1$ . It means that if an ion  $Y$  of lower mobility, e.g., iodate, enters due to diffusion the zone of faster ions  $X$  of periodate, its migration velocity  $v = E_Y \bar{u}_Y$  decreases to  $v = E_X \bar{u}_Y$ , and as a result the iodate ion is again caught up by its own zone. In contrast, if periodate ion appears in the zone of iodate with higher intensity of electric field, it is accelerated and pushed forward to join again the zone of periodate ions. Thus, the step change of intensity of electric field in zones ensures that the boundaries between individual zones remain sharp and physical parameters of the zones, e.g., concentration, conductivity, potential gradient, absorbance, and temperature change also stepwise as demonstrated in Figure 2.
4. The concentrations of substances in their own isotachopheretic zones are adjusted according to the composition of the leading zone, as also illustrated in Figure 2. The situation may be advantageously described with the help of the well-known regulating function derived by Kohlrausch in 1897 for uni-univalent strong electrolytes. Later, the Kohlrausch regulation function was redefined also for strong mono- and multicharged strong electrolytes, and singly charged weak electrolytes within the pH range 3–9. The value of the Kohlrausch regulation function is given by the composition of the original electrolyte along the migration path prior to onset of electromigration, i.e., before the current is switched on:

$$\omega(x) = \sum_i \frac{c_i(x) |z_i|}{|u_i|} = \text{const} (x) \quad [7]$$

where  $z_i$  is the relative charge and  $|u_i|$  is the absolute value of ionic mobility of the  $i$ th ionic species present at the given point  $x$ . The value of  $\omega(x)$  is retained after the current is switched on and remains constant at a certain point  $x$  during the electromigration. If step changes of  $\omega(x)$  occur in the electrolyte composition along the migration path prior to electromigration, then the concentrations of the migrating substances vary during electromigration so that the value of  $\omega(x)$  stays constant. The number of values of Kohlrausch regulation function along the migration axis corresponds to the number of phases (zones) that were in the system before electric current started



**Figure 2** Features of isotachopheretic zones. (A) Distribution of ion concentration along the migration path before isotachopheretic separation. (B) Distribution of ion concentration along the migration path after the isotachopheretic steady-state was reached. The concentrations of sample ions  $X$  and  $Y$  as well as of terminating ions are adjusted according to Kohlrausch regulation function. (C) Distribution of concentration,  $c$ , intensity of electric field,  $E$ , and inverse value of specific conductivity,  $1/\kappa$ , as detected with a detector placed as shown in scheme B. The height of an isotachopheretic step in the record,  $h_X$ , is the qualitative parameter reflecting the effective mobility of the ion migrating in its zone, the length of the isotachopheretic step,  $l_X$ , corresponds to the analyte amount in the sample and is the quantitative parameter. (D) Changes of absorbance followed with an optical detector placed as shown in scheme B. The height of an isotachopheretic step in the record is given by the extinction coefficient of the analyte, the length of the isotachopheretic step,  $l_X$  (or area of the step), corresponds to the analyte amount in the sample and is the quantitative parameter.

to flow. For strong uni-univalent electrolytes, where the electroneutrality condition in the zones is valid in the form  $c_{R,L} = c_L$ ,  $c_{R,A} = c_A$ , the application of Kohlrausch regulation function to zones  $L$  and  $A$  gives the explicit equation for the calculation of the concentration of an ion  $A$  in its zone adjusted according to the composition of the



leading zone:

$$c_A = c_L \frac{u_A |u_L| + |u_R|}{u_L |u_L| + |u_R|} \quad [8]$$

Obviously, the adjusted concentration  $c_A$  is independent of the original concentration of A in the sample. As a result of the adjustment of the sample concentration, a diluted sample is concentrated or a concentrated one is diluted when passing through the stationary boundary. The concentration of diluted samples by this ITP effect is commonly named stacking, see also later.

## Qualitative and Quantitative Analysis

The above-mentioned adjustment of concentrations and the same velocity of all zones have serious consequences for the qualitative and quantitative analyses. Its principle is different not only from that of zone electrophoresis but also from that of common elution chromatographic techniques. Thus, the separation of substances and the analytical interpretation of records are specific to this technique. In **Figure 2C**, a draft of a record from a universal detector of the analysis of a mixture of two ions is shown. Response on y-axis corresponds to resistance in the zone when a conductivity detector is used or to potential gradient when a potential gradient detector is used. The step height  $h$  reflects the reciprocal value of the effective mobility of ions in their zone and represents the qualitative value. The step length corresponds to the time when the given zone passes through the detector and therefore is proportional to the zone size. The step length is thus the quantitative value. In **Figure 2D**, the same mixture is detected with an optical, i.e., a selective detector. The step height is proportional to extinction coefficient in this case; step length, or better the area of the step, corresponds to the amount of the analyte and is used for quantitative evaluation. When the zone length is too short to create the plateau in the record and an optical detector is used, then only peaks (called spikes) are registered. Both their height and area are proportional to the concentration of the analyte in its zone and quantitative evaluation is still possible.

The practical qualitative analysis means identification of the steps in the record. It is mostly based on comparison between the analysis of standard substances and of sample components under the same conditions. If the standard substance forms a step in the record that is different from all the step heights in the sample, then this substance is not present in the sample. If the step height for the standard is identical to that for one of the components, then this substance may be present in the sample. When simultaneous

recording is performed with a selective detector and a universal one, e.g., conductivity and ultraviolet (UV) detectors, two independent data values are obtained for a single zone. If these values for the analyte in the sample and the standard are the same then the substance is most probably present in the sample.

The practical quantitative analysis is based on measuring the zone lengths,  $l$ , see **Figure 2C** and **D**. In practice, the length of the step in the record is measured, and calibration graphs are employed. Calibration involves plotting the amount of the standard (in moles) against the step length of the given zone in the form

$$N_X = kl_X \quad [9]$$

where  $N_X$  is the amount of the substance X injected,  $l_X$  is the step length in the record, and the constant  $k$  includes the operational parameters of the analysis. In practice, a blank must always be run to reveal the isotachophoretic boundary between the leading and terminating electrolytes. Often even analytical grade reagents contain impurities that also form zones (e.g., sodium in potassium acetate) that can coincide with the analyte zones.

There is still one important item to be mentioned here. If too much of a sample is injected, then the separation is not completed prior to detection and transient mixed zones are detected and registered. Decrease in the sample amount shortens the mixed zone until complete separation is achieved prior to detection. There is, however, still a potential complication that stems from the existence of stable mixed zones, i.e., zones that contain nonseparable substances irrespective of the injected sample amount. These zones can be avoided only by using another electrolyte system that ensures sufficient separability of sample components, see later. Obviously, correct quantification assumes complete separation, i.e., the absence of mixed zones.

## Instrumentation

For single isotachophoretic separation, usually a capillary made from fused silica or a suitable organic polymer of 0.1–0.8 mm ID is used. The ends of the capillary are connected to electrode chambers containing leading and terminating electrolytes, and platinum electrodes for connecting the high-voltage source. The high-voltage source provides a constant direct current up to 500  $\mu$ A and voltage up to 20 kV. A semipermeable membrane placed at the connection of the capillary and the leading electrolyte in the electrode chamber prevents hydrodynamic flow of the electrolyte. The part of the capillary between the chamber with the leading electrolyte and the

injection point is filled with the leading electrolyte; the part between the injection point and the chamber with the terminating electrolyte is filled with the terminating electrolyte. The sample (usually 1–30  $\mu\text{l}$ ) is injected either by microsyringe or through the valve. A detector detects separated zones and a PC processes the response. The most frequent detectors used in ITP are universal conductivity or potential gradient detectors or selective optical detectors, such as UV absorbance detector. It should be mentioned that simple isotachopheretic separations can be performed also in instrumentation used for CZE. It should be stressed, however, that in these apparatuses usually the capillary is not separated from the electrode chambers with a membrane. This is the so-called open system and the electroosmotic flow, if present, influences the migration velocity and direction of analytes in contrast to the closed system where only dispersion of zones can be expected.

Coupling of two capillaries of wider and narrower diameter connected by a bifurcation block enables to

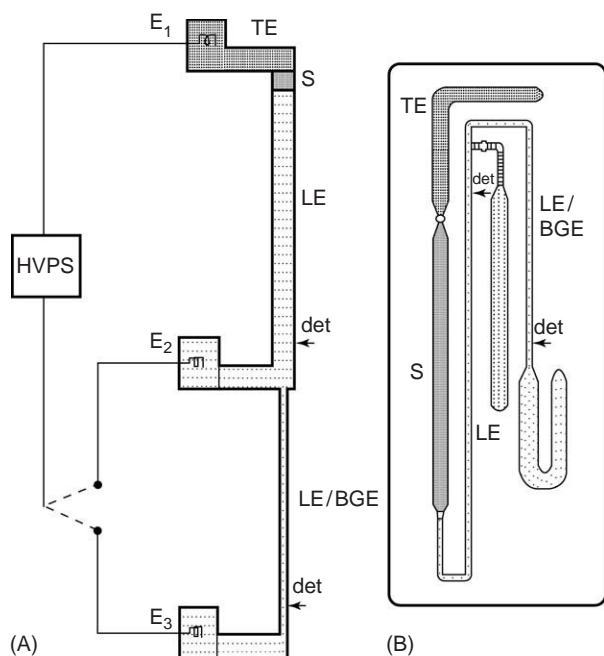
perform more sensitive analyses with high loads. The scheme is shown in **Figure 3A**. In the first capillary, separation of a larger volume of the sample occurs and bulk components are driven to the auxiliary electrode. The analysis of interesting analytes then proceeds online in the analytical capillary of narrower diameter. Thus, higher sensitivity is reached resulting from the fact that the same zone volume in the capillary of a narrower diameter is longer and thus migration through the detector is longer. The system of column coupling is also used in miniaturized systems on chips made from suitable organic polymers. **Figure 3B** shows a scheme for one of the arrangements used on chips. The same principle of arrangement can be used also for the hyphenation of ITP and zone electrophoresis. The only difference is that the analytical capillary is filled with a suitable background electrolyte (BGE) where analytes can migrate in the CZE mode.

## Electrolyte Systems and Separation

The key problem to obtain correct results in ITP both for qualitative and quantitative analysis consists in the selection of a suitable electrolyte system, in which the analytes are completely separated and form stable isotachopheretic zones. Electrolytes are selected on the basis of data on the effective mobilities of the relevant substances and on ways of changing these effective mobilities by employing variable factors, especially the pH and complexation reactions. This variation is based on the definition of the effective mobility  $\bar{u}$ , which permits calculation of the effective mobility of the substance under the given conditions in the presence of acid–base and complexation equilibria. The tabulated values of the effective mobilities for a series of tested electrolyte systems over a wide range of pH values of the leading electrolyte are very useful.

Basic rules for the selection of the electrolyte system are listed in **Table 1**.

The leading ion is usually a suitable fast ion of a completely dissociated substance, e.g., chloride (ionic mobility  $-79.1 \times 10^{-9} \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$ ). The terminating ion is usually  $\text{H}_3\text{O}^+$  for cations or  $\text{OH}^-$  for anions, which must always form the last zone in the system (at least theoretically). If the terminating ion consisting of  $\text{H}_3\text{O}^+$  ( $\text{OH}^-$ ) is too slow (the electric potential gradient is too high), it is preferable to use another faster substance as the terminating ion having usually the effective mobility higher than  $10 \times 10^{-9} \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$ . **Table 2** shows ionic mobilities and  $\text{pK}$  values of several cations and anions. The effective mobility can easily be calculated for the given pH of the milieu by using the eqn [3].



**Figure 3** Scheme of instrumentation for isotachopheresis in column-coupling arrangement. (A) The capillary version. (B) Arrangement on a chip made from an organic polymer. LE, leading electrolyte, TE, terminating electrolyte, BGE, background electrolyte,  $E_1$ ,  $E_2$ ,  $E_3$ , electrodes, HVPS, high-voltage power source, S, sampling point valve or channel, det, detector. (A) (Reprinted with permission from Křivánková L and Boček P (1998) High performance capillary electrophoresis. In: Khaledi MG (ed.) *Capillary Isotachopheresis*, vol. 146, p. 261. Wiley.) (B) (Reprinted with permission from Bodor R, Kaniánsky D, Masár M, Silleová K, and Stanislawski B (2002) Determination of bromate in drinking water by zone electrophoresis – isotachopheresis on a column-coupling chip with conductivity detection, *Electrophoresis*, vol. 23, p. 3631. Wiley-VCH.)

**Table 1** Basic rules for composition of isotachophoretic electrolyte system

Analysis	Anionic	Cationic
Leading ion (3–20 mmol l <sup>-1</sup> )	Cl <sup>-</sup> , NO <sub>3</sub> <sup>-</sup>	K <sup>+</sup> , Na <sup>+</sup> , NH <sub>4</sub> <sup>+</sup>
Terminating ion	OH <sup>-</sup> or weak acid	H <sub>3</sub> O <sup>+</sup> or weak base
Counterion	Weak base with $pK_{BH} = pH_L \pm 0.5$	Weak acid with $pK_{HA} = pH_L \pm 0.5$
pH <sub>L</sub>	$pH \geq pK_{HA}(\text{analyte}) - 1$	$pH \leq pK_{BH}(\text{analyte}) + 1$

pH<sub>L</sub> = pH of the leading electrolyte.

**Table 2** Ionic mobilities and dissociation constants of some ions at 25°C

Ion	Ionic mobility (10 <sup>-9</sup> m <sup>2</sup> V <sup>-1</sup> s <sup>-1</sup> )	pK
H <sub>3</sub> O <sup>+</sup>	+ 362.5	
OH <sup>-</sup>	- 205.5	
K <sup>+</sup>	+ 76.2	
NH <sub>4</sub> <sup>+</sup>	+ 76.1	9.25
Imidazole	52	7.15
Na <sup>+</sup>	+ 51.9	
Ethanolamine (+ 1)	+ 44.3	9.49
Lithium	+ 40.1	
Creatinine	+ 33.1	4.89
Histidine (+ 1)	+ 29.6	6.04
o-Chloraniline	+ 25	2.66
Tetrabutylammonium	+ 16.9	
OH <sup>-</sup>	- 205.5	
Cl <sup>-</sup>	- 79.1	
Nitrate	- 74.1	
Sulfate (- 1)	- 52.0	
Sulfate (- 2)	- 82.9	1.92
Pyruvate	- 42.3	2.49
Lactate	- 36.5	3.86
Phosphate (- 1)	- 35.1	2.12
Phosphate (- 2)	- 61.5	7.47
Phosphate (- 3)	- 71.5	12.36
Methionine	- 29.3	9.34
Hippurate	- 26.1	3.7
HEPES	- 21.8	7.51
LeuGlyPhe	- 19.3	7.94

HEPES = *N*-(2-Hydroxyethyl)piperazine-*N'*-(2-ethanesulfonic acid).

If the separated substances are weak bases or acids, proper selection of the pH of the leading electrolyte and pK<sub>a</sub> of the counterion such that the conditions given in the Table 1 are approximately fulfilled should ensure sufficient ionization. Examples of electrolyte systems for ITP are shown in Table 3.

Another problem is to determine whether the substance of interest forms a stable individual zone in the selected system, i.e., the composition and volume

**Table 3** Examples of electrolyte systems for isotachophoresis

pH <sub>L</sub>	Counterion	Terminating ion
<i>Anionic analysis</i>		
3.1–4.1	β-Alanine	Propionate
4.1–5.1	EACA	Pivalate
4.5–5.5	Creatinine	MES
5.5–6.5	Histidine	HEPES
6.6–7.6	Imidazole	TES
7.6–8.6	Tris	TAPS
8.3–9.3	Ammediol	Glycine
9.0–10.0	Ethanolamine	EACA
<i>Cationic analysis</i>		
3.2–4.2	Formate	H <sub>3</sub> O <sup>+</sup> (formic acid)
4.2–5.2	Acetate	H <sub>3</sub> O <sup>+</sup> (acetic acid)
5.7–6.7	MES	Histidine
6.9–7.9	Veronal	Imidazole
8.3–9.3	Asparagine	Tris
9.1–10.1	Glycine	Triethanolamine
9.8–10.8	β-Alanine	Ethanolamine

Ammediol = 2-amino-2-methyl-1,3-propanediol; EACA = ε-amino-caproic acid; HEPES = *N*-(2-hydroxyethyl)piperazine-*N'*-(2-ethanesulfonic acid); MES = morpholinoethanesulfonic acid; TAPS = *N*-[tris(hydroxymethyl)methyl]-3-aminopropanesulfonic acid; TES = [tris(hydroxymethyl)methyl]-2-aminoethanesulfonic acid; Tris = tris(hydroxymethyl)aminomethane; Veronal = 5,5-diethylbarbituric acid.

of a zone do not change with time. Experimental verification of zone stability is very simple and consists in testing whether the calibration graph constructed for the substance passes through the origin. Theoretical considerations are based on effective mobilities; for the stable zone of a substance X it must hold

$$|\bar{u}_{T,T}| < |\bar{u}_{X,T}| \quad \text{and} \quad |\bar{u}_{X,X}| < |\bar{u}_{L,X}| \quad [10]$$

which means that the effective mobility of the substance X in the terminating zone,  $|\bar{u}_{X,T}|$ , is higher than the effective mobility of the terminator in its own zone,  $|\bar{u}_{T,T}|$ , and the effective mobility of the leading ion in the zone of the substance X,  $|\bar{u}_{L,X}|$ , is higher than the effective mobility of the substance X in its own zone,  $|\bar{u}_{X,X}|$ .

The third task is to verify whether the substances of interest are separable, i.e., whether they form their individual zones separated from each other. Experimental verification of the separability is simple but laborious and is based on measuring calibration graphs for standard solutions of individual substances and for their mixtures. For theoretical evaluation, it is not sufficient to compare the tabulated or calculated effective mobilities of the substances in their own zones (parameters of the  $\bar{u}_{i,i}$  type) since some pairs of substances of the same effective mobilities in their own zones can be completely separated

and migrate separately while other pairs of different effective mobilities in their own zones do not separate. It must be based on evaluation of separability of substances on the basis of parameters of  $\bar{u}_{i,j}$  (effective mobility of an analyte  $i$  in the neighboring zone  $j$ ) and is identical with the principle of unambiguous determination of the migration order of substances. If the migration order can be determined unambiguously for a pair of substances  $i$  and  $j$  in a given electrolyte system, then substances  $i$  and  $j$  can be separated. A simple criterion can be employed to evaluate separability based on the migration order. Two cases representing separation are possible:

1. the zone of a substance  $i$  migrates ahead of the zone of a substance  $j$  if

$$|\bar{u}_{i,j}| > |\bar{u}_{j,j}| \quad \text{and} \quad |\bar{u}_{j,i}| < |\bar{u}_{i,i}| \quad [11]$$

2. the zone of a substance  $j$  migrates ahead of the zone of a substance  $i$  if

$$|\bar{u}_{i,j}| < |\bar{u}_{j,j}| \quad \text{and} \quad |\bar{u}_{j,i}| > |\bar{u}_{i,i}| \quad [12]$$

If conditions [11] and/or [12] are not valid for a given pair of substances then a third case takes place where

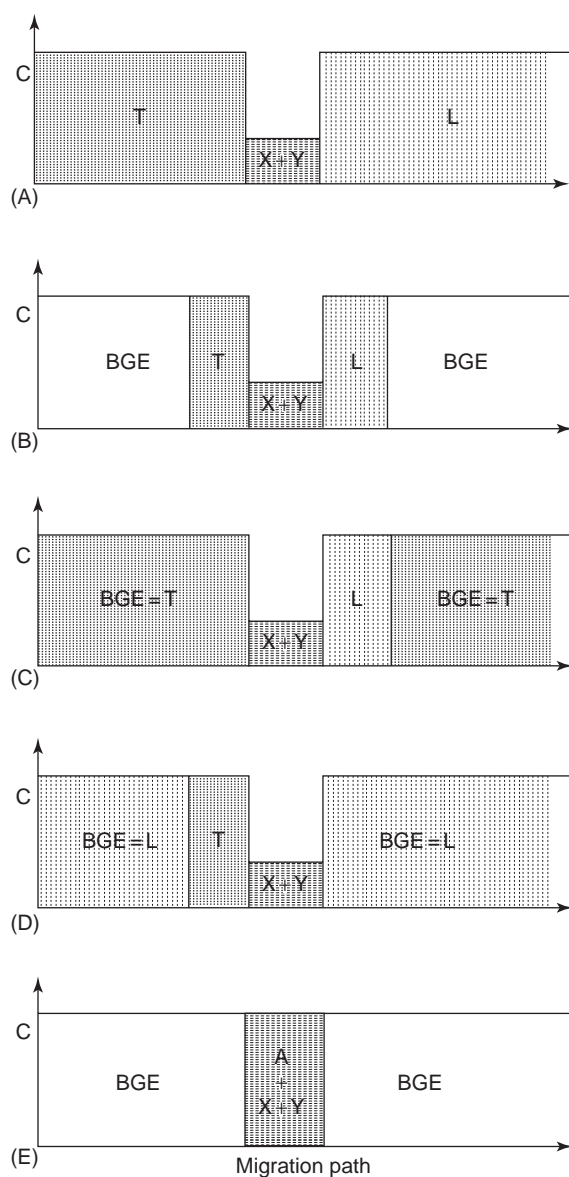
$$|\bar{u}_{i,j}| < |\bar{u}_{j,j}| \quad \text{and} \quad |\bar{u}_{j,i}| < |\bar{u}_{i,i}| \quad [13]$$

Here, the migration order cannot be determined unambiguously and the two substances form a stable mixed zone and cannot be separated.

## Isotachophoretic Stacking and Combination ITP-CZE

### Single Capillary System

Arrangement for ITP, i.e., the discontinuous electrolyte system where mobilities of sample components are lower than the mobility of the leading ion, and higher than the mobility of terminating ion, is usually created prior to the separation in an instrumentation with a capillary where only a leading electrolyte, a terminating electrolyte, and a sample plug were injected, **Figure 4A**. Isotachophoretic stacking power, however, can also be used within zone electrophoretic separation. In such a case, the capillary is filled with a BGE of a constant composition and prior and after the sample plug, zones of leading electrolyte and terminating electrolyte of a proper composition and length are loaded, **Figure 4B**. If the co-ion of the



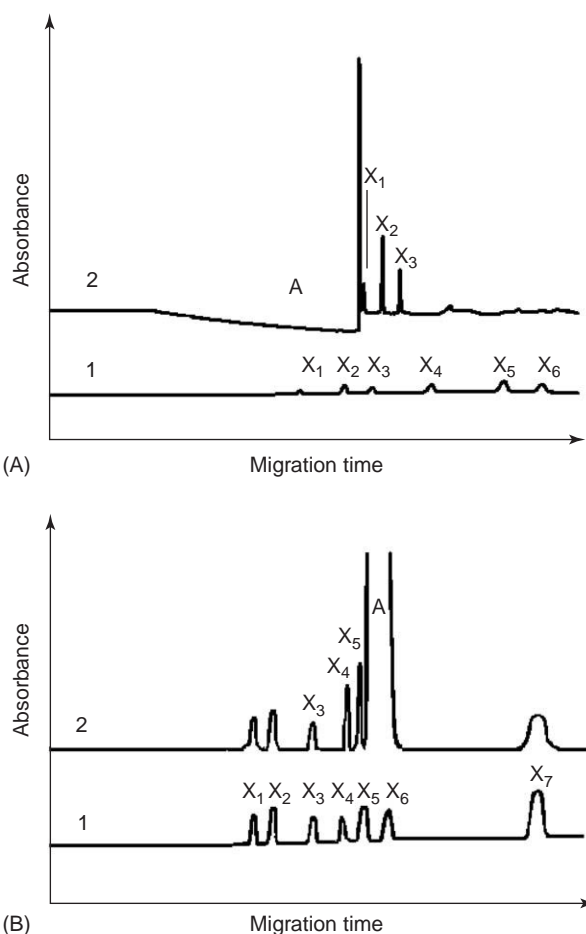
**Figure 4** Arrangement of discontinuous electrolyte systems for isotachopheresis. (A) The capillary is filled with the leading electrolyte containing a leading ion L, the sample plug X + Y is followed with the terminating electrolyte with a terminating ion, T. (B) The capillary is filled with a background electrolyte, BGE, then a plug of a leading electrolyte is injected followed with the sample plug and the plug of a terminating electrolyte. The end of the capillary is immersed again into the BGE solution. (C) As in B, but BGE plays the role of the terminating electrolyte and thus only a plug of a leading electrolyte has to precede the sample zone. (D) As in B, the BGE co-ion is the fastest in the system and ensures the role of the leading ion. The injection of the sample plug is followed with the injection of the zone of a terminating electrolyte prior the capillary is immersed into the solution of BGE. (E) The role of an isotachophoretic stacker is ensured by presence of a macrocomponent A in the sample. The stacker can be faster than analytes in the sample and plays then the role of the leading ion. The BGE co-ion has the function of the terminator. If the macrocomponent is slower than the analytes, it can act as the terminator and simultaneously, the BGE co-ion acts as the leading ion.



BGE is selected in such a way that it can play the role of a terminating ion for a sample, then only the zone of a suitable leader has to be injected in front of the sample zone, **Figure 4C**. Similarly, if the co-ion of the BGE can serve as a leading ion, then the sample plug is followed by the zone of terminator only and the separation proceeds, **Figure 4D**.

In all above cases, the ITP migration is only temporary and is called transient ITP. It has, however, great analytical significance. The great advantage of involving transient ITP into the CZE separation process is enhancement of sensitivity due to the isotachopheretic concentrating effect. The analytes present in the sample in low concentrations are stacked into narrow zones of the concentration corresponding to the Kohlrausch  $\omega$  function. As can be seen from **Figure 4E**, a suitable arrangement for transient ITP can arise also when the sample itself contains macrocomponents that can act as isotachopheretic stackers. When the BGE is constructed so that the co-ion can act as the terminator and a macrocomponent A present in the sample in a concentration higher than a minimum needed value can play the role of the leading ion for an analyte of interest, X, i.e.,  $|u_A| > |u_X| > |u_{BGE}|$ , see **Figure 5A**, or vice versa the BGE co-ion is the fastest ion in this transient ITP mode and acts as the leader and a macrocomponent has the role of the terminator, i.e.,  $|u_{BGE}| > |u_X| > |u_A|$ , see **Figure 5B**, the analyte is concentrated with respect to the concentration of the transient leader according to the Kohlrausch regulation function. This is a very frequent case in biological samples, where, e.g., a bulk concentration of NaCl ensures the presence of a fast ion possessing properties of the leader for both anions (chloride) and cations (sodium), and arrangement for transient ITP can simply be created by the proper choice of the composition of the BGE, especially as far as mobility and the concentration of the BGE co-ion are concerned. Analytes stacked during transient ITP between the leading type (terminating type) stacker and the BGE co-ion are leaving the ITP stack gradually. The analyte with the mobility closest to that of the stacker migrates in the ITP mode the longest and its zone is the sharpest, as can be seen also in **Figure 5**. In this case, a sufficiently long migration path is recommended to ensure destacking of all analytes from ITP mode and to obtain suitable resolution so as to employ also the advantage of CZE migration.

If more macrocomponents are present in a complex sample matrix, generally two different situations can be distinguished. First, macrocomponents can act as stackers of either leading or terminating type and the resulting effect is a sum of their stacking



**Figure 5** Illustration of isotachopheretic effect in CZE. (A) A macrocomponent present in the sample plays the role of the leading ion. Trace 1, analysis of six analytes of  $10^{-5} \text{ mol l}^{-1}$  concentration in a BGE with the mobility of the co-ion equal to that of the analyte  $X_4$ . Trace 2, analysis of the same sample containing a macrocomponent A with the mobility higher than that of the fastest ion  $X_1$  at the concentration  $10^{-2} \text{ mol l}^{-1}$ . Conditions for isotachopheretic stacking of analytes with mobilities  $|u_A| > |u_X| > |u_{BGE}|$  are ensured, the macrocomponent acts as the leader and BGE co-ion as the terminator, enhancement of analyte concentration in their zones reflects in sharpening the peaks. (B) A macrocomponent present in the sample plays the role of the terminating ion. Trace 1, analysis of seven analytes of  $10^{-5} \text{ mol l}^{-1}$  concentration in a BGE with the mobility of the co-ion equal to that of the analyte  $X_3$ . Trace 2, the concentration of the analyte  $X_6$  is increased to  $5 \times 10^{-3} \text{ mol l}^{-1}$ , conditions for isotachopheretic stacking of analytes with mobilities  $|u_{BGE}| > |u_X| > |u_A|$ , i.e., for  $X_4$  and  $X_5$  are ensured with the macrocomponent A acting as the terminator and BGE co-ion as the leader. The peaks of the isotachopheretically stacked analytes were sharpened, sensitivity was increased. (Reprinted with permission from *Journal of Chromatography A* (2001), vol. 916, p. 173. Copyright (2001), © Elsevier.)

effects. Second, one macrocomponent acts as the stacker (A) and the other one acts against stacking as the so-called destacker (B). This happens when  $|u_A| > |u_X| > |u_B| > |u_{BGE}|$ ,  $|u_A| > |u_X| > |u_B| = |u_{BGE}|$

or  $|u_A| > |u_X| > |u_{BGE}| > |u_B|$  in the leading type of stacking, and similarly for the terminating type of stacking when  $|u_{BGE}| > |u_B| > |u_X| > |u_A|$  or  $|u_{BGE}| = |u_B| > |u_X| > |u_A|$  or  $|u_B| > |u_{BGE}| > |u_X| > |u_A|$ , and simultaneously, the concentration ratio of the stacker to the destacker is lower than a critical value,  $c_A/c_B < a_{crit}$ , where for  $a_{crit}$  it holds

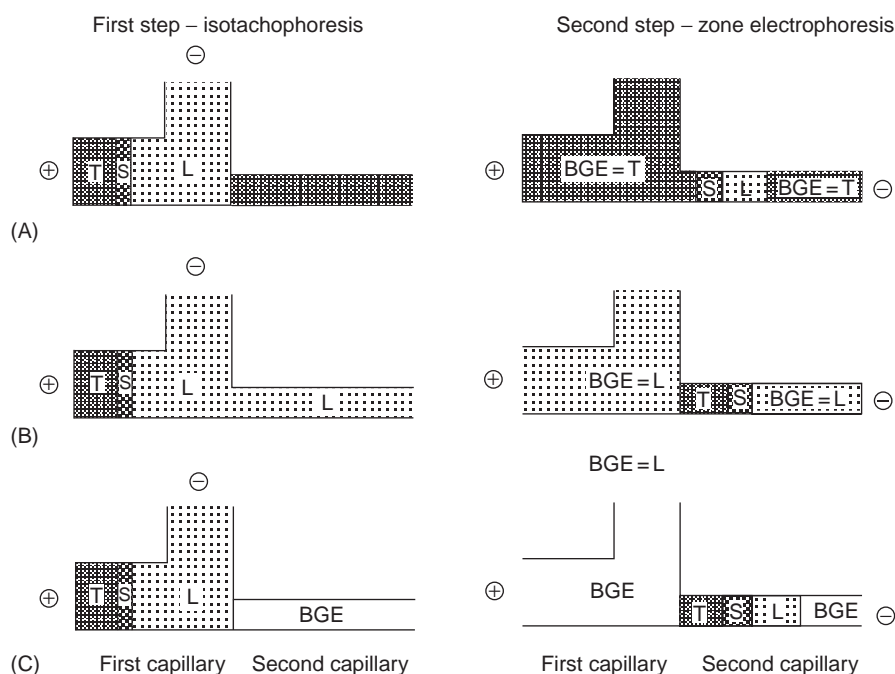
$$a_{crit} = \frac{|u_B| + |u_R|}{|u_A| + |u_R|} \cdot \frac{|u_A| - |u_X|}{|u_X| - |u_B|} \quad [14]$$

When  $c_A/c_B > a_{crit}$ , the stacker has a sufficient potential to ensure stacking of analytes the mobilities of which are lying between  $|u_A|$  and  $|u_B|$  (for  $|u_B| > |u_{BGE}|$  or  $|u_B| = |u_{BGE}|$ ) or between  $|u_A|$  and  $|u_{BGE}|$  (for  $|u_B| < |u_{BGE}|$ ) for the leading type of stacking and analogously in the reversed way for the terminating type of stacking. It should be stressed here that not only a sample macro component can act as the destacker but also the BGE co-ion when present in the sample acts against stacking.

### Column Coupling System

The instrumentation for column coupling arrangement has already depicted in **Figure 3A** and **B**. This arrangement offers not only stacking but also elimination of interfering bulk components, provided that suitable combination of electrolyte systems is used. The combinations of electrolytes for both capillaries in the ITP–CZE arrangement are shown in **Figure 6**.

The simplest is the first combination where the terminator from ITP step is used also as the BGE in the second step. In this configuration, the arrangement for the analysis is set completely before the analysis starts and the only task to do is to switch the voltage across the system of both capillaries (see  $E_1$  and  $E_3$  in **Figure 5A**) at a proper time. This prevents the loss of an interesting part of the sample by its migration to the auxiliary electrode in the reservoir of the leading electrolyte in the first capillary (see  $E_2$  in **Figure 5A**). In both hyphenations, ITP–ITP and ITP–CZE, the proper switching of current between



**Figure 6** Combination of electrolyte systems for the hyphenation of capillary isotachopheresis with zone electrophoresis in column coupling instrumentation. (A) T–S–T system, where the separation in the CZE step proceeds in the terminating electrolyte from the ITP step. A short zone of the leading electrolyte was allowed to enter the second capillary to avoid loss of sample components and responds for temporary survival of isotachophoretic mode of migration. The result is enhanced separation efficiency of faster sample components. (B) L–S–L system, where the separation in the CZE step proceeds in the leading electrolyte from the ITP step. A short zone of the terminating electrolyte was allowed to enter the second capillary to avoid loss of sample components and responds for temporary survival of isotachophoretic mode of migration. The result is enhanced separation efficiency of slower sample components. (C) BGE–S–BGE system, where the separation in the CZE step proceeds in a background electrolyte different from leading and terminating electrolytes. Short zones of leading and terminating electrolytes were allowed to enter the second capillary to avoid loss of sample components and respond for temporary survival of isotachophoretic mode of migration. The result is enhanced separation efficiency of faster and slower sample components.



electrodes  $E_2$  and  $E_3$  enables very precise cleaning of the sample resulting in a shorter time of analysis and more sensitive and reliable results.

## Applications

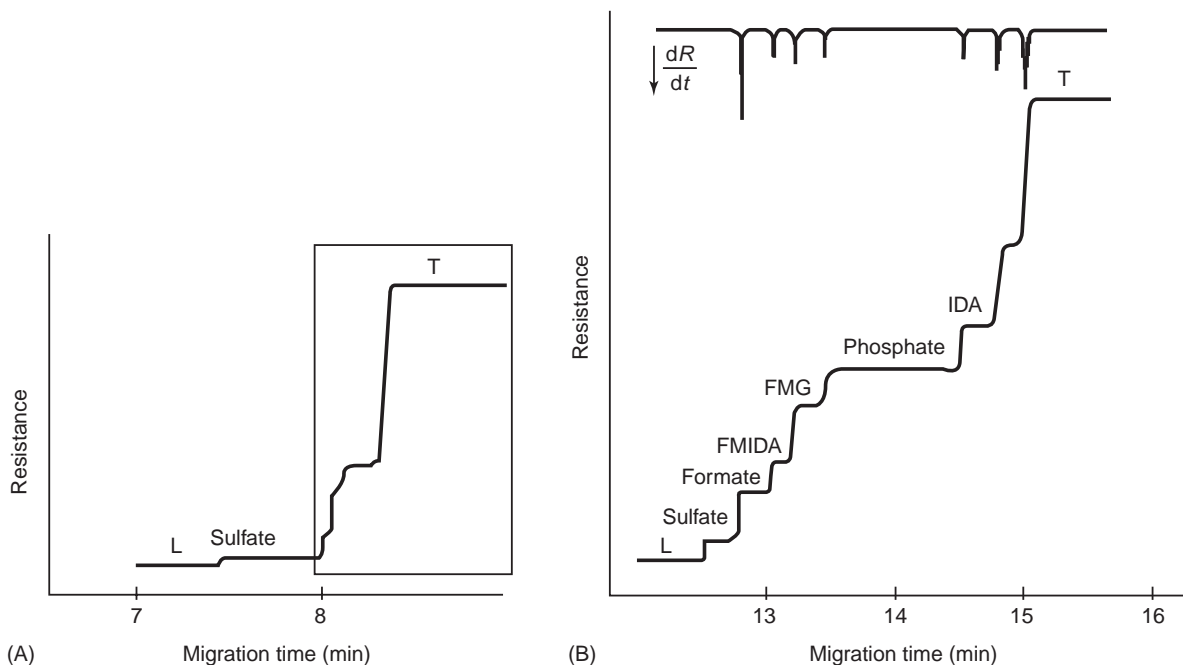
A basic characteristic of ITP is the limitation of its use to ionogenic substances. In a single run either anions or cations are analyzed; however, electrolyte systems that enable simultaneous analysis of both groups can be found and so called bi-directional ITP employing two detectors can be performed. Non-ionogenic substances that often form the bulk of a sample do not interfere with the analysis. Advantages of ITP include its independence of derivatization and deprotection of the sample and this feature can be very effectively evaluated when ITP is used as a pre-separation method. ITP is mostly carried out in aqueous solutions or in aqueous mixtures with solvents and therefore it is most useful for substances that are water-soluble or can be included into water-soluble complexes, e.g., with micelles of surfactants.

Most of the practical applications of ITP are found in the analysis of ionogenic organic substances,

especially in biochemistry, industrial chemistry, and environmental analysis, as reflected in the examples given below.

**Figure 7** demonstrates an analysis of 5  $\mu\text{l}$  of 100 times diluted mother liquor from the synthesis of herbicide glyphosate (*N*-phosphonomethylglycine). In **Figure 7A**, the record from the pre-separation capillary is shown and the part of the sample that is transferred to the second narrower capillary is denoted. The zone of sulfate is removed by migrating to the auxiliary electrode in the first step. For more precise reading of the step length, the recorded curve from the second capillary, **Figure 7B**, is differentiated and inflex points of the curve are read as distances between maxima of the differential record,  $dR/dt$ . The leading electrolyte was  $0.01 \text{ mol l}^{-1}$  HCl + histidine, pH 6.0, the terminating electrolyte was  $0.01 \text{ mol l}^{-1}$  morpholinoethanesulphonic acid (MES).

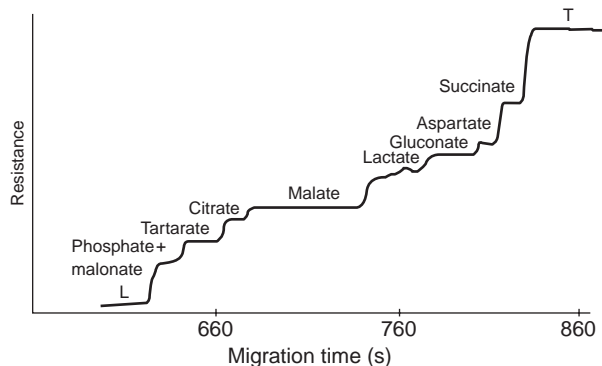
In **Figure 8**, an example of an isotachopheretic separation on a poly(methyl methacrylate) chip is shown demonstrating an analysis of organic acids and inorganic ions present in red wine. A 0.9  $\mu\text{l}$  (the fixed volume of the sampling channel) of 20 times diluted red wine was analyzed using of  $10 \text{ mmol l}^{-1}$



**Figure 7** ITP–ITP analysis of 5  $\mu\text{l}$  of 100 times diluted sample of the mother liquor from the synthesis of herbicide glyphosate (*N*-phosphonomethylglycine, FMG). (A) Preseparation capillary, ID 0.8 mm. The zone of sulfate is transported to the auxiliary electrode  $E_2$ , see **Figure 5**, and by proper timing of switching current between electrodes  $E_1$ – $E_2$ , and  $E_1$ – $E_3$ , the demarcated part is transferred to the second capillary. (B) The record from conductivity detector in the second capillary, ID 0.3 mm. For more precise reading of the step length, the recorded curve from conductivity detector is differentiated and inflex points of the curve are read as distances between maxima of the differential record,  $dR/dt$ . FMIDA, *N*-phosphonomethyliminodiacetic acid; IDA, iminodiacetic acid; L, leading ion (chloride); T, terminating ion (MES). The composition of electrolytes was: LE  $0.01 \text{ mol l}^{-1}$  HCl + histidine, pH 6.0, TE:  $0.01 \text{ mol l}^{-1}$  morpholinoethanesulphonic acid (MES). (Reprinted with permission from Křivánková L and Boček P (1986) Analytical control of the production of herbicides and growth regulators glyphosate and glyphosine by capillary isotachopheresis, *Electrophoresis* 7: 102; © Wiley-VCH.)

HCl +  $\beta$ -alanine, pH 2.9, with 0.1% w/v methylhydroxycellulose as the leading electrolyte and  $5 \text{ mmol l}^{-1}$  glutamic acid + histidine, pH 5.0, as the terminating electrolyte.

Figure 9 shows the application of the hyphenation technique ITP-ITP for the analysis of  $5 \mu\text{l}$  of untreated heparin plasma as a method to follow changes in concentrations of pyruvate, acetoacetate, lactate, and  $\beta$ -hydroxybutyrate in plasma of patients with diabetes mellitus. In the column switching

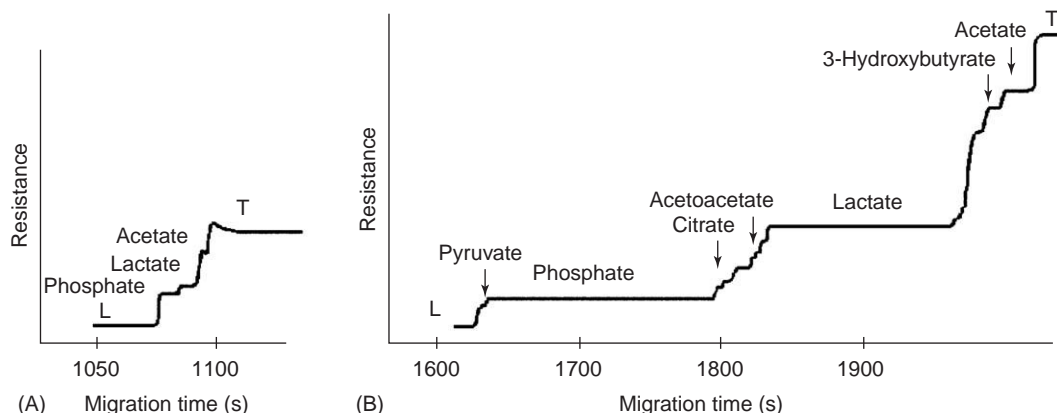


**Figure 8** Isotachopheretic separation of organic acids and inorganic ions present in red wine on a poly(methylmethacrylate) chip. A  $0.9 \mu\text{l}$  (the fixed volume of the sampling channel) of 20 times diluted red wine sample was analyzed using of  $10 \text{ mmol l}^{-1}$  HCl +  $\beta$ -alanine, pH 2.9, with 0.1% w/v methylhydroxycellulose as the leading electrolyte and  $5 \text{ mmol l}^{-1}$  glutamic acid + histidine, pH 5.0, as the terminating electrolyte. L, leading ion – chloride; T, terminating ion – glutamate. (Reprinted with permission from Masár M, Kaniánský D, Bodor R, Jöhnnck M, and Stanislawski B (2001) Determination of organic acids and inorganic anions in wine by isotachopheresis on a planar chip. *Journal of Chromatography A*, vol. 916, p. 173; © Elsevier.)

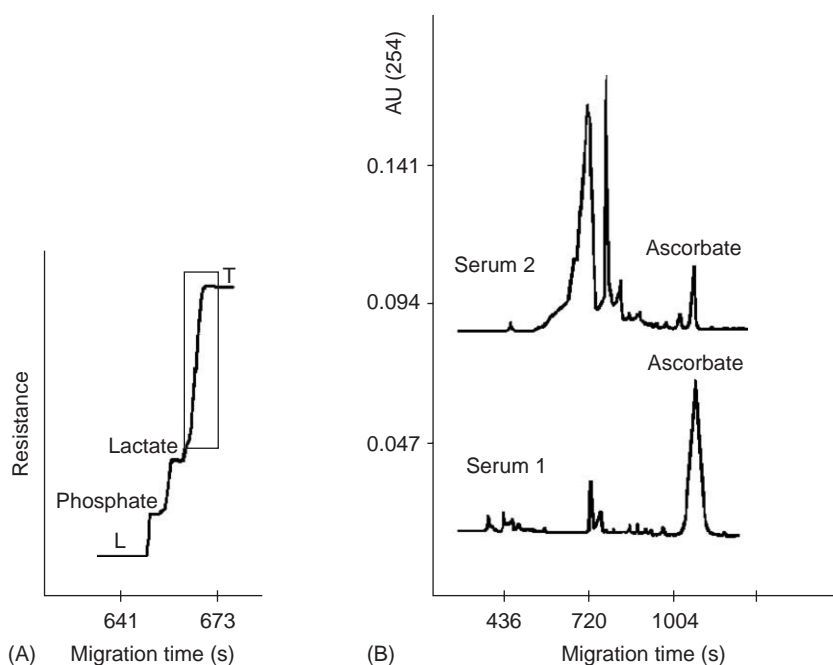
system,  $10 \text{ mmol l}^{-1}$  HCl +  $\beta$ -alanine, pH 4.2, was used as the leading electrolyte for the pre-separation stage, and  $5 \text{ mmol l}^{-1}$  HCl + glycine, pH 3.0, was used as the leading electrolyte for the analytical stage. A volume of  $10 \text{ mmol l}^{-1}$  nicotinic acid served as the terminating electrolyte.

In Figure 10, hyphenation of ITP with zone electrophoresis is demonstrated for an analysis of ascorbic acid in serum. Figure 10A shows the isotachopheretic step of a serum sample analyzed in the electrolyte system: leading electrolyte:  $10 \text{ mmol l}^{-1}$  HCl +  $\beta$ -alanine, pH 3.3; terminating electrolyte:  $10 \text{ mmol l}^{-1}$  propionic acid. Only components migrating behind lactic acid are driven to the second capillary filled with  $50 \text{ mmol l}^{-1}$  propionic acid adjusted with  $\beta$ -alanine to pH 3.8 (corresponding to the pH of terminating electrolyte adjusted according to Kohlrausch equation) serving as BGE for the CZE step. This is the combination of electrolytes denoted as T-S-T, cf. Figure 6. The record in Figure 10B shows an excellent reproducibility of migration time of two samples from different persons.

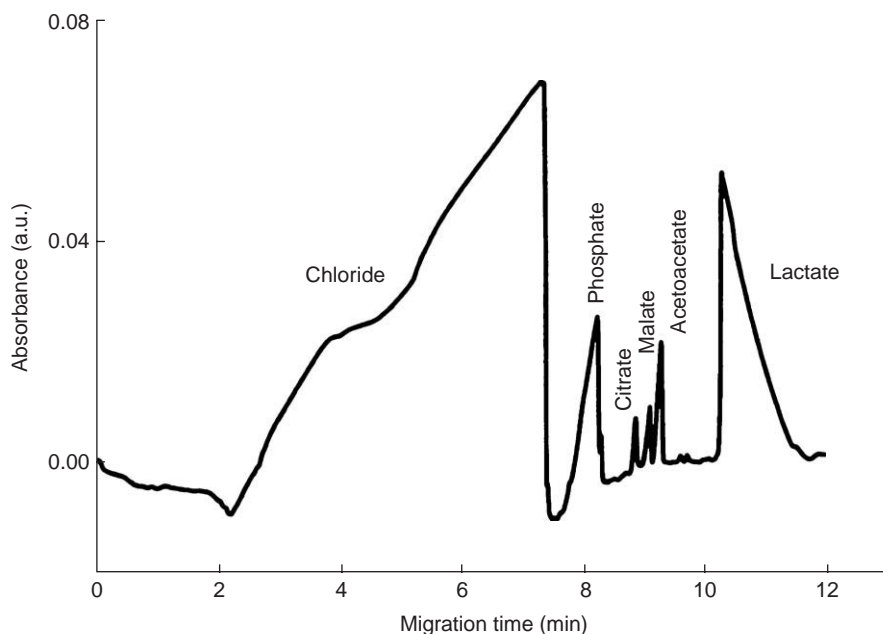
Figure 11 shows an analysis of a human serum by capillary zone electrophoresis with inherent transient isotachopheretic stacking of analytes. The excess of chloride present naturally in the sample acts as the stacker for analytes citrate, malate, and acetoacetate with mobilities lower than the mobility of the BGE co-ion, mandelate, in the BGE composed of  $5 \text{ mmol l}^{-1}$  mandelic acid +  $\epsilon$ -aminocaproic acid, pH 3.8. Lactate has lower mobility than the BGE co-ion; however, the predominating stacking effect of chloride eliminates its destacking effect.



**Figure 9** ITP-ITP analysis of  $5 \mu\text{l}$  of untreated heparin plasma of patients with diabetes mellitus. (A) The record from the conductivity detector in the pre-separation capillary, ID  $0.8 \text{ mm}$ , LE  $10 \text{ mmol l}^{-1}$  HCl +  $\beta$ -alanine, pH 4.2. (B) The record from the analytical capillary, ID  $0.3 \text{ mm}$ , LE  $5 \text{ mmol l}^{-1}$  HCl + glycine, pH 3.0. L, leading ion – chloride; T, terminating ion – nicotinate. (Reprinted with permission from Křivánková L and Boček P (1990) Determination of pyruvate, acetoacetate, lactate, and 3-hydroxybutyrate in plasma of patients with diabetes mellitus by capillary isotachopheresis, *Journal of Microcolumn Separation* 2: 83.)



**Figure 10** ITP-CZE analysis of 10  $\mu\text{L}$  of five times diluted untreated serum. (A) ITP step in the preseparation capillary, ID 0.8 mm, LE  $10 \text{ mmol L}^{-1}$  HCl +  $\beta$ -alanine, pH 3.3, TE  $10 \text{ mmol L}^{-1}$  propionic acid. Denoted components migrating behind lactic acid are driven to the second capillary. (B) CZE step in the analytical capillary, ID 0.2 mm, BGE  $50 \text{ mmol L}^{-1}$  propionic acid +  $\beta$ -alanine, pH 3.8. The combination of electrolytes corresponds to the T-S-T<sub>2</sub> system as shown in **Figure 6**. LOD:  $5 \times 10^{-7} \text{ mol L}^{-1}$  ascorbate. Estimated concentration of ascorbic acid in serum 1:  $5.74 \text{ mg L}^{-1}$ ; in serum 2:  $12.15 \text{ mg L}^{-1}$ . (Reprinted with permission from Procházková A, Křivánková L, and Boček P (1998) Quantitative trace analysis of L-ascorbic acid in human body fluids by on-line combination of capillary isotachopheresis and zone electrophoresis. *Electrophoresis*, vol. 10, p. 302; © Wiley-VCH.)



**Figure 11** Analysis of a human serum by capillary zone electrophoresis with inherent transient isotachopheretic stacking of analytes. The BGE co-ion, mandelate, acts as terminator in the system where chloride surplus present naturally in the sample acts as the leading type stacker and conditions for transient isotachopheretic stacking of phosphate, citrate, malate, and acetoacetate with mobilities between chloride and mandelate are ensured. BGE is composed of  $5 \text{ mmol L}^{-1}$  mandelic acid +  $\epsilon$ -aminocaproic acid, pH 3.8. (Reprinted with permission from Křivánková L, Pantůčková P, Gebauer P, *et al.* (2003) Chloride present in biological samples as a tool for enhancement of sensitivity in capillary zone electrophoretic analysis of anionic trace analytes, *Electrophoresis* 24: 515.)

See also: **Capillary Electrochromatography**. **Capillary Electrophoresis**: Overview. **Electrophoresis**: Overview.

## Further Reading

- Boček P, Deml M, Gebauer P, and Dolník V (1988) *Analytical Capillary Isotachopheresis*. Weinheim: VCH Verlagsgesellschaft.
- Gebauer P and Boček P (1997) Recent application and developments of capillary isotachopheresis. *Electrophoresis* 18: 2154–2161.
- Gebauer P and Boček P (2000) Recent progress in capillary isotachopheresis. *Electrophoresis* 21: 3898–3904.
- Gebauer P and Boček P (2002) Recent progress in capillary isotachopheresis. *Electrophoresis* 23: 3858–3864.
- Everaerts FM, Beckers JL, and Verheggen TPEM (1976) *Isotachopheresis, Theory, Instrumentation and Applications*. Amsterdam: Elsevier.
- Hirokawa T, Nishino M, Aoki N, *et al.* (1983) Table of isotachopheretic indices. 1. Simulated qualitative and quantitative indices of 287 anionic substances in the range pH 3–10. *Journal of Chromatography* 271: D1–D100.
- Křivánková L and Boček P (1997) Synergism of capillary isotachopheresis and capillary zone electrophoresis. *Journal of Chromatography B* 689: 13–34.
- Křivánková L, Gebauer P, and Boček P (1995) Some practical aspects of utilizing the on-line combination of isotachopheresis and capillary zone electrophoresis. *Journal of Chromatography A* 716: 35–48.
- Křivánková L, Gebauer P, and Boček P (1996) Isotachopheresis. In: Karger BJ (ed.) *Methods in Enzymology*, vol. 270. *High Resolution Separation Methods of Biological Macromolecules*, Part A, pp. 375–401. New York: Academic Press.
- Křivánková L, Gebauer P, Thormann W, Mosher RA, and Boček P (1993) Options in electrolyte systems for on-line combined capillary isotachopheresis and capillary zone electrophoresis. *Journal of Chromatography* 638: 119–135.
- Křivánková L, Pantůčková P, and Boček P (1999) Isotachopheresis in zone electrophoresis. *Journal of Chromatography A* 838: 55–70.
- Křivánková L, Pantůčková P, Gebauer P, *et al.* (2003) Chloride present in biological samples as a tool for enhancement of sensitivity in CZE analysis of anionic trace analytes. *Electrophoresis* 24: 505–517.
- Urbánek M, Křivánková L, and Boček P (2003) Stacking phenomena in electromigration. From basic principles to practical procedures. *Electrophoresis* 24: 466–485.

## Isoelectric Focusing

P G Righetti, University of Verona, Verona, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Isoelectric focusing (IEF) represents a unique electrokinetic method, in that it is based on steady-state patterns attained by amphoteric species (mostly proteins and peptides) along a pH gradient under the influence of an electric field. Amphoteric compounds migrate until they align themselves at their isoelectric positions, where a dynamic equilibrium is produced between electrokinetic concentrating processes and dispersive effects that include diffusion. Extremely sharp zones are thus obtained, characterized by a very high resolving power. We will deal with conventional IEF in soluble, amphoteric buffers and with immobilized pH gradients (IPG) in insolubilized, nonamphoteric buffers in gel media. Additionally, we will cover IEF in capillaries, the most recent evolution of this technique.

## Conventional Isoelectric Focusing

In principle, pH gradients could be obtained by diffusion of nonamphoteric buffers but such ‘artificial’ gradients would be altered by changes in electric migration and diffusion of the buffer ions. Thus, in 1961, Svensson introduced the concept of ‘natural’ pH gradients, created and stabilized by the electric current itself. The buffers used in this system required two fundamental properties: (1) amphotericism, so that they could reach an equilibrium position along the separation column and (2) ‘carrier’ ability. This latter concept is subtler but just as fundamental. Any ampholyte cannot simply be used for IEF; only a carrier ampholyte (CA), which is a compound capable of transporting the current (a good conductor) and capable of carrying the pH (a good buffer). With this notion, and with Vesterberg’s elegant synthesis of such ampholytes, present-day conventional IEF was born.

## Some Basic Theoretical Concepts

We will survey here some basic equations governing the IEF process. The most important one regards the distribution profile of an ampholyte about its

See also: **Capillary Electrochromatography**. **Capillary Electrophoresis**: Overview. **Electrophoresis**: Overview.

## Further Reading

- Boček P, Deml M, Gebauer P, and Dolník V (1988) *Analytical Capillary Isotachopheresis*. Weinheim: VCH Verlagsgesellschaft.
- Gebauer P and Boček P (1997) Recent application and developments of capillary isotachopheresis. *Electrophoresis* 18: 2154–2161.
- Gebauer P and Boček P (2000) Recent progress in capillary isotachopheresis. *Electrophoresis* 21: 3898–3904.
- Gebauer P and Boček P (2002) Recent progress in capillary isotachopheresis. *Electrophoresis* 23: 3858–3864.
- Everaerts FM, Beckers JL, and Verheggen TPEM (1976) *Isotachopheresis, Theory, Instrumentation and Applications*. Amsterdam: Elsevier.
- Hirokawa T, Nishino M, Aoki N, *et al.* (1983) Table of isotachopheretic indices. 1. Simulated qualitative and quantitative indices of 287 anionic substances in the range pH 3–10. *Journal of Chromatography* 271: D1–D100.
- Křivánková L and Boček P (1997) Synergism of capillary isotachopheresis and capillary zone electrophoresis. *Journal of Chromatography B* 689: 13–34.
- Křivánková L, Gebauer P, and Boček P (1995) Some practical aspects of utilizing the on-line combination of isotachopheresis and capillary zone electrophoresis. *Journal of Chromatography A* 716: 35–48.
- Křivánková L, Gebauer P, and Boček P (1996) Isotachopheresis. In: Karger BJ (ed.) *Methods in Enzymology*, vol. 270. *High Resolution Separation Methods of Biological Macromolecules*, Part A, pp. 375–401. New York: Academic Press.
- Křivánková L, Gebauer P, Thormann W, Mosher RA, and Boček P (1993) Options in electrolyte systems for on-line combined capillary isotachopheresis and capillary zone electrophoresis. *Journal of Chromatography* 638: 119–135.
- Křivánková L, Pantůčková P, and Boček P (1999) Isotachopheresis in zone electrophoresis. *Journal of Chromatography A* 838: 55–70.
- Křivánková L, Pantůčková P, Gebauer P, *et al.* (2003) Chloride present in biological samples as a tool for enhancement of sensitivity in CZE analysis of anionic trace analytes. *Electrophoresis* 24: 505–517.
- Urbánek M, Křivánková L, and Boček P (2003) Stacking phenomena in electromigration. From basic principles to practical procedures. *Electrophoresis* 24: 466–485.

## Isoelectric Focusing

P G Righetti, University of Verona, Verona, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Isoelectric focusing (IEF) represents a unique electrokinetic method, in that it is based on steady-state patterns attained by amphoteric species (mostly proteins and peptides) along a pH gradient under the influence of an electric field. Amphoteric compounds migrate until they align themselves at their isoelectric positions, where a dynamic equilibrium is produced between electrokinetic concentrating processes and dispersive effects that include diffusion. Extremely sharp zones are thus obtained, characterized by a very high resolving power. We will deal with conventional IEF in soluble, amphoteric buffers and with immobilized pH gradients (IPG) in insolubilized, nonamphoteric buffers in gel media. Additionally, we will cover IEF in capillaries, the most recent evolution of this technique.

## Conventional Isoelectric Focusing

In principle, pH gradients could be obtained by diffusion of nonamphoteric buffers but such ‘artificial’ gradients would be altered by changes in electric migration and diffusion of the buffer ions. Thus, in 1961, Svensson introduced the concept of ‘natural’ pH gradients, created and stabilized by the electric current itself. The buffers used in this system required two fundamental properties: (1) amphotericism, so that they could reach an equilibrium position along the separation column and (2) ‘carrier’ ability. This latter concept is subtler but just as fundamental. Any ampholyte cannot simply be used for IEF; only a carrier ampholyte (CA), which is a compound capable of transporting the current (a good conductor) and capable of carrying the pH (a good buffer). With this notion, and with Vesterberg’s elegant synthesis of such ampholytes, present-day conventional IEF was born.

## Some Basic Theoretical Concepts

We will survey here some basic equations governing the IEF process. The most important one regards the distribution profile of an ampholyte about its



isoelectric point. Under steady-state conditions (as obtained by balancing the simultaneous electrophoretic and diffusional mass transports), Svensson, in 1961, derived the following differential equation describing the concentration profile of a focused zone:

$$C\mu/qk = D(dC/dx) \quad [1]$$

where  $C$  is the concentration of a component in arbitrary mass units per arbitrary volume unit;  $\mu$  the electrophoretic mobility in  $\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$  of ion constituent except  $\text{H}^+$  and  $\text{OH}^-$ , with positive sign for cationic and negative sign for anionic migration;  $i$  the electric current in A;  $q$  the cross-sectional area in  $\text{cm}^2$  of electrolytic medium, measured perpendicularly to the direction of current;  $k$  the conductance of medium, in  $\Omega^{-1} \text{cm}^{-1}$ ;  $D$  the diffusion coefficient in  $\text{cm}^2 \text{s}^{-1}$  of a given ionic component with mobility  $\mu$ ;  $x$  the coordinate along the direction of current,  $x$  increasing from 0 at the anode to higher values toward the cathode.

Each term in eqn [1] expresses the mass flow per second and square centimeter of the cross-section, that on the left being the electric and that on the right the diffusional mass flows. If eqn [1] is written in the form:

$$(i\mu/q)(dx/k) = D(dC/C) \quad [2]$$

it is seen that it is possible to integrate it if  $\mu$  is known as a function of pH and  $D$  as a function of  $C$ . Specifically, if the conductance, the diffusion coefficient, and the derivative:

$$p = -d\mu/dx = -[d\mu/d(\text{pH})][d(\text{pH})/dx] \quad [3]$$

(where  $p$  is the ratio between the protein titration curve and the slope of the pH gradient over the separation axis) can be regarded as constant within the focused zone, then  $\mu = -px$  and one obtains the following analytical solution:

$$C = C_0 \exp[-(pix^2)/(2qkD)] \quad [4]$$

where  $x$  is now defined as being equal to zero at the concentration maximum  $C_0$ . This is a Gaussian concentration distribution with the inflection points at:

$$x_i = \pm\sqrt{(qkD)/(pi)} \quad [5]$$

where  $x_i$  denotes the width of the Gaussian distribution of the focused zone measured from the top of the distribution of the focused ampholyte to the inflection point (one standard deviation). The course of the pH gradient is  $d(\text{pH})/dx$  and  $d\mu/d(\text{pH})$  represents the titration curve of the ampholyte. It should be

borne in mind that this Gaussian profile holds only if and as long as the conductivity of the bulk solution comprised within the zone is constant. Constant conductivity along a pH gradient is quite difficult to maintain, especially as one approaches pH extremes (below pH 4 and above pH 10), if for no other reason, because the non-negligible concentration of  $\text{H}^+$  and  $\text{OH}^-$  present in the bulk liquid begin to strongly contribute.

Another important equation regards the resolving power in IEF, expressed in  $\Delta(\text{pI})$  units, i.e., in the minimum difference of surface charge between two adjacent proteins that the IEF technique is able to resolve. If two adjacent zones of equal mass have a peak-to-peak distance three times larger than the distance from peak to inflection point, there will be a concentration minimum approximating the two outer inflection points. Taking this criterion for resolved adjacent proteins, Rilbe has derived the following equation for minimally but definitely resolved zones:

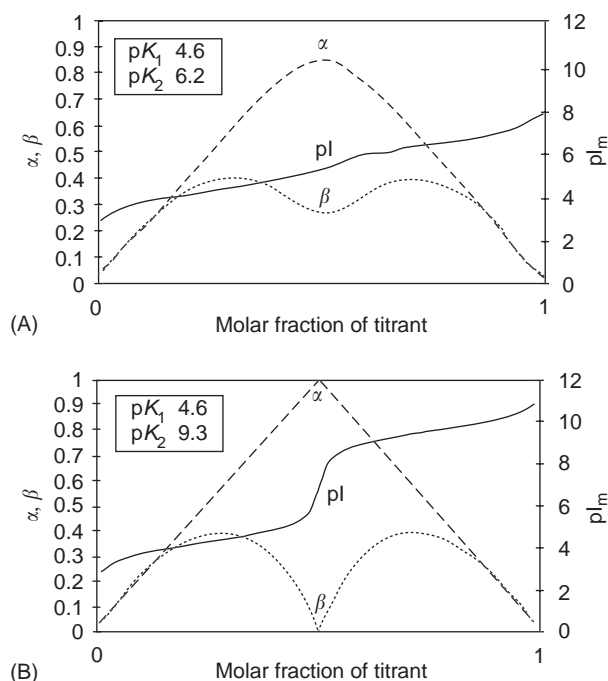
$$\Delta(\text{pI}) = 3\sqrt{\{D[d(\text{pH})/dx]\}/\{E[-d\mu/d(\text{pH})]\}} \quad [6]$$

Equation [6] shows that good resolution should be obtained with substances with a low diffusion coefficient ( $D$ ) and a high mobility slope  $[d\mu/d(\text{pH})]$  at the isoelectric point – conditions that are satisfied by all proteins. Good resolution is also favored by a high field strength ( $E$ ) and a shallow pH gradient  $[d(\text{pH})/dx]$ . It will be seen that, whereas in conventional IEF the limit to the resolving power is  $\sim 0.01$ , in IPGs it is 0.001 pH units.

### The Carrier Ampholyte Buffers

We recall here that the buffer capacity of an ampholyte in the isoprotonic state decreases with increasing  $\Delta\text{pK}$  across the isoprotonic point, linearly at first, then exponentially. Let us take as an example a hypothetical ampholyte, with  $\text{pK}_1 = 4.6$  (a carboxyl group) and  $\text{pK}_2 = 6.2$  (an amino group), having thus  $\text{pI} = 5.4$  and  $\Delta\text{pK} = 1.6$ . If we titrate this ampholyte in the pH 4–7 range, encompassing the two  $\text{pK}$ s, and if we plot the accompanying buffering power ( $\beta$ ), degree of dissociation ( $\alpha$ ), and slope of the pH gradient, we will have the plot of **Figure 1A**. It can be seen that there is still a substantial buffering power at the  $\text{pI}$  of the ampholyte, with a corresponding degree of ionization less than unity, and that the titration curve is smooth throughout the pH gradient explored, with only a small deviation about the  $\text{pI}$  of the ampholyte, indicating that this species is indeed a 'good' CA. Let us now take an ampholyte with  $\text{pK}_1 = 4.6$  but with  $\text{pK}_2 = 9.3$ , thus with a  $\text{pI} = 6.95$  and  $\Delta\text{pK} = 4.7$ . If we now titrate it in the pH 4–10 range, again encompassing the two  $\text{pK}$  values, we will have the graph of **Figure 1B**. It can be seen now





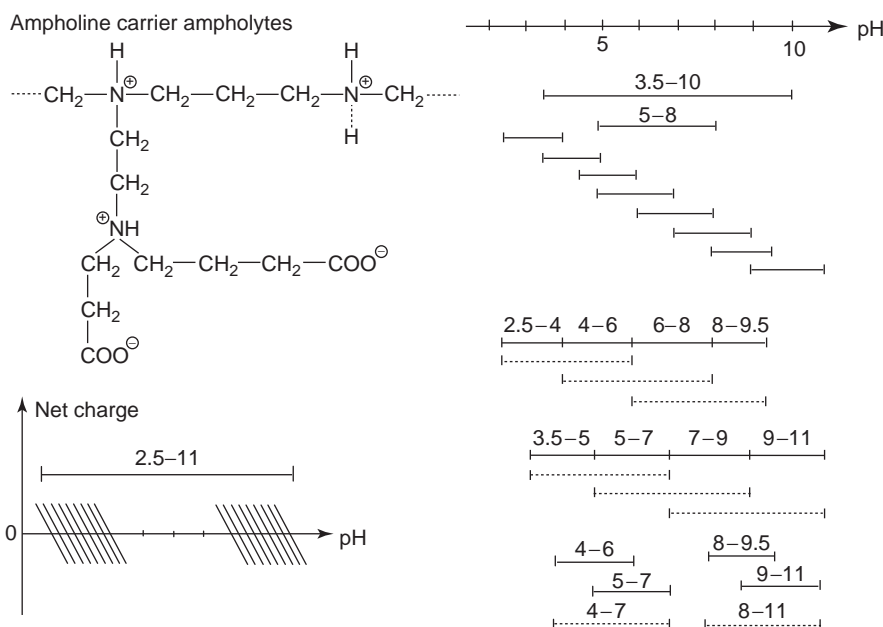
**Figure 1** Degree of ionization ( $\alpha$ ) and buffering power ( $\beta$ ) of a good (A) and a poor (B) carrier ampholyte. (A) Computer simulations obtained assuming  $pK_1 = 4.6$  and  $pK_2 = 6.2$  ( $pI = 5.4$ ). The ampholyte was titrated in the pH 4–7 interval. (B) Computer simulation obtained by assuming  $pK_1 = 4.6$  and  $pK_2 = 9.3$  ( $pI = 6.95$ ). The ampholyte was titrated in the pH 4–10 interval. Note the sharp sigmoidal transition in the pI region in B, suggesting total lack of buffering power (P Wenger and PG Righetti, unpublished).

that at the theoretical pI value the ampholyte does not have any appreciable buffering power and that it is fully ionized. In addition, it is not only isoelectric at pH 6.95, but indeed almost at any pH in the interval 5–9, as seen by the abrupt sigmoidal shape in the pI environment. This species will be a ‘bad’ CA, useless for a well-behaved IEF fractionation.

An important prerequisite for a good CA is that it has a high conductivity at its pI. Regions of low conductivity will absorb much of the applied voltage, thus reducing the field strength and hence the potential resolution in other parts of the gradient. It has been demonstrated that good conductivity is associated with small values of  $pI - pK$ . This is also true for the buffering capacity of an ampholyte. Thus, the parameter  $pI - pK$  (equivalent to  $\frac{1}{2}\Delta pK$ ) becomes the most important factor in selecting CAs exhibiting both good conductivity and buffering capacity ( $\beta$ ).

### Methodology

The structure of CAs (trade names Ampholine, Pharmalyte, Servalyte, and Biolyte, to name but a few) and their general properties are illustrated in **Figure 2**. CAs are oligoprotic amino carboxylic acids, each containing at least four weak protolytic groups, at least one being a carboxyl group and at least one a basic nitrogen atom, but no peptide bonds. In a typical synthesis, a mixture of oligoamines (four to six nitrogens in length, linear and branched) is reacted

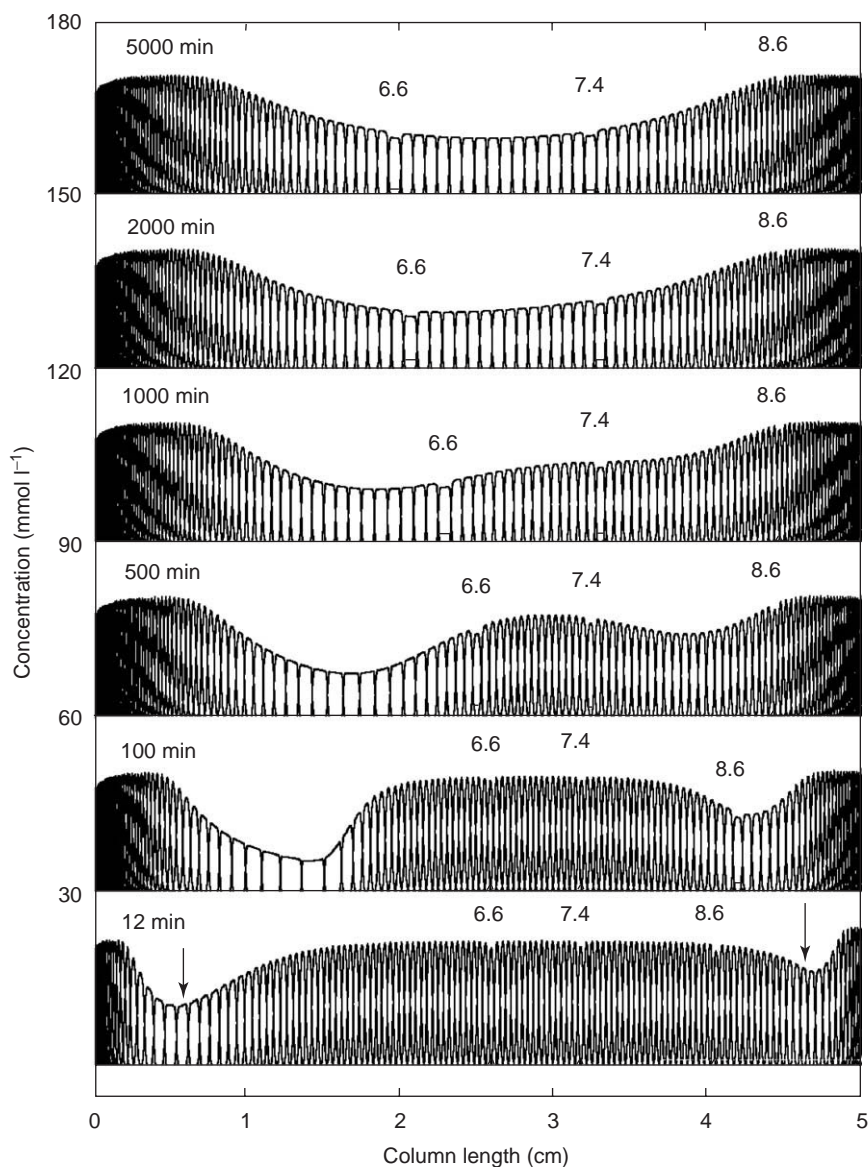


**Figure 2** Composition of Ampholine. On the upper left side a representative chemical formula is shown (aliphatic oligoamino oligocarboxylic acids). On the lower left side, portions of hypothetical titration curves of carrier ampholytes are depicted. Right: different pH cuts for wide and narrow range ampholytes.

with an  $\alpha,\beta$ -unsaturated acid (typically acrylic or itaconic acid), at a nitrogen:carboxyl ratio of 2:1.

The dynamics of IEF are illustrated in the simulations of Figure 3. This simulator is quite realistic, since it accepts 150 components and voltage gradients ( $300 \text{ V cm}^{-1}$ ) employed in laboratory practice. Before passage of the current, the column is at constant pH and the multitude of amphoteric buffers is randomly distributed, resulting in reciprocal neutralization. However, each individual CA species will have its own titration curve (see Figure 2, lower left

side) defining different mobilities in the electric circuit. After starting the experiment, the focusing process appears to proceed in two phases, a relative fast separation phase followed by a slow stabilizing phase during which a steady state should be reached. As shown in the bottom tracing, a pH 3–10 gradient appears to be fully established within 10–12 min, with the CAs forming Gaussian-like, overlapping zones of high concentration ( $>100$ -fold increase compared to the initial stage) with the three samples (dyes with pIs of 6.6, 7.4, and 8.6) well focused too.



**Figure 3** Computer-simulated distributions of 140 carrier ampholytes and of three dyes after 12, 100, 500, 1000, 2000, and 5000 min of current flow. The numbers refer to the pI values of the dyes and the arrowheads point to their locations. Successive graphs are presented with a y-axis offset of  $30 \text{ mmol l}^{-1}$ . The arrows at the bottom graph mark the two transient concentration valleys that are characteristic for the stabilizing phase. The cathode is to the right. (Reproduced with permission from Mosher RA and Thormann W (2002) High-resolution computer simulation of the dynamics of isoelectric focusing using carrier ampholytes: The postseparation stabilizing phase revisited. *Electrophoresis* 23: 1803–1814.)

Close to the column ends, large changes of carrier concentrations (valleys) are marked with arrows. This initial distribution of CAs generates an almost linear pH profile. This simulation, however, shows also the major drawback of all CA-based focusing processes: steady state can never be achieved. The two valleys at the extremes migrate inwards, flattening the focused CA from Gaussian into square-shaped peaks, and displacing them toward the terminal electrodes (see the simulated profiles from 100 min to 5000 min). As a result, the pH gradient keeps changing slopes, till it assumes a sigmoidal profile, with marked flattening of the central region with inflection point around neutrality (called plateau phenomenon). This results in a continuous shift of the foci of the focused samples along the column length (see the diverging positions of the three dyes from bottom to top in **Figure 3**).

By and large, most analytical IEF runs are performed in horizontal chambers: the polyacrylamide gel slab rests on a cooling block (in general made of glass or coated aluminum or even beryllium oxide, used as the heat shield of the space shuttle). This horizontal configuration allows one to dispose of electrode reservoirs and of all the hydraulic problems connected with vertical chambers (tight seals, etc.): in fact, anolyte and catholyte are soaked in filter paper strips resting directly on the open gel surface. In addition, most modern chambers contain a cover lid with movable electrodes, which can be adjusted to any gel length (in general from 10 to 25 cm electrode distances). Since thick gels (e.g., 2 mm thick) generate thermal gradients through the gel thickness, resulting in skewed zones (essentially all horizontal chambers have cooling only on one gel face) ultrathin gels (0.2–0.5 mm thin) supported onto a reactive polyester foil (Gel Bond PAG) are today preferred.

## Immobilized pH Gradients

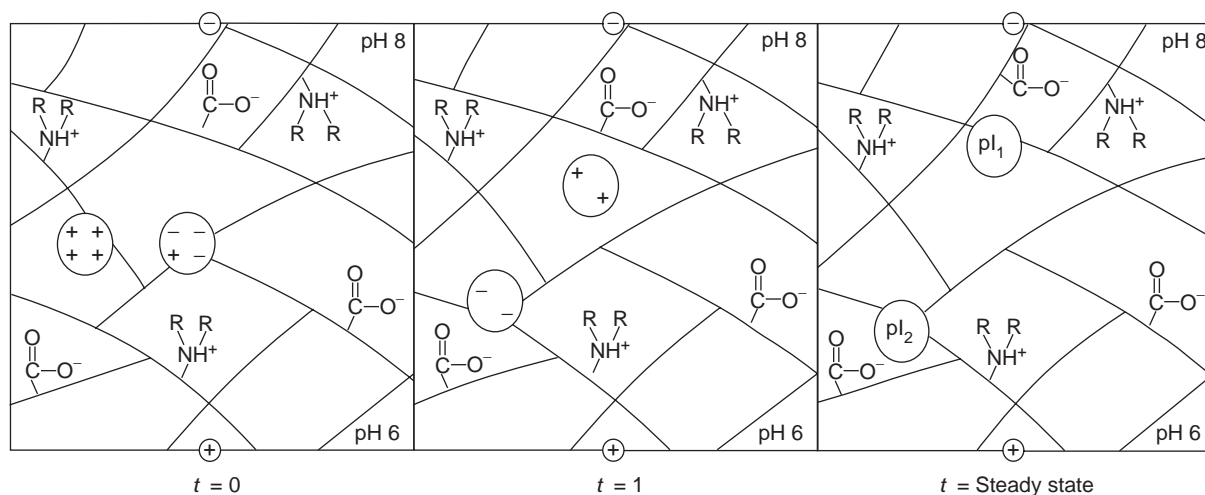
IPGs are based on the principle that the pH gradient, which exists prior to the IEF run itself, is copolymerized, and thus insolubilized, within the polyacrylamide matrix. This is achieved by using, as buffers, a set of up to 10 nonamphoteric, weak acids and bases, called Immobilines, having the following general chemical composition:  $\text{CH}_2=\text{CH}-\text{CO}-\text{NH}-\text{R}$ , where R denotes either three different weak carboxyls, with pKs 3.1, 3.6, and 4.6 (**Table 1**), or five tertiary amino groups, with pKs 6.2, 7.0, 8.5, 9.3, and 10.3 (**Table 2**). This set of eight weak buffers is complemented by one strong acid (pK  $\sim 1$ , 2-acrylamido-2-methyl propane sulfonic acid) and a strong base (pK  $> 12$ , quaternary amino ethyl acrylamide), used only as titrants. During gel polymerization, these buffering species are incorporated into the gel (84–86% conversion efficiency at 50°C for 1 h), by the same free radical reaction used to activate the acrylamide double bond. **Figure 4** depicts a segment of a hypothetical structure of an Immobililine matrix and the process of focusing two proteins in it. It is seen that only the proteins migrate to their steady-state position, whereas the Immobilines remain affixed at their original grafting position in the gel, where a fixed ratio buffering/titrant ions defines the pH locally. This means that, unlike conventional CA-IEF, the pH gradient is stable indefinitely (but it has to pre-exist before the onset of polymerization) and can only be destroyed if and when the polyacrylamide gel is hydrolyzed. Given the sparse distribution of Immobilines in the gel, they behave as isolated charges, able to effectively contribute to the ionic strength of the medium. In conventional IEF, on the contrary, at steady state the ionic strength is exceedingly low ( $< 1$  mequiv.  $\text{l}^{-1}$ ) since the focused carrier ampholytes

**Table 1** Acidic acrylamido buffers

pK	Formula	Name	$M_r$
1.2	$\text{CH}_2=\text{CH}-\text{CO}-\text{NH}-\begin{array}{c} \text{CH}_3 \\   \\ \text{C}-\text{CH}_3 \\   \\ \text{CH}_2-\text{SO}_3\text{H} \end{array}$	2-Acrylamido-2-methylpropane sulfonic acid	207
3.1	$\text{CH}_2=\text{CH}-\text{CO}-\text{NH}-\begin{array}{c} \text{CH}-\text{COOH} \\   \\ \text{OH} \end{array}$	2-Acrylamidoglycolic acid	145
3.6	$\text{CH}_2=\text{CH}-\text{CO}-\text{NH}-\text{CH}_2-\text{COOH}$	N-acryloylglycine	129
4.6	$\text{CH}_2=\text{CH}-\text{CO}-\text{NH}-(\text{CH}_2)_3-\text{COOH}$	4-Acrylamidobutyric acid	157

**Table 2** Basic acrylamido buffers

<i>pK</i>	<i>Formula</i>	<i>Name</i>	<i>M<sub>r</sub></i>
6.2	$\text{CH}_2=\text{CH}-\text{CO}-\text{NH}-(\text{CH}_2)_2-\text{N} \begin{array}{c} \diagup \diagdown \\ \diagdown \diagup \end{array} \text{O}$	2-Morpholinoethylacrylamide	184
7.0	$\text{CH}_2=\text{CH}-\text{CO}-\text{NH}-(\text{CH}_2)_3-\text{N} \begin{array}{c} \diagup \diagdown \\ \diagdown \diagup \end{array} \text{O}$	3-Morpholinopropylacrylamide	199
8.5	$\text{CH}_2=\text{CH}-\text{CO}-\text{NH}-(\text{CH}_2)_2-\text{N} \begin{array}{c} \text{CH}_3 \\   \\ \text{CH}_3 \end{array}$	<i>N,N</i> -dimethylaminoethyl acrylamide	142
9.3	$\text{CH}_2=\text{CH}-\text{CO}-\text{NH}-(\text{CH}_2)_3-\text{N} \begin{array}{c} \text{CH}_3 \\   \\ \text{CH}_3 \end{array}$	<i>N,N</i> -dimethylaminopropyl acrylamide	156
10.3	$\text{CH}_2=\text{CH}-\text{CO}-\text{NH}-(\text{CH}_2)_3-\text{N} \begin{array}{c} \text{C}_2\text{H}_5 \\   \\ \text{C}_2\text{H}_5 \end{array}$	<i>N,N</i> -diethylaminopropyl acrylamide	184
> 12	$\text{CH}_2=\text{CH}-\text{CO}-\text{NH}-(\text{CH}_2)_2-\text{N}^+ \begin{array}{c} \text{C}_2\text{H}_5 \\   \\ \text{C}_2\text{H}_5 \end{array}$	<i>N,N,N</i> -triethylaminoethyl acrylamide	198

**Figure 4** Hypothetical structure of an Immobiline gel and mechanism of the focusing process. The acrylamido acid and basic groups are shown grafted onto the polyacrylamide matrix. Two proteins are shown migrating in the gel at the times  $t=0$ , at  $t=1$  and finally at the steady state, where they reach their respective  $pI$  values ( $pI_1$  and  $pI_2$ ) as points of zero net charge.

form an inner salt and this often results in protein precipitation and smears both at the  $pI$  and in its proximity. In IPGs, the high ionic strength existing in the matrix (typically 10 mequiv.  $l^{-1}$ ) induces protein solubilization at the  $pI$  value (thus CA-IEF is similar to a 'salting-out' milieu and IPGs to a 'salting-in' environment).

Immobiline-based pH gradients can be cast in the same way as conventional polyacrylamide gradient gels, by using a density gradient to stabilize the Immobiline concentration gradient, with the aid of a standard, two-vessel gradient mixer. As shown in the above formula, these buffers are no longer amphoteric, as in conventional IEF, but are bifunctional: at

one end of the molecule the buffering group is located, and at the other end is the acrylic double bond, which will disappear during the grafting process. The three carboxyl Immobilines have rather small temperature coefficients of ionization ( $\text{dpK/dT}$ ) in the 10–25°C range, due to their small standard heats of ionization ( $\sim 1 \text{ kcal mol}^{-1}$ ) and thus exhibit negligible  $\text{pK}$  variations in this temperature interval. On the other hand, the four basic Immobilines exhibit rather large  $\Delta\text{pK}$ s in the same temperature range (as much as  $\Delta\text{pK} = 0.44$  for the  $\text{pK}$  8.5 species) due to their larger heats of ionization ( $6\text{--}12 \text{ kcal mol}^{-1}$ ). Therefore, for reproducible runs and pH gradient calculations, all the experimental parameters have been fixed at 10°C. Temperature is not the only variable that will affect Immobililine  $\text{pK}$ s (and therefore the actual pH gradient generated): additives in the gel that will change the water structure (chaotropic agents, such as urea) or will lower its dielectric constant, and the ionic strength of the solution itself will alter  $\text{pK}$  values.

### Narrow and Ultranarrow pH Gradients

We define the gradients (in the gel slab) from 0.1 to 1 pH unit as ultranarrow and narrow gradients, respectively. Within these limits, in general, one can work on a 'tandem' principle, i.e., one chooses a 'buffering' Immobililine (e.g., a base or an acid), having its  $\text{pK}$  within the desired pH interval, and a 'nonbuffering' Immobililine (e.g., an acid or a base, respectively), having its  $\text{pK}$  at least 2 pH units removed from either  $\text{pH}_{\min}$  or  $\text{pH}_{\max}$  of the pH range. The latter will thus provide equivalents of acid or base, respectively, to titrate the buffering group but will not itself buffer in the desired pH interval. For these calculations, one can resort to modified Henderson–Hasselbalch equations and to rather complex nomograms or simply adopt tabulated recipes, 1 pH unit wide, which start with the pH 3.8–4.8 interval and end with the pH 9.5–10.5 span, separated by 0.1 pH unit increments (58 such recipes have been tabulated). If a narrower pH gradient is needed, this can be derived from any of the 58 pH intervals tabulated by a simple linear interpolation of intermediate Immobililine molarities.

### Extended pH Gradients

For wider pH intervals, one has to mix several buffering species and the situation becomes considerably more complex. This has been solved with the aid of computer programs designed specifically for this purpose. The basic findings are: (1) for generating a linear pH gradient the buffering power has to be

constant throughout the desired pH interval (this is best achieved when the  $\text{pK}$  values are spaced at 1 pH unit interval); (2) to avoid deviations from linearity, the titrants should have  $\text{pK}$ s well outside  $\text{pH}_{\min}$  and  $\text{pH}_{\max}$  of the wanted pH range (in general, at least 2 pH units removed from the limits of the pH interval). As a consequence of this, for pH ranges wider than 4 pH units, two additional Immobilines are needed as titrants: one strongly acidic ( $\text{pK} < 1$ ) and one strongly basic ( $\text{pK} > 12$ ). There are two ways for generating extended pH intervals. In one approach, the concentration of each buffer is kept constant throughout the span of the pH gradient and 'holes' of buffering power are filled by increasing the amounts of the buffering species bordering the largest  $\Delta\text{pK}$ s; in the other approach (varying buffer concentration), the variation in concentration of the various buffers along the width of the desired pH gradient results in a shift of their apparent  $\text{pK}$ s with a concomitant evening-out of the  $\Delta\text{pK}$  values. With the available recipes, preparation of any Immobililine gel is now a trouble-free operation, as all the complex computing routines have been already performed and no further calculations of any type are required.

### Nonlinear, Extended pH Gradients

IPG formulations have been given only in terms of rigorously linear pH gradients. While this has been the only solution adopted so far, it might not be the optimal one in some cases. Altering the pH slope in some portions of the gradient might be required in those pH regions overcrowded with proteins. The reasons for resorting to nonlinear pH gradients are given in the histogram of Figure 5. With the relative abundance of different species, it is clear that an optimally resolving pH gradient should have a gentler slope in the acidic portion, and a steeper course in the alkaline region. Such a general course has been calculated by assigning to each 0.5 pH unit interval in the pH 3.5–10 region a slope inversely proportional to the relative abundance of proteins in that interval; by such a procedure, the ideal (dotted) curve in Figure 5 was obtained. What is also important here is the establishment of a new principle in IPG technology, namely that the pH and density gradients stabilizing it need not be colinear, the possibility existing of modulating the former, by locally flattening of pH gradients for increased resolution, while leaving unaltered the latter. Though we have given here only one example of a nonlinear extended pH gradient, clearly the possibility exists of modulating in the same fashion any narrower pH interval.



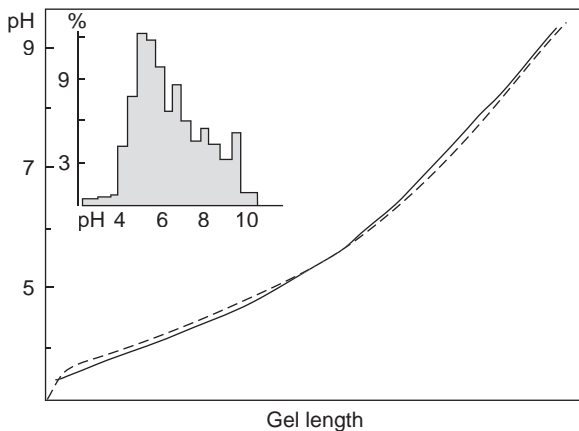
### Examples on the Resolving Power

What can IPGs achieve in practice? One interesting example is given in **Figure 6**: here a preparation of bovine tubulin is resolved into at least 21 isoforms, in the IPG pH 4–7 interval. The advantage of the IPG technique is that, unlike in CA-IEF, true steady-state patterns are obtained, rendering the method highly reproducible. The very low conductivity of the IPG strips, in addition, allows much higher voltage

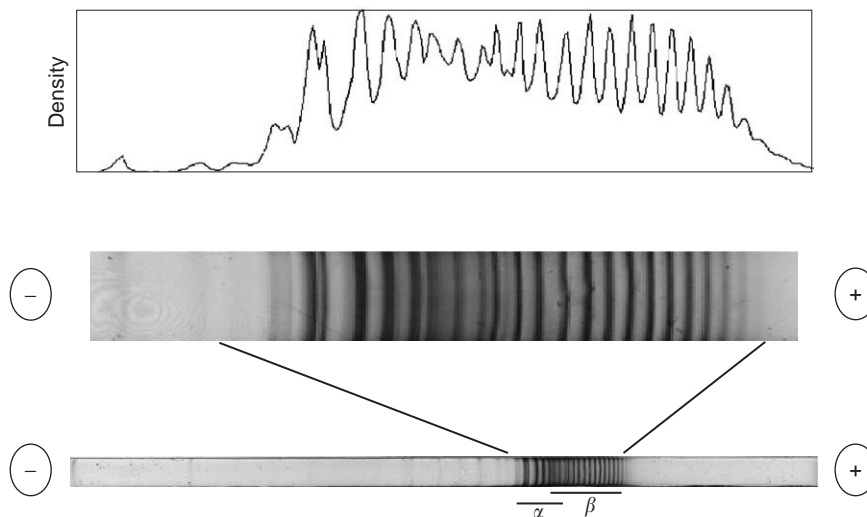
gradients to be delivered along the focusing axis, thus producing much sharper protein zones and ultimately leading to much higher resolution (it is recalled here that the resolving power of CA-IEF is  $\sim 0.01$  pH units, whereas that of IPG is greater by one order of magnitude, i.e., 0.001 pH units).

### Capillary Isoelectric Focusing

Although capillary isoelectric focusing (cIEF) still is beleaguered by some problems (notably how to cohabit with the ever present hazard of electroosmotic flow, EOF), the technique is gaining momentum and is becoming quite popular, especially in the analysis of r-DNA products and of heterogeneity due to differential glycosylation patterns and to 'protein aging,' i.e., asparagine (Asn) and glutamine (Gln) deamidation *in vitro*. A spin-off of cIEF is zone electrophoresis in zwitterionic, isoelectric buffers, a technique that exploits all the basic concepts of IEF and offers unrivalled resolution due to fast analysis in high-voltage gradients. Capillary electrophoresis offers some unique advantages over conventional gel slab techniques: the amount of sample required is truly minute (a few microliters at the injection port, but only a few nanoliters in the moving zone); the analysis time is, in general, very short (often just a few minutes) due to the very high voltages applicable; analyte detection is online and is coupled to a fully instrumental approach (with



**Figure 5** Nonlinear pH 4–10 gradient. Ideal (----) and actual (—) formulation courses. The shape for the ideal profile was computed from data on the statistical distribution of protein pIs. The relevant histogram is redrawn in the figure inset. (Reproduced with permission from Gianazza E, Giacom P, Sahlin B, and Righetti PG (1985) Non-linear pH courses with immobilized pH gradients. *Electrophoresis* 23: 1803–1814.)



**Figure 6** IEF of 25 µg bovine tubulin in an IPG pH 4–7 gradient. Bottom: the entire IPG strip; middle: enlarged detail of the principal protein-containing region; top: scan of this region. At least 21 bands are visible. Zones in which  $\alpha$ - and  $\beta$ -tubulin are known to focus are indicated. The 11 cm long IPG strip was rehydrated for 14 h at 22°C. Focusing was for 1 h at 500 V, 1 h at 1000 V, and 3 h at 5000 V, all at 20°C. Current was limited to 50 µA per strip. The positions of positive and negative electrodes are indicated. (Reproduced with permission from Williams RC, Jr., Shah C, and Sackett D (1999) Separation of tubulin isoforms by isoelectric focusing in immobilized pH gradients gels. *Analytical Biochemistry* 275: 265–267.)

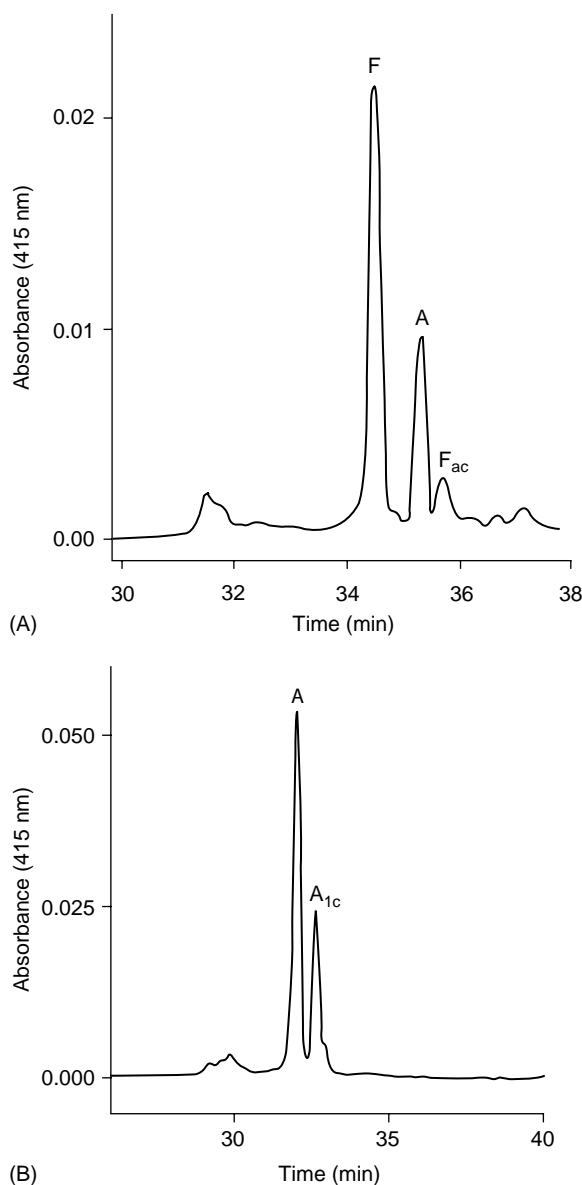


automatic storage of electropherograms on a magnetic support).

A principal difference between IEF in a gel and in a capillary is that, in the latter, mobilization of the focused proteins past the detector has to be carried out if an online imaging detection system is not being used. Mainly three techniques are used: chemical and hydrodynamic flow mobilization (in coated capillaries) and mobilization utilizing the EOF (in uncoated or partially coated capillaries). The last approach is troublesome, though, since the transit times of the focused zones change severely from run to run; thus, it is preferable to perform cIEF in well-coated capillaries, where EOF is completely suppressed.

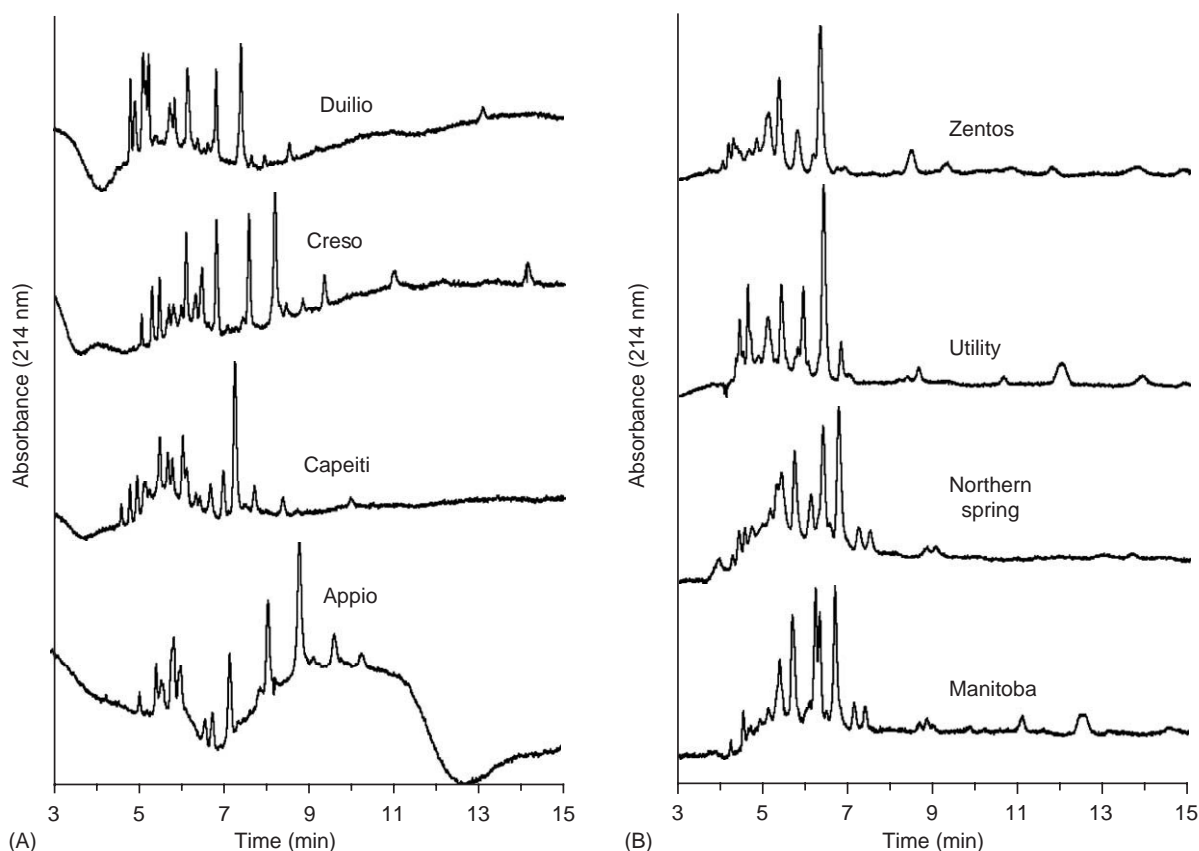
Basically, most of the rules that apply to CA-IEF in gel slabs are also valid in cIEF. It should be noted, though, that direct measurements of pH gradients cannot be performed in cIEF; thus, one has to resort to the use of pI markers. Of particular interest is a set of 16 synthetic peptides, engineered so as to cover the pH 3.5–10 interval; they have the advantage, over protein pI markers, of being stable and not generating spurious peaks due to 'aging,' as typical of macromolecules. In addition to conventional cIEF in standard capillaries (typically with a lumen of 50–100  $\mu\text{m}$  and with lengths of 25–50 cm) the recent trend is to adopt miniaturized chambers microfabricated in plastic microfluidic devices, such as polymethylmethacrylate, poly(ethylene terephthalate), and polycarbonate. In such devices, the separation channel can be just a few centimeters long; longer channels (e.g., 16 cm long) have to be machined in a serpentine fashion. Finally, cIEF is used increasingly as the first dimension of 2D maps, by which the eluate from the capillary separation is interfaced with mass spectrometry, for mass separation and characterization. The final result is a true 2D map, having pI/ $M_r$  coordinates, just like in gel-based techniques. It has been claimed that cIEF displays a peak capacity of  $\sim 800$  and sample concentration factors of  $\sim 500$ -fold, with a resolving power approaching  $\Delta(\text{pI})$  of  $\sim 0.005$ .

Some examples of separations that can be obtained by cIEF are shown in Figure 7. For species with minute  $\Delta\text{pIs}$ , one can resort to nonlinear pH gradients, by addition of zwitterions that alter the pH locally. In Figure 7A, one can appreciate an outstanding separation of fetal (F), adult (A), and acetylated fetal (Fac) hemoglobins. Another difficult separation, between adult and glycated hemoglobins, is shown in Figure 7B. Both separations are of interest in clinical chemistry: in the first case, for screening of thalassemias in umbilical cord blood, in the second instance for screening for diabetic patients.



**Figure 7** cIEF of hemoglobins. (A) separation of Hb F, A, and  $\text{F}_{\text{ac}}$ . Background electrolyte: 5% Ampholine, pH 6–8, added with 3% short-chain polyacrylamide and  $100 \text{ mmol l}^{-1}$   $\beta$ -Ala. (B) separation of an artificial mixture of 65% HbA and 35% Hb A1c in 5% Ampholine, pH 6–8, added with an equimolar mixture of separators,  $0.33 \text{ mol l}^{-1}$   $\beta$ -Ala and  $0.33 \text{ mol l}^{-1}$  6-amino caprioc acid. Anolyte:  $20 \text{ mmol l}^{-1}$   $\text{H}_3\text{PO}_4$ ; catholyte:  $40 \text{ mmol l}^{-1}$  NaOH. Sample loading: by pressure, for 60 s. Focusing run: 20 kV constant at  $7 \mu\text{A}$  (initial) to  $1 \mu\text{A}$  (final current),  $20^\circ\text{C}$  in a Bio Rad Bio Focus 2000 unit. Capillary: coated with poly(*N*-acryloyl amino ethoxy ethanol),  $25 \mu\text{m}$  ID, 23.6/19.1 total/effective length. Mobilization conditions: with  $200 \text{ mmol l}^{-1}$  NaCl added to anolyte, 22 kV. Detection at 415 nm.

As a corollary of focusing techniques, one should mention capillary zone electrophoresis in isoelectric buffers. This is an interesting development of cIEF, whereby zone electrophoresis can be performed in



**Figure 8** Representative CZE run of gliadins, obtained by sequential extraction of flour, from four different cultivars of durum (panel A) and of soft (panel B) wheat. Runs performed in  $40 \text{ mmol l}^{-1}$  Asp buffer, added with  $7 \text{ mol l}^{-1}$  urea and 0.5% HEC (apparent pH: 3.9). Conditions:  $50 \mu\text{m}$  ID, 30 cm long capillary, run at  $1000 \text{ V cm}^{-1}$  at room temperature. Detection at 214 nm (from Capelli *et al.*, 1998).

isoelectric, very-low-conductivity buffers, allowing the highest possible voltage gradients and thus much improved resolution of peptides and proteins due to reduced diffusion in short analysis times. Although originally described in alkaline pH values (notably in lysine,  $\text{pI}=9.87$ ) and at neutral pH (notably histidine buffers,  $\text{pH}=\text{pI}=7.6$ , albeit mostly for DNA and oligonucleotides), it has been found that acidic zwitterions offer an extra bonus: they allow the use of uncoated capillaries, due to protonation of silanols at the prevailing pH of the background electrolyte. A number of such acidic, isoelectric buffers have been described: cysteic acid ( $\text{pI}=1.80$ ); iminodiacetic acid ( $\text{pI}=2.23$ ); aspartic acid (Asp,  $\text{pI}=2.77$ ), and glutamic acid (Glu,  $\text{pI}=3.22$ ). **Figure 8** gives an example of such separations, performed in  $50 \text{ mmol l}^{-1}$  Asp, for the screening of gliadins in extracts from durum and soft wheat. Voltages as high as  $1000 \text{ V cm}^{-1}$  can be applied, with analysis times of only a few minutes, in 30-cm long capillaries.

**See also:** **Capillary Electrophoresis:** Overview. **Electrophoresis:** Overview; Principles; Isotachopheresis; Polyacrylamide Gels; Two-Dimensional Gels; Clinical Applications.

## Further Reading

- Righetti PG (1983) *Isoelectric Focusing: Theory, Methodology and Applications*. Amsterdam: Elsevier.
- Righetti PG (1990) *Immobilized pH Gradients: Theory and Methodology*. Amsterdam: Elsevier.
- Righetti PG, Stoyanov AV, and Zhukov MY (2001) *The Proteome Revisited*. Amsterdam: Elsevier.
- Righetti PG, Gelfi C, and Conti M (1997) Current trends in capillary isoelectric focusing of proteins. *Journal of Chromatography B* 699: 91–104.
- Righetti PG, Gelfi C, Bossi A, *et al.* (2000) Capillary electrophoresis of peptides and proteins in isoelectric buffers: an update. *Electrophoresis* 21: 4046–4053.
- Rilbe H (1976) Theory of isoelectric focusing. In: Catsimpoolas N (ed.) *Isoelectric Focusing*, pp. 14–52. New York: Academic Press.

- Rilbe H (1996) *pH and Buffer Theory: a New Approach*. Chichester: Wiley.
- Shen Y and Smith RD (2002) Proteomics based on high-efficiency capillary separations. *Electrophoresis* 23: 3106–3124.
- Shimura K, Wang Z, Matsumoto H, and Kasai K (2000) Synthetic oligopeptides as isoelectric point markers for capillary isoelectric focusing with ultra-violet absorption detection. *Electrophoresis* 21: 603–610.
- Svensson H (1961) Isoelectric fractionation, analysis and characterization of ampholytes in natural pH gradients. I: Solution of the equation for the steady state. *Acta Chemica Scandinavica* 15: 325–341.
- Tan W, Fan ZH, Qiu CX, Ricco AJ, and Gibbons I (2002) Miniaturized capillary isoelectric focusing in plastic microfluidic devices. *Electrophoresis* 23: 3638–3645.
- Vesterberg O (1969) Synthesis of carrier ampholytes for isoelectric focusing. *Acta Chemica Scandinavica* 23: 2653–2666.

## Electrophoresis in Ionic Liquids

**C F Poole**, Wayne State University, Detroit, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

viscosity (although viscosity is generally higher than for traditional organic solvents).

### Introduction

Ionic liquids are solvents comprised entirely of ions. This includes materials also known as molten salts, fused salts, liquid organic salts, nonaqueous ionic liquids, and room-temperature ionic liquids, but excludes aqueous and nonaqueous solutions of salts. Ionic liquids are inorganic or organic salts and their mixtures at a temperature above their melting point where they form stable liquids with favorable solvent properties. Roughly, half of all common inorganic salts have melting points below 500°C and many of these form eutectic mixtures with even lower melting points. Most organic salts have melting points below 400°C. Many of these, however, decompose at temperatures close to their melting points. An exception is the large number of alkylammonium, alkylphosphonium, imidazolium, and other heterocyclic cations with anions of low nucleophilicity (e.g., sulfonate, tetrafluoroborate, hexafluorophosphate, and bis(trifluoromethylsulfonyl)amide). These salts possess stable liquid ranges that are often thermally stable at temperatures more than 100°C above their melting point. A subset of these salts, known as room-temperature ionic liquids, has melting points below room temperature and is particularly promising for chromatographic and electrophoretic applications. At present, over 200 room-temperature ionic liquids are known, but only a few are fully characterized, explained by the fact that most were identified during the 1990s. Favorable properties include: virtually no vapor pressure; low flammability; high solubility for organic and inorganic compounds; high conductivity and a large window for electrochemical reactions; good thermal stability; and relatively low

### Inorganic Salts

Electrophoresis in fused inorganic salts and their eutectic mixtures were developed in the 1960s for the separation of inorganic ions and isotopes. Later, electrophoresis in ionic liquids was used for fundamental studies of ion solvation, to gain insight into the mechanism of ion transport, and to study complexation equilibria in the absence of competition from water molecules. As a separation method, it is virtually unused today, as more convenient techniques for separating ions, including isotopes, are now available. As a tool to probe the fundamental physical properties of ionic liquids, it finds occasional use. Electrophoresis in inorganic ionic liquids requires a special apparatus that can tolerate the high temperatures and corrosive nature of the fused salts and their electrolysis products.

A suitable support for electrophoresis should exhibit the following properties: (1) have good thermal stability to allow its use at high temperatures; (2) be chemically inert to allow its use with corrosive ionic liquids; (3) be a poor electrical conductor; (4) have a low obstruction factor; and (5) should not compete with electrophoretic separations by adsorbing sample ions. For flatbed separations glass fiber or asbestos paper strips are generally used. Powdered thin layers made by spraying aqueous suspensions of ceramic oxides onto sintered ceramic strips can be used as well. The support is first homogeneously impregnated with the ionic liquid. Then a small amount of the sample is spotted on the support (usually at the center). Separations are then carried out as in normal flatbed electrophoresis, except that here evaporation of the ionic liquid is negligible and high temperatures

- Rilbe H (1996) *pH and Buffer Theory: a New Approach*. Chichester: Wiley.
- Shen Y and Smith RD (2002) Proteomics based on high-efficiency capillary separations. *Electrophoresis* 23: 3106–3124.
- Shimura K, Wang Z, Matsumoto H, and Kasai K (2000) Synthetic oligopeptides as isoelectric point markers for capillary isoelectric focusing with ultra-violet absorption detection. *Electrophoresis* 21: 603–610.
- Svensson H (1961) Isoelectric fractionation, analysis and characterization of ampholytes in natural pH gradients. I: Solution of the equation for the steady state. *Acta Chemica Scandinavica* 15: 325–341.
- Tan W, Fan ZH, Qiu CX, Ricco AJ, and Gibbons I (2002) Miniaturized capillary isoelectric focusing in plastic microfluidic devices. *Electrophoresis* 23: 3638–3645.
- Vesterberg O (1969) Synthesis of carrier ampholytes for isoelectric focusing. *Acta Chemica Scandinavica* 23: 2653–2666.

## Electrophoresis in Ionic Liquids

**C F Poole**, Wayne State University, Detroit, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

viscosity (although viscosity is generally higher than for traditional organic solvents).

### Introduction

Ionic liquids are solvents comprised entirely of ions. This includes materials also known as molten salts, fused salts, liquid organic salts, nonaqueous ionic liquids, and room-temperature ionic liquids, but excludes aqueous and nonaqueous solutions of salts. Ionic liquids are inorganic or organic salts and their mixtures at a temperature above their melting point where they form stable liquids with favorable solvent properties. Roughly, half of all common inorganic salts have melting points below 500°C and many of these form eutectic mixtures with even lower melting points. Most organic salts have melting points below 400°C. Many of these, however, decompose at temperatures close to their melting points. An exception is the large number of alkylammonium, alkylphosphonium, imidazolium, and other heterocyclic cations with anions of low nucleophilicity (e.g., sulfonate, tetrafluoroborate, hexafluorophosphate, and bis(trifluoromethylsulfonyl)amide). These salts possess stable liquid ranges that are often thermally stable at temperatures more than 100°C above their melting point. A subset of these salts, known as room-temperature ionic liquids, has melting points below room temperature and is particularly promising for chromatographic and electrophoretic applications. At present, over 200 room-temperature ionic liquids are known, but only a few are fully characterized, explained by the fact that most were identified during the 1990s. Favorable properties include: virtually no vapor pressure; low flammability; high solubility for organic and inorganic compounds; high conductivity and a large window for electrochemical reactions; good thermal stability; and relatively low

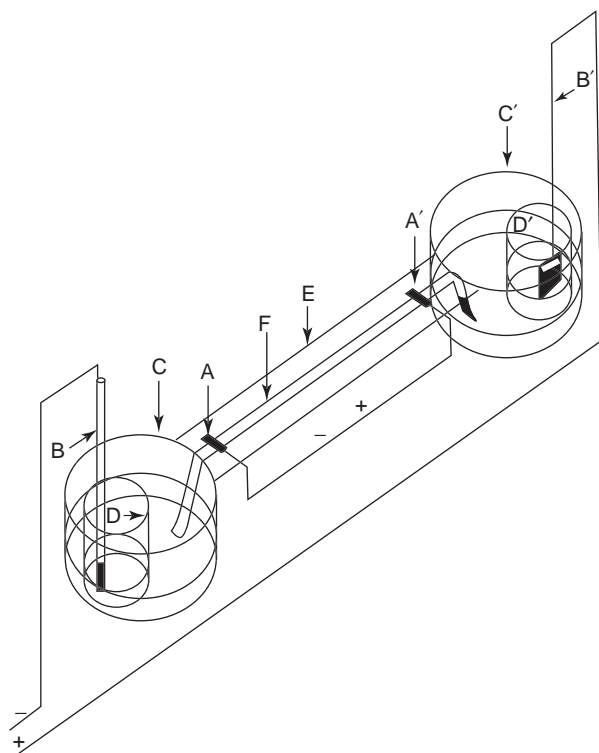
### Inorganic Salts

Electrophoresis in fused inorganic salts and their eutectic mixtures were developed in the 1960s for the separation of inorganic ions and isotopes. Later, electrophoresis in ionic liquids was used for fundamental studies of ion solvation, to gain insight into the mechanism of ion transport, and to study complexation equilibria in the absence of competition from water molecules. As a separation method, it is virtually unused today, as more convenient techniques for separating ions, including isotopes, are now available. As a tool to probe the fundamental physical properties of ionic liquids, it finds occasional use. Electrophoresis in inorganic ionic liquids requires a special apparatus that can tolerate the high temperatures and corrosive nature of the fused salts and their electrolysis products.

A suitable support for electrophoresis should exhibit the following properties: (1) have good thermal stability to allow its use at high temperatures; (2) be chemically inert to allow its use with corrosive ionic liquids; (3) be a poor electrical conductor; (4) have a low obstruction factor; and (5) should not compete with electrophoretic separations by adsorbing sample ions. For flatbed separations glass fiber or asbestos paper strips are generally used. Powdered thin layers made by spraying aqueous suspensions of ceramic oxides onto sintered ceramic strips can be used as well. The support is first homogeneously impregnated with the ionic liquid. Then a small amount of the sample is spotted on the support (usually at the center). Separations are then carried out as in normal flatbed electrophoresis, except that here evaporation of the ionic liquid is negligible and high temperatures

are necessary to maintain the electrolyte (fused salt) in the liquid state.

Figure 1 shows a typical apparatus for flatbed electrophoresis. It is fabricated from glass or fused silica. An electric field is applied to the impregnated support by two electrodes immersed in separate reservoirs at either end of the support. The choice of

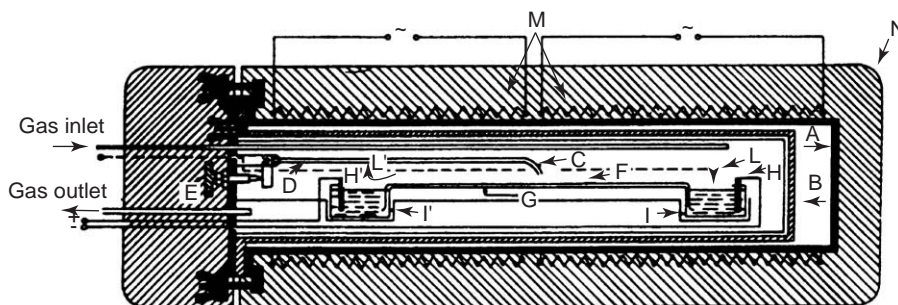


**Figure 1** Apparatus for zone electrophoresis in inorganic ionic liquids. A and A', platinum wires for measurement of the electric field; B and B' working electrodes; C and C' electrolyte reservoirs (fused salt); D and D' electrode compartments with sintered disks at the bottom; and E glass plate that supports the electrophoretic medium, F. (Reproduced with permission from Alberti G, Allulli A, and Modugno G (1964) *Journal of Chromatography* 15: 420–429.)

the electrodes depends exclusively upon their resistance to corrosion under the experimental conditions. The anode is generally platinum, and other materials such as tungsten, nickel, copper, or graphite are used as the cathode. Platinum may be attacked by the alkali metals formed at the cathode during electrophoresis. This kind of apparatus is best housed in an oven, as shown in Figure 2. Provision is also made to circulate a gas to remove volatile electrolysis products such as chlorine. At the end of the experiment, the substance is located by usual spot test reactions.

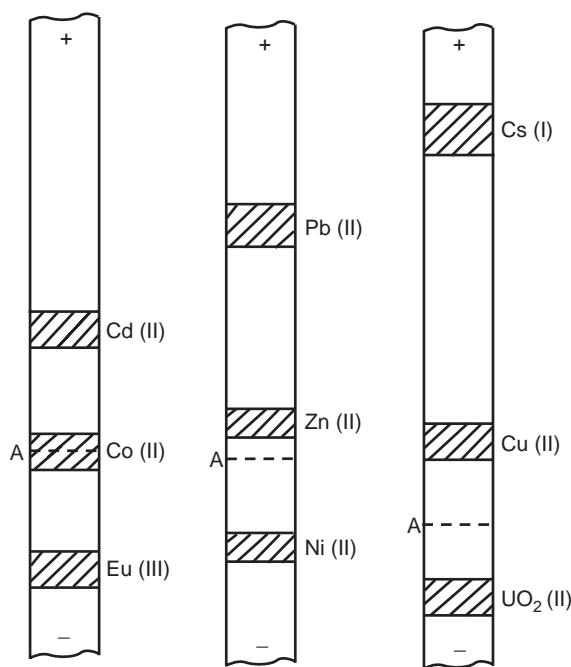
The technique is fairly rapid and simple; but when accurate data are required, several precautions are necessary. Auxiliary electrodes (A and A' in Figure 1) are required for an accurate measurement of the applied field due to polarization of the working electrodes. Temperature gradients (external or Joule heating) are minimized by choice of the operating conditions. The reservoirs have to be leveled to avoid flow of ionic liquid by siphon effects. Relatively low temperatures, i.e., between 150°C and 300°C, are commonly used. Suitable eutectic mixtures include lithium nitrate–potassium nitrate (43:57) at 160°C, sodium nitrate–potassium nitrate (50:50) at 250°C, and lithium chlorate–potassium chlorate (76:24) at 300°C. Potassium or sodium nitrates can be used at 350°C. Additives can be added to prevent precipitation of metal oxides and complexing agents to adjust selectivity. Some representative separations of metal ions on glass fiber paper are shown in Figure 3.

Column electrophoresis is carried out in an apparatus such as that shown in Figure 4. Glass powder, alumina, splinters, or quartz powder is used as a support. The support is first impregnated with the ionic liquid, then a solution of the sample is applied at the center of the separation column and the electric field applied. Column electrophoresis is commonly used for isotope separations.



**Figure 2** Cross-sectional view of an electrophoretic apparatus. A, Furnace; B, electrophoresis chamber; C, capillary; D, tube supporting the capillary; E, screw to adjust the height of the tube; F, electrophoretic support medium; G, heat-resistant glass plate; H and H' graphite electrodes; I and I', reservoirs; L, thermal couple; M, heating wire, and N, insulating jacket. (Reproduced with permission from Alberti G, Allulli A, and Modugno G (1964) *Journal of Chromatography* 15: 420–429.)



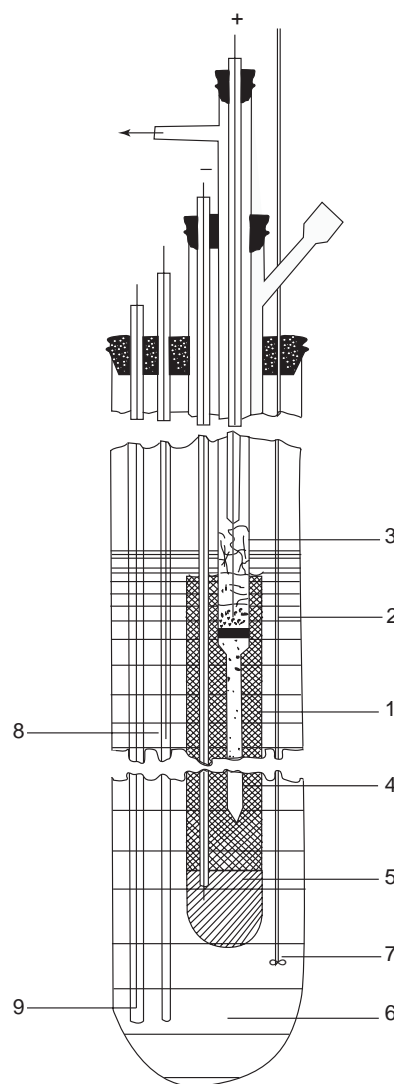


**Figure 3** Separation of inorganic ions by zone electrophoresis at 160°C in molten lithium nitrate–potassium nitrate eutectic containing 10% ammonium nitrate to inhibit precipitation of metal oxides. Sample applied at A. (Reproduced with permission from Alberti G and Allulli S (1968) *Chromatography Reviews* 10: 99–128.)

The relative mass difference of isotopes in ionic liquids is higher than in aqueous solution. This is because solvation of ions by water tends to diminish mass differences. Consequently, greater differences in electrophoretic mobility are possible in ionic liquids. Countercurrent electrophoresis in columns allowed isolation of highly enriched isotope mixtures without great difficulty. Some examples of the mass difference effect are summarized in **Table 1**. These results depend on the temperature as well as the composition of the ionic liquid.

## Organic Salts

Some typical room temperature ionic liquids and their physical properties are summarized in **Table 2**. The viscosity of most ionic liquids is substantially greater than traditional solvents. Elevated temperatures or dilution with conventional solvents of high dielectric constant can be used to lower the viscosity into a more desirable range. High viscosity, variable purity, and interference in absorbance detection at low wavelengths are the primary problems for the wider use of ionic liquids in electrophoresis at present. On the other hand, new ionic liquids are introduced at frequent intervals, and knowledge is



**Figure 4** Apparatus for column electrophoresis. 1, Separation column; 2, anode compartment; 3, foam at the platinum anode; 4, cathode compartment; 5, molten zinc cathode; 6, fused salt bath; 7, stirrer; 8 and 9, thermocouples. (Reproduced with permission from Ljublimov V and Lunden A (1966) *Zeitschrift für Naturforschung* 21a: 1592–1956.)

**Table 1** Effect of mass on electromigration of metals

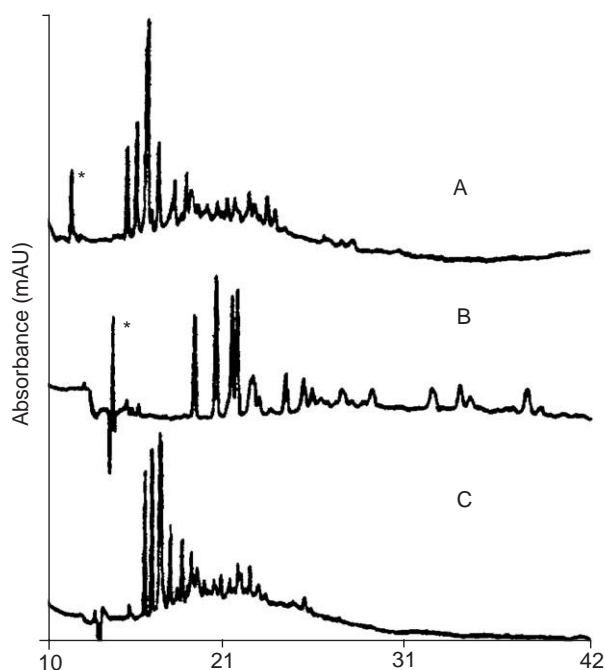
Isotopes of:	Ionic liquid	Effect of mass <sup>a</sup>
Li	LiCl	0.14
Li	LiBr	0.26
Li	LiNO <sub>3</sub>	0.05
K	KNO <sub>3</sub>	0.037
Cu	CuCl	0.080
Ag	AgCl	0.064
Pb	PbCl <sub>2</sub>	0.024
Zn	ZnBr <sub>2</sub>	0.11

<sup>a</sup>Mass effect =  $\Delta m M / \Delta M m$  where  $\Delta m$  is the difference in the mobility,  $\Delta M$  the difference in mass, and  $M$  the average mass of the isotopes. The external mobility  $m$  is the mobility with respect to the support as a reference.



**Table 2** Physical properties of some room-temperature ionic liquids (20–25°C)

Ionic liquid	Melting point (°C)	Density (g ml <sup>-1</sup> )	Viscosity (cP)	Refractive index
n-Butylammonium thiocyanate	20.5	0.95	97	1.526
1-Butyl-3-methylimidazolium				
bis(trifluoromethylsulfonyl)amide	– 4	1.43	69	1.427
heptafluorobutylate		1.33	182	1.414
hexafluorophosphate	10	1.37	330	
tetrafluoroborate	– 81	1.21	219	
trifluoroacetate		1.21	83	1.449
trifluoromethanesulfonate	16	1.29	90	1.438
Ethylammonium nitrate	12.5	1.12	32	1.454
1-Ethyl-3-methylimidazolium				
bis(trifluoromethylsulfonyl)amide	14	1.45	35	1.426
tetrafluoroborate	5.8	1.25	67	
trifluoroacetate		1.25	43	1.443
trifluoromethanesulfonate	– 9	1.39	50	1.433
n-Propylammonium nitrate	4	1.16	67	1.456



**Figure 5** Separation of poly(phenols) using (A) 1-ethyl-3-methylimidazolium tetrafluoroborate, (B) 1-butyl-3-methylimidazolium tetrafluoroborate, and (C) 1-ethyl-3-methylimidazolium hexafluorophosphate. The electrolyte solution was a 150 mmol l<sup>-1</sup> concentration of the ionic liquid in water. The fused-silica capillary column was 50 cm × 50 μm ID, operating voltage 16 kV, and absorbance detection at the anode end at 240 nm. (Reprinted with permission from Yanes EG, Gratz SR, Baldwin MJ, Robison SE, and Stalcup AM (2001) Capillary electrophoretic application of 1-alkyl-3-methylimidazolium-based ionic liquids. *Analytical Chemistry* 73: 3838–3844; © American Chemical Society.)

building on the design of ionic liquids for specific applications, which might address these problems at a future date.

Ionic liquids have been used as background electrolytes in aqueous and nonaqueous capillary

electrophoresis. The miscibility of the ionic liquids with water and organic solvents is determined by the structure of the salt. While some are immiscible with water, most are miscible with polar organic solvents. Bulky organic groups incorporated in either the cation or anion tend to diminish solubility in water. The most common co-solvents for capillary electrophoresis are water, acetonitrile, and mixtures of acetonitrile and methanol. In solvents of low polarity, a lack of free ions diminishes the effectiveness of ionic liquids for electrophoresis. 1,3-Dialkylimidazolium-based ionic liquids are adsorbed onto the capillary wall from aqueous solution. In so doing they reverse the direction of electroosmotic flow, more or less independent of pH. A number of different sample types (DNA fragments, metal ions, ionic dyes, poly(phenols), etc.) were separated using 1,3-dialkylimidazolium-based ionic liquids as a major component of the electrolyte solution without significant operational difficulties. The separation of poly(phenols) is shown in Figure 5. A contributing factor to this separation is the association of the poly(phenols) with the 1,3-dialkylimidazolium cations coating the capillary wall and/or as components of the electrolyte solution.

**See also:** Electrophoresis: Overview. Liquid Chromatography: Isotope Separations.

## Further Reading

Alberti G and Allulli S (1968) Chromatography and electrophoresis of inorganic ions in fused salts. *Chromatographic Reviews* 10: 99–128.

- Janz GJ (1967) *Handbook of Fused Salts*. New York: Academic Press.
- Lederer M (1994) *Chromatography for Inorganic Chemistry*. Chichester: Wiley.
- Poole CF (2004) Chromatographic and spectroscopic methods for the determination of solvent properties of

room temperature ionic liquids. *Journal of Chromatography A* 1037: 49–82.

- Vaher M, Koel M, and Kaljurand M (2002) Application of 1-alkyl-3-methylimidazolium-based ionic liquids in non-aqueous capillary electrophoresis. *Journal of Chromatography A* 979: 27–32.

## Polyacrylamide Gels

**P G Righetti**, University of Verona, Verona, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Zone electrophoresis became extremely popular in the early 1960s, with the introduction of polyacrylamide matrices and the discovery of the concept of discontinuous buffers. We will first review the chemistry of acrylamide monomers and cross-linkers and then discuss the following: disk electrophoresis, pore gradient electrophoresis, and sodium dodecyl sulfate electrophoresis. While these techniques are typically applied to protein analysis, there are at least two important fields of application of polyacrylamide gels to nucleic acid analysis: DNA-sequencing gels and temperature gradient gels. We will conclude with the most recent trend, especially in capillary zone electrophoresis, the use of ‘physical’ (i.e., uncross-linked) versus ‘chemical’ (i.e., cross-linked) gels. Polyacrylamide gel electrophoresis is routinely abbreviated as PAGE.

### On the Chemistry of Polyacrylamides

The popularity of polyacrylamide gels stems from several fundamental properties: (1) optical clarity, including ultraviolet (UV) (280 nm) transparency (but it absorbs strongly at 210 nm, a region of good response of the peptide bond; thus, its use is not recommended as sieving polymer in capillary electrophoresis of proteins and peptides at this wavelength, where UV-transparent polymers, such as celluloses, should be preferred); (2) electrical neutrality, due to the absence of charged groups; and (3) availability in a wide range of pore sizes. Their chemical formula, as commonly polymerized from acrylamide and *N,N'*-methylene bisacrylamide (Bis), is shown in **Figure 1**, together with that of two most widely employed catalysts, peroxydisulfate (ammonium or potassium) and *N,N,N',N'*-tetramethylethylenediamine. Normal polyacrylamide gels (i.e., cross-linked with

standard amounts of Bis, 3–5%) have porosities that decrease linearly with the total amount of monomers (%T: total monomer concentration per 100 ml of solution) in them and which are substantially smaller than the corresponding agarose gels. However, highly cross-linked gels (>10%C) (%C: grams of cross-linker per 100 g of total monomers) can, at very high cross-link levels (>30%C), have porosities comparable with, or even greater than, agarose matrices. High %C polyacrylamide matrices, however, are brittle, opaque, and tend to collapse and exude water above 30%C. Thus, their use is at present limited.

Hydrophilic gels are considered to be a network of flexible polymer chains, into whose interstices macromolecules are forced to migrate by an applied potential difference, according to a partition governed by steric factors. Large molecules can only penetrate into regions where the meshes in the net are large, while small molecules find their way into tightly knit regions of the network, even close to the cross-links. Different models of the gel structure have been described: geometric, statistical, and thermodynamic. Perhaps the best way to envision a hydrophilic matrix is to consider it as being composed of two interlacing fluid compartments endowed with different frictional coefficients. No matter how we describe the gel structure, it should be remembered that, in general, agaroses are more porous than polyacrylamides and that these two matrices are used complementary to each other. A number of new monomers and cross-linkers, with unique properties, have been synthesized over the years. They are listed in **Tables 1** and **2**, respectively, and will be described below.

### Trisacryl Gels

These gels are derived from the polymerization of the monomer *N*-acryloyl-2-amino-2-hydroxymethyl-1,3-propanediol. The trisacryl monomer creates a microenvironment that favors the approach of hydrophilic solutes (proteins) to the gel polymer surface, since the polyethylene backbone is buried underneath a layer

- Janz GJ (1967) *Handbook of Fused Salts*. New York: Academic Press.
- Lederer M (1994) *Chromatography for Inorganic Chemistry*. Chichester: Wiley.
- Poole CF (2004) Chromatographic and spectroscopic methods for the determination of solvent properties of

room temperature ionic liquids. *Journal of Chromatography A* 1037: 49–82.

- Vaher M, Koel M, and Kaljurand M (2002) Application of 1-alkyl-3-methylimidazolium-based ionic liquids in non-aqueous capillary electrophoresis. *Journal of Chromatography A* 979: 27–32.

## Polyacrylamide Gels

**P G Righetti**, University of Verona, Verona, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Zone electrophoresis became extremely popular in the early 1960s, with the introduction of polyacrylamide matrices and the discovery of the concept of discontinuous buffers. We will first review the chemistry of acrylamide monomers and cross-linkers and then discuss the following: disk electrophoresis, pore gradient electrophoresis, and sodium dodecyl sulfate electrophoresis. While these techniques are typically applied to protein analysis, there are at least two important fields of application of polyacrylamide gels to nucleic acid analysis: DNA-sequencing gels and temperature gradient gels. We will conclude with the most recent trend, especially in capillary zone electrophoresis, the use of ‘physical’ (i.e., uncross-linked) versus ‘chemical’ (i.e., cross-linked) gels. Polyacrylamide gel electrophoresis is routinely abbreviated as PAGE.

### On the Chemistry of Polyacrylamides

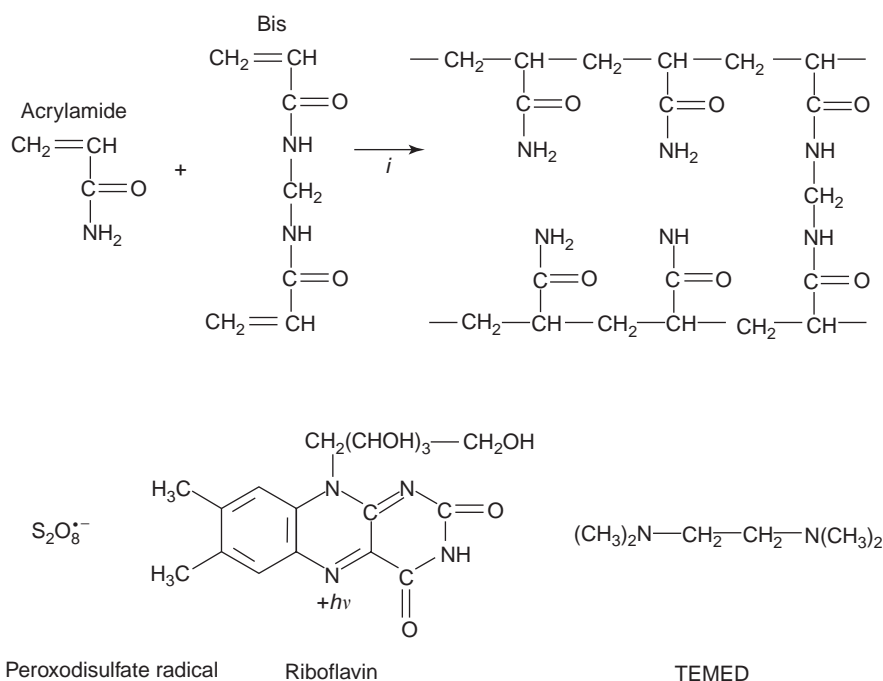
The popularity of polyacrylamide gels stems from several fundamental properties: (1) optical clarity, including ultraviolet (UV) (280 nm) transparency (but it absorbs strongly at 210 nm, a region of good response of the peptide bond; thus, its use is not recommended as sieving polymer in capillary electrophoresis of proteins and peptides at this wavelength, where UV-transparent polymers, such as celluloses, should be preferred); (2) electrical neutrality, due to the absence of charged groups; and (3) availability in a wide range of pore sizes. Their chemical formula, as commonly polymerized from acrylamide and *N,N'*-methylene bisacrylamide (Bis), is shown in **Figure 1**, together with that of two most widely employed catalysts, peroxydisulfate (ammonium or potassium) and *N,N,N',N'*-tetramethylethylenediamine. Normal polyacrylamide gels (i.e., cross-linked with

standard amounts of Bis, 3–5%) have porosities that decrease linearly with the total amount of monomers (%T: total monomer concentration per 100 ml of solution) in them and which are substantially smaller than the corresponding agarose gels. However, highly cross-linked gels (>10%C) (%C: grams of cross-linker per 100 g of total monomers) can, at very high cross-link levels (>30%C), have porosities comparable with, or even greater than, agarose matrices. High %C polyacrylamide matrices, however, are brittle, opaque, and tend to collapse and exude water above 30%C. Thus, their use is at present limited.

Hydrophilic gels are considered to be a network of flexible polymer chains, into whose interstices macromolecules are forced to migrate by an applied potential difference, according to a partition governed by steric factors. Large molecules can only penetrate into regions where the meshes in the net are large, while small molecules find their way into tightly knit regions of the network, even close to the cross-links. Different models of the gel structure have been described: geometric, statistical, and thermodynamic. Perhaps the best way to envision a hydrophilic matrix is to consider it as being composed of two interlacing fluid compartments endowed with different frictional coefficients. No matter how we describe the gel structure, it should be remembered that, in general, agaroses are more porous than polyacrylamides and that these two matrices are used complementary to each other. A number of new monomers and cross-linkers, with unique properties, have been synthesized over the years. They are listed in **Tables 1** and **2**, respectively, and will be described below.

### Trisacryl Gels

These gels are derived from the polymerization of the monomer *N*-acryloyl-2-amino-2-hydroxymethyl-1,3-propanediol. The trisacryl monomer creates a microenvironment that favors the approach of hydrophilic solutes (proteins) to the gel polymer surface, since the polyethylene backbone is buried underneath a layer



**Figure 1** The polymerization reaction of acrylamide. The structure of acrylamide, *N,N'*-methylenebisacrylamide (Bis) and of a representative segment of cross-linked polyacrylamide are shown. Initiators, designated by *i*, shown are peroxodisulfate, riboflavin, and *N,N,N',N'*-tetramethylethylenediamine (TEMED). Light is designated as  $h\nu$ . (Reprinted with permission from Chrambach A and Rodbard D (1971) The polymerization reaction of acrylamide. *Science* 172: 440–451.)

of hydroxymethyl groups. Thus, this type of matrix has an obvious advantage over polyacrylamide supports, which, lacking this protective hydrophilic layer, have a more pronounced hydrophobic character (Trisacryl, with a partition coefficient,  $P$ , of 0.01, is in fact the most hydrophilic monomer ever reported, bearing in mind that even acrylamide, considered to be rather hydrophilic, has a  $P = 0.2$ ). However, this monomer degrades with zero-order kinetics; thus, its use in electrophoresis is quite limited. The reason for this instability is the fact that Trisacryl constantly forms hydrogen bonds between the  $-OH$  groups and the carbonyl of the amido group; this activates a mechanism of 'N–O acyl transfer', which leads to quick hydrolysis of the amido bond even under mild alkaline conditions. The same drawbacks apply to the other very hydrophilic monomer, *N*-acryloyl-1-amino-1-deoxy-D-glucitol, which undergoes autocatalytic hydrolysis at alkaline pH values and is thus just as unstable as Trisacryl.

### ACM-BAP Gels

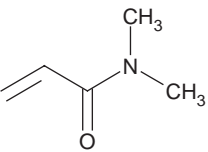
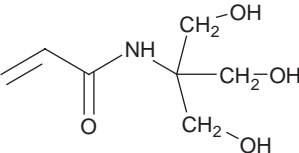
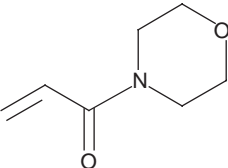
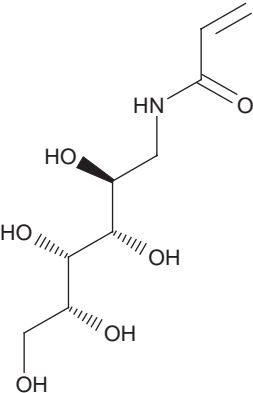
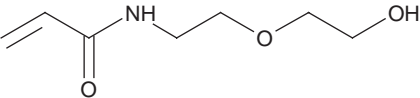
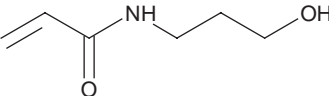
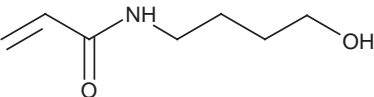
A couple of novel monomers imparting peculiar properties to polyacrylamide-type gels were reported in 1984: acryloyl morpholine (ACM) and bisacrylyl piperazine (BAP, a cross-linker). These gels exhibited some interesting features: due to the presence of the

morpholino ring on the nitrogen involved in the amido bond, the latter was rendered stable against alkaline hydrolysis, which bedevils conventional polyacrylamide matrices. In addition, such matrices are fully compatible with a host of hydro-organic solvents, thus allowing electrophoresis in mixed phases.

### Other Monomers Extremely Resistant to Hydrolysis

At the other extreme of the hydrophilicity scale is located DMA (*N,N'*-dimethylacrylamide): very high hydrophobicity ( $P = 0.5$ ), but extreme resistance to alkaline hydrolysis (approximately three orders of magnitude more stable). *N*-Acryloylaminoethoxyethanol (AAEE) is a novel monomer combining high hydrophilicity with extreme hydrolytic stability. As a free monomer, it has 10 times higher resistance to hydrolysis than plain acrylamide, but as a polymer its stability (in 0.1 M NaOH, 70°C) is 500 times higher. The stratagem: the  $\Omega$ -OH group in the *N*-substituent was removed enough from the amido bond so as to impede formation of hydrogen bonds with the amido carbonyl; if at all, the oxygen in the ethoxy moiety of the *N*-substituent would act as a preferential partner for *H*-bond formation with the  $\Omega$ -OH group. Acryloylaminopropanol (AAP), in which the ether group on the *N*-substituent was removed and the

**Table 1** Structural data on different types of *N*-substituted acrylamide monomers

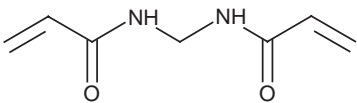
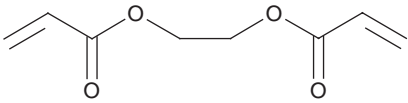
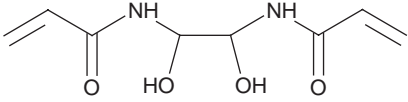
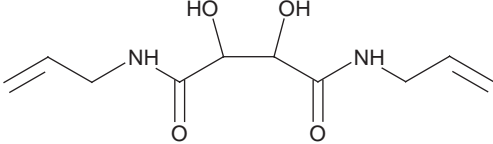
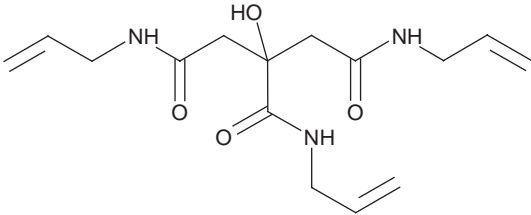
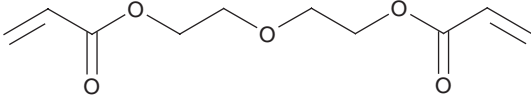
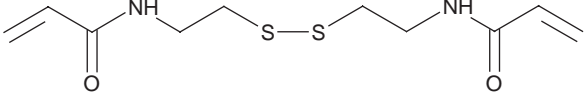
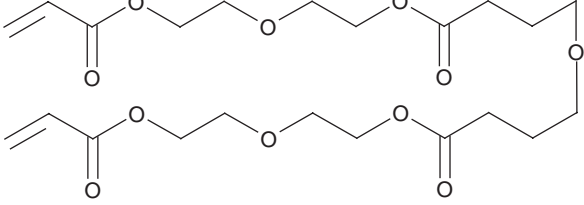
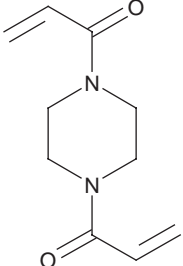
Name	Chemical formula	$M_r$
<i>N,N</i> -Dimethylacrylamide (DMA)		99
<i>N</i> -Acryloyl-2-amino-2-hydroxy methyl-1,3-propane (Trisacryl)		175
<i>N</i> -Acryloylmorpholine		141
<i>N</i> -Acryloyl-1-amino-1-deoxy-D-glucitol		235
<i>N</i> -Acryloylaminoethoxyethanol		159
<i>N</i> -Acryloylaminopropanol		129
<i>N</i> -Acryloylaminobutanol		143

chain shortened, appears to be even more hydrophilic than AAEE ( $P = 0.1$  for AAP versus  $P = 0.13$  for AAEE) and highly resistant to hydrolysis as compared to free acrylamide (hydrolysis constant of AAP  $0.008 \text{ l mol}^{-1} \text{ min}^{-1}$  versus  $0.05 \text{ l mol}^{-1} \text{ min}^{-1}$  for acrylamide, in an alkaline milieu).

#### Different Cross-Linkers

The versatility of polyacrylamide gels is also shown by the large number of cross-linkers, besides Bis, which can be used to cast gels with peculiar properties for different fractionation purposes. They are

**Table 2** Structural data on different types of cross-linkers

Name	Chemical formula	$M_r$	Chain length
<i>N,N'</i> -Methylene bisacrylamide (Bis)		154	9
Ethylene diacrylate (EDA)		170	10
<i>N,N'</i> -(1,2-Dihydroxyethylene) bisacrylamide (DHEBA)		200	10
<i>N,N'</i> -Diallyltartardiamide (DATD)		228	12
<i>N,N',N''</i> -Triallylcitric triamide (TACT)		309	12–13
Poly(ethylene glycol diacrylate) 200 (PEGDA <sub>200</sub> )		214	13
<i>N,N'</i> -Bisacrylyl cystamine (BAC)		260	14
Poly(ethylene glycol diacrylate) 400 (PEGDA <sub>400</sub> )		400	25
<i>N</i> -Bisacrylylpiperazine (BAP)		194	10



listed in Table 2. DHEBA can be used for casting reversible gels, since the 1,2-diol structure of DHEBA renders them susceptible to cleavage by oxidation with periodic acid. The same principle should also apply to DATD gels. Alternatively EDA gels could be used, since this cross-linker contains ester bonds that undergo base-catalyzed hydrolytic cleavage. The poly(ethylene glycol diacrylate) cross-link belongs to the same series of ester derivatives. As the latest addition to the series, BAC gels, which contain a disulfide bridge cleavable by thiols, have been described. These gels appear to be particularly useful for fractionation of RNA, and are the only ones that can be liquefied under mild and almost physiological conditions. Practically any cross-linker can be used, but definitely DATD and TACT should be avoided since, being allyl derivatives, they are inhibitors of gel polymerization when admixed with acrylic double bonds. Their use at high %C is simply disastrous. BAC is extremely hydrophobic, so its use in protein separations should be discouraged.

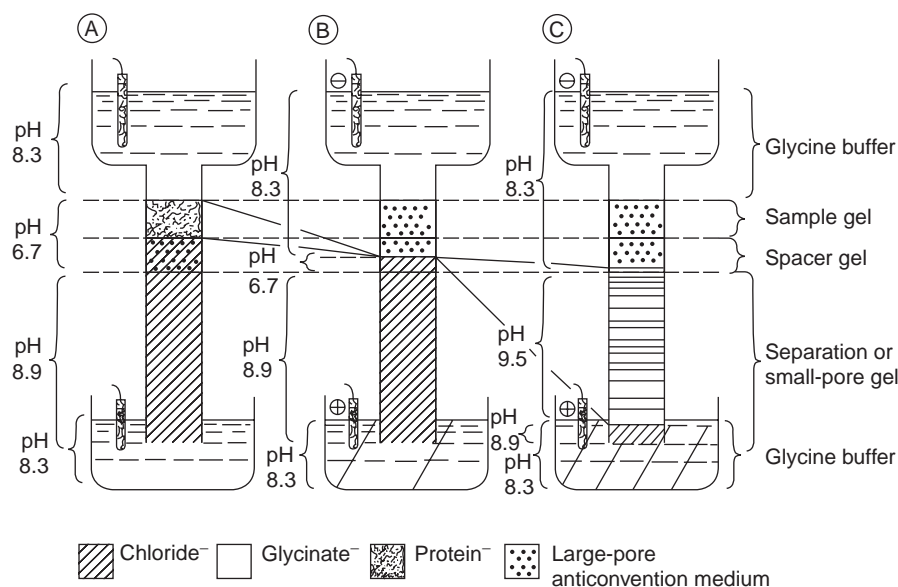
## Disk Electrophoresis

This method originated in the early 1960s, due to the theoretical and methodological breakthroughs of Ornstein and Davis. Before this time, zone electrophoresis was essentially performed in 'continuous' buffers, having the same composition (in buffering ions and titrants) and same pH along the migration

path and the electrolyte reservoirs. The amazing resolving power of disk electrophoresis is due to four such discontinuities: (1) in the gel matrix; (2) in the pH of the buffers; (3) in the buffer conductivity; and (4) in the types of ions present in the buffers (leading and trailing ions).

The matrix in Figure 2 is divided into three sections (from bottom): a 'separation' or 'running' gel; a 'spacer' or 'stacking' gel; and a sample gel. A few discontinuities exist at the running/stacking interface: the bottom gel is of small pore size, while the second and third layers consist of wider pore structures. Within the running/stacking interface, a pH discontinuity also exists. In fact, the running gel is buffered at pH 8.9, while spacer and sample gels are titrated to pH 6.7. The spacer and sample gel segments are also low-conductivity regions (third discontinuity), which means that a higher voltage gradient will be generated in these zones upon passage of the electric current. Below (i.e., in the running gel) and above (i.e., in the cathodic reservoir) these two gel segments, high-conductivity regions are to be found. A fourth discontinuity exists at the sample gel/upper electrode buffer interface: below it only  $\text{Cl}^-$  (leading) ions are present, above it only glycinate (trailing or terminating) ions are to be found.

This intricate setup has to satisfy an equation derived in 1897 by Kohlrausch (the regulating function), which links the concentration ratio of the fastest (leading, L) and slowest (terminating, T) ions to



**Figure 2** Set-up for disk electrophoresis. (A) Tree-layer polyacrylamide gel ready for the run; (B) isotachophoretic sample component stacking during the first few minutes of run (the process starts in the sample gel and has to be completed when leaving the spacer gel); (C) progress of separation in the running gel (Reproduced from Ornstein L (1964) *Annals New York Academy of Sciences* 121: 321–349).

their respective mobilities:

$$C_L/C_T = [M_L/(M_L + M_R)][(M_T + M_R)/M_T] \quad [1]$$

where  $C$  is the concentration,  $M$  the mobility, and  $R$  denotes the common counterion. If all the ions in the system are arranged in such a way that:  $M_{(\text{leading ion})} > M_{(\text{proteins})} > M_{(\text{trailing ion})}$  (where all the ions have the same charge, i.e., migrate to the same electrode) then, upon applying a voltage gradient, the ions will migrate down the cylinder with equal velocity and the boundary between each adjacent species will be maintained. This process has been called steady-state stacking or isotachopheresis. In more detail, as soon as the circuit is closed,  $\text{Cl}^-$  (fastest moving) ions are swept down the column toward the anode. Just behind this boundary, all protein ions will start arranging themselves in order of their mobilities, the lowest pI component next to the  $\text{Cl}^-$  boundary, the highest pI species coming at the end. This stacking process begins in the sample gel and is terminated in the spacer gel. It is important that these two gels are highly porous, so that protein ions in them move solely according to their charge (mobility), thus satisfying eqn [1], and not as a function of their size. The end of this series is the glycinate ion, which is why sample and spacer gels are buffered at pH 6.7. Gly has a theoretical isoelectric point (pI) = 6.0, but, indeed, as shown by its titration curve, is almost isoelectric even at pH 6.7, so that its anionic mobility is extremely small; in any event smaller than the slowest protein ion. As the isotachopheretic series moves down the sample and spacer gels, the width of the pH 6.7 zone shrinks gradually (see Figure 2) since the pH 8.3 glycinate buffer (catholyte) is drawn into the gel, following the pH 6.7/8.3 boundary.

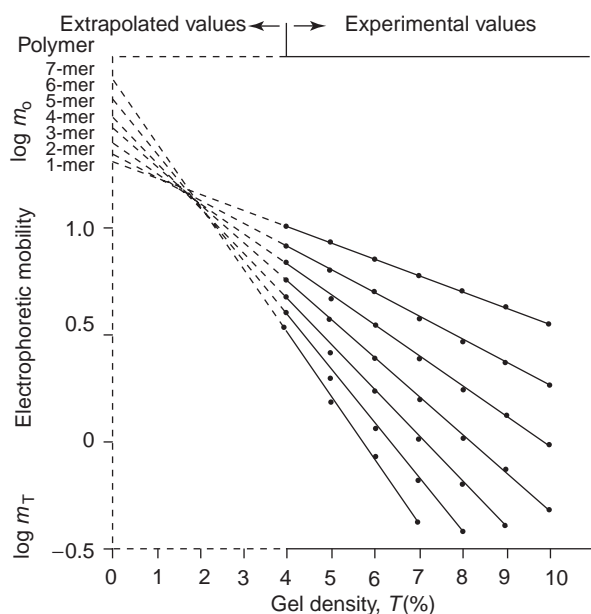
As the last protein zone enters the running gel, the various protein disks now overrun each other, since they experience a strong frictional force, due to the sieving matrix in the small-pore gel, so that now their velocity is a function of their charge/mass ratio. Only the  $\text{Cl}^-$  is unaffected. In addition, as the zwitterionic glycinate ion enters the pH 8.9 zone, its negative charge density strongly increases, so that it moves ahead and closely follows the  $\text{Cl}^-$  leading ion. As Gly<sup>-</sup> sweeps down the running gel (and increases its concentration in this zone due to its mobility increment) the pH increases from pH 8.9 to 9.5 (the pK of the Gly amino group) so that now the net charge on Gly is  $-0.5$ . As a last consequence of this further pH jump, all proteins experience an additional mobility increment. To summarize these events, in the sample and spacer gels two basic phenomena occur: (1) all protein ions are sorted out and physically separated

according to their pIs; (2) each protein ion is strongly concentrated in extremely thin starting zones (disks of barely a few micrometer thickness; a concentration process of more than 1000-fold).

Besides the obvious applications for the analysis of purity of a protein sample, disk electrophoresis can be used for physicochemical determinations. If the protein under analysis is run in a series of gels of different concentrations (% $T$ ) and the relative mobility ( $R_m$ : defined as the ratio between the distances migrated by a given protein and a tracking dye) in each gel measured, a linear plot of  $\log R_m$  versus % $T$  can be constructed (Ferguson plot). This plot can yield a wealth of information: the slope of this line is a measure of molecular size and is designated as the retardation coefficient ( $K_R$ ). The antilog of the y-intercept of this line ( $Y = R_m$  when  $T = 0$ ) is a measure of the free electrophoretic mobility (i.e., when no gel is present) and therefore of net charge. Since it has been demonstrated that for globular proteins there is a linear relationship between  $\sqrt{K_R}$  and the molecular radius ( $R$ ), the latter can be calculated from  $K_R$  measurements. Once the free mobility ( $m_o$ , calculated from  $Y_o$ ) and  $R_m$  (calculated from  $K_R$ ) are known, one can calculate the net charge on the molecule by using the classical theory of electrophoresis. Moreover, if Ferguson plots are measured for all the proteins in a mixture, the slopes of the different lines will immediately tell if the different components have similar size but different charge, or similar charge but different size, or if they differ in both size and charge. An example of a Ferguson plot is given in Figure 3: when bovine serum albumin (BSA) is run in a series of gels from 4% to 10% $T$ , a family of curves of different slopes is obtained, representing oligomers (from monomer to heptamer) of BSA.

## Electrophoresis in Porosity Gradients

Running macromolecules not in a gel of uniform concentration, but in a continuously varying matrix concentration (which results in a porosity gradient), has the effect of compacting the protein zones along the running track, as the band front is, at any given time, at a gel concentration somewhat higher than the rear of the band, so that the former is decelerated continuously. A progressive band sharpening thus results. There are other reasons for resorting to gels of graded porosity. Disk electrophoresis separates macromolecules on the basis of both size and charge differences. If the influence of molecular charge could be eliminated, then clearly the method could be used with a suitable calibration for measuring molecular size ( $M_r$ ). This has been accomplished with two main



**Figure 3** Plot of the logarithm of electrophoretic mobility of BSA versus gel density (%T) at constant 2%C. The straight lines were calculated from the data by means of least squares regression. (Reproduced from Thorun W (1971) *Zeitschrift für Klinische Chemie und Klinische Biochemie* 9: 3–11.)

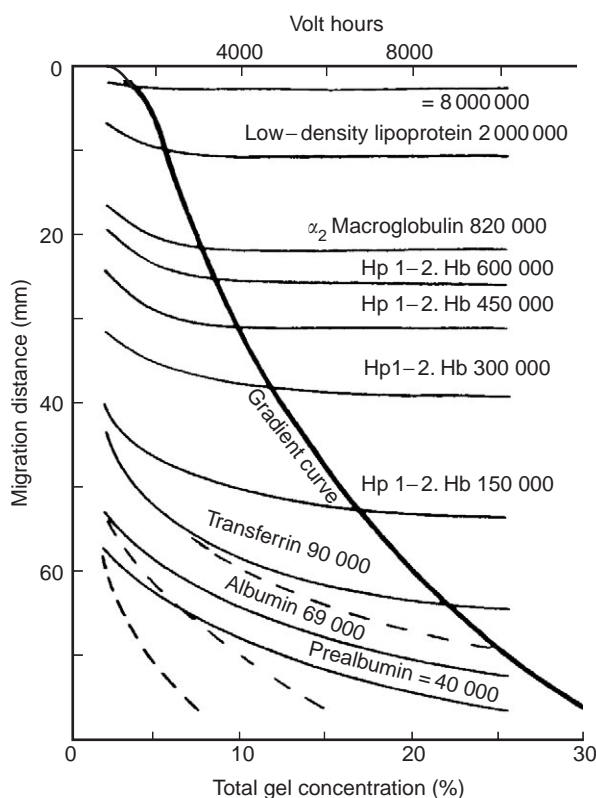
ways of overcoming charge effects. In one, a relatively large amount of charged ligand, such as sodium dodecyl sulfate (SDS), is bound to the protein, effectively swamping the initial charges present on the protein molecules and giving a quasi constant charge-to-mass ratio. However, in SDS electrophoresis proteins are generally dissociated into the constituent polypeptide subunits and the concomitant loss of functional integrity and antigenic profiles cannot be prevented. Therefore, the size of the original, native molecule has to be evaluated in the absence of denaturing substances.

In the second method for  $M_r$  measurements, this can be done by relying on a mathematical canceling of charge effects following measurements of mobility of native proteins in gels of different concentrations. This is the 'Ferguson plot' described above:

$$\log R_m = \log Y_0 - K_R T \quad [2]$$

However, this approach to  $M_r$  determination has never found widespread application as it is laborious, tedious, and expensive in terms of time and sample material.

This leads on to a third method for molecular size measurements: the use of gels of graded porosity. The principle of the technique is shown in **Figure 4**; it is seen that, given enough time (at least  $10 \text{ kV} \times \text{h}$ ) the mobility of most serum proteins becomes constant and eventually ceases as each constituent reaches a



**Figure 4** Plots of the mobilities of serum proteins against time using pore gradient gels. Solid lines: uniform cross-linking (5%C); dashed lines: variable cross-linking. (Reproduced with permission from Margolis J and Wrigley CW (1975) *Journal of Chromatography* 106: 204–210; © Elsevier.)

gel density region in which the average pore size approaches the diameter of the protein (pore limit). Thus, the ratio between migration distance of a protein to that of any other becomes a constant after the proteins have all entered a gel region in which they are subjected to drastic sieving conditions. This causes the electrophoretic pattern to become constant after prolonged runs in a gel gradient. The gel concentration at which the migration rate for a given protein becomes constant is called the 'pore limit': if this porosity is properly mapped with the aid of a suitable set of marker proteins, then it is possible to correlate the migration distance to the molecular mass of any constituent in the mixture.

After the electrophoretic run has been performed, and the proper experimental data gathered, there are two basic ways for mathematical handling of data: two-step and one-step methods. In the former case, the migration distance of proteins is plotted against the square root of electrophoresis time. If now the slopes of the various regression lines thus obtained are plotted against the respective molecular masses, a good linear fit is obtained, which allows  $M_r$  measurements

of unknown proteins between  $2 \times 10^5$  and  $10^6$ . The shape of the proteins (globular or fibrillar), their carbohydrate content (up to 45%), and their free electrophoretic mobilities (between  $2.1$  and  $5.9 \times 10^{-5} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ ) do not seem to be critical for proper  $M_r$  measurements by this procedure. In the one-step methods, when  $\log M_r$  is plotted versus  $\%T'$  or  $D'$ , a linear regression line is obtained, which allows accurate estimation of  $M_r$ s of unknown proteins. The following empirical equations, linking the above parameters, have been derived:

$$\log M_r = -e'\%T + y' \quad [3]$$

$$\log M_r = -e''D' + y'' \quad [4]$$

where  $-e'$  and  $-e''$  represent the slopes of the various regression lines while  $y'$  and  $y''$  (i.e., the  $y$ -intercepts) are indeed a single point at which the gel concentration equals the starting gradient concentration. The  $\log(M_r \text{ max})$  or  $\log(R_s \text{ max})$  corresponding to this concentration is of the order of  $1.3 \times 10^6 \text{ Da}$  ( $R_s = 9.3 \text{ nm}$ ). The correlations  $\log M_r - \%T'$  or  $\log M_r - D'$  are not significantly altered by the duration of electrophoresis. Therefore, a constant  $M_r$  value should be obtained for a stable protein, no matter how long electrophoresis is performed. It has also been demonstrated that the same mathematical relationship given above ( $\log M_r$  versus  $D'$ ) is also applicable to SDS electrophoresis in linear polyacrylamide gel gradients.

## Sodium Dodecyl Sulfate Electrophoresis

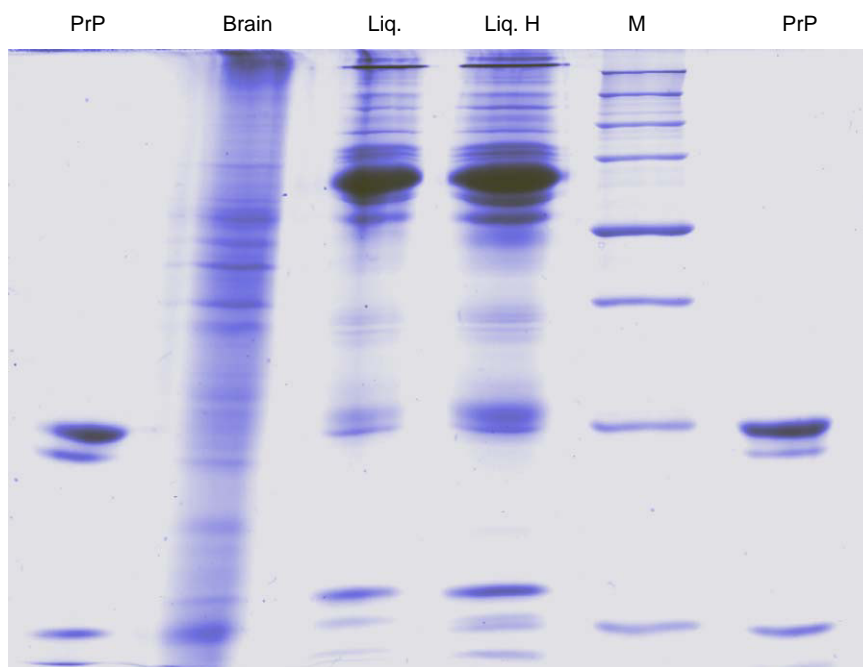
SDS electrophoresis was the next logical step after disk electrophoresis. While the latter discriminates macromolecules on the basis of both size and surface charge, SDS electrophoresis fractionates polypeptide chains essentially on the basis of their size. It is therefore a simple, yet powerful and reliable method for molecular mass ( $M_r$ ) determination. In 1967, it was first reported that electrophoretic migration in SDS is proportional to the effective molecular radius and thus to the  $M_r$  of the polypeptide chain. This result means that SDS must bind to proteins and cancel out differences in molecular charge, so that all components then migrate solely according to size. Surprisingly large amounts of SDS appear to be bound (an average of 1.4 g SDS per gram of protein), which means that the number of SDS molecules bound is of the order of half the number of amino acid residues in a polypeptide chain. This amount of highly charged surfactant molecules is sufficient to overwhelm effectively the intrinsic charges of the polymer coil, so that their net charge per unit mass

becomes approximately constant. If migration in SDS (and disulfide reducing agents, such as 2-mercaptoethanol, in the denaturing step, for a proper unfolding of the proteins) is proportional only to molecular mass, then, in addition to canceling out of charge differences, SDS also equalizes molecular shape differences as well (e.g., globular versus rod-shaped molecules). This seems to be the case for protein-SDS mixed micelles: these complexes can be assumed to behave as ellipsoids of constant minor axis ( $\sim 1.8 \text{ nm}$ ) and with the major axis proportional to the length in amino acids (i.e., to molecular mass) of the protein. The rod length for the 1.4 g SDS/g protein complex is of the order of 0.074 nm per amino acid residue.

The protein sample ( $\sim 1 \text{ mg ml}^{-1}$ ) is usually denatured in  $100 \text{ mmol l}^{-1}$  phosphate pH 7.0, containing 1% SDS, 1% 2-mercaptoethanol, 5–10% sucrose (for incrementing the sample density for gel loading purposes) and traces of bromophenol blue (as a tracking dye for monitoring boundary formation in discontinuous systems and for checking run termination). After heating at  $100^\circ\text{C}$  for 5 min, the sample is usually deposited in pockets preformed in the gel slab. Originally, continuous gel matrices and buffers were used, but today discontinuous buffers and gel phases are preferred, since they produce sharper bands and better resolution. In SDS electrophoresis, the proteins can be prelabeled with dyes that covalently bind to its  $-\text{NH}_2$  residues. They can be conventional (like the blue dye Remazol) or fluorescent, like dansyl chloride, fluorescamine, *o*-phthaldialdehyde, and MDPF (2-methoxy-2,4-diphenyl-3(2 H)-furanone). Prelabeling is compatible with SDS electrophoresis, as the size increase is minimal, but would be anathema in disk electrophoresis or isoelectric focusing, as it would generate a series of bands of slightly altered mobility or pI from an otherwise homogeneous protein. Among post-labeling techniques, silver staining has become quite popular, since the amount of detectable protein per band can be as little as 1 ng, and this allows resolution of bands differing by only 300 Da in size. An example of SDS electrophoresis of prion proteins is given in Figure 5.

Two classes of proteins show anomalous behavior in SDS electrophoresis: glycoproteins (because their hydrophilic oligosaccharide units prevent hydrophobic binding of SDS micelles) and strongly basic proteins (e.g., histones) (because of electrostatic binding of SDS micelles through their sulfate groups). The first can be partially alleviated by using Tris-borate buffers at alkaline pH, which will increase the net negative charge on the glycoprotein, thus producing migration rates well correlated with molecular size.





**Figure 5** Analysis of prion proteins by SDS-PAGE. Extreme left and right lanes: prion protein (PrP) purified by affinity capture with specific antibodies; brain homogenate; two tracks with cerebrospinal fluid (Liq. and Liq. H); M: molecular mass markers. Gel: 7.5–20% *T* porosity gradient, with a Laemmli's discontinuous buffer system. Staining: Coomassie Brilliant Blue. (Unpublished experiments with Castagna A, Monaco S, and Zanusso P.)

Migration of histones can be improved by using pore gradient gels and allowing the polypeptide chains to approach the pore limit.

## DNA Sequencing

The first step in sequencing a purified DNA molecule is to break it specifically into a number of well-defined species with a restriction endonuclease and then to select and prepare the piece to be sequenced. The second step is to prepare families of related fragments from this species. For this, two basic methods are available: (1) the chain terminator approach of Sanger and (2) the chemical or Maxam and Gilbert DNA sequencing method. Once this family of fragments has been produced by either method, their analysis can be performed by PAGE, in matrices capable of resolving oligonucleotide chains with one end identical and the other end varying in length by a single nucleotide. The resulting map is interpreted by a computer-linked gel scanner and data handling equipment. The gels used for sequencing are usually thin (0.2–0.4 mm) and rather large, e.g., 25 cm × <1 m; this enables a sequence of ~400 nucleotides to be read off a single polyacrylamide gel slab.

The chain terminator method depends on the primed synthesis of a complementary radioactive copy of a single-stranded DNA template by a DNA polymerase. The base-specific ends of the oligodeoxynucleotides thus synthesized are produced by incorporating 2',3'-dideoxynucleoside triphosphates into the newly synthesized DNA. The dideoxy analogs lack the 3'-hydroxyl group necessary for the formation of the next phosphate bond during chain growth, so that whenever a dideoxynucleotide is incorporated into the growing polynucleotide chain, growth stops. During normal chain extension, all four deoxynucleoside triphosphates are available to DNA polymerase. Usually one of them has a radioactive ( $^{32}\text{P}$ ) phosphate, allowing the newly synthesized DNA to be labeled. If one of these is completely replaced by its dideoxy analog, chain growth will stop at the first site of incorporation of the dideoxynucleotide. Since in the incubation mixture the dideoxy analog is present with some amount of its normal deoxynucleoside triphosphate counterpart, a partial incorporation of the former occurs at each possible site. Thus, by adding one of the dideoxy analogs to each of the four separate incubation mixtures, four different groups of fragments, ending at a particular base, are produced. The other end of all the fragments is identical, so that parallel

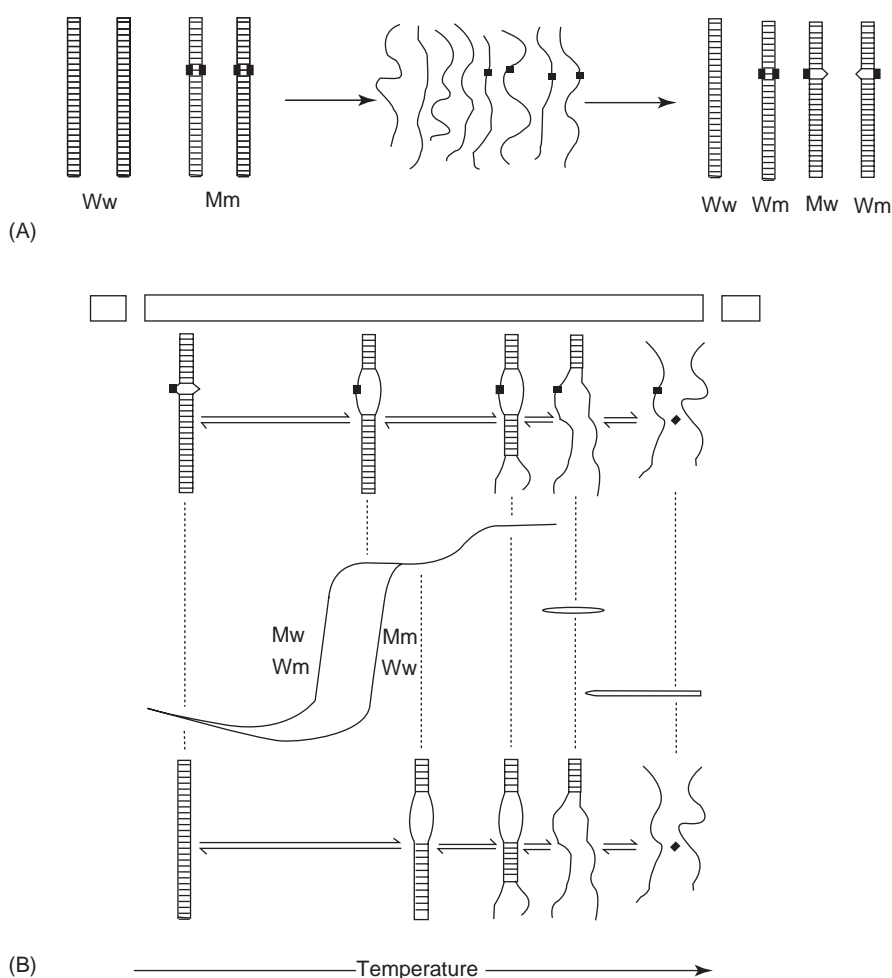
size-fractionation of the denatured fragment groups by analytical PAGE allows the sequence to be deduced.

In alternative to endpoint methods with radionucleosides, it is possible to use four different fluorophores, having four different emission wavelengths, for tagging the primer. The purified DNA fragment to be sequenced is thus divided into four aliquots, each of which is then confronted with a primer, tagged with one of the four fluorophores. The primer extension is then obtained with the Klenow fragment of DNA polymerase I from *Escherichia coli*, or with T7 polymerase or Taq polymerase. The four reaction mixtures are then electrophoresed simultaneously in a gel equipped with a detection system, which comprises a scanning argon laser beam, which sweeps the gel at a fixed position, perpendicular to the band

migration. The excited fluorescence is converted and amplified by a photomultiplier tube, and the digital signal is analyzed by a computer.

## Temperature Gradient Gels

This is a unique mode of performing electrophoresis under conditions of continuous conformational changes. This method is especially suitable for nucleic acid analysis. A 1-mm-thick gel slab is made to contain a long trench at the side of the negative electrode. A linear temperature gradient is established perpendicular to the direction of the electric field. Thus, as the nucleic acid sample is moving toward the opposite electrode, it will migrate against progressively increasing temperatures in the gel.

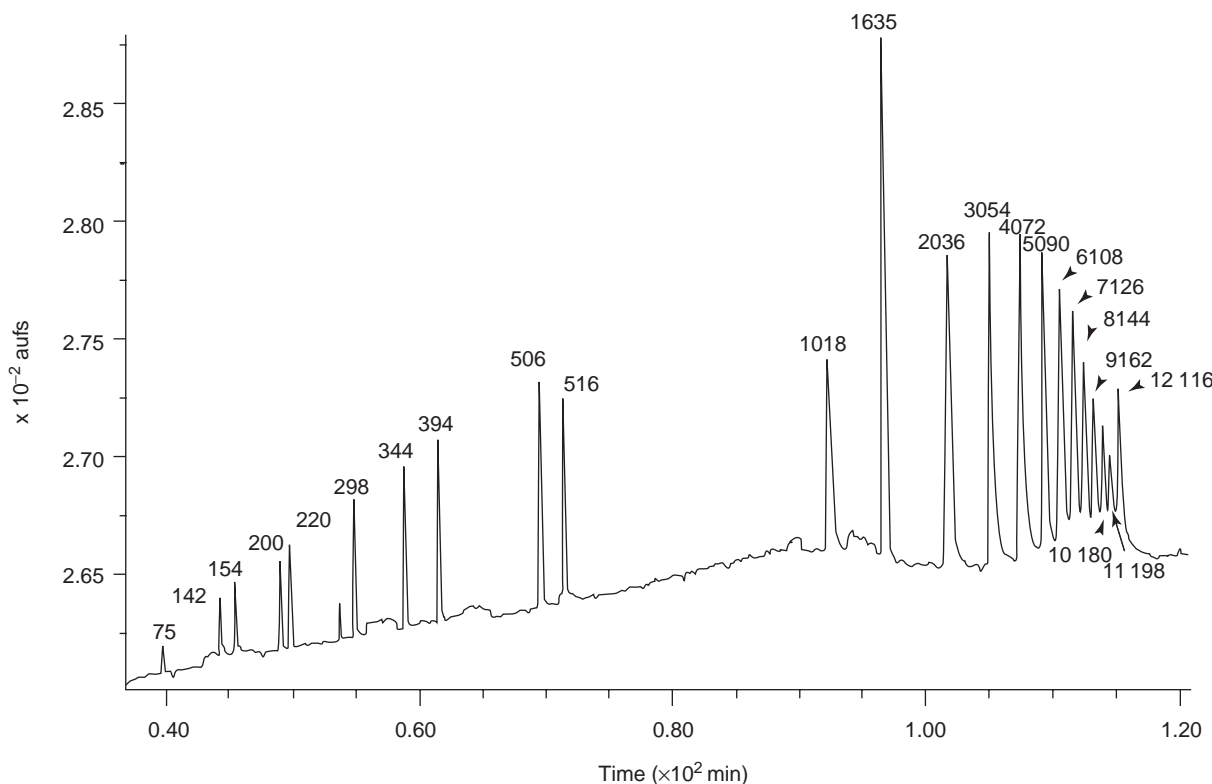


**Figure 6** Example of detection of mutations in form of mismatches in ds-DNA by temperature gradient electrophoresis. (A) Denaturation–renaturation cycle, with a schematic representation of the mechanism demonstrating that a base-pair exchange (■–■) is transferred into a mismatch (■>, <■). (B) Schematic drawing of the transition curves, showing that the denaturation temperature of the segment, which carries the mutation is lowered significantly. W, wild-type sequence, (+) strand; w, wild-type sequence, (–) strand; M, mutated sequence, (+) strand; m, mutated sequence, (–) strand; ♦, point mutation. (Reproduced with permission from Riesner D, Henco K, and Steger G (1991) *Advances in Electrophoresis*, vol. 4, pp. 169–250; VCH.)



A double-stranded (ds) DNA will undergo two highly cooperative conformational transitions: one from the native, base-paired state to a partially denatured state (in general sigmoidal) and the second one to the totally denatured state after strand separation. This helix-coil transition of ds nucleic acids can be described quantitatively by statistical thermodynamics. This method can be applied to a number of analyses: e.g., detection of mutations in the form of base-pair exchanges in ds-RNA and DNA, analysis of cDNA clones with point mutations from site-directed mutagenesis, detection of mutations in the form of mismatches in ds-DNA. An example on the last case is given in **Figure 6**: if a mutant ds-DNA is present in the same solution as the wild-type ds-DNA, hybrids between both can be formed in a denaturation-renaturation cycle (**Figure 6A**). In such an event a mismatch can form in the double helix at the site of mutation. Mismatches always decrease the  $T_m$  value and will be easily detected by electrophoresis in a thermal gradient by a sigmoidal transition occurring at a lower temperature as compared with the wild-type (**Figure 6B**). In general, one expects, from this type of analysis, in the presence of a point mutation,

the formation of four zones: two homo- (Ww and Mm) and two hetero- (Mw and Wm) duplexes. This kind of analysis is known under the acronym DIGE (denaturing gradient gel electrophoresis) and it is better performed in presence of the so-called GC-clamps, i.e., of 30–40 base pair-long GC regions added at one extremity of the amplified DNA region under analysis. Such clamps prevent total strand separation, due to high resistance to denaturation (either chemical or thermal) of GC-rich regions and thus permit the detection of the four zones of homo- and hetero-duplexes. It should be noted that DIGE can be performed not only in temperature gradients, but also in chemical denaturing gradients, such as gradients from 0 to 8 mol l<sup>-1</sup> urea, with identical results. The latter gradients are in fact preferred due to the much easier experimental setup. As a latest evolution of DIGE, a double gradient technique has been reported, by which the denaturing gradient (either thermal or chemical) is co-linear with a porosity gradient in the gel matrix. The latter gradient is needed to recompact the diffuse bands (especially of hetero-duplexes), which melt over a broader range of denaturants and are thus diffuse.



**Figure 7** Analysis of the 1-kb ladder by capillary zone electrophoresis. Conditions: coated capillary of 100  $\mu$ m i.d., 47.4 cm length, filled with 4.5% polyacrylamide strings in 100 mmol l<sup>-1</sup> Tris–borate–EDTA, pH 8.0. Sample: a 0.25  $\mu$ g ml<sup>-1</sup> solution of the 1-kb ladder was injected electrophoretically for 3 s at 4 kV; run: at 4 kV and 11.4  $\mu$ A. The numbers on each peak represent the length of each fragment, in number of bases. Tris, 2-amino-2-hydroxymethylpropane-1,3-diol; EDTA, ethylenediaminetetraacetic acid.

## Physical versus Chemical Gels

This is perhaps the newest and most exciting development in zone electrophoresis. While so far most runs have been performed in 'chemical' gels, cross-linked (see **Figure 1**) so as to form a fixed chain distribution in space, it appears that analogous results can be obtained in 'physical' gels (or polymer networks), i.e., in viscous solutions of long, linear strings of monomer, in the absence of cross-linker. Sieving in 'physical' gels is just as effective as in 'chemical' gels, provided that the polymer solution is above the entanglement threshold, which, for polyacrylamides, is set at  $\sim 0.5\%$  concentration. The larger the object to be separated, the more dilute will be the polymer solution and vice versa. Thus, for separating large DNA fragments, up to 22 kbp, a 4.5% polymer concentration will be adequate. But, for DNA sequencing gels, up to 100 bp, a 10% polymer solution will be required. For running native proteins, solutions up to 15% and higher are often required. The advantage of physical gels, especially in capillary zone electrophoresis, is the possibility of filling the separation chamber (often of very narrow bore, e.g., 20  $\mu\text{m}$  i.d.) without trapping air bubbles. In addition, 'physical' gels can be reused several times (up to 50) without trapping or precipitating the sample at the injection port, a phenomenon that occurs regularly in 'chemical' gels, due to the large fiber diameter and thus to a greater surface available for sample adsorption and trapping. **Figure 7** shows the power of separation in physical gels, when analyzing a large spectrum of DNA sizes, ranging from as little as 75 bp up to 12 kbp.

See also: **Electrophoresis**: Overview; Isoelectric Focusing; Two-Dimensional Gels; Blotting Techniques; Proteins. **Forensic Sciences**: DNA Profiling. **Proteins**: Traditional Methods of Sequence Determination. **Proteomics**.

## Further Reading

- Andrews AT (1986) *Electrophoresis: Theory, Techniques and Biochemical and Clinical Applications*. Oxford: Clarendon Press.
- Chiari M, Micheletti C, Nesi M, Fazio M, and Righetti PG (1994) Towards new formulations for polyacrylamide matrices: N-acryloylamino ethoxyethanol, a novel monomer combining high hydrophilicity with extreme hydrolytic stability. *Electrophoresis* 15: 177–186.
- Cremonesi L, Firpo S, Ferrari M, Righetti PG, and Gelfi C (1997) Double-gradient DGGE for optimized detection of DNA point mutations. *BioTechniques* 22: 326–330.
- Davis BJ (1964) Disc electrophoresis II: method and application to human serum proteins. *Annals of the New York Academy of Sciences* 121: 404–427.
- Dunn MJ (ed.) (1986) *Gel Electrophoresis of Proteins*. Bristol: Wright.
- Gelfi C, De Besi P, Alloni A, and Righetti PG (1992) Properties of novel acrylamido-monomers as investigated by capillary zone electrophoresis. *Journal of Chromatography* 608: 333–341.
- Gelfi C, Alloni A, De Besi P, and Righetti PG (1992) Properties of acrylamido-bifunctional monomers (cross-linkers) as investigated by capillary zone electrophoresis. *Journal of Chromatography* 608: 343–348.
- Jorgenson JW and Phillips M (eds.) (1987). *New Directions in Electrophoretic Methods*. Washington: American Chemical Society Symposium Series, vol. 335.
- Maurer HR (1972) *Disc Electrophoresis*. Berlin: De Gruyter.
- Riesner D, Henco K, and Steger G (1991) Temperature gradient gel electrophoresis. In: Chrambach A, Dunn MJ, and Radola BJ (eds.) *Advances in Electrophoresis*, vol. 4, pp. 169–250. Weinheim, VH:
- Righetti PG, van Oss CJ, and Vanderhoff JW (1979) *Electrokinetic Separation Methods*. Elsevier: Amsterdam.
- Righetti PG (1989) Of matrices and men. *Journal of Biochemical and Biophysical Methods* 19: 1–20.
- Simò-Alfonso E, Gelfi C, Sebastiano R, Citterio A, and Righetti PG (1996) Novel acrylamido monomers with higher hydrophilicity and improved hydrolytic stability. II: properties of N-acryloyl amino propanol. *Electrophoresis* 17: 732–737.

## Two-Dimensional Gels

P G Righetti, University of Verona, Verona, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Two-dimensional, polyacrylamide gel electrophoresis (2D PAGE) is an open assay approach, which permits the separation and detection of proteins from a wide variety of sources without the need for any prior

knowledge of function. Two-dimensional PAGE is capable of profiling many thousands of proteins on a single matrix with exquisite resolution, separating very discrete isoforms differing by post-translational modifications as subtle as a single deamidation event. The basic elements for the development of 2D PAGE were laid out in the mid-1960s, with the description of isoelectric focusing (IEF), by which proteins would be separated essentially on the basis of their net charge (at a given value of their titration curve, the

## Physical versus Chemical Gels

This is perhaps the newest and most exciting development in zone electrophoresis. While so far most runs have been performed in 'chemical' gels, cross-linked (see **Figure 1**) so as to form a fixed chain distribution in space, it appears that analogous results can be obtained in 'physical' gels (or polymer networks), i.e., in viscous solutions of long, linear strings of monomer, in the absence of cross-linker. Sieving in 'physical' gels is just as effective as in 'chemical' gels, provided that the polymer solution is above the entanglement threshold, which, for polyacrylamides, is set at  $\sim 0.5\%$  concentration. The larger the object to be separated, the more dilute will be the polymer solution and vice versa. Thus, for separating large DNA fragments, up to 22 kbp, a 4.5% polymer concentration will be adequate. But, for DNA sequencing gels, up to 100 bp, a 10% polymer solution will be required. For running native proteins, solutions up to 15% and higher are often required. The advantage of physical gels, especially in capillary zone electrophoresis, is the possibility of filling the separation chamber (often of very narrow bore, e.g., 20  $\mu\text{m}$  i.d.) without trapping air bubbles. In addition, 'physical' gels can be reused several times (up to 50) without trapping or precipitating the sample at the injection port, a phenomenon that occurs regularly in 'chemical' gels, due to the large fiber diameter and thus to a greater surface available for sample adsorption and trapping. **Figure 7** shows the power of separation in physical gels, when analyzing a large spectrum of DNA sizes, ranging from as little as 75 bp up to 12 kbp.

See also: **Electrophoresis**: Overview; Isoelectric Focusing; Two-Dimensional Gels; Blotting Techniques; Proteins. **Forensic Sciences**: DNA Profiling. **Proteins**: Traditional Methods of Sequence Determination. **Proteomics**.

## Further Reading

- Andrews AT (1986) *Electrophoresis: Theory, Techniques and Biochemical and Clinical Applications*. Oxford: Clarendon Press.
- Chiari M, Micheletti C, Nesi M, Fazio M, and Righetti PG (1994) Towards new formulations for polyacrylamide matrices: N-acryloylamino ethoxyethanol, a novel monomer combining high hydrophilicity with extreme hydrolytic stability. *Electrophoresis* 15: 177–186.
- Cremonesi L, Firpo S, Ferrari M, Righetti PG, and Gelfi C (1997) Double-gradient DGGE for optimized detection of DNA point mutations. *BioTechniques* 22: 326–330.
- Davis BJ (1964) Disc electrophoresis II: method and application to human serum proteins. *Annals of the New York Academy of Sciences* 121: 404–427.
- Dunn MJ (ed.) (1986) *Gel Electrophoresis of Proteins*. Bristol: Wright.
- Gelfi C, De Besi P, Alloni A, and Righetti PG (1992) Properties of novel acrylamido-monomers as investigated by capillary zone electrophoresis. *Journal of Chromatography* 608: 333–341.
- Gelfi C, Alloni A, De Besi P, and Righetti PG (1992) Properties of acrylamido-bifunctional monomers (cross-linkers) as investigated by capillary zone electrophoresis. *Journal of Chromatography* 608: 343–348.
- Jorgenson JW and Phillips M (eds.) (1987). *New Directions in Electrophoretic Methods*. Washington: American Chemical Society Symposium Series, vol. 335.
- Maurer HR (1972) *Disc Electrophoresis*. Berlin: De Gruyter.
- Riesner D, Henco K, and Steger G (1991) Temperature gradient gel electrophoresis. In: Chrambach A, Dunn MJ, and Radola BJ (eds.) *Advances in Electrophoresis*, vol. 4, pp. 169–250. Weinheim, VH:
- Righetti PG, van Oss CJ, and Vanderhoff JW (1979) *Electrokinetic Separation Methods*. Elsevier: Amsterdam.
- Righetti PG (1989) Of matrices and men. *Journal of Biochemical and Biophysical Methods* 19: 1–20.
- Simò-Alfonso E, Gelfi C, Sebastiano R, Citterio A, and Righetti PG (1996) Novel acrylamido monomers with higher hydrophilicity and improved hydrolytic stability. II: properties of N-acryloyl amino propanol. *Electrophoresis* 17: 732–737.

## Two-Dimensional Gels

P G Righetti, University of Verona, Verona, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

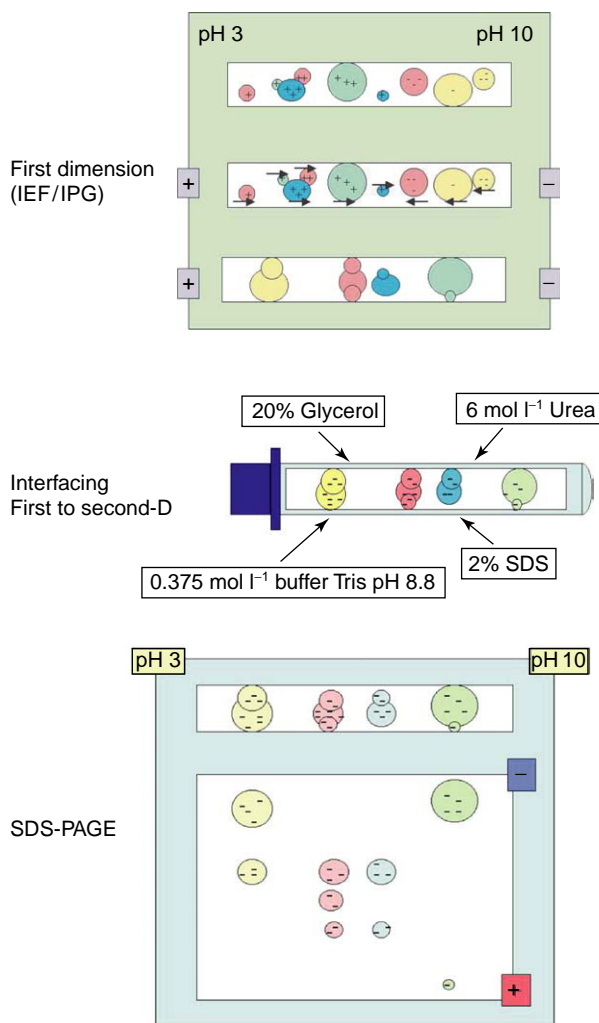
Two-dimensional, polyacrylamide gel electrophoresis (2D PAGE) is an open assay approach, which permits the separation and detection of proteins from a wide variety of sources without the need for any prior

knowledge of function. Two-dimensional PAGE is capable of profiling many thousands of proteins on a single matrix with exquisite resolution, separating very discrete isoforms differing by post-translational modifications as subtle as a single deamidation event. The basic elements for the development of 2D PAGE were laid out in the mid-1960s, with the description of isoelectric focusing (IEF), by which proteins would be separated essentially on the basis of their net charge (at a given value of their titration curve, the

isoelectric point, pI). Almost coeval was the discovery of sodium dodecyl sulfate (SDS) electrophoresis, by which surfactant-laden proteins would be separated essentially via their mass, rather than by combination of mass and charge effects, as customary in conventional zone electrophoresis. Now, the elements for the next major quantum leap were laid out on the table. In the mid-1970s, O'Farrell took advantage of this previous know-how and reported the creation of 2D maps, by combining orthogonally a first dimension, based on pure charge fractionation, with a second dimension, based on size discrimination of denatured and fully dissociated polypeptide chains. Although the original O'Farrell approach was very advanced at the methodologically level, it took, nevertheless, many years of developments and refinements for bringing the technique to the present-day highly sophisticated level. Instrumental to that were major contributions from the field of informatics, which had to develop new algorithms for mapping the 2D space, acquiring the images, cleaning the background, matching different maps among themselves. Informatics has also greatly contributed by laying out a format for protein databases, which today represent a formidable tool in protein characterization. Another major contribution was the introduction, in 1982, of immobilized pH gradients (IPG), which offered a new, most powerful view of the field, guaranteeing much increased resolution and much higher reproducibility in spot position. Just as fundamental was the introduction, in the early 1990s, of mass spectrometry (MS) as a tool for sequencing small peptides and for identifying proteins. Up to the present, 2D mapping was a synonym for proteome analysis, although today a host of 2D chromatographic approaches (or hybrid, hyphenated chromatographic and electrophoretic techniques) are rivaling with the old, well-ingrained 2D PAGE. With a proviso, though: most chromatographic methods start with a fully digested tissue extract, and thus have to deal with a tremendously heterogeneous peptide population; 2D maps, on the contrary, deal only with intact polypeptide chains, as a genuine snapshot of the proteomic patrimony of the tissue under investigation.

## Sample Solubilization and Pretreatment

Figure 1 shows what is perhaps the most popular approach to 2D map analysis today. The first dimension is preferably performed in individual IPG strips, laid side by side on a cooling platform, with the sample often adsorbed into the entire strip during



**Figure 1** Pictorial representation of the steps required for obtaining a 2D map, with the first dimension typically done in IPG strips, followed by equilibration of the focused strips in SDS-interfacing solution and finally by the second dimension run in SDS-PAGE.

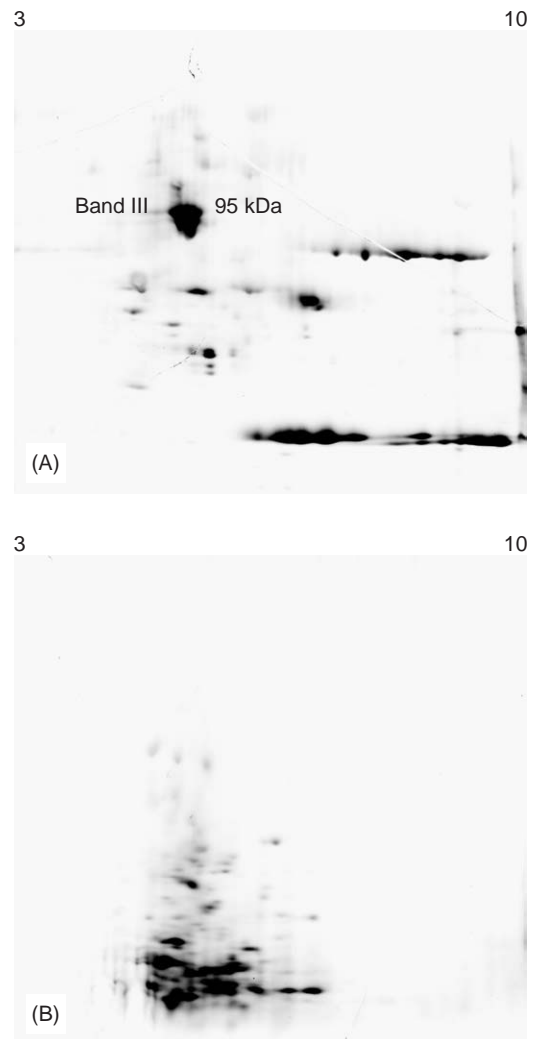
rehydration. At the end of the IEF step, the strips have to be interfaced with the second dimension, almost exclusively performed by mass discrimination via saturation with the anionic surfactant SDS. After this equilibration step, the strip is embedded on top of an SDS-PAGE slab, where the second-dimensional run is carried out perpendicular to the first-dimensional D migration. The 2D map displayed at the end of these steps is the stained SDS-PAGE slab, where polypeptides are seen, after staining, as (ideally) round spots, each characterized by an individual set of pI/ $M_r$  coordinates (pI, isoelectric point;  $M_r$ , molecular mass).

Instrumental to the success of 2D PAGE is sample treatment prior to any electrophoretic step. To start with, one should recall that, when cells are lysed,



hydrolases (phosphatases, glycosidases, and especially proteases) are in general liberated or activated. In the presence of chaotropes, glycosidases and phosphatases are quickly denatured, but the denaturation of proteases might have slower kinetics. This could result in much altered spots in the subsequent 2D analysis, to the point that no band recognition is any longer possible. Proteases are in general inhibited when the tissue is disrupted directly in strong denaturants, such as  $8 \text{ mol l}^{-1}$  urea, 10% trichloroacetic acid, or 2% SDS. But there are still two major problems: (1) some proteases might be more resistant to denaturation than the majority of the other cellular proteins, so that, while the former have been unfolded, proteases might have some time for attack before being denatured too; (2) in other cases (e.g., red blood cells, RBCs) the cells might have to be lysed under native conditions, in order to eliminate some cellular components, so that there is ample time for strong proteolytic aggression. In any event, proteases are less active at lower temperature, which is why, during cell disruption, low temperatures ( $2\text{--}4^\circ\text{C}$ ) are in general recommended. In addition, most tissue proteases are inactive above pH 9.5, so that proteolysis can often be inhibited by preparing the sample in Tris-free base, sodium hydrogen carbonate, or basic carrier ampholytes (CAs) at pH values  $\sim 10$ . Some proteases, however, might retain activity even when all the above precautions are taken. Thus, as a safety precaution, it is advisable to use a cocktail of protease inhibitors. Such combinations are available from a number of commercial sources and in general comprise both chemical and proteinaceous inhibitors (e.g., phenylmethylsulfonyl fluoride, ethylenediaminetetraacetic acid, aprotinin). The importance of safeguarding the sample against accidental proteolytic attack during preparation should never be underestimated. A case in point is the dramatic example of 2D maps of RBC membranes, given in **Figure 2**, when processed in the absence of presence of proper inhibitor cocktails. The primary concern when preparing RBC membranes for electrophoresis is to remove hemoglobin, which is present at extremely high levels in the RBC and can cause severe streaking in 2D maps. Most of the procedures in the literature involve washing the RBCs in physiological buffered saline followed by lysis and washing in hypotonic buffer. However, there has been a lack of attention paid to the activation of proteases upon RBC lysis. It is seen here that, in the absence of protease inhibitors, a peptide map (B) is generated, rather than a map of intact protein species (A).

The second major advance in 2D maps has been a much improved solubilization cocktail, after the classical O'Farrell mixture, comprising  $9 \text{ mol l}^{-1}$



**Figure 2** Two-dimensional map of RBC membranes: (A) treated with protease inhibitors during the lysis and washing steps; (B) control, untreated with protease inhibitors. Note, in this last case, the massive proteolytic action. The numbers at the gel top indicate the pH interval (pH 3–10, nonlinear). (Reproduced with permission from Olivieri E, Herbert B, and Righetti PG (2001) The effect of protease inhibitors on the two-dimensional electrophoresis pattern of red blood cell membranes. *Electrophoresis* 22: 560–565.)

urea, 2% Nonidet P-40, 2%  $\beta$ -mercaptoethanol, and 2% CAs. Membrane proteins require a much stronger solubilizing power, which has been ensured, over the years, by resorting to more powerful mixtures. The consensus cocktail in vogue today comprises  $5\text{--}7 \text{ mol l}^{-1}$  urea,  $2\text{--}3 \text{ mol l}^{-1}$  thiourea, and 2% amidosulfofetaines. Thiourea, in view of its marked hydrophobicity (it is soluble in water only up to  $1 \text{ mol l}^{-1}$  level; thus urea acts as a co-solvent), dramatically increases the solubilizing power of this mixture. So does the surfactant amidosulfofetaine, a modified dodecylsulfofetaine. This last class of surfactants had been found to be by far the class

with the highest solubilizing potency, perhaps due to their resemblance with the anionic SDS moiety, but it was precipitated by levels of urea  $>4 \text{ mol l}^{-1}$ . By breaking the tail hydrophobicity with the introduction of an amido group, the sulfobetaines were made compatible with levels of urea up to  $10 \text{ mol l}^{-1}$ .

Another factor that influences protein solubilization is the choice of the reducing agent. It is agreed that to solubilize a complex mixture of proteins in preparation for IEF, it is necessary to completely break inter- and intrachain disulfide bonds by reduction. This is commonly achieved using cyclic reducing agents, such as dithiothreitol or dithioerythritol, which are added to the solubilizing cocktails in large excess ( $20\text{--}40 \text{ mmol l}^{-1}$ ), so as to shift the equilibrium toward oxidation of the reducing agent with concomitant reduction of the disulfide bridges. Because this is an equilibrium reaction, loss of the reducing agent through proteins migration away from the sample application zone, during the first IEF dimension run, can result in reoxidation of free Cys, an effect that has two possible consequences: horizontal streaking and formation of spurious bands due to scrambled  $\text{--S--}$  bridges and their cross-linking various polypeptide chains. This situation would be even worse when using  $\beta$ -mercaptoethanol, which has a lower  $pK$  value; thus, it will be more efficiently depleted in the alkaline region and will form a concentration gradient toward pH 7, with a distribution in the gel according to its extent of ionization at a given pH value along the IEF strip. Recently, it was demonstrated that the use of a non-charged reducing agent, such as tributyl phosphine (TBP), greatly enhanced protein solubility during the IEF and resulted in a much improved transfer to the second dimension. It is worth noting that the contribution of thiourea to solubility follows a different route from that of TBP: the first increases the chaotropic capability of the sample solution, while TBP ensures the necessary reducing conditions during the IEF stage. Curiously, though, whereas the need for adding a reducing agent has been recognized since the early days of O'Farrell, up to recent times no 2D PAGE protocol reported the other complementary step, i.e., alkylation of free  $\text{--SH}$  groups following their reduction. Failure to reduce and fully alkylate proteins prior to the first dimension can result in a large number of spurious spots in the alkaline pH region, due to 'scrambled' disulfide bridges among like and unlike chains. This potential series of artifactual spots can comprise dimers, trimers, up to nonamers, and perhaps higher oligomers. This statement can be easily supported by the data in **Figures 3A** and **3B**, which refer to 2D maps of globin chains. Map (B), which was obtained following

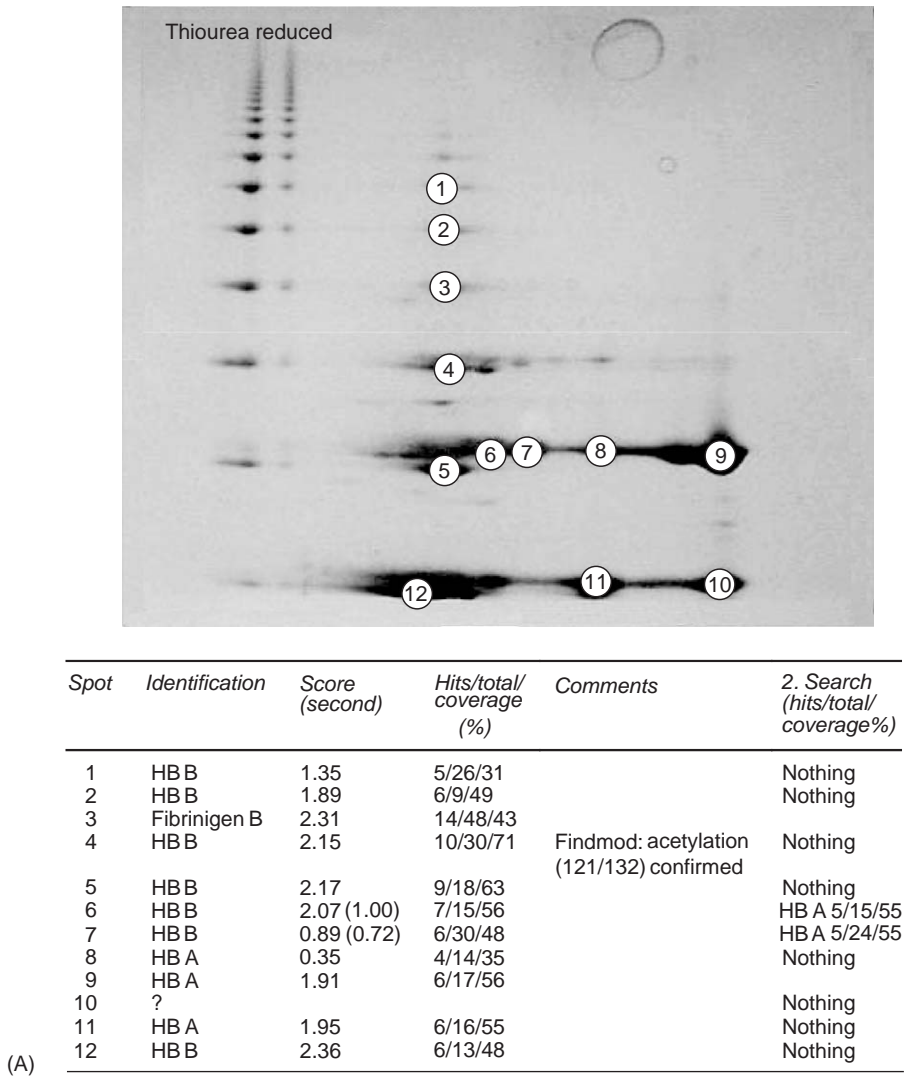
reduction and alkylation prior to the first dimension, presents the expected simple pattern, composed of a few spots aligned at  $\sim 16 \text{ kDa}$ ; the map in (A), on the other hand, was obtained without alkylation, which explains the presence of a number of strings, which MALDI-TOF-MS (matrix-assisted laser desorption ionization time-of-flight MS) identified as multiples of the above  $M_r$ .

Another important problem is the proper choice of the alkylating agent. By far the most popular chemical, up to the present, has been iodoacetamide. Yet, this compound, in presence of thiourea, is quickly destroyed, thus jeopardizing the alkylation reaction. An alternative solution would be to alkylate not with an iodinated agent, but with acrylamide, by exploiting the reactivity of  $\text{--SH}$  group toward addition to double bonds. An even better solution is to alkylate at neutral pH with 2- (or 4-)vinylpyridine, since this reagent has been found to offer 100% reactivity coupled to 100% specificity. The specificity is ensured by lowering the reaction pH from 9.5 (as customary for alkylation of the  $\text{--SH}$  group) to pH 7, where Lys and other side chains cannot possibly react; the sustained reaction is ensured by the fact that this pH value is half a way between the  $pK$ s of the reactants, thus providing fractional positive and negative charges onto the two reactants. **Figure 4** shows that indeed, with vinylpyridine, 100% alkylation of all reacting  $\text{--SH}$  groups is obtained, contrary to alkylation with other chemicals.

## The First Dimension

In the past, the vast majority of 2D maps were run by conventional IEF with soluble CAs for the first dimension. However, due to poor reproducibility in both gradient shape and CA synthesis, there has been a marked shift toward the use of IPGs. Precast, dried Immobiline gels, encompassing a number of ranges, are now available from Amersham Pharmacia Biotech, Bio-Rad, and Proteome Systems. They all contain 4%T (total monomer concentration, expressed as grams acrylamide + grams Bis per 100 ml of solution) and they span the following narrow pH ranges: 3.5–4.5; 4.0–5.0; 4.5–5.5; 5.0–6.0; 5.5–6.7. In addition, there are a number of wide pH ranges: 4–7L; 6–9; 6–11; 3–10L, and 3–10NL (L = linear; NL = nonlinear). Some of them are available in 7, 11, 13, and 18 cm in length, whereas all the narrow ranges are cast only as long (18 cm) strips (more recently, the trend is toward longer IPG strips, e.g., 24 cm in length). All of them are 3 mm wide and, when reswollen, 0.5 mm thick (gel layer). Precast, dried IPG gels in alkaline narrow ranges should be





**Figure 3** Two-dimensional maps of globin chains: (A) reduced but nonalkylated sample; (B) reduced and alkylated sample prior to the focusing step. Note, in panel A, the presence of higher-order homooligomers of  $\beta$ -chains, detected up to the dodecamer level. Note additionally the presence of homodimers of  $\alpha$ -chains, as well as of a heterodimer  $\alpha$ - $\beta$ . Note, in panel B, the complete removal of all higher-order oligomers and the detection of minor components of RBC lysates, such as carbonic anhydrase and thioreduction reductase, which would have been hidden under the string of dimers in nonalkylated samples. (Reproduced with permission from Herbert B, Galvani M, Hamdan M, *et al.* (2001) Reduction and alkylation of proteins in preparation of two-dimensional map analysis: why, when and how? *Electrophoresis* 22: 2046–2057.)

handled with care, because at high pHs the hydrolysis of both the gel matrix and the Immobiline chemicals bound to it is much more pronounced. Such gels, when cast onto a Gel-Bond PAG foil, have dimensional stability and can be handled quite safely. The fact that they are rather dilute (4%T) improves their performance, since there is less steric hindrance to large proteins, thus permitting shorted focusing times. In addition, if required, even more dilute matrices can be cast, down to  $\sim 3\%$ T (with variable amounts of cross-linker, in general from 3–5%C) (C, grams of cross-linker per 100g of total monomers). Although the most popular cross-linker is Bis

(*N,N'*-methylene bisacrylamide), good performance is also given by adopting bisacrylylpiperazine as crosslinking agent, since the latter is more stable to hydrolysis and is immune from the strongly yellowish backgrounds produced during silver staining in bis-gels. Routinely, IPG gels are cast to provide an average buffering power ( $\beta$ ) of 3 mequiv. $l^{-1}$ , which corresponds to total Immobiline molarities in the gel (buffering and titrant ions) of the order of 10 mmol $l^{-1}$ . However, when working at pH extremes (e.g., in the pH 2.5–4.5 and 9.5–11.5 intervals), this molarity should be doubled, so as to counteract the  $\beta$  power of the bulk solvent (water).

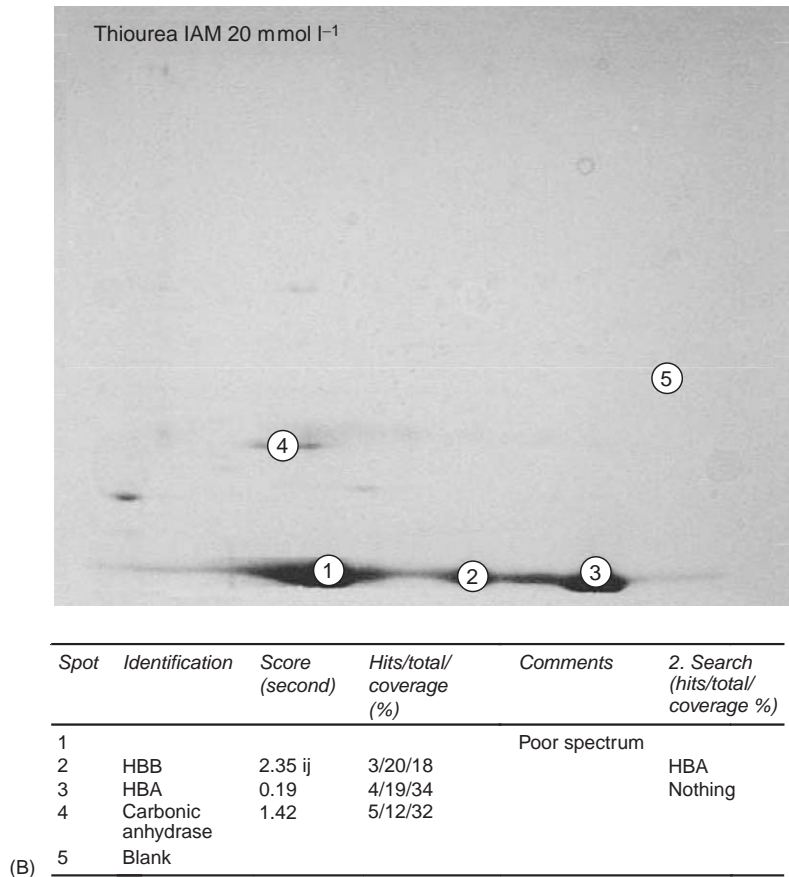


Figure 3 (Continued).

Since the IPG strips are supplied as dried films, there are different types of sample loading, as illustrated below.

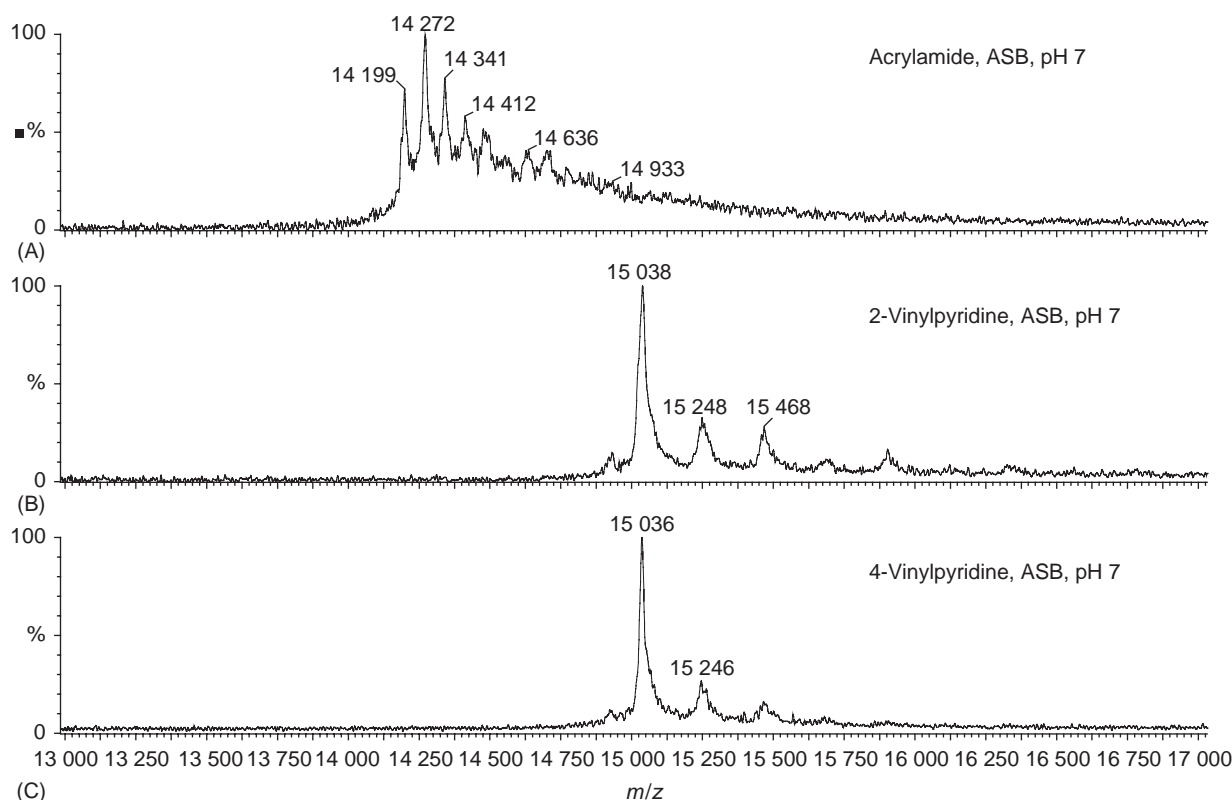
**Pulse Loading**

This is the method typically adopted in all fractionation techniques, including chromatography, since due to entropic forces tending to dissipate the sample zone, the latter have in general to be applied as a very thin zone. This, of course, does not hold true for any focusing procedure, which reaches steady-state conditions. Nevertheless, it has been customary, for years, to apply the sample (especially in the old procedure of CA-IEF) in pockets precast in the gel, or in surface basins or adsorbed onto granulated materials (e.g., Sephadex, Paratex films, and the like). Sample cups did seem to offer a valid alternative to loading, since they permitted higher volume loadings (up to 100 µl) than with any other conventional means. To this purpose, Amersham Pharmacia Biotech devised a bar holder, able to position of battery of cups on a linear array of pre-hydrated IPG strips. This holder would gently press the cups against the gel surface, so

as to prevent sample leakage; each cup would then be filled with the desired volume of any sample. This method too is being used less and less, since it was found that significant amounts of proteins were lost at the application point, due to protein aggregation and precipitation as they migrated from the free liquid phase into the gel, with the ultimate result being poor resolution and ‘underloading’. This effect has been explained as a concentration phenomenon occurring at the gel-sample interface, where proteins massively accumulate as they are dramatically slowed down in their transit from a liquid phase to a gel phase. At this interface also anomalous ionic boundaries will form and the two concomitant phenomena would favor protein aggregation and precipitation.

**In-Gel Rehydration Loading**

This method is today by far preferred since it facilitates higher protein loads and reduces focusing times. With this technique, the dehydrated IPG strip is directly reswollen with the protein sample dissolved in the rehydration solution (notice that, in this last case,



**Figure 4** MALDI-TOF mass spectra of bovine  $\alpha$ -lactalbumin after 1 h incubation with 2% amidosulfobetaine-14 and: (A) acrylamide, (B) 2-vinylpyridine, and (C) 4-vinylpyridine. In all cases the reaction has been carried out at pH 7.0. Note, in both panels B and C, a single reaction channel of LCA with 2VP and 4VP, respectively. The higher-order peaks ( $m/z$  15 248 and 15 468 in B and C) represent adducts LCA with the MALDI matrix, sinapinic acid (unpublished data).

the same solution is used to solubilize the protein and rehydrate the IPG strip). After a suitable rehydration time ( $\sim 5$ – $6$  h), the IPG strip is ready for the first dimension IEF, with the proteins already uniformly distributed within the gel matrix. The clear advantage of in-gel rehydration is the large volume of samples that can be applied (up to  $450\ \mu\text{l}$  for a 18 cm long, 4 mm wide, 0.5 mm thick IPG strip) compared to  $100\ \mu\text{l}$  for conventional cup loading. The other major advantage is minimization of sample aggregation and precipitation: since the sample is diluted through the entire gel strip, no local, accidental build ups of proteins can occur. The third advantage of this protocol is that, due to dilution of proteins throughout the gel, much accelerated focusing protocols can be adopted, utilizing higher voltage gradients at the very start of the IEF/IPG run.

#### In-Gel Hydration Loading under a Low Voltage

This technique, also called active sample loading, is distinguished from the previous one (called passive), in that, during IPG strip rehydration, a gentle voltage gradient (typically 50 V) is applied, usually for the

entire duration of the reswelling (overnight). It is believed that this procedure would further facilitate sample entry. Recently, in fact, an interesting aspect of such active sample loading was reported: its ability to facilitate entry of high  $M_r$  proteins, such as spectrins, large (280 kDa for the  $\alpha$ -chains), filamentous proteins, which represent the major constituents of the cytoskeletal network underlying the RBC plasma membrane and are typically associated with band 4.1 and actin to form the cytoskeletal superstructure of the RBC plasma membrane.

## The Second Dimension

As shown in Figure 1, after the focusing step, the gel strip is transferred to a small test tube and bathed in a 10 volume excess of interfacing solution, for up to 15 min, so as to begin to saturate the isoelectric proteins with the SDS anionic surfactant. Thiol agents (e.g., DTT, TBP) can be omitted if the protein mixture has been properly reduced and alkylated prior to the focusing step. The equilibrated strips are then individually transferred to the upper edge of a

vertically standing SDS-gel cassette and cemented *in situ* by pouring over it a hot agarose solution. The overlay agarose recipe consists of 25 mmol l<sup>-1</sup> Tris base, 192 mmol l<sup>-1</sup> glycine, 0.1% SDS and 0.8% agarose (traces of bromophenol blue can be added, to mark the leading ion boundary during the SDS run).

The SDS-gel for the second dimension run can be cast in different ways. Quite a number of users prefer discontinuous gels, made after the classical Laemmli recipe. The stacking gel (usually 1–2 cm, long, made of a 3–4%T, 2.6%C, porous matrix) typically contains 125 mmol l<sup>-1</sup> Tris base, titrated to pH 6.8 with HCl, and 0.1% SDS. The running gel is in general a small pore gel, usually 10–12%T (2.6%C) cast in 375 mmol l<sup>-1</sup> Tris base, titrated to pH 8.8 with HCl, in presence of 0.1% SDS. The upper buffer reservoir (cathode) contains the terminating ion, 192 mmol l<sup>-1</sup> glycine, titrated with 25 mmol l<sup>-1</sup> Tris to pH 8.3, in the presence of 0.1% SDS. Since the anodic reservoir does not play any role in formation of the isotachophoretic train, it is usually filled with the same buffer employed as catholyte. In this experimental setup, the running gel is a constant-concentration matrix.

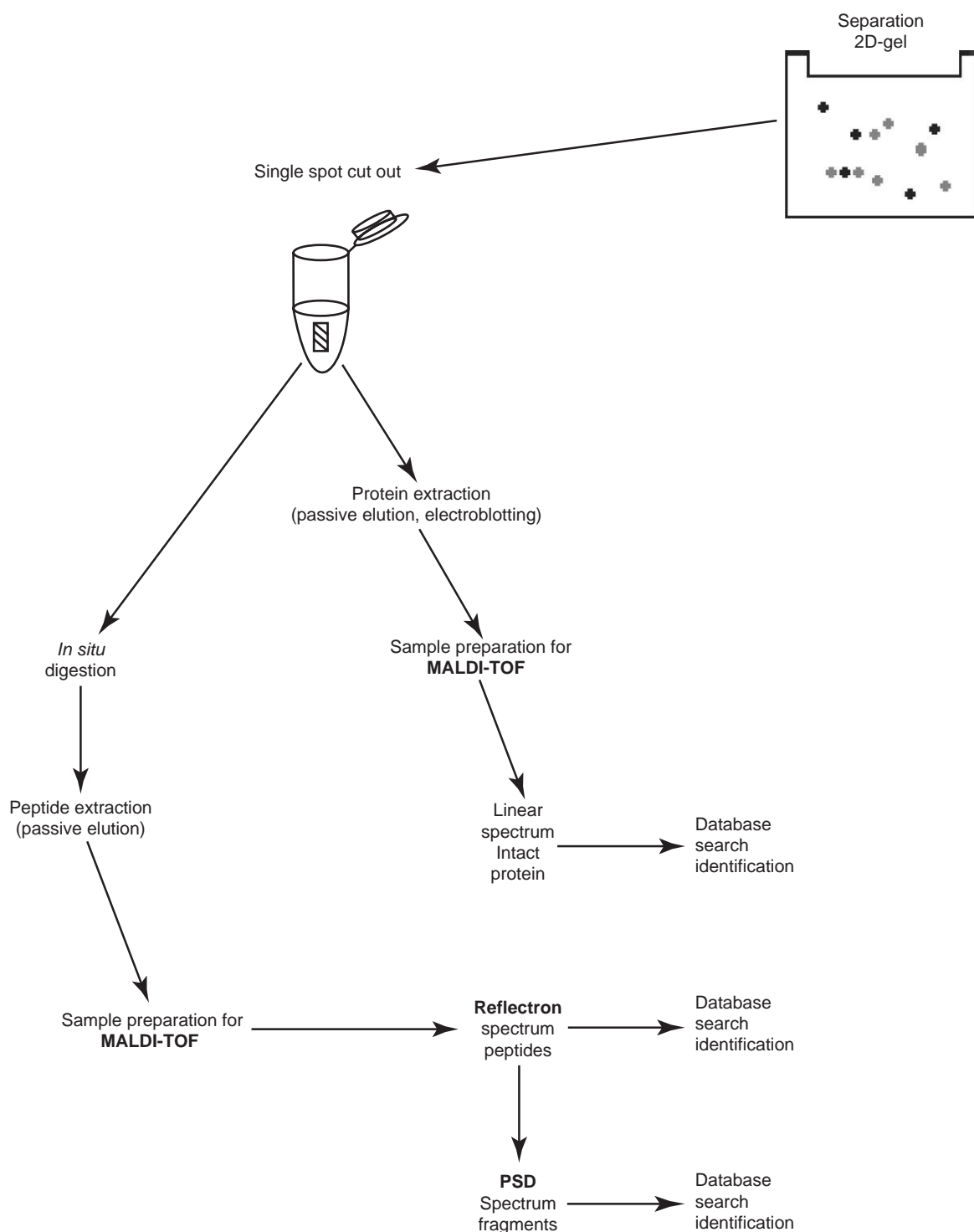
In another variant, the discontinuous Laemmli buffer is coupled to a gel of graded porosity as a running gel. In general, this last gel is cast with the aid of a two-vessel gradient mixer and the porosity gradient typically spans a 6%T to 20%T acrylamide gradient, although higher upper limits have been reported, up to 30%T in case of mixtures of rather small  $M_r$  proteins. A number of authors, though, deem the coupling of discontinuous buffers with a porosity gradient gel an extra burden in running the second dimension of 2D maps. Thus, in general, one or the other technique is used and rarely a combination of the two. It should be remembered, in fact, that porosity gradient gels also have a band-sharpening effect, since the leading edge of each zone is at any given time point migrating at a slower rate than its trailing edge, due to the higher sieving produced by the continuously diminishing pore size along the migration path.

The SDS gel slab, when stained at the end of the second dimension run, is the final 2D map displaying all the polypeptide chains spread randomly in the pI/ $M_r$  plane. For high sensitivity, this gel slab is in general stained with silver stains, deemed to be the most sensitive today available. However, as the 2D map has to be further processed, with elution of the various spot and their analysis by a variety of (MS methods, the silver stains employed cannot use glutaraldehyde or other aldehydes that could irreversibly modify the proteins, impeding proper

MS identification. MS-compatible silvering protocols have been described, but those markedly loose sensitivity, by as much as one order of magnitude. Thus, many laboratories prefer the use of colloidal Coomassie stains, which are fully compatible with MS analysis and have sensitivities similar to those of MS-compatible silver protocols. Fluorescent stains are also gaining momentum, in particular Sypro Ruby, a very high sensitivity stain, fully compatible with any MS approach. Imaging gels tagged with fluorophores, though, requires special (and costly) equipment, since regular gel scanners are not equipped with fluorescent-spot reading devices.

### Mass Spectrometry Analysis of Spots Eluted from 2D Maps

Obtaining a good 2D map, with nice, round spots well spread out in the 2D plane is certainly a good start, but by all means is only a part of the vast job required in proteome analysis, namely the discovery of the identity of as many polypeptide chains as possible in the map thus generated. Here, MS plays a fundamental role. **Figure 5** outlines one of the most popular procedures adopted in view of MS analysis. Single spots are cut out from the SDS gel slab (or from a blot onto, e.g., cellulose nitrate membranes) and eluted by various means or even digested *in situ*. In the last case, the eluted peptides are subjected directly to MS analysis, e.g., via MALDI-TOF. In the former case, the intact protein, once eluted, is injected directly in the mass spectrometer, which results in the  $M_r$  value of the undigested macromolecule. Obtaining both the precise  $M_r$  of the intact polypeptide chain as well as of its tryptic fragments greatly facilitates searching and identification of the species under analysis in the various databases. In order to improve data quality, a number of procedures are routinely adopted. First of all, when eluting a spot (or its tryptic fragments) directly from a gel piece, it is customary to adsorb the liquid onto a Zip-Tip, i.e., a tiny plastic cone of automatic pipettors filled with a C<sub>18</sub>-resin, for purification purposes. The tip is washed with water, so as to remove salt and other contaminants. The adsorbed polypeptides or peptides are then eluted with acetonitrile, a solvent fully compatible with subsequent MS analysis. Finally, it should be borne in mind that the data thus generated are subjected to a number of filtering procedures, so as to eliminate spurious data, such as those that could correspond to contaminants. As an example, all mass values that could be attributed to trypsin (or its fragments), Coomassie Blue adducts, chaperones, are automatically rejected. In addition,

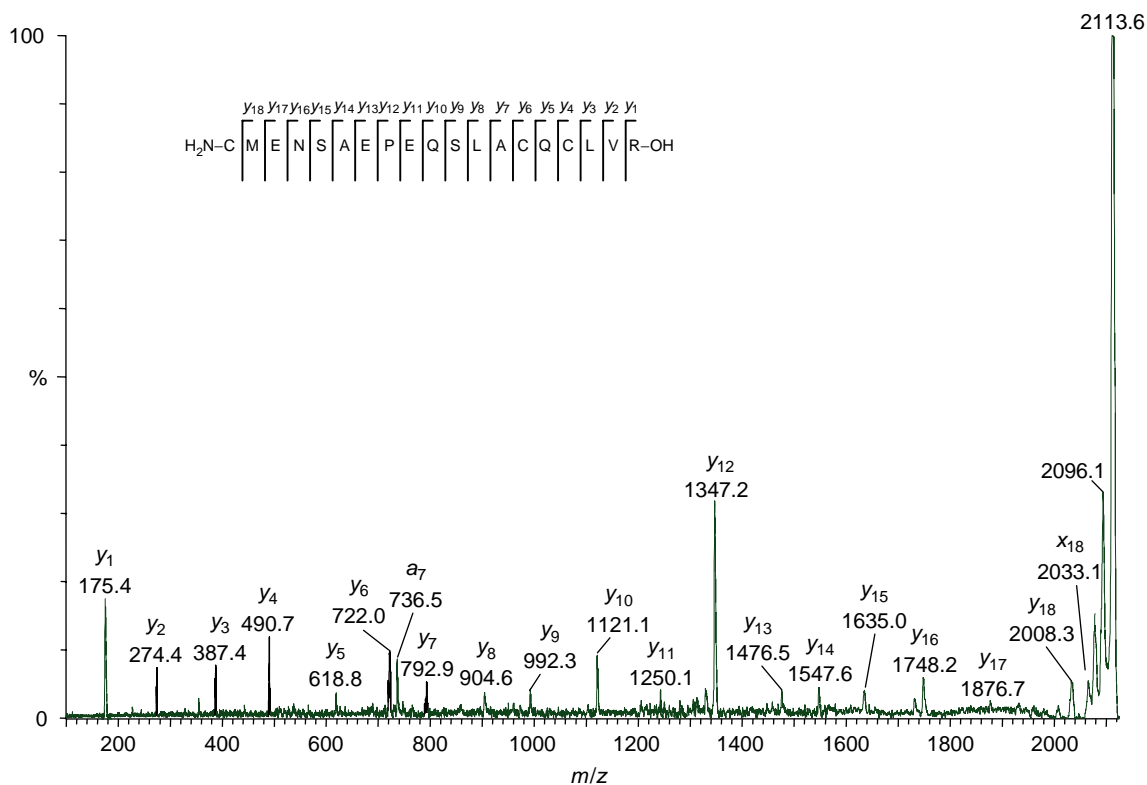


**Figure 5** Scheme of the series of steps to be taken for identifying a polypeptide chain eluted from SDS-PAGE or from a 2D map. Note that, in addition to enzyme digestion and searching of databases via peptide masses (and partial peptide sequences obtained via PSD spectra), it is suggested here that an aliquot of the intact protein should be taken for assessment of native  $M_r$ . (From Hamdan M and Righetti PG, unpublished.)

each peptide list is compared with an analogous list obtained by analyzing a control spot, i.e., the eluate from a small gel segment that does not appear to contain any detectable polypeptide spot: in

this way, all peaks belonging to contaminants are rejected.

MS analysis can be exploited in more subtle ways, for obtaining not simply the precise mass of entire



**Figure 6** Example of sequence determination by PSD of a peptide of  $m/z$  2113.6 obtained by tryptic digestion of  $\beta$ -lactoglobulin B. On the basis of the PSD spectrum and the sequence determined, it was possible to assign this fragment to the amino acid region (106–124) of the protein chain. For a detailed explanation, see text. (Reproduced from Bordini E, Hamdan M, and Righetti PG (1999) Probing the reactivity of S–S bridges to acrylamide in some proteins under high pH conditions by matrix assisted laser-desorption/ionization. *Rapid Communication in Mass Spectrometry* 13: 1818–1827.)

proteins and of their tryptic fragments, but also for generating sequence data. For example, modern MALDI-TOF instruments are equipped with a device, called reflectron, which can select any peptide giving a strong signal and send it back to a reflectron detector placed near the ion source. This parental ion is thus subjected to a so-called post-source decay (PSD), by which it generates a series of ions produced by background collision. By properly searching for a series of  $y_n$ -fragment ions (i.e., those generated from the  $\text{NH}_2$  terminus of the parental peptide) it is possible to reconstruct the amino acid sequence of the said peptide. An example of this procedure is given in Figure 6, which reports the entire sequence of a peptide with  $m/z$  2113.6 obtained by tryptic digestion of  $\beta$ -lactoglobulin B.

## Quantitative Proteomics

2D map analysis, ultimately, should lead to quantitative profiling, i.e., to evaluation of up- and down-regulation in control and pathological tissues under analysis. Although up to present times this has been

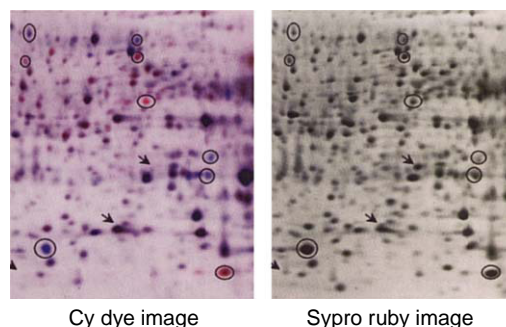
done by profiling the mRNA expression, this does not seem to be quite satisfactory, considering that the correlation between the levels of mRNA and their expressed protein, at any given time in the life cell cycle, is quite poor, barely 0.4 (on a scale from zero to one). Comparison of 2D maps, separately run, by powerful software is one of the oldest and most popular methods in the electrophoretic approach to proteome analysis. For statistical significance three to five tissue samples, for each state, should be collected; the pooled samples will then be processed in preparation for 2D mapping. Four to five replicas of such 2D maps should be run simultaneously, so as to maximize spot reproducibility. From the replicas of the control and pathological states, master maps are produced, which contain all spots found in the individual gels. Spot intensities are normalized in each gel and a statistical test is adopted to evaluate significant differences between the control and treated groups, thus eliminating artifacts due to gel running. The comparison between the two master maps offers a clue about polypeptide chains whose expression is either up- or downregulated or which are newly expressed and unique to one or the other



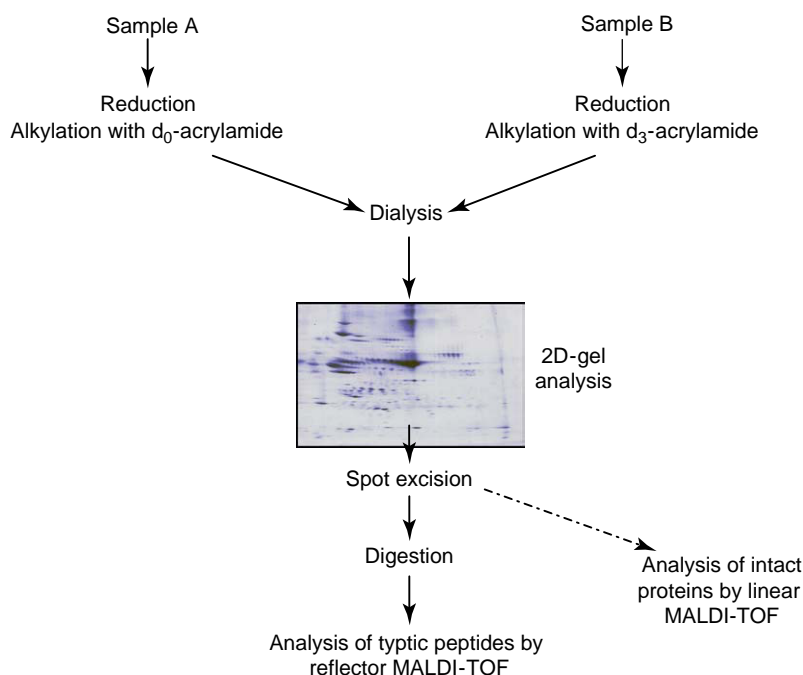
tissues. Once this differential analysis has been performed, all the spots of interest are excised, in-gel digested, and subsequently characterized by MS (e.g., via MALDI-TOF or liquid chromatography coupled to electrospray ionization).

This approach, amply used to the present, gives highly reliable results, but it is terribly time consuming and truly labor intensive. An alternative to this protocol is known under the acronym of DIGE, differential in gel electrophoresis. It is based on differential labeling with *N*-hydroxysuccinimide ester-modified cyanine fluors, the most popular couple being named Cy3 and Cy5. Cy3 is excited at 540 nm and has an emission maximum at 590 nm, while Cy5 is excited at 620 nm and emits at 680 nm. The two samples to be compared are separately labeled with either Cy3 or Cy5, which covalently modify Lys residues in proteins. These dyes have positive charges to replace the loss of charge on the  $\epsilon$ -amino group of Lys, and the masses of the dyes are similar to each other. The reaction is carried out so as to label only a few Lys residues per macromolecule. As long as the extent of the reaction is similar between the samples to be compared, the mass shift will be uniform and the pI should be essentially unaltered. Given the distinguishable spectra of the two fluorophores, the two samples can then be combined and run in a

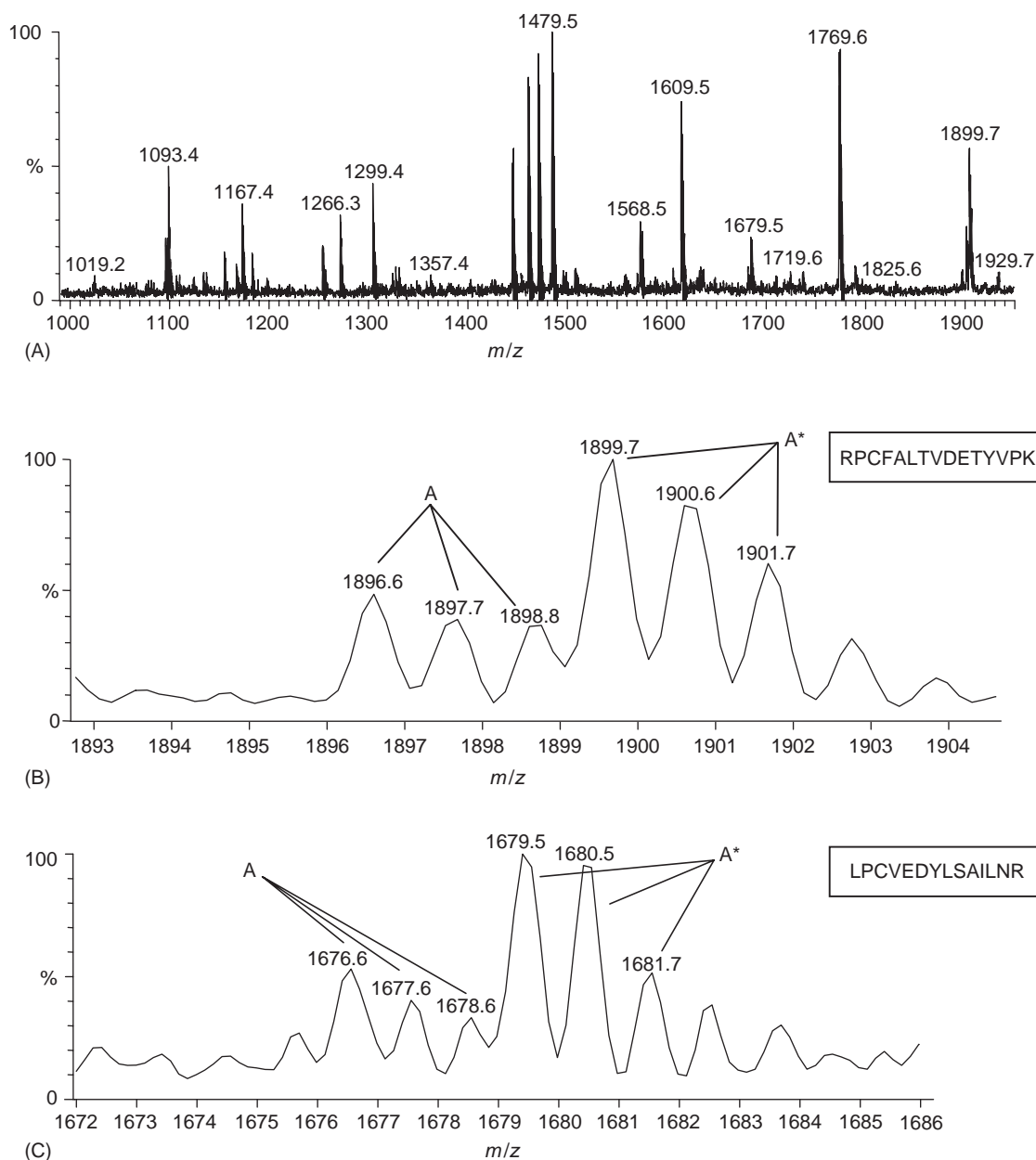
single 2D gel. The differences between the quantities of the individual proteins from each sample can then be determined using specialized 2D image analysis software. Since both samples to be compared are separated in a single gel, this eliminates gel-to-gel variation, resulting in improved spot matching. As a corollary, the number of parallel and replicate gels required for obtaining reliable results is greatly



**Figure 7** Comparison of 2D DIGE imaging and Sypro Ruby poststaining. Left panel: merged Cy dye image of HB4a lysate labeled with Cy3 (red) and HBC3.6 lysate labeled with Cy5 (blue). The same gel was poststained with Sypro Ruby (right panel). (Reproduced from Gharbi S, Gaffney P, Yang A, *et al.* (2002) Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer system. *Molecular Cell Proteomics* 2: 91–98.)



**Figure 8** Scheme for differential labeling of two samples with  $d_0/d_3$  acrylamide (alkylation of Cys residues). The central map refers to rat sera, labeled separately with either  $d_0$ - or  $d_3$ -acrylamide and mixed in a 30:70% ratio. (Reproduced from Gehanne S, Cecconi D, Carboni L, *et al.* (2002) Quantitative analysis of two-dimensional gel-separated proteins using isotopically marked alkylating agents and matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* 16: 1692–1698.)



**Figure 9** (A) Reflectron MALDI mass spectrum of an *in situ* digest of apo-transferrin taken from the 2D map of rat sera displayed in **Figure 4**, which were alkylated with  $d_0$ -acrylamide and  $d_3$ -acrylamide and mixed in a 30/70% ratio. (B) and (C) are two short intervals taken from (A), and are associated with the two indicated peptide sequences. (Reproduced from Gehanne S, Cecconi D, Carboni L, *et al.* (2002) Quantitative analysis of two-dimensional gel-separated proteins using isotopically marked alkylating agents and matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* 16: 1692–1698.)

reduced. Furthermore, fluorescence imparts the ability of detecting proteins over a much broader linear dynamic range of concentrations than visible gel stains. **Figure 7** gives an example of the DIGE technique, as applied to the analysis of breast cancer cells ErbB-2-transformed. Proteins that are present at equal levels in the two cell populations give a uniform violet hue. Proteins present in only one of the two tissues under comparison are either purely red or

blue in color, according to the Cy3/Cy5 label that they carry. Proteins up- and downregulated give intermediate hues that are properly quantified by specialized software.

In another variant for quantitative proteome analysis by 2D PAGE, one could exploit the technique of stable isotope tagging. In this approach, the labeling strategy involves light/heavy forms of the same tagging molecule. An example of such an approach

is the use of  $d_0/d_3$  acrylamide for blocking Cys residues in intact protein molecules. The basic steps in such an approach are depicted in **Figure 8**. Basically, relative quantification of individual proteins in two different samples is achieved by alkylating one sample with  $d_0$ -acrylamide, and the second with its  $d_3$ -counterpart; the two samples are then combined with predetermined ratios, dialyzed, and subjected to 2D gel electrophoresis. Following visualization of the separated proteins, each spot can be excised, digested with trypsin, and examined by MALDI-TOF. The relative quantification of a number of proteins would then be obtained by comparing the relative peak heights within a reflector MALDI spectrum of two adjacent isotopic envelopes that happen to differ by  $m/z = 3$ . The application of this approach to quantitation of various proteins within the 2D map of rat serum shown in **Figure 8** is illustrated below. The map in **Figure 8**, covering the pH 3–10 IPG interval, was obtained by mixing in different proportions two fractions of rat sera, the first (30%) being alkylated with  $d_0$ -acrylamide, and the second (70%) reacted with  $d_3$ -acrylamide. A representative example of a reflector MALDI spectrum that pertains to apotransferrin is given in **Figure 9(A–C)**. The spectrum of the entire digest is given in (A), whereas (B) and (C) display two short intervals of the same spectrum and show two isotopic distributions marked A and A\* in which a difference of 3 Da in the  $m/z$  values of the corresponding peaks is clearly evident. A database search yielded the two indicated peptides, each of which contains a single cysteine. Considering the relative peak heights in both isotopic distributions, a ratio of 34:66 was obtained, which is in good agreement with the labeling ratio 30:70 prior to 2D separation.

The methodology of 2D mapping, for proteome research, has been extensively developed, it is highly reproducible and quite standardized, and surely appears to be one of the best methods still available, at present time, for analyzing very complex tissue extracts, with the ability of displaying up to 12 000 spots in large format slabs, detected by autoradiography, a resolution unmatched by any other separation technique. Yet, on the horizon, seems to be rising the star of multidimensional (as pompously

called, in reality two-dimensional) chromatography, coupling, e.g., a strong cation exchanger with a RP-LC (reversed phase, liquid chromatography). With a proviso, though: most 2D chromatography approaches do not analyze the intact proteome (as 2D electrophoresis does) but a tryptic digest of the solubilized protein. This enormously increases the sample complexity, e.g., up to half a million peptides from a mixture of 10 000 proteins, thus calling for special affinity techniques able to reduce sample complexity.

**See also:** **Electrophoresis:** Overview; Principles; Iso-tachophoresis; Isoelectric Focusing; Polyacrylamide Gels; Blotting Techniques; Clinical Applications.

## Further Reading

- Gygi SP, Rist B, Gerber SA, *et al.* (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* 17: 994–999.
- Hamdan M and Righetti PG (2002) Modern strategies for protein quantification in proteome analysis: advantages and limitations. *Mass Spectrometry Reviews* 21: 287–302.
- James P (ed.) (2001) *Proteome Research: Mass Spectrometry*. Berlin: Springer.
- Kellner R, Lottspeich F, and Meyer HE (eds.) (1999) *Microcharacterization of Proteins*. Weinheim: Wiley-VCH.
- Rabilloud T (ed.) (2000) *Proteome Research: Two-Dimensional Gel Electrophoresis and Identification Methods*. Berlin: Springer.
- Righetti PG (1990) *Immobilized pH Gradients: Theory and Methodology*. Amsterdam: Elsevier.
- Righetti PG, Stoyanov AV, and Zhukov MY (2001) *The Proteome Revisited: Theory and Practice of all Relevant Electrophoretic Steps*. Amsterdam: Elsevier.
- Sebastiano R, Citterio A, Lapadula M, and Righetti PG (2003) A new deuterated alkylating agent for quantitative proteomics. *Rapid Communications in Mass Spectrometry* 17: 2380–2386.
- Wang AH and Hanash S (2003) Multi-dimensional liquid phase based separations in proteomics. *Journal of Chromatography B* 787: 11–18.
- Wilkins MR, Williams KL, Appel RD, and Hochstrasser DF (eds.) (1997) *Proteome Research: New Frontiers in Functional Genomics*. Berlin: Springer.

## Affinity Techniques

**T C Bøg-Hansen**, Panum Institute, Copenhagen, Denmark

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

This chapter deals with detection and exploitation of noncovalent reversible molecular interactions during

is the use of  $d_0/d_3$  acrylamide for blocking Cys residues in intact protein molecules. The basic steps in such an approach are depicted in **Figure 8**. Basically, relative quantification of individual proteins in two different samples is achieved by alkylating one sample with  $d_0$ -acrylamide, and the second with its  $d_3$ -counterpart; the two samples are then combined with predetermined ratios, dialyzed, and subjected to 2D gel electrophoresis. Following visualization of the separated proteins, each spot can be excised, digested with trypsin, and examined by MALDI-TOF. The relative quantification of a number of proteins would then be obtained by comparing the relative peak heights within a reflector MALDI spectrum of two adjacent isotopic envelopes that happen to differ by  $m/z = 3$ . The application of this approach to quantitation of various proteins within the 2D map of rat serum shown in **Figure 8** is illustrated below. The map in **Figure 8**, covering the pH 3–10 IPG interval, was obtained by mixing in different proportions two fractions of rat sera, the first (30%) being alkylated with  $d_0$ -acrylamide, and the second (70%) reacted with  $d_3$ -acrylamide. A representative example of a reflector MALDI spectrum that pertains to apotransferrin is given in **Figure 9(A–C)**. The spectrum of the entire digest is given in (A), whereas (B) and (C) display two short intervals of the same spectrum and show two isotopic distributions marked A and A\* in which a difference of 3 Da in the  $m/z$  values of the corresponding peaks is clearly evident. A database search yielded the two indicated peptides, each of which contains a single cysteine. Considering the relative peak heights in both isotopic distributions, a ratio of 34:66 was obtained, which is in good agreement with the labeling ratio 30:70 prior to 2D separation.

The methodology of 2D mapping, for proteome research, has been extensively developed, it is highly reproducible and quite standardized, and surely appears to be one of the best methods still available, at present time, for analyzing very complex tissue extracts, with the ability of displaying up to 12 000 spots in large format slabs, detected by autoradiography, a resolution unmatched by any other separation technique. Yet, on the horizon, seems to be rising the star of multidimensional (as pompously

called, in reality two-dimensional) chromatography, coupling, e.g., a strong cation exchanger with a RP-LC (reversed phase, liquid chromatography). With a proviso, though: most 2D chromatography approaches do not analyze the intact proteome (as 2D electrophoresis does) but a tryptic digest of the solubilized protein. This enormously increases the sample complexity, e.g., up to half a million peptides from a mixture of 10 000 proteins, thus calling for special affinity techniques able to reduce sample complexity.

**See also:** **Electrophoresis:** Overview; Principles; Iso-tachophoresis; Isoelectric Focusing; Polyacrylamide Gels; Blotting Techniques; Clinical Applications.

## Further Reading

- Gygi SP, Rist B, Gerber SA, *et al.* (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* 17: 994–999.
- Hamdan M and Righetti PG (2002) Modern strategies for protein quantification in proteome analysis: advantages and limitations. *Mass Spectrometry Reviews* 21: 287–302.
- James P (ed.) (2001) *Proteome Research: Mass Spectrometry*. Berlin: Springer.
- Kellner R, Lottspeich F, and Meyer HE (eds.) (1999) *Microcharacterization of Proteins*. Weinheim: Wiley-VCH.
- Rabilloud T (ed.) (2000) *Proteome Research: Two-Dimensional Gel Electrophoresis and Identification Methods*. Berlin: Springer.
- Righetti PG (1990) *Immobilized pH Gradients: Theory and Methodology*. Amsterdam: Elsevier.
- Righetti PG, Stoyanov AV, and Zhukov MY (2001) *The Proteome Revisited: Theory and Practice of all Relevant Electrophoretic Steps*. Amsterdam: Elsevier.
- Sebastiano R, Citterio A, Lapadula M, and Righetti PG (2003) A new deuterated alkylating agent for quantitative proteomics. *Rapid Communications in Mass Spectrometry* 17: 2380–2386.
- Wang AH and Hanash S (2003) Multi-dimensional liquid phase based separations in proteomics. *Journal of Chromatography B* 787: 11–18.
- Wilkins MR, Williams KL, Appel RD, and Hochstrasser DF (eds.) (1997) *Proteome Research: New Frontiers in Functional Genomics*. Berlin: Springer.

## Affinity Techniques

**T C Bøg-Hansen**, Panum Institute, Copenhagen, Denmark

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

This chapter deals with detection and exploitation of noncovalent reversible molecular interactions during



electrophoresis. Its focus is immunoelectrophoresis and affinity electrophoresis (AE) in gels, but reference will also be made to electrophoresis in free solutions, i.e., affinity capillary electrophoresis (ACE). Crossing immunoelectrophoresis (CIE) and AE are a group of procedures in which proteins and other macromolecules are characterized by their electrophoretic migration in a gel and by their binding properties toward antibodies or other ligands.

Characteristically an interaction between an analyte and a ligand takes place during electrophoresis. The noteworthy advantage of AE is that macromolecular characterization is allowed independent of the purity of the sample and does indeed work well with small amounts of crude samples. There are even fewer requirements of the amount of the sample and ligand with capillary electrophoresis.

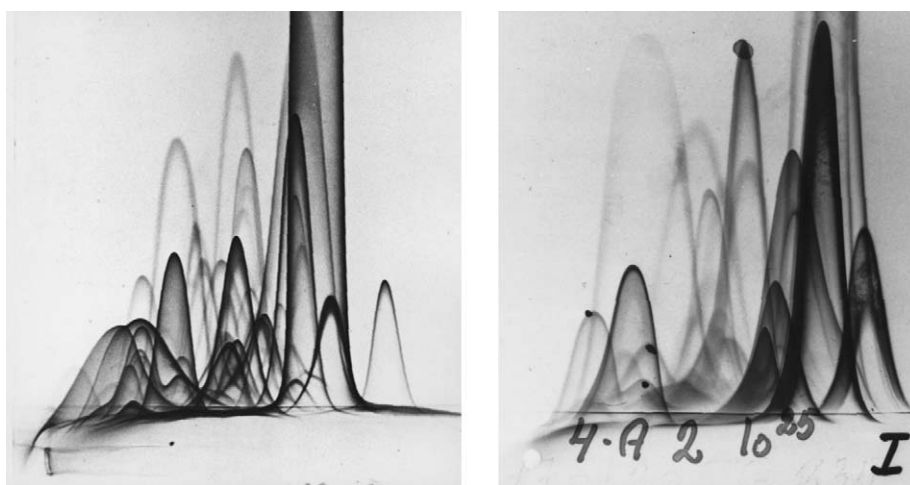
Gel procedures originate from the classical immunoelectrophoresis of P. Grabar, often referred to as IEA, an abbreviation for immunoelectrophoretic analysis, and from the crossing diagrams of K. Nakamura. A distinction may be made between the two-dimensional CIE of Clarke and Freeman (**Figure 1**) and the one-dimensional rocket immunoelectrophoresis of Laurell, often referred to as electroimmunoassay (EIA). Numerous technical modifications of these methods have been devised in order to obtain a certain characterization of one or more proteins. The methods are suitable for establishing an inventory of proteins and for molecular characterization of

proteins and other macromolecules. Electrophoretic methods have been developed successfully as analytical methods; the preparative aspects of electrophoresis are still waiting to be explored and generally exploited, except for a few specific applications. The principal advantages of ACE compared with gel techniques are the even smaller quantity of sample and ligand and the speed of analysis.

## Principles

Interaction during electrophoresis between an analyte and antibodies or other ligands, macromolecular or otherwise, has proven advantageous from an analytical and quantitative viewpoint. Methods have primarily been developed as gel methods for the characterization of proteins and have been expanded to other substances that migrate in an electric field and can react with a ligand, e.g., migration shift assay or charge shift electrophoresis of nucleic acid fragments.

Thus these methods contain an element of biological variation. Immunoelectrophoretic methods contain the greatest element of irreproducibility as they need antibodies as an integral part of the analytical system. Therefore results may vary enormously from one batch of antibodies to the next, and results should be interpreted with caution. The antibodies are generally polyclonal, either monospecific or polyspecific. Monoclonal antibodies may be used, depending on the purpose of the analysis. Especially



**Figure 1** CIE of serum proteins. The patterns indicate the level to which it is possible to number, identify, and characterize proteins in crude mixtures of proteins with commercial antibodies (left) and with self-prepared antibodies (right). Each curved line is a protein precipitated with its specific antibody. From left to right the major proteins are Immunoglobulin A and IgG (partly fused precipitates), complement C3, transferrin,  $\alpha_2$ -macroglobulin, albumin,  $\alpha_1$ -antitrypsin, orosomucoid (AGP), and prealbumin. The patterns also indicate the experimental variation of two-dimensional quantitative immunoelectrophoresis. (Left) Normal human serum (2  $\mu$ l) was electrophoresed for 1 h at  $10 \text{ V cm}^{-1}$  in the first dimension. Into the second dimension gel was added 0.5 ml antibodies against human serum proteins (DAKO-Cytomation). Electrophoresis was performed overnight at  $2 \text{ V cm}^{-1}$ . Agarose gel on a  $7 \times 10 \text{ cm}$  glass plate and stained with Coomassie Brilliant Blue. (Right) Normal mouse serum, amounts and conditions as above except that antibodies against mouse serum produced by us were used. Agarose gel on a  $5 \times 7 \text{ cm}$  glass plate.

in immunoblotting techniques, monoclonal antibodies are applied for specific identification of a protein and after a primary one- or two-dimensional electrophoretic separation.

In classical IEA *ad modum* Grabar, proteins are separated through zone electrophoresis in an agar or agarose gel and subsequently diffuse toward polyclonal antibodies, resulting in complex formation between protein and antibodies, which leads to precipitation in a macromolecular network of protein and antibodies. Macroscopically, precipitation lines are formed individually for each protein with their corresponding antibodies.

CIE (Figure 1) is analogous to classic IEA apart from the second step, where electrophoresis into an antibody-containing gel substitutes for the diffusion step, giving a greatly increased separation. This method allows detection and characterization of individual proteins as well as quantification. Rocket immunoelectrophoresis (EIA) is a one-dimensional electrophoresis of an antigen into an antibody-containing gel; this method is now substituted by enzyme-linked immunosorbent assay (ELISA) and other automated methods for quantification of specific antigens.

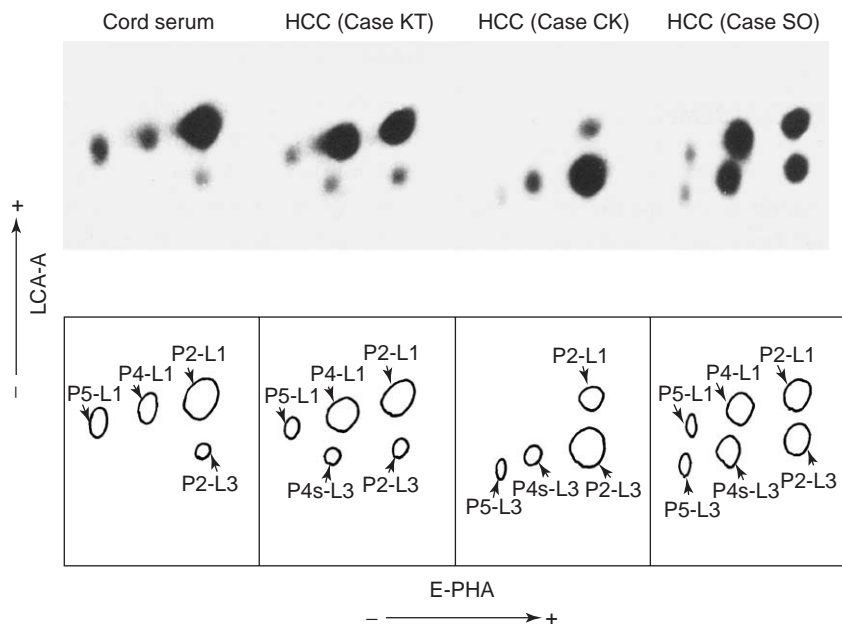
For molecular characterization of macromolecules, gel electrophoresis, especially some of the variants of the basic methods, may be used with great advantage. Without prior purification molecular

characteristics may be investigated and molecular variants like specific glycoforms may be quantified.

A biospecific interaction between two macromolecules or between one macromolecule and a smaller ligand may be exploited in AE (Figure 2). Molecular interaction takes place during electrophoresis and results in a state of dynamic complexation, resulting in the formation of one or more complexes with electrophoretic properties different from that of the uncomplexed macromolecule, leading to a change in the banding pattern. Generally the resulting band pattern is dependent on the character and amount of ligand used. The ligand may be immobilized in the gel matrix or may be present in free solution. A ligand in free solution is often preferred because the electrophoretic system will be less disturbed. Exploiting the variation in ligand concentration, kinetic binding constants may be measured, leading to a thermodynamic description of the interacting molecules.

## Immunoelectrophoresis

The ease with which individual proteins may be assessed and quantified using immunoelectrophoretic methods has resulted in widespread use of these methods when there is a requirement for specific assessment of a protein. This has been seen especially in clinical chemistry, where numerous measurements of



**Figure 2** AE shows the microheterogeneity of a glycoprotein. Glycoform identification and semi-quantitative estimation in two-dimensional lectin AE in agarose gels;  $\alpha$ -fetoprotein (AFP) from serum of patients with hepatocellular carcinoma (HCC) compared with AFP in newborn cord serum. The first dimension was run with erythroagglutinating phytohemagglutinin (E-PHA) and the second dimension with lentil LCA (LCA-A). The separated AFP spots were detected by antibody-affinity blotting. The patterns show increased E-PHA-reactive AFP, increased LCA-A-reactive AFP, and both increased E-PHA-reactive and increased LCA-A-reactive AFP glycoforms. (Reproduced with permission from Taketa K. Okayama.)



specific disease markers are required. However, since the appearance of highly automated methods (like ELISA, radioimmunoassay (RIA), and nephelometry), immunoelectrophoresis is most often used in nonroutine settings, such as in research laboratories, the methods being rather work intensive and skill requiring. Serum and plasma proteins, as well as proteins from other body fluids, have been quantified and characterized using immunoelectrophoresis as disease markers and genetic variants for use in clinical monitoring of patient condition. An early review with many references to the relationship between individual proteins and diseases has been given by Verbruggen.

### Instrumental Requirements

Gel electrophoresis is normally carried out in thin gel layers, for immunoelectrophoresis in agarose gels, usually 1 mm thick or less. Since the molecular interaction is temperature dependent, the system is sensitive to changes in temperature. This condition imposes some restrictions on the type of equipment to be used. Equipment and procedures have been designed to maintain the native conditions of macromolecules until the biospecific interaction has taken place.

Generally a power pack capable of delivering 200–300 V and 200 mA may be used to drive several electrophoresis tanks, preferably with individually regulated outlets. The electrophoresis tank consists principally of a water-cooled surface to support the gel and buffer vessels with electrodes. The gel in which the electrophoretic separation of the analytes takes place is usually cast on a plastic film or glass plate and is placed on the cooling surface in connection with the buffer vessels through several layers of filter paper soaked in electrophoresis buffer or through agarose gel slabs on a support of dialysis film.

The temperature of the instrument is maintained at the desired level by circulating water from a thermostat. For electrophoresis at a field strength greater than  $10 \text{ V cm}^{-1}$ , the temperature of the coolant might be  $10\text{--}14^\circ\text{C}$ , depending on the humidity of the air. For a field strength below  $10 \text{ V cm}^{-1}$ , a temperature of  $16\text{--}20^\circ\text{C}$  is adequate. A variety of suitable instruments are commercially available.

Accessories for easy casting and handling of gels before and after electrophoresis as well as computerized scanning and image processing systems are available commercially.

### Reagent and Substrate Requirements

Agarose is a strong and versatile matrix for gels. Many grades are commercially available: high- and

**Table 1** Some selected examples of applications of AE

Enzymes	Substrates and inhibitors
Glycoproteins	Lectins
Membrane proteins	Detergents
Monoclonal antibodies	Antigens and haptens
Nucleic acid fragments	Nucleic acid-binding proteins
Polyclonal antibodies	Antigens and haptens
Proteins	Drugs
Proteins	Dyes
Receptors	Ligands

low-electroendosmosis (EEO), high- and low-melting point, high- and low-gel strength. It is most important to choose a gel with the proper EEO. For most analyses performed as crossed immunoelectrophoresis with rabbit antibodies at pH  $\sim 8.6$  a gel with EEO in an  $M_r$  interval from 0.13 to 0.18 will be suitable.

Antibodies are usually polyclonal rabbit (or goat) antibodies, either monospecific or polyspecific preparations. Numerous antibodies are available commercially, and several companies offer custom immunization. Monoclonal antibodies are not well suited to precipitation assays and are not generally used. However, antibodies from species other than rabbit and goat may be freely applied.

Immobilized ligands as well as precursor molecules activated for immobilization for binding studies are commercially available and may be applied in AE or may be prepared using the usual methods for affinity chromatography (Table 1).

### Visualization and Detection Techniques

On termination of the immunoelectrophoresis the gel may be visually inspected for the presence and pattern of bands and precipitation lines. This is best done using oblique illumination against a dark background. The plates may be photographed or scanned at this stage, but normally the detection and further evaluation take place after staining. Numerous staining procedures have been described. Digital image processing has been developed for evaluation, comparison, and storage of pattern information, and complete computer systems are commercially available.

### Crossed Radioimmuno-electrophoresis

Crossed radioimmuno-electrophoresis (CRIE) is a technique specially designed for identifications of allergens and for determination of the antigenic specificities of patients' immunoglobulin E (IgE). The proteins toward which the patients' IgE is directed may be determined by incubating the immunoprecipitated proteins in CRIE with patient serum and subsequently visualizing the reaction by adding

radiolabeled anti-IgE and performing autoradiography. This method has been used mainly for detection of patients' IgE response to individual allergenic molecules – 'allergo-prints' – and for identification of allergens in complex allergenic mixtures through performance of 'allergo-grams'. Thus it has been established that the quantitative and qualitative binding of a specific IgE to the various antigens in an allergen extract varies significantly from patient to patient. The summation of the individual patient's specific IgE binding can be correlated to the results of other diagnostic tests, i.e., histamine release and the radio-allergo-sorbent test. The CRIE analysis of a group of patients allergic to the same allergen extract revealed that the binding of IgE differs both in frequency and in strength for various allergens. Through such analyses allergens may be identified, and the results are normally given as allergo-grams, on the basis of which major allergens can be defined.

### Fused Rocket Immunelectrophoresis

This method is a modification of rocket immunelectrophoresis. Small volumes from fractions obtained in separation experiments or samples taken through a biological process are applied in a row of sample wells across the plate in an antibody-free gel, where the samples are allowed to diffuse for up to 1 h. The proteins are then electrophoresed into the antibody-containing gel. The resulting precipitation pattern of the fused proteins from the various fractions will show the fate of each protein and will be an immunological profile of each protein, either representing the elution profile from the fractionation experiment or a profile showing the appearance and disappearance during a biological process.

### Preparative Immunelectrophoresis

An interesting application of CRIE is the preparative aspect. The separation of proteins into clearly positioned and well-spaced precipitates provides an easy means of obtaining purified antigens for immunization of animals. By loading the gel maximally with a protein sample, there will be sufficient material for successful immunization.

## Affinity Electrophoresis

In AE the analyte interacts during electrophoresis with a ligand that may be either in a bound or in a soluble state. Generally the complexation of interacting molecular partners will lead to a change in the net charge-to-size ratio. Thus in free solution the complexed macromolecule will have an electrophoretic mobility different from that of the free (or

uncomplexed) macromolecule. Binding (immobilization) of the ligand to the gel (polyacrylamide or agarose) is used when the ligand is a small molecule or is without a noteworthy electric charge. During electrophoresis the immobilized ligand will slow down the net migration of the interacting macromolecule.

A small soluble ligand with a high negative or positive charge is called an affinophore. Either by using an affinophore directly or by coupling the ligand to an affinophore it is possible to design a soluble system that is well suited for the separation in question, be it analytical or preparative. A prerequisite for using a soluble system is that the mobility of the ligand is different from that of the affinant. As the process of immobilizing a ligand may alter the binding characteristics of the ligand and may make the precise determination of total ligand concentration difficult, most often systems using soluble ligands are required for relevant kinetic determinations.

Crossed affinity immunelectrophoresis (CAIE) is very useful for identification, characterization, and quantification of isoforms and glycoforms. CAIE generally combines analysis of crude protein mixtures with ligand binding and immunological characterization. Thus the binding of a ligand to more than one macromolecule can be studied simultaneously specifically interacting macromolecules in complex mixtures can be separated and identified from those that are nonreacting. Isoforms of the same protein will react slightly differently with the ligand.

With a free unmodified ligand the binding takes place under native conditions, and the interaction is revealed with high sensitivity by changes in position as well as band characteristics. Since electroimmunoprecipitation methods are quantitative, each isoform or glycoform may be quantified directly from a crude molecular mixture.

### Glycoforms

The microheterogeneity of glycoproteins has been strikingly visualized and the amounts of individual glycoforms have been quantified using affinity immunelectrophoresis and AE in polyacrylamide gels. In human serum, some proteins, among others orosomucoid ( $\alpha_1$ -acid glycoprotein (AGP)) and  $\alpha_1$ -antitrypsin, with free concanavalin A (con A, a mannose-binding lectin) in the first dimension have shown up as several glycoforms. The glycoform pattern has revealed that the microheterogeneity of human serum glycoproteins in general is remarkably constant and is probably a general characteristic of serum glycoproteins. These studies show that the carbohydrate side-chain structures in a specific

glycoprotein vary systematically, disfavoring the original 'gadget' concept, in which a presumed sloppiness of glycosylation reactions was postulated to give rise to a statistical microheterogeneity. The microheterogeneity of AGP has been most widely studied in relation to various pathological conditions and pregnancy. In CAIE it results in the separation of this protein into three subfractions interacting differentially with con A.

Microheterogeneous glycoproteins investigated using AE include  $\alpha_1$ -fetoprotein (two glycoforms with con A) and fetal malformations,  $\alpha_1$ -antichymotrypsin and testicular and colorectal cancer,  $\alpha_1$ -fetoprotein and cancer (Figure 2), and transferrin and alcoholism.

The reproducible patterns found for serum glycoproteins has prompted the comparison of glycoform profiles between normal serum and patients' serum. This has been investigated especially with the lectins con A and LCA (lens culinaris agglutinin). The reproducibility of the glycoform pattern may indicate that glycoform variation is a part of the normal function of a glycoprotein and may be involved in the expression of the biological or biochemical function of the protein.

### Binding Constant Calculations

Apparent binding constants of ligand–affinant complexes may be calculated from the change in electrophoretic mobilities seen upon electrophoresis of an analyte in the presence of varying concentrations of ligand. The theory for calculation of binding constants was developed for AE with an immobilized ligand; the simple affinity equation for an immobile ligand is

$$1/R_{mi} = (1/R_{m0})(1 + (c/K_d))$$

where  $R_{mi}$  is the relative migration distance of a component in the presence of immobile (e.g., immobilized) ligand in the gel,  $R_{m0}$  is the relative migration distance of the uncomplexed component (i.e., the distance migrated in a gel without ligand under corresponding electrophoretic conditions),  $c$  is the total concentration of ligand (expressed in molarity of binding sites), and  $K_d$  is the apparent dissociation constant. In this equation, it is presumed that the mobility of the ligand–affinant complex is zero, that  $c$  is much higher than the concentration of the affinant, that  $K_d$  is unaffected by the electrophoresis, that the ligand is distributed randomly in the electrophoresis gel and is 100% accessible, and that the electrophoresis medium is without influence on the electrophoretic mobility of the ligand–affinant complex. By measurement of the migration distances

(relative to an internal marker) at varying ligand concentrations during electrophoresis, the apparent dissociation constant can be found by plotting  $1/R_{mi}$  as a function of  $c$ . Then the intercept on the abscissa is  $-K_d$ .

As immobile ligands may give rise to practical problems, a general affinity equation has been derived that holds for ligands with nonzero mobility:

$$(R_{m0} - R_{mi})^{-1} = (R_{m0} - R_{mc})^{-1}(1 + K_d c^{-1})$$

where  $R_{mc}$  is the migration distance of the complex relative to a nonreacting internal marker (other symbols as above) and where, again,  $c$  is presumed to be much higher than the concentration of the affinant.  $R_{mc}$  may be called the 'eigenmobility' of the complex. Here a plot of  $(R_{m0} - R_{mi})^{-1}$  as a function of  $c^{-1}$  yields a straight line intercepting the abscissa at  $-(K_d)^{-1}$  and the ordinate axis at  $(R_{m0} - R_{mc})^{-1}$ . For most lectins and glycoproteins,  $R_{mc} < R_{m0}$  as the electrophoretic mobility of the lectin at pH > 7 is most often less than that of the glycoprotein. The only other requirement of this equation is that the electrophoretic mobility of the complex should be different from that of the uncomplexed affinant; the accuracy of this method obviously increases the greater this difference is. The original simple equation derives from the general equation by setting  $R_{mc} = 0$ .

These plots allow both  $K_d$  and  $R_{mc}$  to be calculated for a given affinant–ligand system. The latter parameter may in fact give an indication of the relative number of ligands bound per molecule of affinant; the more ligand molecules bound, the smaller  $R_{mc}$  and the smaller  $(R_{m0} - R_{mc})^{-1}$ , provided that  $R_{lectin} < R_{glycoprotein}$ . This is most helpful when elucidating the structural basis for the microheterogeneity of glycoproteins. The practical range of binding constants that can be determined this way is  $10^{10} \text{ mol}^{-1}$ . For the practical calculation of lectin–glycoprotein binding constants using this method, it is obviously pivotal that the values for  $c$ ,  $R_{m0}$ , and  $R_{mi}$  be known with the highest possible accuracy. Using these equations, binding constants have been estimated for binding of soluble glycoproteins to lectins, membrane glycoproteins to lectins, monoclonal antibodies to antigens, and enzymes to substrates and inhibitors.

### Affinity Capillary Electrophoresis

In affinity capillary electrophoresis (ACE) the interaction between the molecular components usually takes place in free solution. The impressive selectivity offered by CE has been further increased by affinity applications, including pharmaceutical compounds,

DNA-binding proteins, lectins, lipids, carbohydrates, and estimation of binding constants and weak interactions.

See also: **Capillary Electrophoresis:** Clinical Applications. **Electrophoresis:** Principles; Clinical Applications.

## Further Reading

- Axelsen NH (ed.) (1983) Handbook of immunoprecipitation-in-gel techniques. *Scandinavian Journal of Immunology* 17 (supplement 10).
- Bøg-Hansen TC (1983) Affinity electrophoresis of glycoprotein. In: Scouten WH (ed.) *Solid Phase Biochemistry: Analytical Synthetic Aspects*, pp. 223–251. New York: Wiley.
- Bøg-Hansen TC (1990) Affinity electrophoresis. The third Hirai memorial prize lecture. *Physical Chemical Biology (Seibutsu-Butsuri-Kagaku (Japan))* 34: 125.
- Bøg-Hansen TC and Hau J (1981) Affinity electrophoresis of glycoproteins. In: Deyl Z (ed.) *Journal of Chromatography Library* 18B: 219.
- Bøg-Hansen TC and Takeo K (eds.) (1989) Paper symposium. Affinity electrophoresis. *Electrophoresis* 10: 811–869.
- Freitag R (1999) Utilization of enzyme-substrate interactions in analytical chemistry. *Journal of Chromatography B* 722: 279–301.
- Galbusera C and Chen DDY (2003) Molecular interaction in capillary electrophoresis. *Current Opinion* 14: 126–130.
- Heegaard NHH and Bøg-Hansen TC (1990) Affinity electrophoresis in agarose gels. Theory and some applications. *Applied and Theoretical Electrophoresis* 1: 249.
- Heegaard NHH (ed.) (2002) Paper symposium. Functional electrophoresis on chips and capillaries. *Electrophoresis* 23: 815–977.
- Heegaard NHH (2003) Applications of affinity interactions in capillary electrophoresis. *Electrophoresis* 24: 3879–3891.
- Heegaard P and Bøg-Hansen TC (1986) Immunoelectrophoretic methods. In: Dunn MJ (ed.) *Analytical Gel Electrophoresis of Proteins*, pp. 262–311. London: Adam Hilger.
- Kim HS, Austin J, and Hage DS (2002) Identification of drug-binding sites on human serum albumin using affinity capillary electrophoresis and chemically modified proteins as buffer additives. *Electrophoresis* 23: 956–963.
- Mackiewicz A and Breborowicz J (eds.) (1992) *Affinity Electrophoresis*. Boca Raton, FL: CRC Press.
- Takeo K (1987) Affinity electrophoresis. In: Crambach A, Dunn MJ, and Radola BJ (eds.) *Advances in Electrophoresis*, vol. 1, pp. 229–279. Weinheim: VCH.

## Blotting Techniques

**B T Kurien and R H Scofield**, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The advent of sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE), with its high resolving power, was a landmark event in basic research methodology. Its full potential was, however, realized only with the advent of another procedure called protein blotting or Western blotting. This technique facilitated the transfer of proteins, separated in a gel matrix, to an adsorbent membrane for subsequent immunodetection; thus becoming a powerful tool for detecting and characterizing a multitude of proteins, especially those proteins that are of low abundance. The term 'Western blotting' was coined in an effort to retain the 'geographic' naming tradition initiated for DNA (Southern blotting) and RNA (Northern blotting) transfer. The blotted proteins formed an exact replica of the gel and became

the starting step for a variety of experiments. The subsequent employment of antibody probes directed against the nitrocellulose bound proteins has revolutionized the field of immunology. Since its inception, protein blotting has been evolving constantly, and now the scientific community is confronted with a plethora of ways and means of transferring proteins.

## Efficiency of Protein Blotting

Successful transfer of proteins from the gel to a solid membrane support depends greatly on the nature of the gel, the molecular mass of the proteins being transferred, and the membrane used. Running the softest gel, in terms of the acrylamide and cross-linker that yields the required resolution, is the best option. Transfer becomes faster and more complete with the use of thinner gels. Use of ultrathin gels, however, may cause handling problems, and a 0.4 mm thickness represents the lower practical limit. Proteins with a high molecular mass blot poorly following SDS PAGE, which results in low levels of detection on immunoblots. However, the efficiency



DNA-binding proteins, lectins, lipids, carbohydrates, and estimation of binding constants and weak interactions.

See also: **Capillary Electrophoresis:** Clinical Applications. **Electrophoresis:** Principles; Clinical Applications.

## Further Reading

- Axelsen NH (ed.) (1983) Handbook of immunoprecipitation-in-gel techniques. *Scandinavian Journal of Immunology* 17 (supplement 10).
- Bøg-Hansen TC (1983) Affinity electrophoresis of glycoprotein. In: Scouten WH (ed.) *Solid Phase Biochemistry: Analytical Synthetic Aspects*, pp. 223–251. New York: Wiley.
- Bøg-Hansen TC (1990) Affinity electrophoresis. The third Hirai memorial prize lecture. *Physical Chemical Biology (Seibutsu-Butsuri-Kagaku (Japan))* 34: 125.
- Bøg-Hansen TC and Hau J (1981) Affinity electrophoresis of glycoproteins. In: Deyl Z (ed.) *Journal of Chromatography Library* 18B: 219.
- Bøg-Hansen TC and Takeo K (eds.) (1989) Paper symposium. Affinity electrophoresis. *Electrophoresis* 10: 811–869.
- Freitag R (1999) Utilization of enzyme-substrate interactions in analytical chemistry. *Journal of Chromatography B* 722: 279–301.
- Galbusera C and Chen DDY (2003) Molecular interaction in capillary electrophoresis. *Current Opinion* 14: 126–130.
- Heegaard NHH and Bøg-Hansen TC (1990) Affinity electrophoresis in agarose gels. Theory and some applications. *Applied and Theoretical Electrophoresis* 1: 249.
- Heegaard NHH (ed.) (2002) Paper symposium. Functional electrophoresis on chips and capillaries. *Electrophoresis* 23: 815–977.
- Heegaard NHH (2003) Applications of affinity interactions in capillary electrophoresis. *Electrophoresis* 24: 3879–3891.
- Heegaard P and Bøg-Hansen TC (1986) Immunoelectrophoretic methods. In: Dunn MJ (ed.) *Analytical Gel Electrophoresis of Proteins*, pp. 262–311. London: Adam Hilger.
- Kim HS, Austin J, and Hage DS (2002) Identification of drug-binding sites on human serum albumin using affinity capillary electrophoresis and chemically modified proteins as buffer additives. *Electrophoresis* 23: 956–963.
- Mackiewicz A and Breborowicz J (eds.) (1992) *Affinity Electrophoresis*. Boca Raton, FL: CRC Press.
- Takeo K (1987) Affinity electrophoresis. In: Crambach A, Dunn MJ, and Radola BJ (eds.) *Advances in Electrophoresis*, vol. 1, pp. 229–279. Weinheim: VCH.

## Blotting Techniques

**B T Kurien and R H Scofield**, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The advent of sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE), with its high resolving power, was a landmark event in basic research methodology. Its full potential was, however, realized only with the advent of another procedure called protein blotting or Western blotting. This technique facilitated the transfer of proteins, separated in a gel matrix, to an adsorbent membrane for subsequent immunodetection; thus becoming a powerful tool for detecting and characterizing a multitude of proteins, especially those proteins that are of low abundance. The term 'Western blotting' was coined in an effort to retain the 'geographic' naming tradition initiated for DNA (Southern blotting) and RNA (Northern blotting) transfer. The blotted proteins formed an exact replica of the gel and became

the starting step for a variety of experiments. The subsequent employment of antibody probes directed against the nitrocellulose bound proteins has revolutionized the field of immunology. Since its inception, protein blotting has been evolving constantly, and now the scientific community is confronted with a plethora of ways and means of transferring proteins.

## Efficiency of Protein Blotting

Successful transfer of proteins from the gel to a solid membrane support depends greatly on the nature of the gel, the molecular mass of the proteins being transferred, and the membrane used. Running the softest gel, in terms of the acrylamide and cross-linker that yields the required resolution, is the best option. Transfer becomes faster and more complete with the use of thinner gels. Use of ultrathin gels, however, may cause handling problems, and a 0.4 mm thickness represents the lower practical limit. Proteins with a high molecular mass blot poorly following SDS PAGE, which results in low levels of detection on immunoblots. However, the efficiency

of transfer of such proteins from a gel matrix has been facilitated using heat, special buffers, and partial proteolytic digestion of the proteins prior to the transfer.

Nitrocellulose, poly(vinylidene difluoride) (PVDF), activated paper, and activated nylon have all been used successfully to bind transferred proteins. The most commonly used membrane support, nitrocellulose, is disadvantageous due to the fact that the proteins are not covalently bound and nitrocellulose is brittle when dry. Small proteins tend to move through nitrocellulose membranes, and only a small fraction of the total amount actually binds. This can be obviated by using membranes with smaller pores. Some investigators have used gelatin-coated nitrocellulose for quantitative retention. In supported nitrocellulose (e.g., Hybond-C Extra), the mechanical strength of the membrane has been shown to be improved by incorporating a polyester support web, thereby making handling easier.

PVDF membranes offer the advantages of high protein binding capacity, physical strength, and chemical stability. One of the advantages of electroblotting proteins onto PVDF membranes is that replicate lanes from a single gel can be used for various purposes such as N-terminal sequencing and proteolysis/peptide separation/internal sequencing, along with Western analysis. In addition, while nitrocellulose membranes cannot be stained with Coomassie Brilliant Blue (CBB), PVDF membranes are amenable to staining with CBB, thus allowing excision of proteins for N-terminal sequencing.

Activated paper (diazo groups) binds proteins covalently but finds its main drawback in that the coupling method is incompatible with many gel electrophoresis systems. Linkage is through primary amines, and therefore systems that use gel buffers without free amino groups must be used with this paper. In addition, the paper is expensive and the reactive groups have a limited half-life once the paper is activated. Nylon has excellent mechanical strength but can only bind a small amount of protein and is not suitable for most applications. Some of

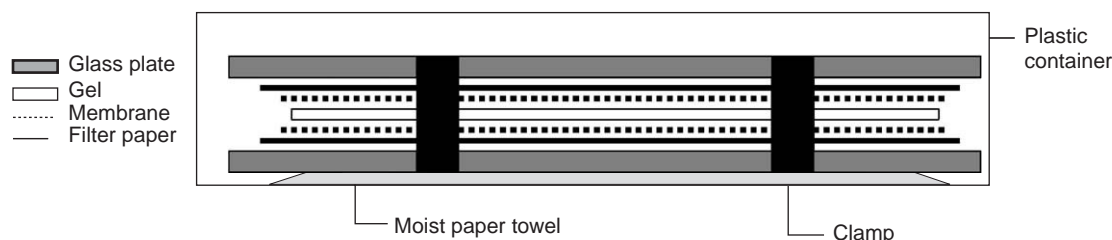
these problems are solved by using activated nylon (positively charged), which however, produces higher nonspecific binding. On account of all these problems, nitrocellulose membranes have remained the best compromise for most situations.

## Methods of Transferring Proteins from Gel to Membrane

Transfer of proteins from native or SDS-PAGE gels to nitrocellulose or PVDF membranes has been achieved in three different ways: (1) simple diffusion; (2) vacuum-assisted solvent flow; and (3) 'Western' blotting or electrophoretic elution.

### Simple Diffusion

Diffusion blotting was originally developed as a means of transferring proteins separated by isoelectric focusing on thin gels to membranes and then later expanded to other gel systems. This procedure requires placing a membrane on the surface of the gel with a stack of dry filter paper on top of the membrane. A glass plate and a weighty object are usually placed on this assembly to facilitate the diffusion process. However, this protocol had not gained widespread acceptance owing to the fact that there was no quantitative transfer of protein. Since the recent demonstration that multiple immunoblots can be generated after nonelectrophoretic bidirectional transfer of a single SDS PAGE gel with multiple antigens, there is considerable interest in diffusion mediated transfer of proteins. This protocol involves incubating the membrane on either side of a gel at 37°C for 2, 4, or 10 h (**Figure 1**). Twelve blots can be obtained using this procedure. Lifts from SDS PAGE gels for immunoblotting using this method are particularly useful in identification of proteins by mass spectrometry. Following diffusion blotting the gel can be stained with CBB. The antigens on the blot are detected by immunostaining, and the immunoblotted target band can be compared with the



**Figure 1** Nonelectrophoretic bidirectional transfer of proteins from SDS PAGE gels to nitrocellulose membranes to obtain up to 12 blots. The PAGE gel is sandwiched between two membranes, filter paper, and glass plates and incubated at 37°C for varying periods to obtain up to 12 blots.



Coomassie stained gel by superimposing the blot and the stained gel, allowing the identification of the band to be excised for tryptic digestion for subsequent mass spectrometric analysis. The greatest advantage of diffusion blotting compared with electroblotting is that several transfers or imprints can be obtained from the same gel and different antisera can be tested on identical imprints.

Quantitative information regarding protein transfer during diffusion blotting was obtained using  $^{14}\text{C}$  labeled proteins. A 3 min diffusion blotting was shown to give a transfer of 10% compared with electroblotting. Diffusion blotting of the same gels carried out multiple times for prolonged periods at 37°C causes the gel to shrink. This can be overcome by using gels cast on plastic supports.

Zymography, or activity gel electrophoresis, has also been studied regarding the utility of diffusion. Zymography involves electrophoresis of enzymes (either nucleases or proteases) through discontinuous polyacrylamide gels (PAGEs) containing an enzyme substrate (either type III gelatin or B-casein). After electrophoresis, the enzyme is renatured by removal of sodium dodecyl sulfate (SDS) and the substrate gets degraded. Staining of the gel with CBB allows the bands of enzyme activity to be detected as clear bands of lyses against a blue background. In this procedure an additional immunoblotting analysis using another gel is often required for examining a particular band that is involved. However, diffusion blotting can be used to circumvent the use of a second gel for this purpose. The activity gel can be blotted onto PVDF for immunostaining and the remaining gel after blotting can be used for routine 'activity staining'. Because the blot and the activity staining are derived from the same gel, the signal localization in the gel and the replica can be easily aligned for comparison.

### Vacuum Blotting

This is a simple blotting method developed as an alternative to diffusion blotting and electroblotting. The suction power of a pump connected to a slab gel dryer system was used to drive the separated polypeptides from the gel to the nitrocellulose membrane. Proteins with both high and low molecular mass could be transferred using membranes with an appropriate pore size.

There is the possibility of the gel drying if the procedure is carried out over longer periods (>45 min), and enough buffer should be used in such cases. In some cases, low concentration PAGEs stuck to the membrane after the transfer. The authors suggest rehydrating these gels, which makes it

swollen, and the nitrocellulose membrane can be easily detached from the gel remnants.

### Electroblotting

Electroblotting is the most commonly used method of transferring proteins from a gel to a membrane. The principal advantages are the speed and the completeness of transfer compared with diffusion or vacuum blotting. Electroelution can be achieved either by (1) complete immersion of a gel-membrane sandwich in a buffer (wet transfer) or (2) placing the gel-membrane sandwich between absorbent paper soaked in transfer buffer (semi-dry transfer).

**Wet transfer** For the wet transfer, the sandwich is placed in a buffer tank with platinum wire electrodes (Figure 2). There are a large number of different apparatuses that will efficiently transfer proteins (or other macromolecules) transversely from gel to membrane. Most of these have vertical stainless steel/platinum electrodes in a large tank (Figure 2).

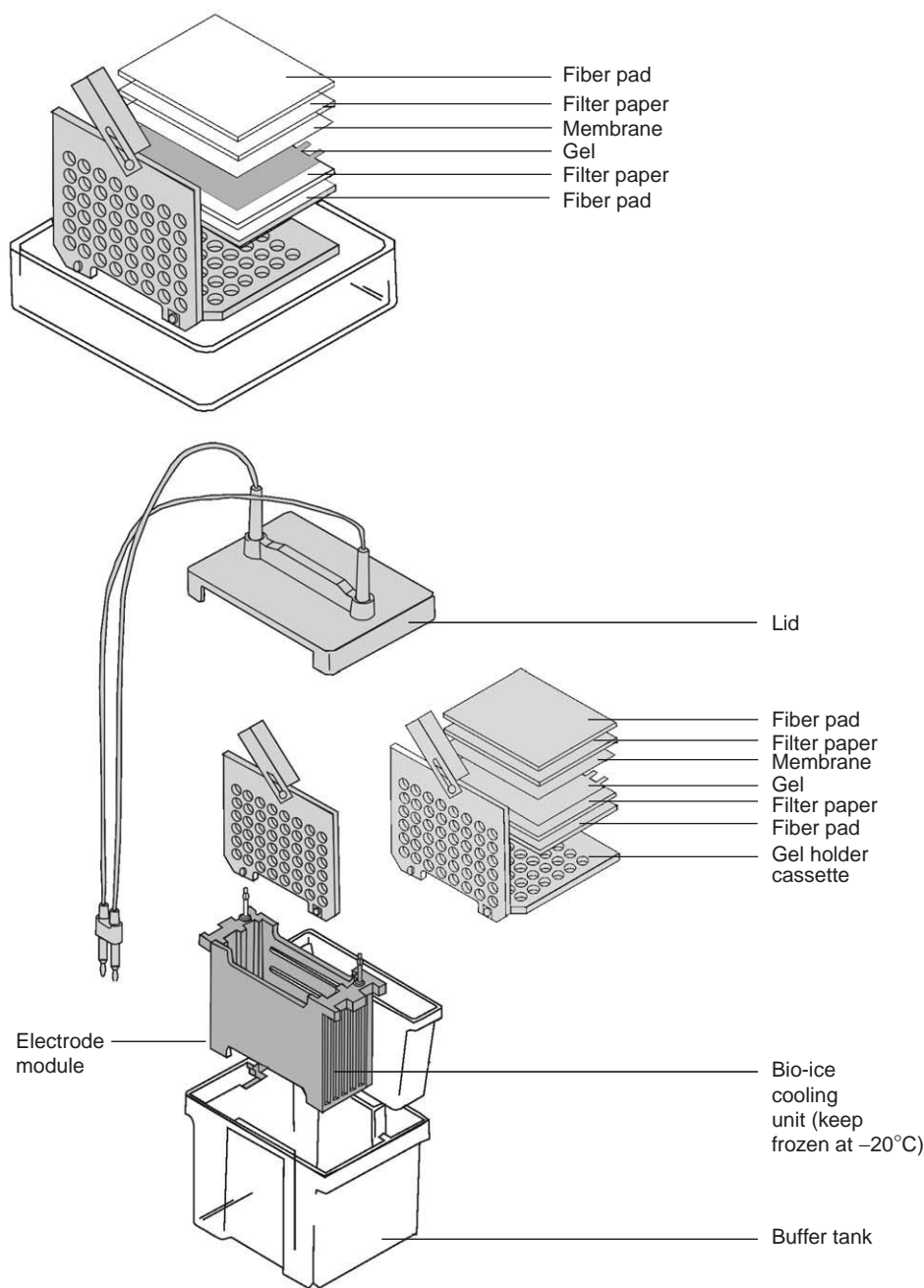
**'Semi-dry' transfer** For the semi-dry transfer, the gel-membrane sandwich is placed between carbon plate electrodes. Semi-dry or horizontal blotting uses two plate electrodes (stainless steel or graphite/carbon) for a uniform electrical field over a short distance, and sandwiches between these of up to six gel/membrane/filter paper assemblies, all well soaked in transfer buffer. The assembly is clamped or otherwise secured on its side and electrophoretic transfer effected in this position, using as transfer buffer only the liquid contained in the gel and filter papers or other pads in the assembly.

There are a number of advantages to this procedure over the conventional upright protocol. Several gels can be blotted simultaneously; electrodes can be cheap carbon blocks; less power is required for transfer (and therefore a simpler power pack).

### Modifications of Electroblotting

Several variations of the basic electroblotting procedure have been developed. Among these are a number of attempts to improve the original protocol, focusing on increasing the amount of protein transferred and retained on the membrane.

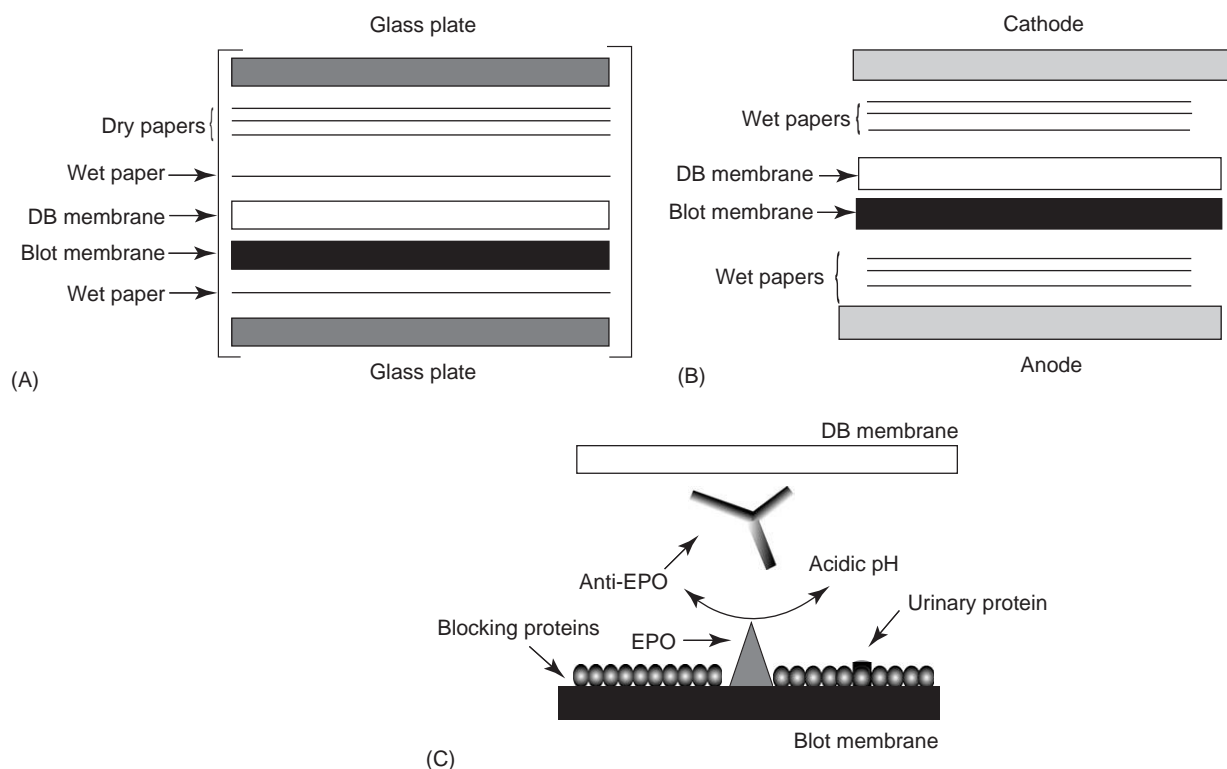
**Double blotting** This procedure was developed to reduce false positives arising as a consequence of nonspecific binding of secondary antibodies. In this protocol the membranes obtained after dot blotting or electroblotting were blocked with nonfat milk as usual, incubated with a primary antibody and washed following conventional procedures. The



**Figure 2** Western blot transfer sandwich assembly for wet transfer (top) and the preparation for transfer using the Mini-Trans-blot apparatus (bottom). The transfer membrane is sandwiched between the gel, filter papers, and support pad, with the transfer membrane facing the anode. The cassette containing the assembled gel-membrane sandwich is inserted into the Trans-blot apparatus containing cold transfer buffer, and the electro-transfer carried out. (Adapted from Bio-Rad Instruction manual.)

principle of double blotting (DB) was to transfer the primary antibody separately from the blot membrane to a new support called the DB membrane (**Figure 3**). Two variants of the DB were developed successfully, (1) pressure DB (**Figure 3A**), which does not use electrophoresis (only simple diffusion) and (2) electro DB, which involves electrophoresis (**Figure 3B**).

**Pressure DB** In this procedure a second PVDF membrane (DB membrane) was cut to the dimensions of the dot-blot membrane and moistened with methanol and  $0.1 \text{ mol l}^{-1}$  glycine/HCl pH 2.5. This membrane was layered on a 3-mm filter paper sheet moistened with the same buffer and laid on three dry filter paper sheets. The blot membrane was placed



**Figure 3** Double blotting (DB). (A) Experimental setup for pressure double blotting, a nonelectrophoretic transfer procedure. The blot membrane was sandwiched between a second PVDF membrane (DB) membrane cut to the dimensions of the blot membrane and filter papers as shown in the assembly. The assembly was clamped between two glass plates, and contact was maintained for 5 min before processing the DB membrane further. (B) Experimental setup for electro DB. The DB membrane was incubated in a 0.7% (v/v) (pH 2.79) acetic acid solution and layered onto the blot membrane between two stacks of filter paper wetted with the same solution. The assembly was placed in a semi-dry electrophoretic blotting instrument with the blot membrane and DB membrane facing the anode and the cathode, respectively, and transferred for 10 min. (C) The principle of DB. The blotting membrane with erythropoietin (EPO) and other urinary proteins was blocked with 5% nonfat milk. The acidic pH of the DB membrane and filter papers induce the dissociation of the primary antibody (anti-EPO), which is transferred to the DB membrane. (Reprinted with permission from Lasne F (2001) Double-blotting: A solution to the problem of non-specific binding of secondary antibodies in immunoblotting procedures. *Journal of Immunological Methods* 253: 125–131; © Elsevier.)

onto the DB membrane and covered with a filter paper sheet wetted with the same buffer. This assembly was placed between two glass plates and secured by clamps as shown in **Figure 3A**, and this contact was maintained for 5 min. Use of pressure results in the capture of the desorbed antibody by the capture DB membrane facing the membrane with the antigen and primary antibody. The decreased pH of the buffer used, a classical desorption method in immunoaffinity chromatography, induces dissociation of the primary antibody from the blotted antigen. Due to the acidic pH, the primary antibody, having a pI much higher than the pH of the acetic acid solution, is positively charged and migrates towards the DB membrane that is in contact with the cathode. Since the acidity does not affect hydrophobicity with PVDF, the antigen and the interfering proteins are retained on the blot membrane.

**Electro DB** This protocol was found to be efficient with electroblotted membranes. The DB membrane was conditioned in a 0.7% (v/v) (pH 2.79) acetic acid solution. It was then layered onto the blot membrane between two stacks of filter paper moistened with the same solution in the semi-dry electrophoretic blotting instrument with the blot membrane and the DB membrane facing the anode and the cathode, respectively (**Figure 3B**). A constant current of  $0.8 \text{ mA cm}^{-2}$  was applied for 10 min. Here again, the low pH used helps in desorption of the primary antibody from the antigen. The application of current thus helped to drive the primary antibody from the primary membrane to the capture membrane. The primary antibody was thus desorbed from its corresponding antigen and transferred onto the DB membrane (**Figure 3C**). The membranes were then disentangled, and the DB membrane was

blocked and further incubated with secondary antibody and developed according to a standard method.

**Slice blotting** This is a simple method of blotting that can map the *in vitro* release of cytokines, growth factors, chemoattractants, chemorepellents, morphogens, enzymes, and other paracrine signals in spatially organized systems, such as brain tissue, under a variety of stimuli and conditions. In this protocol, a brain slice is placed onto a transfer membrane (nitrocellulose) that can trap signal molecules, such as neurotransmitters, proteins, and peptides. Once they are released, molecules diffuse to the nitrocellulose membrane, where they are bound and immobilized. The membrane will accumulate the secreted substances over time, in a pattern reflecting the distribution of active release sites within the slice, and the standard Western blotting techniques can be used to develop this picture.

**Tissue printing (pressure)** This procedure allows the transfer of proteins to nitrocellulose membranes when a fresh cut organ or the cross sections of seeds, stems, tubers, leaves, and fruits are pressed against the membrane surface. This procedure was further modified to detect plant virus coat proteins on whole leaf blots. In this procedure, plant leaves to be blotted were placed on dry nitrocellulose membrane. The leaf and the nitrocellulose membrane were sandwiched on either side with blotting paper or one 0.75 inch thick block of plywood. The sandwich was positioned into a Carver laboratory press, and 10 000 psi of pressure was applied for 2 min. After disassembly, the membrane was air-dried for 20 min prior to standard immunodetection procedures. This procedure was shown to transfer protein uniformly. Some amount of chlorophyll was found to be transferred along with the protein, but this was not found to interfere significantly with the subsequent visualization of results.

**Native electrophoresis and Western blot analysis** Under native PAGE electrophoresis conditions, proteins retain their higher-order structure and often retain their biological activity. SDS, dithiothreitol or  $\beta$ -mercaptoethanol, and heat are omitted from the standard Laemmli SDS protocol. Many factors, including size, shape, and native charge determine the migration of proteins. The resolution of non-denaturing electrophoresis is generally not as high as that of SDS PAGE, but the technique is useful when the native structure or enzymatic activity of a protein must be assayed following electrophoresis.

**Grid immunoblotting** Grid immunoblotting is a procedure that allows the simultaneous testing of up to 20 different samples for specific antibodies against up to 20 different antigens or allergens in a single nitrocellulose membrane sheet. Only  $\sim 150$ – $200\ \mu\text{l}$  of total sample volume is required for the assay, compared with 10 times more than is required by standard enzyme-linked immunosorbent assay (ELISA). This method is especially useful when the sample volume is the limiting factor especially in cases where blood has to be drawn from allergic children to be tested against various allergens.

Grid immunoblotting consists of three basic steps. First, the proteins are immobilized on the carrier membrane. For this, cyanogen bromide activated (to improve protein binding) nitrocellulose membrane is placed onto a multichanneled manifold of the Surf-blot apparatus (Idea Scientific, Minneapolis, MN). The various proteins are then applied to the channels and incubated for 1 h. After washing, the grid is disassembled and the unoccupied binding sites on the membrane are blocked. Second, the blot is incubated with primary antibody (e.g., serum or monoclonal antibody) or other specific probes such as saccharide-specific lectins; and third, the specific binding is sought with a detection system.

**Multiple antigen blot assay** This assay, termed MABA, is an adaptation of the cross-dot system and checkerboard immunoblotting. The antigens (up to 28 different antigens), such as synthetic peptides, are immobilized on a nitrocellulose membrane in 28 parallel lanes delineated by a template (Multiscreen apparatus). Strips are then excised perpendicular to the antigen lanes, individually immersed in immune sera, and developed according to standard procedures. This method allows the screening of the reactivity of a serum against 28 different antigens and was found to be twice as sensitive as conventional ELISA. In a single assay, MABA can be used to analyze qualitatively as many as 26 different sera against 28 individual antigens, the equivalent of eight ELISA plates (728 wells).

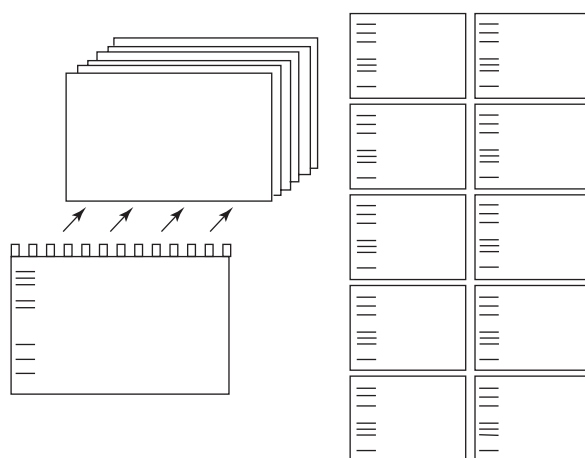
**Multiple tissue Western (MTW) blot** Human MTW blots provide a new immunological tool for the investigation of tissue-specific protein expression. These are premade immunoblots prepared using proteins isolated from adult human tissue. The proteins are isolated from whole tissue under conditions engineered to minimize proteolysis and to ensure maximal representation of tissue-specific proteins. SDS solubilized proteins are fractionated by SDS PAGE and electroblotted onto PVDF membranes to generate blots ready for incubation with

specific antibodies. Several specific antigens can be detected in separate cycles of antibody incubations using the same MTW blot. This approach was possible using a stripping procedure that allowed selective removal of both primary and secondary antibodies in a single incubation. The use of multiple reprobing makes this a very useful tool for studying human tissue-specific proteins.

**Dot-immunobinding assay (Dot-Iba)** This procedure entails placing the antigen on the surface of nitrocellulose cellulose membrane (NCM) disks (1 cm diameter) and incubating at 4°C for 12 h. The disks were developed using a standard procedure, and a positive reaction was indicated by the development of an insoluble purple color in the central portion of the NCM disks. In contrast to conventional bacteriological methods, Dot-Iba required only 6 h to be performed in the laboratory. This has been used to measure mycobacterial antigens in cerebrospinal fluid (CSF) specimens for laboratory diagnosis of tuberculous meningitis with a sensitivity level of 100 ng of antigen per milliliter. NCM disks containing the 14 kDa (100 ng ml<sup>-1</sup>) mycobacterial antigen were used as a positive control.

**High-efficiency blotting of high-molecular mass proteins** Prolonged electrotransfer (16–21 h) at a high current density coupled with inclusion of SDS in the transfer buffer, to enhance protein elution, has been used to transfer proteins efficiently with a high molecular mass. Special blotting buffers (40 mol l<sup>-1</sup> Tris, 20 mol l<sup>-1</sup> sodium acetate, 2 mol l<sup>-1</sup> ethylene diamine tetraacetic acid, pH 7.4, 20% (v/v) methanol, 0.05% (v/v) SDS), novel gels (50:1 ratio of acrylamide: bisacrylamide), composite agarose-PAGEs containing SDS and urea, a heated transfer buffer (without methanol), or partial proteolytic digestion of high-molecular-weight proteins prior to transfer have also been used for this same purpose.

**Western blot with kits** The KODAK BIOMAX multiblot kit for proteins applies the ease and convenience of Western blotting to a multiplex format. The proteins are transferred, after electrophoresis, with a conventional transfer apparatus from the gel to the BIOMAX multiblot kit's proprietary membrane stack using a specially formulated transfer buffer (Figure 4). Because the membrane has high affinity protein binding and a low capacity, proteins can be transferred from a single gel onto the BIOMAX multiblot kit's membrane stack, producing up to 10 blots. Experiments have demonstrated that at least a 50-fold difference



**Figure 4** The KODAK BIOMAX multiblot kit for transfer of proteins. The membrane has a high affinity protein binding and low capacity. The proteins are transferred from a single gel onto the BIOMAX multiblot kit's proprietary membrane stack, producing up to 10 blots.

(10–500 ng of protein could be detected) in amount of protein can be measured on all 10 blots.

**Blotting of stained proteins from PAGE gels to transparencies** CBB stained proteins were transferred nonelectrophoretically from PAGE gels to transparencies similar to those used in plain or regular photocopiers. Proteins are first separated by regular SDS PAGE, stained with CBB, and destained. The destained gel is pressed against the transparency for 2 h at room temperature or for 30 min at 40°C. The blue bands corresponding to the protein-dye complexes are transferred to the transparency. Since the transparencies are poreless and hence not conductive in an electric field, it is not possible to blot dye-protein complexes electrophoretically from PAGE gels to transparencies. Protein patterns present in several types of gels – SDS PAGE, nondenaturing PAGE, isoelectric focusing PAGE, and SDS-agarose – all were found to be transferred successfully following staining with the dye. Proteins visualized with other organic dyes such as Fast Green FCF2, Uniblue A3, and Procion Blue MX-R4 were also found to transfer, but proteins stained with Stains-all 5 or silver did not. This transfer method provides a simple, economical way of preserving data and a convenient method of displaying data using an overhead projector. The protein pattern obtained is sharper using this method of transfer, presumably due to reduced diffusion during the transfer process, and there is no change in the antigenicity of the proteins.



**Treatment of nitrocellulose blots to improve detection** A large number of proteins bind strongly to nitrocellulose under different experimental conditions. However, milk, Triton X-100, Nonidet P-40, and Tween-20 have all been shown to remove bound proteins. Milk, however, has been generally used to block unoccupied sites on the nitrocellulose membrane and thus prevent nonspecific binding. The sensitivity of nondenaturing blots can be increased by soaking the membrane in acidic buffer after transfer. In addition milk-stripping of antibody from membrane can be completely eliminated by exposure to acidic buffer.

### Antigen detection

**Antigen detection by protein stains** Proteins have been detected after immunoblotting onto membrane supports directly by use of organic dyes (Ponceau red, amido black, fast green, CBB), using fluorescent labels (fluorescamine, coumarin), various silver staining methods, and colloidal particles such as gold, silver, copper, iron, or India ink. Metal chelate stains such as the Ferrozine/ferrous (iron(II)) complex and the ferrocyanide/ferric (iron(III)) complex have been used successfully to stain one-dimensional and two-dimensional electroblots. Ferrozine and Ferene S, when combined with ferrous (iron(II)) ions, effectively stain proteins with a sensitivity comparable with amido black stain, producing purple and blue stains, respectively.

**Immunodetection of antigens** Radioactive and enzyme-linked reagents have been used for immunodetection of antigens after blotting. Radioactive methods use iodinated ( $^{125}\text{I}$ ) staphylococcal protein A or streptococcal protein G to bind to the primary antibody that is bound to the desired antigen on the membrane blot. The radiolabeled antigen can then be detected using autoradiography. Enzyme-linked reagents include secondary antibodies coupled to either horseradish peroxidase or alkaline phosphatase, which bind to antigen-antibody complexes on the membrane. Using appropriate substrates (those that produce color or light), depending upon the degree of sensitivity required, the desired protein bands on the blot can be visualized. Protein A bound gold particles that bind to the primary antibody to produce a red coloration have also been used. The biotin-streptavidin system has also been used for detecting immune complexes.

See also: **Electrophoresis**: Principles; Polyacrylamide Gels; Proteins. **Proteins**: Overview.

### Further Reading

- Alwine JC, Kemp DJ, and Stark GR (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proceedings of the National Academy of Sciences of the United States of America* 74: 5350–5354.
- Burnette WN (1981) 'Western Blotting': Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Analytical Biochemistry* 112: 195–203.
- Egger D and Bienz K (1994) Protein (Western) blotting. *Molecular Biotechnology* 1: 289–305.
- Gershoni JM and Palade GE (1983) Protein blotting: Principles and applications. *Analytical Biochemistry* 131: 1–15.
- Harlow E and Lane D (eds.) (1988) Immunoblotting. In: *Antibodies. A Laboratory Manual*, pp. 471–510. Cold Spring Harbor: Cold Spring Harbor Press.
- Harper DR, Kit ML, and Kangro HO (1990) Protein blotting: Ten years on. *Journal of Virological Methods* 30: 25–39.
- Kurien BT and Scofield RH (1997) Multiple immunoblots after non-electrophoretic bidirectional transfer of a single SDS-PAGE gel with multiple antigens. *Journal of Immunological Methods* 205: 91–94.
- Kurien BT and Scofield RH (2002) Heat mediated, ultra-rapid electrophoretic transfer of high and low molecular weight proteins to nitrocellulose membranes. *Journal of Immunological Methods* 266: 127–133.
- Kurien BT and Scofield RH (2003) Protein blotting: A review. *Journal of Immunological Methods* 274: 1–15.
- Laemmli UK (1970) Cleavage of structural proteins during assembly of the head of bacteriophage T4. *Nature* 227: 680–685.
- Patton WF, Lam L, Su Q, *et al.* (1994) Metal chelates as reversible stains for detection of electroblotted proteins: Application to protein microsequencing and immunoblotting. *Analytical Biochemistry* 220: 324–335.
- Peferoen M, Huybrechts R, and De Loof A (1982) Vacuum-blotting: A new simple and efficient transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to nitrocellulose. *FEBS Letters* 145: 369–372.
- Reinhart MP and Malamud D (1982) Protein transfer from isoelectric focusing gels: The native blot. *Analytical Biochemistry* 123: 229–235.
- Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* 98: 503–517.
- Towbin H, Staehelin T, and Gordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and applications. *Proceedings of the National Academy of Sciences of the United States of America* 76: 4350–4354.



# Carbohydrates

H Scherz, C W Huck, and G K Bonn, University of Innsbruck, Innsbruck, Austria

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Carbohydrates are a large family of vicinal polyhydroxy compounds and can be classified into mono-, oligo-, and polysaccharides.

Monosaccharides contain carbonyl groups and they are – like their deriving oligo- and polysaccharides – without charge and therefore electrically neutral. Derivatives of monosaccharides are alduronic acids, aldonic acids, and sugar phosphates having carboxyl or phosphate groups. These compounds as well as their oligomers and polymers possess electric charges and migrate in the electric field. As electrically neutral mono-, oligo-, and polysaccharides do not migrate in the electric field, they have to be transferred to charged complex compounds by reaction with inorganic polyhydroxy anions like tetrahydroxyborate, arsenite, germanate, molybdate, and vanadate. These charged complex compounds migrate in the electric field with different mobilities.

The anionic complex between carbohydrates and borate is the most studied and has been known for a long time. The borate anion exists in aqueous solution as tetrahydroxyborate with tetrahedral symmetry, and reacts with vicinal hydroxyl groups to the following charged complex compounds shown in **Figure 1**. The formation of such a charged complex can take place only when the distance between the hydroxyl groups in the organic molecules is similar to that in the borate anion, which is 2.36–2.39 Å. The strength of such borate polyol complexes

depends on the number of such –O–O– units as well as on their conformation in the resulting molecules. They have a positive correlation with their mobilities in the electric field. The same situation is valid using phenyl boric acid and diphenyl boric acid. Similarly, oligohydroxyanion–carbohydrate complexes are formed with germanate, hydrogenarsenite acid, molybdate, and vanadate. The strength of the resulting polyol complex strongly depends on the structure of the carbohydrate part.

For the practical purposes two modes of electrophoretic operations are used:

1. zone electrophoresis on a flat inert support;
2. zone electrophoresis in fused silica capillaries (length: 20–100 cm, 25–75 µm i.d.) with potentials of 10–30 kV (capillary zone electrophoresis).

## Electrophoresis of Carbohydrates on a Flat Inert Support

### Chromatographic Paper

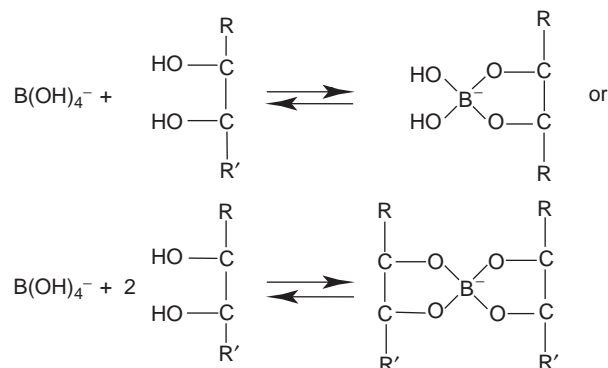
Cellulose is the oldest support for these electrophoretic separations. The paper is soaked with the electrolyte before spotting the sample on the wet surface. A special disadvantage is the need to use high voltage in order to obtain satisfactory separation (30–100 V cm<sup>−1</sup>) with sharp zones and reduced separation time. The use of expensive equipment is necessary as intense cooling is required. A further disadvantage of paper electrophoresis is the limited possibility to visualize the separated sugar zones.

The following reagents can be used for the visualization of zones containing low molecular mass-carbohydrates:

- Alkaline silver nitrate, sodium metaperiodate-benzidine (or *o*-dianisidine) for all vicinal polyhydroxy compounds like sugars, sugar alcohols, aldonic acids, etc.
- Aniline phthalate, aniline trichloroacetate, *p*-anisidine hydrochloride, *p*-anisidine trichloroacetate, *m*-phenylenediamine, etc., for reducing sugars and uronic acids.
- Acetylacetone-*p*-dimethylaminobenzaldehyde for amino sugars (Morgan Elson reaction).

### Glass Fiber Paper

This support overcomes the limited possibilities of visualization of the separated sugar zones. It enables the use of sulfuric acid containing carbohydrate



**Figure 1** Reaction of tetrahydroxyborate with vicinal hydroxyl groups of the corresponding carbohydrate.

reagents. The mobilities of the zones are higher than when using cellulose matrices of the chromatographic paper due to their reduced adsorption. A disadvantage is the high electroendosmosis due to an enhanced charge of the glass fiber surface. This can be avoided by silanization of the surface with dichlorodimethylsilane, which leads to a decrease in mechanical stability. Separations of mono-, oligo-, and polysaccharides can be successfully carried out using this material.

### Silanized Silica Gel

This material was developed to overcome the disadvantage of the low stability of the silanized glass fiber paper. After the surface of the silanized silica gel (for thin layer chromatography) is covered with a thin film of 1-octanol, it is slurried with the electrolyte containing polyvinylpyrrolidone and is then spread on thin layer plates (20 × 20 cm) giving a weak quasigel layer. After storing for 3 h in a humid atmosphere, the sample can be spotted on these plates and electrophoresis can be carried out using a total potential of 200–400 V. This support can be used for the separation of mono- and oligosaccharides, sugar alcohols, uronic acids, sugar phosphates, and polysaccharides (Figure 2).

Silanized silica gel is stable against aggressive reagents and in addition to those reported in the section 'Chromatographic paper', the following

reagents can be preferably applied for visualization:

- Alkaline potassium permanganate for all vicinal polyhydroxy compounds (sugar, sugar alcohols, aldonic acids, alduronic acids, etc.).
- Aniline–diphenylamine phosphoric acid, orcinol–Fe(III)–sulfuric acid for mono- and oligosaccharides as well as uronic acids.
- 1,3-Dihydroxynaphthalene (naphthoresorcinol) – sulfuric acid with or without the addition of 1-octanol for mono-, oligo-, and polysaccharides.
- 1,3-Dihydroxynaphthalene–hydrochloric acid for uronic acids.

### Cellulose Acetate Membranes

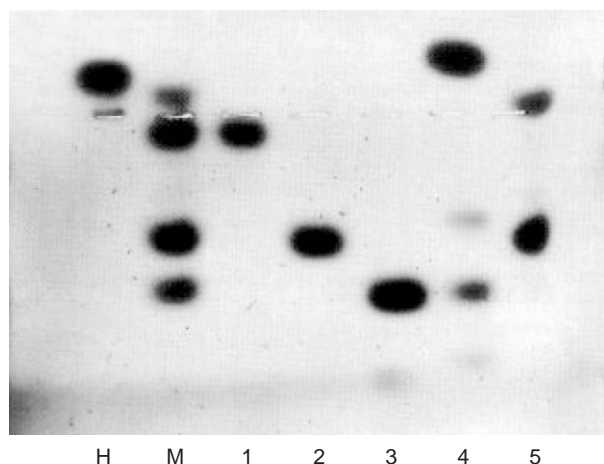
This support, which is used in the field of proteins, was introduced for the separation of glycosaminoglycans (e.g., heparin, hyaluronic acid, etc.), which is important in the field of medical diagnosis (e.g., for the detection of mucopolysaccharidosis). This group of compounds contains negative and positive charged units (e.g., uronic acids, hexosamines) in different proportions. The visualization of the separated glycosaminoglycans is carried out by staining with dyes, which have strong affinities to such polysaccharides. The most frequently used ones are Alcian blue (a copper-containing phthalocyanine) and Toluidine blue. Furthermore, the separation of acid and neutral polysaccharides containing thickening agents on this support has been described in the field of food analysis. The zones of acidic polysaccharides can be visualized with basic dyes like Methylene blue, Amido black 10B, or Toluidine blue O. The neutral polysaccharides can be visualized by the reaction with Fuchsin–sulfurous acid.

### Agarose Gel

As an alternative to cellulose acetate strips, this support can be used for the separation of glycosaminoglycans. It offers the possibility to separate these compounds into two dimensions. For visualization, the glycosaminoglycans are at first fixed in the gels by Cetavlon (cetyltrimethylammonium bromide) and then stained with Toluidine blue.

### Polyacrylamide Gel

This support, which is of widespread use in the field of protein and DNA electrophoresis, is applied for the separation of oligosaccharides, labeled by fluorophores like 2-aminoacridone or 8-aminonaphthalene-1,3,6-trisulfonate, formed by enzymatic hydrolysis of glycoproteins. For this purpose small-pore gel material is used either with uniform pores or as



**Figure 2** Thin-layer electrophoresis of uronic acids on silanized silica gel, pretreated with octanol-1 in 0.07 mol l<sup>-1</sup> barium acetate solution. Electrophoresis: 400 V, 150 min, 16°C. Lanes: (1) galacturonic acid, (2) mannuronic acid, (3) glucuronic acid, (4) hydrolysate of gum arabic (glucuronic acid), (5) hydrolysate of sodium alginate (mannuronic and guluronic acids), (M) mixture of (1–3) and guluronic acid, (H) omega-hydroxymethylfurfural. Sample loading: 10 µg of each individual compound. Visualization with naphthoresorcinol–sulfuric acid. Origin indicated.

gradient gel. The visualization of the separated zones can be done by illumination under an ultraviolet lamp.

**Application** The electrophoresis on flat inert supports is used for:

- Separation of glycosaminoglycans like hyaluronic acid, heparin, chondroitin sulfate, keratan sulfate, etc., in biological materials (like urine). This method is important in the field of medical diagnosis (e.g., identification of mucopolysaccharidosis). The supports most often used for this purpose are cellulose acetate and agarose gel.
- Separation of acid and neutral polysaccharides, e.g., those of food hydrocolloids (thickening agents, e.g., guar, gum arabic, carrageen, etc.) and pectins. The best supports are silanized glass fiber paper or silanized silica gel-1-octanol.
- Mono- and oligosaccharides can be easily separated on silanized glass fiber paper and on silanized silica gel-1-octanol plates. Both systems offer, as a rapid screening method, an alternative to thin-layer chromatography especially for a clear identification of uronic acids in a mixture with neutral monosaccharides, as it occurs in the case of hydrolysates of acid

**Table 1** Some examples for the electrophoretic separation of carbohydrates on flat inert supports

<i>Carbohydrates</i>	<i>Support</i>	<i>Complexing agent; electrolyte; applied voltage</i>	<i>Detection</i>
Monosaccharides, disaccharides, methyl sugars, glycosides, cyclitols, sugar alcohols	Chromatographic paper (Whatman 3)	0.2 mol l <sup>-1</sup> Sodium tetraborate, pH 10.0; 0.05 mol l <sup>-1</sup> sodium tetraborate, pH 9.5; 500 V, 1300–1400 V	Aniline–hydrogenphthalate
Monosaccharides, disaccharides, sugar alcohols	Chromatographic paper (Whatman 4)	0.1 mol l <sup>-1</sup> Sodium arsenite, pH 9.6	Aniline–phosphates, H <sub>2</sub> O <sub>2</sub> , CrO <sub>3</sub> –H <sub>2</sub> SO <sub>4</sub>
Monosaccharides, sugar alcohols, amino sugars	Chromatographic paper (Whatman 3, Schleicher-Schüll 2043b)	Sodium molybdate (25 g per 1000 ml), pH 5.0, 1.5% sodium metavanadate, pH 8.7; 3000 V	Silver nitrate–ethanol, sodium hydroxide, <i>p</i> -anisidine–HCl; acetone–KMnO <sub>4</sub>
Mono- and disaccharides, alduronic acids, methylated sugars, polysaccharides (thickening agents)	Silanized glass fiber paper	0.05 mol l <sup>-1</sup> Sodium tetraborate, pH 9.0, 0.1 mol l <sup>-1</sup> barium acetate, 0.1 mol l <sup>-1</sup> zinc acetate, 0.1 mol l <sup>-1</sup> calcium acetate	1,3-Dihydroxynaphthalene (naphthoresorcinol)–H <sub>2</sub> SO <sub>4</sub>
Mono- and oligosaccharides, sugar alcohols, sugar acids, sugar phosphates, polysaccharides (thickening agents)	Silanized silica gel (for thin-layer chromatography)–1-octanol	0.3 mol l <sup>-1</sup> Borate buffer, pH 10.0 containing 0.02 mol l <sup>-1</sup> EDTA and 0.44% polyvinylpyrrolidone (K-90: MM: 350 000)	1,3-Dihydroxynaphthalene–H <sub>2</sub> SO <sub>4</sub> –1-octanol
Mono- and oligosaccharides, esp. those obtained from glycoproteins) labeled with fluorophores like 2-aminoacridone or 8-amino-naphthalene-1,3,6-trisulfonic acid); lipopolysaccharides	Small porous polyacrylamide gel (with or without pore gradient)	0.1 mol l <sup>-1</sup> Tris–borate/boric acid buffer, pH 8.3, 0.01 mol l <sup>-1</sup> Tris–glycine buffer (according to Laemmli), 0.1 mol l <sup>-1</sup> Tris–HCl buffer, pH 6.8 containing 2% SDS for lipopolysaccharides	UV-light 265–365 nm (UV- or fluorescence) for labeled oligosaccharides; periodic acid–silver nitrate for lipopolysaccharides
Glycosaminoglycans (hyaluronic acid, heparin, chondroitin sulfate, etc.)	Cellulose acetate membranes	0.2 mol l <sup>-1</sup> Barium acetate, pH 5.0 or pH 8.0, 0.05 mol l <sup>-1</sup> lithium chloride, 0.01 mol l <sup>-1</sup> HCl (pH 2.0). Two dimensional: 1st direction: 0.1 mol l <sup>-1</sup> pyridine, 0.47 mol l <sup>-1</sup> formic acid, pH 9.0; 2nd direction: 0.1 mol l <sup>-1</sup> barium acetate, pH 8.0	Alcian blue
Glycosaminoglycans (heparin, hyaluronic acid, chondroitin sulfate, etc.)	Agarose (0.9% agarose)	0.06 mol l <sup>-1</sup> Barbitol (5,5'-diethylbarbituric acid) buffer, pH 8.6, 0.05 mol l <sup>-1</sup> diaminopropane–acetate (pH 8.5)	Cetavlon-Alcian blue, Toluidine blue

polysaccharides such as gum arabic (Figure 2). The advantage of the silica gel-1-octanol layer is that, contrary to thin-layer chromatography, no additional preparations of the sample before the electrophoresis procedure is necessary, as it was found, for example, for the separation of lactulose and lactose in processed milk. A drop of the sample is put on the layer and the electrophoresis carried out immediately to obtain clear distinct zones for visualization.

Table 1 lists some examples for the electrophoretic separation of carbohydrates on the supports described above. Table 2 lists the electrophoretic mobilities of some common mono-, di-, oligosaccharides, sugar alcohols, and aldonic acids. Table 3 summarizes the electrophoretic mobilities of sugar acids, sugar phosphates, and amino sugars. In Table 4, the chromatographic conditions used in Tables 2 and 3 can be deduced by the use of the corresponding reference letters (A–G).

## Capillary Electrophoresis of Carbohydrates

Capillary electrophoresis (CE) is a very effective method for the separation and determination of carbohydrates. It offers higher resolution of separation, requires lesser analysis time, and can be automated. CE can be used to separate mono-, oligo-, and polysaccharides, their acid and basic derivatives as well as glyco-conjugates like glycoproteins, proteoglycans, and glycolipids. A characteristic advantage of CE is the requirement of extremely small amounts of sample as the detection limits range from pico- to attomoles depending on the derivatization conditions used. The mode of separation is almost the same as described in the first two sections, but instead of flat inert supports separations are carried out in fused silica capillaries of lengths of 20–100 cm and having an internal diameter of 25–100  $\mu\text{m}$ . One of the fundamental processes that drive CE is the

**Table 2** Electrophoretic mobilities of mono-, di-, oligosaccharides, sugar alcohols, and aldonic acids

Compound	Relative mobility		Compound	Relative mobility	
	A*	B*		A*	B*
<i>Monosaccharides</i>			<i>1,6-Disaccharides</i>		
D,L-Glyceraldehyde	0.79		Gentiobiose	0.75	0.73
1,3-Dihydroxyacetone	0.78		Melibiose	0.80	0.79
L-Arabinose	0.96	1.00	Isomaltose	0.69	0.62
D-Ribose	0.77	0.70	<i>Tri-/tetrasaccharides</i>		
D-Xylose	1.00	1.01	Raffinose		0.29
2-Deoxy-D-ribose	0.33	–	Melezitose		0.28
L-Fucose	0.89	0.95	Maltotriose		0.26
L-Rhamnose	0.52	0.53	Stachyose		0.37
D-Galactose	0.93	0.92	<i>Sugar alcohols</i>		
D-Glucose	1.00	1.00	Meso-erythritol		0.68
D-Mannose	0.72	0.69	Threitol		0.87
D-Fructose	0.90	0.86	Adonitol (Ribitol)		0.87
L-Sorbose	0.95	0.95	Xylitol		0.79
<i>1,1-Disaccharides</i>			Arabitol	0.90	0.75
$\alpha$ , $\alpha'$ -Trehalose	0.19	0.22	Mannitol	0.90	0.90
$\alpha$ , $\beta'$ -Trehalose	0.23		Galactitol (dulcitol)	0.98	0.94
$\beta$ , $\beta'$ -Trehalose	0.19		Sorbitol (glucitol)	0.89	0.80
<i>1,2-Disaccharides</i>			Myo-inositol		0.48
Sopharose	0.24	0.35	<i>Aldonic acids</i>		
Sucrose	0.18	0.20	D-Gluconic acid		1.21
<i>1,3-Disaccharides</i>			D-Galacturonic acid		1.23
Nigerose	0.69		D-Arabinic acid		1.30
Laminaribiose	0.69		D-Ribonic acid		1.17
Turanose		0.64			
<i>1,4-Disaccharides</i>					
Maltose	0.32	0.39			
Lactose	0.38	0.42			
Cellobiose	0.23	0.32			

\*A, B: reference letters.

electroosmotic flow (EOF). EOF is a consequence of the surface charge on the wall of the capillary. The fused silica capillaries that are typically used for separations have ionizable silanol groups in contact with the buffer contained within the capillary. Separations can be carried out in two modes: either with the electroosmotic flow (EOF) or against it (Figure 3), Co-EOF allows the shortening of analysis time.

**Table 3** Electrophoretic mobilities of sugar acids, sugar phosphates, and amino sugars

Compound	C*	D*	E*	F*	G*
<i>Uronic acids</i>					
D-Glucuronic acid	1.44	1.00	1.00		
D-Galacturonic acid	1.18	0.42	0.26		
D-Mannuronic acid	1.00	0.79	0.74		
L-Guluronic acid	0.85	0.26	0.06		
<i>Sugar phosphates</i>					
D-Ribose-5-phosphate				0.69	
D-Ribose-1,5-diphosphate				1.04	
D-Fructose-1-phosphate				0.63	
D-Fructose-6-phosphate			0.81	0.63	
D-Fructose-1,6-diphosphate			1.38	0.95	
D-Glucose-1-phosphate			0.86	0.64	
D-Glucose-6-phosphate			0.77	0.61	
D-Mannose-6-phosphate				0.63	
D-Galactose-6-phosphate				0.63	
<i>Amino sugars</i>					
N-Acetylglucosamine					0.23
N-Acetylgalactosamine					0.33
N-Acetylmannosamine					0.65
N-Acetylxylosamine					0.23
N-Acetylfucosamine					0.14

\*C–G: reference letters.

The detection of the separated zones in the tube can be carried out in different modes: detection of underivatized carbohydrates through direct UV detection, indirect UV detection, or by electrochemical methods. Carbohydrates with labeled chromophore groups can be detected by UV or fluorescence detectors.

### Detection of Underivatized Carbohydrates

**Direct UV detection** Generally, carbohydrates have no chromophores in their molecules and they absorb UV light only at wavelengths below 200 nm. Exceptions are acidic disaccharides obtained by the enzymatic digestion of glycosaminoglycans, which have an unsaturated uronic acid unit in their molecules. They can be detected by direct UV detection at 232 nm. The oligosaccharides containing *N*-acetylglucosamine, *N*-acetylgalactosamine, and sialic acid residues can be detected at 200 nm but at this range the UV absorbances are not selective.

As an interesting effect, a 2–50-fold increase in signal intensity for mono- and oligosaccharides using UV detection at 195 nm was observed by applying a borate buffer at pH values between 9 and 10, which enables the detection of these sugars by direct UV absorption. The highest reproducibility could be obtained at elevated temperatures (40–60°C).

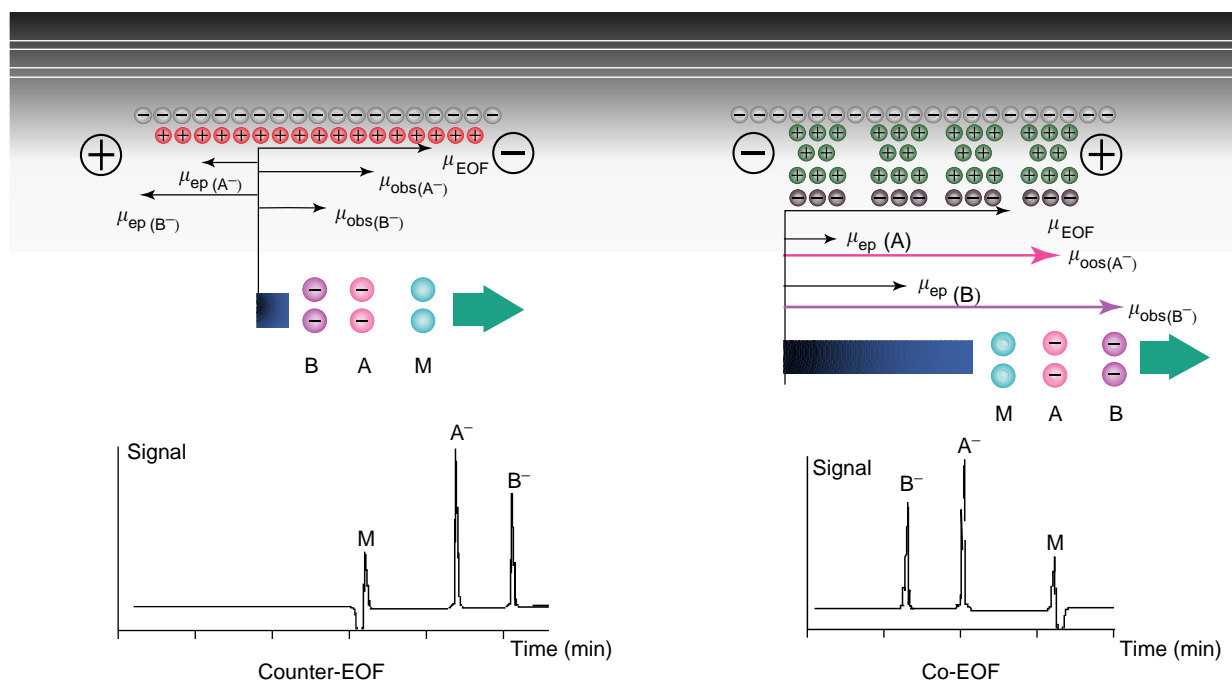
Table 5 gives some examples for the practical application of CE using direct UV detection.

**Indirect UV detection** In strong alkaline medium and in the presence of alkali-metal hydroxides (LiOH, NaOH, KOH), the carbohydrates possess slight acidic properties due to the ionization of the hydroxyl groups, yielding in the so-called alcoholates. (The ionization constants *pK* for mono- and oligosaccharides in water are between 11.98 and 12.74, and for sugar alcohols ~13.5.) Under these conditions, these compounds migrate in the electric field. When using

**Table 4** Chromatographic conditions corresponding to reference letters in Tables 2 and 3

Reference letter	Electrolyte	Support	Voltage	Reference substance
A	0.2 mol l <sup>-1</sup> Sodium borate, pH 10	Chromatographic paper (Whatman 3)	500 V (15 V cm <sup>-1</sup> )	D-Glucose
B	0.3 mol l <sup>-1</sup> Sodium borate (EDTA), pH 10	Silanized silica gel–1-octanol	400 V (20 V cm <sup>-1</sup> )	D-Glucose
C	0.01 mol l <sup>-1</sup> Sodium borate–0.005 mol l <sup>-1</sup> CaCl <sub>2</sub> , pH 9.2	Chromatographic paper (Schleicher-Schüll 2043b)	0.5 mA cm <sup>-1</sup>	D-Mannuronic acid
D	0.1 mol l <sup>-1</sup> Barium acetate, pH 7.8	Chromatographic paper (Whatman)	600 V	D-Glucuronic acid
E	0.07 mol l <sup>-1</sup> Barium acetate	Silanized silicagel–1-octanol	400 V (20 V cm <sup>-1</sup> )	D-Glucuronic acid
F	Pyridine–acetic acid–water 20:64:916, pH 3.9	Chromatographic paper (Whatman 3 washed)	2000 V (35 V cm <sup>-1</sup> )	Orthophosphate
G	Borate buffer, pH 10 (23.4 g sodium tetraborate + 30 ml 1 M NaOH/l)	Chromatographic paper	1300 V	D-Glucose





**Figure 3** Counter and co-electroosmotic flow of negatively charged analytes A, B, and a neutral analyte M in capillary electrophoresis.  $\mu_{\text{EOF}}$ , electroosmotic flow (EOF);  $\mu_{\text{obs}}$ , observed flow of the individual analyte;  $\mu_{\text{ep}}$ , flow of analyte without EOF.

indirect UV detection, a UV-absorbing co-ion with high molar absorbance or with high fluorescence is added to the electrolyte used to provide a constant detector response. Due to the maintenance of charge neutrality, the analyte with the same charge as the co-ion displaces the co-ion resulting in a decrease in detector response. Indirect photometric or fluorometric detection has the advantage that there is no need for pre- or postcolumn derivatization of the compounds to be studied. The following compounds are used as co-ion for the application of indirect detection: sorbic acid, tryptophan, 1-naphthylacetic acid,  $\beta$ -indoleacetic acid, 2,6-pyridinedicarboxylic acid, benzoic acid in the case of alkaline medium. For the separation of negatively charged species the following compounds are used: heparin-derived oligosaccharides in neutral or slight acidic medium (pH 2.5–3.5), 5-sulfosalicylic acid, or 1,2,4-tricarboxybenzoic acid.

For indirect fluorescence detection coumarin or fluorescein are used. This indirect UV (or fluorescence) detection of the separated carbohydrates appears simple, but it has some drawbacks such as the instability of the detection system itself resulting in a drift of the baseline, the unfavorable relation between the analyte and the co-ion giving low efficiencies at higher sample concentrations of the sample, and the lack of specificity for the detection. Nevertheless, this method was applied for the

determination of sugars in food samples such as fruit juices (Figure 4).

**Electrochemical detection** There are two variations of this technique:

1. amperometric detection at a constant potential (ADCP), and
2. pulsed amperometric detection (PAD).

*Amperometric detection at a constant potential* The materials best suited for use as working electrodes are elementary copper and copper(I) oxide. Using a microelectrode consisting of a Cu wire of 25  $\mu\text{m}$  diameter, Ag/AgCl–3 mol l<sup>-1</sup> KCl as reference electrode, and a Pt wire as an auxiliary electrode at +0.6 V, 15 different mono- and oligosaccharides were detected in a single run using a fused-silica capillary (length: 73 cm, 50  $\mu\text{m}$  i.d.) at 100 mmol l<sup>-1</sup> NaOH and 11 kV. The limit of detection was below 50 fmol and the linear response of the detector signal/concentration was over four decades. The copper electrode maintains a constant potential even after hundreds of runs. This system proved to be useful for the determination of sugars in carbonated soft drinks like Coca Cola. An alternative system using the ADCP technique is the use of copper(I) oxide (Cu<sub>2</sub>O) embedded in a matrix of conductive carbon cement with Ag/AgCl as the



**Table 5** Examples for the practical application of CE for the analysis of carbohydrates

<i>Analytes</i>	<i>Derivatization method; Detection (D)</i>	<i>Capillary (C); electrolyte (E), voltage (V)</i>	<i>Application</i>
Oligosaccharides derived from enzymatic cleavage of hyaluronan chondroitin sulfate, etc.	Direct UV detection D: $\lambda = 200\text{--}232\text{ nm}$	C: Fused silica (72 cm $\times$ 50 $\mu\text{m}$ i.d.) E: 40 mmol l <sup>-1</sup> phosphate, 40 mmol l <sup>-1</sup> SDS, 10 mmol l <sup>-1</sup> borate, pH 9 V: 15 kV	Study of digests of connective tissues; diagnosis of arthritis
Monosaccharides, oligosaccharides, uronic acid, especially D-glucose, D-fructose, sucrose	Indirect UV detection D: $\lambda = 256\text{ nm}$	C: Fused silica (90 cm $\times$ 50 $\mu\text{m}$ i.d.) E: 6 mmol l <sup>-1</sup> potassium sorbate, pH 12.2 V: 20 kV	Determination of sugars in fruit juices; comparison of the results to those of liquid chromatography methods
D-Glucose, D-glucosamine, D-glucosamine phosphate	Pulse amperometry, gold working electrode D: E <sub>1</sub> : +325 mV/199 ms; E <sub>2</sub> : +800 mV/166 ms; reactivation: E <sub>3</sub> : -600 mV/249 ms	C: Fused silica (95 cm $\times$ 75 $\mu\text{m}$ i.d.) polyimine coated E: 10 mmol l <sup>-1</sup> NaOH-8 mmol l <sup>-1</sup> Na <sub>2</sub> CO <sub>3</sub> V: 25 kV	Determination of D-glucose in blood samples
Monosaccharides; oligosaccharides	Constant potential amperometry copper-working electrode E: +600 mV	C: Fused silica (73 cm $\times$ 50 $\mu\text{m}$ i.d.) E: 100 mmol l <sup>-1</sup> NaOH V: 11 kV	Determination of sugars in soft drinks
Oligosaccharides from glycopeptides by hydarazinolyses	Derivatization: 2-pyridylamine; dimethylamine/borane D: FI: $\lambda_{\text{ex}} = 320\text{ nm}$ $\lambda_{\text{em}} = 380\text{ nm}$	C: Fused silica (60 cm $\times$ 50 $\mu\text{m}$ i.d.) polyimine coated E: 100 mmol l <sup>-1</sup> phosphate, pH 2.5; 200 mmol l <sup>-1</sup> borate V: 20 kV	Glycopeptides obtained by trypsin digestion of human transferrin, serum fetuin, human immunoglobulin, bovine, spleen ribonuclease, yeast invertase
Monosaccharides: rhamnose, xylose, glucose, arabinose, galactose, glucuronic acid, galacturonic acid	Derivatization: 4-aminobenzoic acid ethyl ester D: UV: $\lambda = 305\text{ nm}$	C: Fused silica (72 cm $\times$ 50 $\mu\text{m}$ i.d.) E: 200 mmol l <sup>-1</sup> borate pH 10.5 V: 25 kV	Analysis of the monosaccharide units of the polysaccharides from <i>Flos matricariae</i>
Monosaccharides, uronic acids and xylose-containing oligosaccharides	Derivatization: 4-aminobenzoic acid ethyl ester; D: UV: $\lambda = 306\text{ nm}$	C: fused silica (44 cm $\times$ 30 $\mu\text{m}$ i.d.) E: 438 mmol l <sup>-1</sup> borate, pH 11.5 V: power 1200 mW	Mono- and oligosaccharides obtained by enzymatic hydrolysis of wood and pulp
Glycoproteins and monosaccharides: sialic acid, amino sugars, and neural sugars	Derivatization: 2-aminoacridone for sialic acid; 8-aminopyrene-1,3,6-trisulfonate (for amino and neutral sugars) D: LIF (laser induced fluorescence): Ar-laser, $\lambda_{\text{ex}} = 488\text{ nm}$ , $\lambda_{\text{em}} = 520\text{ nm}$	C: Fused silica (37 cm $\times$ 25 $\mu\text{m}$ i.d.) E: 25 mmol l <sup>-1</sup> Li-borate pH 10.0 V: 750 V cm <sup>-1</sup>	Oligosaccharides obtained by trifluoroacetic hydrolysis (0.1 mol l <sup>-1</sup> ; 2 mol l <sup>-1</sup> ; 4 mol l <sup>-1</sup> ) of bovine fetuin
Oligomers of D-glucose and D-galacturonic acid (up to 30 units)	Derivatization: 8-aminonaphthalene-1,3,6-trisulfonate D: UV: $\lambda = 233\text{ nm}$ LIF: He-Cd laser, $\lambda_{\text{ex}} = 325\text{ nm}$ , $\lambda_{\text{em}} = 520\text{ nm}$	C: Fused silica (37-42 cm $\times$ 50 $\mu\text{m}$ i.d.) E: 50 mmol l <sup>-1</sup> phosphate pH 2.5 V: 20-25 kV	Enzymatic hydrolysis of polygalacturonic acid, dextrin, and dextran
Enantiomers of monosaccharides (D- and L-sugars like: D- and L-galactose, D- and L-glucose, etc.)	Derivatization: S-( - )-1-phenylethylamine D: UV: $\lambda = 200\text{ nm}$	C: Fused silica (77 cm $\times$ 50 $\mu\text{m}$ i.d.) E: 50 mmol l <sup>-1</sup> borate buffer + 30% acetonitrile, pH 10.3 V: 28 kV	Determination of the monosaccharide enantiomers in beverages
Monosaccharides; units of polysaccharides; co-electroosmotic separation	Derivatization: 4-aminobenzoic acid ethyl ester (ABEE) and 4-aminobenzo-nitrile (ABN) D: UV: 280 nm V: 15-18 kV	C: Fused silica (57.5 cm $\times$ 50 $\mu\text{m}$ i.d.) E: 500 mmol l <sup>-1</sup> borate 20% methanol, pH 10.0, 0.001% hexadimethrine bromide (HDB) (for ABEE) 350 mmol l <sup>-1</sup> borate, 5% methanol, 5% acetonitrile, 0.001% HDB pH 10.5 (for ABN)	Analysis of the monomeric units of sugarcane bagasse and of plant polysaccharides

reference electrode and stainless steel fiber as the auxiliary electrode.

**Pulsed amperometric detection** The principle of this technique is based on the electrochemical oxidation of the carbohydrates at surfaces of noble metal microelectrodes such as gold or platinum but the surfaces of these electrodes strongly adsorb the oxidation products resulting in a rapid decrease in analyte response. For cleaning the surfaces of these electrodes by oxidation, the positive potential is enhanced yielding a layer of oxides of the noble metal, which are then immediately reduced by changing the potential to negative values. When using a gold working electrode the range of the voltage waves are as follows: (1) detection step, 200–300 mV (110–220 ms); (2) 900–1200 mV (55–110 ms) for the cleaning step (oxidation); and (3) 500–1000 mV (110–330 ms) for restoring the electrodes (reduction step). Using  $0.1 \text{ mol l}^{-1}$  NaOH as an electrolyte and a voltage of 30 kV the common sugar and sugar alcohols can be separated with this system. This method has proved to be useful for the determination of glucose in human blood and for the analysis of oligosaccharide residues resulting from glycoproteins and glycopeptides. A drawback of PAD is the necessity of special pulse sequences requiring expensive instrumentation.

**Mass spectrometry (MS) detection** In the case of MS detection, there must be a distinction between online and offline CE–MS modes.

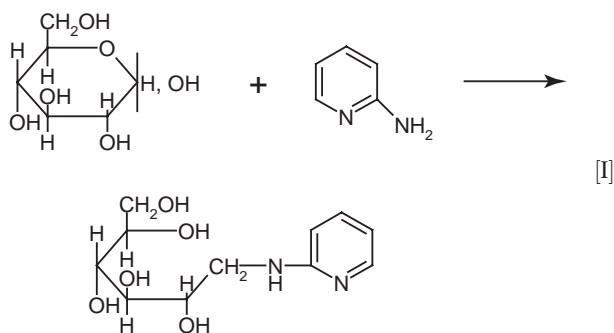
In the online mode the fused silica capillary used in CE is connected to the mass spectrometer through an electrospray ionization interface. This system is used for the identification of oligosaccharides, released by enzymatic digestion of glycosaminoglycans (e.g. heparin), glycopeptides, and glycoproteins. The offline mode is used for the identification of the structure of glycopeptides. Data are collected by a micropreparative CE system for the mass determination by matrix-assisted laser desorption ionization mass spectrometry.

**Refractive index detection** A microvolume refractive index detection method has been developed for CE consisting of a micro-interferometer back-setter system, illuminated by a He–Ne laser giving a characteristic interference pattern. The change of such a pattern is due to the change of the refractive index caused by the separated substance. This detection system was tested for the separation of mono- and oligosaccharides.

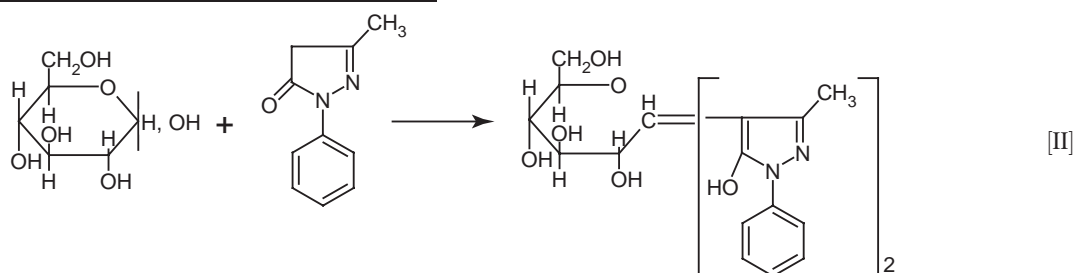
### Detection of Derivatized Carbohydrates (Precolumn Derivatization)

Labeling of carbohydrates by the introduction of a chromophore or a fluorophore group leads to a fundamental improvement of detection of the separated carbohydrate zones after the electrophoretic run. For this purpose the following chemical reactions are used:

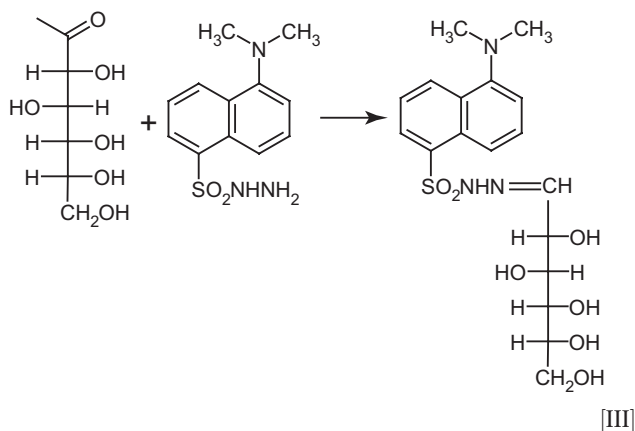
1. Reductive amination: reaction of the reducing carbohydrate with primary aromatic or heterocyclic amines in the presence of sodium cyanoborohydride yielding the corresponding aminoalditols, e.g.,



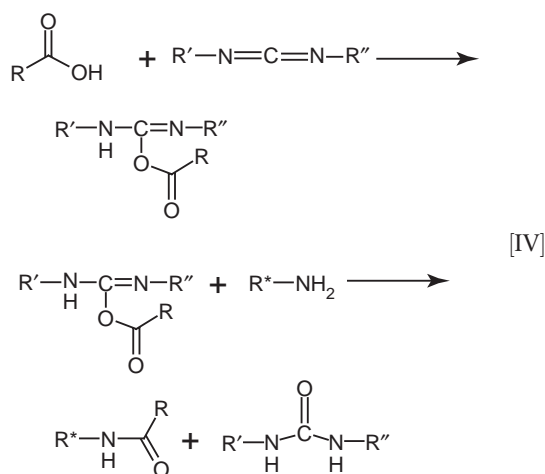
2. Condensation of the reducing carbohydrate with 1-phenyl-3-methyl-5-pyrazolone (PMP) in alkaline medium yielding the corresponding bis-PMP-compound:



3. Condensation of reducing sugars with aromatic amines yielding the corresponding hydrazones, in reaction [III], and



4. Condensation of carboxylated carbohydrates with UV-absorbing amines in the presence of carbodiimide:



The reductive amination (reaction [I]) is the most frequently used labeling reaction. Regarding the chromophore (respectively) fluorophore group there must be distinction between (1) neutral or ionizable and (2) permanently charged derivatives. In Table 6, the compounds belonging to groups (1) and (2) are listed. The chemical structures of these compounds for labeling are shown in Figure 5.

For the most frequent practical use, stock solutions of the required amine in methanol or tetrahydrofuran containing up to 10% acetic acid are prepared. Immediately before use, an excess of sodium cyano-

**Table 6** Chromophore (UV) and fluorophore (FL) compounds for labeling the reductive carbohydrates by reductive amination

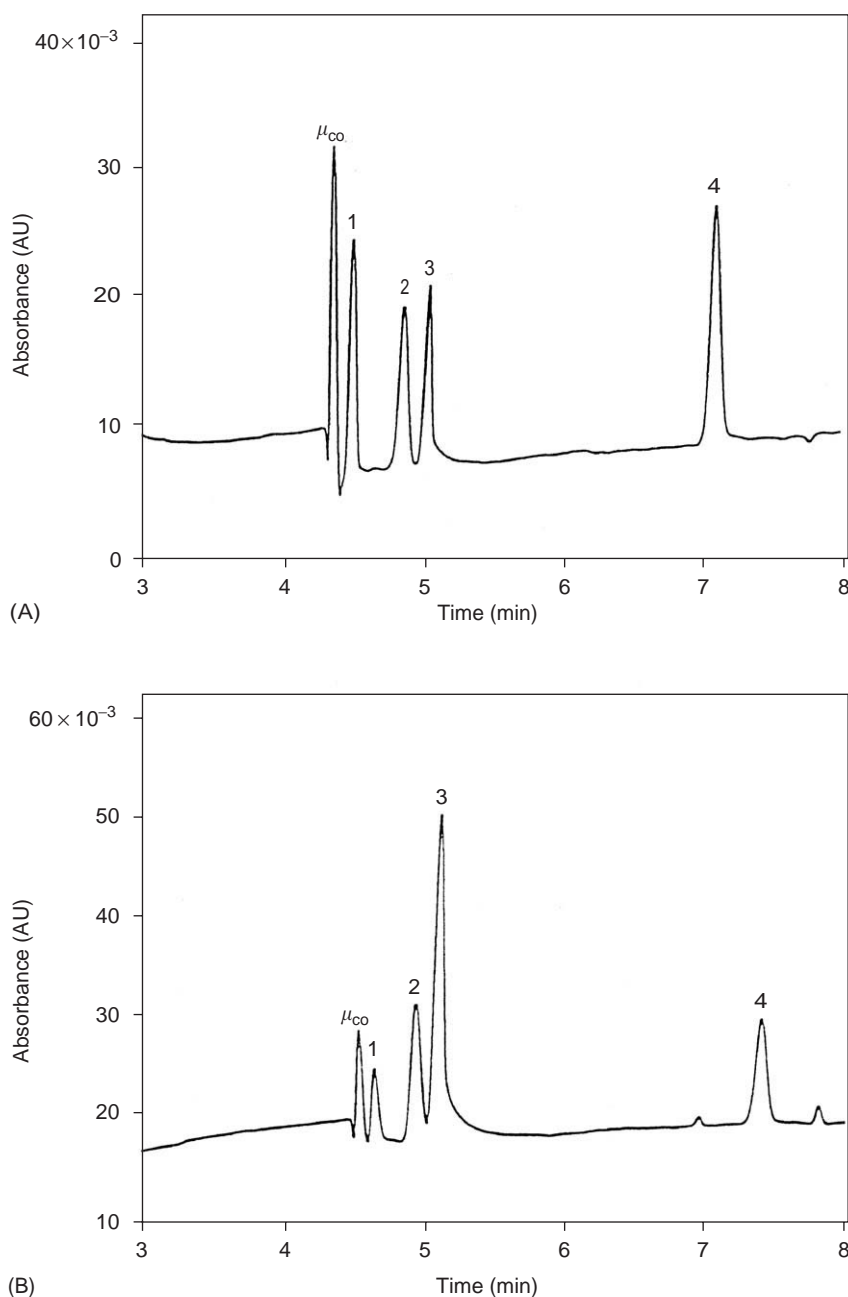
*A: Compounds giving neutral or ionizable carbohydrate compounds*

- (1) 2-Aminoacridone (LIF: Ar-ion laser;  $\lambda_{\text{exc}} = 425/488$  nm,  $\lambda_{\text{em}} = 550/520$  nm)
- (2) 2-Aminobenzamide ( $\lambda_{\text{exc}} = 330$  nm,  $\lambda_{\text{em}} = 420$  nm)
- (3) 2-Aminopyridine (FL:  $\lambda_{\text{exc}} = 320$  nm,  $\lambda_{\text{em}} = 380\text{--}400$  nm; LIF: He-Cd laser;  $\lambda_{\text{exc}} = 325$  nm, UV:  $\lambda_{\text{max}} = 240$  nm)
- (4) 6-Aminoquinoline (LIF: He-Cd laser or Ar-ion laser;  $\lambda_{\text{exc}} = 325$  nm,  $\lambda_{\text{em}} > 495$  nm, UV:  $\lambda_{\text{max}} = 270$  nm)
- (5) 4-Aminobenzonitrile (UV:  $\lambda_{\text{max}} = 285$  nm)
- (6) Ethyl-4-aminobenzoate (UV:  $\lambda_{\text{max}} = 305$  nm)
- (7) S-(−)-1-phenylethylamine (UV:  $\lambda = 200$  nm)
- (8) 3-(4-Carboxybenzoyl)-2-quinolinecarbaldehyde (CBQCA) (LIF:  $\lambda_{\text{exc}}$ : Ar-ion laser: 457 or 488 nm; He-Cd laser: 442 nm,  $\lambda_{\text{em}} = 552$  nm)

*B: compounds giving permanently charged carbohydrate compounds*

- (9) 4-Aminobenzoic acid (UV:  $\lambda_{\text{max}} = 285$  nm)
- (10) Sulfanilic acid (UV:  $\lambda_{\text{max}} = 247$  nm)
- (11) 2-Aminonaphthalene-1-sulfonic acid (2-ANSA) (UV:  $\lambda_{\text{max}} = 235$  nm)
- (12) 5-Aminonaphthalene-2-sulfonic acid (5-ANSA) (LIF: He-Cd laser:  $\lambda_{\text{exc}} = 325$  nm,  $\lambda_{\text{em}} = 475$  nm; UV:  $\lambda_{\text{max}} = 235$  nm)
- (13) 7-Aminonaphthalene-1,3-disulfonic acid (ANDSA) (FL: Xe-Hg lamp,  $\lambda_{\text{exc}} = 315$  nm,  $\lambda_{\text{em}} = 420$  nm; UV:  $\lambda_{\text{max}} = 247$  nm)
- (14) 4-Amino-5-hydroxynaphthalene-2,7-disulfonic acid (AHNS) (FL: He-Cd laser:  $\lambda_{\text{exc}} = 325$  nm,  $\lambda_{\text{em}} = 475$  nm)
- (15) 3-Aminonaphthalene-2,7-disulfonic acid (3-ANDSA) (UV:  $\lambda_{\text{max}} = 235$  nm)
- (16) 8-Aminonaphthalene-1,3,6-trisulfonic acid (ANTS) (FL:  $\lambda_{\text{exc}} = 370$  nm,  $\lambda_{\text{em}} = 570$  nm, LIF: He-Cd laser 325 nm; UV: 223 nm)
- (17) 9-Aminopyrene-1,4,6-trisulfonic acid (FL:  $\lambda_{\text{exc}} = 455$  nm,  $\lambda_{\text{em}} = 512$  nm; Ar-ion laser 488 nm)

borohydride is added to an aliquot of this solution. This solution is mixed with the sample containing the reducing carbohydrates. Finally, this mixture is heated up to 50–80°C from 15 min to 2 h. After the dilution of the reaction mixture, mostly with water, the sample is ready for electrophoresis. By accurately controlling the reaction conditions, the yields of labeled carbohydrates are reproducible, in many cases quantitative. In special cases it is necessary to label the carboxylic group of acid mono- or oligosaccharides with chromophore- or fluorophore-bearing compounds, which is carried out by forming the corresponding amide due to reaction [IV], using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloric as reactive agents. This method was



**Figure 4** Capillary zone electrophoresis of fruit juices: (A) standard mixture (sugars,  $2.22\text{--}2.42\text{ mmol l}^{-1}$ ) and internal standard (glucuronic acid,  $1.15\text{ mmol l}^{-1}$ ); (B) apple juice (diluted 1:50). Running conditions: buffer  $6\text{ mmol l}^{-1}$  sorbate; capillary,  $42\text{ cm}$  ( $35\text{ cm}$  effective length)  $\times 50\text{ }\mu\text{m}$  i.d.;  $230\text{ V cm}^{-1}$ ;  $15^\circ\text{C}$ ; indirect UV detection at  $256\text{ nm}$ ; 1-s injection. 1 = Sucrose; 2 = glucose; 3 = fructose; 4 = glucuronic acid.

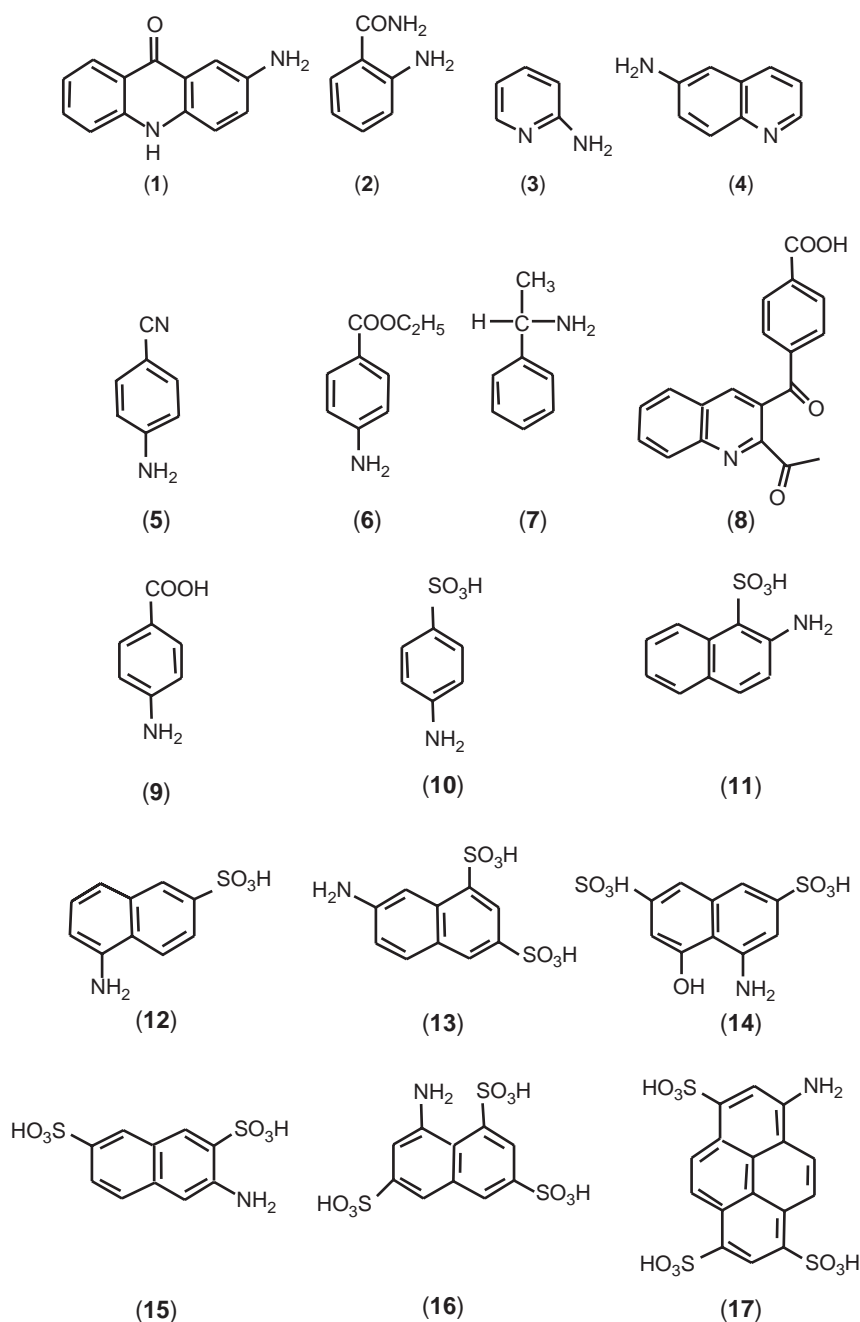
applied for labeling of oligosaccharides derived from enzymatic digestion of glycosaminoglycans.

For electrophoretic separation two different kinds of electrolytes are generally used:

1. borate containing electrolytes with pH  $8.5\text{--}10.5$  for the separation of the single-, mono-, and oligosaccharides

2. neutral to slightly acidic electrolytes pH  $2.0\text{--}5.0$  for the separation of the oligosaccharides according to their number of monomeric units.

Figures 6 and 7 show an example for the separation of sugars in both kinds of electrolytes.



**Figure 5** Chemical structures of compounds cited in Table 6 for labeling.

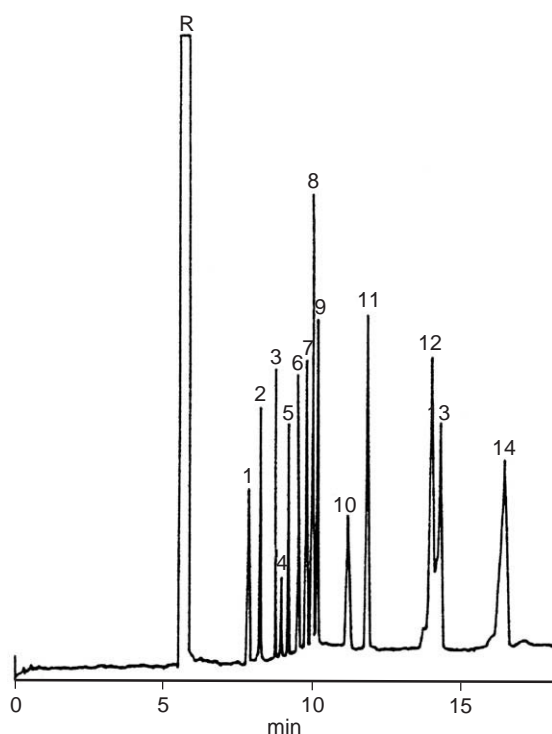
### Capillary Electrochromatography

Capillary electrochromatography (CEC) is a hybrid of CE and liquid chromatography. The separation is carried out in capillaries that are either fully packed with octadecyl-(ODS)-silica particles or partially packed having an open segment. The separation of the analytes is based on the difference in the magnitude of the distribution between the mobile and stationary phase. The flow of mobile phases is generated

by the electroosmosis (EOF) due to the negative charged surface of the silica particles.

CEC was recently applied to the separation of derivatized mono- and oligosaccharides, namely:

- separation of the phenyl-methyl-pyrazolone derivatives of mono- and oligosaccharides;
- separation of 2-aminobenzamide of mono- and oligosaccharides;



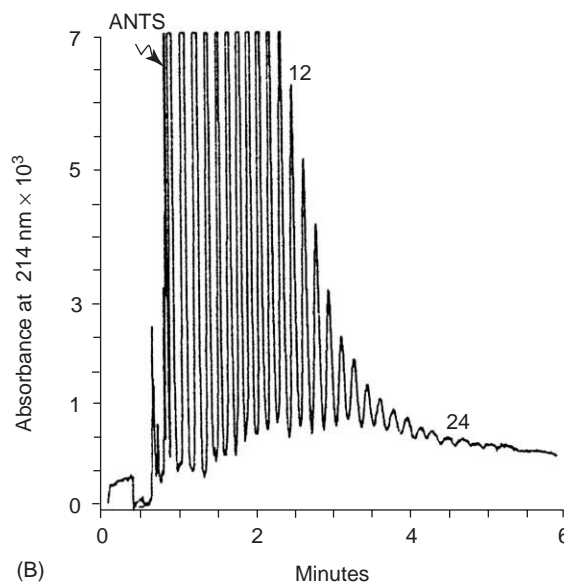
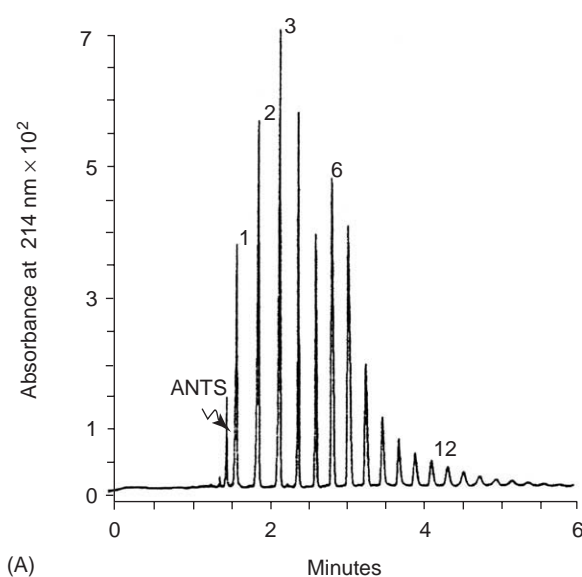
**Figure 6** Separation of mono- and oligosaccharides derivatized with ethyl *p*-aminobenzoate. Capillary: fused silica,  $L=72$  cm,  $l=50$  cm,  $50\text{ }\mu\text{m}$  i.d.; carrier:  $175\text{ mmol l}^{-1}$  borate, pH 10.5; voltage: 25 kV; detection: UV, 305 nm; temperature: 30°C; injection: Vacuum, 1.0 s. Zone identification: R = reagent, 1 = 2-deoxy-D-ribose, 2 = maltotriose, 3 = rhamnose, 4 = cellobiose, 5 = xylose, 6 = ribose, 7 = lactose, 8 = glucose, 9 = arabinose, 10 = fucose, 11 = galactose, 12 = mannuronic acid, 13 = glucuronic acid, 14 = galacturonic acid.

- separation of some *p*-nitrophenylglycosides; and
- separation of glycolipids (neutral and acid glycosphingolipids).

## Microchip Capillary Electrophoresis of Carbohydrates

Microfluidic separation devices have generated considerable interest due to the fact that sample preparation, injection, separation, derivatization, and detection can be integrated into one miniaturized device. A significant amount of effort has been dedicated to the development of diagnostic tools for the benefit of diabetic patients, which resulted in new biosensors and bioassays for monitoring glucose. A glucose/alcohol chip for monitoring glucose and ethanol based on the reaction with glucose oxidase and alcohol dehydrogenase was developed. The detection limits were  $10\text{ }\mu\text{mol l}^{-1}$  for glucose and  $500\text{ }\mu\text{mol l}^{-1}$  for ethanol and the analysis time was less than 100 s.

The catalytic oxidation of the carbohydrates in the presence of Cu(II) for the detection of sucrose, galactose, and fructose was exploited using a Teflon-coated platinum wire plated with copper as the working electrode. Therefore, the addition of copper ions in the run buffer increased the sensitivity to an order of magnitude compared to run buffer without copper. Detection limits were  $\sim 1\text{ }\mu\text{mol l}^{-1}$ .



**Figure 7** Electropherogram of ANTS-derivatized malto-oligosaccharides, obtained at acidic pH at two different temperatures with a  $270\text{ mm} \times 50\text{ }\mu\text{m}$  capillary. (A)  $50\text{ mmol l}^{-1}$  phosphate buffer, pH 2.5; temperature, 25°C; voltage, 15 kV; 22 ng of sample. (B)  $30\text{ mmol l}^{-1}$  phosphate buffer, pH 2.5; temperature, 50°C; voltage, 17 kV; 31 ng of sample. In both cases, the samples were introduced at the cathodic end.



See also: **Capillary Electrochromatography**. **Capillary Electrophoresis**: Overview. **Electrophoresis**: Overview; Principles.

## Further Reading

- Breen M, Weinstein HG, Blacik LJ, Borcharding MS, and Sittig RA (1976) Microanalysis and characterisation of glycosaminoglycans from human tissue via zone electrophoresis. In: Whistler RE (ed.) *Methods of Carbohydrate Chemistry*, vol. 7, pp. 101–115. New York: Academic Press.
- Chiesa C and Horvath C (1993) Capillary zone electrophoresis of maltooligosaccharides derivatized with 8-aminonaphthalene-1,3,6-trisulfonic acid. *Journal of Chromatography A* 645: 337–352.
- Dahlman O, Jacobs A, Liljenberg A, and Olsson AI (2000) Analysis of carbohydrates in wood and pulps employing enzymatic hydrolysis and subsequent capillary zone electrophoresis. *Journal of Chromatography A* 891: 157–174.
- El Rassi Z (2002) Capillary electrophoresis and electrochromatography of carbohydrates. *Journal of Chromatography Library* 66: 597–676.
- Guo-Fu H (1995) Fluorophore-assisted carbohydrate electrophoresis: technology and applications. *Journal of Chromatography A* 705: 89–103.
- Hase S (1996) Pre-column derivatization for chromatographic and electrophoretic analysis of carbohydrates. *Journal of Chromatography A* 720: 173–182.
- Honda S, Iwase S, Makino A, and Fujiwara S (1989) Simultaneous determination of reducing monosaccharides by capillary zone electrophoresis as borate complexes of N-2-pyridylglycamines. *Analytical Biochemistry* 176: 72–77.
- Huck C and Bonn G (2001) Capillary electrochromatographic methods for the separation of p-nitrophenyl and 1-phenyl-3-methyl-5 pyrazolone derivatised mono- and oligosaccharides. *Journal of Carbohydrate Chemistry* 20: 1–16.
- Jandik P and Bonn G (1993) *Capillary electrophoresis of small molecules and ions*. New York: VCH.
- Klockow A, Paulus A, Figueredo V, Amadó P, and Widmer HM (1994) Determination of carbohydrates in fruit juices by capillary electrophoresis and high performance liquid chromatography. *Journal of Chromatography A* 680: 187–200.
- La Course WR and Owens GS (1996) Pulse electrochemical detection of non chromophoric compounds following capillary electrophoresis. *Electrophoresis* 17: 310–318.
- Lacher NA, Garrison KE, Martin RS, and Lunte SM (2001) Microchip capillary electrophoresis/electrochemistry. *Electrophoresis* 22: 2526–2536.
- Oefner P and Scherz H (1994) Capillary electrophoresis and thin-layer electrophoresis of carbohydrates. In: Chrambach A, Dunn MJ and Radola BJ (eds.) *Advances in Electrophoresis*, vol. 7. New York: VCH Publications.
- Scherz H and Bonn G (1998) *Analytical Chemistry of Carbohydrates*. Stuttgart: G. Thieme Verlag.
- Suzuki S and Honda S (1998) A tabulated review of capillary electrophoresis of carbohydrates. *Electrophoresis* 19: 2539–2560.
- Suzuki S, Kazuaki K, and Susumu H (1992) Two-dimensional mapping of N-glycosidically linked asialo-oligosaccharides from glycoproteins as reductively pyridylaminated derivatives using dual separation modes of high-performance capillary electrophoresis. *Analytical Biochemistry* 205(2): 227–236.
- Vorndran A, Grill E, Huber C, Oefner P, and Bonn G (1992) Capillary zone electrophoresis of aldoses, ketoses and uronic acids derivatised with ethyl-p-aminobenzoate. *Chromatographia* 34: 109–113.
- Weigel H (1963) Paper electrophoresis of carbohydrates. *Advances in Carbohydrate Chemistry* 18: 61–97.

## Clinical Applications

J A Beeley, University of Glasgow, Glasgow, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The analysis of human body fluids and tissue extracts by electrophoresis and clinical applications of the technique is one of the most rapidly developing areas of biological research. Whilst analysis of serum proteins by one-dimensional electrophoresis on gel-media or even cellulose acetate still continues to produce

valuable information in the clinical laboratory, capillary electrophoresis and proteomics, including microchip technology, are increasingly becoming the focus of current methodologies and research. Computer analysis of gels produced by sophisticated two-dimensional separation procedures followed by identification of the proteins by mass spectrometry is facilitating the generation of databases on hundreds of proteins in various body tissues and fluids. Changes in sometimes only a few minor components can be correlated with pathological disorders. This article will review techniques currently used in the

See also: **Capillary Electrochromatography**. **Capillary Electrophoresis**: Overview. **Electrophoresis**: Overview; Principles.

## Further Reading

- Breen M, Weinstein HG, Blacik LJ, Borcharding MS, and Sittig RA (1976) Microanalysis and characterisation of glycosaminoglycans from human tissue via zone electrophoresis. In: Whistler RE (ed.) *Methods of Carbohydrate Chemistry*, vol. 7, pp. 101–115. New York: Academic Press.
- Chiesa C and Horvath C (1993) Capillary zone electrophoresis of maltooligosaccharides derivatized with 8-aminonaphthalene-1,3,6-trisulfonic acid. *Journal of Chromatography A* 645: 337–352.
- Dahlman O, Jacobs A, Liljenberg A, and Olsson AI (2000) Analysis of carbohydrates in wood and pulps employing enzymatic hydrolysis and subsequent capillary zone electrophoresis. *Journal of Chromatography A* 891: 157–174.
- El Rassi Z (2002) Capillary electrophoresis and electrochromatography of carbohydrates. *Journal of Chromatography Library* 66: 597–676.
- Guo-Fu H (1995) Fluorophore-assisted carbohydrate electrophoresis: technology and applications. *Journal of Chromatography A* 705: 89–103.
- Hase S (1996) Pre-column derivatization for chromatographic and electrophoretic analysis of carbohydrates. *Journal of Chromatography A* 720: 173–182.
- Honda S, Iwase S, Makino A, and Fujiwara S (1989) Simultaneous determination of reducing monosaccharides by capillary zone electrophoresis as borate complexes of N-2-pyridylglycamines. *Analytical Biochemistry* 176: 72–77.
- Huck C and Bonn G (2001) Capillary electrochromatographic methods for the separation of p-nitrophenyl and 1-phenyl-3-methyl-5 pyrazolone derivatised mono- and oligosaccharides. *Journal of Carbohydrate Chemistry* 20: 1–16.
- Jandik P and Bonn G (1993) *Capillary electrophoresis of small molecules and ions*. New York: VCH.
- Klockow A, Paulus A, Figueredo V, Amadó P, and Widmer HM (1994) Determination of carbohydrates in fruit juices by capillary electrophoresis and high performance liquid chromatography. *Journal of Chromatography A* 680: 187–200.
- La Course WR and Owens GS (1996) Pulse electrochemical detection of non chromophoric compounds following capillary electrophoresis. *Electrophoresis* 17: 310–318.
- Lacher NA, Garrison KE, Martin RS, and Lunte SM (2001) Microchip capillary electrophoresis/electrochemistry. *Electrophoresis* 22: 2526–2536.
- Oefner P and Scherz H (1994) Capillary electrophoresis and thin-layer electrophoresis of carbohydrates. In: Chrambach A, Dunn MJ and Radola BJ (eds.) *Advances in Electrophoresis*, vol. 7. New York: VCH Publications.
- Scherz H and Bonn G (1998) *Analytical Chemistry of Carbohydrates*. Stuttgart: G. Thieme Verlag.
- Suzuki S and Honda S (1998) A tabulated review of capillary electrophoresis of carbohydrates. *Electrophoresis* 19: 2539–2560.
- Suzuki S, Kazuaki K, and Susumu H (1992) Two-dimensional mapping of N-glycosidically linked asialo-oligosaccharides from glycoproteins as reductively pyridylaminated derivatives using dual separation modes of high-performance capillary electrophoresis. *Analytical Biochemistry* 205(2): 227–236.
- Vorndran A, Grill E, Huber C, Oefner P, and Bonn G (1992) Capillary zone electrophoresis of aldoses, ketoses and uronic acids derivatised with ethyl-p-aminobenzoate. *Chromatographia* 34: 109–113.
- Weigel H (1963) Paper electrophoresis of carbohydrates. *Advances in Carbohydrate Chemistry* 18: 61–97.

## Clinical Applications

J A Beeley, University of Glasgow, Glasgow, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The analysis of human body fluids and tissue extracts by electrophoresis and clinical applications of the technique is one of the most rapidly developing areas of biological research. Whilst analysis of serum proteins by one-dimensional electrophoresis on gel-media or even cellulose acetate still continues to produce

valuable information in the clinical laboratory, capillary electrophoresis and proteomics, including microchip technology, are increasingly becoming the focus of current methodologies and research. Computer analysis of gels produced by sophisticated two-dimensional separation procedures followed by identification of the proteins by mass spectrometry is facilitating the generation of databases on hundreds of proteins in various body tissues and fluids. Changes in sometimes only a few minor components can be correlated with pathological disorders. This article will review techniques currently used in the

analysis of human body fluids (serum and plasma, saliva, cerebrospinal fluid (CSF), urine, and lacrimal fluid) as well as cells and tissues. Uses of electrophoresis in clinical aspects of genetics, virology, and microbiology are also reviewed. Veterinary medicine too is included as this now relies increasingly on this methodology.

Consideration will be given to techniques requiring simple equipment that is inexpensive and readily available in most biomolecular science laboratories, as well as to the more sophisticated procedures that require highly expensive equipment that is available in only a few relatively specialized departments. Although electrophoretic analysis of nucleic acids also continues to be of considerable importance in clinical studies, this review will be concerned mainly with proteins, electrophoresis of nucleic acids having been considered in a separate article.

## Sample Preparation

In the electrophoretic analysis of biological molecules, it is essential to ensure as far as possible that no degradation or modification has occurred between sample collection and analysis. Proteolytic cleavage, deglycosylation, or other degradation by endogenous or bacterial enzymes will give misleading results; cleavage of a single molecule could lead to erroneous conclusions. Samples should, therefore, be stored on ice or frozen immediately after collection (unless freezing also causes changes) and where necessary inhibitors of degradative enzymes or bacterial growth should be added. Samples must not be collected into formal saline or fixed in any way prior to analysis. Because of the sensitivity of some staining procedures, it is also essential to ensure that samples are not contaminated with bacteria (especially in the case of saliva and urine) and that samples are collected into clean sterile containers, otherwise some of the components revealed may originate from contaminant bacteria. The human body is subject to marked circadian variation in its metabolism; accordingly samples should be collected at the same time of day unless changes arising from this variable are known not to occur. Finally, it is essential to ensure that the same size of sample load is applied in all cases where electrophoretic patterns are to be quantified or compared. Unless this is rigorously adhered to, a difference in the presence of a component or its intensity may reflect not a true variation but merely a difference in sample load. Recent improvements in the sensitivity of protein detection methods have largely precluded the need for sample concentration prior to analysis

but when necessary this can be achieved by dialysis followed by lyophilization, ultrafiltration, or acetone precipitation. Problems may still arise from the presence of a component in high concentration masking the component of interest. This can be overcome by removal of the component causing problems by the use of antibody or other affinity-based techniques.

## Technical Considerations

Electrophoretic analysis of clinical samples is usually carried out in a solid support matrix or in liquid systems (capillary electrophoresis). Cellulose acetate has been largely superseded by polyacrylamide gels and agarose gels as a solid support matrix. Although polyacrylamide is probably the most widely used gel medium, the pouring and polymerization of gels can be tedious. Nowadays, an increasingly wide variety of readymade gels is commercially available. Whilst they may be more expensive, they give very reproducible results and avoid the tedium of gel preparation.

Electrophoresis of proteins can be carried out in native (nondenaturing) or denaturing conditions. In the former, the species are separated as a function of charge. In denaturing conditions, usually facilitated by boiling with sodium dodecyl sulfate (SDS) in the presence or absence of a reducing agent such as dithiothreitol, proteins are generally separated as a function of molecular weight. However, whilst this is true for many proteins, unusual amino compositions or post-translational modifications (e.g., glycosylation) can markedly change the migration rate in denaturing gels. Accordingly, this technique is not always a valid indicator of the molecular weight of a protein, especially in the case of clinical samples where many of the proteins may be unusual in some way.

Whilst nucleic acids are normally detected by means of ultraviolet fluorescence in the presence of ethidium bromide or by autoradiography of radio-labeled components, proteins are generally detected by staining procedures or immunoblotting. Coomassie Brilliant Blue (G-250 or R-250) is probably still the most popular protein stain and destaining is relatively rapid. The sensitivity of protein staining can be increased 10–1000-fold by the use of silver-based staining methods. The sensitivity of the staining of proteins by silver-based procedures varies substantially; glycoproteins tend to stain yellow orange and some proteins may not stain at all. More recently, fluorescent stains have become increasingly popular, especially the SYPRO<sup>®</sup> protein stains.

Automated destaining gives more reproducible results than manual. A wide variety of equipment is now available for computerized scanning of stained gels and recording of results. This also enables the storage of vast amounts of information and comparison of gels. Because of the variety of proteins in clinical samples, however, quantitative densitometric measurements on stained components should initially be interpreted with caution. The sensitivity and specificity of protein staining can be increased still further by the use of immunoblotting (Western blotting) and amplification procedures. This involves the probing of gels with a specific antibody, which is then visualized enzymatically or by means of a second radiolabeled antibody.

Although one-dimensional procedures are the easiest and quickest and involve the least sophisticated equipment, substantially more information is produced by two-dimensional procedures. Indeed, it is two-dimensional analysis that has yielded so much new information about the composition of human body fluids and tissues. Successful two-dimensional techniques usually involve separation in the first dimension as a function of charge and in the second as a function of molecular weight. Although separation in the first dimension has generally involved isoelectric focusing in carrier ampholytes, much improved resolution can be obtained by means of immobilized pH gradients, especially at alkaline pH values. Two-dimensional gels of body fluids and cell extracts can reveal several hundred proteins, the complexity of which is such that computer analysis is essential for proper interpretation of the results. It is this technique that has formed the basis of proteomics and facilitated the setting up of databases on the proteins of human body fluids and tissues. With one-dimensional techniques, it cannot be assumed that a single band is a pure component and resolution in a second dimension must always be carried out in order to assess the true degree of purity. The amino acid sequence of pure components from electrophoresis gels can be determined by blotting onto membranes such as poly(vinylidene difluoride) (PVDF), staining, and carrying out sequence analysis on the stained membrane. However, peptides from proteins of interest are now mainly identified by mass spectrometry and the protein characterized by matching these peptides to already characterized proteins in internationally available databases.

Capillary electrophoresis is also now increasingly used in clinical studies. The procedure involves rapid separation in a liquid or gel medium in a microcapillary and requires fairly expensive equipment. Its main advantages are speed of analysis, high sample

throughput, low running costs, versatility, and small sample size (only a few microliters), all of which make it an attractive technique in clinical studies. It can be automated for the analysis of large numbers of samples and the results can be computer recorded and analyzed. Although originally it was best suited for the study of small molecules and ions, its versatility has been extended to include proteins and nucleic analysis and it is now used widely for routine analysis in clinical laboratories. With the advent of microchip electrophoresis, high-resolution separation of microliter samples can be performed in a few hundred seconds.

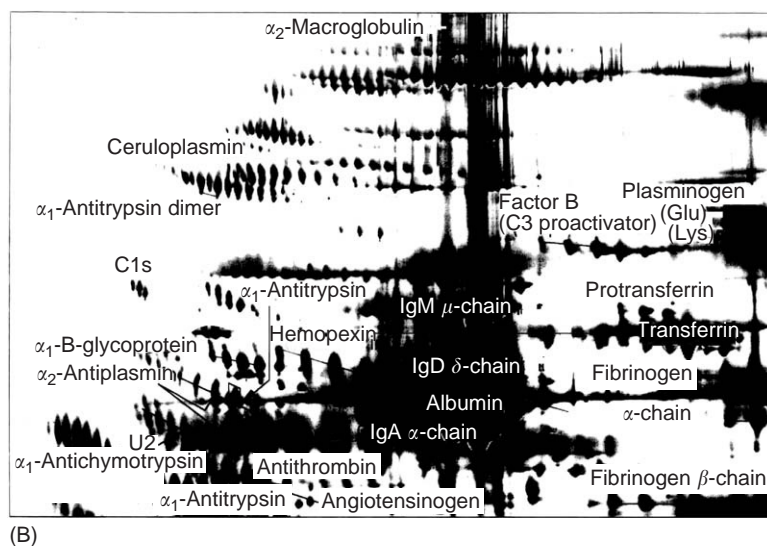
## Human Body Fluids

### Plasma/Serum

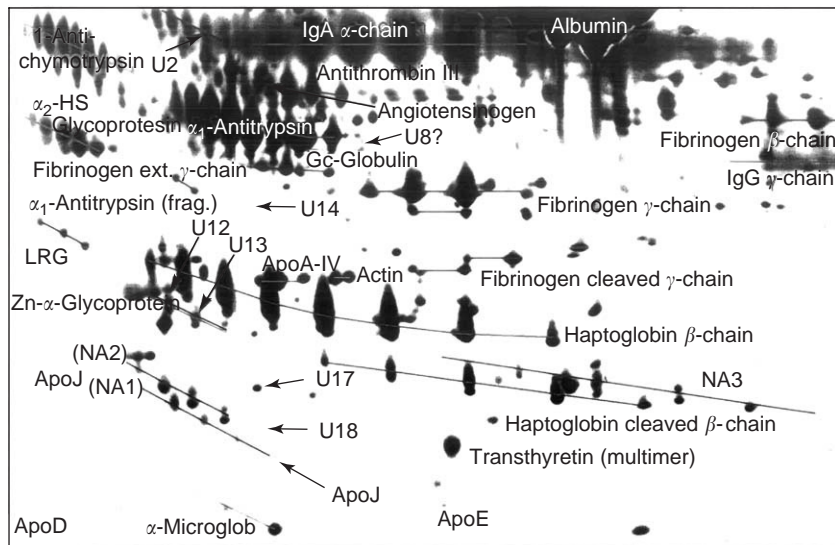
Quantitative and qualitative analysis of plasma and serum proteins forms an essential part of the clinical diagnosis of an increasing number of diseases. Electrophoretic separation of these proteins is an integral part of such analyses. Because of its role in tissue-to-tissue transport, changes in plasma protein composition may reflect metabolic disorders in all parts of the body.

One-dimensional analysis using discontinuous buffers in polyacrylamide gel electrophoresis (PAGE) results in 30 or more bands, and a similar number of components can be identified on crossed immunoelectrophoresis. Resolution can be increased to ~60 components by the use of PAGE on gradient gels or isoelectric focusing. The major components are  $\beta$ -lipoproteins,  $\alpha_2$ -macroglobulins, transferrin, albumin,  $\gamma$ -globulins, and paraproteins, with immunoglobulin abnormalities probably being diagnostically the most important. Electrophoresis is used routinely to detect monoclonal gammopathies (paraproteins) in both serum and urine and there are automated systems designed for this purpose. The true complexity of plasma protein composition only becomes properly apparent on two-dimensional analysis. Using isoelectric focusing in the first dimension and SDS gradient-PAGE in the second, followed by silver staining and blotting onto PVDF membranes, ~4000 polypeptides and proteins have now been identified with many, including immunoglobulins, consisting of multiple spots differing in glycosylation or amino acid sequence (**Figure 1**). Spots are characterized by their isoelectric point (pI) and relative molecular mass ( $M_r$ ) in a database or more recently by peptide analysis and mass spectrometry. A high proportion has been identified as known proteins or their derivatives. Despite the complexity of the pattern, which consists of several thousand spots, certain modifications of this protein pattern can be





**Figure 1** Human plasma proteins. (A) Two-dimensional PAGE of human plasma proteins using a pH gradient of 3.5–10.0 in the first dimension and silver stained. Note the focusing of basic proteins as well as the high relative-molecular-mass proteins ( $M_r$ ). (B) Enlargement of the high  $M_r$  area from (A) and (C) the more acidic medium  $M_r$  region from (A). (Reproduced with permission from Hughes GJ *et al.* (1992) Human plasma protein. *Electrophoresis* 13: 707–714; © Wiley-VCH.)



(C)

**Figure 1** (Continued).

associated with specific pathological disorders. The modifications include the presence or absence of components, decreases or increases in intensity of components, or changes in microheterogeneity. The disorders included monoclonal gammopathies, hyper- and hypo- $\gamma$ -globulinaemia, hepatic failure, chronic renal failure, haemolytic anaemia, etc. Monoclonal and polyclonal gammopathies can be differentiated easily on two-dimensional analysis. The literature on detailed changes in plasma/serum proteins associated with a variety of pathological disorders is extensive. The procedures used also involve, for example, zymograms to study isozymes as well as protein staining and antibody-based detection procedures. Capillary zone electrophoresis of serum proteins is now becoming an attractive alternative to agarose- and acrylamide-based techniques.

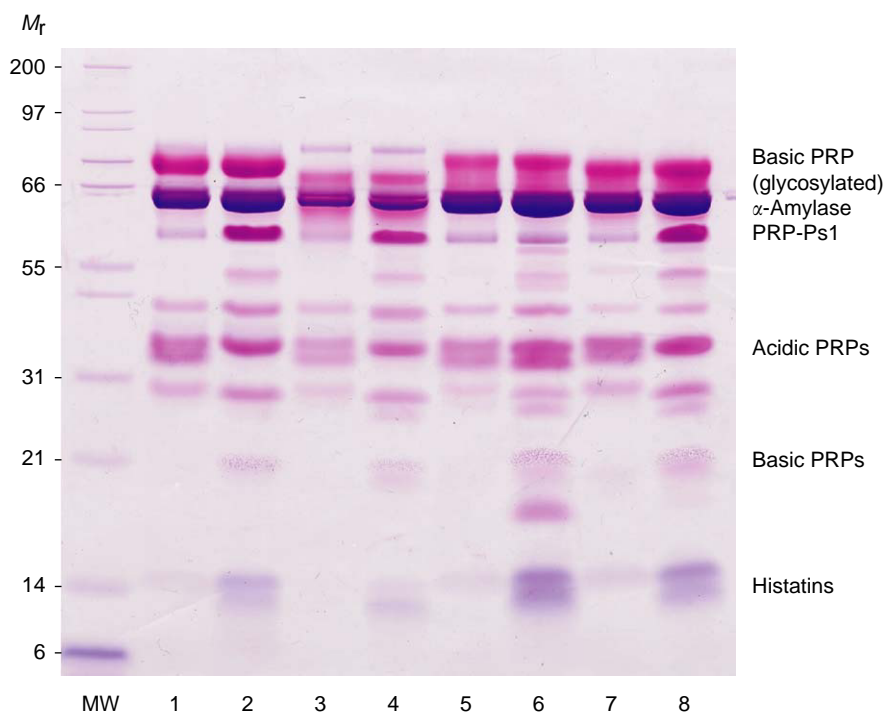
### Saliva

Human saliva is secreted by three pairs of major glands (parotid, submandibular, and sublingual) and numerous minor ones (labial, buccal, palatine, and glossopalatine). Whilst the parotid glands are the major contributors to stimulated saliva, in the resting state much of the fluid present in the mouth is the viscous submandibular/sublingual secretion, together with the products of the minor glands. Mixed or whole saliva (oral fluid) consists of the secretions of the major and minor glands together with gingival exudate (crevicular fluid), microorganisms, and cell debris. The composition of saliva is modulated on stimulation and subject to circadian variation. The

major proteins in duct saliva are  $\alpha$ -amylase and proline-rich proteins, the latter being comprised of three groups, acidic, basic, and glycosylated. These, together with histidine-rich polypeptides (histatins) and mucins, constitute  $\sim 90\%$  of the protein in parotid duct saliva (**Figure 2**). Antibacterial proteins (lactoferrin, lysozyme, lactoperoxidase, secretory-IgA, etc.), statherin, cystatins, secretory component, and proteins derived from serum (albumin, IgG) comprise the remainder. The major proteins in human saliva are unusual in that they are polymorphic families of proteins that are well suited to separation by electrophoresis. Proline-rich proteins, the functions of which are still unclear, can easily be separated on SDS-PAGE, followed by staining with Coomassie Blue R250 by which they are stained pink-violet. Histidine-rich polypeptides (histatins), which are antifungal and may constitute a non-humoral defense mechanism, are also polymorphic and can be separated by gel or capillary electrophoresis. Because of its viscosity, submandibular saliva is more difficult to collect but electrophoresis is beginning to produce useful new information (**Figure 3**).

Although salivary proteins can be resolved by denaturing or nondenaturing electrophoresis, proline-rich proteins are difficult to detect and currently the only satisfactory method for their analysis is SDS-PAGE. Early attempts to analyze salivary proteins by two-dimensional electrophoresis were unsuccessful, partially because the samples were bacterially contaminated and partially because the pI of the basic proline-rich proteins is too high (pI  $> 8$ ) to remain on





**Figure 2** Human parotid saliva. SDS-PAGE of parotid saliva from eight different individuals (1–8) followed by staining with Coomassie Blue R-250, proline-rich proteins (PRPs) stain pink. MW: molecular weight standards. There are considerable subject-to-subject variations in protein profile; these reflect genetic polymorphisms.

the gel when carrier ampholyte isoelectric focusing is employed in the first dimension. Introduction of immobilized pH gradients in the first dimension has overcome this problem and two-dimensional maps of the major antigens in human parotid saliva have now been produced. Clinical applications of electrophoresis of human salivary protein have been complicated by the fact that all of the major proteins are polymorphic families of proteins that display substantial individual to individual genetic variation.

Connective tissue disorders such as rheumatoid arthritis and Sjögren's syndrome are associated with lymphocyte infiltration of the salivary glands and two-dimensional gels have revealed substantially increased levels of a heterogeneous group of proteins in many of these individuals. Western blots have demonstrated the presence of antibodies to HIV antigens in mixed saliva. One of the basic proline-rich proteins appears to have anti-HIV activity. Although routine clinical applications of electrophoresis of saliva are still limited, this appears to be an area in which there will be considerable developments in the future.

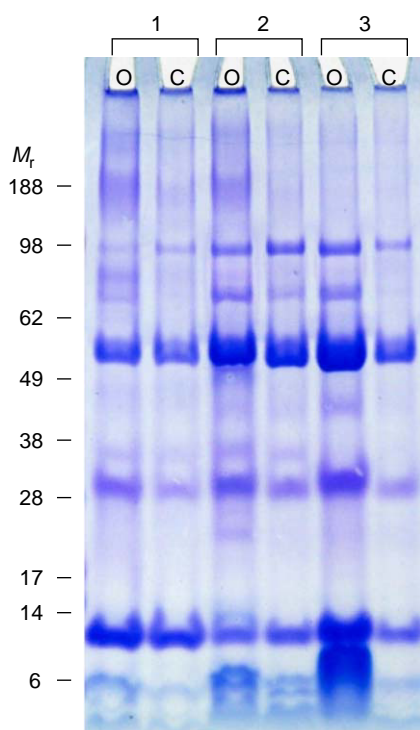
### Cerebrospinal Fluid

The choroid plexus produces ~70% of the CSF, the remainder being produced by the brain and spinal

chord parenchyma. Normally, 80% of the proteins are derived from plasma by diffusion across the blood–CSF barrier and the remainder are synthesized locally. The analysis of CSF protein by electrophoresis is of considerable diagnostic value in neurology. The protein concentration range is ~0.05–0.6 g l<sup>-1</sup>; accordingly, samples may need concentration prior to analysis.

In many cases, a substantial amount of useful information about the protein profile of CSF and any abnormalities can be obtained on one-dimensional electrophoresis. Agarose gels give good results and are easy to prepare. SDS-PAGE gives higher resolution and isoelectric focusing gives even better resolution, especially in the case of immunoglobulins. The gels required are now commercially available. Because some of the oligoclonal bands have high pIs, systems involving immobilized pH gradients give even better results. As with most body fluids and tissue extracts, the time, nature, and complexity of CSF proteins is only properly revealed on two-dimensional analysis, with several hundred or more spots having been reported.

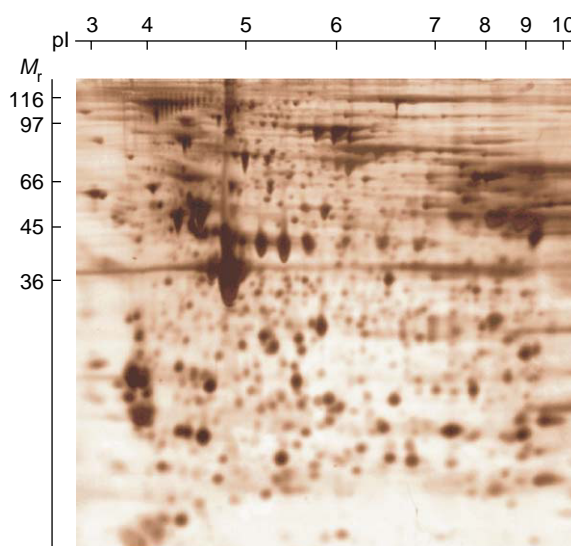
The proteins in normal CSF are similar to those in serum but at lower concentrations. Albumin is the most abundant. Any defect in the blood–brain or blood–CSF barrier leads to an increase in the level of serum proteins. CSF-specific proteins are also



**Figure 3** Human submandibular saliva. This gel shows the proteins in human submandibular saliva separated on a 4–12% Bis–Tris NuPAGE gel and stained with Coomassie Brilliant Blue R250. Samples 1, 2, and 3 are from three different subjects, in which the sample in track ‘O’ is from a gland from which obstructive calculi had been removed 6 months previously and ‘C’ is the contralateral normal control. Equal volumes of sample were applied to each track. On recovery from obstruction, the protein profile is similar to the controls, but the increased band intensity results from a higher protein concentration. (Courtesy of Proctor GB and Osailan SM, Kings College, University of London.)

present, the levels of which may be increased, decreased, or altered in neurological disorders. Immunoglobulins are not normally synthesized in CSF; they enter by diffusion, constitute ~10% of the total protein, and are predominantly IgG. In inflammatory conditions, lymphocytes migrate into the CNS leading to increased levels of IgG, which are readily resolved by isoelectric focusing. In multiple sclerosis, structurally restricted IgG of reduced heterogeneity is produced in the CSF.

Diagnostic studies on oligoclonal IgG, however, are better carried out by two-dimensional analysis. Characteristic changes in CSF proteins derived from the CNS have been associated with CNS disorders such as amyotrophic lateral sclerosis and Huntington’s chorea. Increased  $\alpha_2$ -haptoglobin levels are associated with Alzheimer’s disease and schizophrenia and an increase in apolipoprotein E with low back pain. The establishment of two-dimensional databases of CSF proteins will be of considerable



**Figure 4** Human platelet proteome. Proteins from human platelets (100  $\mu$ g protein) were separated by two-dimensional gel electrophoresis using isoelectric focusing in the first dimension in a pH 3–10 gradient and SDS-PAGE in the second using a 10% 1.5 mm gel followed by silver staining. The relative molecular masses ( $M_r$ ) and isoelectric points of the proteins (pI) separated are indicated. (Courtesy of Maguire PB, Royal College of Surgeons of Ireland.)

assistance in the further development of clinical analysis of CSF.

### Urine

Although it was more than two centuries ago that proteinuria was associated with various diseased states, only since the advent of modern electrophoretic analysis has the potential diagnostic information on the proteins present in urine become apparent. Normal urine has a very low protein content, which necessitates concentration and dialysis of samples prior to analysis. This may lead to some protein loss prior to electrophoretic analysis by some of the earlier procedures. Microscale SDS-PAGE techniques followed by silver staining, however, do not require prior sample concentration and preparation and in some cases of proteinuria dilution may be necessary. One microliter samples of normal urine (protein concentration  $<0.01 \text{ mg ml}^{-1}$ ) reveal one minor band corresponding to albumin. Orthostatic proteinuria is associated with increased excretion of glomerular proteins (albumin, transferrin, IgG), whilst in postrenal haematuria, the additional proteins are similar to those of blood haemolysate, all of the plasma proteins being present. Other postrenal proteinurias are associated with increased levels of proteins such as IgG, monoclonal gammopathies, changes in Tamm–Horsfall glycoprotein, and the

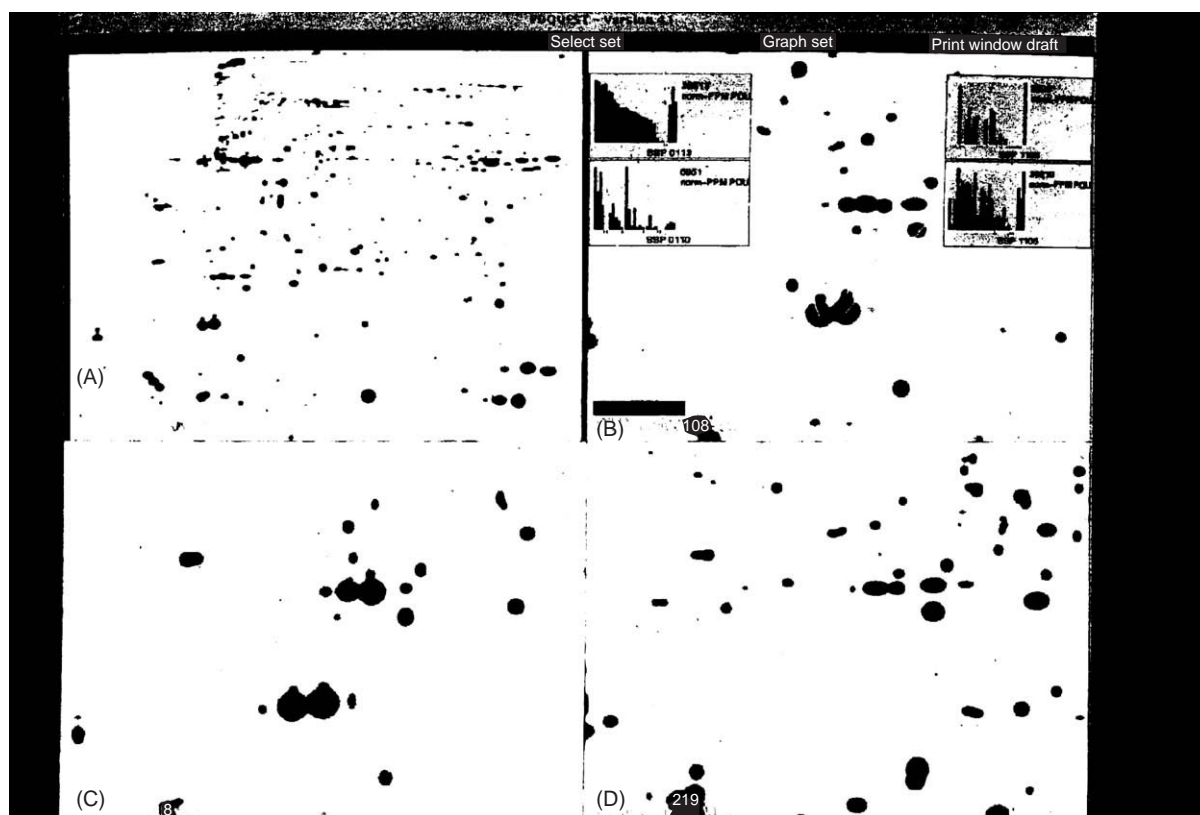
presence of Bence-Jones proteins. The presence of  $\alpha_1$ -microglobulin and increased levels of albumin and transferrin would be indicative of renal complication. Lower and upper urinary tract infections (cystitis and pyelonephritis) can also be associated with increased levels of specific proteins in urine, the nature of which is indicative of the site of infection. In type 1 diabetes mellitus, microalbuminuria occurs in the early stages of nephropathy and the changes in the nature of the proteinuria that occur as the disease progresses can be conveniently monitored by electrophoretic analysis. Electrophoresis (both gel based and capillary) is now used for the analysis of nephrolithiasis-related compounds. Accordingly, micro-SDS-PAGE now affords a fast, simple, and reliable technique for the routine clinical analysis of urine samples without any prior treatment. However, capillary electrophoresis, which offers many advantages over conventional procedures, also gives high resolution of urinary proteins and is becoming increasingly used for clinical analysis of this fluid also.

### Tears (Lacrimal Fluid)

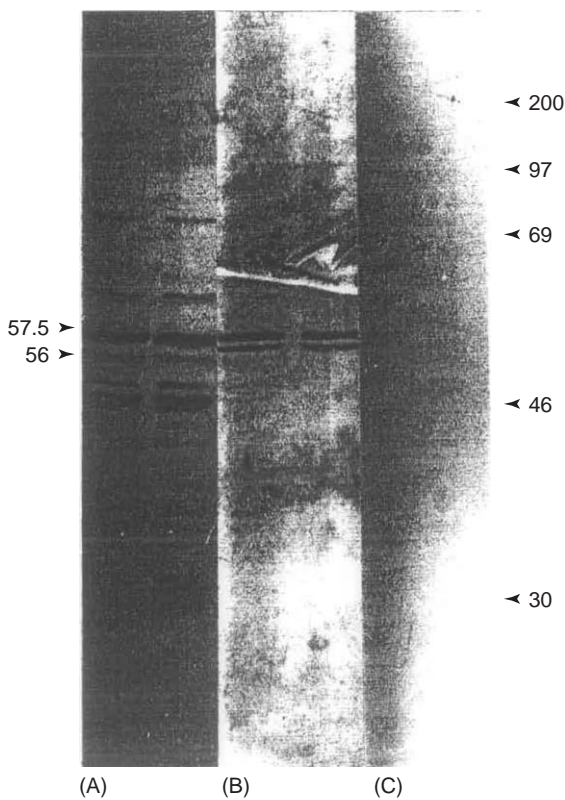
Two-dimensional electrophoretic analysis of tear proteins followed by immunoblotting has shown that the major components are lactoferrin, lysozyme, albumin, secretory IgA, and five tear-specific proteins. There are significant differences in the amounts of lactoferrin and two kinds of tear-specific proteins differ between the sexes. A comparison of the two-dimensional electrophoretic patterns of tear proteins from patients with conjunctivitis with normal individuals shows reductions in the levels of IgA, lactoferrin, and tear-specific proteins whilst the intensity of albumin is increased; in addition, haptoglobin and IgG are present in the patient samples.

### Cells and Tissues

Although electrophoresis of body fluids can produce valuable clinical diagnostic information, changes



**Figure 5** Human heart proteins. Computer analysis of two-dimensional gels to characterize changes in protein expression associated with human heart disease. (A) Reference two-dimensional gel of myocardial proteins from human ventricle. (B) Enlargement of a region of the pattern in (A) to show variation in the quantitative expression of myosin light chain 2 (MLC2) proteins (arrows) associated with dilated cardiomyopathy (DCM); in the histogram normal samples are shown on the right. (C) Region of gel showing normal pattern of expression of MLC2 and (D) showing absence of MLC2 in DCM. (Courtesy of Dunn MJ, Corbett LM, and Yacoub MH, Human heart proteins: 2D gel showing changes associated with human heart disease. National Heart & Lung Institute, Harefield Hospital, London, used by permission.)

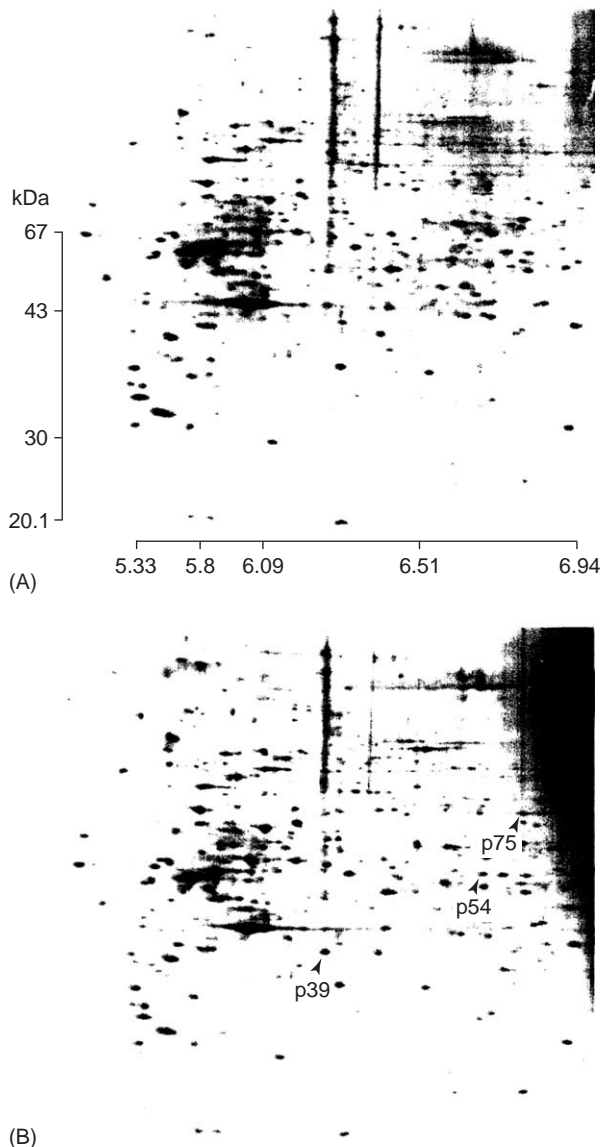


**Figure 6** Antiendothelial antibodies associated with coronary heart disease. Human umbilical vein endothelial cell proteins were separated by SDS-PAGE and transferred electrophoretically onto nitrocellulose membranes that were then probed with human serum samples. (A) Coronary heart disease patient with several endothelial antigens, (B) serum reactive with only the 57.5 and 56 kDa doublet, and (C) normal serum showing no bands of reactivity. (Reproduced with permission from Dunn *et al.* (1983) Antiendothelial antibodies associated with coronary heart disease. *Lancet* 339: 1566–1570.)

in the composition of these fluids merely reflects metabolic changes in the tissues from which the components originate. Tissue and cell analysis, therefore, offers the potentiality of being more specific in pathological investigations and characteristic patterns have been associated with them. High-resolution two-dimensional-PAGE of tissue extracts results in over 2000 proteins being separated from only a few micrograms of tissue, and some of the patterns produced are disease associated. **Figure 4** shows the proteome of a human platelet. Databases have been generated for most body cells and tissues. Two areas of specific interest in tissue analysis are cancer and heart disease.

### Cancer Cells

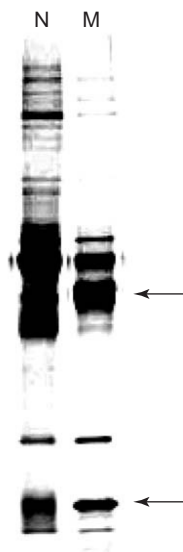
Whilst alterations in cancer cells are likely to be at the DNA or chromosomal level, proteome analysis of



**Figure 7** Protein profiles of virally infected cells. Two-dimensional protein profiles of rhabdomyosarcoma cells (A) uninfected and (B) infected with Cocksackie B5 virus. The proteins were labeled and located by autoradiography. The three proteins p39, p54, and p75 are specific for the infected cells. (Reproduced with permission from Argo E, Gimenez B, and Cash P (1992) Protein profiles of virally infected cells. *Archives of Virology* 126: 215–229; Vienna: Springer-Verlag.)

tumor cells and body fluids is now widely used in cancer research as a tool for the identification of tumor markers and the development of molecular diagnostic markers for its early detection. The technique complements other aspects of cancer research and produces information that cannot be obtained otherwise. Because of the complexity of the polypeptides in whole cell extracts, analysis of subsets of proteins (e.g., membrane, cytoplasmic, nuclear) not



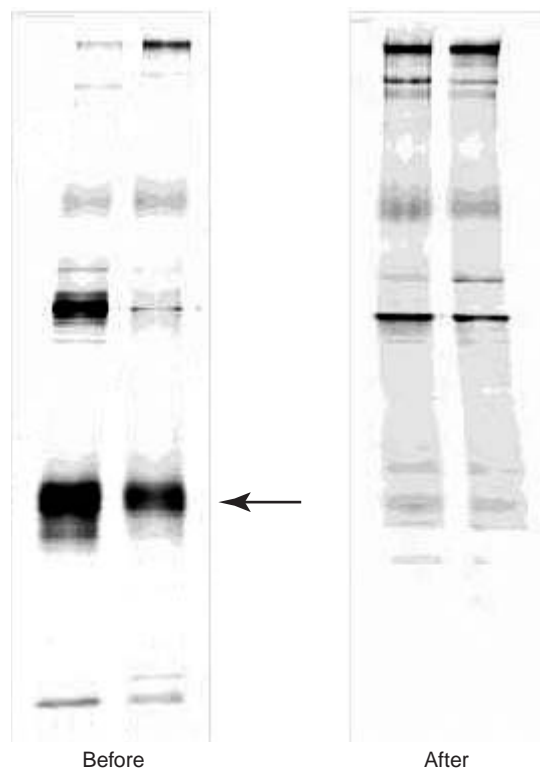


**Figure 8** Canine serum. One-dimensional gel electrophoresis of dog serum separated on SDS-PAGE followed by silver staining. Track 'N' is from a normal healthy dog and track 'M' is from a dog with a monoclonal gammopathy (IgA); Prominent H and L chains can be clearly seen. (Courtesy of Miller I, University of Veterinary Medicine, Vienna.)

only simplifies the results but can also facilitate detection of proteins that are below the levels of sensitivity in whole cell preparations. When comparing transformed and nontransformed cells qualitatively, there can be remarkable similarity in their polypeptide patterns but quantitatively there are substantial variations. The procedure can also be applied in the detection of lineage markers in cancer cells and the study of polypeptides secreted by them, as well as monitoring the progress of the disorder.

### Cardiac Disease

Databases have been produced for myocardial tissue proteins in order to elucidate the nature of protein expression in normal and disease conditions. **Figure 5** shows a computer-generated image of a two-dimensional-PAGE separation of human heart proteins with regions showing alterations in the pattern that are associated with heart disease. The most serious complication of cardiac transplantation is accelerated coronary artery disease that may result from endothelial cell damage mediated by antiendothelial cell antibodies produced as a result of the transplant. Western blotting of human umbilical cord endothelial cells with sera from post-transplant patients with coronary artery disease showed the presence of antiendothelial cell antibodies that were rare in post-transplant patients without coronary artery disease (**Figure 6**).



**Figure 9** Canine urine. One-dimensional SDS-PAGE electrophoresis of the proteins in dog urine from two different animals before and after castration. After castration the pattern is similar to that of females. Differences in levels of the male-specific protein (indicated by an arrow) may be associated with different prostate disorders. (Courtesy of Miller I, University of Veterinary Medicine, Vienna.)

### Genetics

Electrophoresis has facilitated substantial developments in the study of genetics ranging from one-dimensional electrophoresis of enzyme polymorphisms to the use of DNA electrophoresis in the human genome project, which culminated in the sequencing of the entire human genome. Gel electrophoresis of DNA fragments is now widely used in the investigation of genetic defects that affect human health and has led to substantial developments in the understanding of inherited disease at a molecular level. Mutations in the base sequence also lead to changes in the size of fragments formed on degradation by restriction nucleases. Where appropriate, specific genes (DNA fragments) can be targeted and amplified by the polymerase chain reaction (PCR).

Electrophoretic separation of proteins (the gene products and their post-translational modifications) continues to be an essential tool in the study of genetic polymorphisms and inherited diseases. Detection methods for the proteins of interest include

protein staining, zymograms, Western blotting, and isotopic labeling. In many cases, the product of the mutant gene is unknown. In such cases it is now possible to clone the gene involved and characterize its product, i.e., 'reverse genetics'.

## Virology

Molecular techniques are now being used increasingly in the study of viral infections. They are especially useful either as a supplement to existing serological techniques or as an alternative when such techniques do not exist.

Electrophoretic analysis of viral nucleic acid has played a major role in investigating and characterizing viruses as well as bacteria. Analysis of virally induced proteins in one-dimensional SDS-PAGE has been used to differentiate clinical isolates of certain viruses; one-dimensional IEF has been used to differentiate foot-and-mouth disease virus isolates from cattle. As in all cases of one-dimensional analysis of proteins, two-dimensional PAGE gives improved resolution and portrays the vast numbers of proteins that are present in the sample. Accordingly, analysis of proteins from virally infected cells has facilitated the identification of proteins characteristic of certain viral infections and yielded new information about gene expression in viruses including foot-and-mouth disease, polio, and Cocksackie B viruses. **Figure 7** shows some of the differences in protein profile of a Cocksackie B infected cell compared with an uninfected one.

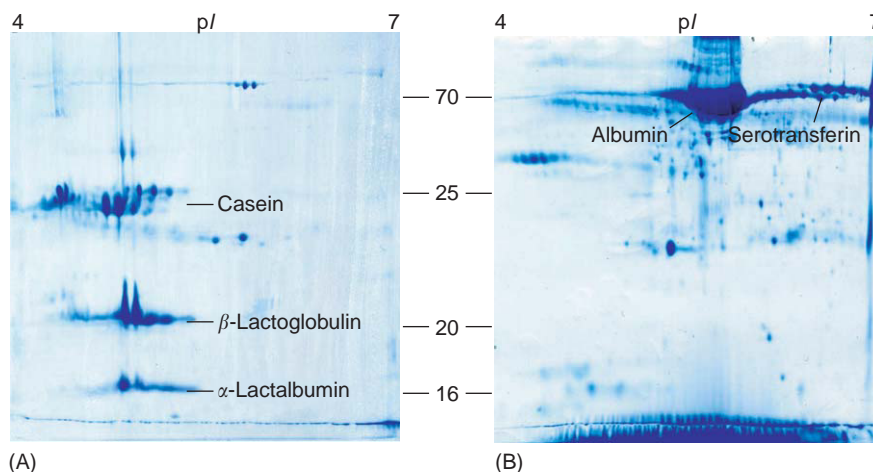
## Microbiology

The majority of bacterial typing in the past has been based on phenotypic or readily observable characteristics. Although these procedures are still the method of choice in many cases in the routine laboratory analysis, typing procedures based on electrophoretic analysis of nucleic acids, especially PCR and gene probing, are becoming increasingly used in bacterial identification.

Electrophoretic analysis of bacterial proteins also produces a pattern that is characteristic of the particular bacterial strain and bacterial proteomes are now well documented. This technique now plays a key role in many areas of medical microbiology including gene expression, epidemiology, taxonomy of pathogens, and drug resistance. Numerous strains of bacteria have now been studied by this procedure. Capillary electrophoresis is also gaining in popularity for the analysis, identification, and characterization of microorganisms. The advantages include small sample size, ease of quantification, and identification and rapid simultaneous analysis.

## Veterinary Clinical Studies

Advances in laboratory-based veterinary clinical studies have been accompanied by the increased use of electrophoresis-based techniques both in the routine clinical and research laboratory. Variations in the composition of serum, urine, and other body fluids and tissues of animal origin are all possible markers of health and disease. They offer



**Figure 10** Bovine milk proteins. Two-dimensional PAGE analysis of proteins from bovine milk: (A) normal milk and (B) milk from animals with mastitis. The proteins were separated in the first dimension by isoelectric focusing in a pH gradient of 4–7 and in the second dimension by SDS-PAGE electrophoresis in a 12% gel followed by staining with Brilliant Blue G colloidal stain. In the mastitic milk, most of the proteins (casein,  $\beta$ -lactoglobulin, and  $\alpha$ -lactalbumin) have been lost, and the major proteins are now of serum origin (albumin, serotransferrin). (Courtesy of Hogarth CG, University of Glasgow.)



considerable potential not only in animal welfare and care, but are also of growing economic importance in livestock management. Animal serum proteomes are now being established and serum protein reference maps have been established for cows, dogs, rats, and pigs. Currently, studies are largely focused on the inflammatory process and gammopathies, where serum analysis can yield valuable information (**Figure 8**). Other body fluids (e.g., urine) are also revealing interesting new findings (**Figure 9**) and the technique may also be of value in the early diagnosis of mastitis (**Figure 10**).

Although one-dimensional electrophoretic analysis of proteins continues to produce valuable information in the clinical laboratory, 2D-analysis of proteins, computer analysis of gels, mass spectrometry based identification procedures, proteomics and the formation of international databases, which are readily accessible on websites have opened up a new dimension in clinical diagnostic procedures. With the completion of the human and bacterial genome projects, although nucleic acid based studies, will continue to be high profile in clinical studies, future developments are likely to be centered increasingly on gene products and their post-translational modifications, i.e., proteins.

*See also:* **Electrophoresis:** Nucleic Acids. **Mass Spectrometry:** Peptides and Proteins.

## Further Reading

- Al-Hashimi I, Haghighat N, and Fox PC (1998) Salivary electrophoresis in the diagnosis of Sjögren's syndrome. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontics* 85(5): 542–547.
- Cash P (2000) Proteomics in medical microbiology. *Electrophoresis* 21(6): 1187–1201.
- Decker J and Reischl U (2004) Molecular diagnosis of infectious diseases. In: Walker J (ed.) *Methods in Molecular Medicine*, 2nd edn. New Jersey: Humana Press.
- Desai MJ and Armstrong DW (2003) Separation, identification and characterization of microorganisms by capillary electrophoresis. *Microbiology and Molecular Biology Reviews* 67(1): 38–51.
- Dwek MV and Rawlings SL (2002) Current perspectives in cancer proteomics. *Molecular Biotechnology* 22(2): 139–152.
- Gromov PS, Ostergaard M, Gromova I, and Celis JE (2002) Human proteomic databases: a powerful resource for functional genomics in health and disease. *Progress in Biophysics and Molecular Biology* 80(1–2): 3–22.
- Hanash SM (2000) Biomedical applications of two-dimensional electrophoresis using immobilized pH gradients: current status. *Electrophoresis* 21(6): 1202–1209.
- Jain KK (2002) Role of proteomics in diagnosis of cancer. *Technology in Cancer Research and Treatment* 1(4): 1–6.
- Landers JP (2003) Molecular diagnostics on electrophoretic microchips. *Analytical Chemistry* 75(12): 2919–2927.
- Marshall T and Williams KM (1998) Clinical analysis of human urinary proteins using high resolution electrophoretic methods. *Electrophoresis* 19(10): 1752–1770.
- Peterson JR, Okorodudu AO, Mohammad A, and Payne DA (2003) Capillary electrophoresis and its applications in the clinical laboratory. *Clinica Chimica Acta* 330 (1–2): 1–30.
- Rhigheiti PG, Castagna A, Antonucci F, et al. (2002) The proteome: anno Domini 2002. *Clinical Chemistry and Laboratory Medicine* 41(4): 425–438.
- Westermeyer R (2001) *Electrophoresis in Practice*, 3rd edn. Weinheim: Wiley-VCH.
- Williams LR (2001) Staining nucleic acids and proteins in electrophoresis gels. *Biotechniques and Histochemistry* 76(3): 127–132.

## Nucleic Acids

**C M Roth**, Rutgers University, Piscataway, NJ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Electrophoresis is an essential technology for the separation and analysis of nucleic acids. Electrophoresis of nucleic acids is used routinely at the lab bench for the isolation and manipulation of cloned DNA fragments. In addition, it is a critical component of many molecular biology protocols that assess the role and interaction of nucleic acids in cells and tissues. Nucleic acid electrophoresis has attained particular importance in the current era of genome

sequencing. In this application, speed and accuracy of nucleic acid analysis are of vital scientific and economic interest, and the practice of nucleic acid electrophoresis has evolved to reflect these needs. In this article, the physicochemical aspects of nucleic acid electrophoresis are outlined and the major techniques utilized for the electrophoresis of nucleic acids are described.

## Nucleic Acid Structure and Electrophoretic Mobility

Due to their phosphate groups, nucleic acids are negatively charged at all but the most acidic of pH

considerable potential not only in animal welfare and care, but are also of growing economic importance in livestock management. Animal serum proteomes are now being established and serum protein reference maps have been established for cows, dogs, rats, and pigs. Currently, studies are largely focused on the inflammatory process and gammopathies, where serum analysis can yield valuable information (**Figure 8**). Other body fluids (e.g., urine) are also revealing interesting new findings (**Figure 9**) and the technique may also be of value in the early diagnosis of mastitis (**Figure 10**).

Although one-dimensional electrophoretic analysis of proteins continues to produce valuable information in the clinical laboratory, 2D-analysis of proteins, computer analysis of gels, mass spectrometry based identification procedures, proteomics and the formation of international databases, which are readily accessible on websites have opened up a new dimension in clinical diagnostic procedures. With the completion of the human and bacterial genome projects, although nucleic acid based studies, will continue to be high profile in clinical studies, future developments are likely to be centered increasingly on gene products and their post-translational modifications, i.e., proteins.

*See also:* **Electrophoresis:** Nucleic Acids. **Mass Spectrometry:** Peptides and Proteins.

## Further Reading

- Al-Hashimi I, Haghighat N, and Fox PC (1998) Salivary electrophoresis in the diagnosis of Sjögren's syndrome. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontics* 85(5): 542–547.
- Cash P (2000) Proteomics in medical microbiology. *Electrophoresis* 21(6): 1187–1201.
- Decker J and Reischl U (2004) Molecular diagnosis of infectious diseases. In: Walker J (ed.) *Methods in Molecular Medicine*, 2nd edn. New Jersey: Humana Press.
- Desai MJ and Armstrong DW (2003) Separation, identification and characterization of microorganisms by capillary electrophoresis. *Microbiology and Molecular Biology Reviews* 67(1): 38–51.
- Dwek MV and Rawlings SL (2002) Current perspectives in cancer proteomics. *Molecular Biotechnology* 22(2): 139–152.
- Gromov PS, Ostergaard M, Gromova I, and Celis JE (2002) Human proteomic databases: a powerful resource for functional genomics in health and disease. *Progress in Biophysics and Molecular Biology* 80(1–2): 3–22.
- Hanash SM (2000) Biomedical applications of two-dimensional electrophoresis using immobilized pH gradients: current status. *Electrophoresis* 21(6): 1202–1209.
- Jain KK (2002) Role of proteomics in diagnosis of cancer. *Technology in Cancer Research and Treatment* 1(4): 1–6.
- Landers JP (2003) Molecular diagnostics on electrophoretic microchips. *Analytical Chemistry* 75(12): 2919–2927.
- Marshall T and Williams KM (1998) Clinical analysis of human urinary proteins using high resolution electrophoretic methods. *Electrophoresis* 19(10): 1752–1770.
- Peterson JR, Okorodudu AO, Mohammad A, and Payne DA (2003) Capillary electrophoresis and its applications in the clinical laboratory. *Clinica Chimica Acta* 330 (1–2): 1–30.
- Rhigheiti PG, Castagna A, Antonucci F, et al. (2002) The proteome: anno Domini 2002. *Clinical Chemistry and Laboratory Medicine* 41(4): 425–438.
- Westermeyer R (2001) *Electrophoresis in Practice*, 3rd edn. Weinheim: Wiley-VCH.
- Williams LR (2001) Staining nucleic acids and proteins in electrophoresis gels. *Biotechniques and Histochemistry* 76(3): 127–132.

## Nucleic Acids

**C M Roth**, Rutgers University, Piscataway, NJ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Electrophoresis is an essential technology for the separation and analysis of nucleic acids. Electrophoresis of nucleic acids is used routinely at the lab bench for the isolation and manipulation of cloned DNA fragments. In addition, it is a critical component of many molecular biology protocols that assess the role and interaction of nucleic acids in cells and tissues. Nucleic acid electrophoresis has attained particular importance in the current era of genome

sequencing. In this application, speed and accuracy of nucleic acid analysis are of vital scientific and economic interest, and the practice of nucleic acid electrophoresis has evolved to reflect these needs. In this article, the physicochemical aspects of nucleic acid electrophoresis are outlined and the major techniques utilized for the electrophoresis of nucleic acids are described.

## Nucleic Acid Structure and Electrophoretic Mobility

Due to their phosphate groups, nucleic acids are negatively charged at all but the most acidic of pH

conditions. In fact, because each base carries a phosphate moiety, the charge-to-mass ratio of nucleic acids is essentially independent of size and form (e.g., single stranded versus double stranded, linear versus plasmid). As a result, nucleic acids migrate readily in an electric field from the cathodic to anodic (negatively to positively charged) ends of the apparatus. For separation and analysis of nucleic acids, therefore, an additional discriminatory parameter is needed. Most frequently, this is accomplished by electrophoresis through a sieving medium, i.e., a gel, through which small molecules will pass more easily and larger ones will pass more slowly, due to the combined effects of sampling a greater fraction of the gel space and hindered transport (volume exclusion) effects.

For a small, spherical particles, the balance of electric and viscous forces results in an electrophoretic mobility ( $u$ , defined as the particle velocity per applied electric field strength) given by

$$u = \frac{q}{6\pi R\eta} \quad [1]$$

where  $q$  is the charge on a particle of radius  $R$  and  $\eta$  is the solvent viscosity. The formation of higher-order structures by nucleic acids will cause considerable deviation from this behavior, and indeed electrophoresis is used to ascertain the physical form of DNA. Condensed (e.g., supercoiled) DNA behaves most closely to a spherical particle, whereas linear DNA exhibits mobility that reflects its polymeric nature. That is, the DNA is thought to reptate through the pores of the gel in a wriggling, snake-like manner. However, double-strand formation and counterion condensation serve to decrease the flexibility of the nucleic acid, and as a result the reptating nucleic acid behaves more as a stiffly jointed rod in its transport through a sieving gel. A consequence of this behavior is that for high molecular weights, the 'tail' of the DNA chain no longer samples a distribution of pore space, and the mobility becomes independent of chain length. Thus, separation of very high (>60 kbp) molecular weight DNA requires other techniques, such as pulsed-field gradient electrophoresis.

Thus, under constant applied electric field, the mobility is determined primarily by the pore size of the gel matrix. In this case the so-called Ferguson plot gives the relationship between the gel concentration and the electrophoretic mobility (sieving effect). In this method, the logarithm of the electrophoretic mobility is plotted as a function of gel concentration. The intercept of this plot may be interpreted as being related to the charge on the solute and the intercept as related to its size or shape. Conversely, the separation and analysis of nucleic

acids can be tuned by selection of an appropriate matrix and composition (Table 1). Other electrophoretic parameters, such as applied field strength and temperature, have an effect on the mobility of the different chain length nucleic acids and on the separation efficiency.

## Detection

Nucleic acids in solution can be detected and quantified readily on the basis of ultraviolet (UV) absorbance. Nucleic acids exhibit strong absorbance with a peak at 260 nm in a manner that is sensitive to chemistry and bonding. For accurate quantification of small oligonucleotides, a sequence-dependent extinction coefficient can be computed, but for all other routine laboratory analysis, approximate values are almost universally employed (Table 2).

For in-gel analysis of double-stranded DNA, ethidium bromide (EtBr) is still the most widely applied dye. EtBr intercalates between base pairs of the DNA, resulting in the production of fluorescence, which can be observed either by broad UV illumination (excitation maxima at 254, 302, and 366 nm) or by laser excitation at 514 nm and detection at 590 nm. Agarose gels may be either cast in the presence of EtBr or stained with the dye solution afterward; acrylamide gels must be stained after electrophoresis as ethidium inhibits the polymerization of acrylamide. The limit of detection with EtBr is on the order of 10 ng. Because EtBr changes the physical properties of DNA, most notably its stiffness, prestaining is not appropriate for DNA in

**Table 1** Separation ranges for agarose and acrylamide gels

% Agarose	Range (kb)	% Acrylamide monomer	Range (bp)
0.3	5–60	3.5	100–1000
0.6	1–20	5.0	80–500
0.7	0.8–10	8.0	60–400
0.9	0.5–7	12.0	40–200
1.2	0.4–6	15.0	25–150
1.5	0.2–4	20.0	6–100
2.0	0.1–3		

**Table 2** Concentration of nucleic acids giving one absorbance unit (i.e., absorbance = 1.0)

Nucleic acid	Concentration for $A_{260} = 1.0$
dsDNA	50 $\mu\text{g ml}^{-1}$
ssDNA	33 $\mu\text{g ml}^{-1}$
RNA	40 $\mu\text{g ml}^{-1}$

assays designed to assess physical form (e.g., supercoiling). On the other hand, the displacement of EtBr can be used as a means to monitor the complexation of DNA with cationic polymers used for nonviral gene delivery.

The most sensitive detection of nucleic acids is accomplished by radiolabeling. Single-stranded DNA is readily 5'-end labeled using T4 polynucleotide kinase. RNA species transcribed *in vitro* or DNA generated by PCR may be radiolabeled by the inclusion of a fraction of isotopic dNTPs in the polymerase reactions. Probes to be used for Northern or Southern blotting can be labeled using a variety of methods; common ones include random priming from a DNA template, primer extension, and end-labeling using the Klenow fragment of *Escherichia coli* DNA polymerase I.

With improvements in chemical dyes and fluorescence scanning instrumentation, DNA electrophoresis is performed increasingly using fluorescent dyes as the means for detection. Single-stranded oligonucleotides are synthesized readily with fluorescent dyes at the 5'- or 3'-ends, albeit that the latter significantly reduces the yield of the synthesis. Most commonly, fluorescein derivatives such as FAM (carboxy fluorescein) are utilized, but many other dyes have been incorporated as well. Plasmid DNA can be labeled using commercial reagents that place amine or carboxyl groups onto guanine residues, which are then used to attach Cy3, rhodamine, or other dyes. As an alternative to EtBr, SYBR Gold and SYBR Green effectively bind to double-stranded DNA and to a lesser extent, single-stranded DNA. Dyes specific for single-stranded oligonucleotides (e.g., OliGreen) exist as well, but they have been tested less extensively for gel detection.

## Slab Gel Electrophoresis

Electrophoresis is most commonly performed in a slab gel format, which can be used in either the vertical or the horizontal configuration. Slab gels are relatively easy to form and provide a mechanically stable medium with an extended surface area to reduced thermal convection (though they are vastly inferior to capillaries in this regard – see following section). It is possible to excise separated bands from a gel, and thus perform batch separations using electrophoresis.

### Materials

The two materials used most commonly for DNA slab gel electrophoresis are polyacrylamide and agarose. Polyacrylamide gels are formed with pore

sizes that depend on the percentage of acrylamide used as well as the proportion of cross-linker, *N,N'*-methylene bisacrylamide. Acrylamide monomer concentrations range from 3.5% to 20%. The polymerization reaction is generally carried out directly to form a slab in a mold. Details and specific protocols are provided in the volume by Sambrook. Excellent – down to single base – resolution of nucleic acids is achieved using polyacrylamide gels, and as a result they are used for DNA sequencing.

Agarose is a polymer of D-galactose and 3,6-anhydrogalactose and is used in concentrations ranging from 0.3% to 2.0% (Table 1). The agarose gel is formed when the constituents are heated and then cooled under ambient conditions to room temperature. Agarose is used for higher molecular weight separations due to the larger pore sizes and the ability to maintain mechanical stability at low weight percentages. An additional advantage of agarose is that it can be easily dried to a thin, transparent film, which makes it quite suitable for autoradiography.

### Operations and Applications

The buffers used for electrophoresis can be categorized as nondenaturing and denaturing. Because double-stranded DNA is dominated by secondary structure rather than by high-order structure under many conditions, nondenaturing electrophoresis is sufficient for many applications. Sometimes, a denaturant such as urea is introduced when it is critical to minimize higher-order effects. This is particularly common in the analysis of oligonucleotides by polyacrylamide gel electrophoresis (PAGE) for applications such as sequencing. Single-base resolution is obtained for sequencing gels, which are run vertically.

For the separation of very large DNA molecules (>20 kb), pulsed field gel electrophoresis (PFGE) is employed. In PFGE, two sets of electrodes are used to alternately generate current in transverse directions. When the electric field changes directions, the DNA molecules must reorient themselves in the gel to begin moving in a new direction. The ability of the DNA to reorient is length dependent, forming the basis for the PFGE separation. As DNA molecules up to 10 Mb can be resolved by PFGE, this technique is finding applicability for gene–chromosome mapping in lower eukaryotes and for long-range restriction mapping in higher eukaryotes, whose complete chromosomes are too large for even PFGE separation.

Separation of nucleic acids is used routinely at the lab bench, for example, to isolate cloned DNA, to purify *in vitro* transcribed RNA, and to recover transcripts differentially expressed in disease or other conditions of interest. The identification of nucleic



acids from cells and tissues, using techniques such as Northern and Southern blotting, and reverse transcriptase-polymerase chain reaction (RT-PCR), is critically dependent upon electrophoresis. In the Northern and Southern blotting techniques, which are used for analysis of RNA and DNA, respectively, the nucleic acids are electrophoresed twice. First, they are separated by electrophoresis on an agarose gel. Then, they are transferred to nitrocellulose or nylon membranes by electrophoresis in the direction perpendicular to the gel. Detection of specific DNA or RNA molecules in the blot is performed using a complementary probe DNA that is labeled either radioactively or fluorescently as described above. The relative size, as determined by position on the gel relative to molecular weight markers, provides confirmation of the identity of the target RNA or DNA.

Electrophoretic mobility shift assays are used to characterize the binding of nucleic acids to other molecules. The traditional application is to protein-DNA interactions, in which proteins are separated in a PAGE gel and probed with a labeled double-stranded oligonucleotide with a sequence corresponding to the consensus for the DNA binding site of interest. This methodology can be applied in additional areas, most notably the characterization of polymer-DNA interactions in nonviral gene and oligonucleotide deliveries. Depending on the characteristics (e.g., chemistry, molecular weight, degree of branching) of a particular delivery polymer, the charge ratio (cationic polymer-anionic DNA) at which complex formation occurs can vary widely. This is assessed by a gel shift assay in which the disappearance of the band owing to labeled (or stained) free DNA is followed.

## Capillary and Capillary Gel Electrophoresis

The advent of high-performance capillary electrophoresis (HPCE) operating at electric field strengths of up to  $1000\text{--}1500\text{ V cm}^{-1}$  has enabled the technique to be used for the rapid multicomponent analysis of single-stranded oligonucleotides and double-stranded DNA fragments. The use of capillaries allows the heat generated by such high field strengths to be dissipated, and allows the use of sample volumes less than a microliter. The method is sensitive enough to detect differences in the secondary structure of nucleic acid molecules that are of the same size, has micropreparative capabilities, and offers the ability to detect samples online. The high efficiency and speed of this approach makes it useful as a means of assessing the purity of nucleic acid

molecules produced by a DNA synthesizer and for PCR products. The method is also seeing extensive application in DNA sequencing.

## Materials and Equipment

In HPCE, samples can be detected on-column (online). Removing the polyimide coating of the fused silica capillary in a small section allows UV or fluorescent detection. A power supply is used to apply a field strength of  $0\text{--}30\text{ kV}$ . Injections can be made by means of pressure or electrokinetic modes that allow the application of automated injection systems.

In polyacrylamide capillary gel electrophoresis, the gel is usually polymerized *in situ* in the capillary tubing. The polyacrylamide is entirely synthesized from monomers, either with or without cross-linking, depending on the application. These gel columns are employed in the size separation of biologically important polymers such as DNA or RNA. The high efficiency and speed of this approach makes it useful as a means of assessing the purity of nucleic acid molecules produced by a DNA synthesizer, for PCR products, and for DNA sequencing.

## Operations and Applications

In the size separations of oligodeoxyribonucleotides and oligoribonucleotides by capillary gel electrophoresis for chain length determination, denaturing agents should be used in the buffer system to avoid intra- and interchain interactions and therefore the formation of secondary structures. In PAGE,  $7\text{--}9\text{ mol l}^{-1}$  urea or  $30\text{--}70\%$  formamide is used as a denaturing agent. The primary sequence of the oligomers can have a significant effect on electrophoretic mobility. By changing the pH of the gel buffer, the separation power of the system can be altered.

Complexing agents, charged or uncharged, can be simply added to the gel-buffer system. Complexation can speed up or slow down the migration velocity of the solute. In the separation of double-stranded DNA fragments, an intercalator additive such as EtBr can be used as a complexing agent in order to enhance the separation power of the method. EtBr is positively charged at the pH of the running buffer; thus, it decreases the effective charge of the DNA molecule. As a result, the DNA-EtBr complex migrates more slowly under the same applied electric field, increasing in this way the separation time window and enhancing separation efficiency.

Capillary electrophoresis (CE) is now the technique of choice for analysis of nucleic acids when small sample volume and high resolution are important considerations. Large-scale sequencing efforts are conducted presently using electrophoresis in arrays

of capillaries. In addition, CE is being applied to the detection of genetic mutations through dideoxy fingerprinting and to DNA adduct formation. For efficient separations of DNA in CE, linear polyacrylamide solutions of low viscosity and coated capillary surfaces are used. These are not ideal, however, and the tailoring of polymer solutions to particular separation needs is an active area of research. In the future, the marked improvements in microtechnology promise to bring about microfabricated sequencing arrays, the proof of principle for which has already been established. This approach can be easily integrated with microfluidics to develop 'lab-on-a-chip' devices. Instruments such as the Agilent Bioanalyzer are already capable of performing rapid, quality control analyses on-chip with 1 µl or less of sample.

**See also:** **Capillary Electrophoresis:** Overview. **Electrophoresis:** Clinical Applications.

## Further Reading

- Blackburn GM and Gait MJ (1996) *Nucleic Acids in Chemistry and Biology*. Oxford: Oxford University Press.
- Cantor CR and Schimmel PR (1980) *Biophysical Chemistry. The Behavior of Biological Macromolecules*. San Francisco: W. H. Freeman and Company.
- Grodzinsky AJ and Yarmush ML (1993) Electrokinetic separations. In: G. Stephanopoulos (ed.) *Biotechnology. Volume 3: Bioprocessing*, pp. 680–692. Weinheim: VCH Verlagsgesellschaft.
- Hu S and Dovichi NJ (2002) Capillary electrophoresis for the analysis of biopolymers. *Analytical Chemistry* 74(12): 2833–2850.
- Righetti PG, Gelfi C, and D'Acunzio MR (2002) Recent progress in DNA analysis by capillary electrophoresis. *Electrophoresis* 23(10): 1361–1374.
- Sambrook J and Russell DW (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

## Proteins

**V Kašička**, Czech Academy of Sciences, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Proteins are, together with nucleic acids, the most important biomolecules. Acting as enzymes, enzyme substrates and inhibitors, hormones, receptors, structural units, antigens, antibodies, drugs, and toxins, they play a vitally important role in all living organisms. In the current era of proteomics, the comprehensive analysis of the proteome, i.e., analysis of the set of all proteins encoded by the genome of a given organism and expressed in cells, tissues, or organs at a particular time and particular conditions, which nowadays represents the main approach for a complete understanding of biological systems and new drug discovery, the importance of proteins is ever increasing.

The huge number of variations of amino acid sequences in protein molecules and variety of attached prosthetic groups result in an extraordinary diversity of proteins that differ in electric charge, size, shape, specific binding capabilities, and, consequently, in electrophoretic mobilities, which allows their separation using different electrophoretic

methods:

1. zone electrophoresis (ZE) and isotachopheresis (ITP) in free-solution media – based on different charge-to-size ratios of proteins;
2. ZE in sieving media (gels and entangled polymer networks) – based on size-dependent retardation of electromigration of proteins by the medium matrix;
3. isoelectric focusing (IEF) – based on different isoelectric points of proteins; and
4. (bio)affinity electrophoresis (BAE) – based on influencing the electromigration of proteins through specific interactions with other (bio)molecules in separation media.

All these separation modes can be performed in different instrumental formats. The first electrophoretic separations of proteins were performed by Tiselius in a free solution in a U-tube macrocuvette, allowing refraction index detection of boundaries of protein zones. Soon it was found out that the separation would be more efficient if performed in anti-convective stabilizing medium; thus electrophoresis started to be performed in different carriers (paper, acetyl cellulose membrane) or in gel media (starch, agar, agarose, polyacrylamide (PAA)). In the last two decades electrophoresis has reverted to the free solution, stabilized by the anticonvective capillary effect of microbore (inner diameter (i.d.) < 100 µm)



of capillaries. In addition, CE is being applied to the detection of genetic mutations through dideoxy fingerprinting and to DNA adduct formation. For efficient separations of DNA in CE, linear polyacrylamide solutions of low viscosity and coated capillary surfaces are used. These are not ideal, however, and the tailoring of polymer solutions to particular separation needs is an active area of research. In the future, the marked improvements in microtechnology promise to bring about microfabricated sequencing arrays, the proof of principle for which has already been established. This approach can be easily integrated with microfluidics to develop 'lab-on-a-chip' devices. Instruments such as the Agilent Bioanalyzer are already capable of performing rapid, quality control analyses on-chip with 1 µl or less of sample.

**See also:** **Capillary Electrophoresis:** Overview. **Electrophoresis:** Clinical Applications.

## Further Reading

- Blackburn GM and Gait MJ (1996) *Nucleic Acids in Chemistry and Biology*. Oxford: Oxford University Press.
- Cantor CR and Schimmel PR (1980) *Biophysical Chemistry. The Behavior of Biological Macromolecules*. San Francisco: W. H. Freeman and Company.
- Grodzinsky AJ and Yarmush ML (1993) Electrokinetic separations. In: G. Stephanopoulos (ed.) *Biotechnology. Volume 3: Bioprocessing*, pp. 680–692. Weinheim: VCH Verlagsgesellschaft.
- Hu S and Dovichi NJ (2002) Capillary electrophoresis for the analysis of biopolymers. *Analytical Chemistry* 74(12): 2833–2850.
- Righetti PG, Gelfi C, and D'Acunzio MR (2002) Recent progress in DNA analysis by capillary electrophoresis. *Electrophoresis* 23(10): 1361–1374.
- Sambrook J and Russell DW (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

## Proteins

**V Kašička**, Czech Academy of Sciences, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Proteins are, together with nucleic acids, the most important biomolecules. Acting as enzymes, enzyme substrates and inhibitors, hormones, receptors, structural units, antigens, antibodies, drugs, and toxins, they play a vitally important role in all living organisms. In the current era of proteomics, the comprehensive analysis of the proteome, i.e., analysis of the set of all proteins encoded by the genome of a given organism and expressed in cells, tissues, or organs at a particular time and particular conditions, which nowadays represents the main approach for a complete understanding of biological systems and new drug discovery, the importance of proteins is ever increasing.

The huge number of variations of amino acid sequences in protein molecules and variety of attached prosthetic groups result in an extraordinary diversity of proteins that differ in electric charge, size, shape, specific binding capabilities, and, consequently, in electrophoretic mobilities, which allows their separation using different electrophoretic

methods:

1. zone electrophoresis (ZE) and isotachopheresis (ITP) in free-solution media – based on different charge-to-size ratios of proteins;
2. ZE in sieving media (gels and entangled polymer networks) – based on size-dependent retardation of electromigration of proteins by the medium matrix;
3. isoelectric focusing (IEF) – based on different isoelectric points of proteins; and
4. (bio)affinity electrophoresis (BAE) – based on influencing the electromigration of proteins through specific interactions with other (bio)molecules in separation media.

All these separation modes can be performed in different instrumental formats. The first electrophoretic separations of proteins were performed by Tiselius in a free solution in a U-tube macrocuvette, allowing refraction index detection of boundaries of protein zones. Soon it was found out that the separation would be more efficient if performed in anti-convective stabilizing medium; thus electrophoresis started to be performed in different carriers (paper, acetyl cellulose membrane) or in gel media (starch, agar, agarose, polyacrylamide (PAA)). In the last two decades electrophoresis has reverted to the free solution, stabilized by the anticonvective capillary effect of microbore (inner diameter (i.d.) < 100 µm)

fused silica capillaries or microchannels on chips. Due to space limitations, the different modes and instrumental formats of electrophoretic methods can be described here only in brief.

## Electromigration Properties of Proteins

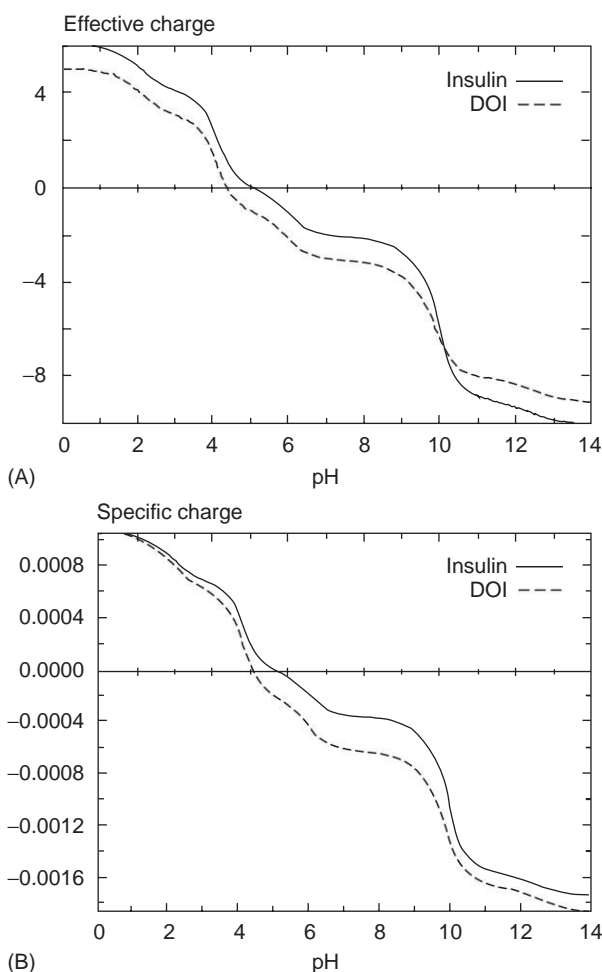
From the physicochemical point of view proteins are amphoteric (poly)electrolytes or (poly)ampholytes containing different types of ionogenic groups, e.g., carboxyl groups of C-terminal amino acids and aspartic and glutamic acids, amino groups of N-terminal amino acids, and side groups of lysine, arginine, and histidine. The basic parameter in electromigration methods of proteins is their electrophoretic mobility,  $m$  ( $\text{m}^2 \text{V}^{-1} \text{s}^{-1}$ ), i.e., their migration velocity related to a unit intensity electric field. The electrophoretic mobility of a protein is a complex function of (1) the properties of the protein itself (charge, size, shape); (2) the properties of the medium (background electrolyte (BGE)) in which the protein is moving (composition, pH, ionic strength, viscosity, permittivity, temperature); and (3) the interactions of the protein with the components of the BGE (solvation, dissociation, complex formation). The protein mobility, related to the given experimental conditions, particularly to the composition, pH, and temperature of the BGE, is called the effective (net) mobility.

The relationship between the effective mobility of a protein,  $m_{\text{ef}}$ , its effective charge,  $z_{\text{ef}}$ , and size (expressed as relative molecular mass,  $M_r$ ) can be approximately described by the equation

$$m_{\text{ef}} = az_{\text{ef}}/[b(M_r)^{1/3} + c(M_r)^{2/3}] \quad (1)$$

where  $a$ ,  $b$ , and  $c$  are empirical constants dependent on the electrolyte system used and on the size and structure of the proteins investigated. As follows from this equation, the mobility of a protein is directly proportional to its charge and indirectly proportional to its relative molecular mass with exponent in the range 1/3 to 2/3.

The effective (net) charge of a protein is given by the sum of the charges (including their signs) of all the ionogenic groups in its molecule. It can be estimated from the amino acid sequence of a given protein if the dissociation constants of the present ionogenic groups are at least approximately known. The calculated pH dependence of the effective charge and the specific charge (effective charge divided by  $M_r$ ) of the 'smallest protein', human insulin, and its polypeptide fragment, des-B23-B30-octapeptide insulin (DOI), is shown in Figure 1. From these curves



**Figure 1** Calculated dependence of (A) effective charge and (B) specific charge (effective charge divided by relative molecular mass) of pig insulin and DOI on pH.

an important characteristic of proteins, the isoelectric point, the pI, i.e., the pH at which the effective charge is zero, can be read, and some other useful information can be obtained, such as the pH regions of the maximal differences of their specific charges, suitable for their separation.

## Electrophoresis in Gels

Among the different gel matrices (starch, agar, agarose, PAA), especially the last two, agarose and PAA, are widely used for electrophoretic separations of proteins.

### Agarose Gel Electrophoresis

Agarose gels are used when nonrestrictive gels with large pores for the separation of macromolecules, proteins and nucleic acids, over 10 nm in diameter, are needed. The gels are prepared by cooling the

agarose (polysaccharide from red seaweed), dissolved in boiling water and poured into layers of different thicknesses, 0.1–2 mm, set on horizontal glass plates or support films. Agarose gels are applied for separations where the sieving effect of the gel matrix should be avoided, i.e., in isoelectric focusing, ITP, and immunoelectrophoretic techniques. In the latter technique a specific antibody is allowed to diffuse or migrate electrophoretically through the gel and to form insoluble immunocomplexes with corresponding antigens, thus detecting them with high specificity and sensitivity due to the high specificity and strength of the antigen–antibody complexes.

Agarose gels with concentrations of 0.7–1% (m/v) are often used in clinical laboratories for analysis of serum proteins and isoenzymes of diagnostic importance such as lactate dehydrogenase and creatine kinase.

### PAA Gel Electrophoresis

PAA gels are particularly suitable for electrophoresis of proteins due to their chemical inertness and controllable porosity. PAA gels are prepared by chemical copolymerization of acrylamide with a crosslinking reagent, *N,N'*-methylenebisacrylamide (or its derivatives). A clear transparent gel with low electroosmosis is obtained. The porosity can be exactly and reproducibly controlled by the total acrylamide concentration,  $T$ , and degree of crosslinking,  $C$

$$T[\%] = (a + b)100/V \quad [2]$$

$$C[\%] = b \times 100/(a + b) \quad [3]$$

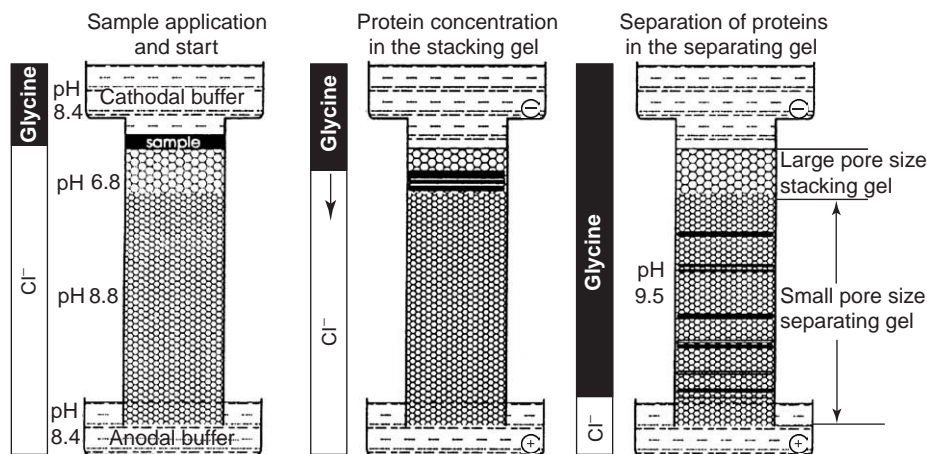
where  $a(b)$  is the mass of the acrylamide (bisacrylamide) in grams and  $V$  is the volume of the monomer

solution in milliliters. With increasing  $T$  at constant  $C$ , the pore size decreases; for constant  $T$  the pores are minimal at  $C = 4\%$ . Gels are usually polymerized in vertical casting chambers, cylindrical gels in glass tubes, and slab gels in moulds formed by two parallel glass plates sealed together along the edges. For electrophoresis in vertical systems the gels in glass rods or cassettes are placed in buffer reservoirs with electrodes, through which the electric field is applied to the gel.

For sample application the wells are formed by insertion of a sample comb between the glass plates at the upper edge of the gel during polymerization. In horizontal gels the sample can be introduced directly on the surface with strips of filter paper or silicone rubber.

At first, PAA gel electrophoresis (PAGE) was mostly performed in vertical cylindrical, rod-like gels polymerized in glass tubes of diameter 3–5 mm; later on and nowadays thin layer (0.1–0.5 mm) gels, both in the vertical and horizontal orientation, with or without cooling, are preferred because of faster, more efficient, and more sensitive separations.

**Discontinuous PAGE** Discontinuous PAGE (Disc-PAGE) is electrophoresis in a discontinuous electrolyte system. This discontinuous (isotachophoretic) or multiphasic buffer system consists of two different buffers, the leading electrolyte (Tris–HCl) and the terminating electrolyte (Tris–glycine), in two different gels. The sample is loaded on a large-pore stacking gel containing a lower pH (6.8) and lower concentration ( $0.125 \text{ mol l}^{-1}$ ) Tris–HCl buffer (see Figure 2). The large-pore stacking gel is polymerized on the top of the small-pore resolving gel containing a higher pH (8.8) and higher concentration



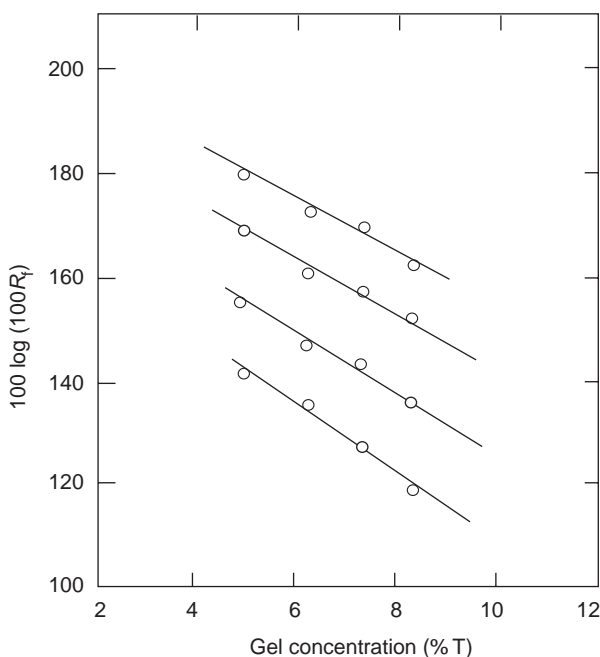
**Figure 2** Principle of discontinuous PAGE. (Reproduced with permission from Westermeier R (1997) *Electrophoresis in Practice: A Guide to Methods and Applications of DNA and Protein Separations*. Weinheim: Wiley-VCH; © Wiley-VCH.)

( $0.375 \text{ mol l}^{-1}$ ) Tris-HCl buffer. In both these gels and in the anode reservoir the chloride anion acts as the leading ion of the anionic ITP electrolyte system. The glycinate anion of the Tris-glycine buffer placed above the stacking gel and in the cathode reservoir serves as the terminating ion.

The main advantage of discontinuous buffer systems over continuous ones, which use a uniform buffer of constant pH, is that relatively large volumes of dilute protein samples without preliminary sample concentration can be applied to the gels. Due to the concentrating and self-sharpening effects of the ITP buffer system, proteins are concentrated into narrow zones during their migration in the large-pore gel, thus forming a suitable starting zone before entering the small-pore gel, in which size-based separation takes place in the ZE mode in the Tris-glycinate buffer. High-resolution separation is achieved due to both anticonvective and sieving effects of the PAA gel. When Disc-PAGE is performed in PAA rods, the well-defined and resolved proteins zones have a form of narrow disks, which is the second origin of the name disk-electrophoresis.

**Pore gradient gel electrophoresis (PGGE)** Soon after electrophoresis was established in the PAA gel, it was understood that the high resolution of proteins in this medium was achieved mainly due to a molecular sieving mechanism, differentially retarding proteins according to their size. This effect is exploited for size-based electrophoretic separation of proteins in gels with changing gel porosity, i.e., in PGGE. Pore gradient gels are prepared by varying the acrylamide concentration in the polymerization solution. The proteins are forced to migrate through progressively smaller pores until they reach a point where, due to their size, they are trapped in the tight gel matrix. PGGE can be used for estimation of the size (diameter or  $M_r$ ) if proper size calibration of the gel porosity is performed.  $M_r$  values can be obtained also from the so called Ferguson's plot, i.e., from the linear dependence of the decadic logarithm of the relative mobility ( $R_f$ ) of the given proteins on the gel concentration (% T), see **Figure 3**. The slope of this dependence, the retardation factor,  $K_R$ , is linearly related to the molecular radius or  $M_r$ .

**PAGE in the presence of sodium dodecyl sulfate (SDS)** Most proteins bind the anionic detergent SDS in a constant mass ratio, 1.4 g of SDS per gram of protein, effectively masking the intrinsic charge of the protein. Consequently, the charge-to-mass ratio of almost all protein-SDS complexes is constant, their mobilities in a free solution are identical, and they can be electrophoretically separated only in



**Figure 3** Ferguson plot of bovine  $\gamma$ -crystallin proteins. Four different gel concentrations (5.0%, 6.25%, 7.3%, 8.3% T) with Tris-glycine gel buffer were used for determination of the slope,  $K_R$ , of each subfraction, which is proportional to the relative molecular mass,  $M_r$ , of the native protein. (Reprinted with permission from Chiou SH and Wu SH (1999) Evaluation of commonly used electrophoretic methods for the analysis of proteins and peptides and their application to biotechnology. *Analytica Chimica Acta* 383: 47–60; © Elsevier.)

media exhibiting the sieving effect. In addition to masking of protein intrinsic charge, the detergent SDS molecules cause unfolding and denaturation of proteins by disruption of hydrogen bonds. Thus SDS-PAGE is characterized as denaturing electrophoresis, unlike the native electrophoresis described above. For complete denaturation and unfolding of proteins, the intra- and interchain disulfide bonds between cysteine residues are cleaved by reduction with mercaptoethanol or dithiothreitol, and the free -SH groups are protected by alkylation with iodoacetamide. There are two main buffer systems used for SDS-PAGE, the continuous phosphate buffer suggested by Weber and Osborn and the discontinuous system of Laemmli, which is in fact the original Ornstein-Davis native discontinuous Tris-HCl/Tris-glycine buffer system in the presence of SDS.

Due to a number of advantages, SDS-PAGE has become one of the most widely applied techniques for analysis and characterization of proteins. SDS solubilizes almost all proteins, and SDS-protein complexes are uniformly and highly negatively charged, i.e., fast migrating in the same anodic direction. In addition, SDS-protein complexes are better fixed and



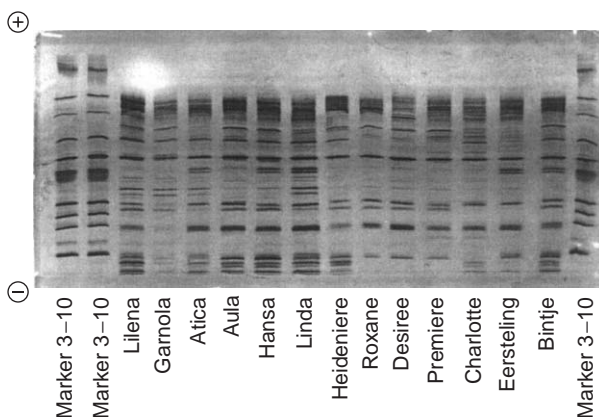
stained in the strongly restrictive gels than are native proteins. Separation is based on a physicochemical parameter, the effective molecular radius. In gels with uniform concentration, there is a linear relationship between the relative mobility ( $R_f$ ) and  $\log_{10} M_r$  of the SDS–polypeptide complex. From this linear dependence (calibration curve of a given gel) the  $M_r$  values of investigated proteins can be read. However, the linear relationship between  $\log_{10} M_r$  and  $R_f$  is valid only over a limited range of molecular mass; e.g., linearity in the  $M_r$  range 14 000–95 000 is achieved in a 5% T stacking gel and 14% T resolving gel. Moreover, some glycoproteins, collagenous polypeptides, and some very basic or acidic proteins behave anomalously, providing false  $M_r$  values.

### IEF in Agarose or PAA Gels

IEF is in fact electrophoresis of amphoteric molecules (proteins and peptides) in an electrolyte system with a pH gradient. In this pH gradient proteins move according to their charge until they reach the region of pH at which their effective charge is zero, i.e., their isoelectric point (pI).

**IEF with carrier ampholytes** In classical IEF the pH gradient is created and maintained by the passage of electric current through a solution of complex mixtures of small synthetic amphoteric molecules with closely spanned isoelectric points due to different amounts and ratios of both basic amino- and acidic carboxylic groups. These carrier ampholytes are electrophoretically transported according to their pI and thus form the pH gradient with the pH increasing from the anode to the cathode. The widest pH range of IEF with carrier ampholytes is 3.5–10. Unlike ZE, IEF is a steady-state technique, i.e., once the proteins have reached their pI, they remain concentrated (focused) in narrow zones at the pH equal to their pI. Because of the focusing effect of the pH gradient, sharp protein zones with high resolution (pI difference up to 0.01) can be obtained. In gel IEF it is important to use gels with a large porosity; in a low concentration (3–5% T) PAA gel proteins with  $M_r$  up to 750 000 can be separated, but for proteins with  $M_r > 500\,000$ , agarose gels (0.8–1.0%) are preferred. The solubility of hydrophobic and other low-solubility proteins, which is lowest at pI, can be improved by addition of chaotropic agents, urea or thiourea, and nonionic (Triton X100) or zwitterionic (CHAPS) detergents.

**IEF with immobilized pH gradient** An important innovation in IEF was the development of special reagents (immobilines) for preparation of PAA gels



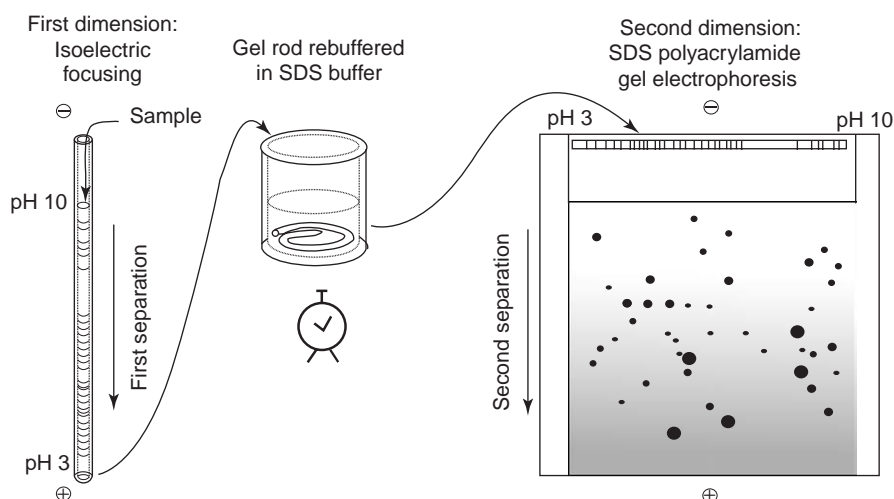
**Figure 4** High resolution IEF of proteins from potatoes of different varieties in slab PAA gel with carrier ampholytes. Staining with CBB. (Reproduced with permission from Westermeier R (1997) *Electrophoresis in Practice: A Guide to Methods and Applications of DNA and Protein Separations*. Weinheim: Wiley-VCH; © Wiley-VCH.)

with immobilized pH gradients (IPGs). The pH gradient is prepared before the IEF run by copolymerization of several acrylamide derivatives containing either an acidic carboxylic group or basic tertiary amino groups. Depending on the amounts and ratio of acidic and basic acrylamide derivatives in the polymerization mixture in different places of the gel matrix, a PAA gel with different acidity (pH gradient) is formed. The main advantage of IPGs is that the buffering groups forming the pH gradient are covalently attached and immobilized to the PAA backbone. The pH gradient they generate is also immobilized and stable in time and space, eliminating the deleterious effects of pH gradient (cathodic) drift commonly observed in conventional IEF with carrier ampholytes.

IEF in horizontal slab gels of ultrathin thickness (0.05–0.5 mm) cast on thin plastic supports with improved heat transfer is largely applied to protein analysis, preparation, and characterization. It is used for determination of isoelectric points of proteins, for evaluation of protein microheterogeneity, for identification of genetic variations (see Figure 4), and for investigating chemical, physical, and biological influences on proteins. One of the most important applications of IEF is using this technique as the first dimension separation of two-dimensional (2D) gel electrophoresis.

### 2D-PAGE

The separation power of 1D electrophoretic or other separation techniques for analysis of complex protein mixtures is rather limited since closely related proteins remain unresolved in overlapping zones. The

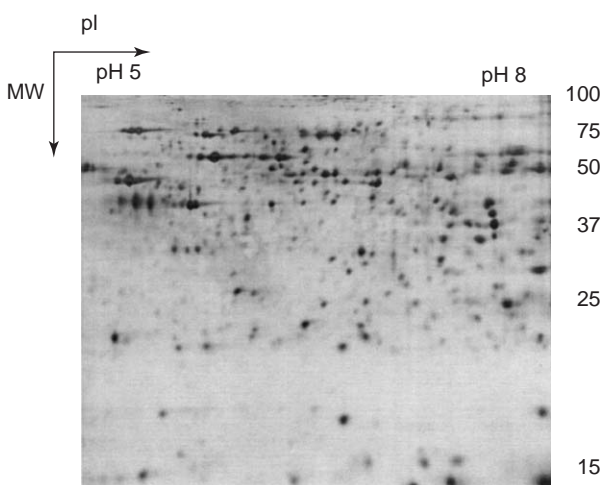


**Figure 5** Principle of high resolution 2D gel electrophoresis. (Reproduced with permission from Westermeier R (1997) *Electrophoresis in Practice: A Guide to Methods and Applications of DNA and Protein Separations*. Weinheim: Wiley-VCH; © Wiley-VCH.)

principle of solution of this problem is to combine two independent separation principles based on two different physicochemical parameters in one method. This was performed in 1975 by O'Farrell, who combined the high resolving power of IEF, separating proteins according to isoelectric points, with the high-resolution SDS-PAGE, separating proteins according to size, and developed 2D-PAGE, which can resolve up to a few thousand proteins.

In the first step the sample is dissolved in a small volume of a denaturing solution containing urea, nonionogenic or zwitterionic detergents, and a reducing agent (dithiothreitol). This sample buffer solution can dissolve and dissociate most proteins without changing their intrinsic charge. The dissolved polypeptide mixture is separated by IEF as described above. In the second step the thin (0.5–1 mm) rod-like or strip gel containing the IEF separated proteins is soaked in SDS-containing buffer (see Figure 5) and the protein–SDS complexes are separated in the second dimension in the slab gel according to size in the SDS-PAGE mode. The result of separation is a pattern of spots of stained proteins in a rectangular area – the 2D protein map.

The separation parameter of the first dimension (pI) is independent of  $M_r$ , which is the separation parameter of the second dimension. Consequently, the 2D IEF  $\times$  SDS-PAGE separation is orthogonal and able to provide the highest resolution of complex protein mixtures, such as proteins of biofluids, cell lysates, and tissue extracts, among all currently available methods. Under optimized separation conditions, with large-size gels, sensitive silver staining, or autoradiography detection, quite a few thousand proteins can be resolved and detected in one



**Figure 6** 2D gel electrophoresis of liver proteins. Approximately 40  $\mu$ g of protein was separated using IEF on pH 5–8 IPG strips for 22 000 V h followed by a second dimension SDS-PAGE on a 12% T PAA gel and was silver stained. (Reproduced with permission from Yarmush ML and Jayaraman A (2002) *Advances in proteomic technologies. Annual Review of Biomedical Engineering* 4: 349–373.)

two-dimensional electrophoresis (2-DE) map (see Figure 6). Naturally, 2-DE has become a basic technology in proteomics.

The high reproducibility of the positions of the spots, which is very important for interpretation of protein 2-DE maps, has been particularly improved with IPGs used in the first, IEF, dimension. For qualitative and quantitative analysis and for intergel comparison the 2-DE maps are converted into digital data using densitometers, scanners, or video cameras and evaluated by computer image analysis.



## Detection

Proteins separated in gels are mostly detected using staining techniques based on the ability of different dyes to form colored complexes with proteins and to visualize them. Coomassie brilliant blue (CBB) is the most commonly used organic stain for detection of proteins in gels. The relatively high staining intensity of Coomassie blue stains is based on the electrostatic interaction between the dye molecules and the amino groups of proteins and, on the secondary bonds, dye-dye interactions between the dye molecules. CBB, in a micellar form, can detect as little as  $\sim 50$  ng of protein and gives a linear response in the densitogram of up to  $20\ \mu\text{g}$ , but the slope of the response is different for different proteins. The sensitivity of detection and resolution of the separation is increased by fixing proteins in gel with trichloroacetic acid. Amido black and fast green are other organic dyes used for protein detection, but their sensitivity is three to six times lower than that of CBB.

More than a hundred times higher detection sensitivity, up to  $0.1$  ng of protein, is achieved by silver staining. It is based on reduction of silver ions to metallic silver due to differences in the reduction-oxidation potential between the sites occupied by proteins and the adjacent sites of the gel or membrane. Densitograms of silver-stained protein electrophoregrams are linear over a 30- to 40-fold range for most proteins, but the slope of the response also differs for different proteins.

The other techniques used for protein detection in gels include direct ultraviolet (UV) densitometry, and autoradiography and fluorography of radio- or fluoro-labelled proteins, respectively. Methods have also been developed for identification of specific types of proteins. Proteins may be stained on the basis of the presence of specific moieties (e.g., glycoproteins) or by their specific enzymatic activity (positive or negative zymograms). Immunochemical methods are used for identification of proteins with specific antigenic sites (e.g., allergens, phosphoproteins).

Both universal staining procedures and specific detection techniques can be performed after (electro) transfer or (electro)blotting of proteins (also called Western blotting) from the gel matrix (which sometimes hinders protein analysis) to a nitrocellulose or polyvinylidenedifluoride membrane, to which they are bound and immobilized. On the membrane, protein molecules are faster and better accessible for the interactions with the applied specific antibodies. The antigen-antibody complexes are visualized by a second antibody (against the first antibody) with an attached enzyme label, which catalyzes the color reaction in the place of the protein zone.

In the latest developments, matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) has been employed for detection of proteins with the capability to determine their  $M_r$  with high accuracy.

## Capillary Electrophoresis

### Basic Characteristics, Advantages, and Problems Encountered

Since the 1980s electromigration methods have started to be performed in thin-wall microbore fused silica capillaries with a typical i.d. of  $25\text{--}75\ \mu\text{m}$  and length  $20\text{--}70$  cm. The capillary format of the separation compartment brings significant advantages over gel electrophoresis.

1. Capillaries possess a stabilizing anticonvective effect, allowing separation in a free solution, under mild conditions, preserving the biological activity of biomolecules.
2. The relatively high surface/volume ratio of the separation compartment ensures efficient Joule heat transfer from the separation space, which allows application of high-intensity electric fields, up to  $1000\ \text{V cm}^{-1}$ , resulting in fast and highly efficient separations of analytes.
3. Online, on-column, or end-column detection makes possible direct monitoring and evaluation of the separation process; the whole analytical procedure can be automated.

During the last decades the different modes of capillary electrophoresis (CE) have developed into high-performance (HPCE) separation methods, offering a high separation efficiency ( $10^5\text{--}10^7$  theoretical plates), high sensitivity (femtomole to zeptomole amounts in nano- to picoliter sample volumes), and short analysis times (typically  $5\text{--}20$  min, in special cases only a few seconds). They are considered as a recognized complement and/or counterpart of liquid chromatography (LC) and gel electrophoretic methods.

The application potential of HPCE methods is extremely large; they can be used for separation of any class of compounds, ionogenic and nonionogenic, small ions and macromolecules, and also for (bio)particles and cells. Proteins belong to the substances most frequently analyzed using HPCE.

CE is mostly performed in fused silica (FS) capillaries, which form a suitable separation compartment because of their reproducible physical dimensions, good electrical isolation properties, high thermal conductivity, and transparency in the short wavelength UV radiation used for detection. However, the dissociation of silanol groups created on a

hydrated FS surface brings complications for CE separation of proteins due to their adsorption to the ionized capillary wall. There are several strategies for suppressing protein adsorption to the FS capillary. In addition to the separations performed in extreme (high or low) pH and high ionic strength BGEs, different FS capillary coatings are employed. Dynamic coatings result from reversible (dynamic) adsorption of either low molecular mass compounds (alkylammonium cations, amines) or hydrophilic polymers such as cellulose derivatives added to the BGE. Static (permanent) modification of the capillary wall is carried out prior to filling the capillary with BGE and may involve the formation of a covalent bond between a suitable derivatized coating agent and the silanol groups of the silica, giving rise to an immobilized polymeric coating, or the bond may be noncovalent but yet 'permanent', resulting from the irreversible adsorption of the polymer, e.g., poly(vinyl alcohol) or poly(ethylene glycol), to the capillary wall.

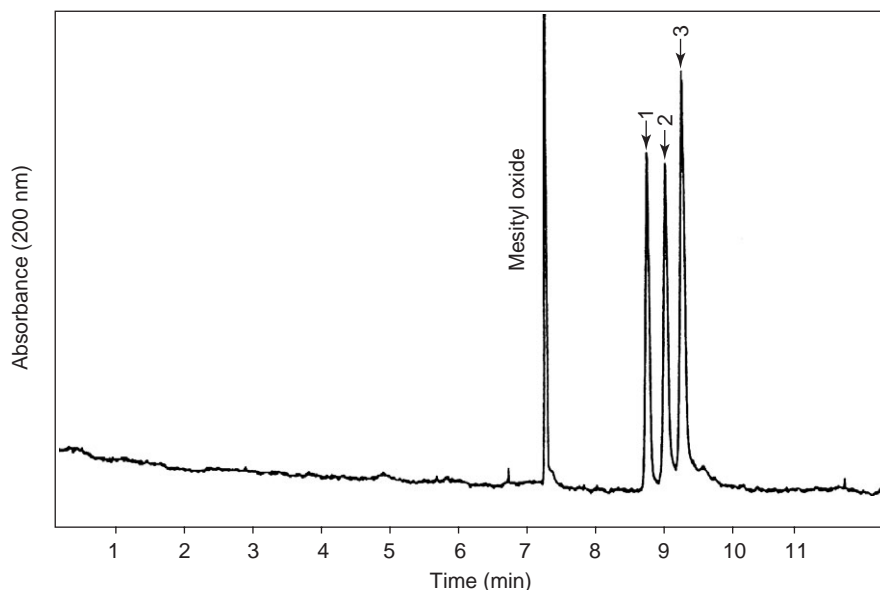
The second important phenomenon induced by the dissociation of silanols is the electroosmotic flow (EOF), i.e., bulk flow of BGE inside the capillary resulting from application of an electric field on a free charge in the diffusion part of the electric double layer close to the capillary wall. The EOF in the bare FS capillary has a cathodic direction, and its velocity

is higher than the electrophoretic velocity of proteins, which allows analyzing both positively and negatively charged proteins in a single experiment.

### Separation Modes

In the capillary format, different electromigration separation principles (separation modes) are performed.

*Capillary zone electrophoresis (CZE)* is the simplest and most universal CE mode applied to separation of proteins. Separation takes place in the homogeneous solution of the background electrolyte (BGE) on the basis of different electrophoretic mobilities of analytes in a free solution. Sample components are separated into zones with gaussian concentration profiles. The selectivity of protein separations is manipulated mainly by changes in their charge-to-size ratio, which can be achieved by changes in the composition, pH, ionic strength, viscosity, permittivity, and temperature of the BGE. The high resolving power of CZE is demonstrated in **Figure 7**, showing the separation of human growth hormone (hGH) and its deamidation products, differing in one elementary charge per 191 amino acid residues. CZE in sieving media, particularly in replaceable solutions of entangled polymer networks (linear PAA, dextran), is broadly applied for SDS-electrophoresis of proteins.



**Figure 7** CZE separation of hGH and its derivatives. CZE performed in FS capillary (i.d. 50  $\mu\text{m}$ , total length 105 cm, effective length (to detector) 81.5 cm; BGE: 10  $\text{mmol l}^{-1}$  Tricine, 5.8  $\text{mmol l}^{-1}$  morpholine, 20  $\text{mmol l}^{-1}$  NaCl, pH 8.0; electric field intensity 300  $\text{V cm}^{-1}$ , current 20  $\mu\text{A}$ , temperature 24°C. 1, hGH; 2, (desamido-149)- and (desamido-152)-hGH; 3, (didesamido-149-152)-hGH. Mesityl oxide, electroosmotic flow marker. (Reprinted with permission from Grossman PD, Colburn JC, Lauer HH, *et al.* (1989) Application of free-solution capillary electrophoresis to the analytical scale separation of proteins and peptides. *Analytical Chemistry* 61(11): 1186–1194; © American Chemical Society.)

*Capillary ITP (CITP)* is suitable for determination of low molecular mass ionic admixtures of proteins, and it is also often used as a preconcentration and/or preseparation step prior to CE analysis of proteins present at low concentrations and/or in complex mixtures, increasing the sensitivity up to several hundred times.

*Capillary IEF (CIEF)* combines the high resolving power of conventional gel IEF with the advantages of CE instrumentation. The capillary format with efficient Joule heat transfer permits the application of high electric field strengths for rapid and high resolution focusing with online detection of separated zones mobilized by EOF, electroelution or hydrodynamic flow. CIEF is used for investigation of protein microheterogeneity and for pI determination. Due to its focusing effect, CIEF is frequently used as the first concentrating step in 2D or multidimensional separations of complex protein mixtures. A 2D separation, analogous to classical IEF/SDS-PAGE, is achieved by a combination of CIEF with mass spectrometric detection, which also provides both pI and  $M_r$  values.

*Capillary (bio)affinity electrophoresis* is mostly used as a mild and sensitive tool for the investigation of interactions of proteins with other molecules and for estimation of the binding constant of their complexes.

*Capillary electrokinetic chromatography* with micellar pseudophases of ionogenic detergents, anionic SDS, or cationic cetyltrimethylammonium bromide, is capable of resolving small proteins or polypeptides ( $M_r < 6500$ ) with minor differences in hydrophobicity resulting, e.g., from replacement of a single nonionogenic amino acid by other electroneutral amino acid residue.

## Detection

One of the main advantages of CE over gel electrophoresis is that the separation is monitored by online, on-column, or end-column detection. In the most frequently employed UV absorption photometric detection, a small part (less than 1 mm) of the capillary serves as a detection cell. Micromolar concentrations of proteins are detectable using the low UV detection wavelength of 200–220 nm. A higher sensitivity, up to nanomolar concentrations, is achieved with fluorescence, particularly laser induced fluorescence (LIF) detection. The disadvantage of the LIF detection of proteins is the necessity for their derivatization using a fluorogenic label. The native fluorescence of proteins, mostly due to the presence of aromatic amino acids residues, tryptophan, and tyrosine, can be utilized only when low UV laser

systems with an excitation wavelength between 270 and 300 nm are available or when two or three visible or infrared photon excitation is possible.

Mass spectrometry (MS) is, due its universality, sensitivity, and selectivity, the most powerful detection principle for CE. The introduction of electrospray ionization (ESI) and MALDI have especially brought tremendous progress in online and offline characterization of proteins separated electrophoretically using MS. A combination of CE with ESI-MS and MALDI-MS allows not only high accuracy determination of the protein  $M_r$  but also provides important structural data on amino acid sequences, posttranslational modifications, peptide mapping, and noncovalent interactions of proteins. The MS detection principle, combined with the high resolution of CE, is becoming of importance, namely for comprehensive analysis of proteins (proteomics) and for protein characterization by peptide mapping.

## Separation on Microchips

Electrophoretic techniques performed on microfabricated devices – microchips – represent the platform for a new generation of miniaturized analyzers where all operations, sample introduction, derivatization, separation, and detection are fully integrated and automated in the so called ‘micro-total analytical systems’ or ‘labs on a chip’. They are expected to become the most powerful tools of analytical chemistry in the coming period, with broad application in life sciences, particularly proteomics, requiring fast, highly efficient, highly sensitive, and high throughput separation and characterization of complex protein mixtures.

## Applications

Both gel electrophoresis (GE) and CE have an extremely broad application potential for analysis, preparation, and characterization of proteins.

## Analysis

There are thousands of applications of both GE and CE for qualitative and quantitative analysis, quality control, or purity determination of proteins isolated from the natural material or prepared using modern recombinant DNA technology, for monitoring the modification of proteins through chemical reactions or enzymatic conversions, for amino acid and sequence analysis of proteins, and for protein characterization by peptide mapping. Only a few of them will be mentioned here as illustrative examples.

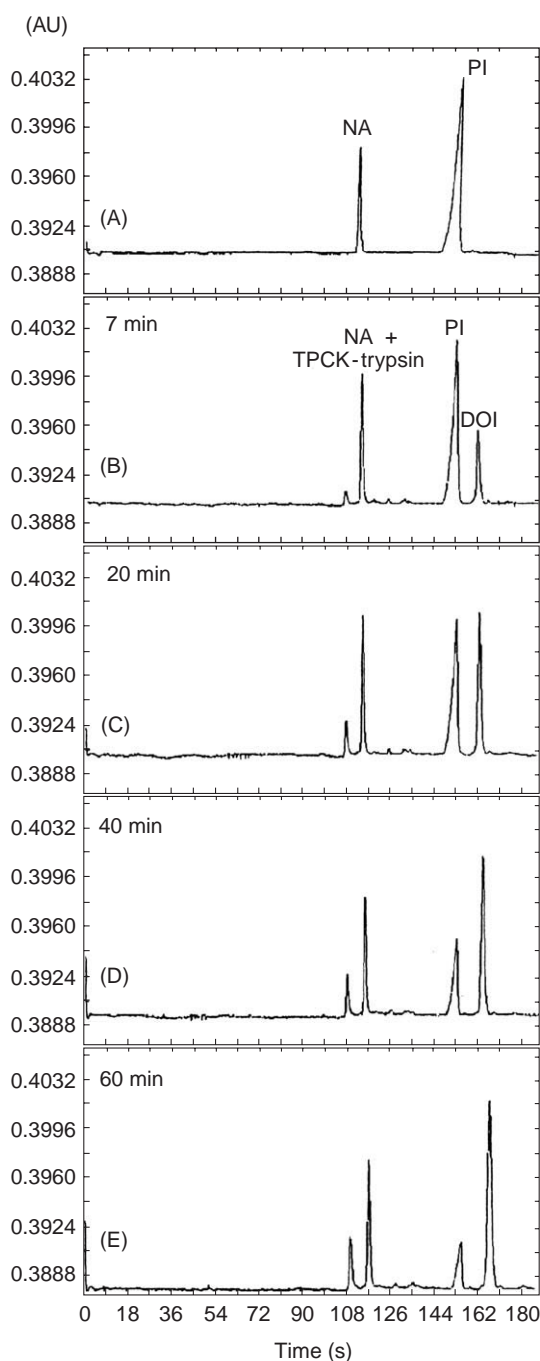
Validated CE and GE methods have been developed for the qualitative and quantitative analysis of

biotechnologically produced protein pharmaceuticals, such as human insulin, growth hormone, tumor necrosis factor, antithrombin III, hirudin, and many others.

The high sensitivity of GE and CE, frequently in combination with online sample cleanup and pre-concentration, allows these methods to be applied to the analysis of proteins present at low concentration levels in complex biomatrices, such as biological fluids (blood, serum, plasma, urine, cerebrospinal fluid), cell lysates, and tissue extracts. Protein biomarkers of physiologically important processes and diseases, such as abnormal prion proteins in the blood of animals and humans infected with a transmissible spongiform encephalopathy,  $\beta$ -amyloid polypeptides in plasma or serum of Alzheimer disease patients, peptidomimetic inhibitors of HIV-1 protease in serum of AIDS patients were determined using CZE. Both GE and CE (CITP) are applied for analysis of lipoproteins and apolipoproteins in plasma and for elucidation of disorders of lipoprotein metabolism related to diseases of the cardiovascular and central nervous systems. Slab gel IEF of proteins, particularly some enzymes, is used for blood group typing in the medico-legal field.

In addition to analysis of 'static' protein preparations, GE and CE are capable of also monitoring dynamic changes in proteins, such as their chemical modifications and reactions (oxidation, reduction, deamidation, hydrolysis), physical changes (denaturation, aggregation, folding/unfolding), and enzymatic conversions, including physiologically important posttranslation modifications as phosphorylation and glycosylation. Several CE applications deal with enzymatic conversions of proteins for studying some details of these processes and/or activities of enzymes and the kinetics of their action on proteins and peptides. An example of CZE monitoring of pig insulin cleavage using trypsin is shown in Figure 8. Another application is CZE monitoring of asparagine deamidation and aspartate isomerization in recombinant human interleukin-11 due to heat stress and determination of activity of angiotensin-converting enzyme, an important regulator of blood pressure.

The high separation power makes CE and GE powerful techniques for peptide mapping, i.e., separation of peptide fragments originating from specific chemical and/or enzymatic hydrolysis of proteins and polypeptides. Peptide mapping serves as an important tool for protein identification, for sequence determination of internal parts of polypeptide chains, for monitoring of posttranslational modifications, and elucidation of the structure of proteins. Due to the high complexity of peptide maps, namely of large proteins, usually multidimensional separations,

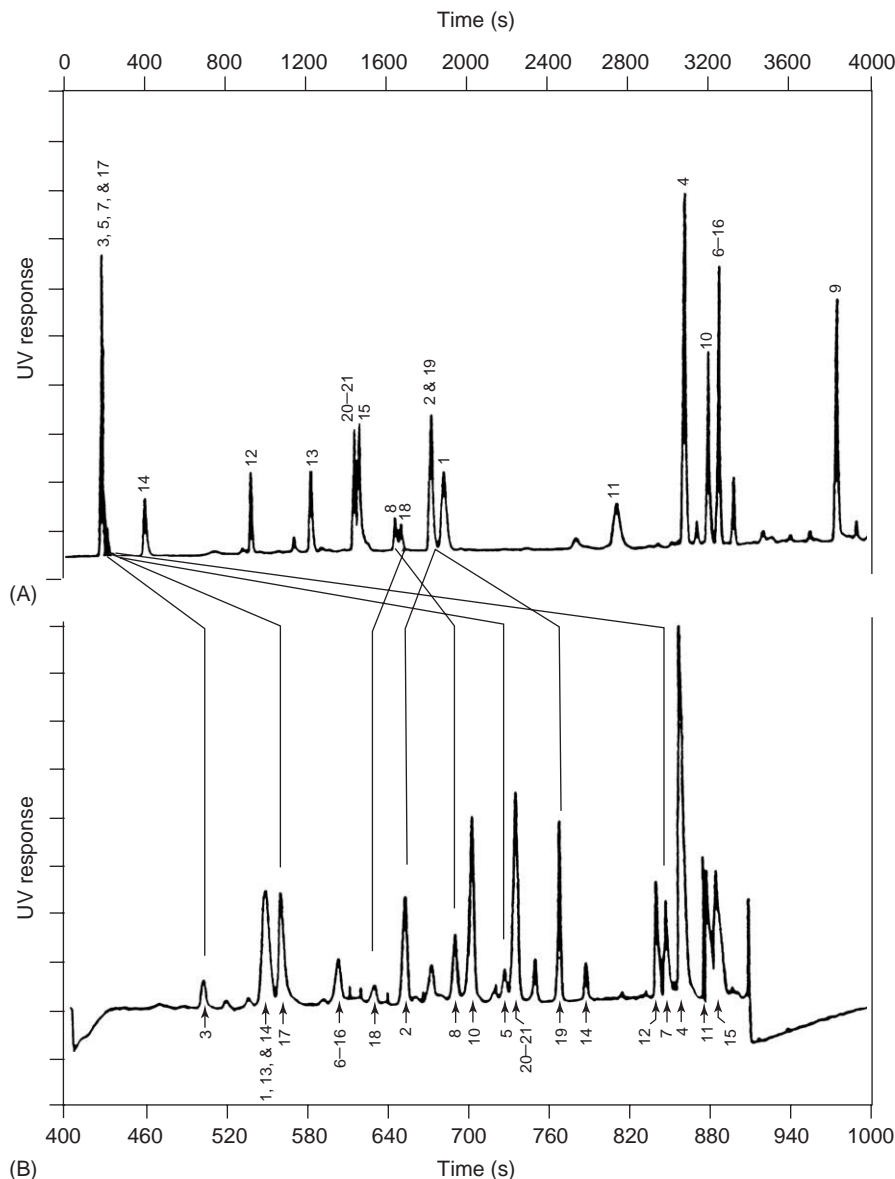


**Figure 8** CZE analysis of pig insulin (A) and monitoring the cleavage of pig insulin using trypsin after 7, 20, 40, and 60 min (B–E). 150  $\mu$ g of pig insulin dissolved in 150  $\mu$ l of BGE and mixed with 50  $\mu$ g of trypsin dissolved in 100  $\mu$ l BGE and analyzed at given time periods using CZE in a FS capillary (i.d. 56  $\mu$ m, total/effective length 31/20 cm; BGE: 10 mmol l<sup>-1</sup> tricine, 5.8 mmol l<sup>-1</sup> morpholine, 20 mmol l<sup>-1</sup> NaCl, pH 8.0; voltage 7.3 kV, current 20.0  $\mu$ A; UV absorption detection at 206 nm. PI, pig insulin; DOI, des-octapeptide-B23-30-insulin; NA, nicotinamide (EOF marker). (Adapted from Kašička V (1999) Analytical and preparative separations of peptides by capillary and free-flow zone electrophoresis. In: Aboul-Enein HY (ed.) *Analytical and Preparative Separation Methods of Biomacromolecules*, pp. 39–97. New York: Dekker.)

2D-PAGE or CE-MS, are necessary for complete resolution of these mixtures. CE separation of tryptic peptides of hemoglobins, i.e., peptides obtained by cleavage of hemoglobin by trypsin, allows identification of hemoglobin variants in clinical laboratories. The comparison of CZE and LC tryptic peptide maps of hGH presented in Figure 9 demonstrates the high separation power and complementarity of reverse-phase (RP) LC and CZE, the latter however being four times faster.

### Preparative Separations

Several special apparatus have been designed for large-scale electrophoretic separation of proteins, but none of them is widely used nowadays. The main reason is that the Joule heat produced in the enlarged separation compartment cannot be sufficiently and evenly dissipated, which results in a high temperature increase and a temperature gradient in the separation compartment and loss of separation power. A



**Figure 9** Comparison of RP-LC (A) and CZE (B) separations of tryptic peptides of hGH. Peak assignments and correlations for selected fragments are noted. CZE: background electrolyte  $0.01 \text{ mol l}^{-1}$  Tricine,  $0.045 \text{ mol l}^{-1}$  morpholine,  $0.02 \text{ mol l}^{-1}$  NaCl, pH 8.0, detection at 200 nm, 316 V cm,  $20 \mu\text{A}$ . LC: column Aquapore RP-300 ( $4.6 \times 250 \text{ mm}$ ), flow rate  $1 \text{ ml min}^{-1}$ , detection 214 nm. Solvents: A, 0.1% trifluoroacetic acid (TFA) in water; B, 0.1% in acetonitrile; gradient 0–20% B in 20 min, 20–25% B in 25 min. (Reprinted with permission from Grossman PD, Colburn JC, Lauer HH, *et al.* (1989) Application of free-solution capillary electrophoresis to the analytical scale separation of proteins and peptides. *Analytical Chemistry* 61(11): 1186–1194; © American Chemical Society.)



compromise between the deleterious effects of Joule heating and the preparative capacity is best achieved in column electrophoresis with an annular geometry with subsequent elution of the separated proteins after they leave the gel.

Due to the inherently low preparative capacity of miniaturized thin layer gels and capillary columns, the application of GE and CE for preparative separations of proteins is limited to cases where their nanogram to microgram amounts are sufficient for further characterization, e.g., using offline MS or using amino acid and sequence analysis. A higher preparative capacity (50–100 mg of proteins per hour) is obtained in free-flow electrophoresis. In this technique the sample is continuously introduced into the laminar flowing thin layer of the BGE in the rectangular flow-through electrophoretic chamber. During the flow-through time the sample components are separated according to their electrophoretic mobilities using an electric field oriented perpendicular to the hydrodynamic flow of the BGE and the sample solution.

### Physicochemical Characterization

GE and CE can also provide important physicochemical characteristics of the separated proteins, such as effective and absolute (limiting) mobilities, effective charges, isoelectric points, relative molecular masses, Stokes radii, acid–base dissociation constants of ionogenic groups, diffusion coefficients, association (dissociation) constants of protein complexes, and rates of protein physical changes and chemical reactions. The size parameters (diameter or  $M_r$ ) of proteins are obtained from CE and GE in sieving media; the effective charge of proteins may be estimated from separation of protein charge ladders; the  $pK_a$  values of ionogenic groups of proteins are determined from CZE measurements of the dependence of the effective mobility on the pH; the pH dependence of proteins' effective mobility and the isoelectric point can be obtained using a special 2D technique, the electrophoretic titration curve, based on the application of protein(s) as the longitudinal zone in the middle of slab gel along the pH gradient, first formed by IEF in one direction, followed by protein electromigration in the direction perpendicular to the direction of the pH gradient. The diffusion coefficients of proteins can be estimated from their 'stopped flow' CZE separations. CE and GE separation of folded and unfolded forms of proteins and polypeptides allows us to study the equilibria and kinetics of conformation transition states during

protein and polypeptide folding–unfolding processes and coil–helix transitions.

**See also:** **Capillary Electrophoresis:** Overview. **Electrophoresis:** Principles; Isotachopheresis; Isoelectric Focusing; Polyacrylamide Gels; Two-Dimensional Gels; Affinity Techniques; Blotting Techniques; Clinical Applications. **Mass Spectrometry:** Overview; Matrix-Assisted Laser Desorption Ionization; Time-of-Flight.

### Further Reading

- Bean SR and Lookhart GL (2001) High-performance capillary electrophoresis of meat, dairy, and cereal proteins. *Electrophoresis* 22: 4207–4215.
- Chiou SH and Wu SH (1999) Evaluation of commonly used electrophoretic methods for the analysis of proteins and peptides and their application to biotechnology. *Analytica Chimica Acta* 383: 47–60.
- Dunn MJ (ed.) (1986) *Gel Electrophoresis of Proteins*. Bristol: Wright.
- Hjertén S (1996) Capillary electrophoretic separation in open and coated tubes with special reference to proteins. *Methods in Enzymology* 270: 296–319.
- Hu BH and Martin LM (1999) Capillary electrophoresis of peptides and proteins. In: Millner PA (ed.) *High Resolution Chromatography. A Practical Approach*, pp. 77–116. Oxford: Oxford University Press.
- Hutterer K and Dolník V (2003) Capillary electrophoresis of proteins 2001–2003. *Electrophoresis* 24: 3998–4012.
- Kašička V (2003) Recent advances in capillary electrophoresis and capillary electrochromatography of peptides. *Electrophoresis* 24: 4013–4046.
- Patrick JS and Lagu AL (2001) Applications of capillary electrophoresis to the analysis of biotechnology-derived therapeutic proteins. *Electrophoresis* 22: 4179–4196.
- Regnier FE and Lin S (1998) Capillary electrophoresis of proteins. In: Khaledi MG (ed.) *High-Performance Capillary Electrophoresis: Theory, Techniques, and Applications*, pp. 683–727. New York: Wiley.
- Righetti PG (2001) Capillary electrophoretic analysis of proteins and peptides of biomedical and pharmaceutical interest. *Biopharmaceutics & Drug Disposition* 22: 337–351.
- Righetti PG, Stoyanov A, and Zhukov M (2001) *The Proteome Revisited: Theory and Practice of the Relevant Electrophoretic Steps*. Amsterdam: Elsevier.
- Wehr T, Rodriguez-Diaz R, and Liu CM (1997). Capillary electrophoresis of proteins. In: Brown PR and Grushka E (eds.) *Advances in Chromatography*, vol. 37, pp. 237–361. New York: Dekker.
- Wiktorowicz JE and Colburn JC (1992) Protein analyses by capillary electrophoresis. In: Grossman PD and Colburn JC (eds.) *Capillary Electrophoresis: Theory and Practice*, pp. 273–300. San Diego: Academic Press.



# ELECTROTHERMAL ATOMIC ABSORPTION SPECTROMETRY

See ATOMIC ABSORPTION SPECTROMETRY: Electrothermal

## ELEMENTAL SPECIATION

Contents

**Overview**

**Waters, Sediments, and Soils**

**Practicalities and Instrumentation**

### Overview

**A Sanz-Medel and M L Fernández-Sánchez,**  
University of Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

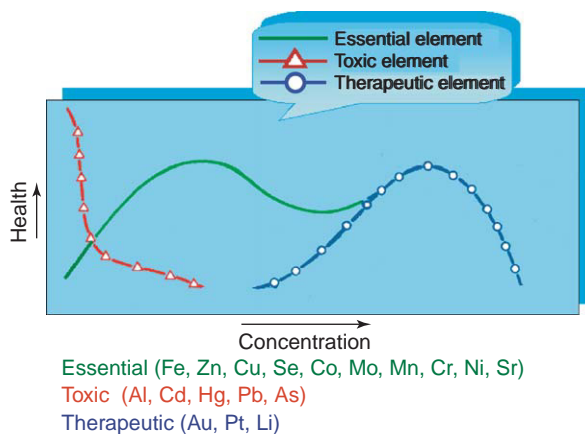
### Chemical Speciation: Concept and Importance

Analytical information about the total amount of trace and ultratrace level concentrations of elements is increasingly demanded in biology, medicine, industry, and environment. The interest in trace and ultratrace elements in living organisms began after discovering the presence of various metals in a number of bio-compounds having biological significance. Moreover, it was demonstrated that trace and ultratrace elements in a living organism could play important roles including: (1) 'Essential' element: the simplest definition is an element required for life in such a way that their absence results in death or severe malfunction of the organism. The relationship between concentration level of an essential element and biological function is shown in **Figure 1**. (2) 'Toxic' element: as can also be seen in **Figure 1**, they are elements that can be harmful or even lethal at very low concentrations. As a toxic element concentration increases, its deleterious effect becomes more harmful. (3) 'Therapeutic' element: an element that can be beneficial to health when administered at higher levels than those of essential elements (see **Figure 1**).

However, metals and semimetals naturally occur in many physicochemical forms or species in the different environmental compartments. They may be found as free metal ions, hydroxides, organometallic compounds, large biomolecules, etc. In fact, today it

is worldwide recognized that the actual distribution, mobility, bioavailability, toxicity, bioaccumulation, and biodegradability of an element will depend not only on its total concentration but on its chemical form in the sample. Therefore, total concentration determinations of an element seem to provide incomplete information on its mobility, ecotoxicity, bioavailability, etc. For this reason the analysis of its physicochemical form(s) then becomes mandatory to understand its biological behavior, its toxicity, or its eventual environmental impact.

A simple and well-known example of speciation importance is arsenic present in marine foodstuffs. Inorganic arsenic species have long been recognized to be highly toxic. However, the major arsenic compound content in such food turns out to be arsenobetaine or arsenosugars, which are totally harmless to humans (as they are not metabolized during digestion). Similarly, all forms of mercury are toxic but their toxicity depends strongly on the metal



**Figure 1** Concentration levels of essential, toxic, and therapeutic elements and their influence on health.

chemical form: methylmercury (MeHg) is one of the most toxic species and the famous Minamata Bay disaster in the 1950s was caused by MeHg poisoning. An industrial effluent forming MeHg was discharged in the Minamata Bay being then bioconcentrated by the fish. Not only animals (e.g., cats) but also humans were severely affected by consumption of the contaminated fish. Thus, MeHg was the origin of the so-called 'Minamata disaster', as the toxicity of the methylated mercury compound (e.g., to the central nervous system) turned out to be much stronger than that of inorganic mercury.

These two examples have been selected only to illustrate the fact that elemental speciation information is today urgently needed as total determination can be even misleading.

According to IUPAC recommendations a 'chemical species' is the specific form of an element defined as to isotopic composition, electronic or oxidation state, and/or complex of molecular structure. 'Speciation analyses' are the analytical activities of identifying and/or measuring the quantities of one or more individual chemical species in a sample, while 'speciation of an element' is the distribution of an element amongst defined chemical species in a system.

Although metal speciation began closely related to the toxicity aspects and to environmental concerns, speciation information is increasingly demanded from most varied areas. The main fields for speciation analysis today are: environment, biology, and medicine, and also industrial analysis. Environmental monitoring (air, water, soils, sediments, and biota) of anthropogenic organometallic contaminants (e.g., organolead or organotin) and of toxic redox states (e.g., Cr(III) is essential while Cr(VI) is carcinogenic) is one of the better established speciation fields. Elucidation of biomethylation metal mechanisms, as methylated compounds show different toxicities than the inorganic forms (e.g., Hg, As, Se), is another important area of chemical speciation activity.

In biology and medicine, identification and characterization of species of essential (e.g., Fe, Cu, Zn, Se), toxic (e.g., Hg, Pb, Cd, As), and therapeutic elements (e.g., Pt, Au) in living organisms is arousing great interest these days. Also, in occupational medicine, speciation provides information about the volatile species (e.g., Hg) and inhalable particles at the workplace (e.g., Cr(VI) in dust particles), providing information about trace element toxicants' absorption, distribution, reactivity, toxicity, and the final excretion after an occupational exposure to Pb, Cr, As, etc.

Distribution, transport, metabolism, and detoxification studies of both essential and toxic trace elements via speciation information are becoming popular in ecotoxicology as well.

On the other hand, the importance of trace element speciation is also increasing so as to help understand the bioavailability of inorganic nutrients (Fe, Cu, Zn, Se) from food. Of course, intelligent administration of food supplements (e.g., the assimilation of selenium in living organisms and its metabolic processes closely depend on Se chemical form) demands more and more information on the actual species present in the supplement.

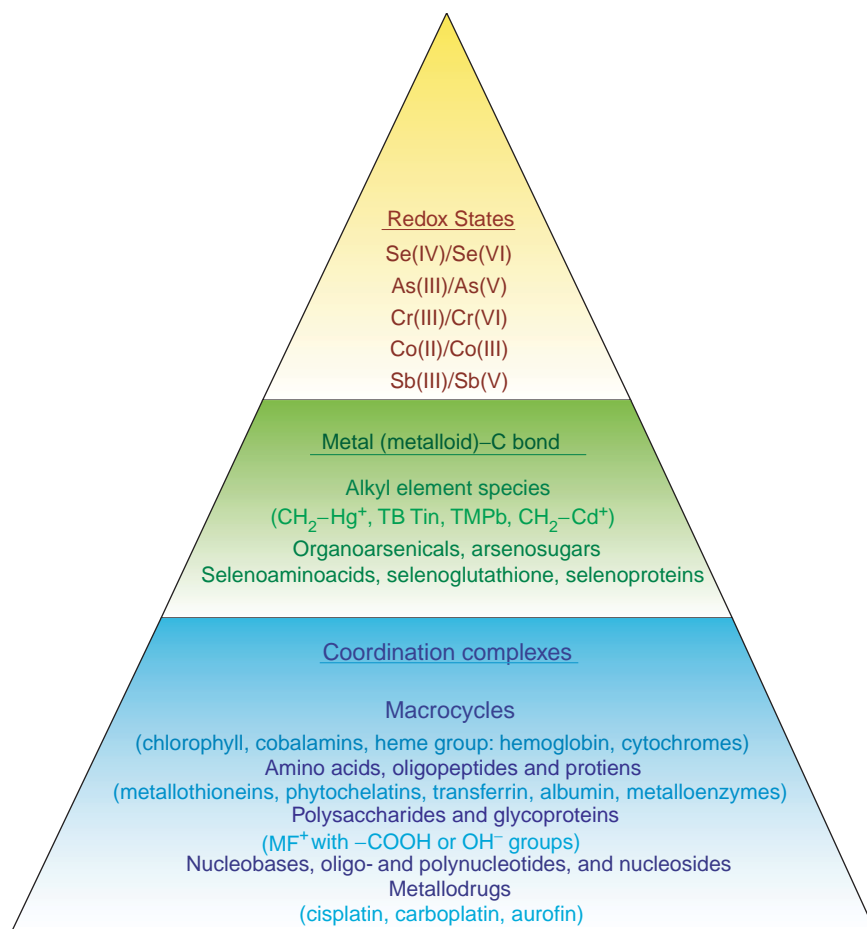
Finally, metal species analysis is becoming of great practical importance in industrial processes. Some examples are food industry speciation studies, required for determining actual toxicity of contaminated foodstuff (e.g., fish by methylmercury, oysters and mussels by organotin or arsenic, or wine by organolead). In the petrochemical industry speciation of Fe, V, Co, or Ni geometalloporphyrins of crude oil is increasingly demanded. Also, in the pharmaceutical industry and in the production of packing materials, speciation and speciation analysis become increasingly important (e.g., organometallic compounds can be used as catalysts or polymer stabilizers like tributyltin (TBT), etc.).

In brief, as shown in Figure 2, analytes or compounds of interest in speciation analysis are becoming important in many fields. They include redox states, organometallic compounds (with a strong metal-carbon covalent bond), and also coordination complexes with any kind of ligands including small compounds (e.g., in metallodrugs) and big metal-biomolecules (formed in the living organisms as a result of the metabolism of the element, e.g., Cd-metallothioneins).

## The Analytical Challenge of Trace Element Speciation

As speciation information is becoming very important, analytical methods and strategies to obtain such information are urgently needed and are now being widely developed.

However, trace element speciation analyses constitute a real challenge for present analytical chemistry, due to their complexity, which stems from the following aspects: (1) The first problem is related to the 'lability' of the target species: changing the ambient conditions from those prevailing in the original systems may often disturb the existing physicochemical equilibria and so the resulting speciation information may not be fully representative of that originally present in the sample. (2) Sensitive and selective detectors are required to measure the extremely low concentrations, naturally found in the real-life samples. This is



**Figure 2** Typical analytes sought in speciation analysis.

particularly difficult after the fractionation of the different metal forms needed for speciation. Of course, except for highly contaminated samples, speciation of toxic metals is even more difficult because their content is usually at micrograms per liter levels or below. (3) The trace amounts of the species sought are 'buried' in a very complex matrix (e.g., serum, sediment, soft tissue, etc.).

According to the chemical nature of the species to be analyzed, three different categories of speciation analysis have been proposed:

- Category I. Thermodynamically stable and kinetically inert species (e.g., organometallics such as MeHg, TBT, selenoproteins, and Se-aminoacids).
- Category II. Thermodynamically stable but kinetically labile or dynamic species. That is, compounds or species having a relatively high stability constant in solution (e.g., Al-transferrin, Cd-metlothionein, oxidation states), which have to be carefully preserved in order to prevent their alterations (reaction).
- Category III. Weakly bound and dynamic species like the so-called 'labile-monomeric aluminum', present in natural water (e.g., Al<sup>3+</sup>, Al(OH)<sup>2+</sup>, Al(OH)<sup>+</sup>, AlCl<sup>2+</sup>, etc.), which are, in fact, a group of chemical species instead of a defined one. All of them behave in a 'group-similar' manner that can provide useful information (e.g., 'labile monomeric aluminum' for the toxicity of acidified waters to fish).

Most of methods developed for analytical speciation involve several steps: species extraction, preconcentration, cleaning, derivatization, separation, and detection. Such process could affect the integrity of the species sought. Of course, accurate speciation analysis requires preserving the integrity of the species sought along the whole analytical process, from sampling to final measurement. Since species lability increases from Category I to III, the analytical requirements needed to preserve the integrity of the analyte (species) increases in the same order.

## Analytical Strategies and Techniques

A large number of analytical techniques and strategies have been developed to tackle modern speciation problems. There are four main approaches possible: computational, direct species-specific detection, hybrid techniques, and physicochemical characterization techniques.

### Computational Approach

This theoretical approach consists of numerical thermodynamic calculations of the species present in the sample. Knowing the pH, ligand and metal concentrations, temperature, redox potential, and formation constants, it is possible to predict the speciation of an element in a given sample. This approach, however, has several limitations because in most cases we do not know all the ligands' concentrations or the thermodynamic constant values. What it is more, kinetic aspects are not considered in such models, the system is usually considered to be homogeneous.

Notwithstanding the above constraints, computational methods are useful for a first approximation for a speciation problem or to predict speciation changes caused by changes in a parameter of the environmental conditions. Published practical examples of speciation by using a computational approach are the aluminum speciation in serum and distribution of trace elements in fresh waters.

### Direct Species-Specific Techniques

The direct species-specific techniques would be the ideal speciation procedure, providing identification and quantification of a given individual species *in situ* without any sample treatment. The techniques of this type, more used for individual metal-containing species determination, are electroanalytical in nature (e.g., ion-selective electrodes have been used for determining major free alkali ion determinations in natural waters and serum). More sensitive voltamperometric redissolution techniques can be useful for speciation of free and 'labile' forms of metals (or their different oxidation states) in waters.

In general, electrochemical techniques lack the required selectivity to be used for most ultratrace and trace element levels present in complex real samples (e.g., environmental and biological matrices).

Molecular techniques based on ultraviolet (UV)–visible spectrophotometry and fluorimetry have been also proposed for direct determination of a specific element form. Colored species, such as those of Cr or Mn, can be directly determined spectrophotometrically. However, most of such species will require the use of a selective color-forming reagent, lacking the sensitivity, and often the selectivity, required for

speciation analysis. More recently, investigations are directed to develop the analytical potential of biosensors that could be adequate for *in situ* monitoring of species without sample disturbance.

## Hyphenated Techniques and Selected Speciation Examples

Without discussion, the tools most commonly used for trace element speciation analysis nowadays are 'hyphenated techniques', which is the coupling of more than one technique: one or more of them facilitating the separation of the various forms of the element of interest and the other for the final sensitive and selective detection of the element. The main separation methods and detectors commonly used in hyphenated or coupled techniques for speciation analysis have been summarized in Figure 3. These species separation techniques can be coupled online or offline with the detector (usually an atomic spectroscopic detector).

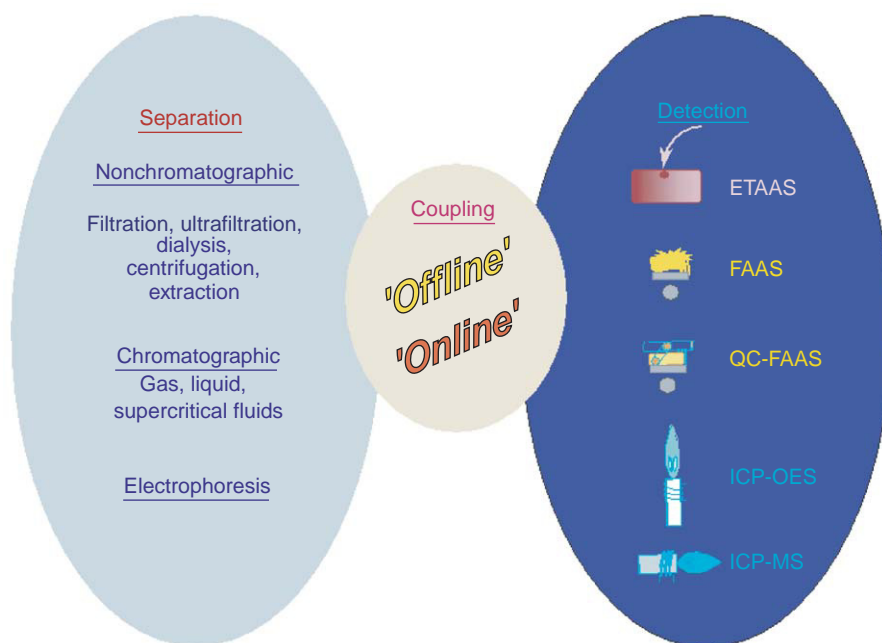
### Speciation by 'Offline' Couplings

Notwithstanding the fact that a continuous or 'online' coupling is desirable for speciation purposes, there are examples of the usefulness of both discontinuous nonchromatographic (e.g., ultrafiltration) and continuous separation techniques (e.g., chromatography), which have been coupled offline with atomic detectors to solve important problems requiring speciated information.

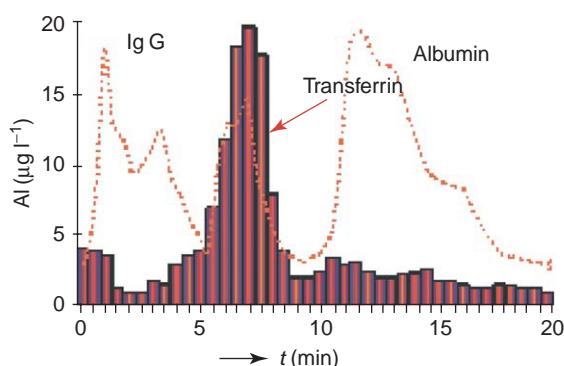
An interesting illustrative example of 'offline' speciation usefulness is aluminum speciation in human serum. Aluminum has been shown to be very toxic to renal failure patients suffering long-term hemodialysis. Ultrafiltration using 30 kDa membranes and electrothermal atomic absorption spectrometry (ETAAS) detection was employed to demonstrate that ~90% of Al in human serum is bound to proteins. Therefore, only 10% of serum aluminum is ultrafiltrable. Moreover, the Al-detoxification mechanism occurring by treating dialysis patients with the drug Desferrioxamine before dialysis was clarified with such offline techniques. Also, the nature of Al and Si binding proteins in serum has been established using 'offline' liquid chromatography (LC) followed by ETAAS. Most of Al present in serum is bound to transferrin, while Si is weakly bound to several serum proteins (Figure 4).

### Speciation by 'Online' Couplings

Nonchromatographic and chromatographic separation methods coupled 'online' with the atomic detector used exhibit great advantages compared with 'offline' systems, including less sample manipulation,



**Figure 3** Possibilities of hyphenation of different separation techniques to different types of atomic detectors for speciation analysis. FAAS=Flame atomic absorption spectrometry; QC=quartz tube concentrator.



**Figure 4** Aluminum speciation by LC-ETAAS in diluted (1 + 4) human serum. (Reprinted from Soldado Cabezuelo AB, Montes Bayon M, Blanco Gonzalez E, *et al.* (1998) Speciation of basal aluminium in human serum by fast protein liquid chromatography with inductively coupled plasma mass spectrometric detection. *Analyst* 123: 865–869.)

less contamination risk, easy automation, and, in general, more precise and reliable results.

**Nonchromatographic separation coupled to atomic detectors** Flow injection analysis (FIA) strategies may provide an efficient, continuous, and automatic method for preconcentration and/or separation of the sought species from a complex matrix, prior to its final determination. Furthermore, FIA systems can be interfaced easily with atomic detectors making them very convenient in speciation analysis. A typical application is selective preconcentration on

minicolumns of the desired species (e.g., of As(III),  $\text{Pb}^{2+}$ ,  $\text{Hg}^{2+}$ ,  $\text{Cd}^{2+}$ ) in seawater. Species contained in a few milliliters of the sample are injected in the minicolumn allowing for straightforward preconcentration factors of 50–100. Several separation techniques, including dialysis, ion-exchange, and liquid–liquid extraction, can be used for speciation applications in nonchromatographic arrangements coupled to the atomic detector via FIA strategies.

In this vein, many applications have appeared in the literature on speciation of oxidation states, such as Cr(III)/Cr(VI), Fe(II)/Fe(III), As(III)/As(V), Se(IV)/Se(VI). Also, FIA techniques have been used to preconcentrate Sn, Hg, and Pb organometallics mainly from natural waters, sediment, and soil extracts. In particular, great interest has been focused into the speciation of chromium oxidation states in water samples at very low level (nanograms per liter). A FI system with a minicolumn of acidic preconcentration and inductively coupled plasma optical emission spectrometry (ICP-OES) for final detection was developed for a rapid speciation of Cr(VI) and Cr(III) in waters. On sample injection, Cr(VI) is retained in the alumina column whilst Cr(III) is not passing directly to the atomic detector. Afterwards, the retained Cr(VI) is eluted by injection of ammonium hydroxide, as shown in Figure 5, and its analytical signal of emission in the ICP-OES is registered.

An alternative nonchromatographic separation such as continuous liquid–liquid extraction can also provide a simple and effective speciation approach.

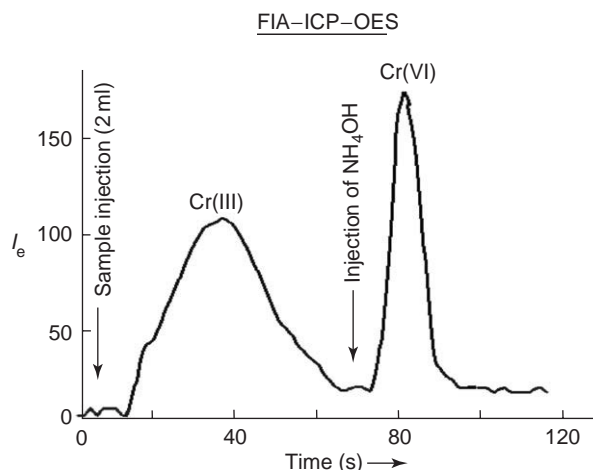


After the liquid separation, hydride generation can be used 'online' to increase the sensitivity of the final detection of the volatile species formed in a continuous manner (e.g., speciation of  $\text{Hg}^{2+}$  and  $\text{MeHg}$ , of  $\text{As(III)/As(V)}$  or  $\text{Sb(III)/Sb(V)}$ ).

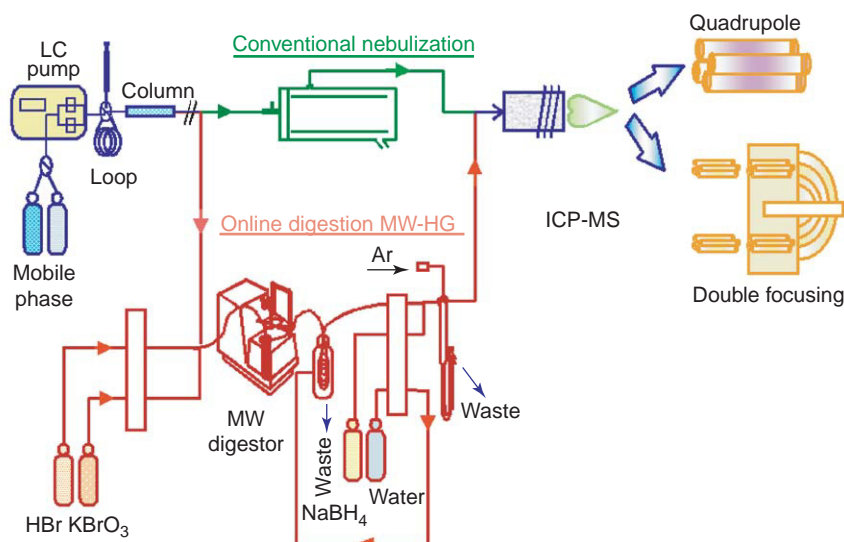
**Chromatographic separation coupled to atomic detectors** Chromatographic separation techniques, both gas chromatography (GC) and LC, have been

coupled to element-specific detectors to form the most powerful and versatile tool for speciation analysis (see **Figure 3**). Flames, heated quartz tubes (QC), microwave plasmas (MIPs), ICP-OES, and ICP-MS have been applied as specific detectors linked to LC or GC. MIP is the source of choice for GC, because of their high excitation efficiency for nonmetallic elements such as S and C, whereas ICP-OES and ICP-mass spectrometry (MS) are the recommended detectors for speciation using LC, because of their multielement capabilities and great sensitivity analyzing liquid samples (in the range of micrograms per milliliter and nanogram per milliliter, respectively, for most of the elements). Moreover, the liquid flow of LC, as well as the gas flow exiting the column in GC, can be directly coupled to an ICP. Also, hydride generation (HG) online atomic detectors can be extremely useful to increase the sensitivity of the final detection in LC-ICP-MS studies. A typical LC-(HG)-ICP-MS coupling proposed for sensitive selenium speciation in urine is given in **Figure 6**.

Whereas the choice of the detection technique is mainly determined by the analyte level in the sample, the selection of the more adequate separation technique is determined by the particular physicochemical properties of the analyte. As shown in **Scheme 1**, for thermally stable and volatile analytes (or able to produce them by easy derivatization) GC is the preferred separation technique. HG or reaction with sodium tetraethylborate is one of the most suitable derivatization procedures. GC has been used

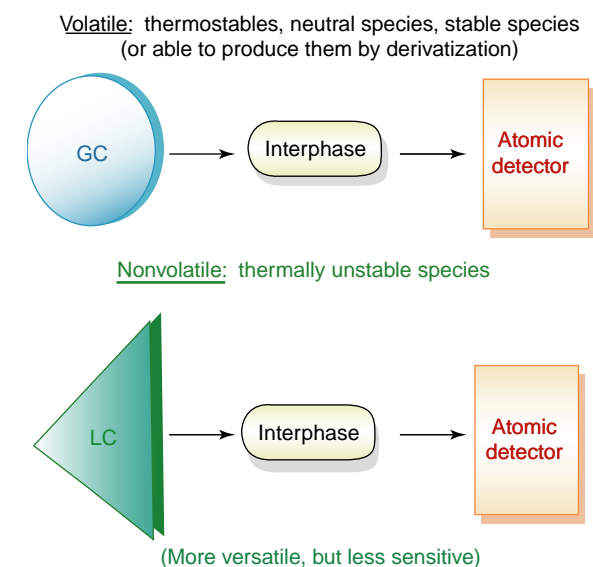


**Figure 5** Chromium speciation in water by FIA-ICP-OES. (Reprinted from Cox AG, Cook IG, and McLeod CW(1985) Rapid sequential determination of chromium(III)-chromium(VI) by flow injection analysis-inductively coupled plasma atomic emission spectrometry. *Analyst* 110: 331-333.)



**Figure 6** Schematic diagram of online coupling LC-ICP-MS via conventional nebulization and via microwave (MW)-HG. (Reprinted from Gonzalez LaFuente JM, Marchante Gayon JM, Fernandez Sanchez ML, and Sanz-Medel A (1999) Urinary selenium speciation by high-performance liquid chromatography-inductively coupled plasma mass spectrometry; advantages of detection with a double focussing mass analyser with a hydride generation interface. *Talanta* 50: 207-217.)

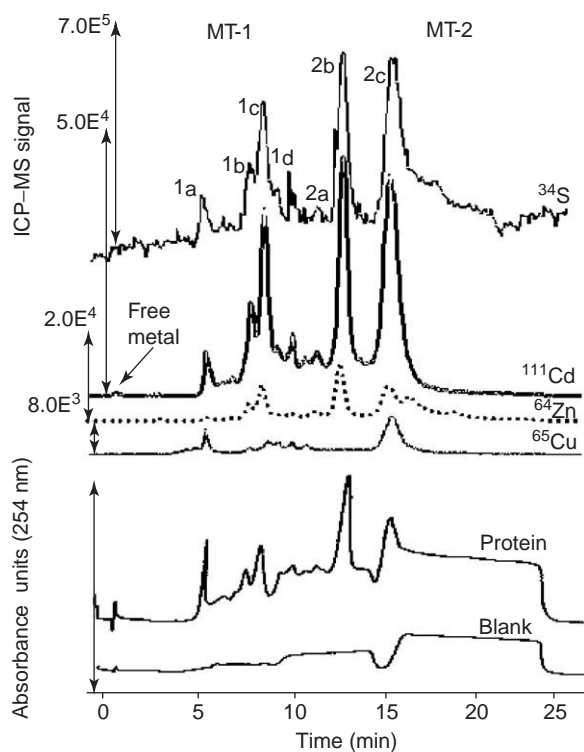




Scheme 1

extensively for identification and determination in environmental samples of different molecular forms of the organometals of Pb, such as tetraalkyllead, dialkyllead ( $R = Et$  or  $Me$ ), used as antiknock compounds in gasoline. Similarly, organotin derivatives (tetraalkyltin, trialkyltin, dialkyltin, and monoalkyltin), widely used as antifouling paints or polymer stabilizers, have been analyzed throughout GC-atomic detectors. As stated before, hydride generation and reaction with sodium tetraethylborate are the most suitable derivatization procedures used for converting the metals and organometals of Pb, Sn, and Hg into volatile compounds for eventual GC separation. Speciation of inorganic- and methyl-mercury in environmental samples has also been carried out using such derivatizations ( $NaBH_4$  forms a 'cold vapor' of mercury with  $Hg^{2+}$  and a hydride with  $CH_3Hg^+$ ).

For nonvolatile, thermolabile, or charged metal-containing compounds LC is preferred. LC is more versatile than GC owing to the diversity of possible separation mechanisms in an LC column: reverse-phase, ion-exchange, size-exclusion chromatography (SEC) are the most used. Ion-pair formation, chiral and affinity chromatography have also been employed. The choice of a particular separation mechanism depends on the analytical problem to be solved: for instance reverse-phase chromatography is recommended for separation of hydrophobic biomolecules and their complexes (e.g., cobalamins, organoarsenic, and organoselenium compounds). Ion exchange is an ideal technique for separation of small charged organometallic ions (e.g., organoarsenic, organoselenium species, etc.) and also for charged



**Figure 7** Multi-element speciation of MTs by FPLC coupled online to ICP-MS and UV-visible spectrometry. (Reprinted from Ferrarello CN, Fernandez de la Campa MR, Carrasco JF, and Sanz-Medel A (2000) Speciation of metallothionein-like proteins from the mussel *Mytilus edulis* at basal levels by chromatographic separations coupled to quadrupole and double focussing magnetic sector. ICP-MS. *Analytical Chemistry* 72: 5874–5880.)

species of biomolecules (e.g., Cd-metallothioneins). Of course, size exclusion should be used for a fractionation in terms of the molecular size of the metal-bioligands (e.g., proteins, polypeptides, etc.). So far other mechanisms such as ion-pair formation, chiral recognition, or affinity chromatography are rather scarce in the speciation analysis literature.

In recent years, many analytical strategies for elemental speciation analysis, particularly in (bio)medical speciation, make use of LC-atomic detector hybrid techniques. Most frequently studied elements include As, Pb, Cd, Sn, Hg, Se, Cr, and Al both in biological and environmental samples. The capability for multi-elemental detection of ICP-MS can be most useful in such analysis. An illustrative example of multi-element trace-element speciation is metallothionein proteins (MT) from rabbit liver by LC-ICP-MS. Comparative profiles for Cd, Zn, Cu, and S obtained for MTs from rabbit liver, using a fast protein liquid chromatography (FPLC) column coupled to ICP-MS are shown in Figure 7, using molecular and specific detectors.

Capillary electrophoresis (CE) offers several advantages as compared to LC for bioligands separation including: low sample volume requirements, high efficiency, ability to separate positive, neutral (with special buffers), and negatively charged species in one single run and at comparatively high speed. In principle, a combination of the advantages of CE for the convenient separation with the high sensitivity and selectivity of ICP-MS detection would provide a powerful technique for elemental speciation. Even if the sensitivity of such CE-ICP-MS systems is poor, this coupling has now been applied to the analysis of a wide variety of compounds, including small metal ions, amino acids, protein digests, protein mixtures, and large neutral or electrically charged biomolecules.

In any case, speciation analyses by LC-ICP-MS and GC-ICP-MS (if possible) seem to provide so far the more straightforward, reliable, and sensitive approach for real-life speciation analysis in complex environmental, biological, or clinical samples.

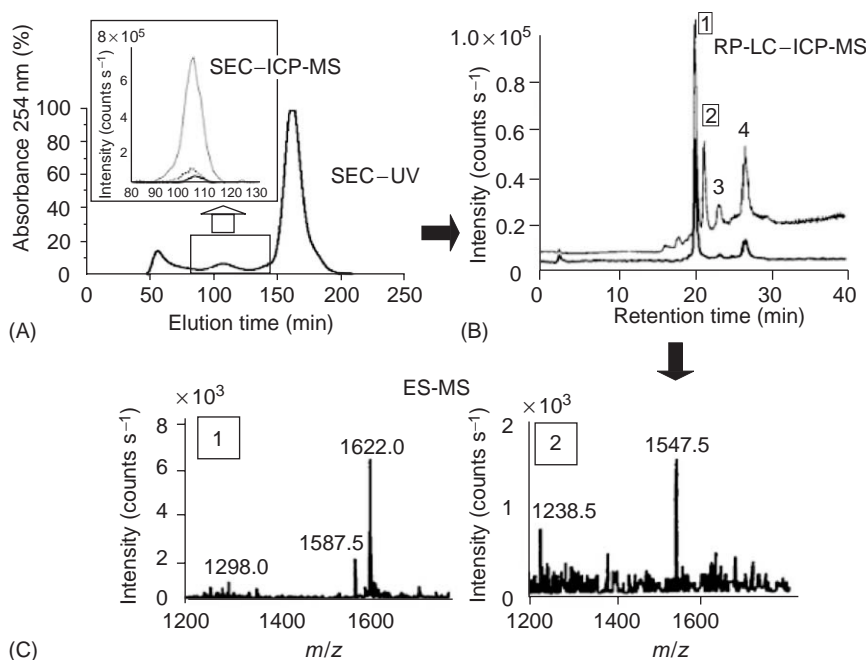
## Likely Developments

There is a worldwide recognition that the chemical, toxicological, biological, and clinical properties of a given element are critically dependent on the actual

chemical form in which the element occurs. The need for 'speciated' information has prompted the development of different analytical strategies and techniques to provide such information. Among them, the combination of a chromatographic separation with a sensitive and element-specific detector has become the more popular and powerful tool for real-life speciation analysis. The initial use of such hybrid techniques to obtain speciated information in environmental issues (i.e., speciation of anthropogenic organometallics of Pb, Hg, or Sn) has been extended now to the difficult speciation of trace elements in biological material of the most varied living organisms.

Speciation analysis has been one of the major developing areas of analytical chemistry during the last 15 years and a great number of new developments are expected in this analytical field in the near future. In particular, some areas where important developments can be envisaged include:

- Development of simpler analytical methodology and commercial instrumentation. Elemental speciation analysis is not yet routine, due to lengthy and cumbersome analytical protocols. Moreover, many methods have been developed for expensive MS detectors, only available in a limited number of laboratories. Simple and low-cost



**Figure 8** New trends in speciation: Separation of Cd-metalllothionein isoforms by: (A) SEC with UV and ICP-MS detection, (B) reversed-phase-LC-ICP-MS of the main Cd containing species eluted by SEC, (C) ES-MS of the peaks labeled as 1 and 2 in the RP-LC-ICP-MS chromatogram. (Reproduced from Polec K, García-Arribas O, Pérez-Calvo M, *et al.* (2000) Identification of cadmium-bioinduced ligands in rat liver using parallel HPLC-ICP-MS and HPLC-electrospray MS. *Journal of Analytical Atomic Spectrometry* 15: 1363–1368.)

instrumentation and reliable methods are needed to achieve routine speciation analysis. Future developments in this area are expected, but they are closely related with the introduction of commercial instrumentation for analytical speciation by instrument manufacturers.

- Developments of quality assurance of the analytical speciation procedures and results are of prime importance at this point. Validation of new methodologies should be carried out if possible with certified reference materials (CRMs) (with a matrix as similar as possible to the real sample certified for the sought species). Unfortunately, very few CRMs are commercially available today for chemical speciation. Thus, alternative approaches for validation using complementary methods based on different mechanisms of separation and complementary detectors are now in order.
- Improvements in the sample pretreatment procedures. Sample cleanup, preconcentration, and preservation procedures of natural species are urgently needed. There is no guarantee that transformation of elemental speciation along the whole analytical procedure does not take place. In this sense, the use of isotopically labeled compounds using an adequate separation technique coupled to ICP-MS is being established. It seems clear that the use of isotopically enriched species, synthesized for the purpose, will become an important tool for validation of speciation results.
- Identification/confirmation of the species. A natural extension of the present popular use of ICP-MS to investigate metal-containing biomolecules (e.g., to 'screen' new metal-biocompounds) is to identify and confirm the whole chemical nature of the new species. To do so, molecular-specific detection techniques such as matrix-assisted laser desorption/ionization time-of-flight or electrospray-tandem mass spectrometry, following a powerful electrophoretic or chromatographic separation, are required. Figure 8 illustrates with a typical example the new trends in speciation analysis to identify/confirm the bioligands associated to the element of interest.

The concept of chemical speciation and speciation analysis is continuously evolving in such a way that the discovery and molecular elucidation of new (semi) metal-biomolecules are now possible via speciation analysis with hyphenated techniques. Of course, a discovered metal-biomolecule can be the signature of a new biochemical reaction mechanism paving

the way to deeper understanding of biochemical pathways of the element. Thus, the fate of trace and ultratrace elements in living organisms and their role in health and disease (even at such minute amounts) could be more properly understood via speciation information.

**See also:** **Capillary Electrophoresis:** Overview. **Elemental Speciation:** Waters, Sediments, and Soils. **Gas Chromatography:** Overview. **Liquid Chromatography:** Overview. **Proteins:** Overview. **Quality Assurance:** Quality Control.

## Further Reading

- Caruso JA, Sutton KL, and Ackley KL (eds.) (2000) *Elemental Speciation. New Approaches for Trace Element Analysis*. Amsterdam: Elsevier.
- Chassaigne H, Vacchina V, and Lobinski R (2000) Review. Elemental speciation analysis in biochemistry by electrospray mass spectrometry. *Trends in Analytical Chemistry* 19: 300–313.
- Ebdon E, Pitts L, Cornelis R, Crews H, and Donard OFX (eds.) (2001) *Trace Element Speciation for Environmental Food and Health*. Cambridge: Royal Society of Chemistry.
- Ferrarelo CN, Fernández de la Campa MR, and Sanz-Medel A (2002) Multielement trace-element speciation in metal-biomolecules by chromatography coupled with ICP-MS. *Analytical Bioanalytical Chemistry* 373: 412–421.
- Harrison RH and Rapsomanikis S (eds.) (1989) *Environmental Analysis using Chromatography Interfaced with Atomic Spectroscopy*. Chichester: Ellis Horwood.
- Lobinski R (1997) Elemental speciation and coupled techniques. *Applied Spectroscopy* 51: 260A–278A.
- Ruiz Encinar J, Rodríguez-González P, García Alonso JJ, and Sanz-Medel A (2003) Isotopically labelled compounds for validation organometallics speciation analysis. *Trends in Analytical Chemistry* 22: 108–114.
- Sanz Medel A (1998) Trace element analytical speciation in biological systems: Importance, challenges and trends. *Spectrochimica Acta B* 53: 197–211.
- Sanz-Medel A (ed.) (1999) *Flow Analysis with Atomic Spectrometric Detectors*. Amsterdam: Elsevier.
- Szpunar J (2000) Review: Bio-inorganic speciation analysis by hyphenated techniques. *Analyst* 125: 963–988.
- Szpunar J, Lobinski R, and Prange A (2003) Review. Hyphenated techniques for elemental speciation in biological systems. *Applied Spectroscopy* 57: 102A–112A.
- Templeton DM, Ariese F, Cornelis R, et al. (2000) Guidelines for terms related to chemical speciation and fractionation of elements. Definitions, structural aspects, and methodological approaches. *Pure and Applied Chemistry* 72: 1453–1470.

## Waters, Sediments, and Soils

**G E Batley and S C Apte**, CSIRO Energy Technology, Bangor, NSW, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Speciation analysis in waters, sediments, and soils is the measurement of the different chemical forms of trace constituents such as metals, organometals, and inorganic ions. Interest in speciation analysis is driven by the fact that analyses of total element concentrations fail to reflect the separate chemical, physical, and biological properties that individual species might possess. Assessing the toxicity and bioavailability of trace elements is a major focus of speciation analysis as transport across the cell membranes of living organisms (bioavailability) and potential toxicity vary greatly with chemical form. Historically, the greatest focus of speciation analyses has been on metals (and metalloids).

Speciation analysis can involve the measurement of discrete chemical compounds (e.g., methylmercury chloride, tributyltin oxide, arsenobetaine), or in some cases different oxidation states; or the identification of common classes of compounds that might be operationally defined on the basis of their reactivity, size, or some other physicochemical property. Speciation analysis methods are generally more complex than procedures used for measuring total metal concentrations. This review focuses primarily on trace elements; the diverse speciation of carbon (i.e., organic chemistry) and the analysis of organic compounds are not covered.

In natural waters, trace elements can be present in a variety of species (**Figure 1**). Colloidal particles containing mixtures of hydrous iron and manganese

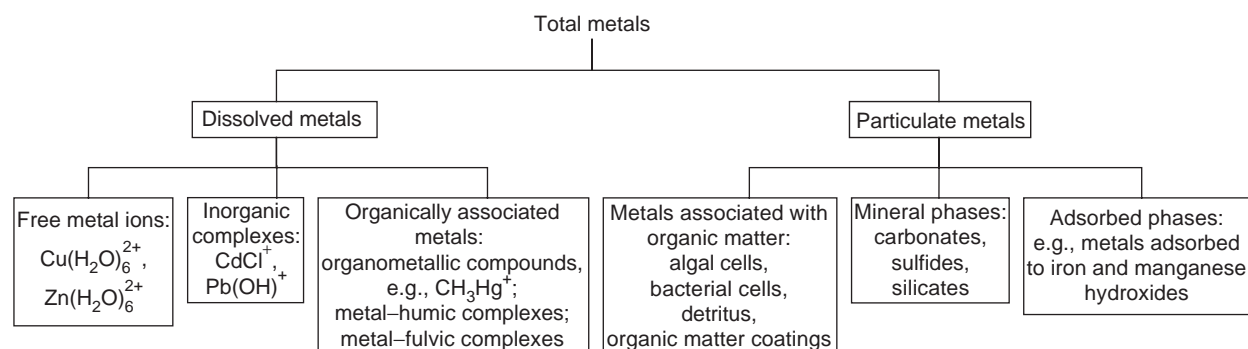
oxides, and organic macromolecules represent an important compartment for many trace elements. Generally a poor correlation exists between total trace metal concentrations and biological effects, and for several metals, bioavailability is better correlated with the concentration of simple aquated ion and inorganic complexes.

In aquatic sediments or soils, there are also a range of trace elements species ranging from ions exchanged to particles, to those bound to organic matter or in various inorganic forms (e.g., oxides, carbonates, sulfides) or as more inert crystalline mineral phases. As in waters, speciation studies in soils and sediments are generally undertaken to better understand the bioavailability of toxic substances and to investigate transport pathways to and from other parts of the ecosystem. Sediment and soil pore waters (soil solutions) are of particular interest because they are in equilibrium with the solid phase and are the medium for contaminant uptake by plants and many other biota. The techniques used for speciation analysis in these aqueous samples differ little from those for waters.

### Analytical Methods

#### Sample Collection and Storage

Trace elements may be present in natural samples at concentrations of parts per billion and lower. The challenge with speciation analysis is frequently the contamination-free sampling of these matrices in a manner that does not perturb solution chemistry and speciation. The stability of samples during storage is a critical issue confronting speciation analysis. Biota and solid samples are generally preserved frozen. For waters, refrigeration at 4°C and storage in the dark is generally thought to be the best storage procedure as



**Figure 1** Types of metal species in natural waters.



it slows chemical reactions and biological processes. Freezing is not generally recommended as it can rupture cells and cause precipitation of some constituents. An alternative approach is the selective preconcentration of certain metal species by immobilization onto selective adsorbents in the field. This is particularly appropriate for the analysis of redox-sensitive species. Given the uncertainties in this area, it is highly recommended that validation tests be carried out to characterize the effects of storage on sample stability.

Concerns over the validity of sample storage have stimulated interest in the development of *in situ* methods for trace element speciation. This includes microelectrodes and various sensors. Examples of this measurement approach are given below.

### Separations Based on Size

One of the easiest separations is that based on size. For waters, these largely divide into the use of filtration or diffusion-based separations.

**Filtration** Traditionally, aquatic chemists have operationally defined 'dissolved' species by filtration through a 0.45 µm pore size membrane. The term 'dissolved' is erroneously used to describe the filterable metal fraction. Filterable and nonfilterable metals are more accurate and appropriate terms. For water samples, filtration through small pore size (0.01–0.45 µm) membranes can be used to separate small particulate and even colloidal species. Ultrafiltration that involves the use of small pore size filters and pressurized filtration has been used to separate macromolecules. Ultrafilters are available with molecular weight cutoffs (MWCO) typically between 1000 and 50 000. Cross-flow ultrafiltration, where water flow is parallel to the surface of the membrane is used for the efficient isolation of colloidal material from large volume samples. Adsorption losses to the large surface area membranes can be an issue and several methods employ correction factors to account for such effects.

**Dialysis** Dialysis membranes can be used to separate species on the basis of their differing rates of diffusion across semipermeable membranes. The size discrimination of commonly used dialysis membranes ranges from typically 14 000 to 1000 MWCO. Dialysis involves equilibration of a large volume of sample (donor) with a small receiving water sample (receptor) contained by a semipermeable membrane. A net diffusive flux occurs until concentrations of the dialyzable species are the same on either side of the membrane, irrespective of any adsorption losses. Equilibrium dialysis where the

sample volume is large and the receptor solution volume is small (e.g., 20–50 ml) is preferred for metal speciation studies as it causes minimal disturbance of bulk solution equilibria. The technique has been applied both in the laboratory and *in situ* in the field, but is not in widespread usage.

**Donnan dialysis** In Donnan dialysis, a cation-exchange membrane separates the donor and receptor solutions. Cationic metal species are transported across the membrane driven by the negative electrostatic potential (the Donnan potential) across the membrane, until equilibrium is achieved. Matching of the ionic strengths of donor and acceptor solutions is necessary. Since cationic species exchange readily compared to neutral and anionic species, it is claimed that the measurement more closely relates to the free metal ion.

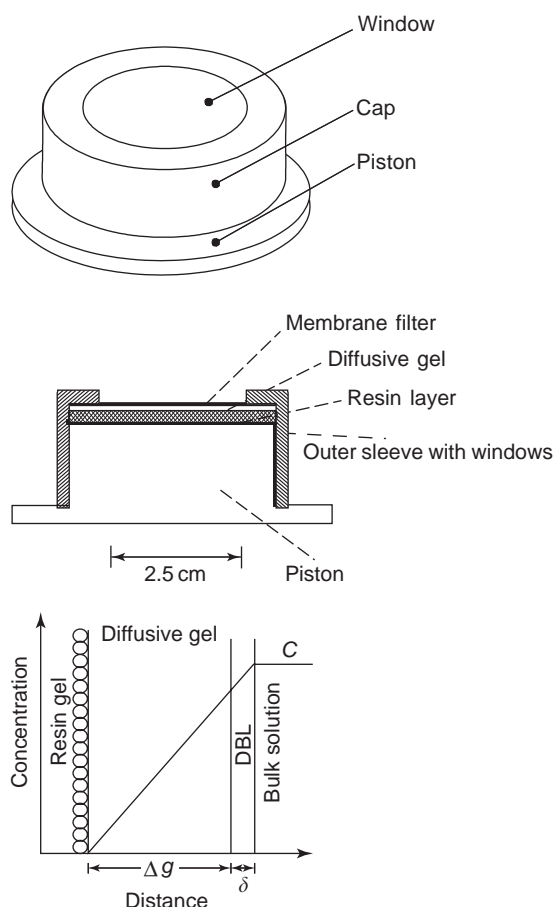
The technique has recently found application in studies of metal speciation in soil solutions and pore waters. A Teflon or Kel-F cell is employed with a 30–50 ml donor solution separated from a 200 µl acceptor solution by a strong acid cation-exchange membrane (e.g., Nafion-117). Unless metal concentrations are high, the necessary detection limits may be difficult to achieve.

**Diffusive gradients in thin films** In addition to size and charge, diffusion-limited separation of species can be achieved using a polymer membrane of adequate thickness and pore size overlying a membrane impregnated with a chelating resin. Diffusive gradients in thin films (DGT) is gaining popular acceptance for speciation studies in both laboratory and field, for waters, aquatic sediments, and wetted soils.

A typical DGT device for use in waters consists of a gel impregnated with a chelating resin, covered by a diffusive hydrogel (0.4–0.8 mm thick polyacrylamide) topped with a 100 µm thick, 0.45 µm pore size cellulose nitrate or polysulfone membrane, and contained in a plastic holder with a 2 cm diameter window (Figure 2). For sediment and soil monitoring another sampler that allows sectioning of fine vertical strips is used. Typically, DGT samplers might be deployed for 24–72 h, at the end of which the Chelex gel is removed and placed in a known volume of dilute nitric acid. The elutriate is subsequently analyzed for trace metals by a sensitive technique such as inductively coupled plasma mass spectrometry (ICP-MS).

In waters, the flux of metal ions through the diffusive gel (= accumulated mass  $M$ /(area of membrane  $A \times$  exposure time  $t$ )) will be related to the bulk solution concentration  $C$ , by the relationship:

$$M/At = DC/\Delta g$$



**Figure 2** Schematic diagram of a DGT sampler designed for water quality monitoring and schematic representation of the free concentration profile of ionic species in the sampler when in contact with an aqueous solution. DBL is the diffusive boundary layer. (Reprinted with permission from Zhang H and Davison W (1995) Performance characteristics of diffusion gradients in thin films of the *in situ* measurement of trace metals in aqueous solution. *Analytical Chemistry* 67: 3391–3400; © American Chemical Society.)

where  $D$  is the diffusion coefficient in the gel and  $\Delta g$  the gel thickness.

The diffusive boundary layer in solution also controls the flux of metals and must be kept as small as possible by stirring the solution, or relying on natural mixing in the field. DGT samplers can be deployed in the field or in more controlled conditions in the laboratory. In sediments, the interpretation of metal flux data is more complex as it has to take into account resupply of metals to pore waters by desorption from sediment particles.

DGT methods measure a labile metal fraction, which is a subset of the total dissolved metals concentration. The proportion of the dissolved metal fraction determined will depend on the rate and extent of metal complex dissociation in the presence of

the chelating resin and the diffusion coefficients of the individual metal species. The species that can be measured by the gel are determined by the diffusion layer thickness and the pore size of the gel. Labile metal complexes that can dissociate within the time taken to diffuse through the gel ( $<1$  min) will be measured, together with any inert metal complexes that bind directly to the chelating resin. By changing the composition of the gel and its pore size, it is possible to retard the diffusion of large molecular weight species. This increases the discriminatory capabilities of DGT.

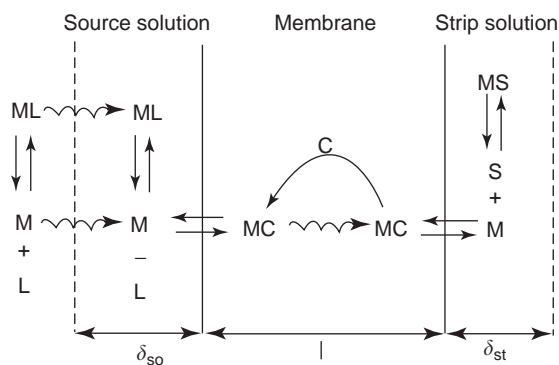
**Permeation liquid membranes** The permeation liquid membrane technique is one of the newest approaches to metal speciation in waters. It attempts to mimic the transport of metal species across a biological membrane, typically using a supported liquid membrane consisting of an organic solvent containing an organic carrier molecule that is selective for the metal of interest and held in the pores of a hydrophobic chemically inert membrane. The membrane is placed between the sample solution and a receiving or stripping solution. One version of this procedure involves the use of a microporous polypropylene membrane (PLM) impregnated with an equimolar mixture of dodecyl-1, 10-diaza crown ether and lauric acid dissolved in a toluene/phenylhexane mixture.

Cyclohexaneethylenediaminetetraacetic acid is used in the stripping solution. Depending on the flow rate and accumulation time used, the flux of metal species across the membrane yields information on the free metal ion and/or labile metal species concentrations in the original solution. PLM has been used in a number of configurations that are rapidly evolving from the original large cell concept (Figure 3) to hollow fiber devices and miniaturized versions. Experiments with model ligands show good agreement with calculated free metal ion concentrations. Measurements using a hollow fiber PLM with a 120 min *in situ* preconcentration have measured picomolar free copper and lead ion concentrations in a range of natural waters.

### Differentiation of Oxidation States

Some elements exist naturally in more than one oxidation state. This applies particularly to metals such as chromium (Cr(III) and Cr(VI)), manganese (Mn(II), Mn(IV), and Mn(VII)), iron (Fe(II) and Fe(III)), antimony (Sb(III) and Sb(V)), thallium (Tl(I) and Tl(III)), arsenic (As(III) and As(V)), selenium (Se(IV) and Se(VI)). The elemental form of many elements can coexist with the ionic forms, but in only a





**Figure 3** Schematic diagram illustrating the operation of a permeation liquid membrane device. (Reprinted with permission from Tomaszewski L, Buffle J, and Galceran J (2003) Theoretical and analytical characterization of a flow-through permeation liquid membrane with controlled flux for metal speciation measurements. *Analytical Chemistry* 75: 893–900; © American Chemical Society.)

few cases is this considered important (e.g.,  $\text{Hg}_0$ ,  $\text{Se}_0$ ). The determination of oxidation states is important for elements such as As, Se, and Cr, because toxicity and reactivity can vary with oxidation state (e.g.,  $\text{As(III)} > \text{As(V)}$ ;  $\text{Cr(VI)}$  is more toxic than  $\text{Cr(III)}$ ). This may be achieved by selective coprecipitation, complexation with oxidation-state specific ligands, electrochemical analysis, or species-selective derivatization and separation by volatilization or liquid–liquid extraction. For example, controlling reaction pH it is possible to selectively evolve hydrides from  $\text{As(III)}$  and  $\text{Se(IV)}$  in the presence of higher oxidation state species. Solid phase adsorbents (e.g., chemically modified silicas) have been used for the selective preconcentration of specific oxidation states.

### Ion-Exchange and Chelating Resin Separations

Equilibration of solutions with various resins has been used to define labile and nonlabile metals species. Contact time is an important factor controlling the split between labile and nonlabile metal species. Batch equilibration, where a fixed mass of resin is added to a sample and then removed after a specified contact time, is suited for timescales of minutes to days whereas passage of samples through resin-filled columns allows much shorter contact times of typically seconds.

Alone, chelating resins such as Chelex-100 in columns have been used to separate weak metal complexes, but have not been widely applied in speciation studies. More recently success has been obtained with the use of tuned Chelex columns, where the column volume is minimized and solution flow rate is maximized, while still giving quantitative recovery of ionic metals. Lability is influenced by

contact time and the nature of the counter ion. Strong cation ion-exchange resins have been used in aluminum speciation studies to separate labile (monomeric) from nonlabile species.

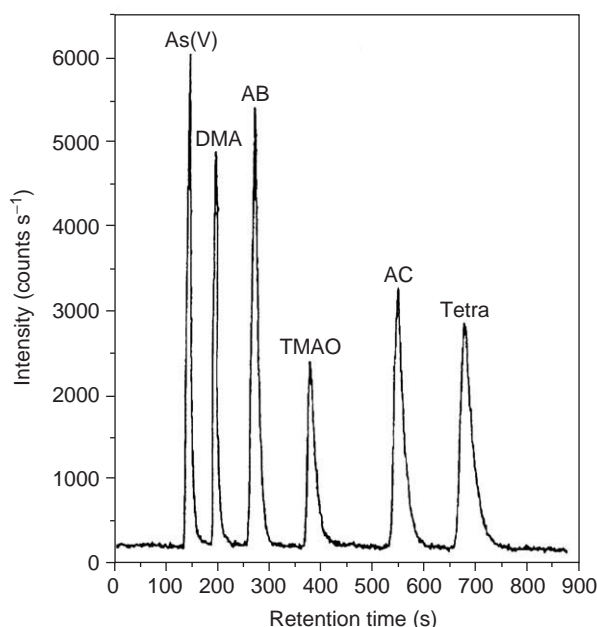
A number of equilibrium ion-exchange methods have been developed to measure free metal ion concentrations in waters. This approach involves the equilibrations of a small volume of resin with a sample. Following the attainment of equilibrium, the amount of metal adsorbed is measured. Calibration is achieved by using matrix-matched solutions of known free metal ion concentration.

### Hyphenated Techniques Involving Chromatographic Separation

A range of chromatographic techniques coupled to element specific detectors has been used in speciation studies to separate individual organometallic species (e.g., butyltins, arsenic species) and to separate metals bound to various biomolecules. The combination of a chromatographic separation with varying instrumental detection systems are commonly called coupled, hybrid, or hyphenated techniques (e.g., liquid chromatography inductively coupled plasma–mass spectrometry (LC–ICP–MS), gas chromatography–atomic absorption spectroscopy (GC–AAS)). The detection systems used in coupled techniques include MS, ICP–MS, atomic fluorescence spectrometry (AFS), AAS, ICP–atomic emission spectrometry (ICP–AES), and atomic emission detection (AED).

Various types of LC have been employed including reversed-phase liquid chromatography (RP–LC), ion chromatography, gel permeation, and affinity chromatography. An important consideration is the compatibility of the eluant with the detection system. Usually aqueous media or low concentrations of miscible organic solvents (e.g., methanol) are used. Using LC, different organometallic species of the same metal are amenable to reversed-phase separations using nonaqueous solvent mixtures and hydrophobic columns (e.g.,  $\text{C}_{18}$ ). Examples include organolead or organotin species analysis. The separation of ionic species, e.g.,  $\text{As(III)}$ ,  $\text{As(V)}$ , monomethylarsonic acid, and dimethylarsinic acid, is best achieved using anion-exchange columns. The separation of various arsenic compounds including arsenobetaine is shown in **Figure 4**. Ion-exchange separations have also been successfully applied to the separation of selenium, chromium, and mercury species in waters.

A major problem with coupled LC procedures is achieving the necessary detection limits for these species, especially in waters. Either some form of preconcentrations is required or the detection system

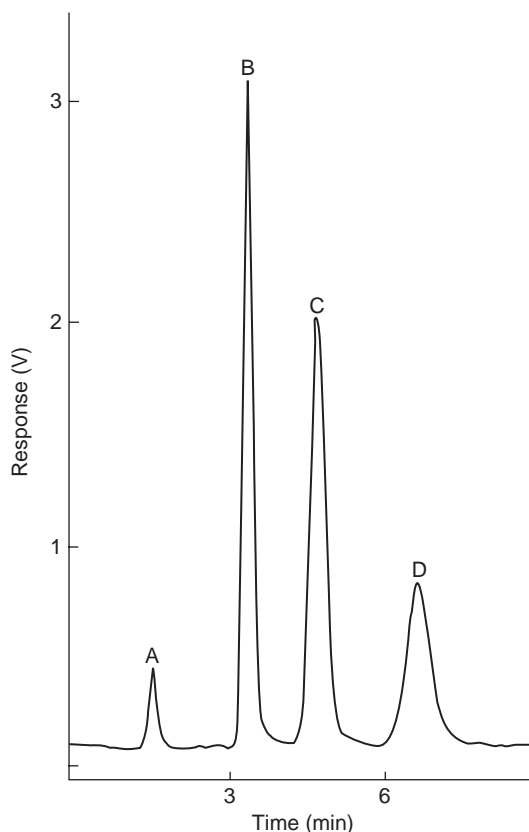


**Figure 4** Separation of arsenic species by LC-ICP-MS. The chromatogram shows the separation of As(V) dimethylarsenic acid (DMA), arsenobetaine (AB), trimethylarsine oxide (TMAO), arsenocholine (AC), and tetramethylarsonium ion (tetra) on a HHPN Nucleosil 5SA cation-exchange column (mobile phase  $30 \text{ mmol l}^{-1}$  pyridine, pH 3.0, injection volume  $20 \mu\text{l}$ , flow rate  $1 \text{ ml min}^{-1}$ ,  $25^\circ\text{C}$ ). The concentration of each species was  $2 \mu\text{g As per liter}$ . (From Falk K and Emons H (2000) Speciation of arsenic compounds by ion-exchange HPLC-ICP-MS with different nebulizers. *Journal of Analytical Atomic Spectrometry* 15: 643–649; reproduced by permission of The Royal Society of Chemistry.)

needs high sensitivity. The most sensitive detection system for metals is ICP-MS.

Capillary electrophoresis has also been used for separation of organometallic species, although detection limits are poor owing to the very small sample volumes used in this technique. Polyacrylamide gel electrophoresis has been coupled to hydride generation (HG)-AFS to separate selenium in various protein fractions.

GC is widely used to separate volatile organometallic species. The favored approach involves derivatization (e.g., ethylation or hydride formation), preconcentration, chromatographic separation followed by element-specific detection using AAS, AES, or AFS. For species such as methylmercury and methylarsenic compounds, derivatization is normally necessary prior to GC in order to form volatile species. The most common approach has been to convert both inorganic and organometal species via hydridization with sodium borohydride, or alkylation, e.g., using sodium tetraethylborate, to form volatile hydrides or alkylated species, respectively. Not all species of an element are amenable to derivatization. For example, the application of hydridization



**Figure 5** Analysis of butyltin species in coastal waters. Typical chromatogram of an MBT, DBT, and TBT (each  $75 \text{ ng Sn}$ ), analyzed by GC-AAS following aqueous phase ethylation with sodium tetraethylborate. Retention times: A, tetraethyltin 1.7 min; B, triethylmonobutyltin 3.4 min; C, diethyldibutyltin 4.7 min; D, ethyltributyltin 6.6 min. (Reprinted from Bowles KC, Apte SC, and Hales LT (2003) Determination of butyltin species in natural waters using aqueous phase ethylation and off-line room temperature trapping. *Analytica Chimica Acta* 477: 103–111; © with permission from Elsevier.)

to arsenic in a fish tissue extract would not convert arsenobetaine to a hydride. Preconcentration of volatile species, either on solid adsorbents at room temperature or colder (cryogenic trapping using liquid nitrogen). Volatile forms, once thermally desorbed, are also more amenable to sensitive spectrometric detection systems, including atomic absorption, atomic emission, or atomic fluorescence. For these systems, atomization of the separated species is first required, and this has been achieved by a variety of furnace systems, most commonly a quartz tube furnace, but ceramic, tantalum, graphite, plasma, and other variants have also been used. The coupling of effective derivatization, preconcentration, and sensitive detection has led to measurements of subnanogram per liter concentrations of species such as tributyltin (Figure 5) and methylmercury in natural waters.

## Electrochemical Techniques

**Voltammetric techniques** Voltammetric techniques have been widely applied to determine metal contaminants in aqueous media. They involve reduction (or oxidation) at a working electrode with the current associated with the reduction being measured as a function of the electrode potential. Techniques such as differential pulse polarography at a dropping mercury electrode or voltammetry at a hanging mercury drop electrode (HMDE) have detection limits near  $10^{-8} \text{ mol l}^{-1}$ . Greater sensitivity can be achieved using stripping voltammetric methods such as anodic stripping voltammetry (ASV) and cathodic stripping voltammetry (CSV).

In ASV, stripping (oxidation) of the reduced metal is achieved by the application of an anodic potential scan. This technique is applicable to metals that are reversibly reduced at mercury electrodes, including Cu, Pb, Cd, Zn, Bi, Sb, In, and Tl. Detection limits will be dependent on the deposition (accumulation) time, the nature of the stripping waveform (usually either direct current, differential pulse, or square wave), and the electrode type. HMDEs are widely used, although greater sensitivity may be obtained using a mercury film deposited on a glassy carbon substrate. Under optimum conditions, detection limits of  $10^{-11} \text{ mol l}^{-1}$  are achievable. Carbon fiber or iridium wire microelectrodes have also been successfully used as mercury film substrates. They are ideal for small volumes and for high resistance (low ionic strength) solutions.

CSV involves the addition of a ligand of known thermodynamic properties to the solution, equilibration and the accumulation of an adsorbed metal complex on a mercury electrode and its reduction by a cathodic potential scan. CSV measures only the metal species that will react with the ligand, either at the electrode or in solution. The method has found application in ligand competition methods for speciation analysis as discussed below.

Both ASV and CSV measure labile metal concentrations (assumed to approximate the concentration of inorganic metal species and weakly bound organic complexes) and have been used to elucidate the extent of complexation by dissolved organic matter for a range of metals. The value of voltammetric techniques in metal speciation studies lies in the fact that discrimination can be achieved on the basis of their reactivity at the electrode. This will include not only the free metal ion, but those in complexed forms (ML), provided the dissociation of the complex in the diffusion layer is very fast compared to diffusion, i.e., the thermodynamic equilibrium of the dissociation of ML is maintained despite consumption of M at the

electrode surface. Complexes that meet this criterion are termed labile, those that do not are termed inert.

**Ion selective electrodes** Measurement of the free metal ion is an important goal of metal speciation studies in natural waters, because of the tenet of the free ion activity model, that bioavailability is proportional to the free metal ion concentration. Ion selective electrodes respond to the free metal ion activity. A range of ion selective electrodes are available, but only the copper electrode has sufficient sensitivity for use in measurement at realistic environmental concentrations in natural waters. The correct use of the copper ion selective electrode remains a matter of debate, especially its calibration at the very low free metal ion concentrations using metal ion buffers and the applicability of this procedure to natural samples.

## Ligand Competition Techniques

There have been a number of approaches to the use of ligand competition (or exchange) reactions for speciation analysis in solution.

Equilibrium ligand competition has been most widely used in combination with stripping voltammetric techniques to determine the free metal ion concentrations and the relative importance of other metal binding ligands in natural waters. For example, for copper, ligand competition with catechol was measured by differential pulse cathodic stripping voltammetry (DPCSV). For cadmium, ethylene diamine was used as the ligand with differential pulse ASV detection, and for nickel, dimethylglyoxime with DPCSV. The methods involve the addition of a known amount complexing ligand, and titrating the sample with added ionic metal, allowing appropriate equilibration before DPCSV measurements. The titration data allow calculation of the free metal ion concentration in the sample and the concentrations and binding constants of the other major ligands in the sample, usually simplified to a one- or two-ligand complexation models. This approach has found greatest application in elucidating metal speciation in seawater.

The reaction of metal species with added colorimetric ligand has also been utilized to gain speciation information. A classic example is the use of pyrocatechol violet as a spectrophotometric reagent to determine labile monomeric (fast reacting) and organic aluminum species.

## Modeling Speciation in Waters

There are a number of computer codes available for calculating solution equilibria and the speciation of trace metals in natural waters. Programs such as

MINEQL and MINTEQA2 are now widely used by aquatic scientists. Similar models (e.g., GEOCHEM and SOILCHEM) are used by soil scientists. The effective use of such computational models requires critical knowledge of key reactions and judicious selection of the appropriate stability constants. Recent research endeavors in this area have been directed toward describing metal interactions with natural organic matter (NOM). This is an immense challenge owing to the polydisperse, heterogeneous nature of NOM. Biotic ligand models (BLM) have recently been developed which predict acute metal toxicity to some aquatic organisms. BLM models assume that metal toxicity is related to free metal ion binding to a specific ligand at the cell surface of unicellular organisms or at the gill surface in fish. The models take into account proton, calcium, and magnesium ion competition for the binding sites and assume a simple relationship between the bound metal concentration and toxicity.

### Speciation of Inorganic Ions

The speciation of nitrogen and phosphorus are important areas of nutrient chemistry in both waters and soils. Most methods are based on colorimetry with sample pretreatment. For waters, flow injection analysis techniques are usually preferred because of their ability to handle high sample throughput. These species are usually individually determined following standard analytical procedures. Analytical methods are available for the determination of nitrate, nitrite, and ammonia. Only ammonia exerts toxic effects on aquatic biota and then only in the unionized form. Phosphorus in waters exists mainly as orthophosphate, but also as polyphosphates or organophosphorus compounds. It is normal to measure total and reactive phosphorus. Sulfur speciation is also of interest, the species including sulfate, sulfite, thiosulfate, sulfur, hydrogen sulfide, and other sulfides and polysulfides. As for nutrients, sulfur speciation analysis has used a range of routine analytical procedures. Most methods employ ion chromatography or colorimetry.

Cyanide speciation is of particular interest in industrial effluents. Weak acid dissociable cyanide species are determined by distillation following acidification of the sample and measurement of the trapped cyanide in alkaline solution. Individual metal cyanide species may also be determined by LC.

## Speciation Analysis in Solids

### Direct Measurement of Speciation

Some surface speciation information may be obtained using X-ray photoelectron spectroscopy, Auger

spectroscopy, and secondary ion mass spectrometry (SIMS). The use of these high vacuum techniques may cause artifacts owing to sample dehydration. X-ray absorption near edge spectrometry and extended X-ray absorption fine structure are very promising direct measurement techniques, which give information on solid and solution phase speciation with reasonable sensitivity. Using a combination of both techniques, it is possible to deduce oxidation states, the identity of nearest neighbor atoms, and bond distances. Sample preparation is minimal, however, these procedures are restricted in their applications as they require a synchrotron to generate high energy radiation.

### Selective Dissolution Procedures

A number of operational procedures are available to determine potentially mobilizable species in solids. These involve the extraction of samples with acid, alkali, or a complexing agent. With respect to metal bio-availability, a more useful extraction is to use either cold  $1 \text{ mol l}^{-1}$  hydrochloric acid or  $0.5 \text{ mol l}^{-1}$  ethylenediaminetetraacetic acid, on the basis that this will not react with the more intractable phases. The measured values probably exceed that which might be released by at the lowest gut pH of any organisms, but it is at least an improvement on 'total' metal measurements. Various enzyme 'cocktail' mixtures have also been used to mimic metal mobilization in the digestive tracts of organisms.

The most sophisticated approaches involve sequential chemical leaches with extractants that have selectivity for particular mineral phases. For metals, those 'fractions' traditionally measured include:

- ion-exchangeable surface metals (removed with a  $\text{MgCl}_2$ , ammonium acetate, or other salt solution);
- carbonate minerals (removed with dilute acetic acid);
- metals associated with easily reducible iron and manganese oxides (using acidified hydroxylamine hydrochloride);
- metals bound to organics or sulfides (released by hydrogen peroxide); and
- a residual inert, mineralized fraction that requires strong acid digestion (nitric and hydrofluoric acids).

There are many problems with carryover and read-sorption that can make these procedures 'poorly selective', and the value of the speciation information is often questionable, although it has been extremely widely applied.

A recent speciation interest with respect to sediments has been the role of reactive sulfides, notable



FeS, in moderating metal bioavailability. Provided it is present in excess, FeS will exchange with soluble metal ions forming insoluble metal sulfides (e.g., Cu, Pb, Zn, Cd). Acid volatile sulfide (AVS) is measured by reaction with dilute acid under nitrogen. Under these conditions, ordered phases such as pyrite do not react. An excess of soluble (potentially bioavailable) metals will only be seen in the pore waters if the exchange capacity of the FeS for these metals is exceeded, so provided there is an FeS excess there should be no toxicity due to pore water metals. It is common to assess this from the difference between the molar AVS concentration and the sum of the molar concentrations of metals that are released simultaneously in the dilute acid treatment.

For organometallic species, the analytical challenge involves quantitative extraction of the species from the solid phase without appreciable transformation to other forms. Various liquid extractants have been used such as methanolic hydrochloric acid (extraction of organotins). Strong alkali digests (KOH or tetramethylammonium hydroxide) will dissolve most tissue without breakdown or transformation of arsenic species. As an alternative to liquid extraction, methylmercury species may be extracted from solids using steam distillation. Specific reaction conditions will affect the recovery of various species and rigorous validation of extraction procedures is therefore recommended.

## Quality Control in Speciation Analysis

Quality control (QC) is as important in speciation analyses as it is in other areas of analysis. While there are a growing number of certified reference materials available that contain specific forms of an element, for example, tributyltin, methylmercury, this is the exception rather than the rule. Instead, other standard QC approaches need to be used. These include the use of spike-recovery tests and internal standards, comparisons with the results from other methods,

and interlaboratory comparisons. Standard protocols have yet to be documented for many speciation methods, and if these are to be incorporated in regulations as they are now starting to be, considerations of appropriate certified reference materials and QC measures need to be given high priority.

**See also:** **Arsenic. Flow Injection Analysis:** Environmental and Agricultural Applications. **Membrane Techniques:** Dialysis and Reverse Osmosis; Ultrafiltration; Liquid Membranes; Pervaporation. **Selenium. Sulfur.**

## Further Reading

- Allen HE, Huang C-P, Bailey GW, and Bowers AR (1999) *Metal Speciation and Contamination of Soil*. Boca Raton, FL: Lewis Publishers.
- Batley GE (ed.) (1989) *Trace Element Speciation: Analytical Methods and Problems*. Boca Raton, FL: CRC Press.
- Buffle J, Parthasarathy N, Djane N-K, and Matthiasson L (2000) Permeation liquid membranes for field analysis and speciation of trace compounds in waters. In: Buffle J and Horvai G (eds.) *In Situ Monitoring of Aquatic Systems: Chemical Analysis and Speciation*, pp. 407–494. New York: Wiley.
- Caroli S (ed.) (1996) *Element Speciation in Bioinorganic Chemistry*. New York: Wiley.
- Cornelis R, Caruso J, Crews H, and Heumann K (2003) *Handbook of Elemental Speciation: Techniques and Methodology*. Chichester: Wiley.
- Davison W, Fones G, Harper M, Teasdale P, and Zhang H (2000) Dialysis, DET and DGT: *In situ* diffusional techniques for studying water, sediments and soils. In: Buffle J and Horvai G (eds.) *In Situ Monitoring of Aquatic Systems: Chemical Analysis and Speciation*, pp. 496–569. New York: Wiley.
- Ebdon L, Pitts L, Cornelis R, *et al.* (2001) *Trace Element Speciation for Environment, Food and Health*. Cambridge: Royal Society of Chemistry.
- Tessier A and Turner DR (1995) *Metal Speciation and Bioavailability in Aquatic Systems*. Chichester: Wiley.
- Ure AM and Davidson CM (eds.) (1992) *Chemical Speciation in the Environment*. London: Blackie.

## Practicalities and Instrumentation

**D J Butcher**, Western Carolina University, Cullowhee, NC, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Elemental speciation is the quantitative determination of individual chemical forms of an element.

Considerable interest has developed in speciation since the early 1960s involving the biological and environmental sciences when it was observed that the toxicity and mobility of metals is dependent upon their chemical form. Speciation may also involve the characterization of the bioavailability of an element from complex matrices such as soil or sediments.

This article provides an overview of the major issues involving elemental speciation in environmental

FeS, in moderating metal bioavailability. Provided it is present in excess, FeS will exchange with soluble metal ions forming insoluble metal sulfides (e.g., Cu, Pb, Zn, Cd). Acid volatile sulfide (AVS) is measured by reaction with dilute acid under nitrogen. Under these conditions, ordered phases such as pyrite do not react. An excess of soluble (potentially bioavailable) metals will only be seen in the pore waters if the exchange capacity of the FeS for these metals is exceeded, so provided there is an FeS excess there should be no toxicity due to pore water metals. It is common to assess this from the difference between the molar AVS concentration and the sum of the molar concentrations of metals that are released simultaneously in the dilute acid treatment.

For organometallic species, the analytical challenge involves quantitative extraction of the species from the solid phase without appreciable transformation to other forms. Various liquid extractants have been used such as methanolic hydrochloric acid (extraction of organotins). Strong alkali digests (KOH or tetramethylammonium hydroxide) will dissolve most tissue without breakdown or transformation of arsenic species. As an alternative to liquid extraction, methylmercury species may be extracted from solids using steam distillation. Specific reaction conditions will affect the recovery of various species and rigorous validation of extraction procedures is therefore recommended.

## Quality Control in Speciation Analysis

Quality control (QC) is as important in speciation analyses as it is in other areas of analysis. While there are a growing number of certified reference materials available that contain specific forms of an element, for example, tributyltin, methylmercury, this is the exception rather than the rule. Instead, other standard QC approaches need to be used. These include the use of spike-recovery tests and internal standards, comparisons with the results from other methods,

and interlaboratory comparisons. Standard protocols have yet to be documented for many speciation methods, and if these are to be incorporated in regulations as they are now starting to be, considerations of appropriate certified reference materials and QC measures need to be given high priority.

**See also:** **Arsenic. Flow Injection Analysis:** Environmental and Agricultural Applications. **Membrane Techniques:** Dialysis and Reverse Osmosis; Ultrafiltration; Liquid Membranes; Pervaporation. **Selenium. Sulfur.**

## Further Reading

- Allen HE, Huang C-P, Bailey GW, and Bowers AR (1999) *Metal Speciation and Contamination of Soil*. Boca Raton, FL: Lewis Publishers.
- Batley GE (ed.) (1989) *Trace Element Speciation: Analytical Methods and Problems*. Boca Raton, FL: CRC Press.
- Buffle J, Parthasarathy N, Djane N-K, and Matthiasson L (2000) Permeation liquid membranes for field analysis and speciation of trace compounds in waters. In: Buffle J and Horvai G (eds.) *In Situ Monitoring of Aquatic Systems: Chemical Analysis and Speciation*, pp. 407–494. New York: Wiley.
- Caroli S (ed.) (1996) *Element Speciation in Bioinorganic Chemistry*. New York: Wiley.
- Cornelis R, Caruso J, Crews H, and Heumann K (2003) *Handbook of Elemental Speciation: Techniques and Methodology*. Chichester: Wiley.
- Davison W, Fones G, Harper M, Teasdale P, and Zhang H (2000) Dialysis, DET and DGT: *In situ* diffusional techniques for studying water, sediments and soils. In: Buffle J and Horvai G (eds.) *In Situ Monitoring of Aquatic Systems: Chemical Analysis and Speciation*, pp. 496–569. New York: Wiley.
- Ebdon L, Pitts L, Cornelis R, *et al.* (2001) *Trace Element Speciation for Environment, Food and Health*. Cambridge: Royal Society of Chemistry.
- Tessier A and Turner DR (1995) *Metal Speciation and Bioavailability in Aquatic Systems*. Chichester: Wiley.
- Ure AM and Davidson CM (eds.) (1992) *Chemical Speciation in the Environment*. London: Blackie.

## Practicalities and Instrumentation

**D J Butcher**, Western Carolina University, Cullowhee, NC, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Elemental speciation is the quantitative determination of individual chemical forms of an element.

Considerable interest has developed in speciation since the early 1960s involving the biological and environmental sciences when it was observed that the toxicity and mobility of metals is dependent upon their chemical form. Speciation may also involve the characterization of the bioavailability of an element from complex matrices such as soil or sediments.

This article provides an overview of the major issues involving elemental speciation in environmental



samples. Following a discussion of several example elements, sampling, sample preparation, and calibration procedures are discussed. Instrumentation for speciation is reviewed, including separation techniques and detection methods. The conclusion is a detailed discussion of a particular application of speciation involving environmental samples.

## Speciation of Example Elements

Metal compounds exist in a wide variety of chemical forms that may be classified as inorganic, complexed ions, or organometallic. A variation in the toxicity of different inorganic oxidation states exists for some metals. Perhaps most widely studied element with this characteristic is chromium. Chromium has been introduced into the environment from a variety of industrial processes that include tanning, chrome plating, stainless steel welding, spray painting, dye manufacturing, and cleaning aircraft and automobile parts. Chromium(III) is an essential element at trace concentrations that is characterized by low uptake by biological organisms. However, chromium(VI) is considerably more available and toxic to biological organisms as chromate or dichromate. It has also been classified as a carcinogen by the US Environmental Protection Agency (EPA).

Organometallic compounds may have very different environmental toxicities than inorganic compounds. Alkylated mercury compounds (e.g., methyl mercury) are far more toxic than inorganic mercury (although these species are also regarded as toxic). A number of anthropogenic activities release mercury into the environment, which may then be methylated by microorganisms. Although the concentrations of mercury may be in the low part-per-billion range in the environment (e.g., the ocean) and the methylation process is relatively inefficient (1%), methyl mercury may accumulate to micrograms per gram levels in top predators in the aquatic food chain.

Organotin compounds (e.g., tributyl tin (TBT)) are of considerable interest for environmental speciation because of their past use as pesticides, algicides, fungicides, and molluscicides. TBT was employed as the active ingredient on boat hulls to prevent the accumulation of barnacles and other fouling organisms that increase fuel consumption and maintenance costs. These compounds accumulated to toxic levels in shellfish and fish, with the first evidence appearing in France in the late 1970s and early 1980s. Legislation was passed in France and other nations that had restricted the use of TBT-based paints. This is in sharp contrast to elemental and inorganic tin, which are characterized by low toxicity and very low rates of accumulation into living organisms.

Arsenic is well known because of its widespread use as a poison in criminal activity. Arsenic has also been widely used for agricultural applications, which include herbicides, pesticides, and wood preservatives. These applications have been reduced in recent years due to concerns about its toxicity, and the US EPA has recently reduced the acceptable limit of arsenic in drinking water. Unlike mercury and tin, organometallic forms of arsenic, such as arsenobetaine, arsenocholine, and some arsenosugars, are relatively nontoxic, while inorganic arsenic(III) (arsenite) and arsenic(V) (arsenate) are highly toxic.

Manganese is the 12th most common element and is characterized by a variety of oxidation states, including Mn(VII), Mn(VI), Mn(IV), Mn(III), and Mn(II). It is most commonly found as the mineral, pyrolusite ( $\text{MnO}_2$ ). Manganese is used in a variety of industrial applications (Table 1). Its most significant applications include metallurgical processes, primarily steel manufacturing, the manufacture of dry-cell batteries, and as a coloring agent for the brick industry. Other applications include glass (coloring agent), chemical industry (dye and paint production), and electronics industry (magnets). An organomanganese compound, methylcyclopentadienyl manganese

**Table 1** Industrial applications of manganese

<i>Chemical form</i>	<i>Application</i>	<i>Function</i>
Mn	Steel manufacture	Scavenger of sulfur and oxygen; increase hardness
$\text{MnO}_2$	Dry-cell batteries	Acts as depolarizer to prevent liberation of hydrogen
$\text{MnO}_2$	Brick	Colorant
$\text{MnO}_2$	Glass	Decolorizer
$\text{MnO}_2$	Hydroquinone production	Oxidizing agent
$\text{MnO}_2$	Ceramic ferrites	Magnets
$\text{KMnO}_4$	Synthesis of chemicals (saccharin, benzoic acid)	Oxidizing agent
Methylcyclopentadienylmanganese tricarbonyl	Gasoline, aircraft fuels	Antiknock agent

tricarbonyl (MMT), has been widely used as an anti-knock agent in aircraft fuel and gasoline.

Although manganese is considered a nutritionally essential trace element, it is known that very high levels of inhaled manganese induce a neurological disorder called manganism. This disease has been most commonly associated with miners, who are exposed to manganese via inhalation, although there have been reports of exposure through food and water. Considerable scientific study of manganism has been performed in recent years because of its similarities with Parkinson's disease. A number of studies have considered the environmental and toxicological impacts of the use of the fuel-additive MMT because of its higher toxicity compared to inorganic manganese compounds.

## Sampling, Sample Preparation, and Calibration

As speciation methodology has developed over the past 40 years, it has become clear that quality control issues such as sampling, sample preparation, and calibration are critical to obtain meaningful results regarding environmental samples. Many early studies were flawed by errors in these methodologies that produced inaccurate data. Here some of the major issues in sample handling and treatment are discussed.

### General Considerations of Environmental Sampling

Considerable planning needs to be performed to develop a sampling procedure that is representative of the environmental system of interest. First, it is necessary to collect samples from sites that are representative for the study. One should also make certain to collect appropriate types of materials (e.g., atmosphere, hydrosphere, biosphere, pedosphere, lithosphere) to answer the environmental question of interest. A representative number of specimens for each material need to be sampled and subsequently analyzed. Although ideally this value should be calculated based on the known values of the standard deviation of the samples, in many cases the variance is unknown and it is necessary to select the number of samples based on an estimated value. Lastly, it is necessary to account for weather conditions and natural seasonal variations in the concentrations of metal species in biological samples.

It is essential to maintain information about each of the chemical species of interest during all analytical procedures. This can be achieved by ensuring that no changes occur during sampling, storage, and analysis, or by quantitative derivatization of the

**Table 2** Concentration of example elements in laboratory glass and plasticware

<i>Material</i>	<i>Concentration range (<math>\mu\text{g per g}</math>)</i>	
	100	10–0.1
Glass	Al, K, Mg, Mn, Sr	Fe, Pb, B, Zn, Cu, Rb, Ti, Ga, Cr, Zn
Polyethylene/polypropylene	Na, Zn, Ca, Al, Ti	K, Br, Fe, Pb, Cl, Si, Sr
PTFE	K, Na	Cl, Na, Al, W

analytes into stable compounds. Also, most speciation studies involve trace or ultratrace analysis, so it is critical to avoid contamination and loss of analytes throughout sampling, storage, and analytical procedures. This includes the use of clean sampling tools and collection vials that will not induce contamination for the analytes and by the use of careful cleaning protocols. Examples of elements present in measurable concentrations in laboratory glass and plasticware are listed in **Table 2**. For example, when doing speciation analysis of zinc, zinc-doped stoppers must be avoided. Samples to be analyzed for mercury should be stored in glass.

One approach to investigate contamination of collection tubes is by the analysis of a certified reference material, after storage for a length of time representative of the study, to see if the certified value is obtained. However, there are relatively few elements certified at these low concentration levels. Alternatively, the sample containers may be rinsed with acid or other suitable solutions, followed by determination of the analyte concentration in the leachate. Although easy to perform, this procedure may not accurately represent the amount of contaminants picked up by the samples.

### Collection and Storage of Specific Environmental Samples

Although it is beyond the scope of this article to consider detailed collection and sample preparation procedures for environmental samples, here are some general considerations for speciation analysis of air, water, biological material, and soils/sediments. The collection of samples in air has focused upon the determination of hydrides and methylated compounds of elements such as arsenic, lead, mercury, and tin. The sample may be introduced into a preevacuated container, or with a pump. The analytes may be isolated using cryotrapping techniques or adsorbent cartridges. For best results, it is recommended to collect the entire gas sample in a container, coated

with an inert material such as polytetrafluoroethylene (PTFE), followed by analysis within 24 h.

The specifics of sampling and storage for water are dependent upon the sample. Clearly, different procedures are required for groundwater, seawater, waste water, and atmospheric precipitation. Samples should be collected at a reproducible location and depth below the surface. In general, the contamination control procedures discussed above should be followed. Although it may be necessary to remove bacteria and other suspended solids from samples using membrane filters, it is necessary to investigate these methods for possible contamination or influence upon the types of species present. Acidification of water samples may also induce chemical transformations of some elements.

Biological samples are particularly challenging because of wide variations in the accumulation of metal compounds in different organisms, differences in accumulation in various tissues, and the potential for chemical transformation of compounds during and subsequent to sampling. Sampling strategies need to include an appropriate number of individuals and a well-defined protocol to isolate the biological tissue of interest. It is normally advisable to remove tissue followed by immediate storage in liquid nitrogen to prevent degradation. Although it is normally necessary to remove surface contamination such as soil or sediment from plant samples, washing procedures may remove or transform some chemical compounds.

A major issue in the speciation of soils and sediments is obtaining a representative sample. It is necessary to collect relatively large quantities of samples using tools (e.g., shovels, corers, sediment traps) and collection containers that do not introduce contamination. As with biological samples, one must consider the effects of collection and storage upon the integrity of the chemical species. For example, air-drying of soil samples that were collected under anoxic conditions may be expected to cause significant changes in sample composition.

### Calibration and Standard Reference Materials

Analytical methods may be classified as absolute or relative methods. 'Absolute methods' have the advantage of producing results without comparative measurements, but are generally restricted to pure samples because of their susceptibility to interferences from sample matrices. 'Relative methods' involve the comparison of the signal response of a detection system ( $S$ ) compared to calibration samples of known concentration. Normally linear calibration graphs are employed because of their ease of

preparation from a few standards and their description by the following simple relationship:

$$S = kc + b \quad [1]$$

where  $c$  is the concentration of the analyte;  $k$  the slope, reflecting the sensitivity of the method; and  $b$  is the  $y$ -intercept, which is normally close to zero.

Ideally the analyst would have a wide variety of samples, called certified reference materials (CRMs) or standard reference materials, with well-defined analyte concentrations. Practically, however, it is too expensive to produce the vast number of these materials that would be needed for all elements and samples. Instead, calibration is typically performed using aqueous standards, and the CRMs are employed as samples to verify accuracy.

The creation of CRMs for speciation analysis was a major analytical challenge because of the instability of many chemical species of interest. However, since the 1980s, five organizations have developed CRMs for environmental speciation: the European Commission (BCR, Brussels, Belgium), International Atomic Energy Agency (IAEA, Vienna, Austria), National Institute for Environmental Studies (NIES, Japan), the National Institute for Standards and Technology (NIST, Gaithersburg, MD), and the National Research Council Canada (NRCC, Ottawa). A list of some available CRMs for speciation analysis is presented in Table 3. It should be noted that a relatively limited number of materials are available for speciation analysis that are certified for a handful of environmentally significant chemical species. It is clearly desirable to have CRMs available for more compounds and also with a wider range of chemical species.

**Table 3** Representative certified reference materials for speciation analysis available from European Commission (BCR), International Atomic Energy Agency (IAEA), National Institute for Environmental Studies (NIES), National Institute for Standards and Technology (NIST), and National Research Council Canada (NRCC)

<i>Producers</i>	<i>Sample and certified compounds</i>
NIST	Cr(III) and Cr(IV) in water
NIST, IAEA, NRCC, BCR	Mercury and methyl mercury in fish or shellfish
BCR, IAEA	Mercury and methyl mercury in sediment
NRCC, BCR	Butyl tin compounds in sediment
BCR	Butyl tin compounds in shellfish
NIES	Total tin and tributyl tin compounds in fish tissue
BCR	Organoarsenic compounds in fish tissue
BCR	Trimethyl lead in urban dust

## Instrumentation for Environmental Speciation Analysis

Instrumentation for speciation requires the ability to quantitatively determine various chemical species of an element. The most widely used instrumental configuration includes a system to separate the chemical species, such as chromatography or electrophoresis, and an atomic spectrometry detector, which provides high sensitivity and specificity for the separated compounds of the element of interest. **Figure 1** illustrates instrumentation that was employed for manganese speciation analysis.

### Separation Systems

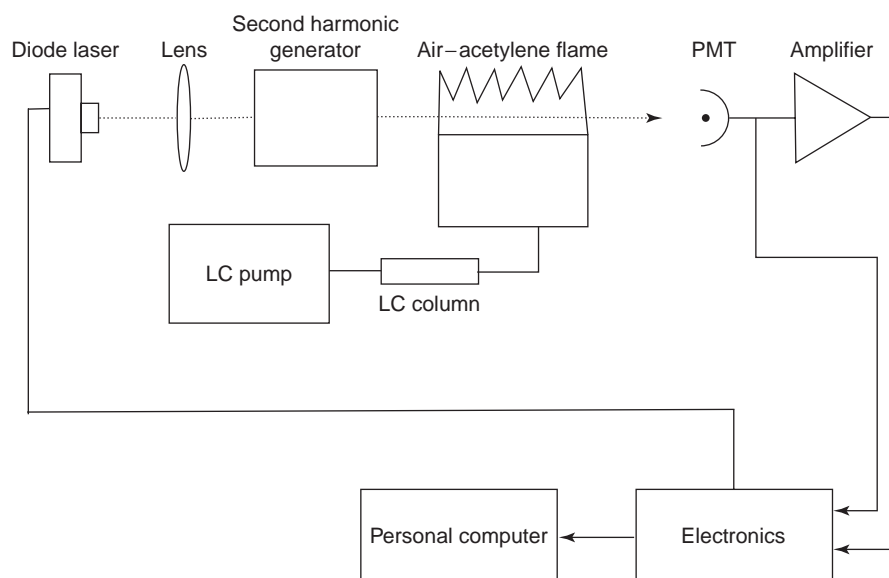
The most commonly used separation techniques for speciation include gas chromatography (GC), liquid chromatography (LC), and capillary electrophoresis (CE). GC has the major advantage that the analytes exiting the separation system are in the gas phase, which allows high sample introduction efficiency (essentially 100%) into the detector, resulting in high sensitivity. Other advantages include the higher chromatographic efficiency of GC compared to liquid chromatography (LC) (narrower peaks) and increased detector stability due to the absence of solvent introduction. Its major limitation is that it is restricted to compounds, which are volatile and thermally stable. A relatively limited number of chemical species can be determined directly by GC. These include tetraalkyl lead, metallic and methyl mercury, and sulfur compounds.

In order to determine nonvolatile elemental species, it is necessary to employ chemical reactions to

produce volatile, relatively nonpolar, thermally stable compounds. These 'derivatization reactions' involve one of the three general classes: production of volatile compounds in aqueous solution, involving hydride generation or ethylation; production of volatiles in organic solution, e.g., butylation using Grignard reactions; or formation of a volatile chelate, such as acetylacetonates or dithiocarbamates. The efficiency of derivatization reactions is highly dependent upon the sample matrix, and hence it is normally necessary to perform recovery studies for each category of sample (e.g., soil, sediment, etc.). Internal standardization or isotope dilution procedures are commonly employed to correct for these effects.

LC has the primary advantage of being applicable to virtually all analytes without derivatization procedures. In addition, a wide variety of stationary phases are available in conjunction with an array of solvent systems to separate analytes. LC techniques routinely used for speciation include reversed phase LC (RP-LC), reversed phase ion pair LC (RPIP-LC), and ion exchange LC (IE-LC). RP-LC, which uses a polar mobile phase with a nonpolar stationary phase (typically silica to which alkyl chains are chemically bonded), is best suited for the separation of nonpolar and slightly polar compounds, such as organometallic species.

RPIP-LC uses the same stationary phases employed in RP-LC, but with the addition of a salt to the mobile phase that contains an ion with a polar head group and a nonpolar tail. These salts are called 'ion pair reagents'. Examples of these compounds include tetraalkylammonium salts and sodium alkyl



**Figure 1** Example of experimental setup of instrumentation employed for manganese speciation.

sulfonates. RPIP-LC is widely used for speciation because it can be used to simultaneously separate ions and uncharged species.

IEC-LC employs a stationary phase that includes ionic functional groups, such as sulfonate or quaternary ammonium, bonded to a supporting material. These functional groups interact electrostatically with analytes having the opposite charge, and hence IEC-LC is employed for the speciation analysis of charged compounds. Aqueous mobile phases are generally employed.

CE is a very powerful tool for chemical separation, providing very high efficiency and the ability to use several separation modes. Moreover, since there is no stationary phase to interact with the analytes, CE has a reduced risk of changing species integrity compared to chromatography techniques. In spite of these potential advantages, CE has not been widely used for speciation, primarily due to the small sample volumes employed, typically a few nanoliters. This results in concentration detection limits 100 times higher than LC methods. The low volume also creates problems in interfacing CE to the detection system. Moreover, artifact peaks have also been reported that may be difficult to distinguish from analyte signals. CE is consequently best suited to environmental samples that include analytes that are difficult to separate by LC and are present at relatively high concentrations.

### Detection Systems

Detection systems for speciation have commonly consisted of atomic spectrometry instrumentation. One of the earliest techniques employed was flame atomic absorption spectrometry (FAAS). Sample is introduced into a flame using a pneumatic nebulizer system. The light source for atomic absorption is a low pressure (a few Torr) hollow cathode lamp (HCL) that includes a ceramic cylinder cathode coated with the pure metal or a compound of the analyte. Application of 150–300 V across the electrodes produces a plasma that results in a narrow atomic emission line that is absorbed by analyte atoms in the flame. FAAS instrumentation is relatively inexpensive and easily interfaced to chromatography systems. However, HCL-FAAS is characterized by relatively poor sensitivity that has limited its use for practical speciation analysis.

An alternative to HCLs for FAAS is the use of semiconductor diode lasers (DLs), which offer higher sensitivity, longer linear calibration graphs, high reliability, small size, and easy operation (**Figure 1**). Diode lasers produce infrared radiation that is converted to visible light using a process called ‘second

harmonic generation’. DL-FAAS is a low cost, high sensitivity detector for speciation.

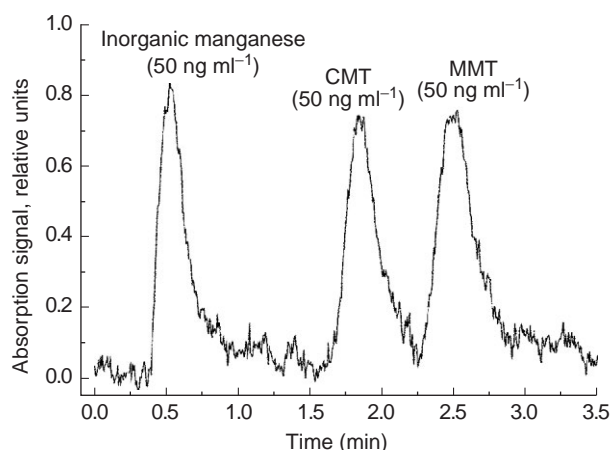
More recently, the inductively coupled plasma (ICP) has become the detection system of choice for most applications, either as a source for optical emission spectrometry (OES) or mass spectrometry (MS). The ICP is produced by coupling energy from a radio-frequency generator (generally 500–2000 W at 27.12 MHz) into argon gas using a water-cooled copper coil. Subsequent to the seeding of the argon with electrons, collisions between charged species occur that cause ionization of gaseous atoms. One of the major advantages of the ICP for real sample analysis is its relatively high temperature (5000–6000 K in the analytical region), which minimizes chemical reactions between the analyte and the sample matrix. Sample introduction is normally performed using a pneumatic nebulizer system. Modern detection systems for ICP-OES employ charge-coupled devices (CCD) that allow nearly simultaneous determination of 20 or more elements. Although ICP-OES provides superior sensitivity to FAAS for speciation analysis, it too is insufficiently sensitive for some applications.

ICP-MS is the most widely used detection system for environmental speciation, based on its excellent sensitivity, long linear calibration graphs, and ability to determine many elements simultaneously. A fraction of the plasma passes through an interface into a mass spectrometer where the ions are distinguished on the basis of mass-to-charge ratio. Although most ICP-MS work has employed a quadrupole filter for mass analysis, other mass analyzers are commercially available, including time-of-flight, ion traps, and double focusing sector instruments. The high concentration of salts employed for LC may introduce problems for ICP-MS, such as blockage and corrosion of the interface, and increased spectral interferences. However, its advantages make ICP-MS the technique of choice for environmental speciation analysis.

### Example of Application of Environmental Speciation

An example of environmental speciation involves the determination of the manganese containing fuel additive, methylcyclopentadienyl manganese tricarbonyl (MMT), and its derivative, cyclopentadienyl manganese tricarbonyl (CMT). As discussed above, the toxicity of this compound and concerns about human exposure to inhaled manganese have raised concerns about the widespread use of this compound in gasoline. Instrumentation for this work (**Figure 1**) involved the combination of LC coupled with





**Figure 2** Chromatogram of inorganic manganese, CMT, and MMT by LC-DL-FAAS.

**Table 4** Analysis of environmental samples spiked with MMT

Sample	Added concentration ( $\mu\text{g ml}^{-1}$ )	Reported solubility in water ( $\mu\text{g ml}^{-1}$ )	Determined value by HPLC-DL-FAAS, ( $\mu\text{g ml}^{-1}$ )
Tap water	—	10–70	$36 \pm 2$
Gasoline	8.3	—	$8.0 \pm 0.8$
Human urine	4.0	—	$4.2 \pm 0.3$

DL-FAAS. The speciation instrumentation included a reversed phase  $\text{C}_{18}$  column for the determination of MMT, CMT, and inorganic manganese. The mobile phase was 65:35 methanol/aqueous pH 4 buffer ( $0.05 \text{ mol l}^{-1}$  ammonium acetate).

An example of a chromatogram is shown in **Figure 2**, illustrating good separation between the various manganese-containing species. It should be pointed out that this analysis was performed in less than 3 min, providing rapid analysis. The manganese detection limit for LC-DL-FAAS was 2 ng per ml, with a linear dynamic range of 3 orders of magnitude.

The suitability of DL-FAAS to perform practical analysis was evaluated by the addition of MMT to samples of tap water, gasoline, and human urine (**Table 4**). Saturated solutions of MMT were prepared to simulate its determination in ground water. MMT was added to gasoline at the maximum allowable level in the United States. Urine was selected because of its significance in toxicological studies. For all samples, good agreement was obtained between the measured and spiked values.

See also: **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Capillary Electrophoresis:** Environmental Applications. **Elemental Speciation:** Waters, Sediments, and Soils. **Environmental Analysis:** Gas Chromatography: Overview.

## Further Reading

- Buffle J and van Leeuwen HP (2000) *In-Situ Monitoring of Aquatic Systems: Chemical Analysis and Speciation*. New York: Wiley.
- Butcher DJ and Sneddon J (1998) *A Practical Guide to Graphite Furnace Atomic Absorption Spectrometry*. New York: Wiley.
- Butcher DJ, Zybin A, Bolshov M, and Niemax K (1999) Speciation of methylcyclopentadienyl manganese tricarbonyl by high performance liquid chromatography – diode laser atomic absorption spectrometry. *Analytical Chemistry* 71: 5379–5385.
- Caroli S (ed.) (1996) *Elemental Speciation in Bioinorganic Chemistry*. New York: Wiley.
- Cornelius R, Caruso J, Crews H, and Heumann K (eds.) (2003) *Handbook of Elemental Speciation: Techniques and Methodology*. New York: Wiley.
- Cornelius R and De Kimpe J (1994) Elemental speciation in biological fluids – invited lecture. *Journal of Analytical Atomic Spectrometry* 9: 945–950.
- Harrison RM and Rapsomanikis S (eds.) (1989) *Environmental Analysis using Chromatography Interfaced with Atomic Spectroscopy*. New York: Halsted Press.
- Hewitt CN (1989) Atomic absorption detectors. In: Harrison RM and Rapsomanikis S (eds.) *Environmental Analysis Using Chromatography Interfaced with Atomic Spectroscopy*, p. 53. New York: Wiley.
- Irgolic KJ (1987) Analytical procedures for the determination of organic compounds of metals and metalloids in environmental samples. *Science of the Total Environment* 64: 61.
- Kramer JR and Allen HE (eds.) (1988) *Metal Speciation: Theory, Analysis, and Application*. Chelsea, MI: Lewis Publishers.
- Montaser A (ed.) (1998) *Inductively Coupled Plasma Mass Spectrometry*. New York: Wiley-VCH.
- Ortiz AIC, Albarrán YM, and Rica CC (2002) Evaluation of different sample pre-treatment and extraction procedures for mercury speciation in fish samples. *Journal of Analytical Atomic Spectrometry* 17: 1595–1601.
- Prichard E, MacKay GM, and Points J (eds.) (1996) *Trace Analysis: A Structured Approach to Obtaining Reliable Results*. Cambridge: Royal Society of Chemistry.
- Quevauviller P (1996) Atomic spectrometry hyphenated to chromatography for elemental speciation: Performance assessment within the standards, measurements and testing programme (Community Bureau of Reference) of the European Union. *Journal of Analytical Atomic Spectrometry* 11: 1225–1231.
- Tessier A and Turner DR (eds.) (1996) *Metal Speciation and Bioavailability in Aquatic Systems*. New York: Wiley.



# ENDOCRINE DISRUPTING CHEMICALS

**F Regan**, Dublin City University, Dublin,  
Republic of Ireland

**B Fogarty**, University of Kansas, Lawrence, KS, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Recent years have seen heightened interest in the investigation of a number of compounds that interfere with the normal action of the endocrine system, collectively known as endocrine disrupting compounds (EDCs). An EDC has been defined as 'an exogenous substance or mixture of substances that alters the function(s) of the endocrine system and consequently causes adverse health effects in an intact organism or its progeny or (sub) population'. Compounds that interfere with the action of endogenous estrogen hormones are known as environmental estrogens, estrogen mimics, estrogenic chemicals, or xenoestrogens.

The significance of these EDCs lies in the biological effects that have been observed in both wildlife and human populations, which have been exposed to them. In wildlife, exposure to elevated levels of environmental estrogens has been associated with sexual disruption of fish populations, eggshell thinning in bird populations, and gonad modification of reptiles. Concern is now focused on the possible implications for humans who have also been exposed to these EDCs, with many of these compounds already implicated in reproductive abnormalities, breast cancer, and male infertility. Compounds identified as priority EDCs include various members of distinct chemical groups such as the bioflavonoids, synthetic estrogens, phenols, alkylphenols, polychlorinated biphenyls, phthalates, brominated flame retardants, organochlorine pesticides, dioxans, and furans. The diverse chemical nature of these compounds is reflected in the diversity of their application, with their widespread use leading to their distribution in various environmental and biological matrices.

In order to appreciate the consequences of exposures to EDCs, it is important to highlight the significance of the endocrine system and of the disruption of its normal function in biological systems.

## The Endocrine System

The body has two important communication systems: (1) the nervous system and (2) the endocrine

system, which regulate and coordinate the activity of the body's muscles and internal organs. A wide range of bodily functions are regulated by the endocrine system, some of which include digestive processes, blood pressure, immune system responses, metabolic rate, and reproductive processes. The major endocrine glands in the body are the hypothalamus and pituitary gland in the brain, the thyroid and parathyroid, the thymus, the adrenal glands, the pancreas, the kidneys, and the gonads (in females the ovaries and in males the testes). The main function of an endocrine gland is to secrete a specific hormone in response to some stimulus, which is generally chemical in nature rather than neural.

The mechanism of action of the steroid hormones is based on a lock-and-key receptor binding procedure. The hormone binds to a complimentary receptor and induced a further cell response. There are, however, additional substances that can interact with the hormone-binding site of the receptor. Such compounds can have agonist, antagonist, or partial agonist activities. Agonists work to mimic the effect of normal hormones through binding to the receptor site and inducing the expected response from the cell. Antagonists block the normal response by binding to the receptor in place of the natural hormone. If the binding is sufficient to prevent binding of the natural hormones then no response will occur. Partial agonists that bind to receptor sites will not induce the full response from the cell but some effect will be observed.

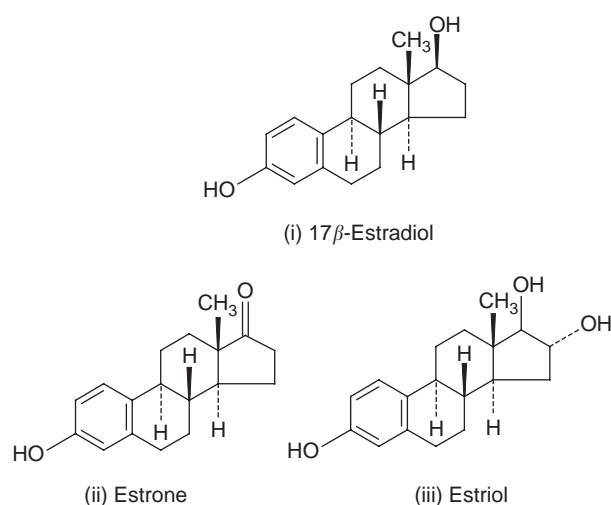
## Estrogens

The estrogen hormones are synthesized in the ovaries in females and in the testes in males. There are at least six distinct estrogen hormones but only three are present in significant amounts in the female body:  $\beta$ -estradiol, estrone, and estriol. All three molecules have a 17-carbon system, which is the steroid with a methyl group at carbon-13 and an aromatic ring with a hydroxyl group at carbon-3.

17 $\beta$ -estradiol is the most biologically active estrogen produced by the ovaries, and is synthesized from androgens by the aromatase complex of enzymes (see Figure 1).

## Androgens

Testosterone is the principal hormone of the androgen group of steroids. Secretion of testosterone



**Figure 1** Chemical structure of three natural estrogens.

increases during puberty and is responsible for the development of male secondary sexual characteristics. In women, androgens are present in the form of androstenedione, which can be converted to testosterone and dihydrotestosterone as needed. Suppression of androgen activity has been seen to improve the survival rate of patients diagnosed with prostate cancer. Antiandrogen treatments are being investigated as therapeutic agents. The synthesis, transport, and metabolism of androgens may be disrupted, by both xenestrogenic and antiandrogenic substances.

Initial research into the area of endocrine disruption focused on the sex steroid hormones, i.e., the estrogens and androgens as several EDCs were found to mimic the action of the natural hormones. Recent research has also included the antiestrogenic and antiandrogenic properties of many of the EDCs. The potential disruption of other hormone signaling systems is also under investigation.

## Endocrine Disrupting Compounds

### Mode of Action

The proposed mode of action of an EDC is to bind to the receptor site in place of the natural hormone. If an EDC succeeds in binding to the hormone receptor it has the potential to activate or block the cellular response characteristic of the hormone it has replaced. Compounds possessing remarkably different structures have been found to compete for the traditional estrogen binding sites on the estrogen receptor and elicit varying degrees of response.

While the molecular structure of EDCs binding to specific receptors can vary greatly, there are often

similarities between these compounds. Many of the estrogen mimicking EDCs share the common structural motif of a phenol or a functional equivalent. They are lipophilic, exhibit resistance to metabolism, and can bioaccumulate.

### Sources of EDCs

Many different types of compounds have been identified as EDCs, all having very different applications. The broad nature of application has resulted in the distribution of these compounds in various environmental and biological matrices and leads to the risk of repeated exposures with compounds originating from different sources. The main sources of exposure to these compounds include the environment, dietary components, and medical and cosmetic products.

Several physicochemical and biological processes determine the concentration, behavior, and ultimate fate of chemical species in the wider environment. Such processes include sorption-desorption, volatilization, and chemical and biological transformations. The chemical characteristics of the compound, including solubility, vapor pressure, and its partition coefficient, will impact upon its residence time in the wider environment.

The aquatic environment acts as a sink for many compounds discharged via the wastewater treatment system. The main input of EDCs to the wider environment is as a result of direct point discharge from both domestic and industrial wastewater treatment facilities. EDCs have been detected in sewage and wastewaters, surface and groundwater, as well as soils and sediments. Conventional wastewater treatment facilities, i.e., those employing primary and secondary treatment technologies, are proving inefficient for the complete removal of all EDCs prior to the discharge of 'treated effluents'. While a significant proportion of the EDCs are retained in the sludge, the stable chemical nature of many of these compounds results in their persistence in the final effluent at low concentrations.

Aside from possible exposure through drinking water sources, dietary exposure to EDCs results from components of diet known as bioflavonoids, phytohormones, or phytoestrogens. These natural plant hormones are less potent than the estrogens found in mammals and in terms of a mass balance approach it has been suggested that the action of such weak or antiestrogens may work to combat the action of elevated environmental estrogen exposure. Some flavones and isoflavones may play an important role in cancer prevention as they are found in plants that are associated with reduced cancer rates.

The use of estrogenic and other hormone-containing preparations in medical and cosmetic applications could constitute another form of exposure to EDCs. One of the main medical applications of the synthetic hormones is in the female contraceptive pill. The detergent nonoxynol, whose main component is nonylphenol (an estrogen mimic), is used as a vaginal lubricant and spermicide.

### Priority Groups of EDCs

Priority groups are estrogens, both natural and synthetic, several phenols, and the polyethoxylated alkylphenol detergents and their breakdown products. Natural estrogens can be considered EDCs after deconjugation restores biological potency. Many of the natural hormones are not removed by conventional wastewater treatment facilities due to their stable chemical composition and resistance to biodegradation (see Figure 2).

Bisphenol-A (BPA) is used in the manufacture of epoxy resins and polycarbonate plastics. These compounds are then used in the manufacture of various other products, including tooth fillings, dental materials, and the coatings on the inside of food tins. BPA was identified as an estrogen mimic after tests with yeast conditioned media (see Figure 3).

Pentachlorophenol (PCP) was introduced in the 1950s and has been included in the list of EDCs. It is used primarily as a biocide for wood preservatives where it is often dissolved in a diesel-type oil and pressure injected into wooden poles (see Figure 4).

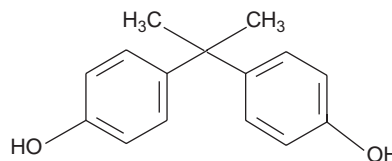
Phenolsulfonephthalein, also known as phenol red, is a pH indicator used in tissue culture media. The

compound has been noted to bear a structural resemblance to some nonsteroidal estrogens, such as cyclofenil and chlorotranisene (see Figure 5).

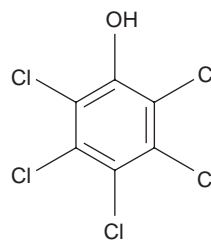
### Alkylphenols

A large number of chemicals known collectively as alkylphenols have exhibited estrogenic behavior and have been identified as components of wastewaters. Alkylphenolic compounds are used in a variety of applications. The alkylphenols octylphenol and nonylphenol are used as antioxidants in plastics. Both nonylphenol and octylphenol have been identified as estrogen mimics using vitellogenin gene expression in trout hepatocytes, gene transcription in transfected cells, and by the growth of breast cancer cell lines (see Figure 6).

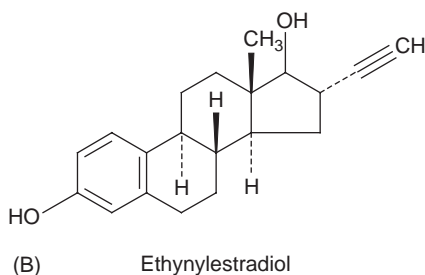
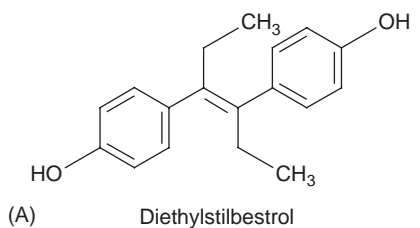
Both octylphenol and nonylphenol are found in the wider environment as a result of the biodegradation of the alkylphenolic polyethoxylates (APEOs) and are themselves recalcitrant breakdown products. The nonionic alkylphenol ethoxylated surfactants, namely



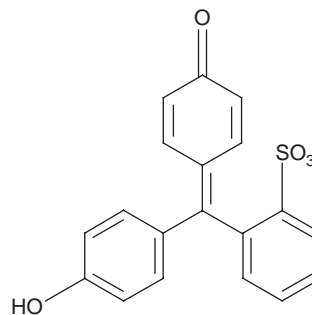
**Figure 3** Chemical structure of bisphenol-A.



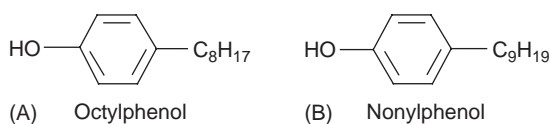
**Figure 4** Chemical structure of pentachlorophenol.



**Figure 2** Chemical structures of synthetic estrogens: (A) diethylstilbestrol and (B) ethynylestradiol.



**Figure 5** Chemical structure of phenolsulfonephthalein (phenol red).



**Figure 6** Chemical structures of (A) octylphenol and (B) nonylphenol.

octylphenol ethoxylate and nonylphenol ethoxylate, are found in domestic and industrial products with some of the applications including shampoos, some detergent-containing petrol, and pesticides formulations. Their structure consists of an alkylphenol with a side chain of several ethoxylate groups, ranging from 2 to 20 units. The biotransformation of the APEOs to the parent alkylphenols involves hydrophilic attack of the ethoxylate chain, which is progressively shortened by one ethoxylate group at a time.

## Determination of EDCs in the Aquatic Environment

The analysis and identification of estrogen-mimicking EDCs is generally achieved using a combination of biological assays and analytical separation techniques. The biological potency of a compound can be assessed using *in vivo* or *in vitro* assays and activity is usually expressed relative to that of estradiol, which is the most common endogenous estrogen.

Chromatographic methods are routinely used for the analysis of environmental estrogens due to their tailored selectivity and high resolving power resulting in efficient analyses. Coupling of chromatography-based separation techniques to sensitive detection methods has allowed the qualitative and quantitative measurement of target analytes in complex matrices. The automation of many devices also allows for continuous analyses and increased sample throughput while the modular design of some systems allows online coupling to pre-separation preparatory units.

Due to the significant interest in the development of analytical techniques for the determination of EDCs in a variety of matrices, several reviews of analytical methodologies used have been published recently.

### Biological Assays

Various biological assays can be employed in order to determine the potency of the compound under investigation. The main methods used for the analysis of xenostrogens have been summarized previously. The techniques can be classified according to biological complexity.

**Receptor binding assays** Receptor binding assays are easy to perform, and allow identification of both agonists and antagonists acting through the receptor but are unable to differentiate the binding of agonists or antagonists. By measuring the direct binding of the chemicals to the fish estrogen receptor can assess the estrogenicity of a selection of environmentally persistent chemicals.

**DNA binding assays** DNA binding of the estrogen receptor only occurs when the hormone-binding sites are occupied by suitable ligands. The kinetics of human estrogen receptor binding to its specific DNA element can be measured using a BIAcore biosensor system.

**Receptor gene assays** The yeast estrogen screen assay has been widely used for the determination of the estrogenic potency of many compounds. It is based on the introduction of two genes into yeast, one of which expresses the estrogen receptor and a reporter plasmid containing a gene coupled to the estrogen response element. The binding of an estrogen or estrogen mimic to the receptor results in the subsequent binding of the complex to the DNA estrogen response element facilitating the expression of  $\beta$ -galactosidase. The resulting enzyme activity can be measured by conversion of a substrate to nitrophenol, which is then detected. While the assay can take a day to complete, it has the advantage of simulating biological effects including the membrane passage of active components.

### Proliferation of estrogen-sensitive cells (E-screen)

Breast cancer cell lines are known to proliferate in response to estrogen stimulation. Charcoal-dextran stripped human serum inhibits the proliferation of MCF-7 breast cancer cells in a dose-dependant fashion. Estrogens overcome this effect. The E-screen test exploits the estrogen-dependant proliferation of MCF-7 cells and has been used previously for the determination of the estrogenic potential of phenol red, estradiol, and tamoxifen among other compounds. Such assays can take 4–7 days.

**Biomarkers** The observation of biological responses can also be used as an indication of EDC exposure. In studies where fish have been exposed to sewage the results have indicated that sewage effluent was estrogenic in nature. Exposed fish experienced a rapid and pronounced increase in their plasma vitellogenin concentrations. Rainbow trout and wild roach populations from various river sites have shown evidence of vitellogenin induction. Quantification of

vitellogenin concentrations can be achieved using an established homologous carp radioimmunoassay.

## Sample Preparation Techniques

EDCs in the aquatic environment are often present in complex matrices such as effluent samples. For this reason, wide ranges of sample preparation and pre-concentration techniques have been employed in conjunction with analytical separations. Preconcentration is often necessary due to sensitivity limitations of the instrumental techniques.

Solid-phase extraction (SPE) has been used in conjunction with separation techniques [gas chromatography (GC), liquid chromatography (LC), and capillary electrophoresis (CE)] for environmental analysis. Recent developments in both SPE and solid-phase microextraction (SPME) have been reviewed. Some of the solid phases investigated previously include graphitized carbon black, octadecylsilica, and C<sub>18</sub> cartridges. Comparisons of the sorbent materials available for the extraction of phenols have been carried out in conjunction with chromatographic separations. A comparison of polycrystalline graphites and SPME fibers [poly(dimethyl siloxane/divinylbenzene), carbowax] showed that both performed equally well for the extraction of nonionic surfactants. SPME fibers have been used in the analysis of short-ethoxy-chain nonylphenols in raw and treated water recently. Semipermeable membrane devices have been used for the identification of a number of hydrophobic organic contaminants.

Additional sample preparation techniques investigated for EDCs include supercritical fluid extractions with the addition of modifiers to aid compound solubility. Extractions of organic compounds can also be accelerated using microwave energy.

## Liquid Chromatography

LC has been used extensively for the determination of many EDCs due to the many favorable aspects of the technique. Some advantages of LC separation systems for the determination of EDCs include:

- good separation efficiency and selectivity;
- nonvolatile analyte separation;
- compatibility with sensitive detection modes; and
- online sample cleanup.

This technique can achieve efficient separations of polar and nonpolar analytes through the variation of both the stationary and mobile phases allowing variation of separation selectivity. LC has been applied to the determination of several phenolic

endocrine disrupting compounds in real samples. Up to 28 phenolic compounds have been separated in a single run using a C<sub>18</sub> stationary phase. Reversed-phase LC has been applied to the determination of BPA and phenol. Isocratic LC has been used for the determination of the 11 priority pollutant phenols, which include phenol, PCP, 2,4,6-trichlorophenol, and several other nitro- and chloro-phenolic compounds. The use of specialized chiral stationary phases can further alter the selectivity of LC methods.  $\beta$ -Cyclodextrin stationary phases have been applied to the analysis of several estrogens, both natural and synthetic forms.

As the system is aqueous based, it is routinely used for the analysis of nonvolatile analytes with no need for derivatization of sample to maintain analyte solubility. A typical LC system experiences lower diffusion rates than GC because the mobile phase has a higher viscosity and density. LC can be coupled online to sensitive electrochemical and fluorescence detection units. This is an important consideration for the analysis of EDCs present at low concentrations in environmental matrices. Both UV and EC detection were applied to the determination of estrogenic compounds in water samples using LC.

Pre- and guard columns can be used for online sample cleanup in conjunction with LC, which helps maintain sample integrity while protecting expensive and fragile stationary phases. Samples may also be preconcentrated prior to elution onto separation column.

The interfacing of LC systems to mass spectrometric (MS) detection units has also allowed further detection strategies. Normal-phase LC with electrospray mass spectrometric detection has been employed for the determination of alkylphenols, including nonylphenol ethoxylates in the aquatic environment and BPA in biological fluids and environmental samples.

Fluorescence detection has also been employed for the detection of estrogens, concentrations of octylphenol and nonylphenol ethoxylates in effluent samples among others. Derivatization can allow the selective detection of analytes when interfering compounds may be present in the sample. Derivatizing agents including 2-(4-carboxyphenyl)-5,6-dimethylbenzimidazole and dansyl chloride have been used.

Conventional methods have more recently been coupled to biological-based assays in order to improve the sensitivity and selectivity of conventional techniques; for example, LC coupled online to receptor affinity detection based on the human estrogen receptor.



The disadvantages of LC systems for the analysis of EDCs include:

- high reagent and sample consumption and
- limited detector compatibility.

LC systems also tend to have a large sample capacity combined with higher solvent demands in the liters per day range as opposed to CE where it is milliliters per day. The large solvent demands of the system lead to increased consumption of expensive chemicals and increased waste for disposal.

The organic nature of both analytes and mobile phases limits detection modes to those that can differentiate between the two. The fragile nature of many of the stationary phase packed columns can lead to further expense in terms of installation, method development, and replacement.

The use of gradient elution methods is often necessary to allow the analysis of analytes of differing polarities, as is the case with mixtures of EDCs, many of which exhibit poor solubility in aqueous systems due to their hydrophobic nature. While gradient schemes can be automated with suitable instrumentation, this adds to the complexity of the separation.

## Gas Chromatography

GC has also been employed extensively for the determination of several EDCs.

The advantages of GC separation systems for the analysis of EDCs include:

- good separation efficiency and selectivity;
- small sample and reagent demands; and
- compatibility with sensitive detection modes.

GC analysis can achieve a high degree of analyte resolution and peak efficiency when long columns are employed. Separation selectivity can be altered by the variation of the stationary phase. This technique is less complex than either CE or LC resulting in straightforward method development options. As with CE methods, GC analysis only demands small sample volumes, which is an advantage when working with toxic or expensive compounds. There is no mobile phase, buffer preparation, or solvent waste to deal with using a GC system. GC methods have been applied to the determination of several phenolic EDCs including BPA and Bisphenol F.

GC has been applied to the determination of several alkylphenols. The determination of 4-nonylphenol (4-NP) in effluent and sludge from sewage treatment plants has been shown while alkylphenol ethoxylates have been determined in industrial and environmental samples using GC–MS.

GC is readily coupled to MS detection, which provides the sensitivity needed to allow quantification of environmental and biological levels of environmental estrogens in real samples, which are often present in the low  $\text{ng l}^{-1}$  to  $\mu\text{g l}^{-1}$  range.

Nonylphenol polyethoxylates and their degradation products have been detected in river water and sewage effluent by GC–ion trap (tandem) MS with electron impact (EI) and chemical ionization, which was followed by direct GC–MS analysis in EI–MS, chemical ionization mass spectrometry (CIMS), or CIMS–MS modes.

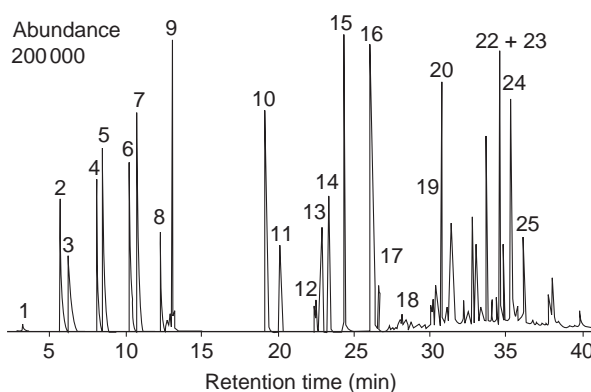
While many methods use offline preconcentration methods, SPME has been used for estrogens and anabolic steroids with on-fiber derivatization prior to analysis by GC–MS analysis (see Figure 7).

The disadvantages of GC systems for the analysis of EDCs include:

- limited to analysis of compatible analytes and
- limited method optimization options.

GC analyses are best suited to those analytes that have significant vapor pressure, are thermally stable at operating temperature, and have molecular weights of less than 1000. Compounds such as the lower chain alkylphenols are difficult to separate using GC due to their reduced volatilities. The polar nature of the extended ethoxylate chains of parent detergent compounds is also an issue.

Temperature programming can aid the separation of some analytes with faster separation at high



**Figure 7** Determination of priority EDCs using GC–MS in scan mode. Temperature ramp conditions –  $80^{\circ}\text{C min}^{-1}$ ; ramp –  $8\text{--}150^{\circ}\text{C}$  (150 for 8 min); ramp –  $11\text{--}280^{\circ}\text{C}$  (280 for 5 min). Peak identification: 1. phenol; 2. ethylphenol; 3. dichlorophenol; 4. 4-propylphenol; 5. 2-sec-butylphenol; 6. 2,4,5-trichlorophenol; 7. biphenyl; 8. methylparaben; 9. 2,3-tertbutyl-4-hydroxyanisole; 10. 4-hexyloxyphenol; 11. hexachlorobenzene; 12. pentachlorophenol; 13. lindane; 14. octylphenol; 15. 4-heptyloxyphenol; 16. octylphenol 2-polyethoxylate; 17. nonylphenol; 18. nonylphenol 2-polyethoxylate; 19. bisphenol A; 20. dieldrin; 21. DES; 22. estrone; 23.  $17\beta$ -estradiol; 24. ethynyl estradiol; 25. NP 12 polyethoxylate. Unlabeled peaks due to column bleed.



**Table 1** A comparison of a GC–MS method and a CE method (in the micellar electrokinetic chromatography (MEKC) mode) for the separation of priority endocrine disruptors. Resolution and limit of detection (LOD) are used as parameters for comparison

<i>Synthetic estrogens (in order of migration)</i>	<i>Resolution with MEKC</i>	<i>LOD (mg l<sup>-1</sup>)</i>	<i>Synthetic estrogens (in order of migration)</i>	<i>Resolution with GC/MS</i>	<i>LOD (mg l<sup>-1</sup>)</i>
Estriol	0	7.2	Phenol	0	0.2
Phenol	12.5	2.0	Butylphenol	20.0	1.0
Trichlorophenol	1.8	4.9	Trichlorophenol	4.2	0.08
Bisphenol-A	2.1	4.0	Biphenyl	1.1	0.05
PCP	0.9	6.7	Nonylphenol	–	0.02
Butylphenol	1.6	3.8	PCP	62.7	0.04
Estrone	7.7	6.8	Bisphenol-A	16.4	0.02
17 $\beta$ -estradiol	5.8	6.8	DES	7.0	1.0
DES	2.9	7.2	17 $\beta$ -estradiol	2.1	2.0
Ethinylestradiol	4.2	7.4	Estrone	0.6	2.0
Nonylphenol	36.5	5.5	Ethinylestradiol	2.0	5.0
			Estriol	5.0	20.0

temperatures offset by increased analysis times. The majority of separations of EDCs with GC employed some form of analyte derivatization procedure in order to achieve detection limits required. A separation of EDCs was achieved without derivatization, however, using online SPE and GC–MS–MS. While GC separations are less complex than LC and CE separations, method development options are limited as stationary phase and temperature are two of the main variables in the separation. Variation of often-fragile stationary phases can be expensive (see Table 1).

### Capillary Electrophoresis

Despite the routine use of GC–MS and LC for the analysis of EDCs there is another analytical technique that has potential for the analysis of these compounds. CE, though not without its own limitations, offers some advantages over existing techniques, which are outlined below:

- good separation efficiency and selectivity;
- method optimization options;
- low sample and reagent consumption;
- wide range of analyte application;
- miniaturization; and
- compatibility with sensitive detection modes.

One of the main advantages that CE offers over existing analytical methods is its ability to efficiently resolve both neutral and charged analytes in a single run. The selectivity of the analysis can be altered through the addition of various components to the run buffer system. These would include organic solutions to aid solubility of hydrophobic compounds, surfactants, and other agents designed to alter the solubility and migration characteristics of the analytes under investigation. The more recently developed technique of capillary electrokinetic chromatography

(CEC), which has been described as a combination of CE and chromatography, would potentially allow an efficient analysis of the environmental estrogens as extremely high peak efficiencies are a characteristic of this technique. CEC is still in the early stages of development, however, with the only alternative to expensive commercial columns involving difficult packing procedures for home-built systems.

CE method development options are cost-effective and simply involve the direct addition of agents to run buffer. The low cost of capillaries allows dedication of individual capillaries to specific methods, which aids reproducibility. This is another advantage of this technique that minimizes the cost of method development and waste generated. CE systems have small sample and reagent demands, which is a definite advantage when working with expensive or hazardous materials. As CE can be used for the separation of a broad spectrum of analytes including polar and nonpolar compounds, there is no need for derivatization to improve compound solubility, as is the case with some GC applications. Hydrophobic analytes can also be separated through the modification of traditional CE buffer systems. Nonaqueous CE modes allow separation of compounds in organic solvents as opposed to aqueous buffers.

There are, however, certain limitations to the technique for the analysis of EDCs including:

- sensitivity of commercial detectors available and
- complexity of separations and reproducibility.

The limited light path and low sample injection volumes further reduce the detection sensitivity of the system. While method development is relatively simple, the addition of agents to the run buffer system alters the complexity of the separation by increasing the number of variables. The presence of

multiple separation components can also influence the reproducibility of the separation. Fluorescence detection and mass spectrometric detection are seeing some use with CE. Analytical applications of these detection techniques with CE are in the early stages of development.

*See also:* **Capillary Electrochromatography.** **Capillary Electrophoresis:** Pharmaceutical Applications; Low-Molecular-Weight Ions; Food Chemistry Applications; Clinical Applications. **Extraction:** Solid-Phase Extraction; Solid-Phase Microextraction. **Gas Chromatography:** Mass Spectrometry. **Liquid Chromatography:** Overview; Liquid Chromatography–Mass Spectrometry.

### Further Reading

- Ahel M, Giger W, and Koch M (1994) Behaviour of alkylphenol polyethoxylate surfactants in the aquatic environment: I. Occurrence and transformation in sewage treatment. *Water Research* 28(5): 1131–1142.
- Baker VA (2001) Endocrine disrupters – testing strategies to assess human hazard. *Toxicology In Vitro* 15: 413–419.
- Brown RP, Greer RD, Mihaic EM, and Guiney PD (2002) A critical review of the scientific literature on potential endocrine-mediated effects in fish and wildlife. *Ecotoxicology and Environmental Safety* 49(1): 17–25.
- Cadbury D (1997) *The Feminisation of Nature: Our Future at Risk*. London: Penguin Books.
- Castillo M and Barcelo D (1997) Analysis of industrial effluents to determine endocrine disrupting chemicals. *Trends in Analytical Chemistry* 16(10): 574–583.
- Colburn T, VomSaal FS, and Soto AM (1993) Developmental effects of endocrine disrupting chemicals in wildlife and humans. *Environmental Health Perspectives* 101(5): 378–384.
- Markey C, Rubin B, Soto A, and Sonnenschein C (2003) Endocrine disruptors: from wingspread to environmental developmental biology. *The Journal of Steroid Biochemistry and Molecular Biology* 83: 235–244.
- Oosterkamp A, Hock B, Seifert M, and Irth H (1997) Novel monitoring strategies for xenoestrogens. *Trends in Analytical Chemistry* 16(10): 544–553.
- Sonnenschein C and Soto A (1998) An updated review of environmental estrogen and androgen mimics. *The Journal of Steroid Biochemistry and Molecular Biology* 65(1–6): 143–150.
- Warhurst AM (1995) *An Environmental Assessment of Alkylphenol Ethoxylates and Alkylphenols*. FOE.

## ENTHALPIMETRIC SENSORS

*See* **SENSORS: Calorimetric/Enthalpimetric**

## ENVIRONMENTAL ANALYSIS

**R Sandford and P Worsfold**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Environmental analysis encompasses matrices from the hydrosphere, atmosphere, lithosphere, and biosphere, and target analytes include naturally occurring chemicals and anthropogenically derived contaminants. In many environmental situations it is important to determine fluxes, which are time integrated loads, as well as concentrations. Particular challenges include the extremely low concentrations of many individual chemicals, the different physicochemical forms of elements (e.g., redox state,

complexed, dissolved, particulate, and colloidal fractions), the heterogeneity of sampling locations, the scale of temporal and spatial variability, access to sampling sites, and the potential instability of stored samples due to factors such as biological activity, atmospheric contamination, and chemical reactivity. Key drivers for environmental analysis are legislation for monitoring anthropogenically derived chemicals, both at the point of emission/discharge and in receiving atmospheres/waters and the need to better understand environmental processes in both pristine and impacted environments. The interface between chemistry and biology is an increasingly important aspect of environmental science in areas such as ecotoxicology, human and ecosystem health, and terrestrial/aquatic ecology and they all require high-quality chemical measurements.

multiple separation components can also influence the reproducibility of the separation. Fluorescence detection and mass spectrometric detection are seeing some use with CE. Analytical applications of these detection techniques with CE are in the early stages of development.

**See also:** **Capillary Electrochromatography.** **Capillary Electrophoresis:** Pharmaceutical Applications; Low-Molecular-Weight Ions; Food Chemistry Applications; Clinical Applications. **Extraction:** Solid-Phase Extraction; Solid-Phase Microextraction. **Gas Chromatography:** Mass Spectrometry. **Liquid Chromatography:** Overview; Liquid Chromatography–Mass Spectrometry.

### Further Reading

- Ahel M, Giger W, and Koch M (1994) Behaviour of alkylphenol polyethoxylate surfactants in the aquatic environment: I. Occurrence and transformation in sewage treatment. *Water Research* 28(5): 1131–1142.
- Baker VA (2001) Endocrine disrupters – testing strategies to assess human hazard. *Toxicology In Vitro* 15: 413–419.
- Brown RP, Greer RD, Mihaic EM, and Guiney PD (2002) A critical review of the scientific literature on potential endocrine-mediated effects in fish and wildlife. *Ecotoxicology and Environmental Safety* 49(1): 17–25.
- Cadbury D (1997) *The Feminisation of Nature: Our Future at Risk*. London: Penguin Books.
- Castillo M and Barcelo D (1997) Analysis of industrial effluents to determine endocrine disrupting chemicals. *Trends in Analytical Chemistry* 16(10): 574–583.
- Colburn T, VomSaal FS, and Soto AM (1993) Developmental effects of endocrine disrupting chemicals in wildlife and humans. *Environmental Health Perspectives* 101(5): 378–384.
- Markey C, Rubin B, Soto A, and Sonnenschein C (2003) Endocrine disruptors: from wingspread to environmental developmental biology. *The Journal of Steroid Biochemistry and Molecular Biology* 83: 235–244.
- Oosterkamp A, Hock B, Seifert M, and Irth H (1997) Novel monitoring strategies for xenoestrogens. *Trends in Analytical Chemistry* 16(10): 544–553.
- Sonnenschein C and Soto A (1998) An updated review of environmental estrogen and androgen mimics. *The Journal of Steroid Biochemistry and Molecular Biology* 65(1–6): 143–150.
- Warhurst AM (1995) *An Environmental Assessment of Alkylphenol Ethoxylates and Alkylphenols*. FOE.

## ENTHALPIMETRIC SENSORS

**See** **SENSORS:** Calorimetric/Enthalpimetric

## ENVIRONMENTAL ANALYSIS

**R Sandford and P Worsfold**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Environmental analysis encompasses matrices from the hydrosphere, atmosphere, lithosphere, and biosphere, and target analytes include naturally occurring chemicals and anthropogenically derived contaminants. In many environmental situations it is important to determine fluxes, which are time integrated loads, as well as concentrations. Particular challenges include the extremely low concentrations of many individual chemicals, the different physicochemical forms of elements (e.g., redox state,

complexed, dissolved, particulate, and colloidal fractions), the heterogeneity of sampling locations, the scale of temporal and spatial variability, access to sampling sites, and the potential instability of stored samples due to factors such as biological activity, atmospheric contamination, and chemical reactivity. Key drivers for environmental analysis are legislation for monitoring anthropogenically derived chemicals, both at the point of emission/discharge and in receiving atmospheres/waters and the need to better understand environmental processes in both pristine and impacted environments. The interface between chemistry and biology is an increasingly important aspect of environmental science in areas such as ecotoxicology, human and ecosystem health, and terrestrial/aquatic ecology and they all require high-quality chemical measurements.

## Environmental Legislation

Environmental protection is a global issue with a plethora of protective legislation, enacted as transnational and national regulations and guidelines, aimed at sustainable use of water, soil, air, and biomass. These laws apply not only to the polluter but also to governments, national institutions, and individuals. They are designed to protect human health and ecosystems from the effects of chemical contaminants, to guarantee the quality of natural resources, e.g., potable water and to facilitate the recovery of degraded environments, e.g., contaminated land.

Within Europe, European Council Directives and Regulations form an overarching legislative framework that requires Member States to take preventative action and specifies that environmental damage should, as a priority, be rectified at source and that the polluter should pay. Importantly, environmental protection is also required to be integrated into other European Community (EC) policies, emphasizing its pervasive nature. Thus, Member States have a strong impetus to enact environmental legislation that incorporates the concepts of responsibility and control. Directives formally bind Member States and specify the outcomes and deadlines but the exact form of the legislation is at the discretion of the individual member country. However, delays in implementation have led to the increased use of EC regulations that bind individual states directly, without requiring national implementation. The principles of subsidiarity and harmonization also apply to environmental law within the EC. Individual Member States have responsibility for enforcement, although overall policing is by the EC and, increasingly, by the European Environment Agency (EEA) and the European Court of Justice.

Within the UK, environmental protection is embodied in overarching statutes, e.g., the Environmental Protection Acts of 1990 and 1995, which require both an integrated approach, as embodied in the European Community Water Framework Directive, and a precautionary approach, as embodied in the 1999 Pollution Prevention and Control Act. Statutes are supported by Government Circulars and statements of policy, codes of practice, and orders and policies from agencies such as the Environment Agency and the Health and Safety Executive.

In the USA, 12 major statutes covering, e.g., Pollution Prevention, Clean Air, Clean Water, and Toxic Substances form the basis of an overarching legislative environmental protection policy generally administered by the US Environmental Protection Agency (EPA). The statutes either set, or require to

be set by the EPA, emission, quality, discharge, disposal, and testing standards with the intention of safeguarding the environment while enabling economic and social development.

International agreements and conventions, e.g., OSPAR and the Rio Summit, define a framework for transnational cooperation on many environmental issues such as sustainable development, biodiversity, protection of endangered species, tropical rain forests, the ozone layer, and means to combat global warming. The establishment of, and compliance with, environmental protection legislation/policies provides a major driving force for many analytical science developments. There is a demonstrable need for highly accurate and precise analytical techniques with increasingly low detection limits for an increasing number of chemicals. This has given impetus to the development of novel environmental analytical methodologies, clean sampling and analytical protocols, and continuous/*in situ* monitors.

## Sampling

Reliable sampling is a key first step in any analytical program. Samples should be representative of the bulk medium from which they are taken and should not change in composition between collection and analysis. Care must therefore be taken to avoid contamination of the sample, e.g., from the atmosphere or the addition of reagents, and loss of analyte, e.g., by volatilization, adsorption, or biological transformation. Sample containers should be non-contaminating, for example, when collecting water samples the use of high density polyethylene for trace metals and glass for organic compounds is recommended. Containers should also be thoroughly cleaned with appropriate solvents before use to prevent contamination. When designing a sampling program, the temporal and spatial variability of environmental samples must also be considered, together with cost, time required, and safety issues. Ultimately the design of the sampling program will depend on the issue that is being addressed. Examples of temporal variability include diurnal cycles, tidal cycles, seasonal trends, and transient human impacts. Spatial variability is often due to point source inputs, such as a sewage treatment works outfall or a stack emission, and mixing processes, e.g., where a tributary joins a river or where air masses converge. The addition of chemicals to the sample at the point of collection, such as chloroform to minimize biological activity and acid to minimize adsorption on the walls of the container, is often required, especially if the sample is stored for some time prior to analysis. However, such additions can



introduce contamination and alter the speciation of the analyte.

## Analytical Techniques

Many 'baseline' environmental measurements, e.g., pH, Eh, conductivity, dissolved oxygen, temperature, flow rate, and turbidity for water samples, can be acquired in the field using simple, cheap, and robust handheld instrumentation. However, most chemical determinands require sampling followed by laboratory analysis.

Many organic and inorganic determinants in environmental matrices require separation and pre-concentration (if concentrations are below instrumental detection limits) prior to detection. This can be done by solvent extraction or solid-phase extraction using selective adsorbents, ion-exchange resins, or chelating resins. In many applications, however, chromatographic techniques with higher resolving power are needed. Gas chromatography (GC) for example, particularly in combination with mass spectrometric detection (GC-MS), is widely used for separating and quantifying individual organic compounds in air, water, and biological matrices. Liquid chromatography (LC), also in combination with MS (LC-MS), is particularly useful for determining low volatility compounds at room temperature, thus minimizing sample decomposition. More recently, capillary electrophoresis (CE) with ultraviolet (UV diode array) or fluorescence detection has been used for rapid and highly efficient separation and quantification of environmental samples.

Atomic spectrometric techniques such as flame atomic absorption spectrometry (FAAS), electrothermal AAS (ETAAS), inductively coupled plasma atomic emission spectrometry (ICP-AES), and ICP-MS are used for the determination of elements, particularly metals. ICP-MS is the most sensitive, typically with microgram per liter detection limits and multielement capability but it has high start-up and operating costs. UV-visible spectrophotometry is also used for the determination of metal ions and anions such as nitrate and phosphate (usually by selective derivatization). It is a low cost and straightforward technique, and portable (handheld) instruments are available for field deployment. Flow injection (FI) provides a highly reproducible means of manipulating solution chemistry in a contamination free environment, and is often used for sample manipulation, e.g., derivatization, dilution, preconcentration and matrix removal, in conjunction with spectrometric detection. Electroanalytical techniques, particularly voltammetry and ion-selective electrodes (ISEs), are

also used for the determination of metal ions in solution. Anodic stripping voltammetry (ASV) and cathodic stripping voltammetry (CSV) are very sensitive and provide data on elemental speciation, whereas ISEs are small, portable, and low-cost sensors that can also be deployed outside the laboratory.

Biological monitoring techniques such as immunoassays are now available for environmental analysis. These form the basis of low cost, rapid, and highly selective kits for the determination of specific compounds or classes of compounds, such as polyaromatic hydrocarbons, polychlorinated biphenyls, pesticides, herbicides, and insecticides, in food, water, and soil matrices. They are simple to use and are well suited to screening sites for contamination prior to sampling and laboratory analysis. Immunoassays can also be applied to blood and urine samples for assessing the exposure of organisms to contaminants.

## Speciation

Legislatively defined chemical exposure limits and quality guidelines often refer to total concentrations of an element. Speciation, however, refers to the specific physicochemical forms of an element that together make up the total metal concentration. Determination of particular chemical forms of an element can provide a much better indicator of bioavailability, toxicity, and mobility than total element concentration. For example, inorganic forms of arsenic are very toxic, whereas arsenobetaine, the form of arsenic found in aquatic organisms, is not. Conversely tributyltin (an antifouling agent) is much more toxic than inorganic tin species. Factors such as pH, redox potential, and particulate/colloidal matter influence the speciation of elements in environmental systems. For example, aluminum and other metals are leached from soils by acid rain due to their higher solubility at low pH. From an analytical perspective the determination of specific physicochemical forms of an element is particularly challenging due to their lower concentrations compared with the total element, and the enhanced selectivity required.

## Continuous Monitoring

Advances in instrumentation and communications (for data transmission) have increased the reliability of continuous monitors. Such systems can be deployed in remote and hazardous locations to provide high temporal resolution chemical data in aquatic and atmospheric environments. Examples include sensors for monitoring nutrients in rivers, instrumented buoys for oceanographic mapping, and

spectroscopic techniques for monitoring stack gas emissions.

## Quality Assurance and Data Handling

Environmental analytical data is often used as a tool for making environmental management decisions or for legislative compliance/enforcement and, therefore, good quality assurance (QA) is essential. This includes the periodic use of performance audits to ensure good laboratory practice, participation in interlaboratory comparison exercises, use of certified reference materials to assess accuracy, determination of blanks, and replicate analyses to assess precision (expressed as relative standard deviation). Regression analysis is often applied to environmental datasets, and tests for outlier rejection (e.g., Grubbs test) are also used. Techniques such as analysis of variance (ANOVA) can be used to assign and compare different sources of variability in a dataset and are therefore useful for separating analytical variability from real environmental differences. Chemometric techniques, e.g., principle component analysis (PCA), are particularly suited for investigating relationships between variables in large datasets such as those generated by environmental monitoring programs. It is essential with such datasets that the information is demonstrably reliable, i.e., statistically robust, before using them for environmental interpretation.

## Safety

Risks to persons or to the environment must be minimized by adherence to safety legislation and protocols, e.g., Occupational Health and Safety (COSSH in the UK), which informs on health risks associated with specific chemicals and appropriate medical and cleanup procedures in the event of exposure or a spillage. Before undertaking field work, a site-specific risk assessment should be carried out to identify potential risks and specify safe practices. Important issues include transport of personnel and equipment, weather conditions, possible health risks from the sample and the general environment, appropriate protective clothing, safe onsite working practices, safe use of equipment, communications, and emergency procedures.

## Air

Monitoring the natural components of air, and of emissions into it, is of global importance due to the need for clean air for respiration and the potential long-range transport of pollutants. For example,

transnational-boundary SO<sub>2</sub> emissions from fossil-fuel-fired power stations can acidify rain that in turn acidifies lakes and rivers, solubilizes metals, and reduces biodiversity. Naturally occurring but anthropogenically enhanced climate change gases, such as CO<sub>2</sub> and CH<sub>4</sub>, as well as stratospheric ozone that protects against damaging UV radiation, also require global monitoring. Air is often analyzed for legislative compliance, e.g., Air Quality Limit Values (AQLVs) and for human exposure limits to toxic gases such as benzene, especially those occurring indoors or in the workplace. Ground level ozone (a plant growth inhibitor) and oxides of nitrogen from combustion engines contribute to smog formation and impact on human health concerns such as asthma. They are therefore also routinely monitored.

## Sampling

Sampling is a critical aspect of the analytical process due to the dynamic nature of atmospheric environments, which result from chemical reactivity, rapid dispersion, and spatial/temporal variability. Accurate measurement of the volume of air sampled, sampling time, and location are important factors. The choice of extractive (onto solids or into liquids) or passive sampling is often dictated by the analyte. Passive sampling is unsuitable for labile atmospheric substances due to their reactivity and adsorption onto container walls; most reactive gases can only be measured *in situ*. However, volatile organic compounds (VOCs) are often sampled by extractive filtration, followed by GC-MS analysis. Most extractive samplers operate for 8–24 h to enable averaging over a suitable time period. There are also specialist sampling methods for aerosols, often with a focus on sub 10 µm (PM<sub>10</sub>) and especially sub 2.5 µm (PM<sub>2.5</sub>) particulates due to their penetration into the respiratory tract and alveoli of lungs.

## Analytical Techniques for Air

There are a number of techniques that can be used in the field. These include electrochemical sensors for gases such as O<sub>2</sub> and SO<sub>2</sub> and diffusive samplers containing immobilized reagents that produce a visible color change with visual detection on exposure to a specific chemical. Passive diffusion tubes can also be used for analyte preconcentration. Subsequent laboratory analysis is usually undertaken by thermal desorption coupled with GC. This approach is particularly useful for trace organic compounds such as polyaromatic hydrocarbons (PAHs) and VOCs. Spectrometric techniques such as Fourier transform infrared (FTIR) spectrometry, correlation spectrometry, and the laser based LIDAR (light detection



and ranging) are sophisticated instrumental techniques that can be transported in a mobile laboratory and used for large-scale remote mapping, e.g., of plumes emitted from a power station or factory stack.

## Waters

The oceans make up 97% of the world's water, with 2.5% present as freshwater. Of this, 75% is in the form of glaciers, 24% is groundwater, and only 1% is in rivers, lakes, and soils. Environmental analysis encompasses all of these different water types as well as dynamic interfacial regions such as estuaries (0–35% salinity), potable waters, waste waters, and interstitial waters. Marine waters include areas of high primary biomass productivity, e.g., coastal and shelf waters, which are often rich in macronutrients (nitrate and phosphate) and metals (from riverine and short-range atmospheric deposition). The water column is usually stratified with a shallow, well mixed upper layer (euphotic zone) and has distinct seasonal influences. Open ocean environments are characterized by less well-differentiated waters, e.g., central gyres, that have a deeper euphotic zone, and for a number of elemental constituents are strongly influenced by long-range atmospheric deposition.

The availability of suitable quantities of high-quality freshwater, when required and in the right location, is a major global environmental issue. National and transnational water management policies should, therefore, be knowledge based and designed to sustain, protect, and enhance waters while maintaining economic, e.g., fisheries or agriculture, and social use, e.g., drinking water, recreational activity, and visual amenity. Within Europe, for example, the 2000 Water Framework Directive provides a comprehensive legal and management framework for an integrated and coordinated approach to sustainable water management rather than using single chemical indicators of water quality such as dissolved oxygen.

## Sample Preservation

It is essential to preserve the integrity of the sample between the time of collection and the time of analysis. There are, however, several processes that can cause changes in the chemical composition. Examples of these include biodegradation (e.g., of nitrogen- and phosphorus-containing compounds), oxidation (e.g., of Fe(II) and organic compounds), absorption (e.g., of CO<sub>2</sub> which affects pH and alkalinity), precipitation (e.g., removal of CaCO<sub>3</sub>, Al(OH)<sub>3</sub>), volatilization (e.g., loss of NH<sub>3</sub>, HCN), and adsorption (e.g., of dissolved metals on the walls of the container).

Changes in composition can be retarded, but not prevented, by refrigeration at  $-4^{\circ}\text{C}$  (which reduces biological activity), freezing at  $-20^{\circ}\text{C}$  (but only for plastic bottles), addition of preserving agents at the point of sampling (e.g., acidification for trace metals, biocides such as HgCl<sub>2</sub>) for biodegradable determinants), and filtration (to remove biotic and abiotic particulate matter but not all bacteria and viruses). Commonly filtration through a 0.45 or 0.2- $\mu\text{m}$  filter (cellulose acetate or polycarbonate) is used to operationally separate the 'dissolved' from the 'particulate' fraction. However, this is not appropriate for separating the 'colloidal' fraction and for this fractionation techniques such as field-flow fractionation can be used.

## Analytical Techniques for Waters

Freshwater and seawater pose different analytical challenges. Matrix effects (enhancing or quenching interferences) are usually greater in seawater, e.g., dissolved solids coating out on ICP cones, or the Schlieren (refractive index) effect in flow-through analyzers with molecular spectroscopic detection, such as in the determination of phosphate using the molybdenum blue chemistry. Dissolved organic matter (DOM) is often complexed with metal ions in natural waters and this can mask their detection to varying degrees with different detectors. Atomic spectroscopic techniques generally determine total metal concentrations, whereas techniques such as voltammetry only determine the electroactive forms of the element.

Physicochemical indicators such as pH, Eh, conductivity, dissolved oxygen, temperature, light, and nutrients are the 'master variables' for water quality and can be determined in the field using portable (handheld) instrumentation. More sophisticated instrumentation is required for other parameters such as trace metals, organic compounds, and generic parameters such as chemical oxygen demand (COD) and biological oxygen demand (BOD). Each class of aquatic elements/compounds poses its own analytical challenges. Organic substances can be separated from matrix interferences by solvent extraction followed by back-extraction into a suitable medium, e.g., nitric acid. Volatile organic compounds can be separated by automated solid-phase microextraction (SPME) or purge-and-trap techniques. GC-MS is commonly used for screening and quantifying organic compounds due to its high specificity and sensitivity.

Trace metals in waters are commonly determined by atomic spectrometry (FAAS, ETAAS, derivatization (hydride/cold vapor), AAS, ICP-AES, and ICP-MS),

voltammetry (CSV and ASV), and flow injection (FI) with molecular spectroscopic detection. It is also possible to predict water quality using biological monitoring techniques to evaluate the interaction between environmental factors and physiological processes operating at different scales. *In situ* analysis using portable instrumentation enables the collection of near real-time data, provides high temporal and/or spatial resolution, and minimizes the need for sample preservation, storage, and transport. Reliable monitors can also reduce the cost of surveillance programs.

## Soils and Sediments

Soils and sediments are important reservoirs in the lithosphere, acting as sinks and sources for many chemical species. Soils are a key component of both natural and agricultural terrestrial ecosystems and are essential for the growth of plants and the degradation and recycling of biomass. They are highly variable but can be broadly characterized by soil texture and hence classified by particle-size distribution into sand (coarse) and silt and clay (fine) fractions. Key soil properties that define its chemical behavior include pH, content of organic matter, mineralogy (aluminosilicates and oxides of iron, manganese, and aluminum), cation exchange capacity, and its three-dimensional structure. Soils also contain water, air, and living biota and can be vertically subdivided into layers or horizons through a soil profile. These consist of an uppermost organic rich, leaf litter layer (L layer or O horizon), overlying an eluvial mineral/organic A horizon from which materials are generally washed downward, an illuvial B horizon of altered parent material in which materials accumulate and a C horizon which is the parent material. If unweathered rock is present below the C horizon it is known as bedrock (R layer).

Sediments are the particulate solid material that accumulates beneath aquatic environments and can be broadly classified into three types. Lithogenous sediments, which are produced directly from the disintegration of preexisting rocks and volcanic debris; biogenic sediments that are the remains of organisms; and hydrogenous sediments formed by *in situ* chemical reactions, e.g., precipitates from seawater. Sedimentary biogeochemistry is therefore determined by lithological composition and by hydrological, biological, geological, and anthropogenic influences. Sediment chemicals can be remobilized through dredging activities or changes in physicochemical parameters such as bottom water currents, pH, or dissolved oxygen concentration. For example Cu, Cd, and Zn can be remobilized by oxidation of the

sulfidic phases and organic matter in contaminated, anoxic benthic sediments. Aquatic systems with high particulate loading such as turbid estuaries can also impact significantly on sediment biogeochemistry.

## Extraction Techniques

The simplest sediment/soil extraction technique involves mixing (e.g., shaking or ultrasonic blending) with an appropriate solvent at room temperature, followed by filtration or centrifugation. For non- and semipolar organic compounds Soxhlet extraction is often used but thorough wetting of the soil/sediment is essential to obtain an efficient extraction. Column extraction using an appropriate solvent and supercritical fluid extraction (SFE) are also used. SFE requires less material and time, and offers a more selective extraction by alteration of the extraction temperature and/or pressure, or the use of modifiers. Since the extraction temperatures are relatively low, thermally labile components are protected from decomposition. As the supercritical fluids used for extraction are gases at ambient temperature, a concentration step after extraction is not required. Volatile organic compounds in soil can be extracted by purge-and-trap or headspace analysis, which provide a matrix free sample for analysis.

For the extraction of metals, sequential extraction schemes such as 'Tessier' and 'BCR' are often used. This approach mimics various natural conditions under which elements are released from sediments and soils. The BCR scheme extracts sequentially with (1) acetic acid (mild acid), (2) hydroxylamine hydrochloride (reducing agent), and (3) hydrogen peroxide and ammonium acetate (oxidizing agent). Each stage is interspersed with centrifugation and washing of the sample to give three extracted fractions. The Tessier scheme (five stages) has been used to separate, for example, the components of coal ash resulting from coal-fired power generation (distilled water for the adsorbed fraction, ammonium acetate for the ion exchangeable fraction, ammonium oxalate/oxalic acid for the oxyhydroxides of iron, manganese, and some aluminosilicates, acidic  $H_2O_2$  for the organic fraction, and concentrated HCl for the remainder of the aluminosilicates). A simple approach to determine the 'bioavailable' fraction of a soil or sediment is extraction with a dilute acid to mimic the pH of the stomach. Sequential extraction has also been used to enable metabolic profiling of plant tissues by separating plant metabolites with similar properties, e.g., using water for salts, amino acids, carbohydrates, and organic acids, methanol for phenolics, flavonoids, and other polar organic compounds, and chloroform for lipids and terpenes.

See also: **Air Analysis:** Sampling; Outdoor Air; Workplace Air. **Geochemistry:** Inorganic; Sediment; Soil, Major Inorganic Components; Soil, Minor Inorganic Components; Soil, Organic Components. **Sample Handling:** Comminution of Samples; Sample Preservation; Automated Sample Preparation; Robotics. **Sampling:** Theory; Practice. **Water Analysis:** Potable Water; Freshwater; Seawater – Organic Compounds; Seawater – Inorganic Compounds; Organic Compounds; Biochemical Oxygen Demand; Chemical Oxygen Demand; Oil Pollution; Microbiological; Algal and Microbial Toxins; Particle Characterization.

## Further Reading

Barcelo D (ed.) (1993) *Environmental Analysis – Techniques, Applications, and Quality Assurance*. Oxford: Elsevier.  
Burden FR, McKelvie I, Forstner U, and Guenther A (eds.) (2002) *Environmental Monitoring Handbook*. New York: McGraw-Hill.

Curran S (1998) *Environmental Handbook*. London: The Stationery Office.  
Day R, Reader JA, and Rowland E (2003) *Health, Safety and Environment Legislation, a Pocket Guide*. Cambridge: Royal Society of Chemistry.  
Dean JR (2003) *Methods for Environmental Trace Analysis*. Chichester: Wiley.  
Fifield FW and Haines PJ (eds.) (2000) *Environmental Analytical Chemistry*, 2nd edn. Oxford: Blackwell Science.  
Garbutt J (1995) *Environmental Law, a Practical Handbook*. Chichester: Wiley.  
Loconto PR (2001) *Trace Environmental Quantitative Analysis, Principles, Techniques and Applications*. New York: Dekker.  
Radojevic M and Bashkin VN (1999) *Practical Environmental Analysis*. Cambridge: Royal Society of Chemistry.  
Reeve RN and Barnes JD (1994) *Environmental Analysis*. Chichester: Wiley.  
Ure AM and Davidson CM (2001) *Chemical Speciation in the Environment*, 2nd edn. London: Blackwell.

# ENZYMES

Contents

## Overview

### Immobilized Enzymes

### Enzyme-Based Electrodes

### Enzymes in Physiological Samples

### Industrial Products and Processes

### Enzyme-Based Assays

## Overview

**A Narváez and E Domínguez**, University of Alcalá, Madrid, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Living systems rely on kinetic stability for their existence. They are all thermodynamically unstable; that is, they would burn up to form carbon dioxide and water if the system reaches thermodynamic equilibrium. In other words, life processes depend upon the ability to restrict these thermodynamic tendencies by controlled kinetics to produce energy as needed. A key feature of living systems is thus the use of catalysts for the controlled release of that energy. Examples of such catalysts are enzymes that control the synthesis and degradation of biologically important molecules.

Preliminary evidence for enzymes can be attributed to the Swedish chemist Jon Jakob Berzelius, who first assumed in 1835 that the reactions of living organisms were the result of a new power coining the word 'catalysis'. In 1878, the term enzyme, from the Greek *enzymé* meaning 'in yeast' was introduced. In 1890, Emil Fischer intuitively proposed that substrate and enzyme behaved like key in lock illustrating that a substrate had a matching shape to fit into the enzyme active site. Experimentation on enzymes began in 1897, when Eduard Büchner demonstrated enzyme activity in a crude cell-free extract. In 1926, the basis for proving that enzymes are proteins was settled when James B. Sumner crystallized urease from jack bean extracts. Fifteen years later, the first trials to incorporate enzymes to industrial use were attempted and 10 years later the first analytical application of an enzyme was to be described.

Like all catalysts, enzymes increase the rate of chemical reactions without altering the equilibrium constant of the catalyzed reaction and without

See also: **Air Analysis:** Sampling; Outdoor Air; Workplace Air. **Geochemistry:** Inorganic; Sediment; Soil, Major Inorganic Components; Soil, Minor Inorganic Components; Soil, Organic Components. **Sample Handling:** Comminution of Samples; Sample Preservation; Automated Sample Preparation; Robotics. **Sampling:** Theory; Practice. **Water Analysis:** Potable Water; Freshwater; Seawater – Organic Compounds; Seawater – Inorganic Compounds; Organic Compounds; Biochemical Oxygen Demand; Chemical Oxygen Demand; Oil Pollution; Microbiological; Algal and Microbial Toxins; Particle Characterization.

## Further Reading

Barcelo D (ed.) (1993) *Environmental Analysis – Techniques, Applications, and Quality Assurance*. Oxford: Elsevier.  
Burden FR, McKelvie I, Forstner U, and Guenther A (eds.) (2002) *Environmental Monitoring Handbook*. New York: McGraw-Hill.

Curran S (1998) *Environmental Handbook*. London: The Stationery Office.  
Day R, Reader JA, and Rowland E (2003) *Health, Safety and Environment Legislation, a Pocket Guide*. Cambridge: Royal Society of Chemistry.  
Dean JR (2003) *Methods for Environmental Trace Analysis*. Chichester: Wiley.  
Fifield FW and Haines PJ (eds.) (2000) *Environmental Analytical Chemistry*, 2nd edn. Oxford: Blackwell Science.  
Garbutt J (1995) *Environmental Law, a Practical Handbook*. Chichester: Wiley.  
Loconto PR (2001) *Trace Environmental Quantitative Analysis, Principles, Techniques and Applications*. New York: Dekker.  
Radojevic M and Bashkin VN (1999) *Practical Environmental Analysis*. Cambridge: Royal Society of Chemistry.  
Reeve RN and Barnes JD (1994) *Environmental Analysis*. Chichester: Wiley.  
Ure AM and Davidson CM (2001) *Chemical Speciation in the Environment*, 2nd edn. London: Blackwell.

# ENZYMES

Contents

**Overview**

**Immobilized Enzymes**

**Enzyme-Based Electrodes**

**Enzymes in Physiological Samples**

**Industrial Products and Processes**

**Enzyme-Based Assays**

## Overview

**A Narváez and E Domínguez**, University of Alcalá, Madrid, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Living systems rely on kinetic stability for their existence. They are all thermodynamically unstable; that is, they would burn up to form carbon dioxide and water if the system reaches thermodynamic equilibrium. In other words, life processes depend upon the ability to restrict these thermodynamic tendencies by controlled kinetics to produce energy as needed. A key feature of living systems is thus the use of catalysts for the controlled release of that energy. Examples of such catalysts are enzymes that control the synthesis and degradation of biologically important molecules.

Preliminary evidence for enzymes can be attributed to the Swedish chemist Jon Jakob Berzelius, who first assumed in 1835 that the reactions of living organisms were the result of a new power coining the word 'catalysis'. In 1878, the term enzyme, from the Greek *enzymé* meaning 'in yeast' was introduced. In 1890, Emil Fischer intuitively proposed that substrate and enzyme behaved like key in lock illustrating that a substrate had a matching shape to fit into the enzyme active site. Experimentation on enzymes began in 1897, when Eduard Büchner demonstrated enzyme activity in a crude cell-free extract. In 1926, the basis for proving that enzymes are proteins was settled when James B. Sumner crystallized urease from jack bean extracts. Fifteen years later, the first trials to incorporate enzymes to industrial use were attempted and 10 years later the first analytical application of an enzyme was to be described.

Like all catalysts, enzymes increase the rate of chemical reactions without altering the equilibrium constant of the catalyzed reaction and without



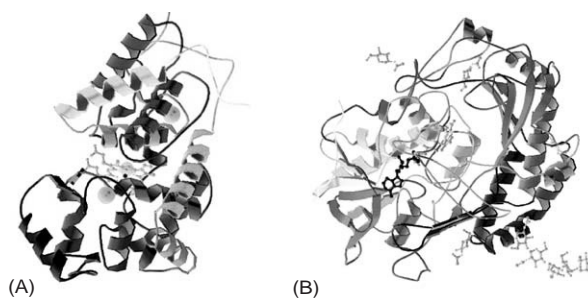
undergoing any net change in their structure. Enzymes are very efficient catalysts, often far superior to conventional chemical catalysts. The relative values over nonenzymatic rates of transformation – acceleration rates – are often  $10^{10}$  as for protease-mediated hydrolysis of peptide bonds, and can reach  $10^{23}$  in the case of orotidine-5'-phosphate decarboxylase (EC 4.1.1.23) in the pyrimidine biosynthetic pathway.

The outstanding characteristic of enzymes over conventional catalysts is their specificity and selectivity. It is this specificity that draws the interest of scientists and technologists for a wide range of applications, from industrial processes to clinical diagnosis or from environmental remediation to drug or food formulation as obvious examples of the key role that enzymes play in our daily life.

Although the fundamentals of enzymes, mainly formulated in the first half of the twentieth century still explain most of the properties of enzymes, multidisciplinary efforts have furnished a realm of new biocatalysts and synthetic enzyme mimics that expand the traditional biocatalytic world toward a new scenario of still nascent technology. The following overview covers all these aspects setting up the fundamentals for the use of enzymes in analytical science.

## Enzyme Structure and General Properties

Enzymes are proteins with the few exceptions of catalytic nucleic acids. Like proteins, enzymes are made up of amino acids linked together by amide bonds. Each peptide chain has one free amino end, the N-terminus, and one free carboxyl end, the C-terminus. The sequence of these amino acids determines the primary structure. Because of the properties of the amino acids, the structure may possess a helical structure ( $\alpha$ -helices) and other possible arrangements such as turns and pleated sheets ( $\beta$ -sheets) resulting in the secondary structure. The tertiary structure of an enzyme refers to its three-dimensional (3D) arrangement in order to minimize nonfavorable thermodynamic interactions with its environment. This 3D structure forms the sites that are directly involved in binding the substrate, the prosthetic group and any other cofactor. The vast majority of known enzymes are composed of two or more polypeptide chains usually linked to each other by noncovalent interactions. These polypeptide chains are termed subunits and may be identical to or different from each other. The combination and spatial arrangement of the subunits is known as the quaternary structure. **Figure 1** shows the 3D

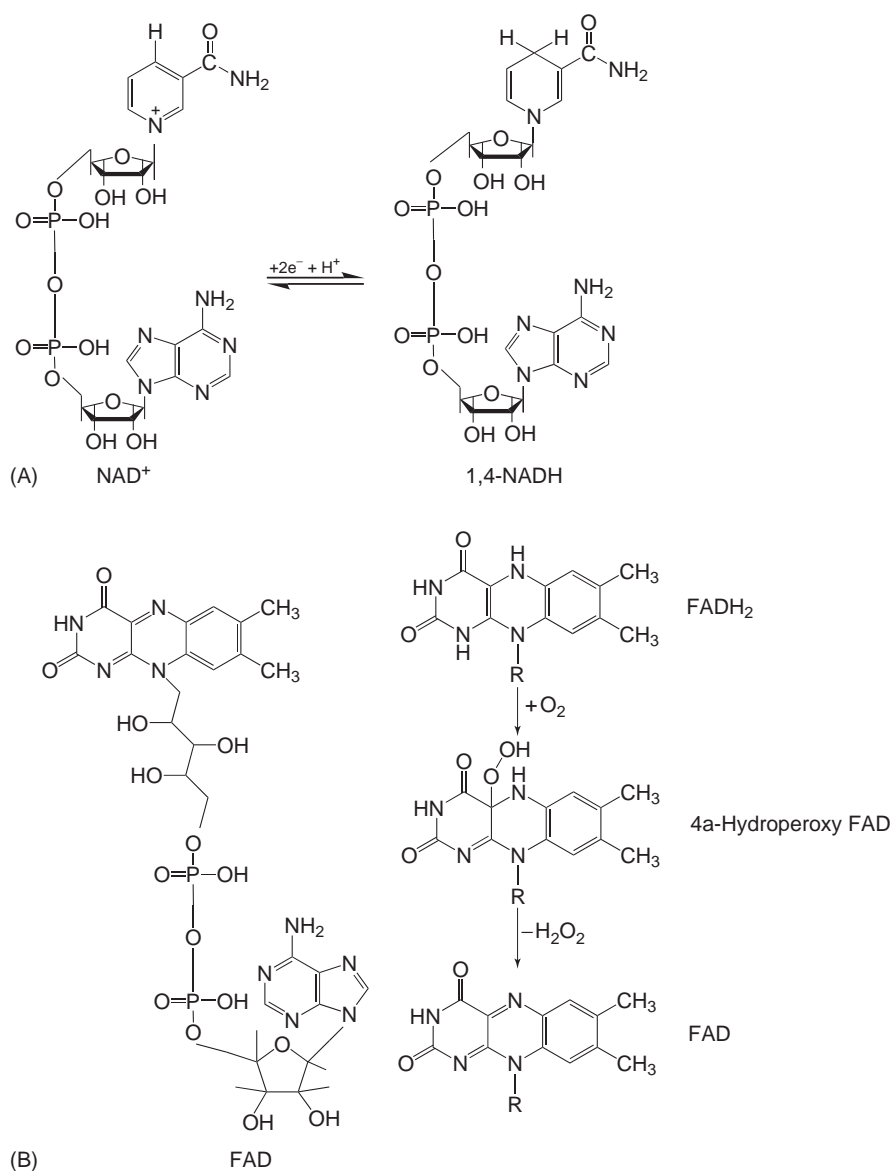


**Figure 1** Three-dimensional view of (A) horseradish peroxidase (HRP 1HCH) and (B) glucose oxidase (GOx 1GAL). HRP from *Armoracia rusticana* is a glycosylated (not shown) monomeric protein with one heme group and two calcium ions. GOx from *Aspergillus niger* is a glycosylated dimeric protein (only one subunit is shown) containing one flavin adenine dinucleotide per monomer. These structures have been obtained from the Protein Data Bank and displayed with the MOLSCRIPT program.

structure of two enzymes of wide application in enzymatic analysis. Many enzymes are glycoproteins containing a large portion of poly- or oligosaccharides derivatives.

For catalytic activity, many enzymes require the presence of a nonprotein component, called a cofactor. The protein part of the enzyme is then termed apoenzyme and the active enzyme, including cofactor, a holoenzyme. Cofactors can range from metal ions, for instance, Zn(II) in alcohol dehydrogenase, to large, complex organic molecules such as nicotinamide adenine dinucleotide (NAD) in some dehydrogenases and flavin adenine dinucleotide (FAD) in some oxidases. The function of the metal ion can be either to provide the catalytic center of the enzyme or it may form a bridging group to link enzyme and substrate together via a coordination complex. If cofactors are tightly bound to apoenzymes, being difficult to remove without damaging the enzyme, they are termed prosthetic groups. The term coenzyme refers to cofactors that are of organic nature. **Figure 2** presents two important cofactors for oxidoreductases.

The catalytic power of enzymes relies on their ability to stabilize the transition state of a reaction relative to that of the ground state ( $\Delta G_{ETS}^{\ddagger} > \Delta G_{ES}^{\ddagger}$ ). This is illustrated in **Figure 3**, also showing that the enzyme–substrate complex is stabilized relative to the free species in solution. Overall, enzymes decrease the Gibbs energy of activation (standard free energy of activation) as compared with non-catalyzed reactions. For instance, the cleavage of hydrogen peroxide normally requires an activation energy of  $75.4 \text{ kJ mol}^{-1}$  and catalase (EC 1.11.1.6) diminishes this energy to  $23 \text{ kJ mol}^{-1}$ .

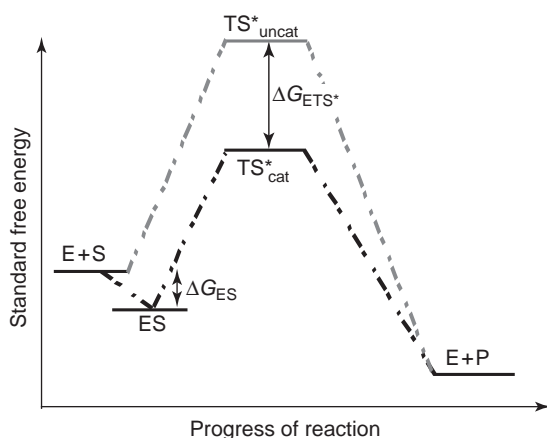


**Figure 2** Structural formulae of two enzyme cofactors: (A) nicotinamide adenine dinucleotide (NAD) and (B) flavin adenine dinucleotide (FAD) with the corresponding redox transformations in dehydrogenases and oxidases catalyzed reaction, respectively. NAD is a soluble cofactor and it has to be added to the reaction mixture. As shown in **Figure 1**, FAD is bound to the flavoprotein.

Enzymes can exhibit catalytic activity for chemical species that have a given functional group (e.g., phosphate or methyl groups) and have group specificity. If they only act on certain types of chemical bond (e.g., C–C) they show linkage specificity. Stereochemical specificity refers to catalytic activity on reaction of particular stereoisomers or chiral isomers. The stereochemical specificity of enzymes calls for at least three different points of interaction between enzyme and substrate. These interactions can have either a binding or a catalytic function. Binding sites, in the binding domain of the protein, link to specific groups in the substrate, ensuring that the

enzyme and substrate are held together in a fixed orientation with respect to each other. Catalytic sites, in the catalytic domain, link the group or groups of the substrate that are directly involved in the catalysis. The substrate binding site and the catalytic site are usually near one another and frequently overlap. The binding site may be hydrophobic or hydrophilic depending on the complementary structure of substrates; the catalytic site is usually hydrophilic. The region of an enzyme's 3D structure that contains the substrate binding and catalytic site is called the active site or active center. The binding of substrates to enzymes requires the interplay of electrostatic,





**Figure 3** Free energy diagram for catalyzed and uncatalyzed reactions. S, substrate; E, enzyme; P, product; ES, enzyme–substrate complex;  $TS^*$ , transition state;  $\Delta G$ , standard free energy of activation.

hydrogen bonds, and van der Waals interactions. This binding of substrates to enzymes is well explained by the induced fit hypothesis, meaning that upon binding the substrate causes an alteration on the enzyme's conformation resulting in an appropriate orientation for catalysis.

### Heteroenzymes and Isoenzymes

Enzymes exhibiting the same catalytic function are known as homologous enzymes and fall into two classes: heteroenzymes and isoenzymes. The first group includes enzymes derived from different sources but catalyzes identical reaction, yet show different chemical and kinetic characteristics. For instance, the enzyme alcohol dehydrogenase (EC 1.1.1.1), which is found in yeast, is different from the enzyme obtained from horse liver. It is also possible to find this enzyme in several bacteria, plants, and insects. The mammalian enzyme is a dimer with a relative molecular mass of 80 kDa. Two zinc ions are firmly bound to each subunit; one is essential in the catalysis, acting as a Lewis acid. Yeast alcohol dehydrogenase is a tetramer of relative molecular mass of 140–150 kDa, made up of four subunits and containing four active sites. It is generally accepted that there is one zinc ion per subunit. Isoenzymes are different molecular forms of the same enzyme that are found in the same animal or organism, although they show a pattern of distribution between tissues. Lactate dehydrogenase (EC 1.1.1.27) is a classic example of an isoenzyme. In humans, it is composed of two different subunits known as A and B arranged into tetramers and giving five different hybrid forms of the two basic units ( $A_4$ ,  $A_3B$ ,  $A_2B_2$ ,  $AB_3$ , and  $B_4$ ).

They all show different substrate selectivity and inhibition and activation effects, permitting one isoenzyme to function under conditions that would reduce the activity of another. It is likely that such variations aid the control of the same reaction under different cellular or tissue conditions.

### Enzyme Activity and Enzyme Concentration

In principle, the amount of a biocatalyst may be expressed as a number of elementary entities, an amount of substance, a mass, or by a catalytic effect. In many cases, due to the lack of information or for practical reasons, the later route is used. Then, the generally accepted term for the expression of enzyme activity is the Unit (U) defined in terms of reaction rate. One unit (U), accordingly to the International System of Units, is defined as the amount of enzyme that catalyzes the conversion of 1  $\mu\text{mol}$  of substrate (or the formation of 1  $\mu\text{mol}$  of product) in 1 min under standard conditions. The specific activity is the number of units per milligram of protein. If the relative molecular mass of the enzyme is accurately known it is then possible to express the activity as the molar catalytic activity, defined as the number of units per micromole of enzyme. This is, the number of moles of product formed, or substrate consumed, per mole of enzyme per minute. This may not correspond to the number of moles of substrate converted per enzyme active site since the enzyme may contain more than one active site. If the number of active sites per mole is known, then the activity may be expressed correspondingly as the catalytic center activity.

The use of katal (kat) is recommended in the fields of medicine and biochemistry. It differs from the previous units in that the conversion of 1 mol of substrate per second is used ( $1 \text{ kat} = 6 \times 10^7 \text{ Units}$ ;  $1 \text{ U} = 16.67 \text{ nkat}$ ).

### Enzyme Nomenclature

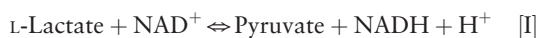
Over 3000 enzymes have so far been identified and this number may be greatly augmented in the wake of genomic and proteomic research. In view of the large number of enzymes and the wide range of catalyzed reactions, it is necessary to have a classification system. Enzymes are generally named for the reaction they catalyze by adding the suffix 'ase'. A systematic scheme of classification was first adopted by the International Union of Biochemistry in 1961 on the basis of the type of catalyzed reaction. Presently, the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology guarantees a strict classification and rational

nomenclature of known and new enzymes. According to this enzyme classification (EC), all enzymes fall into six general groups and they can be identified through four numbers. The first one refers to the main type of reaction. The second and third numbers indicate the subclass and sub-subclasses, respectively, on basis of nature of the actual reaction that is catalyzed. The fourth number is the serial number in its particular subclass. The description of these numbers within the six different classes of enzymes is given in Table 1. For example, the systematic name of lactate dehydrogenase is L-lactate: NAD oxidoreductase with EC number: EC 1.1.1.27. The first number '1' indicates oxido-reductase; the second number '1' of the subclass indicates that it acts on CH–OH groups; the sub-subclass or subgroup is denoted by the third number '1' indicating that NAD is used as H-acceptor; and the last number '27' shows that it is the 27th enzyme in subgroup 1.

The six classes of enzymes are:

**EC1. Oxidoreductases:** to this class belong all enzymes catalyzing oxido-reductions. The substrate oxidized is regarded as hydrogen or electron donor. The classification is based on 'donor:acceptor oxidoreductase'. This class is very important in analytical applications and the most relevant reactions can be summarized as:

**Dehydrogenases.** They catalyze hydrogen transfer from a substrate to an acceptor that is not molecular oxygen; or vice versa. Example: lactate dehydrogenase (EC 1.1.1.27)



**Oxidases.** They catalyze hydrogen transfer from the substrate to molecular oxygen. Example: glucose oxidase (EC 1.1.3.4)



**Peroxidases.** They catalyze the oxidation of a substrate by hydrogen peroxide. Example: horseradish peroxidase (EC 1.11.1.7)

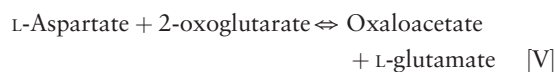


**Oxygenases.** They catalyze substrate oxidation by molecular oxygen with a hydrogen donor that may be the substrate itself. Example: monophenol monooxygenase (EC 1.14.18.1)

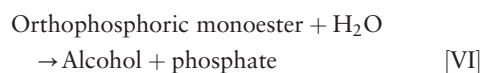


**EC2. Transferases:** enzymes transferring a group, for example, the methyl group or a glycosyl

group, from one compound – generally regarded as the donor – to another compound – generally regarded as the acceptor. The classification is based on the scheme 'donor:acceptor grouptransferase'. Transaminases, transketolases, transaldolases, and transmethyldases belong to this class. Example: aspartate aminotransferase (EC 2.6.1.1)



**EC3. Hydrolases:** these enzymes catalyze the hydrolysis of various bonds. Some of these enzymes pose problems because they have a broad selectivity. According to the type of bond cleaved, a distinction is made between peptidases, esterases, glycosidases, phosphatases, etc. Example: alkaline phosphatase (EC 3.1.3.1)



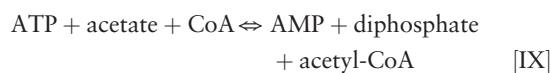
**EC4. Lyases:** enzymes cleaving C–C, C–O, C–N, and other bonds by other means than by hydrolysis or oxidation. They differ from other enzymes in that two substrates are involved in one reaction direction, but only one in the other direction. Example: mandelonitrile lyase (EC 4.1.2.10)



**EC5. Isomerases:** these enzymes catalyze changes within one molecule. They include racemases, epimerises, mutases, *cis-trans*-isomerases, etc. Example: aldose mutarotase (EC 5.1.3.3)

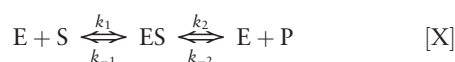


**EC6. Ligases:** enzymes that catalyze the joining of two molecules with concomitant hydrolysis of the diphosphate bond in ATP or a similar triphosphate. 'Ligase' is commonly used for the common name, but, in a few cases, 'synthase' or 'carboxylase' is used. Example: acetate CoA ligase (EC 6.2.1.1)



## Fundamentals of Enzyme Kinetics

The simplest model of an enzyme-catalyzed reaction is illustrated by:



**Table 1** International classification of enzymes*EC 1. Oxidoreductases: Acting on*

EC 1.1. CH–OH group of donors	EC 1.2. Aldehyde or oxo group of donors	EC 1.3. CH–CH group of donors	EC 1.4. CH–NH <sub>2</sub> group of donors	EC 1.5. CH–NH group of donors
EC 1.6. NADH or NADPH	EC 1.7. Other nitrogenous compounds as donors	EC 1.8. Sulfur group of donors	EC 1.9. Heme group of donors	EC 1.10. Diphenols and related substances as donors
EC 1.11. Peroxide as acceptor	EC 1.12. Hydrogen as donor	EC 1.13. Single donors with incorporation of molecular oxygen (oxygenases)	EC 1.14. Paired donors, with incorporation or reduction of molecular oxygen	EC 1.15. Superoxide radicals as acceptor
EC 1.16. Oxidizing metal ions	EC 1.17. CH <sub>2</sub> groups	EC 1.18. Iron–sulfur proteins as donors	EC 1.19. Reduced flavodoxin as donor	EC 1.20. Phosphorus or arsenic in donors
EC 1.21. X–H and Y–H to form an X–Y bond	EC 1.97. Other oxidoreductases			

*EC 2. Transferases: Transferring*

EC 2.1. One-carbon groups	EC 2.2. Aldehyde or ketonic groups	EC 2.3. Acyl group	EC 2.4. Glycosyl group	EC 2.5. Alkyl or aryl groups, other than methyl groups
EC 2.6. Nitrogenous groups	EC 2.7. Phosphorus-containing groups	EC 2.8. Sulfur-containing groups	EC 2.9. Selenium-containing groups	

*EC 3. Hydrolases: Acting on*

EC 3.1. Ester bonds	EC 3.2. Glycosylases	EC 3.3. Ether bonds	EC 3.4. Peptide bonds (peptidases)	EC 3.5. Carbon–nitrogen bonds, other than peptide bonds
EC 3.6. Acid anhydrides	EC 3.7. Carbon–carbon bonds	EC 3.8. Halide bonds	EC 3.9. Phosphorus–nitrogen bonds	EC 3.10. Sulfur–nitrogen bonds
EC 3.11. Carbon–phosphorus bonds	EC 3.12. Sulfur–sulfur bonds	EC 3.13. Carbon–sulfur bonds		

*EC 4. Lyases*

EC 4.1. Carbon–carbon lyases	EC 4.2. Carbon–oxygen lyases	EC 4.3. Carbon–nitrogen lyases	EC 4.4. Carbon–sulfur lyases	EC 4.5. Carbon–halide lyases
EC 4.99. Other lyases				

*EC 5. Isomerases*

EC 5.1. Racemases and epimerases	EC 5.2. <i>cis</i> – <i>trans</i> -isomerases	EC 5.3. Intramolecular isomerases	EC 5.4. Intramolecular transferases: mutases	EC 5.5. Intramolecular lyases
EC 5.99. Other isomerases				

*EC 6. Ligases: Forming*

EC 6.1. Carbon–oxygen bonds	EC 6.2. Carbon–sulfur bonds	EC 6.3. Carbon–nitrogen bonds	EC 6.4. Carbon–carbon bonds	EC 6.5. Phosphoric ester bonds
EC 6.6. Nitrogen–metal bonds				

This mechanism consists of two steps; the first one is the binding of enzyme (E) to substrate (S) forming a complex (ES), and the second one includes both the catalysis of the reaction and the release of the

product (P). The formation of an intermediate enzyme–substrate complex has been demonstrated experimentally and is determined by the ratio of two rate constants  $k_1$  and  $k_{-1}$ . These rate constants typically

show values of  $10^5$ – $10^8$   $\text{l mol}^{-1} \text{s}^{-1}$  and  $1$ – $10^4$   $\text{s}^{-1}$ , respectively. For the sake of simplicity, the reverse reaction, determined by  $k_{-2}$ , is not considered in the following treatment. This can be readily assumed at the beginning of the reaction when very little product is formed or when the reaction is not reversible. Thus, the rate of the reaction ( $v$ ) can be calculated as:

$$v = \frac{d[P]}{dt} = k_2[ES] \quad [1]$$

where  $k_2$  is the catalytic rate constant, which is also called  $k_{\text{cat}}$  or turnover number, and it shows typical values of  $1$ – $10^5$   $\text{s}^{-1}$ . The turnover number represents the maximum number of substrate molecules that the enzyme can convert into product per active site and per second.

The evolution of the concentration of the enzyme–substrate complex can be calculated as:

$$\frac{d[ES]}{dt} = k_1[E][S] - (k_{-1} + k_2)[ES] \quad [2]$$

Because the total enzyme concentration  $[E]_t$  equals the free enzyme concentration  $[E]$  and the concentration of the ES complex, eqn [2] can be denoted as:

$$\frac{d[ES]}{dt} = k_1([E]_t - [ES])[S] - (k_{-1} + k_2)[ES] \quad [3]$$

Rearranging terms:

$$\frac{d[ES]}{dt} = k_1[E]_t[S] - (k_1[S] + k_{-1} + k_2)[ES] \quad [4]$$

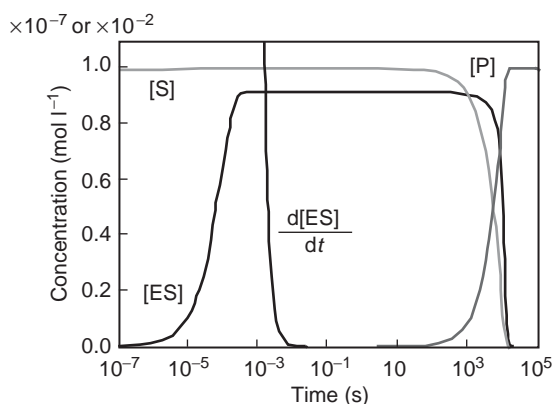
This gives:

$$\frac{d[ES]/dt}{k_1[S] + k_{-1} + k_2} + [ES] = \frac{k_1[E]_t[S]}{k_1[S] + k_{-1} + k_2} \quad [5]$$

This equation can be greatly simplified in the left-hand side assuming the steady-state approximation by which  $d[ES]/dt$  equals zero. This is, the rate of formation of ES equals its rate of disappearance. However, this is a rather restrictive and unnecessary restriction because eqn [5] can also be simplified when its differential term can be neglected in relation to the concentration of the enzyme–substrate complex:

$$\frac{d[ES]/dt}{k_1[S] + k_{-1} + k_2} \ll [ES] \quad [6]$$

This occurs during a substantial part of the reaction time-course over a wide range of kinetic rate constants and substrate concentrations and at low to moderate enzyme concentrations. An illustration of this assumption is depicted in **Figure 4** where the variation in  $[ES]$ ,  $d[ES]/dt$ ,  $[S]$ , and  $[P]$  with the time-course of the reaction is simulated for a set of values. As seen in the figure, the simplified equation is valid throughout most of the reaction.



**Figure 4** Computer simulation of the progress curves of  $d[ES]/dt$  ( $0$ – $10^{-7}$   $\text{mol l}^{-1}$  scale),  $[S]$  ( $0$ – $10^{-2}$   $\text{mol l}^{-1}$  scale), and  $[P]$  ( $0$ – $10^{-2}$   $\text{mol l}^{-1}$  scale) for a simple Michaelis–Menten kinetics with  $k_1 = 10^6$   $\text{l mol}^{-1} \text{s}^{-1}$ ,  $k_{-1} = 10^3$   $\text{s}^{-1}$ ,  $k_2 = 10$   $\text{s}^{-1}$ ,  $[E] = 10^{-7}$   $\text{mol l}^{-1}$  and  $[S] = 10^{-2}$   $\text{mol l}^{-1}$ . The simulation shows three distinct phases to the reaction time-course, an initial transient phase that lasts for about a millisecond followed by a longer steady-state phase of  $\sim 30$  min when  $[ES]$  remains constant but only a small portion of the substrate reacts. This is followed by the final phase taking  $\sim 6$  h during which the substrate is completely converted to product. (Reproduced with permission from Martin F Chaplin Enzyme Technology.)

The well-known Michaelis–Menten equation is derived from eqns [1] and [5] by introducing the Michaelis constant:

$$K_m = \frac{k_{-1} + k_2}{k_1} \quad [7]$$

Thus, the general reaction rate can be calculated as:

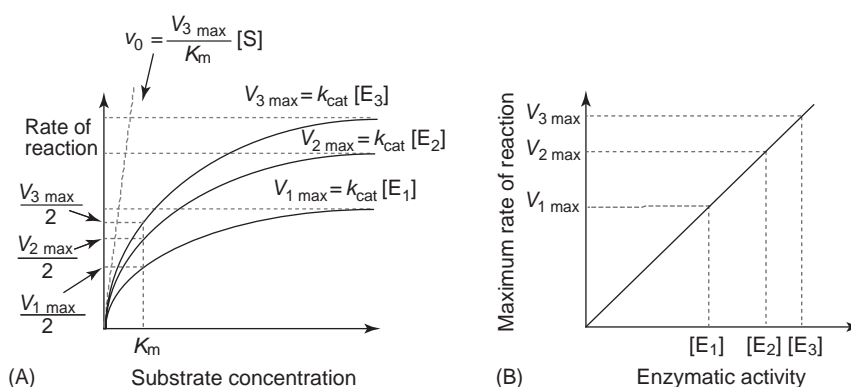
$$v = k_2[ES] = \frac{k_2[E]_t[S]}{[S] + K_m} \quad [8]$$

$k_2[E]_t$  can be considered the maximum rate of reaction ( $V_{\text{max}}$ ) as the maximum value that  $[ES]$  can have is  $[E]_t$  if  $[E]_t < [S]$ . This is, the enzyme is completely saturated with substrate that always occurs when  $[S] \gg K_m$ .

Thus, the reaction rate becomes:

$$v = \frac{V_{\text{max}}[S]}{[S] + K_m} \quad [9]$$

This Michaelis–Menten equation gives the rectangular hyperbolic dependence of the reaction rate on the substrate concentration (**Figure 5A**). This plot represents the actual  $[S]$  over the time-course of the reaction, which only equals the initial substrate concentration at the very beginning of the reaction. Consequently, rate of reactions can only be accurately estimated as initial reaction rates ( $v_0$ ) representing the tangent to the curve at the origin. Also, from this graphical representation, the



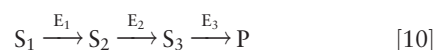
**Figure 5** Michaelis–Menten dependence of the reaction rate ( $v$ ) on the substrate concentration  $[S]$ . The hyperbolic relationship is shown in (A) for three different concentrations of enzymes ( $E_1$ ,  $E_2$ , and  $E_3$ ). A typical plot for the determination of the enzyme concentration or activity is shown in (B).

Michaelis constant can be easily interpreted as the concentration of substrate where the rate of the reaction is half of its maximum. This is, at this concentration, half of the enzyme molecules bind the substrate and, thus, it should be independent of the enzyme concentration. However, this is a rather simplistic conclusion because it considers that all enzymes strictly follow the Michaelis–Menten behavior over a wide range of experimental conditions. If not, the  $K_m$  is likely to equal a much more complex relationship between the different rate constants involved. The clear dependence between  $V_{\text{max}}$  and enzyme concentration enables a direct way for the estimation of enzyme concentration (see Figure 5B). This is the case of many clinical determinations targeting key enzymes for diagnostic purposes or the basis of many other assays, e.g., immunoassays, where enzymes are used as reporters or tracers for easy and sensitive transduction of noncatalytic interactions.

It seems clear then that enzymes can be kinetically characterized through  $V_{\text{max}}$  and  $K_m$  values although a more informative value is given by the ratio of both because it includes saturating and nonsaturating conditions of substrate concentration, respectively. For the purpose of comparison,  $k_{\text{cat}}$  instead of  $V_{\text{max}}$  is introduced and then, the ratio  $k_{\text{cat}}/K_m$  determines the relative rate of reaction at low substrate concentration. This ratio is known as the specificity constant since the discrimination between competing substrates is determined by this value that combines both the rate and the binding parameters. The higher the value, the more specific is the enzyme for the particular substrate. Values in the  $10^8$ – $10^9$   $\text{l mol}^{-1} \text{s}^{-1}$  range are claimed to be very close to the maximum rate (kinetic perfection) that could be expected from calculations of the diffusion rates of molecules.

This is to say that enzymes evolve by maximization of this ratio approaching the diffusion-control limit, which is obviously dependent on the equilibrium constant of the catalyzed reaction. Arguably this reasoning is only valid for conditions of initial rates, which are not the typical situation in living cells, where an enzyme is embedded in a chain of reactions with no initial velocities and thus, the driving force of natural selection could be contemplated as reaching the precise  $k_{\text{cat}}/K_m$  value to maximize the conversion of substrate fluxes. Table 2 summarizes the most important parameters that define the kinetics of a monomolecular enzyme-catalyzed reaction.

Similarly to complex enzymatic pathways in living systems, coupled enzyme reactions are very often used for analytical purposes. In a simplified case:



where  $S_1$  and  $P$  denote the analyte and the transducing compounds, respectively. In this case the intermediate reactions should not limit the overall reaction and thus:

$$\frac{k_{\text{cat}}}{K_m} (E_1) \leq \frac{k_{\text{cat}}}{K_m} (E_2) \leq \frac{k_{\text{cat}}}{K_m} (E_3) \quad [11]$$

The kinetics of reversible reactions demands a more laborious mathematical treatment. The reader can find a detailed description in the recommended reading. It should be stressed that enzymes can favor a reaction in one direction rather than its reverse. This is not contradictory with the fact that enzymes can never alter the equilibrium constant of a catalyzed reaction because they can favor one direction by binding strongly, as enzyme–reactants complexes, the

**Table 2** Kinetic terms used in enzyme catalyzed reactions:  $E + S \xrightleftharpoons[k_{-1}]{k_1} ES \xrightleftharpoons[k_{-2}]{k_2} E + P$ 

Definition	Symbols	Mathematical expression	Units
Maximum rate of reaction	$V_{\max}$	$k_2[E]_t$	$\text{mol l}^{-1} \text{s}^{-1}$
Rate of reaction	$v$	$V_{\max}[S]/([S] + K_m)$	$\text{mol l}^{-1} \text{s}^{-1}$
Initial rate of reaction	$v_0$	$V_{\max}[S]/K_m$	$\text{mol l}^{-1} \text{s}^{-1}$
Michaelis constant	$K_m$	$k_{-1} + k_2/k_1$	$\text{mol l}^{-1}$
Catalytic rate constant or turnover number	$k_2(k_{\text{cat}})$	$V_{\max}/[E]_t$	$\text{s}^{-1}$
Specificity constant	$k_{\text{cat}}/K_m$		$\text{l mol}^{-1} \text{s}^{-1}$
Acceleration rate	$k_{\text{cat}}/k_{\text{uncat}}$		
Equilibrium constant of the reaction	$K$	$k_1 k_2/k_{-1}$	

reactants in the preferred direction whilst binding weakly the products. Consequently, it is unlikely that the same enzyme would be optimum for catalyzing a reversible reaction in both directions.

The knowledge of these kinetic constants is then of high importance in order to choose the right enzyme and the right conditions for an analytical application. Either endpoint or kinetic methods can be applied. If a nearly complete conversion of substrate in a reasonable time is pursued, then the  $K_m$  of the enzyme should be as small as possible. This is of special importance to maintain a relatively high rate of reaction for the determination of analytes at low concentration ( $[S] < K_m$ ). Because the reaction rate is directly proportional to the amount of enzyme (eqn [9]), a high concentration of enzyme, represented by  $V_{\max}$ , is desirable to shorten the time of analysis or response time:

$$t = \frac{X[S]}{V_{\max}} - \frac{K_m \ln(1 - X)}{V_{\max}} \quad [12]$$

where X denotes the fractional conversion  $[S] - [P]/[S]$ .

Alternatively, if the equilibrium constant is not favorable or the  $K_m$  is too high in relation to the target concentrations ( $[S] \ll K_m$ ), then a kinetic method can be applied. Equation [9] can then be simplified to:

$$v = \frac{V_{\max}[S]}{K_m} \quad [13]$$

In general, to maintain a direct proportionality between the rate of reaction and the concentration of analyte, this should be much smaller than the  $K_m$  ( $[S] < K_m$ ). The specificity constant and the amount of enzyme will then determine the sensitivity of the assay whilst its dynamic range will be limited by the value of  $K_m$ .

For the estimation of  $K_m$  and  $V_{\max}$ , different graphical plots can be obtained after measurement of the reaction rates over a wide range of substrate concentration. These methods are illustrated in Figure 6. The Lineweaver–Burk plot or double reciprocal

( $1/v$  versus  $1/S$ ), or the Eadie–Hofstee plot ( $v$  versus  $v/S$ ) should never be used to calculate  $K_m$  and  $V_{\max}$  values for publication. The Lineweaver–Burk plot is preferred to test for the qualitative correctness of the enzyme mechanism and the Eadie–Hofstee plot is preferred for discovering deviations from linearity. The Hanes plot ( $S/v$  versus  $S$ ) is less dangerous but, even so, for statistical reasons these three plots should not be used for anything other than display purposes. There is only one statistically sound method for the determination of the best fit  $K_m$  and  $V_{\max}$  values by visual inspection of a graph. This is the direct linear plot that uses a nonparametric criterion of best fit, in which the median values of a series of estimates of  $K_m$  and  $V_{\max}$  are shown to be the best. This method depends upon the assumption that any errors in the experimental data are as likely to be positive (too high) as negative (too low). The usual objection to any graphical method of calculating these values is that confidence limits or standard deviations are not determined whilst the direct linear plot does not suffer from this disadvantage.

### Allosteric Enzymes

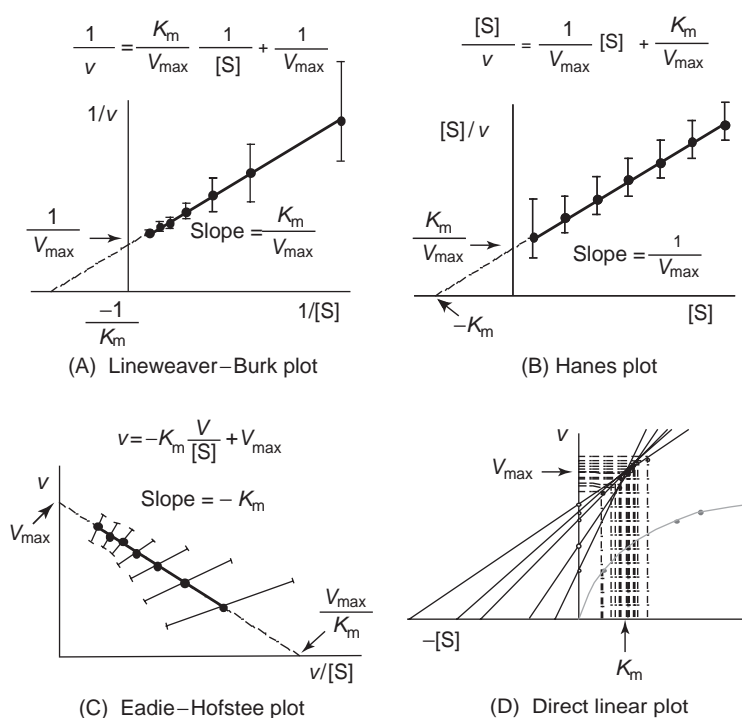
A large family of enzymes that deviate from hyperbolic kinetics (Michaelis) is the allosteric enzymes. These enzymes contain two or more topologically distinct binding sites that interact functionally with each other. Most commonly, sigmoidal or S-shaped curves are obtained, being indicative of positive substrate cooperativity. The reaction rate for these enzymes can be calculated by the Hill equation:

$$v = \frac{V_{\max}[S]^b}{K^b + [S]^b} \quad [14]$$

Rearrangement of this equation to a linear form gives:

$$\log \frac{v}{V_{\max} - v} = b \log [S] - \log K \quad [15]$$





**Figure 6** Different graphical representations of the estimated reaction rates at different concentrations of substrate. The different plots show the relative weight of standard deviations at the given concentrations assayed.

$K$  and  $h$  represent a kind of Michaelis constant and the Hill coefficient, respectively. If  $V_{\max}$  is known, these constants can be calculated from the slope and intercept of the Hill plot (eqn [15]). If the Hill coefficient results equal to one, then there is no cooperativity and the graph is hyperbolic. An increasing value of  $h$  will show an increasingly sigmoidal curve with positive cooperativity for the substrate. A value less than one shows negative cooperativity. The reaction rate of these enzymes is easily controlled by allosteric effectors, activators, or inhibitors, this mechanism being of crucial importance for the control of metabolic pathways. Equally important, this mechanism can also be used for the detection of analytes.

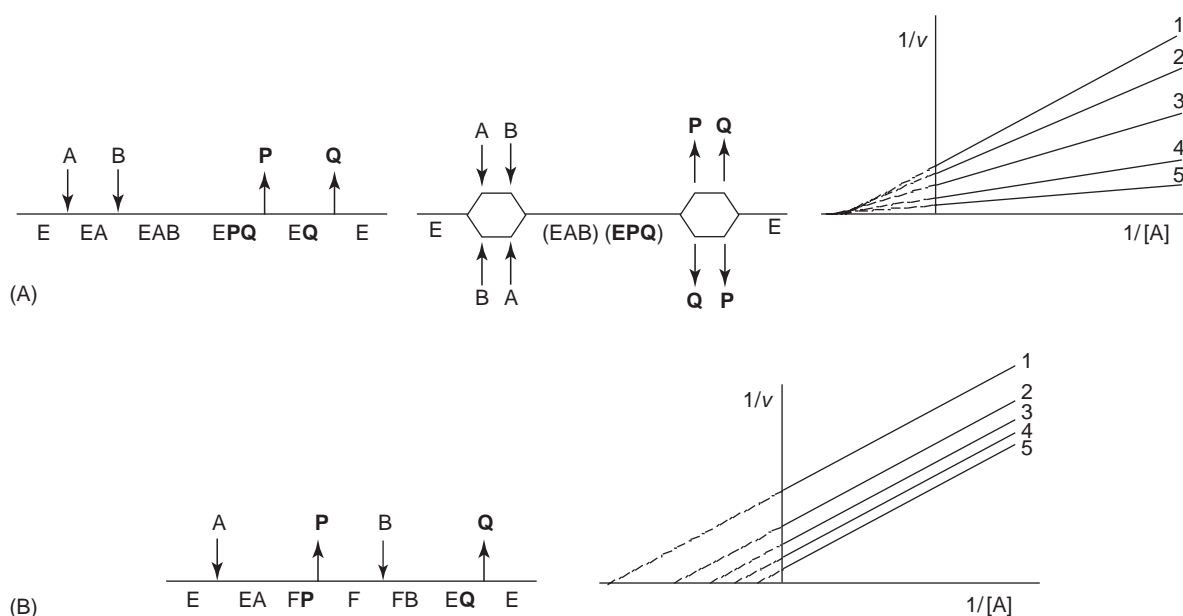
## Kinetic Mechanisms of Multisubstrate Systems

Many enzyme-catalyzed reactions involve two or more substrates and enzyme complexes and hence they deviate from the Michaelis-Menten treatment. A direct simplification of this situation consists of keeping one substrate at nonlimiting concentration. For instance, reactions with  $\text{NAD}^+/\text{NADH}$  dependent dehydrogenases or oxidases using molecular oxygen as electron donor can be readily

simplified to single substrate kinetics by using a saturating concentration of the soluble cofactor or dissolved oxygen, respectively (pseudo-first-order conditions).

Kinetic mechanisms of multisubstrate reactions can be broadly divided into two main types, sequential and ping-pong (also known as double displacement). These mechanisms can be easily illustrated by Cleland plots where the enzyme is represented as a horizontal line and arrows denote the addition of substrates and release of products (Figure 7). In sequential reactions, all substrates bind to the enzyme before the first product is released. They can be further subdivided into ordered and random reactions. The ordered sequential mechanism imposes that the binding of substrates to the enzyme and the release of products follow a defined sequence. In contrast, the random sequential mechanism does not demand any specified order and alternate reaction sequences exist. The distinguishing feature of enzymes under a ping-pong mechanism is that at least one product is released from the enzyme before all of substrates have bound and the enzyme will exist in two or more stable forms between which it oscillates during the reaction.

These mechanisms can be distinguished by keeping one substrate (e.g., A in Figure 7) at a constant concentration (fixed concentration) and varying the concentration of the other substrate. Then, if increased



**Figure 7** Cleland plots of the different mechanisms in multisubstrate enzyme-catalyzed reactions: (A) sequential mechanism, ordered (left) and random (right); (B) ping-pong mechanism. The double reciprocal plots identify either mechanism. Ordered and random show identical data display. 1, 2, 3, ..., 5 denote increasing concentrations of the second substrate.

concentrations of the fixed substrate are assayed and data obtained are represented as Lineweaver–Burk plots, the different pattern of lines is indicative of the enzyme mechanism.

## Factors Affecting Enzyme Activity

Most enzymes are adapted to function in a rather limited set of physiological conditions: chemical and ionic composition of the medium, pH, temperature, pressure, or other chemical species.

### Temperature

Like all reaction rates, temperature increases the reaction rate according to the Arrhenius equation:

$$k = Ae^{-E_a/RT} \quad [16]$$

where  $k$  is the kinetic rate constant,  $A$  the Arrhenius constant,  $E_a$  the standard free energy of activation,  $R$  the gas constant, and  $T$  the absolute temperature. From eqn [16], the energy of activation of the enzyme-catalyzed reaction can be directly calculated:

$$\ln k = \ln A - \frac{E_a}{R} \frac{1}{T} \quad [17]$$

The increase of rate constant with temperature is limited by the inherent thermal instability of enzymes that can be categorized into two types. The first one

is time dependent and consists of gradual and irreversible loss of enzymatic activity on exposure to high temperatures. The second is heat-induced, cooperative unfolding of enzyme molecules, usually almost instantaneous and reversible (melting temperature). Time-dependent thermal inactivation is a rather complex process that may be simplified by one-stage first-order kinetics, which proposes the transition of a fully active native enzyme to a fully inactivated species in a single step. This mechanism leads to a model of exponential decay enabling the calculation of the half-life:

$$t_{1/2} = \frac{0.693}{k_d} \quad [18]$$

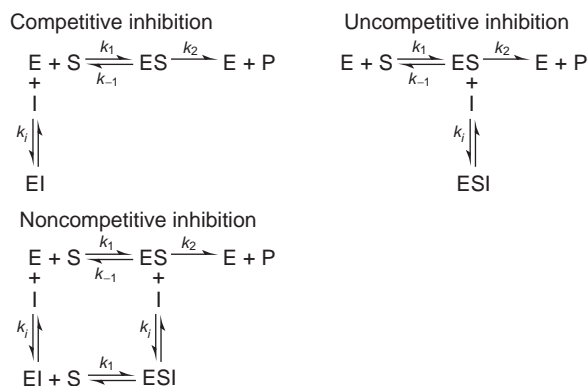
where  $k_d$  is the rate constant of enzyme inactivation. The half-life ( $t_{1/2}$ ) of an enzyme is the time it takes for the activity to reduce to half of the original activity. Although temperature may not be the only inactivating factor for an enzyme, a clear distinction should be made between operational and shelf-life (storage) stability. Both should be determined under  $V_{\max}$  conditions but the first one clearly denotes the time for an enzyme to decay in its activity under continuous operation. In order to minimize loss of activity on storage, dry conditions, concentrated enzyme solutions, or addition of additives preventing freezing (e.g., glycerol) extend the shelf-life of enzymes for extended periods of time.

## pH and Ionic Strength

Enzymes are very sensitive to pH because they are amphoteric molecules. Typically and unless other factors are taken into account, they show a rate of reaction with respect to pH that follows a bell-shaped curve. At the pH optimum, the conformation of the enzyme at that particular environment, the pattern of charges in the active site, and the ionization of substrates, products, and enzyme–substrate complex enzymes (if charged molecules) are most favorable for a maximum rate. In choosing a buffer for an analytical application, zwitterionic buffers are more suitable because they do not bind appreciably to divalent ions, very often essential for catalysis, and are chemically stable. In addition to pH, the ionic strength of the medium may substantially alter the rate of the reaction. High ionic strength will diminish water activity in addition to detrimental effects on the binding of charged substrates to enzymes and on the movement of charged groups within the active site. Thus, it is best to use buffers at the lowest concentration necessary to give adequate buffering capacity for the enzyme-catalyzed reaction.

## Other Chemical Species

Enzyme-catalyzed reactions can be strongly inhibited by the presence of other compounds. This effect sets the basis for many enzyme assays enabling, in the presence of a substrate, the determination of the concentration of the inhibitor (analyte). There are two mechanisms of inhibition: reversible and irreversible. Reversible inhibition can be further distinguished as: competitive, noncompetitive, uncompetitive, and mixed type of inhibition. A brief scheme of the most important mechanisms is depicted in Figure 8.



**Figure 8** Scheme of different enzyme inhibition mechanisms, where  $k_i$  and  $I$  denote the inhibition constant and the inhibitor, respectively.

## Traditional Enzymes in New Environments and New Enzymes in Traditional Environments

Enzymes have had billions of years to evolve as needed in living systems but none of these needs are spontaneously driven by technological or analytical challenges. Thus, the most obvious disadvantage in exploiting the exquisite specificity and rate of accelerations of enzymes is their inherent instability in addition to a rather narrow set of conditions for optimal activity. Much effort is being invested in search of: (1) different performance of classical enzymes, (2) new biocatalysts, and (3) synthetic molecules exhibiting the essence of the biocatalytic activity.

## Enzymes in Nonaqueous Solvents

Over the past 15 years, it has been firmly established that many enzymes can work in organic solvents containing little or no water. In general, hydrophobic solvents are preferred to hydrophilic ones because the latter have a greater tendency to strip tightly bound water – essential for catalytic activity – from the enzyme molecule. Although much of the interest for enzymes in organic media is driven by synthetic purposes, challenging analytical applications could be easily implemented. Examples include: online coupling of sample pretreatment and biocatalytic detection, pre- or postcolumn biocatalytic transformations within chromatographic separation, or direct biocatalytic detection of nonsoluble analytes.

In general, reaction rates of enzymes are far lower in neat organic solvents than in water. Enzyme structural changes and conformational mobility, pH constraints, and the energetics of substrate desolvation and transition state destabilization are claimed to be the major reasons for lowering the enzyme activity in organic media. These factors contribute very differently to the catalytic activity and may not be strictly ascribed to the effect of the organic media or are very dependent on the polarity of the solvent. For instance, structural changes of enzymes are rather related to the prior enzyme dehydration during lyophilization than to the direct contact with an organic solvent. Contrarily, organic solvents reduce the conformational mobility of enzymes due to the lack of multiple hydrogen bonds with the solvent and to stronger electrostatic intramolecular interactions and, thus, resulting in a much more rigid enzyme conformation with decreased catalytic activity. In fact, enzymes can only exhibit catalytic activity in organic media because they can retain some essential water in anhydrous solvents. Substrate desolvation from the reaction media to the active site is the first

step for the formation of the enzyme–substrate complex. The more energetically favorable this desolvation is, the greater the net binding energy becomes ( $\Delta G_{ES}$ , see **Figure 3**). Many enzymes have hydrophobic active centers and work best with hydrophobic substrates because they tend to partition more favorably from aqueous media into the active center. Replacement of water with an organic solvent stabilizes the ground state of hydrophobic substrate and hence slowing down the reaction rate. The energy of the transition state in organic solvents can also contribute to this detrimental reaction rate, particularly if the transition state is highly polar and is exposed to the solvent. Water will then tend to stabilize it much better than less polar solvents also resulting in lower activity in nonaqueous media. In contrast to biocatalysis in aqueous solution, the influence of pH in organic media is meaningless. Enzymes have shown a pH memory effect meaning that their catalytic behavior in organic media reflects the pH of the last aqueous solution to which they were exposed.

All these considerations should be contemplated in light of the intriguing properties that enzymes can show in nonaqueous media. For instance, enzymes are insoluble in almost all organic media providing thus a direct and simple way of immobilization. Similarly to immobilized systems, diffusional limitations may account for lower catalytic activity compared with that in homogeneous systems. The two types of thermal instabilities can be suppressed to a certain extent by organic media. In fact, thermal unfolding of some enzymes may require higher temperatures in anhydrous media than in water and, likewise, the half-lives of enzymes at high temperatures can be significantly extended. In this regard, the conformational rigidity of enzymes in anhydrous media turns now to be of great advantage. Of particular importance is the fact that classical organic reactions traditionally prohibited to biocatalytic pathways can now be easily performed and, equally important, the enzymatic selectivity can be changed dramatically on switching from one solvent to another.

### Extremozymes

Most of the enzymes used to date are obtained from mesophilic organisms and, thus, their limited stability to temperature, pH, or ionic strength. Extremophiles are organisms that have evolved to exist in a variety of extreme environments. **Table 3** lists the range of habitats where extremophiles have been found and some of the identified extremozymes. Adaptation to extreme conditions means that these enzymes have the same order of magnitude of activity and stability but at different temperatures, ionic

strengths, etc. Different factors contribute to the adaptation of extremozymes. Among them: changes in specific amino acid residues, increased number of ion pairs, reduced ratio of surface area to volume, reduction in the size of loops and in the number of cavities, increased hydrophobic interaction at subunit interfaces, changes in solvent-exposed surface areas, increase in the extent of secondary structure formation, and truncated amino and carboxyl termini.

Development of the polymerase chain reaction has been to a large extent facilitated by the availability of thermostable DNA polymerases, enzymes that catalyze the elongation of the primer DNA strand. So far, this is the largest application of thermophilic enzymes.

### Protein Engineering

By the use of genetic or/and chemical techniques, the structure and function of an enzyme can be deliberately altered resulting in new properties. In contrast to molecular biology tools, chemical modification of enzymes enables an almost unlimited introduction of chemical groups, but the reactions used for this purpose are typically nonspecific. Thus, nonspecific modifications of multiple sites create mixtures of enzymes as a result of poorly discriminating or insufficiently efficient chemistry. An example of successful chemical modification of cationic horseradish peroxidase is illustrated in **Figure 9**. By chemical binding of sulfonate derivatives to the glycosidic residues of the enzyme, its pI is shifted to yield an anionic peroxidase that shows strong electrostatic interaction with positively charged redox mediators for electrochemical communication.

Genetic modification falls into two categories: random mutagenesis and site-directed mutagenesis. The first one constitutes the basis of directed evolution (see below) whilst the second one changes – in a rational manner – a specific amino acid into another by changing the gene sequence for the protein. Thus, the technique is basically limited to the 20 amino acids – encoded by the Genetic Code – at each position of the peptide chain and it requires a precise knowledge of the relationship between sequence, structure, and mechanism. A large variety of DNA technology is nowadays available producing engineered enzymes that deeply contribute to the understanding of enzyme mechanisms.

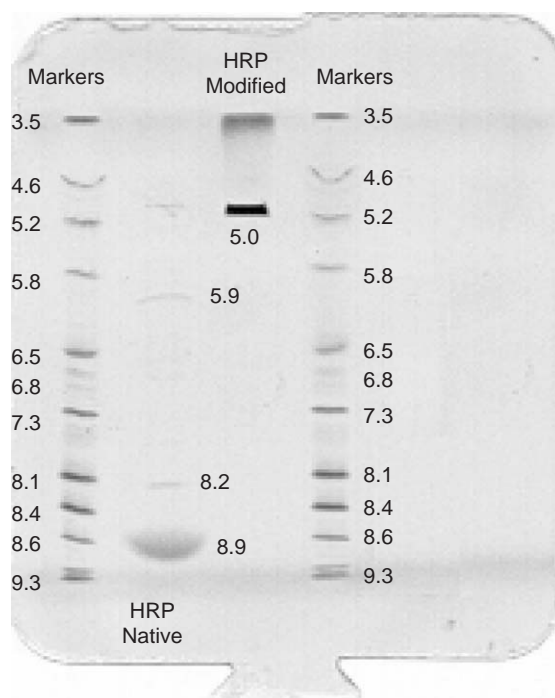
### Directed Evolution of Enzymes

This technique mimics in the laboratory the natural evolution processes of enzymes, including random mutagenesis and recombination for engineering enzymes in search of specific capabilities, especially

**Table 3** Different kinds of enzymes found in extreme conditions

Phenotype	Habitat	Enzymes	Typical genera
Thermophilic	High temperatures: Moderate thermophiles (45–65°C) Thermophiles (65–85°C) Hyperthermophiles (> 85°C)	Amylases, xylanases, proteases, DNA polymerases	<i>Methanobacterium</i> , <i>Thermoplasma</i> , <i>Thermus</i> <sup>a</sup> , some <i>Bacillus</i> <sup>a</sup> species, <i>Aquifex</i> <sup>a</sup> , <i>Archaeoglobus</i> , <i>Hydrogenobacter</i> <sup>a</sup> , <i>Methanothermus</i> , <i>Pyrococcus</i> , <i>Pyrodictium</i> , <i>Pyrolobus</i> , <i>Sulfolobus</i> , <i>Thermococcus</i> , <i>Thermoproteus</i> , <i>Thermotoga</i> <sup>a</sup>
Psychrophilic	– 2 to 20°C	Proteases, dehydrogenases, amylases	<i>Alteromonas</i> <sup>a</sup> , <i>Psychrobacter</i> <sup>a</sup>
Acidophilic	pH < 4	Sulfur oxidation, chalcopryrite concentrate	<i>Acidianus</i> , <i>Desulfurolobus</i> , <i>Sulfolobus</i> , <i>Thiobacillus</i> <sup>a</sup>
Alkalophilic	pH > 9	Cellulases	<i>Natronobacterium</i> , <i>Natronococcus</i> , some <i>Bacillus</i> <sup>a</sup> species
Halophilic	2–5 mol l <sup>–1</sup> NaCl	Malate dehydrogenase, <i>p</i> -nitrophenylphosphate phosphatase	<i>Haloarcula</i> , <i>Halobacterium</i> , <i>Haloferax</i> , <i>Halorubrum</i>
Piezophilic	High pressure	Whole microorganism	
Metalophilic	High metal concentration	Whole microorganism	
Radiophilic	High radiation levels	Whole microorganism	

<sup>a</sup>Genera of the domain Bacteria; all others are Archaea.



**Figure 9** Isoelectrofocusing of native horseradish peroxidase and chemically modified. The pI of the modified enzyme becomes anionic after conjugation with sulfonate residues. Experiments performed at the author's laboratory.

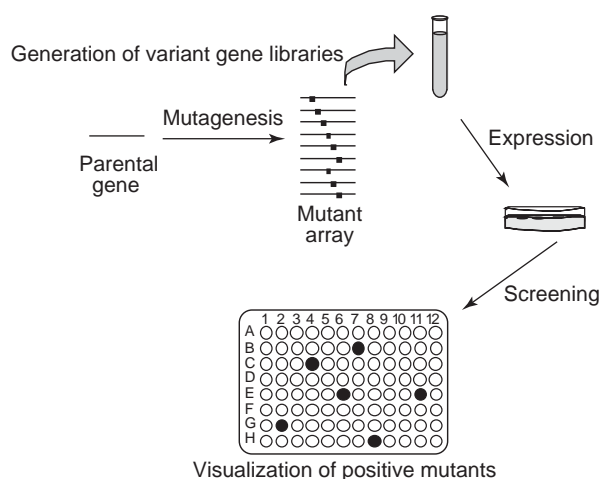
increased catalytic activity, enhanced selectivity, or improved stability. Similarly to the process by which Nature has produced enzymes without using any knowledge of structure and function, direct

evolution of enzymes generates a library of proteins. The two key steps are the generation of molecular diversity and the identification of the variants that possess specific characteristics. This is schematically illustrated in **Figure 10**. After the generation of the library of mutant genes using molecular biology methods (e.g., error-prone polymerase chain reaction), they are inserted in an appropriate microorganism. Mutant enzymes (variants) are then expressed, after being individually screened for the specific property. The mutant gene of the optimal enzyme variant is then subjected once more to the process: mutagenesis/expression/screening. This creates a pressure leading to the formation and identification of an ever better enzyme. The high throughput screen of enzyme variants is challenging and detrimental for successful process because the frequency of improved mutants might be 1 in 1000. Given that the most common enzymatic assays are not sensitive enough for single-cell analysis, screening of populations comprised of colonies is predominantly used. This technique can be viewed in parallel with the emergence of combinatorial methods in organic synthesis for lead identification and optimization of new drugs or synthetic enzymes (see below).

### Catalytic Antibodies

In common with enzymes, antibodies can specifically bind a large range of chemical structures and using the same type of molecular interactions, but, in





**Figure 10** Key steps of a typical directed enzyme evolution experiment.

contrast, the immunoglobulins bind tightly to the target molecule in their ground state and then are unable to catalyze reactions. Hence, by producing an antibody specific for the transition state, the same catalytic effect should be observed. This is the basic idea, already formulated in 1969 by Bill Jencks, behind catalytic antibodies (abzymes) that can be produced through classical immunization procedures or can also be isolated from sera of patients with several diseases. As the transition state is by definition unstable, it is necessary to use an analog of this structure to raise the abzymes. In general, catalytic antibodies display a high degree of stereospecificity although their rates of accelerations ( $10^3$ – $10^4$ ) are much smaller when compared to their enzymatic counterparts. In principle, catalytic antibodies represent a promising field for analytical applications if the vast repertoire of antibodies produced by the immune system and the possibilities of DNA technology for their production is balanced versus the number of reactions catalyzed by enzymes. So far, more than 100 reactions have been accelerated by antibodies, including reactions for which no enzyme exists.

### Nonprotein Enzymes: Catalytic RNA

Since 1982, it is recognized that nucleic acids and particularly RNA may catalyze a broad range of reactions such as formation, cleavage, and rearrangements of some chemical bonds. Catalytic RNAs are known as ribozymes and they can be found in organelles of plants and lower eukaryotes, in amphibians, in prokaryotes, in bacteriophages, and in viroids or satellite viruses. Arguably they may not

be considered true biocatalysts since they suffer a net change in the catalyzed reaction. The exception is the RNase P, which is an endoribonuclease responsible of the removal of 5' leader sequences from transfer RNA precursors. In bacteria, this RNase P consists of two subunits, a large ( $\approx 140$  kDa, 400 nucleotides) RNA and a small ( $\approx 14$  kDa, 120-amino acids) protein. *In vivo*, both subunits are strictly necessary to exhibit catalytic activity whilst *in vitro* the RNA subunit is itself catalytically active at high ionic strength and in the absence of the protein.

The use of engineered nucleic acids has allowed to convert most of the ribozymes in true biocatalyst resulting in specificity constants ( $k_{\text{cat}}/K_m$ ) up to  $10^8 \text{ l mol}^{-1} \text{ min}^{-1}$  although the acceleration rate still remains  $\sim 10^3$ -fold lower than those provided by protein enzymes catalyzing comparable reactions. All known ribozymes require a doubly charged cation, usually  $\text{Mg}^{2+}$ , which is required, in some cases, for proper assembly of a complex 3D structure.

*In vitro* selection methods permit the generation of ribozymes and deoxyribozymes (catalytic DNA) with allosteric properties, wherein the binding of an effector molecule controls the catalytic function. This feature provides challenging opportunities for the development of new analytical and very specific tools using traditional assay formats or for the development of entirely new biosensing principles.

### Synthetic Enzymes

Presently, there is great effort to produce synthetic equivalents to enzymes that being inherently stable could rival rate of accelerations and specificity. Basically, there are three different synthetic routes to mimic enzymes: (1) the design approach, (2) the transition state analog (TSA)-selection approach, and (3) the catalytic activity-selection approach. In the first one, a host molecule is modified with a functionality that being present in the natural counterpart is expected to render catalytic properties (e.g., catalytic cyclodextrins). This may demand a substantial synthetic effort for a given enzyme mimic. In contrast, the TSA-selection approach relies on the generation of a library of hosts in the presence of TSA and the best host is then selected recalling the natural abzymes generation. Catalytic molecularly imprinted polymers (MIPs) are a much younger field than its biological counterpart but the basic concepts remain very similar. Molecular imprinting is a polymerization process resulting in porous polymers with binding sites that can selectively bind the molecule with which they were imprinted. If the imprint



molecule is a TSA, the resulting MIP should then show catalytic properties. The present scope of catalytic MIPs is still relatively limited. These two methods face the common failure that the desired property of catalysis is not in the screening criteria. In contrast, the catalytic activity-selection approach takes advantage of the combinatorial chemistry revolution wherein a library of possible catalysts is generated and screened directly for enzyme-like activity.

See also: **Enzymes:** Immobilized Enzymes; Enzyme-Based Electrodes; Enzyme-Based Assays.

## Further Reading

BRENDA: The comprehensive enzyme information system. It contains the main collection of enzyme functional data available to the scientific community. It is available free of charge for academic, nonprofit users via the Internet: <http://www.brenda.uni-koeln.de>

Chaplin MF and Bucke C (1990) *Enzyme Technology*. Cambridge: Cambridge University Press. This book is out

of print, an electronic version is updated by M. Chaplin and available at: <http://www.sbu.ac.uk/biology/enztech>

Copeland RA (2000) *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*, 2nd edn. New York: Wiley-VCH.

Fersht A (1999) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. New York: WH Freeman & Co.

Hilvert D (2000) Critical analysis of antibody catalysis. *Annual Reviews in Biochemistry* 69: 751–793.

Motherwell WB, Bingham MJ, and Six Y (2001) Recent progress in the design and synthesis of artificial enzymes. *Tetrahedron* 57: 4663–4686.

Narlikar GJ and Daniel Herschlag D (1997) Mechanistic aspects of enzymatic catalysis: Lessons from comparison of RNA and protein enzymes. *Annual Reviews in Biochemistry* 66: 19–59.

Nature. Insight: Biocatalysis (2001) *Nature* 409, 225–246. A compendium of five outstanding reviews including the state of the art of biocatalysis.

Protein data bank showing 3D biological macromolecular structure data. Available at: <http://www.rcsb.org/pdb>

White JS and White DC (1997) *Source book of enzymes*. Boca Raton: CRC Press.

## Immobilized Enzymes

R M Twyman, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by G A Marko-Varga, pp. 1182–1187, © 1995, Elsevier Ltd.

## Introduction

Enzymes are exploited as catalysts in many industrial, biomedical, and analytical processes. There has been considerable interest in the development of carrier systems for enzyme immobilization because immobilized enzymes have enhanced stability compared to soluble enzymes, and can easily be separated from the reaction. This leads to significant savings in terms of reduced enzyme consumption, and the ability to use such enzymes in continuous processes. The activity and stability of enzymes depends largely on the particular operating and storage conditions, and is strongly influenced by factors such as the chemical environment, temperature, pH, and solvent properties. Most enzymes are water soluble and a certain amount of water is always required for their solubilization. Additionally, many enzymes require the presence of a cofactor, which may be attached firmly to the enzyme or may need to be added separately as a coenzyme. When the immobilization of an enzyme

is considered, the above factors need to be taken into consideration.

Immobilized enzymes are defined as biocatalysts that are restrained or localized in a microenvironment yet retain their catalytic properties. Immobilization often increases stability and makes the reuse of the enzyme preparation very simple. The repeated analysis that can be performed by the use of enzymes in an immobilized form reduces the cost of the analysis. The ideal immobilization procedure for a given enzyme is one which permits a high turnover rate of the enzyme yet also retains a high catalytic activity over time. There are two major applications for immobilized enzymes in analytical systems. In one, the enzyme is immobilized onto a particulate solid support matrix, which is then packed into a small column and incorporated into a flow system. The other involves immobilization within or on the surface of an electrode, whereby the electrochemical transduction of enzymatic product is monitored.

Many experiments have shown that the advantages of immobilized enzymes extend beyond their reusability. Their physicochemical properties are often significantly modified compared to those of the corresponding enzyme in free solution. The immobilized enzyme usually has a lower activity than the same mass of enzyme in solution. Nevertheless,

molecule is a TSA, the resulting MIP should then show catalytic properties. The present scope of catalytic MIPs is still relatively limited. These two methods face the common failure that the desired property of catalysis is not in the screening criteria. In contrast, the catalytic activity-selection approach takes advantage of the combinatorial chemistry revolution wherein a library of possible catalysts is generated and screened directly for enzyme-like activity.

See also: **Enzymes:** Immobilized Enzymes; Enzyme-Based Electrodes; Enzyme-Based Assays.

## Further Reading

BRENDA: The comprehensive enzyme information system. It contains the main collection of enzyme functional data available to the scientific community. It is available free of charge for academic, nonprofit users via the Internet: <http://www.brenda.uni-koeln.de>

Chaplin MF and Bucke C (1990) *Enzyme Technology*. Cambridge: Cambridge University Press. This book is out

of print, an electronic version is updated by M. Chaplin and available at: <http://www.sbu.ac.uk/biology/enztech>  
Copeland RA (2000) *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*, 2nd edn. New York: Wiley-VCH.

Fersht A (1999) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. New York: WH Freeman & Co.

Hilvert D (2000) Critical analysis of antibody catalysis. *Annual Reviews in Biochemistry* 69: 751–793.

Motherwell WB, Bingham MJ, and Six Y (2001) Recent progress in the design and synthesis of artificial enzymes. *Tetrahedron* 57: 4663–4686.

Narlikar GJ and Daniel Herschlag D (1997) Mechanistic aspects of enzymatic catalysis: Lessons from comparison of RNA and protein enzymes. *Annual Reviews in Biochemistry* 66: 19–59.

Nature. Insight: Biocatalysis (2001) *Nature* 409, 225–246. A compendium of five outstanding reviews including the state of the art of biocatalysis.

Protein data bank showing 3D biological macromolecular structure data. Available at: <http://www.rcsb.org/pdb>

White JS and White DC (1997) *Source book of enzymes*. Boca Raton: CRC Press.

## Immobilized Enzymes

R M Twyman, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by G A Marko-Varga, pp. 1182–1187, © 1995, Elsevier Ltd.

## Introduction

Enzymes are exploited as catalysts in many industrial, biomedical, and analytical processes. There has been considerable interest in the development of carrier systems for enzyme immobilization because immobilized enzymes have enhanced stability compared to soluble enzymes, and can easily be separated from the reaction. This leads to significant savings in terms of reduced enzyme consumption, and the ability to use such enzymes in continuous processes. The activity and stability of enzymes depends largely on the particular operating and storage conditions, and is strongly influenced by factors such as the chemical environment, temperature, pH, and solvent properties. Most enzymes are water soluble and a certain amount of water is always required for their solubilization. Additionally, many enzymes require the presence of a cofactor, which may be attached firmly to the enzyme or may need to be added separately as a coenzyme. When the immobilization of an enzyme

is considered, the above factors need to be taken into consideration.

Immobilized enzymes are defined as biocatalysts that are restrained or localized in a microenvironment yet retain their catalytic properties. Immobilization often increases stability and makes the reuse of the enzyme preparation very simple. The repeated analysis that can be performed by the use of enzymes in an immobilized form reduces the cost of the analysis. The ideal immobilization procedure for a given enzyme is one which permits a high turnover rate of the enzyme yet also retains a high catalytic activity over time. There are two major applications for immobilized enzymes in analytical systems. In one, the enzyme is immobilized onto a particulate solid support matrix, which is then packed into a small column and incorporated into a flow system. The other involves immobilization within or on the surface of an electrode, whereby the electrochemical transduction of enzymatic product is monitored.

Many experiments have shown that the advantages of immobilized enzymes extend beyond their reusability. Their physicochemical properties are often significantly modified compared to those of the corresponding enzyme in free solution. The immobilized enzyme usually has a lower activity than the same mass of enzyme in solution. Nevertheless,

unless the activity losses are severe this disadvantage is not serious, and can be compensated by using more of the immobilized preparation, knowing that it can be used repeatedly.

A significant advantage of immobilization is the increased thermal stability conferred on many enzymes, allowing their use for longer periods at higher temperatures than would be possible for the soluble molecule. An early experiment showed that immobilized papain (a protease derived from papaya latex) retained over half its room-temperature activity at 80°C in conditions under which the soluble enzyme was almost entirely denatured.

Also interesting and valuable are the changes in the pH-activity curves that often accompany immobilization. A common result is broadening of the curve, i.e., the enzyme is active over a wider pH range than its soluble counterpart. This is normally attributed to the range of microenvironments of different enzyme molecules in or on the solid matrix. Depending on the charge properties of this matrix, the optimum pH may undergo significant shifts. The optimum pH for an enzyme bound to a negatively charged carrier such as carboxymethylcellulose is shifted to higher values, while immobilization on a cationic matrix such as diethylaminoethyl-cellulose (DEAE-cellulose) has the opposite effect. These effects may reflect the change in the enzyme's microenvironment brought about by neighboring charged groups. Immobilization on a neutral carrier would not be expected to change the pH optimum.

Despite the advantages of enzymes immobilized on noncatalytic matrices, the yield and productivity of the reaction can be reduced simply due to the presence of the noncatalytic mass of the carrier. There has therefore been much interest in the development of carrier-free systems, in which enzyme molecules are linked to each other to form large complexes. These are inherently immobilized because individual enzyme molecules are no longer free to diffuse in solution, but they are largely undiluted by inert molecules and therefore retain a greater degree of activity than carrier-bound enzymes. This article discusses strategies for enzyme immobilization in carrier-based and carrier-free systems and considers some of their major applications.

## Immobilized Enzyme Reactors

Enzymes may be immobilized onto a number of different types of matrix, including beads, fibers, films, membranes, and even enclosed capsules (Figure 1A). The means of association may be covalent or non-covalent. There is no general rule as to the optimum

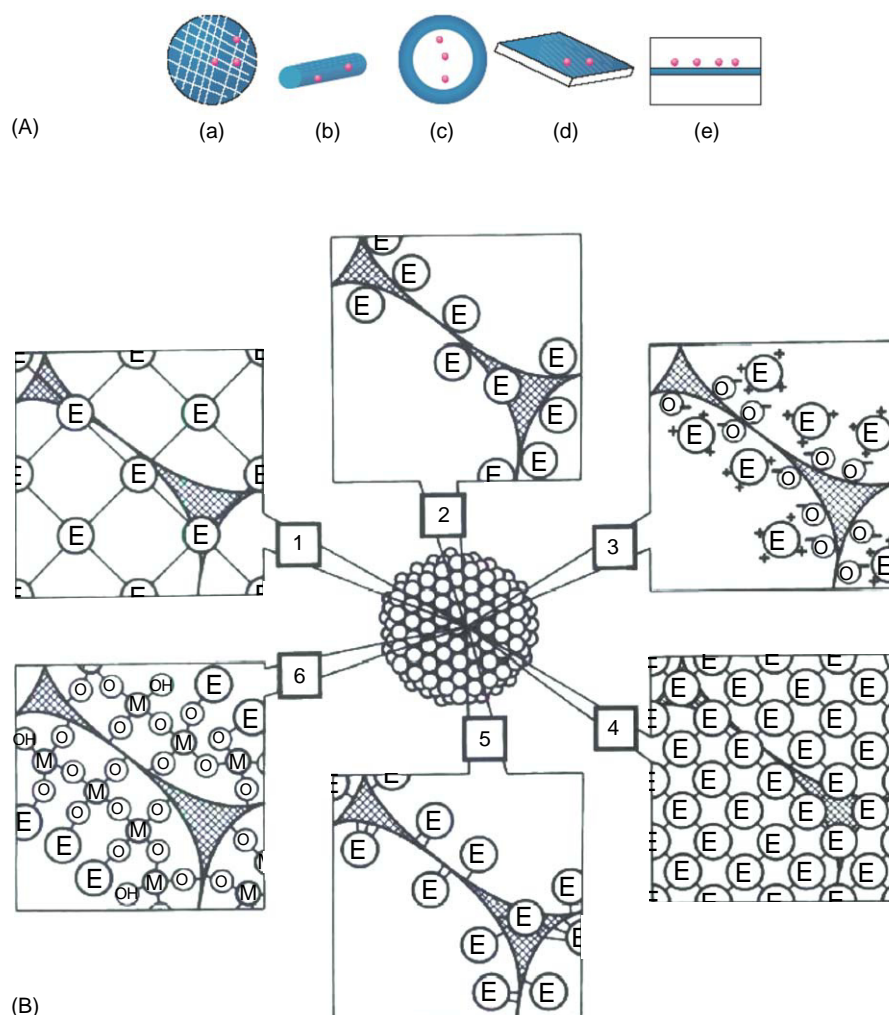
immobilization procedure and an empirical approach is always required. Various alternative coupling reactions have been used, the most common of which are adsorption, covalent binding, electrostatic binding, intermolecular cross-linking, gel entrapment, and chelation (Figure 1B).

Immobilization of enzymes by means of covalent coupling to a reactive insoluble support or to the surface of an electrode is the method of choice if a long operational performance is required. The covalent binding should involve functional groups on the enzyme that are not required for catalytic activity and do not lie near the active site. In general, binding is achieved by a nucleophilic group on the enzyme reacting with an activated functional group on the support material. Amino, hydroxyl, and thiol groups on the enzyme may participate in the linkage. Cysteine, lysine, tyrosine, and histidine residues are most commonly involved in such coupling reactions.

The matrices most widely used for enzyme immobilization are made from controlled-pore glass. This is a macroporous, high-silica glass commonly used in both flow injection and liquid chromatography systems. The support comprises particles or beads (typically ~10 µm diameter) with bonded functionalities, such as C<sub>18</sub> (octadecyl), alkylamino-bonded (–NH<sub>2</sub>), or zirconium-treated silica matrices. Polymeric supports are also commonly used as carriers. Some typical polymers include vinylsulfone-hydroxymethyl methacrylate, poly(vinylalcohol) beads, polyacrylamide gels, 2-fluoro-1-methylpyridinium salt-activated Fractogel, hexamethylacrylate polymer (Sepharon Herma E, an anion exchange polymer), chlormethylated polystyrene beads, and beaded cellulose. Polysaccharide stationary phases like agarose have also been used as enzyme supports, e.g., cyanogen bromide-activated Sepharose 4B.

Some of these enzyme supports are available from the manufacturers as dry or ready-to-use preparations or can be purchased in the form of prepacked columns. Generally, these phases are already activated and immobilization is achieved by recirculating the enzyme solution for a given time at a pH appropriate for the binding reaction. The activation and immobilization conditions are generally provided by the suppliers.

Matrix functionalities that are often used in covalent binding reactions for enzyme immobilization include aminoalkyl, arylamine, 3-glycidopropyl, carboxyl, 3-mercaptopropyl and alkylhalide groups. The most commonly used immobilization techniques are illustrated in Figure 2. Alkylamine- and arylamine-functionalized supports are shown with glutaraldehyde and azo coupling, respectively. For silica or glass supports, the functionalities are



**Figure 1** (A) Carrier-bound immobilized enzymes of defined size and shape. Insoluble carriers vary in their geometric parameters. Different shapes and types of enzyme carrier are illustrated: (a) bead, (b) fiber, (c) capsule, (d) film, and (e) membrane. (B) Methods used for immobilizing enzymes onto a spherical solid support matrix: 1, physical absorption; 2, covalent binding; 3, electrostatic binding; 4, intermolecular cross-linking; 5, gel entrapment; 6, chelation and/or metal binding. E, enzyme; M, metal.

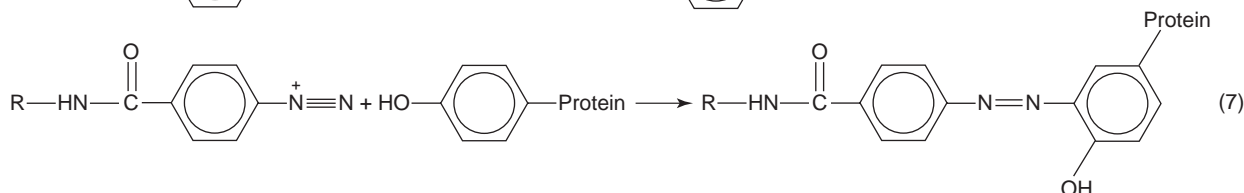
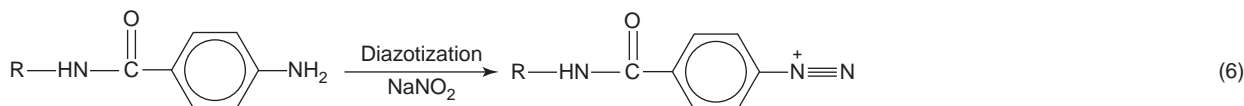
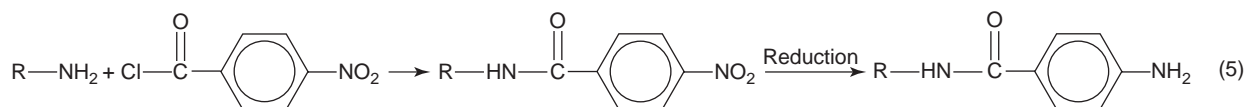
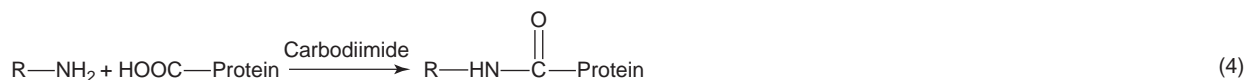
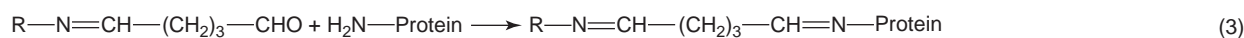
usually obtained through silanization of the support with a silane derivative. The best known covalent linkage methods include the use of cyanogen bromide (CNBr) or chloro-*sym*-triazinyl groups to bind  $\text{-NH}_2$  groups to hydroxylic matrices such as agarose or cellulose, and the use of carbodiimides to link amine and carboxylic acid functional groups between enzymes and solid phases. Such covalent linkages are not infinitely stable, and even the best immobilized enzyme reactors run down after several months through slow leakage of the enzyme into solution.

Controlled-pore glass with aminopropyl functional groups has proven to be the most successful support for enzyme immobilization. Most enzymes immobilized on this support, e.g., with glutaraldehyde, retain their activities. The goal is to achieve maximal loading of the active enzyme, resulting in optimal

conversion efficiency per unit volume of the non-catalytic matrix.

Activation of the functionalized support is most often achieved with glutaraldehyde, which introduces a carbonyl residue that subsequently reacts with nucleophilic groups on the enzyme. The activation of the functional material and the binding of the enzyme result in the formation of a Schiff base. The optimum pH for these reactions is a compromise between the preferred acidic catalytic conditions, the  $\text{pK}_a$  of the nucleophilic groups (in such a way that the unprotonated form is desirable) and the stability of the enzyme at different pH values. The reversibility of Schiff base formation is suppressed by the addition of sodium cyanotrihydroborate, a weak reducing agent, which does not reduce aldehydes or ketones at neutral pH but readily reduces the imine form to the





**Figure 2** Schemes of the most commonly used immobilization techniques. (1) Coupling of aminopropyl functionality to a silica-based support, (2) Glutaraldehyde coupling, activation of amine-modified surface with glutaraldehyde, (3) Immobilization of enzyme on glutaraldehyde-activated support, (4) Carbodiimide coupling, (5) Arylamine support prepared by *p*-nitrobenzoyl chloride with alkylamine-modified support followed by reduction (6) and (7) Enzyme immobilization by azo coupling through the tyrosine moiety to the arylamine after diazotization.

amine form. Another important aspect of the addition of cyanotrihydroborate is that the imine product itself is a reactive species and is prone to further reaction, resulting in the addition of new nucleophilic agents to the imine bond. If water is the nucleophile, the transimination reaction may result in unsuccessful activation of the functionalized material.

There is also a trend toward the utilization of polymer-based supports. These are known to be more stable under alkaline conditions, resulting in reduced band broadening effects compared with silica and glass phases. These effects are most clearly seen in flow systems (flow injection and liquid chromatography) in the determination of solutes that are catalyzed at a higher pH.

Successful enzyme immobilization depends on the structure of the porous support. Small pores provide a large total surface area because of their larger internal surface, while larger pores provide a smaller surface area. The efficiency of the immobilized enzyme can be improved by chromatographic purification of the crude enzyme. The amount (mg) per unit of activity (U) of enzyme can easily be increased, even with the purified preparations that are commercially available.

To achieve these goals, the choice of enzyme support and the binding chemistry should be optimized. Small particle diameter and small pore sizes will result in large specific surface areas, which are important for high enzyme loading and low additional

band-broadening. All these factors have to be optimized empirically since they are difficult to predict and determine theoretically.

## Amperometric Biosensors

A general definition of a biosensor is a system of two transducers, biochemical and physical, in intimate contact, where there is a stoichiometric relation between the concentration of the analyte and the electrical signal. The biochemical transducer converts the analyte into a chemical or physical property, which is sensed and converted into an electrical signal by the physical transducer. Enzymatic redox reactions are particularly closely related to electrochemical transducers since an electron transfer step is involved in the natural cycle of the redox enzyme. The coupling is, however, far from straightforward. All redox enzymes depend on a cofactor as the natural partner in the exchange process.

The physical combination of immobilized enzymes and electrodes can be realized in several different ways. The most common are those in which the enzyme is immobilized in a membrane that is held in close proximity to the electrode surface, immobilized directly on the electrode surface with retained activity, or immobilized in an electrically conducting paste comprising a carbon-oil mixture, known as a carbon paste electrode. The enzyme substrate (and cosubstrate, if required) will diffuse from the

solution to the electrode surface, reach the enzyme, and undergo the appropriate reaction. The reaction product will then diffuse to the electrode surface, where the actual analytical reaction will take place. A membrane is often deposited onto the electrode surface. This membrane can be used as a matrix, the material support for a proper immobilization. It can also be used as a barrier to protect the enzyme from interfering compounds present in the sample. The selectivity of the membrane can in the latter case be controlled by electrostatic shielding, using charged membranes and/or by a size exclusion effect controlled by the pores in the membrane. Optimal membranes should adhere well to the surface of the electrode, and they should be thin, hydrophilic, and porous. Cellulose acetate, cellulose nitrate, poly(vinyl alcohol), and polyurethane are examples of commonly used enzyme membranes.

## Immobilized Enzymes in Analytical Flow Systems

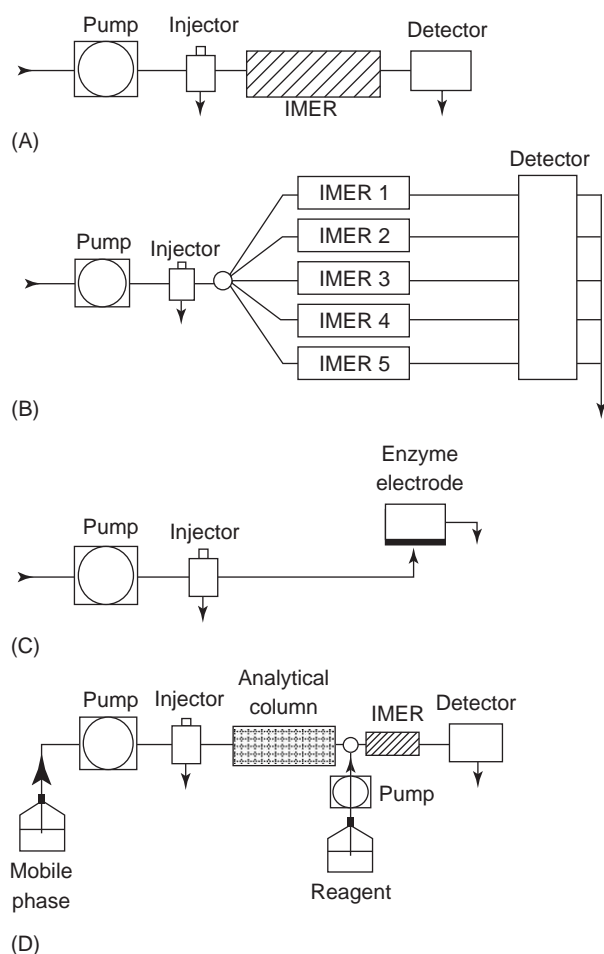
The amount of enzyme that can be immobilized and contained in a reactor of moderate size (5–500  $\mu$ l) is very large (up to 1000 U). Thus, when a solution containing the enzyme substrate passes through the reactor, equilibrium of the enzymatic process can be reached. In many cases, the equilibrium strongly

**Table 1** Use of immobilized enzymes in analytical flow systems

Analyte(s)	Enzyme(s) <sup>a</sup>	Flow techniques <sup>b</sup>
Acetaldehyde	Aldehyde DH	FI, LC
Acetylcholine/choline	Acetylcholine OX Choline ES	FI, LC
Aliphatic alcohols	Alcohol OX, alcohol DH	FI, LC
Amino acids (L and D)	L-Amino acid OX D-Amino acid OX	FI, LC
Ascorbate	Ascorbate OX	FI
Bile acids	Hydroxysteroid DH	FI, LC
Fructose	Fructose DH Hexokinase/glucose-6-phosphate DH	FI
Cholesterol and derivatives	Cholesterol OX	FI, LC
Creatinine	Creatinine iminohydrolase, creatininase	FI
Glucosides	$\beta$ -Glucuronidase	LC
Glucose	Glucose OX Glucose DH	FI, LC
Inorganic/organic peroxides	Peroxidase	FI
Inositol phosphates	Alkaline phosphatase	LC
Lactic acid	Lactate OX	FI
L-Leucine	L-Leucine DH	FI
Malto-oligosaccharides	Malto-oligosaccharide DH	FI, LC
L-Malate	L-Malate DH	FI
Penicillins	Penicillinase, $\beta$ -lactamase	FI
Peptides	Peptidases	FI
Phenol and derivatives	Tyrosinase, laccase	FI, LC
Phenolic glycosides	$\beta$ -Glucuronidase	FI, LC
Pyruvate, lactate	Lactate OX	FI, LC
Starch	Amyloglucosidase/glucose OX Amyloglucosidase/glucose DH	FI, LC
Sucrose	Invertase/mutarotase/glucose OX	FI
Urea	Urease	FI
Xanthine, hypoxanthine	Xanthine OX	FI
D-Xylitol		
D-Sorbitol	Sorbitol DH	LC
D-Ribitol		

<sup>a</sup>Enzymes: OX, oxidase; DH, dehydrogenase; ES, esterase.

<sup>b</sup>Techniques: FI, flow injection; LC, liquid chromatography.



**Figure 3** Flow injection and liquid chromatographic flow configurations using immobilized enzymes: (A) single immobilized enzyme reactor (IMER) in a flow injection system; (B) multiple immobilized enzyme reactors in a flow injection system; (C) system with enzyme electrode; and (D) postcolumn liquid chromatography system with reagent flow and immobilized enzyme reactor.



favors the product side (chemically irreversible reaction), which means that virtually 100% of the substrate can be converted into a detectable product. This is never the case with an enzyme electrode, where the amount of immobilized enzyme in the membrane or on the electrode surface is more restricted. It will usually be too low ( $1\text{--}3\text{ U cm}^{-2}$ ) to achieve equilibrium of the catalyzed reaction within a reasonable time, particularly when used in a flowing solution. With a reactor containing excess enzyme, slight variations in the flow rate, pH, ionic strength or temperature, and the presence of small concentrations of inhibitors and activators will not affect the conversion efficiency of the reactor.

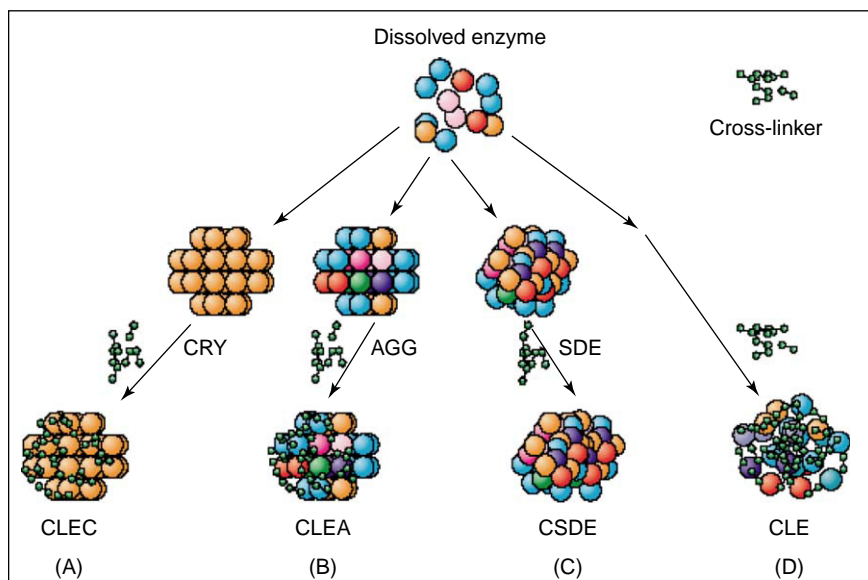
Nevertheless, both immobilized enzyme reactors and biosensors are used as detection units in flow injection analysis and liquid chromatography. Some of these flow systems are illustrated in Figure 3. The choice of flow system will be determined both by the number of analytes to be measured and by the complexity of the sample. Flow injection in combination with immobilized enzymes is used for single solute determinations, as in the systems shown in Figure 3A–C. Multiple solute determination requires a separation step whereby the chromatography column is introduced (Figure 3D). There is also an additional need for separation power when interfering matrix components in complex samples need to be eliminated in order to permit accurate quantitation of the analyte. The use of multiple flow lines with immobilized enzyme reactions in flow injection systems has been demonstrated, whereby each flow line measures a

single analyte in the sample. This approach has an efficiency similar to that obtained using chromatographic separation.

Immobilized enzymes are used to detect many different solutes in analytical flow systems (Table 1). Among the most common targets are glucose (detected with immobilized glucose oxidase), acetylcholine and choline (acetylcholine/choline oxidase), lactate (lactate oxidase or dehydrogenase), and pyruvate (pyruvate oxidase). The types of applications range from those involving relatively clean matrices to those involving complex samples such as clinical tests (e.g., blood, urine), food, feed, water, and environmental specimens.

## Carrier-Free Systems

The use of a carrier or polymeric matrix to immobilize enzymes introduces a large noncatalytic component into the system which has the potential to interfere with the catalytic properties of the enzyme and reduce its activity compared to the same mass of free enzyme. Although this disadvantage is balanced by the reusability of immobilized enzymes, it would be even more beneficial if the activities of immobilized enzymes could match those of free enzymes. Carrier-free systems, in which enzyme molecules are linked to each other to form large complexes, may provide a solution to this problem. In a carrier-dependent system, up to 99.9% of the mass is taken up by the noncatalytic matrix. In noncarrier systems, 100% of the complex has the potential to retain catalytic activity.



**Figure 4** The different approaches to the production of carrier-free immobilized enzymes; (A) crystallization; (B) aggregation; (C) spray-drying; and (D) direct cross-linking. AGG, aggregates; CLE, cross-linked dissolved enzyme; CLEA, cross-linked enzyme aggregate; CLEC, cross-linked enzyme crystal; CSDE, cross-linked spray-dried enzyme; CRY, crystals; SDE, spray-dried enzyme.

The simplest way to achieve carrier-free immobilization is to cross-link enzymes in solution. However, the success of this process is highly dependent on the ambient conditions (temperature, pH, etc., which must be established empirically) and often results in an unacceptable loss of activity. Furthermore, the gelatinous complexes formed in this manner can be difficult to handle. Some degree of control may be introduced by cross-linking the enzymes within a gel matrix or following adsorption to a membrane, but this reintroduces a carrier into the system, and hence increases the noncatalytic mass and reduces the efficiency. Alternatives to the in-solution cross-linking strategy include the prior formation of enzyme crystals or aggregates, which are much more stable, easier to handle, and retain a higher level of activity. Several cross-linked crystals are used in commercial catalysts, mainly for stereospecific synthesis reactions. Enzymes may also be spray dried prior to cross-linking, but such complexes tend to show a relatively low activity compared to crystals and aggregates. The methods for preparing carrier-free immobilized complexes are shown in Figure 4.

See also: **Enzymes:** Overview; Enzyme-Based Electrodes; Enzymes in Physiological Samples; Industrial Products and Processes; Enzyme-Based Assays.

## Further Reading

Bickerstaff G (ed.) (1996) *Immobilization of Enzymes and Cells (Methods in Molecular Biotechnology)*. New York: Humana Press.

- Cabral JMS and Kennedy JF (1993) Immobilisation techniques for altering thermal stability of enzymes. In: Gupta MN (ed.) *Thermostability of Enzymes*, pp. 163–179. Berlin: Springer.
- Cao LQ, van Langen L, and Sheldon RA (2003) Immobilised enzymes: Carrier-bound or carrier-free? *Current Opinion in Biotechnology* 14: 387–394.
- Clark DS (1994) Can immobilisation be exploited to modify enzyme activity? *Trends in Biotechnology* 12: 439–443.
- Hartmeier W (1998) *Immobilised Biocatalysts*. Berlin: Springer.
- Katchalskikatzir E (1993) Immobilized enzymes – learning from past successes and failures. *Trends in Biotechnology* 11: 471–478.
- Lam S and Malikin G (eds.) (1994) *Analytical Applications of Immobilized Enzyme Reactors*. Dordrecht: Kluwer Academic Publishers.
- Liang JF, Li YT, and Yang VC (2000) Biomedical application of immobilized enzymes. *Journal of Pharmaceutical Sciences* 89: 979–990.
- Messing RA (1975) *Immobilized Enzymes for Industrial Reactors*. London: Academic Press.
- Mosbach K (ed.) (1987) *Immobilised Enzymes and Cells (Part B) (Methods in Enzymology)*. London: Academic Press.
- Svec F and Gemeiner P (1996) Engineering aspects of carriers for immobilized biocatalysts. *Biotechnology and Genetic Engineering* 13: 217–235.
- Tischer W and Kasche V (1999) Immobilized enzymes: crystals or carriers? *Trends in Biotechnology* 17: 326–335.
- Tischer W and Wedekind F (1999) Immobilized enzymes: methods and applications. *Topics in Current Chemistry* 200: 95–126.
- Zarborsky O (1973) *Immobilised Enzymes*. Cleveland: CRC Press.

## Enzyme-Based Electrodes

**F W Scheller**, Universitaet Potsdam, Golm, Germany

**D Pfeiffer**, BST Bio Sensor Technologie GmbH, Berlin, Germany

**F Schubert**, Physikalisch-Technische Bundesanstalt, Berlin, Germany

**U Wollenberger**, Universitaet Potsdam, Golm, Germany

© 2005, Published by Elsevier Ltd.

## Introduction

Electrochemical sensors are appropriate for the determination of different gases (e.g., oxygen, hydrogen, and carbon dioxide), nitrogen compounds (e.g., NO<sub>x</sub> and NH<sub>3</sub>), ion activities (alkali and heavy metals), and oxidizable and reducible organic substances down to the submicromolar concentration

range. The determination of many other important substances requires the coupling of sensors with a chemical or biochemical reaction to generate an electroactive product.

A problem of paramount importance to analytical chemistry is selectivity, particular at trace concentrations where potential interferents might be present at higher concentrations than the analyte. Enzymes allow the determination of trace concentrations even in complex media due to their excellent chemical selectivity, thus avoiding the need to use highly sophisticated instrumental techniques such as chromatography or mass spectrometry. Furthermore, when enzymes are employed as tracers in binding assays using antibodies, binding proteins, lectins, etc., the inherent chemical amplification properties of the enzyme's catalytic activity can be fully exploited

The simplest way to achieve carrier-free immobilization is to cross-link enzymes in solution. However, the success of this process is highly dependent on the ambient conditions (temperature, pH, etc., which must be established empirically) and often results in an unacceptable loss of activity. Furthermore, the gelatinous complexes formed in this manner can be difficult to handle. Some degree of control may be introduced by cross-linking the enzymes within a gel matrix or following adsorption to a membrane, but this reintroduces a carrier into the system, and hence increases the noncatalytic mass and reduces the efficiency. Alternatives to the in-solution cross-linking strategy include the prior formation of enzyme crystals or aggregates, which are much more stable, easier to handle, and retain a higher level of activity. Several cross-linked crystals are used in commercial catalysts, mainly for stereospecific synthesis reactions. Enzymes may also be spray dried prior to cross-linking, but such complexes tend to show a relatively low activity compared to crystals and aggregates. The methods for preparing carrier-free immobilized complexes are shown in Figure 4.

See also: **Enzymes:** Overview; Enzyme-Based Electrodes; Enzymes in Physiological Samples; Industrial Products and Processes; Enzyme-Based Assays.

## Further Reading

Bickerstaff G (ed.) (1996) *Immobilization of Enzymes and Cells (Methods in Molecular Biotechnology)*. New York: Humana Press.

- Cabral JMS and Kennedy JF (1993) Immobilisation techniques for altering thermal stability of enzymes. In: Gupta MN (ed.) *Thermostability of Enzymes*, pp. 163–179. Berlin: Springer.
- Cao LQ, van Langen L, and Sheldon RA (2003) Immobilised enzymes: Carrier-bound or carrier-free? *Current Opinion in Biotechnology* 14: 387–394.
- Clark DS (1994) Can immobilisation be exploited to modify enzyme activity? *Trends in Biotechnology* 12: 439–443.
- Hartmeier W (1998) *Immobilised Biocatalysts*. Berlin: Springer.
- Katchalskikatzir E (1993) Immobilized enzymes – learning from past successes and failures. *Trends in Biotechnology* 11: 471–478.
- Lam S and Malikin G (eds.) (1994) *Analytical Applications of Immobilized Enzyme Reactors*. Dordrecht: Kluwer Academic Publishers.
- Liang JF, Li YT, and Yang VC (2000) Biomedical application of immobilized enzymes. *Journal of Pharmaceutical Sciences* 89: 979–990.
- Messing RA (1975) *Immobilized Enzymes for Industrial Reactors*. London: Academic Press.
- Mosbach K (ed.) (1987) *Immobilised Enzymes and Cells (Part B) (Methods in Enzymology)*. London: Academic Press.
- Svec F and Gemeiner P (1996) Engineering aspects of carriers for immobilized biocatalysts. *Biotechnology and Genetic Engineering* 13: 217–235.
- Tischer W and Kasche V (1999) Immobilized enzymes: crystals or carriers? *Trends in Biotechnology* 17: 326–335.
- Tischer W and Wedekind F (1999) Immobilized enzymes: methods and applications. *Topics in Current Chemistry* 200: 95–126.
- Zarborsky O (1973) *Immobilised Enzymes*. Cleveland: CRC Press.

## Enzyme-Based Electrodes

**F W Scheller**, Universitaet Potsdam, Golm, Germany

**D Pfeiffer**, BST Bio Sensor Technologie GmbH, Berlin, Germany

**F Schubert**, Physikalisch-Technische Bundesanstalt, Berlin, Germany

**U Wollenberger**, Universitaet Potsdam, Golm, Germany

© 2005, Published by Elsevier Ltd.

## Introduction

Electrochemical sensors are appropriate for the determination of different gases (e.g., oxygen, hydrogen, and carbon dioxide), nitrogen compounds (e.g., NO<sub>x</sub> and NH<sub>3</sub>), ion activities (alkali and heavy metals), and oxidizable and reducible organic substances down to the submicromolar concentration

range. The determination of many other important substances requires the coupling of sensors with a chemical or biochemical reaction to generate an electroactive product.

A problem of paramount importance to analytical chemistry is selectivity, particular at trace concentrations where potential interferents might be present at higher concentrations than the analyte. Enzymes allow the determination of trace concentrations even in complex media due to their excellent chemical selectivity, thus avoiding the need to use highly sophisticated instrumental techniques such as chromatography or mass spectrometry. Furthermore, when enzymes are employed as tracers in binding assays using antibodies, binding proteins, lectins, etc., the inherent chemical amplification properties of the enzyme's catalytic activity can be fully exploited

to realize extremely sensitive assay methods. In addition, biologically related parameters such as taste, odor, fatigue substances, mutagenicity, and nutrient content are quantifiable by using multienzyme systems, intact organelles, or cells.

## Development of Enzymatic Measurements

Enzymes accelerate the attainment of equilibrium of chemical reactions by a factor of  $10^6$ – $10^{12}$  as compared with uncatalyzed reactions. Several enzymes, e.g., catalase, acetylcholinesterase, and fumarase, have achieved virtual catalytic perfection approaching the diffusion-controlled limit. Thus, the splitting of  $H_2O_2$  is accelerated by a factor of  $3 \times 10^{11}$  in the presence of catalase.

Traditionally, enzymes have been used as analytical reagents to measure substrate molecules by catalyzing the turnover of these species to detectable products. In addition, groups modifying the rate of the enzyme reaction – activators, prosthetic groups, inhibitors, and also enzymes themselves – are accessible to measurement.

To simplify early enzymatic measurements the principle of litmus paper was adopted – a filter paper was impregnated with the analyte-converting enzymes and coreactants. Enzyme ‘test strips’ for the determination of more than 15 different metabolites and enzymes are commercially available. In the early 1960s, Clark introduced a new sensor concept of the ‘enzyme electrode’. He placed an enzyme solution (an oxidase) immediately in front of an oxygen electrode and prevented loss of the enzyme by dispersion into the background solution by covering the reaction chamber with a semipermeable membrane. The same enzyme preparation could then be used for several samples. The next stage was reached in 1967 by Updike and Hicks, who entrapped the enzyme in a gel, thus increasing the working stability and simplifying the sensor preparation. In 1975, Yellow Springs Instruments Co. (USA) commercialized a glucose analyzer that was based on Clark’s invention.

The choice of the ‘indicator electrode’ is largely determined by the species involved in the sensing reaction. Oxygen and  $H_2O_2$ , which are the cosubstrate and product of oxidases, as well as NAD(P)H, the cosubstrate of  $\sim 300$  pyridine nucleotide-dependent dehydrogenases, can be determined amperometrically. Hydrolases are mostly coupled to ion-selective electrodes. Based on these principles, many enzyme sensors have been developed and commercialized (Table 1).

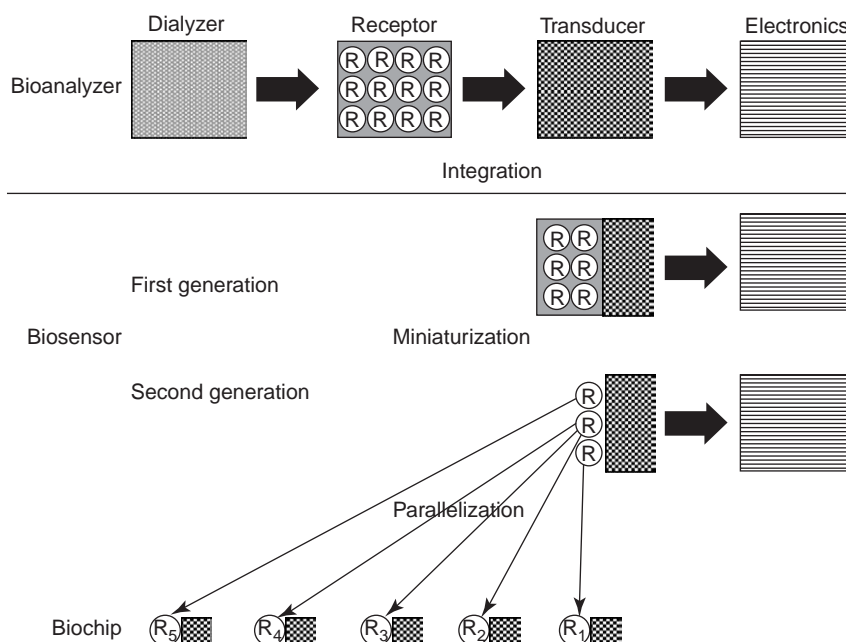
**Table 1** Coupling of enzymes with electrochemical sensors in enzyme electrodes

<i>Biocomponents</i>	<i>Indicated species</i>	<i>Electrode type</i>
<i>Oxidoreductases</i>	<i>Cosubstrates</i>	<i>Amperometric</i>
Dehydrogenases	NAD(P)H	
Oxidases	$O_2/H_2$	
Peroxidases	Mediators	
Electron-transferases		
	<i>Products</i>	
	Phenols	<i>Potentiometric</i>
	Redox dyes	
	<i>Prosthetic groups</i>	
	Heme	
	PQQ	
	FAD	
	$Cu^{2+}$	
<i>Hydrolases</i>	<i>Products</i>	
Proteases	$H^+$	
Esterases	$HCO_3^-$	
Glycosidases	$NH_4^+$	
	$I^-$	
	$F^-$	

IUPAC proposed a definition of a biosensor in 2001: “An electrochemical biosensor is a self-contained integrated device which is capable of providing specific quantitative or semi-quantitative analytical information using a biological recognition element (biochemical receptor) which is in direct spatial contact with a transduction element. A biosensor should be clearly distinguished from a bioanalytical system which requires additional processing steps, such as reagent addition. A device that is both disposable after one measurement, i.e., single use, and unable to monitor the analyte concentration continuously or after rapid and reproducible regeneration, should be designated a single use biosensor.”

According to their level of integration, the enzyme electrodes described in the literature can be subdivided into three generations. In the simplest approach (first generation) the biocatalyst is entrapped between or bound to membranes and this arrangement is fixed at the surface of the transducer. The immediate adsorptive or covalent binding of the biocatalyst at the electrode surface or the inclusion within the electrode body permits the elimination of semipermeable membranes (second generation). The direct binding of the biocatalyst to the transducer is the basis for further miniaturization. Independently addressable microelectrodes or ion-selective field-effect transistors (ISFETs) are the transducer units of these electrochemical biochips that allow for the ‘parallel’ determination of a broad spectrum of analytes (Figure 1).





**Figure 1** Generations of biosensor technology.

A typical example of the first generation would be a glucose or lactate electrode comprising the appropriate oxidase entrapped in or bound to a membrane that is fixed at an oxygen or hydrogen peroxide detecting electrode and covered by a semipermeable membrane.

The 'membrane' directed toward the (usually ideally mixed or flowing) measuring solution fulfills a number of functions. First, it provides the sensor with a certain degree of selectivity. The pore size and, perhaps, charge permits the exclusion of deleterious or interfering molecules, such as proteins or electroactive compounds and may provide useful partitioning of other compounds. Furthermore, the thickness and pore size permit us to affect the measuring range of the sensor by controlling the actual rate of reagents reaching the reaction layer.

After permeation of analyte, cosubstrates, and effectors through the membrane to the underlying enzyme layer the analyte is converted therein with formation or consumption of a detectable species. In the illustrated case this is the formation of a reduced mediator or cosubstrate (e.g., hydrogen peroxide from oxygen) that is oxidized at the electrode or the generation of protons. The enzyme membrane is characterized by 'enzyme loading' reflecting the interplay of enzyme kinetics and mass transport. The loading is crucial for the response characteristics and the stability of the sensor whereas the choice of enzyme determines the chemical selectivity of the measurement.

## Combination of Electrodes with Immobilized Enzymes

### Immobilization of Enzymes on Electrodes

To allow the repeated use of enzymes, cells, and other biologically active agents in analytical devices, numerous techniques have been developed for their fixation. In many cases the enzyme is stabilized by immobilization and so may easily be separated from the sample. Its stable and largely constant activity makes the enzyme an integral part of the analytical device. Methods to immobilize enzymes entail physical and chemical techniques and a combination thereof. The primary physical methods are adsorption direct to the electrode and entrapment in water-insoluble polymeric gel. Chemical immobilization is effected by covalent coupling to reactive groups at the transducer surface or by intermolecular cross-linking of the biomolecule using bifunctional reagents.

Structured deposition of enzymes on the surface of microelectrodes or ISFETs is a major problem during mass production. Layers of polymers, e.g., polypyrrole or polyaniline, are deposited on conducting areas by electropolymerization. Biomolecules are either entrapped in the polymer matrix while the layer is being formed or are coupled to the layer via specific chemical reactions. The structuring of a uniform layer containing a biocomponent is possible when passive regions are photodeactivated. Alternatively,

enzymes are deposited only at the sensitive region by using an enzyme solution in a photoresist.

Using these microspotting techniques the active areas of arrays are covered with different receptor molecules or cells.

### Transfer of the Chemical Signal from the Enzyme to the Electrode

The high selectivity of enzyme electrodes is well recognized in the literature. Prerequisites for the selectivity are (1) the high selectivity of the biocatalyst and (2) interference-free electrochemical indication of the reaction product. With regard to selectivity, two groups of enzymes are available in nature as analytical reagents: (1) group-specific enzymes converting a family of analogous substances and (2) enzymes possessing high chemical selectivity that allows them to discriminate even between stereoisomeric compounds.

Amperometric enzyme electrodes are at the leading edge of biosensors as far as the body of scientific publications as well as the development of the commercially available devices is concerned. Only a few conductimetric enzyme electrodes and EN(zyme)-FETs have been described.

Ion-sensitive electrodes based on ion exchangers or neutral carriers (e.g., nonactin, valinomycin) have been coupled with ammonia-liberating enzymes, e.g., urease and creatininase, and amino acid oxidases to build probes for urea, creatinine, and different amino acids. The most important ion-selective electrode using a 'solid membrane' is the glass electrode for pH measurement. By combination with enzymes that catalyze reactions that change the pH (especially hydrolases, e.g., urease, penicillinase), sensors are obtained for the respective substrates. Despite their selectivity for hydrogen ions their applicability is restricted because changes in the sample buffer capacity may affect the indicated pH signal. The formation or consumption of ammonia or carbon dioxide in the enzyme-catalyzed substrate conversion can be indicated by membrane-covered pH glass electrodes. The membranes allow the permeation of gases from the enzyme layer into the buffer-containing layer adjacent to the pH sensor. To obtain sufficient sensitivity the pH of the measuring solution should be above 9 and below 5 for indication of ammonia and carbon dioxide, respectively. Generally, these pH values differ substantially from the pH optima of deaminases and decarboxylases. Therefore, for the respective enzyme electrodes a compromise pH has to be found or the enzyme reaction has to be spatially separated from the potentiometric indicator and pH shift included between them (reactor electrode).

Platinum oxide and iridium oxide probes have been coupled to immobilized enzymes to give miniaturized biosensors. Furthermore, to construct microbiosensors, ISFETs sensitive to hydrogen or fluoride ions have been covered with a layer of enzyme. The light-addressable biosensor indicates the photocurrent generated by illumination of the rear by an array of light-emitting diodes. The signal reflects the local pH or redox potential at the surface. In this way, changes of the surface potential caused by enzyme reactions, microorganisms, or the action of receptors can be measured with high sensitivity.

In order to eliminate erroneous readings due to differences of the sample pH, dual sensor systems have been developed that use one enzyme-loaded pH sensor and one blank sensor as a reference. Problems arise due to independent fluctuations of the sensor's characteristics and changes in buffer capacity of the sample, which cannot be compensated for by this arrangement. The 'pH-stat' enzyme sensor circumvents this interference. The hydrogen or hydroxyl ions generated in the enzyme reaction are consumed by water electrolysis. The charge necessary in order to maintain a constant pH represents the amount of ions formed and it is independent of the buffer capacity. The device contains, in addition to the enzyme-covered pH FET, a pair of amperometric electrodes for oxygen or hydrogen generation, which are controlled by an on-chip pH-stat coulometer system.

Oxygen and hydrogen peroxide, which are the cosubstrate and product, respectively, of oxidase catalyzed reactions, as well as NAD(P)H, the coenzyme for ~300 dehydrogenases, can be determined amperometrically. The Clark-type oxygen electrode can detect oxygen or hydrogen peroxide. Furthermore, artificial redox mediators, such as hexacyanoferrate(III), *N*-methylphenzinium sulfate ( $\text{NMP}^+$ ), ferrocene, and benzoquinone, are suitable electroactive reaction partners. The electrode potential is decisive for the selectivity of the sensor. Any electroactive substance being converted at lower potential than hydrogen peroxide contributes to the total current. Thus, at an electrode potential of +600 mV for hydrogen peroxide measurement, ascorbic acid is also oxidized. Such interferences may in part be eliminated by covering of the anode with a thin, anion-bearing membrane, such as cellulose acetate. Alternatively, the difference between the signal obtained from a naked reference electrode and an enzyme-loaded sensor can be measured. In spite of these problems, enzyme electrodes based on oxidases combined with amperometric hydrogen peroxide measurement are still the most common type of enzyme electrodes.



To eliminate interferences by co-oxidizable sample constituents, the electrode potential is kept as low as possible. Therefore, the reaction partner chosen to be electrochemically indicated should be the one that is converted at the lowest possible potential. The natural electron acceptors of many oxidoreductases, for example, methanol dehydrogenase, alcohol dehydrogenase, and cytochrome  $b_2$ , and glucose, lactate, pyruvate, glycolate, sarcosine, and galactose, oxidases can be replaced by electron mediators, e.g., hexacyanoferrate(III),  $NMP^+$ , ferrocene, and benzoquinone. With these mediators an electrode potential of  $\sim +200$  mV can be applied; this decreases the interference of ascorbic acid and also enables such enzymes to be coupled with electrodes in oxygen-free solution. For this reason chemically modified electrodes, where the mediator is integrated with the amperometric electrode, have been combined with immobilized, mediator-dependent enzymes. In the simplest approach the solid mediator was introduced in crevices of a carbon electrode body and a dialysis membrane used to prevent it from leaching out. In the next stage of development the mediator was mixed with the carbon paste of the electrode body. Finally, all reagents – enzyme, mediator, and cofactor – were integrated into the top of the electrode.

As with mediators, enzymes may be bound covalently to the electrode surface, thus producing enzyme-chemically modified electrodes. Adsorption of redox polymers containing benzoquinone or heavy-metal ion complexes at carbon electrodes results in the catalysis of the electron transfer by ‘wiring’ the enzyme molecules to the electrode. When enzymes and mediators are co-immobilized at the surface or within the electrode, addition of auxiliary substances during the measuring process can be avoided; thus a reagentless measuring regime becomes feasible. An alternative to the application of mediators is direct electron transfer between the prosthetic group of the enzyme and the redox electrode.

Heterogeneous electron transfer reactions have been realized with more than 50 different proteins, mainly electron transferases, and also substrate-converting oxidoreductases. At bare metal electrodes irreversible adsorption accompanied by denaturation prevents a fast electron transfer to the protein molecules. Adsorption of modifiers that promote an appropriate orientation of the protein results in a facilitated direct electron transfer with different redox enzymes, for example, cytochromes and ferredoxins.

A symbiosis of mediated and direct electron transfer is the covalent fixation of electron-tunneling relays to the protein moiety of oxidoreductases.

The use of an electrode reaction offers the possibility to affect the enzyme reaction. In the oxygen-stabilized enzyme electrode the influence of the oxygen content of the sample solution is eliminated. For this purpose the same amount of oxygen as is consumed in the enzyme reaction is generated at a net electrode arranged within the enzyme layer. By using voltage pulses instead of a constant potential, both sensitivity and specificity of microenzyme electrodes can be improved.

## Measurable Analytes and Analytical Characteristics

Electrochemical enzyme sensors for  $\sim 140$  analytes have been described. Among them are low (relative) molecular mass substances (metabolites, drugs, nutrients, gases, metal ions, coenzymes, enzyme activators, and vitamins) as well as macromolecules (enzymes, lectins, nucleic acids, polymeric carbohydrates like starch and cellulose), viruses, and microorganisms. The analytes may be roughly subdivided as follows:

- Carbohydrates: amygdalin, cellobiose, cellulose, fructose, galactose, glucose, glucose-1-phosphate, glucose-6-phosphate, lactose, maltose, pullulane, starch, sucrose.
- Steroids: androstenedione, cholestenone, cholesterol, cholesterol esters, testosterone.
- Alcohols and carboxylic acids: acetate, ascorbate, benzoate, benzoquinone, catechol, chlorophenol, ethanol, fatty acids, formate, glycerol, gluconate, glycolate, glyoxylate, hexanol, hydroquinone, hydroxybutyrate, isocitrate, lactate, malate, methanol, oxalate, phenol, phosphoenolpyruvate, pyruvate, succinate, salicylate.
- Nitrogen compounds: acetylcholine, aminopyrine, aniline, bilirubin, cadaverine, caprolactam, epinephrine, guanosine, hypoxanthine, inosine, lecithin, monoamines, penicillin, spermine, spermidine, theophylline, urea, uric acid, xanthine.
- Amino acids: alanine, asparagine, aspartate, cysteine, glutamate, glutamine, glutathione, histidine, iso-leucine, leucine, lysine, methionine, phenylalanine, sarcosine, serine, tryptophan, tyrosine, valine.
- Cofactors: AMP, ADP, ATP, FAD, IMP,  $NAD^+$ , NADH,  $NADP^+$ , NADPH, hydrogen peroxide, pyridoxal phosphate, thiamine pyrophosphate.
- Proteins, peptides, and enzymes:  $\alpha$ -amylase, angiotensin, arginase, asparatame, cholinesterase, creatine kinase, glucoamylase, serum albumin, lactate dehydrogenase, peroxidase, pyruvate kinase, transaminases, alkaline phosphatase,  $\beta$ -galactosidase.

- Gases: CH<sub>4</sub>, CO, CO<sub>2</sub>, H<sub>2</sub>, NH<sub>3</sub>, NO, O<sub>2</sub>, SO<sub>2</sub>.
- Inorganic ions: cadmium, copper, fluoride, manganese, mercury, nitrate, nitrite, phosphate, sulfate, sulfide, sulfite, zinc, cyanide.

The signal-concentration dependence, i.e., measuring range, for electrochemical biosensors is linear over between one and three concentration decades. The lower limit of detection is  $2 \times 10^{-4} \text{ mol l}^{-1}$  with potentiometric electrodes and  $1 \times 10^{-8} \text{ mol l}^{-1}$  with amperometric monoenzyme electrodes. The use of the amplification reactions permits the detection limit to be decreased below the nanomole per liter range. Whereas the response time of potentiometric enzyme electrodes averages 2–10 min, with amperometric electrodes an assay can be conducted from within a few seconds up to one minute. This permits up to several hundred determinations per hour to be performed. Increasing complexity of the biochemical reaction system, e.g., by use of coupled enzyme reactions, brings about an increase in the overall measuring time.

Usually the imprecision found with enzyme electrodes ranges from 0.5% to 5%, which is, of course, dependent on the analytical setup used (e.g., manual sample injection versus flow injection). The lifetime of enzyme electrodes depends on several factors including the specific enzyme activity, the eventual formation of inactivating reaction products, and the operation conditions of the sensor. Generally, a high enzyme excess in the biocatalytic membrane is desirable in order to obtain a high stability. Lifetimes are in the range of several days or months, or between some hundreds and several thousand measurements.

## Coupled Enzyme Reactions in Enzyme Electrodes

Since not all enzyme-catalyzed reactions involve compounds detectable at an electrode, such as hydrogen ions, oxygen, or hydrogen peroxide, only a limited number of substances can be determined with one-enzyme sensors. In order to broaden the range of measurable substances, coupled enzyme reactions for analyte conversion in enzyme electrodes have been devised. In 'enzyme sequences' the primary product of the analyte conversion is further converted enzymatically with the formation of a measurable secondary product or reaction effect. On this basis, families of electrodes have been developed, which combine glucose-, lactate-, or alcohol-producing primary enzyme reactions with the respective oxidases. Such enzyme sequence electrodes are known for sucrose, lactose, maltose, glucose itself, gluconate, glucosinolate, bilirubin, ATP, glucose-6-phosphate,

cholesterol esters, fatty acid esters, acetylcholine, creatine, hypoxanthine, glycerol, lactate, pyruvate, inosine.

The application of 'competition' schemes realized by coupled enzyme reactions also provides access to analytes not determinable with the usual enzyme electrodes. Here the analytical information is gained either from the competitive action of two enzymes, of which one produces the signal, on the same substrate; or from the competition of the analyte with the substrate for the same (signal-generating) enzyme, the latter approach resembling that of inhibitor determination.

The coupling of enzyme pairs in recycling schemes provides a means for signal amplification, i.e., the enhancement of the sensitivity of enzyme electrodes. This method works analogously to the cofactor recycling known from biochemical analysis using dissolved enzymes. In a bienzyme electrode the analyte (lactate) is converted in the reaction of enzyme 1 (lactate oxidase) to a product (pyruvate), which is the substrate of enzyme 2 (lactate dehydrogenase). The latter catalyzes the regeneration of the analyte that thus becomes available for enzyme 1 again, and so forth. One of the coreactants (e.g., oxygen) is detectable directly or via additional reactions. (It is evident that the terms 'analyte' and 'product' are interchangeable here; this may be useful, but may also cause problems in real samples.) If enzyme 1 is present in sufficiently high concentration to ensure diffusion control, signal amplification is achieved by switching on enzyme 2 by addition of its cosubstrate (i.e., NADH). In such systems the analyte acts as a catalyst, being shuttled between both enzymes in the overall reaction of both cosubstrates. In this way significantly more cosubstrate will be converted than analyte is present in the enzyme membrane. Hence, the change in the parameter indicated at the electrode will greatly exceed that obtained with one-way analyte conversion. The ratio of the sensitivity in the linear measuring range of the amplified and the unamplified regime is termed the amplification factor. According to theory, the amplification factor  $G$  is given by

$$G = k_1 k_2 L^2 / 2(k_1 + k_2)D$$

where  $k_1$  and  $k_2$  are kinetic parameters,  $D$  is the diffusion coefficient of the analyte, and  $L$  is the membrane thickness.

Specifically in enzyme electrodes, where the sensitivity limit is determined by diffusion, the use of analyte recycling gives rise to a sensitivity enhancement by overcoming just this limit. On the other hand, the upper limit of the linear measuring range is decreased. Furthermore, since the enzyme excess

present in the membrane is included in the analyte conversion, the amplification factor will decrease with progressive enzyme inactivation during operation of the sensor.

Enzyme electrodes based on analyte recycling have been developed for the following compounds: lactate, pyruvate, glucose, gluconolactone, NADH,  $\text{NAD}^+$ , ethanol, benzoquinone, hydroquinone, ADP, ATP, glutamate, and malate. Depending on the enzymes and membrane materials used, amplification factors range from 3 to 48 000. Detection limits down to  $50 \text{ pmol l}^{-1}$  have been achieved. However, when dealing with extremely high amplification it should be borne in mind that the sensor signal becomes highly susceptible to minute amounts of contaminants affecting the enzyme reactions or the diffusion of the reactants.

Another coupling principle used for amplifying the response of enzyme electrodes is based on 'biocatalytical preconcentration' of intermediates. The method relies on an enzyme sequence. In the first step of the measurement the conversion of the analyte (glycerol) by the first enzyme (glycerol dehydrogenase) is permitted to proceed for a certain time (usual to equilibrium) during which an intermediate (NADH) accumulates in the enzyme layer. Then the second step, the actual measurement, is initiated by addition of an excess of cosubstrate (pyruvate) of the indicator enzyme reaction (lactate dehydrogenase + lactate monooxygenase). The conversion of the accumulated intermediate results in increased oxygen consumption. This method leads to amplification between 2- and 60-fold and has been used in the determination of glucose, glucose-6-phosphate, hypoxanthine, formate, glycerol, and  $\text{NADP}^+$ . It is applicable to all analytes being converted in sequential enzyme systems, provided the terminal reaction depends on a cofactor and leads to an electrochemical signal.

The selectivity of enzyme electrodes can be improved by means of another coupling principle that is capable of filtering chemical signals by eliminating interferences of the enzyme or electrode reaction caused by constituents of the sample. Compounds that interfere with signal transduction, e.g., ascorbic acid with anodic oxidation of hydrogen peroxide, can be transformed into inert products by reaction with an eliminator or 'anti-interference' enzyme (e.g., ascorbate oxidase). Since the conversion of analyte and interferent proceed in parallel, both the eliminator and the indicator enzyme may be co-immobilized in one membrane. On the other hand, constituents of the sample that are at the same time intermediates of coupled enzyme reactions can be eliminated before they reach the indicator enzyme layer. For this purpose several different enzyme

membranes have to be used. Of particular importance are enzymatic anti-interference systems for glucose because this compound is contained in most real samples where di- and polysaccharides have to be determined with enzyme sequence sensors based on glucose oxidase as indicator enzyme. Systems have been devised to eliminate lactate, ascorbic acid, ammonia, and oxygen.

In general, enzymes to be used in multienzyme electrodes should fulfill the following requirements:

- Their pH optima should be reasonably close to each other and to that of the indicator reaction.
- They should not be inhibited by necessary cofactors, effectors, or intermediates.
- Their cofactors or effectors should not react with each other.

## Application of Enzyme Electrodes

The most relevant fields of practical application of enzyme electrodes are medical diagnostics, followed by process control, food analysis, and environmental monitoring. The first commercial enzyme electrode-based analyzer (1975) was developed to meet the high demand for glucose determination in the blood of diabetic patients. Since 1975, analyzers for ~12 different analytes have been commercialized. As compared with conventional enzymatic analysis, the main advantages of such analyzers are the extremely low enzyme demand (a few milliunits per sample), their simplicity of operation, the high sample throughput rate, and high analytical quality.

In the medical field a trend toward hand-held devices based on either reusable or disposable glucose and lactate enzyme electrodes for home control of diabetics and for onsite monitoring of surgery and exercise is evident.

The selective determination of blood glucose is of extraordinary importance for the screening and treatment of diabetes. Johnson & Johnson, which is currently the market leader in blood glucose self-testing, launched its first biosensor product – the FastTake<sup>TM</sup> system – in April 1998. All the biggest suppliers in this market now sell biosensor-based products. Abbott Laboratories market ExacTech<sup>TM</sup>, Precision QID<sup>TM</sup> ranges, and recently Sof.Tact; Bayer the Glucometer<sup>TM</sup>, Elite, and Esprit; and Roche Diagnostics the Accu-check Advantage<sup>TM</sup>.

A remarkable new product is the FreeStyle<sup>TM</sup> glucose measuring system from TheraSense Inc. Because of the capillary-type sensor and coulometric measuring principle a 300 nl drop of whole blood is sufficient for one test. Even less painful is the ionophoretic detection with the GlucoWatch<sup>TM</sup>

automatic glucose biographer from Cygnus. This device controls blood glucose over range of  $2.2\text{--}22.2\text{ mmol l}^{-1}$  ( $40\text{--}400\text{ mg dl}^{-1}$ ) for up to 12 h using a single-point calibration. The automatic, frequent, and noninvasive measurements provide more information about glucose levels than the current standard of care.

For a continuous *in vivo* monitoring of glucose long-term stability, selectivity and sensitivity in the millimolar range is needed. To address these problems mass-transport limiting layers were assembled on  $250\text{ }\mu\text{m}$  diameter gold wires coated with glucose oxidase and redox polymer. The multiple sequential adsorption of polyanions and polycations results in an increase of dynamic range up to  $30\text{ mmol l}^{-1}$  glucose and at the same time a reduction of interferants. When a sufficient number of layers were applied glucose could be monitored over several hours. Extremely fast responding glucose sensors with sub-micrometer tip diameter have been constructed by using platinum-deposited flame etched carbon fiber electrodes additionally coated with glucose oxidase.

Other important analytes in medical diagnostics for which enzyme electrodes appear well suited are urea (for dialysis monitoring), uric acid, lactate, and enzyme activities, particularly transaminases and cholinesterase.

In process control and the food industry enzyme electrodes are employed mainly for the determination of carbohydrates, ascorbic acid, amino acids, and penicillin. However, the application of such sensors *in situ* in bioreactors is associated with significant difficulties:

- The sensor has to be sterilized, which, up to now, is virtually impossible.

- Calibration has to be performed by discrete measurements or *in situ*.
- Often the analyte concentration exceeds the linear range of the sensor.
- Various interfering compounds have to be dealt with.
- Thermal and mechanical stress inactivate the sensor.

Environmental control benefits from enzyme electrode-based measurements of toxic compounds, such as heavy metals, fume from organic wastes and pesticides in water, air, and soil.

**See also:** Amperometry. Blood and Plasma. Clinical Analysis: Glucose. Enzymes: Immobilized Enzymes. Ion-Selective Electrodes: Overview.

## Further Reading

- Cass T and Ligler FS (1998) *Immobilized Biomolecules in Analysis: A Practical Approach*. Oxford: Oxford University Press.
- Eggins BR (1996) *Biosensors: An Introduction*. Chichester & Stuttgart: Wiley & Teubner.
- Gizeli E and Lowe CR (2002) *Biomolecular Sensors*. London: Taylor & Francis.
- Hall EAH (1991) *Biosensors*. Englewood Cliffs, NJ: Prentice Hall.
- Lottspeich F and Zorbas H (1998) *Bioanalytik*. Heidelberg: Spektrum Akademischer Verlag.
- Scheller FW and Schubert F (1992) *Biosensors*. Amsterdam: Elsevier.
- Schmid RD (2002) *Taschenatlas der Biotechnologie und Gentechnik*. Weinheim: Wiley-VCH.

## Enzymes in Physiological Samples

**N Kiba**, University of Yamanashi, Takeda, Kofu, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Enzymes are proteins with catalytic activity, and many are holoenzymes, which are conjugated proteins, such as glyco-, hemo-, metallo-, and flavoproteins. Enzyme-catalyzed metabolic reactions take place in complexes of the enzyme with the reactants (substrates). The concentration of such complexes at any moment is dependent on the concentrations of

the enzyme and its substrate and the affinity of the enzyme for the substrate (Michaelis constant,  $K_m$ ). Measurements of enzyme activities (in micromoles per minute) in human body fluids are essential in clinical chemistry for diagnosis of diseases. The diagnosis is based on the fact that the activity of a particular enzyme in the fluids reflects the concentration of all enzyme molecules, inactive as well as active, which are released from damaged tissues or tumors. Therefore, the determination of enzyme concentrations (milligrams per liter) is of importance in providing a reference point for the measurement of enzyme activity. Though more than 800 kinds of enzyme responsible for



automatic glucose biographer from Cygnus. This device controls blood glucose over range of 2.2–22.2 mmol l<sup>-1</sup> (40–400 mg dl<sup>-1</sup>) for up to 12 h using a single-point calibration. The automatic, frequent, and noninvasive measurements provide more information about glucose levels than the current standard of care.

For a continuous *in vivo* monitoring of glucose long-term stability, selectivity and sensitivity in the millimolar range is needed. To address these problems mass-transport limiting layers were assembled on 250 µm diameter gold wires coated with glucose oxidase and redox polymer. The multiple sequential adsorption of polyanions and polycations results in an increase of dynamic range up to 30 mmol l<sup>-1</sup> glucose and at the same time a reduction of interferants. When a sufficient number of layers were applied glucose could be monitored over several hours. Extremely fast responding glucose sensors with sub-micrometer tip diameter have been constructed by using platinum-deposited flame etched carbon fiber electrodes additionally coated with glucose oxidase.

Other important analytes in medical diagnostics for which enzyme electrodes appear well suited are urea (for dialysis monitoring), uric acid, lactate, and enzyme activities, particularly transaminases and cholinesterase.

In process control and the food industry enzyme electrodes are employed mainly for the determination of carbohydrates, ascorbic acid, amino acids, and penicillin. However, the application of such sensors *in situ* in bioreactors is associated with significant difficulties:

- The sensor has to be sterilized, which, up to now, is virtually impossible.

- Calibration has to be performed by discrete measurements or *in situ*.
- Often the analyte concentration exceeds the linear range of the sensor.
- Various interfering compounds have to be dealt with.
- Thermal and mechanical stress inactivate the sensor.

Environmental control benefits from enzyme electrode-based measurements of toxic compounds, such as heavy metals, fume from organic wastes and pesticides in water, air, and soil.

**See also:** Amperometry. Blood and Plasma. Clinical Analysis: Glucose. Enzymes: Immobilized Enzymes. Ion-Selective Electrodes: Overview.

## Further Reading

- Cass T and Ligler FS (1998) *Immobilized Biomolecules in Analysis: A Practical Approach*. Oxford: Oxford University Press.
- Eggins BR (1996) *Biosensors: An Introduction*. Chichester & Stuttgart: Wiley & Teubner.
- Gizeli E and Lowe CR (2002) *Biomolecular Sensors*. London: Taylor & Francis.
- Hall EAH (1991) *Biosensors*. Englewood Cliffs, NJ: Prentice Hall.
- Lottspeich F and Zorbas H (1998) *Bioanalytik*. Heidelberg: Spektrum Akademischer Verlag.
- Scheller FW and Schubert F (1992) *Biosensors*. Amsterdam: Elsevier.
- Schmid RD (2002) *Taschenatlas der Biotechnologie und Gentechnik*. Weinheim: Wiley-VCH.

## Enzymes in Physiological Samples

**N Kiba**, University of Yamanashi, Takeda, Kofu, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Enzymes are proteins with catalytic activity, and many are holoenzymes, which are conjugated proteins, such as glyco-, hemo-, metallo-, and flavoproteins. Enzyme-catalyzed metabolic reactions take place in complexes of the enzyme with the reactants (substrates). The concentration of such complexes at any moment is dependent on the concentrations of

the enzyme and its substrate and the affinity of the enzyme for the substrate (Michaelis constant,  $K_m$ ). Measurements of enzyme activities (in micromoles per minute) in human body fluids are essential in clinical chemistry for diagnosis of diseases. The diagnosis is based on the fact that the activity of a particular enzyme in the fluids reflects the concentration of all enzyme molecules, inactive as well as active, which are released from damaged tissues or tumors. Therefore, the determination of enzyme concentrations (milligrams per liter) is of importance in providing a reference point for the measurement of enzyme activity. Though more than 800 kinds of enzyme responsible for

metabolic reactions have been found in the human body, among others ~30 kinds of enzyme are of importance in clinical enzymology. Further, among them only several enzymes are of significance in clinical diagnosis and now measured routinely.

In this article, the principles of sample preparation and measurement of the catalytic concentration of enzymes of clinical importance and of the absolute enzyme concentration are discussed.

## Preparation of Samples

Most physiological materials must be processed to obtain the optimal sample for measurement of enzyme activity and enzyme concentration. The treatments are obligatory for obtaining accurate and precise analytical results.

Enzymes are relatively fragile substances, with a tendency to undergo inactivation and denaturation under unsuitable conditions. In handling these materials, the first consideration must always be to avoid inactivation.

Serum from whole blood is used as the sample analyzed in order to obtain accurate analytical results because, with relatively few exceptions, the intracellular concentration of an enzyme greatly exceeds its extracellular value. Centrifugal techniques are most frequently employed for separation of blood. In spite of their general use, no standard recommendations exist regarding the centrifugal force, time, and temperature limits; the generally accepted conditions for 10 ml serum samples are 1500–2000 *g* for 10–15 min at room temperature. Fibrin aggregates are spun down with the blood cells. A sufficient clotting time (not less than 2 h) at room temperature provides clear separation of the clot.

Lymph, cerebrospinal fluid (CSF), and synovial fluid are centrifuged to remove any contaminating blood cells, fibrin, and particulate matter.

Saliva is centrifuged at 10 000 *g* for 30 min in a refrigerated centrifuge. Uncentrifuged saliva is used directly for assay of  $\beta$ -glucosidase (EC 3.2.1.21).

Urine is centrifuged and then dialyzed and gel filtered to remove enzyme inhibitors.

Tears are analyzed without any treatment.

In many cases, minced tissue is extracted with a buffer solution and then centrifuged. In some cases, membrane enzymes that are more firmly bound to the structural framework of mitochondria require special methods; these include drying with acetone, treatment with butanol–water mixtures, extraction of dried mitochondria with various organic solvents, extraction using aqueous solutions of detergents, treatment with chaotropic ions, or exposure to the action of hydrolytic enzymes.

## Measurement of Catalytic Concentration

Enzyme-catalyzed reactions usually follow the rate of appearance of a product or disappearance of a substrate. The catalytic activity of an enzyme is measured in terms of the catalyzed rate of conversion, not rate of reaction; in the physicochemical sense, the unit for the reaction rate is moles per liter per second, whereas for enzyme-catalyzed reactions it is defined by the unit moles per second (this quantity is called a katal (kat)). The activity has until now been expressed in units of micromoles per minute (unit (U) or international unit (IU)). One unit (IU) is  $16.67 \times 10^{-9} \text{ kat} = 16.67 \text{ nkat}$ .

In the measurement of enzyme activity, a high substrate concentration that is greatly in excess of the  $K_m$  value is always used, and the enzyme sample to be investigated is correspondingly diluted; under the conditions, the rate of the enzyme-catalyzed reaction depends only on the enzyme concentration, i.e., it is a zero order reaction. Even under conditions of substrate saturation, the measured catalytic activities are influenced by slight differences in reaction conditions, such as the temperature, composition and concentration of the buffer, pH value, nature of the substrate and its concentration, coenzymes, and protein content in the sample. Therefore, the results of measurement of the catalytic activity of an enzyme are in principle method dependent; direct comparison of the results between laboratories is made difficult by the use of different methods in different laboratories.

## Measurement of Enzyme Activity in Serum

Though various methods for the determination of enzyme activity are used routinely in clinical laboratories, many of them are covered by patents, and the components of their test kits are not commercially available. In clinical chemistry, reference methods and certified enzyme reference materials (enzyme calibrators) for determination of activities of enzymes in human blood serum are recommended by international commissions (ISOTC/212, International Federation of Clinical Chemistry IFCC, Institute for Reference Materials and Measurements EU, etc.) at present for improved accuracy and establishment of traceability chains of the enzyme measurement system. By using these reference materials, the inaccuracy (imprecision between laboratories) has been minimized to within several per cent in several enzymes, such as lactate dehydrogenase (LDH) (EC 1.1.1.27),  $\gamma$ -glutamyltranspeptidase ( $\gamma$ -GT) (EC 2.3.2.2), alanine aminotransferase (ALT) (EC 2.6.1.2), creatine kinase (CK) (EC 2.7.3.2), alkaline phosphatase (ALP) (EC 3.1.3.1), acid phosphatase (EC 3.1.3.2), and  $\alpha$ -amylase (AMY)



(EC 3.2.1.1.). The measuring temperature for reference methods is 30°C. When a result is obtained at a temperature other than 30°C, it should be related to that of the reference method. This can be done by using the reference materials or a temperature conversion factor.

### Measurement of Enzyme Activity in Body Fluids Other than Serum

Measurement of the catalytic activity of enzymes in body fluids other than blood, such as lymph, CSF, saliva, tears, and urine, is used as an aid to clinical diagnosis. Since lymph is recognized as a transport vehicle for enzymes, the lymph-serum ratio of enzyme activity is useful in estimating the origin and fate of a particular enzyme in lymph and serum. Methods usually applied to serum are applicable to measuring enzyme activity in lymph.

The diagnostic potential of measurement of the activity in CSF may not have been fully pursued owing to a certain risk to the patient by the puncture and to the small sample volume.

Saliva and tears can be obtained easily using harmless sampling techniques. Measurement of lysosomal enzymes (mainly hydrolases) in these fluids is applied for diagnosis of enzyme deficiency. Since the activities of many enzymes are comparatively low in their fluids, substrates have been used for fluorimetry, which is more sensitive and selective than spectrophotometry.

Enzymes are physiological constituents of urine, and most of them originate from renal tissue. The measurement of urinary enzymes has not yet been accepted as a routine clinical diagnostic method since urine contains many kinds of enzyme inhibitors and activators that must be removed before measurement of enzyme activity; the most commonly used separation methods are gel filtration and dialysis.

There is as yet no clear evidence that any measurement of the enzyme activity in synovial fluid is useful for diagnosis of disease.

### Methods Following Enzyme Reactions

Today, spectrophotometric methods are almost solely used for determination of the catalytic activities of enzymes in physiological samples in both manual and automated techniques because of their ease, simplicity, and sensitivity. No sampling is required, no reagents are necessary other than the actual reactants, and the result can be obtained with one small sample.

Many substrates and products of enzyme reactions absorb light in the visible region of the spectrum

and/or ultraviolet (UV) region. In spectrophotometric methods, the change in concentration with time (reaction rate) is directly proportional to the change in the absorbance ( $\Delta A$ ) with time ( $\Delta t$ ); the unit of measurement is  $\Delta A/\Delta t$ . These methods are directly applicable to oxidizing enzymes in physiological samples in which the second reactant (hydrogen acceptor) undergoes a change in absorbance on reduction by the substrate; e.g., for dehydrogenases the absorbance changes of the coenzymes nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP), which have an absorption band at 340 nm in the reduced state, not in the oxidized state, and of hexacyanoferrate(III) (420 nm) and dyes such as dichloroindophenol (600 nm) and methylene blue (610 nm), which absorb light in their oxidized forms, can be followed using these methods. Even when there is no absorbance change with the normal substrates, it is sometimes possible to find an artificial substrate that gives absorption changes at a suitable wavelength. For hydrolyzing enzymes such as phosphatases, nitrophenyl derivatives have been widely employed.

Even when none of the reactants or products of the enzyme reaction of interest lends itself to an appropriate absorption change, it is often possible to determine one of these components by adding another enzyme or color reagent that acts on the product of the reaction in such a way as to bring about the change; one of the products of the enzyme reaction is made a substrate or reactant of a second reaction (indicator reaction) that can be followed spectrophotometrically.

By coupling NAD(P)-dependent dehydrogenase reactions with the indicator reaction in which tetrazolium salts are reduced using reduced NAD (NADH) and reduced NADP (NADPH) to formazan in the presence of diaphorase (EC 1.6.4.3), the sensitivity is increased by a factor of  $\sim 2.5$  with respect to measurement at 340 nm.

Manometric methods are convenient and accurate methods for following reactions in which one of the reactants or products is a gas; they are adapted for measurement of the catalytic activity of carbonate hydrolyase (EC 4.2.1.1.) in erythrocytes and salivary glands.

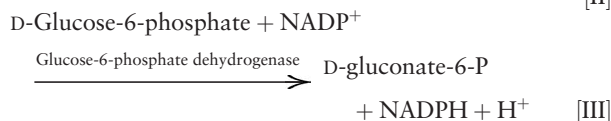
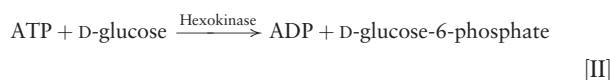
### Applications

**Determination of CK activity** CK catalyzes the reversible *N*-phosphorylation of creatine to form creatine phosphate by  $\text{Mg}^{2+}$ -ATP/ $\text{Mg}^{2+}$ -ADP complexes and thus is an important enzyme regulator of high-energy phosphate production and utilization within contractile tissues. CK occurs predominantly

in striated and smooth muscles, and brain, but erythrocytes, liver, and kidneys display almost no CK activity. Serum CK activity is measured as a sensitive indicator of injuries to the skeletal muscle and myocardium. High levels of serum CK activity are characteristic of acute myocardial infarction, muscular dystrophy, brain damage, and malignant tumors of the gastrointestinal tract, though the level of damage to the skeletal muscle or myocardium may depend partially on lymphatic transport of CK. CK has three isoenzymes (CK-MM, CK-BB, and CK-MB) in cytoplasm and two isoenzymes (sarcomeric and non-sarcomeric) in mitochondria. The diagnosis of each disease is accomplished by interpretation of the isoenzyme data because of their specific tissue distribution.

The activity can be measured spectrophotometrically. An internationally agreed reference method has been developed for routine determination in clinical chemistry.

The reactions in the method are shown in reactions [I]–[III].



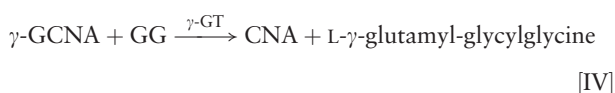
The rate of increase in absorbance at 340 nm due to the formation of NADPH is directly proportional to the CK activity. The CK reaction (reaction [I]) is coupled by the auxiliary hexokinase (EC 2.7.1.1) reaction (reaction [II]) to the glucose-6-phosphate dehydrogenase (EC 1.1.1.49) indicator reaction (reaction [III]). In these reactions,  $\text{Mg}^{2+}$  and *N*-acetylcysteine are required as activators to form complexes and a reactivator, respectively. To avoid the inhibition of  $\text{Ca}^{2+}$  and adenylate kinase (EC 2.7.4.3) in serum samples, ethylenediaminetetraacetic acid and di(adenosine-5-) pentaphosphate must be present. The activity in serum sample decreases to 98% of initial value after a week in a refrigerator.

**Determination of  $\gamma$ -GT**  $\gamma$ -GT is a glycoprotein and a membrane-bound enzyme that is located primarily on the outer surface of plasma membrane.  $\gamma$ -GT plays a significant role in the glutathione metabolism and the transfer of the  $\gamma$ -glutamyl moiety to certain amino acids that may be involved in the uptake of amino acids by a cell. The body's  $\gamma$ -GT is present in order of decreasing activity in kidneys (relative activity 100), pancreas (24), and liver (7). The significant elevation of  $\gamma$ -GT activity in serum is

associated almost with hepatobiliary disease, and the activity in urine increases in renal diseases; total  $\gamma$ -GT activities in patients' sera can be regarded as being liver specific. In heavy drinkers without hepatobiliary disease, the levels of activity in blood are kept higher than those of normal subjects.

A well-defined reference method is recommended as the routine method in clinical chemistry.

In this method, an artificial substrate, L- $\gamma$ -glutamyl-3-carboxy-4-nitroanilide ( $\gamma$ -GCNA), and glycylglycine (GG) are used as the donor and acceptor, respectively. Since the product of the reaction, 3-carboxy-4-nitroaniline (CNA), has an absorption maximum at 380 nm and  $\gamma$ -GCNA does not absorb above 410 nm, the activity, which is directly proportional to the amount of CNA liberated per unit time, is measured by monitoring the increase in absorbance at 410 nm.

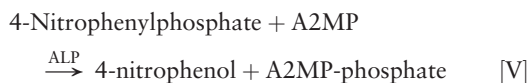


Under these conditions, the autotransfer reaction of  $\gamma$ -GCNA accounts for 1% of the product formed.

The activity in a serum sample is stable for 1 month in a refrigerator.

**Determination of ALP** ALP catalyzes the hydrolysis of phosphate monoesters such as esters of alcohols, phenols and amines, and inorganic pyrophosphates. ALP occurs in cell membranes throughout most of the tissues of the body. High levels of ALP activity in serum arise from hepatobiliary disease, bone disease, and transient hyperphosphatasemia in children. The value of the  $\gamma$ -GT activity is often useful to differentiate, whether an elevation of ALP activity comes from the liver or the bones; high levels of both suggest liver origin. ALP has many isoenzymes that are useful for laboratory diagnosis.

ALP activity is determined spectrophotometrically using an artificial substrate, 4-nitrophenylphosphate, and 2-amino-2-methyl-1-propanol (A2MP) as a phosphate acceptor and buffer agent. The rate of increase in absorbance at 405 nm due to the formation of 4-nitrophenol is proportional to the ALP activity.



In these measurements,  $\text{Mg}^{2+}$  and  $\text{Zn}^{2+}$  are required as activators because the enzyme is a metalloprotein and the activity is stimulated by  $\text{Mg}^{2+}$  and ALP contains a less tightly bound zinc atom that is involved in the catalytic process.

The activity in a serum sample is stable for 1 week in a refrigerator.

**Determination of LDH** Lactate dehydrogenase (LDH) catalyzes the equilibrium reaction of pyruvate to lactate. The activity of serum LDH is due to the presence of the enzyme released from damaged organs and tissues such as liver, heart, skeletal muscle, erythrocytes, etc. because LDH is located in the cytoplasm of the cells. Therefore, the activity of LDH is useful for screening for the existence of cell injuries, estimation of damaged tissues, and evaluation of treatment of diseases. LDH has five isoenzymes, and their patterns are of diagnostic importance.

The LDH activity is measured by determining spectrophotometrically the amount of NADH formed.

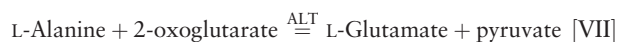


The amount of L-lactate oxidized per unit time, indicated by the continuous increase in absorbance at 340 nm due to the reduction of NAD, is directly proportional to the LDH activity. The measurement has to carry out in an alkaline medium and at relative high concentrations of lactate and NAD to obtain reasonable reaction velocities for accurate readings because the reverse reaction is strongly favored.

The activity in serum sample decreases to 97% of the initial value in a day in a refrigerator.

**Determination of ALT** ALT (formerly glutamate pyruvate transaminase) catalyzes the equilibrium transfer reaction of the amino group from L-alanine to 2-oxoglutarate to form L-glutamate and pyruvate; ALT requires pyridoxal phosphate as coenzyme, which acts as an amino carrier. It is found in the main organs, such as the liver, kidney, and heart. The ALT activity in serum is elevated in diseases of the liver.

The spectrophotometric method for measurement of the ALT activity involves two reactions, a reaction catalyzed by ALT and an indicator reaction.

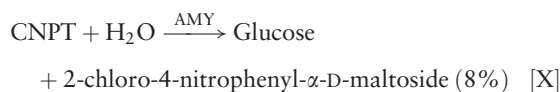
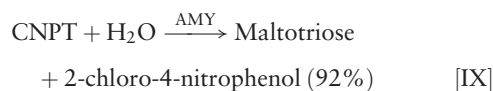


ALT catalyzes the reaction (reaction [VII]), and the pyruvate formed is reduced by NADH in a reaction catalyzed by LDH (indicator reaction, reaction [VIII]). The activity is measured by monitoring the decrease in absorbance of 340 nm due to the oxidation of NADH. The substrate concentrations are optimized theoretically, based on the kinetics of a two-substrate inhibited reaction because the ALT is subject to both substrate and product inhibitions.

The activity in serum samples decreases to 93% of the initial value after 3 days in a refrigerator.

**Determination of AMY** AMY catalyzes the hydrolysis of  $\alpha$ -1,4-glucan linkages to produce oligosaccharides. AMY is produced and stored in salivary glands and the pancreas and secreted into the digestive tract and eventually into the bloodstream. It is rapidly excreted into the urine because of its relatively small mass. Elevated levels of AMY activity in serum and urine are characteristics of acute pancreatitis or salivary lesions.

Many AMY reference methods based on malto-oligosaccharides with auxiliary enzymes and/or indicator enzymes have been recommended. For the direct method, an artificial substrate, 2-chloro-4-nitrophenyl- $\alpha$ -D-maltotriose (CNPT), is used for continuous monitoring of AMY activity. AMY hydrolyzes CNPT to maltotriose and 2-chloro-4-nitrophenol at high thiocyanate concentrations. The reaction is followed by monitoring the increase in absorbance due to the production of 2-chloro-4-nitrophenoxide at 405 nm. A side reaction (reaction [X]) occurs with formation of glucose and 2-chloro-4-nitrophenyl- $\alpha$ -D-maltose, which accounts for 8% of the product formed.



In the reaction,  $\text{Ca}^{2+}$  and  $\text{Cl}^-$  are required as activators.

The activity in a serum sample is stable for 5 days at room temperature.

## Determination of Enzyme Concentration

Determination of the concentration of all enzyme molecules, inactive as well as active, is essential in studying the clinical chemistry for diagnosis of disorders of metabolism because activity measurements alone are not sufficient for distinguishing whether the increase represents an increase in the number of all enzyme molecules present in sample; in some cases, inactive molecules are released from damaged tissues and a change in activity may not necessarily be directly related to the change in the number of molecules of enzyme present.

For determination of the enzyme concentration, methods of protein assay are frequently used; these are based on the general properties and behavior of proteins (nonspecific methods). For accurate determinations, the enzyme protein must be carefully purified because all methods are susceptible to

interferences from a large number of different enzymes and proteins that perhaps are present in physiological samples, and biological molecules, such as nucleic acids and amino acids.

Enzyme proteins have a number of general properties in common. However, of course, different enzyme proteins vary greatly in their amino acid composition and sequence as well as in their size, shape, and charge. Since the general properties and behavior of enzyme proteins are not always identical for different enzyme proteins, ideally a calibration curve should be constructed using the purified enzyme under study. In most cases bovine serum albumin is used as the standard or reference; therefore, enzyme protein quantities are obtained as 'serum albumin units' rather than as absolute amounts. Since different assay methods dependent on different general properties of enzymes, consistency between methods may be low. Furthermore, the various methods can hardly be expected to give equal responses with different enzymes; variability with different enzymes is large. Thus, no single method is optimal for every application.

### Enzyme Purification Methods

The body fluids and the extracts containing the enzyme will also contain numerous other substances. Small molecules in the sample can be easily removed through gel filtration and dialysis. The remaining large molecules are proteins and polysaccharides. The main part of purification consists of a series of fractionations by which the protein enzymes is separated from the other proteins present.

Although there are many fractionation methods, such as fractional precipitation using salts, organic solvents, or a change in pH, fractional denaturation by heating, and fractional adsorption, the purification of enzymes through chromatography on a column is the most effective of all the separation methods. The physicochemical properties of the enzyme that should be utilized in chromatography include the hydrophobicity, charge, biological recognition, and size. The mechanism of the separation depends in different cases on the adsorption, ion exchange, affinity to immobilized ligands, or molecular sieving, but the technique is practically the same for all.

Adsorption chromatography usually employs microcrystalline hydroxyapatite columns that adsorb proteins under conditions above pH 5. Gradient or step-wise elution with increasing concentration of eluting agent in the mobile phase may be used to release the enzyme required.

Hydrophobic chromatography of an enzyme employs cellulose derivatives with attached linear alkyl groups

$(-(CH_2)_n-CH_3, n = 1, 3, 5, 7, 9)$ , alkylamino groups  $(-(CH_2)_nNH_2, n = 2, 4, 6, 8, 10)$ , or the phenyl group.

In ion exchange chromatography cation exchangers, such as carboxymethyl cellulose, sulfoethyl cellulose, and sulfopropyl cellulose, are commonly used. The anion exchangers most commonly used are diethylaminoethyl cellulose and triethylaminoethyl cellulose.

Affinity chromatography involves the immobilization of a ligand (substrate, analog, inhibitor or effector) on agarose as a support that can be used selectively to adsorb the enzyme from a sample solution passed over it.

A mixture of enzymes can be separated by molecular size using the techniques of size exclusion chromatography. Packed bed columns of cross-linked dextran (Sephadex series), granulated polyacrylamide gels (Bio-Gel P series), or bead-formed agarose gels (Sephacrose series, Bio-Gel A series) which act as molecular sieves, are employed; enzyme proteins are eluted from the column in decreasing order of size, i.e., molecule weight.

In each technique, the eluate from the separation column is collected as a series of fractions by a fraction collector. The resulting enzyme solutions are concentrated to a smaller volume for the following determination. For the purpose of reduction in volume, ultrafiltration can be easily and rapidly performed; membranes with different porosities, which will retain the chosen macromolecule while being permeable to the solvent and to small molecules, are commercially available.

### Quantitation of Enzyme

**Nonspecific methods** Nonspecific methods for the determination of enzyme proteins are summarized in Table 1.

**Dry-weight analysis.** If a solution of highly pure enzyme is available, a dry-weight analysis may be possible. Though this should provide an exact value for the enzyme protein content, a great expense in labor and materials has to be devoted to the enzyme purification. Determination of the dry weight of a protein is important for investigations of protein structure.

**UV absorption method.** Most enzyme proteins absorb UV light at an absorption maximum at  $\sim 275\text{--}280\text{ nm}$  due to their content of tyrosine and tryptophan. If a pure enzyme is available, the absorbance at  $280\text{ nm}$  can be used for quantitation; for many enzyme proteins, a solution containing  $1\text{ mg}$  of protein per milliliter has an absorbance of  $1.0$  at  $280\text{ nm}$ . However, others differ in their absorbance with more than an order of magnitude due to the

**Table 1** Nonspecific methods for enzyme determination

<i>Method</i>	<i>Principle</i>	<i>Determination range</i>
Dry-weight method	Solution of the pure enzyme is evaporated to dryness and weight	> 0.1 µg
UV absorption method	Enzymes exhibit UV absorption maxima ~280 nm, based on their content of tyrosine and tryptophan residues. To increase the detectability, measurements at 215/225 nm are used because of the peptide bond	100–1000 µg ml <sup>-1</sup> (280 nm); 5–100 µg ml <sup>-1</sup> (215/225 nm)
Biuret method	Peptide bonds in a enzyme react with the Cu(II) ion in a strongly alkaline solution and produce a complex that can be determined spectrophotometrically at 540 or 310 nm	0.2–10 mg ml <sup>-1</sup> (540 nm); 30–500 µg ml <sup>-1</sup> (310 nm)
Bicinchoninic acid method	Bicinchoninic acid is added to the reaction products of the biuret reaction to form the copper–bicinchonate complex. The absorbance of the complex is measured spectrometrically at 560 nm	0.5–50 µg ml <sup>-1</sup>
Lowry method	The biuret reaction is incorporated with the reduction reaction of Folin-Ciocalteu's reagent and the resulting phosphomolybdenum blue is quantitated spectrophotometrically at 750 nm	2–100 µg ml <sup>-1</sup>
Protein–dye binding methods	Enzymes are stained with the dye Coomassie Brilliant Blue G-250 to produce color yields that exhibit absorption maxima at 595 nm	2–100 µg ml <sup>-1</sup>
Kjeldahl method	Ammonia produced by the oxidation of the enzyme by concentrated sulfuric acid is measured using the indophenol technique	0.5–10 µg ml <sup>-1</sup>
Fluorimetry	Amino groups in an enzyme derivatize with fluorescamine or <i>o</i> -phthalaldehyde, which is not fluorescent, and the resulting derivative is measured fluorimetrically	0.005–10 µg ml <sup>-1</sup>
Hydrolysis method	After alkaline hydrolysis, amino acids produced are reacted with ninhydrin or with <i>o</i> -phthalaldehyde and then measured spectrometrically or fluorimetrically	(ninhydrin) 20–500 µg ml <sup>-1</sup> ; ( <i>o</i> -phthalaldehyde) 10–1000 ng ml <sup>-1</sup>
Turbidimetry	Trichloroacetic acid is added to a solution of the enzyme to produce a suspension in the presence of a surface-active agent. The absorbance of the suspension is measured at 340 nm	5–120 µg ml <sup>-1</sup>
Refractometry	Refractive index change for the enzyme solution is measured using a differential refractometer	0.1–5 mg ml <sup>-1</sup>

different contents of the aromatic amino acids. Measurements at 280 nm can be used effectively only for monitoring proteins in effluents from chromatographic columns. To increase the detectability, measurement at below 220 nm is performed using shorter wavelength pairs, 215/225 nm; the absorption below 220 nm is based on the peptide bond ( $-\text{C}(=\text{O})-\text{NH}-$ ). UV absorption methods are non-destructive and can be easily and rapidly performed.

**Biuret method.** The biuret method is based on the reaction of peptide bonds with the copper(II) ion in

an alkaline solution leading to the formation of a violet colored complex. The variability with different proteins is smaller than that in the UV absorption method because the color development depends on the number of peptide bonds rather than the amino acid residues. The sensitivity of the biuret method is rather low. It appears that another method is more appropriate when better detectability is required. To improve the detectability, many modifications of the original procedure have been proposed that make use of the intense light absorption of copper complexes



with chromophores such as bicinchoninic acid (4,4'-dicarboxy-2,2'-biquinoline). The biuret method is practically suited for use as a reference method for determination of total enzyme protein owing its high precision.

**Lowry method.** The Lowry method is the most generally accepted owing to its sensitivity, simplicity, and reproducibility; it is  $\sim 20$  times more detectable than the UV absorption method and 100 times more sensitive than the biuret method. Since the color development in the Lowry method depends on the content of aromatic amino acid residues, the variation of response with different proteins is surely greater than in the biuret method but generally less than in UV adsorption measurement.

**Protein-dye binding method.** The protein-dye binding method is based on the shift in the absorption maximum of Coomassie Brilliant Blue G-250 from 465 to 595 nm that occurs upon binding of the dye to protein. The main disadvantage of this dye binding method is the variation of response with various proteins. The relative standard deviation in determinations of concentrations of 23 kinds of protein using the present method is about twice the value obtained using the Lowry method. Since bovine serum albumin gives a much higher color yield than most other proteins, the purified enzyme under study must be used as standard protein. If a sufficient quantity of the enzyme in purified form is not available in order to construct the calibration curve, bovine gamma globulin is recommended as a better standard that gives a more normal response.

**Kjeldahl method.** The Kjeldahl method is an important method for accurately determining nitrogen in proteins. Since a protein generally has a nitrogen content of  $\sim 16\%$  (12–19%), the quantity of protein can be calculated from a knowledge of the per cent nitrogen contained in it. Although, as mentioned above, rapid methods for determining proteins exist, the Kjeldahl method is the standard on which all other methods are based. The protein is digested with sulfuric acid to decompose it and convert the nitrogen to the ammonium ion, followed by spectrophotometric quantitation of the ammonium ion by reaction with hypochloride and phenol, forming indophenol blue.

**Other methods.** Other methods of high detectability (determination range  $0.005\text{--}10\ \mu\text{g ml}^{-1}$ ) include fluorimetric techniques such as binding fluorescamine, o-phthalaldehyde, dansyl chloride, 4-fluoro-7-nitrobenzo-2-oxa-1,3-diazole, 1-anilinonaphthalene-8-sulfonic acid, and hypochlorite-thiamine. Estimation of protein using ninhydrin ( $20\text{--}500\ \mu\text{g ml}^{-1}$ ) or o-phthalaldehyde ( $10\text{--}1000\ \text{ng ml}^{-1}$ ) after alkaline hydrolysis, turbidimetry ( $5\text{--}12\ \mu\text{g ml}^{-1}$ ), or refractometry ( $0.1\text{--}5\ \text{mg ml}^{-1}$ ) has been proposed.

**Specific methods** Immunochemical methods have been developed for identification and specific determination of proteins that have no catalytic activity. The method is based on the special relationship between a protein and a specific antibody that has been raised against the protein. The advantages of an immunoassay lie in its high specificity and detectability. Since the specificity is independent of the activity of the antigen protein, the method measures all enzyme molecules, inactive as well as active. The methods depend on the availability of specific antisera directed against the particular enzymes under study. The antibody is raised in an experimental animal by injection of a sample of protein antigen. Selective precipitation method and enzyme linked immunosorbent assay (ELISA) are commonly used for the specific detection and reliable measurement of a particular enzyme in a raw sample, such as a tissue extract or body fluids.

**Selective precipitation methods.** The principle of the selective precipitation method is that excess antibody is added to a solution of enzyme to precipitate it as an insoluble antigen-antibody complex. The precipitate may be separated and analyzed for its enzyme content. This method permits the specific measurement of several nanograms of enzyme protein.

**ELISA** ELISA is based on the principle of competition between enzyme-labeled and unlabeled antigens for a specific antibody present in a limited concentration in the assay mixture. The labeled antigen and antibody form a reversible complex. The complex is then detectable by assay of the linked enzyme. Since a pure sample of antigen is required for preparation of standards, the method has not been applied extensively to the determination of enzyme concentrations.

**See also:** **Blood and Plasma. Clinical Analysis:** Sample Handling. **Electrophoresis:** Clinical Applications. **Enzymes:** Overview; Industrial Products and Processes; Enzyme-Based Assays. **Immunoassays, Techniques:** Enzyme Immunoassays. **Liquid Chromatography:** Size-Exclusion; Affinity Chromatography; Clinical Applications. **Peptides. Proteins:** Physiological Samples. **Quality Assurance:** Reference Materials.

## Further Reading

Bais R and Philcox M (1994) Approved recommendations on IFCC methods for the measurements of catalytic concentrations of enzymes. Part 8. IFCC method for lactate dehydrogenase (L-lactate: NAD<sup>+</sup> oxidoreductase, EC 1.1.1.27). *European Journal of Clinical Chemistry and Clinical Biochemistry* 32: 639–655.



- Bergmeyer HY, Bergmeyer J, and Grassl M (1986) *Methods of Enzymatic Analysis*, 3rd edn., vols. I–V. Weinheim: VCH.
- Bergmeyer HY, Horder M, and Rej R (1986) Approved recommendations on IFCC methods for the measurements of catalytic concentrations of enzymes. Part 3. IFCC method for alanine aminotransferase (L-alanine: 2-oxoglutarate aminotransferase, EC 2.6.1.2). *Journal of Clinical Chemistry and Clinical Biochemistry* 24: 481–495.
- Bergmeyer HY, Horder M, and Rej R (1986) Approved recommendations on IFCC methods for the measurements of catalytic concentrations of enzymes. Part 2. IFCC method for aspartate aminotransferase (L-aspartate: 2-oxoglutarate aminotransferase, EC 2.6.1.1). *Journal of Clinical Chemistry and Clinical Biochemistry* 24: 497–510.
- Bowers GN, Bergmeyer HU, Horder H, and Moss DW (1980) Approved recommendations on IFCC methods for the measurements of catalytic concentrations of enzymes. Part 1. General considerations concerning the determination of the catalytic concentration of an enzyme in the blood serum or plasma of man. *Journal of Clinical Chemistry and Clinical Biochemistry* 18: 89–95.
- Dubach UC and Le Hir M (1984) Critical evaluation of the diagnostic use of urinary enzymes. *Contributions to Nephrology* 42: 74–80.
- Hoder M, Elser RC, Gerhardt W, Mathieu M, and Sampson EJ (1991) Approved recommendations on IFCC methods for the measurements of catalytic concentrations of enzymes. Part 7. IFCC method for creatine kinase (ATP: creatine N-phosphotransferase, EC 2.7.3.2). *Journal of Clinical Chemistry and Clinical Biochemistry* 29: 435–456.
- IFCC Scientific Division, Working Group on Enzymes (1999) IFCC methods for the measurements of catalytic concentrations of enzymes. Part 9. IFCC method for  $\alpha$ -amylase [1,4- $\alpha$ -glucan 4-glucanohydrolase, EC 3.2.1.1]. *Clinica Chimica Acta* 281: S5–S39.
- Ishikawa E (1999) Ultrasensitive and rapid enzyme immunoassay. In: Work TS and Work E (eds.) *Laboratory Techniques in Biochemistry and Molecular Biology*, vol. 27. Tokyo: Elsevier.
- Janson JC and Ryden L (eds.) (1989) *Protein Purification: Principles, High Resolution Methods, and Applications*. New York: VCH.
- Nakajima T, Yanaihara N, and Miyazawa T (eds.) (1988) *Physiologically Active Peptides*. Tokyo: Kodansha.
- Scopes RK, (1994) Protein purification: principles and practice. In: Cantor CR (ed.) *Springer Advanced Texts in Chemistry*, 3rd edn. New York: Springer.
- Shaw LM, Stromme JH, London JL, and Theodorsen L (1983) IFCC methods for the measurements of catalytic concentrations of enzymes. Part 4. IFCC method for  $\gamma$ -glutamyltransferase, EC 2.3.2.2. *Clinica Chimica Acta* 135: 315F–338F.
- Tietz NW, Rinker AD, and Shaw (1983) IFCC methods for the measurements of catalytic concentrations of enzymes. Part 5. IFCC method for alkaline phosphatase (orthophosphoric monoester phosphohydrolase, alkaline optimum, EC 3.1.3.1). *Clinica Chimica Acta* 135: 339F–367F.
- Wild D (ed.) (1994) *The Immunoassay Handbook*. New York: Stockton Press.

## Industrial Products and Processes

P Williams, Mould Green, Huddersfield, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Mikhail V Rekharsky and Alexei M Egorov, pp. 1212–1216, © 1995, Elsevier Ltd.

### Introduction

The world enzyme market is currently valued in the region of US\$1.3 billion and is made up of markets such as detergents (37%), textiles (12%), starch (11%), baking (8%), and animal feed (6%).

There has been much recent corporate investment in the field of biocatalysis with leading fine chemical companies now beginning to supplement traditional chemical and metal catalysis with biocatalysis. This will lead to more and more practical industrial uses

of enzymes in the twenty-first century. A comprehensive review of present industrial processes using enzymes can be found in the book *Industrial Enzymology*.

The pharmaceutical industry is driven primarily by the need for innovative medicines with new vaccines occupying only a small area of the market compared to small molecule products for oral ingestion. The variable activity of different optical isomers and regulatory pressures has led to the desire to manufacture only chirally pure therapeutics. Thus, biocatalysis with its high selectivity offers great opportunities. The top two drugs in terms of global sales – Lipitor (Pfizer) and Zocor (Merck), with combined sales of almost \$14 billion in 2002 – are single-enantiomer small-molecule drugs. Worldwide sales of single-enantiomer drugs have now reached more than \$159 billion.

- Bergmeyer HY, Bergmeyer J, and Grassl M (1986) *Methods of Enzymatic Analysis*, 3rd edn., vols. I–V. Weinheim: VCH.
- Bergmeyer HY, Horder M, and Rej R (1986) Approved recommendations on IFCC methods for the measurements of catalytic concentrations of enzymes. Part 3. IFCC method for alanine aminotransferase (L-alanine: 2-oxoglutarate aminotransferase, EC 2.6.1.2). *Journal of Clinical Chemistry and Clinical Biochemistry* 24: 481–495.
- Bergmeyer HY, Horder M, and Rej R (1986) Approved recommendations on IFCC methods for the measurements of catalytic concentrations of enzymes. Part 2. IFCC method for aspartate aminotransferase (L-aspartate: 2-oxoglutarate aminotransferase, EC 2.6.1.1). *Journal of Clinical Chemistry and Clinical Biochemistry* 24: 497–510.
- Bowers GN, Bergmeyer HU, Horder H, and Moss DW (1980) Approved recommendations on IFCC methods for the measurements of catalytic concentrations of enzymes. Part 1. General considerations concerning the determination of the catalytic concentration of an enzyme in the blood serum or plasma of man. *Journal of Clinical Chemistry and Clinical Biochemistry* 18: 89–95.
- Dubach UC and Le Hir M (1984) Critical evaluation of the diagnostic use of urinary enzymes. *Contributions to Nephrology* 42: 74–80.
- Hoder M, Elser RC, Gerhardt W, Mathieu M, and Sampson EJ (1991) Approved recommendations on IFCC methods for the measurements of catalytic concentrations of enzymes. Part 7. IFCC method for creatine kinase (ATP: creatine N-phosphotransferase, EC 2.7.3.2). *Journal of Clinical Chemistry and Clinical Biochemistry* 29: 435–456.
- IFCC Scientific Division, Working Group on Enzymes (1999) IFCC methods for the measurements of catalytic concentrations of enzymes. Part 9. IFCC method for  $\alpha$ -amylase [1,4- $\alpha$ -glucan 4-glucanohydrolase, EC 3.2.1.1]. *Clinica Chimica Acta* 281: S5–S39.
- Ishikawa E (1999) Ultrasensitive and rapid enzyme immunoassay. In: Work TS and Work E (eds.) *Laboratory Techniques in Biochemistry and Molecular Biology*, vol. 27. Tokyo: Elsevier.
- Janson JC and Ryden L (eds.) (1989) *Protein Purification: Principles, High Resolution Methods, and Applications*. New York: VCH.
- Nakajima T, Yanaihara N, and Miyazawa T (eds.) (1988) *Physiologically Active Peptides*. Tokyo: Kodansha.
- Scopes RK, (1994) Protein purification: principles and practice. In: Cantor CR (ed.) *Springer Advanced Texts in Chemistry*, 3rd edn. New York: Springer.
- Shaw LM, Stromme JH, London JL, and Theodorsen L (1983) IFCC methods for the measurements of catalytic concentrations of enzymes. Part 4. IFCC method for  $\gamma$ -glutamyltransferase, EC 2.3.2.2. *Clinica Chimica Acta* 135: 315F–338F.
- Tietz NW, Rinker AD, and Shaw (1983) IFCC methods for the measurements of catalytic concentrations of enzymes. Part 5. IFCC method for alkaline phosphatase (orthophosphoric monoester phosphohydrolase, alkaline optimum, EC 3.1.3.1). *Clinica Chimica Acta* 135: 339F–367F.
- Wild D (ed.) (1994) *The Immunoassay Handbook*. New York: Stockton Press.

## Industrial Products and Processes

P Williams, Mould Green, Huddersfield, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Mikhail V Rekharsky and Alexei M Egorov, pp. 1212–1216, © 1995, Elsevier Ltd.

### Introduction

The world enzyme market is currently valued in the region of US\$1.3 billion and is made up of markets such as detergents (37%), textiles (12%), starch (11%), baking (8%), and animal feed (6%).

There has been much recent corporate investment in the field of biocatalysis with leading fine chemical companies now beginning to supplement traditional chemical and metal catalysis with biocatalysis. This will lead to more and more practical industrial uses

of enzymes in the twenty-first century. A comprehensive review of present industrial processes using enzymes can be found in the book *Industrial Enzymology*.

The pharmaceutical industry is driven primarily by the need for innovative medicines with new vaccines occupying only a small area of the market compared to small molecule products for oral ingestion. The variable activity of different optical isomers and regulatory pressures has led to the desire to manufacture only chirally pure therapeutics. Thus, biocatalysis with its high selectivity offers great opportunities. The top two drugs in terms of global sales – Lipitor (Pfizer) and Zocor (Merck), with combined sales of almost \$14 billion in 2002 – are single-enantiomer small-molecule drugs. Worldwide sales of single-enantiomer drugs have now reached more than \$159 billion.

## Application of Enzymes in the Detergent, Textile, Leather, Paper, Animal Feed, Baking, and Alcohol Industries

The detergent industry saw the first large-scale use of microbial enzymes, namely, proteinases. However, in the late 1980s lipases were introduced to break down fats into water-soluble components of glycerol and fatty acids. Amylases remove starch-based stains, and since the early 1990s cellulases have been used to remove cotton-based cellulose. The textile industry is rapidly increasing its dependence on enzymology where amylases now strip the fabrics' protective coating of starch whilst cellulases provide the 'worn look' denim. The other industries also use hydrolytic enzymes to manufacture their products (see Table 1).

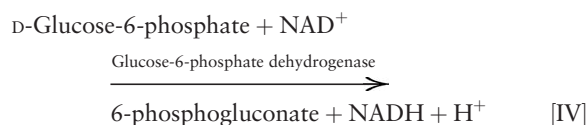
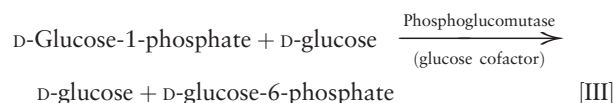
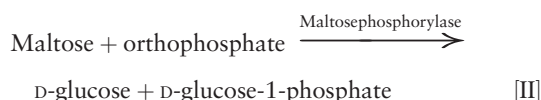
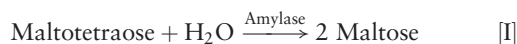
### Determination of $\alpha$ - and $\beta$ -Amylase

Amylases are responsible for hydrolysis of starch to oligosaccharides.  $\alpha$ -Amylase hydrolyzes the 1,4- $\alpha$ -glucoside bonds in compounds involving three or more molecules of glucose.  $\beta$ -Amylase liberates (mainly)  $\beta$ -maltose from starch and other compounds. The simplest way to determine amylase activity requires the determination of the time required to change the starch iodine color from blue

to reddish brown. This time is proportional to the amylase activity.

Two different kinetic methods are available for the determination of amylase activity used in clinical diagnostics. These methods can be applied for the analysis of industrial products.

The reactions involved in the first method are



where NADH is the abbreviation for (reduced) nicotinamide adenine dinucleotide.

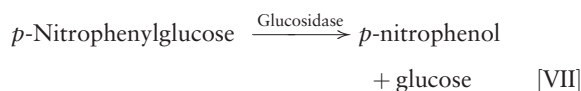
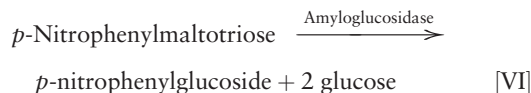
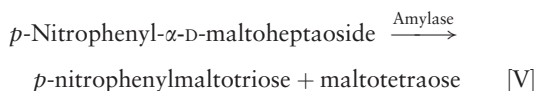
The rate of increase in absorbance at 340 nm due to the formation of NADH is directly proportional to

**Table 1** Applications of enzymes in the detergent, textile, leather, paper and pulp, and alcohol industries

Enzyme	Substrate	Industry	Source
$\alpha$ - and $\beta$ -Amylases	Starch	Detergent, textile, alcohol, paper, and pulp	<i>Bacillus amyloliquefaciens</i>
		Baking	<i>Bacillus licheniformis</i> <i>Bacillus polymyxa</i> <i>Aspergillus oryzae</i>
Amyloglucosidase	Dextrins	Alcohol	<i>Aspergillus niger</i>
		Baking	<i>Rhizopus nivens</i> <i>Aspergillus niger</i>
Cellulases (alkaline, neutral, and acidic)	Cellulose	Detergent, textile	<i>Phanerochaete chrysosporium</i> <i>Trichoderma reesei</i> <i>Bacillus</i> spp., <i>Trichoderma longibrachiatum</i>
Lipases	Lipids	Detergent	<i>Candida ruqusa</i>
		Baking	<i>Geotrichum parasitica</i> <i>Aspergillus oryzae</i>
Proteinases (alkaline, neutral, and pepsin- or rennin-like)	Proteins	Detergent, textile, alcohol, leather	<i>Aspergillus oryzae</i>
		Baking	<i>Endothia parasitica</i> <i>Bacillus subtilis</i>
Xylanases	Xylan	Animal feed, baking, paper, and pulp	<i>Trichoderma longibrachiatum</i>

amylase activity. Multienzyme preparations catalyzing these reactions are commercially available.

The reactions involved in the second method are



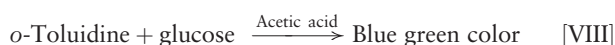
The increase in absorbance at 405 nm is proportional to the amylase activity. It might be pointed out that many different assay procedures and 'units of amylase activity' are used in the literature.

#### Determination of Amyloglucosidase

Amyloglucosidase removes consecutive glucose molecules from starch and some other compounds. Reactions [VI] and [VII] are also suitable for detection of amyloglucosidase activity. The best method is generally to determine the rate of glucose formation.

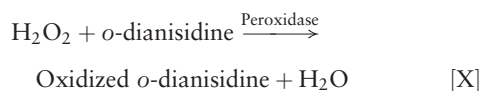
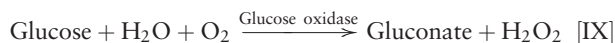
There are many different experimental procedures for glucose detection. One nonenzymatic and two enzymatic methods for glucose detection are described below.

The nonenzymatic method for glucose detection involves reaction [VIII] carried out in a boiling water bath:



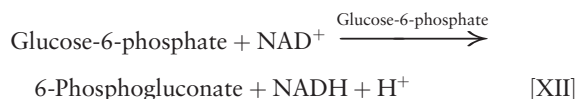
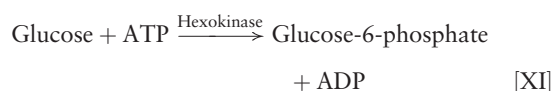
The absorbance, measured spectrophotometrically at 620–650 nm, is directly proportional to the glucose concentration.

The first enzymatic method for glucose detection involves the reactions



The absorbance of the brown product, measured at 425–475 nm, is proportional to the total glucose concentration.

The second enzymatic method for glucose detection makes use of reactions [XI] and [XII]:



The absorbance at 340 nm due to the formation of NADH is proportional to the glucose concentration.

One unit of amyloglucosidase activity is defined as the amount of enzyme that releases 1 mg of glucose from starch in 1 min at pH 4.5 and 55°C.

#### Determination of Cellulase

Cellulase catalyzes the hydrolysis of 1,4- $\beta$ -glucosidic bonds in cellulose, lichenin, and other polysaccharides. One unit of cellulase activity releases 1  $\mu\text{mol}$  of glucose from cellulose in 1 h at pH 5.0 and 37°C. Useful methods for determining the rate of glucose formation are described above, and may be employed for cellulase determination.

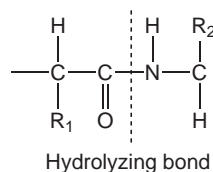
#### Determination of Lipase

Lipase liberates fatty acids from triglycerides. The rate of fatty acids formation is measured by titration with standard sodium hydroxide solution. Microcalorimetric methods can also be used; the enzyme is determined by the heat generated in the reaction in Tris buffer at pH 8.0. One unit of lipase activity liberates 1 microequivalent of fatty acid from a triglyceride in 1 h at pH 7.7 and 37°C.

#### Determination of Proteinases (Proteases)

Proteinases catalyze the hydrolysis of peptide bonds in proteins. The selectivity of action of some proteinases is shown in Figure 1.

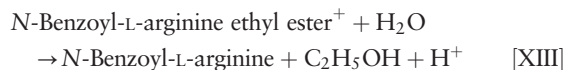
Pepsin, pepsin-like enzymes, chymosin, rennin, and other acid proteinases have an activity optimum at pH 2.0–3.5; papain, trypsin, chymotrypsin, and similar enzymes are most active at neutral pH (pH 6–8). Subtilisin BPN, pancreatic elastase, leucine



**Figure 1** Scheme showing the selectivity of action of three proteinases: trypsin,  $R_1 = \text{Lys, Arg}$ ; chymotrypsin,  $R_1 = \text{Phe, Trp, Tyr}$ ; and pepsin,  $R_2 = \text{Phe, Trp, Tyr, Leu, Asp, Glu}$ .

aminopeptidase (cytosol), and other alkaline proteinases work best at pH > 8.

Many proteinases cleave not only peptide bonds but also ester bonds. Determinations of trypsin, chymotrypsin, papain, and some other enzymes are based on this property. The standard substrate for activity determination of trypsin and papain is *N*-benzoyl-L-arginine ethyl ester (BAEE), and for chymotrypsin, *N*-benzoyl-L-tyrosine ethyl ester. For example:



The reaction may be monitored by spectrophotometry or microcalorimetry.

Some proteinases use high relative molecular mass substrates as standards. For example, hemoglobin is a standard substrate for pepsin, angiotensinogen is a substrate for renin, elastin for elastase, and casein for chymosin.

### Determination of Xylanase

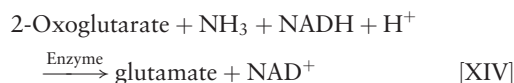
Soluble oat spelt xylan is used as the substrate for xylanase activity. The assay is carried out for 30 min at 40°C in 100 ml potassium phosphate buffer at pH 6.0. The release of reducing sugar from the breakdown of xylan is measured using the dinitrosalicylic (DNS) reagent method. After the addition of DNS reagent, each assay should be covered and placed in a bath of boiling water for exactly 5 min, followed by rapid cooling in cold water to room temperature. The absorbance of each tube should then be determined relative to the blank at 540 nm. One unit (U) of xylanase activity is defined as the amount of enzyme that produces reducing sugars corresponding to 1 μmol of xylose per minute.

## Enzyme Determination in Pharmaceutical Products

The enzymes that are successfully applied in pharmaceuticals are shown in Table 2. Those enzymes catalyzing the conversion of low relative molecular mass substrates will be considered first. They are asparaginase, rhodanese, uricase, and trypsin. These enzymes (except for trypsin) are used to carry out the conversion of lower relative molecular mass substrates not only under laboratory conditions, but also in living systems.

The reaction catalyzed by asparaginase (glutaminase) is shown in Table 2. The ammonia released is used to monitor enzyme activity. It can be detected spectrophotometrically by the classical color analytical reaction with phenol and sodium hypochlorite, or

by using L-glutamate dehydrogenase:



Alternatively, the small differences in ultraviolet spectra of asparagine and aspartate may also be used for developing a spectrophotometric method of detection.

In the case of rhodanese, the rate of thiocyanate ( $\text{SCN}^-$ ) formation should be measured by usual chromogenic reactions for thiocyanate ions.

Hydrogen peroxide is one of the products of the reaction catalyzed by uricase. It may be detected spectrophotometrically by reaction with some dyes, e.g., *o*-dianisidine, or 4-aminoantipyrine + *p*-hydroxybenzene sulfonate, and the enzyme peroxidase. The reaction with *o*-dianisidine has been described above.

High relative molecular mass substrates are the standards for collagenase, hyaluronidase, lysozyme, ribonuclease, streptokinase, and urokinase. It should be emphasized that small variations in the quality of 'high relative molecular mass standards' or conditions can result in sharp changes in the apparent composition of reaction products and can lead to nonreproducible results.

Simple, low relative molecular mass substrates are needed for detecting enzyme activity to obtain reproducible results. For example, the method for collagenase activity determination using furylacryloyl-Leu-Gly-Pro-Ala (FALGPA) is preferable to the method using collagen as a substrate. First, low relative molecular mass substrates are more selective with respect to different enzymes (e.g., FALGPA, unlike collagen, is not hydrolyzed by nonspecific proteases, which often exist in collagenase preparations). Second, the value of the enzyme activity measured by low relative molecular mass standards has an exact chemical meaning (e.g., the activity in hydrolyzing the Leu-Gly bond in FALGPA or in 4-phenylazobenzyloxycarbonyl-Pro-Leu-Gly-Pro-D-Arg).

The application of low relative molecular mass substrates (e.g., for trypsin) allows the assignment of enzyme activity values not only in the units of absorbance change at 253 nm ( $\Delta A_{253}$ , see Table 2), but also in the units micromoles of substrate per minute (hydrolysis of a particular bond); this second value can also be obtained from microcalorimetric determination.

This should not be interpreted as a call for the immediate rejection of high relative molecular mass substrates for enzyme activity determination. In all likelihood, two types of standards (low and high relative molecular masses) may be widely used in the near future. However, it is likely that the low relative



**Table 2** Application of enzymes as therapeutic agents

Enzyme	Reaction catalyzed	Therapeutic application	Method for detection of activity	Definition of activity unit <sup>a</sup>
Asparaginase (glutaminase)	$\text{L-Asparagine}^{+ -} (\text{L-glutamine}^{+ -}) + \text{H}_2\text{O} \rightarrow \text{L-aspartate}^{-} (\text{L-glutamate}^{-}) + \text{NH}_4^{+}$	Leukemia	Spectrophotometry, potentiometry (ammonia-selective electrode), or microcalorimetry	One unit liberates 1.0 $\mu\text{mol}$ of ammonia from L-asparagine (or L-glutamine) per min at pH 8.6 and 37°C
Collagenase	Hydrolysis of peptide bonds in spiral portions of collagen molecules	Skin ulcers	Spectrophotometry	<i>Chromatographically pure</i> : one unit liberates peptides from collagen equivalent in ninhydrin color to 1.0 $\mu\text{mol}$ of leucine in 5 h at pH 7.4 and 37°C in the presence of calcium ions <i>Crude; nonspecific proteases and clostripain-containing</i> : one unit hydrolyzes 1.0 $\mu\text{mol}$ of furylacryloyl-Leu-Gly-Pro-Ala per min at 25°C at pH 7.5 in the presence of calcium ions. This substrate is specific for collagenase and is not hydrolyzed by proteases
Hyaluronidase	Hydrolysis of 1,4 bonds between 2-acetamido-2-deoxy- $\beta$ -D-glucose and D-glucuronic acid in hyaluronic acid. Enzyme also hydrolyzes 1,4-bonds in some other polysaccharides containing sulfate groups	Heart attack	Spectrophotometry	One unit produces absorbance equivalent to 1.0 $\mu\text{g}$ of glucuronic acid per hour from hyaluronic acid, using 3,5-dinitrosalicylic acid to develop the color.
Lysozyme	Cell-wall hydrolysis of some bacteria. Hydrolysis of 1,4 bonds between 2-acetamido-2-deoxy- $\beta$ -D-glucose and N-acetylmuramic acid	Antibiotic	Spectrophotometry	One unit produces a change in absorbance at 450 nm of 0.001 per min at pH 6.2 and 25°C, using a suspension of <i>Micrococcus lysodeikticus</i> as substrate, in a 2.6 ml reaction volume (1 cm light path)
Rhodanese	$\text{S}_2\text{O}_3^{2-} + \text{CN}^{-} \rightarrow \text{SO}_3^{2-} + \text{SCN}^{-}$	Cyanide poisoning	Spectrophotometry	One unit converts 1.0 $\mu\text{mol}$ of cyanide to thiocyanate per min at pH 8.6 and 25°C
Ribonuclease	Hydrolysis of pyrophosphate bonds in ribonucleic acids	Antiviral	Spectrophotometry	One unit produces soluble oligonucleotides equivalent to a change in absorbance at 260 nm of 1.0 in 30 min at pH 7.5 and 37°C in a 1.5 ml reaction volume
Streptokinase + urokinase	Conversion plasminogen to plasmin	Blood clots	Spectrophotometry	<i>Streptokinase</i> : one unit will liquefy a standard clot of fibrinogen, plasminogen, and thrombin at pH 7.5 and 37°C in 10 min <i>Urokinase</i> : one unit activates the amount of plasminogen which produces a change of absorbance of 1.0 per ml per min at pH 7.5 and 37°C when measuring perchloric acid-soluble products from $\alpha$ -casein (1 cm light path)
Trypsin	Protein hydrolysis (for details see industrial section above)	Inflammation	Spectrophotometry or microcalorimetry	One BAEE ( <i>N</i> -benzoyl-L-arginine ethyl ester) unit is an absorbance change at 253 nm of 0.001 per min with BAEE as substrate at pH 7.6 and 25°C
Urease	$\text{Uric acid} + 2\text{H}_2\text{O} + \text{O}_2 \rightarrow \text{Allantoin} + \text{CO}_2 + \text{H}_2\text{O}_2$	Gout	Spectrophotometry	One unit converts 1.0 $\mu\text{mol}$ of uric acid to allantoin per min at pH 8.5 and 25°C

<sup>a</sup>Definition adopted from Sigma Chemical Co. (1992) Biochemicals, Organic Compounds for Research and Diagnostic Reagents. St. Louis: Sigma Chemical Co.



molecular mass standards will be more useful in the future.

See also: **Enzymes:** Overview; Immobilized Enzymes; Enzyme-Based Electrodes; Physiological Samples; Enzyme-Based Assays.

## Further Reading

Biocatalysis – Current Applications and Best Practice. Technology Report August 2003. Pro-Bio Faraday Partnership.

Chaplin MF and Buche C (1990) *Enzyme Technology*. Cambridge: Cambridge University Press.

Dordick JS (1991) *Biocatalysts for Industry*. New York: Plenum Press.

Godfrey T and West S (1997) *Industrial Enzymology*, 2nd edn., London: Macmillan.

Rekharsky MV and Egorov AM (1992) *Thermodynamics of Biotechnological Processes*. Moscow: Moscow University Press (in Russian).

Schmid A, Dordick JS, *et al.* (2001) Industrial biocatalysis today and tomorrow. *Nature* 409: 258–268.

## Enzyme-Based Assays

**K Matsumoto**, Kyushu University, Fukuoka, Japan

**H Ukeda**, Kochi University, Nankoku, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

## Principle

Enzymatic analysis involves the determination of sample constituents, which can be both substrates and inhibitors of enzymes, and the determination of enzyme activity occurring in the sample. Such analyses are carried out on a wide variety of materials, particularly foodstuffs and their precursor raw materials.

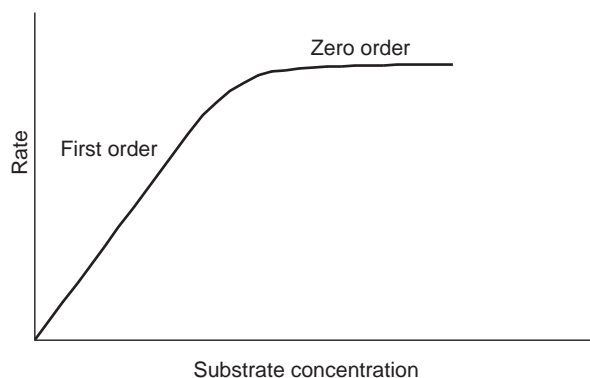
Figure 1 illustrates the dependence of an enzyme-catalyzed reaction rate on the substrate concentration. An enzyme is characterized by the number of molecules of substrate that it can convert to the product; that is, turnover number. As long as the substrate concentration is small enough with respect to the enzyme concentration that the turnover number is not exceeded, the reaction rate is directly proportional to the substrate concentration; that is, it is first order with respect to substrate. If the enzyme concentration is held constant, then the overall

reaction is first order and directly proportional to substrate concentration. This serves as the basis for substrate determination.

When the amount of substrate exceeds the turnover number corresponding to the amount of the enzyme present, the enzyme becomes saturated with respect to the number of molecules that it can complex and the reaction rate reaches a maximum value. If the enzyme is saturated with respect to substrate, then the overall reaction is first order with respect to enzyme concentration. Since a linear relationship between reaction rate and enzyme activity will exist, this becomes the basis for enzyme determination.

## Enzyme Activity Determination

In the determination of the substrate concentration, the accuracy of the results obtained by a method can be checked by comparison with a weighed-in standard. In contrast, the accuracy of the determination of the catalytic activities of enzymes can be achieved only on the basis of defined conditions of measurement. The weighing in of a crystalline enzyme with a definite catalytic activity would by no means lead to identical, accurate results under different conditions of measurement. The measurement conditions refer to the temperature, the type of the buffer, its concentration, the pH, co-factors, activators, substrate(s), and their concentration(s). For example, the temperature coefficient of the reaction rate can be as large as 10% per degree centigrade or more. This means that for a temperature rise of 1°C, the value found for the catalytic activity of an enzyme is ~10% higher. The substrate should be in excess so that the reaction rate depends only on the enzyme concentration. In order to achieve a value of  $0.99V_{\max}$  (99% of a maximum reaction rate),  $99K_m$  (99-fold substrate concentration



**Figure 1** Dependence of enzyme-catalyzed reaction rate on substrate concentration.

molecular mass standards will be more useful in the future.

See also: **Enzymes:** Overview; Immobilized Enzymes; Enzyme-Based Electrodes; Physiological Samples; Enzyme-Based Assays.

## Further Reading

Biocatalysis – Current Applications and Best Practice. Technology Report August 2003. Pro-Bio Faraday Partnership.

Chaplin MF and Buche C (1990) *Enzyme Technology*. Cambridge: Cambridge University Press.

Dordick JS (1991) *Biocatalysts for Industry*. New York: Plenum Press.

Godfrey T and West S (1997) *Industrial Enzymology*, 2nd edn., London: Macmillan.

Rekharsky MV and Egorov AM (1992) *Thermodynamics of Biotechnological Processes*. Moscow: Moscow University Press (in Russian).

Schmid A, Dordick JS, *et al.* (2001) Industrial biocatalysis today and tomorrow. *Nature* 409: 258–268.

## Enzyme-Based Assays

**K Matsumoto**, Kyushu University, Fukuoka, Japan

**H Ukeda**, Kochi University, Nankoku, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

## Principle

Enzymatic analysis involves the determination of sample constituents, which can be both substrates and inhibitors of enzymes, and the determination of enzyme activity occurring in the sample. Such analyses are carried out on a wide variety of materials, particularly foodstuffs and their precursor raw materials.

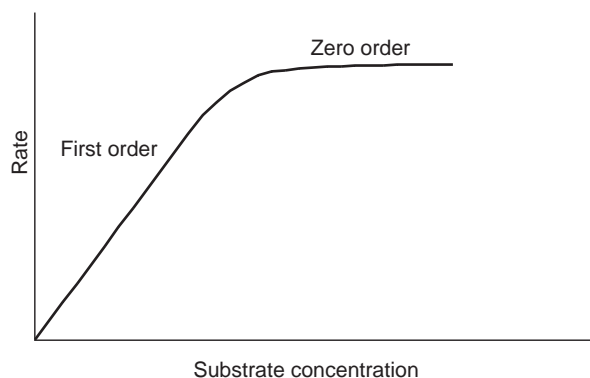
Figure 1 illustrates the dependence of an enzyme-catalyzed reaction rate on the substrate concentration. An enzyme is characterized by the number of molecules of substrate that it can convert to the product; that is, turnover number. As long as the substrate concentration is small enough with respect to the enzyme concentration that the turnover number is not exceeded, the reaction rate is directly proportional to the substrate concentration; that is, it is first order with respect to substrate. If the enzyme concentration is held constant, then the overall

reaction is first order and directly proportional to substrate concentration. This serves as the basis for substrate determination.

When the amount of substrate exceeds the turnover number corresponding to the amount of the enzyme present, the enzyme becomes saturated with respect to the number of molecules that it can complex and the reaction rate reaches a maximum value. If the enzyme is saturated with respect to substrate, then the overall reaction is first order with respect to enzyme concentration. Since a linear relationship between reaction rate and enzyme activity will exist, this becomes the basis for enzyme determination.

## Enzyme Activity Determination

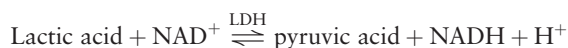
In the determination of the substrate concentration, the accuracy of the results obtained by a method can be checked by comparison with a weighed-in standard. In contrast, the accuracy of the determination of the catalytic activities of enzymes can be achieved only on the basis of defined conditions of measurement. The weighing in of a crystalline enzyme with a definite catalytic activity would by no means lead to identical, accurate results under different conditions of measurement. The measurement conditions refer to the temperature, the type of the buffer, its concentration, the pH, co-factors, activators, substrate(s), and their concentration(s). For example, the temperature coefficient of the reaction rate can be as large as 10% per degree centigrade or more. This means that for a temperature rise of 1°C, the value found for the catalytic activity of an enzyme is ~10% higher. The substrate should be in excess so that the reaction rate depends only on the enzyme concentration. In order to achieve a value of  $0.99V_{\max}$  (99% of a maximum reaction rate),  $99K_m$  (99-fold substrate concentration



**Figure 1** Dependence of enzyme-catalyzed reaction rate on substrate concentration.

of Michaelis constant) is required theoretically. Similarly, for 98, 97, 96, and 95% of  $V_{\max}$ , the substrate concentrations can be calculated as  $49K_m$ ,  $32.3K_m$ ,  $24K_m$ , and  $19K_m$ , respectively. The optimum substrate concentration is practically selected by technical conditions (solubility of the substrate, absorption of light, and so on) or the properties of the enzyme (inhibition by substrates or products).

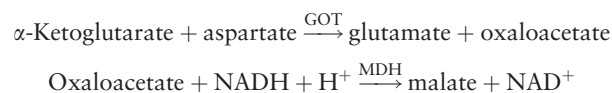
Spectrophotometric methods are most widely employed to monitor enzyme reactions. The reaction product may have an absorption spectrum different from the substrate, allowing simple measurement of the product or substrate. The typical example is the determination of dehydrogenase activity. The co-enzyme of dehydrogenases, nicotinamide adenine dinucleotide, exhibits marked differences in ultraviolet absorption spectrum between the reduced (NADH) and oxidized ( $\text{NAD}^+$ ) forms (Figure 2).  $\text{NAD}^+$  has a negligible absorption at 340 nm, while NADH has a characteristic absorption maximum. Thus, the absorption is widely used for following the dehydrogenase reactions. The example is the determination of the enzyme lactate dehydrogenase (LDH).  $\text{NAD}^+$  is required in the LDH-catalyzed oxidation of lactic acid to pyruvic acid:



The reaction is reversible and can be employed in either direction. In the forward reaction, a sample containing an unknown amount of LDH would be added to a solution containing lactic acid and  $\text{NAD}^+$  at the enzyme saturating concentrations, and the increase in absorbance at 340 nm would be measured as a function of time.

This convenient spectrophotometric property of NADH can be useful for monitoring enzyme reactions in which it is not directly involved by using it as

a coupling agent in a secondary reaction with the product. For example, glutamate-oxalate transaminase (GOT) catalyzes the reaction of  $\alpha$ -ketoglutarate and aspartate to glutamate and oxaloacetate and the product is reduced by NADH in the presence of another enzyme, malate dehydrogenase (MDH):



The second reaction should be fast compared to the first in the presence of an excess of MDH, and so the rate of decrease in NADH concentration is directly proportional to the GOT activity.

## Substrate Determination

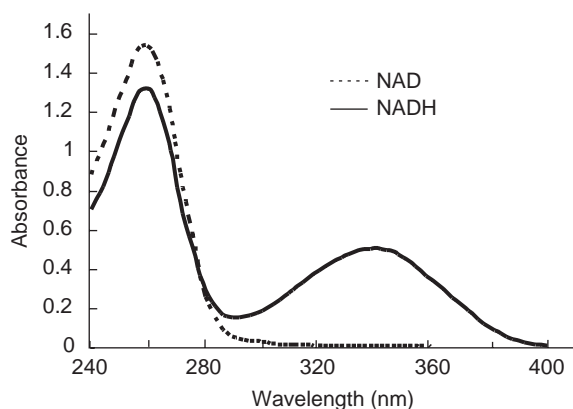
Two general techniques, endpoint method and kinetic method, can be used for measuring substrates. The endpoint method utilizes the completion conversion of the substrate. Before and following completion of the enzymatically catalyzed reaction, a product is analyzed or the depletion of a reactant that was originally in excess is measured. The analyzed substrate (net change) is then related to the original substrate concentration. These reactions are often not stoichiometric with respect to the substrate concentration because of possible side reactions or instability of products or reactants. Also, the reaction may require extraordinarily long times for completion. For these reasons, the analytical procedure is usually standardized by preparing a calibration curve of some type in which the measured quantity is related to known concentrations or quantities of the substrate.

The kinetic method employed for substrate determination is the measurement of the rate of an enzymatically catalyzed reaction, as is used to determine enzyme activity. Rate methods are generally more rapid than the endpoint method. Complete conversion reactions, on the other hand, are less subject to interference from enzyme inhibitors or activators as long as sufficient time is allowed for completion conversion.

## Endpoint Method

### One-Step Reaction

A substrate S is to be determined. S shall be completely converted into the product P in an enzymatic reaction. If S has a characteristic light absorption in the ultraviolet or visible range different from P, S can be directly determined even in the



**Figure 2** Absorbance of NAD and NADH as a function of wavelength. NADP and NADPH show nearly the same curves.

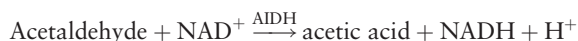
presence of other absorbing substances in the spectrophotometer cuvette. The absorbance decreases by an amount corresponding to the quantity of S converted. Various measuring techniques can be used as well as absorption photometry, such as fluorimetry, luminometry, calorimetry, and potentiometry. If dissolved oxygen is consumed by the reaction, especially such as oxidase reaction, the oxygen consumed is monitored by a Clark-type oxygen electrode amperometrically. This measuring technique is advantageous for colored or turbid sample solution.

The most widely used enzymatic reactions are those with NAD(P)-dependent dehydrogenases. By monitoring the absorbance at 340 nm, the enzymatic conversion of the substrate can be followed directly in the photometer cuvette without influencing the chemical process. The substrate concentration can be calculated using the molar extinction ( $\epsilon_{340} = 6.3 \times 10^3 \text{ l mol}^{-1} \text{ cm}^{-1}$ ) of NADH.

All enzymatic reactions involving co-enzymes are two-step reactions. However, if one of the substrates is present in a very high concentration in relation to its Michaelis constant, two-substrate reactions can be treated kinetically as one-substrate reactions. These conditions are also desirable for endpoint determinations with co-enzymes, simply to achieve a high reaction rate. However, if NAD(P)H is the second substrate, the degree to which its concentration can be increased is limited by its high absorbance. On the basis of experience, relatively large quantities of enzyme are used with relatively little substrate, so that the reaction proceeds rapidly to completion. The absorbance should thus be easily readable (neither too low nor too high).

Generally, enzyme reactions are equilibrium reactions. If the conversion is incomplete because of an unfavorable equilibrium, a determination is only possible with shifting the equilibrium by special experimental techniques. The equilibrium can be shifted by the increase of substrate concentration, variation of pH, and by trapping agents. The equilibrium constant may be altered by several orders of magnitude. The use of trapping agents, for example, semicarbazide and hydrazine when ketones and aldehydes are reaction product, is convenient for achieving the purpose. However, this technique is not always useful. Trapping agents can inhibit the catalyzing enzyme, as in the case of 3-hydroxybutyrate dehydrogenase (EC 1.1.1.30). Semicarbazide and hydrazine react slowly with NAD to form a compound that absorbs in the longwave ultraviolet and overlaps the absorption spectrum of NADH. The more elegant way is to use a trapping enzyme reaction, whose equilibrium is far

in favor of the reaction products. Total substrate conversion can be achieved by coupling such a trapping reaction with the primary reaction. Moreover, if the trapping reaction itself is catalyzed by an NAD(P)-dependent enzyme, in many cases the measuring signal is doubled, as in the case of the determination of ethanol using alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (AIDH):



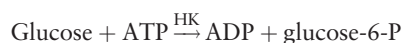
The equilibrium of the alcohol dehydrogenase reaction is far in favor of ethanol; however, ethanol conversion can be completed by continuous oxidation of acetaldehyde dehydrogenase reaction.

### Coupled Reaction

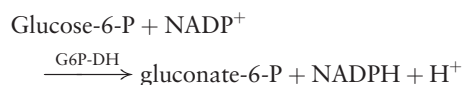
In some enzyme reactions, none of the reactants or products can be readily monitored by physical and chemical measurements. In such cases, it is often possible to determine one of these components enzymatically. This considerably increases the number of substances that can be determined by these methods.

The reaction in which the substance to be determined is transformed is known as the auxiliary, while the reaction used for the actual measurement is called the indicator reaction. Both reactions can generally be carried out in a single assay mixture (coupled reactions). If a product of the auxiliary reaction is measured in the indicator reaction, one speaks of a succeeding indicator reaction. In this case, the function of the auxiliary enzyme is to convert the substance to be analyzed into a substance that can be measured in the indicator reaction. An example is the determination of glucose using hexokinase (HK) and glucose-6-phosphate dehydrogenase (G6P-DH).

Auxiliary reaction:

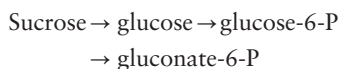


Indicator reaction:

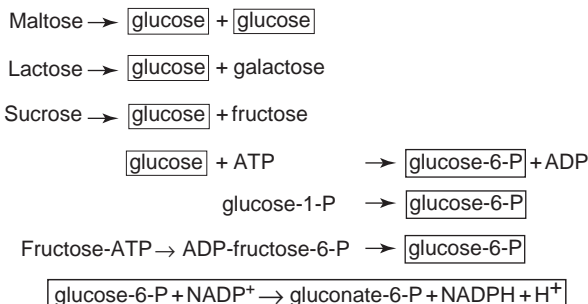


In the auxiliary reaction, glucose is phosphorylated by ATP to glucose-6-phosphate (glucose-6-P), which is the substrate of the NADP-dependent indicator reaction.

Several auxiliary reactions may also be carried out in succession, for example, in the sequence



If different auxiliary reactions give rise to the same reaction product such as the analysis of sugars, only one indicator reaction is required:



These reactions finally yield glucose or glucose-6-phosphate, and therefore all of these compounds can be determined by the glucose-6-phosphate dehydrogenase as indicator reaction. These assay methods have attracted much interest in the analysis of foods.

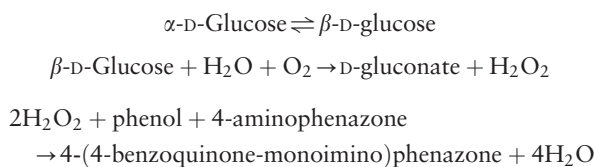
## Kinetic Method

The second technique employed for substrate determination is the measurement of the rate of an enzymatically catalyzed reaction, as is used to determine enzyme activity. The kinetic assay may be successfully used for determination of substances that do not react in a manner that can be detected directly, or in cases where the quantity of the substrate is too low to measure with the simple endpoint method. The determination of substrate using kinetic methods is possible only when the substrate concentration is low enough compared with the Michaelis constant. Thus, the following is required to perform the assay:

1. For a two-substrate reaction, the concentration of the second reactant must be so high that the rate of reaction depends only on the concentration of the substrate that is being analyzed.
2. Enzymes with high Michaelis constants are required; this enables relatively high substrate concentrations to be determined.
3. If enzymes with high Michaelis constants are not available, the apparent  $K_m$  is increased by using competitive inhibitors.

Typical examples of the kinetic determination are the determination of glucose with the use of pseudo-first-order reactions. Glucose can be determined by

glucose oxidase and Trinder's color reagent according to the following reaction scheme:



Glucose oxidase oxidizes only the  $\beta$ -form of glucose. However, this kinetic method permits the determination of the total amount of glucose present in the solution as a mixture of isomers. The  $\alpha$ - and  $\beta$ -forms of glucose undergo spontaneous interconversion by mutarotation. By choosing appropriate conditions of measurement, one can ensure that the rate of the overall reaction will be determined first by the conversion of the  $\beta$ -form already present in the solution by glucose oxidase, and then by the subsequent production of the  $\beta$ -form by mutarotation of  $\alpha$ -form. Since the Michaelis constant of glucose oxidase, with a value of  $7 \times 10^{-2} \text{ mol l}^{-1}$ , is always substantially greater than the glucose concentrations encountered in assay mixture in practice ( $\sim 8 \times 10^{-6}$  to  $5 \times 10^{-4} \text{ mol l}^{-1}$ ), the oxidation of glucose by glucose oxidase obeys pseudo-first-order kinetics with respect to  $\beta$ -glucose. The kinetics of mutarotation is of the same order. This is because here only the concentration of the glucose isomer changes, while the concentration of water as reactant remains practically constant. This results in a sequence of coupled reactions that are pseudo-first-order in glucose concentration, and to which one can therefore apply the principle of the kinetic 'fixed-time' measurement to determine the glucose concentration. The parameter measured in this method is the change in the absorbance of the dye formed during a fixed time interval.

## Activator and Inhibitor Determinations

Enzymes are involved in many of the chemical reactions that take place in biological processes. Enzymes generally have high specificity for a given substrate and do not attack closely related substrate molecules. This selectivity of enzyme and their ability to catalyze reactions of substrate at low concentrations are of great use in chemical analysis. The activities of enzymes were influenced by many kinds of factors, for instance, pH, temperature, ionic strength, activators, and inhibitors. By use of activity changes of enzymes between the presence and absence of activator (or inhibitor), one can determine the concentration of the activator (or inhibitor). An



activator is either a substance that is required for an enzyme to be active, i.e., it converts the inactive apoenzyme to the active holoenzyme, or one that increases the efficiency of an active enzyme. An inhibitor is a substance that decreases the reaction rate, either by reacting with the enzyme to form a complex or by reacting with one of the substrates.

There are two main inhibition types that one can use for inhibitor determinations, competitive and noncompetitive inhibitions. In case of competitive inhibition, inhibitors structurally related to the substrate may be bound to the enzyme active center and compete with the substrate. Because the formation of the complex between the enzyme and the inhibitor is a reversible reaction, the inhibitor can be displaced by a high concentration of the normal substrate. In the case of noncompetitive inhibition, the inhibitor combines with the enzyme at a site that is often different from the substrate binding site, and as a result inhibits the formation of the product by the breakdown of the normal enzyme–substrate complex. It is generally not reversed by the addition of excess substrate.

Competitive and noncompetitive inhibitions change enzyme kinetics differently. A competitive inhibitor does not change  $V_{\max}$  (maximum velocity) but increases  $K_m$ ; a noncompetitive inhibitor decreases  $V_{\max}$  and  $K_m$  remains unchanged.

Because enzymes are catalysts, which affect the rate and does not affect the equilibrium or endpoint of a reaction, their concentration and activity must be measured by a kinetic method. In such a kinetic method it is not necessary to wait for equilibrium to be established.

### Assay of Activators

The effect of activator concentration on the initial rate of an enzyme reaction is similar to that of substrate concentration. At low concentrations of activator there is a first-order dependence of the rate on activator concentration. At higher concentrations the enzyme is maximally activated and the rate is independent of activator concentration. Consequently, lower concentration is preferred for assay of activators. The basic technique for determination of metal ions is the use of metalloenzymes converted to apoenzymes, the latter being reconverted to the holoenzymes by binding with the metal ions. In this case, the metal ion in the holoenzyme is first removed by adding a chelating agent such as ethylenediaminetetraacetic acid. For instance, one can determine  $\text{Ca}^{2+}$  or  $\text{Zn}^{2+}$  by use of apoalkaline phosphatase and  $\text{Co}^{2+}$  or  $\text{Cu}^{2+}$  by use of apoaminopeptidase. Many researchers have reported the activation of

isocitrate dehydrogenase by metal ions, such as  $\text{Mn}^{2+}$ ,  $\text{Mg}^{2+}$ , and  $\text{Zn}^{2+}$ :

- $\text{Ca}^{2+}$  and  $\text{K}^{+}$ :  $\text{Ca}^{2+}$  activates a number of enzyme systems: pancreatic lipase is activated by  $\text{Ca}^{2+}$  in the presence of bile and insoluble triglycerides, and bacterial 2-oxyglutarate dehydrogenase is also activated by  $\text{Ca}^{2+}$ . The apoenzyme of calf intestinal alkaline phosphatase is activated by  $\text{Ca}^{2+}$ .  $\text{K}^{+}$  activates phosphate fructokinase and pyruvate kinase, respectively.
- $\text{Mg}^{2+}$  and  $\text{Mn}^{2+}$ : Isocitrate dehydrogenase (ICDH) is activated by  $\text{Mg}^{2+}$ . The rate of the ICDH reaction is proportional to magnesium under the condition of constant amounts of ICDH and nonrate-limiting concentration of isocitrate and NADP. The formation of NADPH ( $\lambda_{\max} = 340 \text{ nm}$ ) can be monitored spectrophotometrically. Both  $\text{Mg}^{2+}$  and  $\text{Mn}^{2+}$  can activate heart pyruvate dehydrogenase and bacterial 2-oxyglutarate dehydrogenase, and can be determined.
- $\text{Zn}^{2+}$ ,  $\text{Cu}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Ni}^{2+}$ : Apoaminopeptidase was prepared by removing zinc from pig kidney amino peptidase, a commercially available metalloenzyme. By use of this apoamino peptidase, one can determine zinc concentration up to a given limit, giving the enzyme activity of the apoenzyme being strictly proportional to the concentration of zinc ions in the assay system. The analytical procedure has been extended to the assay of  $\text{Cu}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Ni}^{2+}$ . The apoenzyme from alkaline phosphatase (zinc-metalloenzyme) recovers its activity by adding zinc ions, and thus  $\text{Zn}^{2+}$  can be determined. Ascorbate oxidase from cucumber contains  $\text{Cu}^{2+}$  ions and the apoenzyme of the enzyme recovers its activity by addition of  $\text{Cu}^{2+}$ .

### Assay of Inhibitors

**Determination of organic compounds** Organophosphorus, carbamates, and chlorinated pesticides inhibit cholinesterase (acetylcholinesterase or butylcholinesterase). This inhibition of cholinesterase accounts to a large extent for the effectiveness of organophosphorus insecticides in the control of harmful insects, and for their toxicity to warm-blooded animals. The activity of cholinesterase decreases in proportion to the amount of pesticides present. The activity of cholinesterase can be assayed by measuring acetic acid produced by use of a pH stat approach or by measuring the choline produced by choline oxidase enzyme reaction (measurement of  $\text{O}_2$  decrease). For fluorimetric detection, *N*-methylindoxyl acetate is suitable for the assay of pesticides as a substrate. The substrate is hydrolyzed by cholinesterase from



**Table 1** A list of enzymes that are inhibited by various inorganic ions

Enzyme used	Ion determined by inhibition
Alcohol dehydrogenase	$\text{Ag}^+$ , $\text{Hg}^{2+}$
Alkaline phosphatase	$\text{Al}^{3+}$ , $\text{Bi}^{3+}$ , $\text{Be}^{2+}$ , $\text{Zn}^{2+}$
Glucose oxidase	$\text{Ag}^+$ , $\text{Hg}^{2+}$ , $\text{Pb}^{2+}$
Hyaluronidase	$\text{Cu}^{2+}$ , $\text{Fe}^{2+}$ , $\text{Fe}^{3+}$ , $\text{CN}^-$
Isocitrate dehydrogenase	$\text{Ag}^+$ , $\text{Al}^{3+}$ , $\text{Ce}^{3+}$ , $\text{Cd}^{2+}$ , $\text{Cu}^{2+}$ , $\text{Fe}^{2+}$ , $\text{Fe}^{3+}$ , $\text{Hg}^{2+}$ , $\text{In}^{3+}$ , $\text{Ni}^{2+}$ , $\text{Pb}^{2+}$
Peroxidase	$\text{Cd}^{2+}$ , $\text{Co}^{2+}$ , $\text{Cu}^{2+}$ , $\text{Fe}^{2+}$ , $\text{Fe}^{3+}$ , $\text{Mn}^{2+}$ , $\text{Pb}^{2+}$ , $\text{Cr}_2\text{O}_7^{2-}$ , $\text{S}^{2-}$
Urease	$\text{Ag}^+$ , $\text{Cd}^{2+}$ , $\text{Co}^{2+}$ , $\text{Cu}^{2+}$ , $\text{Hg}^{2+}$ , $\text{Mn}^{2+}$ , $\text{Ni}^{2+}$ , $\text{Pb}^{2+}$
Xanthine oxidase	$\text{Ag}^+$ , $\text{Hg}^{2+}$
Liver esterase	$\text{F}^-$

various sources to the highly fluorescent *N*-methyl-indoxyl ( $\lambda_{\text{ex}} = 430 \text{ nm}$ ;  $\lambda_{\text{em}} = 510 \text{ nm}$ ).

Dichlorodiphenyltrichloroethane (DDT) inhibits carbonic anhydrase at the amount of micrograms at which other inhibitors are inactive. Carbonic anhydrase catalyzes the hydration of carbon dioxide to carbonic acid. The activity of the enzyme is assayed by measuring the  $\text{CO}_2$  liberated by dehydration in a Warburg manometer or a  $\text{CO}_2$  gas-selective electrode. DDT inhibits the enzyme, causing a decrease in the rate of  $\text{CO}_2$  liberation.

Ascorbic acid inhibits catalase in concentrations as low as  $2 \times 10^{-6} \text{ mol l}^{-1}$ . Very low concentration of  $\text{Cu}^{2+}$  ions accelerates the rate of inhibition. The inhibition of ascorbic acid to catalase occurs in the presence of  $\text{Cu}^{2+}$  ions. The activity of catalase can be monitored by the decrease of  $\text{H}_2\text{O}_2$  substrate or increase of  $\text{O}_2$ . For fluorimetric detection of  $\text{H}_2\text{O}_2$ , the decrease in fluorescence of scopoletin or the increase in fluorescence of diacyldichlorofluorescein upon reaction with  $\text{H}_2\text{O}_2$  are preferably monitored.

**Assay of inorganic compounds** There have been many methods that are based on the inhibitions of inorganic compounds for enzyme-catalyzed reactions. A list of enzymes that are inhibited by various inorganic ions is shown in **Table 1**. Generally, the specific determination of one inorganic compound inhibitor in the presence of others is impossible. Usually, a prior separation or masking procedures are performed. Thiosulfate and cyanide are generally

used as masking agents for  $\text{Ag}^+$  or  $\text{Hg}^{2+}$ , and  $\text{Ni}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Fe}^{2+}$ , respectively. The choice of masking agent, however, will depend on interfering inorganic compounds, pH, the substrate and enzyme, and so on. The use of appropriate masking agents will allow the researcher to develop specific procedures for very low concentration of the inorganic compounds of interest.

## Enzyme Immunoassay (EIA)

Enzymes have become one of the most popular labels for immunoassay. Nowadays, enzyme-labeled antigen and antibodies were employed as alternative to radioisotope-labeled materials in immunoassay. The class of immunoassays that employ an enzyme as a label is known as enzyme immunoassays. This is described in a separate article.

*See also:* **Enzymes:** Overview; Immobilized Enzymes; Enzyme-Based Electrodes; Enzymes in Physiological Samples; Industrial Products and Processes.

## Further Reading

- Belitz H-D and Grosch W (1987) *Food Chemistry*, 2nd edn. Berlin: Springer.
- Bergmeyer HU, Bergmeyer J, and Grass M (1983) *Methods of Enzymatic Analysis*, 3rd edn, vol. 1. Weinheim: VCH.
- Christian GD (1994) *Analytical Chemistry*, 5th edn. New York: Wiley.
- Guilbault GG (1976) *Handbook of Enzymatic Methods of Analysis*. New York: Dekker.

# ESSENTIAL OILS

**R P W Scott**, Great Sanders House, Sedlescombe, UK

## Introduction

The first recorded use of essential oils was probably by the Egyptians who employed them for a range of purposes including medicinal, cosmetic, in religious

**Table 1** A list of enzymes that are inhibited by various inorganic ions

Enzyme used	Ion determined by inhibition
Alcohol dehydrogenase	$\text{Ag}^+$ , $\text{Hg}^{2+}$
Alkaline phosphatase	$\text{Al}^{3+}$ , $\text{Bi}^{3+}$ , $\text{Be}^{2+}$ , $\text{Zn}^{2+}$
Glucose oxidase	$\text{Ag}^+$ , $\text{Hg}^{2+}$ , $\text{Pb}^{2+}$
Hyaluronidase	$\text{Cu}^{2+}$ , $\text{Fe}^{2+}$ , $\text{Fe}^{3+}$ , $\text{CN}^-$
Isocitrate dehydrogenase	$\text{Ag}^+$ , $\text{Al}^{3+}$ , $\text{Ce}^{3+}$ , $\text{Cd}^{2+}$ , $\text{Cu}^{2+}$ , $\text{Fe}^{2+}$ , $\text{Fe}^{3+}$ , $\text{Hg}^{2+}$ , $\text{In}^{3+}$ , $\text{Ni}^{2+}$ , $\text{Pb}^{2+}$
Peroxidase	$\text{Cd}^{2+}$ , $\text{Co}^{2+}$ , $\text{Cu}^{2+}$ , $\text{Fe}^{2+}$ , $\text{Fe}^{3+}$ , $\text{Mn}^{2+}$ , $\text{Pb}^{2+}$ , $\text{Cr}_2\text{O}_7^{2-}$ , $\text{S}^{2-}$
Urease	$\text{Ag}^+$ , $\text{Cd}^{2+}$ , $\text{Co}^{2+}$ , $\text{Cu}^{2+}$ , $\text{Hg}^{2+}$ , $\text{Mn}^{2+}$ , $\text{Ni}^{2+}$ , $\text{Pb}^{2+}$
Xanthine oxidase	$\text{Ag}^+$ , $\text{Hg}^{2+}$
Liver esterase	$\text{F}^-$

various sources to the highly fluorescent *N*-methyl-indoxyl ( $\lambda_{\text{ex}} = 430 \text{ nm}$ ;  $\lambda_{\text{em}} = 510 \text{ nm}$ ).

Dichlorodiphenyltrichloroethane (DDT) inhibits carbonic anhydrase at the amount of micrograms at which other inhibitors are inactive. Carbonic anhydrase catalyzes the hydration of carbon dioxide to carbonic acid. The activity of the enzyme is assayed by measuring the  $\text{CO}_2$  liberated by dehydration in a Warburg manometer or a  $\text{CO}_2$  gas-selective electrode. DDT inhibits the enzyme, causing a decrease in the rate of  $\text{CO}_2$  liberation.

Ascorbic acid inhibits catalase in concentrations as low as  $2 \times 10^{-6} \text{ mol l}^{-1}$ . Very low concentration of  $\text{Cu}^{2+}$  ions accelerates the rate of inhibition. The inhibition of ascorbic acid to catalase occurs in the presence of  $\text{Cu}^{2+}$  ions. The activity of catalase can be monitored by the decrease of  $\text{H}_2\text{O}_2$  substrate or increase of  $\text{O}_2$ . For fluorimetric detection of  $\text{H}_2\text{O}_2$ , the decrease in fluorescence of scopoletin or the increase in fluorescence of diacyldichlorofluorescein upon reaction with  $\text{H}_2\text{O}_2$  are preferably monitored.

**Assay of inorganic compounds** There have been many methods that are based on the inhibitions of inorganic compounds for enzyme-catalyzed reactions. A list of enzymes that are inhibited by various inorganic ions is shown in **Table 1**. Generally, the specific determination of one inorganic compound inhibitor in the presence of others is impossible. Usually, a prior separation or masking procedures are performed. Thiosulfate and cyanide are generally

used as masking agents for  $\text{Ag}^+$  or  $\text{Hg}^{2+}$ , and  $\text{Ni}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Fe}^{2+}$ , respectively. The choice of masking agent, however, will depend on interfering inorganic compounds, pH, the substrate and enzyme, and so on. The use of appropriate masking agents will allow the researcher to develop specific procedures for very low concentration of the inorganic compounds of interest.

## Enzyme Immunoassay (EIA)

Enzymes have become one of the most popular labels for immunoassay. Nowadays, enzyme-labeled antigen and antibodies were employed as alternative to radioisotope-labeled materials in immunoassay. The class of immunoassays that employ an enzyme as a label is known as enzyme immunoassays. This is described in a separate article.

*See also:* **Enzymes:** Overview; Immobilized Enzymes; Enzyme-Based Electrodes; Enzymes in Physiological Samples; Industrial Products and Processes.

## Further Reading

- Belitz H-D and Grosch W (1987) *Food Chemistry*, 2nd edn. Berlin: Springer.
- Bergmeyer HU, Bergmeyer J, and Grass M (1983) *Methods of Enzymatic Analysis*, 3rd edn, vol. 1. Weinheim: VCH.
- Christian GD (1994) *Analytical Chemistry*, 5th edn. New York: Wiley.
- Guilbault GG (1976) *Handbook of Enzymatic Methods of Analysis*. New York: Dekker.

# ESSENTIAL OILS

R P W Scott, Great Sanders House, Sedlescombe, UK

## Introduction

The first recorded use of essential oils was probably by the Egyptians who employed them for a range of purposes including medicinal, cosmetic, in religious

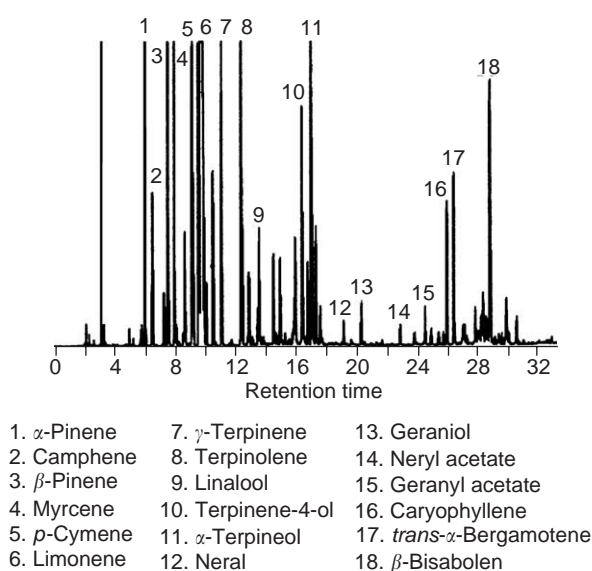
ceremonies, and for embalming the dead. About the same time, the Chinese were using herbs and aromatic plant extracts for medicinal purposes and their procedures were, eventually, assimilated into Indian Ayurvedic medicine. After the fall of the Egyptian empire, the ancient Greeks assumed much of the Egyptian knowledge and essential oils were employed by Hippocrates in aromatherapy. In due course, Greek medicine was taken up by the Romans who were impressed by the power of fragrances and also used them enthusiastically in aromatherapy. Subsequent to the dissolution of the Roman Empire and the resurgence of the Arabian empire, the use of essential oils was continued in the manner of the Romans and during this period an efficient distillation process was developed for their extraction.

During the dark ages, the monks in their monasteries persisted with the use of herbs for medicinal purposes and, although they did not recognize the nature of their activity as such, they used the antibacterial and antipesticide properties of many essential oils to cleanse wounds and control infestation. During the renaissance, the holistic approach to medicine, pioneered by Hippocrates, came back into vogue and Paracelsus used herbs (the knowledge of which largely came from folk remedies) to effect a cure for leprosy. The term aromatherapy was introduced by the French perfumer Gattefosse and appeared to have arisen from a personal experience after burning his hand in the laboratory. He treated his hand with pure lavender oil, which, apparently, immediately eased the pain and the burn subsequently healed with no infection and little or no scarring. This technique was subsequently used with great success in treating soldiers wounded in the First World War. Today, essential oils are widely used as perfumes, food flavors, and for medicinal purposes, but many of their bactericidal and antipesticide uses have been replaced by antibiotics and synthetic antipesticide and antifungal agents.

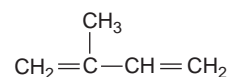
## The Chemical Nature of Essential Oils

Essential oils are composed of a very complex mixture of hundreds (sometimes thousands) of different substances, many often chemically related to one another. An indication of the complexity of an essential oil is afforded by the chromatogram of lime oil shown in **Figure 1**.

Only the major components are labeled in **Figure 1**, but the physiological activity of an oil (be it odor, bactericidal, fungicidal, etc.) is the net synergistic effect of all components including those present at the trace level. A synthetic lavender oil can be made up



**Figure 1** A chromatogram of lime oil (Superco, Inc.).



**Figure 2** The structure of isoprene.

from the major components, lavandol, borneol, terpineol, geraneol, and linalool mixed in the correct proportion to that of the natural oil. However, it will be found that the resulting mixture will not have nearly the same desirable effect on burns as the natural oil.

Many of the components of an essential oil are multiple derivatives of the hydrocarbon isoprene, the structure of which is shown in **Figure 2**.

Isoprene is a hydrocarbon and is the basic building block of many essential oils, either in the form of a terpene or in an oxygenated form (e.g., alcohol, aldehyde, etc.). The components of an essential oil can be roughly classed into the following types: terpenes (monoterpenes and sesquiterpenes), oxygenated compounds, alcohols, phenols, aldehydes, ketones, esters, lactones, coumarins, ethers, and oxides.

### Terpene Hydrocarbons

**Monoterpenes** A terpene molecule contains 10 carbon atoms (derived from two isoprene units) and at least one double bond. Terpene hydrocarbons are thermally labile and easily oxidized and, thus, citrus oils, which contain a high level of terpenes, do not keep well. Some terpenes are thought to have

anti-inflammatory, antiseptic, antiviral, and antibacterial properties, but, overall, the effect of the essential oil is a combination of the physiological activity of all the components. Some terpenes are considered to have analgesic properties or can act as stimulants. Some are said to stimulate mucous secretion and are consequently effective decongestants. It is clear that over the ages essential oils have been used extensively in medicine.

**Sesquiterpenes** Sesquiterpenes normally contain 15 carbon atoms and can have very complex pharmacological activity. An example of a sesquiterpene, farnesene, is shown in **Figure 3**. Farnesene is often found in chamomile and rose and many other floral essential oils. Farnesene has significant anti-inflammatory and antiallergy properties and, thus, in the past essential oils containing farnesene were used for the same purposes that aspirin and ibuprofen are used today.

Chamazulene, the structure of which is also shown in **Figure 3**, has similar pharmacological properties and is found in German chamomile.

### Terpene Alcohols

**Monoterpene alcohols** Examples of these types of essential oil component are linalool, citronellol, and terpinol; the structure of the latter is shown in **Figure 4**.

These substances are commonly found in lavender, rose, geranium, and juniper oil. The monoterpene alcohols have antiseptic, antiviral, and antifungal

properties and are claimed to give an uplifting or energizing impact on the user.

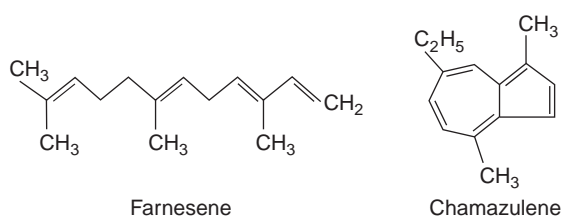
**Sesquiterpene alcohols** Sesquiterpene alcohols are not commonly found in essential oils but, like *bisabolol* (found in the German chamomile) are anti-allergen, anti-inflammatory, and can also act as liver and glandular stimulants. Other essential oils that contain sesquiterpene alcohols are sandalwood oil (the active alcohol being  $\alpha$ -sanatol, the structure of which is included in **Figure 4**), oil of ginger, patchouli oil, vetiver oil, and carrot seed oil.

**Aldehydes** It is the aldehydes present in melisa, lemon grass, and citronella oils that give them their citrus-like fragrance. They have antifungal, anti-inflammatory, and disinfectant properties together with some sedative effects. They are usually employed in aromatherapy at less than 1% levels contained in a neutral solvent; as at higher concentrations they can cause skin irritation. Aldehydes are generally unstable as they are easily oxidized and tend to be thermally labile; as a consequence, the oils must be stored with care.

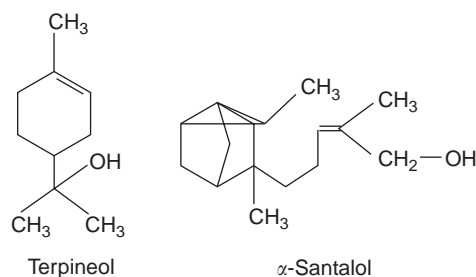
**Ketones** In general, the ketones found in essential oils can be very toxic (e.g., thujone, the structure of which is shown in **Figure 5**, and pinopamphone). However, at appropriate concentrations they have useful therapeutic properties, promoting the secretion of mucous and inducing cell and tissue regeneration as well.

Hyssop, eucalyptus, and rosemary contain moderate amounts of ketones and are employed extensively in aromatherapy. The ketone italdione is used to aid skin regeneration, wound healing, and to help reduce wound scars, stretch marks, etc., as well act as a mutolytic (promoting mucous secretion).

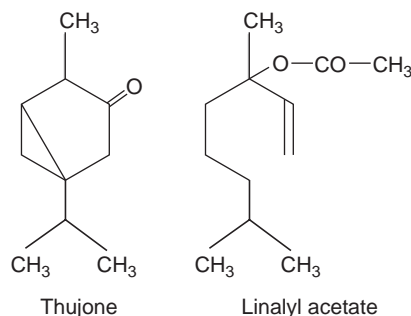
**Esters** The esters that naturally occur in essential oils usually provide the fruity and fragrant character to the oil. Some esters are reported to have antifungal



**Figure 3** The structures of farnesene and chamazulene.



**Figure 4** The structures of terpineol and  $\alpha$ -santalol.



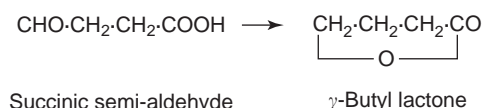
**Figure 5** The structures of thujone and linalyl acetate.

and antimicrobial activity as shown by geranium oil. Others have been shown to act as sedatives and antispasmodics. An example of an ester commonly present in many essential oils is linalyl acetate, the structure of which is also included in **Figure 5**. Most of the therapeutic effects are relatively mild and such esters can be used in moderation with some confidence.

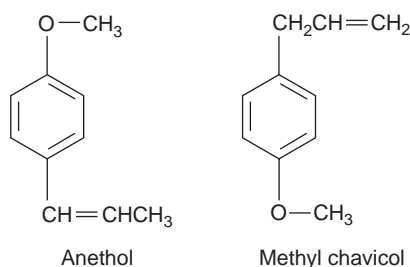
**Ethers** There are very few ethers commonly found in essential oils and those that do are phenolic in nature. The two most important are anethol, which is found in aniseed, and methyl chavicol, which is found in basil and tarragon. The structures of anethol and methyl chavicol are shown in **Figure 6**.

**Oxides** The oxides in essential oils, e.g., 1,8-cineol (eucalyptol) in eucalyptus oil, are largely responsible for the expectorant characteristics of such oils. The structure of eucalyptol is given in **Figure 7**.

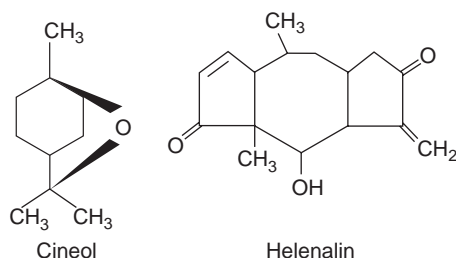
**Lactones and coumarins** Lactones are the result of molecular ring formation involving an ester group by the elimination of water, an example of which is the formation of  $\gamma$ -butyl lactone given below:



The amount of lactones and coumarins contained in essential oils is relatively small. Many can act as



**Figure 6** The structures of anethol and methyl chavicol.



**Figure 7** The structures of cineol and helenalin.

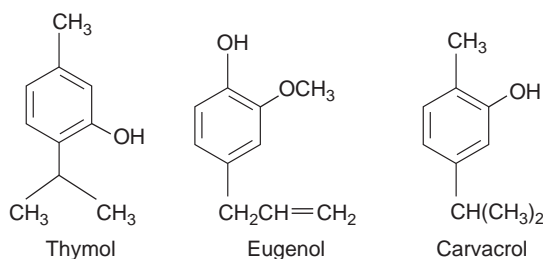
sedatives and antispasmodics and some lactones, coumarins, and ketones exhibit neurotoxic effects and can cause skin sensitizing and irritation. Helenaline (the structure of which is included in **Figure 7**), is a component of arnica oil and is probably the agent present in the oil that is responsible for its anti-inflammatory properties. Some coumarins (e.g., furocoumarin and bergaptene that are found in bergamot oil) can be phototoxic and, thus, if used topically, great care should be taken to avoid the treated area from being exposed to bright sunlight.

**Phenols** Phenols that occur in essential oils usually contain an aliphatic side chain; for example, thymol, eugenol, and carvacrol, whose structures are shown in **Figure 8**.

Phenols are inherently antiseptic, antibacterial, and, thus, act as disinfectants. Consequently, those essential oils that have high phenol content must be used at low concentrations and over relatively short periods of time. As they have good antiseptic properties, phenols are also skin and mucous membrane irritants, examples of which are cinnamon and clove oil.

## Raw Materials

Essential oils can be stored in many different parts of a plant. The different botanical sources of essential oils together with examples are summarized in **Table 1**.



**Figure 8** The structures of thymol, eugenol, and carvacrol.

**Table 1** Storage sources of essential oils

<i>Morphological site</i>	<i>Essential oil</i>
Buds	Black current, clove
Flowers	Hyacinth, jasmine, lavender, rose, ylang-ylang
Fruits	Citrus, juniper berry, vanilla
Seeds	Anis, cardamom, coriander, nutmeg
Leaves	Buchu, geranium, lavender, patchoili, peppermint
Bark	Cinnamon
Wood	Cederwood, sandalwood
Roots, rhizomes	Angelica, orris, vetiver
Resins	Galbanum, laudanum, myrrh, olibaum



The actual oils are stored in specific plant organs from which they need to be released and then extracted. There are three plant organs in which oils are stored and these are glands, oil cells, and cavities.

### Glands

Glands are the name given to cellular excrescences on some surface area of the plant that holds the essential oil. Examples of plants that have these are thyme, marjoram, rosemary, sage, and savory. It is interesting to note that all these plants are members of the Lamiaceae family.

### Oil Cells

Oil cells are individual plant cells that are filled with components of the oil and which must be broke open to release the oil.

Examples of plants that have oil cells are laurel and cinnamon of the family Lanraceae and lemon, orange, and other citrus plants of the family Rutaceae.

### Cavities

Cavities are tubular spaces developed in plants due to cell deterioration. Cavities occur in anis, caraway, coriander, and fennel (members of the Umbelliferae family) and the needle oils of the Abies and Pinus.

## Extraction Techniques for Essential Oils

There are three general methods of extraction: distillation, expression (largely employed for the extraction of essential oils from citrus fruits), and solvent extraction. The selected process depends on the nature of the raw material more than on the nature of the essential oil. An exception would be supercritical extraction with liquid carbon dioxide, which is an ideal method of extraction but, unfortunately, extremely expensive and thus can only be used in very special circumstances. The process chosen for extraction can have a profound effect on the quality of the oil produced.

### Distillation

There are a number of different types of distillation, which, although basically the same process, vary in detail to suit specific types of raw material and specific types of oil.

**Water distillation** The raw material is completely immersed in water and the temperature raised to boiling point. The oil and the water then distill from the matrix (largely by steam distillation) but steam

*per se* is not introduced into the still. On cooling the distillate, the oil and water separate and the oil is selectively removed. One advantage of this method is that the oil is never raised above a temperature of 100°C and, thus, thermal decomposition is minimized. For thermally labile oils, the temperature can further be reduced by distilling under reduced pressure. The separated water is also a product and sold as 'floral waters', e.g., rose water, lavender water, etc. If the basic oil contains a large number of esters, they will probably break down under extended exposure to hot water by hydrolysis and, consequently, an alternative extraction method would be appropriate.

**Steam distillation** In steam distillation, hot steam is forced through the matrix of raw material, opening the cavities in which the oil is held, and volatilizing the oils. This process is, in fact, pure steam distillation. The oil separates from the water in the distillate and is collected. As the steam is under pressure, the temperature can be carefully adjusted to provide the maximum rate of extraction with minimum thermal decomposition. Lavender oil, which is thermally, labile can be extracted by steam distillation without thermal damage.

The two processes can be combined by filling the still with the raw material and water and distillation carried out, not by heating the still, but by passing steam through it. This combined process has the combined advantages and disadvantages of both techniques.

**Hydrodiffusion** Hydrodiffusion is basically a modification of steam distillation that provides a shorter processing time, a higher yield, and greater steam economy. It differs from the conventional steam distillation in that the steam is passed into the top of the container, passes through a bed of botanical material supported on a grill, and is condensed below the extraction bed. Other than the improvements mentioned above, the process has the same properties as the conventional form of steam distillation.

**Cohobation** Some oils are partially soluble in water and so when the oil separates from the water distillate a portion of the oil still remains in the aqueous fraction. This typically happens with rose oil as one of the important components is phenyl ethyl alcohol, which remains in the water after separation. Thus, phenyl ethyl alcohol and any other water-soluble components must be either distilled from the distillate or extracted by some other means and added back to the separated oil. This process has been given the name of cohobation.



## Expression

Expression techniques largely involve crushing or macerating the botanical material to release the essential oils so that they can be washed free from the plant matrix. There are a number of methods used but, basically, they all involve maceration followed by washing.

**Cold pressing** In this process, the oils are expressed under high pressure and not heated at all. This process is simple, inexpensive, and particularly suitable for oils that are thermally labile. Some of the oil is inevitably retained by the plant matrix but the amount may not be significant.

**Écuelle à Piquer extraction process** The Écuelle à Piquer process (the prick/stick/prod process) is mainly used for the extraction of citrus fruits. The fruit is placed in a spinning container that has spikes on the sides which puncture the oil cells in the fruit. The oil from the punctured cells, together with the juice and pigments, pass down the center of the device into a collection area. The liquid is removed and the oil that floats on the surface of the aqueous layer is separated by decantation.

**Machine abrasion** This method is similar to that of the Écuelle à Piquer process and is used largely for citrus fruits. The machine strips off the outer peel, which is taken away by a stream of water and then passed into a centrifugal separator. Although the separation is relatively fast the oils obtained are in contact with the broken cell contents (which contain different enzymes) for a significant time and, thus, the oil can easily be degraded enzymatically.

**The maceration process** The maceration method also includes a type of solvent extraction process. The flowers are immersed in hot oil that causes the membranes to become ruptured, releasing the essential oil. The essential oil dissolves in the hot oil and the botanical residues separated from the mixture by filtration or centrifugation and the essential oil recovered by distillation.

## Solvent Extraction

Solvent extraction is a very gentle way of isolating the essential oils from a botanical matrix. Consequently, they are employed with thermally labile materials that cannot accommodate elevated temperatures such as jasmine, hyacinth, narcissus, and tuberose. An added advantage is that in many cases solvent extraction also reduces enzymatic

degradation. The solvents commonly employed are hexane, petroleum ether, methanol, and ethanol, all of which are readily available and relatively inexpensive. A solvent-extracted oil is highly concentrated and is close to a true representation of the natural fragrance of the parent flower.

There is some criticism of the solvent-extracted oil in that it may contain high boiling contaminants present in the original solvent that may amount to between 2% and 20% (this was particularly so when benzene was used as the extracting solvent). Consequently, for aromatherapy, solvent-extracted oils may not be acceptable. However, if hexane is used as the extracting solvent, solvent residues in the essential oil can be reduced to less than  $10 \text{ mg l}^{-1}$ , which is generally acceptable.

**Supercritical carbon dioxide extraction** Supercritical carbon dioxide extraction is a relatively recent process and yields products of extremely high quality. The process is, however, relatively expensive due to high cost of liquid carbon dioxide. The critical temperature of carbon dioxide is  $31^\circ\text{C}$  and the critical pressure just over 1000 psi. Critical carbon dioxide is an excellent solvent for essential oils. Due to the relatively low temperature of the extraction, it can easily handle thermally labile oils without degradation and in addition it is chemically very inert and so does not react with any of the essential oil components. The essential oil is easily recovered from the extract by reducing the pressure in a controlled manner and allowing the carbon dioxide to evaporate. The extraction is carried out in a pressurized container constructed from heavy duty stainless steel at  $\sim 35^\circ\text{C}$  and  $\sim 1000$  psi. The equipment can also be very expensive.

**Enfleurage extraction** Enfleurage extraction involves the use of a high boiling semisolid solvent in the form of highly purified, odorless animal fat. The apparatus consists of an enclosure frame carrying glass shelves that are covered with purified lard. The freshly picked flowers or other plant material are laid on the shelves and pressed into the fat. The flowers are allowed to remain in the enclosure for a few days to allow the essential oil components to diffuse into the fat. The extracted flowers are then removed and replaced with a fresh supply.

When the concentration of essential oil in the fat has reached a maximum, the fat and essential oil are extracted with alcohol removing the oil from the fat, which is then used to make soap. The alcoholic solution is then carefully evaporated leaving the essential oil as the residue. This process is labor-intensive and obviously very costly, but is still used

for specifically extracting the essential oils of tuberose and jasmine.

## The Analysis of Essential Oils

Today, essential oils are almost exclusively analyzed by chromatographic techniques; mostly by gas chromatography, but if very high boiling components are present, by liquid chromatography. Chromatographic techniques provide very high resolution (necessary for highly complex mixtures) and very high sensitivity so that traces of materials can be separated and quantitatively measured with a fair degree of accuracy. If the chromatograph is associated with a mass spectrometer, then the trace component can also be identified. Chromatographs are expensive, but are a necessary analytical tool for the essential oil chemist.

### The Basic Gas Chromatograph

A block diagram representing the basic gas chromatograph is shown in Figure 9. The instrument consists of a number of units: the gas supplier, the flow controllers, the sample injector, the column situated in an oven, the detector (also situated in an oven), the detector electronics, and finally an appropriate display unit.

**Gas supplier** The mobile phase or carrier gas supply may consist of a single gas such as helium, or may include other gases (depending on the detector chosen), such as hydrogen and air for the flame ionization detector.

**Flow controllers** Many detectors are flow sensitive and, thus, a flow controller may be essential. The controllers are simple pneumatic devices that will maintain a constant mass flow rate independent of the column temperature or column impedance.

**The sample injector** The sample injector can be a simple septum injector and a hypodermic syringe that can place the sample directly on to the column or a high-pressure sample valve (that can be manually or mechanically operated) fitted with an appropriate sample loop.

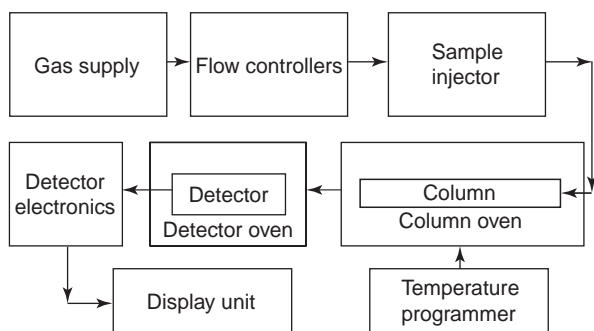
**The column and column oven** The column oven controls the column temperature and, thus, the rate at which the solutes migrate through the column.

**The temperature programmer** Increasing the column temperature increases the speed of solute migration through the column and shortens the analysis time for mixtures that contain solutes that extend over a wide range of molecular weight or polarity. The temperature of the column must be continuously increased during chromatographic development in a carefully controlled manner. This device can increase the temperature of the oven at a chosen rate and holds the oven at a specified temperature for a defined time before returning to a preset starting temperature.

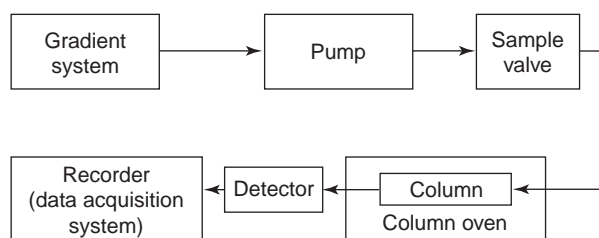
**The detector and detector electronics** The detector and its electronic processor provide an electrical response that is linearly related to solute concentration as it leaves the column. Thus, the elution time can be measured and the concentration profile of an eluted solute can be monitored. The most common detectors are the flame ionization detector, the katharometer, and the electron capture detector.

**The display unit** The display units can be a potentiometric recorder, a computer printer, or, for very fast separations, a cathode ray tube. Modern instruments digitize the detector signal and pass it to a computer where the data are stored and processed and the results printed out directly on the printer. There are, however, many laboratories that still use potentiometric recorders and manually process their data.

**The analytical procedure** In a gas chromatographic analysis, the column temperature and temperature program is set appropriately for the sample to be analyzed. A solution of the sample is made up in a volatile solvent that will elute very early in the chromatogram and well away from the solutes of interest, the concentration usually being between 1% and 5% v/v. The sample (normally between 0.5 and 5  $\mu$ l in volume) is placed on the column, either with a hypodermic syringe directly through the septum of the injector onto the column, or into a sample loop of a sample valve, which by rotation allows the carrier gas to carry the sample onto the column. The



**Figure 9** The basic gas chromatograph.



**Figure 10** The basic liquid chromatograph.

temperature program is initiated, the separation developed, and the chromatogram monitored by a recorder or stored on the computer disk. The data can be processed by the computer or the analysis calculated from the recorder chart.

### The Basic Liquid Chromatograph

A diagram of the basic liquid chromatograph is shown in **Figure 10**.

The chromatograph consists of six units: the gradient system, the pump, the sample valve, the column, the detector, and the recorder.

**The solvent supply and gradient system** Solvents are stored in glass or stainless steel vessels (usually two or three in number) and are best sparged with helium to eliminate dissolved air. Some systems employ a vacuum-assisted permeable membrane device for degassing. This method, however, is inferior in performance to helium sparging, which, in particular, must be used with microbore and capillary columns. If the mixture contains substances that cover a wide range of polarity, then gradient elution must be employed. There are two types of gradient programmers: programmers that have a pump for each solvent and are programmed to mix the solutes at high pressure and programmers that have a mixing valve prior to a single pump and, thus, the different solvents are mixed at low pressure.

**Sample valve** There are two types of sample valves: the internal loop valve and the external loop valve. The former are for use with small bore columns (columns having internal diameters of less than 1.5 mm) and the latter are used for larger diameter columns up to semipreparative columns. Sample volumes generally range from 0.5 to 5  $\mu$ l.

**Column ovens** The thermal capacity of a liquid chromatographic column is much higher than that of the gas chromatograph as the specific heats of liquids are much greater than those of a gas. As a result it is often necessary to preheat the mobile phase before it

enters the column to maintain good temperature control. Air is usually employed as the thermostating medium although a liquid thermostating medium would be more thermally effective but renders column changing more difficult.

**Detectors** The most common liquid chromatography detectors are the ultraviolet detector, the fluorescence detector, and the refractometer. Liquid chromatography detectors have less sensitivity (defined as the minimum detectable concentration) than gas chromatography detectors (in fact several orders of magnitude less). The fluorescence detector is probably one of the most sensitive detectors and the refractive index detector the least sensitive. In special applications (e.g., carbohydrate analysis), the light scattering detector is growing in popularity but, unfortunately, although universal in response, has a sensitivity (as defined previously) that is only slightly better than that of the refractive index detector.

**Analytical procedure** The analytical procedure is very similar to that used in gas chromatography, except that sample valves are exclusively used to inject samples onto the column.

See also: **Extraction:** Solvent Extraction Principles. **Gas Chromatography:** Overview. **Liquid Chromatography:** Overview.

### Further Reading

- Bedoukian PZ (1992) 48th Annual report on perfumery and flavor materials. *Perfumer and Flavorist* 17(3): 1–30.
- Croteau R and Karp F (1991) Origin of natural odorants. In: Müller PM and Lamparksy D (eds.) *Perfumes: Art, Science and Technology*, pp. 101–126. London: Elsevier.
- Kaiser R (1991) Trapping, investigation, and reconstitution of flower scents. In: Müller PM and Lamparksy D (eds.) *Perfumes: Art, Science and Technology*, pp. 213–250. London: Elsevier.
- Lamparksy D (1992) Methodes d'analyse des arômes alimentaire-chromatographie en phase gazeuse. In: Richard H and Multon H (eds.) *Les Arômes Alimentaires*, pp. 278–316. Paris: Lavoisier Tec & Doc.
- Schreier P (1984) *Chromatographic Studies of Biogenesis of Plant Volatiles*. Heidelberg: Dr. A H Hüthig.
- Scott RPW (1995) *Techniques of Chromatography*. New York: Dekker.
- Scott RPW (1998) *Chiral Chromatography*. New York: Wiley.
- Scott RPW (1998) *Tandem Techniques*. New York: Wiley.
- Scott RPW (2001) *Quantitative Chromatographic Analysis*. New York: Dekker.

# ESTROGENS

See ENDOCRINE DISRUPTING CHEMICALS

# ETAAS

See ATOMIC ABSORPTION SPECTROMETRY: Electrothermal

# ETHANOL

W J Criddle, University of Wales, Cardiff, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1225–1232, © 1995, Elsevier Ltd.

## Introduction

Alcohol ( $\text{CH}_3\text{CH}_2\text{OH}$ ), more correctly known as ethanol or ethyl alcohol to distinguish it from the generic group of alcohols, is one of the most important individual organic compounds readily available in virtually unlimited quantities. It has been described as one of the most versatile of compounds, being widely used as a solvent, a germicide, a beverage, an antifreeze, a fuel, and perhaps most importantly, as an intermediate for the synthesis and production of other organic compounds.

Fermentation has been used to produce ethanol since ancient times, alcoholic beverages certainly having been known to the Pharaohs of ancient Egypt. It has been suggested that the name 'alcohol' derives from the Arabic words 'al' and 'kohl' used to describe an ancient preparation used to darken hair that may have contained ethanol. There is also some suggestion that Noah had a vineyard where a form of alcoholic beverage was produced.

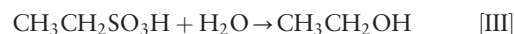
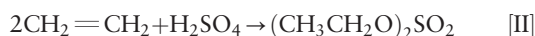
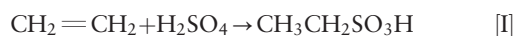
## Ethanol Manufacture

Ethanol is manufactured by three main processes: indirect hydration, direct hydration, and fermentation.

### Indirect Hydration

In this process, ethene (ethylene) is passed into sulfuric acid to form mono- and diethyl sulfates, which

may then be hydrolyzed to give ethanol. Reactions [I]–[IV] show how this occurs:



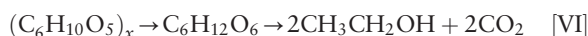
### Direct Hydration

Ethene and water are mixed and passed over a catalyst (phosphoric acid on silica) at a temperature of  $300^\circ\text{C}$  and a pressure of  $\sim 70$  atm:



### Fermentation

The process known as fermentation involves the breakdown of sugar-based substances using yeasts of various types and is the oldest method known:



Although the total volume of ethanol produced worldwide by this process is less than that produced by the methods described above, it is probably the most important from an analytical viewpoint as most analytical work is associated with fermentation ethanol in the form of alcoholic beverages.

## Physiological Properties

The physiological properties of ethanol are well known. It has an odor and taste that are generally considered pleasant. Ethanol has marked hypnotic properties and depresses activity in the upper brain

**Table 1** Some physical properties of ethanol

Property	Data
Boiling point (at 1 atm, °C)	78.32
Freezing point (°C)	−114.1
Density ( $d_4^{20}$ ) (g ml <sup>−1</sup> )	0.7893
Viscosity (20°C, cP)	1.17
Refractive index ( $n_D^{20}$ )	1.36143
Solubility (water)	Miscible, all proportions
Flash point (°C)	12

even though it gives the initial impression of being a stimulant. Ethanol can be toxic if consumed to excess (the lethal dose in rats is 13.7 g per kg of body weight) but it is markedly less toxic than the other members of the homologous series.

## Physical Properties

Ethanol is a clear, colorless liquid. Its physical properties derive mainly from the presence of the hydroxyl group, infrared (IR) studies indicating that ethanol exists in the liquid state mainly in a dimeric form. The most important physical properties of ethanol are given in Table 1, some of which are important in the analytical chemistry of ethanol.

## Chemical Properties

The chemical properties are also of importance in certain of the procedures for ethanol determination. The hydroxyl group is the important component in the structure of ethanol, since all of the chemical properties of ethanol derive essentially from that group. From an analytical viewpoint, the most important property is the relative ease with which the molecule can be oxidized to give acetaldehyde (ethanal) and, in some cases, acetic acid (ethanoic acid). The pathway is shown in reaction [VII]:



## Environmental Considerations

Ethanol is a flammable liquid, its vapor concentrations being explosive in the range 3.3–19.0% (v/v) in air. However, flammability is not a serious problem in the industrial environment providing there is reasonable ventilation, which also prevents ethanol from causing severe problems as an industrial poison. The compound is reported to be detectable by nose at ~350 ppm, while higher levels of ~5000–10 000 ppm can cause irritation to the eyes and mucous membranes of the upper respiratory tract.

Concentrations in excess of these values soon become intolerable, causing severe coughing and laceration. However, as ethanol is readily oxidized in the body to water and carbon dioxide, it is not an accumulative poison. Ethanol poisoning normally results from the consumption of liquid ethanol in alcoholic beverages rather than from inhalation. It is this distinction that governs the main areas of application required for ethanol determination, namely measuring concentration levels in alcoholic beverages of all types and measuring blood ethanol levels in humans after the consumption of alcoholic beverages. These two principal areas will be considered in detail.

## Types of Alcoholic Beverage

### Beer

There are four areas in the brewing of beer where ethanol measurement is of vital importance. These are: (1) duty payment, (2) trade returns, (3) quality control, and (4) high-gravity beer cracking. Note that the level of ethanol in beer (including high-gravity beer) is 2.5–12.0% (v/v), while ethanol levels in some other beverages, i.e., ‘shandy’ and ‘alcohol-free’ types of beer and all soft drinks, fall in the range 0–1.5% (v/v).

**Duty payment** In most countries, excise duty is payable on the basis of the original gravity (OG) of beer.

**Trade returns** In the UK, the ethanol content of spoiled beer must be assessed by the customs and excise authorities prior to return of excise duty.

**Quality control** The modern brewing process requires that beer be passed through extensive lengths of piping, which can lead to dilution of beer with washing solutions.

**High-gravity beer cracking** It is now the case that most beer is initially brewed to an alcoholic strength up to three times the required level. The high-gravity beer is diluted with water to give the required strength and this dilution is controlled by the ethanol content, which must be monitored regularly during the dilution process.

### Wines and Spirits

European Community (EC) regulations require that the ethanol concentration in beverages falling in this group should be displayed on bottle labels. The excise duty payable on wine and fortified wine depends



on the ethanol concentration range, the ranges being: (1) <15% (v/v), (2) 15–18% (v/v), (3) 18–22% (v/v), and (4) >22% (v/v). Similar rules apply to distilled spirits where the range system applies up to 96% (v/v), although most spirits contain between 35% and 50% (v/v).

## Analytical Methodology

The analytical chemistry of alcoholic beverages, of which the determination of ethanol is only one facet, continues to develop. However, several methods of analysis for ethanol are available and tradition, government legal requirements, and convenience, taken together or separately, usually govern the choice of method. These methods include: (1) distillation and specific gravity measurement, (2) gas–liquid chromatography, (3) IR spectrometry, (4) headspace techniques, (5) enzyme methods, (6) chemical (colorimetric) methods, and (7) refractometry. These methods are now considered in turn.

### Distillation

The distillation method is the benchmark for all other procedures and is officially approved by excise authorities in most countries, particularly when any form of litigation is involved. The procedure described below is that approved in the UK for the determination of OG values of beer. Essentially the same procedure is used for other beverages. Details are as follows:

A sample of beer at 60°C is shaken and then filtered to remove dissolved gases (primarily carbon dioxide). The beer (100 ml) is distilled using standard distillation equipment to give a distillate volume of ~85 ml. The distillate is made up to 100 ml with water, the temperatures being maintained at 60°C. The specific gravity of the distillate is measured using traditional methods, i.e., the density bottle, pycnometer or hydrometer, or, as is increasingly the case, using electromagnetically induced oscillation techniques (see below). In addition, the residue in the distillation vessel is made up to 100 ml and its specific gravity also measured giving the 'Residue Gravity' value. The 'Spirit Indication' is calculated from the reduction in specific gravity from 1.000 00 resulting from the ethanol present in the distillate being multiplied by 1000. This value is then converted from statutory tables to the 'Gravity Lost' value. The OG value is calculated by adding the 'Residue Gravity' to the 'Gravity Lost' value.

For samples other than beer, i.e., wines and spirits, the experimental procedure, while essentially the same as for beer, is modified in two important

aspects, namely the temperature for density measurement is 20°C and the ethanol value is obtained from UK Customs and Excise Laboratory Alcohol Tables.

The determination of the specific gravity of the various distillates and residues is clearly a key element in the determination of ethanol by distillation. Such determinations by specific gravity bottle or pycnometer are well established and will not be described here. However, the use of hydrometers for this purpose now requires that only hydrometers complying with BS 5471:1977 can be used for official analyses in the UK.

More important at the present time is the increasing use of electronic devices for density measurement. Such instruments allow the liquid density to be determined with high precision and accuracy using the principle that when a column of liquid is electromagnetically excited, its period of oscillation is related to its density by the relationship:

$$D = A(T^2 - B) \quad [1]$$

where  $D$  is the density,  $T$  is the period of oscillation, and  $A$  and  $B$  are calibration constants. The procedure requires that a small U-tube is filled with 0.7 ml of the distillate and the sample temperature controlled to  $\pm 0.01^\circ\text{C}$ . The sample is electromagnetically excited and the period of oscillation determined by comparison with a highly accurate reference frequency produced by a crystal oscillator. It is claimed that the results of this procedure are unaffected by viscosity, volatility, or turbidity. This sometimes allows the direct determination of ethanol, although such results must be treated with caution depending on the degree of dissolved solids, e.g., sugars, in the sample, which could affect the density of the sample and thus affect the ethanol value. In the context of ethanol determination, the instrument microprocessor allows the direct display of ethanol content.

An assessment has been reported of the distillation procedure that suggests that the overall accuracy of the method is  $\pm 2\%$ . However, the procedure is laborious (~30–40 min per determination) and requires a high degree of skill on the part of the analyst. However, it is the method preferred by many analysts and is the official method by which disputes regarding ethanol content are judged.

### Gas–Liquid Chromatography

The use of gas–liquid chromatography (GLC) as a general analytical technique is now well established and it is rare to find a modern analytical laboratory without such a facility, although many smaller breweries and wineries still do not have such equipment. As an important analytical technique in its own right,



it is discussed elsewhere in this encyclopedia and as such will not be considered here in detail. However, ethanol can be determined by standard GLC techniques in most situations appertaining to alcoholic beverages. Ethanol is a volatile substance that lends itself to analysis by GLC, flame ionization detection normally being chosen because of its insensitivity to water. Polar columns are the norm, poly(ethylene glycol) types usually being chosen. Internal standard techniques are normally adopted, usually using one of the propanols as an internal standard. Advantages of GLC are the high degree of selectivity achievable, especially using modern capillary and megabore columns, and the high speed of analysis compared with distillation. Accuracy and precision are about the same or marginally better than distillation. GLC is now accepted as a standard procedure for beer analysis in the US but the rather high cost of the equipment and the necessity for relatively high technical skills on the part of the analyst still preclude its use by small beer and wine makers.

### Infrared Spectroscopy

Infrared spectrometry is now extensively used both to characterize and determine organic compounds. As a technique it is considered elsewhere in this Encyclopedia. Its specific use to determine ethanol in alcoholic beverages is not widespread although near-IR instruments dedicated to this purpose are available. However, such instruments are extremely costly and can normally only be justified by a large throughput of samples. However, IR spectroscopy is widely used for breath alcohol determination, particularly in the UK, where the method is approved as the definitive procedure for evidential use. In this context, the absorption bands normally utilized are the asymmetric C–H stretching band at  $2950\text{ cm}^{-1}$  and the complex C–O stretching band at  $1050\text{ cm}^{-1}$ .

### Headspace Techniques

The volatile nature of ethanol makes it eminently suitable for headspace analysis. The principle on which the method is based is that of partition theory, i.e., that the concentration of a volatile species in the headspace above a solution is proportional to the concentration of that species in solution. Procedures based on this principle must therefore be classed as 'indirect', but nevertheless they have some advantages over direct methods, the most important being that the procedure separates volatile species from the numerous nonvolatile compounds commonly found in alcoholic beverages. As most sensors used in instrumentation for the determination of ethanol are at best semispecific and are usually oxidative in

principle, this feature is most advantageous, as many of the nonvolatiles, e.g., sugars, are readily oxidized.

Several instruments are available that operate on this principle. All involve the passing of a flow-controlled gas (usually air or nitrogen) stream through the beverage (which may require dilution). The gas stream is fed to a suitable sensor, comparisons usually being made with similar streams generated from standard ethanol solutions. In one type of instrument, the sensor is a ceramic material coated with a catalytic agent that is sensitive to oxidizable substances. The sensor forms part of a Wheatstone bridge circuit. When an oxidizable species is present the ceramic surface of the sensor is heated, causing a change of resistance in the sensor, which in turn generates an out-of-balance current in the Wheatstone bridge circuit. In another instrument based on the headspace principle, the sensor is a fuel cell that generates a potential when exposed to readily oxidizable compounds. This sensor is somewhat more selective than the ceramic types, as, of the numerous chemical species found in alcoholic beverages, it is sensitive only to aldehydes and alcohols.

Instruments of the above type are nowadays usually microprocessor controlled and give a direct reading of the ethanol content of the sample. They are relatively costly instruments, much of the cost resulting from the necessity for precise control of sample temperatures, as the headspace concentration of ethanol is very susceptible to temperature change, the variation being  $\sim 6\%$  per  $^{\circ}\text{C}$ . However, such instruments allow a high throughput of samples, and have the added advantage of being portable, allowing on-site ethanol determinations as and when required.

### Enzyme Methods

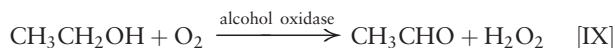
In recent years, numerous methods of ethanol determination have been developed, several of which are a variation on the following theme. Ethanol is oxidized to ethanal using nicotinamide adenine dinucleotide ( $\text{NAD}^+$ ) and alcohol dehydrogenase according to reaction [VIII]:



Normally this equilibrium lies well toward ethanol but is easily displaced quantitatively to ethanal formation by increasing the pH or removing the ethanal from the system. The NADH produced is usually determined spectrophotometrically by its absorbance at 334, 340, or 365 nm.

More recently, the use of electrodes incorporating alcohol oxidase has been used for ethanol assay. In this situation, the ethanol is oxidized in the presence of oxygen to give hydrogen peroxide according to

reaction [IX]:



The hydrogen peroxide formed can be measured amperometrically with an appropriate electrode.

Many publications have appeared on the themes outlined above, but the approach has been dogged by the instability of the enzymes, leading to the need for frequent electrode calibration. Work is continuing and it is possible that in due course this approach might gain acceptance for ethanol assay. To date, however, the enzymatic approach has had little impact on the more well-established procedures of gas chromatography and distillation.

### Chemical Methods

Ethanol can be assayed in solution using oxidizing reagents, potassium dichromate being the most commonly used reagent. The color change from yellow (chromium(VI)) to green (chromium(III)) is used, usually with a spectrophotometric endpoint. The ethanolic solution is refluxed with excess acidified ( $\text{H}_2\text{SO}_4$ ) potassium dichromate until no further reaction occurs and the chromium(III) assayed as indicated.

Although this method is little used in the alcoholic beverage industries, mainly because of its lack of specificity, it has found extensive application in breath and blood ethanol analyses associated with drink-driving cases. This method will be considered in more detail below.

### Refractometry

Refractometers have long been used to determine the alcoholic strength of beer in particular, usually in conjunction with specific gravity determinations. It has been reported that it is possible to determine OG, ethanol content, and Residue Gravity of beers by simultaneously measuring the density of the beer and its refractive index and applying a set of empirical equations to the data. Fully automated instruments are available utilizing this approach.

## Breath and Blood Ethanol Determination

Since 1960, the need for reliable methods of ethanol determination for blood samples taken from vehicle drivers has been driven by the introduction of legislation in many of the developed countries and an increasing number of Third World countries. This legislation introduced legal limits for the level of blood ethanol for car, lorry, and public vehicle drivers. In the UK, this value is currently 80 mg of

ethanol per 100 ml of blood, but the value varies considerably around the world.

The legislation brought with it the need for two ethanol determinations to be made, the first a screening test that established a *prima facie* case that the suspect had a blood ethanol value at or in excess of the legal limit and would allow the law enforcement agencies to apply a second, more stringent, assay that could be used evidentially, i.e., a value that would be acceptable as evidence in a Court of Law.

### The Breath and Blood Ethanol Relationship

It has been shown that, under controlled conditions, there is a direct relationship between the level of ethanol in the breath and that in the blood. Ethanol, after ingestion, is absorbed into the bloodstream largely through the small intestine. Blood circulation carries the ethanol to the lungs where the ethanol equilibrates rapidly with lung air. The relationship established is such that at body temperature (which essentially thermostats the sample), the UK legal limit of 80 mg of ethanol per 100 ml of blood gives rise to a breath sample containing 35  $\mu\text{g}$  of ethanol per 100 ml of breath. It is, therefore, essential that the sample taken be such that it represents a deep lung sample and it is necessary to ensure that sufficient time has elapsed between the subject's last drink and sampling, as mouth samples taken immediately after a drink would be artificially high. This elapsed time varies worldwide, but is usually  $\sim 20$  min.

### Screening Procedures

Screening of suspects in drink-driving cases is normally carried out by police officers with little or no knowledge of analytical chemistry. It is, therefore, of paramount importance that the procedure adopted demands the minimum of experimental skill, is easily applied under difficult circumstances, and gives a result with a relative accuracy of  $\sim \pm 10\%$  (w/v). Two methods are currently adopted, both involving the sampling of breath, one involving a version of the chemical test (potassium dichromate) referred to above, and the other, more recent procedure, utilizing a sensor and microelectronics to both sample and assay the breath sample.

**'Blow-in-the-bag' screeners** Early breath screeners consisted of a glass tube containing a precise amount of silica gel impregnated with acidified ( $\text{H}_2\text{SO}_4$ ) potassium dichromate. The packing (which is yellow) was contained between graduation marks, with a further graduation within the outer marks indicating a 'fail' situation. When a sample of breath containing ethanol was blown through the tube, the dichromate color changed from yellow to green (chromium(III)). If the

green color extended beyond the graduation mark, the subject was considered to have failed the test, and could be further examined by more sophisticated procedures. The volume of breath used was clearly critical, and this was measured by the subject blowing through the tube, ideally with one deep breath, until a standard volume plastic bag attached to the tube was completely filled.

**Electronic screeners** Difficulties with the above methods, arising mainly from the problems deriving from the rather poor color change of the dichromate tubes, and uncertainty about an individual's ability to fill the plastic bag reproducibly, provided the impetus for the development of a more sophisticated approach to breath screening.

Most developed countries now use an electronically controlled breath sampling and monitoring system that incorporates a sampling tube into which the suspect blows (Figures 1 and 2). Incorporated into the tube are a pressure transducer and indicator light

to ensure that a deep lung sample has been taken. When the indicator light shows that a sufficient exhalation of breath has occurred, a sampling valve is activated which takes a fixed volume of the breath sample and feeds it to the sensing device, which in most cases is a platinum catalyst fuel cell.

This type of fuel cell sensor provided a major breakthrough in breath screening analysis for several reasons. These include relatively low cost, robustness, sensitivity, and perhaps, above all, the fact that the cell generates its own electrical output, only requiring power for signal amplification purposes. A fuel cell thus consists of two catalytic (usually platinum) surfaces supported on a suitable porous material (commonly sintered poly(vinyl chloride)) impregnated with an electrolyte such as aqueous sulfuric acid. Oxidation occurs at the working surface, and an equivalent reduction (usually of atmospheric oxygen) occurs at the counter-electrode.

The ethanol in the breath sample is oxidized in the fuel cell as follows according to reaction [X]:



This gives an electrical output that is linearly proportional to the ethanol content up to four to five times the legal limits. The results are usually displayed digitally as blood ethanol values in whatever units are appropriate for the particular country, although in the UK a system of colored lights is used to represent different levels of ethanol.

### Evidential Methods

**Gas chromatography** Ethanol may be assayed directly in blood samples using gas chromatography, the procedure being similar to that used for alcoholic beverages. An internal standard procedure is usually adopted, propan-2-ol normally being used as internal standard. However, the method has fallen out of favor due to the invasive nature of the sampling procedure (blood sample) and it is now used only sparingly. However, in cases where the data from other methods are disputed, it is still used as the final arbitration procedure and can usually be demanded by a suspect who wishes to challenge analytical data obtained by other methods.

**Infrared spectroscopy** Infrared spectroscopy has become the most widely used, although not the only, method of evidential breath and blood analysis. It has the considerable advantage of being extremely rapid and lends itself well to vapor (breath) analysis. These features are of particular importance as most governments specify that the time taken for the total analytical process, which includes the required calibration procedure, should not exceed ~5 min. In



**Figure 1** A roadside breath screening test being administered. (Photograph from The Scotsman Publications Ltd.)



**Figure 2** A fuel cell-based evidential breath analyzer.

essence, the suspect will be required to provide a deep-lung breath sample in much the same way as described for screening, although the sample delivery tube is heated to prevent condensation of water in the breath. The breath sample is fed to an IR cell where absorbance measurements are made, frequencies being selected using optical filters. The instrument most widely used in the UK utilizes such a system, the asymmetric stretching frequency at  $2950\text{ cm}^{-1}$  being used. The possibility of interfering substances (only acetone (propanone) is likely to be present naturally in sufficiently high levels to cause interference) is obviated by the incorporation of an additional sensor, usually of the Taguchi (tin oxide) type. This latter sensor produces an offset signal and the test procedure is aborted. Direct blood ethanol determination by gas chromatography is usually resorted to in such circumstances.

Organisation Internationale de Metrologie Legale (OIML) regulations have in recent times demanded a considerably higher degree of selectivity and this has resulted in the development of more sophisticated optical systems allowing measurement at lower frequencies, particularly at  $1050\text{ cm}^{-1}$ , where selectivity for ethanol is high. Detector sensitivity at this frequency is low but modern electronics appear to have coped with this situation, several instruments with measurements made at this lower frequency now being available.

**Fuel cell-based evidential monitors** Although the IR spectroscopic approach to evidential breath analysis appears to have gained general acceptance in most developed countries, the need for a more robust instrumental option has been apparent in the needs of some developing countries. To this end, some instrumentation is currently available using a fuel cell sensor of the type used in screening devices. These units are essentially the same as screeners, but with more sophisticated electronics for signal analysis, automatic (usually motorized) sampling systems, and a means of printing data in a form acceptable for evidential use. Their main advantage is that, compared to the relatively fragile optical system of an IR unit, the fuel cell sensor is extremely robust and instrumentation of this type can withstand a considerable degree of maltreatment.

## Other Areas Requiring Alcohol Analysis

### Yeast Production

Bakers yeasts used in the making of bread are produced by propagation in closed vessels. The yeast is

held as a suspension by continuous stirring and aeration, molasses and suitable nutrients (nitrogen source) being used as the growth medium. Under these conditions (pH  $\sim 4.4$ ,  $32^\circ\text{C}$ ), the food source is metabolized to carbon dioxide. However, should the rate of aeration become insufficient to sustain this process, the fermentation will become anaerobic and ethanol will be produced with a diminution in the efficiency of yeast growth. Headspace techniques are usually applied to monitor ethanol formation, i.e., GLC and fuel cell-based methods, although solution analysis can also be used.

### Vinegar Manufacture

The production of vinegar involves an initial anaerobic alcoholic fermentation followed by an aerobic fermentation using acetic acid bacteria to convert the ethanol to acetic acid. The ethanol concentration is vital to the overall efficiency of the process and must be monitored regularly using procedures similar to those indicated for yeast production.

### Controlled Atmosphere Fruit Storage

The large-scale storage of fruit, particularly apples and pears, is now carried out under refrigeration ( $\sim 4^\circ\text{C}$ ) and in a chemically controlled atmosphere where the oxygen concentration is reduced to below 4% (v/v). The onset of fruit deterioration is heralded by the production of ethanol, which may be detected by online analyzers of the headspace type.

*See also:* **Distillation.** **Enzymes:** Overview; Enzyme-Based Electrodes. **Food and Nutritional Analysis:** Antioxidants and Preservatives; Alcoholic Beverages. **Forensic Sciences:** Alcohol in Body Fluids; Blood Analysis. **Headspace Analysis:** Static; Purge and Trap. **Infrared Spectroscopy:** Overview. **Optical Spectroscopy:** Refractometry and Reflectometry. **Quality Assurance:** Quality Control. **Sensors:** Amperometric Oxygen Sensors.

## Further Reading

- Barzana E, Klibanov AM, and Karel M (1989) A colorimetric method for the enzymic analysis of ethanol and formaldehyde vapors using solid alcohol oxidase. *Analytical Biochemistry* 182: 109–115.
- Clark LC, Jr. (1973) Family of polarographic enzyme electrodes and the measurement of alcohol. *Biotechnology and Bioengineering Symposium* 3: 377–394.
- Criddle WJ, Jones TP, and Neame MJH (1984) Current trends in alcohol measurement and control. Part 3. *Journal of the Institute of Measurement and Control* 17: 109–111.
- Criddle WJ, Jones TP, and Parry KW (1989) A new approach to ethanol measurement in alcoholic beverages. *Journal of the American Society of Brewing Chemists* 47: 1–3.



HM Customs and Excise (1979) *Laboratory Alcohol Tables*. RD C80/267/04. London: HMSO.

HM Customs and Excise (1981) *Tariff and Overseas Trade Classification*. London: HMSO.

Leopold H, Jelinek R, and Tilz GP (1977) The application of the mechanical oscillator technique for the determination

of the density of physiological fluids. *Biomedizinische Technik* 22: 231–234.

Noordzij PC and Roszbach R (eds.) (1987) *Alcohol, Drugs and Traffic Safety T86*. Amsterdam: Elsevier.

OIML (1988) *International Recommendations Relating to Evidential Breath Analysers*. Reporting Secretariat OIML (SP 30–SR 13), France.

## EXPERT SYSTEMS

See **CHEMOMETRICS AND STATISTICS: Expert Systems**

## EXPLOSIVES

See **FORENSIC SCIENCES: Explosives**

## EXTRACTION

Contents

**Solvent Extraction Principles**

**Solvent Extraction: Multistage Countercurrent Distribution**

**Microwave-Assisted Solvent Extraction**

**Pressurized Fluid Extraction**

**Supercritical Fluid Extraction**

**Solid-Phase Extraction**

**Solid-Phase Microextraction**

### Solvent Extraction Principles

**A M Sastre**, Universitat Politècnica de Catalunya, Barcelona, Spain

**J Szymanowski**, Poznan University of Technology, Poznan, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Solvent extraction is a classical analytical technique used to determine the contents of various inorganic and organic species. Inorganic compounds are usually extracted after complexation with organic ligands. The technique also enables preconcentration of solutes and their separation and finds practical applications in various industries, including nuclear, metal processing, petrochemistry, pharmaceutical,

and others. Depending upon the extraction system considered, different ligands may be used, normally dissolved in organic diluents. Due to environmental problems connected with the use of organic solvents, novel systems with more environment-friendly solvents have been proposed for various new preconcentration methods.

Various phenomena should be considered in solvent extraction. The analytical purposes and determination of complex stoichiometry are the classical problems. The amphiphilic character of many extraction reagents must also be taken into consideration when studying kinetics and extraction mechanism.

### Types of Reagents and Extraction Reactions

Physical extraction is seldom used in practice; extraction reagents are normally used. These can be

HM Customs and Excise (1979) *Laboratory Alcohol Tables*. RD C80/267/04. London: HMSO.

HM Customs and Excise (1981) *Tariff and Overseas Trade Classification*. London: HMSO.

Leopold H, Jelinek R, and Tilz GP (1977) The application of the mechanical oscillator technique for the determination

of the density of physiological fluids. *Biomedizinische Technik* 22: 231–234.

Noordzij PC and Roszbach R (eds.) (1987) *Alcohol, Drugs and Traffic Safety T86*. Amsterdam: Elsevier.

OIML (1988) *International Recommendations Relating to Evidential Breath Analysers*. Reporting Secretariat OIML (SP 30–SR 13), France.

## EXPERT SYSTEMS

See **CHEMOMETRICS AND STATISTICS: Expert Systems**

## EXPLOSIVES

See **FORENSIC SCIENCES: Explosives**

## EXTRACTION

Contents

**Solvent Extraction Principles**

**Solvent Extraction: Multistage Countercurrent Distribution**

**Microwave-Assisted Solvent Extraction**

**Pressurized Fluid Extraction**

**Supercritical Fluid Extraction**

**Solid-Phase Extraction**

**Solid-Phase Microextraction**

### Solvent Extraction Principles

**A M Sastre**, Universitat Politècnica de Catalunya, Barcelona, Spain

**J Szymanowski**, Poznan University of Technology, Poznan, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Solvent extraction is a classical analytical technique used to determine the contents of various inorganic and organic species. Inorganic compounds are usually extracted after complexation with organic ligands. The technique also enables preconcentration of solutes and their separation and finds practical applications in various industries, including nuclear, metal processing, petrochemistry, pharmaceutical,

and others. Depending upon the extraction system considered, different ligands may be used, normally dissolved in organic diluents. Due to environmental problems connected with the use of organic solvents, novel systems with more environment-friendly solvents have been proposed for various new preconcentration methods.

Various phenomena should be considered in solvent extraction. The analytical purposes and determination of complex stoichiometry are the classical problems. The amphiphilic character of many extraction reagents must also be taken into consideration when studying kinetics and extraction mechanism.

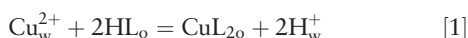
### Types of Reagents and Extraction Reactions

Physical extraction is seldom used in practice; extraction reagents are normally used. These can be



divided into acidic, basic, and solvating (neutral) extractants.

Acidic extractants can be further subdivided into simple organic acids, e.g., carboxylic acids, phosphoric acids, phosphonic acids, phosphinic acids, thiophosphorus acids, and sulfonic acids, and chelating reagents, e.g., hydroxyoximes, 8-hydroxy-(thio)quinolines,  $\beta$ -diketones, nitrosophenols, and pyridylazonaphthols. These reagents transfer metal cations into the organic phase in the form of uncharged complexes, e.g., copper(II) with hydroxyoximes (HL):



where subscripts 'w' and 'o' denote the aqueous and organic phases, respectively.

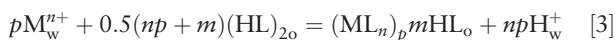
The distribution ratio ( $D = C_o/C_w$ ) is the function of reagent concentration and equilibrium pH:

$$\log D = \log K_{\text{ex}} + 2 \log [\text{HL}]_o + 2\text{pH} \quad [2]$$

An increased acidity of the aqueous phase shifts the equilibrium of eqn [1] to the left. Due to this, acidic extractants cannot transfer large amounts of metal cations, and the back-extraction is accomplished with mineral acids of appropriate concentration.

Extraction can be used to separate and pre-concentrate the analyte component(s), to recover valuable components from aqueous solutions (hydro-metallurgy, biotechnology), and also to determine the stoichiometry of extracted complexes. In the latter case, the relations  $\log D$  versus pH and  $\log D$  versus  $\log \text{HL}$  are considered, and the composition of the complex can be determined from the slopes of the linear parts of these relationships.

The extractant can form associated forms, usually dimers, and the extracted complex can be solvated by undissociated extractant molecules and can undergo polymerization. Thus, the more complex eqn [3] can be presented:



Values of  $n$ ,  $m$ , and  $p$  can be determined from extraction data by slope analysis or by using computer programs.

The extraction constant  $K_{\text{ex}}$  present in eqn [2] is the function of several individual constants characteristic of steps such as the partitioning of the ligand ( $P_{\text{HL}}$ ), the dissociation of the ligand ( $K_a$ ), the overall complex stability constant ( $\beta_{\text{CuL}_2}$ ), and the partitioning of complex ( $P_{\text{CuL}_2}$ ):

$$K_{\text{ex}} = \beta_{\text{CuL}_2} K_a^2 P_{\text{CuL}_2} P_{\text{HL}}^{-2} \quad [4]$$

Thus, solvent extraction can also be used as a research method to determine certain physical-chemical data.

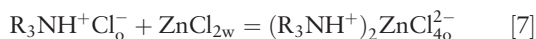
Basic reagents are usually primary, secondary, and tertiary amines or their ammonium salts with a different length and the structure of the alkyl group(s). Amines require protonation before they can act as extractants. For example,



They can only extract anions from acidic solutions, by transferring bulky anions to the organic phase in the form of ion-pairs. For example,



An alternative mode of extraction involves an addition reaction. For example,



The interpretation of extraction data must take into account the formation of various chlorocomplexes  $\text{ZnCl}_i^{2-i}$ , where  $i = 1, 2, 3, 4$ , and a more complex equation for  $\log D$  is obtained. For example, for eqn [7]:

$$\log D = \log(K_{\text{ex}} \cdot \beta_2) + 2 \log [\text{R}_3\text{NHCl}]_o + 2 \log [\text{Cl}^-]_w - \log \left( 1 + \sum_{i=1}^4 \beta_i \cdot [\text{Cl}^-]_w^i \right) \quad [8]$$

The stability constants of various metal anion complexes  $\beta_i$  are given in the literature. Usually, reliable data can be obtained for typical analytical diluted solutions, but such estimations are not reliable for concentrated solutions that contain various ions.

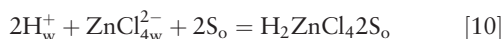
Quaternary ammonium salts react without protonation and the extraction is not sensitive to the acidity of the aqueous phase.

Solvating reagents such as neutral compounds (e.g., ethers, ketones, alcohols, and esters of carboxylic acids, alkyl phosphates, phosphonates, phosphinates, phosphine oxides, and phosphine sulfides) and also esters of pyridinecarboxylic acids and certain benzimidazole derivatives that have electron-withdrawing substituents form solvates with extracted neutral species. For example, dialkylpyridine-3,5-carboxylates extract Cu(II) at pH 3 from concentrated chloride solutions:



Solvating reagents can also extract charged species together with counterion(s). For example, tributyl

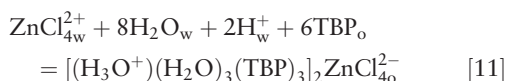
phosphate or trialkylphosphine oxides extract zinc(II) from hydrochloric acid solutions:



Instead of protons, metal cations can also be transferred to the organic phase.

As in the case of ion-pair extraction with basic reagents, the formation of various chlorocomplexes must be taken into consideration.

Solvating reagents can also extract bulky anions, e.g.,  $\text{CrO}_4^{2-}$ ,  $\text{MoO}_4^{2-}$ , and anionic chlorocomplexes, by means of the hydration mechanism. For example,



The preprotonation of solvating reagents, such as alcohols, and the extraction of anionic species are also possible:



The parallel transfer of extracted species by means of various mechanisms also occurs.

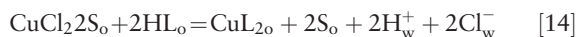
The solvating reagents are usually less selective than acidic extractants and the transferred complexes are not as well defined. A high concentration of salts and/or the presence of complexing anions are needed to obtain the efficient extraction of metal species. The drawbacks are coextraction of acids and water. The back-extraction is usually carried out with water.

Certain reagents can be considered bifunctional, i.e., 8-hydroxyquinoline and its derivatives and alkylpyridine carboxylates. The former, depending on the pH of the aqueous phase and the type of metal cation, can act as a chelating acidic extractant, forming chelates with copper(II) and zinc(II), or after protonation as a basic reagent, forming ion-pairs with anionic zinc(II) chlorocomplexes. Alkylpyridine carboxylates act as solvating reagents that extract copper(II) from chloride solutions at pH 3, but after protonation they extract palladium(II) efficiently from HCl solutions. The use of bifunctional extractants or the mixtures of basic and chelating extractants or solvating (S) and chelating (HL) extractants enables the extraction of metal species from chloride solutions as ion-pairs or solvates. Scrubbing with water causes a transfer of metal cation into chelate from which the cation can be stripped with sulfuric acid. Thus, mixed extractants enable a transfer of metals from chloride to sulfate

solutions. For example, for copper(II) extraction: extraction step:



scrubbing of chlorides with water:



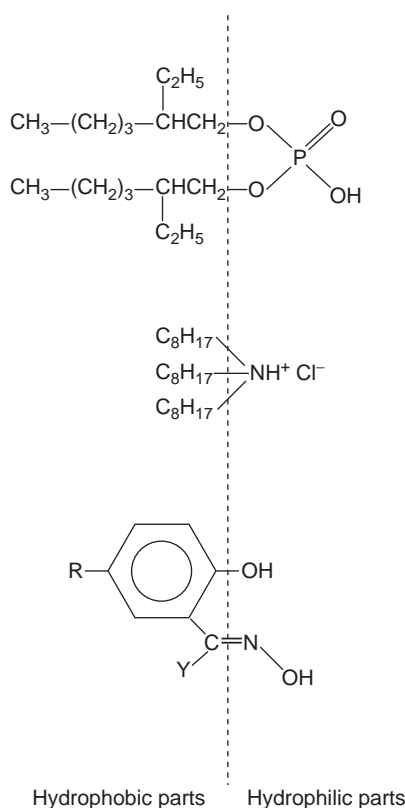
back-extraction:



Extractants are often used with modifiers (e.g., hydrophobic alcohols, alkylphenols, sterically hindered esters of carboxylic acids, and tributyl phosphate). Modifiers are used for two different reasons: first, to increase the solubility of extractants and their complexes and to avoid the formation of the third phase, and second, to modify the extraction properties, i.e., the extraction and stripping abilities. The first option is usually exploited in systems that contain various amines. The second option is associated with hydroxyoximes and the formation of tailored blends, which optimize the extraction-stripping properties and adjust them to the aqueous feed, i.e., to the acidity and concentration of copper(II). Modifiers that form hydrogen bonds with hydroxyoxime molecules decrease their extraction ability but facilitate back-extraction with acids. Weak hydroxyoxime or  $\beta$ -diketone extractants (e.g., 2-hydroxy-5-alkylbenzophenone oxime) can also act as modifiers of strong hydroxyoxime extractants (e.g., salicylaldoxime and its alkyl derivatives).

## Extractants as Amphiphilic Compounds

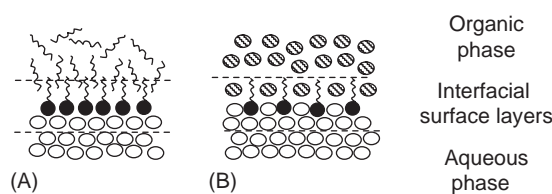
Extractants that have a hydrophobic substituent(s) and a hydrophilic fragment(s) (Figure 1) show an amphiphilic character, adsorb at water/organic solvent interfaces, and decrease the interfacial tension. Their interfacial activity depends on the type of hydrophilic and hydrophobic groups, the number and structure of the alkyl group, and the type of solvent. Thus, they behave in a similar way to hydrophobic surfactants in surface tension reduction and aggregation. However, in typical extraction systems, chelating extractants show moderate interfacial activity and decrease the interfacial tension to only  $15\text{--}25\text{ mN m}^{-1}$ . Strong acidic extractants and protonated alkylamines exhibit the highest interfacial activity, while solvating reagents and neutral amines are on the opposite side of the scale.



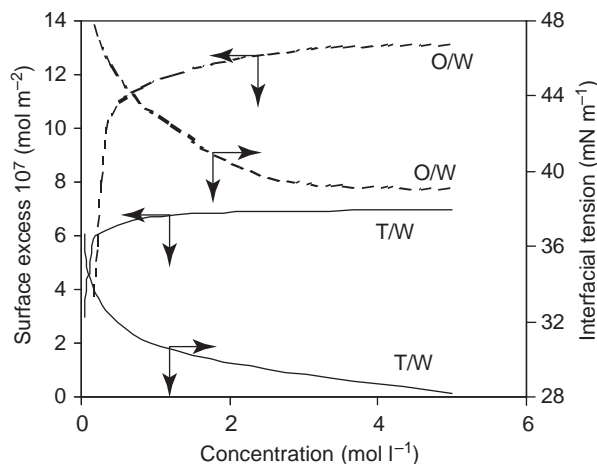
**Figure 1** Extractants as amphiphilic substances.

All factors that lead to hydration, protonation, and dissociation of the polar group decrease the hydrophobicity of extractant molecules and polarize them. The decrease in hydrophobicity makes the extractant molecules less compatible with the nonpolar organic phase, and enhances the attractive interaction between extractant head group(s) and water. Thus, the forces that draw the extractant molecules from the bulk organic phase to the interfacial region increase adsorption. On the other hand, phenomena such as solvation or aggregation of the extractant molecules result in the formation of species that are less surface-active than monomeric extractants. The adsorption, however, needs an appropriate balance of forces acting on both sides of the interface, and excessively strong interactions of the extractant head groups with water molecules favor the dissolution of extractant in the aqueous phase and inhibit the adsorption at the hydrocarbon/water interface.

The density of population and orientation of surfactants at the interface depends upon their structure and concentration and the type of diluent. The effect of the latter is decisive. Molecules of aromatic hydrocarbons that have  $\pi$  electrons compete with extractant molecules for access to the hypothetical interface, and solvating diluents disturb the adsorption of extractant molecules and the penetration of



**Figure 2** Interfacial zone with adsorbed extractant molecules (A and B denote systems with aliphatic and aromatic hydrocarbons).

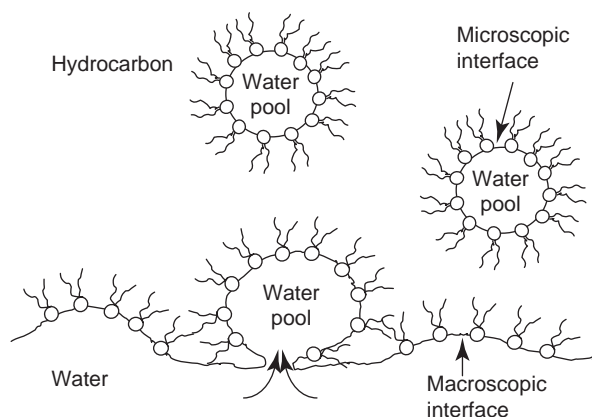


**Figure 3** Isotherms of interfacial tension and surface excess for acidic chelating extractant at octane/water (O/W) and toluene/water (T/W) interface.

extractant hydrophilic groups in the aqueous layers (Figure 2). The density of extractant population, or the surface excess  $\Gamma$ , can be calculated from the Gibbs isotherm alone or by the coupling of this isotherm with various equations that correlate the surface tension with extractant concentration in the organic phase, e.g., the Szyszkowski equation. The density of population increases with the extractant concentration only for a limited range of low concentrations, assuming a constant value at the saturated interface (Figure 3). The interface already becomes saturated at bulk extractant concentrations of  $10^{-3}$ – $10^{-2}$  mol l $^{-1}$ , and even earlier in the presence of solvating diluents.

Modifiers also have some hydrophilic groups and adsorb at liquid–liquid interfaces. Thus, they compete with extractant molecules for access to the interface and presence in the adsorption layer, thus decreasing the interfacial concentration of extractant molecules. The same phenomenon occurs when different extractants are present in the extraction system.

Some extractants that have acidic groups, e.g., sulfonic, phosphonic, phosphinic, and also some basic and solvating extractants in the systems containing acidic aqueous phase, form reverse micelles (Figure 4). The formation of reverse micelles appears



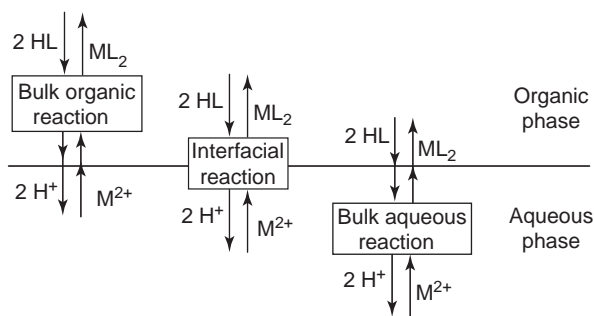
**Figure 4** Reverse micelles in organic phase. (Adapted from Nitsch W and Plucinski P (1990) Two-phase kinetics of the solubilization in reverse micelles. *Journal of Colloid and Interface Science* 136: 338–351.)

to be a general phenomenon for acidic extractants and occurs even at very low concentrations, e.g.,  $10^{-5}$  and  $10^{-3} \text{ mol l}^{-1}$  for dinonylnaphthalene sulfonic acid and di(2-ethylhexyl)phosphoric acid, respectively. Even tributyl phosphate forms reverse micelles in contact with solutions of HCl and  $\text{HNO}_3$ . The presence of the water pool in the core of reverse micelles causes nonspecific transfer of water-soluble components, thus decreasing the selectivity of extraction and disturbing the interpretation of the equilibrium extraction data. However, the presence of reverse micelles may increase the rate of extraction. This phenomenon is exploited in the so-called micellar extraction, especially for the extraction of biochemicals that need an aqueous environment to keep them stable.

Under certain conditions, in the presence of a cosurfactant, e.g., butanol, and/or alkaline aqueous phase, reverse micelles swell, accepting a large amount of water, and transparent and stable microemulsions are formed. Microemulsion can also be obtained when typical surfactants with cosurfactant are present in the organic phase. Such systems enable the rapid extraction of germanium(IV) and gallium(III) with alkyl derivatives of 8-hydroxyquinoline, while classical extraction needs days of phase contact.

## Kinetics and Mechanism of Extraction

Extraction is performed by the diffusional flow of a metal across the interface between two liquid phases, accompanied by the chemical reaction of metal complex formation. The diffusion of the species that participate in the chemical reaction is of great significance, as are the locations in which the chemical reaction proceeds, and the rate of complexation.

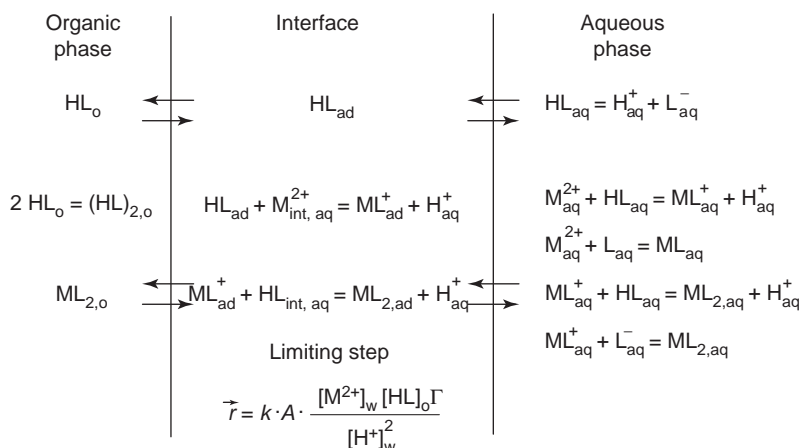


**Figure 5** Extraction pathways for acidic extractant with different location of complexing step.

Depending on the relative rates of the chemical and diffusion steps, the reaction can proceed in the kinetic, diffusion, or mixed regime, the entire process being controlled by the rate of the chemical step, a diffusion process, or by both kinetics and diffusion. Thus, under very good hydrodynamic conditions, e.g., upon vigorous agitation, the influence of the diffusion can be substantially eliminated and the kinetic results can be used to discuss the reaction mechanism. This conclusion is not always true, and the use of typical surfactant micellar aqueous solutions with extractants dissolved (solubilized) in micellar pseudophase (micelles) and inorganic species dissolved in aqueous pseudophase mimic the extraction systems effectively and the diffusion processes are totally eliminated.

The chemical step can occur in the bulk of the aqueous phase or aqueous pseudophase, in the bulk of the organic phase or in the micelle and/or at the actual interface, or at the hypothetical interface of the micelles. **Figure 5** shows the possible extraction pathways in typical extraction for metal recovery with an acidic extractant. Metal ions, such as hydrophilic and charged species, are insoluble in typical organic phases. Thus, the version of complex formation in the organic phase can be disregarded as long as the formation of reverse micelles and W/O microemulsions is avoided. However, such an option is possible in the extraction of organic compounds. The interfacial reaction can be understood as the reaction in the real interfacial zone of  $\sim 2 \text{ nm}$  thickness or as the reaction in a thin boundary layer in the aqueous phase.

The location of the reaction step depends on extractant surface activity and hydrophobicity, metal concentration, and hydrodynamics (agitation intensity). Low metal concentration and high hydrodynamics, together with high surface activity and the low solubility of the extractant in the aqueous phase (i.e., high hydrophobicity), favor the interfacial reaction. However, the reaction place may also depend on experimental conditions, i.e., the method used for the studies. In general, case reaction



**Figure 6** Reaction scheme of metal complexing from acidic sulfate solutions with hydroxyoxime extractants. (Adapted from Szymanowski J (1993) *Hydroxyoximes and Copper Hydrometallurgy*. Boca Raton: CRC Press.)

proceeds simultaneously both in the bulk of the aqueous phase and at the interface (Figure 6). The contribution of the interfacial reaction increases in a homologous series of extractants, e.g., with an increase in the alkyl group length.

In systems with reverse micelles/microemulsion, the transfer of extracted species occurs both through the macroscopic interface between dispersed and continuous phases and through the large microscopic interface between water pools and extractant hydrophilic groups in the cores of the micelles. The transfer of water with hydrophilic species is possible due to reverse micelles forming at the interface and sucking the aqueous phase.

Although adsorbed molecules of extractants are well oriented at the interface, the interfacial reactions occur more slowly, as recalculated for the same volume, than reactions in the aqueous phase. The extraction of palladium(II) from HCl solutions with dialkyl sulfides is a classic example of a very slow process. Modifying the extractant structure by adding a hydrophilic hydroxyl group and/or a phase transfer catalyst, e.g., trialkylamine or quaternary ammonium salt, increases the rate of extraction. The addition of sulfonic and phosphoric acids to hydroxyoximes may enhance the rate of copper extraction due to the formation of reverse micelles and the development of the microscopic interface. The  $\alpha$ -acyloin oximes that form an intermediate complex 'that has a five-member ring' with copper(II) increase the rate of extraction of aromatic hydroxyoximes. The addition of surfactants may cause both retardation and acceleration of extraction. If the interfacial tension is decreased, surfactants cause an increase in the interfacial surface area in dispersed systems. When they adsorb at the interface, they cause an additional

interfacial resistance, but when the charge is opposite to that of the intermediate metal complex, they enhance the transfer due to electrostatic attraction.

## New Methods of Extraction and Solute Preconcentration

Sample pretreatment is still currently the weakest link, the time-determining step in the whole analytical procedure and the primary source of errors. The increasing need for the determination of numerous analytes at lower and lower levels and in increasingly complex matrixes makes sample pretreatment, especially trace enrichment, an indispensable step in many analytical procedures.

Trace enrichment can be performed by liquid-liquid extraction (LLE) or solid-phase extraction (SPE). Although SPE is a relatively more recent and popular technique, LLE is still one of the most versatile techniques for enrichment because of its advantages: it is simple to perform, requires simple and inexpensive equipment, has a high capacity for interfering compounds, has a high potential for chemical tuning, and separates by incorporating different specific reagents. The main drawbacks of LLE are the high consumption of solvents, the difficulty of automation, and the formation of emulsion.

Many efforts have been made to overcome these inherent drawbacks. There is also a tendency to develop analytical procedures for sample preparation that reduce or completely eliminate the use of organic solvents and decrease the number of operations. New processes are now being developed that employ aqueous-based systems. These use water effectively and with a marked reduction in the risk to



human health and the environment. The following new extraction processes can be considered:

- cloud point extraction (CPE),
- micellar enhanced ultrafiltration (MEUF),
- extraction in aqueous biphasic systems (ABS),
- homogenous liquid–liquid extraction (HLLE),
- extraction to ionic liquids (IL),
- supercritical fluid extraction (SFE),
- continuous flow liquid–liquid extraction (CPLLE),
- supported liquid membrane (SLM) extraction, and
- microporous membrane liquid–liquid extraction (MMLLE).

When surfactants aggregate in aqueous solutions, they form micelles even at relatively low concentrations, i.e., above the critical micelle concentration. An increase in temperature causes the dehydration of nonionic surfactants that have a hydrophilic polyoxyethylene chain, e.g., Triton X-114, dissolved in aqueous solutions, the aggregation of micelles, and therefore the clouding of the solution at the cloud point (CP) and the separation of the surfactant-rich phase (SRP), as in **Figure 7**. The process can be reversed by cooling. The volume of SRP is 1–2 orders lower than the remaining aqueous micellar phase and depends on the surfactant concentration and the temperature of separation. The SRP volume increases with an increase in the surfactant concentration but decreases with an increase in temperature above the CP. An increase in temperature also causes a decrease in water content in the SRP, which becomes more hydrophobic. The phase can be considered a water solution in the surfactant that is micelle-like in structure. The CP of nonionic surfactants can be modified by the surfactant structure and the addition of electrolytes. An increase in surfactant hydrophobicity and an addition of salting-out electrolyte, e.g., NaCl, decreases CP. Salting-in electrolytes, such as KSCN, have the opposite effect. The cooling of solutions that contain zwitterionic surfactants, e.g.,  $R(CH_3)_2N^+-(CH_2)_3SO_3^-$ , also causes the clouding and separation of the SRP. Species that have some hydrophobicity (e.g., organic compounds, biochemicals, and

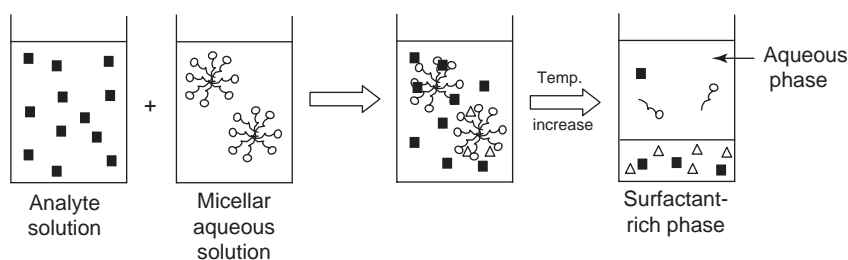
proteins) can be first solubilized in micelles and then transferred to the SRP. Ions must first be neutralized with hydrophilic complexing reagents. The CPE technique is used for the extraction and preconcentration of metal ions after the formation of sparingly water-soluble complexes as an initial step for their determination by various instrumental methods as well as for the separation and purification of various biochemicals, especially proteins.

Coacervation, the clouding of solutions that contain typical anionic surfactants, can be induced by the addition of an electrolyte or by acidification and can also be used for preconcentration. However, the phenomenon cannot be reversed as easily as the CP.

Substances that are solubilized in surfactant micelles can be separated by ultrafiltration through membranes whose pores are smaller in diameter than the micelle size. For a membrane molecular weight cut-off from 1 to 50 kDa, the rejections are  $\sim 98\%$ . The stream of water-containing monomeric molecules of surfactant (permeate) flows through the membrane. The remaining solution (retentate) contains solutes solubilized in micelles. The MEUF process is used for the separation of organic substances and various ions, the latter after their previous complexation.

Aqueous biphasic systems that contain two or more hydrophilic polymers, especially polyoxyethylene glycol–dextran–water or polymer and salt, and mainly polyoxyethylene glycol–inorganic salt–water, are classical techniques for recovering and separating biochemicals such as proteins and nucleic acids. They also appear (especially the polyoxyethylene glycol–inorganic salt–water system) to have some utility for solvent replacement in the extraction of metal ions and small organic molecules. The benefits of ABS are connected with the aqueous environment of both separated phases, which are formed without the involvement of any organic solvent.

In HLLE, a small volume (microliter range) of water-immiscible liquid is separated from 10 to 100 ml of homogeneous aqueous solution containing perfluorooctanoate ion. This phenomenon is induced by a change in pH or by the addition of counter-cations, such as quaternary ammonium ions.



**Figure 7** Cloud point extraction.



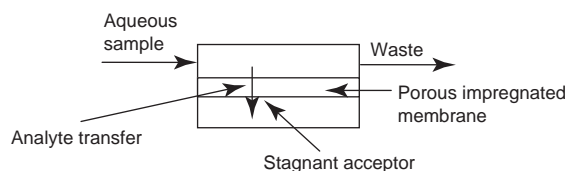
The solute species in the sample solution is concentrated in the sedimented phase, and a 100–1000-fold increase in concentration is achieved. This method is used to concentrate vital compounds such as hemoglobin, vitamin B<sub>12</sub>, and various metal ions. HLLE seems to be a simple and powerful preconcentration method for several instrumentation techniques, including capillary gas chromatography and capillary electrophoresis.

ILs, mainly quaternary ammonium salts with hydrophobic anions, e.g., 1-n-octyl-3-methylimidazolium hexafluorophosphate and tetrafluoroborate, are nonvolatile solvents that can be also applied in extraction. The selection of anions decides the IL hydrophobicity, the formation of a two-phase system with water, the solubility of ILs in water and of water in ILs, and the IL's hygroscopic properties. Up to now, the hydrophobic ILs that form separated phases with water contain potentially dangerous anions, such as PF<sub>6</sub><sup>−</sup> and BF<sub>4</sub><sup>−</sup>, which may decompose and produce fluoride ions. ILs are more viscous than typical solvents and may interfere with the transfer of ionic species. New applications for small-scale separations may emerge in the future.

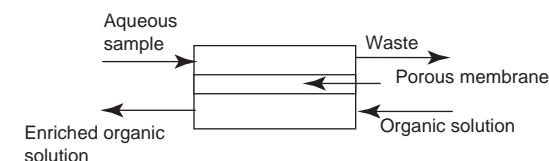
SFE is a well-established process for the recovery of different organics, mainly nonpolar substances from various solid matrices. It allows the selective extraction of different chemicals without an additional clean-up step and uses small sample amounts. Supercritical carbon dioxide (SC-CO<sub>2</sub>) is a convenient solvent, due to its moderate constants ( $T_c = 31^\circ\text{C}$ ,  $P_c = 7.3\text{ MPa}$ ), nontoxicity, nonflammability, sufficient solvation power, and availability in pure form. SC-CO<sub>2</sub> is also used to extract various organic compounds from environmental aqueous samples. Quantitative removal requires the addition of low molecular organic modifiers such as alcohols to enhance the polarity of SC-CO<sub>2</sub>.

Automation and online connection to analytical instruments requires extraction to be performed in a continuous way and simultaneous with the flows of phases. The most successful techniques are MPLLE and SLM extraction (Figure 8). SLM enables extraction and back-extraction to be carried out in one continuous process due to the separation of flowing aqueous phases (the acceptor phase is stagnant in analytical applications) by a hydrophobic membrane, which is usually of polytetrafluoroethylene, and the organic phase immobilized in the pores of the membrane. The organic phase, used in small amounts in comparison with LLE, may contain sophisticated extractants called carriers. Membranes are impregnated by soaking in the organic phase. The analyte present in the donor phase (aqueous sample) is dissolved in the membrane phase or complexed

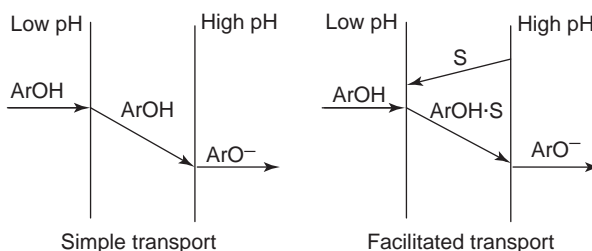
#### SLM extraction



#### MMLLE



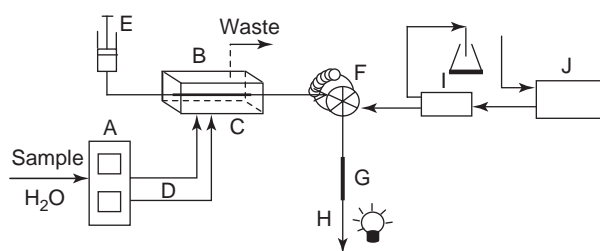
**Figure 8** Supported liquid membrane and microporous membrane liquid–liquid extraction. (Adapted from Jonsson JA and Mathiasson L (1999) Liquid membrane extraction in analytical sample preparation. *Trends in Analytical Chemistry* 18: 318–324.)



**Figure 9** Simple and facilitated transport with solvating carrier of an aromatic compound having a phenolic group.

with the carrier, then transported via diffusion through the pores on the other side of the membrane and transferred to the acceptor phase. The aqueous phases have different acidities to keep analytes in neutral and charged forms in the donor and acceptor phases, respectively, in simple transport, e.g., of amines, phenols, or organic acids, or to form and decompose the complex in facilitated transport with carriers (Figure 9). The latter must be used for extraction of ionic species that are not soluble in the membrane without prior complexation. The SLM extraction is a dynamic process that is usually controlled by diffusion in the pores and that gives separations that are different to those for LLE, which is an equilibrium process. By incorporating specially designed carriers, a very high selectivity can be obtained with an enrichment level of over 100 in environmental samples. Using hollow fibers, enrichments of up to a few thousands were reported.

SLM extraction is well suited to automated interfacing with liquid chromatography and capillary



**Figure 10** Enrichment and separation setup for SLM extraction. A – peristaltic pump, B – membrane device with hollow fiber installed into fused silica capillary, C – inlet on the donor side, D – washing line used after enrichment, E – microinjection pump to transport the acceptor into loop F, G –  $C_{18}$  packed capillary column, H – detector window, I – splitter unit, J – high pressure pump. (Adapted from Thordarson E, Palmarsdottir S, Mathiasson L, and Jonsson JA (1996) Sample preparation using a miniaturized supported liquid membrane device connected on-line to packed capillary liquid chromatography. *Analytical Chemistry* 68: 2559–2563.)

electrophoresis. A miniature SLM device (Figure 10) consists of polypropylene hollow fiber with pores filled with an organic solvent, inserted and fastened in a cylindrical channel of the module. The sample is pumped on the outside of the hollow fiber, and the analyte is selectively enriched and trapped in the fiber lumen. The volume of the stagnant acceptor solution can be as low as 1–2  $\mu\text{L}$ , and is transferred through capillaries to chromatographic systems. Using devices with a flat membrane, the volume of the enriched sample is 10–15  $\mu\text{L}$ .

A new version of CPLLE can be performed as cross- and countercurrent flows of aqueous and organic phases separated by the microporous membrane (MMLLE) in the same devices used for SLM extraction (Figure 8). The organic phase wets the membrane better than an aqueous solution and enters the membrane pores. A leakage of the organic phase can be prevented by applying a suitably higher pressure on the aqueous side.

See also: **Analytical Reagents:** Specification; Purification. **Sampling:** Theory; Practice. **Surfactants and Detergents.**

## Further Reading

- Gopalan AS, Wai CM, and Jackobs HK (eds.) (2003) *Supercritical Carbon Dioxide. Separations and Processes*. ACS Symposium Series 860. Washington: Oxford University Press.
- Huldestone JG, Willauer HD, Griffin ST, and Rogers RT (1999) Aqueous polymeric solutions as environmentally benign solvent extraction. *Industrial & Engineering Chemistry Research* 38: 2523–2539.
- Pramauro E and Pelizzetti E (1996) Comprehensive analytical chemistry. In: Weber SG (ed.) *Surfactants in Analytical Chemistry. Applications of Organized Amphiphilic Media*, vol. 31. Amsterdam: Elsevier.
- Prochaska K (2002) Interfacial activity of metal ion extractants. *Advances of Colloid and Interface Science* 95: 51–72.
- Rogers RD and Seddon KR (eds.) (2002) *Ion Liquids Industrial Applications to Green Chemistry*. ACS Symposium Series 818. Washington: Oxford University Press.
- Rydberg J, Cox M, Musikas C, and Choppin GR (2004) *Solvent Extraction. Principles and Practices*. New York: Dekker.
- Sastre AM, Kumar A, and Singh RK (1998) Improved techniques in liquid membrane separations: An overview. *Separation and Purification Methods* 27: 213–298.
- Szymanowski J (1993) *Hydroxyoximes and Copper Hydrometallurgy*. Boca Raton: CRC Press.
- Szymanowski J (2000) Kinetics and interfacial phenomena. *Solvent Extraction and Ion Exchange* 18: 703–727.
- Takagai Y and Igarashi S (2002) Homogenous liquid-liquid extraction as a simple and powerful preconcentration method for capillary gas chromatography and capillary electrophoresis. *American Laboratory July*: 29–30.
- Thornton JD (ed.) (1992) *Science and Practice of Liquid-Liquid Extraction*. Oxford: Oxford Science Publications, Clarendon Press.
- Zaslavsky BV (1994) *Aqueous Two Phase Partitioning: Physical Chemistry and Bioanalytical Applications*. New York: Dekker.

## Solvent Extraction: Multistage Countercurrent Distribution

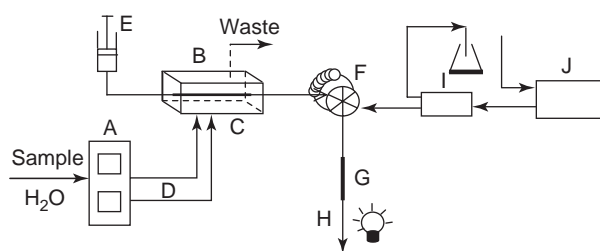
C F Poole, Wayne State University, Detroit, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Separations by partitioning between two immiscible liquids can be made more effective by using a cascade process. Suitable processes can be viewed as

continuous or discontinuous depending on how the liquid phases are contacted. Discontinuous processes are characterized by a number of separate equilibration steps followed by transfer of either or both liquid phases for further contact with the fresh liquid phase. This multistage partitioning process is referred to as a countercurrent distribution. Manual countercurrent distribution separations require little equipment beyond suitable vessels, such as test tubes or separating funnels, and use either decantation or



**Figure 10** Enrichment and separation setup for SLM extraction. A – peristaltic pump, B – membrane device with hollow fiber installed into fused silica capillary, C – inlet on the donor side, D – washing line used after enrichment, E – microinjection pump to transport the acceptor into loop F, G –  $C_{18}$  packed capillary column, H – detector window, I – splitter unit, J – high pressure pump. (Adapted from Thordarson E, Palmarsdottir S, Mathiasson L, and Jonsson JA (1996) Sample preparation using a miniaturized supported liquid membrane device connected on-line to packed capillary liquid chromatography. *Analytical Chemistry* 68: 2559–2563.)

electrophoresis. A miniature SLM device (Figure 10) consists of polypropylene hollow fiber with pores filled with an organic solvent, inserted and fastened in a cylindrical channel of the module. The sample is pumped on the outside of the hollow fiber, and the analyte is selectively enriched and trapped in the fiber lumen. The volume of the stagnant acceptor solution can be as low as 1–2  $\mu\text{L}$ , and is transferred through capillaries to chromatographic systems. Using devices with a flat membrane, the volume of the enriched sample is 10–15  $\mu\text{L}$ .

A new version of CPLLE can be performed as cross- and countercurrent flows of aqueous and organic phases separated by the microporous membrane (MMLLE) in the same devices used for SLM extraction (Figure 8). The organic phase wets the membrane better than an aqueous solution and enters the membrane pores. A leakage of the organic phase can be prevented by applying a suitably higher pressure on the aqueous side.

See also: **Analytical Reagents:** Specification; Purification. **Sampling:** Theory; Practice. **Surfactants and Detergents.**

## Further Reading

- Gopalan AS, Wai CM, and Jackobs HK (eds.) (2003) *Supercritical Carbon Dioxide. Separations and Processes*. ACS Symposium Series 860. Washington: Oxford University Press.
- Huldestone JG, Willauer HD, Griffin ST, and Rogers RT (1999) Aqueous polymeric solutions as environmentally benign solvent extraction. *Industrial & Engineering Chemistry Research* 38: 2523–2539.
- Pramauro E and Pelizzetti E (1996) Comprehensive analytical chemistry. In: Weber SG (ed.) *Surfactants in Analytical Chemistry. Applications of Organized Amphiphilic Media*, vol. 31. Amsterdam: Elsevier.
- Prochaska K (2002) Interfacial activity of metal ion extractants. *Advances of Colloid and Interface Science* 95: 51–72.
- Rogers RD and Seddon KR (eds.) (2002) *Ion Liquids Industrial Applications to Green Chemistry*. ACS Symposium Series 818. Washington: Oxford University Press.
- Rydberg J, Cox M, Musikas C, and Choppin GR (2004) *Solvent Extraction. Principles and Practices*. New York: Dekker.
- Sastre AM, Kumar A, and Singh RK (1998) Improved techniques in liquid membrane separations: An overview. *Separation and Purification Methods* 27: 213–298.
- Szymanowski J (1993) *Hydroxyoximes and Copper Hydrometallurgy*. Boca Raton: CRC Press.
- Szymanowski J (2000) Kinetics and interfacial phenomena. *Solvent Extraction and Ion Exchange* 18: 703–727.
- Takagai Y and Igarashi S (2002) Homogenous liquid–liquid extraction as a simple and powerful preconcentration method for capillary gas chromatography and capillary electrophoresis. *American Laboratory July*: 29–30.
- Thornton JD (ed.) (1992) *Science and Practice of Liquid–Liquid Extraction*. Oxford: Oxford Science Publications, Clarendon Press.
- Zaslavsky BV (1994) *Aqueous Two Phase Partitioning: Physical Chemistry and Bioanalytical Applications*. New York: Dekker.

## Solvent Extraction: Multistage Countercurrent Distribution

**C F Poole**, Wayne State University, Detroit, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Separations by partitioning between two immiscible liquids can be made more effective by using a cascade process. Suitable processes can be viewed as

continuous or discontinuous depending on how the liquid phases are contacted. Discontinuous processes are characterized by a number of separate equilibration steps followed by transfer of either or both liquid phases for further contact with the fresh liquid phase. This multistage partitioning process is referred to as a countercurrent distribution. Manual countercurrent distribution separations require little equipment beyond suitable vessels, such as test tubes or separating funnels, and use either decantation or

pipettes for liquid transfer between vessels. This is a rather slow and tedious process whenever a significant number of equilibration stages are required for the separation. Martin and Synge introduced automated machines for countercurrent distribution in the 1940s. Craig and others improved upon these machines in the 1950s, resulting in widely used apparatuses with up to 1000 cells, although machines in the range of 50–250 cells were more common. At a later date, Albertsson introduced the thin-layer countercurrent distribution apparatus to reduce the settling time for liquid phases of similar density, such as those used in two-phase aqueous systems for the separation of biopolymers and particles. Akerlund improved upon this design by using centrifugal force for the separation of the two liquid phases. The technical development of automated countercurrent distribution machines had stopped by the 1990s, however, and these machines find little use today.

There are several reasons for the demise of countercurrent distribution as a separation technique. Shortly after the development of automated countercurrent distribution machines, Martin and Synge demonstrated that liquid–liquid partition separations could be performed more efficiently and faster by partition chromatography. Initially, paper chromatography, and later thin-layer chromatography, was adopted for analytical applications and columns for preparative separations. Eventually, the emergence of high-pressure liquid chromatography as a major separation technique for both analytical and preparative separations ended interest in multistage countercurrent distribution separations. Almost in parallel, in the 1970s, Ito began the development of several smaller and more efficient devices for countercurrent chromatography. Countercurrent chromatography differs from countercurrent distribution in the use of mechanical forces created by planetary motion of a coiled tube to stabilize one liquid phase, a stationary phase, while the other, a mobile phase, is continuously pumped through it. Distinct separation stages are not employed and the separation system need not be at equilibrium. In addition, it was found that many of the applications traditionally performed by countercurrent distribution could be transferred to countercurrent chromatography without difficulty.

Countercurrent distribution and countercurrent chromatography are distinguished from column chromatography by the absence of a solid stationary phase or support. Sorptive surfaces present few problems for the separation of small molecules but for macromolecules, multiple surface interactions can result in a loss of selectivity and/or denaturation of biopolymers. Cost savings for preparative

applications may be significant for liquid–liquid compared with liquid–solid systems. In addition, the range of additives that can be used in liquid–liquid systems is generally less restrictive and mechanisms more easily understood than in the case of column chromatography.

The general theory of countercurrent distribution is still of general interest since manual liquid–liquid extraction remains a common laboratory technique. In addition, countercurrent distribution models often provide a starting point for the explanation of separations by chromatography and continuous liquid–liquid extraction processes. A favorable feature of countercurrent distribution systems is that separations are completely predictable, once the solute distribution ratios and the phase ratio are known.

## Theory

Liquid–liquid extraction is an equilibrium process between two immiscible phases, described by a equilibrium constant, usually called the distribution constant or partition coefficient. The distribution constant,  $K$ , is defined as the ratio of the concentration of the substance in the two phases at equilibrium ( $K = C_U/C_L$ , where  $C_U$  is the concentration of analyte in the less dense (upper) phase and  $C_L$  its concentration in the more dense (lower) phase). The distribution or partition ratio,  $G$ , defined as the ratio of the solute masses in the two phases ( $G = m_U/m_L$ , where  $m_U$  is the mass of analyte in the less dense phase and  $m_L$  its mass in the more dense phase), is often more useful. The distribution constant and distribution ratio are related through the phase ratio,  $V$ , by  $G = KV$ . The phase ratio is the volume ratio of the liquid phases ( $V = \text{volume of the less dense phase}/\text{volume of the more dense phase}$ ).

A convenient way of analyzing the countercurrent distribution process is to calculate the amounts (in fractions) of a single component in the two phases based on the distribution ratio. In the initial two-phase system (number 0), the substance is distributed with the fraction  $p$  in the upper phase and  $q$  in the lower phase. By definition  $p + q = 1$ . Transfer of the upper phase from cell 0 to the lower phase of the next cell (number 1) results, after equilibration, in the distribution of the transferred fraction according to  $G = p/q$ . The fraction in the upper phase of cell 1 will become  $p^2$  and in the lower phase  $pq$ . A fresh upper phase is added to cell 0. The fraction  $q$  remaining in the lower phase will distribute itself into the upper phase and at equilibrium the fraction in the upper phase will be  $pq$  and in the lower phase  $q^2$ . The resulting distribution after 10 transfers in the



$pq^{10}$	$10p^2q^9$	$45p^3q^8$	$120p^4q^7$	$210p^5q^6$	$252p^6q^5$	$210p^7q^4$	$120p^8q^3$	$45p^9q^2$	$10p^{10}q$	$p^{11}$	Amount in upper phase
$q^{11}$	$10pq^{10}$	$45p^2q^9$	$120p^3q^8$	$210p^4q^7$	$252p^5q^6$	$210p^6q^5$	$120p^7q^4$	$45p^8q^3$	$10p^9q^2$	$p^{10}q$	Amount in lower phase
0	1	2	3	4	5	6	7	8	9	10	$i$
$q^{10}$	$10pq^9$	$45p^2q^8$	$120p^3q^7$	$210p^4q^6$	$252p^5q^5$	$210p^6q^4$	$120p^7q^3$	$45p^8q^2$	$10p^9q$	$p^{10}$	Amount in each tube

**Figure 1** Distribution of a component that partitions in the ratio  $p/q$  between the upper and lower phases after a countercurrent distribution with 10 transfers. The amounts are given in fractions, i.e.,  $p + q = 1$ . The amount in each cell is the combined mass in the upper and lower phases for each cell.

11 cells 0–10 is shown in **Figure 1**. Generally, the amount of material,  $T_{n,i}$  (in fractions), in cell number  $i$  after a countercurrent distribution with  $n$  transfers is given by

$$T_{n,i} = [(n!)/i!(n-i)!][G^i/(1+G)^n] \quad [1]$$

This relation can be used to calculate the distribution profiles for 100 and 1000 transfers with  $G$  values varied in steps between 0.05 and 20 (**Figure 2**). The resolution of peaks with various  $G$  values increases with the increase in the number of transfers ( $n$ ). This is due to the fact that the difference in the position (cell number) of the peaks is proportional to the number of transfers while the peakwidth only increases with the square root of the number of transfers.

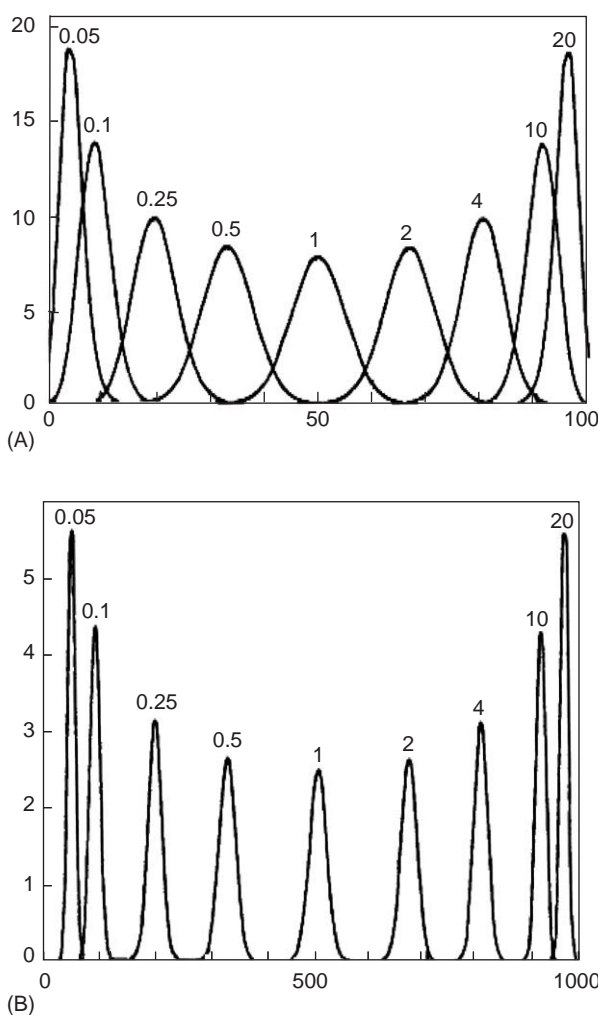
Since virtually all the pure substance is contained in a region extending for  $3\sigma$ -values either side of the peak maximum, the peakwidth,  $W$ , can be approximately calculated by

$$W = 6\sqrt{nG}/(1+G) \quad [2]$$

This is true for symmetrical peaks, which will be the case when either  $n$  is large or  $G \sim 1$ .

The relative peakwidth expressed both as a number of cells and as a percentage of the total number of cells for different values of  $G$  and  $n$ , is summarized in **Table 1**. The space occupied by a single peak as a fraction of the total space available rapidly decreases as the number of transfers increases. As  $G$  becomes either large or small compared to 1, the separation factor approached the limiting values of  $n$  or 0. Thus, substances with large or small values of  $G$  will crowd together at the ends of the countercurrent train.

The most effective separation of two substances with distribution constants  $K_1$  and  $K_2$  is achieved when their peak maxima are symmetrically arranged



**Figure 2** The resulting countercurrent distribution profiles after (A) 100 and (B) 1000 transfers for substances with the indicated  $G$  values. The x-axis indicates the number of transfers ( $n$ ) and the y-axis the value for  $T_{n,i}$  expressed in fractions (see eqn [1]). The x-axis also relates the relative position of a substance with a defined value of  $G$  to the cells which it will occupy after a given number of equilibration steps.

**Table 1** The absolute peak width as the number of cells occupied,  $W$ , and percent of the total number of cells occupied by a peak ( $100W/n$ ) calculated for several values of  $G$  and  $n$ 

$n$	$W$	$100W/n$
$G = 1$		
50	21	42
100	30	30
1000	95	9.5
$G = 5$		
50	16	32
100	23	23
1000	71	7
$G = 10$		
50	12	24
100	17	17
1000	54	5.4

**Table 2** The number of transfers,  $n$ , in a countercurrent distribution necessary to separate two substances, 1 and 2, with a given separation factor  $\alpha$ . The phase ratio is selected such that the equality  $G_1 \cdot G_2 = 1$  is satisfied

Separation factor ( $\alpha$ )	Distribution ratios		Number of transfers ( $n$ )
	$G_1$	$G_2$	
2	0.709	1.414	264
3	0.577	1.732	110
4	0.500	2.00	70
9	0.333	3.00	30
16	0.250	4.00	20
25	0.200	5.00	16
36	0.167	6.00	14
49	0.143	7.00	12
81	0.111	9.00	10
400	0.05	20.0	6

around the middle of the distribution train, i.e., cell  $i = n/2$ . This occurs when the relation  $G_1 \cdot G_2 = 1$  at an optimum phase ratio of  $V = 1/\sqrt{(K_1 K_2)}$ . The ratio of  $K_1/K_2$  is the separation factor,  $\alpha$ , and is a measure of the ease of separating the two substances. The number of transfers necessary for virtually complete (more than 95%) separation for various  $\alpha$  values is calculated in Table 2.

The  $G$  value of a substance that has its maximum peak value in cell number  $i$  after  $n$  transfers is given approximately by  $G = i/(n - i)$ . The  $G$  value can also be obtained by comparing the amount of a substance in two consecutive cells, numbers  $i$  and  $i + 1$  by combining the expressions for  $T_{n,i}$  and  $T_{n,i+1}$

$$G = [T_{n,i+1}/T_{n,i}][(i + 1)/(n - i)] \quad [3]$$

The determination of  $G$  values from experimental data is of considerable importance, as it is the first step in the calculation of theoretical distribution curves to fit the data. When the  $G$  value is known,

the entire distribution curve can be calculated for any chosen value of  $n$  from eqn [1] using successive values of  $i$  from 0 to  $n$ . Comparison of the apparent  $G$  values over a distribution peak allows detection of the presence of impurities with slightly different  $G$  values.

## Operational Aspects

The standard method of operation in countercurrent distributions is known as the fundamental procedure. This basic process can be modified or extended in several ways.

### Fundamental Procedure

Shaking or agitation is used to equilibrate the sample between the phases in the first cell of the machine so that one phase becomes finely dispersed in the other. The fine dispersion facilitates the attainment of equilibrium by greatly increasing the interfacial area between the liquid phases. When the equilibration process is complete, agitation is stopped and the liquid phases allowed to separate. One phase (usually the upper) is then transferred quantitatively to the next cell of the apparatus, where it is brought into contact with a fresh volume of the stationary phase. Simultaneously, a fresh portion of upper (or lower) phase is introduced into the first cell of the apparatus, and agitation recommenced to bring about a second equilibrium of the sample between the phases in both cells. The alternate equilibration and transfers with introduction of the fresh mobile phase at each transfer are repeated until the initial portion of mobile phase has reached the last cell in the apparatus. The distribution is then complete and the cells emptied individually or grouped according to their contents.

### Recycle

If an appreciable section of the apparatus contains no sample components after completion of the fundamental procedure, the resolving power of the apparatus is not being fully utilized. This sample-free section, however, may be used for separation after completion of the fundamental procedure by making one more transfer operation. The mobile phase from the last ( $n$ th) cell being reintroduced into the first (cell 0) of the apparatus. The process can be continued in principle until the leading boundary of the fastest-moving component is about to overtake the trailing boundary of the slowest component. At this stage, all the cells of the apparatus contain sample components, and its potential resolving power is fully utilized. The recycle mode can be combined with



any of the withdrawal procedures (see below). In this way, completely resolved components are removed continuously from the apparatus in the mobile phase with fresh mobile phase added to replace the withdrawn phase. The calculation of theoretical curves for the recycling method are identical with those for the fundamental procedure, except that the number of transfers applied is greater than the number of cells in the machine.

### Single and Double Withdrawals

In the single withdrawal procedure, also known as elution countercurrent distribution, the number of transfers is unlimited. The fundamental procedure is completed as usual followed by one more equilibration and settling of the phases, with transfer and withdrawal of a portion of upper (or lower) phase. Unlike the method of complete withdrawal, a new portion of upper (or lower) phase is introduced into cell 0. The processes of equilibration, settling, and transfer are carried out again and a second portion of the upper (or lower) phase is introduced into cell 0. The process is repeated for any required number of cycles, or until all the sample components have emerged in the withdrawn phase. Double and alternate withdrawals of upper and lower phases are also possible. The combination of recycle and single withdrawal procedures is the most important countercurrent distribution technique.

The different method of operation makes the single withdrawal procedure suitable for the separation of samples with a different range of distribution ratios to those commonly separated by the fundamental procedure. Substances with high  $G$  values are easily eluted in the upper phase by this method while those with low  $G$  values need a greater number of transfers to be carried through the countercurrent distribution train. The opposite is true when the lower phase is withdrawn. The greater the number of transfers, the more diluted the sample components are when eluted from the countercurrent distribution apparatus.

### Continuous Feed

To increase sample throughput for preparative applications requires a continuous feed of sample solution into the distribution apparatus. In O'Keefe's method, the sample is introduced at or near the center of the distribution apparatus. After equilibration and settling, all upper phases are moved to the next cell at the left and all lower phases to the next cell at the right. The upper and lower phases from the extreme cells on the left and right are collected. Fresh upper and lower phases are added to these cells to restore the original number of systems. After

addition of a new portion of sample to the center of the apparatus, the cycle is repeated. Substances with  $G > 1$  will be recovered in the upper phases collected to the right and substances with  $G < 1$  in the lower phases removed to the left. Typically, the number of cells employed is small (e.g., 11 or less). The method is used primarily to purify and concentrate a single component in a complex mixture or to separate two-component mixtures (pure components withdrawn in the upper and lower phases).

A similar process, called Watanabe–Morikawa partitioning, employs a sample feed at the end of the apparatus. This is useful when only components with high or low  $G$  values are to be isolated.

### Gradient

In gradients, for example, ionic strength, pH, complexing agent, affinity ligands, etc., can be introduced using the single withdrawal method by sequentially adjusting the composition of the new upper (or lower) phase added to cell 0. A gradient can be started with the first cycle or at any cycle number in the separation.

### Solvent Systems

Solvent systems suitable for countercurrent distribution must always form two discrete liquid phases, and a finally divided suspension in one phase in the other must separate quickly into two bulk-liquid layers. These properties must be retained in the presence of appreciable amounts of sample. The two phases should, therefore, differ in density, neither should have a high viscosity, while the interfacial tension should not promote the formation of stable emulsions. Formation of emulsions, which separate only with difficulty with real samples, is the most serious practical problem. Solvent systems should be selected based on their selectivity for the separation. In addition, distribution ratios can be optimized by changing the phase ratio or distribution properties of the phases by using additives (e.g., pH, complexing agents, salts, polymers, etc.).

### Two-Phase Aqueous Systems

Separation by partition between two immiscible aqueous phases has been successfully applied to a large number of proteins, nucleic acids, and cellular particles such as mitochondria, chloroplasts, membrane vesicles, and whole cells. The phase systems are made up of two different water-soluble polymers, dextran and poly(ethylene glycol), dissolved in water and complemented with suitable buffers, salts, and sucrose. Highly selective adjustment of the distribution process is possible by linking an affinity

ligand to one of the phase-forming polymers. This then serves to concentrate the selected substances in one of the phases. In two-phase aqueous systems, particles are separated according to their surface properties, for example, surface charge, hydrophobicity, and biospecific interactions. In this case, the particles are induced to partition between the two liquid phases and the interface between them. Particles with similar properties are accumulated in the interface region. The main devices for countercurrent distribution on a laboratory scale using two-phase aqueous systems are the thin-layer and centrifugal countercurrent distribution apparatuses. The viscosity of the phases and long settling times sometimes limits the proper functioning of traditional distribution machines. Countercurrent distribution methods complement more traditional centrifugation techniques for the separation of particles.

## Apparatus

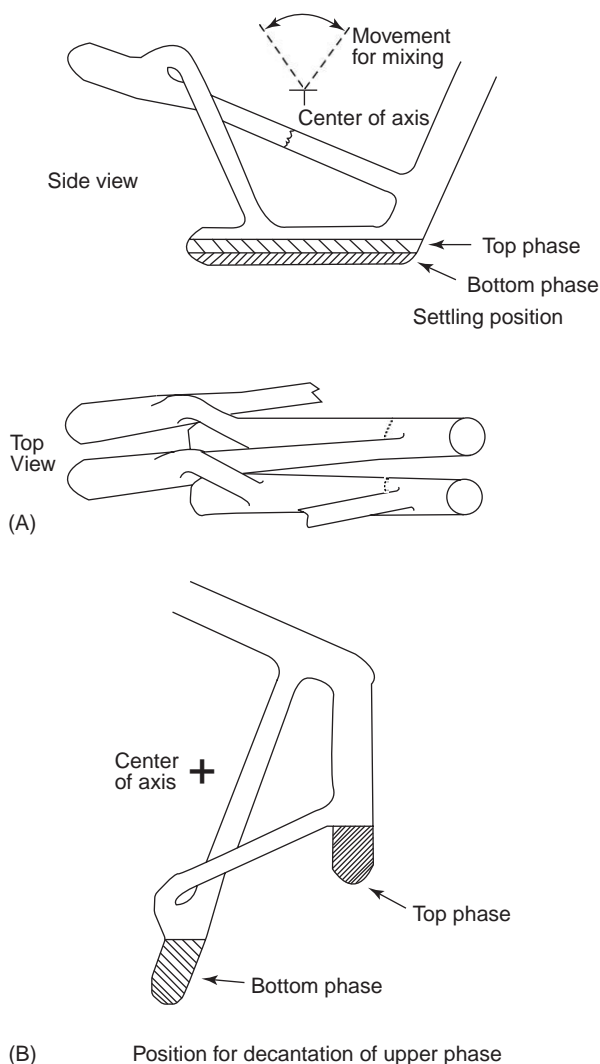
Countercurrent distribution with only a few transfers (up to  $\sim 10$ ) is easily carried out using a set of separating funnels for mixing, settling, and phase transfer. For small phase volumes, 0.5–5 ml, the separating funnels can be replaced by test tubes and the upper phases transferred by pipette. For countercurrent distributions with more than 10 transfers, some kind of automated apparatus is recommended. Some examples of suitable apparatuses are presented below.

### Craig-Type Apparatus

Several different types of countercurrent distribution machines have been constructed for use with organic solvents and water. The most widely known design is the all-glass machine of Craig and Post. Several glass units, allowing mixing, settling, and phase transfers are arranged in batteries on a horizontal axis (Figure 3). Movement about the axis can be used to gently mix the phases, to put the glass units in position for settling of the phases, and for decanting the upper (or lower) phases, respectively. The time for each segment of the operation cycle as well as the number of transfers can be programmed. Standard designs contain from 25 to 250 cells, while more advanced machines with up to 1000 cells have been described. The glass cells are available in various sizes. Standard-size cells have space for  $\sim 2$  ml (fixed amount) of lower phase and up to 5 ml of upper phase.

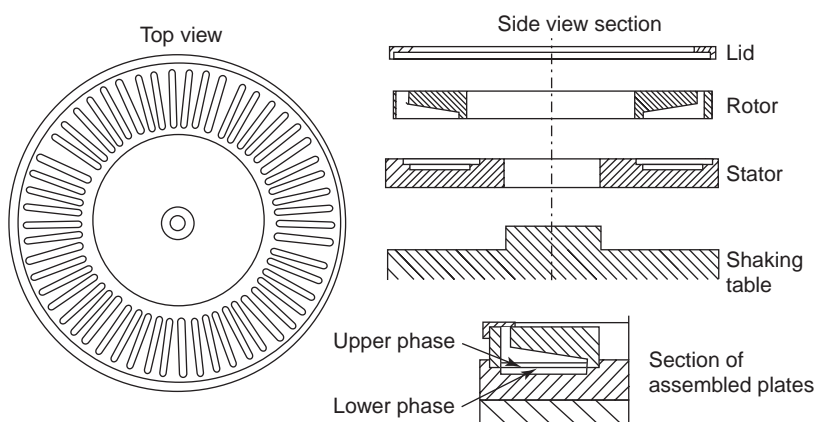
### Thin-Layer Apparatus

The thin-layer apparatus were developed for use with two-phase aqueous systems for both liquid–liquid



**Figure 3** Section of the Craig-type apparatus for countercurrent distribution. (A) Two segments (or cells) are shown from side and top views. (B) By turning the axis, the upper phases can be decanted. When returning to the original position, the upper phases are transferred to the neighboring cell.

and liquid–interface countercurrent distributions. A drawback of the Craig-type apparatuses for use with two-phase aqueous systems is the long time needed for phase separation to occur, which is a result of the small density difference between phases and their high viscosity. To circumvent this problem Albertsson constructed an apparatus in which the phases form thin layers of  $\sim 2$  mm, thereby reducing the settling time by a factor of  $\sim 10$ . Thin-layer countercurrent distribution takes place in a partition cell block consisting of two cylindrical Plexiglass plates (Figure 4). The lower plate (stator) is fixed while the upper one (rotor) can be rotated stepwise. Both plates have radially oriented cavities, which pair-wise form containers for the two-phase system. The cavities of the stator contain the lower phases

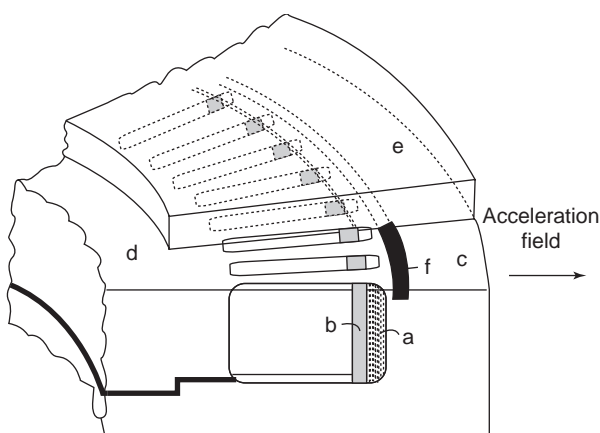


**Figure 4** Thin-layer countercurrent distribution apparatus. The two Plexiglass disks contain matching cavities for the upper and lower phases and by rotation the upper plate can move all upper phases relative to the lower ones.

while the upper phases are situated in the cavities of the rotor and therefore can be transferred relative to the lower phases by rotating the upper plate. The partition block rests on a shaking table, which also contains a drive for the upper plate rotation. The upper cavities are deeper than the lower cavities to allow air space for efficient mixing. A planetary movement of the shaking table achieves mixing of the phase systems. The phase systems are added to each container (chamber) via openings in the upper side of the rotor. These openings, during the run, are covered with a ring-formed lid. All procedures including shaking, settling, and phase transfer are automated by a control unit. After completion of the countercurrent distribution, the two-phase systems are collected by decanting into a fraction collector.

### Centrifugal Apparatus

Akerlund developed a fully automated apparatus utilizing the increased speed of phase separation obtained in a centrifugal field. Instead of using chambers with low height, the time of phase separation is reduced by centrifugation at  $\sim 100g$  after each equilibration step. The operating unit of the machine consists of an outer ring with 60 cavities for the lower phase and an inner plate with corresponding cavities for the upper phase (Figure 5). After mixing, by shaking on a platform in the machine, the plates start to rotate to achieve centrifugal force. While still rotating, the transfers of all the upper phases with respect to the lower phases are carried out by a rotational movement of the inner plate relative to the ring. After the transfer, the centrifugation is stopped and a new cycle begins with the shaking process. After completion of the distribution process, the phases are collected by pipette. With this machine, countercurrent distributions can be carried out with



**Figure 5** Centrifugal countercurrent distribution apparatus. A section of the circular separation unit is shown. It is comprised of four units: c, the outer ring with cavities for the lower phase; d, the inner plate with cavities for the upper phase; e, the lid; and f, an O-ring for sealing. The position of the two-phase system during centrifugation is shown with the upper phase, a, and the lower phase, b. (Reproduced with permission from Johansson G, Akerlund H-E, and Olde B (1984) *Journal of Chromatography* 311: 277–289).

systems of high viscosity or when the difference in phase densities is quite low.

**See also:** **Countercurrent Chromatography:** Overview; Solvent Extraction with a Helical Column. **Extraction:** Solvent Extraction Principles.

### Further Reading

Albertsson P-A and Hultin E (1987) Elution countercurrent distribution. *Analytical Biochemistry* 168: 218–223.  
Ito Y and Mandava NB (eds.) (1988) *Countercurrent Chromatography: Theory and Practice*. New York: Dekker.

Morris CJOR and Morris P (1976) *Separation Methods in Biochemistry*, 2nd edn, pp. 638–702. London: Pitman.

Sutherland IA, de Folter J, and Wood P (2003) Modelling CCC using an eluting countercurrent distribution model. *Journal of Liquid Chromatography and Related Technologies* 26: 1449–1474.

Truost H and Johansson G (1996) Fractionation of wheat proteins by counter-current distribution using aqueous two-phase systems containing propionic acid. *Journal of Chromatography B* 680: 71–80.

Walter H and Johansson G (eds.) (1994) Aqueous two-phase systems. In: *Methods in Enzymology*, vol. 228. San Diego: Academic Press.

## Microwave-Assisted Solvent Extraction

**J L Luque-García**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

From simply heating up a meal to preparing a sample for analysis via a multistep process, microwave technology finds various uses in both homes and analytical laboratories. It can help reduce the time taken to complete many processes, and so it is very useful with many types of samples. How microwaves speed up some processes is not fully understood, however. The main factors to be considered here include sample transport properties, molecular agitation, the heating of solvents above their boiling points, and, in some cases, product selectivity.

Analytical chemists began using microwaves in 1975 to digest biological samples. They used domestic microwave ovens then to heat a mixture of a sample and digestion acids rapidly to its atmospheric boiling point in an Erlenmeyer flask. Since this, microwave-assisted methods have been used widely in analytical chemistry. The latest advances in the use of microwaves in various chemical fields include sample digestion for elemental analysis, solvent extraction, sample drying, moisture measurement, analyte adsorption and desorption, cleaning up samples, chromogenic reactions, solid-phase retention, elution, distillation, microwave plasma atomic spectrometry, and synthetic reactions, among others.

Microwave-assisted extraction (MAE) is the process by which microwave energy is used to heat solvents in contact with samples (mainly solid samples) and to partition compounds of interest from the sample into an extractant. Extractions are done in closed or open microwave-transparent vessels where a solvent and sample are combined and then exposed uniformly to microwave energy. Partitioning may occur by any one or several of the following three heating mechanisms: (1) a single solvent or mixture of solvent possessing high dielectric loss coefficients;

(2) solvent mixtures of high and low dielectric loss; and (3) susceptible samples with a high dielectric loss in a solvent of low dielectric loss.

Although the traditional Soxhlet and solvent extraction techniques are widely accepted, they have inherent limitations and problems. Thus, Soxhlet extraction requires 12–24 h in most cases and uses high volumes of organic solvents (hundreds of milliliters). In contrast to conventional methods, MAE can reduce the extraction time to less than 30 min and solvent consumption to under 50 ml. Moreover, the recoveries obtained with MAE are mostly comparable with those provided by alternative methods.

Unlike digestion, MAE is used mainly with soil and sediments rather than with biological samples. As regards analytes, MAE is applied primarily to organic pollutants such as polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), and pesticides rather than to metal species, where other extraction alternatives, such as ultrasound-assisted extraction, are usually more effective.

This article describes the principal applications of microwaves to sample extraction, with special emphasis on solid samples, which have been the most used so far. The description is preceded by a discussion of the equipment used, which can be of the open or closed type, depending on whether they operate at atmospheric pressure or at a higher level and whether they use multimode or focused microwaves. Selected special laboratory designs are also commented on, as is the possibility of coupling of some of these devices to subsequent steps of the analytical process. Before applications are dealt with, the main variables affecting MAE are also examined.

### Microwave-Assisted Extractors

Microwave-assisted extractors can be classified into two groups, according to the way microwave energy is applied to the sample. In multimode systems, the microwave radiation is dispersed randomly in a cavity, and so every zone in the cavity and the sample



Morris CJOR and Morris P (1976) *Separation Methods in Biochemistry*, 2nd edn, pp. 638–702. London: Pitman.

Sutherland IA, de Folter J, and Wood P (2003) Modelling CCC using an eluting countercurrent distribution model. *Journal of Liquid Chromatography and Related Technologies* 26: 1449–1474.

Truost H and Johansson G (1996) Fractionation of wheat proteins by counter-current distribution using aqueous two-phase systems containing propionic acid. *Journal of Chromatography B* 680: 71–80.

Walter H and Johansson G (eds.) (1994) Aqueous two-phase systems. In: *Methods in Enzymology*, vol. 228. San Diego: Academic Press.

## Microwave-Assisted Solvent Extraction

**J L Luque-García**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

From simply heating up a meal to preparing a sample for analysis via a multistep process, microwave technology finds various uses in both homes and analytical laboratories. It can help reduce the time taken to complete many processes, and so it is very useful with many types of samples. How microwaves speed up some processes is not fully understood, however. The main factors to be considered here include sample transport properties, molecular agitation, the heating of solvents above their boiling points, and, in some cases, product selectivity.

Analytical chemists began using microwaves in 1975 to digest biological samples. They used domestic microwave ovens then to heat a mixture of a sample and digestion acids rapidly to its atmospheric boiling point in an Erlenmeyer flask. Since this, microwave-assisted methods have been used widely in analytical chemistry. The latest advances in the use of microwaves in various chemical fields include sample digestion for elemental analysis, solvent extraction, sample drying, moisture measurement, analyte adsorption and desorption, cleaning up samples, chromogenic reactions, solid-phase retention, elution, distillation, microwave plasma atomic spectrometry, and synthetic reactions, among others.

Microwave-assisted extraction (MAE) is the process by which microwave energy is used to heat solvents in contact with samples (mainly solid samples) and to partition compounds of interest from the sample into an extractant. Extractions are done in closed or open microwave-transparent vessels where a solvent and sample are combined and then exposed uniformly to microwave energy. Partitioning may occur by any one or several of the following three heating mechanisms: (1) a single solvent or mixture of solvent possessing high dielectric loss coefficients;

(2) solvent mixtures of high and low dielectric loss; and (3) susceptible samples with a high dielectric loss in a solvent of low dielectric loss.

Although the traditional Soxhlet and solvent extraction techniques are widely accepted, they have inherent limitations and problems. Thus, Soxhlet extraction requires 12–24 h in most cases and uses high volumes of organic solvents (hundreds of milliliters). In contrast to conventional methods, MAE can reduce the extraction time to less than 30 min and solvent consumption to under 50 ml. Moreover, the recoveries obtained with MAE are mostly comparable with those provided by alternative methods.

Unlike digestion, MAE is used mainly with soil and sediments rather than with biological samples. As regards analytes, MAE is applied primarily to organic pollutants such as polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), and pesticides rather than to metal species, where other extraction alternatives, such as ultrasound-assisted extraction, are usually more effective.

This article describes the principal applications of microwaves to sample extraction, with special emphasis on solid samples, which have been the most used so far. The description is preceded by a discussion of the equipment used, which can be of the open or closed type, depending on whether they operate at atmospheric pressure or at a higher level and whether they use multimode or focused microwaves. Selected special laboratory designs are also commented on, as is the possibility of coupling of some of these devices to subsequent steps of the analytical process. Before applications are dealt with, the main variables affecting MAE are also examined.

### Microwave-Assisted Extractors

Microwave-assisted extractors can be classified into two groups, according to the way microwave energy is applied to the sample. In multimode systems, the microwave radiation is dispersed randomly in a cavity, and so every zone in the cavity and the sample

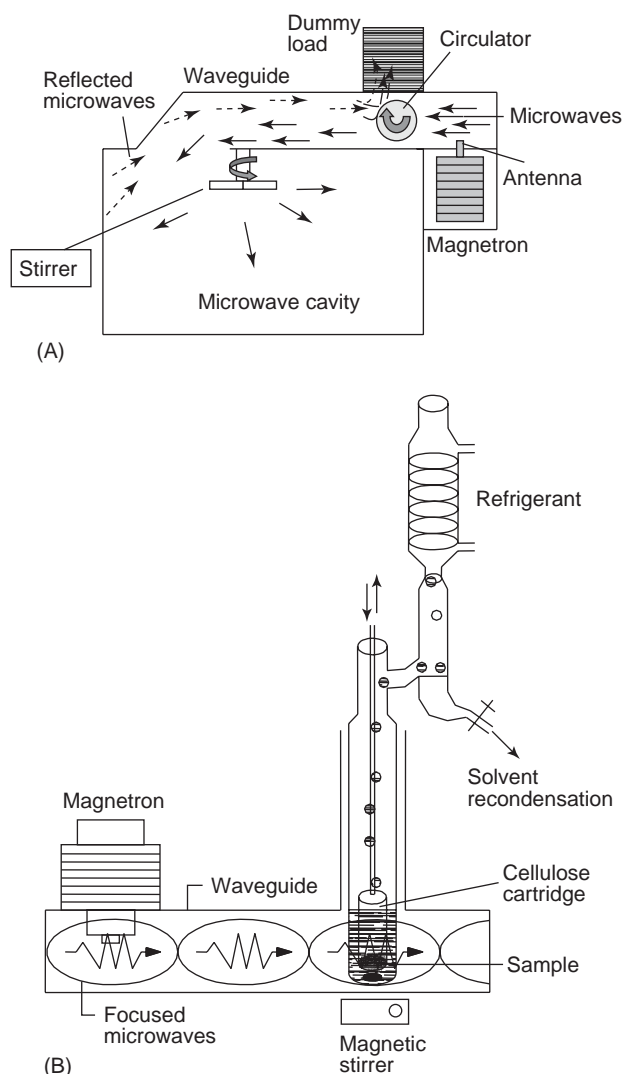
it contains is irradiated. The single-mode or focused system allows focusing of the microwave radiation on a restricted area where the sample is placed for subjection to a much stronger electromagnetic field than in the previous case. **Figure 1** shows a general scheme of both types of extractors. Usually multimode systems (**Figure 1A**) are of the closed-vessel type, where the microwave-assisted treatment is conducted at a high pressure, whereas focused systems (**Figure 1B**) are of the open-vessel type, in which microwaves are applied at atmospheric pressure. This mutual association, however, is incorrect as some devices that operate at a high pressure use focused microwaves, and domestic ovens have been used with multimode sample irradiation but at atmospheric pressure.

Closed-vessel systems have not been used so much for sample extraction as they are primarily used for

sample digestion, whereas the main application of open-vessel systems has been solid sample extraction.

Both multimode and focused microwave devices consist of four major components:

1. The 'magnetron' or microwave generator, where the microwave energy is produced.
2. The waveguide, used to propagate the microwaves from the magnetron to the microwave cavity.
3. The applicator, where the sample is placed. It can be a multimode cavity where microwaves are randomly dispersed or the waveguide itself. In the latter case the sample vessel is placed directly inside to focus the microwave radiation onto the sample.
4. The circulator, which allows microwaves to pass only in the forward direction.



**Figure 1** Scheme of a (A) multi-mode and (B) focused microwave system. (Reproduced with permission from Springer.)



### Closed-Vessel Microwave Systems

The earliest microwave systems for analytical purposes were closed vessels with a multimode cavity. They usually allowed processing of several samples at the same (in a carousel) under pressure and temperature feedback control. Closed vessels exist basically in two different forms. One encompasses noninsulated, relatively thin, single-walled fluoropolymer vessels. These vessels have minimal insulating characteristics and allow large amounts of heat to escape. The other type of closed vessel is a well-insulated container, usually of very thick-walled fluoropolymer, or one with a very thick outer layer or casing (or both). These vessels retain heat very efficiently, and so they do not allow rapid cooling when ambient air is forced over them within the microwave cavity.

Most commercially available closed-vessel microwave systems are based on multimode microwaves; however, the advantages of high-pressure vessels and focused microwave heating have led to the development of systems that combine both assets. These focused high-pressure, high-temperature microwave extractors consist of an integrated closed vessel and a focused microwave-heated system operating at a very high pressure and temperature.

Dynamic systems for high-pressure microwave extraction were developed much later than open-vessel systems because operating under a high pressure reduces the flexibility afforded by working at atmospheric pressure. In 1999, a microwave-assisted flow digestion system was developed from a novel pressure equilibration system that affords temperatures up to 250°C and pressures in the region of 35 bar with previously slurried solid samples. More recently, Ericsson and Colmsjö developed a dynamic high-pressure microwave-assisted extraction device that allows extraction of solid samples, while avoiding the principal drawback of the previous approach (the relative complexity of the assembly and the inability to introduce unslurried solid samples directly).

### Open-Vessel Microwave Systems

The first completely re-engineered laboratory-focused microwave system was introduced by Prolabo in 1986. Most commercial open-vessel microwave systems manufactured since then are of the focused-microwave type, i.e., they use the waveguide as a single-mode cavity. Since their introduction, they have widely been used for sample extraction, substituting in most cases the closed-vessels systems, which as of now are used mainly for carrying out sample digestions.

The principle of focused microwaves is efficient in terms of energy transfer. The electromagnetic energy

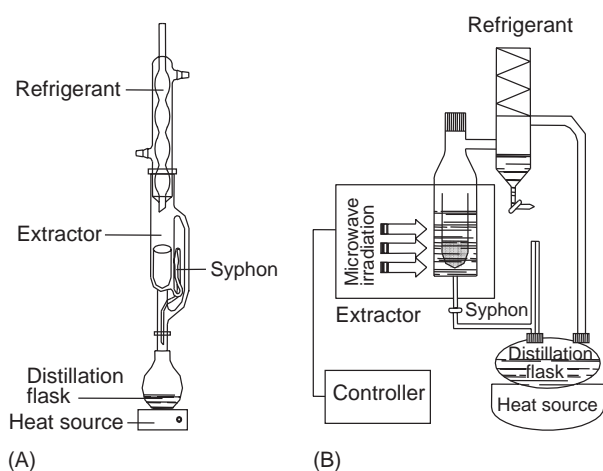
and the sample are highly efficiently coupled, and a high density is obtained as a result. In fact, according to Prolabo, the coupling efficiency is increased by a factor of 10 compared with the use of a multimode cavity. Moreover, these open systems allow users to add reagents or fresh solvent during the extraction or to connect a device such as a reflux system. In these open devices, a continuously adjustable percentage of the maximum power is used over a given period of time. This is different from a multimode cavity, where pulsed power is used predominantly. One of the main drawbacks of these designs is that they allow the use of only one flask at a time. Automation in some systems can enable sequential use of flasks. One recent development involves the use of four flasks at a time by splitting the microwave energy symmetrically among the flasks at the end of each waveguide.

As noted earlier, not all open-vessel systems are of the focused type. A number of reported applications used a domestic multimode oven to process the samples for analytical purposes, usually with a view to coupling the MAE to some other step of the analytical process (usually the determination step) as described below.

The increased flexibility of open-vessel systems has promoted the design of new microwave-assisted sample treatment units based on focused or multimode (domestic) ovens adapted to the particular purpose. Regarding MAE, examples of these new units include the microwave-ultrasound combined extractor and the focused microwave-assisted Soxhlet extractor.

The microwave-ultrasound combined system has been constructed from a Prolabo Maxidigest 350 single-mode microwave oven, in which a sonotrode is placed at the base for indirect ultrasonic agitation of the sample. The ultrasonic probe is placed at a sufficient distance from the electromagnetic field to avoid interactions and short-circuits. The use of this system has allowed the digestion of  $\text{Co}_3\text{O}_4$  and olive oil in shorter times than those required by the conventional method (from 3 to 1 h and from 45 to 30 min, respectively).

The focused microwave-assisted Soxhlet extractor design is based on the same principles as a conventional Soxhlet extractor (Figure 2A) modified to facilitate accommodation of the sample cartridge compartment in the irradiation zone of a microwave oven. The latter has also been modified by making an orifice at the bottom of the irradiation zone, thus enabling connection of the cartridge compartment to the distillation flask through a siphon as illustrated in Figure 2B. The use of this device has provided better results than conventional Soxhlet extraction



**Figure 2** (A) Conventional Soxhlet extractor. (B) Focused microwave-assisted Soxhlet extractor from Prolabo. (Reproduced with permission from Ericsson M and Colmsjö A (2000) Dynamic microwave-assisted extraction. *Journal of Chromatography* 897: 279; © Elsevier.)

in much shorter periods of time. It enables focused microwave-assisted Soxhlet extraction (FMASE), retains the advantages of conventional Soxhlet extraction while overcoming restrictions such as a long extraction time, unsuitability for automation, waste of large volumes of organic solvents, and nonquantitative extraction of strongly retained analytes due to the easier cleavage of analyte–matrix bonds by interactions with focused microwave energy, and, unlike a conventional Soxhlet extractor, allows recycling up to 75–85% of the total extractant volume. In addition, solvent distillation in the FMAS extractor is achieved by electrical heating, which is independent of the extractant polarity, thus avoiding the principal problem of commercial focused microwave devices such as those of the Soxwave series from Prolabo.

## Variables Influencing MAE

The performance in microwave-assisted processes depends on a number of variables including the microwave power output, exposure time, solvent used, and sample size used.

### Microwave Power and Exposure Time

The principal variables to be considered in developing an MAE method are the applied power and the irradiation (exposure) time. The power and the corresponding time will depend on the type of sample and solvent used. Usually, these two variables have opposing effects: for a given process, the use of high microwave power affords a decreased exposure time;

on the other hand, the use of low power entails irradiating the sample for a longer time to apply the same amount of energy.

In theory, one should use a high microwave power to reduce the exposure time as much as possible. Occasionally, however, using a very high power is inadvisable; such as in the case of extraction processes, where complete dissolution of the sample should be avoided in order to minimize the effect of potential interference. In some cases, using a very high microwave power decreases the extraction efficiency through degradation of the sample or its analytes or rapid boiling of the solvent in open-vessel systems (which hinders contact with the sample). Sometimes, the application of microwaves over short, intermittent periods results in substantially increased extraction efficiency relative to the continuous application of microwaves, which causes the solvent to boil and hinders sample–extractant contact.

### Temperature and Pressure

Temperature is a key variable in most analytical processes. In MAE, it plays a prominent role and affects the rate of some reactions, the degradation of thermolabile species and the solubilization of some substances, among others. A number of devices have been developed for monitoring or even controlling the temperature.

Pressure is a highly influential factor in closed-vessel systems. The development of vessels capable of withstanding pressures as high as 41 bar has enabled extraction at very high temperatures, thereby increasing dramatically the extraction efficiency and decreasing the exposure time.

### Type of Solvent

The choice of solvent in a microwave-assisted sample treatment is a direct function of the type of treatment used. Thus, digestion and hydrolysis are best done with aqueous solutions of pure or mixed acids, whereas selective extraction of organic compounds from a sample matrix usually requires an organic solvent.

MAE of nonpolar compounds usually entails the adoption of a compromise since, although these compounds are more readily dissolved in nonpolar solvents such as n-hexane, the interaction of microwaves with the solvent depends on its dielectric constant (the greater  $\epsilon$  is, the stronger the interaction); this requires the use of a mixture of solvents of different polarity in many cases. For example, the most suitable solvent for MAE of organic compounds such as PAHs and PCBs is a mixture of hexane and acetone. In specific cases such as in the extraction

of organometal compounds, the use of methanol acidified with acetic acid allows organometals to be extracted in a rapid, efficient manner with no degradation. In some cases such as the extraction of organic compounds of medium or high polarity, pure water can be used as an efficient 'clean' extractant. Although its power of solubility for organic compounds is not as high as those of organic solvents, its interaction with microwaves is much higher due to its high dielectric constant, thus providing high extraction efficiencies and environmentally friendly methods.

### Influence of Sample Viscosity and Sample Size on Microwave Heating

The viscosity of a sample reflects its ability to absorb microwave energy because it affects molecular rotation. The effect of viscosity is best illustrated by considering ice. When water is frozen, the water molecules become locked in a crystal lattice. This restricts molecular mobility greatly and makes it difficult for the molecules to align with the microwave field. Thus, the dielectric dissipation factor of ice is low. When the temperature of the water is increased to 27°C, the viscosity decreases and the dissipation factor rises to a much higher value.

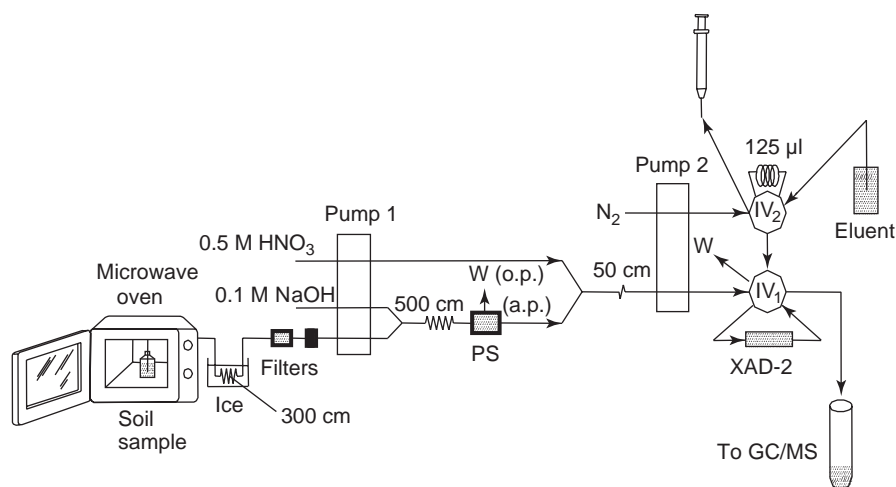
The input microwave frequency also affects the penetration depth of the microwave energy. At a given input frequency, the greater the dissipation factor of a sample is, the less it will be penetrated by microwave energy. In large samples with high dissipation factors, the heating that occurs beyond the penetration depth of the microwave energy is due to

thermal conductance through molecular collisions. Therefore, temperatures at or near the surface will be higher. Because boiling and other agitation increase the rate of thermal conductance, surface heating is not a problem unless penetration is very low and sample heating is extremely superficial. In such a case, heat losses through the vessel walls can be significant and increase the sample heating time.

Although the small sample size used in most analytical dissolution processes has some advantages, it also has at least one disadvantage: the amount of microwave energy absorbed decreases with decreasing sample size. With small sample sizes, a substantial amount of energy is not absorbed but is reflected. Reflected energy can damage the magnetron, and so, in using small samples for analytical work, it is advisable to employ microwave systems designed to protect the magnetron from reflected power.

### Coupling of MAE with Other Operations of the Analytical Process

Online systems using domestic microwave ovens are mainly used for sample digestion. Slurried samples are introduced in a Teflon coil placed in the microwave oven, and after digestion, the solutions are driven to the detector (usually of the atomic type). However, some methods use a domestic microwave oven for extraction purposes, in which the sample is not introduced as a slurry but rather placed in an appropriate vessel that is then introduced into the oven. **Figure 3** shows one of these systems, where the microwave extraction is coupled to subsequent steps



**Figure 3** Experimental design for continuous MAE, liquid-liquid extraction, sorption/cleanup of phenol compounds in soil samples. IV, injection valve; PS, membrane phase separator; o.p. and a.p., organic and aqueous phases; W, waste. (Reproduced with permission from Ericsson M and Colmsjö A (2000) Dynamic microwave-assisted extraction. *Journal of Chromatography* 877: 141; © Elsevier.)

via a flow injection (FI) manifold, thus providing partial automation of the analytical process.

Commercial focused microwave devices have been used for coupling the extraction with other steps such as filtration, preconcentration, and/or detection. The experimental setup showed in **Figure 4** uses a piston pump as the interface between the extractor and an FI manifold. Extraction is performed by supplying an appropriate extractant in an automatic manner by means of the piston pump. Then, the sample is irradiated with microwaves for a preset time and the extract aspirated through the other pump channel to a vessel connected to the FI manifold, where the analytes are filtered, preconcentrated, and determined chromatographically in an automatic manner.

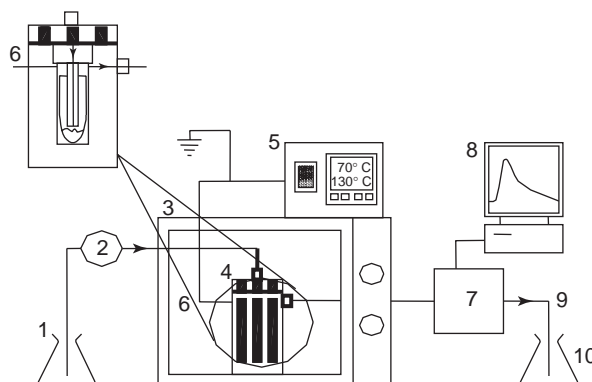
The dynamic high-pressure MAE system described previously also allows the extraction to be coupled to a subsequent stage by connecting the extractor outlet to the unit concerned. **Figure 5** illustrates one possible coupling: connecting the extractor to a fluorimetric detector allows the kinetics of the extraction of PAHs from sediments to be monitored.

## Applications of MAE

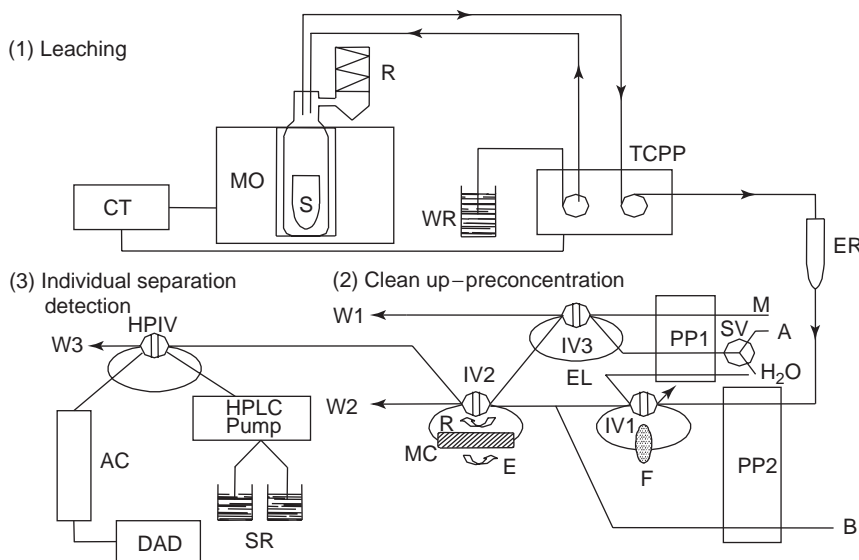
As stated before, MAE has been used mainly for extraction of solid samples. The use of MAE is a continuously expanding area of research at present. As a

result, it is difficult to provide a completely updated overview. Rather, selected groups of compounds of interest that have been subjected to MAE are discussed here.

Traditionally, the extraction of PAHs from solid samples has been conducted using a Soxhlet method, which, while being efficient, usually takes more than 6 h and uses large volumes of solvent. Several reports have indicated that PAHs can be extracted from solid matrices such as soil and plant and animal tissues



**Figure 5** Manifold for dynamic MAE. 1, Solvent; 2, pump; 3, microwave oven; 4, extraction chamber; 5, temperature set-point controller; 6, thermocouple; 7, fluorescence detector; 8, recording device; 9, restrictor; 10, extractor. (Reproduced with permission from Luque de Castro MD and Luque-García JL (2002) *Acceleration and Automation of Solid Sample Treatment*; © Elsevier.)



**Figure 4** Experimental setup used to integrate MAE with the subsequent steps of the analytical process. (1) Leaching step. CT, controller; MO, microwave oven; S, sample; R, refrigerant; WR, water reservoir; TCPP, two-channel piston pump; ER, extract reservoir; (2) Clean-up/preconcentration step. SV, selecting valve; M, methanol; A, air; B, buffer; PP1 and PP2, peristaltic pumps; F, filter; EL, elution loop; MC, minicolumn; R, retention direction; E, elution direction; IV1–IV3, injection valves; W, waste. (3) Individual separation–detection step. HPIV, high-pressure injection valve; AC, analytical column; DAD, diode array detector; SR, solvent reservoirs. (Reproduced with permission from Luque de Castro MD and Luque-García JL (2002) *Acceleration and Automation of Solid Sample Treatment*; © Elsevier.)



relatively easily using MAE. The most effective solution used for leaching PAHs is an acetone–hexane mixture. The use of water as an extractant has also been tested with a view to developing clean methods using no organic solvents. Because of the high hydrophobicity of PAHs, however, the recoveries were very low in spite of the strong interaction between water and microwaves. One way of using water effectively for MAE of nonpolar compounds (PAHs included) is through micelle formation.

Most studies about MAE of PAHs from solid samples have been conducted using closed-vessel systems and only a few with open-vessel focused microwave devices. Because open-vessel systems operate at atmospheric pressure, the extraction vessel has been used as a reactor in order to perform online purification pretreatments of the total extracts (reagents can be added readily to the medium) or introduce the extracts directly into the determination instrument, as in the focused microwave-assisted extractor with online fluorescent monitoring, which provides a matrix-independent approach to the extraction of PAHs.

MAE has also become a frequent choice for the extraction of PCBs from solid matrices. In fact, this technique has been used to extract PCBs from a wide range of samples including soils, sediments, and animal tissues. Normally, the extractant used is the same as that employed with PAHs; micelle formation have also provided results similar to those obtained using conventional methodologies.

A wide variety of pesticides including organochlorine and organophosphorus compounds, triazines, herbicides, and imidazolinones have been extracted using microwaves, with recoveries ~100% in most cases.

Several MAE methods have also been developed for extraction of organometal compounds such as 3-nitro-4-hydroxyphenylarsonic acid and methylmercury from sediments and animal tissues. The MAE recoveries were consistently higher than those achieved with sonication or Soxhlet extraction. All species recoveries were matrix-dependent, as previously found with other pollutants, but using methanol acidified with acetic acid under optimal MAE conditions afforded rapid, efficient extraction of all analytes.

MAE has also been used as a solid sample treatment prior to speciation analysis, leaving the organometal compound moiety intact. This is a prerequisite for a successful extraction procedure to be applied before speciation analysis can be met by careful optimization of the conditions of the microwave attack. Open-vessel treatment is preferred to the pressurized bomb systems commonly used in analysis of total metals because it offers milder reaction conditions (the increase in temperature is governed to a great

extent by the boiling point of the solvent) and easier control of process variables.

Other organic compounds such as phenols, hydrocarbons, polymer additives, and natural compounds have also been extracted from a wide range of samples with good results.

One of the most interesting fields of application of MAE in food analysis is the extraction of lipids. This step, traditionally performed with conventional Soxhlet extraction, has been performed with the focused microwave-assisted Soxhlet extractor prototype of **Figure 2B**. Extraction of oil from olives, sunflower seeds, and soyabeans; extraction of the lipid fraction of dairy products (milk and cheese); and extraction of fatty acids from precooked and sausage foods have significant advantages over conventional methods, including dramatically reduced extraction times, lower degradation of thermolabile analytes, and acceleration of other analytical steps such as hydrolysis in milk samples, in addition to completeness of analyte extraction, which is not always achieved with conventional methods.

The extraction of metals and metalloids has also been assisted by microwaves, most often for speciation analysis. Metals such as Fe, Zn, Cu, As, Se, Hg, Pb, Co, Cr, Ni, and Cd have been extracted using acid solutions (usually  $\text{HNO}_3$ ) in a focused open-vessel system for periods ranging from 3 min for soil samples to 50 min for coal (where the metals are much more strongly retained). MAE has been used both as a pretreatment for chemical speciation (mainly to discriminate As and Se species) and as a means of accelerating fractionation of metals in sewage sludge for operation and/or functional speciation. The former type of application has usually been developed using open-vessel systems coupled to various types of equipment for individual separation and/or determination (e.g., hydride generation–atomic fluorescence detection, electrochemical detection, and liquid chromatography–hydride generation–inductively coupled plasma–mass spectrometry) following extraction.

**See also:** **Countercurrent Chromatography:** Solvent Extraction with a Helical Column. **Extraction:** Solvent Extraction Principles; Solvent Extraction: Multistage Countercurrent Distribution; Pressurized Fluid Extraction; Supercritical Fluid Extraction.

## Further Reading

García-Ayuso LE, Sánchez M, Fernández de Alba A, and Luque de Castro MD (1998) Focused microwave assisted Soxhlet: An advantageous tool for sample extraction. *Analytical Chemistry* 70: 2426–2431.

Luque-García JL and Luque de Castro MD (2003) Where is microwave-based analytical equipment for solid sample pretreatment going? *Trends in Analytical Chemistry* 22: 90–98.

Kingston HM and Haswell SJ (1997) *Microwave-Enhanced Chemistry*. Washington DC: ACS Professional Reference Books.

Kingston HM and Jassie LB (1998) *Introduction to Microwave Sample Preparation*. Washington DC: ACS Professional Reference Books.

Paré JRJ and Bélanger JMR (1994) Microwave-assisted process (MAP): A new tool for the analytical laboratory. *Trends in Analytical Chemistry* 13: 176–184.

Szpunar J, Schmitt VO, Donard OFX, and Lobinski R (1996) Low-power focused microwave technology as a new tool for rapid preparation of solid samples for speciation analysis. *Trends in Analytical Chemistry* 15: 181–187.

## Pressurized Fluid Extraction

**J L Luque-García**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Over the past few years, a technique based on the use of solvents at a high pressure and high temperature without reaching the critical point has gradually emerged as an efficient means for increasing automatability, shortening process times, and reducing the amount of solvent required to digest or leach solid samples. Such features as the low consumption of organic solvents and the expeditiousness of determinations make this technique especially suitable for environmental analysis, the area where it has expanded to the greatest extent.

The principal aspects of the technique, in its two modes (static and dynamic), the devices typically employed by each, and their amenability to coupling with subsequent operations of the analytical process, which is one of the most interesting aspects of this technique, are discussed in this article. The main variables affecting the extraction process as well as the applications are also commented on.

### Operational Pressurized Fluid Extraction Modes

Depending on the way the sample and extractant are brought into contact, pressurized fluid extraction (PFE) can be implemented in three different operational modes, static, dynamic, and static–dynamic.

In the static extraction mode, the sample is soaked in the extractant and the liquid flow is halted over a preset interval, after which it is propelled to the collection vessel, usually by a nitrogen stream. This extraction mode allows penetration of the extractant, and it is especially useful when the analytes cannot be removed expeditiously from the matrix.

In the dynamic extraction mode, the extractant is pumped through the sample at a preset flow rate. This mode allows the analyte to be exposed continuously to the pure (clean) solvent, thus favoring displacement of the analyte's partitioning equilibrium to the extractant.

Which extraction mode is better remains a controversial issue. The static mode provides longer contact between the sample and solvent, swells the matrix, and facilitates penetration of the extractant in its interstices. However, it is not suitable when the partitioning equilibrium is displaced slightly to the solubilization of the analytes in the extractant. The dynamic mode has the advantage that it puts the sample into contact with fresh extractant continuously but requires a larger amount of extractant and causes dilution of the extracts, which usually calls for a preconcentration step prior to determination of the analytes. The static–dynamic mode, which consists of a combination of the two previous modes, can be used to enjoy the advantages of both modes while circumventing their shortcomings.

### Pressurized Fluid Extractors

PFE has been carried out using both commercial and laboratory-built extractors, depending mainly on the operational mode used. While static PFE is usually carried out using commercial equipment (e.g., Dionex accelerated solvent extraction (ASE) 200 and ASE 300 or, more recently, the one commercialized by Applied Separation), the dynamic mode has been implemented only in laboratory-built extractors as nowadays there is no commercially available equipment. **Figure 1A** and **1B** shows schematic diagrams of a static and dynamic pressurized fluid (PF) extractor, respectively. The former (**Figure 1A**) consists of the following parts: (1) a reservoir for storing the liquid extractant; (2) a high-pressure pump for propelling the extractant through the system; (3) an extraction chamber, consisting of a



Luque-García JL and Luque de Castro MD (2003) Where is microwave-based analytical equipment for solid sample pretreatment going? *Trends in Analytical Chemistry* 22: 90–98.

Kingston HM and Haswell SJ (1997) *Microwave-Enhanced Chemistry*. Washington DC: ACS Professional Reference Books.

Kingston HM and Jassie LB (1998) *Introduction to Microwave Sample Preparation*. Washington DC: ACS Professional Reference Books.

Paré JRJ and Bélanger JMR (1994) Microwave-assisted process (MAP): A new tool for the analytical laboratory. *Trends in Analytical Chemistry* 13: 176–184.

Szpunar J, Schmitt VO, Donard OFX, and Lobinski R (1996) Low-power focused microwave technology as a new tool for rapid preparation of solid samples for speciation analysis. *Trends in Analytical Chemistry* 15: 181–187.

## Pressurized Fluid Extraction

**J L Luque-García**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Over the past few years, a technique based on the use of solvents at a high pressure and high temperature without reaching the critical point has gradually emerged as an efficient means for increasing automatability, shortening process times, and reducing the amount of solvent required to digest or leach solid samples. Such features as the low consumption of organic solvents and the expeditiousness of determinations make this technique especially suitable for environmental analysis, the area where it has expanded to the greatest extent.

The principal aspects of the technique, in its two modes (static and dynamic), the devices typically employed by each, and their amenability to coupling with subsequent operations of the analytical process, which is one of the most interesting aspects of this technique, are discussed in this article. The main variables affecting the extraction process as well as the applications are also commented on.

### Operational Pressurized Fluid Extraction Modes

Depending on the way the sample and extractant are brought into contact, pressurized fluid extraction (PFE) can be implemented in three different operational modes, static, dynamic, and static–dynamic.

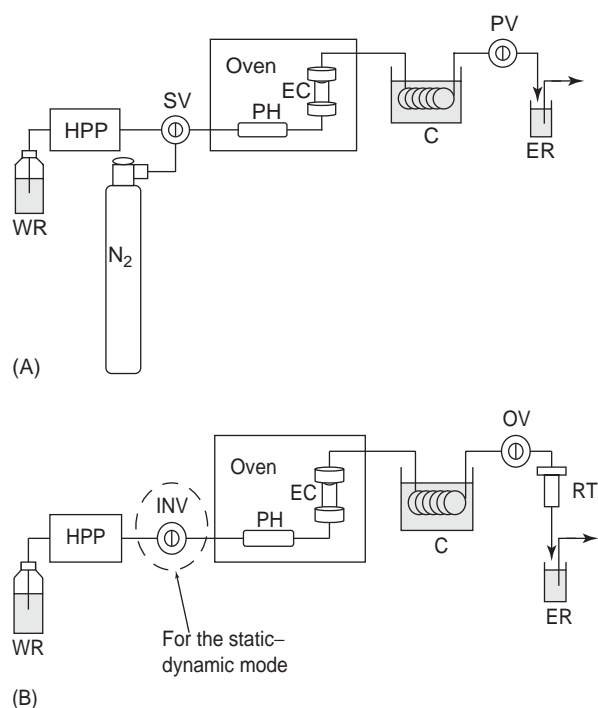
In the static extraction mode, the sample is soaked in the extractant and the liquid flow is halted over a preset interval, after which it is propelled to the collection vessel, usually by a nitrogen stream. This extraction mode allows penetration of the extractant, and it is especially useful when the analytes cannot be removed expeditiously from the matrix.

In the dynamic extraction mode, the extractant is pumped through the sample at a preset flow rate. This mode allows the analyte to be exposed continuously to the pure (clean) solvent, thus favoring displacement of the analyte's partitioning equilibrium to the extractant.

Which extraction mode is better remains a controversial issue. The static mode provides longer contact between the sample and solvent, swells the matrix, and facilitates penetration of the extractant in its interstices. However, it is not suitable when the partitioning equilibrium is displaced slightly to the solubilization of the analytes in the extractant. The dynamic mode has the advantage that it puts the sample into contact with fresh extractant continuously but requires a larger amount of extractant and causes dilution of the extracts, which usually calls for a preconcentration step prior to determination of the analytes. The static–dynamic mode, which consists of a combination of the two previous modes, can be used to enjoy the advantages of both modes while circumventing their shortcomings.

### Pressurized Fluid Extractors

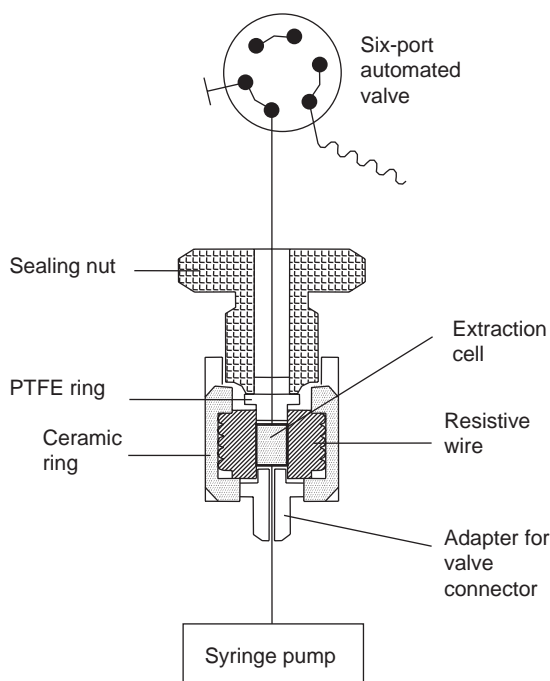
PFE has been carried out using both commercial and laboratory-built extractors, depending mainly on the operational mode used. While static PFE is usually carried out using commercial equipment (e.g., Dionex accelerated solvent extraction (ASE) 200 and ASE 300 or, more recently, the one commercialized by Applied Separation), the dynamic mode has been implemented only in laboratory-built extractors as nowadays there is no commercially available equipment. **Figure 1A** and **1B** shows schematic diagrams of a static and dynamic pressurized fluid (PF) extractor, respectively. The former (**Figure 1A**) consists of the following parts: (1) a reservoir for storing the liquid extractant; (2) a high-pressure pump for propelling the extractant through the system; (3) an extraction chamber, consisting of a



**Figure 1** General setup used to carry out PFE. (A) Static mode and (B) dynamic and static-dynamic modes. WR, water reservoir; HPP, high pressure pump; SV, selection valve; PH, pre-heater; EC, extraction chamber; C, cooler; PV, pressure valve; ER, extract reservoir; INV, inlet valve; RT, restrictor.

stainless steel cylinder closed with screws at either end, which permits circulation of the leaching fluid through the sample; (4) an electrically heated oven for placing the extraction chamber; (5) an on/off pressure valve for producing and keeping an appropriate pressure in the system; (6) a cylinder of gas (usually  $N_2$ ) for purging the system after extraction; (7) a valve located between the high-pressure pump and the oven, which allows flushing the system with the gas; and (8) a collection vessel (usually a vial).

The dynamic pressurized liquid extractor is similar to the static one but with the following modifications, as shown in **Figure 1B**: (1) an outlet valve located outside the oven used to produce the over-pressure required at the beginning of the extraction; (2) a cooler system used to cool the fluid from the oven temperature to  $\sim 25^\circ\text{C}$  (it is usually present in the laboratory-built static PF extractors as well); (3) a restrictor coupled to the outlet of the cooler (substituting the pressure valve) with the aim of maintaining the pressure constant in the system during dynamic extractions; (4) the selecting valve located after the high-pressure pump and the  $N_2$  stream are not present in dynamic extractors as no purging of the system is required in this extraction mode.



**Figure 2** Miniaturized device for dynamic-PFE of solid and semisolid samples. (Reproduced with permission of Elsevier.)

The static-dynamic mode is used with the assembly described for the dynamic mode but including an inlet valve (see INV in **Figure 1B**) for developing the static extraction step prior to the dynamic one. Recently, a laboratory-built miniaturized device for dynamic PFE has been developed (**Figure 2**). The most salient element of this device is a heatable stainless steel holder that serves as the extraction chamber, which is placed between a syringe pump (used to deliver the extractant) and an automated six-port valve for direct control of the pressure inside it.

## Variables Influencing PFE

### Temperature

Temperature is the most important variable affecting the efficiency of mass transfer in PFE as the use of high temperatures increases the extracting power of solvents. A high temperature has a favorable effect, decreasing the viscosity of liquid solvents, thus facilitating penetration of the extractant in the matrix particles. Temperature changes also affect the surface equilibrium. In fact, they alter strong analyte-matrix interactions caused by van der Waals forces, hydrogen bonding, and dipole attractions of the analyte molecules and active sites of the matrix and can lead to disruption of the surface equilibrium in response to positive changes in temperature. Thermal energy can overcome cohesive (analyte-analyte) and

adhesive (analyte–matrix) interactions by decreasing the activation energy required for desorption of the analytes from a solid particle.

Notwithstanding the foregoing, using a high temperature does not always result in increased extraction efficiency. Adverse effects are especially outstanding when thermolabile species are presented.

### Pressure

The pressure is usually a minor variable for the resulting efficiency in PFE, provided the level used is high enough to maintain the solvent in the liquid state. In some cases, however, pressure can be a key to ensuring complete analyte removal. The use of high pressures facilitates extraction from samples where the analytes have been trapped in matrix pores. The pressure increment forces the solvent into areas of the matrices that would not normally be contacted by them under atmospheric conditions.

### Matrix Composition

The composition of the sample (physical–chemical properties, type of compounds it contains, and its particle size) can be an important aspect related to the extraction efficiency. Between these factors the particle size is the most influential. Its influence depends on the particular variable that governs the extraction efficiency. Obviously, if the PFE rate is determined by diffusion of the analytes, it can be increased greatly by decreasing the particle size. However, the risk of losing volatile and reactive analytes during grinding is higher with samples consisting of particles less than 2 mm in size. These two opposing effects must thus be taken into account in reducing the particle size of samples to be extracted.

### Sample Size

Sample sizes for PFE usually range from 0.5 to 10 g. Obviously, the amount of sample used must be large enough to ensure homogeneity and obtain adequate sensitivity for trace analyses. A small sample size (<5 g) is preferred so as to avoid compaction of the sample in the extraction cell. Moreover, the smaller the amount of sample used is, the greater is the cell volume that can be occupied by the extractant and hence the higher is the extractant/analyte ratio and the more markedly shifted is the partitioning equilibrium toward solubilization of the analytes, which facilitates their extraction.

### Solvent Type

Optimal development of an extraction process entails an appropriate choice of solvent. The PFE technique is compatible with a wide range of solvents

other than those with an auto-ignition temperature of 40–200°C (e.g., carbon disulfide, diethyl ether, 1,4-dioxane) or a strong polar character. As a rule, strong bases and acids should also be avoided as solvents because they are corrosive, and their corrosion power increases with increasing pressure and temperature. The PFE technique provides an opportunity for the use of a variety of solvents or solvent mixtures, even those that are not effective in conventional methods. This is done by adjusting the temperature and pressure during the process, which generally increases the solubilization power of the solvents.

The static mode uses both organic solvents such as toluene, methanol, and acetone and solvent mixtures (usually in a 1:1 ratio) including dichloromethane–acetone, acetone–hexane, heptane–acetone, acetone–isohexane, and methanol–water. Water has rarely been used as the extractant in the static mode; however, it is the extractant most frequently used in the dynamic mode, both alone and with modifiers such as organic solvents, surfactants, and derivatizing reagents.

### Solvent Volume

The amount of solvent required depends strongly on the extraction mode used. In the static mode, the sample is extracted with a minimum volume of solvent (usually <15 ml), with no outflow. When the solvent volume used in the static mode does not ensure quantitative extraction of the target analytes, several extraction cycles or the dynamic mode must be used. In the dynamic mode, the extractant flows continuously through the extraction cell, and so the volume of solvent that comes into contact with the sample is a direct function of both the flow rate of the circulating extractant and the extraction time. Obviously, the dynamic mode uses larger volumes of solvent than the static mode, and so it is less well suited for trace analysis, although there are various ways of minimizing this dilution effect, as shown below.

### Solvent Flow Rate

This is a characteristic variable of the dynamic extraction mode. The effect of the flow rate on the efficiency and expeditiousness of PFE can be used to determine whether the extraction is limited by analyte solubility in the extractant, diffusion in the particles, and/or transfer from the particle surface to the extractant. If the extraction efficiencies do not change when the flow rate is increased, then the extraction is limited neither by the solubility nor by the equilibrium of mass transfer between the matrix and extractant, and so the rate-determining step of the

process is diffusion inside the solid particles. In such a case, the extraction rate must be increased by raising the extraction temperature. If the limiting factor is analyte solubility, then doubling the flow rate doubles the amount of analyte that is extracted during the same time. On the other hand, if the analyte undergoes different readsorption/desorption steps during elution from the extraction chamber, doubling the flow rate also doubles the extraction rate.

### **Extraction Time**

Extraction times in PFE are very short relative to those required by conventional solid-liquid extraction techniques such as Soxhlet extraction and depend on which particular process determines the extraction rate. The extraction times in PFE usually range between 5 and 30 min, but they can be as long as 90 min in highly unfavorable cases such as the extraction of selenium, mercury, and arsenic from coal.

### **Coupling PFE with Other Steps of the Analytical Process**

One of the most interesting aspects of this technique (apart from the good results it provides) is the possibility of coupling to other steps of the analytical process, thus enabling partial or total automation of the analytical process and solving problems such as dilution of the extracts when the dynamic extraction mode is used, the lack of selectivity under the extraction conditions, or the presence of solid suspended particles in the extracts.

#### **Coupling Static PFE Mode**

Static pressurized liquid extraction, as implemented in commercial equipment, is basically discrete in nature, and so it is rarely coupled to other operations of the analytical process. In fact, only in a few reported applications is the static mode coupled online to other operations such as chromatographic separation, preconcentration, and detection. In any case, custom extractors are used as the compact design of the commercial models precludes their adaptation.

Static PFE has been coupled to solid-phase extraction (SPE) with both SPE disks placed directly into the extraction chamber jointly with the sample and by using a minicolumn packed with the sorbent material and connected at the end of the extractor by means of a flow injection manifold. The minicolumn can be placed either in the transport tube or in the loop of an auxiliary injection valve of a flow-injection manifold, thus enabling elution in the direction opposite to retention. Analytes can be eluted manually, by disconnecting the column from the extractor, or

automatically (either by passing a volume of organic solvent with the aid of a pump and collecting the eluate, which is made to volume in a vial, or by using a second injection valve to control the eluted volume, thus ensuring reproducibility). The latter mode also has the possibility of direct transfer of the eluate to a liquid or gas chromatograph also in an automatic manner. In both modes, the solid phase must be allowed to dry before the analytes are eluted; to this end, an air or nitrogen stream is passed after the extract has been passed through the column.

A laboratory-made static PF extractor has also been coupled to a high-performance liquid chromatograph directly with the aim of passing the extract through the analytical column in an automatic manner after static extraction. This coupling has been performed by just connecting the injection valve of the liquid chromatograph to the outlet tube of the extractor, thus allowing injection of a fraction of the extract.

The coupling of a static PF extractor to a fluorimeter has been performed primarily to develop an independent matrix extraction method allowing the extraction kinetics to be monitored. This approach has been followed with a view to determining the number of static extraction cycles required for quantitative extraction of the analytes, irrespective of the kind of matrix concerned and whether the species concentrations in it were known. The extract obtained after each cycle is driven to a graduated reservoir from which it is aspirated into a flow-injection manifold connected to the fluorimeter.

#### **Coupling Dynamic PFE**

The potential of the dynamic PFE technique for coupling to subsequent operations of the analytical process is only limited by the analyst's ingenuity and material resources. The systems described below can be combined by altering the sequence of steps and introducing appropriate modifications to develop fully automated systems for specific purposes.

Two different approaches to the liquid-liquid extraction of leached analytes have been reported. One, which is discrete in nature, uses a vial containing the immiscible organic solvent, through which the aqueous leached phase is percolated. The other operates in a continuous manner, by percolating the effluent from a restrictor through a flow cell located at the detector and filled with an organic phase that is lighter than water and allows partitioning of the extracted analytes into the organic solvent. In the latter case, by dissolving an appropriate complexing agent in the organic phase, it has been possible to monitor indirectly the extraction kinetics.

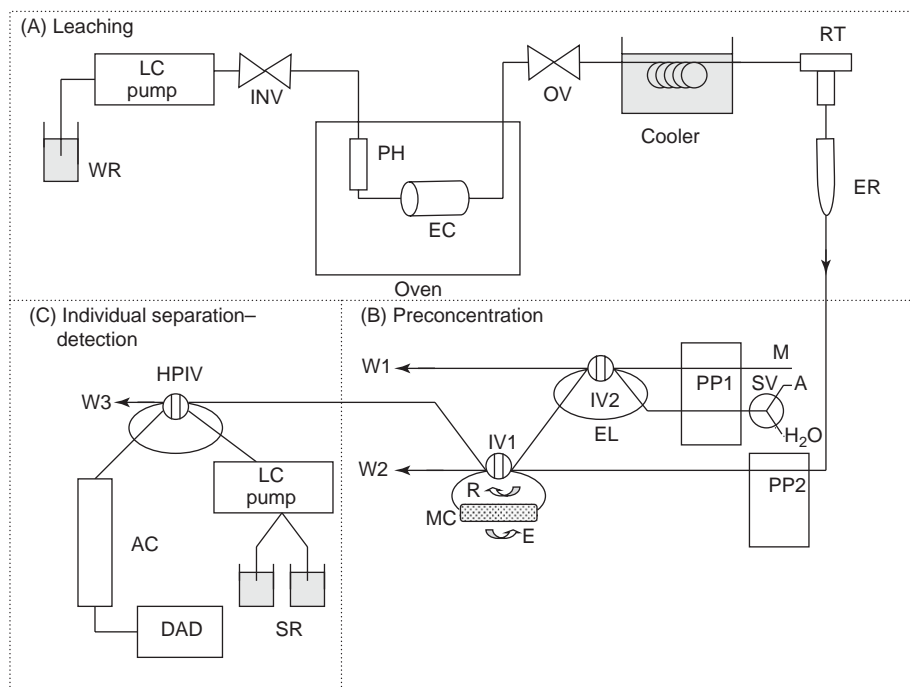
Dynamic PFE usually requires implementing a concentration step prior to the determinative step, and because the extracted analytes are dissolved in a liquid (usually aqueous) phase, SPE is a highly useful tool for avoiding the dilution effect. For this purpose, SPE cartridges and columns packed with appropriate sorbents and coupled online to the extractor outlet can be employed in the same way as commented on for static PFE. Miniaturized retention has also been developed by using solid-phase microextraction (SPME).

One very common, undesirable occurrence in dynamic PFE is the presence of solid particles in the liquid extract due to oversaturation during cooling. This shortcoming can be circumvented by fitting a filtration unit in line with the extractor. The unit can be placed either in the transport tube or in the loop of an auxiliary injection valve; the latter choice allows the filter to be cleaned between samples by passing a rinsing solution at a high flow rate in the direction opposite to filtration.

When the extracted analytes provide no response on passing through a flow cell located in a detector connected online to the continuous extractor via a dynamic manifold, the extractor outlet can also be connected to a manifold where the extract is merged with a stream of an appropriate reagent to derivatize the analytes as they are extracted, thereby enabling their subsequent online determination.

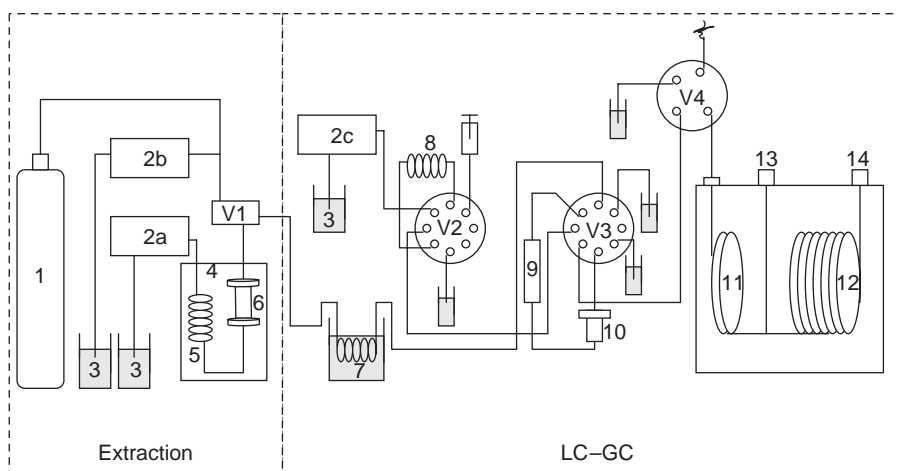
Coupling of dynamic PFE with supercritical water oxidation has been carried out in order to oxidize (destroy) organic compounds previously extracted using PFE such as polyaromatic hydrocarbons (PAHs), phenol, and benzaldehyde, among others, as a way of treating waste. Two ovens, one containing the extraction vessel and another containing the reaction tube, are connected by a tube that is insulated. Two high-pressure pumps are used, one to deliver the extractant (water) to the extraction vessel and another to deliver the oxidant (hydrogen peroxide) to the stream at a T confluence in the second oven.

The dynamic PFE–SPE–liquid chromatography (LC) coupling has been implemented with two different configurations. In one, the chromatographic mobile phase also acts as the eluent for the analytes retained on the SPE column; in the other, elution is done by using an injection valve that allows the eluent volume to be controlled in a reproducible manner. In the latter configuration (Figure 3), the eluate is swept to the high-pressure injection valve in the chromatograph using air as carrier to avoid dilution. In both assemblies, the SPE column is placed in the loop of a six-way valve to allow continuous passage of the mobile phase through the chromatographic column and detector in order to establish the baseline. Also, in the latter, elution is done in the direction opposite to



**Figure 3** Experimental setup used for the coupling dynamic PFE–SPE–LC. WR, water reservoir; INV, inlet valve; PH, preheater; EC, extraction chamber; OV, outlet valve; RT, restrictor; ER, extract reservoir; M, methanol; A, air; SV, selection valve; PP1 and PP2, peristaltic pumps; IV1 and IV2, injection valves; EL, elution loop; MC, minicolumn; R, retention; E, elution; W1–W3, wastes; SR, solvent reservoirs; HPIV, high-pressure injection valve; AC, analytical column; DAD, diode array detector.





**Figure 4** Scheme of a pressurized liquid extraction-LC-GC setup. 1, nitrogen; 2a-c, pump; 3, eluent; 4, oven; 5, preheating coil; 6, extraction vessel; 7, cooling coil; 8, eluent coil; 9, trapping and LC column; 10, restrictor; 11, precolumns; 12, analytical column; 13, solvent vapor exit; 14, flame ionization detector; V1, extraction valve; V2-V4, multiport valves. (Reproduced with permission from Hyötyläinen T, Andersson T, Hartonen K, Kuosmanen K, and Riekkola ML (2000) Pressurized hot water extraction coupled online with LC-GC: Determination of polyaromatic hydrocarbons in sediment. *Analytical Chemistry* 72: 3070-3076; © American Chemical Society.)

retention to avoid changes in compaction of the sorbent material.

PFE has been combined directly with a liquid chromatography-gas chromatography (LC-GC) system. **Figure 4** shows the experimental setup used. The analytical procedure developed with this system is divided into five steps: (1) PF extraction. The sample is extracted and the analytes are trapped in the Tenax column. After extraction, the water flow is stopped by closing valve V1 and pump 2a. (2) Drying. After extraction the column is dried with nitrogen to remove water. The nitrogen flows through valve V1. (3) LC clean-up. Valve V3 is switched, and pump 2c pushes the eluent (pentane) through the column. After this, the eluent is changed to pentane/ethyl acetate, which is delivered from the loop by switching valve 2. (4) Transfer to GC. When the fraction containing the analytes is eluted from the Tenax column, valve V4 is switched, and the fraction is transferred to GC. During the transfer, partially concurrent solvent evaporation is applied. (5) After transfer, the tubing attached to the extraction vessel and the trap column is flushed. After rinsing the lines and column, they are dried with a flow of nitrogen. The next extraction can then be started while the GC analysis of the previous extraction is still proceeding. The main benefit of this coupling is that the concentration step is highly efficient, so that the sensitivity is ~800 times better than that obtained with traditional methods.

Sorbent materials can also be used in the flow cell of a nondestructive detector for continuous monitoring of the retention step, thus providing an indirect way of controlling the extraction kinetics. The

kinetics can be monitored in a direct manner if the analytes possess appropriate intrinsic properties or follow online derivatization. In both cases, completion of extraction is shown by the obtainment of a plateau (a constant signal) as a result of the constancy in the analyte content on the solid support. Once the signal levels off, the analytes can be eluted to waste – thus making the manifold ready for processing a new extract – or a portion of the extract can be inserted into a gas or liquid chromatograph for its individual determination as the sorbent is regenerated for a new extraction.

## Applications of PFE

PFE has so far been applied exclusively to solid samples, especially in environmental analysis, which accounts for over 80% of its reported uses. In recent years, however, PFE has been used increasingly in the pharmaceutical and industrial fields as well (especially for extraction of additives from polymers).

Soil and sediments constitute the principal environmental target samples, and organic pollutants, the main target analytes. Static and dynamic PFE modes have been used for the extraction of PAHs, polychlorinated biphenyls, phenols, chlorophenols, hydrocarbons such as diesel range organics, waste oil organics, surfactants, etc., as well as inorganic compounds such as those of phosphorus, heavy metals, and organometals. As commented before, organic solvents have been used mainly in the static mode; meanwhile, water has been used in the dynamic one both alone and with some modifiers



such as organic solvents, derivatization agents, and surfactants. In most of the cases, the results obtained have been comparable with those obtained using other techniques such as Soxhlet extraction, supercritical fluid extraction, and microwave-assisted extraction.

PFE has been used in the analysis of biological samples for two main purposes, namely to determine contaminants in food and animal tissues and to extract target species from animal and plant tissues. In the last case, dynamic-PFE has been used widely for isolation of edible essential oils from plants such as fennel, marjoram, laurel, rosemary, oregano, and eucalyptus, using water as the extractant.

In the pharmaceutical field, PFE has also been used for two primary purposes, namely to extract pharmacologically active substances from plants and, especially, for quality control of tablets and medical foods. In this field, PFE clearly surpasses classical extraction methods such as those endorsed by pharmacopeias (which use as official standards techniques such as Soxhlet extraction, percolation, maceration, digestion, extraction under reflux, and steam distillation) and others based on ultrasonication or turboextraction. PFE is a firm candidate for use in high-throughput screening programs for natural product discovery, where large numbers of small-scale extractions have to be performed in an efficient, reproducible manner.

PFE has been used to extract various additives from different types of polymers, all with good results and the intrinsic advantages of this technique, which include a high throughput, solvent economy, and the production of extracts that require no

filtration. Thus, PFE has enabled the efficient extraction of flame-retarded thermoplastic additives, plasticizers such as dioctyl adipate, trioctyl phosphate, dioctyl phthalate, and trioctyl trimellitate, and antioxidants such as Irganix and Irgafos from various polymers including poly(vinyl chloride), polyethylene, and polypropylene.

**See also:** **Countercurrent Chromatography:** Solvent Extraction with a Helical Column. **Extraction:** Solvent Extraction Principles; Solvent Extraction: Multistage Countercurrent Distribution; Microwave-Assisted Solvent Extraction; Supercritical Fluid Extraction.

## Further Reading

- Björklund E, Nilsson T, and Bowadt S (2000) Pressurized liquid extraction of persistent organic pollutants in environmental analysis. *Trends in Analytical Chemistry* 19: 434–445.
- Hawthorne SB, Yang Y, and Miller DJ (1994) Extraction of organic pollutants from environmental solids with sub- and supercritical water. *Analytical Chemistry* 66: 2912–2920.
- Luque de Castro MD and Luque-García JL (2002) *Acceleration and Automation of Solid Sample Treatment*. Amsterdam: Elsevier.
- Ramos L, Vreuls JJ, and Brinkman UATH (2000) Miniaturized pressurized liquid extraction of polycyclic aromatic hydrocarbons from soil and sediment with subsequent large-volume injection-gas chromatography. *Journal of Chromatography* 891: 275–286.
- Richter BE, Jones BA, Ezzell JL, and Porter NL (1996) Accelerated solvent extraction: A technique for sample preparation. *Analytical Chemistry* 68: 1033–1039.

## Supercritical Fluid Extraction

**E Björklund and C Sparr-Eskilsson**, Lund University, Lund, Sweden

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by A A Clifford, pp. 4724–4725, © 1995, Elsevier Ltd.

## Introduction

Supercritical fluid extraction (SFE) is today an important technique for the extraction of a large number of differing analytes from various matrices such as pollutants from soils, fats from food stuffs, active substances from plants, and additives from polymers. The obtained extracts are normally analyzed on a chromatographic system even though other

techniques are applied (e.g., gravimetry and spectroscopy). Compared to many methods based on conventional liquid extraction, SFE offers speed and cleaner extracts, decreasing the need for external clean-up steps. This cuts the total costs of analysis due to reduced labor-intensiveness. There is also an environmental advantage, as harmless extraction fluids are normally used. This is important as regulations are getting stricter on the use of certain chlorinated organic solvents, with increasing costs for solvent disposal. The reason for the decreased interest in SFE, starting at the end of the last decade, was mainly caused by an initial over enthusiasm in SFE, when first introduced in the mid-1980s. In some early attempts, SFE was also applied on analytes and matrices for which it was not well suited.

such as organic solvents, derivatization agents, and surfactants. In most of the cases, the results obtained have been comparable with those obtained using other techniques such as Soxhlet extraction, supercritical fluid extraction, and microwave-assisted extraction.

PFE has been used in the analysis of biological samples for two main purposes, namely to determine contaminants in food and animal tissues and to extract target species from animal and plant tissues. In the last case, dynamic-PFE has been used widely for isolation of edible essential oils from plants such as fennel, marjoram, laurel, rosemary, oregano, and eucalyptus, using water as the extractant.

In the pharmaceutical field, PFE has also been used for two primary purposes, namely to extract pharmacologically active substances from plants and, especially, for quality control of tablets and medical foods. In this field, PFE clearly surpasses classical extraction methods such as those endorsed by pharmacopeias (which use as official standards techniques such as Soxhlet extraction, percolation, maceration, digestion, extraction under reflux, and steam distillation) and others based on ultrasonication or turboextraction. PFE is a firm candidate for use in high-throughput screening programs for natural product discovery, where large numbers of small-scale extractions have to be performed in an efficient, reproducible manner.

PFE has been used to extract various additives from different types of polymers, all with good results and the intrinsic advantages of this technique, which include a high throughput, solvent economy, and the production of extracts that require no

filtration. Thus, PFE has enabled the efficient extraction of flame-retarded thermoplastic additives, plasticizers such as dioctyl adipate, trioctyl phosphate, dioctyl phthalate, and trioctyl trimellitate, and antioxidants such as Irganix and Irgafos from various polymers including poly(vinyl chloride), polyethylene, and polypropylene.

**See also:** **Countercurrent Chromatography:** Solvent Extraction with a Helical Column. **Extraction:** Solvent Extraction Principles; Solvent Extraction: Multistage Countercurrent Distribution; Microwave-Assisted Solvent Extraction; Supercritical Fluid Extraction.

## Further Reading

- Björklund E, Nilsson T, and Bowadt S (2000) Pressurized liquid extraction of persistent organic pollutants in environmental analysis. *Trends in Analytical Chemistry* 19: 434–445.
- Hawthorne SB, Yang Y, and Miller DJ (1994) Extraction of organic pollutants from environmental solids with sub- and supercritical water. *Analytical Chemistry* 66: 2912–2920.
- Luque de Castro MD and Luque-García JL (2002) *Acceleration and Automation of Solid Sample Treatment*. Amsterdam: Elsevier.
- Ramos L, Vreuls JJ, and Brinkman UATH (2000) Miniaturized pressurized liquid extraction of polycyclic aromatic hydrocarbons from soil and sediment with subsequent large-volume injection-gas chromatography. *Journal of Chromatography* 891: 275–286.
- Richter BE, Jones BA, Ezzell JL, and Porter NL (1996) Accelerated solvent extraction: A technique for sample preparation. *Analytical Chemistry* 68: 1033–1039.

## Supercritical Fluid Extraction

**E Björklund and C Sparr-Eskilsson**, Lund University, Lund, Sweden

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by A A Clifford, pp. 4724–4725, © 1995, Elsevier Ltd.

## Introduction

Supercritical fluid extraction (SFE) is today an important technique for the extraction of a large number of differing analytes from various matrices such as pollutants from soils, fats from food stuffs, active substances from plants, and additives from polymers. The obtained extracts are normally analyzed on a chromatographic system even though other

techniques are applied (e.g., gravimetry and spectroscopy). Compared to many methods based on conventional liquid extraction, SFE offers speed and cleaner extracts, decreasing the need for external clean-up steps. This cuts the total costs of analysis due to reduced labor-intensiveness. There is also an environmental advantage, as harmless extraction fluids are normally used. This is important as regulations are getting stricter on the use of certain chlorinated organic solvents, with increasing costs for solvent disposal. The reason for the decreased interest in SFE, starting at the end of the last decade, was mainly caused by an initial over enthusiasm in SFE, when first introduced in the mid-1980s. In some early attempts, SFE was also applied on analytes and matrices for which it was not well suited.

Additionally, unsatisfactory extraction efficiencies were often blamed on poor extraction conditions, while the true cause was inadequate collection. Consequently many laboratories regarded SFE as too complex and abandoned the technique despite its excellent features. However, at present there are many well-conducted scientific applications and comprehensive text books that aids in the development of potent and competitive methods for the replacement of many conventional liquid-based methods such as Soxhlet. The basic concepts of SFE and important factors governing the extraction process will be outlined. Several applications suitable for SFE will also be presented.

## **Alternative Extraction Techniques**

SFE is today not the only hyphenated extraction technique available. Both microwave-assisted solvent extraction (MAE) and pressurized fluid extraction (PFE) have evolved in parallel with SFE during the last decade. A thorough comparison of the different techniques can be found in the analytical literature, and will assist in choosing an appropriate technique for a specific application. These newer techniques offer both speed and simplicity, while some of the main benefits of SFE are still inherent to the usage of supercritical fluids as MAE and PFE utilize liquid solvents. More recently, subcritical water extraction has also been introduced as an environmental friendly alternative but the reader is referred to the analytical literature on this research topic.

## **Coupling of SFE to Other Techniques**

There are a number of online methods coupling SFE to other analytical techniques such as gas chromatography (GC), capillary or packed column supercritical fluid chromatography (SFC), and liquid chromatography (LC). The online techniques are rather sensitive because the entire portion of extracted components is transferred to the column. A disadvantage with online approaches is that coextracted fat, water, and sometimes modifier may interfere in the final analysis. Additionally overloading of the analytical column must be avoided. Therefore, only small samples are usually extracted with online methods in SFE. A comprehensive discussion of online methods can be found in Further Reading.

## **Supercritical Fluids**

A supercritical fluid refers to a substance above its critical temperature ( $T_c$ ) and pressure ( $p_c$ ), where

only fluids exist, and the distinction between gases and liquids disappear. As the pressure is increased the fluid changes from being gas-like to become liquid-like in its properties. Compared to liquids, diffusion is faster and viscosity is lower, which results in a more rapid matrix penetration and analyte transport out of the sample. Surface tension is also low resulting in a good matrix wetting. The density of a supercritical fluid is close to a liquid and can be adjusted by simply changing the pressure. This compressibility of supercritical fluids allows the density, and thereby, the solvent strength at a given temperature, to be easily adjusted, and admits some selectivity to the extraction. Solvent strength is often expressed (qualitatively) using the Hildebrand solubility parameter ( $\delta$ ) and allows for some comparison between supercritical fluids and liquid solvents. However,  $\delta$  is not always sufficient to describe a supercritical fluid, since other physicochemical properties of the molecule might affect the extraction efficiency, such as permanent dipole moments. It is also important to point out a major misconception that density alone governs analyte solubility. Temperature also plays an important role and often increases analyte solubility for certain compounds (see the section 'Solubility, temperature, and pressure'). In fact the simplest way to describe extraction conditions is to give information on pressure and temperature, as these define the density of the supercritical fluid. Additionally, most instruments require that pressure and temperature are set prior to starting the extraction.

Even though the extraction efficiency for some analytes can be increased by change of extraction fluid, carbon dioxide is by far the most common compound used (98% of all applications). It has low critical parameters, it is nonexplosive, nontoxic, and environmentally benign. Alternatives have been proposed such as alkanes and freons but they have never been widely accepted due to health and safety risks for the former and ozone depletion by the latter. One of the few competitors to carbon dioxide is nitrous oxide, however, it might cause explosion in contact with high amounts of organic material. Supercritical carbon dioxide has a polarity similar to that of n-hexane, and consequently for the extraction of more polar analytes an organic modifier such as methanol or acetonitrile (1–5%) has to be added to increase the polarity (see the section 'Modifiers'). To maintain supercriticality for two-component fluids, somewhat different conditions have to be applied, but normally there is no problem at the conditions under which SFE is normally carried out.

## Instrumentation

### Basic SFE Setup

A simplified picture of an instrumental SFE setup is shown in Figure 1.

The SFE instrument consists of a solvent supply, a pump to maintain the pressure and the flow rate, an extraction cell mounted in a heated oven in order to maintain the temperature at the specified value, a restrictor connected to the outlet of the cell, and a collection device. There are different ways of collecting analytes as discussed in the section 'Collection Devices'. In a representative experiment, the cell has a volume of  $\sim 10$  ml and is loaded with 1–5 g of sample matrix. Carbon dioxide is pumped at a rate of  $0.5\text{--}4\text{ ml min}^{-1}$ , measured as liquid at the pump. The pressure is 150–450 bar and the temperature  $40\text{--}150^\circ\text{C}$ , with a dynamic extraction time of 10–60 min. The extracting fluid is transported through a restrictor. At the restrictor, the pressure of the extracting fluid drops as the extraction fluid expands, releasing the extracted compounds. The gaseous extraction fluid is vented and the analytes are collected in a small volume of solvent (5–20 ml) or onto a solid-phase trap (commonly octadecyl-silica), which is rinsed with solvent in a subsequent step. The extract is often ready for direct analysis. An internal standard is added to the vial to correct for differences in final extract volumes.

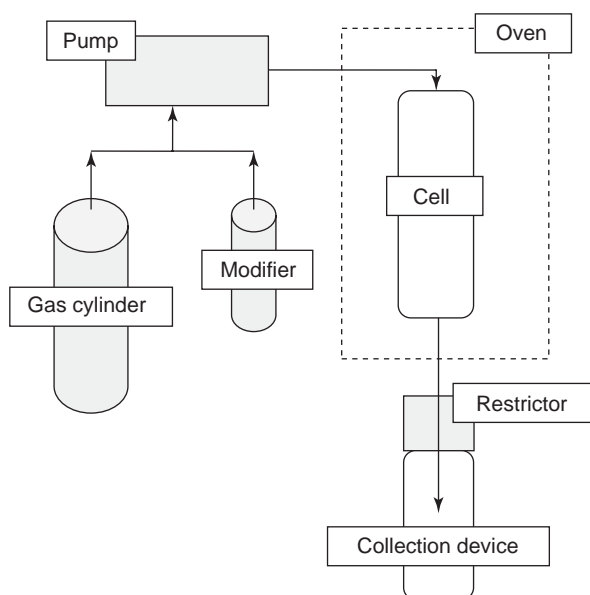
In certain applications, static extractions can be performed. In the static mode, the outlet valve is closed, the cell filled with fluid, and the inlet valve is closed. The fluid is then left in contact with the

matrix for 10–30 min. The valves are opened and the fluid in the cell is swept into the analyte trap by performing a dynamic extraction. Static extraction is less common than dynamic extraction, but has been considered preferable for some applications. When the extraction process is controlled by a slow removal of analytes out of the matrix into the extracting fluid (see the section 'Analyte/matrix considerations and sample types'), an initial static extraction step may be advantageous. A combination of such a static and dynamic step then usually leads to higher recoveries in a shorter time than using a single dynamic step.

### Equipment Items

**Pumps and fluid supply** For analytical SFE, two pumps types are common: syringe pumps and reciprocating pumps. Conventional LC pumps can also be used for SFE. Though syringe pumps are generally reliable with a limited need of maintenance, a disadvantage is the limited capacity and need of refilling regularly. A suitable SFE pump should measure both pressure and flow rate. Additionally, the maximum pressure (at least 400 bar) should be maintained at the maximum flow rate required (at least  $4\text{ ml min}^{-1}$ ). Moreover, the set pressure should be reached fast when the inlet valve to the extraction cell is opened. It is also important that the pump is sufficiently cooled. Cooling can be performed by water or by using an additional carbon dioxide cylinder for this purpose. The fluid substance is contained in a cylinder with a dip-tube and used so that liquid is supplied to the pump. Also available are gases, which have an overpressure of helium of  $\sim 100$  bar, which removes the necessity of cooling the pump. Unfortunately carbon dioxide supplied in this way contains  $\sim 3\%$  of helium, and studies have shown that these cylinders preferably should be avoided. They are also relatively expensive.

LC-grade solvents are normally used as modifiers. The modifier can be added directly into the extraction cell prior to extraction. In this case no additional pump is needed. A disadvantage is that the entire volume of modifier is extracted together with the analytes during the dynamic mode of the extraction. Using premixed cylinders is another option. A range of mixtures of carbon dioxide with various modifiers added (1–10%) are commercially available. There are severe drawbacks of premixed cylinders as there is less flexibility in the choice of the solvent used, and the composition of the liquid supplied from mixed cylinders changes over the lifetime of the gas mixture. The most convenient and most commonly preferred way is the use of an additional LC pump for



**Figure 1** Schematic apparatus for carrying out SFE.



modifier delivery. When using a gradient LC pump in automated SFE systems, several modifiers with different concentrations can easily be tested.

**Extraction cells** Different cell types are available, but the best choice is cells that can be sealed properly without any tools. In commercial systems hand-tightened extraction cells are standard. Cells are available in sizes from 1 to 100 ml. The material in the cells should be inert and should not absorb the extracting media. Stainless steel or crystalline polymers are common materials for this purpose. There have been discussions on the orientation of the extraction cells, but the most logical way is a vertical position with the fluid entering from the top. This means that there is no delay in emergence of the extracted analytes. Solid matrices should be uniformly packed into the cell, with no easy passages for the fluid. Care must be taken when assembling the cell to keep particles away from the surfaces of the connectors.

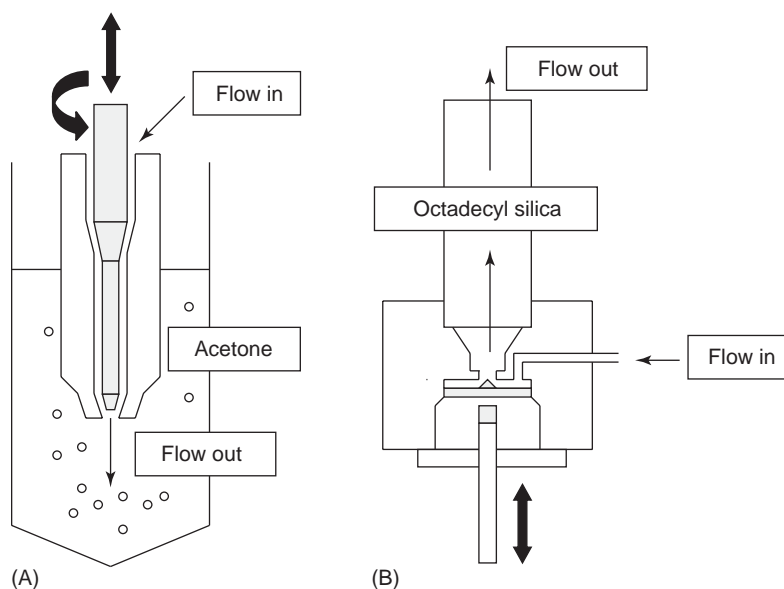
**Restrictors** Several different types of restrictors have been used in analytical SFE. The simplest method of providing pressure or flow control is to use a tubing of fused silica capillary as a linear restrictor. Tubing of 20–30  $\mu\text{m}$  internal diameter and lengths of 10–50 cm are usual. One major drawback with linear restrictors may be blockage caused by coextraction of matrix components, i.e., water, sulfur, lipids, waxes, and small particles. Water gives rise to plugging by freezing in the outlet caused by cooling due to the expansion of the fluid and the evaporation of the

trapping solvent. To avoid the problem with clogging, the restrictor can be heated and/or some kind of water-adsorbing material is mixed with or placed downstream the sample. Blockage may also be avoided by heating the solvent collection vial, taking care not to evaporate the trapping solvent, which can lead to loss of volatile analytes. The advantage of the fused silica restrictor is the low cost. However, care must be taken in order to obtain reproducible results when complex samples are extracted. Alternatively, linear restrictor can be made of stainless steel or inert polymeric materials.

In a needle valve restrictor, the flow rate is controlled by changing the valve position. This can be performed during the extraction to avoid blockage. The drawbacks with needle valves are virtually the same as for capillary restrictors. The preferred choice is electrically controlled restrictors, which maintains a back-pressure or flow rate set by the operator. These devices are not very often subject to blockage, which makes them highly reliable and often form part of commercial SFE instrument. Two types of restrictor designs are shown in **Figure 2**.

A coaxial heated restrictor consisting of a syringe with a movable piston position controls the flow, or alternatively a small metal disk is pushed toward an orifice in order to regulate the flow rate. The former is suitable for collection in organic solvents, while the latter has been utilized for solid-phase trapping.

**Collection devices** The importance of a proper collection of analytes in SFE has been emphasized by



**Figure 2** Two types of restrictors, combined with organic solvent collection or solid-phase trapping. (A) Solvent collection; (B) solid-phase trapping.

numerous investigations. Faulty collection rather than nonquantitative extraction could explain low extraction recoveries. Collection in a solvent is most commonly achieved by keeping the restrictor outlet immersed into a vessel containing a small volume of an organic solvent (Figure 2A). There are several parameters to consider, including solvent type, solvent volume, solvent temperature, restrictor flow rate, restrictor temperature, and pressurization of the collection vessel. Solvent collection is rather insensitive to analyte breakthrough losses when the sample contains large amounts of fat or water that can be coextracted, or when a modifier is used. However, volatile analytes are more difficult to trap in a solvent than on an adsorbing solid-phase trap.

In solid-phase trapping, the decompressed extraction fluid passes a device filled with a sorbent material (e.g., chromatographic material such as octadecylsilica, diol, silica, and Florisil<sup>®</sup>, Figure 2B). After extraction is completed, the analytes are eluted from the solid-phase trap with a suitable solvent. The choice of trapping material (single or a combination of adsorbents), trapping temperature, and the eluting solvent or solvent mixture is of great concern. The capacity of the trapping material may be a limiting factor, resulting in breakthrough losses of analytes. Even if the target analytes do not exceed the limit of the trap, losses may occur because of overloading of the trapping material with coextracted matrix components. However, one way of avoiding capacity problems is to perform a fractionated extraction/elution procedure, i.e., to rinse the trap at certain time intervals during the extraction. There is a major drawback when using solid-phase trapping together with water-containing samples or when high concentrations of modifier are added to the extraction fluid. During these conditions condensation of water or modifier in the trap may cause breakthrough losses of analytes. A common solution is to keep the trapping temperature above the boiling point of the modifier. Collection with the use of an adsorbent has the potential for selective trapping. Compared to solvent collection, the solid-phase trapping technique offers high trapping efficiency for substances with high vapor pressures since the trap temperature can easily be reduced to  $-30^{\circ}\text{C}$ . Collection in an empty vessel has also been employed in analytical-scale SFE, mostly for fat extractions, even though it may result in lower recoveries compared to collection in a vessel containing solvent.

### Commercial Equipment

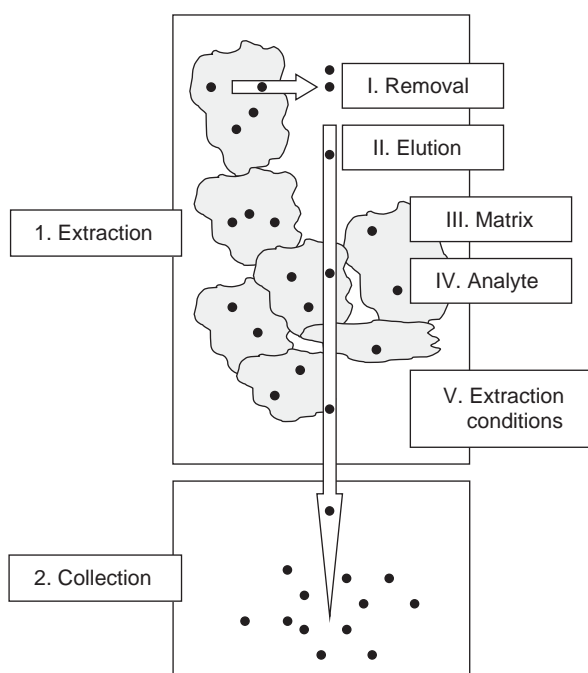
A number of companies produce SFE equipment offering convenience and easy control of extraction

conditions. Other features are simultaneous extraction of samples (2–4) or sequential sample extraction (8–24) as well as direct extraction of liquid samples. In the last few years, smaller lab/pilot-scale extraction units, specialized for certain applications have entered the market. These are designed for the extraction of oils and aroma compounds from natural products, active components from pharmaceuticals, polymer fractionation, and precision parts cleaning. Some of these are also utilizable in the analytical laboratory. Certain devices offer a phase monitor for visual observation of materials under conditions that may be controlled by the researcher. Thereby, the solubility parameters of compounds and mixtures in subcritical and supercritical fluids can be determined.

### Factors Affecting SFE

The basic requirement for SFE to be useful is that it should be quantitative. Consequently, reliable methods must be developed based on good understanding of the extraction process. In principle, an extraction method consists of an extraction step (1) and a collection step (2), both equally important (Figure 3).

The collection step is often overlooked but it is the first thing to be evaluated in a new method. Secondly, the extraction step must be considered, since several factors influence analyte recovery. Much is still unknown about the precise mechanisms involved in the



**Figure 3** Major factors affecting the overall SFE process. See text for detailed discussion of the various factors.



extraction process (which applies to all extraction methods) but in principle the analyte must be removed from the matrix (I) followed by elution of the analyte from the extraction cell (II), where both matrix and analyte characteristics (III, IV) will influence the extraction process (Figure 3). Consequently the extraction conditions (V) have to be optimized to obtain a quantitative extraction within a reasonable time.

### **Collection Step**

The evaluation of the collection step requires that the analyte is spiked on an inert matrix such as stainless steel beads or sand, and that preliminary extraction conditions are chosen. If low recoveries are achieved it is either caused by insufficient solubility of analytes (often nonvolatiles) or poor trapping conditions (normally volatiles or semivolatiles). This is conveniently verified by reextraction of the residual with an organic solvent to check the presence of remaining analytes in the extraction cell. Once the collection has been set, the extraction step can be evaluated in more detail. Modest solubility of the analyte in the fluid is often sufficient for eluting the analytes from the extraction cell, and consequently for matrices where mild conditions can be applied, the collection step is rather uncomplicated. However, for certain analyte/matrix combinations stronger conditions are needed, and when major changes are made to the initial mild conditions (e.g., rise in temperature or modifier addition), the collection step must be verified once more.

### **Extraction Step**

**Analyte removal** For analyte removal to be successful, the applied extraction conditions must be strong enough to overcome analyte/matrix interactions such as adsorption of analytes on active surfaces, and speed up transport processes of analytes within the matrix. Often an increase in temperature or addition of a modifier significantly improves these processes. Analyte transport may be normal diffusion of the solute or may involve diffusion of the analyte through the fluid along pores in the matrix. Adsorption and desorption may take place during transport, but often the precise process will not be known even though a transport process similar to diffusion occurs.

**Analyte elution** The solubility of the analyte is the main factor controlling analyte elution, and to be successful the analyte must be sufficiently soluble in the fluid. Analytes with low affection for the matrix and with high solubility are eluted fast. There are different ways of increasing the solubility to achieve

a satisfying transport out of the cell such as decreasing or increasing the temperature and increasing the pressure. Also the addition of modifier may drastically increase solubility, but changes to these factors will also affect the initial removal of analytes from the matrix.

**Analyte/matrix considerations and sample types** Since both the transport of analytes from the matrix to the bulk fluid (removal) and the transport of analytes from the extraction cell to the collection device (elution) govern the extraction process, but to different extents for different analyte/matrix combinations, attempts have been made to classify samples. Two extreme groups of sample types can be identified based on flow rate studies; those where the analytes are recovered at a rate proportional to the fluid flow through the extraction cell, and those completely unaffected by the flow rate. In the former group of samples, the elution of the analytes is controlling the extraction speed, while in the latter group of samples, analyte removal is limiting the overall extraction procedure.

In general, samples controlled by the elution process contain high amounts of weakly bound analytes such as oil in oil seeds or bulk fat in pig meat. These samples will gain from increased flow rates and extraction times. Small sample sizes are preferable as is dynamic extraction processes. It is also important to increase analyte solubility. Samples controlled by the removal process often contain trace levels of very hardly bound analytes such as pesticides in soils or flame retardants in sediments. This group of sample will be relatively unaffected by flow rate and extraction time. Sample size is of less importance, and static extraction steps are often as efficient as dynamic. It is, however, important that the cell void volume is satisfactorily swept by the dynamic step. For these sample types, the most important thing to consider is the analyte/matrix interactions and to find ways to break these. This above description is of course a rough simplification. There are examples of fat components present at bulk levels in fishes that are hard to extract, just as there are soil and sediment samples containing trace levels of organic contaminants that are relatively easy to extract. Moreover, sample classification gives some insight into the various problems encountered for different sample types and the reader is referred to the Further Reading section for a more thorough discussion on sample classification and extraction behavior.

**Solubility, temperature, and pressure** Solubility depends not only on the nature of the fluid and its density, but also on the temperature. At a given temperature an increase in pressure will cause an

increase in density and therefore normally an increase in solubility. However, at a constant pressure an increase in temperature will cause a decrease in density, but not necessarily a decrease in solubility. For analytes with relatively high vapor pressures, the drop in density at increased temperatures is by far outweighed by solubility increases caused by vapor pressure effects at higher temperatures. Only compounds with insignificant vapor pressures will suffer from decreased densities at increased temperatures. A more detailed discussion on this topic can be found in the Further Reading section.

Depending on the type of application different combinations of pressure and temperature must be considered. For matrices where analyte elution from the cell is the main controlling factor, an increase in pressure might lead to faster extractions, as more analytes can be dissolved in the fluid when density is increased. However, for other matrices where the transport from the matrix to the bulk fluid is the critical factor, the temperature is often a better parameter to speed up the extraction as it will affect both desorption and diffusion processes. Care must, however, be taken when increasing the temperature so that thermal degradation of heat sensitive analytes is avoided.

**Modifiers** In some cases, the solvent strength of the supercritical fluid (e.g., carbon dioxide) is too low and, therefore, a modifier (e.g., methanol) can be added. The modifier will then cause increased solubility of more polar analytes. For example, carbon dioxide does not extract phospholipids to a very great extent, but by adding 5–10% methanol to the fluid, quantitative recoveries can easily be achieved. In this case the modifier is improving the transportation of the analyte out of the cell (elution). However, modifiers may also assist in the breaking of strong analyte/matrix interactions for certain applications, thereby, improving the transportation of the analyte from the matrix to the bulk fluid (removal). In these cases the analytes are normally sufficiently soluble in the fluid even without modifier present, and therefore, the main role of the modifier is to interact with certain regions of the matrix to which the analytes are bound by specific bonds. The nature of these interactions is often not known and various modifiers with differing chemical characteristics should preferably be tested. Once an appropriate modifier has been identified it can normally be added in relatively small concentrations (1–2%). It is, however, crucial to understand that the addition of modifiers will cause an increased coelution of unwanted matrix components. Additionally the collection efficiencies might decrease dramatically unless

analyte trapping is evaluated separately with these new experimental settings. Especially solid-phase trapping can suffer from severe analyte losses by modifier condensation in the trap. This can often easily be avoided by increasing the trap temperature.

## Sample Pretreatment

Normally solid samples are extracted and it is important to grind them in order to obtain homogeneity. Grinding can conveniently be performed with a mortar and pestle or a kitchen blender. Grinding often facilitates the extraction process, but in certain cases grinding might negatively influence the extraction. There is always a risk of losing analytes on the surface device or by evaporation, and carry-over effects must also be eliminated. Another important factor to consider is the water fraction of the sample. In some applications for more polar analytes, water might act as a modifier, but generally large amounts of water hinder the extraction process. It is common to handle the water fraction by mixing with an adsorbent such as sodium sulfate or Hydromatrix. Another option is freeze drying, but with a risk of losing volatiles. The analyst should consult the analytical chemical literature for specific applications as well as the Further Reading section for a more in-depth discussion of sample pretreatment.

## Verifying a Final Method

Once a method has been developed including a quantitative collection step combined with an exhaustive extraction process, it is necessary to test the method's wider applicability. This is unfortunately a critical step (and applies to all extraction techniques) since it is impossible to know the total amount of a certain analyte present in a real-world sample. There are certain strategies that should be adopted to verify the quantitateness of the method. The applied extraction conditions can be altered by increasing the temperature for matrices known to bind analytes hard or increase pressure for matrices known to contain bulk levels of analytes. If this improves extraction recoveries, the method is not final. Another important option is to compare SFE recoveries with recoveries obtained with conventional extraction procedures (Soxhlet) or other modern extraction techniques (PFE). If the same results are achieved, it is likely that the developed method is quantitative. It is also possible to reextract the SFE residual with another extraction method, thereby, detecting remaining analyte residues in the matrix. A third alternative is to utilize certified reference materials where

**Table 1** Examples of the use of SFE in sample preparation

Matrices	Examples of analytes extracted
Soils, sediments, sewage sludge, fly-ash	Polychlorinated biphenyls, polycyclic aromatic hydrocarbons, dioxins, phenols, pesticides, metals
Foods	Fats, drugs, additives, pesticides, dioxins, mycotoxins, vitamins
Biological fluids/tissues	Drugs, steroids, narcotics, polychlorinated biphenyls, polycyclic aromatic hydrocarbons, dioxins, metals
Polymers, food packaging, cosmetics	Low oligomers, surfactants, additives
Plant material	Essential oils, bioactive compounds, flavors, fragrances, pesticides, metals
Water samples, beverages	Phenols, metals, polycyclic aromatic hydrocarbons, alcohols, flavors, fragrances, bioactive compounds

concentrations have been independently determined by several laboratories.

## Applications

When considering SFE for a specific application, a literature search is always necessary. SFE has been successfully carried out on a wide variety of matrices as seen in **Table 1**. There are also reference methods available for certain applications, for example, the US Environmental Protection Agency (EPA) methods for extracting various pollutants from solid environmental matrices (EPA method 3560, 3561, and 3562). Many SFE manufacturers also provide their customers with application notes focused on specific analytes and matrices. Useful information can also be found in the Further Reading section.

See also: **Chromatography**: Overview; Principles. **Dioxins**. **Extraction**: Solvent Extraction Principles; Microwave-Assisted Solvent Extraction; Pressurized Fluid Extraction. **Food and Nutritional Analysis**: Overview. **Forensic Sciences**: Overview. **Gas Chromatography**: Overview. **Lipids**: Overview. **Liquid Chromatography**: Overview. **Pesticides**. **Pharmaceutical Analysis**:

Overview. **Polychlorinated Biphenyls**. **Polycyclic Aromatic Hydrocarbons**: Determination. **Polymers**: Natural Rubber; Synthetic; Polyurethanes. **Quality Assurance**: Quality Control; Instrument Calibration; Interlaboratory Studies; Reference Materials; Production of Reference Materials; Method Validation; Accreditation; Clinical Applications; Water Applications. **Sample Handling**: Comminution of Samples; Sample Preservation; Automated Sample Preparation; Robotics. **Sampling**: Theory; Practice. **Solvents**. **Supercritical Fluid Chromatography**: Overview; Applications. **Vitamins**: Overview; Fat-Soluble.

## Further Reading

- Burford MD, Bartle KD, and Hawthorne SB (1997) Directly coupled (online) SFE-GC: instrumentation and applications. *Advances in Chromatography* 37: 163–204.
- Caude M and Thiebaut D (eds.) (1999) *Practical Supercritical Fluid Chromatography and Extraction*. Amsterdam: Harwood Academic Publishers.
- Chester TL, Pinkston JD, and Raynie DE (1992) Supercritical fluid chromatography and extraction. *Analytical Chemistry* 64: 153R–170R.
- Chester TL, Pinkston JD, and Raynie DE (1994) Supercritical fluid chromatography and extraction. *Analytical Chemistry* 66: 106R–130R.
- Chester TL, Pinkston JD, and Raynie DE (1996) Supercritical fluid chromatography and extraction. *Analytical Chemistry* 68: 487–514.
- Chester TL, Pinkston JD, and Raynie DE (1998) Supercritical fluid chromatography and extraction. *Analytical Chemistry* 70: 301R–319R.
- Clifford T (1998) *Supercritical Fluids*. Oxford: Oxford University Press.
- Lee ML and Markides KE (eds.) (1990) *Analytical Supercritical Fluid Chromatography and Extraction*. Provo: Chromatography Conferences Inc.
- McHugh M and Krukonis V (1986) *Supercritical Fluid Extraction: Principles and Practice*. Stoneham: Butterworth-Heinemann.
- Smith RM (1999) Supercritical fluids in separation science – the dreams, the reality and the future. *Journal of Chromatography A* 856: 83–115.
- Smith RM and Hawthorne SB (eds.) (1997) *Supercritical Fluids in Chromatography and Extraction*. Oxford: Elsevier.
- Westwood SA (1994) Supercritical fluid extraction and its use in chromatographic sample preparation. *Toxicological and Environmental Chemistry* 43: 121.

## Solid-Phase Extraction

J S Fritz, Iowa State University, Ames, IA, USA

## Introduction

Despite highly selective and sensitive instrumentation for quantitative chemical analysis, a preliminary

**Table 1** Examples of the use of SFE in sample preparation

Matrices	Examples of analytes extracted
Soils, sediments, sewage sludge, fly-ash	Polychlorinated biphenyls, polycyclic aromatic hydrocarbons, dioxins, phenols, pesticides, metals
Foods	Fats, drugs, additives, pesticides, dioxins, mycotoxins, vitamins
Biological fluids/tissues	Drugs, steroids, narcotics, polychlorinated biphenyls, polycyclic aromatic hydrocarbons, dioxins, metals
Polymers, food packaging, cosmetics	Low oligomers, surfactants, additives
Plant material	Essential oils, bioactive compounds, flavors, fragrances, pesticides, metals
Water samples, beverages	Phenols, metals, polycyclic aromatic hydrocarbons, alcohols, flavors, fragrances, bioactive compounds

concentrations have been independently determined by several laboratories.

## Applications

When considering SFE for a specific application, a literature search is always necessary. SFE has been successfully carried out on a wide variety of matrices as seen in **Table 1**. There are also reference methods available for certain applications, for example, the US Environmental Protection Agency (EPA) methods for extracting various pollutants from solid environmental matrices (EPA method 3560, 3561, and 3562). Many SFE manufacturers also provide their customers with application notes focused on specific analytes and matrices. Useful information can also be found in the Further Reading section.

See also: **Chromatography**: Overview; Principles. **Dioxins**. **Extraction**: Solvent Extraction Principles; Microwave-Assisted Solvent Extraction; Pressurized Fluid Extraction. **Food and Nutritional Analysis**: Overview. **Forensic Sciences**: Overview. **Gas Chromatography**: Overview. **Lipids**: Overview. **Liquid Chromatography**: Overview. **Pesticides**. **Pharmaceutical Analysis**:

Overview. **Polychlorinated Biphenyls**. **Polycyclic Aromatic Hydrocarbons**: Determination. **Polymers**: Natural Rubber; Synthetic; Polyurethanes. **Quality Assurance**: Quality Control; Instrument Calibration; Interlaboratory Studies; Reference Materials; Production of Reference Materials; Method Validation; Accreditation; Clinical Applications; Water Applications. **Sample Handling**: Comminution of Samples; Sample Preservation; Automated Sample Preparation; Robotics. **Sampling**: Theory; Practice. **Solvents**. **Supercritical Fluid Chromatography**: Overview; Applications. **Vitamins**: Overview; Fat-Soluble.

## Further Reading

- Burford MD, Bartle KD, and Hawthorne SB (1997) Directly coupled (online) SFE-GC: instrumentation and applications. *Advances in Chromatography* 37: 163–204.
- Caude M and Thiebaut D (eds.) (1999) *Practical Supercritical Fluid Chromatography and Extraction*. Amsterdam: Harwood Academic Publishers.
- Chester TL, Pinkston JD, and Raynie DE (1992) Supercritical fluid chromatography and extraction. *Analytical Chemistry* 64: 153R–170R.
- Chester TL, Pinkston JD, and Raynie DE (1994) Supercritical fluid chromatography and extraction. *Analytical Chemistry* 66: 106R–130R.
- Chester TL, Pinkston JD, and Raynie DE (1996) Supercritical fluid chromatography and extraction. *Analytical Chemistry* 68: 487–514.
- Chester TL, Pinkston JD, and Raynie DE (1998) Supercritical fluid chromatography and extraction. *Analytical Chemistry* 70: 301R–319R.
- Clifford T (1998) *Supercritical Fluids*. Oxford: Oxford University Press.
- Lee ML and Markides KE (eds.) (1990) *Analytical Supercritical Fluid Chromatography and Extraction*. Provo: Chromatography Conferences Inc.
- McHugh M and Krukonis V (1986) *Supercritical Fluid Extraction: Principles and Practice*. Stoneham: Butterworth-Heinemann.
- Smith RM (1999) Supercritical fluids in separation science – the dreams, the reality and the future. *Journal of Chromatography A* 856: 83–115.
- Smith RM and Hawthorne SB (eds.) (1997) *Supercritical Fluids in Chromatography and Extraction*. Oxford: Elsevier.
- Westwood SA (1994) Supercritical fluid extraction and its use in chromatographic sample preparation. *Toxicological and Environmental Chemistry* 43: 121.

## Solid-Phase Extraction

J S Fritz, Iowa State University, Ames, IA, USA

## Introduction

Despite highly selective and sensitive instrumentation for quantitative chemical analysis, a preliminary



sample preparation step often is necessary for sample cleanup or to isolate components of interest from a sample matrix. In the past, liquid–liquid extraction (LLE) has played a major role in sample preparation. However, recovery of sample components by liquid extraction tends to be slow and labor intensive. The disposal of organic solvents used in liquid extraction is an environmental concern.

In solid-phase extraction (SPE), solutes are extracted from a liquid phase into a solid phase. Most commonly, the liquid phase is a predominately aqueous sample solution and the extractive solid phase consists of small porous particles of silica with a bonded organic outer layer or else it is an organic polymer, such as cross-linked polystyrene. The extraction can take place in a batch mode in which the solid extractant particles are intimately mixed with the sample and then filtered off. However, in chemical analysis it is more common to use a flow-through mode in which the liquid sample is passed through a bed of the solid extractant packed in a small tube. This technique gives more complete extraction of the desired analytes than the batch mode, in which there is only a single equilibration of analytes between the liquid and solid phases.

Substances that have been extracted by the solid particles can be removed by washing with an appropriate liquid solvent. For example, most organic analytes are eluted from an SPE tube (column) with an organic solvent such as acetone, acetonitrile, or methanol. Usually, the volume of solvent needed for complete elution of the analytes is much smaller than the original sample volume. An increased concentration of the analytes is thus achieved.

Extracted molecules can often be removed from solid particles by heating in a gentle stream of a non-reactive carrier gas. This is a convenient way to transfer the molecules into a gas chromatograph for analysis.

In some respects the mechanism of SPE is similar to that of LLE where an extractive organic solvent is added to the aqueous sample solution and the vessel is agitated vigorously to create a temporary emulsion. The emulsion consists of very small spherical solvent droplets suspended in the aqueous phase. The interfacial contact area between the two phases must be quite large in order to promote rapid mass transfer of the analytes from the aqueous to the organic phase. To complete the desired extraction, a finite amount of time is required for the emulsion to break and the phases coalesce into separate layers. The lower layer is carefully drawn off to complete the separation.

In SPE, the solid particles must be small (typically 10–50  $\mu\text{m}$  in diameter) and porous with a large surface area (often 200–800  $\text{m}^2\text{g}^{-1}$ ) to provide an intimate interfacial contact for fast, efficient mass

transfer. A major attribute of SPE is that excellent contact is accomplished simply by allowing the liquid sample to flow through a bed of spherical solid particles contained in a tube or membrane.

The advantages of solid-phase extraction over conventional LLE are:

1. Faster, with less manipulation. A sample can be quickly passed through an SPE column or cartridge by means of a pump or with gentle pressure or suction. After a quick rinse, the extracted substances can be washed from the column by a small volume of an organic solvent or another appropriate eluent. These steps can be automated readily. By contrast, simple solvent extraction requires a considerable amount of manipulation in adding the extractive liquid, shaking, waiting for the emulsion to break, and carefully separating the two liquid phases.
2. Much smaller amounts of liquid organic solvents are used. The large quantities of organic solvents used in analytical separations have become an environmental concern. Aqueous samples become contaminated with organic solvents and evaporative concentration of the extracts pollutes the air with organic vapors.
3. Less stringent requirements for separation. SPE is a multistage separation method and as such requires only a reasonable difference in extractability to separate two solutes. Separation of two solutes by LLE requires complete extraction of one solute in a single equilibration and essentially zero extraction of the other.
4. Higher concentration factors. The concentration factor tells how many times more concentrated a substance is in the extract than it is in the original sample. In SPE, concentration factors of 1000 or more are possible.

## SPE Particles

Bonded-phase silica materials are the most widely used particles for SPE. Since the technology of bonded-phase silica packings for LC columns was already well developed, it was relatively easy for suppliers to offer similar materials in a form suitable for SPE. The most common types are listed in **Table 1**. Octadecylsilane, or sometimes octylsilane, particles are used for retention of relatively hydrophobic organic analytes from predominately aqueous samples. A more polar particle, such as cyanopropylsilane, is selected for SPE of less hydrophobic analytes. Cation- and anion-exchange resins are used in small precolumns to concentrate sample ions from very dilute aqueous samples prior to their

**Table 1** Bonded-phase silica particles for SPE, listed in approximate order of increasing polarity

Phase	Designation
Octadecyl	C18
Octyl	C8
Cyclohexyl	CH
Phenyl	PH
Cyanopropyl	CN
Diol	2OH
Carboxymethyl	CBA, Weak cation exchanger
Aminopropyl	NH <sub>2</sub> , Weak anion exchanger
Propylbenzene sulfonic acid	SCX, Strong cation exchanger
Trimethylaminopropyl	SAX, Strong anion exchanger

determination by ion chromatography. Cations, such as protonated amine analytes, are retained with some selectivity by particles with a strong- or weak-acid functional group. Anionic analytes are taken up by a material with amino or quaternary ammonium groups.

Excellent polymeric particles for SPE are now available commercially. Most of these are cross-linked polystyrene materials, but some polymethacrylate particles are also available. Organic polymers are more stable than silica materials at alkaline pH values and also tend to have a larger surface area, which enhances their extractive ability. Introduction of  $-\text{CH}_2\text{OH}$  or  $-\text{COCH}_3$  groups into polystyrene resins enhances their ability to extract more polar organic analytes, and also improves mass transport of analytes from the aqueous to the solid phase. A co-polymer of divinylbenzene and vinylpyrrolidone, known as 'Oasis', is also well suited for SPE of relatively polar analytes.

Restricted access polymers are finding extensive use in clinical analysis. Large biomolecules pass quickly through a small column while smaller organic solutes are retained and concentrated by SPE.

Various activated carbon sorbents were used extensively in the early days of SPE. Although activated carbon has very strong sorptive ability, subsequent elution of analytes is often kinetically slow and incomplete. Nowadays, porous graphitic carbon particles are available that work very well for SPE, especially for phenolic compounds that are poorly retained by other sorbents.

A class of compounds known as molecular sieves has found numerous uses in chemical analysis. Molecular sieves are silica or metal-silica particles that contain cavities that are usually 5–6 Å in diameter. Compounds small enough to enter the cavities are sorbed by the molecular sieve while larger molecules are excluded. Virtually all molecular sieves are hydrophilic. However, a molecular sieve known as Silicalite has the rather unique property of being

**Table 2** Recovery of organic analytes by SPE on a Silicalite column

Class	Compound	Recovery (%)
Aldehydes	<i>trans</i> -Crotonaldehyde	91
	n-Valeraldehyde	100
Ketones	Benzaldehyde	89
	Acetones	94
	2-Butanone	92
	2-Pentanone	90
	3-Pentanone	41
	4-Methyl-2-pentanone	104
Esters	Methyl formate	74
	Methyl acetate	85
	Ethyl formate	83
	Ethyl acetate	91
	Ethyl butyrate	90
	Methyl benzoate	68
	Pentyl benzoate	82
	Chloroform	82
Chlorinated alkanes	1,2-Dichlorethane	83
	1,1-Dichlorethane	80
	1,2-Dichloropropane	85
	Carboxylic acids	
Amines	Acetic	2
	Propionic	62
	Butyric	78
	Valeric	95
	Propylamine	63
	n-Butylamine	72
	<i>tert</i> -Butylamine	64
	Pyridine	77

hydrophobic, at least within the cavities of the sieve. Silicalite is commercially available from UOP and has been used very successfully in SPE to retain small, polar organic analytes. Even ethanol and propanol are retained to some extent. Percentage recoveries of several test compounds by SPE on a short Silicalite column and subsequent elution are given in **Table 2**.

Of the many sorbents available for SPE of organic solutes from predominately aqueous samples, bonded-phase silica particles are undoubtedly the most widely used. But are they best? The relative extractive abilities of the various solid-phase extractants can perhaps best be compared by measuring the retention factors ( $k$ ) with analytes of different chemical structure. Because  $k$  values are often too high to measure readily in water alone, they can be measured in a series of methanol–water solutions and the value in water alone ( $k_w$ ) obtained by linear extrapolation. The values of  $k_w$  for some representative solutes are compared in **Table 3** for C18 silica, a cross-linked polystyrene polymer (PRP-1), and pyrolytic graphitized carbon (PGC). These and other results reveal that polystyrene particles give higher  $k_w$  values than C18 silica in virtually every case. PGC gives lower values than polystyrene sorbents for most organic solutes except for some polyhydroxy compounds.



**Table 3** Comparison of  $k_w$  values with C18 silica, PRP-1, and PGC sorbents

Solute	$k_w$ value		
	C18 silica	PRP-1	PGC
Acetophenone	63	1250	—
Aniline	12	200	22
Benzene	160	3300	28
Benzoic acid	80	2000	250
1,4-Dihydroxybenzene	—	7	140
Phenol	35	250	63
Toluene	560	14 000	—

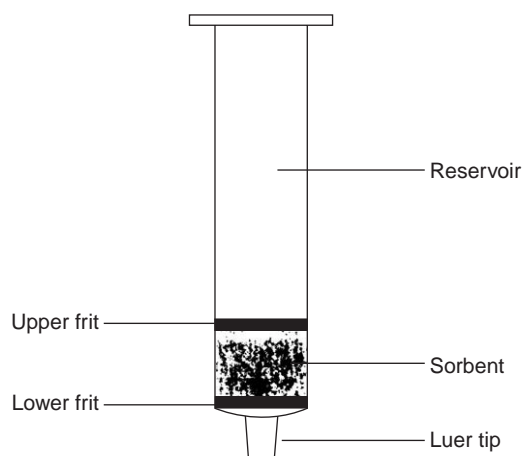
## Steps in a Solid-Phase Extraction

In its simplest form, SPE can be carried out in a small column packed with a short bed of solid extractive particles (see **Figure 1**). Commercial cartridges are very popular for SPE. Typical cartridges contain ~300, 600, or 900 mg of packed bed. As a general guideline, cartridges can extract up to ~1 mg of sample components per 100 mg of packing.

The SPE process can be divided into four main steps:

1. **Conditioning.** Without some pretreatment an aqueous sample will flow in small channels through the hydrophobic solid phase without making good interfacial contact. Conditioning the column with a little organic solvent such as methanol before introduction of the sample will largely avoid this difficulty.
2. **Adsorption.** The liquid sample is passed at a reasonably constant rate through the column with the aid of suction, applied pressure, or a pump. Columns or cartridges packed with 50–100  $\mu\text{m}$  particles generally necessitate the use of a resin bed 50 mm or more in length. Columns packed with smaller particles are more efficient and can be used with a shorter packing bed.
3. **Washing.** When the sample has passed through the column, a small volume of a wash liquid is used to remove the remaining, unextracted sample. Water, or water containing a low percentage of an organic solvent, is a typical wash solution.
4. **Elution.** In the elution step the adsorbed analytes are removed from the solid extractant and are returned to a liquid phase that is suitable for analytical measurement. Most commonly, the eluting phase is an organic liquid, although it is often possible to thermally desorb analytes with the aid of a gas stream.

The eluting solvent must be compatible with the analytical measurement methods to be used. For

**Figure 1** A simple SPE column.

example, when gas chromatography is to be used, the eluting solvent should have a fairly low boiling point so that the large solvent peak will not interfere with the sample peaks. The eluting solvent should be mostly free from impurities that might give interfering chromatographic peaks.

Many different organic solvents and mixtures of solvents have been proposed for elution of organic analytes in SPE. A study of the elution of aromatic analytes (benzene, naphthalene, anthracene, and chrysene) showed methanol and acetonitrile to be inefficient eluents. The best solvents for these aromatics were tetrahydrofuran and ethyl acetate. Methylene chloride was also very effective but formed an immiscible layer with the residual water in the column.

## Formats

Packed cartridges and mini-columns are often used for SPE. A typical cartridge includes a plastic or glass tube with porous metal or plastic frits at both ends. It is filled with 100–500 mg of 40  $\mu\text{m}$  particles. Although SPE cartridges are popular and easy to use, the relatively large particle size of the sorbent and mediocre packing efficiency necessitate the use of a relatively slow flow.

In 1989, a new family of sorbents for SPE was introduced by the 3 M Company under the trade name Empore. The sorbent particles are a nonspherical bonded-phase silica or a porous poly(styrene-DVB) held together by PTFE fibrils. Approximately 90% of the weight of the membrane is made up of the sorbent particles. The particles are close together in the membrane but not necessarily touching one another. The resin-impregnated membranes are flexible and generally are ~0.5 mm thick.

Sample solutions flow evenly through the thin ( $\sim 0.5$  mm thick) membrane making intimate contact with the extractive particles. It would be very difficult to avoid uneven flow (channeling) through a 0.5 mm bed of loose particles packed into a SPE tube. Studies with very dilute solutions of a dye show that the dye is taken up by the uppermost part of the membrane and that the dye uptake is very even over the entire area of the circular disk. Owing to the small volume and extraction efficiency of the membrane disk, elution of retained analytes requires a much smaller volume of eluting solvent than SPE cartridges.

In laboratories that have a large number of samples to analyze, some degree of automation has become essential. SPE is amenable to automation and is used widely for sample pretreatment. A popular system consists of an array of extraction plates with wells that hold 96 collection tubes, usually in rows of eight tubes each. An assembly of eight automatic pipettes applies samples to the system one row at a time. Semiautomatic elution is also employed. With this system the time needed to process 96 samples is reduced to  $< 1$  h.

See also: **Extraction:** Solvent Extraction Principles; Solid-Phase Microextraction.

## Further Reading

- Ambrose DM and Fritz JS (1997) Silicalite as a sorbent for solid-phase extraction. *Journal of Chromatography* 770: 301–310.
- Chambers TK and Fritz JS (1996) High-performance liquid chromatography with nonaqueous solvents. *Journal of Chromatography A* 728: 271–278.
- Fritz JS (1990) *Analytical Solid-Phase Extraction*. New York: Wiley-VCH.
- Hennion MC and Coquart V (1993) Comparison of reversed-phase extraction sorbents for the on-line trace enrichment of polar organic compounds in environmental aqueous samples. *Journal of Chromatography* 642: 211–224.
- Pawliszyn J (ed.) (2002) *Sampling and Sample Preparation for Field and Laboratory*. Amsterdam: Elsevier.
- Pawliszyn J (2003) Sample preparation: quo vadis. *Analytical Chemistry* 75: 2543–2558.
- Poole CF (2003) New trends in solid-phase extraction. *Trends in Analytical Chemistry* 22: 362–373.
- Simpson N (1997) *Solid-Phase Extraction: Principles, Strategies, and Applications*. New York: Dekker.
- Sun JJ and Fritz JS (1990) Chemically modified resins for HPLC. *Journal of Chromatography* 522: 95–105.
- Thurman EM and Mills MS (1998) *Solid-Phase Extraction, Principles and Practice*. New York: Wiley.

## Solid-Phase Microextraction

**C Ross**, Brock University, St Catharines, ON, Canada

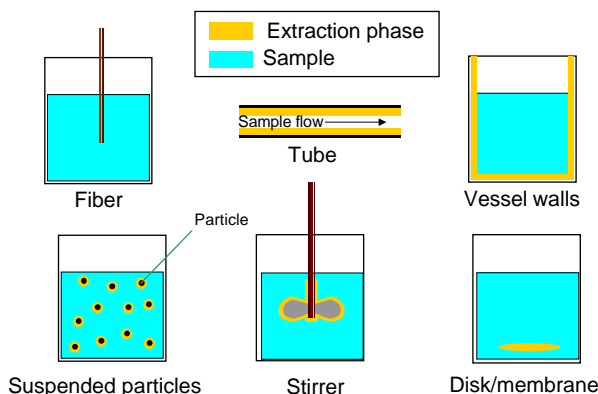
**J Pawliszyn**, University of Waterloo, Waterloo, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Solid-phase microextraction (SPME) was developed to address the need to facilitate rapid sample preparation both in the laboratory and onsite where the investigated system is located. In this technique, a small amount of extracting phase dispersed on a solid support is exposed to the sample for a well-defined period of time. In one approach, a partitioning equilibrium between the sample matrix and extraction phase is reached. In this case, convection conditions do not affect the amount extracted. In the second approach utilizing short time pre-equilibrium extraction, if convection/agitation is constant, the amount of analyte extracted is related to extraction time. Quantification can then be performed based on timed

accumulation of analytes in the coating. **Figure 1** illustrates several implementations of SPME that have been considered. They include primarily open bed extraction concepts such as coated fibers, vessels, agitation mechanism disks, but in-tube approaches are also considered. Some implementations better address issues associated with agitation while other



**Figure 1** Configurations of solid-phase microextraction.

Sample solutions flow evenly through the thin ( $\sim 0.5$  mm thick) membrane making intimate contact with the extractive particles. It would be very difficult to avoid uneven flow (channeling) through a 0.5 mm bed of loose particles packed into a SPE tube. Studies with very dilute solutions of a dye show that the dye is taken up by the uppermost part of the membrane and that the dye uptake is very even over the entire area of the circular disk. Owing to the small volume and extraction efficiency of the membrane disk, elution of retained analytes requires a much smaller volume of eluting solvent than SPE cartridges.

In laboratories that have a large number of samples to analyze, some degree of automation has become essential. SPE is amenable to automation and is used widely for sample pretreatment. A popular system consists of an array of extraction plates with wells that hold 96 collection tubes, usually in rows of eight tubes each. An assembly of eight automatic pipettes applies samples to the system one row at a time. Semiautomatic elution is also employed. With this system the time needed to process 96 samples is reduced to  $< 1$  h.

See also: **Extraction:** Solvent Extraction Principles; Solid-Phase Microextraction.

## Further Reading

- Ambrose DM and Fritz JS (1997) Silicalite as a sorbent for solid-phase extraction. *Journal of Chromatography* 770: 301–310.
- Chambers TK and Fritz JS (1996) High-performance liquid chromatography with nonaqueous solvents. *Journal of Chromatography A* 728: 271–278.
- Fritz JS (1990) *Analytical Solid-Phase Extraction*. New York: Wiley-VCH.
- Hennion MC and Coquart V (1993) Comparison of reversed-phase extraction sorbents for the on-line trace enrichment of polar organic compounds in environmental aqueous samples. *Journal of Chromatography* 642: 211–224.
- Pawliszyn J (ed.) (2002) *Sampling and Sample Preparation for Field and Laboratory*. Amsterdam: Elsevier.
- Pawliszyn J (2003) Sample preparation: quo vadis. *Analytical Chemistry* 75: 2543–2558.
- Poole CF (2003) New trends in solid-phase extraction. *Trends in Analytical Chemistry* 22: 362–373.
- Simpson N (1997) *Solid-Phase Extraction: Principles, Strategies, and Applications*. New York: Dekker.
- Sun JJ and Fritz JS (1990) Chemically modified resins for HPLC. *Journal of Chromatography* 522: 95–105.
- Thurman EM and Mills MS (1998) *Solid-Phase Extraction, Principles and Practice*. New York: Wiley.

## Solid-Phase Microextraction

**C Ross**, Brock University, St Catharines, ON, Canada

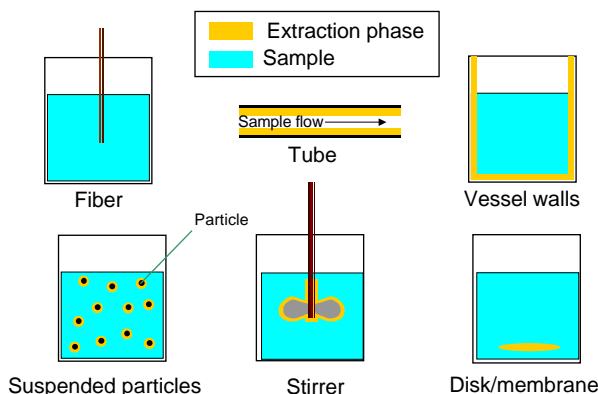
**J Pawliszyn**, University of Waterloo, Waterloo, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Solid-phase microextraction (SPME) was developed to address the need to facilitate rapid sample preparation both in the laboratory and onsite where the investigated system is located. In this technique, a small amount of extracting phase dispersed on a solid support is exposed to the sample for a well-defined period of time. In one approach, a partitioning equilibrium between the sample matrix and extraction phase is reached. In this case, convection conditions do not affect the amount extracted. In the second approach utilizing short time pre-equilibrium extraction, if convection/agitation is constant, the amount of analyte extracted is related to extraction time. Quantification can then be performed based on timed

accumulation of analytes in the coating. **Figure 1** illustrates several implementations of SPME that have been considered. They include primarily open bed extraction concepts such as coated fibers, vessels, agitation mechanism disks, but in-tube approaches are also considered. Some implementations better address issues associated with agitation while other



**Figure 1** Configurations of solid-phase microextraction.

implementations address sample introduction into the analytical instrument. It should be noted that SPME was originally named after the first experiment using an SPME device that involved extraction onto solid fused silica fibers contained in a syringe needle. However, the term SPME broadly defines the appearance of the small amount of the solidified extraction phase, frequently involving polymers relative to a liquid or gaseous donor phase (matrix), even though it is recognized that the polymeric extraction phase is not always technically a solid (for example, polydimethylsiloxane). In this article, the principles of SPME and the theoretical aspects of SPME method optimization will be discussed.

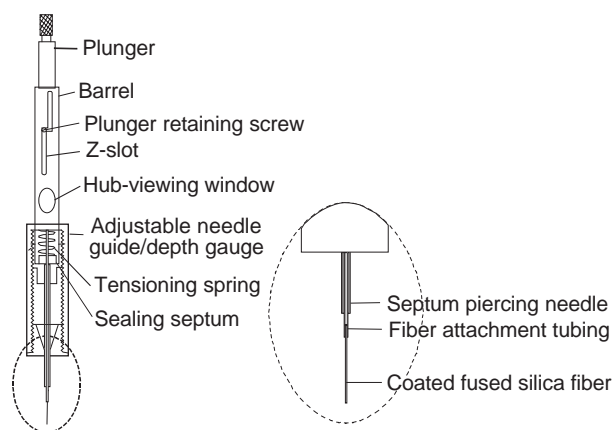
SPME has been introduced as a modern alternative to traditional sample preparation technology and it is able to address many of the requirements put forward by analytical researchers. This technique eliminates the use of organic solvents, substantially shortens analysis time and allows for convenient automation of the sample preparation step. SPME can integrate sampling with sample preparation, making it suitable for onsite analysis and process monitoring. The configurations and operation of the SPME devices are very simple. The technology is designed to greatly simplify sample preparation. This feature, however, creates a false impression that the extraction is a simple, almost trivial process. This misunderstanding frequently results in disappointments. It should be emphasized that the fundamental processes involved in SPME are similar to more traditional techniques and therefore challenges to develop successful methods are analogous. The nature of target analytes and complexity of sample matrix determine the level of difficulties in accomplishing a successful extraction. The simplicity, speed, and convenience of the extraction devices primarily impacts the costs of practical implementation and automation of the developed methods.

## Principles of Solid-Phase Microextraction

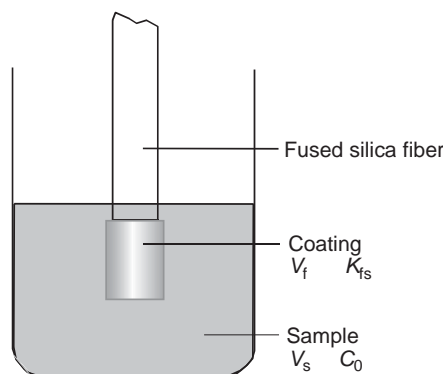
In SPME, a small amount of extracting phase associated with a solid support is placed in contact with the sample matrix for a predetermined amount of time. If the time is long enough, a concentration equilibrium is established between the sample matrix and the extraction phase. When equilibrium conditions are reached, exposing the fiber for a longer time does not accumulate more analytes. Two different implementations of the SPME technique have been explored extensively to date. One is associated with a tube design and the other with a fiber design. The tube design can use very similar arrangements as

solid-phase extraction, but the primary difference, in addition to volume of the extracting phase, is that the objective of SPME is never an exhaustive extraction. This substantially simplifies the design of systems. For example, in-tube SPME for liquid analysis uses 0.25 mm ID tubes and  $\sim 0.1 \mu\text{L}$  of extraction phase as related to breakthrough is not relevant as exhaustive extraction is not an objective. In fact, the objective of the experiment is to produce full breakthrough as soon as possible as this indicates equilibrium extraction has been reached.

A more traditional approach to SPME involves the use of coated fibers. The first commercially available SPME fiber device is shown in Figure 2. The transport of analytes from the matrix into the coating begins as soon as the coated fiber has been placed in contact with the sample (Figure 3). Typically, SPME extraction is considered to be complete when the analyte concentration has reached distribution equilibrium between the sample matrix and the fiber coating. In practice, this means that once equilibrium is reached, the extracted amount is constant within



**Figure 2** Design of the first commercial SPME device made by Supelco.



**Figure 3** Microextraction with SPME.  $V_f$ , volume of fiber coating;  $K_{fs}$ , fiber/sample partition coefficient;  $V_s$ , volume of sample;  $C_0$ , initial concentration of analyte in the sample.



the limits of experimental error and it is independent of further increase of extraction time. The equilibrium conditions can be described as:

$$n = \frac{K_{fs} V_f V_s C_0}{K_{fs} V_f + V_s} \quad [1]$$

where  $n$  is the number of moles extracted by the coating,  $K_{fs}$  is a fiber coating/sample matrix distribution constant,  $V_f$  is the fiber coating volume,  $V_s$  is the sample volume, and  $C_0$  is the initial concentration of a given analyte in the sample.

Strictly speaking, this discussion is limited to partitioning equilibrium involving liquid polymeric phases such as polydimethylsiloxane. The method of analysis for solid sorbent coatings is analogous for low analyte concentration, since the total surface area available for adsorption is proportional to the coating volume if constant porosity of the sorbent is assumed. For high analyte concentrations, saturation of the surface can occur, resulting in nonlinear isotherms. Similarly, high concentration of a competitive interference compounds can displace the target analyte from the surface of the sorbent. Equation [1], which assumes that the sample matrix can be represented as a single homogeneous phase and that no headspace is present in the system, can be modified to account for the existence of other components in the matrix by considering the volumes of the individual phases and the appropriate distribution constants. The extraction can be interrupted and the fiber analyzed prior to equilibrium. However, in order to obtain reproducible data, constant convection conditions and careful timing of the extraction are necessary.

Simplicity and convenience of operation make SPME a superior alternative to more established techniques in a number of applications. In some cases, the technique facilitates unique investigations. Equation [1] indicates that, after equilibrium has been reached, there is a direct proportional relationship between sample concentration and the amount of analyte extracted. This is the basis for analyte quantification. The most visible advantages of SPME exist at the extremes of sample volumes. Because the setup is small and convenient, coated fibers can be used to extract analytes from very small samples. For example, SPME devices are used to probe for substances emitted by a single flower bulb during its life span. The use of submicrometer diameter fibers permits the investigation of single cells. Since SPME does not extract target analytes exhaustively, its presence in a living system should not result in significant disturbance. In addition, SPME facilitates speciation in natural systems since the presence of a minute fiber, which removes small amounts of analyte, is not likely to disturb chemical equilibria in the system. It should

be noted that the fraction of analyte extracted increases as the ratio of coating to sample volume increases. Complete extraction can be achieved for small sample volumes when distribution constants are reasonably high. This observation can be used to advantage if exhaustive extraction is required as it is very difficult to work with small sample volumes using conventional sample preparation techniques. SPME also allows rapid extraction and transfer to an analytical instrument. SPME can be used for studies of the distribution of analytes in a complex multiphase system and speciate different forms of analytes in a sample. These features result in an additional advantage when investigating intermediates in the system.

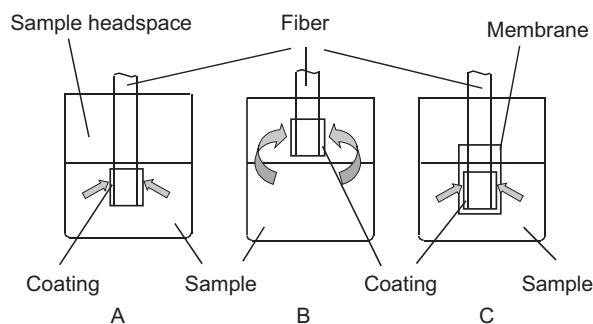
In addition, when the sample volume is very large, eqn [1] can be simplified to:

$$n = K_{fs} V_f C_0 \quad [2]$$

which points to the usefulness of the technique for field applications. In this equation, the amount of extracted analyte is independent of the volume of the sample. In practice, there is no need to collect a defined sample prior to analysis as the fiber can be exposed directly to the ambient air, water, production stream, etc. The amount of extracted analyte will correspond directly to its concentration in the matrix without being dependent on the sample volume. When the sampling step is eliminated, the whole analytical process can be accelerated and errors associated with analyte losses through decomposition or adsorption on the sampling container walls will be prevented.

### Extraction Modes with Coated Fiber

Three basic types of extractions can be performed using SPME: direct extraction, headspace configuration, and a membrane protection approach. **Figure 4** illustrates the differences among these modes. In the direct extraction mode (**Figure 4A**), the coated fiber is inserted into the sample and the analytes are transported directly from the sample matrix to the extracting phase. To facilitate rapid extraction, some level of agitation is required to transport analytes from the bulk of the solution to the vicinity of the fiber. For gaseous samples, natural convection of air is sufficient to facilitate rapid equilibration. For aqueous matrices, more efficient agitation techniques, such as fast sample flow, rapid fiber or vial movement, stirring, or sonication is required. These conditions are necessary to reduce the effect caused by the 'depletion zone' produced close to the fiber as a result of fluid shielding and slow diffusion coefficients of analytes in liquid matrices.



**Figure 4** Modes of SPME operation: (A) direct extraction, (B) headspace SPME, and (C) membrane-protected SPME.

In the headspace mode, the analytes need to be transported through the air barrier before they can reach the coating. This modification serves primarily to protect the fiber coating from damage by high molecular mass and other nonvolatile interferences present in the sample matrix, such as humic materials or proteins. The headspace mode also allows modification of the matrix, such as a change of the pH, without damaging the fiber. Amounts of analyte extracted from the same vial at equilibrium into the coating using direct and headspace sampling are identical as long as sample and gaseous headspace volumes are the same. This is due to the fact that the equilibrium concentration is independent of fiber location in the sample/headspace system. If the above condition is not satisfied for very volatile analytes, a significant difference in sensitivity will be observed between the direct and headspace methods.

The choice of sampling mode has a significant impact on extraction kinetics. When the fiber coating is in the headspace, the analytes are removed from the headspace first, followed by indirect extraction from the matrix as shown in **Figure 4B**. Overall mass transfer to the fiber is typically limited by mass transfer rates from the sample to the headspace. Therefore, volatile analytes are extracted faster than semivolatiles since they are at a higher concentration in the headspace and contribute to faster mass transport rates through the headspace. Temperature has a significant effect on the kinetics of the process by determining the vapor pressure of analytes. In fact, the equilibration times for volatiles are shorter in the headspace SPME mode than for direct extraction under similar agitation conditions. This outcome is produced by two factors: a substantial portion of the analyte is in the headspace prior to extraction and diffusion coefficients in the gaseous phase are typically four orders of magnitude larger than in liquid media. As concentrations of semivolatiles in the gaseous phase at room temperature are typically small, overall mass transfer rates are substantially lower

and result in longer extraction times. Extraction times can be improved by using very efficient agitation or by increasing the extraction temperature.

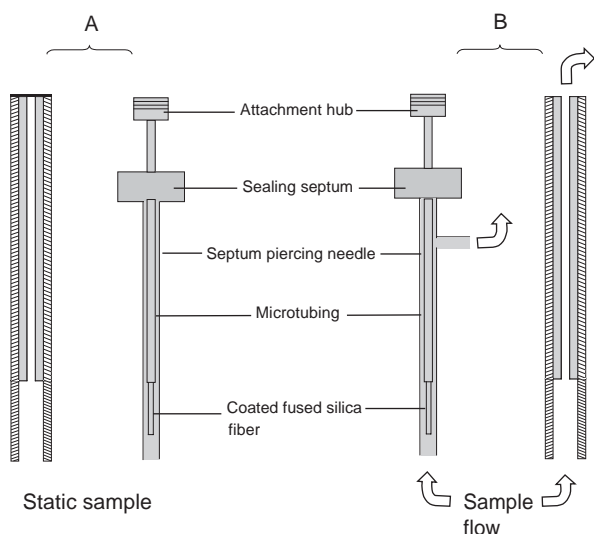
The effect of elevated sampling temperature is a lowering of fiber sample partition coefficients. This decreases the amount extracted at equilibrium but it may be acceptable if target limits of detection can still be reached. This effect is demonstrated dramatically with the analysis of amphetamines. At room-temperature extraction, a very long equilibrium extraction time is observed, but ultimately the highest amount was extracted. Conversely, the highest temperature tested (73°C) produced a very short equilibration time (~5 min) but a significantly lower equilibrium amount extracted. The combination of an increase in Henry constant and diffusion coefficient to improve kinetics at higher temperature, with a decrease in fiber capacity determined by the distribution constant, contribute to a substantial decrease in extraction time. Therefore, it is important to ensure constant agitation conditions in experiments and acceptable extraction times in order to obtain good precision.

In the third mode, indirect SPME extraction through a membrane (**Figure 4C**), the fiber is separated from the sample with a selective membrane. The main purpose of the membrane barrier is to protect the fiber against damage. Membrane protection is advantageous for determination of analytes having volatilities too low for the headspace approach. In addition, a membrane made from appropriate material can add a certain degree of selectivity to the extraction process. The kinetics of membrane extraction is substantially slower than for direct extraction because the analytes must diffuse through the membrane before they can reach the coating. The use of thin membranes and increased extraction temperatures will result in faster extraction times. Thicker membranes can be used to slow down the mass transfer through the membrane resulting in the time-weighted average (TWA) measurement for aqueous and gaseous samples.

### Extraction Modes with In-Tube SPME

**Figure 5** illustrates the two fundamental approaches to in-tube SPME: (1) active or dynamic, when the analytes are passed through the tube; and (2) passive or static, when the analytes are transferred into the sorbent using diffusion. In both of these approaches, the coating may be supported on a fused silica rod, or coated on the inside of a tube or capillary. The theoretical aspects of the extraction processes that use these geometric arrangements will be discussed below.





**Figure 5** Comparison of (A) passive versus (B) dynamic modes of in-tube extraction.

**Dynamic in-tube SPME** In this system, we assume the use of a piece of fused silica capillary, internally coated with a thin film of extracting phase (a piece of open tubular capillary gas chromatography (GC) column), or that the capillary is packed with extracting phase dispersed on an inert supporting material (a piece of micro-LC capillary column).

The front of analyte migrates through the capillary with a speed proportional to the linear velocity of the sample and inversely related to the partition ratio. For short capillaries with small dispersion, the extraction time can be assumed to be similar to the time required for the center of the band to reach the end of the capillary. The extraction time is proportional to the length of the capillary and inversely proportional to the linear flow rate of the fluid. Extraction time also increases with an increase in the coating/sample distribution constant and with the thickness of the extracting phase but decreases with an increase in the void volume of the capillary. An increase in the coating/sample distribution constant produces an increase in absolute amount extracted. It has been observed in many cases that increases in extraction amounts can be achieved by preconditioning the capillary with methanol or another appropriate solvent prior to extraction.

It should be emphasized that the above discussion is valid only for direct extraction when the sample matrix passes through the capillary. This approach is limited to particulate-free gas and clean water samples. The headspace SPME approach can broaden the application of in-tube SPME. In that case, careful consideration of the mass transfer between sample and headspace should be given in order to describe the process properly. Also, if the flow is very rapid

producing turbulent behavior and the coating/sample distribution constant is not high, then perfect agitation conditions are met and eqn [3] can be used to estimate equilibration times. In this case equilibration time,  $t_e$ , is assumed to be achieved when 95% of the equilibrium amount of analyte is extracted from the sample:

$$t_e = t_{95\%} = \frac{b_t^2}{2D_f} \quad [3]$$

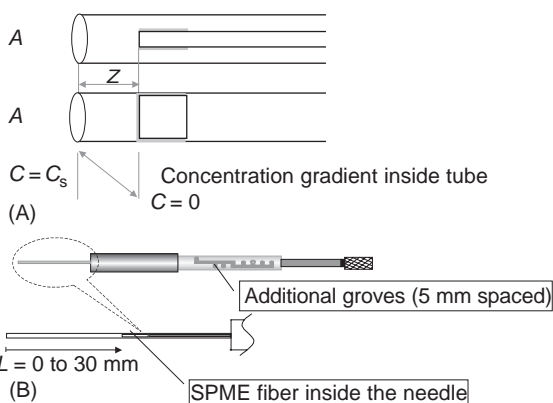
In this equation,  $b_t$  refers to the thickness of the sorbent material, and  $D_f$  refers to analyte diffusion coefficient in the sorbent.

Removal of analytes from a tube is an elution problem analogous to frontal chromatography and has been discussed in detail. In general, if the desorption temperature of a gas chromatogram is high and thin coatings are used, then all the analytes are in the gas phase as soon as the coating is placed in the injector. The desorption time then corresponds to the elution of two void volumes of the capillary. For liquid desorption, the desorption volume can be even smaller since the analytes can be focused at the front of the desorption solvent.

**Static in-tube SPME time-weighted average sampling** An integrated sampling is possible with a simple SPME system. This is particularly important in field measurements when analyte concentrations change with time and location.

When the extracting phase is not exposed directly to the sample but is contained in a protective tubing (needle) without any flow of the sample through it (Figure 5), the extraction occurs through the static gas phase present in the needle. The integrated system can consist of extraction phase coating the interior of the tubing, or it can be an externally coated fiber withdrawn into the needle. These geometric arrangements represent a very powerful method able to generate a response proportional to the integral of the analyte concentration over time and space (when the needle is moved through the space). In these cases, the only mechanism of analyte transport to the extracting phase is diffusion through the gaseous phase contained in the tubing. During this process, a linear concentration profile (shown in Figure 6) is established in the tubing between the small needle opening, characterized by surface area (A) and the distance (Z) between the needle opening, and the position of the extracting phase. The amount of analyte extracted ( $dn$ ), during time interval ( $dt$ ), can be calculated by considering Fick's first law of diffusion:

$$dn = AD_g \frac{dc}{dz} dt = AD_g \frac{\Delta C(t)}{Z} dt \quad [4]$$

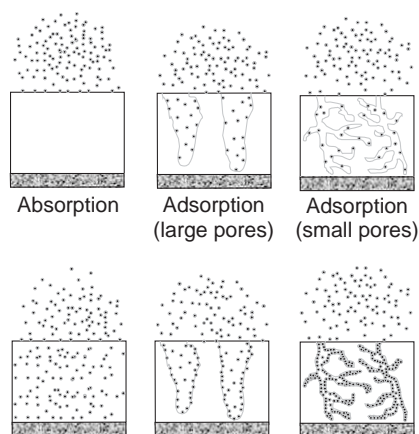


**Figure 6** Use of SPME for in-tube time-weighted average sampling where  $C$  = concentration,  $C_s$  = concentration of analyte in the sample,  $Z$  = distance between needle opening and position of extraction phase, and  $L$  = length of SPME needle: (A) schematic, (B) adaptation of commercial SPME manual extraction holder.

where  $\Delta C(t)/Z$  is a value of the gradient established in the needle between needle opening and the position of the extracting phase,  $Z$ ;  $\Delta C(t) = C(t) - C_z$ , where  $C(t)$  is a time dependent concentration of analyte in the sample in the vicinity of the needle opening, and  $C_z$  is the concentration of the analyte in the gas phase in the vicinity of the coating.  $C_z$  is close to zero for a high coating/gas distribution constant capacity, then:  $\Delta C(t) = C(t)$ . The concentration of analyte at the coating position in the needle,  $C_z$ , will increase with integration time but it will be kept low compared to the sample concentration because of the presence of the sorbing coating. Therefore, the accumulated amount over time can be calculated as:

$$n = D_g \frac{A}{Z} \int C(t) dx \quad [5]$$

As expected, the extracted amount of analyte is proportional to the integral of the sample concentration over time, the diffusion coefficient of analytes in gaseous phase ( $D_g$ ), in the area of the needle opening ( $A$ ), and inversely proportional to the distance of the coating position in respect of the needle opening ( $Z$ ). It should be emphasized that eqn [5] is valid only in a situation where the amount of analyte extracted onto the sorbent is a small fraction (below RSD of the measurement, typically 5%) of the equilibrium amount with respect to the lowest concentration in the sample. To extend integration times, the coating can be placed further into the needle (larger  $Z$ ), the opening of the needle can be reduced by placing an additional orifice (smaller  $A$ ), or a higher capacity sorbent can be used. The first two solutions will result in low measurement sensitivity. An increase of sorbent capacity presents a more attractive opportunity. It can be achieved by either increasing the



**Figure 7** Schematic representation of absorptive versus adsorptive extraction, and adsorption in small versus large pores.

volume of the coating or its affinity toward the analyte. An increase of the coating volume will require an increase of the device size. The optimum approach to increased integration time is to use sorbents characterized by large coating/gas distribution constants.

The exploitation of restricted access to the absorbing medium allows the implementation of SPME for TWA sampling. Where diffusion to the sorbent surface is limited, the sorbent can act as a sort of 'zero sink' such that extraction is very far from equilibrium under normal sampling conditions. In practice then, any analytes reaching the sorbent surface are absorbed, essentially exhaustively. However, the rate of diffusion is still dependent on the sample concentration so the total amount absorbed by the coating is proportional to the average analyte concentration over time, resulting in TWA sampling. This has been implemented to date with the conventional fiber assembly, by retracting the fiber a known distance inside the needle (Figure 6). The small size of the needle orifice limits diffusion to the sorbent surface and the ultimate diffusion rate is also a function of the distance between the fiber tip and the end of the needle. Depending on the volatility and concentration of the analyte of interest, the fiber may be positioned either closer to or further away the end of the needle to achieve the desired degree of nonequilibrium extraction and sensitivity. To date, the retractable needle implementation has gained the most attention, due to its ease of use and adjustability for the analyte and sample at hand.

**Fiber coatings** Substantial differences in performance between the liquid and solid coatings have been observed (Figure 7). In the case of liquid coatings, the analytes partition onto the extraction phase where the molecules are solvated by the coating molecules.

The high diffusion coefficient in the liquid coating allows the molecules to penetrate the whole volume of the coating within a reasonable extraction time, if the coating is thin. In the case of solid sorbents, the coating has a well-defined crystalline structure, which, if dense, substantially reduces the diffusion coefficients within the structure. Therefore, the extraction occurs only within the experimental time on the surface of the coating. There is only limited surface area on the coating available for adsorption. If this area is substantially occupied, the displacement effects occur and the equilibrium amount extracted will vary with concentrations of both the target and other analytes.

One way to overcome the fundamental limitation of the porous coatings is to use an extraction time much less than the equilibrium time so that the total amount of analytes accumulated onto the fiber is substantially below the saturation value. When performing such experiments, is it critical to precisely control extraction times and convection conditions to ensure that they are constant or can be compensated for. One way of eliminating the need for compensation of convection is to normalize agitation conditions.

The short time exposure SPME measurement described has an advantage associated with the fact that the rate of extraction is defined by diffusivity of analytes through the boundary layer of the sample matrix, and their corresponding diffusion coefficients, rather than distribution constants.

The amount of analytes accumulated on the fiber as a function of time for fiber geometry can be estimated as:

$$n = \frac{2\pi D_s L C_s t}{\ln((b + \delta)/b)} \quad [6]$$

where  $n$  is the mass of extracted analyte over sampling time ( $t$ ) in ng;  $D_s$  is the analyte molecular diffusion coefficient in sample matrix ( $\text{cm}^2 \text{s}^{-1}$ );  $b$  is the outside radius of the fiber coating (cm);  $L$  is the length of the coated rod (cm);  $\delta$  is the thickness of the boundary layer surrounding the fiber coating (cm); and  $C_s$  is analyte concentration in sample matrix ( $\text{ng ml}^{-1}$ ).

## Theoretical Aspects of Method Optimization

SPME is a multiphase equilibration process. Frequently, the extraction system is complex, as in a sample consisting of an aqueous phase with suspended solid particles having various adsorption interactions with analytes, plus a gaseous headspace. In some cases, specific factors have to be considered,

such as analyte losses by biodegradation or adsorption on the walls of the sampling vessel. In the discussion below, three phases will be considered: the fiber coating, the gas phase or headspace, and a homogeneous matrix such as pure water or air. During extraction, analytes migrate between all three phases until equilibrium is reached.

The mass of an analyte extracted by the liquid polymeric coating is related to the overall equilibrium of the analyte in the three-phase system. Since the total mass of an analyte should remain constant during the extraction, we have:

$$C_0 V_s = C_f^\infty V_f + C_h^\infty V_h + C_s^\infty V_s \quad [7]$$

where  $C_0$  is the initial concentration of the analyte in the matrix;  $C_f^\infty$ ,  $C_h^\infty$ , and  $C_s^\infty$  are the equilibrium concentrations of the analyte in the coating, the headspace, and the matrix, respectively;  $V_f$ ,  $V_h$ , and  $V_s$  are the volumes of the coating, the headspace, and the matrix, respectively. If we define the coating/gas distribution constant as  $K_{fh} = C_f^\infty / C_h^\infty$  and the gas/sample matrix distribution constant as  $K_{hs} = C_h^\infty / C_s^\infty$ , the mass of the analyte absorbed by the coating,  $n = C_f^\infty V_f$ , can be expressed as:

$$n = \frac{K_{fh} K_{hs} V_f C_0 V_s}{K_{fh} K_{hs} V_f + K_{hs} V_h + V_s} \quad [8]$$

Also:

$$K_{fs} = K_{fh} K_{hs} = K_{fg} K_{gs} \quad [9]$$

since the fiber/headspace distribution constant,  $K_{fh}$ , can be approximated by the fiber/gas distribution constant,  $K_{fg}$ , and the headspace/sample distribution constant,  $K_{hs}$ , by the gas/sample distribution constant,  $K_{gs}$ , if the effect of moisture in the gaseous headspace can be neglected. Thus, eqn [2] can be rewritten as:

$$n = \frac{K_{fs} V_f C_0 V_s}{K_{fs} V_f + K_{hs} V_h + V_s} \quad [10]$$

The equation states, as expected from the equilibrium conditions, that the amount of analyte extracted is independent of the location of the fiber in the system. It may be placed in the headspace or directly in the sample as long as the volumes of the fiber coating, headspace, and sample are kept constant. There are three terms in the denominator of eqn [9] that give measures of the analyte capacity of each of the three phases: fiber ( $K_{fs} V_f$ ), headspace ( $K_{hs} V_h$ ), and the sample itself ( $V_s$ ). Therefore, it is always important to optimize volumes of all the phases present in the system. If we assume that the vial containing the sample is completely filled (no headspace), the term  $K_{hs} V_h$  in the denominator, which is related to the capacity ( $C_h^\infty V_h$ ) of the headspace, can be eliminated, resulting in eqn [1]. Equation [1] describes the

mass absorbed by the polymeric coating after equilibrium has been reached in the system.

Strictly speaking, the above discussion is limited to partitioning equilibrium involving liquid polymeric phases such as polydimethylsiloxane. The method of analysis for solid sorbent coatings is analogous for low analyte concentration, since the total surface area available for adsorption is proportional to the coating volume if we assume constant porosity of the sorbent. For high analyte concentrations, the saturation of the surface can occur resulting in nonlinear isotherms. Similarly, high concentration of competitive interference compound can displace the target analyte from the surface of the sorbent. The simplest way to consider these high concentration effects is to replace the volume of the fiber coating,  $V_f$ , in the above equations as a measure of the total fiber surface area by a fraction of the original coating volume corresponding to a free surface area available for adsorption.

### Prediction of Distribution Constants

In many cases, the distribution constants present in the above equations determine the sensitivity of SPME extraction that can be estimated from physicochemical data and chromatographic parameters. This approach eliminates the need for calibration. For example, distribution constants between a fiber coating and gaseous matrix (e.g., air) can be estimated using isothermal GC retention times on a column with a stationary phase identical to the fiber coating material. This is possible because the partitioning process in GC is analogous to the partitioning process in SPME, and there is a well-defined relationship between the distribution constant and the retention time. The nature of the gaseous phase does not affect the distribution constant, unless the components of the gas, such as moisture, swell the polymer and change its properties.

A useful method for determining coating-to-gas distribution constants employs the linear temperature programmed retention index (LTPRI) system, which indexes compounds' retention times relative to the retention times of *n*-alkanes. This system is applicable to retention times for temperature-programmed gas-liquid chromatography. The logarithm of the coating-to-air distribution constants of *n*-alkanes can be expressed as a linear function of their LTPRI values. The LTPRI values for many compounds are available in the literature; hence, this method allows estimation of  $K_{fg}$  values without experimentation. If the LTPRI value for a compound is not available from published sources, it can be determined from gas chromatographic analysis.

Estimation of the coating/water distribution constant can be performed using eqn [9]. The appropriate coating/gas distribution constant can be found by applying techniques discussed above and the gas/water distribution constant (Henry's constant) can be obtained from physicochemical tables or can be estimated by the structural unit contribution method.

Some correlations can be used to anticipate trends in SPME coating/water distribution constants for analytes. For example, a number of investigators have reported the correlation between octanol/water distribution constant  $K_{ow}$  and  $K_{fw}$ . This is expected since  $K_{ow}$  is a very general measure of the affinity of a compound to the organic phase. It should be remembered, however, that the trends are valid only for compounds within homologous series, such as aliphatic hydrocarbons, aromatic hydrocarbons, or phenols; they should not be used to make comparisons between different classes of compounds, because of different analyte activity coefficients in the polymer.

### Effect of Extraction Parameters

Thermodynamics theory predicts the effects of modifying certain extraction conditions on partitioning and indicates parameters to control for reproducibility. The theory can be used to optimize the extraction conditions with a minimum number of experiments and to correct for variations in extraction conditions, without the need to repeat calibration tests under the new conditions.

Extraction conditions that affect  $K_{fs}$  include temperature, pH, salting, and organic solvent content in water. Extraction temperature increase causes an increase in the extraction rate, but simultaneously a decrease in the distribution constant. In general, if the extraction rate is of major concern, the highest temperature which still provides satisfactory sensitivity should be used. Adjustment of the pH of the sample can improve the sensitivity of the method for basic and acidic analytes. This is related to the fact that unless special coatings are used, SPME can only extract neutral (nonionic) species from water. By properly adjusting the pH, weak acids and bases can be converted to their neutral forms, in which they can be extracted by the SPME fiber. To make sure that at least 99% of the acidic compound is in the neutral form, the pH should be at least two units lower than the  $pK_a$  of the analyte. For the basic analytes, the pH must be larger than  $pK_b$  by two units.

Addition of salt to aqueous samples generally increases the fiber/matrix distribution constants of neutral organic molecules. However, when the analytes are in the dissociated form, a decrease in the amount extracted is observed. This is related to the

fact that the activity coefficients of ionic species in water decrease with the increase in the ionic strength of the solution. It is important, therefore, to first convert the analytes to neutral forms.

*See also:* **Air Analysis:** Sampling. **Chromatography:** Overview; Principles. **Clinical Analysis:** Sample Handling. **Drug Metabolism:** Metabolite Isolation and Identification. **Extraction:** Solid-Phase Extraction. **Food and Nutritional Analysis:** Sample Preparation. **Forensic Sciences:** Volatile Substances. **Headspace Analysis:** Purge and Trap. **Perfumes.** **Sample Handling:** Sample Preservation; Automated Sample Preparation. **Sampling:** Theory.

## Further Reading

- Bruheim I, Liu X, and Pawliszyn J (2003) Thin-film microextraction. *Analytical Chemistry* 75: 1002–1010.
- Chen Y and Pawliszyn J (2003) Time-weighted averaged passive sampling with a solid-phase microextraction device. *Analytical Chemistry* 75: 2004–2010.
- Furton KG, Wang J, Hsu YL, Walton J, and Almirall JR (2000) The use of solid-phase microextraction–gas chromatography in forensic science. *Journal of Chromatographic Science* 38: 297–306.
- Kataoka H (2002) Automated sample preparation using in-tube solid-phase microextraction and its application: A review. *Analytical and Bioanalytical Chemistry* 373: 31–45.
- Kataoka H, Lord HL, and Pawliszyn J (2000) Applications of solid-phase microextraction in food analysis. *Journal of Chromatography A* 880: 35–62.
- Lord HL, Grant RP, Walles M, *et al.* (2003) Development and evaluation of a solid-phase microextraction probe for in vivo pharmacokinetic studies. *Analytical Chemistry* 75: 5103–5115.
- Lord H and Pawliszyn J (2000) Microextraction of drugs. *Journal of Chromatography A* 902: 17–63.
- Martos PA and Pawliszyn J (1997) Calibration of solid-phase microextraction for air analyses based on physical chemical properties of the coating. *Analytical Chemistry* 69: 206–215.
- Mullett WM and Pawliszyn J (2003) The development of selective and biocompatible coating for solid-phase microextraction. *Journal of Separation Science* 26: 251–260.
- Pawliszyn J (1997) *Solid-Phase Microextraction Theory and Practice*. Toronto: VCH.
- Pawliszyn J (1999) *Applications of Solid-Phase Microextraction*. Cambridge: Royal Society of Chemistry.
- Pawliszyn J (2002) *Sampling and Sample Preparation for Field and Laboratory*. New York: Elsevier.
- Theodoridis G, Koster EHM, and deJong GJ (2000) SPME for the analysis of biomedical samples. *Journal of Chromatography B* 745: 49–82.



# F

## FAAS

See **ATOMIC ABSORPTION SPECTROMETRY: Flame**

## FATS

See **FOOD AND NUTRITIONAL ANALYSIS: Oils and Fats**

## FATTY ACIDS

See **LIPIDS: Fatty Acids**

## FERTILIZERS

**J H Falls and S A Siegel**, CF Industries Inc., Plant City, FL, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

There are three primary elements in fertilizers that require analytical determinations: nitrogen (N); available phosphate (AP) for plant uptake expressed as  $P_2O_5$ ; and water-soluble potassium expressed as  $K_2O$ . Procedures are examined to determine values based on the best methods available for the primary elements, secondary elements, micronutrients in fertilizers, and other trace elements. Uniform methods can eliminate problems associated with fertilizer analyses, resulting in agreement between laboratories on a given sample. By following the exact procedures as outlined in an acceptable method, one can expect a reliable and reproducible chemical analysis.

### Types of Fertilizer Materials

#### Dry Fertilizers

There are six main concentrated dry fertilizers produced and used in the majority of all mixed fertilizers. Diammonium phosphate (DAP), mono-ammonium phosphate (MAP), granular triple super-phosphate (GTSP), urea, ammonium nitrate, and ammonium sulfate are currently being manufactured with the following 'grades': 18-46-0, 11-52-0, 0-46-0, 46-0-0, 34-0-0, and 21-0-0, respectively. Water-soluble potassium (0-0-60) and potassium nitrate (13-0-44) are added to concentrated fertilizers to produce mixtures of 10-10-10, 8-8-8, 21-7-14, and many other grades. The formulation of a fertilizer with a 15-15-15 grade represents 15 mass% nitrogen (N), 15 mass% available  $P_2O_5$  (AP), and 15 mass%  $K_2O$ . One example of how a mixed fertilizer can be prepared is shown in **Table 1**.

When mixing dry fertilizers, the size of the fertilizer particles in each material will have a major influence on the chemical analysis owing to the



**Table 1** Example of a dry fertilizer (15-15-15)

Concentrated material	Grade	Weight (kg)	Resulting grade contribution		
			N	P <sub>2</sub> O <sub>5</sub>	K <sub>2</sub> O
Urea	46-0-0	456.5	10.5	0	0
DAP	18-46-0	500.0	4.5	11.5	0
GTSP	0-46-0	152.2	0	3.5	0
Potash	0-0-60	500.0	0	0	15.0
Filler	0-0-0	391.3	0	0	0
Total		2000.0	15.0	15.0	15.0

connection between size and uniform sampling. Chemical fertilizer manufacturers will furnish the size guide number (SGN) and the uniformity index (UI) of each material used in the fertilizer mixture. Mixing dry fertilizers with the same SGN and UI values will result in a well-blended product. When fertilizer products with nonuniform particles are mixed, the blended particles will segregate causing erratic analytical results. Sampling of a dry blend fertilizer is commonly done from a conveyor belt, a conical pile, or a prepackaged container. When obtaining a sample, it is imperative to take a representative sample from the population being analyzed. In the case of a confined population such as a 22 kg bag, the sample can be riffle cut from the total population using a riffing procedure, such as that described by the Association of Official Analytical Chemists (AOAC) or the Association of Fertilizer and Phosphate Chemists (AFPC) Methods Manual. For a large sample such as a silo or conical pile, multiple samples from the sample are taken at varying depths and locations around the perimeter of the pile or silo. A single grab sample is a poor representation of the pile due to particle segregation from sloughing due to gravity. The multiple samples are then composited and riffle cut to obtain a sample for analysis. For conveyor belt sampling, multiple cross-section samples must be taken from the belt and riffle cut after composting to obtain a representative sample. To analyze a dry blend fertilizer a riffle cut fraction of ~227 g is ground using a sample mill that will not subject the sample to undue heat or humidity during and after grinding. Samples of dry blend fertilizer containing DAP will deaminate and lose moisture under heat from grinding, resulting in low nitrogen determinations and high phosphate results. High humidity interferes with a uniform grind, resulting in sample deposited on the mill and caking of the ground sample. After grinding, the sample must be homogenized by rolling the powder on a cloth and sealed in a moisture-resistant container. For a sample to be homogeneous it should pass through a 35-mesh screen after grinding and all the sample ground must be retained for analysis.

Calculations of SGN and UI were originally developed by the Canadian Fertilizer Institute as means for describing the average particle size as well as the distribution of particle sizes in granular products and bulk-blended ingredients. SGN provides for the calculation of average size in millimeters multiplied by 100. UI is the ratio of the 'fine' particles retained at the cumulative 95th percentile to 'coarse' particles retained at the cumulative 10th percentile with the result then multiplied by 100. These two indices are particularly significant when considering bulk blended ingredients since ingredients not within acceptable SGN and UI ranges may segregate in storage potentially resulting in nonuniform application. The accepted uncertainty of  $\pm 10\%$  is generally used when considering granular formulations. The overall average and  $\pm 10\%$  uncertainty are then calculated after which each of the blend ingredients are compared to these values. Ingredients whose SGN and UI values are within the 10% uncertainty are acceptable.

### Liquid Fertilizers

Liquid fertilizers commonly used are UAN (urea and ammonium nitrate mixture) containing 28, 30, and 32 mass% N; anhydrous ammonia, which contains 82 mass% N; superphosphoric acid, which is a concentrated form of phosphoric acid with a grade of 0-70-0; and ammonia reacted with superphosphoric acid to produce a liquid fertilizer of 10-34-0 or 11-37-0 grade. Different liquid grades of fertilizer can be produced using a variety of chemicals for grades similar to the dry fertilizer grades.

## Analytical Methods for Nitrogen

### Formaldehyde Titration – Method 1 (Used for Ammonium Nitrate or Sulfate Nitrogen Only)

Method 1 determines the percentage of ammonium nitrate by reacting formaldehyde with ammonium nitrate, which produces nitric acid. The nitric acid is titrated with a standardized sodium hydroxide solution to a phenolphthalein endpoint or a pH endpoint. The determined percentage ammonium nitrate can be converted to %N by multiplying the percentage ammonium nitrate by 0.349978. The percentage ammonium sulfate can be determined by the reaction of formaldehyde, which produces sulfuric acid. The sulfuric acid can be titrated with sodium hydroxide to determine the amount of ammonium sulfate present. The %N from ammonium sulfate is determined by multiplying the percentage ammonium sulfate by 0.212. Ammonium nitrate in UAN can be determined by this method since the urea will not react. *Note:* Method 1 is of less importance now

and used less because formaldehyde can be a safety hazard if not used properly.

#### **Urease Digestion/Titration – Method 2 (Used for Urea Nitrogen Only)**

Method 2 determines the urea nitrogen by the action of the enzyme urease or Jack Bean on urea, producing ammonia and carbon dioxide. The ammonia is titrated with sulfuric acid to a neutral endpoint using methyl purple as the indicator. A pH meter can be used for a better endpoint determination because of the obscure visual endpoint when using methyl purple. To determine the correct endpoint in the titration, a known sample of urea should be titrated several times to determine the pH endpoint ( $\sim 4.2$ ). Both urea alone and urea in UAN solutions can be determined by this method. The conversion of urea into %N can be accomplished by multiplying the percentage urea by 0.4665. In the case of UAN, the total nitrogen can be found by adding the %N found in the formaldehyde and urease titration methods plus any free ammonia nitrogen present. *Note:* Jack bean meal can be substituted for urease at a much lower cost since it contains sufficient amounts of urease to accomplish the conversion.

#### **Nitrogen Combustion – Method 3 (Used for all Forms of Nitrogen)**

The nitrogen combustion process can analyze both liquid and dry fertilizer samples. In the nitrogen combustion process, all forms of nitrogen are converted to nitrogen gas by incineration. Its effective filament resistance measures the nitrogen gas concentration. Known standards are used to calibrate the instrument before use. There are several manufacturers offering instrumentation for this purpose. This method offers a rapid way to determine the total nitrogen for all forms of nitrogen fertilizers and generating infinitesimally small amounts of chemical waste as compared to the classic Kjeldahl nitrogen method. High cost of maintenance of equipment is a drawback in this method.

#### **Comprehensive Kjeldahl – Method 4 (Used for all Forms of Nitrogen)**

The comprehensive method determines the nitrogen content for all forms of nitrogen compounds (nitrite, nitrate, urea, organic nitrogen, and ammonia), which upon distillation can be collected in a known amount of sulfuric acid and titrated with sodium hydroxide. A great deal of time is consumed in performing this classical nitrogen method. *Note:* Boric acid may be used for collection of ammonia in the distillation step rather than sulfuric acid.

#### **Spectrophotometric Salicylate–Nitroprusside – Method 5 (Used for Ammonia Nitrogen)**

The salicylate–nitroprusside method determines the ammonia nitrogen ( $\text{NH}_3\text{-N}$ ) by reacting ammonia with a salicylate–nitroprusside solution to produce color intensity proportional to the concentration. Automation can be used to analyze multiple samples rapidly for ammonium products and urease conversion digestions. Often, Methods 5 and 9 in combination are analyzed spectrophotometrically for ammonia nitrogen and phosphate using an autoanalyzer.

#### **Ammonia (Nitrogen by Difference) – Method 6 (Used for Anhydrous Ammonia Only)**

Method 6 determines the moisture content of anhydrous ammonia by evaporation and the nitrogen content is determined by subtracting the water content from 100% and multiplying by 0.8224:  $\%N = (100.0\% - \% \text{moisture}) \times 0.8224$ . *Note:* \* % Moisture for anhydrous ammonia can be determined by the AFPC, The Fertilizer Institute (TFI), and AOAC procedures.

#### **Kjeldahl Digestion – Method 7 (Used for Ammonia Nitrogen)**

The Kjeldahl digestion method is used to determine only the ammonia nitrogen content. Sodium hydroxide or magnesium oxide is used to raise the pH to above 12 to liberate the ammonia upon distillation, which is collected in a known amount of sulfuric acid. The excess sulfuric acid is titrated with sodium hydroxide and the percentage ammonia nitrogen is determined. Only ammonia-based fertilizer material should be analyzed using this method.

#### **Refractive Index and Specific Gravity – Method 8 (Used for UAN Solutions)**

Using known concentrations of UAN solutions with various concentrations of water, ammonium nitrate, and urea, a triangular graph can be plotted using refractive index and specific gravity. Each manufacturer develops a curve related specifically to products produced at each location. Owing to slight variations in the manufacturing processes, the triangular graph from one producer may not work for other producers or independent laboratories.

*Note:* Other determinations could be used for nitrogen compounds in fertilizer, such as ion chromatography, spectrophotometric (dye approach), gas diffusion, potentiometric, and indophenol blue methods. These methods have not been approved as official methods by credible organizations such as AOAC International and AFPC. Potentiometric

measurements are typically not as precise and accurate as approved procedures.

*Note:* When determining any nitrogen ingredients in fertilizer, known and blank standards should be used before the final official analysis is stated. Corrections should be made for the blank on all samples and the known standards should be within 0.1% N. If a known NIST (National Institute of Standards and Technology) standard reference material is available, it should be used to verify the method and to verify other internal check samples.

## Analytical Methods for $P_2O_5$

### Indirect Available $P_2O_5$ – Method 9 (Used on all Forms of Phosphate Fertilizers)

Method 9 determines the available  $P_2O_5$  or available phosphate in all types of fertilizer samples. The total  $P_2O_5$  and the citrate-insoluble  $P_2O_5$  are initially determined, and the citrate-insoluble  $P_2O_5$  is subtracted from the total  $P_2O_5$  to give the percentage available  $P_2O_5$  or %AP. Citrate-insoluble  $P_2O_5$  (CI) is the amount not theoretically soluble in a neutral ammonium citrate solution that simulates agronomic conditions. Spectrophotometric and gravimetric methods are used for the determination in all  $P_2O_5$  methods. The indirect AP method has been used longer than any other procedure and demonstrates the best accuracy and precision. The spectrophotometric method utilizes a yellow molybdovanadate complex and the gravimetric method involves a quimociac precipitate. The spectrophotometric method can be completed faster than the gravimetric method using an autoanalyzer system.

### Direct Available $P_2O_5$ – Method 10 (Used on all Forms of Phosphate Fertilizers)

Direct AP methods involve the determination of citrate-soluble  $P_2O_5$  by 'direct' AP analysis after the water-soluble portion is removed. The failure to remove the water washing step before addition to the neutral ammonium citrate will result in oversaturation of the ammonium citrate, resulting in low results. Generally, direct methods are less precise and results have a higher standard deviation. Problems encountered with direct methods are technique, interference of ammonium citrate in the spectrophotometric method, and not using proper acid digestion methods.

### Inductive Coupled Plasma – Method 11 (Used for all Types of Phosphate Fertilizers)

Inductive coupled plasma optical emission spectrometry (ICP-OES) involves elemental excitation through electromagnetic collisional energy. Concentration

measurements are accomplished by conventional spectroscopic instrumentation. Laboratories have used ICP-OES in the past, and the results have been erratic, with lower precision and accuracy (results that are 1–2% lower) than older established methods. As in any method development improvements have been made using the ICP-OES by members of the AFPC.

Determination of phosphorus by using a simultaneous ICP-OES has been achieved as an assay method for rock and high concentrate phosphate products with accuracy and precision comparable to existing spectrophotometric and gravimetric methods. This has been accomplished by matrix matching the standards with the sample, using an internal standard, and buffering the interference effects of calcium and magnesium by introducing lithium nitrate as a buffer. An additional tool that has reduced the interference effects of easily ionizable elements has been robust plasma conditions. These are characterized as low center tube flow of  $0.61 \text{ min}^{-1}$  and a high radiofrequency of 1400 kW. Choice of nebulizer, analyte concentration, and sample introduction protocol varies with each instrument and requires the chemist and the applications specialist to optimize based on signal-to-noise ratios and relative standard deviations of the replicates. Instrument manufacturers are an invaluable source of information when optimizing a new method. The best candidates for internal standard are scandium and beryllium. Yttrium is found in phosphate rock in varying quantities, introducing variability into the results. Further information on the ICP-OES method as described can be obtained from the members of AFPC.

## Analytical Methods for Water-Soluble Potassium

### Water-Soluble Potassium Determinations – Method 12 (Used for Water-Soluble Potassium Fertilizers)

Water-soluble potassium (expressed as  $K_2O$ ) can be determined by manual or automated flame photometric and tetraphenylboron titration methods for all types of fertilizer samples. Atomic absorption can be used for samples below 5%  $K_2O$  to maintain the highest degree of precision and accuracy. (*Note:* Tetraphenylboron method for potassium is being used less often because of safety issues associated with formaldehyde.) Methods using ICP-OES for high concentration of potassium are currently being developed. *Note:* Potentiometric measurements can be used to determine potassium in fertilizers but are not official approved methods and do not meet the accuracy and precision requirements as current methods.

## Slow Release Fertilizers

### Slow Release Fertilizers – Method 13 (Used for all Types of Fertilizers)

The commercial development of slow release fertilizer materials has been incremental and based on the use of several unique technologies over the last ~50 years. While each technology has found a niche in certain specialty markets they have not found widespread use in broad-based agricultural markets. Each technology was addressed, as it was developed, in terms of the regulation and analysis of the specific material. This prompted Association of American Plant Food Control Officials (AAPFCO) to form a taskforce addressing this problem. They identified a need to develop an effective method to evaluate a broad range of materials instead of a number of product-specific methods. Equipped with this, regulators can begin to monitor these materials efficiently in a regulatory laboratory setting. Key components in the method's evolutionary process were developing an accelerated laboratory procedure and a stable laboratory soil leaching system mimicking real-life biological conditions. Having accomplished this, the correlation of the output data from the two new methods can give analysts a statistically reliable method to analyze, characterize, and regulate slow release fertilizer materials.

The laboratory method can be summarized as follows: A representative unground fertilizer sample is exposed to increasingly aggressive solvent extractions. The procedure is designed to rapidly accelerate the release of materials in a matter of hours instead of the normal days or weeks normally experienced. The first extraction is with distilled water at 25°C, subsequent extractions are carried out with 0.2% citric acid at 60°C for increasing periods of time. Each extraction is designed to extract and isolate nutrients that become available over 74 h of extraction time. Extracts are preserved and then analyzed by traditional AOAC procedures for the nutrient of interest. Along with analysis of total nutrients in the sample, data are used to develop information specific to the cumulative percentage of nutrient released over the entire extraction period.

A final report was published by AAPFCO slow release taskforce in August 2004. (The report can be obtained from AAPFCO.) The report summarizes both lab and soil methods as well as correlates the data. Following this presentation the method will be submitted to AOAC for review and publication.

## Trace Elements in Fertilizers

Trace elements for the fertilizer industry are elements occurring at less than 1000 ppm in samples analyzed.

Several classes of elements are routinely analyzed in fertilizers. Secondary nutrients and micronutrients in fertilizers are two classes of trace elements beneficial to plants. Common micronutrients analyzed in fertilizers are Zn, Mn, B, Fe, Mo, and Cu, and secondary nutrients include Ca, Mg, and S.

Trace elements such as As, Cd, Co, Pb, Hg, Mo, Ni, Se, and Zn found in fertilizers causing possible health problems in the transfer of these trace elements into the food chain are under investigation. A scientific risk-based study was performed independently by the US Environmental Protection Agency (EPA), State of California, and The Fertilizer Institute to determine safe levels of trace elements in fertilizer materials. All three studies provided similar results, concluding that the vast majority of fertilizers produced currently in the USA are considered safe when properly applied. Where issues were found, these were exclusively within a small group of micronutrient products that EPA has chosen to regulate under the Recycled Zinc Micronutrient Rule. AAPFCO has developed a model bill SU1P 25 for determining safe levels of trace elements in fertilizers based on the scientific risk-based studies. Several US States, Canada, and other countries have adopted limits on trace elements in fertilizers with some variation in limits. For information on SU1P 25 contact AAPFCO (<http://www.aapfco.org>). For information on metal limits worldwide go to website <http://afpc.net> for the latest data.

### Inductive Coupled Plasma (ICP-OES) – Method 14 (Trace Metals in all Fertilizers)

Matrix matching is the most commonly used tool for overcoming the complex interactions when analyzing secondary, micronutrients, and trace elements under regulation. The procedural tools described in Method 11 are useful in achieving accurate results when analyzing trace elements. In addition to those procedures, using multiple lines for an element is a necessary precaution to prevent a weak line from iron or other elements with multiple spectral lines from influencing the result of a trace element through spectral interferences. When analyzing near the limits of quantification, use of known addition to the samples will confirm a signal response and recovery for a specific element to increase precision.

For the past several years a metals forum has been established to review four different areas associated with trace metals. The four areas include: (1) digestion and reference materials, (2) detection, instrumentation, and data handling, (3) soil analysis, metal availability, and plant uptake, and (4) regulation, enforcement, risk management, and education.



The event is sponsored by AAPFCO, FLDACS, AFPC, TFI, and IMC. For further and current information contact Bill Hall of IMC Global (e-mail: wlhall@imcglobal.com).

## Methods for Different Types of Fertilizers

The previously discussed analytical methods are grouped according to application to specific fertilizer types for nitrogen, phosphate, and potassium. The preferred methods for analyzing N, P, and K fertilizers are listed in Table 2. The specific claims or guarantees on fertilizer products such as ammonia nitrogen, nitrate nitrogen, and urea nitrogen will require different nitrogen procedures beyond a total nitrogen procedure, which may be the method of choice for most analytical determinations.

### Production Application for Nitrogen and Available Phosphate

In the manufacturing plant the grades of MAP, DAP, and GTSP are controlled by analyzing the nitrogen and total phosphate content every 2 h and the citrate insoluble (CI)  $P_2O_5$  every 8 h. Expected CI in these

products is as follows:

	Expected range
MAP	0.0–0.2%
DAP	0.0–0.2%
GTSP	1.0–3.5%

*Note:* GTSP continues to react for several weeks after being produced, reducing the CI content to 0.5–1.5%, or less.

For over 100 years the empirical method for determining AP has been used to make sure that no manufacturer was misrepresenting the products produced. Many years ago, some unscrupulous manufacturers were adding phosphate rock to normal superphosphate (0-20-0) to obtain the final product. In the manufacture of DAP, MAP, and GTSP today it would be impossible to add another product to these high-grade materials and still achieve the grades 18-46-0, 11-52-0, and 0-46-0. The use of phosphate rock, which contains 30–35%  $P_2O_5$ , would dilute both the nitrogen and phosphate grades.

The empirical method for the present concentrated fertilizers would benefit from a new concept for determining the relationship of plant agronomics with regard to phosphate. Agronomists state that only 5–15% of a year's available phosphate application will be available in the first growing season. The remaining phosphate will react with the soil and form insoluble compounds and will be available in future years. The empirical method for AP today restricts these materials from being sold as AP. This so-called insoluble phosphate behaves similarly to the reaction of soil complexation. Over time the CI breaks down and becomes available to plants. A simple solution for high-grade materials would be to sell it on the basis of total phosphate. DAP and MAP were not produced until 1959 and of the previous work on available phosphate none was used in testing of these products.

**Table 2** Recommended methods for specific types

<i>Nitrogen products – percentage nitrogen</i>	
Ammonium nitrate	Methods 1, 3, 4
Ammonium sulfate	Methods 1, 3, 4
Ammonia	Method 6
UAN	Methods 1, 2, 3, 4, 7, 8
MAP	Methods 3, 4, 5, 7
Urea	Methods 2, 3, 4
Fertilizer mixtures	Methods 3, 4
<i>Phosphate products – percentage available and total <math>P_2O_5</math></i>	
MAP	Methods 9, 11
DAP	Methods 9, 11
GTSP	Methods 9, 11
Mixed fertilizers	Methods 9, 11
Superphosphoric acid	Methods 9, 11
10-34-0 or 11-37-0	Methods 9, 11
<i>Water-soluble potash products – percentage <math>K_2O</math></i>	
Potash, KCl	Method 12
Mixed fertilizers	Method 12
<i>All fertilizers – slow release properties</i>	
All types of fertilizers	Method 13
<i>All fertilizers – trace elements, ppm</i>	
All types of fertilizers	Method 14

## Quality Assurance

In precisely determining the N- $P_2O_5$ - $K_2O$  concentrations in fertilizer samples, a well-designed quality assurance program is essential. The program should contain the following as a minimum: (1) internal standards; (2) duplicate analysis of samples; (3) standard additions; (4) NIST standards; and (5) participation in one or more check sample programs.

Internal standards are prepared from reagent-grade mixtures of potassium dihydrogen phosphate and ammonium sulfate 'primary standards'. Both primary standards are cross-checked with NIST #194 ammonium phosphate to verify the N and  $P_2O_5$  grade and NIST #193 potassium nitrate to verify the

N and K<sub>2</sub>O grades. Internal and 'working' standards are prepared for both N and P<sub>2</sub>O<sub>5</sub> for simultaneous determination using an autoanalyzer. Internal, NIST, and purchased standards should be randomly placed between samples during a group of analytical determinations. Duplicate samples should check within  $\pm 0.2\%$  P<sub>2</sub>O<sub>5</sub> and  $\pm 0.1\%$  N or should be rechecked before the official analysis is given. This is a guideline that must be followed for a high degree of accuracy and precision to be maintained. A laboratory must maintain a good check sample program. There are several excellent check programs dealing with fertilizer materials, and these are detailed below.

### Association of Fertilizer and Phosphate Chemists Check Sample Program

The Association of Fertilizer and Phosphate Chemists administers two check sample (or proficiency) programs that are utilized by manufacturers, customers, and regulatory agencies throughout the world. The AFPC Rock Check Program, formerly known as the IMC Rock Check Program, distributes a 75-g sample of phosphate rock ground to -35-mesh/+200-mesh for determination of major, minor, and trace elements by participating laboratories. The samples are sent monthly at no cost to the participating laboratory. AFPC members from throughout the United States submit rock check samples for distribution. The results submitted by the participating labs are collectively analyzed and the median and standard deviation calculated for each parameter reported. The standard deviation is calculated using the interquartile range, to mitigate the effect of outliers and new participants on the statistical results. The check sample program for high concentrate fertilizers distributes samples of MAP, DAP, GTSP, and Feed Grade Material on a monthly basis and are analyzed for major, minor, and trace elements. Samples are prepared by AFPC members and are distributed at no cost to participants. Results are submitted by participants, and treated statistically as previously described. Information and results on the check sample programs is available on the AFPC website. Both check programs are updated on the web page for the AFPC (<http://afpc.net>) monthly. The AFPC provides archived check samples of rock and fertilizer products for purchase. Further information about the check sample programs can be obtained from the web site or by contacting the AFPC, P.O. Box 1645, Bartow, Florida 33831, USA.

### Magruder Fertilizer Check Sample Program

The Magruder check sample program consists of mixed fertilizers and a few concentrated fertilizers. Example grades such as 13-13-13, 6-24-24, 0-0-50,

6-6-8, 10-10-10, etc., are used in the program. Each month's sample(s) contain the primary plant nutrients N, P, and K; and, about all samples each year contain secondary, micronutrients, and trace elements under regulation. Samples are analyzed by methods chosen by the individual laboratories. A comprehensive statistical summary is prepared, noting interlaboratory bias, precision, and ranking of coded individual laboratories. Subscribers receive this summary to evaluate their performance.

There are ~140 laboratories that subscribe to the Magruder program. All of the US state control laboratories participate, and the remaining number is about equally divided between US industrial and foreign laboratories, the annual subscription fee is \$204 (US) plus postage for foreign countries. For information contact Magruder Check Sample Program (<http://www.magruderchecksample.org/>).

### Internal Goal

Each laboratory using the check sample programs should set an internal goal. Many production and nonproduction laboratories set a goal of falling within one pooled standard deviation on each analysis. A well-established quality assurance program can aid in increasing analytical accuracy and precision to the desired levels. All three check sample programs are highly recommended.

*See also:* **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Geochemistry:** Soil, Major Inorganic Components. **Nitrogen. Quality Assurance:** Primary Standards; Internal Standards. **Sampling:** Theory.

### Further Reading

- AOAC (2002) *Official Method of Analysis of AOAC International*, 17th edn., current through first revision.
- Association of American Plant Food Control Officials Laboratory Services Committee (2004) *Guidelines for Preparing Fertilizer Laboratory and Analytical Samples*. Visit website <http://www.aapfco.org> to obtain a copy.
- Falls JH (1991) Comparison and review of three available P<sub>2</sub>O<sub>5</sub> methods for diammonium phosphate and triple superphosphate. *Fertilizer Research* 28: 239–249.
- Falls JH, Barnacascel WR, and Britt PW (1976) Direct potentiometric titration of urea with urease. *Journal of Association of Official Analytical Chemists* 59(5): 1045–1047.
- Falls H, Siegel S, Longest J, and Boggs B (2001) Comparative methods for metals in fertilizers and 'some' pitfalls to avoid. Paper presented to the *Phosphate Conference*, 19 October, Lakeland, Florida. Visit website <http://afpc.net> to obtain a free copy.
- Hall WL Jr. and Robarge WP (eds.) *Environmental Impact of Fertilizer on Soil and Water*. Oxford: Oxford



- University Press. ISBN 0-8412-3811-1. Developed from a symposium at the 222nd National Meeting of the American Chemical Society, Chicago, Illinois, August 26–30.
- Martin D, Falls H, Hayward P, and Vessey J (2002) Practical analytical methods for quality control in fertiliser plants. Paper presented to *The International Fertiliser Society*, Brussels, 18 April. Order from <http://www.fertiliser-society.org>
- Methods Used and Adopted by the Association of Fertilizer and Phosphate Chemists* (formerly known as the Association of Florida Phosphate Chemist, AFPC), 8th edn., 2001.
- Mortvedt JJ, Murphy LS, and Follett RH (1999) Fertilizer Technology and Application. Order from [www.meisterpro.com](http://www.meisterpro.com). Library of Congress Catalog Card Number; 99-63148.
- Official Publication of AAPFCO, No. 57, 2004. Order from <http://www.aapfco.org>
- Prepared by AAPFCO's BMP Committee and The Fertilizer Institute's Product Quality and Technology Professionals (2002) *Fluid Fertilizer Quality Control Manual*, 1st edn. Order from <http://www.aapfco.org>
- Siegel S, Boggs B, and Falls H (2003) Utilization of ICP-AES for the analysis of phosphate fertilizers. Paper presented to the *Phosphate Conference*, October 16, Lakeland, Florida, and Federation of Analytical Chemistry & Spectroscopy Societies in Ft. Lauderdale, Florida, October 19–23. Visit website <http://afpc.net> to obtain a free copy.
- Soil Fertility Handbook*. Ministry of Agriculture, Food and Rural Affairs of The Fertilizer Institute of Ontario Inc., 1998. ISBN 0-7778-7730-9. Order at 1-888-466-2372 or [www.gov.on.ca/](http://www.gov.on.ca/)
- The Fertilizer Institute Product Quality Committee (1982) *Fertilizer Sampling and Analytical Methods*, 4th edn. Order from <http://www.tfi.org>
- The Fertilizer Institute Product Quality Committee (1996) *Bulk Blend Quality Control Manual*, 1st edn. Order from <http://www.tfi.org>
- Trenkel ME (1997) *Improving Fertiliser Use Efficiency – Controlled Release & Stabilized Fertilisers in Agriculture*. Paris: IFA.
- US Environmental Protection Agency (1999) *Estimating Risk From Contaminants Contained in Agricultural Fertilizers*, August.
- Weinberg Group for The Fertilizer Institute (2001) *Scientific Basis for Risk-Based Acceptable Concentrations of Metals in Fertilizers and their Applicability as Standards*, July 9. Can be found on website: <http://www.aapfco.org>

### Website Links

- The Association of Fertilizer and Phosphate Chemists Check Sample Programs – <http://afpc.net>
- Comparative Methods for Metals in Fertilizers and ‘Some’ Pitfalls to Avoid – <http://afpc.net/Methods%20Committee.html>
- Scientific Basis for Risk-Based Standards – <http://www.aapfco.org/RBCWhite.pdf>
- Estimating Risk From Contaminants Contained in Agricultural Fertilizers – <http://www.epa.gov/epaoswer/hazwaste/recycle/fertiliz/risk/index.htm>
- The Fertilizer Institute – <http://www.tfi.org>
- Fertilizers Are Safe – <http://www.tfi.org/Issues/metalsinfertilizers.asp>
- Florida Institute of Phosphate Research – <http://fipr.state.fl.us>
- Florida Phosphate Council – <http://www.flaphos.org>
- The International Fertilizer Society – <http://www.fertiliser-society.org>
- National Institute of Standards and Technology – <http://www.ts.nist.gov>
- Association of Analytical Chemist International – <http://www.aoac.org>
- Association of American Plant Food Control Officials – <http://www.aapfco.org>
- Magruder Check Sample – <http://www.magrudercheck-sample.org/>

## FFF

See **FIELD-FLOW FRACTIONATION**

## FIA

See **FLOW INJECTION ANALYSIS: Principles; Instrumentation; Detection Techniques; Environmental and Agricultural Applications; Clinical and Pharmaceutical Applications; Industrial Applications**

## FIBERS

See **FORENSIC SCIENCES: Fibers. TEXTILES: Natural; Synthetic**

# FIELD-FLOW FRACTIONATION

R Beckett, Monash University, Melbourne, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Separations play a central role in the chemical and biological sciences. They are used to prepare products of the required purity, to eliminate unwanted interfering components in analysis, to identify unknown compounds, and in some cases for the measurement of physical properties.

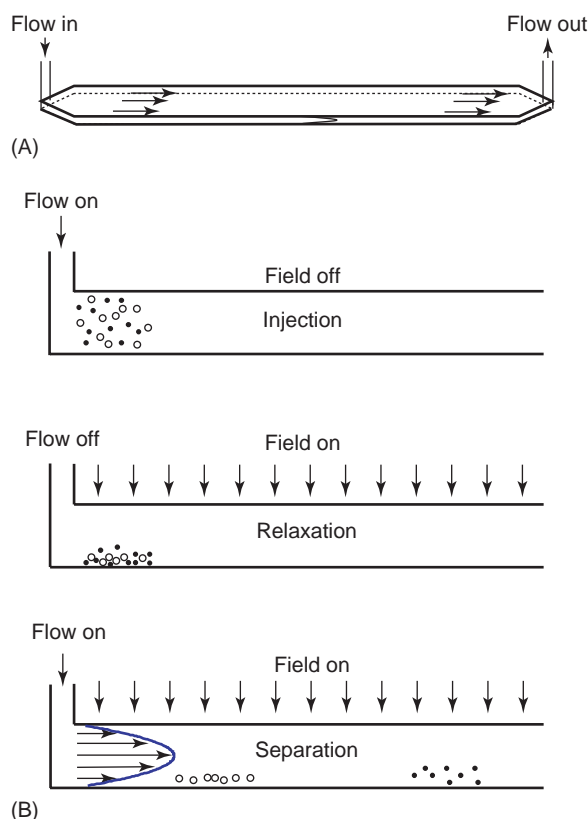
All separation methodologies are based on the selective transport of components in a mixture. This is either due to migration at different velocities in the medium caused by the application of a field or gradient, as in most settling and zone electrophoresis methods, or differences in the distribution of components either across or in the vicinity of a phase boundary, as in liquid extraction or most forms of chromatography.

Field-flow fractionation (FFF) refers to a suite of separation methods that fall into the latter class of separations mentioned above. Basically the solution or suspension is injected into a thin channel, and a field is applied to drive the sample to one wall, where different components take up different distributions across the fluid layer adjacent to the channel wall (**Figure 1**). This partial separation is then magnified by the flow of carrier down the channel, and zones are separated in a manner analogous to chromatography. The applied field may consist of a gravitational, centrifugal, hydrodynamic, or acoustic force, a temperature gradient, or an electrical or magnetic field.

FFF channels are typically 20–100 cm long and 1–4 cm wide but very thin (0.05–0.5 mm). They use the laminar flow of the carrier to effect separation. The construction of a simple FFF channel is depicted in **Figure 2**.

The concept was first developed by the late J Calvin Giddings from 1965 onward, although a similar system was independently reported by Berg and Percell in 1967. The methodology has steadily grown in popularity in recent years, and a range of instrumentation is now commercially available (e.g., Postnova Analytics, Wyatt Technology).

FFF is capable of separating macromolecules and small particles over a very large size range. Molecules from 1000 kDa in molecular weight and particles from ~1 nm up to 50 µm or so can be processed. Although mainly an analytical method for detecting

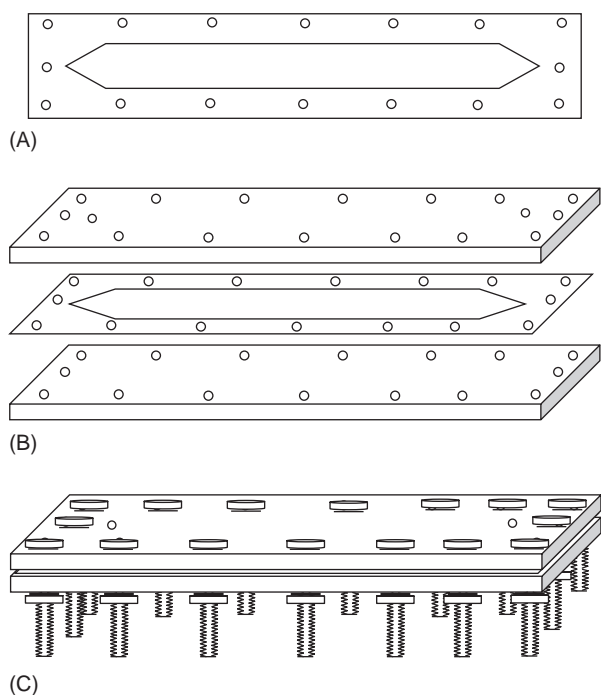


**Figure 1** (A) Schematic diagram of an FFF channel. (B) Side view of an FFF channel showing the steps involved in separation.

components in a mixture and measuring properties such as size, diffusion coefficients, and mass, it can also be used to collect small fractions for analysis. FFF is ideally suited for connecting to online detectors such as ultraviolet (UV)–visible, fluorescence, and light scattering detectors and inductively coupled plasma (ICP)–atomic emission spectrometers (AESs) and ICP–mass spectrometers (MSs), which can be used to generate distributions of physical and chemical characteristics of heterogeneous mixtures.

## The FFF Concept

FFF is a very flexible set of techniques that can be applied to various sample types with a large range of particle or molecular sizes. This is due to the fact that many different force fields and gradients can be employed and several distinct separation mechanisms are possible. In most cases the field can be varied, often during the course of a particular run (cf. gradient elution chromatography), which extends the size range that can be analyzed. The current size range covered is



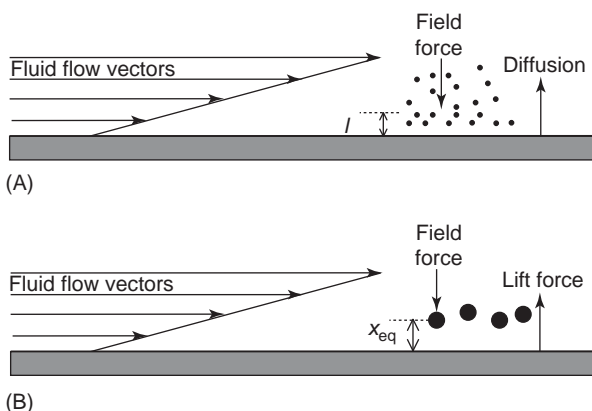
**Figure 2** Construction of a simple FFF channel. (A) Plastic sheet spacer; (B) spacer and blocks for forming channel walls; (C) complete FFF channel.

from  $\sim 1$  nm to  $50\text{ }\mu\text{m}$  and for macromolecules from 1000 Da to more than 20 MDa. Brief descriptions of the major modes of operation (separation mechanisms) and fields (FFF subtechniques) that have been used are given in the following sections. Equations are not given in this review, although areas where robust calculations can be made are indicated. Complete listings of FFF equations can be found in the sources given in Further Reading.

### Normal or Brownian Mode FFF

When a sample is injected at one end of the channel the field drives the particles or molecules to the accumulation wall of the channel (Figure 1B). For small particles (say  $<1\text{ }\mu\text{m}$ ), where the Brownian diffusion is comparable with the field-induced flux, diffuse clouds form for each component in the mixture (Figure 3A). The mean thickness of each cloud depends on the balance of the field force and diffusion, which generally depends on the particle size and other particle properties. These clouds are typically  $1\text{--}10\text{ }\mu\text{m}$  thick, compared with the channel thickness of  $50\text{--}500\text{ }\mu\text{m}$ . This relaxation step is usually carried out with the channel flow turned off.

When the flow is resumed, each component cloud is moved down the channel at a different rate, depending on the cloud thickness. This occurs due to the parabolic flow profile in the channel, where the fluid velocity will be zero at the wall and will increase



**Figure 3** FFF separation mechanisms for (A) Brownian (normal) mode and (B) steric/hyperlayer or focusing mode.

with distance from the wall as shown in Figure 3A. Particles with greater interaction with the force field or gradient and/or lower diffusivity (smaller size) will tend to have clouds compressed closer to the accumulation wall and be retained in the channel longer. This was the first FFF mechanism investigated and is most commonly referred to as the normal mode or NIFFF.

At slow flow rates the sample particles have time to diffuse across the cloud many times, ensuring that each will migrate with the mean velocity. At higher flow rates this does not have time to occur and non-equilibrium band broadening results.

One big advantage of normal mode FFF is that certain particle properties such as the buoyant mass, diameter, density, concentration and thermal diffusion coefficients, and electrophoretic mobility can be calculated directly from the elution time or volume of the sample peaks. For example, with flow FFF (see below) the diffusion coefficient and hence equivalent spherical hydrodynamic diameter can be estimated. Other fields enable different properties to be obtained.

### Steric, Hyperlayer, or Focusing Mode FFF

For particles larger than  $\sim 1\text{ }\mu\text{m}$ , Brownian diffusion becomes negligible compared with the field-induced flux. In this case the particle components form layers very close to the wall at positions where the field force is just balanced by a hydrodynamic lift force, which tends to push the particles away from the accumulation wall (Figure 3B). Each sample component is positioned in its individual hyperlayer and is swept down the channel by the corresponding flow lamina, resulting in separation. Particles that have larger sizes and as a consequence experience greater lift forces will be found in hyperlayers that are more

elevated from the accumulation wall and will thus elute earlier than smaller particles. This separation mechanism has variously been called the steric/hyperlayer, lift-hyperlayer, or focusing mode, but for simplicity it is often just referred to as steric FFF.

Due to the complexities of hydrodynamic lift forces, the particle diameter and other physical properties cannot be calculated directly from the measured retention. For this to be done a calibration must be performed using standard particles. The standards must have known characteristics and often need to be of the same material as the samples being analyzed. Separation is based on a number of physical properties such as the diameter, density, and shape.

Sample components can be focused into hyperlayers by forces other than the hydrodynamic lift. Examples include counteracting gravitational forces with a field operating in the opposite direction, as discussed below in the sections on dielectrophoretic FFF and acoustic FFF. The term 'focusing FFF' or FoFFF is sometimes used for this separation mechanism.

### Other FFF Separation Modes

A few other separation mechanisms that can be classified as FFF have been subjected to preliminary investigation. Alternating gravitational and electrical fields have been employed in cyclical FFF (CyFFF). A change in the relative direction of the gravitational field was achieved by periodically flipping the channel 180°. In CyFFF the particles move away from the accumulation wall and then back again: thus different components will move across the parabolic channel flow profile at different speeds and this will affect their migration rate down the channel, resulting in separation.

A chromatographic-FFF hybrid mode has been demonstrated for separating cells. Here the attractive interaction of cells to the accumulation wall, which is normally destructive for FFF operation, is used to separate B and T lymphocytes. Although the cells have very similar sizes and densities, selectivity was achieved because the attachment force was different and this could be balanced by the lift and drag forces of the carrier fluid.

The disadvantage of the small scale of FFF separations can be overcome in two-dimensional continuous devices. Here one FFF subtechnique is employed along the flow axis of the channel and another force vector generates a component of motion at right angles to this (cf. free flow or curtain electrophoresis). The result is that different components move along and across the planar accumulation wall, tracing out different trajectories.

This enables the sample mixture to be continuously fed into the channel, and components can be collected at different outlet ports.

One such device uses a combination of gravitational steric/hyperlayer FFF along the channel axis, and since the channel is mounted at an angle, and not flat on the bench, gravity also causes movement of particles sideways across the channel. A series of outlet ports along the lower edge of the channel continuously removes the separated components.

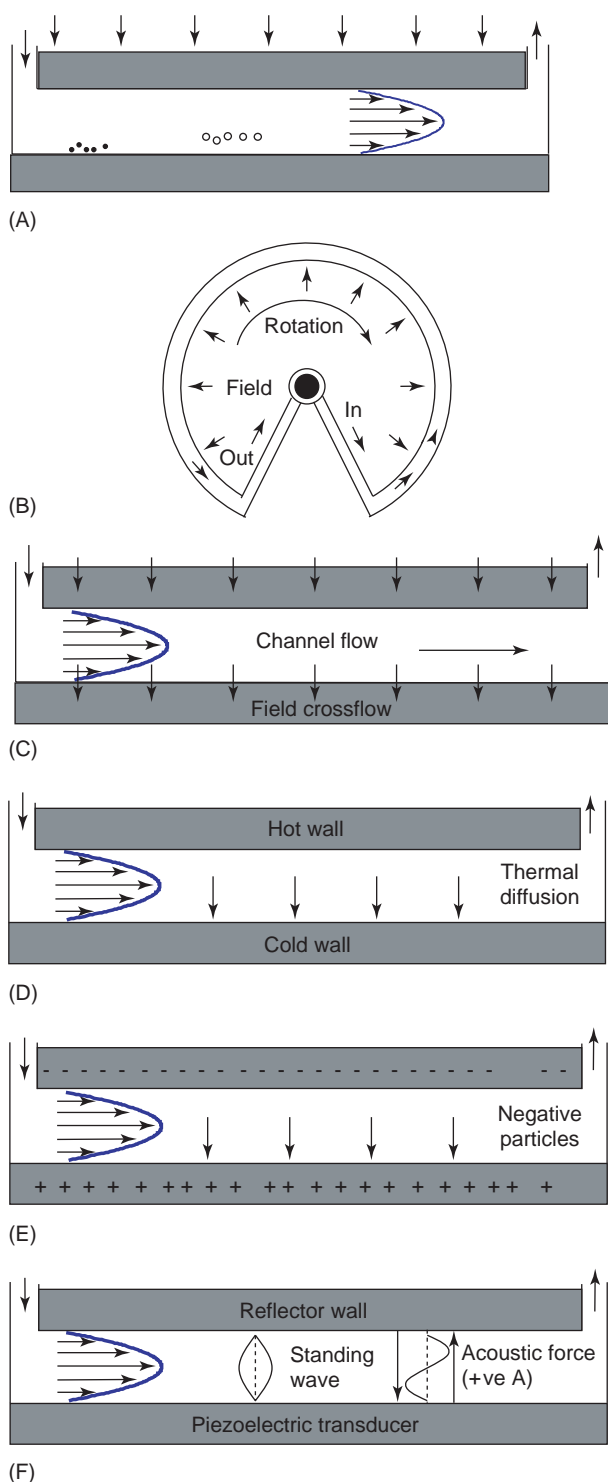
Another recent application uses thermal FFF in the radial direction of a circular channel and fluid shear generated by rotating one wall relative to the other. Polymer mixtures were continuously fed into the channel near the centre, and different molecular weight fractions were removed from the channel at points around the outer circular edge.

### Subtechniques of FFF

Many different fields and gradients have been used to perform FFF experiments, each giving rise to a particular FFF subtechnique. Since each field or gradient will interact with the particles or molecules differently, the resulting separation will depend on different sample characteristics. Consequently, certain subtechniques may be more suited to certain types of samples. In addition, the separation capabilities of normal mode FFF can often be used to measure particle or molecular parameters such as the size, mass, density, and diffusion coefficient. The exact property that can be measured will depend on the field being used. Some common subtechniques of FFF are listed below.

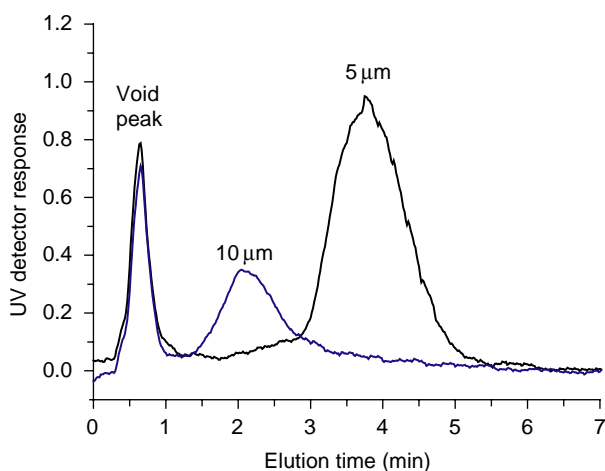
**Gravitational FFF (GrFFF)** GrFFF utilizes the earth's gravitational field, which is only strong enough for particles greater than  $\sim 1\mu\text{m}$  in diameter (**Figure 4A**). Although GrFFF channels are the simplest and cheapest to construct, they are limited to steric/hyperlayer runs of micrometer-sized particles. Examples of GrFFF runs of 5 and  $10\mu\text{m}$  silica particles are given in **Figure 5**. The void peak contains unretained material, and the larger particles elute before the smaller ones.

**Sedimentation FFF** In these instruments the channel is inserted inside a centrifuge basket and the centrifugal acceleration generates the force field (**Figure 4B**). The form of the force exerted on a particle is similar to that produced by gravity, but the magnitude can be much greater, and it can be varied easily by changing the rotational speed. With a centrifuge capable of  $\sim 2500\text{ rpm}$ , particles down to perhaps  $50\text{ nm}$  in diameter can be separated. The resolution is



**Figure 4** Schematic diagrams of the channels used in various FFF subtechniques. (A) Gravitational FFF; (B) Sedimentation FFF; (C) Symmetrical FIFFF; (D) Thermal FFF; (E) electrical FFF; and (F) acoustic FFF.

also the highest for any FFF subtechnique, with the diameter-based selectivity being 3. This means that increasing the diameter by only 10% should increase



**Figure 5** GrFFF fractograms of 5 and 10  $\mu\text{m}$  silica particles.

the elution time by 33%, often resulting in good separation of such peaks.

Both Brownian and steric/hyperlayer FFF are possible, enabling particulate samples between  $\sim 0.05$  and  $50 \mu\text{m}$  to be processed. Normal sedimentation FFF (SdFFF) elution depends directly on the buoyant mass of the particles and hence on a combination of the equivalent spherical diameter (ESD) and density.

**Flow FFF** If the channel is constructed with porous walls (e.g., ceramic or metal frits covered with an appropriate membrane), a second flow stream can be pumped across the channel in order to drive the sample toward the accumulation wall (**Figure 4C**). On reaching the wall, back diffusion in the case of the Brownian mode or hydrodynamic lift forces for larger particles cause the sample components to take up their equilibrium positions. The parabolic fluid flow then causes selective migration along the channel. Molecules as small as 1 kDa and particles 1 nm in diameter can be separated in the Brownian mode. For small particles or macromolecules, elution depends entirely on the diffusion coefficient of the particles: thus the hydrodynamic diameter can be determined, independent of density.

The majority of studies have used a symmetrical flow FFF (SyFIFFF) channel where the cross flow is pumped through a frit at the top of the channel and is removed through the membrane and frit at the accumulation wall as shown in **Figure 4C**. The channel and field flows are usually driven by two independent pumps. In recent times a variant termed asymmetric FIFFF (AsFIFFF) has gained popularity. Here the top wall is an impervious glass or plastic plate and there is a single channel inlet flow. Some of the carrier fluid is taken out through the lower membrane and frit that produces the hydrodynamic 'field', causing the



sample to migrate to the accumulation wall. The remaining flow exits the end of the channel and carries the separated sample components to the detector.

Although the equations governing AsFIFFF are much more complicated than for SyFIFFF, asymmetric channels have the distinct advantage that the operator can see through the top wall and using injected dyes can test the flow pattern in the channel. In addition only one inlet pump is needed and injected samples can be focused into a narrow band during the relaxation period by pumping carrier from both ends of the channel, which reduces band broadening. However, the results from both SyFIFFF and AsFIFFF are comparable.

**Thermal FFF** Macromolecules and particles migrate along a temperature gradient, from hot regions to cold, through a poorly understood phenomenon called thermal diffusion (Figure 4D). If a channel is made of metal blocks, one which is heated and the other cooled, large temperature gradients of  $\sim 10^5$ – $10^6$  °C m<sup>-1</sup> can be generated across the thin dimension of the channel. The cold block becomes the accumulation wall, and components are separated on the basis of the ratio of the concentration diffusion coefficient ( $D$ ) to the thermal diffusion coefficient ( $D_T$ ). Thus if  $D$  is known, thermal FFF (ThFFF) can be used to estimate  $D_T$  values for polymer–solvent combinations. ThFFF has been used extensively for polymer characterization in organic solvents. Since  $D_T$  is independent of molecular weight ( $M$ ) but  $D$  is related to the size and hence  $M$ , ThFFF elution profiles can yield molecular weight distributions, provided a suitable calibration is performed to take into account the variations in molecular conformation that affect the conversion of  $D$  to  $M$ . This calibration procedure is similar to that required for size exclusion chromatography (SEC), and the two techniques are rather complimentary. ThFFF is very effective at high  $M$ , being less prone to artifacts such as stationary phase interactions (adsorption and pore rejection) and shear degradation. Above  $\sim 100$  kDa, ThFFF has a higher resolution, which tends to increase with  $M$ , but separation becomes more difficult below  $\sim 20$  kDa. In recent years ThFFF has been extended to include particles and aqueous phase separations.

**Electrical FFF** A fairly obvious extension is to use an electrical field to separate charged particles and macromolecules on the basis of their electrophoretic mobility and size (Figure 4E). In contrast to the very widely used electrophoresis techniques, the field is used to cause a partial separation of sample components near the accumulation wall, which is magnified

as they are moved along the channel by the fluid flow. Typically quite low voltages (1–2 V) are required since the spacing between the channel wall electrodes of  $\sim 0.1$  mm ensures high voltage gradients are produced. Practical problems such as production of gaseous electrolysis products in the channel and poorly understood double layer effects near the accumulation wall have reduced the separation efficiency. It is likely that some technological advances will appear soon, making this a valuable addition to the FFF family, particularly for protein and cell characterization.

**Magnetic FFF** Magnetic properties of some particles can generate forces in a magnetic field. This principle has been demonstrated with ferromagnetic materials used in tape recording but is of course rather limited. Another approach is to selectively adsorb magnetic nanoparticles to the component of interest. For example, antibody-coated magnetic beads could be used to label important mammalian cell types.

**Dielectric FFF** Dielectrophoretic forces result from the interaction of field-induced polarized particles and a nonuniform AC electrical field. The magnitude of the force depends on the frequency of the applied field and the polarizability of the particles, which in turn is affected by the chemical composition. Gascoyne and coworkers in Texas have performed a separation of human leukocytes in focusing mode FFF, where the gravitational force on the cells toward the accumulation wall was balanced by the upward dielectrophoretic force. Thus the separation illustrated the combined use of two different fields (DI–Gr/FoFFF). This is surely an exciting new development in FFF.

**Acoustic FFF** If an acoustic standing wave is generated across the thin dimension of an FFF channel, particles in the carrier will experience a force, either toward the walls or toward the center of the channel, depending on the sign of the acoustic contrast factor ( $A$ ) (Figure 4F). A ceramic transducer, driven by a megahertz AC voltage, produces the ultrasound wave, and the frequency is adjusted to produce a standing wave between the transducer and reflector plates. The wavelength is chosen to be twice the distance between these plates. For particles with negative  $A$  values (e.g., O/W emulsions, polyethylene) the transducer and reflector become the channel walls and fractionation occurs along each as the force is away from the center line. For particles with positive ( $A$ ) values (e.g., polystyrene, silica, steel) the gap between the transducer and reflector can be at least half-filled with an acoustically transparent gel,



the surface of which becomes the accumulation wall. The acoustic FFF/FoFFF separation principle has been demonstrated using micrometer-sized polystyrene latex, yeast cells, and an O/W emulsion. The acoustic field has also been combined with gravitational forces (Ac-Gr/FoFFF), analogous to DI-Gr/FoFFF, described above. The magnitude of the acoustic force depends on the particle size, density, and compressibility, which again suggests the possibility of separations based on the material properties as well as physical size. New applications for these techniques are being sought in medicine, environmental science, and nanotechnology.

### Field Programming in FFF Runs

With samples that have high polydispersity, it is often advantageous to employ field programming to ensure reasonable run times are achieved. The approach is similar to temperature programming in gas-liquid chromatography and solvent programming in high-performance liquid chromatography. Field programming is sometimes essential for normal mode SdFFF due to the very high-size selectivity ( $S_d = 3$ ), which can result in the larger particles being spread along the elution axis to the extent that they become too dilute to detect. Fortunately, programming is quite simple to achieve in SdFFF as the field is controlled by the centrifuge speed. It should be noted that field programming does not benefit separations of narrowly dispersed samples.

Many different forms of the field program have been employed. They often consist of an initial period of a constant high field followed by a decay of the field to speed up elution of the larger particles. Decay functions have included linear, parabolic, and exponential expressions. One popular decay function, introduced by Williams and Giddings, is referred to as the power program. It has the advantage of producing separations with a constant fractionating power, which means the fractional increase in diameter is proportional to the fractional increase in elution volume.

Field programming is also used in the separation of macromolecules using FIFFF and ThFFF, although since the selectivity is much less than for SdFFF it is not usually essential unless the sample molecular weight (MW) polydispersity is well over an order of magnitude.

### Hyphenated FFF Techniques

The separation capabilities of FFF can be capitalized upon by making physical and chemical measurements

on the fractionated sample. This may be achieved by obtaining discrete subsamples using a fraction collector and performing offline analysis using appropriate instruments. However, a more efficient approach is to connect the channel outlet to one of the various types of online detectors that are available. A few examples of such hyphenated FFF techniques are given below.

**FFF-UV** Traditionally UV absorbance detectors are used to monitor the mass of particles or macromolecules in the eluent. Errors will arise if the size dependence of the attenuation of the light is neglected. For particles that scatter the light beam, this requires correcting the signal using Mie theory, but this is usually not attempted due to the complexity of the computation or lack of information such as the refractive index. For macromolecules where light absorbance occurs, it is assumed that the chromophores are uniformly distributed across the MW distribution. These data are usually plotted in the form of fractograms (analogous to chromatograms), which are plots of detector response versus elution time or volume. The fractograms obtained from UV detectors can be converted to particle size or MW distributions either using the FFF equations or when necessary calibration standards to convert the elution time axis to diameter or MW.

**FFF-multiangle light scattering (MALS)** MALS detectors enable the MW of eluted macromolecules to be measured online. When combined in series with a mass detector such as a UV or refractive index detector, the MW distribution is generated directly without any calibration with standards. Furthermore, MALS also yields the radius of gyration of the molecules and the variation of this size parameter with MW can be used to deduce molecular conformation information.

**FFF-ICP-MS** The ICP-MS is a multielement analysis tool ideally suited for direct coupling with FFF. The ICP torch is capable of vaporizing and ionizing particles in the eluent up to  $\sim 10\ \mu\text{m}$ , and the plasma is then fed into an MS for simultaneous detection of many elements. Quadripole, mass-sector, and time of flight MSs are now available, depending on the sensitivity, mass resolution, and response time required. FFF-ICP-MS yields element-based size distributions. Other element detection systems that have been used include ICP-AESs, electrothermal atomic absorption spectrometers, and very recently laser-induced breakdown spectrometers.

## Applications of FFF

### Particle Separation

The range of particles that have been analyzed using FFF is huge, ranging from minerals to polymer latexes, emulsions, metals, carbon black, carbon nanotubes and onions, pigments, bacteria, viruses, casein micelles, starch granules, and yeast and mammalian cells, to name a few.

Perhaps the biggest issue that must be confronted in particle separation is the normal–steric transition that occurs at  $\sim 1\ \mu\text{m}$ . The problem arises because the elution order of particles in the normal and steric modes is opposite (increasing size with increasing elution time for normal mode, but the opposite trend in the steric mode). Thus for broad populations spanning the transition diameter it is possible for fractions eluted at a specific time (or volume) to contain two populations, submicrometer particles that have eluted in the normal mode and micrometer particles eluted in the steric mode.

It is not possible to specify the exact diameter where the normal–steric transition occurs as a loss of selectivity and resolution will be observed in the approximate range  $0.5\text{--}2\ \mu\text{m}$  and this will be affected by the field strength. The simplest general advice that can be given is that if the sample spans  $1\ \mu\text{m}$ , then it should be split into two fractions ( $<1\ \mu\text{m}$  and  $>1\ \mu\text{m}$ ) before attempting FFF separation. Available methods for achieving this prefractionation include settling, centrifugation, filtration, split-flow thin-channel fractionation, and elutriation.

For the above reason, we will consider applications of FFF for submicrometer and micrometer particles separately.

**Submicrometer particles** Particles below  $\sim 1\ \mu\text{m}$  in diameter can be separated using normal mode FFF using a variety of subtechniques, the most widely used being sedimentation and flow. FIFFF can be used down to  $1\ \text{nm}$  and yields the diffusion coefficient and hence the equivalent spherical hydrodynamic diameter distribution. SdFFF instruments usually operate up to  $\sim 2500\ \text{rpm}$ , which can resolve silica particles down to  $\sim 50\ \text{nm}$ . The quantity directly measured is the effective or buoyant mass, and particle size distributions can be generated if the density of the particles is known.

Shape is not a huge concern in these measurements, although for accurate work it should be considered. Shape will influence the conversion of the diffusion coefficient ( $D$ ) to the hydrodynamic diameter in FIFFF, although the discrepancy in  $D$  between spheres and ellipsoids of the same volume will seldom exceed 50%. For SdFFF the buoyant mass

determined will not be influenced by shape unless the particles are very highly retained so that the cloud thickness is less than the largest particle dimension.

An interesting extension of SdFFF measurements is the determination of the particle density. Since the elution time enables the buoyant mass to be estimated from the FFF equations, then if the size of the particles eluted at a given time is measured independently, for example, using electron microscopy or dynamic light scattering, then the density can be calculated. Alternatively if the elution times of a sample are determined in a series of carriers that have different densities, the particle density can be extracted.

FFF methods are certainly developing into important techniques for colloidal particles. They compare well with other particle sizing techniques for resolution and accuracy, and usually calibration standards are not necessary. The fact that separation is employed helps in obtaining high quality particle size data for polydisperse samples and enables fractions to be examined using other physical and analytical techniques. A drawback is the fact that the separation is truly microscale, with typically  $10\text{--}100\ \mu\text{g}$  being processed in a single run.

**Micrometer particles** As outlined above, micrometer size particles are separated in the steric/hyperlayer mode of FFF, and calibration using suitable standards is required (examples of typical fractograms are given in Figure 5). Shape will be a crucial factor as the lift force is highly shape dependent and can make a difference of two to three times in the elution time of particles with the same volume and mass. For SdFFF and GrFFF, the separation density will also be important, and standards with the same density as the samples must be used or careful adjustment of the field (rpm) applied for the standards, and the sample can be used to compensate for the difference in density. Fortunately, if FIFFF channels are operated with a vertical orientation, so the cross flow is horizontal and the channel flow is vertical, the particle density has no effect on the retention.

Calibration lines are often plotted as  $\log t_r$  versus  $\log d$ , where  $t_r$  is the elution time and  $d$  is the particle diameter, and a straight line is commonly obtained. This enables the  $t_r$  axis of the sample fractogram to be converted to a size axis. Further manipulation of the detector response will be required to generate the particle size distribution as outlined in detail in Further Reading.

There is probably less incentive to resort to FFF analysis for micrometer-size particles than there is for colloids as other adequate methods are available in this range. Examples include single-particle

counters (electrozone, optical blocking, flow cytometers), Fraunhofer diffraction, settling, and elutriation. For FFF the biggest source of concern is the effect of shape on the elution time, which introduces uncertainty in the size determinations. However, separations have high resolution, can cope with broad distributions, and provide small fractions for subsequent analysis. In particular, GrFFF is cheap, can easily be constructed from readily available components already used in liquid chromatography and flow injection analysis, and is simple to operate.

### Macromolecule Characterization

**Thermal FFF** Polymers dissolved in nonaqueous solvents have usually been analyzed using ThFFF. The MW range has varied from  $\sim 5000$  Da up to possibly 100 million. ThFFF is a direct competitor to SEC as a method of MW determination, although they are best seen as complimentary techniques. SEC has better resolution at low MW values, but ThFFF excels at the high MW end. The crossover in terms of resolution probably occurs  $\sim 50$ – $100$  kDa, although this will depend on a number of factors.

ThFFF fractograms can be converted to MW distributions using standards of the same polymer. Universal calibration curves analogous to those used in SEC are possible only for polymers with known thermal diffusion coefficients ( $D_T$ ).

The dependence of retention on  $D_T$  has been used to develop a method for determining this constant, which is difficult to do using other methods. This has resulted in an expansion in the number of polymer types that now have tabulated  $D_T$  values that may contribute to the development of theories for the thermal diffusion process.

It is now known that  $D_T$  does not depend on the MW but is related to the chemical nature of the polymer and solvent. This can be used to advantage to separate different polymers with the same diffusion coefficient and size, which cannot be achieved using SEC. It was found that for random copolymers  $D_T$  is linearly related to the proportion of the amount of the two monomers present, providing a way of measuring the composition of such copolymers. Further, for block copolymers,  $D_T$  apparently depends on the polymer block occupying the outer layer of the copolymer molecules. This could perhaps provide a method of studying the conformation of block copolymers in solution.

Two very important advantages of ThFFF lies in the low shear experienced in the unpacked channel and the lower chance of wall interactions compared with the problems experienced in SEC with pore rejection and stationary phase adsorption. These

effects are of particular significance for ultrahigh MW samples, where shear degradation and low sample recovery are often encountered in SEC.

Another application where ThFFF is very useful is in the characterization of polymers that contain both dissolved molecules and gel particles. Such samples cannot be injected into SEC columns but can be analyzed using ThFFF, where the channel is open and will not be clogged.

ThFFF is most often used for polymers dissolved in organic solvents as samples dissolved in water often have low  $D_T$  values and hence poor retention. However, there are some studies on particle separations and aqueous systems, and this may expand the range of ThFFF applications in the future.

**Flow FFF** Water-soluble macromolecules are usually characterized using FIFFF. The diffusion coefficient and hydrodynamic diameter are determined directly, but these can only be converted to MW using calibration standards of the same type as the sample to ensure the molecular conformation is similar.

FIFFF has been found to be very useful for separation of proteins. In addition to MW and size, molecular conformation and aggregation can be investigated.

Most difficulties encountered with FIFFF are associated with the correct choice and installation of the membrane. Adsorption or repulsion of the sample can cause delayed or early elution, resulting in errors in size and MW, poor resolution, or low sample recovery. No doubt these problems will be reduced as a better understanding of membrane-sample interactions is obtained.

**See also:** **Liquid Chromatography:** Overview; **Mass Spectrometry:** Overview; Mass Separation.

### Further Reading

- Beckett R and Hart BT (1993p) Use of field-flow fractionation techniques to characterize aquatic particles, colloids and macromolecules. In: Buffle J and van Leeuwen HP (eds.) *Environmental Particles*, vol. 2, pp. 165–205. Boca Raton, FL: Lewis.
- Caldwell KD (1988) Field-flow fractionation. *Analytical Chemistry* 60: 959A–971A.
- Giddings JC (1973) The conceptual basis of field-flow fractionation. *Journal of Chemical Education* 50: 667–669.
- Giddings JC (1981) Field flow fractionation. A versatile method for the characterization of macromolecular and particulate materials. *Analytical Chemistry* 53: 1170A–1178A.
- Giddings JC (1988) Field-flow fractionation. *Chemical and Engineering News* 66: 34–45.

Giddings JC (1993) Field-flow fractionation: analysis of macromolecular, colloidal, and particulate materials. *Science* 260: 1456–1465.

Schimpf ME, Caldwell K, and Giddings JC (eds.) (2000) *Field-Flow Fractionation Handbook*. New York: Wiley-Interscience.

## FILTRATION

See **MEMBRANE TECHNIQUES: Ultrafiltration**

## FINGERPRINTS

See **FORENSIC SCIENCES: Fingerprint Techniques**

## FIRE ASSAY

**G L Everett**, Johnson Matthey Plc,  
Royston, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1282–1288, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

### Introduction

Fire assay describes a group of separation methods in which precious metals (PMs) are separated from other species by dissolution in other molten metals, usually lead, nickel, or tin. Fire assay has been and continues to be fundamental to the determination of PMs. These are defined as gold (Au), silver (Ag), and the platinum group metals (PGMs) platinum (Pt), palladium (Pd), rhodium (Rh), iridium (Ir), ruthenium (Ru), and osmium (Os). In the context of fire assaying other elements are referred to as base elements or metals.

The Further Reading section lists some of the excellent texts available that cover the practical operation of these methods. Bugbee's book, although written 70 years ago, is still relevant to the classical lead, scorification, and cupellation assays. The more recent book by Lenahan and Murray-Smith provides comprehensive cover of practice in the South African mining industry and includes very good discussions on laboratory layout and practice in the calculation of charges for the fire assay. The books by Beamish and Beamish and Van Loon give a very good overview of fire assay in general PM analysis and in

the former there is an excellent section on the calculation of the composition of the assay charge **Figure 1**.

The text by Van Loon and Barefoot contains a chapter on fire assay in the context of instrumental analysis. In this there is comprehensive information on the charges used for a wide range of minerals and ores and several detailed procedures combining fire assay with an instrumental finish are described. Even though modern instrumentation is fairly tolerant of the many interferences found in PM-containing samples, fire assay is still essential at the start of many analytical procedures. Moreover, whether the finish is gravimetric or instrumental, fire assay can offer greater accuracy and precision in sample preparation than many other procedures. This is because it can be used as a separation or preconcentration technique, or both, and it also allows the use of large samples (trials).



**Figure 1** Fire assaying being performed in a laboratory. (Courtesy of Johnson Matthey Plc.)



Giddings JC (1993) Field-flow fractionation: analysis of macromolecular, colloidal, and particulate materials. *Science* 260: 1456–1465.

Schimpf ME, Caldwell K, and Giddings JC (eds.) (2000) *Field-Flow Fractionation Handbook*. New York: Wiley-Interscience.

## FILTRATION

See **MEMBRANE TECHNIQUES: Ultrafiltration**

## FINGERPRINTS

See **FORENSIC SCIENCES: Fingerprint Techniques**

## FIRE ASSAY

**G L Everett**, Johnson Matthey Plc,  
Royston, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1282–1288, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

### Introduction

Fire assay describes a group of separation methods in which precious metals (PMs) are separated from other species by dissolution in other molten metals, usually lead, nickel, or tin. Fire assay has been and continues to be fundamental to the determination of PMs. These are defined as gold (Au), silver (Ag), and the platinum group metals (PGMs) platinum (Pt), palladium (Pd), rhodium (Rh), iridium (Ir), ruthenium (Ru), and osmium (Os). In the context of fire assaying other elements are referred to as base elements or metals.

The Further Reading section lists some of the excellent texts available that cover the practical operation of these methods. Bugbee's book, although written 70 years ago, is still relevant to the classical lead, scorification, and cupellation assays. The more recent book by Lenahan and Murray-Smith provides comprehensive cover of practice in the South African mining industry and includes very good discussions on laboratory layout and practice in the calculation of charges for the fire assay. The books by Beamish and Beamish and Van Loon give a very good overview of fire assay in general PM analysis and in

the former there is an excellent section on the calculation of the composition of the assay charge **Figure 1**.

The text by Van Loon and Barefoot contains a chapter on fire assay in the context of instrumental analysis. In this there is comprehensive information on the charges used for a wide range of minerals and ores and several detailed procedures combining fire assay with an instrumental finish are described. Even though modern instrumentation is fairly tolerant of the many interferences found in PM-containing samples, fire assay is still essential at the start of many analytical procedures. Moreover, whether the finish is gravimetric or instrumental, fire assay can offer greater accuracy and precision in sample preparation than many other procedures. This is because it can be used as a separation or preconcentration technique, or both, and it also allows the use of large samples (trials).



**Figure 1** Fire assaying being performed in a laboratory. (Courtesy of Johnson Matthey Plc.)

## Types of Fire Assay

There are three major categories within the overall description of fire assay.

1. Procedures variously termed crucible fusion, pot fusion, or pot running.
2. Scorification.
3. Cupellation.

One or more of these procedures may be used in any individual analytical scheme, depending on the type of sample being assayed and the final elements to be determined and their concentration.

### Pot Fusion

This procedure involves the melting of the sample with a variety of powdered reagents to give a metallic (collector) button (e.g., lead) containing the elements of interest or the metalliferous portion, separated from a slag that contains the unwanted portion of the sample, known as the gangue.

Of the reagents added there are first those that produce the collector. This is usually an oxide with the attendant reducing agent to produce the metallic collector. Second, there are those fluxing reagents (the charge) that help break down the sample and carry away the gangue into the slag portion of the melt. If these components are intimately mixed, as the temperature rises a series of reactions takes place in the molten and semimolten state. First the collector is reduced to tiny globules of metal that are distributed throughout the charge. As the temperature rises further the charge begins to react, the slag begins to form, and the collector begins to function. As the temperature reaches its maximum the reaction is completed and the collector falls to the bottom of the pot with a fluid molten slag on top. Thus, after ensuring a successful fusion the charge can be separated by pouring the molten mixture into an iron mold and allowing it to separate and solidify prior to a physical separation. The separation process is by gravity, the heavier metalliferous portion containing the PMs settling to the bottom of the mould.

The chemistry of the fusion process is complex and in many cases not fully characterized owing to the variety of samples assayed. However, good approximations of the reactions can be made and these help in the estimation of the content of charges as described later. Nevertheless, pot fusion is essentially an empirical process, which accounts for laboratories' different assaying conditions for similar materials.

In the past the assaying conditions would have been determined by a series of trial assays based on

one's knowledge of the sample. This still occurs but many laboratories now have an inductively coupled plasma-atomic emission spectrometer (ICP-AES), whereby a fast and fairly accurate assessment of the composition of the gangue can be made by analyzing solutions prepared from fusion of the material in sodium peroxide.

Once the gangue composition is known, by reference to the molten state chemistry involved one can establish the appropriate amount of reagent required to obtain a successful assay.

Several factors affect the success of a pot fusion and they will now be discussed in turn.

**The sample under test** This is the primary factor in the assay as it is the nature of the sample and the information required that will govern all decisions that will subsequently be made.

Samples will be acidic, basic, or amphoteric, and it is this breakdown that will determine the fluxing reagents added to give a good slag. Along with the elements requiring determination the gangue content will affect the collector chosen. There is also the question of any pretreatment necessary. The components of a sample can be broken down as shown in **Table 1**.

Other factors affecting the assay that are imparted by the sample are its reducing or oxidizing power. As pot fusion is a reducing process, the reducing or oxidizing power of a sample will affect the amount of metallic collector formed in the reaction. This will in turn affect the quantities of reagent added as the final collector button size has to be controlled. For instance, samples containing sulfides, arsenides, antimonides, and carbonaceous matter will be reducing and will reduce litharge (PbO) to lead in the reaction mixture. Samples containing iron(III) oxide and manganese(IV) oxide will oxidize a proportion of the lead (or other collector) or will react with the reductants added, thus decreasing their effectiveness. Generally, siliceous, oxide, and carbonate ores remain neutral with respect to oxidation and reduction.

**Table 1** Sample components

#### *Acidic components*

SiO<sub>2</sub>, B<sub>2</sub>O<sub>3</sub>, WO<sub>3</sub>, As<sub>2</sub>O<sub>5</sub>, Sb<sub>2</sub>O<sub>5</sub>, Bi<sub>2</sub>O<sub>5</sub>, V<sub>2</sub>O<sub>5</sub>

#### *Amphoteric components*

Al<sub>2</sub>O<sub>3</sub>, TiO<sub>2</sub> can take on amphoteric or acidic properties depending on the conditions

Bi<sub>2</sub>O<sub>3</sub>, As<sub>2</sub>O<sub>3</sub>, SnO<sub>2</sub>

ZnO can take on amphoteric or basic properties depending on the conditions

#### *Basic components*

FeO, Fe<sub>2</sub>O<sub>3</sub>, Cr<sub>2</sub>O<sub>3</sub>, CoO, CaO, MgO, MnO, PbO, CuO, Na<sub>2</sub>O, K<sub>2</sub>O



The reducing components of a sample can often be removed by roasting in air prior to the fusion reaction. This will remove sulfur dioxide and carbon dioxide and may partly volatilize arsenic, antimony, and tellurium or otherwise turn them from the reducing chalcogenides to relatively neutral oxides in terms of oxidizing properties.

**Vanning** It was mentioned above that ICP-AES can be used to give an estimate of the composition of the sample so that suitable charges of reagents can be prepared. A time-honored, alternative method of estimating mineral content is by the procedure of vanning. A vanning pan has a spherical surface and could be a watch glass or a purpose-made, white enamel, vanning shovel. About 1 g of sample is placed in the center of the pan and is wetted with water. By manipulation of the pan the water is made to flow through the sample in one direction and effectively separates the minerals by gravity. Clay particles will be suspended easily and can be washed out whilst heavy minerals such as galena will resist any movement with the water. This movement allows the minerals to stratify in a fan-like shape because of their differing specific gravities. One can then apply simple tests to identify the major components of the sample. For instance, the white minerals calcite and quartz differ in their hardness, whereas carbonates and some sulfides will react with acid.

**The collector** The choice of collector is made with a view of not only the PMs of interest but also of the subsequent analytical measurements. The four major collectors in use are lead, nickel sulfide, copper, and tin.

**Lead** This is by far the most common collector in use worldwide. It can be used as red lead,  $Pb_3O_4$ , or more commonly nowadays as litharge,  $PbO$ . Litharge is widely available and can be obtained in a purer state than red lead.

The litharge is intimately mixed with the assay charge. When reduced to lead the fine particles of metal fall through the charge 'like rain' collecting the PMs at the bottom of the melt.

Whilst lead will collect all PMs it is best for determining gold, silver, platinum, and palladium. Although rhodium, ruthenium, iridium, and osmium can be collected, the chemical separation after the fire assay is rather long and tedious.

With lead collection sulfur can prove a problem if not properly dealt with. It can be absorbed into the lead button and form a matte layer, or regulus, into which some PMs will collect, and this will give subsequent problems in the analysis. Lead oxide can be used as a basic flux and if added in excess can

react with sulfur to prevent the formation of a regulus. However, for high sulfur levels alternative steps must be taken to remove sulfur, such as roasting in air prior to the fusion.

**Nickel sulfide** This matte is preferred as a collector when the insoluble (in aqua regia) platinum metals are being sought, i.e., rhodium, ruthenium, iridium, and osmium. It is also a very good collector for platinum and palladium, but it is less effective than lead for the collection of gold and great care must be taken with the fluxing agents to maximize recovery.

Nickel is often associated with the PMs in nature, a good example being the ore body at Sudbury, Ontario, so one must be very careful to obtain a supply of nickel that contains a very low and known content of PMs. The usual chemicals used are nickel powder, nickel carbonate, or nickel oxide with flowers of sulfur, but for the determination of very low PM contents 'carbonyl' nickel (prepared by thermal decomposition of  $Ni(CO)_4$ ) or ceramic grade  $NiO$  are preferred.

When the fusion takes place the nickel species and the sulfur combine to form a matte button of nickel sulfide.

Many of the platinum group minerals exist as sulfides, such as cooperite ( $Pt(AsS)_2$ ) and bragite ( $Pt \cdot Pd \cdot NiS$ ), which undoubtedly enhances their collection.

A practical issue with nickel sulfide buttons is that although appearing metallic to the eye they are very brittle and must be carefully separated from the slag. Tapping gently with an old-fashioned toffee hammer provides excellent separation.

**Tin** This is used as a collector after addition as  $SnO_2$  in an analogous manner to  $PbO$ . It is not widely used but it does effectively collect all PMs. However, there may be problems with the subsequent analysis for all but platinum, palladium, and gold, which may explain its lack of popularity.

**Copper** This is used on its own and also in conjunction with iron and nickel. It can be used for all PMs but it is not as effective in general terms as the other collectors mentioned. However, copper does give good results for gold, silver, platinum, and palladium and has the advantage of being easily melted into a form suitable for an X-ray fluorescence finish with addition of extra copper to standardize the weight of the button. For details of the  $CuFeNi$  collection the reader is referred to the book by Beamish.

**Use of co-collectors** Sometimes it is useful to add small quantities of PM to help in the collection or

subsequent analytical process. In this respect silver can be added in the lead collection of gold and platinum to help with the collection of other PGMs. If this procedure is used one must pay attention to the purity of the co-collector and run concurrent blanks, otherwise results, especially at very low levels, will have a high bias.

**Reducing and oxidizing agents** The term ‘reducing power’ as used in fire assay refers essentially to the removal of oxygen. For a sample or reagent it is that amount of collector that will be formed as a button when an excess is fused with 1 g of sample or reagent. Similarly, the ‘oxidizing power’ is that amount of collector or the equivalent of reducing agent that will be oxidized by 1 g of sample or oxidizing agent. With respect to lead the reactions that occur with some common reducing agents are as follows.

With charcoal as reductant there are two reactions, at low temperature, as in reaction [I],



and at high temperature, as in reaction [II],

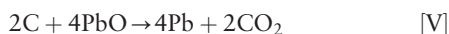


Thus, 1 g of charcoal should provide a button of 34.5 g of lead if excess litharge is used.

With wheat flour as reductant the reaction provides 15.3 g of lead from 1 g of flour (reaction [III]):



The reaction with cream of tartar is more complex and provides 5.5 g of lead per gram of reagent (reactions [IV]–[VI]):



In practice, the air above a fusion mixture will oxidize an amount of the carbon produced and thus the actual power of a reducing agent is somewhat less than theoretical; but reasonable estimates can be given as in Table 2.

Besides these reductants the components of a sample can reduce the collecting agent that could lead to a larger button than planned. For lead the most important of these are probably sulfur, iron, and sulfides.

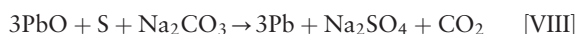
The reducing power of sulfur and sulfides depends on the other constituents of the melt. For instance, the simple reaction is as in reaction [VII]:



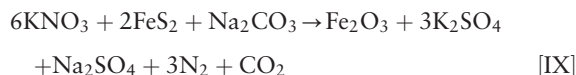
**Table 2** Reducing and oxidizing power

<i>Reducing agents</i>	<i>Oxidizing agents</i>	<i>Sample components</i>
Flour 10–12	$\text{KNO}_3$ 4.2	<i>Reducing</i>
Charcoal 18–30		$\text{FeS}_2$ 9–13
Tartar (KH tartrate) 5–6		$\text{FeS}$ 9
		$\text{Fe}$ 4
		$\text{PbS}$ 3.4
		<i>Oxidizing</i>
		$\text{Fe}_2\text{O}_3$ 1.3
		$\text{MnO}_2$ 2.4
		$\text{Fe}_3\text{O}_4$ 0.9

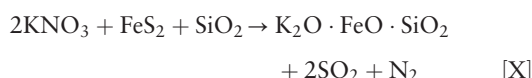
But in the presence of alkaline carbonates the reaction is modified as shown in reaction [VIII]:



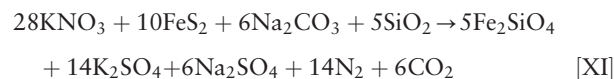
Hence, the reducing power of sulfur can be from 12.9 to 19.4 (grams of lead) depending on the conditions of fusion. Potassium nitrate, or nitre, is the most commonly used reagent deliberately added to an assay charge, usually to control the effect of sulfides. In these cases one must consider the other components of the charge or else very undesirable reactions can occur. For instance, if no silica is present the nitre will react with  $\text{FeS}_2$  to produce iron[III] oxide, an undesirable component to have in a fusion mixture (reaction [IX]):



If alkali metal carbonate is absent the sulfur is partially oxidized via reaction [X]:

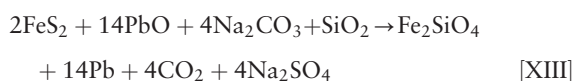


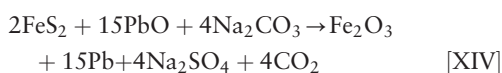
but if the components are balanced the iron may be reduced and removed completely into the slag along with the sulfur via reaction [XI]:



A practical point to note when using nitre is that it decomposes to produce oxygen. Hence, if used in excess it can produce frothing in the pot, sometimes violently with associated losses, and this must be avoided.

Samples containing species such as  $\text{FeS}_2$  will affect the amount of collector produced. Depending on the conditions lead can be produced by reactions [XII]–[XIV]:





The relevant fact is the reducing power for these reactions. Approximate figures for reducing and oxidizing power are given in **Table 2**.

The point of these calculations concerns the production of a successful fusion and correctly sized button. The usual mass of lead button aimed for is 30–60 g. If the button is too small there is the possibility of incomplete recovery of PMs; if it is too large the problem may be in the subsequent analytical work.

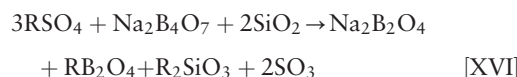
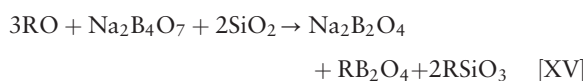
**Fluxing agents and slag formation** The qualities sought of a good assay slag are as follows:

1. It should have a relatively low formation temperature.
2. It should be fairly thick on formation to hold up particles of sample and collector thus allowing total breakdown of the sample and efficient separation and collection of the PMs.
3. It should be thin and fluid when heated above its melting point to allow adequate mixing, good settling of the collector, and an even pour.
4. It should not attack the crucible walls to any degree.
5. It should have a very low capacity for the PMs.
6. Its relative density should be low to give good separation of slag and button.
7. It should physically separate readily from the button after cooling.
8. It should contain all unwanted material from the sample and be clear, indicating complete decomposition.

Various agents may be used to give a good slag. The common fluxing agents are given in **Table 3** and again they can be divided according to acidity.

The fluxing agents and the sample components react to form a series of silicates and borates. Various compounds can be formed with differing acidities but the preferred slag is the balanced bisilicate and metaborate as shown in **Table 4**. The acidity increases going down each series but the typical reactions to consider, for a metal oxide, RO, and a sulfur-containing sample, are as in reactions [XV]

and [XVI]:



Excess lead oxide will form  $\text{PbSiO}_3$  with silica.

**Borax (sodium borate)** It is important to use the dehydrated form for, if water of crystallization is present, there is a danger of the charge being pushed out of the pot. Molten borax is an excellent solvent for nearly all the metal oxides. When it first melts at  $\sim 750^\circ\text{C}$  it is quite viscous but as the temperature increases it becomes a free-flowing liquid that is strongly acidic. In this phase it lowers the melting point of the slag and is particularly effective in dissolving inert materials like bone ash and clay. At  $\sim 500^\circ\text{C}$  it decomposes to  $\text{Na}_2\text{O}$  and  $\text{B}_2\text{O}_3$  with the boron oxide readily forming borates with basic oxides. The different borates are classified above but here one must consider the ratio of basic to acidic oxygens; the best ratio is 1:3 as in the metaborate.

If too large an excess of borax is added to the charge then the fluidity of the melt is diminished with subsequent retention of lead or other collector, and thus PMs, in the slag.

**Alkali metal carbonate** Sodium carbonate is a very powerful basic flux and will readily form silicates. It melts at  $850^\circ\text{C}$  whilst at  $950^\circ\text{C}$  it decomposes to  $\text{Na}_2\text{O}$  and  $\text{CO}_2$ . It can hold carbon and other infusibles in suspension and act as a desulfurizing agent. When initially introduced to the furnace it can form dust, with possible losses. This can be prevented by covering the charge with a thin layer of NaCl. Potassium carbonate behaves similarly, melting at  $890^\circ\text{C}$ . It has the disadvantage of deliquescing and it is more expensive.

**Silica** This is a strongly acidic flux and is usually added as silver sand. If there is high silica content in the sample it may not be necessary to add any, but if the charge is deficient in silica,  $\text{PbO}$ ,  $\text{CuO}$ ,  $\text{FeO}$ , and borax can attack the walls of the crucible. Also, a deficiency can lead to the formation of a matte in

**Table 3** Fluxing agents

Basic	Acidic	Neutral
$\text{Na}_2\text{CO}_3$	$\text{SiO}_2$	$\text{CaF}_2$
$\text{K}_2\text{CO}_3$	$\text{Na}_2\text{B}_4\text{O}_7$	$\text{NaCl}$
$\text{PbO}$		$\text{LiBO}_2$
$\text{Na}_2\text{B}_4\text{O}_7 = \text{Na}_2\text{O} \cdot 2\text{B}_2\text{O}_3$ and $\text{LiBO}_2 = \text{Li}_2\text{O} \cdot \text{B}_2\text{O}_3$ .		

**Table 4** Silicates and borates of bivalent base RO

Subsilicate	$4\text{RO} \cdot \text{SiO}_2$	Orthoborate	$3\text{RO} \cdot \text{B}_2\text{O}_3$
Monosilicate	$2\text{RO} \cdot \text{SiO}_2$	Pyroborate	$2\text{RO} \cdot \text{B}_2\text{O}_3$
Sesquisilicate	$4\text{RO} \cdot 3\text{SiO}_2$	Sesquiborate	$3\text{RO} \cdot 2\text{B}_2\text{O}_3$
Bisilicate	$\text{RO} \cdot \text{SiO}_2$	Metaborate	$\text{RO} \cdot \text{B}_2\text{O}_3$
Trisilicate	$2\text{RO} \cdot 3\text{SiO}_2$	Tetraborate	$\text{RO} \cdot 2\text{B}_2\text{O}_3$

sulfurous samples. Conversely, if the silica content is too high a viscous slag will result, preventing the collection of the PM and giving difficulty in pouring.

**Other agents** Fluorspar,  $\text{CaF}_2$ , melts at  $900^\circ\text{C}$  without dissociation. It is a neutral flux used to dissolve high-melting-point materials such as alumina, calcium phosphate, and barium compounds. A small amount may be added to any charge as it will aid in the separation of the button.

Sodium chloride volatilizes at red heat but it can be used to prevent dusting, exclude air, and wash down the sides of the crucible.

Lithium metaborate is useful in the assay of chromium-containing samples that do present particular difficulties and, as mentioned in earlier sections, litharge and nitre have fluxing properties.

**Temperature and furnace conditions** One of the commonest sources of error during pot fusion is poor mixing. It is essential, first, that the collector and reducing agent are intimately mixed and, second, that the sample is intimately mixed with these. This is best achieved by using a spatula and sulfite paper or by grinding them together in a mortar and pestle. Some assayers blend in the crucible with an iron rod. The fluxes can then be added, again ensuring a good mix. Finally, a covering of  $\text{NaCl}$  may be applied.

The furnace is brought to temperature (normally  $1000$ – $1200^\circ\text{C}$ ). Lead collections are normally run in the range  $1000$ – $1100^\circ\text{C}$  although temperatures as high as  $1300^\circ\text{C}$  may be used if alumina or baria ( $\text{BaO}$ ) are present. Nickel sulfide is run at  $\sim 1150^\circ\text{C}$  and copper at  $1250^\circ\text{C}$ . On charging the sample the temperature will fall to  $800$ – $900^\circ\text{C}$  and a very viscous melt forms that will help hold up the collector and give an efficient reaction. The temperature then rises again to the set temperature and the melt becomes fluid.

The time in the furnace will depend on whether it is a reducing (gas) fire or oxidizing (electrical) fire and on the sample size and type. For a lead collection of 25 g, 25–30 min is usually sufficient, increasing to 40–60 min for 100 g. Nickel sulfide runs take  $\sim 45$  min, and copper collection  $\sim 1$  h.

On completion of the melt the contents of the crucible are swirled to bring any unattacked material to the surface. If there is any residue in the crucible, it must be returned to the furnace to complete the fusion. Upon completion of the fusion the contents are poured into a lightly oiled iron mold, ensuring the collector goes into the mold apex and not down the sides. After cooling the slag is separated from the button and the slag retreated in the same crucible if appropriate. The button is physically cleaned of any

adhering slag and placed aside for the next stage of analysis. For nickel sulfide this will be grinding and chemical attack but for a lead button this could be any or all of scorification, cupellation, and a chemical attack, depending on the assay in question.

**Other considerations** There is no single, universal flux but the general flux referred to as ‘white flux’ can prove very useful. The composition is equal parts (by weight) of borax, sodium carbonate, potassium carbonate, and sodium chloride. This can be mixed and kept in bulk. Samples of 1–10 g are blended with 55 g of litharge and 5 g of fluor and 100 g of the white flux is added. This suffices for many materials of the sweep-type especially, and the addition of silica is unnecessary as corrosion of the crucible supplies all that is required to the fusion.

An obvious exception to the use of this type of flux is for sulfur-containing materials such as copper concentrates and, in the case of nickel sulfide, collection. Nickel sulfide is used when one is interested in PGMs rather than in gold and silver. The general type of charge is a sample of 20–50 g mixed intimately with 20 g of  $\text{NiO}$ , 12.5 g flowers of sulfur, and 10 g of silver sand placed in the crucible. The fluxes of 60 g of borax plus 30 g of sodium carbonate are placed on top and fused. The slag is rerun with a half-charge of collector and flux and the buttons combined for the ensuing analysis.

The charge for a copper collection is fairly standard, being:

- Sodium carbonate 80 g,
- Borax 40 g,
- Silver sand 12 g,
- Copper(I) oxide 50 g,
- Deactivated charcoal 2 g, and
- Sample 1–10 g.

Such a fusion gives a button of  $\sim 40$  g, but depends on the copper content of the sample.

### Scorification

This process is an oxidizing fusion and can be performed in two main areas.

It can be used as a replacement for pot fusion but is little used for accurate work. The sample, after oxidation, is mixed with lead, which acts as a collector and a fluxing agent. The sample is placed in a shallow impervious fireclay dish, covered with borax and placed in a furnace under an air draught. The lead is slowly oxidized to  $\text{PbO}$  with some being volatilized. During the fusion the borax reacts with the  $\text{PbO}$  to form a slag initially on the edges of the molten



charge. As the reaction proceeds the slag moves toward the center where the exposed lead forms an 'eye'. The scorifier is withdrawn from the furnace as the 'eye' closes and is poured into a mold as in pot fusion. This procedure can be useful for sulfides as the air decomposes the sulfide and the resulting oxide dissolves in the PbO.

The second and more widely used area of scorification is in the decrease in the size of lead buttons from pot fusion to a more manageable size whether they are going to be treated further by a chemical attack or cupellation (see below). This procedure will also be used if more than one button needs to be combined for subsequent operations. It proceeds in exactly the same manner as described above.

Bugbee gives a comprehensive description of this process.

### Cupellation

This is another oxidizing fusion that can be used in two main areas. The process is that a PM-containing lead button is placed on a cupel, which is essentially a shallow depression in a magnesia (MgO) or bone ash block.

The blocks are placed in a preheated furnace under an air draught; the lead melts and is oxidized. The molten PbO collects the non-PM-bearing part of the charge, small amounts of base metal oxide, and is absorbed into the porous cupel with ~1–2% being volatilized. A cupel can hold about its own weight in PbO. The PMs form a bead and due to their high surface tension are not absorbed.

The cupel is cooled, the PM bead removed, and taken for subsequent analysis. In some procedures the bead is simply weighed and a total PM content reported. This is generally used in mining laboratories where such results are sufficient for control purposes. In other procedures the bead is chemically analyzed to determine the individual PMs as required.

If gold is the element of interest then use is made of the 'parting' procedure. The bead is heated with nitric acid to dissolve away the silver, leaving the gold. In this case the Ag–Au weight ratio needs to be ~5:1 in order to obtain efficient parting. If the sample does not give this ratio then the inquartation method is used. This is the deliberate addition of silver at the cupellation stage to bring the ratio to the correct level for parting. The PGMs can impart a particular appearance to a silver bead. This is amply described by Bugbee. A very relevant practical point in cupellation is the cooling of the beads. Molten silver absorbs oxygen and if the cooling of the beads is not even there may be spitting, which gives the

beads small protuberances that may impair further processing or lead to losses.

The second cupellation procedure is used in the analysis of Dore bullion and fine gold alloys and bullion. In this case the sample has no preliminary fire assay and goes straight to the cupel. Generally, samples are of 0.5 g, weighed on a six-place balance. The sample is placed on lead foil along with ~0.1 g of copper wire to help carry away impurities and a piece of silver to give an Ag–Au weight ratio of ~2.5. These pieces can be prepared in bulk by cutting pieces of wire or sheet with a template. The lead foil is wrapped up as a small packet and the samples placed on preheated cupel blocks in a furnace at 1000°C.

The cupels are removed upon completion of the reaction, which can be monitored visually, the bead 'brightening' when the final thin layer of oxide has been absorbed.

The beads are then physically conditioned for parting. The bead is cleaned physically and tapped so that it can be rolled to a fairly long narrow strip. This is then coiled loosely and placed in a parting vessel. The silver is dissolved away by using, in turn, 1.16 relative density nitric acid, a boiling water wash, 1.284 relative density nitric acid, and a final wash to remove the acid. The resulting coils of gold, or cornets, are then dried, annealed to prevent them from breaking up, and weighed as gold.

With this procedure standards of 99.999% proof gold are included in the furnace run to cover the area of the furnace used. This is because there will be variations in each run and small gains may well be observed. The proof gold standard allows correction for these gains.

The accuracy obtained with this method is perhaps the highest of any analytical method one can use and is within 0.02% at very high levels of gold. For the analysis of fine gold at 99.99% the accuracy and precision will be better than 0.01%.

It is perhaps paradoxical in these days of high technology that one of the most accurate analyses carried out in the world originated in the time of the pharaohs with a basically unchanged methodology.

**See also: Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Precious Metals.**

### Further Reading

- Beamish FE (1966) *The Analytical Chemistry of the Noble Metals*. Oxford: Pergamon Press.
- Beamish FE and Van Loon JC (1977) *Analysis of Noble Metals, Overview and Selected Methods*. London: Academic Press.

- Bugbee EE (1983) *A Textbook of Fire Assaying*. Colorado: School of Mines.
- Bugbee EE (1991) *A Textbook of Fire Assaying*, 3rd edn. Colorado: School of Mines.
- Haffty J, Riley LB, and Goss WD (1977) *A Manual on Fire Assaying and Determination of the Noble Metals in Geological Materials*.
- Lenahan HC and Murray-Smith R de L (1986) *Assay and Analytical Practice in the South African Mining Industry*. South African Institute of Mining and Metallurgy.

- Methods Manual – vol. 1: Gold Analysis in Alkaline Cyanide Solutions*. Society of Mineral Analysts Publication.
- Methods Manual – vol. 2: Gold Analysis in Mill Process Solids*. Society of Mineral Analysts Publication, 1989.
- Shepard O. *Fire Assaying*. Met-Chem Research.
- Smith E (1992) *Sampling and Assay of the Precious Metals*. Alfred Waller.
- Van Loon JC and Barefoot RR (1991) *Determination of the Precious Metals: Selected Instrumental Methods*. New York: Wiley.

## FLAME ATOMIC ABSORPTION SPECTROMETRY

See ATOMIC ABSORPTION SPECTROMETRY: Flame

## FLAME PHOTOMETRY

See ATOMIC EMISSION SPECTROMETRY: Flame Photometry

## FLOTATION

See FOAM FRACTIONATION AND FLOTATION

## FLOW ANALYSIS

### Overview

**E A G Zagatto**, University of São Paulo, Piracicaba SP, Brazil

**P J Worsfold**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Flow analysis refers to quantitative analytical chemistry in flowing streams, and was devised in 1951 by L.T. Skeggs. The main driving forces for its inception and early development were the increasing demand for chemical assays at the Veterans Administration Hospital (Cleveland, OH, USA) and the consensus

that robotic systems at that time were too cumbersome to be practical. During development, it was realized that flow systems could play a wider role in laboratory automation and solution management, i.e., carrying out wet chemical analyses. Nowadays, the number of flow-based analytical methods, commercially available instruments, end users, and publications in the field continues to grow, demonstrating the worldwide acceptance of flow analysis. This article provides an overview of fundamentals, characteristics, flow analysis modes, theory, and applications.

### Fundamentals

Flow analysis is an approach to mechanized analytical chemistry usually carried out inside narrow bore tubing. An aliquot of an aqueous sample is



- Bugbee EE (1983) *A Textbook of Fire Assaying*. Colorado: School of Mines.
- Bugbee EE (1991) *A Textbook of Fire Assaying*, 3rd edn. Colorado: School of Mines.
- Haffty J, Riley LB, and Goss WD (1977) *A Manual on Fire Assaying and Determination of the Noble Metals in Geological Materials*.
- Lenahan HC and Murray-Smith R de L (1986) *Assay and Analytical Practice in the South African Mining Industry*. South African Institute of Mining and Metallurgy.

- Methods Manual – vol. 1: Gold Analysis in Alkaline Cyanide Solutions*. Society of Mineral Analysts Publication.
- Methods Manual – vol. 2: Gold Analysis in Mill Process Solids*. Society of Mineral Analysts Publication, 1989.
- Shepard O. *Fire Assaying*. Met-Chem Research.
- Smith E (1992) *Sampling and Assay of the Precious Metals*. Alfred Waller.
- Van Loon JC and Barefoot RR (1991) *Determination of the Precious Metals: Selected Instrumental Methods*. New York: Wiley.

## FLAME ATOMIC ABSORPTION SPECTROMETRY

See ATOMIC ABSORPTION SPECTROMETRY: Flame

## FLAME PHOTOMETRY

See ATOMIC EMISSION SPECTROMETRY: Flame Photometry

## FLOTATION

See FOAM FRACTIONATION AND FLOTATION

## FLOW ANALYSIS

### Overview

**E A G Zagatto**, University of São Paulo, Piracicaba SP, Brazil

**P J Worsfold**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

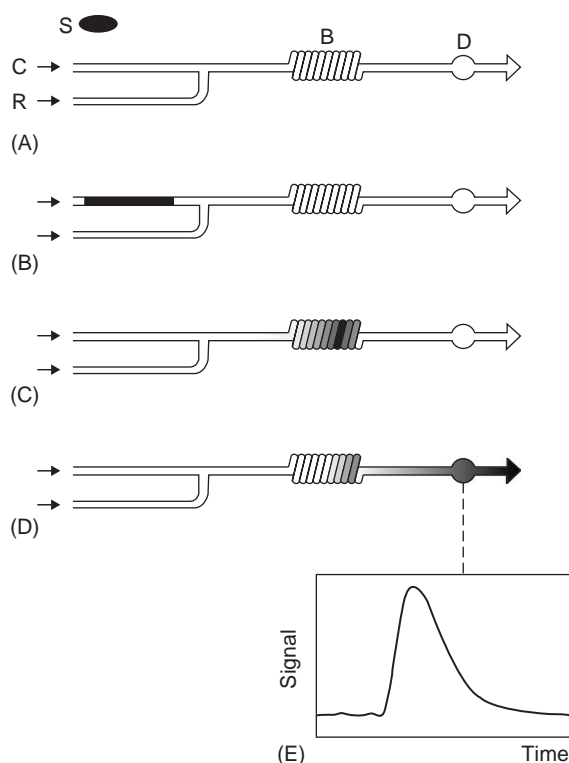
### Introduction

Flow analysis refers to quantitative analytical chemistry in flowing streams, and was devised in 1951 by L.T. Skeggs. The main driving forces for its inception and early development were the increasing demand for chemical assays at the Veterans Administration Hospital (Cleveland, OH, USA) and the consensus

that robotic systems at that time were too cumbersome to be practical. During development, it was realized that flow systems could play a wider role in laboratory automation and solution management, i.e., carrying out wet chemical analyses. Nowadays, the number of flow-based analytical methods, commercially available instruments, end users, and publications in the field continues to grow, demonstrating the worldwide acceptance of flow analysis. This article provides an overview of fundamentals, characteristics, flow analysis modes, theory, and applications.

### Fundamentals

Flow analysis is an approach to mechanized analytical chemistry usually carried out inside narrow bore tubing. An aliquot of an aqueous sample is



**Figure 1** Didactic presentation of a flow analyzer. S, sample; C, sample carrier stream; R, reagent stream; B, reactor; D, detector. For details, see text.

introduced into the flow system (see Figure 1), and pushed toward the detector by the carrier/wash stream. During transport through the analytical path, the sample undergoes dispersion and dilution, resulting in a well-defined sample zone that undergoes reproducible, online physical and chemical treatment, e.g., dilution, reagent addition, and dialysis. Sample passage through the detector results in a transient signal that is recorded as a peak, the height of which is proportional to the analyte concentration in the sample.

## Characteristics of the Flow System

### The Sample Volume Is Small

The sample aliquot is precisely selected (see Figure 1A), usually by means of either a loop-based or a time-based injection device. Typical sample volumes are 5–500  $\mu\text{L}$ , although volumes in the nanoliter range can be handled in microfluidic systems. Volume-limited samples such as biological fluids are therefore easily analyzed; *in vivo* assays are efficiently performed and operations such as dialysis, gas diffusion, and ion exchange are efficiently accomplished in-line. Reagent consumption is also low, typically 0.5–2.5  $\text{mL min}^{-1}$  per channel, thus

minimizing waste and allowing ‘environmentally friendly’ chemical analysis.

### The Analytical Path Is a Closed Environment

The selected sample volume is inserted into the flow system (see Figure 1B) and, while inside the analytical path, there is no physical contact of the sample with the external environment (and vice versa). Consequently, analyte losses and/or sample contamination, which lead to unreliable results, and environmental pollution, are avoided. The flow analyzer can therefore be regarded as a ‘clean room’ inside which the sample is manipulated under reproducible conditions. Drawbacks related to the use of hazardous and/or volatile reagents are also minimized.

### Dispersion/Dilution Is Inherent to Flow Analysis

During transport through the analytical path, the sample undergoes a continuous dispersion process and sudden dilution at the points where confluent streams are added (see Figure 1C), and control of these processes is a key parameter for system design. The degree of dispersion/dilution should be minimized when sample integrity needs to be preserved and increased when controlled sample dilution is needed in order to, e.g., match the dynamic range of the detector.

### Sample Passage Through the Analytical Path Is Fast

The mean sample residence time inside the analytical path is usually 10–180 s. As a consequence, some of the processes involved, e.g., derivatization, may not reach completion, allowing partial but reproducible development of chemical reactions, quantification of unstable chemical species, and easy implementation of catalytic methods. Flow-based procedures where chemical equilibria are not reached have often been proposed; this aspect should be considered with care however, particularly in relation to masking, which should always be quantitative.

Another feature arising from the short sample residence time is the feasibility of rapid and high throughput analysis, and this is relevant for process control, screening, and studies involving concentration-orientated feedback mechanisms. Alternatively, if a relatively long sample processing time is required, e.g., for procedures involving slow chemical reactions, specific strategies such as stream segmentation or sample stopping can be exploited.

### Sample Management Is Reproducible

Conditions for sample handling are reproducible from one sample to the next, and this feature is

particularly important for applications involving slurry processing, in-line sample digestion, exploitation of micellar media, monitoring of suspensions, and the use of renewable sensors. In some circumstances, e.g., in the analysis of sample batches with high variability in matrix composition, sample processing may not be reproducible from one sample to another. In such cases, the standard addition method or matrix matching procedures to circumvent matrix effects are easily implemented.

### The Analytical Signal Is Transient

When the processed sample passes through the detector, a transient modification of the monitored signal is obtained and is usually recorded as a peak (see Figure 1D). For most applications, the height (or area) of this peak is proportional to the analyte concentration and is the measured parameter. The detector is permanently situated in the flow manifold, thereby maintaining detection geometry during the analysis of entire sample batches, which leads to an improvement in repeatability. Moreover, monitoring the detector output in the presence (analytical signal) and absence (baseline) of the flowing sample is a useful diagnostic tool for assessing the proper operation of the system.

### Sample Throughput Is High

Flushing time, the interval between achievement of the maximum analytical signal and baseline restoration, is typically 12–120 s. The next sample can therefore be introduced without a long delay time and this permits a high sample throughput, typically 30–300 h<sup>-1</sup>. This is particularly advantageous in relation to, e.g., repetitive measurements, simultaneous determinations, titrations, standard additions, and quality control.

### Concentration Gradients Can Be Exploited

The recorded peak (see Figure 1E) represents an ‘infinite’ number of fluid elements in the flowing sample, each one associated with a particular mean sample residence time, and hence a particular condition for sample manipulation and a particular analyte concentration. A number of different analytical signals are therefore obtained during monitoring of the sample zone, allowing the easy implementation of analytical procedures involving exploitation of concentration gradients. To this end, algorithms for multiparametric analysis are generally used. Moreover, this permits the generation of an analytical curve using only one standard solution.

The concentration gradients established may also be undesirable, as they can impair measurement.

A typical example is the Schlieren effect arising from the presence of neighboring fluid elements with different compositions, and therefore refractive indices, which results in a myriad of ‘lenses’ and ‘mirrors’ randomly established inside the flowing sample. A common example of this is the introduction of a saline sample (high ionic strength) into an aqueous carrier stream. Occurrence of the Schlieren effect can limit the signal-to-noise ratio in some flow-based spectrophotometric procedures; it can be reduced and/or compensated by improving system design, exploiting matrix matching, and/or using a dual-wavelength spectrophotometer.

### A ‘Time Window’ Is Available

The mean available time intervals for the development of all the steps to which the sample is subjected are strictly constant because rigid timing control is inherent to flow analyzers. In view of this ‘time window’, the samples and reference solutions are manipulated under reproducible conditions, and this aspect leads to more precise results. Moreover, exploitation of this feature permits the easy implementation of analytical procedures even when analogous batch procedures do not exist.

Another favorable aspect inherent to flow-based analytical methods is that the possibility of random errors due to operator intervention is significantly reduced. Classical glassware, e.g., volumetric flasks, pipettes, funnels, and burettes, designed a century ago are less intensively used. Attention should be paid, however, to the possibility of systematic errors, which are generally increased when a traditional analytical procedure is carried out in a mechanized fashion. Regarding this aspect, detecting and circumventing potential sources of inaccuracy can be accomplished in-line.

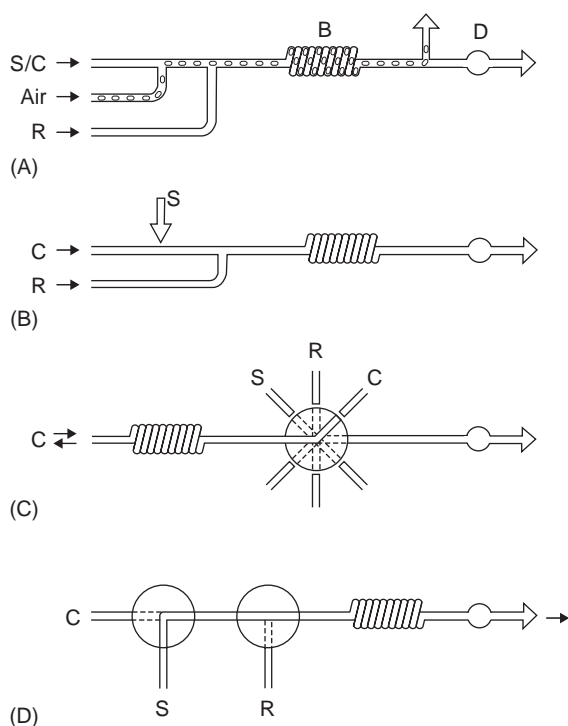
In addition to the above-mentioned aspects, the present trend toward full automation, miniaturization, and *in situ* analysis reflects the attractive characteristics of flow-based methodologies.

## Types of Flow Analysis

Flow analysis can be classified into three different types, each characterized by the distinct nature of the flowing streams, the versatility of the flow analyzer, and the degree of external control.

### Segmented Flow Analysis

In segmented flow analysis, either the sample or a wash solution is aspirated to merge with an air (or immiscible solvent) stream (see Figure 2A), resulting in a segmented stream comprising a number of

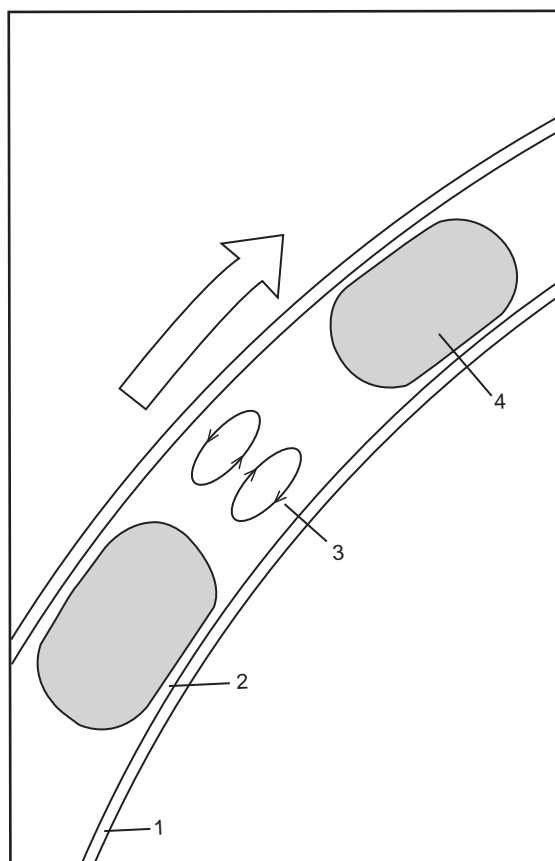


**Figure 2** Flow diagrams of different flow analyzers comprising only the sample, carrier, and reagent solutions. (A), (B), (C), (D) correspond to segmented flow, flow injection, sequential injection, and multi-commuted flow analyzers, respectively. Symbols as in **Figure 1**.

discrete plugs. As a consequence of stream segmentation, broadening of the sample zone (thus carry-over) is minimized, mixing conditions within each plug are improved, and the tubing inner walls are scrubbed. These favorable features are a consequence of the typical flow pattern inside the plugs (see **Figure 3**). The segmented sample zone is manipulated inside the analytical path under conditions of limited dispersion. As two phases are involved, sample manipulation is, however, somewhat restricted by compression/expansion phenomena (often temperature dependent) that can limit versatility and sample throughput.

The air phase is generally physically removed immediately before detection or, in modern analyzers, electronically removed with a bubble-gating device. A steady state situation is approached at the central portion of the monitored sample; therefore, the recorded peak is flat topped.

The initial development of air-segmented flow analysis (also known as CFA, from 'continuous flow analysis') was restricted to Technicon Corporation Inc., owner of the main patents until the mid-1970s. The first air-segmented systems were marketed in clinical chemistry with the AutoAnalyzer<sup>®</sup> trade



**Figure 3** Flow pattern of a segmented flow analyzer. Figure refers to a part of a coiled reactor encompassing a liquid segment between two of the successive gaseous bubbles. 1, Tubing wall; 2, thin liquid layer adherent to it; 3, movement inside the liquid phase; 4, gaseous bubble. Overall displacement is specified by the outer large arrow.

name. Expansion of the original CFA concept led to the appearance of multi-channel segmented flow analyzers, where several peaks (each for a different analyte) are simultaneously recorded per sample. In clinical chemistry, this multiple output represents the so-called 'multiple analysis chart', recognized as a very important tool for clinical diagnosis.

The dimensions of the analytical path cannot be reduced at will, as the bubble pattern has to be preserved. Moreover, a large sample volume is required in order to provide an almost steady-state situation at the center of the flowing sample zone. As a consequence of these features, reagent consumption is relatively high. As sample broadening is low, several samples are simultaneously manipulated inside the analytical path and this is important for achieving long sample residence times without impairing sample throughput. Passage of the sample zone through the pumping unit is also a drawback.

System operation is not very amenable to external control, as the only moving element in the flow analyzer is the sampler arm. This does, however, mean that such systems are rugged and reliable. Segmented flow analyzers played a dominant role in automated chemical assays for several decades and are still the basis for several commercial analyzers, particularly in clinical and environmental (water quality) laboratories.

### Unsegmented Flow Analysis

In this type, the sample is manipulated inside a continuously flowing unsegmented stream that acts as an incompressible and homogeneous column. This liquid column can be efficiently managed, allowing different processes such as stream splitting, stream redirecting, flow reversal, sample stopping, sample slicing, zone sampling, and merging zones to be easily accomplished. Consequently, system design is simplified, versatility is enhanced, and timing is more precise compared with segmented flow analyzers. As segmentation is not involved, a higher degree of external control can be attained, and hyphenation with other techniques is more efficiently accomplished. System miniaturization is easier with unsegmented flow analysis and has led to the appearance of microfluidic flow systems. These features are the main reasons for the rapid development and acceptance of unsegmented flow analysis, especially flow injection and sequential injection analysis.

Flow injection analysis was conceived in the mid-1970s independently by Ruzicka and Hansen and by Stewart *et al.*, who demonstrated the feasibility of sample injection into a continuously flowing unsegmented carrier stream for analytical purposes. The potential and versatility of so simple an analyzer amazed the scientific community. Since then, it has been by far the most investigated and applied unsegmented flow system, as demonstrated by more than 18 000 scientific articles and at least 12 manufacturers. In flow-injection analysis, the sample is inserted into an unsegmented carrier stream (often a reagent) and the established sample zone undergoes a continuous dispersion process, resulting in reproducible mixing of the sample with the carrier stream and, eventually, with other confluent streams (see **Figure 2B**). Control of sample dispersion/dilution enables its application to many different analytical methods. The recorded peak shape tends toward Gaussian in view of the efficiency of the dispersion process. Operation of the flow injection system relies on three cornerstones: sample injection, controlled dispersion, and reproducible timing. As flow is unsegmented, commutation (i.e., the ability to switch

fluidic zones between different channels) is an important feature of the flow injection analyzer and enhances versatility. Although manual operation of the system is possible, computer control is usually used.

Sequential injection analysis was proposed by Ruzicka and Marshall in 1990 as the answer to the increasing demand for system versatility and automation. In this system, a multiposition selection valve – the heart of the system – selects the sample and reagent aliquots to be sequentially aspirated, generating a stack of well-defined zones that is directed toward a holding coil. After a predefined time interval, the flow is reversed and the multi-position valve is switched in order to direct the stack toward the detector. During transport through the analytical path, the zones penetrate one another and mixing among them gives rise to the detectable species to be monitored. Versatility is an important characteristic of sequential injection analyzers, as several changes to the methodology can be achieved through software control of the system parameters rather than actual physical changes to the system architecture. Each port of the valve is dedicated to a specific purpose; therefore, combinations of sample, standards, reagents, sample treatment devices, and detectors around the valve are easily modified to suit a particular analysis.

### Zone Fluidics

This type encompasses a number of different concepts and variations of analyzers that have been developed as a logical consequence of the increasing presence of the computer in the laboratory, the need for enhanced versatility, and the diversity of applications demanding *in situ* (and often *in vivo*) assays. Such analyzers are usually small, lightweight, and portable. As a rule, they are also very versatile systems with a high degree of external control, involving both segmented and unsegmented (and also monosegmented) flows. Operation of discrete devices such as pumps, injectors, bidirectional valves, and multidirectional valves is usually controlled in real-time through feedback mechanisms. Batch injection analysis, bead injection analysis, discontinuous flow analysis, ‘lab-on-valve’, multisyringe flow analysis, and multi-commuted flow systems are examples of this type of flow analysis.

The multi-commuted system (see **Figure 2D**) is one of the most versatile systems, and exploits loop-based injection, time-based injection, or binary sampling for introducing the samples, reagents, and diluents in a programmable fashion. With discrete operation of its components, several approaches such as random reagent access, computer-controlled



splitting, multisite detection, and in-line sample preparation are efficiently accomplished. Zone fluidics (i.e., control and manipulation of flowing zones) is the underlying principle of this type of flow analyzer.

## Theoretical Aspects

Understanding how the sample zone interacts with the surrounding solution is of the utmost importance for system design and method implementation. In this regard, the flow regime – also alterations in flow pattern due to implementation of different strategies – and sample dispersion/dilution are key aspects to be considered.

### Flow Regime

In view of the typical flow rates and tubing dimensions of a flow analyzer, laminar flow is the prevailing regime, Reynolds number often being  $<2000$ . Laminarity has also been verified near confluence points where turbulence is more likely to occur.

Laminar flow in open tubes is characterized by the typical parabolic profile of the axial linear velocity distribution. Consequently, the linear velocity of a fluid element at the center of the tube is about twice the average linear flow velocity, whereas the linear velocity of the fluid elements near the tubing inner walls tends to zero. Therefore, after sample introduction, the chemical species in the frontal region of the sample zone, traveling at higher speed, diffuse toward less concentrated regions (and vice versa) whereas chemical species at the trailing edge, near the tube inner wall, diffuse toward faster moving central regions (and vice versa). This results in the so-called zone broadening effect with its deleterious influence on mixing conditions and, thus, sample throughput. In order to minimize this effect, different strategies such as segmentation (see Figure 3), reactor coiling, improvement in reactor design, and the use of supercritical flows have been proposed.

### Sample Dilution/Dispersion

A common feature of flow analysis is that during sample transport toward the detector, the combined effects of dilution and dispersion decrease the sample concentration, and the extent of these processes is very important for system design. Dilution occurs instantly at every confluence point due to mixing of the sample with the confluent stream. Under ideal mixing conditions, the concentration  $C$  of a chemical species in a fluid slice immediately before a confluence point is modified to  $C'$  according to

$$C' = C[N_1/(N_1 + N_2)] = C(N_1/N_3) \quad [1]$$

where  $N_1$ ,  $N_2$ , and  $N_3$  are flow rates of carrier, confluent, and resulting streams, respectively, the latter flowing at  $N_3$  [or  $(N_1 + N_2)$ ]. The enlargement of the sample zone by dispersion is also a consequence of confluent stream addition. These phenomena are often masked by downstream sample dispersion. Dilution predominates in segmented flow analysis and is also an important parameter in unsegmented flow systems with large sample volumes.

Dispersion occurs continuously during transport of the sample through the analytical path, due to the redistribution of material from the sample zone into the carrier stream, and is caused by convective and diffusive mass transport. Convective mass transport is a consequence of the parabolic distribution of the linear velocities of every fluid element, whereas diffusive mass transport is dependent on both the concentration differences between neighboring fluid elements and the diffusion coefficient. Although diffusive mass transport occurs in an isotropic fashion, only its radial component is relevant as a factor influencing dispersion inside a straight tubular reactor. Dispersion is an important parameter in unsegmented flow analysis, and can be efficiently controlled to give highly reproducible physical dispersion and hence reproducible analytical results. This effect is also relevant – although to a lesser extent – in segmented flow analysis, where the film of the solution established on the tube inner wall (see Figure 3) is the main source of carryover between successive segments. Dispersion can be experimentally manipulated by varying parameters such as sample volume, reactor geometry, and flow rates. Different models have been proposed for a quantitative description of sample dispersion in flow analysis and the convective–diffusive model (eqn [2]) has often been used to describe the concentration of any fluid element as a function of its spatial and temporal coordinates:

$$\begin{aligned} D_m(M^2C/Mx^2 + M^2C/Mr^2 + 1/rMC/Mr) \\ = MC/Mt + \Phi_0(1 - r^2/a^2)MC/Mx \end{aligned} \quad [2]$$

where  $D_m$  is the molecular diffusion coefficient,  $C$  the concentration of the specified fluid element,  $x$  the distance along the tube,  $r$  the radial distance from the tube axis,  $a$  the tube radius,  $t$  the time, and  $\Phi_0$  the maximum linear flow velocity at the center of the tube.

Equation [2] is derived from the mass balance in a differential volume of the fluid and considers the axial and radial concentration gradients as well as the linear velocities established in a laminar flow pattern in the absence of chemical reactions. The first term ( $M^2C/Mx^2$ ) refers to the axial diffusion and the other terms to the radial diffusion. The expression



$\Phi_0(1 - r^2/a^2)$  refers to the parabolic shape of the sample plug due to the different velocities of each fluid element. The expression  $MC/Mt$  refers to the temporal variation in concentration due solely to concentration gradients and is the main contribution to the very low sample dispersion occurring in situations where the flow is stopped. Equation [2] is presented in a differential fashion, so numerical calculations are required and boundary conditions need to be well defined in order to apply them to practical situations. This convective–diffusive model permits a quantitative description of sample dispersion in a flow analyzer, and provides a good simulation to assist further system design and method optimization. Equation [2] should be modified in order to consider the presence of reactors other than the open tubular reactor (e.g., coiled, knotted, packed bead), mixing chambers, micro-columns, and other specific devices in the analytical path. Other quantitative models have also been proposed for specific applications, particularly the tanks-in-series model for describing sample dilution/dispersion in relation to mixing chambers.

## Instrumentation

The flow analyzer involves simple apparatus such as samplers, liquid drivers (peristaltic pumps, piston pumps, solenoid pumps), injection devices (rotary valves, injector-commutators), reactors and flow lines (usually narrow bore tubing), mixing chambers, and flow-through detectors. As a rule, these devices are readily available in most laboratories devoted to chemical analysis. Regarding detection, almost all analytical techniques have been used in flow analysis; a small flow cell volume and a short response time that is compatible with system dynamics are important detector parameters.

## Applications

The number of applications in flow analysis is increasing, and several are related to modern concepts such as multi-commutation, multi-pumping, zone fluidics, and bead manipulation. This is perhaps the most important consequence of the favorable characteristics of the flow system, often leading to improved sample treatment and quantification.

Flow analysis is used mostly in fields characterized by a high demand for chemical assays; therefore, reported applications predominate in agronomic, environmental, geological, biomedical, and pharmaceutical areas, as well as industrial quality control. The notorious expression “My system processes a

number of samples but I only have a few” – the driving force toward the development of early flow analyzers – is sometimes no longer valid. In fact, there are situations where flow analysis is a viable solution for assaying a single sample.

Spectrophotometry and related techniques requiring intensive manipulations of liquid solutions are the most commonly used detection techniques in flow analysis because the flow analyzer is an excellent solution manager. Applications involving electro-analytical techniques, especially potentiometry and amperometry, have also been extensively reported. Techniques requiring sample monitoring under flowing conditions, e.g., atomic absorption/emission spectrometry (especially with hydride generation), flow enthalpimetry, and thermal lens spectrometry, as well as chemi- and bio-luminescence, are all enhanced by flow analysis as a means of online sample processing. There is also a trend toward using flow analysis in conjunction with online concentration/separation and with biosensors. Improved sample handling associated with flow analysis systems means that this family of techniques will continue to have a major impact on many application areas of Analytical Science.

**See also:** **Flow Injection Analysis:** Principles; Instrumentation; Detection Techniques; Environmental and Agricultural Applications; Clinical and Pharmaceutical Applications; Industrial Applications. **Segmented Flow Analysis.** **Sequential Injection Analysis.**

## Further Reading

- Burguera JL (ed.) (1989) *Flow Injection Atomic Spectroscopy*. New York: Dekker.
- Coakley W (1981) *Handbook of Automated Analysis Continuous Flow Techniques*. New York: Dekker.
- Fang Z (1993) *Flow Injection Separation and Preconcentration*. Weinheim: VCH.
- Karlberg B and Pacey GE (1989) *Flow Injection Analysis. A Practical Guide*. Amsterdam: Elsevier.
- Martinez Calatayud J (1996) *Flow Injection Analysis of Pharmaceuticals: Automation in the Laboratory*. London: Taylor and Francis.
- Ruzicka J and Hansen EH (1975) Flow injection analyses. Part I. A new concept of fast continuous flow analysis. *Analytica Chimica Acta* 78: 145.
- Ruzicka J and Hansen EH (1988) *Flow Injection Analysis*, 2nd edn. New York: Wiley-Interscience.
- Ruzicka J and Marshall GD (1990) Sequential injection: a new concept for chemical sensors, process analysis and laboratory assays. *Analytica Chimica Acta* 237: 329.
- Sanz Medel A (1999) *Flow Analysis with Atomic Spectrometric Detectors*. Amsterdam: Elsevier.
- Skeggs LT Jr. (1957) Automatic method for colorimetric analysis. *American Journal of Clinical Pathology* 28: 311.

Trojanowicz M (2000) *Flow Injection Analysis: Instrumentation and Applications*. London: World Scientific.

Ukeda H (2002) FIA bibliography (37). *Journal of Flow Injection Analysis* 19: 52.

Valcarcel M and Luque de Castro MD (1987) *Flow Injection Analysis. Principles and Applications*. Chichester: Ellis Horwood.

Valcarcel M and Luque de Castro MD (1989) *Automatic Methods of Analysis*. Amsterdam: Elsevier.

# FLOW INJECTION ANALYSIS

Contents

**Principles**

**Instrumentation**

**Detection Techniques**

**Environmental and Agricultural Applications**

**Clinical and Pharmaceutical Applications**

**Industrial Applications**

## Principles

**M D Luque de Castro**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Automatic methods applied to the analysis of series of samples can be grouped into three categories, namely batch, robotic, and continuous-flow methods. Flow injection (FI) methods fall into the last category. Although the first acronym for this technique was FIA, it was later shortened to FI both because rarely the overall analysis process is developed in an FI manifold and because FIA is the acronym used for fluoroimmunoassay. FI is a special mode of continuous analysis in which there are no bubbles separating the transport and/or reaction zones of samples successively introduced into the FI analyzer. Liquid samples (and, less commonly, solid or gaseous samples) are directly inserted into the dynamic system, along which they are carried and subjected or not subjected to one or more (bio)chemical reactions and/or separation processes. The dispersion or dilution of the analyte – or its reaction product – can be controlled through the geometric and dynamic features of the system. A continuous detector, usually furnished with a flow cell, is used to obtain transient signals. Neither physical equilibrium (homogenization of a portion of the flow) nor chemical equilibrium (reaction completeness) has been reached by the

time the signal is detected. Hence, FI methods can be regarded as fixed-time kinetic methods. Operational timing must always be highly reproducible because measurements are carried out under nonsteady physical and chemical conditions, so small changes may seriously affect the results obtained. In short, FI has four essential features: (1) unsegmented flow, (2) direct insertion, (3) controlled partial dispersion, and (4) reproducible operational timing.

This article provides an overview of this unsegmented flow technique. The units that make up an FI manifold and the basic analytical features are described. The types of recording and measurement provided by FI methods and the main similarities to and differences from other dynamic techniques are discussed, together with their theoretical and practical aspects. Elementary FI modes are briefly described in order to illustrate the versatility of this technique.

## The Basic FI System

A scheme of a basic FI manifold is shown in **Figure 1**. This consists of five main parts, namely:

1. The propulsion system, which establishes the flow at a rate as constant as possible of one or several solutions either containing dissolved reagents or merely acting as carriers. This function can be performed by a peristaltic pump (the propulsion system being more frequently used), a syringe or piston pump, a pressurized reservoir, a gravity-based unit, an electroosmotically pumped device, an osmotic pump, etc.

Trojanowicz M (2000) *Flow Injection Analysis: Instrumentation and Applications*. London: World Scientific.

Ukeda H (2002) FIA bibliography (37). *Journal of Flow Injection Analysis* 19: 52.

Valcarcel M and Luque de Castro MD (1987) *Flow Injection Analysis. Principles and Applications*. Chichester: Ellis Horwood.

Valcarcel M and Luque de Castro MD (1989) *Automatic Methods of Analysis*. Amsterdam: Elsevier.

# FLOW INJECTION ANALYSIS

Contents

**Principles**

**Instrumentation**

**Detection Techniques**

**Environmental and Agricultural Applications**

**Clinical and Pharmaceutical Applications**

**Industrial Applications**

## Principles

**M D Luque de Castro**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Automatic methods applied to the analysis of series of samples can be grouped into three categories, namely batch, robotic, and continuous-flow methods. Flow injection (FI) methods fall into the last category. Although the first acronym for this technique was FIA, it was later shortened to FI both because rarely the overall analysis process is developed in an FI manifold and because FIA is the acronym used for fluoroimmunoassay. FI is a special mode of continuous analysis in which there are no bubbles separating the transport and/or reaction zones of samples successively introduced into the FI analyzer. Liquid samples (and, less commonly, solid or gaseous samples) are directly inserted into the dynamic system, along which they are carried and subjected or not subjected to one or more (bio)chemical reactions and/or separation processes. The dispersion or dilution of the analyte – or its reaction product – can be controlled through the geometric and dynamic features of the system. A continuous detector, usually furnished with a flow cell, is used to obtain transient signals. Neither physical equilibrium (homogenization of a portion of the flow) nor chemical equilibrium (reaction completeness) has been reached by the

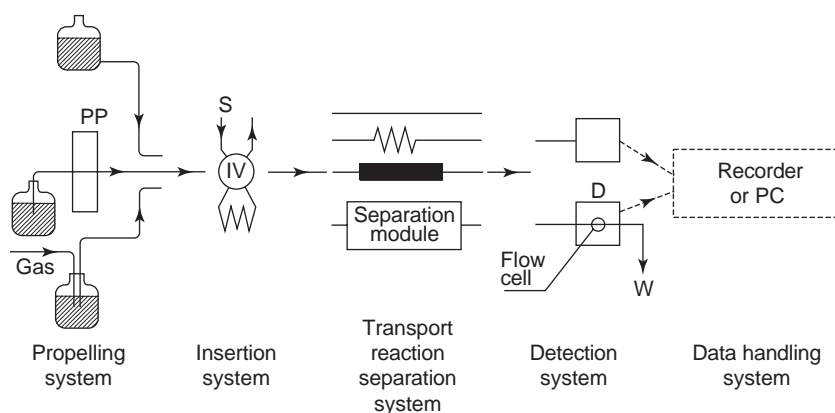
time the signal is detected. Hence, FI methods can be regarded as fixed-time kinetic methods. Operational timing must always be highly reproducible because measurements are carried out under nonsteady physical and chemical conditions, so small changes may seriously affect the results obtained. In short, FI has four essential features: (1) unsegmented flow, (2) direct insertion, (3) controlled partial dispersion, and (4) reproducible operational timing.

This article provides an overview of this unsegmented flow technique. The units that make up an FI manifold and the basic analytical features are described. The types of recording and measurement provided by FI methods and the main similarities to and differences from other dynamic techniques are discussed, together with their theoretical and practical aspects. Elementary FI modes are briefly described in order to illustrate the versatility of this technique.

## The Basic FI System

A scheme of a basic FI manifold is shown in **Figure 1**. This consists of five main parts, namely:

1. The propulsion system, which establishes the flow at a rate as constant as possible of one or several solutions either containing dissolved reagents or merely acting as carriers. This function can be performed by a peristaltic pump (the propulsion system being more frequently used), a syringe or piston pump, a pressurized reservoir, a gravity-based unit, an electroosmotically pumped device, an osmotic pump, etc.



**Figure 1** General scheme of an FI configuration. S denotes sample, PP peristaltic pump, IS injection system, D detector, W waste, and PC personal computer.

2. The sample introduction unit, which is usually an inexpensive low-pressure rotary valve that allows reproducible insertion of accurately measured sample microvolumes without the need for halting the flow – as in liquid chromatography (LC). Liquid samples can also be inserted by time-controlled aspiration at a merging point.

3. The transport–reaction system, which is often improperly called the ‘reactor’. This can be a straight, coiled, or knotted tube (0.3–0.8 mm i.d.), a length of tubing packed with a chemically active material (e.g., ion exchanger, sorbent material, redox reagent, an immobilized enzyme), a mini-mixing chamber, or separation module in which mass transfer across a gas–liquid, liquid–liquid, or solid–liquid interface takes place.

4. The detection system, which allows signals to be acquired and transduced as the analyte or the reacting plug passes through it. Nondestructive detectors such as spectroscopic (photometers, fluorimeters) and electrochemical detectors (potentiometric, voltammetric) use flow cells of various shapes and in different configurations, while destructive detectors (e.g., atomic spectrometric detectors) use no flow cell. The coupling of FI manifolds to high-resolution detectors such as Fourier-transform infrared and mass detectors is also possible.

5. The data acquisition system, a recorder or computer that receives the transduced signal from the detector.

The FI system can either be operated manually and the data acquisition performed by a recorder, or it can be readily automated to different degrees according to particular requirements. Presently, the integration of all the components of the FI manifold by an adequate software program is possible. All the

desirable functions can be controlled by a time- or an alert-based system.

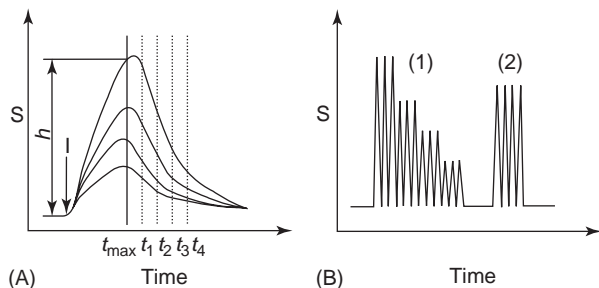
In addition to the above essential elements, FI configurations may include other elements or some modifications of the basic units for developing special FI modes, generally to achieve special aims. Manipulation of the flow with a programmable propelling device – usually a peristaltic pump – can involve stopping the flow, changing the flow rate, changing the flow direction for passage of a given plug through the detection point as many times as required, precise measurements of the working time, etc., thus giving rise to approaches such as stopped-flow, iterative change of the flow direction, temporal variation of the flow rate, and time-based injection. Also, the possibilities for coupling rotary injection valves (in-series, in-parallel, interiorly) and the capability of their loops for holding different FI devices enhances the versatility of this technique.

The present trend toward miniaturization in modern analytical chemistry has affected FI in two main ways: (1) size reduction of the different units, especially tubing diameters; and (2) the design and commercialization of the so-called ‘integrated microconduits’, which include in a single plastic block smaller than a cigarette box almost the whole FI manifold (injection, reaction/transport, and detection units).

Notwithstanding the fact that FI is a relatively recent continuous-flow technique (the earliest reference to it appeared in 1975), it has gained widespread use, especially in the last decade, in which it has been the subject of a number of research papers, reviews, and monographs. However, despite its immense potential for routine analysis, FI has not yet been as widely commercialized or as greatly used as it deserves.

## Signal and Signal Measurement

When recorded, FI systems provide a transient signal, acquired by a detector on passage of the dispersed sample zone, which is similar to the signals provided by other analytical techniques such as chemiluminescence and electrothermal vaporization–atomic absorption spectrometry. A typical signal–time recording includes an abrupt rise from the baseline (see **Figure 2A**, in which the timescale is expanded). The most typical FI measurement is the peak height. However, gradient dilution techniques based on selecting a suitable element of the dispersed zone (usually at times longer than that of the maximum) are also used to obtain complementary information. The timescale of FI as used for routine monitoring is narrower and the signals obtained are similar to those shown in **Figure 2B**. These signals are yielded (1) by successive injections performed in order to construct a calibration graph (three injections per standard) and (2) by insertion of four identical microvolumes of sample. As can be seen, the reproducibility achieved is quite high, despite the typically kinetic, physical, and chemical, nature of the FI technique.



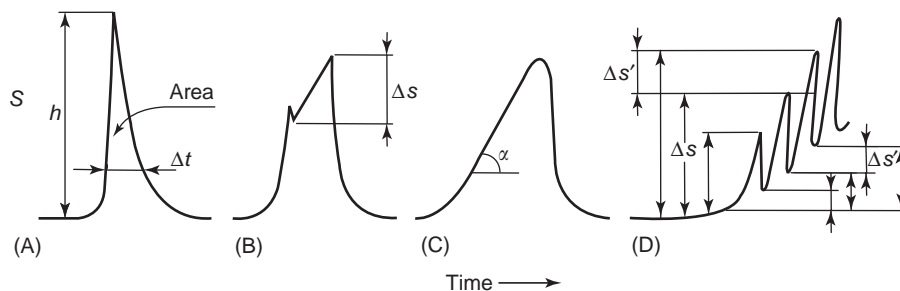
**Figure 2** Types of transient signals yielded by an FI system: (A) expanded conventional recording and measurement of the height at the maximum and at a time longer than that of the maximum (dilution mode); (B) usual recording for routine analyses (low chart speed).

The wide variety of FI applications developed so far provides a correspondingly wide variety of signals, the most outstanding of which are shown in **Figure 3**. The measurement parameters used in this context are also quite varied. Thus, in addition to the height, the area of an ordinary FI peak (**Figure 3A**) is also often related to the concentration, as in LC. The width at a given signal level (in time units) can also be related to the logarithm of the analyte concentration and permits calibrations to be performed at different sensitivity levels suited to the sample concentration, which is of great use to large-scale routine analyses.

The signals typically obtained by halting the flow when the reactant plug reaches the flow cell of a nondestructive detector (stopped-flow methods) allow signal increments ( $\Delta S$ ) to be measured throughout the standstill period and hence measurements of reaction rate to be obtained (**Figure 3B**). Because of their relative, rather than absolute, nature, such measurements are of particular use with catalytic methods employing dissolved enzymes and in avoiding interferences from sample matrices; however, they are only effective with slow analytical reactions (half-lives between 10 s and a few minutes).

The signals provided by retention(reaction)–detection systems are somewhat different from typical FI signals (**Figure 3C**). Retention is followed by elution and the signal includes an ordinary maximum that can be used to obtain the desired information. In addition, the slope of the rising portion of the signal is directly related to the analyte concentration.

FI multippeak recordings are provided by a variety of configurations including open–closed systems and those in which the iterative change of the flow direction causes the sample–reagent plug to pass many times through a nondestructive detector. Such signals provide richer information (**Figure 3D**). The sum of the signals from various peaks and the monitoring of the minima enables the calibration range be scaled up



**Figure 3** Type of measurements and recordings in FI systems: (A) peak height, peak area, or peak width at a preset height; (B) signal increment during the stop interval in the stopped-flow mode; (C) signal from flow-through sensors (measurement of the slope of the rising portion of the curve); (D) multippeak recordings from an FI system with iterative passage of the reactant plug through the detection point showing the different types of measurements. For details, see text.



or down, respectively, to suit each sample by using appropriate software and the same set of standards.

## FI versus LC, Segmented Continuous-Flow Analysis, and Sequential-Flow Analysis

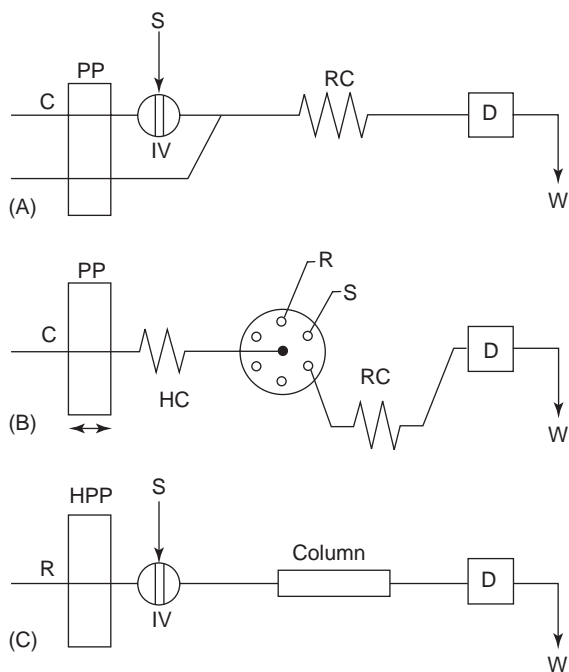
FI has some aspects in common with other dynamic techniques such as segmented continuous-flow systems (SCF), LC, and sequential-flow injection analysis (SIA). **Figure 4** highlights the similarities and differences between these techniques.

The most ostensible difference between SCF and FI lies in the absence of air bubbles in the latter, which are used in SCF to prevent carryover. This shortcoming is overcome in FI by the special geometric (small diameter of the tubing) and dynamic conditions (laminar flow). The absence of air bubbles in FI results in major advantages such as the ability to introduce reproducible sample volumes into the system by injection, reproducible handling with little pulsation, no requirement for predetector bubble separator, the possibility of reproducibly controlling the flow rate of the carrier and reagent streams (including preset intermittent stop and go periods), and that of including improved separation steps. Other outstanding differences lie in the lack of physical and chemical equilibria in FI, which enables kinetic

discrimination, measurement of transient or metastable species, and, in general, the development of methods unfeasible by SCF.

In relation to LC, FI has the following significant similarities and differences: the similarities include miniaturization capability, injection, unsegmented flow, small sample volume, signal profile, and the fact that both lack the characteristic lag phase of SFC. The most important difference between the two techniques probably lies on their foundation: in LC there is always an interface that affects the separation of a mixture of substances passing through the column. Other differences include the working pressure, which is low in FI, and sufficiently high in LC for the pump to overcome the dynamic resistance of the column. Also, the scope of application of the two techniques is very different since the basic aim of an LC instrument is to separate and analyse a complex mixture of substances, whereas the aim of FI is the rapid determination of one or two species in a large number of samples.

SIA, conceived as a single pump, single valve (frequently selection valve), single channel technique, was introduced as a feasible and mechanically simpler alternative to FI. Although SIA is based on the same principles as FI, the differences between these techniques are remarkable, especially when considering the dispersion patterns inside the two systems. In FI, reagents are normally added to the injected sample zone through confluence points, resulting in a concentration gradient of analyte within a constant background of reagent. In SIA, an initial sharp boundary is formed between the adjacent sample/reagent zones stacked in the holding coil; even after flow reversal, only a partial overlap of the analyte and reagent zones is achieved in these systems.



**Figure 4** Schemes of the common (continuous line) and uncommon parts (dotted line) of an FI manifold (A) as compared with an SIA system (B) and a chromatograph (C).

## Dispersion in FI Systems

### Types of Dispersion and Operational Modes

Theoretical studies on the FI technique have so far been aimed at defining the dispersion and dilution undergone by a sample plug on injection into an FI system, which ultimately defines the profile of the transient signal obtained and coincides with the dispersion at the detection point (flow cell). A number of authors have sought mathematical relationships between the basic parameters of an FI peak (travel or residence time, coordinates of the maximum, width) and the experimental features of the FI system, both of dynamic (flow rate) and of geometric nature (length, diameters of tubing, connectors, etc.). Despite of the endeavors made in this respect, no generic expressions for such potential relationships



have so far been obtained. A detailed description of the mathematical aspects of FI is provided in the monographs by Ruzicka and Hansen, and by Valcárcel and Luque de Castro. The most solid conclusions in this respect are summarized below:

1. Most FI systems operate under laminar-flow conditions, even though the limited carry-over or rather lack of it between successively injected samples was formerly ascribed to the occurrence of a turbulent regime.
2. The dispersion of the injected sample is a function, among others, of the time taken in the transport process. Initially, dilution is chiefly convective in nature; later, it becomes convective–diffusional and, finally, takes on purely diffusional character. The intermediate situation (i.e., convective–diffusional transport) is by far the most common in FI.
3. Radial dispersion contributes more significantly to the peak profile (dilution of the sample plug) than does axial dispersion. This type of fluid motion results in a washout effect accounting for the small carry-over between successively injected samples. This advantageous phenomenon in turn is a result of the use of low flow rates and small tubing bores, and leads to decreased peak widths and hence increased sample throughputs.
4. A number of theoretical approximations have been applied to the typical non-Gaussian shape of FI peaks, such as Taylor's and the tanks-in-series and mixing-chambers models. However, the most accurate description of the behavior of a solute injected into an FI system is provided by the so-called 'general model', which is based on a general equation describing convective–diffusional transport and takes account both axial and radial concentration gradients, the contribution of convective transport, and the parabolic shape of the velocity profile of a laminar-flow regime:

$$D \left[ \frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial r^2} + \frac{1}{r} \frac{\partial^2 C}{\partial r} \right] = \frac{\partial C}{\partial t} + u_0 \left[ 1 - \frac{r^2}{a^2} \right] \frac{\partial C}{\partial x}$$

where  $D$  is the molecular diffusion coefficient ( $\text{cm}^2 \text{s}^{-1}$ ),  $C$  is the analyte concentration,  $t$  is the time (s),  $x$  the distance along the tube,  $r$  is the tube radius (cm), and  $u_0$  is the linear flow velocity ( $\text{cm s}^{-1}$ ). This equation is far from easy to solve and requires the use of approximations that vary from author to author.

### Factors Affecting Dispersion

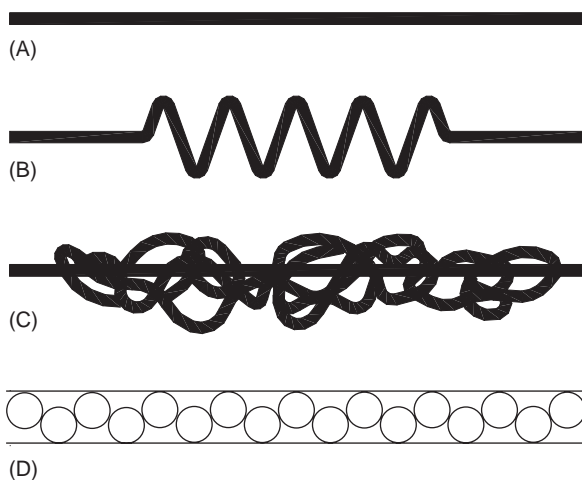
Dispersion in an FI manifold can be roughly modified by acting on the (1) dynamic or (2) geometric

variables, or (3) by modifying the kinetics of the biochemical reactions involved.

1. Provided other variables are constant, dispersion increases with decreased flow rate, which also increases the residence time of the plug in the continuous-flow system. Conversely, increased flow rates in the normal working range of FI ( $0.5\text{--}5 \text{ ml min}^{-1}$ ) decrease dispersion concurrently with increased sensitivity, as the height of the peak increases.
2. An increase in the reactor length has a similar effect on dispersion to that of decrease in the flow rate because both parameters have a similar influence on the residence time. However, the dispersion can be greatly altered by changing the geometric form of the reactor. Possible reactor geometries are shown in **Figure 5**. Dispersion decreases from (A) to (D) as an additional 'secondary flow' occurs which increases the radial dispersion in nonstraight tubes. This effect increases with the length traveled by the plug, and as a result the sensitivity also increases. Also, decreasing the inner diameter of the tube results in decreasing dispersion because of the increasing occurrence of secondary flow. On the other hand, inclusion of a mini-mixing chamber in a continuous-flow system increases dispersion and increases the difference between the height of the signal recorded in the absence of dispersion (obtained by aspirating the sample instead of the carrier) and the maximum of the FI peak:

$$D_{\text{max}} = h_0/h_{\text{max}}$$

By definition,  $D$  will always be greater than unity, unless the system incorporated facilities for preconcentration. In routine FI applications,  $D_{\text{max}}$  ranges between 1 and 15. As  $D$  increases, the resulting FI



**Figure 5** Reactor geometries most frequently used in FI: (A) straight open tube; (B) coiled tube; (C) knotted reactor; and (D) single bead string reactor.

peak becomes flatter, even though the enclosed area remains constant. The larger the  $D$  is, the lower will be the sensitivity achieved if the maximum height is used as measurement parameter. The sample throughput also decreases with increasing  $D$ , which results in increased peak width and hence in a higher risk of peak overlap. This is a clear symptom of undesirable carryover between samples.

## General FI Modes

Single-channel manifolds are used when no chemical reactions are required prior to detection (e.g., use of electrochemical or atomic spectroscopic detectors). However, even in these cases, some treatments or conditioning makes it necessary to use additional channels. Thus, from the most basic FI system (Figure 1) setups can vary widely depending on the chemical system involved and/or the particular aim. Some of the more elementary FI modes deriving from the basic scheme are illustrated in Figure 6 and described below.

### Multichannel Manifolds

Two channels merging before the injection system can be required to generate a reagent in *status nascenti* or to obtain a reaction medium in which the reagent is unstable (Figure 6A). Other additional merging channels can be required for masking purposes and/or to sequentially develop the different steps of an overall reaction (Figure 6B).

### Reversed FI

The roles of sample and reagent can be exchanged in an FI manifold (Figure 6C). This reversal is particularly useful when the sample is abundant and inexpensive (e.g., drinking- or wastewater), as it saves reagent. The sequential determination of several analytes can be accomplished by inserting into the sample channel as many injection valves as required, through which selective reagents for different analytes can be inserted into the sample stream (Figure 6D).

### Merging-Zone Mode

This approach allows saving both the sample and the reagent. A dual-injection system is used to inject the sample and the reagent simultaneously into two carriers or additional reagent streams (Figure 6E). The length ratio of the channels between the injection and merging points give rise to two merging-zone modes: the symmetric mode, when both channels have the same length and both injected plugs merge completely (Figure 6F); and the asymmetric mode (the

most versatile), which requires different sample and reagent volumes in addition to different lengths of the channels between the injection and merging points in order to ensure mixing of the sample with the carrier, or an additional reagent stream on one of its ends with the injected reagent on the other (Figure 6G). In this way, it is possible to develop simultaneous determinations of two analytes.

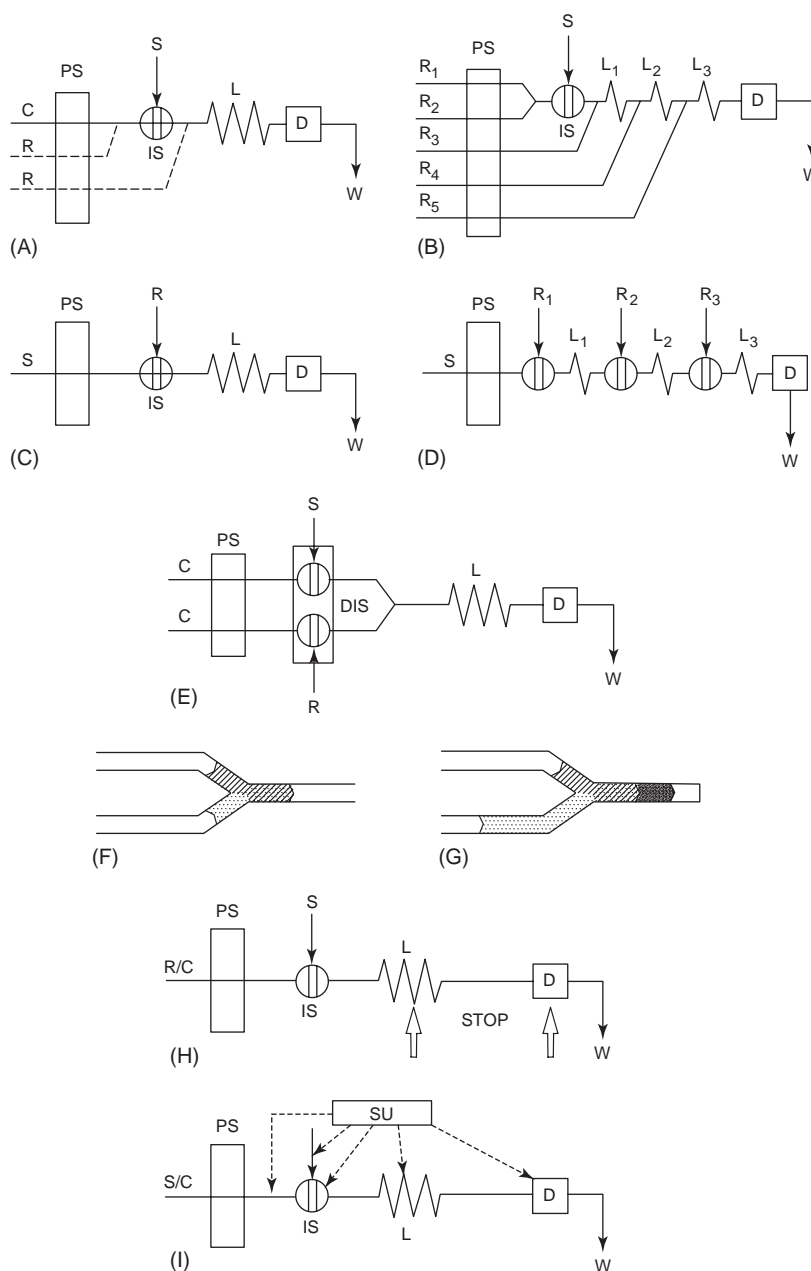
### Stopped-Flow Mode

By synchronizing the propulsion and the injection systems through a timer or microcomputer (Figure 6H), the flow can be halted when the injected plug reaches a preset zone of the manifold. Most commonly, the injected plug is stopped at the flow cell in order to monitor the development of a biochemical reaction between the analyte in the sample and some reagent contained in the carrier or merged with the main channel prior to the detection point. In this way, it is possible to perform reaction rate measurements. However, the flow can also be halted while the reacting plug is in the reactor, thus allowing a longer sample–reagent contact time prior to measurement.

### Inline Coupling with Nonchromatographic Continuous Separation Techniques

Enhanced sensitivity (through a preconcentration step) and/or selectivity (by removing interferences from the sample matrix or, less commonly, from reagents) can be achieved by using a nonchromatographic continuous separation technique/flow injection arrangement. Nonchromatographic separation techniques involving all types of interfaces (gas–liquid, liquid–liquid, and solid–liquid) can be coupled inline to FI manifolds at locations depending on the pursued aim (Figure 6I): placing the separation unit in the main channel prior to injection is intended to effect reagent purification, while placement in the sample stream, in the injection system or between this and the detector, is aimed at preconcentration and/or interference removal. Finally, if the separation unit is in the flow cell, integration of separation and detection usually provides increased sensitivity and selectivity by *in situ* concentration and kinetic monitoring of the separation process, respectively.

Nonchromatographic continuous separation techniques involving gas–liquid interfaces (e.g., gas diffusion, hydride generation, pervaporation) prior to detection by an atomic spectrometer have frequently been coupled to FI manifolds to develop excellent methods of interest, mainly to the clinical and industrial fields.



**Figure 6** Elementary FI modes: (A) and (B) multichannel manifolds; (C) and (D) reversed mode; (E) merging-zone manifold; symmetric (F) and unsymmetric (G) merging patterns; (H) stopped-flow mode; (I) possible locations of a separation unit in an FI manifold. C denotes carrier, R reagent, PS propulsion system, S sample, IS injection system, L reactor, D detector, W waste, DIS dual-injection system, and SU separation unit.

Liquid–liquid interfaces have been involved in FI configurations mainly through dialysis and extraction. Excellent yields are obtained through the separation process as the FI manifold can be altered to match the performance of the method concerned to the features of the sample and/or analyte to be determined.

Ion exchange and sorption systems in general are the most common types of nonchromatographic

continuous separation techniques involving liquid–solid interfaces that have been coupled to FI. In addition to the dramatic enhancement in sensitivity and/or selectivity achieved inline by using microcolumns of an active material, the ease with which they can be included in FI systems has fostered their usage in virtually all fields of applied analysis. Also, unconventional liquid–solid separation techniques such as stripping techniques have been successfully

implemented in FI systems as their development in a continuous fashion circumvents many of the shortcomings involved in batch stripping methods. Typically, manual liquid–solid separation processes, such as those involving precipitation/filtration/dissolution, have also been implemented in continuous FI manifolds, thus allowing automation of these conventional, tedious procedures.

### Online Coupling with Chromatographic Separation Techniques and Capillary Electrophoresis

The coupling of an FI system to a liquid chromatograph in a pre- or postcolumn arrangement saves reagents when derivatizing reactions are required either for facilitating the chromatographic separation or for detection. In addition, the sensitivity can be matched to particular requirements.

The coupling of an FI manifold to capillary electrophoresis equipment enables sample preparation prior to introduction into the capillary. FI preconcentration by liquid–solid steps, dialysis, and liquid–liquid separations have been developed in a precapillary position.

### Basic Analytical Features of FI

A comparison of the main analytical features of FI with those of its sequential injection and manual counterparts highlights its advantages and disadvantages.

Two typical features of FI methods are: (1) the reaction time is rather short, so equilibrium is not attained; and (2) the physical dispersion or dilution of the sample in the carrier results in a signal of lower intensity than that corresponding to the undiluted plug. However, the dispersion pattern inside an FI system results in a concentration gradient of analyte within a constant background of reagent; in an SI system the initial sharp boundary formed between the adjacent sample/reagent zones stacked in the holding coil, and even after the flow reversal, only a partial overlap of analyte and reagent is achieved. This fact gives rise to lower dispersion, but it is also a source of inaccuracy, especially when the sample is contaminated with interferents.

The lower sensitivity of FI is amply offset by its versatility, which allows sensitivity to be decreased or enhanced at will. A decrease in sensitivity can be achieved by decreasing the flow rate or injected volume, increasing the length of the reactor or using auxiliary channels, using under-optimal chemical conditions, or performing measurements at points far from the maximum of the FI peak. The different ways of enhancing sensitivity in FI can be classified into three groups: (1) by altering the FI manifold or

the working conditions (using large sample volumes, increasing the residence time without increasing the dispersion by stopping the flow in the reactor, reducing tubing diameters, using solid reagents, or introduction of membrane-based reagents or using micellar media); (2) by inline coupling with a separation/preconcentration technique (any of those described above) and by using integrated separation/detection approaches; and (3) by using the so-called ‘scale expansion techniques’, based on chemometrics, which allow calibration curves to be expanded at either end.

Enhanced selectivity is usually achieved when either a manual or air-segmented method is adapted to FI as a result of its kinetic nature: unwanted reactions do not develop to a significant extent over the short residence times of FI methods. Another positive aspect of FI is that the use of a closed system inherent in the technique prevents reagents and sample from coming into contact with the atmosphere, thereby avoiding undesirable reactions with air components.

Although FI involves kinetic, physical, and chemical aspects which may *a priori* lead readers to conclude that results will not be too good in terms of precision, the reproducibility is comparable to, or better than, that of SI and manual methods, which is probably one of the reasons of the success of this methodology. This feature makes FI methods reliable for use in intercomparison studies.

Rapidity is one of the most outstanding features of FI methods, which are usually much faster than their manual and automatic counterparts. The time spent in tuning an FI system is similar to that of an SI system because the 5–10 min required for injection depends on the time required by the propulsion system for regulating the flow. As a rule, the start-up time is dictated by the warm-up time of the detector. However, the rapidity of FI is apparent in the sample rate or frequency, which is higher than that of manual and SI. Finally, when a multisyringe approach is used, this is limited by the necessity of reloading the syringes after each cycle. Nevertheless, the sampling frequency of FI methods strongly depends on the kinetics of the chemical reaction as an acceptable amount of the monitored product is needed to ensure appropriate sensitivity.

Simplicity is also a salient feature of both FI and SI, and arises from their simple foundation: the experimental arrangements are simple, in clear contrast to the new analytical methodologies; and they are systems of easy design and operation.

As FI and SI share the same principles (precise sample introduction, controlled dispersion, and reproducible timing), the costs are very similar and lower than those of commercial analyzers providing a similar (or even poorer) performance. For the same reason, the

versatility provided by both is according to the modular nature that allows them to meet a wide variety of analytical needs. Nevertheless, the highest versatility of FI is clearly demonstrated in the widest number of modes it is capable of implementing, most of them are unable to be implemented by the SI technique.

See also: **Flow Analysis:** Overview. **Flow Injection Analysis:** Instrumentation; Detection Techniques; Environmental and Agricultural Applications; Clinical and Pharmaceutical Applications; Industrial Applications.

## Further Reading

Burguera JL and Burguera M (2001) Flow injection-electrothermal atomic-absorption spectrometry configurations: recent developments and trends. *Spectrochimica Acta B* 56(10): 1801–1829.

Karlberg B and Pacey GE (1989) *Flow Injection Analysis: A Practical Guide*. Amsterdam: Elsevier.

Kuban P and Karlberg B (1998) Interfacing of flow-injection pretreatment systems with capillary electrophoresis. *Trends in Analytical Chemistry* 17(1): 34–41.

Luque de Castro MD and Luque-García JL (2002) *Acceleration and Automation of Solid Sample Treatment*. Amsterdam: Elsevier.

Matsuoka S and Yoshimura K (1995) Application of solid-phase spectrometry to flow analysis. *Journal of Flow Injection Analysis* 12(1): 8–22.

Ruzicka J and Hansen RH (1988) *Flow Injection Analysis*. New York: Wiley-Interscience.

Valcárcel M and Luque de Castro MD (1987) *Flow Injection Analysis: Principles and Applications*. Chichester: Ellis Horwood.

Valcárcel M and Luque de Castro MD (1990) *Non-Chromatographic Continuous Separation Techniques*. Cambridge: Royal Society of Chemistry.

## Instrumentation

**M D Luque de Castro**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

One of the key features of flow injection (FI), in addition to its versatility and simplicity, is its low cost. The cost is low partly because the working pressures used in FI methods are usually lower than 70 kPa (~10 psi), which means that the different units making up the FI manifolds are easy to design and inexpensive to assemble. Although several FI instruments are already commercially available, the individual basic and dedicated components needed to construct FI equipment can be purchased separately. This is especially useful in those cases where innovations are intended.

The basic units of an FI manifold – the propulsion, reaction, and detection/data acquisition systems – are described in detail together with the most salient types and modifications of the propulsion, injection, and reaction units. Details are also given of other units that can sometimes be dispensed with, but are commonly used in the implementation of a variety of FI methods, especially in connection with rather complex samples. These special components may be

separation/preconcentration units (gas diffusers, pervaporators, dialyzers, liquid–liquid extractors, etc.), mini-mixing chambers, special valves, etc., in addition to the elements required for full or partial automation of an FI analyzer, namely samplers and interfaces.

## Basic Units

Three essential components of an FI system are shown in **Figure 1**, which also shows different possible modifications, simple or complex, that could be made depending on the problem to be solved. The fourth essential unit – the detector – is not part of FI manifolds, as it can be coupled to any type of molecular detector – provided the appropriate flow cell is available – or atomic detector.

## Propulsion Units

One of the most outstanding features of FI – versatility – is partially supported on the propelling unit. The requirements for an ideal functioning of this unit in FI may be summarized as follows:

1. Reproducible flow rates on a short- (hours) or long-term (days) basis, in order to provide a predictable residence time and a constant dispersion (directly related to the flow rate) throughout the manifold.



versatility provided by both is according to the modular nature that allows them to meet a wide variety of analytical needs. Nevertheless, the highest versatility of FI is clearly demonstrated in the widest number of modes it is capable of implementing, most of them are unable to be implemented by the SI technique.

See also: **Flow Analysis:** Overview. **Flow Injection Analysis:** Instrumentation; Detection Techniques; Environmental and Agricultural Applications; Clinical and Pharmaceutical Applications; Industrial Applications.

## Further Reading

Burguera JL and Burguera M (2001) Flow injection-electrothermal atomic-absorption spectrometry configurations: recent developments and trends. *Spectrochimica Acta B* 56(10): 1801–1829.

Karlberg B and Pacey GE (1989) *Flow Injection Analysis: A Practical Guide*. Amsterdam: Elsevier.

Kuban P and Karlberg B (1998) Interfacing of flow-injection pretreatment systems with capillary electrophoresis. *Trends in Analytical Chemistry* 17(1): 34–41.

Luque de Castro MD and Luque-García JL (2002) *Acceleration and Automation of Solid Sample Treatment*. Amsterdam: Elsevier.

Matsuoka S and Yoshimura K (1995) Application of solid-phase spectrometry to flow analysis. *Journal of Flow Injection Analysis* 12(1): 8–22.

Ruzicka J and Hansen RH (1988) *Flow Injection Analysis*. New York: Wiley-Interscience.

Valcárcel M and Luque de Castro MD (1987) *Flow Injection Analysis: Principles and Applications*. Chichester: Ellis Horwood.

Valcárcel M and Luque de Castro MD (1990) *Non-Chromatographic Continuous Separation Techniques*. Cambridge: Royal Society of Chemistry.

## Instrumentation

**M D Luque de Castro**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

One of the key features of flow injection (FI), in addition to its versatility and simplicity, is its low cost. The cost is low partly because the working pressures used in FI methods are usually lower than 70 kPa (~10 psi), which means that the different units making up the FI manifolds are easy to design and inexpensive to assemble. Although several FI instruments are already commercially available, the individual basic and dedicated components needed to construct FI equipment can be purchased separately. This is especially useful in those cases where innovations are intended.

The basic units of an FI manifold – the propulsion, reaction, and detection/data acquisition systems – are described in detail together with the most salient types and modifications of the propulsion, injection, and reaction units. Details are also given of other units that can sometimes be dispensed with, but are commonly used in the implementation of a variety of FI methods, especially in connection with rather complex samples. These special components may be

separation/preconcentration units (gas diffusers, pervaporators, dialyzers, liquid–liquid extractors, etc.), mini-mixing chambers, special valves, etc., in addition to the elements required for full or partial automation of an FI analyzer, namely samplers and interfaces.

## Basic Units

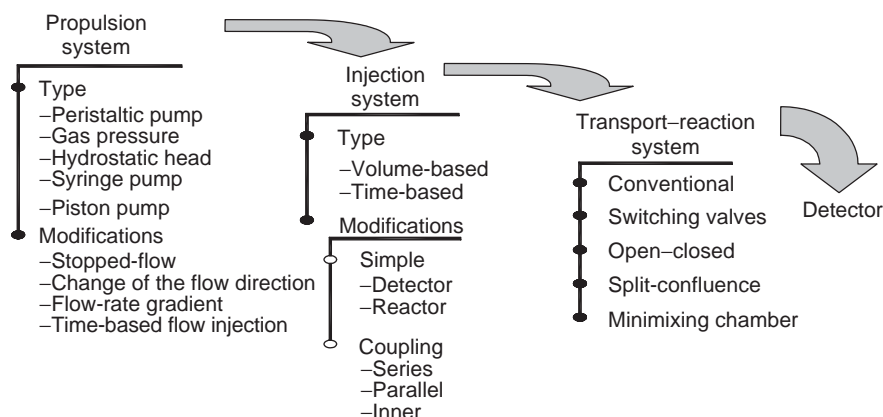
Three essential components of an FI system are shown in **Figure 1**, which also shows different possible modifications, simple or complex, that could be made depending on the problem to be solved. The fourth essential unit – the detector – is not part of FI manifolds, as it can be coupled to any type of molecular detector – provided the appropriate flow cell is available – or atomic detector.

## Propulsion Units

One of the most outstanding features of FI – versatility – is partially supported on the propelling unit. The requirements for an ideal functioning of this unit in FI may be summarized as follows:

1. Reproducible flow rates on a short- (hours) or long-term (days) basis, in order to provide a predictable residence time and a constant dispersion (directly related to the flow rate) throughout the manifold.





**Figure 1** Basic units of a flow injection manifold and possible modifications.

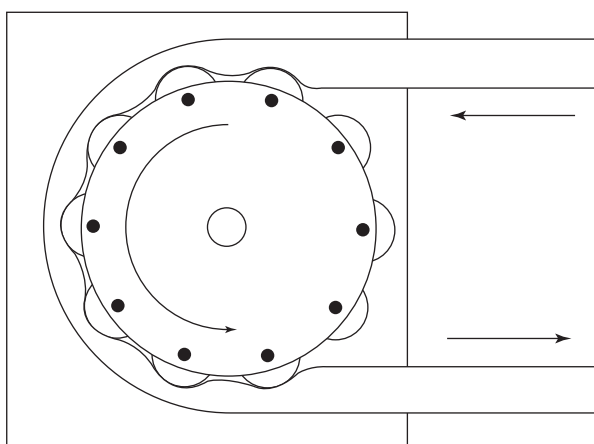
2. Multichannel capability for providing at least four parallel streams to ensure versatility.
3. Pulse-free flow that is not completely fulfilled in commercially available propelling systems but can be avoided by using a suitable attenuator.
4. Resistance to aggressive reagents and solvents.
5. Readily adjustable flow rates, and low initial investment and running cost.

A large variety of these devices have been used but, unfortunately, none of them meets all the requirements listed above. One way to minimize the limitations of the propelling systems is by computer-controlled flow programming. Also, it is possible to reach ideal conditions under specific circumstances where some of the requirements are not critical, or by a combination of different propelling units.

A distinctive feature of the propelling devices in FI systems is the low working pressure required for obtaining suitable flow rates. FI systems usually operate at pressures lower than 10 psi. When higher working pressures are required for special applications, high-performance liquid chromatography (HPLC) pumps can be used.

The usual location of the propelling unit is before the injection system. However, in some FI systems the propelling unit is placed after the detection system, to probably achieve improved pulse attenuation.

The general trend toward miniaturization in all branches of science can also be seen in propelling units. In addition to easier transportation of micro-propelling devices, especially those required in portable and submersible FI analyzers, other advantages afforded are low reagent and sample consumption, fast sample–reagent mixing, and decreased undesired broadening of the sample zone, thus increasing both sampling frequency and sensitivity. However,



**Figure 2** Details of the drum of a peristaltic pump showing the rollers squeezing the flexible tubing.

limitations such as power requirements and periodic replacement of some elements (tubes, usually) are needed when they are operating in a continuous mode.

The propelling systems normally used in FI are peristaltic, piston, syringe, and solenoid pumps, gravity-driven sources, electro-osmotic flow (EOF) devices, and osmotic pumps.

Peristaltic pumps are the ones most commonly used in FI. They consist of a motor-driven wheel with peripheral placed rollers and an adjustable compression cam (or band) that is squeezed against the rollers (**Figure 2**). The successive peristaltic movements of the tubes force various fluids through the tubes. The number of channels depends on the roller length and usually ranges from 1 to 16. Pump tubes are made of poly(vinyl chloride) (PVC), silicone, or any other similar plastic material. Two collars are usually glued near the end of each pump tube with a dual purpose: first, to serve as a point of attachment between the

harnesses of the pump and, second, to identify the inner diameter and hence the flow rate. Tygon<sup>®</sup> tubes are used to propel aqueous solutions and alcoholic solutions, wherein the alcohol content should not exceed 5% in order to ensure a long tube lifetime. Strongly acidic or basic solutions should be pumped through other tubes made from other materials. Thus, Solvaflex<sup>®</sup> yellow pump tubes are used to propel alcoholic solutions and some alkanes. Also, pump tubes made from Viton material are used to propel strong acidic solutions and some chlorinated organic solvents.

According to the preferable usage pattern of peristaltic pumps in FI, major changes have been introduced in them. Peristaltic pumps usually work in one direction and at a constant flow rate. Any change in some characteristic requires appropriate programming. By increasing computerization in the laboratory environment, new FI methodologies have evolved on pump functioning: stopped-flow, iterative change of the flow direction, flow variation as a function of time, time-based injection, etc. Start-and-stop, variable pump speed, and forward-backward flow direction are some modes in which these devices work. Also, improvements in the performance of peristaltic pumps have been aimed at exploiting flow pulsation for increasing sensitivity in voltammetric detection and improving both sensitivity and selectivity in atomic detection. Strategies for injection/pumping synchronization have also been exploited both for improving measurement precision and providing scope for new FI approaches.

In addition to peristaltic pumps, other propelling devices have been used in FI. Syringe pumps are durable devices and capable of generating a pulse-free, perfectly controlled flow. However, they are too expensive to be used in multireagent systems and their less expensive versions (such as infusion pumps) are not designed to stop or reverse the flow precisely enough. Piston pumps operate by moving the plunger (piston) back and forth in a pump head, and using one-way directional check valves (usually ball-style valves). The movements produced by the rotation of a circular cam provide a pipeline flow-rate diagram and, therefore, provide changes in pumping speed. This makes it less readily adapted to FI requirements, as they are usually designed for use in HPLC. Solenoid pumps operate by the same mechanism as piston pumps with the only difference that the movements are produced by a diaphragm, usually made from polyether ether ketone (PEEK) and Viton. Some miniaturized solenoid pumps can replace peristaltic pumps for *in situ* applications to minimize size, weight, maintenance, and power requirements. Gas-pressurized reservoirs are simple, inexpensive, do

not contain any moving parts, and produce pulseless flow. However, it is very difficult to adjust the flow rate in a multiple line manifold. Gravity-based propelling units rely on the difference in height between the reagent and/or carrier reservoir and the FI line. A pulse-free flow is obtained and the different heights determine the flow rate. The most significant shortcoming is the need for maintaining a constant level of the liquid(s) in the reservoir(s). EOF-based pumps are made of materials whose surfaces can carry a net charge when in contact with aqueous solutions. The oppositely charged counterions present in the aqueous solutions are attracted by these surfaces. Application of an external electric field results in a net migration of these counterions toward the oppositely charged electrodes. The major application of EOF propelling systems is in capillary dynamic systems. Analyte enrichment, limited dispersion, and removal of matrix interferences are remarkable features of EOF systems. Osmotic pumps provide a flow by the osmotic pressure created across a rigid semipermeable membrane separating a saturated salt solution from a solution of lower salinity. The diffusion of water through the membrane results in a net flow from the diluted to more saline solution. Temperature dependence and low flow rates reduce the usage of these propelling units in FI.

### Injection Systems

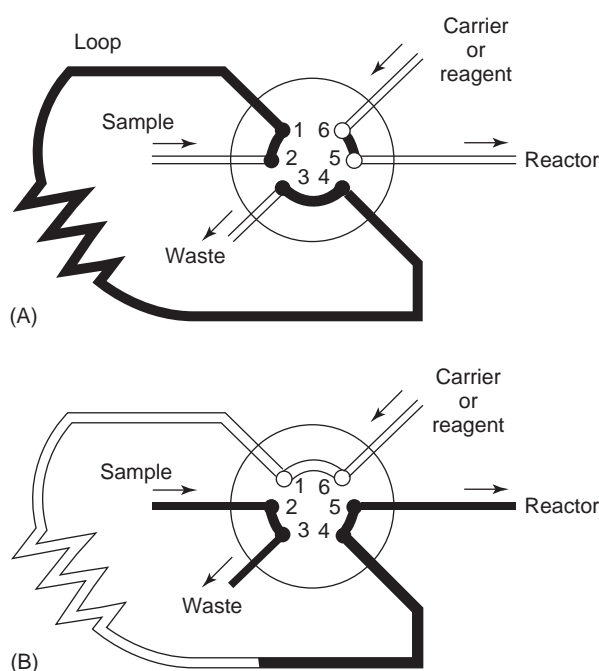
The injection system is intended to insert a well-defined plug into the flowing carrier stream and should meet a number of requirements imposed by the intrinsic features of the FI technique, namely:

1. It should introduce accurately measured volumes (usually within 25–100  $\mu\text{l}$ ) of a sample or reagent in a reproducible manner. The wider the volume range, the more versatile the system.
2. The sample should be incorporated into the carrier stream in such a way that no disturbances are caused to the latter.
3. It should allow for convenient and fast operation in order to provide a high sampling rate.

In addition, it should be controllable by an electric motor. The injection devices designed for this purpose can be divided into the following categories: volume-based injectors, time-based injectors, or a combination of these two. The first type of device is based on the typical entrapment of the sample solution into a geometrically well-defined volumetric cavity and subsequent transfer (injection) of the sample zone thus formed in the unsegmented carrier stream. The second category is based on the

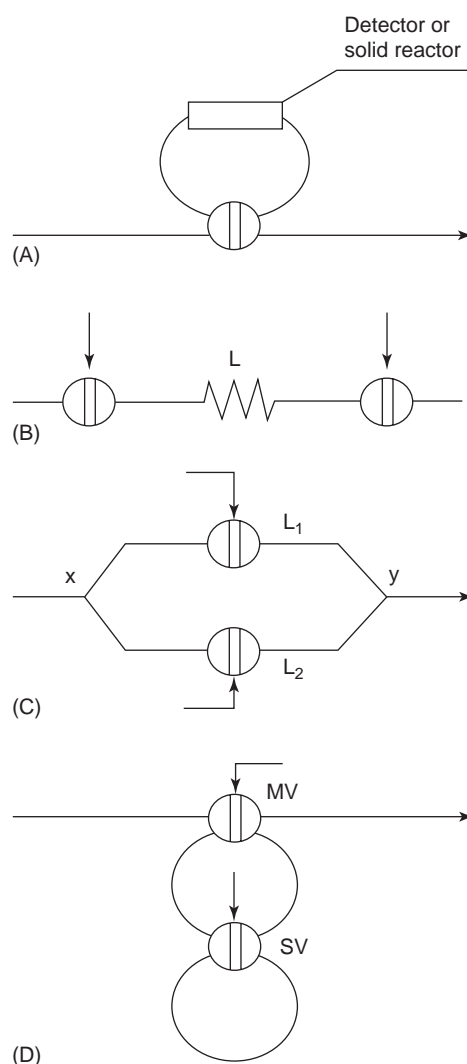
aspiration of the sample solution at a constant flow rate for a fixed period of time into a well-defined section of a flow-through channel from which the metered sample solution is injected into the carrier stream by a combination of hydrostatic and hydrodynamic forces. Whereas in time-based injection procedures a crucial role is the control of the injection volume by the selected time interval existing – and thus, the possibility of computer control – in volumetric methods, the sample volume is changed physically by changing the volume dimensions of the sample loop.

The injection unit first employed in FI was a syringe and a hypodermic needle. The injection system most frequently used now is a volume-based injection unit such as the typical six-port rotary valve shown in **Figure 3**. In the fill position, the sample (or reagents) enters the valve through port 2, fills the loop between ports 1 and 4, and goes to waste through port 3. Meanwhile, the carrier stream enters the valve through port 6 and passes out through port 5. In the inject position, port 6, through which the carrier enters the valve, is connected internally with port 1, sweeping the sample plug toward the detector through ports 4 and 5, while the sample flows into the valve through port 2 and is sent to waste through port 3. This type of valve gives high reproducibility in the injected volume, a wide range of variation in the loop volume, rapidity, and easy manual operation; in addition, it can be readily automated.



**Figure 3** Typical six-port rotary injection valve: (A) fill position and (B) inject position.

The facility for manipulating injection devices has multiplied their use, not only for such simple injection devices, but also for multiple injection approaches for fulfilling the different purposes of use. Some of the most common modifications and arrangements are shown in **Figure 4**. In addition to their use for single injection, their loops can hold the most varied flow devices (**Figure 4A**); they can be coupled in series (**Figure 4B**), in parallel (**Figure 4C**), and interiorly (**Figure 4D**). Even several of these couplings can be included in a single manifold, thus enlarging the number of steps to be developed in a single, cheap, and simple approach. The most salient of these approaches can be divided into two: use of a single injection valve and use of multiple injection valves, in the same manifold.



**Figure 4** Modifications of rotary valves: (A) inclusion of a flow cell or solid reactor in the loop; (B) serial and (C) parallel arrangements; and (D) internal coupling. L denotes the reactor, x and y the splitting and confluence points, respectively, and MV and SV the main and secondary valves, respectively.

When using a single injection valve, the most immediate use is the insertion of a preset liquid volume of sample into a carrier or reagent stream that is also a liquid. In the reverse FI mode, abundance of sample acts as a carrier where the reagent is injected. Also, there are possible FI configurations into which gaseous samples are inserted into liquid or vice versa. The insertion of this volume into a stream can be either time based or volume based (the latter being the most frequently used). Despite these general functions of the devices, other multiple functions of volume-based injection systems demonstrate their versatility. These functions are performed with the help of some devices working in a continuous fashion and located in the valve loop. These devices, inserted into the dynamic system, range from microcolumns packed with different materials to detector cells, and also include other continuous separation units such as dialyzers, gas-liquid separators, liquid-liquid extraction cells, filters, and pervaporation chambers.

The easy manipulation and arrangement capacity of single injection valves enable their coupling in different manners. In-parallel coupling enables sample-sample injection and sample-reagent injection facilitating the symmetrical or asymmetrical modes of the merging-zone approach. In-series coupling of these valves allows zone sampling and a variety of manifolds for simultaneous determinations. Finally, the interior coupling, the so-called nested loop approach, has demonstrated high versatility. Multi-determinations, calibration and speciation studies, construction of pH gradients, and blank-sample measurements are the more remarkable applications of this configuration.

Solenoid valves and commutators are among the volume-based injection systems less commonly used in FI. The latter have been used by their inventors – a Brazilian team – but they are not practical devices as they easily deteriorate.

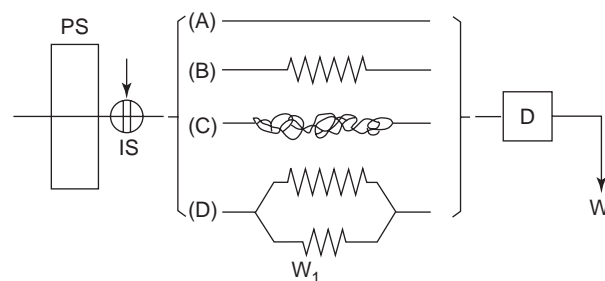
### Transport-Reaction Systems

In addition to its primary function of carrying the flowing stream along the manifold and allowing, where appropriate, a reaction to develop to a suitable extent, the transport and reaction system serves to link the different parts of the FI system. This component normally consists of small-bore tubes of internal diameter between 0.1 and 1.0 mm, but usually 0.3, 0.5, and 0.7 mm internal diameters for low, medium, and high dispersion requirements, respectively. Large bores are of little use as the remaining FI components are not designed to match them – tubing expansion and compression of the other components result in irregularities that

significantly affect the dispersion. On the other hand, bores smaller than 0.3 mm internal diameter are only used in the miniaturized FI mode. The reactor, a major component of the transport system, influences the residence time and profile of the sample plug, and is designed to meet the particular requirements of the system concerned. Three basic types of open reactors are used in FI (**Figure 5**): (1) straight tubes of variable length of diameter; (2) coiled tubes, which are pieces of tubing coiled helically around a rigid cylinder of the desired diameter; and (3) knotted tubes, which are much more effective than coiled or straight tubes as far as reduction of dispersion is concerned. However, producing knotted tubes requires skill, and the tubing must be flexible enough and have walls thick enough to prevent collapse or narrowing at tight bends.

The connectors used in FI setups serve the purpose of joining the tubes to one another and to the other parts of the system. There is a wide range of connectors available for each type of application, but basically they are either dual (linear or V-shaped), triple (T-, Y-, or W-shaped), or quadruple (usually in the shape of an arrowhead). The distance between the ends of the two tubes brought together by the linear dual connector should be as small as possible in order to minimize the dead volume, since this acts as a small mini-mixing chamber that results in increased peak tailing. Triple and quadruple connectors are commonly employed for both merging and splitting more than two streams (**D** in **Figure 5**). They make very useful devices for the development of reaction rate methods as the injected plug is split into several subplugs that reach the detector at different times.

The most suitable material for making reactor tubes is polytetrafluoroethylene (PTFE), which, in addition to being chemically resistant, absorbs less solutes on its surface. Polyethylene or polypropylene is inexpensive and easy to flange. As these three materials are highly permeable to oxygen, they pose serious problems in dealing with strong reductants or



**Figure 5** Types of simple open reactors: (A) straight, (B) coiled, (C) knotted, and (D) with splitting and confluence points.

air-catalyzed reactions. Connectors are normally PTFE or polymethylmethacrylate (PMMA).

## Other Units

One of the most interesting features of FI is its great versatility, i.e., the ease with which the manifold lends itself with alteration depending on the type of analyte, sample, or particular requirements involved. Alterations can involve one of the basic units described in the previous section or inclusion of units other than the basic ones, such as special valves, separation modules, packed reactors, or units required for the partial or full automation of the manifold.

## Special Valves

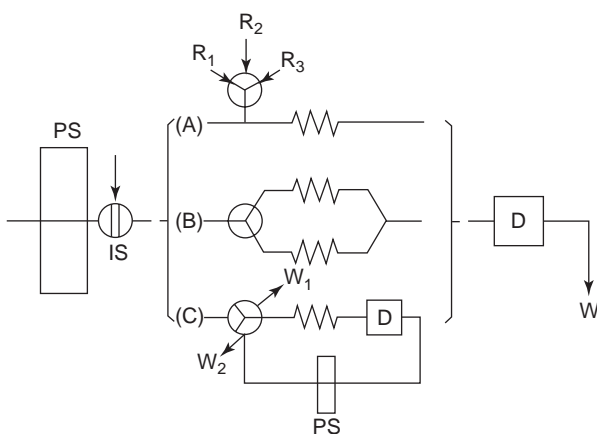
Two special types of valves are often used in FI: (1) special injection valves different from those described above, which allow special FI modes to be implemented; and (2) selecting or switching valves, which enable the streams traveling through the transport–reaction system to be handled as required.

A single eight-port injection valve can perform the same functions as two six-port valves as it can accommodate two loops. This type of valve is commonly used in gradient and sandwich techniques, where the sample is sandwiched between the carrier and a solution containing a reagent or a standard, and then returned to the carrier.

The heart of the sequential injection mode is a multiposition selection valve that is operated in synchronization with a pump. This valve can be connected to various sample and reagent reservoirs and when the pump is operated in reverse, aliquots are sequentially drawn into a holding coil, and, upon forward propulsion, the resultant stack of sample and reagents is dispersed into a zone of detectable product.

Switching valves provide great flexibility in the transport–reaction zones of FI configurations. Thus, the use of one such valve (A in Figure 6) allows selection of different reagents or reaction media for the sequential determination of as many analytes as desired, provided the reagents used are selective enough. A selection valve, placed at the point of splitting of two channels that are merged prior to the detector, allows the sample plug to be driven through one or the other channel, each of which normally includes a different reaction (B in Figure 6).

Open–closed systems are probably the most useful applications of switching valves. The valve keeps the system open in one of its positions, operating as a normal FI configuration (see C in Figure 6). Once the



**Figure 6** Reactors involving additional components such as valves for (A) selection of different reagents, (B) selection of different channels, and (C) closing and opening the circuit. PS denotes the propulsion system, IS the injection system, R the reagent, D the detector, and W the waste.

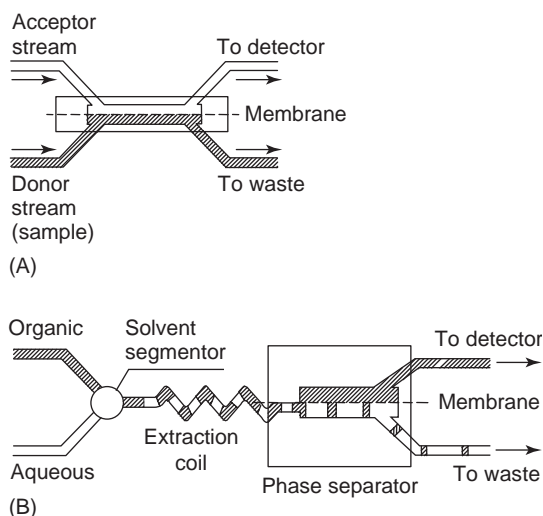
injected plug is inside, the valve is switched and the system is closed. The plug is then circulated through the manifold, which includes an ordinary detector and an auxiliary propulsion system, and is passed many times through the detection point, thereby yielding a multipeak recording from which large amount of information of theoretical and practical interests can be drawn. Some applications of these systems include determination of reaction orders and rate constants, stoichiometries, multicomponents, etc.

Mixing chambers are used when a high degree of mixing (merging point or unstirred chamber) or complete homogenization of sample and reagent (well-stirred sample) is required, which is usually the case when using an electrochemical detector or when the carrier and sample are considerably different in viscosity, relative density, temperature, or detergent content. Mixing chambers are commonly used in FI titrations.

## Separation Modules

Separation steps are quite often included in the analytical process in order to increase the selectivity and sensitivity of the overall determinations by removing undesirable interferents and preconcentrating the analyte(s), respectively. By appropriately coupling the separation technique and an FI manifold, separation/preconcentration processes can be performed automatically. The major separation techniques that are coupled online with FI manifolds are discussed, and the basic units required in each case are described.





**Figure 7** Continuous separation modules: (A) gas diffuser or dialyzer and (B) liquid-liquid extractor with extraction of the analyte(s) into the organic phase.

Figure 7A shows the scheme of a typical gas diffuser for use in FI, whether a chemical reaction is needed from the gas (e.g., carbon dioxide, ammonia, sulfur dioxide, or hydrogen sulfide) or the gas occurs as such in the sample (e.g., nitrogen oxides,  $\text{ClO}_2$ , ethanol, or ozone). The unit comprises a sandwich cell consisting of two blocks (PMMA or PTFE) with identical internally engraved conduits that make up an inner chamber, the geometry and dimensions of which vary from model to model. A membrane (usually Teflon) is placed between the two blocks, which must be joined tightly in order to avoid leakage. Each microconduit has two holes at its ends that connect it with the manifold tubing. The dimensions of the conduits engraved in the blocks vary over the following ranges: 0.1–0.3 mm depth, 1–3 mm width, and 3–10 cm length. Tubular cells including PTFE tubes and membranes are less commonly used for this purpose. The separation efficiency of the gas-diffusion process can be improved by including a derivatizing reaction in the acceptor stream in order to effect continuous removal of the analytes from the vicinity of the membrane, thereby also creating a useful concentration gradient.

In a dialysis process, separation of the different solutes relies on the difference between the rates at which they are transferred across a semipermeable membrane located between the two liquid phases and across which mass transfer takes place. The type of the membrane is the key difference between gas-diffusion and dialysis units; both use identical modules. Dialyzers are normally placed in the transport-reaction zone, and they are also occasionally positioned in the loop of the injection valve for

increasing dialysis efficiency. In such instances, the acceptor stream remains static or is recirculated, as required. Dialyzers accommodated in flow cells are not intended to effect the separation proper, but to facilitate the introduction of a reagent in order to accomplish integrated reaction and detection. On the other hand, placing a dialyzer at a point in the sample channel prior to the injection valve is quite common in online process monitoring, particularly in biotechnological applications to avoid contamination of fermentation tanks by microorganisms. For this purpose, laboratory-built or commercially available modules are furnished with sterilizable membranes.

A continuous liquid-liquid extractor receives separate streams of two immiscible phases and establishes a segmented flow within. Ordinary liquid-liquid extractors consist of three basic elements that are connected online (Figure 7B) in the following sequence: a solvent segmenter, an extraction coil, and a phase separator. The solvent segmenter – a minimixing chamber or a merging point – receives the streams of the two immiscible phases and establishes an outgoing flow consisting of alternate, regular segments of the two phases, which is driven to the extraction coil. The phase separator is the most important element of a liquid-liquid extractor in relation to smooth functioning. It receives the segmented liquid-liquid flow from the extraction coil and separates the segments in such a way that two independent streams of the two phases emerge from it. The separation of the immiscible phases can be based on density differences or the selective wetting of a special membrane by one of the phases – normally the organic phase. The operation of all three elements relies on two fundamental principles: (1) the density difference between the immiscible phases and (2) the selective wetting of the inner component surfaces by both organic and aqueous phases. As a rule, organic solvents wet PTFE and other plastic surfaces, whereas aqueous phases wet glass surfaces; conversely, organic solvents are repelled by glass surfaces and aqueous phases are rejected by PTFE surfaces. Indeed, complete phase separation is accomplished rarely and only when special requirements are met. Thus, the actual purpose of this element is to provide a flow of one of the two phases containing none of the other. The phase in question is usually the one containing the analyte or its reaction product, and also includes one that is subsequently driven to a continuous detector.

The extractor can be located at two different points in the FI manifold:

1. Before the injection valve in the sample stream, which is usually the aqueous phase. The



outgoing organic phase containing the analyte(s) fills the loop of an injection valve through which a carrier continuously passing through the detector is circulated. In the inject position, the organic content of the valve loop is directly transferred to the detector.

2. After the injection valve, where the sample is injected into an aqueous carrier that can continuously merge with one or several reagent streams. This and another stream carrying an organic solvent converge in the extractor. Finally, the outgoing phase stream is monitored in a continuous fashion.

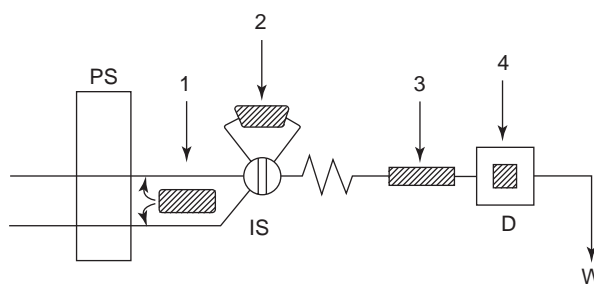
Some atypical FI liquid–liquid extraction systems do not use any proper separation unit, but an emulsion, an organized medium, a minicolumn capable of retaining one of the phases involved in the extraction process, or a reversed-flow system. A number of ingenious approaches for FI liquid–liquid extraction without phase separation have been designed.

### Packed Reactors

The use of solid-phase reactors coupled online to FI manifolds significantly increases the potential of FI by enhancing such basic analytical parameters as sensitivity and selectivity and allowing implementation of specific reactions. These reactors involve (bio)chemical reactions – whether enzymatic, immunoassay, ion exchange, or redox – or act as sorbent extractants or reagent releasers.

The minicolumn – tubular or conical in shape – is usually made of glass, PTFE, Tygon<sup>®</sup>, or PVC, and is 10–50 mm long and has an internal diameter of 1.5–3.0 mm. The location of these reactor(s) in an FI manifold depends on the type of material with which they are packed, which in turn is determined by their function in the overall process. **Figure 8** shows the different possible locations of these units:

1. Before the injection system, whether in the carrier or in the sample stream, for removal of reagent



**Figure 8** Possible locations for a minicolumn containing a solid reagent in an FI manifold.

impurities and release of reagents, or for matrix interference removal or preconcentration purposes, respectively.

2. In the loop of the injection valve, for preconcentration, multianalyte determinations, or simultaneous blank and sample measurements.

3. Between the injection and detection units, the most common location for a solid reactor in an FI manifold. The aim in this case can be any of those described above.

4. In the detection system, the main aim being integration of the reaction (retention) and detection in order to exploit some of the typical advantages of these approaches: enhanced sensitivity, selectivity, sampling rate, miniaturization, etc.

In addition to the use of immobilized enzymes on suitable supports, which results in decreased costs in routine analyses, the solid forms of these biocatalysts provide a number of additional advantages, the most immediate of which are: (1) simplification of the FI manifold required (use of an additional channel with a point of merging with the sample stream or an additional valve for simultaneous sample–enzyme injection is thus avoided) and (2) increased sensitivity as a result of the lower dilution of the sample on mixing with the enzyme solution. This latter advantage is even more apparent when several enzymatic steps involving different biocatalysts are required to obtain a measurable product. A serial arrangement of as many reactors as steps to be developed is very useful for this purpose. Immuno-reactors with either membranes or magnetic particles as the solid phase for the bound antibodies can also be used.

Ion exchange or sorption microcolumns coupled in-line in an FI manifold are used for two main purposes: preconcentration of the analyte(s) and removal of interferences using a support suited to the species to be retained. The two most serious shortcomings arising from the use of this type of microreactor are: (1) the increased compactness of the packed material resulting from continuous circulation of the flow in the same direction and (2) the appearance of parasitic signals at the detector as the sample matrix passes through the detection point during the preconcentration process. Both shortcomings are overcome by introducing a very simple change in the FI system, i.e., by placing the reactor in the loop of an injection valve in order to perform retention in the direction opposite to elution.

Redox reactions are used in FI systems for two main purposes: (1) to convert the analyte to an appropriate oxidation state for development of the subsequent derivatization reaction, thus avoiding

dilution arising from the merging with a redox solution and (2) for in-line formation *in statu nascendi* of strong reductants, thus making possible the use of redox agents in unstable oxidation states.

The most commonly used material for packing adsorptive reactors included in FI manifolds is C<sub>18</sub>-bonded silica beads and, to a lesser extent, alumina and silica. Preconcentration is by far the most common function of these reactors (especially of metal ions prior to their determination by atomic absorption or emission spectrometry). Enrichment factors of up to 50 are readily obtained with the aid of conical reactors, the direction of which in the manifolds is arranged in such a way that the adsorbed analytes are eluted from the wider to the narrower end of the column in order to minimize dispersion in the column.

Reagent-releaser reactors make very useful means of increasing sensitivity through substantially decreased dilution. In addition, they simplify the FI manifold as merging points and auxiliary channels are made redundant. These reagents are packed in cartridges or bound to resins.

## Detectors

Most of the detectors coupled to FI manifolds have been conventional instruments (e.g., molecular, atomic optical, or electrochemical) capable of providing only two-dimensional information. In this case, a simple flow cell or direct aspiration to the atomizer is the only interface required. In a parallel but delayed development of hyphenated systems in chromatography, the coupling of FI with high discrimination instruments pointed out the maturity of this technique. This step in FI started with the FI-ICP-AES coupling. More recent and less numerous have been the arrangements of FI with instruments that enable three-dimensional information and high discrimination capabilities, like mass spectrometry – by direct coupling or through an inductively coupled source – Fourier-transform infrared and nuclear magnetic spectroscopies, among the most important. The design of the interface for these couplings clearly limits the performance of the overall system.

## Automation of a Flow Injection Analyzer

Strictly speaking, FI is not an automatic or automated technique as only the transport and detection systems work continuously in any FI manifold. Indeed, for any of these systems to be categorized

as automatic or automated, the injection operation itself must be performed automatically. In addition, automation of the go-stop sequence of the propulsion unit is mandatory in order to implement stopped-flow methods, while automation of the data acquisition system is required when fast detectors are used, and that of switching valves is required in open-closed systems and other types of manifolds using them.

The old-fashioned components needed for full automation of an FI analyzer include electrical step motors for mechanically actuating valves and controlling the go-stop sequence of the propulsion system. Presently, RS-232 interfaces are used to allow for computer control and frame programs make operation very easy by including the corresponding steps of a given process in the general frame. The inclusion of triggers enables decisions to be taken by the analyzer, which thus becomes automated.

Automated sample introduction requires a sampler in addition to automated functioning of the injection valve when the analyzer is not devoted to online monitoring. These samplers are of the same type as those used for air-segmented systems and flame atomic absorption spectrometers. The liquid samples are transferred into individual vessels that are placed either on a circular turntable or in a square-shaped rack. The sampling time is preset so that the injection loop is completely filled. The washing or rinsing time – i.e., the time during which the sample probe either resides in a wash or rinse bath or is aspirating air – is also predetermined. The cycle time is the sum of the sampling time and the washing or rinsing time. The most commonly used samplers for FI work through the samples position by position on the turntable – random access samples are not normally used in FI analyzers. Carryover is usually a minor problem for most sampler types. A washing solution can be aspirated in between aspiration of different samples. Aspiration of air during the wash cycle is often quite sufficient. If the sample volume is a matter of concern when multiple analyses are to be performed, then the pump aspirating the sample should be programmed to be halted over the analysis time.

*See also:* **Enzymes:** Immobilized Enzymes. **Ion Exchange:** Overview. **Membrane Techniques:** Dialysis and Reverse Osmosis.

## Further Reading

Fang ZL (1998) Trends and potentials in flow-injection online separation and preconcentration techniques for

- electrothermal atomic-absorption spectrometry. *Spectrochimica Acta, Part B* 53(10): 1371–1379.
- Hansen EH (1992) Exploiting kinetic-based flow-injection methods for quantitative chemical assays. *Analytica Chimica Acta* 261(1–2): 125–136.
- Luque de Castro MD and Tena MT (1995) Hyphenated flow injection systems and high discrimination instruments. *Talanta* 42: 151–169.
- Prados-Rosales RC, Luque García JL, and Luque de Castro MD (2002) Propelling devices: the heart of flow injection approaches. *Analytica Chimica Acta* 461: 169–180.
- Prados-Rosales RC, Luque-García JL, and Luque de Castro MD (2003) Valves and flow injection manifolds: an excellent marriage with unlimited versatility. *Analytica Chimica Acta* 480: 181–192.
- Ruzicka J and Hansen EH (1988) *Flow Injection Analysis*. New York: Wiley Interscience.
- Valcárcel M and Luque de Castro MD (1987) *Flow Injection Analysis: Principles and Applications*. Chichester: Ellis Horwood.
- Wang JH and Hansen EH (2003) Sequential injectionlab-on-valve: the third generation of flow-injection analysis. *Trends in Analytical Chemistry* 22(4): 225–231.

## Detection Techniques

**M Miró**, University of the Balearic Islands, Palma de Mallorca, Spain

**W Frenzel**, Technische Universität Berlin, Berlin, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Flow injection analysis (FIA) is a popular methodology with widespread use largely because it can employ any detection system that is capable of accepting a flowing stream. In many instances, the flow pattern of a liquid stream passing through a particular detector is not disrupted, so that additional detection systems can be arranged in series for multiparametric measurements in a single sample zone injected. In recent years, daughter flow techniques such as sequential injection analysis (SIA), multi-commuted flow injection analysis, and multisyringe flow injection analysis have emerged as competitive methodologies of the former, especially in terms of versatility and online coupling to different detection instruments for sequential monitoring of key analytes. Moreover, the discontinuous operation of these novel techniques has expanded the applicability of flow systems, so that discrete, non-flow-through detection principles can also be hyphenated.

In general, detectors can be classified according to the way in which the analyte or reaction product is probed, i.e., whether it is done by bulk or surface sensing. It is also possible to draw a distinction between flow-through detectors measuring an intrinsic property of the analyte and those requiring prior online derivatization. From the point of view of instrumental requirements, it is more appropriate to

make the classification according to the underlying detection principle as presented in this article.

An ideal detector integrated in a flow manifold should possess several attributes, which in many respects are identical to those demanded for liquid chromatography. Accordingly, fast response, small dead volumes, wide dynamic ranges, high sensitivity, low noise level, reproducible and stable response, independence of the signal with fluctuations in flow velocity, simplicity of design, miniaturized size, robustness, and moderate cost are desirable. No existing detector fulfils the entire set of aforementioned requirements, and therefore a compromise between the various factors should be considered.

This article provides some general remarks on detection requirements for FIA and related techniques and outlines the basic features of the most commonly used detection principles, including optical methods (namely, ultraviolet (UV)–visible spectrophotometry, spectrofluorimetry, chemiluminescence (CL), infrared (IR) spectroscopy, and atomic absorption/emission spectrometry) and electrochemical techniques such as potentiometry, amperometry, voltammetry, and stripping analysis methods. Very few flowing stream applications involve other detection techniques. In this respect, measurement of physical properties such as the refractive index, surface tension, and optical rotation, as well as the  $\alpha$ -,  $\beta$ -, or  $\gamma$ -emission of radionuclides, should be underlined. Piezoelectric quartz crystal detectors, thermal lens spectroscopy, photoacoustic spectroscopy, surface-enhanced Raman spectroscopy, and conductometric detection have also been coupled to flow systems, with notable advantages in terms of automation, precision, and sampling rate in comparison with the manual counterparts.

Special emphasis is given in the bulk of the text to the design criteria for both flow-through cells and

- electrothermal atomic-absorption spectrometry. *Spectrochimica Acta, Part B* 53(10): 1371–1379.
- Hansen EH (1992) Exploiting kinetic-based flow-injection methods for quantitative chemical assays. *Analytica Chimica Acta* 261(1–2): 125–136.
- Luque de Castro MD and Tena MT (1995) Hyphenated flow injection systems and high discrimination instruments. *Talanta* 42: 151–169.
- Prados-Rosales RC, Luque García JL, and Luque de Castro MD (2002) Propelling devices: the heart of flow injection approaches. *Analytica Chimica Acta* 461: 169–180.
- Prados-Rosales RC, Luque-García JL, and Luque de Castro MD (2003) Valves and flow injection manifolds: an excellent marriage with unlimited versatility. *Analytica Chimica Acta* 480: 181–192.
- Ruzicka J and Hansen EH (1988) *Flow Injection Analysis*. New York: Wiley Interscience.
- Valcárcel M and Luque de Castro MD (1987) *Flow Injection Analysis: Principles and Applications*. Chichester: Ellis Horwood.
- Wang JH and Hansen EH (2003) Sequential injectionlab-on-valve: the third generation of flow-injection analysis. *Trends in Analytical Chemistry* 22(4): 225–231.

## Detection Techniques

**M Miró**, University of the Balearic Islands, Palma de Mallorca, Spain

**W Frenzel**, Technische Universität Berlin, Berlin, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Flow injection analysis (FIA) is a popular methodology with widespread use largely because it can employ any detection system that is capable of accepting a flowing stream. In many instances, the flow pattern of a liquid stream passing through a particular detector is not disrupted, so that additional detection systems can be arranged in series for multiparametric measurements in a single sample zone injected. In recent years, daughter flow techniques such as sequential injection analysis (SIA), multi-commuted flow injection analysis, and multisyringe flow injection analysis have emerged as competitive methodologies of the former, especially in terms of versatility and online coupling to different detection instruments for sequential monitoring of key analytes. Moreover, the discontinuous operation of these novel techniques has expanded the applicability of flow systems, so that discrete, non-flow-through detection principles can also be hyphenated.

In general, detectors can be classified according to the way in which the analyte or reaction product is probed, i.e., whether it is done by bulk or surface sensing. It is also possible to draw a distinction between flow-through detectors measuring an intrinsic property of the analyte and those requiring prior online derivatization. From the point of view of instrumental requirements, it is more appropriate to

make the classification according to the underlying detection principle as presented in this article.

An ideal detector integrated in a flow manifold should possess several attributes, which in many respects are identical to those demanded for liquid chromatography. Accordingly, fast response, small dead volumes, wide dynamic ranges, high sensitivity, low noise level, reproducible and stable response, independence of the signal with fluctuations in flow velocity, simplicity of design, miniaturized size, robustness, and moderate cost are desirable. No existing detector fulfils the entire set of aforementioned requirements, and therefore a compromise between the various factors should be considered.

This article provides some general remarks on detection requirements for FIA and related techniques and outlines the basic features of the most commonly used detection principles, including optical methods (namely, ultraviolet (UV)–visible spectrophotometry, spectrofluorimetry, chemiluminescence (CL), infrared (IR) spectroscopy, and atomic absorption/emission spectrometry) and electrochemical techniques such as potentiometry, amperometry, voltammetry, and stripping analysis methods. Very few flowing stream applications involve other detection techniques. In this respect, measurement of physical properties such as the refractive index, surface tension, and optical rotation, as well as the  $\alpha$ -,  $\beta$ -, or  $\gamma$ -emission of radio-nuclides, should be underlined. Piezoelectric quartz crystal detectors, thermal lens spectroscopy, photoacoustic spectroscopy, surface-enhanced Raman spectroscopy, and conductometric detection have also been coupled to flow systems, with notable advantages in terms of automation, precision, and sampling rate in comparison with the manual counterparts.

Special emphasis is given in the bulk of the text to the design criteria for both flow-through cells and

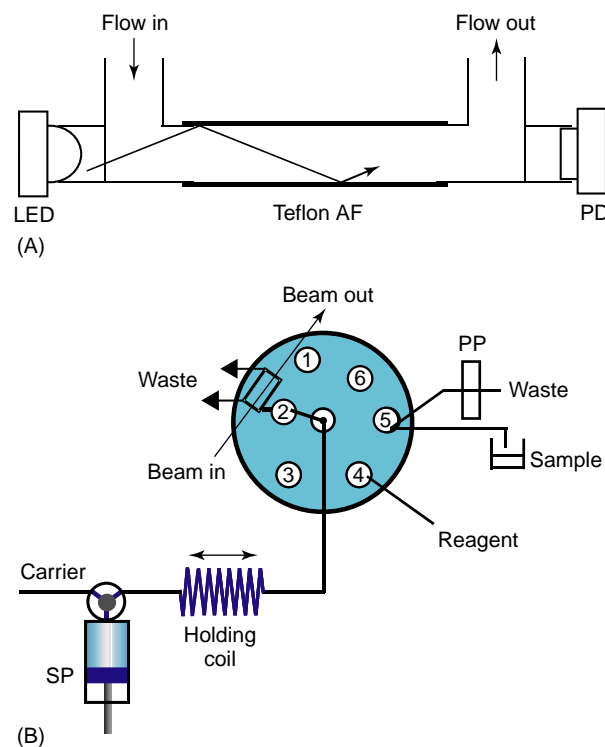
bio(chemical)sensors and also to the particular and unique benefits that arise from the hyphenation of flow systems with detection techniques.

## Optical Detection

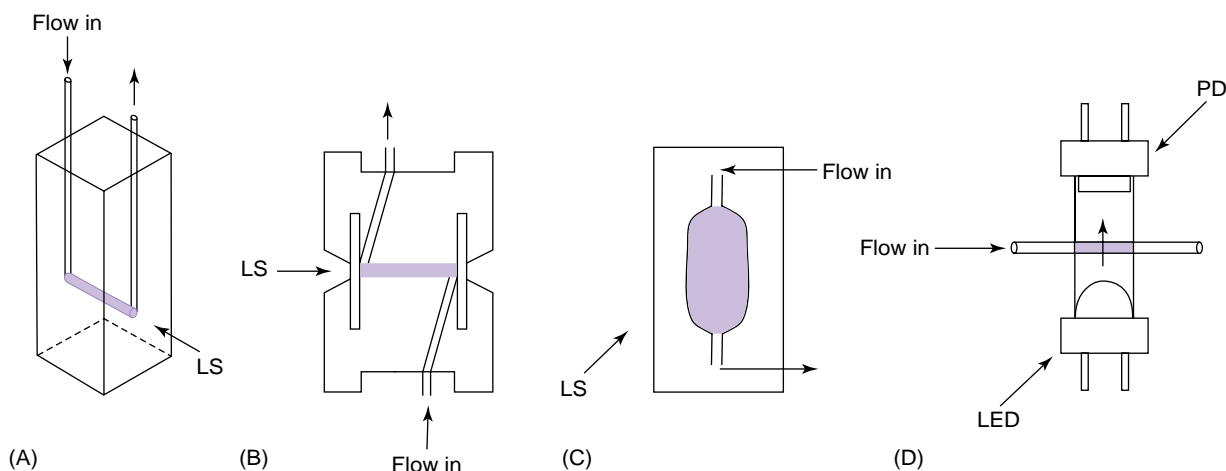
### UV-Visible Spectrophotometry

Through all the 30 years of development of flow injection analysis, photometry has been, and currently is, the detection technique most commonly interfaced with flow systems. Traditionally, common batchwise-operating spectrophotometers have been adapted easily to the flow requirements. Different commercially available U-, Z-, or oval-shaped flow-through cells (typically of 10 mm path length) are in widespread use (see Figure 1A–1C). Additional efforts have been focused recently on improvement of the sensitivity of the UV-visible spectrophotometric methods through the design of novel cells with optical paths longer than 20 mm. Although different materials have been assessed, all long path length cells suffer severe interference problems due to light attenuation and scattering. The novel synthesized cladding material Teflon AF-2400 warranted the use of liquid-core waveguide (LCW) flow cells (depicted in Figure 2A) with lengths as high as 200 cm. Since the refractive index of this polymeric material is lower than that of water, total reflection of light can be achieved by careful selection of the incident angle on the water/Teflon interface. The particular feature of this material with respect to fused silica, employed as a total reflection surface in organic media, is its unique behavior as a gas-diffusion membrane, which enables detection of volatile species.

A typical problem encountered in flow injection photometry is the influence of the Schlieren effect due to changes in the refractive index between the different zones delivered to the detector. The



**Figure 2** Universal flow-through cells in flow systems adaptable to absorbance, fluorescence, or chemiluminescence measurements. (A) Long path length water-core waveguide; (B) LOV coupled to sequential injection schemes. In the figure, both cells are assembled for spectrophotometric detection. SP, syringe pump; PP, peristaltic pump; LED, light-emitting diode; PD, photodetector.



**Figure 1** Typical flow-through cells used for spectrophotometric detection in flowing stream methods. (A) U-shaped cell for implementation in standard spectrophotometers; (B) Z-shaped cell as commonly employed in liquid chromatography; (C) oval-shaped cell used for both liquid-phase measurements and solid-phase optosensing detection; (D) solid-state detector with radially transmitted light beam. LS, light source; LED, light-emitting diode; PD, photodetector.



appearance of artifact signals over the transient profile deteriorates the sensitivity and repeatability of the measurements. This effect has been minimized using capillary flow cells wherein the light beam is directed to the optical detector perpendicular to the flow direction (see **Figure 1D**). Dual-wavelength detection is a commonly exploited alternative to compensating instrumentally the spectral shift caused by refractive index gradients. In recent years, this aim was accomplished using optical fiber-based micro-computer plug-in or charge coupled device spectrophotometers, which in combination with solid-state light emitting diodes (also available as multidiode light sources) enabled the design of portable and miniaturized detectors. The hyphenation of FIA with downsized photometers has been shown to be well suited for in-field environmental monitoring without the need for discrete sample collection and storage. A further advantage of fast scanning array spectrophotometers is the capability to exploit multidetermination through instantaneous full spectral acquisition and chemometric data processing. Simultaneous discrimination of various species in a single sample plug based on the reaction rate differences is also attained using laboratory-made serial or combined flow cells.

In the context of automation of the microfluidic handling of solutions and microminiaturization, the novel concept of lab-on-valve (LOV) in combination with SIA is worth mentioning. It is based on integrating a set of microconduits within a multipurpose valve, which allows a plethora of unit operations to be implemented. Optical detection in the microflow analyzer is performed easily via optical fibers assembled to the flow conduits. The sensitivity of the photometric measurements can be adjusted to the requirements of the assay through reconfiguration of the integrated system, so that different path lengths are available as shown schematically in **Figure 2B**.

In recent years, different flow cell designs have been proposed aiming to prevent the entrapment of bubbles of out-gassed air within the light path junctions, which disturbs the signal profile and worsens the signal-to-noise ratio. In this respect, the construction of multireflection flow cells consisting of a straight tube (made for example from Teflon AF-2400) that is illuminated from one side should be pointed out. In case gas bubbles enter, they are not retained and exit with the same velocity as the flowing solutions. An elegant strategy for overcoming the adherence of bubbles to the flow walls consists of exploiting a liquid drop as an optical cell. The windowless nature of the drop-based cell with a single collector fiber also enhances the capabilities of flow injection systems to accommodate turbidimetric/nephelometric detection

since the elimination of the windows overcomes the shortcoming of deposition of precipitate into the flow-through cell.

### Luminescence Detection

Despite the limited number of species that exhibit luminescent properties, fluorescence and (electro)chemiluminescence detection have been used extensively as detection techniques in flowing stream systems. This is a consequence of the much wider range of response, better selectivity, and lower detection limits achieved in comparison with molecular absorption spectroscopic methods. The outstanding benefits of flow injection systems for fluorescence measurements are the controlled chemical conditions that can be created in a highly repeatable manner and the fact that oxygen quenching may be readily minimized in the closed environment provided by the automated technique. Moreover, commercially available spectrofluorimeters furnished with flow-through quartz cells can be used without additional modifications. In recent years, spectrofluorimetry in combination with FIA/SIA can be regarded as a powerful hyphenated tool for performing multiparametric detection exploiting variable angle scanning fluorescence. The feasibility of implementing the stopped-flow mode required for recording the total luminescence data matrix and also pH gradient techniques for improving both the quantification limits and the spectral discrimination of the set of analytes is specially worth mentioning. The unique optical property of Teflon AF previously mentioned permitted the construction of a simple, portable, and stable transversely illuminated LCW laser-based fluorescence detector.

The analytical application of CL detection has been given a strong impetus by the advent of flow systems. The reproducible mixing of streams and precise timing control inherent to FIA are indispensable conditions for reliable CL measurements. Flow injection should be highlighted as a unique strategy for accommodating fast CL reactions, very often involving catalyzed-luminol oxidation, wherein light emission occurs instantaneously upon merging the sample and chemiluminogenic reagent with half-life times of only a few seconds. The coupling of enzymatic heterogeneous reactions using packed-bed or open-tubular reactors with CL detection has been presented as a useful online approach for indirect determination of several organic substrates. Immobilization procedures warrant low consumption of the expensive catalysts and enhanced enzyme stability. Yet, handling of soluble enzymes using multi-commutation protocols or merging-zones in FIA is



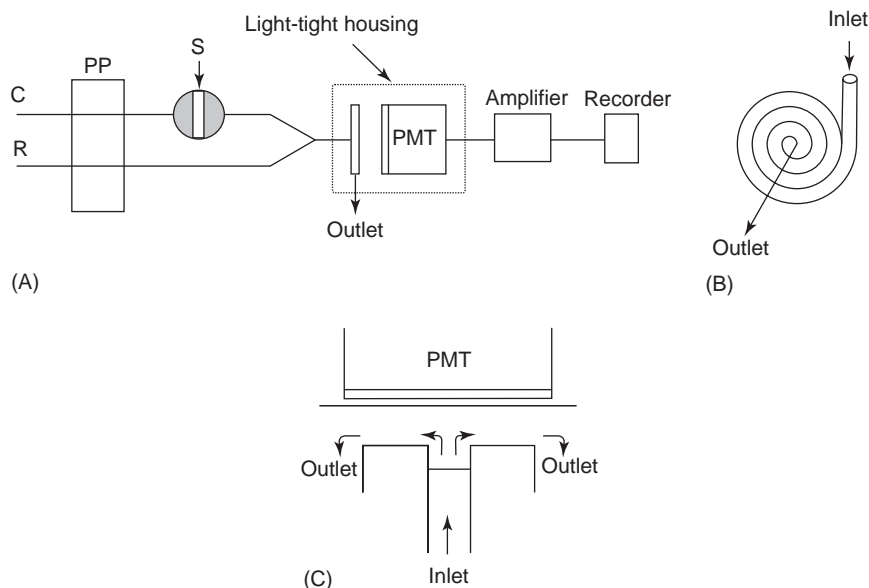
demanding for those preparations whose active groups are susceptible to being blocked during immobilization. It should also be stressed that the novel multisyringe flow injection principle, which includes the noteworthy features of FIA and SIA schemes, has been proved to be an excellent alternative to performing automated CL assays with different kinetic requirements without requiring manifold reconfiguration.

Instrumentally, CL measurements are extremely simple since no wavelength-restricting devices are necessary. The detector consists of a suitable flow-through cell, typically a spiral-shaped cell, placed in front of the window of a photomultiplier tube (PMT) as shown in **Figure 3A** and **3B**. In recent years, new designs with a larger surface, such as the fountain cell illustrated in **Figure 3C**, have been explored for CL detection. This cell geometry enables suitable collection of emitted light as a consequence of the accommodation of the entire sample volume within the observation area. High sensitivity is also attained with LCW-based detectors because of the effective transfer of the CL emission. Recording of the overall intensity–time profile is ensured with sandwich-type units furnished with a reading optical fiber since the analyte and reagents are mixed where the detection takes place. With respect to luminol electrochemiluminescence, novel configurations relying upon the working electrode facing the PMT have been devised during recent years. FIA has also been shown to be suitable for gas-phase CL involving the reaction of

nitric oxide in dried air (generated from nitrate, nitrite, or ammonium) with ozone supplied by an ozone generator.

### IR Spectroscopy

IR spectroscopy is a direct probe of molecular vibrations so that with few exceptions, practically all molecular species exhibit a unique IR absorption spectrum. In spite of this fact, the applicability of this detection technique for quantitative measurements in flowing stream systems has been rather limited in comparison with UV–visible and luminescence detection. This is a consequence of the greater complexity of the spectra, the overlapping of bands between moieties of different compounds, and the limitations of IR instruments. The potential of this technique has been enhanced via application of multiplex systems based on the Fourier transform (FT) algorithm. Fast spectra scanning and high resolution measurements are warranted exploiting FT-IR (FTIR) spectroscopy. Due to the powerful spectral information supplied by FTIR spectroscopy and the fact that the Lambert–Beer law is obeyed, many flow injection applications are devoted to the simultaneous determination of active constituents or pesticides in complex mixtures using multivariate data analysis (mainly partial least squares regression models). A plethora of flow methodologies use the information-rich mid-IR spectra in combination with organic carriers with a large absorption-free area such as chloroform. Flow-through



**Figure 3** A common configuration for flow injection chemiluminescence measurements. (A) Entire experimental setup; (B) enlarged view of the conventional flow-through cell; (C) schematic illustration of the fountain cell and its coupling to an end-on photomultiplier tube. R, reagent; S, sample; C, carrier; PMT, photomultiplier tube.

cells with sodium chloride or potassium bromide windows assembled in a liquid nitrogen-cooled mercury–cadmium–telluride detector have been employed commonly.

The applicability of mid-FTIR spectroscopy to aqueous solutions typically handled in flow manifolds is restricted to the spectral window of the range of  $1400\text{--}1000\text{ cm}^{-1}$  owing to the strong absorption of water. In this region, novel pH-modulation approaches for organic acids based on the laminar flow of aqueous streams (viz., alkaline sample and acidic solutions) through a sheath-flow cell configuration have been developed. The analytical readout is taken from the spectral changes induced by the pH change of the sample.

In case of measurements in the zone of the bending vibration of water, which also corresponds to the information-rich amide-I band of proteins, optical paths below  $10\text{ }\mu\text{m}$  are required to prevent large interference effects. Such short path lengths are available using either dedicated transmission cells or attenuated total reflection (ATR) approaches. However, both strategies have poor sensitivity as a result of the short optical path and also present practical problems for repeatable measurements. To circumvent these shortcomings, new IR light sources of high intensity, such as quantum cascade lasers, which enable transmission measurements at extended path lengths, have been adapted to FIA/SIA.

It should also be emphasized that the coupling of FIA/SIA with diamond-based ATR probes for online bioprocess monitoring was found to be an excellent alternative to *in situ* measurements aiming to overcome biofilm formation and gas bubble adherence to the surface of the ATR element.

### Atomic Spectrometry

Although originally FIA was conceived as a special technique for delivery of a sample segment into the instrument, the combination of flow injection as a sample pretreatment tool with atomic spectrometry has been shown to be of great potential for enhancing the selectivity and sensitivity of the measurements. Moreover, contamination problems are reduced due to the closed system used, making this interface suitable for ultratrace determination of metal species. Hyphenated techniques such as FIA/SIA with flame atomic absorption spectrometry, inductively coupled plasma (ICP)–optical emission spectrometry, and ICP–mass spectrometry (MS) have been exploited extensively in recent years. The major attraction of FIA–ICP–MS is its exceptional multi-elemental sensitivity combined with high speed of analysis. In addition, the possibility of

performing isotope and isotope-ratio analysis offers unique applications. The capabilities of FIA–ICP–time-of-flight–MS as a fast transient detector for simultaneous determination of a large number of isotopes in a FIA peak have also been evaluated thoroughly.

Flow rates of FIA systems for sample introduction are usually compatible with the continuous working mode of the aforementioned techniques and permit online operation using relatively straightforward interfaces. The interfacing of FIA with ICP–MS, for example, is attained readily by connecting the outlet of the FIA manifold with the liquid flow inlet of the detector nebulizer through a short length of narrow-bore tubing. It has also been reported that an improvement in nebulization efficiency and a facile adjustment of the signal magnitude may be accomplished via optimized automatic methods. Besides, sample injection into a continuously pumped carrier solution at a constant flow rate (instead of the conventional aspiration) minimizes transport interferences due to differences in viscosity of the samples and eliminates carry-over effects between consecutive samples owing to the periodical rinsing of the nebulizer. The ability to handle sample volumes at the microliter level also prevents some interfering processes, such as matrix deposition in the nebulizer. Hyphenated techniques allow a multitude of different injection modalities to be assessed. In this respect, merging-zones approaches, microsampling systems, and hydrodynamic injections have been proposed for decreasing sample and/or reagent consumption with regard to manual procedures. Monosegmented sample injection using air-segmentation was proved to be effective in restricting axial dispersion and in improving sampling throughput. It should also be highlighted that atomic spectrometry is the field where the largest amount of research effort on developing new FIA calibration methods has been focused. In this context, methodologies relying upon reproducible formation of the concentration gradient (namely, gradient dilution or gradient ratio methods) as well as clever designs for automatic standard additions calibration using merging-zones, zone penetration, gradient dilution, or interpolative strategies are particularly worth mentioning.

The practical applicability of absorption and emission spectrometric techniques is frequently restricted by spectral and nonspectral interferences due to their inherently low tolerance limit to sample constituents present in complex matrices. The most adequate way of alleviating the interfering effects consists of performing appropriate online sample pretreatment prior to analysis using flowing stream schemes. In many

instances, matrix separation is accompanied concurrently with analyte preconcentration. Hence, different FIA/SIA-hyphenated techniques involving co-precipitation with metal chelates, solvent extraction/stripping approaches, sorption on polytetrafluoroethylene knotted reactors, and microcolumn solid-phase extraction in a renewable fashion via the bead-injection (BI)-LOV principle have proven very effective.

The entire set of separation/preconcentration methodologies is equally applicable to electrothermal-atomic absorption spectrometry (ETAAS). It should be emphasized that the discontinuous operation of the SIA technique is better suited than FIA systems for hyphenation with ETAAS as a consequence of the discrete, non-flow-through nature of the detector. Moreover, the replacement of peristaltic pumps with syringe pumps as liquid drivers enables handling of minute, well-defined volumes with higher precision. Thus, the extremely low sample/eluent volumes able to be accommodated in the graphite tube ( $< 50 \mu\text{l}$ ) are manipulated easily with the automated SIA assembly and injected generally into the furnace via air-segmentation.

Vapor generation (VG) techniques involving hydride formation and cold vapor methods for monitoring metalloid species and mercury, respectively, have profited greatly from FIA in terms of a more than 90% reduction in sample and reagent consumption, two- to threefold higher sampling frequencies, better coefficients of variation, and significantly enhanced tolerance to interferences due to kinetic discrimination. Many efforts have recently been made to design novel gas-liquid separators for improved gas-phase separation. In the case of metalloid determination, FIA has been presented as a powerful tool for implementing speciation protocols through careful control of the experimental variables, such as the tetrahydroborate concentration for the formation of the volatile species; the selective complexation of trivalent or pentavalent species; the sequential volatilization of hydride species following cryogenic trapping; or the operational control of focused microwaves. *In situ* trapping of hydride species onto a Pd- or Ir-coated graphite tube of an ETAAS detector and mercury enrichment as a gold amalgam prior to detection enabled lowering the detection limits by more than order of magnitude. Very recently, the coupling of FIA-VG with atomic fluorescence spectrometry has been presented as a promising alternative to atomic absorption methods. The equipment is simpler and less expensive, and improved tolerance to organic compounds along with better determination limits is obtained usually.

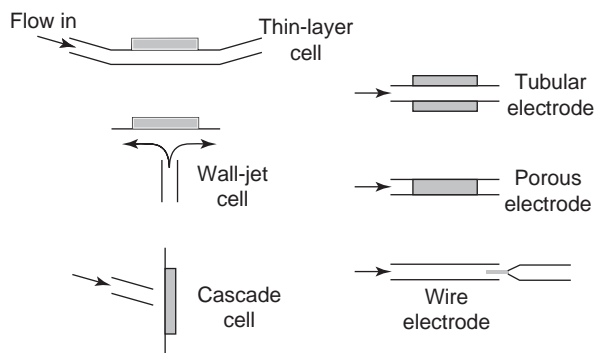
## Electrochemical Detection

The hydrodynamically well-defined conditions of flow systems are an ideal environment for electrochemical detectors, resulting in enhanced performance characteristics. The surface sensing properties of most electrochemical methods require particular attention in the construction of suitable flow-through cells. Efficient and repeatable mass transport toward the electrode surface is necessary, and dead volumes should be small. Various flow-through cells have been designed for electrochemical detection, all of which can be derived from the basic configurations depicted in Figure 4.

Electrochemical methods are subjected markedly to the interfering effects caused by adsorption of high molecular weight organic species on the electrode surfaces. Yet, matrix interferences can be alleviated considerably with the beneficial features resulting from interfacing FIA with membrane separation processes. As an alternative, online matrix removal with concomitant analyte preconcentration prior to the measurement of the electrical property may be performed via sorbent extraction procedures.

### Potentiometry with Ion-Selective Electrodes

Potentiometry is the most common method of electrochemical detection used in flow injection analysis. This methodology creates more favorable conditions for potentiometric measurements when compared to batch procedures. In FIA measurements, it is easier to avoid incidental contamination or that resulting from leakage of the solution from the reference electrode. Moreover, deactivation of the sensing surface of the indicator electrode due to adsorption, precipitation, or corrosion is minimized greatly as a consequence of the short contact time with the sample



**Figure 4** Basic configuration of flow-through cells applied in electrochemical flow methods. The position of the working/indicator electrode is marked by gray areas.

plug. An additional advantage resulting from the continuous flow of the sample zone is the improved selectivity attained exploiting the mechanism of kinetic discrimination between the analyte and the interfering species. The response time limitation of potentiometric detectors under dynamic regime for low analyte concentrations, which is observed mainly as a slow return of the potential value to the baseline level, can be overcome readily in FIA by addition of a small concentration of the sensed ion to the carrier stream. Flow systems are also useful schemes for proper sample presentation to the selective sensor since common strategies such as buffering, ionic-strength adjustment, pH control, and addition of masking agents are easily performed.

Another outstanding feature of using potentiometric detection in combination with FIA is the relative ease of construction of flow cells for implementing conventional electrodes or dedicated flow-through detectors. Careful consideration of the position of the indicator and reference electrode is essential to avoid artifact transient signals due to the influence of junction and streaming potentials. The most common types of flow-through cells used in FIA systems with commercially available ion-selective electrodes (ISEs) or photocured coated wire ISEs involve cascade designs, thin-layer cells, or wall-jet configurations (see **Figure 4** for further details). Although the latter arrangement provides the best dynamic performance and smaller carry-over effects, thin-layer cells are specially adapted to accommodating an array of sensing electrodes for multidetermination. In some instances, combination of the signals of the electrodes via multiplexed readings aiming to produce a more intense signal and an improvement in the precision of individual measurements has been explored.

Amongst dedicated potentiometric detectors, tubular cells with solid-state crystalline homogeneous membranes or with polyvinyl chloride or epoxy resin layers containing appropriate ionophores have been implemented successfully in flow manifolds. This configuration circumvents the problems that the incorporation of conventional-shaped units in FIA assemblies have in terms of mechanical stability. It should be stressed that difficulties in fixing the electrode in a rigid way can lead to an increase in noise and dead volumes, affecting the repeatability and return to the baseline. Further advantages include the maintenance of the sample dispersion profile and the avoidance of air bubble entrapment. Enhancement of the sensitivity of potentiometric tubular detectors in FIA has been achieved using double-membrane systems coupled with external electronic summing devices or dual detector systems in series wherein

one of the electrodes works alternatively as a reference electrode.

Progress in analytical potentiometry in recent years has brought a miniaturization of detection systems via the hyphenation of the sensor with a large extent of integration electronics. In this context, chemically sensitive layers are placed directly onto field effect transistors (FETs) to produce ion-selective FETs (ISFETs). The use of two identical ISFETs in a flow-through cell offers the possibility of differential measurements without the influence of the liquid junction potential.

### Amperometric and Voltammetric Detection

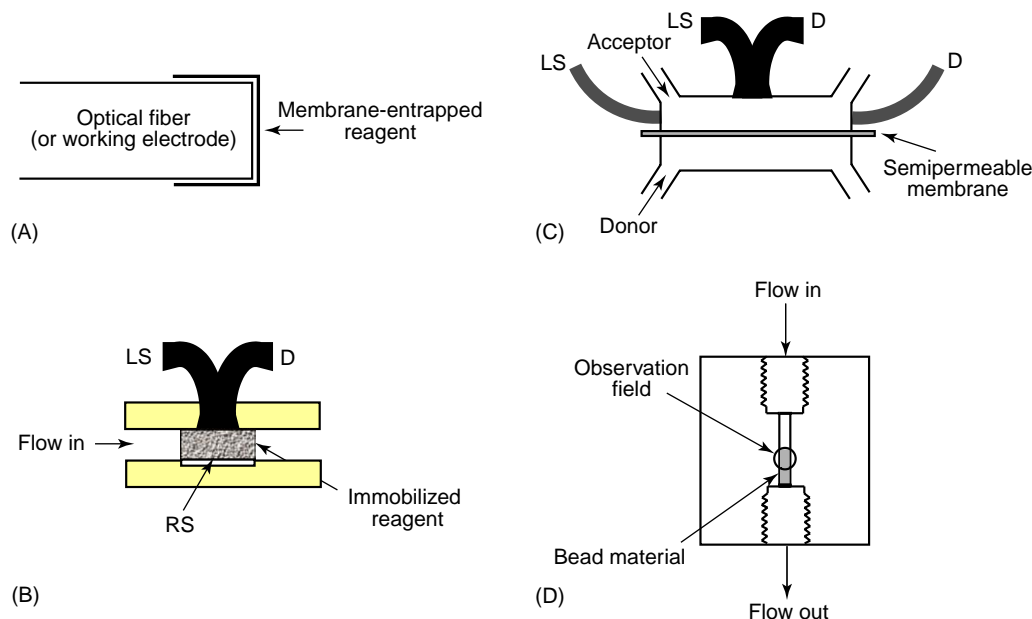
Despite the limited selectivity achievable with amperometric and voltammetric detection, their application in FIA has attracted the attention of many researchers. Flow systems provide suitable schemes for isolation of the analyte prior to detection and also the capability to implement selective biocatalytic reactions. The majority of developed flow injection amperometric procedures are performed in a three-electrode configuration, with constant polarization of the working electrode. Although glassy carbon is the material par excellence employed, modification of the electrode surface with species of electrocatalytic behavior is frequently demanded for minimizing overpotential effects. Selectivity improvement for electrochemical measurements is usually carried out by suitable conditioning of the electrode or by implementation of a polymeric layer exhibiting selective permeability or electrocatalytic properties. Furthermore, a marked increase in sensitivity can be achieved by exploiting pulse polarization techniques that enhance the faradaic current of the electroactive species. With regard to the flow cells used in flow injection amperometry, wall-jet and thin-layer arrangements should be highlighted. Different configurations are reported in the literature aiming to minimize the ohmic drop by proper positioning of the set of electrodes. It should be stated that the wall-jet design can be used not only with solid disk working electrodes but also with the hanging mercury drop electrode (HMDE). Very recently, an ultra-simple adaptor for converting batch cells with the HMDE in detectors for flow injection amperometric analysis, compatible with all the different brands and operating modes of the HMDE, has been described. A crucial parameter to take into account in these electrochemical techniques is the severe dependence of the flow rate on the current measurements, and thus a periodical recalibration of the liquid driver is required to prevent drastic oscillations of the hydrodynamic parameters.

The advantage of voltammetric methods, wherein current is measured as a function of the varying polarization of the working electrode, is the likelihood of performing multicomponent detection. Yet, a careful choice of the sweep rates in FIA is needed to avoid large background signals due to the changing current. In order to obtain the desired resolution of the voltammetric peaks, a fast scan rate with differential pulse or square-wave voltammetry has been used. When overlapping of several voltammetric signals is detected, data processing using a Kalman filter is known to be effective for deconvolution of the overall signal.

Metal traces are commonly determined in electrochemical procedures by flow injection anodic stripping voltammetry. Preconcentration of the desired species with the subsequent isolation from the matrix constituents is generally performed using mercury film electrodes. Stripping of the reduced species through an appropriate potential scan allows simultaneous determination of various metal ions. The basic advantages of performing stripping measurements in a dynamic mode are the ability to change the reaction medium conditions, which may be needed for the different stages of the stripping procedure, and the easy contamination control via the automation of the analyses.

## (Bio)chemical Sensors

In recent years, there has been increasing interest in the design of extrinsic optical fiber chemical (bio)sensors (i.e., (bio)optrodes) relying upon monitoring the optical properties of a reagent phase attached at the end of the optical fiber (as detailed in Figure 5A) that change on exposure to the analyte or an active derivative. They combine the advantages of optical fibers with the specificity and selectivity of the sensing element. Similarly, the immobilization of an enzyme or a component of an immunochemical assay at the surface of an electrochemical or piezoelectric detector, and also within the electrode material, allows the design of new detection devices for analytical purposes, so-called biosensors. Although these bio(sensors) have been traditionally used as probes for batch measurements, their implementation in flowing stream systems has opened new perspectives in the field of optical/electrochemical sensing owing to the ability to automate both the sample conditioning and sensor regeneration steps as well as to the ready miniaturization, facile handling, and individual optimization of the unit operations involved. It should be mentioned that though ISEs are a generic kind of chemical sensor, their fundamentals and application in flowing stream systems have been included in a separate section.



**Figure 5** Schematic illustrations of several bio(sensor) configurations assembled in flowing stream systems. (A) Probe-type configuration based on the modification of the optical fiber or working electrode with a reagent/catalyst containing membrane; (B) sandwich-type cell for optical-fiber-based reflectance measurements on a pad or thin layer with immobilized reagent; (C) membrane-based flow through sensor for absorbance or reflectance measurements in the receiver stream; (D) dedicated sorbent-packed planar-wall flow-through cell designed for solid-phase absorptiometry. LS, light source; D, detector; RS, reflecting surface.



The incorporation of the active microzone within the flow manifold has launched the development of flow-through (bio)sensors based on integrated (bio)chemical reaction and detection. Since the first application to reflectance measurements for pH monitoring using the scheme depicted in Figure 5B, optosensing techniques in flow injection systems have been exploited using very different sensing surfaces, such as thin membranes, cellulose pads or strips, or glass/polymeric beads accommodating indicators, chromogenic ligands, ionophores, or catalysts. In order to achieve the sensitivity requirements for many assays, optosensing preconcentration strategies have been adopted by several researchers. In the case of volatile analytes, a suitable means of increasing the enrichment factors consists of using flow-through gas-diffusion optosensors (see Figure 5C) in an acceptor stopped-flow alternative.

A common approach to integrating separation, preconcentration, and detection in flow-through sensors relies upon solid-phase optical detection onto sorbent bead materials packed in commercially available flow-cells. Yet, novel dedicated configurations with different path lengths involving tubular microcolumns, prismatic-shaped cells (shown in Figure 5D), and sandwich-type arrangements (especially useful for disk-based extraction procedures) have been evaluated for spectrophotometric and reflectance measurements. The pursued aim is minimizing the background attenuation, collecting the maximum amount of scattered light, and diminishing the optical dilution of the analyte retained. The noteworthy features of the on-column detection principle with regard to liquid-phase readings are the elimination of the artifact signals resulting from the Schlieren effect and the prevention of dispersion of the preconcentrated analyte, which is the most severe pitfall of sorbent extraction/elution microcolumn-based flow systems with eluate sensing.

The novel concept of BI commonly associated with sequential injection systems represents a remarkable progress in the field of flow-through solid-phase (bio)optosensors. It has also been introduced as a powerful tool for automation of immunoassays in both the jet-ring cell and lab-on-valve configurations. As a consequence of its renewable nature, BI presents several advantages for proper sensor performance with respect to optosensing strategies with permanent packed-bed reactors. Amongst them, the suitability for implementing irreversible reagent-based assays as well as the long-term operational applicability due to the absence of both sorbent compaction

(or clogging) and alteration of sorption behavior as a result of surface deactivation should be highlighted.

**See also:** **Amperometry.** Principles and Instrumentation; Flame; Electrothermal; Vapor Generation. **Atomic Emission Spectrometry:** Principles and Instrumentation; Inductively Coupled Plasma. **Atomic Fluorescence Spectrometry.** **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Chemiluminescence:** Liquid-Phase. **Enzymes:** Enzyme-Based Electrodes. **Fluorescence:** Instrumentation. **Ion-Selective Electrodes:** Overview. **Optical Spectroscopy:** Detection Devices. **Sensors:** Overview. **Voltammetry:** Overview.

## Further Reading

- Calokerinos AC and Palilis LP (2001) Chemiluminescence in flow injection analysis. In: García-Campaña AM and Baeyens WRG (eds.) *Chemiluminescence in Analytical Chemistry*, pp. 321–348. New York: Dekker.
- Fang Z-L (1995) *Flow Injection Atomic Absorption Spectrometry*. Chichester: Wiley.
- Izquierdo A and Luque de Castro MD (1995) Ion-sensitive field-effect transistors and ion-selective electrodes as sensors in dynamic systems. *Electroanalysis* 7: 505–519.
- Kolhed M, Haberkorn M, Pustogov V, *et al.* (2002) Assessment of quantum cascade lasers as mid-infrared light sources for measurement of aqueous samples. *Vibrational Spectroscopy* 29: 283–289.
- Li Q-Y, Morris KJ, Dasgupta PK, Raimundo IV, and Temkin H (2003) Portable flow-injection analyzer with liquid-core waveguide based fluorescence, luminescence and long path length absorbance detector. *Analytica Chimica Acta* 479: 151–165.
- Miró M and Frenzel W (2004) Flow-through sorptive preconcentration with direct optosensing at solid surfaces. *TrAC – Trends in Analytical Chemistry* 23: 11–20.
- Nagels LJ and Staes E (2001) Polymer (bio)materials design for amperometric detection in liquid chromatography and flow injection analysis. *TrAC – Trends in Analytical Chemistry* 20: 178–185.
- Sanz-Medel A (1999) *Flow Analysis with Atomic Spectrometric Detectors*. Amsterdam: Elsevier.
- Trojanowicz M (2000) *Flow Injection Analysis: Instrumentation and Applications*. Singapore: World Scientific.
- Trojanowicz M, Szweczyńska M, and Wcisło M (2003) Electroanalytical flow measurements – recent advances. *Electroanalysis* 15: 347–365.
- Valcárcel M and Luque de Castro MD (1994) *Flow-Through (Bio)Chemical Sensors*. Amsterdam: Elsevier.
- Wang J-H and Hansen EH (2003) Sequential injection lab-on-valve: the third generation of flow injection analysis. *TrAC – Trends in Analytical Chemistry* 22: 225–231.



## Environmental and Agricultural Applications

**M Miró**, University of the Balearic Islands, Palma de Mallorca, Spain

**W Frenzel**, Technische Universität Berlin, Berlin, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Flow techniques are being consolidated as powerful tools for the routine control of parameters in various samples of environmental and agricultural origin. The outstanding features of flow injection analysis (FIA) for analytical applications are its extremely straightforward configuration, easy operation, and low costs. Automated analyses with high sample throughput are attained using relatively simple instrumentation. The microformat of FIA results in considerable reduction of sample and reagent consumption compared to classical wet-chemical methods of analysis. Furthermore, as opposed to air-segmented flow analysis (SFA) methodologies, the reproducible timing and readily controllable dispersion of the analyte zone allows the performance of kinetic measurements without demanding steady-state conditions. Other differential features of FIA in relation to SFA are the injection of smaller sample volumes (of the order of 10–100  $\mu\text{l}$ ), the suppression of the washing cycles, the increase of the sampling throughput up to 120 injections  $\text{h}^{-1}$ , the repeatability improvement, and especially the ease of manifold construction and flexibility. The ready adaptation of batch methods permits FIA to be used in compliance with many official standard methods (namely, US Environmental Protection Agency (EPA), International Standards Organization, and European Norms; see comments in Table 1). Environmental safety is another attribute of FIA. The closed-system chemistry not only decreases contamination of the sample but also prevents technicians coming into contact with hazardous chemicals. Moreover, as a consequence of its inherent miniaturization, a considerable decrease in waste generation, and hence also in costs for waste disposal, is achieved. FIA has also attracted the attention of many researchers owing to its ability to fulfill the current demands in environmental analysis. The fast response of FIA makes the analytical information available in real time, which is especially desirable in monitoring schemes, some of the flow analysers being included in EPA directives. FIA has

also been exploited for in-field analysis aiming to obtain high temporal and spatial resolution data without the need for discrete sample collection and storage. This is of great interest whenever offline analysis or excessive sample handling is unacceptable due to the rapid transformation of the target analyte.

Despite the aforementioned benefits of FIA with regard to manual procedures and precedent flow techniques, the incompatibility of the elastic tubes of peristaltic pumps with concentrated acid or bases and organic solvents usually forces the periodical recalibration of the system or the incorporation of more expensive reagent-resistant tubing. The physical adsorption of organic analytes onto the Tygon tubes is another practical limitation commonly described. Though performance of multiparametric determinations using a single multidetection system or several detectors arranged in series has been reported in several studies, the simultaneous monitoring of various environmental parameters have been rather limited to particular applications. It should be also stressed that the continuous flowing of solutions according to FIA philosophy results in the consumption of carrier, reagent, and sample during the overall analysis protocol.

Sequential injection analysis (SIA) was proposed in 1990 as a new concept for handling solutions alternatively to FIA. Whereas in conventional FIA manifold the sample volume is inserted into the carrier stream and subsequently merged downstream with reagents, SIA methodology is based on the software-controlled sequential aspiration of precise volumes of sample and reagent, which are afterwards dispensed into the reaction coil by flow reversal. Therefore, a considerable sample and reagent saving is achieved – of the order of 20 times in relation to FIA. The replacement of the FIA injection valve with a rotary selection valve has enabled the incorporation of additional reactors/modules, the performance of backward–forward flow movements, the exploitation of stopped-flow schemes, and the simple development of multiparametric determinations by coupling different detectors and appropriate reagents to the multiposition valve. The use of syringe pumps as liquid drivers has allowed the manipulation of sample and reagent volumes at the low microliter level with high precision. The microanalytical features of SIA, along with the handling of the exact volumes required for the particular assays, have opened new avenues in the monitoring of environmental parameters. For example, atmospheric

**Table 1** Analytical features of flowing-stream methods applied to the determination of relevant parameters in natural, tap, wastewaters, and industrial effluents

Analyte	Method/reagents/ derivatization reaction	Detection	Working range	Particular features
Ammonium	Berthelot	Spectrophotometry	0.010–1.0 mg N l <sup>-1</sup>	Replacement of phenol by salicylate
	Nessler	Spectrophotometry	0.050–0.50 mg N l <sup>-1</sup>	Coating of the colloidal product on the tubing walls
	GD – boric acid, HCl, or distilled water as acceptors	Conductimetry	0.010–0.50 mg N l <sup>-1</sup>	Conductivity changes
	GD – pH indicator	Spectrophotometry	0.010–1.0 mg N l <sup>-1</sup>	Standard method EN/ISO 11732
	ISE containing the nonactin ionophore	Potentiometry	0.8–140 mg N l <sup>-1</sup>	Gas-diffusion/use of activated carbon to remove interfering organic compounds
	Ion-selective coated wire electrode	Potentiometry	0.14–1400 mg N l <sup>-1</sup>	Portable battery-powered analyzer
	Clinoptilolite-modified electrode	Amperometry	0.28–14 mg N l <sup>-1</sup>	Copper-doped electrode
	NaClO-luminol	Chemiluminescence	0.05–0.4 mg N l <sup>-1</sup>	Indirect measurement
	Hypochlorite-ozone	Chemiluminescence	0.0014–1.4 mg N l <sup>-1</sup>	Gas-phase detection
	OPA	Spectrofluorimetry	0.0004–0.7 mg N l <sup>-1</sup>	Use of nucleophilic species
Nitrite	Griess-Ilosvay	Spectrophotometry	0.010–0.70 mg N l <sup>-1</sup>	Simultaneous determination of nitrate using a Cu–Cd column/standard method EN/ISO 13395
	Griess-Ilosvay/reversed-phase extraction	SPS	0.3–30 µg N l <sup>-1</sup>	Solid-phase optrode
	Gallocyanine	Spectrophotometry	0.003–0.80 mg N l <sup>-1</sup>	Catalytic reaction
	Naphthol Green B	Spectrophotometry	0.6–60 µg N l <sup>-1</sup>	Catalytic reaction
	3-Aminonaphthalene-1,5-disulfonic acid	Spectrofluorimetry	> 14 µg N l <sup>-1</sup>	Formation of an azoic acid salt
	3,3',4,4'-Biphenyltetramine	Spectrofluorimetry	9–30 µg N l <sup>-1</sup>	Direct determination
	Rhodamine 6G/bromate	Spectrofluorimetry	2.4–30 µg N l <sup>-1</sup>	Kinetic method
	Iodide	Biamperometry	0.02–14 mg N l <sup>-1</sup>	Detection of the tri-iodide generated
	Carbon paste electrode modified with a ruthenium polymer	Amperometry	0.0007–7 mg N l <sup>-1</sup>	Decrease of the overpotential for nitrite oxidation
	Griess/Cu–Cd	Spectrophotometry	0.060–11.30 mg N l <sup>-1</sup>	Online reduction to nitrite
Nitrate	No reagent	Spectrophotometry	0.10–3.10 mg N l <sup>-1</sup>	Direct UV detection-multivariate analysis
	Griess/photo-induced reduction	Spectrophotometry	> 0.42 µg N l <sup>-1</sup>	Use of EDTA as activator
	Luminol/UV irradiation	Chemiluminescence	0.001–1.4 mg N l <sup>-1</sup>	Generation of peroxynitrite species
	Tubular ISE composed of a DPNi containing film	Potentiometry	> 3.4 mg N l <sup>-1</sup>	Assessment of various ionic strength modifiers
	Sodium peroxodisulfate/hydrazine/Griess	Spectrophotometry	0.20–20 mg N l <sup>-1</sup>	Online microwave-assisted digestion
Total nitrogen	Alkaline peroxodisulfate/Cd-column/Griess	Spectrophotometry	0.03–3.0 mg N l <sup>-1</sup>	Online photo-oxidation
	Sodium peroxodisulfate/Cu–Cd column/Griess	Spectrophotometry	> 5.0 µg N l <sup>-1</sup>	Thermal digestion using a capillary tube containing a Pt wire
Orthophosphate	Molybdenum blue	Spectrophotometry	0.010–0.25 mg P l <sup>-1</sup>	Use of antimony as a catalyst
	Molybdenum blue	Spectrophotometry	> 0.2 µg P l <sup>-1</sup>	Solid-phase extraction
	Molybdate/vanadate	Spectrophotometry	0.15–18 mg P l <sup>-1</sup>	Detection of the yellow complex without reduction
	Molybdate/malachite green	Spectrophotometry	0.010–0.10 mg P l <sup>-1</sup>	Ion-pair association

Table 1 Continued

Analyte	Method/reagents/ derivatization reaction	Detection	Working range	Particular features
Phosphorus species	Molybdate/rhodamine B	Spectrofluorimetry	0.0003–0.095 mg P I <sup>-1</sup>	Quenching of fluorescence
	Molybdate/thiamine	Spectrofluorimetry	0.007–6.6 mg P I <sup>-1</sup>	Generation of fluorescent thiochrome
	Pyruvate oxidase G/luminol	Chemiluminescence	0.003–1.0 mg P I <sup>-1</sup>	Generation of hydrogen peroxide via enzymatic reactions
	Glassy carbon electrode	Amperometry	0.006–0.32 mg P I <sup>-1</sup>	Free of silicate interference
	Cobalt wire electrode	Potentiometry	3.1–310 mg P I <sup>-1</sup>	Indirect determination
	Sodium peroxodisulfate in acidic medium/molybdate/ascorbic acid	Spectrophotometry	0.010–1.0 mg P I <sup>-1</sup>	Determination of dissolved organic phosphorus/online UV-photo-oxidation/standard method EN/ISO 15681
Silicate	Thermal (or microwave-assisted) digestion in acidic medium/molybdate/tin(II) chloride	Spectrophotometry	0.020–2.0 mg P I <sup>-1</sup>	Determination of total dissolved phosphorus previous acidic hydrolysis of condensed phosphates
	Molybdenum blue	Spectrophotometry/ICP-AES	0.1–200 mg P I <sup>-1</sup>	Serial arrangement of detectors for the determination of orthophosphate and total phosphorus
	Molybdenum blue	Spectrophotometry	0.14–1.4 mg Si I <sup>-1</sup>	Addition of masking agents (tartaric or oxalic acids)/standard method EN/ISO 16264
	Molybdate/rhodamine B	Spectrophotometry	0.17–2.0 mg Si I <sup>-1</sup>	Ion-pair association
Sulfate	Molybdate	Spectrophotometry	3.5–80 mg Si I <sup>-1</sup>	Simultaneous determination of orthophosphate and silicate exploiting the differences of reaction rates
	Molybdate	Spectrophotometry	0.8–15 mg P I <sup>-1</sup>	
	Molybdate	Spectrophotometry	0.9–30 mg Si I <sup>-1</sup>	Selection of appropriate acidity and sample segmentation with oxalic acid
	Molybdenum blue	Spectrophotometry	0.2–12 mg P I <sup>-1</sup>	
	Molybdenum blue	Spectrophotometry	5.0–50 mg Si I <sup>-1</sup>	Selection of appropriate concentration of oxalic acid
	Molybdenum blue	Spectrophotometry	0.2–7.0 mg P I <sup>-1</sup>	
	Barium chloride in a polyvinyl alcohol medium	Turbidimetry	5.0–200 mg SO <sub>4</sub> <sup>2-</sup> I <sup>-1</sup>	Addition of alkaline EDTA to avoid the build-up of precipitate
	Barium dimethylsulfonazo-III	Spectrophotometry	0.2–14 mg SO <sub>4</sub> <sup>2-</sup> I <sup>-1</sup>	Color-fading reaction
Sulfide	Ferric nitrate	Spectrophotometry	10–1000 mg SO <sub>4</sub> <sup>2-</sup> I <sup>-1</sup>	Sample pretreatment with activated charcoal is required to remove interfering organic species
	Methylthymol blue (MTB)	Spectrophotometry	0.025–1.0 mg SO <sub>4</sub> <sup>2-</sup> I <sup>-1</sup>	Solid-phase extraction/detection of the MTB–Ba complex
	Barium chloranilate	Spectrophotometry	4.0–100 mg SO <sub>4</sub> <sup>2-</sup> I <sup>-1</sup>	Displacement reaction
	BANH-Zr	Spectrofluorimetry	1.5–150 mg SO <sub>4</sub> <sup>2-</sup> I <sup>-1</sup>	Comparison of various manifolds
	MTB-Zr	Spectrophotometry	0.5–20 mg SO <sub>4</sub> <sup>2-</sup> I <sup>-1</sup>	Catalytic reaction
Sulfide	Lead nitrate	Potentiometry	100–1000 mg SO <sub>4</sub> <sup>2-</sup> I <sup>-1</sup>	Use of a Pb-ISE
	Nitroprusside	Spectrophotometry	0.5–10 mg S <sup>2-</sup> I <sup>-1</sup>	No concentrated aggressive reagents are required

Continued

Table 1 Continued

Analyte	Method/reagents/ derivatization reaction	Detection	Working range	Particular features
Chloride	Methylene blue	Spectrophotometry	$0.01\text{--}15\text{ mg S}^{2-}\text{ l}^{-1}$	30-fold sensitivity increase with regard to nitroprusside method
	Methylene blue and nitroprusside	Spectrophotometry	$>0.2\text{ }\mu\text{g S}^{2-}\text{ l}^{-1}$	Gas-diffusion separation/preconcentration using a halted recipient solution
	Direct determination on a glassy carbon electrode	Amperometry	$0.03\text{--}24\text{ mg S}^{2-}\text{ l}^{-1}$	Adaptable to on-board measurements
	Sulfide-sensitive electrode	Potentiometry	$1\text{--}1000\text{ mg S}^{2-}\text{ l}^{-1}$	Tubular electrodes with homogeneous crystalline membranes or treated silver foils
	2,6-Dichlorophenol-indophenol modified electrode	Amperometry	$0.65\text{--}32\text{ mg S}^{2-}\text{ l}^{-1}$	Long-term stability
	Mercury(II)thiocyanate/Fe(III)	Spectrophotometry	$1.0\text{--}1000\text{ mg Cl}^{-}\text{ l}^{-1}$	Standard method EN/ISO 15682
	Hg-EDTA	Spectrophotometry	$>0.2\text{ mg Cl}^{-}\text{ l}^{-1}$	UV detection
	Mercury(II)thiocyanate	Spectrophotometry	$0.16\text{--}2000\text{ mg Cl}^{-}\text{ l}^{-1}$	Absence of Fe(III)
	Ag (or Hg)-chloranilate	Spectrophotometry	$1.0\text{--}20\text{ mg Cl}^{-}\text{ l}^{-1}$	Use of a solid reagent reactor
	AgNO <sub>3</sub>	Turbidimetry	$3.0\text{--}30\text{ mg Cl}^{-}\text{ l}^{-1}$	Recirculation of reagent in a closed system
Residual chlorine	Chloride-ISE	Potentiometry	$0.1\text{--}10\,000\text{ mg Cl}^{-}\text{ l}^{-1}$	Wide dynamic range/standard method EN/ISO 15682
	Dedicated Ag-ISE	Potentiometry	$2.0\text{--}90\text{ g Cl}^{-}\text{ l}^{-1}$	Online titration
	Methyl orange	Spectrophotometry	$0.1\text{--}10\text{ mg Cl}^{-}\text{ l}^{-1}$	Color fading reaction
	4-Nitrophenylhydrazine	Spectrophotometry	$0.05\text{--}40\text{ mg Cl}^{-}\text{ l}^{-1}$	Generation of azo-dye species (or measurement of decrease of reagent absorbance)
	<i>o</i> -Tolidine	Spectrophotometry	$0.5\text{--}5.0\text{ mg Cl}^{-}\text{ l}^{-1}$	Formation of an unstable hydroquinone
	<i>o</i> -Dianisidine	Spectrophotometry	$0.05\text{--}1.30\text{ mg Cl}^{-}\text{ l}^{-1}$	Gas-diffusion/tandem-flow approach
	Gold electrode	Amperometry	$>7.0\text{ }\mu\text{g Cl}^{-}\text{ l}^{-1}$	Useful for monochloramine determination
	Rhodamine 6G	Chemiluminescence	$0.07\text{--}7.0\text{ mg Cl}^{-}\text{ l}^{-1}$	Electrostatic immobilisation of the reagent
	Disproportionation of chlorine into HClO and Cl <sup>-</sup>	Potentiometry	$0.7\text{--}7.0\text{ mg Cl}^{-}\text{ l}^{-1}$	Gas-diffusion separation/Cl-ISE
	<i>N,N</i> -diethyl- <i>p</i> -phenylenediamine/KI	Spectrophotometry	$0.1\text{--}8.0\text{ mg Cl}^{-}\text{ l}^{-1}$	Sequential determination of free and combined chlorine
Fluoride	KI/UV detection	Spectrophotometry	$0.03\text{--}3.0\text{ mg Cl}^{-}\text{ l}^{-1}$	Determination of total chlorine
	Fluoride-ISE	Potentiometry	$0.005\text{--}50\text{ mg F l}^{-1}$	High selectivity, wide dynamic ranges, and low detection limits
	Lanthanum(III)/alizarin complexone	Spectrophotometry	$0.03\text{--}3.5\text{ mg F l}^{-1}$	Sensitivity improvement via addition of cationic surfactants
	Zr(IV)-SPADNS	Spectrophotometry	$0.02\text{--}3.5\text{ mg F l}^{-1}$	Inhibitory effect
Bromide	Lanthanum(III)/alizarin complexone	ICP-AES	$0.03\text{--}1.3\text{ mg F l}^{-1}$	Liquid-liquid extraction/indirect determination
	Phenol red	Spectrophotometry	$0.16\text{--}2.4\text{ mg Br l}^{-1}$	Formation of the bromophenol blue dye

Table 1 Continued

Analyte	Method/reagents/ derivatization reaction	Detection	Working range	Particular features
Iodide	Tetrapase/chloramine T	Spectrophotometry	1.0–40 $\mu\text{g Br I}^{-1}$	Catalytic reaction
	Permanganate/Pt-electrode	Amperometry	> 0.08 mg Br I <sup>-1</sup>	Gas-diffusion
	Iron(III) thiocyanate	Spectrophotometry	0.75–150 $\mu\text{g I I}^{-1}$	Fading effect
	Dichromate/KI (receiver)	Spectrophotometry	> 0.2 mg I I <sup>-1</sup>	Gas-diffusion
	Ce(IV)/As(III)	Spectrophotometry	0.002–0.5 mg I I <sup>-1</sup>	Sandell–Kolthoff's reaction
	Sodium nitrite	MIP-AES	> 2.3 $\mu\text{g I I}^{-1}$	Generation of volatile iodine
Bromate	Cr(VI)	FAAS	6.0–220 $\mu\text{g I I}^{-1}$	Indirect method/sorption of formed Cr(III) onto a chelating column
	Chlorpromazine	Spectrophotometry	1.0–30.0 $\mu\text{g BrO}_3^- \text{ I}^{-1}$	Masking strategies are required to minimize interference from nitrite and chlorite
Chlorate/chlorite	Ion-exchanger/ammonium hydroxide	ICP-MS	0.06–50 $\mu\text{g BrO}_3^- \text{ I}^{-1}$	Selective elution of bromide and bromate from an activated alumina microcolumn
	KI/HCl	Spectrophotometry	0.1–10.1 mg ClO <sub>2</sub> <sup>-</sup> I <sup>-1</sup> 0.1–8.3 mg ClO <sub>3</sub> <sup>-</sup> I <sup>-1</sup>	Speciation is possible due to the different acidic requirements for iodide oxidation
Cyanide	Chloramine-T/pyridine-barbituric acid	Spectrophotometry/ Amperometry	0.002–2.0 mg CN <sup>-</sup> I <sup>-1</sup>	Gas-diffusion preconcentration/UV-photolysis for total cyanide determination/ standard method ISO 14403
Surfactants	Metallic silver-wire electrode	Potentiometry	0.05–250 mg CN <sup>-</sup> I <sup>-1</sup>	Gas-diffusion/use of a non-selective sensor
	Luminol/Cu	Chemiluminescence	0.005–2.0 mg CN <sup>-</sup> I <sup>-1</sup>	Luminol oxidation by Cu(CN) <sub>4</sub> <sup>2-</sup>
	Hypochlorite/rhodamine B	Chemiluminescence	0–50 mg I <sup>-1</sup>	Determination of amine ethoxylate-based nonionic surfactant
	Alizarin fluorine blue	Spectrophotometry	0.2–12.0 mg I <sup>-1</sup>	Determination of Triton-type surfactants
	Bromocresol purple	Spectrophotometry	> 10 <sup>-6</sup> mol I <sup>-1</sup>	Determination of cationic surfactants
	Luminol/ <i>N</i> -bromosuccinimide (or <i>N</i> -chlorosuccinimide)	Chemiluminescence	10 <sup>-6</sup> –10 <sup>-4</sup> mol I <sup>-1</sup> (CPC, DTAB, CTAB)	Decrease of chemiluminescence emission at surfactant concentrations below the critical micelle concentration
DIC	Methylene blue (standard method EN/ISO 16265) or rhodamine B	Spectrophotometry (or spectrofluorometry)	> 0.03 mg I <sup>-1</sup> (C <sub>12</sub> -alkylbenzene sulfonate)	Solvent extraction/on-tube detection
	GD-Cresol Red or 4-(2',4'-dinitrophenylazo)-1-naphthol	Spectrophotometry	10 <sup>-6</sup> –10 <sup>-3</sup> mol I <sup>-1</sup>	Color change of acid–base indicators
	GD-NaOH (acceptor stream)	Conductimetry	> 3 × 10 <sup>-6</sup> mol I <sup>-1</sup>	Measurement of signal decrease
	GD-H <sub>2</sub> SO <sub>4</sub> (donor stream)	QCM	5 × 10 <sup>-4</sup> to 2 × 10 <sup>-2</sup> mol I <sup>-1</sup>	Transient variation of the monitored frequency of QCM on CO <sub>2</sub> passage
TOC	GD-HNO <sub>3</sub> (donor stream)/N <sub>2</sub> (recipient stream)	FTIR	> 7.5 × 10 <sup>-5</sup> mol I <sup>-1</sup>	Gas-phase detection
	GD-carbon dioxide detection	Spectrophotometry/ infrared spectrometry	10 <sup>-6</sup> –10 <sup>-4</sup> mol I <sup>-1</sup>	Photochemical oxidation

Continued

Table 1 Continued

Analyte	Method/reagents/ derivatization reaction	Detection	Working range	Particular features
COD	GD-carbon dioxide detection	BAWIS	$10^{-5}$ – $2 \times 10^{-2} \text{ mol l}^{-1}$	Wet chemical oxidation
	Permanganate oxidation	Spectrophotometry	$> 5.0 \text{ mg l}^{-1}$	40 m coil, 100°C
	Dichromate oxidation	Spectrophotometry	$2.0$ – $100 \text{ mg l}^{-1}$	Microwave-assisted digestion
	Cerium(IV) oxidation Dichromate oxidation via microwave-assisted digestion	Spectrophotometry FAAS	$0.5$ – $130 \text{ mg l}^{-1}$ $25$ – $5000 \text{ mg l}^{-1}$	20 m coil, 100°C Retention of the excess of dichromate on an anionic exchange column prior atomic absorption detection
BOD	Trichosporon cutaneum/tris(4,7-diphenyl-1,10-phenanthroline) ruthenium(II)	Spectrofluorimetry	$0$ – $110 \text{ mg l}^{-1}$	Design of a three-layer fiber optic microbial sensor
pH	Flat-headed combined glass electrode	Potentiometry	pH 1.0–14.0	FIA titration
	N,N-dioctadecyl-methylamine	Potentiometry	pH 4.0–11.0	The ionophore was photo-cured on a silver wire
	Immobilized indicator	Spectrophotometry	pH 4.0–10.0	Optosensing method
	Thymol blue	Spectrophotometry	pH 7.6–8.1 ( $\pm 0.0007 \text{ pH units}$ )	High temporal resolution
Alkalinity	Methyl orange or methyl red	Spectrophotometry	$10$ – $200 \text{ mg l}^{-1}$	Online titration
	HCl as a titrant	Potentiometry	No calibration required	Determination of partial and intermediate alkalinity

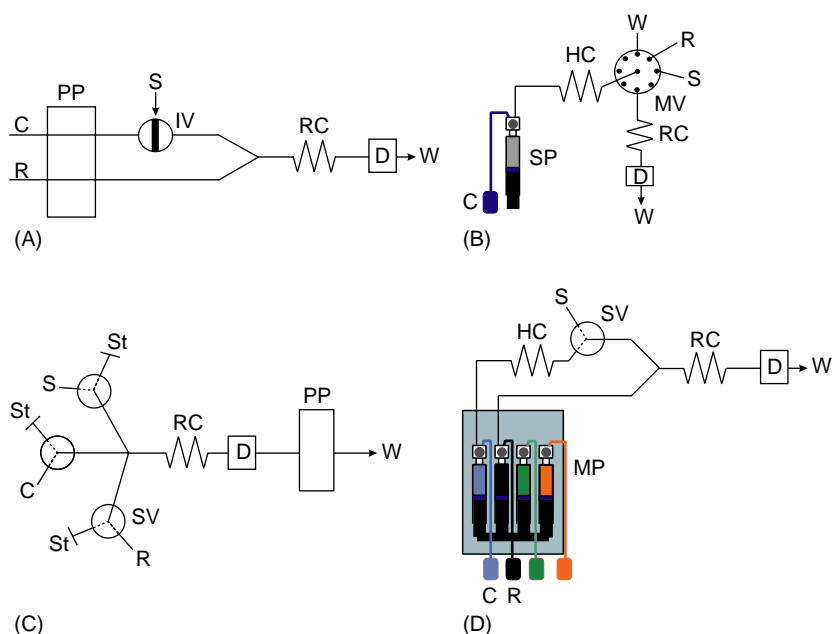
DIC, dissolved inorganic carbon; TOC, total organic carbon; COD, chemical oxygen demand; BOD, biochemical oxygen demand; ISE, ion-selective electrode; SPS, solid-phase spectrophotometry; Tur, turbidimetry; FAAS, flame atomic absorption spectrometry; ICP-AES, inductively coupled plasma atomic emission spectrometry; ICP-MS, inductively coupled plasma mass spectrometry; MIP-AES, microwave-induced plasma atomic emission spectrometry; QCM, quartz crystal microbalance; FTIR, Fourier transform infrared spectrometry; BAWIS, bulk acoustic wave impedance sensor; GD, gas-diffusion; EDTA, ethylenediaminetetraacetic acid; OPA, o-phthalaldehyde; MTB, methylthymol blue; BANH, biacetylmonoximinicotinylhydrazone; SPADNS, 2-(p-sulfophenylazo)-1,8-dihydroxynaphthalene; Tetrabase, 4,4'-bis(dimethylamino)diphenylmethane; CPC, cetylpyridinium chloride; DTAB, dodecyltrimethylammonium bromide; CTAB, cetyltrimethylammonium bromide; DPNi, tris(4,7-diphenyl-1,10-phenanthroline)nickel(II).

precipitates are often only present in limited quantities or long collection times are required, so that microsampling is extremely beneficial. The main drawback of SIA for analytical applications is the lower sampling rate achieved in relation to FIA as a consequence of the periodical filling of the liquid driver and the stacking of the entire set of segments in the holding coil. Besides, the sequential and non-simultaneous propelling of the solutions that takes place in SIA results in longer residence times for proper interdispersion between zones. This shortcoming has been already solved by replacing the multiposition SIA valve with solenoid valves or miniaturized integrated conduits, so that novel flow techniques such as the so-called multicommutated flow injection analysis, multisyringe flow injection analysis (MSFIA), and sequential-injection lab-on-valve have been recently designed. These methodologies, which include the outstanding features of the parent FIA and SIA schemes, have proven effective tools for the miniaturization and automation of

environmental assays and also for performing reliable sample pretreatments. **Figure 1** illustrates various flowing-stream configurations devised for the spectrophotometric or spectrofluorimetric monitoring of several analytical parameters in aqueous samples.

In this article, flowing-stream systems assembled for environmental and agricultural analysis are classified according to their application area, i.e., water, air, soil, and plant analysis. Within each area relevant flow methodologies for the determination of individual analytes are briefly reviewed and specific features of particular methods are outlined. Additional information about the analytical performance of several flow assemblies is listed comprehensively in the tables. The likelihood of direct introduction and treatment of solid samples in an automated fashion is also highlighted in the bulk of the text. Finally, attention is also paid to the different schemes available for online speciation studies, which are of increasing significance in environmental assays.





**Figure 1** Schematic illustration of flowing stream manifolds devised for the automatic determination of relevant environmental parameters in water samples using various flow techniques: (A) flow injection analysis, (B) sequential injection analysis, (C) multicommutation flow injection analysis, (D) multisyringe flow injection analysis. The sequential injection-lab-on-valve configuration is depicted in previous article: Flow injection analysis-detection techniques. S, sample; R, reagent; C, carrier; D, detector; IV, injection valve; HC, holding coil; RC, reaction coil; MV, multiposition valve; SV, solenoid valve; SP, syringe pump; MP, multisyringe pump; PP, peristaltic pump; St, stopper; W, waste.

## Determination of Nutrients, Anionic Species, and Related Compounds in Waters

Flow techniques are particularly suitable for monitoring environmental parameters in waters. Moreover, flow methods offer unique possibilities for automatic sample pretreatment, involving inline dilution or filtration, adjustment of viscosity, ionic strength, or pH as well as removal of solid and colloidal matrix constituents via in-line dialysis separation techniques.

Most of water matrices (viz. drinking, marine, salty, river, waste, ground, estuarine, cooling, and thermal power plant water) may be easily handled with flow systems. Since the analytical methods performed to date are likely suitable for the determination of the specific parameter in different kinds of matrices, the applications of flow systems will be reviewed according to the target analyte – following the practical guidelines for water-quality surveillance, instead of attending the sample source. The analytical features of relevant flowing-stream methods applied to environmental water analysis are comprehensively compiled in Table 1.

## Nitrogen Compounds

**Ammonium** The determination of ammonium is of significant interest in pollution control and environmental evaluation of water. Current flow injection spectrophotometric strategies are based upon the Berthelot reaction since it is highly selective for ammonium, its repeatability is excellent, and in the presence of catalysts, such as sodium nitroprusside, low detection limits are attained. The Nessler reaction has also been implemented in flow systems, yet its applicability was limited by the precipitation of the colloidal product onto the inner walls of the poly(tetrafluoroethylene) (PTFE) tubing. One of the most promising methodologies for the selective determination of ammonium involves the implementation of gas-diffusion units in flow injection devices coupled mainly to spectrophotometric detection, notwithstanding conductometric and potentiometric detections have also been proposed. In the latter case, the membrane-based ammonium ion-selective electrode (ISE) containing the nonactin ionophore has been widely used. Nevertheless, photo-cured membranes on silver wires are recommended owing to the excellent adhesion to the metal substrate and improved mechanical strength. With regard to

amperometric detection, ion-doped clinoptilolite or polyaniline-modified electrodes have been successfully assembled in flow injection devices. Sensitive chemiluminescence (CL) methods include the measurement of light emission decrease from NaClO–luminol system or the gas-phase monitoring of the reaction product between ozone and nitric oxide resulting from analyte oxidation. Luminescence methods relying on derivatization reactions with o-phthalaldehyde in the presence of nucleophilic species have been reported as convenient procedures for ammonium determination in a variety of water matrices.

**Nitrite** Nitrite is an important indicator of fecal pollution in natural waters as well as a potential precursor of carcinogenic species. A rush of flow and sequential injection spectrophotometric method based on Griess-type reactions has been proposed, also coupled to online sorbent enrichment schemes. The catalytic effect of nitrite on the oxidation of various organic species constitutes the basis of fairly sensitive spectrophotometric methods. Fluorometric methods based on the formation of aromatic azoic acid salts, quenching of Rhodamine 6G fluorescence, and direct reaction with substituted tetramine or naphthalene species have been also reported. Indirect CL methods usually involve conversion into nitric oxide and gas-phase detection as mentioned in the foregoing section. The redox reaction between nitrite and iodide in acidic media is the fundamental of a plethora of flow injection methodologies with spectrophotometric, CL, or biamperometric detection. New electrochemical sensors with chemically modified carbon paste electrodes containing ruthenium sites, or platinum electrodes with cellulose or naphthalene films, have recently attracted special attention for amperometric detection.

**Nitrate and total nitrogen** One of the primary steps for environmental studies concerning water quality, sanitation, and public health is the determination of nitrate. Runoff from fields treated with fertilizers is the major cause of nitrate pollution of river and underground waters. High levels of nitrate are toxic to children. Flow and sequential injection spectrophotometric procedures for nitrate as well as total oxidized nitrogen determination are almost entirely based on the heterogeneous online reduction to nitrite using a copperized cadmium reactor. Yet, the homogeneous reduction of the analyte by hydrazine or ultraviolet (UV) irradiation has also been investigated by several authors. Photochemical activation is also well suited for CL detection, since peroxo-nitrite species, which are the oxidizing agents for

luminol, are formed during the process. Highly sensitive enzymatic methods based on the spectrophotometric monitoring of the nicotinamide adenine dinucleotide phosphate reduced by the nitrate reductase enzyme have been presented. Potentiometric detection using tubular-shaped electrodes or heterogeneous membrane-based ISEs is also commonly exploited in flowing systems.

Suitable alternatives to the batch digestion methods for total nitrogen determination include online UV photodegradation in the presence of oxidizing agents and online microwave-assisted chemical digestion with alkaline peroxodisulfate. The resulting inorganic nitrogen species are determined using the methods described above.

### Phosphorus and Silicon Compounds

Determination of orthophosphate is of great importance in water-quality surveillance since this compound is regarded as the best indicator of the nutrient status of natural waters. The chemical basis of spectrophotometric flow procedures for monitoring dissolved reactive phosphorus lies in the reduction of the molybdophosphate complex (molybdenum blue method) in the presence of antimony to enhance the reaction yield, although detection of the yellow molybdovanadophosphate complex has also been reported. Spectrophotometric alternatives comprise the formation of ion pairs of heteropoly acid species with cationic complexes or basic dye compounds such as Rhodamine B, Crystal Violet, and Malachite Green. Fluorometric determinations relying upon the generation of the fluorescent thiochrome or on the quenching of fluorescence of various Rhodamine species are frequently performed. Moreover, highly sensitive CL-based biosensors involving enzyme-packed bead reactors have been very recently devised. Several researchers described the electrochemical reduction of the heteropoly acid to molybdenum blue at a glassy carbon electrode without silicate interference. Indirect potentiometric determination of orthophosphate in wastewaters has been proven effective using a Co-ISE. The overall methodologies have also been applied to dissolved organic phosphorus determination following mild inline photo-oxidative decomposition. Yet, monitoring of total dissolved phosphorus requires previous acidic hydrolysis of condensed phosphates.

Silicate is an essential element and even a limiting micronutrient for some aquatic microorganisms. Its concentration drives the growth of a large part of the siliceous phytoplankton. Determination of dissolved silica (viz. orthosilicic acid and its dimer) is generally performed by flow injection spectrophotometric

procedures based on the formation of the yellow  $\beta$ -silicomolybdic acid, its reduction product, or ion-associate complexes. As a consequence of the close chemistry between orthophosphate and silicate, both species severely interfere with the determination of each other. The interference of orthophosphate in silicate monitoring methods is overcome by the addition of tartaric or oxalic acids, which are capable of decomposing selectively the molybdophosphate complex. The concentration of masking agent may be also adjusted to avoid the molybdosilicic acid formation in the determination of orthophosphate. The different rates of formation of both heteropoly acids and the dependence of the color development on the acidity of the reaction medium have also been exploited as discrimination procedures.

### Sulfur Compounds

Though sulfate is one of the major components of natural waters, its determination is of particular interest with respect to the problems of acid rain and water pollution. The widespread industrial use of sulfuric acid makes the monitoring of sulfate decisive for environmental protection. Five spectrophotometric methodologies commonly exploited in the flow or sequential injection fashion are the barium (or lead) sulfate turbidimetric method, the barium dimethylsulfonazo-III approach, the UV-detection of the iron(III) sulfate complex, the methylthymol blue procedure, and the barium chloranilate (or barium chromate) displacement reaction. The major shortcoming of the above methods is their limited selectivity owing to the nonspecific reaction involved. The implementation of online cation-exchange separation processes and the addition of masking agents overcome this drawback. The participation of sulfate in both the formation of fluorophore ternary complexes and the catalysis of the slow reaction between methylthymol blue and zirconium constitutes the basis of fairly sensitive methods. Water quality assessment has been frequently performed using both lead and barium-coated ISEs, although improved detectability is readily attained following in-line solid-phase preconcentration onto anion exchangers.

Sulfide is mainly produced in natural waters from the oxidation of organic matter, and it is highly toxic to most aquatic organisms. Extensive sample manipulation should be avoided in order to minimize loss of analyte by air oxidation or volatilization, and hence flow methodologies are advantageously used. Both methylene blue and nitroprusside methods are the most popular spectrophotometric procedures to measure the sulfide content in various water matrices. Isolation and preconcentration steps for matrix

removal and determination of sulfide at trace levels usually involve sorbent extraction strategies following analyte derivatization or permeation of hydrogen sulfide through porous membranes into a stationary acceptor solution. Electrochemical techniques well adapted to portable flow devices such as amperometry with integrated calibration units or potentiometry using tubular ISEs with homogeneous crystalline membranes, or treated silver foils have been proposed by other researchers. The electrocatalytic oxidation of the target species using an indophenol-modified glassy carbon electrode has been evaluated and successfully applied to the monitoring of sulfide in urban wastewater treatment plants.

### Halogen and Pseudohalide Compounds

One of the most useful chemical analysis to assess the salinity, river and ocean flow patterns, and seawater contamination of estuarine areas, is the determination of the concentration of chloride in environmental waters. The most frequently exploited methodology relies upon chloride reaction with mercury(II) thiocyanate in an iron(III)-containing medium. Spectrophotometric detection in the UV region of either the Hg-ethylenediaminetetraacetic acid-Cl ternary chelate or the  $\text{Hg}(\text{SCN})_2\text{-Cl}$  complex in the absence of iron(III) has also been investigated. Additional indirect photometric methods based on the principle of ion exchange of easily detectable anions versus chloride from suitable mercury salts, such as the UV determination of chloranilate ions, should be mentioned. Aiming to prevent the use of mercury-containing reagents, automated turbidimetric procedures involving silver chloride precipitation have been very recently reported. Amongst the electrochemical methods, the direct or titration-based potentiometric flow systems using dedicated ISEs have attracted the interest of many researchers owing to the wide dynamic ranges and low detection limits attainable as well as the easy adaptation of the electrodes in the flow network.

The inherent toxicity of chlorine and that derived from the production of carcinogenic trihalomethanes has led to the increasing importance of devising robust and simple analyzers for free and combined residual chlorine monitoring. Most of the existing procedures are based upon flow injection spectrophotometric methods comprising color-fading reactions with methyl orange or nitrophenylhydrazine, formation of an unstable hydroquinone with *o*-tolidine, or oxidation of various chromogenic reagents. Amperometric approaches relying on detection of the reduction wave of hypochlorous acid at a gold electrode have been designed and successfully applied to

different kinds of water matrices. The oxidation of organic molecules such as xanthene or rhodamine 6G with free chlorine composes the fundamental of sensitive luminescence procedures. The most widely accepted methodology for measuring total chlorine in waters involves the reaction of chloramines and free chlorine with a potassium iodide stream, and further determination of the iodine generated by UV detection or by reaction with *N,N*-diethyl-*p*-phenylenediamine aiming to form a colored semiquinoid.

Almost all flowing-stream methods reported to date for free and total fluoride determination and applied to quality control and environmental monitoring are related to potentiometric detection using the fluoride ISE owing to its high selectivity. Spectrophotometric alternatives based either on the Alizarin complexone methodology or on the inhibitory effect of fluoride on the formation of different metal chelates have also accomplished a great approval.

Selective determinations of bromide and iodide have been performed by implementation of gas diffusion units for halogen permeation after analyte oxidation with permanganate or dichromate, followed by the phenol red reaction or amperometric measurements, and spectrophotometric detection of triiodide, respectively.

Trace amounts of cyanide are usually determined by flow injection spectrophotometric procedures. The target species is first halogenated with chloramine-T, after which it reacts with a mixture of pyrazolone or barbituric acid and isonicotinic acid or pyridine to form a bluish-violet polymethine dye. The implementation of gas-diffusion modules in the flow setup for hydrogen cyanide separation avoids matrix interferences and enables the adaptation of inherently nonselective detectors, such as metallic silver-wire electrodes for potentiometric measurements. Total inorganic cyanide, including free and complexed species, such as iron-cyanide complexes, may be determined by sample decomposition with UV irradiation and further photometric or amperometric analysis.

### Nonspecific Water Quality Parameters

The presence of pollutants in environmental waters may be detected by the measurement of the chemical oxygen demand (COD). This parameter, which represents the oxygen required for complete oxidation of the organic matter, is, likewise to the batch procedure, evaluated in a flow system as the consumption of a strong chemical oxidant (namely, permanganate, dichromate, or cerium(IV)). The intrinsic drawback of making the COD determination fully automated in flow assemblies owing to the short

residence time of the sample zone as opposed to the conventional procedure may be alleviated by performing online microwave-assisted digestion. Determination of the amount of oxygen taken by microorganisms that decompose organic waste matter, i.e., biochemical oxygen demand, under flow conditions requires the construction of miniaturized microbial sensors based on the combination of an oxygen or optical fiber probe with a culture of microorganisms able to assimilate a wide spectrum of pollutants after immobilization. The measurement of the dissolved inorganic carbon (DIC) after sample pretreatment by wet-chemical oxidation or UV photolysis allows determination of TOC, which is the simplest approach to evaluate the degree of organic pollution without interference of reducing inorganic compounds. DIC is usually monitored using various detection techniques (viz. conductimetry, spectrophotometry, infrared spectrometry, or photoacoustic spectroscopy) following transformation into carbon dioxide, which is trapped in the recipient stream of a gas-diffusion module.

Conductivity measurements are normally performed using purpose-built flow-through cells aimed to avoid flow rate influence. Total ionic concentration is determined by transformation of the salts present in the water sample into their corresponding acids on passing through a cationic exchanger in the proton form, whereupon spectrophotometric titration is carried out. Potentiometric methods for pH measurement using either ionophore-containing photo-cured or poly(vinyl chloride) membrane-based sensors or flat-headed combined electrodes have been readily adapted to automated flow analyzers. To circumvent the classical shortcomings of potentiometric methodologies, several researchers exploited UV-visible spectrophotometric detection with different acid-base indicators. Alkalinity is conventionally measured by online titration with hydrochloric acid as titrant and either methyl orange or methyl red as indicators.

### Metalloid Species and Trace Metals

Metalloid compounds are usually determined by flowing-stream techniques hyphenated with hydride generation (HG)-atomic absorption or atomic fluorescence spectrometry. The continuous operation mode inherent to flow injection is specially suited for the latter detection technique as the tetrahydroborate reagent is a potential source of hydrogen for supporting the flame. Analyte preconcentration is frequently needed to detect the typical levels of metalloid species found in water matrices. In this context, cold trap collection of generated hydrides, sorbent extraction microcolumn methods, sorption

of metalloid derivatives onto the inner walls of PTFE tubes, co-precipitation with lanthanum hydroxide, or retention of volatile compounds onto Pd- or Ir-coated graphite tubes before electrothermal atomization are worth mentioning. Speciation of As and Se has gained increasing importance as the toxic effects of the above species are directly related to their oxidation state. Differentiation between trivalent and pentavalent arsenic is made possible by careful control of the chemical variables, namely tetrahydroborate concentration and pH for arsine generation. Speciation of inorganic selenocompounds has been demonstrated by simple on-off operation of a focused microwave connected to the HG flow set-up. Electrochemical assays capitalized on the strict control of both the electrolysis potential of thin mercury film-based working electrodes and the acidic medium in the potentiostatic deposition have also been reported for metalloid speciation.

The hyphenation of flowing systems, mainly flow injection, with atomic spectrometric techniques, such as flame or plasma emission spectroscopy (FES or ICP-OES), inductively coupled plasma mass spectrometry (ICP-MS), flame (FAAS) or electrothermal (ETAAS) atomic absorption spectrometry, and cold vapor atomic absorption spectrometry for mercury has proven to be a powerful analytical tool for the determination of ultratrace levels of metals without risk of sample contamination, since an entirely enclosed environment is attained. The benefits of interfacing sequential injection schemes with the discontinuous operation of an ETAAS detector have been recently demonstrated. A noteworthy feature of ICP-MS is the ability of performing isotopic measurements as well as simultaneous multielemental detection, which is of particular importance for fast screening of metal profiles in environmental waters. Yet, the foregoing detection techniques, mainly ICP-MS, inherently show low tolerance to sample constituents present in environmental matrices, including high concentrations of easily ionized elements, salts, and mineral acids (after sample treatment), which might cause severe matrix effects and spectral interferences. In this respect, the isolation of the metal of interest from the matrix components using online separation techniques readily adaptable to flowing-stream methods is called for. Among the different reported schemes, those based on incorporation of column reactors packed with a sort of chelating sorbents (namely, 8-hydroxyquinoline, iminodiacetate, dithiocarbamate, or dithiophosphate) or with nonpolar materials such as copolymeric resins or octadecyl chemically modified silica gel should be highlighted. Co-precipitation/adsorption procedures involving entrapment of the

precipitate/chelate on the inner walls of a knotted PTFE reactor, sorption of volatile species onto coated-graphite tubes of an electrothermal atomizer, and solvent extraction methods including both phase-segmentation or wetting-film modalities should also be featured as worthwhile strategies for online analyte enrichment with concomitant matrix removal.

In recent years, important efforts have been made to adapt electrochemical techniques such as anodic/cathodic stripping voltammetry and potentiometric stripping analysis to flowing systems for monitoring trace levels of transition metals in aqueous samples. Significant progress has also been made in the development of spectrophotometric catalytic methods for trace metal determinations as these approaches are extremely suitable for implementation in portable devices and, hence, for onsite and real-time analysis. Most of such systems are based on well-known catalytic oxidation reactions of phenol or aniline derivatives as well as on the oxidative coupling of amines. Flow injection with inline matrix removal coupled to CL detection is another noteworthy option owing to the high sensitivity and selectivity, wide dynamic ranges, rapid analysis (seconds), robustness, portability, and low costs.

A plethora of speciation schemes for discrimination between different oxidation states and chemical forms (namely, Fe(II)/Fe(III), Cr(III)/Cr(VI), exchangeable/labile/acid soluble aluminum, and inorganic/organic complexes) has been developed exploiting selective complexation reactions prior to spectrophotometric measurements. Multidetector in flow systems with a single sample injection involving the arrangement of several detectors in series, mainly spectrophotometric and AAS detection, has also attracted special attention in speciation studies owing to the inherent advantages of simultaneous determination of species and reduced sample consumption. **Table 2** compiles the noteworthy features of flowing-stream systems hyphenated with various detection techniques for the determination and speciation of metal and metalloid species.

### Organic Compounds

In recent years, the determination of trace amounts of organic compounds in environmental samples for routine water quality control has been performed using flow systems as an economical and straightforward alternative to chromatographic and electrophoretic methods. Great attention is specially paid to the development of FIA/SIA methods for the determination of anthropogenic pollutants such as polycyclic aromatic hydrocarbons, polychlorinated biphenyls, and different families of pesticides



**Table 2** Application of flowing stream systems to the determination of metal and metalloid species in water matrices

<i>Detection technique</i>	<i>Typical elements determined</i>	<i>Water sample</i>	<i>Remarks</i>
Atomic spectrometry	Na, K, Li, Ca	Drinking, tap, ground, surface, natural	An internal standard (namely, Li) is commonly incorporated in samples and standards
FP			A low-temperature flame is used to prevent the excitation of most other metals Wide dynamic range with online dilution and/or preconcentration steps
FAAS	Transition metals, alkaline, alkaline-earth, Al, Pb	Sea, tap, waste, surface, ground, drinking, sewage	Flow rate in the flow system should be carefully selected and compared with the aspiration rate of the nebulizer Sample spends short time in contact with nebulizer and burner, so that the measuring system is readily rinsed between consecutive injections High sample throughput High tolerance to saline matrices Better adapted to SIA technique Improved sensitivity with regard to FAAS due to the rapid atomization of the entire sample
ETAAS	Transition metals, metalloid, Al, Pb	River, surface, lake, drinking, seawater	Analyte isolation/preconcentration with solid-phase extraction, adsorption or (co)precipitation Metalloid determination following entrapping of hydride species in precoated graphite tubes
ICP-AES/ICP-MS	Transition metals, alkaline earth, metalloid, Al, Pb, Bi	Drinking, tap, river, estuarine, wastewater	Fast multielemental analysis and isotope determinations Low tolerance to sample constituents (namely, ionized elements, mineral acids, organic solvents) Analyte isolation with solid-phase extraction, adsorption or (co)precipitation to overcome matrix interferences and enhance sensitivity
HG-AAS/HG-AFS	Metalloids, Pb, Sn, Cd, and Zn	Sea, lake, tap, drinking, surface, wastewater	Different dedicated configurations for gas-liquid separations FIA is preferred in HG-AFS aiming to feed the flame continuously with tetrahydroborate, which is a hydrogen source Implementation of various speciation strategies in flowing systems
CV-AAS/CV-AFS	Hg	River, sea, drinking, tap, surface, wastewater	Enrichment via gold-trap collection Speciation between inorganic mercury and organomercurials
Electrochemical ASV/PSA	Cu, Cd, Zn, Pb, In, Hg	Tap, sea, river, surface, drinking	Multielemental analysis and trace determinations SIA improves the sensitivity using flow reversals and also selectivity exploiting its inherent ability to perform medium exchange procedures
Potentiometry	Alkaline metals, Ag, Ca, Cd, Cu, Hg, Pb	Drinking, natural, tap, effluent	High tolerance to colored and turbid samples Coated tubular electrodes are specially suited for measurements in flowing systems High tolerance to colored and turbid samples Long-term performance due to the short contact time of the sample zone with the active surface



**Table 2** Continued

<i>Detection technique</i>	<i>Typical elements determined</i>	<i>Water sample</i>	<i>Remarks</i>
<i>Spectrophotometry</i>	Earth alkaline, Al, Fe, Mn, Co, Cr, Ni, Cu, Zn	River, lake, well, ground, tap, drinking, waste	Kinetics methods are commonly explored for minimizing interfering effects
			Typical single-element determination A plethora of kinetic catalytic procedures for trace metal monitoring has been developed Sorbent preconcentration sometimes used A variety of speciation schemes based upon the selective complexation of particular oxidation states has been designed
<i>Spectrofluorimetry</i>	Al, earth-alkaline metals, Cu, Fe, Mn, Zn, Hg	Tap, mineral, natural, drinking	Metal traces are determined exploiting the enhancement of the quantum yield of chelating agents via complexation with the analytes
<i>Chemiluminescence</i>	Cu, Co, Fe, Mn, Zn, Cd	Mineral, tap, drinking, seawater, estuarine	Use of the catalytic or inhibitory effect of metal traces on the luminol reaction with hydrogen peroxide or dissolved oxygen Kinetic discrimination and masking strategies are widely used for selectivity improvement Low detection limits are frequently achieved

FP, flame photometry; FAAS, flame atomic absorption spectrometry; ETAAS, electrothermal atomic absorption spectrometry; ICP-AES, inductively coupled plasma-atomic emission spectrometry; ICP-MS, inductively coupled plasma-mass spectrometry; HG, hydride generation; CV, cold vapor; AFS, atomic fluorescence spectrometry; ASV, anodic stripping voltammetry; PSA, potentiometric stripping analysis.

(namely, carbamates, benzimidazoles, and coumarins). For this purpose, chemical procedures based on inhibition of the activity of appropriate enzymes and immunochemical assays have been explored. Though both methods are used for screening of pollutants, the latter exhibits extremely high selectivity to be applied to quantitative determinations in natural waters. More recently, a plethora of luminescence flow injection systems capitalized on the intrinsic fluorescence or (electro)chemiluminescence of the target species has been proposed. The performance of online sorbent-extraction preconcentration procedures or solid-phase based flow-through optosensors provided determination limits down to the risk levels demanded by regulatory authorities. The selectivity of spectrofluorimetric methods is generally improved with the application of different strategies such as constant wavelength synchronous fluorescence, constant energy synchronous fluorescence, or variable angle scanning fluorescence aiming to record characteristic fluorescence spectra. The potential of these methodologies is even enhanced by the application of multivariate chemometric algorithms for treatment of the total luminescence data matrix, with the consequent resolution of multicomponent mixtures.

Surfactants are usually determined as a group of compounds, so that sophisticated chromatographic procedures are not needed. Ion-pair formation between both anionic and cationic surfactants with polar dyes, such as methylene blue and bromocresol purple, respectively, with or without subsequent inline solvent extraction has been assessed using spectrophotometric or spectrofluorimetric detection as shown in **Table 1**. Polyoxyethylene and amine ethoxylate-based nonionic surfactants in environmental waters have been determined by spectrophotometry following polar interactions with chromogenic agents, and by CL detection in the presence of sensitizers, respectively.

The determination of traces of phenolic species has been successfully accomplished by adaptation of batch standard methods based on the oxidative coupling of the analytes with 4-aminoantipyrine. The online preconcentration of the reaction products using solvent extraction, sorption/elution, or sorbent extraction-based optosensing techniques rendered detection limits below the maximum level recommended by EPA Directives. Interferences caused by colored aromatic amines can be successfully circumvented by analyte enrichment on reversed-phase materials prior to their oxidative coupling.

The potentialities of flowing stream systems for the determination of organic pollutants are thoroughly presented in the critical review entitled *Flow Injection Methods of Analysis for Waters: Organic Pollutants* by Danet *et al.* Readers are referred to this exhaustive review and references therein for further details and applications.

## Air Pollution Analysis

The analysis of precipitates (rain and snow), atmospheric water (cloud, fog, and dew), particulate matter, and gaseous compounds is an essential task for air pollution investigations. The application of FIA, and more recently SIA, for such purposes has found considerable interest because many environmental parameters can be determined with high reproducibility and sensitivity. In the analysis of precipitates and atmospheric water attractive features are the high sample throughput and the low consumption of sample, so that microsampling techniques can be fruitfully exploited in monitoring schemes. The micro-analytical capability of FIA has been demonstrated in the determination of hydrogen ion concentrations in single rain drops.

Compounds frequently monitored in precipitate samples are chloride, nitrate, fluoride, hydrogen peroxide, and sulfur species. The determination of sulfur compounds deserves particular attention in atmospheric analysis, especially in the elucidation of the generation, transportation, and conversion of the different species, as well as in understanding the formation of acid rain. Free and combined (namely, hydroxymethane sulfonic acid) sulfite has been commonly determined by flow injection spectrophotometric procedures using chromogenic reagents such as pararosaniline and organic disulfides. Excellent precision, reliability, and selectivity have been achieved through photometric, potentiometric, or spectrofluorimetric approaches coupled to in-line gas-diffusion separation of the generated sulfur dioxide after sample acidification. Iodimetric in-line determinations of sulfite with spectrophotometric or amperometric detection have also been reported. Improved selectivity and sensitivity are readily obtained via enzymatic methods comprising the electrochemical detection of hydrogen peroxide formed during sample passage through an immobilized sulfite oxidase-filled microcolumn. The detection chemistry used for other relevant analytes in precipitates and atmospheric waters is generally the same as that used in water analysis applications. Therefore, the readers are referred to the above sections as well as to Table 3, which summarizes

several flow methods successfully applied to air pollution surveillance.

Filter-collected airborne particulates can be analyzed by flow systems with great advantage since only minute liquid volumes are required. Thus, the amount of solvent used for dissolution, extraction, or digestion can be significantly reduced, consequently creating higher analyte concentration in the eluate, leachate, or extract. Sensitivity enhancement and improved time-resolution of the sampling procedure are then achieved. Flow techniques are also proven excellent avenues to determine anionic species and metal traces from a single filter following sequential treatment with water and acid in a pressurized digestion bomb.

The determination of gaseous compounds (namely, sulfur dioxide, nitrogen oxides, hydrogen fluoride, and ammonia) at atmospheric levels is generally performed ensuing batch trapping/preconcentration in suitable absorber liquids, which are, whereupon, injected in the flow manifold. Yet, a particularly promising approach for gas analysis in the presence of concomitant particulate matter is the exploitation of membrane-based gas sampling coupled to flow systems. In this method, the gas sample is fed through the donor side of a gas-diffusion module, while a liquid absorber is used as a trapping solution. The acceptor liquid is either delivered continuously through the unit toward a downstream detector or is halted for a preset time by stopping the liquid-driver or by implementation of a closed-loop. The latter option not only enables the isolation of the gaseous species from matrix constituents but also the analyte enrichment within the receiver solution. Direct introduction of a gas sample plug into a flowing liquid reagent stream followed by detection of the reaction products formed at the gas-liquid interface has been successfully assessed for both sulfur dioxide and nitrogen dioxide monitoring.

## Analysis of Soils, Sediments, Fertilizers, Animal Feed, and Plant Materials

Flow techniques, mainly FIA, have become consolidated as one of the most efficient and versatile methods in agricultural laboratories. The high sample throughput of FIA meets admirably the requirements of soil laboratories, which are overwhelmed in the planting season. The wide concentration ranges found in agricultural samples, both within the same compound and from one to another, can be easily covered by flowing-stream systems in making use of in-line dilution and preconcentration schemes in

**Table 3** Analytical features of relevant flowing-stream methods applied to air pollution surveillance

Analyte	Sample	Method/derivatization reaction	Detection	Typical working range	Features/sample collection
Sulfur dioxide	Air	Pararosaniline	Photometry	0.01–20 mg m <sup>-3</sup>	Trapping in formaldehyde
	Air	<i>N</i> -Acridinylmaleimide	Spectrofluorimetry	> 2.5 µg l <sup>-1</sup>	Trapping in formaldehyde
	Air	Barium-dimethylsulphonazo-III	Photometry	0.2–14 mg l <sup>-1</sup>	Hydrogen peroxide absorber
	Air	5,5'-Dithiobis(2,2'-dinitrobenzoic acid)	Photometry	> 35 µg m <sup>-3</sup>	Online sampling/gas permeation denuder
	Air	Direct detection	Conductimetry	0.16–3.2 mg m <sup>-3</sup>	Capillary denuder/preconcentration
	Air	Electrogenerated Mn(III) or Ag(II)	Chemiluminescence	0.02–0.64 mg l <sup>-1</sup>	
Ammonia	Air	Direct detection	Potentiometry (NH <sub>4</sub> -ISE)	5.0–100 µg m <sup>-3</sup>	Membrane-based gas sampling
	Air	Berthelot	Photometry	> 0.15 µg m <sup>-3</sup>	Phosphoric acid coated diffusion denuder
Nitrogen oxides	Air	Oxidation of ferroin	Amperometry	1.0–10 mg m <sup>-3</sup>	Direct gas injection
	Air	Griess reaction	Photometry	2.0–500 µg m <sup>-3</sup>	Membrane-based gas sampling
	Air	Direct detection	Coulometry	0.01–2.0 µg m <sup>-3</sup>	Direct gas sampling
Hydrogen chloride	Gas emissions	Direct detection	Potentiometry	1.0–60 mg l <sup>-1</sup>	Use of a Cl-ISE
Volatile arsenic compounds	Air	Arsenite detection after hydrolysis	PSA	0.5–50 µg m <sup>-3</sup>	Gas permeation membrane-based sampling unit
Hydrogen peroxide	Air	Hematin- <i>p</i> -cresol	Spectrofluorimetry	> 10 ng m <sup>-3</sup>	Diffusion scrubber, preconcentration
	Rain	<i>N</i> -Ethyl- <i>N</i> -(sulfopropyl)aniline	Photometry	7.0 × 10 <sup>-3</sup> –1.4 mg l <sup>-1</sup>	
	Rain	Luminol/oxidase (or hematin)	Chemiluminescence	1.0 × 10 <sup>-4</sup> –0.34 mg l <sup>-1</sup>	
Formaldehyde	Air/fog	5,5-Dimethylcyclohexane-1,3-dione/ammonium acetate	Spectrofluorimetry	25–100 µg l <sup>-1</sup>	Interference from sulfite is eliminated by oxidation
Sulfite	Air	Pararosaniline/sulfite	Photometry	0.1–20 mg l <sup>-1</sup>	Masking transition metal ions with chelating agents Gas diffusion separation in an acidified donor stream Membrane-based separation
	Cloud	Pararosaniline/formaldehyde	Photometry	> 13 µg l <sup>-1</sup>	
	Rain	4,4-Dithiodipyridine	Photometry	0.25–10 mg l <sup>-1</sup>	
Nitrate	Rain, fog, snow	Direct measurement	Amperometry	0.08–0.8 mg l <sup>-1</sup>	Membrane-based separation
	Rain	Hydrazine reduction, Griess reaction	Photometry	1.0–10 mg l <sup>-1</sup>	
	Rain	2,3-Diaminonaphthalene/UV irradiation	Spectrofluorimetry	0.03–0.7 mg l <sup>-1</sup>	
Ammonium	Rain	Luminol/chlorine	Electrochemiluminescence	0.014–1.4 mg l <sup>-1</sup>	
	Rain	<i>o</i> -Phthaldialdehyde/2,4,6-trichlorophenyl oxalate	Chemiluminescence	0.01–5.0 mg l <sup>-1</sup>	
Sulfate	Rain	Barium chloride	Turbidimetry	0.1–2.0 mg l <sup>-1</sup>	

Continued

Table 3 Continued

Analyte	Sample	Method/derivatization reaction	Detection	Typical working range	Features/sample collection
Acidity	Atmospheric precipitations	Barium/orthanilic acid	Photometry	0.5–20 mg l <sup>-1</sup>	Single rain drop analysis
	Rain	Barium chloranilate	Photometry	4.0–100 mg l <sup>-1</sup>	
	Rain	Lead-nitrate ISE	Potentiometry	1.0–100 mg l <sup>-1</sup>	
	Rain	Acid-base indicator, titration	Photometry	pH 3–6	
Anionic species (NH <sub>4</sub> <sup>+</sup> , NO <sub>3</sub> <sup>-</sup> , Cl <sup>-</sup> , SO <sub>4</sub> <sup>2-</sup> )	Rain	Liquid-membrane ISE	Potentiometry	pH 2.7–5	Microsampling/filter extraction
	Airborne particulate matter	ISE and standard photometric methods	Potentiometry/photometry	µg l <sup>-1</sup> to mg l <sup>-1</sup>	
	Airborne particulate matter	Direct injection	FAAS	µg l <sup>-1</sup> to mg l <sup>-1</sup>	
	Airborne particulate matter	Direct injection	FAAS	Low µg l <sup>-1</sup>	
Metal ions (Ca, Mg, Fe, Al, Pb, Cd, Zn)	Airborne particulate matter	Direct injection	ETAAS/ICP-MS	>0.1 mg m <sup>-3</sup>	Ultrasonic acidic leaching from membrane filter
Lead/cadmium	Airborne particulate matter	Direct injection	Photometry	7 × 10 <sup>-3</sup> –12 mg l <sup>-1</sup>	Speciation protocols/sequential extraction procedures
Nickel	Airborne particulate matter	2,2'-Dipyridyl-2-pyridylhydrazone			Online solvent extraction
Palladium	Airborne particulate matter				

ISE, ion-selective electrode; PSA, potentiometric stripping analysis; FAAS, flame atomic absorption spectrometry; ETAAS, electrothermal atomic absorption spectrometry; ICP-MS, inductively coupled plasma mass spectrometry.

combination with almost any detection technique. Flow methodologies are well accepted since the underlying detection chemistry of established standard methods can be maintained and the extremely low sample volumes required simplify greatly the analytical protocols and reduce the chemical consumption. Since flow techniques are inherently liquid-phase analysis methodologies, constituents in solid samples are usually transferred into the liquid state prior to analysis using appropriate batch pretreatment procedures. Then, the methods described in the foregoing sections for monitoring environmental parameters in waters are equally applicable to the determination of key species in soil, sediments, and plant materials, such as nutrients and metal traces. The coupling of flow systems with atomic absorption spectrometric techniques is especially beneficial for the latter analytes. The high tolerance exhibited to elevated salt and acid contents and differing sample viscosities makes it possible to admit extracts and digests directly without further sample manipulation. Readers are referred to Table 4 for further details related to particular methods devoted to agricultural and solid sample analysis.

In recent years, many efforts have been focused on the design and characterization of sample processing units coupled with flowing systems aiming to enable the direct introduction and treatment of solid samples of environmental origin in an automated fashion. In this respect, various sample pretreatment techniques including electrolytic dissolution, online dialysis/filtration, dynamic-gas extraction, liquid-liquid extraction, ultrasound-assisted extraction, and pervaporation-based procedures have been successfully implemented in flow systems. Very recently, the interfacing of flow injection/continuous flow systems with microwave-assisted extraction, supercritical carbon dioxide extraction, or subcritical fluid extraction treatments has proven useful for the determination of organic pollutants (namely, polycyclic aromatic hydrocarbons, phenolic species, substituted carbamates, and chlorinated biphenyls) as well as metal/metalloid traces in soils, sediments, seeds, and plant tissues as shown in Table 4. These methods usually provide similar efficiencies than US-EPA methods but with dramatic reduction in both the extraction time and sample handling.

It should be also stressed that online dynamic leaching protocols are currently being employed by several researchers as an alternative to steady-state measurements, since kinetic information from the extraction process is thus obtained. With respect to sampling modalities, the exploitation of slurry sampling strategies as well as the implementation of microcolumn reactors packed with solid materials should be underlined.

**Table 4** Analytical features of relevant flowing-stream methods applied to the determination of organic pollutants, nutrients, anionic species, and metal traces in agricultural samples

Analyte	Sample	Method/derivatization reaction	Detection	Typical working range	Features/sample pretreatment
Polycyclic aromatic hydrocarbons	Soil	HPLC/intrinsic properties	Spectrofluorimetry	0.030–0.375 mg l <sup>-1</sup>	Continuous subcritical extraction (micellar medium-water)/reversed-phase preconcentration
	Soil	Intrinsic properties	Fiber-optic spectrofluorimetry	2.0–24 µg per g	Supercritical extraction
N-Methylcarbamates	Soil and sediments	HPLC/intrinsic properties	Spectrofluorimetry	0.2–1.2 mg l <sup>-1</sup>	Online subcritical water extraction
	Soil and fruits	LC/o-Phthaldialdehyde (postcolumn)	Spectrofluorimetry	0.5–10 mg l <sup>-1</sup>	Online ultrasound assisted extraction/inline filtration/solid-phase extraction
Chlorophenoxy acids	Soil	HPLC/intrinsic properties	UV-spectrophotometry	10–50 mg l <sup>-1</sup>	Continuous subcritical water extraction/inline filtration/reversed-phase extraction
Phenol compounds	Soil	Direct injection of eluates	GC–MS	0.05–5.0 µg per g	Online leaching with alkaline solution/ filtration/reversed-phase extraction (semiautomated method)
Ammonium	Fertilizer and soil	pH indicator/gas-diffusion	Spectrophotometry	0.5–20 mg l <sup>-1</sup>	Extraction in a hydrochloric acid medium
Nitrate	Plant tissue	Reduction to nitrite/Griess–Illosvay reaction	Spectrophotometry	> 0.1 mg l <sup>-1</sup>	Hot water extraction
Organic nitrogen (Kjeldahl)	Soil and fertilizer	Direct detection	Potentiometry (NO <sub>3</sub> -ISE)	6.2–62 mg l <sup>-1</sup>	Aqueous extraction
	Plant materials	AgCl microcolumn/displacement reaction	FAAS	1.0–200 µg l <sup>-1</sup>	Kjeldahl digestion
	Fertilizer and plant	Berthelot reaction	Spectrophotometry	0.5–5% (w/w) N	Kjeldahl digestion
	Animal feed, plant, and fertilizer	Gas-diffusion	Conductimetry	0.5–50 mg l <sup>-1</sup>	Kjeldahl digestion
Orthophosphate	Soil	Molybdenum blue	Spectrophotometry	0.05–100 mg l <sup>-1</sup>	Extraction in a hydrogencarbonate medium
Total dissolved phosphorus	Plant tissue	Zinc chloride	Turbidimetry	5.0–60 mg l <sup>-1</sup>	Acid digestion
	Fertilizer and animal feed	Molybdenum blue	Spectrophotometry	5.0–250 mg l <sup>-1</sup>	Online UV photo-oxidation
Sulfate	Soil (or plant)	Barium chloride or lead phosphate	Turbidimetry	1.0–50 mg l <sup>-1</sup>	Extraction in a calcium phosphate medium (or acid digestion)
Sulfide	Soil	Direct detection	Potentiometry	0.3–50 µg per g	Online pervaporation
Boron	Soil and fertilizer	Azomethine H	Spectrophotometry	5.0–50 µg per g	Solid sample introduction, online ultrasonic leaching
pH	Plant	Azomethine H	Spectrophotometry	0.1–6.0 mg l <sup>-1</sup>	Dry-ashing
	Soil	Capillary glass electrode	Potentiometry	pH 4–11	Extraction in a calcium chloride medium
Silicate	Soil	Molybdenum blue	Spectrophotometry	> 30 µg l <sup>-1</sup>	Various extraction procedures
Chloride	Soil	Potentiometric pseudotitration	Potentiometry	1.0–100 mg l <sup>-1</sup>	Aqueous extraction
	Soil	Dialysis separation	Potentiometry (Cl-ISE)	5.0–5000 mg l <sup>-1</sup>	Slurry sampling/online removal of high molecular weight matrix constituents

*Continued*

**Table 4** Continued

<i>Analyte</i>	<i>Sample</i>	<i>Method/derivatization reaction</i>	<i>Detection</i>	<i>Typical working range</i>	<i>Features/sample pretreatment</i>
Chloride, nitrate, and sulfate Potassium	Soil Fertilizer and soil	Capillary electrophoresis Direct detection	Spectrophotometry Potentiometry (K-ISE)	0.1–100 mg l <sup>-1</sup> 1.6–400 mg l <sup>-1</sup>	Online water extraction/filtration Extraction in a sodium acetate solution
Calcium	Animal feed Soil (or plant)	$\alpha$ -Cresolphthalein complexone Glyoxal-bis(2-hydroxyanil)	Spectrophotometry Spectrophotometry	0.05–5% (w/w) Ca < 100 mg l <sup>-1</sup>	Dry-ashing Extraction in a potassium chloride medium (or acid digestion)
Potassium, calcium, nitrate, and chloride	Soil	Multi-ion sensor	Potentiometry	mmol l <sup>-1</sup> level	Extraction in a sodium acetate medium
Potassium, calcium, and magnesium	Plant	Direct injection	FAAS (or FES)	mg l <sup>-1</sup> level	Dry-ashing
Aluminum	Soil	Pyrocatechol violet	Spectrophotometry	0.1–2.0 mg l <sup>-1</sup>	Extraction in a potassium chloride medium
	Plant	Eriochrome cyanine R	Spectrophotometry	0.5–15 mg l <sup>-1</sup>	Acid digestion
	Plant	Direct injection	FAAS	> 75 $\mu$ g l <sup>-1</sup>	Acid digestion/solid-phase preconcentration
Iron	Plant	1,10-Phenanthroline and ascorbic acid	Spectrophotometry	0.1–30 mg l <sup>-1</sup>	Acid digestion
	Soil (or plant)	Direct injection	FAAS	0.4–5.0 mg l <sup>-1</sup>	Aqueous extraction (or acid digestion)
Copper	Plant	Dithiocarbamate or dithiophosphate chelating agents	FAAS	2–900 $\mu$ g l <sup>-1</sup>	Online solvent (or sorbent) extraction of chelates
	Plant	2,2'-Dipyridyl ketone hydrazone and hydrogen peroxide	Spectrofluorimetry	0.2–300 $\mu$ g l <sup>-1</sup>	Catalytic oxidation
	Animal feed	Dialysis with diffusate preconcentration	FAAS	0.025–0.25% (w/w) Cu	Aqueous extraction/injection of particle containing samples
Nickel	Plant tissue	Br-PADAP	Spectrophotometry	0.025–0.5 mg l <sup>-1</sup>	Masking agents and preconcentration on PAN-C <sub>18</sub> column
Chromium	Soil	1,5-Diphenylcarbazide	Spectrophotometry	0.25–10 mg l <sup>-1</sup>	Online ultrasound-assisted extraction/anion-exchange preconcentration
Manganese	Plant (or soil)	Periodate	Spectrophotometry	0.14–1.3 mg l <sup>-1</sup>	Acid digestion (or aqueous extraction)
	Plant	Diphenylcarbazone/O <sub>2</sub> in the presence of triethanolamine	Spectrophotometry	0.1–40 mg l <sup>-1</sup>	Catalytic oxidation
Lead	Plant	Direct injection	ETAAS	8.0–40 $\mu$ g l <sup>-1</sup>	Online acidified subcritical water or ultrasound-assisted digestion (semiautomatic procedure)



Zinc	Plant	NaCl and Zincon	Spectrophotometry	0.1–2.0 mg l <sup>-1</sup>	Dry-ashing/preconcentration of zinc chlorocomplexes
Iron, zinc, manganese, calcium, and magnesium	Plant	Direct injection	FAAS	µg g <sup>-1</sup> level	Mild calcinations/slurry sampling/online dilution for Ca and Mg
Cadmium	Plant	Direct injection	ETAAS	0.1–1.0 µg l <sup>-1</sup>	Online ultrasound-assisted digestion or acidified subcritical water (semiautomatic procedure)
	Fertilizer and plant	Iodide and malachite green	Spectrophotometry	10–100 µg l <sup>-1</sup>	Masking agents/kinetic discrimination/ion-exchange separation
Molybdenum	Plant	Thiocyanate complex	Spectrophotometry	0.05–1.0 mg l <sup>-1</sup>	Liquid–liquid extraction
	Plant (or soil)	Iodide/hydrogen peroxide	Spectrophotometry (or BI)	1.0–40 µg l <sup>-1</sup>	Catalytic oxidation/cation-exchange for metal ion removal
Cobalt	Plant	1-Nitroso-2-naphthol-3,6-disulphonate	Spectrophotometry (or FAAS)	> 3.5 µg l <sup>-1</sup> (or > 0.44 µg l <sup>-1</sup> )	Microwave-assisted digestion/solid-phase extraction
Metalloid species	Plant	Sodium tetrahydroborate	HG-AAS	µg l <sup>-1</sup> level	Acid digestion
	Soil	Sodium tetrahydroborate	HG-AFS	ng l <sup>-1</sup> to µg l <sup>-1</sup>	Online pervaporation/speciation studies
Mercury	Soil	Tin(II) chloride in a hydrochloric acid medium	CV-AAS	1.0–20 µg l <sup>-1</sup>	Microwave assisted digestion or wet-oxidant digestion for total mercury determination
	Soil	Tin(II) chloride in a hydrochloric acid medium	CV-AAS	50–500 ng l <sup>-1</sup>	Online preconcentration/reductive elution
	Soil	Tin(II) chloride in a hydrochloric acid medium	CV-AFS	> 0.01 µg g <sup>-1</sup>	Online acidified subcritical water

ISE, ion-selective electrode; FAAS, flame atomic absorption spectrometry; ETAAS, electrothermal atomic absorption spectrometry; FES, flame emission spectrometry; HG, hydride generation; CV, cold vapor; AFS, atomic fluorescence spectrometry; BI, biamprometry; HPLC, high-performance liquid chromatography; LC, liquid chromatography; GC–MS, gas chromatography–mass spectrometry; Br-PADAP, 2-(5-bromo-2-pyridylazo)-5-(diethylamino)phenol; C<sub>18</sub>, octadecyl-chemically modified silicagel; PAN, 1-(2-thiazolylazo)-2-naphthol.

See also: **Air Analysis:** Sampling. **Arsenic.** **Elemental Speciation:** Waters, Sediments, and Soils. **Environmental Analysis.** **Extraction:** Microwave-Assisted Solvent Extraction; Pressurized Fluid Extraction; Supercritical Fluid Extraction; Solid-Phase Extraction. **Fertilizers.** **Flow Injection Analysis:** Detection Techniques. **Fluorescence:** Environmental Applications. **Geochemistry:** Soil, Major Inorganic Components; Soil, Minor Inorganic Components; Soil, Organic Components. **Ion-Selective Electrodes:** Water Applications. **Lead.** **Membrane Techniques:** Pervaporation. **Mercury.** **Nitrogen.** **Pesticides.** **Phosphorus.** **Quality Assurance:** Water Applications. **Sample Handling:** Automated Sample Preparation. **Selenium.** **Sequential Injection Analysis.** **Spectrophotometry:** Inorganic Compounds; Organic Compounds. **Sulfur.** **Surfactants and Detergents.** **Tin.** **Voltammetry:** Inorganic Compounds. **Water Analysis:** Potable Water; Freshwater; Seawater – Organic Compounds; Seawater – Dissolved Organic Carbon; Seawater – Inorganic Compounds; Industrial Effluents; Sewage; Organic Compounds; Biochemical Oxygen Demand; Chemical Oxygen Demand.

## Further Reading

Danet AF, Cheregi M, Martínez-Calatayud J, García-Mateo JV, and Enein HYA (2001) Flow injection

methods of analysis for waters: Inorganic species. *Critical Reviews in Analytical Chemistry* 31: 191–222.

Danet AF, Cheregi M, Martínez-Calatayud J, García-Mateo JV, and Enein HYA (2003) Flow injection methods of analysis for waters: Organic pollutants. *Critical Reviews in Analytical Chemistry* 33: 57–68.

Fang Z-L (1995) Environmental and agricultural applications. In: *Flow Injection Atomic Absorption Spectrometry*, ch. 12, pp. 229–244. Chichester: Wiley.

Lenahan CE, Barnett NW, and Lewis SW (2002) Sequential injection analysis. *Analyst* 127: 997–1020.

Miró M, Estela JM, and Cerdà V (2003) Application of flowing stream techniques to water analysis. Ionic species: Dissolved inorganic carbon, nutrients, and related compounds. *Talanta* 60: 867–886.

Miró M, Estela JM, and Cerdà V (2004) Application of flowing stream techniques to water analysis. General quality parameters and anionic compounds: Halogenated, sulphur and metalloid species. *Talanta* 62: 1–15.

Trojanowicz M (2000) Applications of flow injection methods in routine analysis. In: *Flow Injection Analysis: Instrumentation and Applications*, ch. 8, pp. 298–338 and 354–370. Singapore: World Scientific.

Zhi ZL, Rios A, and Valcárcel M (1996) Direct processing and analysis of solid and other complex samples with automatic flow injection systems. *Critical Reviews in Analytical Chemistry* 26: 239–260.

## Clinical and Pharmaceutical Applications

**J M Calatayud**, Universidad de Valencia, Burjassot, Spain

**M C Icardo**, Polytechnic University of Valencia, Valencia, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Recent analytical literature reveals that a large number of researchers are interested in flow injection analysis (FIA) as an analytical ‘working tool’. One of the early goals of research on FIA was to develop an analytical application that was continuous, and not batch-wise. The alternative to classical analysis was desired to be cost-effective, rapid, simple, and flexible, to allow a ‘universal’ application. The pioneering advances in this period led to new possibilities for implementing various assays and solutions for problems that were very difficult to solve using ‘classical’ batch methods. Once FIA and other automated techniques were in place, research shifted to the development of new analytical procedures.

At present, there is no lack of commercially available FIA instruments that can be interfaced with sampling and data acquisition devices or used in combination with other flow methodologies or more recent flow devices. Many of these instruments (fluorimeters, atomic absorption spectrometers, etc.) offer compact setup suited to specific purposes.

The techniques of flow analysis can be applied to the pharmaceutical and clinical analysis. This entails employing ‘classical’ detectors and chemical strategies. Solution chemistry and ultraviolet (UV)–visible (Vis) spectrophotometry tools are frequently used. The selection of the actual procedure is determined by the type of detector used, and may in general include use of spectrometric detectors, followed by electroanalytical techniques, and separation and biochemical procedures. Special attention is paid to ‘new’ devices used in flow manifolds and to detectors such as chemiluminescence detectors (in which researchers have shown increasing interest). This article discusses these procedures in some detail.

See also: **Air Analysis:** Sampling. **Arsenic.** **Elemental Speciation:** Waters, Sediments, and Soils. **Environmental Analysis.** **Extraction:** Microwave-Assisted Solvent Extraction; Pressurized Fluid Extraction; Supercritical Fluid Extraction; Solid-Phase Extraction. **Fertilizers.** **Flow Injection Analysis:** Detection Techniques. **Fluorescence:** Environmental Applications. **Geochemistry:** Soil, Major Inorganic Components; Soil, Minor Inorganic Components; Soil, Organic Components. **Ion-Selective Electrodes:** Water Applications. **Lead.** **Membrane Techniques:** Pervaporation. **Mercury.** **Nitrogen.** **Pesticides.** **Phosphorus.** **Quality Assurance:** Water Applications. **Sample Handling:** Automated Sample Preparation. **Selenium.** **Sequential Injection Analysis.** **Spectrophotometry:** Inorganic Compounds; Organic Compounds. **Sulfur.** **Surfactants and Detergents.** **Tin.** **Voltammetry:** Inorganic Compounds. **Water Analysis:** Potable Water; Freshwater; Seawater – Organic Compounds; Seawater – Dissolved Organic Carbon; Seawater – Inorganic Compounds; Industrial Effluents; Sewage; Organic Compounds; Biochemical Oxygen Demand; Chemical Oxygen Demand.

## Further Reading

Danet AF, Cheregi M, Martínez-Calatayud J, García-Mateo JV, and Enein HYA (2001) Flow injection

methods of analysis for waters: Inorganic species. *Critical Reviews in Analytical Chemistry* 31: 191–222.

Danet AF, Cheregi M, Martínez-Calatayud J, García-Mateo JV, and Enein HYA (2003) Flow injection methods of analysis for waters: Organic pollutants. *Critical Reviews in Analytical Chemistry* 33: 57–68.

Fang Z-L (1995) Environmental and agricultural applications. In: *Flow Injection Atomic Absorption Spectrometry*, ch. 12, pp. 229–244. Chichester: Wiley.

Lenahan CE, Barnett NW, and Lewis SW (2002) Sequential injection analysis. *Analyst* 127: 997–1020.

Miró M, Estela JM, and Cerdà V (2003) Application of flowing stream techniques to water analysis. Ionic species: Dissolved inorganic carbon, nutrients, and related compounds. *Talanta* 60: 867–886.

Miró M, Estela JM, and Cerdà V (2004) Application of flowing stream techniques to water analysis. General quality parameters and anionic compounds: Halogenated, sulphur and metalloid species. *Talanta* 62: 1–15.

Trojanowicz M (2000) Applications of flow injection methods in routine analysis. In: *Flow Injection Analysis: Instrumentation and Applications*, ch. 8, pp. 298–338 and 354–370. Singapore: World Scientific.

Zhi ZL, Rios A, and Valcárcel M (1996) Direct processing and analysis of solid and other complex samples with automatic flow injection systems. *Critical Reviews in Analytical Chemistry* 26: 239–260.

## Clinical and Pharmaceutical Applications

**J M Calatayud**, Universidad de Valencia, Burjassot, Spain

**M C Icardo**, Polytechnic University of Valencia, Valencia, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Recent analytical literature reveals that a large number of researchers are interested in flow injection analysis (FIA) as an analytical ‘working tool’. One of the early goals of research on FIA was to develop an analytical application that was continuous, and not batch-wise. The alternative to classical analysis was desired to be cost-effective, rapid, simple, and flexible, to allow a ‘universal’ application. The pioneering advances in this period led to new possibilities for implementing various assays and solutions for problems that were very difficult to solve using ‘classical’ batch methods. Once FIA and other automated techniques were in place, research shifted to the development of new analytical procedures.

At present, there is no lack of commercially available FIA instruments that can be interfaced with sampling and data acquisition devices or used in combination with other flow methodologies or more recent flow devices. Many of these instruments (fluorimeters, atomic absorption spectrometers, etc.) offer compact setup suited to specific purposes.

The techniques of flow analysis can be applied to the pharmaceutical and clinical analysis. This entails employing ‘classical’ detectors and chemical strategies. Solution chemistry and ultraviolet (UV)–visible (Vis) spectrophotometry tools are frequently used. The selection of the actual procedure is determined by the type of detector used, and may in general include use of spectrometric detectors, followed by electroanalytical techniques, and separation and biochemical procedures. Special attention is paid to ‘new’ devices used in flow manifolds and to detectors such as chemiluminescence detectors (in which researchers have shown increasing interest). This article discusses these procedures in some detail.

## Spectrometric Methods

### UV-Vis Spectrophotometric Methods: Homogeneous Systems

Drug analyses based on absorption spectrophotometry usually involve an enzyme catalyzed reaction, ion pair formation (with the aid of liquid-liquid extraction or turbidimetry) or a redox reaction (normally oxidation in solution or an oxidizing bed). Examples of homogeneous applications are listed in Table 1. Zn determination (Table 1) can be used with high selectivity against other metal ions through a differential demasking technique. In the first step, all metal ions (Zn included) are masked by adding KCN; then the Zn(II) is quickly released (kinetic selectivity) using formaldehyde. In another application a double-injection valve is used to insert an enzyme and the analyte fosfestrol into a single carrier stream; then both boluses merge due to dispersion. This approach reduces enzyme consumption.

Most multiparametric procedures are restricted to two active components, and the main challenge is in using different reagents and detection wavelengths. The use of photodiode array spectrophotometers and, more recently, multichannel fiber optic spectrophotometers facilitates this approach. Solenoid valves, to control the flow system, have proved to be very useful in the development of techniques for simultaneous determination of Zn, Fe, Cu, Ca, and Mg in pharmaceutical drugs. Derivative

spectrophotometry should be used to analyze non-complex matrices, with components having overlapped spectra; examples include simultaneous determination of adrenaline and noradrenaline in synthetic mixtures, and sulfamethoxazole and trimethoprim in pharmaceutical formulations using the first-order derivative. Another example is the hydrochlorothiazide-losartan pair, which can be determined in an alcoholic medium using the fourth-derivative spectrum.

A different strategy has been used in the simultaneous determination of ascorbic acid and cysteine. The system uses a double-beam spectrophotometer. When the sample converges with Fe(III) and phenanthroline, the signal obtained corresponds to ascorbic acid. After passing through the flow cell, there is mixing with Cu(II), which accelerates the reaction between Fe(III) and cysteine, with practically no effect on the reaction between Fe(III) and ascorbic acid. When the mixture arrives at the second flow cell, the signal corresponds to the total amount of ascorbic acid and cysteine.

The simultaneous injection of two sample volumes into the same carrier ('single-line double-injection') allows simultaneous determination of Fe(II) and Fe(III) using the colored complex Fe(II)-2,2'-dipyridyl-2-pyridylhydrazone. Fe(III) is reduced online with ascorbic acid before being injected into the carrier. In this way, two peaks are obtained, one from the Fe(II) and other from the total Fe.

**Table 1** Some spectrophotometric (homogeneous systems) applications of FIA

Analyte	Sample	Procedure
Thiamine	Pharmaceutical formulation	Oxidation with hexacyanoferrate(III)
Ascorbic acid	Pharmaceutical formulation	Oxidation with permanganate
Ascorbic acid; vitamin K <sub>3</sub> ; promazine	Pharmaceutical formulation	Oxidation with Fe(III). Determination of released Fe(II) through a colored complex
Cefadroxil; cephalosporins	Pharmaceutical formulation	Hydrolysis in basic medium + oxidation with Fe(III). Determination of released Fe(II) by a colored complex
Fosfestrol; forfomycin	Pharmaceutical formulation and urine	Alkaline-enzymatic hydrolysis or digestion with ammonium persulfate. Determination of orthophosphate through the molybdenum blue reaction
Leucogen	Tablets	Hydrolysis in basic medium
Amoxicillin and ampicillin	Pharmaceutical formulation	Hydrolysis in micellar medium and catalyzed by Cu(II)
Zn(II)	Serum, human hair, and Pharmaceutical formulation	Complex formation with 2,2'-dipyridyl-2-pyridylhydrazone in strong alkaline medium (selective demasking of the KCN complex with formaldehyde)
Creatinine	Urine	Jaffe reaction
Phenylephrine hydrochloride	Pharmaceutical formulation	Ferricyanide + 4-aminoantipyrine
Tyrosine	Urine	$\alpha$ -Nitroso- $\beta$ -naphthol in acid medium
Captopril	Pharmaceutical formulation	Inhibition of the complex Co(II)-2,2'-dipyridyl-2-pyridylhydrazone
Ascorbic acid	Pharmaceutical formulation	Inhibition of the oxidation of pyrogallol red by iodate
Sulfanilamides	Pharmaceutical formulation and blood	As 4,6-dinitrobenzofuroxane derivatives
Doxazosin mesylate	Pharmaceutical formulation	Native absorption in UV at pH 4 in 10% methanol
Se(IV)	Tablets and pig kidney	Catalysis of the gallocyanine reduction by Na <sub>2</sub> S

Methods such as standard addition only provide good results with a relatively simple matrices. One of the main problems when a first-order multivariate model is used is the presence of unknown interferences. Mathematical models have become very important for solving this problem; an example is the determination of five pollutants of the chlorophenol family in urine. The effect of the matrix is minimized by including, in the calibration step, standard samples containing the analytes in the presence of the interfering matrix. The calibration set includes 60 standard samples: 50 samples of chlorophenols in water and 10 of lyophilized urine.

### Heterogeneous Solid-Liquid Systems (Turbidimetry)

The suspension (precipitate) is usually formed by injecting the sample directly into a buffered carrier-reagent stream. The required manifold is essentially a single channel, and the reactor is normally short. This method provides a straightforward and inexpensive alternative for measuring the active components in pharmaceutical dosages using various organic reagents or metal complexes such as counterions. No liquid-liquid separation or subsequent extraction of the organic phase is required. A surfactant can be added to stabilize the suspension and eliminate frequent flushing injections.

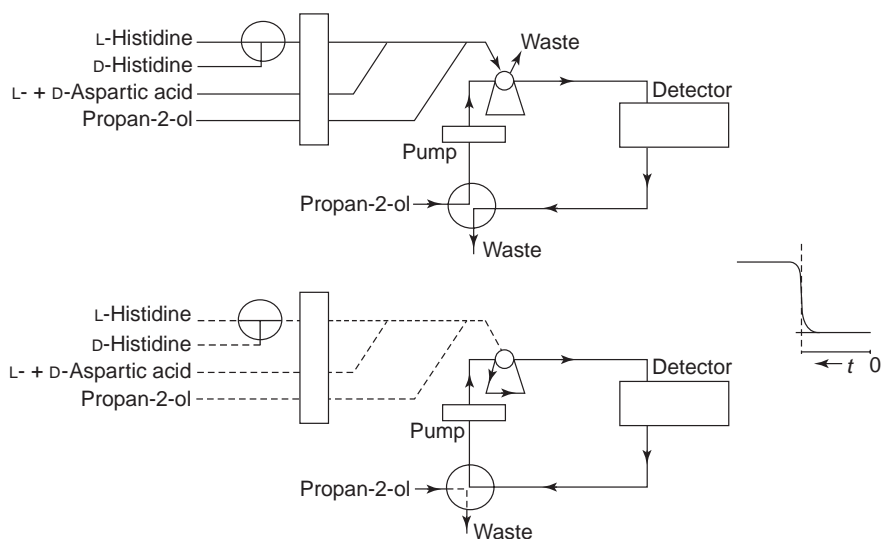
The precipitation of fibrinogen can be used to analyze human blood samples by injecting the sample into a carrier stream of guanidine hydrochloride/titriplex III that merges with ammonium sulfate (the reagent) dissolved in the same buffer.

A final example is depicted in **Figure 1**. The enantiomeric discrimination of l- and d-aspartic acids is based on the inhibition of the process of crystallization of l- and d-histidine. A sample with the pH adjusted at 3–10 (the variation in pH resulted in a plateau) is mixed with l- or d-histidine, and the resulting mixture is injected into a carrier of propan-2-ol and entrapped in a closed-loop circuit. The absorbance versus time plot, which reflects the crystal growth, is recorded at 550 nm. The induction period before the onset of crystallization is used to calculate the analyte concentration. A similar strategy can be used employed for the resolution of structural isomers of arginine and ornithine in pharmaceutical formulations.

### Fluorimetric Methods

Fluorimetry was considered in the 1950s as the ‘natural’ detector for pharmaceuticals, due to its improved selectivity and sensitivity compared with UV-Vis absorption. Recent FIA applications include the determination of diazepam, nitrazepam, and oxazepam in pharmaceutical formulations using acidic hydrolysis and fluorimetric detection. Oxidation with Ce(IV) and measurement of the fluorescence from the released Ce(III), which can be considered as a classical strategy, is an appropriate technique for mixtures of amoxycillin and clavulanic acid where kinetic data are used in combination with partial least-squares multivariate calibration.

Another classical procedure is used for many pharmaceutical compounds: when a primary or secondary amine is present, it reacts with o-phthaldehyde and a thiol compound (like mercaptoethanol or N-acetylcysteine), producing a fluorescent isoindol.



**Figure 1** Flow systems for enantiomeric discrimination of l- and d-aspartic acids.



The secondary amine or amides require a prior reaction with sodium hypochlorite to be converted into a primary amine.

One of the FIA fluorimetric methods for clinical applications involves the reduction of the herbicide diquat using sodium dithionite in a stopped-flow modality; the analyte yields a highly fluorescent product and it can be applied to the analysis of blood serum. Cysteine has also been determined through fluorescence quenching of a complex of Cu(II) with 5-(3-fluoro-4-chlorophenylazo)-8-benzenesulfonamidoquinoline.

### Chemiluminescence Methods

The iterative strategy of investigating the effect of an analyte on luminol or any other effective light emitter, has given way to direct study of the oxidation of the analyte by a strong oxidative reagent. The main goal is to find new chemiluminescent reactions for simpler and more sensitive procedures. The limits to selectivity are imposed by the high chemical activity of the oxidant compared to other components present in the sample. The simplicity of the procedure and the fast kinetics of oxidation provide the additional advantage of a high sample throughput compared with batch procedures.

Most applications are based on the use of potassium permanganate in acidic media. The extensive use of this reagent, besides as an oxidizing agent, relies on its active role in chemiluminescent emission.

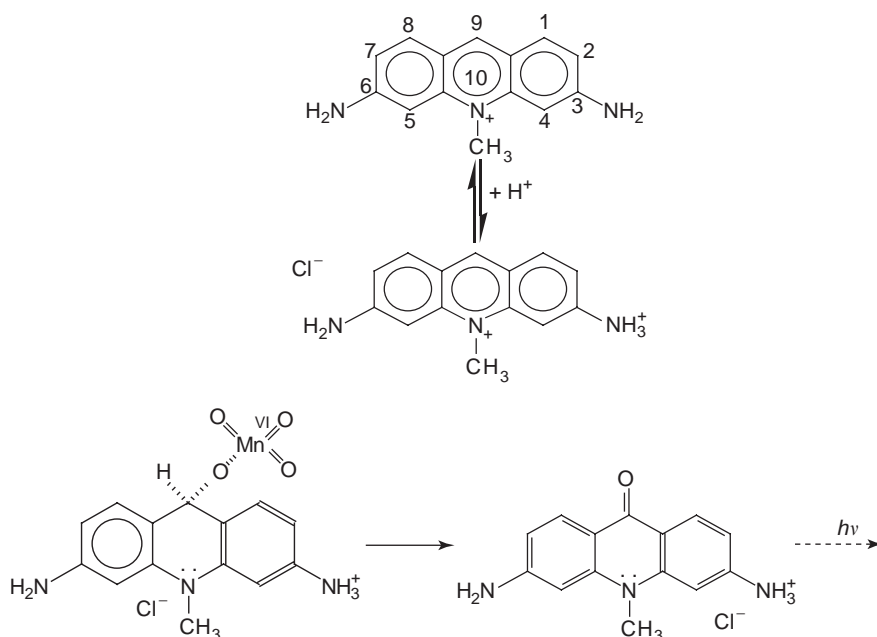
Recent studies give new insight into this by revealing the formation of intermediate complexes with the tested organic compounds proflavine and acriflavine. **Figure 2** illustrates this sequence and the light emission in the process.

The application of molecular connectivity to predict the chemiluminescent behavior of substances when they react with common oxidants in a liquid phase is useful for finding new chemiluminescent reactions. A large empirical screening, together with a review of the literature, can provide baseline data of positive and negative chemiluminescent reactions with strong inorganic oxidants. The discriminant function can be applied to different groups of pharmaceuticals and other organic compounds, namely, polyphenols, ergot alkaloids, and pesticides.

Pharmaceuticals displaying chemiluminescent behavior through direct oxidation are depicted in **Table 2**.

The online electrogenerated oxidative agent is theoretically more selective (kinetic selectivity) because it is used with small quantities of oxidant in a small volume, acting over a short period of time.

Indirect methods in chemiluminescence have long been used in pharmaceutical applications. The methodology is based on the light emission of some (very few) effective light emitters: the presence of the analyte affects the light emission by enhancing or inhibiting the yield. The use of the 'universal' light emitter luminol is still widely used (**Table 3**). The majority of listed procedures are based on the oxidation of luminol by hydrogen peroxide.



**Figure 2** Intermediate complexes in the oxidation of proflavine and acriflavine by potassium permanganate.



**Table 2** Chemiluminometric procedures using direct oxidation of the analyte

	Oxidant	Drug (sample)
Reactive in solution	Permanganate	Tetrahydropalmatine (tablets); reserpine (Ph); rescinnamine (Ph); yohimbine (Ph); medazepam (Ph); cefadroxil monohydrate (Ph and BF); salicylamide (Ph and urine); perphenazine (synthetic formulations); dipyrindamole (tablets and injections); etilefrine hydrochloride (Ph and BF); isoxsuprine hydrochloride (Ph and BF); prenalterol hydrochloride (Ph and BF); amidopyrine (Ph); pipemidic acid (Ph); ethamsylate (Ph); folic acid (Ph); sulphonamides (Ph); chloramphenicol (Ph)
	Ce(IV)	Captopril (tablets); tiopronin (Ph); penicillamine (Ph); phenothiazines (fluphenazine; levomepromazine; methotrimeprazine; trimeprazine tartrate) (Ph, urine, and plasma); furosemide (Ph); hydrochlorothiazide (Ph); analgin (tablets); phenacetinum (tablets); folic acid (Ph)
	Ce(IV)/sulfite	Fluoroquinolones (ofloxacin; norfloxacin; ciprofloxacin; lomefloxacin) (Ph); benzamides (sulpiride; sultopride; tiapride) (Ph and BF); ofloxacin (Ph).
	Tris(2,2'-bipyridyl) ruthenium(II)/Ce(IV)	Fluoroquinolone (ofloxacin, norfloxacin, ciprofloxacin) (Ph); thioxanthenes (zuclopenthixol, flupentixol, thiothixene) (Ph and BF); flufenamic and mefenamic acid (Ph)
	Tris(2,2'-bipyridyl) ruthenium(III)	Ranitidine (Ph); salbutamol (Ph)
	Periodate	Isoniazide (tablets); lincomycin hydrochloride (injections)
	Peroxides	Tetracyclines (tetracycline, chlortetracycline, oxytetracycline) (Ph); vitamin B <sub>12</sub> (Ph); albumin (Ph)
	Sodium hypochlorite	Persantin (injections and tablets)
	Hypobromite (generated <i>in situ</i> )	Urea (urine)
	Dissolved oxygen	Menadione (Ph and BF)
	Hexacyanoferrate(III)	Thiamine (tablets); isoniazide (Ph); ergonovine maleate (Ph); ergotamine (Ph)
	N-Bromosuccinimide	Amiloride (Ph); streptomycin (Ph); epinephrine (Ph)
Electrogenerated	Co(III)	Gentamycin (Ph); pipemidic acid (Ph); sodium phosphate dexamethasone (injections)
	Mn(III)	Adreson (tablets and injections); quinine (Ph); captopril (Ph)
	Ag(II)	Captopril (Ph)
	Bromine	Tetracyclines (tetracycline, oxytetracycline, chlortetracycline) (Ph)

Ph, pharmaceutical formulations; BF, biological fluids.

Lucigenin is the next most frequently used emitter for the determination of isoprenaline or kanamycin by inhibition of the lucigenin–hydrogen peroxide–cobalt(II) system, and also for the determination of the trace amounts of dopamine. The chemiluminescent determination of protein in a micellar medium using FIA employs offline labeling of albumin by sonicating a mixture of 1 ml of albumin in NaHCO<sub>3</sub> (pH 11.5) and fluorescamine in acetone for 1 min at room temperature.

The wide spectrum of chemiluminescent determination of pharmaceuticals also includes the use of immobilized reagents. However, the luminescence obtained through procedures other than those involving solved oxidants is not widely used, except in electroanalytical procedures.

Electrogenerated chemiluminescence in a flow cell containing a working electrode can be applied to the detection of codeine, heroin, and dextromethorphan. For the antibiotic cefadroxil, the flow cell is equipped with a Pt disk working electrode.

### Photochemical Methods

The easy control of photoreaction by regulating the flow (the usual photoreactor is a polytetrafluoroethylene (PTFE) tubing helically coiled on the lamp) makes it a very effective procedure. The most frequently used irradiation sources are mercury low-pressure lamps radiating in the UV range. Some authors propose the use of light in the visible range, and a Nd:YAG laser lamp has been used to irradiate chloroquine at 335 nm. The variety of commercially available lamps permit work with different 'sensitivity' and 'selectivity' (power and wavelength, respectively) levels of this 'analytical reagent'.

Fluorimetric detection of some pharmaceuticals is based on coupled irradiation–fluorimetry (Table 4). UV–Vis absorption techniques also use photolysis as the sole 'chemical' reaction of the sample before the determination or, in other cases, as an aid to a chemical derivatization reaction (Table 4).

**Table 3** Chemiluminescence analytical procedures based on the light emission of luminol

	<i>Drug (effect)</i>
H <sub>2</sub> O <sub>2</sub>	Vitamin B <sub>12</sub> (catalysis); ascorbic acid; morphinan alkaloids (morphine, sinomenine, and codeine); pyridoxine; sodium nitroprusside; sulbactam sodium (enhanced CL); clavulanic acid (enhanced CL); captopril (enhanced CL)
Potassium periodate	Isoniazide (enhanced CL); ascorbic acid; adrenaline and isoprenaline(immobilized); captopril
Fe <sup>2+</sup> ( <i>in situ</i> generated Fe(III) in HClO <sub>4</sub> )	Chlorpromazine, trimeprazine, and trifluoperazine (all emitting light by reduction of Fe(III); the Fe(II) was the catalyst)
Potassium permanganate	Paracetamol (inhibition); oxymetazoline (inhibition)
Potassium hexacyanoferrate(III)	Isoniazide (inhibition); menadione sodium bisulphite; rutin (inhibition); calcium debesilane (inhibition)
Hypochlorite electrogenerated	Etamsylate (inhibition); catecholamines (inhibition); isoniazide (sensitizing); methotrexate; cytarabini
Hypobromite electrogenerated	Isoniazide

**Table 4** Photodegradation analytical flow procedures for pharmaceuticals

<i>Detection</i>	<i>Drug (matrix)</i>	<i>Reagent or medium</i>
Fluorescence	Sulfadiazine (pharmaceutical preparations)	Aqueous solution
	Phenothiazine derivatives (urine)	In water or in ethanol
	Chlorpromazine (pharmaceutical preparations)	In slightly acidic medium
	Reserpine (tablets)	In acetate buffer of pH 3.4 and acetone
	Paracetamol (pharmaceutical preparations)	In presence of acetone and borax buffer at pH 9.6 or in Na <sub>2</sub> CO <sub>3</sub> /NaOH medium
	Sulfamethazine (pharmaceutical preparations)	By using sodium sulfite as sensitized and borate buffer at pH 8
	Menadione sodium bisulfite (pharmaceutical preparations)	In a mixture of acetone, NaOH, and sodium sulphite
	Thiamine (pharmaceutical preparations and serum)	In alkaline medium and presence of sodium sulfite
	Tianeptine and some of its metabolites (pharmaceutical preparations)	In aqueous solution containing 30% propan-2-ol at pH 2.3
	Ascorbic acid (pharmaceutical preparations)	With thionine blue in phosphate buffer at pH 3
Spectrophotometry	Nitroprusside (serum and pharmaceutical preparations)	With M-phloxin buffered at pH 6 with phosphate and in the presence of EDTA
	Sulfamethoxazole (tablets)	In ethanol
	Thioridazine (pharmaceutical preparations)	In nitric acid
	Thiamine (pharmaceutical preparations)	In HCl
	Diethylstilbestrol (pharmaceutical preparations)	In aqueous dipotassium hydrogen orthophosphate medium
	Riboflavine and riboflavine 5'-phosphate (multivitamin preparations)	With dianisidine and Mn <sup>2+</sup>
	Ascorbic acid (tablets)	With methylene blue
	Lactate (serum and pharmaceutical preparations); oxalate, citrate, and tartarate (pharmaceutical preparations)	By means of Fe(III); the resulting Fe(II) was determined using ferroine
	Vitamin K <sub>3</sub> (pharmaceutical preparations); chloramphenicol (pharmaceutical preparations); lactate (serum and pharmaceutical preparations); citrate (pharmaceutical preparations); riboflavine and riboflavine 5'-phosphate (pharmaceutical preparations); DNA (rabbit tissues); glucose (serum and urine)	Luminol
	Chloramphenicol (pharmaceutical preparations); sulfamethoxazole (pharmaceutical preparations); thiamine (pharmaceutical preparations)	Permanganate in acid medium
Biamperometry	Hydrochlorothiazide (pharmaceutical preparations)	Ce(IV) in acid medium
	Chloramphenicol and metronidazole (pharmaceutical preparations)	By an indirect procedure; reaction in acidic media and iodine release
Amperometry	Ascorbic acid (tablets)	With methylene blue
	DNA	Modified carbon paste electrode
AAS	Organoarsenic (urine)	Basic medium

The use of irradiation combined with chemiluminescence is not unusual; the idea is to use the photofragments as 'chemiluminescent analytes' to be oxidized directly or to form part of a luminol system. The photoproduct of sparfloxacin forms a complex with  $Tb^{3+}$ ; the complex sensitizes the chemiluminescence intensity of the Ce(IV)/sulfite system. Electrochemical detectors have also been employed to detect the resulting photofragments (Table 4).

### Atomic Spectrometric Methods

Atomic absorption spectrometry (AAS) can be used for the analysis of clinical samples, which often involves the determination of presence of metals in fluids and tissues, whether for toxicological investigation or for therapeutic indications. Most of these studies are carried out on urine, although determinations are also made in whole blood, blood serum, hair, biological tissues, or saliva (Table 5).

Other techniques used in this context include graphite furnace analysis, hydride generation, and cold-vapor absorption. Other applications involve the indirect FIA-AAS determination of some drugs in pharmaceutical formulations based on a prior reaction or precipitation, or redox or complex formation. Examples are shown in Table 6.

### Electroanalytical Techniques

The development of potentiometric detectors based on ion-selective electrodes continues to expand the scope of clinical and pharmaceutical applications of FIA. The small surface area of the sensor avoids adsorption problems and extends the service life of the electrode. The surface can be readily renewed periodically by alternating the washing cycles with the sampling cycles. The selectivity thus achieved is usually very good as it relies on differences in the

responsiveness of the detector to different components of a sample.

Some of the advantages of this type of sensor were already anticipated in one of the earliest FIA applications, namely, the joint determination of sodium and potassium in serum. Important parameters include the angle at which the carrier impinges on the active surface, the electrode placement angle, the immersion depth and surface area, the flow rates, and the cell void volume.

### Potentiometry

Several electrodes are used for the direct determination of pharmaceuticals. For example, an electrode is prepared using a polyvinyl chloride (PVC) membrane and bis(2-ethylhexyl)sebacate has been applied to tetracyclines in pharmaceutical formulations. The system has also been used to obtain the dissolution profile of some tetracycline formulations. A membrane for the determination of dopamine was synthesized using a  $CuSO_4$ -poly(ethylene-co-vinyl acetate) mixture in tetrahydrofuran, and the resulting solution was dropped into a tubular graphite-epoxy electrode. The drug was oxidized by Cu(II) immobilized on the membrane.

The miniaturization of sensors is an advantage for *in vivo* measurements. A miniaturized ion-selective sensor chip for potassium, mounted in a catheter or cannula and integrated onto an Si chip containing valinomycin with a Ag/AgCl reference electrode coated with a poly(hydroxyethyl methacrylate) layer, has been used in *in vivo* determination of K in blood or serum.

Indirect potentiometric measurements of pharmaceuticals can be performed by different methods,

**Table 5** FIA-determination of metals in biological and pharmaceutical samples

Sample	Metal
Serum	Ni, Se, Sr, Sn, Zn, Pb, Mg, Cu, Fe, Li, Cu
Whole blood	Hg, Se, Bi, Cd, Ni, Sb, Mn, Zn, Co, Pb, Cu
Urine	As, Gd, Co, Se, Ni, Bi, Sb, Ge, Pb, Cd, La, Ca, Mg, Li, Hg, Cu
Hair	Hg, Se, Cd, Ni, Sn, Pb, Cr, Co, Mo
Saliva	Hg
Pharmaceutical preparations	Cu, Fe, Zn, Pb, Bi, Cr, Co, Cd, Ni, As, Hg, Se

**Table 6** FIA indirect determination of pharmaceuticals (precipitation, complex formation, and solid-phase reactions) using AAS

Analyte	Reagent
Captopril and analgin	Cu(II)/KSCN buffered at pH 5.5
Norfloxacin	Cu(II)/sulfosalicylic acid in buffer at pH 5
Amoxicillin	Permanganate and barium chloride in basic medium
Quinine	Sodium tetraphenylborate in basic medium
Saccharin	Ag(I)
Papaverine and cocaine	Reinecke's salt
Ascorbic acid	Cr(III) in acid medium (redox)
Cyanide	Ag(I); solid-phase reactor
Vitamin B <sub>6</sub>	MnO <sub>2</sub> at pH 4–5; solid-phase reactor
Salicylic acid	Cu(II); solid-phase reactor
Ondasetron	Pb(II); solid-phase reactor

such as flow titration or by measuring the product of a prior chemical process. A flow titration for ascorbic acid is based on the reduction of iodate and determination of the released iodine using a flow-through ion-selective electrode. The FIA titration of epinephrine in pharmaceutical formulations is based on a PVC tetraphenylborate tubular electrode.

An example of the second modality is the hydrolysis of penicillin through a reactor containing immobilized penicillinase in which penicillinoic acid is detected using a pH electrode.

The combination of spectrometry and potentiometry enables multiple measurements, such as that of calcium and pH in serum using the spectral absorption of the former and the potentiometric signal of the latter. Also, stripping potentiometry has been employed for measuring mercury in urine, sediments, and biological materials using a gold electrode. Other inorganic analytes that have been studied include permanganate in pharmaceutical preparations (by incorporating an electrode prepared by dissolving a gold(I) complex), 2-nitrophenyl octyl ether, and PVC in tetrahydrofuran.

### Voltammetric Methods

The adsorption–preconcentration sequence has been applied to the determination of chlorpromazine in the presence of nonabsorbable species with similar redox potentials at concentrations up to 100 times higher than that of the analyte. Similarly, doxorubicin has been determined in urine samples subjected to no pretreatment; the drug is rapidly adsorbed over various carbon electrodes, where it is oxidized at fairly low potentials (0.57 V) and pH 4.5.

Te(IV) in urine has been determined through differential pulse voltammetry; the electropolymerization of 3,3'-diaminobenzidine on a gold surface gives a film of poly(3,3'-diaminobenzidine), which can form a complex with the Te(IV).

The measurement of tin in human hair, canned fruit juice, and spiked sea water has been carried out using flow injection anodic stripping voltammetry. The system uses an epoxy–carbon powder 8-hydroxyquinoline composite electrode and nonelectrode preconcentration in an open circuit in which Sn(II) is accumulated in the wall-jet electrode.

There have been recent reports of a glucose biosensor based on electroenzyme-catalyzed oxidation of glucose using a horseradish peroxidase–glucose oxidase lying on polymeric membrane. The biosensor has been used with human serum and pharmaceutical samples.

### Amperometric Methods

Most amperometric detection relies on enzymatic activity. A biosensor based on encapsulated glucose oxidase within a sol–gel glass has been used to determine glucose in an electrolytic solution for intravenous administration and human serum samples.

The isocitrate dehydrogenase activity in human serum has been determined by kinetic assay at 37°C, with isocitrate and NADP as substrates. The rate of formation of NADPH was amperometrically monitored and the selectivity was improved by covering the active surface of the electrode with a poly(*o*-phenylenediamine) film. The electroimmobilization of the enzyme monoamine oxidase into a polypyrrole film around a platinum electrode has been used to determine the antidepressant drug fluoxetine in pharmaceutical preparations.

Ascorbic acid in pharmaceutical preparations has been determined using a gold electrode made from compact disks (CDs) modified with platinum. The CDs consisted of a polycarbonate base on which a thin layer of photosensitive organic dye had been deposited. On this layer a mirror-like film of gold was fixed. Due to the low price of the CDs, each sensor can be used as gold electrode. A slice of the CDs was modified by electrochemical deposition of Pt.

## Online Separation Techniques

### Liquid–Liquid Extraction

Earlier FIA liquid–liquid extraction assemblies were limited to a few types of segmenters for merging solutions (aqueous and nonaqueous) and phase separators; most were home-made from glass, and the phase separation did not reach satisfactory levels. This led to the development of a variety of new and more sophisticated segmenters and phase separators by combining different types of models and hydrophilic and hydrophobic materials. This in turn provided improved phase separation when gravity was combined with hydrophobic PTFE membranes allowing 'filtration' of only the nonaqueous phase. This methodology has been frequently used for complex samples, e.g., in clinical studies. Other research lines have been devoted to avoiding the phase-separation step by 'electronic separation' or by homogenizing both phases.

A 'dual-conical gravitational phase separator' has been proposed for determining Cu and Pb in urine. A citrate-buffered medium (pH 3) serves to form analyte–ammonium pyrrolidinedithiocarbamate neutral complexes. These are then extracted into isobutyl ketone, and the organic phase separated by means of a dual-conical gravitational phase separator and

stored in a PTFE holding coil. Subsequently, an aqueous back extraction is performed with Pd(II) ions as the stripping agent using a second separation unit.

For the measurement of theophylline and diphenhydramine in pharmaceutical preparations, the sample is injected into an aqueous phase, and a hexanol-water segmented flow is established in the extraction coil. The wetting film of the organic phase forms on the walls of the PTFE tubing, reducing the linear velocity of the organic flow, and the differential flow velocities yield a separation in time between the unextracted and extracted components. The segmented flow is merged with a methanol stream and enters a mixing coil for homogenization before passage through the spectrophotometric flow cell.

Among the most frequently used organic solvents is chloroform, used for the determination of many anions in various samples: free cyanide and thiocyanate ions in human saliva; chlorophenols in urine; and Bi(III), ethambutol hydrochloride, diclofenac, phenothiazine derivatives, imipramine, warfarin, flufenamic acid, and ethylmorphine in pharmaceutical preparations. Another proposed solvent is dichloroethane, used for the determination of bromhexine in serum; and imipramine, ambroxol, berberine, bromhexine, neostigmine, 8-chlorotheophylline, and diphenhydramine in pharmaceutical samples.

## Dialysis

The dialyzer unit employed in FIA is formed by two plates of inert material (methacrylate) that are screwed to each other through a crescent bore in each. To decrease dispersion, the channels are often packed with inert glass beads. In this way, the sample is injected into the donor stream, crosses the dialyzer, and is brought into contact with a membrane that is crossed by the analyte. The dialyzer can be nesting at virtually any point in the flow assembly, according to the requirements of the method, and planar spiral arrangements have been proved to be the most efficient.

Several systems include one or two reactors (usually enzyme reactors) to remove interferences. The determination of Zn(II) in multivitamin tablets employs an electro-dialyzer fitted with a passive neutral membrane to enhance mass transfer together with the removal of interferences. Two electrodes are placed at the sides of the membrane, and the applied potential forces the analyte to the acceptor stream.

Methods based on converting an analyte into a gaseous phase to be separated through membranes are used, e.g., ammonia in urine is released by

merging the sample solution with the alkaline stream and then it crosses membrane to an acidic acceptor. The same separator has been used for creatinine and carbon dioxide in blood after the sample flows through an enzymatic solid-phase reactor that releases ammonia.

The earliest FIA applications of dialysis were focused on developing microdialysis units for *in vivo* measurements such as the determination of glucose using a biosensor connected to a microdialysis probe. Other *in vivo* and microdialysis processes have been developed to study the bovine serum albumin (BSA) binding of streptomycin sulfate using a luminometer detector; to monitor calcium in blood and for fluorimetric detection; and to monitor calcium in blood through chemiluminometry.

The dialyzer also serves as a means of either removing interferences or diluting the sample, as has been shown for the photometric determination of phosphate and chloride and sulfate in blood serum and urine. For electrochemical determination of glucose in parenteral solutions, two dialyzer units are inserted, the first one for diluting the matrix and to eliminate interferences and the other for the enzymatic reaction with glucose oxidase circulating through a closed circuit to minimize the required reagent (Figure 3).

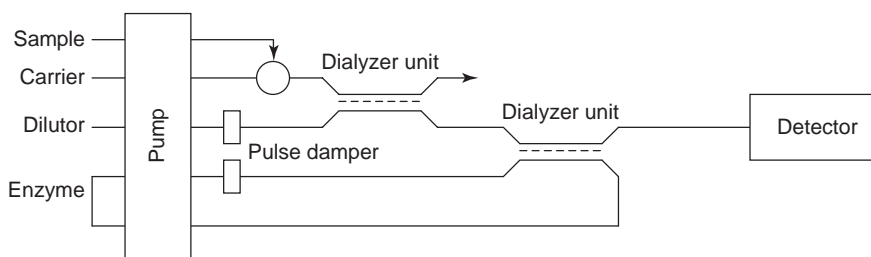
## Precipitation

As mentioned previously, the simplest possible way of inducing precipitation in FIA is to use a single-channel manifold and inject the sample into a buffered carrier-reagent stream, typically containing a metal ion. The manifold includes a filtration device for measuring the 'excess' of reagent, which will give rise to a negative transient signal; or the precipitate should be dissolved to measure the product of the reaction.

Off-line precipitation of captopril in pharmaceuticals using the Cu(II)-SCN<sup>-</sup> system buffered at pH 5 gives a suspension to be injected into a buffered carrier that is transferred to a conical filter packed with cotton, and further diluted in the excess Cu(II) in a mixer prior to air/acetylene flame AAS. Systems analogous to online precipitation-dissolution can be found for saccharinate in pharmaceuticals; Co, Ni, and Cd in urine; and immunoglobulin G in serum.

Reverse FIA has been proposed for papaverine and cocaine in pharmaceutical preparations. The reagent Reinecke's salt is injected into the HCl carrier, and the observed FIA peak is due to chromium. The peak difference (HCl carrier or sample carrier) represents the precipitated chromium and the analytical output.





**Figure 3** Determination of glucose in parenteral solutions.

## Solid-Phase Reactors and Online Sample Conversion

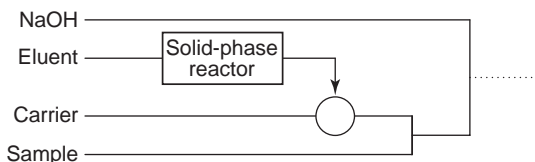
A reactive bed can be prepared using 'insoluble' materials, provided they have the required mechanical and chemical resistance; a more frequent procedure is to anchor the reagent to a solid support. The reactor can be nested at different points along an FIA manifold, depending on the purpose they serve. Some examples below are illustrative of some of the operational modes used in this context.

On redox- or complex-forming reaction with a solid-phase reactor, some drugs release a constituent metal to an extent proportional to the concentration of the drug in the sample. The released ion can thus be directly monitored, usually through AAS, or by using any other detector, after a chemical reaction.

The bed-reactor can act as a source for the reagent; a reverse FIA as in **Figure 4** has been used for thiamine, rutin, and riboflavin measurement in pharmaceuticals and in urine. The column was prepared with luminol and periodate on an anionic exchange resin. The drug content was related to the diminution or increase of the emitted light. The resorcinol chemosensor is another example, in which the light emission was generated by the reaction of  $K_3Fe(CN)_6$  and luminol in an alkaline medium with the reagents immobilized on an anion-exchange resin and eluted by sodium phosphate. Similar flow systems can be found for folic acid and rutin in human urine and pharmaceuticals.

The same resin has been employed to fabricate a sensor. Immobilized resin is packed in a glass tube ( $50 \times 3$  mm inner diameter) placed in front of a photomultiplier tube. The system can be used for the determination of isoniazide, adrenaline, and isoprenaline in pharmaceuticals by insertion of the sample in a carrier of water.

The use of a sensor with a solid-phase reactor in the same cell minimizes the analyte dispersion. The solid bed (resin or a gel ion exchange) contains the immobilized reagent or can be used to retain the analyte, either directly or after forming a complex or an ion-paired compound, which is retained. These



**Figure 4** Flow assembly with a solid-phase reactor as a source for the reagent.

sensors have also been used for resolution of mixtures such as paracetamol, caffeine, and acetyl salicylic acid.

Sequential spectrophotometric measurements of two drugs (ascorbic acid–thiamine, acetyl salicylic acid–thiamine, or paracetamol–salicylamide) and fluorimetric measurements (warfarin–thiabendazole) have been performed using two solid-phase reactors: first, a precolumn retains one of the analytes and the nonretained one is monitored in a flow cell containing a solid-phase reactor; anionic exchange for the spectrophotometric measured couples or C18 bonded silica for the fluorimetric measurements. The next step is to elute the formerly retained analyte to be monitored.

The major difficulty of using an 'insoluble' reagent lies in the particle sizes of commercially available chemicals. A reactor of  $CoCO_3$  prepared by entrapping the salt into a polymeric material and filling a packed-bed reactor has been employed for cysteine determination. The same strategy has been used in the measurement of many drugs: the immobilization of  $FePO_4$  as a source of  $Fe(III)$  allows determination of nalidixic acid, pipemidic acid, thioridazine, or chlorpromazine using a merging-zones flow system. A reactor containing cupric salts allows spectrophotometric determination of salicylic acid; a reactor of  $AgCl$  or  $AgI$  is used for ammonium or cyanide ion, respectively. The trapped oxides of manganese and lead allow determination of isoniazide, phenothiazines and adrenaline, and metamizol and ondasetron, respectively, through AAS, spectrophotometry, or fluorimetry.

Immobilization on a membrane is occasionally an effective alternative to solid-phase reactors. Thus, a



cationic fluorescent probe of hexadecyl-acridine orange is immobilized on a PVC membrane containing valinomycin with the membrane fixed to the inner walls of the flow cell. In this way, potassium ion is selectively determined in plasma samples. A flow-through bulk optrode membrane that incorporates 1-(2-pyridylazo)-2-naphthol in a plasticized PVC membrane entrapped by a cellulose support is useful for the determination of Zn(II) in pharmaceutical samples.

### ***In Vitro* Availability or Dissolution Profiles**

The *in vitro* availability test, also known as the solution profile of a pharmaceutical formulation, is an established mandatory test in international pharmacopoeias.

The original idea of implementing FIA in a dissolution assembly was to obtain 'individual' profiles versus the 'global' profile test mandatory in pharmacopoeias. The flow assembly, prepared for the determination of an active principle, is connected to the solution vessel, and the resulting 'curve' (a set of increasing FIA peaks) is the concentration versus time of a single active principle. An FIA system for the determination of thioridazine based on the oxidation of the drug by Fe(III) obtained from a solid-phase reactor has been used to obtain the dissolution profile of tablets containing this drug.

Any kind of detector other than the UV-Vis spectrophotometer, implementing the FIA assembly, is useful for obtaining the 'individual profile' of a drug in a pharmaceutical formulation. The dissolution profile of tablets containing ondasetron has been obtained using AAS, monitoring the Pb(II) released from a solid-phase reactor when the ondasetron goes through it. Chemiluminescence has been used to obtain the dissolution profile of sulfamethoxazole by oxidation with permanganate after photodegradation of the drug.

The simultaneous dissolution profile of two compounds with overlapped spectra has been solved using different mathematical approaches. Derivative spectrophotometry is a solution used for the couples sulfadiazine-trimethoprim, amitriptyline-perphenazine, sulphamethoxazole-trimethoprim, amoxicillin-bromhexine, and amoxicillin-clavulanic acid. This strategy has been used to obtain three dissolution profiles, namely, the 'standard' (global) profile and two individual profiles such as sulphamethoxazole-trimethoprim or hydrochlorothiazide-captopril.

The simultaneous monitoring of three drugs requires chemometric methods. The fluorimetric determination of vitamins B<sub>1</sub>, B<sub>2</sub>, and B<sub>6</sub> from multicomponent solid

preparations has also been used to monitor the dissolution of these drugs.

A microdialysis sampling has been used for multivessel dissolution testing of isoniazide tablets. The dialyzer is produced using regenerated hollow cellulose dialysis fibers, and the detection is at 254 nm.

## **Biochemical (Enzyme and Immunoassay) Methods**

### **Enzymatic Assays**

The earliest FIA enzymatic analysis involved the use of a solution or an immobilized enzyme reactor (IMER). At present, most FIA users prefer the marriage of convenience FIA-IMER, clearly due to the advantages presented by this modality, namely, high reactivity and high reproducibility. Further, the enzyme is regenerated via a catalytic cycle, in variations of the enzyme activity are reflected on transient FIA outputs, and small variations in the system behavior (pH, flow rates, temperature, or traces of inhibitors or activators) are not seen in the results.

An assay for angiotensin has been developed with the aid of FIA-IMER and by converting enzyme-inhibitory activity. The quantification of hippuric acid liberated from hippuryl-histidyl-leucine with the aid of the d-amino acid oxidase in the carrier solution resulted in a constant level of H<sub>2</sub>O<sub>2</sub> until the reaction was inhibited by benzoate produced by the action of aminoacylase on hippuric acid. This reduction was monitored amperometrically, allowing indirect measurement of the hippuric acid.

The simultaneous determination of galactose and galactose-1-phosphate in dried blood uses an enzyme-immobilized column and has been applied to preliminary neonatal mass screening for galactosemia. Galactose dehydrogenase and alkaline phosphatase have been immobilized into a gel and packed into a column.

The immobilization can be performed with a modality different from that of the reactor filled with the immobilized enzyme. The spectrophotometric control of hemodialysis treatment is based on an optical flow-through biosensor based on Prussian Blue film with chemically linked urease forming a monomolecular layer of the enzyme. This pH-enzyme optrode-FIA system has been used for the selective determination of postdialyze urea in real clinical samples.

A novel immobilization device for an FIA system is an enzyme reactor fabricated from an anisotropically etched Si chip and a glass plate (determination of

glucose). The V-shaped groove side of the Si wafer is anodically bonded to a glass plate to produce an Si capillary. A solution of 3-aminopropyltriethoxysilane in toluene is passed through the capillary, followed by heating overnight at 115°C. The capillary is treated with glutaraldehyde at pH 7.1 for 1 h and then filled with glucose oxidase at pH 7.1 overnight for immobilization of the enzyme. The enzyme reactor is then incorporated into an FIA system.

The classic peroxidase immobilization has been extended to electrodes, e.g., determination of hydrogen peroxide based on a peroxidase/ferrocene-embedded carbon paste electrode covered with a Nafion-coated cellulose acetate membrane. The system was successfully used for glucose and urate in serum.

In pharmaceutical formulations, acetylcholine can be determined by employing a detector where portions of the analyte are injected into a carrier at pH 7.4 and passed through a solid-phase reactor containing immobilized acetylcholinesterase on glass beads. The solution merges with a stream of H<sub>2</sub>SO<sub>4</sub> and passes through a PTFE membrane diffusion cell. The acetic acid that diffuses through the membrane is carried by the acceptor water to the detector cell.

Immobilization of enzymes in the detector flow cell is of paramount importance in the performance of electrodes. The fabrication and performance of a reticulated vitreous carbon-based tyrosinase flow-through electrode, in which the enzyme was covalently immobilized, was tested as an amperometric detector for phenolic compounds. The measurement of uric acid in plasma can be carried out through closed-loop FIA using a co-immobilized enzyme flow cell and chemiluminescence detection.

### Immunoassays

The specific interaction antigens (analytes) and antibodies (reagents) used for analytical methods generally known as immunoassays are reflected quantitatively by changes in the activity or signal yielded by an appropriate label. The inception of FIA in the immunological field results in faster and more reproducible assays. Examples are given in Table 7.

Since antibody-antigen binding does not produce a detectable signal by itself, several techniques involve a tag such as an enzyme, radioisotope, or fluorophore to monitor the reaction with a variety of optical and electrochemical detectors. One group of detectors generally uses fluorescence and chemiluminescence while another employs amperometry and potentiometry. Electrochemical immunoassays are emergent alternatives to existing immunochemical

methods, especially for opaque or optically dense matrices.

Heterogeneous flow systems are generally preferred as they show a higher sensitivity and flexibility than the homogeneous ones. A solid-phase reactor binds the species of interest and allows the rest of the matrix to pass through the reactor; this is the separation step. Then, the bound analyte must be removed without affecting the activity of the reaction surface of the immobilized antibody. Both characteristics, the antibody affinity and the detectability of the label, are essential in the design of the final heterogeneous immunoassay system.

Chemiluminescence continues to be a very selective and sensitive detection tool in FIA immunoassays. The main component in the most commonly used heterogeneous systems is an immunoreactor column, consisting of PTFE tubing packed with immobilized antibodies or haptens. The immunoreaction and the chemiluminescence reaction takes place within the immunoreactor, and the emitted light is collected directly from the cell.

As the antibodies are immobilized and the immunoreactor can be used repeatedly, the consumption is minimum; on the other hand, the binding reaction also represents a preconcentration of the sample, and pretreatment is usually not necessary.

A comparison of fluorescent immunoassay and enzyme-multiplied immunoassay techniques (EMITs) concluded that they are similar in terms of accuracy, precision, and simplicity; however, the sensitivity of fluorescent immunoassay is better by a factor of  $\sim 8$  than that of EMITs, but the assay time for fluorescent immunoassay is longer. The comparison was performed on the measurement of serum carbamazepine.

A pH-enzyme optode spectrophotometric flow injection system has been used for monitoring clinical hemodialysis. The role of the incorporated dialyzate urea detector is played by an optical flow-through biosensor based on Prussian Blue film with chemically linked urease forming a monomolecular layer of the enzyme. This pH-enzyme optode-FIA system is useful for selective determination of postdialyzate urea in the range of concentrations corresponding to its level in real clinical samples ( $2\text{--}16\text{ mmol l}^{-1}$ ) at  $\sim 15$  samples of spent dialyzate per hour.

Potentiometric, capacitive, and amperometric transducers have been used for direct and indirect electrochemical immunoassays. However, due to their fast detection, broad linear range, and low detection limit, amperometric transducers are preferred.

Developments in the sensor fabrication techniques coupled to a flow injection assembly have been incorporated in real-time monitoring of biospecific interactions. The assembly utilizes screen-printed

**Table 7** FIA-immunoassays

Detection	Drug (matrix)	Procedure
Chemiluminescence	Digoxin and triiodothyronine (serum)	Acridinium system. Solid-phase reactor in flow cell
	Uric acid (plasma)	Luminol; carbonate buffer at pH 9; uricase and peroxidase on polymer beads in the flow cell
	Rabbit IgG	System sodium tetraphenylborate and <i>p</i> -phenylphenol + luminol/H <sub>2</sub> O/horseradish peroxidase
Fluorescence	Phenytoin (serum)	Competitive binding of fluorescein-labeled and unlabeled phenytoin with a phenytoin antibody; gel filtration
	Atrazine in rabbit anti-atrazine antisera	Protein A/G bonded to an azlactone-activated polymeric support
	4-Nitrophenol	Competitive labeled 4-nitrophenol derivative (tracer) and the analyte
	2,4-Dinitrobenzenesulfonate (DNP)-BSA	Sephacrose. Cation exchange column; fluorescein isothiocyanate-labeled antibody to DNP
	Total disopyramide in serum	Equilibrium dialysis of samples before analysis
	Gentamicin (serum) <sup>a</sup>	Enzyme immunoassay versus a radioimmunoassay
	Carbamazepine (serum) <sup>a</sup>	Fluorescent immunoassay versus EMIT
	Cotinine (saliva and serum) <sup>a</sup>	Monoclonal and polyclonal antibodies versus cotinine in nonisotopic and isotopic immunoassays
Electrochemical	Simultaneous determination of galactose and galactose-1-phosphate in a dried blood disk <sup>a</sup>	Enzyme-immobilized column
	Chorionic gonadotrophin (human blood serum)	Ferrocenecarboxylic acid conjugated with anti-human chorionic gonadotrophin immunoglobulin G antibody. Separation in a capillary column
	IgG	Amperometric detection using antibodies immobilized by physical absorption label, substrate alkaline phosphatase naphthyl phosphate
	4-Nitrophenol	Glucose dehydrogenase amperometry biosensor
	Atrazine	Electrochemiluminescence; atrazine antibodies, atrazine labeled with glucose oxidase
	Cocaine	Online detection of the label alkaline phosphatase by a substrate-recycling biosensor
	Angiotensin	Amperometric quantification of hippuric acid from hippuryl-histidyl-leucine
	<i>Escherichia coli</i> and <i>Salmonella</i>	Porous nylon membranes as a support for anti- <i>Escherichia coli</i> or anti- <i>Salmonella</i> antibodies
	Staphylococcal protein A	Amperometric immunosensor. Anti-protein A antibodies immobilized on dispersed C-powder
	Digoxin	Tyrosinase label – detecting the phenolic product liberated by an alkaline phosphatase

<sup>a</sup>Comparative assays.

electrodes incorporating the electroactive polymer, polyaniline, which brings about mediator-less redox coupling between the electrode and biomolecular components attached to the polymer surface. This system also utilizes enzyme-labeled antibodies as the biomolecular recognition component for the analysis of the test analyte, biotin.

See also: **Chemiluminescence:** Overview. **Flow Analysis:** Overview. **Flow Injection Analysis:** Principles. **Spectrophotometry:** Overview.

## Further Reading

- Gómez-Taylor Corominas B, Antón Fos GM, García Mateo JV, Lahuerta Zamora L, and Martínez Calatayud J (2003) Prediction of the chemiluminescent behaviour of phenols and polyphenols. *Talanta* 60(2–3): 623–628.
- Hosse M, Ballesteros E, Gallego M, and Valcarcel M (1996) Turbidimetric flow method for the enantiomeric discrimination of l- and d-aspartic acid. *Analyst* 121: 1397–1400.
- Lahuerta Zamora L, Fuster-Mestre Y, Duart MJ, *et al.* (2001) Prediction of the chemiluminescence behaviour of

- pharmaceuticals and pesticides. *Analytical Chemistry* 73: 4301–4306.
- Martínez Calatayud J (1996) *Flow Injection Analysis of Pharmaceuticals. Automation in the Laboratory*. London: Taylor and Francis.
- Moreno Gálvez A, García Mateo JV, and Martínez Calatayud J (2002) Simultaneous dissolution profiles of two drugs in solid oral dosage forms by a FIA manifold provided with a single detector. *Journal of Pharmaceutical and Biomedical Analysis* 27: 1027–1036.
- Ruiz Medina A, Fernández de Córdova ML, and Molina-Díaz A (1999) Simultaneous determination of paracetamol, caffeine and acetylsalicylic acid by means of a FI ultraviolet psl multiptosensing device. *Journal of Pharmaceutical and Biomedical Analysis* 21: 983–992.
- Sanfeliu Alonso MC, Lahuerta Zamora L, and Martínez Calatayud J (2003) Determination of tyrosine through a FIA-direct chemiluminescence procedure. *Talanta* 60(2–3): 369–376.
- Solich P, Sklenářová H, Polasek M, and Karlicek R (2001) Application of flow-injection technique in pharmaceutical analysis. Part I. Spectrophotometric and chemiluminescence detection. *Journal of Flow Injection Analysis* 18(1): 13–20.
- Trojanowicz M (2000) *Flow Injection Analysis. Instrumentation and Application*. Singapore: World Scientific.
- Vela MH, de Jesús DS, Couto CMCM, Araujo AN, and Montenegro MCBSM (2003) Electroimmobilisation of MAO into a polypyrrole film and its utilisation for amperometric flow detection of antidepressant drugs. *Electroanalysis* 15(2): 133–138.

## Industrial Applications

**A F Danet**, Bucharest University, Bucharest, Romania  
**M Badea**, National Institute for Chemistry and Petrochemistry – ICECHIM, Bucharest, Romania

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

In the last 25 years it has been demonstrated that flow injection analysis (FIA) is a powerful tool for real-time monitoring and control of industrial processes. Its versatility makes possible the setting up of FIA devices for measuring a number of physical, chemical, and biochemical parameters.

Here, we present an overview of the instrumentation and operational modes used in FIA for industrial applications. Different applications of this technique, which can be applied for specific purposes in various industries, are discussed, with particular emphasis on the chemical industry, biotechnology, the food and beverages industry, and wastewater control.

### General Considerations

FIA has many features that make it a valuable technique for industrial analysis and for online process monitoring and control. Most FIA techniques used in the laboratory can be adapted to industrial analysis. Sampling is of utmost importance in industrial analysis since very often the analyzed liquids are not homogeneous. FIA cannot directly improve the

sampling capabilities of an analytical system. Nevertheless, due to the large pool of data that it can provide, greater statistical confidence can be placed in the results obtained.

### Instrumentation and Operational FIA Modes in Industrial Analysis

The instruments used for continuous monitoring and control of industrial plant process must be much more robust than those used in the laboratory. They are intended to work in a nonprotected, often hostile environment, over long periods of time and with no operator present to survey them. Frequently, the industrial samples analyzed contain suspended particles that can easily block the small-bore (0.5 mm inner diameter (ID)) tubes used in the construction of laboratory FIA instruments. This problem can be solved either by inline filtration of samples or by using larger (2 mm ID) tubes. The latter solution is often chosen because frequent change of filters is involved in the former. Polytetrafluoroethylene and stainless steel are used to make the tubes. Gas-driven and peristaltic pumps are the usual pumping systems, the former finding particular use in FIA process monitoring and control. Injection of samples into the FIA devices used in industrial analysis is carried out by means of injection valves similar to those used in low-pressure liquid chromatography, or injection is carried out hydrodynamically. Almost any flow-through detector can be used in FIA devices for industrial analysis.

In the last two decades the major chemical instruments producers have introduced in their product lists

- pharmaceuticals and pesticides. *Analytical Chemistry* 73: 4301–4306.
- Martínez Calatayud J (1996) *Flow Injection Analysis of Pharmaceuticals. Automation in the Laboratory*. London: Taylor and Francis.
- Moreno Gálvez A, García Mateo JV, and Martínez Calatayud J (2002) Simultaneous dissolution profiles of two drugs in solid oral dosage forms by a FIA manifold provided with a single detector. *Journal of Pharmaceutical and Biomedical Analysis* 27: 1027–1036.
- Ruiz Medina A, Fernández de Córdova ML, and Molina-Díaz A (1999) Simultaneous determination of paracetamol, caffeine and acetylsalicylic acid by means of a FI ultraviolet psl multiptosensing device. *Journal of Pharmaceutical and Biomedical Analysis* 21: 983–992.
- Sanfeliu Alonso MC, Lahuerta Zamora L, and Martínez Calatayud J (2003) Determination of tyrosine through a FIA-direct chemiluminescence procedure. *Talanta* 60(2–3): 369–376.
- Solich P, Sklenářová H, Polasek M, and Karlicek R (2001) Application of flow-injection technique in pharmaceutical analysis. Part I. Spectrophotometric and chemiluminescence detection. *Journal of Flow Injection Analysis* 18(1): 13–20.
- Trojanowicz M (2000) *Flow Injection Analysis. Instrumentation and Application*. Singapore: World Scientific.
- Vela MH, de Jesús DS, Couto CMCM, Araujo AN, and Montenegro MCBSM (2003) Electroimmobilisation of MAO into a polypyrrole film and its utilisation for amperometric flow detection of antidepressant drugs. *Electroanalysis* 15(2): 133–138.

## Industrial Applications

**A F Danet**, Bucharest University, Bucharest, Romania  
**M Badea**, National Institute for Chemistry and Petrochemistry – ICECHIM, Bucharest, Romania

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

In the last 25 years it has been demonstrated that flow injection analysis (FIA) is a powerful tool for real-time monitoring and control of industrial processes. Its versatility makes possible the setting up of FIA devices for measuring a number of physical, chemical, and biochemical parameters.

Here, we present an overview of the instrumentation and operational modes used in FIA for industrial applications. Different applications of this technique, which can be applied for specific purposes in various industries, are discussed, with particular emphasis on the chemical industry, biotechnology, the food and beverages industry, and wastewater control.

### General Considerations

FIA has many features that make it a valuable technique for industrial analysis and for online process monitoring and control. Most FIA techniques used in the laboratory can be adapted to industrial analysis. Sampling is of utmost importance in industrial analysis since very often the analyzed liquids are not homogeneous. FIA cannot directly improve the

sampling capabilities of an analytical system. Nevertheless, due to the large pool of data that it can provide, greater statistical confidence can be placed in the results obtained.

### Instrumentation and Operational FIA Modes in Industrial Analysis

The instruments used for continuous monitoring and control of industrial plant process must be much more robust than those used in the laboratory. They are intended to work in a nonprotected, often hostile environment, over long periods of time and with no operator present to survey them. Frequently, the industrial samples analyzed contain suspended particles that can easily block the small-bore (0.5 mm inner diameter (ID)) tubes used in the construction of laboratory FIA instruments. This problem can be solved either by inline filtration of samples or by using larger (2 mm ID) tubes. The latter solution is often chosen because frequent change of filters is involved in the former. Polytetrafluoroethylene and stainless steel are used to make the tubes. Gas-driven and peristaltic pumps are the usual pumping systems, the former finding particular use in FIA process monitoring and control. Injection of samples into the FIA devices used in industrial analysis is carried out by means of injection valves similar to those used in low-pressure liquid chromatography, or injection is carried out hydrodynamically. Almost any flow-through detector can be used in FIA devices for industrial analysis.

In the last two decades the major chemical instruments producers have introduced in their product lists



FIA analyzers or FIA components. One company produces a series of instruments that are flow injection systems with atomic absorption spectrometric detection dedicated to determination of mercury. Some companies produce flow injection analyzers for a large number of ions. One supplier has an analyzer that comprises three separate units: a basic analytical module, an automatic sample module, and a data capture module, all these units being completely automated. The instrument is capable of analyzing nutrients, ions, and metals. It offers a wide analytical choice using ion-selective electrodes (ISEs), chemiluminescence, or fluorescence. With analysis speeds up to 120 samples per hour and detection limits down to parts per billion levels, this flow injection analyzer performs determinations well compared with other techniques.

The FIA techniques used in industrial analysis are mostly based on modification of the chemical matrix of the sample to be analyzed. This modification can be achieved by passing the analyte into another medium, by dilution, or by matrix matching. Transfer of the analyte into a different medium is the method most commonly used for modifying the sample matrix, the purpose of the modification being to increase the selectivity and, in some cases, the sensitivity of the method. It can be carried out by solvent extraction, gas diffusion through membranes, dialysis, etc. Sample dilution, which is often necessary in industrial analysis, is usually done by sample splitting and zone sampling. Matrix matching is required when certain types of detector are used. The parameters that depend on the chemical matrix of samples and which must be considered include the ionic strength, refractive index, pH value, viscosity, etc. For FIA control of some biotechnological process, inline conversion techniques based on enzymatic microreactors are used.

Two operational FIA modes are used for analyzing industrial samples, particularly for continuous process monitoring and control. In the first mode, which is the basic FIA technique, the sample to be analyzed is continually drawn out from the process to be monitored and periodically injected into the carrier stream and thus into the FIA channel. The system is first calibrated using one or several standards of known concentration; the process is then monitored via periodic sampling. The instrument may be rechecked by injecting another series of standards. In the second operational mode it is the reactant that is injected. The sample solution is pumped continuously through the FIA system and the reactant, which generates the product whose property is measured, is injected into this stream at predetermined time intervals. This technique, known as reversed FIA, has the

advantage of conserving the reagent. It has, however, the disadvantage that the detector may eventually become fouled due to a lack of carrier stream, and, moreover, the amount of sample needed is relatively large.

### **Advantages of Applying FIA to Industrial Analysis**

Among the advantages of using FIA in industrial analysis are the following: rapid changes of methods are possible, the throughput is large, the data are acquired in real time, and there is the possibility of performing matrix modification. Real-time or near-real time acquisition of data is critical to the efficiency of a process control system. Normally the process stream has a matrix, which makes it difficult to measure the analyte, and hence matrix modification is necessary in order to develop industrial analytical methods. Pretreatment (matrix modification) and analysis, including chemical conversion, can be carried out very rapidly. The entire operation usually takes less than 1 min.

The FIA system can be used to automate existing methods or to introduce minor modifications and improvements. Thus, the FIA system can utilize selectivity enhancement, kinetic discrimination, and matrix optimization to create new methods that are not applicable to batch or segmented continuous-flow systems. Reproducibility is excellent, partly due to the elimination of human intervention and bias. The FIA system is closed, making it suitable for use with volatile or toxic substances, and offers the advantage of removing, or at least minimizing, human exposure. The equipment is simple, rugged, and inexpensive. The hardware is computer compatible, and therefore the analyses can be automated easily. Calibration is also easy, baseline conditions are continuously monitored, and the detector is continuously cleaned by the carrier. High pressures and temperatures can be used.

### **Industrial Processes**

Process monitoring and control is one of the chief fields of interest in modern analytical chemistry. In spite of the advantage to be gained by using FIA for online process monitoring and control, the number of applications proposed so far is very low, given the technique's potential. Until the mid-1980s hardly any reference could be found to the industrial application of FIA for real process monitoring. This was at least partly due to problems associated with the design and construction of adequate sampling and sample conditioning systems. The use of membranes for dialysis, gas diffusion, and other separative and



preconcentration techniques merits more attention for online process analysis. Several novel process FIA instruments have come into the market in the past few years, and other companies working in this area expected to introduce systems in the near future. In some cases the interested companies have built their own FIA instruments devoted to process monitoring. Many FIA process control applications have not yet been published, to avoid divulging proprietary data.

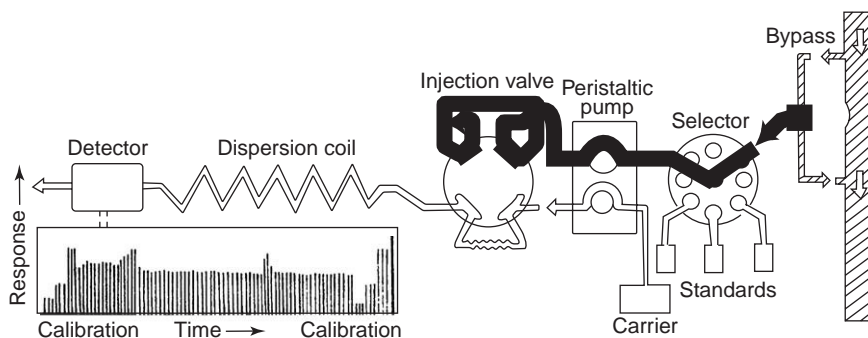
The FIA analyzers available for online process monitoring are sturdy and have been field tested, with proven design of components. The functional parameters, as well as the data acquisition and reporting functions, are microprocessor controlled. These analyzers still preserve all the FIA advantages, they are easy to service, and the systems can be safely placed in corrosive or hazardous environments. One company, for instance, produced FIA instruments designed specifically for online process monitoring and control of industrial plant processes with the features mentioned above. These instruments are designed to reside at the process line and can accommodate a high-pressure sample line. Several other firms produce FIA instruments that are able to analyze industrial samples with complex composition. These devices were originally designed to work under laboratory conditions but can be adapted easily to process monitoring and control. **Figure 1** shows an FIA device for online process monitoring with automatic calibration that can also be utilized for analyzing separate samples. Often these types of analyzer are modular. The modules can be combined in a variety of ways, and switching between analytical methods is simple. The analyzers can be equipped with a wide variety of detectors, gas-sensitive electrodes, etc. Using interchangeable analyzers, the time between individual analyses can be set at an optimal value in order to allow both accurate control of the process and a saving of the reagents. The

calibration can be programmed to be carried out as frequently as necessary.

By making use of FIA process monitoring and control, the measured parameters can be maintained at the desired optimal values. This leads to an increase in the quality of the products obtained and also to significant savings by allowing a decrease in the number of personnel required and reducing expenditure on materials and energy. Some important examples of application of FIA to online process monitoring and control are listed in **Table 1**.

## Biotechnological Processes

The need for better monitoring and control of bioreactors has become greater as more and more processes are studied. The control and optimization of fermentation requires online information about the biological processes involved. Thus the concentrations of the various biological factors the substrate, some enzymes, some ions, and the product must all be determined in order to provide precise feedback for better control. Conventionally, most industrial biotechnological processes are determined offline in the laboratory, after withdrawal of separate samples. This method, however, is not satisfactory due to the time delay, the limited reliability, and the manpower needed for analysis of a large number of samples. Inline sensors appear to be the ideal choice, but although their use is increasing, they show some reliability problems. For this reason, analytical approaches based on continuous sampling, followed by automatic online analysis, are currently being used. The traditional automatic analyzers used for this purpose are increasingly being supplanted by FIA devices. The FIA systems offer fast, selective analysis on request, based on classical wet chemistry, and allow integration of automated sterile sample taking

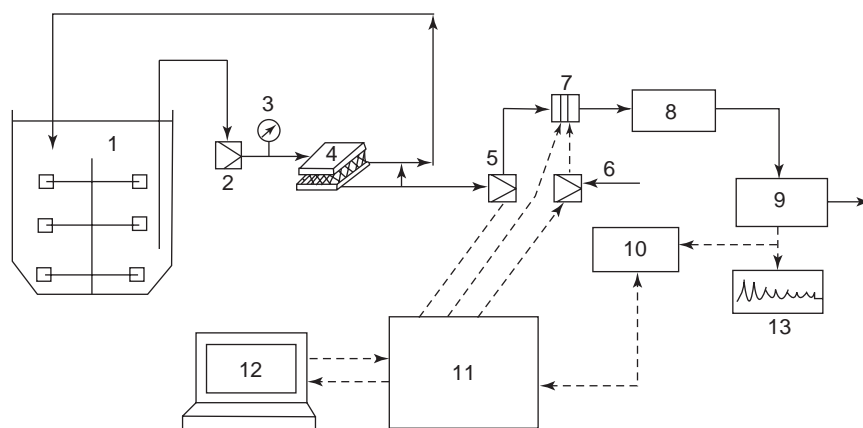


**Figure 1** A single-line FIA analyzer for online process monitoring with automatic calibration. Inset: The shape of the output signal from the analyzer.

**Table 1** Features of FIA online process monitoring and control

Analyzed species	Measuring range	Type of sample	Detection
Sodium hydroxide	0–4 g l <sup>-1</sup>	Caustic process stream	Potentiometry
Sodium carbonate	0–7 g l <sup>-1</sup>	Potentiometry	Potentiometry
Hydrochloric acid	9–38% (m/m)	Concentrated HCl production plants	ISE
PH	4.5–9.5	Water in power plant cycles	pH electrode
Ammonia	0.2–3 mg of N per liter	Water in power plant cycles	Spectrophotometry
Hydrazine	0.025–0.3 mg l <sup>-1</sup>	Water in power plant cycles	Spectrophotometry
Fe	0.01–10 mg l <sup>-1</sup>	Water in power plant cycles	Spectrophotometry
Cu	0.02–0.2 mg l <sup>-1</sup>	Water in power plant cycles	Spectrophotometry
Zn	0.2–60 g l <sup>-1</sup>	Process stream in hydrometallurgical plants	Spectrophotometry
Zn	mg l <sup>-1</sup> up to 110 g l <sup>-1</sup>	Galvanizing preflux solutions	Atomic absorption spectrometry
Ti	2.5–30 mg l <sup>-1</sup>	Brine	Spectrophotometry
Fe(III)	0.5–180 mg l <sup>-1</sup>	Mineral process liquors	Spectrophotometry
Fe(II)	0.5–120 mg l <sup>-1</sup>	Mineral process liquors	Spectrophotometry
Ni(II)	0.17–0.24 mg l <sup>-1</sup>	Electroplating bath	Spectrophotometry
NH <sub>3</sub>	70–2000 mg l <sup>-1</sup>	Organic solvents	Spectrophotometry
Peroxide (active oxygen)	0–1000 mg l <sup>-1</sup>	Organic process streams	Spectrophotometry
Sulfuric acid	20–100 mg l <sup>-1</sup>	Phenol/acetone matrix	Conductimetry
Free fatty acids	0.01–0.3% (m/m)	Edible oils	Spectrophotometry
Hg	0.1–100 µg l <sup>-1</sup>	Zinc battery anode	Atomic fluorescence spectrometry
Cu	0.2–1.0 mg l <sup>-1</sup>	Effluent streams	Spectrophotometry
Carbonyl	0.05–1.0 carbonyl number	Organic intermediates	Spectrophotometry
Water	0.001–0.1% (m/m)	Organic solvents	Spectrophotometry
Cyanides	0.1–1.0 mg l <sup>-1</sup>	Industrial effluents	Spectrophotometry
	20–1000 mg l <sup>-1</sup>	Process solutions	ISE
Sulfate	0.05–0.5 g l <sup>-1</sup>	Basic solutions	Spectrophotometry

From Features of FIA/on-line process monitoring. Puchades R, Maquieira A, Atienza J, and Herrero MA (1990) State of the art in on-line techniques coupled to flow injection analysis FIA/on-line – a critical review. *Journal of Automatic Chemistry* 12: 163–173.



**Figure 2** A simple device used for monitoring quality in a fermentation process. The scheme shows the arrangement of the fermentation vessel, sampling device, analytical system, actuator, and monitoring equipment. Key: —, flow lines; —, electrical connections; 1, fermenter; 2, pump; 3, pressure meter; 4, microfiltration equipment; 5, sample pump; 6, reagent pump; 7, injection valve; 8, FIA equipment; 9, spectrophotometer; 10, voltmeter; 11, analog-to-digital converter system; 12, computer; 13 recorder. (Reproduced with permission from Valero F, Lafuente J, Poch M, *et al.* (1990) On-line fermentation monitoring using flow injection analysis. *Biotechnology and Bioengineering* 36: 647–651; © Wiley.)

and preparation during analysis. The combination of suitable sampling devices with FIA systems is the prerequisite for online control of bioreactor processes. By using an FIA manifold such as that illustrated

in **Figure 2**, a system can be constructed that, in addition to having a high sampling frequency, offers the benefit of real-time continuous monitoring by aspirating the sample solution regularly (e.g., from a

batch-type bioreactor) into a carrier/wash solution. The signals, recorded as a series of peaks (the injected samples superposed upon a constant background of the carrier/wash solution), can provide positive control of both the detector and the process. Fouling of the detector is prevented by the periodic wash sequence of the carrier stream. Injection of calibration samples from time to time allows an additional control for correct functioning of the detector.

Different FIA procedures have been developed for biotechnology applications, especially for online glucose analysis. Some methods use continuous enzymatic reagent consumption, and some use enzyme immobilization techniques. The former are more versatile, easier to assemble, and more robust. The latter are simpler and more economical, but due to the nature of enzyme kinetics, it is difficult to measure substrate concentrations above  $1\text{ g l}^{-1}$  without the use of dilution techniques, which may affect the final result.

The measurement of enzyme activity has been used as a direct control parameter for achieving optimal fermentation. At present, the most common enzymes are determined offline in the laboratory, after withdrawal of a separate sample. With FIA it should be easy to develop automatically operated enzyme determination procedures, based on reaction schemes, that can be used for fast and efficient process monitoring and control. For instance, the feasibility of online measurement during the fermentation of extracellular hydrolases (proteases or glucosidases) on a 30 or 70 l pilot scale has been demonstrated. The changes in the concentrations of the solutes caused by the enzyme activity were monitored spectrometrically. An FIA system devoted to online monitoring of formate dehydrogenase and leucine dehydrogenase in the control of enzyme production process has also been proposed. An online FIA system has also been developed for intracellular ATP measurement in a culture fluid of yeast and *Escherichia coli*; it uses a luminescent reaction for determination of the microorganism cell number.

A flow injection immunoanalysis system connected via a sterile sampling unit to a continuous bioreactor has been used for online monitoring of monoclonal antibodies in the course of a hybridoma cell fermentation. Mouse IgG and rabbit antimouse IgG immobilized in a macroporous network have been used for an antigen inhibition assay and sandwich assay, respectively. The product of the enzymatic indicator was measured fluorimetrically.

There are many applications of FIA in the control of biotechnological processes, some of which are listed in Table 2.

## Food and Beverage Industry

FIA affords a means of automating many traditional wet-chemical methods of food and beverage analysis, incorporating into them the rapid throughput and high reproducibility that characterize FIA methods. Analysis using standard or official chemistries may be achieved in a very short time (typically 30 s).

For example, Figure 3 shows an automated FIA system used for quality control of olive oil through the use of a distributed expert system. This system consists of four nodes: node 1 controls pump 1 and the spectrophotometer and connects the system to the PC; node 2 measures the total acidity; node 3 measures the peroxide; and node 4 is responsible for ultraviolet spectrophotometric measurements.

The use of FIA with immobilized enzymes combines the advantage of enzymatic analysis with the rapidity and versatility of flow techniques. Numerous methods have been based on this combination of techniques, and the number of applications in food and beverage analysis is increasing. Approximately 30% of the proposed methods use enzymes, 99% of which are immobilized. Very good results have been obtained by combining the specificity of the enzymatic reactions with the high sensitivity of chemiluminescent detection. Many oxidases react with their substrates to produce hydrogen peroxide, which can be quantified by chemiluminescence. Chemiluminescent (CL) monoenzyme sensors based on optical fibers have been developed for FIA determination of hydrogen peroxide, xanthine, and hypoxanthine using microbial peroxidase or xanthine oxidase. To develop bienzyme sensors for glutamate, lysine, and xanthine the microbial peroxidase was coimmobilized with the corresponding oxidases. FIA systems with CL detection and minicolumns with immobilized glucose oxidase, alcohol oxidase, cholesterol oxidase, glutamate oxidase, and lysine oxidase for glucose, ethanol, cholesterol, glutamate, and amino acid determination in foodstuffs and beverages show extremely promising features.

The lack of ready-made FIA methods of food and beverage analysis in the literature should not be a discouragement since it is relatively easy to adapt an existing classical method to FIA. All that is required is the empirical selection of manifold designs and reagent concentrations to produce outputs of suitable magnitude over the working range. On the negative side, the total analysis time for foods will be dictated by the need for sample extraction and possibly clean-up prior to FIA, but this need not to be a serious drawback and is equally applicable to other techniques.

**Table 2** FIA applications in the control of some biotechnological processes

Analyzed species	Detection <sup>a</sup>	Linear range	Sample
Total glucose	Amperometry	10–600 $\mu\text{mol l}^{-1}$	Fermentation process
Intracellular ATP	Chemiluminescence	0.01–1.0 $\text{mmol l}^{-1}$	Kefir fermentation
Glucose	Chemiluminescence	10 $\text{mg l}^{-1}$ –2 $\text{g l}^{-1}$	Lactic acid fermentation
Glucose	Spectrophotometry	500 $\text{mg l}^{-1}$ –3 $\text{g l}^{-1}$	Mammalian cell culture
Glucose	Amperometry	2–100 $\text{g l}^{-1}$	Fermentation broth
Fructose	Thermistor	0.1–0.75 $\mu\text{g ml}^{-1}$	Fermentation broth
Sucrose	Amperometry	0–2.5 $\text{g l}^{-1}$	Effluents from pharmaceutical industry
Ammonia	Spectrophotometry	0.1–2 $\text{mmol l}^{-1}$	Hybridoma cell culture
Glutamine	Spectrophotometry	0.07–2 $\text{mmol l}^{-1}$	Hybridoma cell culture
Glutamine	Chemiluminescence	0.001–1 $\text{mmol l}^{-1}$	Animal cell culture
Glutamate	Chemiluminescence	0.0001–1 $\text{mmol l}^{-1}$	Animal cell culture
L-Methionine	Thermistor	10–60 $\text{mmol l}^{-1}$	Fermentation broth
IgG	Fluorimetry	5–500 $\mu\text{g ml}^{-1}$	Mammalian cell culture
Lactic acid	Fiber optic biosensor	0–50 $\text{mmol l}^{-1}$	Lactic acid fermentation
Lactic acid	Chemiluminescence	0.1–2 $\text{g l}^{-1}$	Lactic acid fermentation
L-Lactate	Amperometry	1–60 $\text{g l}^{-1}$	Effluents from pharmaceutical industry
Extracellular proteins	Spectrophotometry	0.1–2 $\text{g l}^{-1}$	<i>Trichoderma</i> fermentation
L-Phenylalanine	Fluorimetry	0.1–1.2 $\text{mmol l}^{-1}$ , 1–30 $\text{mmol l}^{-1}$	<i>Rhodococcus</i> cultivation
Optical density	Spectrophotometry	< 200 $\text{mg l}^{-1}$	Biotechnological process
Extracellular hydrolases	Spectrophotometry	50–500 $\text{U ml}^{-1}$	Fermentation broth
Intracellular $\beta$ -Galactosidase activity	Spectrophotometry	50–5100 $\text{U ml}^{-1}$	<i>Escherichia coli</i> fermentation
Lipase activity	Spectrophotometry	0.6–5 $\text{U ml}^{-1}$	<i>Candida rugosa</i> fermentation
$\beta$ -Galactosidase activity	Spectrophotometry	25–2150 $\text{U ml}^{-1}$	<i>Escherichia coli</i> culture
Formate dehydrogenase activity	Spectrophotometry	0.05–20 $\text{U ml}^{-1}$	<i>Candida beldinii</i> culture
Esterase activity	Spectrophotometry	0.01–1.0 $\text{U ml}^{-1}$	Fermentation broth

Table 3 presents details of several FIA methods, together with some of their features, that have already been applied to the analysis of foods and beverages.

Beverages were the first area of application of FIA in the beverage and food industry since being liquids they require little or no sample preparation prior to analysis. The analysis of soft drinks for quality control purposes normally includes the determination of pH, sugars, acidity, preservatives, and trace metals (e.g., lead and cadmium); moreover, colas are also analyzed for caffeine, and soda water for hydrogen-carbonate. FIA may be utilized to carry out many of these determinations. FIA has also been employed for determinations more specific to beer, wine, and milk. FIA titration has been used to determine volatile acids in spoiled beer samples. Bitterness in beer is determined by solvent extraction from an acidified carrier stream and measurement of the absorbance of the solvent stream at 275 nm. The reducing sugars in wine have been determined by automation of classical colorimetric reactions using FIA.

A number of FIA applications are made possible by the fluorimetric determination of nicotinamide adenine dinucleotide in its reduced form at 365 nm (excitation) and 400 nm (emission). This way ethanol in wine and pyruvate in milk can be determined.

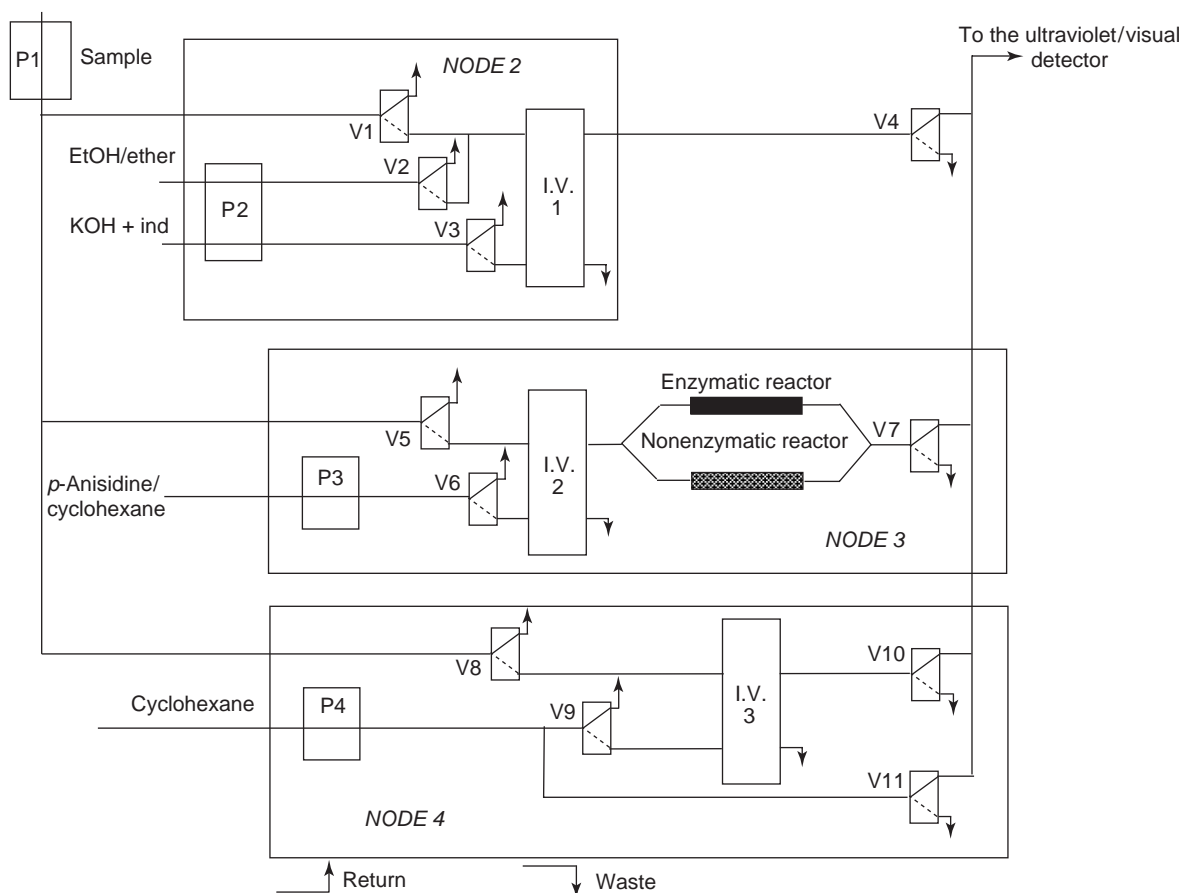
Table 3 gives some examples of FIA determinations in beverage and food samples in this context.

FIA offers a means of carrying out beverage and food analysis based on classical chemistry that is simple, precise, and quick. Rapid development of a method is possible, and changeover from one determination to another is quick and easy.

## Wastewater Monitoring

Nowadays, control and treatment of industrial effluents has become one of the most important steps of productive processes since the regulatory laws have been very rigorous about environment protection. As a consequence, automated, robust, and multiparametric methods capable of reducing drastically both the reagent consumption and the analyst manipulation together with the feasibility of achieving a high sample throughput need to be designed.

Although originally flow injection techniques were developed for off-site measurements, their outstanding feature is the capability for both *in situ* and real-time monitoring of chemical parameters in water. This way, the simultaneous determination of the relevant parameters in water samples such as ammonium, nitrite, nitrate, total nitrogen, orthophosphate, total phosphate, and total organic carbon is possible (Table 4).



**Figure 3** Configuration of a distributed FIA system for quality control of olive oil. Ind, indicator (phenolphthalein); P, peristaltic pump; I.V., injection valve; V, electrovalve. (Reproduced with permission from Bonastre A, Ors R, and Peris M (2004) Advanced automation of a flow injection analysis system for quality control of olive oil through the use of a distributed expert system. *Analytica Chimica Acta* 506: 189–195.)

Different miniaturized flow injection analyzers have been constructed using spectrophotometry as a leading detection technique. Besides, electrochemical techniques such as amperometry and potentiometry with chemically modified solidstate electrodes and tubular membrane-based ISEs, respectively, have been proved to be well adapted to multiparametric measurements of inorganic species present in wastewater, using multiple sensor arrays.

The advantage of the FIA technique lies in the possibility of implementing heterogeneous reactions, dilution procedures, or separation steps using inline gas-permeable and dialysis units.

Phenolic compounds and especially their chlorinated derivatives are important water pollutants that are present in wastewater discharged from a variety of industrial sources. A fast method for direct determination of phenols in wastewater based on solid-phase extraction coupled online to a flow injection manifold has been developed. The phenol

preconcentration is carried out in a column packed with Amberlite XAD-4 resin and inserted in the FIA manifold. The eluted phenols are spectrophotometrically determined as products of reaction with 4-aminoantipyrine and potassium ferricyanide.

A rapid flow injection analyzer for successive determination of ammonia, nitrite, and nitrate in water using gas-phase chemiluminescence has been used with good results for water quality control in sewage plants and water purification plants.

A low-cost and robust FIA method for total phosphorus in wastewater is based on a two-stage photooxidation/thermal digestion procedure together with a mixed oxidizing/hydrolyzing reagent for converting phosphorus compounds to orthophosphate. The formed orthophosphate is then determined by the phosphomolybdenum blue reaction.

One company has installed complete water monitoring stations based on the flow injection principle with colorimetric, fluorimetric, and ISE detection, and these have been functioning for many years now.

**Table 3** Features of FIA determinations in beverage and food samples

Analyzed species	Measuring range	Type of sample	Detection
Glucose	0.01–3 mmol l <sup>-1</sup>	Foodstuffs	Amperometric biosensor
Sucrose	0.04–2 mmol l <sup>-1</sup>	Soft drinks	Chemiluminescence
Sucrose	11–14% (m/m)	Sugarcane	Spectrophotometry
Glucose	0.2–1.0 mmol l <sup>-1</sup>	Soft drinks	Spectrophotometry
Fructose	0.02–2 mmol l <sup>-1</sup>	Common fruits	Amperometry
Lactose	0.1–1.5 mmol l <sup>-1</sup>	Milk	Spectrophotometry
Lactose	25–250 mmol l <sup>-1</sup>	Milk	Amperometry
Raffinose	0.5–6 mmol l <sup>-1</sup>	Foodstuffs	Amperometry
Glucose	0.001–0.01 mmol l <sup>-1</sup>	Fruit juice	Fluorimetry
Starch	0.1–1000 mg l <sup>-1</sup>	Potatoes	Amperometry
Ethanol	6–12% (m/m)	Wine, spirits	Enthalpimetry
Ethanol	0.002–0.016%	Wine	Photodiode array
Glycerol	2200 mmol l <sup>-1</sup>	Wine	Amperometry
Acetaldehyde	0.5–11 mg l <sup>-1</sup>	Wines	Spectrophotometry
SO <sub>2</sub> , free total	1–16 mg l <sup>-1</sup>	Wines	Spectrophotometry
Sulfite	1–10 mg l <sup>-1</sup>	Juices, vegetables	Spectrophotometry
Tartaric acid	0.02–4 g l <sup>-1</sup>	Wines	Spectrophotometry
Glutamic acid	0.01–0.5 mmol l <sup>-1</sup>	Cheese	Fluorimetry
Ascorbic acid	0.140 mg l <sup>-1</sup>	Fruit juice, jam	Spectrophotometry
L-Lactate	20–650 mg l <sup>-1</sup>	Milk, foodstuffs	Spectrophotometry
Tartaric acid	0.6–4 g l <sup>-1</sup>	Wine	Potentiometry
Citric acid	0.1–10 mmol l <sup>-1</sup>	Soft drinks	Potentiometry
NH <sub>4</sub> <sup>+</sup>	1–8 mmol l <sup>-1</sup>	Rice, orange, coffee, soya	Conductimetry
Ammonia	0.1–40 mg l <sup>-1</sup>	Beer	Spectrophotometry
Urea	0.02–0.04 mmol l <sup>-1</sup>	Cow's milk	Spectrophotometry
Selenium	2–10 mg l <sup>-1</sup>	Wheat, corn	Atomic absorption spectrometry
Lead	0.05–0.5 mg kg <sup>-1</sup>	Powdered food	Inductively coupled plasma–mass spectrometry
Iodide	5–2000 mg kg <sup>-1</sup>	Common salt	Amperometry
Chloride	250–5000 mg l <sup>-1</sup>	Milk	Amperometry
Boron	0–3 mg l <sup>-1</sup>	Plants	Spectrophotometry
Protein	1.0–4.0%	Dairy products	Amperometry
Cholesterol	0–800 mg kg <sup>-1</sup>	Butter	Amperometry
Gentamicin	0–100 µg kg <sup>-1</sup>	Milk	Amperometry

**Table 4** Features of FIA determinations in wastewater

Analyzed species	Detection	Measuring range	Throughput (samples per hour)
Ammonia	Spectrophotometry	1–100 mg of N per liter	13
Ammonia	Conductimetry	0–49 mg of N per liter	9
Ammonia	Potentiometry	0.8–140 mg of N per liter	25
Ammonia	Chemiluminescence	1.4 µg of N per liter to 1.4 mg of N per liter	20
Nitrate	Spectrophotometry	0.056–11.3 mg of N per liter	8
Nitrate	Spectrophotometry	0.022–0.790 mg of N per liter	30
Nitrite	Spectrophotometry	0.015–1.520 mg of N per liter	8
Nitrite	Spectrophotometry	0.003–0.670 mg of N per liter	30
Phosphorus species	Spectrophotometry	0.05–4.0 mg of P per liter	30
Phosphorus species	Potentiometry	3.1–310 ng of P per liter	20
Cationic surfactants	Spectrophotometry	0.5–30 mg of CTMB <sup>a</sup> per liter	100
Phenol	Spectrophotometry	0.01–1 µg ml <sup>-1</sup>	12
Nickel	Spectrophotometry	0.03–4 µg ml <sup>-1</sup>	60
Copper	Spectrophotometry	0.04–4 µg ml <sup>-1</sup>	60
Iron(II)	Chemiluminescence	5 nmol l <sup>-1</sup> –1 µmol l <sup>-1</sup>	36
Monochloroacetic acid	ISE	10 <sup>-4</sup> –10 <sup>-1</sup> mol l <sup>-1</sup>	120
ClO <sub>2</sub>	Spectrophotometry	0.005–1.5 µg ml <sup>-1</sup>	15

<sup>a</sup>CTMB, cetyltrimethylammoniumbromide.

With this system it is possible to determine a large category of water contaminants and pollutants, e.g., heavy metals (iron, lead, manganese, chromium,

cadmium), phenols, ammonia, nitrite, nitrate, hydrazine, toluene, cyanide, sulfite, sulfide, phosphate, calcium, magnesium.



See also: **Clinical Analysis:** Glucose. **Enzymes:** Immobilized Enzymes. **Fluorescence:** Instrumentation. **Food and Nutritional Analysis:** Overview; Soft Drinks; Alcoholic Beverages. **Process Analysis:** Overview; Bioprocess Analysis.

## Further Reading

- Fang Z (1993) *Flow Injection Separation and Preconcentration*. Weinheim: VCH.
- Hansen EH (1990) Flow injection analysis. A versatile tool for biotechnological applications. *Wissenschaftliche Zeitschrift TH Leuna-Merseburg* 32: 373–387.
- Karlberg B and Pacey GE (1989) *Flow Injection Analysis: A Practical Guide*. Amsterdam: Elsevier.
- Keay PJ and Wang W (1997) Applications of flow injection analysis to analytical biotechnology. *Trends in Biotechnology* 15: 76–81.
- Lemieux L, Puchades R, and Simard RE (1989) Application of FIA techniques to food analysis. *Lebensmittel-Wissenschaft und Technologie* 22: 254–263.
- Miró M, Estela JM, and Cerdà V (2003) Application of flowing stream techniques to water analysis. Part I. Ionic species: dissolved inorganic carbon, nutrients and related compounds. *Review. Talanta* 60: 867–886.
- Osborne BG and Tyson JF (1988) Review: flow injection analysis – a new technique for food and beverage analysis. *International Journal of Food Science and Technology* 12: 541–554.
- Růžička J and Hansen EH (1988) *Flow Injection Analysis*, 2nd edn. New York: Wiley.
- Růžička J and Hansen EH (1998) Flow injection analysis – where are we heading?. *Trends in Analytical Chemistry* 17: 69–73.
- Schügerl K (2001) Progress in monitoring, modeling and control of bioprocesses during the last 20 years. *Journal of Biotechnology* 85: 149–173.
- Sosnitza P, Irtel F, Ulber R, et al. (1998) Flow injection analysis system for the supervision of industrial chromatographic downstream processing in biotechnology. *Biosensors & Bioelectronics* 13: 1251–1255.
- Trojanowicz M (2000) *Flow Injection Analysis: Instrumentation and Applications*. Singapore: World Scientific.
- Valcárcel M and Luque de Castro MD (1994) *Flow-Through (Bio)chemical Sensors*. New York: Elsevier.
- Valcárcel M, Luque de Castro MD, and Losada A (1987) *Flow Injection Analysis. Principles and Applications*. Chichester: Ellis Horwood.
- Van der Linden WE (1986) Flow injection analysis in online process control. *Analytica Chimica Acta* 179: 91–101.

# FLUORESCENCE

Contents

**Overview**

**Instrumentation**

**Multidimensional Fluorescence Spectrometry**

**High-Resolution Techniques**

**Time-Resolved Fluorescence**

**Derivatization**

**Fluorescence Labeling**

**Quantitative Analysis**

**Clinical and Drug Applications**

**Environmental Applications**

**Food Applications**

## Overview

**M E Díaz-García and R Badía-Laiño**, University of Oviedo, Oviedo, Spain

## Introduction

Molecular photoluminescence processes can be classified according to the mode in which a molecule is promoted to an electronically excited state and to the type of molecular excited state (see **Table 1**). As the excited molecule decays back to the ground state, or to a lower-lying excited electronic state, light is

See also: **Clinical Analysis:** Glucose. **Enzymes:** Immobilized Enzymes. **Fluorescence:** Instrumentation. **Food and Nutritional Analysis:** Overview; Soft Drinks; Alcoholic Beverages. **Process Analysis:** Overview; Bioprocess Analysis.

## Further Reading

- Fang Z (1993) *Flow Injection Separation and Preconcentration*. Weinheim: VCH.
- Hansen EH (1990) Flow injection analysis. A versatile tool for biotechnological applications. *Wissenschaftliche Zeitschrift TH Leuna-Merseburg* 32: 373–387.
- Karlberg B and Pacey GE (1989) *Flow Injection Analysis: A Practical Guide*. Amsterdam: Elsevier.
- Keay PJ and Wang W (1997) Applications of flow injection analysis to analytical biotechnology. *Trends in Biotechnology* 15: 76–81.
- Lemieux L, Puchades R, and Simard RE (1989) Application of FIA techniques to food analysis. *Lebensmittel-Wissenschaft und Technologie* 22: 254–263.
- Miró M, Estela JM, and Cerdà V (2003) Application of flowing stream techniques to water analysis. Part I. Ionic species: dissolved inorganic carbon, nutrients and related compounds. *Review. Talanta* 60: 867–886.
- Osborne BG and Tyson JF (1988) Review: flow injection analysis – a new technique for food and beverage analysis. *International Journal of Food Science and Technology* 12: 541–554.
- Růžička J and Hansen EH (1988) *Flow Injection Analysis*, 2nd edn. New York: Wiley.
- Růžička J and Hansen EH (1998) Flow injection analysis – where are we heading?. *Trends in Analytical Chemistry* 17: 69–73.
- Schügerl K (2001) Progress in monitoring, modeling and control of bioprocesses during the last 20 years. *Journal of Biotechnology* 85: 149–173.
- Sosnitza P, Irtel F, Ulber R, et al. (1998) Flow injection analysis system for the supervision of industrial chromatographic downstream processing in biotechnology. *Biosensors & Bioelectronics* 13: 1251–1255.
- Trojanowicz M (2000) *Flow Injection Analysis: Instrumentation and Applications*. Singapore: World Scientific.
- Valcárcel M and Luque de Castro MD (1994) *Flow-Through (Bio)chemical Sensors*. New York: Elsevier.
- Valcárcel M, Luque de Castro MD, and Losada A (1987) *Flow Injection Analysis. Principles and Applications*. Chichester: Ellis Horwood.
- Van der Linden WE (1986) Flow injection analysis in online process control. *Analytica Chimica Acta* 179: 91–101.

# FLUORESCENCE

Contents

**Overview**

**Instrumentation**

**Multidimensional Fluorescence Spectrometry**

**High-Resolution Techniques**

**Time-Resolved Fluorescence**

**Derivatization**

**Fluorescence Labeling**

**Quantitative Analysis**

**Clinical and Drug Applications**

**Environmental Applications**

**Food Applications**

## Overview

**M E Díaz-García and R Badía-Laiño**, University of Oviedo, Oviedo, Spain

## Introduction

Molecular photoluminescence processes can be classified according to the mode in which a molecule is promoted to an electronically excited state and to the type of molecular excited state (see **Table 1**). As the excited molecule decays back to the ground state, or to a lower-lying excited electronic state, light is

emitted at a characteristic wavelength. All the processes that involve the emission of electromagnetic radiation are called luminescence.

Following absorption, a number of vibrational levels of the excited state are populated. Molecules in these higher vibrational levels then decay to the lowest vibrational level of the excited state (vibrational relaxation). Fluorescence is the radiation released when a molecule, which has been promoted to an excited singlet state by light absorption, rapidly relaxes from the lowest vibrational mode of the electronically excited state  $S_1$  (Kasha's rule) to a vibrational mode of the electronic ground state  $S_0$ . The energy transitions associated with photoluminescence can be represented on Morse energy surface diagrams (Figure 1). In addition to fluorescence, the energy of the excited state may be dissipated by

nonradiative decay processes, such as fluorescence energy transfer, collisional quenching, or intersystem crossing.

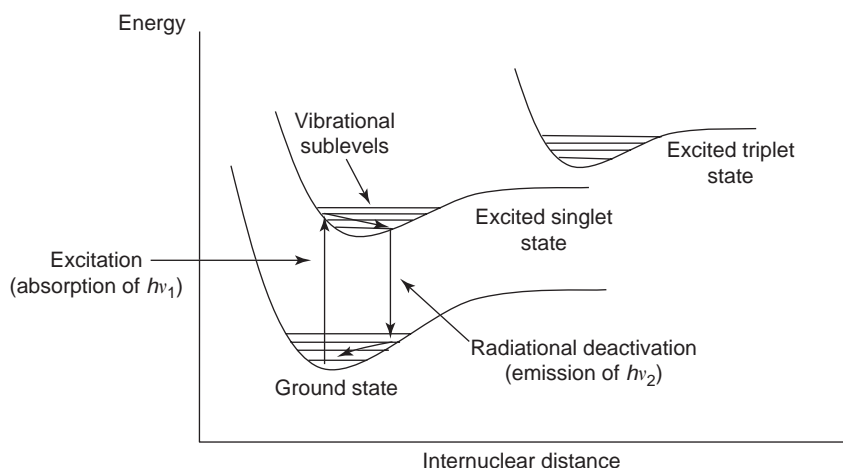
Fluorescence spectrometry has become established as a routine technique in many specialized applications due to its high sensitivity. Examples of the current applications of fluorescence range from simple fluorimetric analysis for biomolecules, metal ions, and organic compounds to identification of specific DNA and/or RNA sequences in tissues. In addition, much of the current research in biochemistry, medicine, and molecular biology involves fluorescence spectroscopy of either intrinsic molecular fluorescence (from tyrosine or tryptophan residues) or exogenous fluorescent probes.

Besides high sensitivity, fluorescence exhibits unique performance characteristics among which selectivity, reproducibility, and temporal resolution are included. It is relatively a simple and rapid analytical technique, specially adequate for quantification of aromatic, or highly unsaturated, organic molecules present at trace concentrations, especially in biological and environmental samples. The technique can also be applied to a wide variety of organic and inorganic compounds via chemical labeling and derivatization procedures and can be used as the detection mode in flow injection analysis, chromatography, capillary electrophoresis, and thin-layer chromatography. Fluorescence output is linear to sample concentration over a very broad range and the technique can be used over three to six decades of concentration without sample dilution or modification of the sample cell. On the other hand, almost any liquid, gaseous, or solid sample can be analyzed by fluorescence. Sample size can be extremely small and nanoliter sample volumes may be analyzed using

**Table 1** Molecular photoluminescence processes

	<i>Luminescence process</i>
<i>(A) Excitation mode</i>	
Absorption of radiation (UV–Vis)	Photoluminescence
Chemical reaction	Chemiluminescence, bioluminescence
Thermally activated ion recombination	Thermoluminescence
Friction	Triboluminescence
Sound waves	Sonoluminescence
High-energy particles, radiation	Radioluminescence
Injection of charge	Electroluminescence
<i>(B) Excited state</i>	
First excited singlet state	Fluorescence, delayed fluorescence
Lowest triplet state	Phosphorescence

UV–Vis, Ultraviolet–visible



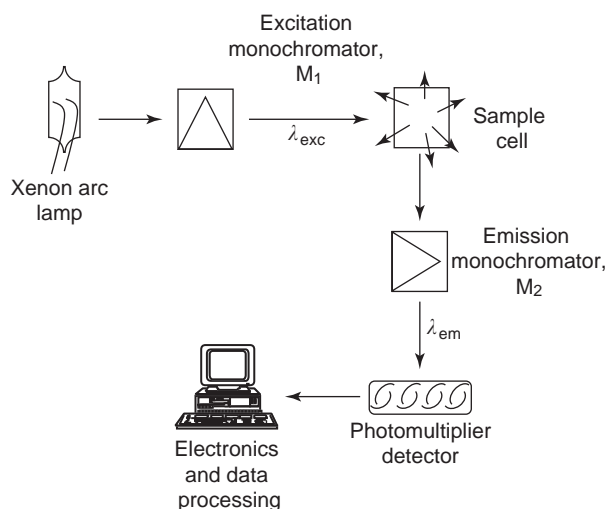
**Figure 1** Simplified energy level diagram of a polyatomic molecule.

specialized techniques. Even, using sophisticated techniques (e.g., flow and image cytometry), single molecules can be studied via fluorescence, while multiple targets can be detected during a single assay by using a combination of fluorophores.

Fluorescent dyes and tailor-made fluorescence-labeled compounds are growing commercially for coupling not only to metal ions but also to biomolecules, as a result of the increasing demand for identification of molecular interactions and for the visualization, recognition, and quantification of molecules, which bind to target molecules like peptides, proteins, nucleic acids, carbohydrates, etc. Companies have built the basic chemistries into user-friendly kits and systems that make labeling easy. Fluorescence spectroscopy has evolved to a routinely used ultra-sensitive detection technology on high-throughput screening, from a 96-well situation up to high-density 384- and 1536-well platforms. An integration of combinatorial chemistry and fluorescence spectroscopy creates a powerful tool to probe biological targets. A consistent theme of fluorescent technologies is their adaptability to users' needs in both industry and academe.

## Basic Instrumentation for Molecular Fluorescence Measurement

A block diagram of a basic fluorimeter is shown in Figure 2. Aside from the optical components shown, modern fluorimeters have dedicated computers, which may control instrumental operating parameters, the acquisition of spectral data and the post-processing of the data.



**Figure 2** Basic component of a spectrofluorimeter.

## Sources

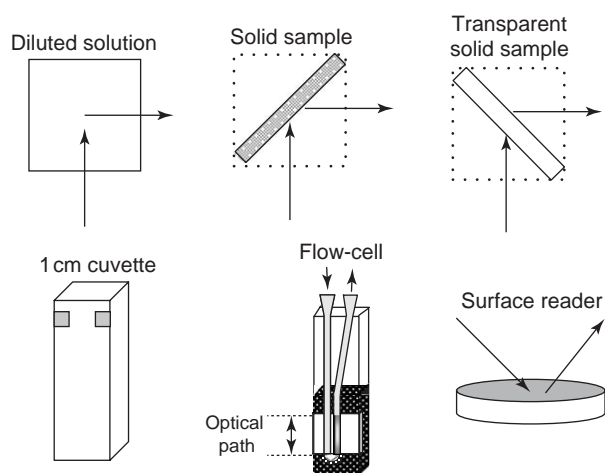
The source used in most commercial fluorimeters is the 75–450 W xenon arc lamp, which covers the range of 200–1000 nm. However, for greater excitation energy at selected wavelengths, the mercury-arc lamp may be used, although excitation spectra obtained with this lamp are usually severely distorted because principal lines (at 254, 302, 313, 546, 587, 691, and 773 nm) are superimposed on a continuum. Lasers are now increasingly used due to the high optical power (number of photons per unit time). A few fluorescence instruments using laser sources are commercially available and most are intended for specific applications such as analyses of uranium in the nuclear industry and for remote measurements using optical fibers. In some applications, such as frequency domain lifetime measurements, it may be necessary to modulate the intensity of the source. In these cases, some classes of laser or light-emitting diodes are suitable sources.

## Wavelength Selection

Isolation of selected wavelengths is carried out with filters, usually in conjunction with white light sources. Filters are used in inexpensive, portable fluorescence spectrometers and their optical quality will often determine the performance of the instrument. Filters allow transmission of a very large number of photons from the source to the sample and from the sample to the detector (fluorescence signal is maximized) but have poor wavelength selectivity. While filters can be used for limited applications, grating monochromators are used in most spectrofluorimeters. For optically clear solutions, single-grating monochromators are adequate while double-grating monochromators are ideal for turbid solutions as stray light and scattering is reduced. The emission wavelength selector system is generally placed at an angle of 90° with respect to the excitation axis to minimize interferences from transmitted and scattered exciting light.

## Sample Illumination

The most common arrangement uses the 90° geometry depicted in Figure 2 rather than the 180° geometry common to absorption spectrophotometers. This geometry is suitable for weakly absorbing solution or diluted samples. For solids (samples adsorbed on solid surfaces such as polymers, paper, etc.) and for solutions that absorb strongly at the excitation wavelength, a front surface geometry is preferable (see Figure 3). Fluorescence is viewed in these cases from the face of the sample on which the exciting light impinges. In the front-face arrangement there is



**Figure 3** Sample geometry and typical cell for fluorescence measurements.

a risk that reflected light from the surface enters the emission monochromator resulting in large amount of stray light. If the solid sample is transparent, back-face illumination is also possible.

In practice, most fluorescence measurements are taken in solution and 1 cm glass or fused silica cuvettes with four polished windows are used. For some specialized applications different types of cuvettes are commercially available. For example, flow cells of different configurations for continuous flow spectrofluorimetric analysis have been designed.

### Detectors

The final basic component of the fluorimeter is the detector, which is placed at the exit slit of the emission-wavelength selection system. A key requirement for a detector is its ability to detect very low radiation levels. Photomultiplier tubes (PMTs) are most widely used because of their high sensitivity. In many cases, PMTs are operated at low temperature to minimize dark counts and to push detection limits lower. The essential principle of the PMT has remained unchanged for many years and its main limitation is that it is a single-channel detector and some important analytical information may be lost when obtaining a spectrum (e.g., in chromatographic applications). An array of detectors, as in many UV-Vis absorption spectrometers, would allow rapid acquisition of full fluorescence spectra. Until now the PMT-based devices remain the most common detectors although new classes of electronic array detectors are beginning to be employed in some fluorimeters. At present, one of the most promising electronic array detectors for fluorimetry is the charge-coupled device (CCD), which essentially is an array of semiconductor photodetectors.

### Information from Fluorescence Measurements

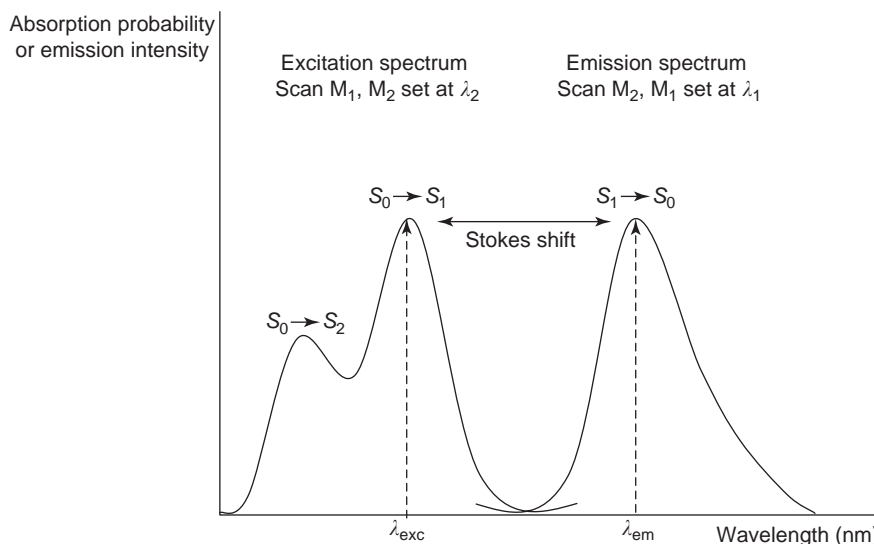
All fluorescent molecules can be characterized by two types of spectra – the excitation spectrum and the emission spectrum. The excitation spectrum is a plot of emitted fluorescence as a function of the excitation wavelength at a fixed emission wavelength. The excitation spectrum represents the relative probability that a fluorescent molecule will be excited by a given wavelength of incident light. If the excitation energy is constant, the fluorescence excitation spectrum is very similar to the absorption spectrum. The photon energy at the maximum of the excitation peak equals the energy difference between the ground state ( $S_0$ ) and a favored vibrational level of the first excited state ( $S_1$ ) of the molecule. In some cases, the excitation spectrum shows a second peak at a shorter wavelength (higher energy) that indicates transition of the molecule from the ground state to the second excited state ( $S_2$ ) (Figure 4).

A plot of the relative intensity of emitted light as a function of the emission wavelength at a fixed excitation wavelength is termed fluorescence emission spectrum. The fluorescence emission is characterized by the transition from the lowest vibrational mode of the electronically excited state ( $S_1$ ) to the ground state. Therefore, the shape of the emission spectrum is always the same and is independent of the wavelength of the exciting radiation.

The shape of the emission spectrum is approximately a mirror image of the longest-wavelength excitation band. Peaks in a fluorescence excitation spectrum usually correspond closely in wavelength to absorption peaks. In theory, the maximum of the emission peak should occur at the same wavelength than that of the excitation peak (commonly called the O–O transition of absorption and emission). In practice, the emission spectrum is always shifted toward a longer wavelength relative to the excitation spectrum, as shown in Figure 4. This difference between the excitation and emission maxima is termed the Stokes shift. The Stokes shift represents the energy lost whilst the molecule was in the excited state and, from a practical point of view, it allows the excitation and emission peaks to be spectrally separated and easily distinguished. When the spectra are scanned, ‘scattering peaks’ usually appear when the wavelengths of excitation and emission monochromators coincide. They can be eliminated if the excitation and emission band passes do not overlap.

The absorption spectrum is reproducible when scanned on different instruments (provided there are no distortions due to inappropriate settings of scan speed, band pass width, source output, etc.).





**Figure 4** Typical features of fluorescence spectra.

However, fluorescence spectra (excitation and emission) are less reproducible because fluorescence spectra are affected by intensity of light source and by the response of photomultiplier detector, both of which vary with the wavelength. Thus, excitation and fluorescence spectra, obtained in the usual nonratio mode with single-beam instruments, are often distorted and not reproducible from instrument to instrument. It is possible to eliminate these variations instrumentally and several commercial instruments are available that provide 'corrected spectra'. These adjust for variations in the source intensity with wavelength and also correct for variations of the detector response.

The emission wavelength and the fluorescence intensity are determined by the structure of the molecule. In principle, any molecule that absorbs radiation of adequate energy could fluoresce. However, many molecules exhibit very weak fluorescence and only a small fraction of molecules exhibit analytically useful fluorescence. Most intensely fluorescent molecules contain highly conjugated  $\pi$ -electron systems. For example, polycyclic compounds such as aromatic hydrocarbons, vitamin K, barbiturates, nucleosides, and conjugated polyenes such as vitamin A are fluorescent. Typically, more the number of conjugated bonds in the molecule, longer is the wavelength of emission observed. The presence of heteroatoms such as O, S, and N, results in  $n \rightarrow \pi^*$  transitions that may promote a change in the spin of the excited electron (intersystem crossing process) and no fluorescence is observed (e.g., pyridine, pyrrole, furan, and thiophene are not fluorescent). Molecules with fused aromatic rings have high molar absorptivity values and, consequently, are highly fluorescent (e.g., naphthalene, pyrene, anthracene). The

presence of heteroatoms in such systems can either decrease or increase fluorescence intensity, depending on the probability of intersystem crossing. Molecules in which it is possible to stabilize charge by resonance are much more fluorescent than those with no resonance structures. A typical example of this is aniline, which is highly fluorescent while anilinium ions are not fluorescent. Finally, more rigid molecules exhibit stronger fluorescence owing to the lower probability of energy dissipation by nonradiative processes such as energy transfer or by transitions between electronic states. So, while biphenyl is weakly fluorescent, fluorene is strongly fluorescent (quantum yield of 1.0). The applicability of fluorescence may be extended to nonfluorescent compounds by converting them to a fluorescent derivative. For example, nonfluorescent steroids may be converted to fluorescent phenolic compounds by dehydration with concentrated sulfuric acid. Antibodies may be made to fluoresce by attaching to it a fluorescent tag or label (e.g., fluorescein isocyanate). Concerning inorganic species, relatively few inorganic metal ions (e.g.,  $\text{UO}_2^{2+}$ , Ce(III), Tl(I)) exhibit intense fluorescence. Nonfluorescent inorganic metal ions may be reacted with aromatic organic ligands to form fluorescent chelates. Considerations about rigidity, conjugated bonds, and aromatic fused rings also apply in these cases.

## Molecular Fluorescence: Practical Considerations

Conventional fluorimetric determinations are carried out in solution with external standardization. When



the native fluorescence of the analyte is measured, minimal sample treatment is necessary and a fluorimetric analysis can be carried out in less than 10 min. However, when fluorimetric derivatization reactions are necessary, the analytical reaction is allowed to reach equilibrium before the fluorescence signal is measured. In any case, the fluorescence signal is related to the analyte concentration.

Solution conditions, such as pH, viscosity, ionic strength, solvents, and reagent concentrations must be carefully adjusted and controlled in order to maximize the fluorescence signal. Background fluorescence from sample matrix or contaminants in solvents, reagents, and/or laboratory glassware can occur over the same wavelength range as the analyte fluorescence (additive interference). This can be a major limitation in achieving optimal detection limits, especially in biological and environmental samples. Background fluorescence can be circumvented through different ways, depending on its origin or nature. For example, the analyte can be separated from the interferent matrix before measurement of fluorescence. Time resolution may be useful when background fluorescence has a different decay time from that of the analyte. Measurement of analyte fluorescence must be done at wavelengths at which other sample components do not absorb or fluoresce.

Fluorescent molecules are subject to intensity variations as a function of temperature. In general, the frequency of collisional deactivation with solvent molecules increases as the temperature increases. These collisions bleed energy from the excited state and a fluorescence intensity decrease results. Temperature coefficients are typically 1–2% per °C. A thermostated sample cell is recommended to ensure temperature control both in samples and in standards.

Photodecomposition can be a serious problem in fluorescence during exposure to the exciting radiation. The degree of photodegradation is proportional to the exposure time and to the incident radiant power. Photodegradation can be minimized through the use of a longer-wavelength excitation band of the analyte, if one is available, or through the use of smaller excitation slit width in order to reduce the incident radiant power.

## Analytical Techniques

### Steady-State Fluorescence

The most widely used analytical approach in fluorescence is the use of steady-state fluorescence, where the intensity of the emitted light by the fluorochrome is measured. In general terms, the intensity of emitted light,  $I_f$ , that is measured by the instrument is

proportional to the fluorochrome concentration and can be described by the following relationship:

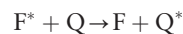
$$I_f = 2.303 \Phi_f I_0 \epsilon b c$$

where  $I_f$  is the observed fluorescence,  $\Phi_f$  is the quantum yield,  $I_0$  is the power of the incoming light,  $\epsilon$  is the molar extinction coefficient,  $c$  is the fluorophore concentration, and  $b$  is the optical pathlength. If  $\Phi_f$ ,  $I_0$ ,  $\epsilon$ , and  $b$  remain constant, the relationship between the fluorescence intensity and the fluorophore concentration is linear. This linearization is valid only for cases where  $\epsilon b c < 0.05$  (diluted solutions). Above this value, nonlinearity occurs. At higher concentrations significant absorption of the excitation beam radiation may take place (primary absorption), thus reducing the number of excited molecules across the light path. This effect causes negative deviation in a concentration–response plot. Also, if the excitation and emission spectra overlap, significant absorption of the emission beam takes place at high analyte concentrations, thus causing further nonlinearity (secondary absorption). The secondary absorption, when due to the analyte, is specifically denoted as self-absorption. Primary and secondary absorption together are called the inner filter effect.

### Fluorescence Quenching

The fluorescence signal generated by a molecule can be strongly altered by its environment. Apart from inner filter effects, the interaction of the fluorochrome with its surroundings is another way in which its fluorescence intensity can be decreased. This process is known as quenching. There are two kinds of quenching processes: (1) dynamic quenching and (2) static quenching. The efficiency of the two processes is dependent on the concentration of the quencher molecules.

Dynamic quenching or collisional quenching normally refers to nonradiative energy transfer from excited species to other molecules:



Here, the excited analyte molecule,  $F^*$ , transfers excitation energy to a quencher molecule,  $Q$ , causing deactivation of  $F$  and forms an excited quencher molecule,  $Q^*$ . For collisional quenching, the decrease in intensity often follows the well-known Stern–Volmer equation:

$$I_0/I = 1 + K[Q] = 1 + k_q \tau_0 [Q] \quad [1]$$

where  $I_0$  and  $I$  are the fluorescence intensities for the analyte in the absence of quencher and presence of quencher at concentration  $[Q]$ , respectively,  $K$  is the

Stern–Volmer quenching constant,  $k_q$  is the bimolecular quenching constant, and  $\tau_0$  is the unquenched lifetime. The Stern–Volmer quenching constant can be estimated from a plot of  $(I_0/I) - 1$  as a function of the quencher concentration. Equation [1] shows that the effect of a quencher decreases as the sample is diluted because the probability of collision between the analyte and the quencher is minimized. Also, the probability of collisions between the excited fluorophore and the quencher is higher as higher is the lifetime of the fluorophore. A wide variety of molecules can act as collisional quenchers, among which small molecules such as oxygen, halogens, amines, and electron-deficient molecules like acrylamide are good quenchers for some fluorophores, particularly if the fluorescence lifetime is greater than 10 ns.

In static quenching, the quencher and the fluorophore in the ground state form a stable complex. The process can be described by the following mechanism:



In this case, fluorescence is only observed from the unbound fluorophore. The decrease in fluorescence intensity is described by eqn [1] where  $K$  must be replaced by  $K_q$ , the constant of complex formation ( $K_q = [FQ]/[F] \times [Q]$ ). An example of this type of interaction is the quenching of ethidium bromide by caffeine.

In contrast to dynamic quenching, static quenching does not show an additional dependence on the lifetime of the excited state of the fluorophore. On the other hand, while formation of a complex just reduces the concentration of the free fluorophores without affecting the lifetime of the excited molecules, the lifetime of the complex is significantly longer than the lifetime of the excited fluorophore. So, measurement of the lifetime provides a means of distinguishing between dynamic and static quenching. Quenching studies provide direct information about diffusion of small molecules in solutions, in cellular tissues and in solid materials, and hence give insights into solvent or solid viscosities and/or accessibilities.

### Energy Transfer

In contrast to fluorescence static or dynamic quenching, for which a coupling between electronic orbitals

is necessary, energy transfer can also take place over large distances (up to 100 Å) without collisions. It occurs between two fluorescent molecules when one, the donor, absorbs a photon, elevating an electron to a higher energy state and, through resonance, the excitation of this electron is passed to another electron in the second, the acceptor. This energy is then re-emitted as a photon that is less energetic, and therefore of longer wavelength than the photon initially absorbed. The emitted radiation from the acceptor is termed sensitized fluorescence as it is observed without direct excitation of the acceptor. For energy transfer, some basic conditions must be satisfied:

1. the donor molecule must be a fluorophore with a sufficiently long fluorescence lifetime;
2. the emission spectrum of the donor and the excitation spectrum of the acceptor must overlap partially; and
3. the distance between donor and acceptor must be within a limiting range (usually 20–50 Å).

The rate of energy transfer ( $k_T$ ) to a specific acceptor is given by

$$k_T = \tau_D^{-1} (R_0/R)^6 \quad [5]$$

where  $\tau_D$  is the luminescence lifetime of the donor,  $R$  is the distance between the donor and the acceptor, and  $R_0$  is the critical distance (Föster distance) between the donor–acceptor pair for which the probability of energy transfer and the deactivation of the donor by radiative and nonradiative processes is the same. When the donor–acceptor distance is equal to the Föster distance, the transfer efficiency is 50%. Energy transfer is widely used in biochemistry to measure protein association, distances between two sites on a macromolecule, and the effects of conformational changes on these distances. It is a powerful method for obtaining both structural and dynamic information about macromolecules and macromolecular complexes.

### Synchronous Fluorescence

A synchronous spectrum, with a constant energy difference ( $\Delta\nu$ ), is performed by scanning the emission monochromator at a slightly faster rate than the excitation monochromator ( $\Delta\lambda = \lambda_{em} - \lambda_{ex}$ ). Synchronous spectra are strongly dependent on the wavelength offset. Ideally, only one peak is obtained when  $\Delta\nu$  is set to the Stokes shift of the fluorophore of interest. The main purpose of synchronous scanning is to generate spectra having decreased

bandwidths. Synchronous scanning may decrease the extent of overlapping in the spectra of mixtures of fluorescent compounds in multicomponent fluorescence applications.

### Time-Resolved Fluorescence

In contrast to steady-state fluorimetry, time-resolved fluorimetry is based on the measurement, at a fixed wavelength, of fluorescence signal as a function of time. For these measurements, the sample is exposed to a pulse of light (pulse fluorimetry), where the pulse width is short in comparison with the excited state lifetime of the fluorescent molecule. The decay of the fluorescence is then recorded with a high-speed detection system (nanosecond timescale). An alternative technique, phase-modulation fluorimetry (often called frequency-domain fluorimetry), uses a source that is amplitude-modulated at one or more frequencies. Measurement of the phase or demodulation of the fluorescence signal can be used to generate fluorescence decay times and time-resolved fluorescence spectra. Commercial instrumentation is available for these types of measurements. Time-resolved spectra or measurement of fluorescence decay times reveal much of the molecular information available from fluorescence. For example, it is possible to distinguish: (1) sample constituents whose fluorescence spectra overlap one another; (2) between static and dynamic quenching; and (3) the fluorescence of an analyte from background scattering.

### Fluorescence Anisotropy

Anisotropy measurements are based on the photoselective excitation of fluorophores by plane-polarized light. In an isotropic medium, the fluorophores are randomly oriented. Upon excitation with polarized light, those fluorophores whose absorption transition dipole is aligned parallel to the electric vector of the excitation, will be preferentially excited. If the molecule rotates and tumbles out of this plane during the excited state, light is emitted in a different plane from the excitation light. The intensity of the emitted light can be monitored in vertical and horizontal planes and thus, fluorescence anisotropy ( $r$ ) and polarization ( $P$ ) are defined by:

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \quad [6]$$

$$P = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}} \quad [7]$$

where  $I_{\parallel}$  and  $I_{\perp}$  are the fluorescence intensities of the vertically ( $\parallel$ ) and horizontally ( $\perp$ ) polarized

emission when the sample is excited with vertically polarized light. Anisotropy and polarization are both expressions for the same phenomenon. Fluorescence polarization measurements are widely used in molecular biophysics and biochemistry for studying rotational motions of electronically excited molecules, to detect the binding of relatively small molecules to macromolecules, to quantify protein denaturation, internal dynamics of proteins as well as in fluoroimmunoassay procedures.

### Fluorescence Performance Characteristics

The major advantage of fluorescence methods is sensitivity. One can see why fluorimetric methods are so sensitive if we compare them with absorption spectroscopy. In absorption methods, the absorbance of a compound in the range of  $1.0 \times 10^{-3}$  absorbance units can be measured using a good absorption spectrophotometer. Assuming a very efficient absorber with an extinction molar coefficient of  $\epsilon \cong 1 \times 10^5 \text{ l mol}^{-1} \text{ cm}^{-1}$ , we can calculate a low detection limit  $C_{DL}$  of:

$$C_{DL} \cong 1.0 \times 10^{-3} / 1 \times 10^5 = 1 \times 10^{-8} \text{ mol l}^{-1}$$

using a 1 cm cell. On the other hand, for efficient luminescent molecules, with a good spectrofluorimeter in a low-background solution, detection limits in fluorescence spectroscopy easily approach concentrations in the range of 1–10 pM. Consequently, detection limits in fluorescence spectroscopy are generally more than three orders of magnitude better than those achieved for the same molecules in absorption spectroscopy. For example, the best experimental detection limit for efficient fluorescein, an efficient fluorophore, is  $\sim 0.1 \text{ pM}$ , using laser excitation and prepurified solvents. Using laser-induced fluorescence detection in capillary electrophoresis, detection limits of 0.1–100 amol (1 nl injection volume) have been reported for amino acids (precolum derivatization with dicarbocyanine fluorophore).

Precision, at concentrations well above the detection limit, is typically in the range 0.5–2% and it is limited by noise. In addition, sample treatment and calibration must be considered.

With external standardization, an equivalent analyte concentration in the standard and sample should yield the same analyte fluorescence signal. The accuracy of the determinations is dependent on interferences, chemical equilibrium involving the analyte, scattering, and quenching. Analyte interference due to absorption, scattering, or quenching may be

eliminated by dilution if the detection limit is low. Also, standard additions can be useful in these cases.

The linear range of molecular fluorescence can be four to six orders of magnitude for efficient fluorophores with low detection limits. As previously described, nonlinearity begins to occur at analyte concentrations where inner filter effects become significant.

## Applications

Fluorescence spectroscopy is one of the most widely used molecular spectroscopic techniques in the fields of molecular biology, biophysics, and biochemistry. Some important applications have already been outlined above, while describing the different fluorescent methods. Major classes of applications of fluorimetry include the following:

1. Determination of trace-level species, in clinical, biological and environmental samples, including inorganic species (metal ions, anions), organic, and biochemical compounds. Direct determination of inorganic species through the formation of fluorescent chelates may be very sensitive. This approach works well for diamagnetic metal ions as paramagnetic ones tend to quench fluorescence through intersystem crossing. Typical chelators include 8-hydroxy-quinoline and derivatives, morine and related flavonols, benzoine, etc. Indirect methods are based on measuring the quenching of fluorophores by inorganic species. This approach lacks sensitivity and selectivity. Among the organic species of environmental concern, polycyclic aromatic hydrocarbons are the most common analytes. Direct detection is possible for analytes of biomedical interest, such as FAD, NAD, porphyrins, and aromatic amino acids (endogenous fluorescence). For nonfluorescent compounds indirect detection is used by derivatization using selective fluorescent probes. Most proteins and peptides can be directly labeled with fluorophores via their available amine (lysine side chains) or thiol (cysteine side chains) groups. Isothiocyanates, such as fluorescein (FITC) and tetramethylrhodamine isothiocyanate (TRITC) are amine-reactive and are widely used for labeling. Succinimidyl esters react with thiol groups and sulfonyl chlorides (e.g., dansyl chloride) are reactive with amines and thiol groups. Fluorescent enzyme and immunoassays are widely used to determine a variety of analytes of biological and clinical interest.

2. Proteins, peptides, and nucleic acid labeling in molecular biology research. The fluorescence modification of nucleic acid molecules can be achieved in a number of ways. For example, fluorescent DNA

probes can be prepared chemically by preparing 5'-amino-end-labeled oligonucleotide probes on a DNA synthesizer. The use of fluorescent probes for *in situ* hybridization is sometimes referred to as FISH. Fluorescent DNA probe can also be prepared from template DNA by enzymatic methods (polymerases can incorporate dUTPs, dATPs, and dCTPs that have fluorescent dyes attached by linker arms to the nucleotide base). An exciting new development is the use of molecular genetic methods to fuse the gene for the green fluorescent protein (intrinsic fluorophore) to other target genes for subsequent expression in living cells.

3. Combination of fluorescence spectroscopy with flow and image cytometry is a powerful tool for diagnostic, prognostic and therapy control procedures in medicine as well as in the study of cellular aspects in immunology, cancer research, molecular biology, and biotechnology.

4. Multicomponent analysis using powerful fluorescence lifetime discrimination approaches.

5. As detection method in separation techniques, especially in thin-layer chromatography, high-performance liquid chromatography (HPLC) and electrophoresis. In HPLC applications for non-fluorescent compounds, either precolumn or postcolumn fluorescent derivatization is used. In capillary electrophoresis, the use of a laser to directly excite fluorescent compounds allows a extremely high sensitivity (femtomole to attomole levels of analytes can be detected in nanoliter volumes).

6. Automated batch and continuous flow analysis, widely used in clinical laboratories where high sample throughput is critical.

7. Studies of the microenvironment of fluorescent probe molecules. Many environmental factors modify the fluorescent properties of a molecule, e.g., solvent polarity, pH, proximity, and concentration of quenching species. The changes experienced by the fluorophore, sometimes subtle, can be used to obtain information about the specific region in which the fluorophore is localized.

8. Fluorescent sensors in solution (chemosensors) and fiberoptic sensors. Fluorescence is a particularly important technique in this field because of its high sensitivity of detection down to a single molecule, subnanometer spatial resolution with submicrometer visualization and submillisecond temporal resolution.

*See also:* **Chromatography:** Overview. **Fluorescence:** Instrumentation; High-Resolution Techniques; Time-Resolved Fluorescence; Derivatization; Fluorescence Labeling; Quantitative Analysis; Clinical and Drug Applications; Environmental Applications; Food Applications. **Sensors:** Photometric.



## Further Reading

Guilbault GG (ed.) (1990) *Practical Fluorescence*, 2nd edn, revised and expanded. New York: Dekker.

Lakowicz JR (ed.) (1999) *Principles of Fluorescence Spectroscopy*, 2nd edn. Dordrecht: Kluwer Academic.

Rettig W, Strehmel B, Schrader S, and Seifert H (1999) *Applied Fluorescence in Chemistry, Biology and Medicine*. Berlin: Springer.

Valeur B and Brochon J-C (eds.) (2001) *New Trends in Fluorescence Spectroscopy: Applications to Chemical and Life Sciences*. Springer Series on Fluorescence. Berlin: Springer.

## Instrumentation

**J N Miller**, Loughborough University of Technology, Loughborough, UK

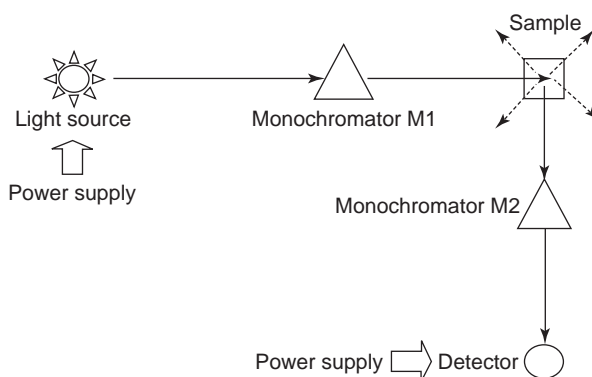
© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Fluorescence spectrometers capable of measuring excitation and emission spectra as well as fluorescence intensities have been commercially available for almost 50 years. While the functions and the overall layout of the main components have not changed very much over that period, the individual components have developed enormously in terms of sensitivity, stability, and longevity, with the result that modern instruments have capabilities far superior to those 50 years ago. Moreover, many accessories, which were either unobtainable or unusual in the 1950s are now routinely available, allowing the selectivity as well as the sensitivity of fluorescence methods to be more fully exploited. Another major change, characteristic of the last two decades, has been the universal use of personal computers to control the instruments as well as to record and manipulate the data. This rapid data processing capability has further enhanced the capabilities of fluorescence spectroscopy. This article summarizes the optical layout used in virtually all commercial instruments, and describes the main components used in modern spectrometers.

## Instrument Layout

The layout of the major fluorescence spectrometer components is shown in **Figure 1**. The characteristic feature of this arrangement is that any fluorescence emitted by the sample is detected at  $90^\circ$  to the incident light beam. This geometry, pioneered by Stokes in the 1850s, is still used in virtually all commercial and laboratory-built fluorescence instruments, inevitably with many additions, variations,



**Figure 1** Block diagram of the principal components of a fluorescence spectrometer.

and accessories for individual applications. Its crucial advantages are that photons transmitted by the sample (the majority of the incident photons when dilute solutions are studied) are not detected, and that Rayleigh scattered light has a minimum intensity in the  $90^\circ$  direction. The key optical elements of the layout are the light source, the excitation monochromator or filter (M1), the sample cell, the emission monochromator or filter (M2), and the detector. Separate power supplies are normally provided for the light source and the detector, and additional power may be needed for such items as a reference detector, sample stirrer, etc. This article deals only with steady-state fluorescence instruments: systems for studying lifetimes are summarized in a separate article.

## Light Sources

The fluorescence signal from a sample in dilute solution is (if photodecomposition effects are ignored) proportional to the incident light intensity. Thus the light source for fluorescence spectroscopy should have an intense and stable output. Other desirables

## Further Reading

Guilbault GG (ed.) (1990) *Practical Fluorescence*, 2nd edn, revised and expanded. New York: Dekker.

Lakowicz JR (ed.) (1999) *Principles of Fluorescence Spectroscopy*, 2nd edn. Dordrecht: Kluwer Academic.

Rettig W, Strehmel B, Schrader S, and Seifert H (1999) *Applied Fluorescence in Chemistry, Biology and Medicine*. Berlin: Springer.

Valeur B and Brochon J-C (eds.) (2001) *New Trends in Fluorescence Spectroscopy: Applications to Chemical and Life Sciences*. Springer Series on Fluorescence. Berlin: Springer.

## Instrumentation

**J N Miller**, Loughborough University of Technology, Loughborough, UK

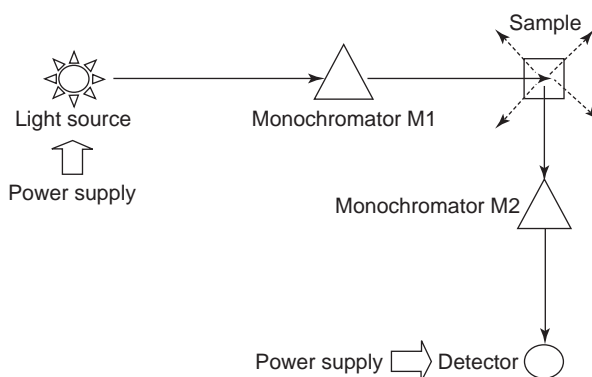
© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Fluorescence spectrometers capable of measuring excitation and emission spectra as well as fluorescence intensities have been commercially available for almost 50 years. While the functions and the overall layout of the main components have not changed very much over that period, the individual components have developed enormously in terms of sensitivity, stability, and longevity, with the result that modern instruments have capabilities far superior to those 50 years ago. Moreover, many accessories, which were either unobtainable or unusual in the 1950s are now routinely available, allowing the selectivity as well as the sensitivity of fluorescence methods to be more fully exploited. Another major change, characteristic of the last two decades, has been the universal use of personal computers to control the instruments as well as to record and manipulate the data. This rapid data processing capability has further enhanced the capabilities of fluorescence spectroscopy. This article summarizes the optical layout used in virtually all commercial instruments, and describes the main components used in modern spectrometers.

## Instrument Layout

The layout of the major fluorescence spectrometer components is shown in **Figure 1**. The characteristic feature of this arrangement is that any fluorescence emitted by the sample is detected at  $90^\circ$  to the incident light beam. This geometry, pioneered by Stokes in the 1850s, is still used in virtually all commercial and laboratory-built fluorescence instruments, inevitably with many additions, variations,



**Figure 1** Block diagram of the principal components of a fluorescence spectrometer.

and accessories for individual applications. Its crucial advantages are that photons transmitted by the sample (the majority of the incident photons when dilute solutions are studied) are not detected, and that Rayleigh scattered light has a minimum intensity in the  $90^\circ$  direction. The key optical elements of the layout are the light source, the excitation monochromator or filter (M1), the sample cell, the emission monochromator or filter (M2), and the detector. Separate power supplies are normally provided for the light source and the detector, and additional power may be needed for such items as a reference detector, sample stirrer, etc. This article deals only with steady-state fluorescence instruments: systems for studying lifetimes are summarized in a separate article.

## Light Sources

The fluorescence signal from a sample in dilute solution is (if photodecomposition effects are ignored) proportional to the incident light intensity. Thus the light source for fluorescence spectroscopy should have an intense and stable output. Other desirables



include the availability of a wide range of excitation wavelengths, long life, low cost, and the absence of hazards. In practice no single light source has all these properties, and several types of source thus find widespread use.

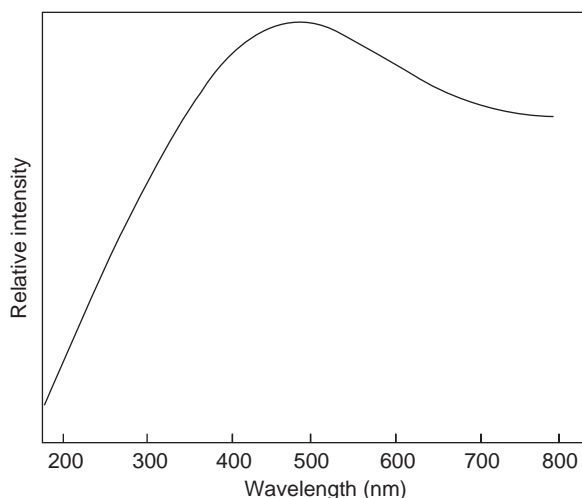
In instruments designed for routine laboratory work, two types of xenon lamps now predominate. Continuous wave (CW) xenon lamps, usually with a power consumption of 150–450 W, provide intense sources covering wavelengths from  $\sim 230$  nm into the near-infrared (**Figure 2**). Such lamps have a thick quartz or silica envelope and are filled with xenon gas at a pressure (at room temperature) of several bars. A high voltage ( $\sim 30$  kV) is used to strike an arc between two metal electrodes a few millimeters apart. Once the arc is established, a stable emission is maintained by running the lamp in low voltage, high current conditions: a 150 W lamp might operate at  $\sim 7.5$  A, 20 V. Clearly such lamps need special power supplies. Their high intensity and wide wavelength range explains their great popularity over  $\sim 50$  years. However, they have several disadvantages, including a lifetime of only a few hundred hours before the arc becomes unstable and impossible to focus on the excitation monochromator. Moreover, the lamps are costly and generate toxic ozone gas (removed by thermal decomposition or an adequate exhaust system). Their operating pressures of 70–100 bar demand the provision of sturdy steel housings for safety reasons. Very occasionally such lamps fail catastrophically while in use: a loud explosion then fills the inside of the spectrometer with tiny fragments of shattered silica and metal, a major problem.

In the last 20 years xenon flash-tubes have become popular alternative light sources. These tubes have a

similar spectral profile to their more formidable cousins (**Figure 2**) but are operated in a pulsed mode. The average power supplied is only a few watts but pulses of up to 10 kW are applied to the lamp at the line frequency (e.g., 50 or 60 Hz, according to the country of use). This causes very bright and short-lived ( $10\text{--}20\ \mu\text{s}$ ) flashes of light to be generated, the resulting pulses of fluorescence being collected, accumulated, and determined by a time-gated detector system (see below). These flash-tubes may have a rather longer life than the high power xenon lamps (though they are still costly,  $\sim$ US\$500), are safer, and generate very little ozone. Their main advantage is that the pulsed source, time-gated system can be used to distinguish short-lived light signals (fluorescence, scattered light) from long-lived ones (phosphorescence and other delayed emissions). In routine analyses, instruments using such sources have sensitivities similar to, or perhaps slightly poorer than, those using CW xenon lamps.

Conventional light sources sometimes used in simpler and older instruments include tungsten and other incandescent lamps, which are only applicable at excitation wavelengths longer than  $\sim 350$  nm, and mercury vapor lamps, which emit light almost solely at characteristic atomic wavelengths. The principal mercury lines are at 254, 313, 365, 405, 436, and 546 nm, depending upon the pressure in the lamp (higher pressure lamps give the longer wavelength lines). These sources are very stable and long-lasting and can be very intense, but they restrict the choice of excitation wavelengths and prohibit the collection of excitation spectra (see below). Some mercury lines are especially useful: many visible region fluorophores can be excited at 365 nm, and the 546 nm line is well matched to rhodamine fluorophores. Moreover, the lines can be used to calibrate monochromator accuracy.

The extremely bright light output of lasers seems to make them ideal sources for fluorescence, but their use has become commonplace only in recent years. This is partly due to concerns over cost and reliability, but other photophysical factors are also important. A single-wavelength laser is only useful if its emission is well matched to the fluorophore(s) under study; and a very intense light source may cause photodecomposition of the sample. The availability of very cheap, reliable, and stable diode lasers has made a big difference. Earlier sources of this type had emission lines in the red and near infrared regions ( $\geq 630$  nm) and their availability coincided with the growth of interest in long wavelength fluorescence, with its advantages of reduced scattered light interference, less background fluorescence from biological samples, and reduced photodecomposition. Diode



**Figure 2** Spectral output of a typical xenon arc lamp.

lasers with wavelengths in the blue and violet regions are now available, and their increased use is inevitable. Also valuable are the latest light-emitting diodes (LEDs), the so-called 'super(ultra)bright' LEDs. Many wavelengths are available, e.g., 375, 400, 420, 430, 450, 470, 505, 525, 570, 610, 630, and 660 nm, with bandwidths of 10–70 nm. A 375-nm LED with an emission bandwidth of only  $\sim 12$  nm is now available, and is marketed with fluorescence excitation specifically intended. Super-bright white LEDs covering a wide range of visible wavelengths (corresponding to a color temperature of  $\sim 5500$  K) are now also available. Most LEDs typically cost only a few dollars each, have lifetimes of many thousands of hours, and can be pulsed for time-gated applications. Such cheap, stable, and simple sources will certainly become more common, especially in compact and robust instruments for use in the field, in doctors' offices, etc.

### Excitation Monochromator or Filter (M1)

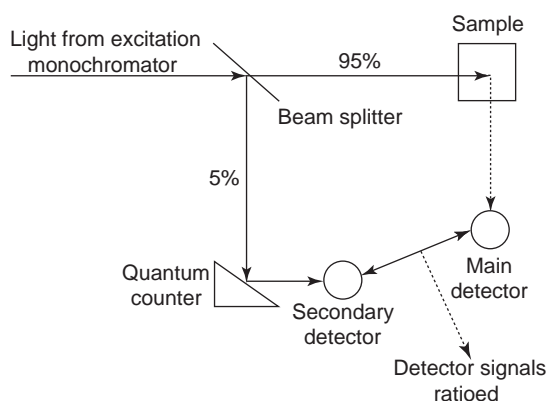
When a broadband light source (xenon lamp, incandescent lamp, pumped dye laser, or LED) is used in fluorescence measurements, it is necessary to restrict the bandwidth of excitation of the sample. This selection is an intrinsic part of the selectivity of the technique. Interference filters can be used to select suitable excitation wavelengths. They have bandwidths (measured between wavelengths at which their transmission efficiency is 50% of the maximum) of 10–20 nm. Mounted in holders or filter wheels, they allow the excitation wavelength to be changed easily according to the fluorophore studied. The maximum transmission of such filters in the visible region is as high as 80% (lower in the ultraviolet region), so they do not seriously compromise sensitivity. Equally they do not permit the study of excitation spectra, or of the best compromise between selectivity and sensitivity for a given fluorophore (see below). In modern instruments a grating monochromator (M1) is thus usually provided, consisting typically of an entrance slit on to which the light source is focused, a holographic reflection grating, and an exit slit. Several different monochromator designs are available, and are described in standard texts on optics and spectroscopy. In some cases a double monochromator is provided, with two gratings, etc., but the resultant reduction in stray light is offset by a loss of light throughput and of measurement sensitivity. A small motor allows M1 to be scanned at a variety of speeds: modern instruments provide speeds up to  $500 \text{ nm s}^{-1}$  to facilitate

'prescan' operations (see below) and allow rapid scans of samples in flow cells. Slit-widths are adjustable, in steps or continuously, to give control of the spectral pass-band of the monochromator. As the latter is decreased, selectivity of excitation of one fluorophore over another may improve, but the light throughput is lowered, so analytical sensitivity is lost. Conversely large slit-widths provide higher sensitivity at the cost of poorer selectivity and poorer rejection of scattered light.

When narrow line light sources are used, a scanning monochromator is largely redundant, so an interference filter will be sufficient, e.g., to isolate one mercury lamp emission line or one line from multi-line laser such as an argon-ion laser. Single-line lasers such as simple diode lasers need no filter or monochromator at all in the excitation beam.

### Reference Beam Systems

A striking feature of the optical layout in **Figure 1** is that a fluorescence spectrometer is essentially a single-beam instrument. Thus fluctuations in light source intensity cause corresponding changes in fluorescence signals. Moreover, the efficiency of the grating in the monochromator M1 will be wavelength dependent (optimum M1 efficiencies at  $\sim 300$  nm are common). So when the excitation spectrum of a sample is measured, by scanning M1 at a fixed emission wavelength, the spectrum obtained will reflect the source and grating properties as well as those of the sample. Such spectra are said to be 'uncorrected', and if routinely used they would complicate comparisons between the results obtained on different instruments. As a result, almost all spectrometers with an excitation monochromator now also include a reference beam system (**Figure 3**). The first essential component of such a unit is a beam-splitter placed between the exit slit of M1 and the sample, which in a wavelength independent manner diverts a small fraction ( $\sim 5\%$ ) of the excitation beam to a quantum counter. The latter contains a concentrated solution of a carefully selected long-wavelength dye, and emits a fluorescence signal that is proportional to the intensity of light falling on it, 'irrespective of the wavelength of the latter'. This emission from the quantum counter falls on to a secondary photodetector. The output of the latter is compared with that of the spectrometer's main detector (see below), thus allowing light source fluctuations to be cancelled out and excitation spectra to be 'corrected' so that results from different instruments should be comparable. (Accuracy and repeatability in a given monochromator should be  $\sim \pm 1$  nm.) These techniques are



**Figure 3** Components of the reference beam system in a spectrometer providing corrected excitation spectra.

described in more detail in texts on fluorescence standards (see Further Reading section).

## Sample Compartment

Light exiting the monochromator M1 is directed by the exit slit and/or a lens on to the sample in a suitable cell. In most cases a vertical slit image is focused on to the cell, but some instruments use a horizontal slit and claim improved sensitivity. Many instruments provide a shutter between M1 and the sample to protect the latter from heating and photodecomposition effects until the moment when measurements begin. In conventional solution measurements, samples are held in silica, glass, or disposable plastic cuvettes with four polished faces and pathlengths of 10 mm or less. Only two adjacent faces of a cuvette are used, but experiments have shown that giving the two remaining surfaces reflective coatings to increase the fluorescence signal is counterproductive, as the scattered light is also increased very greatly. Since fluorescence is emitted in all directions by the sample, only 5–10% of the emitted photons are in practice detected in a conventional instrument.

It is useful for the cell holder position to be adjustable so that the crucial portion of the sample that is illuminated by the light source and observed by the detector is offset from the center of the cuvette: this helps to minimize inner filter effects. Some commercial instruments provide this facility. A great range of flow cells for use in flow analysis and high-performance liquid chromatography detection is available: these generally have very short optical pathlengths, so inner filter effects are much less likely.

Since fluorescence intensities are temperature dependent by up to  $-3\%$  per degree Celsius, the provision of a thermostat system to provide control to within  $\pm 0.1^\circ\text{C}$  is desirable, and both electrical and water circulation systems are in use. Among the

range of other sampling and measurement accessories available, the most important are:

1. Polarization facilities. Fluorescence polarization measurements are now very common. Dedicated instruments may provide a fixed excitation polarizer with a vertical orientation, and an electrooptic emission polarizer with either vertical (i.e., parallel) or horizontal (crossed) orientations. Routine instruments provide polarizing filters allowing a greater range of orientation combinations. The light transmission properties of these filters are wavelength dependent, and this restricts the ranges over which they can be used.
2. Fiber-optic links to provide microtiter plate reading facilities. The plate-reader may fit into the sample compartment itself, or be a separate external unit with fiber-optic connections. The sensitivity of such devices is generally thought to be less good than that of dedicated plate-readers. It is worth adding that fiber-optic based fluorescence instruments using only simple excitation and emission filters have often been used for remote sensing, field-work, etc. Fibers may be made of silica, glass, or plastics, according to the wavelength range under study, and many different optical layouts have been described.
3. A magnetic stirrer motor mounted under the cuvette holder. Stirring may facilitate rapid mixing, e.g., in kinetic studies, but if the stirring is too vigorous, the resulting turbulence in the sample will increase scattered light interference. Also useful is a sample compartment door with an injection port for adding reagents to a cuvette.
4. Systems to allow low temperature fluorescence (or phosphorescence) studies, with liquid nitrogen or other coolants. In some cases a Dewar flask with transparent walls is provided, so that the coolant lies in the light beam and surrounds the sample contained in a narrow silica capillary. Alternatively the sample capillary is cooled by conduction, being mounted against a copper rod whose upper end is immersed in the coolant.
5. Sample holders for thin solid samples such as pieces of paper, thin layer chromatography (TLC) plates, etc. The ability to study such materials is a major advantage of fluorescence spectroscopy, but the optical arrangement used requires careful design. If a solid surface is placed in the sample compartment at a  $45^\circ$  angle to both excitation and emission beams, a great deal of incident light is liable to be reflected directly into the emission beam, giving a very large background signal. This can be avoided by off-setting the angle at which the surface is observed, a range of devices with adjustable optics and orientations being available.

6. Stopped flow accessories, allowing kinetic measurements in the millisecond time domain. These quite complex accessories provide the necessary injection syringes, a mixer, usually in a T-configuration and with a short dead-time, low volume flow cells, etc.

## Emission Monochromator or Filter (M2)

After leaving the sample compartment, fluorescence and unwanted scattered light signals enter the emission monochromator or filter (M2) before passing to the detector. The function of M2 is to isolate the signal from the desired fluorophore and minimize signals from Rayleigh and Raman scattering and background fluorescence. As in the case of M1, this is possible using either a grating monochromator or one or more filters. In commercial instruments a reflection grating monochromator is usual, with its optimum efficiency at 400–500 nm, and again it will have adjustable slits to control its pass-band and be motor-driven at a variety of speeds. Such instruments often also include a selection of cutoff filters in the emission beam to eliminate ‘second-order scattered light’ effects. If the excitation wavelength is 300 nm, light at that wavelength will be scattered by the sample and will pass through M2 when the latter is set to 600 nm, using the second-order properties of the M2 grating. This artifact is easily removed by inserting a filter into the emission beam that will absorb the 300 nm light, but pass light of higher wavelengths. In practice a series of filters is offered, according to the excitation wavelength used.

The fluorescence spectrum (emission spectrum) of a sample is obtained by scanning M2 while keeping the excitation wavelength constant. Since the efficiency of M2 is wavelength dependent, and the detector (see below) also has a sensitivity that varies with wavelength, such a fluorescence spectrum is normally ‘uncorrected’, i.e., it will depend upon the instrument as well as the sample. The correction of emission spectra is less easy to achieve than for excitation spectra, and is less often performed, so published emission spectra often vary from instrument to instrument. At least three methods are available for emission correction. A sound but tedious method is to calibrate the emission system (i.e., M2 plus the detector) with a standard light source of known emission profile. Such devices are available from NIST and other standards bodies. Comparison of the output of the fluorescence spectrometer with the certified output of the lamp then provides a correction factor at each wavelength, which can be applied to subsequent sample spectra. A related technique is to

use the output of the instrument’s own (CW) light source, as measured by the quantum counter (see above) as the standard source, reflecting it into the emission beam using a special sample compartment accessory. This method suffers from the disadvantages that it is difficult to achieve wavelength-independent reflection, and that any errors made in calibrating the excitation system are naturally carried into the emission correction process. Lastly it is possible to measure the fluorescence spectrum of a standard solute (e.g., quinine in perchloric acid) and compare the result with the published spectrum of the standard to give the correction factor at each wavelength. This method is only strictly valid if subsequent sample spectra are measured in the same optical conditions as the standard. Moreover, no single standard material can cover the wavelength range used in modern fluorescence applications. Spectral correction is covered in more detail in books given in the Further Reading section.

A fluorescence spectrometer with two monochromators with separate motor drives allows the measurement of several spectra in addition to the conventional excitation and fluorescence spectra previously described. The most important of these variants are synchronous spectra, obtained by scanning M1 and M2 simultaneously at the same scan rate and with a fixed wavelength interval,  $\Delta\lambda$ , between them (with M2 naturally at the higher wavelength). Compared with conventional spectra synchronous scans, with a judicious choice of  $\Delta\lambda$ , can be greatly narrowed and simplified, with a corresponding gain in selectivity. In principle a fluorophore mixture might be resolved using a single  $\Delta\lambda$  value, the synchronous spectral maxima for each component occurring at different wavelengths, or by using different values of  $\Delta\lambda$  for each component. Another method involves scanning the two monochromators at different rates, chosen so that the ‘frequency’ difference between them is constant. All these spectra, including conventional excitation and emission scans, can be regarded as sections through the excitation–emission matrix (EEM), also known as the total luminescence spectrum (TLS). This is a three-dimensional (3D) representation of the properties of a fluorescent sample, the  $x$ -,  $y$ -, and  $z$ -axes corresponding to excitation wavelength, emission wavelength, and intensity, respectively. In practice an EEM can be presented as a 3D picture, viewed from any angle, or by a mathematically equivalent contour plot. Both these facilities are commonly available on modern, PC-interfaced spectrometers.

In simple detectors, filters may replace the emission monochromator. Interference filters, with or without cutoff filters, are often satisfactory in



instruments dedicated to the measurement of a single fluorophore or a few known fluorophores. With laser light sources, holographic notch filters are very valuable, their sharp cutoff characteristics allowing measurements of fluorescence at wavelengths only a few nanometers above that of the laser. Another device sometimes used is a linear variable interference filter. In this device the spacing between its optical elements changes along its length. As it moves through the emission beam the wavelengths it transmits are continually changing, so it acts as a simple monochromator. Such devices can be used in the range 400–700 nm, their maximum transmission being typically ~40%.

## Detectors

Since only a small fraction of the fluorescence emitted by a sample is normally collected by the emission optics in a spectrometer, it is clearly essential for the instrument to have a sensitive and stable detector with a good response over the whole wavelength range provided by the monochromators. In conventional instruments, the detector is almost always a photomultiplier tube (PMT). These devices are ubiquitous in many areas of spectrometry and so will only be described briefly. They consist of an evacuated quartz or glass envelope containing a photocathode. When a photon falls on this cathode it emits an electron. This electron is focused on to a series of electron multipliers (dynodes) held at increasing potential differences, causing an amplifying shower of electrons, which eventually arrive at the anode and yield a measurable current. Such a detector requires its own power supply to provide the high potentials (several hundred volts) applied. A conventional PMT has its maximum response at ~450 nm, with a quantum efficiency of ~25%; the sensitivity usually falls off sharply above ~550 nm. 'Red-sensitive' PMTs giving good responses of 850–900 nm are now often used: they may have reduced sensitivities in the visible spectrum.

As noted above, instruments using pulsed light sources use a gated detector system to distinguish rapid from delayed light signals. The detector is synchronized to the source pulses and the user can control the delay time,  $t_d$ , between the start of the pulse and the opening of the detector gate, and the gate time,  $t_g$ , for which the detector is on in the period before the next pulse. Manipulation of these times allows the lifetimes of delayed fluorescence and phosphorescence to be estimated.

Miniature PMTs have recently become popular for use in compact or portable instruments. They can be

driven from relatively low voltage supplies. In routine work PMTs are normally used in DC mode, the total anode current they produce being amplified and recorded. But low light levels can be recorded using photon-counting PMT systems, in which each secondary electron pulse arriving at the anode in response to a single incoming photon is recorded on a high-speed counter. Such systems inevitably fail when the photon flux generates cathode electron pulses more quickly than the counter can operate.

Two types of solid-state detector have found many applications in fluorescence detection. Charge coupled device (CCD) arrays are well established in other areas of spectrometry, and will not be described in detail here. They allow the detection of a complete emission spectrum in a few milliseconds, the fluorescence monochromator output being dispersed across the array elements simultaneously. Arrays that plug directly into PC-interface cards are now available. A major advantage of these detectors is that their optimum sensitivity is at longer wavelengths (700–800 nm) than that of most PMTs, so they are well suited to near infrared fluorescence detection. Avalanche photodiodes (APDs) also have good long wavelength sensitivity and have been used successfully, especially in compact or portable detectors. They can be regarded as developments of simple silicon photodiodes, with semiconductor layers and their applied potentials arranged to produce many electrons and holes from each incident photon. Gains of several hundred are feasible with silicon APDs operating between 400 and 1100 nm, and are compact detectors, hard to overload.

## Instrument Control and Data Handling

Most modern fluorescence spectrometers are entirely PC-controlled, the case holding the optical components having only an on-off switch. The interfaced PC, equipped with suitable software, directs all routine operations of the spectrometer, including the type of spectrum to be scanned, scan rate, slit-widths, the delay, and gate times used with pulsed sources, the insertion of polarizing or other filters into the light beams, etc. Signals from the instrument are received and stored by the computer and can be further processed using smoothing or averaging methods, the calculation of derivative spectra, the presentation of EEM spectra, the analysis of mixtures by chemometric resolution of their overlapping spectra, and so on. Other facilities include the presentation of polarization spectra (the variation of fluorescence polarization with wavelength); recording fluorescence as a function of time, for example in kinetic assays; the

determination of the correction factors needed for corrected spectra; and the useful 'prescan' function, which provides rapid scans of both monochromators to suggest excitation and emission wavelengths for brand new samples. As in most areas of spectroscopy, the use of a PC-controlled instrument is now obligatory for all but the most routine analyses, and as PCs grow in speed and power, their value increases correspondingly.

*See also:* **Fluorescence:** Time Resolved Fluorescence; Quantitative Analysis.

## Further Reading

- Ingle JD and Crouch SR (1988) *Spectrochemical Analysis*. Englewood Cliffs, NJ: Prentice-Hall International.
- Lakowicz JR (ed.) (1999) *Principles of Fluorescence Spectroscopy*, 2nd edn. Dordrecht: Kluwer Academic Publishers.
- Miller JN (ed.) (1981) *Standards in Fluorescence Spectroscopy*. London: Chapman and Hall.
- Olson ED (1975) *Modern Optical Methods of Analysis*. New York: McGraw Hill.
- Valeur B (2002) *Molecular Fluorescence – An Introduction: Principles and Applications*. New York: Wiley-VCH.

## Multidimensional Fluorescence Spectrometry

**T T Ndou**, Gillette Research Institute, Gaithersburg, MD, USA

**I M Warner**, Louisiana State University, Baton Rouge, LA, USA

**J B Zung**, Pfizer Central Research, Groton, CT, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1395–1401, © 1995, Elsevier Ltd.

## Introduction

Fluorescence spectroscopy has been established as an analytical technique for both qualitative and quantitative analysis because of its inherent sensitivity, selectivity, and versatility. The selectivity of fluorescence measurements can be attributed to the multiparametric properties of the measurements. This is because even the simplest of luminescence measurements will involve the use of two parameters: an excitation wavelength and an emission wavelength. In a conventional fluorescence measurement, the excitation wavelength is set at a particular value and the emission wavelengths are scanned. Various analytical methods that employ conventional fluorescence measurements exist for the detection and quantification of single-component analytes. These analytes can range from environmental pollutants to pharmaceutical compounds, from marine algae to biological fluids. The fluorescence-based methods usually have impressive limits of quantification. However, these measurements tend not to be highly selective for multicomponent samples (two or more components) since conventional fluorescence measurement usually involves the use of a single excitation and emission wavelength setting in order to achieve the best signal response.

In the case of a multicomponent sample (real-world samples), the resultant fluorescence emission spectrum is often a combination of those of the components in the mixture that are excitable at the chosen excitation wavelength. This can result in broad and severely overlapping emission spectra. Qualitative and quantitative information is all but lost from such spectra. Furthermore, a single-wavelength scan does not offer the possibility of suppressing the fluorescence contributions of some compounds while analyzing only selected components in the mixture.

It is apparent from the scenario described above that conventional fluorescence methods are of limited applicability for the analysis of multicomponent mixtures. Therefore, for routine laboratory analysis of a multicomponent mixture, the most important prerequisite is to develop a simple and easy-to-use qualitative and quantitative fluorescence-based method that is not limited to single-component samples. Additionally, the fluorescence technique adopted should not sacrifice the sensitivity and selectivity of conventional fluorescence measurements and should minimize the problem of overlap in emission spectra of complex multicomponent samples.

To overcome the limitations of single-wavelength measurements in evaluating multicomponent samples, multidimensional luminescence measurements can be used. The multidimensional approach allows for a better estimation of minor differences between the individual components of the mixture and can ultimately lead to well-resolved spectra of multicomponent mixtures. Of significant importance is the fact that multidimensional luminescence measurement can be a viable technique for detecting minute and distinct spectral perturbations from small analytical samples.

As far as the potential application of multidimensional fluorescence measurements is concerned, the



determination of the correction factors needed for corrected spectra; and the useful 'prescan' function, which provides rapid scans of both monochromators to suggest excitation and emission wavelengths for brand new samples. As in most areas of spectroscopy, the use of a PC-controlled instrument is now obligatory for all but the most routine analyses, and as PCs grow in speed and power, their value increases correspondingly.

*See also:* **Fluorescence:** Time Resolved Fluorescence; Quantitative Analysis.

## Further Reading

- Ingle JD and Crouch SR (1988) *Spectrochemical Analysis*. Englewood Cliffs, NJ: Prentice-Hall International.
- Lakowicz JR (ed.) (1999) *Principles of Fluorescence Spectroscopy*, 2nd edn. Dordrecht: Kluwer Academic Publishers.
- Miller JN (ed.) (1981) *Standards in Fluorescence Spectroscopy*. London: Chapman and Hall.
- Oslen ED (1975) *Modern Optical Methods of Analysis*. New York: McGraw Hill.
- Valeur B (2002) *Molecular Fluorescence – An Introduction: Principles and Applications*. New York: Wiley-VCH.

## Multidimensional Fluorescence Spectrometry

**T T Ndou**, Gillette Research Institute, Gaithersburg, MD, USA

**I M Warner**, Louisiana State University, Baton Rouge, LA, USA

**J B Zung**, Pfizer Central Research, Groton, CT, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1395–1401, © 1995, Elsevier Ltd.

## Introduction

Fluorescence spectroscopy has been established as an analytical technique for both qualitative and quantitative analysis because of its inherent sensitivity, selectivity, and versatility. The selectivity of fluorescence measurements can be attributed to the multiparametric properties of the measurements. This is because even the simplest of luminescence measurements will involve the use of two parameters: an excitation wavelength and an emission wavelength. In a conventional fluorescence measurement, the excitation wavelength is set at a particular value and the emission wavelengths are scanned. Various analytical methods that employ conventional fluorescence measurements exist for the detection and quantification of single-component analytes. These analytes can range from environmental pollutants to pharmaceutical compounds, from marine algae to biological fluids. The fluorescence-based methods usually have impressive limits of quantification. However, these measurements tend not to be highly selective for multicomponent samples (two or more components) since conventional fluorescence measurement usually involves the use of a single excitation and emission wavelength setting in order to achieve the best signal response.

In the case of a multicomponent sample (real-world samples), the resultant fluorescence emission spectrum is often a combination of those of the components in the mixture that are excitable at the chosen excitation wavelength. This can result in broad and severely overlapping emission spectra. Qualitative and quantitative information is all but lost from such spectra. Furthermore, a single-wavelength scan does not offer the possibility of suppressing the fluorescence contributions of some compounds while analyzing only selected components in the mixture.

It is apparent from the scenario described above that conventional fluorescence methods are of limited applicability for the analysis of multicomponent mixtures. Therefore, for routine laboratory analysis of a multicomponent mixture, the most important prerequisite is to develop a simple and easy-to-use qualitative and quantitative fluorescence-based method that is not limited to single-component samples. Additionally, the fluorescence technique adopted should not sacrifice the sensitivity and selectivity of conventional fluorescence measurements and should minimize the problem of overlap in emission spectra of complex multicomponent samples.

To overcome the limitations of single-wavelength measurements in evaluating multicomponent samples, multidimensional luminescence measurements can be used. The multidimensional approach allows for a better estimation of minor differences between the individual components of the mixture and can ultimately lead to well-resolved spectra of multicomponent mixtures. Of significant importance is the fact that multidimensional luminescence measurement can be a viable technique for detecting minute and distinct spectral perturbations from small analytical samples.

As far as the potential application of multidimensional fluorescence measurements is concerned, the

method is highly selective and gives enhanced information about the components of a complex sample. Multidimensional fluorescence measurements have found wide application for the identification and quantification of polycyclic aromatic hydrocarbons (PAHs) in environmental samples, the fingerprinting and qualitative identification of fluorescent PAHs contained in oil-spill samples, the characterization and mapping of motor oil and gasoline samples in forensic studies, the characterization and analysis of pharmaceutical compounds, the resolution of complex clinical samples, bacterial fingerprinting and classification, and for the spectral fingerprinting of marine phytoplankton samples.

This technique also allows for the resolution of spectrally similar components and severely overlapping components in a multicomponent mixture. In addition, multidimensional fluorescence measurements provide significant spectral information for distinguishing between closely related samples. The variations in the excitation and emission profiles add to a full understanding of the minimum number of fluorescing components in the sample matrix.

A detailed account of progress in the applications of multidimensional fluorescence analysis in analytical chemistry can be found elsewhere. The purpose of this article is to outline the scope and potential practical applications of multidimensional fluorescence in analytical chemistry and other related research areas.

## Instrumentation

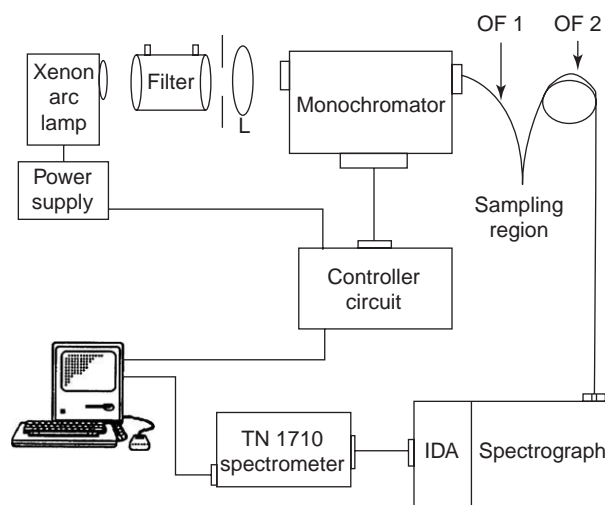
Instrumentation capable of acquiring conventional emission spectra can also be used to acquire data in a multidimensional format. However, the use of conventional fluorescence spectrophotometers can be very time consuming owing to the type of detector (photomultiplier) and wavelength selector (monochromator) that are often used. To reduce the time necessary to acquire the spectra and increase the efficiency of data collection, parallel detectors may be used. Parallel detection allows for the simultaneous acquisition of spectral data over a range of wavelengths (e.g., 200–700 nm). Parallel detection can significantly reduce data acquisition time and can result in an increase in the signal-to-noise ratio owing to the advantage of multichannel operation.

Instrumentation capable of acquiring data in a multidimensional format using parallel detectors is similar to that of conventional spectrofluorimeters. The major differences are the type of detector, light source, and wavelength selectors used. In a conventional fluorimeter, a photomultiplier tube is used as

the photon detector. Although the photomultiplier tube is a very sensitive detector, it is a single-channel detector and lacks the ability to monitor multiple wavelengths simultaneously. Thus, this type of detector is not suited for rapid acquisition of data over a range of wavelengths. Instead of using a photomultiplier tube as the photon detector, multichannel detectors that are capable of acquiring a range of wavelengths in a snapshot are used in multidimensional instrumentation. Multichannel detectors that have been used successfully for the collection of fluorescence spectra include the vidicon, linear photodiode array (PDA), silicon intensified target vidicon, and charge-coupled device (CCD).

Each of these detectors has its own unique advantages and disadvantages. The first generation of multidimensional fluorescence instruments ('videofluorimeters') used the vidicon for collecting the total luminescence spectrum. The vidicon, however, suffers from poor sensitivity and detector lag and blooming. To overcome this, the second generation of multidimensional instruments used linear PDAs. A schematic of a multidimensional fiber-optic instrument that incorporates a PDA for detection is shown in Figure 1. Commercial instruments now offer CCDs as detectors.

It is also important that the light source in a multidimensional instrument be able to provide relatively continuous output over a range of wavelengths, since multiple excitation wavelengths are needed in the multidimensional measurement. In order to use multiple excitation wavelengths, a white light source, such as a high-pressure xenon arc lamp,



**Figure 1** Schematic of a fiber-optic-based multichannel fluorometer: IDA = 512-element intensified linear photodiode array detector; L is the lens; and OF 1 and OF 2 are the excitation and emission fibers, respectively.

is used. The xenon arc provides intense and relatively continuous output throughout the ultraviolet and visible regions (200–1000 nm).

The new computerized spectrofluorimeters in combination with the availability of selective multi-channel detectors allow for a reliable and quick collection and analysis of spectral information that was difficult to accomplish using a single-wavelength scan. The new multichannel detectors used for making multiwavelength measurements also allow the ability to study a wide variety of multicomponent samples.

## Spectral Presentation

In multidimensional fluorescence measurement, the measured luminescence,  $I_L$ , is a function of the excitation ( $\lambda_{ex}$ ) and emission wavelengths ( $\lambda_{em}$ ), i.e.,

$$I_L = f(\lambda_{ex}, \lambda_{em}) \quad [1]$$

By varying either or both of the two wavelengths, the measurement of an individual fluorophore or complex mixture can be achieved.

It is often convenient to format multidimensional fluorescence data in the form of a matrix of intensities called the excitation–emission matrix (EEM). The EEM can be represented as the vector product of the excitation and emission spectra. Each EEM has the following characteristics:

1. Every row represents an emission spectrum excited at a specific excitation wavelength.
2. Each column represents the excitation spectrum at a specific emission wavelength.
3. Scattered light appears as a diagonal line across each EEM.

Data acquired in the form of an EEM or total luminescence spectrum can be conveniently presented either as an isometric projection or in the form of a contour map. When displayed as isometric projections, the EEM provides a three-dimensional intensity profile of the sample over the range of excitation and emission wavelengths scanned. One axis represents the emission wavelengths, while the second axis represents the excitation wavelengths. The third axis is the intensity. The diagonal ridge displayed across the EEM spectrum is due to first-order light scattering. A drawback of displaying the data in the form of isometric projections is that small peaks adjacent to more intense peaks tend to be partially or completely obscured. This prevents accurate viewing of the EEM.

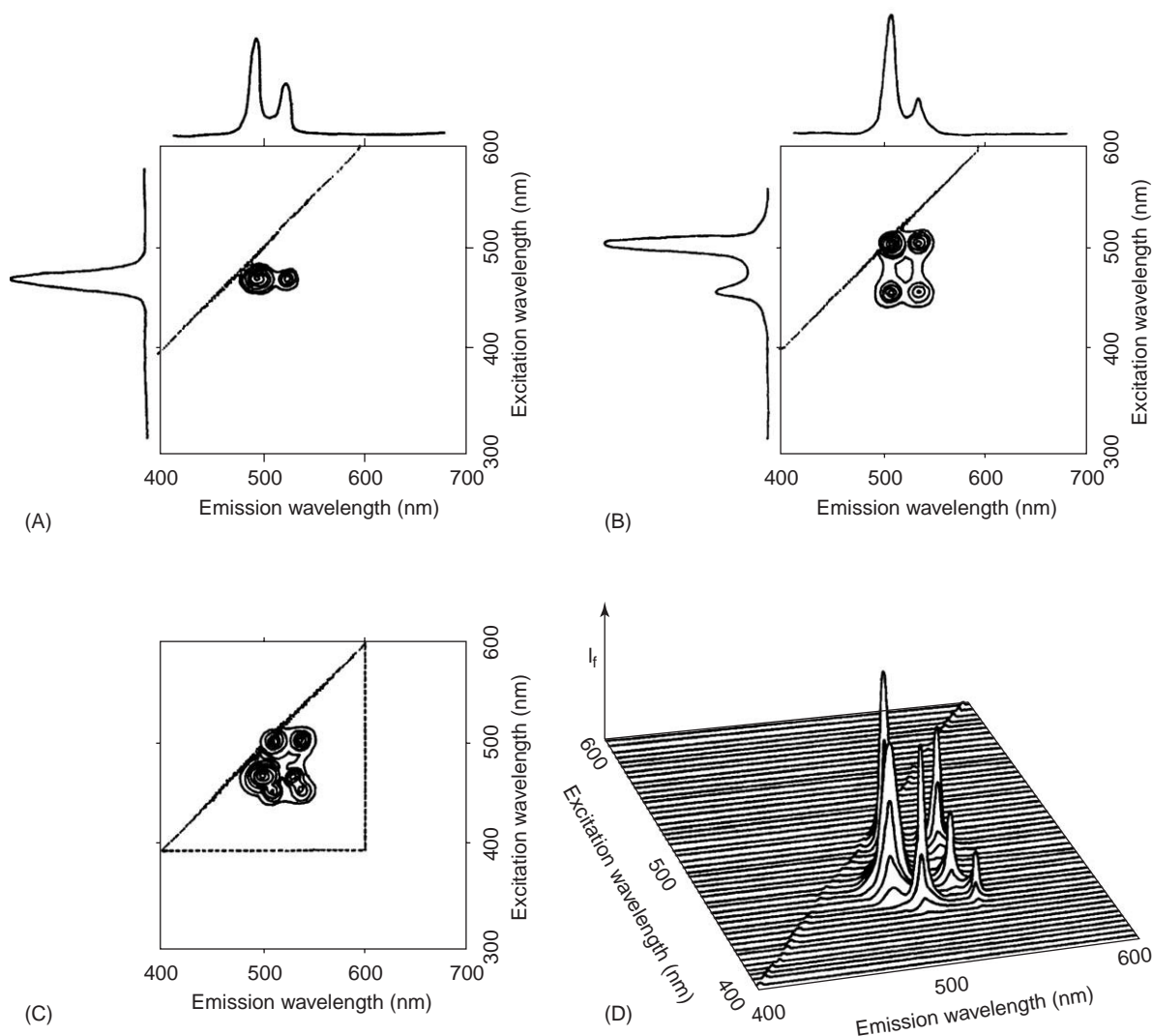
To overcome display problems, the data matrix can be represented in a more informative way, such as a contour map. In the contour plot presentation, the spectral information is plotted as a series of iso-emissive concentric lines in the excitation and emission wavelength plane. Often this technique permits the detection of possible peak overlap and allows for a better view of the spectral information. The primary advantage of the two-dimensional contour plot is that it successfully eliminates the problem of obscured peaks and thereby increases the possibility of identifying individual components in the mixture. Figure 2 shows hypothetical EEMs illustrating both the isometric projection and contour plots. The advantage of viewing the data matrix in a contour plot is illustrated.

## Spectral Data Processing

This section provides a brief overview of methods available for the processing of multidimensional data. Further evidence relating to multidimensional applications is in the use of rapid scanning modern instrumentation for monitoring time-dependent phenomena such as chemical equilibria, reaction kinetics, flow injection analysis, and enzymatic processes. These techniques allow the use of different experimental strategies to capture and store spectral data, thereby facilitating numerous possibilities for data processing and graphical representation. This in turn permits the use of various techniques to deconvolute overlapping spectra.

The acquisition of a single-wavelength fluorescence spectrum of a complex multicomponent mixture is more tedious for many analytical applications that involve the measurement of some essential intermediate components or transient species. The determinations of such systems become more critical if the spectral information for those species is important to the understanding and elucidation of the properties of the final product. Similarly, this is true for chemical measurements of components whose fluorescence intensity decreases continuously with time. A rapid multidimensional data-processing technique becomes exceedingly important in the analysis of these species in a chemical reaction.

The wavelength and time resolution of the EEMs of a multicomponent mixture are particularly important for the deconvolution of the mixture into individual constituents. A simple approach involves application of a ratio deconvolution method after obtaining a series of EEMs for the mixture of components by altering the relative intensities of components in each EEM. The EEM containing a mixture



**Figure 2** Generation of fingerprint excitation–emission matrix: (A) EEM of pure component, compound A; (B) EEM of pure component, compound B; (C) fingerprint EEM of a mixture of compounds A and B; (D) isometric projection of fingerprint in (C). (Reprinted with permission from Shelly DC, Quarles JM, and Warner IM (1980) Identification of fluorescent *Pseudomonas* species. *Clinical Chemistry* 26(8): 1127–1132.)

of components of different lifetimes will provide the characteristics necessary for ratio deconvolution. This will in turn allow for the necessary criterion to isolate individual components whose original spectra contained severely overlapping emission bands. The combination of multidimensional fluorescence measurements and ratio deconvolution is used in an attempt to accomplish satisfactory qualitative separation of mixtures containing components that have strong spectral overlap.

The emergence of new rapid scanning analytical instruments requires the processing of data with high dimensionality. This requires the development of multidimensional data filtering techniques and other modern processing methods. In many cases, the

interpretation of multicomponent chemical data is based on spectral matching and fingerprinting. The fingerprinting technique is employed for rapid qualitative characterization and identification of the properties and origin of the samples.

An added feature of total fluorescence techniques is found in systems whose emission and/or excitation wavelength maxima change with the formation of a complex. Multidimensional fluorescence analysis allows the analyst to detect the differences in sensitivity arising from changes in excitation and emission maxima of the analyte. In addition, the technique finds wide application in examining the corrected fluorescence spectra of clinical compounds and their derivatives. In studies involving biological



fluids, the sensitivity of the method is strongly influenced by changes in the excitation or emission peaks of the drugs when bound to target cells.

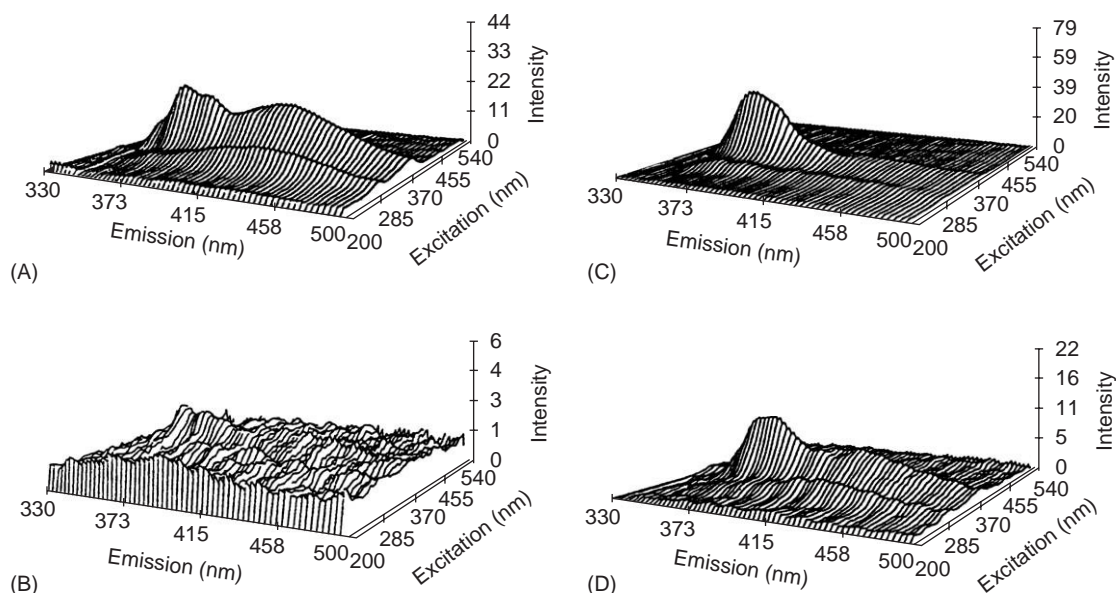
## Applications in Environmental Analysis

Multidimensional fluorescence analysis is well suited for routine qualitative separation and identification of complex mixtures of PAHs in environmental samples. The separation and extraction of the individual component spectra of PAH mixtures affords valuable information about these hazardous compounds. Such information may allow for a detailed study of the means to remove their influence from the environment. The differences in the intensity contributions from a mixture of PAHs in a multicomponent system could be related to the differences in concentrations of the individual components and their respective quantum efficiencies. The problem is exacerbated by the fact that most solutions containing mixtures of PAHs display broad fluorescence spectral bands at room temperature, resulting in significant overlap and limited specificity for multicomponent analysis. This makes the success of multidimensional fluorescence measurements for complete spectral resolution of a complex multicomponent mixture dependent upon the degree of spectral overlap among the individual components and their fluorescence

emission spectra. In contrast, however, the problem of spectral resolution allows for a simple and convenient means for the incorporation of other useful chemical measurements in fluorimetric analysis.

The combination of multidimensional fluorescence analysis and quenching studies has allowed component resolution of PAHs complexed with  $\beta$ -cyclodextrin ( $\beta$ -CD). For example, upon the addition of iodide, the fluorescence intensity of fluanthrene and pyrene in aqueous solution is quenched to less than 10% of its original intensity. The quenching is dramatically reduced upon addition of  $\beta$ -cyclodextrin to a 1% *t*-butanol solution containing pyrene–fluanthrene. **Figure 3** illustrates the utility of a multidimensional fluorescence scheme in combination with cyclodextrin complexation to achieve resolution of a two-component mixture of pyrene and fluanthrene.

Many petroleum-based samples contain complex mixtures of PAHs, which are known or suspected carcinogens and mutagens. Multidimensional fluorescence analysis is suitable for the identification of fluorophores contained in oil samples. The basis of this technique is to employ a unique spectral fingerprint for the selective characterization and mapping of certain PAHs in the oil. Such a procedure is likely to provide a detection scheme for selective analysis of environmental samples for analytically important hazardous compounds. The primary goal will be to expand this technique to create a library containing



**Figure 3** Isometric plots of fluorescence showing the EEMs of (A) a pyrene–fluanthrene mixture in 1% *t*-butyl alcohol; (B) a pyrene–fluanthrene mixture in 1% *t*-butyl alcohol and 100 mmol L<sup>-1</sup> iodide; (C) a pyrene–fluanthrene mixture in 1% *t*-butyl alcohol, 100 mmol L<sup>-1</sup> iodide, and 1.5 mmol L<sup>-1</sup>  $\beta$ -CD; (D) a pyrene–fluanthrene mixture in 1% *t*-butyl alcohol, 100 mmol L<sup>-1</sup> iodide and 3.8 mmol L<sup>-1</sup>  $\gamma$ -CD. (Reprinted with permission from Nelson G, Neal SL, and Warner IM (1986) Resolution of mixtures by cyclodextrin complexation and multidimensional data analysis. *Spectroscopy* 3: 24–28.)

spectral fingerprints with specific characteristics unique to each individual oil sample. The added feature of this technique is that the identification of individual components is dependent on the relative concentration of each PAH in the oil sample. The identification of important emission bands is successfully accomplished by the use of contour plots. Thus, in areas of environmental pollution, fingerprinting techniques should be recognized as a screening method for PAH analysis.

The utility of multidimensional fluorescence is demonstrated using EEMs for comparative analysis in the determination of compounds that share a common source. This technique is effectively applied in characterization and extraction of important information from gasoline and motor oil samples. Multidimensional fluorescence of gasoline samples has definite potential importance in forensic applications. For example, it is suggested that this technique could provide the analyst with tentative evidence in a criminal or civil case by matching the fluorescence emission spectra from a suspect's car to that from the scene of a crime. In addition to forensic motor oil and gasoline identification, fingerprinting techniques have a greater application for determination of the effects of oil spills on marine life and the environment. The technique shows great potential in that it is precise and convenient for forensic and environmental analyses.

Another feature of multidimensional fluorescence is that it provides a good potential for a better spatial resolution, especially when using CCD detectors. The CCD detector is able to improve the detection of a fluorescent mixture of PAHs and provides a large dynamic range and reproducibility of spectral properties of the sample.

## Applications in Phytoplankton Analysis

The selectivity and specificity of multidimensional fluorescence measurements can be beneficial for the characterization and spectral fingerprinting of different classifications of marine phytoplankton or algae. Different and characteristic pigments in the algae produce spectral differences that may be used for the classification of different types of algae. Multidimensional fluorescence measurements are well suited for the spectral fingerprinting of phytoplankton since the entire visible region of the spectrum can be analyzed simultaneously. The individual pigments found in the algae are classified as either a type of chlorophyll, carotenoid, or phycobilin. The distribution of these pigments in the algae provides a unique distribution

between greens, blue-greens, and those that contain a carotenoid complex. By visually examining the EEMs of the different algae, it is possible to differentiate between species within a class of algae. The fingerprinting of algae has been accomplished using a portable multichannel fluorescence spectrometer (PMFS). The PMFS shows enhanced capabilities for rapidly acquiring and fingerprinting fluorescence emission spectra of algae over conventional fluorimeters. Multidimensional fluorescence analysis has the potential for indirect monitoring of the presence of pollutants in marine environments using algae as a bioindicator. This technique is especially suited for rapid measurement of algal fluorescence.

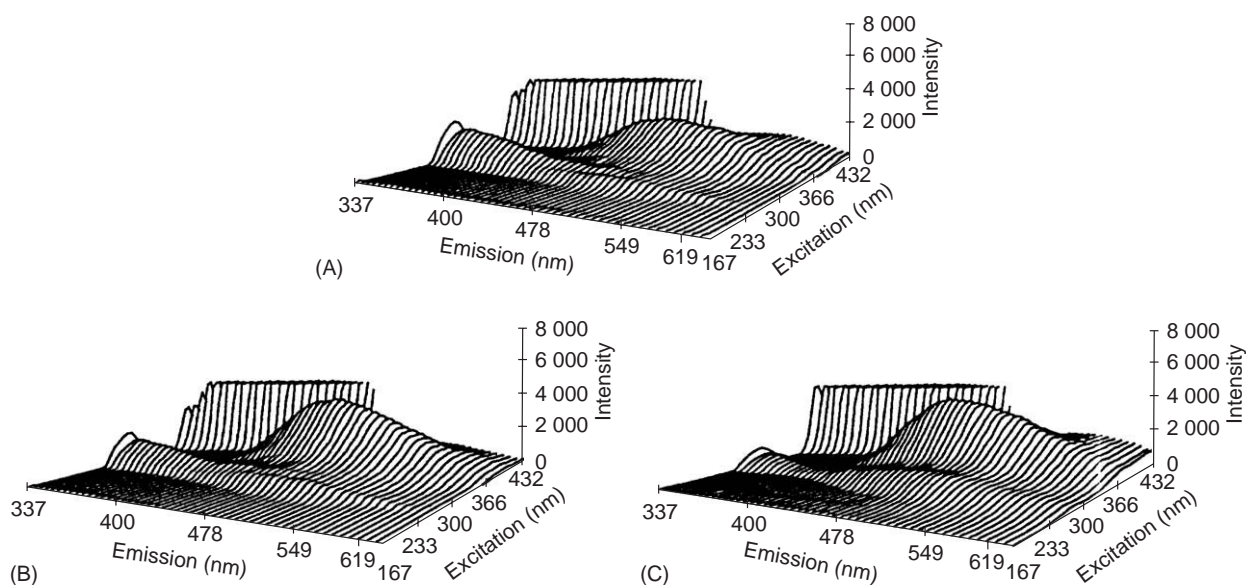
The versatility of multidimensional fluorescence techniques can be further improved by incorporating optical fibers into the instruments. Optical fibers offer the advantages of geometrical flexibility, environmental versatility, and small size and enable online monitoring of systems or detection of species in awkward or remote regions. In this approach, the optical fiber is placed directly in the sampling region, which precludes the need for sampling. The fluorescence is collected through the fiber and transferred to the detector via the optical fiber. Multidimensional fluorescence measurements for the analysis of marine phytoplankton in remote locations using fiber-optic-based sensing have been described previously. The presence of different characteristic pigments allows a remarkably simple classification of algae.

## Applications in Biological Fluid Analysis

Multidimensional fluorescence measurement is also very useful for applications that involve the separation and structural elucidation of potential therapeutic drugs. The capability for rapid acquisition of spectral data and enhanced detection limits achieved by use of fluorescence measurements is an important advantage for multidimensional fluorescence measurements. This detection is important for the screening of unknown compounds in complex mixtures, such as in the area of Pharmaceuticals. The utility of a video fluorimeter in this regard is demonstrated for the analysis of unused growth medium, typical cancer cell, and typical normal cell samples (Figure 4).

Investigation of complex biological samples, such as the analysis of bovine serum albumin, could be best achieved by employing the high resolving power of multidimensional fluorescence spectroscopy. An isometric projection will allow for a simultaneous observation of all the data points in each of the scans. In addition to a more pictorial representation of the





**Figure 4** Isometric projections of EEMs: (A) unused growth medium; (B) supernatant from noncancerous cell medium; (C) supernatant from cancerous cell medium. (Reprinted with permission from Rossi TM, Quarles JM, and Warner IM (1982) Analytical studies of fluorescent metabolites of cancer cells. *Analytical Letters* 15(B13): 1083–1098.)

experimental data, multidimensional fluorescence measurements are efficient, versatile, and provide enhanced information. Furthermore, in this field of biological and biochemical applications, multidimensional fluorescence accelerates the collection of important information on the spectral differentiation of the complexes formed when therapeutic drugs interact with their target cells on tissues in the body. This technique shows much improvement in the ability to quantify and check the purity of drugs and to select the optimal conditions for their analysis.

This technique also provides a better application of pattern recognition. Pattern recognition techniques have been used for the analysis of mixtures of fluorescent compounds with overlapping bands. The fluorescence contours of human sera have been explored as a new variant of pattern recognition methods in clinical chemistry. The rationale for this type of study is based on the fact that biomolecules are highly fluorescent when strongly bound to protein as compared to when free in solution. It is apparent that such an approach will provide the ability to achieve low limits of detection, which are critical, especially for pH-sensitive biological studies where pH change could have a profound influence on the resulting contour map. The limitation of pattern recognition is that the technique does not lead to analytically quantitative measurement of each individual component in the sample mixture. Thus, pattern recognition does not provide complete fluorescence spectral information about all of the individual components contributing to the total fluorescence of the multicomponent mixture.

It is clear that increased dimensionality of measurement effectively improves the resolution of overlapping fluorescence emission spectra. However, it should be emphasized that multidimensional fluorescence alone does not always result in complete resolution of spectrally overlapping fluorescence emission bands. In order to be able to successfully extract the most important information from the spectral data, the analyst should be able to observe and understand real changes as opposed to artifacts in the sample. These types of spectral changes are often clearly shown in the graphical or contour display of three-dimensional fluorescence spectra.

## General Applications

The selectivity of multidimensional fluorescence measurements, coupled with excellent detection limits provided by fluorescence techniques, will allow for fruitful investigations of a wide range of multicomponent mixtures. The information obtained from these types of measurements are valuable for the identification of individual components in complex mixtures and in the investigation and formulation of equilibrium binding models. It should be noted that in biological fluid systems, selective detection is extremely important in the final formulation of the equilibrium binding model and in the identification of the individual analyte.

The advent of multidimensional fluorescence measurements is also useful in the screening of inclusion complexes. Thus, the multidimensional

approach need not be limited to the identification and quantification of almost every analyte in the sample mixture. For example, in the case of an oil spill it is significant to examine the entire sample to produce a fingerprint for use in the identification of patterns relating to its classification. The goal of multidimensional fluorescence measurements is to enhance selectivity. Furthermore, the resolution and accuracy of this technique improve the sensitivity of the method. Despite severe spectral overlap, the results always indicate that multidimensional analysis is superior to conventional fluorimetric analysis.

**See also:** **Fluorescence:** Instrumentation; Derivatization; Clinical and Drug Applications; Environmental Applications. **Fuels:** Oil-Based. **Polycyclic Aromatic Hydrocarbons:** Determination; Environmental Applications.

## Further Reading

- Alexander J, Mashak G, Kapitan N, and Siegel AJ (1987) Fluorescence of petroleum products II: Three-dimensional fluorescence plots of gasoline. *Journal of Forensic Sciences* 32: 72–86.
- Bentz AP (1976) Oil spill identification. *Analytical Chemistry* 48: 454A–472A.
- Christian GD, Callis JB, and Davidson ER (1981) Array detectors and excitation emission matrices in multicomponent analysis. In: Wehry EL (ed.) *Modern Fluorescence Spectroscopy*. New York: Plenum.
- Lakowicz JR (1983) *Principles of Fluorescence Spectroscopy*. New York: Plenum.
- Ndou TT and Warner IM (1991) Multidimensional luminescence and absorption spectroscopy in analytical chemistry. *Chemical Reviews* 91: 493–507.
- Rossi TM and Warner IM (1985) Bacterial identification using fluorescence spectroscopy. In: Nelson WH (ed.) *Rapid Detection and Identification of Microorganisms*, pp. 1–50. New York: Pergamon.
- Talmi Y (1975) *Multichannel Image Detectors*. ACS Symposium Series 102. Washington: American Chemical Society.
- Warner IM, Patonay G, and Thomas MP (1985) Multidimensional luminescence measurements. *Analytical Chemistry* 57: 463A–483A.
- Wolfbeis OS and Leiner M (1985) Mapping of the total fluorescence of human blood serum as a new method for its characterization. *Analytica Chimica Acta* 167: 203–215.

## High-Resolution Techniques

**C Gooijer, F Ariese, and N H Velthorst**, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Fluorescence spectrometry is a well-known method for quantitative analysis of various classes of molecules. Compared to absorption spectrometry it provides in general higher sensitivity and selectivity. However, conventional room-temperature fluorescence excitation and emission spectra are usually broad (e.g., 10 nm or more) and show no or hardly any fine structure. Obviously, its potential for qualitative analysis would be strongly extended if an increase in spectral resolution could be obtained, so that the vibrational fine structure of the spectra became visible (see **Figure 1**). This has been realized in high-resolution fluorescence spectroscopy, which has a sensitivity similar to that of conventional fluorescence spectroscopy and a selectivity comparable to that of infrared (IR) spectroscopy.

Why are conventional fluorescence spectra of molecules in liquid or solid solutions almost featureless?

First, it should be realized that for a particular analyte molecule the interaction with its surrounding cage strongly influences the energy distance between the electronic excited state ( $S_1$ ) and the electronic ground state ( $S_0$ ). In a solid, amorphous solution the specific cage conformation and thus the associated  $S_0$ – $S_1$  energy distance differs from molecule to molecule. In a fluid solution, the cage configurations also fluctuate strongly in time. Second, in fluids frequent inelastic collisions of dissolved molecules and solvent molecules occur, leading to considerable shortening of the excited-state lifetime of the analyte. Finally, there can be numerous vibronic (where vibrational and electronic states change simultaneously) fluorescence transitions, separated by relatively small energy distances, that will merge into featureless bands because they are all subject to the band-broadening processes mentioned above.

Lowering the temperature is a first step to enhance the spectral resolution. Although some resolution may be gained this way, in most cases cooling alone is not sufficient to obtain narrow-banded electronic emission spectra. As noted above, the sample will contain a large ensemble of molecules, each with a different solvent cage and a statistical distribution of

approach need not be limited to the identification and quantification of almost every analyte in the sample mixture. For example, in the case of an oil spill it is significant to examine the entire sample to produce a fingerprint for use in the identification of patterns relating to its classification. The goal of multidimensional fluorescence measurements is to enhance selectivity. Furthermore, the resolution and accuracy of this technique improve the sensitivity of the method. Despite severe spectral overlap, the results always indicate that multidimensional analysis is superior to conventional fluorimetric analysis.

**See also:** **Fluorescence:** Instrumentation; Derivatization; Clinical and Drug Applications; Environmental Applications. **Fuels:** Oil-Based. **Polycyclic Aromatic Hydrocarbons:** Determination; Environmental Applications.

## Further Reading

- Alexander J, Mashak G, Kapitan N, and Siegel AJ (1987) Fluorescence of petroleum products II: Three-dimensional fluorescence plots of gasoline. *Journal of Forensic Sciences* 32: 72–86.
- Bentz AP (1976) Oil spill identification. *Analytical Chemistry* 48: 454A–472A.
- Christian GD, Callis JB, and Davidson ER (1981) Array detectors and excitation emission matrices in multicomponent analysis. In: Wehry EL (ed.) *Modern Fluorescence Spectroscopy*. New York: Plenum.
- Lakowicz JR (1983) *Principles of Fluorescence Spectroscopy*. New York: Plenum.
- Ndou TT and Warner IM (1991) Multidimensional luminescence and absorption spectroscopy in analytical chemistry. *Chemical Reviews* 91: 493–507.
- Rossi TM and Warner IM (1985) Bacterial identification using fluorescence spectroscopy. In: Nelson WH (ed.) *Rapid Detection and Identification of Microorganisms*, pp. 1–50. New York: Pergamon.
- Talmi Y (1975) *Multichannel Image Detectors*. ACS Symposium Series 102. Washington: American Chemical Society.
- Warner IM, Patonay G, and Thomas MP (1985) Multidimensional luminescence measurements. *Analytical Chemistry* 57: 463A–483A.
- Wolfbeis OS and Leiner M (1985) Mapping of the total fluorescence of human blood serum as a new method for its characterization. *Analytica Chimica Acta* 167: 203–215.

## High-Resolution Techniques

**C Gooijer, F Ariese, and N H Velthorst**, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

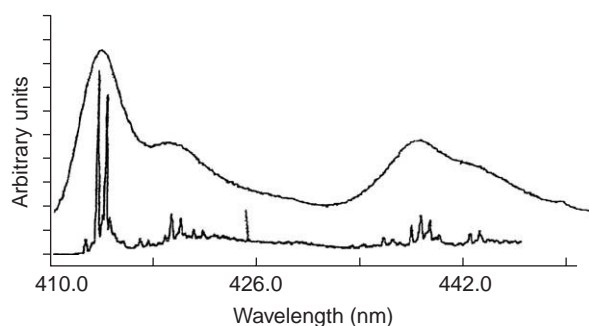
## Introduction

Fluorescence spectrometry is a well-known method for quantitative analysis of various classes of molecules. Compared to absorption spectrometry it provides in general higher sensitivity and selectivity. However, conventional room-temperature fluorescence excitation and emission spectra are usually broad (e.g., 10 nm or more) and show no or hardly any fine structure. Obviously, its potential for qualitative analysis would be strongly extended if an increase in spectral resolution could be obtained, so that the vibrational fine structure of the spectra became visible (see **Figure 1**). This has been realized in high-resolution fluorescence spectroscopy, which has a sensitivity similar to that of conventional fluorescence spectroscopy and a selectivity comparable to that of infrared (IR) spectroscopy.

Why are conventional fluorescence spectra of molecules in liquid or solid solutions almost featureless?

First, it should be realized that for a particular analyte molecule the interaction with its surrounding cage strongly influences the energy distance between the electronic excited state ( $S_1$ ) and the electronic ground state ( $S_0$ ). In a solid, amorphous solution the specific cage conformation and thus the associated  $S_0$ – $S_1$  energy distance differs from molecule to molecule. In a fluid solution, the cage configurations also fluctuate strongly in time. Second, in fluids frequent inelastic collisions of dissolved molecules and solvent molecules occur, leading to considerable shortening of the excited-state lifetime of the analyte. Finally, there can be numerous vibronic (where vibrational and electronic states change simultaneously) fluorescence transitions, separated by relatively small energy distances, that will merge into featureless bands because they are all subject to the band-broadening processes mentioned above.

Lowering the temperature is a first step to enhance the spectral resolution. Although some resolution may be gained this way, in most cases cooling alone is not sufficient to obtain narrow-banded electronic emission spectra. As noted above, the sample will contain a large ensemble of molecules, each with a different solvent cage and a statistical distribution of



**Figure 1** Fluorescence of 1-methoxybenz[a]pyrene in *n*-octane at room temperature and at 10 K. Room temperature spectrum (upper line): concentration,  $2.5 \times 10^{-5} \text{ mol l}^{-1}$ ; bandpass, 0.8 nm. Shpol'skii spectrum (lower line): concentration,  $5 \times 10^{-6} \text{ mol l}^{-1}$ ; band pass, 0.15 nm. The spectra were recorded using otherwise the same experimental settings; lamp excitation at 295 nm.

transition energies is observed. In an amorphous frozen solution, the spectral bandwidths are in general of the same order of magnitude as measured in fluid media. In order to obtain sufficiently narrow fluorescence spectra the inhomogeneous broadening of the spectral bands has to be strongly reduced. This can be achieved in mixed crystal, matrix isolation, and Shpol'skii spectroscopy, solid-state techniques aimed at the creation of equal solute-solvent cage interactions for all the analyte molecules. They differ with respect to the type of matrix utilized, the manner in which the guest molecules are incorporated in the host, and the preparation of the sample.

Fluorescence line-narrowing (FLN) is another technique with which highly resolved fluorescence spectra can be obtained. The method is in principle more universally applicable and makes use of selective excitation, preferably with a laser. Only those analyte molecules that have a specific  $S_1-S_0$  energy difference exactly corresponding to that of the laser wavelength are selected. Ideally, in the solid-state matrix the population thus created remains conserved and will produce a highly resolved FLN spectrum.

In this article, we will concentrate on Shpol'skii spectroscopy (SS), matrix isolation (MI) spectroscopy in relation to SS, and FLN spectroscopy with emphasis on analytical applications. Mixed crystal spectroscopy will not be considered because virtually no analytical applications have been realized thus far. In this method the guest molecules have to be incorporated substitutionally in the host crystal without appreciable distortion of the crystal lattice. This means that the molecular geometries of guest and host are very critical and that the preparation of a good-quality mixed crystal is tedious. Finally, it is

noted that solid-state techniques are not the only techniques to achieve high resolution. In the past a lot of effort has been devoted to supersonic expansion fluorimetry: under favorable conditions analyte molecules seeded in supersonic jets can be considered as isolated and extremely cold, thus producing highly resolved fluorescence spectra. Very sharp spectra can be obtained, yielding information on vibrational and rotational transitions. However, the analytical potential of the technique seems limited since the detection limits achieved are relatively unfavorable. In this section, only solid-state techniques will be discussed.

## Principles

The principles of three cryogenic high-resolution fluorescence techniques will be outlined, namely MI spectroscopy, SS, and FLN spectroscopy.

### Matrix Isolation Spectroscopy

In this technique, originally developed in the 1950s to study transient (especially inorganic) species, the sample is vaporized and mixed with a large amount of an inert gas, the so-called matrix gas, generally in a molar ratio of  $10^4$ – $10^8$ . Next, the mixture is deposited on a cryogenic surface held at a temperature of 15 K or lower, which can easily be obtained with helium cryostats. Ideally, in this way the solute species are matrix isolated. Inert matrix gases originally used were the noble gases neon, argon, and xenon; later, especially for organic compounds, nitrogen and even *n*-alkanes and *n*-perfluoroalkanes were applied as matrix gas (see Figure 2). In the ideal situation the characteristics of these so-called low-temperature matrix-isolated solutions are the absence of solute aggregates, no diffusion of the solute molecules, no rearrangement of the matrix lattice, and minimal interaction between the matrix and the solute molecules. Considering these features, it is obvious that the MI technique appears very suitable for obtaining highly resolved spectra of organic compounds for quantification and qualitative analysis. However, especially if lamp excitation is used, the spectral resolution observed for organic compounds in conventional matrices of nitrogen and argon is often only slightly better than that obtained in glassy frozen solutions. Apparently, the reduction of inhomogeneous broadening in these examples is rather small. For this reason, lasers have been introduced to reduce the inhomogeneous broadening by the FLN effect (see below) and 'Shpol'skii-like' matrix gases such as *n*-alkanes and *n*-perfluoroalkanes have been applied.



In order to utilize MI for quantitative analysis an internal standard should be included to compensate for incomplete recovery, i.e., nonquantitative analyte deposition. Important criteria for the internal standard are that its fluorescent spectrum should not overlap with the spectra of the analyte(s), it should sublime at approximately the same temperature as the analyte(s), and the excitation wavelength should be the same as for the analyte(s).

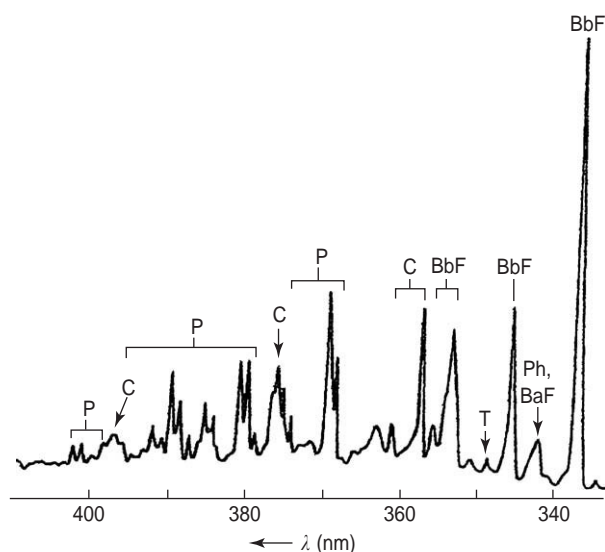
### Shpol'skii Spectroscopy

In 1952, Shpol'skii, a Soviet scientist, reported the observation of narrow spectral lines in the emission spectra of aromatic molecules dissolved in *n*-alkanes (*n*-C5 to *n*-C12), measured at low temperatures ( $T \leq 77$  K). By freezing the *n*-alkanes a polycrystalline

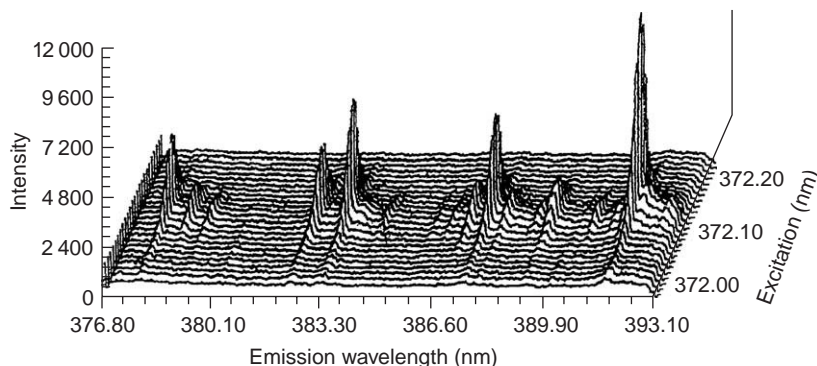
rather than a glassy solid matrix is obtained, in which the analyte molecules are trapped at the point of solidification. The compatibility of guest and host geometry is less stringent than in mixed crystals. In the polycrystalline matrix a reduction in the inhomogeneity of the environment of the guest molecules is achieved, and at temperatures below 20 K narrow spectral bands with bandwidths  $\sim 0.1$  nm may be observed. The Shpol'skii spectra often display several identical spectra that are displaced with respect to each other over a few nanometers (see the doublet spectrum in Figure 1). This so-called site structure is due to subpopulations of chemically identical molecules that are differently oriented inside the crystalline matrix, and is strongly dependent on the matrix used. This phenomenon is unfavorable in terms of the sensitivity of the method, but sometimes profitable for identification purposes.

As the Shpol'skii effect is induced by the matrix, both the absorption and emission spectra measured at low temperatures will be narrow-banded. This holds especially for the long wavelengths (i.e.,  $S_0$ - $S_1$  absorption); shorter-wavelength bands are broadened as a result of the lifetimes of the higher excited states (the Heisenberg uncertainty principle). Thus, using a laser as excitation source (laser-excited Shpol'skii spectroscopy, LESS) an enhancement of the selectivity can be realized. Figure 3 shows a three-dimensional excitation emission spectrum of pyrene in *n*-octane. Emission spectra were continuously recorded as the laser was tuned through the 0-0 transition. Pyrene- $d_{10}$  was also present in the mixture but, although its excitation wavelength is shifted over only 1 nm, it is completely invisible, as the laser line does not match it exactly. It is clear that we have here a powerful tool for increasing the emission of a particular analyte and at the same time reducing interferences from other compounds.

A serious drawback of SS is that the choice of matrices is restricted and that for many analytes



**Figure 2** Matrix isolation spectra for a six-component mixture of polycyclic aromatic hydrocarbons (PAHs); lamp excitation at 313 nm. P=pyrene, C=chrysene, BbF=benzo[*b*]fluorene, T=triphenylene, Ph=phenanthrene, BaF=benzo[*a*]fluorene.



**Figure 3** Excitation – emission plot of pyrene ( $5 \times 10^{-7}$  mol  $l^{-1}$  in *n*-octane) at different laser excitation wavelengths.

there may not be suitable solvents to obtain narrow-line spectra. Further, the shape of the spectra is influenced by experimental parameters, such as the cooling rate, the thermal treatment, the concentration of the sample, the solvents chosen, and the sample temperature during the spectroscopic measurement (preferably 20 K or lower). Fortunately, many large planar molecules yield Shpol'skii spectra that are not very sensitive to the thermal history of the sample. Furthermore, several types of internal standards can be employed to correct for such effects (see below).

### Fluorescence Line-Narrowing Spectroscopy

In this type of spectroscopy, first applied in 1972 by Personov, a reduction of the inhomogeneous broadening is realized by optical selection of the molecules that are excited. An excitation source is used with a spectral bandwidth that is considerably narrower than the inhomogeneous bandwidth of the molecules that are probed. With such a selective excitation process a well-defined subset of excited molecules is created; only molecules with an energy difference between ground state and excited state that is exactly equal to the energy of the exciting photons are excited (see Figure 4). It is clear that in this case lasers are the most appropriate excitation source. If during the lifetime of the excited state no spectral diffusion can occur, only this well-defined subset or 'isochromat' will give fluorescence. This means that the measurements have to be performed at low temperatures, at least at 20 K or lower. Excitation may occur in the 0-0 electronic transition (i.e., from the lowest vibrational state in  $S_0$  to the lowest

vibrational state in  $S_1$ ), or into vibrational states of  $S_1$ . The latter approach may lead to multiplets, as several vibronic bands are probed simultaneously and decay to their corresponding purely electronic states, thus creating several well-defined subsets of excited molecules. These multiplets have been used for identification purposes of closely related isomers. Usually, excitation into high vibrational states more than  $1600\text{ cm}^{-1}$  above the 0-0 transition or into higher electronic states does not lead to good line-narrowed spectra.

Although FLN has been applied in various solvents, there is no total freedom of solvent choice. The main prerequisite is that there is no strong interaction between the matrix and guest molecules. Strong electron-phonon coupling (phonons are vibrational motions throughout the matrix) prohibits the observation of line narrowing in the spectrum.

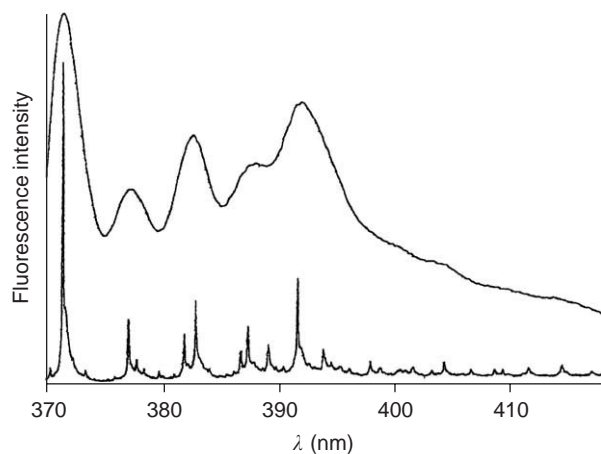
High analyte concentrations should be avoided since intermolecular energy transfer can lead to line broadening. The same holds for local heating of the matrix, which may lead to site interconversion due to diffusion of the matrix and to a loss of the line-narrowing effect. Finally, the occurrence of 'hole-burning' effects in FLN spectra should be mentioned. Two types are discerned, i.e., photochemical and nonphotochemical hole burning. The former is due to photochemical reactions of the excited molecules. In nonphotochemical hole burning, reorientation of guest molecules in the excited state with respect to their solvent cage plays a role. Both effects result in a gradual decrease in the ground-state population that can be selectively excited by the laser and hence in a reduction of the narrow-line intensities in the spectrum upon prolonged irradiation.

Because of the broader choice in solvents compared to SS, the applicability of FLN is much wider. On the other hand, the sensitivity of FLN is in principle not as high as that of LESS since, due to the optical selection, only a fraction of the analyte molecules is detected.

### Instrumentation

To date no complete, dedicated experimental setups are commercially available; only a few research groups have been able to apply the low-temperature high-resolution techniques. The instrumentation is regarded as exotic in analytical laboratories: lasers, cryogenic cooling apparatus, and high-resolution detection devices are not common equipment.

In this section, some instrumental aspects that are relevant for the achievement of highly resolved fluorescence spectra will be discussed. Particular



**Figure 4** Fluorescence emission spectra of 20 ng pyrene on a C-18 modified silica HPTLC plate. Upper spectrum: lamp excitation at room temperature. Lower spectrum: laser excitation at 363.5 nm;  $T = 10\text{ K}$ .



attention will be paid to excitation, cooling, and detection systems.

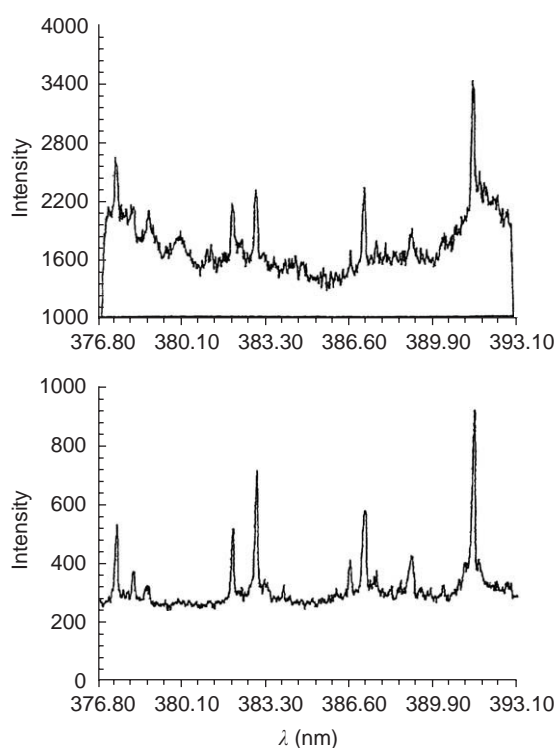
### Excitation

In the past, the most frequently used excitation sources for matrix-induced line-narrowing (as in SS) were conventional light sources such as xenon and mercury lamps in combination with a monochromator of sufficient spectral resolving power. Obviously, these sources are inappropriate if selective excitation is essential and bandwidths down to  $\sim 0.2$  nm are required. These can only be realized by using narrow monochromator slits, with, as a consequence, extremely low light intensities. The availability of robust and user-friendly laser systems has had a strong impact on high-resolution fluorescence spectroscopy. Considerable improvements in sensitivity and selectivity have been achieved by using a tunable laser instead of a lamp (LESS), since not only the fluorescence emission but also the long-wavelength part of the excitation spectra is composed of narrow lines. For FLN spectroscopy laser excitation is essential.

Within this context it is appropriate to discuss the impact of lasers in more general terms. In general, the exchange of classical light sources for lasers has greatly extended the analytical potential of molecular fluorescence techniques. Some inherent features of lasers are of extreme importance. First, the directionality of the radiation, which not only enables collimation down to spot sizes of  $1\text{ }\mu\text{m}$  but also guarantees a sufficiently high sample irradiance to perform analysis far away from the light source (as in remote sensing). The inherent high monochromaticity of lasers is of only limited use for the analysis of fluid or frozen solutions with broadband excitation spectra, since only a small fraction of analytes would have adequate spectral overlap with the laser line. In contrast it is very important, even essential, in FLN and LESS. It must be realized that the fluorescence signal of the analyte increases with the irradiance (power of the laser in watts per unit area of the sample), but that excessive irradiances can cause problems, e.g., detector cell damage due to plasma formation, saturation effects, and photochemical decomposition of the fluorophores. These effects most often occur with pulsed lasers with high peak powers.

Finally, the ability to produce pulsed laser radiation with a short lifetime on the fluorescence timescale opens up the possibility of applying time-resolved detection principles to improve the signal-to-background ratio (see Figure 5).

For laser application in analytical spectroscopy it is important that laser lines are available over a large wavelength region. In particular, laser output in the



**Figure 5** Shpol'skii fluorescence spectra of pyrene in crude bird extract, *n*-octane at 28 K; laser excitation 372.10 nm. Upper spectrum: without time resolution. Lower spectrum: with time-resolved detection; delay = 50 ns.

ultraviolet (UV) is necessary for the ability to detect all types of compounds. UV laser radiation can be obtained by means of dye lasers or nonlinear techniques such as frequency doubling. Nevertheless, laser wavelength selection is still far less convenient than wavelength selection in a conventional scanning fluorimeter.

### Cooling Systems

For the high-resolution fluorescence techniques described here, samples are measured under cryogenic conditions. Cooling of the sample to 10–30 K can be performed in different ways: by means of a bath cryostat, a closed-cycle helium refrigerator, or a flow-through cryostat. In a bath cryostat the sample is immersed in liquid helium so that a temperature of 4.2 K can be achieved almost instantaneously. Disadvantages are the consumption of liquid helium and the presence of boiling helium in the optical pathways of most cryostat designs, which may disturb the measurements.

A very convenient cooling apparatus is the closed-cycle helium refrigerator, in which repeated compression and expansion of gaseous helium takes place. The sample is placed in a vacuum and attached

to a cold finger in contact with the cryogenic medium. Good thermal contact between the sample holder and the cold tip is crucial. Nowadays, temperatures as low as 4 K can be achieved. This apparatus has the advantages that it requires only line electrical power and that no expensive liquid helium is consumed.

The helium flow-through cryostat uses liquid helium that is pumped from a container and sprayed onto the bottom of the cold tip. In this case, temperatures of 3.5 K can be reached. The cooling process takes a long time (typically 1 h), both with the closed-cycle helium refrigerator and the flow-through cryostat. Several samples should preferably be cooled simultaneously.

### Detection

High-quality (emission) monochromators are necessary to disperse the emitted light and record narrow-banded spectra. Particularly if there is only a small difference between excitation and emission wavelengths, the interference from scattered excitation light (likely to be strong in the case of solid samples) has to be minimized, so that the use of a triple monochromator is preferred. In many of the earlier setups, detection was done with a (cooled) photomultiplier tube (PMT), measuring the averaged photocurrent, coupled to a lock-in amplifier. Low-level light intensities can be detected more effectively with a fast PMT, followed by a photon counting system. However, in order to record a full high-resolution spectrum the scanning process may take very long.

In recent years, multichannel detectors have become more and more popular, especially intensified linear diode arrays and CCD cameras. These devices offer a sensitivity similar to that of PMTs, while a complete spectrum can be recorded simultaneously. Apart from the substantial reduction in measurement time, another advantage is the fact that in the case of signal fluctuations (e.g., due to photodegradation, hole burning) all peaks in the spectrum are equally affected. For time-resolved detection a pulsed excitation source and a fast gateable intensified camera can be used.

### Applications

In this section, some representative examples of analytical applications are presented. It is not intended to give a complete overview.

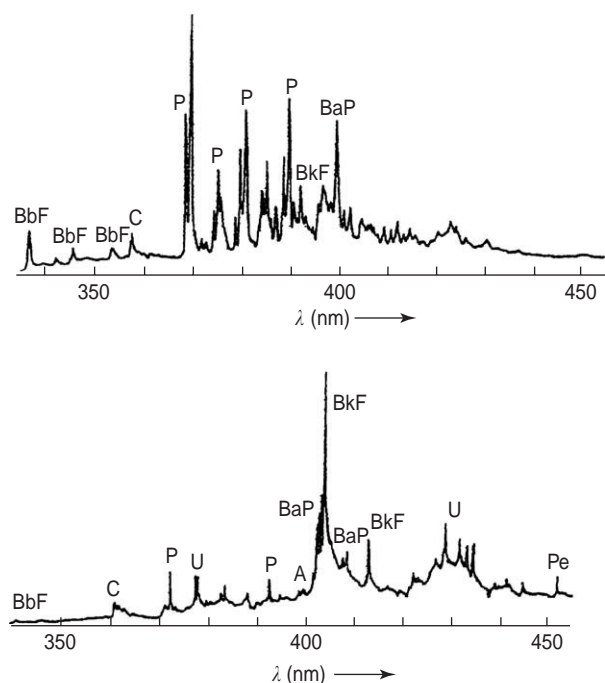
#### Matrix Isolation

Currently, MI sample preparation techniques are used mainly in combination with vibrational

spectroscopy, especially Fourier transform IR, while fluorescence spectroscopy is involved to a minor degree. Fluorescence studies of organic compounds have been devoted to unstable, reactive species. Applications in analytical chemistry have been published mainly by Wehry's group, focusing especially on polycyclic aromatic hydrocarbons (PAHs).

An example is the analysis of very complex samples such as chromatographic fractions of coking plant wastewater. **Figure 6** shows the MI high-resolution fluorescence spectra obtained by applying broadband excitation by a xenon-mercury lamp at 15 K. In the upper spectrum nitrogen is the matrix gas; in the lower one *n*-heptane is utilized and a 'Shpol'skii-type' spectrum is obtained. The selectivity of the MI technique is obvious: even under lamp excitation six PAHs could be simultaneously identified in the complex sample.

Within the context of the development of coupled techniques the combination of gas chromatography and MI (GC-MI) utilizing fluorescence spectroscopy for detection and identification is interesting. Until recently this development was hampered by the long scan times required to obtain high-resolution fluorescence spectra of deposited species. Optical multi-channel detection can now be invoked to tackle this problem.



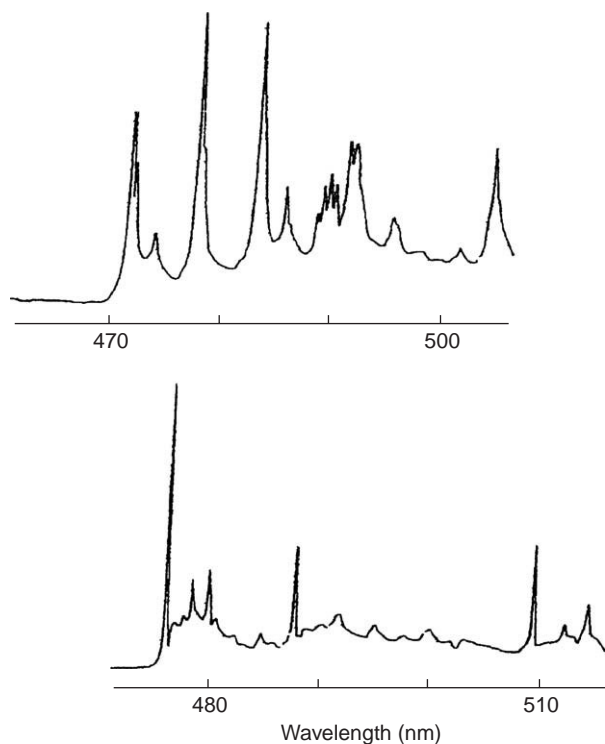
**Figure 6** Lamp-excited MI spectra of chromatographic fraction of coking plant wastewater, in nitrogen (upper spectrum) and *n*-heptane (lower spectrum) matrices at 15 K. BbF = benzo[*b*]fluororene, C = chrysene, P = pyrene, BkF = benzo[*k*]fluoranthene, BaP = benzo[*a*]pyrene, Pe = perylene, U = unidentified.

Finally, it is noted that the MI technique is not exclusively applicable to analytes exhibiting native fluorescence. In the literature much attention has been paid to laser-photolytic gas-phase fragmentation reactions producing fragments that can be detected by means of fluorescence. Instead of fluorescence, molecular phosphorescence has also been utilized. This is illustrated in **Figure 7**, showing laser-induced phosphorescence spectra of 2-chloronaphthalene. In this experiment, the appropriate matrix was composed of naphthalene, the parent molecule. In fact, such a matrix will behave like a mixed crystal, which explains the strong improvement of the spectrum quality on going from *n*-pentane to naphthalene. The approach enables high selectivity for halogenated naphthalenes in general.

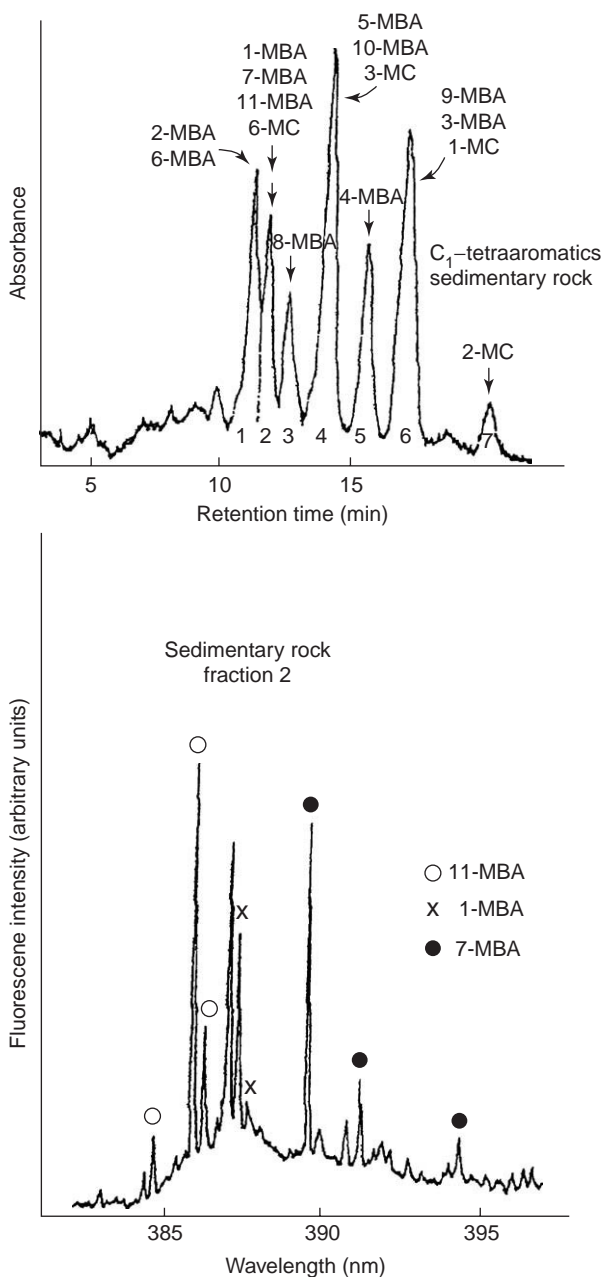
### Shpol'skii Spectroscopy

To date hundreds of papers have been published dealing with the application of SS in chemical analysis, often devoted to PAHs in environmental matrices (e.g., air, river, and marine sediments, coal and oils). The main object has been qualitative analysis, which is not surprising in view of the detailed vibrational structures exhibited by the spectra in combination with the characteristic wavelength of the

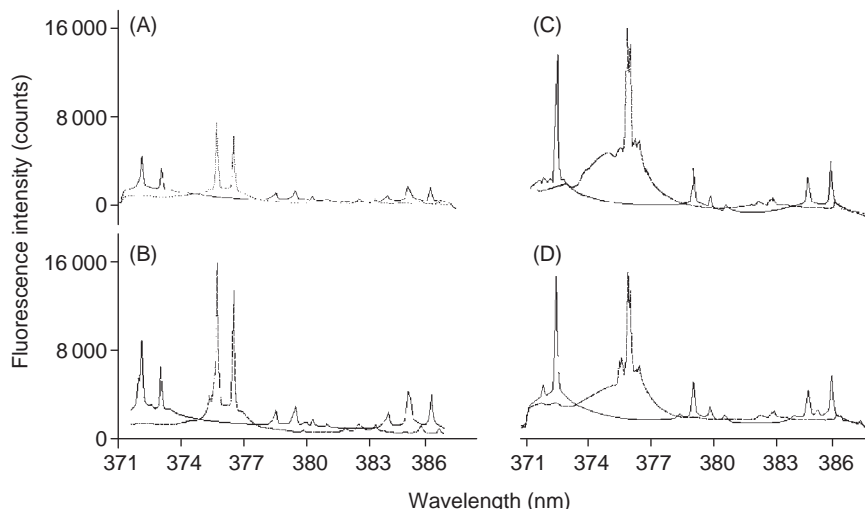
electronic origin in a specific matrix. (It is to be noted that the latter feature is not provided by FLN since the origin of FLN spectra is largely determined by the laser wavelength applied.) The selectivity of SS has been demonstrated by lamp- and laser-excited SS of structurally related methyl isomers of PAHs, using both fluorescence and phosphorescence measurements. As an example, **Figure 8** shows the



**Figure 7** Laser-induced phosphorescence spectra of 2-chloronaphthalene, vapor-deposited in an *n*-pentane matrix (upper; 320 nm exc.) and in a naphthalene matrix (lower; 324 nm exc.).



**Figure 8** Top frame: reversed-phase LC analysis of a tetra-aromatic fraction from a sedimentary rock with tentative assignment of methylbenz[a]anthracenes (MBA). Bottom frame: Shpol'skii fluorescence spectra ( $\lambda_{\text{ex}} = 294 \text{ nm}$ ) showing unambiguously the presence of 11-MBA, and 7-MBA.



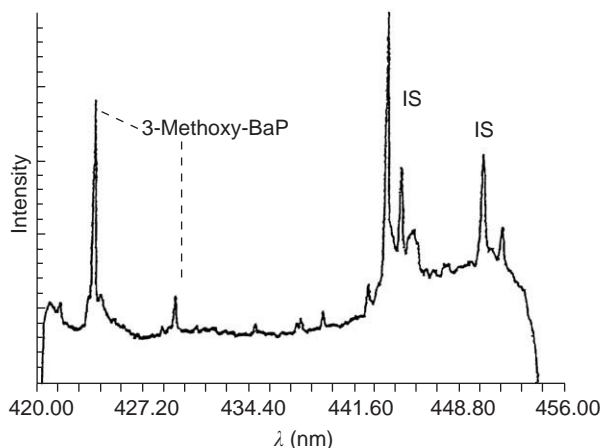
**Figure 9** Shpol'skii spectra of pyrene (—) and 1-fluoropyrene (·····) under different experimental conditions: (A) *n*-hexane matrix, slow cryocooling; (B) *n*-hexane matrix, fast precooling; (C) *n*-octane matrix, slow cryocooling; (D) *n*-octane matrix, fast precooling. Temperature, 11 K; analyte concentration,  $1 \mu\text{mol l}^{-1}$ ; lamp excitation at 338 nm.

identification of isomers of the mutagen methylbenz[*a*]anthracene (MBA) in a chromatographic fraction from a sedimentary rock.

Quantitative analysis by SS requires the use of an internal standard or a standard addition procedure in combination with an internal standard. This is necessary in view of irreproducibilities in sample preparation (including the cooling procedure) and optical alignment. Deuterated analogs of the analytes that show small though detectable spectroscopic shifts and have identical chemical behavior have proved very useful; in general, reproducibilities better than 10% can be realized. In recent years, an alternative set of internal standards were synthesized. Mono-fluorinated PAHs were found to yield high-quality Shpol'skii spectra (see **Figure 9**) and proved to be very useful as internal standards to correct for differences in alignment and thermal history.

With SS (and especially with LESS) extremely low detection limits can be achieved. To quote an example, for the highly fluorescent PAH benzo[*a*]pyrene in a 'synthetic' solution in *n*-octane at 10 K and using lamp excitation a detection limit of  $0.075 \text{ ng ml}^{-1}$  has been published; with laser excitation this was improved to  $0.0012 \text{ ng ml}^{-1}$ .

It is emphasized that, owing to the inherent selectivity of (LE)SS, sample cleanup procedures can often be minimized. An important example is the analysis in a crude river sediment extract of ultralow traces of dibenzo[*a,l*]pyrene, the most carcinogenic PAH known to man. Another interesting application is the identification and quantification of PAH metabolites in samples for biological monitoring. Although SS is not appropriate for highly polar compounds it can be applied to monohydroxy PAHs.



**Figure 10** Shpol'skii spectrum of methylated bile extract from fish from Wadden Sea mesocosm (least polluted control group). Site-selective laser excitation at 418.36 nm; IS (internal standard) = perylene- $d_{12}$ .

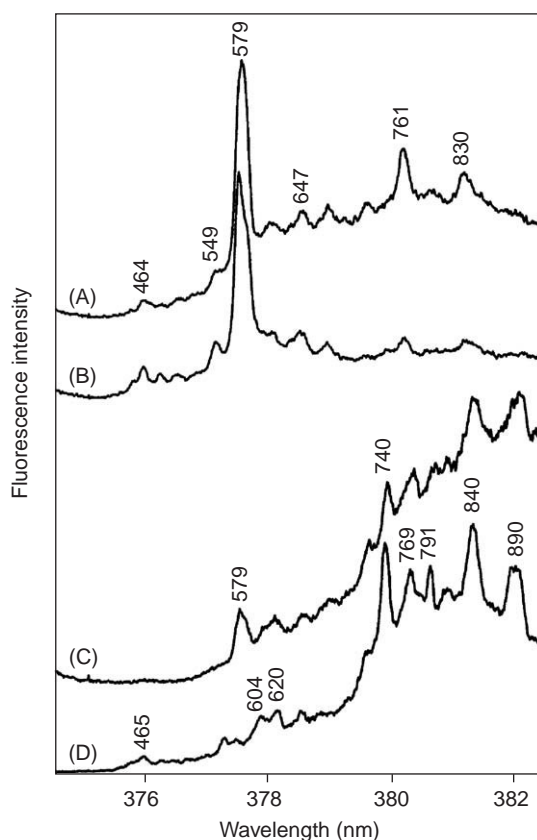
An important improvement of sensitivity was obtained by methylation of the hydroxy groups. As a result even in the bile of fish caught in relatively clean areas (as the Wadden Sea) BaP metabolites could be measured and quantified; this is illustrated in **Figure 10**, showing a LESS spectrum utilizing selective excitation and time-resolved detection. A similar procedure has been applied to determine 3-hydroxybenzo[*a*]pyrene in the urine of coke oven workers.

An interesting recent example is the use of Shpol'skii matrices to study ultrafast phenomena. The excited state proton transfer reaction in 3-hydroxyflavone could be calculated from the (relatively) broad bands in the Shpol'skii excitation spectrum, using the Heisenberg uncertainty principle.

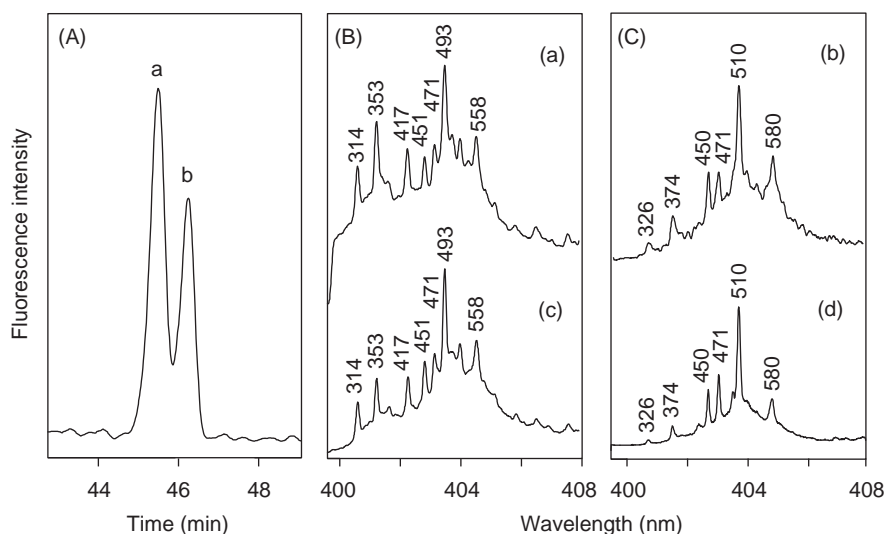
## Fluorescence Line-Narrowing

Early reports on analytical applications of FLN were devoted to PAHs in environmental samples such as cigarette tar, soot, solvent-refined coal, and gasoline. For this type of problem LESS would be the (more appropriate) alternative, provided that the analytes can be brought into *n*-alkane solvents. Such an alternative is not available for the study of polar PAH metabolites (the activated forms of PAHs playing a suspected role in carcinogenesis) and their adducts with DNA, as they are poorly soluble in alkanes. Elucidation of their detailed structures is crucial to acquire insight into the mechanism of adduct formation and DNA damage. As shown by Small, Jankowiak, and co-workers, FLN can be of significant help, especially when adducts isomers with different stereochemistries need to be distinguished. **Figure 11** shows the FLN spectra of four different benzo[*a*]pyrene-DNA adducts in short pieces of synthetic double-stranded DNA. *cis* and *trans* adducts can be distinguished based on different vibrational frequencies, while the (+) and (−) stereochemistries can only be distinguished due to the different modes of embedding in the DNA helix. Since the FLN spectra are very sensitive to the nature of the surrounding microenvironment, the technique can also be used to study the interactions inside a receptor site or inside an antibody.

An interesting recent finding is that FLN is applicable to analytes on various types of thin-layer chromatographic (TLC) plates, after cooling the plate to 10–20 K. This enables the application of FLN



**Figure 11** FLN spectra of *anti*-BPDE-N<sup>2</sup>-dG adducts of different stereochemistries in duplex oligonucleotides. Curve A: (+)-*trans* adduct; curve B: (−)-*trans* adduct; curve C: (+)-*cis* adduct; curve D: (−)-*cis* adduct.  $\lambda_{\text{ex}} = 369.48$  nm. The FLN peaks are labeled with their excited-state vibrational frequencies in  $\text{cm}^{-1}$ .



**Figure 12** (A) Room-temperature fluorescence electropherogram for a mixture of (a) B[a]P-d<sub>12</sub> and (b) B[a]P. Frames (B) and (C): FLN spectra obtained after freezing the CE capillary to 4.2 K,  $\lambda_{\text{ex}} = 395.7$  nm, obtained for the CE-separated analytes (a) and (b). Spectra (c) and (d) are reference spectra recorded solutions of the pure compounds in aqueous buffer. The FLN peaks are labeled with their S<sub>1</sub> vibrational frequencies, in  $\text{cm}^{-1}$ .



identification after TLC separation. Even more importantly, the combination with liquid chromatography (LC) is possible, since LC and TLC can be coupled successfully: the effluent from the LC column is deposited on a TLC plate moving with a constant speed so that an 'immobilized' chromatogram is obtained. Efficient eluent elimination and conservation of chromatographic integrity during the deposition process are crucial. Another promising approach is the online coupling of FLN and capillary electrophoresis. This is done by freezing the separated compounds inside the capillary to 4K and recording FLN spectra of the separate analyte zones (see Figure 12).

**See also:** **Fluorescence:** Instrumentation; Derivatization; Quantitative Analysis. **Laser-Based Techniques.** **Phosphorescence:** Principles and Instrumentation. **Polycyclic Aromatic Hydrocarbons:** Determination; Environmental Applications. **Quality Assurance:** Internal Standards.

## Further Reading

Ariese F (1993) *Shpol'skii Spectroscopy and Synchronous Fluorescence Spectroscopy*. PhD Thesis, Free University Amsterdam.

Bader AN, Ariese F, and Gooijer C (2002) Proton transfer in 3-hydroxyflavone studied by high-resolution 10 K laser-excited Shpol'skii spectroscopy. *Journal of Physical Chemistry A* 106: 2844–2849.

Gooijer C, Ariese F, and Hofstra JW (2000) *Shpol'skii Spectroscopy and Other Site-Selection Methods*. New York: Wiley.

Hofstra JW, Gooijer C, and Velthorst NH (1988) Highly resolved molecular luminescence spectroscopy. *Chemical Analysis Series* 77(2): 283–400.

Hofstra JW, Gooijer C, and Velthorst NH (1993) Laser-excited molecular fluorescence in analytical sciences. *Chemical Analysis Series* 77(3): 325–443.

Luthe G, Ariese F, and Brinkman UAT (2002) Monofluorinated aromatic hydrocarbons as standards in environmental chemistry. In: Neilson AH (ed.) *Handbook of Environmental Chemistry: Organic Fluorine Compounds*, pp. 249–275. Heidelberg: Springer.

Personov RI (1983) Site selection spectroscopy of complex molecules in solutions and its applications. *Spectroscopy and Excitation Dynamics of Condensed Molecular Systems* 10: 555–619.

Scott Cooper R, Jankowiak R, and Small GJ (1991) Bioanalytical applications of fluorescence line narrowing spectrometry. *Practical Spectroscopy Series* 12: 201–236.

Wehry EL (1986) Cryogenic molecular fluorescence spectrometry. *Chemical Analysis Series* 87: 211–271.

## Time-Resolved Fluorescence

**M E Díaz García and R Badía Laíño**, University of Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Fluorimetry has experienced an explosive growth since the early 1980s, much of which has been driven by the use of fluorescence as a noninvasive technique for biology and biochemistry. Fluorescence techniques are widely used to quantify molecular parameters of different chemical, biochemical, and biological processes because of their inherent sensitivity, specificity and temporal resolution. In fact, the luminescence lifetime is an important characteristic of a fluorescent molecule and its environment. Many intra- and intermolecular processes are able to modulate the molecule emission which cannot be investigated by steady-state fluorescence measurements. For example, rotational diffusion,

resonance-energy transfer, or dynamic quenching occur on the same timescale as the fluorescence decay. Lifetime measurements reveal dynamic information on the nanosecond timescale that is useful in fundamental studies of quenching and energy transfer and can be used in analytical methodologies to enhance selectivity and/or to perform multicomponent determinations. For example, fluorescence quenching may occur due to a ground state reaction or an excited state reaction. Only by measuring the fluorescence lifetime may one determine which process causes the quenching.

There are several experimental approaches to obtain lifetime data. The primary objective of these approaches is to obtain data representing the time dependence of the decay of the luminescence. For a single component sample in which the excited state decays by first-order processes the radiative fluorescence lifetime  $\tau_r$  represents the time the fluorophore remains on average in the excited state, before emission takes place. If there are  $N$  molecules in the excited state, the decrease  $dN$  of molecules that revert



identification after TLC separation. Even more importantly, the combination with liquid chromatography (LC) is possible, since LC and TLC can be coupled successfully: the effluent from the LC column is deposited on a TLC plate moving with a constant speed so that an 'immobilized' chromatogram is obtained. Efficient eluent elimination and conservation of chromatographic integrity during the deposition process are crucial. Another promising approach is the online coupling of FLN and capillary electrophoresis. This is done by freezing the separated compounds inside the capillary to 4 K and recording FLN spectra of the separate analyte zones (see Figure 12).

**See also:** **Fluorescence:** Instrumentation; Derivatization; Quantitative Analysis. **Laser-Based Techniques.** **Phosphorescence:** Principles and Instrumentation. **Polycyclic Aromatic Hydrocarbons:** Determination; Environmental Applications. **Quality Assurance:** Internal Standards.

## Further Reading

Ariese F (1993) *Shpol'skii Spectroscopy and Synchronous Fluorescence Spectroscopy*. PhD Thesis, Free University Amsterdam.

Bader AN, Ariese F, and Gooijer C (2002) Proton transfer in 3-hydroxyflavone studied by high-resolution 10 K laser-excited Shpol'skii spectroscopy. *Journal of Physical Chemistry A* 106: 2844–2849.

Gooijer C, Ariese F, and Hofstraat JW (2000) *Shpol'skii Spectroscopy and Other Site-Selection Methods*. New York: Wiley.

Hofstraat JW, Gooijer C, and Velthorst NH (1988) Highly resolved molecular luminescence spectroscopy. *Chemical Analysis Series* 77(2): 283–400.

Hofstraat JW, Gooijer C, and Velthorst NH (1993) Laser-excited molecular fluorescence in analytical sciences. *Chemical Analysis Series* 77(3): 325–443.

Luthe G, Ariese F, and Brinkman UAT (2002) Monofluorinated aromatic hydrocarbons as standards in environmental chemistry. In: Neilson AH (ed.) *Handbook of Environmental Chemistry: Organic Fluorine Compounds*, pp. 249–275. Heidelberg: Springer.

Personov RI (1983) Site selection spectroscopy of complex molecules in solutions and its applications. *Spectroscopy and Excitation Dynamics of Condensed Molecular Systems* 10: 555–619.

Scott Cooper R, Jankowiak R, and Small GJ (1991) Bioanalytical applications of fluorescence line narrowing spectrometry. *Practical Spectroscopy Series* 12: 201–236.

Wehry EL (1986) Cryogenic molecular fluorescence spectrometry. *Chemical Analysis Series* 87: 211–271.

## Time-Resolved Fluorescence

**M E Díaz García and R Badía Laíño**, University of Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Fluorimetry has experienced an explosive growth since the early 1980s, much of which has been driven by the use of fluorescence as a noninvasive technique for biology and biochemistry. Fluorescence techniques are widely used to quantify molecular parameters of different chemical, biochemical, and biological processes because of their inherent sensitivity, specificity and temporal resolution. In fact, the luminescence lifetime is an important characteristic of a fluorescent molecule and its environment. Many intra- and intermolecular processes are able to modulate the molecule emission which cannot be investigated by steady-state fluorescence measurements. For example, rotational diffusion,

resonance-energy transfer, or dynamic quenching occur on the same timescale as the fluorescence decay. Lifetime measurements reveal dynamic information on the nanosecond timescale that is useful in fundamental studies of quenching and energy transfer and can be used in analytical methodologies to enhance selectivity and/or to perform multicomponent determinations. For example, fluorescence quenching may occur due to a ground state reaction or an excited state reaction. Only by measuring the fluorescence lifetime may one determine which process causes the quenching.

There are several experimental approaches to obtain lifetime data. The primary objective of these approaches is to obtain data representing the time dependence of the decay of the luminescence. For a single component sample in which the excited state decays by first-order processes the radiative fluorescence lifetime  $\tau_r$  represents the time the fluorophore remains on average in the excited state, before emission takes place. If there are  $N$  molecules in the excited state, the decrease  $dN$  of molecules that revert

to the ground state in a given interval of time,  $dt$ , is expressed by the relation

$$\frac{-dN}{dt} = k_f N \quad [1]$$

where  $k_f$  is the fluorescence constant rate.

The number of excited molecules therefore decreases exponentially with time:

$$N(t) = -N_0 \exp(-k_f t) \quad [2]$$

The time after which the number of excited molecules decreases from  $N_0$  to  $N_0/e$  is called the fluorescence lifetime. Taking into account the dependence of the luminescence signal on analyte concentration, the above equation can be expressed as

$$I(t) = I_0 \exp(-k_f t) \quad [3]$$

where  $I(t)$  is the luminescence signal after time  $t$  and  $I_0$  is the initial luminescence signal. If all the quanta of energy are lost as fluorescence, the fluorescence lifetime is given by the reciprocal fluorescence transition rate

$$\tau_f = \frac{1}{k_f} \quad [4]$$

In contrast to this radiation or intrinsic lifetime, the real lifetime,  $\tau_f$ , includes the nonradiative deactivation processes:

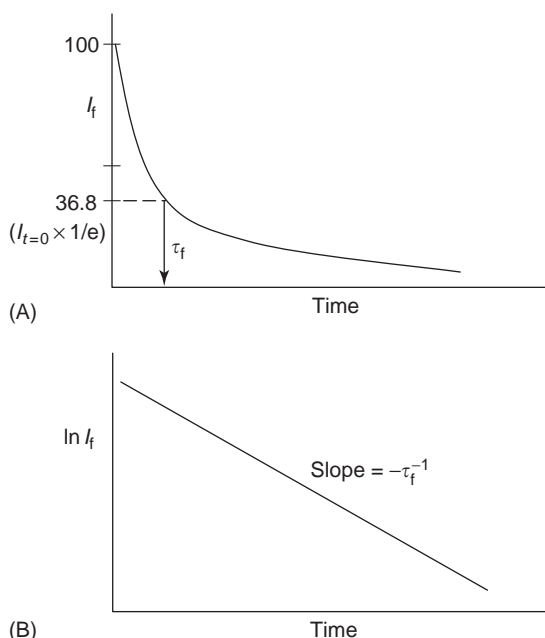
$$\tau_f = \frac{1}{k} = \frac{1}{k_f + k_{ISC} + k_Q} \quad [5]$$

The numerical value of  $\tau_f$  can be found from a plot of fluorescence intensity versus time, following termination of excitation (see **Figure 1A**).  $\tau_f$  is the time at which the intensity has decreased to  $1/e$  (36.8%) of its initial value. It is important to remember that fluorescence lifetime is an average time for a molecule to remain in the excited state before emitting a photon. Each individual molecule emits randomly after excitation. Many excited molecules will fluoresce before the average lifetime while some will also do long after the average lifetime.

Before the availability of laboratory computers, lifetimes were commonly determined by graphical procedures, for example, by taking the natural logarithm of both sides of eqn [3], we obtain

$$\ln I(t) = \ln I_0 - t/\tau_f \quad [6]$$

The negative reciprocal of the slope of a semilogarithmic plot of  $I(t)$  versus  $t$  yields the lifetime (see **Figure 1B**). This method assumes that the luminescence signal decays to zero for long times (background signals are supposed to be negligible). Other data-treatment approaches are available to extract the lifetime information and to compensate for residual background signal. For example, the natural



**Figure 1** (A) Hypothetical decay of fluorescence from an electronically excited state. (B) Logarithmic slope method to determine fluorescence lifetime.

logarithm of the difference in the luminescence signal between pairs of data points along the decay curve, evenly time-spaced, is plotted versus time (Guggenheim method). The lifetime is obtained from the negative reciprocal of the slope.

Although the calculated lifetime may have little physical meaning, for nonexponential decay and for multicomponent mixtures, the decay curve can be fitted to a weighted sum of exponential terms to obtain the lifetime of each component:

$$I(t) = \sum_{i=1}^n \alpha_i \exp(-t/\tau_i) \quad [7]$$

where  $\alpha_i$  is the pre-exponential factor denoting the fractional contribution to the total time-resolved decay of the component with lifetime  $\tau_i$ .

## Time-Domain Lifetime Measurements

Time-resolved fluorescence measurements are performed in two functionally equivalent ways: time domain and frequency domain. In time-domain or pulse-excitation fluorimetry, the fluorophore is excited with a short-duration pulse of light. The resulting fluorescence emission decay signal after the excitation pulse is captured with a fast recorder and the decay time is calculated from the slope of a plot of  $\log I(t)$  versus  $t$ . As the fluorescence species decay involves a process that occurs in a few nanoseconds,

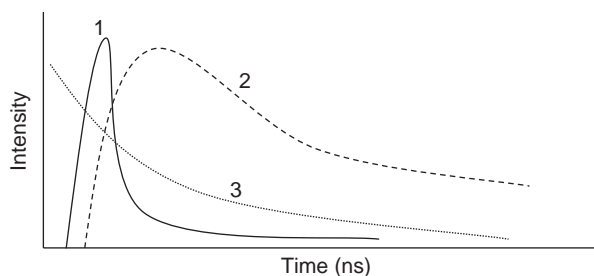
the width of the pulse should be as short as possible and it is preferably to be shorter than the decay time of the analyte. If the excitation pulse width ( $t_p$ ) is shorter than the luminescence lifetime ( $\tau_f$ ), an exponential decay of the luminescence is observed. With nanosecond excitation pulse widths, the decay curve represents a convolution of the source and fluorescence decay, as illustrated in **Figure 2**. In this case, the following equation applies:

$$I_f(t) = \int_0^t I_r(t') I_f(t - t') dt' \quad [8]$$

where  $I_r(t')$  is the reference signal related to the time profile of the excitation pulse,  $I_f(t)$  is the measured decay curve and  $t'$  is a dummy integration variable ( $t > t'$ ). The term  $(t - t')$  indicates that the impulse response is started at  $t = t'$  and it is supposed that no emission is observed before excitation ( $t < t'$ ). The true fluorescence decay curve may be extracted from the experimental decay curve by deconvolution methods, so that the convolution integral is solved for  $I_f(t)$ . Most use least-squares methods in concert with an iterative reconvolution scheme. In this approach, one picks a test model (eqn [7]). The convolution integral (eqn [8]) is then calculated on the basis of an initial set of  $\alpha_i$  and  $\tau_i$  values and the measured instrument response function. The calculated response ( $C(t)$ ) is compared with the observed data ( $I_f(t)$ ) and the  $\alpha_i$  and  $\tau_i$  terms adjusted until a best fit is obtained. The quality of the fit is judged by chi-square ( $\chi^2$ ) test

$$\chi^2 = \sum_{j=1}^{np} W_j [I_f(t) - C(t)]^2 \quad [9]$$

where np is the number of data points in the decay file and  $W_j$  is a weighting factor. These mathematical analyses have two inherent drawbacks: (1) any gain shift or any drift in the instrumentation is reflected in the data as changes in the intensity of light, distorting the true nature of the data; and (2) any electrical

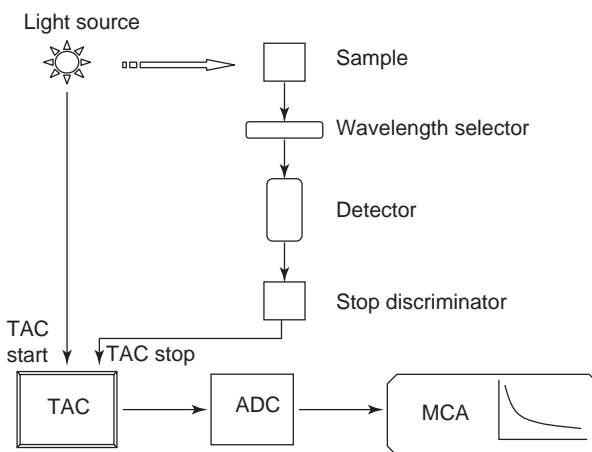


**Figure 2** Source and luminescence decay curves: 1, source time profile; 2, fluorescence decay curve; 3, true fluorescence decay curve. Assuming  $t_p \approx \tau_f$ , the initial part of the decay curve is distorted.

noise present in the instrumentation gets added to the signal, again distorting the data.

At present almost all time-domain measurements are performed using time-correlated single photon counting (TCSPC), an alternative analytical technique based on the ability to detect and count individual quanta, or photons, of light. The basics of single photon counting (SPC) can be understood by considering a block diagram of the instrumentation used (**Figure 3**). The light source delivers the excitation energy required to make the sample fluoresce. A pulsed light source, which generates an excitation pulse train, stimulates absorption of the sample molecule at repetition rates in the several kilohertz to several megahertz range, depending on the light source. Every time the light source sends a pulse of excitation energy, it also sends a start signal to the time-to-amplitude converter (TAC), a unit that measures the time elapsed from the start of the excitation pulse until a photon is detected. The sample is repetitively excited and each pulse is optically monitored by a detector, which is usually a high-speed photodiode or photomultiplier. The detector must have a very narrow entrance slit to ensure that only a single photon of light per 100 source pulses enters and is detected. For each photon collected, the photomultiplier generates an amplified signal which is sent to the stop discriminator, along with the signal due to the inherent photomultiplier noise (additive dark current).

The stop discriminator is an important element of the technique as it allows SPC to be essentially insensitive to the noise and drift problems which limit the analog approaches of light analysis. Basically, a discriminator is designed to differentiate between levels of electrical signals. For instance, if the input



**Figure 3** Basic components of a time-correlated single-photon counting system.

signal to the discriminator is below a specified threshold level, it produces an average output of zero; an input signal greater than the specified threshold level is recognized, thus producing an output pulse. By setting the discriminator level greater than the background signal noise but lower than the noise plus signal level, it is possible to eliminate noise and drift on the experiment.

The TAC may be considered as a stopwatch: the light source provides the start signal and the stop discriminator the stop timing. When the TAC receives the stop signal, it outputs a voltage pulse with an amplitude proportional to the elapsed time between starting and stopping. The events are collected in a memory location with an address proportional to the detection time. Thus, for one photon detected, we can have an accurate representation of just how long it has been since the sample was excited. A multichannel pulse height analyzer (MCA) converts the TAC voltage output to a time channel using an analog-to-digital converter (ADC). This process, which takes only a few microseconds from start to finish, is repeated until collecting more than 10 000 counts in the peak channel. Summing over many pulses, the MCA builds up a probability histogram of counts versus time channels, which represents the fluorescence decay of the sample (see Figure 4).

As was mentioned earlier, there can be no more than one photon detected per 50–100 source pulses. Electronics for SPC only allows detection of the first arriving photon. Once this photon is detected, the dead time in the electronics prevents detection of another photon resulting from the same excitation pulse. The apparent decay time becomes shorter as the number of arriving photons increases because the TAC is stopped by the first arriving photon. The pulse pileup problem is being solved to some extent by the development of dedicated electronic detection

systems and multiplexing methods for detection of multiple photons.

The time-resolved mode of measuring fluorescence lifetime is a digital or photon counting method that shows good signal-to-noise ratio. Advantages of time correlated SPC include a high dynamic range and the measured decay profile is independent of fluctuations in the excitation pulse intensity. Additionally, single photon detection theory is based upon well-documented statistics for which the precision, data weights, goodness-of-fit, and other parameters can be easily calculated.

## Time-Correlated Single Photon Counting Instrumentation

The main criterion for determining the overall system performance of the TCSPC system is the timing resolution and the sensitivity of the device. These characteristics vary from system to system and are largely based on the performance of the individual components. A schematic diagram of a simple TCSPC setup is shown in Figure 5. The instrumentation can be divided into three main groups of components:

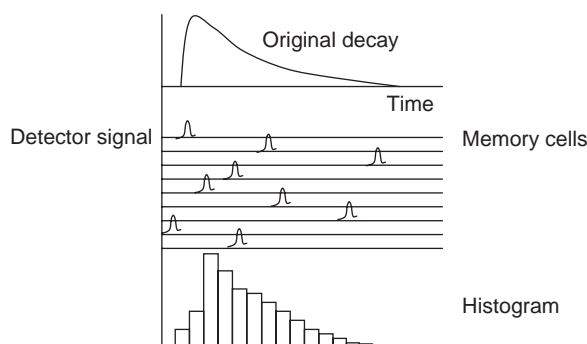
1. the pulsed excitation source;
2. optical spectrometer components (lenses, monochromators, etc.); and
3. single-photon detector and timing electronics (discriminators, TAC, MCA).

### Light Sources

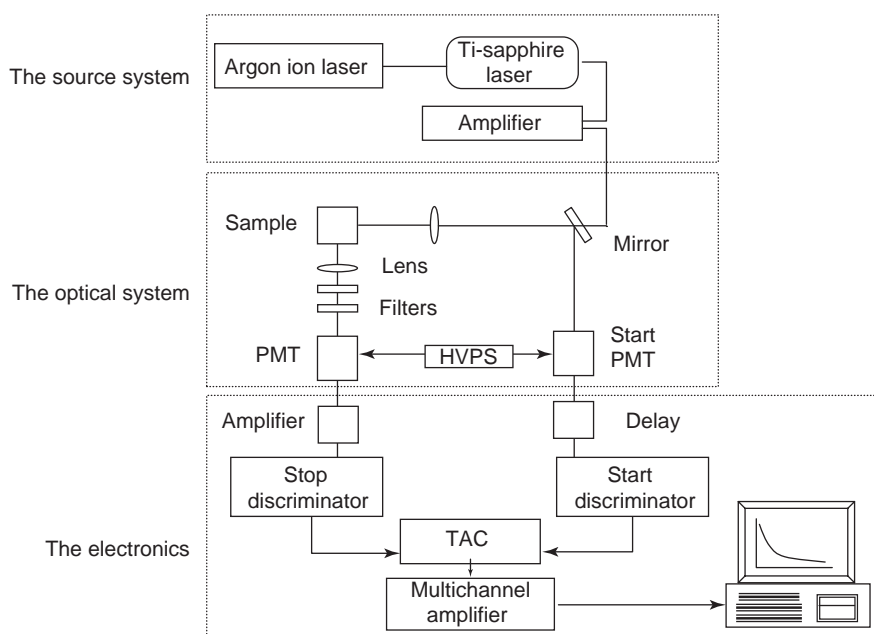
There are basically three choices for pulsed excitation in TCSPC measurements: flashlamps, synchrotron sources, and lasers. Flashlamps and lasers are the most commonly used (see Table 1).

At present, almost all the flashlamps used are gated and their spectral output depends on the gas within them. On the other hand, the pulse width typically depends on both the type of gas and the pressure (Table 1). The main drawback of using a flashlamp for TCSPC is the low repetition rate. The fastest flashlamps have repetition rates up to 100 kHz with 20 kHz being more common.

Lasers are now fast replacing flashlamps. The formation of a pulse for TCSPC by a laser is carried out by a process known as mode-locking, which produces pulses of picosecond duration or less. For a mode-locked laser, a periodic loss is introduced into the laser cavity by some modulator that acts as a shutter, opening once every round trip for the light in the resonator. Among mode-locked lasers, Ti:sapphire lasers pumped by an Argon ion laser is an ultrafast



**Figure 4** TCSPC measurement principle. The random nature of the events recorded yields a luminescence decay histogram. The memory cells are memory addresses to record a photon at a specified time interval.



**Figure 5** Basic TCSPC setup.

**Table 1** Light sources suitable for fluorescence lifetime measurements

Light source	Characteristics
Flashlamps	Hydrogen, deuterium: UV wide range, low intensity Nitrogen: high intensity Ar/H <sub>2</sub> mixtures: nanosecond pulse widths, output up to 750 nm Xenon: for long lifetime samples, UV–IR, intense Nanosecond pulsed xenon commercially available
Picosecond dye lasers	Nd:YAG: fundamental output at 1064 nm, 355 nm (third harmonic), 266 nm (fourth harmonic) Mode-locked argon-ion: output at 514 nm
Femtosecond lasers	Nitrogen/dye: compact and reliable, output at 337 nm, pulse widths between 300 ps and 10 ns Ti:sapphire: solid-state device, long operational life, long-wavelength output 720–920 nm
Solid-state lasers	Simpler, low cost, 600–700 nm (limited range)
Ultrabright LEDs	Nanosecond pulsed, 1–200 MHz, very stable in operation, UV (370–390 nm), blue (470 nm), green (530 nm)

laser system that uses a Ti:sapphire crystal as the gain medium for the near-infrared (NIR) and is tunable from 720 to 990 nm. The Ti:sapphire laser is a solid-state device, with a repetition rate of 76 MHz with 100–150 fs pulses and an extremely long operational life.

Semiconductor diode lasers offer the advantages of low cost, relatively high powers, stable output, and long lifetimes (>40 000 h). These devices, which use a semiconductor material (e.g., GaAlAs) as the gain medium for light production, have also been shown to be adequate excitation sources for TCSPC measurements. A drawback of diode lasers is that the beam-shape is elliptical due to the longitudinal separation between the diode laser's emission points

parallel and perpendicular to the p–n junction. The wavelengths are usually limited to 600–700 nm (some can be frequency-doubled to 410 nm), typical repetition rates up to 80 MHz and pulses of 100 ps.

### Detector Devices

The detector is perhaps the most critical component of the instrumentation. Ultimately, it is its characteristics that dictate the overall timing response of the TCSPC instrument. The main requirements for an ideal SPC detector are:

1. Fast response time, typically <1.5 ns rise time;
2. High gain, typically in excess of  $10^6$ ;



3. Low noise, dominated by dark count and ideally  $<100$  cps;
4. Negligible temporal response dependence on wavelength, ideally  $<0.5$  ps nm $^{-1}$ ;
5. Low detector-generated artifacts;
6. Low transit time jitter across the device's detecting area, ideally  $<20$  ps mm $^{-1}$ ; and
7. Sensitive area compatible with other optical components.

Some of these requirements conflict with each other, such as in the loss of noise performance as the gain is increased. As result, the sensitivity of the TCSPC method is mainly limited by the dark count of the detector. Not all the detectors fulfill the above requirements. Photomultiplier tubes (PMTs) are the most popular type of detector employed in fluorescence lifetime determination by TCSPC. There are two principal types of PMTs used in SPC: the side-on PMT and the linear focused PMT. They operate on the principle of light interacting with either an opaque or semitransparent photocathode liberating photoelectrons which are then multiplied by a dynode structure to provide an electrical output. Each of the PMT geometries has its own advantages and drawbacks, the linear focused type being the more widely used one. In general terms, PMTs exhibit gains  $\sim 10^7$ , fast rise time followed by a rapid and smooth return to the baseline, and low dark counting rate (which is enhanced by cooling).

Defining the sensitivity  $S$  as the intensity at which the signal equals the noise of the dark signal, the following equation applies:

$$S = \frac{(Dkc \times N/T)^{1/2}}{q} \quad [10]$$

where  $Dkc$  is the dark count rate,  $N$  the number of time channels,  $q$  the quantum efficiency of the detector, and  $T$  the overall measurement time. Typical values for a PMT with multialkali cathode without cooling, are  $Dkc = 300$  s $^{-1}$ ,  $N = 260$ ,  $q = 0.1$ , and  $T = 100$  s. This yields a sensitivity of  $S = 280$  photons s $^{-1}$ , a value which is  $10^{15}$  smaller than the intensity of a typical laser ( $10^{18}$  photons s $^{-1}$ ). Then, when a sample is excited by a laser and the emitted light is measured, the emission is still detectable for a conversion efficiency of  $10^{-15}$ .

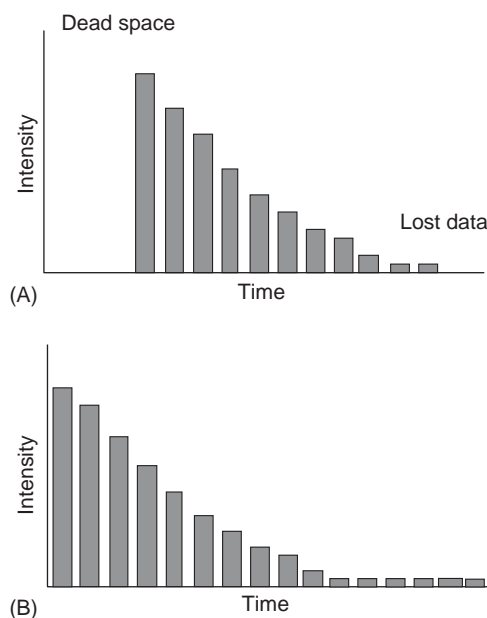
There are also two detector devices that are gaining popularity for fluorescence lifetime studies due to their fast time response: the microchannel plate photodetectors and streak cameras. Microchannel plate photodetectors, similar to PMTs, are based on the use of multiplication of photoelectrons. Instead of using discrete dynodes, continuous semiconductor-coated glass multiplier tubes of  $\sim 10$   $\mu$ m diameter

are employed. An electric field is applied across the faces of the microchannel plate so that electrons entering the multiplier tubes are accelerated and, on striking the walls, produce secondary electrons. These are then accelerated and cause more electrons to become liberated thus producing electron amplification. These electrons are collected by the anode to deliver an output pulse as in a conventional PMT. Microchannel plate detectors offer extremely fast response at high quantum efficiency throughout UV and visible range. They are best used with laser sources which have a more comparable time profile while conventional PMTs can be used effectively with flashlamp excitation sources.

The photoelectronic streak camera is capable of measuring time dependence of electromagnetic radiation from X-ray to NIR at  $10^{-7}$ – $10^{-13}$  ps resolution. Basically, its operation is based on the conversion of the emitted photon into a photoelectron. The photoelectron beam is then spatially dispersed on a phosphor by a synchronized deflection voltage applied across two plates. In contrast to other detectors, streak cameras provide the additional features of data timing and acquisition. However, their main limitation is the relatively low dynamic range and the high cost. Streak cameras can provide simultaneous measurements of both wavelength and time-resolved decays. Such data are particularly valuable in those situations where very short time resolution is required or in the study of samples that contain fluorophores emitting at different wavelengths.

On examining **Figure 5**, it is observed that the fluorescence channel looks very much like the theoretical system discussed above (**Figure 3**), while the excitation channel is significantly different. Light from the pulsed source is detected by the start photomultiplier, which eventually generate the start pulse. The start signal is then delayed – or slowed down – to prevent it from starting the TAC until some point in time later than it would if the delay device were not present. This delay is necessary because the total light travel time is significantly shorter in the excitation channel than it is in the fluorescence channel. In the absence of the delay device, there is a void space at the low edge of the fluorescence decay profile and a loss of data at the high end. Adding the delay device and adjusting it to eliminate the void space, the lost data at the high end are now recovered (see **Figure 6**). After being delayed, the excitation pulse is sensed by the start discriminator. By setting its high threshold level above the noise level of the detector, the valid output signal will start the TAC. In the fluorescence channel, the same components are used, except for the delay device, and





**Figure 6** Effect of the delay module: (A) no delay module, (B) correctly adjusted start delay.

comments concerning their functions apply in this case as well.

Discriminators are used to aid in removing random noise pulses resulting from background photons, to ensure that the timing definition of the start and stop pulses is largely independent of the signal pulse height and to improve the signal-to-noise ratio. There are two types of discriminators: leading edge (LE) and constant fraction (CF) discriminators. The CF is particularly suited to SPC studies compared to LE as with the LE photodetector pulses of different heights will not cross the discriminator level at the same time, introducing errors into the measurement.

As described above, the TAC functions to determine the time interval between the excitation pulse and the subsequent fluorescence photon arriving at the detector. The MCA consists of an ADC, a memory consisting of channels for storing data, and data input and output facilities. A standard instrument incorporates lower and upper discriminator levels and two modes of data collection: pulse height analysis mode for displaying fluorescence decay profiles ( $\geq 1000$  channels) and multichannel scaling mode to bin the data into given time increments. Data are usually displayed on an oscilloscope or on a computer terminal.

An important aspect associated with lifetime determinations is related with the calculation algorithms or data processing to extract the lifetime value from the resulting decay. There are two algorithms

that are commonly used. The rapid determination method estimates the fluorescence lifetime by integrating the number of counts within the decay profile over a specified time interval from:

$$\tau_f = -\frac{\Delta t}{\ln(D_1/D_0)} \quad [11]$$

where  $\Delta t$  is the time interval over which the counts are integrated,  $D_0$  represents the integrated counts in the early time interval, and  $D_1$  is the integrated number of counts in the later time interval of the decay spectrum. This method can extract only a single lifetime value, which in cases of multiexponential profiles represents an average value of the different components comprising the decay.

Nonlinear least squares method allows to compare the measured data  $N(t_k)$  with values predicted  $N_c(t_k)$  from a model and the parameters of the model to be varied to yield the minimum deviation from the data through minimization of the goodness-of-fit parameter  $\chi^2$ , calculated from

$$\begin{aligned} \chi^2 &= \sum_{k=1}^n 1/\sigma_k^2 [N(t_k) - N_c(t_k)]^2 \\ &= \sum_{k=1}^n [N(t_k) - N_c(t_k)]^2 / N(t_k) \end{aligned} \quad [12]$$

where the sum extends over the number of channels ( $n$ ) or data points used in the analysis and  $\sigma_k$  is the standard deviation associated with each data point. Nonlinear least squares method can analyze multiexponential decays and has the capability to remove the contribution of the instrument response function through deconvolution.

## Phase-Resolved Measurements

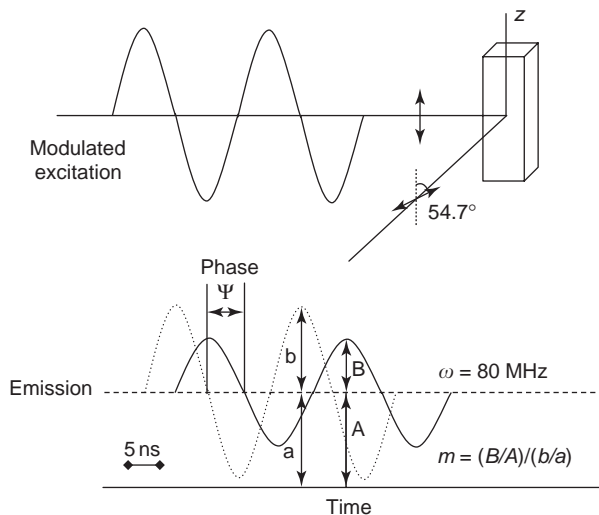
Phase-resolved, phase-modulation, or phase-sensitive lifetime measurements are based on the use of a continuous, sinusoidally modulated excitation source and phase-sensitive detection (Figure 7). The experimentally measured parameters are the modulation ( $m$ ) and the frequency-dependent phase shift ( $\Psi$ ). The modulation of the excitation is given by  $b/a$ , where  $a$  is the average intensity and  $b$  is the modulated amplitude of the incident light. For emission, the modulation is similarly defined, except using the intensities of the emission,  $B/A$ , relative to the modulation of the excitation,  $m = (B/A)/(b/a)$ . The phase delay or phase angle ( $\Psi$ ) is usually measured from the zero-crossing time of the modulated components. For an exponential decay, the fluorescence lifetime  $\tau_f$  can be calculated from the phase shift or

demodulation factor as follows:

$$\tau_f = \tan \Psi / \omega \quad [13]$$

$$\tau_f = (m^{-2} - 1)^{1/2} / \omega \quad [14]$$

where  $\omega$  is the angular modulation frequency ( $\omega = 2\pi f$ ,  $f$  = linear modulation frequency in hertz).



**Figure 7** Phase-modulation technique. The ratios  $B/A$  and  $b/a$  represent the modulation of the emission and of the excitation, respectively. (Reproduced with permission from Lakowicz TR (1999) *Principles of Fluorescence Spectroscopy*, 2nd edn., ch. 4, p. 96. New York: Kluwer Academic; © Kluwer Academic.)

Both  $\Psi$  and  $m$  are measured relative to the signal from a scattering solution or a reference fluorophore of known lifetime (see Figure 8). In the former case,  $m$  is the ratio of the modulation depth of the AC waveform of the fluorophore to that for a scattering solution. The reference signal  $I_r(t)$  that tracks the time-dependent excitation radiance is obtained by directing part of the excitation beam to the scattering solution and monitoring the scattering signal. In this case

$$I_r(t) = (I_r)_{DC}(1 + d_m \sin \omega t) \quad [15]$$

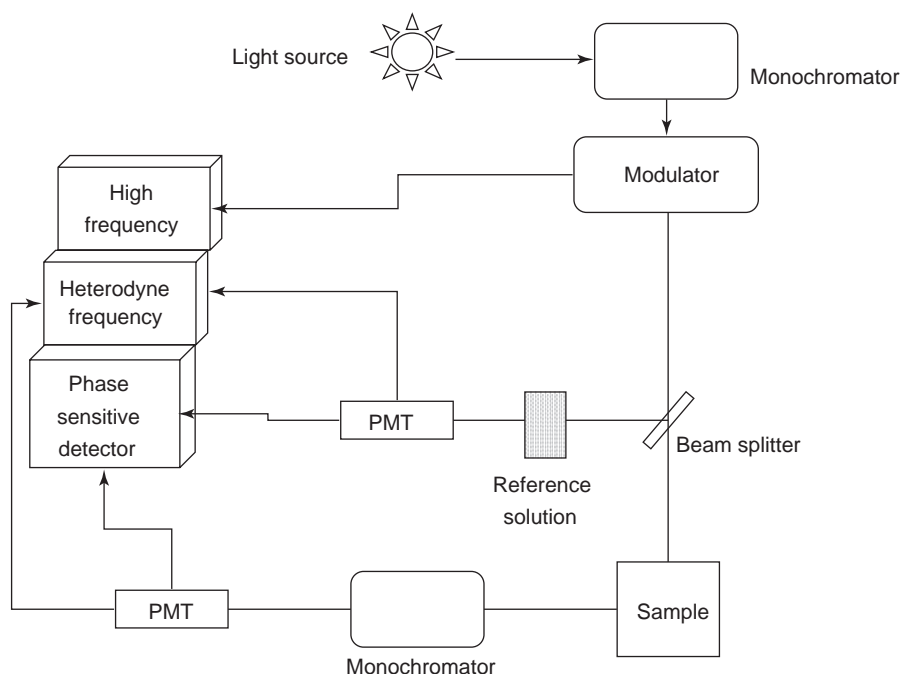
where  $(I_r)_{DC}$  is the magnitude of the DC component,  $d_m$  is the modulation depth of the AC component (i.e., the ratio of the AC to DC amplitude of the excitation radiance). If the modulation period is on the order or smaller than the fluorescence lifetime, the resulting fluorescence signal is demodulated and phase shifted as shown in Figure 7.

The time-dependent fluorescence signal  $I_f(t)$  is described by

$$I_f(t) = (I_f)_{DC}[1 + d_m m \sin(\omega t - \Psi)] \quad [16]$$

where  $(I_f)_{DC}$  is the DC component of the fluorescence signal.

The values of  $\Psi$  and  $m$  can also be compared by nonlinear regression to the values predicted from an assumed decay law. So, for any time-domain decay law (eqn [7]), the frequency-domain data are related



**Figure 8** Block diagram of a phase fluorimeter with phase-sensitive detection.

by the sine and cosine Fourier transforms

$$N_{\omega} = \frac{\int_{-\infty}^0 I(t) \sin \omega t \, dt}{\int_{-\infty}^0 I(t) \, dt} \quad [17]$$

$$D_{\omega} = \frac{\int_{-\infty}^0 I(t) \cos \omega t \, dt}{\int_{-\infty}^0 I(t) \, dt} \quad [18]$$

These are related to the experimentally measured parameters through the expressions:

$$\tan[\Psi(\omega)] = [N(\omega)/D(\omega)] \quad [19]$$

$$[m(\omega)] = [N^2(\omega) + D^2(\omega)]^{1/2} \quad [20]$$

The decay terms ( $\alpha_i$  and  $\tau_i$ ) are recovered by minimization of the  $\chi^2$  function

$$\begin{aligned} \chi^2 = 1/\nu \sum_{\omega} [\Psi_{\omega} - \Psi_{c\omega}/\delta\Psi]^2 \\ + 1/\nu \sum_{\omega} [m_{\omega} - m_{c\omega}/\delta m]^2 \end{aligned} \quad [21]$$

where  $\nu$  is the number of degrees of freedom, and  $\delta\Psi$  and  $\delta m$  are the uncertainties in the measured phase and modulation, respectively. The subscript c denotes the frequency-dependent phase and modulation calculated on the basis of  $\alpha_i$  and  $\tau_i$ .

The excitation beam (He–Cd laser, laser diode, LED) is usually modulated with an electro-optic modulator and modulation frequencies vary typically from 10 to 200 MHz. The modulated fluorescence signal is extracted by synchronous detection with a lock-in amplifier. The detector of choice in frequency-domain fluorescence is a PMT. In order to process the signals, the PMT bias voltage and hence the PMT gain is modulated at frequency  $\omega + \Delta\omega$  to produce a heterodyne signal at  $\Delta\omega$ . In this way, high-frequency signals are shifted to low-frequency signals, which are easier to process. Resolution as good as 1–100 ps is possible.

## Analytical Applications

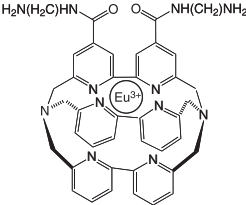
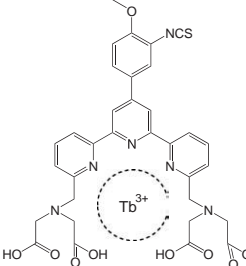
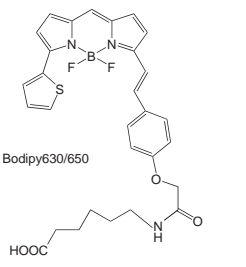
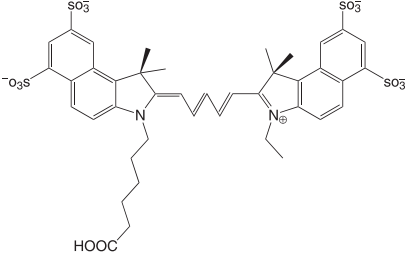
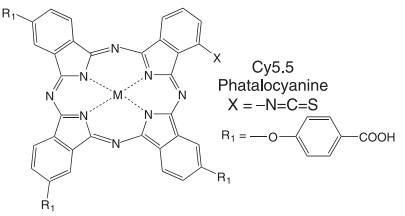
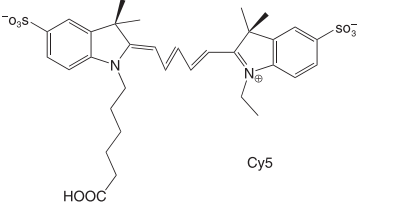
Fluorescence is unique among spectroscopic techniques due to its inherently multidimensional nature, the emission process containing a wealth of orthogonal information that is related to the fluorophore and its surroundings. Time- and frequency-domain fluorescence methods are instrumentally sophisticated, but they improve both the sensitivity and selectivity of fluorimetry. Any dye that is used for steady-state fluorescence detection can be used for time-resolved detection as well. Most of these fluorophores display lifetimes from 1 to 10 ns, which requires fast electronics for time-domain lifetime measurements or modulation frequencies from 10

to 100 MHz for frequency-domain measurements. However, fluorophores with considerably longer lifetimes have been used in assays based on time-resolved fluorescence techniques. For instance, fluorescent lanthanoid chelates have some unique features such as long lifetime sometimes exceeding 1 ms, large Stokes shifts ( $>250$  nm) and very sharp peak profiles with full-width at half-maximum of  $\sim 10$  nm. These fluorescent chelates can be bound to proteins, nucleic acids, and other molecules. Also, some NIR fluorescent tricarbo-cyanine dyes can be used in time-resolved identification of multiple dyes in capillary electrophoresis applications. These dyes show absorption/emission maxima  $\sim 765$ – $794$  nm with lifetimes ranging from 947 to 847 ps. The attractive feature associated with fluorescence in the NIR is the low backgrounds observed during signal collection and the simple instrumentation required for ultrasensitive detection. On the other hand, by incorporation of a heavy atom group into the dye structure the fluorescence lifetime is altered. This is due to the intramolecular heavy-atom effect and internal conversion that result in an increase in  $\tau_f$  with the increasing molecular weight of the substituent group. In **Table 2**, the structure of some chelators and the fluorescent properties of some long-lived fluorophores are shown.

Time- and frequency-domain fluorescence methods have been used in different fields for several purposes, such as distinguishing sample constituents whose fluorescence spectra overlap one another, to distinguish the fluorescence of an analyte from background scattering or luminescence of other sample constituents, in combination with fluorescence resonance energy transfer (FRET) assays to combine the benefits of FRET and time-resolved fluorescence. A listing of some selected applications is given in **Table 3**.

The future of fluorescence spectrometry is indeed promising. The study of liquid/liquid, liquid/solid, and solid/liquid interfaces is of critical importance to protein adsorption, bioadhesion, biosensors, and separation sciences. The analytical potential of time- and frequency-domain techniques in this area is waiting to be explored. The advent of high power pulsed lasers, fast electronics, electro-optics and/or frequency upconversion techniques may allow to study femtosecond processes, such as solvent relaxation and in high-speed problems, for instance, in genome sequencing and capillary separations there is a push to achieve single molecule detection. Today it is possible to obtain time- and frequency-domain decay traces on the fly in a time frame substantially less than 1 s. At present, many fluorescence experiments in these challenging fields may become almost prohibitive. The key factor is to bring down the cost

**Table 2** Long-lived fluorophores

Fluorophore	$\lambda_{exc}$ (nm)	$\lambda_{em}$ (nm)	$\tau_f$ (ns)
 <p><math>H_2N(H_2C)HN-C(=O)-</math> and <math>-C(=O)-NH(CH_2)NH_2</math></p>	337	620, 665	300 <sup>a</sup>
 <p><math>Tb^{3+}</math></p>	325	543	2681 <sup>a</sup> 1353 <sup>a,b</sup>
 <p>Bodipy630/650</p>	629	646	3.89
 <p><math>SO_3^-</math>, <math>HOOC</math></p>	675	692	0.83
 <p>Cy5.5 Phthalocyanine <math>X = -N=C=S</math> <math>R_1 = -O-C_6H_4-COOH</math></p>	683	700	1.5–4 (depending on the central metal ion)
 <p><math>SO_3^-</math>, <math>HOOC</math>, Cy5</p>	647	663	0.91

<sup>a</sup> Milliseconds.

<sup>b</sup> Europium chelate.

**Table 3** Selected applications of time- and frequency-domain fluorescence

Field	Information/analytes	Comments
Photochemistry, photophysics	Solvent–solute interactions, dynamics of micelles, polymer structure and dynamics, characterization of excited states, excimer and exciplex formation	Capability to provide information on rapid phenomena
Multicomponent analysis	Polycyclic aromatic hydrocarbons, environmental analysis, petrochemical analysis	Solving mixtures of spectrally overlapped lumiphores
HPLC detection	Estrogens	Lanthanoid chelate fluorescent labels, detection limits in the picogram range
Clinical analysis, immunoassays	Proteins, hormones, herbicides, drugs	Sensitivity is enhanced combined with fluorescence resonance energy transfer, homogeneous assays
Fluorescence lifetime imaging	Study of single cells, bulk tissue, dynamics of intercellular and intracellular processes, bimolecular interactions	Noninvasive technique, highly sensitive
Nucleic acids	Dynamics and structure, hybridization assays, DNA polymerase reactions, detection limits about picomolar level	Microanalysis of sample arrays, high-throughput screening
Single molecule detection	Counting molecules, rare event detection order of events, probe/target binding, distribution functions (e.g., number of base pairs for DNA sizing), molecular dynamics	Sensitivity is enhanced combined with fluorescence resonance energy transfer
Biochemistry/medicine	Protein structure and folding, dynamics and structure of membranes, Donor–acceptor distances, photochemistry of vision	Background reduction Fluorescence lifetime allows the identification of a single molecule Limitations in sensitivity due to natural prompt fluorescence are reduced

of this technology and to make the instrumentation more reliable.

See also: **Fluorescence**: Overview; Multidimensional Fluorescence Spectrometry; High-Resolution Techniques; Fluorescence Labeling.

## Further Reading

Bazin H, Prédauat M, Trinquet E, and Mathis G (2001) Homogeneous time resolved fluorescence and resonance energy transfer using rare earth cryptates as a tool for

probing molecular interactions in biology. *Spectrochimica Acta* 57: 2197–2211.

Lakowicz JR (ed.) (1999) *Principles of Fluorescence Spectroscopy*, 2nd edn. Dordrecht: Kluwer Academic.

Hungerford G and Birch DJS (1996) Single-photon timing detectors for fluorescence lifetime spectroscopy. *Measurement Science and Technology* 7: 121–135.

<http://cfs.umbi.umd.edu/cfs/software/download.htm>.

Time- and/or frequency-domain analysis programs are available at this address.

Siegel J, Elson DS, Webb SED, *et al.* (2003) Studying biological tissue with fluorescence lifetime imaging: Microscopy, endoscopy and complex decay profiles. *Applied Optics* 42(16): 2995–3004.

## Derivatization

**M E Díaz-García and R Badía-Laíño**, University of Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Fluorimetric analytical methods generally have detection limits from one to four orders of magnitude

lower than the corresponding methods based on absorption. A method to further improve detection limits along with increasing selectivity is by fluorescence derivatization procedures through which weakly fluorescent or nonfluorescent species are converted into highly luminescent products. Also, derivatization procedures are important for the determination of substances that tend to decompose during the analysis or for those substances that do

**Table 3** Selected applications of time- and frequency-domain fluorescence

Field	Information/analytes	Comments
Photochemistry, photophysics	Solvent–solute interactions, dynamics of micelles, polymer structure and dynamics, characterization of excited states, excimer and exciplex formation	Capability to provide information on rapid phenomena
Multicomponent analysis	Polycyclic aromatic hydrocarbons, environmental analysis, petrochemical analysis	Solving mixtures of spectrally overlapped lumiphores
HPLC detection	Estrogens	Lanthanoid chelate fluorescent labels, detection limits in the picogram range
Clinical analysis, immunoassays	Proteins, hormones, herbicides, drugs	Sensitivity is enhanced combined with fluorescence resonance energy transfer, homogeneous assays
Fluorescence lifetime imaging	Study of single cells, bulk tissue, dynamics of intercellular and intracellular processes, bimolecular interactions	Noninvasive technique, highly sensitive
Nucleic acids	Dynamics and structure, hybridization assays, DNA polymerase reactions, detection limits about picomolar level	Microanalysis of sample arrays, high-throughput screening
Single molecule detection	Counting molecules, rare event detection order of events, probe/target binding, distribution functions (e.g., number of base pairs for DNA sizing), molecular dynamics	Sensitivity is enhanced combined with fluorescence resonance energy transfer
Biochemistry/medicine	Protein structure and folding, dynamics and structure of membranes, Donor–acceptor distances, photochemistry of vision	Background reduction Fluorescence lifetime allows the identification of a single molecule Limitations in sensitivity due to natural prompt fluorescence are reduced

of this technology and to make the instrumentation more reliable.

See also: **Fluorescence**: Overview; Multidimensional Fluorescence Spectrometry; High-Resolution Techniques; Fluorescence Labeling.

## Further Reading

Bazin H, Prédauat M, Trinquet E, and Mathis G (2001) Homogeneous time resolved fluorescence and resonance energy transfer using rare earth cryptates as a tool for

probing molecular interactions in biology. *Spectrochimica Acta* 57: 2197–2211.

Lakowicz JR (ed.) (1999) *Principles of Fluorescence Spectroscopy*, 2nd edn. Dordrecht: Kluwer Academic.

Hungerford G and Birch DJS (1996) Single-photon timing detectors for fluorescence lifetime spectroscopy. *Measurement Science and Technology* 7: 121–135.

<http://cfs.umbi.umd.edu/cfs/software/download.htm>.

Time- and/or frequency-domain analysis programs are available at this address.

Siegel J, Elson DS, Webb SED, *et al.* (2003) Studying biological tissue with fluorescence lifetime imaging: Microscopy, endoscopy and complex decay profiles. *Applied Optics* 42(16): 2995–3004.

## Derivatization

**M E Díaz-García and R Badía-Laíño**, University of Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Fluorimetric analytical methods generally have detection limits from one to four orders of magnitude

lower than the corresponding methods based on absorption. A method to further improve detection limits along with increasing selectivity is by fluorescence derivatization procedures through which weakly fluorescent or nonfluorescent species are converted into highly luminescent products. Also, derivatization procedures are important for the determination of substances that tend to decompose during the analysis or for those substances that do



not exhibit the desired properties for a selected transduction scheme. Thus, fluorescence derivatization includes all those procedures that involve reacting the species of interest to yield a fluorescent product. An ideal derivatizing reaction should fulfill some specific analytical criteria that depend on the intended application. In general, desirable properties for fluorescence derivatizing reactions include the following:

1. Fast under mild conditions, quantitative, specific for the compound of interest and simple to perform.
2. The resulting derivative should be stable, not undergo rapid decomposition, and no occurrence of side reactions. These characteristics are especially important in chromatographic precolumn derivatization reactions because side-products may cause interfering peaks in the chromatogram.
3. Measurable differences in spectroscopic properties between the label and its reaction product.
4. The resultant derivative should exhibit reasonably good Stokes' shift and emit fluorescence at long excitation and emission wavelengths in order to avoid background fluorescence from substances present in the sample.
5. The fluorescence efficiency of the derivative should be high for maximum sensitivity.
6. In the analysis of biological samples, it is desirable that the derivatization reagent is water soluble or miscible in aqueous solvents.
7. When used in combination with chromatographic systems, the separation of the resulting derivative and the label must be relatively easy to perform.

Various reactions may be used to convert nonfluorescent species into fluorescent ones. The main advantages as well as the drawbacks of each method will be overviewed. The scope of this article covers those reaction systems not considered elsewhere in this encyclopedia.

## Survey of Derivatization Reactions

Fluorescent derivatization reactions may be classified into seven groups according to the type of reaction involved.

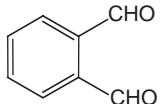
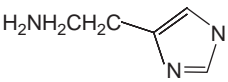
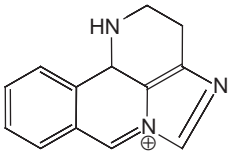
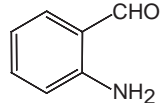
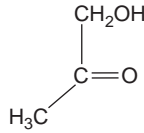
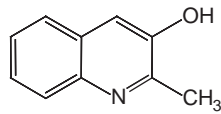
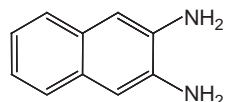
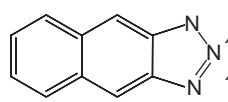
### Reactions That Lead to the Formation of an Extended Aromatic $\pi$ -Electron System

Relatively few aliphatic or alicyclic compounds show analytically useful fluorescence in the ultraviolet (UV) or visible region. On the contrary, efficient fluorescence occurs in molecules having extended

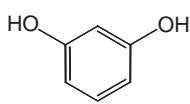
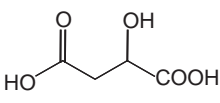
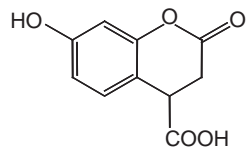
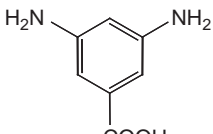
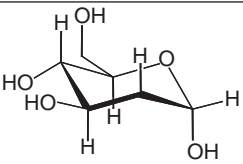
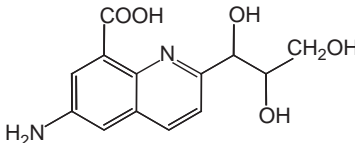
$\pi$ -electron systems. Generally, the efficiency of the fluorescence increases with the rigidity of the molecular structure as well as with the planarity of the structure, both of which are considered to be requirements for effective  $\pi$ -electron delocalization. A practical consequence of these facts for fluorescence derivatization is that most important fluorescence derivatization reactions involve the formation of compounds with at least two condensed aromatic rings. Representative examples are given in Figures 1–3; they are classified according to the manner in which the aromatic  $\pi$ -electron system is formed. Thus, in Figure 1, the reaction of aromatic hydrocarbon compounds with two adjacent substituents, both of which combine with the analyte, produce a new ring. Another method to extend the  $\pi$ -electron system is by coupling the analyte to a substituent and to an unsubstituted position on an aromatic ring (Figure 2). Substituted aromatic compounds may be the analyte, giving rise to reactions based on the same principles (Figure 3). In general, these kinds of reactions are performed in hard conditions (strong acid, elevated temperatures, long reaction times) and, as a result, the structure of the fluorescent product is, in many cases, unknown. Most of the reactions result in the formation of heterocyclic compounds, for which the lowest energy electronic transition is  $n, \pi^*$ . Accordingly, we can expect such compounds to show low fluorescence efficiency. However, the presence of electron-donating groups and the extended  $\pi$ -systems minimizes the influence of the lone pair on the heteroatom, and fluorescence from the lowest singlet state  $\pi, \pi^*$ , rather than  $n, \pi^*$ , may be observed. The formation of the extended  $\pi$ -system is accompanied by a red-shift in the fluorescence band with respect to that of the reagents, which eliminates the need for the removal of residual fluorescent reagents. This fact is directly translatable into lower background signals. Many of these reactions are not specific for an analyte, but the selectivity for a group of compounds is rather good.

### Oxidation/Reduction

Fluorescence derivatization reactions that involve oxidation and/or reduction steps have been employed for determining compounds of biological and environmental interest. Common oxidizing agents include hydrogen peroxide, potassium permanganate, cyanogen bromide, and strong mineral acids, depending on the specific application. Typical examples include the fluorimetric determination of cholesterol, based on the well-known Liebermann–Bouchard reaction, in which cholesterol is reacted with chloroform and acetic anhydride in concentrated sulfuric acid. The method is not specific and

Reagent	Analytes	Reaction conditions	Fluorescent product
 O-Phthalaldehyde	 Histamine, indoles, peptides, proteins	Basic	 $\lambda_{\text{exc}} = 360 \text{ nm}$ $\lambda_{\text{em}} = 490 \text{ nm}$
 O-Amino-benzaldehyde	 Carbohydrates	KOH/boiled	 Blue emission
 2, 3-Diamino-naphthalene	$\text{NO}_2^-$ Nitrite, nitrate	HCl/extraction	 $\lambda_{\text{exc}} = 355 \text{ nm}$ $\lambda_{\text{em}} = 410 \text{ nm}$

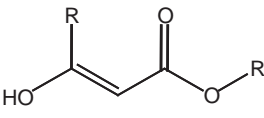
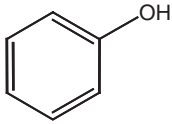
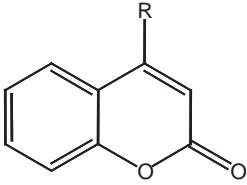
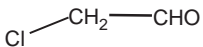
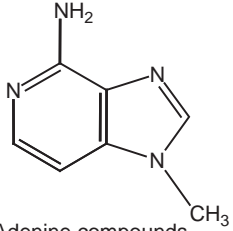
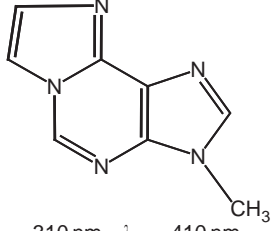
**Figure 1** Formation of extended aromatic  $\pi$ -electron systems for nonaromatic analytes.

Reagent	Analytes	Reaction conditions	Fluorescent product
 Resorcinol	 Keto-hexoses, ketoacids, dicarboxylic acid	$\text{H}_2\text{SO}_4$ 50% 20 min 60°C	 $\lambda_{\text{exc}} = 430 \text{ nm}$ $\lambda_{\text{em}} = 480 \text{ nm}$
 3, 5-Diamdiamo-benzoic acid	 2-Deoxy-D-glucose Acetaldehyde DNA	$\text{H}_3\text{PO}_4$ 15 min 100°C	 $\lambda_{\text{exc}} = 410 \text{ nm}$ $\lambda_{\text{em}} = 500 \text{ nm}$

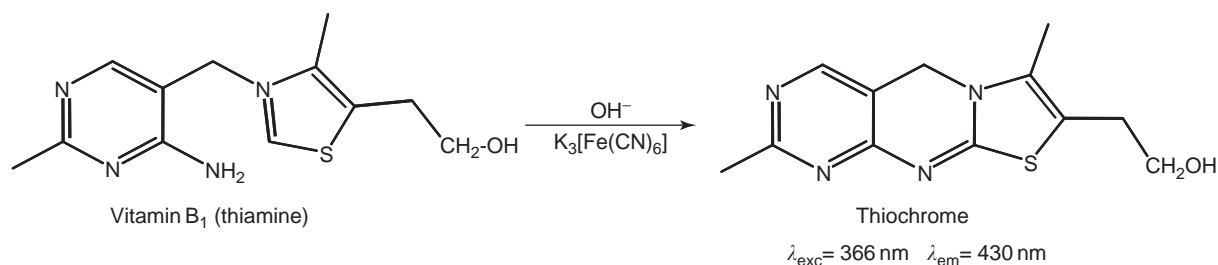
**Figure 2** Extended aromatic  $\pi$ -electron systems by coupling to an unsubstituted position of an aromatic ring.

other steroids also give a similar reaction. This procedure has been applied for the determination of cholesterol in serum samples. Vitamin B<sub>1</sub> (thiamine) is also easily decomposed by oxidizing agents

(ferricyanide, cyanogen bromide, mercuric oxide) and by UV light. The procedure involves the oxidation of thiamine to form the fluorescent thiochrome fluorophore (Figure 4).

Reagent	Analytes	Reaction conditions	Fluorescent product
 Ethyl acetoacetate	 Phenols	H <sub>2</sub> SO <sub>4</sub>	 $\lambda_{\text{exc}} = 366 \text{ nm}$ $\lambda_{\text{em}} = 460 \text{ nm}$
 Chloro-acetaldehyde	 Adenine compounds	pH 6 30 min 100°C	 $\lambda_{\text{exc}} = 310 \text{ nm}$ $\lambda_{\text{em}} = 410 \text{ nm}$

**Figure 3** Derivatization reactions based on extended aromatic  $\pi$ -electron systems for substituted aromatic analytes.

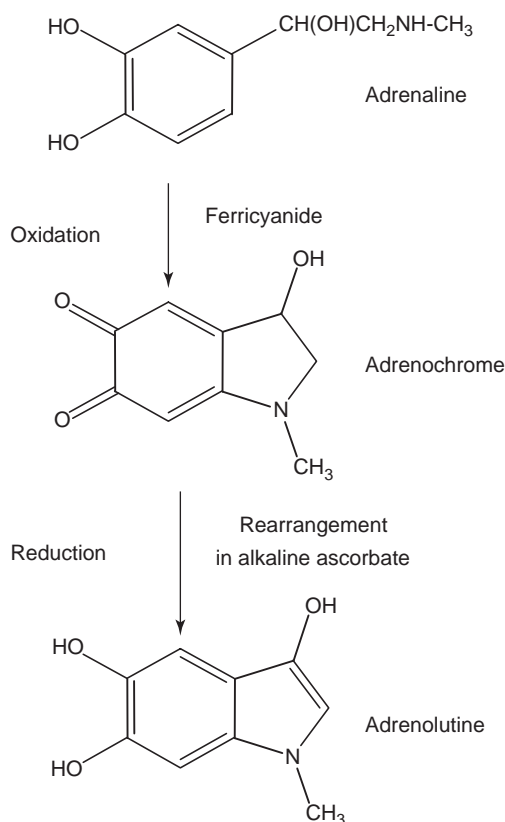


**Figure 4** Derivatization of vitamin B<sub>1</sub> by reaction with oxidizing agents.

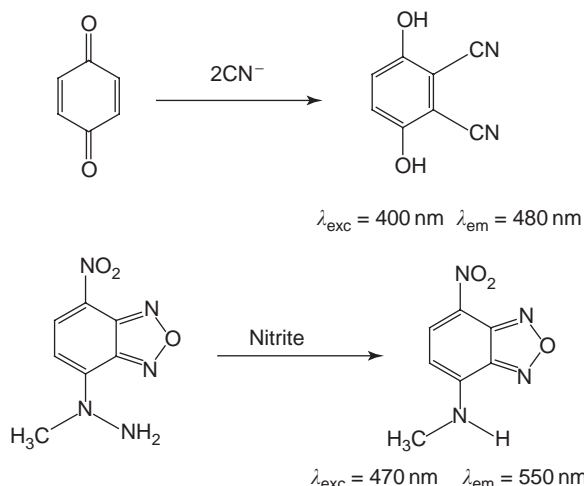
The direct fluorimetric determination of catecholamines based on their native fluorescence (excitation at 275 nm, emission at 318 nm) lacks sensitivity and is susceptible to chemical interference from biological sample matrices (serum, urine). Therefore, in order to improve sensitivity and selectivity, catecholamines can be derivatized by oxidation with ferricyanide or iodine followed by addition of an alkali containing an antioxidant. The alkali causes the rearrangement of the oxidized compounds to yield 3,5,6-trihydroxyindole, an intensely fluorescent product (**Figure 5**). This is one of the most sensitive and widely used methods for derivatizing catecholamines and has been used as postcolumn detection mode after chromatographic separation of the amines. Detection limits in the picomole range have been reported. In some instances, the oxidation is followed by a condensation with an adequate reagent

in order to extend the  $\pi$ -electron system. For example, after oxidation of catecholamines with ferricyanide, condensation with ethylenediamine results in a fluorescent product that shows maximal excitation at 420 nm and maximal fluorescence at 490 and 520 nm.

Anion fluorimetric assays may also be developed through redox reactions. For instance, cyanide ions react with *p*-benzoquinone to form a highly fluorescent derivative, with the cyanide ion acting as a reducing agent. Other anions such as thiocyanate and sulfide do not interfere and a sensitivity range of 0.2–50  $\mu\text{g ml}^{-1}$  has been reported. In a same way, nitrite converts quantitatively a nonfluorescent reagent, 7-amine-methyl-amino-4-nitrobenzoxadiazole, into a highly fluorescent product. Nanomolar concentrations of nitrite can be determined in water samples (see **Figure 6**).



**Figure 5** Fluorescence derivatization of catecholamines by oxidation.



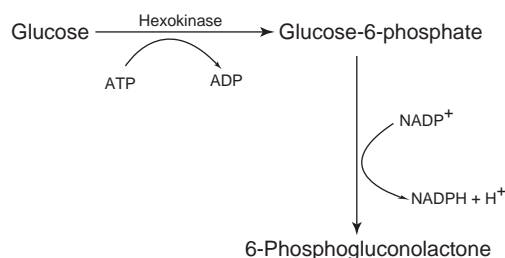
**Figure 6** Anion fluorimetric derivatization through redox reactions.

It is also feasible to use enzymatic assays as an alternative to chemical oxidation/reduction. Many of these enzymatic assays are based on the fluorescence of the reduced form or the oxidized form of the

enzyme cofactor. For example, NADH and NADPH are highly fluorescent, while the oxidized form of NADH and  $\text{NAD}^+$  is nonfluorescent. The excitation and emission wavelengths commonly used are 340 and 450 nm, respectively. The cofactor pyridoxal phosphate is also fluorescent and its absorption and emission spectra are dependent on its chemical structure in the protein. Ribo flavin, flavin mononucleotide, and flavin adenine dinucleotide absorb light in the visible range ( $\sim 450 \text{ nm}$ ) and emit around 525 nm and, in contrast to NADH, the oxidized forms of flavins are fluorescent. The derivatization approach is based either upon the forward reaction when the cofactors (e.g., NADH or NADPH) are the products or upon the reverse reaction in which they are oxidized (e.g., to NAD and NADP) or reduced. Two basic approaches may be distinguished: timed and kinetic. In the timed method, fluorescence is measured after a fixed time; in the kinetic method, the rate of fluorescence change is measured. Depending on the reaction conditions, the kinetic approach can be used for the determination of substrate, enzyme, activator, or inhibitor concentrations, while the timed method is useful for the determination of substrate only, in which a large excess of enzyme over the substrate (analyte) concentration is used to ensure a relatively complete reaction. Many enzymatic assays involve NADH as a reactant or product and a number of analyses based on NADH measurements are feasible. These reactions are employed in many areas. For example, a frequent scheme for fluorimetric determination of ethanol in blood plasma, various foods, and beverages is the reaction:



(in presence of alcohol dehydrogenase as an enzyme). The potential sensitivity of these reactions can be enhanced by a procedure known as coupled assays. A simple coupled assay for the determination of hexokinase activity by coupling the formation of glucose-6-phosphate to the reduction of  $\text{NADP}^+$  in presence of glucose-6-phosphate dehydrogenase is:



The fluorimetric indication of NADH has also been used in optical biosensors for lactate, pyruvate, and ethanol, where the respective dehydrogenase was immobilized at the tip of an optical NADH sensor.

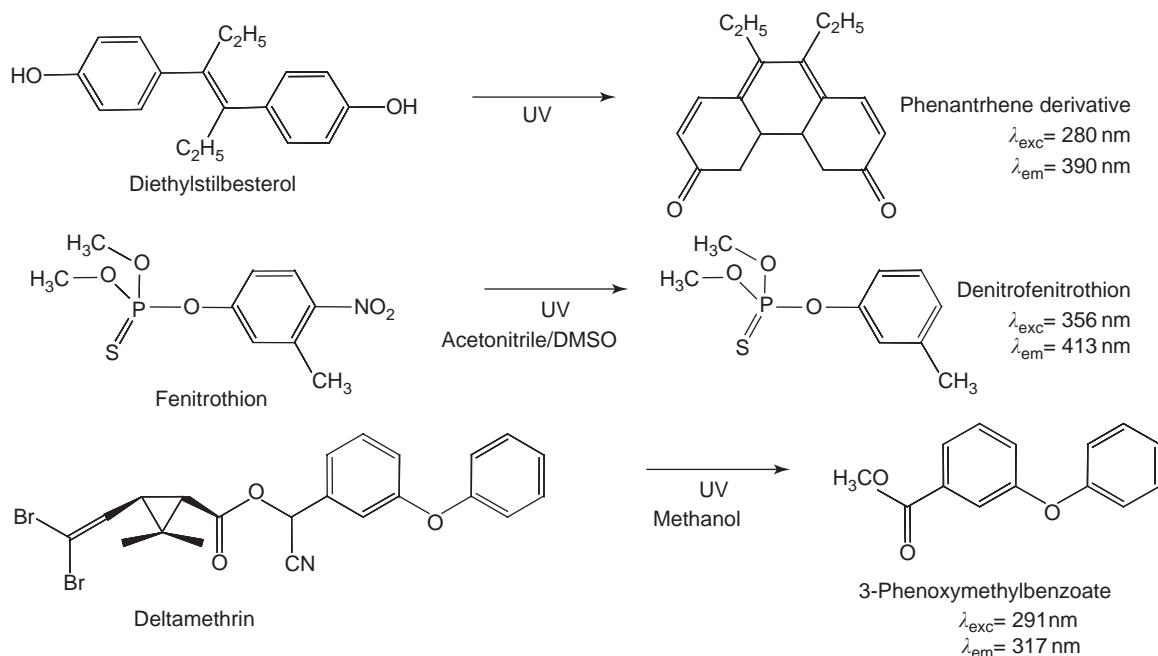
### Hydrolysis Reactions

Hydrolysis is one of the most simple treatment methods to convert nonfluorescent compounds to fluorescent ones. It is usually accomplished in a strongly alkaline aqueous medium and, in some cases, at a high temperature, resulting in the formation of fluorescent anions. Thus, acetylsalicylic acid (aspirin) has weak native fluorescence, but its base-hydrolysis conjugate, the salicylate ion, strongly fluoresces at  $\sim 400$  nm after it has been excited at about 310 nm. This property has been used to determine aspirin and salicylates directly in serum, urine, and plasma samples. In the same way, hydrolysis has been successfully applied to the determination of carbamate pesticides and organophosphorus insecticides, with detection limits in the low nanogram range. The sensitivity could be improved after heat treatment of the hydrolysis products even for many pesticides.

### Photoconversions

After exposition to UV radiation some compounds can be converted to highly fluorescent species.

Typical examples of this approach are the fluorimetric determination of the anabolic agent diethylstilbestrol in urine and the fluorimetric determination of different pesticides (see Figure 7). After a simple extraction step, the analyte may be separated from interfering substances using high-performance liquid chromatography (HPLC) and the derivative detected fluorimetrically. In spite of its simplicity, photochemical derivatization has been much less applied than chemical derivatization. The main advantages include: (1) there is no need to use chemicals, and therefore dilution is avoided; (2) fast reaction rates since most photochemical reactions take place via free radicals; (3) easy to perform, requiring inexpensive equipment; (4) the use of room temperature (compared to thermally initiated derivatization) makes the technique suitable for various experimental conditions (e.g., flow-injection analysis (FIA), dynamic systems, HPLC, thin-layer chromatography (TLC), and stationary liquid solutions); and (5) the structure of the fluorescent compound need not necessarily be known provided that reproducible fluorescence signals are obtained. Photoconversion has also some drawbacks, relative to conventional derivatization methods, such as the possible formation of nonfluorescent photoproducts, secondary thermal reactions, and/or photodegradations leading to nonstable and difficult to identify photoproducts, particularly those responsible for fluorescence emission.



**Figure 7** Fluorimetric derivatization by photoconversion.

**Table 1** Pesticide fluorescence derivatization by photoconversion

Analytical technique	Pesticide	Solvent	Photolysis time (s)	Detection limit (ng ml <sup>-1</sup> )
FBM	Fenvalerate	Acetonitrile	600	8.7
FBM	Deltamethrin	Methanol	1380	33
FBM	Deltamethrin	SDS	720	7
FBM	2,4-D	Methanol/pH 5	900	103
FIA	Fenvalerate	Methanol	52	4
FIA	Fenvalerate	Acetonitrile	95	10
FIA	Deltamethrin	Ethanol	95	18
FIA	2,4-D	Methanol/pH 5	720	98
HPLC	2,4-dichlorophenol	Methanol/pH 5	60	50
HPLC	Alachlor	Methanol	168	5
HPLC	Napropamide	Acetonitrile/water	78	0.6

FBM, fluorimetric batch method; FIA, flow-injection analysis; SDS, sodium dodecyl sulfate; 2,4-D, 2,4-dichlorophenoxyacetic acid.

Some parameters affect the conversion of nonfluorescent compounds into fluorescent ones, e.g., UV irradiation time and the nature of the solvent used (see Table 1). For example, pesticides such as fenvalerate, diflubenzuron, and deltamethrin are efficiently converted into fluorescent products (high signal with short irradiation time) in protic solvents, while for fenitrothion and chlorpyrifos polar aprotic solvents are the choice. Comparing the analytical performances of the photoconversion method with other derivatization approaches, it can be considered a rapid, sensitive, and precise method for quantifying several classes of pesticides in stationary media and as postcolumn detection mode in HPLC and in FIA.

### Complexation

Complexation involves the combination of a metal ion with a weakly fluorescent or a nonfluorescent compound containing chelate-forming, electron-donating functional groups to form highly fluorescent metal chelates. Thus, aromatic molecules that have  $n-\pi^*$  lowest excited singlets when they are not complexed, have  $\pi-\pi^*$  lowest excited singlets when they complex with a metal ion. Thus, a nonfluorescent structure becomes fluorescent. This effect can be used for the analysis of either the metal ion or the organic substance, although metal ion determinations enjoy the widest application.

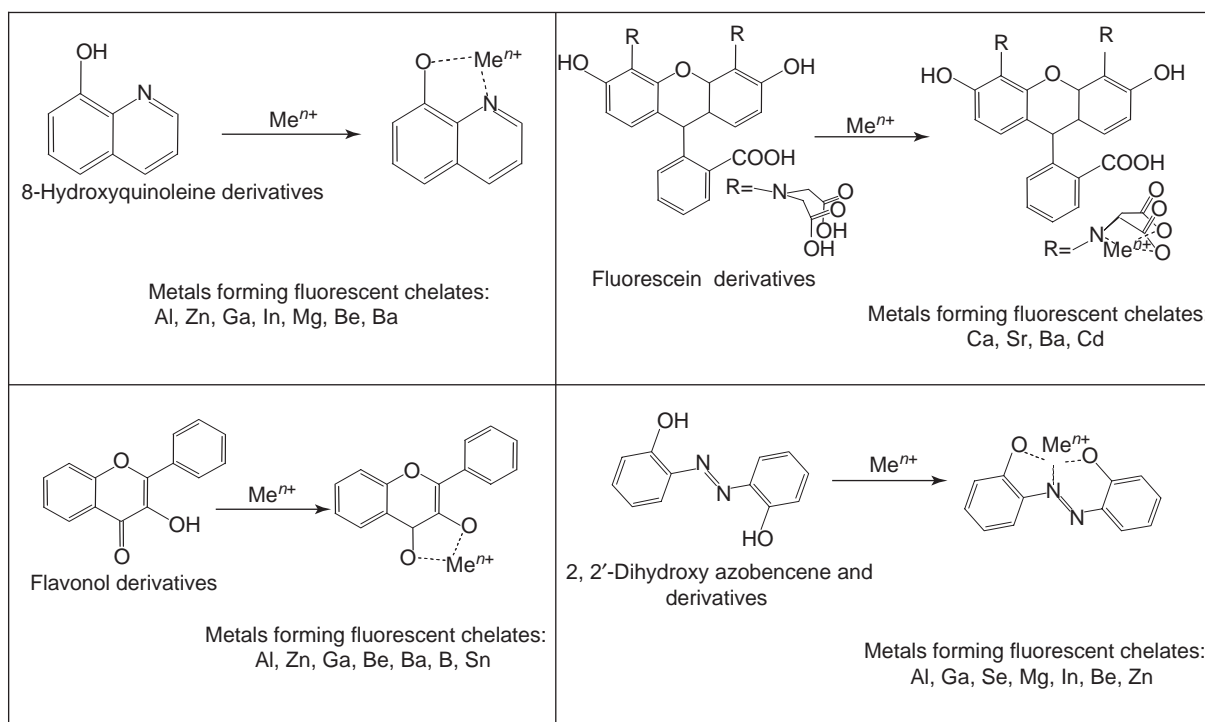
There are some interesting features related with metal complexation. If the metal is paramagnetic (e.g.,  $\text{Fe}^{3+}$ ,  $\text{Co}^{2+}$ ,  $\text{Ni}^{2+}$ ,  $\text{Cu}^{2+}$ ), an increase in intersystem crossing from the  $S_1$  state to the  $T_1$  state of the aromatic ligand takes place, which causes a decrease in fluorescence. Another phenomenon that operates in metal complexation to prevent fluorescence is the heavy atom effect. Heavy diamagnetic atoms such as  $\text{Hg}^{2+}$ ,  $\text{Pb}^{2+}$ ,  $\text{Ag}^+$ , or  $\text{Tl}^{3+}$  increase

spin-orbit coupling, which increases the rate of intersystem crossing. On the other hand, some group VIII  $d^6$  transition metal ions (e.g., ruthenium(II), rhodium(III), iridium(III), and osmium(II)) form fluorescent complexes with strong-field ligands such as 1,10-phenanthroline, 2,2-bipyridine, and 2,2,2-terpyridine. These chelates exhibit low-energy charge-transfer  $d \rightarrow \pi^*$  absorption and  $\pi^* \rightarrow d$  emission bands. Ruthenium(II) polypyridine complexes have been extensively used for luminescence oxygen-sensing applications due to the fact that they are very stable, exhibit high quantum yields, and are readily quenched by oxygen.

The basic requirement of the organic chelating reagent for fluorescence derivatization is its capacity to covalently bound with the metal ion through at least two functional groups to form a rigid structure. Ions most frequently measured by fluorescence are nontransition-metal ions, in particular those in group IIIA – aluminum(III), gallium(III), indium(III), and thallium(III) – which form metal chelates with a number of organic ligands (see Figure 8). The ligands are usually weakly fluorescent when uncomplexed and the major means of exciting the chelates is by using the  $\pi-\pi^*$  absorption bands of the chelated ligand. It is noteworthy that sensitivity of traditional analytical fluorimetric techniques for metal ions is enhanced in the presence of micelles. This effect arises from a number of interactions, among which those of the metal complex with the special microenvironment provided by micelles diminish vibrational quenching, enhance solubility, and alter the intrinsic photochemical properties of the chelate.

Under optimal reaction conditions, some organic compounds are highly specific for certain metal ions, thus resulting in highly selective fluorimetric methods for those ions, and the preconceived idea that fluorescence methods for metals do not offer significant



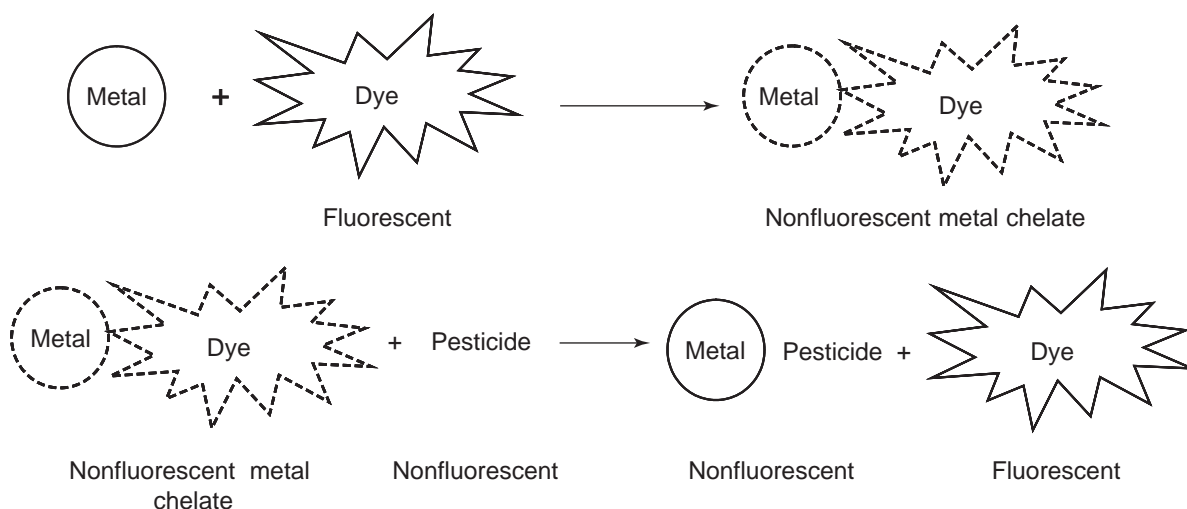


**Figure 8** Fluorescent metal complex formation.

advantages over other competing techniques, such as atomic methods (e.g., atomic absorption or inductively coupled plasma) is misleading in many situations. For example, fluorescence is the best technique for noninvasive measurements in living cells. The advent of ion-sensitive fluorescent dyes has unleashed the power of fluorescence for biomedical and life science research to identify and measure biologically significant nonfluorescent substances. Thus, calcium, magnesium, sodium, and other metal ion concentrations can be fluorimetrically measured inside cells by complexation with fluorescent dyes designed to exhibit a high affinity for a specific ion (e.g., Fura-2 is specific to calcium ions). The introduction of these ion-sensitive dyes in 1985 gave birth to the quantitative ratio fluorescence technique. The dye binds to a negligible amount of ions as compared to the total quantity of the ion species, and so it does not significantly alter the statistical equilibrium of bound and free ions. When the free ion concentration changes, the equilibrium between free indicator and ion-bound dye also changes, resulting in a change in the fluorescence properties of the indicator dye. The key characteristic of these indicators is a wavelength shift that may occur in either the excitation or the emission spectrum. An example of a dual excitation indicator is Fura-2. Its fluorescence excitation maximum

shifts from 372 nm in the absence of calcium to 340 nm when bound to calcium. In both cases, the fluorescence emission intensity is measured at 510 nm. Calcium quantitation can be performed by rationing the fluorescence intensities at the two excitation wavelengths. Since this measurement approach is based on intensities derived from the excitation spectrum, the calcium concentration measurement is independent of dye concentration and optical pathlength, parameters that cannot be controlled within a cell. Also, photodegradation of the dye due to prolonged illumination and variations in the excitation source do not compromise ratio fluorescence measurements.

In addition to direct metal analysis, indirect analysis for ligands that pull away the metal ion from the fluorescent metal chelate, is also possible. In this case, a decrease in fluorescence is observed. For example, fluoride causes the fluorescence of the aluminum–Eriochrome Red B complex to decrease. In a similar way, nonfluorescent metal chelates can be used for the fluorimetric analysis of nonfluorescent analytes via a ligand-exchange reaction, according to the scheme shown in **Figure 9**. In the first step a nonfluorescent complex is formed between a fluorescent chelator and a metal ion. In the next step the addition of the pesticide induces the displacement of the fluorescent ligand and then combines with



**Figure 9** Indirect fluorimetric analysis via ligand-exchange reactions.

the free metal ion. The pesticide concentration is then proportional to the free fluorescent dye. Using palladium as the metal ion and calcein as the fluorophore, detection limits in the nanogram range have been reported for organophosphorus pesticides using TLC. In a similar way, the use of the fluorescent  $\text{Al}^{3+}$ -Morin complex as a postcolumn reagent in HPLC allows the indirect sensitive detection of phosphonic pesticides by monitoring the  $\text{Al}^{3+}$ -Morin fluorescence decrease in the analyte band due to the presence of the phosphoric group.

Alternatively, direct analysis of nonfluorescent organic ligands by adding an excess of metal ion to form fluorescent compounds is also possible. Although this derivatization strategy has attractive analytical potential for the analysis of nonfluorescent aromatic compounds, it has received little attention. A typical example includes the fluorimetric determination of anthracyclines and tetracyclines as metal chelates with  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ , and  $\text{Eu}^{3+}$  under neutral or alkaline solutions, and with  $\text{Al}^{3+}$  or  $\text{Zr}^{4+}$ , under acidic conditions.

Metal complexation reactions are very convenient for fluorescence derivatization as they are fast and can be driven to completion by adding excess of reagent. These characteristics made them especially suitable for postcolumn derivatization in chromatographic separations, HPLC, and capillary electrophoresis (CE). An example of the use of complexation reaction is a CE postcolumn method developed for the determination of catecholamines and related compounds. The method is based on the ternary complex formed with  $\text{Tb}^{3+}$  ions, ethylenediaminetetraacetic acid and the catecholic

compound. Excitation of such a ternary complex at the absorption wavelength of the catecholic compound (the donor) results in the typical  $\text{Tb}^{3+}$  emission (sensitized luminescence of the acceptor). Detection limits are in the  $10^{-7} \text{ mol l}^{-1}$  range. Also, metal complexation can be used as a sensitive fluorimetric detection approach to metal ion speciation. For example, morin (2',3,4',5,7-pentahydroxyflavone) forms highly fluorescent complexes with mono- and di-organotin compounds in organic nonpolar media, while those with tri-organotin complexes are weakly fluorescent. As the reactions involved are simple and the kinetics very fast, derivatization is usually coupled to HPLC in the postcolumn mode.

### Quenching Reactions

Derivatization methods based on quenching make use of the heavy atom effect, the paramagnetic nature of the analyte, and/or any event that affects one or more of the photophysical and photochemical aspects related to a fluorophore. The quencher is usually a nonfluorescent species and a major obstacle in fluorescence-quenching-based methods is to find a fluorophore that is selectively quenched by the target analyte. Additionally, as a consequence of the quenching process itself, fluorescence quenching is much less sensitive than the direct emission-based approach. The decrease in intensity with increasing quencher concentration and the short fluorescence decay time mean that the magnitude of the Stern-Volmer constant  $K_{SV}$  is  $\sim 100$ . In this situation, dynamic quenching is measurably observed when the quencher concentration exceeds  $\sim 10^{-4} \text{ mol l}^{-1}$ .

Quenching methods have been used to analyze inorganic anions. Thus, halide ions have been analyzed using indicator fluorescence quenching, with iodide being the most efficient quencher and chloride the least effective. In the same way, electrophilic bromination of fluorescein to yield the low fluorescent 1,3,6,8-tetrabromofluorescein has been used for bromine determination, and a displacement reaction using a nonfluorescent  $\text{Hg}^{2+}$  metal chelate has been described for sulfide determination.

Oxygen is known to be a potent quencher of fluorophores due to its paramagnetic nature. There are a host of fluorophores that are efficiently quenched by oxygen among which the above-mentioned complexes of ruthenium and several polycyclic aromatic hydrocarbons are commonly used. This fact has been exploited to measure oxygen dissolved in water, in blood, and in gas phase as the approach allows oxygen measurement over a wide range of concentrations. Besides oxygen, quenching of fluorescence has also been used to determine a number of gases, including sulfur dioxide, ammonia, hydrogen chloride, oxides of nitrogen, as well as various volatiles such as chloroform, water vapors, and halothane. Moreover, several gas sensors and biosensors based on fluorescence quenching have been developed. Sensors based on fluorescence quenching are particularly suitable for dissolved gas measurements due to the high diffusibility of the gases through suitable polymer membranes used to immobilize the fluorophore.

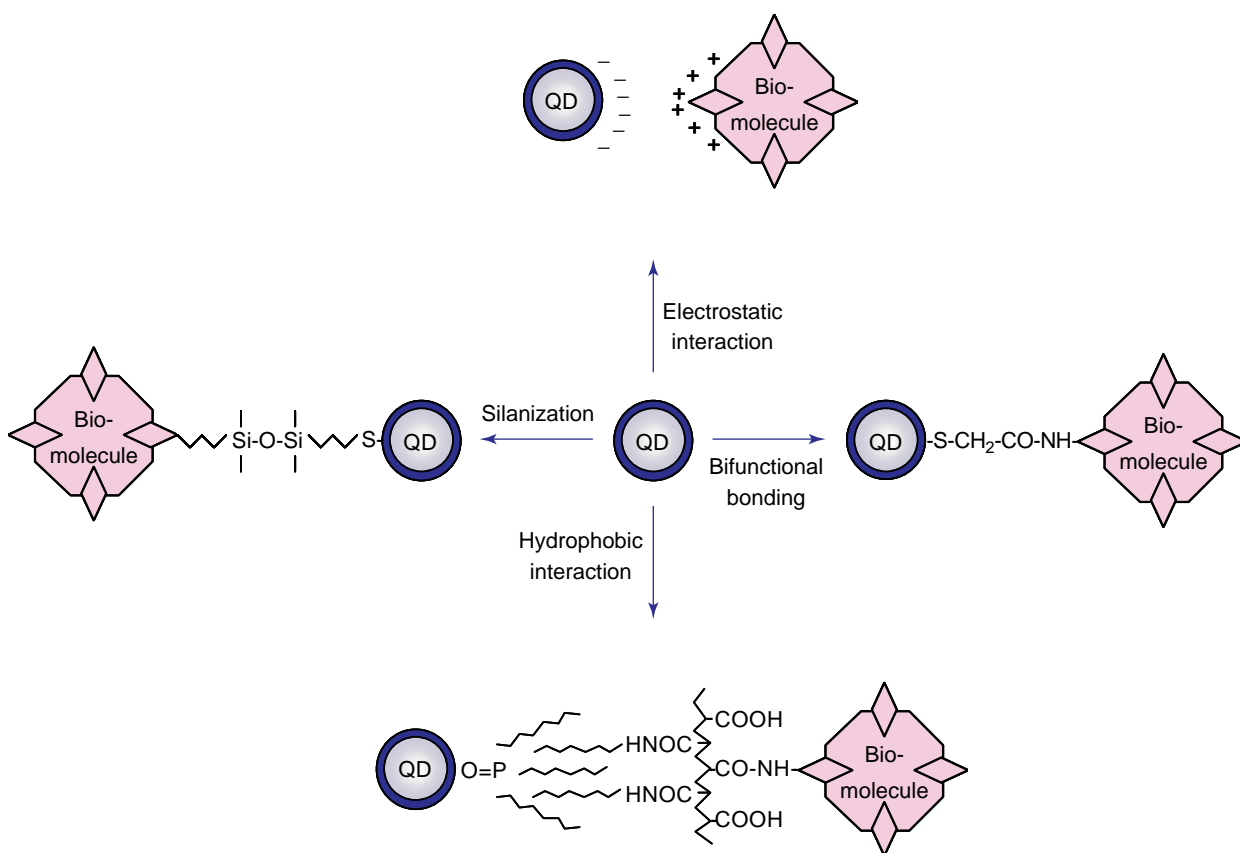
### Extrinsic Fluorescent Tagging

As many biologically important analytes do not exhibit efficient detection properties (UV or visible light absorption, fluorescence, or electrochemical activity), their detection limits are relatively low. For example, drugs with chiral centers exist naturally in racemic mixtures that are optically inactive due to the nearly equal proportion of the enantiomers. The determination of enantiomeric purity is of paramount importance in the pharmaceutical industry as each enantiomer may have different therapeutic characteristics. Currently, a method that offers multiple advantages for chiral separations is by converting enantiomers to diastereomers by precolumn derivatization with a pure fluorescent enantiomer. For instance, propranolol existing in racemic form may be analyzed by precolumn derivatization with (+)-1-(9-fluorenyl)ethyl chloroformate. Well correlated calibration plots were found up to 400 pmol and a reproducibility of <2% for each derivative.

Two types of fluorescence derivatization reactions can be distinguished for biologically important molecules: covalent and noncovalent reactions. In covalent procedures a covalent bond is formed between the analyte and the fluorescent tag. All other types of chemical interactions (e.g., ion-pairing, intercalation, electrostatic interaction) are grouped into noncovalent reactions. Covalent derivatization reactions of interest for biological or medical assays are covered elsewhere in this encyclopedia. Most of these reactions have been used as HPLC and CE precolumn derivatization methods and most of the fluorogenic reagents have been synthesized on the basis of fundamental structures such as coumarin, quinoxaline, benzofurazan, xanthene, etc. Postcolumn fluorescence derivatization is usually carried out when multiple products are obtained or when the resultant fluorescent derivative is not stable enough during the chromatographic process, but the rate of derivatization is rapid. In summary, a wide variety of fluorogenic derivatization reactions are available for applications to organic and inorganic substances in a wide range of disciplines, e.g., chemistry, biochemistry, medicine, pharmaceutical, forensic, and environmental. With increasing availability of suitable fluorogenic reagents and detectors, fluorescence derivatization is also being increasingly applied. Fluorophores currently used as derivatization reagents offer wide variations of chemical and spectrochemical characteristics to the analyst in tuning the chemistry to his/her particular needs for improved detection (direct or indirect methods, steady-state measurements, kinetic methods, etc.) or to meet requirements imposed by available instrumentation (e.g., wavelength of excitation).

### Quantum Dots as Alternative Labels

The last few years have seen a growing interest in the use of quantum dots (QDs) in biological and chemical sensing applications, as well as for electronic devices and quantum dot lasers. QDs are a new type of fluorescent derivatization/labeling reagents. QDs are nanometer size particles ( $10^{-9}$  m) with special optical properties, composed of a semiconductor material, cadmium selenide (CdSe), which has been coated with an additional semiconductor zinc sulfide (ZnS) shell. This core-shell structure is further coated with a polymer shell that allows to be conjugated to biomolecules, while retaining the optical properties of the QD. The interactions of the QDs with target molecules or with their environment depend strictly on the nature (functionalization) of this polymer layer. For example, the QDs can be capped



**Figure 10** Fluorescent quantum dot covalent derivatization.

with an organic layer such as mercaptoacetic acid, trioctyl phosphine/trioctylphosphine oxide mixture, or with a biomolecular (antibody, oligonucleotide, streptavidin) shell. To date, biomolecules can be linked to QDs mainly by covalent attachment, electrostatic attraction, hydrogen bonding interactions, and hydrophobic forces (see **Figure 10**). As an alternative to conventional fluorophores, QDs have several attractive features, including high emission quantum yields, sharp emission spectra, highly resistant to chemical and metabolic degradation, high photostability, tunable color through the whole spectrum, long lifetimes, and shelf life. The color and optical properties of QDs can be tuned to any chosen wavelength by simply preparing them in a different size or composition. QDs can be detected at concentrations comparable to organic dyes by conventional fluorescence methods and individual bioconjugated QDs are observable by confocal microscopy. Novel QD-conjugates are potentially useful in numerous fluorescence tagging applications, for developing homogeneous bioassays, and as multicolor fluorescent labels for ultrasensitive detection and imaging. In conclusion, the future of QD labeling concept

appears promising for developing novel biosensing technologies, for noninvasive molecular imaging, and for medical diagnostics, pharmacogenomics, and drug discovery.

See also: **Chromatography:** Overview; Principles.  
**Fluorescence:** Fluorescence Labeling.

## Further Reading

- Bruchez M, Jr., Moronne M, Gin P, Weiss S, and Alivisatos AP (1998) Semiconductor nanocrystals as fluorescent biological labels. *Science* 281: 2013–2015.
- Coly A and Aaron JJ (1998) Fluorimetric analysis of pesticides: methods, recent developments and applications. *Talanta* 46: 815–843.
- Fukushima T, Usui N, Santa T, and Imai K (2003) Recent progress in derivatization methods for LC and CE analysis. *Journal of Pharmaceutical and Biomedical Analysis* 30: 1655–1687.
- Zhu R and Kok WTh (1998) Post-column derivatization for fluorescence and chemiluminescence detection in capillary electrophoresis. *Journal of Pharmaceutical and Biomedical Analysis* 17: 985–999.

## Fluorescence Labeling

**M E Díaz-García and R Badía-Laiño**, University of Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Fluorescence labeling procedures convert nonluminescent or weakly luminescent molecules into highly fluorescent products. With the aid of specific probes, and in combination with appropriate detector systems, chemical and biochemical systems become amenable to fluorescence spectroscopy. The fluorescence labeling technology has opened new horizons in biosciences, especially in biomedical sciences. It is worth mentioning that fluorescence labeling is made use of in fields ranging from genomics to proteomics, and in areas such as DNA sequencing, protein quantification, polymerase chain reaction (PCR) product analysis, or Western, Northern, and Southern blotting.

The first methods used to label molecules and thus make them quantifiable, employable radioactive isotopes.  $^{125}\text{I}$  isotope is still in use today due to its small size and due to its low detection limit of  $\sim 10$  amol ( $1 \text{ amol} = 10^{-18} \text{ mol}$ ). In spite of the advantages of radioactive labeling, there are a number of disadvantages. For example,  $^{125}\text{I}$ -labeled antibodies cannot be stored for longer than a month, and  $^{32}\text{P}$ -labeled nucleotides decay significantly in about a week. On the contrary, fluorescent labeled biomolecules can be stored for 6 months or longer. In fact, fluorescent labeled reagents can be prepared in a large scale, standardized, and used for extended periods, which minimizes reagent variability between assays. Handling of radioactive materials requires adequate protection of lead or acrylic shields and disposal of radioactive wastes require shielded storage, long-term decay, or regulated landfill disposal. Most fluorescent molecules are easy to handle; the simple use of gloves affords adequate protection and disposal problems are minimal. As a consequence of its several distinct advantages, fluorescent labeling is in many cases less expensive than radiolabeling. Fluorescence labeling is a powerful tool to obtain information from chemical and biochemical systems for a variety of experimental, analytical, and quality control applications, several of which are described below.

### Labels and Reactions

In general, a label is a molecule capable of emitting a signal, which is designed for labeling and to localize

a specific region of a biological system, or to respond to a specific stimulus. It contains a signaling unit (the fluorophore), an anchor group for bonding to the molecule to be labeled, and a spacer. The signaling moiety acts a signal transducer, i.e., changes in the photophysical characteristics of the fluorophore due to different processes (e.g., energy transfer, polarization) are expressed as optical signals. As regards the spacer, it minimizes undesirable steric interactions between the signaling unit of the label and the labeled molecule. A substance labeled in this way is usually designated as a tracer.

The label must fulfill some specific requirements to be useful for labeling:

- Coupling to the target molecule must be simple and gentle to perform, specific for the compound of interest, fast, and quantitative. It is important also that the fluorescence of the label should be soluble in aqueous solutions and should be insensitive to environmental pH and polarity. Primary amines, thiols, and aldehydes are common functional groups available for biomolecules to covalently attach fluorescent labels.
- The label should not alter the properties of the target molecule or the functions of the cell or the organelle by reacting with key groups in active sites or by causing nonspecific binding of a labeled protein, or by causing steric perturbations because of its size. The probe should also not be phototoxic.
- The labeled compound should be stable and not undergo rapid decomposition.
- The luminescent properties of the label should not change significantly after the coupling.
- The labeled compound should have high fluorescence efficiency for maximum sensitivity. So, the value of molar extinction coefficient among fluorescent probes of current practical importance is  $\sim 5000 \text{ mol l}^{-1} \text{ cm}^{-1}$  to  $2 \times 10^5 \text{ mol l}^{-1} \text{ cm}^{-1}$ . On the other hand, for practical purposes, the quantum yield should be  $\sim 0.4$  or greater once the fluorescent probe is bound to the target and is in the solvent environment, where the measurement is to be made.
- Biological systems (e.g., proteins, cells) excited at wavelengths below 500 nm produce considerable autofluorescence that arises mainly from flavins, flavoproteins, NADH, etc. Labels that can be excited at wavelengths above 500 nm are much less susceptible to optical interference from biological chromophores in the test sample.



The term 'probe' is normally used as a synonym of label. Thus, in contrast to labels, probes respond to their microenvironment or to a chemical species. Those probes responding to a chemical species such as oxygen, an ion, or to pH are also referred to as indicators. In this article, both terms will be used indistinctly. Different classes of fluorescent labels are available. Low molecular weight dyes include xanthene (rhodamines, fluoresceins), cyanine, coumarins, sulfonated pyrenes, and metal phthalocyanine compounds, while high molecular weight labels include phycobiliproteins and other luminescent proteins.

## Intercalative Fluorescent Labels

Labeling is usually performed by the formation of a covalent bond between the fluorescent label and the target molecule. However, fluorescent dyes without a reactive group can be employed for some purposes. Due to their particular structural nature, fluorescent probes can bind noncovalently to special biomolecules in cells in a specific or a nonspecific fashion. In this case, the term 'stain' is used to refer to the label and 'staining' to the noncovalent attachment of the label to the studied object. There are a large number of fluorescent dyes that bind to DNA and RNA forming complexes via intercalation. Intercalating dyes are planar aromatic cations that insert between stacked base pairs on the DNA duplex. Their structure and chemical properties provide appropriate size and van der Waals interactions favorable for insertion between bases.

Intercalation provides an environmentally dependent fluorescence change, giving rise to a large fluorescence signal increase relative to the free dye in solution. This signal enhancement provides a concentration proportional response, allowing direct quantitative DNA measurements. Intercalating dyes can also bind nucleic acids inside whole cells, correlating the fluorescence response to enumeration and identification. Ethidium bromide (2,7-diamino-9-phenyl-10-ethylphenanthridinium bromide), related phenanthridinium dyes, and other intercalative dyes (Figure 1) illustrate the pattern of dye-nucleic acid interactions. The primary, intercalative mode of dye binding depends on double-helical polynucleotide structure. Such binding, which is associated with a shift in dye absorption from 480 nm to 520 nm and a marked increase in fluorescence efficiency, can occur with either DNA or RNA. In addition to intercalative binding, ethidium bromide can undergo an external electrostatic interaction with double helices. This secondary dye binding, which tends to reduce the

fluorescence quantum yield and lifetime of dye bound to primary sites, can be reduced by working in solutions of moderate ionic strength and low excess dye concentrations. Intercalative ethidium bromide depends little on nucleic acid base composition.

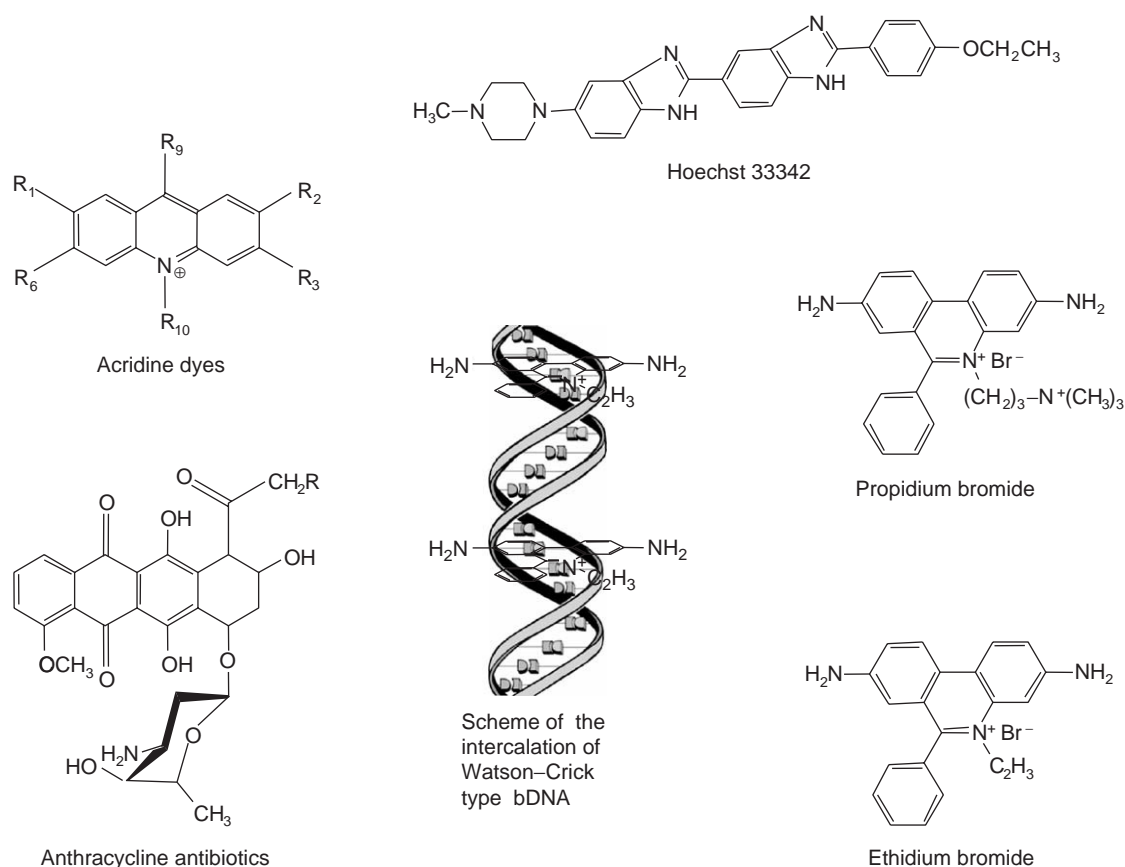
Important information about structural and functional aspects of DNA can be obtained from intercalative fluorescent dyes. For example, conformational factors, such as the presence of Z-DNA or superhelical DNA may affect dye binding, particularly for those intercalating dyes sensitive to topological constraints. Propidium iodide is not appreciably taken up by living cells, so this dye has been useful in discriminating live from dead cells. In contrast, certain bisbenzimidazole dyes, such as Hoechst 33342, are taken up by living cells and have become important fluorescent dyes for staining live cells. Dyes such as Hoechst 33258 preferentially bind to DNA regions that are rich in A-T base pairs while Chromomycin A<sub>3</sub> binds to regions that are rich in G-C. On the contrary, quinacrine and anthracene dyes have little base preference in binding, but their fluorescence quantum yield varies with base composition. The main limitation of intercalating dyes is their toxic/carcinogenic characteristics.

Fluorescence-based methods can detect nucleic acids in solution at concentrations in the nanogram per milliliter range and can be achieved using dyes such as PicoGreen for double-stranded DNA, OliGreen for single-stranded DNA and oligonucleotides, and RiboGreen for RNA. Usually, RNA samples must be treated with Dnase in order to remove DNA contamination as no dye is yet available that exhibits fluorescent enhancement specifically by binding to RNA. In Table 1, fluorescence characteristics of some common nucleic acid probes are summarized.

## Covalent Binding

For covalent labeling processes a few reactions have achieved practical significance. Primary amines, thiols, aldehydes, and ketones are common functional groups available for biomolecules to covalently attach fluorescent labels. In Figure 2, primary amine reactive groups and their reactions are shown. Amine-reactive labels are widely used to modify biomolecules and to prepare bioconjugates for immunochemistry, fluorescence *in situ* hybridization, receptor labeling, fluorescent analog cytochemistry, and cell tracing. Isothiocyanates are widely used to label proteins to form thiourea derivatives for fluorescence microscopy and fluorescence immunoassays. There are a number of amine-reactive fluorescent isothiocyanate labels, among which fluorescein isothiocyanate and tetramethylrhodamine





**Figure 1** Molecular structures of some intercalative dyes for nucleic acids labeling.

**Table 1** Fluorescence and chemical features of nucleic acid intercalative labels

Label	Excitation maximum (nm)	Emission maximum (nm)	Molar extinction coefficient ( $\text{mol}^{-1} \text{cm}^{-1}$ ) ( $\times 10^{-4}$ )	Comments
Ethidium bromide	526	605	0.52	Classic general-purpose nucleic acid label
Propidium iodide	535	617	0.54	Impermeant; dead-cell stain
Hoechst 33342	350	460	4.5	Cell permeant; sequence dependant; dsDNA-selective binding
TOTO-1	514	533	11.7	High affinity DNA labeling; dead-cell stain
YOYO-1	491	509	5.2	Impermeant; ultrasensitive high affinity DNA labeling
Acridine	500*	526*	5.3	Permeant; DNA/RNA
Orange	460 <sup>+</sup>	650 <sup>+</sup>		Discrimination
Vistra Green	490	520	—	Ultrasensitive gel stain for ss or dsDNA or RNA

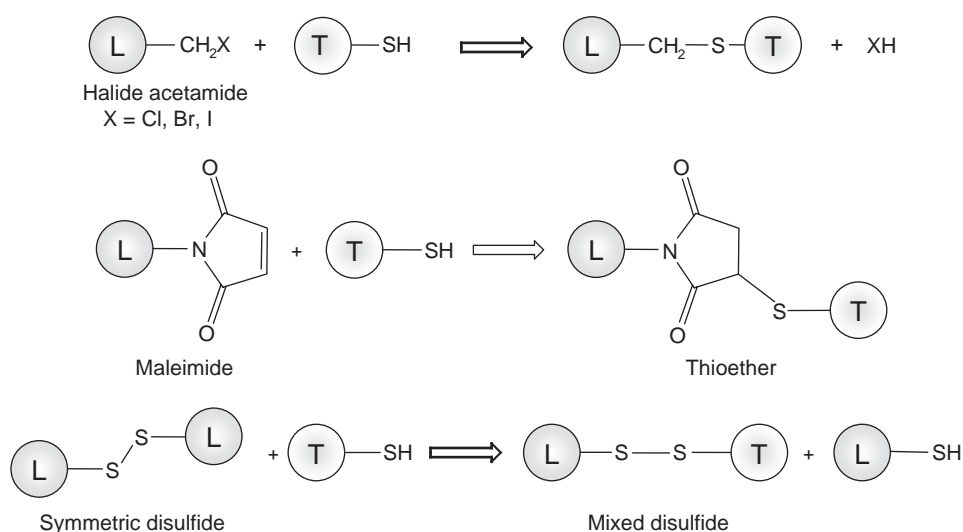
\*Dye bound to ds-DNA; <sup>+</sup>Dye bound to RNA.

ssDNA: single-strand DNA; dsDNA: double-strand DNA.

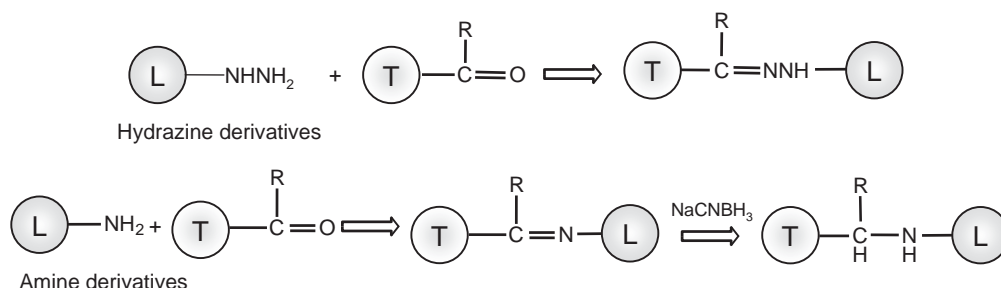
isothiocyanate are widely used for preparing fluorescent antibody conjugates. The *N*-hydroxysuccinimide ester group has several advantages over other reactive groups: (1) it can be readily synthesized from

carboxylic acid derivatives; (2) the corresponding labels can be obtained in high purity, so labeling can be a reproducible process; (3) the dry reagent can be stored for long periods of time without loss of





**Figure 3** Thiol fluorescent reactive labels (L, label; T, target).



**Figure 4** Carbonyl reactive groups (L, label; T, target).

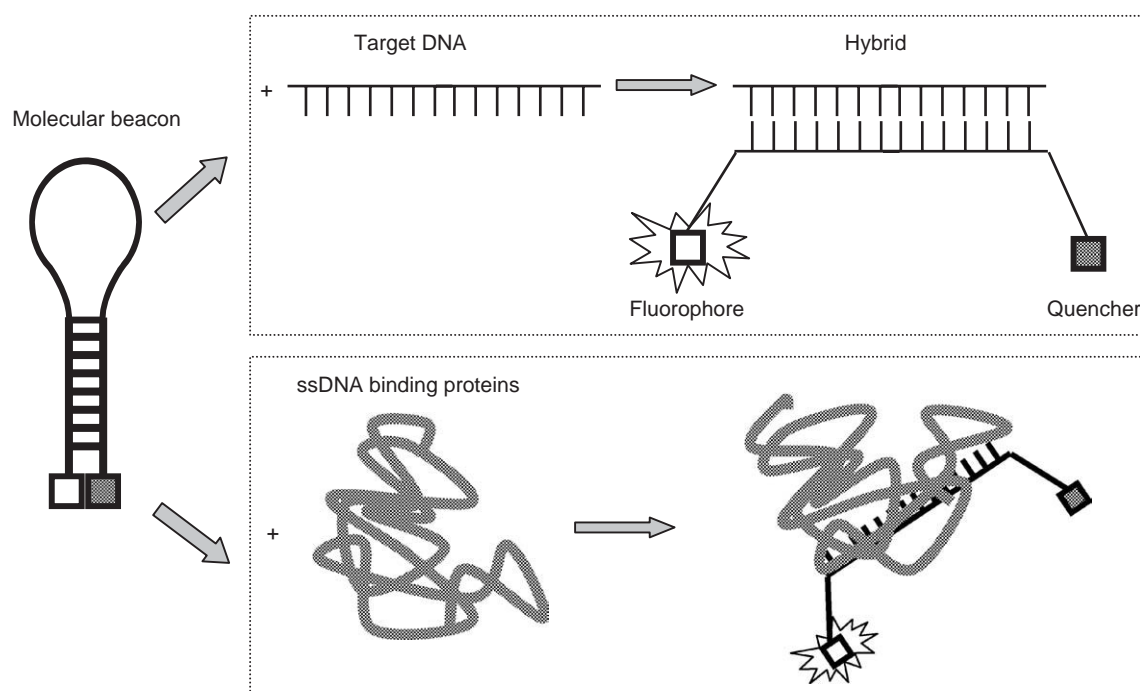
## Molecular Beacons

Molecular beacons are oligonucleotide probes that contain both fluorophore and quencher moieties and can report the presence of specific nucleic acids in homogeneous solutions. First reported by Tyagi and Kramer in 1996, molecular beacons appear to be an invaluable tool for quantitative genomic studies. These probes possess a loop and stem structure (hairpin-shaped molecules), as schematically shown in **Figure 5**. The stem is formed by the annealing of complementary arm sequences on the ends of the probe sequence. A fluorescent moiety (e.g., tetramethylrhodamine, 5-iodoacetamidofluorescein) is attached to the 5'-end of the oligonucleotide and a quenching moiety (e.g., DABCYL, 4-(4'-dimethylaminophenylazo)benzoic acid) to the group at its 3'-end. The stem keeps these two moieties in close proximity to each other, thus causing the emission of the fluorophore to be quenched by energy transfer (direct fluorescence energy transfer or resonance energy transfer). When a target DNA hybridizes to

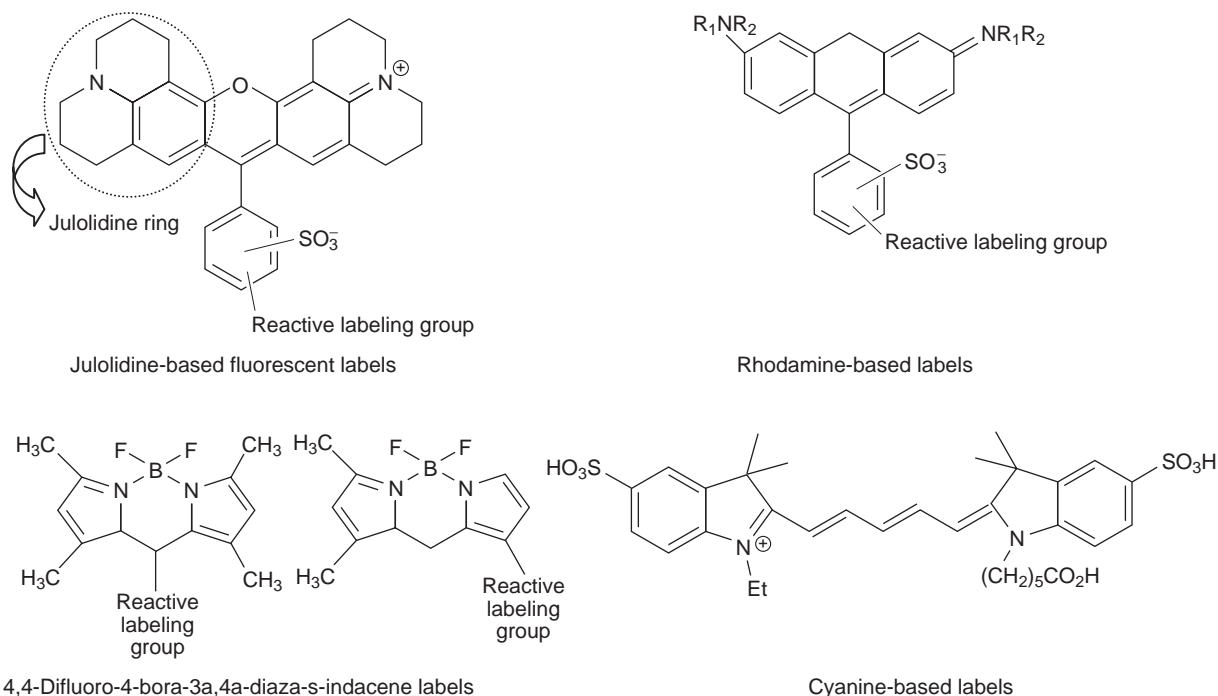
a molecular beacon, the conformational reorganization forces the stem apart and causes the fluorophore and the quencher to move away from each other, so the fluorophore is no longer quenched and the molecular beacon fluoresces. Molecular beacon technology can be used in a variety of PCR applications, for the detection of genetic mutations, to assay enzymatic cleavage of ssDNA (see **Figure 5**), and for RNA detection in living cells.

## Visible-Near-Infrared Labels

Fluorescein and rhodamine dyes have been for many years the dyes of choice in fluorescence labeling. These dyes belong to the xanthene class of dyes containing an oxygen atom in a central six-member ring. The fluorescence of fluorescein is pH sensitive and bleaches rapidly, thus limiting its use in many applications. Efforts to obtain more photostable fluorescent labels that can be excited in the long-wave region and also emit revealed that bioconjugatable



**Figure 5** Schematic representation of a molecular beacon.



**Figure 6** Visible–NIR labels.

arylsulfonated cyanine fluorophores, 4-bora-3a,4a-diaza-s-indacene-based fluorophores and tetramethyl-rhodamines and julolidine-based rhodamines include molecules that possess absorption and

emission maximums in the visible–near infrared (NIR) region of the spectrum (**Figure 6**) and are not affected by changes in pH between 4 and 10. These dyes find many uses in fluorescence microscopy,

**Table 2** Fluorescence characteristics of some NIR dye labels

Label	Absorption maximum (nm)	Fluorescence emission maximum (nm)	Molar extinction coefficient ( $\text{mol l}^{-1} \text{cm}^{-1}$ ) ( $\times 10^{-4}$ )	Fluorescence quantum yield
Cy 3	550	570	15	0.04
Cy 3B	558	572	13	0.67
Cy 5	649	670	25	0.28
Cy 7	743	767	20	0.28
TO-PRO-3	642	661	1	—
TO-PRO-5	747	770	7.7	—
LDS-751	543	712	4.6	—
SYTO-17	621	634	8.8	—
TOTO-3	642	660	16	—

immunodiagnostics, in whole-body imaging, and microarray analysis.

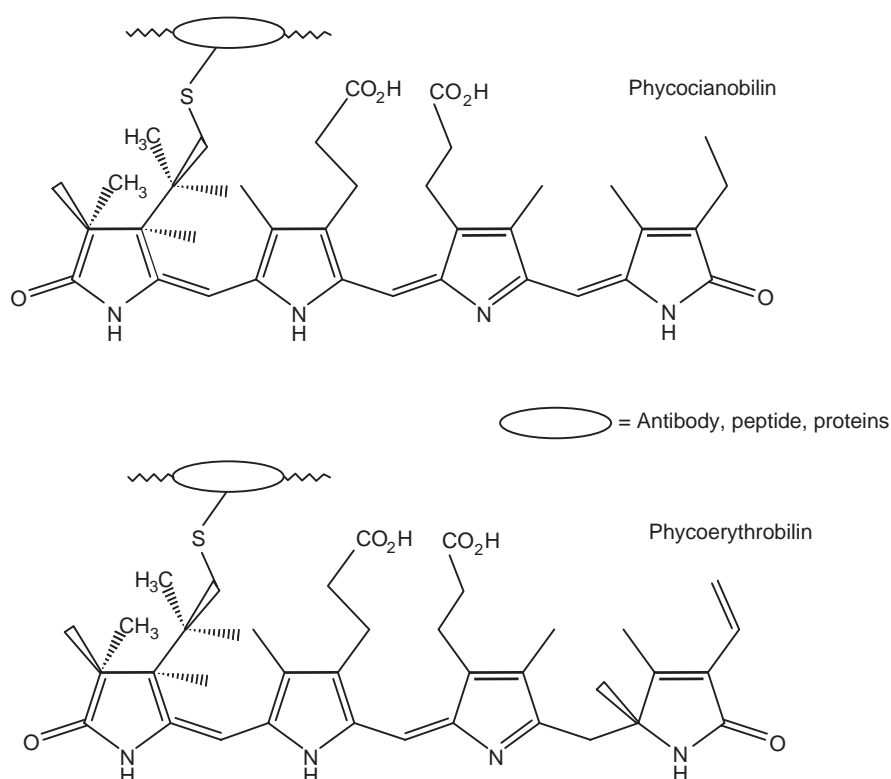
Cyanine-based dyes have a number of advantages over other dyes: the high molar extinction coefficient combined with a good quantum yield results in a very good brightness, which gives high sensitivity and direct labeling capability. Also, the addition of  $-\text{C}=\text{C}-$  units to the methine bridge between the two heterocyclic ring leads to a shift of  $\sim 100$  nm per unit in the excitation wavelength, leading to a wide range of multicolor dyes. The cyanines and related compounds, merocyanins, oxonols, and styryls, have high photostability, bind by intercalating between the bases of nucleic acid molecules, and are not pH sensitive. In Table 2 are reported the fluorescence characteristics of some commonly used cyanine, phenanthridine, and acridine NIR labels. These dyes found interesting applications in DNA microarray technology to investigate global changes in gene expression of cells and tissues.

Most bright fluorophores have usually small Stokes shifts and narrow absorption and emission bands and, consequently, cannot be used in combination with single laser lines for multicolor detection. A technique to provide fluorescent labels with large Stokes' shifts makes use of 'energy transfer cassettes'. In this approach a conjugate of a fluorescent donor (sensitizer) is covalently linked to a fluorescent acceptor. In the design of such constructs an inert, covalent linker between the donor and the acceptor holds the two dyes in close proximity, in such a way that efficient energy transfer from one dye to the other takes place when contiguous dyes are oriented parallel to each other, as opposed to any other arrangement. The overall Stokes' shift is dependent upon the wavelength of excitation of the donor fluorophore and the wavelength of emission of the acceptor one. Using cyanine dyes as both donor and acceptor molecules, a range of energy transfer cassettes has been synthesized. The labels used are the activated *N*-hydroxy-succinimidyl ester, maleimide, or other more specific labeling groups.

The multichromophore phycobiliproteins represent an additional series of labels that exhibit far-red absorption. Phycobilisomes are protein assemblies in cyanobacteria (blue-green algae) and red algae that enable them to harvest light that is not absorbed efficiently by chlorophyll, serving as light-absorbing antennae to funnel excitation energy into the reaction centers of photosystem II. The structures of the two prosthetic groups are given in Figure 7 and spectral data in Table 3. These proteins contain covalently linked tetrapyrrole groups that play a biological role in collecting light and, through fluorescence resonance energy transfer, conveying it to a special pair of chlorophyll molecules allocated in the photosynthetic reaction center. The compounds exhibit very high molar extinction coefficients ( $2 \times 10^6 \text{ mol l}^{-1} \text{cm}^{-1}$ ) and high quantum yields up to 0.98. Other several advantages of phycobiliproteins as fluorescent labels include: very high water solubility, multiple sites for stable conjugation to many biological and synthetic materials, relatively free of spectral interference from other biological materials, and a homogeneous structure with defined molecular weights. They were initially employed in fluorescence microscopy and thereafter methods were developed for flow cytometry and for coupling these large protein-chromophore complexes to antibodies studies. Conjugates of phycobiliproteins with other proteins is generally carried out using pyridyldisulfide derivatives, which can be directly reacted with thiolated antibodies, enzymes, and other biomolecules to form a disulfide linkage.

## Green Fluorescent Protein

Green fluorescent proteins (GFPs) are found in a variety of bioluminescent organisms such as the jellyfish *Aequorea victoria* and the anemone *Discosoma*. While most of other proteins require a cofactor to fluoresce, GFPs exhibit intrinsic fluorescence thanks to three amino acids that cyclize

**Figure 7** Phycobiliproteins.**Table 3** Spectral data of selected phycobiliproteins

Protein	Molecular weight ( $\times 10^{-3}$ )	Absorption maximum (nm)	Emission maximum (nm)	Molar extinction coefficient ( $\text{mol l}^{-1} \text{cm}^{-1}$ ) ( $\times 10^{-4}$ )	Fluorescence quantum yield
B-phycoerythrin	240	546, 565	575	240	0.98
R-phycoerythrin	240	496, 546, 565	578	200	0.82
Allophycocyanin	104	650	660	70	0.68

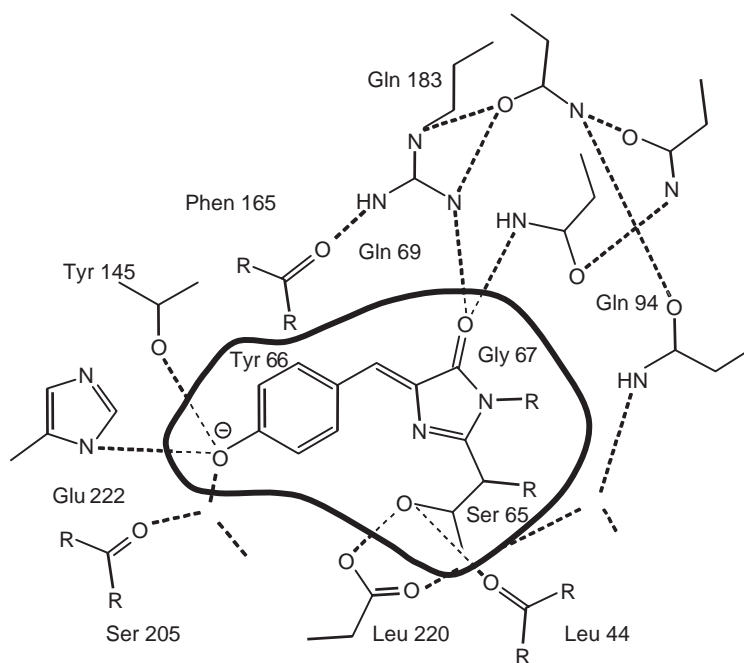
(Ser<sub>65</sub>-Tyr<sub>66</sub>-Gly<sub>67</sub>) and then undergo an oxidation step during a complex process of maturation. This fluorophore (see **Figure 8**) is tightly packed within the hydrophobic core of GFP that consists of a compact barrel of 11  $\beta$ -sheets with a central  $\alpha$ -helix containing the fluorophore.

GFPs are excellent markers for studying the expression and functional dynamics of proteins in real time in living cells using fluorescence methods such as photobleaching and chemical labeling with different colors of biarsenic compounds. For example, in fluorescence recovery after photobleaching (FRAP), GFP-tagged proteins within a specific region of a cell are photobleached using a high-power laser. The rate at which fluorescence recovers in the bleached region is related to the ability of nonbleached proteins to

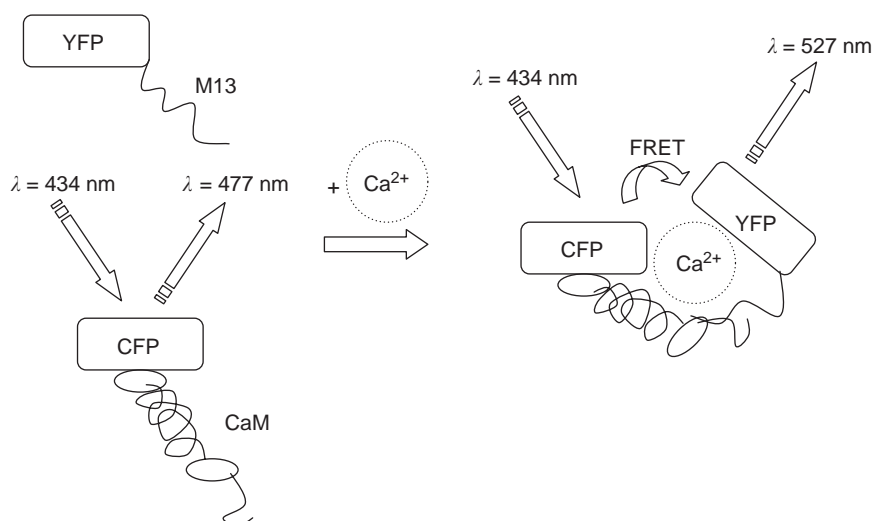
move back into that area. This approach has been successfully applied to study protein dynamism in the Golgi region and the nucleus.

As wild-type GFPs are not optimal for some reporter-gene applications, genetically modified fluorescence-enhanced GFP variants, cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP) have been generated in order to overcome the weak emission intensity of wild GFP and some photochemical drawbacks. The adequate Stokes' shift of these proteins (CFP,  $\lambda_{\text{ex}}=434 \text{ nm}$ ,  $\lambda_{\text{em}}=477 \text{ nm}$ ; YFP,  $\lambda_{\text{ex}}=514 \text{ nm}$ ,  $\lambda_{\text{em}}=527 \text{ nm}$ ) and the possibility of efficient energy transfer from CFP to closely apposed YFP chromophores made them particularly useful for studying simultaneous protein expression, protein-protein interactions, and localized enzyme activation.





**Figure 8** Green fluorescent protein fluorophore and its environment.



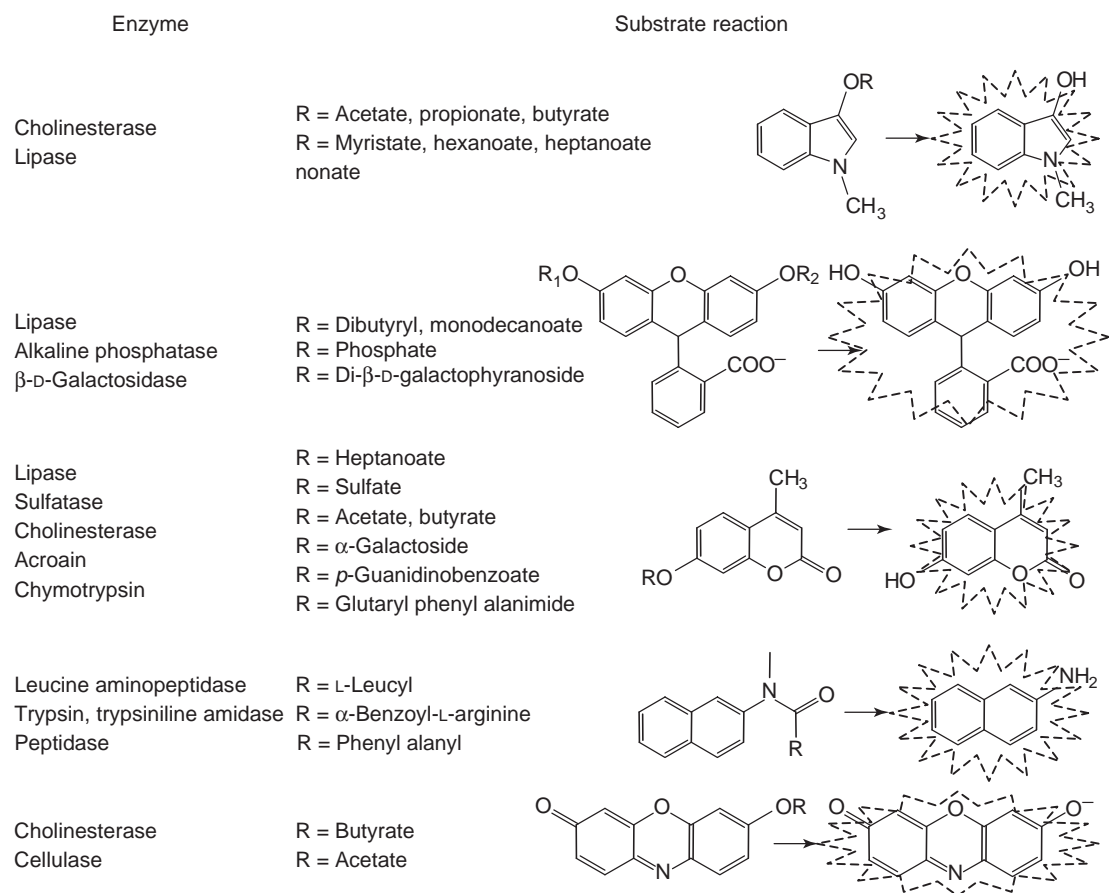
**Figure 9** Chameleon calcium reporter.

Recently, a 'chameleon calcium reporter' has been designed by R.Y. Tsien to sense calcium in cells by FRET. The structure of the system is based on a CFP separated from an YFP by the calmodulin calcium binding protein (CaM) and a calmodulin binding peptide (M13). If  $\text{Ca}^{2+}$  ions are bound, CaM wraps around M13 allowing high efficiency of excitation transfer from the donor CFP to the acceptor YFP (see **Figure 9**). The degree of FRET in chameleon is a sensitive ratiometric reporter (ratio 527/434 YFP/CFP) of the calcium concentration in cells and in solution. The beauty of this approach is that the

reporter does not have to be loaded into the cell under study but the organism is transfected with gene so that all cells within the organism will express the calcium fluorescent indicator.

## Enzyme Labels

The first alternative to radioactive labeling was addressed using enzymes as labels. Although enzymes are no signal-generating compounds, they can produce many signal-generating species. The three more



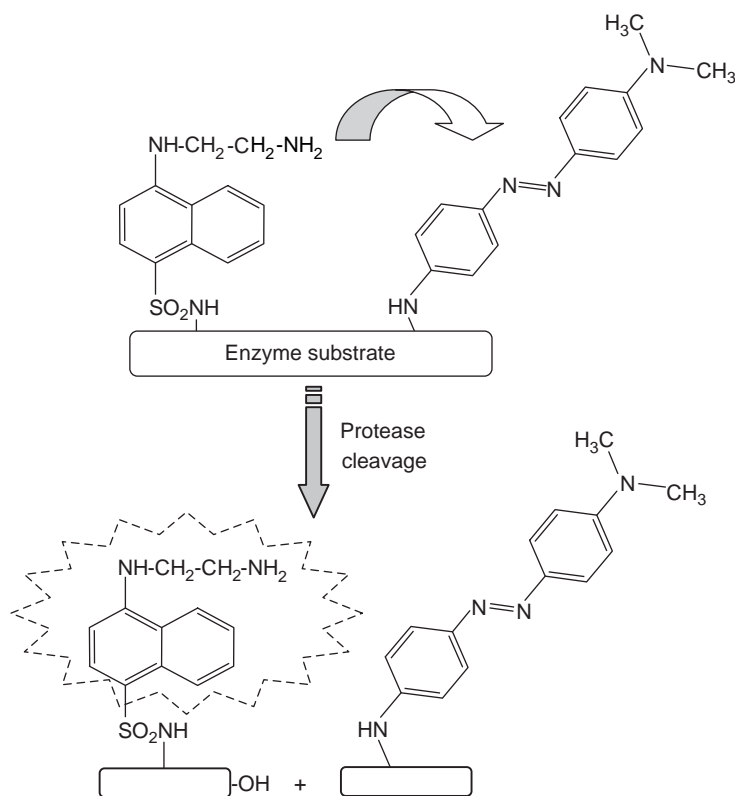
**Figure 10** Fluorogenic enzyme substrates.

important enzymes that are commonly used as labels are alkaline phosphatase (AP),  $\beta$ -D-galactosidase (GAL), and horseradish peroxidase (HRP). Among them, HRP with a molecular weight of  $\sim 40$  kDa, presents the fewest labeling steric problems and AP displays the highest catalytic activity. The coupling of enzymes to antibodies or to other biomolecules is often achieved by bifunctional coupling reagents. On the other hand, fluorogenic enzyme substrates are compounds that are converted by specific enzymes into products that have either increased fluorescence or shifted spectra. Several fluorogenic enzyme substrates, many of them applicable to various hydrolytic enzymes, including esterases, phosphatases, and peptidases, are presented in **Figure 10**. In the listing chemiluminescent enzyme substrates such as luminol or dioxetane derivatives were not included. The principle of excited-state energy transfer can also be used to generate fluorogenic substrates; for example, when a donor fluorophore is effectively quenched by a nearby acceptor. The enzymatic cleavage of the substrate results in spatial separation of the fluorophore and acceptor, thus restoring the fluorophore's

emission (see **Figure 11**). The donor and acceptor must be carefully chosen to ensure efficient quenching through excited-state energy transfer. The use of fluorogenic enzyme substrates has been particularly useful for monitoring enzyme activities and in situations where enzyme activities are low.

## Fluorescent Labeling Based on Substituent Effects

Some reactions used for fluorescence labeling are based on changing the nature of a substituent on an aromatic structure. For example, the reaction of fluorescamine with amines is quite similar to a labeling reaction, is fast, and specific for primary amines. The reagent itself is not fluorescent and does not hydrolyze to a fluorescent product. Reaction of cyanide with benzoquinone is another example where modification of a substituent eliminates  $n - \pi^*$  transitions into the basic aromatic structure. Addition of an electron-donating substituent results in a modification of fluorescence properties, such as in the



**Figure 11** Enzymatic suppression of fluorescence energy transfer.

reaction of primary or secondary amines with 1,2-naphthoquinone-4-sulfonic acid. Some of these fluorescence labeling reactions are summarized in Figure 12.

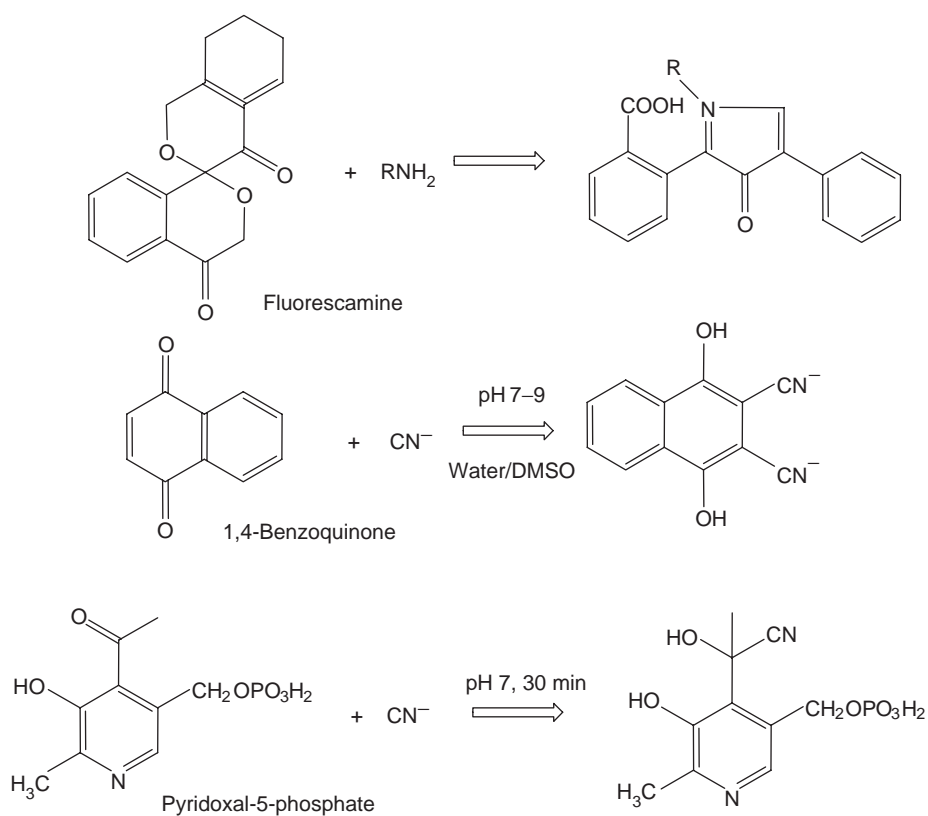
## Labeling Based on Metal Centered Dyes

The use of conventional fluorophores in biological fluids and samples may lead to serious limitations in sensitivity due to the natural fluorescence from proteins and other matrix compounds present. Several lanthanoid fluorescent chelates have spectrochemical features that typical fluorophores do not have: long lifetimes exceeding 1 ms in some cases, large Stokes shifts (over 250 nm), and very sharp peak profiles. The background signals from biological samples being short-lived, the binding of lanthanoid labels to proteins, nucleic acids, and many other molecules, and the use of time-resolved fluorescence techniques may reduce prompt fluorescence interferences.

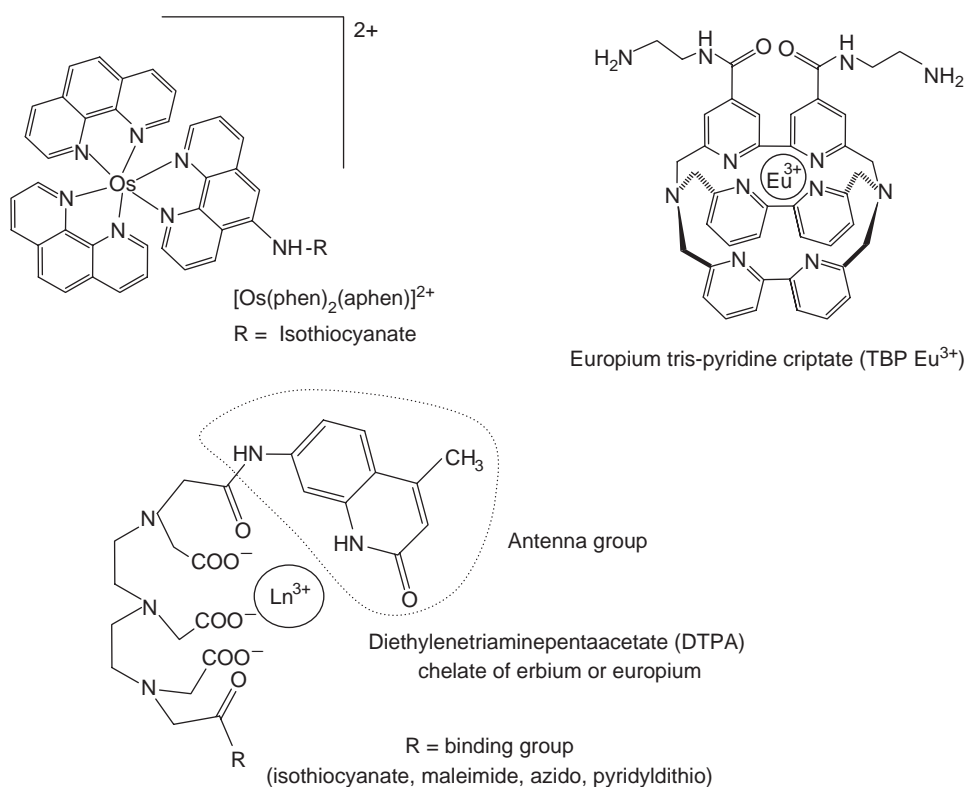
As the emission of radiation occurs from the lowest state of the lanthanide metal ion the chelator must have an excited triplet state with an energy level higher than the lowest emissive level of the metal.

Fluorescent lanthanoid chelates absorb radiation at the wavelength characteristic of the chelator and emit with the wavelength characteristic of the metal, due to energy transfer from the ligand to the metal ion. Lanthanoid chelates are mainly based on bipyridine, polyaminocarboxylate, and cryptate chelators. A typical antenna attached to some polyaminocarboxylate chelators is carbostyryl 124 (7-amino-4-methyl-2(1*H*)-quinolinone) (see Figure 13), which absorbs at a wavelength of 337 nm and the main emission will be at 615 nm for europium and 543 nm for terbium.

Europium and terbium chelates acting as donors offer many technical advantages: (1) the efficiency of energy transfer is considerably higher compared to conventional fluorescent donors, leading to larger signals and longer measurable distances; (2) because lanthanide chelate emission is generally unpolarized, energy transfer is not dependent on their relative orientation but on the distance between donor and acceptor; (3) due to the long-lived fluorescence, lifetime measurements can be made with high precision; (4) the system can provide highly sensitive detection limits by two to four orders of magnitude relative to those of conventional techniques; (5) the sensitized emission of the acceptor due to energy transfer can



**Figure 12** Fluorescent labeling based on substituent effect.



**Figure 13** Fluorescent metal chelates used as labels.

be measured because it is being fed in by the long-lived lanthanide chelate donor; (6) the ligand field around the metal center prevents quenching by water molecules and the consequent nonradiative loss of energy; (7) photobleaching resistance.

Due to their long lifetime of emission and their spectral properties lanthanide complexes are used in time-resolved fluorescence techniques as well in FRET and time-resolved FRET assays. Acceptors can be conventional organic fluorophores, so the emission is dominated by the acceptor spectral characteristics. Coupling of lanthanide chelates to biomolecules can be carried out using thiol-, amine-, ATP-binding reactive derivatives. Lanthanoid fluorescent chelate labels have been used for to explore processes where distance relationship or supramolecular structures are involved: protein associations, immunoanalysis, nucleic acid hybridization, and DNA polymerase reactions. New potential applications are envisaged, including screening of libraries from automated peptide and combinatorial chemical synthesis.

In addition to lanthanide chelates, metalloporphyrins form another class of potentially interesting fluorescent (and phosphorescent) labels. Their luminescence properties depend on the metal substitution, Pt- and Pd-substituted complexes being the most useful. These labels have high molar extinction coefficients (generally above  $1.5 \times 10^5$ ), lifetimes in the range 1–100  $\mu$ s, and exhibit absorption maximum at  $\sim 400$  nm and long wavelength fluorescence emission maximum above 600 nm. The Pd and Pt complexes have also high phosphorescence quantum yield, even in an aqueous environment at room temperature. Metalloporphyrins are chemically and photochemically stable compounds and proved to be useful for time-resolved fluorescence microscopy as labels for antibodies and (strep)avidin.

## Fluorescent Derivatization Based on Metal Complex Formation

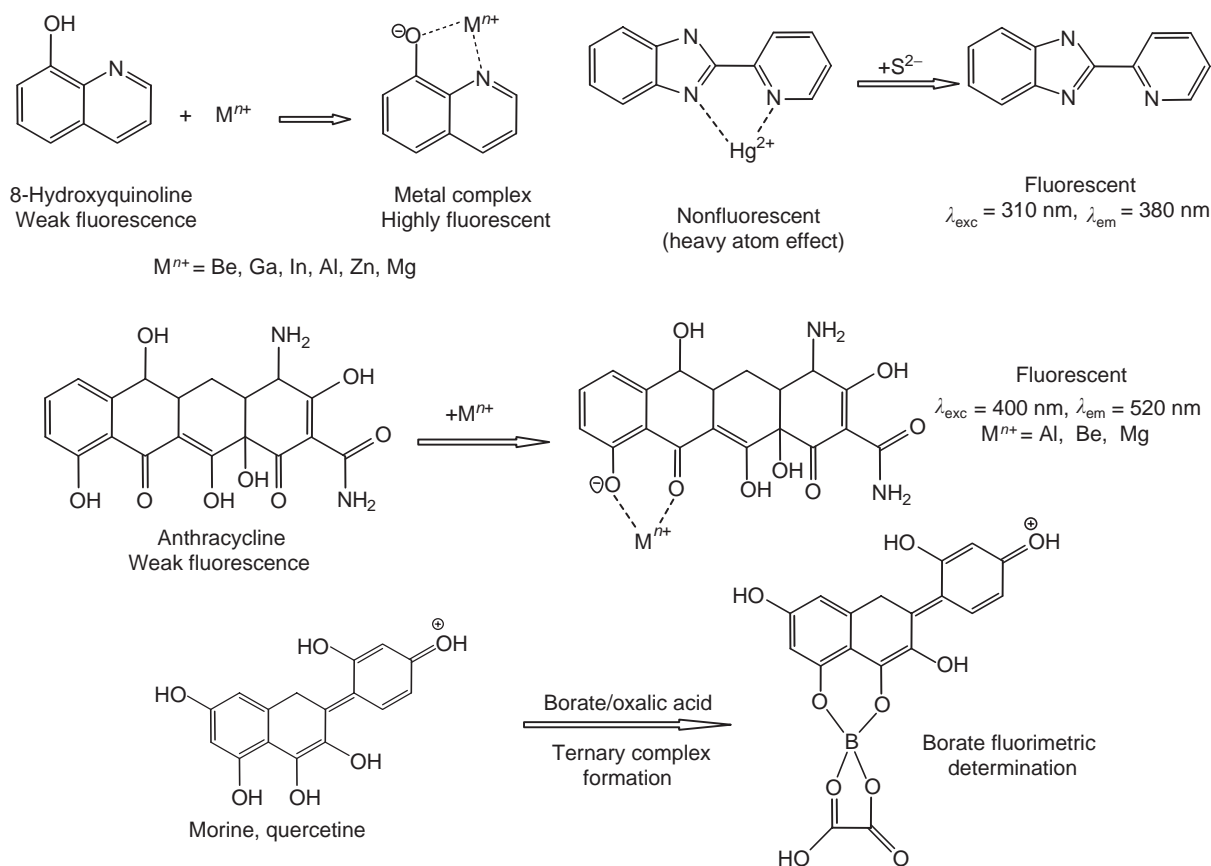
Metal complex formation may cause a nonfluorescent structure to become fluorescent, thus allowing the determination of either metal ion or organic compounds. The process involves the change of  $n - \pi^*$  lowest excited singlets in noncomplexed organic compounds to  $\pi - \pi^*$  lowest excited singlets upon complexation. In Figure 13, a listing of the most widely used reagents is shown. A lot of work has been done in this field with the aim to improve the selectivity for particular metal ions determination, from the synthesis of new reagents to the use of solvent-solvent extraction, and of ordered media

(micelles, cyclodextrins) to protect singlet excited states. For example, during the last decade or so, a number of fluorescent labels have been designed for the analysis of intracellular important metal ions like  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Zn}^{2+}$ , and others. These fluorescent compounds are variations of the nonfluorescent caged chelators EGTA (ethylene glycol-bis-(2-aminoethyl)- $N,N,N',N'$ -tetraacetic acid) and BAPTA (1,2-bis(*o*-aminophenoxy)ethane- $N,N,N',N'$ -tetraacetic acid) and fluorescent signaling units (rhodamine, aminonaphthalene), which combine the metal binding ability with advantageous chemical/photochemical properties. There are many of these fluorescent probes commercially available (fura-2, indo-1, quin-2, fluo-3, rhod-2, etc.) as salt, acetoxymethyl ester, or dextran forms. Salt and dextran forms are typically microinjected or scrape loaded into cells while the ester derivatives are cell membrane permeant. These derivatives are converted by cellular esterases into the free-acid impermeant form of the dye. The trapped dye is then able to chelate the metal ion (e.g., calcium, magnesium). The fluorescence enhancement and/or the luminescence spectral shifts upon metal binding are used for detecting the metal ion concentration and for imaging the ion levels by fluorescence imaging microscopy with high sensitivity and selectivity.

Some metal ions do not form fluorescent complexes due to their paramagnetic nature (e.g.,  $\text{Fe}^{3+}$ ,  $\text{Co}^{2+}$ ) or due to the heavy atom effect (e.g.,  $\text{Pb}^{2+}$ ,  $\text{Hg}^{2+}$ ) both of which promote intersystem crossing, thus deactivating the excited singlet. While fluorescence methods for metal ions are highly sensitive and have wide linear dynamic ranges, the Achilles' heel of the approach is the lack of selectivity, thus making competing techniques such as atomic absorption and inductively coupled plasma the methods of choice. In contrast, it is possible to analyze nonfluorescent organic compounds by forming fluorescent complexes with metal ions. Although this approach offers analytical potential for the analysis of nonfluorescent aromatic compounds, relatively little attention has been paid to it. Typical examples of this approach include the fluorescent analysis of antibiotics and vitamins (e.g., anthracyclines) (Figure 14).

## From Solution Labels to Microarray

During the last 15 years there has been a renaissance of fluorescence, mainly driven by: (1) recent advances in fluorescence detection techniques; (2) the availability of powerful data acquisition and data analysis systems; (3) the development of an array of improved fluorescent dyes and labels; and (4) the increasing



**Figure 14** Fluorescent metal complex formation.

demand for understanding the fundamental principles behind life sciences, specially in genetics, proteomics, molecular, and cell biology. The present account illustrates the great variety of fluorescent labeling approaches that are possible for obtaining chemical and biochemical information. Although not specifically addressed in this article, a wide range of physical and photophysical parameters (fluidity, membrane potential, viscosity, polarity, dielectric constant, lifetime, quantum yields, excited states energy) are also measurable, given the appropriate environmental sensitive fluorescent probe (e.g., 1-anilino-8-naphthalene sulfonate, pyrene, 7-alkoxycoumarin, 6-propionyl-2-(dimethylamino)naphthalene), and using diverse fluorescence techniques such as steady-state methods, time-resolved fluorescence polarization, fluorescence quenching, excimer formation, or energy transfer.

Fluorescence labeling provides a powerful tool for the selective and sensitive determination of a plethora of chemical and biochemical species in a variety of conditions, including live cells, in real time. Immobilization of a fluorescent label on a solid support, such as microparticles or beaded polymer resins, have found interesting applications in the

development of fluorescent optical sensors, in flow-cytometric analysis, in fluorometric microvolume assays, for testing of multiple analytes in one reaction, and high-throughput screening approaches (sensor arrays). For small molecules or short peptides, immobilization often requires covalent linkage of the compounds onto the solid support, while for larger molecules binding may be accomplished by either covalent linkage or noncovalent adsorption. Covalent attachment by methods as those described above is conventionally used. For instance, amine-containing small molecules or proteins can be covalently attached to aldehyde derivatized glass slides to form Schiff's base linkages. Thanks to the possibility of immobilizing thousands of small molecules, peptides, and/or proteins in a small area on a solid-support, to the quality of the chemical chips, and the instrumentation, chemical microarray technology using fluorimetric detection methods has resurfaced as a popular research approach, especially in the fields of genomics and proteomics. It is expected that the development of new ways of introducing fluorescent labels into solid support, particles, proteins, and other biological molecules and the continued development of new enhanced fluorimetric probes



will further expand the applicability of fluorescence techniques and chemical microarray into inexpensive and highly efficient medical diagnostics and environmental monitoring tools in the future.

See also: **Fluorescence**: Overview; Multidimensional Fluorescence Spectrometry; Time-Resolved Fluorescence; Derivatization; Quantitative Analysis; Environmental Applications; Food Applications.

## Further Reading

Epstein JR, Biran I, and Walt DR (2002) Fluorescence-based nucleic acid detection and microarrays. *Analytica Chimica Acta* 469: 3–36.

Haugland RP (2002) *Handbook of Fluorescent Probes and Research Products*, 9th edn. Eugene OR: Molecular Probes.

<http://laxmi.nuc.ucla.edu:8248/M248.98/synprob/lec5.html>  
Mayer A and Neuenhofer S (1994) Luminescent labels – more than just an alternative to radioisotope? *Angewandte Chemie International Edition English* 33: 1044–1072.

Tyagi S and Kramer FR (1996) Molecular beacons: probes that fluoresce upon hybridization. *Nature Biotechnology* 14: 303–308.

Valeur B (1993) Fluorescent probes for evaluation of local physical and structural parameters. In: Schulman SG (ed.) *Molecular Luminescence Spectroscopy, Part 3*. Chemical Analysis Series, vol. 77. USA: Wiley-Interscience.

## Quantitative Analysis

**J N Miller**, Loughborough University of Technology, Loughborough, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Fluorescence analysis methods are now widely used because of their extreme sensitivity, which provides detection limits at picomolar levels and below, and the great variety of sample presentation methods available. Flowing liquids, solid surfaces, concentrated solutions, and suspensions can all be studied in addition to measurements in dilute solution. This article is concerned with the quantitative relationships underpinning steady-state fluorescence measurements: time resolved fluorescence is considered separately.

## Quantitation in Dilute Solution

Since the absorption of photons is the primary process in a fluorescence experiment, the starting point for quantitative studies must be the Beer–Lambert law. For a single absorbing solute, this takes the form:

$$\log_{10} \frac{I_0}{I_t} = \epsilon bc = A \quad [1]$$

Here  $I_0$  and  $I_t$  are, respectively, the intensities of light incident on the sample and transmitted by the sample,  $\epsilon$  is the molar absorptivity of the solute (with units of  $\text{l mol}^{-1} \text{cm}^{-1}$ ),  $b$  is the optical path length in

the sample cell (cm), and  $c$  is the molar concentration of the solute ( $\text{mol l}^{-1}$ ). The product  $\epsilon bc$  is the absorbance,  $A$ , of the sample. This equation can be re-written:

$$I_t = I_0 10^{-\epsilon bc} \quad [2]$$

In fluorescence experiments, interest lies in the amount of light absorbed,  $I_a$ . If the assumption is made that all the light not transmitted is absorbed (i.e., scattered and stray light phenomena are ignored):

$$I_a = I_0 - I_t = I_0(1 - 10^{-\epsilon bc}) \quad [3]$$

Only a fraction of the molecules electronically excited by the absorption of photons subsequently emit fluorescence photons. This fraction is known as the quantum yield,  $\phi_f$  ( $0 \leq \phi_f \leq 1$ ). The intensity of fluorescence,  $I_f$ , is thus given by:

$$I_f = I_a \phi_f = I_0 \phi_f (1 - 10^{-\epsilon bc}) \quad [4]$$

Expansion of this equation gives:

$$I_f = I_0 \phi_f \left\{ 1 - \left( 1 - 2.303 \epsilon bc + \frac{(2.303 \epsilon bc)^2}{2!} - \frac{(2.303 \epsilon bc)^3}{3!} + \dots \right) \right\} \quad [5]$$

If the absorbance,  $\epsilon bc$ , is small – say, less than 0.02 (see below) – then all the terms in the expansion after  $-2.303 \epsilon bc$  are assumed to be negligible, giving:

$$I_f = 2.303 I_0 \phi_f \epsilon bc \quad [6]$$

will further expand the applicability of fluorescence techniques and chemical microarray into inexpensive and highly efficient medical diagnostics and environmental monitoring tools in the future.

See also: **Fluorescence**: Overview; Multidimensional Fluorescence Spectrometry; Time-Resolved Fluorescence; Derivatization; Quantitative Analysis; Environmental Applications; Food Applications.

## Further Reading

Epstein JR, Biran I, and Walt DR (2002) Fluorescence-based nucleic acid detection and microarrays. *Analytica Chimica Acta* 469: 3–36.

Haugland RP (2002) *Handbook of Fluorescent Probes and Research Products*, 9th edn. Eugene OR: Molecular Probes.

<http://laxmi.nuc.ucla.edu:8248/M248.98/synprob/lec5.html>  
Mayer A and Neuenhofer S (1994) Luminescent labels – more than just an alternative to radioisotope? *Angewandte Chemie International Edition English* 33: 1044–1072.

Tyagi S and Kramer FR (1996) Molecular beacons: probes that fluoresce upon hybridization. *Nature Biotechnology* 14: 303–308.

Valeur B (1993) Fluorescent probes for evaluation of local physical and structural parameters. In: Schulman SG (ed.) *Molecular Luminescence Spectroscopy, Part 3*. Chemical Analysis Series, vol. 77. USA: Wiley-Interscience.

## Quantitative Analysis

**J N Miller**, Loughborough University of Technology, Loughborough, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Fluorescence analysis methods are now widely used because of their extreme sensitivity, which provides detection limits at picomolar levels and below, and the great variety of sample presentation methods available. Flowing liquids, solid surfaces, concentrated solutions, and suspensions can all be studied in addition to measurements in dilute solution. This article is concerned with the quantitative relationships underpinning steady-state fluorescence measurements: time resolved fluorescence is considered separately.

## Quantitation in Dilute Solution

Since the absorption of photons is the primary process in a fluorescence experiment, the starting point for quantitative studies must be the Beer–Lambert law. For a single absorbing solute, this takes the form:

$$\log_{10} \frac{I_0}{I_t} = \epsilon bc = A \quad [1]$$

Here  $I_0$  and  $I_t$  are, respectively, the intensities of light incident on the sample and transmitted by the sample,  $\epsilon$  is the molar absorptivity of the solute (with units of  $\text{l mol}^{-1} \text{cm}^{-1}$ ),  $b$  is the optical path length in

the sample cell (cm), and  $c$  is the molar concentration of the solute ( $\text{mol l}^{-1}$ ). The product  $\epsilon bc$  is the absorbance,  $A$ , of the sample. This equation can be re-written:

$$I_t = I_0 10^{-\epsilon bc} \quad [2]$$

In fluorescence experiments, interest lies in the amount of light absorbed,  $I_a$ . If the assumption is made that all the light not transmitted is absorbed (i.e., scattered and stray light phenomena are ignored):

$$I_a = I_0 - I_t = I_0(1 - 10^{-\epsilon bc}) \quad [3]$$

Only a fraction of the molecules electronically excited by the absorption of photons subsequently emit fluorescence photons. This fraction is known as the quantum yield,  $\phi_f$  ( $0 \leq \phi_f \leq 1$ ). The intensity of fluorescence,  $I_f$ , is thus given by:

$$I_f = I_a \phi_f = I_0 \phi_f (1 - 10^{-\epsilon bc}) \quad [4]$$

Expansion of this equation gives:

$$I_f = I_0 \phi_f \left\{ 1 - \left( 1 - 2.303 \epsilon bc + \frac{(2.303 \epsilon bc)^2}{2!} - \frac{(2.303 \epsilon bc)^3}{3!} + \dots \right) \right\} \quad [5]$$

If the absorbance,  $\epsilon bc$ , is small – say, less than 0.02 (see below) – then all the terms in the expansion after  $-2.303 \epsilon bc$  are assumed to be negligible, giving:

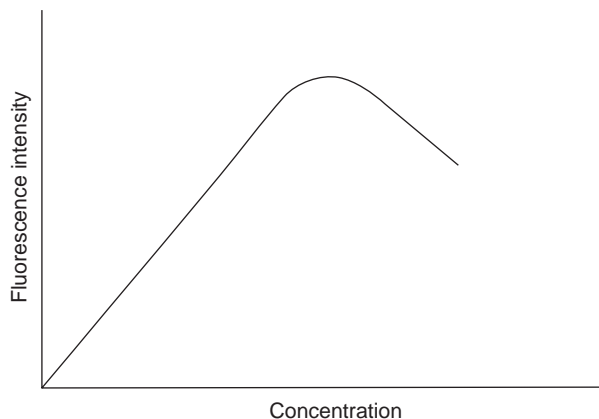
$$I_f = 2.303 I_0 \phi_f \epsilon bc \quad [6]$$

In practice, however, only a small fraction,  $K$ , of the emitted photons is detected in most instruments, so this extra term should be included in the equation for the measured fluorescence:

$$I_f = 2.303I_0\phi_f K\epsilon bc \quad [7]$$

This equation deserves close study. High fluorescence intensities depend on the solute having a high molar absorptivity,  $\epsilon$ , as well as a high quantum yield,  $\phi_f$ . The product  $\epsilon\phi_f$  has been used as a measure of the sensitivity of a solute in fluorescence work, though in reality much will depend on the wavelengths of excitation and emission, the Stokes shift (see below), and the instrumentation used. Two solutes with similar  $\epsilon\phi$  values will not necessarily have similar limits of detection if their excitation and emission wavelengths are different. Fluorescence intensity is also proportional to the incident light intensity, so light sources should be bright and stable. However, most solutes suffer from photodecomposition to a greater or lesser extent, and in practice limits of detection are usually controlled by background signals (see below), so the use of brighter light sources will not be beneficial *ad infinitum*. Anything that can be done to increase the light collection factor  $K$  is beneficial. Finally, the equation shows that the fluorescence intensity should be proportional to concentration, which is obviously desirable when establishing calibration plots for quantitative analysis.

In practice, when fluorophore solutions of increasing concentration are studied the calibration plot will be similar to that shown in Figure 1. At low concentrations the fluorescence intensity is proportional to the concentration as expected, but at higher concentrations negative deviations from linearity occur, and eventually the curve may 'roll over'. The latter effect only occurs at such high levels that it is unlikely that an error would be made in using the

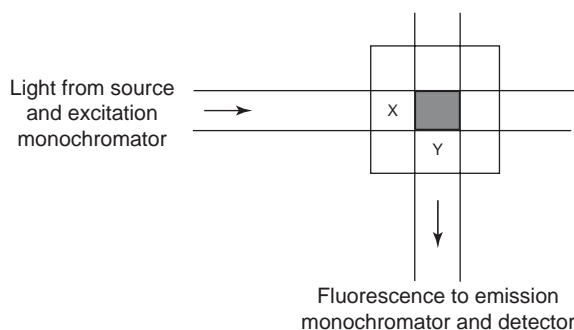


**Figure 1** Typical relationship between fluorescence intensity and concentration for a single fluorophore in solution.

graph to estimate concentrations in a calibration experiment, but the reasons for the lack of linearity demand examination. There are three distinct explanations for a nonlinear response.

The first factor is the mathematical approximation involved in deriving eqns [6] and [7]. If the absorbance of the sample at the excitation wavelength used is, for example, 0.02, then in eqn [5] the term  $2.303\epsilon bc$  has a value of 0.04606. The next term,  $-(2.303\epsilon bc)^2/2!$ , has a value of  $-0.00106$ , i.e., it reduces the effect of the first term, and thus the fluorescence intensity, by  $\sim 2.3\%$  although it is ignored in eqns [6] and [7]. Similarly, if the absorbance reaches 0.05, the corresponding error is  $(- )5.5\%$ . In fluorescence measurements in a 1 cm cuvette the relative standard deviation of replicate intensities for the same sample is commonly about 2%. This figure can be improved substantially by the use of a constant temperature (see below), extreme cleanliness, and care in positioning the cuvette in the light beam (e.g., always using the same faces of a single cuvette). Better repeatability is also usually obtained in the detection of flowing liquids, e.g., in flow injection analysis. It is therefore recommended that the absorbance of the sample at the excitation wavelength should not exceed 0.02, more concentrated samples being diluted if necessary. This is not a serious drawback to the technique. A solute having a relative molecular mass of 500 and a molar absorptivity,  $\epsilon$ , of  $10^4 \text{ l mol}^{-1} \text{ cm}^{-1}$ , gives an absorbance of 0.02 in a 1 cm cuvette at a concentration of  $1 \mu\text{g ml}^{-1}$ . Most fluorescence analyses are performed at concentrations well below this level, so the upper limit on the absorbance value is not of concern.

However, the absorbance of a sample also has an important bearing on inner filter effects, which in practice are the major causes of curved fluorescence calibration plots. Inner filter effects arise for the reasons summarized in Figure 2. In most instruments light from the exit slit of the excitation monochromator falls directly on to the cuvette, in some cases with the aid of a focusing lens. The result is that the beam illuminating the sample is much narrower than the cuvette itself. (For simplicity, this light beam is assumed to be parallel in Figure 2.) Similarly, the entrance slit of the emission monochromator effectively interrogates only a narrow section of the cuvette, at right angles to the illuminated region. So fluorescence is only excited and observed in a small central region of the cuvette. This arrangement has some merits. For example, photodecomposed sample molecules may diffuse out of this region, to be replaced by fresh ones, thus reducing the measured effects of any photodecomposition phenomena. But its drawback is that optical artifacts can occur in the

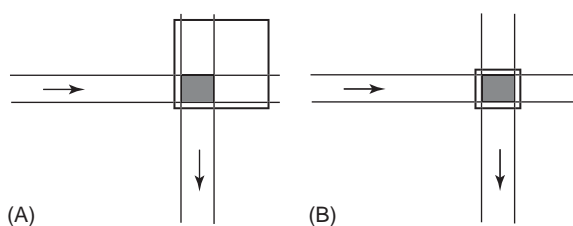


**Figure 2** Origin of inner filter effects in solution fluorimetry. Exciting and emitting light beams intersect only in the center of the cuvette (shaded). Pre- and postfilter effects occur in regions X and Y, respectively.

regions of the cell marked X and Y. In region X, photons can be absorbed by the solute (or other sample components). The light intensity reaching the central region of the cuvette is thus reduced, and according to eqn [7] the fluorescence emitted is also reduced. If the absorbance of the sample at the excitation wavelength is very high few photons will reach the central region and little or no fluorescence will be seen: this is the cause of the 'roll-over' effect described above. In region Y, fluorescence photons emitted from the central region of the cuvette can be re-absorbed, either by other solute molecules (because the excitation and fluorescence spectra of a given material often overlap) or by other components of the sample. Again, the result is a reduction in the observed fluorescence. Equations have been derived that detail such effects, and they have even been used to estimate quantum yields (see below). In practice, however, it is much more important to minimize inner filter effects by ensuring that samples have low absorbances at the excitation and emission wavelengths in use.

Inner filter effects can also be reduced by minimizing the dimensions of the regions X and Y, either by using short pathlength cuvettes, or by off-setting the cuvette in the light beam so that its corner rather than its center is illuminated and observed by the instrument optics (**Figure 3**). These methods run the risk that the excitation or emission beams might strike the side-walls of the cuvette, increasing the scattered light interference. Flow cells used for high-performance liquid chromatography or flow injection detection normally have a short pathlength, and so do not show significant inner filter effects, but they have to be carefully matched to the optical characteristics of the instrument if this scattered light problem is to be avoided.

The third reason for negative curvature in fluorescence calibration graphs is the occurrence of



**Figure 3** Cuvette arrangements to minimize inner filter effects: (A) use of corner of the cuvette and (B) use of a short pathlength cuvette.

bimolecular processes. Many organic fluorophores are large planar molecules with extensive  $\pi$ -electron systems, and as their concentration increases they tend to aggregate through  $\pi$ - $\pi$  interactions. Such aggregates often have reduced or different fluorescence properties compared with the monomeric molecules. In particular, the formation of excimers (excited state dimers) with a rather weaker fluorescence at longer emission wavelengths than the monomers is quite frequently observed. This is an example of self-quenching, i.e., the reduction of the fluorescence of one molecule by another identical one. Such phenomena do not normally occur in very dilute solutions (though some long-wavelength fluorophores do aggregate even at sub-ppm levels), so a linear relationship between fluorescence and concentration would normally be expected for materials studied at sub-micromolar levels. Sadly, the effects described above are often poorly explained and named in the fluorescence literature. Inner filter effects are often referred to as self-quenching (though no molecule-molecule interactions are involved), and the mathematical approximation problem is sometimes called an inner filter effect, even when there is no such optical artifact present.

## Quantitation for Other Sample Types

Although fluorescence methods are frequently applied to the analysis of very dilute solutions, their uses with other sample types are often important. One example is the situation that arises when the concentration of the fluorescent solute is so high that virtually all the light in a given wavelength range striking the sample is absorbed. In that case the intensity of fluorescence emitted is independent of concentration. If the quantum yield and the emission spectrum of the fluorophore are independent of the excitation wavelength, the intensity of the emission is proportional to the intensity of the incident light, and the sample functions as a quantum counter. Such devices are used in spectral correction systems and other areas of photochemistry.

Fluorescence measurements on solid surfaces are extremely important, as the method is used for scanning thin-layer chromatography plates, zone electropherograms, solid-phase immunoassay media, etc. Room-temperature phosphorescence is also often observed at solid surfaces. The optical situation in such systems is more complex than that in dilute solution for several reasons. Emitted fluorescence can be detected in reflection or transmission modes; the solid phase may exhibit its own light absorption and/or fluorescence properties; solid-phase thickness and the presence of any backing material will affect the results; scattering and reflection of the incident light is substantial; and the angle of incidence of the exciting light beam is critical. It is therefore not surprising that in most theoretical studies simplifying approximations are made. For example, an angle of incidence of  $90^\circ$  and uniformity of the solid phase are commonly assumed. A number of equations have been derived by various authors, but there is general agreement that the signals emitted from a given illuminated area by low levels of fluorescent material are proportional to the amount of fluorophore present, and also proportional to the intensity of the incident light. (In these respects surface fluorescence and fluorescence from dilute solutions are similar.) The intensity of the reflected fluorescence is higher than that of the transmitted fluorescence, but the linear range will be poorer for the former than the latter.

### Limits of Detection in Fluorescence Spectroscopy

The ability of fluorescence methods to detect very low concentrations derives from the basic photochemistry involved in the measurement process. Fluorescence is emitted at longer wavelengths than that of the exciting light, and can be observed without serious interference from the latter. The detected signal is usually only a fraction of the total emitted fluorescence, but detecting it against a 'dark' background using the  $90^\circ$  optics of fluorescence spectrometers is critical to the attainment of low limits of detection. By contrast, absorptiometric measurements at trace levels rely on the detection of a small difference between two signals of similar high intensity, i.e., the signal reaching the detector when only the solvent blank is present, and the corresponding signal when low levels of analyte are present. A modern absorption spectrometer carefully used might detect an absorbance,  $A$  (or an absorbance difference), of 0.0001. For a solute with a high molar absorptivity,  $\epsilon$ , of  $10^5 \text{ l mol}^{-1} \text{ cm}^{-1}$  this corresponds to a detection limit of  $10^{-9} \text{ mol l}^{-1}$  ( $1 \text{ nmol l}^{-1}$ ).

Theoretical estimates and experimental evidence agree that a fluorophore with this  $\epsilon$  value and a quantum yield,  $\phi_f$ , of 0.50 could be detected in a good fluorescence spectrometer at levels of  $1 \text{ pmol l}^{-1}$  or less: the fluorescence method gives limits of detection at least three orders of magnitude lower (better) than the absorption spectrometry method.

### Background Signals and Their Reduction

In practice, low limits of detection are only obtained if the background signal (i.e., the signal in the absence of the fluorophore) is as close to zero as possible. The limit of detection for a solute (determined by any methodology) is conventionally defined as the concentration that gives a signal three standard deviations above the background level (some authorities use a factor of two rather than three). In practice, as a solution of a fluorophore is progressively diluted and measured, there comes a point where the observed signal is not significantly different from the background. In such conditions instrumental sensitivity improvements such as the use of laser sources or photon counting detectors may be of no help. The background signal is often much worse with 'real' samples (blood plasma, urine, environmental materials, etc.) than with pure solutions of the fluorophore, and much attention has been paid to the need to minimize it by optical and/or chemical means. Similar problems arise in the analysis of fluorophore mixtures. Unless the excitation and/or emission wavelengths of the mixture components are well separated, analysis of one component will be complicated by background signals from the others. Some of the techniques used for background reduction are thus also applicable to mixture analysis.

A major potential source of background signals is scattered light. For materials in dilute solution Rayleigh scattering is the dominant phenomenon. It results from elastic photon-molecule collisions, so occurs at the same wavelength as the incident light. But fluorescence instruments conventionally use monochromators with pass-bands (measured at half-maximum intensity) of 5–10 nm, so when fluorophores with small Stokes shifts (emission minus excitation maximum wavelengths) are studied the possibility of interference is very real. A good example is the common fluorophore fluorescein, which has a Stokes shift of only  $\sim 23 \text{ nm}$  ( $\lambda_{\text{ex}} 490 \text{ nm}$ ;  $\lambda_{\text{f}} = 513 \text{ nm}$  at pH 9). Other widely used xanthine dyes suffer from the same problem, and many long-wavelength dyes have even smaller Stokes



shifts. Possible remedies include the use of polarizing filters (since Rayleigh scattered light is strongly polarized), or cut-off filters to supplement the performance of the emission monochromator. Probably the simplest remedy is to use an excitation wavelength that is suboptimal in terms of fluorescence intensity, but results in much reduced scattered light. For example, exciting fluorescein (pH 9) at  $\sim 480$  nm causes a loss of fluorescence intensity of 25% compared with the optimum wavelength of 490 nm. But the reduced scattered light signal in the emission beam might well more than compensate for this.

Scattering by the Raman mechanism is a very weak phenomenon in solution, so that only solvent Raman bands are observed, but they can nonetheless present a problem. For any solvent, Raman bands occur at a predictable frequency. The main Raman band of water is at  $\sim 3400\text{ cm}^{-1}$ , so for excitation wavelengths of (for example) 280, 350, and 490 nm, the corresponding Raman emissions are at 309, 397, and 588 nm, respectively. (The signal-noise ratio of the water Raman band excited at 350 nm is often used as a criterion of fluorescence spectrometer sensitivity.) Examples of problems caused by Raman scattering are studies of tyrosine and its derivatives in aqueous solutions, using an excitation wavelength of 280 nm. The water Raman band at 309 nm strongly overlaps with the tyrosine fluorescence at *c.* 307 nm (and Rayleigh interference may also occur). In this and other cases it is not usually possible to remove the effects of scattered light by running a blank and subtracting it from the sample fluorescence. Scattering is enhanced by the presence of proteins and other macromolecules in the sample, so it is impossible to provide a realistic blank. There should not be any confusion between Raman bands and fluorescence signals, partly because Raman shifts are predictable, and partly because of a simple test that will distinguish them immediately. Shifting the excitation wavelength by (say) 10 nm will cause a shift in the wavelength of the Raman band, but no shift in the wavelength of fluorescence, though the intensity of the latter may change. The intensities of Rayleigh and Raman scattered light become much less in the near-infrared region of the spectrum, as they vary as the inverse fourth power of the incident light wavelength. Moreover, the solvent Raman shifts in this region are very large in wavelength terms. Excitation of an aqueous solution at 650 nm produces a Raman band at 834 nm, in contrast with the very small Stokes shifts of typical long-wavelength fluorophores. This reduction of scattered light interference is one of the major reasons for the much increased use of long-wavelength fluorescent labels

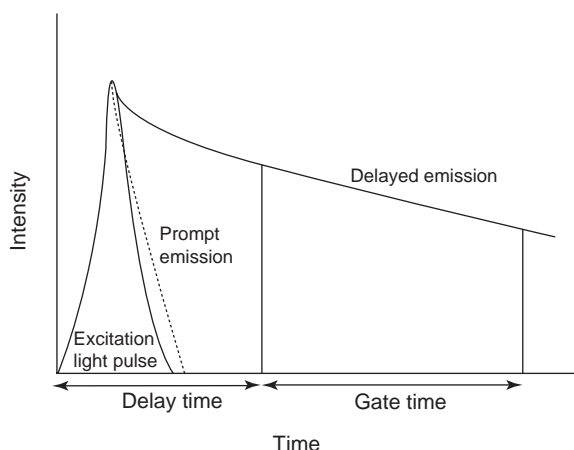
in recent years. (Raman spectroscopists also favor working in the near-infrared region: there is reduced fluorescence background for their samples, as there are not many long-wavelength fluorophores.)

Solvents and other materials used in fluorescence measurements are further sources of background signals. Some water samples contain parts per billion levels of strongly fluorescent aromatic hydrocarbons and other fluorophores. These contaminants are not removed by conventional purification methods involving ion exchange, so extra purification steps are necessary. Organic solvents are even more prone to such interferences (many of them come originally from oil fractions). Some plastic materials give background fluorescence problems, from plasticizers or transition metal compounds used in polymerization processes. Careful choice of solvents and reagents, and extreme cleanliness and care in laboratory procedures are essential if the full sensitivity of fluorescence is to be achieved.

Many natural samples show strong 'autofluorescence' signals. The fluorescence of typical blood plasma specimens and of urine is well documented, for example. In some cases fluorescence analyses of components of such matrices are carried out only after a separation or extraction process, so that the final determination is performed on an essentially pure, or much purified, compound. In many applications (for example, in enzyme- and antibody-based assays) this is entirely impractical, so alternative methods of background reduction are used. Time-gated systems in conjunction with labels or probes based on long-lived fluorophores or phosphors are now common. The instrumentation involved is summarized in a separate article, and the principle is illustrated in **Figure 4**. The sample is illuminated with pulses of light 10–20  $\mu\text{s}$  long, and a gating electronics system allows the detector to start recording the emission after each light pulse has ended. Any fluorophore with a lifetime of  $\sim 100\mu\text{s}$  or more can thus be detected without background interference from the rapidly decaying scattered light and conventional fluorescence signals, the latter having typical lifetimes of only a few nanoseconds. This approach has encouraged the development of analytical methods based on the use of lanthanide ions or ruthenium complexes as labels.

The reduction of autofluorescence is another major reason (alongside the ready availability of suitable solid-state optical components) for the increased use of long-wavelength fluorescence. Fluorescent labels, probes, and enzyme substrates with excitation wavelengths in the 600–800 nm region, where there are very few naturally occurring fluorophores, provide methods with very low background signals.





**Figure 4** Use of time-gated detector system to minimize background effects in fluorescence. The detector does not respond during the delay time, but only during the gate time: by then the source pulse and the scattered light and prompt fluorescence resulting from it are not detected. The gate time ends before the next source pulse begins.

Other spectroscopic methods for background reduction include the use of derivative fluorescence spectra and synchronous scanning. Derivative spectra (the first or second derivatives are mostly used) are readily calculated using PC-interfaced spectrometers, and in principle offer a number of possibilities. For example, the fluorescence of one component in a two-fluorophore mixture can be studied without interference from the second component at the maximum fluorescence wavelength of the latter, where its first derivative spectrum has zero intensity. Synchronous scanning is described elsewhere: it produces simpler and narrower spectra with reduced overlap between signals from different fluorophores, and thus may also help in reducing background effects. The disadvantage of these approaches is that they require spectral scans of every sample, rather than measurements at a single pair of excitation and emission wavelengths. Scanning spectrometers or array detectors are thus essential, and each sample may take longer to analyze. In recent years, however, the use of diode array detectors with chemometric data handling (e.g., partial least squares) methods has allowed the rapid simultaneous analysis of two or more fluorophores, even when their emission spectra are very similar.

## Fluorescence Quenching Methods

Many analytical procedures have been developed in which the analyte is a quencher of the emission of a fluorescent reagent. One application area is the detection and determination of nitro compounds such

as carcinogenic nitro-aromatics, explosives. The detection of oxygen through fluorescence quenching in simple sensors is another important application. The quantitative aspects of such studies are quite different from those of other fluorescence measurements. The principal equation describing dynamic quenching processes (those in which the excited state of a fluorophore interacts with the quencher: is the Stern–Volmer equation:

$$\frac{\phi_0}{\phi} - 1 = k_Q \tau_0 [Q] \quad [8]$$

Here,  $\phi_0$  and  $\phi$  are the quantum yields of the fluorophore in the absence and presence, respectively, of the quencher  $Q$ ,  $k_Q$  is the quenching constant,  $\tau$  is the fluorescence lifetime of the fluorophore in the absence of the quencher, and  $[Q]$  is the quencher concentration. The product  $k_Q \tau_0$  is sometimes called the Stern–Volmer constant,  $k_{SV}$ . It has a typical value of  $\sim 100 \text{ l mol}^{-1}$  for conventional fluorophores in solution at room temperature. If a quenching effect of 1% is the smallest that could be detected, i.e.,  $(\phi_0/\phi - 1) = 0.01$ , then from [8] the lowest concentration of  $Q$  detectable is  $10^{-4} \text{ mol l}^{-1}$ . Some quenching interactions have been found with  $k_{SV}$  values significantly greater than 100 (e.g., using long-lifetime fluorophores or phosphors). In other cases arrays of dozens of individual sensors have been used, with signal processing methods applied to evaluate the modest quenching effects. But quenching analyses are evidently much less sensitive than conventional fluorescence methods. This is, of course, because all quenching methods rely on detecting a small reduction in a large fluorescence signal: in sensitivity terms they are thus analogous to absorption spectrometry methods.

## Other Quantitative Methods

Amongst the many applications of fluorescence several other quantitative approaches are used. Many types of immunoassay using fluorescent labeled antibodies or antigens have been developed. In some cases the intensity of fluorescence measured should be proportional to the amount of analyte present, but in others increasing analyte levels give rise to diminishing fluorescence signals. Moreover, fluorescence polarization may be measured rather than a straightforward intensity. Enzyme activities are often determined using fluorogenic substrates, non-fluorescent, or weakly fluorescent species converted by enzyme action to strongly fluorescent products. The presence of inhibitors will cause a loss of fluorescence intensity, so these materials can also be

determined. In all these analyses the use of constant temperatures is obligatory, as the activities of the biomolecules are temperature dependent, as are any fluorescence signals generated.

## Temperature Effects

The effects of temperature on the fluorescence of individual solutes have been studied repeatedly. In virtually all cases fluorescence intensities fall as the temperature rises, the higher frequency of bimolecular collisions and higher energy vibrational modes increasing the probability of nonradiative energy loss from excited molecules. The radiative properties of the excited state are also affected indirectly by temperature dependent changes in the refractive index of the solvent, and the latter influences the geometry of the fluorescence emitted from the sample cuvette. The absorbance of the sample is also likely to be affected by temperature changes, though to a less extent than the fluorescence quantum yield. It is to be expected that the effect of temperature on the latter might take an exponential form, and that it would become greater as the energy difference between the ground state and the first excited singlet state of the molecule decreases. Equations of the following form have been suggested:

$$\frac{1}{\phi_f} - 1 = k_1 e^{-E/RT} + k_2 \quad [9]$$

In this equation,  $k_1$  and  $k_2$  are constants and  $E$  is an activation energy related to both the structure of the solute and the viscosity of the solvent (which is also temperature dependent!). In practice, perhaps because of the numerous effects summarized above, the following results have been obtained:

- Over limited temperature ranges (e.g., 15–50°C) the observed fluorescence intensity for many species is approximately inversely proportional to temperature.
- For different fluorophores the intensity change may lie in the range  $c. (-)1$ –5% per °C. For example, tryptophan has a temperature coefficient of  $\sim -3.5\%$  per °C over the range 20–30°C, while for quinine the coefficient is closer to  $-2\%$  per °C.
- Molecules with apparently similar fluorophores but with different substituents often show markedly different temperature coefficients. Internal flexibility and steric hindrance effects limiting vibrations may be responsible.
- There is no evidence that long-wavelength fluorophores, despite small ground state–excited state energy differences, show large temperature dependence effects.

Maintaining the temperature of the sample cuvette with variations of not more than  $\pm 0.1^\circ\text{C}$  should ensure that temperature effects are negligible compared with other sources of variation. Fluorescence polarization values are particularly susceptible to temperature effects and control at the level of  $\pm 0.05^\circ\text{C}$  or better may be desirable.

## Fluorescence Quantum Yields

Equations [4]–[7] include the term  $\phi_f$ , the quantum yield of fluorescence, and it is evident that, other factors being equal, fluorophores with high quantum yields will provide more sensitive analyses. Moreover, quantum yield values provide important information on electronic transitions in molecules. In practice, however, absolute quantum yield measurements are quite difficult to make, for a variety of technical reasons. Several approaches have been used. The most direct approach involves comparing the intensity of light emitted from a fluorescent solution in a cuvette with that from an ideal scattering solution, such as a colloidal matrix. Such a scatterer can be assumed to scatter all photons incident on it, i.e., to have an effective quantum yield of 1. This method is complicated by the strong polarization of scattered light, and the sensitivity of the detection components in a spectrometer to this phenomenon. A second approach is a calorimetric one, i.e., the energy *not* emitted by the sample as light is measured, for example, by photoacoustic techniques. This method gives results in good agreement with purely optical approaches, but has not been applied frequently as specialized equipment is required. The method most commonly used is a secondary one, in which the fluorophore under study is compared with a standard material of known quantum yield. If the two samples, studied in the same instrumental conditions, have integrated fluorescence intensities (i.e., areas under their corrected fluorescence spectra) of  $S_1$  and  $S_2$ , absorbances at their own excitation wavelengths of  $A_1$  and  $A_2$ , and quantum yields  $\phi_1$  and  $\phi_2$ , then:

$$\frac{S_2}{S_1} = \frac{A_2}{A_1} \frac{\phi_2}{\phi_1} \quad [10]$$

Unhappily, relatively little agreement has been reached on a range of fluorophores covering a range of wavelengths, which can be used as standards. One manufacturer (Molecular Probes<sup>®</sup>) sells a set of reference solutions (quinine in 0.1 mol l<sup>-1</sup> sulfuric acid, fluorescein in 0.1 mol l<sup>-1</sup> aqueous NaOH, 5-carboxytetramethylrhodamine in methanol, sulforhodamine 101 in ethanol, and Nile Blue in ethanol),

which cover the excitation wavelengths  $\sim 350$ – $650$  nm, and emission wavelengths  $\sim 450$ – $670$  nm. The first two of these materials are assigned standard quantum yields of 0.55 and 0.92, respectively. These values would be accepted by most workers in the fluorescence community and can be used to estimate other quantum yields. However, many quantum yield values reported in the literature are of rather doubtful accuracy.

See also: **Fluorescence:** Overview; Instrumentation; Time-Resolved Fluorescence.

## Further Reading

- Ingle JD and Crouch SR (1988) *Spectrochemical Analysis*. Englewood Cliffs, NJ: Prentice Hall International.
- Lakowicz JR (1999) *Principles of Fluorescence Spectroscopy*, 2nd edn. Dordrecht: Kluwer Academic Publishers.
- Miller JN (ed.) (1981) *Standards in Fluorescence Spectrometry*. London: Chapman & Hall.
- Parker CA (1968) *Photoluminescence of Solutions*. Amsterdam: Elsevier.
- Valeur B (2002) *Molecular Fluorescence – An Introduction: Principles and Applications*. New York: Wiley-VCH.

## Clinical and Drug Applications

A Gómez-Hens, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The usefulness of fluorescence spectroscopy for clinical and drug analysis is beyond doubt. In fact, judging by the number of analyses done per day, the most extensive use of this technique is probably in clinical and pharmaceutical analysis. A large number of substances of biomedical interest possess intrinsic fluorescence at either room temperature or low temperatures, which has been widely exploited for performing basic studies and developing a host of analytical methods. Also, many therapeutic and abuse drugs exhibit native fluorescence in organic and neutral solutions or even in strong acids and bases, which has been used for analytical purposes.

Nonfluorescent or weakly fluorescent analytes can be modified to obtain fluorescent products. The changes involved in the process can be of a chemical or physicochemical nature. The former, usually in the form of derivatization reactions, include chemical treatments using redox, complex formation, and substitution reactions or incorporation of a fluorophore through a chemical (e.g., condensation) reaction. Physicochemical modification of a nonfluorescent or weakly fluorescent substance can be accomplished in a number of ways including changes in the molecular environment arising from a solvent changeover, addition of a chemical compound such as a surfactant or cyclodextrin to protect fluorophores from their molecular environment, and energy

transfer processes between fluorescent and nonfluorescent molecules.

A systematic description of the scope of fluorescence spectroscopy in clinical and drug analysis is rendered difficult by the very large number of determinations reported so far. For reasons of space, this article can only present an overview of the wide variety of applications of this technique in these analytical fields, such as chromatographic and electrophoretic methods, its use in enzymatic and immunochemical assays and in biosensors, and its usefulness in new research areas such as proteomic studies.

### Selectivity and Sensitivity of Fluorescence Spectroscopy in Clinical and Drug Analysis

Direct fluorescence measurements in biological samples are hindered by several factors including concentration effects, background effects arising from Rayleigh or Raman scattering, solvent effects (e.g., interfering nonspecific fluorescence and quenching), and sample matrix effects. Thus, a serum or urine sample may contain a variety of fluorescing substances, such as proteins and bilirubin, that are a potential source of unwanted background fluorescence. Light scattering by proteins and lipoproteins can also contribute to this negative effect. Selectivity is also a major problem in the determination of drugs in their dosage forms, which are often multicomponent preparations. The lack of selectivity in fluorescence analysis of clinical and pharmaceutical samples usually calls for a prior sample cleanup step using a separation technique,

which cover the excitation wavelengths  $\sim 350$ – $650$  nm, and emission wavelengths  $\sim 450$ – $670$  nm. The first two of these materials are assigned standard quantum yields of 0.55 and 0.92, respectively. These values would be accepted by most workers in the fluorescence community and can be used to estimate other quantum yields. However, many quantum yield values reported in the literature are of rather doubtful accuracy.

See also: **Fluorescence:** Overview; Instrumentation; Time-Resolved Fluorescence.

## Further Reading

- Ingle JD and Crouch SR (1988) *Spectrochemical Analysis*. Englewood Cliffs, NJ: Prentice Hall International.
- Lakowicz JR (1999) *Principles of Fluorescence Spectroscopy*, 2nd edn. Dordrecht: Kluwer Academic Publishers.
- Miller JN (ed.) (1981) *Standards in Fluorescence Spectrometry*. London: Chapman & Hall.
- Parker CA (1968) *Photoluminescence of Solutions*. Amsterdam: Elsevier.
- Valeur B (2002) *Molecular Fluorescence – An Introduction: Principles and Applications*. New York: Wiley-VCH.

## Clinical and Drug Applications

A Gómez-Hens, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The usefulness of fluorescence spectroscopy for clinical and drug analysis is beyond doubt. In fact, judging by the number of analyses done per day, the most extensive use of this technique is probably in clinical and pharmaceutical analysis. A large number of substances of biomedical interest possess intrinsic fluorescence at either room temperature or low temperatures, which has been widely exploited for performing basic studies and developing a host of analytical methods. Also, many therapeutic and abuse drugs exhibit native fluorescence in organic and neutral solutions or even in strong acids and bases, which has been used for analytical purposes.

Nonfluorescent or weakly fluorescent analytes can be modified to obtain fluorescent products. The changes involved in the process can be of a chemical or physicochemical nature. The former, usually in the form of derivatization reactions, include chemical treatments using redox, complex formation, and substitution reactions or incorporation of a fluorophore through a chemical (e.g., condensation) reaction. Physicochemical modification of a nonfluorescent or weakly fluorescent substance can be accomplished in a number of ways including changes in the molecular environment arising from a solvent changeover, addition of a chemical compound such as a surfactant or cyclodextrin to protect fluorophores from their molecular environment, and energy

transfer processes between fluorescent and nonfluorescent molecules.

A systematic description of the scope of fluorescence spectroscopy in clinical and drug analysis is rendered difficult by the very large number of determinations reported so far. For reasons of space, this article can only present an overview of the wide variety of applications of this technique in these analytical fields, such as chromatographic and electrophoretic methods, its use in enzymatic and immunochemical assays and in biosensors, and its usefulness in new research areas such as proteomic studies.

### Selectivity and Sensitivity of Fluorescence Spectroscopy in Clinical and Drug Analysis

Direct fluorescence measurements in biological samples are hindered by several factors including concentration effects, background effects arising from Rayleigh or Raman scattering, solvent effects (e.g., interfering nonspecific fluorescence and quenching), and sample matrix effects. Thus, a serum or urine sample may contain a variety of fluorescing substances, such as proteins and bilirubin, that are a potential source of unwanted background fluorescence. Light scattering by proteins and lipoproteins can also contribute to this negative effect. Selectivity is also a major problem in the determination of drugs in their dosage forms, which are often multicomponent preparations. The lack of selectivity in fluorescence analysis of clinical and pharmaceutical samples usually calls for a prior sample cleanup step using a separation technique,

such as liquid-liquid or solid-phase extraction. Dialysis is used in some cases to deproteinize serum. As regards sensitivity, the facts that many drugs are administered in very low doses and that biological fluids usually contain a large number of substances at low concentrations call for highly sensitive procedures.

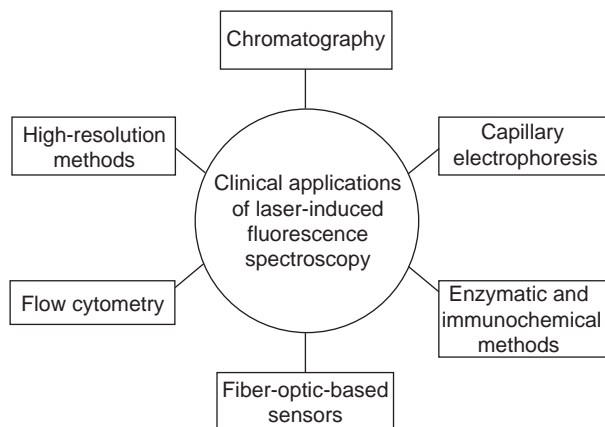
The selectivity and sensitivity of fluorimetric analysis can be enhanced by use of a number of alternatives to conventional fluorimetry. The usefulness of long-wavelength fluorophores in avoiding the interference of the sample matrix has been shown in numerous applications. Advances in fluorescence instrumentation allow interferences to be minimized or even eliminated altogether, thereby providing higher resolutions. Stray and scattered light can now be avoided. Also, sample preparation and handling techniques have been improved in parallel, and a number of enhanced spectral calibration and correction procedures have been developed. In addition, computerized fluorimetric instrumentation endows assays with higher precision and obviates the need for interpretation of vast amounts of data.

The determination of extremely low concentrations call for laser-induced fluorescence (LIF) spectroscopy, which is hardly affected by stray light thanks to the high monochromaticity of the beams used. Although laser instruments are still far from commonplace in the clinical laboratory, they have been used in a number of interesting applications (Figure 1) such as chromatographic and electrophoretic methods since the good collimation of the laser beam allows its full intensity to be focused onto a very small area. Also, the detection limits of laser-based enzymatic and immunochemical methods are typically lower than those afforded by conventional radiation sources. Other interesting applications of LIF in biomedical analysis involve the use of

fiber-optic sensors, which enable *in vivo* measurements, and the application of fluorescence line-narrowing spectrometry, a low-temperature solid-state spectroscopy, to a variety of biomolecular studies including cellular macromolecular damage and chemical carcinogenesis. Lasers are also very useful in fluorescence microscopy and flow cytometry.

Fluorescence spectroscopy has been used in a number of classical diagnostic tests, the capabilities of many of which have been enhanced by exploiting the advantages offered by the development of fluorimetric instrumentation. Thus, screening of newborns with normal tyrosine levels for phenylketonuria and hyperphenylalaninemia is routinely carried out using fluorimetric assays. Fluorimetric procedures for noradrenaline, adrenaline, dopa, dopamine, and their metabolites allow the specific type of catecholamine and its concentration to be determined, which in turn permits characterization of pheochromocytomas and neuroblastomas. Many vitamins and their metabolites are best quantified for diagnosis of avitaminosis by fluorimetric assay. The mineral metabolism of calcium and magnesium has been studied fluorimetrically in many laboratories. Fluorimetric assays for porphyrins are very simple and sensitive. Adrenocortical function is usually assessed by using a fluorimetric procedure for corticosteroids. Also, urinary estrogens are typically assayed fluorimetrically in clinical laboratories.

Although fluorescence spectroscopy is usually applied in solution, fluorescence measurements on solid surfaces are becoming a major choice for screening of drugs of abuse, pharmaceuticals, and clinical substances. In addition to thin-layer chromatographic plates, other solid-surface attachments such as electrophoresis steps, silicone rubber pads, and sodium acetate layers can be used for this purpose. Also, a number of fluorimetric methods use immobilized enzymes or immunoreagents. In this particular analytical field, measurements can be made on specially designed fluorimetric pads containing all the reagents required for quantitative assay, with substantial time savings.



**Figure 1** Scope of application of LIF spectroscopy in clinical analysis.

## Chromatographic and Electrophoretic Methods

Fluorescence spectroscopy has reached a prominent position among the detection systems used in conjunction with separation techniques, mainly liquid chromatography (LC) and capillary electrophoresis (CE). The literature abounds with references to clinical and pharmaceutical applications of LC with fluorimetric detection. Thus, as can be seen in



Table 1, a host of methods for the determination of therapeutic and abuse drugs using LC and native fluorescence measurements have been described. There are also numerous methods for drug determination in clinical samples involving pre- or postcolumn derivatization reactions (Table 2). Fluorescence labeling through covalent binding is the type of reaction most commonly used in this context. Postcolumn photochemical reactions induced by ultraviolet irradiation are other useful means for converting some substances such as phenothiazines into strongly fluorescent products. Another possibility

is to exploit the dependence of the emitted radiation intensity on the particular environment of a fluorescent compound, which is added to the eluent. Thus, lipids have been determined by mixing a continuous stream of aqueous solution containing a fluorescent dye with the column effluent, which results in a certain level of baseline fluorescence that is increased by the presence of lipids, probably because of the enhancing effect of the micelles formed by the lipids.

LC with fluorimetric detection is also suitable for therapeutic drug monitoring (TDM) purposes. For

**Table 1** Drug families determined in clinical samples using LC and intrinsic fluorescence measurements

<i>Family</i>	<i>Compound</i>	<i>Sample</i>	<i>Family</i>	<i>Compound</i>	<i>Sample</i>
Amino acids	Tryptophan	Brain tissue, plasma, urine	Antineoplastics	Daunorubicin	Plasma, urine
	Tyrosine	Plasma		Dexorubicin Melphalan	Plasma Plasma
Analgesics	Ibuprofen	Plasma	Antituberculous agents	Aminosalicilic acid	Plasma, urine
	Naproxen	Serum	Blocking agents ( $\beta$ -adrenoceptors)	Acebutorol	Plasma, serum, urine
	Salicylamide	Plasma			
Antibacterial agents	Ciprofloxacin	Plasma, serum, urine, and others		Alprenolol	Plasma
	Norfloxacin	Plasma, urine		Atenolol	Plasma
	Ofloxacin	Plasma, urine		Bunobronol	Plasma, urine
	Pefloxacin	Serum		Celipronol	Plasma, urine
	Difloxacin	Plasma, urine		Metoprolol	Plasma, urine
	Sulphapyridine	Plasma		Penbutorol	Plasma, urine
	Sulphasalazine	Plasma		Pindolol	Plasma, urine
				Propranolol	Plasma, urine
				Sotalol	Plasma, urine
Anticonvulsants	Carbamazepine	Plasma			
Antidepressants	Desipramine	Plasma	Dermatological agents	Salicylic acid	Plasma, urine, saliva
	Imipramine	Plasma	Diuretics	Amiloride	Plasma
	Trimipramine	Plasma		Bumetanide	Plasma, urine
				Furosemide	Plasma, urine
				Metolazone	Plasma, urine
Antidiabetics	Glyburide	Serum			
Anthelmintics	Thiabendazole	Serum	Hallucinogens	Lynergic acid diethylamide	Urine
			Narcotic analgesics	Codeine	Plasma
Antihypertensives	Indoramin	Plasma	Sympathomimetics Tranquillizers	Prenalterol	Plasma, urine
	Labetalol	Plasma		Pipothiazine	Plasma, urine
	Prazosin	Plasma		Sulpiride	Serum, urine
Antimalarials	Chloroquine	Plasma, urine	Vitamins	Pyridoxine	Blood
	Pyrimethamine	Plasma		Riboflavin	Blood
				Tocopherols	Plasma



**Table 2** Examples of derivatization reactions used in LC for the fluorimetric determination of drugs in clinical samples

<i>Reaction/reagent<sup>a</sup></i>	<i>Analyte</i>	<i>Sample</i>
<i>Pre/o-phthaldialdehyde</i>	Gentamicin	Plasma, urine
	Histamine	Plasma, urine, tissues
	Amphetamine	Plasma, urine, tissues
	Phenylpropanolamine	Urine, plasma
<i>Pre/dansyl chloride</i>	Maprotiline	Plasma, urine
	Perbexiline	Plasma
	Trimetazinide	Plasma
	Barbiturates	Serum, blood
	Cannabinoids	Urine
	Tocainide	Plasma
	Hydrocortisone	Plasma
<i>Pre/dansyl hydrazine</i>	Penicillamine	Serum
<i>Pre/dansyl azidine</i>	Gentamicin	Plasma, urine
<i>Pre/fluorescamine</i>	Tocainide	Plasma
<i>Pre/7-chloro-4-nitrobenz-2-oxa-1,3-diazole</i>	Amphetamine	Blood, urine
<i>Pre/N-(1-pyrene) maleimide</i>	Captopril	Plasma
<i>Pre/naphthacylbromide</i>	Valproic acid	Plasma
<i>Pre/4-bromomethyl-7-methoxycoumarin</i>	Barbiturates	Serum, blood
<i>Pre/deacylation</i>	Indomethacin	Plasma, urine
<i>Pre/oxidation</i>	Dihydromorphine	Urine
	Emetine	Plasma
	Morphine	Urine
	Reserpine	Plasma
	Methotrexate	Plasma
	Morphine	Blood, urine
	Nalorphine	Blood, urine
	Isoprenaline	Plasma, urine
	Thiamine	Blood
	Benzodiazepines	Serum
<i>Post/fluorescamine</i>	Cephadrine	Serum
	Cephtracrine	Urine
	Amphetamine	Urine
	Indomethacin	Plasma, urine
<i>Post/hydrolysis</i>	Clobazam	Serum
<i>Post/irradiation</i>	Demoxepam	Serum, urine
	Penbendazole	Serum
	Phenothiazines	Serum, urine
	Clomiphene	Plasma
	Stilboestrol	Urine
	Cannabinoids	Urine

<sup>a</sup> Precolumn (*Pre*) or postcolumn (*Post*) derivatization.**Table 3** Features of methods for therapeutic drug monitoring of selective serotonin reuptake inhibitors

<i>Analytes</i>	<i>Sample</i>	<i>Extraction<sup>a</sup></i>	<i>Derivatization</i>	<i>LOQ<sup>b</sup> (ng ml<sup>-1</sup>)</i>
Citalopram	Plasma	LL	No	5
Desmethylcitalopram				
Didesmethylcitalopram				
Citalopram propionic acid				
Fluoxetine	Serum	SPE	No	20
Norfluoxetine	Plasma	LL	Yes	3
Fluvoxamine	Plasma	LL	Yes	3
Paroxetine	Serum	SPE	No	5

<sup>a</sup> Liquid-liquid (LL) or solid-phase (SPE) extraction.<sup>b</sup> Limit of quantitation.

instance, it has been applied to the determination of selective serotonin reuptake inhibitors, which are used as antidepressants. Table 3 shows some features of these methods, which involve a sample pretreatment

by liquid-liquid or solid-phase extraction. An interesting application of LC with fluorimetric detection to drug analysis is the separation of enantiomers. For instance, cardiovascular agents

represent one of the largest prescribed class of pharmaceuticals since cardiovascular diseases are one of the main causes of mortality in the world. Many of these compounds are still administered as a racemic mixture, although (R)- and (S)-individual enantiomers usually possess different pharmacological and toxicological activities. Table 4 shows some examples of the usefulness of fluorimetry in this area.

Probably the main analytical application of CE is, for the present, the determination of drug substances and, indeed, the penetration of this technique into the pharmaceutical industry, after a slow start, is now a reality. Also, CE is becoming a tool complementary to established LC methods for drug determination in biological samples. The separation efficiency of CE can be greater than that of LC, and CE can be less expensive because organic solvent consumption is lower and the capillaries are cheaper. However, CE cannot yet be regarded as a routine technique in clinical and drug analysis. There are still several problems such as the lack of sensitivity and the need for sample pretreatment to remove interfering species and particulate matter that can easily clog the CE system. Sometimes CE is not sensitive enough to determine low concentrations because of the small sample injection volume, which is typically only a few nanoliters and the short optical path length for on-capillary detection.

Fluorescence detection in CE is generally performed using laser sources for excitation instead of conventional lamps, with the aim of concentrating enough radiation energy inside the capillary. However, although the use of LIF provides extremely high sensitivity, it is only applicable for some analytes as lasers are commercially available only with a limited number of wavelengths. For instance, some drugs like anthracyclines fluoresce when excited at 488 nm, which is the typical emission wavelength of the argon ion laser. Thus, detection limits of 0.5 and 2 ng ml<sup>-1</sup> have been obtained for idarubicin and doxorubicin,

respectively, using a sample volume of 0.1 nl plasma with a commercial CE instrument equipped with an argon laser. The use of a He–Cd laser emitting at 325 nm has allowed also the determination of several drugs, such as ciprofloxacin, methotrexate, and zolpiden, in biological fluids using CE–LIF. An alternative to direct fluorescence detection is the use of a derivatization reaction, although it can involve a loss of selectivity. In spite of these limitations, the use of LIF detection is one of the preferred options for improving sensitivity in the analysis of biological samples using CE. The interest in the near-infrared region in CE using diode lasers and long-wavelength fluorophores has grown in the last years as it is a useful option for avoiding background interferences. This approach has been applied to the determination of amino acids, peptides and proteins, and DNA sequencing and fragment analysis.

The usefulness of LIF in DNA analysis has been demonstrated using fluorescent dyes as labels in CE and denaturing LC (DHPLC) techniques. It has been suggested that CE is preferable for DNA sizing and sequencing analyses and DHPLC is effective for screening of genetic mutations and polymorphisms.

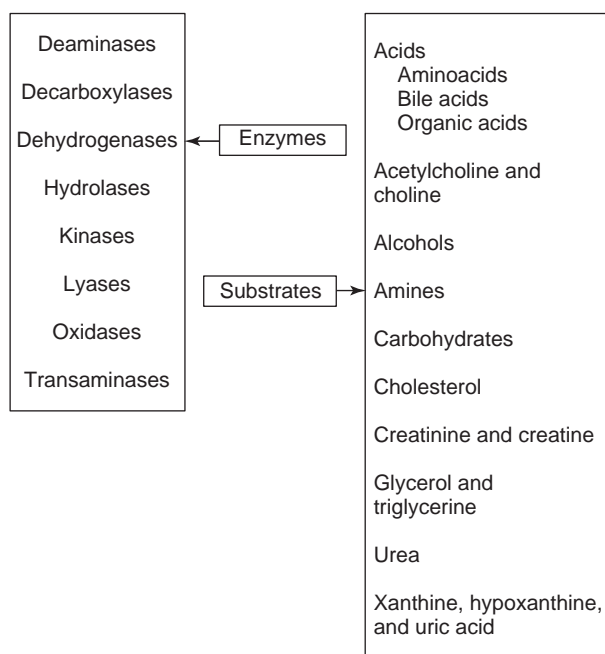
## Enzymatic Assays

The high sensitivity of fluorescence spectroscopy and the selectivity of enzymatic assays are responsible for the increasing use of fluorimetric methods in enzymology. Enzyme determinations usually involve the use of kinetic methodology for measuring the rate of formation of the fluorescent product, while both equilibrium and kinetic methods are used to determine the substrates. Fluorimetric measurements on enzyme-catalyzed reactions have been used for a long time to determine a variety of enzymes and substrates (Figure 2).

NADH and NADPH, the respective reduced forms of NAD<sup>+</sup> and NADP<sup>+</sup>, are highly fluorescent; hence, all NAD<sup>+</sup>- and NADP<sup>+</sup>-dependent reactions involved in enzymatic assays can be monitored fluorimetrically with a sensitivity higher than those of absorptiometric techniques by two or three orders of magnitude. Reactants must be used at much lower concentrations, and detection limits of 10<sup>-9</sup> mol l<sup>-1</sup> or even lower can be achieved. Numerous methods for enzyme determination involve the use of substrates that are transformed into highly fluorescent products. For instance, *p*-hydroxyphenylacetic acid has been used widely as a substrate for the determination of several oxidases such as glucose oxidase and xanthine oxidase.

**Table 4** Chromatographic enantioseparation of cardiovascular drugs using fluorescence detection

Analyte	Sample
Sotalol	Plasma, urine
Propranolol	Urine
4-Hydroxypropranolol	Urine
Albuterol	Serum
Verapamil and norverapamil	Plasma
Gallopamil	Plasma
Mexiletine	Serum
Ibutilide	Plasma
Artilede	Plasma



**Figure 2** Fluorimetric determinations based on enzyme-catalyzed reactions.

Enzymatic substrates yielding fluorescent products that emit at long wavelengths are a useful alternative for avoiding background signals. Thus, naphtho-fluorescein diphosphate has been proposed as a substrate for alkaline phosphatase, which hydrolyses the substrate yielding fluorescent naphthofluorescein. This system has been used for the determination of theophylline at therapeutic levels based on the inhibitory effect of this drug on the enzyme activity. Fluorogenic substrates for proteases containing Cresyl Violet have been synthesized to visualize protease activity in single cells. Also, a tetra-substituted amino aluminum phthalocyanine has been described as a red-region substrate for peroxidase. The system has been applied to the determination of glucose in serum by coupling a glucose oxidase-catalyzed reaction.

## Fluoroimmunoassays

Immunochemical methods are routinely used in the clinical laboratory on account of their special features, particularly their high selectivity and the commercial availability of automatic instrumentation for their implementation. Since the earliest immunoassay involving radioisotopes was reported, a large variety of alternative immunoassays have been developed to circumvent the shortcomings of radioimmunoassay, many of them using fluorescence detection. Some fluoroimmunoassays (FIAs) developed with this purpose, such as fluorescence-quenching

**Table 5** Scope of application of FPIA in clinical analysis

<i>Therapeutic drugs</i>	<i>Illicit drugs</i>	<i>Hormones</i>
Antibiotics	Amphetamines	Thyroid agents
Antiarrhythmics	Barbiturates	Steroids
Anticonvulsants	Cannabinoids	
Antiasthmatics	Benzodiazepines	
Cardiac glycosides	Opiates	
Antineoplastics		
Tricyclic antidepressants		

and fluorescence-enhancing immunoassays, are scarcely sensitive enough or applicable, which hinders their use by clinical laboratories. However, there are some FIAs widely used at present using automated instrumentation.

Fluorescence polarization spectroscopy is applied in homogeneous immunoassay for the determination of small molecules such as hormones and therapeutic and abuse drugs (Table 5). It is routinely used in TDM to quantify drug levels in plasma and serum specimens. The most serious problem faced in applying fluorescence polarization immunoassay (FPIA) to serum samples is its relatively poor sensitivity, which is limited by three factors: (1) the background fluorescence of serum samples, which arises partly from scattered light and partly from the intrinsic fluorescence due to the presence of fluorescent compounds such as NADH and bilirubin; (2) nonspecific binding of the tracer to serum proteins such as albumin, which is probably the chief agent responsible for this interaction because of its versatile ligand properties and affinity for many anionic dyes (e.g., fluorescein, the most commonly used label in FPIA); and (3) the presence of polarizers in the excitation and emission beams, which reduces the light intensity and potential sensitivity by a factor of up to ~10. In spite of these shortcomings, the advent of microprocessor technology, improvements in optics and detectors, and advances in tracer design and immunological techniques have turned FPIA into a practical technique for use in the clinical laboratory.

Fluorescence energy transfer immunoassay is another homogeneous format based on a nonradiative fluorescence resonance energy process. One of the immunoreagents is labeled with a donor fluorophore, while another immunoreagent is labeled with an acceptor chromophore, which may or may not be fluorescent. There should be considerable overlap between the emission spectrum of the donor and the absorption spectrum of the acceptor. The formation of the antigen-antibody complex gives rise to the quenching of fluorescence of the donor, while the fluorescence of the acceptor increases if the acceptor is fluorescent. Besides fluorescence intensity changes,

the energy transfer process also induces fluorescence lifetime changes, which can be measured by using pulsed or modulated excitation sources. Because the fluorescence properties of the unbound species essentially remain unchanged, homogeneous formats are possible.

Among homogeneous fluoroimmunoassays based on conventional fluorimetry, substrate-labeled fluoroimmunoassay is another alternative for both hapten and protein determination using a fluorescent label. As in other homogeneous competitive immunoassays, the sensitivity is limited by the serum background fluorescence and the limited amount of tracer that can be used.

The combined use of time-resolved fluorescence spectroscopy and lanthanide (particularly europium(III)) chelates as labels has given rise to time-resolved fluoroimmunoassay (TRFIA). It provides a high sensitivity as a result of several assets of europium chelates, such as their large Stokes shifts, which make separation of the excitation and emission wavelengths very easy; their narrow emission bandwidths, which allow the use of narrow bandpass emission filters with no loss of energy; and the long life of the fluorescence they emit, such that the short-lived (1–20 ns) background fluorescence of serum, solvents, cuvettes, and reagents obtained upon flash excitation of the sample can be allowed to decay so that the only possible fluorescent signal will be due to the europium chelate ( $\sim 500 \mu\text{s}$ ). As a result, interfering background fluorescence, the greatest hurdle for other FIAs, is reduced by two or three orders of magnitude. TRFIA requires the use of a xenon flash lamp or a laser excitation source for obtaining time-resolved measurements. The process is carried out in two steps: the reagent (antigen or antibody) labeled with a nonfluorescent europium chelate is subjected to the immunoreaction, and then the europium is released and bound to another chelating agent, which produces an intense fluorescence arising from an energy transfer process. This requires the use of coated tube technology, in which usually a europium-labeled antibody reacts with the immobilized antibody–antigen complex on the solid phase (i.e., the sandwich technique). TRFIA is the basis for a host of applications, including in gynecology, thyroid disease, cancer diagnosis, viral infections, cytotoxicity, and enzyme activity measurements. Many of these applications are now used in routine clinical analyses implemented with commercially available kits and automated instrumentation.

Another application of lanthanide ions in immunoassay, in this case terbium(III), is enzyme-amplified lanthanide luminescence. This approach involves the use of the substrate of an enzyme that does not form

a fluorescent chelate with terbium(III), but is converted to a product that does form such a chelate. Alkaline phosphatase (ALP) has been used for labeling, and several compounds, such as salicyl phosphate, fluorosalicyl phosphate, and diflunisal phosphate, as substrates. These compounds cannot coordinate efficiently with the terbium(III)–ethylene-diaminetetraacetic acid complex, but the fluorescent ternary complex is formed in the presence of ALP, which splits the phosphate ester and releases the terbium ligand. This approach has been applied in immunoassay to the determination of several species, such as  $\alpha$ -fetoprotein, thyrotropin, and interleukin 6, leading to very low detection limits.

## Fluorosensors

Optical sensors based on fluorescence spectroscopy and immobilized enzymes or immunoreagents (biosensors) are widely used at present because of the high selectivity they provide. The earliest fluorosensor for glucose was developed in 1980 and was used to measure the amount of oxygen consumed during its enzymatic oxidation by glucose oxidase in the presence of catalase, which is covalently immobilized. Since this biosensor was developed, this methodology has been used for the determination of a large number of species of clinical interest. Fluorosensors have also been widely applied to pharmaceutical analysis. Thus, penicillin can be assayed by using a pH fiber sensor onto which penicillinase is immobilized. The enzyme catalyses the hydrolysis of penicillin to penicilloic acid, and the resulting increase in the hydrogen ion concentration is monitored. Oxygen sensors based on the immobilization of ruthenium(II) complexes, which show long-wavelength fluorescence, have been modified into biosensors for the determination of clinical parameters such as glucose or free cholesterol by immobilizing the corresponding enzyme also.

The development of antibody-based biosensors presents more difficulties than enzyme-based biosensors as the antigen–antibody interactions are not readily reversible because of the high values of the affinity constants. Another limitation is that the physicochemical changes resulting from the immunochemical reaction are often insufficient to provide detection limits comparable with those of conventional analysis. As a consequence, indirect systems have been developed that rely on the use of enzyme- or fluorescent-tagged reagents. Both competitive and sandwich formats are used. Evanescent wave-induced fluorescence is frequently chosen to avoid possible interferences from the bulk media. For

instance, an immunosensor based on the use of the dye Cy5, a long-wavelength fluorophore, has been described for the determination of cocaine metabolites in urine using a competitive format.

Fiber-optic sensors involving the use of LIF measurements are of great interest in biomedical analysis as they can be used for measuring analytes remotely, continuously, and directly within the fluid to be analyzed. This technique has been employed in cytological applications for diagnosing tumors and atherosclerosis. The ability to focus intense beams tightly to subcellular dimensions is significant in obtaining measurable signals from weakly fluorescent cells and essential when morphological information is required. There is a growing interest in the miniaturization of the sensors, which allows the reduction of the sample volume and the response time. The analytical properties of fluorescent nanosensors allow accurate analysis of biological systems at the single cell level. To prevent cytotoxicity upon insertion of the sensor into the cell, the sensing reagent must be isolated from the cellular environment by a biocompatible matrix that is selective to the intracellular analyte. The application of these sensors has been greatly assisted by recent advances in digital fluorescence imaging instrumentation.

## Proteomics Studies

Proteomics includes a series of key technologies that are being used to identify proteins and map their interactions in a cellular context. With the sequencing of the human genome, the scope of proteomics has shifted from protein identification and characterization to include protein structure, protein function, and protein-protein interactions. Fluorescence-based and mass spectrometry (MS)-based methodologies offer new capabilities in the field of proteomics through the performance of multiplex quantitative analysis. Two-dimensional gel electrophoresis (2D-GE) is a separation technique used widely in proteomics. Visualization of proteins in gels is accomplished through staining techniques. Several fluorescent stains have proved to be suitable for incorporation into integrated proteomics platforms for global analysis of protein regulation. The linear dynamic range of detection using fluorescent stains is usually superior to colorimetric or reverse stains. Fluorescence measurements in 2D-GE are obtained using charge-coupled cameras or laser-based detectors as image-acquisition devices. For proteomics work, the protein stains must be compatible with MS as the coupling of 2D-GE to MS provides both precise mass determination and protein identification.

## Other Applications

There are other fluorimetric techniques of interest to clinical and pharmaceutical analyses. Thus, the performance of total fluorescence spectrometry in multicomponent and complex analyses has been assessed. The overall fluorescence of human serum and urine can be used as a diagnostic tool in clinical chemistry, particularly in view of its selectivity towards even quite small changes in the location of the peaks and their relative intensities. This is the foundation of pattern recognition methods for detection of pathological deviations from normal status. This technique is also suitable for rapid screening of a variety of drugs since the samples require virtually no treatment.

Derivative and synchronous fluorescence spectrometry, both individually and jointly, have been used in clinical and drug analyses as a means of enhancing selectivity. Thus, derivative synchronous fluorescence spectrometry allows the direct determination of catecholamines and porphyrins in urine and pyridoxal and its derivatives and cephalosporins in serum. In addition to its use in immunoassay, fluorescence polarization spectroscopy has been employed to study macromolecules and to carry out structural investigations of the binding of small molecules to proteins as well as to assess fetal lung maturity. It has also been applied for single nucleotide polymorphisms genotyping and in drug discovery research using targets such as kinases, phosphatases, proteases, G-protein coupled receptors, and nuclear receptors. Fluorescence microscopy for characterizing biological surfaces is increasingly being cited in the analytical literature. For instance, it is routinely used to study the location and movement of intracellular species.

Fluorescence probes are widely used in flow-cytometric applications in the clinical laboratory. Some flow cytometers include argon lasers and enable right-angle fluorescence, right-angle light scatter, and forward light scatter measurements. In this way, the cellular DNA content, one of the most commonly determined parameters, can be measured fluorimetrically using DNA-specific fluorescent dyes. Another interesting application of this approach is the determination of intracellular ionized calcium concentrations. Several strongly fluorescing  $\text{Ca}^{2+}$ -sensitive dyes can be used for single-cell measurements. In multilaser flow cytometers, several fluorescent dyes are used in order to correlate the intracellular  $\text{Ca}^{2+}$  concentration with those of cell surface antigens or some other parameters. A number of immunofluorescence studies based on flow cytometry have been carried out by directly conjugating a fluorophore



(usually fluorescein isothiocyanate) to an antibody (direct immunofluorescence) or a second antibody that reacts with the first (indirect immunofluorescence). This allows one to ascertain the presence of cell surface antigens.

Much work has been carried out in the development of new fluorescent probes suitable for the detection and quantification of biologically significant molecules both *in vivo* and *in vitro*. For instance, aptamers are oligonucleotides that in a particular conformation exhibit molecular recognition. They show affinity for both small molecules and macromolecules. Aptamer beacons are aptamers that incorporate a fluorescence-quenching pair, which is used to indicate a consequent change in conformation upon binding. These compounds are selected for specific binding to a variety of molecular targets, ranging from small organics to proteins. Their usefulness in molecular recognition is being studied using affinity chromatography, capillary electrophoresis, and biosensors, although there are still relatively few examples of the use of aptamer beacons in analysis, which can be explained by their relatively recent development.

See also: **Clinical Analysis:** Overview. **Electrophoresis:** Clinical Applications. **Immunoassays, Applications:**

**Clinical. Liquid Chromatography:** Clinical Applications; Pharmaceutical Applications. **Microscopy:** Overview. **Pharmaceutical Analysis:** Overview. **Sensors:** Overview.

## Further Reading

- Baeyens WRG, Keukelerie DD, and Korkidis K (1991) *Luminescence Techniques in Chemical and Biochemical Analysis*. New York: Dekker.
- Guilbault GG (1990) *Practical Fluorescence*. New York: Dekker.
- Haugland RP (1996) *Handbook of Fluorescence Probes and Research Chemicals*. The Netherlands: Molecular Probes<sup>TM</sup> (also see [www.probes.com](http://www.probes.com)).
- Hemmilä IA (1991) *Applications of Fluorescence in Immunoassays*. New York: Plenum.
- Lakowicz JR (1992) *Topics in Fluorescence Spectroscopy. Biochemical Applications*. New York: Plenum.
- Lakowicz JR (1994) *Topics in Fluorescence Spectroscopy. Probe Design and Chemical Sensing*. New York: Plenum.
- Lunte SM and Radzik DM (1996) *Pharmaceutical and Biomedical Applications of Capillary Electrophoresis*. Oxford: Elsevier.
- Whorter SM and Soper SA (2000) Near-infrared laser-induced fluorescence detection in capillary electrophoresis. *Electrophoresis* 21: 1267–1280.

## Environmental Applications

**S R Bravo**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Environmental analyses often involve the determination of low concentrations (viz., a few nanograms or picograms per milliliter) that require the sensitivity of fluorescence analysis. Most environmental applications of fluorimetric methods involving air, water, or soil pollution are concerned with polycyclic aromatic compounds, which feature high fluorescence quantum yields. However, in recent years, the scope of fluorimetric techniques has been successfully extended to other major nonfluorescent pollutants such as isocyanates after conversion into fluorescent species. But fluorescence detection and determination of pollutants is still confronted with problems arising from a lack of selectivity. Such problems are currently

being addressed using a number of approaches, mainly involving the use of separation techniques (liquid chromatography, LC, or capillary electrophoresis, CE), multiparametric information (e.g., spectral change, intensity, lifetime, polarization, etc., of fluorescence), low-temperature fluorimetry, and immunoanalytical techniques. These approaches are not incompatible, so two or three are often used in combination. Their main applications to the determination of pollutants are briefly discussed in this article.

## LC-Fluorescence Detection

### Polycyclic Aromatic Hydrocarbons

Polycyclic aromatic hydrocarbons (PAHs) occur widely in the environment and, even though they come from both natural (volcanic eruptions and forest and prairies fires) and anthropogenic (combustion of fossil fuels, waste incineration, coke and asphalt



(usually fluorescein isothiocyanate) to an antibody (direct immunofluorescence) or a second antibody that reacts with the first (indirect immunofluorescence). This allows one to ascertain the presence of cell surface antigens.

Much work has been carried out in the development of new fluorescent probes suitable for the detection and quantification of biologically significant molecules both *in vivo* and *in vitro*. For instance, aptamers are oligonucleotides that in a particular conformation exhibit molecular recognition. They show affinity for both small molecules and macromolecules. Aptamer beacons are aptamers that incorporate a fluorescence-quenching pair, which is used to indicate a consequent change in conformation upon binding. These compounds are selected for specific binding to a variety of molecular targets, ranging from small organics to proteins. Their usefulness in molecular recognition is being studied using affinity chromatography, capillary electrophoresis, and biosensors, although there are still relatively few examples of the use of aptamer beacons in analysis, which can be explained by their relatively recent development.

See also: **Clinical Analysis:** Overview. **Electrophoresis:** Clinical Applications. **Immunoassays, Applications:**

**Clinical. Liquid Chromatography:** Clinical Applications; Pharmaceutical Applications. **Microscopy:** Overview. **Pharmaceutical Analysis:** Overview. **Sensors:** Overview.

## Further Reading

- Baeyens WRG, Keukelerie DD, and Korkidis K (1991) *Luminescence Techniques in Chemical and Biochemical Analysis*. New York: Dekker.
- Guilbault GG (1990) *Practical Fluorescence*. New York: Dekker.
- Haugland RP (1996) *Handbook of Fluorescence Probes and Research Chemicals*. The Netherlands: Molecular Probes<sup>TM</sup> (also see [www.probes.com](http://www.probes.com)).
- Hemmilä IA (1991) *Applications of Fluorescence in Immunoassays*. New York: Plenum.
- Lakowicz JR (1992) *Topics in Fluorescence Spectroscopy. Biochemical Applications*. New York: Plenum.
- Lakowicz JR (1994) *Topics in Fluorescence Spectroscopy. Probe Design and Chemical Sensing*. New York: Plenum.
- Lunte SM and Radzik DM (1996) *Pharmaceutical and Biomedical Applications of Capillary Electrophoresis*. Oxford: Elsevier.
- Whorter SM and Soper SA (2000) Near-infrared laser-induced fluorescence detection in capillary electrophoresis. *Electrophoresis* 21: 1267–1280.

## Environmental Applications

**S R Bravo**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Environmental analyses often involve the determination of low concentrations (viz., a few nanograms or picograms per milliliter) that require the sensitivity of fluorescence analysis. Most environmental applications of fluorimetric methods involving air, water, or soil pollution are concerned with polycyclic aromatic compounds, which feature high fluorescence quantum yields. However, in recent years, the scope of fluorimetric techniques has been successfully extended to other major nonfluorescent pollutants such as isocyanates after conversion into fluorescent species. But fluorescence detection and determination of pollutants is still confronted with problems arising from a lack of selectivity. Such problems are currently

being addressed using a number of approaches, mainly involving the use of separation techniques (liquid chromatography, LC, or capillary electrophoresis, CE), multiparametric information (e.g., spectral change, intensity, lifetime, polarization, etc., of fluorescence), low-temperature fluorimetry, and immunoanalytical techniques. These approaches are not incompatible, so two or three are often used in combination. Their main applications to the determination of pollutants are briefly discussed in this article.

### LC-Fluorescence Detection

#### Polycyclic Aromatic Hydrocarbons

Polycyclic aromatic hydrocarbons (PAHs) occur widely in the environment and, even though they come from both natural (volcanic eruptions and forest and prairies fires) and anthropogenic (combustion of fossil fuels, waste incineration, coke and asphalt

production oil refining, etc.) sources, human pollution contributes the larger proportion by far. Due to their environmental concern, PAHs are included in the US EPA and in the European Union priority lists of pollutants. US EPA has identified 16 unsubstituted PAHs as priority pollutants, some of which possess carcinogenic and mutagenic activities. The European list contains six target PAHs.

A number of analytical techniques have been used for the determination of PAHs in complex environmental samples. Gas chromatography with either flame ionization or mass spectrometry detection and LC with ultraviolet (UV) or fluorimetric detection are the techniques most widely used. Fluorescence spectroscopic detection provides both a sensitive and, often more important, a selective means for monitoring PAHs in liquid chromatographic effluents. PAHs make a virtually ideal class of fluorophores. In fact, PAHs possess large molar absorptivities (usually greater than  $10^4 \text{ mol}^{-1} \text{ cm}^{-1}$ ) and exhibit maximum absorption in the near-UV or visible region, in addition to high fluorescence quantum yields (normally above 0.2 and often approaching 1.0) and long fluorescence lifetimes (a few tens to hundreds of nanoseconds). The extremely high sensitivity in fluorescence emission has lowered the minimum detectable concentration of most PAHs to near sub-picogram levels (detection limits as calculated from the signal-to-noise ratios for the individual peaks in the chromatograms). An additional feature is that the fluorescence intensity of PAHs is linearly dependent on their concentration over ranges of as many as six orders of magnitude.

Reversed-phase LC on chemically bonded  $\text{C}_{18}$  (octadecylsilane) stationary phases is by far the most popular liquid chromatographic mode for separation of PAH compounds. Resolution is greatly influenced by the type of synthesis used to prepare the bonded phase. Good resolution can be achieved for the 16 US EPA PAHs on polymeric  $\text{C}_{18}$  phases. However, some isomers are unresolved (chrysene and benzo[*a*]anthracene) or only partially resolved (benzo[*ghi*]perylene and indeno[1,2,3-*cd*]pyrene, benzo[*k*]fluoranthene, and benzo[*b*]fluoranthene, or fluoranthene and acenaphthene) when monomeric  $\text{C}_{18}$  phases are used.

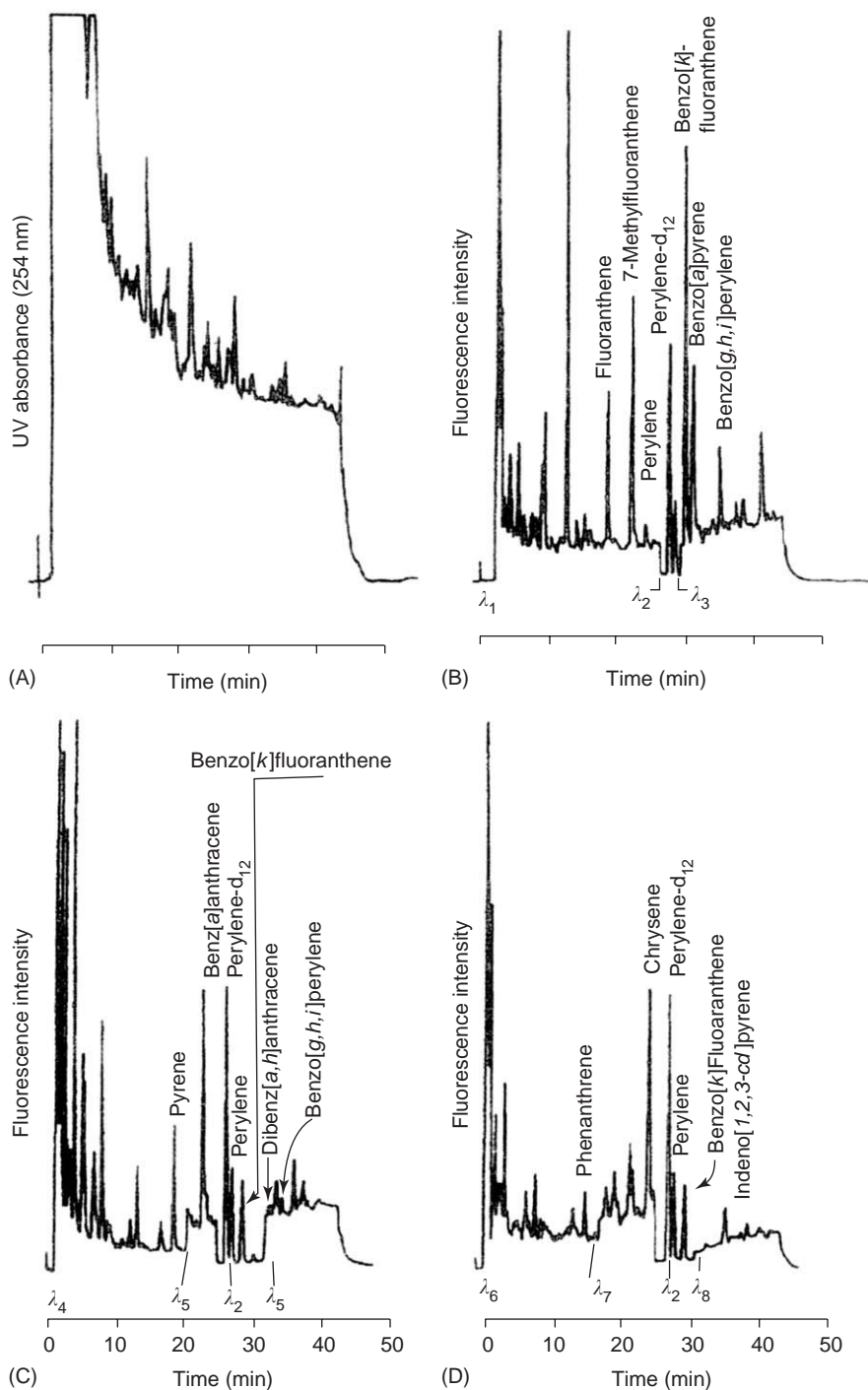
High selectivity is generally obtained for the determination of PAHs in environmental samples using LC-fluorescence detection. Since the retention times obtained using a  $\text{C}_{18}$  LC column and the optimal excitation and emission wavelengths for individual PAHs are known, time programming of the fluorimeter wavelength settings can be used to minimize interference from coeluting compounds. Obviously, the excitation and emission wavelengths to be used

with a given PAH need not be their optimal wavelengths, but instead should be chosen so as to obtain the most selective fluorescence response possible for that PAH in the presence of anticipated coeluting compounds. In this way, the selectivity of fluorescence spectrometry can be used to simplify the separation process. **Figure 1** illustrates results typically obtained using this operational procedure.

Retention in reversed-phase LC increases with the number of aromatic carbons and the extent of alkylations. Therefore, the identification of PAHs based on chromatographic retention is confronted with a serious problem: coelution of alkyl-substituted homologs of smaller PAHs. Thus, some alkylphenanthrenes and alkylanthracenes elute in the region where fluoranthene and pyrene elute; also, alkylpyrenes and alkylfluoranthenes elute in the same region as benzo[*a*]anthracene, chrysene, and triphenylene. This shortcoming has been circumvented in the past by initially fractionating complex PAH mixtures into isomer families on an aminosilane column prior to isolation of individual isomers on a  $\text{C}_{18}$  column in the reversed-phase mode. The resulting chromatograms are free of any interference from alkylated homologs of smaller aromatic carbon numbers.

Quantitation of PAHs requires the use of internal standards with the same number of aromatic carbons as the analytes. Perdeuterated PAHs make excellent internal standards in this respect. Even though the deuterated and nondeuterated analogs of a given PAH are virtually indistinguishable by solution fluorimetric, the  $\text{C}_{18}$  column can isolate deuterated and nondeuterated PAHs with sufficiently high resolution for the fluorescence of each compound to be measured with no appreciable interference from the other. Probably the most suitable internal standards for each isomer family are the perdeuterated analogs of the isomers that elute first, since perdeuterated PAHs would then yield the first elution peak in the reversed-phase liquid chromatogram, thereby minimizing the risk of coelution with alkyl-substituted PAHs within a given isomer group. Thus, the ideal internal standards for analysis of isomer families would be perdeuterated phenanthrene (14 aromatic carbons, ACs), fluoranthene (16 ACs), triphenylene (18 ACs), benzo[*i*]fluoranthene (20 ACs), and benzo[*ghi*]perylene (22 ACs).

The combined use of LC and fluorescence detection for the determination of PAHs has some disadvantages, foremost among which is the long sequence of steps involved and the overall time required to complete analyses. By the same token, the hazards of loss of sample components (e.g.,



**Figure 1** Comparison of UV absorption (A) and fluorescence detection (B–D) in the reversed-phase liquid chromatographic separation of PAHs extracted from an urban particulate sample. The different wavelengths in the fluorescence chromatograms were optimal for excitation of the indicated PAH. The chromatograms were acquired by a computer-controlled fluorescence detector, enabling rapid changes in the excitation and emission wavelengths. (Reprinted with permission from May WE and Wise SA (1984) Liquid chromatographic determination of polycyclic aromatic hydrocarbons in air particulate extracts. *Analytical Chemistry* 56: 225–232; © American Chemical Society.)

through volatilization or photodecomposition) or contamination of samples with traces of fluorescent impurities are manifold unless considerable care is exercised.

Other fluorescence systems, in addition to that of variable-wavelength fluorescence detection, are used combined with LC for increasing sensitivity and selectivity. Thus, by using laser-induced

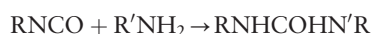
time-resolved fluorimetry, the detection capability for PAHs is improved by one or two orders of magnitude (e.g., the detection limit for benzo[*a*]pyrene is 180 fg). Fluorescence line-narrowing spectroscopy (FLNS) has been also interfaced to LC for online characterization of PAHs. In addition to the high spectral resolution afforded by FLNS, analyzing the separated components at 4.2 K minimizes photo-degradation from the excitation source and provides indefinite detection times for signal averaging. With femtomole detection limits, LC-FLNS can be used for analysis of complex mixtures.

### Isocyanates

Isocyanates, both in monomeric and in polymeric forms, are extensively used for manufacturing polyurethane foams, paints, adhesives, and a number of other products. The health hazards posed by the presence of isocyanates at workplaces, chiefly respiratory tract diseases, have become a source of major concern for occupational health officials. Atmospheric pollution by isocyanates has so far been assessed using two different approaches. One, widely employed to date, involves the determination of several specific isocyanates compounds for which threshold values have been established; the other has become a priority choice for air monitoring programs in recent years and entails determining the concentration of all isocyanate species present.

In determining specific isocyanates by LC one should take into account a few practical considerations. Thus, the extreme reactivity of the -NCO group precludes collection of isocyanates from air without immediate derivatization in order to avoid unwanted reactions of the analytes collected. For this reason, isocyanates are collected in an appropriate derivatization solution held in a bubbler or impinger. The derivatizing reagents used are dissolved in a high-boiling solvent such as toluene in order to avoid excessive volatilization during sampling, after which an aliquot of the absorbing solution is evaporated to dryness and the residue is collected in a solvent compatible with the mobile phase to be used in the liquid chromatographic analysis.

The fluorescent derivatizing reagents most commonly used in this context are primary or secondary amines that react with isocyanates to form ureas according to the reaction given below:



1-Naphthalenemethylamine (NMA), *N*-methyl-1-naphthalenemethylamine (MNMA), and 9-(*N*-methylamino-methyl)anthracene (MAMA) are the three fluorophores most frequently used for this purpose.

The sensitivity of these methods is more than adequate to meet the required detection limits for isocyanates. Thus, the detection limits for hexamethylene-1,6-diisocyanate (HDI) are quoted to be 0.5, 1, and  $0.2 \mu\text{g m}^{-3}$  when using NMA, MNMA, and MAMA, respectively. However, these methods have some major drawbacks. The NMA method results in the appearance of double peaks in the chromatogram for isocyanates for reasons that are still unclear. Also, some NMA derivatives are difficult to dissolve in ordinary liquid chromatographic solvents. On the other hand, MNMA is relatively unstable, so stock solutions of this reagent must be stored refrigerated and prepared freshly every month. The same is the case with MAMA, solutions of which must be freshly prepared from its hydrochloride.

LC methods using fluorescence detection, like most reported LC methods for the determination of isocyanates, have mainly been applied to monomer diisocyanates. In as much as they provide a different detector response for each isocyanate derivative, they require individual isocyanate standards for quantification. As a result, their application to polymeric forms is subject to some drawbacks shared by determinations of polymers using chromatographic procedures, viz., the lack of appropriate standard materials. In fact, such important data as purity, the precise identity of the polymeric components, and the nature of the additives used in the isocyanate polymers available for calibration are often lacking.

Evaluation of airborne polymeric isocyanates has become a major task as hygiene standards are increasingly frequently being defined in terms of total isocyanate concentrations (monomers and polymers) in air. Given the formidable analytical difficulties involved in identifying and quantifying each individual component present in an isocyanate mixture, hygiene standards tend to be based on total (-NCO) ( $\mu\text{g m}^{-3}$ ) data. As a result, isocyanate research in recent years has chiefly been aimed at finding reagents capable of providing a detector response that is proportional only to the amount of isocyanate function (-NCO) being derivatized, irrespective of the type of moiety attached to the -NCO group. In this way, quantification of various isocyanates can be made by use of a single standard for comparison and integration of the individual peak areas after LC separation of the constituent compounds. Tryptamine (3-(2-aminomethyl)indole) has become a standard reagent for this application. This reagent undergoes fluorescence and amperometric oxidation from the conjugate  $\pi$  system of the indolyl moiety, which is located two carbon atoms away from the derivatizing site of the amino group, so the electronic structure of the  $\pi$  system is not perturbed



whatever the types of isocyanates being derivatized. As a result, responses from both fluorimetric and amperometric detectors depend only on the amount of reacted tryptamine, which is stoichiometrically proportional to that of the isocyanate group. The net amount of isocyanate ( $w_{\text{NCO}}$ ) can be readily calculated from

$$w_{\text{NCO}} = \frac{w_{\text{Tryp}}}{160} \times 42$$

where  $w_{\text{Tryp}}$  is the amount of reacted tryptamine, and the divisor 160 and multiplier 42 are the relative molecular masses of tryptamine and isocyanate, respectively.

Although tryptamine derivatives of isocyanates are both fluorimetrically and amperometrically active, fluorescence detection is preferred because of the higher stability and selectivity of the fluorimetric detector. All isocyanates except HDI can be detected at concentrations as low as  $0.5 \text{ ng ml}^{-1}$  in 120 l air samples.

### Pesticides

The use of fluorimetry for the determination of pesticides has been limited by the fact that relatively few of these compounds are strongly fluorescent and by the lack of selectivity of the technique for pesticide screening. In combination with LC (normal, reversed, and ionic phases), direct fluorimetric detection is used for the determination of pesticides exhibiting intrinsic fluorescence strong enough to be analytically useful (e.g., anticoagulant rodenticides such as indanediones and coumarins, carbamates insecticides such as aminocarb, bendiocarb, benomyl, carbaryl, and carbofuran, and some organophosphorus insecticides, including coumaphos and pirimiphos-methyl). Detection limits between 2 and  $20 \text{ ng ml}^{-1}$  are generally obtained.

Online postcolumn derivatization for LC-fluorimetric analysis of nonfluorescent or weakly fluorescent pesticides has mainly involved fluorogenic labeling of *N*-methyl carbamates and metabolites. The labeling reaction is a two-step process consisting of alkaline hydrolysis and posterior reaction of the hydrolysis product with *o*-phthalaldehyde and 2-mercaptoethanol to yield the highly fluorescent 1-hydroxyethylthio-2-methylisindole (EPA method 531.1). The detection limits for surface waters are between 20 and  $30 \text{ ng l}^{-1}$ .

### CE-Laser-Induced Fluorescence Detection

The applicability of CE in environmental analysis is usually rather limited because of poor concentration detection limits in CE compared with LC; however,

combination of CE with laser-induced fluorescence (LIF) detection has dramatically improved the limits of detection for fluorophores. CE-LIF detection of native fluorescence pollutants has focused on the determination of aromatic sulfonates such as naphthalene sulfonates and their amino and hydroxyl derivatives. The technique provides information about their structure, determining them at submicrogram per liter levels, and has been proved to be adequate for real-world environmental monitoring. The emission characteristics of the derivatives are strongly influenced by the nature and the position of the substituents.

On the other hand, derivatization procedures intended to employ LIF detection of nonfluorescent pollutants are numerous. Thus, various schemes for pesticide derivatization have been successfully used in order to determine carbamates (e.g., on-capillary derivatization with quaternary ammonium surfactants that catalyze their thermal decomposition to yield methylamine and use of a separation buffer containing *o*-phthalaldehyde/2-mercaptoethanol) and phenoxy acid herbicides, pyrethroid insecticides, or anilines (e.g., using precolumn derivatization with different fluorescent tags).

### LIF Spectroscopy for Onsite Measurements

The development of portable analytical instrumentation for onsite analysis of contaminants in soil and groundwater has grown significantly over the last decade. In this context, the combination of LIF using fiber optics with cone penetrometer test (LIF-CPT) systems has proven to reduce the costs and increase the efficiency to delineate contamination plumes involving fuels and other petroleum products.

Cone penetrometer testing is an effective means to deliver *in situ* subsurface sensor probes. It is based on driving into subsurface environment a conical steel tip under hydraulic pressure at a nearly constant rate. The tip is advanced by sequentially attaching a series of cylindrical steel extensions or 'push rods' to the assembly that are linked to one another by threadings. The thick-walled CPT push rods are hollow so that they can accommodate the optical fiber bundles.

The basic elements of a LIF-CPT system consist of a laser source, silica-clad silica optical fibers to conduct laser light to the sample and fluorescence to the detector, a sapphire optical window inserted into a brazed or welded channel in the push rod wall, and a detector.

In a typical LIF measurement, wavelength or excited-state decay time are fixed to produce a fluorescent signal from the sample. Calibration is done by

comparing the overall fluorescence pattern with those generated in the laboratory by using simulated samples. The LIF measurements are relatively nonselective, and the information does not necessarily correspond to common environmental designations of contaminants, such as total petroleum hydrocarbon, which are routinely used in soil characterization. Selectivity can be enhanced using multidimensional fluorescence measurements, for example, simultaneous excitation and emission wavelength, decay time, etc.

To date, three LIF-CPT systems have been developed and have proved their field applicability. The pioneering prototype was developed by Lieberman and co-workers at the Naval Research and Development (NRaD) Center in collaboration with the Army's WES laboratories. Later, GD Gillispie's group at North Dakota State University, sponsored by the US Air Force's Armstrong Laboratory, and JE Kenny's group at Tufts University have introduced novel variations with the aim of improving the performance of LIF.

Lieberman's system uses a pulsed nitrogen laser (output wavelength of 337 nm) and two silica-clad silica optical fibers to conduct laser light to the sample and fluorescence to a linear photodiode array detector. By operating the laser at 20 Hz, a fluorescence spectrum can be gathered in 1 s, providing a depth resolution of less than 5 cm at standard push rates. The data produced are most commonly reported as average fluorescence intensities versus depth, which effectively suppresses one dimension of information, the fluorescence emission wavelength, although all emission spectra are stored and may be retrieved for further analysis. The gated photodiode array permits lifetime data to be acquired by scanning the time delay of laser pulses.

In Gillispie's system an Nd:YAG laser is used to pump a rhodamine 6 G dye laser whose pulsed output (0.1 s) is frequency doubled to produce tunable UV light between 260 and 300 nm. It is capable of measuring decay time on the nanosecond scale with the use of a photomultiplier tube and digital oscilloscope. The system may also be stopped at a desired depth to permit the measurement of fluorescence decay curves at multiple emission wavelengths. The resulting data are arranged into a three-dimensional plot representing the wavelength-time matrix and can assist in the identification of species in the contamination mixture being investigated.

The LIF system developed by Kenny's group permits also three-dimensional fingerprint of the sample although it consists of fluorescence intensities as a function of excitation and emission wavelength rather than decay time and emission wavelength.

For this purpose, the system uses a Raman shifter pumped by the fourth harmonic of an Nd:YAG laser to generate multiple laser beams simultaneously in the UV-visible region of the spectrum. Ten beams are separated by a set of prisms and launched into separate silica-clad silica optical fibers and delivered to the sample. Separate collection fibers deliver the fluorescent signals to specific positions on an imaging spectrographic detector.

In field applications, LIF-CPT measurements are primarily used for screening purposes with the aim of delineating the approximate boundaries of a contamination plume so that permanent wells can be efficiently placed for long-term monitoring. For this purpose, the advantages of using the simplicity of Lieberman's LIF-CPT system are clear, since although single-ring aromatic compounds cannot be excited to fluorescence with the single excitation wavelength used, and two-ring aromatic compounds such as naphthalene only produce weak fluorescence, it would be unusual if a site were contaminated by smaller aromatics but not the less volatile, more refractory multiringed organics, which can be discerned by this probe. Gillispie's and Kenny's LIF-CPT systems can obtain plume delineation with increased information content; however, they require more time and more complex equipment.

## Low-Temperature Fluorescence

### Polycyclic Aromatic Hydrocarbons

The exceptionally sharp, quasilinear spectra provided by PAHs under cryogenic sampling have aroused special interest in low-temperature fluorescence measurements on this type of pollutant. It has become apparent that low-temperature fluorimetry can be used to fingerprint specific compounds present at very low levels in PAH mixtures in various environmental matrices (e.g., sediments and biota) without the need for extensive prior cleanup. Low-temperature fluorimetric methods used for the analysis of PAH mixtures rely largely on the Shpol'skii effect (in the frozen-solution or matrix-isolation mode).

The selectivity of the Shpol'skii effect lies in the inherently sharp absorption bandwidths of PAHs (FWHM  $\sim 1\text{--}10\text{ cm}^{-1}$ ) in appropriate frozen *n*-alkane hosts. Techniques based on this effect are usually very sensitive; provided an optimal photomultiplier tube is used, the detection limit for pyrene in a mixture of 10 PAHs in *n*-octane can be made as low as  $5 \times 10^{-10}\text{ mol l}^{-1}$  using an ordinary excitation lamp. The detection limit for benzo[*a*]pyrene is even lower by one order of magnitude; two orders



if laser excitation is employed, which means 20 amol in a 4  $\mu$ l sample volume. Such a high sensitivity is especially significant when samples are available in very small amounts or preconcentration methods are suspected to result in contamination. Shpol'skii fluorimetry in the frozen solution mode has been used successfully to distinguish isomeric alkylated PAHs in various environmental matrices with no prior separation. Thus, it has been applied to the characterization of the alkylated PAH content of diverse marine sediments of geochemical interest: the determination of the concentration ratio of various mono- and dimethylphenanthrenes or of total methylphenanthrenes to phenanthrene itself is a good measure of the maturity of organic matter in sediments.

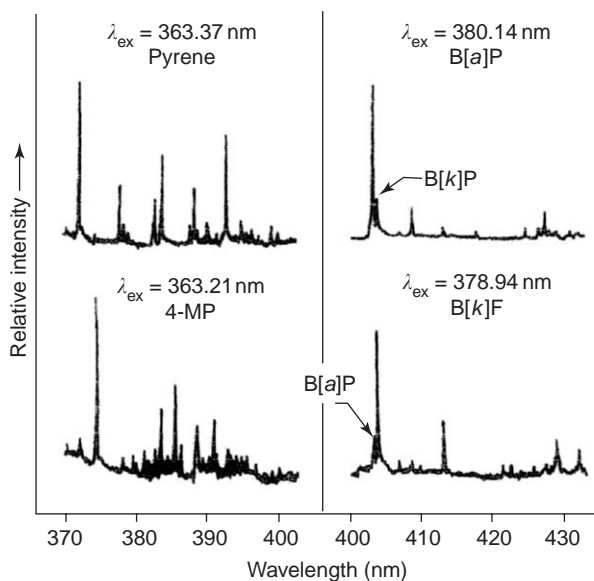
The need to use a sterically compatible solvent for each of the analytes in a mixture when implementing Shpol'skii fluorimetry poses no special problems since a single solvent such as *n*-heptane or *n*-octane usually suffices to obtain the sharp-line spectra of PAH mixtures containing compounds of three to six rings.

In quantitative analyses, the inner filter and enhancing effects may introduce a bias in the analytical results provided by Shpol'skii fluorimetry in the frozen-solution mode. Such a bias can usually be eliminated by internal standardization. Since the fluorescence in Shpol'skii matrices is so highly resolved, the deuterated analog of a PAH can be used as an internal standard for determining the PAH concerned by simply spiking the original sample solution with small volumes of alkane solutions of the required deuterated PAH prior to freezing. Computerized acquisition of data for a spectral scan of 200 nm can be accomplished in barely 5 min against 1–5 h by manual means. There are, however, some additional problems that have a direct bearing on the quantitative usefulness of the Shpol'skii effect in the frozen-solution mode. Thus, the linear dynamic range usually spans less than three orders of magnitude. In addition, since most compounds are less soluble at lower temperatures, solute aggregation can occur – and hence the precision be diminished – unless freezing is accomplished fairly rapidly and the analyte concentration is relatively low.

Matrix-isolation Shpol'skii fluorimetry has been shown to eliminate quantitative artifacts from quenching and aggregation phenomena occurring in Shpol'skii low-temperature solid matrices. The linear dynamic range for the fluorescence of PAHs in vapor-deposited alkane matrices surpasses that of solid-state Shpol'skii fluorimetry. Thus, the analytical calibration graph for the fluorescence of benzo[*a*]anthracene in *n*-heptane is linear from 30 pg (the detection limit) to a maximum amount of over 35  $\mu$ g. The

usual upper limit of linearity for calibration graphs is 1–10  $\mu$ g. PAHs with subpicogram detection limits typically exhibit linear dynamic ranges of six decades. *n*-Alkenes used as matrices in the matrix-isolation mode yield quasilinear spectra similar to those provided by Shpol'skii frozen solutions. This technique has been used to resolve isomeric alkylated PAHs in various environmental matrices. The chief advantage of frozen-solution methods over matrix-isolation methods lies in their greater simplicity: preparing a sample by freezing a liquid solution is much less time consuming than performing matrix isolation.

Very complex mixtures such as coal-derived materials often result in undesirably overlapping spectra for PAHs in Shpol'skii matrices when a xenon lamp is used as the radiation source. In such cases, provided a sufficiently monochromatic source is used, the fluorescence may be selectively excited from one specific compound in a complex mixture of very similar PAHs. In practice, selective excitation usually entails using a laser for excitation. LIF in Shpol'skii matrices has been used to identify and quantify individual PAHs in highly complex coal-derived materials. By way of example, **Figure 2** shows the site-selective fluorescence of four individual PAHs in a solvent-retained coal liquid sample. The spectra were recorded following a 1000-fold dilution of the



**Figure 2** Site-specific, laser-induced fluorescence spectra of pyrene, 4-methylpyrene (4-MP), B[a]P, and B[k]F present in a solvent-retained coal liquid (SRC-11) sample measured in a frozen *n*-octane solution at 15 K. The excitation wavelengths for the PAHs are given. (Reprinted with permission from Yang Y, O'Silva AP, Fassel VA, and Iles M (1980) Direct determination of polynuclear aromatic hydrocarbons in coal liquids and shale oil by laser excited Shpol'skii spectrometry. *Analytical Chemistry* 52: 1350–1351; © American Chemical Society.)

sample in *n*-octane. The excitation wavelengths used are shown in the figure. It should be noted that the selected excitation wavelength for benzo[*a*]pyrene and benzo[*k*]fluoranthene [B[*k*]F] resulted in weak emission from B[*k*]F and B[*a*]P, respectively, owing to partial overlap of primary absorption lines of these compounds at such a wavelength. However, the emission lines are sufficiently resolved to enable the quantitative determination of these compounds in mixtures. Note the high resolution and clean appearance of the spectra – despite the high complexity of the coal liquid sample – and the need for no fractionation steps.

In summary, there seems to be no fluorimetric alternative to excel low-temperature techniques in analytical information throughput if chemical separations are to be avoided. Massive use in the environmental field has so far been restricted by the need to use fairly sophisticated devices such as high-resolution monochromators, cryostats, and, perhaps, dye lasers. On the other hand, a number of commercially available spectrofluorimeters provide inadequate spectral resolution for use in low-temperature spectroscopic techniques. However, given the high selectivity of such techniques, their use in the characterization of complex environmental samples will continue to grow in the foreseeable future.

## Immunoanalytical Techniques

### Pesticides

The development and application of immunoassays and immunosensors based on fluorescence detection for environmental monitoring has grown significantly in recent years, with particular emphasis on monitoring of pesticide levels. Methods with fluorescent labels have the potential to achieve the low limits of detection (LOD) imposed by legislation, but high background signals have restricted in the past the LOD obtained to concentration ranges between  $10^{-9}$  and  $10^{-12}$  mol l<sup>-1</sup>. The development of solid-phase separation systems, new fluorescent probes, and new instrumentation has lowered the background such that sensitivity is now comparable with that of radioactive methods.

Background signals are mainly caused by light scattering and background fluorescence. They can be avoided by use of fluorescent probes with a large Stokes shift and maximal excitation and emission in the far-red wavelength region. Also, measuring time-resolved fluorescence can remove scattered light interferences since scattering has no fluorescence lifetime. On the other hand, reduced background fluorescence can be obtained from probes with

fluorescence lifetimes different from the fast decaying background. In addition to employing suitable labels, different detection techniques and instrumentation can contribute to the reduction of background fluorescence.

Basic test formats used in pesticide monitoring include homogeneous and heterogeneous immunoassays in batch format, flow injection immunoanalysis, and immunosensors. Phase-resolved, modulation-resolved, and kinetic measurements are used for minimizing the effect of sample matrix and improving the LODs in homogeneous assays. Fluorescence polarization is the most widely used detection technique for this test format, although fluorescence quenching and enhancement, and fluorescence lifetime have also achieved progressive applicability.

In heterogeneous immunoassays in batch format the problem of background fluorescence is reduced by separating the analyte from other fluorescent species present in the sample matrix. The use of enzymes with fluorogenic substrates in fluorescence enzyme-linked immunosorbent assays (F-ELISA) and Eu(III)-chelates in time-resolved fluoroimmunoassay (TRFIA) permits to achieve very low LODs for pesticides (e.g., between 0.023–0.3 and 0.05–0.4 µg l<sup>-1</sup> for F-ELISA and TRFIA, respectively).

In flow-injection methods and sensors the background effects are minimized, either because the matrix does not pass the detection unit or due to evanescent field techniques, where only surface bound material is excited. An additional advantage is reusability of the devices. This enables automation and online monitoring, but stability of both the reagents and the transducers has to be considered. LODs obtained by using flow-injection methods and immunosensors are comparable although detection limits can be improved by approximately one order of magnitude in flow methods that use enzyme labels because of the amplification effect of the enzyme reaction. Trends on these test formats are focusing on miniaturization. Especially evanescent field immunosensors have a great potential to become small and cost-effective devices for onsite environmental monitoring.

One can conclude that fluorimetry is not as widely used for the analysis of environmental samples as it is in other areas (particularly the life sciences). However, the high sensitivity and, under appropriate conditions, high selectivity, make fluorimetric techniques appealing candidates for the characterization of environmental materials.

**See also:** Environmental Analysis. Fluorescence: Instrumentation; Quantitative Analysis. Immunoassays,

**Techniques:** Luminescence Immunoassays. **Laser-Based Techniques.** **Polycyclic Aromatic Hydrocarbons:** Determination; Environmental Applications.

## Further Reading

- Aldstadt JH and Martin AF (1997) Analytical chemistry and the cone penetrometer: *In situ* chemical characterization of the subsurface. *Mikrochimica Acta* 127: 1–18.
- Ariese F, Gooijer C, and Velthorst VH (1993) Application of fluorescence spectroscopic techniques in the determination of PAHs and PAH metabolites. In: Barceló D (ed.) *Environmental Analysis: Techniques, Applications and Quality Assurance*, pp. 449–480. Amsterdam: Elsevier.
- Coly A and Aaron JJ (1998) Fluorimetric analysis of pesticides: Methods, recent developments and applications. *Talanta* 46: 815–843.
- Garrigues P and Budzinski H (1995) Recent analytical advances in Shpol'skii spectroscopy. *Trends in Analytical Chemistry* 14: 231–239.

- Henderson-Kinney A and Kenny JE (1995) Spectroscopy in the field. Emerging techniques for on-site environmental measurements. *Spectroscopy* 10: 32–38.
- Manoli E and Samara C (1999) Polycyclic aromatic hydrocarbons in natural waters: Sources, occurrence and analysis. *Trends in Analytical Chemistry* 18: 417–428.
- Pumell CJ and Walker RF (1995) Methods for the determination of atmospheric organic isocyanates. A review. *Analyst* 110: 893–905.
- Schobel U, Barzen C, and Gauglitz G (2000) Immunoanalytical techniques for pesticide monitoring based on fluorescence detection. *Fresenius Journal of Analytical Chemistry* 366: 646–658.
- Wu WS, Stoyanoff RE, Szklar RS, Gai ND, and Rakanovic M (1990) Application of tryptamine as a derivatizing agent for airborne isocyanates determination: Part 3. Evaluation of total isocyanates analysis by high-performance liquid chromatography with fluorescence and amperometric detection. *Analyst* 115: 801–807.

## Food Applications

**A Gómez-Hens**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Fluorescence spectroscopy plays an important function in modern food analysis as can be seen from its wide use in the determination of numerous food components, contaminants, additives, and adulterants. This technique has made available very sensitive and selective methods that satisfy the requirements of food analysis, which are usually very complex, taking into account the large number of species to be determined, frequently at very low concentrations, and the wide variety of foodstuffs available. Initially, the use of fluorescence spectroscopy in food analysis was limited to the determination of species with intrinsic fluorescence (e.g., vitamins, aflatoxins, and some polycyclic aromatic hydrocarbons (PAHs)), but now it is widely applied to nonfluorescent species, using several physicochemical means such as chemical or photochemical derivatization reactions. Numerous techniques involve fluorescence detection in liquid chromatography (LC), frequently using pre- or postcolumn derivatization. In addition to conventional fluorimetry, which is commonly chosen for this purpose, other fluorimetric techniques such as laser-induced

fluorescence (LIF) spectroscopy, fluoroimmunoassays (FIAs), time-resolved fluorescence spectroscopy, and derivative synchronous fluorescence spectroscopy have been shown to be useful in food analysis. There has been recent interest in the use of capillary electrophoresis (CE) with LIF as a detection system. Although the application of immunochemical techniques with fluorimetric detection in food analysis is still relatively limited, there are interesting examples of their suitability for the detection of some proteins and for the control of contamination by toxins and pesticide residues.

## Determination of Food Components

### Vitamins

Vitamins are the foodstuff components most often quantified using fluorimetric means. There are several official fluorimetric methods for the determination of three water-soluble vitamins: vitamin B<sub>1</sub> (thiamine) (AOAC 942.23 and 957.17), B<sub>2</sub> (riboflavine) (AOAC 970.65 and 981.15), and C (ascorbic acid) (AOAC 984.26). Thiamine is determined by oxidation to fluorescent thiochrome with alkaline hexacyanoferrate(III) or an alternative oxidant (Figure 1). The method is quite simple, reproducible, and selective and provides good recoveries. Many LC methods for thiamine determination in foods have

**Techniques:** Luminescence Immunoassays. **Laser-Based Techniques.** **Polycyclic Aromatic Hydrocarbons:** Determination; Environmental Applications.

## Further Reading

- Aldstadt JH and Martin AF (1997) Analytical chemistry and the cone penetrometer: *In situ* chemical characterization of the subsurface. *Mikrochimica Acta* 127: 1–18.
- Ariese F, Gooijer C, and Velthorst VH (1993) Application of fluorescence spectroscopic techniques in the determination of PAHs and PAH metabolites. In: Barceló D (ed.) *Environmental Analysis: Techniques, Applications and Quality Assurance*, pp. 449–480. Amsterdam: Elsevier.
- Coly A and Aaron JJ (1998) Fluorimetric analysis of pesticides: Methods, recent developments and applications. *Talanta* 46: 815–843.
- Garrigues P and Budzinski H (1995) Recent analytical advances in Shpol'skii spectroscopy. *Trends in Analytical Chemistry* 14: 231–239.

- Henderson-Kinney A and Kenny JE (1995) Spectroscopy in the field. Emerging techniques for on-site environmental measurements. *Spectroscopy* 10: 32–38.
- Manoli E and Samara C (1999) Polycyclic aromatic hydrocarbons in natural waters: Sources, occurrence and analysis. *Trends in Analytical Chemistry* 18: 417–428.
- Pumell CJ and Walker RF (1995) Methods for the determination of atmospheric organic isocyanates. A review. *Analyst* 110: 893–905.
- Schobel U, Barzen C, and Gauglitz G (2000) Immunoanalytical techniques for pesticide monitoring based on fluorescence detection. *Fresenius Journal of Analytical Chemistry* 366: 646–658.
- Wu WS, Stoyanoff RE, Szklar RS, Gai ND, and Rakanovic M (1990) Application of tryptamine as a derivatizing agent for airborne isocyanates determination: Part 3. Evaluation of total isocyanates analysis by high-performance liquid chromatography with fluorescence and amperometric detection. *Analyst* 115: 801–807.

## Food Applications

**A Gómez-Hens**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Fluorescence spectroscopy plays an important function in modern food analysis as can be seen from its wide use in the determination of numerous food components, contaminants, additives, and adulterants. This technique has made available very sensitive and selective methods that satisfy the requirements of food analysis, which are usually very complex, taking into account the large number of species to be determined, frequently at very low concentrations, and the wide variety of foodstuffs available. Initially, the use of fluorescence spectroscopy in food analysis was limited to the determination of species with intrinsic fluorescence (e.g., vitamins, aflatoxins, and some polycyclic aromatic hydrocarbons (PAHs)), but now it is widely applied to nonfluorescent species, using several physicochemical means such as chemical or photochemical derivatization reactions. Numerous techniques involve fluorescence detection in liquid chromatography (LC), frequently using pre- or postcolumn derivatization. In addition to conventional fluorimetry, which is commonly chosen for this purpose, other fluorimetric techniques such as laser-induced

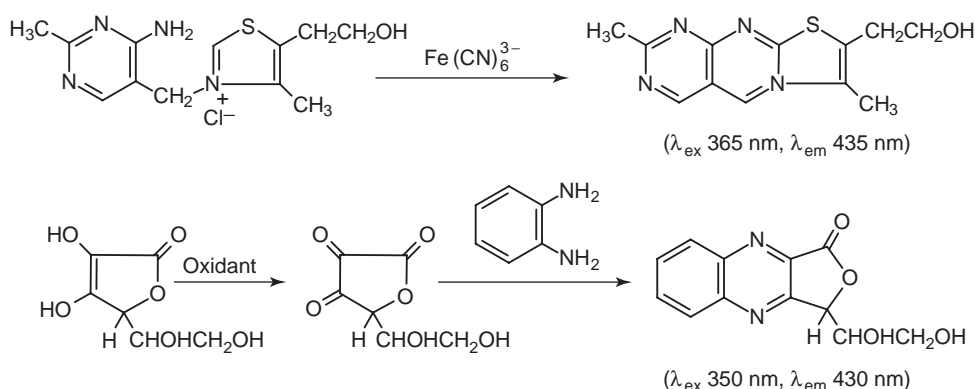
fluorescence (LIF) spectroscopy, fluoroimmunoassays (FIAs), time-resolved fluorescence spectroscopy, and derivative synchronous fluorescence spectroscopy have been shown to be useful in food analysis. There has been recent interest in the use of capillary electrophoresis (CE) with LIF as a detection system. Although the application of immunochemical techniques with fluorimetric detection in food analysis is still relatively limited, there are interesting examples of their suitability for the detection of some proteins and for the control of contamination by toxins and pesticide residues.

## Determination of Food Components

### Vitamins

Vitamins are the foodstuff components most often quantified using fluorimetric means. There are several official fluorimetric methods for the determination of three water-soluble vitamins: vitamin B<sub>1</sub> (thiamine) (AOAC 942.23 and 957.17), B<sub>2</sub> (riboflavine) (AOAC 970.65 and 981.15), and C (ascorbic acid) (AOAC 984.26). Thiamine is determined by oxidation to fluorescent thiochrome with alkaline hexacyanoferrate(III) or an alternative oxidant (Figure 1). The method is quite simple, reproducible, and selective and provides good recoveries. Many LC methods for thiamine determination in foods have





**Figure 1** Formation of fluorescent derivatives of vitamins B<sub>1</sub> and C.

been developed using pre- or postcolumn conversion to thiochrome, which entail a wide variety of chromatographic modes and elution techniques.

Riboflavin is commonly determined fluorimetrically, for instance in milk, by its strong native fluorescence at pH 7, which arises from the extended conjugation and rigidity of the nonribose portion of the molecule. Another fluorimetric method involves conversion of riboflavin into its fluorescent derivative lumiflavin using ultraviolet (UV) irradiation. Mixtures of thiamin and riboflavin in foods such as cereal products have been resolved using LC with fluorimetric detection.

Vitamin C is another soluble vitamin frequently determined in foodstuffs using fluorimetric detection. It is most often oxidized to dehydroascorbic acid with  $\text{HgCl}_2$  or 2,6-dichloroindophenol, which is subsequently reacted with *p*-phenylenediamine to form a fluorescent quinoxaline derivative (Figure 1). The method can be implemented in an automated fashion. The determination of total vitamin C (ascorbic and dehydroascorbic acid) in foods (vegetables and fruits) is also of great interest. The recommended method uses reversed-phase ion-pair LC and fluorimetric detection. Total vitamin C and total isovitamin C (erythorbic and dehydroerythorbic acids) can be determined similarly by prior extraction with trichloroacetic acid and precolumn derivatization, wherein ascorbic and erythorbic acids are oxidized enzymatically and total dehydroascorbic and dehydroerythorbic acids are treated with *o*-phenylenediamine.

The fluorimetric determination of vitamin B<sub>6</sub> compounds (pyridoxine, pyridoxamine, pyridoxal, pyridoxamine 5'-phosphate, pyridoxal 5'-phosphate, and 4-pyridoxic acid) in foods by measurement of their native fluorescence entails a prior LC isolation. Some of these determinations involve postcolumn derivatization with sodium bisulfite in a phosphate

buffer. Mixtures of riboflavin and pyridoxine in infant formula products can be determined simultaneously using ion-pair LC with fluorescence detection.

Fluorimetric LC methods for the determination of fat-soluble vitamins in foods have been described for vitamins A (retinol), E, and K. These methods usually entail saponification and extraction steps. The determination of vitamin A, which shows intrinsic fluorescence, in foods containing large amounts of carotenoids, requires chromatographic purification. This technique has also been used in studying the extent of vitamin A isomerization during manufacture and storage of liquid and dried milk products, which involves prior isolation of the different isomers using LC.

Estimation of true vitamin E in foods requires quantitative determination of all its components since they vary in their biological potency. This vitamin consists of four tocopherols ( $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$ ) and four tocotrienols ( $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$ ), but the three major constituents responsible for vitamin E activity are the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -tocopherols. While these compounds are fluorescent, their esters must be reduced to free alcohols for total tocopherol assays. Total vitamin E can be directly obtained through fluorimetry, but the determination of individual components is carried out using LC with fluorimetric detection. This procedure has been used to determine the composition of vitamin E in seed oils from maize, olives, soya beans, sesame, safflower, and sunflower by measuring the content of all the four tocopherols plus  $\alpha$ -tocotrienol. The simultaneous determination of tocopherols, carotenes, and retinol in cheese has been carried out using LC with two programmable detectors connected in series, a spectrophotometer and a fluorimeter. Carotenes have been determined photometrically, and fluorimetric measurements have been obtained for tocopherol and retinol.

The vitamin K group is usually determined in foods such as fats, oils, and milk through LC with fluorimetric detection after postcolumn zinc reduction. Samples are usually digested with lipase and extracted into hexane.

### Proteins, Peptides, Amino Acids, and Amines

Proteins in foods such as milk have been the subject of much fluorimetric study involving measurement of the intrinsic fluorescence of tryptophan and tyrosine; however, since proteins vary in their content of these amino acids, the fluorescence intensity also varies markedly from protein to protein and, for a single protein, with the experimental conditions. Also, the natural fluorescence of peptides is limited to peptide-containing tryptophans and tyrosines. However, the joint use of LC and derivatizing fluorogenic agents such as dansyl chloride, ninhydrin, fluorescamine, and *o*-phthaldialdehyde (OPA) has allowed the development of a number of methods for the determination of proteins, peptides, amino acids, and amines in food samples.

Dansyl chloride has the disadvantage that any excess of reagent or its degradation products must be removed before the derivatives can be chromatographically separated. Ninhydrin derivatives are reportedly superior and lend themselves readily to automated use. Because ninhydrin is not fluorescent, any excess remaining in solution poses no interference, but the derivatization process involves some heating. Fluorescamine reacts very rapidly at an alkaline pH with primary amino groups giving rise to fluorescent derivatives and can be used in either aqueous solutions or organic solvents; however, the fluorescence yield of its derivatives is ~10 times lower than those of OPA derivatives. Peptide determination methods involving LC have been described using pre- or postcolumn derivatization with fluorescamine. OPA is a highly versatile reagent, even though it does not react with proline or hydroxyproline, and gives a very weak fluorescence with cysteine. It reacts rapidly with primary amino groups at pH 9–11 in the presence of a thiol such as 2-mercaptoethanol (2-ME), 3-mercaptopropionic acid, 3-mercapto-1-propanol, or ethanethiol, yielding is-indole fluorescent derivatives. OPA has been widely used as a derivatizing reagent in food analysis using LC separations. For instance, through precolumn derivatization, polar amino acids are converted into hydrophobic compounds that can be rapidly separated on a reversed-phase column. The amino acids can also be isolated on an ion-exchange column, followed by postcolumn derivatization. It has also been used to determine peptidase activity, based on the

fact that the fluorescence yield of the OPA derivatives of peptides is much lower than it is for amino acids. Some methods for peptide determination in foods involve the use of OPA fluorescence detection online coupled with absorbance detection. The usefulness of this coupling for obtaining additional information has been shown for the selective determination of peptides in blue cheeses and skimmed milk.

Some fluorimetric methods for the individual determination of amino acids in foods have also been reported. Thus, the native fluorescence of tryptophan has been used for its determination in food and feed hydrolysates using ion-exchange chromatography. Also, 3-methylhistidine has been determined in meat and meat products using LC and precolumn derivatization with fluorescamine or postcolumn derivatization with OPA and 2-ME.

Although the use of CE in food analysis is limited compared with LC, many applications have been reported in recent years in the study of food proteins and peptides. It has been shown that the sensitivity provided by CE using LIF detection for bovine whey proteins is twice as good as that provided by UV detection, under the same separation conditions. LIF detection has also been used together with affinity interactions to enhance the detection sensitivity in CE. For instance, an immunoassay detection method has been described using a polyclonal antibody marked with fluorescein to determine lactoferrin. CE–LIF has also been applied to the determination of amines and amino acids in several samples such as wine and milk products.

The usefulness of kinetic methodology together with a long wavelength fluorophor has been shown for the determination of gliadins and caseins in foods. These methods involve the use of a cationic (Cresyl Violet) or anionic fluorophor (Indocyanine Green) in the presence of an anionic (sodium dodecyl sulfate) or cationic cetyltrimethylammonium bromide (CTAB) surfactant, respectively. The variation of the fluorescence intensity with time is the result of ionic forces between oppositely charged molecules, which depend on the protein concentration. These methods combine the temporal and spectral discrimination of the analytical signal, which allows the required selectivity to be obtained.

One of the few reported uses of fiber-optic biosensors in food analysis in combination with flow injection analysis is the determination of L-glutamate in soups and sauces. This amino acid plays a central role in the oxidative determination of other amino acids and is capable of sensitizing gustatory nerves. Although glutamate occurs naturally in some foods, it is also used as a flavor enhancer. The biosensor was



an oxygen-sensitive optrode with immobilized glutamate oxidase and decacyclene, being the fluorescence of decacyclene. The fluorescence of decacyclene is markedly dependent on the oxygen concentration.

Histamine in foods is primarily the result of microbial decarboxylation of histidine, and so it is a customary constituent of fermented foodstuffs such as wine and cheese, and foods with a fairly high histidine concentration (e.g., tuna fish) that have been exposed to microbial degradation. It is routinely determined in foods using fluorimetric methods involving condensation with OPA. Second-derivative synchronous fluorescence spectrometry has also been used for the direct quantification of histamine in wine and canned tuna.

### Inorganic Species

Although metal ions in foods are usually determined using atomic absorption spectrometry, there are a few fluorimetric methods based on the formation of fluorescent complexes. Thus, very low concentrations of selenium in milk, eggs, cheese, meat, and vegetables can be determined using an official method (AOAC 974.15) that involves the use of 2,3-diaminonaphthalene as a reagent. The fluorescent piazselenol formed is extracted into cyclohexane using ethylenediaminetetraacetic acid as a masking agent. Aluminum has been determined in various foodstuffs and tap water using different fluorimetric ligands such as Calcon and 2,4-dihydroxybenzaldehyde iso-nicotinoylhydrazone.

Zinc is determined in milk using an indirect kinetic fluorimetric method based on its activating effect on the Mn(II)-catalyzed oxidation of 2-hydroxybenzaldehyde thiosemicarbazone by hydrogen peroxide. The Mn(II)-catalyzed oxidation of a similar reagent, 2-hydroxynaphthaldehyde thiosemicarbazone, by hydrogen peroxide has been exploited for the kinetic fluorimetric determination of the latter in various types of coffee, tea, and milk. Another application of kinetic methodology in fluorimetric food analysis is the determination of iodide in table salt and milk using its catalytic action on the Ce(IV)–As(III) system, which involves measurement of the fluorescence of the Ce(III) formed and uses the stopped-flow mixing technique for automatic implementation.

## Determination of Food Contaminants

### Polycyclic Aromatic Hydrocarbons

These compounds are primarily found in smoked foodstuffs and roasted products including malt and

coffee as they arise from combustion of some organic materials. Since PAHs with rigid, extensively conjugated structures exhibit strong native fluorescence, they are largely determined fluorimetrically. LC coupled to a spectrofluorimeter detector is a powerful technique used widely for the determination of PAHs in foods. In this sense, benzo[*a*]pyrene is the PAH most commonly present in contaminated foods and hence that is most frequently analyzed for.

Owing to the typical complexity of the sample matrix, the fluorimetric determination of PAHs in foodstuffs entails extraction with *n*-hexane, cyclohexane, or acetone–toluene, either directly or, usually, after alcoholic–alkaline saponification. The residue is cleaned up on a column packed with silica gel, Florisil, or Sephadex, and the eluted fraction containing the PAH is analyzed using LC with, usually, acetonitrile as the mobile phase. Both isocratic and gradient elution have been described. The best way of resolving a mixture of PAHs in the eluate is through programmed fluorescence detection, i.e., use of variable excitation and emission wavelengths. An LC method for the determination of 16 PAHs in different nonfatty foods has been proposed using a gradient of acetonitrile–water as the mobile phase, together with a program of 10 excitation and emission wavelength pairs. LC with fluorimetric detection has been used widely to analyze a variety of foodstuffs (smoked foods, edible vegetable oils, shellfish, coffee, tea, malt, cereals, mineral water, spices, herbs) with good recoveries and very low detection limits (of a few nanograms per kilogram). A method proposed for the individual determination of benzo[*a*]pyrene in foods involves treatment of the sample with sodium sulfate, extraction into *n*-hexane/ethyl ether and cleanup with alumina. Phenanthrene derivatives have been determined in wood smoke from a food-smoking compartment by injecting the sample, after extraction and cleanup, into a gas chromatography (GC)–mass spectrometry system, but the detection limit thus achieved can be lowered 10-fold for benzo[*a*]pyrene using the Shpol'skii low-temperature fluorescence technique.

Fluorodensitometry has been used fairly regularly for the determination of PAHs in foods after separation using thin layer chromatography (TLC). Although the resolving power of this technique is lower than that of LC, it provides similar detection limits. There is an official method (AOAC 991.06) in which PAHs are extracted from the food sample after saponification with alcoholic KOH, purified using solvent partition and column chromatography, separated using TLC, determined using UV spectrophotometry, and confirmed using spectrofluorimetry.

## Mycotoxins

There are probably a few hundred mycotoxins, encompassing a wide range of structural and chemical types, although only a limited number occur frequently at significant concentrations in foods. These compounds are secondary metabolites produced by molds (fungi) that have a toxic effect on man and/or animals. Many mycotoxins show native fluorescence, so that the fluorescence of fungal materials is frequently used in screening raw materials in order to reject contaminated items or batches.

An important group of mycotoxins is formed from aflatoxins, which are produced by some *Aspergillus* varieties and occur in many crops used as food and animal feed ingredients, particularly grains and nuts. The aflatoxins more frequently analyzed are B<sub>1</sub>, B<sub>2</sub>, G<sub>1</sub>, and G<sub>2</sub>, which are highly fluorescent as they have a highly conjugated five-ring cyclic structure. In fact, their names are based on their fluorescent color (B for blue and G for green). TLC with fluorimetric detection has been used extensively for aflatoxin determination as there are several firmly established official methods based on this technique (AOAC 982.25, 993.17, and 998.03). There is also a similar official method for the determination of sterigmatocystin (AOAC 973.38), which is a precursor of aflatoxins. However, the present trend is the use of LC with fluorimetric detection using normal or reverse-phase methods. A limitation of this approach is the fluorescence dependence of these compounds on the composition of the solvent. For instance, the fluorescence of aflatoxins B<sub>1</sub> and B<sub>2</sub> is markedly quenched by solvents frequently used in normal-phase LC. Also, the use of reversed-phase LC is limited by the fact that the fluorescence of aflatoxins B<sub>1</sub> and G<sub>2</sub> is quenched in aqueous medium.

Several derivatization procedures have been proposed to circumvent these shortcomings, involving the use of trifluoroacetic acid (TFA), iodine, bromine, or pyridinium bromide perbromide. TFA transforms aflatoxins B<sub>1</sub> and G<sub>1</sub> into their highly fluorescent hemiacetals, B<sub>2a</sub> and G<sub>2a</sub>, while aflatoxins B<sub>2</sub> and G<sub>2</sub> are not affected by this reagent due to their saturate structure. A limitation of TFA derivatization is the low stability of the products in methanol, probably due to the formation of methyl acetals. Official methods for aflatoxin determination in several foods are based on TFA precolumn derivatization (AOAC 990.33 and 994.08). One of these methods involves the use of a multifunctional column containing a mixture of reverse-phase, ion-exclusion, and ion-exchange sorbents, which retain interfering substances such as fats, proteinaceous compounds, pigments,

**Table 1** Performance characteristics for aflatoxin determination using LC with fluorimetric detection

Aflatoxin	Matrix	Dynamic range (ng per g)	Recovery (%)	RSD <sup>a</sup> (%)
Total	Almonds	5–30	116	16
	Brazil nuts	5–30	97–140	9
	Corn	5–30	98–102	6–20
	Peanuts	5–30	74–91	7
	Pistachios	5–30	91–92	23
	Almonds	3–15	91–95	9
B <sub>1</sub>	Brazil nuts	3–15	94–140	9
	Corn	3–15	97–103	4–21
	Peanuts	3–15	77–87	7
	Pistachios	3–15	88–92	28
B <sub>2</sub>	Almonds	1–3	88–92	–
	Brazil nuts	1–3	83–98	10
	Corn	1–3	91–98	10
	Peanuts	1–3	89–93	8
G <sub>1</sub>	Pistachios	1–3	72–89	10
	Almonds	2–9	89–103	16
	Brazil nuts	2–9	111–153	13
	Corn	0.9–9	102–108	9
	Peanuts	0.9–9	71–96	15
	Pistachios	3–9	96–100	21
G <sub>2</sub>	Almonds	1–3	98–116	–
	Brazil nuts	1–3	97–153	16
	Corn	1–3	108–115	18
	Peanuts	1–3	93–101	14
	Pistachios	1–3	76–80	21–27

<sup>a</sup>RSD, relative standard deviation.

and carbohydrates extracted from food and feed ingredients. Aflatoxins are eluted from the column, derivatized with TFA, and quantified using LC with fluorescence detection. **Table 1** shows some features of the application of this method to several samples. The availability of automated LC systems with post-column derivatization is preferable for reducing the manipulations required for each sample. However, the use of TFA for postcolumn derivatization is limited because of its toxicity and its corrosive properties on the pumping devices. Alternative postcolumn derivatization methods have been proposed. Thus, aflatoxin hemiacetals have been generated with post-column in-line photochemical UV irradiation, which can be detected at subpicogram levels using LIF. Also, postcolumn derivatization with iodine has been adopted as an official method (AOAC 991.31), which involves the use of an immunoaffinity column containing an immobilized antibody to aflatoxins B<sub>1</sub>, B<sub>2</sub>, G<sub>1</sub>, and G<sub>2</sub>, in which aflatoxins are isolated, purified, and concentrated before their LC separation.

In addition to aflatoxins, there are other mycotoxins that can be determined fluorometrically. Official methods have been established for deoxynivalenol (AOAC 986.17), ochratoxin A (AOAC

973.37 and 991.44), zearalenone (AOAC 985.18), and fumonisins (AOAC 995.15), using TLC or LC and fluorescent detection. All these compounds exhibit native fluorescence, except fumonisins, which require a previous derivatization step using OPA and 2-ME.

The usefulness of immunochemical techniques for the fluorimetric determination of mycotoxins has been widely demonstrated. In addition to the use of immunoaffinity column cleanup followed by LC with fluorescent detection, several FIAs have been described for the determination of aflatoxins, deoxynivalenol, and fumonisins using fluorescence polarization immunoassay or time-resolved fluoroimmunoassay. Also, although the application of fluoroimmunosensors to food analysis has been relatively limited, there are some examples of the use of this approach to the determination of some mycotoxins such as aflatoxin B<sub>1</sub> and fumonisin B<sub>2</sub>.

### Phycotoxins

Phycotoxins, also named shellfish toxins, are produced by free-living micro-algae upon which the shellfish feed. The toxins are concentrated in the shellfish, which act as a vector transferring the toxic compounds to the food chain. Control of the presence of these toxins in food is required as they can cause neurotoxic, diarrhetic, paralytic, or amnesic poisoning. LC methods with fluorescence detection are now available for the determination of some of these compounds, such as domoic acid, saxitoxins, okadaic acid, and ciguatoxins. Also, an enzyme inhibition assay has been described for the determination of okadaic acid in mussels using fluorescent endpoint detection.

### Nitrite, Nitrate, and Nitrosamines

The nitrate content in foods of animal and plant origin is significant because nitrate can be reduced to nitrite, which reacts with hemoglobin to form methemoglobin. Nitrite can also react with secondary amines to form *N*-nitroso compounds, which have been shown to be carcinogenic, mutagenic, and teratogenic. While volatile dialkyl nitrosamines are amenable to determination using GC, nonvolatile derivatives are commonly assayed using LC. Thus, *N*-nitrosopyrrolidine, which has been detected at the nanogram per gram level in fried bacon, is determined using LC with fluorimetric detection. The nitrosamine is treated with hydrogen bromide in acetic acid, and the resulting amine is derivatized with 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole to give a

highly fluorescent product that is injected into the chromatographic system. Nitrosamines can also be extracted from foods with ethyl ether, reduced to secondary amines with Ni–Al alloy, and derivatized with fluorescamine.

Very low concentrations of nitrite in milk have been determined through the formation of 1-[*H*]-naphthotriazole using 2,3-diaminonaphthalene as a reagent. Another fluorimetric method for nitrite determination in cured meat products relies on diazotization of 2-aminophenol and coupling with resorcinol. The resulting 2,2',4-trihydroxyazobenzene derivative is reacted with gallium to give a strong orange-red fluorescence. Nitrate can also be determined fluorimetrically in tap water by reaction with 2-phenylbenzothiazole.

### Pesticide Residues

Pesticide residues in foods are basically determined using GC, although this technique shows some limitations for some pesticides. For instance, in the case of *N*-methylcarbamates, most of them are either retained on the chromatographic column or decomposed to their corresponding phenols. Also, GC derivatization procedures for the determination of aromatic carbamates with electron-capture detection, which involve initial hydrolysis to the corresponding phenols or amines and reaction with halogen-rich reagents, are often subject to some limitations that can detract from their sensitivity and applicability. LC procedures are reportedly more convenient than GC procedures for a wide range of carbamates, both in uncombined form through UV detection or after derivatization, usually with OPA and 2-ME, through fluorescence detection. There is an official method (AOAC 985.23) based on this approach that is applicable to residues of aldicarb, bufenarb, carbaryl, carbofuran, methiocarb, methomyl, oxamyl, and the metabolites aldicarb sulfone and 3-hydroxycarbofuran in grapes and potatoes. Some features of these methods are summarized in Table 2. LIF detection has been employed to quantify *N*-methylcarbamates through LC after precolumn derivatization with 7-chloro-4-nitrobenzoxa-1,3-diole.

Substituted ureas, such as phenylureas, are generally determined using LC, as they can be degraded to isocyanates in GC. Although detection is usually through UV absorption, phenylurea residues in various foods including maize and some vegetables and fruits have also been determined through postcolumn UV photolysis and OPA derivatization.

Several pesticides have been determined by measuring the fluorescence of their hydrolysis products after chromatographic separation. Figure 2 shows

the products obtained for the pesticides guthion, coumaphos, and zinophos. This procedure has been used to quantify guthion and coumaphos in milk and zinophos in vegetables. The intrinsic fluorescence of some pesticides allows their direct determination

**Table 2** Analytical performance of some high-performance liquid chromatography (HPLC) methods with an OPA-2ME derivatization reaction for carbamate pesticide determination

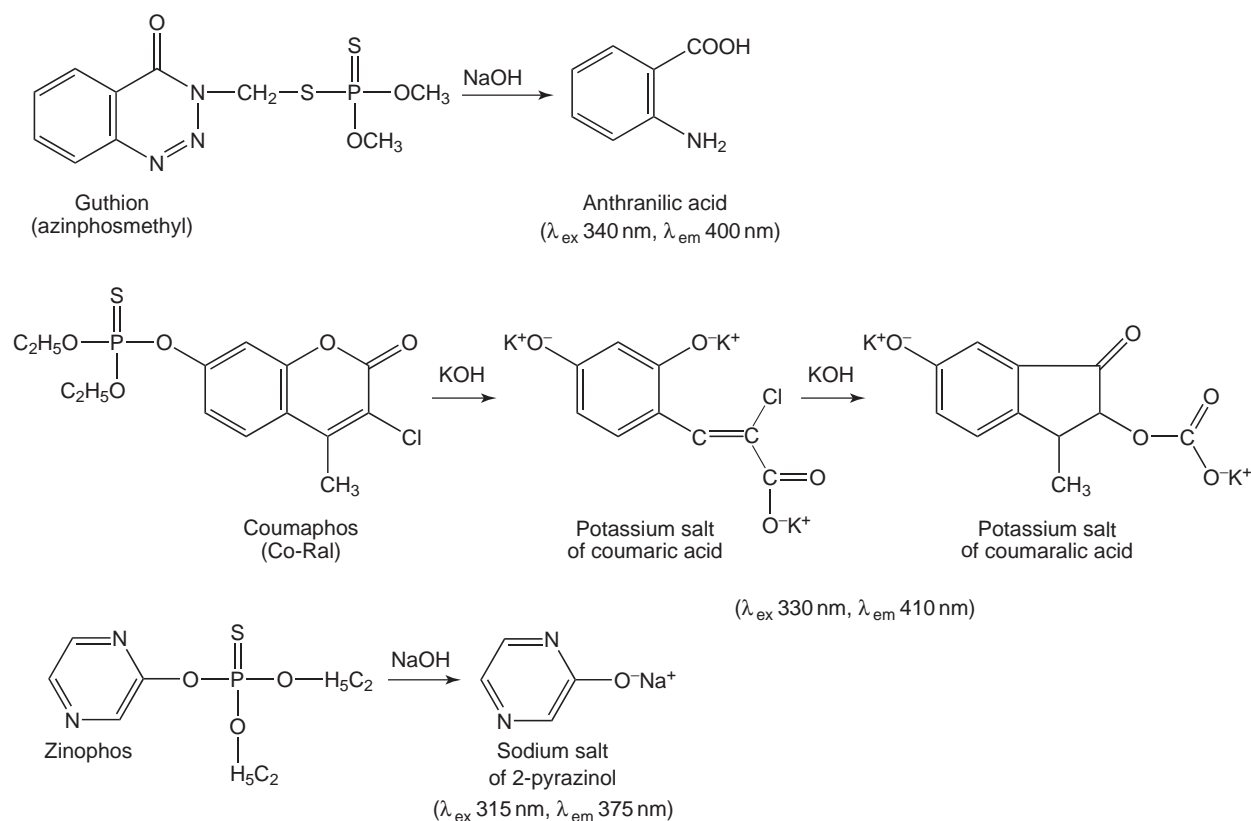
Analyte	LOD (ng ml <sup>-1</sup> )	Sample	Recovery (%)
Aldicarb	10	Potatoes	91–93
Bufencarb	10	Potatoes	96–98
Carbaryl	10	Vegetables	85–90
	10	Vegetables	82–93
Carbofuran	–	Crops	83–97
	10	Grapes	98
	10	Potatoes	97–103
Methiocarb	10	Vegetables	78–85
		Grapes	97–99
Methomyl	–	Crops	80–96
	10	Potatoes	92–96
Oxamyl	–	Crops	77–99
	10	Grapes	88–92
	20	Potatoes	90–94

without the need for derivatization. Thus, the fungicides thiabendazole and carbendazim have been simultaneously determined using derivative synchronous spectroscopy in vegetable samples with no cleanup.

Although the use of FIAs for the determination of pesticide residues in food analysis has been relatively limited, there are some examples of their suitability in this analytical field, as Table 3 shows. Various fluorescence polarization immunoassay (FPIA) methods involving equilibrium or kinetic measurements have been described for the determination of several pesticides. A homogenous FIA has been developed for the determination of 2,4,5-T using Cresyl Violet as the label, which allows long-wavelength measurements to be made in order to avoid background fluorescence signals. This pesticide has also been determined using enzyme-linked immunosorbent assay (ELISA) using a fluorogenic substrate of horseradish peroxidase.

### Veterinary Drug Residues

Fluorescence spectroscopy is frequently used in the control of pharmaceutical residues in foods, such as



**Figure 2** Hydrolysis reaction of some pesticides giving fluorescent products.

**Table 3** Analytical performance of some FIA methods for pesticide determination

Analyte	Method	Label	LOD (ng ml <sup>-1</sup> )	Sample	Recovery (%)
2,4-D <sup>a</sup>	FPIA <sup>b</sup>	Fluorescein	4	White wine	105–110
	QFIA <sup>c</sup>	Fluorescein	3	Grape juice	92–104
				Orange juice	91.3–102
2,4,5-T <sup>d</sup>	EFIA <sup>e</sup>	Cresyl Violet	0.03	Grape juice	94.4–103
				White wine	
				Grape juice	94.4–103
Atrazine	F-ELISA <sup>f</sup>	HRP <sup>g</sup> + HPAA <sup>h</sup>	0.02	Apple juice	93.9–109.9
	FPIA	Fluorescein	0.2	White wine	80–100
				Red wine	80–99
				Orange juice	80–100
Dichlorprop	FPIA	Fluorescein	100	Tea	90–104
DME <sup>i</sup>	F-ELISA	HRP + homovanillic acid	0.3	Apple	117–125
				Apple	96.7–103.5

<sup>a</sup>2,4-Dichlorophenoxyacetic acid.<sup>b</sup>Fluorescence polarization immunoassay.<sup>c</sup>Quenching FIA.<sup>d</sup>2,4,5-Trichlorophenoxyacetic acid.<sup>e</sup>Enhancement FIA.<sup>f</sup>Fluorescence-enzyme-linked immunosorbent assay.<sup>g</sup>Horseradish peroxidase.<sup>h</sup>3-(*p*-Hydroxyphenyl)acetic acid.<sup>i</sup>2-(2,4-Dichlorophenoxy)propionic methyl ester.**Table 4** Some examples of derivatizing reagents used in HPLC methods for antibiotic determination in milk samples

Antibiotic	Reagent
Gentamicin	OPA + mercaptoacetic acid
Streptomycin + dihydrostreptomycin	1,2-Naphtoquinone-4-sulfonate
Ampicillin	Salicylaldehyde
Amoxicillin	Salicylaldehyde
Ampicillin	Formaldehyde + trichloroacetic acid
Tetracycline	Zirconium ion

antibiotics, as they are usually used for veterinary purposes and, sometimes, as animal feed additives, so that there is a risk that residues of these drugs may be present in edible tissues. Numerous methods for the determination of tetracyclines, aminoglycosides,  $\beta$ -lactams, and fluoroquinolones have been described using LC with fluorimetric detection. **Table 4** shows some derivative reagents used for this purpose. In the case of fluoroquinolones, they are detected by measuring their native fluorescence.

Although the application of CE-LIF to this analytical field is still limited, there are some examples, such as the determination of tetracycline, oxytetracycline, and chlortetracycline in milk samples based on the formation of the corresponding fluorescent chelates with magnesium ions.

## Microbial Contamination

Fluorimetry has been used to check for microbial contamination in foods such as eggs and dairy products. This technique is of use in detecting organisms that produce fluorescent pigments (e.g., *Pseudomonas fluorescens*). For example, pyoverdine is a brilliantly green fluorescent pigment present in *Pseudomonas*-infected eggs, the fluorimetric estimation of which is fairly simple, although the simultaneous occurrence of riboflavin calls for the use of blanks. The particular fluorescent pigments encountered can help us to identify the type of microbial contamination present, and their concentration can provide an index of bacterial counts in contaminated samples.

## Food Additives

Many food additives can be determined fluorimetrically, although few are actually assayed routinely using this technique. One of the best examples in this context is the determination of quinine in tonic water and lemonade based on the measurement of its intrinsic fluorescence in sulfuric acid. Although the simultaneous determination of food additives usually involves a chromatographic separation, an alternative method for the determination of two antioxidants, propyl gallate (PG) and butylated hydroxyanisole (BHA), in various food samples has been described using a T-format luminescence



spectrometer and kinetic methodology. Two independent reactions have been used: the formation of the complex terbium(III)–PG and the reaction of BHA with the oxidized form of Nile Blue. The absence of overlap in the emission spectra of the two systems has made it possible to measure the additives separately in each channel of the instrument, avoiding the use of a chromatographic separation.

Gelling agents based on polysaccharides are important food additives, especially used in freeze-dried products to control the consistency of the reconstituted food. CE has been utilized to investigate the monosaccharide composition of hydrolyzed gelling agents in food samples, using 2-aminoanthracene as the derivative reagent. CE has also been applied to the separation of carrageenans, which has been derivatized with 9-aminopyrene-1,4,6-trisulfonic acid and detected using LIF.

### Fluorescence Spectroscopy in Adulteration and Quality Control of Foods

The high sensitivity and simplicity of fluorescence spectroscopy justify the common use of this technique in these fields. It has been widely applied to characterizing citric juices and investigating their alteration, such as the adulteration of lemon oil with grapefruit oil. While grapefruit oil contains an ether of umbelliferone, which is released on acid hydrolysis, lemon oil contains other nonfluorescent hydroxycoumarins. The adulteration of saffron, which is relatively frequent because of its high price, can be detected using TLC with fluorimetric detection. Saffron can be differentiated from some substitutes because it is not fluorescent under UV radiation.

Although edible oils and fats are usually assayed using GC, spectrofluorimetry has been used for straightforward screening of their quality. Thus, vegetable oils exhibit a characteristic fluorescence due to the presence of carotenoids, chlorophyll, and PAH residues. Numerous methods have been developed to distinguish olive oils of different quality. Also, the presence of glycerin, which is an indication of rancidity since it is released during enzymatic breakdown of fat, can be detected fluorimetrically by condensation with  $\beta$ -naphthol, 2,7-dihydroxynaphthalene, and anthrone.

Fluorescence spectroscopy has been used routinely to detect decay and spoilage in foods, the occurrence of which is occasionally signaled by the appearance of fluorescence. Thus, orange rot appears as light

fluorescence on a dark background of the fruit. Other fruits possess a characteristic fluorescence that is either altered or removed by spoilage. Fresh milk and cream emit greenish-yellow fluorescence that is mainly due to riboflavin and changes to blue fluorescence, typical of lumichrome, on exposure to daylight for long periods of time or on irradiation by other means. Detection of the production of biogenic amines, such as cadaverine and agmatine, in fish using LC and postcolumn derivatization with OPA has been proposed as a freshness indicator of the stored fish. An example of the application of FIAs in food quality control is the determination of skatole, a compound causing boar-taint in pig back fat samples. The method has been developed in a homogeneous format and is based on the use of europium(III) as a label together with kinetic methodology.

**See also:** **Amino Acids.** Electrophoresis: Proteins. **Food and Nutritional Analysis:** Overview; Antioxidants and Preservatives. **Immunoassays, Applications:** Food. **Immunoassays, Techniques:** Luminescence Immunoassays. **Kinetic Methods:** Principles and Instrumentation. **Liquid Chromatography:** Food Applications. **Proteins:** Foods. **Vitamins:** Overview.

### Further Reading

- Frazier RA (2001) Recent advances in capillary electrophoresis methods for food analysis. *Electrophoresis* 22: 4197–4206.
- Gilbert J and Anklaam E (2002) Validation of analytical methods for determining mycotoxins in foodstuffs. *Trends in Analytical Chemistry* 21: 468–486.
- Lynch PLM and Young IS (2000) Determination of thiamine by high-performance liquid chromatography. *Journal of Chromatography A* 881: 267–284.
- Moret S and Conte LS (2000) Polycyclic aromatic hydrocarbons in edible fats and oils: Occurrence and analytical methods. *Journal of Chromatography A* 882: 245–253.
- Schenck FJ and Callery PS (1998) Chromatographic methods of analysis of antibiotics in milk. *Journal of Chromatography A* 812: 99–109.
- Schobel U, Barzen C, and Gauglitz G (2000) Immunoanalytical techniques for pesticide monitoring based on fluorescence detection. *Fresenius Journal of Analytical Chemistry* 366: 646–658.
- Wallace JM and Fox PF (1998) Rapid spectrophotometric and fluorimetric methods for monitoring nitrogenous (proteinaceous) compounds in cheese and cheese fractions: A review. *Food Chemistry* 62: 217–224.



# FOAM FRACTIONATION AND FLOTATION

M Hiraide, Nagoya University, Nagoya, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Foam fractionation and flotation are the terms given to the separation processes by which inorganic and organic ions, molecules, colloidal particles, and suspended solids are floated from aqueous solutions with a rising stream of bubbles. Flotation is sometimes used to mean the removal of particulate materials by bubbling, whereas foam fractionation indicates the foaming off of dissolved materials by adsorption on bubble surfaces. Overlap and inconsistency in terminology are often encountered.

In general, the substance to be floated should be hydrophobic. When the substance is hydrophilic, surfactants are added to render it (or its surface) hydrophobic before flotation.

The best-known bubble separation technique is probably ore flotation, which is used for the recovery of valuable minerals from suspensions of crushed ores. For example, sulfide particles are selectively rendered hydrophobic with xanthate ions and floated by bubbling, leaving oxides and silicates in the solution. Another well-known application is the separation of surface-active materials such as surfactants and proteins by simple bubbling.

In 1959, Sebba proposed the technique of ion flotation, in which the desired metal ions or their complexes were allowed to react with surfactant ions and the compounds were floated by bubbling. Selective separation was achieved by using appropriate surfactants and complexing agents, and by adjusting their concentrations and flotation pHs. The first analytical application was reported by Tomlinson and Sebba. An excess of cationic dyestuff (crystal violet) was added to a solution containing milligram quantities of oleate ions, and the resulting reaction product was floated with bubbles. The oleate ions in the original solution were determined by measuring the decrease of absorbance of the dye in the solution. In 1968, Mizuike *et al.* tried to apply the bubble separation technique to the separation and preconcentration of trace heavy metals from matrix elements.

Flotation mechanisms and experimental conditions have since been studied in detail, and the flotation apparatus has been greatly improved. Today, the bubble separation technique is an

indispensable method, especially for the separation and preconcentration of trace constituents in environmental, industrial, and biological samples.

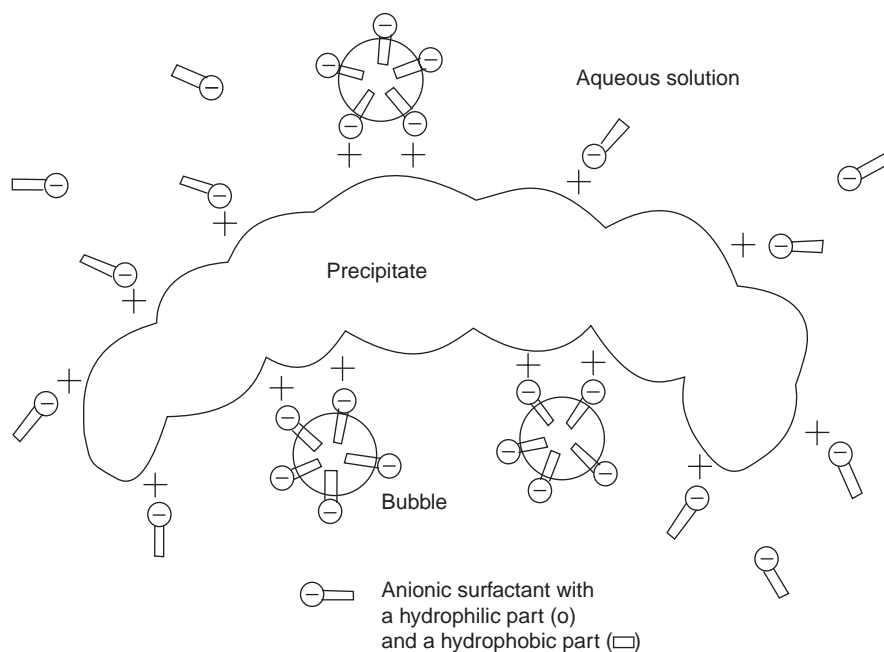
## Principles

Bubble separation techniques can be grouped into three classes: precipitate flotation, colloid flotation, and ion flotation (involving foam fractionation). Precipitate flotation is the most powerful of these separation techniques and is popular in the field of analytical science. Therefore, the principles and important experimental factors in bubble separation techniques are explained with reference to precipitate flotation.

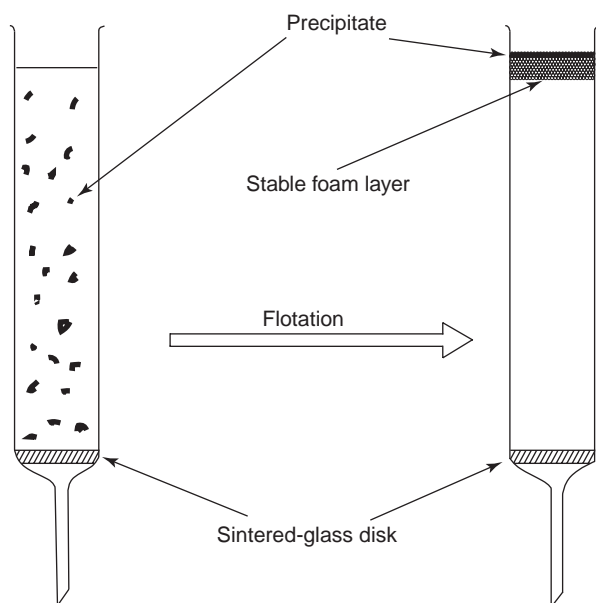
### Precipitate Flotation

Figure 1 is a schematic representation of precipitate flotation with small bubbles and surfactant ions having opposite charge to the precipitate surface. The hydrophilic groups of surfactant ions are oriented toward the surface of the precipitate, and hence the precipitate is made effectively hydrophobic. Small bubbles are easily trapped on the surface of the precipitate and flotation is achieved. An excess of surfactant serves to form a stable foam layer that prevents the redispersion of the floated precipitate into the bulk solution (see Figure 2).

The flotation behavior of hydrated iron(III) oxide is shown in Figure 3. The charge of the precipitate surface changes from positive to negative with increasing pH of the solution. Therefore, flotation is achieved with sodium oleate (anionic surfactant) up to the isoelectric pH, and with benzalkonium chloride (cationic surfactant) at higher pHs. In the absence of ethanol, flotation recovery is not complete because large bubbles (1–1.5 mm in diameter) ascend rapidly and disperse the floated precipitate. By adding small amounts (~1% v/v) of ethanol, however, numerous tiny bubbles (0.1–0.5 mm in diameter) are generated and successful flotation (complete flotation followed by formation of a stable foam layer) is achieved. Both anionic and cationic surfactants are useful for flotation around the isoelectric pH, because tiny bubbles are easily trapped in the interstitial spaces of the flocculent hydrated iron(III) oxide precipitate. Ethanol, which is often used as solvent for a surfactant, prevents the coalescence of bubbles appearing from adjacent pores of the sintered-glass disk (porosity 4) at the bottom of a flotation cell.



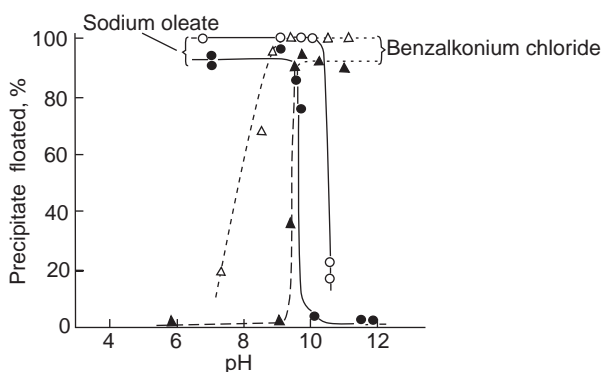
**Figure 1** Flotation of precipitate.



**Figure 2** Before and after flotation of precipitate.

Similar results are obtained with other water-miscible organic solvents (e.g., methanol, acetone, and methyl cellosolve).

The combined use of two kinds of surfactants is sometimes recommended. Sodium oleate plus sodium dodecyl sulfate (SDS) is one of the best combinations for flotation from seawater samples. Sodium oleate functions in the flotation of the precipitate and



**Figure 3** Flotation of iron(III) hydroxide in the presence (○, Δ) or absence (●, ▲) of ethanol.

SDS in the formation of a stable supporting foam layer. With only sodium oleate, a stable foam layer is not formed on the solution surface, probably due to the reaction of an excess of sodium oleate with magnesium ions in seawater. With only SDS, most of the precipitate does not float.

The following precipitates are successfully floated from aqueous solutions with suitable surfactants and ethanol:  $\text{Mg}(\text{OH})_2$ ,  $\text{Al}(\text{OH})_3$ ,  $\text{Ti}(\text{OH})_4$ ,  $\text{Cr}(\text{OH})_3$ ,  $\text{Fe}(\text{OH})_3$ ,  $\text{Fe}(\text{OH})_2$ ,  $\text{Co}(\text{OH})_2$ ,  $\text{Ni}(\text{OH})_2$ ,  $\text{Cu}(\text{OH})_2$ ,  $\text{Zn}(\text{OH})_2$ ,  $\text{Zr}(\text{OH})_4$ ,  $\text{In}(\text{OH})_3$ ,  $\text{Sn}(\text{OH})_4$ ,  $\text{Sb}(\text{OH})_3$ ,  $\text{Bi}(\text{OH})_3$ ,  $\text{Th}(\text{OH})_4$ ,  $\text{CdS}$ , and  $\text{PbS}$ .

Flotation without a surfactant is possible in the separation of organic precipitates. A water-insoluble organic precipitant is first dissolved in a

water-miscible organic solvent, and then the organic solution is added to an aqueous sample solution. After stirring for several minutes, flocculent precipitates are formed; these are nearly completely floated by bubbling because tiny bubbles are easily trapped in their interstitial spaces. The resulting precipitate layer is relatively stable. Flotation with surfactants is, of course, possible. However, after the collection of precipitates, the foam mixed with the precipitates cannot be destroyed by adding organic solvents (e.g., ethanol and diethyl ether), because the solvents also dissolve the precipitates. The following organic precipitates are successfully floated without any surfactant: *p*-dimethylaminobenzylidenerhodamine, dithizone, thionalide, 2-mercaptobenzothiazole, 2-mercaptobenzimidazole,  $\alpha$ -benzoin oxime, and 1-nitroso-2-naphthol.

Precipitate flotation is more rapid and convenient than conventional filtration and centrifugation. Furthermore, precipitate flotation needs less experimental skill and is more reliable and rapid, compared with colloid and ion flotation.

### Colloid Flotation

Colloidal particles are floated by adding surfactant ions of opposite charge to the colloids. During flotation, a scum is generally formed on the solution surface, which is then supported by a stable foam layer. Compared with precipitate flotation, the flotation efficiency is greatly affected by the experimental conditions (e.g., pH, quantity of surfactants, and bubbling time). Therefore, colloidal particles are sometimes coagulated to form bulky precipitates and then floated rapidly. For example, sulfate ions are useful to coagulate hydrated iron(III) oxide colloids at pH 4–5.

### Ion Flotation

Ion flotation originally meant the process by which surface-inactive ions are floated with surfactants of opposite charge to the ions to form a scum on the solution surface. This concept has been extended to include foam fractionation, where dissolved substances (e.g., ions, organic and inorganic molecules) are floated by adsorption on bubble surfaces and copious foam is generated on the solution surface.

The optimum operating conditions for ion flotation vary widely from system to system. The flotation pH, bubbling time, and gas flow rate should be carefully optimized, because in most cases it is difficult to observe when the flotation is complete. The quantities of surfactants should be greater than the stoichiometric amounts, but large excesses may decrease the flotation efficiency. An increase in the quantity of

coexisting ions generally decreases the flotation recovery, probably due to competition between the desired ions and coexisting ions for surfactants.

Ion flotation, in which the desired ions are floated and immediately collected in a water-immiscible organic solvent placed over the sample solution, is also useful. This method is often called solvent sublation.

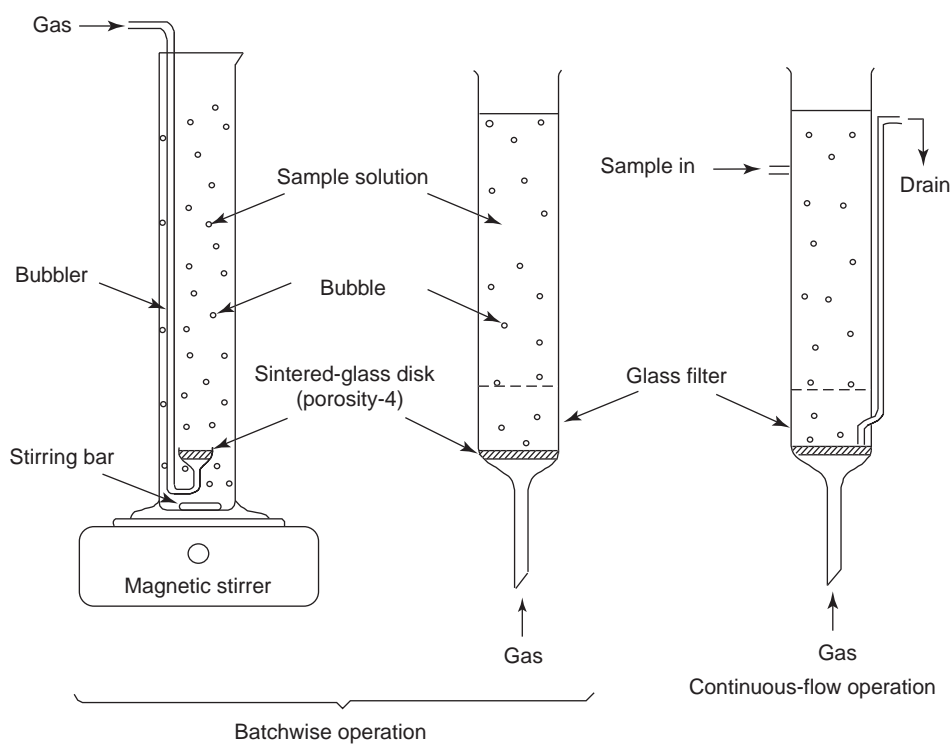
## Instrumentation

Figure 4 shows typical flotation cells for batchwise and continuous-flow operations. Flotation cells are easily prepared by inserting a bubbler into a graduated cylinder or by connecting a long glass tube to a glass filter (porosity 4 commonly used, nominal pore size 5–10  $\mu\text{m}$ ). A rising stream of tiny bubbles is produced by passing nitrogen or air through the sintered-glass disk. A magnetic stirrer is used to ensure that the bubbles are attached to the precipitate that forms under the bubbler. For continuous-flow operation, the dimensions of the flotation cell should be carefully optimized. When the diameter of the cell is reduced, the sample stream disturbs the flotation. The height of the sample inlet should be determined in order that both redispersion of floated substances and loss through a drain do not occur. Electrolysis can be used for the flotation, where tiny bubbles of oxygen and hydrogen are produced at the anode and cathode installed in a flotation cell.

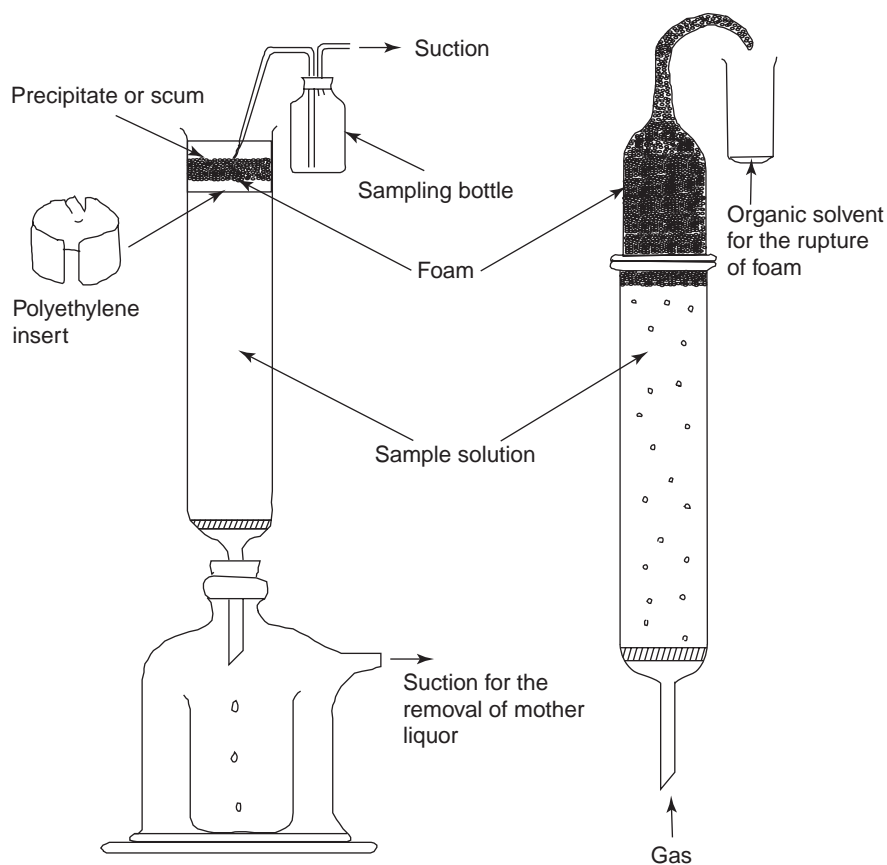
Some useful techniques for the collection of flotation products (precipitate, scum, and foam) are shown in Figure 5. A sampling bottle, spatula, and pipette are useful for direct collection. A detachable polyethylene insert is helpful for the complete recovery of flotation products, which often adhere strongly to the inner wall of the flotation cell. Suction through the sintered-glass disk is also convenient for the separation of precipitate and scum from the mother liquor. Copious foams are ruptured by contact with small amounts of organic solvents (e.g., ethanol, 1-butanol, and diethyl ether).

## Applications

Precipitate flotation has mainly been used for the separation of collector precipitates for trace analysis. Trace elements at the low microgram per milliliter level or less are nearly completely coprecipitated with small amounts of inorganic or organic collector precipitates, which are then floated by bubbling with or without the aid of surfactant ions. This method is



**Figure 4** Typical flotation cells.



**Figure 5** Collection of precipitate, scum, and foam.

often called coprecipitation flotation. Some typical examples are described below.

### **Simultaneous Preconcentration of Heavy Metals in Natural Water**

Trace heavy metals in 1 l of sample can be coprecipitated with indium hydroxide at pH 9–9.5 by using 1 ml of indium nitrate solution ( $100 \text{ mg In ml}^{-1}$ ) and  $0.3 \text{ mol l}^{-1}$  sodium hydroxide solution. After adding 2 ml of sodium oleate solution ( $1 \text{ mg ml}^{-1}$  in 70% ethanol) and 1 ml of SDS solution ( $4 \text{ mg ml}^{-1}$  in 70% ethanol), the precipitate is floated with tiny bubbles and collected in a sampling bottle. The precipitate is dissolved in 2 ml of  $14 \text{ mol l}^{-1}$  nitric acid and 3 ml of water, and analyzed by inductively coupled plasma atomic emission spectrometry for chromium(III), manganese(II), cobalt(II), nickel, copper(II), cadmium, and lead.

### **Preconcentration of Silver in Seawater**

Traces of silver can be concentrated by adding 50 ml of 2-mercaptobenzothiazole solution ( $40 \text{ mg ml}^{-1}$  in acetone, purified by the amalgamation method) to 3 l of sample (acidified to pH 1). By stirring the solution, the precipitate is coagulated and then floated to the solution surface. After the suction of the mother liquor through the sintered-glass disk, the precipitate is destroyed with  $14 \text{ mol l}^{-1}$  nitric acid and 30% hydrogen peroxide. The resulting residue is dissolved in 5 ml of  $8 \text{ mol l}^{-1}$  nitric acid and analyzed by atomic absorption spectrometry for silver.

### **Radiochemical Separation of Cobalt-60 in Seawater**

To a 100 l sample, 500 ml of  $12 \text{ mol l}^{-1}$  hydrochloric acid, 10 ml of cobalt carrier solution ( $1 \text{ mg Co ml}^{-1}$ ), and 400 ml of 1-nitroso-2-naphthol solution ( $5 \text{ mg ml}^{-1}$  in ethanol) are added. With peristaltic pumps, the sample solution,  $1 \text{ mol l}^{-1}$  aqueous ammonia, and anionic surfactant solution ( $0.8 \text{ mg ml}^{-1}$  in 70% ethanol, sodium oleate–SDS = 1:3) are fed into a 1 l reaction bottle at flow rates of 500, 35, and  $10 \text{ ml min}^{-1}$ , respectively. In the reaction bottle, the cobalt chelate is captured by a flocculent precipitate of indium hydroxide at pH 8.5–9. The contents of the bottle are continuously transferred to a flotation cell and flotation continues until the completion of sample feeding. The precipitate is sucked into a sampling bottle and then dissolved in 100 ml of  $6 \text{ mol l}^{-1}$  hydrochloric acid. The cobalt chelate is separated from indium ions by extraction with chloroform and destroyed by dry oxidation. After dissolving the residue in 5 ml of  $8 \text{ mol l}^{-1}$  hydrochloric

acid, the cobalt is extracted into 5 ml of tri-*n*-octylamine ( $50 \text{ mg ml}^{-1}$  in xylene) and the  $\beta$ -activity of cobalt-60 is counted with a liquid scintillation counter.

Some examples of applications of precipitate flotation, using both inorganic and organic collector precipitates, are summarized in Tables 1 and 2. Organic collector precipitates are generally much more selective in the separation of metal ions than inorganic collector precipitates.

Based on the principles of precipitate flotation, a rapid and convenient separation technique has been developed for the determination of toxic heavy metals adsorbed on suspended solids in freshwater. Because suspended solids are negatively charged species, they are rendered hydrophobic and coagulate to form bulky flocs with a cationic surfactant and sodium chloride (to increase the ionic strength). The flocs are easily floated by bubbling and are then treated in nitric acid to determine the desorbed heavy metals (e.g., chromium, manganese, copper, cadmium, and lead) by graphite furnace atomic absorption spectrometry.

Colloid flotation can also be used for separation and preconcentration in trace analysis, by adsorbing trace elements on colloidal particles. This method is known as adsorbing colloid flotation. In most cases, however, the collectors used are visible precipitates (not colloidal particles). Therefore, these examples should properly be included in precipitate flotation.

Ion flotation has the advantage that the desired ions are separated without collector precipitates. Three typical examples of ion flotation are described below.

### **Preconcentration of Uranium(VI) in Seawater**

To a sample of 500 ml, 50 ml of arsenazo-III solution ( $1 \text{ mg ml}^{-1}$ ) and 50 ml of  $1 \text{ mol l}^{-1}$  acetate buffer solution (pH 3.5) are added. The resulting uranium(VI)–arsenazo-III anion complex is floated with 4 ml of tetradecyldimethylbenzylammonium chloride solution ( $10 \text{ mg ml}^{-1}$ ) and a scum is formed on the solution surface. After the suction of the mother liquor through the sintered-glass disk, the scum is ignited and then dissolved in 5 ml of  $9 \text{ mol l}^{-1}$  hydrochloric acid for the spectrophotometric determination of uranium.

### **Preconcentration of Heavy Metals in Hot Spring Water**

To a sample of 1 l, 2 ml of  $9 \text{ mol l}^{-1}$  sulfuric acid, 5 ml of hydroxylammonium sulfate solution ( $100 \text{ mg ml}^{-1}$ ), and 3 ml of  $0.01 \text{ mol l}^{-1}$  1,10-phenanthroline solution are added. The pH of the solution is adjusted to 4 with  $6 \text{ mol l}^{-1}$  aqueous

**Table 1** Flotation of inorganic collector precipitates for the separation of trace constituents

<i>Samples</i>	<i>Trace constituents</i>	<i>Collector precipitates</i>	<i>pH</i>	<i>Surfactants</i>	<i>Determination<sup>a</sup></i>
Natural water	Cr(III), Mn(II), Fe(III), Co(II), Ni, Cu(II), Zn, Cd, Pd	Indium hydroxide or aluminum hydroxide	9–9.5	Sodium oleate + SDS	ICP-AES or AAS
Natural water	P(V), As(V), Se(IV), Mo(VI), Sn(II, IV), Sb(III, V), Bi	Iron(III) hydroxide	4	Sodium oleate + SDS	AAS or SP
Natural water	Te(IV)	Preformed hydrated iron(III) oxide	8–9	Sodium oleate + SDS	AAS
Seawater	Cobalt-60 <sup>b</sup>	Indium hydroxide	8.5–9	Sodium oleate + SDS	LSC
Seawater	Hg(II)	Cadmium sulfide	1	Octadecyltrimethylammonium chloride	AAS
Seawater	U(VI)	Preformed hydrated titanium(IV) oxide	6.5	SDS	SP
Freshwater	P(V)	Aluminum hydroxide	8.5	Sodium oleate	SP
Freshwater	Cd, Sn(IV), Sb(III), Bi	Zirconium(IV) hydroxide	9	Sodium oleate	AAS
Freshwater	Humic acid	Iron(III) hydroxide	7	Sodium oleate + SDS	SP
Wastewater	Sn(II), W(VI)	Zirconium(IV) hydroxide	8	Sodium oleate	PG or SP
Water	Phorate <sup>c</sup>	Iron(III) hydroxide	6	SDS	GC
Zinc nitrate solution	Fe(III), Pb	Bismuth hydroxide	Ammoniacal solution	Sodium oleate	AAS

<sup>a</sup>ICP-AES, inductively coupled plasma-atomic emission spectrometry; AAS, atomic absorption spectrometry (including graphite furnace atomic absorption spectrometry); SP, spectrophotometry; LSC, liquid scintillation counting; PG, polarography; GC, gas chromatography.

<sup>b</sup>In the presence of cobalt carrier and 1-nitroso-2-naphthol.

<sup>c</sup>Phosphorodithioic acid *O,O*-diethyl *S*-[(ethylthio) methyl] ester.

**Table 2** Flotation of organic collector precipitates for the separation of trace constituents

<i>Samples</i>	<i>Trace constituents</i>	<i>Collector precipitates</i>	<i>pH</i>	<i>Determination<sup>a</sup></i>
Natural water	Ni, Cu(II), Cd, Pb	Indium pyrrolidinedithiocarbamate	2	ICP-AES
Seawater	Ag	2-Mercaptobenzothiazole	1	AAS
Copper(II) nitrate solution	Ag	<i>p</i> -Dimethylaminobenzylidenerhodanine	1	SP
Lead nitrate solution	Cu(II), Ag	Dithizone	1	AAS
Zinc nitrate solution	Co(II)	1-Nitroso-2-naphthol	3	AAS

<sup>a</sup>For explanation of abbreviations, see **Table 1**.

ammonia to form anionic chelates of heavy metals (e.g., copper(II), cobalt(II), zinc, and cadmium). Nitrogen is passed through the sintered-glass disk, while SDS solution (10 mg ml<sup>-1</sup>) is continuously added

from the lower part of the flotation cell. The resulting copious foam is ruptured with 1-propanol and diluted to 10 ml with water for analysis by atomic absorption spectrometry.



**Table 3** Ion flotation for the separation of trace constituents

Samples	Trace constituents	Complexing agents and surfactants	pH	Determination <sup>a</sup>	Flotation products
Seawater	U(VI)	Arsenazo-III + tetradecyldimethylbenzylammonium chloride	3.5	SP	Scum
Seawater	Fe(III) <sup>b</sup>	3-(2-Pyridyl)-5,6-diphenyl-1,2,4-triazine + sodium dodecylbenzene sulfonate	3	SP	(Solvent sublation)
Fresh water	Co(II), Cu(II)	Nitroso-R-salt + cetyltrimethylammonium bromide	5	AAS	Scum
Fresh water	SCN <sup>-</sup>	Hexadecyltrimethylammonium chloride + carboxymethyl cellulose	5–8	SP	Scum
Fresh water	Cr(VI)	Diphenylcarbazine + SDS	1	SP	Copious foam
Fresh water	Cr(VI)	Cetyldimethylammonium bromide	5.5	SP	Copious foam
Fresh water	Cu(II), methyl mercury(II) chloride	Potassium <i>n</i> -butyl xanthate + cetyltrimethylammonium bromide	9	AAS or GC	Copious foam
Fresh water	Anionic surfactants	Ethyl violet	5	SP	(Solvent sublation)
Drinking water	Ni	1-(2-Pyridylazo)-2-naphthol (PAN)	5	SP	(Solvent sublation)
Hot spring water	Fe(III), Co(II), Ni, Cu(II), Zn, Cd	1,10-Phenanthroline + SDS	4	AAS	Copious foam
Water	<i>tert</i> -Butylphenol	Cetylpyridinium chloride		SP	Copious foam
Water	Pentachlorophenol	Hexadecyltrimethylammonium bromide	8.9	SP	(Solvent sublation)
Sodium iodide solution	Cd	Ethylhexadecyldimethylammonium bromide	2–10	AAS	Scum
Serum	Cu(II) <sup>b</sup>	Bathocuproine + SDS	7	SP	(Solvent sublation)
Serum	Zn	2-(5-Bromo-2-pyridylazo)-5-( <i>N</i> -propyl- <i>N</i> -sulfoethylamino) phenol + tetrahexylammonium chloride	9	SP	(Solvent sublation)

<sup>a</sup> For other abbreviations, see **Table 1**.<sup>b</sup> Reduced by adding hydroxylammonium sulfate.

### Preconcentration of Anionic Surfactant Ions in Freshwater

To a sample of 200 ml, acetate buffer solution (pH 5) and 0.5 ml of ethyl violet solution (1 mg ml<sup>-1</sup>) are added. The resulting ion pairs of anionic surfactants are floated by bubbling and captured in 5 ml of organic solvent (toluene–methyl isobutyl ketone = 9:1). The organic phase is analyzed by spectrophotometry for anionic surfactants (mainly SDS).

Some examples of the application of ion flotation are summarized in **Table 3**. The recovery of valuable proteins and pharmaceutically important compounds by using foam fractionation and ion flotation is becoming increasingly important because of their economical and simple procedures.

See also: **Environmental Analysis**. **Flow Analysis**: Overview. **Radiochemical Methods**: Overview. **Surfactants and Detergents**. **Water Analysis**: Overview.

### Further Reading

- Caballero M, Cela R, and Perez-Bustamante JA (1990) Analytical applications of some flotation techniques. *Talanta* 37: 275–300.
- Hiraide M and Mizuike A (1982) Flotation techniques in analytical chemistry. *Reviews in Analytical Chemistry*. VI: 151–168.
- Lemlich R (ed.) (1972) *Adsorptive Bubble Separation Techniques*. New York: Academic Press.
- Mizuike A (1983) *Enrichment Techniques for Inorganic Trace Analysis*. Berlin: Springer.
- Mizuike A and Hiraide M (1982) Separation and preconcentration of trace substances: III. Flotation as a preconcentration technique. *Pure and Applied Chemistry* 54: 1555–1563.
- Sebba F (1962) *Ion Flotation*. Amsterdam: Elsevier.
- Smith JS, Valsaraj KT, and Thibodeaux LJ (1996) Bubble column reactors for wastewater treatment. *Industrial & Engineering Chemistry Research* 35: 1688–1699.

# FOOD AND NUTRITIONAL ANALYSIS

Contents

**Overview**

**Sample Preparation**

**Additives**

**Antioxidants and Preservatives**

**Contaminants**

**Water and Minerals**

**Pesticide Residues**

**Mycotoxins**

**Soft Drinks**

**Coffee, Cocoa, and Tea**

**Alcoholic Beverages**

**Wine**

**Meat and Meat Products**

**Dairy Products**

**Vegetables and Legumes**

**Oils and Fats**

**Fruits and Fruit Products**

**Packaging Materials**

## Overview

**V Jain and K Gupta**, Haryana Agricultural University, Hisar, India

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The science of food and nutritional analysis has developed rapidly in recent years. Food scientists analyze foods to obtain information about their composition, appearance, texture, flavor, shelf life, etc., and also to guarantee the quality of the product. Nevertheless, the term food and nutritional analysis is often thought only to be concerned with the determination of food composition and its nutritive value/quality.

The analysis of food began in the nineteenth century utilizing microscopy. It was the first analytical technique used by analysts such as Accum and Hassall to identify food components and to detect adulteration. The development of improved analytical methods to determine the composition of foods (such as Kjeldahl method

for nitrogen estimation in 1885), together with the concern over adulteration resulted in the introduction of statutory legislation over the composition of important food products in many countries. These forms of controls continued for ~100 years until the 1970s, when rapid advances in food science and technology revolutionized the manufacture of food. Media attentions have also brought a public awareness of the links and causation between diet and health. As a result, today's sophisticated consumers have become more concerned over the quality and compositions of their food purchases, the contained ingredients, and the presence of additives and contaminants. Therefore, knowledge of the chemical and biochemical composition of foods is important to the health, well-being, and safety of the consumers. Analytical characterization is also important for compliance with legal standards, quality assurance, and determination of nutritional value.

Various types of modern food analytical techniques have been developed, including electrophoresis, chromatography, spectroscopy, rheological techniques, and sensory evaluation, to meet the challenge of providing information on the diverse components of these complex food materials.

## Information Sources

### Standard Official Methods and Modern Information Retrieval Systems

Standard analytical methods approved by professional associations/scientific agencies such as Association of Official Analytical Chemists, The American Association of Cereal Chemists, and American Oil Chemist Society are used in food industry. (These methods are being improved and updated from time to time.) The computerized systems for storing and retrieving scientific information are now well developed (Table 1).

### Sampling

Food is a heterogeneous mixture of chemical substances; therefore, its analysis requires adequate methods of sampling and preserving. The analytical

method selected must be specific, accurate, precise, and sensitive for the substance to be analyzed. The sampling includes sample selection, preparation, and labeling. If it is not carried out properly/correctly it leads to variability of results. Ideally, the sample analyzed should have exactly the same properties as the bulk of the sample (population) from which it was taken. The major steps in sampling are identification of the population from which the sample is to be taken, location from which the sample should be collected, and method of obtaining the gross samples.

## Techniques and Methods

### Spectroscopic Techniques

A variety of spectroscopic techniques has been developed to analyze food materials (Table 2).

**Table 1** Organizations that issue standard food analysis methods and modern information retrieval systems

<i>Name of organizations/acronym</i>	<i>Area of analysis</i>
<i>Name of organizations</i>	
American Association of Cereal Chemists (AACC)	Cereals
American Oil Chemists Society (AOCS)	Oil seeds
Analytical Methods Committee of the Royal Society of Chemistry, UK (AMC)	Various
Association Francaise de Normalization (French Standard Organization) (AFNOR)	General
Association Internationale de l'Industrie des Bouillons et Potages (International Association of the Stock and Soup Industry) (AIIBP)	Soups
Association of Official Analytical Chemists, USA (AOAC)	Food, agriculture
Association of Public Analysts, UK (APA)	Various
British Standards Institution (BSI)	General
Corn Industries Research Foundation Inc. (CIRF)	Starch products
Deutsche Gesellschaft fur Fettwissenschaft (DGF)	Oils and fats
European Economic Community (EEC)	Various
Food and Agriculture Organization, UN (FAO)	Agriculture
Federation of Oils, Seeds and Fats Association (FOSFA)	Oils and fats
International Association of Seed Crushers (IASC)	Vegetables oils and fats
International Association for Cereal Chemistry (ICC)	Cereals
International Commission on Microbiological Specification for Food (ICMSF)	Food (general)
International Commission for Uniform Methods of Sugar Analysis (ICUMSA)	Sugar
International Dairy Federation (IDF)	Dairy products
International Office of Cocoa, Chocolate and Sugar Confectionery (IOCCC)	Cocoa, confectionery
International Organization of the Flavour Industry (IOFI)	Flavor
International Organization for Standardization (ISO)	General
International Union of Pure and Applied Chemistry (IUPAC)	General
Institute of Brewing, UK (IOB)	Beer
Nordisk Metodik-kommittee for Livsmedel (NMKL)	Food
Nederlands Normalisatie-Instituut (NNI)	General
Office Internationale de la Vigne et du Vin (OIV)	Wine
<i>Modern information retrieval systems</i>	
<i>Database</i>	
AGRICOLA, BIOSIS PREVIEWS, CAB abstracts, CRIS, Food science and technology abstracts, Science citation index	
<i>Website</i>	
http://agrifor.ac.uk, www.altavista.com, www.biacore.com, www.bmn.com, www.rediff.com, www.cabi.com, www.google.com, www.yahoo.com, http://www.worldmedicus.com	

**Table 2** Spectroscopic methods for food analysis

<i>Type of spectroscopy</i>	<i>Information obtained</i>	<i>Nature of material required</i>
Atomic absorption and emission	Analysis of elements	Solution
Atomic fluorimetry	Analysis of elements	Solution
Circular dichroism	Conformation of proteins	Solution of optically active molecules
Dielectric	Moisture content	Native
Electron spin resonance	Physical state	Native material
Nuclear magnetic resonance	Physical state and chemical analysis	Native material or solid or liquid extract
Near-IR and mid-IR	Physical state and chemical structure	Native material depending on type of IR
Mass spectroscopy	Analysis of elements and structure determination	Usually solid extracts
Raman	Physical state	Molecules
X-ray	Structure composition	Native material
UV-visible absorption	Chemical quantitative analysis	Native

**Table 3** Techniques for biological evaluation of food

<i>Analysis using live system</i>			<i>Analysis using biochemical techniques</i>	
<i>Whole animal as analytical tool</i>	<i>Microbiological analysis</i>	<i>Tissue and cell culture analysis</i>	<i>Immunochemical techniques</i>	<i>Enzymatic techniques</i>
Assay on body	Turbidimetric	Cell culture	Agglutination	<i>Substrate assay methods</i>
Body weight changes	Gravimetric method	Organ or tissue culture	Immunodiffusion	Endpoint methods
Digestibility assays	Serial dilution assays		Radioimmunoassay	Kinetic methods
Analysis of food protein and energy (PER, NPR), and net energy assay	Acidimetric and metabolic product assay		Enzyme immunoassay	<i>Immobilized enzymes assay</i>
	Diffusion plate assay			Tubular enzyme reactions
<i>Changes in nutrient content of body</i>				Enzyme electrodes
Energy balance				Rapid analysis strips
Bv, NPU				<i>Immunoenzymatic methods</i>
				Enzyme-linked immunosorbent assay
				Enzyme multiplied immunoassay technique

### Rheological Techniques

Rheology is the study of flow of matter and deformation and these techniques are based on their stress and strain relationship and show behavior intermediate between that of solids and liquids. The rheological measurements of foodstuffs can be based on either empirical or fundamental methods. In the empirical test, the properties of a material are related to a simple system such as Newtonian fluids or Hookian solids. The Warner-Bratzler technique is an empirical test for evaluating the texture of food materials. Empirical tests are easy to perform as any convenient geometry of the sample can be used. The relationship measures the way in which rheological properties (viscosity, elastic modulus) vary under a

specific system of applied forces. These techniques provide a measure of baking quality of flours, thickening properties of biopolymers, texture, hardness, flakiness, consistency in food, tenderness of meat, and texture of fruits and vegetables. Rheological techniques have now become essential tools in food science and technology.

### Biological Techniques

Biological techniques encompass two methodological approaches: live systems and biochemical techniques. Live systems employ (1) whole animal, (2) microorganisms, and (3) cell and tissue culture methods as tools in food analysis. Biochemical techniques include uses of enzymes and immunochemical techniques (Table 3).

**Live systems as analytical tools** Different life systems have been utilized in assessing composition and nutritional quality of foods.

The whole animal techniques are used in the nutritional and toxicological evaluation of foods by providing a combined assessment of digestibility, metabolism of food, and nutritional evaluation of quality of foods. These approaches are aimed at determining the changes in the nutrient content in animal body in response to feeding a diet or meal containing that nutrient. The animal used in conducting food analysis assays include albino rats and mice, and guinea pigs, chicken, ducks, hamsters, gerbils, dogs, and monkeys.

The changes in nutrient content and energy value of foods required for animal body involves the measurements on animal body itself. These assays are based on body weight changes for analysis of food protein and energy and include protein efficiency ratio (PER), net protein ratio (NPR), and multilevel assays. The second approach to estimate change in the nutrient content of the body involves measurements of intake and output of nutrients from the body and assay includes nitrogen or energy balance, net protein utilization (NPU), biological value (BV), and net energy.

The animal assay for evaluating toxic constituents (natural or added) of food uses lethality as an index and determines the dose of that toxic chemical to kill/affect 50% of the test animal, referred to as LD<sub>50</sub> dose. The response criteria can be toxicity, carcinogenicity, mutagenicity, reproduction, and metabolism.

**Microbiological food assay** Microbiological food assays involve the use of microorganisms as a substitute for a higher animal. Since some of the nutritional requirements of microorganisms and experimental animals are similar, it is possible to use microbiological food assays to determine the substances that are essential constituents of living cells. These are based on the principle that in the presence of limiting amount of specific nutrients, the amount of growth is a function of concentration of this nutrient. These are used for determination of amino acids, vitamins, nucleic acids, heavy metals, growth factors, and the nutritional value of proteins and antibodies. The microorganisms used for the assay are bacteria, fungi, algae, yeast, and protozoa. The test organism selected is based on nonpathogenicity, sensitivity, and specificity to the nutrient to be assayed, its rapid and reproducible growth, and ease in making growth measurements. It can be assayed after adding the food sample extract to a liquid medium or gel medium, inoculated with the microbial culture followed by

growth stimulation, and then by turbidimetry measurement, gravimetry measurement, diffusion plate assay, and metabolic product assay.

**Cell and tissue culture methodology** This method uses cell, tissue, or organs for analytical purposes. These have been used in food toxicology and safety. Toxic chemicals like pesticides, food additives, and chemicals produced during processing and cooking of foods are tested for their toxicity, mutagenicity, and oncogenic properties using tissue culture techniques.

**Biochemical techniques** These techniques include immunochemical and enzymatic techniques.

Immunochemical techniques are based on specific interactions of antibodies with antigens. Since macromolecules are present in microorganisms, food, and agricultural products and are good antigens, the antibodies against these molecules can be obtained in response to the immunization with antigens, the substances foreign to animals. The antibodies produced in response to a particular antigen are capable of recognizing and binding due to the complementary binding site on the antibody with that of antigen. The methods employed in food analysis include hemagglutination, precipitation, immunodiffusion, immunofluorescence, immunoelectrophoresis, radioimmunoassays, and enzyme immunoassays. Enzymatic analysis means analysis with the aid of enzymes using specific enzymatic reactions. The enzymatic techniques are highly useful because of specificity and sensitivity for measuring the concentration of compounds. This reduces the time of analysis and avoids lengthy separations of the components. The immobilized enzyme systems with electrochemical detection methods are used for analytical determinations of food components that would be laborious or otherwise impossible. These techniques require small amounts of sample and do not require components to be extracted from their matrix. Enzymatic methods are used in the determination of structure of polymeric macromolecules, quality indices of foods, and nutritional and physical induced changes in human beings.

### Sensory Evaluation

Sensory evaluation is a scientific discipline used to evoke, measure, analyze, and interpret reaction to those characteristics of food material as they are perceived by the senses of sight, smell, taste, touch, and hearing (sound). The sensory attributes of quality of food are measured to determine consumer acceptance/preference in order to manufacture an



acceptable and affective product at maximum production economy. The sensory attributes include appearance (color, size, shape, and consistency of liquid and semisolid products), kinesthetic (texture, consistency, and viscosity), and flavor (taste and odor). The evaluation is done to determine quality criteria by which raw materials and finished products may be graded and classified. The sensory techniques help food scientists to determine the conformity of a food with established government or trade standards and food grades and product development while maintaining desirable sensory characteristics. The technique of sensory evaluation uses humans as data generators; hence, it is influenced by cultural, psychological, religious, social, and climatic factors, physical and educational status of an individual, availability, and nutritional knowledge. Highly trained groups of experts are employed for evaluation to minimize the effects of such factors. For complete sensory evaluation of food, two types of tests – analytical sensory tests and affective tests (like-dislike) are performed (Table 4).

The analytical sensory tests are confined to sensitivity, discriminatory, quantitative (intensity), and descriptive or qualitative analysis of the sensory attributes of food.

A sensitivity test is a measure of sensitivity to a human stimulus and is expressed as threshold, which is a fixed value at a given moment of time. The threshold may be absolute or differential, which is measured using various psychological methods. The threshold is determined by approaching and receding from standard stimulus by short steps and the threshold is that step where the judge's response shifts from one category to another. In the method of average error (or the method of adjustment), the judge adjusts the concentration of the comparison with the

standard stimulus. The frequency method records each comparison stimulus against a standard stimulus many times, thereby creating a difference response.

### **Discrimination Tests**

Food analysis is based on difference testing, the fundamental approach to sensory analysis of food. A simple difference test permits one of the two responses 'Yes, there is a difference' or 'No, there is a no difference'. In directional difference testing, a judge is asked which sample is more in a predefined characteristic. The predesignated standard must be similarly understood and used by all the judges. A large number of difference tests, single stimulus, paired comparison, paired difference, triangle, dual standard, multiple standard, and multiple pairs, are used to detect the sensory differences (such as sweetness, softness, color, etc.) between the two or more samples.

### **Quantitative Test (Intensity)**

This approach of sensory evaluation is based on description of scaling, i.e., ranking category (or internal) and magnitude estimation (or ratio). Ranking is used for grouping of the products based on their quality or order of preference. The scoring on category scale involves the use of a limited number of categories, designated in terms of numbers, letters, or points on a line. The magnitude estimation is a type of ratio scaling that measure the relationship between physical and sensory criteria. Any number is assigned to the first stimulus. A proportional number is assigned to reflect its strength in comparison to the first.

### **Descriptive Analysis**

In descriptive analysis small groups of highly trained judges with considerable experience with the commodity under study develop adjectives to characterize the qualitative properties of the product (attributes like appearance, aroma, texture, taste, etc., are analyzed). The data of sensory analysis experiments are subjected to statistical analysis to get reliable results. Instrumental methods are also used to correlate physicochemical measurements with sensory judgments (Table 5).

**Effective testing/acceptance testing** By acceptance testing we mean measuring liking or preference for a product. Preference is that expression of appeal of one product versus another that can be measured directly by comparing two or more products with each other. The nine-point hedonic scale method

**Table 4** Human response of sensory systems corresponding to selected physical properties of foods

<i>Physical property (stimulus)</i>	<i>Sensory system</i>	<i>Human sensation</i>
Density	Kinesthetic, haptic	Heavy, light
Moisture content	Haptic, thermal	Dry, wet, soggy
pH	Gustation, pain	Harsh, sharp, sour
Radiant energy	Visual	Appearance, color
Shear	Kinesthetic, haptic	Hard
Surface abrasion	Haptic, pain	Rough, prickly
Solubility	Gustation	Taste
Temperature	Thermal system	Cold, hot
Texture	Olfaction	Odor
Vapor pressure	Kinesthetic, haptic	Thin, thick
Vibration/pitch	Auditory	Sound, e.g., crisp, crunchy, sizzle



**Table 5** Analytical methods and consumer measurements of sensory attributes

<i>Analytical tests</i>				
<i>Stimulus concentration</i>				
<i>Sensitivity</i>		<i>Quantitative</i>		<i>Qualitative</i>
<i>Threshold</i>	<i>Discrimination</i>	<i>Scaling</i>	<i>Duration</i>	<i>Descriptive analysis</i>
Methods of limits	Paired comparison	Ordering or ranking	Time intensity	Texture profile
Methods of adjustments or average error	Paired difference	Category (internal)		Flavor profile
Frequency method	Duo-trio	Magnitude estimation (ratio)		Dilution profile
	Dual standard			Quantitative descriptive analysis
	Multiple standard			Other methods
	Triangle			
<i>Consumer tests</i>				
<i>Effective</i>				
<i>Acceptance</i>	<i>Preference</i>	<i>Hedonic test</i>		
Accept/reject what is available	Select one over the other	Degree of like and dislike		

occupies a niche for the sensory evaluation of product acceptance/preference. The measure of liking preference is a sensory evaluation model for measuring product acceptance and it represents the final phase of test resources with discrimination and descriptive analysis test. The sensory acceptance test is a cost-effective resource that has a major role in the development of successful products.

The scale was developed to access the degree of acceptability of food items, beverages, cosmetics, paper products, etc., by measuring the degree/magnitude of like/dislike.

### Nutritional Evaluation of Food Processing

Although processed foods are not considered to be as nourishing as fresh and unprocessed foods, there is an increasing demand for processed foods, which are available throughout the year. Food processing involves a wide variety of industrial processes with a correspondingly large variety of products. The optimum quality maintenance of the product attributes are compared in relation to the origin as 'fresh like' characteristics of preserved fresh products and 'just cooked like' attributes of preserved cooked products. New techniques such as modified atmosphere packing, edible and biodegradable packing materials, high-pressure short-duration treatment, biopreservation, high-voltage inputs, and a combination of different mild treatments help in preserving the quality attributes of the processed food.

## Food Components

Knowledge regarding the major constituents of food (e.g., carbohydrates, amino acids, proteins, lipids, dietary fiber, and nutritional and antinutritional components) is necessary so as to gauge the extent of any structural change that occurs during processing and storage of foods, which may affect the quality and safety of the food product.

### Major Components

**Moisture** Water content/moisture content is the most ubiquitous substance in nature, the largest single constituents of all living things and affects quality, value, and freshness of food and is of major concern in food, paper, and plastic industries. Moisture determination is a widely used fundamental analytical operation, which satisfies the technological, analytical, commercial, and regulatory necessities in the processing, testing, and storage of food products and is an index of economic value, stability, and nutritional quality of food products.

Removal of water for processing/storage purposes either by conventional dehydration or freezing and drying alters the native functional properties of foods. Simple, rapid, and accurate methods for moisture determination in raw, processed, and stored food products are used to know the nutritive value of food products. A homogenous food sample should be prepared using a number of electrical/mechanical

devices like blenders, graters, grinders, homogenizers, and mincers for the determination of moisture by any of the analytical methods given in **Table 6**, which are classified as direct and indirect procedures. The weight of sample is taken before and after it is dried, and the moisture content is calculated. Instruments used for moisture determination are simple to use and provide rapid and reliable measurements and are suitable for routine quality control applications.

**Carbohydrates** Carbohydrates are a major source of energy for humans and are present in all foods (grains, vegetables, fruits, and milk), and vary in form from simple monosaccharides (fructose, glucose, galactose, sorbitol) to oligosaccharides (maltose, sucrose, lactose, raffinose, stachyose, verbascose), and more complex polysaccharides (starch, cellulose, etc.).

The complete analysis of foodstuffs may require the determination of simple sugars and reducing sugars (fructose, glucose, galactose, sorbitol, maltose, lactose), oligosaccharides (sucrose, raffinose, stachyose, verbascose), polysaccharides (starch, cellulose, etc.), and fibers, which vary in amount/form and all of which may play an important role in the quality of the product. Methods of

carbohydrate determination in foods are summarized in **Table 7**.

Neutral monosaccharides, uronic acids, hexosamines, and sialic (neurominic acids) are identified and determined by specific colorimetric reactions. The principle behind the techniques rests on the condensation of the degraded products of the neutral monosaccharides (hexose, pentose and methyl pentose) by sulfuric acid with anthrone, cysteine hydrochloride, orcinol, and phenol reagents. Uronic acids may be determined by colorimetric and manometric procedures. While sialic acids are determined after chemical/enzymatic hydrolysis, gravimetric and Van-Soest detergent based methods are used to determine cellulose, hemicellulose, and fiber.

**Amino acids and proteins** A large variety of proteins, either from animal origin (eggs, fish, meat, and milk) or plant origin (cereal, pulses, fruits, and vegetables) is available from food sources. There is a practical need to determine amino acids, proteins, and protein quality, and to monitor and regulate protein quality in food production, processing, storage, and marketing, which emphasizes the importance of protein quality. The comprehensive nutritional evaluation of protein quality of food and food products begins with the determination of nitrogen content, essential amino acid concentrations, amino acids, protein types, and assessment of

**Table 6** Classification of analytical methods for moisture determination

<i>Direct methods</i>	<i>Indirect methods</i>
Direct methods usually yield accurate and absolute value and are manual and time consuming	Indirect methods are rapid, nondestructive and easily automated
<i>Drying method</i>	<i>Physical and electrical methods</i>
Chemical desiccation	AC and DC conductivity
Freeze-drying	Dielectric capacitance
Oven drying	Microwave absorption
Vacuum drying	<i>Spectroscopic methods</i>
<i>Distillation method</i>	IR absorption
Azotropic distillation	Near-IR reflectance
<i>Chemical methods</i>	Nuclear magnetic resonance
Generation of acetylene	Neutron and $\gamma$ -ray scattering
Heat on mixing with $H_2SO_4$	Refractometry
Karl Fischer titration	Sonic and ultrasonic
<i>Extraction method</i>	
<i>Gravimetric method</i>	
Thermogravimetric analysis	
<i>Absolute methods</i>	
Dew point method	
Gas chromatography	
Manometric method	
Psychrometry	
Volumetric	

**Table 7** Major methods of carbohydrate analysis

<i>Traditional methods</i>	<i>Recent methods</i>
<i>Physical methods</i>	<i>Physicochemical methods</i>
Determine the quantity of a particular sugar present in food	Provide more rapid analysis with a greater precision and specificity
Hydrometry	<i>Chromatography and electrophoresis</i>
Polarimetry	Paper chromatography
Refractometry	Thin-layer chromatography
<i>Chemical methods</i>	Gas chromatography
Classical methods	Liquid chromatography
Ferricyanide method	Ion-exchange chromatography
Iodometry method	
<i>Colorimetric methods</i>	<i>Biochemical methods</i>
Based on chemical reagents used	Enzymatic methods: used for starch and cell wall carbohydrates
Anthrone	
Clegg-anthrone	<i>Flow injection analysis</i>
Copper reduction	Automated method
Dinitrosalicylate	
Phenol/sulfuric acid	
Nelson-Somogyi	
Neocuproine	
Picric	

nutritional value including digestibility (*in vitro* and *in vivo* assays), and biological evaluation (PER, NPU, BV, growth parameters, etc.) by taking albino rats, guinea pigs, and human beings as experimental models. Enzymatic and microbiological tests may also evaluate nutritional quality of foods. Amino acids can be determined by colorimetric and enzymatic methods after hydrolysis of proteins in a specific medium used to diminish destruction of essential amino acids after separation by different techniques (column chromatography, thin-layer chromatography (TLC), gas chromatography (GC), high-performance liquid chromatography, etc.). The availability of all the essential amino acids in food proteins can be measured with the enzymatic ultrafiltration digest.

The determination of protein in food/food products depends on a measurement of a specific element or chemical group in the proteins, which may be carried out either directly by using chemical or physical properties. Based on nitrogen content, protein content can be estimated by Kjeldahl's method (AOAC). Ammoniacal nitrogen produced after destruction of organic matter of foods is multiplied by a coefficient of 6.25 for animal, vegetable, and pulse proteins and by 5.7 for cereals to obtain protein contents. Ammonia content can also be determined by using an ammonia-specific electrode, based on the difference in potential between a reference electrode and a measurement electrode, and with the help of Nessler's reagent. Protein content in food/food products may be determined by various chemical and physical methods as given in Table 8.

Specific protein components in food/food products can be determined by chromatography (ion-exchange chromatography, PC, TLC, GC, LC, etc.), electrophoresis, and immunology or their combination. It is also necessary to use high-resolution techniques such as reversed phase LC, ion-exchange LC, size-exclusion LC, slab electrophoresis, and capillary electrophoresis (CE).

All protein determination methods described are not absolute and demand some form of calibration. The Kjeldahl's method remains the only official method currently available for calibration purposes and maintains its position as the most frequently used technique for the determination of organic nitrogen in food products. CE and immunochemical (enzyme-linked immunosorbent assay) techniques are most suitable for rapid separation and quantification of individual food proteins and are promising for widespread use in food protein analysis due to their high sensitivity, specificity, and simplicity of operation. There are numerous methods for the evaluation of the nutritional quality of food proteins,

**Table 8** Methods of protein determination

Name of method	Principle used
<i>Chemical methods</i>	
Biuret method	Based on binding of copper(II) to a peptide bond in protein molecules at alkaline pH values (measured between 540 and 650 nm)
Lowry method	Based on reduction of Folin–Ciocalteu reagent by oxidation of tyrosine, tryptophan on polypeptide side chain (measured at 750 nm)
Dye binding	Based on the measurements of excess dye remaining in solution after removal of the precipitated protein–dye complex
<i>Physical methods</i>	
UV spectrophotometry	Based on absorption maximum at 280 nm, due to presence of a side chain of aromatic amino acids (Tyr and Try)
Fluorimetry	Based on emission of fluorescence excited at 275 or 345 nm
IR reflectance	Based on characteristics absorbance spectra at different wavelengths
Refractometry	Based on change in refractive index by the displacement of proteins
Turbidimetry	Based on change in intensity of light due to protein diffusion
<i>Immunological method</i>	Based on interaction between an antigen and its corresponding antibody

which are already listed as determination of amino acid composition and essential amino acids, and enzymatic tests, microbiological methods, and biological measures.

**Lipids** The analysis of lipids in food has three distinctive objectives: to determine (1) total lipid content, (2) the composition, and (3) the quality of lipids.

1. *Total lipid content.* There are physical and chemical procedures for oil/fat estimation. In physical procedures, lipids are not isolated and samples are used directly, and lipids are estimated by nuclear magnetic resonance (NMR) spectroscopy. In chemical procedures, lipids are extracted by refluxing the sample in suitable solvents (like petroleum ether) by using standard methods (AOAC or AOCS). Different classes of lipids such as neutral lipids (by nonpolar solvents such as petroleum ether, hexane), phospho- or glycolipids (by polar solvents such as methanol) can be extracted and used for

quantitative determination. TLC, column chromatography, HPLC are specific techniques used for separation of various classes of lipid.

2. *Composition of oil/fat.* To study the composition of oil/fat it is essential to test the purity of an oil/fat for adulteration, accidentally or voluntarily. The specific fatty acid in fat can be determined by GC by preparing methyl esters with sodium methoxide. Mass spectrometry coupled to GC (GC-MS) is the most powerful tool for identification of fatty acids separated by GC. Free fatty acids in oil (index of rancidity) can be determined by titration against standard alkali. Infrared (IR) spectroscopy, Raman spectroscopy, and ultraviolet (UV) spectroscopy (200–400 nm) are used to detect isomers (*trans* and *cis*) of unsaturated fatty acids and conjugated double bonds. It is important to study saponification value (depict fatty acid chain length), iodine value (give the degree of unsaturation), and hydroxyl value (free fatty acids present in fats).

### Ash and Mineral Components

The determination of ash/minerals in food contributes to assessment of a food's nutritional value and refers to elements and verifying if the food contains some minerals in quantities dangerous to the health of the consumer, whether their presence is natural or adulteration of certain food stuffs/processed or stored food products.

Ash content is a measure of total minerals and is a reliable index of nutritional value for many foods (tea, flour, edible gelatin, etc.) and feed (for poultry and cattle) and is recognized as a useful tool in determining the nature and distribution of mineral constituents of food. Ash is the inorganic material left after complete oxidation of organic material at high temperatures (500–600°C).

Ash is prepared by two direct procedures (dry and wet ashing) and an indirect technique (conductometric method that can determine the total electrolyte content of foods), governed by the following purpose: the particular constituents to be determined and the method of analysis to be used. Dry ashing is the standard method to determine ash content of a sample (AOAC). For wet ashing the sample is digested in a mixture of HNO<sub>3</sub> and HClO<sub>4</sub> (ratio 4:1).

Gravimetry (for sulfur), titrimetric (chloride), fluorimetry (Se, Al, F), colorimetric methods, or atomic spectroscopy are the traditional classical chemical methods for the determination of individual elements. Atomic spectroscopy techniques have a much higher sensitivity and specificity and provide a complete profile of elements in a food/feed. Emission

spectrometry, polarography, X-ray spectrometry, mass spectrometry, and activation analysis are important physical methods used for the determination of minerals, and are suitable for determination of alkaline metals, provide accurate and reproducible results, and offers simultaneous determination of numerous elements.

**Other food components** It is also important to analyze the toxic constituents/antinutritional factors, which are present naturally in foods, produce a deleterious effect when ingested by humans or animals. The toxic constituents/antinutritional factors, which are present naturally in foods are flatous producing factors, protease inhibitors, hemagglutinins (lectin), and saponins present in legumes, and glucosinolates (in oil seed crops), cyanogens (in bitter almond, lima beans), gossypol (in cotton), and lathyrigen (seeds of *Lathyrus* species). A large number of standard (AOAC) analytical techniques are adopted to determine the food components such as bitter compounds like anthocyanins, carotenoids (pigments), flavonoids, phenols, terpenoids, vitamins, aroma compounds, alcohols, organic acids, antioxidants and preservatives etc.

See also: **Amino Acids.** **Carbohydrates:** Overview. **Food and Nutritional Analysis:** Sample Preparation; Packaging Materials. **Liquid Chromatography:** Overview. **Thin-Layer Chromatography:** Overview.

### Further Reading

- AACC (1983) *Approved Methods of Analysis*, 8th edn. St. Paul, MN: American Association of Cereal Chemists.
- AOAC (1995) *Official Methods of Analysis*, 16th edn. Association of Official Analytical Chemists, Arlington, Virginia, USA.
- AOCS (1990) *Official Methods of Recommended Practices*. Chicago, IL: American Oil Chemists Society.
- Fennema OR (1996) *Food Chemistry*. New York: Dekker.
- Gruenwedel DW and Whitaker JR (eds.) (1984) *Food Analysis: Principles and Techniques*, vols 1–4. New York: Dekker.
- James CS (1994) *Analytical Chemistry of Foods*. London: Blackie Academic and Professional.
- Kulp K and Ponte JG Jr (2000) *Handbook of Cereal Science and Technology*, 2nd edn. New York: Dekker.
- Linden G (ed.) (1996) *Analytical Techniques for Foods and Agricultural Products*. New York: VCH.
- Multon JL (ed.) (1997) *Analysis of Food Constituents*. New York: Wiley-VCH.
- Nollet LML (ed.) (1996) *Handbook of Food Analysis*, vols 1–2. New York: Dekker.
- Nollet LML (ed.) (2000) *Food Analysis by HPLC*, 2nd edn. New York: Dekker.

Oliveira FAR, Oliveira JC, Hendrickx ME, Korr D, and Gorris LGM (eds.) (1999) *Processing Food: Quality Optimization and Process Assessment*. New York: CRC Press.

Poumeranz Y and Meloan CE (1994) *Food Analysis: Theory and Practice*, 3rd edn. New York: Chapman and Hall.

Sorensen (2000) *Chromatography and Capillary Electrophoresis in Food Analysis*. New York: Springer.

Wilson RH (ed.) (1994) *Spectroscopic Techniques for Food Analysis*. New York: Wiley-VCH.

## Sample Preparation

**Kiyoshi Matsumoto**, Kyushu University, Fukuoka, Japan

**Hiroyuki Ukeda**, Kochi University, Nankoku, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

### Sample Preparation

In many respects, the sampling and preparation of a sample are critical steps in any technique of food analysis. The objective of sampling is to ensure that the sample taken for analysis is representative of a defined whole, and the method of sampling depends on the size and nature of the defined whole.

Obtaining a sample for analysis that is representative of the whole is referred to as sampling, and the total quantity from which a sample is obtained is called the population. An accurate and precise estimate of the quality of a population will only be achieved by using an adequate sampling technique because only a portion of the population is used for analysis. With a proper sampling technique, the estimated value will reflect the characteristics of the population.

Special care, however, is required in the sampling and sample preparation for food analysis because of the peculiarities of food. Foods are derived from biological substances and contain various organic compounds. Many of these components are influenced by various external and internal influences such as temperature, light, moisture, microorganisms, metabolism, and ripening. Therefore, rapid preparation of food samples enough to prevent composition change is very important in food analysis. Preparation at low temperatures during processing or storage in the frozen state is also effective for preventing the composition change of food samples.

In food analysis, analyses are generally performed on the edible portion of the defined food, discarding the nonedible portion, unless the compositions with respect to the total sample weight are required. Variations in moisture of food are relatively large, even among the same kinds of food. Therefore, values

corrected on the basis of dry matter (values of anhydrous material) or values of moisture contents are used in evaluating the analyzed value.

The general steps for sample preparation and sample processing are (1) sampling and size reduction, (2) comminution and homogenization, (3) pretreatments of the sample, such as predrying or defatting, (4) preservation of the prepared sample, and (5) cleanup of the analytical sample for instrumental analysis. This article covers the practical considerations of sample preparation necessary for food analysis.

### Sampling and Size Reduction

It is important to define clearly the population that is to be sampled. Populations are generally finite, such as the size of a defined lot, except for evaluating a process. The sampling methods selected depend on the purpose of the inspection, the nature of the product, the nature of the test method, and the nature of the population being investigated. The basic principle for sampling is probability sampling. This sampling plan provides a statistically sound basis for obtaining representative samples while eliminating human bias and, therefore, is the most desirable. The most useful sampling method is random sampling in which samples are simply taken at random from the whole population. The random sampling process allows all parts of the population to have the same chance of being sampled. The goal of providing the same chance can be achieved by means of random number tables or computer-generated random numbers. Sample size is determined by lot size, the degree of accuracy required, and the expense of the test method. However, a general standard for the number to be sampled is proposed. For instance, to select the sample flour from sacks, the number of sacks to be sampled is determined by the square root of the number of sacks in the lot.

If the mass of the collected samples is too large for analysis, it is subjected to size reduction to obtain a smaller quantity for analysis. For granular or powdered material, such as cereals, pulses, nuts, and



Oliveira FAR, Oliveira JC, Hendrickx ME, Korr D, and Gorris LGM (eds.) (1999) *Processing Food: Quality Optimization and Process Assessment*. New York: CRC Press.

Poumeranz Y and Meloan CE (1994) *Food Analysis: Theory and Practice*, 3rd edn. New York: Chapman and Hall.

Sorensen (2000) *Chromatography and Capillary Electrophoresis in Food Analysis*. New York: Springer.

Wilson RH (ed.) (1994) *Spectroscopic Techniques for Food Analysis*. New York: Wiley-VCH.

## Sample Preparation

**Kiyoshi Matsumoto**, Kyushu University, Fukuoka, Japan

**Hiroyuki Ukeda**, Kochi University, Nankoku, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

### Sample Preparation

In many respects, the sampling and preparation of a sample are critical steps in any technique of food analysis. The objective of sampling is to ensure that the sample taken for analysis is representative of a defined whole, and the method of sampling depends on the size and nature of the defined whole.

Obtaining a sample for analysis that is representative of the whole is referred to as sampling, and the total quantity from which a sample is obtained is called the population. An accurate and precise estimate of the quality of a population will only be achieved by using an adequate sampling technique because only a portion of the population is used for analysis. With a proper sampling technique, the estimated value will reflect the characteristics of the population.

Special care, however, is required in the sampling and sample preparation for food analysis because of the peculiarities of food. Foods are derived from biological substances and contain various organic compounds. Many of these components are influenced by various external and internal influences such as temperature, light, moisture, microorganisms, metabolism, and ripening. Therefore, rapid preparation of food samples enough to prevent composition change is very important in food analysis. Preparation at low temperatures during processing or storage in the frozen state is also effective for preventing the composition change of food samples.

In food analysis, analyses are generally performed on the edible portion of the defined food, discarding the nonedible portion, unless the compositions with respect to the total sample weight are required. Variations in moisture of food are relatively large, even among the same kinds of food. Therefore, values

corrected on the basis of dry matter (values of anhydrous material) or values of moisture contents are used in evaluating the analyzed value.

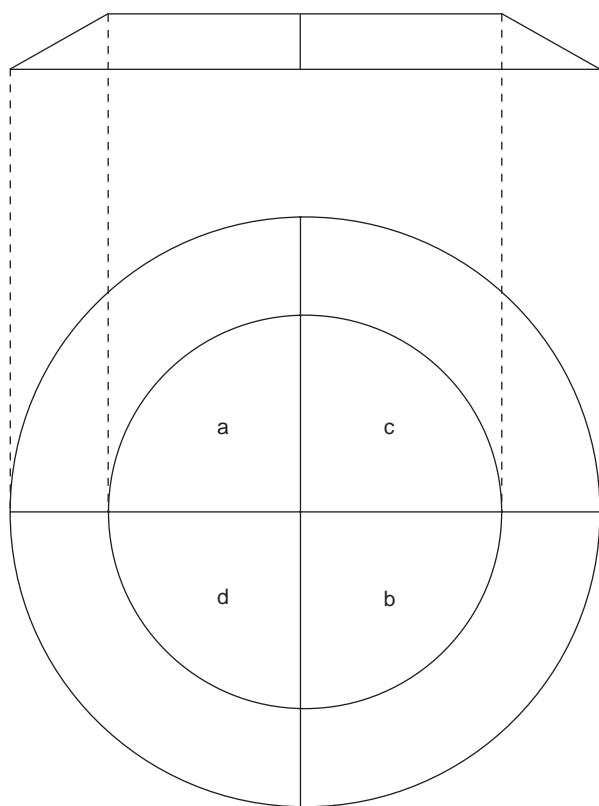
The general steps for sample preparation and sample processing are (1) sampling and size reduction, (2) comminution and homogenization, (3) pretreatments of the sample, such as predrying or defatting, (4) preservation of the prepared sample, and (5) cleanup of the analytical sample for instrumental analysis. This article covers the practical considerations of sample preparation necessary for food analysis.

### Sampling and Size Reduction

It is important to define clearly the population that is to be sampled. Populations are generally finite, such as the size of a defined lot, except for evaluating a process. The sampling methods selected depend on the purpose of the inspection, the nature of the product, the nature of the test method, and the nature of the population being investigated. The basic principle for sampling is probability sampling. This sampling plan provides a statistically sound basis for obtaining representative samples while eliminating human bias and, therefore, is the most desirable. The most useful sampling method is random sampling in which samples are simply taken at random from the whole population. The random sampling process allows all parts of the population to have the same chance of being sampled. The goal of providing the same chance can be achieved by means of random number tables or computer-generated random numbers. Sample size is determined by lot size, the degree of accuracy required, and the expense of the test method. However, a general standard for the number to be sampled is proposed. For instance, to select the sample flour from sacks, the number of sacks to be sampled is determined by the square root of the number of sacks in the lot.

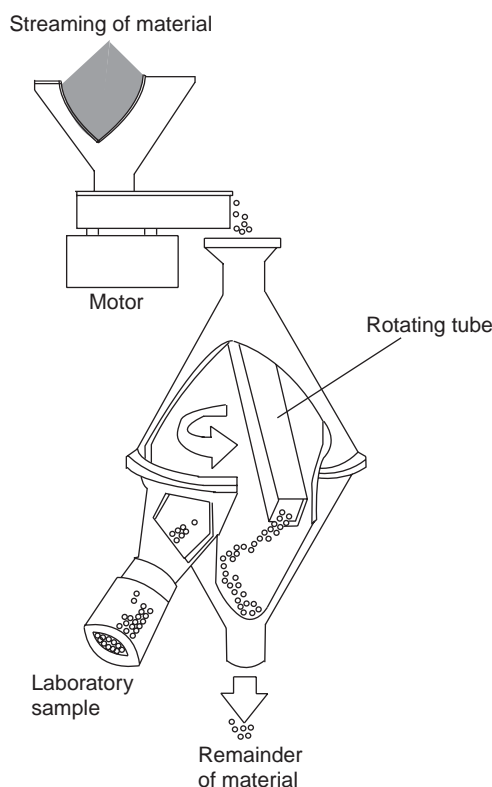
If the mass of the collected samples is too large for analysis, it is subjected to size reduction to obtain a smaller quantity for analysis. For granular or powdered material, such as cereals, pulses, nuts, and





**Figure 1** Method for size reduction (granular or powdered material).

seeds, the sample can be spread on a clean surface and divided into quarters. The two opposite quarters are combined as shown in **Figure 1**. If the mass is still too large for analysis, the process is repeated until an appropriate amount is obtained. Sample dividers that mechanically mix and divide powdered or granular materials may also be useful for the size reduction. Rotary tube dividers shown in **Figure 2** or equivalents are generally used for this purpose. For a large-sized fish, such as salmon, yellowtail, or bonito, the whole body is filleted, and the fillet from one side is sliced into pieces of equal width, and every third slice is selected, as shown in **Figure 3**. For fruits such as apples, oranges, or pears, the whole fruit is cut as shown in **Figure 4**, and pieces a, b, c, and d are selected. For liquid or viscous materials, such as milk, sugar syrup, or honey, the material should be mixed thoroughly immediately before sampling. Small quantities of liquefied materials can be mixed by rotating and shaking in a closed container that has a volume at least twice that of the sample. This kind of sampling for apparently homogeneous materials is simple, as a sample can be taken from any location, and the analytical data obtained will be representative of the whole.



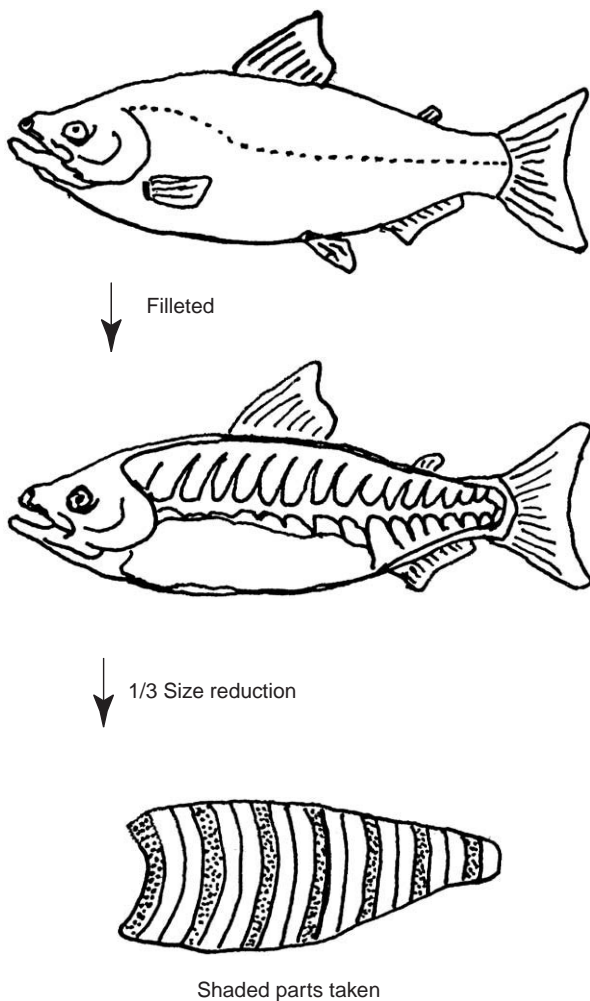
**Figure 2** A rotating tube divider for reducing a large sample of dry, free flowing material to a laboratory size sample.

## Comminution and Homogenization

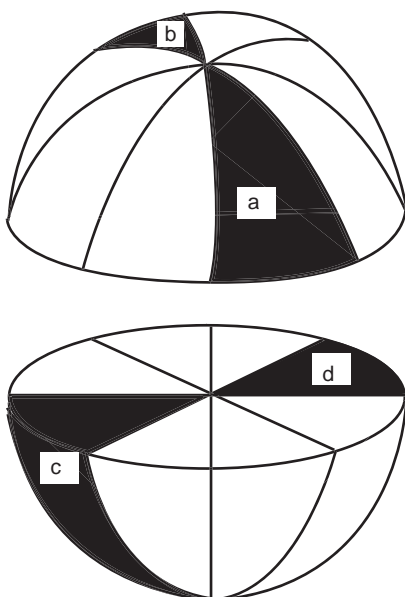
In order to prevent differences between samples each portion, should be ground. For the determination of moisture, total protein, and mineral concentrations, dry foods are generally ground to pass through a 0.75 mm sieve screen. For assays involving extraction (lipids, carbohydrates, various forms of protein), samples are ground to pass through a 0.40 mm sieve screen. Generally, foods grind better after predrying to the moisture level of air-dried materials. If the materials are rich in fat, defatting is recommended before grinding to avoid losses of fat as described in the section 'Defatting'.

### Grinding

Mechanical methods range from the simple mortar and pestle to power-driven hammer mills. The usable grinding devices among them are roller mills, food processors, coffee mills, and ultracentrifugal mills. Hammer mills wear well, and reliably and effectively grind cereals and dry foods. An ultracentrifugal mill grinds by beating, impacting, and shearing. The food is fed to a grinding chamber from an inlet and is reduced in size by a rotor. The small ground particles



**Figure 3** Method for size reduction (large-sized fish).



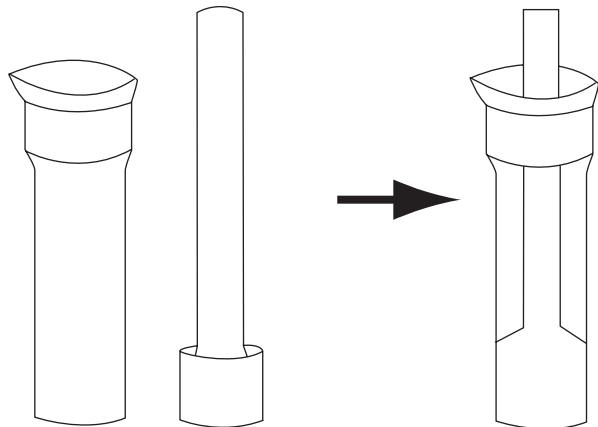
**Figure 4** Method for size reduction (fruits).

are delivered by centrifugal force into a collection pan. For small samples, especially if fine grinding is required, a ball mill is used. A ball mill grinds by rotating the material in a container half-filled with ceramic balls, which exert an impact grinding action. A chilled ball mill can be used to grind frozen foods without predrying. Grinding foods while frozen reduces undesirable chemical changes (i.e., the Maillard reaction) that can occur during milling.

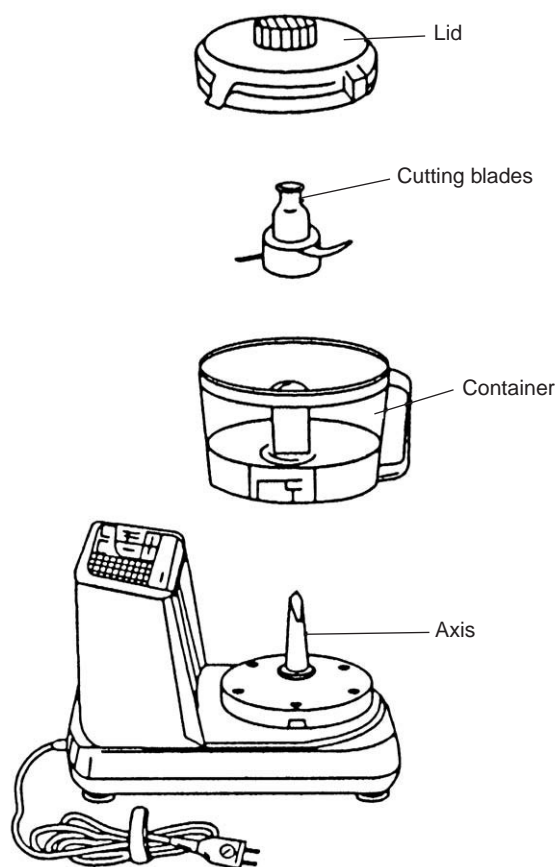
The grinding process should not heat the sample, and therefore the grinder should not be overloaded to prevent the heat produced through friction. Contamination by metals from the abrasives of the mill should be avoided if trace metal analysis is to be performed. In such cases, the working parts of the grinder should be made of an inert material (glass, ceramic, or agate). After grinding the sample, all the powder ground should be collected using a clean brush.

### Homogenization

For the disintegration of moist material such as tissues, tubers, vegetables, fruits, roots, or meats, various fine-slicing devices are available. The common devices are mortars and pestles, meat mincers or grinders, mixers or blenders, bowl cutters, and tissue grinders. Bowl cutters can be used for leafy vegetables and fleshy tubers, while meat mincers may be better suited for fruit, roots, and meat. Power-driven mortars and pestles can finely subdivide and are used for making fish meat paste. Tissue grinders (**Figure 5**) are used for small samples of soft materials. Waring blenders are effective in grinding soft and flexible foods and suspensions. Knives rotating at up to 25 000 rpm disintegrate a sample into a suspension. Blenders are routinely used for the disintegration of almost all moist food samples (**Figure 6**).



**Figure 5** A tissue grinder (homogenizer with piston-type, ground glass pestle).



**Figure 6** A blender-disintegrator apparatus equipped with stainless cutting blades.

## Pretreatment of Sample

Foods rich in fat or moisture are difficult to grind or disintegrate in their intact state. Predrying and defatting are often used as a pretreatment, except for the determination of heat-sensitive and easily oxidated components such as vitamins.

### Predrying

If the components of a low-moisture, high-fat sample to be analyzed are heat-stable, predrying is recommended to the moisture level of air-dried material before grinding. For instance, an appropriate amount of the sample is cut (or loosened) into pieces and weighed. The material is placed in a porcelain dish, heated on a warm water bath (or in an electric dryer at 60–80°C), and dried. The dried material is kept for 24 h at room temperature to equilibrate it to the humidity of the environmental air. The equilibrated material is weighed and then ground.

### Defatting

When nuts and seeds, such as walnuts or sesame seeds, and processed foods containing a high level of

fat are ground, the fat oozes out and clings to the surface of the grinder, and this consequently leads to a loss of fat. Therefore, defatting is required for samples rich in fat before grinding. In brief, the weighed material is roughly ground and dried. The material is placed in a flask containing diethyl ether and kept overnight. The content of the flask is filtered, and the solid material on the filter paper is air-dried. After a thorough evaporation of the ether, the solid material is dried at 60–80°C. The dried solid material is equilibrated with air and weighed. The fat in the filtrate (ether phase) should be measured separately. Defatting is suitable for fat-rich foods. However, it should be avoided when fat-soluble vitamins are to be analyzed.

## Preparation for Specified Components

A general procedure cannot always be applied for a particular component such as dietary fiber and vitamins. Special attention should be paid to these components; a sample should be ground as finely as possible for dietary fiber analysis, and light and oxidation should be avoided for vitamin analyses.

### Dietary Fiber

Estimates of fiber are most consistent when the samples contain little fat (less than 5–10% fat), are dry, and finely ground. If the sample contains more than 10% fat, fat should be extracted by mixing the sample with 25 parts (vol/wt) of petroleum ether or hexane. After centrifugation and decantation of the organic solvent, the lipid extraction is repeated two more times. Then the sample is dried overnight in a vacuum oven at 70°C and ground finely enough to pass through a 0.3–0.5 mm mesh screen. It should be noted that loss of weight due to fat and moisture removal must be recorded and appropriate correction to the final percentage dietary fiber value found in the analysis must be made.

Nonsolid samples containing less than 10% fiber are best analyzed after lyophilization and treated as described above. Nonsolid samples containing greater than or equal to 10% fiber can be analyzed without drying if the sample is homogeneous and low in fat and if particle size is sufficiently small to allow the efficient removal of digestible carbohydrate and protein.

### Vitamins

Because some vitamins are adversely affected by conditions such as light, oxygen, pH, and heat, proper precautions need to be taken to prevent any deterioration throughout the analytical process,

regardless of the type of assay used. Most vitamin assays involve the extraction of the vitamin from its biological matrix prior to analysis. This generally includes one or more of the following treatments: heat, acid, alkali, solvent, and enzyme.

In general, extraction procedures are specific for each vitamin and designed to stabilize the vitamin. In some cases, certain procedures are applicable to the combined extraction of more than one vitamin, e.g., for thiamine and riboflavin and some of the fat-soluble vitamins. Typical extraction procedures are as follows:

- Ascorbic acid: Cold extraction with metaphosphoric acid/acetic acid.
- Vitamins B<sub>1</sub> and B<sub>2</sub>: Boiling or autoclaving in acid plus enzyme treatment.
- Niacin: Autoclaving in acid (nongrain products) or alkali (grain products).
- Vitamins A, E, or D: Organic solvent extraction, saponification, and reextraction with organic solvents. Antioxidants are routinely added to these unstable vitamins to protect them from oxidation.

Analysis of fat-soluble vitamins may require saponification, generally either overnight at room temperature or by refluxing at 70°C. In the latter case, an air-cooled reflux vessel provides an excellent control of conditions conducive to oxidation.

## Preservation of Prepared Sample

Prepared samples should be analyzed immediately. If immediate analysis is not possible, samples should be chilled or dried, depending on the storage period and conditions, the analyte, and the analyses that are to be performed. The following factors should be controlled during storage: (1) moisture changes, (2) oxidation, (3) light decomposition, (4) enzymatic changes, (5) heat denaturation, and (6) microbial changes and contamination.

To minimize oxidative changes, preservation at low temperatures under nitrogen is recommended for most foods. Compositional changes (desiccation or moisture absorption, volatilization) of relatively dry foods are reduced by storage in a closed container at ~4°C. If such containers contain powdered materials and are opened frequently to remove samples, it is advisable to check whether moisture changes have taken place. Frozen foods should be wrapped in plastic material or placed in air-tight containers to reduce dehydration.

Many food materials contain enzymes that may change the food components being analyzed. Enzyme

activity should therefore be controlled using methods that depend on the nature of the food. Heat denaturation for enzyme inactivation (e.g., use of a microwave oven) and freezer storage (–20°C to –30°C) to limit enzyme activity are common methods. However, some enzymes are more effectively controlled by changing the pH or by salting out in the sample preparation. Oxidative enzymes may be controlled by adding reducing agents.

To reduce or eliminate microbial attack, several methods can be used. They include freezing, drying, and the use of a preservative or a combination of any of the three. Commonly used preservatives include sorbic acid or sorbate, sodium benzoate, sodium salicylate, tylosin, formaldehyde, toluene, and thymol. The selection of a method or preservative will depend on the nature of the food, expected contamination, storage period and conditions, and the analyses that are to be performed.

When the prepared sample is to be used repeatedly, it is recommended that the sample be stored by dividing it into several containers in order to avoid any deterioration of the sample by repetitive freezing and thawing or exposure to air. The container should be labeled so as to identify the sample name and number, the preparation date, and the person in charge of preparation even after a long storage period.

## Novel Cleanup Techniques

Often, the crude sample extract is partially purified before the analyses. A cleanup procedure is frequently required for analyses based on various kinds of chromatography. The necessity for and the degree of cleanup required depend, to a large extent, on the instrumental detector to be used and, to a lesser extent, on the type of chromatography in the automated separation stage of the analysis. In general, the sample cleanup is the most time-consuming, labor-intensive, and error-prone step of standard analytical methods. Recently, several novel methods have been employed to decrease the time required to complete the operation and the quantities of solvents consumed in the extraction procedure.

### Solid-Phase Extraction

Solid-phase extraction (SPE) typically involves small quantities of dry 30–40 µm diameter packing material contained in a single-use polypropylene or glass cartridge (30 mg to 10 g of sorbent) or disk that can be used to purify, concentrate, or, less frequently, derivatize the analytes prior to analyses. For cartridge materials, typical retention capacities are ~5% of the sorbent mass. However, one must also account

for the other components of the crude extract that can be retained on the sorbent material. Minimum elution volumes are only 120  $\mu\text{L}$  per 100 mg of packing material. Although most packing materials are silica-based, polymeric packings can also be used. Many of the common liquid chromatography packing materials are available in one or more of the SPE formats as well. Given the limited resolving power of these systems, conventional strategies are to either selectively retain the analytes on the packing material (during the loading and rinsing cycles) or conversely retain interfering impurities while selectively eluting analytes from the packing material (elution cycle). An intermediate wash/rinse cycle can serve to separate other contaminants from the adsorbed analytes (Figure 7).

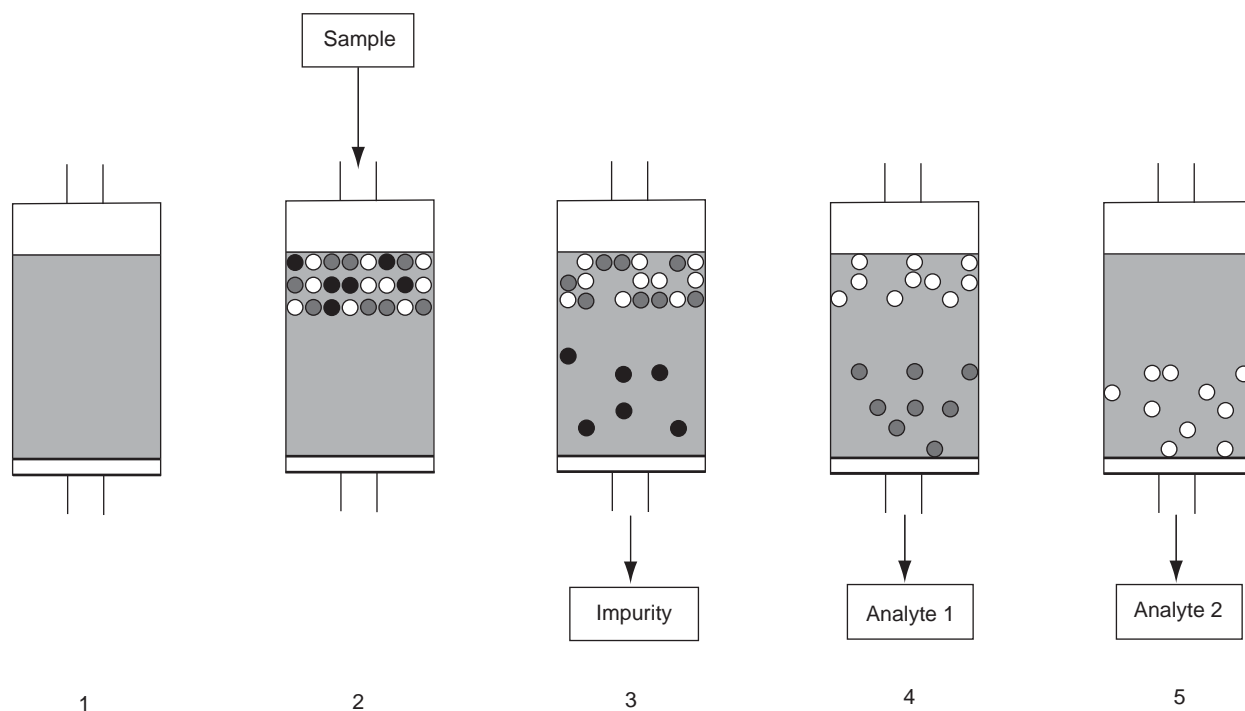
SPE has numerous advantages over traditional liquid-liquid extractions: (1) less solvent required, (2) faster, (3) less glassware needed (lower cost and potential for contamination), (4) better precision and accuracy, (5) minimal solvent evaporation for further analysis (e.g., GC), and (6) readily automated.

The most recent version of this method is called solid-phase microextraction (SPME). In this adaptation, the phase is bound onto a fine fused silica filament (*ca.* the size of a 10  $\mu\text{L}$  syringe needle). The

filament can be immersed in a sample or in the headspace above a sample. After the desired loading time, the filament is pulled into a protective metal sheath, removed from the sample, and forced through the septum of a gas chromatograph where the adsorbed volatiles are thermally desorbed from the filament. For SPME, one can readily see the advantages of simplicity and freedom from solvent usage. The primary concerns are for sensitivity limitations, precision, and the life of the filament. If the filament must be replaced because of breakage, the reproducibility of the characteristics of the old filament by the new one can be an issue.

### Gel Permeation Chromatography

Extraction processes for determining semivolatile organic compounds in food samples are not very discriminating, and the other constituents of a sample can be detrimental to the performance of the analytical method, causing deposits in the injectors of GC and GC-MS instrumentation. However, gel permeation chromatography (GPC) cleanup extracts the analyte very well from any number of other constituents.



**Figure 7** SPE based on reverse-phase chromatography. 1. Solvate the bonded phase with six to ten cartridge holdup volumes of methanol or acetonitrile. Flush the cartridge with six to ten holdup volumes of water or buffer. Do not allow the cartridge to dry out. 2. Load the sample dissolved in a strongly polar solvent. 3. Elute interfering impurities with the strongly polar solvent. 4. Elute weakly held analytes (Analyte 1) of interest with a less polar solvent. 5. Elute more tightly bound analytes (Analyte 2) with progressively more nonpolar solvents.



Various cleanup techniques can be used to remove these contaminants, but they are time-consuming, may remove target compounds, and cannot remove all interfering compounds. One of the primary classes of problem compounds is lipid in food samples. GPC cleanup is a technique that separates target compounds from a sample based on their molecular size. Since lipids are very large molecules compared to the target compounds in these methods, they are effectively removed from the extract prior to analysis. The packing gel used in GPC cleanup is porous and is characterized by the range or uniformity of that pore size. GPC operates on the principle of loading all components in an extract on the gel bed, and then selectively removing the components of larger molecular size. This procedure provides for the efficient separation of typical semivolatile and pesticide components from various higher-molecular weight compounds. This also improves method performance and extends GC column life, leading to more efficient analyses. With autosampling and automatic fraction collection as part of the GPC

system, cleanups can be performed with little or no operator intervention, requiring only that the final extracts be concentrated.

**See also:** **Carbohydrates:** Dietary Fiber Measured as Non-starch Polysaccharides in Plant Foods. **Extraction:** Solvent Extraction Principles; Solid-Phase Extraction; Solid-Phase Microextraction. **Food and Nutritional Analysis:** Overview. **Sampling:** Theory; Practice. **Sensory Evaluation.**

## Further Reading

- AOAC International (1995) *Official Methods of Analysis*, 16th edn. Gaithersburg, MD: AOAC International.
- Horwitz W (1988) Sampling and preparation of samples for chemical examination. *Journal of the Association of Official Analytical Chemists* 71: 241–245.
- Pomeranz Y and Meloan CE (1994) *Food Analysis: Theory and Practice*, 3rd edn, pp. 11–29. New York: Chapman and Hall.
- Proctor A and Meullenet J-F (1998) Sampling and sample preparation. In: Nielsen SS (ed.) *Food Analysis*, 2nd edn, pp. 71–82. Gaithersburg, MD: Aspen Publishers.

## Additives

**M Gallego**, University of Córdoba, Córdoba, Spain

**M González**, Instituto Canario de Investigaciones Agrarias, La Laguna, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Foods naturally provide a series of substances that maintain the vital processes and normal development of the body, but sometimes it is desirable to add foreign substances to foods in order to modify some aspect of its properties; these substances are called food additives. The advantages of food additives are undeniable and so no country prohibits them entirely. Currently, the establishment of international positive lists and the acceptable daily intake (ADI) instituted by international organizations ensure that these additives, used at a normal dosage, will not have toxic effects on the human body. Analytical methods for the determination of additives have been developed so that the stability of these substances until the food's consumption can be monitored and also to ensure that the additives are within the limits established by legislation. The sample treatment

methods (which consume more than 70% of the analytical process) and the determinative and discriminative methodologies for the determination of additives in foods have gone through remarkable changes in the last few years as a consequence of the spectacular advances in analytical instrumentation. However, in this regard it has to be pointed out that these modern methodologies (with sophisticated instrumentation) are uncommon in many food industries.

## General Information about Food Additives

### Definition of Additive

Among the various definitions of food additives that can be found in international legislation and accords is the one adopted by the European Union (EU) in its European Community Directive 89/107/EEC of the Council of 21 December 1988: "any substance not normally consumed as a food in itself and not normally used as a characteristic ingredient of food whether or not it has nutritive value, the intentional addition of which to food for a technological purpose in the manufacture, processing, preparation,



Various cleanup techniques can be used to remove these contaminants, but they are time-consuming, may remove target compounds, and cannot remove all interfering compounds. One of the primary classes of problem compounds is lipid in food samples. GPC cleanup is a technique that separates target compounds from a sample based on their molecular size. Since lipids are very large molecules compared to the target compounds in these methods, they are effectively removed from the extract prior to analysis. The packing gel used in GPC cleanup is porous and is characterized by the range or uniformity of that pore size. GPC operates on the principle of loading all components in an extract on the gel bed, and then selectively removing the components of larger molecular size. This procedure provides for the efficient separation of typical semivolatile and pesticide components from various higher-molecular weight compounds. This also improves method performance and extends GC column life, leading to more efficient analyses. With autosampling and automatic fraction collection as part of the GPC

system, cleanups can be performed with little or no operator intervention, requiring only that the final extracts be concentrated.

**See also:** **Carbohydrates:** Dietary Fiber Measured as Non-starch Polysaccharides in Plant Foods. **Extraction:** Solvent Extraction Principles; Solid-Phase Extraction; Solid-Phase Microextraction. **Food and Nutritional Analysis:** Overview. **Sampling:** Theory; Practice. **Sensory Evaluation.**

## Further Reading

- AOAC International (1995) *Official Methods of Analysis*, 16th edn. Gaithersburg, MD: AOAC International.
- Horwitz W (1988) Sampling and preparation of samples for chemical examination. *Journal of the Association of Official Analytical Chemists* 71: 241–245.
- Pomeranz Y and Meloan CE (1994) *Food Analysis: Theory and Practice*, 3rd edn, pp. 11–29. New York: Chapman and Hall.
- Proctor A and Meullenet J-F (1998) Sampling and sample preparation. In: Nielsen SS (ed.) *Food Analysis*, 2nd edn, pp. 71–82. Gaithersburg, MD: Aspen Publishers.

## Additives

**M Gallego**, University of Córdoba, Córdoba, Spain

**M González**, Instituto Canario de Investigaciones Agrarias, La Laguna, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Foods naturally provide a series of substances that maintain the vital processes and normal development of the body, but sometimes it is desirable to add foreign substances to foods in order to modify some aspect of its properties; these substances are called food additives. The advantages of food additives are undeniable and so no country prohibits them entirely. Currently, the establishment of international positive lists and the acceptable daily intake (ADI) instituted by international organizations ensure that these additives, used at a normal dosage, will not have toxic effects on the human body. Analytical methods for the determination of additives have been developed so that the stability of these substances until the food's consumption can be monitored and also to ensure that the additives are within the limits established by legislation. The sample treatment

methods (which consume more than 70% of the analytical process) and the determinative and discriminative methodologies for the determination of additives in foods have gone through remarkable changes in the last few years as a consequence of the spectacular advances in analytical instrumentation. However, in this regard it has to be pointed out that these modern methodologies (with sophisticated instrumentation) are uncommon in many food industries.

## General Information about Food Additives

### Definition of Additive

Among the various definitions of food additives that can be found in international legislation and accords is the one adopted by the European Union (EU) in its European Community Directive 89/107/EEC of the Council of 21 December 1988: "any substance not normally consumed as a food in itself and not normally used as a characteristic ingredient of food whether or not it has nutritive value, the intentional addition of which to food for a technological purpose in the manufacture, processing, preparation,

treatment, packaging, transport or storage of such food results, or may be reasonably expected to result, in it or its by-products becoming directly or indirectly a component of such foods.” This definition emphasizes that additives are substances added intentionally to food to produce permanent functional effects and that they are generally not considered themselves to be foods.

Another food additive definition is that proposed by the Codex Alimentarius Commission: “any substance not normally consumed as a food itself and not used as a typical ingredient of the food, whether or not it has nutritive value, the intentional addition of which to food for a technological (including organoleptic) purpose in the manufacture, processing, preparation, treatment, packing, packaging, transport or holding of such food results, or may be reasonably expected to result (directly or indirectly), in it or its by-products becoming a component of or otherwise affecting the characteristics of such foods. The term does not include ‘contaminants’ or substances added to food for maintaining or improving nutritional qualities.” This definition indicates the difference between an additive that is added intentionally and a contaminant (e.g., micro-organisms and their toxins, heavy metals) whose presence in a food is unintentional.

These two examples illustrate that different food additive definitions tend to be very similar because a general consensus has been reached among the majority of countries as to what an additive is. It should also be pointed out that a special statute is normally reserved for dietetic products, nutritive substances added to food products (e.g., vitamins, amino acids, minerals, etc.), technological processing aids, and aromas.

**Classification of Food Additives**

There are many classification methods for additives: alphabetical order, numeration of each substance within a collection of chemicals, or classification by chemical function. The most common method is to group them by functional categories, that is, by anticipated use. This is the method used by the EU, which classifies additives in 24 categories, and by the Codex Alimentarius Commission, which uses 21 categories. The juxtaposition between the types of additives classified by the EU and the Codex Alimentarius Commission is not perfect, given that the number of categories is different and some categories that exist in one do not exist in the other. Furthermore, there is a large amount of multi-functional additives that could be classified in

**Table 1** Comparison between the classifications of food additives established by EU and Codex Alimentarius

<i>Improved characteristic</i>	<i>EU legislation</i>	<i>Codex Alimentarius</i>
Color	Color additives	Color additives
Color	–	Color stabilizers
Preservation	Preservatives	Preservatives
Preservation	Antioxidants	Antioxidants
Texture	Emulsifiers	Emulsifiers
Texture	Emulsifying salts	Emulsifying salts
Texture	Thickeners	Thickeners
Texture	Gelling agents	Gelling agents
Texture	Stabilizers	Stabilizers
Flavor	Flavor enhancers	Flavor enhancers
Flavor	Acids	–
Flavor	Acidity regulators	Acidity and pH regulators
Texture	Anticaking agents	Anticaking agents
Flavor	Sweeteners	Sweeteners
Texture	Modified starch	–
Texture	Raising agents	Raising agents
Texture	Antifoaming agents	Antifoaming agents
Texture	–	Foaming agents
Texture	Glazing agents and lubricants	Lubricants
Texture	Firming agents	–
Texture	Flour treatment agents	Flour treatment agents
Texture	Humectants	Humectants
Texture	Sequestrants	–
Texture	Enzymes	–
Texture	Bulking agents	Bulking agents
Texture	Propellants and packaging gases	Propellants

various categories. **Table 1** shows a comparison between the two different classifications.

In practical terms, a more general classification could be created that includes the additive categories previously mentioned, separating them by the functions that they fulfill in the foods:

1. additives that increase the shelf life or stability of a food,
2. additives that improve sensorial qualities without altering the nature or quality of the food in such a way that could deceive the consumer,
3. additives that improve the nutritional value of the foods, and
4. additives that aid in the manufacture, transformation, preparation, treatment, packaging, transportation, or storage of the foods.

**Food Additive Regulations**

Although additives are generally associated with the modern food industry, in reality they have been used for centuries. Preserving food has always been a

concern for humans and to achieve this end, our ancestors used salt to preserve meats and fish; preserved fruit with sugar; and pickled or canned vegetables with vinegar. During the nineteenth century and the beginning of the twentieth century, new chemical substances were discovered that could be used to preserve and color foods. In the twentieth century the need to combat commercial fraud and abuse led to the first regulation of chemical food additives by France in 1912, giving rise to the concept of a 'positive list' of chemicals allowed to be added to foods. For the first time there was explicit prohibition to the addition of chemicals or colorants, which had not been previously approved by law, to foods. The concept of positive lists of additives has lasted until today and is still in use in modern international legislation.

### **Regulations in Europe and in the United States**

Since 1988, the EU, using safety evaluations done by its Scientific Committee for Food (SCF), has harmonized the national legislation of its Member States through European Community Directive 89/107/EEC, concerning additives that can be used in products destined for human consumption. The role of the SCF has grown in importance since the adoption of directives on sweeteners (Directive 94/35/EC), colorants (Directive 94/36/EC), and other additives (Directive 95/2/EC) in 1994 and 1995. The committee, created in 1974 by the European Commission, has dedicated itself to evaluating the possible risks of food additives and elaborating guidelines for their safe use. Meanwhile, the European Commission has established concrete criteria on the purity of additives, the general or specific categories of foods with which additives may be used, and the maximum amounts that can be added to them.

In the United States, the regulation of additives is done by the Food and Drug Administration (FDA). As a response to the widespread use of many substances not approved by the FDA, in 1958 the agency created the 'GRAS lists' (Generally Recognized As Safe), comprising substances that were unanimously recognized as safe. These lists, the US equivalent to the 'positive lists' in European legislation, included only substances that had been proven by scientific practice and evidenced to be innocuous to human beings. In the same year, the 'Food Additive Amendment, to the Federal Food, Drug and Cosmetic Act (FD&C Act) was passed, requiring the explicit authorization of an additive before it could be added to a food. It also obligates the manufacturer to prove that the additive is safe in all of the forms in which it will be used. This amendment excluded two groups of substances from the additive regulation process:

those that the FDA or the United States Department of Agriculture (USDA) had deemed safe for use in foods previous to the 1958 amendment and those on the GRAS lists. In 1960, the US Congress approved the 'Color Additive Amendments' to the FD&C Act, which state that all color additives to be used in foods, cosmetics, and drugs (including externally applied drugs and cosmetics) must be approved by the FDA before being sold on the market. Since 1958, the FDA and USDA have continued to monitor all prior sanctioned and GRAS substances in response to new scientific information and evidence on the safety of these substances.

**International organizations** Every country needs to have access to reliable evaluations of the risks associated with the chemical substances used in foods, but relatively few have the technical capability and the economic means to do individual evaluations on the risks posed by the huge number of existing additives. The work done by the Joint FAO/OMS Expert Committee on Food Additives (JECFA) and the Codex Committee on Food Additives and Contaminants (CCFAC) is extremely important in this regard. Both committees were jointly created by the World Health Organization (WHO) and the Food and Agriculture Organization of the United Nations (FAO) and are dedicated to studying additives with the objective of producing international regulations for them. The JECFA works as the scientific advisor to the FAO, the WHO, the member governments of the FAO, and the WHO and the Codex Alimentarius Commission. The JECFA establishes guidelines to evaluate the safety of additives, makes toxicological evaluations, determines the ADI of each of the additives, prepares their purity specifications, and evaluates the frequency of consumption of foods that contain additives. The CCFAC advises the Codex Alimentarius Commission on food additives, contaminants, and naturally present toxic substances; examines the purity characteristics and the identity of the additives; establishes the maximum level of contaminants in foods; and advises the Products Committee of the Codex Alimentarius in which foods an additive can be used and the maximum amounts that can be added to them.

**Labeling and identification of food additives** In addition to the strict criteria applied to the evaluation of risk and the obligation of having a demonstrable use, the regulations also require that additives appear on the labels of the packaging of the foods and drinks that contain them. These labels must explain the function of the additive and its name or assigned number.

The identification of food additives using a number is an alternative to stating the specific name of the additive, which is normally quite long and has a complex chemical structure. In 1960, the EU assigned conventional numbers to food additives to better identify them. It was decided that each substance would be identified with a number (of three or four digits) preceded by the letter E. It was thought that the additives could be separated into nine categories and that each category would have 100 numbers: the first 100 numbers was assigned to color additives (100–199); the second to preservatives (200–299); the third to antioxidants (300–399); and the fourth to emulsifiers, thickeners, and gelling agents (400–499). The initial letter E indicates that the additive has been certified by the SCF as safe for use in the entire EU. In the framework of the Codex Alimentarius, in 1989 the CCFAC created the International Numbering System (INS). This numbering system was created to reach an international agreement on the identification number of additives and establish a numerically ordered list of additives that included the number of each additive and its technological function. The European Commission and that of the Codex Alimentarius have agreed to adopt the same identification numbers, although the INS system does not include the E prefix, which is reserved for the EU.

### Criteria Used to Establish the Maximum Food Additive Dose

To establish the maximum food additive dose that can be used, international authorities consider two basic criteria: technology and toxicology. On the one hand, they assess the amount of the additive that is technologically required to satisfy the objective need to include the additive in the food; on the other hand, they evaluate the innocuousness of the additive.

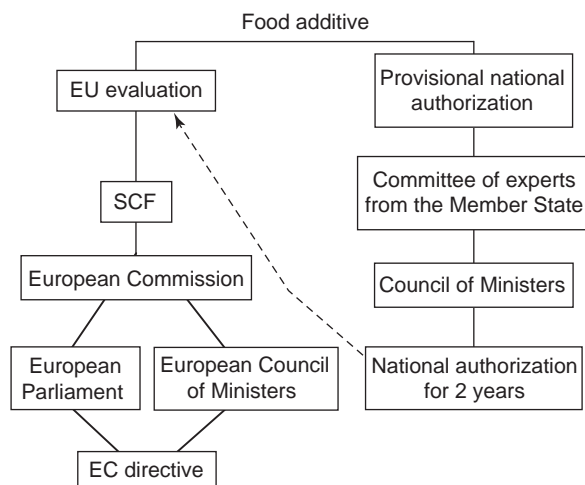
The technologically useful dose of the additive is that which accomplishes the desired effect on the food. The tests that determine the technologically useful dose must follow the guidelines established in the Good Manufacturing Practice used by the food industry.

The ADI is established by international organizations like the JECFA and the SCF after completing toxicological evaluations. The toxicological evaluation of additives is done using as reference the maximum dose that does not have a toxic effect on animals or the maximum dose that in short- and long-term toxicological studies has not had any significant toxic effects. The ADI for man is defined as the additive dose that, when consumed daily during a lifetime, does not cause a noticeable risk and is

expressed in milligrams of additive per kilogram of body weight per day.

### Legal Procedures for the Authorization of New Food Additives

The EU has established a legislative structure for the authorization of new additives proposed by either the EU or any of the individual Member States. This authorization requires a safety evaluation by the SCF, which evaluates the information provided by the manufacturers of the new food additive, including toxicity data and the function of the new additive. If the additive passes the safety standards, the committee establishes its ADI. The European Commission then develops legislation based on the committee's recommendations and then presents it to the European Council of Ministers and the European Parliament. The final result is published in a European Community Directive that obligates each Member State to make the necessary changes in their national legislation to include the new additive. Alternatively, the manufacturers can ask one of the Member States for a provisional national authorization while the EU deliberates on whether or not to legalize the additive. In this case, a committee of experts from the Member State evaluates the request, following procedures similar to those of the SCF. If the Member State establishes an ADI, the additive can be approved for use in that country for a period of 2 years, while the authorization request is studied in the EU. If the SCF authorizes the additive it is incorporated in a European Community Directive. If the SCF denies the authorization request, the additive must immediately be banned in all of the Member States. **Figure 1** summarizes the legislative procedures required to authorize additives in the EU.



**Figure 1** Legislative procedures to authorize additives in the EU.



**Table 2** European Community Directives (updated on December 2003)

<i>EC Directive</i>	<i>Content</i>
<i>General legislation</i>	
Directive 89/107/CEE	Additives that can be used in products destined for human consumption
<i>Sweeteners</i>	
Directive 94/35/CE	Sweeteners
Directive 95/31/CE	Criteria on the purity of sweeteners
<i>Color additives</i>	
Directive 94/36/CE	Color additives
Directive 95/45/CE	Criteria on the purity of color additives
Directive 99/75/CE	Modification of the Directive 95/45/CE
Directive 2001/50/CE	Modification of the Directive 95/45/CE
<i>Food additives other than colors or sweeteners</i>	
Directive 95/2/CE	On food additives other than colors or sweeteners
Directive 96/85/CE	Modification of the Directive 95/2/CE
Directive 98/72/CE	Modification of the Directive 95/2/CE
Directive 2001/5/CE	Modification of the Directive 95/2/CE
Directive 96/77/CE	Criteria on the purity of food additives other than colors or sweeteners
Directive 98/86/CE	Modification of the Directive 96/77/CE
Directive 2000/63/CE	Modification of the Directive 96/77/CE
Directive 2001/30/CE	Modification of the Directive 96/77/CE

In the US, a request to authorize a new food additive is presented to the FDA. A petition for an additive or a colorant must include data pertaining to the toxicity of the additive accompanied by toxicological tests on animals and even on humans. To evaluate whether an additive should be approved, the agency first considers the composition and attributes of the substance, the probable amount that will be ingested, the possible long-term effects on consumers and various additional safety factors. If the additive is approved, the FDA publishes regulations that may include the types of foods in which it can be used, the maximum quantities that can be added, and how it should be identified in the food labels.

### Food Additive Legislation

European legislation is compiled in European Community Directives that are published in the Bulletin of the EU. European Community Directives affect the entire EU, setting a time limit for their integration in the legislation of each Member State, after which they are required to be enforced in the entire EU. Table 2 is a catalog of European Community Directives, updated in December 2003, that regulate positive lists of additives, standards for the purity of additives, and the maximum amounts of additives that can be added to different categories of foods.

The US additive legislation is incorporated in the Code of Federal Regulations, Title 21, Chapter 1. The content of the US legislation is organized in parts, shown in Table 3.

**Table 3** US additive legislation

<i>Part</i>	<i>Content</i>
70–82	Regulation of color additives
170–173	General regulation of food additives
174–178	Regulation of indirect food additives
180	Regulation of food additives permitted in food or in contact with food on an interim basis pending additional study
182–186	Regulation of the substances affirmed as generally recognized as safe
189	Regulation of the substances prohibited from use in human food

### Analytical Methods for the Determination of Food Additives

The development of analytical methods for the determination of food additives responds to the need to check the stability of the additives during processing and storage or, primarily, to legal requirements such as: (1) establishing whether there are additives present in foods and whether they are permitted, (2) determining their concentration to ensure that they are within the legally established limits, and (3) confirming the absence of illegal additives in foods.

#### Sample Pretreatment: Extraction and Cleanup

The pretreatment of foods before additive determination varies greatly depending on the type of sample (liquid or solid foods, samples with high content of lipids, proteins or carbohydrates) or the analytical

technique employed for the determination of the additive. Often, very simple samples do not require prior treatment or at the most need uncomplicated pretreatments, but most foods need to be treated before the determination of the additive. Frequently, the first step is to use liquid-liquid extraction (LLE) to eliminate interferences. LLE can be simple or repetitive (with the same solvent and then combining all the extracts). The extractant and the extraction method used depend fundamentally on the kind of sample being analyzed. The extraction can be done with a single solvent, principally *n*-hexane, water or acidified water, acetonitrile, dichloromethane, ethanol or methanol, diethyl or petroleum ether, acetone, and chloroform, or with a binary or even ternary mixture of the solvents, containing a very polar solvent, like water or methanol, and other less polar solvents.

However, other matrix compounds are extracted together with the analytes, making it necessary to include an extract-cleanup step. The thoroughness of this step depends on the detection system: if the detector being used is selective enough, an exceedingly thorough cleanup of the extract is not necessary; but if the detector is not very selective, this step must eliminate as many as possible of the coextracts that can interfere in the determination.

Moreover, because the additive concentration in this extract is not sufficient for the detection of legal limits, it is necessary to preconcentrate the analytes. LLE is commonly used to separate additives from their coextracts. Nevertheless, the recent trend is to replace LLE with solid-phase extraction (SPE). It is possible to use SPE as a single cleanup/preconcentration step or as an additional step after LLE. The use of SPE has become popular because of the great variety of polar and nonpolar sorbents and ionic exchange resins that exist. The sorbents most often used to preconcentrate food additives are: RP-C<sub>18</sub>, silica gel, sorbents with amino groups, XAD-2 polymeric sorbent, polyamide and wool.

Supercritical fluid extraction is another technique used to extract food additives. Its main advantages are that it saves both time and solvents; the solvent most often employed is carbon dioxide, for its moderate critical temperature (31°C) and pressure (73 atm). In addition, carbon dioxide is not inflammable, toxic, or very expensive. Methanol is sometimes used as a modifier to adjust the dissolution power of this fluid.

Quite recently, other extraction techniques have been developed and are being employed for the determination of food additives, but are not yet widely used. These techniques include countercurrent chromatography, gel permeation chromatography,

distillation, membrane separation, dialysis and microdialysis, microwave-assisted extraction, solvent extraction assisted by ultrasound, solvent microextraction, solid-phase microextraction, and stir bar sorptive extraction.

It is important to emphasize that specific difficulties arise when extracting antioxidants and preservatives from fatty foods. Most of the permitted antioxidants and preservatives are fat soluble, so it may first be necessary to extract the fat from the food and then extract the antioxidants from the fat. As a result, the determination of these additives implies the removal of triglycerides; the techniques most frequently used to remove triglycerides are saponification of the extracts, esterification of the extracts in the presence of lipase, and the separation of the analytes by preparative liquid chromatography.

### **Determination, Separation, and Detection of Food Additives**

For the determination of food additives several analytical methods are used based on spectroscopic, electrochemical, and separation techniques, the latter with detectors included.

For the determination of individual additives several spectroscopic techniques are normally employed: ultraviolet-visible spectrophotometry, spectrofluorimetry, luminescence, and photoacoustic spectrometry. Sometimes, to increase method selectivity, a combination of spectroscopic techniques is used. Another alternative to determine individual additives are electrochemical techniques: voltammetry, polarography, amperometry, and potentiometry.

However, additives are normally combined to complement and promote their activity; as a result, it is necessary to develop analytical methods for the determination of additive mixtures. Although some spectroscopic and chemical methods are used, it is preferable to use separation methods for this purpose. Most analytical methods used to determine food additives are based on chromatographic techniques, although several recent papers have demonstrated the usefulness of electrophoresis for the analysis of food colors, sweeteners, antioxidants, and/or preservatives. The separation of food colors has received most attention, with a number of articles published on both capillary zone electrophoresis and micellar electrokinetic chromatography.

Thin-layer chromatography, in its normal and high resolution forms, is used for the separation and quantification of additives with minimal sample manipulation. The most common detection technique



is ultraviolet–visible spectrophotometry, although other techniques such as spectrofluorimetry and mass spectrometry are also used.

Liquid chromatography is the preferred technique to determine food additives, which in general are not very volatile and sometimes are thermolabile.

**Table 4** Official AOAC methods of analysis applicable to antioxidants

<i>AOAC method</i>	<i>Antioxidant</i>	<i>Food</i>	<i>Analytical technique</i>
948.26	$\alpha$ -Tocopheryl acetate	Foods	Colorimetry
942.11	Quaternary ammonium compounds	Milk	Qualitative test
942.13	Quaternary ammonium compounds	Foods, beverages	Bromophenol blue method
952.09	Propyl gallate	Foods	Colorimetry
952.10	Quaternary ammonium compounds	Milk	Optical crystallographic properties of the Reineckates
954.06	Quaternary ammonium compounds	Milk	Eosin yellowish method
965.28	Antioxidants	Foods	Qualitative test
967.21	Ascorbic acid	Juices	Titrimetry
968.17	Butylated hydroxyanisole, butylated hydroxytoluene	Cereals	Gas chromatography
971.30	$\alpha$ -Tocopherol, $\alpha$ -tocopheryl acetate	Foods	Colorimetry
975.43	$\alpha$ -Tocopherol	Foods	Polarimetry
983.15	Antioxidants	Oils, fats	Liquid chromatography
984.26	Ascorbic acid	Foods	Semiautomated method – fluorimetry

**Table 5** Official AOAC methods of analysis applicable to preservatives

<i>AOAC method</i>	<i>Preservative</i>	<i>Food</i>	<i>Analytical technique</i>
892.02	Sulfurous acid free	Meat	Titrimetry
910.02	Benzoic acid	Foods	Qualitative test
931.08	Formaldehyde	Foods	Qualitative test
935.34	Benzoic acid	Flour	Photometry
950.24	Benzoate, caffeine, saccharin	Soda beverages	Liquid chromatography
953.12	Dehydroacetic acid	Cheese	Qualitative test
953.13	Dehydroacetic acid	Cheese	Spectrophotometry
959.09	Boric acid	Meat	Semiquantitative test
960.27	Preservatives	Milk	Qualitative test
960.38	Benzoic acid	Non-solid foods, beverages	Spectrophotometry
961.09	Sulfites	Meat	Qualitative test
962.16	Sulfurous acid total	Foods	Modified Monier–Williams method
963.19	Benzoic acid	Foods	Titrimetry
963.20	Sulfurous acid total	Dried fruit	Colorimetry
964.13	Nitrites	Curing preparations	Titrimetry
967.15	Benzoic acid	Foods	Thin-layer chromatography
968.16	Fumaric acid	Foods	Polarography
969.26	Boric acid	Caviar	Spectrophotometry
970.33	Boric acid, borates	Foods	Qualitative test
970.34	Boric acid	Foods	Titrimetry
971.15	Sorbic acid	Cheeses	Oxidation method
972.19	Boric acid	Foods	Atomic absorption spectrometry
974.08	Sorbic acid	Wines	Spectrophotometry
974.10	Sorbic acid	Dairy products	Spectrophotometry
975.10	Sorbic acid	Wines	Colorimetry
975.26	Boric acid	Foods	Emission spectroscopy
975.30	Salicylic acid	Foods, beverages	Qualitative test
975.31	Salicylic acid	Foods, beverages	Colorimetry
975.32	Sulfurous acid	Foods	Qualitative test
980.17	Preservatives	Ground beef	Spectrophotometry
983.16	Benzoic acid, sorbic acid	Foods	Gas chromatography
987.04	Sulfites total	Foods	Differential pulse polarography
990.28	Sulfites	Foods	Optimized Monier–Williams method
990.29	Sulfites total	Foods, beverages	Flow injection method – spectrophotometry
990.30	Sulfites free	Wines	Flow injection method – spectrophotometry
990.31	Sulfites	Foods, beverage	Ion exclusion chromatography
993.03	Nitrates	Baby foods	Spectrophotometry

In most of the methods that have been developed, the stationary phase used for the additive separation is a polygosyl-bonded silica reversed-phase sorbent with octadecylsilane functional groups (RP-C<sub>18</sub>), although other stationary phases such as reverse RP-C<sub>30</sub> and ionic-exchange phases have also been used. The detection of additives, after their separation by liquid

chromatography, has been done using ultraviolet-visible, fluorescence, electrochemical, and mass spectrometry detectors.

Gas chromatography, with or without derivatization, is also used to selectively determine individual or mixtures of food additives. The additive separation is done using nonpolar stationary phases consisting

**Table 6** Official AOAC methods of analysis applicable to sweeteners

<i>AOAC method</i>	<i>Sweetener</i>	<i>Food</i>	<i>Analytical technique</i>
934.04	Saccharin	Nonalcoholic beverages	Semiquantitative method
941.10	Saccharin	Foods	Qualitative test
947.10	Saccharin	Foods	Sublimation
950.24	Benzoate, caffeine, saccharin	Soda beverages	Liquid chromatography
957.09	Cyclamate salts	Nonalcoholic beverages	Sodium nitrite test
957.10	Cyclamate salts	Nonalcoholic beverages	Gravimetry
957.11	Dulcin	Fruits	Qualitative test – spectrophotometry
969.27	Nonnutritive sweeteners	Nonalcoholic beverages	Thin-layer chromatography
969.28	Cyclamate salts	Canned fruits	Colorimetry
973.28	Sorbitol	Foods	Gas chromatography
973.29	Saccharin	Foods	Gravimetry
980.18	Saccharin	Foods	Differential pulse polarography

**Table 7** Official AOAC methods of analysis applicable to color additives

<i>AOAC method</i>	<i>Color additive</i>	<i>Food</i>	<i>Analytical technique</i>
900.01	Color additives	Milk	Qualitative test
920.114	Color additives	Cream	Qualitative test
920.119	Color additives	Butter fat	Qualitative test
920.209	Natural color additives	Foods	Qualitative test
925.28	Color additives	Ice cream, frozen desserts	Qualitative test
930.17	Artificial color additives	Distilled liquors	Qualitative test
930.38	Synthetic organic color additives	Foods	Column chromatography
938.04	Carotenoids	Macaroni products	Colorimetry
942.19	Synthetic color additives	Oils, fats	Spectrophotometry
948.07	Caramel	Wines	Qualitative test
955.18	Natural and artificial organic and water-soluble color additives	Distilled liquors	Spectrophotometry
966.21	Oil-soluble color additives	Foods	Column chromatography
970.65	Riboflavin	Foods	Fluorimetry
981.15	Riboflavin	Foods	Flow injection method – fluorimetry
985.31	Riboflavin	Ready-to-feed milk-based infant formula	Fluorimetry
988.13	FD&C color additives	Foods	Spectrophotometry – thin-layer chromatography

**Table 8** Official AOAC methods of analysis applicable to emulsifying agents

<i>AOAC method</i>	<i>Emulsifying agent</i>	<i>Food</i>	<i>Analytical technique</i>
920.106	Gelatin	Milk, milk products	Qualitative test
935.61	Gums	Salad dressings	Qualitative test
945.51	Gelatin	Ice cream, frozen desserts	Qualitative test
959.06	Alginates	Chocolate products	Qualitative test
960.33	Gums	Ice creams, frozen desserts	Infrared spectroscopy
960.34	Alginates	Chocolate frozen desserts	Qualitative test
963.25	Alginates	Food dressings	Colorimetry
968.18	Sodium lauryl sulfate	Egg white	Colorimetry
974.11	Polysorbate 60	Shortening, oils, dressings	Gravimetry
991.30	Polydimethylsiloxane	Pineapple juice	Atomic absorption spectrometry

of 100% poly(dimethylsiloxane), intermediate polarity phases consisting of 5% diphenyl–95% dimethylsiloxane or 50% diphenyl–50% dimethylsiloxane, and polar phases consisting of 100% poly(ethyleneglycol). The detection of these analytes is usually done using a flame ionization detector, which is able to detect compounds with  $-CH_3$  groups, or universal detectors like a mass spectrometry detector.

### Identification of Food Additives

When analyzing food additives it is frequently necessary to identify the additives that have been determined quantitatively using the above cited detectors that, while supplying high sensitivity (essential for measuring the established legal limits), are not able to identify these compounds. To confirm the obtained results two strategies are used: on the one hand, the use of detectors that are able to unmistakably identify the analytes; on the other hand, the use of two different detectors to compare their responses against the same mixture of additives.

Mass spectrometry is an excellent tool that can be used coupled to either a liquid or a gas chromatograph to confirm the identity of the additives present in foods. A complete mass spectrum of an additive provides structural information that is unique to the additive. Most of the studies that have been done use the ionization technique for electronic impact to a certain energy (normally 70 eV).

Detection by a diode array detector can be used to confirm the identity of an additive. This detector is usually coupled with a liquid chromatograph to record the ultraviolet–visible spectra for each of the chromatographic peaks so that they can be compared with the spectra of the pure compounds. Moreover, the purity of each of the chromatographic peaks can be checked.

The use of two connected detectors, in succession or in parallel, is widely employed, coupling them to either a liquid or gas chromatograph.

The Association of Official Analytical Chemists (AOAC) publishes reference methods for the analysis of additives in foods. The following **Tables 4–8** summarize the AOAC Official Methods of Analysis applicable to antioxidants (**Table 4**), preservatives (**Table 5**), sweeteners (**Table 6**), color additives (**Table 7**), and emulsifying agents (**Table 8**).

*See also:* **Food and Nutritional Analysis:** Overview; Antioxidants and Preservatives. **Gas Chromatography:** Overview. **Liquid Chromatography:** Instrumentation. **Sweeteners.**

### Further Reading

- Abidi SL (2000) Chromatographic analysis of tocol derived lipid antioxidants. *Journal of Chromatography A* 881: 197–216.
- Belitz HD and Grosch W (eds.) (1999) *Food Chemistry*, 2nd edn. Berlin: Springer.
- Boyce MC (2001) Determination of additives in foods by capillary electrophoresis. *Electrophoresis* 22: 1447–1459.
- Burdock GA (ed.) (1996) *Encyclopedia of Food and Color Additives*. Boca Raton, FL: CRC Press.
- Clydesdale FM (ed.) (1996) *Food additives: Toxicology, Regulation and Properties*. Boca Raton, FL: CRC Press.
- Code of Federal Regulations Title 21, Chapter 1, Parts 70–82, 170–189. Washington, DC: US Government Printing Office, 2003; <http://www.access.gpo.gov/cgi-bin/cfrassemble.cgi?title=200321>.
- Horwitz W (ed.) (2000) *Official Methods of Analysis of the Association of Official Analytical Chemists*, 17th edn. Arlington: AOAC International.
- Karovicová J and Simko P (2000) Determination of synthetic phenolic antioxidants in food by high performance liquid chromatography. *Journal of Chromatography A* 882: 271–281.
- Macrae R, Robinson RK, and Sadler MJ (eds.) (1993) *Encyclopaedia of Food Science, Food Technology and Nutrition*, vols. I and II. San Diego: Academic Press.
- Pokorny J, Yanishlieva N, and Gordon MH (eds.) (2001) *Antioxidants in Food: Practical Applications*. Boca Raton, FL: CRC Press.

## Antioxidants and Preservatives

**M H Gordon**, The University of Reading, Reading, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Antioxidants and preservatives are added to foods to retard their deterioration. Antioxidants retard

chemical oxidation of the lipids whereas preservatives retard microbiological deterioration. In the UK, the permitted antioxidants and preservatives are listed in the Miscellaneous Food Additives Regulations (1995). Common allowed antioxidants and preservatives are listed in **Tables 1** and **2**. The allowed additives vary from country to country with

of 100% poly(dimethylsiloxane), intermediate polarity phases consisting of 5% diphenyl–95% dimethylsiloxane or 50% diphenyl–50% dimethylsiloxane, and polar phases consisting of 100% poly(ethyleneglycol). The detection of these analytes is usually done using a flame ionization detector, which is able to detect compounds with  $-CH_3$  groups, or universal detectors like a mass spectrometry detector.

### Identification of Food Additives

When analyzing food additives it is frequently necessary to identify the additives that have been determined quantitatively using the above cited detectors that, while supplying high sensitivity (essential for measuring the established legal limits), are not able to identify these compounds. To confirm the obtained results two strategies are used: on the one hand, the use of detectors that are able to unmistakably identify the analytes; on the other hand, the use of two different detectors to compare their responses against the same mixture of additives.

Mass spectrometry is an excellent tool that can be used coupled to either a liquid or a gas chromatograph to confirm the identity of the additives present in foods. A complete mass spectrum of an additive provides structural information that is unique to the additive. Most of the studies that have been done use the ionization technique for electronic impact to a certain energy (normally 70 eV).

Detection by a diode array detector can be used to confirm the identity of an additive. This detector is usually coupled with a liquid chromatograph to record the ultraviolet–visible spectra for each of the chromatographic peaks so that they can be compared with the spectra of the pure compounds. Moreover, the purity of each of the chromatographic peaks can be checked.

The use of two connected detectors, in succession or in parallel, is widely employed, coupling them to either a liquid or gas chromatograph.

The Association of Official Analytical Chemists (AOAC) publishes reference methods for the analysis of additives in foods. The following **Tables 4–8** summarize the AOAC Official Methods of Analysis applicable to antioxidants (**Table 4**), preservatives (**Table 5**), sweeteners (**Table 6**), color additives (**Table 7**), and emulsifying agents (**Table 8**).

*See also:* **Food and Nutritional Analysis:** Overview; Antioxidants and Preservatives. **Gas Chromatography:** Overview. **Liquid Chromatography:** Instrumentation. **Sweeteners.**

### Further Reading

- Abidi SL (2000) Chromatographic analysis of tocol derived lipid antioxidants. *Journal of Chromatography A* 881: 197–216.
- Belitz HD and Grosch W (eds.) (1999) *Food Chemistry*, 2nd edn. Berlin: Springer.
- Boyce MC (2001) Determination of additives in foods by capillary electrophoresis. *Electrophoresis* 22: 1447–1459.
- Burdock GA (ed.) (1996) *Encyclopedia of Food and Color Additives*. Boca Raton, FL: CRC Press.
- Clydesdale FM (ed.) (1996) *Food additives: Toxicology, Regulation and Properties*. Boca Raton, FL: CRC Press.
- Code of Federal Regulations Title 21, Chapter 1, Parts 70–82, 170–189. Washington, DC: US Government Printing Office, 2003; <http://www.access.gpo.gov/cgi-bin/cfrassemble.cgi?title=200321>.
- Horwitz W (ed.) (2000) *Official Methods of Analysis of the Association of Official Analytical Chemists*, 17th edn. Arlington: AOAC International.
- Karovicová J and Simko P (2000) Determination of synthetic phenolic antioxidants in food by high performance liquid chromatography. *Journal of Chromatography A* 882: 271–281.
- Macrae R, Robinson RK, and Sadler MJ (eds.) (1993) *Encyclopaedia of Food Science, Food Technology and Nutrition*, vols. I and II. San Diego: Academic Press.
- Pokorny J, Yanishlieva N, and Gordon MH (eds.) (2001) *Antioxidants in Food: Practical Applications*. Boca Raton, FL: CRC Press.

## Antioxidants and Preservatives

**M H Gordon**, The University of Reading, Reading, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Antioxidants and preservatives are added to foods to retard their deterioration. Antioxidants retard

chemical oxidation of the lipids whereas preservatives retard microbiological deterioration. In the UK, the permitted antioxidants and preservatives are listed in the Miscellaneous Food Additives Regulations (1995). Common allowed antioxidants and preservatives are listed in **Tables 1** and **2**. The allowed additives vary from country to country with

**Table 1** Permitted antioxidants in the UK

Serial number	Name
E 300	L-ascorbic acid
E 301	Sodium L-ascorbate
E 302	Calcium L-ascorbate
E 304	6-O-Palmitoyl-L-ascorbic acid (ascorbyl palmitate)
E 306	Extracts of natural origin rich in tocopherols
E 307	Synthetic $\alpha$ -tocopherol
E 308	Synthetic $\gamma$ -tocopherol
E 309	Synthetic $\delta$ -tocopherol
E 310	Propyl gallate (propyl 3,4,5-trihydroxybenzoate)
E 311	Octyl gallate (octyl 3,4,5-trihydroxybenzoate)
E 312	Dodecyl gallate (dodecyl 3,4,5-trihydroxybenzoate)
E 315	Erythorbic acid
E 316	Sodium erythorbate
E 320	Butylated hydroxyanisole (BHA)
E 321	Butylated hydroxytoluene (BHT) (2,6-di- <i>tert</i> -butyl- <i>p</i> -cresol)

Miscellaneous Food Additives Regulations 1995.

**Table 2** Permitted preservatives in the UK

Serial number	Name
E 200-3	Sorbic acid (plus sodium, potassium, calcium salts)
E 210-3	Benzoic acid (plus sodium, potassium, calcium salts)
E 214,5	Ethyl 4-hydroxybenzoate (plus sodium salt)
E 216,7	Propyl 4-hydroxybenzoate (plus sodium salt)
E 218,9	Methyl 4-hydroxybenzoate (plus sodium salt)
E 220	Sulfur dioxide (plus sodium sulfite, sodium hydrogensulfite, sodium 'metabisulfite', potassium metabisulfite, calcium sulfite, calcium hydrogensulfite)
E 228	Potassium hydrogensulfite
E 230	Biphenyl
E 231	2-Hydroxybiphenyl
E 232	Sodium biphenyl-2-yl oxide
E 233	Thiabendazole (2-(thiazol-4-yl)-benzimidazole)
E 239	Hexamine
E 249	Potassium nitrite
E 250	Sodium nitrite
E 251	Sodium nitrate
E 252	Potassium nitrate
E 280-3	Propionic acid (plus sodium, calcium, potassium salts)
E 234	Nisin

Miscellaneous Food Additives Regulations 1995.

*tert*-butylhydroquinone (TBHQ) and 2,4,5-trihydroxy-butyrophenone (THBP) being allowed antioxidants in the USA.

The Association of Official Analytical Chemists (AOAC) describes several methods for the analysis of antioxidants or preservatives. These include methods developed over many years. Most methods involve analysis by gas chromatography (GC), high-performance liquid chromatography, or spectrophotometry.

Capillary electrophoresis has been shown to be a very versatile technique for the analysis of a wide range of additives including both antioxidants and preservatives. Methods employing mixed micellar electrokinetic chromatography have been reported. However, the AOAC has not yet developed an official method based on this technique.

## Antioxidants

Antioxidants are mainly analyzed by colorimetric methods, liquid chromatography (LC), or GC. Method 965.28 of the AOAC describes qualitative color tests for detecting propyl gallate (PG), nordihydroguaiaretic acid (NDGA), butylated hydroxyanisole (BHA), and butylated hydroxytoluene (BHT). After solvent and aqueous extractions PG forms a pink color with ammonia solution, NDGA forms a blue color with barium hydroxide, BHA forms a red-purple color with Ehrlich reagent (diazobenzene sulfonic acid), whereas BHT is separated by column chromatography and develops a chloroform-soluble pink to red color in the presence of dianisidine (3,3'-dimethoxybenzidine) and sodium nitrite.

PG may also be quantified by a colorimetric procedure in which quantification depends on light absorption at 540 nm, after color development by reaction with ammonium acetate and iron(II) tartrate (AOAC method 952.09).

A reversed-phase LC method can be used for the quantitative analysis of mixtures containing PG, 2,4,5-tri-hydroxybutyphenone (THBP), *tert*-butylhydroquinone (TBHQ), NDGA, BHA, BHT, and 2,6-di-*tert*-butyl-4-hydroxymethylphenol (Ionox-100), as described in AOAC method 983.15. Baseline separation of the seven antioxidants is achieved using a reversed-phase column and stepwise gradient of 30% acetonitrile–acetic acid (95:5; v/v), 70% water–acetic acid (95:5; v/v) changing to acetonitrile–acetic acid (95:5; v/v). Detection of the antioxidants is by ultraviolet (UV) absorption at 280 nm. Octyl gallate, if present, may coelute with Ionox-100, but can be separated by a linear gradient of 30% methanol–acetic acid (95:5; v/v), 70% water–acetic acid (95:5; v/v) changing to methanol–acetic acid (95:5; v/v). Page and Charbonneau determined the antioxidants in a variety of foods by a method similar to this.

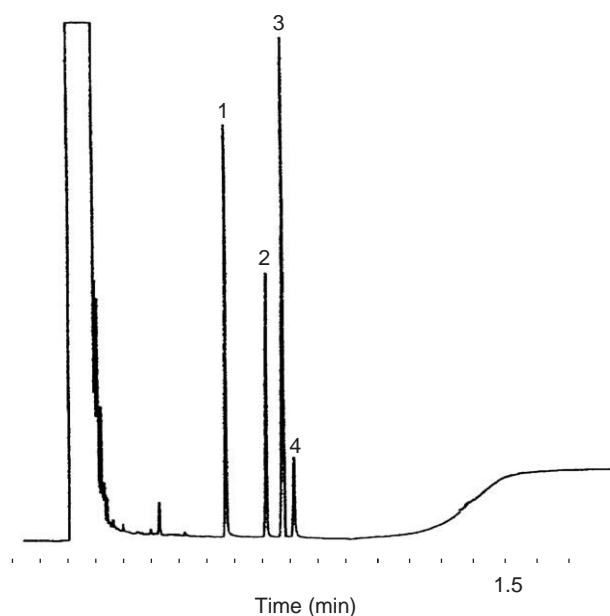
BHA and BHT are commonly determined by GC. AOAC method 968.17 describes a procedure for the analysis of these antioxidants in cereals. The antioxidants are extracted with carbon disulfide and analyzed by GC using packed glass columns containing Apiezon L or QF-1 fluorosilicone fluid as



stationary phases. The order of elution from the Apiezon column is BHA, BHT, di-BHA, whereas the order of elution from the QF-1 column is BHT, BHA, di-BHA. Analysis is performed with the oven at 160°C, with BHT eluting in 3–4 min from the QF-1 column and BHA eluting in 3–4 min from the Apiezon column. The analytes are detected with a flame ionization detector.

Gas chromatographic procedures have been described for the determination of several antioxidants. Glass columns packed with OV-17 as stationary phase are able to separate TBHQ, TDPA, PG, THBP, and NDGA. Oil samples containing antioxidants can either be injected onto a precolumn, which prevents the nonvolatile oil entering the analytical column, or alternatively the antioxidants can be isolated by steam distillation or liquid–liquid partition prior to GC analysis. Solid-phase extraction cartridges have been used to remove traces of triacylglycerols remaining after liquid–liquid extraction prior to GC analysis.

Although the standard methods describe the use of packed columns, capillary columns are widely used for the analysis since they improve the resolution of sample components, and reduce inaccuracies in the quantification of antioxidants arising from interference of other oil components. Figure 1 shows the resolution of antioxidants by capillary GC.



**Figure 1** Gas chromatogram of antioxidants. Peaks: 1, 8-quinolinol; 2, BHA; 3, BHT; 4, TBHQ. (Reproduced with permission from Yang M-H, Lin H-J, and Choong Y-M (2002) A rapid gas chromatographic method for direct determination of BHA, BHT and TBHQ in edible oils and fats. *Food Research International* 35: 627–633; © Elsevier.)

Besides the GC analysis of underivatized antioxidants, antioxidants have also been converted to trimethylsilyl ethers, benzoyl esters, trifluoroacetates, and heptafluorobutyl derivatives prior to GC analysis.

Vitamin C (ascorbic acid) and its fat-soluble derivative ascorbyl palmitate also have antioxidant activity under certain conditions. The determination of vitamin C is discussed elsewhere.

## Preservatives

Preservatives have been analyzed by a wide variety of methods including titration, thin-layer chromatography (TLC), GC, LC, and spectrophotometric methods.

### Benzoic Acid

Benzoic acid and its salts may be determined by titration with sodium hydroxide after extraction of the benzoic acid from an aqueous food suspension into chloroform, and evaporation of the chloroform and any acetic acid present (AOAC method 963.19). Vanillin interferes with this determination and a more selective method involves the determination of benzoic acid in an ether extract by UV absorption at 272 nm, as described in AOAC method 960.38. An alternative method of isolating benzoic acid from food involves the use of steam distillation and TLC separation. These sample preparation techniques are used in AOAC method 967.15 prior to the determination of benzoic acid by UV absorption.

A gas chromatographic method may be employed for analysis for both benzoic acid and sorbic acid in food, as described in AOAC method 983.16. The acids are extracted from an acidified aqueous homogenized food sample into ether. The acids are then extracted from the ether solution into aqueous alkali, and after acidification they are re-extracted into dichloromethane. Treatment of the residue with *N*-methyl-*N*-trimethylsilyltrifluoroacetamide converts the benzoic and sorbic acids into trimethylsilyl esters. These esters are then analyzed with a glass column packed with a support coated with OV-1, in a temperature programmed analysis. Phenylacetic and caproic acids are used as internal standards for benzoic and sorbic acids, respectively. An LC method has also been described for analysis of benzoic acid with simultaneous determination of sorbic acid, and methyl, ethyl, propyl, and butyl *p*-hydroxybenzoates. The preservatives were extracted from meat or seasonings with 70% ethanol, and analyzed by reversed-phase LC using a C<sub>18</sub> stationary phase with a linear gradient from 10 to 70% methanol in 1.5% aqueous ammonium acetate and 1.5% aqueous acetic acid.



Good baseline separation of the analytes was achieved in less than 13 min.

The International Dairy Federation has adopted an LC method as a standard method for the determination of the benzoic acid and sorbic acid content of milk, dried milk, yogurt, and other fermented milks (IDF standard 139: 1987). Fats and proteins are removed initially by treatment of the product with aqueous alkali, addition of methanol, and filtration prior to LC analysis.

Ion chromatography has also been described for the analysis of benzoic acid and sorbic acid, together with artificial sweeteners, caffeine, theobromine, and theophylline. The preservatives were analyzed on an anion-exchange analytical column operated at 40°C.

### Sorbic Acid

Sorbic acid may be isolated from cheese by steam distillation from an acidified sample. Oxidation of the sorbic acid by heating in a boiling water bath with acidified dichromate solution, followed by treatment with thiobarbituric acid solution, converts the analyte to a colored product absorbing at 532 nm (AOAC method 971.15). Alternatively, sorbic acid may be determined in dairy products by absorption at 250 nm after treatment of a sample with metaphosphoric acid and extraction into ether, as described in AOAC method 974.10.

### Sulfur Dioxide

Over the pH range of foods (pH 2–7), sulfur dioxide is found as one or more of the species  $\text{SO}_2 \cdot \text{H}_2\text{O}$ ,  $\text{HSO}_3^-$ ,  $\text{S}_2\text{O}_5^{2-}$ , and  $\text{SO}_3^{2-}$ . Acidification is required to allow all sulfur(IV) oxoanions to be determined as sulfur dioxide.

The simplest method available for the determination of sulfur dioxide in foods is iodimetric titration in the presence of starch after addition of acetone, glyoxal, or formaldehyde. Acidification before the titration allows the determination of both free and reversibly bound sulfur dioxide. Amperometric methods for detecting the endpoint allow the method to be extended to colored samples.

A long-established procedure for the determination of sulfur dioxide is the Monier–Williams method in which sulfur dioxide is distilled from an acidified solution and the vapor passes through U-tubes containing hydrogen peroxide which oxidizes it to sulfuric acid. The acidity produced is then titrated with dilute standard alkali. This procedure is preferred to an iodimetric titration when volatile sulfur compounds are present as in mustard, onion, and horseradish. Sulfur dioxide may also be

determined spectrophotometrically by methods based on the reaction between *p*-rosaniline, formaldehyde, and sulfur(IV) oxoanions. This reaction is employed in AOAC method 963.20 for sulfurous acid (total) in dried fruit. The sulfur(IV) oxoanions are extracted into water and treated with sodium tetrachloromercurate solution, before addition of acid-bleached *p*-rosaniline hydrochloride and formaldehyde solution. The absorbance at 550 nm is determined and quantification is achieved by an external standard calibration plot.

Sulfur dioxide in shrimps, orange juice, peas, dried apricots, and dehydrated potatoes may be determined by differential pulse polarography, as described in AOAC method 987.04. The sulfur dioxide is purged with nitrogen from an acidified sample, and collected in electrolyte trapping solution comprising ammonium acetate buffer prior to analysis by differential pulse polarography.

Ion-selective electrodes using a sulfur dioxide gas-sensing membrane probe have been developed into commercial models for the determination of sulfur(IV) oxo-species. A typical procedure for the determination of free sulfur dioxide requires the probe to be immersed in a stirred solution at pH 1, whereas total sulfur dioxide requires the sample to be treated with strong alkali prior to acidification and analysis at pH 1. Molecular sulfur dioxide passes through the membrane into a solution of hydrogen sulfite ion and the pH of the  $\text{SO}_2 \cdot \text{H}_2\text{O}/\text{HSO}_3^-$  buffer can be determined with a glass electrode.

Sulfur dioxide may be determined after removing it from foods by distillation, and titration with sodium hydroxide in the presence of methyl red. GC headspace analysis with a flame photometric detector has also been used for the quantification of this food additive.

LC methods have been developed for the determination of sulfur dioxide. In one procedure, the sulfur dioxide was collected by distillation from strongly acidic solutions of food samples and oxidized to sulfate with hydrogen peroxide. The sulfate solutions were neutralized, diluted, and analyzed by LC with a strong anion-exchange column, and a mobile phase of potassium hydrogenphthalate solution at pH 5.7. UV detection at 280 nm gave a good detection limit of  $\sim 5 \text{ mg kg}^{-1}$ .

Ascorbates, benzoates, sorbates, and sulfites in ground beef may be determined spectrophotometrically in AOAC method 980.17 by extracting the preservatives into aqueous solutions. The ascorbates are determined by the indophenol reaction; benzoates and sorbates are determined by absorption at 225 and 250 nm, respectively, in light petroleum solution; and sulfites are determined at 550 nm after

reaction with *p*-rosaniline in the presence of sodium tetrachloromercurate.

### Nitrates and Nitrites

Nitrates and nitrites occur ubiquitously in nature and are added to some foods, for example, meat products, as preservatives. They inhibit the growth of microorganisms, especially *Clostridia* in canned-comminuted and cured meats.

Nitrite may be determined by its absorbance at 522 nm after extraction from foods such as cheese or flour and reaction with a color reagent comprising sulfanilic acid and 1-naphthylamine in aqueous acetic acid. Nitrate is determined by this method after reduction to nitrites using cadmium in a modified Jones Reductor as described in AOAC method 976.14.

Nitrites and nitrates in meat may be determined by oxidizing nitrite to nitrate with potassium permanganate. The nitrate is then reacted with *m*-xylenol in tungstophosphoric acid. Silver-ammonia solution is added to precipitate all the chloride and excess tungstophosphoric acid present. The nitroxylenol is distilled and determined by UV absorption at 450 nm.

Nitrate may also be determined by LC with an anion-exchange column. A comparison has been made between the traditional method of nitrate determination using a reducing cadmium column and spectrophotometric determination with a reversed-phase LC method with orthophosphoric acid adjusted to pH 3.5 with sodium hydroxide as the mobile phase. A high correlation was observed between the nitrate content determined by the two methods. However, LC was found to be more precise, reproducible, and appropriate for routine work.

Various GC methods have also been developed for the determination of nitrate, but the ions must be derivatized to achieve suitable volatility before GC analysis. Nitrate is commonly determined as nitrobenzene, which can be detected with an electron-capture detector or a thermal energy analyzer at levels of 0.05 or 0.1  $\mu\text{g ml}^{-1}$ , respectively.

Nitrite can be oxidized to nitrate and determined as nitrobenzene, but several alternative derivatization techniques have been developed. Nitrate can be converted into tetrazolophthalazine or the trimethylsilyl derivative of benzotriazole, or the ion can be derivatized with 3,4-dichlorobromobenzene or *p*-bromochlorobenzene by the Sandmeyer reaction. An elegant one-step derivatization procedure involves the reaction of pentafluorobenzyl bromide with nitrite in alkali at 50°C for 90 min. This procedure allows analysis with an electron-capture

detector, and good agreement between this method and spectrophotometric methods has been found.

### Biphenyl, 2-Hydroxybiphenyl, and Thiabendazole

Biphenyl, 2-hydroxybiphenyl, and thiabendazole are used almost exclusively to prevent fungal growth on fruits, usually citrus fruits, after harvesting. Different countries have set varying limits as to the residual levels of these preservatives, which may carry through into the food chain.

Biphenyl is usually determined by chromatographic analysis of an extract prepared using an organic solvent. In the TLC method, biphenyl is located under UV radiation, extracted with methanol, and determined spectrophotometrically at 248 nm. If the silica gel layer is impregnated with the fluorescent indicator GF254, biphenyl can be detected by densitometry. LC analysis with UV detection at 254 nm and fluorimetric detection at 350 nm with excitation at 285 nm are good alternatives. Derivative infrared spectrometry has also been developed as an alternative procedure.

Steam distillation of biphenyl and 2-hydroxybiphenyl has been used to isolate the preservatives prior to analysis by GC. Simultaneous analysis of biphenyl, 2-hydroxybiphenyl, and thiabendazole in an organic extract from homogenized citrus fruits has been achieved with a column packed with FFAP as stationary phase. However, thiabendazole has commonly been converted into the methyl ester or pentafluorobenzyl ester before GC analysis. LC methods are also used for the determination of thiabendazole.

**See also:** **Amperometry.** Derivatization of Analytes. **Food and Nutritional Analysis:** Meat and Meat Products; Dairy Products. **Liquid Chromatography:** Food Applications. **Nitrogen.** **Polarography:** Inorganic Applications. **Quality Assurance:** Primary Standards. **Spectrophotometry:** Organic Compounds. **Sulfur.** **Vitamins:** Fat-Soluble; Water-Soluble.

### Further Reading

- Ali MS (1985) Rapid quantitative method for simultaneous determination of benzoic acid, sorbic acid, and four parabens in meat and non-meat products by liquid chromatography. *Journal of the Association of Official Analytical Chemists* 68(3): 488–492.
- AOAC International (2000) In: Horwitz W (ed.) *Official Methods of Analysis*, 17th edn. Gaithersburg: Association of Official Analytical Chemists.
- Boley NP (1990) Food additives. In: Gordon MH (ed.) *Principles and Applications of Gas Chromatography in Food Analysis*. Chichester: Ellis Horwood.

- Boyce MC (2001) Determination of additives in food by capillary electrophoresis. *Electrophoresis* 22: 1447–1459.
- Cheng QC and Wang J (2001) Simultaneous determination of artificial sweeteners, preservatives, caffeine, theobromine and theophylline in food and pharmaceutical preparations by ion chromatography. *Journal of Chromatography A* 937: 57–64.
- Dezman DL, Nagy S, and Brown GE (1986) Postharvest fungal decay control chemicals: treatments and residues in citrus fruits. *Residue Reviews* 97: 37–92.
- Gonzalez M, Gallego M, and Valcarcel M (1999) Gas chromatographic flow method for the preconcentration and simultaneous determination of antioxidant and preservative additives in fatty foods. *Journal of Chromatography A* 848: 529–536.
- Gordon MH (ed.) (1990) Amino acids and other nitrogen-containing compounds. In: *Principles and Applications of Gas Chromatography in Food Analysis*. Chichester: Ellis Horwood.
- Page BD and Charbonneau CF (1989) Liquid chromatographic determination of several antioxidants in dry foods. *Journal of the Association of Official Analytical Chemists* 72(2): 259–265.
- Pizzoferrato L, Quattnici E, and di Lucco G (1990) Evaluation of an HPLC method for the determination of sulfiting agents in foods. *Food Additives and Contaminants* 7(2): 189–195.
- Wedzicha BL (1984) *Chemistry of Sulphur Dioxide in Foods*. London: Elsevier Applied Science.
- Yang M-H, Lin H-J, and Choong Y-M (2002) A rapid gas chromatographic method for direct determination of BHA, BHT and TBHQ in edible oils and fats. *Food Research International* 35: 627–633.

## Contaminants

**M O’Keeffe**, The National Food Centre (Teagasc), Dublin, Republic of Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The common understanding of food contaminants is very broad, covering topics as diverse as natural toxins through agrochemical and veterinary drug residues to industrial contaminants and extraneous matter such as glass and metal objects introduced during food processing.

A European Community Council Regulation (No. 315/93) defines a food contaminant as: “any substance not intentionally added to food which is present in such food as a result of the production (including operations carried out in crop husbandry, animal husbandry and veterinary medicine), manufacture, processing, preparation, treatment, packing, packaging, transport or holding of such food, or as a result of environmental contamination. Extraneous matter, such as, for example, insect fragments, animal hair, etc. is not covered by this definition.”

This definition of food contaminants encompasses a very broad range of chemical substances, many of which are covered by other articles in this Encyclopedia, particularly natural toxins, mycotoxins, pesticides, antibiotics, and metals (both individually and collectively). Consequently, this article will give an overview of chemical contaminants in foods, dealing particularly with agrochemical and veterinary drug

residues in food and residues of industrial contaminants, leaving the more in-depth treatment of particular classes of contaminants to the other articles.

Chemical contaminants in food are of concern to the consumer, to food manufacturers, and to regulatory authorities. Contamination of food with chemicals may arise from the use of agrochemicals and veterinary drugs in food production or from environmental pollutants and industrial chemicals during food processing and packaging. A wide variety of analytical techniques are used for the detection and quantitation of contaminants in food ranging from microbiological tests through immunoassays to all forms of chromatography. Analyses for contaminants in foods are characterized by the requirements for intricate extraction/cleanup and for sensitivity. This article describes the methodologies used for contaminant residue analyses in food, covering screening and confirmatory techniques, the range of extraction and clean-up methodologies used, and both specific and multiresidue procedures.

### Major Analytes

The major source of potential contaminants in food is the direct application of agrochemicals to plants and animals and the use of veterinary drugs for treating food-producing animals.

Pesticides, including fungicides, herbicides, and insecticides, are routinely used in most food production to control pests that would otherwise destroy or reduce the food production. Because of the very large

- Boyce MC (2001) Determination of additives in food by capillary electrophoresis. *Electrophoresis* 22: 1447–1459.
- Cheng QC and Wang J (2001) Simultaneous determination of artificial sweeteners, preservatives, caffeine, theobromine and theophylline in food and pharmaceutical preparations by ion chromatography. *Journal of Chromatography A* 937: 57–64.
- Dezman DL, Nagy S, and Brown GE (1986) Postharvest fungal decay control chemicals: treatments and residues in citrus fruits. *Residue Reviews* 97: 37–92.
- Gonzalez M, Gallego M, and Valcarcel M (1999) Gas chromatographic flow method for the preconcentration and simultaneous determination of antioxidant and preservative additives in fatty foods. *Journal of Chromatography A* 848: 529–536.
- Gordon MH (ed.) (1990) Amino acids and other nitrogen-containing compounds. In: *Principles and Applications of Gas Chromatography in Food Analysis*. Chichester: Ellis Horwood.
- Page BD and Charbonneau CF (1989) Liquid chromatographic determination of several antioxidants in dry foods. *Journal of the Association of Official Analytical Chemists* 72(2): 259–265.
- Pizzoferrato L, Quattnici E, and di Lucco G (1990) Evaluation of an HPLC method for the determination of sulfiting agents in foods. *Food Additives and Contaminants* 7(2): 189–195.
- Wedzicha BL (1984) *Chemistry of Sulphur Dioxide in Foods*. London: Elsevier Applied Science.
- Yang M-H, Lin H-J, and Choong Y-M (2002) A rapid gas chromatographic method for direct determination of BHA, BHT and TBHQ in edible oils and fats. *Food Research International* 35: 627–633.

## Contaminants

**M O’Keeffe**, The National Food Centre (Teagasc), Dublin, Republic of Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The common understanding of food contaminants is very broad, covering topics as diverse as natural toxins through agrochemical and veterinary drug residues to industrial contaminants and extraneous matter such as glass and metal objects introduced during food processing.

A European Community Council Regulation (No. 315/93) defines a food contaminant as: “any substance not intentionally added to food which is present in such food as a result of the production (including operations carried out in crop husbandry, animal husbandry and veterinary medicine), manufacture, processing, preparation, treatment, packing, packaging, transport or holding of such food, or as a result of environmental contamination. Extraneous matter, such as, for example, insect fragments, animal hair, etc. is not covered by this definition.”

This definition of food contaminants encompasses a very broad range of chemical substances, many of which are covered by other articles in this Encyclopedia, particularly natural toxins, mycotoxins, pesticides, antibiotics, and metals (both individually and collectively). Consequently, this article will give an overview of chemical contaminants in foods, dealing particularly with agrochemical and veterinary drug

residues in food and residues of industrial contaminants, leaving the more in-depth treatment of particular classes of contaminants to the other articles.

Chemical contaminants in food are of concern to the consumer, to food manufacturers, and to regulatory authorities. Contamination of food with chemicals may arise from the use of agrochemicals and veterinary drugs in food production or from environmental pollutants and industrial chemicals during food processing and packaging. A wide variety of analytical techniques are used for the detection and quantitation of contaminants in food ranging from microbiological tests through immunoassays to all forms of chromatography. Analyses for contaminants in foods are characterized by the requirements for intricate extraction/cleanup and for sensitivity. This article describes the methodologies used for contaminant residue analyses in food, covering screening and confirmatory techniques, the range of extraction and clean-up methodologies used, and both specific and multiresidue procedures.

### Major Analytes

The major source of potential contaminants in food is the direct application of agrochemicals to plants and animals and the use of veterinary drugs for treating food-producing animals.

Pesticides, including fungicides, herbicides, and insecticides, are routinely used in most food production to control pests that would otherwise destroy or reduce the food production. Because of the very large



number of these substances used in food production, analyses are mainly multiresidue, directed to particular classes of pesticides such as organochlorine, organophosphorus, or carbamate substances.

Veterinary drugs are used to treat animals for bacterial infections (antibiotics) and parasitic infestations (anthelmintics, coccidiostats), to enhance growth (antibiotics, anabolics, partitioning agents, thyrostats), to control fertility and reproduction (steroid hormones), or to alter behavior (tranquillizers and sedatives). Within each of the drug types there are different classes and methods for residue analysis are typically class specific (e.g., methods for the class of tetracycline antibiotics). In certain cases, multiresidue methods are available that are broader than specific classes.

In many cases, maximum residue limits (MRLs) exist for veterinary drugs and the analytical methods must be capable of measuring below such MRLs. Since MRLs vary considerably, the sensitivity requirements of the various methods also vary widely. Typically, MRLs for veterinary drug residues vary from  $1 \text{ mg kg}^{-1}$  (ppm) to  $1 \mu\text{g kg}^{-1}$  (ppb) (Table 1) or, in the case of banned substances, to 'no detectable residue'. Where 'no detectable residue' is specified, it is common to declare an 'action limit' or a 'limit of decision' for use by regulatory agencies; this is not strictly scientifically based and takes into account other aspects such as the sensitivity of a suitable confirmatory technique and policy requirements.

A wide range of industrial contaminants may occur in food arising from environmental contamination, from food processing procedures, or from food packaging materials. These substances include polychlorinated biphenyls (PCBs), dioxins and heavy metals, nitrosamines, polycyclic aromatic hydrocarbons (PAHs), and acrylamide, and packaging material monomers and plasticizers.

## Sampling and Preparation

### Samples and Sampling

Sampling of food for residue analysis is governed by the rules that apply to all sampling, namely that the analytical sample should contain the residues in the forms and at the levels corresponding to those present in the lot of which the sample is representative. Sampling of food for residue analyses may be very directed (e.g., to establish whether residues of a particular pesticide are present in a particular shipment of fruit) or very broad (e.g., as part of a national monitoring program designed to establish the incidence of use of banned substances in meat-producing animals). For the latter circumstances exact

**Table 1** Maximum residue limits (MRLs) specified by EC Regulation (2377/90) for some veterinary drugs

<i>Veterinary drug</i>	<i>Target tissues</i>	<i>MRL</i> ( $\mu\text{g kg}^{-1}$ )
<i>Anti-infectious agents – chemotherapeutics</i>		
Sulfonamides	Muscle, liver, kidney, milk	100
Trimethoprim	Muscle, liver, kidney, fat, milk	50
<i>Anti-infectious agents – antimicrobials</i>		
Benzylpenicillin, Ampicillin, Amoxicillin	Muscle, liver, kidney, fat	50
	Milk	4
Oxacillin, Cloxacillin, Dicloxacillin	Muscle, liver, kidney, fat	300
	Milk	30
Tetracyclines	Kidney	600
	Liver	300
	Eggs	200
	Muscle, milk	100
Spiramycin	Liver, kidney, fat	300
	Muscle, milk	200
Tylosin	Muscle, liver, kidney, fat	100
	Eggs	200
	Milk	50
<i>Antiparasitic agents</i>		
Ivermectin	Liver	15–100 <sup>a</sup>
	Fat	20–40 <sup>a</sup>
Febantel, Fenbendazole, Oxfendazole	Liver	500
	Muscle, kidney, fat	50
	Milk	10
Levamisole	Liver	100
	Muscle, kidney, fat	10
<i>Agents acting on the nervous system</i>		
Azaperone	Muscle, liver, kidney, fat	100
Carazolol	Liver, kidney	15–25 <sup>a</sup>
	Muscle, fat	5
<i>Glucocorticoids</i>		
Dexamethasone, Betamethasone	Liver	2
	Muscle, kidney	0.75
	Milk	0.3

<sup>a</sup> Depending on species.

sampling plans are required which are statistically based so that an accurate representation of the situation may be obtained from the analyses.

The appropriate sample depends to a large extent on the nature of the food. In the case of plant foods such as grain, a composite sample is constructed

from a number of samples taken at specified points throughout the load, as described in ISO Standard 7002:1986. In the case of plant foods such as fruit and vegetables, the composite sample is constructed from individual items selected according to a defined procedure and is described in ISO Standard 874:1980.

In the case of foodstuffs of animal origin, sampling from milk and milk products is covered by a number of ISO standards – 707:1997; 5538:1987; 8197:1988. The purpose of these defined procedures is to ensure representativeness of the sample for a variety of nonhomogeneous and homogeneous solid and liquid, bulk and packaged foods. For meat and edible tissues, samples taken from liver or kidney are usually representative, although there may be differences in residue content between the two sections of the kidney (medulla and cortex). Samples from carcasses are normally taken from the same position (muscle or fat) on each carcass and chosen so as to reduce commercial loss as much as is possible.

Samples of body fluids or excreta are often used to monitor illicit use of prohibited substances or for indirect determination of the presence of contaminants in foods of animal origin. For example, analyses of urine, blood, bile, feces may be used to establish whether specified veterinary drugs have been used on a particular animal. Where well-established ratios exist between the levels of drug residue in body fluids and edible tissues (e.g., sulfadimidine levels in blood or urine and in muscle or kidney), analyses of body fluid samples may be used to indicate the levels in edible tissues.

### **Sample Storage**

Samples of foodstuffs taken for analysis for contaminant residues must be stored in such a way as to prevent deterioration of the sample and reduction in the residue content. In the case of relatively dry samples, such as milk powders, or samples protected by an outer skin, such as fruit and vegetables, storage may be at ambient temperature for short periods or at refrigeration temperature. In the case of most animal food materials, storage at refrigeration temperature is satisfactory if analyses are to commence in the short term, but frozen storage ( $\leq -18^{\circ}\text{C}$ ) is recommended, and is essential for prolonged storage. Even lower storage temperatures of  $-40^{\circ}\text{C}$  to  $-70^{\circ}\text{C}$  may be recommended for certain contaminant residues. Homogenization and division of the sample, for example, where required for contra-analysis, may be performed before freezing to avoid the cycle of thawing-freezing-thawing, which may be detrimental to the stability of some contaminant

residues. However, homogenization may increase the contact between contaminant residues and endogenous enzymes in the sample matrix (e.g., chloramphenicol in liver, the effect of  $\beta$ -lactamase on penicillin). Sometimes, addition of preservatives to the sample, such as the antimicrobial sodium azide or enzyme inhibitors, may be useful but only where these agents do not interfere with the subsequent analysis.

### **Sample Preparation**

The analytical sample is generated as a subsample from the composite or bulk sample. In the case of fruit and vegetables, cutting/chopping of the composite sample, mixing, and subsampling is required. Grains are normally subsampled, and then ground before analysis. Milk and milk powders, being already very homogeneous, usually require simple mixing before an aliquot is taken for analysis. Milk products, such as cheese, require chopping/grinding and mixing before subsampling. Tissue samples are minced/homogenized and subsampled. Urine samples are mixed and may be centrifuged, to remove solid particles, prior to sampling for analysis. Blood may be clotted and serum removed or collected in heparinized containers and the blood cells separated by centrifugation and plasma removed.

## **Residue Analysis**

### **Sample Extraction**

The challenge to the residue analyst is to extract the residue of interest from the sample and to purify the resultant extract to the extent necessary to allow for unequivocal identification and accurate quantification by the selected determination technique. The sample extraction and cleanup procedures are defined, therefore, by the requirements of the determination step. These requirements can vary from as little as no extraction, e.g., use of an intact piece of tissue on an agar plate with a test organism for determination of inhibitory substances; or minimal treatment, e.g., dilution of urine and direct assay for residues of growth enhancing substances by immunoassay; to very extensive extraction and cleanup procedures, e.g., multiple extractions with organic solvent, filtration of extracts, evaporation of combined extracts, liquid-liquid partitioning into acid, alkalization, and liquid-liquid partitioning with organic solvent, neutralization, and multiple liquid-liquid partitioning into organic solvent, evaporation and derivatization, as required in the Association of Official Analytical Chemists (AOAC) method for determination of sulfamethazine residues in tissues by gas chromatography-mass



spectrometry (GC–MS) or gas chromatography-electron capture detection (GC–ECD).

With the exception of some limited assays where determination can be made on the intact or diluted sample, most residue determination procedures commence with extraction of the residue(s) from the sample matrix. Organic solvents are commonly used, with selection of the appropriate solvent being based, largely, on the polarity of the residues of interest. Organic solvents cause denaturation of protein and decrease the interaction between residues and proteins. Aqueous extraction is also used particularly for relatively polar residues and has the advantage that lipids present in the sample matrix are not extracted. Modification of the sample homogenate through pH changes may be necessary to make the residue of interest available to the organic solvent extractant; by adjusting pH, the residues may become largely non-ionized and hence more readily extracted. Typically, solvent-sample ratios of 2:1 to 10:1 are used in the extraction procedure and repeated extractions (two or more) are carried out on the sample. The extraction procedure may be by mechanical blending, by manual (separatory funnel), or mechanical (shaker) mixing, by ultrasonics, or by accelerated solvent extraction, which provides efficient extraction under controlled temperature/pressure conditions. In the case of animal tissues, protease digestion prior to extraction may be used to provide better accessibility of the solvent to the residue molecules and to release residues weakly bound to proteins. Where conjugated forms of residues arise in samples, a deconjugation step using enzymes is undertaken prior to extraction; many residues occur as glucuronide or sulfate conjugates, which, being hydrophilic, are difficult to extract.

Having extracted the residue into the extractant, it may be separated from the matrix material by centrifugation, filtration, freezing of the aqueous phase, or binding of the aqueous phase using anhydrous salts. Where the extract is an organic solvent, the volume may be reduced or brought to dryness, by controlled heating and distillation of solvent or evaporation under nitrogen or vacuum, prior to cleanup of the extract. Where the extract is aqueous, transfer of the residues to an organic solvent by liquid–liquid partitioning or isolation of the residues directly from the aqueous phase by a procedure such as immunoaffinity chromatography may be the first step in cleanup of the extract.

### Extract Cleanup

Cleanup of the extract is undertaken by a range of procedures such as liquid–liquid partitioning or

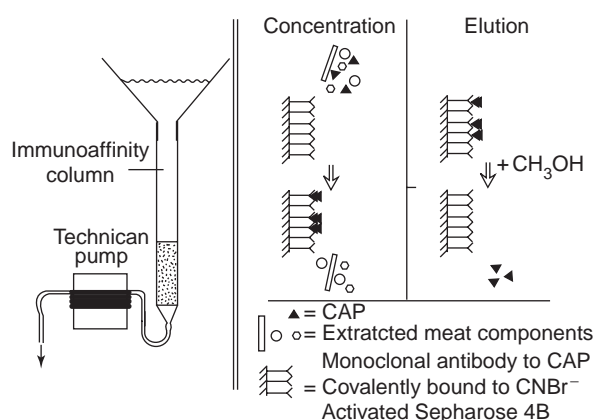
column chromatography. Liquid–liquid partitioning consists of the intimate mixing of two immiscible solvents such that the residue of interest is accumulated selectively in one phase while certain of the contaminating coextractives are accumulated into the other phase. The classical procedure for liquid–liquid partitioning is by manual mixing in a separatory funnel. A form of liquid–liquid partitioning, sometimes termed liquid–solid partitioning, may also be undertaken on columns or cartridges packed with inert material, such as diatomaceous earth. An aqueous extract is introduced on to the column and distributes over the packing material into which the water is absorbed. When the second solvent (phase) is added to the column, it mixes intimately with the inert material and the residue is extracted from the column.

Some cleanup procedures involve a number of liquid–liquid partitioning steps, typically to remove fat and nonpolar coextractives in one step and polar coextractives in another. Movement from one liquid–liquid partitioning step to another may be by separation of phases, by evaporation of one phase and resolubilization in another solvent, by freezing of an aqueous phase and decanting of an organic phase, or by changing the pH to alter the affinity of the residue of interest for the aqueous phase.

Column chromatography is very widely used as a cleanup procedure. The classical glass column chromatography based on a variety of size-exclusion, polar (normal), nonpolar (reversed phase), and ion-exchange materials has been superseded to a considerable extent by microcolumn cartridges available commercially with a very broad range of packing materials or sorbents. This cartridge-based chromatography, known as solid-phase extraction (SPE) is found as a step in many cleanup procedures for residues in foodstuffs. There have been considerable developments during the last decade both in the range of sorbents and the formats available for SPE. In addition to the classical modified silica reversed-phase sorbents a number of polymeric sorbents have been developed that have multiphase characteristics. Formats for automated SPE, including 96-well plate systems, are available.

Automated systems for size-exclusion, or gel permeation, chromatography are very popular, particularly in pesticide residue analyses. Another form of column chromatography, based on the use of specific antibodies to the analyte(s) of interest, is immunoaffinity chromatography (IAC). In this technique, antibodies specific for one or more analytes are covalently bound to an activated support such as Sepharose, which is then packed in a column (**Figure 1**). This technique is particularly suited for coupling

with aqueous extraction. Dialysis, particularly in an automated format, has been used to separate the analytes of interest from larger-sized coextractives in an aqueous extract. However, this technique has not been applied very widely in residue analysis. Column-switching techniques, whereby semipurified sample extracts may be cleaned-up and/or concentrated on an SPE-type precolumn, prior to automated switching into the flow to a high-performance liquid chromatography (HPLC) analytical column, are being used and further developed (Figure 2).

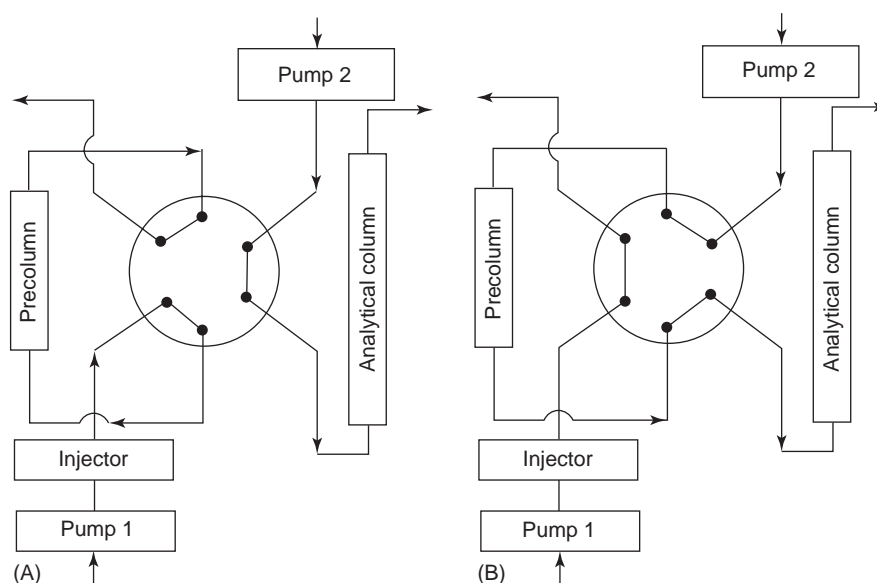


**Figure 1** Schematic representation of immunoaffinity column cleanup. CAP, chloramphenicol (an antimicrobial). (Reproduced with permission from Haagsma N (1990) Sample pretreatment in drug residue analysis. In: Haagsma N, Ruiter A, and Czedik-Eysenberg PB (eds.) *Residues of Veterinary Drugs in Food*, Proceedings of the *EuroResidue Conference*, Noordwijkerhout, The Netherlands, 21–23 May 1990, p. 47. Utrecht: Rijksuniversiteit.)

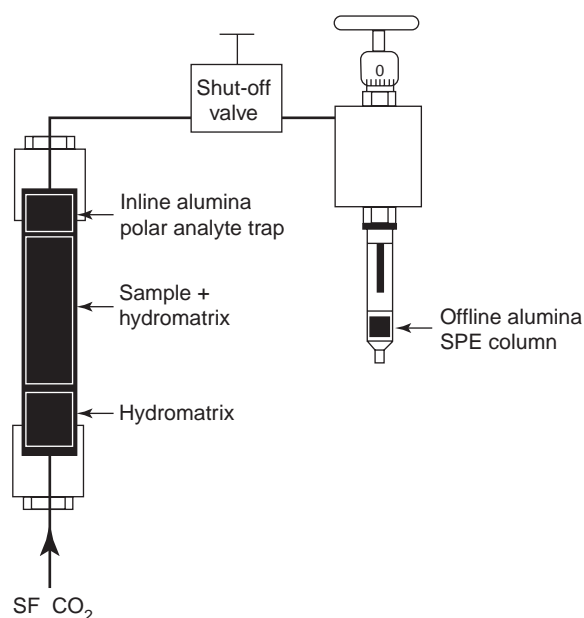
Some techniques that combine the properties of extraction and cleanup are supercritical fluid extraction (SFE) and matrix solid-phase dispersion (MSPD). Supercritical fluids, i.e., at a temperature and pressure in excess of their critical point, have unique properties for selective extraction of analytes from a sample. Solid samples are mixed with an inert dispersant, such as hydromatrix, and the mixture packed into the cell of the SFE apparatus. The sample is extracted with supercritical CO<sub>2</sub>, with or without addition of organic modifier, and the extracted analytes may be collected inline or offline on suitable adsorbents (Figure 3). Further cleanup of the sample extract may be performed using SPE. MSPD is based on intimate mixing of animal tissue sample with a bonded silica, such as C<sub>18</sub>, and packing of the blended material into a column from which interferences can be eluted by washing with solvents and the analytes eluted using a selective solvent.

### Residue Determination

Residues are determined in the purified extracts by chromatographic or immunochemical techniques. In the chromatographic systems, thin-layer chromatography (TLC), liquid chromatography (LC), and GC, the analytes are separated on plates or columns and determined by colorimetry, by spectrophotometry (ultraviolet (UV), infrared (IR, Fourier transform infrared (FTIR)), by fluorescence, by selective detectors (in GC analysis: ECD, flame photometric (FPD), nitrogen/phosphorus (NPD, TSD), etc.), or by MS. Separations may also be achieved by



**Figure 2** Column switching flow diagram: (A) preconcentration and (B) desorption and chromatographic separation. Reproduced with permission from O'Keefe, *Residue Analysis in Food*, p. 78 (2000) Taylor and Francis.



**Figure 3** SFE system configured for offline analyte collection using a standard 6 ml SPE column and inline using a packed sorbent bed. Reproduced with permission from O'Keefe, *Residue Analysis in Food*, p. 96 (2000) Taylor and Francis.

capillary zone electrophoresis, coupled with UV or MS detectors.

In the case of immunoassays, the binding of the analyte of interest by an appropriate antiserum is the basis of the determination procedure, using radio-labeled, enzyme, or fluorescent markers. Apart from the classical immunoassays in tube or 96-well plate formats, a variety of 'dip-stick' or sol particle systems are available for rapid testing. There have been considerable advances in the application of sensor technologies to high throughput immunoassays for contaminant residues, particularly using surface plasmon resonance (SPR). In SPR-based biosensor systems, antigen-antibody binding reactions on a thin gold film surface may be measured as shifts in the angle of light. Fully automated SPR-based biosensor systems are available – e.g., Biacore<sup>TM</sup> – on which a broad range of rapid analyses have been developed for drug classes such as sulfonamides, coccidiostats, and beta-agonists.

### Quality Assurance

Methods are validated using samples fortified with the analyte of interest at appropriate levels (i.e., normally levels similar to levels found in the samples), levels related to MRLs for permitted substances, or levels related to minimum required performance limits for prohibited or unapproved substances. Following validation of the method and its

**Table 2** Food analysis performance assessment scheme (FAPAS) – contaminants in food

Series Title	Analytes
2 Veterinary drug residues	Sulfonamides, tetracyclines, chloramphenicol, $\beta$ -agonists, nitrofurans, malachite green
4 Aflatoxins	Aflatoxins B, G, M <sub>1</sub>
5 OC pesticides and PCBs	Organochlorine pesticides and PCBs
6 Environmental contaminants	Dioxins, marine toxins, PAHs
7 Metallic contaminants	Pb, Cd, As, Hg, Sn, Fe, Cu, Zn
9 OP and pyrethroid pesticides	Organophosphorus and pyrethroid pesticides
12 Specific migration	Semicarbazide
15 Nitrate analysis	Nitrate, nitrite
16 Patulin	Patulin
17 Ochratoxin A	Ochratoxin A
19 Pesticides	Pesticides
22 Fusariums	Deoxynivalenol, fumonisins, zearalenone

establishment as a routine method in the laboratory, ongoing quality control is maintained by assay of negative and positive controls and/or of fortified samples in the assay. These controls may be generated within the laboratory and are supplemented by external certified reference materials (CRMs), certified for specified levels of the analyte(s) of interest. Such CRMs are produced by the EC Institute for Reference Materials and Measurements, by international and national agencies, such as NIST, and by commercial companies. Quality assurance for residue analysis may be provided also by involvement of a laboratory in appropriate proficiency testing schemes, operated according to ISO standards (ISO Guides 43-1, 43-2: 1997). One example of such is the Food Analysis Performance Assessment Scheme organized by the Central Science Laboratory, UK, which produces samples on a number of occasions each year covering a particular analyte or class of analytes (Table 2). Quality assurance of residue analysis is frequently maintained within a laboratory accreditation framework, to standards such as ISO 17025.

## Quantitative Methods

### Pesticides

This topic is covered in detail elsewhere in the Encyclopedia so only brief mention of relevant methods will be made here. A number of multiresidue procedures for organochlorine (OC), organophosphorus

(OP), and carbamate pesticides are available as standard methods by the AOAC covering both non-fatty and fatty foods. These methods are based on solvent extraction and a variety of column chromatographic cleanup procedures with determination by GC using selective detectors. The International Dairy Federation has approved similar methods for OC and OP pesticide residues in milk and milk products. The AOAC have standard methods for specific pesticides or groups of pesticides that are approved for certain foodstuffs; these methods involve colorimetric, spectrophotometric, or gas chromatographic determination. In the case of methylcarbamate pesticides, a liquid chromatographic method is approved.

### Veterinary Drugs

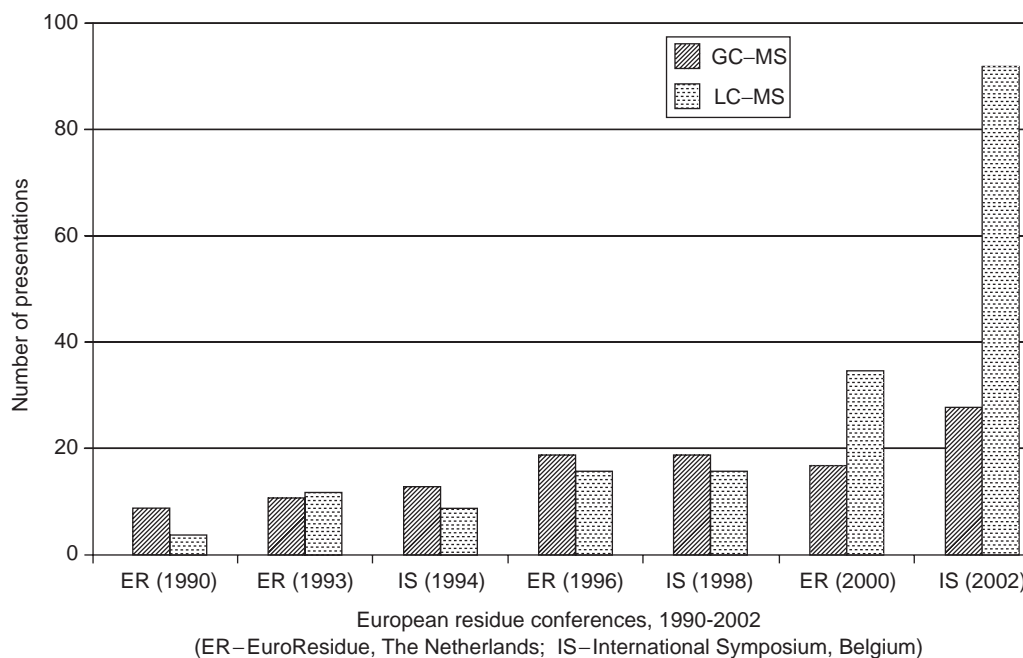
There are relatively few standard methods for veterinary drug residues although very large numbers of methods are available for the different classes of compounds. The United States Department of Agriculture's (USDA) Food Safety and Inspection Service publish a Chemistry Laboratory Guidebook that, in addition to some AOAC methods, contains a number of validated methods.

Within the European Union, an alternative approach has been adopted to that of standard methods. Commission Decision 2002/657/EC specifies a range of performance criteria for analytical methods to be used to test for the presence of

veterinary drug and other residues in foods of animal origin, under Council Directive 96/23/EC. This decision defines suitable analytical methods and the performance criteria that should be determined for methods, depending on whether they are classified as qualitative or quantitative and screening or confirmatory. Validation procedures are specified for performance characteristics such as specificity, trueness, ruggedness, stability, calibration, recovery, repeatability, reproducibility, decision limit, and detection capability.

Confirmatory methods should preferably be based on MS, providing direct information on the molecular structure of the analyte(s), especially for residues of prohibited substances or substances for which no MRLs are set. However, validated chromatographic methods with specific detectors (e.g., photodiode array, fluorescence) or using two or more different chromatographic separation systems may be used, particularly for contaminants for which MRLs are specified.

There have been major developments in LC coupled with MS (LC-MS/MS) during the last 5–10 years and this technology has been applied very widely for veterinary drug residue analysis (Figure 4). LC-MS/MS is particularly suitable for veterinary drug residues in that derivatization, commonly required for GC-MS analysis, is not necessary. In addition, it can provide robust and sensitive determination.



**Figure 4** Increase in use of mass spectrometry in veterinary drug residue analysis (1990–2002), presentations at major European conferences.

**Antibiotics** The AOAC has listed methods for sulfamethazine residues in swine tissues with determination either by GC-MS or GC-ECD of methylated derivatives and for sulfamethazine (and for the class of sulfonamides) in milk with determination by HPLC-UV. There is an AOAC method for the class of sulfonamide antimicrobials in animal tissues using solvent extraction and liquid partitioning with determination by TLC and fluorimetric scanning. For analysis of tetracyclines, AOAC describes methods based on buffer extraction from tissue samples and SPE (C<sub>18</sub>) cleanup, or metal chelate affinity binding from milk samples, with determination in both cases by HPLC-UV. USDA/FSIS methods include: (1) a method (similar to the AOAC GC-MS method for sulfamethazine) for confirmation of sulfonamide residues in edible tissues using solvent extraction and multiple liquid partitioning with determination of the methylated derivatives by GC-MS; (2) methods for determination and confirmation of chloramphenicol in muscle by solvent extraction, liquid partitioning, and determination of the trimethylsilane (TMS) derivative by GC-ECD and GC-MS, respectively; and (3) a method for determination of the beta-lactam antibiotic amoxicillin by aqueous extraction, cleanup by tricarboxylic acid precipitation, and ether extraction and formation of a fluorescent derivative for determination by LC.

Methods for the determination of residues of most of the important classes of antimicrobials (sulfonamides, chloramphenicol, tetracyclines, beta-lactams, quinolones, and nitrofurans) have been developed based on HPLC or LC-MS/MS determination. These methods are applicable to edible tissues, milk, eggs, urine, and other matrices. The methods typically involve sample workup procedures of solvent extraction and single or multiple SPE cleanup prior to chromatographic determination. For example, extraction of sulfonamides and chloramphenicol may be achieved using aqueous methanol, acetonitrile, or ethyl acetate with cleanup using SPE on reversed phase (e.g., C<sub>18</sub>) and/or cation exchange (e.g., benzenesulfonic acid) sorbents. In the case of tetracyclines, extraction is commonly performed with EDTA/McIlvane buffer and cleanup on C<sub>18</sub>, Oasis HLB<sup>TM</sup> and/or SCX (propylsulfonic acid) SPE cartridges. Determination may be by HPLC with UV detection, by fluorescence detection following postcolumn derivatization, or by MS. In the case of the prohibited nitrofurans, which occur as protein-bound metabolites in edible tissues, methods are directed at releasing these metabolites by acid hydrolysis, derivatization with nitrobenzaldehyde, extraction with ethyl acetate, and determination by LC-MS/MS.

**Growth-enhancing substances** An AOAC method is approved for the growth promoter melengestrol acetate in animal tissues by solvent extraction of fatty tissue and cleanup on a Florisil column with determination by GC-ECD.

Numerous methods have been validated for the determination of anabolic agents (natural and synthetic estrogens, androgens, and gestagens) in edible tissues and in other samples that can be used for food monitoring such as urine, bile, and feces. These methods involve determination of analytes by LC-MS/MS, by GC-MS of suitable derivatives, or by immunological assay, with or without prior LC separation. Typical steps in the extraction, cleanup, and isolation of the analytes are the following:

1. enzymatic hydrolysis of conjugated residues;
2. mechanical homogenization, or enzymatic digestion with proteolytic enzymes;
3. solvent extraction direct, or on diatomaceous earth;
4. cleanup using SPE, e.g., coupled C<sub>18</sub> and amino cartridges for anabolic steroids in urine;
5. analyte isolation by IAC or HPLC;
6. determination of residues by LC-MS/MS, or by GC-MS following derivatization.

Methods for determination of beta-agonists (repartitioning agents) in biological samples involve aqueous extraction using ultrasonics, enzymatic hydrolysis of conjugated residues, solvent extraction, isolation of analytes by IAC, and determination by LC-MS/MS, or of derivatives by GC-MS.

For example, a semiautomated quantitative method is described which is capable of screening and confirmation of 22 steroids in urine. This method involves enzymatic deconjugation followed by dual SPE on C<sub>18</sub> and amino cartridges, using an automated system (ASPEC XL4, Gilson Ltd.). Screening is undertaken by LC-MS/MS, with an electrospray interface, and the column effluent is split between the mass spectrometer and a fraction collector. Where the presence of steroid residues is determined, confirmation is achieved by reanalysis of the collected fraction using either LC-MS/MS, while monitoring additional transition products, or GC-high resolution MS, after derivatization.

**Anthelmintics** The major classes of anthelmintics, or antiparasitic drugs, are benzimidazoles and macrocyclic lactones (avermectins and milbemycins). USDA/FSIS have approved a method based on LC/fluorescence detection for determination of the anthelmintics albendazole and ivermectin in tissues (albendazole extracted with ethyl acetate and cleanup



by liquid partitioning and SPE on a C<sub>18</sub> cartridge; ivermectin extracted with iso-octane and cleanup by liquid partitioning, formation of a fluorescent derivative by heating with imidazole reagent and purification by SPE on a silica cartridge).

Multiresidue methods for benzimidazoles and macrocyclic lactones are applied in many testing laboratories. For example, benzimidazoles may be determined by extraction from tissue with ethyl acetate, removal of fat by liquid/liquid partitioning between acidic ethanol and hexane, cleanup by automated SPE on C<sub>18</sub> or C<sub>8</sub> cartridges, and HPLC chromatography with UV detection at 298 nm. In the case of the macrocyclic lactones, samples are typically extracted using acetonitrile and modified with water–triethylamine prior to purification on C<sub>8</sub> or C<sub>18</sub> SPE cartridges. Residues may be determined, after derivatization, by HPLC with fluorescence detection (excitation 365 nm, emission 475 nm) or determined without derivatization by LC–MS/MS.

**Other veterinary drugs** Other veterinary drugs of importance are the anticoccidial feed additives such as the ionophores, narasin, monensin, salinomycin, apramycin, lasalocid, and nicarbazine. The USDA/FSIS have a method for the major ionophores in tissue samples, which is based on purification of sample extracts by silica gel, alumina, or ion-exchange column chromatography and determination by TLC with detection by bioautography. A number of alternative methods based on immunoassays, biosensor technology, and HPLC have been developed.

Tranquillizers and beta-blockers are important potential contaminants in edible tissues, particularly of pigs, where they may be used to counteract the stress susceptibility of some breeds during transport, which results in poor quality meat and even mortality. Some of these substances are prohibited for use in food-producing animals while others have MRL values set. A simple method for seven of these compounds (azaperol, azaperone, carazolol, chlorpromazine, acepromazine, propionylpromazine, xylazine) involves extraction of kidney or muscle samples with acetonitrile, modification of the extract with a sodium chloride solution (10%), and SPE on an Oasis HLB<sup>TM</sup> cartridge. The eluate (acetonitrile) is evaporated, reconstituted in aqueous acetonitrile, and residues determined by LC–MS/MS, with atmospheric pressure chemical ionization. Two additional steps are incorporated into this method to increase the numbers of sample extracts that may be analyzed without affecting the mass spectrometer source; first, an inline column (Biomatrix<sup>®</sup>) is used ahead of the analytical column to remove any macromolecules in the sample extracts and, second, a switching valve is

used ahead of the mass spectrometer that allows the eluate from the analytical column to pass through the mass spectrometer only during analyte elution.

### Industrial and Processing Contaminants

Typical contaminants of concern are PCBs and polychlorinated dibenzo-*p*-dioxins (dioxins), which are highly persistent environmental contaminants arising from industrial processes, detergents, and disinfectants (especially in milk) that are used in cleaning processes, process contaminants such as PAHs and acrylamide, and food packaging monomers and residual solvents from extraction processes or printed packaging material.

The AOAC standard method for PCBs in foods is similar to that for organochlorine pesticide residues, involving extraction with lipophilic solvent and cleanup by liquid partitioning and column adsorption chromatography with determination by GC with ECD and/or potassium chloride thermionic detector. Current methods for dioxins and coplanar PCBs involve determination by GC coupled with high-resolution MS, to achieve the sensitivity and specificity required. Typically, the sample preparation consists of fat/lipid extraction and isolation of dioxins/PCBs by gel permeation chromatography with extract purification by column chromatography. A method that may be used to screen samples for the presence of dioxins/PCBs is the CALUX<sup>®</sup> cell-based assay. This bioassay technique – Chemical Activated Luciferase eXpression – is based on use of rat or mouse hepatoma cells, modified to show an increased production of luciferase in response to binding of dioxins and dioxin-like PCBs to a particular cell receptor.

Because many detergents and disinfectants contain phosphate, chloride, or iodide, it is difficult to determine the presence of these substances against the naturally occurring levels of these salts in milk or other food. Contamination with detergents or disinfectants is suspected, therefore, by the determination of much higher levels than normally would be found in the food. Ion-selective electrodes are commonly used.

Food packaging monomers and solvents are assayed by headspace GC linked with various detection systems such as FID, AFID, and MS.

Nitrosamines are carcinogenic substances that may form in foods preserved with high levels of sodium nitrite. The AOAC have approved two methods for analysis of nitrosamines in fried bacon, one for volatile *N*-nitrosamines by mineral oil vacuum distillation and extraction into dichloromethane, and the other for *N*-nitrosopyrrolidine by extraction



from a column of Celite with dichloromethane. Nitrosamine levels in the extracts are determined by chemiluminescence, using a gas chromatograph fitted with a thermal energy analyzer.

Acrylamide was identified as a contaminant in heat-treated starchy foods (e.g., potatoes, bread, snack foods) in 2002. Methods for determination of acrylamide in food are based mainly on GC-MS (without derivatization or with derivatization by bromination) or LC-MS/MS. Acrylamide is generally extracted from the food matrix with water, sometimes with defatting using hexane, and cleanup of the extract is achieved using mixed-mode or multiple SPE cartridges.

## Screening Methods

Apart from the methods described above that are used for monitoring of residues in food, meat and milk supplies are assayed routinely for the presence of antimicrobial or 'inhibitory' substances.

In the EC the microbiological 'Four Plate Test' method is used widely as a screening technique for antimicrobial residues in edible tissues (mainly kidney). The principle of this method is a measurement of the extent of inhibition of the growth of test organisms due to the presence of antimicrobial residues in the tissue sample. Modifications and extensions of this methodology have been developed such as a single plate New Dutch Kidney Test and a 12-plate antimicrobial residue identification procedure used in Denmark. A tube-based test, the Premitest<sup>TM</sup>, also uses the principle of inhibition of growth of a test organism but with a colorimetric endpoint, similar to the widely used tests for antimicrobial residues in milk. A sample of fluid is obtained from the tissue sample and incubated in the tube, containing agar seeded with the test organism and a colored substrate sensitive to lowering of pH caused by microbial growth. Presence of inhibitory substances (antimicrobial residues) in the tissue fluid prevents growth of the test organism with no consequent lowering of pH and color change.

The USDA/FSIS have a number of rapid tests for determining the presence of antimicrobial substances in carcasses or in live animals. The 'swab test on premises' and the 'live animal swab test' are developed for kidney (or liver) and urine samples, respectively, using an inhibition agar plate assay. The 'sulfa on site' test is used on farm or at slaughter plant to screen samples of serum, urine, or feed for sulfadimidine residues. The test is based on TLC and uses well-defined ratios between concentrations of sulfadimidine in serum or urine and in edible tissues

to establish whether the concentration in edible tissues is above the MRL of 100  $\mu\text{g kg}^{-1}$ .

Enzyme immunoassays, in a card-test format, are also used for screening carcasses by testing of urine for specific antimicrobial residues such as sulfadimidine and chloramphenicol. The sol particle immunoassay uses adsorption of specific antibodies to dyed colloidal particles (e.g., carbon) as a mechanism for making directly visible the presence of residues (i.e., antigen) in the sample. A lateral flow membrane device is used to identify residue-positive samples through the absence of color formation at the test capture zone (Figure 5).

Milk samples are screened for the presence of antimicrobial residues with inhibition tests in different formats and using a variety of test organisms. Specific antimicrobial identification tests, based on more extensive inhibition tests, enzymatic tests, immunoassays, and receptor assays (Table 3) are also used for milk samples. These tests show considerable



Positive result

Negative result

**Figure 5** Lateral flow membrane device for residue testing. S, sample well; T, test line; C, control line. (Reproduced from O'Keeffe M, Crabbe P, Salden M, *et al.* (2003) Preliminary evaluation of a lateral flow immunoassay device for screening urine samples for the presence of sulphamethazine. *Journal of Immunological Methods* 278: 117–126.)

**Table 3** Detectable concentrations ( $\mu\text{g ml}^{-1}$ ) of antimicrobials in milk for various identification tests

Antimicrobial (class)	Penzyme	Delvotest P	Three-plate	Six-plate	Charm II
$\beta$ -Lactams	0.003–0.04	0.003–0.02	0.003–0.03	0.004–0.04	0.003–0.03
Tetracyclines	–	0.3–0.7	0.4	0.4–0.5	0.1–0.5
Chloramphenicol	–	8	2	15	0.1
Aminoglycosides	–	0.5–15	0.3–20	13–28	0.1
Macrolides	–	0.4–5	0.1	2	0.02
Sulfonamides	–	–	0.1–1	1	0.01

variation in sensitivity for different classes of antimicrobials and for individual antimicrobials within a class.

### Limitations of Current Analytical Procedures

The aim and purpose of residue analysis for contaminants in food is to ensure the safety of food to the consumer. Because of the large number of potential contaminants and the relatively difficult, expensive, and lengthy procedures required for residue analysis in complex matrices such as food, it is difficult to achieve this aim. Multiresidue procedures, use of a combination of screening and confirmatory techniques, use of robotics and automated systems, and soundly based statistical sampling are all factors that contribute to enhanced control of food safety.

Particular problems are (1) the need for sensitive, multiresidue techniques, giving equivalent, high recovery for a range of analytes of diverse chemistry, (2) the difficulties associated with extracting and determining polar metabolites and/or bound residues, and (3) the inability of the analytical procedures to keep pace with food manufacturing. Real-time and/or online analysis, based on highly automated systems, has the potential for solving some of these problems. In other aspects of food analysis (e.g., near-infrared spectroscopy for online compositional analysis) such systems have been developed. Biosensor-based techniques offer good possibilities here.

In the area of contaminants in food, developments in automated systems using rapid extraction and cleanup based on SPE, IAC, etc., and determination by MS offers the possibility for real-time and multiresidue analyses.

**See also:** **Dioxins.** Extraction: Solvent Extraction Principles; Supercritical Fluid Extraction. **Food and Nutritional Analysis:** Antioxidants and Preservatives; Pesticide Residues; Mycotoxins; Packaging Materials. **Gas Chromatography:** Overview. **Hormones:** Steroids.

**Immunoassays, Applications:** Food. **Liquid Chromatography:** Food Applications. **Nitrosamines. Polychlorinated Biphenyls. Quality Assurance:** Reference Materials; Production of Reference Materials. **Sampling:** Theory. **Sensors:** Overview. **Surfactants and Detergents.**

### Further Reading

- Botsoglou NA and Fletouris DJ (2001) *Drug Residues in Food*. New York: Dekker.
- Gilbert J (ed.) (1996) *Progress in Food Contaminant Analysis*. London: Blackie Academic & Professional.
- Haagsma N and Ruiter A (eds.) (1996) *Residues of Veterinary Drugs in Food, Proceedings of the EuroResidue III Conference*, Veldhoven, The Netherlands, 6–8 May 1996. Utrecht: University of Utrecht.
- Moats WA and Medina MB (eds.) (1996) *Veterinary Drug Residues: Food Safety*. Washington DC: American Chemical Society.
- O'Keeffe M (ed.) (2000) *Residue Analysis in Food: Principles and Applications*. Amsterdam: Harwood Academic Publishers.
- Stephany RW and Bergwerff AA (eds.) (2004) *Residues of Veterinary Drugs in Food, Proceedings of the EuroResidue V Conference*, Noordwijkerhout, The Netherlands, 10–12 May 2004. Bilthoven: National Institute of Public Health and the Environment (RIVM).
- Turnipseed SB and Long AR (eds.) (1998) *Analytical Procedures for Drug Residues in Food of Animal Origin*. West Sacramento: Science Technology System.
- Van Ginkel LA and Ruiter A (eds.) (2000) *Residues of Veterinary Drugs in Food, Proceedings of the EuroResidue IV Conference*, Veldhoven, The Netherlands, 8–10 May 2000. Bilthoven: National Institute of Public Health and the Environment (RIVM).
- Van Peteghem C (ed.) (1998) *Third International Symposium on Hormone and Veterinary Drug Residue Analysis, Proceedings of Symposium*, Bruges, Belgium, 2–5 June 1998. *The Analyst* 123(12): 2401–2801.
- Van Peteghem C (ed.) (2003) *Fourth International Symposium on Hormone and Veterinary Drug Residue Analysis, Proceedings of Symposium*, Antwerp, Belgium, 4–7 June 2002. *Analytica Chimica Acta* 473(1–2): 1–196; 483(1–2) 1–432.
- Watson DH (ed.) (2001) *Food Chemical Safety, Volume 1: Contaminants*. Cambridge: Woodhead Publishing.

## Water and Minerals

**M González**, Instituto Canario de Investigaciones Agrarias, La Laguna, Spain

**V González**, University of La Laguna, La Laguna, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Water

Water determination is one of the most important and frequent measurements in the processing and analysis of foods and responds to different necessities. The first is commercial because it is more profitable to buy products based on their dry weight. The second one is legal because for hygienic and commercial reasons the law limits the water content in many foods. The third reason is technological: several processes of industrial transformation need to know this value. Finally, the fourth reason is analytical because the food composition is generally expressed with respect to dry matter to facilitate comparison between samples.

The water content of foods varies enormously, as can be seen in **Table 1**. Water is the main component in most foods. It constitutes the medium in which chemical reactions occur and is one of the substrates in hydrolysis. So, eliminating water, or immobilizing it by increasing sugar or salt concentration, inhibits many reactions and the development of microorganisms, thereby increasing food shelf-life. Water also contributes significantly to food texture because of its physical interactions with proteins, polysaccharides, lipids, and salts.

Knowing the water content is not enough to evaluate food stability as foods with similar water content differ in their perishability. Because water activity is a very important factor in the stability and quality of foods, it is frequently measured along with water content. The water activity ( $a_w$ ) of a product in equilibrium with air, in an airtight container at a given temperature and pressure, is defined as the ratio of the partial pressure of water in the air in equilibrium with the product to the vapor pressure of pure water at the same temperature and pressure. Foods with  $a_w$  values between 0.2 and 0.4 are most stable (**Table 2**). The quality of these foods is practically unspoiled by microorganism development, nonenzymatic browning, or lipid autooxidation. Intermediate moisture foods ( $a_w$  value between 0.6 and 0.9) are protected against most of the alterations produced by microorganisms. However, most fresh foods have an  $a_w$  value of 0.97, being susceptible to alteration.

Water in foods is found in three different forms: free, absorbed, and bounded. The ease with which water is eliminated from food depends on the form of the water. To ensure similar results between different water determination methods, standard methods have been developed in which work conditions are specified. In each case the precision of the analytical results must be checked in order to decide if their precision can be substantially improved by using other methods.

### Methods for the Determination of Water in Food

When determining water some precautions must be taken to minimize the losses and gains of moisture that can be produced during sampling and sample preparation. In general, any exposure of the sample to open air and heating of the sample by friction during homogenization should be minimized. During storage the empty space in the sample container must be kept to a minimum because water is transferred to this space to equilibrate water content. It is also necessary to minimize temperature fluctuations because water migrates to the coldest part of the sample. To reduce this potential error, homogenization of the sample is required before its analysis.

Choosing an analytical method depends on the expected water content, volatility or sensitivity to heat of other food components, instrument availability, speed requirements, necessary accuracy, and aim of the analysis.

‘Water content’ is defined as the amount of water lost by a food when it reaches the true equilibrium against zero water vapor pressure. From this definition arises a quantification method of water content called the absolute reference method, a determination that is only possible in specialized laboratories. The food industry only uses practical reference methods, calibrated against the absolute reference method. Moreover, water content standards are not available because a product’s water content depends on the humidity and temperature of the environment, which makes participating in intercomparison exercises between laboratories essential for detecting possible experimental errors.

The Association of Official Analytical Chemists (AOAC) publishes reference methods for the analysis of moisture in foods. Some of these methods are summarized in **Table 3**.

**Table 1** Water, macroelement (calcium, magnesium, phosphorus, potassium, and sodium) and oligoelement (iron, zinc, copper, manganese, and selenium) content in some foods

Food	Amount in 100 g of edible portion									
	Water (g)	Ca (mg)	Fe (mg)	Mg (mg)	P (mg)	K (mg)	Na (mg)	Zn (mg)	Cu (mg)	Se ( $\mu$ g)
<i>Fruits, vegetables, legumes and derived products</i>										
Broccoli, raw	89.30	47	0.73	21	66	316	33	0.41	0.049	2.5
Bananas, raw	74.91	5	0.26	27	22	358	1	0.15	0.078	1.0
Bananas, dehydrated	3.00	22	1.15	108	74	1491	3	0.61	0.391	3.9
Oranges, raw, California, Valencias	86.34	40	0.09	10	17	179	0	0.06	0.037	—
Orange juice, raw	88.30	11	0.20	11	17	200	1	0.05	0.044	0.1
Potatoes, white, flesh and skin, raw	84.58	9	0.52	21	62	407	6	0.29	0.116	0.3
Tomatoes, red, ripe, raw	94.50	10	0.27	11	24	237	5	0.17	0.059	0.0
<i>Legumes</i>										
Beans, navy, mature seeds, sprouted, raw	79.15	15	1.93	101	100	307	13	0.89	0.356	0.6
Lentils, sprouted, raw	67.34	25	3.21	37	173	322	11	1.51	0.352	0.6
<i>Nuts</i>										
Almonds	5.25	248	4.30	275	474	728	1	3.36	1.110	2.8
Pistachio nuts	3.97	107	4.15	121	490	1025	1	2.20	1.300	7.0
<i>Cereals and derived products</i>										
Pasta, corn, dry	10.00	4	0.93	119	253	294	3	1.79	0.202	7.9
Rice, brown, long-grain, raw	10.37	23	1.47	143	333	223	7	2.02	0.277	23.4
Wheat flour, white, unenriched	11.92	15	1.17	22	108	107	2	0.70	0.144	33.9
Wheat flour, whole grain	10.27	34	3.88	138	346	405	5	2.93	0.382	70.7
Wheat, soft white	10.42	34	5.37	90	402	435	2	3.46	0.426	—
Bread, whole-wheat, commercially prepared	37.70	72	3.30	86	229	252	527	1.94	0.284	36.6
<i>Fats and oils</i>										
Butter, light, stick, with salt	42.10	48	1.09	5	34	71	450	0.06	0.000	1.0
Margarine, regular, stick, composite, 80% fat, with salt	17.17	3	0.12	1	5	18	654	0.11	0.000	0.0
Oil, vegetable, sunflower	0.00	0	0	0.03	0	0	0	0.00	0.000	0.0
<i>Beverages</i>										
Beer, regular	94.32	5	0.02	6	13	25	4	0.01	0.005	0.7
Carbonated beverage, cola	89.10	3	0.02	1	13	1	4	0.01	0.006	0.1
<i>Meat, poultry, and fish</i>										
Pork, fresh, leg (ham), whole, raw	62.47	5	0.85	20	199	315	47	1.93	0.065	29.4
Beef, chuck, arm pot roast, raw	57.92	7	2.01	18	166	289	57	3.81	0.077	15.9
Chicken, broilers or fryers, breast, meat only, raw	74.76	11	0.72	28	196	255	65	0.80	0.041	17.8
Egg, whole, raw, fresh	75.84	53	1.83	12	191	134	140	1.11	0.102	31.7
Fish, tuna, fresh, bluefin, raw	68.09	8	1.02	50	254	252	39	0.60	0.086	36.5
<i>Dairy products</i>										
Brie cheese	48.82	184	0.50	20	188	152	629	2.38	0.019	14.5
Milk, whole, 3.3% fat	88.32	101	0.03	10	84	133	43	0.38	0.023	3.7
Yogurt, plain, skim milk	85.23	199	0.09	19	157	255	77	0.97	0.015	3.6

**Absolute Reference Methods**

Among these methods, the Karl Fischer method is the most used. It is a chemical method based on iodine reduction by sulfur dioxide in the presence of water.

The sample is titrated with the Karl Fischer reagent, consisting of a mixture of sulfur dioxide, iodine, and pyridine in methanol. Because pyridine has an unpleasant odor and is toxic, it has been replaced by imidazole, and the methanol has been replaced by



methoxymethanol to stabilize the reagent. The method is precise because the water is determined specifically and selectively by chemical reaction. Moreover, this method determines both free and bounded water.

### Practical Reference Methods

This section examines oven-drying methods. In these methods the sample is heated under specific conditions and the weight loss is used to calculate the water content of the sample. For this determination three types of ovens can be used: convection, forced draft, and vacuum. The determined water quantity is highly dependent upon the type of oven used, the conditions inside the oven, and the time and temperature of drying.

Excess errors can be produced in these analyses due to the loss of volatile compounds different from

water (acids, alcohols, esters, and aldehydes) and to water formation in oxidation and nonenzymatic browning reactions. On the other hand, other errors can be produced as a consequence of not allowing enough time or temperature to evaporate all the water from the sample.

These methods are simple and can analyze many samples simultaneously. The analysis time can be between a few minutes and 24 h. Recently other thermic methods have been developed that use microwave ovens, infrared lamps, halogen lamps, or ceramic heating elements, which reduce analysis time.

To complete the analysis, distillation procedures must also be considered. These techniques consist in the codistillation of sample water with a high-boiling point solvent that is immiscible in water (toluene, xylene, benzene). The distilled mixture is collected, and the sample volume is measured. These methods produce fewer losses by food decomposition than does oven drying at high temperatures. However, measuring by water volume can be less exact than measuring by weight.

Table 4 compares some of the methods used to analyze water in food by principle, type of sample, and analysis time, including some methods based on the physical characteristics of water in food.

### Methods of Determination of Water Activity in Foods

Methods used to measure water activity can be direct and absolute or indirect and calibrated with regard to the former.

**Table 2** Typical water activity of some foods

<i>Food</i>	<i>Water activity (<math>a_w</math>)</i>
Fresh meat and fish	0.95–1.00
Bread	0.95
Cheese	0.90–0.95
Margarine	0.85–0.90
Salami	0.80–0.85
Jams and jellies	0.80
Dried fruits	0.70–0.80
Nuts	0.65–0.75
Honey	0.60–0.65
Cookies	0.30
Milk powder	0.20
Dried vegetables	0.20

**Table 3** Official AOAC analysis methods applicable to moisture

<i>AOAC method</i>	<i>Food</i>	<i>Analytical technique</i>
931.04	Cacao products	Gravimetry
977.10; 984.20; 967.19	Cacao products; oils and fats; dried vegetables	Karl Fischer titrimetry
935.29	Malt	Gravimetry – convection oven
920.116; 984.25	Butter; frozen potatoes	Convection oven
981.05	Malting barley	Convection oven – forced-draft oven
925.45	Sugars	Convection oven – vacuum oven
948.12	Cheese	Steam bath + forced-draft oven
926.07	Macaroni products	Forced-draft oven – vacuum oven
950.46	Meat	Air drying – vacuum oven
968.11; 979.12; 927.05; 926.08; 934.06; 962.12; 920.186; 977.21	Roasted coffee; dried milk; cheese; dried fruits; oils and fats; maple products; corn syrups and sugars	Vacuum oven
969.38	Honey	Vacuum oven – refractometry
977.11	Cheese	Microwave oven
985.14	Meat and poultry products	Rapid microwave oven
969.19; 986.21	Cheese; spices	Distillation
978.13	Milk	Vapor pressure osmometric (VPO) method
972.20	Prunes and raisins	Moisture meter



**Table 4** Comparison of methods used to determine water in foods

<i>Method</i>	<i>Principle</i>	<i>Nature of sample</i>	<i>Analysis time</i>
Karl Fischer	Chemical reaction of the water	All foods, especially foods very low in moisture or high in fats and sugars	5 min
Convection oven	Removal of water and weight of the remaining solids	Not valid for foods that contain volatiles or have components that undergo chemical reactions at high temperatures	0.75–24 h
Forced-draft oven	Removal of water and weight of the remaining solids	Not valid for foods that contain volatiles or have components that undergo chemical reactions at high temperatures	0.75–24 h
Vacuum oven	Removal of water and weight of the remaining solids		3–6 h
Distillation	Separation of water from the solids and volume measurement	Valid for foods that contain volatiles	
Dielectric and conductivity	Change in capacitance or resistance to an electric current that passes through the sample	Foods that contain less than 30–35% of water	
Hydrometry	Relationship between specific gravity and moisture content	Liquid samples: drinks and sugar solutions	
Refractometry	Water in the sample affects light refraction	Liquid samples, condensed milk, sugar solutions, fruits, and fruit products	
Infrared drying Near-infrared spectroscopy	Penetration of heat into the sample Specific absorption of near-infrared radiation (1400–1450, 1920–1950 nm) by the water molecules of the food		10–25 min

### Absolute Methods

These methods are based on measuring air temperature in equilibrium with a product and some other air characteristic like water content, wet bulb temperature, or dew temperature. With these data and the use of an enthalpy diagram of moist air or tables, its hygrometric level can be measured and thus the  $a_w$  value of the product in equilibrium with this air. There are chromatographic, psychometric, and manometric methods.

### Calibrated Methods

These methods use mechanical probes or hygrometers that respond to the variations in the relative humidity of air that are caused by variations in dielectric resistance or capacity. These probes directly measure the relative humidity of air, and the most commonly used is lithium chloride. In general, these methods require frequent calibration and should only be used within a narrow  $a_w$  range.

On the other hand, the reference salts method can indirectly determine the  $a_w$  value of a product using graphic interpolation. This method consists in introducing an aliquot of the sample in an airtight container where different saturated saline solutions or dilutions of sulfuric acid are placed, creating a relative humidity and therefore a known  $a_w$  value. Once equilibrium has been reached, the samples are weighed and the weight variations are represented

according to relative humidity. An interpolation of the curve obtained this way allows the determination of relative humidity and, by extension, the  $a_w$  value of the food.

### Importance of Minerals in the Diet and in Processing Food

Minerals are the inorganic elements constituting foods (excluding carbon, hydrogen, oxygen, and nitrogen) that remain as ashes when foods are incinerated. They are significant for their nutritional value, toxicological potential, and interaction with the texture and processing of foods. For these reasons it is necessary to know and control their concentration levels in foods.

Among the 50 known minerals, between 15 and 20 minerals are natural components of foods that are part of at least one vital biological system of a plant or animal. Some of them are denominated macroelements because of their abundance in foods; these include calcium, phosphorus, sodium, potassium, magnesium, and chlorine. Others are called oligoelements or trace elements due to their minimal concentration; among these are iron, iodine, zinc, copper, chromium, manganese, molybdenum, fluoride, and selenium.

Minerals are divided into two categories according to their biological importance: those with a

known biological role, called essential minerals, and those with an unknown biological role, referred to as nonessential minerals. Some minerals in the nonessential group are being investigated because there is some evidence that they have a biological function, although this biochemical role is not yet clear. The nonessential minerals being investigated are vanadium, tin, nickel, arsenic, and boron. There are also toxic elements, which should be avoided in the diet. This group includes lead, mercury, cadmium, and aluminum. Some essential minerals, like fluoride and selenium, are known to be harmful if consumed in excessive quantities, even though they have biochemical functions that are beneficial when consumed at safe levels.

Some minerals are found naturally in foods in variable amounts and can be modified, or others can be added, during industrial processing of food. **Table 1** indicates the amounts of some of the macro- and microelements that can be found in various raw and processed foods. Moreover, almost all these minerals can be found as pollutants in foods due to environmental or industrial factors, and are analyzed for hygienic reasons, and not as natural components of foods.

## Food Treatments Prior to Mineral Analysis

There are various methods of mineral determination available, and they can be used according to the analytical characteristic that best suits the objectives of the analyst: accuracy, sensitivity, detection limit, specificity, and interferences. Other factors to take into account are the costs of the complete analysis, instrumental availability, and time necessary for analysis.

Before mineral analysis it is usually necessary to treat the sample to ensure that the sample is homogeneous and also to prepare it for the analytical procedure that follows. Various processes may be necessary, but among the most important is sample mineralization, often associated with the need to destroy organic matter present in the sample and always necessary to make the sample soluble. Moreover, the treatment of a sample may entail reduction and homogenization of its size or elimination of interferences. In any case, contamination of the sample or loss of volatile compounds may occur during these steps of the analytical process, affecting the quality of the analytical results.

### Mineralization

Mineralization is usually a crucial step before the analysis of specific minerals because most analytical

methods require that minerals be freed from their matrix. Nevertheless, some analytical techniques, such as near-infrared spectroscopy, sometimes allow mineral estimation without destroying the carbon matrix that constitutes foods.

The destruction of the matrix, generally organic matter, can be done by wet or dry ashing. The selected mineralization procedure depends on what the ashes will be used for and on limitations based on cost, time, and number of samples.

Dry ashing, consisting in incinerating the sample at high temperatures (450°C or higher) in a muffle furnace, converts most minerals into oxides, sulfates, phosphates, chlorides, and silicates. Before measuring the analytes, the residuum is redissolved in a minimal amount of acid and is diluted with distilled water to a known volume. It is a safe method that does not require the addition of a reagent, and once ignited it requires little of the analyst's attention, allowing the treatment of many crucibles simultaneously. The main disadvantage of this method is the long treatment time (12–18 h). Moreover, some elements such as selenium, lead, and mercury can partially volatilize with this procedure. For this reason alternative methods must be used to determine these minerals.

In wet ashing, organic matter is oxidized by the addition of acids and oxidants or their combinations. A strong oxidant such as concentrated nitric acid, sulfuric acid, or perchloric acid is needed to destroy organic matter. In general, reagent mixtures are chosen in accordance with the type of food and the quantification method. Perchloric acid is an excellent oxidant, but its use runs the risk of forming explosive organic perchlorates if the mixture dries completely. Some precautions must be followed when fat-rich foods are treated.

Minerals remain in solution during wet digestion without volatilization losses because the temperatures used are lower than in dry ashing. The oxidation time is short and requires a hood, hot plate, long tongs, and safety equipment. However, this method requires the continual attention of the operator, uses corrosive reagents, and only a reduced number of samples can be treated simultaneously. If perchloric acid is used in the wet digestion, all work must be done in a perchloric acid fume hood.

Microwave radiation systems are now available that accelerate the mineralization process substantially. This process has obvious advantages in reducing decomposition time, acid needs, and the risks of contamination or foaming.

Other alternatives include closed reactor digestion at a high temperature and pressure; combustion in a vessel containing oxygen; the Shöniger method,

which is especially suitable when determining halogens and sulfur; or combustion in an oxygen–argon plasma at a low temperature.

### **Other Pretreatments**

Most solid foods must be ground up in order to obtain a homogeneity compatible with the previously described procedures. In this process it is necessary to adjust the grinding fineness according to the mineralization method that is going to be used. Dry ashing needs particles of diameter between 0.5 and 1 mm to avoid losses by mechanical drag during the incineration. Oxygen combustion methods and especially direct fusion methods need a grade of fineness that avoids particle volatilization losses and artificial enriching of samples.

According to their nature, the ground and homogenized samples are stored cold or at room temperature, in airtight containers made of a material that does not allow the transfer of materials between the container and the product. Generally, a polyethylene or Teflon container is a good solution.

On the other hand, mineralized and solubilized samples sometimes need treatments that eliminate interferences. Factors such as pH, sample matrix, some pretreatments, or the incorporation of a reagent can interfere with the ability of an analytical method to quantify a mineral. In this situation, it is common to turn to the use of masking reagents or separation processes using selective precipitation techniques, liquid–liquid extraction, ion-exchange resins, etc.

### **Sample Contamination**

One of the biggest problems that can occur during mineral analysis is contamination of the sample. The selection of the equipment used to treat the samples is closely related to the mineral that is going to be quantified. Stainless steel enriches the sample in chromium and nickel; agate mortars and mills notably increase the calcium content of the ground product, etc. Teflon has some important advantages, but due to its low resistance to abrasion not all treatments can be done using this material. Sometimes glass is inadequate; this must be taken into account when determining sodium because the glassware used may enrich the solution. Additionally, the repetitive use of glassware can be a contamination source, and for this reason glassware used in sample preparation and analysis must be cleaned scrupulously using acids and the purest water.

Since solvents can contain large amounts of minerals, it is necessary to use the purest reagents and distilled–deionized water to analyze minerals. As

reagents can be very expensive, a possible alternative is to work with a reagent blank or a sample of the reagents used in the sample analysis in the same proportion as in the sample but without the mineral that is going to be analyzed.

## **Methods of Determination of Minerals in Foods**

### **Chemical Methods**

The determination of food minerals can be done using chemical methods. Among them, volumetric methods stand out because they are fast and inexpensive while still being adequately precise. Their main disadvantages are that they have low sensitivity and selectivity.

Ethylenediaminetetraacetic acid (EDTA) complexometric titrations are based on the fact that many metallic ions form stable complexes with this tetradentate ligand, EDTA. The endpoints are detected using complexing agents capable of forming complexes with the metallic species to be determined and having lower coordination constants than those of the complexes that are formed with EDTA and that also have different colors in their free and complexed states.

In precipitation titrations the titration reaction produces an insoluble precipitate. Despite the many known precipitation reactions, very few of them have the necessary requirements to be the basis of a volumetric method. Determining chloride using silver ion precipitation, using several methodologies, is the most appropriate application of this group of titrations.

The most important applications of these methods in food analysis are limited to the determination of certain metals such as calcium in water and beer or chlorides in different types of samples. Table 5 lists some methods used to determine food minerals based on volumetry.

### **Molecular Absorption Spectrophotometry**

These methods, which include colorimetric methods, are based on the measurement of radiation absorption by molecular species at a specified wavelength. In mineral analysis, the absorbent species are usually compounds, mostly coordination complexes, formed in a reaction between the mineral and a chromogen ligand. These methods are more sensitive and selective than the previously discussed methods. Some of the methods that use ultraviolet (UV)–visible spectrophotometry can be combined with dynamic techniques like flow injection systems or with a previous separation using liquid chromatography,

**Table 5** Mineral determination using methods different from atomic spectroscopic methods

<i>Mineral</i>	<i>Foods</i>	<i>Pretreatments</i>	<i>Analytical technique</i>
<i>Titrimetry</i>			
Chloride	Oils	Dry ashing	Precipitation titrimetry
Calcium	Legumes	Wet ashing (HClO <sub>4</sub> :HNO <sub>3</sub> )	EDTA titrimetry
Calcium	Tubers	Lyophilization, dry ashing	EDTA titrimetry
<i>Colorimetry</i>			
Phosphorus	Eggs, vegetables		
Phosphorus	Oils, yogurts, infant foods, fishes, grains, nuts, vegetables, tubers, fruits	Dry ashing	
Phosphorus	Meat, legumes	Wet ashing (HNO <sub>3</sub> :H <sub>2</sub> SO <sub>4</sub> :HClO <sub>4</sub> )	
Phosphorus	Fish, grains, legumes	Wet ashing (HNO <sub>3</sub> :HClO <sub>4</sub> )	
Phosphorus	Grains	Wet ashing (20% HCl)	
Phosphorus	Nuts	Wet ashing (HNO <sub>3</sub> :H <sub>2</sub> SO <sub>4</sub> )	
Iron	Wine	Wet ashing (HNO <sub>3</sub> :HClO <sub>4</sub> )	
Iron	Tubers	Lyophilization, dry ashing	
Iodine	Dairy products	Dry ashing	
Boron	Fruits	Dry ashing	
<i>Electroanalytical techniques</i>			
Selenium and lead	Wine	Wet ashing (HNO <sub>3</sub> :HClO <sub>4</sub> )	Voltametry
Selenium	Vegetables	Wet ashing (HNO <sub>3</sub> :HClO <sub>4</sub> )	Voltametry
Fluorine	Grains		Voltametry
<i>Chromatography</i>			
Sodium, potassium, magnesium, and calcium	Dairy products	Dry ashing	Ionic chromatography
Selenium	Nuts	Wet ashing (HNO <sub>3</sub> :H <sub>2</sub> SO <sub>4</sub> )	Gas chromatography – mass spectrometry
<i>Other techniques</i>			
Sulfates	Oils	Dry ashing	Turbidimetry
Manganese, bromine, cobalt, vanadium, arsenic, antimony, copper, selenium, aluminum, and lanthanum	Tubers		Neutron activation analysis

allowing the quantification of many elements simultaneously. **Table 5** summarizes some of the main colorimetric methods used regularly in mineral analysis.

### Atomic Spectroscopy

These techniques are based on the atomization of the analyte present in a solution, using a flame or a plasma. The amount of an element present in the sample is determined from the absorption or emission of the visible or UV radiation of its atoms in the gaseous phase. The high sensitivity and selectivity that can be achieved with these techniques and the increasing necessity to respond to requests for mineral quantification in foods at trace levels explain the increasing use of some of the atomic absorption and emission spectroscopic techniques. **Table 6** summarizes some

methods applicable to the determination of minerals using atomic spectroscopy.

Among atomic emission spectrometry (AES) methods, the classic flame photometric technique is still favored for determination of sodium and potassium in foods.

The use of an inductively coupled plasma (ICP) allows temperature and stability conditions that eliminate most of the interferences found in combustion sources. The sensitivities that can be obtained and the speed of this technique, which can be used to determine several elements simultaneously, make it in an interesting, although expensive, alternative for the analysis of metals found in foods. Coupling ICP with a mass spectrometer gives the best analytical results, although it is currently a technique restricted to specialized laboratories.

**Table 6** Mineral determination using atomic spectroscopic techniques

<i>Mineral</i>	<i>Foods</i>	<i>Pretreatments</i>
<i>Flame photometry</i>		
Sodium, potassium and calcium	Honey	Infrared lamp drying, dry ashing
Sodium and potassium	Oils, vegetables, fruits, yogurts, fish	Dry ashing
Sodium, potassium, and calcium	Grains	
Sodium and potassium	Meat	Wet ashing ( $\text{HNO}_3\text{:H}_2\text{SO}_4\text{:HClO}_4$ )
<i>Flame atomic absorption spectrometry</i>		
Sodium, potassium, zinc, iron, calcium, and copper	Honey	Wet ashing ( $\text{H}_2\text{SO}_4\text{:HNO}_3$ )
Zinc, cadmium, and lead	Honey	Microwave drying, dry ashing
Iron	Wine	Wet ashing ( $\text{HClO}_4\text{:HNO}_3$ )
Iron	Wine	Without treatment
Manganese, chromium, iron, nickel, and copper	Juices	Wet ashing ( $\text{H}_2\text{SO}_4\text{:HNO}_3$ )
Cadmium, copper, chromium, cobalt, iron, nickel, lead, and zinc	Beverages, dairy products	Dry ashing
Copper, iron, and zinc	Milk	Wet ashing ( $\text{HNO}_3\text{:H}_2\text{O}_2$ )
Copper, iron, zinc, and manganese	Yogurts	Dry ashing
Calcium, magnesium, and zinc	Yogurts	Dry ashing
Copper and zinc	Vegetables, fruits, meat, fish, legumes, cereals, spices, dairy products, tubers, sweeteners, canned foods	Dry ashing
<i>Copper and zinc</i>		
Sodium, potassium, calcium, magnesium, iron, zinc, manganese, and copper	Fats, oils	Wet ashing ( $\text{HNO}_3\text{:HClO}_4$ )
Cadmium and lead	Fish	Wet ashing ( $\text{HNO}_3\text{:HClO}_4$ )
Iron, zinc, aluminum, titanium, and vanadium	Fish	Wet ashing (30% $\text{H}_2\text{O}_2\text{:HNO}_3$ )
Selenium, lead, and cadmium	Baby foods	Wet ashing ( $\text{HNO}_3\text{:H}_2\text{SO}_4$ ) 0.1% Triton X-100 + 1% $\text{HNO}_3$ + 30% $\text{H}_2\text{O}_2$
Cadmium, cobalt, chromium, copper, iron, manganese, molybdenum, nickel, lead, and zinc	Seafood	Wet ashing ( $\text{HNO}_3$ , $\text{H}_2\text{O}_2$ )
Cadmium, calcium, iron, magnesium, manganese, and zinc	Meat	Wet ashing ( $\text{HNO}_3\text{:H}_2\text{SO}_4\text{:HClO}_4$ )
Lead, cadmium, iron, copper, manganese, zinc, and cobalt	Mushrooms	Wet ashing ( $\text{HNO}_3\text{:H}_2\text{SO}_4\text{:H}_2\text{O}_2$ )
Copper, magnesium, lead, sodium, silver, bismuth, manganese, nickel, lithium, cobalt, antimony, calcium, zinc, iron, chromium, aluminum, and potassium	Mushrooms	Wet ashing ( $\text{HNO}_3\text{:HClO}_4\text{:HCl}$ )
Potassium, sodium, calcium, magnesium, iron, manganese, copper, and zinc	Vegetables	Wet ashing ( $\text{HNO}_3$ )
Sodium, potassium, magnesium, zinc, and copper	Tubers	Lyophilization, dry ashing
Calcium, magnesium, iron, zinc, copper, and manganese	Fruits	Microwave drying, wet ashing ( $\text{HNO}_3$ : 30% $\text{H}_2\text{O}_2$ )
<i>Inductively coupled plasma</i>		
Aluminum, barium, calcium, copper, iron, potassium, iron, copper, iron, potassium, magnesium, manganese, sodium, strontium, and zinc	Beverages	Without treatment
Calcium, cobalt, copper, chromium, iron, magnesium, manganese, molybdenum, sodium, zinc, silver, aluminum, arsenic, barium, beryllium, cadmium, mercury, nickel, lead, antimony, tin, strontium, titanium, thallium, uranium, and vanadium	Milk, flour	Microwave drying, dry ashing
Calcium, phosphorous, magnesium, sodium, potassium, aluminum, iron, zinc, and copper	Seafood, meat	Dry ashing
Chromium, iron, manganese, selenium, and zinc	Vegetables	Wet ashing ( $\text{HNO}_3\text{:HClO}_4$ )
Aluminum, boron, calcium, iron, magnesium, phosphorus, potassium, sodium, and titanium	Fruits, fruit products	Microwave drying, wet ashing ( $\text{HNO}_3$ )
Hydride/cold vapor		
Selenium	Flour, flour products	Wet ashing ( $\text{HNO}_3\text{:HClO}_4\text{:H}_2\text{SO}_4$ )
Selenium	Milk	Wet ashing ( $\text{HNO}_3\text{:HClO}_4$ )
Mercury	Fish	Wet ashing ( $\text{HCl}$ , $\text{HNO}_3$ , $\text{H}_2\text{SO}_4$ )
Mercury	Mushrooms	Wet ashing ( $\text{HNO}_3$ )



Atomic absorption spectrometry (AAS) is the most common technique for the determination of metals in food products.

Most commonly used instruments use a flame (flame AAS (FAAS)) produced by combustion of an air/acetylene or dinitrogen oxide/acetylene mixture. The few interferences are easy to avoid, and the sensitivities that are reached are adequate for the metals of greatest interest to the food industry. Variants of this technique, such as the coupling of hydride generation (HG) systems (HG-AAS), increase its scope to higher-sensitivity determination of elements like selenium, arsenic, tin, and other elements that form hydrides. In a similar vein, the determination of mercury using the cold vapor technique should be highlighted.

An alternative to FAAS, particularly suited to the determination of metals with higher sensitivity, is electrothermal AAS (ET-AAS), in which a graphite furnace substitutes the flame as the atomization source. Generally speaking, the detection limits obtained with a graphite furnace are two orders of magnitude lower than those reached with flame atomization.

Comparing ET-AAS and ICP, techniques with similar scopes in food analysis, confirms that absorption techniques require less expensive instruments and usually achieve higher sensitivities. Since ICP is a particularly fast method, a laboratory that has to analyze many metals per sample and/or many samples can justify the investment and maintenance associated with ICP.

Other intriguing spectrometric methods used in metallic species analysis, such as X-ray fluorescence spectrometry and neutron activation analysis, are unusual in food analysis.

### Electrochemical Methods

In food analysis, the most common electrochemical methods are potentiometric and voltametric.

In potentiometric methods, the potential between a reference and an indicator electrode is measured, which corresponds to the analyte activity. Because of their usefulness in food analysis, ion-selective electrodes (ISEs) that measure anions like bromide, chloride, and fluoride or cations like potassium, sodium, and calcium stand out among indicator electrodes. The characteristics and advantages of ISE include the ability to measure different anions and cations directly, the fact that they do not consume the analyte, the fact that analyses are independent of sample volume when taking direct measurements, and that moreover turbidity, color, and viscosity do not affect the measurement. Potentiometric methods are also

fast, easy to use, and inexpensive. The disadvantages of ISEs include the following: they have a relatively low sensitivity, proteins or other organic solutes can interfere in the determination, and some ions can act like ligands or can poison the electrodes.

Voltametric techniques are based on the relation between current and voltage in an electrochemical process. Among them, anodic stripping voltammetry permits metallic species determination with detection limits of parts per billion or lower. The equipment used with these techniques is much more inexpensive than that used with spectroscopic techniques that are also used in trace analysis.

Table 5 summarizes some of the electroanalytical methods usually employed in mineral analysis of foods.

### Separation Methods: Electrophoretic and Chromatographic Methods

For multielemental analysis of foods, separation methods are also used, although less frequently. Capillary electrophoresis is used to separate metallic cations like sodium, potassium, calcium, manganese, and magnesium. Among chromatographic methods, determination of anionic and cationic species using ionic chromatography, with or without the use of an ionic suppressor, and conductimetric detection should be highlighted. Table 5 reviews some separation methods applicable to the determination of minerals.

The AOAC publishes reference methods for the analysis of minerals in foods. Table 7 summarizes some of the AOAC official methods of analysis applicable to minerals.

### Analytical Quality Assurance

Quality control is of the utmost importance in the case of mineral analyses because of the low concentrations of the elements normally found in foods and the ubiquitous presence of significant levels of many of them in the environment. In addition to the standard techniques of working in a clean laboratory to reduce the potential for accidental contamination to a minimum, it is essential that procedures be validated and results checked against appropriate certified reference materials (CRMs). CRMs for most of the trace and other minerals of interest in foods are available from international reference centers such as the Community Bureau of Reference of the European Union, the National Institute of Standards and Technology of the United States, and the International Atomic Energy Agency in Vienna.

**Table 7** Official AOAC analysis methods applicable to the determination of minerals in foods

<i>AOAC method</i>	<i>Mineral</i>	<i>Food</i>	<i>Chemical reaction</i>
<i>Titrimetry</i>			
976.09; 976.10	Calcium	Beer	EDTA + eriochrome black T
944.03	Calcium	Flour	Oxalic acid + bromocresol green
948.09	Phosphorus	Flour	Molybdate + phenolphthalein
983.19; 968.31	Calcium	Poultry and beef; canned vegetables	EDTA + hydroxyl naphthol blue
960.29	Chloride	Butter	Mohr method
915.01	Chloride	Plant material	Volhard method
<i>Colorimetry</i>			
944.02; 945.40; 950.39; 955.21	Iron	Flour; bread; macaroni products; beer	<i>o</i> -Phenantroline
960.17; 967.17	Copper	Beer; distilled liquors	Zinc dibenzylthiocarbamate
970.18; 972.12	Copper	Wines; beer	Diethanolamine + carbon disulfide
955.19; 962.11; 970.39; 986.24; 991.27; 991.25	Phosphorus	Distilled liquors; wines; fruits; milk-based infant formulas; meat and meat products; cheese	Molybdovanadate
<i>Flame photometry</i>			
963.08; 965.30	Potassium	Distilled liquors; fruit and fruit products	—
963.09; 966.16	Sodium	Distilled liquors; fruit and fruit products	—
963.13; 969.23; 990.23	Potassium and sodium	Wines; seafood; dried milk	—
<i>AAS</i>			
967.08; 970.18	Copper	Distilled liquors; wines	—
970.19	Iron	Wines	—
972.06	Aluminum	Baking powders	—
987.02	Potassium	Beer	—
987.03	Sodium	Beer	—
985.35	Minerals	Milk-based infant formulas	—
991.25	Calcium and magnesium	Cheese	—
<i>ICP emission spectroscopy</i>			
984.27	Calcium, copper, iron, magnesium, manganese, phosphorus, potassium, sodium, and zinc	Infant formulas	—

See also: **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Principles and Instrumentation. **Elemental Speciation:** Overview. **Food and Nutritional Analysis:** Sample Preparation. **Ion-Selective Electrodes:** Overview. **Quality Assurance:** Reference Materials. **Sample Dissolution for Elemental Analysis:** Dry Ashing; Oxygen Flask Combustion; Wet Digestion; Microwave Digestion. **Spectrophotometry:** Inorganic Compounds. **Titrimetry:** Overview.

## Further Reading

- Belitz HD and Grosch W (eds.) (1999) *Food Chemistry*, 2nd edn. Berlin: Springer.
- Nielsen SS (ed.) (2003) *Food Analysis*, 3rd edn., pp. 81–100, 103–111, 191–202. New York: Kluwer Academic/Plenum Publishers.
- Horwitz W (ed.) (2000) *Official Methods of Analysis of the Association of Official Analytical Chemists*, 17th edn. Arlington: AOAC International.
- Macrae R, Robinson RK, and Sadler MJ (eds.) (1993) *Encyclopedia of Food Science, Food Technology and Nutrition*. San Diego: Academic Press.
- Pomeranz Y and Meloan CE (1994) *Food Analysis: Theory and Practice*, 3rd edn., pp. 575–624. New York: Chapman & Hall.
- USDA Agricultural Research Service (2003) USDA nutrient database for standard reference. [http://www.nal.usda.gov/fnic/cgi-bin/nut\\_search.pl](http://www.nal.usda.gov/fnic/cgi-bin/nut_search.pl).

## Pesticide Residues

**K Mastovska**, Drexel University, Philadelphia, PA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

According to the globally adopted definition of the Food and Agriculture Organization (FAO) of the United Nations (UN), ‘pesticide means any substance or mixture of substances intended for preventing, destroying, attracting, repelling or controlling any pest including unwanted species of plants or animals during the production, storage, transport, distribution and processing of food, agricultural commodities, or animal feeds or which may be administered to animals for the control of ectoparasites. The term includes substances intended for use as a plant growth regulator, defoliant, desiccant, fruit thinning agent, or sprouting inhibitor and substances applied to crops either before or after transport’. Pesticides can be divided into several groups based on their specific biological effect on target pest species. Table 1 presents major chemical classes of the most important groups of pesticides, insecticides, fungicides, and herbicides, which are used against insects, fungi, and weeds (and other plants that grow where they are not wanted), respectively.

Although intended to control pests and weeds, pesticides can also be toxic to nontarget organisms, including humans. Real and perceived concerns about the impact of pesticides on human health and the environment have led to strict regulation of both their application and residue levels in food and water supplies. Maximum residue limits of pesticides in food commodities and drinking water are commonly set by national regulatory authorities and international bodies, such as the Codex Alimentarius. For regulatory purposes, pesticide residue definitions include the parent compound and any specified derivatives such as degradation and conversion products, metabolites, and impurities considered to be of toxicological significance.

Suitable (reliable and cost-effective) analytical methods are needed mostly in order to enforce pesticide residue regulations, to provide toxicological risk assessment data, and to study pesticide fate in the food chains and the environment.

### Analytical Strategies and Trends in Pesticide Residue Analysis

In a large majority of cases, multiresidue methods (MRMs), capable of simultaneously determining

more than one residue in a single analysis, usually provide the most efficient approach to pesticide residue analysis. The reasons for their application include: (1) the need for analysis of a large (and continuously increasing) number of pesticides that should be determined together with their toxicologically important metabolites/degradation products; (2) usually unknown ‘history’ of the analyzed sample, i.e., unavailable information on the field treatment and/or postharvest application of pesticides, as well as, on the character of contamination in the locality of the sample origin; and (3) economic aspects (costs, time, and labor). A possible drawback of MRMs represents a potential worsening of some parameters such as accuracy (both trueness and precision), sensitivity (limit of detection, LOD), selectivity, and ruggedness for some individual analytes. Any MRM must be carefully optimized to provide results that meet specific (fit-to-purpose) requirements for each individual analyte, thus often compromising between the analytical scope of the method and the quality of the results for all analytes.

In practice, MRMs consist of sample preparation followed by determination of residues. The sample preparation step usually involves (1) sample homogenization, (2) isolation of residues from a representative sample (extraction), and (3) separation of residues from coextracted matrix components that interfere in the determinative step (clean-up). The determinative step comprises analyte identification and quantification, which is ideally carried out by a confirmatory technique (elucidating analyte structure), otherwise a confirmation step should follow.

Gas chromatography (GC) is traditionally a prevailing determinative technique employed in multiresidue pesticide analysis. The increasing availability of GC in combination with mass spectrometry (MS) leads to continuous replacement of GC with conventional detectors by GC–MS for simultaneous quantitative and confirmatory analysis covering a wider range of analytes. However, many pesticide residues are not directly amenable for GC and their permanently increasing number reflects a trend in pesticide product development, i.e., transition from the use of persistent and less polar compounds to more readily degradable, more (sometimes very) polar, low volatile, and/or thermolabile pesticides. Determination of these ‘modern’ pesticides and their metabolites has been rather difficult until recent advances in liquid chromatography–mass

**Table 1** Major chemical classes of insecticides, fungicides, and herbicides

<i>Pesticide group</i>	<i>Chemical class</i>	<i>Representatives</i>
Insecticides	Carbamate	Aldicarb, carbaryl, carbofuran, fenoxycarb, methiocarb, methomyl, oxamyl, pirimicarb
	Organochlorine	Chlordane, DDT, dicofol, endosulfan, endrin, heptachlor, lindane, methoxychlor
	Organophosphorus	Acephate, azinphos-methyl, chlorpyrifos, diazinon, dimethoate, malathion, parathion, phosmet
	Neonicotinoid	Acetamiprid, imidacloprid, thiamethoxam, clothianidin, dinotefuran, nitenpyram
	Pyrethroid	Allethrin, bifenthrin, cyfluthrin, cyhalothrin, cypermethrin, deltamethrin, fenvalerate
Fungicides	Chlorine-substituted aromatics	Chlorothalonil, dicloran, quintozone (PCNB), tecnazene (TCNB)
	Dithiocarbamate	Ferbam, mancozeb, maneb, metiram, propineb, thiram, zineb, ziram
	Dicarboximide	Chlozolinate, iprodione, procymidone, vinclozolin
	Imidazole	Benomyl, carbendazim, imazalil, prochloraz, thiabendazole, thiophanate-methyl
	<i>N</i> -trihalomethylthio	Captan, captan, folpet, dichlofluanid, tolylfluanid
Herbicides	Triazole	Bitertanol, fenbuconazole, hexaconazole, metconazole, myclobutanil, penconazole
	Aryloxyalkanoic (chlorophenoxy) acid	2,4-D, 2,4-DB, dichlorprop, MCPA, MCPB, mecoprop, 2,4,5-T
	Aryloxyphenoxypropionate	Cyhalofop-butyl, diclofop-methyl, fenoxaprop-ethyl, fluazifop-butyl, haloxyfop-methyl
	Chloroacetamide	Acetochlor, alachlor, metazachlor, metolachlor
	Dinitroaniline	Benfluralin, dinitramine, pendimethalin, trifluralin
	Imidazolinone	Imazamox, imazapyr, imazaquin, imazethapyr
	Quaternary ammonium	Diquat, paraquat
	Phenylurea	Chlorotoluron, diuron, fenuron, isoproturon, linuron, metoxuron, monolinuron, neburon
	Sulfonylurea	Chlorimuron-ethyl, chlorsulfuron, metsulfuron-methyl, sulfometuron-methyl, triasulfuron
	Thiocarbamate	Butylate, cycloate, EPTC, molinate, pebulate, thiobencarb, tri-allate, vernolate
	Triazine	Ametryn, atrazine, cyanazine, prometryn, propazine, simazine, terbutryn
	Uracil	Bromacil, lenacil, terbacil

spectrometry (LC–MS) brought this powerful technique to routine laboratories and opened the door to direct, selective, and sensitive multiresidue analysis of these analytes.

Ideally, all existing pesticide residues would be analyzed by a single method; however, no current method is capable of that. Even if we neglect a much smaller group of inorganic pesticides that require different analytical methodology, the major group of organic pesticides still represents compounds of rather diverse physicochemical properties, mainly in terms of polarity, volatility, and stability (to pH, temperature, etc.) that make the ideal ‘all residue’ method impossible in practice. Based on the chemical diversity of analytes, MRMs can be further divided into (1) ‘multiclass’ MRMs (MMRMs) and (2) ‘single class’ or ‘selective’ MRMs (SMRMs), involving multiple residues from different pesticide classes or only chemically related pesticides (from

one class), respectively. Most of the SMRMs that include derivatization of specific analytes to enhance their detection sensitivity and selectivity in LC (e.g., postcolumn derivatization of *N*-methylcarbamates for fluorescence detection) or volatility and stability (and also sensitivity) for GC analysis (e.g., methylation or pentafluorobenzoylation of chlorophenoxy acid herbicides, pentafluorobenzoylation of carbendazim) are being replaced by MMRM(s) employing LC–MS in the determinative step. However, some special procedures, such as conversion of dithiocarbamates to carbon disulfide for head-space GC analysis, are still conveniently used in current practice. In some cases, the determinative step may be based on a multiclass concept, but some of the parts of the sample preparation step may be selective and vice versa. This article reviews the main sample preparation and determinative techniques applicable in analysis of pesticide residues in food matrices with

a particular emphasis on the multiclass, multiresidue strategy.

## Sample Preparation

### Sample Types

The selection of sample preparation methodology is highly dependent on both analyte and sample (matrix) characteristics. To approach the matrices in the same cost-effective ('multi') way as analytes, food, commodities, and in relation to that also sample preparation methods, are primarily classified based on the fat content; the limit between 'fatty' and 'nonfatty' food matrices being usually set at 2%.

Nonfatty foods of plant origin represent the most frequently analyzed samples for pesticide residues. Among this group, special attention is paid to fruits and vegetables that are mostly consumed raw or after minimum postharvest processing, thus potentially containing the highest residue levels and posing the highest risk to consumers. Furthermore, fruits and vegetables represent the main ingredients in products intended for infants and small children ('baby food') in which pesticide residues are especially strictly regulated and monitored (e.g., pesticide residues in baby food should not exceed 0.010 mg per kg in most European countries).

In the case of fatty foods, the analysis is commonly aimed at lipophilic pesticides, such as organochlorines (OC), that accumulate in lipidic tissues and may be persistent, posing a great risk to both consumers and the environment.

### Sample Homogenization

The analytical sample is generally a part (subsample) of a larger amount of material that arrives to a laboratory as a result of a sample collection procedure. Thorough homogenization of the original laboratory sample is necessary for obtaining a representative subsample of as low as practicable size (mostly less than 50 g). Samples should be comminuted at low temperatures reducing the risk of analyte losses via degradation or evaporation. Furthermore, milling/grinding of frozen samples at extremely low temperatures, i.e., cryogenic milling in the presence of dry ice (solid carbon dioxide) or cooled by liquid nitrogen, results in fine, highly homogeneous powder that can be easily subsampled.

### Extraction

Table 2 provides a brief overview of extraction techniques applicable to isolation of pesticide residues from food samples. Currently, classical solvent extraction (at room temperature and normal pressure)

represents the most popular extraction approach due to its simplicity, relatively low costs (mainly of investment), wide applicability to both solid and liquid food samples, and, when applied properly, a low risk of analyte losses due to degradation, evaporation, and/or irreversible adsorption. Most current MMRMs employ acetonitrile, ethyl acetate, acetone, or methanol for efficient isolation of a wide range of pesticide residues from nonfatty samples. Acidic or basic analytes may require pH adjustment. In special instances, such as isolation of quaternary ammonium herbicides (diquat, paraquat), a hydrolytic extraction using concentrated acids is applied to break strong matrix-analyte interactions. Ultrasonic extraction (USE) can facilitate extraction of analytes with less strong, yet still relatively high affinity to the matrix. Sonication is also a simple and convenient approach; however, losses of more volatile and less thermally stable pesticide residues may occur due to the increased extraction temperature.

In the case of fatty samples, the target analytes usually accumulate in fat and their selective isolation from lipidic compounds is not feasible due to the similar solubility in the extraction solvent. Thus, the analytical strategy usually involves initial isolation of fat from the matrix using less polar solvents (e.g., hexane or hexane-acetone mixture) followed by clean-up step(s) separating analytes from the lipidic fraction. Soxhlet extraction (SOX) represents a traditional, reference method for isolation of fat from various food samples. SOX is more efficient than a blending procedure; therefore, it can also be used in applications that require enhanced extraction power for overcoming analyte-matrix interactions. The major drawback of this method is its long time consumption (up to 24–48 h), which is the main reason for replacing SOX by more rapid techniques in most modern laboratories. Less time- and also solvent-consuming alternatives to SOX (and also to USE) mainly include (1) microwave-assisted (solvent) extraction (MAE or MASE) and (2) pressurized liquid (or solvent) extraction (PLE or PSE), which is often referred by its commercial name: accelerated solvent extraction (ASE<sup>TM</sup>). Nevertheless, like SOX and other procedures performed at elevated temperatures, the application of these techniques is limited to relatively stable analytes.

In PLE, the increased pressure (up to 20 MPa) helps to keep extraction solvents in a liquid state at temperatures above their atmospheric boiling points, taking advantage of their enhanced solvation power and lower viscosities and hence higher diffusion coefficients. Based on a similar principle, supercritical fluid extraction (SFE) employs even more penetrative and solvating supercritical fluids. CO<sub>2</sub> is most



**Table 2** Extraction techniques applicable in the analysis of pesticide residues in foods

<i>Extraction technique</i>	<i>Basic principle</i>	<i>Materials/chemicals and devices</i>
Classical solvent extraction	Blending (mixing, shaking) with a solvent at room temperature and normal pressure	Organic solvent(s); blenders, probes, shakers, or manual shaking
SOX	Continuous extraction at elevated temperatures with repeatedly redistilled (fresh) solvent	Organic solvent(s); dispersant, drying agent, or sorbent permeable extraction vessel (thimble) Soxhlet apparatus
USE	Blending with a solvent in the presence of ultrasound waves	Organic solvent(s); ultrasonic bath or probe
MAE	Blending with a solvent heated by microwave energy	Organic solvent(s); special extraction containers – a special microwave oven
PLE	A combination of static and dynamic (flow-through) solvent extraction at a higher pressure and usually also elevated temperatures	Organic solvent(s); dispersant, drying agent, or sorbent Special extraction vessel PLE system
SFE	A combination of static and dynamic extraction using supercritical fluids	Supercritical fluid, modifiers dispersant, drying agent, or sorbent special extraction vessel SFE system
SPE	Adsorption of analytes from liquid samples (solutions) on a sorbent followed by selective elution of analytes using suitable solvent(s)	Sorbent (usually in a cartridge or disk); solvents for conditioning, washing, and elution special device for the flow control (vacuum manifold, automatic SPE system)
SPME	Adsorption of analytes from liquid samples (solutions) on a coated fiber followed by their desorption in the GC or LC injection ports	Fused-silica fiber coated with a sorbent
SBSE	Adsorption of analytes from liquid samples (solutions) on a coated stir bar followed by their thermal desorption for GC analysis	Glass-linked magnetic stir bar coated with a sorbent thermodesorption unit
MSPD	Dispersion of a sample with a sorbent, packing of this mix in a column followed by a selective elution of analytes	Sorbent empty column/cartridge elution solvent(s)

commonly used as a supercritical fluid in SFE mainly due to its suitably low critical point (31.2°C and 7.4 MPa), safety, relatively low cost, and presence in a gaseous state at normal conditions. The first feature may offer mild extraction conditions for less stable analytes and the latter one enables an easy removal of the extraction medium (by reducing the pressure), thus a convenient concentration of analytes prior to the determinative step. Compared to the techniques employing liquid extraction solvents, SFE allows a higher degree of selectivity through better control of physicochemical properties of the supercritical fluid. In MMRMs, however, increased extraction selectivity may turn into a disadvantage, limiting the range of pesticide residues that can be isolated under the same conditions. This and other limitations, including high investment and operating costs, a relatively small sample size, and demanding method optimization, precluded a widespread implementation of SFE in pesticide residue analysis and other applications.

Extraction of pesticide residues from liquid samples can be performed using a solid sorbent material. Currently available sorbent extraction techniques include: (1) solid-phase extraction (SPE), (2) solid-phase microextraction (SPME), and (3) stir-bar sorptive extraction (SBSE). In the case of solid samples, a liquid extraction of pesticide residues (transfer into a solution) usually precedes the sorption step; thus, it should be considered rather as a clean-up than an extraction. Matrix solid-phase dispersion (MSPD) represents a unique SPE approach that combines extraction and clean-up of solid or semisolid food samples in one step. In MSPD, the sample is mixed with a sorbent (Florisil, C<sub>18</sub>, C<sub>8</sub>) that serves as a solid support in sample disruption and dispersion. The resulting mix is packed into a column from which the analytes are eluted while separated matrix components are retained by the sorbent. The main drawbacks of this approach comprise rather small sample sizes (~0.5 g) and a relatively high consumption of expensive sorbents.

SPE dominates in extraction of pesticides from water samples and it represents presumably the most flexible technique for clean-up of food extracts. SPME and SBSE use a fused-silica fiber or a stir bar, respectively, coated by a sorptive material (usually polydimethylsiloxane, PDMS, or its modifications) for partition (i.e., not complete isolation) of analytes. These techniques are convenient and fast; however, their application in quantitative MMRMs for analysis of pesticides in complex food samples is rather limited. The main disadvantages relate to strong matrix effects (matrix-dependent partition of analytes and deterioration of the coating by irreversible adsorbed matrix components), insufficiently wide polarity range to extract diverse pesticide residues, and variability of method sensitivity for different analytes depending on their partition coefficients.

### Cleanup

A highly selective extraction isolating only the analytes from the sample is hardly feasible in practice,

especially in MMRMs. Thus, food sample extracts contain various matrix components posing serious difficulties (generally called 'matrix effects') in the determinative step. The cleanup step aims to remove matrix coextractives without impairing analyte recoveries. Similarly, a thorough cleanup is not feasible in MMRMs; however, some cleanup is generally better than none because it improves method accuracy, ruggedness, and potentially decreases LODs. Table 3 gives an overview of basic cleanup techniques used for separation of pesticide residues from coextracted food components. The widely applicable approaches in MMRMs include: (1) liquid-liquid partitioning (LLP), (2) gel permeation chromatography (GPC), and (3) SPE. Stable analytes, mainly OC pesticides, can tolerate applications of more aggressive procedures, such as sweep co-distillation (SCD) or chemical (degradative) cleanup, principally serving for removal of fat from final extracts.

LLP represents a cost-effective and easy approach utilized in many MMRMs despite its imperfections including mainly the use of large volumes of

**Table 3** Basic clean-up techniques used for separation of pesticide residues from coextracted food components

<i>Clean-up technique</i>	<i>Basic principle</i>	<i>Separated matrix components</i>
LLP	Separation of compounds based on their diverse solubility in immiscible solvent layers	Compounds more polar than pesticides that separate from pesticides in the organic phase to the aqueous layer; e.g., removal of sugars and some other polar coextractives together with water from acetone and acetonitrile extracts by LLP induced by addition of nonpolar solvent, such as dichloromethane, or salts (potential loss of more polar pesticides) Compounds less polar than pesticides that separate from pesticides in a more polar layer to a less polar one; e.g., removal of lipids from acetonitrile extracts by LLP with hexane (loss of lipophilic pesticides) Ionic compounds from neutral pesticides or neutral compounds from ionic pesticides by pH adjustment in the aqueous layer
GPC	Separation of molecules by their size	Compounds larger than pesticides, mainly pigments and lipids (usually removal of low-volatile components prior to the GC analysis)
SPE	Separation of compounds based on their diverse sorption power on/affinity to the used sorbent(s)	Compounds with high affinity from pesticides with no or lower affinity to the sorbent = chemical filtration or selective elution of analytes Compounds with no or lower affinity from pesticides with higher affinity to the sorbent = separation of matrix components during sorption or washing steps
SCD	Separation of compounds based on their diverse volatility	Compounds less volatile than pesticides that undergo thermal cracking while pesticides are carried to a trap by a stream of nitrogen (mainly separation of lipids from thermally stable pesticides)
Chemical treatment	Conversion of matrix components into compounds no longer soluble in the final extract or easily removable from it	Compounds (usually lipids) that undergo degradative reactions, such as mineralization by concentrated sulfuric acid or saponification by hydroxides

potentially hazardous solvents. The widely applied Luke and Specht methods (the official US Food and Drug Administration and German DFG-S19 methods, respectively), using dichloromethane or dichloromethane–petroleum ether to remove water and polar coextractives from primary acetone extracts (with certain sacrifice in recoveries of more polar pesticides), may serve as typical examples. The recently introduced QuEChERS (QuEChERS stands for Quick, Easy, Cheap, Effective, Rugged, and Safe) method avoids the use of nonpolar solvents inducing LLP by addition of  $\text{MgSO}_4$  and  $\text{NaCl}$  to acetonitrile extracts, which leads to removal of majority of water and highly polar matrix components, yet achieving high recoveries of wide range of both GC- and LC-amenable pesticides (additional SPE clean-up is performed to remove mainly sugars and fatty acids).

GPC (also known as size-exclusion chromatography, SEC) takes advantage of similar sizes of pesticide molecules for their effective, polarity-independent separation from larger matrix molecules. The reusable character of GPC columns enables easy automation and sequential runs. However, practical disadvantages include necessary periodical controls of elution patterns, the need for post-GPC sample concentration, high solvent consumption, and waste generation. Most GPC set-ups utilize ethyl acetate–cyclohexane (1:1, v/v) mobile phase in combination with styrene–divinylbenzene gels that can be either (1) soft, such as popular Bio-Beads SX-3, with high sample capacity (used mainly for removal of higher lipid contents) or (2) rigid (PL gel, etc.) enabling faster analysis and better reproducibility of elution volumes.

SPE offers a wider spectrum of clean-up options through the use of various sorbent–solvent combinations. A traditional approach (often called adsorption chromatography), employing normal-phase separation on Florisil, silica, alumina, or charcoal sorbents, effectively separates lipids and pigments from less polar pesticides (OC) that can be selectively eluted by low-polarity solvents. However, this technique is less suitable for more polar analytes that require elution by more polar eluents resulting in decreased clean-up efficiency or that may even be irreversibly retained. Similarly, adsorption of planar matrix molecules (pigments, sterols) using graphitized carbon black (GCB) sorbent is an excellent approach provided that the pesticides are not too planar.

Strong anion or cation (SAX/SCX) exchange sorbents offer highly selective clean-up in single residue/class approach to analysis of acidic or basic pesticides. Reversed-phase sorbents (e.g.,

$\text{C}_{18}$ ,  $\text{C}_8$ , and divinylbenzene polymers) are predominantly used for retention of pesticides from water samples followed by their selective elution by small solvent volumes. In food applications, another SPE approach called ‘chemical filtration’, i.e., adsorption of matrix components, such as sugars and fatty acids on aminopropyl ( $-\text{NH}_2$ ) or primary secondary amine (PSA) sorbents, while analytes remain in the extract, can be conveniently employed in cartridge or disk formats or as a novel dispersive SPE clean-up (mixing of a small amount of sorbent with the extract).

## **Determination – Identification and Quantitation**

As mentioned above, GC and LC techniques (and GC–MS and LC–MS in particular) represent the major determinative approaches used in current MMRMs. Other, much less widely applicable and applied techniques include: (1) capillary electrophoresis and capillary electrochromatography; (2) thin-layer chromatography; and (3) an array of immunoassays, such as immunosensors and enzyme-linked immunosorbent assay (ELISA). Immunoassays are, however, relatively useful for sensitive and rather rapid and inexpensive screening (followed by a confirmatory method in the case of a positive response) of selected pesticides for which ELISA kits or sensors are available.

### **Gas Chromatography**

Currently, a conventional approach to GC analysis in pesticide MMRMs employs capillary (open tubular) columns with low-bleed stationary phases mostly consisting of dimethylpolysiloxane with 5% of phenyl (other possible methyl substituents include cyanopropyl, cyanopropylphenyl or increased phenyl content up to 50%). Relatively long capillary columns (30–60 m) of 0.25–0.32 mm inner diameters (i.d.) are commonly used in routine practice, with GC analysis times usually approaching or even exceeding 1 h. However, the rising number of required analyses leads to attempts to decrease analysis time using fast (high-speed) GC techniques (mostly in combination with MS detection), thus increasing sample throughput and potentially reducing laboratory operating costs. Currently, practical approaches to fast GC–MS applicable in pesticide residue analysis employ: (1) short, microbore capillary GC columns (i.d.  $<0.2$  mm); (2) fast temperature programming; (3) low-pressure GC–MS; (4) supersonic molecular beam for MS at high carrier

gas flow rates; and (5) pressure-tunable (also called stop-flow) GC–GC.

**Sample introduction and matrix effects in gas chromatography** Table 4 lists the sample introduction (injection) techniques applied in GC pesticide residue analysis along with their positive and negative features. Injection usually represents the most crucial part (weakest link) of the GC analysis. Moreover, in real-world analysis of pesticide residues, it is the injection step that is mainly associated with adverse matrix effects, caused by the unavoidable presence of coextracted matrix components in the injected sample. Generally, matrix effects impact all steps in GC analysis (injection, separation, and detection), potentially resulting in inaccurate quantitation, decreased method ruggedness, high LODs, and/or even reporting of false positive or negative results.

During classical splitless injection, thermal degradation and/or adsorption of certain analytes may occur. When a real sample is injected, the coextracted matrix components tend to block active sites in the inlet and column (mainly free silanol groups), thus reducing losses of susceptible analytes. This phenomenon results in higher analyte signals in matrix versus matrix-free solutions, thus in an effect called ‘matrix-induced response enhancement’. As a consequence, if analyte solutions only in solvent are used for calibration standards, the concentrations of the affected analytes in sample extracts are systematically overestimated. The extent of the matrix-induced enhancement effect is related to both the chemical structure of the analyte and type of the matrix. Organophosphorus pesticides containing P=O bonds (e.g., methamidophos, acephate, omethoate, dimethoate) are particularly prone to this matrix effect. Another possible problem associated

**Table 4** GC injection techniques applied in pesticide residue analysis

<i>Injection technique</i>	<i>Advantages</i>	<i>Disadvantages</i>
Classical splitless	Easy operation Certain protection of analytical column against nonvolatile matrix components	Degradation/adsorption of susceptible analytes Matrix-induced response enhancement Small injection volumes Potential discrimination of volatile analytes due to the liner overflow
Pulsed splitless	Reduction of the residence time in the liner resulting in the decreased analyte degradation/adsorption and reduced matrix-induced response enhancement Larger injection volumes (limited by the liner size and solvent expansion volume)	Potential transfer of nonvolatile matrix components further into the column
Cold on-column	Minimum discrimination and thermodegradation	Minimum protection of the analytical column against nonvolatile matrix components – unsuitable for real-world food samples
PTV injection	Large injection volumes (limited by the injection port liquid trapping capacity – cooling efficiency, liner surface size) Reduction of thermal degradation resulting in decreased matrix-induced response enhancement Certain protection of analytical column against nonvolatile matrix components	Potential losses of volatile analytes (evaporation with the solvent) Time-consuming optimization of operating parameters (solvent trapping, solvent and analyte evaporation)
DSI/DMI <sup>a</sup>	Large injection volumes (limited by the microvial size) Effective protection of the liner and analytical column against nonvolatile matrix components resulting in minimized matrix-induced diminishment and improved ruggedness	Potential losses of volatile analytes (evaporation with the solvent) Potential discrimination of low-volatile analytes (insufficient transfer from the microvial) Careful optimization of solvent and analyte evaporation

<sup>a</sup>DSI, direct sample introduction; DMI, difficult matrix injection.

with matrix injections involves accumulation of nonvolatile matrix components in the GC system, resulting in the formation of new active sites and gradual decrease in analyte responses in both solvent and matrix solutions ('matrix-induced response diminishment effect').

In theory, the matrix-induced response enhancement effect would be overcome by eliminating of active sites or matrix components. However, both absolute GC system deactivation and thorough clean-up are virtually impossible in practice. Alternative injection techniques, such as programmed temperature vaporizing (PTV) or pulsed splitless injection, may significantly reduce the matrix-induced enhancement due to decreased thermal degradation and/or residence time in the injection port; however, they do not eliminate it. Therefore, the efficient way to avoid this problem is to compensate for it. In this respect, matrix-matched standardization is thus far the most widely used and recommended method. Unfortunately, this approach has several drawbacks including a rather time-consuming and laborious preparation of matrix-matched standards from blanks. A novel concept of 'analyte protectants' (compound additives) can offer more convenient and effective solution. Analyte protectants are compounds that strongly interact with active sites in the GC system. When added to sample extracts and matrix-free standards alike, they can induce the same response enhancement in both instances, resulting in effective equalization of the matrix-induced response enhancement effect. Thus, this approach takes advantage of response enhancement and optimizes it rather than trying to eliminate it (or merely compensate for it).

In an attempt to decrease LODs, particularly when no concentration step is conducted in sample preparation, it is desirable to use a large volume injection (LVI) technique, such as PTV in solvent vent (split) mode or direct sample introduction (DSI). DSI is a novel LVI approach, which also has a major advantage to prevent nonvolatile matrix components accumulating in the GC system, thus potentially eliminating matrix-induced response diminishment effect, reducing the maintenance, extending column life, and generally improving method ruggedness.

**Detection in gas chromatography** A traditional approach is to employ GC with element selective detectors followed by GC-MS for confirmation of the results. However, GC-MS is rapidly becoming the preferred primary tool for determination of pesticide residues in most modern MMRMs.

*Element selective detectors* Element selective detectors applicable in pesticide residue analysis include electron capture detector (ECD), electrolytic conductivity detector (ELCD), halogen-specific detector (XSD), nitrogen phosphorus detector (NPD), flame photometric detector (FPD), pulsed flame photometric detector (PFPD), sulfur chemiluminescence detector (SCD), and atomic emission detector (AED). To cover a wider range of pesticide residues, a halogen-selective detector (ECD, ELCD, XSD) in conjunction with a phosphorus- (NPD, FPD), nitrogen- (NPD), and/or sulfur-selective detector (FPD, SCD) is commonly used. A practical approach is to split the column flow to two detectors that reduces the number of injections; however, the reduced amount of analyte that reaches the detector must be considered.

The AED would represent a useful tool for the simultaneous selective detection of a number of elements occurring in pesticide molecules; however, its rather high cost and maintenance requirements virtually ruled this detector out of the routine practice. The PFPD is another detector that can be tuned to selectively detect more elements (up to two simultaneously) based on their time-delayed chemiluminescence emission versus carbon. Thus, in contrast to the FPD that can only be employed either in phosphorus or sulfur mode, the PFPD offers simultaneous detection of these (or other) two elements. However, no element selective detector, even if it detects several elements simultaneously, is confirmatory without additional information because structural isomers with the same molecular formula may exist.

*Gas chromatography-mass spectrometry* Advantages of GC-MS over traditional methods that use multiple element selective detectors include: (1) simultaneous quantitation and confirmation of detected analytes; (2) detection of a wide range of analytes independent of elemental composition; (3) possibility to spectrometrically resolve (deconvolute) coeluting peaks; and, as a consequence, (4) reduced personnel, laboratory space, and instrumentation needs, and (5) increased sample throughput. On the other hand, GC-MS instruments are still more expensive than GC instruments with element selective detectors, and sometimes, in very complex matrices, element selective detectors may permit detection of lower concentrations of residues masked by matrix interferences in GC-MS.

An ideal approach is to screen for all potential GC-amenable pesticide residues using a sensitive, rugged, and fast MS instrument in full scan mode that enables automated (1) mass spectral deconvolution, (2) identification of pesticide residues at sufficiently



**Table 5** The main (A) ionization techniques and (B) mass analyzers currently used in GC–MS analysis of pesticide residues

<i>MS techniques</i>	<i>Application strengths</i>	<i>Application limitations</i>
<i>(A) Ionization techniques</i>		
Electron impact	Controlled and reproducible fragmentation pattern Mass spectral libraries for identification of unknowns Confirmation of analyte identity (consistent ion abundance ratios) Determination of molecular structure	Difficult determination of the molecular weight (molecular ion not always present in the spectrum)
Chemical ionization	Determination of molecular weight Increased sensitivity (less fragmentation) and selectivity (reduced chemical noise) resulting in improved detectability	Reduced ability to identify/confirm (little or no fragmentation) Decreased analytical scope
<i>(B) Mass analyzers</i>		
Quadrupole	Rugged, reliable, and highly quantitative Highly compatible with existing spectral libraries Increased sensitivity in selected ion monitoring (SIM) mode Affordable	Scan mode relatively slow (max. 10–20 spectra/s) and insufficiently sensitive SIM decreases analytical scope (targeted analysis only) and requires inconvenient time-window set-ups Low mass resolution
Ion trap	Relatively sensitive full scan mode Increased detection selectivity and confirmation ability using MS <sup>n</sup> option Affordable	Quantitation problems and space charge effects Relatively slow scanning (max. 10–20 spectra/s) MS <sup>n</sup> decreases analytical scope (targeted analysis only), sensitivity, and speed and requires time-consuming optimization Low mass resolution
Time-of-flight (TOF)	No spectra skewing (optimal for deconvolution)	Reduced ruggedness Expensive (especially the high-resolution variant) Large data files
(a) High-speed TOF	Fast spectral acquisition rate (up to 500 spectra/s) for very and ultrafast GC–MS and GC × GC techniques	Lower mass resolution
(b) high-resolution TOF	High resolution (7000 FWHM at $m/z$ 614) improving detection selectivity	Relatively slow spectral acquisition rate (up to 10 spectra/s)

low concentrations using mass spectral libraries/pesticide databases, and (3) reliable quantification of detected residues. Table 5 indicates how the capabilities of the main MS (ionization and ion separation) techniques currently used in GC–MS analysis of pesticide residues approximate this ideal.

### High-Performance Liquid Chromatography

**Traditional detectors and postcolumn derivatization** Traditionally, LC served mostly for analysis of single or small groups of non-GC-amenable pesticide residues for which derivatization for GC was unsuitable or complicated. Besides a lower number of possible theoretical plates, the main reason for its inferior position as compared to GC consists in the

limited variety and capabilities of traditional LC detectors, such as ultraviolet/visible (UV/VIS), diode array (DAD), or fluorescence (FLD) detectors. Although DAD offers improved selectivity versus a single-wavelength UV/VIS detector, they are both restricted to the detection of analytes with chromophores and usually do not provide sufficiently low LODs for residue analysis in complex matrices. FLD is a substantially more selective and sensitive option; however, only few pesticides have a fluorophore in their molecular structure. In some instances, postcolumn derivatization may be performed to convert the analytes to fluorescent derivatives, such as hydrolysis of *N*-methylcarbamates to methylamine and its subsequent reaction with *o*-phthalaldehyde (and 2-mercaptoethanol).

**Liquid chromatography–mass spectrometry** The initial attempts to couple LC with MS lacked important attributes for trace analysis: sensitivity, robustness, and reliable quantitation. Moreover, the cost of the early LC–MS instruments was prohibitive for most laboratories. The revolutionary introduction of atmospheric pressure ionization (API) techniques, mainly electrospray (ESI) and atmospheric pressure chemical ionization (APCI), resulted in greater applicability of LC–MS and manufacture of more reliable, affordable, and user-friendly instruments. Thus, LC–MS is now becoming an indispensable part of the analytical strategy in many routine laboratories, enabling direct, selective, and sensitive multiclass, multiresidue analysis of more polar, low volatile, and/or thermolabile pesticides, such as carbamates, phenylureas, sulfonylureas, imidazoles, triazoles, imidazolinones, chlorophenoxy acids, and many others.

A drawback of using soft API techniques involves reduced spectral information (often only a protonated  $[M + H]^+$  or deprotonated  $[M - H]^-$  molecular ion in positive or negative ionization mode, respectively) resulting in decreased selectivity and limited confirmation ability. Thus, in most applications, the use of tandem MS, either tandem-in-time (ion trap MS<sup>n</sup>) or more expensive and reliable tandem-in-space (combination of two mass analyzers for MS–MS, such as triple quadrupole, quadrupole-TOF) instruments, becomes essential for achieving both low LODs (increased signal-to-noise ratios due to improved selectivity) and almost unambiguous confirmation. High-resolution MS may also offer greater selectivity and confidence in analyte identity; however, at substantially higher costs.

Similarly to LC with traditional detectors, LC–MS also mainly employs reverse-phase LC with modified silica stationary phases; however, the composition of mobile phases is rather limited to methanol or acetonitrile as principal organic modifiers and to volatile additives (buffers).

Matrix effects complicate not only GC but also LC–MS analysis. Matrix components may impact selective determination of coeluting analytes (a rare effect in tandem MS) and mainly interfere in the spraying, ionization, and evaporation processes leading to analyte signal suppression or less often enhancement. To compensate for the latter effects, matrix-matched standardization is mostly used after failure of more convenient approaches, such as application of a different ionization technique (APCI is usually better than ESI in this respect), altered LC separation, or improved sample preparation selectivity. In the case of a limited number of analytes, the use of deuterated internal standards or echo-peak

technique (injection of a standard solution shortly after the sample resulting in closely eluting peaks of the same analyte) represent less laborious compensation methods than matrix-matched standardization.

**See also:** **Capillary Electrophoresis:** Overview. **Counter-current Chromatography:** Solvent Extraction with a Helical Column. **Extraction:** Solvent Extraction Principles; Solvent Extraction: Multistage Countercurrent Distribution; Microwave-Assisted Solvent Extraction; Pressurized Fluid Extraction; Supercritical Fluid Extraction; Solid-Phase Extraction; Solid-Phase Microextraction. **Gas Chromatography:** Overview; Principles; Column Technology; Multidimensional Techniques; High-Speed Techniques; Instrumentation; Detectors; Mass Spectrometry. **Immunoassays, Applications:** Food. **Liquid Chromatography:** Overview; Principles; Column Technology; Mobile Phase Selection; Reversed Phase; Instrumentation; Food Applications. **Mass Spectrometry:** Overview; Principles; Ionization Methods Overview; Electron Impact and Chemical Ionization; Atmospheric Pressure Ionization Techniques; Electrospray; Mass Separation; Ion Traps; Time-of-Flight; Food Applications. **Thin-Layer Chromatography:** Overview; Principles; Instrumentation; Method Development.

## Further Reading

- Ahmed FE (2001) Analyses of pesticides and their metabolites in foods and drinks. *Trends in Analytical Chemistry* 20: 649–661.
- Barker SA (2000) Matrix solid-phase dispersion. *Journal of Chromatography A* 885: 115–127.
- Beltran J, López FJ, and Hernández F (2000) Solid-phase microextraction in pesticide residue analysis. *Journal of Chromatography A* 885: 389–404.
- Fong WG, Moye HA, Seiber JN, and Toth JP (1999) *Pesticide Residues in Foods; Methods, Techniques, and Regulations*. New York: Wiley.
- Food and Drug Administration (1994) *Pesticide Analytical Manual*, 3rd edn. Washington, DC: US Department of Health and Human Services.
- Hajslova J and Zrostlikova J (2003) Matrix effects in (ultra)trace analysis of pesticide residues in food and biotic matrices. *Journal of Chromatography A* 1000: 181–197.
- Lee PW (ed.) (2002) *Handbook of Residue Analytical Methods for Agrochemicals*. Chichester: Wiley.
- Lehotay SJ (2000) Multiclass, multiresidue analysis of pesticides, strategies for. In: Meyers RA (ed.) *Encyclopedia of Analytical Chemistry*, pp. 6344–6384. Chichester: Wiley.
- Lehotay SJ and Schenck FJ (2000) Multiresidue methods: extraction. In: Wilson ID, Adlard TR, Poole CF, and Cooks M (eds.) *Encyclopedia of Separation Science*, pp. 3409–3415. London: Academic Press.
- Menzinger F, Schmitt-Kopplin Ph, Freitag D, and Kettrup A (2000) Analysis of agrochemicals by capillary electrophoresis. *Journal of Chromatography A* 891: 45–67.

- Nunes GS, Toscano IA, and Barceló D (1998) Analysis of pesticides in food and environmental samples by enzyme-linked immunosorbent assays. *Trends in Analytical Chemistry* 17: 79–87.
- Picó Y, Font G, Moltó JC, and Mañes J (2000) Pesticide residue determination in fruit and vegetables by liquid chromatography–mass spectrometry. *Journal of Chromatography A* 882: 153–173.

- Sherma J (2001) Pesticide residue analysis (1999–2000): a review. *Journal of AOAC International* 84: 1303–1312.
- Sherma J (2000) Recent advances in the thin-layer chromatography of pesticides: a review. *Journal of AOAC International* 86: 602–611.
- van der Hoff GR and van Zoonen P (1999) Trace analysis of pesticides by gas chromatography. *Journal of Chromatography A* 843: 301–322.

## Mycotoxins

**M Driffield**, Central Science Laboratory, Sand Hutton, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Maurice O Moss, pp. 1490–1499, © 1995, Elsevier Ltd.

### Introduction

The mycotoxins are fairly low relative-molecular-mass secondary metabolites of filamentous fungi that are toxic to humans and domesticated animals, causing illness, or even death, when contaminated foods or animal feeds are consumed. There are many toxic metabolites produced by fungi but only a fraction of these are currently considered to be important in foods. Similarly, a very broad range of fungal species may produce toxic metabolites, but in human food the most important genera are *Aspergillus*, *Penicillium*, and *Fusarium*. Although the presence of a mycotoxin in a food usually implies the previous infection by, and growth of, the producing mold on the food, or a constituent of the food, there are two significant instances when that is not so. Mycotoxins can be passed through the food chain ending up in a commodity that was not itself moldy. Thus, aflatoxin M<sub>1</sub> is secreted into milk following consumption by a cow of feed contaminated by aflatoxin B<sub>1</sub>. Ochratoxin A can pass unmetabolized into the kidney and muscle tissue of the pig after feeding on a contaminated feed such as barley.

### Range of Mycotoxins Encountered in Foods

Major problems encountered in the analysis of mycotoxins arise because of their diversity of chemical structure and the wide range of foods and animal feeds in which some of them may occur. This often means the development of quite different analytical methods for each mycotoxin in each food

commodity. **Table 1** lists the most important mycotoxins that have been found in food. The structures of these mycotoxins are shown in **Figures 1–11**.

The diversity in chemical structure of the mycotoxins is reflected in a diversity of biosynthetic pathways and biological activity. **Table 2** provides information on the toxicity of the more important toxins. Recognition of the importance of some mycotoxins is reflected in the production of statutory and advisory regulations for maximum tolerated levels in foods and animal feeds in many countries. In the case of the aflatoxins the low levels set for maximum tolerated concentrations (**Table 3**) reflect the assumption that these compounds are carcinogenic to humans, coupled with the ability of the analyst reproducibly to detect such low levels in a wide range of foods and animal feeds. At least 50 countries have regulations in place or proposed for the control of aflatoxins in food or animal feeds, and about a one-fourths of these countries also have regulations governing permitted levels for other mycotoxins such as ochratoxin, patulin, deoxynivalenol, and zearalenone. Many countries have especially stringent levels for aflatoxin M<sub>1</sub> in milk and dairy products ranging from 0.01 µg per kg in pasteurized fresh milk for infants in Austria to 0.5 µg per kg for milk and milk products in the USA. This, in part, reflects concern for exposure of the very young and elderly, who would be more likely to be both more sensitive and have a higher level of exposure. International trade in many of the commodities most at risk from contamination requires the establishment of agreed methods of analysis and many countries use those described in the Official Methods of Analysis of the Association of Official Analytical Chemists.

### Sampling

Before an analysis can be attempted a sample must be obtained. The analyst is frequently presented with a

- Nunes GS, Toscano IA, and Barceló D (1998) Analysis of pesticides in food and environmental samples by enzyme-linked immunosorbent assays. *Trends in Analytical Chemistry* 17: 79–87.
- Picó Y, Font G, Moltó JC, and Mañes J (2000) Pesticide residue determination in fruit and vegetables by liquid chromatography–mass spectrometry. *Journal of Chromatography A* 882: 153–173.

- Sherma J (2001) Pesticide residue analysis (1999–2000): a review. *Journal of AOAC International* 84: 1303–1312.
- Sherma J (2000) Recent advances in the thin-layer chromatography of pesticides: a review. *Journal of AOAC International* 86: 602–611.
- van der Hoff GR and van Zoonen P (1999) Trace analysis of pesticides by gas chromatography. *Journal of Chromatography A* 843: 301–322.

## Mycotoxins

**M Driffield**, Central Science Laboratory, Sand Hutton, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Maurice O Moss, pp. 1490–1499, © 1995, Elsevier Ltd.

### Introduction

The mycotoxins are fairly low relative-molecular-mass secondary metabolites of filamentous fungi that are toxic to humans and domesticated animals, causing illness, or even death, when contaminated foods or animal feeds are consumed. There are many toxic metabolites produced by fungi but only a fraction of these are currently considered to be important in foods. Similarly, a very broad range of fungal species may produce toxic metabolites, but in human food the most important genera are *Aspergillus*, *Penicillium*, and *Fusarium*. Although the presence of a mycotoxin in a food usually implies the previous infection by, and growth of, the producing mold on the food, or a constituent of the food, there are two significant instances when that is not so. Mycotoxins can be passed through the food chain ending up in a commodity that was not itself moldy. Thus, aflatoxin M<sub>1</sub> is secreted into milk following consumption by a cow of feed contaminated by aflatoxin B<sub>1</sub>. Ochratoxin A can pass unmetabolized into the kidney and muscle tissue of the pig after feeding on a contaminated feed such as barley.

### Range of Mycotoxins Encountered in Foods

Major problems encountered in the analysis of mycotoxins arise because of their diversity of chemical structure and the wide range of foods and animal feeds in which some of them may occur. This often means the development of quite different analytical methods for each mycotoxin in each food

commodity. **Table 1** lists the most important mycotoxins that have been found in food. The structures of these mycotoxins are shown in **Figures 1–11**.

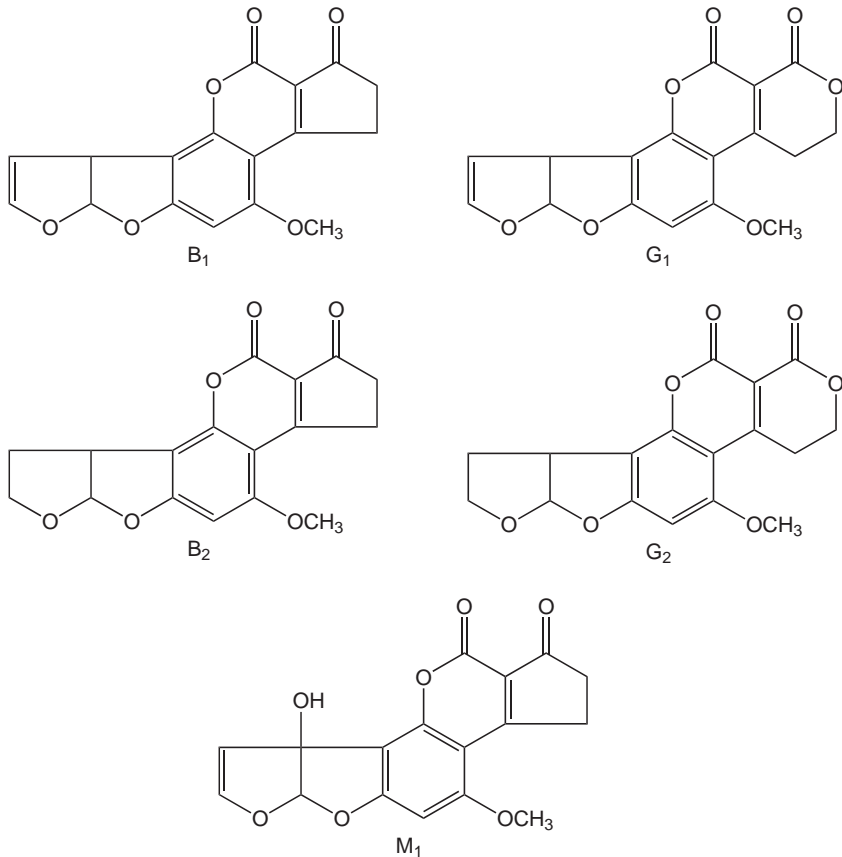
The diversity in chemical structure of the mycotoxins is reflected in a diversity of biosynthetic pathways and biological activity. **Table 2** provides information on the toxicity of the more important toxins. Recognition of the importance of some mycotoxins is reflected in the production of statutory and advisory regulations for maximum tolerated levels in foods and animal feeds in many countries. In the case of the aflatoxins the low levels set for maximum tolerated concentrations (**Table 3**) reflect the assumption that these compounds are carcinogenic to humans, coupled with the ability of the analyst reproducibly to detect such low levels in a wide range of foods and animal feeds. At least 50 countries have regulations in place or proposed for the control of aflatoxins in food or animal feeds, and about a one-fourths of these countries also have regulations governing permitted levels for other mycotoxins such as ochratoxin, patulin, deoxynivalenol, and zearalenone. Many countries have especially stringent levels for aflatoxin M<sub>1</sub> in milk and dairy products ranging from 0.01 µg per kg in pasteurized fresh milk for infants in Austria to 0.5 µg per kg for milk and milk products in the USA. This, in part, reflects concern for exposure of the very young and elderly, who would be more likely to be both more sensitive and have a higher level of exposure. International trade in many of the commodities most at risk from contamination requires the establishment of agreed methods of analysis and many countries use those described in the Official Methods of Analysis of the Association of Official Analytical Chemists.

### Sampling

Before an analysis can be attempted a sample must be obtained. The analyst is frequently presented with a

**Table 1** The major mycotoxins found in foods

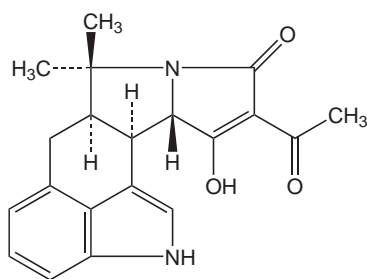
<i>Mycotoxin</i>	<i>Figure</i>	<i>Major foods</i>	<i>Common producing spp.</i>
Aflatoxins B <sub>1</sub> , B <sub>2</sub> , G <sub>1</sub> , and G <sub>2</sub>	1	Maize, groundnuts, figs, tree nuts	<i>Aspergillus flavus</i> , <i>A. parasiticus</i>
Aflatoxin M <sub>1</sub>	1	Milk, milk products	(Secreted by cow after metabolism of aflatoxin B <sub>1</sub> )
Cyclopiazonic acid	2	Cheese, maize, groundnuts, Kodo millet	<i>Aspergillus flavus</i> , <i>Penicillium aurantiogriseum</i>
Deoxynivalenol	3	Cereals	<i>Fusarium graminearum</i> , <i>F. culmorum</i>
Fumonisin B <sub>1</sub> , B <sub>2</sub> , B <sub>3</sub>	4	Maize	<i>Fusarium moniliforme</i>
Nivalenol	3	Cereals	<i>Fusarium graminearum</i>
Ochratoxin	5	Maize, cereals, coffee beans	<i>Penicillium verrucosum</i> , <i>Aspergillus ochraceus</i>
Patulin	6	Apple juice	<i>Penicillium expansum</i>
Penitrem A	7	Walnuts	<i>Penicillium aurantiogriseum</i>
Sterigmatocystin	8	Cereals, coffee beans, cheese	<i>Aspergillus versicolor</i>
T-2 and HT-2 toxins	11	Cereals	<i>Fusarium graminearum</i>
Tenuazonic add	9	Tomato paste	<i>Alternaria tenuis</i>
Zearalenone	10	Maize, barley, wheat	<i>Fusarium graminearum</i>



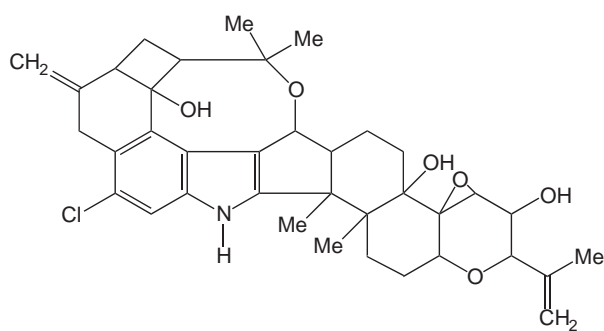
**Figure 1** Aflatoxins B<sub>1</sub>, B<sub>2</sub>, G<sub>1</sub>, G<sub>2</sub>, and M<sub>1</sub>.

single sample, adequate in quantity for the analytical methodology, but without any guarantee that it is representative of the consignment of food of animal feed from which it came. Many of the commodities

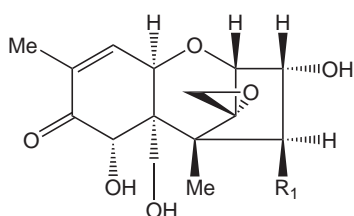
to be analyzed are particulate in structure and molds do not grow uniformly through such materials. Thus, mycotoxins that are not usually distributed evenly may be aflatoxins in milk and patulin in fruit juice,



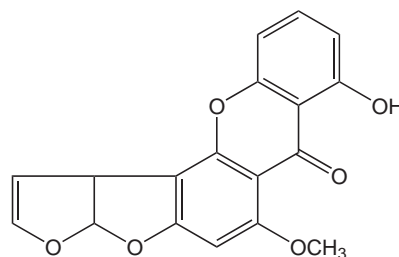
**Figure 2** Cyclopiazonic acid.



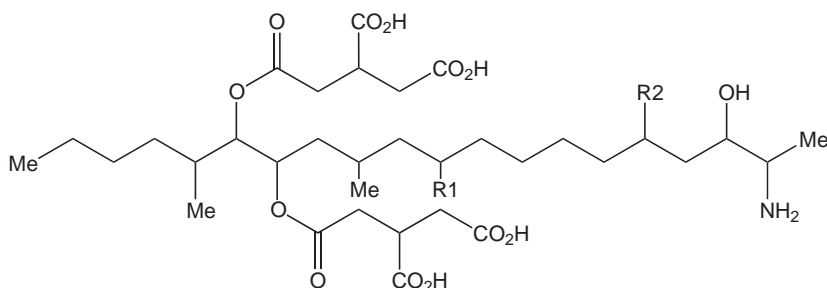
**Figure 7** Penitrem A.



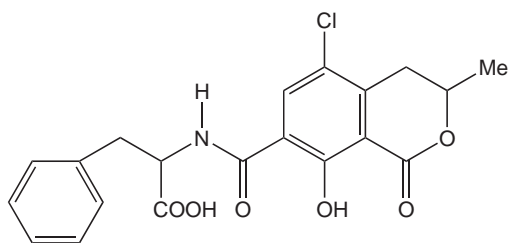
**Figure 3** Deoxynivalenol (R1 = H) and nivalenol (R1 = OH).



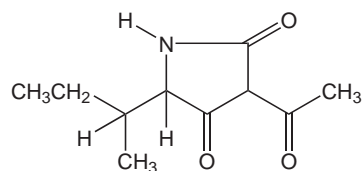
**Figure 8** Sterigmatocystin.



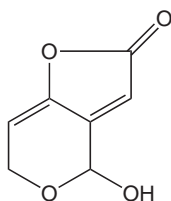
**Figure 4** Fumonisin B<sub>1</sub> (R1 = OH, R2 = OH), fumonisin B<sub>2</sub> (R1 = H, R2 = OH), and fumonisin B<sub>3</sub> (R1 = OH, R2 = H).



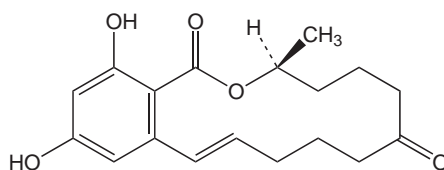
**Figure 5** Ochratoxin A.



**Figure 9** Tenuazonic acid.



**Figure 6** Patulin.



**Figure 10** Zearalenone.

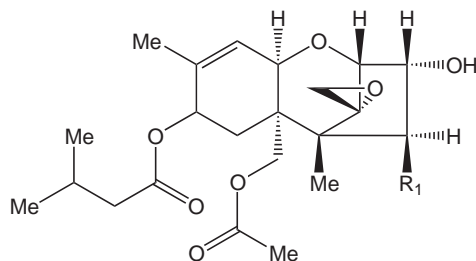


and it is important to agree to a sampling plan that adequately protects the consumer without unnecessarily penalizing the producer. There is not even a normal distribution of mycotoxins within a batch of a commodity. For example, in the contamination of groundnuts, maize, cotton seed, or palm kernels with aflatoxin, it is likely that a relatively small number of kernels or seeds are contaminated with very high levels and many are not contaminated at all, giving a very skewed distribution.

Satisfactory sampling plans, such as those devised for groundnuts by the US Department of Agriculture (USDA) and the Tropical Products Institute (now the Natural Resources Institute) of the UK, should define the size of the sample, how it is to be collected, a method for dividing the sample, and an analytical procedure. The USDA sampling plan for the determination of aflatoxins in groundnuts requires that 66 kg of material be collected from at least 100 sacks for each 20 ton lot. The 66 kg is divided into three subsamples of 22 kg and the first of these is ground, thoroughly mixed, and analyzed. At a time when the maximum tolerated level was 25 µg per kg, if the

result of this first analysis was less than 16 µg per kg the lot was accepted, if greater than 75 µg per kg it was rejected. If, however, the first analysis gave a result between these two figures, then the second subsample had to be analyzed in the same way. If the mean of the two is less than 22 µg per kg the lot was accepted and if greater than 38 µg per kg rejected but, again, if the result fell between these two figures then the third subsample had to be analyzed. In this example, the mean of all three now needs to be less than or equal to 25 µg per kg for the lot to be accepted.

For many years sampling was not taken sufficiently seriously but detailed research on the distribution of mycotoxins in commodities and an appropriate statistical analysis of the results have led to the increased use of acceptable sampling plans. The recognition of the extremely skewed distribution of mycotoxins has also led to the design of methods for recognizing and eliminating the heavily contaminated kernels from consignments, thus ensuring that they do not contribute to the overall contamination of the batch after further treatment such as mixing, grinding, and oil extraction.



**Figure 11** T-2 toxin ( $R_1 = \text{OCOMe}$ ) and HT-2 toxin ( $R_1 = \text{OH}$ ).

## Extraction Procedures and Cleanup

Mycotoxins are produced in a complex matrix of macromolecules, including proteins, polysaccharides, and lipids, and many of the plant commodities most at risk of contamination contain a complex range of low-relative-molecular-mass organic compounds. Although many mycotoxins are readily soluble in water-immiscible organic solvents, it is not always

**Table 2** Toxicity and biological effects of major mycotoxins

<i>Mycotoxin</i>	<i>Biological activity</i>	<i>LD<sub>50</sub> (mg per kg)</i>
Aflatoxin B <sub>1</sub>	Hepatotoxic, carcinogenic	0.5 (dog), 9.0 (mouse)
Cyclopiazonic acid	Convulsions	36 (rat)
Deoxynivalenol	Vomiting, feed refusal	70 (mouse)
Fumonisin B <sub>1</sub> , B <sub>2</sub> , B <sub>3</sub>	Equine encephalomalacia, pulmonary edema and hydrothorax in pigs, esophageal carcinoma	?
Nivalenol	Gastrointestinal disturbances, cytotoxic and immunosuppressive	4.1 (mouse)
Ochratoxin	Nephrotoxic	20–30 (rat)
Patulin	Oedema, haemorrhage, possibly carcinogenic	35 (mouse)
Penitrem A	Tremorgen	1.05 (mouse)
Sterigmatocystin	Hepatotoxic, carcinogenic	166 (rat)
T-2 and HT-2 toxins	Gastrointestinal disturbances, cytotoxic and immunosuppressive	5.2 (T-2 mouse) 9 (HT-2 mouse)
Tenuazonic acid	Convulsions, haemorrhage	81 (female mouse) 186 (male mouse)
Zearalenone	Oestrogenic	Not acutely toxic

**Table 3** Maximum acceptable levels for a range of mycotoxins in foodstuffs for a selection of countries

<i>Mycotoxin</i>	<i>Country</i>	<i>Limit (<math>\mu\text{g per kg}</math>)</i>	<i>Matrix</i>
Aflatoxin B <sub>1</sub>	EU	2	Nuts or dried fruit intended for direct human consumption or as an ingredient in foodstuffs
Aflatoxin B <sub>1</sub> , B <sub>2</sub> , G <sub>1</sub> , G <sub>2</sub>	EU	4	Nuts or dried fruit intended for direct human consumption or as an ingredient in foodstuffs
Aflatoxin B <sub>1</sub> , B <sub>2</sub> , G <sub>1</sub> , G <sub>2</sub>	Canada	20	Feeding stuffs for animals
Aflatoxin M <sub>1</sub>	EU	0.025	Infant formulae including infant milk. The maximum level for infant formulae refers to the product ready to use after reconstitution
Aflatoxin M <sub>1</sub>	US	0.5	Milk for human consumption
DON	US	5000	Grain by product not to exceed 40% of animal diet. Includes animals other than feedlot cattle older than 4 months, beef cattle older than 4 months, chickens, and swine
Fumonisin B <sub>1</sub> , B <sub>2</sub> , and B <sub>3</sub>	US	3000	Cleaned corn intended for popcorn
Ochratoxin A	EU	3	All products derived from cereals including processed cereal products and cereal grains intended for direct human consumption
Ochratoxin A	Switzerland	0.5	Processed cereal based baby foods
Ochratoxin A	Hungary	15	Green coffee
T-2 toxin	Bulgaria	100	Cereals intended for direct consumption or as an ingredient in foodstuffs

easy to remove them from the macromolecular food matrix by direct extraction with simple solvents such as chloroform or ethyl acetate. It may be necessary to remove oils and fats by a preliminary extraction with hexane before extracting the mycotoxin.

In the case of aflatoxins, detailed methods will vary from one commodity to another. Thus, defatted hazel nuts, pistachios, Brazil nuts, groundnuts, or groundnut meal, can be extracted directly with chloroform after the addition of a small volume of water, whereas cottonseed and cottonseed products are best extracted initially with aqueous acetone and the aflatoxins subsequently partitioned into chloroform. Established procedures for aflatoxins that illustrate these principles include:

(1) 'Groundnuts'. Comminuted material (50 g) is extracted with hexane in a Soxhlet extractor for 3–4 h. Residual solvent is removed from the solid in a warm oven with an appropriate extractor fan. To 10 g of defatted material, in a 250 ml stoppered flask, 10 ml of water is added and mixed thoroughly using a glass rod. Chloroform (100 ml) is added, the flask stoppered, and the mixture shaken on a wrist shaker for 30 min. The extract is filtered, a measured volume taken to dryness under a stream of nitrogen, transferred to a vial in a known volume of chloroform, and retained for subsequent analysis.

(2) 'Cottonseed meal'. Comminuted material (25 g) is extracted with 250 ml acetone–water (85:15; v/v) on

a wrist shaker for 30 min. The crude extract is filtered and 100 ml treated with 20 ml lead acetate solution (20% neutral lead acetate trihydrate in deionized water with 3 ml glacial acetic acid added per liter) and 80 ml deionized water. The mixture is stirred well and allowed to stand for 5 min before treating with filter aid (5 g) and filtering through a Whatman No. 4 filter paper. A measured volume of this filtrate (100 ml) is extracted with chloroform (50 ml) in a separating funnel. The chloroform layer is allowed to drain through a column containing a layer (2 cm) of anhydrous sodium sulfate, a layer of 5 g acidic alumina, and topped off with an additional layer (2 cm) of anhydrous sodium sulfate. The aqueous residue is extracted with a second portion of chloroform (50 ml) and the chloroform layer drained through the same column. The combined filtrate is taken to dryness under nitrogen and transferred quantitatively to a vial for subsequent analysis.

Those mycotoxins, such as ochratoxin A, which contain a carboxyl group, can be readily extracted into an organic solvent such as chloroform after acidifying the commodity to ensure that it exists entirely in the undissociated form. Aqueous phosphoric acid ( $0.1 \text{ mol l}^{-1}$ ) is frequently used. Ochratoxin A can then be back-extracted from the chloroform using sodium hydrogencarbonate solution or trapped on a column of diatomaceous earth

impregnated with a basic aqueous solution such as sodium hydrogencarbonate. Esters and fat can be removed with hexane and chloroform and the acidic ochratoxin A eluted with a formic acid–chloroform mixture.

Some mycotoxins are sufficiently hydrophilic that they cannot be recovered efficiently with a simple chloroform extraction even though they are intrinsically soluble in chloroform. Amongst the trichothecenes, deoxynivalenol is best extracted initially with aqueous methanol, before subsequently partitioning into chloroform. Indeed, the dependence on relatively nonpolar organic solvents for the convenient extraction of mycotoxins from foods, animal feeds, and raw materials may be one of the reasons why the fumonisins have only recently been isolated and characterized, despite their toxicological importance and widespread occurrence. Fumonisins are associated with encephalomalacia in horses, edema of the lung in pigs, liver carcinoma in rats, and may be responsible for esophageal carcinoma in humans. They are produced by *Fusarium moniliforme* and related species of the *Liseola* group of *Fusarium* and it is now realized that they are widespread in cereals such as maize and can occur naturally in very high concentrations. Their extraction from maize can be carried out in acetonitrile–methanol–water (25:25:50; v/v/v) by shaking for 20 min or by homogenization or blending for 5 min.

Cleanup procedures vary considerably depending on both the mycotoxin(s) and the commodity to be analyzed. For many commodities slurries of freshly precipitated iron(III) hydroxide and copper(II) carbonate have proved effective in removing co-extractives. Flocculation of impurities with aqueous lead acetate or zinc acetate has also been used. Some mycotoxins can be adsorbed onto silica gel and/or florosil and subsequently eluted. Indeed, a range of commercially available prepared columns, such as Sep-Pak, have proved very convenient as a cleanup stage for crude extracts of mycotoxins before subsequent analysis.

More recently cleanup procedures have been developed purposely for mycotoxins. The principle of immunoaffinity, where an antibody specific to a particular mycotoxin is immobilized onto an inert solid support and packed into a column, is most commonly used. Such a column can be used to absorb a mycotoxin from a known volume of a food extract or liquid with very little, if any, preliminary cleanup. Most of the constituents of the extract pass through the column or can be washed off the column with water. Because the mycotoxins are not covalently bound to the column they can be displaced with an organic solvent such as methanol. This produces a

very clean, concentrated extract that is then usually analyzed by liquid chromatography (LC), gas chromatography (GC), or directly by fluorimetry for those mycotoxins that fluoresce. A major disadvantage of this technique is the cost of the columns as they can only be used once because the antibodies are destroyed when releasing the mycotoxins. Immunoaffinity columns are commercially available for total aflatoxins (B<sub>1</sub>, B<sub>2</sub>, G<sub>1</sub>, G<sub>2</sub>), aflatoxin B<sub>1</sub>, aflatoxin M<sub>1</sub>, ochratoxin A, deoxynivalenol, fumonisin B<sub>1</sub>, B<sub>2</sub>, and B<sub>3</sub>, T-2 toxin, zearalenone, and total aflatoxins and ochratoxin A together in one column.

There is also a range of cleanup columns for mycotoxins that allow rapid one-step purification with no washing or eluting. The sample extract is passed through a packing material consisting of charcoal, alumina, and silica where the impurities are retained whilst the mycotoxins pass through the packing material with a cleaner solution. As the packing material is impregnated with impurities each column can only be used once. These columns will cleanup extracts for single mycotoxins or ones with similar properties including fumonisins, ochratoxin A, moniliformin, and trichothecenes. Others exist for purification of samples for a range of mycotoxins, for example, one column for aflatoxins, zearalenone, and patulin.

The recovery of mycotoxins from all extraction and cleanup procedures needs to be validated and this is usually done using carefully spiked material prepared from clean samples of the commodity under study. It must, however, be recognized that it is usually easier to recover compounds from spiked samples than it is to recover the same mycotoxins from naturally contaminated material, and for some mycotoxins there are available samples of naturally contaminated material the mycotoxin content of which has been established by international collaborative studies. The extraction and cleanup procedures will still usually provide a complex mixture requiring further separation before a specific physicochemical method for the quantitative or qualitative analysis of individual mycotoxins can be applied. This further separation is usually carried out by some form of chromatography.

## Chromatographic Methods

The separatory power of a wide range of chromatographic methods can be used as a final cleanup stage before determination of a specific mycotoxin, with thin-layer chromatography (TLC), in either one- or two-dimensional mode, LC, and GC being the most common.

### Thin-Layer Chromatography

All the aflatoxins fluoresce intensely under long-wave UV (360 nm) when absorbed on a solid substrate such as the silica gel of a TLC plate (this is not always true in solution). Thus, TLC proved to be especially useful during the early days in the study of the etiological agents of turkey X disease leading to the recognition, isolation, and characterization of the aflatoxins. Indeed, the two groups of aflatoxins, B and G, are so labeled because of the bluish or greenish fluorescence that they produce.

The problem with the simple TLC analysis of aflatoxins is that many natural products fluoresce blue or green under UV and many of these compounds have an  $R_F$  value close to, or even identical with, that of one of the four common aflatoxins (Figure 1). The early literature contains a number of false-positive reports of the presence of aflatoxins, and even their production by species of mold other than *Aspergillus flavus*, *A. parasiticus*, and *A. nomius*, so it is essential that a confirmatory test be carried out such as derivatization coupled with two-dimensional TLC. In the case of the aflatoxins, a useful confirmatory test is to derivatize as the hemiacetals with trifluoroacetic acid by superimposing 2  $\mu$ l of a 50% solution of the reagent in chloroform onto the spot of both the unknown and a standard marker before chromatography. Aflatoxins B<sub>1</sub> and G<sub>1</sub>, but not B<sub>2</sub> and G<sub>2</sub>, form hemiacetals that have the same fluorescence as the parent compound but run significantly more slowly

in the commonly used solvent systems. The fluorescence of all four aflatoxins, and the derivatives of B<sub>1</sub> and G<sub>1</sub>, change from blue or green to yellow after spraying the plate with 25% aqueous solutions of an inorganic acid such as sulfuric or nitric acids.

Table 4 lists a selection of TLC methods used for some of the more important mycotoxins listed in Table 1.

### Liquid Chromatography

Although many well-validated methods existed for TLC-based analysis of several mycotoxins, once the equipment became readily available LC has frequently been the technique of choice for quantitative determination. A major advantage of LC is the possibility of automation and the use of autosamplers allowing the unattended operation of LC equipment. Online clean-up of crude extracts and online derivatization before or after separation are additional advantages. A range of detectors can be used, although the most widespread are based on UV absorption or the detection of fluorescence, the latter being potentially 30–40 times more sensitive for aflatoxins. Although the four common aflatoxins fluoresce well in the solid state on TLC plates, the fluorescence of aflatoxins B<sub>1</sub> and G<sub>1</sub> is weak in many solvents. This can be overcome in a number of ways. The aflatoxins can be converted to hemiacetal derivatives with trifluoroacetic acid before chromatography occurs. Postcolumn treatment with iodine-saturated water, or more commonly pyridinium bromide perbromide solution, is increasingly becoming

**Table 4** TLC methods for some mycotoxins

Mycotoxin	Solvent (v/v)	Detection method	Detection limit ( $\mu$ g per kg)
Aflatoxins	Acetone–CHCl <sub>3</sub> (1:9)	Fluorescence under UV (360 nm)	1–5
Aflatoxin M <sub>1</sub>	CHCl <sub>3</sub> –acetone–2-PrOH (87:10:3)	Fluorescence under UV (360 nm)	0.05
Cyclopiazonic acid	EtOAc–2-PrOH–NH <sub>3</sub> aq (50:15:10)	<i>p</i> -Dimethylaminobenzaldehyde spray (blue fluorescence)	125
Deoxynivalenol	CHCl <sub>3</sub> –2-PrOH–acetone (8:1:1)	AlCl <sub>3</sub> spray, heat 7 min, 120°C, blue fluorescence	40
Ochratoxin A	Benzene–acetic acid–MeOH (90:5:5)	Exposure to NH <sub>3</sub> –MeOH vapor, blue fluorescence (360 nm)	10
Patulin	Toluene–EtOAc–CHCl <sub>3</sub> –formic acid (105:45:48:3)	MBTH <sup>a</sup> spray, heat 5 min, 130°C, yellow brown fluorescence (360 nm)	120–130
Penitrem A	CHCl <sub>3</sub> –acetone (93:7)	Spray 1% FeCl <sub>3</sub> in BuOH gentle warming, purple blue color	–
Sterigmatocystin	Benzene–acetic acid–MeOH (90:5:5)	AlCl <sub>3</sub> spray, heat 5 min, 110°C, yellow–green fluorescence (360 nm)	5–10
Tenuazonic acid	Benzene–acetic acid–MeOH (90:5:5)	Fluorescence quenching on Hf <sub>254</sub> plates (250 nm)	2 $\mu$ g per spot
Zearalenone	CHCl <sub>3</sub> –MeOH (97:3)	Blue fluorescence under UV (25 nm)	10

<sup>a</sup>MBTH, 3-methyl-2-benzothiazolinone hydrazine hydrochloride (0.5 g in 100 ml water).

the method of choice. Postcolumn derivatization can also be afforded by addition of potassium bromide and nitric acid to the mobile phase, which is passed through an electrochemical (Kobra) cell to produce the bromine derivatizing agent. The change in intensity of the peaks with and without postcolumn treatment is a useful confirmation of the presence of aflatoxin B<sub>1</sub> and G<sub>1</sub>.

The electrochemical detector has been used for the analysis of roquefortine and the zearalenones and refractive index detection has been used for trichothecenes such as T-2 toxin with no native fluorescence or useful UV absorption. For molecules such as T-2 toxin, with a free hydroxyl group, it is possible to derivatize with a reagent such as an aromatic acid, which confers UV absorption on the derivative.

There are well-established methods for the LC analysis of all the aflatoxins, ochratoxin A, zearalenone and its metabolites, the 8-keto trichothecenes, patulin, and the fumonisins. The fumonisin method uses precolumn derivatization with *o*-phthaldialdehyde and mercaptoethanol, despite the limited stability of the resulting compounds. Naphthalene-2,3-dicarboxaldehyde/potassium cyanide has been proposed as a viable alternative, although handling requires additional safety precautions.

### LC-MS (MS/MS)

The use of a mass spectrometer as a detector for LC analysis brings a number of benefits to mycotoxin analysis. There is no need for chromophores or fluorophores in the analytes so derivatization can be avoided. The chemical structure of the analytes can be confirmed from molecular mass and fragmentation information and the use of tandem MS (MS/MS) allows greater selectivity. Multiple reaction monitoring and selected ion monitoring modes mean that chromatographic separation of all analytes is not necessary, as differentiation is carried out by the different ion transitions measured, and many multiresidue mycotoxin LC-MS methods now exist. These data acquisition modes can also increase the sensitivity of the method as the background noise is often reduced.

Methods exist for many matrix/mycotoxin combinations including patulin in apple juice using atmospheric pressure photoionization, zearalenone and its metabolites in fish tissue and porcine urine, muscle, and liver tissues using electrospray ionization, trichothecenes (T-2 toxin, HT-2 toxin, acetyl T-2 toxin, diacetoxyscirpenol, neosolaniol, and monoacetoxyscirpenol) in grain using atmospheric pressure chemical ionization, and trichothecenes (T-2 toxin, HT-2 toxin, deoxynivalenol, nivalenol), zearalenone and its metabolites ( $\alpha$ - and  $\beta$ -zearalenol and  $\alpha$ - and  $\beta$ -zearalanol) in eggs.

### Gas Chromatography

For GC the mycotoxins need to be sufficiently thermostable and volatile to be converted into volatile derivatives. The most widely used detectors in mycotoxin GC analysis are flame ionization (FID), electron capture (ECD), and mass spectrometry (MS). These do not require that a compound show any fluorescence or UV absorption and the groups of mycotoxins for which GC is the most suited are the trichothecenes and the fumonisins.

Many mycotoxins, including the trichothecenes and fumonisins, have polar hydroxyl groups that can be readily derivatized to trimethylsilyl ethers with such reagents as trimethylsilyl chloride or hexamethyldisilazane and *N*-trimethylsilylimidazole. Alternatively, polyhalogenated acyl derivatives can be made, using reagents such as trifluoroacetic anhydride or heptafluorobutyric anhydride, which are especially useful in conjunction with the ECD.

There are a large number of different trichothecenes and ideally their analysis requires confirmation using GC-MS and analysis of the macrocyclic trichothecenes, which includes the verrucarins, roridins, and satratoxins, requires considerable experience and a specialized laboratory. However, well-established methods are available for individual trichothecenes of particular importance, such as T-2 toxin, diacetoxyscirpenol, deoxynivalenol, and neosolaniol. The large number of *Fusarium* trichothecenes are acylated derivatives of a much smaller number of parent alcohols, such as T-2 tetraol, scirpentriol, deoxynivalenol, and nivalenol, so another approach to their analysis is to hydrolyze the possibly complex mixture of trichothecenes to the parent alcohols and analyze these as their trimethylsilyl ethers. However, when this is done, there is often a poor correlation between total trichothecene content and observed toxicity, reflecting the large differences in the toxicity of different acyl derivatives even of the same parent alcohol.

Methods for the GC analysis of the fumonisins require the hydrolytic removal of the two very polar tricarboxylic acid residues and subsequent trimethylsilylation of the several hydroxyl groups and amino acid groups present in the parent molecule. Ideally confirmation with GC-MS is still desirable but only available in a very limited number of laboratories.

### Immunological Methods

Modern LC and GC apparatus in conjunction with automated sample cleanup allow multiple samples to be analyzed in one run, increasing the throughput of

analyses greatly. Immunological approaches to mycotoxin determination can also be used if a large number of analyses are required. Although mycotoxins are relatively small molecules, and hence not themselves immunologically active, they can form the haptens of an antigen by coupling to a suitable macromolecule such as bovine serum albumin. They may then present suitable epitopes to which specific antibodies can be raised.

During the last few years immunological methods have been developed for a wide range of mycotoxins including the aflatoxins, T-2 toxin, zearalenone, ochratoxin, cyclopiazonic acid, and, most recently, fumonisins B<sub>1</sub>, B<sub>2</sub>, and B<sub>3</sub>.

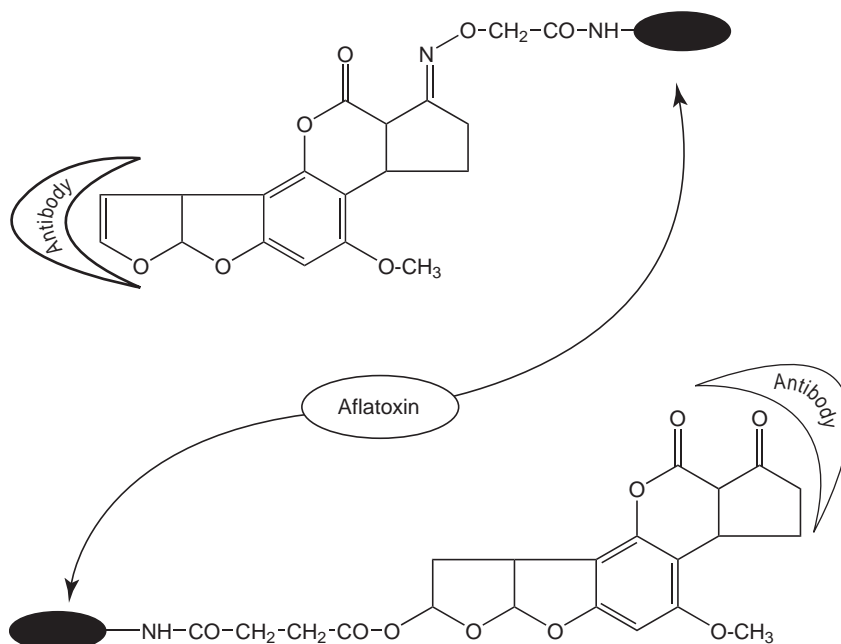
The manner in which the mycotoxin is linked to the macromolecular carrier can have a profound effect on the activity and specificity of any antibody raised, especially if it is monoclonal. Ochratoxin already has a carboxyl group that can be used directly to form a derivative with a carrier but many mycotoxins need to be linked to the carrier via a reactive linker molecule. Different linking molecules may be attached to different sites of a polyfunctional mycotoxin, thus exposing different parts of the molecule as potential epitopes. Aflatoxin B<sub>1</sub> can be derivatized through oxime formation on the pentenone ring leaving the difuran ring system exposed, or through the hemiacetal on the difuran ring leaving the  $\delta$ -lactone/pentenone ring system exposed (Figure 12). In the first case the antibody would be expected to distinguish between the unsaturated aflatoxins

(B<sub>1</sub> and G<sub>1</sub>) and their dihydrocompounds (B<sub>2</sub> and G<sub>2</sub>). In the second case the antibody would be expected to distinguish the B group from the G group.

The specificity of monoclonal antibodies make it possible to develop an analytical method for a single mycotoxin, such as aflatoxin B<sub>1</sub> in maize and groundnut meal, or aflatoxin M<sub>1</sub> in milk and milk products. Even within chemically closely related structures such as the *Fusarium* trichothecenes there is very little cross-reactivity between a monoclonal raised to a single toxin such as T-2 toxin and other members of the family. Thus, a monoclonal raised against 3-acetyl-deoxynivalenol showed negligible cross-reactivity with deoxynivalenol, nivalenol, or T-2 toxin.

### Enzyme-Linked Immunosorbent Assay

Enzyme-linked immunosorbent assay (ELISA) methods are widely used in medical and food analysis laboratories and it seems appropriate that such methods could be used for the detection of mycotoxins in foods. In the direct ELISA method the antibody is bound to a solid surface, such as a microtiter plate, and it is necessary to prepare a conjugate of the mycotoxin to be analyzed with the enzyme, such as phosphatase or peroxidase, to be used for color development. The sample or standard is mixed with the enzyme-linked mycotoxin and the two allowed to compete for the antibody bound to the surface. After washing away soluble material, the



**Figure 12** The possibility of forming two distinct antibodies to aflatoxins.



quantity of enzyme bound to the antibody, which is inversely proportional to the concentration of mycotoxin in the original sample, is determined by adding the chromogenic reagent.

Although the indirect competitive ELISA seems to be more complex it is in fact more widely used, an advantage being the stability of the mycotoxin conjugate prepared using a protein such as keyhole limpet hemocyanin, or a synthetic polypeptide such as polylysine, compared with the mycotoxin enzyme conjugate required for the direct assay. The conjugate is used to coat the surface of the assay system that might be a microtiter plate for a quantitative assay carried out in a laboratory or a preprinted plastic card for a qualitative assay suitable for application in the field. These coated surfaces are stable and can be stored for long periods of time. In the indirect method the surface-coated mycotoxin conjugate is incubated with a mixture of the sample and the specific rabbit antibody. Once soluble material has been washed away any bound antibody (which will again be inversely proportional to the concentration of mycotoxin in the original sample) is visualized with a goat antirabbit IgG enzyme complex, several of which are readily available commercially, and subsequent color development. Commercial ELISA analytical kits are available for most mycotoxins including total aflatoxins, aflatoxin B<sub>1</sub> aflatoxin M<sub>1</sub>, ochratoxin A, citrinin, zearalenone, T-2 toxin, deoxynivalenol, fumonisins, patulin, and moniliformin, and AOAC standard procedures exist.

## Biological and Biochemical Assays

Before confidence was gained in the specificity and sensitivity of physicochemical methods of analysis the acceptable assay for the presence of aflatoxins in foods and animal feeds was a biological assay using 1-day-old ducklings in which test samples were injected into the gizzard via a plastic tube. The acute LD<sub>50</sub> doses for 1-day-old ducklings have been estimated to be (mg per kg) B<sub>1</sub>, 0.36; B<sub>2</sub>, 1.69; G<sub>1</sub>, 0.78; G<sub>2</sub>, 2.45. Physicochemical and immunological methods have replaced this type of bioassay for individual mycotoxins but they also can only be applied to the analysis of known mycotoxins. If the question is more general, and there is a need to assess the presence of mycotoxins, some of which may be of unknown structure, in a sample, then a bioassay may be needed. An appropriate bioassay may alert the analyst to the presence of unknown toxic metabolites, or of synergistic interactions between several toxic metabolites in a mixture. The most

widely used bioassays involve crustacea such as brine shrimp larvae, protozoa such as *Tetrahymena*, or a range of tissue cultures of mammalian origin. Such methods are not specific and may give false-positive results because of the sensitivity of the test organism to a normal food component. Thus, brine shrimp larvae are sensitive to long-chain fatty acids.

Cytotoxicity tests, using a range of tissue cultures, would seem to provide a system more closely related to the whole animal. Baby hamster kidney cells, and human cells lines such as HE<sub>p</sub>-11, which are especially sensitive to trichothecenes, and Chang liver cells, have been successfully used for screening for the presence of toxic metabolites in food extracts.

An interesting biochemical assay has been developed for ochratoxin A, which can be split into ochratoxin  $\alpha$  and phenylalanine by the enzyme carboxypeptidase A with a loss of fluorescence intensity at 380 nm. It is a sensitive assay and has been adapted for the analysis of ochratoxin A in blood samples and seta.

## Standard Procedures

The following accepted methods, which have been evaluated and validated through collaborative studies, are extensively described by Scott, for the AOAC, who also provides detailed information on the apparatus, reagents, the preparation of standards, and sampling:

- 968.22: Aflatoxins in peanuts and peanut products – CB method
- 970.45: Aflatoxins in peanuts and peanut products – BF method
- 970.46: Aflatoxins in green coffee – TLC method
- 971.23: Aflatoxins in cocoa beans – modification of CB method
- 971.24: Aflatoxins in coconut, copra, and copra meal – TLC method
- 972.26: Aflatoxins in corn – TLC method
- 974.16: Aflatoxins in pistachio nuts – TLC method
- 975.36: Aflatoxins in foods and feeds – Romer minicolumn method
- 972.27: Aflatoxins in soy beans – TLC method
- 979.18: Aflatoxin in corn and peanuts – Holaday-Velasco minicolumn method
- 973.37: Ochratoxins in barley – TLC method
- 973.38: Sterigmatocystin in barley and wheat – TLC method
- 974.17: Aflatoxin M<sub>1</sub> in dairy products – TLC method
- 974.18: Patulin in apple juice – TLC method

- 975.38: Ochratoxin A in green coffee – TLC method
- 976.22: Zearalenone in corn – TLC method
- 978.15: Aflatoxin B<sub>1</sub> in eggs – TLC method
- 980.21: Aflatoxin M<sub>1</sub> in milk and cheese – TLC method
- 980.20: Aflatoxins in cottonseed products – TLC and liquid chromatography methods
- 982.24: Aflatoxins B<sub>1</sub> and M<sub>1</sub> in liver – TLC method
- 982.25: Aflatoxins B<sub>1</sub> and M<sub>1</sub> in liver – confirmation method
- 982.26: Aflatoxin M<sub>1</sub> in liver – confirmation method
- 985.17: Identification of aflatoxin B<sub>1</sub> – confirmation method
- 985.18:  $\alpha$ -Zearalenol and zearalenone in corn – liquid chromatography method
- 986.16: Aflatoxins M<sub>1</sub> and M<sub>2</sub> in fluid milk – liquid chromatography method
- 986.17: Deoxynivalenol in wheat – TLC method
- 986.18: Deoxynivalenol in wheat – GC method
- 989.06: Aflatoxin B<sub>1</sub> in cottonseed products and mixed feed – ELISA screening method
- 990.33: Aflatoxins in corn and peanut butter – liquid chromatography method
- 990.34: Aflatoxin B<sub>1</sub>, B<sub>2</sub>, and G<sub>1</sub> in corn, cottonseed, peanuts, and peanut butter – ELISA method
- 991.31: Aflatoxins in corn, raw peanuts, and peanut butter – immunoaffinity column (Aflatest) method
- 991.44: Ochratoxin A in corn and barley – liquid chromatography method
- 991.45: Total aflatoxin levels in peanut butter – ELISA method
- 993.16: Total aflatoxins (B<sub>1</sub>, B<sub>2</sub>, and G<sub>1</sub>) in corn – ELISA method
- 993.17: Aflatoxins in corn and peanuts – TLC method

## Safety Procedures

Mycotoxins should be handled as toxic substances and special care should be taken when handling pure toxins in a dry form, or samples (including TLC plates), which may be a source of dust. Accidental spillages of material containing toxin can generally be treated with 1% sodium hypochlorite bleach, left for 10 min, and swabbed down with 5% aqueous acetone initially and subsequently with plenty of water.

Glassware exposed to aflatoxins can be rinsed with methanol soaked in 1% sodium hypochlorite

solution to which, after 2 h, acetone is added to 5% of the total volume. The acetone is left for 30 min to destroy any potentially harmful but reactive intermediates formed by hypochlorite and then washed thoroughly. It is known, for example, that aflatoxin B<sub>1</sub> can form a carcinogenic derivative if treated with hypochlorite alone.

A very detailed discussion of the laboratory decontamination and destruction of aflatoxins in laboratory wastes has been prepared for the International Agency for Research on Cancer.

**See also:** **Bioassays:** Overview. **Derivatization of Analytes.** **Extraction:** Solvent Extraction Principles. **Fluorescence:** Food Applications. **Food and Nutritional Analysis:** Contaminants. **Gas Chromatography:** Detectors; Mass Spectrometry. **Immunoassays:** Production of Antibodies. **Immunoassays, Applications:** Food. **Immunoassays, Techniques:** Enzyme Immunoassays. **Liquid Chromatography:** Instrumentation; Liquid Chromatography–Mass Spectrometry; Food Applications. **Sampling:** Theory.

## Further Reading

- Bennett JW and Klich M (2003) Mycotoxins. *Clinical Microbiology* 16: 497–516.
- Cole RT (1986) *Modern Methods in the Analysis and Structural Elucidation of Mycotoxins*. Orlando, FL: Academic Press.
- Krogh P (1987) *Mycotoxins in Food*. London: Academic Press.
- Scott PM (1995) Natural poisons: Mycotoxins. In: *Official Methods of Analysis*, 16th edn., ch. 49, pp. 1–49. Arlington, VA: Association of Official Analytical Chemists, Inc.
- Smith JE and Moss MO (1985) *Mycotoxins: Formation, Analysis and Significance*. Chichester: Wiley.
- Stoloff L, Van Egmond HP, and Park DL (1991) Rationales for the establishment of limits and regulations for mycotoxins. *Food Additives and Contaminants* 8: 213–222.
- Trucksess MW (2003) Committee on natural toxins and food allergens – mycotoxins. *Journal of AOAC International* 86: 129–138.
- Trucksess MW and Pohland AE (2000) Methods and method evaluation for mycotoxins. *Molecular Biotechnology* 22: 287–292.
- Van Egmond HP (1989) Current situation on regulations for mycotoxins. Overview of tolerances and status of standard methods of sampling and analysis. *Food Additives and Contaminants* 6: 139–188.
- Van Egmond HP (1989) *Mycotoxins in Dairy Products*. London: Elsevier.
- Yiannikouris A and Jouany JP (2002) Mycotoxins in feeds and their fate in animals: A review. *Animal Research* 51: 81–99.

## Soft Drinks

**J Sadecka and J Polonsky**, Slovak University of Technology, Bratislava, Slovakia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

There are many kinds of soft drinks (fruit and vegetable juices, fruit and vegetable nectars, and water-based flavored drinks including 'sport' or 'electrolyte' drinks) that are formulated with different recipes and not all the ingredients are found in all soft drink products. Typically, the ingredients include water, sugars, sweeteners, preservatives, antioxidants, colorants, flavors, and pH adjusters. Ingredient contents of major types of soft drinks are given in Table 1. Most regular (nondiet) soft drinks contain ~90% of water from a pure source, while diet soft drinks contain up to 99% of water. Regular soft drinks are sweetened with cane or beet sugar, high fructose corn syrups, or blends of these sweeteners. These sugars result in a mixture of sucrose and/or fructose and glucose in the final product. Most diet soft drinks are sweetened with aspartame. The ingredients that are used in soft drinks are approved and closely regulated by the US Food and Drug Administration (FDA) and the national laws and regulations. The main additives allowed in soft drinks are listed in Table 2.

The soft drink samples are analyzed in order to monitor the quality of products, determine the contaminants, or verify the authenticity of products.

### Quality

#### Sugars

Global sugar content is obtained by indirect refractometry ('Brix') or Fourier transform infrared spectroscopy (FTIR). Although sugars are nonultraviolet (UV) absorbing molecules, the UV/UV method allows the simple determination of the global sugar content. Under the influence of UV radiation sugars are oxidized (in buffer, pH 9.0) to UV-absorbing carbonyl compounds. The formation of these compounds is monitored by UV absorption spectrometry at the wavelength of maximum absorbance of 268 nm.

Individual sugars are estimated using thin-layer chromatography (TLC) and high-performance liquid chromatography (HPLC). The main advantage of TLC is that it is rapid and inexpensive; its disadvantage is that it only provides semiquantitative information on all of the principal sugars typically encountered in drinks, e.g., monosaccharides, fructose, glucose and the disaccharides, sucrose, maltose, and lactose and also higher oligosaccharides consisting of 3–10 monosaccharides. These preliminary results are then confirmed using other techniques, e.g., enzymatic.

The most commonly used analytical technique for sugars is HPLC with a refractive index detector (RID). Although the HPLC–RID method is simple, the RID lacks sensitivity and selectivity. Therefore, UV and fluorescence detection is frequently used, coupled with pre- or postcolumn derivatization, for analysis with higher sensitivity. Liquid chromatography–mass spectrometry (LC–MS) using electrospray ionization also requires pre- or postcolumn derivatization. LC–MS using atmospheric pressure chemical ionization does

**Table 1** Ingredient content of major types of soft drinks

Flavor type	Total sugars ( $\text{g l}^{-1}$ )	Phosphorous ( $\text{mg l}^{-1}$ )	Caffeine ( $\text{mg l}^{-1}$ )	Aspartame ( $\text{mg l}^{-1}$ )	Acid (%)
<i>Regular (nondiet soft drinks)</i>					
Cola	110–13	120–200	30–140	0	0.06
Lemon–lime	110–120	0–3	0	0	0.7
Orange	120–150	0–180	0	0	0.7–1
Other citrus	90–140	0–3	0–190	0	1.1–1.6
Tonic water	90–100	Trace	0	0	0
Other regular	110–160	0–270	0–140	0	0.2–0.5
Juice added	110–150	0–220	0	0	0.5
<i>Diet soft drinks</i>					
Cola	0–trace	70–170	0–170	0–560	0.06
Lemon–lime	0	0–trace	0	0–560	0.2–0.7
Other	0–50	0–trace	0–200	0–600	0.2–0.7
Club soda	0	0–3	0	0	0
Juice added	3–20	0–170	0	400–560	0.15–0.4

**Table 2** Additives allowed in soft drinks

<i>E-number</i>	<i>Additive</i>	<i>Maximum level (mg l<sup>-1</sup>)</i>
<i>Acidity regulators</i>		
E330	Citric acid	3000
E296	Malic acid	GMP
E334–337	Tartrates	4000, GMP
<i>Antioxidants</i>		
E300–303	Ascorbic acid and its salts	GMP
E220	Sulfur dioxide	50
<i>Carbonating agent</i>		
E290	Carbon dioxide	GMP
<i>Preservatives</i>		
E210–213	Benzoates	1000
E200–203	Sorbates	1000
<i>Stabilizers</i>		
E440	Pectins	3000
<i>Sweeteners</i>		
E950	Acesulfame K	350
E951	Aspartame	600
E952	Cyclamic acid and its salts	400
E954	Saccharin and its salts	80
E955	Sucralose	300
E959	Neohesperidine dihydrochalcone	30
<i>Colorants</i>		
E102	Tartrazine	100
E104	Quinoline yellow	100
E110	Sunset yellow FCF	100
E122	Azorubin	100
E123	Amaranth	100
E124	Ponceau 4R	100
E129	Allura red	100
E132	Indigotine	100
E133	Brilliant blue	100
E143	Fast green FCF	100
E150	Caramel, classes III, IV	GMP

GMP limited by Good Manufacturing Practices.

not require pre- or postcolumn derivatization; however, this method requires  $\text{CHCl}_3$  and  $\text{CH}_2\text{Cl}_2$  to be added in the postcolumn stage. As an alternative technique, HPLC with pulsed amperometric detection is used for nonderivatized sugars.

Nonderivatized sugars including mono- and disaccharides, aminosugars, and sugar alcohols can be determined by capillary zone electrophoresis (CE) with indirect UV detection. CE is not yet in use in routine analysis. However, the method's simplicity and stability make it suitable for the application in routine labs.

### Artificial Sweeteners

Beverages that are popular as diet soft drinks are made using low-calorie artificial sweeteners such as

saccharin, aspartame, acesulfame-K, and cyclamate. Many methods are used for analyzing individual sweeteners (titration, photometric, fluorimetric, and enzymatic), but relatively few methods are capable of simultaneously analyzing several sweeteners.

HPLC with reversed-phase (RP) or ion-pair RPLC and UV detection is the most popular choice for the determination of aspartame, saccharin, and acesulfame-K. However, the lack of a chromophore in the cyclamate molecule makes its determination difficult using the common HPLC–UV detection mode. A high-performance anion-exchange chromatographic method with UV and conductivity detectors connected in series offers an attractive alternative for simultaneous determination of the four sweeteners and also citric acid.

### Preservatives

The use of benzoic and sorbic acids as preservatives is widespread. Although both acids are highly efficient against yeasts, molds, and to a lesser extent against bacteria, sorbic acid has the advantage of being active in less acidic media (unlike pH values of 6) and in addition has no taste. HPLC including reversed-phase, ion-exclusion, and ion-exchange chromatography is the most common analytical procedure for the determination of these acids.

### Antioxidants

Antioxidants are usually classified into two groups: natural (e.g., ascorbic acid, flavonoids, tocopherols) and synthetic.

HPLC with UV-based diode array detection (DAD–UV) or electrochemical detection is normally used to determine ascorbic acid. Many types of electrochemical determinations of ascorbic acid have been proposed. Although the electrochemical determinations using enzyme-based biosensors exhibited high specificity and sensitivity, these methods suffer in the fabrication of the electrodes and in automatic analysis. Recently, chemically modified screen-printed electrodes have been constructed for the determination of ascorbic acid. This is one of the most promising routes for mass production of inexpensive, reproducible, and reliable electrochemical sensors.

The analyses of flavonoids are usually carried out by HPLC with UV or DAD–UV detection. Since flavonoids are found as complex mixtures, hyphenated techniques are needed. Among those, LC–MS represents a rapid and reliable technique.

### Colorants

The colorants permitted for use in drinks can be divided in two categories: synthetic (azo or

triarylmethane dyes) and natural (anthocyanins or caramel).

Today, synthetic colorants have largely replaced natural colorants and there are laws that strictly control their use. The maximum amounts allowed and the acceptable daily intakes have been decreased. As these values differ for various colorants, it appears necessary to identify and quantify the colorants present in drinks.

Usually, only two or three colorants are added. Due to this reason and taking into account the economic feasibility and rapidity, double division-ratio spectra derivative, inverse least-squares, and principal component regression methods are reliable for the simultaneous determination of the colorants in the drinks without a priority procedure such as separation, extraction, and preconcentration.

The chromatographic methods are applied when the sample contains many different colorants. RPLC and ion-pair RPLC with DAD or MS detection are the most widely used methods. Both isocratic and gradient systems are used and the latter is preferred for the separation of more complex mixtures. Although some samples can be injected directly, for most LC methods colorants must be extracted first from the matrices and purified as well as concentrated prior to chromatography. All the pretreatment techniques are time consuming and sometimes expensive, making them unsuitable for routine analysis. High-performance ion chromatography on an anion-exchange column with very low hydrophobicity proved to be a promising method for the separation of colorants. No time-consuming pretreatment, as used in conventional chromatography, is needed.

Anthocyanins are responsible for the blue, red, and purple colors of fruits. The anthocyanin profiles of some products derived from red fruits are used to verify the authenticity of these products. RPLC is the most common method for the separation of anthocyanins.  $C_{18}$  columns are commonly used for the separation and single wavelength detectors can selectively monitor anthocyanins between 520 and 546 nm, where no other plant phenolics show absorption. The separation is accomplished more effectively with gradient elution using methanol or acetonitrile as an organic modifier. The pH of the gradient system is normally kept below two. Anthocyanins can also be separated by micellar electrokinetic chromatography (MEKC). The MEKC analysis requires lesser time not only for separation, but also for reconditioning the capillary.

The caramels used in soft drinks are classified as Class IV caramels. The established method for determination of caramels is a simple spectrometric measurement at 610 nm. This method cannot distinguish

between caramel classes or between caramel and other compounds that absorb at 610 nm. CE at acidic pH can discriminate between Class I, III, and IV caramels. A reliable, robust method for the analysis of Class IV caramels by CE has recently been developed.

### Aroma and Flavor

Natural aroma and flavors in soft drinks come from spices, natural extracts, and oil. There are also some artificial flavorings (quinine, caffeine, naringenin, and hesperidin).

Aroma and flavor compounds usually occur at low concentrations and consist of a wide range of organic compounds. Most aroma and flavor compounds are volatile and procedures for isolation from samples such as steam distillation, solvent extraction, and trapping of the volatiles on adsorbent require a long time prior to chromatographic separation. The volatile compounds are also analyzed using either direct headspace or dynamic headspace purge-and-trap methods. The headspace solid-phase microextraction (SPME) in combination with gas chromatography (GC) or GC-MS can be used for the analysis of various juices. The SPME methods coupled with HPLC or LC-MS are used for the analysis of less volatile or thermally labile compounds. These techniques provide accurate measurements and define the right chemical nature of the aroma compounds. However, they are complicated and expensive, unsuitable for real-time, online, and *in situ* sensing.

Sensorial analyses based on the sensor arrays and pattern recognition algorithms seem to be very promising in encountering the mentioned advantages. Various kinds of gas sensors are available, but only metal oxide semiconductors and conducting organic polymers are used in commercialized electronic noses. The electronic nose can only perform quick 'yes' or 'no' test to compare products. Therefore, it cannot totally replace either the human nose or other analytical methods.

Following the development of the electronic nose for gas analysis, the electronic tongue (an array of potentiometric chemical sensors) has been devised for the analysis of complex liquid samples. If properly configured and trained (calibrated), the electronic tongue is capable of recognizing the qualitative and quantitative composition of multicomponent solutions of different natures. The electronic tongue and electronic nose, in combination, can be used to predict the sensory characteristics and their relationship to the quality of, e.g., apple juices measured by a trained sensory panel and consumers.

Quinine is a flavoring agent used in soft drinks. Several methods are available for the determination

of quinine (UV-vis spectrometry, fluorimetry, ion-selective electrode) but these methods are generally not very selective. There are certain drawbacks to the spectrophotometric methods such as strong influence from benzoic acid and the presence of coloring materials. The same problems will be expected when the fluorimetric method is applied. The use of ion-selective electrode is advantageous over the spectral methods because no interference from those species is expected. Chromatographic techniques, especially LC, are widely used with quinine detection by means of its fluorescence or UV absorption.

In some cases, small amount of caffeine is added to soft drinks. Caffeine has a classic bitter taste that enhances other flavors. UV spectrometry, HPLC, and CE are used for the determination of caffeine in beverages. All of the above methods require sample preparation. FTIR spectroscopy with chemometrics, using partial least squares first derivative spectrum in the region between 2800 and 3000  $\text{cm}^{-1}$ , have been developed to directly determine the caffeine in soft drinks.

Naringenin and hesperidin are examples of bittering compounds used in soft drinks. RPLC with UV-DAD is the method of choice for the analysis of naringenin and hesperidin.

## Acids

Acids are used in soft drink to balance the sweeteners and to inhibit the growth of microorganisms such as yeasts, molds, and bacteria. Soft drinks are acidified either by the addition of fruit juices or by the inclusion of an acid. The three most commonly used acids are citric, malic, and phosphoric. Other acids such as tartaric and fumaric are also used. Citric acid is found in citrus fruits, blackcurrants, strawberries, and raspberries. Malic acid is found in apples, cherries, plums, and peaches. Phosphoric acid is generally used in cola drinks.

Titrimetry is routinely used to determine total titratable acids (results are expressed as gram anhydrous citric acid per kilogram).

HPLC and enzymatic methods are used for the determination of citric and malic acids in routine analysis. Tartaric, fumaric, and phosphoric acids are normally determined by HPLC. The choice of a given HPLC method depends on the nature of acids and sample matrix. The most general method is ion-exclusion chromatography followed by RP chromatography because these two methods provide better, faster, and more reliable data compared to ion-exchange and ion-pair RPLC. The better-shape peaks are obtained employing ion-exclusion

chromatography, and faster analysis times are achieved using RPLC.

## Carbon Dioxide

Carbon dioxide is an essential characterizing ingredient in all 'carbonated' beverages. It can be measured volumetrically, manometrically, or automatically using a thermal conductivity detector.

A new method for the determination of dissolved  $\text{CO}_2$  in beverages is based on passing the liquid over the retentate side of a membrane that is at least partially permeable to the gas, measurement of the volumetric flow of the gas to the permeate side, measurement of liquid temperature, and calculation of the gas concentration using these data. Carbonate can also be determined using online coupling of gas diffusion to a CE system.

## Miscellaneous

HPLC is used increasingly in the analysis of soft drink samples to separate and detect natural compounds and additives. Differences in the respective molecular characteristics of this range of compounds would make the simultaneous separation of these compounds by HPLC quite difficult although ion-pair RP or ion-exchange methods could be used (Table 3). The high separation capabilities of CE, and ease with which the running buffer and/or the micellar phase can be modified, suggests it may be suitable for complex determinations (Table 4). However, both these methods do not allow easy, continuous monitoring, because they are 'slow'. Biosensors offer advantages as alternative to conventional methods due to their specificity, simplicity, and quick response.

## Contaminants

### Microbial Spoilage

Soft drinks have high water activity ( $a_w$ ) that permits microbial growth. Combinations of hurdles, such as pH and chemical preservatives, prevent the growth of most microorganisms. A variety of microorganisms may be found in soft drinks; however, only a few acid-tolerant spoilage organisms are of significance. Molds grow only when dissolved oxygen is present, as may be the situation in certain noncarbonated drinks. Acid-tolerant bacteria grow well at a soft drink pH level. Since yeasts can tolerate extreme conditions better than bacteria, they are often found in acidic products and products containing preservatives to such an extent that bacteria cannot grow. Especially, yeast spoiling has increased in recent years as a result of lower doses of preservatives



**Table 3** Simultaneous separations by HPLC methods

<i>Compounds</i>	<i>Column type</i>	<i>Mobile phase</i>	<i>Detection</i>
Inorganic anions and carboxylic acids	Ion-exchange	Sodium hydroxide gradient in water–methanol–ethanol	Suppressed conductivity
Sugars, ascorbic acid, and furanic compounds	Ion-exchange	Aqueous sulfuric acid, acetonitrile	RI and diode-array UV
Carboxylic acids, sugars, glycerol, and ethanol	Ion-exchange	Aqueous sulfuric acid	RI
Organic acids and phenolic compounds	Reversed phase	Methanol gradient in sulfuric acid	Diode-array UV
Antioxidants, benzoic and sorbic acids, and sweeteners	Reversed phase	$\alpha$ -Hydroxyisobutyric acid, hexadecyltrimethylammonium bromid, acetonitrile	UV

RI, refractive index.

**Table 4** Simultaneous separations by CE methods

<i>Compounds</i>	<i>Method</i>	<i>Electrolyte</i>	<i>Detection</i>
Inorganic ions, organic acids, amino acids, and carbohydrates	CE	2,4-Pyridine dicarboxylic acid, cetyltrimethylammonium hydroxide, pH 12.1	Indirect UV
Antioxidants, benzoic and sorbic acids, and sweeteners	MEKC	Borate, sodium cholate, sodium dodecyl sulfate, methanol, pH 9.3	Direct UV
Colorants, preservatives, and sweeteners	MEKC	Carbonate, sodium dodecyl sulfate, pH 9.5	Direct UV–vis

CE, capillary zone electrophoresis; MEKC, micellar electrokinetic chromatography.

required for high standards of food quality. The most known alternatives are characterized by abundant gas production, cloudiness, sediment and off-flavor formation, and off-tastes.

The analytical methods used by the industry to evaluate yeasts present in drinks are still, generally, laborious and provide poor information. The most common microbial indicator is still the ‘yeasts and molds’, resulting from the utilization of a rich culture medium (supplemented with antibiotics to prevent bacterial growth) to enumerate yeasts and molds. However, this indicator does not distinguish between the dangerous spoilage yeasts and others.

Several methods have been developed for the differentiation of yeasts. Traditional plating techniques may be adapted using selective and/or differential plating media. Such media have been developed for *Zygosaccharomyces bailii*, the most important of all food spoilage yeasts. However, yeast differentiation by differential media is poorly developed when compared with similar works applied to bacteria. Other phenotyping methods include conventional yeast identification by means of assimilation or fermentation tests and the use of morphological characteristics. Conventional methods are not suited to industrial laboratories even when these procedures are automated and computerized.

Another approach concerns the use of biomarkers, such as the composition of cellular molecules like fatty acids and nucleic acids (DNA and RNA). Fatty acid composition has been used successfully in differentiating the most important yeasts. A gas chromatograph is required to implement this technique. Based on the presence or absence of polyunsaturated C18 fatty acids it is possible to separate yeasts in groups of three levels. The most dangerous yeasts for fruit concentrates and juice industry are in group III. The yeasts of group I may be dangerous to the fruit concentrates, fruit juices, and soft drinks industry. The yeasts of group II are considered as dangerous. The difficulty for the food industry is the absence of a database of long-chain fatty acid profiles for the main food contaminant yeasts. From a technological point of view, fatty acid profiling of yeasts is more easily applicable than nucleic acid methods. Fatty acid profiles would give the probable identification of a given yeast strain, which could be confirmed using a molecular primer of the species expected to have been found.

### **Mycotoxins**

Mycotoxins are highly toxic compounds produced by fungi. Thus, in fruit juices several mycotoxins

have been identified: alternariol is produced by *Alternaria alternata* in apple, cranberry, grape, and raspberry juice. *Byssoschlamys nivea* in grape, pineapple, apple, orange, and tomato juices produced by ochlamic acid. Citrinin is produced by *Penicillium expansum* in apple and grape juices. Ochratoxin A is produced by *Aspergillus carbonarius* in grape juices. Patulin is produced by *Byssoschlamys fulva*, *Byssoschlamys nivea*, and *Penicillium expansum* in fruit juices.

Currently most mycotoxins are still assayed using TLC, which permits effective compound separation and characterization. When the compounds to be detected are fluorescent, a fluorodensitometry reads the plates quantitatively and objectively. However, the use of RPLC has expanded rapidly in recent years. The diversity of detection systems (DAD, fluorescence, and MS) permits identification and accurate assay of a great variety of these compounds. Recently, a database for 474 mycotoxins and fungal metabolites for dereplication by standardized LC–UV–MS has been reported. Sample preparation techniques include: extraction of mycotoxins using mostly chloroform, acetone, or methanol, purification of the extract with cleanup methods, and concentration of the extract.

Alternariol can be determined using gradient RPLC with UV detection at 254 nm, although fluorescence or MS detection have also been used. Sample preparation involves extraction with dichloromethane and purification of the extract by silica column chromatography using diethyl ether–hexane in the washing step and ethyl acetate–methanol for elution of analyte. Besides, cleanup of the sample can be carried out in two steps using first a polypropylene C<sub>18</sub> SPE column, followed by aminopropyl SPE column.

The usual methods for the assessment of ochratoxin A are immunochemical detection or RPLC coupled with fluorescence detection ( $\lambda_{\text{ex}} = 247 \text{ nm}$ ,  $\lambda_{\text{em}} = 480 \text{ nm}$ ). For ochratoxin sample preparation, the sample is mixed with HCl and MgCl<sub>2</sub>, followed by extraction into toluene. The toluene supernatant is passed through a silica gel column and the column is washed with hexane, toluene–acetone, and toluene. Ochratoxin A is eluted with toluene–acetic acid and dried down to 40°C. The residue is dissolved in a mobile phase, filtered, and analyzed by HPLC. The limit of ochratoxin A in grape juice is proposed to be fixed at 3 µg per kg.

The patulin level can be determined using gradient RPLC with UV detection (270 nm). For patulin sample preparation, two procedures are described: (1) juice sample can be cleaned on an Extralut cartridge followed by patulin extraction on a silica gel column

with toluene–ethyl acetate; (2) patulin can be extracted into ethyl acetate followed by extraction of phenolic acids into carbonate solution. Next, the carbonate layer is back-extracted with ethyl acetate. After evaporation of the combined ethyl acetate layers, the residue is dissolved in methanol–ethyl acetate and it is to be analyzed by RPLC. Researchers recommended that apple juice should not contain residual patulin level in excess of 50 µg l<sup>-1</sup>.

Detection of mycotoxins usually requires a time- and solvent-consuming sample preparation step. A new, simple procedure based on diphasic dialysis extraction coupled with GC–MS has been demonstrated to be effective in reducing solvent consumption for the determination of patulin in apple juices. Using a semipermeable membrane, patulin was extracted from an apple juice sample into a methane chloride organic phase. An improvement in *in situ* derivatization and higher extraction recoveries have been obtained by means of the addition of 4-*N,N*-dimethylaminopyridine as derivatization activator and acetic anhydride as derivatizing agent to the sample and to the dialyzing solvent, respectively.

### Anionic Toxins

A rapid determination of anionic toxins (cyanide, arsenite, arsenate, selenate, and azide) in beverages is possible using CE. These toxins can be detected within 15 min, requiring only minimal sample pretreatment. Compared to ion chromatography, the advantages of CE are the shorter analysis time, the minimal sample pretreatment, and the simultaneous analysis of azide and arsenate together with cyanide and arsenite.

### Residues of Pesticides

In most countries, growing concern about the residue of pesticides in beverages is evident. Regulations limiting the concentrations of pesticides have been introduced. Several methods are used to control these limits. TLC or bioanalytical techniques involving inhibition of enzyme reactions, bioassays using fungal spores, or immunological techniques may be used as an initial screening to detect whether a residue is present before a sample is subjected to a more complex instrumentation analysis (GC–MS or HPLC).

### Authenticity

Adulteration of fruit juices and concentrates is a problem that has received wide publicity. Common methods of adulteration either alone or in combination include addition of water, corn syrup, cane and beet invert sugar, peel extract, pulp wash, cheaper

**Table 5** Authenticity methods for juice

<i>Method</i>	<i>Provision</i>	<i>Adulteration</i>
IRMS	Carbon stable isotope ratio of sugars	Addition of corn syrup or cane sugar to most juice with the exception of pineapple
IRMS	Oxygen stable isotope ratio in fruit juice water	The fraudulent sale of juice made from concentrated as freshly squeezed
IRMS/NMR	Combination of five isotope ratio	Addition of beet sugar
Pyrolysis-MS	Pectin	Addition of pulp wash and/or peel extract
Liquid chromatography	Ratio hesperidin/narirutin, summary flavonoid glycoside and hesperidin	Addition of pulp wash to orange juice
Liquid chromatography	Organic acids profile, minor acids, citric/isocitric acid ratio	Addition of cheaper juice, e.g., grapefruit added to orange juice can be detected by tartaric acid
Liquid chromatography/enzymatic	Sugar profile of sucrose, glucose and fructose or amino acids, e.g., D-malic acid	Addition of cane and beet invert sugar, addition of D,L-malic acid

MS, mass spectrometry; NMR, nuclear magnetic resonance.

juices, colorants, and other undeclared additives, often to mimic the composition profiles of pure juices. Because of the diversity in adulteration techniques, a variety of test methods may be necessary (Table 5). Multiple-component chemical analyses are required to reliably evaluate the differences between adulterated and pure juice. This approach, however, is both time consuming and expensive. Near-infrared spectroscopy seems to be a very useful tool to reduce the cost of authenticity checking. In the case of a good database, near-infrared spectroscopy could be proposed as a rapid preliminary screening method to find the out-layers differing from the other authentic samples. Out-layers only have to be analyzed using the more expensive and time-consuming analytical procedure.

Carbon stable isotope ratio (CSIR) analysis of sugar utilizing an isotope ratio mass spectrometer (IRMS) is routinely used to detect the addition of cane sugar and corn syrup to most juices with the exception of pineapple. The inability of CSIR analysis to detect beet sugar in orange and apple juices because of metabolic similarities between adulterant and authentic products has required alternative methods, e.g., combination of IRMS and nuclear magnetic resonance (NMR) spectroscopy. The method requires the determination of five isotope ratios after fermentation of the juice: the  $^2\text{H}/^1\text{H}$  and  $^{18}\text{O}/^{16}\text{O}$  ratios of the fermentation water plus the total  $^{13}\text{C}/^{12}\text{C}$  and  $^2\text{H}/^1\text{H}$  ratios on the methyl and methylenic groups of the alcoholic distillate.

Flavonoid glycosides provide an excellent fingerprint for juices that are not subjected to extensive pretreatment to remove them. There are a large number of these compounds in citrus juice. Hesperidin and narirutin are typical of sweet orange, whereas naringin and neohesperidin can be used to

**Table 6** Hesperidin/narirutin ratios and naringin/neohesperidin ratios in citrus juices

<i>Citrus juice</i>	<i>Hesperidin/ narirutin ratio</i>	<i>Naringin/ neohesperidin ratio</i>
Orange	4–12	–
Orange pulp wash	2.5–3.9	–
Murcott	0.3–0.5	–
Tangerine	2–3	–
Lemon	> 7	–
Grapefruit	< 0.1	15–80
Pommerans	> 4	1
Bergamot	> 4	1.5

identify grapefruit (Table 6). The ratio hesperidin/narirutin (H/N) in pulp wash differs from that of orange juice and if this ratio is low, ranging from 2.5 to 3.9, the orange juice can contain pulp wash. For increasing amounts of murcott juice, or grapefruit juice in orange juice, narirutin increases, resulting in a lower H/N ratio. Blending with grapefruit, pommeranse, or bergamot juices may clearly be proved by the presence of naringin and neohesperidin. On the basis of the naringin concentrations, it even may be possible to trace the addition of very low percentages (<2%) of these juices in orange juice.

Fruit juices each have very distinct organic acid profiles that can be used as fingerprint for establishing authenticity. HPLC methods that compare ratios of the major acid components are limited in most cases because the adulterators can add synthetic acids so that the ratios are consistent with those found in pure juice. Some minor acids analysis is a more powerful technique, since it is not economically feasible to adjust the levels of all acids, many of which are expensive. D-isocitric acid in lower

**Table 7** The content of D-isocitric acid and citric acid/D-isocitric acid ratio in different fruit juices

Fruit juice	D-isocitric acid content (mg l <sup>-1</sup> )	Citric acid/D-isocitric acid ratio
Orange	65–200	80–130
Grapefruit	140–350	50–90
Lemon	240–420	170–260
Mandarin	70–90	120–140
Raspberries	60–220	80–200
Apricot	70–200	15–150

concentrations can be found in various fruits (Table 7). The determination of D-isocitric acid is important especially in the evaluation of citrus juices. Simple falsification of orange concentrate or juice by addition of sugars, citric acid, and water can be detected from the citric acid/D-isocitric acid ratio, which is usually lower than 130 in authentic orange juice.

See also: **Carbohydrates:** Sugars – Chromatographic Methods. **Fluorescence:** Food Applications. **Immunoassays, Applications:** Food. **Ion-Selective Electrodes:** Food Applications. **Mass Spectrometry:** Food Applications. **Nuclear Magnetic Resonance Spectroscopy**

**Applications:** Food. **Radiochemical Methods:** Food and Environmental Applications.

## Further Reading

- Gilbert J and Anklaam E (2002) Validation of analytical methods for determining mycotoxins in foodstuffs. *Trends in Analytical Chemistry* 21: 468–486.
- Heftmann E (ed.) (2000) *Chromatography and Electrophoresis in Food Analysis*. Amsterdam: Elsevier.
- Loureiro V and Querol A (1999) The prevalence and control of spoilage yeast in foods and beverages. *Trends in Food Science & Technology* 10: 356–365.
- Mello LD and Kubota LT (2002) Review of the use of biosensors as analytical tools in the food and drink industries. *Food Chemistry* 77: 237–256.
- Nollet LML (2000) *Food Analysis by HPLC*. New York: Marcel Dekker.
- Sancho T, Gimenez-Jurado G, Malfeito-Ferreira M, and Loureiro V (2000) Zymological indicators: A new concept applied to the detection of potential spoilage yeast species associated with fruit pulps and concentrates. *Food Microbiology* 17: 613–624.
- Schaller E, Bosset JO, and Escher F (1998) 'Electronic noses' and their application to food. *Lebensmittel-Wissenschaft und-Technologie* 31: 305–316.
- Wetzel DLB and Charalambous G (eds.) (1999) *Instrumental Methods in Food and Beverage Analysis*. Amsterdam: Elsevier.

## Coffee, Cocoa, and Tea

S Y Tse, Wyeth-Ayerst Research, Princeton, NJ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1512–1517, © 1995, Elsevier Ltd.

## Introduction

The beverages coffee, cocoa, and tea can be considered complex mixtures of naturally occurring substances (for example, methylxanthine alkaloids, flavonoids, organic acids, pigments, and nutritional elements) and compounds generated during industrial processes, such as roasting, fermentation, and decaffeination. This article will provide a brief overview of the major chemical components commonly encountered in coffee, cocoa, and tea, together with the standard procedures for the extraction and analyses of these major components.

## Survey of Major Analytes

The major analytes of coffee include caffeine, chlorogenic acids, and flavor and volatile aromatic components. The major analytes in cocoa are methylxanthines, mainly theobromine and trace amounts of caffeine, cocoa fat, and lipids. Other analytes of interest in cocoa are tannins, pigments, and aroma components. The major analytes of tea are the methylxanthine alkaloids, including caffeine and theophylline, polyphenols (catechins, tannins, and related flavanols), and volatile and aromatic components. Analysis of black tea would also include theaflavins and thearubigens, which are oxidation and condensation products of polyphenols.

## Sample Preparation and Extraction Techniques

### Sample Preparation

Green coffee beans are ground to pass a No. 40 sieve before extraction and analysis. Roasted coffee beans

**Table 7** The content of D-isocitric acid and citric acid/D-isocitric acid ratio in different fruit juices

Fruit juice	D-isocitric acid content (mg l <sup>-1</sup> )	Citric acid/D-isocitric acid ratio
Orange	65–200	80–130
Grapefruit	140–350	50–90
Lemon	240–420	170–260
Mandarin	70–90	120–140
Raspberries	60–220	80–200
Apricot	70–200	15–150

concentrations can be found in various fruits (Table 7). The determination of D-isocitric acid is important especially in the evaluation of citrus juices. Simple falsification of orange concentrate or juice by addition of sugars, citric acid, and water can be detected from the citric acid/D-isocitric acid ratio, which is usually lower than 130 in authentic orange juice.

See also: **Carbohydrates:** Sugars – Chromatographic Methods. **Fluorescence:** Food Applications. **Immunoassays, Applications:** Food. **Ion-Selective Electrodes:** Food Applications. **Mass Spectrometry:** Food Applications. **Nuclear Magnetic Resonance Spectroscopy**

**Applications:** Food. **Radiochemical Methods:** Food and Environmental Applications.

## Further Reading

- Gilbert J and Anklaam E (2002) Validation of analytical methods for determining mycotoxins in foodstuffs. *Trends in Analytical Chemistry* 21: 468–486.
- Heftmann E (ed.) (2000) *Chromatography and Electrophoresis in Food Analysis*. Amsterdam: Elsevier.
- Loureiro V and Querol A (1999) The prevalence and control of spoilage yeast in foods and beverages. *Trends in Food Science & Technology* 10: 356–365.
- Mello LD and Kubota LT (2002) Review of the use of biosensors as analytical tools in the food and drink industries. *Food Chemistry* 77: 237–256.
- Nollet LML (2000) *Food Analysis by HPLC*. New York: Marcel Dekker.
- Sancho T, Gimenez-Jurado G, Malfeito-Ferreira M, and Loureiro V (2000) Zymological indicators: A new concept applied to the detection of potential spoilage yeast species associated with fruit pulps and concentrates. *Food Microbiology* 17: 613–624.
- Schaller E, Bosset JO, and Escher F (1998) 'Electronic noses' and their application to food. *Lebensmittel-Wissenschaft und-Technologie* 31: 305–316.
- Wetzel DLB and Charalambous G (eds.) (1999) *Instrumental Methods in Food and Beverage Analysis*. Amsterdam: Elsevier.

## Coffee, Cocoa, and Tea

S Y Tse, Wyeth-Ayerst Research, Princeton, NJ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1512–1517, © 1995, Elsevier Ltd.

## Introduction

The beverages coffee, cocoa, and tea can be considered complex mixtures of naturally occurring substances (for example, methylxanthine alkaloids, flavonoids, organic acids, pigments, and nutritional elements) and compounds generated during industrial processes, such as roasting, fermentation, and decaffeination. This article will provide a brief overview of the major chemical components commonly encountered in coffee, cocoa, and tea, together with the standard procedures for the extraction and analyses of these major components.

## Survey of Major Analytes

The major analytes of coffee include caffeine, chlorogenic acids, and flavor and volatile aromatic components. The major analytes in cocoa are methylxanthines, mainly theobromine and trace amounts of caffeine, cocoa fat, and lipids. Other analytes of interest in cocoa are tannins, pigments, and aroma components. The major analytes of tea are the methylxanthine alkaloids, including caffeine and theophylline, polyphenols (catechins, tannins, and related flavanols), and volatile and aromatic components. Analysis of black tea would also include theaflavins and thearubigins, which are oxidation and condensation products of polyphenols.

## Sample Preparation and Extraction Techniques

### Sample Preparation

Green coffee beans are ground to pass a No. 40 sieve before extraction and analysis. Roasted coffee beans

and tea samples are ground to pass a No. 30 sieve before analysis. Soluble coffee solids, instant tea, and cocoa powder are analyzed after thorough mixing to ensure homogeneity. Cocoa products other than cocoa are extracted with petroleum ether to remove fat before extraction and analysis of alkaloids and other nonvolatile components.

## Extraction Techniques

### Methylxanthine alkaloids

*Caffeine in coffee and tea* The Bailey–Andrew method, the Power–Chesnut method, and the Levine method have been adopted by the Association of Official Analytical Chemists (AOAC) for the extraction and determination of caffeine in coffee and tea.

In the Bailey–Andrew method, 5–10 g of ground sample is weighed into a 1 l Erlenmeyer flask with 500 ml of water. The mixture is heated to boiling, and 10 g of magnesium oxide (MgO) is added. The mixture is boiled over low flame for a period of 2 h, water being added periodically to prevent frothing. After cooling, water is added to make to weight (1000 g) and the mixture is filtered. Twenty milliliters of sulfuric acid is added to 200 ml of clear filtrate. The acidified sample is extracted successively with 25, 20, 15, 10, 10, 10 ml portions of chloroform. Five milliliters of potassium hydroxide solution (1%) is added to the pooled chloroform extract. The remaining aqueous potassium hydroxide layer is washed with two 10 ml portions of chloroform. All chloroform portions are collected into a Kjeldahl flask. After evaporation of the chloroform, the amount of caffeine extracted is determined by measuring the total nitrogen (Kjeldahl method).

In the Power–Chesnut method, 10 g of coffee sample is transferred into a Soxhlet extractor with ethanol, and extracted with ethanol for up to 8 h. The ethanolic extract is transferred into a porcelain dish containing 10 g of high-density MgO and 100 ml of water. Ethanol in the mixture is evaporated by slow heating on a steam bath. The residue is moistened with hot water and filtered into a 1 l flask. The filtrate is acidified with 20 ml of sulfuric acid and brought to boiled for 30 min. After cooling, the acid extract is filtered into a separator. The filtrate is extracted with chloroform (six 25 ml portions). The combined chloroform extract is washed with 5 ml of 1% KOH solution. The chloroform phase is drained into an Erlenmeyer flask and the KOH portion is washed with two portions of chloroform (10 ml each). All chloroform extracts are pooled, and the volume of extract is reduced to 10 or 15 ml by distillation. The remaining chloroform extract is transferred to a weighing beaker, and the residual solvent is

evaporated. The residue is dried for 30 min at 100°C. The amount of caffeine is determined by weighing or by nitrogen determination using the Kjeldahl method.

The Levine method is used in preparing samples for the spectrophotometric determination of residual caffeine in decaffeinated coffees. An acidic Celite column is prepared by adding 2 ml of sulfuric acid ( $4 \text{ mol l}^{-1}$ ) to 2.0 g of Celite 545. The mixture is transferred into a 25 mm ID  $\times$  250 mm glass column plugged with fine glass wool at both ends. A basic column is prepared by mixing 3 g of Celite 545 with 2 ml of sodium hydroxide ( $2 \text{ mol l}^{-1}$ ); this is then transferred into a 25 mm ID  $\times$  250 mm glass column with a fine glass wool plug at the bottom. Coffee sample (1.0 g decaffeinated coffee or 0.5 g decaffeinated soluble coffee) is treated with 5 ml of ammonia solution and heated over a steam bath for 2 min. Six grams of Celite 545 is added to adsorb the alkaline mixture and the alkali–coffee–Celite mixture is applied directly onto the basic Celite column. The basic column is mounted on top of the acidic column. Water-saturated ether (150 ml) is passed through both basic and acidic columns. The acidic column is washed with an additional 50 ml of ether. The ether portions are discarded. Caffeine is eluted from the acidic column with 50 ml of water-saturated chloroform solution.

*Theobromine in cocoa* The Wadsworth method is used for the extraction of theobromine from cocoa powder or defatted cocoa products. Ten grams of sample is put in a porcelain dish and mixed with 2–3 g of fresh calcined MgO. Water (9–20 ml) is added, a few milliliters at a time, and mixed thoroughly to ensure uniform consistency and wetting of all particles. The mixture is heated on a steam bath for 30 min, with mixing at regular intervals to maintain uniform consistency and to prevent drying. After heating, the mixture is transferred into a 250 ml flask and refluxed with 150 ml of tetrachloroethane for 30 min. The hot extract is filtered into a 200 ml flask. The residue is refluxed with three additional portions of tetrachloroethane (120 ml) for up to 30 min. Solvent in the 200 ml flask is removed by distillation after each reflux and filtration step, until the volume is reduced to between 3 ml and 5 ml. The residue is cooled and treated with 65 ml of ether, mixed by swirling the flask. The ether mixture is stoppered and left standing until the supernatant is clear (up to an hour or more). The ether mixture is filtered with preweighed filtered paper and dried at 100°C. The weight of precipitated theobromine collected after filtration and removal of ether is determined by weighing. A correction of 0.004 g is added to account for the amount of theobromine dissolved in the ether.



An aqueous extraction method is used to extract both the theobromine and caffeine from defatted cocoa products for analysis by liquid chromatography (LC). In this method, 0.6 g of cocoa is placed in an extraction tube with a polytetrafluoroethylene-lined screw cap. The sample is extracted twice with 30 ml portions of petroleum ether, centrifuged, and the ether phase, which contains fat, is discarded. The residue, after removing all residual solvent, is transferred into an Erlenmeyer flask with boiling chips and added with 95 ml of water. The mixture is boiled for 25 min. After cooling, several portions of the mixture are transferred into screw-capped centrifuge tubes and centrifuged at 2000 g for 5 min. The supernatant is filtered through a 0.45  $\mu\text{m}$  membrane filter, and the filtrate is analyzed by LC.

### Organic Acids, polyphenols, and lipids

**Chlorogenic acids in coffee** One gram of roasted ground coffee is transferred into a 750 ml Erlenmeyer flask. Boiling water (400 ml) is added, and the mixture is boiled gently for 15 min, then cooled rapidly under cold tap water. The mixture is transferred into a 500 ml volumetric flask, diluted to volume, and filtered. The first 25–50 ml of the filtrate is discarded. Chlorogenic acid in the filtrate is determined by a spectrophotometric or chromatographic method. For soluble coffee solids, 0.35 g of sample is transferred into a 500 ml flask and diluted to volume. For green coffee, 0.7 g of ground sample is extracted with petroleum ether (25 ml) in a 50 ml centrifuge tube. After mixing and centrifugation, petroleum ether is removed by decanting. The residue is washed with two additional portions of petroleum ether. Residual solvent is removed by drying under a stream of dry air. The residue is transferred with a few milliliters of water into the 750 ml Erlenmeyer flask. Boiling water is added and the sample is extracted following the same procedure as in roasted coffee. Alternative extraction procedures using aqueous alcohols have also been employed.

**Cocoa fat and lipids** The lipid components in cocoa are extracted into petroleum ether by either the Knorr Tube method or by Soxhlet extraction. Both are adopted by AOAC, and the latter method by both AOAC and the Office International du Cacao et du Chocolat. In the Knorr Tube method, 2–3 g of cocoa powder or grated chocolate product is transferred into a Knorr Tube fitted with 6 mm of tightly packed mat of washed asbestos filter. The extraction tube is connected to suction by a two-way stopcock, and the stem of the extraction tube is connected to a preweighed Erlenmeyer flask. The extraction tube is filled up to two-thirds capacity by petroleum ether.

After thorough mixing, the ether phase is separated by standing, and drained into the Erlenmeyer flask by suction. The extraction process is repeated up to 10 times for complete extraction of fat. The solvent is evaporated and the sample is dried at 100°C. The amount of fat in the sample is determined by weighing.

In the Soxhlet method, 4–5 g of cocoa powder is placed in a 500 ml beaker and treated with 45 ml of boiling water with stirring. The mixture is acidified with hydrochloric acid (8 mol l<sup>-1</sup>, 55 ml). Antibumping beads are added and mixture is boiled for 15 min. The acid digest is filtered after boiling and cooling. The beaker is rinsed several times with water until the filtrate is free of chloride, as tested by adding 0.1 mol l<sup>-1</sup> silver nitrate. The filter paper with the collected residue is put in an extraction thimble and dried in a small beaker at 100°C for 6–18 h. The dried extraction thimble is placed in a Soxhlet extractor. The digestion beaker and drying beaker are washed with 50 ml portions of petroleum ether into the extraction thimble. The sample is refluxed for 4 h with gentle heat. The extraction solvent is siphoned into a preweighed Erlenmeyer flask. The solvent is subsequently removed by distillation on a steam bath, and the flask is dried at 100°C for up to 2 h. The amount of fat is determined by weighing.

**Tannins and pigments in cocoa** Tannins and coloring matters in defatted cocoa powder and related products can be extracted by shaking consecutively with 150 ml and 100 ml portions of acidified 82% ethanol (10 ml HCl and 432 ml ethanol, diluted to 500 ml with water), at 55°C. The solvent is separated by centrifugation. Ethanolic extracts are pooled and evaporated. Alternatively, tannins and catechins in acetone extract can be precipitated with cinchonine sulfate and measured gravimetrically. In this extraction method, 25 g of sample is extracted overnight with 40% acetone (200 ml, cold). The acetone extract is filtered through a Whatman No. 41 paper. The residue is washed with 20 ml of 40% acetone. Water is added to make up to 250 ml. Saturated cinchonine sulfate (150 ml) is added to 25 ml of tannin-containing filtrate. The precipitate formed is filtered off and dried. The dried weight is multiplied by 0.534 to obtain the tannin content.

**Catechins, polyphenols, and tannins in tea** Catechins and related flavanols are water soluble, and can be extracted by boiling with water, or by shaking with aqueous acetone.

**Volatile and aromatic components** Volatile and aromatic components in coffee are concentrated by

passing purified helium gas directly over the sample (50 g of ground coffee) or the beverage (made from 2 g of coffee). The gas is aspirated through an adsorption tube (60 mm  $\times$  4 mm ID) containing 90 mg of a porous polymer (Tenax, 60/80 mesh, preconditioned for 24 h at 300°C in helium). Aromatic components are desorbed by heating and swept into the injection port of a gas chromatograph for the analysis of individual components.

Volatile and aromatic components of tea and tea infusion are also concentrated by aqueous distillation and the cold-trapping method. In this method, the sample is suspended in water in a 2 l round-bottomed flask connected to a rotary evaporator. The condenser is connected to cold traps containing freezing mixture (–15°C to –18°C) or dry ice–acetone mixture (–78°C). The aqueous mixture is distilled twice at 70°C, 25–30 mmHg. The distillate is saturated with sodium chloride and extracted with ether. The ether extract is dried over anhydrous sodium sulfate and concentrated by distillation of the solvent at 40°C. Components in the ether extract are analyzed by gas chromatography.

## Analytical Procedures

### Coffee

**Caffeine** The analyses of methylxanthines in coffee and tea are conducted: (1) by measuring the total nitrogen in extracts, either derived from the Bailey–Andrew extraction in the case of caffeine or by the Wadsworth method in the case of theobromine; (2) by spectrophotometric methods; and (3) by chromatographic methods, including thin-layer chromatography (TLC), LC, and gas chromatography (GC). The major advantage of chromatographic methods over spectrophotometric methods is the ability of the former to separate and quantify several methylxanthines (caffeine, theophylline, and theobromine) simultaneously.

*Spectrophotometric methods for the analysis of caffeine* Caffeine is quantified spectrophotometrically by measuring the ultraviolet (UV) absorbance at wavelengths between 270 and 280 nm. The major drawback of such methods is the presence of interfering substances. Such interferences can be reduced, but not eliminated, by using derivative techniques. Sample cleanup by various types of columns and background correction procedures is used to purify caffeine before spectrophotometric measurement. For example, in the Levine method, sample cleanup is achieved by a series of acidic and basic Celite columns. This method is adopted by AOAC for the analysis of

caffeine in decaffeinated coffees. Other methods using alumina columns (Kum–Tatt method), alkaline extraction, and combinations of various types of columns have also been used in sample cleanup.

*Thin-layer chromatography* Methylxanthines analysis by TLC has the advantages of high sample throughput and low cost. The major disadvantage of TLC methodology is its nonquantitative nature. Separation is achieved by chromatography on silica gel plates, using a variety of organic solvent mixtures as eluting solvents. Some typical solvent mixtures are: chloroform–methanol (9:1; v/v); benzene–acetone (3:7; v/v); chloroform–carbon tetrachloride–methanol (8:5:1; v/v/v); chloroform–ethanol–formic acid (88:10:2; v/v/v). Separation of methylxanthines has also been achieved using paper chromatography or with cellulose plate using a butanol–hydrochloric acid–water (100:11:28; v/v/v) mixture.

*Liquid chromatography* Analysis of methylxanthines by LC is carried out with C<sub>18</sub> reversed-phase columns, with acidic buffer mobile phases, and monitoring by UV absorption at ~270–280 nm, or at 254 nm. The LC procedure adopted by the AOAC for the analysis of caffeine utilizes a  $\mu$ Bondapak (Waters) C<sub>18</sub> column, with a mobile phase of water–methanol–acetic acid (74:25:1; v/v/v) and detection at 280 nm. Other solvent systems, such as water–acetonitrile–acetic acid (81:18:1; v/v/v) and phosphoric acid (7 mmol l<sup>–1</sup>)–acetonitrile–methanol (92:4:4; v/v/v), have also been used successfully in separating methylxanthines in foods and beverages. LC methods using normal-phase columns, such as Lichrosorb Si-60, with mobile phase composition of chloroform–2-propanol–acetic acid (96:2:2 or 92:7:1; v/v/v) and detection at 275–280 nm have also been applied to the analysis of methylxanthines, but are less robust. Quantification of methylxanthines is achieved by using either external standards or by using xanthine derivatives as internal standards; examples are 8-chlorotheophylline and 8-hydroxy-ethyltheophylline.

*Gas chromatography* The AOAC method for the analysis of caffeine by GC uses a packed glass column, 6 ft (~180 cm) in length by 4 mm ID, packed with 10% DC-200 oil on 80–100 mesh Gas Chrom Q, using nitrogen carrier gas and a thermionic potassium chloride detector. The injector temperature is 220°C, the column temperature is 190°C, and the detector temperature is 220°C. A wide variety of column-packing materials and detection modes can be used. For example, separation of caffeine, theophylline, and theobromine can be achieved with

3% OV-17 on Gas Chrom Q (100–120 mesh) by isothermal or temperature programming, with electron-capture detector, flame ionization detector (FID), or nitrogen–phosphorus detector (NPD).

### Chlorogenic acids

**Spectrophotometric analysis** Spectrophotometric determination of total chlorogenic acid content in green coffee extract is conducted by measuring the absorbance at 324 nm, as in the AOAC standard procedures. Modification of this method by purification and extraction of the sample has been applied to eliminate interferences in roasted coffee. The formation of colored complexes with borates, molybdates, and periodates enables more accurate determination of specific classes of chlorogenic acids.

**Gas chromatography** The complete separation and analysis of chlorogenic acids by capillary GC requires derivatization of the phenolic or carboxylic acid functions by silylating agents, such as hexamethyldisilazane or N,O-bis(trimethylsilyl)acetamide, or a mixture of both.

**Liquid chromatography** Analyses of chlorogenic acids are carried out with reversed-phase LC systems, both in gradient and isocratic modes. The typical analytical system employs C<sub>18</sub> columns of 5 µm packing, with a mobile-phase gradient of 25–47.5% methanol in pH 2.5 citrate buffer (0.025 mol l<sup>-1</sup>), or isocratic elution with 30% methanol in pH 2.5 citrate buffer (0.025 mol l<sup>-1</sup>), with UV detection at 310–330 nm.

**Volatile and aromatic components** Separation of volatile components is achieved on either fused silica capillary columns or packed columns. Individual volatile components are detected with a FID and identified by the use of reference standards. Methods using specific detectors, such as the NPD, sulfur-specific flame photometric detector, and mass-selective detector (MSD) have also been used. The MSD has the additional advantage of providing structural identification of the individual components.

### Cocoa

**Theobromine and caffeine** Separation and quantitative analyses of theobromine and caffeine are carried out using C<sub>18</sub> reversed-phase columns, with monitoring by UV absorption at 280 nm. In the AOAC method, a Waters LC system is used, with a µBondapak C<sub>18</sub> column, 30 cm length × 4 mm ID and 10 µm packing. The mobile phase is methanol–acetic acid–water (20:1:79; v/v/v) at 1 ml min<sup>-1</sup> with

a nominal back-pressure of 2000 psi (~14 000 kPa). Concentrations of theobromine and caffeine in cocoa extracts are determined from a calibration graph, using the external standard method.

**Cocoa fat and lipids** In AOAC methodologies, the composition of fat and fatty acids in cocoa fat extracts are analyzed and characterized by physicochemical methods, including refractive index, melting point, saponification value and unsaponifiable material, iodine number, Reichert–Meissl and Polenske values.

**Volatile and aromatic components** Cocoa volatile and aromatic components are analyzed by capillary GC, with samples derived from headspace enrichment or cold-trapping techniques. Analysis is similar to that of coffee volatiles; FIDs are used and identification is by the use of reference compounds.

### Tea

**Methylxanthines: caffeine and theophylline** The amount of caffeine in tea is determined by spectrophotometric and chromatographic methods, as described above for coffee. Separation of caffeine from theophylline is achieved by chromatographic methods.

**Polyphenols: catechins and flavanols** Flavanols and tea polyphenols are commonly analyzed using paper chromatography, TLC, and LC. GC has also been used, but derivatization of the hydroxyl groups with silylation is required to increase the volatility of these analytes.

**Paper chromatography and thin-layer chromatography** Tea flavanols can be separated by two-dimensional paper chromatography, using water as the first dimension solvent, and a mixture of butanol–acetic acid–water (4:1:5; v/v/v) as the second dimension. A similar procedure can be carried out using TLC with cellulose plates. Different flavanol components are visualized by fluorescence under UV radiation, either before or after exposure to fuming ammonia. Other visualizing agents are bis-diazotized benzidine and vanillin.

**Liquid chromatography** Separation and quantitative analysis of flavanols and polyphenols may be conducted by C<sub>18</sub> reversed-phase LC methods, using either isocratic or gradient modes. In one example of the isocratic separation method, the solvent system used consists of acetic acid–methanol–dimethylformamide–water (1:2:40:157; v/v/v/v). Eluting peaks are monitored by UV absorption at 254 nm.

*Gas chromatography* The utility of GC for the analysis of catechins and polyphenols is quite limited, given the polar and nonvolatile nature of these compounds. Derivatization of a polyphenol extract with silylating agent and subsequent separation of the trimethylsilyl derivatives of flavanols can be achieved by chromatographing on a glass column 5 ft (~150 cm) in length  $\times$  4 mm ID packed with 3% OV-1 on diatomite.

**Tannins, theaflavins, and thearubigens** Tannins and theaflavins are pigmented products in black tea formed after oxidation and condensation of green tea flavanols during the fermentation process. These compounds are analyzed by spectrophotometric, colorimetric, and chromatographic methods, using LC and GC.

*Spectrophotometric and colorimetric methods* Spectrophotometric analysis of theaflavins is carried out at 380 and 460 nm on isobutyl methyl ketone or ethyl acetate extracts of black tea infusions. Theaflavins are also determined by using Flavognost reagent (diphenyl boric acid ethanolaamine), which forms a green chromophore by reacting with the benzotropolone nucleus of theaflavins. The color developed is measured at 600 nm.

*Liquid chromatographic methods* Separation and quantitative analysis of theaflavins and thearubigens has also been achieved by gel permeation chromatography on Sephadex LH-20, monitoring the eluting peaks at 380 nm. Reversed-phase LC separation has also been conducted using a C<sub>18</sub> column, with a mobile phase consisting of 29% aqueous acetone and 1% acetic acid.

## **Analysis of Coffee, Cocoa, and Tea for Quality Assurance Purposes**

### **Coffee**

The flavor and aroma qualities of the common coffee beverages are dependent on the source of the coffee bean used, soil and climatic conditions, as well as the duration and temperature of roasting. The determination of the quality of roasted coffee is mainly by tasting of the brew by professional tasters. Such processes are qualitative only. Chemical analysis of coffee quality is based on headspace analysis of the aroma components of roasted or brewed coffee. For instance, drastic decreases in methylfuran and methylethylketone in coffee aroma have been observed in roasted and ground coffee within a few days, which may be correlated to coffee staleness. A

decrease in the ratio of 2-methylfuran to 2-butanone in coffee aroma with time has also been observed. However, the use of such ratios of individual aroma components as measures of coffee quality has not been officially established.

### **Cocoa**

Quality evaluation of cocoa bean and cocoa powder is by visual inspection for contamination, moldiness, and by aroma/flavor and tasting. Physical analysis of cocoa bean and cocoa powder includes analysis for total moisture (<8%) and fat (<55%). Additionally, the quality of cocoa is characterized by the iodine number (degree of unsaturation of the fatty acid components), unsaponifiable matter, and GC analysis (for volatile and aroma components).

### **Tea**

Quality assessment of tea is conducted by evaluation of the color and appearance of the tea leaves and the tea infusion formed, as well as by GC analysis of the aroma of both the tea leaves and the infusion. For example, (Z)-3-hexenyl hexanoate is a major contributor to the aroma of fresh green tea. This component is found to decrease with storage. The astringent taste of tea is mainly due to the presence of unoxidized flavanols and polyphenolics (e.g., catechins). Gradual and continued oxidation of these components with storage time would result in changes in both color and taste of the product. The brothy taste of green tea is due to the presence of amino acids, notably L-threonine. A decrease in the amino acid contents during storage would also result in deterioration of quality.

## **Limitations of Current Technology**

Despite the fact that major components in coffee, cocoa, and tea can be analyzed and measured by the methodologies outlined above, there are many minor components in these beverages that have not been extensively studied. For example, roasting of coffee induces caramelization, condensation, pyrolysis, and Maillard reactions; the nature and composition of these reaction products (collectively termed melanoidins) are dependent on the temperature and duration of roasting. Although these reaction products in roasted coffee constitute as much as 30% of dry weight, their chemical identities are largely unknown. The relationship of melanoidin content to quality, taste, and flavor of the product is not well understood and analytical procedures for determining these components are lacking. The manufacturing processes of black tea involve extensive oxidation and polymerization of

polyphenols and results in the formation of the aflavins and thearubigens. The formation of these condensation products is dependent on the duration of oxidation during fermentation. Analytical methods for the characterization of aroma components remain to be established and the relationship between individual components and product quality has yet to be elucidated.

*See also:* **Derivatization of Analytes.** **Extraction:** Solvent Extraction Principles. **Food and Nutritional Analysis:** Contaminants; Oils and Fats. **Gas Chromatography:** Detectors. **Lipids:** Fatty Acids. **Liquid Chromatography:** Food Applications. **Sample Handling:** Comminution of Samples. **Sensory Evaluation.** **Spectrophotometry:** Organic Compounds.

## Further Reading

AOAC (1990) *Official Methods of Analysis*, 15th edn., vols. I and II. Arlington, VA: Association of Official Analytical Chemists, Inc.

- Chatt EM (1953) *Cocoa: Cultivation, Processing, Analysis*. New York: Interscience.
- Charalambous G (1978) *Analysis of Foods and Beverages, Headspace Techniques*. London: Academic Press.
- Charalambous G (1986) *Handbook of Food and Beverage Stability: Chemical, Biochemical, Microbiological, and Nutritional Aspects*. London: Academic Press.
- Charalambous G and Inglett G (1983) *Instrumental Analysis of Foods, Recent Progress*, vol. 2. London: Academic Press.
- Clarke RJ and Macrae R (1985) *Coffee, vol. 1, Chemistry*. London: Elsevier Applied Science.
- Sivetz M (1963) *Coffee Processing Technology*. Westport: AVI Publishing Co.
- Sivetz M and Desrosier NW (1979) *Coffee Technology*. Westport: AVI Publishing Co.
- Spiller GA (1984) *The Methylxanthine Beverages and Foods: Chemistry, Consumption, and Health Effects*. New York: Alan R. Liss.
- Willson KC and Clifford MN (1992) *Tea: Cultivation to Consumption*. London: Chapman & Hall.

## Alcoholic Beverages

**J F Cacho and R Lopez**, University of Zaragoza, Zaragoza, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Xuejun Zhang, pp. 1504–1512, © 1995, Elsevier Ltd.

## Introduction

Alcoholic beverages comprise a large group of beverages that contain varying amounts of alcohol (ethanol). Alcoholic beverages produced on an industrial scale include beer, wine, and China rice wine, and distilled spirits such as brandy, whisky, rum, gin, cognac, vodka, tequila, pisco, and China distilled spirit. The components of alcoholic beverages are highly complex and over 1300 compounds have been identified in various beverages.

The constituents of each alcoholic beverage can be divided into major, minor, or trace components. The major constituents usually consist of ethanol and water. The minor or trace constituents are fusel alcohols, organic acids, carbonyl compounds, esters, aldehydes, lactones, sulfur compounds, sugars, preservatives, and colorants. There are two general approaches to analyzing the components of alcoholic beverages. The first, and most usual, is by using

chemical and physiochemical analysis. The other is to use sensory evaluation. In recent years, instrumental analytical methods have become an important tool for analyzing minor and trace constituents of alcoholic beverages. The most widely used methods are ultraviolet (UV)–visible spectrophotometry, gas chromatography (GC), liquid chromatography (LC), mass spectrometry (MS), and paper chromatography, and thin-layer chromatography.

This article will review the methods applied for the determination of major and minor components in alcoholic beverages.

## Sensory Assessment

Sensory assessment is a scientific discipline used to evoke, measure, analyze, and interpret responses to those properties of a substance (food or beverage) that are perceived by the five human senses: sight, smell, taste, touch, and hearing. Even though individual judgments are subjective, the techniques of sensory evaluation use objective scientific testing methods principally based on working with a panel of assessors. Two types of analytical objective tests can be observed: (1) The discriminative tests: is there a difference between several products? (2) The

polyphenols and results in the formation of the aflavins and thearubigens. The formation of these condensation products is dependent on the duration of oxidation during fermentation. Analytical methods for the characterization of aroma components remain to be established and the relationship between individual components and product quality has yet to be elucidated.

*See also:* **Derivatization of Analytes.** **Extraction:** Solvent Extraction Principles. **Food and Nutritional Analysis:** Contaminants; Oils and Fats. **Gas Chromatography:** Detectors. **Lipids:** Fatty Acids. **Liquid Chromatography:** Food Applications. **Sample Handling:** Comminution of Samples. **Sensory Evaluation.** **Spectrophotometry:** Organic Compounds.

## Further Reading

AOAC (1990) *Official Methods of Analysis*, 15th edn., vols. I and II. Arlington, VA: Association of Official Analytical Chemists, Inc.

- Chatt EM (1953) *Cocoa: Cultivation, Processing, Analysis*. New York: Interscience.
- Charalambous G (1978) *Analysis of Foods and Beverages, Headspace Techniques*. London: Academic Press.
- Charalambous G (1986) *Handbook of Food and Beverage Stability: Chemical, Biochemical, Microbiological, and Nutritional Aspects*. London: Academic Press.
- Charalambous G and Inglett G (1983) *Instrumental Analysis of Foods, Recent Progress*, vol. 2. London: Academic Press.
- Clarke RJ and Macrae R (1985) *Coffee, vol. 1, Chemistry*. London: Elsevier Applied Science.
- Sivetz M (1963) *Coffee Processing Technology*. Westport: AVI Publishing Co.
- Sivetz M and Desrosier NW (1979) *Coffee Technology*. Westport: AVI Publishing Co.
- Spiller GA (1984) *The Methylxanthine Beverages and Foods: Chemistry, Consumption, and Health Effects*. New York: Alan R. Liss.
- Willson KC and Clifford MN (1992) *Tea: Cultivation to Consumption*. London: Chapman & Hall.

## Alcoholic Beverages

**J F Cacho and R Lopez**, University of Zaragoza, Zaragoza, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Xuejun Zhang, pp. 1504–1512, © 1995, Elsevier Ltd.

## Introduction

Alcoholic beverages comprise a large group of beverages that contain varying amounts of alcohol (ethanol). Alcoholic beverages produced on an industrial scale include beer, wine, and China rice wine, and distilled spirits such as brandy, whisky, rum, gin, cognac, vodka, tequila, pisco, and China distilled spirit. The components of alcoholic beverages are highly complex and over 1300 compounds have been identified in various beverages.

The constituents of each alcoholic beverage can be divided into major, minor, or trace components. The major constituents usually consist of ethanol and water. The minor or trace constituents are fusel alcohols, organic acids, carbonyl compounds, esters, aldehydes, lactones, sulfur compounds, sugars, preservatives, and colorants. There are two general approaches to analyzing the components of alcoholic beverages. The first, and most usual, is by using

chemical and physiochemical analysis. The other is to use sensory evaluation. In recent years, instrumental analytical methods have become an important tool for analyzing minor and trace constituents of alcoholic beverages. The most widely used methods are ultraviolet (UV)–visible spectrophotometry, gas chromatography (GC), liquid chromatography (LC), mass spectrometry (MS), and paper chromatography, and thin-layer chromatography.

This article will review the methods applied for the determination of major and minor components in alcoholic beverages.

## Sensory Assessment

Sensory assessment is a scientific discipline used to evoke, measure, analyze, and interpret responses to those properties of a substance (food or beverage) that are perceived by the five human senses: sight, smell, taste, touch, and hearing. Even though individual judgments are subjective, the techniques of sensory evaluation use objective scientific testing methods principally based on working with a panel of assessors. Two types of analytical objective tests can be observed: (1) The discriminative tests: is there a difference between several products? (2) The



descriptive tests: what is the difference and how big is it?

The sensory assessment of alcoholic beverage, such as wine or beer, has, in general, three basic steps (except the decision on rating method):

1. Visual inspection: Checking if the container is bottle full and examination for transparency or sediment. After pouring the liquid into the glass (for wine there exist ISO standardized glasses), an additional appearance inspection follows. The sample is oriented against a bright background. In the case of wine, the sample is viewed at an angle of 30–45°; in the case of beer it is agitated to get a generous head. The appearance is scored later. One examines absence of haze, color shade and depth, viscosity, effervescence, or head retention. Visual inspection may seriously affect the subsequent evaluation steps, so one needs to be careful not to prejudge the beverage based on a visual inspection of the bottle.

2. Smell: There are various modes of odor evaluation. In the case of wine, it is recommended not to swirl first, as the most volatile compounds are perceived in this way. The nature and intensity of odors is recorded according to the rating system. Successive swirling should promote release of the heavier polar compounds that are retained in the hydroalcoholic solution more than in the apolar ones. Some tasters hold the glass between the hands when the beverage is too cold. The impressions are recorded yet again.

3. In-mouth sensations: The most complex task is the evaluation of simultaneously appearing sensations while tasting the beverage in the mouth. After coating the inside of the mouth with the liquid, one should score: (1) Taste and mouth-feel (body) – presence of taste sensations (sweet, acid, and bitter taste: their duration and changes in quality and intensity) and tactile sensations (astringency, prickling, body, temperature, and heat). (2) Odor – rating of beverage odors at the warmer temperatures of mouth, which increases volatility of odorants, before swallowing. Mainly the differences between the in-mouth and in-glass odor features are observed. (3) Aftersmell – rating of odor sensations after swallowing the liquid and expiring through the nose. The final stage of wine tasting is termed finish. It is the final impression from the wine (good or bad), may include odor-related, hedonic (e.g., flourish), or time-related terms (e.g., short finish). Typically, the longer the finish the more highly rated the wine.

Sensory richness of alcoholic beverages, especially of wine, promoted an intense effort to unveil the

chemical background behind these percepts. As the impression derived from sensory evaluation is mainly governed by olfaction, the area of volatile compounds was the most investigated. Especially the performance and wide accessibility of GC and MS allowed identifying many volatiles, which could potentially contribute to the sensations experienced during wine or beer tasting. However, it was the introduction of GC–olfactometry in the 1970s that radically advanced the discovery of the real key compounds.

The establishment of odor units (a ratio of actual concentration of odorant to its threshold concentration) was also beneficial, even if in reality more emphasis is given to intensity ratings.

The most useful flavor descriptive procedure has so far turned out to be the development of flavor vocabulary precisely defined for each product. It is important to find sufficiently refined terms for the characteristic sensory properties of the specific alcoholic beverage. There are various intensity rating strategies in the sensory assessment practice. Perhaps the most widely used are the aroma wheels. The first was M.C. Mailgaard's 'Beer Flavor Wheel', jointly adopted in the 1970s by the European Brewery Convention, the American Society of Brewing Chemists, and the Master Brewers Association of America. A similar attempt was made in the 1980s by A. Nobel's 'Wine Aroma Wheel'. These systems provide standardized flavor descriptors and recipes for their reconstruction. Various methods were proposed for flavor profiling, among which the Quantitative Descriptive Analysis (QDA<sup>TM</sup>) and Free Choice Profiling are prominent. The reliability of the results must be tested statistically even when using threshold values or relative values like odor unit values. Statistics are also needed in the different profile techniques, both to examine how judges employ the methods and to develop the terminology.

Endeavors have been made to find a link between two data sets (sensory versus instrumental data). The common goal of these tools is to discover the components or parameters whose variation explains the variation of sensory characteristics. The most useful statistical methods used for such purpose are partial least squares regression and generalized procrustes analysis. From a practical point of view, the models can be used to complement sensory assessment in routine quality control or in product and process development work. Regression-based statistical techniques are often used in conjunction with GC to distinguish well-known brands of alcoholic beverages from less expensive ones to detect counterfeit products.

## Proof and Extract

The proof of an alcoholic beverage is a measurement of its ethanol content. The proof of spirit and wine is defined as the percentage by volume of ethanol. The traditional method for determination of proof uses a pycnometer or a small, accurately graduated hydrometer. Recently, the European Community has adopted a reference method for ethanol determination in spirits, based on two steps. The first step is a prescribed distillation stage and the second step allows a choice from three different methods of measuring the density of the distillate: (1) pycnometry: time consuming and susceptible to error; (2) electronic densimetry: measurement of the oscillations of a vibrating U-tube, this method is simple and fast; (3) densimetry using hydrostatic balance: this method is also fast with the new generation of instrumentation.

The extract of an alcoholic beverage is a measurement of its nonvolatile compounds. The extract of a wine is the most important among the alcoholic beverages. Wine extract is composed of sugars, fixed acids, organic salts, and other substances. Its value is a measure of wine quality. The traditional method of determining wine extract is drying at 100°C. The method has the disadvantages of being time consuming and limited to wines having an extract not greater than 6 g per 100 ml. It is being used less frequently today.

Of the two AOAC methods (Official Methods of Analysis of the Association of Official Analytical Chemists) to determine wine extract, the more convenient one relates the extract of wine to the specific gravity (relative density) of the dealcoholized wine. In spirits, the reference method adopted by EU legislation is gravimetry, involving weighing the residue left by evaporation of the spirit on a boiling water bath and drying in a drying oven. The method is not suitable for liqueurs because of their high sugar content.

Some recent and convenient procedures have been used to determine both ethanol content and extract. Regarding ethanol, an AOAC-approved method exists for the determination of ethanol in wine based on GC and the technique is very precise and accurate. An increasingly popular technique is the use of near-infrared spectroscopy. Instruments are commercially available allowing the determination of ethanol and other common parameters in wine.

## Determination of Alcohols

Alcohols in alcoholic beverages mainly consist of aliphatic monohydric alcohols and di- and trihydric alcohols. Methanol and fusel oil in alcohols are

determined but ethanol is determined as a proof in a routine analysis of alcoholic beverages.

Methanol is ubiquitous and is usually a minor component in alcoholic beverages, but some distilled alcoholic beverages produced from potato and stone fruits, such as cherries and plums, are exceptions and may contain rather large amounts. Large amounts of methanol have also been determined in some fortified wine, China distilled spirit, distilled wine, and brandy.

Higher alcohols from propanol up to 3-methyl-1-butanol (isoamyl alcohol), frequently encountered in so-called fusel oil fractions, are considered to be an important aroma factor, and for this reason they have been determined in a large number of beers, wines, and distilled alcoholic beverages. Because higher alcohols are believed to have a marked effect on the volatile flavor of distilled beverages, numerous papers have been published describing both the relative and absolute compositions of fusel oil fractions.

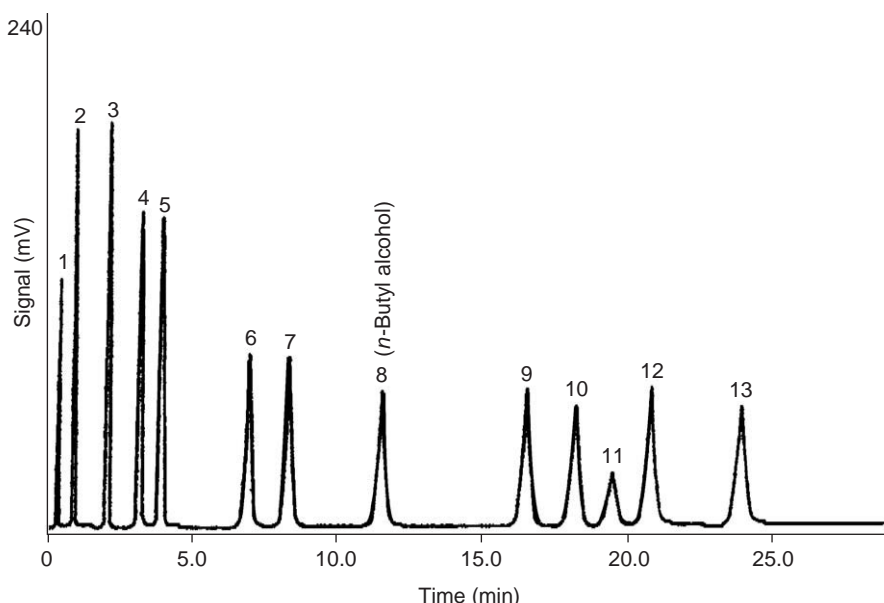
Generally, colorimetric methods and GC are used for the determination of methanol and fusel oil in distilled alcoholic beverages and spirits.

According to the procedure presented in the AOAC method, methanol is oxidized to formaldehyde by a potassium permanganate solution. The aldehyde formed is allowed to react with chromotropic acid in strongly acidic solution to form a colored product. Its absorbance is measured spectrophotometrically at 575 nm.

The colorimetric methods recommended for the determination of total fusel oil contents in beverages are based on a color reaction. In this reaction, 2-methyl-1-propanol, 2-methyl-1-butanol, and 3-methyl-1-butanol lose water during heating in a strongly acidic solution and the unsaturated hydrocarbons formed yield colored complexes with vanillin, salicylaldehyde, and 4-dimethyl-aminobenzaldehyde. The absorbance is measured spectrophotometrically at 445 and 560 nm.

GC provides a means of evaluating small amounts of methanol in beer, wine, and distilled alcoholic beverages. It is the most convenient method to determine single fusel oil and methanol in beverages simultaneously.

Although, direct injection with GC-flame ionization detection (FID) is sensitive enough to quantify methanol and fusel oils, it is advisable only for spirits because of the dirt accumulated in the injection port when beer, wine, or liqueurs are analyzed. Thus, liquid-liquid extraction has been widely applied to separate and concentrate fusel oil from beer, wine, and distilled alcoholic beverages, followed by GC determination of the compounds. Nowadays, many simpler and online extraction techniques are applied



**Figure 1** Chromatogram showing baseline resolution of the 13 alcohols in the order listed in **Table 1**. Chromatographic conditions: 6 ft ( $\sim 180$  cm)  $\times$  2 mm ID glass column packed with 0.2% Carbowax 1500 on Carbowax C (80–100 mesh); column temperature held at 55°C for 4 min, then programmed to 95°C at 2°C min<sup>-1</sup>.

in alcohol determination such as liquid microextraction, solid-phase extraction (SPE), and solid-phase microextraction (SPME).

The direct GC method is the simplest and fastest for simultaneous determination of methanol, acetaldehyde, and fusel oil. The efficiency and resolution of GC has improved with the development of more efficient supports and liquid phases. Liquid phases, such as Carbowax and combinations of two liquid phases, have been most useful for determining fusel oil. **Figure 1** shows the results obtained by using a single direct injection procedure for separating and quantifying methanol and fusel oil whisky. Baseline resolution was obtained for each of the 13 alcohols (**Table 1**) in 25 min. 1-Butanol was chosen as the internal standard with a relative retention time of 1.0.

Alcohols are also analyzed by LC; in most cases the choice of the column falls on an ionic-exchange resin of the styrene-divinyl benzene polymer type, to which has been linked a functional group of the sulfonic type in the form H<sup>+</sup> or with another suitable cation. Since the separation of alcohols is performed very often together with other analytes, such as organic acids and sugars, the choice of the type of detector also takes into account their chemical-physical properties. Usually, the choice is the UV detector, the refractive index detector, or the electrochemical detector (EC). The use of LC allows the determination of less volatile alcohols such as glycerol in wine and beer, which could be difficult to analyze by GC.

**Table 1** Gas chromatographic retention times (RRTs) of alcohols (relative to 1-butanol)

IUPAC name	Common name	RRT
Methanol	Methyl alcohol	0.03
Ethanol	Ethyl alcohol	0.08
2-Propanol	Isopropyl alcohol	0.18
1-Propanol	<i>n</i> -Propyl alcohol	0.26
2-Methyl-2-propanol	<i>t</i> -Butyl alcohol	0.34
2-Butanol	<i>s</i> -Butyl alcohol	0.60
2-Methyl-1-propanol	Isobutyl alcohol	0.72
1-Butanol	<i>n</i> -Butyl alcohol	1.00
3-Pentanol		1.43
2-Pentanol	<i>s</i> -Amyl alcohol	1.58
2-Methyl-1-butanol	Active amyl alcohol	1.69
3-Methyl-1-butanol	Isoamyl alcohol	1.81
1-Pentanol	<i>n</i> -Amyl alcohol	2.08

(Reproduced with permission from Wilson LA, Ding JH, and Woods AE (1991) Gas-chromatographic determination and pattern-recognition analysis of methanol and fusel oil concentrations in whiskeys. *Journal of the Association of Official Analytical Chemists* 74: 248–256.)

## Aliphatic Carbonyl Compounds

Aliphatic carbonyl compounds consist mainly of aldehydes and ketones. A large number of aliphatic aldehydes have been shown to be present in alcoholic beverages. Among the aldehydes determined quantitatively acetaldehyde is the major component, and generally constitutes more than 90% of the total aldehyde content of the beverage. For this reason it could be expected that acetaldehyde is of importance

to the aroma and is often associated with improved quality.

Ketones form a minor fraction in alcoholic beverages. Ketones can be said to be potential aroma compounds. In particular, diketones such as 2,3-butanodione and 2,3-pentanodione are of great importance to the aroma of alcoholic beverages because of their low sensory threshold values. Those aldehydes with 8–10 carbon atoms, such as (*E*)-2-nonenal, octanal, nonanal, decanal, or (*E,Z*)-nonadienal, are also strong odorants, related with important off-flavors.

Titrimetric and colorimetric methods are available for the determination of aldehyde contents in alcoholic beverages. An iodimetric method has been used for the determination of the total aldehyde content of wines. A modified iodimetric titration method is included in the methods of AOAC. In order to avoid the reactions of interfering substances, aldehydes are separated by distilling an aliquot of wine into a neutral hydrogensulfite solution. After the excess of hydrogensulfite has been titrated with an iodine solution at pH 2, hydrogensulfite is released from its aldehyde addition compound at pH 9 and titrated with iodine solution.

Different colorimetric methods have been developed for the determination of aldehyde and diketone contents in alcoholic beverages. Fuchsin, piperidine, 3-methylbenzothiazole-2-one hydrazone, 2-naphthol, and creatine have generally been used as reagents to produce colored compounds, the absorbances of which are measured at a wavelength between 630 and 660 nm or 500 and 550 nm.

As has been previously said, 2,3-butanodione (diacetyl) is an important aroma of alcoholic beverages, it has not been studied and measured extensively in the past because of analytical difficulties in the quantitation caused by its highly volatile nature, chemical instability, and interference of other compounds. Colorimetric methods to measure diacetyl have been widely used in the past. These methods involve steam distillation to isolate diacetyl from the matrix. However, distillation has the disadvantage of incomplete isolation of diacetyl from other closely related compounds that will result in an overestimation of its concentration. A fluorometric method was developed to improve upon the lengthy distillation methods that involve derivatization. Although acetaldehyde and its acetal can be determined by direct injection GC-FID in spirit drinks (EU reference method for spirits), most chromatographic methods for minor aldehydes implicate also derivatization. While a very sensitive and accurate method based on SMPE without derivatization and MS detection has been developed, it requires the use of

deuterated diacetyl- $d_6$  as an internal standard, not easily available.

Due to the poor chromatographic and MS properties of higher aldehydes and ketones, most analytical methods are based on GC-MS or GC-ECD analysis of chemical derivatives. An added analytical problem is the high content of major carbonyl compounds, such as acetaldehyde, pyruvic acid, acetone, or diacetyl compared with C8-C10 aldehydes. Most of the methods previously developed for the analysis of carbonyl components in wine are based on the GC-MS or GC-ECD analysis of oximes formed by reaction with *O*-(2,3,4,5,6-pentafluorobenzyl)hydroxylamine hydrochloride, although other derivatization reagents have been explored such as 2,4-dinitrophenyl-hydrazine or cysteamine. On the other hand, the analysis of (*E*)-2-nonenal in wine, or of some higher methyl ketones in cognac, has been carried out by performing the chemical reaction over an organic extract obtained from the wine or cognac. All these methods use tedious liquid-liquid extractions. Nevertheless, several authors have shown the potential of direct derivatization of carbonyls on a solid phase, particularly in SPME- and SPE-based strategies.

## Organic Acids

Organic acids in alcoholic beverages can be classified into volatile and fixed acids and may be of great importance. It has been found that some organic acids significantly influence the odor and taste of alcoholic beverages. Any substantial increase in the volatile acidity in wines seems to be a consequence of the activity of spoilage microorganisms. For this reason statutory restrictions exist in many countries for the maximum amounts of the volatile acids in wine.

Organic acids consist of aliphatic monobasic carboxylic acids from formic acid up to the  $C_{18}$  acids; aliphatic monobasic hydroxyl acids such as lactic acid; aldonic and uronic acids such as gluconic, glucuronic, and galacturonic acids; aliphatic monobasic oxo acids, mainly including succinic, malic, citric, and tartaric acids.

Distilled and undistilled beverages contain acetic acid in abundance. Frequently it amounts to 40–95% of the total volatile acids. Although the greatest part of the volatile acidity in wine consists of acetic acid, the remaining acids are nonetheless important aroma compounds. Many short-chain and long-chain acids, and in particular the branched-chain acids, have an appreciably strong odor.

Methods employed for the determination of organic acids in alcoholic beverages include

titrimetric, colorimetric, enzymatic, paper chromatographic, thin-layer chromatographic, gas chromatographic, and liquid chromatographic procedures, as well as GC combined with MS. Titrimetric methods are mainly employed to determine the total volatile acids from formic acid up to C<sub>18</sub> acids. In the procedure, the volatile acids are isolated from alcoholic beverages by steam distillation followed by titration with a strong base. In order to concentrate the volatile acids methods such as microdiffusion, reduced pressure, or vacuum steam distillation can be used.

Titrimetric methods have also commonly been proposed for the determination of fixed and total acids in wines. After treating the wine with a cation-exchange resin, oxalic acid can be precipitated as its calcium salt and titrated with potassium permanganate. For determination of tartaric acid in wine the acid can be precipitated as calcium tartrate or potassium hydrogentartrate.

Colorimetric methods are used for the determination of lactic and tartaric acids in wines. In the determination of lactic acid the procedure consists of the conversion of lactic acid into acetaldehyde by heating in the presence of sulfuric acid or by oxidation with cerium(IV) sulfate and subsequent formation of a colored compound with *p*-hydroxydiphenyl or piperidine and sodium nitroprusside. The red color formed is measured at 560–570 nm.

In the determination of tartaric acid in wines, the tartaric acid produces a colored complex with ammonium metavanadate, which is measured at 530 nm.

Major acids in distilled alcohol beverages are usually analyzed by high-performance liquid chromatography (HPLC). Several methods based on LC have been developed allowing simultaneous determination in alcoholic beverages of citric, tartaric, malic, succinic, lactic, and acetic acids in a single run using cation-exchange or C18 columns and refractive index, UV, or electrochemical detectors. If a previous step for cleaning or concentration is required, SPE based on a strong anion-exchange is the most selective technique for the enrichment of the ionizable acid analytes. It is preferred instead of the conventional solvent extraction method, which is time consuming and uses relatively large volumes of solvents.

Methods for the determination of organic acids by GC consist of direct and derivative techniques. The methods of extraction can take advantage of the acid properties of the acids by using an anion-exchange resin, or they can use a more unspecific strategy carrying out the extraction with an apolar solvent or resin. In the latter case, the pH of the beverage must

be adjusted to 7 to assure complete recovery of the acids from the aqueous phase.

With modern capillary columns derivatization of low molecular weight acids is not compulsory and peaks without tailing can be obtained. In this case, it is important to keep a permanent check on the inertness of the whole GC system, because acids are very sensitive to active sites on the column. However, to analyze heavier and more polar acids by GC, derivatization must be applied. Derivative methods involve the isolation of the acids from the matrix and afterward the conversion of the acids into the more volatile benzyl or trimethylsilyl esters.

A recent alternative to HPLC or GC in the analysis of organic acids is capillary zone electrophoresis. It provides completely different selectivities from chromatographic methods, shorter analysis times, and no derivatization is required. As a drawback, the lack of UV absorbance of aliphatic organic acids above 220 nm requires the addition of substances showing strong UV absorption for indirect UV detection.

## Esters

The esters of aliphatic monocarboxylic acids are the most numerous of the neutral volatile aroma compounds in alcoholic beverages. Because of the relatively high volatility of the esters it has often been proposed that esters have a noticeable influence on the odor of a beverage.

Esters of aliphatic di- and tribasic carboxylic acids occur frequently as aroma components in alcoholic beverages. Because these compounds are mildly odored, they are apparently not very significant aroma factors. In addition to aliphatic fatty acid esters, several hydroxyl- and oxo-acid esters have been determined by GC/MS. Unlike the fatty acid esters, the esters of hydroxyl and oxo acids are slightly volatile compounds. However, they do not play an important role in the formation of the odor.

Titrimetric and colorimetric methods are two classical methods for determining the ester contents of alcoholic beverages. The titrimetric procedure involves the treatment of esters with an excess of base followed by determination of the quantity of base required to hydrolyze the esters present in the sample. The colorimetric procedure involves the esters reacting with hydroxylamine in an alkaline solution to form a hydroxamic acid, which forms a colored complex with iron(III).

It becomes clear that these chemical methods are too long and not specific, and are very rarely used today; instead, analysts prefer to use GC-based procedures. GC has many advantages over the classical

methods. The direct method of GC can give good reproducibility and enough sensitivity, especially when large volume injection can be performed. But regular direct injections of alcoholic beverages accumulate nonvolatile material in the injection port, leading to losses in chromatographic performance, appreciable after a few injections in the cases of wine and beer.

In most investigations, however, procedures for isolation and enrichment are required. Generally, liquid-liquid extraction with a high sample/solvent ratio is used prior to GC analysis. Among the solvents used are Freon 113, isooctane, ether, pentane, and dichloromethane. Enrichments of  $\sim 20$  times of the original sample are enough to quantify a few milligrams per liter of the esters using splitless injection and FID detection. Many other procedures have been widely applied including distillation and headspace analysis but currently SPE and SPME are the techniques of choice for the possibility of automation and because of the low volumes of solvents used.

Summarizing, volatile compounds with concentrations  $\sim 0.1 \text{ mg l}^{-1}$ , including major esters and higher alcohols acetates, can be determined after a single isolation step and GC-FID detection. Some minor esters, at concentrations between  $0.1 \mu\text{g l}^{-1}$  and  $0.1 \text{ mg l}^{-1}$  will require a powerful isolation-preconcentration step and further GC-MS analysis.

## Other Important Volatile Compounds

There is an important part of the aroma of beverages elicited by compounds that present very serious analytical problems. This is a heterogeneous group formed by compounds whose analysis is very difficult due to different reasons, such as bad chromatographic behavior and poor chemical stability, extremely low concentrations, or very poor analytical properties. Volatile sulfur compounds, alkyl methoxy-pyrazines, furaneol, sotolon, and some aromatic thiols are well-known examples of this group. In general, the analysis of these compounds requires the development of specific methods of isolation, or detection, or the use of chemical derivatives. Currently, a great effort is devoted to developing new analytical methods to characterize these flavor components, among them GC-olfactometry is a useful choice.

See also: **Ethanol. Food and Nutritional Analysis:** Overview; **Wine. Gas Chromatography:** Overview. **Indicators:** Complexometric, Adsorption, and Luminescence

**Indicators. Liquid Chromatography:** Overview. **Sensory Evaluation. Spectrophotometry:** Organic Compounds.

## Further Reading

- Amelio M (2000) HPLC analysis of alcohols in foods and beverages. In: Nollet LML (ed.) *Food Analysis by HPLC*, 2nd edn., pp. 303–320. New York: Dekker.
- AOAC International (1995) *Official Methods of Analysis of AOAC International*. Arlington, VA: AOAC International.
- Brereton P, Hasnip S, Bertrand S, Wittkowski R, and Guillou C (2003) Analytical methods for the determination of spirit drinks. *Trac – Trends in Analytical Chemistry* 22: 19–25.
- Brewers Publications (1993) *Evaluating Beer*. Boulder, CO: Brewers Publications.
- Commission Regulation (EC) No. 2870/2000 of 19 December 2000 laying down Community reference methods for the analysis of spirit drinks, N.L. *Official Journal of the European Communities*, No. L333, pp. 20–46.
- Demuyttenaere JCR, Martinez JIS, Verhe R, Sandra P, and De Kimpe N (2003) Analysis of volatiles of malt whisky by solid-phase microextraction and stir bar sorptive extraction. *Journal of Chromatography A* 985: 221–232.
- Ferreira V, Aznar M, and López R (2001) Olfactometry and aroma extract dilution analysis of wines. In: Jackson J, Linskens H, and Inman R (eds.) *Molecular Methods of Plant Analysis*, pp. 89–122. Heidelberg: Springer.
- Ferreira V, Cullere L, Lopez R, and Cacho J (2004) Determination of important odor-active aldehydes of wine through gas chromatography-mass spectrometry of their O-(2,3,4,5,6-pentafluorobenzyl)oximes formed directly in the solid phase extraction cartridge used for selective isolation. *Journal of Chromatography A* 1028: 339–345.
- Hayasaka Y and Bartowsky EJ (1999) Analysis of diacetyl in wine using solid-phase microextraction combined with gas-chromatography mass-spectrometry. *Journal of Agricultural and Food Chemistry* 47: 612–617.
- Jackson RS (2001) *Wine Tasting: a Professional Handbook*. San Diego, CA: Academic Press.
- Lawless HT and Heymann H (1998) *Sensory Evaluation of Food: Principles and Practices*. New York: Chapman & Hall.
- Ng LK, Lafontaine P, and Harnois J (2000) Gas chromatographic-mass spectrometric analysis of acids and phenols in distilled alcohol beverages – application of anion-exchange disk extraction combined with in-vial elution and silylation. *Journal of Chromatography A* 873: 29–38.
- Nykänen L and Suomalainen H (1983) *Aroma of Beer, Wine, and Distilled Alcoholic Beverages*. Berlin: Akademie-Verlag.
- Ough CS and Amerine MA (1988) *Methods for Analysis of Musts and Wines*, 2nd edn. New York: Wiley.



## Wine

**M T Kelly and A Blaise**, Universite Montpellier I, Montpellier, France

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The European Union (EU) directive 1493 (1999) defines wine as a product obtained exclusively from the total or partial alcoholic fermentation of crushed or intact fresh grapes or grape musts. During the process of vinification, grapes undergo two types of transformation:

1. The alcoholic fermentation whereby fermentable sugars (glucose and fructose) are converted into ethanol by the anaerobic action of yeast:



This reaction also produces  $25.4 \text{ kcal mol}^{-1}$  of energy that heats the mass in fermentation. The must contains between 150 and  $250 \text{ g l}^{-1}$  of sugar, which in the case of dry wines drops to less than  $2 \text{ g l}^{-1}$  when fermentation is completed after 4–10 days. If the fermentation is deliberately stopped, for example, by the addition of ethanol or sulfur dioxide, the resulting sweet wine will contain residual sugars of up to  $60 \text{ g l}^{-1}$  and an alcohol content slightly higher than that of a dry wine (15–17 vol.%).

2. The malolactic fermentation (MLF), where as a result of bacterial action malic acid is converted into lactic acid. This process is not, in reality, a ‘fermentation’ process, rather a biochemical process involving the removal of one of the carboxylic acid groups of malic acid by the action of bacteria of the *Pediococcus*, *Leucomastoc*, and *Lactobacillus* genera. Since the MLF reduces the acidity of a wine, it is a process that is encouraged in red wines and frequently prevented (by the addition of sulfur dioxide) in white and rosé wines in order to preserve the perception of freshness on the palate.

The analysis of a wine at all stages of its production is essential to ensure both the quality of the product from the consumer's point of view, and to establish that it conforms to the legislative requirements. It is for this reason that a series of analyses are routinely carried out – some on the wine during production, some on the finished product, and some on both. In both cases, chemical analyses are frequently complemented

by organoleptic analyses (tasting), though this aspect of wine analysis will not be treated here.

A compendium of analytical methods for wines, musts, and products used in wine technology was first published by the international wine organization (OIV: Organisation internationale de la vigne et du vin) in 1962. The compendium is updated on a yearly basis and most recent edition (see Further Reading) contains in excess of 70 monographs. The OIV works in close association with the European commission (EC) and in several cases where an EC method does not exist for a particular analysis it is recommended the recognized methods of the OIV be used.

In September 2000, the OIV adopted the system of method classification used by the Codex Alimentarius Commission for Method of Analysis and Sampling. In this system, methods of analysis are classified into four groups as follows:

**Category I (Criterion benchmark method):** This type of method determines a value that can be arrived at only by that method *per se* and which by definition serves as the only method for establishing the accepted value of the parameter measured.

**Category II (Benchmark method):** A category II method is designated as the benchmark method in cases where category I methods cannot be used. It should be selected from category III methods (as defined below). Such methods should be recommended for use in cases of disputes and for calibration purposes.

Categories I and II are reference methods that provide the most accurate results. They are used in case of litigation in international transactions, and should be employed for standard reference materials intended for subsequent standardization of other (nonreference) methods.

**Category III (approved alternative methods):** A category III method meets all the criteria specified by the Sub-Committee on Methods of Analysis and is used for monitoring, inspection, and regulatory purposes (e.g., enzymatic determinations of glucose and fructose).

**Category IV (auxiliary method):** A category IV method is a conventional (usual) or recently implemented technique for which the Sub-Committee on Methods of Analysis has not yet specified the requisite criteria (e.g., synthesized coloring agents, measurement of oxidation–reduction potential).

Only methods that have been adopted since September 2000 have been classified in this manner; existing analytical methods are still classified by the old

system (reference, usual, and rapid test methods). Since the analyses described in this article have not yet been re-classified by the OIV, they will, where appropriate, be referred to as simply 'reference' or 'usual' methods.

The analyses that are routinely carried out in wine are alcohol content (vol.%), sugars ( $\text{g l}^{-1}$ ), glucose and fructose ( $\text{g l}^{-1}$ ), glycerol ( $\text{g l}^{-1}$ ,  $\text{mg l}^{-1}$ ), pH, total acidity ( $\text{meq l}^{-1}$ ), lactic, malic, and tartaric acids, volatile acidity ( $\text{meq l}^{-1}$ ), malvidol diglucoside, and polyphenols. The total and free sulfur dioxide content is also routinely analyzed in wine; however, since it is not a native constituent of wine its analysis will not be described in this article.

## Alcohol

Next to water, alcohol is the most abundant constituent of wine. It is estimated that it was first analyzed by Arnauld de Villeneuve in Montpellier in the early part of the fourteenth century, though the notion of alcoholic strength in terms of degrees of alcohol was first proposed by Gay-Lussac some 200 years ago, a concept that has changed little in the intervening period. The alcohol content is usually expressed in terms of percentage volume (vol.%), specifically the number of liters of ethyl alcohol contained in 100 l of the product, measured at  $20^{\circ}\text{C}$ . The number of grams per liter of ethanol in the product are obtained by multiplying the vol.% by  $0.78924 \text{ g ml}^{-1}$ . There are various reasons for analyzing the alcohol content of wine. The first is to ensure that a wine conforms to legislative criteria currently in force. According to the EU directive 1493 (1999), a wine may only be classified as such if it contains a minimum amount of alcohol, which varies from 6% to 10% volume depending on the climatic production zone and whether the must has been enriched with grape or beet sugar. Dry wines may not contain more than 15% alcohol by volume and wines originating in countries outside the EU for sale within the Union must contain not less than 9% and not more than 15% alcohol by volume. Secondly, since alcoholic beverages are subject to taxes and duty that depend on the alcohol content, the accurate determination of this parameter will have significant commercial implications, especially for transactions involving large volumes of wine.

There are a variety of methods in which the alcohol content of a wine may be determined, which may be as **Table 1** shows, loosely divided into direct and indirect methods. The latter involves prior separation of the alcohol by distillation, usually by steam distillation, whereas the direct methods may be carried out on wine without any sample pretreatment. Within both categories there exist both physical, and less widely used, chemical methods.

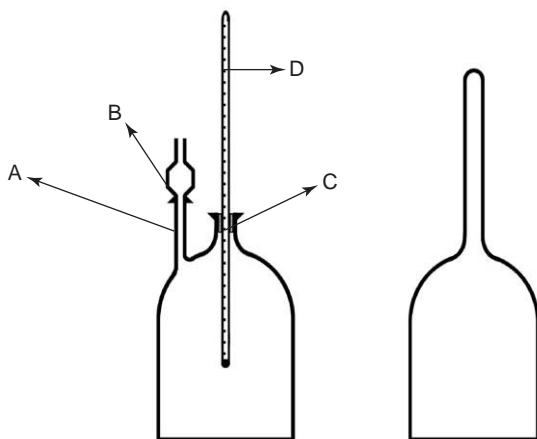
**Table 1** Methods used for the determination of alcohol

	<i>Physical</i>	<i>Chemical</i>
Indirect (distillation)	Hydrometry  Pycnometry Hydrostatic balance Paar densitometer Refractometry	
Direct	Ebullimetry  NIR reflectance Heat of reaction (ethanol- $\text{K}_2\text{SO}_4$ ) Gas chromatography Direct Headspace Liquid chromatography	Enzymatic (biosensors)

### Indirect Methods

Distillation/pycnometry is a type I method for determination of alcoholic strength; it is used for labeling discrepancies and dispute resolution, since if correctly executed it gives a highly accurate result (to within 0.03 vol.%). The wine is steam distilled to collect the alcohol fraction with measures taken to avoid collecting other volatile substances. The distillate is then transferred to a glass recipient (a pycnometer – **Figure 1**) and weighed to four decimal places. The density of the distillate is calculated from the weight of the distillate and the alcoholic strength is then calculated from a series of tables (in the OIV compendium) relating the density to the vol.% at different temperatures. This method gives highly accurate results and does not require a standard but is time consuming, requires the availability of an accurate balance, and highly trained personnel.

Distillation/hydrometry is less time consuming, less cumbersome, and less expensive than distillation/pycnometry, but it yields less accurate results, in the order of 0.1 vol.%. Furthermore, it tends to give lower results and the hydrometer must be kept extremely clean, as any material adhering to it will falsify the results. The hydrostatic balance (which is also an OIV type I method) is based both on gravimetry and the principle of Archimedes: a plunger attached to a balance is totally immersed in the liquid (the distillate for wine alcohol analysis). The volume of the displaced liquid is constant; its density, obtained simply by weighing, is directly converted in automated instruments to alcoholic strength. Refractometry, also an OIV 'usual' method is much less widely used. Measurements of the



**Figure 1** A pycnometer and its matching tare. A, lateral capillary tube; B, dilation chamber; C, ground glass neck; and D, thermometer.

distillate are made using a refractometer capable of reading refraction indices between 1.330 and 1.346. The temperature is recorded, the index of refraction at ambient temperature is converted to its equivalent value at 20°C, and the alcoholic strength is read from appropriate conversion tables. The vibrating tube density detector is an OIV type I method and is extremely and widely used. It consists of a U-shaped tube that creates mechanical resonant vibrations within the liquid it contains. The square of the resonance frequency is inversely proportional to the density, which can be determined in a single measurement and is directly converted into alcohol content in automated devices.

### Direct Methods

Alcoholic strength analysis by near-infrared reflectance. This technology is extremely and widely used and is the industry standard for both larger wineries and for commercial oenological laboratories. This fast and accurate method of measuring alcoholic strength is calibrated using distillation/pycnometry.

Enzymatic methods for alcohol determination frequently rely on the activity of alcohol oxidase alone or combined with horseradish peroxidase alcohol, quinoxinoprotein alcohol dehydrogenase, or pyrroloquinoline quinine-dependent dehydrogenases. The enzymes are usually incorporated into biosensors for automated and routine application. The reaction may be followed by monitoring the concentration of either NADH or hydrogen peroxide produced in the reaction. The concentration of NADH is determined by its ultraviolet (UV) absorbance at 340 nm, and amperometric or

chemiluminescence detection is used to measure the hydrogen peroxide produced.

In order to automate the analysis, these methods frequently combine immobilized enzymes with flow or sequential injection techniques. These methods may include a separation step such as solid-phase extraction, gas diffusion, or pervaporation. The latter is a nonchromatographic separation technique, which selectively separates a liquid mixture by partial vaporization through a nonporous polymeric membrane. Separation is not based on relative volatilities as in distillation, but rather on the relative rates of permeation through the membrane.

The advantage of enzyme-based methods is that they are highly specific – methanol and higher alcohols are not detected; however, as previously stated, these compounds are present in such minor quantities compared to ethanol that the accuracy of the result is not significantly improved. On the other hand, since enzymatic methods are highly sensitive, dilution factors ranging from 50 to 1000 are required, which may lead to significant error in the final determination. For this reason, also due to the fact that these methods are costly, the majority of oenological laboratories do not use enzymatic methods to analyze ethanol in wine.

High-performance liquid chromatography (HPLC) may be used for the simultaneous separation of ethanol and other wine constituents such as carbohydrates, glucose, fructose, and saccharose, coupled with either refractive index or Fourier transform infrared (FTIR) detection. Ethanol may also be code-terminated with other volatile compounds in wine by gas chromatography with flame ionization detection. It is important to note that chromatographic methods for the analysis of ethanol are not suitable where accurate determinations are required.

### Sugars

In a dry wine, the concentration of residual sugars, i.e., the amount of sugar remaining after the completion of the alcoholic fermentation, is in the region of 0.3–3 g l<sup>-1</sup>. Above this concentration there is a risk that bacteria such as *Leuconostoc oenos* (one of the bacteria involved in the malolactic fermentation) will degrade the sugars leading eventually to the formation of acetic acid and thus spoilage of the wine. For that reason the criteria of the Appellation Contrôlée system in France require that a dry wine does not contain more than 2 g l<sup>-1</sup> of (reducing) sugars and why wines that contain significantly more than this are frequently sterilized by filtration or pasteurization before bottling.

### Total Reducing Sugars

For the reason outlined above, the analysis of residual sugar forms part of the routine analysis of wine. The OIV reference method involves a redox reaction that determines the total reducing sugar content (composed primarily of glucose and fructose). The method requires a sample preparation step to eliminate other reducing compounds such as polyphenols and uronic acids. The wine is neutralized, dealcoholized, and applied to an anion-exchange column, which fixes the uronic acids. The eluate is then treated with a mixture of lead acetate and calcium carbonate to precipitate the tannins. The mixture is filtered and the filtrate is then analyzed by redox titration by the Luff method. This analysis is precise but requires a certain level of skill and extreme attention must be paid to factors such as the reaction time and dilutions in the case of sugar-rich wines.

To obtain an estimate of the quantity of sugar in wine, it is possible to use a proprietary method (Clinitest<sup>®</sup> AMES) that was originally designed to measure sugar in urine. The method, also based on the reducing properties of the sugars, involves reacting a small volume of wine with a Clinitest tablet and comparing the color obtained with a scale provided by the manufacturer. It is necessary to modify the scale as follows in order to apply it to wines:

AMES values	0	0.25	5	7.5	10	20	g l <sup>-1</sup>
Modified values	1	2	3	5			g l <sup>-1</sup>

This method is useful for distinguishing wines in which the alcoholic fermentation has finished and those that yet contain several g l<sup>-1</sup> of sugar; for accurate analyses alternative methods should be used. Its advantages of low cost and simplicity of use, however, render it accessible to even the smallest wineries.

Automated or semiautomated versions of classic chemical reactions for sugars (e.g., Fehlings solution, picric acid potassium hexacyanoferrate(III)) are also used. A so-called UV/UV system was recently described to follow the evolution of sugar concentration during winemaking, but which may also be applied to finished wines. The principle of this system is that the sugars are converted by exposure to UV light at 254 nm into UV-absorbing aldehydes and ketones characterized by a peak at ~268 nm. Interference by ethanol in the analysis is removed by dilution of the sample.

### Determination of Individual Sugars

The OIV compendium describes enzymatic individual methods for the determination of glucose and fructose. These methods are based on the conversion

of glucose-6 phosphate (G6P) to gluconate-6-phosphate in the presence of G6P dehydrogenase. The concomitant reduction of NAD to NADH (related to the concentration of G6P) is measured spectrophotometrically at 340 nm. Fructose-6-phosphate is converted to G6P by hexokinase before reaction of G6P dehydrogenase. Sensor-based enzymatic methods for the determination of sugars have also been described, usually involving immobilized glucose oxidase or pyrroloquinoline quinone dependent (oxygen independent) glucose dehydrogenase.

Individual sugars can also be measured by separation techniques such as liquid chromatography (LC) and capillary electrophoresis (CE). Separations are usually ion-exchange based with refractive index detection. With this type of method it is also possible to determine simultaneously other wine constituents such as carboxylic acids, sugars, glycerol, and ethanol in wines. Different detection modes may be used in conjunction with postcolumn derivatization or alternatively the eluate may be directly analyzed by FTIR.

### Acids and Acidity

Wine is one of the most acidic alcoholic beverages with pH values ranging from 2.8 to 3.8; white and rosé wines are, in general, more acidic than red wines due to the natural desacidification brought about by the MLF (see Introduction). It is also possible to adjust the pH of a wine, a procedure that may be required for technological reasons or to protect against bacterial growth. Tartaric acid (the most abundant acid in grapes and wines) is normally used for acidification, whereas desacidification – a more unusual and delicate procedure – involves the addition of either potassium tartrate, or calcium or potassium hydrogen carbonate.

The acidity of a wine encompasses a number of parameters

1. pH,
2. Individual wine acids,
3. Total acidity, and
4. Volatile acidity.

Since the measurement of pH is a standard procedure, it will not be treated here. The other three parameters will each be discussed in turn.

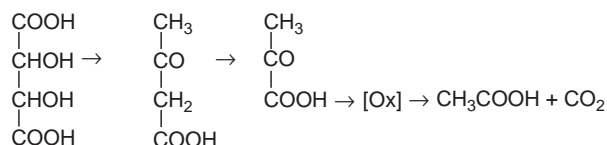
### Total Acidity

The total or titratable acidity of a wine is defined as the fraction of nonionized acids in the wine; officially, it is determined by the volume of sodium hydroxide required to bring the pH of a specified

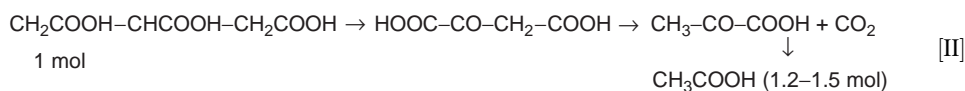
volume of wine to 7. This means, of course, that a proportion of the acids will not be neutralized (as the vast majority of the acids involved are weak acids); however, it would be impossible to determine the true endpoint of the titration given the variety of acids (and thus  $pK_a$ s) in question. The OIV methods for the determination of total acidity use 10 ml of wine; a solution of  $0.1 \text{ mol l}^{-1}$  sodium hydroxide and the titration is effected either by pH measurement (reference method) or with bromthymol blue indicator (usual method). The results are expressed in  $\text{meq l}^{-1}$  (titer volume  $\times 10$ ), in terms of  $\text{g l}^{-1}$  tartaric acid (titer volume  $\times 0.75$ ), or more usually in terms of  $\text{g l}^{-1}$  sulfuric acid (titer volume  $\times 0.49$ ). Common values are in the region of  $60\text{--}70 \text{ meq l}^{-1}$ . Since carbon dioxide is not considered as part of the total acidity, it is first removed from the wine by agitation under vacuum.

### Volatile Acidity

The volatile acidity is a test of the state of health of a wine. This is because it is represented principally by acetic acid that may be formed by the oxidation of ethanol by *Mycoderma acetii*, the degradation of tartaric acid by *Lactobacillus plantarum* (reaction [I]) or by the degradation of citric acid by *Leuconostoc oenos* of (reaction [II]).



[I]



Acetic acid then combines with ethanol to form ethyl ethanoate, which contributes to the pungency perceived wines with high volatile acidity.

Although limits are imposed for the amount of this acid in a wine (OIV:  $20 \text{ meq l}^{-1}$ ) it is a significant increase in the concentration of acetic acid (rather than its absolute value) during the course of vinification or aging that signals a problem with the wine. In order to correctly determine the concentration in volatile acidity, it is necessary to eliminate  $\text{CO}_2$  and subtract the concentration of various antiseptics or preservatives that may have been added, i.e., sulfur dioxide, sorbic acid, salicylic acid, and sodium salicylate. Volatile acidity is expressed in terms of

$\text{meq l}^{-1}$ , or in terms of  $\text{g l}^{-1}$  sulfuric acid, or more frequently acetic acid. The method involves steam distillation since 100% recovery of acetic acid would not be possible with direct distillation.

### Individual Organic Wine Acids

The acids present in a given wine are determined by the grape variety, climate, presence of gray rot (*Botrytis cinerea*), yeasts, bacteria, and various treatments to which the wine may be subjected (sulfur dioxide, ascorbic acid, acidification, desacidification). There are at least 50 different acids in wine ranging in concentration from 1 or  $2 \text{ g l}^{-1}$  (tartaric, malic, succinic acids) to hundreds of  $\text{mg l}^{-1}$  (citric, lactic, acetic acids) to tens of  $\text{mg l}^{-1}$  (pyruvic, shikimic acids). However, tartaric, malic (depending on the MLF), acetic, and succinic acids constitute 80–90% of total complement of wine acids.

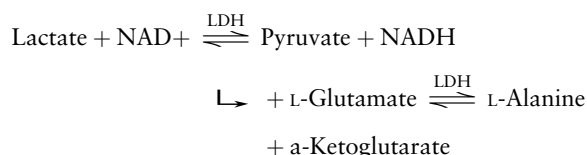
Simultaneous determination of individual organic wine acids may be obtained by either LC or CE. In the case of the former, either reversed-phase liquid chromatography (RPLC) or ion-exchange chromatography may be used if only the principal organic acids are to be determined; a combination of both chromatographic modes may be necessary if the lesser organic acids, such as shikimic acid, are to be analyzed simultaneously. In either case, the mobile phase consists of a dilute aqueous acidic solution; even in RPLC no organic modifier is added due to the polarity of the organic acids. Various detection methods are employed, for example, direct UV detection at 254 nm in the case of HPLC, indirect UV in CE, and conductimetric or refractive index detection with ion-exchange chromatography. Some methods do

not require sample preparation but greater sensitivity and selectivity is obtained if the wine is first passed through a reversed-phase solid-phase extraction cartridge to remove interfering compounds such as sugars, polyphenols, and anthocyanins. The OIV compendium gives an HPLC method for the simultaneous determination of citric, lactic, and acetic acids (strong ion-exchange chromatography) or tartaric, malic, malic shikimic, succinic, and fumaric acids (reversed-phase chromatography). In the latter case, it is necessary to use two 25 cm (octadecylsiloxane-bonded silica) columns in series. Detection in both cases is at 210 nm and sample cleanup is by solid-phase extraction.

Chemical methods also exist for the various organic wine acids. For tartaric acid, the OIV describes a method involving sample cleanup on an ion-exchange column followed by reaction between tartaric acid and vanadic acid to give a red complex that is measured spectrophotometrically. This reaction may be incorporated into a flow injection system, which eliminates the need for the ion-exchange sample cleanup step. Lactic acid is first oxidized to ethanol and measured by colorimetry following reaction with nitroprusside and piperidine; for citric acid the chemical method requires controlled oxidation to form acetone which is subsequently separated by distillation and determined by iodometric titration.

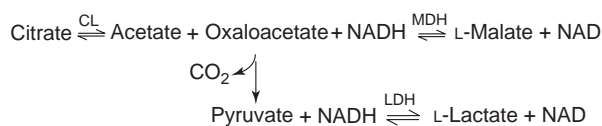
Enzymatic methods rely on similar principles: redox reactions involving the acids associated with either the oxidation of NADH or the reduction of NAD. The reaction is followed by the appearance or disappearance of NADH, which is measured spectrophotometrically at 340 nm. The primary reaction with the acid is often coupled to a second reaction in order to remove the product and thus drive the reaction equilibria to completion. Some of the possible reactions are summarized in the following schemes:

#### Lactate



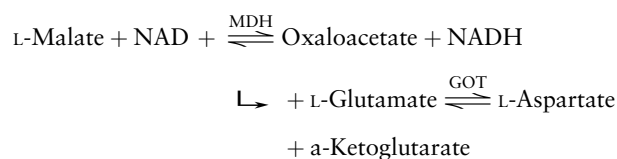
LDH: lactate dehydrogenase (L-LDH for L-lactate and for the conversion of pyruvate; D-LDH for D-lactate).

#### Citrate



CL: citrate lyase; MDH: malate dehydrogenase.

#### L-Malate



GOT: glutamate-oxaloacetate transaminase.

Paper chromatography is extensively used to follow the progress of the malolactic fermentation and the simplicity of this method is such that practically all wineries can carry it out. It is based on the difference in  $R_f$  values between malic and lactic acids using a mobile phase consisting of acetic acid and butanol. Bromothymol blue is incorporated into the mobile phase as indicator and thus the acids appear as yellow spots on a blue background. The absence of a spot for malic or lactic acid indicates that the MLF has or has not, respectively, taken place.

## Phenolic Compounds

The phenolic compounds are amongst the most important grape and wine constituents. They are responsible for all the differences in color and taste between red and white wines. They represent a diverse group of compounds for which a number of different classification systems have been applied.

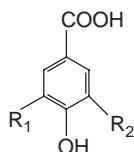
Wine phenols are commonly referred to as 'polyphenols', due as may be seen in **Figure 2**, to the presence of multiple phenolic groups in their structures, which confer on these compounds various properties linked to health benefits, specifically the antioxidant properties attributed to the consumption of moderate amounts of red wine. Much has been published on this subject and the area and references to reviews may be found in 'Further reading' section.

The phenolic acids of wine (**Figure 2A**) consist principally of derivatives of benzoic and cinnamic acids. The cinnamic acids are extensively esterified with tartaric acid (e.g., caffeoyltartaric or caftaric acid) and they also form acylated derivatives of anthocyanin monoglycosides (see below). These compounds do not play a significant role in the organoleptic properties of wine, though caftaric acid is highly sensitive to oxidation, which turns the must or the wine brown. Bacterial action can lead to the conversion of cinnamic acid derivatives (principally *p*-coumaric and ferulic acids) into volatile phenols such as ethyl phenol, ethyl gaiacol, or vinyl gaiacol, which even in minor amounts lend intensely disagreeable aromas and flavor to the wine.

The flavonoids (**Figure 2B**) are yellow-colored pigments that give the pale yellow color to white wines. They are also present in red wines but their color is masked by the anthocyanins (**Figure 2C**), the coloring substances of red wines. Like the anthocyanins, the flavonoids exist primarily as heterosides where the 3-position on the heterocycle contains a sugar moiety. The latter may in turn be acylated with cinnamic acid, which confers greater stability on the molecule.

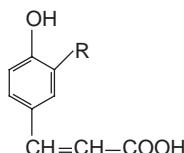


## (A) Phenolic acids

(i) Derivatives of *o*- and *p*-hydroxybenzoic acid

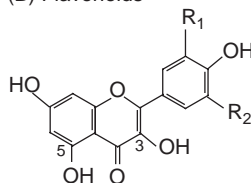
R = H	R' = OH	Catechic acid
R = OH	R' = OH	Gallic acid
R = OCH <sub>3</sub>	R' = OH	Vanillic acid

## (ii) Derivatives of cinnamic acid



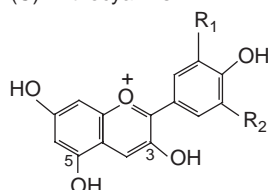
R = H	Coumaric acid
R = OH	Caffeic acid
R = OCH <sub>3</sub>	Ferulic acid

## (B) Flavonoids



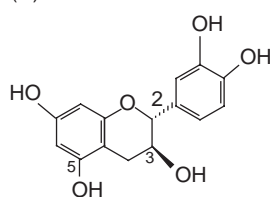
R <sub>1</sub>	R <sub>2</sub>	Aglycone
H	H	Kaempferol
OH	H	Quercetin
OH	OH	Myricetin

## (C) Anthocyanins



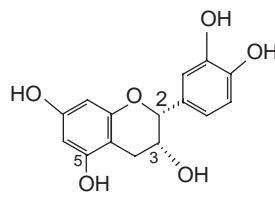
R <sub>1</sub>	R <sub>2</sub>	Aglycone
OH	H	Cyanidin
OCH <sub>3</sub>	H	Paeonidin
OH	OH	Delphinidin
OH	OCH <sub>3</sub>	Petunidin
OCH <sub>3</sub>	OCH <sub>3</sub>	Malvidine

## (D) Tannins



(+) Catechin 2R, 3S

(–) Catechin 2S, 3R



(+) Epicatechin 2S, 3S

(–) Epicatechin 2R, 3R

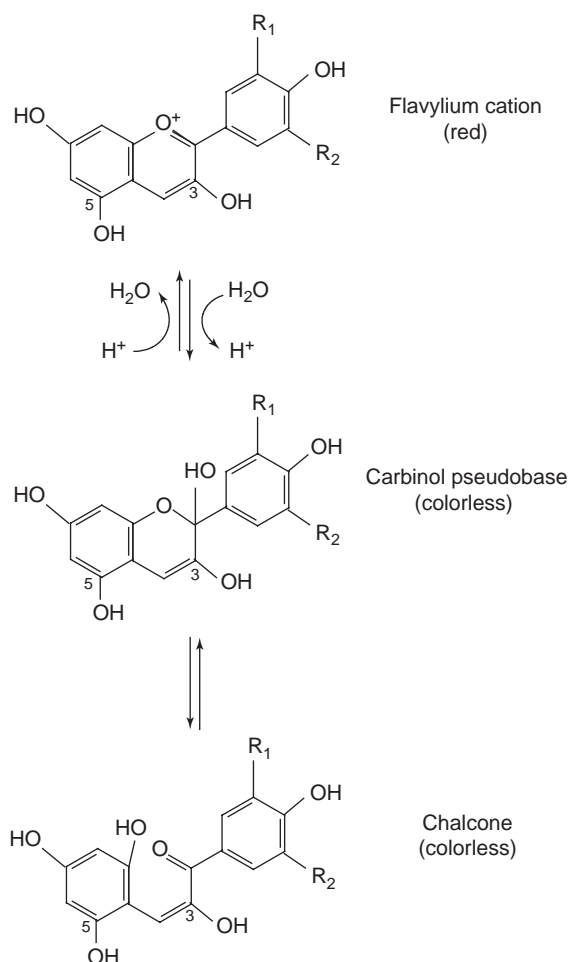
**Figure 2** Structure of phenolic compounds. (A) Phenolic acids: (i) derivatives of *o*- and *p*-hydroxybenzoic acid, (ii) derivatives of cinnamic acid; (B) flavonoids; (C) anthocyanins; and (D) Tannins.

The sugar moiety, which in wine anthocyanins is usually D-glucose, or more rarely L-rhamnose or D-galactose, occurs principally in the 3-position. Malvidne-3-glucoside is by far the most abundant of the anthocyanins in wine produced from *Vitis vinifera* varieties. It should be noted that certain hybrid species of *Vitis* (*Vitis vinifera* × *V. berlandieri* or *V. rupestris*) contain malvidne-3,5-diglucoside and these species

are not permitted in wines produced or intended for sale in the European Union on account of the elevated levels of methanol they produce during vinification. Identification of the presence of such a hybrid variety in a wine is by determination of the quantity of malvidine-3,5-diglucoside, an analysis that is rapidly carried out by thin-layer chromatography.

The tannins naturally occurring in wine are known as condensed tannins, polymeric structures resulting from the polymerization of elementary phenol compounds, polyhydroxyflavan-3-ols known as the catechins (**Figure 2D**). Depending on the configuration of the chiral carbons in positions 2 and 3 of the heterocycle, four different stereoisomers are possible, (+) and (–) catechin and (+) and (–) epicatechin; (+) catechin and (–) epicatechin are the principal building blocks of wine tannins; these monomers rapidly condense to form dimers, trimers, oligomers, and polymers during the course of vinification and wine aging. They also condense with anthocyanins, a process that during vinification is important in the stabilization of wine color. Unlike the flavonols, the pyran ring of the flavan-3-ols is saturated and non-planar and so these compounds have a maximum UV–visible absorbance at 280 nm, as opposed to ~370 and 520 nm for the flavonoids and the anthocyanins, respectively.

The methods employed for the determination of wine phenolic compounds are highly varied, given the diverse nature of their structure, polarity, hydrophobicity, and UV and visible-light absorption characteristics. It is possible to separate simultaneously the majority of the principal nonpolymerized wine phenolic compounds by HPLC or CE. In the case of the former, a reversed-phase column, usually octadecyl carbon (C18), is used. If the analysis is to include separations of the anthocyanins, the aqueous component of the mobile phase requires a pH of <2 (pH values as low as 1.5 have been used); either formic or acetic acids or a combination of both in concentrations ranging from 2% to 10% are used for this purpose. Above pH 2, severe peak broadening results from the slow interconversion between the various anthocyanin species leading to poor resolution and poor detection limits. In addition, the spectral properties of the anthocyanins are poor at pH values greater than 2 owing to the conversion of the flavylium cation to colorless carbinol pseudobase form (**Figure 3**). This means that chromatographic columns specifically designed to withstand extreme pH values are recommended for this analysis; in addition, peak tailing due to interaction between the hydroxyl groups of the polyphenols and unreacted silanol groups on the silica surface is minimized by using base-deactivated columns. The organic



**Figure 3** Different forms of anthocyanins.

component of the mobile phase is usually acetonitrile, though methanol is occasionally used. In order to achieve satisfactory separation of compounds of such diverse polarities, elaborate binary or even ternary gradients are required. Analysis time ranges from 50 to 120 min depending on the exact operating conditions and the degree of separation required. With a combination of judiciously chosen gradient and UV-visible photodiode array detection, it is possible to simultaneously analyze tannin monomers, dimers, and trimers ( $\lambda_{\max}$  280 nm), phenolic acids ( $\lambda_{\max}$  ~330 nm), flavonoids ( $\lambda_{\max}$  ~370 nm), and anthocyanins ( $\lambda_{\max}$  ~525 nm) in a single run. However, the presence of various combinations of these compounds, in varying concentrations, combined with the unavailability of many of the standards renders interpretation of the chromatograms rather difficult. Partially for this reason, mass spectrometry is more and more frequently used as a detection method for polyphenol analysis – atmospheric pressure-ionization techniques tend to be the most widely employed.

CE is not as widely used as LC for the analysis of wine polyphenols. As regards CE–UV analysis of the anthocyanins, separations must be carried out at very acidic pHs (due to the poor spectral properties of these compounds at neutral or basic pH values) and this presents problems in terms of method selectivity. CE is, on the other hand, better suited to the non-anthocyanin fraction of wine polyphenols.

As outlined earlier, a certain fraction of the wine polyphenols exist in condensed forms (tannin–tannin, anthocyanin–anthocyanin, tannin–anthocyanin), this fraction varying with the conditions of vinification and increasing with the age of the wine. For this reason, the above separation techniques remain relatively limited for the analysis of polyphenols given that only elementary nonpolymerized molecules, which play a relatively minor role in the overall properties of the polyphenols, may be determined. In order to analyze the condensed or polymerized fraction by one of these separation techniques it is necessary to carry out a thiolysis beforehand.

Except for specific applications, the requirements for routine polyphenol analysis can be adequately met by the application of a number of tests that are rapidly carried out without the need for elaborate instrumentation. It is difficult to determine or to express the results in terms of concentration, so therefore various indices are used that represent either total polyphenols or different groups of polyphenolic compounds. The tests briefly described below are useful in following the progress of a vinification, in making comparisons between different wines, and in seeking to quantify certain organoleptic properties of wines (notably bitterness and astringency).

The total polyphenol index is obtained by diluting the wine and measuring its absorbance at 280 nm. The index is given as the absorbance multiplied by the dilution factor (100 in the case of red wines). This test does not take into account the colorless chalcone forms of anthocyanins (**Figure 3**) nor the phenolic acids that absorb at 230 nm, though these compounds are present in such small quantities that they contribute little to the overall polyphenol content.

The permanganate index is based on the reduction of permanganate by the polyphenols using the indicator carmino indigo, the color of which varies between blue and yellow according to its oxidation state. The correct endpoint – when the polyphenols are oxidized but not the sugars – is difficult to identify, and furthermore the reproducibility of the method is poor between operators. Nonetheless, it is possible to classify wines according to their

permanganate index (milliequivalents of permanganate used in the titration) as follows:

Red press wine	100–150
Tannic red wine	75
Soft red wine	60
Rosé wine	27–35
White wine	1–15

The Folin–Ciocalteu (F–C) index is a more reliable indicator of polyphenol content and is extensively used. The reaction mixture consists of phosphotungstic and phosphomolybdic acids, which in alkaline medium, are reduced by the polyphenols to a mixture of blue oxides of tungsten and molybdenum, the absorbance of which is read at 750 nm. The F–C index is expressed as 100-fold the absorbance of the sample; typical values are in the region of 25–35.

There are number of chemical methods for the determination of the total anthocyanin content in a wine, the most simple being variations in the color of a wine as a function of pH. At very acidic pH, anthocyanins are in their flavinium cation form, whereas under mildly acidic conditions the carbinol pseudobase (**Figure 3**) is formed. The absorbance of the wine is read at pH 0.6 and at pH 3.5 and the difference in absorbance has been empirically shown to correlate with the anthocyanin concentration. Anthocyanin determination using the difference in absorbance before and after the addition of a known quantity of sulfur dioxide is based on the same principle.

It is also possible to fractionate wine anthocyanins on the basis of their degree of polymerization on a hand-packed column consisting of polyvinylpyrrolidone, silica gel G, and silica 60. The free anthocyanins elute in the first fraction (methanol–HCl 999:1), slightly polymerized anthocyanins in the second fraction (formic acid–water 1:1), and the condensed polymeric forms are eluted in 100% formic acid. The absorbance of the three fractions is read spectrophotometrically (538 nm for the first fraction and 525 nm for the second and third fractions) and their relative proportions may thus be calculated. This test is a very good indicator of the age of a wine: the younger the wine the higher the absorbance of the first fraction due to the presence of anthocyanin monomers; conversely, the absorbance of the third (polymerized) fraction will be greatest in older wines.

Quantitation of total tannin concentration is still carried out by methods based on the classic Bate-Smith reaction: heating in strong acid medium leads to the formation of carbocations, which in an oxidant medium forms cyanidin (an anthocyanin). The reaction mixture thus turns red (hence the old name

for the tannins – proanthocyanidins) and its absorbance is read at 550 nm. Later refinements of the original 1954 method led to the incorporation of further measures in the technique. Various empirically derived calculations based on the absorbance at 470, 520, 550, and 570 nm can give a reasonably accurate measure of the total tannin content of a wine. These measurements do not, however, distinguish between tannin monomers, oligomers, or polymers. Tannins produce a sensation of bitterness or astringency on the palate due to their interaction with proline-rich proteins and glycoproteins in saliva. The tannin molecules most active in this respect are those whose molecular mass is between 600 and 3500; tannins of greater molecular mass are too cumbersome to access the active sites of the proteins, and this explains why during wine ageing, as tannins become more polymerized amongst themselves or with anthocyanins, they appear less aggressive to the palate. Eventually, this polymerization leads to molecules that are sufficiently large to precipitate, explaining the deposit often seen in old wines. This principle of the reaction between tannins and proteins forms the basis of a test, known as the gelatine index, for the degree of astringency of a wine. A specified volume of wine is mixed with a solution of gelatin (MM 5000–30 000) and allowed to react for 3 days and centrifuged. The tannin concentration of the wine before and after reaction with gelatin is determined as described above, and the index is expressed in terms of the percentage difference. The higher the gelatin index the greater the concentration of ‘astringent’ tannins in the wine. This test correlates reasonably well with the degree of astringency perceived by expert tasters.

## Automated Methods

To conclude this article, it is important to state that, in general, commercial oenological laboratories are equipped with automated instrumentation that carry out the above analyses (and others besides) in a single step. The most widely used instrumental technique is based on FTIR analysis. The infrared spectrum of an organic solution such as wine presents complex absorption spectra characteristic of the different wine components. The Michelson interferometer, which is at heart of the FTIR method, is based on the division of a polychromatic band of infrared radiation into two beams which then follow different optical pathways; one beam traverses the sample cell directly, while the other is reflected on a mobile mirror before arriving at the sample cell. For each elemental wavelength arriving at the detector cell there will be a phase difference, which is continuously varied

by the changing position of the moveable mirror. The two out-of-phase signals of the same wavelength arriving at the detector results in an interference signal, the intensity of which depends on the magnitude of phase difference. This variation in signal intensity as a function of the phase difference is called an interferogram. Fourier transform is a mathematical procedure that enables the construction of an infrared spectrum on the basis of the interferogram. A typical FTIR instrument used in wine analysis measures spectra between 1000 and  $5000\text{ cm}^{-1}$  (10 000–2000 nm), which lies between the near- and mid-infrared. Once the spectrum for each of the components in the wine have been established, specific wavelengths (up to 15) are selected for their quantitation; the wavelengths selected are those that show the best correlation between absorption intensity and concentration for a given parameter. With 15 different wavelengths, up to 97% of the 'information' relating to a given parameter may be obtained. Based on the absorption intensity at specific wavelengths for the spectra of reference standards, a polynomial calibration model is calculated by a statistical method, the most commonly used of which is partial least squares.

The instrument must be calibrated using reference standards for each of the parameters that are to be analyzed, which is a fundamental step in assuring the quality of the future results. In order to calibrate the instrument correctly it is necessary to analyze a large number of reference standards containing each of the analytes at known concentrations; it is considered that up to 150 wines of the same type are needed to establish a good calibration. The concentration of each of the parameters in the reference standards should be determined by a reference method, such as the methods described in this article, and obviously the quality of the calibration is directly related to the accuracy of the results obtained by the reference method. With good quality calibrations, FTIR is capable of excellent repeatability and reproducibility, frequently superior to those obtained with official 'usual' methods and in some cases to official 'reference' methods. There is usually a good correlation between reference methods and FTIR analyses; however, if the wine contains a substance for which the instrument was not calibrated the results obtained will differ enormously from those obtained by a standard reference method for that substance.

The technique is remarkably simple in its execution and the sample to be analyzed does not require any particular sample pretreatment, though if the wine is turbid, clarification by centrifugation or filtration may be appropriate. If the concentration in carbon dioxide exceeds  $\sim 750\text{ mg l}^{-1}$  it is

recommended to remove it under vacuum or by ultrasonication to avoid degassing in the analysis circuit. The sample is aspirated by a sampling needle and transferred by a peristaltic pump into a chamber where the sample is heated to  $40^\circ\text{C}$ . After passage through a filter, the sample enters the measurement cell which is composed of fluorosilicate. A typical sample cycle is of the order of 30 s, automated instruments can therefore analyze up to 120 samples per hour.

In spite of its incontestable advantages, the method is subject to certain limitations. First, calibration is based on a sample population of wine bearing certain common characteristics, such as region of origin, sugar concentration, and acidity. This implies that a given calibration will not be valid for all types of wines (or musts); in other words, the matrix plays a major role, and parameters measured in matrices differing significantly from that used in the calibration will generate erroneous results. To combat this problem a calibration may be developed using several different wine matrices; however, this will degrade the analysis precision and so a compromise must be sought between robustness and precision if several types of wines are to be analyzed using the same calibration. Many oenological laboratories circumvent this problem by establishing specific calibrations for different 'types' of wines (dry wines, sweet wines, musts, etc.). Apart from the quality of the calibration, the absorption characteristics of the individual parameter are also determining factors. Some compounds that present highly specific absorption characteristics will be selectively determined whereas others – even though they may be present in high concentrations – will be less selectively determined and thus the sensitivity of their analysis will be compromised. As this is a macroanalytical method, precision drops dramatically at concentrations below  $100\text{ mg l}^{-1}$ . Although many laboratories routinely apply FTIR to the analysis of several wine analytes (alcohol, sugars, glucose, fructose, total acidity, major organic acids, total polyphenols, etc.) there are some parameters for which the method is either not adapted (e.g., mineral elements or compounds present in small concentrations, such as the antiseptic, sorbic acid) whereas other parameters are either extremely matrix-dependent (total sulfur dioxide) or yield results of unacceptable precision (free sulfur dioxide, volatile acidity). Automated analysis of these compounds may be accomplished by a multiparameter sequential analyzer, the basis of which is the UV-visible absorbance of the analytes following chemical (iron, copper) or enzymatic (malate, lactate, citrate) reaction with appropriate reagents.

See also: **Carbohydrates:** Sugars – Chromatographic Methods. **Ethanol. Food and Nutritional Analysis:** Alcoholic Beverages. **Infrared Spectroscopy:** Near-Infrared. **pH. Quality Assurance:** Primary Standards; Spectroscopic Standards. **Spectrophotometry:** Organic Compounds.

## Further Reading

- Barrus NW and Evans JA (1991) *A Handbook for Must and Wine Analysis*. University of Texas System.
- Boulton RB, Singleton VL, Bisson LF, and Kunkel RE (1996) *Principles and Practices of Winemaking*. Dordrecht: Kluwer Academic.
- Gump BH and Pruett DJ (eds.) (1993) *Beer and Wine Production: Analysis, Characterization and Technological Advances* (ACS Symposium Series). American Chemical Society.
- Peynaud E and Spencer AFG (1984) *Knowing and Making Wine*. New York: Interscience.
- Ribereau-Gayon P (ed.) (2000) *The Handbook of Enology*. New York: Wiley.
- Margalit Y (1990) *Winery Technology and Operations Handbook*. Wine Appreciation Guild.
- OIV (1999) *Compendium of International Methods of Analysis of Wines and Musts*. Paris: OIV Publications.
- Ough CS and Amerine MA (1988) *Methods for Analysis of Musts and Wines*. New York: Wiley.
- Watkins TR (ed.) (1997) *Wine Nutritional and Therapeutic Benefits* (ACS Symposium Series). American Chemical Society.
- Zoecklein BW, Fugelsang KC, Gump BH, and Nury FS (1995) *Wine Analysis and Production*. Dordrecht: Kluwer Academic.

## Meat and Meat Products

**M O’Keeffe**, The National Food Centre (Teagasc), Dublin, Republic of Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The analysis of meat and meat products is a significant activity in the general area of food analysis since they represent an important and relatively expensive component of the diet. Characterization of meat through chemical analysis is of importance to meat buyers in the food processing industry and is the subject of a large amount of regulatory control in most countries. Analysis of meat products is of importance for the food processing industry in product development, quality control/assurance, and for nutritional labeling. In addition, meat products are covered by a range of regulatory requirements to which adherence can be checked by chemical analyses. This article describes the analyses used for meat and meat products, covering both major and minor constituents, and addresses the range of methods available for particular analytes.

## Major Analytes

Fresh or unprocessed meat is characterized normally by microbial tests, by measurement of physical attributes such as tenderness and color, and by ‘proximate’ analysis; that is, the proportions of the major constituents of moisture, protein, fat, and ash

(inorganic material) in the meat. In the case of carcass meat, other measurements such as pH and color are commonly made to give an indication of quality. Meat is often tested for its freshness through rancidity tests such as peroxide value and thiobarbituric acid (TBA) number, which give a measure of oxidative rancidity in the fat, and the free fatty acid value, which gives a measure of hydrolytic rancidity. Spoilage of meat can be measured through assay of the total volatile basic nitrogen (TVBN), which is related to protein breakdown.

In the case of comminuted meat (e.g., burgers), the purpose of analysis is to characterize the main constituents (moisture, protein, fat, ash) and also to determine to what extent it differs from the intact meat. For example, typical analyses applied to comminuted meat samples might include meat species identification and determination of collagen and carbohydrate content. As the meat products move further away from the original entire meat, a range of analyses is used to characterize the food, particularly assays for nonmeat proteins, for additives such as salts, and for preservatives. In the case of cured meat and cured meat products, the important analyses are for salt, nitrite and nitrate, and other additives such as sugars and phosphorus (polyphosphates).

Meat products, under regulations, are defined according to ‘meat content’ and this may be calculated from compositional data. The percentage ‘fat-free meat’ is calculated from the total nitrogen content (corrected for any nonmeat nitrogen) using nitrogen factors specific for meat of each species. ‘Meat content’ is calculated as the ‘fat-free meat’ plus the fat.

See also: **Carbohydrates:** Sugars – Chromatographic Methods. **Ethanol. Food and Nutritional Analysis:** Alcoholic Beverages. **Infrared Spectroscopy:** Near-Infrared. **pH. Quality Assurance:** Primary Standards; Spectroscopic Standards. **Spectrophotometry:** Organic Compounds.

## Further Reading

- Barrus NW and Evans JA (1991) *A Handbook for Must and Wine Analysis*. University of Texas System.
- Boulton RB, Singleton VL, Bisson LF, and Kunkel RE (1996) *Principles and Practices of Winemaking*. Dordrecht: Kluwer Academic.
- Gump BH and Pruett DJ (eds.) (1993) *Beer and Wine Production: Analysis, Characterization and Technological Advances* (ACS Symposium Series). American Chemical Society.
- Peynaud E and Spencer AFG (1984) *Knowing and Making Wine*. New York: Interscience.
- Ribereau-Gayon P (ed.) (2000) *The Handbook of Enology*. New York: Wiley.
- Margalit Y (1990) *Winery Technology and Operations Handbook*. Wine Appreciation Guild.
- OIV (1999) *Compendium of International Methods of Analysis of Wines and Musts*. Paris: OIV Publications.
- Ough CS and Amerine MA (1988) *Methods for Analysis of Musts and Wines*. New York: Wiley.
- Watkins TR (ed.) (1997) *Wine Nutritional and Therapeutic Benefits* (ACS Symposium Series). American Chemical Society.
- Zoecklein BW, Fugelsang KC, Gump BH, and Nury FS (1995) *Wine Analysis and Production*. Dordrecht: Kluwer Academic.

## Meat and Meat Products

**M O’Keeffe**, The National Food Centre (Teagasc), Dublin, Republic of Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The analysis of meat and meat products is a significant activity in the general area of food analysis since they represent an important and relatively expensive component of the diet. Characterization of meat through chemical analysis is of importance to meat buyers in the food processing industry and is the subject of a large amount of regulatory control in most countries. Analysis of meat products is of importance for the food processing industry in product development, quality control/assurance, and for nutritional labeling. In addition, meat products are covered by a range of regulatory requirements to which adherence can be checked by chemical analyses. This article describes the analyses used for meat and meat products, covering both major and minor constituents, and addresses the range of methods available for particular analytes.

## Major Analytes

Fresh or unprocessed meat is characterized normally by microbial tests, by measurement of physical attributes such as tenderness and color, and by ‘proximate’ analysis; that is, the proportions of the major constituents of moisture, protein, fat, and ash

(inorganic material) in the meat. In the case of carcass meat, other measurements such as pH and color are commonly made to give an indication of quality. Meat is often tested for its freshness through rancidity tests such as peroxide value and thiobarbituric acid (TBA) number, which give a measure of oxidative rancidity in the fat, and the free fatty acid value, which gives a measure of hydrolytic rancidity. Spoilage of meat can be measured through assay of the total volatile basic nitrogen (TVBN), which is related to protein breakdown.

In the case of comminuted meat (e.g., burgers), the purpose of analysis is to characterize the main constituents (moisture, protein, fat, ash) and also to determine to what extent it differs from the intact meat. For example, typical analyses applied to comminuted meat samples might include meat species identification and determination of collagen and carbohydrate content. As the meat products move further away from the original entire meat, a range of analyses is used to characterize the food, particularly assays for nonmeat proteins, for additives such as salts, and for preservatives. In the case of cured meat and cured meat products, the important analyses are for salt, nitrite and nitrate, and other additives such as sugars and phosphorus (polyphosphates).

Meat products, under regulations, are defined according to ‘meat content’ and this may be calculated from compositional data. The percentage ‘fat-free meat’ is calculated from the total nitrogen content (corrected for any nonmeat nitrogen) using nitrogen factors specific for meat of each species. ‘Meat content’ is calculated as the ‘fat-free meat’ plus the fat.



Related analyses in the meat and meat products area are the analysis of curing brines and of special meat fractions such as mechanically recovered meat (MRM). Completing the list of analyses on meat and meat products are the mineral and vitamin assays used to characterize the nutritional status of these foods.

## Sampling and Preparation

Meat and other edible tissues of animals are difficult samples to analyze because of the nonuniformity and inherent instability of the sample. Because of the high water content ( $\geq 70\%$ ) of raw meat, meat samples are particularly prone to change either in storage or during preparation for analysis. This is more of a problem for raw meat than for meat products, which are often 'stabilized' by some of the constituents added to the meat or by the processing procedures (cooking, curing) used.

In the case of raw carcass meat, another problem is that of obtaining a representative sample. Because the muscles of the animal (the primal cuts) are irregular in shape and in fat content, samples for analysis must be clearly defined in terms of anatomical position. In the case of boxed beef, which is used in meat processing, it is important to have a sample representative of the total content, particularly in terms of lean and fat.

Samples of meat are taken either from defined locations, as multiple subsamples over the entire material, which are combined to form the analytical sample, or as a single sample after comminuting and mixing to ensure homogeneity. Preferably raw meat samples should be analyzed immediately but, if this is not possible, refrigerated or frozen storage is required to prevent (reduce) alteration to the sample.

Having defined the batch that is to be sampled, the sampling must be such as to ensure that the batch will be characterized adequately from the analyses. The number of samples to be taken from a batch depends on criteria such as the number of discrete units in the batch, the importance of the analyte, and the confidence required from the results. Since these are subjective or variable criteria, the number of samples is determined initially and then particular procedures of sampling are followed.

The sampling of meat and meat products is covered by an International Standard (ISO 3100-1: 1991). This standard gives general instructions and specifies procedures to be followed for taking primary samples and is intended primarily for commercial, rather than regulatory purposes. It distinguishes between units not exceeding 2 kg in weight, which

are treated as a sample, and carcasses or meat cuts exceeding 2 kg in weight, from which secondary samples of 0.5–1 kg may be taken. The above standard deals with issues such as the inertness of the containers and the integrity of closures and the necessity to store and transport samples at 0–2°C (for analyses to be completed within 24 h) or frozen (where longer storage is required). The standard also describes the use of seals and the information to be attached to the sample container label.

Before analysis, meat samples must be ground in a suitable mincer or bowl-cutter to produce a homogeneous sample. Connective tissue is a particularly difficult component of the meat sample to homogenize and this is of importance in collagen analysis. Suitable comminution of meat samples is specified as a mincer (meat chopper) plate size of 4 mm or less.

Equipment used for sample preparation must be dry and clean and may be cooled or chilled. Speedy sample preparation is required to reduce heating and consequent evaporation of water, which would affect the determination of moisture content. Blades must be sharp to achieve consistent and fast comminution. Samples (preferably well chilled or slightly frozen) may be prepared by passing through a meat chopper or mincer multiple ( $3 \times$ ) times, often using plates of decreasing size where the sample is very large. Often, quite small samples are received for analysis and a bottom-driven bowl cutter may be more appropriate.

In the case of meat products that are a composite (such as meat-filled pies), the analysis may require separation of the sample into its constituent parts, weighing of each part, and analysis as separate samples.

## Standard Methods

### Moisture

Moisture is determined by measuring the loss in weight of the sample on heating. A standard method (ISO 1442: 1997) used consists of drying a sample of 5–8 g, mixed with predried sand, to constant weight with 2 h periods in an oven at  $103 \pm 2^\circ\text{C}$ . Methods approved by the Association of Official Analytical Chemists (AOAC) for moisture include air-drying of a 2 g sample in a convection oven at 100–102°C for 16–18 h, at 125°C to constant weight, or under vacuum at 95–100°C. An alternative procedure for moisture determination is microwave drying; integrated systems exist consisting of an electronic balance and control microprocessor that can give results after a drying period of 3–5 min.

In all cases, moisture determination is based on recording the mass that evaporates from the sample

under strictly defined circumstances of temperature and time. It is possible in some samples (1) that components other than water may be lost by evaporation or (2) that incomplete removal of water may occur due to formation of a 'skin' on the surface of the sample. Moisture determination, therefore, requires close attention to the exact conditions used and maintenance of constant conditions for each analysis.

### Protein

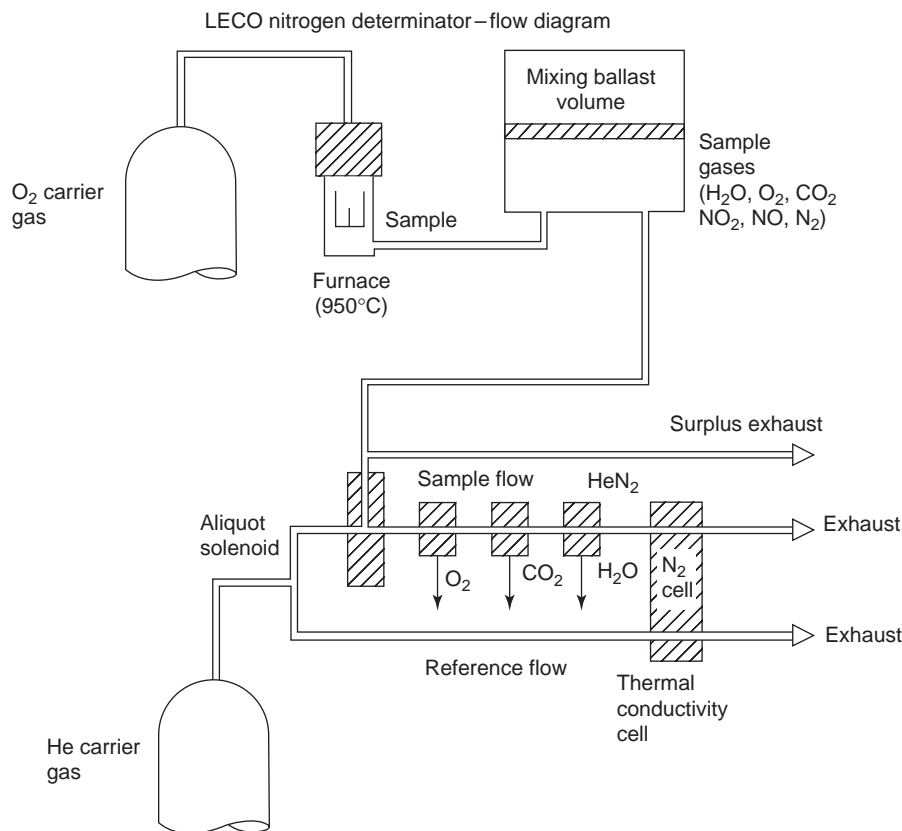
The total nitrogen content of the sample is determined after either reduction of all nitrogen to ammonia or liberation of all nitrogen as nitrogen gas. The 'crude protein' content of the sample is derived from the ammonia/nitrogen content by use of appropriate conversion factors. In the case of meat and meat products, the factor 6.25 is used to convert ammonia/nitrogen to crude protein, on the basis that 16% of protein is nitrogen.

Kjeldahl nitrogen is the classical assay for crude protein in meat, consisting essentially of a catalyst-aided (sulfate salts and copper or mercury) digestion with concentrated sulfuric acid at elevated

temperature, rendering of the resultant digest (containing ammonium sulfate) alkaline with concentrated sodium hydroxide, distillation of the liberated ammonia into an excess of acid, and back-titration with acid to determine the ammonia content. This ISO procedure (ISO 937: 1978) has been developed from a macrosystem in special flasks to a microsystem in tubes and in a variety of forms of automated and semiautomated systems – many of which are AOAC-approved methods.

An AOAC variation, still based on determination of ammonia, is an automated system of digestion with determination using the colorimetric reaction of ammonia with hypochlorite and phenol to produce an indophenol absorbing at 630 nm.

The Dumas method, based on determination of gaseous nitrogen in the sample, has been approved by the AOAC for crude protein in an automated form for a variety of foodstuffs. The modern versions of this system have fully automated sample introduction, combustion (950°C), removal of non-nitrogenous gases, and reduction of nitrogenous gases to nitrogen, and determination of nitrogen content by a thermal conductivity cell (Figure 1). The Dumas



**Figure 1** Crude protein determination by the Dumas method for total nitrogen with the 'LECO' apparatus. (Reproduced from O'Keeffe M (1992) *Chemical analysis of animal feed and human food*. In: Smyth MR (ed.) *Chemical Analysis in Complex Matrices*, pp. 241–287. Chichester: Ellis Horwood.)

**Table 1** Performance of Kjeldahl and Dumas methods for determination of protein in meat products (12 different laboratories, 30 different samples)

Performance measurement	Kjeldahl method	Dumas method
Repeatability ( $S_r$ )	0.11–0.40	0.12–0.41
Reproducibility ( $S_R$ )	0.29–0.49	0.18–0.46
Protein, overall mean ( $n=360$ )	15.59%	15.75%

From King-Brink M and Sebranek JG (1993) Combustion method for determining crude protein in meat and meat products. *JAOAC International* 76: 787–793.

method is attractive because it is a fully automated system, avoids use of corrosive and contaminant chemicals, and is a rapid method. However, the small sample size (<1 g) used in the Dumas method places critical requirements for homogeneity of sample that may not be as easily attained with meat samples as with other sample types. The Dumas method compares well with the Kjeldahl method both in terms of the performance criteria of repeatability and reproducibility and protein content determined, as shown from the results of an AOAC collaborative study among 12 different laboratories and using 30 different meat and meat product samples (Table 1). However, the Dumas method gives slightly higher protein values overall compared with the Kjeldahl method, which may be due to atmospheric nitrogen and/or additional nitrogen detected from basic amino acids. In samples with relatively low protein concentration (<30%), the difference between the two methods may not be of practical significance.

There may be a requirement to characterize the proteins in a meat product in terms of their amino acid content. This may be determined by hydrolysis with concentrated hydrochloric acid ( $6 \text{ mol l}^{-1}$ ) and separation and identification of the individual amino acids either (1) by ion exchange chromatography and colorimetric measurement using ninhydrin, or (2) by high-performance liquid chromatography of fluorescent derivatives of the amino acids.

## Fat

Standard methods for the determination of fat in meat and meat products are based on extraction of the fat from the sample using a lipophilic solvent and determination of the extracted fat by weighing or by measurement of specific gravity (relative density) changes in the extractant solvent. There are two ISO methods for fat in meat and meat products, one for 'free fat' (ISO 1444: 1996), which consists of repeated extraction of the sample with *n*-hexane or petroleum spirit in a Soxhlet apparatus, and the other for

'total fat' (ISO 1443: 1973), which involves an hydrolysis with hydrochloric acid prior to the solvent fat extraction. For some meat product samples, simple extraction of fat with solvent may be incomplete; digestion with hydrochloric acid liberates the fat by hydrolyzing proteins and carbohydrates. In both cases, the sample is oven-dried at  $103 \pm 2^\circ\text{C}$  prior to extraction with solvent. The fat content is determined by weighing the residue after evaporation of the solvent.

The primary AOAC method is a Soxhlet method, similar to the ISO 'free fat' method. An automated form of the Soxhlet petroleum spirit (petroleum ether) method is approved by the AOAC, using equipment such as the Soxtec<sup>TM</sup> apparatus. In this procedure, multiple samples, mixed with sand, are predried at  $125^\circ\text{C}$  (1 h) prior to a rapid solvent extraction step. There are two other AOAC-approved methods that are widely used for fat determination in meat and meat products. One method uses the Fosslet<sup>TM</sup> apparatus comprising a shaker or reactor that extracts the fat from the sample by very vigorous shaking in a metal container with perchloroethylene for 2–3 min. The specific gravity of the extract is measured in a magnetic float cell controlled by a potentiometer. The fat content can be determined from the difference in specific gravity between pure perchloroethylene and the extract. The second method is based on microwave-predrying of the sample, extraction of fat from the dried sample with dichloromethane in a custom-built extractor, and microwave-drying of the residue prior to weighing the extracted fat.

Both the Fosslet and the microwave systems are relatively rapid, a complete procedure taking not more than 10–15 min. Both these methods are very attractive for situations where rapid analyses are required, such as online for product control. However, where large batches of samples are to be analyzed and where results are not required within 24 h, the Soxhlet solvent extraction, or the more rapid Soxtec<sup>TM</sup> system, may be preferable. These latter techniques do not require constant attention, as is the case for the rapid methods. Good comparison between results obtained by Soxhlet extraction and the Fosslet technique has been reported. In the case of the microwave-based system, adjustment factors have been defined for different classes of meat products to overcome a negative bias in the method.

Supercritical fluid extraction (SFE) has been applied also for determination of fat in meat and meat products. Supercritical carbon dioxide, produced using elevated pressure (7.4 MPa) and temperature ( $31^\circ\text{C}$ ), extracts the fat from samples that have been mixed with moisture-absorbing agents, such as

diatomaceous earth. The advantage of this technique is that there is no waste solvent disposal problem and a number of samples may be extracted simultaneously. A number of studies have shown good agreement between the SFE and the Soxhlet methods.

X-ray absorption technology (e.g., Anyl-Ray<sup>TM</sup>) is used in the meat industry to determine indirectly the fat content of meat. The meat sample (~7 kg) is transilluminated with X-rays and the absorption of energy is related to the mineral content of the sample. Fat content is calculated from the relationship between fat, protein, water, and mineral content of meat.

Where the fat in a meat or meat product sample is to be characterized in terms of the fatty acid profile, extraction of the fat with chloroform/methanol is required. This solvent mixture, while it may not give complete fat extraction, is used to ensure no chemical change to the lipids and the extraction of phospholipids. Fatty acid analysis of the extracted fat is undertaken by formation of volatile methyl esters of the fatty acids (ISO 5509: 2000) and determination by gas chromatography (ISO 5508: 1990).

### Ash

Ash is the residue remaining after incineration of the sample in a furnace at temperatures in excess of 500°C. The AOAC method specifies drying of the sample and charring before ignition at 525°C. The ISO method (ISO 936: 1998) specifies incineration of the sample at  $550 \pm 25^\circ\text{C}$ , in a muffle furnace, following drying and carbonization of the sample either (1) separately in an oven and on a hot-plate or (2) in the furnace by gradually raising the temperature. In all cases, the ash is determined by weighing of the residue and care must be taken in handling the silica dishes to avoid losses.

A common practice is to determine ash content on samples that have been oven-dried for moisture determination. While ash may be used for determination of metals, salts, etc., attention must be given to the possibility of losses due to volatility of these components of the sample at the very high ashing temperatures used.

### pH

The pH of meat is an important measure of quality, a higher than normal pH (>pH 6.2 at 24 h after slaughter) or a lower than normal pH (<pH 5.9 at 45 min after slaughter) indicating conditions known as DFD (dark, firm, dry) and PSE (pale, soft, exudative) meat, respectively. The determination of such meat is important as the uses to which it can be put are limited. Such pH measurements are made on the carcass with a spear probe.

For laboratory measurements, an ISO method for pH (ISO 2917: 1999) specifies either homogenization of the sample in potassium chloride solution and insertion of the electrode(s) into the extract, or introduction of the electrode into prepared holes in the nonhomogenized sample. Other methods specify preparation of a slurry of the ground sample with an equal weight of water, or preparation of a homogenate in the ratio of one part meat to five parts sodium iodoacetate ( $5 \text{ mmol l}^{-1}$ )/potassium chloride ( $150 \text{ mmol l}^{-1}$ ).

### Collagen

Determination of the collagen content of meat products is important in controlling the eating quality (toughness) and for ensuring that too high an amount of low-grade meat, with a high level of connective tissue, is not in a product. Collagen content is determined from an analysis for the amino acid L-hydroxyproline, which occurs at much higher levels in collagen than in other muscle proteins.

The ISO method for L-hydroxyproline (ISO 3496: 1994) involves hydrolysis of the sample with sulfuric acid, at  $105^\circ\text{C}$  for 16 h. After filtration and dilution of the hydrolysate, the L-hydroxyproline content is determined by oxidation with chloramine-T and reaction with *p*-dimethylaminobenzaldehyde to form a colored complex that is measured spectrophotometrically at 558 nm. This method requires close attention to detail, particularly the color reaction which must be carried out under exact temperature and time conditions ( $60 \pm 0.5^\circ\text{C}$  for 20 min). The repeatability of the determination is assured by preparation of a calibration graph with each batch of samples. Using a specific microwave oven for digestion can reduce the time for this procedure to 30 min, so that the total time for analysis can take less than 3 h.

The collagen content is determined by multiplication of the L-hydroxyproline result by a factor of 8 (corresponding to the hydroxyproline content of collagen being equal to 12.5%, where the nitrogen-to-protein factor is 6.25).

### Additives

In the production of meat products, additives such as salts, sugars, and preservatives are used. Analyses for these substances are important in that maximum levels allowed in foods are specified in various national regulations such as the US Food and Drugs Administration and the European Commission DG SANCO.

The standard methods for sodium chloride in meat products are titrimetric methods in which the salt is extracted from the sample, reacted with excess silver

nitrate to form a silver chloride precipitate, and the residual silver nitrate titrated with thiocyanate. The ISO method (ISO 1841-1: 1996) specifies an extraction with hot water and precipitation of proteins with Carrez reagents (potassium hexacyanoferrate(II) and zinc acetate solutions), followed by acidification of the filtered extract with nitric acid before addition of excess silver nitrate. The silver chloride precipitate is coagulated with nitrobenzene and the residual silver nitrate titrated with potassium thiocyanate. In the AOAC method, the sample is treated with excess silver nitrate before extraction by boiling with dilute nitric acid and titration of residual silver nitrate is with ammonium thiocyanate. This method is appropriate for samples with a sodium chloride content  $\geq 1\%$ .

Another ISO method (ISO 1841-2: 1996), based on potentiometric titration with a silver nitrate solution using a silver electrode, is appropriate for samples with a sodium chloride content  $\geq 0.25\%$ . The sample is dispersed in water and an aliquot is acidified by addition of nitric acid, prior to titration.

Alternative methods for salt determination include use of a chloride ion-selective electrode and a semi-quantitative indicating strip method based on the reaction of chloride ion with silver chromate to produce silver chloride.

The salts associated with cured meats, nitrates and nitrites, are assayed by aqueous extraction and colorimetric determination. The ISO methods involve extraction in hot water containing borax solution and deproteinization of a portion of the extract with Carrez reagents. Nitrite (ISO 2918: 1975) is determined directly by formation of an azo dye with sulfanilamide and *N*-(1-naphthyl)ethylenediamine dihydrochloride which is measured at 538 nm. Nitrate (including nitrite) (ISO 3091: 1975) is determined by cadmium reduction to nitrite and azo dye formation. Reduction of nitrate to nitrite may be carried out on a special column (as in the ISO method) or using 'spongy' cadmium. Care must be taken to prevent interference from ascorbic acid if it is present in the sample at relatively high levels ( $> 20 \mu\text{g ml}^{-1}$ ).

Other methods for nitrates and nitrites include the AOAC procedure in which *m*-xylenol is nitrated and the nitro-xylenol is distilled into sodium hydroxide solution to form a colored sodium salt that is determined with ultraviolet spectrophotometry at 450 nm. Gas chromatography with electron capture detection and high-performance liquid chromatography have been used for determination of nitrate and nitrite derivatives.

Of particular interest in cooked cured meat products is the determination of nitrosamines,

carcinogenic products resulting from the reaction of secondary amines with the nitrites used for curing. The AOAC official method is based on distillation under vacuum of the nitrosamines from the sample in mineral oil into a trap. The trapped nitrosamines are collected in methylene chloride, concentrated by evaporation, and determined by gas chromatographic analysis using a thermal energy analyzer. Recently, a method was approved by the AOAC for the following three volatile *N*-nitrosamines in minced meat: *N*-nitrosodimethylamine, *N*-nitrosopyrrolidine, and *N*-nitrosomorpholine. In this method, the sample is mixed with anhydrous sodium sulfate and Celite, packed into a glass chromatographic column, and eluted with pentane-dichloromethane. The eluate passes through a second column containing acid Celite, on which the nitrosamines are trapped. The nitrosamines are eluted from this second column with dichloromethane, the eluate is concentrated, and the nitrosamines determined by gas chromatographic analysis using a thermal energy analyzer.

Phosphate and polyphosphates are determined as total phosphorus in the ISO method (ISO 2294: 1974). The sample is digested with nitric and sulfuric acids and the orthophosphate is precipitated as quinoline 12-molybdophosphate. The precipitate is collected on a glass filter, dried, and weighed and expressed as phosphorus pentoxide. A similar procedure described by the AOAC is ashing of the sample at  $550^\circ\text{C}$ , boiling of the ash in dilute nitric acid and filtration before precipitation of quinoline 12-molybdophosphate.

An alternative approach to these gravimetric methods is determination procedure based on the formation of colored complexes. Orthophosphate in acidic solution reacts with molybdic acid and vanadic acid to form yellow-orange vanadomolybdophosphoric acid which has maximum absorption at 330 nm. Another approach is a two-stage reaction in which yellow phosphomolybdate is produced from the addition of ammonium molybdate to the orthophosphate in acidic solution, and the phosphomolybdate is then reduced by ascorbic acid to a molybdenum blue, which is measured at 890 nm. This last method is the basis for an AOAC-approved spectrophotometric method.

In some countries, sulfur dioxide is used as a preservative in meat products. Sulfur dioxide occurs in many forms in food but quantitative tests for the material in meat products are normally for total (free plus bound) sulfur dioxide. The AOAC method applied to meat products consists of liberation of sulfur dioxide from the sample by heating with hydrochloric acid, distillation into hydrogen peroxide solution,

which then reacts with the sulfur dioxide to form sulfuric acid, and titration with sodium hydroxide using methyl red indicator. Gravimetric determination is also possible by addition of barium chloride to the titrated liquid resulting in a precipitate of barium sulfate, which can be isolated by filtration, dried, and weighed. Special apparatus for the AOAC procedures (the modified Monier-Williams method and the optimized Monier-Williams method) have been developed.

An alternative to the hydrogen peroxide/sulfuric acid titration with alkali is distillation of sulfur dioxide from the acidified sample into excess iodine solution. A proportion of the iodine, equivalent to the amount of sulfur dioxide, is reduced to iodide, and the residual iodine is titrated with sodium thiosulfate using starch as indicator. This more rapid method is suitable for meat products where other volatile sulfur compounds do not interfere, and it has been developed for use with automated distillation systems.

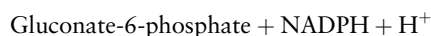
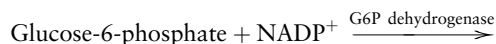
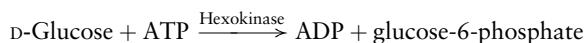
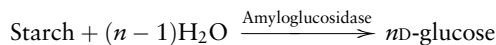
Another AOAC approved method for sulfites in meat is the qualitative test based on decolorization of malachite green solution when mixed with a sample containing sulfites. This may be used to screen samples prior to quantitative determination as described above.

Two other methods have been approved by AOAC for determination of sulfites in food. One is a quantitative assay based on malachite green decolorization using flow injection analysis. Sulfite is released from a sample slurry with alkali, then the test stream is acidified to produce sulfur dioxide gas which diffuses across a Teflon membrane into a flowing stream of malachite green, and the extent of decolorization is measured at 615 nm. The other method is based on ion exclusion chromatography with sulfur dioxide being released by alkali extraction, and the diluted filtrate is injected onto an anion exclusion column linked to an electrochemical detector.

The presence of added starch in meat products is determined by a titrimetric method for reducing sugars. The ISO (ISO 5554: 1978) and AOAC describe methods for starch which involve extraction with water, clarification with Carrez reagents, and acid hydrolysis of the isolated starch. Determination of the resultant reducing sugars is by reaction with Fehling's solution (containing copper sulfate), oxidation of potassium iodide with the excess copper(II), and titration of the liberated iodine with thiosulfate, according to the Luff-Schoorl method.

Starch, and individual sugars, may be determined in meat products using enzymatic kit methods. Starch is extracted from the sample and hydrolyzed by the enzyme amyloglucosidase to D-glucose. The

quantity of D-glucose is determined by a two-stage reaction involving hexokinase and dehydrogenase enzymes. The specific glucose-6-phosphate dehydrogenase utilizes the coenzyme nicotinamide adenine dinucleotide phosphate (NADP). The reaction of the enzyme with the sugar involves reduction of NADP to NADPH, which is measured spectrometrically at 340 nm:



## Metals

Analyses for metals in meat and meat products are carried out by atomic spectrometry, generally after dry-ashing and solubilization of the ash in acid. Sodium and potassium, being present at relatively high levels in meats, are determined by atomic emission spectrometry, while other metals, such as cadmium, copper, iron, lead, and zinc, are determined by atomic absorption spectrometry. An alternative method applicable to most metals is inductively coupled plasma atomic emission spectrometry.

The AOAC method for arsenic involves ashing at 600°C, dissolving the ash in dilute hydrochloric acid with the addition of metallic zinc, and the arsenic is distilled as arsine (AsH<sub>3</sub>) into an iodine solution. Ammonium molybdate is added to the solution and, through a series of reactions assisted by heating, molybdenum blue is formed which is determined spectrophotometrically at 840 nm.

The determination of calcium in meat products, especially in mechanically recovered (or separated) meat (MRM/MSM), is an important estimation of the amount of bone material included in the meat. The AOAC method describes an acid digestion of the sample and reaction of an aliquot of the filtered digest with excess ethylenediaminetetraacetic acid (EDTA) under alkaline conditions to form a chelated complex with calcium ion. The excess EDTA is then titrated with calcium carbonate using hydroxynaphthol blue as indicator.

## Kit Methods for Additives and Metals

A variety of qualitative or semiquantitative kit methods based on indicating test strips (e.g., Merckoquant<sup>®</sup>), color comparators (e.g., Aquaquant<sup>®</sup>, Microquant<sup>®</sup>), and photometric determinations (e.g., Spectroquant<sup>®</sup>) are available. These kit methods



may be used for onsite measurements or for screening assays. Typical analytes measured are nitrite/nitrate, sulfite, phosphate, and metals.

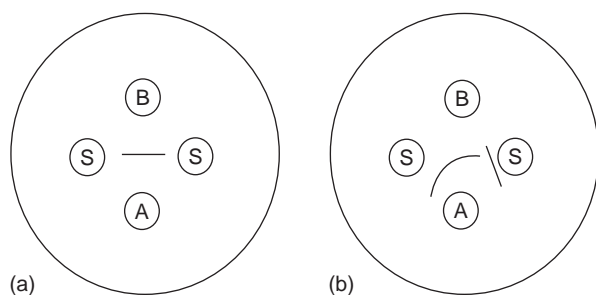
## Analyses for Adulteration

### Meat Species Identification

The determination of the animal species contributing to the meat(s) in a meat product is important for marketing purposes. It may be necessary, in some cases, to ensure the absence of meat of a particular species, such as pork. A number of methods for meat species identification are used which are based on immunological antigen–antibody reactions, on protein isolation techniques, such as electrophoresis, or on DNA analysis.

The AOAC methods are based on agar gel immunodiffusion tests for beef and for poultry meat in meat products. These tests use stabilized reagent paper disks containing beef (or poultry) antibody and beef (or poultry) antigen. The samples are applied to the agar plate by absorbing fluids on to paper disks, which are then placed on the plate. Diffusion of antigen and antibody components occurs during incubation for 18–24 h and the immunoprecipitin lines are examined for presence of beef (or poultry) in the samples (**Figure 2**). These tests are limited in that they are suitable only where the adulterant is present at  $\geq 10\%$ .

Agar gel immunodiffusion methods have been produced as commercial kits and a further development is dipstick assays for meat species identification. The immunoreagents are immobilized on a dipstick and a color change, occurring in less than 1 h, identifies the presence of a particular meat species, with a detection limit of  $\sim 1\%$  lean meat.



**Figure 2** Agar immunodiffusion technique. (a) Illustrates the reference line formed between beef reference antigen B and anti-beef antibody A and two negative samples S. (b) Illustrates a positive beef sample on the left, shown by complete fusion of its immunoprecipitin line with the reference line. The presence of the nonspecific line shows that the sample on the right is negative for beef. (Reprinted with permission from *Official Methods of Analysis* (2000) 17th edn., AOAC International, Gaithersburg, MD, Sec. 39.1.36, AOAC International.)

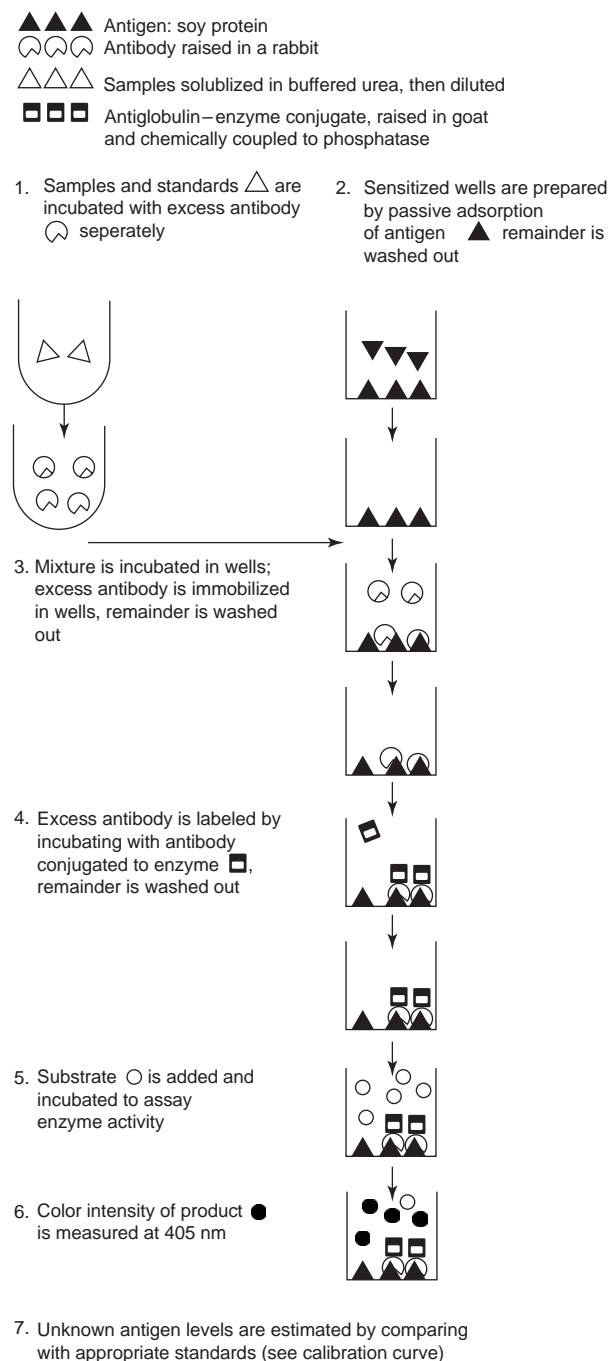
Enzyme-linked immunosorbent assays (ELISAs) in kit form are most widely used giving relatively rapid and inexpensive methods for multispecies identification. A typical format for such an ELISA is to coat different strips of the normal  $12 \times 8$ -well plate with antisera formed against serum albumin of the various species of interest. An extract of the meat product is added to the antibody-coated wells, incubated to ensure antibody binding of the serum albumin, and, after washing, a second antibody coupled with enzyme is introduced. The ‘sandwich’ is visualized by addition of a substrate to the enzyme. ELISAs have been developed, also, for meat species identification in cooked meat products. These ELISAs are quite specific and sensitive ( $\sim 1\%$  of each species can be detected) but are qualitative, or at best, semiquantitative.

Methods for meat species identification based on DNA analysis benefit from the heat stability of the DNA molecule and its high specificity. Originally, DNA methods consisted of immobilization of partially purified and denatured DNA, extracted from the meat product sample, on a nylon membrane, followed by hybridization of a species-specific segment of labeled (colorimetric, fluorescent, or chemiluminescent) DNA with any complementary sequences of DNA present on the membrane. More recently, a DNA amplification method – the polymerase chain reaction – has been used, but this is a relatively expensive and technically demanding technique.

### Nonmeat Proteins

Assays to detect the presence of nonmeat proteins, such as soya protein, are based on similar antigen–antibody methodologies as are used for meat species identification. An AOAC method for soya protein in raw and heat-processed meat products uses an indirect ELISA technique. The soya protein is extracted from the sample as acetone powder and renatured by dilution with buffer. The soya protein reacts with an excess of antibody and the unbound antibody is immobilized in wells that have been coated with antigen. A second antibody conjugated to enzyme labels the immobilized antibody and substrate is added to give a color reaction, the extent of which is inversely related to the amount of soya protein extracted from the sample (**Figure 3**). This procedure is very lengthy and test kits utilizing a direct ELISA give more rapid results. Qualitative microscopic methods are also used to detect the presence of soya in meat products.

Skim milk powder or milk proteins may be used in meat products. The presence of milk powder is



**Figure 3** Schematic diagram of soya protein determination by the ELISA procedure. (Reprinted with permission from Official Methods of Analysis (2000) 17th edn., AOAC International, Gaithersburg, MD, Sec. 39.1.3; © AOAC International.)

determined by a quantitative assay for the distinctive milk sugar, lactose. The AOAC method describes the determination of lactose as a reducing sugar using Benedict solution. Prior to the determination of lactose, all other reducing sugars in the sample are fermented with Baker's yeast. Lactose may be determined directly in a procedure that involves hot-water

extraction of the sugar, freeze-drying to remove the water, formation of a silyl derivative of the lactose followed by determination using gas–liquid chromatography with flame ionization detection. Alternative methods for direct determination of the milk proteins in meat products are based on enzyme immunoassays or electrophoresis.

## Assessment of Spoilage

A number of methods are used to determine spoilage of meat. These methods are directed at measuring changes in protein and fat that indicate deterioration due to spoilage. Protein breakdown is determined through measurement of TVBN. The meat sample is extracted with trichloroacetic acid, the extract made alkaline with sodium hydroxide, the volatile bases distilled into standard acid, and the TVBN determined by back-titration with standard alkali.

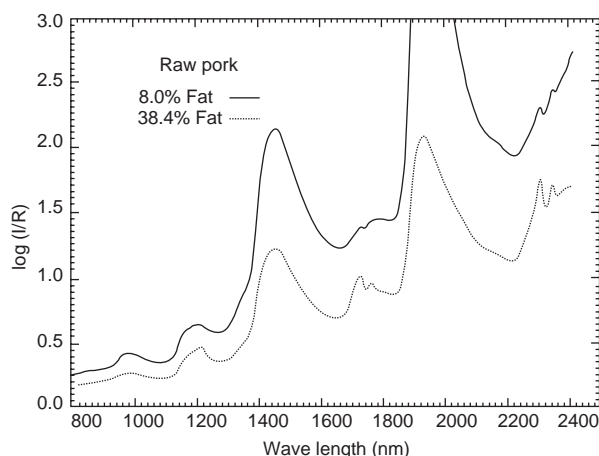
Deterioration of fat causes rancidity of which there are two types – hydrolytic rancidity, caused by a combination of microorganisms and moisture, and oxidative rancidity, caused by oxygen interacting with unsaturated fatty acids. Hydrolytic rancidity is measured as the quantity of free fatty acids (FFA) in the fat; the fat is extracted from the meat into chloroform, a portion of the extract is mixed with neutral alcohol, and the FFA titrated directly with alkali (using phenolphthalein as indicator). A number of methods are available to measure oxidative rancidity including peroxide value (PV), TBA value, and anisidine value (AnV). The PV is determined on a chloroform extract by measuring the oxidation of potassium iodide to iodine through titration with thiosulfate. TBA and AnV are determined as colorimetric reactions between aldehydes and 2-thiobarbituric acid and between carbonyl compounds and *p*-anisidine, respectively.

The suitability of meat is established by ensuring that the values for TVBN, FFA, PV, TBA, and AnV are below specified levels which have been established as correlating with off-flavors in the meat.

## Limitations of Current Analytical Procedures

Analysis of meat using standard methods is relatively slow and does not give results that can be used to influence the manufacturing process. Major developments in meat analysis technology have occurred in both offline and online systems that give rapid analysis capability.

Systems based on microwave energy have been developed for determination of moisture, fat, and

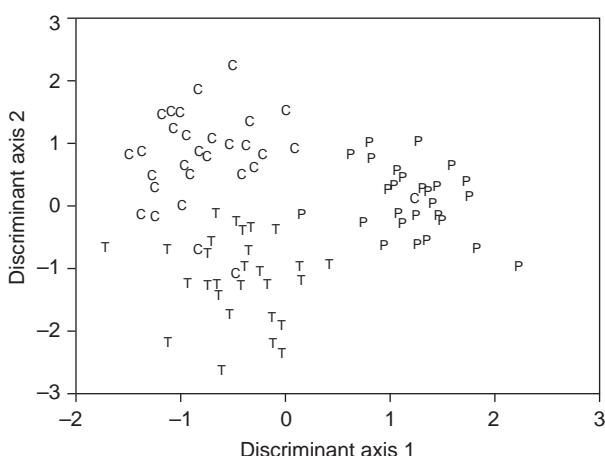


**Figure 4** Spectra of homogenized raw pork at two levels of fat. (From Norris KH (1984) Reflectance spectroscopy. In: Stewart KK and Whitaker JR (eds.) *Modern Methods of Food Analysis*, pp. 167–186. Westport, CT: AVI Publishing.)

protein in meat. Automated systems are available that determine moisture and fat content gravimetrically and calculate protein using a factor, giving results within 5–10 min. Microwave energy may be used, also, to give rapid Kjeldahl digestion (10 min) for direct determination of crude protein.

A very promising technology for both offline and online applications is near-infrared (NIR) spectroscopy. NIR radiation energy causes bending or stretching of bonds such as  $-C-H$ ,  $-O-H$ , and  $-N-H$  to produce absorbance spectra (Figure 4). The absorption of energy at wavelengths in the NIR region (750–2500 nm) can be used to analyze for constituents of meat and meat products. A requirement for use of NIR analysis is the development of calibrations for the analytes of interest, involving chemical analysis of many ( $\geq 100$ ) samples representative of the ranges of analytes and variations in the product. Successful applications of NIR technology in meat analysis have been done for moisture, protein, and fat. Salt has been determined by NIR in meat products based on a displacement of the water absorbance peak at 1806 nm. The particular advantage of NIR analysis is that all analytes may be determined simultaneously. Limitations in the use of NIR analysis are the time and expense required for the development of robust calibrations and the high product specificity of calibrations, which may limit the capability of the technology to give accurate results when there are changes in product composition.

A particular attraction of NIR analysis is the possibility of using it online to monitor meat constituents continuously. Fiber optic probe technology has been successfully adapted to deliver the incident energy and record the reflected energy from the sample. A



**Figure 5** Discriminant scores plot for homogenized chicken (C), turkey (T), and pork (P) meats. (From Downey G (2003) The National Food Centre (Teagasc), personal communication.)

higher energy wavelength range of 750–1100 nm is often used to achieve sufficient penetration of energy into intact meat surfaces for nondestructive analysis online. Apart from quantitative compositional analysis, NIR has been applied to qualitative analysis, to predict tenderness of beef carcasses, meat quality of pork carcasses, and to confirm the animal species of comminuted meat (Figure 5).

A variety of technologies are being applied to online analysis of carcasses and meat cuts, total body electrical conductivity (TOBEC), video image analysis (VIA), and ultrasound scanning. TOBEC measures the interference produced in an electromagnetic field by the carcass/meat cut and this is related to moisture content and, indirectly, to lean meat content. TOBEC has been used for online measurement of the lean meat content of boxed boneless beef for over a decade and is being used also for the grading of pork carcasses and hams. VIA has been used for determining the fat/lean content of boxed beef and has been applied also to grading beef, pork, lamb, and turkey carcasses. VIA is being developed to predict the eating quality of beef. Ultrasound scanning has also been applied to automate pig carcass grading. For example, the AUTOFOM<sup>TM</sup> equipment may be more accurate at predicting the lean content of carcasses than are conventional grading probes used on the slaughterline, and the equipment gives information also on the leanness of individual cuts.

**See also:** Atomic Absorption Spectrometry: Principles and Instrumentation. Atomic Emission Spectrometry: Principles and Instrumentation. Carbohydrates: Starch. Clinical Analysis: Glucose. Extraction: Solvent Extraction Principles. Food and Nutritional Analysis: Dairy Products; Oils and Fats. Gravimetry. Immunoassays,

**Applications:** Food. **Immunoassays, Techniques:** Enzyme Immunoassays. **Infrared Spectroscopy:** Near-Infrared. **Lipids:** Fatty Acids. **Liquid Chromatography:** Food Applications. **Microscopy Applications:** Food. **Nitrogen. Nitrosamines. pH. Phosphorus. Proteins:** Foods. **Sampling:** Theory; Practice. **Sulfur. Water Determination. X-Ray Absorption and Diffraction:** X-Ray Absorption.

## Further Reading

- Baltes W (1990) *Rapid Methods for Analysis of Food and Food Raw Material*. Hamburg: Behr's Verlag.
- Bauer F (1991) Microwave digestion of proteins for rapid determination of hydroxyproline in meat products. In: Baltes W, Eklund T, Fenwick R, *et al.* (eds.) *Strategies for Food Quality Control and Analytical Methods in Europe*, vol. II, pp. 552–556. Hamburg: Behr's Verlag.
- Bauer F (1996) Chemical analysis to monitor the quality of meat and meat products. In: Taylor SA, Raimundo A, Severini M, and Smulders FJM (eds.) *Meat Quality and Meat Packaging*, pp. 183–194. Utrecht: ECCEAMST.

- Kirk RS and Sawyer R (1991) *Pearson's Composition and Analysis of foods*, 9th edn. Harlow: Longman Scientific & Technical.
- Lawrie RA (1998) *Lawrie's Meat Science*, 6th edn. Cambridge: Woodhead Publishing Ltd.
- Lumley ID (1996) Authenticity of meat and meat products. In: Ashurst PR and Dennis MJ (eds.) *Food Authentication*, pp. 108–139. London: Blackie.
- Pomeranz Y and Meloan CE (1994) *Food Analysis: Theory and Practice*, 3rd edn. New York: Chapman Hall.
- Savell JW (2000) *Meat Science Laboratory Manual*, 7th edn. Boston: American Press.
- Sebranek JG (1998) Rapid methods for compositional analyses of meat and meat products. In: Tunick MH (ed.) *New Techniques in the Analysis of Foods*, pp. 161–169. New York: Kluwer Academic Publishers.
- Swatland HJ (1995) *On-line Evaluation of Meat*. Lancaster: Technomic Publishing Co.
- Young OA, Frost DA, West J, and Braggins TJ (2001) Analytical Methods. In: Hui YH, Nip W-K, Rogers RW, and Young OA (eds.) *Meat Science and Applications*, pp. 104–126. New York: Dekker.

## Dairy Products

**R J Marshall**, London Metropolitan University, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Many of the components in milk and dairy products can be analyzed by standard methods approved by the International Dairy Federation (IDF). In December 2000, the IDF and the International Organization for Standardization (ISO) agreed to publish standards jointly and these are printed, distributed, and sold exclusively by the ISO Secretariat. Many of the methods are also published by the Association of Official Analytical Chemists (AOAC). The World Health Organization (WHO) has approved certain methods. Some of the methods have been used for many years but, particularly in microbiology, there have been recent advances that have improved our ability to determine microorganisms more specifically.

An analysis of dairy products includes the proximates: total solids, protein, fat, energy, ash, acidity, and specific gravity, and the specifics: lactose, sodium, potassium, calcium, copper, chloride, phosphate, citrate, lactose, preservatives and antibiotics, added dyes, detergent residues, organic residues, and

microorganisms. In addition, milk itself may be analyzed for freezing point and dirt. Methods can be used for different products with a little adaptation in most cases.

## Physical Analysis of Milk

### Freezing Point

Water may get into milk from milking machinery, accidentally or fraudulently. Dilution of milk changes its freezing point. The WHO method uses a thermostatically controlled water bath cooled by electrical refrigeration and a thermistor probe. There are two types of instrument: the first determines a plateau in the freezing curve and the other, used for routine screening, reads at a fixed time from the start of freezing.

In the approved method, supercooled milk is induced to freeze by mechanical vibration that causes the temperature to rise to a plateau corresponding to the freezing point. Standard solutions of sodium chloride are used for calibration. The method gives excellent agreement between different laboratories testing the same sample of milk ( $<0-005^{\circ}\text{C}$ ). The addition of 1% water in milk raises the freezing point by about  $\sim T/100$ , where  $T$  is base freezing point of authenticated samples. Dietary, daily,

**Applications:** Food. **Immunoassays, Techniques:** Enzyme Immunoassays. **Infrared Spectroscopy:** Near-Infrared. **Lipids:** Fatty Acids. **Liquid Chromatography:** Food Applications. **Microscopy Applications:** Food. **Nitrogen. Nitrosamines. pH. Phosphorus. Proteins:** Foods. **Sampling:** Theory; Practice. **Sulfur. Water Determination. X-Ray Absorption and Diffraction:** X-Ray Absorption.

## Further Reading

- Baltes W (1990) *Rapid Methods for Analysis of Food and Food Raw Material*. Hamburg: Behr's Verlag.
- Bauer F (1991) Microwave digestion of proteins for rapid determination of hydroxyproline in meat products. In: Baltes W, Eklund T, Fenwick R, *et al.* (eds.) *Strategies for Food Quality Control and Analytical Methods in Europe*, vol. II, pp. 552–556. Hamburg: Behr's Verlag.
- Bauer F (1996) Chemical analysis to monitor the quality of meat and meat products. In: Taylor SA, Raimundo A, Severini M, and Smulders FJM (eds.) *Meat Quality and Meat Packaging*, pp. 183–194. Utrecht: ECCEAMST.

- Kirk RS and Sawyer R (1991) *Pearson's Composition and Analysis of foods*, 9th edn. Harlow: Longman Scientific & Technical.
- Lawrie RA (1998) *Lawrie's Meat Science*, 6th edn. Cambridge: Woodhead Publishing Ltd.
- Lumley ID (1996) Authenticity of meat and meat products. In: Ashurst PR and Dennis MJ (eds.) *Food Authentication*, pp. 108–139. London: Blackie.
- Pomeranz Y and Meloan CE (1994) *Food Analysis: Theory and Practice*, 3rd edn. New York: Chapman Hall.
- Savell JW (2000) *Meat Science Laboratory Manual*, 7th edn. Boston: American Press.
- Sebranek JG (1998) Rapid methods for compositional analyses of meat and meat products. In: Tunick MH (ed.) *New Techniques in the Analysis of Foods*, pp. 161–169. New York: Kluwer Academic Publishers.
- Swatland HJ (1995) *On-line Evaluation of Meat*. Lancaster: Technomic Publishing Co.
- Young OA, Frost DA, West J, and Braggins TJ (2001) Analytical Methods. In: Hui YH, Nip W-K, Rogers RW, and Young OA (eds.) *Meat Science and Applications*, pp. 104–126. New York: Dekker.

## Dairy Products

**R J Marshall**, London Metropolitan University, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Many of the components in milk and dairy products can be analyzed by standard methods approved by the International Dairy Federation (IDF). In December 2000, the IDF and the International Organization for Standardization (ISO) agreed to publish standards jointly and these are printed, distributed, and sold exclusively by the ISO Secretariat. Many of the methods are also published by the Association of Official Analytical Chemists (AOAC). The World Health Organization (WHO) has approved certain methods. Some of the methods have been used for many years but, particularly in microbiology, there have been recent advances that have improved our ability to determine microorganisms more specifically.

An analysis of dairy products includes the proximates: total solids, protein, fat, energy, ash, acidity, and specific gravity, and the specifics: lactose, sodium, potassium, calcium, copper, chloride, phosphate, citrate, lactose, preservatives and antibiotics, added dyes, detergent residues, organic residues, and

microorganisms. In addition, milk itself may be analyzed for freezing point and dirt. Methods can be used for different products with a little adaptation in most cases.

## Physical Analysis of Milk

### Freezing Point

Water may get into milk from milking machinery, accidentally or fraudulently. Dilution of milk changes its freezing point. The WHO method uses a thermostatically controlled water bath cooled by electrical refrigeration and a thermistor probe. There are two types of instrument: the first determines a plateau in the freezing curve and the other, used for routine screening, reads at a fixed time from the start of freezing.

In the approved method, supercooled milk is induced to freeze by mechanical vibration that causes the temperature to rise to a plateau corresponding to the freezing point. Standard solutions of sodium chloride are used for calibration. The method gives excellent agreement between different laboratories testing the same sample of milk ( $<0-005^{\circ}\text{C}$ ). The addition of 1% water in milk raises the freezing point by about  $\sim T/100$ , where  $T$  is base freezing point of authenticated samples. Dietary, daily,



and seasonal variations in the composition of milk affect the freezing point but the osmotic pressure does not change very much, so the freezing point normally remains within the range  $-0.530^{\circ}\text{C}$  to  $-0.570^{\circ}\text{C}$  (mean  $-0.540^{\circ}\text{C}$ ).

### Specific Gravity

The specific gravity of milk can be used to determine the total solids (TS) content and is usually measured by either a standard hydrometer or with a pycnometer (density bottle) after gentle mixing. The effect of different temperatures can be corrected using a table. The TS of the milk can be calculated if the fat content is known by using Richmond's formula:

$$\text{TS, } T(\%) = 0.25G + 1.2F + 0.72$$

where  $G$  is the lactometer reading at  $20^{\circ}\text{C}$  and  $F$  is the percentage of fat present. Milk has to be aged for several hours and chilled.

### Chemical Analysis

Individual methods are available for the determination of the proximates (TS, fat, protein, ash, etc.) in milk and dairy products. However, many laboratories use an infrared analyzer that measures the protein, fat, and lactose content of milk simultaneously and almost instantly. The latest instruments can analyze proximates in solid and semisolid dairy products in 50 s (Figure 1).

These instruments use near-infrared transmittance to give rapid results. For fat, infrared analyzers measure the radiation absorbed by the carbonyl groups of ester bonds in the acylglycerols (at  $\sim 5.73\text{ }\mu\text{m}$ ) or by the  $-\text{CH}$  groups (at  $\sim 3.48\text{ }\mu\text{m}$ ). For protein, the absorption at  $6.46\text{ }\mu\text{m}$  by secondary amide groups is

measured and for lactose, the absorption at  $9.61\text{ }\mu\text{m}$  by the hydroxyl groups.

### Fat

The standard method for determining fat in milk is the Gerber method. A measured quantity of milk is added to a standard amount of concentrated sulfuric acid in a special test bottle (butyrimeter). Isoamyl alcohol is added to extract the fat and then the bottle is centrifuged to separate the aqueous and fat phases. The fat phase separates into the graduated neck of the bottle and the percentage of the total volume may be read off.

The Babcock method uses a tube that is much larger than the Gerber bottle. After adding concentrated sulfuric acid and centrifuging, water is added to bring the fat column into the graduated neck of the bottle.

The Rose–Gottlieb method is also used. This method gives excellent results but requires very careful attention to detail to achieve the best accuracy. The sample is treated with concentrated ammonia and then the fat is extracted using organic solvents. Finally, the extract is dried, weighed, and expressed as a percentage of the original weight of sample.

The Werner–Schmid method is useful for sour milk and other milk-based foods. This method uses ammonia (relative density 0.88) in which the sample is dispersed followed by the addition of hydrochloric acid. The casein is dissolved by heating in boiling water. The fat is then extracted with diethyl ether, dried, and weighed.

The same methods can be used for other dairy products, cheese, cream, etc. There are different butyrimeters for the different products that allow for the different ranges of fat contents.

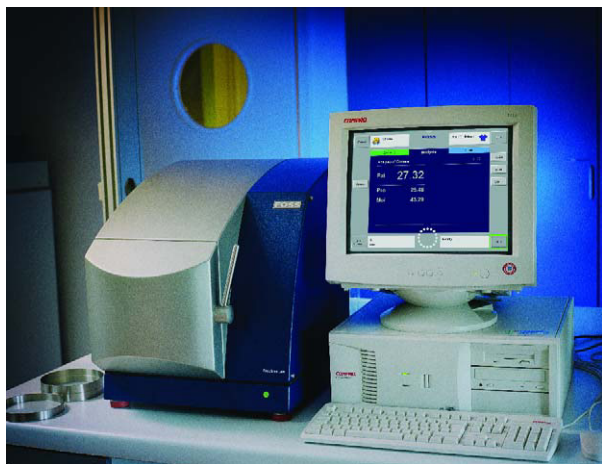
### Total Solids

The TS content is the mass remaining after complete removal of water and volatiles. In the standard method, a weighed sample of milk is dried under controlled conditions to constant weight. The TS is expressed as a percentage by mass. The TS of cheese is measured in the same way but washed, dried sand is placed in the drying tin and used to break up the sample into small pieces. This prevents spitting during drying which would cause losses.

The TS and thus the nonfatty solids (solids-not-fat, SNF) can also be calculated after determination of the density and fat contents.

### Protein

A number of different methods are commonly used to determine the protein content.



**Figure 1** FOSS FoodScan Lab Infra-red analyser (From *In Focus* (2002), vol. 26(1), p. 13, with permission).



In the classical Kjeldahl method, the proteins are 'digested' (wet oxidized) in sulfuric acid with a catalyst (mercury and selenium tablets now succeed by the much safer potassium and copper sulfate tablets). An acid solution of ammonium sulfate is formed which is then diluted in water. The solution is made alkaline with sodium hydroxide and heated to distil off ammonia into excess standard acid sulfuric acid. The excess acid is back-titrated with standard sodium hydroxide to determine the amount of ammonia. It is more usual now to use boric acid in which to collect the ammonia and titrate with standard hydrochloric acid. The Kjeldahl procedure has been partially automated in systems such as the Kjeltec Analyzer. Total protein can be calculated as nitrogen content  $\times 6.38$ .

The total protein, casein protein, and noncasein protein are determined by measuring total nitrogen (TN) and noncasein nitrogen (NCN) after precipitation of the casein fraction at its isoelectric point by acetic acid-sodium acetate. Casein N = TN - NCN.

Dye-binding methods use the ability of dyes to bind to charged amino acid residues of the proteins. The proteins may be separated as previously described and then treated with the dye. Dyes used include amido black and orange G. On binding to the protein they form an insoluble complex that may be removed by centrifugation. The amount of residual color in the supernatant is measured by spectrophotometry and calibrated by reference to the Kjeldahl method.

### **Titrateable Acidity and pH**

The acidity of milk, expressed as percentage of lactic acid, is normally measured by titration of milk diluted to two times its own volume with carbon dioxide-free water against  $0.111 \text{ mol l}^{-1}$  sodium hydroxide using phenolphthalein indicator. One milliliter of sodium hydroxide is equivalent to 0.0010 g lactic acid.

The pH of milk is measured directly using a pH meter. Care is needed to keep the electrode free of fat which may block the porous plug and give false readings.

The titrateable acidity and pH of other dairy products, such as cheese and during cheesemaking, are measured in the same way. For cheese, it is usual to make a slurry of 10 g cheese in 10 ml distilled water and measure the acidity or pH.

### **Lactic Acid**

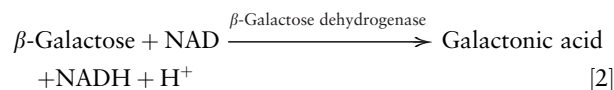
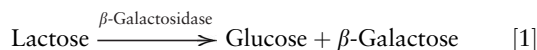
In the standard method for lactic acid in milk, the proteins are removed by precipitation with phosphotungstic acid and filtration. The filtrate is acidified with sulfuric acid and lactic acid extracted in a

special extractor using ether. After adding water, the ether is evaporated off and the remaining solution neutralized. After removal of the ethanol, used to wash the solution, it is treated with ferric chloride in hydrochloric acid and a blue color develops which may be quantified at  $\sim 365 \text{ nm}$  using a spectrophotometer. Gas-liquid chromatography (GLC) and enzymatic methods are also used.

### **Lactose**

Several methods are available to measure lactose content. As lactose is optically active and rotates the plane of polarized light in proportion to its concentration, polarimetry may be used to quantify it. The proteins are first removed by precipitating them with phosphotungstic acid. This method is much less used now.

Lactose can also be measured enzymatically by hydrolysis in the presence of nicotinamide adenine dinucleotide (NAD) which is reduced to NADH (see eqns [1] and [2]). The amount formed is determined spectrophotometrically at 340 nm:



The NADH is produced in stoichiometric amounts and has an absorption coefficient of  $6.3 \text{ m mol}^{-1} \text{ cm}^{-1}$  at 340 nm. Kits are available that contain all the reagents and instructions (for example, see: <http://www.r-biopharm.com/>. AOAC International (Association of Official Analytical Chemists) also has an index of rapid test kits: <http://www.aoac.org/testkits/tkdata8.htm>).

The Lane and Eynon titration, which uses Fehling's solution (equal parts of Fehling's 1 and 2; Fehling's 1: 69.3 g copper sulfate pentahydrate in water to 1 l; Fehling's 2: 100 g sodium hydroxide, 345 g sodium potassium tartrate in water to 1 l), can also be used for the quantitative determination of lactose (and other reducing sugars). The sample has to be cleared first using standard reagents for precipitating the protein. Then the clarified sample is run from a burette into boiling Fehling's solution containing methylene blue indicator. The Fehling's is standardized using standard reducing sugar solution in the burette.

### **Ash**

Ash is the residue after incineration at  $<550^\circ\text{C}$  and comes from both the organic and inorganic

constituents. Chlorides are easily volatilized if the ashing temperature is too high. Citrate is lost completely. Samples are weighed into acid-washed crucibles that are then placed in a muffle furnace. Prior to combustion, they can be heated gently over a flame taking care to avoid inflaming the sample, or by infrared lamp, to dryness.

### Individual Proteins and Free Amino Acids

The caseins are solubilized with urea and separated by electrophoresis on polyacrylamide slab gels containing urea to prevent them from aggregating. Standards of purified caseins are used for quantification. However, this method is not very satisfactory. The whey proteins ( $\beta$ -lactoglobulin and  $\alpha$ -lactalbumin) are determined in casein-free extracts by polyacrylamide slab or disk electrophoresis. Proteins in milk may also be analyzed by fast protein liquid chromatography in which they are separated on a short column of ion exchange resin (e.g., Mono Q anion exchange column, Pharmacia Fine Chemicals, Sweden), eluted, and detected by their adsorption of ultraviolet (UV) light.

Individual amino acids can be determined in solubilized extracts using high performance liquid chromatography (HPLC).

### Minerals

Metallic cations in milk and dairy products (including sodium, potassium, calcium, and heavy metal contamination) may be measured by flame photometry ( $\text{Na}^+$ ,  $\text{K}^+$ ) or atomic adsorption spectrometry (AAS) (other elements) in which the sample is atomized in the gas flame and the adsorption is measured at the characteristic wavelength of each element.

Samples are 'wet oxidized' in concentrated sulfuric acid plus concentrated nitric acid by heating to  $\sim 400^\circ\text{C}$ . The solutions are then diluted to standard volume and used for subsequent analysis.

### Calcium

Calcium in wet-oxidized samples is determined by AAS. In principle, it can be measured by flame photometry but the lower sensitivity of instruments for the calcium flame limits the usefulness of this method.

The standard method for determining total calcium involves precipitating the proteins with trichloroacetic acid and filtering to obtain the soluble calcium. This is converted to the oxalate by adding saturated ammonium oxalate and precipitated by adding acetic acid. The calcium oxalate is washed into water, acidified with sulfuric acid, and heated over a boiling water bath to dissolve it. The hot solution ( $>60^\circ\text{C}$ )

is titrated with potassium permanganate to a first permanent pink color.

### Chloride

Chloride may be determined by the Volhard method. The sample is acidified with concentrated nitric acid and silver nitrate is added, which reacts with the chloride, and then the mixture is boiled. After cooling, the pale yellow solution is diluted and the excess silver nitrate titrated with potassium thiocyanate solution using ammonium ferric sulfate as indicator. The ash may also be used for this determination.

### Phosphate and Other Salts

Phosphate and other salts may be measured by various titrimetric and colorimetric methods. Phosphate is measured titrimetrically by precipitation of quinoline phosphomolybdate that is collected and dissolved in a small known excess of alkali and then back-titrated with standard acid. A colorimetric method based on the reaction of the phosphate with ammonium molybdate followed by partial reduction to give 'molybdenum blue' has been developed for use in auto-analyzers.

### Energy

The energy content can be measured by combustion of the residue after drying in a ballistic bomb calorimeter. However, this gives the total energy in the sample, which is not the same as the energy available for digestion. This can be approximated by calculating the metabolizable energy contribution of the fat, protein, and carbohydrate.

## Contaminants

### Contamination of Milk by Soil and Other Dirt

The standard method for determining contamination of milk by soil particles and similar dirt is in two forms. The reference method involves centrifuging after standing for 72 h and so is not used routinely. The rapid method uses a filtration technique that takes  $\sim 3$  min. The amount of visible dirt on the filter is compared visually with standard photographic disks and gives values in milligrams per liter.

### Microorganisms

Microbial contamination of milk is measured by culturing the microorganisms on suitable growth media. The most effective method is to make appropriate dilutions of milk, or suspensions of dairy products in saline or Ringer's solution, which are spread over the surface of nutrient agar in sterile Petri dishes. The

agar containing nutrients (purified meat extracts, salts, and a carbohydrate source such as glucose), is melted, allowed to cool somewhat, and poured into the Petri dishes. These are allowed to cool to set the agar into a gel. After spreading the samples onto plates they are incubated at a suitable temperature (e.g., 30°C) for 2–3 days. Colonies of microorganisms are counted by placing the dish over an indirect light source and marking each one on the outside of the dish with a felt-tipped pen. This process can be automated using image analysis equipment.

Total viable microorganisms are measured by culturing them in a complete growth medium that provides nutrients for all types (e.g., milk plate count agar). Specific types of microorganisms are determined by using media that promote their growth but retard or inhibit others. For example, 'selective' media such as violet red bile agar are used to measure total coliforms (indicator bacteria originating either in the gut of animals or in the environment). Yeasts and molds are not of importance in milk but may occur in yogurt and high-moisture low-fat spreads. Bacterial growth is inhibited by adding an antibiotic, usually chloramphenicol, to the growth medium. Its not usual practice to examine milk for pathogens but *Campylobacter*, *Listeria monocytogenes* and *Salmonella* can be detected by isolation in various enrichment and selective media before being recognized and confirmed biochemically and serologically. Mold-ripened soft cheeses can be examined for *Listeria monocytogenes* and dried milk may be examined for *Salmonella*. A wide range of rapid methods is also available. The most widely used are those based on molecular or serological techniques. Rapid methods proposed for raw milk include the adenosine triphosphate (ATP)-Limulus (horse-shoe crab) lysate test.

With the need to prevent contaminated milk from being consumed or processed, rapid tests have been developed. The direct epifluorescent filter technique is an accepted ISO test, used routinely by some laboratories. A fluorescent marker dye is attached chemically to the nucleic acid of living and dead cells. Marked microorganisms that are viable fluoresce orange under UV light and are counted automatically by an image analyzer through a microscope.

Bioluminescent methods are used to detect microbiological and somatic cell contamination of milk and milk-processing equipment. They are based on the reaction of ATP with luciferin catalyzed by luciferinase producing light, which can be then be measured in a bioluminometer. The amount of light produced is proportional to the cellular activity. Kits are available from a number of suppliers (e.g., Bio-Trace, Opto-Comp, Hy-Lite, Lumac).

## Psychrotrophs

Psychrotrophs are bacteria that grow at temperatures of 7°C and below. In milk, the main psychrotrophs are pseudomonads with *P. fluorescens* being the commonest, but *Alcaligenes*, *Flavobacterium*, and some coliforms have also been found to be psychrotrophic. They are measured by using a growth medium similar to that described above but incubated at 6.5°C for 10 days. A recent modification to the method allows the number of psychrotrophs to be measured after 25 h at 21°C, but other nonpsychrotrophs may interfere reducing the accuracy of this method. The most effective method of removing this interference is to add inhibitors of gram positive bacteria to milk and preincubate.

Pseudomonads produce extracellular enzymes that breakdown proteins, glycoproteins, fats, glycolipids, and phospholipids in milk. The enzyme linked immunosorbent assay (ELISA) method can be used to measure levels of thermostable psychrotrophic proteases and lipases in milk as a determinant of its suitability for cheesemaking or ultrahigh-temperature processing.

## Microbiological Quality

Most milk for direct sale or for use in the manufacture of dairy products is pasteurized to ensure it is free from pathogens. The quality of raw milk can be measured by the reducing effects of microorganisms on sensitive dyes. Heat treatment of the milk not only destroys potential pathogens but also the milk phosphatase enzyme.

## Dye Reduction Tests

These are not now considered sufficiently reliable because of the possible presence of psychrotrophs that do not have the enzyme systems to reduce these dyes.

**Methylene blue reduction test** This test is used to estimate the keeping quality of pasteurized milk. In principle, a measured quantity of methylene blue solution is added to 10 ml of milk in a tube. This is mixed by slow inversion, then the tube is sealed and incubated in a water bath at 37°C. A satisfactory test is given when there is no decoloration in 30 min.

**Resazurin test** The resazurin test is much quicker than the methylene blue test. Milk is mixed to aerate it and 10 ml placed in a tube. Resazurin dye solution (made from a standard tablet) is added, the tube sealed, inverted twice, and incubated at 37.5°C for 10 min. If actively reducing bacteria are present, the

resazurin dye changes from blue through pink to colorless.

### Phosphatase Activity

The enzyme alkaline phosphatase is naturally present in milk. As it is sensitive to heat it is used to measure the effectiveness of pasteurization. Milk is diluted with a buffer (pH 10.6) containing disodium phenol phosphate which is hydrolyzed by the enzyme releasing phenol. The phenol is reacted with 2,6-dibromoquinonechloroimide producing a color that is measured spectrophotometrically at 610 nm. An earlier version of this test uses *p*-nitrophenyl phosphate and the amount of yellow *p*-nitrophenyl which is released is compared with standard comparator disks.

Alternative enzymes can be measured for enhanced pasteurization, for example,  $\gamma$ -glutamyl transferase.

### Antibiotics and Other Microbial Inhibitors

There are at least 40 antibiotics and sulfonamides that may occur in milk. They can be detected and identified by both microbial and chemical methods.

**Microbial methods** Any inhibitory substance in the milk will reduce the growth of test bacteria (e.g., *Bacillus stearothermophilus*, *Streptococcus thermophilus*, *Bacillus subtilis*, *Sarcina lutea*, *Lactobacillus bulgaricus*, *Bacillus megaterium*, etc.). The disk-assay plate method uses a disk of absorbent paper that is impregnated with the milk sample to be examined and placed on the surface of an agar plate preinoculated with a suitable test organism. After incubation, the growth of the bacteria is seen as turbidity in the agar. The presence of inhibitory substances such as antibiotics which diffuse into the agar is indicated by a clear zone around the edge of the disk. The inhibitor may be identified by adding an agent that blocks its action. For example, the effect of penicillin is prevented by the addition of penicillinase, while *p*-aminobenzoic acid blocks sulfonamides. With such substances in the growth medium, the antibiotic no longer prevents the test organism from growing. Commercially available methods include Intertest, Delvotest, and TTC.

On the farm, a test kit may be used. The test organism is inoculated in tablet form into liquid medium in a test tube with a sample of the milk. An indicator in the buffer changes color if the bacteria are able to grow (i.e., no antibiotic is present). The Charm analyzer can be used for several different tests (Figure 2).



**Figure 2** Charm Sciences ROSA<sup>®</sup> portable incubator/analyzer (Reproduced with permission from Charm Sciences Inc.).

**Chemical methods** Inhibitory substances are extracted from the milk and then detected chemically by thin-layer chromatography, HPLC, fluorescence liquid chromatography, gas chromatography (GC), electrophoresis, bioautoradiography, ELISA, radioimmunoassay, enzymatic methods, and other forms of immunological methods. They are usually more specific than the microbiological methods.

### Preservatives

Although preservatives are not permitted in milk, occasionally they may be present if added accidentally or illegally.

Formaldehyde may be determined qualitatively or quantitatively, although it is only the former that is usually applied to milk. Qualitative tests include Hehner's test in which 90% sulfuric acid containing a trace of ferric chloride is poured gently into the sample. If formaldehyde is present a violet color is formed where the layers meet. This test can be carried out in conjunction with the Gerber test for fat.

Qualitative determination of boric acid is carried out by making the milk just alkaline with lime water, drying, igniting the residue, acidifying with hydrochloric acid, and placing a drop of the liquid onto

turmeric paper. The paper is then dried and the presence of boric acid is indicated by a pink or red coloration, depending on the amount present.

Benzoic acid can be determined by HPLC.

### **Pesticide Residues**

There is considerable difficulty in identifying pesticide residues because they may be present in very low concentrations. To measure organochlorine (e.g., 1,2,1-trichloro-2,3-bis-(4'-chlorophenyl)ethane) and organophosphorus (e.g., Bromophos, Malathion) residues, the fat is first extracted using a suitable nonpolar solvent. The residues are then extracted from this into acetonitrile and diluted with water. The residues are extracted into petroleum ether, purified through a column of Florisil, and eluted with petroleum ether and diethyl ether. The residues in the eluate are measured by GC with mass spectrometry (GC-MS). The software controlling the MS uses a database to identify signatures from specific groups in molecules.

### **Polychlorinated Biphenyls**

Two similar methods are used to measure polychlorinated biphenyls (PCBs). In the first they are isolated by solvent extraction along with organochlorine pesticides and measured by GLC. In the second method, they are extracted by solvents and purified on alumina or on Fluorosil columns before determination by GC-MS.

### **Dioxins**

After treatment with ammonia and ethanol, the fat phase containing the dioxins is extracted into ether-pentane, then dried and concentrated. It is cleaned up by passing it through a series of columns of base silica, alumina, and activated carbon. The eluate is analyzed by high-resolution GC-MS. The method is calibrated by using standard dioxin labeled with  $^{13}\text{C}$ .

### **Radionuclides**

Radionuclides are either naturally occurring or come from nuclear accidents such as the Chernobyl reactor disaster. Strontium-89 and strontium-90 ( $^{89}\text{Sr}$ ,  $^{90}\text{Sr}$ ), cesium-137 ( $^{137}\text{Cs}$ ), and iodine-131 ( $^{131}\text{I}$ ) are probably the most important contaminants. Strontium is adsorbed into bone, cesium mainly into muscle, and iodine only into the thyroid. These radionuclides are measured by scintillation counting of their radioactivity. Samples have to be enriched, unless the contamination is extremely high, to get counts that are significantly above background. Scintillation counting consists of placing a clarified sample into a special glass bottle with a reagent (scintillant) that

absorbs the radiation and emits light in proportion. The light is detected by a sensitive photomultiplier and expressed as counts per unit time or time to standard count.

### **Quaternary Ammonium Compounds**

These residues from cleaning agents and other fluids containing surfactants (e.g., cetyl pyridinium chloride and cetyl trimethyl ammonium bromide) may have bacteriostatic effects or cause flavor taints. To determine quaternary ammonium compounds (QACs), they are first extracted into solvent (1,1,2,2-tetrachloroethane) and then detected by the formation of a pink color with dioctyl sodium sulfosuccinate (Aerosol T), which is quantified by spectrophotometry.

Some sanitizing agents used in the dairy industry contain iodine and nonionic detergent (iodophors). Free iodine can be detected by adding an equal volume of concentrated hydrochloric acid to cold milk followed by a few drops of 2-phenylbenzo(h)chromen-4-one(7,8-benzoflavone) solution and a similar quantity of hydrogen peroxide. This is shaken and the presence of iodine is indicated by a lilac or blue color.

### **Added Dyes**

Flavored milks and other dairy-based foods may have certain added colors. The natural dye annatto is used to give a yellow/golden color. It can be detected by adding sodium hydrogen carbonate to a small amount of milk in a shallow dish and into this is placed a length of filter paper. If annatto is present, the paper is stained brown and turns pink when a drop of stannous chloride solution is added.

### **Analysis of Other Dairy Products**

**Cream** There are statutory requirements for the composition of cream. The fat content has to reach certain minima in order for cream to qualify as single, double, etc. The fat content is best determined by the Rose-Gottlieb method described above but the Gerber method may also be used.

Other components of cream, such as TS, protein, lactose, sucrose, and chloride, can be determined as described above. When acidity is determined, a control of cream, water, and dilute cobalt sulfate is used. Otherwise the method involves titration with  $0.111 \text{ mol l}^{-1}$  sodium hydroxide as before.

Some creams may contain thickeners. Starch may be detected by adding iodine. Gelatin may be detected by mixing the cream with Stokes reagent (mercury dissolved in concentrated nitric acid), filtering, and then adding saturated picric acid solution. A yellow precipitate indicates the presence of gelatin.

**Condensed and evaporated milk** The TS, fat, protein, ash, and acidity of these milks can be measured by the standard methods with slight modification in some cases. Sucrose content can be determined by polarimetry before and after inversion but more readily by HPLC.

**Dried milk products** TS, fat, protein, ash, lactose, and acidity can be measured by the standard methods. Vitamins A and D can be measured by HPLC. The burnt particles formed in the drying process can be determined after filtration by comparing the appearance of the filter pads with photographic reference standards.

**Yogurt** Most of the methods described above can be used to characterize yogurt. The fruit content can be estimated from the citric and malic acid contents using HPLC.

**Butter** Butter should be analyzed for water, salt, and butterfat content. Water can be determined by the usual drying procedure and salt by Volhard titration. Alternatively, salt may be measured by diluting it in hot water and titrating with silver nitrate using potassium chromate indicator.

The genuine butterfat content may be determined by GC or by determining the Reichert, Polenske, and Kirschner values for volatile fatty acids in butter. These are obtained by titration of various distillation fractions against  $0.1 \text{ mol l}^{-1}$  sodium hydroxide.

**Cheese** The standard methods described previously are used to determine the composition of cheese with minor modifications. The Werner-Schmid method is generally preferred for the determination of fat but the Gerber method is also used. In both cases, the cheese has to be mixed very thoroughly with the reagents.

The emulsifying salts in processed cheese (sodium, potassium, calcium and ammonium phosphate, citrate, and tartrate) can be measured from the phosphate, citrate, and tartrate content, after removing the protein with phosphotungstic acid, using standard methods.

**Nitrate and nitrite** The nitrate and nitrite content of cheeses that have been treated with these salts can be determined by reduction with cadmium followed by reaction with sulfonamide and N-1,naphthylethylenediamine hydrochloride. This produces a colored compound that can be quantified spectrophotometrically.

**Ice cream** Most of the methods described for the analysis of milk can be similarly applied to ice cream. As with cheese, moisture is determined by drying samples in dishes containing predried sand into which the ice cream is mixed. Ice cream contains considerable amounts of air that is whipped in to give the appropriate texture. This gives the product a greater volume than the starting mixture and this increase, known as 'overrun', is expressed as a percentage of the original volume. This can be calculated according to the formula:

Percent overrun

$$= \frac{\text{Volume of ice cream} - \text{Volume of mix}}{\text{Volume of mix}} \times 100$$

$$= \frac{\text{SG melted ice cream} - \text{SG ice cream before melting}}{\text{SG melted ice cream}} \times 100$$

where SG is the specific gravity.

When the ice cream is melted, all the air must be removed before measuring the specific gravity.

*See also: Food and Nutritional Analysis: Overview; Sample Preparation; Meat and Meat Products; Oils and Fats; Packaging Materials.*

## Further Reading

- BSI International, British Standards On-line; <http://bson-line.techindex.co.uk/> (website for BS analytical methods).
- Harrigan WF (1998) *Laboratory Methods in Food and Dairy Microbiology*, 3rd edn. London: Academic Press.
- International Dairy Federation (2002) List of Standards. Available on line at <http://www.fil-idf.org/standard-senglish.htm>
- James CS (1998) *Analytical Chemistry of Foods*. London: Kluwer Academic/Plenum Press.
- Kirk RS and Sawyer R (1991) *Pearson's Composition and Analysis of Foods*, 9th edn. Harlow: Longman Scientific & Technical.
- Nielsen SS (2003) *Food Analysis*. London: Kluwer Academic/Plenum Press.
- Nollet LML (2000) *Food Analysis by HPLC*. New York: Dekker.
- Pomeranz Y and Meloan CE (2002) *Food Analysis: Theory and Practice*, 3rd edn. London: Kluwer Academic/Plenum Press.
- Tortorello ML and Gendel ST (eds.) (1997) *Food Microbiological Analysis: New Technologies* (IFT Basic Symposium series). New York: Dekker.
- Wehr M (2004) *Standard Methods for the Examination of Dairy Products*. Washington: American Public Health Association.



## Vegetables and Legumes

**Rachael J Pither**, Campden and Chorleywood Food Research Association, Chipping Campden, Gloucestershire, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1553–1561, © 1995, Elsevier Ltd.

Vegetables form an important part of a balanced diet, supplying appreciable quantities of nutrients in a relatively low-energy source food. Their attractive appearance and their range of flavors and textures make them valued foodstuffs.

The preservation of vegetables in order to provide year-round availability is not a new idea. The Indians of Bolivia and Peru have been freezing and drying potatoes as a means of preservation for many hundreds of years, with the salting of vegetables and preserving in vinegar also predating the industrial era. However, it was the development of drying, freezing, and canning techniques in the nineteenth century and improvements of these in the twentieth century, as well as modern rapid transport systems, that have led to the now year-round availability of a very large range of vegetable products.

Vegetables are important sources of vitamin C (ascorbic acid),  $\beta$ -carotene, certain B vitamins, minerals, dietary fiber and protein, with different vegetable types being particularly rich in particular nutrients.

### Major Nutrient Components of Vegetables

In common with fruits the major component of fresh vegetables is water, with contents varying from approximately 80% to 95% in the edible portion. The nutrient composition of different vegetables varies considerably and can also be influenced by agronomic factors, storage and processing or preservation technique.

#### Green Vegetables

In general, green vegetables are high in water content, low in carbohydrate, fat and protein, and moderate to high in dietary fiber and a range of vitamins and minerals. They are particularly good sources of vitamin C,  $\beta$ -carotene, folic acid, iron, calcium, phosphorus, and potassium. Many green vegetables can be eaten in the raw state, particularly leafy salad vegetables, thereby avoiding any nutrient losses due to cooking or processing.

#### Root Vegetables

Root vegetables also have a high water content and relatively low carbohydrate content; one exception to this, however, is the parsnip, which contains on average 12–13% carbohydrate, comprising almost equal proportions of sugars and starch. Carrots are the most widely cultivated of root crops and are a particularly important source of  $\beta$ -carotene, with 100 g of raw carrot containing more than the US recommended daily dietary allowance for adult men. Although carrots are poor sources of vitamin C, other root vegetables contain reasonable amounts. Root vegetables also contain significant quantities of magnesium, iron, calcium, and potassium.

#### Potatoes

Due primarily to the quantities in which they are consumed, potatoes provide the main source of vitamin C in many diets, with a survey by the National Food Survey Committee (London) showing that on average over 25% of the daily requirement of ascorbic acid in the UK diet is derived from the potato. They are also good sources of thiamin, riboflavin and niacin, as well as the minerals iron, magnesium and potassium. On average ~75% of the total solids of potatoes are made up of carbohydrates, primarily starch, making the potato a good low-fat energy source. Although potato protein levels are reasonably low, they have a high biological value and are an important vegetable protein source.

#### Legumes

Food legumes are important sources of many nutrients, particularly for people following a largely vegetarian diet. They are rich in protein and carbohydrate, and provide more energy, B vitamins and dietary fiber than green or root vegetables. Many legumes also contain a small amount of oil, which is mostly unsaturated, with soybeans being a particularly good source. Soya flour and other products derived from beans are increasingly being used in manufactured foods, with soya flour especially useful in the diet of diabetic patients.

### Sampling and Preparation Techniques

Many of the issues relevant to sampling and preparation techniques for vegetables and legumes are as discussed in the article on Fruit and Fruit Products. Vegetables are naturally heterogeneous and great

consideration must be given to sampling technique. When sampling from the field there are several considerations that have to be taken into account:

- (a) Representative samples of the crop must be taken by a recognized procedure, e.g., ISO 874: 1980.
- (b) The individual units taken must be typical of those taken in a commercial harvest; undersized, diseased or immature product that would not normally be harvested must be avoided.
- (c) Samples must be taken in a way that is reasonably representative of typical harvesting procedure.
- (d) Quantity – the following values should be considered as a minimum, where possible:
  - (i) Root crops (large) – 5 kg, not less than five items.
  - (ii) Leafy or stem vegetables (large) – 5 kg, not less than five items.
  - (iii) Leafy or stem vegetables (small) – 2 kg.
  - (iv) Fruiting vegetables (large) – 5 kg, not less than five items.
  - (v) Fruiting vegetables (small) – 2 kg.
  - (vi) Legumes – 2 kg.

Suitable preparation procedures for vegetables and legumes are detailed in Table 1. For moist materials, fine slicing devices, e.g. domestic bowl-type food processors, Waring blenders etc., produce well-homogenized samples. Chilled ball mills can be used to grind frozen vegetable samples; this reduces the risk of undesirable chemical changes occurring. Grinding is the most suitable preparation method for dry legumes or other dried vegetables, with power-driven hammer mills most often used; domestic-type electric

coffee mills are also suitable. It is important to ensure that there is no excessive heat generation during any grinding or milling process. These methods are unsuitable for wet samples owing to losses of moisture and the possibility of chemical changes resulting from heat generation. The Godex Alimentarius sampling plans for prepackaged foods (CAC/RM 42-1969) are suitable for processed and quick-frozen vegetables and legume products.

Immediately after homogenization, the sample should be transferred to a storage vessel, which should be of glass or rigid plastic, have a tightly fitting lid, and be easily labeled. If the product is not to be analyzed immediately, it should be stabilized to prevent any oxidative, enzymatic, or microbial changes. For most vegetable products the most practical effective way of achieving this is by rapid freezing and storage at  $-20^{\circ}\text{C}$ .

## Analytical Methodologies

### Moisture and Total Solids

Accurate determination of water content presents many problems due to the wide variation in food matrix and the different forms in which water occurs in foods. Standard analytical techniques, therefore, are generally chosen for their reproducibility and practicability rather than for their high accuracy.

Several standard methods exist for use in quantification of moisture content in vegetables, legumes, and their products. Full descriptions of these can be found in the Association of Official Analytical Chemists' (AOAC) *Official Methods of Analysis* and other texts on food analysis.

**Table 1** Preparation procedures for vegetables, legumes and their products

<i>Vegetable type</i>	<i>Preparation technique</i>
Leafy (e.g., cabbages, lettuce) Root (e.g., carrots, parsnips, potatoes, onions, leeks)	Select a number of individual vegetables (4–6 for large items, 10–15 for smaller items). Remove any loose soil by and brushing, and remove any extraneous matter, dead leaves, stalks, etc. Roughly chop into quarters along the longitudinal axis and take a selection of these quarters. Mechanically finely chop and mix to prepare a laboratory sample.
Legumes (e.g., peas, beans)	Take a random selection of individual seeds from the sample (ideally at least 100 g) and mechanically finely chop and mix to prepare a laboratory sample.
Stalk (e.g., celery)	Remove inedible portions (leaves, etc.), remove any extraneous matter (soil, etc.). Chop into quarters, take a selection of these quarters and mechanically finely chop and mix to prepare a laboratory sample.
Packaged vegetables (e.g., cans, jars, bottles)	For whole contents: ensure that all contents are removed from the container and macerate thoroughly until a homogeneous mixture is obtained. For drained contents: ensure that all contents are removed from the container onto a sieve of suitable mesh size. Incline sieve and drain for 2 min. Macerate total solid portion until a homogeneous mixture is obtained.

**Table 2** Drying oven methods for the determination of moisture content

<i>Sample type</i>	<i>Methodology</i>
Fresh, canned, and frozen	Dry at 69–71°C in a vacuum oven $\leq 50$ mm Hg for 2 h and repeat drying and weighing until successive weighings vary by $\leq 1$ mg.
Frozen spinach	Dry at 105°C in air circulation or gravity convection oven for 4 h and repeat drying and weighing until weight loss in 1 h is $\leq 1$ mg.
Frozen french-fried potatoes	Dry at $103 \pm 2^\circ\text{C}$ in convection drying oven for 16 h and repeat drying and weighing until successive weight loss is $\leq 0.5$ mg.
Dried vegetables and legumes	Dry at $103 \pm 2^\circ\text{C}$ in convection drying oven for 5 h and repeat drying and weighing until successive weight loss is $\leq 1$ mg.

### Drying Oven Methods

These methods measure the weight of water lost due to evaporation and give accurate comparative results, although this may not be a true measure of total water content. A summary of suitable drying times and temperatures is given in **Table 2**.

**Karl Fischer method** This is primarily used for low-moisture products and is suitable for determining moisture in dried vegetables and legumes. This sensitive titration method is based on the nonstoichiometric reaction of water with iodine and sulfur dioxide in pyridine/methanol solution.

**Near-infrared spectrometric method** This method has also been shown to be particularly suited to the determination of moisture in dried vegetables. Absorption is measured at wavelengths characteristic of molecular vibration in water and compared with a calibration equation derived from reference samples of known composition.

**Rapid methods** Many more rapid techniques are used to monitor moisture content for quality control purposes. These include rapid infrared (IR) drying, microwave drying, and electrical methods. For both the IR and microwave methods, equipment is available with built-in balances, digital displays and microcomputers to calculate moisture content.

### Carbohydrates

**Sugars** Vegetables contain substantially less glucose and fructose than is present in fruits, and sugar beets are the only significant source of sucrose. Not all methods therefore are suitable for determination of sugars in vegetables and legumes because of these lower concentrations.

It is generally necessary with vegetables and legumes to clarify the hot water extract prior to sugar determination. This is achieved by the addition of heavy metals, typically lead acetate or Carrez I and II solutions ( $\text{K}_4\text{Fe}(\text{CN})_6 \cdot 3\text{H}_2\text{O}$  and

$\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ ), which precipitate the proteins and can also decolorize the solutions. Analysis of the water extract can then be carried out either by copper reduction methodology, enzymatically, or by liquid chromatography.

**Copper(II) reduction** Methods based on the ability of certain sugars, including all monosaccharides, to reduce copper(II) salts to copper(I) are recommended for measuring sugars in vegetable products. In the USA the Munson–Walker method as detailed by the AOAC is most commonly used, whereas in Europe the modified Lane–Eynon titration method is standard.

**Enzymatic methodology** Because of the specificity and accuracy of enzymatic reactions, and the availability of easy-to-use analytical kits, enzymatic methodologies are becoming ever more widely used in the analysis of sugars in a wide range of foodstuffs including vegetables. In common with the copper reduction methods, enzymatic methods have the advantage of simple sample preparation techniques. The enzymatic reactions involved cause the reduction of NADP to NADPH, the concentration of which is determined by spectrophotometry at 340 or 366 nm. Great care is needed in their use because of the very exact quantities involved; however, the results obtained compare favorably with more traditional methods and enzymatic methods are now accepted widely as a standard analytical technique.

**Liquid chromatography (LC)** This is now a popular and useful method of analysis which has the advantage that different sugars can be determined simultaneously. The separation technique can be carried out using strongly acidic cation exchange chromatography, or a modified silica gel amino column, with refractive index detection. This method does have certain disadvantages in that initial equipment costs are high and, owing to the sensitivity of the detectors, they are best when used exclusively and continuously for the same analysis. The latest development in LC analysis of carbohydrates uses improved anion exchange

techniques combined with pulsed amperometric detection, which is both more sensitive and selective than traditional LC methodology.

**Gas-liquid chromatography** This technique has the advantage of allowing food sugars to be determined with very high sensitivity and is the basis of an AOAC recommended method. The sample sugars have to be converted to volatile and heat-stable derivatives before analysis and this technique is not widely used in routine analysis because of the accuracy and relative simplicity of alternative methods.

**Starch** Starch is the main dry matter constituent of potatoes and legumes and is also found in many other vegetables, being important as an energy source and for its texture properties. Many methods are available for its quantification.

*Dispersion method* This method is suitable for potatoes and certain other vegetables but is not suitable for foods containing modified or presolubilized starches. It relies on the dispersion of the starch in an appropriate solution followed by determination by polarimetry. The most commonly used dispersants are hydrochloric acid and acidified calcium chloride. The solutions are then clarified and the optical rotation of the solutions measured. The specific rotation for potato starch is  $+185.7^{\circ}\text{C}$ .

*Hydrolysis methods* These involve the hydrolysis of starch to glucose which can be achieved using acid, enzymes, or a combination of these. Both acid and enzymatic hydrolysis methods can give low recoveries for modified starches; however, dispersion with dilute alkali prior to enzymatic hydrolysis can give better results. Enzymatic hydrolysis has the advantage of being specific to the starch components, whereas acid hydrolysis can lead to the breakdown of nonstarch polysaccharides.

The glucose can be determined enzymatically with a colorimetric end point or titrimetrically; commercial kits based on enzymatic hydrolysis of starch and measurement of glucose are currently available.

Although there are many variations on these two basic techniques, the accurate measurement of starch, especially in processed products where the starches may be modified or pregelatinized, remains an analytical problem.

**Structural polysaccharides** The quantification of the two main groups of complex structural polysaccharides, cellulose and pectins, involves many difficulties, partly because of the wide range of polymer size and composition that occurs.

*Cellulose and hemicellulose* These are among the most abundant carbohydrates found in vegetables and legumes. Methods for quantification generally involve the extraction of soluble carbohydrate from the defatted foodstuff, the extraction of lignin, and the hydrolysis of the remaining cellulose and hemicelluloses, followed by measurement of the resulting reducing sugars. The Englyst method for dietary fiber (see below) allows for the quantification of total nonstarch polysaccharides (NSP), soluble NSP, insoluble NSP, resistant starch, and cellulose.

*Pectin* Most standard procedures for pectin analysis in vegetables and legumes involve extraction of the pectin as pectic acid followed by precipitation. Aqueous extraction of the pectin is followed by saponification with cold alkali, acidification, and the quantification of the precipitate. The pectic acid can be precipitated as calcium pectate or, as in the AOAC procedure, using ethanol.

*Crude fiber* Crude fiber analysis is a measure of food components that are indigestible and is determined as material insoluble in dilute acid and dilute alkali under standard conditions. The basic method involves the defatted sample being boiled, first in 1.25%  $\text{H}_2\text{SO}_4$  and then in 1.25%  $\text{NaOH}$ , with the residue being filtered, dried, weighed, incinerated, and reweighed. The loss in weight on incineration is taken as the crude fiber content.

Many methods exist, mainly modifications of this basic technique, with the AOAC detailing a widely used version. Automated systems also exist, e.g. the Fibretec system in which the sample remains in the same glass crucible throughout the analysis. Crude fiber analysis is commonly carried out in cereals, animal feeds, including legumes, and flours; however, it is also used as an indicator of succulence in fresh vegetables.

**Dietary fiber** The determination of dietary fiber has increased greatly in importance in recent years; however, due partly to lack of a clear definition for dietary fiber, several standard methods exist.

*AOAC enzymatic gravimetric method* The samples are dried, with fat extracted if necessary, gelatinized with heat-stable  $\alpha$ -amylase and enzymatically digested with protease and amyloglucosidase to remove protein and starch. The soluble dietary fiber is then precipitated using ethanol, the residue filtered and washed, and corrections made for indigestible protein and ash. This method is widely used in the EC and the USA.

*Englyst method* Samples are defatted if required and any starch present is gelatinized and then removed by enzymatic digestion. The remaining polysaccharides are hydrolyzed by sulfuric acid and the resulting sugars can be measured either colorimetrically or by using gas chromatography. This method can be used to determine both total NSPs and resistant starch and is a recognized method in the UK and Canada.

### **Vitamin C (Ascorbic Acid)**

There are two standard procedures for vitamin C determination in vegetables and legumes. The sample preparation is common to both and involves maceration/dilution of the sample in a stabilizing solution, such as 5% metaphosphoric acid or trichloroacetic acid, followed by filtration. The ascorbic acid can then be determined by titration with 2,6-dichloroindophenol, where the ascorbic acid reduces the oxidation/reduction indicator dye to a colorless solution; or by fluorimetric detection, in which the ascorbic acid is oxidized to dehydroascorbic acid, which then reacts with *o*-phenylenediamine (1,2-diamino benzene) to produce a fluorophore. The latter method has the advantages that it is suitable for colored solutions and can be used to measure levels of naturally occurring dehydroascorbic acid as well as ascorbic acid. Many other methods are available for determination of vitamin C with the use of LC techniques, currently the subject of much interest. Reversed-phase LC techniques can be used to determine dehydroascorbic acid, ascorbic acid, and their isomers.

### **Carotenoids**

$\beta$ -Carotene, the most important carotenoid precursor of vitamin A, is found in carrots and many dark-green or yellow vegetables. In the average diet,  $\beta$ -carotene provides approximately one-third of the vitamin A intake.

For accurate determination of carotenoids, LC is the preferred technique and many, LC procedures have been published. Generally the sample is extracted into a solvent before being analyzed by reversed-phase LC with UV absorbance detection. These methods have the advantage of being able to detect and quantify individual carotenoids, allowing for more accurate calculation of vitamin A activity.

UV-visible spectrophotometric methods can also be used, such as are detailed in the AOAC *Official Methods of Analysis*. They are less accurate and less selective than LC methods as they measure total carotenoids; however, they offer a rapid, relatively simple analytical procedure.

### **B Vitamins**

Vegetables and legumes are good sources of many B vitamins, including thiamine and niacin (potatoes, legumes), pyridoxine (potatoes, green vegetables, peas), folic acid (green leafy vegetables), and pantothenic acid (legumes). The analysis of B vitamins is a very complex and detailed subject. Many analytical methods are available, although much work is still required in order to provide standard, reliable, reproducible measurement in all food types.

Several B vitamins, including folic acid, niacin, pyridoxine, and pantothenic acid, are routinely determined using microbiological assays, details of which can be found in the AOAC *Official Methods of Analysis*. Standard methods for thiamine determination using fluorimetric detection are also detailed in the AOAC methods; in addition, LC techniques are now being used routinely for thiamine and other B vitamins, e.g., riboflavin.

### **Protein**

The protein content of vegetables and legumes is most commonly estimated from the organic nitrogen content, determined using the Kjeldahl procedure. This can lead to slight inaccuracies due to the presence of nonprotein nitrogen compounds; however, the content of these is usually very small when compared to protein content. The Kjeldahl method is based on the wet combustion of the sample, by heating in sulfuric acid in the presence of a metallic catalyst, in order to reduce the organic nitrogen to ammonia, which is retained in solution as ammonium sulfate. The digest is made alkaline by the addition of sodium hydroxide and steam distilled to release the ammonia, which is then trapped and titrated. Many modifications of the basic method have since been carried out. The AOAC details both macro-Kjeldahl and micro-Kjeldahl methods. The former uses mercury as a catalyst during digestion; the ammonia is trapped, after distillation, in a standard acid solution which is then titrated against sodium hydroxide. The latter uses boric acid to trap the released ammonia which is then titrated against dilute acid. Many automated systems have been developed based on the Kjeldahl method. These instrumental systems include digestion in an electrically heated aluminum block digester, followed by rapid steam distillation of ammonia into acid and titration, either manually or automatically, with colorimetric or potentiometric detection.

Rapid automated systems, employing high-temperature combustion for removal of the nitrogen from the sample and a thermal conductivity detection system for measurement of the released

nitrogen, are also now available, allowing protein measurements to be carried out in <10 min.

Analysis of the individual amino acids is not normally a part of routine nutritional analysis; however, as proteins from plant sources, legumes, etc. can be deficient in certain amino acids, it is sometimes required. For this the proteins are extracted and assayed by thin-layer chromatography, gas chromatography, or LC.

## Fat

Most vegetables and legumes contain little naturally occurring lipid, the exceptions to this being soy bean, groundnut, and winged bean. Lipid is present in both free and bound forms and the amount extracted will depend on the method of analysis used. Free lipid consists essentially of neutral fats (triglycerides) and free fatty acids, which can be determined by extraction of the dried material with petroleum ether or diethyl ether. This method is suitable only for frozen fried vegetable products, e.g. french-fried potatoes, where a 'free' or 'crude' fat content is sufficient. Bound lipid can be released by acid hydrolysis, the extract filtered, and the lipid extracted as for free lipid. The latter method is suitable for a very wide range of foodstuffs including vegetables and legumes. Automated systems are widely available based on these methods, most comprising an acid hydrolysis unit and a Soxhlet extraction unit. The finely ground sample is boiled in hydrochloric acid, filtered, dried, and extracted using light petroleum ether or diethyl ether.

The Bligh and Dyer technique is based on the homogenization of wet foods with chloroform and methanol. Further addition of chloroform and water leads to the separation of the phases, with the lipid contained in the chloroform layer. The AOAC details a similar extraction method based on chloroform-methanol extraction, the advantage of these methods being that no heating stage is involved and the lipid extract can then be analyzed to determine the individual fatty acids.

Many rapid methods are now available for estimation of fat content. The Foss-Let fat analysis system determines the fat content as a function of relative density of a perchloroethylene extract of the sample. It was designed for analysis of oilseeds, but is also applicable to other foods. The method is very rapid and its accuracy is comparable to other standard methods. Nuclear magnetic resonance (NMR) spectrometry is a rapid, nondestructive method also suitable for determining the fat content of oilseeds. The NMR value is related to the total hydrogen content of the lipid, with actual lipid content calculated from calibration tables.

## Fatty Acids

Legumes contain predominantly unsaturated fatty acids; there are exceptions to this, with certain legumes containing appreciable proportions of saturated fats. In order to determine accurately the fatty acid profile of legumes, the fat portion must be extracted without heating; chloroform-methanol extraction methods are therefore widely used, e.g., those of Bligh and Dyer, Karow.

Analysis of the fatty acid profile of the extracted fat is now almost exclusively carried out using gas-liquid chromatography (GLC). Standard methodology for this technique is detailed by the International Union of Pure and Applied Chemistry (IUPAC). It involves the saponification of the extracted fat to break down glycerides, with the liberated fatty acids being esterified in the presence of methanol and boron trifluoride. Fatty acid methyl esters are then extracted with heptane and analyzed using GLC with flame ionization detection.

This allows for the percentage composition of individual fatty acids present in the sample to be determined, enabling levels of saturated, polyunsaturated, and mono-unsaturated fats to be calculated.

## Salt

Sodium chloride levels are routinely determined in canned vegetables and legumes. Titrimetric methods are most commonly used with the sodium chloride first isolated by either ashing at 500–550°C, followed by aqueous dissolution of the ash and titration with silver nitrate solution (Mohr method), or by boiling the food in dilute nitric acid, adding excess silver nitrate and back-titration with potassium thiocyanate (Volhard method). Of these, the latter method is generally more accurate than the former method; however, it is also more time-consuming.

As well as these titrimetric methods, the AOAC details a potentiometric method that is also recommended in Codex Alimentarius standards. The vegetables or legumes are dispersed in water and acidified, with the soluble chlorides titrated potentiometrically with silver nitrate. Many instruments are also now available with ion-selective electrodes that can be applied to the determination of salt in foods. They offer reasonable accuracy with ease of use and are particularly suitable for quality assurance purposes.

## Minerals (Metal Compounds)

The AOAC and Codex Alimentarius methods for determination of potassium and sodium by flame atomic emission spectrometry are suitable for the analysis of many wet vegetables. These methods have the advantage of quick and simple sample



preparation, aqueous dilution of sample, filtration and rapid, reasonably accurate detection. The samples can be aspirated directly into the flame and the readings quantified by comparison with a range of known standards. Atomic absorption spectrometry (AAS), however, has many advantages; increased accuracy and low-level sensitivity and a far wider range of detectable elements. The dry-ashing sample preparation technique required, though simple, can take a relatively long time; however, it is suitable for all types of samples.

Many elements are essential in the diet at trace levels, although at higher concentrations some possess toxic properties, examples being zinc, copper, and selenium. Vegetables and legumes are routinely tested for trace element content, e.g., canned products for tin, iron, copper, and zinc; vegetable juices for copper, zinc, and lead; potatoes for copper, zinc, and iron. Statutory limits exist for many trace elements in foods, e.g., lead and arsenic. AAS is widely used for trace metal analysis as multi-element determinations can be made on a single sample; however, other methodologies are applicable, including inductively coupled plasma-atomic emission spectrometry and colorimetric methods such as the determination of copper as its diethyldithiocarbamate.

### **Phosphate**

Certain vegetables, especially potatoes and cabbage, are good sources of phosphorus. In order to determine the phosphorus content, the sample is ashed and dissolved in hydrochloric acid; the phosphate can then be determined titrimetrically or colorimetrically. The most widely accepted method is a colorimetric (visible spectrophotometric) method based on Mission's reagent, which relies upon the reaction between the acidified sample solution and an acidic reagent containing molybdic acid and vanadic acid, to produce the stable orange-yellow molybdovanadophosphoric acid.

### **Quality Assurance**

A range of analyses and tests are carried out on vegetables, legumes, and their products as a means of quality control, and many regulations and specifications exist in relation to this. These include the Joint FAO/WHO Food Standards Programme (Codex Alimentarius), the US Department of Agriculture (USDA) standards for grades of processed products, the British Fruit and Vegetable Canners Association Codes of Practice, and EC directives.

The range of analyses carried out to determine compliance with quality and specification

requirements can include pH, titratable and volatile acidity; total solids, water-insoluble solids, alcohol-insoluble solids; sugars; ash, sodium chloride, trace metals, heavy metals; preservatives, e.g., sulfur dioxide, sorbic acid, benzoic acid, and benzoates; nitrates and enzymatic activity, e.g., peroxidase.

In general, the specification or quality document specifies methods to be used; however, the following are standard recommended methods as detailed in AOAC and Codex standards.

#### **Alcohol-Insoluble Solids**

This is usually measured in both canned and frozen peas and is an indication of maturity. The well-homogenized peas are boiled in ethanol, and filtered using a Buchner funnel. The residue is washed repeatedly with 80% ethanol to remove all soluble substances. The alcohol-insoluble solids are dried, weighed, and their percentage content calculated.

#### **Organic Acids, Acidity**

Measurement of titratable acidity is routinely carried out in many canned or pickled vegetable products, with the acidity calculated as the predominant acid. The sample is diluted in distilled water and titrated against dilute sodium hydroxide to either the phenolphthalein end point or to pH 8.1. Volatile acidity, generally expressed as acetic acid, can be measured by distilling the sample in a steam distillation apparatus, with titration of the distillate as above. Individual acids can be determined using a range of techniques including chromatographic and titrimetric methods fully described in the AOAC document, and enzymatic methodologies with commercial enzyme-based test kits now available for many acids. LC techniques are also available, facilitating identification and quantification of a range of organic acids.

#### **Preservatives**

The use of preservatives in fresh and processed vegetables and legumes is tightly controlled worldwide, e.g., in the UK under the Preservatives in Food Regulations (1989), while in the EC a directive is currently being prepared on Food Additives (excluding colors and sweeteners).

Sulfur dioxide (SO<sub>2</sub>) is generally measured by indirect methods. The AOAC describes a modified Monier Williams method, while the Codex Alimentarius standards and EC directive recommend the modified Tanner method; both are based on similar principles, involving the distillation of the SO<sub>2</sub> from an acidified aqueous suspension of the food. The evolved gas is subsequently trapped in

hydrogen peroxide where it reacts to form sulfuric acid, which is then titrated against dilute sodium hydroxide.

Benzoic acid and sorbic acid can both be extracted using modified Monier Williams distillation methods and quantified spectrophotometrically, or in the case of benzoic acid, by thin-layer chromatography. LC and GLC methodologies are available and routinely used, having the advantage of simultaneous quantification of both benzoic acid/benzoates and sorbic acid.

### Chemical Residues

Fresh vegetables and legumes in particular may also be analyzed for the presence of chemical residues. These may be present due to specific horticultural and agricultural practices (pesticides, herbicides, etc.), environmental contamination, or through the use of certain permitted additives. Detection and quantification of these residues is a complex analytical problem. Most methods used are based on solvent separation coupled with chromatographic determination using either LC, GLC or GC-mass spectrometry (GC-MS).

*Pesticide residue screening* A procedure for determining levels of organophosphorus and organochlorine pesticide residues suitable for vegetables and legumes can be found in the USA Food and Drug Administration Pesticide Analytical Manual and in AOAC methods. These methods involve the extraction of the sample into acetonitrile, with the residues being partitioned into petroleum ether. Depending on the nature of the extract, residues can then be analyzed directly, or cleaned up using Florisil or alumina columns. Analysis is by GC with nitrogen-phosphorus or flame photometric detection for organophosphorus residues, and electron capture detection for organochlorine residues.

*Nitrate and nitrite* Considerable attention is currently being paid to the levels of nitrate being found in vegetables with surveyed data ranging from 1 to 10 000 mg kg<sup>-1</sup>. Ion chromatography (IC) is a widely used method for measuring nitrate and nitrite in vegetables and their products; however, for rapid routine screening procedures electrometric measurement systems are available. A standard LC technique involves the extraction of the sample with hot water, filtration, and a sample clean-up procedure using a solid-phase extraction cartridge. The sample is then analyzed for nitrate and nitrite by IC to separate the ions, with detection by UV absorption at 214 nm.

### Enzymatic Activity

The quality of vegetables and legumes, their color, appearance, flavor, aroma and texture, can be greatly affected by endogenous enzyme activity. The effectiveness of heat-processing, e.g., blanching of vegetables for freezing, in terms of enzyme inactivation, is generally assessed by testing for peroxidase activity, peroxidase being one of the more heat-stable enzymes.

The AOAC details a titrimetric procedure for the quantification of peroxidase activity; however, for routine quality assurance a rapid spot test method is generally used. This is based on the color reaction catalyzed by peroxidase which involves guaiacol in the presence of hydrogen peroxide. The development of a brown/red color on the vegetable, after addition of the substrate, indicates that the concentration of active enzymes may be sufficient to cause a noticeable deterioration in quality.

An additional method for the assessment of adequacy of blanching is based on the detection of residual lipoxygenase. The lipoxygenase catalyzes a reaction between linoleic acid and oxygen to form a hydroperoxide. This product reacts with potassium iodide to form iodine, giving a brown coloration which can be measured spectrometrically.

### Rapid Methods

Research into new analytical techniques for foodstuffs continues, striving for greater accuracy, sensitivity or simplicity, for more rapid methods, for simultaneous multielement analysis, etc. Chromatographic techniques, e.g., LC, GLC, GC-MS, have led to great improvements in the levels of accuracy, sensitivity, and detection that can be achieved for many analytes including carbohydrates, certain vitamins, chemical residues, and additives. Work is still required, for instance, in the area of vitamin analysis in order to provide standard techniques that are applicable to all food types and that would enable concurrent multi-vitamin analysis to take place. Many of the microbiological assays currently used for vitamin determination involve long incubation times and more rapid techniques are needed.

This need for rapid analytical systems extends across most of the routine nutritional and quality assurance analyses. Recent advances in near-IR technology have led to instrumentation capable of rapid quantification of moisture, fat, and protein levels in certain foods; however, this area is still quite limited.

*See also:* **Atomic Absorption Spectrometry:** Flame. **Atomic Emission Spectrometry:** Flame Photometry;

Inductively Coupled Plasma. **Carbohydrates:** Sugars – Spectrophotometric Methods; Starch. **Food and Nutritional Analysis:** Overview; Antioxidants and Preservatives; Oils and Fats. **Ion-Selective Electrodes:** Food Applications. **Liquid Chromatography:** Amino Acids. **Proteins:** Foods. **Quality Assurance:** Quality Control. **Sampling:** Theory. **Vitamins:** Fat-Soluble; Water-Soluble. **Water Determination.**

## Further Reading

Association of Official Analytical Chemists (AOAC) (1990) *Official Methods of Analysis*, 15th edn. Arlington, VA: AOAC.

Codex Alimentarius Commission (1981) *Codex Standards for Processed Fruits and Vegetables and Edible Fungi*.

Rome: Food and Agriculture Organization/World Health Organization.

Eskin NAM (1989) *Quality and Preservation of Vegetables*. Boca Raton, FL: CRC Press.

Kirk RS and Sawyer R (1991) *Pearson's Composition and Analysis of Foods*, 9th edn. London: Longman Scientific and Technical.

Paquot C and Hautfenne A (1987) *Standard Methods for the Analysis of Oils, Fats and Derivatives*, 7th edn. Oxford: International Union of Pure and Applied Chemistry (IUPAC)/Blackwell Scientific Publications.

Pomeranz Y and Meloan CE (1987) *Food Analysis, Theory and Practice*, 2nd edn. New York: AVI.

Robinson DS (1987) *Food Biochemistry and Nutritional Value*. London: Longman Scientific and Technical.

## Oils and Fats

**E W Hammond**, Greenisle-Consulting, Northants, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Major Analytes

Lipids that are solid at ambient temperatures are referred to as 'fats' while those that are liquid are called 'oils'. Two main groups, having significantly different natural functions and structures, can be classified as the 'neutral lipids' (acylglycerides, fatty acids, alcohols, hydrocarbons, waxes) and 'polar' lipids (glycolipids, phospholipids). Neutral lipids generally form part of an energy store, while polar lipids are functional molecules in the structure of membranes.

Fatty acids (FAs) form homologous series of 'straight' chain carboxylic acids, with generally even numbers of carbon atoms. These vary typically from C4 to C30, with the majority lying between C10 and C22. Regio-specific desaturase enzymes yield an important series of FAs that contains one or more double bonds, of the *cis* isomeric form. For multiple double bonds, each pair is normally separated by a single methylene unit ( $-\text{CH}_2-$ ). The series of unsaturated FAs synthesized in plants is different from that in animals. It is important analytically to recognize this fact, particularly where nutritional profiling might be being done, as the difference creates the requirement by animals for the 'essential dietary fatty acids' produced by plants. By combining desaturation and chain elongation, animals create two main

series of polyunsaturated fatty acids, the  $n-6$  and  $n-3$  (measuring from the methyl end) with chain lengths usually up to C22.

Exceptions to the generality of straight chains are FAs substituted with hydroxy, keto (oxo), and epoxy (oxirane) groups and FAs with unusual unsaturation. Methyl branched FAs occur mainly in bacteria, but also naturally in the fat of ruminant animals together with a series of odd-chain fatty acids.

Triacylglycerides (TAGs) are fatty acid triesters of glycerol. The hydrolysis products of TAGs are FAs together with diacyl- and monoacyl-glycerides. Due to acyl transfer diacylglycerides exist in the 1:2(2:3) and 1:3 forms. Monoacylglycerides can exist in the 1( $\alpha$ ) and the 2( $\beta$ ) forms, but are predominantly in the  $\alpha$  form. Despite the potential structural complexity, most of the TAGs that the food analyst encounters are composed of a normal range of FAs. There is considerable conformity of structure in natural TAGs. This is a bonus for the analyst making differentiation between species of, for instance, seed oils easier, even after allowing for varietal (cultivar) differences. **Table 1** shows typical data for some edible fats.

Polar lipids are conveniently divided into phospholipids and glycolipids. This is not entirely adequate, but it will serve here to describe the structural types found in food. Phospholipids fall into two main groups. The phosphoglycerides carry an *sn*-3 phosphate with a nitrogenous base such as choline or ethanolamine (other groups linked to this phosphate may be inositol or glycerol). In the sphingosyl phosphatides sphingosine is esterified with, for

Inductively Coupled Plasma. **Carbohydrates:** Sugars – Spectrophotometric Methods; Starch. **Food and Nutritional Analysis:** Overview; Antioxidants and Preservatives; Oils and Fats. **Ion-Selective Electrodes:** Food Applications. **Liquid Chromatography:** Amino Acids. **Proteins:** Foods. **Quality Assurance:** Quality Control. **Sampling:** Theory. **Vitamins:** Fat-Soluble; Water-Soluble. **Water Determination.**

## Further Reading

Association of Official Analytical Chemists (AOAC) (1990) *Official Methods of Analysis*, 15th edn. Arlington, VA: AOAC.

Codex Alimentarius Commission (1981) *Codex Standards for Processed Fruits and Vegetables and Edible Fungi*.

Rome: Food and Agriculture Organization/World Health Organization.

Eskin NAM (1989) *Quality and Preservation of Vegetables*. Boca Raton, FL: CRC Press.

Kirk RS and Sawyer R (1991) *Pearson's Composition and Analysis of Foods*, 9th edn. London: Longman Scientific and Technical.

Paquot C and Hautfenne A (1987) *Standard Methods for the Analysis of Oils, Fats and Derivatives*, 7th edn. Oxford: International Union of Pure and Applied Chemistry (IUPAC)/Blackwell Scientific Publications.

Pomeranz Y and Meloan CE (1987) *Food Analysis, Theory and Practice*, 2nd edn. New York: AVI.

Robinson DS (1987) *Food Biochemistry and Nutritional Value*. London: Longman Scientific and Technical.

## Oils and Fats

**E W Hammond**, Greenisle-Consulting, Northants, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Major Analytes

Lipids that are solid at ambient temperatures are referred to as 'fats' while those that are liquid are called 'oils'. Two main groups, having significantly different natural functions and structures, can be classified as the 'neutral lipids' (acylglycerides, fatty acids, alcohols, hydrocarbons, waxes) and 'polar' lipids (glycolipids, phospholipids). Neutral lipids generally form part of an energy store, while polar lipids are functional molecules in the structure of membranes.

Fatty acids (FAs) form homologous series of 'straight' chain carboxylic acids, with generally even numbers of carbon atoms. These vary typically from C4 to C30, with the majority lying between C10 and C22. Regio-specific desaturase enzymes yield an important series of FAs that contains one or more double bonds, of the *cis* isomeric form. For multiple double bonds, each pair is normally separated by a single methylene unit ( $-\text{CH}_2-$ ). The series of unsaturated FAs synthesized in plants is different from that in animals. It is important analytically to recognize this fact, particularly where nutritional profiling might be being done, as the difference creates the requirement by animals for the 'essential dietary fatty acids' produced by plants. By combining desaturation and chain elongation, animals create two main

series of polyunsaturated fatty acids, the  $n-6$  and  $n-3$  (measuring from the methyl end) with chain lengths usually up to C22.

Exceptions to the generality of straight chains are FAs substituted with hydroxy, keto (oxo), and epoxy (oxirane) groups and FAs with unusual unsaturation. Methyl branched FAs occur mainly in bacteria, but also naturally in the fat of ruminant animals together with a series of odd-chain fatty acids.

Triacylglycerides (TAGs) are fatty acid triesters of glycerol. The hydrolysis products of TAGs are FAs together with diacyl- and monoacyl-glycerides. Due to acyl transfer diacylglycerides exist in the 1:2(2:3) and 1:3 forms. Monoacylglycerides can exist in the 1( $\alpha$ ) and the 2( $\beta$ ) forms, but are predominantly in the  $\alpha$  form. Despite the potential structural complexity, most of the TAGs that the food analyst encounters are composed of a normal range of FAs. There is considerable conformity of structure in natural TAGs. This is a bonus for the analyst making differentiation between species of, for instance, seed oils easier, even after allowing for varietal (cultivar) differences. **Table 1** shows typical data for some edible fats.

Polar lipids are conveniently divided into phospholipids and glycolipids. This is not entirely adequate, but it will serve here to describe the structural types found in food. Phospholipids fall into two main groups. The phosphoglycerides carry an *sn*-3 phosphate with a nitrogenous base such as choline or ethanolamine (other groups linked to this phosphate may be inositol or glycerol). In the sphingosyl phosphatides sphingosine is esterified with, for

**Table 1** Fatty acid and triacylglyceride composition of typical edible fats

Edible fat	% Fatty acid (chain length and number of double bonds)														
	8:0	10:0	12:0	14:0	16:0	16:1	18:0	18:1	18:2	18:3	20:0	20:1	22:0	22:1	24:0
Palm kernel oil	4.9	4.0	47.5	15.8	7.7	–	2.4	14.3	3.1						
Palm oil	–	–	–	1.0	43.8	0.5	5.0	38.5	10.5	0.3	0.4		–	–	–
Soya bean oil	–	–	–	–	10.0	0.2	3.5	21.0	55.6	9.2	0.5	0.3	–	–	–
Rapeseed oil	–	–	–	tr	4.7	0.3	1.7	59.3	21.4	9.9	0.6	1.4	0.4	0.3	0.2

Edible fat	% Triacylglyceride (carbon number)																	
	C <sub>30</sub>	C <sub>32</sub>	C <sub>34</sub>	C <sub>36</sub>	C <sub>38</sub>	C <sub>40</sub>	C <sub>42</sub>	C <sub>44</sub>	C <sub>46</sub>	C <sub>48</sub>	C <sub>50</sub>	C <sub>52</sub>	C <sub>54</sub>	C <sub>56</sub>	C <sub>58</sub>	C <sub>60</sub>	C <sub>62</sub>	C <sub>64</sub>
Palm kernel oil	1.7	6.9	9.2	21.6	16.5	9.9	8.8	6.5	5.3	5.7	2.4	2.5	2.8	0.2	–	–	–	–
Palm oil	–	–	–	–	–	–	–	–	0.4	9.2	42.8	38.9	8.4	0.3	–	–	–	–
Soya bean oil	–	–	–	–	–	–	–	–	0.2	0.7	5.1	27.1	61.3	4.0	1.2	0.3	tr	–
Rapeseed oil	–	–	–	–	–	–	–	–	–	0.3	1.1	15.1	74.2	6.8	1.3	0.6	0.4	0.2

Note: Carbon number relates to the sum of the carbon atoms in the three fatty acid chains of a triglyceride. Thus, C<sub>54</sub> might be made up of 18 + 18 + 18, 16 + 18 + 20, etc. No distinction is made, in this analysis, between saturated and unsaturated fatty acids of the same chain length. Data can vary between batches of fat or oil. Large differences can be found between different cultivars and varieties of the same plant.

instance, phosphorylcholine. Glycolipids feature widely in nature. The structure is a 1:2 diacylglyceride with the *sn*-3 hydroxyl glycosidically linked to a mono-, di-, or trisaccharide. The mono- and di-galactosyldiglycerides are very abundant, particularly in cereal lipids.

Other common food lipids are the sterols and tocopherols. Sterols are generally present in natural oils as both the free sterol and sterol fatty esters. Animal fats exhibit solely cholesterol while plant lipids contain a range of phytosterols. Tocopherols are found in vegetable oils as a mixture of four isomers:  $\alpha$ -tocopherol (true vitamin E) and  $\beta$ -,  $\gamma$ -, and  $\delta$ -tocopherol. Four unsaturated analogs, tocotrienols, exist with the same numbering terminology and are particularly abundant in palm oil.

## Sampling, Preparation, Extraction of Lipids

Obtaining the lipid in suitable form for analysis may be the most difficult stage of sample analysis. This step can be responsible for the degradation of lipids and the formation of artifacts. There are a number of important considerations to be made if the analytical results are to give a true representation of fat composition:

- the sampling procedure,
- the extraction procedure,
- processing temperatures,
- the possible exclusion of air,
- the lighting conditions,
- the surface area of lipid films during concentration,

- the possible addition of antioxidants, and
- the storage temperature.

Representative sampling of a bulk of material is not easy. Only gram quantities are required for analysis, but this size of sample may have to represent many tons. The analyst may have to make assumptions about the contents and a conclusion about homogeneity. Bulk fat storage tanks ideally will be heated to maintain a temperature of 5–15°C above the clear point of the fat (the point at which no crystalline fat exists). The contents are unlikely to have been mixed, other than by agitation during transport and pumping from one tank to another. Bulk oils can exhibit layering due to density gradients and if storage temperature drops to, or slightly below, the clear point, the oil will begin to crystallize causing sedimentation.

Sampling at several different levels and in different quadrants of a tank is normal. The International Organization for Standardization developed the International Standard–ISO-5555. Animal and Vegetable Fats and Oils Sampling. The British Standard BS 627 (1982) forms the basis of the International Standard. The Standard includes: sampling technique; methods and apparatus for sampling; packaging and labeling of samples; dispatch of samples; preparation of a sampling report. Any sampling from a large bulk of fat must follow the procedures laid out in the Standard as the sample and the report are considered to be legal evidence in disputes.

Smaller quantities of fat (0.5–25 kg) should be melted and mixed very well before sampling. For measurement of the oxidation level, a core sample should be taken, before melting, with a trier tube

(refer to BS 809) or other hollow tube. This sample is gently melted under nitrogen and mixed before determination of, e.g., peroxide value (PV).

Emulsified and plasticized fat spreads are already homogeneous, easily sampled, and have published standard methods for their processing (IUPAC 1979). For other food products, it is important that a homogeneous sample of the food is obtained by careful blending.

The ester bonds and unsaturation in lipids are vulnerable to hydrolysis or oxidation, so that any manipulations or procedures (and subsequent storage) must take account of these facts. Lipid hydrolysis is mainly an issue of residual active lipase enzymes or microbial contamination. If fresh samples become bruised or are frozen, tissue damage releases hydrolytic enzymes. Lipid oxidation is a worse problem, particularly with highly unsaturated marine oils. Susceptibility to oxidation is enhanced in the presence of even trace ( $\text{mg kg}^{-1}$ ) amounts of copper, nickel, and iron. Lipid oxidation is a free-radical process catalyzed by light of wavelengths below 400 nm. Therefore, care should be taken during process and storage of highly unsaturated samples.

The extraction of the lipids is an important procedure in analysis. The procedure used will depend upon the structure of the matrix containing the lipid and the composition and physical state of the lipids. For instance, the matrix may contain emulsified water as in a low-calorie fat spread. This can generally be separated into fat and aqueous layers by warming at 45°C. Fresh meat contains a significant amount of water and will require homogenization in mixtures of methanol and chloroform, followed by partition of the extract against salt solution. The lipids are found in the lower chloroform phase. However, the partial solubility of some of the more polar lipids needs to be considered as does partition at the water/organic interface. These can result in quantitative losses if care is not exercised when collecting the organic extracts.

Fresh green vegetables should first be homogenized in 2-propanol to deactivate lipases, which otherwise might be activated by some organic solvents. The solids are further extracted with chloroform/2-propanol. The lipids in both extracts are combined.

The fat in coffee whiteners is encapsulated within a water-soluble carbohydrate matrix, making it unavailable to extracting solvents. Dry powder extraction will typically achieve only 1–5% of the fat. To obtain the total fat:

1. Disperse powder in warm (40°C) water and follow by solvent extraction. This will extract 95–100% of the fat in a gentle way. Or

2. Strong acid hydrolysis (Weibull–Berntrop procedure) followed by solvent partition. This is aggressive and can lead to fat oxidation.

The lipids in cereal flours (e.g., wheat) are highly structural and difficult to extract quantitatively:

1. Total lipid is estimated by gas chromatography after acid hydrolysis of the flour in the presence of a known amount of heptadecanoic acid internal standard.
2. Nonstarch lipid is extracted from flour with cold, water-saturated, *n*-butanol under standard conditions.
3. The residue from extraction (2) can be used to obtain the starch lipid by resuspending the solids in methanol, centrifuging, and recovering the solids.
4. The final solids are extracted with water-saturated *n*-butanol at 95–100°C for 60 min, repeating this five times to yield the starch lipids.

Marine organisms present particular problems due to high levels of lipid unsaturation together with high water and protein contents. The following procedure can provide a protected environment under a carbon dioxide blanket: The tissue is cut into small pieces at water/ice temperatures and homogenized in a kitchen food processor, together with enough pelleted solid carbon dioxide to fully freeze the material to  $\sim -40^\circ\text{C}$ . The powdered sample is extracted with chloroform–methanol mixtures that have been pre-cooled with cardice to  $-50^\circ\text{C}$ . This whole mixture is left to reach ambient temperature with occasional mixing and then processed using precautions against lipid oxidation.

## Standard Methods and Quality Assurance

Analytical precision and accuracy require standardization of methodology. Techniques require proving with standards and reference materials and should be tested regularly by blind inclusion of these materials into normal analytical practice. In this way the statistical variability of results can be built up providing improved confidence. The analyst can draw upon standard methodology from international authorities such as the International Union of Pure and Applied Chemistry (IUPAC), British Standards Institution, American Oil Chemists' Society (AOCS), International Organization for Standardization (ISO), Deutsche Gesellschaft für Fettwissenschaft, and Association of Official Analytical Chemists (AOAC). The priority is to standardize the chosen



methodology in the laboratory and ensure consistent practice.

Quality assurance methods commonly used are for the determination of free fatty acid (FFA), PV, iodine value (IV), anisidine value (AnV), Rancimat induction time, and *trans* value (TV).

Crude oils may be up to 15% FFA, while refined oils will be <0.1%. Measurement of FFA is normally done by titration of an ether-ethanol solution of the fat with standardized aqueous sodium or potassium hydroxide, in the presence of phenolphthalein indicator (see ISO 660: 1983). This method is very accurate, using the molecular weight of oleic acid (282) for all calculations. Accuracy is improved by using the average molecular mass of the fatty acids in the fat, calculated from the FA composition (e.g., for palm oil 256 is used, for palm kernel/coconut oil 200 is used). The results are expressed either as acid value; the number of milligrams of potassium hydroxide required to neutralize 1 g of the fat, or, FFA%; the percentage concentration of oleic acid equivalent to the free acids present.

PV is normally assessed using the Wheeler procedure: AOCS (1984) method CD 8-53 (73); and IUPAC method 2.501. The determination should ideally be carried out under tungsten lighting as ultraviolet light liberates iodine from potassium iodide in solution, in the presence of oxygen, creating high blanks. The liberated iodine is titrated with standard sodium thiosulfate. Results are expressed in Standard International (SI) Units. Some laboratories express the result in different units, which may be related to the SI units by the factors shown in the Table 2.

The interpretation of results requires some knowledge of the sample and sample processing history. For instance, low values do not necessarily indicate 'good' quality fats; organic lipid peroxides are not stable at temperatures above 50°C so that in deep fat fryers operating at up to 180°C, one would not expect to find much peroxide. The half-life of peroxides can be extremely short under frying conditions and the sample requires a standard amount of time after removal from the fryer (typically 30 min) before analysis providing more meaningful results.

IV measures the level of unsaturated components in a fat equivalent to a weight of iodine reacting with 100 g of the sample. The sample is reacted under standard conditions with a known amount of iodine (iodine trichloride as Wijs solution) in the dark (BS684 section 2.13, 1981; and ISO 3961, 1979). Unreacted iodine is liberated chemically and back-titrated with sodium thiosulfate to yield an IV. A minimum IV can be calculated from the FA composition determined by gas chromatography. The titrated IV may be higher due to the presence of other iodine-positive components. The technique assays unreacted reagent therefore, it becomes essential that the amount of Wijs reagent used is very accurately measured. A purpose-designed, exact-volume dispenser is better than using a burette. Precision of technique is very important to the reproducibility of IV measurement. A common reason for error is incorrect ratio of fat and reagent. For the reaction to be complete and the back-titration to yield a satisfactory value, there must be a minimum excess level of Wijs reagent of ~50%. The weight of sample fat used for the estimation depends upon the expected IV and may be calculated from the equation below. For a totally unknown sample this amount may be difficult to judge requiring a test IV to be carried out:

$$\text{Weight of sample (gm)} = \frac{12.69 \times M \times V_R}{\text{approx IV}}$$

where  $M$  is the molarity of sodium thiosulfate and  $V_R$  the volume of Wijs reagent used.

Alternatively, where an estimated IV is available, the weight of sample (25 ml Wijs reagent used) may be obtained from Table 3.

AnV measures the level of carbonyl compounds, particularly aldehydes, in a fat (IUPAC method 2.504). The fat, in 2,2,4-trimethylpentane solution, is treated with *p*-anisidine reagent prepared with anhydrous acetic acid. The level of reaction products is determined spectrophotometrically at 350 nm. The reagent is particularly toxic and skin contact must be

**Table 2** Relationship between SI units and other units used for peroxide value

Units of PV	Factor
Millimole active oxygen per 2 kg lipid – SI units	× 1.0
Milliequivalents of active oxygen per kilogram lipid	× 1.0
Millimole active oxygen per kilogram lipid (Lea value)	× 0.5
Milligram active oxygen per kilogram lipid	× 8.0

**Table 3** Weight of sample to be used for IV determination (25 ml Wijs reagent used)

Expected IV	Weight sample (g)	Weighing accuracy (± g)
0–5	4.0	0.0005
6–20	1.0	0.0005
21–60	0.4	0.0005
61–80	0.25	0.0001
81–130	0.15	0.0001
> 130	0.10	0.0001

avoided absolutely. Good-quality fats should show values of less than 10.

The Rancimat induction time is the time taken for a fat under test conditions to reach an approximately maximal rate of oxidation. The test is a variant of the Swift active oxygen method using a device called the Rancimat (Metrohm, Switzerland). A constant rate of air is bubbled through the fat sample held in a thermostatically controlled cell normally at 100°C. The data produced give a measure of potential shelf-life for a fat. At 100°C, a typical biscuit shortening gives a time of 50 h, cocoa butter >140 h, and sunflower oil ~12 h. The technique is reliable. All glassware must be very clean and a blank (i.e., without fat) analyzed. Samples are determined in duplicate and any showing large deviations repeated.

As part of the nutritional analysis of fat in food, it may be important to determine the level of *trans* unsaturation (TV) present. The measurement of total *trans* fatty acids is conveniently done by infrared (IR) spectroscopy but needs careful calibration with known standard solutions of pure trielaidin and tristearin (99%). Above 10% *trans*, the accuracy is good and the measurement is done on the TAGs. Below 10% *trans*, interference is common, reducing accuracy and TV is measured either by IR of the methyl esters, or, preferably by gas chromatography. Calibration in this case must also use methyl esters. The *cis* absorption is small in IR, but isolated *trans* double bonds show a relatively strong IR band absorbing at  $\sim 967\text{ cm}^{-1}$ . The AOAC published a method (AOAC 28.052–28.067) for TV. Isomerized or oxidized fats cause errors because conjugated species show strong absorption close to the isolated *trans* bond and interfere with the correct allocation of baseline. Samples and standards are usually analyzed in solution ( $200\text{ mg ml}^{-1}$ ), traditionally in carbon disulfide. A calibration line is drawn for different amounts of trielaidin mixed with tristearin, a solution of pure tristearin being in the reference beam at all times. The line should be straight. For each standard and sample record, a straight baseline is drawn joining the minima at 1000 and  $925\text{ cm}^{-1}$ . TV of samples are read from the calibration graph.

For TV analysis by gas-liquid chromatography (GLC) of fat to methyl esters (FAMEs) the official AOCS method prescribes the cyanopropylpolysiloxane phase SP2340. The column should be  $60\text{ m} \times 0.25\text{ mm}$  fused silica with a 1 m retention gap of silanized fused silica with a film thickness of  $0.2\text{ }\mu\text{m}$ . Such columns will provide adequate separation of *cis* and *trans* FAME for most situations involving food fats (except fats containing hydrogenated fish oils). However, there is peak overlap and for exacting analyses, where information is required on the iso-

meric nature of the *trans* FAME, even these columns are not adequate. For separation that is more complete, a column of SP2560,  $0.2\text{ }\mu\text{m}$  film on  $100\text{ m} \times 0.25\text{ mm}$  fused silica should be used. In both column cases, the carrier gas can be helium (linear velocity  $20\text{ cm s}^{-1}$ ) or hydrogen ( $40\text{ cm s}^{-1}$ ). Column temperature should be optimized, using standards, between 175°C and 200°C isothermal. Otherwise, faster results can be obtained by programming the temperature between 100°C and 225°C with some sacrifice to optimum resolution. The samples should be injected using the on-column technique in preference to split injection. *Cis/trans* ratios calculated from this type of chromatography will, in general, be more accurate than those obtained via IR techniques. However, this is true only below 15% *trans* content.

## Additional Analyses

GLC is applied to the analysis of fats and fat components in several ways. Typical data for GLC is shown in Table 1. This includes the analysis of FA composition after derivatizing the FAME (or propyl esters for butter fat) and the direct analysis of TAG composition. GLC is used for the analysis of sterols, hydrocarbon contaminants, pesticides, and volatile products produced during fat processing and refining.

For general FAME analysis it is adequate for most work to use wide bore ( $0.53\text{ mm ID}$ ) silica (quartz) columns. A  $30\text{ m}$  length with bonded FFAP phase (film thickness  $1.0\text{ }\mu\text{m}$ ) produces a robust column of excellent long-term stability. Other phases are suitable, such as silar 5C, silar 10C, CPSil88. Better resolution may be obtained with columns of  $0.32$  and  $0.25\text{ mm ID}$  (film thickness  $0.5\text{ }\mu\text{m}$ ) but sample loading capacity will be lower and they might be more prone to damage by mishandling (see Figures 1 and 2).

Capillary GLC columns are excellent for the analysis of TAGs. It is essential that a precolumn (retention gap) of uncoated, silanized silica tube ( $1.0\text{ m} \times 0.53\text{ mm ID}$ ) be attached to the analytical column. The injection technique of choice for both FAME and TAGs should be on-column. Split injection almost always leads to quantitative errors in these analyses. Both helium and hydrogen are adequate carrier gases. A  $7\text{--}10\text{ m}$  bonded phase (OV1, film thickness  $0.1\text{ }\mu\text{m}$ )  $0.53\text{ mm ID}$  column is used with a temperature stability of at least 350°C, but aluminum-clad columns are avoided. A temperature program is used, starting at  $\sim 100^\circ\text{C}$  rising to  $350^\circ\text{C}$  at  $\sim 10^\circ\text{C min}^{-1}$ . While the carrier gas velocity for

FAME columns should be optimal for maximum resolution, it is advantageous to have the carrier gas velocity for TAG analysis at twice optimal to minimize on-column polymerization.

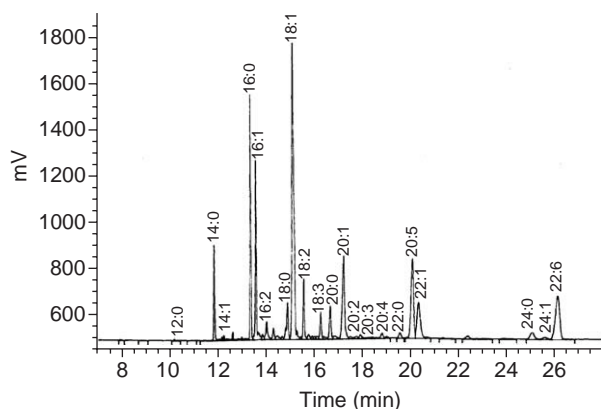
Free and esterified sterols generally occur together in natural oils but are analyzed after hydrolysis as part of the nonsaponifiable fraction. To reduce chromatographic losses the sterols are derivatized to the trimethyl silyl ethers prior to GLC. Trimethylsilylimidazole is used for the preparation of TMS derivatives. A 25 m  $\times$  0.53 mm (or 0.34 mm) fused silica column with a 1.0  $\mu$ m bonded film of 5% phenyl-methyl silicone, fitted with a 1 m  $\times$  0.53 mm silanized retention gap, is adequate for good

resolution and reasonable quantitative analysis. Samples are injected using cold on-column injection. The column start temperature is 275°C with a 2 min hold, rising to 325°C at 2°C min<sup>-1</sup>.  $\alpha$ -Cholestane is used as an internal standard in this analysis.

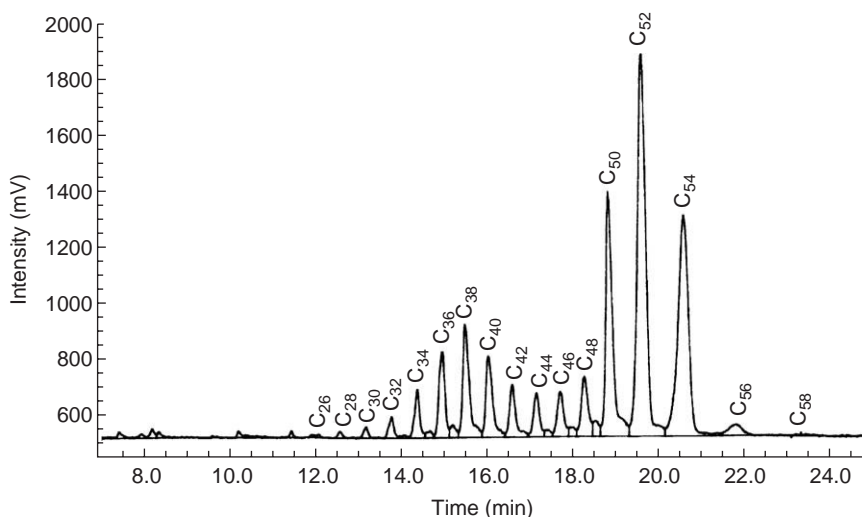
High-performance liquid chromatography (HPLC) is routinely used for the compositional analysis of lipid classes, TAGs, and tocopherols. However, there can be problems for the quantitative detection of lipids other than tocopherols because most lipids do not have a suitable chromophore, and therefore cannot be detected spectrophotometrically. Evaporative technology such as nebulizing 'mass' detectors and the quartz/flame ionization transport system has to be employed.

For lipid class separation a 10 cm  $\times$  4 mm plain silica (3  $\mu$ m particle size) column is adequate, providing separation across the range from sterol esters and TAGs through to phospholipids and galactolipids in  $\sim$ 30 min.

For TAG separation by reversed-phase HPLC, two 10 cm Lichrosphere RP-18, 5  $\mu$ m columns are linked together. Ideally, column temperature should be kept constant at 25°C. It is advisable to analyze as many known materials as possible and construct a graph of equivalent chain length by plotting total carbon number against elution time. Peak types overlap because a double bond is almost equivalent to a chain shortening of two carbon atoms. The rather complex separation achieved yields much information about the TAG structure and is a composite of chain length and unsaturation, being more complex than the high-temperature polar phase GLC procedure. These data can be used to 'finger print' fats.



**Figure 1** FAME gas chromatography: fatty acid methyl esters of a fish oil; 30 m  $\times$  0.53 mm FFAP (see text for details). Peak names are shown as numbers separated by a colon. The number before the colon is the chain length, that after is the number of double bonds.



**Figure 2** Carbon number gas chromatography: triglycerides of a mixture of 1:1 butter fat and cocoa butter; 8 m  $\times$  0.53 mm OV1 (see text for details). Peak names are shown as carbon numbers which are the sum of the carbons in the three fatty acid chains, excluding the glycerol carbons.

For chocolate fats in particular, it is often necessary to know the composition of the TAGs based simply upon the level and position of unsaturation. The SOS (1-saturated, 2-monounsaturated, 3-saturated) type TAGs are responsible for the particular physical properties of cocoa butter and its equivalents. The information can be obtained from the reversed-phase method mentioned above, but a simpler profile is obtained by using 5% silver nitrate-impregnated silica as a 25 cm column.

Tocopherols are analyzed in their unesterified form and separations are very easy by HPLC. The use of fluorescence detection provides very accurate results. The oil is diluted with eluent (3% tetrahydrofuran in isooctane) to  $\sim 10 \text{ mg ml}^{-1}$ . Pure standards are analyzed ( $2 \mu\text{g ml}^{-1}$ ) to obtain retention data, while pure  $\alpha$ -tocopherol is used as an external standard for every two samples. A 15 cm column of 5  $\mu\text{m}$  Chromegasphe is equilibrated with the isocratic solvent at  $1.5 \text{ ml min}^{-1}$ . Injection volume is 50  $\mu\text{l}$ . The spectrofluorimeter excitation is 290 nm and emission is 330 nm. Full separation should be readily achieved, but the column will require conditioning by pumping chloroform-methanol (2:1; v/v) for 10 min every 30 samples. After this conditioning, the column activity must be regenerated by pumping normal eluent for at least 20 min.

**See also:** **Carbohydrates:** Starch. **Derivatization of Analytes.** **Extraction:** Solvent Extraction Principles. **Fluorescence:** Food Applications. **Gas Chromatography:** High-Temperature Techniques. **Lipids:** Fatty Acids; Polar Lipids. **Liquid Chromatography:** Food Applications. **Mass Spectrometry:** Food Applications. **Quality Assurance:** Quality Control; Internal Standards. **Sample Handling:** Comminution of Samples. **Sampling:** Theory.

## Further Reading

- Firestone D (Associate Chapter Editor) (1997) AOAC *Official Methods of Analysis chapter 28, Oils and Fats*, p. 507. Washington: AOAC.
- Duchateau GSMJE, van Oosten HJ, and Vasconcellos MA (1996) Analysis of *cis*- and *trans*-fatty acid isomers in hydrogenated and refined vegetable oils by capillary gas-liquid chromatography. *JAOCS* 73(3): 275–282.
- Gurr MI and James AT (1980) *Lipid Biochemistry: An Introduction*, 3rd edn. London: Chapman & Hall.
- Hammond EW (1993) *Chromatography for the Analysis of Lipids*. Boca Raton, FL: CRC Press.
- Hamilton RJ and Hamilton S (eds.) (1992) *Lipid Analysis, A Practical Approach*. New York: IRL Press, Oxford University Press.
- Standard Methods for the Analysis of Oils, Fats and Derivatives* (1987) 7th Revised and Enlarged Edition, International Union Pure and Applied Chemistry. Oxford: Blackwell Scientific Publications.

## Fruits and Fruit Products

**R J Pither**, Campden and Chorleywood Food Research Association, Chipping Campden, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1525–1531, © 1995, Elsevier Ltd.

well as being good sources of dietary fiber, mineral salts, and vitamins, especially vitamin C. The organic acids found in fruits essentially balance with the sugars to give them their characteristic flavors and are therefore also a vital component.

## Introduction

Fruits have long been valued as a foodstuff owing to their pleasant flavor and aroma, and attractive appearance; however, the use of fruits as a staple in the diet has only become possible in the past century through the development of canning and other bulk preservation techniques, and improvements in transportation and storage systems. The increased awareness of the importance of a nutritionally balanced diet has also contributed to the increased consumption of fruit and fruit products.

Fruits are an important low-fat energy source containing varying proportions of sugars and starches as

## Major Components

The major component of fresh fruit is water, with levels ranging from 65% (in avocado) to 93% (e.g., in melons, strawberries). Dried fruit products can contain up to 25% water while fruit juices can vary between 85% and 93%. The second most important component in terms of percentage content is carbohydrate, with reducing sugars, primarily a mixture of glucose and fructose, being the main soluble carbohydrate. Most fruits contain relatively low levels of sucrose; however, it is the primary sugar in peaches, apricots, bananas, pineapple, and certain citrus fruits. Starch, cellulose, hemicellulose, and pectin are also found in varying proportions in fruit.

For chocolate fats in particular, it is often necessary to know the composition of the TAGs based simply upon the level and position of unsaturation. The SOS (1-saturated, 2-monounsaturated, 3-saturated) type TAGs are responsible for the particular physical properties of cocoa butter and its equivalents. The information can be obtained from the reversed-phase method mentioned above, but a simpler profile is obtained by using 5% silver nitrate-impregnated silica as a 25 cm column.

Tocopherols are analyzed in their unesterified form and separations are very easy by HPLC. The use of fluorescence detection provides very accurate results. The oil is diluted with eluent (3% tetrahydrofuran in isooctane) to  $\sim 10 \text{ mg ml}^{-1}$ . Pure standards are analyzed ( $2 \mu\text{g ml}^{-1}$ ) to obtain retention data, while pure  $\alpha$ -tocopherol is used as an external standard for every two samples. A 15 cm column of 5  $\mu\text{m}$  Chromegasphe is equilibrated with the isocratic solvent at  $1.5 \text{ ml min}^{-1}$ . Injection volume is 50  $\mu\text{l}$ . The spectrofluorimeter excitation is 290 nm and emission is 330 nm. Full separation should be readily achieved, but the column will require conditioning by pumping chloroform-methanol (2:1; v/v) for 10 min every 30 samples. After this conditioning, the column activity must be regenerated by pumping normal eluent for at least 20 min.

**See also:** **Carbohydrates:** Starch. **Derivatization of Analytes.** **Extraction:** Solvent Extraction Principles. **Fluorescence:** Food Applications. **Gas Chromatography:** High-Temperature Techniques. **Lipids:** Fatty Acids; Polar Lipids. **Liquid Chromatography:** Food Applications. **Mass Spectrometry:** Food Applications. **Quality Assurance:** Quality Control; Internal Standards. **Sample Handling:** Comminution of Samples. **Sampling:** Theory.

## Further Reading

- Firestone D (Associate Chapter Editor) (1997) AOAC *Official Methods of Analysis chapter 28, Oils and Fats*, p. 507. Washington: AOAC.
- Duchateau GSMJE, van Oosten HJ, and Vasconcellos MA (1996) Analysis of *cis*- and *trans*-fatty acid isomers in hydrogenated and refined vegetable oils by capillary gas-liquid chromatography. *JAOCS* 73(3): 275–282.
- Gurr MI and James AT (1980) *Lipid Biochemistry: An Introduction*, 3rd edn. London: Chapman & Hall.
- Hammond EW (1993) *Chromatography for the Analysis of Lipids*. Boca Raton, FL: CRC Press.
- Hamilton RJ and Hamilton S (eds.) (1992) *Lipid Analysis, A Practical Approach*. New York: IRL Press, Oxford University Press.
- Standard Methods for the Analysis of Oils, Fats and Derivatives* (1987) 7th Revised and Enlarged Edition, International Union Pure and Applied Chemistry. Oxford: Blackwell Scientific Publications.

## Fruits and Fruit Products

**R J Pither**, Campden and Chorleywood Food Research Association, Chipping Campden, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1525–1531, © 1995, Elsevier Ltd.

well as being good sources of dietary fiber, mineral salts, and vitamins, especially vitamin C. The organic acids found in fruits essentially balance with the sugars to give them their characteristic flavors and are therefore also a vital component.

## Introduction

Fruits have long been valued as a foodstuff owing to their pleasant flavor and aroma, and attractive appearance; however, the use of fruits as a staple in the diet has only become possible in the past century through the development of canning and other bulk preservation techniques, and improvements in transportation and storage systems. The increased awareness of the importance of a nutritionally balanced diet has also contributed to the increased consumption of fruit and fruit products.

Fruits are an important low-fat energy source containing varying proportions of sugars and starches as

## Major Components

The major component of fresh fruit is water, with levels ranging from 65% (in avocado) to 93% (e.g., in melons, strawberries). Dried fruit products can contain up to 25% water while fruit juices can vary between 85% and 93%. The second most important component in terms of percentage content is carbohydrate, with reducing sugars, primarily a mixture of glucose and fructose, being the main soluble carbohydrate. Most fruits contain relatively low levels of sucrose; however, it is the primary sugar in peaches, apricots, bananas, pineapple, and certain citrus fruits. Starch, cellulose, hemicellulose, and pectin are also found in varying proportions in fruit.

In general, fruits are poor sources of protein and lipid, the major exceptions to this being avocado and olives, which can contain ~11% and between 10% and 40% lipid, respectively.

Most fruits are particularly rich in organic acids, the principal ones being citric, tartaric, and malic acids; however, the range is extensive.

By far the most important acid present in fruits is L-ascorbic acid, vitamin C. Fruits, together with vegetables, are the main sources of this vitamin in the human diet. Its importance was first reported by a Hungarian physician in 1732, who documented the relationship between citrus fruit and antiscorbutic activity. The concentration of L-ascorbic acid can vary between only 2 and 30 mg per 100 g in apples and pears, to over 1000 mg per 100 g in rose hips and certain cherry varieties. The majority of fruits, however, including citrus, have values in the region of 30–100 mg per 100 g. Today, orange juice alone provides 60% of the US Recommended Daily Allowance (RDA) for vitamin C in the American diet.

Several fruits are good sources of  $\beta$ -carotene, for example, apricots, peaches, melons, and cherries, while a range of B vitamins (pantothenic acid, folic acid, biotin, thiamine, and riboflavin) can also be found. In general, however, the concentrations of the latter vitamins are low when compared with other animal and vegetable sources.

Fruits and fruit products contain a wide range of mineral salts including potassium, calcium, magnesium, phosphorus, and the trace elements, copper, zinc, iron, and manganese, with potassium being the most abundant of these in a majority of fruits. Whereas a serving of a fruit or fruit product may provide only a small percentage of the RDA of many of these mineral nutrients, they are still good sources when calculated in terms of their calorific contribution.

## Sampling and Preparation for Analysis

Sampling and preparation techniques are the first, and some of the most important, steps toward achieving an accurate analytical result. There are many sources of error in any analytical technique, e.g., analyst's error, contaminated reagents or equipment, inappropriate methodology, and calculation or data handling error. Most of these can be controlled or detected by the use of standards, blanks, reference samples, and other quality control procedures. An unrepresentative sample, however, may not be so easily detected. Foodstuffs are naturally heterogeneous materials and, therefore, obtaining a homogeneous and representative sample is a major problem. For most fruits and fruit products sampling

will be carried out from bulk materials. For such sampling it is important, first, to identify the size of the population or batch, then to select and remove gross samples from this population, and finally to reduce these samples to a suitable laboratory sample size, all whilst maintaining as far as possible the representative nature of the sample. This is generally achieved by selecting several samples from the bulk, either randomly or according to a plan, and thoroughly mixing these to form a composite sample from which a subsample for laboratory analysis can be taken.

There is much information available on sampling plans for foodstuffs. The Codex Alimentarius Commission (document CAC/RM 42-1969) deals with sampling plans for prepackaged foods, and there are standards available from the International Standards Organization (ISO) for sampling of various foodstuffs including fresh fruits and vegetables (ISO 874:1980). Details of the mathematical principles involved are dealt with in many books on statistics.

Due both to the economic aspects of statistical sampling plans and the difficulties encountered owing to the natural variation in foodstuffs, most analyses, however, are carried out on randomly selected samples.

The method of preparation used to produce a homogeneous laboratory sample varies considerably with food type. Examples of laboratory sampling and preparation procedures for fresh fruit products can be seen in **Table 1**.

There is a wide range of fine-slicing devices available that are suitable for homogenization of fruit and fruit products: liquidizers, domestic-type food processors, Waring blenders. Care must be taken to ensure that excessive heat generation does not occur during the homogenization process that can lead to loss of moisture, and that the sample is transferred immediately to a suitable storage vessel. These should be of glass or rigid plastic, have a tightly fitting lid and be easily labeled.

Stabilization of the product after preparation is also an important consideration, with the sample susceptible to change due to enzymatic, oxidative, or microbial action. In most cases for fruit and fruit products the most practical way of achieving this is by rapid freezing and storage at  $-20^{\circ}\text{C}$ . Inactivation of enzymes can also be achieved by treatment with 80% methanol or ethanol, or ice-cold 10% perchloric or trichloroacetic acid, while oxidation may be reduced if reducing agents such as hydrogensulfite or dithiothreitol are added during the homogenization process. The most appropriate method will depend on the analyses to be performed and the exact nature of the sample.



**Table 1** Preparation procedures for fresh fruit products

<i>Fruit type</i>	<i>Preparation technique</i>
Apples, pears, etc.	Select 10–15 fruits Remove extraneous matter: leaves, stalks, etc. Cut into quarters along the longitudinal axis Take a selection of the quarters and mechanically finely chop and mix
Plums, peaches, apricots, etc.	Select 15–20 fruits Remove extraneous matter and destone Proceed as for apples, pears, taking care to collect any released juices
Berries, cherries, olives, etc.	Remove a random selection of individual berries from the sample (minimum of 100 g) Remove stones (preferably using a cherry stoner) while carefully avoiding loss of juice and fruit flesh Mechanically finely chop and mix
Exotic fruits: bananas, avocados, kiwi, etc.	Remove inedible portions: skin, peel, stones, etc., and extraneous matter Cut into quarters Take a selection of these quarters and mechanically chop and mix
Citrus fruits: oranges, grapefruit, lemon, lime, etc.	

## Analytical Methods

### Moisture/Total Solids

Water is present in foods as both bound and available free water, the quantification of which is dependent on the methodology used.

The standard procedure for measuring moisture in fruits and fruit products is by weight loss on drying. References to this methodology can be found in many books; however, they all follow the same basic technique, i.e., Association of Official Analytical Chemists (AOAC) methodology:

Accurately weigh a sample into a large flat-bottomed dish and dry at 70°C under vacuum (pressure  $\leq 100$  mmHg) until consecutive weighings vary by less than 3 mg.

This method has the disadvantage of a considerable drying time (usually 4–6 h) and many rapid techniques are now available. Infrared drying ovens and microwave ovens have been shown to produce comparable results to vacuum oven drying methods for various fruit products. Nondestructive techniques such as low-resolution nuclear magnetic resonance

(NMR) and near-infrared (NIR) spectroscopy can also be used for moisture determination and are of particular use for online checking during food manufacture and processing.

### Carbohydrates

**Sugars** A wide range of methods is available for the determination of sugars in fruit and fruit products. These include relatively simple refractometry and hydrometry techniques, polarimetry, copper(II) reduction, enzymatic/color detection, and various chromatographic techniques. These vary greatly in their sensitivity, specificity, and in the level of equipment and expertise required, and these considerations generally dictate the method chosen.

**Refractometry** This allows for the rapid, reasonably accurate estimation of sugar levels based on the measurement of soluble solids. Correction tables exist allowing for other dissolved substances, e.g., acids, and for changes in temperature; however, many modern refractometers make automatic allowance for this.

**Hydrometry** This technique relies on the measurement of relative density but is generally less accurate than refractometry.

**Polarimetry** The determination of sucrose and reducing sugars in fruit and fruit products by polarimetry is an AOAC recommended method. The polarimeter measures the rotation of plane polarized light caused by a solution containing an optically active compound. Separate measurements for the quantification of sucrose and reducing sugars can be made with the Clerget–Hertzfeld double-polarization method.

**Copper(II) reduction** These methods rely on the ability of certain sugars, i.e., all monosaccharides and the disaccharides maltose and lactose, to act as reducing agents, owing to the presence of free aldehyde or ketone groups in their structure. Fehling's solution consists of alkaline copper(II) tartrate and it is the ability of reducing sugars to convert this to insoluble copper(I) oxide that forms the basis of a number of analytical procedures. The most widely accepted of these are the Lane and Eynon method, which has a titrimetric endpoint using a redox indicator and the Munson and Walker method (AOAC) where the copper(I) oxide is filtered off and weighed. The Luff–Schoorl method uses alkaline carbonate solution and sodium citrate and employs an iodimetric determination of the excess copper(II) with

thiosulfate. This method is recommended by the Codex Alimentarius Commission for the analysis of fruit juices.

All of the copper reduction methods can be used to determine sucrose, after inversion (conversion to glucose and fructose), as well as reducing sugars. These methods are therefore particularly suitable for fruits and fruit products.

**Liquid chromatography (LC)** LC is now a popular and useful analytical technique that has the advantage that different sugars can be determined simultaneously. The separation technique can be carried out using strong cation-exchange chromatography, or a modified silica gel-amino column, with detection by refractometry. This method does have certain disadvantages in that initial equipment costs are high and, owing to the sensitivity of the detectors, they are best when used exclusively and continuously for the same analysis. The latest development in LC analysis of carbohydrates uses improved anion-exchange techniques combined with pulsed amperometric detection, which is both more sensitive and specific than the traditional LC methodology.

**Gas chromatography (GC)** This technique has the advantage of allowing food sugars to be determined with very high sensitivity and provides an AOAC recommended method. The sample sugars have to be converted to volatile and heat-stable derivatives before analysis and this technique is not widely used in routine analysis because of the accuracy and relative simplicity of alternative methods.

**Enzymatic reactions** Enzymatic reactions have the advantage of being specific; for example, hydrolysis of sucrose is more accurate using invertase than an acid. A wide range of sugar analysis kits is available commercially; for example, sucrose/glucose/fructose kits are particularly suitable for fruit and fruit products. They, in common with the copper reduction methods, have the advantage of simple sample preparation techniques. The enzymatic reactions involved cause the reduction of  $\text{NADP}^+$ , the concentration of which is determined by spectrophotometry at 340 or 366 nm. Great care is needed in their use because of the very precise quantities involved; however, the results obtained compare favorably with more traditional methods and the methods are now widely accepted.

**Starch** The determination of starch in fruits and fruit juices is normally achieved by the hydrolysis of the starch to glucose and the subsequent analysis of glucose levels. Acid hydrolysis methods can be

subject to error because of the possibility of hydrolysis of nonstarch polysaccharides or the destruction of glucose. Enzymatic hydrolysis of the starch can also be employed using amyloglucosidase, and commercial kits are available that allow for concurrent analysis of the prepared sample for glucose content, with and without the addition of the amyloglucosidase.

**Dietary fiber** The determination of dietary fiber has increased greatly in importance in recent years; however, due partly to lack of a clear definition for dietary fiber, several standard methodologies exist.

**AOAC enzymatic gravimetric method** The samples are dried, with fat extracted if necessary, gelatinized with heat-stable  $\alpha$ -amylase, and enzymatically digested with protease and amyloglucosidase to remove protein and starch. The soluble dietary fiber is then precipitated using ethanol, the residue filtered and washed, and corrections are made for indigestible protein and ash. This method is widely used in the EC and the USA.

**Englyst method** Samples are defatted if required and any starch present is gelatinized and removed by enzymatic digestion. The remaining polysaccharides are hydrolyzed by sulfuric acid and the resulting sugars can be measured either colorimetrically or by using gas chromatography. This method can be used to determine both total nonstarch polysaccharides and resistant starch, and is a recognized method in the UK and Canada.

**Pectin** There are many methods for the estimation of pectic substances. The most commonly used for fruit products are those based on aqueous extraction, saponification with cold alkali, acidification and the precipitation of the pectin as the calcium salt; alternatively, precipitation using acetone or alcohol can be utilized.

### Minerals: Sodium and Potassium

The AOAC and Codex recommended methods for determination of potassium and sodium are by flame atomic emission spectrometry. These methods have the advantage of quick and simple sample preparation, aqueous dilution of sample and filtration, with rapid, reasonably accurate detection. The samples can be aspirated directly into the flame and the readings quantified by comparison with a range of known standards. Atomic absorption spectrometry can also be used, giving increased accuracy and low-level sensitivity and a far wider range of detectable

minerals. It has the disadvantage of higher equipment costs and requires a dry-ashing sample preparation technique, which, though simple, can take a relatively long time; however, if trace metals or other minerals are also being determined the ashed sample can be used for all analyses.

### **Vitamin C**

There are two standard (AOAC) procedures for vitamin C (ascorbic acid) determination in fruits and fruit products. The sample preparation is common to both and involves maceration/dilution of the sample in a stabilizing solution such as 5% metaphosphoric acid or trichloroacetic acid followed by filtration. The ascorbic acid can then be determined by titration with 2,6-dichloroindophenol, where the ascorbic acid reduces the redox indicator dye to a colorless solution, or by fluorimetric detection, in which the ascorbic acid is oxidized to dehydroascorbic acid, which then reacts with *o*-phenylenediamine to produce a fluorophore. The latter method has the advantages that it is suitable for colored solutions and can also be used to measure levels of naturally occurring dehydroascorbic acid as well as ascorbic acid. Many other methods are available for determination of vitamin C, with the use of LC techniques currently the subject of much interest. Reversed-phase LC techniques can be used to determine both dehydroascorbic acid, ascorbic acid and their isomers.

### **Phosphate**

The sample is first ashed and dissolved in hydrochloric acid; the phosphate can then be determined titrimetrically or colorimetrically. The most widely accepted method is a colorimetric method using Mison's reagent, which is based on the reaction between the acidified sample solution and an acid reagent containing molybdic acid and vanadic acid, to produce the stable orange-yellow vanadomolybdophosphoric acid.

### **Organic Acids/Acidity**

For routine analysis a measurement of titratable acidity is usually sufficient, with the acidity of the fruit/fruit product calculated as the predominant acid, e.g., as citric acid in citrus fruits, as malic acid in apples. The sample is diluted in distilled water and titrated against dilute sodium hydroxide to either a phenolphthalein endpoint or to pH 8.10. Volatile acidity, generally expressed as acetic acid, can be measured by distilling the sample using a steam distillation apparatus, with titration of the distillate as above. Individual acids can be determined using a

range of techniques including chromatographic and titrimetric methods fully described by the AOAC, and enzymatic methodologies with commercial enzyme-based test kits now available for many fruit acids. LC techniques are also available facilitating identification and quantification of a range of organic acids.

## **Quality Assurance of Fruits and Fruit Products**

Many analyses and tests are carried out on fruits and fruit products as a means of quality assurance and many regulations and specifications exist in relation to this. These include the Joint Food and Agriculture Organisation of the United Nations/World Health Organization (FAO/WHO), Food Standards Programme-Codex Alimentarius, the US Department of Agriculture (USDA) standards for grades of processed fruit products, the UK Fruit Juices and Fruit Nectars Regulations, and EC directives.

The range of chemical analyses carried out to determine compliance with quality and specification requirements can include: pH, titratable, and volatile acidity; total solids, soluble solids ( $^{\circ}$ Brix), water-insoluble solids, sugars; acid-insoluble ash, alkalinity of ash, mineral impurities, trace metals (tin, copper, zinc, iron, aluminum), heavy metals, e.g., lead; ethanol; preservatives such as sulfur dioxide, sorbic acid, benzoic acid, and benzoates.

Many specifications also define standards for physical characteristics, e.g., color, flavor, appearance, and defects. These are generally assessed by trained personnel; however, a range of equipment is available for color measurement.

Many of the chemical analyses used for quality assurance have already been discussed; however, there are other important tests worth brief consideration.

### **Essential Oils**

Citrus juices make up the bulk of the retail fruit juice market, with citrus essential oil being one of the primary constituents contributing to the flavor of the juice. Both the Codex and USDA standards legislate an essential oil content, the official method for determination of the recoverable oil being the Scott Oil Method.

### **Ethanol**

Low levels of alcohol, which can result from the fermentation of juices, are permitted in the Codex standards; however, they must be monitored. The levels can be determined using distillation and

dichromate oxidation methodology, enzymatic methods, and GC, with the last now widely accepted.

### Mold Contamination

The Howard mold count is the traditional method for assessing mold content in tomato products, but has also been applied to other fruit products. The samples are studied microscopically using a Howard cell and the number of positive fields, those where the aggregate length of not more than three mold filaments exceeds one-sixth of the diameter of the field, is counted. There are regulations governing the percentages of positive fields that are acceptable in the various products. The *Geotrichum* mold count method suitable for citrus juices, canned fruits and juices, and pureed fruit products is also used.

### Preservatives

Many of the standards governing fruit and fruit products contain guidelines on acceptable levels of preservatives, e.g., sulfur dioxide, benzoic acid/benzoates, and sorbic acid.

Sulfur dioxide (SO<sub>2</sub>) is generally measured using indirect methods. The AOAC describes a modified Monier Williams method, while the Codex standards and EC directive recommend the Tanner modified method. Both are based on similar principles, involving the distillation of the SO<sub>2</sub> from an acidified aqueous suspension of the food. The evolved gas is subsequently trapped in hydrogen peroxide where it reacts to form sulfuric acid which is then titrated against dilute sodium hydroxide.

Benzoic acid and sorbic acid can both be extracted using modified Monier Williams distillation methods and quantified spectrophotometrically, or in the case of benzoic acid using thin-layer chromatography (TLC). LC and GC methodologies are now available and routinely used, having the advantage of simultaneous quantification of both benzoic acid/benzoates and sorbic acid.

### Trace Metals

Trace metals present in foods can have nutritional or toxicological significance and in the case of certain metals, e.g., copper and zinc, both of these. Fruit products are routinely tested for trace metal contents, e.g., canned fruits for tin, iron, copper, and zinc; fruit juices for copper, zinc, and lead; tomato puree for copper and also lead, tin, and arsenic. Statutory limits exist for many metals, e.g., lead, arsenic, zinc, and copper. Atomic emission spectrometry is now widely used for trace metal analysis as multielement determinations can be made on a single sample; however, other methodologies are

applicable. AOAC details colorimetric methods for determination of copper as diethyldithiocarbamate, with colorimetric methods for tin also available.

### Chemical Residues

Fresh fruit in particular may also be analyzed for the presence of chemical residues. These may be present due to specific horticultural and agricultural practices (pesticides, herbicides, etc.), environmental contamination, or through use of certain permitted additives. Detection and quantification of these residues is a complex analytical problem. Most methods used are based on solvent separation coupled with chromatographic determination using LC, GC, or GC-mass spectrometry.

### Added Colors

Fresh fruits must not contain any added color; however, certain colors are permitted in processed products. The range and use of colors are strictly governed, e.g., by UK and USDA regulations, and EC directives. Many methods exist for the extraction, separation, and identification of colors. A standard method suitable for fruit juices and other products involves the separation of colors from an acidified aqueous solution of the food using reversed-phase C<sub>18</sub> cartridge and identification using TLC or UV-visible spectrophotometry by comparison with standard color solutions.

### Adulteration

By far the largest area of adulteration occurring in fruit and fruit products is in relation to fruit juices. The size and value of the fruit juice market has meant that falsification of juice has become a lucrative business. The main areas of juice adulteration are the addition of water, sugar and water, or other fruit-derived products. Water addition is most effective when used in single-strength juices, as concentrates have to be diluted to a certain Brix value to meet specifications. The addition of sugar and water is difficult to detect, because of the possibility of adding appropriate sugar mixes to duplicate natural proportions. The addition of other constituents, e.g., organic acids, minerals, amino acids, carotenoids, to maintain the chemical balance of the fruit juice further complicates detection. The most complex form of adulteration is the use of materials derived from fruit but which cannot legally be used in the fruit juice. This can range from the addition of other fruits, to the use of peel extract, pulp wash, and the water extract of rag, pulp cell membranes, seeds, etc., in citrus juices. The wide range of fruit-derived

materials available with different properties can be combined in such a way as to make detection extremely difficult.

The methods available for detecting adulteration can be divided into two types: those which detect a specific indicator of the added material and those which rely on the disturbance of a natural analytical balance.

### **Stable Isotope Analysis**

The ratios of stable isotopes in biological materials provide a valuable tool for the analyst in detecting adulteration.

**Carbon isotope ratios** The ratio of the carbon isotopes ( $^{13}\text{C}/^{12}\text{C}$ ) in plant sugars varies according to whether the plant has a  $\text{C}_3$  or  $\text{C}_4$  photosynthetic carbon cycle. Most fruit juices, including citrus and apple, are derived from plants with  $\text{C}_3$  metabolism; however, cane sugar, corn invert sugar, and high fructose corn syrup all originate from the less common  $\text{C}_4$  plants.

**Oxygen isotope ratios** The use of oxygen isotope ratios is largely restricted to the measurement of water in fruit juices. Water in fruit is assimilated from groundwater and reflects the isotope ratios present in it; variations between juices are therefore naturally derived from variations in the groundwater. Analysis for carbon and oxygen isotopes is carried out using mass spectrometry; however, analysis is not simple.

**Hydrogen isotope ratios** The ratio of  $^2\text{H}/^1\text{H}$  can provide information on water addition but can also detect the addition of sugars from other plants. Measurement is usually carried out by the formation of sugar derivatives, which eliminates the oxygen-bound hydrogen from the molecule. This is most simply achieved by the fermentation of the sugars to ethanol. Fourier transform NMR spectroscopy has been used to examine the internal  $^2\text{H}/^1\text{H}$  ratios in the ethanol and has been successful in the detection of beet sugar in orange juice.

**Oligosaccharide analysis** The technique is based on the detection of specific oligosaccharides that are not normally present in the juice but are introduced by the addition of sugars that have undergone inversion. Analysis relies upon liquid chromatographic detection.

### **Analytical Pattern Methods**

There are several analytical systems that rely on authentic juice having a consistent pattern based on its composition.

**Matrix method** This method relies primarily on sugar profiles, UV-visible spectrofluorimetry, metals, calcium stable isotopes, and flavanoid glycosides to detect the addition of sugar, other fruits, or pulp wash.

**RSK system** This, the best known of the analytical systems, was developed in Germany in 1977 and is used to control commercial practice in that country. RSK stands for Richtwerte und Schwankungsbreiten Bestimmter Kennsahlem, or guide values and ranges of specific reference numbers. The juices are analyzed for a very broad range of analytes for which guide values, the ranges, and central value are given. Because of the natural variation in juice composition the ranges given are necessarily broad, and the detailed interpretation of the analytical data is an integral part of the system.

Other systems also exist including the French national system, the AFNOR standards, and on a European basis, the Association of the Industry of Juices and Nectars from Fruits and Vegetables of the European Economic Community has drawn up accepted analytical characteristics or reference standards.

**Visible and ultraviolet spectra** The use of visible and UV absorption and fluorescent spectral characteristics to detect juice adulteration has also been developed, with the presence or absence of specific absorption maxima at selected wavelengths indicating adulteration.

**Prospective methods** Several new techniques for the detection of fruit juice adulteration are still under development. These include pyrolysis mass spectrometry, size-exclusion chromatography, NIR spectrometry, and an expert system for data interpretation. With the methods of falsifying juices being so numerous and diverse, however, it is unlikely that any single method will ever be capable of detecting all forms of falsification.

*See also:* **Atomic Absorption Spectrometry:** Flame. **Atomic Emission Spectrometry:** Flame Photometry. **Carbohydrates:** Starch; Dietary Fiber Measured as Non-starch Polysaccharides in Plant Foods. **Chiroptical Analysis.** **Essential Oils.** **Ethanol.** **Food and Nutritional Analysis:** Antioxidants and Preservatives; Contaminants. **Isotope Ratio Measurements.** **Liquid Chromatography:** Food Applications. **Nuclear Magnetic Resonance Spectroscopy Applications:** Food. **Optical Spectroscopy:** Refractometry and Reflectometry. **Pesticides.** **Sampling:** Theory. **Vitamins:** Fat-Soluble; Water-Soluble. **Water Determination.**

## Further Reading

- AOAC (1990) *AOAC Official Methods of Analysis*, 15th edn. Arlington, VA: AOAC.
- Codex Alimentarius (1969) *Codex Alimentarius Sampling Plan for Prepackaged Foods*. Arlington, Virginia CAC/RM 42-1969.
- Hulme AC (1970) *The Biochemistry of Fruits and Their Products*. London: Academic Press.

- Kirk RS and Sawyer R (1991) *Pearson's Composition and Analysis of Foods*, 9th edn. London: Longman Scientific and Technical.
- Kratochvil B and Taylor JK (1981) Sampling for chemical analysis. *Analytical Chemistry* 53(8): 924A–938A.
- Nagy S and Attaway JA (1980) *Citrus Nutrition and Quality*. Washington DC: American Chemical Society.
- Pomeranz Y and Meloan CE (1987) *Food Analysis, Theory and Practice*, 2nd edn. New York: AVI.

## Packaging Materials

**A W Lord**, Pira International, Leatherhead, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Philip Tice, pp. 3698–3706, © 1995, Elsevier Ltd.

## Introduction

Analytical measurements on food packaging materials are generally carried out for three main purposes:

- To identify the components of the packaging.
- To identify and measure substances present that could migrate into the packaged foods and cause a health hazard to the consumer of the food. This work is often accompanied by measurements of the migration of particular substances into either the actual packaged foods, or into food simulants.
- To identify and measure substances present that could migrate into the packaged foods and result in adverse effects on the organoleptic properties, such as taste or odor.

## Food Packaging Materials

The main categories of basic materials used for food packaging are:

- plastics,
- regenerated cellulose films,
- paper and board,
- metal, and
- glass.

Of these, plastics are the most widely used and within this category there are also the largest numbers of variants. Many packaging materials are, however, multilayered with either different layers of plastics or combinations of plastics with paper/board, metal, or

glass. The individual properties of the different materials are used to produce food packaging with the required characteristics. For example, in a packaging material with two layers of different plastics, one layer might provide the basic strength whilst the other layer enables the packaging to be easily heat-sealed. Coatings are also often added to the basic plastic packaging material to provide additional barriers to the permeation of oxygen and water vapor. These coatings can be polymeric or vacuum deposited aluminum.

With some metal cans used for foods and beverages there is an inner lacquer (plastics) coating for the purpose of either preventing corrosion of the metal by the food/beverage, or preventing contamination of the food/beverage by the can metal.

A combination of a polymer layer with a board is used to package liquids such as milk, where the plastic layer provides the barrier to contain the milk within the package and the board the basic strength. Where it is necessary to store the beverage for long periods, such as fruit juices, additional barrier properties are required to prevent permeation of oxygen into the food product. To achieve this additional protection, an aluminum layer is incorporated within a plastic/board composite.

With many of the multilayer packaging materials adhesives are used to bind the layers together. The printing on the outside is a further important component of food packaging.

## Identifying the Components of Packaging Materials

It is often necessary to identify or confirm the basic composition of the packaging materials. This applies particularly with plastics due to the range of polymer types that are used. Six major polymer types are used for packaging and these are shown in Table 1, with



## Further Reading

- AOAC (1990) *AOAC Official Methods of Analysis*, 15th edn. Arlington, VA: AOAC.
- Codex Alimentarius (1969) *Codex Alimentarius Sampling Plan for Prepackaged Foods*. Arlington, Virginia CAC/RM 42-1969.
- Hulme AC (1970) *The Biochemistry of Fruits and Their Products*. London: Academic Press.

- Kirk RS and Sawyer R (1991) *Pearson's Composition and Analysis of Foods*, 9th edn. London: Longman Scientific and Technical.
- Kratochvil B and Taylor JK (1981) Sampling for chemical analysis. *Analytical Chemistry* 53(8): 924A–938A.
- Nagy S and Attaway JA (1980) *Citrus Nutrition and Quality*. Washington DC: American Chemical Society.
- Pomeranz Y and Meloan CE (1987) *Food Analysis, Theory and Practice*, 2nd edn. New York: AVI.

## Packaging Materials

**A W Lord**, Pira International, Leatherhead, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Philip Tice, pp. 3698–3706, © 1995, Elsevier Ltd.

## Introduction

Analytical measurements on food packaging materials are generally carried out for three main purposes:

- To identify the components of the packaging.
- To identify and measure substances present that could migrate into the packaged foods and cause a health hazard to the consumer of the food. This work is often accompanied by measurements of the migration of particular substances into either the actual packaged foods, or into food simulants.
- To identify and measure substances present that could migrate into the packaged foods and result in adverse effects on the organoleptic properties, such as taste or odor.

## Food Packaging Materials

The main categories of basic materials used for food packaging are:

- plastics,
- regenerated cellulose films,
- paper and board,
- metal, and
- glass.

Of these, plastics are the most widely used and within this category there are also the largest numbers of variants. Many packaging materials are, however, multilayered with either different layers of plastics or combinations of plastics with paper/board, metal, or

glass. The individual properties of the different materials are used to produce food packaging with the required characteristics. For example, in a packaging material with two layers of different plastics, one layer might provide the basic strength whilst the other layer enables the packaging to be easily heat-sealed. Coatings are also often added to the basic plastic packaging material to provide additional barriers to the permeation of oxygen and water vapor. These coatings can be polymeric or vacuum deposited aluminum.

With some metal cans used for foods and beverages there is an inner lacquer (plastics) coating for the purpose of either preventing corrosion of the metal by the food/beverage, or preventing contamination of the food/beverage by the can metal.

A combination of a polymer layer with a board is used to package liquids such as milk, where the plastic layer provides the barrier to contain the milk within the package and the board the basic strength. Where it is necessary to store the beverage for long periods, such as fruit juices, additional barrier properties are required to prevent permeation of oxygen into the food product. To achieve this additional protection, an aluminum layer is incorporated within a plastic/board composite.

With many of the multilayer packaging materials adhesives are used to bind the layers together. The printing on the outside is a further important component of food packaging.

## Identifying the Components of Packaging Materials

It is often necessary to identify or confirm the basic composition of the packaging materials. This applies particularly with plastics due to the range of polymer types that are used. Six major polymer types are used for packaging and these are shown in Table 1, with

**Table 1** Types of plastics used for food packaging and typical uses

<i>Plastic type</i>	<i>Typical uses</i>
Polyethylene	Bags and bottles
Polypropylene	Wrapping films and pots
Poly(vinyl chloride), unplasticized	Trays, bottles, and containers
Poly(vinyl chloride), plasticized	Wrapping film and cling film
Polystyrene	Trays, pots, and containers
Poly(ethylene terephthalate)	Lidding films and oven containers
Polyamide (nylon)	Laminated with polyethylene, 'boil-in-bag' pouches

typical uses. In addition to these basic polymer types, various copolymers are also used. For example, ethylene is copolymerized with vinyl acetate to produce ethylene vinyl acetate, styrene is copolymerized with both acrylonitrile and butadiene to produce the terpolymer ABS.

On simple, one- or two-layer structures, identification of the polymer type is conveniently achieved using Fourier transform infrared spectroscopy (FTIR). Each of the major polymer classes or copolymers has unique infrared spectra and are easily identified by comparison of the spectra to reference spectra.

The infrared spectrum of a plastic packaging material is most easily obtained when the sample is in the form of a film sufficiently thin to allow the infrared spectrum to be obtained through the film. However, with most plastic packaging, even those that are used in the form of films, the thickness is usually too great to obtain a good transmission spectrum. With some film materials stretching can adequately reduce the thickness. Alternatively, it may be possible to produce a solvent solution and cast a film of the required thickness. Thin films may also be pressed from the sample by careful melting on a hot-plate. Caution should be exercised with melt pressing, as apart from polymer degradation, there is the risk of altering the structure, for example, by skewing a polymer layer away from the region of analysis.

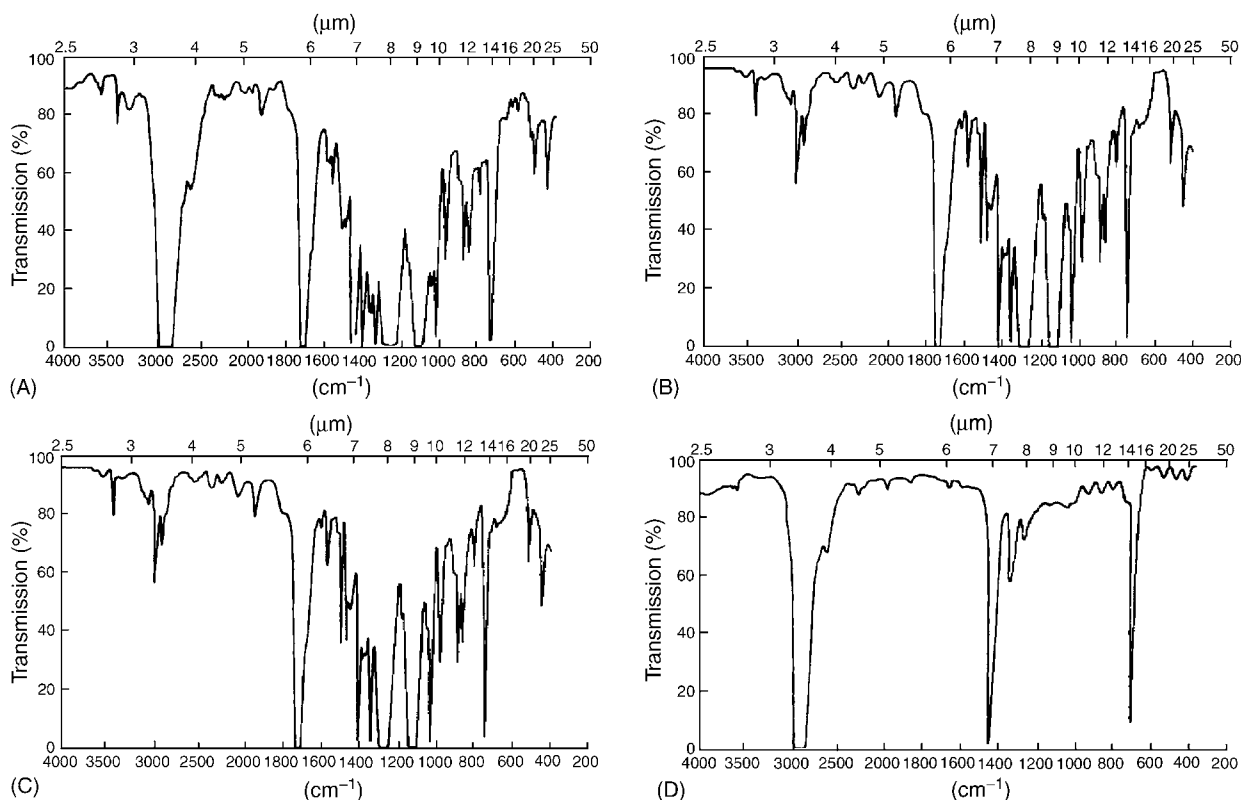
Infrared spectra may be obtained from surfaces using a variety of techniques. These included attenuated total internal reflectance (ATR) and specular and diffuse reflectance. These techniques involve the infrared beam passing through only the outer few micrometers of the sample. The most widely applicable is ATR. A typical two-layer plastic material used for lidding on plastic food trays consists of polyethylene and poly(ethylene terephthalate) bound together with an adhesive. ATR infrared spectra of the two surfaces will easily identify one surface as

polyethylene and the other surface as poly(ethylene terephthalate) by the very different spectra obtained. With the standard KRS5 (thallium bromide/iodide) ATR crystal the depth of penetration of the infrared radiation is a few micrometers. This is less than the individual polymer layers and consequently the adhesive is not 'shown' in either of the ATR spectra.

Where the polymer material is a copolymer it is often possible to obtain a measurement of the relative amounts of the various monomer components from an infrared spectrum. For example, with an ethylene-vinyl acetate copolymer the relative heights of absorption bands from both the ethylene and vinyl acetate are measured and ratioed with the spectrum recorded in the absorbance mode. The most convenient absorbance bands are:  $720\text{ cm}^{-1}$  for polyethylene and  $1235$  or  $1740\text{ cm}^{-1}$  for vinyl acetate. Copolymers of known composition are required for calibration. It is possible to obtain an assessment on the butadiene and acrylonitrile contents in styrene/butadiene/acrylonitrile copolymers. The bands usually used are: for styrene  $1600\text{ cm}^{-1}$ , for acrylonitrile  $2240\text{ cm}^{-1}$ , and for butadiene  $996\text{ cm}^{-1}$ .

Modern packaging materials are very often multiple-layered structures. If the packaging material is a laminate or coextrusion each layer will produce an infrared spectrum. The resulting composite spectrum becomes difficult to interpret. In most cases laminates are manufactured using adhesive to bond the layers together. It is sometimes possible to select a solvent to dissolve the adhesive thereby enabling the individual polymer layers to be separated. The separated polymers can then be identified by their infrared absorption spectra. Spectra from a polyethylene/poly(ethylene terephthalate) laminate and the separated layers are shown in **Figures 1A–1D**. Polyurethane-based adhesives are widely used to bond poly(ethylene terephthalate) to polyolefins. Hot benzylalcohol is a good solvent for a range of polyurethanes. Other solvents include tetrahydrofuran and chloroform for acrylate-based adhesives. This approach also enables the adhesive to be identified.

An approach that can be applied to laminates and coextrusions is to selectively dissolve and remove polymer layers by careful selection of solvents. Thus, the nylon layer in a polyethylene/nylon/polyethylene coextrusion can be isolated by boiling in xylene. Alternatively, the nylon could be removed by boiling in formic acid. Solvents for the selective removal of polymers are listed in **Table 2**. Acids or alkalis should be avoided on some polymers where there is a risk of reaction with the polymer. An example would be the use of concentrated sodium hydroxide solution on a metallized film comprising certain acrylic/ethylene



**Figure 1** Infrared transmission spectra of: (A) polyethylene/poly(ethylene terephthalate) laminate; (B) separated poly(ethylene terephthalate) layer; (C) separated polyethylene layer plus adhesive; and (D) separated polyethylene layer with adhesive removed.

**Table 2** Solvents for plastics

Plastic type	Solvents
Polystyrene and copolymers	Chloroform, ethyl acetate, ketones
Polyethylene	Decalin, hot toluene
Polypropylene	Decalin, hot xylene
Poly(vinyl chloride)	Tetrahydrofuran, cyclohexanone
Poly(ethylene terephthalate)	<i>o</i> -Chlorophenol, trichloroacetic acid
Polyamide	Formic acid, phenols

ionomer types, where marked alterations to the infrared spectra can result from such treatment.

It is sometimes difficult to obtain thickness measurements on layers due to swelling of layers with the solvent or partial break up of thin layers. If the density of the polymer is known or measured the thickness of a layer may be calculated from the weight of the polymer.

Although infrared spectroscopy is a useful technique for identifying polymers in packaging materials it is important to emphasize that as a result of the development of packaging technology, it is now

rarely possible or wise to use the technique on its own without recourse to other analytical techniques. There has in recent years been a trend toward the use of coextrusions, and away from adhesive bonded laminates. Coextrusions of a wide range of polymers can be produced using thin tie layers (a few micrometers) of polymer with compatibility for the different polymer types. For example, polypropylene and ethylene vinyl alcohol can be extruded into films and bottles. Layers within coextrusions cannot be readily separated using solvents. In addition, the time and cost constraints upon the analysis of coextrusions mean that physical isolation of layers is often impractical.

The most efficient approach to establishing the structure of packaging is a combination of optical microscopy, differential scanning calorimetry (DSC), and infrared spectroscopy. The first stage in establishing the construction of an unknown packaging material is to subject it to the following optical microscopy techniques. A section (typically 5–10  $\mu\text{m}$ ) is cut from a 10  $\times$  10 mm area using a microtome. It is important that the sample is held rigid but strain free and cut with a very sharp knife. The best knife for packaging material is usually a freshly made glass

knife. Thicker sections ( $>2\text{ mm}$ ) require use of a steel knife. It is sometimes beneficial to cool samples below their glass transition temperature ( $T_g$ ) in order to provide a more rigid structure to section. It may also be necessary to mount samples in potting resin for support prior to sectioning through the sample and resin.

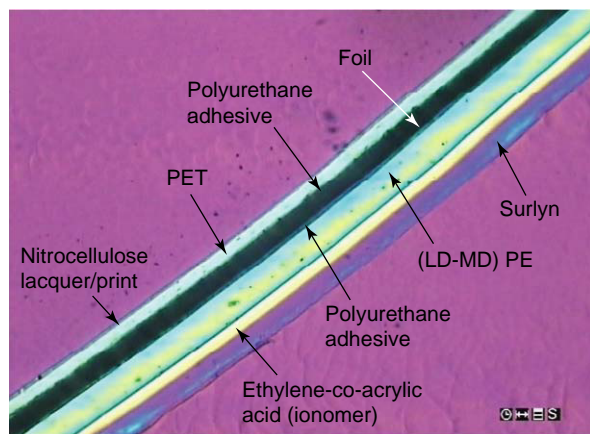
The section is then examined under polarized light (using cross-polars). A tint plate is useful to provide color differentiation of layers. This enables the number and thickness of each layer to be established. Thickness measurements are made by calculation after measuring the total thickness of the sample using a micrometer. The ratio of total thickness to layer thickness is calculated in arbitrary units using the scale graduations on the microscope. The thickness of each layer is then calculated from the ratios knowing the true total thickness. An example of a typical cross-section of a coextrusion is shown in Figure 2.

Information on the composition of individual layers in the structure can then be obtained by observations of the layer under the microscope. Different types of polymers have a recognizable morphology. Polypropylene has very large spherulites distinct from many other polymers. Figure 3 shows a photograph obtained from polypropylene cooled slowly from a melt. Compounded poly(vinyl chloride) (PVC) can also be quite characteristic due to the different phase regions arising from the presence of impact modifiers as well as pigment specks from color adjusters. Calcium carbonate filler and talc anti-blocking agent have recognizable morphologies. It is also possible to determine whether a layer contains

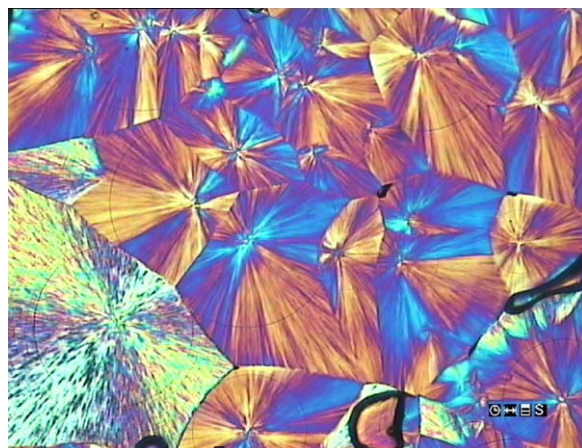
recycled process scrap. It is quite common for off-cuts to be recycled in an inner layer of a coextrusion.

Infrared spectra are then obtained from the surfaces of the packaging material after solvent removal of any print or lacquer. Spectra are best obtained by ATR. A further portion of the packaging material is then subjected to DSC. This is a technique where a few milligrams of the sample is subjected to a programmed temperature ramp in a specified atmosphere inside a sample chamber. The heat flow (power) to the sample is monitored against temperature as the sample is subjected to the heating ramp. For the purposes discussed here this provides a trace showing the melting points of the polymers present. Typical melting ranges for common packaging polymers are tabulated in Table 3.

The technique cannot be used to obtain melting points for amorphous polymers. The sample polymer is heated and cooled and then reheated at a controlled rate to record the melting points. This procedure removes hysteresis effects that may be present in the polymer as a result of the manufacturing process and which may alter the perceived melting point. DSC is capable of identifying polymers and polymer blends



**Figure 2** Cross-section of a coextrusion viewed through a tint plate on an optical microscope. PET = Poly(ethylene terephthalate), (LD-MD) PE = ? (Reprinted with permission from Mr R Musgrove, Pira International.)



**Figure 3** Polypropylene spherulites viewed through a tint plate. (Reprinted with permission from Mr R Musgrove, Pira International.)

**Table 3** Typical melting ranges for common polymers

Polymer	Melting range ( $^{\circ}\text{C}$ )
Linear low-density polyethylene	115–130
Low-density polyethylene	100–115
Ethylene vinyl acetate	100–110
Polyamide	210–260
Poly(ethylene terephthalate)	240–260
Poly(vinylidene chloride)	220
Polypropylene	160–170
Ethylene propylene random co polymer	149

not readily identifiable in packaging materials using infrared spectroscopy. Examples of these include, low-density polyethylene, linear low-density polyethylene, high-density polyethylene, and blends of these polymers. The blend ratios of these polyolefins can be estimated after calibration using the pure polymers.

The construction of the packaging material is then determined by comparing the data obtained from all the analytical techniques. Any layers that are difficult to identify are then identified by either detailed analysis on the isolated layer using FTIR and other analytical techniques, or by applying additional analytical techniques to the whole structure. For example, pyrolysis gas chromatography–mass spectrometry (GC–MS) can confirm the presence of a styrene/acrylate copolymer adhesive or vinylidene chloride/acrylate copolymer coating. The pyrolysis causes depolymerization, often to the starting monomers, which are then identified from their mass spectra. The technique can be applied to the whole construction. This is useful when FTIR analysis is not conclusive, or where the layer cannot be isolated for FTIR analysis. In the absence of a pyrolyser instrument, it is possible to perform the technique by briefly heating the sample in an inert atmosphere in a sealed headspace vial over a gentle Bunsen burner flame. Static headspace GC–MS analysis of the pyrolysate is then carried out. Hot stage microscopy is a particular useful technique for identifying layers. A key advantage of the technique is the ability to identify layers that cannot be isolated for analysis. In this technique, the packaging material cross-section is heated at a controlled rate under the microscope. The melting ranges of the individual layers can be observed and compared with the melting ranges observed by DSC. It should be noted that there is a difference of a few degrees centigrade in the melting ranges observed by DSC and hot stage microscopy:

The strategy for identification of a packaging material construction is summarize:

1. Examine a cross-section by optical microscopy; determine the number of layers and their thickness, tentatively identify the polymers in the layers.
2. Obtain a DSC of the whole material. This will identify all the crystalline polymers present in the structure.
3. Obtain infrared spectra from the surfaces. This will confirm the composition of the outer layers. Isolate fractions of the construction and obtain their infrared spectra to confirm the identifications made by DSC and optical microscopy.

Optical microscopy techniques can also be applied to packaging failure problems. Sections can be taken through a heat seal region to establish the integrity of the seal. Molecular orientation, melt flow, blend homogeneity, and crystallinity can be observed that can reveal the cause of stress cracking and other types of packaging failures.

## Analysis of Substances Related to Food Safety

It is sometimes necessary for technological reasons to use chemicals that have toxic properties in the manufacture of a food packaging materials. Also, there is the possibility that some of the chemicals and components used for food packaging materials can contain trace levels of toxic contaminants. Where toxic substances are unavoidably present in a food packaging material for any of these reasons, it is necessary to ensure that levels of these substances are restricted so that any transfer to packaged food does not exceed safety limits. The national regulations of individual countries control the safety of food packaging with respect to the substances with known toxic properties. In some countries the specific restrictions are contained in official recommendations or codes of practice. The primary restrictions are on the levels that migrate or transfer to the packaged food and are designated ‘specific migration limits’. However, in some cases the restriction is a permitted level in the packaging material, while for others the restriction is a limit on the quantity, which can be extracted. As might be expected where the safety of food is concerned, the set limits are often low requiring sensitive analytical methods.

### Food Contact Plastics

Vinyl chloride monomer used for the manufacture of PVC plastics intended for contact with foods provides an example where there is a low ‘specific migration limit’, plus a low limit on the level allowed to be present in the packaging material. These limits are contained in an EC Directive 78/142/EEC on PVC plastics and are: 0.01 milligrams per kilogram of food (10 ppb), and 1 milligram per kilogram of polyvinyl chloride.

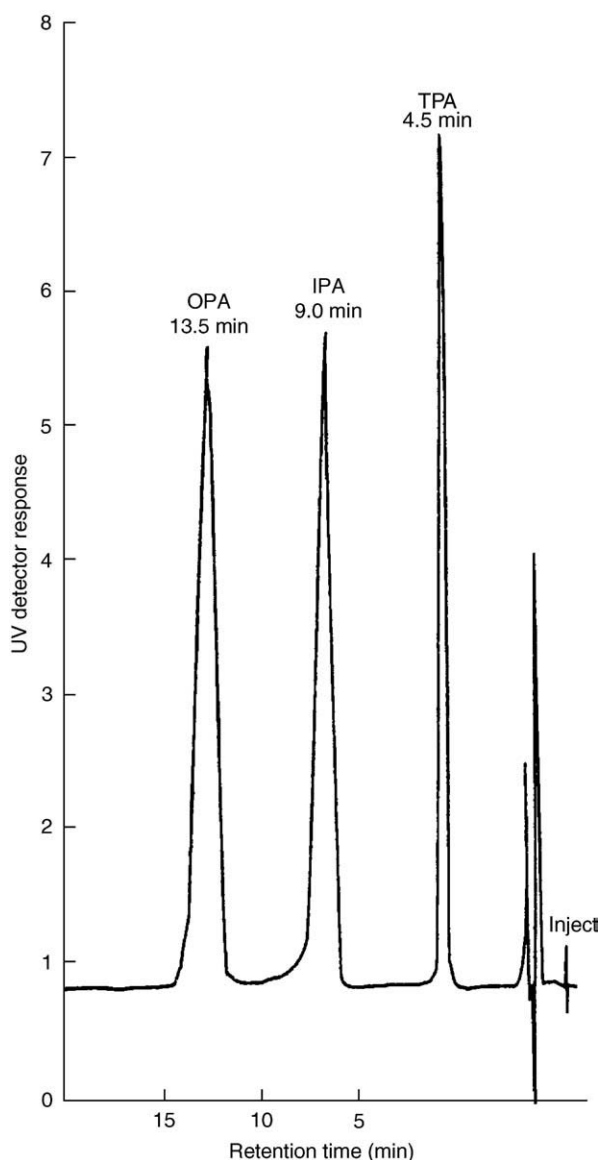
Vinyl chloride is a gas at ambient temperature and the official EC analytical methods for both determinations use headspace GC with a flame ionization detector (FID). Where a determination exceeds the legislation limit, confirmation is required with headspace GC using either a different chromatography column, or a different detector, or with the gas chromatograph coupled to a mass spectrometer.

For the determination of free vinyl chloride monomer in plastics, the test sample is dissolved or dispersed in *N,N*-dimethylacetamide in a sealed vial and then equilibrated at 60°C before sampling the headspace. When determining vinyl chloride monomer in foods or food simulants, *N,N*-dimethylacetamide is again used with the sample in a sealed vial, with equilibration at 60°C.

Other volatile plastics monomers with similar migration limits, such as acrylonitrile and butadiene, are also determined by the headspace GC technique. For the measurement of 'nonvolatile' monomers in plastic food packaging and in foods or food simulants due to migration, high-performance liquid chromatography (HPLC) and ion-exchange chromatography techniques can be employed. Food and food simulants may give rise to interference problems with the analysis. Sample clean-up procedures are widely used such as solid-phase extraction of interference from extracts or size exclusion chromatography to remove fats and oils. Selective detectors such as mass spectrometers are now widely used for both liquid and gas chromatography.

Figure 4 is an ion-exchange chromatogram of phthalic acids. Terephthalic acid and isophthalic acid are monomers of polyester plastics. Orthophthalic acid is the internal standard. There are migration limits for both terephthalic acid and isophthalic acid in EC regulations of 7.5 and 5.0 mg per kg of food. The other main classes of substances in the 'safety' category, which can be present in food packaging plastics and for which analytical measurements are required, are the plastics additives. These substances perform the functions of plasticizers, antioxidants, antistatics, slip agents, and stabilizers. Of these additives, the plasticizers have been most extensively studied and analytical methods developed for their determination in the plastics and in packaged foods. The techniques are usually based on GC either with an FID or with a mass spectrometer as the detector. Key advantages offered by a mass spectrometer are selectivity of response and the ability to add a deuterated internal standard to the sample to compensate for the incomplete and variable recovery of the analyte in the analysis.

Migration testing of plastics packaging prior to use for compliance with any legislation limits is usually carried out with food simulants rather than actual foods. First, the analytical task is more often simple and, second, testing with a food simulant or simulants for a class of foods covers use with all foods in that class. The food simulants are simple liquids that represent different classes of foods. For foods where the liquid phase is largely water, distilled water is used as the simulant. For acidic foods (typically pH



**Figure 4** Separation of terephthalic acid (TPA), isophthalic acid (IPA), and orthophthalic acid (OPA) by ion-exchange chromatography. (From Ashby *et al.*, 1992.)

4.5 or less) the simulant is an acetic acid aqueous solution, and for alcoholic beverages and other foods containing alcohol, the simulant is an ethanol aqueous solution with strengths more or less equal to the alcoholic concentration in the beverage or food.

Selecting a simulant for foods containing fats and oils has not been easy. In the USA, *n*-Heptane is specified as the fatty food simulant in the Food and Drug Administration (FDA) regulations, although it is now recognized to give migration levels of specific substances well in excess of those that occur with the foods even after applying a suitable reduction factor. In Europe, olive oil has been selected as the fatty



food simulant with alternatives of sunflower oil or a synthetic triglyceride, known as HB307, developed specifically for migration testing. Alternatives to olive oil were considered necessary as it was known that various analytical problems arise with olive oil. The food simulants required for regulatory migration testing in the EC Member States on plastics packaging intended for use with foods, are listed in EC Directives 97/48/EEC and 85/572/EEC. These are shown in **Table 4** together with the classes of foods and beverages that they represent. The 85/572/EEC Directive also contains a table that specifies the simulant or simulants to be used for individual categories foods and beverages, and the 97/48/EEC Directive gives the test conditions of time and temperature corresponding to the intended conditions of use. Fatty foods contain various amounts of oils and fats. For those that have high oil or fat content the extent of migration of substances from the plastics is often higher than for those with low oil or fat contents. Consequently, reduction factors are applied to migration values obtained with olive oil to allow for the various levels of oil and fat in particular foods.

It has been generally accepted that migration from packaging materials into dry foods will be low compared to moist, liquid, or fatty foods. Accordingly, more work has been completed in developing simulants for these foods than dry foods. At the present time the only regulatory or generally recognized simulant for dry foods is Tenax (poly(2,6-diphenyl-*p*-phenylene oxide)). Tenax has been adopted as a dry food simulant because it is a dry porous polymer with a large surface area that exhibits high-adsorption characteristics for a wide range of volatile organic compounds. Tenax is therefore considered to be a worst-case dry food simulant for a wide range of dry foods. Other food simulants have been investigated for use with paperboard. A semisolid food simulant consisting of a mixture of diatomaceous earth, water, and olive oil has been used, as well as filter paper impregnated with olive oil.

**Table 4** EC food simulants for migration tests and the corresponding classes of foods

<i>Food simulant</i>	<i>Class of food or beverage</i>
Distilled water	Aqueous foods and beverages
3% (w/v) aqueous solution of acetic acid	Foods and beverages with pH 4.5 or less
10% (v/v) aqueous solution of ethanol <sup>a</sup>	Foods and beverages with 15% or less alcohol
Olive oil or sunflower oil or HB307	Foods containing fats and oils

<sup>a</sup> For a food or beverage with an alcohol content greater than 10% (v/v), a simulant with a similar ethanol concentration is used.

In addition to the requirement to measure the levels of individual specific substances that have migrated from plastic food packaging into foods or food simulants and have toxic properties, there is sometimes also the necessity to measure the total quantity of substances that are likely to transfer from the plastics to the food. This is called overall migration or global migration. The tests are carried out with food simulants as they are impracticable with foods. No attempt is made to identify the nature of the substances that have migrated from the plastic material. In the current EC 'food contact' plastics regulations there is an overall migration limit of 60 mg of substances from the plastics material per kilogram of food simulant ( $60 \text{ mg kg}^{-1}$ ), or expressed alternatively as 10 mg of plastic substances from  $1 \text{ dm}^2$  surface area of plastics test specimen ( $10 \text{ mg dm}^{-2}$ ). With the three aqueous based food simulants – distilled water, acetic acid, and ethanol solutions – the overall migration is measured by determining gravimetrically the nonvolatile residue in the simulant following exposure to the plastic test specimen. The values normally obtained with these aqueous-based food simulants are usually well below the regulation limit and in the region of 6–18  $\text{mg kg}^{-1}$  ( $1\text{--}3 \text{ mg dm}^{-2}$ ). As the tests are most often carried out on test specimens with a surface area of  $1\text{--}3 \text{ mg dm}^{-2}$  the total quantities of migrating substances are typically a few milligrams. Consequently, care has to be taken with the gravimetric measurement to ensure reliable results are obtained. After evaporating the simulant to dryness the nonvolatile residue is dried in an oven at  $110^\circ\text{C}$  until constant weight is obtained. It has been found, however, that particular care has to be taken when using glass evaporating dishes to ensure that there is adequate time allowed for both the heating period in the oven and the cooling period in the desiccator for the mass to stabilize before each weighing. With metal evaporating dishes the mass of the dish and residue stabilize more quickly, but with the acetic acid simulant it is necessary to use dishes made of platinum, or a metal with similar chemical resistance, to prevent additional errors from corrosion products.

To measure overall migration with olive oil or alternative simulants, the method used with the aqueous-based food simulants is obviously not applicable. With oil-type simulants the test is carried out by measuring the loss in mass of the test specimen after exposure to the food simulant. However, most plastics absorb some of the oil that then has to be extracted and quantitatively measured before the true loss in mass can be calculated. The extraction solvent that has been most commonly used in the past is

1,1,2-trichlorotrifluoroethane, as it is a good solvent for olive oil but does not dissolve most plastics. As this solvent is a chlorofluorocarbon its main use and supply is being phased out and pentane and diethyl ether are now used instead for the extractions.

Once the extraction has been completed the olive oil is quantitatively analyzed by hydrolyzing to the fatty acids, methylating to form the methyl esters, and measurement by GC. The value obtained is subtracted from the mass of the test specimen after exposure to the olive oil, to allow the overall migration value to be calculated. The test method has a reputation of poor precision and reliability not only due to the complex procedure, described above, but also due various other factors that are known to influence results, such as moisture absorption by the plastic test specimens, incomplete extraction of the olive oil, and change in composition of the extracted olive oil. The overall and specific migration analytical test methods have been established as Standards by the European Committee for Standardisation (CEN). Reference plastics are also available with certified migration values in olive oil (Institute for Reference Materials and Measurements BCR, Geel, Belgium, sales@irmm.jrc.be).

### **Paper and Board**

Most paper and board food packaging materials are not used in direct contact with liquid foods and consequently migration tests with liquid food simulants are not considered appropriate. Paper and board packaging does, however, come into direct contact with various moist and fatty foods where migration of substances into the food can sometimes occur. No simulants have yet been selected to specifically represent these classes of foods in migration tests on paper and board. To ensure that the paper and board material is suitable and safe to package foods, extraction tests are often carried out. The extraction test is performed with cold or hot water, or sometimes with dilute acids and solvents and is considered to be a more severe test than a migration test. Analytical measurements are then carried out for specific substances on the extraction liquid. For example, tests are carried out for free formaldehyde, which can arise from wet-strength additives of the melamine formaldehyde or urea formaldehyde types. A typical analytical method for the determination of free formaldehyde in the extracts is based on colorimetric procedures using chromotropic acid or acetylacetone (pentane-2,4-dione). With paper and board products there is concern that there could be toxic contaminants present that in turn could transfer to the food when used as packaging. These

possible contaminants include: the toxic heavy elements arsenic, mercury, lead, cadmium, and chromium, plus chlorophenols and polychlorinated biphenyls (PCBs). With extraction liquids such as water or dilute aqueous acid solutions, the toxic heavy elements can be analytically determined using atomic absorption spectroscopy. Arsenic can be measured with the hydride generation technique, mercury with the cold vapor technique, and the other metals by the standard flame technique. Inductively coupled plasma atomic emission spectroscopy is now also widely used. Pentachlorophenol and other chlorophenols can be determined by either GC or HPLC. When using GC the chlorophenols are best derivatized to form the methyl or acetyl derivatives in order to improve the chromatographic performance and the analytical precision. These analytical techniques have also been used in the detection and analysis of chlorophenols suspected of being responsible for odors and food tainting, as described later. The PCBs are determined in extracts from paper and board materials by GC with an electron capture detector or mass spectrometer.

Two compounds are currently of particular interest in paper and board. Diisopropyl naphthalene (DIPN) is a mixture of isomers that until recently were widely used in carbonless copy papers as ink solvents. Although it is currently being replaced it occurs as a persistent contaminant in recycled paper and board. Various studies have shown that it is able to migrate from paperboard into food. There is a draft CEN analytical method available. This method involves acetone extraction and quantification by GC-MS using diethyl naphthalene as an internal standard. There is currently no limit for DIPN but levels are being monitored to reduce concentrations in recycled paperboard.

Two related compounds are 3-monochloropropane-1,2-diol (3-MCPD) and dichloropropanol. These arise in paper board due to the hydrolysis of epichlorohydrin-based wet strength agents. 3-MCPD can occur in food from hydrolyzed vegetable protein. The limit in food is 120 ppb. In packaging the specific migration limit is 12 ppb in the food. The dichloropropanol does not at present have a limit. However, the German BGVV recommendations (widely accepted as useful guidelines) list a limit of 2 ppb in a hot water extract. A convenient method of analysis is to extract the two compounds with water. The water extract is then totally absorbed onto a diatomaceous earth cartridge. The cartridge is then washed with a large volume of diethyl ether. The water is retained on the cartridge and the 3-MCPD and dichloropropanol extracted and eluted by the diethyl ether. The ether is then concentrated by

evaporation and the two compounds derivatized and injected for analysis by GC-MS.

### Metal Packaging

Cans are widely used to pack food. In some cases tinplated steel cans are used, for example, for packing fruit. Predominantly, the cans are internally coated with a polymeric coating to prevent corrosion or food spoilage. A considerable amount of work has been done in recent years investigating the extent to which compounds present in the lacquers migrate into food.

Attention has focused on the chemical compounds bisphenol A, BADGE, and BFDGE. Bisphenol A is manufactured from the reaction of phenol with acetone. The bisphenol A is further reacted with epichlorohydrin to produce bisphenol A diglycidyl ether (BADGE). BADGE is then polymerized and cross-linked in a stoving process to produce an epoxy phenolic coating that has high chemical and mechanical resistance. These coating are called bisphenol A-epoxies.

Alternatively, phenol may be reacted with formaldehyde to generate bisphenol F. Unlike bisphenol A, bisphenol F is a mixture of isomers rather than a discrete compound. Bisphenol F can be subjected to a condensation process in which a polymeric resin called Novolac is produced. The Novolac can be reacted with epichlorohydrin to produce a polyglycidyl ether and these are called novolac glycidyl ethers (NOGE). NOGE is not used to produce epoxyphenolic coatings, as is the case with BADGE.

Organosol coatings are dispersions of PVC in softener, solvent, and other resins. Solid contents are typically 40–80%. The coating is stoved to evaporate off the solvents and cure the resin. BADGE is often added as an additive to scavenge for hydrochloric acid generated from the PVC during curing. Alternatively, NOGE is used as an additive instead of BADGE.

The most widely used lacquer types for food cans, where the food is retorted in the can to ensure preservation are:

- epoxyphenolic and
- organosol

Epoxyphenolic lacquers are universally used for both can bodies and ends for two- and three-piece constructions, although more usually for shallow draw cans. Beverage can bodies are commonly epoxyamino coated, and the 'easy-open' end and deeper draw two-piece cans are organosol coated. Coatings may contain residual BFDGE and bisphenol F arising from the NOGE in organosol coatings and bisphenol

A as well as BADGE arising from the use of BADGE in organosol and epoxy phenolic coatings. Possible residues remaining in the coatings are listed below, all of which have been found to contaminate the food:

- BFDGE
- BADGE
- Bisphenol A
- Bisphenol F

BADGE and BFDGE undergo hydrolysis and addition of hydrogen chloride released from the PVC organosol in aqueous foods and a series of reaction products result. Concern has been raised over these reaction products. These are listed below:

- BADGE · HCl
- BADGE · 2HCl
- BADGE · H<sub>2</sub>O · HCl
- BADGE · H<sub>2</sub>O

These decomposition products result from the ring opening on the epoxy group of which there are two. The legislation (Directive 2002/16/EC, February 20, 2002, on epoxy food contact materials) specifies a migration limit of 1 ppm in the food. This limit is the total of all the reaction products and BADGE added to the BFDGE and its reaction products. In addition, there is a requirement of no detectable migration of NOGE at a detection limit of 0.2 mg kg<sup>-1</sup> in the food or 0.2 mg/6 dm<sup>2</sup> in the can. The decomposition product BADGE · 2H<sub>2</sub>O in food is ignored as this is not of toxicological significance. However, it must be included if the migration test is done on food simulants as there is the risk of forcing decomposition through to the BADGE · 2H<sub>2</sub>O and underestimating the other compounds. The legislation is due for review in 2004 as the toxicity of the chlorohydrins is not at present established.

### Analysis of Substances Causing Taint

Taint from food packaging is very rare when one considers the tonnage of packaged food consumed each year. Tainting chemical compounds present or derived from the food packaging are often volatile compounds. With plastics these odorous volatiles can be: monomer residues, reaction by-products from the polymerization process, breakdown products of certain additives and contaminants. For example, with polystyrene plastics high levels of styrene monomer produce a very characteristic odor and a number of incidents of tainting from styrene monomer have been reported. With polyethylene terephthalate (PET) plastics, acetaldehyde can be formed during

the polymerization process and when PET is used for beverage bottles the acetaldehyde can cause tainting of the beverage. With paper and board materials the volatiles arise mainly from natural lipids and resins originating from the wood raw material, but some can come from synthetic resins used in surface coatings that are applied for improved printability and appearance. The predominant volatiles originating from the wood lipids and resins are usually aldehydes, carboxylic acids, and alcohols. The odors of some of these aldehydes are not unpleasant being described as 'grassy', but others have 'rancid' odors. Many of the carboxylic acids and alcohols have strong sharp odors. The synthetic resin binder used in the surface coatings is typically a styrene/butadiene copolymer that can contain odorous reaction by-products such as 4-phenyl cyclohexene. If solvent-based adhesives are used in sealing the packaging or to bind layers together, and if the finished packaging is printed with solvent-based inks, solvent residues can add to the list of volatile substances. If the more odorous of these print solvents and volatile substances are present in sufficient quantities they can cause the packaging to be odorous and in turn result in tainted foods.

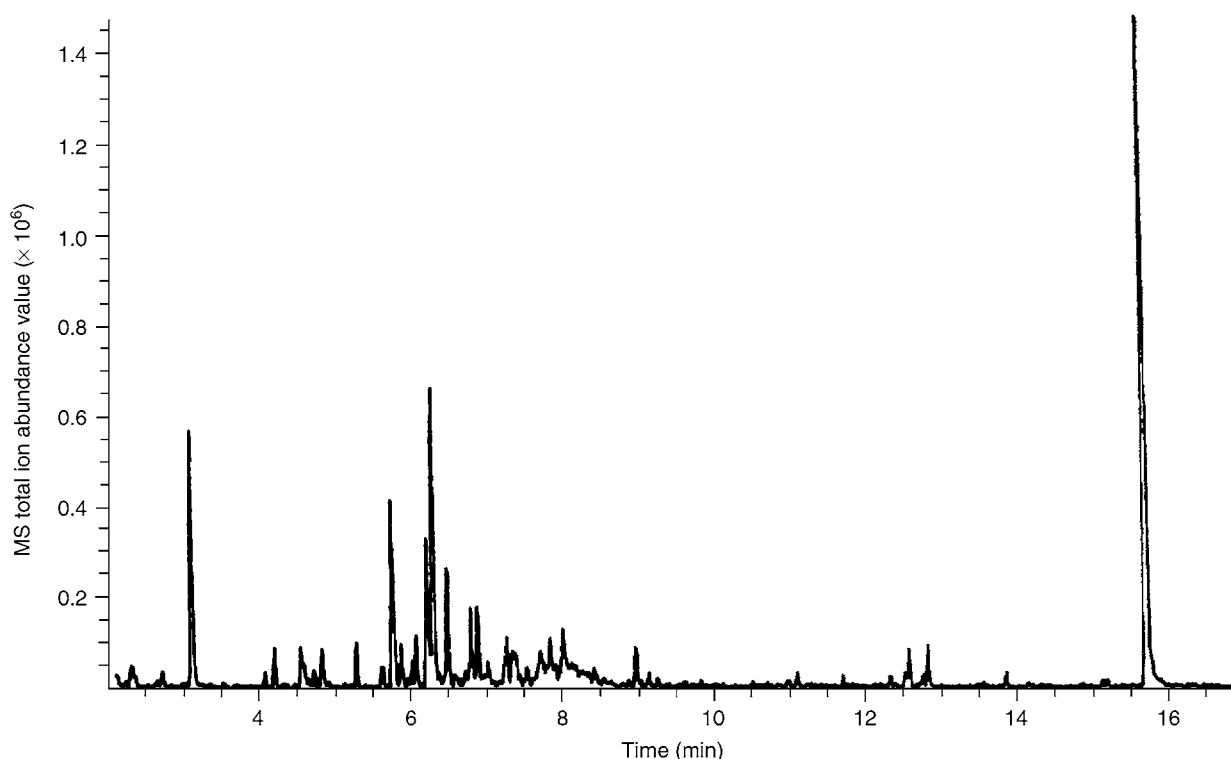
Two classes of highly odorous substances that have been known to contaminate food packaging and result in food tainting are the halophenols and the related haloanisoles (methylated chlorophenols and bromophenols). In the past contamination was invariably with the chlorophenols and corresponding anisoles from wooden pallets and surfaces treated with wood preservers or phenol-based disinfectants. The anisoles are generated by microorganisms such as molds from the phenols. Recently, there has been a noticeable trend toward increased contamination with the bromophenol and corresponding anisoles. This reflects the substitution of bromophenols for chlorophenols in wood treatments. The odor threshold for tribromoanisole in water is  $8 \text{ pg l}^{-1}$  ( $8 \times 10^{-12} \text{ g l}^{-1}$  or 8 parts per trillion, ppt). Low-detection limits are therefore required for such taint investigations. Concentrations above 1 ppb in the packaging are often sufficient for tainting to occur. Polyethylene is the most widely used polymer in contact with food, usually in the form of a thin inner layer of a food pack. It therefore only requires contamination of a few sacks of polyethylene granules with the anisole to result in tainting of a large amount of food.

Analytical measurements and investigations are therefore carried out to detect and measure volatile odorous substances in food packaging either for quality control purposes or when odor and taint problems arise. The technique of choice is GC-MS.

For odor investigation the chromatograph is fitted with an odor port so that the flow from the analytical capillary column is split via a T piece to an odor port. By this means it is possible to smell compounds eluting from the capillary column simultaneously with their detection by the mass spectrometer. This enables an odorous compound to be identified and quantified in food and packaging.

Isolation and concentration of tainting compounds from the rejected food and packaging prior to analysis is usually the most challenging step in the investigation. Dynamic headspace sampling is widely used. In this technique, a sample of the packaging material is placed in a vessel that is closed, heated to a temperature of  $\sim 70^\circ\text{C}$ , and then purged with an inert gas such as nitrogen or helium. Volatiles released from the packaging are removed by the purge gas, trapped, and concentrated on a porous polymer such as Tenax. Transfer of the volatiles from the porous polymer to the gas chromatograph is performed by thermal desorption or by solvent elution and injection as a solvent solution. The chromatogram in Figure 5 shows volatile substances that have been collected by the dynamic headspace technique from a printed carton-board that had caused tainting in a packaged cake. The tainting was attributed to the benzophenone that appears as the large peak at just below 16 min. Benzophenone is used as an initiator in ultraviolet radiation cured printing inks. The peak at 3.1 min is the aldehyde, hexanal, which originated from the pulp used to make the board. The cluster of peaks from  $\sim 5.5$  to  $\sim 7$  min is volatiles from the synthetic resin binder in the board coating. None of these substances produced detectable odors.

The Likens-Nickerson extraction technique can also be used as a concentration technique, particularly for those volatile substances that are steam volatile such as the chlorophenols and chloroanisoles, and also when carrying out an analysis for the packaging volatiles in foods. The sample is boiled in a flask with water. Consideration must be given to the pH of the sample in the water. Basic compounds will be present as water-soluble involatile salts in boiling water at low pH, and acids as the corresponding salts in boiling water at high pH. The procedure is therefore best carried out under basic conditions and then repeated after acidification with a few drops of nitric acid. The steam is condensed and continuously extracted with a suitable nonwater-miscible solvent, any solvent-soluble volatile substances being transferred to the solvent. After concentration of the solvent solution by evaporation of the solvent with a Kuderna-Danish apparatus, the analysis is again performed using a gas chromatograph coupled to a mass spectrometer.



**Figure 5** Chromatogram of volatile substances from a carton-board food packaging printed with a UV-cured ink. Benzophenone, the printing ink component responsible for tainting of packaged food, is represented by the peak at 15.7 min.

Solid-phase micro extraction is a useful technique in which volatiles are partitioned from the sample onto fibers coated with polar or nonpolar bonded phases. The fiber is then placed directly into the heated injection port of the GC where compounds are volatilized and carried onto the capillary column. The technique is less sensitive than the techniques described above as it is an equilibrium process. However, modern ion trap mass spectrometers have increased in sensitivity and the technique is becoming widely used.

### Residual Solvents

Solvent residues from printing and adhesives have the potential to cause food tainting. Typical solvents used in printing with characteristic, easily detectable odors are aliphatic esters, such as ethyl acetate, isopropyl acetate, and *n*-propyl acetate, the alcohols isopropyl alcohol and *n*-propyl alcohol and hydrocarbon mixtures, particularly aromatics. Regular tests are carried out by printers of food packaging to ensure that the concentrations of solvent residues are maintained below the odor and tainting threshold levels. The most widely used analytical technique to measure the levels of solvent residues is GC headspace analysis. Portions of the packaging are placed in sealed vials or other suitable containers and heated

at a set temperature for a prescribed period of time. The headspace is then sampled by means of a gas syringe or automatic sampling unit and injected into a suitable gas chromatograph with an FID. As the measurements are often carried out for quality control purposes, short heating times are sometimes used with external calibration and the measurements do not always accurately determine the solvent residues in the packaging, but do give reproducible results. This is the case with UK, BSI Standard BS6455 – Monitoring the levels of residual solvents in flexible packaging materials – and also the corresponding American ASTM Standard F 151-86. There are two draft EN standard methods in existence, prEN 13628-1 (absolute method) and prEN 13628-2 (industrial quality control monitoring method).

**See also:** **Adhesives and Sealants.** **Food and Nutritional Analysis:** Oils and Fats. **Infrared Spectroscopy:** Overview; Sample Presentation; Industrial Applications. **Liquid Chromatography:** Food Applications. **Plastics.** **Sensory Evaluation.**

### Further Reading

Ashby R, Cooper I, Harvey S, and Tice P (1997) *Food Packaging Migration and Legislation*. Leatherhead, UK: Pira International.

- Bradbury S and Bracegirdle B (1998) *Introduction to Light Microscopy*. Oxford: BIOS Scientific Publishers.
- Bradbury S and Evannett PJ (1996) *Contrast Techniques in Light Microscopy*. Oxford: BIOS Scientific Publishers.
- Briston JA and Katan LL (1974) *Plastics in Contact with Food*. London, UK: Food Trade Press Ltd.
- Crosby NT (1981) *Food Packaging Materials – Aspects of Analysis and Migration of Contaminants*. London: Applied Science Publishers Ltd.
- Food Contact Materials, Practical Guide (March 2002) *A Practical Guide for Users of European Directives*. European Commission, Health and Consumer Protection Directorate-General, <http://cpf.jrc.it/webpack/>.
- FSA (April 2001) Survey of Bisphenols in Canned Foods. Food Surveillance Information Sheet Number 13/01, Food Standards Agency.
- Haslam J, Willis HA, and Squirell DCM (1972) *Identification and Analysis of Plastics*. London: Iliffe Books.
- Kolb B (1984) Analysis of food contamination by head-space gas chromatography. In: Gilbert J (ed.) *Analysis of Food Contaminants*, pp. 117–156. Barking, UK: Elsevier.
- Krause, Lange, and Ezrin (1983) Schultheis KR (trans.) *Plastics Analysis Guide – Chemical and Instrumental Methods*. Munich: Carl Hanser Verlag.
- Lord AWT (2003) Packaging materials as a source of taints. In: Baigrie B (ed.) *Taints and Off-Flavours in Food*, pp. 64–111. London: Woodhead Publishing.
- MAFF (January 1999) Diisopropylnaphthalenes in Food Packaging Made from Recycled Paper and Board. Food Surveillance Information Sheet Number 169, MAFF Joint Food Safety Standards Group.
- Synoptic Document (updated 15 January 2002) European Commission, Health and Consumer Protection Directorate-General, <http://cpf.jrc.it/webpack/>.
- Tice PA (1993) Packaging as a source of taints. In: Saxby MJ (ed.) *Food Taints and Off-Flavours*, pp. 202–233. Glasgow: Blackie Academic Professional.

## FORENSIC SCIENCES

Contents

### Overview

### Alcohol in Body Fluids

### Arson Residues

### Blood Analysis

### Carbon Monoxide and Cyanide from Fire and Accident

### DNA Profiling

### Drug Screening in Sport

### Explosives

### Fibers

### Fingerprint Techniques

### Glass

### Gunshot Residues

### Hair

### Illicit Drugs

### Paints, Varnishes, and Lacquers

### Questioned Documents

### Systematic Drug Identification

### Thin-Layer Chromatography

### Volatile Substances

## Overview

**P Margot**, Université de Lausanne, Lausanne, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Forensic sciences group the scientific principles and technical methods applied to the investigation of crimes, litigations in civil matters, or regulatory and state administrative matters. Results are presented as



- Bradbury S and Bracegirdle B (1998) *Introduction to Light Microscopy*. Oxford: BIOS Scientific Publishers.
- Bradbury S and Evannett PJ (1996) *Contrast Techniques in Light Microscopy*. Oxford: BIOS Scientific Publishers.
- Briston JA and Katan LL (1974) *Plastics in Contact with Food*. London, UK: Food Trade Press Ltd.
- Crosby NT (1981) *Food Packaging Materials – Aspects of Analysis and Migration of Contaminants*. London: Applied Science Publishers Ltd.
- Food Contact Materials, Practical Guide (March 2002) *A Practical Guide for Users of European Directives*. European Commission, Health and Consumer Protection Directorate-General, <http://cpf.jrc.it/webpack/>.
- FSA (April 2001) Survey of Bisphenols in Canned Foods. Food Surveillance Information Sheet Number 13/01, Food Standards Agency.
- Haslam J, Willis HA, and Squirell DCM (1972) *Identification and Analysis of Plastics*. London: Iliffe Books.
- Kolb B (1984) Analysis of food contamination by head-space gas chromatography. In: Gilbert J (ed.) *Analysis of Food Contaminants*, pp. 117–156. Barking, UK: Elsevier.
- Krause, Lange, and Ezrin (1983) Schultheis KR (trans.) *Plastics Analysis Guide – Chemical and Instrumental Methods*. Munich: Carl Hanser Verlag.
- Lord AWT (2003) Packaging materials as a source of taints. In: Baigrie B (ed.) *Taints and Off-Flavours in Food*, pp. 64–111. London: Woodhead Publishing.
- MAFF (January 1999) Diisopropylnaphthalenes in Food Packaging Made from Recycled Paper and Board. Food Surveillance Information Sheet Number 169, MAFF Joint Food Safety Standards Group.
- Synoptic Document (updated 15 January 2002) European Commission, Health and Consumer Protection Directorate-General, <http://cpf.jrc.it/webpack/>.
- Tice PA (1993) Packaging as a source of taints. In: Saxby MJ (ed.) *Food Taints and Off-Flavours*, pp. 202–233. Glasgow: Blackie Academic Professional.

## FORENSIC SCIENCES

Contents

### Overview

### Alcohol in Body Fluids

### Arson Residues

### Blood Analysis

### Carbon Monoxide and Cyanide from Fire and Accident

### DNA Profiling

### Drug Screening in Sport

### Explosives

### Fibers

### Fingerprint Techniques

### Glass

### Gunshot Residues

### Hair

### Illicit Drugs

### Paints, Varnishes, and Lacquers

### Questioned Documents

### Systematic Drug Identification

### Thin-Layer Chromatography

### Volatile Substances

## Overview

**P Margot**, Université de Lausanne, Lausanne, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Forensic sciences group the scientific principles and technical methods applied to the investigation of crimes, litigations in civil matters, or regulatory and state administrative matters. Results are presented as

scientific expert evidence in the courts of law. The essence of forensic science is the study of traces, which, as a result, become information in the form of physical evidence or intelligence to lead investigation. The scientist assists the magistrates in determining the existence of a break in a rule (mostly crime), associate the evidence to a source (hopefully the author) or to a type of activity (determining that the action is the cause of the observed result). The traditional use of forensic science evidence is the identification of the author of a crime, but in recent years the value of physical evidence has taken a new dimension because of its value in reconstructing events (forensic intelligence) or because of its physical nature, it has become the only type of evidence that can be tested again to demonstrate wrongful convictions.

Physical evidence presents essential and unique features that make it so useful. It exists! This means that it can be tested repeatedly if necessary, as long as it is available, unlike statements (of witnesses, victims, or suspects which vary notoriously over time, or which may even be lies). The fact that it is physical evidence also means that there is no need for personal data for it to be useful. This allows extensive possibilities for data exchange without infringing personal data protection regulations. A trace in case A may be linked to a trace in case B across administrative borders. If the source of the trace of case A is known, then the source for the trace in case B may be known. With this information, investigation authorities may require international cooperation through the usual legal channel (bilateral agreement, Interpol, etc.). The study of physical evidence could be compared to the interrogation of a silent witness who never lies and protects human rights because it is not handling persons but only traces.

Locard in 1920 expressed the fundamental axiom, sometimes referred to the 'Locard exchange principle', that, in essence, every contact leaves a trace: any person committing a crime leaves something at the scene that was not there before and carries away traces that were not associated with him when he arrived. When an item is submitted to the laboratory for analysis, the usual questions are:

1. What is it? The answer to this question covers the physicochemical nature of the trace material.
2. Can it be associated with a source (usually a suspect) without any possible dispute? This relates to the specificity of the analyses or in terms usual in forensic science, to the identification or to the individualization of the source.
3. Is it the result of the action (the crime)? This question relates to the pertinence of the trace, whether the investigated action is its cause or the trace belongs to an unrelated action that happened before or after the investigation.
4. Is there another instance with the same type of trace, potentially from the same source, that demonstrates a potential link between two cases or more cases?
5. Finally, What are the risks that this identity might be accidental? That is, that another source may match the trace, other than the purportedly identified source. This relates to population studies used to find out the spread of specific attributes commonly analyzed and the risk of wrongful attribution to a specific source.

To answer these questions, there is a great need for all possible forms of chemical analyses to determine (1) the chemical make-up of a sample; (2) the frequency distribution of this type of material; (3) the permanence and stability of the trace; and (4) the variation in the composition over the population of materials of this type. The ultimate purpose is chemical identification, comparison, and, finally, interpretation of the results obtained in the context of the events.

One essential factor that differentiates this type of analyses from other analytical procedures is that the trace is the result of a past and uncertain action and its quantity cannot be modified. One can only assume that it is representative of the chemical make-up of the source and this assumption has to be taken into account when interpreting results. An essential limiting factor therefore comes from the sample that cannot be selected as in a true sampling procedure except when mass produced crime material is seized (illicit drugs, counterfeit).

The items may range from materials of natural origin (biological traces such as blood, hair, seminal stains, polymorphic enzymes, or drugs such as opiates, hallucinogens); materials of natural origin manufactured into a finished product (such as glass, paper, petroleum products), to completely synthesized and manufactured products (such as plastics, paints, fibers, drugs). Furthermore, large amounts may be available for analysis (e.g., in drug trafficking) or minute, submicrogram samples may be all that is available to the analyst (paint chip, arson residue, etc.). This requires precise and accurate analytical work on increasingly smaller amounts of material. Complex and modern methods of instrumental analysis therefore equip most modern forensic science laboratories. However, the methodology requires a stepwise progression from simple, general, rapid, widely applicable, and nondestructive screening methods, ranging from simple visual comparative examinations to the more discriminating and specific

methods, or combinations thereof, in order to reach the conclusion of identity.

For the results to be valid, it is essential that not only the items (crime or comparison samples) must be collected in a controlled, reproducible fashion, but also following an admissible legal procedure that varies from country to country (e.g., chain of evidence in common law; seals in other formal legal systems).

## General, Nondestructive Techniques

### Optical Examinations

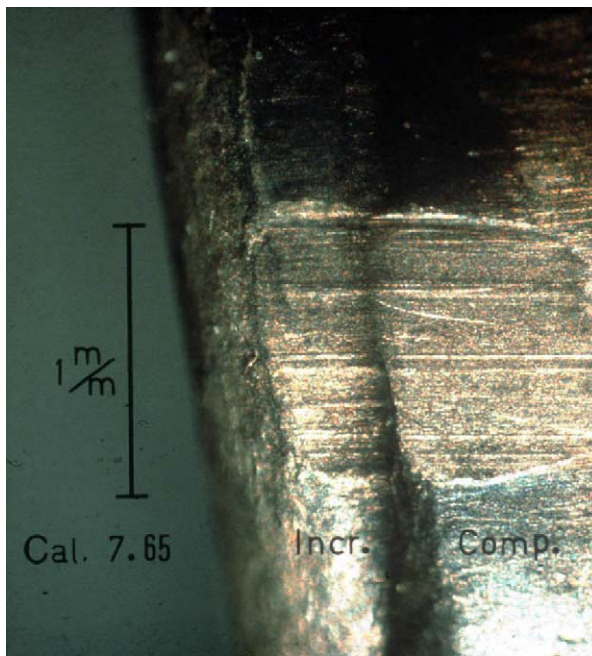
The first methods are simple optical examinations with the unaided eyes (transmitted, reflected, or grazing light) under natural daylight and filtered or unfiltered artificial light. These observations may be extended from ultraviolet (UV) to infrared (IR) using light converters (attached to video cameras). Attributes such as color, shape, design, dimensions, and surface quality allow rapid selection and discrimination of a large majority of products encountered in the investigation. This may lead to exclusion (elimination of suspects) or information useful in criminal investigation (criminal intelligence) and is an essential part of operational forensic science.

Microscopy extends the capabilities of our eyes; the morphology of small objects is precisely delineated (in the micrometer range for optical microscope down to the picometer range for the electron microscope), the fine details of their physical structure are analyzed, physical characteristics are precisely measured (length, width, angles, thickness, quantity), and optical attributes analyzed (color, opacity, refraction index, reflectance, phase change, photoluminescence, birefringence, phase contrast, etc.). Furthermore, microcomparisons by juxtaposition or superposition are obtained with specialized comparison microscopes (Figures 1 and 2). Not the least advantage is that microscopy produces an enlarged image that can be recorded photographically to document, demonstrate, identify, and measure objects to help the court (judge, jury) understand and objectively visualize what the forensic scientist has discovered.

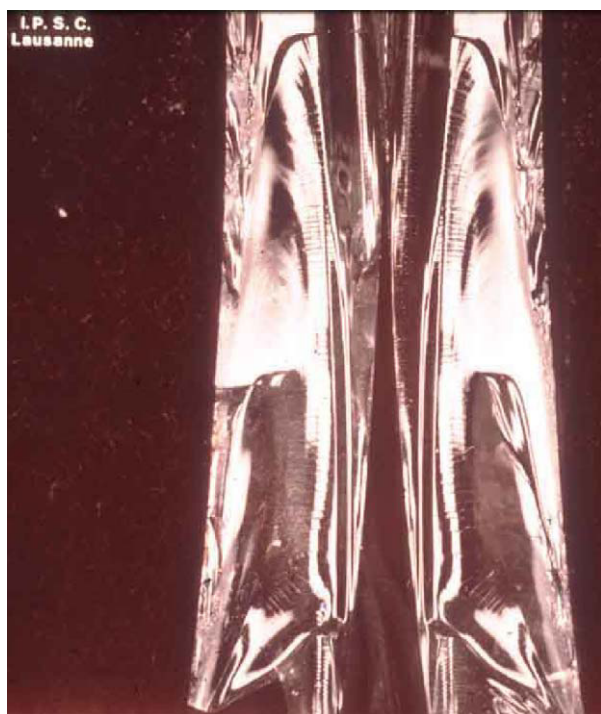
All evidence types showing complex morphological designs or patterns do not warrant further analyses: it is generally all that is needed to discriminate, eliminate, classify, and ultimately prove individual identity of broken glass (Figure 3), paint and paper by physical fit, as well as toolmarks, firearms, footmarks, handwriting, signatures, and fingerprints by comparison of complex accidental designs (although with the latter a chemical reaction is often needed to make the print visible).



**Figure 1** Comparison microscope with camera system. (Courtesy of Leica.)



**Figure 2** Identification of a firearm by comparison of the microstriations marked of the bullet. Left crime sample, right comparison sample.



**Figure 3** Conchoidal structure of glass fracture. Face view of the two sides of fractured glass shown side by side. The physical fit proves the singular identity of source.

### Spectroscopic Techniques

Depending on the interaction of electromagnetic energy with the matter constituting the evidential items, techniques that give simple spectra (such as with UV absorption spectrometry) permit discrimination only on the class level, i.e., showing the presence or absence of certain chromophores within the sample analyzed. Methods that result in complex spectra are much more selective, to the extent that they may individualize a chemical product. The morphologic complexity of the spectra is often abusively referred to as chemical 'fingerprinting', reflecting the uniqueness of a given product (e.g., the 'fingerprint' region of the IR spectrum between 500 and 2000 wave numbers (6–10  $\mu\text{m}$ )).

Three main types of nondestructive spectroscopic techniques are commonly used: absorption, emission, and physical techniques based on the modification of rays of electromagnetic energy (refraction, scattering, polarization, and refraction).

UV and visible absorption (also reflectance) microspectrometry has an extensive use in screening pigments and colors in paints, inks, and fibers, whereas it has found applications mostly for quantitative measurements in drug analysis (in particular for barbiturates). The development of sensitive IR spectrometers with Fourier transformation and

microspectrometers has meant rapid generalization of the technique for the analysis of organic materials from fibers, paint, glue, polymers, drugs, greases, and oils to inks, papers, and toners. Many IR databases have been produced that are extensively used in forensic analyses. One such database is the car paint database, now in use extensively throughout the world. Such databases allow exchanges because they concern material information rather than personal information covered by national personal data protection regulations.

Emission spectroscopic methods, due to their high intrinsic sensitivities, are techniques of choice much sought after in forensic science. Fluorimetry has been applied to the analysis of mineral oils, grease, inks, and drugs (e.g., LSD), whereas X-ray fluorescence has found extensive applications in the analysis of inorganic materials (e.g., paint pigments and glass).

Raman spectroscopy has shown great potential in forensic science because it is nondestructive, it has a high selectivity, and the analyses are fast. Methods to minimize interference and enhance the Raman signal caused this renewed interest.

The determination of refractive indices of glass is routinely performed and allows extensive discrimination and valuable evidence in cases of break-ins, aggressions, and hit-and-run accidents. The determination is obtained by immersion of the sample in an oil whose refractive index is calibrated on a hot plate under a microscope. The disappearance of the delineaation line (Becke's line) around the glass sample gives the temperature at which the glass and oil have the same refractive index.

The comparison of fibers and soil particles (and other anisotropic crystals) between two crossed polarizing filters under a microscope is a powerful and discriminating tool. The birefringence measurement gives important qualitative information on the products analyzed.

Scanning electron microscopy with energy dispersive X-ray spectrometry has found extensive applications in the analysis of glass, paint, fibers, and especially in the identification of gunshot residues (where it is the method of choice). This technique offers high resolution, great depth of field, and qualitative and quantitative information on small samples. However, although this technique has much to offer, its use in crime detection is often overrated.

In the late 1960s and the 1970s, neutron activation analysis has received much attention in forensic science because of its ultimate sensitivity combined with specificity. The early optimism has been dampened by the difficulty in interpreting results due to the variation of trace elements within samples at a given time and over a length of time. It has been used mainly in

glass, gunshot primer residue (determination of Pb, Ba, and Sb), and toxicological (As, Th) analyses.

## Destructive Techniques

All techniques based on a chemical reaction or which separate the components of a mixture, even if the reaction is reversible or the separation does not chemically modify the components of the mixture, are considered destructive in forensic science because the integrity of the sample in its original form is often paramount in the criminal trial (it offers the warranty that the analyses can be tested and reproduced by an independent analyst).

## Spectroscopic Techniques

Atomic absorption spectroscopy has been at the forefront of inorganic analyses in forensic sciences from the late 1960s. With the development of non-flame atomizers (carbon furnace) and high-intensity lamps (electrodeless discharge, hollow cathode), many elements of interest, due to their high toxicity, are currently routinely analyzed (As, Th, Sb, Cd, etc.). It has found applications in hair, fingernail, paint, fuels, lubricants, fibers, polymers, papers, pharmaceutical, gunshot residue, forged and counterfeited coins analyses. However, because of the high sensitivity and the need for sample preparation, background noise and high impurity levels hamper a more routine use of this technique.

Laser and inductively coupled plasma (ICP) emission spectroscopy constitute powerful multielement qualitative and quantitative tools for the analysis of metals, paints, and glass. Finally, ICP with mass spectrometry (MS) makes an increasing impact on the analyses of trace evidence from pollution source determination to traditional trace exploitation.

MS is accepted as an indispensable technique in the analysis of drugs and their fate in the body. Lately, there has been an increase in the frequency of xenobiotic studies that have taken advantage of tandem MS to screen complex biological samples for the presence of drug-related products. But mostly MS is used in combination with a separation technique (e.g., chromatography) in forensic science.

## Chemical Reactions

The need for rapid results, often in the field, in order to detect dangerous products (e.g., explosives) and/or products subject to trafficking (e.g., drugs), has led to the development of many chemical screening tests that are commonly used by police technicians. This involves mostly color or crystal tests that require confirmation by laboratory analyses.

Other color reactions employed in forensic science are simply used to create a contrast and make structural entities visible for taxonomical or identification purposes, especially in connection with microscopical examination (e.g., coloration of sperm heads with Baecchi's stain – acid fuchsin S and methyl blue).

Immunological methods have become the standard screening methods in forensic toxicology laboratories and have provided sensitive qualitative and quantitative means of testing blood and urine. Radioimmunoassay kits exist for the testing and measurement of opiates, amphetamines, cocaine, barbiturates, cannabinoids, benzodiazepines, etc. The introduction of monoclonal antibodies has allowed the development of increasingly specific tests. Various immunological techniques have been devised using enzyme mediated amplification to avoid the drawbacks of using radioactive materials (e.g., enzyme multiplied immunological technique and enzyme linked immunosorbant assay). All require analytical confirmation for legal purposes.

## Analytical Separations

Most products or contact traces found as evidence are made up of complex combinations of molecules and atoms that constitute our environment. This complexity makes that the world is infinitely variable and each item is unique in its make-up and design. This is exploited to individualize, categorize, or exclude the identity of samples from contacts or any other event connected with a crime. This process necessitates the extraction and separation of often minor components of material and requires careful and optimal separation techniques. From simple sample preparation (such as liquid–liquid extraction) in toxicological analyses, the forensic scientist may use the whole range of separation techniques from thin layer chromatography (inks, fiber dyes, drugs, etc.), gas chromatography (GC) (arson accelerants, drugs, solvents, ethanol, etc.), and high-performance liquid chromatography (drugs, paints, inks, oils, etc.) to capillary electrophoresis (inks, proteins, enzymes, DNA fragments), isoelectric focusing (blood products, serology, proteins), isotachopheresis, etc. The separation may be based on differences in solubility, size, polarity, volatility, electric charge, etc. Field tools based on separation methodology have been developed and are now in common use at scenes of crime or for security checks (airports, public administrations, military, police, or custom offices, etc.) to detect materials such as illicit drugs and explosives.

All separation techniques are used in combination with other analytical methods designed to make products visible (on a thin-layer chromatogram, for

example) or to give an electrical signal at the end of a capillary in GC (e.g., flame ionization detection). With spectrometric detection, an additional identification capability (such as GC–MS or high-performance liquid chromatography (HPLC)–UV/visible diode array detection, etc) increases the information content of the analysis. This allows often powerful discrimination that may require interpretation.

Developments in the analysis of genetic material (DNA profiling) consist of first isolating genetic material (DNA) by extraction followed by enzymatic digestion with restriction enzymes of specific polymorphic sections of the DNA molecule. These fragments are then multiplied by splitting the two DNA strands and synthesizing their complementary strands with a natural polymerase using thermal cycles. After a determined number of cycles, the multiplied DNA fragments are separated and detected by employing an electrophoretic method. In that manner, complex biochemical mechanisms are used to produce an extremely selective and sensitive analytical result to be presented before the Court of Justice.

## Interpretation

In order to distinguish between samples of different origin that contain fairly similar amounts of the same compounds, accurate and precise quantitative measurements of very low concentrations (in the ppb range) are often necessary. Statistical analysis can therefore represent a challenging problem usually mastered by most analysts.

When samples are not discriminated, the problem of the interpretation of evidence becomes primordial and much more difficult. There is some need to know the risk or the probability that two items come from different sources, but are not discriminated. This is expressed by the discrimination power of the attribute analyzed that defines the effectiveness of that attribute to differentiate samples in forensic science. Extensive databases are used to compute the probability that two randomly selected samples will be discriminated when the attribute is analyzed or measured. This probability is, in effect, the discrimination power that has been calculated for uncorrelated, correlated, and continuous variables such as blood grouping systems, drugs, glass, paints, etc.

This probability is useful; however, it only gives an indication to the forensic scientist as to his ability to distinguish between closely related, but different, materials. It does not give an indication of the value of the evidence in a given situation, i.e., the probability of finding this type of evidence material on a person connected or inversely, unconnected with a

crime. Therefore, other data such as the preeminence or rarity of given materials, their retention capacity on the victim or assailant in a crime situation, and their variation over a population of interest are necessary. These data are collected to estimate the risk of finding a particular type of evidence on an innocent person as compared to the chance of finding it on the guilty party. It all comes down to the smaller the coincidence probability, the greater the evidential value of a match. This is the object of intensive and extensive research in a number of laboratories around the world, and it is here that real progress can be expected in the near future, rather than in the development of new analytical techniques. The treatment of the data is the object of Bayesian statistics that are now being applied to intelligent knowledge-based systems to interpret the value of evidence under a given set of circumstances.

In forensic intelligence and crime analysis, data may be treated differently. It is less essential to avoid false incrimination. The essence in investigation is to have reliable information that shows potential links between crimes or events over time and space. Extracting cases from complex files overloaded with fuzzy information and showing potential links allows the review of these cases seen as a group (important in serial crimes). The demonstration of links creates a picture that becomes clearer with each new data of what can be described as a crime series whose profile can determine regularities in the activities of the criminal or criminal group. This has developed into a very profitable and proactive methodology to fight against serial crimes. Forensic science offers solid and reliable evidence that helps remove linkage blindness unlike many attempts at profiling that increase fuzziness by combining complex information whose reliability is often unknown. Forensic intelligence is making strong advances and is credited with promising successes.

*See also:* **Forensic Sciences:** Alcohol in Body Fluids; Arson Residues; Blood Analysis; Carbon Monoxide and Cyanide from Fire and Accident; DNA Profiling; Drug Screening in Sport; Explosives; Fibers; Fingerprint Techniques; Glass; Gunshot Residues; Hair; Paints, Varnishes, and Lacquers; Questioned Documents; Systematic Drug Identification; Thin-Layer Chromatography; Volatile Substances. **Gas Chromatography:** Forensic Applications. **Mass Spectrometry:** Forensic Applications. **Microscopy Applications:** Forensic.

## Further Reading

Aitken CGG and Taroni F (2004) *Statistics and the Evaluation of Evidence for Forensic Scientists*, 2nd edn. Chichester: Wiley.



- Inman K and Rudin N (2000) *Principles and Practice of Criminalistics: The Profession of Forensic Science*. Boca Raton, FL: CRC Press.
- James SH and Nordby JJ (2002) *Forensic Science: An Introduction to Scientific and Investigative Techniques*. Boca Raton, FL: CRC Press.
- Locard E (1920) *L'enquête criminelle et les méthodes scientifiques*. Paris: Flammarion.
- Ribaux O and Margot P (1999) Inference structures for crime analysis and intelligence using forensic science data: the example of burglary. *Forensic Science International* 100: 193–210.
- Ribaux O and Margot P (2003) Case based reasoning in criminal intelligence using forensic case data. *Science and Justice* 43: 135–143.
- Robertson B and Vignaux GA (1995) *Interpreting Evidence*. Chichester: Wiley.
- Saferstein R (ed.) (2001) *Forensic Science Handbook*, 2nd edn, vol. I. Englewood Cliffs, NJ: Prentice Hall.
- Saferstein R (ed.) (1988) *Forensic Science Handbook*, vol. II. Englewood Cliffs, NJ: Prentice Hall.
- Saferstein R (ed.) (1993) *Forensic Science Handbook*, vol. III. Englewood Cliffs, NJ: Prentice Hall.

## Alcohol in Body Fluids

F Moriya, Kochi University, Kochi, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Alcohol commonly refers to ethanol that is found in beverages made since ancient times. Alcoholic beverages are consumed worldwide except in some countries where drinking is prohibited for religious and cultural reasons. The excessive consumption of alcoholic beverages has a significant impact on society, and ethanol is the leading toxic substance handled in clinical and forensic toxicology. Traffic accidents associated with alcohol consumption are a serious social problem, and more than 1 million driver's licenses per year have been revoked worldwide because of inebriation. Annual fatalities associated with alcohol consumption occur in more than 15 000 people in the European Union and 10 000 people in the USA. In Japan, 8000–9000 people have been involved in fatal traffic accidents annually and ~15% of them were victims of drunken driving. Legal alcohol limits for driving are different among countries. In the USA, 49 states (all except Massachusetts) have laws defining it as a crime to drive with a blood alcohol concentration at or above  $0.8 \text{ mg ml}^{-1}$  in 34 states and  $1.0 \text{ mg ml}^{-1}$  in 15 states. In other countries, the legal limits are in the range  $0.2$ – $0.8 \text{ mg ml}^{-1}$  or mg per g. In Sweden and Japan, drunken driving is more strictly defined as blood alcohol concentrations of  $0.2 \text{ mg per g}$  or more and  $0.3 \text{ mg ml}^{-1}$  or more, respectively.

Widmark reported the first analytical method for measuring ethanol in body fluids in 1922. This method was based on the oxidation of ethanol with potassium dichromate followed by titrimetric analysis.

Later, an enzymatic method based on oxidation of ethanol by alcohol dehydrogenase (ADH) and nicotinamide adenine dinucleotide (NAD) followed by spectrophotometric analysis was reported. Current methods for detecting ethanol in body fluids are predominantly based on physicochemical techniques. A gas–liquid chromatographic (GLC) method is the most widespread because it is easy, rapid, and has high specificity and accuracy. Analytical methods used to determine alcohol in body fluids are summarized in Table 1.

### Sample Collection and Handling

#### Living Persons

Blood is usually collected from the cubital vein using evacuated glass tubes (5 or 10 ml vacutainer tubes)

**Table 1** Analytical methods used to determine alcohol in body fluids

Analytical method	Principle of detection
Chemical oxidation	Reduction of potassium dichromate (yellow–orange) to chromium(III) ion (green) in the presence of alcohol and sulfuric acid
Enzymatic oxidation	Reduction of nicotinamide adenine dinucleotide ( $\text{NAD}^+$ ) to NADH (UV absorbed at 340 nm) in the process of alcohol oxidation catalyzed by alcohol dehydrogenase (ADH)
Gas chromatography (direct injection)	Injection of a liquid sample mixed with an internal standard and detection with flame ionization detector
Gas chromatography (headspace injection)	Injection of headspace vapor after incubation of a closed vial containing a liquid sample mixed with an internal standard and detection with flame ionization detector

- Inman K and Rudin N (2000) *Principles and Practice of Criminalistics: The Profession of Forensic Science*. Boca Raton, FL: CRC Press.
- James SH and Nordby JJ (2002) *Forensic Science: An Introduction to Scientific and Investigative Techniques*. Boca Raton, FL: CRC Press.
- Locard E (1920) *L'enquête criminelle et les méthodes scientifiques*. Paris: Flammarion.
- Ribaux O and Margot P (1999) Inference structures for crime analysis and intelligence using forensic science data: the example of burglary. *Forensic Science International* 100: 193–210.
- Ribaux O and Margot P (2003) Case based reasoning in criminal intelligence using forensic case data. *Science and Justice* 43: 135–143.
- Robertson B and Vignaux GA (1995) *Interpreting Evidence*. Chichester: Wiley.
- Saferstein R (ed.) (2001) *Forensic Science Handbook*, 2nd edn, vol. I. Englewood Cliffs, NJ: Prentice Hall.
- Saferstein R (ed.) (1988) *Forensic Science Handbook*, vol. II. Englewood Cliffs, NJ: Prentice Hall.
- Saferstein R (ed.) (1993) *Forensic Science Handbook*, vol. III. Englewood Cliffs, NJ: Prentice Hall.

## Alcohol in Body Fluids

F Moriya, Kochi University, Kochi, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Alcohol commonly refers to ethanol that is found in beverages made since ancient times. Alcoholic beverages are consumed worldwide except in some countries where drinking is prohibited for religious and cultural reasons. The excessive consumption of alcoholic beverages has a significant impact on society, and ethanol is the leading toxic substance handled in clinical and forensic toxicology. Traffic accidents associated with alcohol consumption are a serious social problem, and more than 1 million driver's licenses per year have been revoked worldwide because of inebriation. Annual fatalities associated with alcohol consumption occur in more than 15 000 people in the European Union and 10 000 people in the USA. In Japan, 8000–9000 people have been involved in fatal traffic accidents annually and ~15% of them were victims of drunken driving. Legal alcohol limits for driving are different among countries. In the USA, 49 states (all except Massachusetts) have laws defining it as a crime to drive with a blood alcohol concentration at or above  $0.8 \text{ mg ml}^{-1}$  in 34 states and  $1.0 \text{ mg ml}^{-1}$  in 15 states. In other countries, the legal limits are in the range  $0.2$ – $0.8 \text{ mg ml}^{-1}$  or mg per g. In Sweden and Japan, drunken driving is more strictly defined as blood alcohol concentrations of  $0.2 \text{ mg per g}$  or more and  $0.3 \text{ mg ml}^{-1}$  or more, respectively.

Widmark reported the first analytical method for measuring ethanol in body fluids in 1922. This method was based on the oxidation of ethanol with potassium dichromate followed by titrimetric analysis.

Later, an enzymatic method based on oxidation of ethanol by alcohol dehydrogenase (ADH) and nicotinamide adenine dinucleotide (NAD) followed by spectrophotometric analysis was reported. Current methods for detecting ethanol in body fluids are predominantly based on physicochemical techniques. A gas–liquid chromatographic (GLC) method is the most widespread because it is easy, rapid, and has high specificity and accuracy. Analytical methods used to determine alcohol in body fluids are summarized in Table 1.

### Sample Collection and Handling

#### Living Persons

Blood is usually collected from the cubital vein using evacuated glass tubes (5 or 10 ml vacutainer tubes)

**Table 1** Analytical methods used to determine alcohol in body fluids

Analytical method	Principle of detection
Chemical oxidation	Reduction of potassium dichromate (yellow–orange) to chromium(III) ion (green) in the presence of alcohol and sulfuric acid
Enzymatic oxidation	Reduction of nicotinamide adenine dinucleotide ( $\text{NAD}^+$ ) to NADH (UV absorbed at 340 nm) in the process of alcohol oxidation catalyzed by alcohol dehydrogenase (ADH)
Gas chromatography (direct injection)	Injection of a liquid sample mixed with an internal standard and detection with flame ionization detector
Gas chromatography (headspace injection)	Injection of headspace vapor after incubation of a closed vial containing a liquid sample mixed with an internal standard and detection with flame ionization detector

containing ethylenediamine tetra-acetic acid, citric acid, or heparin as an anticoagulant and/or sodium fluoride as a preservative. Disposable plastic syringes and needles are suitable for obtaining samples and transferring them to appropriate containers containing a suitable anticoagulant and preservative. The skin should be cleaned with soap and water or an alcohol-free swab (e.g., 0.1–0.5% chlorhexidine gluconate solution) before inserting the needle. In our experience, the risk of contaminating blood with alcohol is negligible even after the skin is swabbed with 80 v/v% ethanol; however, to avoid unnecessary battles in the courtroom swabs containing ethanol or isopropanol should not be used before taking blood samples. Ethanol in blood is relatively stable at 4°C for as long as 1 month when tubes are sealed. A negligible amount of ethanol, 0.02–0.03 mg ml<sup>-1</sup> a month, is converted to acetaldehyde due to a non-enzymatic oxidation by oxyhemoglobin. The rate of ethanol loss actually depends on how often and how long the tubes have been opened at room temperature. In our laboratory, the maximum loss of ethanol from blood stored at 4°C, 25°C, and 37°C for 30 min in containers without caps was ~1%, 3%, and 7%, respectively.

Urine is also a good specimen for detecting ethanol because it is easy to obtain and handle. However, the ethanol concentration represents an average value between the last voided time and sampling time. Although ethanol is stable in urine, it is preferable to transfer urine to a container containing sodium fluoride at a final concentration of 1–2% and store at 4°C prior to analysis.

Saliva is used for on-site ethanol testing using commercially available testing devices. Saliva is absorbed in a pad on the test strip under the tongue until it is saturated. The volume of saliva needed for the current commercial tests is small and can be collected in 1 or 2 min or less.

### Dead Persons

Various body fluids, in addition to blood and urine, are available in forensic autopsy cases. These fluids include cerebrospinal fluid, vitreous humor, pericardial fluid, and bile. Samples can be collected with disposable syringes, with or without needles, and transferred to glass containers (5–10 ml) containing sodium fluoride at a final concentration of 1–2% as a preservative. As samples from living persons, the containers should be preserved at 4°C until analysis. Special care should be taken while testing for ethanol in postmortem samples since intestinal bacteria diffuse through the body after death via vascular system and induce putrefaction. In moderate to heavily

decomposed bodies, proliferating bacteria can produce ethanol levels as high as 1–2 mg ml<sup>-1</sup> due to anerobic metabolism of carbohydrates. Unfortunately, this complicates the interpretation of analytical results for ethanol. There is, so far, no definitive method to determine how much ethanol is produced after death. Currently, the most reliable indicator of postmortem ethanol production is n-propanol, which is not found in alcoholic beverages. Bacteria produce both n-propanol and ethanol after death; however, the amount of n-propanol in blood is usually at or less than 1/20 times the amount of ethanol. Thus, if detectable amounts of n-propanol exist in postmortem samples, it is obvious that postmortem ethanol production also occurred. Postmortem ethanol production in vitreous humor and urine is slower than other body fluids. However, it is often difficult to obtain vitreous humor in decomposed bodies. Postmortem ethanol production in urine of diabetic patients is increased due to bacterial fermentation of urinal glucose.

Another issue in postmortem testing for ethanol is postmortem diffusion of ethanol from the stomach into the surrounding tissues. This phenomenon becomes evident within 48 h after death of individuals who died shortly after drinking sessions. In these cases, cardiac blood, blood in the thoracic and abdominal large vessels, pericardial fluid, and bile show falsely elevated levels of ethanol. Femoral venous blood, urine, and cerebrospinal fluid are relatively spared from ethanol diffusion from the stomach. However, vitreous humor is the best specimen of choice due to its anatomical location.

### Breath Alcohol Testing

In 1955, breath analysis for alcohol was first used for traffic law enforcement in the USA. This technique is now in routine use in many countries as a substitute for the invasive blood sampling in apprehended drivers. Police officers can obtain on-the-spot results of breath alcohol tests using portable evidential breath testing devices. The distribution of alcohol between blood and alveolar air obeys Henry's law, which states that at a given temperature, a direct relationship exists between the amount of alcohol dissolved in blood and the amount of it in the alveolar air. Although the blood–breath alcohol ratio is affected by several factors such as the arterial–venous differences in alcohol concentrations during alcohol absorption, hematocrit, body temperature, and pulmonary diseases, it is usually in the range of 2000:1–2300:1. For forensic purposes, the value of 2000:1 is employed in Austria, France, and Japan, the value of 2100:1 in Germany, Norway, Sweden,

and the USA, and the value of 2300:1 in the Netherlands and the UK.

Breath alcohol analysis is also useful for diagnostic purposes especially at the emergency room. Based on bedside results of breath alcohol tests, emergency doctors can determine whether observed symptoms, signs, or injuries are drinking-related.

An erroneous result may be obtained when contaminated breath with residual alcohol from vomiting or eructating in the mouth is delivered. However, these can be avoided by monitoring the subject for ~15 min prior to breath sampling to make sure that nothing enters the mouth. Residual alcohol is rapidly removed by salivary flow, reflex swallowing, and absorption into the buccal mucosa. Analyzing two breath samples from a subject may be another safeguard. The results agreeing within  $\pm 0.02$  g per 210 l are acceptable.

## Chemical Methods

Owing to its volatility, alcohol can be separated from biological samples and subsequently measured using oxidizing agents such as potassium dichromate, potassium permanganate, and osmic acid. For example, the yellow-orange dichromate ion is reduced to the blue-green chromium(III) ion in strong sulfuric acid solution in the presence of alcohol. Widmark in 1922 developed one of the earliest diffusion methods based on this reaction.

### Widmark Microdiffusion Method

Two milliliters of mixed solution of potassium dichromate and sulfuric acid is added to a 50 ml Widmark flask and 0.2 ml of a body fluid is placed in a glass spoon whose stem is fused to the center of a glass stopper that seals the flask. The stoppered flask is heated at 50–60°C for 2 h to allow the simultaneous distillation and oxidation reaction to be completed, and cooled at room temperature. The flask is unsealed and 25 ml of distilled water is added to the flask. After cooling, 0.5 ml of potassium iodide solution is added and precipitated iodine is titrated with 0.01 N sodium thiosulfate using 1% starch in sodium chloride-saturated water as an indicator. The amount of alcohol (mg) is calculated by the following formula:  $0.113 \times (\text{consumed volume (ml) of sodium thiosulfate solution in the sample} - \text{consumed volume (ml) of sodium thiosulfate solution in a diluted intact dichromate solution})$ .

### Conway Cell Microdiffusion Method

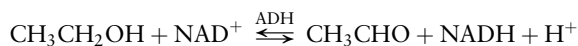
A Conway cell is a porcelain dish containing two concentric wells. A biological sample is added to the

outer well and a color reagent (the combined solution of potassium dichromate and sulfuric acid) is added to the inner well. The cell is covered and is heated in an oven for an hour or allowed to remain overnight at room temperature. Ethanol diffuses out of the sample into the center well where the chemical reaction occurs. This color test can be observed visually, making this a simple screening method. Semi-quantitation can be made by preparing standards with known amounts of ethanol and comparing the intensity of color of a sample with the standards. Sample can also be quantified spectrophotometrically at 450 nm.

However, this method is not specific for ethanol. Various alcohols, aldehydes, and ketones that are produced in dead bodies during putrefaction also convert dichromate to the chromic ion and give falsely elevated levels of ethanol.

## Enzymatic Methods

ADH is used to measure ethanol in body fluids. ADH oxidizes ethanol to acetaldehyde in the presence of a coenzyme  $\text{NAD}^+$ , which is reduced to NADH that is measured spectrophotometrically at 340 nm:



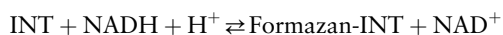
Bonnichsen and Theorell in 1951 were the first to use ADH for analyzing ethanol in blood, and for several commercial kits, such as Blood Alcohol UV Test 'BMY' (Boehringer Mannheim) and Alcohol Kit (Sigma), are available for measuring ethanol in whole blood, serum, plasma, and urine by using this method.

### A Representative Spectrophotometrical Method

To a glass centrifuge tube, containing 4 ml of 0.33 M perchloric acid, 0.5 ml of a body fluid is added. The capped tube is vigorously vortexed and centrifuged at 3000 rpm for 5 min. To a test tube containing 4.8 ml of pH 8.7 semicarbazide buffer and 0.1 ml of NAD are added 0.1 ml of supernatant of the deproteinized sample and 0.02 ml of ADH (32 IU or more). The tube is capped and incubated at 37°C for 25 min, and absorption is measured at 340 nm. Distilled water is used as a negative control. Ethanol levels are determined with a standard curve prepared by using known amounts of ethanol.

On-site kits for measuring saliva ethanol are available from Chematics, Inc. (ALCO-SCREEN 02), Roche Diagnostic Systems (On-Site Alcohol) and STC Diagnostics Inc. (Q.E.D. A 150 or 350 Saliva Alcohol Test). All of these kits use the same methodology for detecting alcohol. NADH reacts with

indonitrotetrazolium violet dye (INT) and a hydrogen acceptor to form a highly colored formazan compound:



Sample should be collected at least 10 min after the consumption of food or beverages, and tests are completed within 3 min, according to the manufacturers' instructions.

Automated enzymatic systems are available for measuring ethanol in body fluids for clinical and forensic purposes. EMIT is a popular technique for detecting drugs of abuse in urine, serum, and plasma. EMIT ETS PLUS ethyl alcohol assay was designed for the Syva ETS PLUS analyzer. EMIT is a homogeneous immunoassay which utilizes the enzyme glucose-6-phosphate dehydrogenase activity, instead of ADH, to reduce NAD to NADH. Abbott Laboratories introduced a radioactive energy attenuation (REA) assay for determining ethanol on the TDx analyzer, microprocessor-controlled automated fluorometer used for fluorescence polarization immunoassay for many therapeutically monitored drugs. In the REA method, the enzyme-catalytic reactions of ADH and diaphorase results in the formation of reduced thiazolyl blue dye (MTT), a purple color chromogen that has an absorbance peak at 565 nm. The fluorescence intensity of fluorescein, which is a fluorescent indicator added to the reaction, decreases logarithmically with increasing concentration of reduced MTT.

ADH does not react with methanol or acetone but reacts to a lesser degree with n-propanol, isopropanol, and n-butanol. EMIT is free of interference from small molecular weight alcohols, aldehydes, ketones, and glycols. Automated enzymatic methods are useful for clinical purpose because a large number of samples can be handled in a short time period. In the forensic laboratory, however, chemical methods for determining ethanol in biological samples are not generally used because they are not specific.

## GLC Methods

Gas-liquid chromatography (GLC) is the most commonly used method for the forensic and clinical analysis of ethanol. Cadman and Johns in 1958 first reported a GLC method for determining blood ethanol concentrations. They extracted ethanol from samples prior to analysis by adding equal volume of n-propyl acetate (1 ml) to a blood sample (1 ml) to extract ethanol. No internal standard was used. A fixed volume of extract (35  $\mu$ l) was injected into a GL chromatograph equipped with a thermal

conductivity detector (TCD) and a column containing 28 g of a mixture consisting of Flexol 8N8, diisodecyl phthalate, and poly(ethylene glycol) (PEG) 600 (15:10:3 by weight on 100 g of 42/60 mesh firebrick). The levels of ethanol were determined using a standard curve prepared by adding known amounts of ethanol to blood. Subsequently, other methods have been developed that use protein precipitation, distillation, direct injection, and headspace techniques.

Current methods use direct injection and headspace techniques because they are easy, very sensitive, and reproducible. While the flame ionization detector (FID) is not very sensitive to water, it is very sensitive to organic compounds and is well suited for analyzing biological samples. Generally, nitrogen is used as a carrier gas and chromatographic separations are carried out isothermally at an oven temperature of 80–100°C. Recently, alcohols were analyzed in blood using capillary or wide-bore columns. PoraPLOT Q, a fused silica, was the first capillary column used for analyzing short-chain alcohols. Graphitized carbon coated with 5% Carbowax (= PEG) 20 M or 0.2% Carbowax 1500 was ideal for analyzing short-chain alcohols. However, packed columns dominate the procedures used in forensic laboratories, because capillary columns require more maintenance and strict conditioning compared with packed columns. In addition, capillary columns with bonded phases are vulnerable to water found in body fluids. Moreover, capillary columns are prone to overloading because of less sample loading capacity compared with packed columns. The columns for GLC analysis of alcohols are usually 1–2 m long by 0.3 cm ID and made of glass or stainless steel. Stationary phases commonly used for alcohol analysis are Carbowax or PEG 400, 600, 1000, 1500, or 20 M at a loading of 5–25 w/w% on such a solid support as Carbowax, Celite, Chromosorb, or Shimalite. Porous polymer materials (styrene-divinylbenzene) such as Porapak P and Q have also been used as packing materials.

## Direct Injection Techniques

The easiest and fastest method involves the direct injection of a diluted sample with an internal standard solution into a GL chromatograph. We recommend an ultramicro-method utilizing direct injection. Aliquots of a body fluid and an internal standard, such as n-propanol, are mixed together in equal volumes. A 1  $\mu$ l aliquot of the mixture is injected into a GL chromatograph equipped with a glass column (2 m long by 0.3 cm ID) packed with 0.2% Carbowax 1500 on 80/100 Carbowax C and an

FID. The column is filled with a precolumn glass insert filled with a loosely packed silanized glass wool plug that acts as a trap for the nonvolatile proteins in the sample. Temperatures of injection port and column are 120°C and 100°C, respectively. Ethanol concentrations in body fluids are determined from a calibration curve constructed by plotting ethanol concentration of standard ethanol solution versus peak area ratios of ethanol to the internal standard. In analyzing body fluids obtained from decomposed bodies, 2-butanol rather than n-propanol is suitable as internal standard because n-propanol is produced postmortem in various degrees. Unfortunately, the repeated injection of biological material into the chromatograph causes a build-up of non-volatile substances in the injector and the front end of the analytical column, which causes more maintenance and column replacement. These problems can be alleviated by using the headspace procedure.

### Headspace Techniques

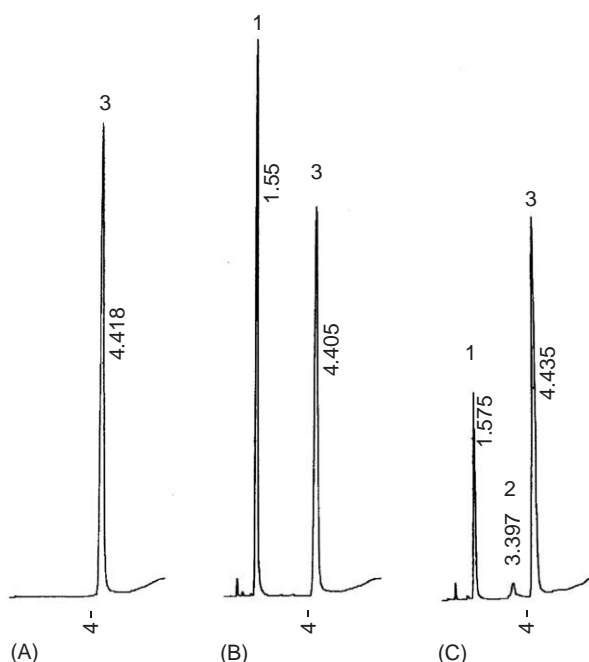
Curry in 1962 first described headspace techniques, which are most common worldwide. Headspace techniques are based on Henry's law, which states that the ratio of a dissolved substance in aqueous solution to the substance in the vapor in a sealed glass vial is constant, depending on temperature. The concentrations of volatile substances, such as alcohols in a body fluid, can be determined by analyzing the vapor, which is in equilibrium with the body fluid, in a sealed glass vial at an increased temperature. A sample of the headspace is removed from the vial using a gas-tight syringe and injected into a GL chromatograph. The following is a procedure for detecting ethanol in body fluids in forensic autopsy cases including decomposed bodies and victims suspected of the existence of alcohols and/or other volatiles.

A 0.5 ml portion of a body fluid and 0.5 ml of an internal standard ( $1 \text{ mg ml}^{-1}$  2-butanol) is placed in a 15 ml glass vial, which is capped with a silicon rubber and sealed with an aluminum cap. The sealed vial is incubated in an aluminum block at 55°C for 15 min and 1 ml of the air phase is injected into a GL chromatograph equipped with a 2 m long by 0.3 cm ID glass column packed with 5% Carbowax 20 M on 60/80 Carbowax B. Temperature of the injection port and detector is 130°C. The column temperature is initially maintained at 80°C for 2 min and then increased to 130°C at a rate of  $10^\circ\text{C min}^{-1}$ . The final temperature is maintained for 1 min. Flow pressure of the carrier gas is 150 kPa. This procedure is summarized in Table 2. In this method, no interfering

peaks appear at retention times for ethanol and the internal standard. Typical gas chromatograms of blood samples from a nondrinker, drinker, and decomposed body free of exogenous ethanol are shown in Figure 1. Additionally, a 1 m long by 0.3 cm ID glass column packed with 25% PEG 1000 on 60/80

**Table 2** An example of headspace GL chromatography for detecting alcohol in body fluids of living and dead persons

Sample preparation	A 0.5 ml portion of a body fluid mixed with an equal volume of internal standard ( $1 \text{ mg ml}^{-1}$ 2-butanol) in a 15 ml glass vial. Sealed with a silicon rubber and an aluminum cap
Incubation	Equilibrated at 55°C for 15 min
Sampling	Approximately 1 ml of headspace vapor drawn using a gas-tight syringe and injected into a GL chromatograph
Analysis	Separated on a 2 m long by 0.3 cm ID glass column packed with 5% Carbowax 20 M on 60/80 Carbowax B. Column temperature programs: 80°C (2 min) at $10^\circ\text{C min}^{-1}$ to 130°C (1 min). Nitrogen used as the carrier gas at a flow pressure of 150 kPa.
Quantitation	Performed by comparing the peak area ratio of ethanol to the internal standard in the sample with that in a standard ethanol solution



**Figure 1** Gas chromatograms obtained from analyses of headspace vapors in equilibrium with blood samples from (A) nondrinker, (B) drinker with ethanol concentration of  $3.34 \text{ mg ml}^{-1}$ , and (C) decomposed body free of exogenous ethanol. 1: ethanol; 2: n-propanol; and 3: 2-butanol (internal standard).



Shimalite is very useful for routine ethanol testing. *t*-Butanol ( $0.2 \text{ mg ml}^{-1}$  in distilled water, 0.5 ml) is used as an internal standard and isothermal analysis is performed at  $80^\circ\text{C}$ . Flow pressure of the carrier gas is 40 kPa. On this column, ethanol elutes right after the internal standard, which completes within 2.5 min. This is advantageous when handling a large number of samples. However, a caution must be exercised because ethanol peak is interfered with peaks of methanol and isopropanol. Results should be confirmed by reanalyzing samples using another stationary phase such as Proopack Q, or Carbowax B containing 5% Carbowax 20 M.

In performing headspace techniques, partition coefficients of analytes vary depending on the sample matrix (matrix effects), which can affect the results. Ethanol distributes in body fluids according to their water contents. If inorganic salts, such as sodium chloride and sodium fluoride, are present in excessive amounts, then the ratio of alcohol in aqueous solution to the vapor decrease (salting-out effects). Utilizing salting-out effects by adding as much as 1.5 g of sodium chloride or potassium carbonate to 1 ml of blood, the sensitivity of the headspace method is enhanced and very low concentration of ethanol can be determined. However, this salting-out step varies among matrices and alcohols, and it does not eliminate matrix effects. The three-carbon *n*-propanol seems to be salted-out more effectively than the two-carbon ethanol. The best way to eliminate matrix effects is to dilute the fluids. However, in routine biological fluids, even if they are decomposed, matrix effects are usually minimal for ethanol and no special steps are required.

Recently, automated headspace GL chromatographs became available. Although they are expensive, the microprocessor control ensures precise thermostating and timing and also automatic unattended injection and analysis of a large number of samples.

## Infrared Spectrometry

Infrared (IR) technology is widely used for determining alcohol levels in exhaled breath. Molecules of ethanol absorb IR energy at several wavelengths such as  $3.4 \mu\text{m}$  corresponding to the stretching frequency of C–H bonds, and  $9.5 \mu\text{m}$  corresponding to the stretching frequency of C–O bonds. Such specific absorption of IR energy allows identification and quantification of ethanol in breath. Normal human breath consists of nitrogen, carbon dioxide, oxygen, vaporized water, and trace amounts of organic volatile substances. Acetone is an endogenous compound, which is considered to be a potential interfering

substance because it absorbs IR energy at  $3.4 \mu\text{m}$ . Elevated levels of acetone may exist in the breath of diabetics and fasting individuals. Acetaldehyde, the first metabolite of ethanol, and industrial solvents such as toluene and xylene also absorb IR energy at  $3.4 \mu\text{m}$ . These compounds have been suggested as potential interfering substances when IR breath testing devices are used to test drunken drivers. However, specificity can be enhanced by measuring IR energy absorption at multiple wavelengths. Intoxilyzer 1400 and 5000 (CMI, Inc.), which appear on the Conforming Products List of the National Highway Traffic Safety Administration (NHTSA), are the most widely employed evidential breath testing devices using IR technology. These devices are equipped with three filters. A  $3.8 \mu\text{m}$  filter is used as a reference wavelength,  $3.47$  or  $3.48$  for alcohol quantification, and  $3.38$ ,  $3.39$ , or  $3.40$  for detection of interfering substances and correction for acetone. Two optional filters ( $3.36$  and  $5.52 \mu\text{m}$ ) are available for enhancing the capability of detecting interfering substances.

## Electrochemical Oxidation/Fuel Cell

Fuel cell technology for detecting breath alcohol was introduced in the 1970s. The fuel cell consists of two platinum electrodes separated by an ion-conducting porous substrate. Alcohol in an exhaled breath sample is converted in the fuel cell to acetic acid, producing two electrons for each ethanol molecule. These electrons produce an electrical current that is directly proportional to the amount of ethanol in the sample. Acetic acid further reacts to form oxygen, carbon dioxide, and water. Since this reaction is slower, acetic acid accumulates in the fuel cell. Thus the time necessary to return to a zero baseline is prolonged after several consecutive analyses of breath samples positive for ethanol. Because the fuel cell can potentially react with other alcohols such as methanol and isopropanol, and with acetaldehyde, it is primarily used for screening tests. Alcolmeter SD-2 (Lion Laboratories, Ltd.), Breathalyzer 7410 (Dräger Safety, Inc.), and RBT IV (Intoximeters, Inc.) are popular fuel cell breath testing devices approved by NHTSA. These devices are suitable for on-site testing because they are handheld, portable, and battery-operated devices. Intox EC/IR (Intoximeters, Inc.), a fuel cell device coupled with an IR detector, is used in evidential breath testing. In this device, a breath sample is first analyzed for ethanol nondestructively by the IR detector and then by the fuel cell. Since the fuel cell is unreactive to acetone, potential acetone interference can be monitored.

## Solid-State Semiconductor Gas Sensor

Breath alcohol screening devices incorporating solid-state semiconductor gas sensors employ Taguchi cells. Taguchi gas sensors are also commonly used in breath alcohol ignition interlock devices. A Taguchi cell uses so-called n-type semiconductor which consists of tin oxide ( $\text{SnO}_2$ ) bead mounted in a ceramic cylinder; n-type means 'negative-type,' a term used to describe an abundance of electrons that carry a negative charge. The bead has a very high surface resistance in fresh air. When exposed to alcohol, the surface conductivity increases in proportion to the concentration of alcohol. Alert J4X.ec (Alcohol Countermeasure System Corp.) is among officially approved Taguchi cell breath testing devices. Intoximeter 3000 (Intoximeters, Inc.) is a Taguchi cell device coupled with an IR detector. The Taguchi cell is more sensitive than the IR detector to acetone and various other combustible gases, and this difference can be used to monitor interfering substances in breath.

## Other Methods

Aliphatic alcohols are easily separated by liquid chromatography (LC) using a reversed-phase column. Unfortunately, LC does not have adequate sensitivity to alcohols, and its application to analysis of ethanol in body fluids has been hindered. Davis *et al.* developed a liquid chromatographic method coupled with an enzymatic method that generated NADH in a precolumn ADH-catalyzed oxidation of ethanol. Since this method is not selective for ethanol, it has only limited value for determining ethanol in body fluids.

Gas chromatography-mass spectrometry is selective and sensitive for measuring ethanol and other alcohols. As a result of its cost and maintenance requirements, it has not been used for routine testing of ethanol in biological fluids. Headspace GLC with FID has sufficient specific, accurate, and easy to detect ethanol in body fluids for both forensic and clinical purposes. In addition, selective ion monitoring (SIM) is also useful for determining trace amounts of endogenous aliphatic alcohols such as ethanol and n-propanol in body fluids.

## Interpretation of Analytical Results

Since the effects of ethanol on a person's mental and physical condition are correlated with its blood concentrations, intoxication is based on blood ethanol concentrations (Table 3). However, some

**Table 3** Blood ethanol concentrations and clinical signs/symptoms

Blood alcohol (mg ml <sup>-1</sup> )	Signs/symptoms
0.5 or less	Subclinical. Slight physiological changes and euphoria
0.5–1.0	Slightly increased blood pressure, increased respiration and heart rate, sociability, talkativeness, self-confidence, and impaired reaction responses, attention and judgment
1.0–1.5	Apparently intoxicated, decreased inhibitions, excitement, emotional instability, and impaired sensory-motor coordination
1.5–2.5	Slurred speech, vomiting, apathy, and impaired sensory-motor activities
2.5–3.5	Inertia, loss of motor functions, impaired consciousness, lethargy, and inability to stand or walk
3.5–4.5	Unconsciousness, abolished reflex, subnormal temperature, and impaired circulation and respiration. Possible death
4.5 or more	Cardiac insufficiency and depressed brain stem function. Death from respiratory arrest

populations are very sensitive to alcohol. For example, some Orientals (Japanese, Chinese, Koreans, Thais, Filipinos, and Malaysians) have a poor ability to metabolize acetaldehyde, which is the first metabolite of ethanol. As a result, acetaldehyde accumulates in their blood, even after they consume a small amount of alcohol, and they experience an intense flushing of their face that can spread to their neck and upper arms. They can also experience throbbing headache, dizziness, accelerated pulse, tachycardia, and difficulties in breathing. Two types of aldehyde dehydrogenase (ALDH), ALDH-1 and ALDH-2, are responsible for metabolizing acetaldehyde. ALDH-2 isozyme is located mainly in the mitochondrial fraction, and it has a low  $K_m$  ( $\sim 1 \text{ mol l}^{-1}$ ), which ensures that a very low concentration of acetaldehyde circulate in the blood. On the other hand, ALDH-1 isozyme, which is located mainly in the cytosol fraction, has a higher  $K_m$  ( $50\text{--}100 \text{ mol l}^{-1}$ ), and it does not effectively metabolize acetaldehyde until higher concentrations are achieved. The ALDH-2 gene exhibits polymorphism:  $ALDH-2^1/ALDH-2^1$  (normal enzyme activity);  $ALDH-2^1/ALDH-2^2$  (extremely low enzyme activity); and  $ALDH-2^2/ALDH-2^2$  (no enzyme activity). Approximately 40% of Japanese people have extremely low ALDH-2 activity and 5–10% completely lack enzyme activity. Therefore, intoxication in this population should be based on analytical results of blood ethanol and their genetic backgrounds.

There are variations in ethanol concentration in blood. For example, ethanol distributes according to water content and serum or plasma contains 12–20% more water than whole blood. Consequently, the serum or plasma concentration of ethanol is 1.1–1.2 times higher than whole blood. Many researchers have tried to correlate blood ethanol concentrations with levels in other body fluids. For example, the ethanol levels in the cerebrospinal fluid, vitreous humor, and pericardial fluid after equilibrium were 1.1–1.3 times higher than levels found in blood. In addition, ethanol concentrations in bile were very similar to blood levels. Urine and blood are routine specimens collected in both living and dead persons. Generally, blood concentrations of ethanol are higher during the absorptive phase and urine levels are higher during the postabsorptive period. For example, urine to blood ethanol ratio is  $\sim 1.3$  during the postabsorptive phase.

As is stated in the 'Sample collection and handling' section of this article, several postmortem factors may affect forensic ethanol testing in dead persons. Special care must be taken when evaluating effects of ethanol from its levels in postmortem fluid samples.

See also: **Enzymes:** Overview. **Forensic Sciences:** Overview; **Blood Analysis.** **Gas Chromatography:** **Liquid Chromatography:** Overview.

## Further Reading

- Cary PL, Whitter PD, and Johnson CA (1984) Abbott radiative energy attenuation method for quantifying ethanol evaluated and compared with gas-liquid chromatography and the Du Pont aca. *Clinical Chemistry* 30: 1867–1870.
- Dubowski KM (2002) Analysis of ethanol in saliva. In: Jenkins AJ and Goldberger BA (eds.) *On-Site Drug Testing*, pp. 77–93. Totowa: Humana Press.
- Garriott JC (ed.) (2003) *Medical-Legal Aspects of Alcohol*, 4th edn. Tucson: Lawyers & Judges Publishing.
- Jain NC and Cravey RH (1972) Analysis of alcohol: I. A review of chemical and infrared methods. *Journal of Chromatographic Science* 10: 257–262.
- Jain NC and Cravey RH (1972) Analysis of alcohol: II. A review of gas chromatographic methods. *Journal of Chromatographic Science* 10: 263–267.
- Johnes AW and Schuberth J (1989) Computer-aided head-space gas chromatography applied to blood-alcohol analysis: Importance of online process control. *Journal of Forensic Sciences* 34: 1116–1127.
- Levine B (1999) Alcohol. In: Levine B (ed.) *Principles of Forensic Toxicology*, pp. 170–184. Washington, DC: AACC Press.
- O'Neal CL and Poklis A (1996) Postmortem production of ethanol and factors that influence interpretation. A critical review. *The American Journal of Forensic Medicine and Pathology* 17: 8–20.
- Redetzki HM and Dees WL (1976) Comparison of four kits for enzymatic determination of ethanol in blood. *Clinical Chemistry* 22: 83–86.

## Arson Residues

M Baron, University of Lincoln, Lincoln, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by H C Lee and J Hubball, pp. 1643–1649, © 1995, Elsevier Ltd.

## Introduction

Arson continues to be a serious problem on a global scale resulting in high costs in terms of property loss, injuries, and fatalities. It can be defined as the deliberate and malicious burning of property having three main elements. First, there has been a burning of property. Second, the burning is incendiary in origin, and finally, the burning was started with the intent of destroying the property. The motives for fire-setting behavior have been classified by The National Center for the Analysis of Violent Crime as vandalism,

excitement, revenge, crime concealment, profit, and extremism.

To prove that the crime of arson has been committed, the cause of the fire has to be determined. Arson is a unique crime in that evidence at the scene can be destroyed; however, a systematic investigation may yield sufficient evidence to determine the cause. It is therefore important that every fire scene is treated as a potential arson crime until proof of natural or accidental cause has been established.

## Nature of Fire Scene Evidence

There are many types of physical evidence present at fire scenes. In general, the physical evidence can be divided into four categories:

1. *Transient evidence.* This type of evidence is temporary or short-lived. The most commonly

There are variations in ethanol concentration in blood. For example, ethanol distributes according to water content and serum or plasma contains 12–20% more water than whole blood. Consequently, the serum or plasma concentration of ethanol is 1.1–1.2 times higher than whole blood. Many researchers have tried to correlate blood ethanol concentrations with levels in other body fluids. For example, the ethanol levels in the cerebrospinal fluid, vitreous humor, and pericardial fluid after equilibrium were 1.1–1.3 times higher than levels found in blood. In addition, ethanol concentrations in bile were very similar to blood levels. Urine and blood are routine specimens collected in both living and dead persons. Generally, blood concentrations of ethanol are higher during the absorptive phase and urine levels are higher during the postabsorptive period. For example, urine to blood ethanol ratio is  $\sim 1.3$  during the postabsorptive phase.

As is stated in the 'Sample collection and handling' section of this article, several postmortem factors may affect forensic ethanol testing in dead persons. Special care must be taken when evaluating effects of ethanol from its levels in postmortem fluid samples.

See also: **Enzymes:** Overview. **Forensic Sciences:** Overview; **Blood Analysis.** **Gas Chromatography:** **Liquid Chromatography:** Overview.

## Further Reading

- Cary PL, Whitter PD, and Johnson CA (1984) Abbott radiative energy attenuation method for quantifying ethanol evaluated and compared with gas-liquid chromatography and the Du Pont aca. *Clinical Chemistry* 30: 1867–1870.
- Dubowski KM (2002) Analysis of ethanol in saliva. In: Jenkins AJ and Goldberger BA (eds.) *On-Site Drug Testing*, pp. 77–93. Totowa: Humana Press.
- Garriott JC (ed.) (2003) *Medical-Legal Aspects of Alcohol*, 4th edn. Tucson: Lawyers & Judges Publishing.
- Jain NC and Cravey RH (1972) Analysis of alcohol: I. A review of chemical and infrared methods. *Journal of Chromatographic Science* 10: 257–262.
- Jain NC and Cravey RH (1972) Analysis of alcohol: II. A review of gas chromatographic methods. *Journal of Chromatographic Science* 10: 263–267.
- Johnes AW and Schuberth J (1989) Computer-aided head-space gas chromatography applied to blood-alcohol analysis: Importance of online process control. *Journal of Forensic Sciences* 34: 1116–1127.
- Levine B (1999) Alcohol. In: Levine B (ed.) *Principles of Forensic Toxicology*, pp. 170–184. Washington, DC: AACC Press.
- O'Neal CL and Poklis A (1996) Postmortem production of ethanol and factors that influence interpretation. A critical review. *The American Journal of Forensic Medicine and Pathology* 17: 8–20.
- Redetzki HM and Dees WL (1976) Comparison of four kits for enzymatic determination of ethanol in blood. *Clinical Chemistry* 22: 83–86.

## Arson Residues

M Baron, University of Lincoln, Lincoln, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by H C Lee and J Hubball, pp. 1643–1649, © 1995, Elsevier Ltd.

## Introduction

Arson continues to be a serious problem on a global scale resulting in high costs in terms of property loss, injuries, and fatalities. It can be defined as the deliberate and malicious burning of property having three main elements. First, there has been a burning of property. Second, the burning is incendiary in origin, and finally, the burning was started with the intent of destroying the property. The motives for fire-setting behavior have been classified by The National Center for the Analysis of Violent Crime as vandalism,

excitement, revenge, crime concealment, profit, and extremism.

To prove that the crime of arson has been committed, the cause of the fire has to be determined. Arson is a unique crime in that evidence at the scene can be destroyed; however, a systematic investigation may yield sufficient evidence to determine the cause. It is therefore important that every fire scene is treated as a potential arson crime until proof of natural or accidental cause has been established.

## Nature of Fire Scene Evidence

There are many types of physical evidence present at fire scenes. In general, the physical evidence can be divided into four categories:

1. *Transient evidence.* This type of evidence is temporary or short-lived. The most commonly

encountered examples of transient evidence at a fire scene include gasoline or other petroleum distillate odors, chemical odor, color of flame, and the temperature of the fire. This type of evidence is created by the type of combustible materials in the fire, availability of accelerants, and actions of the arsonist (if applicable). Transient evidence at a fire scene is determined by smell and vision.

2. *Pattern evidence.* This type of evidence is produced by a physical contact between the fire and objects or persons. Common burn patterns found at a fire scene include: V pattern (the pattern at the origin(s) of the fire), multiple origin burn patterns, low burn patterns, and the depth and direction of charring. Other patterns that can be observed at the fire scene are glass fractures, concrete spalling, and the presence of 'trailers' (fire promoters from one point to another). It is also worth noting that if a compartment fire goes to flashover then burn patterns can be difficult to interpret.

3. *Conditional evidence.* This type of evidence is produced by an event or action and includes: direction of the smoke stains, the speed at which the fires spread, the amount, type, and degree of melted material, the condition of the electrical system and fuse box, the condition of the fire alarm and/or smoke detectors, and the location of the safe, important documents, and valuables. This type of evidence often can be used to determine the cause of the fire and/or be used to reconstruct the fire.

4. *Transfer evidence.* This is the classical type of physical evidence and is produced by physical contact between persons or objects, or between persons and objects. Transfer evidence may be classified according to the nature of the matter being transferred, such as biological evidence (e.g., blood, tissue, hair, bone, and body fluids), physical impressions (e.g., fingerprints, footprints, toolmarks, handwriting, identifications, and imprints), and chemical evidence (e.g., ignitable liquid, gunshot or explosive residue, drugs, soil, glass, fibers, plastics, metals, or chemicals). Transfer evidence requires additional laboratory analysis to identify the material and to trace its origin. The main objective of the examination of transfer evidence is to provide a link between victim, suspect, and the crime scene.

In theory, all four types of physical evidence can be found at a fire scene. In practice, however, much or all of the physical evidence may be destroyed by the fire itself or by the efforts to suppress the fire.

## Chemical Evidence at the Fire Scene

An accelerant can be defined as an agent used to initiate a fire or increase the rate of growth

or spread of fire. The detection and identification of an accelerant at a fire scene demonstrates that the crime of arson has been committed. Accelerants can be grouped according to their physical state:

1. Gases: e.g., propane, butane, natural gas.
2. Liquids: e.g., gasoline, paint thinners, kerosene.
3. Solids: e.g., gunpowders, flares, high explosives, flashpowder.

The most commonly detected arson means is the pouring of an ignitable liquid with subsequent ignition by a naked flame such as a match. Even though the ignitable liquid is vaporized and consumed in the fire, its involvement can be established by the detection and identification of ignitable liquid residues (ILRs) in fire debris. The most popular ignitable liquids used as accelerants are gasoline and the petroleum distillates. However, there is a wide range of commercial products that are petroleum derived and therefore potential accelerants. The American Society for Testing and Materials (ASTM) has divided ignitable liquids into eight major classes as shown in **Table 1**. Each class (except for gasoline) is further divided into light ( $C_4$ – $C_9$ ), medium ( $C_8$ – $C_{13}$ ), and heavy ( $C_8$ – $C_{20+}$ ) as determined by the carbon number of their constituent compounds. Each class of liquid is characterized by the types of compounds that constitute the major and minor components.

## Properties of Ignitable Liquids

Ignitable liquids do not burn. The important properties of ignitable liquids relate to the generation of a flammable vapor mixture above the surface of the liquid. These are flashpoint, boiling point, and vapor pressure. The flashpoint is the lowest temperature at which the vapors above the liquid form an ignitable mixture. When a liquid is below its flashpoint, ignition of the vapors does not occur. The vapor pressure of a liquid is the pressure exerted by the vapors above a liquid at a given temperature. When the vapor pressure reaches atmospheric pressure then the liquid is at its normal boiling temperature. Above this temperature the liquid will fully convert to vapor. Other important properties are ignition temperature and lower/upper flammable limits (LFLs/UFLs) of the vapor (**Table 2**). Ignition temperature is the temperature needed to initiate the combustion reaction of the vapor/oxidant mixture. For ignition to occur the vapor/oxidant mixture needs to be within the concentration range of the

**Table 1** Ignitable liquid classification scheme (adapted from ASTM E1618)

<i>ASTM class</i>	<i>Examples</i>	<i>Characteristic compounds</i>
Gasoline		Abundant: aromatics Present: alkanes, cycloalkanes, PNAs <sup>a</sup>
Petroleum distillates	Cigarette lighter fluids Paint thinners Kerosene	Abundant: alkanes Present: cycloalkanes, aromatics, PNAs <sup>a</sup>
Isoparaffinic products	Aviation gas Charcoal starters Specialty solvents	Abundant: branched alkanes
Aromatic products	Xylenes, toluene-based products Specialty cleaning solvents Insecticide vehicles	Abundant: aromatics
Naphthenic paraffinic products	Cyclohexane-based solvents/products Lamp oils	Abundant: branched alkanes, cycloalkanes
n-Alkane products	Solvents, pentane, hexane, heptane Copier toners Candle oils	Abundant: alkanes
Dearomatized distillates	Camping fuels Paint thinners Odorless kerosene	Abundant: alkanes Present: cycloalkanes
Oxygenated solvents	Alcohols, ketones Lacquer thinners	Composition may vary
Others – miscellaneous	Single-component products Turpentine products	

<sup>a</sup>PNAs are polynuclear aromatics.

**Table 2** Properties of common ignitable liquids

	<i>Flashpoint (°C)</i>	<i>Boiling range (°C)</i>	<i>Ignition temperature (°C)</i>	<i>LFL (%)</i>	<i>UFL (%)</i>
Gasoline (100 octane)	– 38	34–184	456	1.4	7.6
Naphtha	– 29	162	134–210	1.1	6.0
Paint thinner	36	134	220	0.8	5.0
Turpentine	33	134	253	0.8	5.0
Kerosene	38	136–270	210	0.7	5.0

lower and upper flammable limits, which is known as the flammable range.

as an accelerant may arise if indicators such as multiple points of origin are established or if unusual burn patterns are observed.

## Recognition of Arson Evidence

Fire investigation is carried out to determine the origin and cause of the fire. The fire investigator uses evidence at the scene to try and establish the point at which the fire started, the first fuel ignited, and the source of ignition. Most fuel available in a structural fire consists of pyrolyzable solids. For this type of fuel to be available for combustion, it must first be pyrolyzed, with the greatest degree of pyrolysis occurring where the heat is most intense. By examining structural damage and tracing the burn pattern, the point of origin can be determined. When investigating a fire, suspicion of the use of an ignitable liquid

## Location of Fire Scene Evidence

Since liquids tend to seek the lowest location, the area in proximity to any heavily charred areas must be searched for an ILR. Other likely areas are those that are shielded from the fire and include: floor cracks, under door thresholds, beneath floor coverings, and areas between baseboard moldings and walls.

The investigator samples debris most likely to contain an ILR in sufficient quantity to enable detection and identification. A range of field detection methods are available to assist the investigator in locating fire debris for laboratory analysis.



Olfactory detection is the oldest method used for detecting an ILR at a fire scene. Experienced investigators can successfully detect an ILR; however, the human sense of smell can fatigue and can be overwhelmed by the inhalation of certain vapors. There are also health and safety concerns in inhaling potentially toxic gases and vapors. Specially trained accelerant dogs can detect sub-part per million levels of ignitable liquids against a clean background. Use of a dog reduces the arduous task of searching the entire fire scene by pin pointing areas where an accelerant may have been used. The results of the onsite canine detection may be useful as an investigative aid, but are nonspecific and may not be admissible in court unless the canine findings are confirmed by laboratory testing.

A variety of portable instruments commonly referred to as 'sniffers' are used. They tend to be nonspecific detectors based on the principles of flame ionization and photoionization. More recently portable gas chromatographs and gas chromatographs coupled with mass spectrometers have become available for use at the fire scene. These provide more information about the sample than sniffers but require more training in their use and interpretation of data.

## Collection of Fire Scene Evidence

Any item or debris suspected of containing accelerant residue should be collected in a tightly sealed, clean, and previously unused paint can, a special evidence bag, or a clean glass jar to prevent further evaporation. The container should be of size similar to the amount of debris but should be no more than three-fourths full, leaving sufficient headspace volume for subsequent laboratory analysis. Care should be taken to avoid contamination as this will compromise the value of the evidence. The container should be labeled with the investigator's name as well as the date, department, location, area from which the sample was taken, case number, and description of the item or contents. The chain of custody must be maintained if the evidence is to be used in court.

There are many factors that contribute to the presence of fire debris that has retained an ILR; however, many of these are beyond the control of the fire investigator. The investigator can only sample from what is available at the scene but needs to consider the properties of the matrix in retaining liquid residues. Porous or absorbent materials such as paper, books, carpeting, and fabric can retain a substantial amount of ignitable liquid and should be collected. If liquid is found in the debris, it should be transferred to a small glass vial using a pipette or

absorbed into an absorbent material and placed into a suitable debris container. It may be possible to compare ignitable liquid samples in the possession of a suspect with unevaporated samples obtained at a fire scene.

## Recovery of Ignitable Liquids from Fire Debris

Chemical analysis of fire debris submitted to the laboratory is usually intended to establish the presence of ILRs. The analysis procedure consists of three separate steps: (1) sample preparation – extraction; (2) analysis of the extracted volatiles to yield a chromatographic profile; and (3) data interpretation – identification of the ILR profile by comparison with known products.

Several groups including the Federal Bureau of Alcohol, Tobacco, and Firearms and ASTM have established procedures for the laboratory analysis of the fire debris. All the procedures involve similar strategies for extraction, analysis, interpretation, and classification of ILRs. Of these, the process of recovering the ILR from the fire debris is the most crucial step in the analysis. Traditional methods for recovering ILRs from debris were based on distillation techniques (steam and vacuum) and solvent extraction; however, these have largely been replaced by headspace sampling techniques that provide both sampling and concentration of the ILR vapors and are capable of detecting sub-microliter levels of ILRs.

Headspace methods work by extracting and concentrating volatiles from the debris on an adsorbent material for subsequent analysis. The methods can be divided into passive or dynamic headspace depending on how the headspace is presented to the adsorbent. Each has its own advantages and disadvantages; however, all headspace methods are limited in that they discriminate against high-boiling residues in the matrix. This can be improved by heating the debris and most methods suggest a temperature range with an upper temperature limit set so as to avoid disproportionate recovery of higher molecular mass compounds and thermal degradation of the matrix. Each of these methods will be considered in more detail.

### Passive Headspace – Activated Charcoal Sampling

In this procedure, volatiles from the fire debris are adsorbed onto a suitable substrate such as a carbon-impregnated polymer commonly referred to as the carbon strip method. The process is determined by diffusion of volatiles to the surface of the adsorbent; hence, it is considered to be a passive or static

headspace technique. The container and adsorbent are stored at an elevated temperature of 50–80°C for 2–24 h with the temperature and the duration being determined by the sample. The adsorbent is then eluted with a minimal amount of a solvent such as carbon disulfide, n-pentane, or diethyl ether and the extract is directly analyzed by gas chromatography (GC). This method is now the most widely used.

### **Passive Headspace – Solid-Phase Microextraction**

Solid-phase microextraction (SPME) is a static headspace method similar to the carbon strip method; however, it does not require a solvent desorption stage. Volatiles are extracted from the headspace by absorption into an absorbent polymer such as polydimethylsiloxane (ASTM method E2154). The absorbent polymer is coated onto a quartz fiber that is housed within a needle similar to a syringe needle. The coated fiber is exposed beyond the tip of the needle in the headspace above the fire debris. As with the carbon strip method, the fiber debris sample can be heated to increase the concentration of volatiles in the headspace. Volatiles are absorbed within the polymer with exposure times for routine screening being within the range 5–15 min. The fiber is retracted within the needle and can then be directly inserted into the injector of a gas chromatograph where the volatiles are thermally desorbed from the polymer onto the column. SPME fibers can be reused but appropriate blanks need to be run to ensure that the fiber is clean.

### **Dynamic Headspace**

Headspace from the evidence container is forced into a tube containing an adsorbent such as activated charcoal or a porous polymer such as Tenax. The flow of headspace through the tube can be achieved by either increasing the headspace pressure with inert gas (positive pressure) or by the application of vacuum to the other end of the adsorbent tube (negative pressure). After the extraction process is completed, the container is removed from the heat source and the tube is removed from the holder. This is a potentially destructive method and samples subjected to this treatment may not be suitable for re-analysis. When charcoal is used as the adsorbent, the trapped organics can be eluted by trickling solvent through the tube into a small vial. The eluent may be analyzed directly or evaporated to a smaller volume before analysis. For Tenax, desorption is best achieved using a thermal desorption system that is directly interfaced to a gas chromatograph. To improve the chromatography a two-stage process is employed, in which the volatiles are thermally

desorbed from the Tenax and then collected in a cold trap. The condensate is then rapidly evaporated onto the column.

### **Direct (heated) Headspace**

The simplest headspace procedure does not use an adsorbent material but involves the direct removal and analysis of a volume of headspace (0.1–3 ml) from the evidence container using a gas-tight syringe fitted with a hypodermic needle. The success of this method is dependent on the concentration of volatiles in the headspace as there is a limit to the volume of headspace that can be directly injected into a gas chromatogram. Heating the debris sample can increase the concentration of volatiles sampled; however, the syringe needs to be heated to avoid condensation problems. Although lacking in sensitivity, this technique is often applied as a screening method because of its speed and ease of use.

### **Solvent Extraction Method**

This method can be applied when the adsorptive nature of the matrix may be problematic in extracting a representative sample. The fire debris is placed in a flask and extracted with an organic solvent. The volatile hydrocarbons that are miscible with the solvent are extracted from the debris. Commonly used solvents are carbon disulfide, dichloromethane, pentane, and hexane. The extract is filtered, evaporated to a small volume, and then analyzed. Sensitivity is similar to steam distillation; however, interfering substances such as pyrolysis products are also extracted.

## **Identification of Ignitable Liquids**

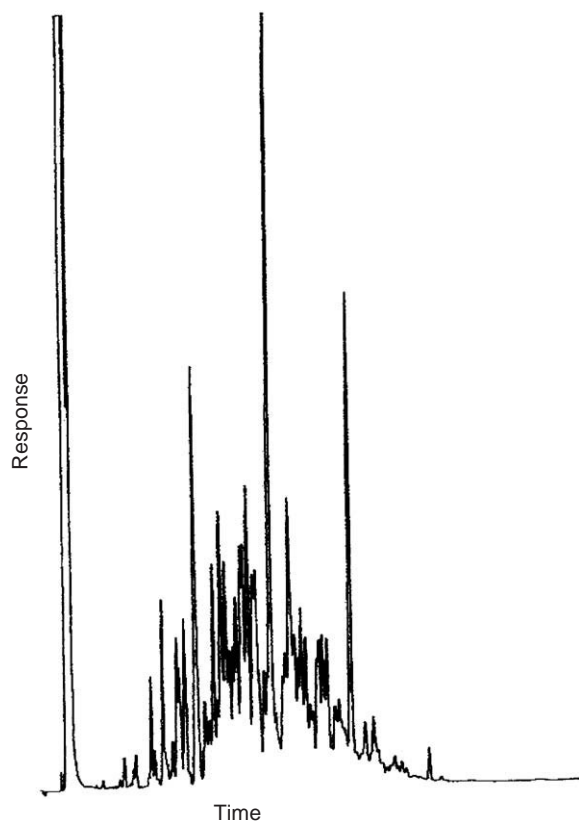
The most commonly used techniques in forensic science laboratories to detect and identify ignitable liquids are gas chromatography–flame ionization detection (GC–FID) and gas chromatography–mass spectrometry (GC–MS). GC–FID is ideally suited for ignitable liquid analysis and has been the method of choice since the early 1960s. It is still widely used as a screening method and can be used for identification; however, the latter is now more commonly carried out by GC–MS. Modern capillary GC columns used with temperature programming can provide good resolution of components in short analysis times. Silicone-based stationary phases provide ‘boiling point’ separations for the aliphatic and aromatic hydrocarbons found in petroleum-based ignitable liquids. However, columns based on chemical interactions can separate aromatic components as a distinct group of peaks from aliphatic

components and polar substances from nonpolar substances.

The flame ionization detector is a universal detector and is capable of detecting all the hydrocarbon components of an ignitable liquid. Specific detectors such as an electron capture detector or nitrogen/phosphorus detector can be used to identify trace substances that may be unique to a specific sample.

During the chromatographic separation the sample is separated into its components, and each peak in the resulting chromatogram represents at least one of the components in the mixture. Each class of ignitable liquids is characterized by a specific chromatographic pattern. An unknown ignitable liquid can be identified by comparison of its chromatographic profile with those in a library of known substances obtained under identical analytical conditions.

Figure 1 shows the chromatogram of a medium product range petroleum distillate recovered from fire debris. This is characteristic of a petroleum distillate in that the predominant peaks are associated with a homologous series of *n*-alkanes in a Gaussian distribution with abundant but less significant isoparaffinic, cycloparaffinic, and aromatic compounds also present. These less significant



**Figure 1** Chromatographic profile for a medium product range petroleum distillate.

components can provide a fingerprint for matching chromatograms. Comparison with a test mixture of *n*-alkanes places the ignitable liquid in the medium product range subclass. Comparison with chromatograms of reference ignitable liquids gave a match with a known mineral spirits (paint thinner) product. Such a result indicates that the fire debris contained a petroleum distillate in the medium boiling range, i.e.,  $C_8$ – $C_{13}$  hydrocarbons. Similar criteria should be applied for assigning an ILR to any of the other ignitable liquid classes.

Such a case as shown by Figure 1 is easy to deal with as the analysis has given a chromatogram that has a characteristic pattern of regularly spaced peaks that are well resolved and distinguishable from the background and interfering compounds. Volatiles from fire debris may be present due to an ILR, the composition of the substrate, and pyrolysis products from the action of heat on the substrate. As the FID does not discriminate between these substances, chromatograms can be difficult to interpret. GC-MS provides the chemical information that can be used to distinguish between ignitable liquid components and background interference. High-resolution capillary GC columns give a more complete separation of petroleum distillates that provides more certain mass spectra on which to base component identifications. Pattern recognition may be carried out using total ion chromatograms; however, this approach may suffer the same problems as GC-FID analysis when dealing with the detection of ILR in the presence of interfering components. This can be overcome by extracted ion profiling as petroleum-based ignitable liquids produce several series of gas-phase ions that are characteristic for the major classes of compounds present. The characteristic molecular and fragment ions for each type of compound found in ignitable liquids are listed in Table 3.

Chromatograms for specific ions indicate the possible presence of a specific class of compound, e.g., the

**Table 3** Characteristic ions for the major types of compounds found in ignitable liquids

Compound type	<i>m/z</i>
Alkanes	43, 57, 71, 85
Cycloalkanes and alkenes	55, 69, 82, 83
Alkylbenzenes	91, 105, 119, 133 (molecular ions 78, 92, 106, etc.)
Indanes	117, 118, 131, 132
Alkyl-naphthalenes	128, 142, 156, 170
Alkylstyrenes	104, 118, 132, 146
Alkylanthracenes	178, 192, 206
Alkylbiphenyls/acenaphthenes	154, 168, 182, 196
Monoterpenes	93, 136

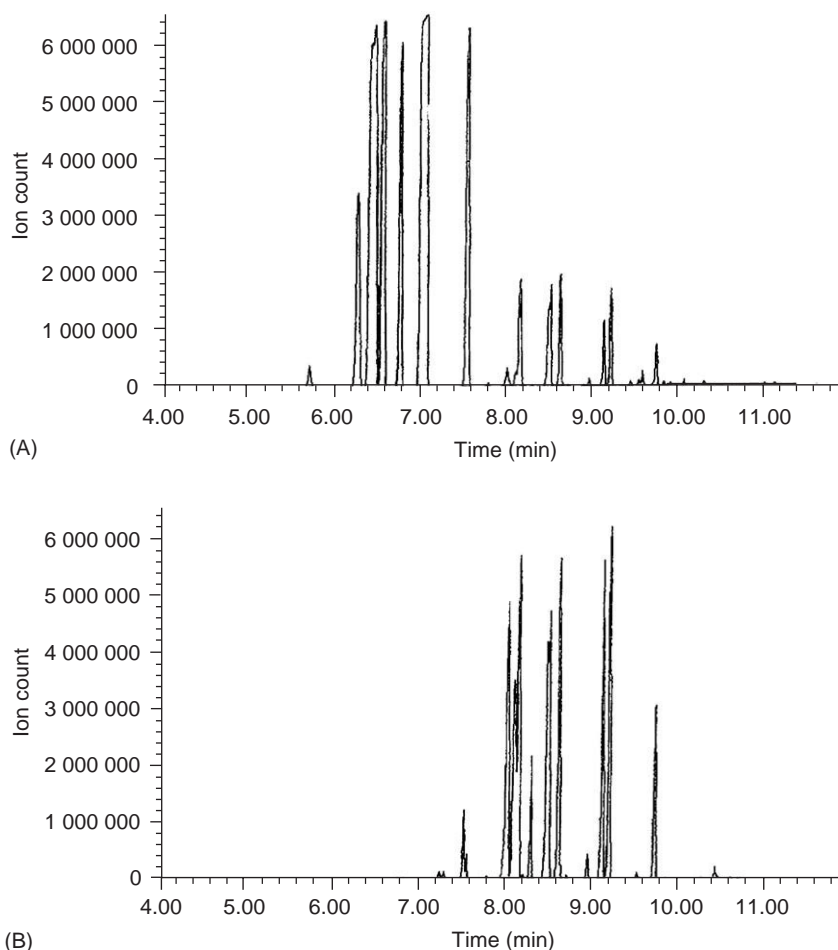
alkanes using the ion  $m/z$  57. This also helps in assigning an ILR to ignitable liquid classes. For example, it is possible to distinguish between petroleum distillates, naphthenic-paraffinic, isoparaffinic products, and dearomatized distillates by comparing extracted ion chromatograms of the ions 55, 57, 83, and 105 (Table 4). Major peaks should be identified from the associated mass spectra. Figure 2 shows how extracted

ion chromatograms can be used to profile trimethyl and tetramethyl benzenes in a gasoline. Extracted ion profiles for unknowns are compared to extracted ion profiles obtained from reference ignitable liquids.

Target compound analysis uses specific compounds that are known to be present in an ignitable liquid. Peakarea ratios can be compared against standards to ensure that chromatographic patterns

**Table 4** Use of extracted ions to differentiate between classes of ignitable liquids

Class	$m/z$ 55	$m/z$ 57	$m/z$ 83	$m/z$ 105
Petroleum distillates	Present – lower abundance than 57	Abundant – Gaussian distribution	Present – lower abundance than 57	Present – lower abundance than 57
Isoparaffinic products	Absent	Abundant – n-alkanes absent	Absent or lower amounts than alkanes	Absent
Naphthenic products	Abundant	Abundant – n-alkanes absent	Abundant	Absent
Dearomatized products	Present	Abundant – n-alkane pattern	Present	Absent



**Figure 2** Extracted ion chromatograms showing the characteristic profiles for (A) the trimethylbenzenes ( $m/z$  120) and (B) the tetramethylbenzenes ( $m/z$  134) found in gasoline.

**Table 5** Target compounds for gasoline and medium petroleum distillate

<i>Gasoline</i>	<i>Medium petroleum distillate</i>
1. 1,3,5-Trimethylbenzene	1. Nonane
2. 1,2,4-Trimethylbenzene	2. Propylcyclohexane
3. 1,2,3-Trimethylbenzene	3. 1,3,5-Trimethylbenzene
4. Indane	4. 1,2,4-Trimethylbenzene
5. 1,2,4,5-Tetramethylbenzene	5. Decane
6. 1,2,3,5-Tetramethylbenzene	6. 1,2,3-Trimethylbenzene
7. 5-Methylindane	7. n-Butylcyclohexane
8. 4-Methylindane	8. <i>trans</i> -Decalin
9. Dodecane	9. Undecane
10. 4,7-Dimethylindane	10. 1,2,3,5-Tetramethylbenzene
11. 2-Methylnaphthalene	11. n-Pentylcyclohexane
12. 1-Methylnaphthalene	12. Dodecane
13. Ethylnaphthalenes (mixed)	13. n-Hexylcyclohexane
14. 1,3-Dimethylnaphthalene	
15. 2,3-Dimethylnaphthalene	

match. In addition, visual pattern matching can be made easier by producing target compound chromatograms (TCCs) that are graphical representations of peak areas for the target compounds. Table 5 lists the target compounds used for the identification of gasoline and medium petroleum distillate. While useful, TCCs should not be used alone.

## Interpretation of Results

The type of physical evidence at an arson scene depends on the method used by the arsonist to light the fire. Generally, it falls into three categories: ignition devices; ignitable liquids; and other associated physical evidence such as toolmarks, fingerprints, footprints, blood, hair, and fibers. Examination of ignition devices is carried out mainly for the purpose of identification. Identification of the type of device and the manufacturer may help in providing useful investigative leads. Macroscopic and microscopic examinations are the most common techniques used to examine an ignition device.

The methods for examination of the different types of associated evidence are varied and are covered in other forensic sections of this encyclopedia.

In establishing the presence of an ILR at the origin of a fire scene careful interpretation of the chromatographic data is required. Consideration needs to be given to the chromatographic distortions that can result from the exposure of the ignitable liquid to the fire scene. Components can be lost from the profile at either end of the pattern and there can also be selective loss of components from within the pattern. Loss of the more volatile components is expected due to combustion and exposure of the liquid to heat. Loss of the less volatile components may be due to incomplete recovery from the debris whereas the loss

of specific components may be due to microbiological degradation.

There is also the difficulty of interpreting patterns in the presence of interfering components that can be present in the substrate or that are generated by pyrolysis of these materials as mentioned in the previous section. These components may mask the presence of an ignitable liquid or suggest the presence of an ignitable liquid as many of these substances are common to ignitable liquids and some can chromatograph as a homologous series of peaks and be mistaken for a petroleum distillate.

It is therefore important to have libraries of chromatograms of ignitable liquids in different stages of evaporation and to also have them for substrates and pyrolysis products. Comparison samples obtained from the fire scene can be used to obtain chromatograms of the substrate background and pyrolysis products. Pyrolysis products that are not identified as target compounds can be excluded from target compound chromatograms. When these products appear on the TCC, it may not be possible to identify the ILR if the pattern becomes overwhelmed.

When an ILR has been identified as belonging to a particular IL class it is appropriate to provide examples of commercial products that belong to this class. It may be possible to match the ILR to a particular example within the class. If the analysis is unable to identify the presence of an ILR it does not necessarily rule out that the fire was incendiary in nature. The analyst should also be aware that any ILR found should not be classified as an accelerant as there may be perfectly legitimate reasons for its presence in the debris submitted for analysis.

**See also:** Forensic Sciences: Overview; Carbon Monoxide and Cyanide from Fire and Accident; Volatile Substances.

## Further Reading

- 2002 *Annual Book of ASTM Standards*, 2002: 14.02, ASTM, Philadelphia, PA.
- Bertsch W and Ren Q (2000) *The Chemical Analysis of Fire Debris for Potential Accelerants. Handbook of Analytical Separations*, vol. 2, ch. 18. Amsterdam: Elsevier.
- DeHaan J (2002) *Kirk's Fire Investigation*. New Jersey: Prentice Hall.
- Lennard C (2001) Fire cause and fire debris analysis. In: Tontarski RE, Jr. (ed.) *Proceedings of the 13th Interpol International Forensic Science Symposium*, 16–19 October. US Department of Justice.
- Newman R, Gilbert M, and Lothridge K (1998) *GC–MS Guide to Ignitable Liquids*. Boca Raton, FL: CRC Press.
- NFPA 921: *Guide for Fire and Explosion Investigations*, 2001 edition. Quincy, MA: NFPA.

## Blood Analysis

**C Brandt-Casadevall and N Dimo-Simonin,**  
Université de Lausanne, Lausanne, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

In this article we deal with the interesting subject of blood analysis in forensic sciences. For historical reasons we give importance to blood group detection. Indeed, even though presently paternity testing and especially blood stains analyses are performed through DNA profile determination, blood groups were for decades the only way of making identifications.

Blood is a biological liquid that represents between 5.5% and 8.0% of the body weight of an individual. It is composed of 60% liquid, or plasma, with the remaining 40% made up of the following cellular elements: erythrocytes, or red blood cells; leukocytes, or white blood cells; and thrombocytes, or platelets.

The plasma contains ~85% water, the remaining 15% being made up of electrolytes, proteins, glucides, and lipids. The plasma proteins are divided into three groups: albumin, fibrinogen, and globulins (three types of globulins are found,  $\alpha$  and  $\beta$ , whose function is essentially the transport of substances, and  $\gamma$ , which principally constitutes the antibodies).

The red blood cells are unnucleated and have a principal function of transport and maintenance of the functional state of the hemoglobin, the respiratory pigment that transports the oxygen and part of the carbonic gas. Hemoglobin is a protein formed from four peptide chains, each one supporting a molecule of hem (a combination of a porphyrin and an atom of iron). In an adult, a major hemoglobin (Hb A) is found in a proportion of 97%, formed from two  $\alpha$ -chains and two  $\beta$ -chains, and a minor hemoglobin (Hb A<sub>2</sub>) represents 3%, formed from two  $\alpha$ -chains and two  $\delta$ -chains. Blood from a newborn (neonate) contains a high percentage (50% at 2 months of age) of fetal hemoglobin (Hb F), composed of two  $\alpha$ -chains and two  $\gamma$ -chains. Certain pathological hemoglobins (S, C, E, and D) probably result from selective mutations that affect the amino acid sequence. These hemoglobin variants are found in very specific subgroups of the population: for example, hemoglobin S is only found in black populations coming from the west of north tropical Africa (sickle cell anemia). Certain illnesses, such as the

thalassemias, will affect the biosynthesis of the hemoglobin chains by mutation of the regulatory genes and do not, therefore, lead to the formation of abnormal chains but to a change in the proportion of the hemoglobins Hb A, Hb A<sub>2</sub>, and Hb F.

The white blood cells are the mobile units of the reticuloendothelial system. There are three different types of white blood cells: granulocytes (which, according to their affinity for stains, are classified as neutrophils, eosinophils, or basophils), lymphocytes, and monocytes.

The platelets are unnucleated cells that play an important role in blood coagulation.

### Blood Preservation and Aging Effects

As a biological material, blood deteriorates rapidly. For analytical purposes, samples must be taken under the most favorable conditions (sterility, cleanliness, container). It must be submitted to immediate treatment in view of the examinations to be conducted; for example, an anticoagulant must be added for certain analyses (blood groups). Some anticoagulants also serve as preservatives (ethylenediaminetetraacetic acid, potassium oxalate, sodium fluoride, etc.) by fixing certain doubly charged cations, inhibiting the action of the complement and other enzymes as well as countering bacterial proliferation. In order to avoid the rapid denaturation of certain substances (e.g., extremely labile proteins), the blood sample must be kept cold: freezing is essential for the proteins and the isoenzymes, while storage at 4°C is necessary to avoid hemolysis in the case of the detection of erythrocyte groups.

When it is a question of bloodstains, the sampling technique will be as important as the storage conditions as far as later examination is concerned. Contamination and high humidity conditions that promote maceration must be absolutely avoided (transport in plastic containers is excluded). A dried bloodstain must be stored frozen for blood group typing. Certain blood characteristics will be preserved if kept cold, even over many years (ABO, Gm), while others will disappear after a much shorter delay.

During preservation, even under optimum conditions, certain enzymes change their electrophoretic profile. It is possible that this results from interactions between the sulfhydryl groups (SH) of some cysteine residues and the glutathione accumulated over the storage period. This effect may be



eliminated by the addition, before analysis, of reductants such as mercaptoethanol or dithiothreitol.

Presently, techniques exist that store bloodstains for the purpose of DNA analysis in a simple and efficient manner. Thus, the FTA Gene Guard system is a method for collecting and storing dried blood samples for DNA testing ensuring a long-term stability of genomic DNA (>5 years) at room temperature. The FTA's filter paper matrix is impregnated with a chaotropic agent that denatures infectious agents and prevents DNA degradation. The blood cells are lysed, and the DNA is linked to the paper matrix. DNA extraction is achieved through successive washing of a paper punch, which removes contaminants and inhibitors.

# Analysis of Blood Components

## Blood Groups

Some blood constituents display, apart from their particular function, the characteristic of being genetically polymorphic, thus permitting the differentiation of individuals. Such is the case with blood groups (ABO, Rhesus, MNSs, Lewis, Kell, Duffy, etc.), the erythrocyte isoenzymes (phosphoglucosaminase (PGM), acid phosphatase (EAP), esterase D, adenylate kinase, adenosine deaminase, glutamate-pyruvate transaminase, glyoxalase, 6-phosphogluconate dehydrogenase, etc.), and the plasma proteins (group-specific component (GC), haptoglobin (Hp), transferrin C,  $\alpha$ -1-antitrypsin, complement component C3, properdin factor B, plasminogen, immunoglobulin groups (Gm and Km), etc.). Meriting a special mention is the group of human leukocyte antigens (HLA), described by Dausset, that exhibits an exceptionally large polymorphism and a remarkable balance in the population.

Some of these groups have the characteristics of being immunogenic, that is to say they provoke an immune response (antibody synthesis, cellular reaction). This explains their medical importance in transfusions and transplantations. But apart from that, the blood groups, by their genetically determined polymorphism, permitted for many decades, on the one hand, paternity testing to be carried out and, on the other hand, the inclusion or exclusion of a person in a criminal case.

As blood groups remain constant throughout life, they provide information that can aid in personal identification. By comparing phenotypes from a child with those of a presumed father, it is possible to determine if there is an exclusion. For this, it is necessary to know the hereditary characteristics and particularities of each group. In the search for

**Table 1** Different types of paternity exclusions

Blood group	Phenotypes		
	Mother	Child	Putative father
EAP	B	BC	B
Group-specific component (Gc)	2-1S	2	1S-1F
PGM	a1	a3-a1	a4-a1
Hp	2-1	1 <sup>a</sup>	2 <sup>b</sup>

<sup>a</sup>Possible genotypes: 1-1 or 1-0 (0 = silent allele).

<sup>b</sup>Possible genotypes: 2-2 or 2-0 (0 = silent allele).

exclusions, it is necessary to take into consideration two fundamental rules.

Rule 1: 'A characteristic found for the child must also be found either with the mother or with the father'. This rule is strict. **Table 1** shows three examples of exclusion according to this rule – for erythrocyte EAP it is essential to know the mother type to exclude the putative father. On the other hand, for the GC and PGM it is possible to exclude the putative father even if the mother phenotype is not known.

Rule 2. 'A child who is homozygotic for a given system cannot have a father who, for the same system, is inverse homozygotic'. This is a strict rule only in the case where the genotype is readily known as it must be possible to exclude the presence of a silent gene (a gene without phenotypic expression). In **Table 1**, we can see one example of exclusion (Hp) according to this rule.

Knowledge of the frequency of occurrence of different phenotypes in each group of a population permits the calculation of paternity probabilities in the case of nonexclusion.

Blood groups are not only found in the blood. For example, ABO antigens may be found in nearly all tissues of the organism as well as in all the secretions (semen, vaginal secretions, sweat, saliva, urine, etc.) for people having the secretor gene (~80% of the population). The same is true for many isoenzymes and certain proteins, which are found in practically all cells of the body. The A, B, and Rhesus antigens appear in the red blood cells of the fetus during the second month of gestation, but in contrast to the Rhesus antigens, the A/B antigens are incompletely developed at birth.

## Blood Group Detection

The erythrocytes groups are detected by means of immunological reactions, of which the most common is the agglutination of antibodies specific to each antigen. The erythrocyte antigens are situated on the membrane of the red blood cells. Depending on the

antibody used for detection (immunoglobulin M (IM) and immunoglobulin G (IgG)) as well as the localization of the antigen on the membrane of the erythrocyte, the agglutination may occur in a sample saline medium at ambient temperature (ABO, Lewis) or require a macromolecular medium at 37°C (Rhesus) or a medium of weak ionic concentration in the presence of an antiglobulin (Duffy by de Coombs test in low ionic strength solution).

The erythrocyte isoenzymes and the plasma proteins are determined by electrophoretic or isoelectrofocalization techniques. These techniques are performed on a large variety of supports (agarose gel, polyacrylamide, acetate sheets). Electrophoresis is a method of separation in an electric field formed using a continuous current with specific conditions for each protein: buffer composition, migration time, voltage, and temperature. Isoelectric focusing is an electrophoretic separation based on migration in a stable linear pH gradient created with the aid of ampholytes. The proteins will migrate and be focused at their pI (isoelectric point). This technique gives better resolution and sensitivity than the former. Detection of protein bands is performed either through specific enzymatic reactions to give fluorescent or chromogenic products or by direct staining with specific protein stains (Figure 1).

The description of Jeffreys *et al.* of the polymorphism that occurs in DNA repetitive base-pair sequences ('DNA fingerprinting') has permitted more informative data to be obtained for the identification of individuals and for paternity testing.

Since the early 1990s, DNA typing for variable number of tandem repeats (VNTR) loci, detected using the Southern method, started being used in complement with blood group analysis in paternity

testing and also in criminal cases. Some years later DNA amplification techniques (polymerase chain reaction (PCR)) for typing of genetic loci also within satellite DNA VNTR and especially STR (short tandem repeats)) stopped relying on the identification of blood groups.

### Hemoglobin

Determination of the hemoglobin variants may give important information in forensic medicine, given that their biosynthesis is genetically determined. Their different electrophoretic mobilities permit easy separation and the data obtained may be applied in cases of disputed paternity and genetic population studies. The detection and quantification of the different hemoglobins may provide information for the detection of certain hemoglobinopathies, which may be interesting in an identification.

### Electrolytes

The quantification of electrolytes, notably through flame photometric (atomic emission) techniques, may have some importance in forensic medicine. Indeed, different authors have reported the possibility of determining the time of death by the increase in concentration of some electrolytes and the decrease in certain others.

### Glucose

The quantification of glucose in the blood, using enzymatic techniques (glucose oxidase) or by ultraviolet (UV)-visible absorption spectrophotometry, may provide some clarification as to the cause of suspicious death (for example, hyper- or hypoglycemic comas). However, the results are only valid if the time between the death and the analysis is relatively short.

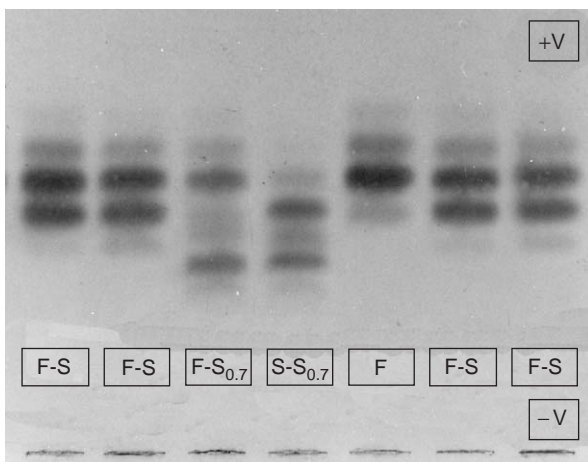
## Analysis for Exogenic Substances

This involves the detection of foreign substances in the blood and is a major field in forensic science, forensic toxicology. It includes analysis for different medicinal toxins, narcotics, etc. The determination of blood alcohol concentrations takes on particular importance due to road traffic regulations and the social repercussions of alcohol intoxication.

## Bloodstains

Blood is the one of the most important biological traces and also one of the most frequently studied in forensic cases.

The macroscopic aspect of bloodstains may provide interesting information and is largely



**Figure 1** Properdin factor-B phenotyping performed using thin-layer high-voltage agarose gel electrophoresis and subsequent immunofixation with specific antiserum.

dependent on the age of the stain and the nature of the support on which it is found. Indeed, according to the form and contour characteristics of a bloodstain, it is sometimes possible to determine the trajectory (angle and distance) of the blood droplet that formed the stains. It is also very important to observe accurately and make a full documentation of the bloodstain pattern before sampling. This will ensure that the results obtained from analysis in the forensic laboratory will answer the questions of the criminal enquiry. Three questions will be posed: (1) It is blood? (2) Is the blood human? (3) Who does the blood belong to? The following sequence of tests is performed in the forensic laboratory to identify bloodstains.

### Presumptive Tests

These tests are extremely sensitive but have the disadvantage of not being entirely specific for blood. Such tests are based on the detection of the pseudoperoxidase activity of hemoglobin. The decomposition of a peroxide liberates oxygen, which oxidizes a reduced leukobase, provoking a color change.

Tetramethylbenzidine and *ortho*-tolidine are the most sensitive tests (capable of detecting blood in dilutions of  $1:10^6$  and  $1:10^5$ , respectively), while phenolphthalein and leucomalachite green are less sensitive but more selective.

### Confirmation Tests

These tests serve to detect an entity specific for blood, either cellular components or hemoglobin. The methods employed may be based on microscopic, microcrystallographic, spectroscopic, chromatographic, immunochemical, or immunochromatographic techniques.

Microscopic techniques aim at detecting blood cells, which are identified by their appearance. For this to be possible, the material must be fresh and well conserved. Among the microscopic techniques are those involving a direct search on the material, with the advantage of being nondestructive, while other methods require preparation of the sample. The latter demand a reconstitution and staining of the cellular elements.

Microcrystallographic or microchemical techniques rely on the fact that certain derivatives of hemoglobin have the tendency to crystallize. It is possible to find either crystals of hemein or crystals of hemochromogen (depending on whether the heme molecule is degraded to an iron(II) or iron(III) pigment).

Spectroscopic methods aim to detect the UV-visible absorption spectrum of hemoglobin or one of its derivatives.

Thin-layer chromatographic (TLC) techniques use very small quantities of material. A sample is deposited in the form of a spot and submitted to the action of a solvent (eluant), which causes component substances to migrate on the TLC plate. Detection is performed using a redox reaction with tetramethylbenzidine or *ortho*-tolidine. Positive and negative controls are tested on the same plate. The retention factor ( $k$ ) is used to identify haemoglobin in the sample.

Immunochemical tests (HemeSelect™ commercial kit) use an antihemoglobin antibody in a reverse passive hemagglutination procedure to detect the presence of human hemoglobin with a high degree of sensitivity.

Currently, the most frequently used tests in forensic analysis of bloodstains are immunochromatographic tests (Hexagon OBTI-Test, commercial kit). Initially, these tests were developed to detect the presence of human hemoglobin in stool samples. The kit consists of a testing device and a sample tube with Tris buffer, pH 7.5, with a mild detergent. Monoclonal antihuman hemoglobin (hHB) antibodies conjugated to blue dye particles are able to bind hemoglobin. The resulting antigen-antibody complexes migrate to the test zone, where they are captured with an immobilized second antibody directed against hHB, forming a blue test line (T) to indicate a positive result. Unbound reagents migrate further and are bound in a second line by immobilized antimouse IgG antibodies. This control line indicates proper functioning and correct handling of the test. The test is easy and quick to perform, with a high degree of sensitivity.

These two immunological tests detect the presence of human (and primates) hemoglobin in the stains and can also confirm/inform of the human origin of the blood.

### Bloodstain Origin

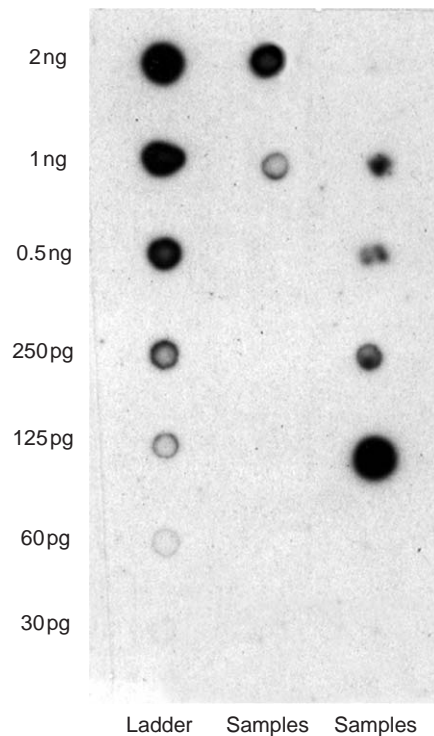
Many methods exist for the determination of the origin, human or otherwise, of a particular bloodstain: by the recognition of cellular elements from their morphology and by detecting different characteristics of human hemoglobin (for example, solubility, electrophoretic profile, antigenicity, sequence, etc.). However, tests based on the detection of plasma proteins are the most often employed. A first group is made up of immunological techniques such as antiprotein or antiglobulin sera (Ouchterlony) or immunoelectrophoresis, which has the advantage of eliminating cross-reactions. The value of these tests is irrefutable when they give positive results. They can, however, give false negative results when there is no

quantitative equivalence between the antigens and the antibody (e.g., an antigen is present in too small a quantity) or when the antigens are denatured (e.g., stains treated with certain detergents or proteolytic enzymes).

The second group aims to detect the presence of globulins in the stains using an agglutination method. This technique is easy to apply, very sensitive, and very selective. It involves the inhibition reaction of the particularity of certain incomplete anti-Rhesus (IgG) antibodies to fix onto red blood cells without agglutination occurring. The test demands an initial contact between the sample to be analyzed and a human antiglobulin serum (Coombs serum). O/Rhesus + red cells sensitized with an antiRh antibody are employed as a detection system. If the material contains human globulins, the sensitized red blood cells are not agglutinated as the antiglobulins have been previously neutralized. An agglutination is therefore synonymous with the absence of human globulins in the analyzed material. The human antiglobulin test is more sensitive than the Ouchterlony test. However, given that the latter not only detects globulins but also all the plasma proteins including albumin, it is capable of detecting human material in traces that have been subjected to extreme degradation (albumin is particularly resistant).

The human origin of a bloodstain can also be determined using the immunochemical test (Heme-Select™ commercial kit) and the immunochromatographic test (Hexagon OBTI-test commercial kit) cited above. The latter is the most frequently used in forensic testing.

The origin of a bloodstain can be ascertained in a dot-blot hybridization test using a primate-specific probe such as D17Z1 (= 40 nucleotides complementary to an alpha satellite DNA sequence located on chromosome 17). The DNA sample is denatured, deposited, and fixed on a positively charged nylon membrane. The membrane is first prehybridized in the presence of sodium dodecyl sulfate (anionic detergent) to saturate fixation sites on the membrane and thus prevent nonspecific hybridization of the probe. The hybridization itself is done with a biotin-labeled probe. The bound probe is detected by chemiluminescence using streptavidine/horseradish peroxidase and luminol. Luminol oxidation causes photon emission, which can be detected on X-ray film. A distinct advantage of this test is that it is semiquantitative: in addition to helping establish the origin of the bloodstain, it can also indicate the amount of DNA present by comparing the strength of the signal with known concentration standards (Figure 2). This test is at present the most used in the forensic laboratories to establish the origin of



**Figure 2** Dot-blot. At left, the ladder with spots corresponding to decreasing DNA concentrations. At right, two lines with different samples to be tested.

bloodstains as well as other kinds of forensic samples (saliva, sperm, etc.).

### Bloodstain Identification

**Erythrocyte groups** Historically, the most important by far in forensic practice has been the ABO group as it can persist in a stain up to several years. The determination of antigens for the ABO group was based on indirect agglutination techniques that can only detect ABO phenotypes. Currently, it is possible to determine the ABO genotype using several DNA-based methods using PCR followed by restriction enzyme digestion, single-strand combination polymorphism, or sequencing analysis.

**DNA analysis** Since Jeffreys's discovery of the polymorphism of certain DNA fragments, this analysis has tended increasingly to replace the analysis of blood groups in stains. This is due to the fact that it has the highest discriminating power and specially the fact that DNA is more resistant to degradation than certain proteins. For the analysis of DNA using the Southern blot method, it is necessary to have high molecular weight DNA in a quantity greater than 50 ng. It is for this reason that DNA analysis by PCR amplification is of particular interest. Techniques



using PCR permit us to establish DNA profiles in very small quantities and in samples that are partly degraded.

## Other Data of Forensic Interest

### Sex Determination

When a sample of suspect blood arrives at the laboratory, it may be useful to determine the sex of the donor (for example, in the case of criminal abortion). Low-performance older techniques (search as for the sex chromatin, Barr's body, Y-chromosome fluorescence, etc) and other methods such as the determination of sex hormone levels (progesterone/testosterone) have been replaced by procedures for detecting sites on the X and Y chromosomes using PCR. At present, the sex determination is carried out at the same time that the genetic profile is determined when using multiplex STR kits.

### Determination of Which Part of the Body the Blood Has Come From

This diagnosis is based on a cytological study of the cellular elements found in the stain. The stain must be relatively fresh in order to detect, through microscopic examination, the desquamation cells typical of different regions of the human body.

### Menstrual Blood

This may be identified from its richness in fibrin degradation products, which may be detected using a hemagglutination inhibition method. It is also possible to identify menstrual blood by an analysis for lactate dehydrogenase. This enzyme shows little phenotypic variation between individuals but shows tissue differences that permit the recognition of menstrual blood.

Several authors have reported the possibility of confirming the presence of menstrual blood in stains by detection of messenger ribonucleic acid specific for epithelial cells using reverse transcriptase-PCR.

### Fetal Blood

It may be important in certain cases involving infanticide or criminal abortion to determine whether the blood is of fetal origin. The concentration of fetal hemoglobin or of  $\alpha$ -fetoprotein may be determined, and good results are generally obtained. Liquid chromatography techniques have been described recently for detecting fetal hemoglobin.

**See also:** Atomic Emission Spectrometry: Flame Photometry. Blood and Plasma. Clinical Analysis: Overview; Sample Handling; Sarcosine, Creatine, and

Creatinine. **Electrophoresis:** Principles; Isoelectric Focusing. **Enzymes:** Enzymes in Physiological Samples; Industrial Products and Processes; Enzyme Assays. **Forensic Sciences:** Alcohol in Body Fluids; DNA Profiling; Systematic Drug Identification; Thin-Layer Chromatography. **Immunoassays Applications:** Forensic. **Microscopy Applications:** Forensic. **Nucleic Acids:** Electrochemical Methods. **Polymerase Chain Reaction. Spectrophotometry:** Overview; Biochemical Applications.

## Further Reading

- Bauer M, Kraus A, and Patzelt D (1999) Detection of epithelial cells in dried blood stains by reverse transcriptase-polymerase chain reaction. *Journal of Forensic Sciences* 44: 1232-1236.
- Boormann KE, Dodd BE, and Lincoln PJ (1988) *Blood Group Serology*. Edinburgh: Churchill Livingstone.
- Budowle B, Smith J, Moretti T, and DiZinno J (2000) *DNA Typing Protocol: Molecular Biology and Forensic Analysis*. Natick, MA: Eaton Publishing, BioTechniques® Books Division.
- Burgoyne LA (1994) *Proceedings of the Eighth International Symposium on Human Identification*. Madison, WI: Promega Corporation.
- Butler JM (2001) *Forensic DNA Typing*. London: Academic Press, Harcourt Science and Technology Company.
- Grunbaum BW (ed.) (1981) *Handbook for Forensic Individualization of Human Blood and Bloodstains*. Göttingen: Library of Congress, Sartorius GmbH.
- Hochmeister MN, Budowle B, Sparkes R, et al. (1994) Validation studies of an immunochromatographic 1-step test for the forensic identification of human blood. *Journal of Forensic Sciences* 44: 597-602.
- Inoue H, Maeno Y, Iwasa M, et al. (1997) Sensitive detection of human globin chains by microbore high-performance liquid chromatography and its forensic applications. *Journal of Chromatography B, Biomedical Sciences & Applications* 688: 221-227.
- Jeffreys A, Wilson V, and Thein SL (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature* 314: 67-73.
- Lee JC-I, Kao LG, and Linacre A (2000) Rapid identification of the ABO genotypes by their SSCP pattern of multiplex PCR products. In: Sensabough GF, Lincoln PJ, and Olaisen B (eds.) *Progress in Forensic Genetics*, vol. 8. Amsterdam: Elsevier.
- Lincoln P (2000) From ABO to DNA. *Medicine, Science and Law* 40: 3-7.
- Waye JS, Presley J, Budowle B, Shutler GG, and Fourney RM (1991) A simple and sensitive method for quantifying human genomic DNA in forensic specimen extracts. *BioTechniques* 7: 852-855.
- Wolson TL (2001) DNA analysis and the interpretation of bloodstains patterns. *Journal of Canadian Society of Forensic Science* 34: 151-157.
- Yamamoto F (2000) Molecular Genetics of ABO. *Vox Sanguinis* 78: 91-103.

## Carbon Monoxide and Cyanide from Fire and Accident

C N Gray, Deakin University, Geelong, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Carbon monoxide and cyanide compounds are found widely in the environment as a result of natural processes and human activity. CO is produced mainly by incomplete combustion of carbon-based fuels. It is a major toxic component of emissions from natural fires, gasoline (petrol) engines, heating plant, explosives, cooking stoves, open fires, barbecues, cigarettes, etc. The main technological source of CO emissions in advanced countries is the automobile. Concentrations in urban air are typically of the order of a few parts per million by volume (ppm), while concentrations in gasoline engine exhausts can be up to ~8% by volume (80 000 ppm; at normal temperature 1 ppm CO is  $\sim 0.9 \text{ mg m}^{-3}$ ). Diesel engines have much lower exhaust concentrations (up to ~0.1%, 1000 ppm). Mainstream cigarette smoke can contain ~5% (50 000 ppm). The CO concentration in the air in dwellings equipped with gas stoves can be several ppm. Human exposure to CO is usually highest among people directly exposed to combustion gases as a result of un-vented space and water heaters, leaking flues, gas stoves, portable electrical generators, automobiles running within closed garages, and direct inhalation of tobacco smoke. There is also toxicologically significant exposure among some working populations, such as coke oven workers, miners, and underground car park attendants.

Inorganic cyanide compounds, such as sodium or potassium cyanide, are released into the environment by a number of industrial processes including mining, electroplating, and metal treatment. Solutions of cyanide salts are used to leach gold, silver, and some other metals from their ores, a practice that can result in large volumes of cyanide-contaminated water and solid waste. Cyanide compounds are used in gold, cadmium, and zinc electroplating. Some metals are hardened by heating them in molten cyanide salts. Cyanide compounds have many other applications including in the manufacture of pigments, in photography and etching, and as pesticides for the control of rabbits, rats, and termites. Hydrogen cyanide is used as a feedstock in the production of a range of chemicals including adiponitrile (used in the manufacture of nylon) and methyl

methacrylate. It is also used as a poison for the control of rodents and other pests; for example, in the fumigation of grain stores. Hydrogen cyanide can also be released during fires involving nitrogen-containing polymers such as polyurethane, nylon, acrylonitrile, and ABS (acrylonitrile-butadiene-styrene) and can be a danger to fire fighters and others exposed to the smoke of such fires. The main source of cyanide in urban air is the internal combustion engines where it is produced by free-radical reactions involving nitrogen from the atmosphere. Coal and some petroleum fuels contain organic nitrogen compounds that can also produce cyanide during combustion, as does tobacco. Significant amounts of cyanide are produced by microorganisms and insects such as millipedes. Cyanide or its precursors (such as cyanogenic glycosides) occur in many plants including cassava, bitter almonds, sorghum, and lima beans.

Human exposure to cyanide can result from inhalation of hydrogen cyanide gas or inhalable dusts of cyanide compounds, the ingestion of cyanide in water or food, or from unintentional hand to mouth transfer in contaminated environments. Away from cities and industry the background air concentration is  $\sim 0.2 \mu\text{g m}^{-3}$ . Concentrations in occupational settings can be several parts per million (as HCN) or several milligrams per cubic meter of air. Typical surface waters uncontaminated by industrial activity can contain cyanide in concentrations up to  $\sim 5 \mu\text{g l}^{-1}$ . Tap water can contain  $\sim 0.5\text{--}0.8 \mu\text{g l}^{-1}$  of cyanogen chloride (produced during chlorination). Wastewaters from mines or industry can contain up to several hundred milligrams per liter. Foodstuffs can contain cyanide or cyanogenic compounds in a wide concentration range: cereals, from  $\sim 1$  to  $500 \mu\text{g per kg}$ ; lima beans, as much as  $3 \text{ g per kg}$ . Cooking tends to reduce or destroy the cyanide content of food.

### Toxic Effects

It is estimated that there are tens of thousands of cases of accidental CO poisoning in the US each year, of which ~600 are fatal, and there are ~2000 cases of suicide utilizing car exhaust emissions. CO owes its toxicity largely to its affinity for the oxygen transport sites in red blood cells thereby interfering with normal oxygen transport to the tissues. CO has about 240 times greater affinity than oxygen for hemoglobin and readily displaces it. About 80% of the CO binds to hemoglobin and 10–15% to



myoglobin in muscles. About 5% binds with the cytochrome enzymes in cells responsible for producing energy from glucose oxidation, thus causing cytotoxic hypoxia. CO can cross the placental barrier and affect the unborn child of an exposed pregnant woman. It is not significantly metabolized and is mainly exhaled unchanged, although ~1% is converted to carbon dioxide. It has an excretion half-life of ~3–4 h for a resting subject breathing air. This can be shortened by the administration of oxygen. CO is also produced endogenously during normal metabolic processes resulting in a background carboxyhemoglobin concentration of up to about 1.2%. CO can also be produced by the metabolism of other substances including dichloromethane (used as a paint stripper). The toxic effects include cardiovascular and neurological damage, the severity of which depends on both the exposure concentration and duration. Concentrations above about 2000 ppm in air are usually fatal after a short time while 1200 ppm is classified as 'immediately dangerous to life or health'. Exposure to about 500 ppm over several hours can result in headache, confusion, and possible collapse; 50 ppm can be tolerated over many hours by healthy individuals with minimal symptoms. Shorter exposure periods cause less effect because it takes some time for the carboxyhemoglobin to equilibrate with the environmental CO concentration. For healthy individuals 200 ppm is generally considered safe over an exposure period of 15 min, 100 ppm for 30 min, and 60 ppm for 60 min. Most countries have set occupational exposure limits of 50 ppm for 8-h daily average exposure as well as short-term exposure limits (refer to local regulations).

Cyanide is recorded as the direct cause of death in only a few tens of cases in the US each year, but it is probably a contributory cause in many other cases such as the 5000–10 000 deaths due to smoke inhalation. It has, of course, been used for legal execution in the US for many years. Cyanide occurs naturally in the blood at concentrations up to about  $10 \mu\text{mol l}^{-1}$  ( $260 \mu\text{g l}^{-1}$  as  $\text{CN}^-$ ). Elevated levels can be cleared from the body quite quickly with a half-life of about 30 min. Cyanide uptake can occur by inhalation of hydrogen cyanide gas or cyanide dust, and by the ingestion of cyanide or cyanogenic compounds. In the acid conditions of the stomach cyanide compounds react to produce hydrogen cyanide, which is rapidly absorbed. Cyanide is transported in the blood stream to all parts of the body where it enters the cells and blocks the cytochrome *c* enzyme responsible for transferring the energy produced by the oxidation of glucose into chemical energy in the form of adenosine triphosphate. This mechanism is similar to

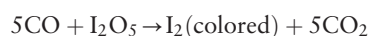
the cytotoxic effect of CO except that oxygen transport to the tissues is not restricted; only its utilization by the cells is impaired. The primary effects of cyanide poisoning are on the nervous system, partly because nerve cells have little energy reserve when oxidative metabolism is blocked. Hydrogen cyanide gas at concentrations above a few hundred parts per million are generally fatal. Death has been reported following exposure to ~200 ppm for 10 min or ~130 ppm for 30 min. The average fatal oral dose in humans is ~100 mg as CN ( $\sim 1.5 \text{ mg per kg}$  of body mass). The lowest fatal human dose is thought to be ~40 mg. Occupational exposure limits in various countries are typically between 2 and 5 ppm for hydrogen cyanide and about  $2\text{--}5 \text{ mg m}^{-3}$  (as  $\text{CN}^-$ ) over an 8-h day. In addition, ceiling limits have been established that should not be exceeded at any time (refer to local regulations).

## Sampling and Analytical Methods

### Carbon Monoxide

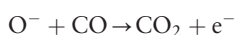
Sampling and analytical methods are available for very high concentrations of CO in exhaust pipes and flues and very low concentrations in the ambient atmosphere. These include colorimetric devices, electrochemical analyzers, infrared spectrometers, and gas chromatographs. A sample may be collected in a suitable container for analysis at a later time in the laboratory, or the analysis may be carried out onsite. Sampling containers can be canisters or either Teflon or Tedlar bags that can be filled using manual or battery-operated pumps. In many parts of the world new houses are now required by law to have smoke detectors and CO detectors are also becoming mandatory in some places. Detectors used for such purposes need only be calibrated at the set point of the alarm system, but must be reliable and must be certified by the relevant authorities.

**Detector tubes** A simple and relatively inexpensive measurement method involves the use of calibrated stain-length detector tubes such as those produced by Draeger, Kitagawa, Gastec, MSA-AUER, and others. CO detector tubes often rely on the reaction of CO with iodine pentoxide, which produces a purple colored iodine stain. The detector tube carries a scale calibrated in parts per million and the length of the color stain that develops after the required number of pump strokes corresponds to the concentration of CO in the sample:



Detector tubes are available for CO covering a measurement range from a few ppm to several thousand ppm. Some are suitable only for short-term sampling whilst others may be operated over a period of hours with the aid of a low-flow battery operated sampling pump or by passive diffusion.

**Electronic and electrochemical detectors** There are several different types of electronic or electrochemical detector for CO of which metal oxide semiconductor (MOS) sensors and electrolytic cells are the most common. In MOS detectors CO is oxidized on a heated tin oxide semiconductor where surface-adsorbed oxygen oxidizes gases such as CO and hydrocarbons producing electrons that enter the conduction band of the semiconductor thereby reducing its resistance:



MOS detectors can be used in alarm systems to trigger a warning when dangerous concentrations of CO (or other reducing gases) are present, or in quantitative instruments. They are not specific to CO and interference can result from either reducing or oxidizing gases. It is possible to make them more selective by using a filter to prevent interfering gases from reaching the detector, or by doping the detector with catalyst metals so that they can be operated at lower temperatures at which some gases do not interfere. They have relatively high power requirements and continuously operating variants are usually connected to mains electricity.

Another type of CO detector uses porous platinum electrodes to catalyze the oxidation of CO at the anode of an electrolytic cell:



Usually the cell is held in a semiporous membrane that allows CO to diffuse through. The cell reactions produce a current that is proportional to the CO concentration in the surrounding air and can be used to give a direct reading of concentration or trigger an alarm at a preset value.

**'Biomimetic' detectors** Another relatively simple device for detecting CO utilizes the reversible reaction of CO with an iron(II) compound held in a porous matrix to produce a color change that can be measured optically. The iron compound may be a synthetic porphyrin, which in a sense mimics the biochemistry of hemoglobin; hence these detectors are

referred to as biomimetic. These devices are most commonly used in dwellings as CO alarms in a smoke/fire alarm system. They are relatively slow to respond to changing CO concentrations and some early models had a tendency to progressively retain their color change and provide false alarms.

**Infrared analyzers** Some fixed point and portable CO analyzers used in industry employ infrared spectrometry or nondispersive infrared (NDIR) photometry. CO strongly absorbs infrared radiation at a wavelength close to  $4.6\ \mu\text{m}$  and can be measured in the presence of some other gases such as water vapor and carbon dioxide. These instruments have a gas cell that allows a beam of radiation to pass through the sample gas either once, or multiple times by reflection from mirrors. A typical portable infrared spectrometer used for gas analysis is the Foxboro MIRAN SapphIRe. This contains a gas cell that by multiple reflections from internal gold-plated mirrors can achieve a single beam pathlength of up to  $12.5\ \text{m}$ . Wavelength selection is by a variable interference filter giving moderate resolution, plus a number of fixed wavelength filters for specific gases including ones centered on  $4.5$  and  $4.7\ \mu\text{m}$ . They are capable of measuring CO down to concentration in the sub-ppm range and give a continuous reading with an averaging time of about  $10\ \text{s}$ . NDIR instruments are similar in many respects to simple single beam gas spectrometers. They may have specific or narrow wavelength detectors and many have narrow band-pass filters to select the wavelength of the beam. Sensitive NDIR instruments can detect sub-ppm concentrations of CO.

**Gas chromatography** For measuring very low background levels in ambient air gas chromatography is more sensitive than most of the preceding methods. This can be automated to give a semicontinuous analysis *in situ*, or samples can be collected in containers, canisters, or inert bags, returned to the laboratory, and transferred to the gas chromatograph by gas-tight syringe. Most of the carbon dioxide present is first removed by alkaline permanganate and residual carbon dioxide and hydrocarbons are removed in a 'stripper column'. CO is then separated from any methane and other gases and vapors present in the sample by passing it through a second column. The eluant from the column is then passed over a catalyst, where CO is reduced to methane, and then passed to a flame ionization detector. An alternative method uses a helium glow discharge ionization detector (DID).

**Calibration** Calibration of analytical instruments for CO is usually achieved using commercially

available calibration gas mixtures. Several of the major suppliers of industrial gases sell certified calibration gases and some such mixtures are also available through the Occupational Safety and Health Administration (OSHA) laboratories in Cincinnati, USA.

**Biological monitoring** It is often useful to be able to measure the CO in the blood of an exposed person, for example, to investigate occupational exposures, to assist in diagnosis and treatment after accidental over-exposure, and in forensic investigations following a suspicious death. If a blood sample can be obtained it is possible to measure the carboxyhemoglobin concentration directly by spectrophotometry. For example, the blood sample is substantially diluted in sodium carbonate solution, with the addition of sodium hydrosulfite solution and sodium hydroxide. The absorbance is then measured at wavelengths of 532 and 558 nm and the carboxyhemoglobin ratio is calculated from these values. There may be some complication in postmortem samples because of the presence of other modified forms of hemoglobin including methemoglobin (iron(III)) and sulfhemoglobin. Carboxyhemoglobin concentration can also be measured using gas chromatography by liberating the CO and collecting it in a headspace device. Where it is not convenient to obtain a blood sample, for example, during routine occupational monitoring, exhaled breath analysis is usually a satisfactory substitute. A breath sample can be collected in a bag or tube using an exhalation valve to collect the end-tidal sample. The sample can later be returned to the laboratory for analysis by one of the techniques described previously. Alternatively, a portable instrument can be used to analyze the exhaled breath directly. It is important to consider the kinetics of uptake and excretion (toxicokinetics) when interpreting the results of biological monitoring; for example, a moderate concentration in blood many hours after exposure could indicate a very high initial concentration.

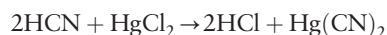
## Cyanide

There is often a need to monitor cyanide in air, water, solid waste, food, and other environmental samples. The cyanide present in these samples may include free (noncomplexed) cyanide such as hydrogen cyanide (or hydrocyanic acid in water solution), cyanogen ( $C_2N_2$ ), cyanogen chloride, cyanide salts, or complexed cyanide such as metal-cyanide complexes of iron, nickel, copper, mercury, silver. Complexed cyanides are less toxic because they are less bioavailable, but they may break down producing free cyanide, for example, as a result of the ultraviolet radiation in daylight. Analytical techniques for free

cyanide may not detect complexed cyanide, but methods are also available for total cyanide, for example, by first distilling the sample with phosphoric acid and, if necessary, ethylenediaminetetraacetic acid to form hydrogen cyanide which is then collected in an alkaline solution.

Direct methods of analysis are available that may be used *in situ*, or samples may be collected for subsequent analysis in the laboratory. Liquid samples may be collected in glass or plastic containers provided that the pH is kept at about 11 or greater in order to prevent the loss of hydrogen cyanide gas. If the pH falls, for example, because of the uptake of carbon dioxide from the atmosphere, significant losses can occur during transport and storage. Air samples can be collected in a liquid-filled bubbler (impinger) containing alkali solution. Alternatively, a filter can be used to collect particulate cyanide compounds while the hydrogen cyanide gas passes through the filter and can be collected in an alkaline bubbler or on a silica gel or soda lime sampling tube; the particulate cyanide is then washed from the filter with an alkaline solution and the trapped gaseous cyanide is washed from the tube in the same way.

**Detector tubes** As with CO, hydrogen cyanide in air can be measured conveniently with reasonable accuracy by stain-length detector tubes. A typical reagent mixture is mercuric chloride and methyl red indicator. The hydrogen cyanide is converted to mercuric cyanide with the production of hydrogen chloride, which produces a pH change and as a result the indicator changes color to red:



The tubes generally have a precleanse layer to remove interfering substances such as acid gases. These tubes are not suitable for particulate cyanide. Detector tubes are available for HCN covering a measurement range from a few ppm to several hundred ppm. Some are suitable only for short-term sampling whilst others may be operated over a period of hours with the aid of a low-flow battery operated sampling pump or by passive diffusion. There are also detector tubes for cyanogen chloride.

**Ion-selective electrodes** Ion-selective electrodes function as a half-cell that in the presence of certain ions produces an electrical potential that can be compared to a reference cell. The selectivity is determined by the nature of the membrane separating the electrode from the test solution. Membranes used in cyanide ion-selective electrodes are usually of the

insoluble inorganic salt type. These electrodes are relatively specific, but interference can result from chloride, bromide, and iodide ions. Metal ions can also be a problem if they can form complexes with cyanide, for example, cadmium, zinc, silver, nickel, copper(I), iron, and mercury. Ion-selective electrodes are relatively cheap and do not require expensive or complex peripheral equipment. They can also be used in potentiometric analysis; the cyanide in solution is titrated against hypochlorite until there is no further change in electrode potential. The hypochlorite forms a complex with the cyanide so that it is no longer available to the detector.

**Spectrophotometry** There are numerous spectrophotometric methods for the determination of cyanide including the pyridine-benzidine method, the pyridine-pyrazole method, and the pyridine-barbituric acid method. As an example, in the latter method the pyridine-barbituric acid coupling agent produces a complex with cyanide that adsorbs principally at a wavelength of 580 nm. This method is frequently used for the measurement of airborne hydrogen cyanide in industrial air. The hydrogen cyanide is collected in an alkaline bubbler, while any particulate cyanide is collected on a prefilter. With a 3 l air sample the range of quantitation is about 0.3–235 ppm ( $\sim 3$ –260 mg m<sup>-3</sup>). In water samples, this method has a lower limit of detection of about 0.02 mg l<sup>-1</sup>.

**Gas chromatography** Gas chromatographic methods for cyanide usually involve using headspace techniques to detect hydrogen cyanide with the use of a nitrogen-phosphorus detector (NPD) or electron capture detector (ECD). Total cyanides can be analyzed in this way after conversion to hydrogen cyanide.

**Flow injection analysis** A standard method of cyanide analysis employs flow injection following ligand exchange and amperometry. The sample is pretreated with ligand exchange reagents in order to release the cyanide from any metalocyanide complexes in which it may be bound. The treated sample is then analyzed by flow injection analysis. Hydrochloric acid converts the cyanide into hydrogen cyanide, which can diffuse through a membrane and dissolve into an alkaline solution. The resultant cyanide ions are measured amperometrically.

**Biological monitoring** As in the case of CO, it is often useful to be able to measure the cyanide in the body tissue of persons suspected of being exposed or poisoned. For cyanide this is possible using tissue, blood, or urine samples. Cyanide is converted in the body to thiocyanate and this may be analyzed as an

alternative to the cyanide itself. Blood or urine samples should be analyzed as soon as possible after collection to minimize biochemical changes that can alter the concentration of cyanide. Bacterial activity, if unchecked, can produce cyanide in the samples during storage. A suitable technique for cyanide in blood analysis is headspace gas chromatography using either an electron capture or nitrogen-phosphorus detector. Spectrophotometric analysis using a suitable reagent, such as pyridine-pyrazolone, is another option. Urine analysis is also possible using spectrophotometry. In this case it is usual to analyze both the cyanide present and its thiocyanate metabolite, for example, by bromination and treatment with pyridine-*p*-phenylene diamine. In postmortem cases tissue samples can be taken as well as stomach contents to analyze for any unabsorbed cyanide or cyanogenic materials.

In fire victims it is not uncommon to find both CO and cyanide. The amounts that are detected can often shed light on the circumstances of the death. The collection and analysis of the samples should be carried out as quickly as possible as the concentration of CO in blood gradually falls while that of cyanide might increase with time.

**See also:** **Flow Injection Analysis:** Principles. **Forensic Sciences:** Overview; Arson Residues. **Gas Chromatography:** Overview. **Ion-Selective Electrodes:** Overview. **Spectrophotometry:** Overview.

## Further Reading

- Agency for Toxic Substances and Disease Registry (ASTDR) (1997) Toxicological Profile for Cyanide. US Department of Health and Human Services, Public Health Service.
- Center for Disease Control (CDC) (2004) The Compressed Mortality Database, Mortality for 1999–2000 with ICD 10 codes, Mortality for 1979–1998 with ICD 9 codes. Center for Disease Control and Prevention, 1600 Clifton Rd, Atlanta, GA 30333, USA.
- NIOSH (1994) *Cyanides, Aerosol and Gas: Method 7904, Manual of Analytical Methods*, 4th edn, issue 2.
- NIOSH (1996) *Carbon Monoxide: Method 6604, Manual of Analytical Methods*, 4th edn, issue 1.
- Russell J, Lewis RJ, Johnson RD, and Canfield DV (2002) An accurate method for the determination of carbon monoxide in postmortem blood using GC/TCD. Office of Aerospace Medicine, Washington, DC 20591, DOT/FAA/AM-02/15.
- Standards Development Branch Ontario Ministry of the Environment (2002) Information Draft on the Development of Air Standards for Hydrogen Cyanide.
- World Health Organization (1999) *Environmental Health Criteria 213: Carbon Monoxide*, 2nd edn. Geneva: Switzerland.



## DNA Profiling

**R Coquoz**, AMS Genetique, Lausanne, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

DNA analysis is now a basic tool in the establishment of links between biological evidence and comparison samples. It has reached such a widespread use and its power is so impressive that it is probably the best-known forensic analytical method in the general public. Because of the need of high-tech instruments and sophisticated analysis kits, standardization has been surprisingly quick to come, favored by the almost monopoly of a couple of providers in this field. As a consequence, forensic DNA analysis is being practiced worldwide using almost identical tools and protocols.

### DNA Polymorphisms

The DNA sequences of any pair of individuals are almost identical. There is, however, a large number of small differences that allow the remarkable variety of mankind. The elements of interindividual variability are also called polymorphisms. Any polymorphism used for forensic identification purposes is called a marker, and the variants at the polymorphic positions are called alleles.

Polymorphisms may come from minor differences at single nucleotide levels. These are called single nucleotide polymorphisms (SNPs) (Figure 1). It is

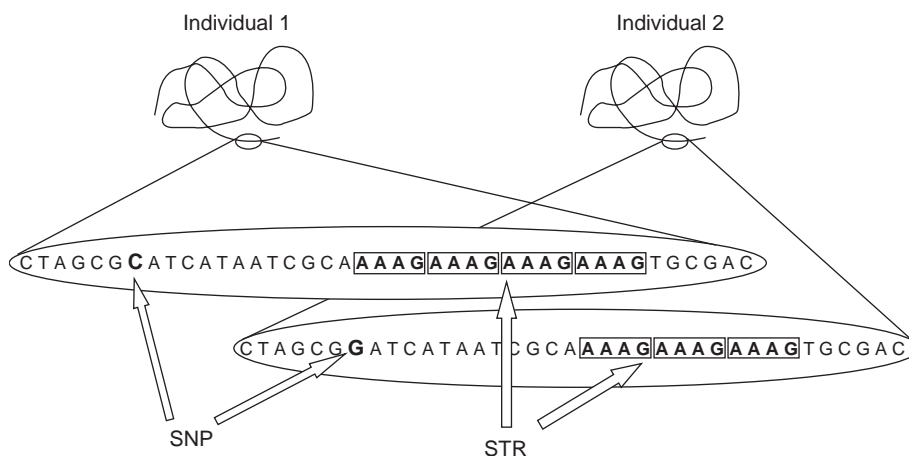
estimated that there is on an average one SNP for every 500 nucleotides. Another source of variability comes from the repetitive sequences. These are made up of a nucleotide sequence (the repeat unit), repeated in tandem like a series of identical beads on a necklace (Figure 1). The repeat unit can be from two to more than 100 nucleotides long. The polymorphism comes from the extreme variation in the number of repeats. The alleles then differ by their size (length polymorphism) and get their name from the number of repeats they possess. When the repeat unit is larger than six-nucleotides long, the repetitive sequence is called a minisatellite or VNTR (variable number tandem repeat). When it is six-nucleotides long or less, the repetitive sequence is called a microsatellite or STR (short tandem repeat).

There is a huge reservoir of thousands of repetitive sequences in the human DNA that can be used as markers. The VNTRs have been the markers of choice for DNA profiling from 1986 to ~1995, through the use of a method called Southern blotting or restriction fragment length polymorphism typing. Nowadays, the DNA-profiling techniques mainly focus on STR markers, and SNP analysis begins to gain some importance.

### DNA-Profiling Procedures

#### Multiplex PCR Analyses of STRs

Virtually all the forensic DNA analyses performed worldwide are multiplex polymerase chain reaction

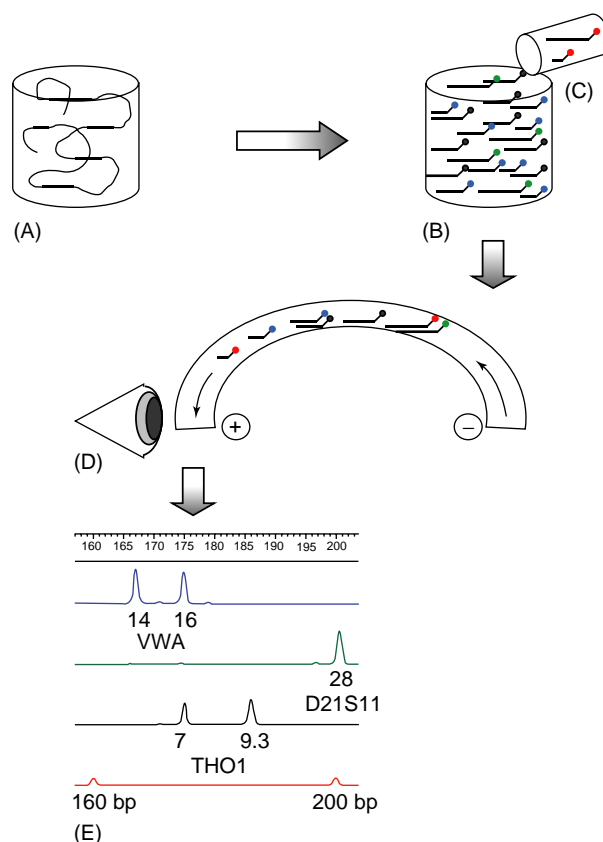


**Figure 1** Schematic presentation of the DNA from two individuals and zoom on a region of their DNA to show that DNA is a linear molecule made of nucleotides, and to illustrate the two main types of polymorphisms used in forensics: the short tandem repeats (STRs) and the single nucleotide polymorphisms (SNPs).

(PCR) analyses of STRs. A basic requirement toward any forensic analysis is a high sensitivity. DNA analysis in forensic science as well as in laboratory medicine has reached a remarkable sensitivity through the use of the so-called PCR. PCR is an ingenious DNA amplification method that cyclically reproduces the natural DNA replication process. Each cycle consists of (1) a denaturation step where the two DNA strands are separated by heating, (2) an annealing step where, after cooling, two short synthetic DNA strands (the primers) are hybridized on both sides of the DNA fragment to be amplified, and (3) an elongation step where a DNA polymerase adds nucleotides to the primers to synthesize the DNA strands complementary to the template DNA. Since the newly synthesized DNA strands can serve as templates for the next PCR cycles, each cycle leads to the doubling of the number of copies of the target sequence. The result of many cycles is millions of copies of the DNA fragment delimited by the primers. The use of a heat-stable polymerase has rendered the whole process very easy and has allowed automation. The nature of PCR makes it naturally extremely sensitive, with a theoretical detection threshold of one molecule! The primers allow the amplification to be highly specific for the desired targets. By using several primer pairs in the same tube, several STRs can be amplified simultaneously; this is called multiplex PCR.

At the end of multiplex PCR amplification, the amplification products of the targeted STR markers must be analyzed by electrophoresis to make visible the nature of the alleles at each marker (Figure 2). Electrophoresis is a separation method for charged molecules, that allows them to get separated because of their varying mobility when they move under the influence of an electric field. DNA fragments are negatively charged molecules and they move toward the anode of the electric field at a speed inversely correlated to their size. The alleles of an STR can thus be separated from each other. The detection of the alleles is rendered simple by using primers labeled with a fluorescent dye. In most laboratories, the electrophoresis is done in capillary electrophoresis instruments that are fully automated. Some are equipped with an array of 100 capillaries, allowing a corresponding number of samples to be analyzed in parallel with a throughput capacity of more than 2000 samples a day.

With an adequate choice of STRs and primers it is possible to analyze simultaneously several STRs displaying alleles in different size ranges. If any overlap of the size ranges is excluded, it is then easy to attribute any allele detected to the correct STR, and to identify its name. By using different fluorescent dye



**Figure 2** A schematic view of multiplex PCR analysis of STRs. In a test tube, some STRs from the sample's DNA (A) are amplified by PCR (B) using fluorescent-labeled primers. After the addition of an internal standard (red-labeled fragments) (C) the DNA fragments obtained at the end of the PCR are separated by capillary electrophoresis according to their size (D) and detected at the end of the capillary. Each peak (E) is then electronically labeled with the name of the corresponding allele. The profiles in blue, green, and black present the alleles detected for the STRs VWA, D21S11, and THO1, respectively. The red profile displays two peaks from the internal standard.

labels, even STRs having alleles in the same size range can be analyzed in multiplex, thus allowing the design of complex multiplex STR analysis kits analyzing simultaneously more than 15 STRs.

What is now called a DNA profile is the result of the analysis of a set of STRs. Two DNA profiles are only comparable if the set of STRs analyzed includes at least a few STRs in common. The key to comparability is standardization. An excellent degree of standardization has been reached worldwide. The US professional organizations have defined a set of 13 STRs that have been considered as obligatory for the US forensic labs. In Europe, there are seven STRs obligatory and they all belong to the set of 13 STRs adopted in USA (Table 1). The companies marketing forensic DNA analysis kits have designed kits to satisfy these convergent needs. As a consequence,



**Table 1** Description of the STRs most frequently analyzed

<i>Name of the STR</i>	<i>Length of the repetitive element</i>	<i>Chromosome location</i>	<i>Allele range</i>	<i>Discrimination power (%)</i>
THO1	4	11	4–13.3	91
VWA	4	12	9–25	94
D21S11	4	21	26–34	95
FGA	4	4	13–51.2	96
D18S51	4	18	8–27	97
D8S1179	4	8	5–19	93
D3S1358	4	3	11–20	92
TPOX	4	2	6–16	80
CSF1PO	4	5	6–15	87
D16S539	4	16	5–15	92
D7S820	4	7	6–15	94
D13S317	4	13	7–15	92
D5S818	4	5	7–16	85
D2S1338	4	2	15–28	97
D19S433	4	19	9–19	92
Penta E	5	15	5–24	97
Penta D	5	21	2.2–17	94

Note: The length of the repeated element is expressed in number of nucleotides and the power of discrimination is calculated for a Caucasian population. The first seven STRs correspond to the minimum European set and the first 13 STRs correspond to the US standard set.

all the kits available on the market offer almost identical sets of STR markers.

A DNA profile appears like the one displayed in **Figure 3**. For each STR, the individual displays a pair of alleles in the form of a pair of peaks. One allele of the pair comes from the mother of the individual and the other allele comes from the father. The two alleles being different, the individual is called heterozygous. A single allele may sometimes be detected for the STRs for which the sample is homozygous, having inherited the same allele from both parents. The DNA profile can be summarized as the list of alleles detected for each STR. The probability that two unrelated individuals taken at random will display different DNA profiles for a set of 15 STRs (the power of discrimination) is quasi-100%. On the contrary, the chance that two unrelated individuals taken at random share a common profile (the random match probability) is usually below  $10^{-15}$ . Such a low number demonstrates how powerful the DNA profiles are. But it reaches the limits of what can be accurately calculated without having to make risky assumptions about the genetic structure of human populations and particularly about the degree of independence of the alleles. Indeed, the random match probability is obtained through the multiplication of the estimated frequencies of the individual alleles of the DNA profile; this multiplication assumes independence of the events, a characteristic that is not completely warranted in the real world. As a consequence, the random match probabilities calculated cannot be defended as correct to the last digit. Only the approximate order of magnitude is correct.

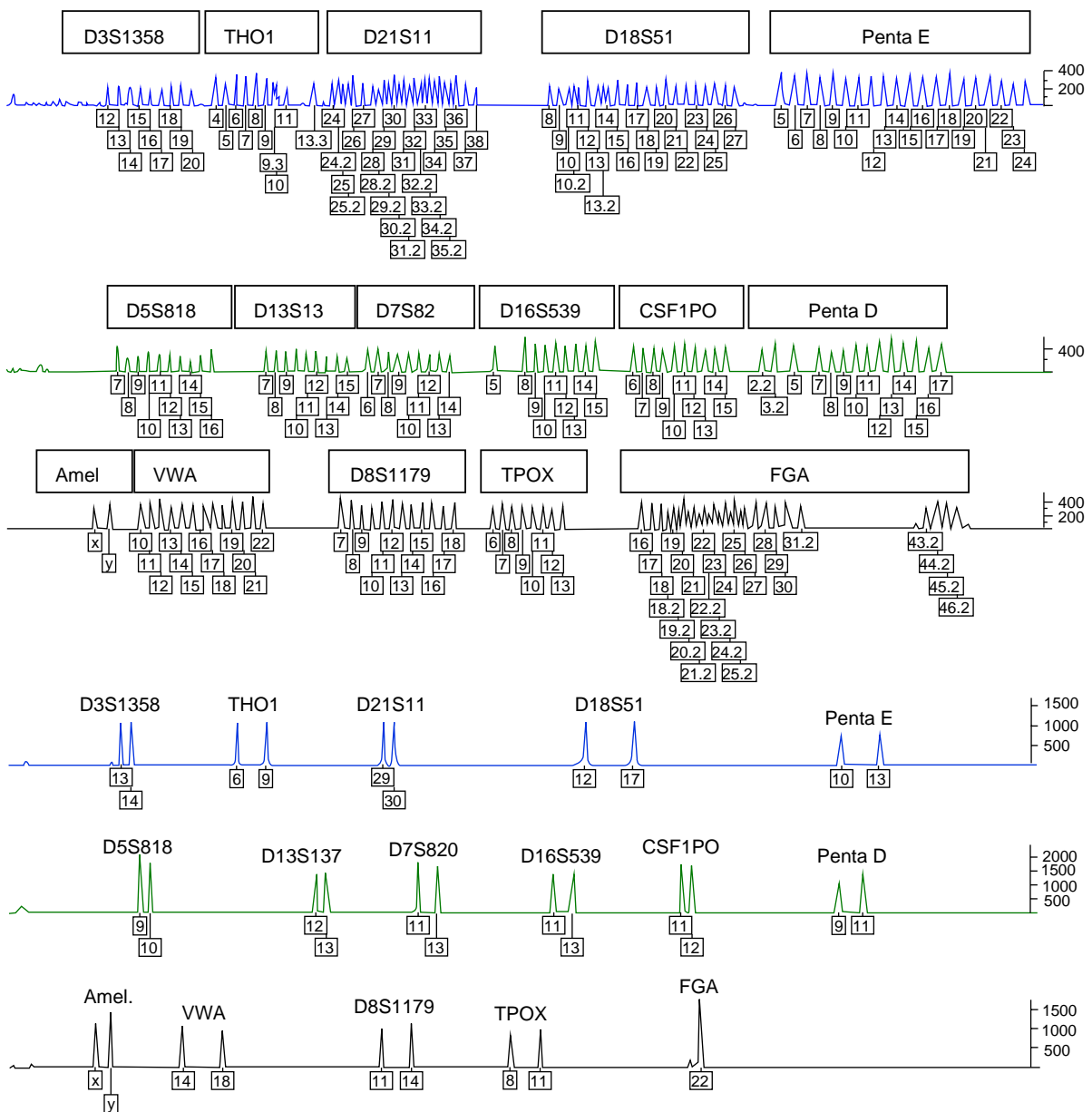
### Ultrasensitive Protocols

By performing enough cycles of PCR amplification, the theoretical sensitivity of any PCR technique is one molecule of target DNA. PCR analysis in forensic science is, however, not adjusted at this level of sensitivity. Indeed, disturbing artifacts dramatically affect the quality of the results when PCR is pushed toward its maximal sensitivity.

Despite these difficulties, ultrasensitive protocols have been designed that use more cycles of PCR amplification to try to retrieve information from even the tiniest traces encountered during the investigations. The key to these protocols is a sophisticated algorithm of interpretation of the results that takes into account the potential artifacts, e.g., the fact that some alleles of the source of the trace may not have been detected (allele drop out) or the fact that some alleles detected may be artifacts (spurious alleles). Such ultrasensitive protocols are also called low copy number (LCN) protocols. The interpretation algorithm relies on a Bayesian evaluation of the two competing hypotheses: the hypothesis of the accusation (the trace comes from the suspect), and the hypothesis of the defense (the trace comes from another unidentified person). This Bayesian estimation forces the laboratory to gather complex experimental data on the risk of allele dropout and the risk of spurious allele appearance with LCN protocols.

### Mitochondrial DNA

Virtually all the DNA of a cell is contained in its nucleus. However, cell compartments called



**Figure 3** Typical example of STR analysis results. The lower part presents the DNA profile of an individual's sample obtained using a multiplex STR analysis kit analyzing 16 markers. The individual analyzed is homozygous for STR FGA and heterozygous for all the other STRs. The upper part presents the DNA profile of a standard containing all the alleles frequently encountered in the population for the set of markers.

mitochondria contain their own small DNA: the mitochondrial DNA or mtDNA, a small 16 569-nucleotides-long DNA. Since mtDNA is present in multiple copies in each mitochondrion and each cell contains many mitochondria, mtDNA is present in thousands of copies per cells in contrast to the only two copies of nuclear DNA contained in the nucleus. For minuscule traces or degraded samples (e.g., hair shafts), mtDNA is thus logically the last chance target when the usual DNA profiling fails. Since

mtDNA does not contain repetitive sequences, the only polymorphisms available are sequence polymorphisms, which are clustered within a 1000-nucleotides long region called the control region. Using a technique called sequencing, the sequence of this control region can be revealed. Since two unrelated individuals display on an average eight nucleotide differences in this region, mtDNA analysis is a useful tool to exclude innocent suspects. Its power of discrimination is, however, limited and it will never be

a means of identification. Moreover, because of its maternal mode of transmission, mtDNA is identical across a whole family tree along the maternal lines of transmission. Although this characteristic is a limitation for the power of discrimination, it allows comparisons to be made between individuals with very distant relationships. This has found some uses in historical or genealogical investigations, or in the identification of remains where close relatives were not available for comparison. A well-known example was the identification of the remains of the family of the Tsar Nicolas II killed in 1917, and identified by comparison with descendants of close relatives.

### Y-STR Analysis

There are STRs on the Y chromosome as well, the so-called Y-STRs. Because of the mode of inheritance of this chromosome, the alleles of the Y-STRs are transmitted unchanged from father to son and further along the paternal lines. Y-STR analysis has thus properties symmetrical to those of the mtDNA. Although its power of discrimination is higher compared to mtDNA, it is anyway much lower than standard multiplex STR analysis. The main usefulness of Y-STR analysis is its ability to provide a DNA profile from mixed stains containing male DNA as a minor component among a larger amount of female DNA. Using standard multiplex STR analysis, the minor male DNA will provide a detectable contribution to the DNA profile only when it represents more than 10% of the total DNA. However, a full Y-STR profile will be obtained even when the stain contains a 1000-fold excess of female DNA. This helps to solve cases where the aggressor is male and the victim female, like sexual assault cases. Like mtDNA analysis, Y-STR analysis has also found some uses in the identification of remains where close relatives were not available for comparison, or in historical or genealogical investigations. A well-known example was the evaluation of the claims that Thomas Jefferson, the third president of the United States, might have fathered the children of his slave Sally Hemings.

### Single Nucleotide Polymorphisms

SNPs have raised tremendous interest in medical research because some SNPs are the basis of the variations in the individuals' sensitivity to diseases or their variable reactions to drugs. Very efficient methods have been designed for high throughput analyses of SNPs. Most of these methods are covered by the name 'DNA chip' analysis or 'DNA microarray' analysis (Figure 4). Indeed, most of the methods allow the typing of large sets of SNPs to be made on

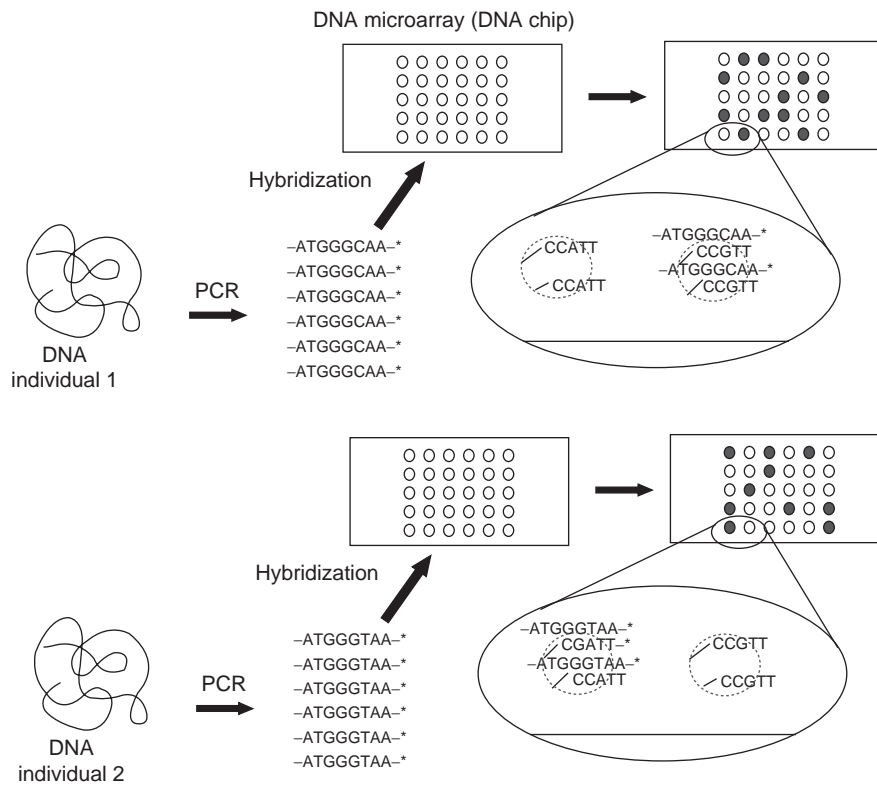
miniaturized devices resembling electronic chips. SNPs might become an interesting alternative to STRs as targets for forensic identification. Most SNPs usually have only two alleles and their power of discrimination is clearly much lower compared to STRs. But this weakness can easily be compensated by the use of a much larger set of SNPs. For a simple stain DNA analysis, a set of 50–100 SNPs is as powerful as a set of 10 STRs. For mixtures, however, the set of SNPs to be used should be much larger. But the technical advances may allow the analysis of thousands of SNPs to be made at a speed and at a price that is a fraction of that required for STR analysis. It is difficult to guess if SNP profiling will be able to replace STR analysis in the near future. However, SNP analysis will certainly be the method to gain morphological or ethnic information on the individual at the origin of a stain. DNA profiling is in essence a comparison process. It is of no use as long as a comparison sample is not available. However, it is clear that some of the morphological aspects of an individual are determined by his DNA. A stain could thus be submitted to a DNA analysis conceived to extract all these morphological informations (hair color, eye color, skin color, etc.). The genetic basis of these morphological polymorphisms are SNPs, whose details will certainly be better understood in the future. Morphology SNP kits will likely be designed to serve to focus the investigations toward the adequate subset of the population.

## Applications

### DNA Profiling in Practice

Any kind of biological material containing DNA may be used. The traditional types of traces can be easily analyzed: blood stains, semen stains, autopsy tissue, bone, etc. But thanks to the sensitivity made possible by the PCR, the largest number of samples submitted to DNA forensic laboratories is now saliva stains and above all contact traces (traces of the contact between skin and other surfaces). This is the main breakthrough of the last few years: the ability to make use of a large set of traces that were not even considered 10 years ago. The contact traces have led to the involvement of DNA profiling in the investigations of any type of crime, even the minor categories like burglary or theft that represent the largest category. As a consequence, the number of samples submitted to the DNA laboratory has exploded.

Although DNA profiling can make use of extremely faint traces, there are limits to its power. Logically, the traces have to contain the DNA of at least one cell. In practice, the requirements are higher. About



**Figure 4** Schematic description of SNP analysis on a DNA chip. The regions of the DNA harboring the SNPs are first amplified by PCR (only one SNP PCR product is displayed in the figure). The PCR is devised to obtain labeled and thus detectable DNA fragments. These are confronted to the set of DNA probes of the DNA chip and allowed to hybridize to their corresponding probes. The details of the hybridization are illustrated by zooming on the region of the DNA chip, which presents a pair of positions displaying the probes detecting two alternative alleles of an SNP.

a hundred cells are necessary to obtain a DNA profile using the standard methods. Since any human cell contains  $\sim 6$  pg of DNA, a trace has to contain  $\sim 0.5$  ng DNA for a successful DNA analysis. The requirements are naturally reduced for ultrasensitive LCN protocols. The DNA concentration is  $\sim 30$  ng DNA  $\mu\text{l}^{-1}$  in blood,  $400$  ng  $\mu\text{l}^{-1}$  in semen,  $2$  ng  $\mu\text{l}^{-1}$  in saliva, and  $0$ – $250$  ng per hair root. About  $0.01$   $\mu\text{l}$  of blood might thus be sufficient to provide a DNA profile. But such a small amount might be extremely hard to detect on a set of clothes and it is evident that some of the main difficulties of the forensic investigation of traces have shifted toward the detection of the minute traces of biological material theoretically exploitable. The availability of sensitive chemical or biochemical tests for blood and semen is sometimes of great help in this respect. For saliva, the available tests are less sensitive. For contact traces, there is no detection method. The sampling of the traces is done simply by assuming where the traces ought to be. A sometimes disturbing consequence is that the DNA profile obtained at the end of the analysis cannot be attributed to a specific material,

and the meaning of the presence of biological material from a suspect on a surface might be widely different if it is skin material or saliva.

An important factor in DNA profiling is the quality of the material. DNA is quite a robust molecule and usable DNA fragments have been recovered from 1000-year-old mummies. However, it can be very quickly degraded into smaller fragments when exposed to humidity, heat, or sunlight. Today's DNA-profiling methods target very short DNA fragments, the STRs. They may still be applied successfully on partially degraded samples. Furthermore, there are all the chemicals such as dyes or proteins contained in the traces or their support that can interfere in the delicate enzymatic steps involved in PCR amplification, the so-called inhibitors. All the DNA typing procedures involve a DNA isolation and purification step, which tries to eliminate the inhibitors. But their complete elimination is never warranted. Negative results (absence of DNA profile at the end of the analysis) then have to be critically scrutinized since they can be caused either by the absence of DNA or by the interference of inhibitors.

The remarkable sensitivity of multiplex PCR analysis of STRs makes it particularly vulnerable to contamination events. Accidental DNA exchange from one sample to the other, from lab personnel to the samples, etc., can occur and be unnoticed. The large-scale use of negative controls should reveal such contamination events. Moreover, the establishment of adequate laboratory design and use of pertinent working guidelines and decontamination procedures should prevent the occurrence of such damaging episodes.

Involuntary transfer of biological material is the cause of the existence of crime scene traces whose analysis has made a DNA-profiling techniques success. But involuntary transfer of biological material is also a matter of concern when it leads to contamination. In the worst scenarios, involuntary transfer of biological material can have even more worrying consequences when innocent sources leave traces that might be wrongly considered as crime scene traces. In such cases the power of DNA profiling may lead to the use of nonpertinent traces to accuse innocent persons of having a link to a crime. This danger is made even more likely by the extension of intelligence databases.

### DNA Databases

The identification power of DNA profiling makes the establishment of databases for the storage of DNA profiles not only possible but also highly desirable. DNA profiles from new or unsolved cases can be compared with the DNA profiles of the individuals stored in the database or can be compared between them with the hope of obtaining useful links. The standardization necessary for such purposes is a goal that has been reached at the end of the 1990s.

Almost all the industrialized countries have created a national DNA database for intelligence purposes. Depending upon the age of the database, and the legal framework limiting its use, between 0.1% and 2% of the population is usually on file. When the size of the database increases, the success rate reaches impressive level. For the largest databases available, the source of a crime scene trace is revealed by the database in more than 50% of the cases. The trend in most countries is then to adjust the limits of the database so as to include all the individuals ever convicted of a crime. It can be predicted that the DNA profiles of more than 10 million persons will be stored in DNA intelligence databases worldwide by 2010.

When such huge databases become connected, the occurrence of adventitious hits is no more an improbable possibility and becomes a real risk, even

with random match probabilities as low as  $10^{-15}$ . As a consequence, the number of markers analyzed might have to be increased.

### Interpretation

The power of the DNA evidence has to be evaluated by comparing two competing hypotheses. According to the Bayesian framework of evidence interpretation, it is usually expressed as a ratio of the probability of the evidence under the hypothesis that the evidential material has the same source as the comparison sample (the hypothesis favored by the accusation) and the probability of the evidence under the hypothesis that somebody else is the source of the evidence (the hypothesis favored by the defense). This likelihood ratio forms one of the core elements of the Bayesian framework of evidence interpretation. When the ratio is greater than 1, it supports the accusation hypothesis, and when it is lower than 1, it supports the defense hypothesis. In simple cases this ratio is inversely proportional to the frequency of the detected genetic characteristics in the adequate reference population. It is as much higher as the characteristics are rare.

The frequency of a DNA profile is calculated by multiplying the frequency of the result obtained at each STR analyzed. The frequency of the heterozygous genotype  $A_1A_2$  at a putative STR A is the product  $2a_1a_2$  (with  $a_1$  and  $a_2$  being the respective allele frequencies). For the homozygous genotype  $B_1B_1$  at STR B, the frequency would be  $(b_1)^2$ . The frequency of the combined genotypes is then  $2a_1a_2 \times (b_1)^2$  and so on for the whole profile. However, all these multiplications can only be done under the assumption of statistical independence between the events whose frequencies are multiplied. This translates into certain assumptions related to the genetic independence of the analyzed STRs and to the genetic homogeneity of the populations (Hardy-Weinberg equilibrium). The accumulated data demonstrate no major contradiction to the independence assumptions. Nevertheless, they show that real human communities do not behave like the ideal groups of population genetics, which are infinitely large, mate at random, and do not segregate. As a consequence, for improved accuracy, the frequency calculation of DNA profiles should include some correction parameters to take into account the imperfections of the real world. More important than the subtleties of population genetics is the dealing with situations where relatives of the suspects might also be potential suspects or situations where the suspect was identified through a database search. It is easy to get a gross overestimation of the power of the



evidence when these elements are not properly taken into account.

Further challenges in the interpretation process are encountered in cases of DNA profiles from complex mixtures of two or more DNA, and naturally with the DNA profiles obtained with the already mentioned ultrasensitive LCN protocols. In such cases, the Bayesian mode of interpretation displays its wonderful capacity to bring complex data into a coherent conceptual framework leading to sound and founded conclusions.

### Paternity Testing

The identification power of DNA profiling has naturally been used in paternity testing. Since a child has to share an allele with his biological father at each STR analyzed, the same multiplex STR kits, which serve so well the forensic science, can also provide powerful answers in paternity testing. But here, the question relates to the transmission of genetic material from one generation to the other. Then mutations have to be taken into account, which can cause apparent exclusions of true fathers for one or more of the STRs analyzed. The polymorphism of the STRs is linked to quite a high mutation rate (0.001–0.005 mutations per generation). These rates are high enough to bring an apparent exclusion of

the true father for at least one STR in every 50 cases of paternity testing. The statistical evaluation of paternity cases has to integrate these mutation rates. But the power of the DNA profiles is high enough to provide clear statistical evidence even with an apparent mutation at one STR of the multiplex set analyzed.

*See also:* **Capillary Electrophoresis:** Overview; Clinical Applications. **DNA Sequencing. Electrophoresis:** Overview. **Polymerase Chain Reaction.**

### Further Reading

- Butler JM (2001) *Forensic DNA Typing*. New York: Academic Press.
- Coquoz R (2003) *La preuve ADN*. Lausanne: Presses polytechniques et universitaires romandes (in French).
- Evvett IW and Weir B (1998) *Interpreting DNA Evidence. Statistical Genetics for Forensic Scientists*. Sunderland, MA: Sinauer Associates.
- National Research Council (1996) *The Evaluation of Forensic DNA Evidence*. Washington: National Academy Press.
- Rudin N and Inman K (2001) *An Introduction to Forensic DNA Analysis*, 2nd edn. Boca Raton, FL: CRC Press.

## Drug Screening in Sport

**L Rivier**, Laurent Rivier Science Consulting, Lausanne, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

In sports, sophisticated pharmacological products and methods have been used over the recent years. Besides stimulants and narcotics, anabolic agents and peptide hormones, many substances are used and abused to improve the mental and/or physical performance of athletes. This is called doping and contradicts both sport and medical ethics. The credibility of the excellence of performance and unpredictability of the outcome of a competition attract much enthusiasm from the public. If drug misuse or pharmacological manipulations of the athletes remove that unpredictability, the public will lose its interest in the outcome of these sports. Sports practitioners are basically healthy and should not be administered

pharmacologically active substances for any sound medical purposes. For the International Olympic Committee (IOC, [www.olympic.org](http://www.olympic.org)), the International Sport Federations, and the World Anti-Doping Agency (WADA, [www.wada-ama.org](http://www.wada-ama.org)), doping is considered as cheating. Such offence can be sanctioned by penalties as high as several years of suspension. The sports authorities take these important decisions after the laboratory has reported an adverse finding, i.e., the presence of a forbidden substance or its metabolite(s) in the athlete's sample (usually urine) at any concentration.

### The Compounds to be Screened

Because of greater concerns about the health of the young athletes (especially of adolescents, both male and female) and because of the obscure origin of the products as well as of the very high doses used, there is urgent need to implement the best screening capacity of the analytical procedures. Thus, all IOC



evidence when these elements are not properly taken into account.

Further challenges in the interpretation process are encountered in cases of DNA profiles from complex mixtures of two or more DNA, and naturally with the DNA profiles obtained with the already mentioned ultrasensitive LCN protocols. In such cases, the Bayesian mode of interpretation displays its wonderful capacity to bring complex data into a coherent conceptual framework leading to sound and founded conclusions.

### Paternity Testing

The identification power of DNA profiling has naturally been used in paternity testing. Since a child has to share an allele with his biological father at each STR analyzed, the same multiplex STR kits, which serve so well the forensic science, can also provide powerful answers in paternity testing. But here, the question relates to the transmission of genetic material from one generation to the other. Then mutations have to be taken into account, which can cause apparent exclusions of true fathers for one or more of the STRs analyzed. The polymorphism of the STRs is linked to quite a high mutation rate (0.001–0.005 mutations per generation). These rates are high enough to bring an apparent exclusion of

the true father for at least one STR in every 50 cases of paternity testing. The statistical evaluation of paternity cases has to integrate these mutation rates. But the power of the DNA profiles is high enough to provide clear statistical evidence even with an apparent mutation at one STR of the multiplex set analyzed.

*See also:* **Capillary Electrophoresis:** Overview; Clinical Applications. **DNA Sequencing. Electrophoresis:** Overview. **Polymerase Chain Reaction.**

### Further Reading

- Butler JM (2001) *Forensic DNA Typing*. New York: Academic Press.
- Coquoz R (2003) *La preuve ADN*. Lausanne: Presses polytechniques et universitaires romandes (in French).
- Evvett IW and Weir B (1998) *Interpreting DNA Evidence. Statistical Genetics for Forensic Scientists*. Sunderland, MA: Sinauer Associates.
- National Research Council (1996) *The Evaluation of Forensic DNA Evidence*. Washington: National Academy Press.
- Rudin N and Inman K (2001) *An Introduction to Forensic DNA Analysis*, 2nd edn. Boca Raton, FL: CRC Press.

## Drug Screening in Sport

**L Rivier**, Laurent Rivier Science Consulting, Lausanne, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

In sports, sophisticated pharmacological products and methods have been used over the recent years. Besides stimulants and narcotics, anabolic agents and peptide hormones, many substances are used and abused to improve the mental and/or physical performance of athletes. This is called doping and contradicts both sport and medical ethics. The credibility of the excellence of performance and unpredictability of the outcome of a competition attract much enthusiasm from the public. If drug misuse or pharmacological manipulations of the athletes remove that unpredictability, the public will lose its interest in the outcome of these sports. Sports practitioners are basically healthy and should not be administered

pharmacologically active substances for any sound medical purposes. For the International Olympic Committee (IOC, [www.olympic.org](http://www.olympic.org)), the International Sport Federations, and the World Anti-Doping Agency (WADA, [www.wada-ama.org](http://www.wada-ama.org)), doping is considered as cheating. Such offence can be sanctioned by penalties as high as several years of suspension. The sports authorities take these important decisions after the laboratory has reported an adverse finding, i.e., the presence of a forbidden substance or its metabolite(s) in the athlete's sample (usually urine) at any concentration.

### The Compounds to be Screened

Because of greater concerns about the health of the young athletes (especially of adolescents, both male and female) and because of the obscure origin of the products as well as of the very high doses used, there is urgent need to implement the best screening capacity of the analytical procedures. Thus, all IOC

and WADA accredited laboratories are asked to systematically test for stimulants, narcotic analgesics, anabolic agents, diuretics, peptide hormones, and other pharmacological agents and/or methods aiming to alter performance in sport, the list of which is automatically made known and regularly updated on the WADA website pages.

The WADA list of prohibited substances and methods is covering the following items on January 1, 2004.

## I. Substances and Methods Prohibited in-Competition

### IA. Prohibited Substances

**S1. Stimulants** All stimulants are prohibited, including both their optical (D- and L-) isomers and other substances with similar chemical structure or similar pharmacological effects.

#### Remarks:

- A few substances should present a concentration higher than a defined threshold (e.g., cathine is prohibited when its concentration in urine is greater than  $5 \mu\text{g ml}^{-1}$ ; each of ephedrine and methylephedrine is prohibited when its concentration in urine is greater than  $10 \mu\text{g ml}^{-1}$ ).

**S2. Narcotics** Only the following narcotics are prohibited: buprenorphine, dextromoramide, diamorphine (heroin), hydromorphone, methadone, morphine, oxycodone, oxymorphone, pentazocine, and pethidine.

**S3. Cannabinoids** Cannabinoids (e.g., hashish and marijuana) are prohibited.

**S4. Anabolic agents** All anabolic agents are prohibited.

1. Anabolic Androgenic Steroids (AAS)
  - a. Exogenous\* AAS and their analogs#.
  - b. Endogenous\*\* AAS including but not limited to androstenediol, androstenedione, dehydroepiandrosterone (DHEA), dihydrotestosterone, testosterone and their analogs.
2. Other anabolic agents: clenbuterol, zeranol.

#### Remarks:

- 'Exogenous' refers to a substance that is not capable of being produced by the body naturally.
- \*\* 'Endogenous' refers to a substance that is capable of being produced by the body naturally.

- # An 'analog' is defined as 'a substance derived from the modification or alteration of the chemical structure of another substance while retaining a similar pharmacological effect.'

Where a prohibited substance (as listed above) is produced by the body naturally, a sample will be considered to contain such prohibited substance when the concentration of that substance or its metabolites or markers and/or any other relevant ratio(s) in the athlete's sample deviates from the range of values normally found in humans (that is 'not to be consistent with normal endogenous production').

**S5. Peptide hormones** The following substances are prohibited, including their mimetics\*, analogs#, and releasing factors:

1. erythropoietin (EPO);
2. growth hormone (hGH) and insulin-like growth factor (IGF-I);
3. chorionic gonadotrophin (hCG) prohibited in males only;
4. pituitary and synthetic gonadotrophins (luteinizing hormone, LH) prohibited in males only;
5. insulin; and
6. corticotrophins.

Unless the athlete can demonstrate that the concentration of any of these endogenous peptides was due to a physiological or pathological condition, a sample will be considered to contain such prohibited substance as for the endogenous anabolic agents.

The presence of analogs, mimetics, diagnostic marker(s), or releasing factors of a hormone listed above or of any other finding which indicate(s) that the substance detected is not the naturally present hormone, will be reported as an adverse analytical finding.

#### Remarks:

- A 'mimetic' is defined as a substance with pharmacological effect similar to that of another substance, regardless of the fact that it has a different chemical structure.
- # An 'analog' is defined as for the anabolic agents.

**S6. Beta-2 agonists** All beta-2 agonists including their D- and L-isomers are prohibited except that formoterol, salbutamol, salmeterol, and terbutaline are permitted by inhalation only to prevent and/or treat asthma and exercise-induced asthma/bronchoconstriction.

When the laboratory has reported a concentration of salbutamol (free plus glucuronide) greater than  $1000\text{ }\mu\text{g l}^{-1}$ , this will be considered as an adverse analytical finding unless the athlete proves that the abnormal result was the consequence of the therapeutic use of inhaled salbutamol.

**S7. Agents with anti-oestrogenic activity** Aromatase inhibitors, clomiphene, cyclofenil, and tamoxifen, are prohibited only in males.

**S8. Masking agents** Masking agents are prohibited. They are products that have the potential to impair the excretion of prohibited substances, to conceal their presence in urine or other samples used in doping control, or to change hematological parameters.

Masking agents include, but are not limited to, diuretics, epitestosterone, probenecid, and plasma expanders (e.g., dextran and hydroxyethyl starch).

**S9. Glucocorticosteroids** Glucocorticosteroids are prohibited when administered orally, rectally, or by intravenous or intramuscular administration. All other administration routes require a medical notification.

## IB. Prohibited Methods

**M1. Enhancement of oxygen transfer** The following items are prohibited.

1. Blood doping. Blood doping is the use of autologous, homologous or heterologous blood, or red blood cell products of any origin, other than for legitimate medical treatment.
2. The use of products that enhance the uptake, transport, or delivery of oxygen, e.g. erythropoietins, modified hemoglobin products, including, but not limited to, hemoglobin-based blood substitutes, microencapsulated hemoglobin products, perfluorochemicals, and efaproxiral (RSR13).

**M2. Pharmacological, chemical, and physical manipulation** Pharmacological, chemical, and physical manipulation is the use of substances and methods, including masking agents, which alter, attempt to alter, or may reasonably be expected to alter the integrity and validity of specimens collected in doping controls. These include, but are not limited to, catheterization, urine substitution, and/or tampering, inhibition of renal excretion and alterations of testosterone and epitestosterone concentrations.

**M3. Gene doping** Gene or cell doping is defined as the nontherapeutic use of genes, genetic elements, and/or cells that have the capacity to enhance athletic performance.

## II. Substances and Methods Prohibited Out-of-Competition

During training, athletes can also abuse prohibited substances. Then, only the so-called long-term acting performance enhancing compounds are targeted during screening. Thus, S1 (stimulants), S2 (narcotics), and S3 (cannabinoids) groups of prohibited substances are not checked in these samples collected outside competition's time.

## III. Substances Prohibited in Particular Sports

Quite understandably, special pharmaceuticals, mainly psychoactive substances, are forbidden 'in-competition only' in specific sports.

**P1. Alcohol** Ethanol is prohibited, only in a few selected sports (e.g., when driving an engine is necessary) at variable levels ranging from 0 to  $0.5\text{ g l}^{-1}$ .

**P2. Beta-Blockers** Unless otherwise specified, beta-blockers are prohibited in a selection of sports like: Aeronautic (FAI), Archery (FITA) (also prohibited out of competition), Automobile (FIA), Billiards (WCBS), Curling (WCF), Football (FIFA), Gymnastics (FIG), Motorcycling (FIM), Shooting (ISSF) (also prohibited out of competition), Skiing (FIS) ski jumping & free style, as well as snow board, etc.

## Dedicated Procedures in Special Laboratories

As doping accusations can have major economical, sportive, and personal consequences, the whole procedure shall be forensic, e.g., strictly controlled, secured, and verified. That is why only accredited laboratories are allowed to proceed with these official analyses.

In practice, when each sample enters the laboratory, it shall be first classified into in- or out-of competition testing, but, for simplification reasons, the same screening procedures are applied to all of them. The interpretation of the results is adapted accordingly when the analyses are finished. Consequently, the existence of the official list of prohibited substances means that urine or sometimes blood samples

have to be searched systematically for more than 250 analytes, often at the low microgram per liter levels.

The analytical procedures for the small (e.g., stimulants) and medium-sized molecules (e.g., derivatized steroids or glucocorticoids) are using high-resolution gas chromatography (HR-GC) with nitrogen-specific detection or liquid chromatography (LC) with diode array detection, combined gas chromatography–mass spectrometry (GC–MS) and combined LC–MS on the full scan mode or selected ion monitoring mode. For larger molecules such as the peptide hormones, immunoassays and electrophoretic separation are preferred.

Recent statistics from the 32 IOC accredited laboratories have shown that more than 150 000 samples are analyzed annually worldwide, out of which less than 2% are declared positive by them. Around 50% of all these adverse findings are anabolic agents related to testosterone (Figure 1).

In general, this strictly controlled procedure for screening in sport doping is complex and includes the following key elements.

**Chain of custody.** As far as chain of custody is concerned, it starts well before the vials are entering the laboratory's area with the proper collection and tampering-proof transportation of the samples. In the laboratory itself, complete knowledge of the follow-up of each sample is recorded. It ends with the correct interpretation and reporting of the results and the disposal of the samples.

**Validated analytical methodology.** Analytical methods are implemented only if they have been correctly validated. The general procedure for method validation is defined in the several ISO (International Standards Organisation) 17025 companion documents. The accumulation of a vast body of analytical data is most often necessary covering both intra- and interlaboratory variations. WADA is the ultimate body that allows the use of a newly

developed procedure as it was demonstrated recently for EPO in 2002.

**Quality assurance including ISO 17025 and WADA accreditation.** Both WADA and ISO (International Standards Organization) 17025 accreditations are aimed at bringing all laboratories to the same level of good practice. Due to these elevated requirements for quality, doping control laboratories have to use highly specific and selective approaches and have to cope with some degree of harmonization of their methods. Thus, if clear rules have been disclosed in certain restricted professional circles, they might become mandatory for all the laboratories through the new WADA accreditation process that will take place starting from 2004.

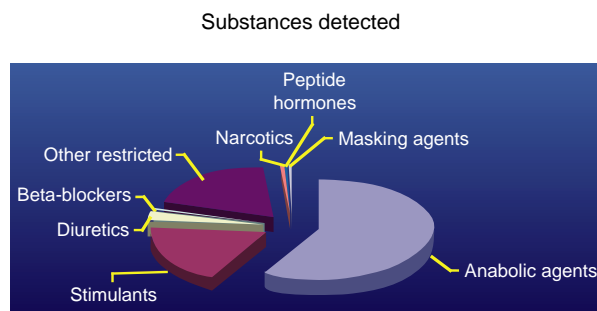
## Short History of Doping Testing in Sport and Actual General Approach

For their own credibility, it is evident that these highly specialized laboratories have to agree between them upon common analytical procedures and interpretation criteria for the confirmation of identity of the compounds of interest. This was developed over the years, starting in 1967 with the first official human doping controls taking place at the Olympic games. At that time, the IOC had banned the use of sympathetico-mimetic amines and other stimulants, narcotics, tranquillizers, and antidepressive substances of the IMAO type. No steroids were included, even with the introduction of the GC–NPD at the Munich Winter Games in 1972, until 1976 at the Montreal Olympic Games with the use of a radioimmunoassay. Beginning 1983, GC–MS was systematically used and, since then, doping control laboratories have used to their advantage any technical advance as soon as they were available to them.

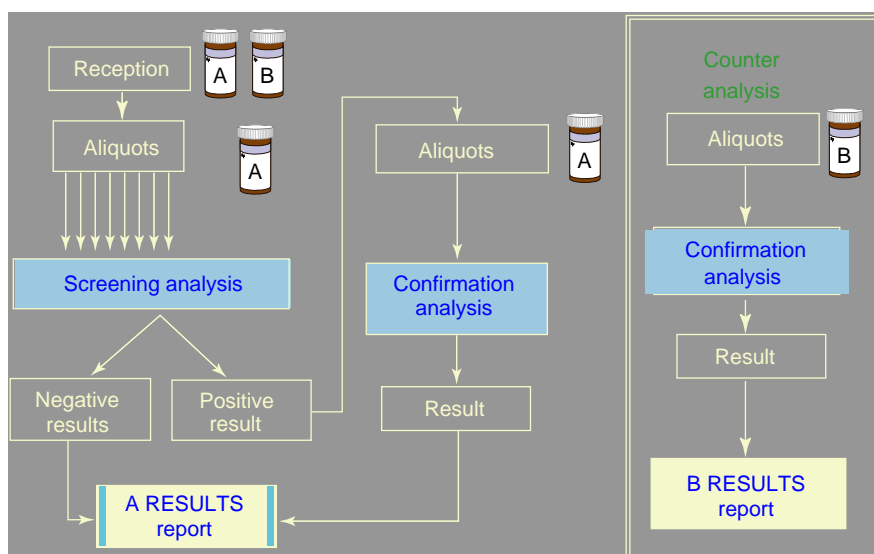
The most advanced analytical approaches are incorporated into doping controls. Today, MS and particularly GC–MS, is so far considered the less likely technique to be challenged in Court. Noticeably, LC–MS has significantly added to the extension of the panel of compounds already detected by GC–MS including large, polar, and/or thermolabile substances.

Now, the general strategy for handling and analyzing the samples within the laboratory is described in Figure 2.

After registering each sample, the volume and the aspect of the urine are recorded. The vial A is then opened and the pH and density are measured together. The analytical approach for the screening is presently restricted to the list of forbidden substances and the most important of their metabolites. It corresponds to approximately 200–250 compounds. Divided by



**Figure 1** Averaged distribution of adverse reports obtained on urine A samples emitted by all the IOC accredited laboratories in the recent years following the classes of prohibited substance. (Adapted from original 1995–2000 data released previously by IOC; see, for example, [www.olympic.org](http://www.olympic.org).)



**Figure 2** Place of the screening in the whole analytical process for doping controls. Positive samples are always confirmed before reporting an adverse finding. For safety reasons, sample B analysis – also called second confirmation or counter analysis – is performed on the athlete's and/or sport authorities' request almost always in the same laboratory as sample A.

groups of substances, the screenings for doping substances can be described as followed:

- Screening A. 5 ml: stimulants and volatile nitrogen-containing compounds – liquid–liquid extraction – GC-NPD analysis;
- Screening B. 5 ml: narcotic analgesics and  $\beta$ -blockers excreted free and/or conjugated – basic hydrolysis – liquid–liquid extraction, concentration – trimethylsilyl (TMS) derivatization – GC–MS full scan analysis;
- Screening C. 2 ml: anabolic steroids, some diuretics, some  $\beta$ -blockers and  $\beta$ -2-agonists – purification – enzymatic hydrolysis – extraction – concentration – TMS, trifluoroacetyl (TFA) derivatization – GC–MS selected ion monitoring (SIM) analysis;
- Screening D. 1 ml: diuretics and acids like THC-COOH – methylative extraction – purification – concentration – GC–MS full scan analysis;
- Screening E. 1 ml: peptide analysis for hCG, FSH, and LH – immunoassays; and
- Screening F. 1 ml: corticosteroids free and conjugated – extraction – LC–MS SIM and full scan analysis.

If specially requested by the sport authorities, separate urine aliquot is collected and screening for EPO and derivatives are done separately:

- Screening G. 20 ml: EPO analysis – ultrafiltration – gel electrophoresis – immunoblotting transfer – revelation and densitometry

This general approach is not permanent, but each time the laboratories improve the efficacy of the detection procedure, cheaters immediately turn into any possible remaining loopholes. As a recent example, tetrahydrogestrinone (THG), a new designer steroid produced specifically for the exclusive purpose to improve athletic performances, posed a real identification challenge to the laboratories. The development or the implementation of rapid, informative, and reliable systematic screening tests, which have the required sensitivity and can withstand the scrutiny of cross-examination at the Court, is constantly needed.

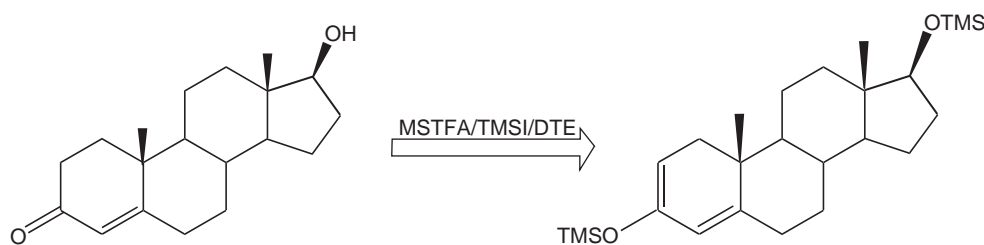
## The Specific Analytical Techniques Involved in Doping Controls

### Sample Preparation

In official doping controls, limited amount of urine (70 ml total) and blood (3–5 ml total) samples are available for the detection of all listed drugs including their metabolites.

The use of urine has been advocated as best sample for doping controls, in part due to the relative easiness of collection and in part because of the higher levels found in urine over blood. Most doping agents, especially steroids, are excreted with their hydroxyl function being substituted with either  $\beta$ -glucuronide or sulfate. Measurements of doping compounds excreted in urine are currently performed either as the





**Figure 3** The quantitative reaction for the derivatization of testosterone is obtained with a mixture of 50  $\mu\text{l}$  *N*-methyl-*N*-trimethylsilyltrifluoroacetamide (MSTFA)/trimethylsilyltrimethylsilane (TMSI)/dithioerythriol (DTE) (1000  $\mu\text{l}$ :5  $\mu\text{l}$ :5 mg) in a stopped vial at 60°C for 30 min.

free form or as their aglycone after hydrolysis of their conjugates.

The analytical criteria for reporting positive findings in screening are the following:

- At the minimum, the analysis should consist of two steps, namely, screening the A-sample and performing an A-sample confirmation.
- The order recommended for injection into the analytical instrument for confirmation analysis is: (1) reagent blank, if pertinent, (2) negative control urine, (3) sample being confirmed, (4) negative control urine, and (5) positive control urine.

At present, the most complex procedure for small and medium-sized molecules concerns the detection of all listed steroids at the low microgram per liter level in routine. It is presented here as a typical example:

- 5 ml urine;
- add 25  $\mu\text{l}$  methylandrostenediol ( $0.02 \text{ mg ml}^{-1}$ ) (Internal Standard);
- apply on  $\text{C}_{18}$  500 mg solid-phase extraction column (prewashed with 5 ml methanol and 5 ml  $\text{H}_2\text{O}$ );
- wash with 5 ml  $\text{H}_2\text{O}$ , elute with 3 ml methanol evaporate at 40°C with rotary vacuum evaporator add 1 ml phosphate buffer  $0.2 \text{ mol l}^{-1}$  pH 7.0, 30  $\mu\text{l}$  *Escherichia coli*  $\beta$ -glucuronidase (5000 units glucur);
- hydrolyze at 50°C for 1 h or at 37°C overnight cool at room temperature, then add 100 mg solid buffer; pH 8.5–9 (disodium carbonate: sodium hydrogen carbonate. 1:10 w/w) and 5 ml diethyl ether;
- shake 10 min, centrifuge 5 min, 2500 rpm;
- save organic phase (dried with sodium sulfate) evaporate to dryness, dry for at least 15 min in the vacuum chamber then add 50  $\mu\text{l}$  MSTFA/TMSI/DTE (1000  $\mu\text{l}$ :5  $\mu\text{l}$ :5  $\mu\text{l}$ ) warm at 60°C for 30 min; and
- inject 1  $\mu\text{l}$  to GC–MS

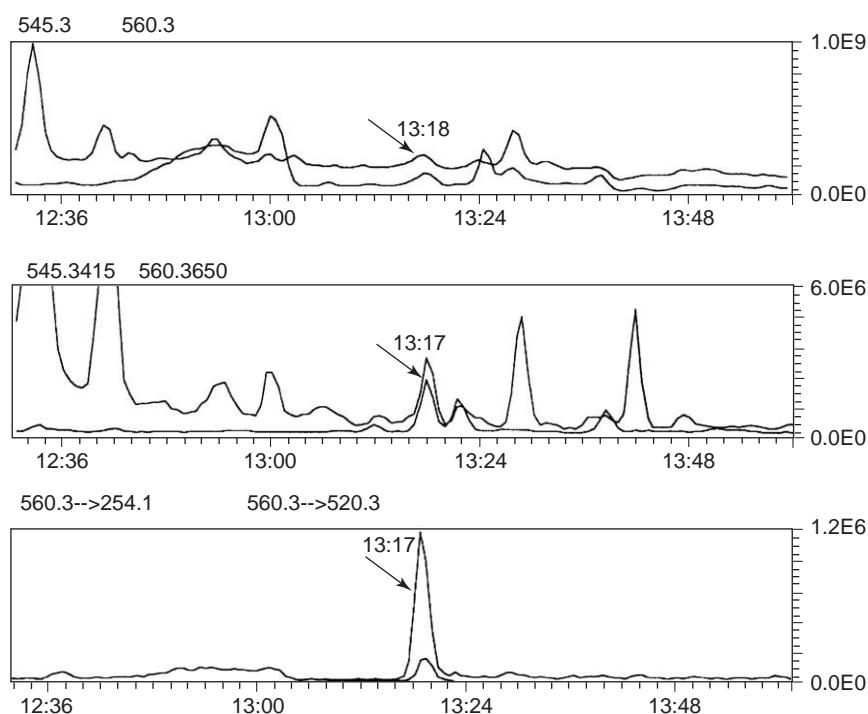
A key part of this procedure is the derivatization step that was optimized in order to give a single derivative of the target molecules with both sufficient thermal stability in GC and favorable fragmentation pattern in MS. This reaction is quantitatively silylating both hydroxy- and keto-groups for all steroids (Figure 3).

#### GC–MS Detection

Automated GC–MS instrumentation is best suited to run the many samples in a short period of time and to eliminate efficiently the many negatives. Quadrupole or ion trap MS instruments allow both automatic injection and treatment of the results. Printout of the results is obtained by running a series of specifically written macro commands. In general, total ion or selected ion traces are generally very complex even at restricted time windows, and targeted substances might be hidden behind the biological matrix background noise. A qualified laboratory technician will read the reports with a second person, usually a senior scientist, verifying the findings of the first. In any case, it is considered as a very tiring process where mistakes are most often occurring with possible false negative and one should not spend more than 60 min on it in a row.

Because of the so many true negative results (>95%) even at low microgram per liter concentration levels, the doping analyst should get best compromise between selectivity of the extraction and high recovery yield, speed, and low cost. As indicated in the steroids' procedure, liquid–liquid partitioning or solid-phase extraction – either manual or automated – are most commonly used. Immunoaffinity chromatography and LC fraction collection have recently been introduced for better removing the biological background most often seen in GC–MS detection. Together with GC–MS/MS, GC–HRMS has significantly improved the detection limit (Figure 4).





**Figure 4** On this illustration, the 'upper' trace shows that identification of 3'OH-stanozolol-*tris*-TMS ( $< 5 \mu\text{g/l}$  in a urine extract) by low resolution MS (upper window) had failed due to the high level of background. If the specificity of detection is enhanced ('middle': high resolution  $R = 10\,000$ , 'below': tandem MS), the limit of detection is essentially improved and the presence of the compound can be much clearly established. All other parameters (extraction, chromatography, ions detection) remained unchanged. Selected ion monitoring (SIM) traces for low nanogram level detection of one TMS derivative of the main Stanozolol metabolite found in urine. Upper part: recorded traces of  $m/z$  545.5 and 560.3 with a quadrupole MS running at the unit resolution; middle part: recorded traces of  $m/z$  545.3415 and 560.3650 with a high resolution MS; and lower part: recorded traces for transitions of  $m/z$  560.3 $\rightarrow$ 254.1 and 520.3 with a ion trap MS running at the unit resolution. Even if the resulting intensities of the ions – as indicated at the right hand side of the illustration – are lower, a higher selectivity provides a better signal to noise ratio. (Adapted from Kusamran T, de la Torre X, *et al.* (1999) Monitoring of low concentration anabolic steroids in urine samples by ion-trap GC/MS/MS. In: *Recent Advances in Doping Analysis*, vol. 7.)

On this illustration, the 'upper' trace shows that identification of 3'OH-stanozolol-*tris*-TMS ( $< 5 \mu\text{g l}^{-1}$  in a urine extract) by low-resolution MS (upper window) had failed due to the high level of background. If the specificity of detection is enhanced ('middle': high resolution  $R = 10\,000$ , 'below': tandem MS), the limit of detection is essentially improved and the presence of the compound can be much clearly established. All other parameters (extraction, chromatography, and ions detection) remained unchanged.

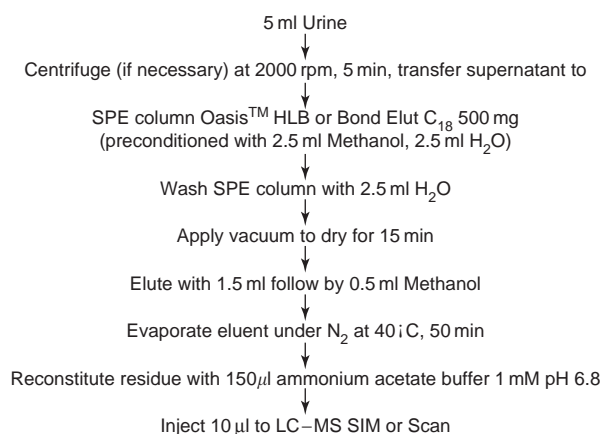
#### LC-MS Detection

The high polarity of the conjugates eliminated in urine limited the use of GC-MS or GC-MS/MS to the hydrolyzed samples only. Such measurements offer targeted molecules with modified structure for identification and rather imprecise quantitative measurements of each individual moiety. Recent advances in LC-MS interface technology have

allowed the detection of a larger pattern of thermolabile and polar compounds or metabolites involved in sport doping. One of the major advantages of liquid introduction in MS over other ways is that biological sample preparation can be greatly simplified. Sometimes, direct injection of the biological sample is even possible. For example, 11 abused corticosteroids can be rapidly extracted from 5 ml urine (Figure 5) and easily screened routinely in one single injection at the  $1\text{--}5 \mu\text{g l}^{-1}$  level by LC-electrospray ionization-MS (LC-ESI-MS). Many more applications of LC-MS can be expected in the near future.

#### Further Discrimination between Endogenous and Exogenous Doping Agents

Pharmaceutical testosterone and natural testosterone mass spectra are identical when measured with



**Figure 5** Extraction schema for LC-MS corticosteroids and confirmation. The procedure, that last ~2 h when automated with a suitable robotic system, can treat several samples in parallel.

ordinary MS detectors. In recent years, gas chromatography-carbon, online combustion-isotope ratio mass spectrometry (GC-C-IRMS) has been developed as a complementary technique to GC-MS. This technique made it possible to discriminate exogenous testosterone from endogenous testosterone by measuring  $^{13}\text{C}/^{12}\text{C}$  ratio of these steroids. Since the 1998 Olympic Winter Games in Nagano, GC-C-IRMS has been accepted in doping controls. However, a relatively large volume of urine (~25 ml) is necessary and time-consuming cleanup procedure is needed for obtaining the necessary highly purified steroid fractions. However, not all doping control laboratories are able to implement it on a routine basis.

## The Availability of Reference Dope Substances

Many doping substances are available at the requested purity for reference purpose through chemical factories. However, most of the metabolites are not. Consequently, all the laboratories are forced to prepare their own certified urines from controlled excretion studies. After the relevant permission from the medical ethical commission is granted, a small and unique dose of the doping agent is given to one volunteer. This healthy individual then delivers several urine samples that allow the laboratory to trace and identify the most relevant metabolites. After being analyzed, the remaining samples are mixed together. Several aliquots are then kept at low temperature until used as reference collection sample.

## Criteria for Identification

Practically, identification of a small molecule, based on its mass spectrum (MS) is usually obtained by passing the corresponding reference compound or standard the same day through the same analytical procedure and instrumentation, and then by comparing both the collected data. In this way, the laboratory will minimize instrumental variations to meet the required criteria for identification:

- Chromatographic separation
  - For capillary GC, retention time (Rt) or relative Rt (RRt) of the analyte shall not differ more than 1% from that of the same substance in the spiked urine sample, reference collection sample, or reference material analyzed in the same batch. In these cases where shifts in retention can be explained, for example, column overload, the Rt criteria can be relaxed. For LC, the Rt or RRt shall not differ more than 2%.
- MS detection
  - Full scan mode: all diagnostic ions with a relative abundance >10% in the reference MS obtained from a positive control urine or reference material analyzed in the same batch must be present in the MS of the unknown. In addition, the relative intensities of at least three diagnostic ions shall not differ more than the amount indicated in the Table 1.
  - If three diagnostic ions with relative abundance >10% are not available, a second derivative shall be prepared, or a second ionization or fragmentation technique shall be used. In any case, a minimum of two diagnostic ions is mandatory in each mass spectrum.
  - SIM mode: at least three diagnostic ions must be acquired. The  $s/n$  ratio must be >3. The relative intensities of any of the ions shall not differ more than the amount indicated in Table 1. For diagnostic ion with relative abundance <5% in the reference, the ion must be present in the unknown.
  - If the laboratory protocol requires three ions to be within a tolerance window to identify a substance, it is not permissible to collect additional ions and select those ion ratios that are within tolerance and ignore others that would not result in meeting identification criteria without valid explanation.
  - If three diagnostic ions are not available, a second derivative shall be prepared, or a second ionization or fragmentation technique shall be used. In any case, a minimum of two diagnostic ions is mandatory in each MS.
  - MS<sup>n</sup>: a minimum of three diagnostic ions must be monitored either as an MS or in the selected

**Table 1** Maximum tolerance windows for relative ion intensities to ensure appropriate uncertainty in identification (adapted from the WADA Laboratory International Standard)

Relative abundance (% of base peak)	EI-GC-MS	CI-GC-MS; GC-MS <sup>n</sup> ; LC-MS; LC-MS <sup>n</sup>
≥50%	± 10% (absolute)	± 15% (absolute)
<50 and ≥25%	± 20% (relative)	± 25% (relative)
<25%	± 5% (absolute)	± 10% (absolute)

Note: for example, absolute difference means 45–55% for a 50% ± (5% absolute) abundance. Relative difference means 40–60% for a 50% ± (20% relative) abundance.

reaction monitoring (SRM) mode. The *s/n* of the diagnostic ions must be greater than 3. The relative intensities of any of the ions shall not differ by more than the amount in Table 1 from the relative intensities of the same ions acquired in a spiked urine, reference collection sample, or reference material. For a diagnostic ion with a relative abundance > 5% in the reference, the ion must be present in the unknown.

## Analytical Approach for Large Molecules and Peptide Detection

These screenings are basically performed by immunoassays directly on the urine or blood, or after specific manipulation of the samples (e.g., ultrafiltration and concentration, blotting transfer, and electrophoretic separation). In principle, formal confirmation shall be obtained by using other immunological tests acting on a different epitope of the targeted molecule.

Today, if no hGH protocol is yet fully validated, an official test for recombinant EPO detection and its derivatives like Darbepoietine is in place since 1999. It is highly innovative and specific, but it is so sophisticated and time-consuming, expensive, and labor intensive that only highly trained technicians are able to perform it.

## Inter- and Intralaboratory Variations

Very few collaborative studies between accredited doping laboratories are available to determine their variances. As such determinations are rather complex, some of the experimental parameters are most of the time left uncontrolled. Great attention should be given to standardization of the methods. Matrix problems are known during calibration and best results can be obtained with the systematic use of good quality deuterium-labeled internal standards. The

sport authorities are fighting against the misuse of doping agents with the direct help of their accredited analytical laboratories, the efficacy of which is limited by the techniques used and the content's quality of the list of prohibited substances and methods.

The continuous implementation of new analytical techniques is mandatory to improve the deterrent effect of these controls. Regular interlaboratory comparisons should be organized to define more precisely the qualification level of each laboratory and to overcome the practical limitation of such methodologies. Better standardization of the analytical procedures and greater level in harmonization would probably add confidence to the results.

*See also:* **Chemometrics and Statistics:** Optimization Strategies. **Chromatography:** Overview; Principles. **Derivatization of Analytes.** **Extraction:** Solvent Extraction Principles; Solid-Phase Extraction. **Forensic Sciences:** Systematic Drug Identification. **Gas Chromatography:** Mass Spectrometry. **Hormones:** Steroids. **Liquid Chromatography:** Liquid Chromatography–Mass Spectrometry. **Mass Spectrometry:** Atmospheric Pressure Ionization Techniques; Forensic Applications.

## Further Reading

- Aguilera R, Chapman TE, Starcevic B, Hatton CK, and Catlin DH (2001) Performance characteristics of a carbon isotope ratio method for detecting doping with testosterone based on urine diols: Controls and athletes with elevated testosterone/epitestosterone ratios. *Clinical Chemistry* 47: 292–300.
- Burns M (2002) *Medical–Legal Aspects of Drugs*. Tucson, AZ: Layers & Judges.
- Flury K, Rivier L, Dienes-Nagy A, *et al.* (2001) Method for confirmation of synthetic corticosteroids in doping urine samples by liquid chromatography–electrospray ionisation mass spectrometry. *Journal of Chromatography B* 926: 87–95.
- Geyer H, Schaezner W, Mareck-Engelke U, Nolteersting E, and Opfermann G (1998) Screening procedure for anabolic steroids – the control of the hydrolysis with deuterated androsterone glucuronide and studies with direct hydrolysis. In: Schaezner W, Geyer H, Gotzmann A, and Mareck-Engelke U (eds.) *Recent Advances in Doping Analysis*, vol. 5, pp. 97–102. Köln: Sport und Buch Strauss.
- International Organization for Standardization (1999) ISO/IEC 17025: General requirements for the competence of testing and calibration laboratories. Geneva, Switzerland.
- Lasne F, Martin L, Crepin N, and de Ceaurriz J (2002) Detection of isoelectric profiles of erythropoietin in urine: Differentiation of natural and administered

- recombinant hormones. *Analytical Biochemistry* 311: 119–126.
- Mullis PE (2000) *Doping in Sport*. Baillière's Best Practice & Research: Clinical Endocrinology & Metabolism, vol. 14, No. 1. Elsevier.
- Rivier L (2003) Criteria for the identification of compounds by liquid chromatography–mass spectrometry and liquid chromatography–multiple mass spectrometry in forensic toxicology and doping analysis. *Analytica Chimica Acta* 492: 69–82.
- World Anti-Doping Agency. *International Standard for Laboratories*. Version 2: February 20, 2003. Loadable PDF File on [www.wada-ama.org](http://www.wada-ama.org).
- World Anti-Doping Agency. *The 2004 Prohibited List Internal Standard*. Version November 25, 2003. Loadable PDF File on [www.wada-ama.org](http://www.wada-ama.org).

## Explosives

**R Sleeman and J F Carter**, Mass Spec Analytical Ltd., Bristol, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by C R Midkiff Jr., pp. 1606–1612, © 1995, Elsevier Ltd.

### Introduction

In the forensic laboratory, explosives-related examinations are conducted to confirm the identity of a suspected explosive, to compare an explosive to one in the possession of a suspect, or to identify the material involved in an explosion. Explosives may be submitted for examination intact or as postexplosion debris. Intact explosives represent recovery of stolen explosives, undercover purchase, or a disarmed live or hoax device. Postblast samples are debris collected at the scene of a suspected accidental explosion or criminal bombing or following intentional detonation of an explosive device. In the examination of postblast debris, the analyst initially seeks to locate particles of the original explosive. When intact or partial particles are located, they are physically removed for testing free of interference from other materials in the debris. If no identifiable particulates are observed under low (10–30×) microscopic examination, residues characteristic of a particular class of explosive are sought. Residues may be combustion/decomposition products or residual traces of components of the original explosive such as nitroglycerin (NG) or nitrates used as the oxidizer.

### Types of Explosives Encountered

Most commonly encountered are low explosives or propellants that must be confined, as in a pipe, to be effective. Among inorganic propellant types is black gun-powder or a commercially modified black powder, Pyrodex. Organic propellants are primarily

smokeless powders; single-base, containing nitrocellulose (NC) only or double-base, containing NC and NG for extra energy and faster burning. Rarely, a low explosive of the commercial blasting agent type may be encountered, but because these require a primer charge for initiation, they are infrequent in criminal bombings. One widely used blasting agent is ammonium nitrate fuel oil (ANFO), ~95% ammonium nitrate and 5% fuel oil. A wide variety of improvised mixtures consisting of an organic fuel and an inorganic oxidizer is encountered, including such common materials as crushed match heads. Oxidizers typically comprise chlorates, perchlorates, or nitrates. Improvised devices may also contain fillers of pyrotechnic materials such as flash powders and flare compositions. Depending on formulation, certain flash powders may detonate and are classified as high explosives. Major components of several propellants or low explosives are shown in Table 1.

High explosives are those that detonate unconfined and are of commercial, military, and, rarely, improvised types. Typical examples of commercial high explosives are dynamite, water gels or emulsions, binaries, and certain plastic explosives such as Semtex or Detasheet. Military explosives are typically

**Table 1** Ingredients of representative propellants and low explosives

Explosives	Oxidizer	Fuel	Other
Black gunpowder	KNO <sub>3</sub>	Charcoal/sulfur	
Pyrodex	KNO <sub>3</sub> /KClO <sub>4</sub>	Charcoal/sulfur	
Single-base smokeless		NC	Stabilizers
Double-base smokeless	NG	NC	Stabilizers
ANFO	NH <sub>4</sub> NO <sub>3</sub>	Fuel oil	
Improvised	ClO <sub>4</sub> <sup>−</sup> or ClO <sub>3</sub> <sup>−</sup>	Sugar	
Flash powder	KClO <sub>4</sub>	Al	S or Sb <sub>2</sub> S <sub>3</sub>
Pyrotechnics	Ba(NO <sub>3</sub> ) <sub>2</sub> or Sr(NO <sub>3</sub> ) <sub>2</sub>	Al or Mg	



- recombinant hormones. *Analytical Biochemistry* 311: 119–126.
- Mullis PE (2000) *Doping in Sport*. Baillière's Best Practice & Research: Clinical Endocrinology & Metabolism, vol. 14, No. 1. Elsevier.
- Rivier L (2003) Criteria for the identification of compounds by liquid chromatography–mass spectrometry and liquid chromatography–multiple mass spectrometry

in forensic toxicology and doping analysis. *Analytica Chimica Acta* 492: 69–82.

- World Anti-Doping Agency. *International Standard for Laboratories*. Version 2: February 20, 2003. Loadable PDF File on [www.wada-ama.org](http://www.wada-ama.org).
- World Anti-Doping Agency. *The 2004 Prohibited List Internal Standard*. Version November 25, 2003. Loadable PDF File on [www.wada-ama.org](http://www.wada-ama.org).

## Explosives

**R Sleeman and J F Carter**, Mass Spec Analytical Ltd., Bristol, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by C R Midkiff Jr., pp. 1606–1612, © 1995, Elsevier Ltd.

### Introduction

In the forensic laboratory, explosives-related examinations are conducted to confirm the identity of a suspected explosive, to compare an explosive to one in the possession of a suspect, or to identify the material involved in an explosion. Explosives may be submitted for examination intact or as postexplosion debris. Intact explosives represent recovery of stolen explosives, undercover purchase, or a disarmed live or hoax device. Postblast samples are debris collected at the scene of a suspected accidental explosion or criminal bombing or following intentional detonation of an explosive device. In the examination of postblast debris, the analyst initially seeks to locate particles of the original explosive. When intact or partial particles are located, they are physically removed for testing free of interference from other materials in the debris. If no identifiable particulates are observed under low (10–30×) microscopic examination, residues characteristic of a particular class of explosive are sought. Residues may be combustion/decomposition products or residual traces of components of the original explosive such as nitroglycerin (NG) or nitrates used as the oxidizer.

### Types of Explosives Encountered

Most commonly encountered are low explosives or propellants that must be confined, as in a pipe, to be effective. Among inorganic propellant types is black gun-powder or a commercially modified black powder, Pyrodex. Organic propellants are primarily

smokeless powders; single-base, containing nitrocellulose (NC) only or double-base, containing NC and NG for extra energy and faster burning. Rarely, a low explosive of the commercial blasting agent type may be encountered, but because these require a primer charge for initiation, they are infrequent in criminal bombings. One widely used blasting agent is ammonium nitrate fuel oil (ANFO), ~95% ammonium nitrate and 5% fuel oil. A wide variety of improvised mixtures consisting of an organic fuel and an inorganic oxidizer is encountered, including such common materials as crushed match heads. Oxidizers typically comprise chlorates, perchlorates, or nitrates. Improvised devices may also contain fillers of pyrotechnic materials such as flash powders and flare compositions. Depending on formulation, certain flash powders may detonate and are classified as high explosives. Major components of several propellants or low explosives are shown in Table 1.

High explosives are those that detonate unconfined and are of commercial, military, and, rarely, improvised types. Typical examples of commercial high explosives are dynamite, water gels or emulsions, binaries, and certain plastic explosives such as Semtex or Detasheet. Military explosives are typically

**Table 1** Ingredients of representative propellants and low explosives

Explosives	Oxidizer	Fuel	Other
Black gunpowder	KNO <sub>3</sub>	Charcoal/sulfur	
Pyrodex	KNO <sub>3</sub> /KClO <sub>4</sub>	Charcoal/sulfur	
Single-base smokeless		NC	Stabilizers
Double-base smokeless	NG	NC	Stabilizers
ANFO	NH <sub>4</sub> NO <sub>3</sub>	Fuel oil	
Improvised	ClO <sub>4</sub> <sup>−</sup> or ClO <sub>3</sub> <sup>−</sup>	Sugar	
Flash powder	KClO <sub>4</sub>	Al	S or Sb <sub>2</sub> S <sub>3</sub>
Pyrotechnics	Ba(NO <sub>3</sub> ) <sub>2</sub> or Sr(NO <sub>3</sub> ) <sub>2</sub>	Al or Mg	

**Table 2** Ingredients of representative high explosives

Type	Explosive		Fuel/binder	Other
	I	II		
Dynamite	NG (GTN)	EGDN	Various	Na/NH <sub>4</sub> NO <sub>3</sub>
Water gel/slurry	NH <sub>4</sub> NO <sub>3</sub>			Amine salt
Emulsions	NH <sub>4</sub> NO <sub>3</sub>			Glass spheres
Binary	NH <sub>4</sub> NO <sub>3</sub>	Nitromethane		
Semtex	RDX	PETN	Oil/polymer	
Comp B	TNT	RDX		
Comp C-4	RDX		Oil/PIB	
Improvised	KClO <sub>3</sub>	Nitromethane		

trinitrotoluene (TNT) or cyclotrimethylenetrinitramine (RDX) and combinations of these. Easily improvised high explosives such as triacetone triperoxide are extremely shock sensitive and until recently have only infrequently been encountered in criminal activity. Other types of improvised high explosives have been used in terrorist bombings including mixtures of chlorate and nitrobenzene. Ingredients of some common high explosives are shown in Table 2.

## Analytical Methods for Low Explosives

### Physical and Chemical Tests

When an intact explosive material is available for examination, its morphology and appearance provide an initial guide to its identity. Black powder grains are usually pyramidal in shape with a shiny surface. Ignition of a single grain gives an instantaneous lavender flash for potassium nitrate (KNO<sub>3</sub>) or a yellow flash in the case of sodium nitrate (NaNO<sub>3</sub>). Pyrodex is a black powder propellant for firearms in which part of the KNO<sub>3</sub> is replaced with potassium perchlorate (KClO<sub>4</sub>). Two stabilizers are added to produce a powder of lower ignition sensitivity that reduces barrel corrosion. Grains of Pyrodex are roughly spherical and dark grey. Ignition is similar to black powder. Smokeless powder grains differ in morphology but intact particles are readily recognizable. Grains may be tubular, disk-shaped, spherical, flattened spheres, or flakes. Disk or tube grains may be perforated. On ignition, these burn smoothly rather than with a flash. Flash powder is typically a fine powder, silver in color, and loose powder burns with sparkles.

Smokeless powders are insoluble in water and are readily distinguished from black or flash powders, Pyrodex, or most improvised low explosives which have significant solubility. With inorganic explosives, qualitative analysis for identification may require only simple chemical spot, flame, or microcrystal

**Table 3** Chemical tests for inorganic low explosives

Anion	Chemical spot	Microcrystal
NO <sub>3</sub> <sup>-</sup>	Griess/Zn, Nitron, Brown Ring	Nitron (needles)
ClO <sub>3</sub> <sup>-</sup>	Aniline Sulfate Blue Ring, DPA	
ClO <sub>4</sub> <sup>-</sup>	Methylene Blue/ZnSO <sub>4</sub> , Nitron Triphenylselenium-chloride	Tetrabutylammonium-chloride Methylene blue ortriphenylselenium-chloride
NO <sub>3</sub> <sup>-</sup>	Diphenylamine/Griess	
SO <sub>4</sub> <sup>-</sup>	HCl/BaCl <sub>2</sub>	
SCN <sup>-</sup>	FeCl <sub>3</sub>	
Cl <sup>-</sup>	HNO <sub>3</sub> /AgNO <sub>3</sub>	
ClO <sup>-</sup>	HCl/Na/CCl <sub>4</sub>	
Cation	Chemical spot or crystal	Flame color
NH <sub>4</sub> <sup>+</sup>	Nessler (orange)	
Ba <sup>2+</sup>	H <sub>2</sub> SO <sub>4</sub> (white ppt.)	Green
Ca <sup>2+</sup>		Red-orange
K <sup>+</sup>	Sodium hexanitrocobaltate(III)	Lavender (cobalt glass)
Na <sup>+</sup>	Zinc uranyl acetate/UV fluorescent	Bright yellow
Sb(III)	Triphenylarsonium iodide	
Sr <sup>2+</sup>		Intense red

DPA, diphenylamine.

tests. A small amount of the suspected explosive is treated with hot water and filtered. Initially, the filtrate is tested for soluble oxidizer ions using the simple tests listed in Table 3. The results and information from Table 1 are used to begin characterizing the explosive material. The filter paper is dried and rinsed with pyridine to dissolve sulfur. The pyridine filtrate is collected and a few drops of 10% sodium hydroxide solution are added. A blue-green to brown color, depending on concentration, indicates sulfur. A residue on the filter, if black, is charcoal which can be identified by drying the filter paper and removing a small portion of the black residue. The residue is heated on the tip of a metal spatula until it glows. On removal from the flame, the charcoal will continue to



glow and slowly disappear, resembling the behavior of charcoal briquettes. If the residue is silvery, it is aluminum or magnesium, used in some pyrotechnic mixtures. These may be distinguished by treating separate portions of the residue on a microscope slide with dilute hydrochloric acid and sodium hydroxide and observing the reactions.

If desired, the approach above may be adapted to semiquantitative analysis of black powder, Pyrodex, or flash powders. A weighed sample, 0.25 g or more, on a tared filter paper is treated several times with a few milliliters of hot water, the filter dried, and the weight loss determined. The loss represents the oxidizer concentration. The filter is treated with pyridine, dried, and reweighed. The weight loss represents sulfur concentration and the residue either charcoal or aluminum as appropriate.

ANFO is typically yellowish spheres that dissolve slowly in water. Suspected ANFO is treated with a volatile solvent such as chloroform and the filtrate concentrated by slow evaporation for analysis by gas chromatography (GC) to identify the fuel oil component. The solid residue is dried and tested for ammonium nitrate by either chemical spot tests or instrumental methods. If weighed prior to solvent extraction, reweighing of the residual ammonium nitrate will demonstrate that the oil concentration is in the proper range for ANFO-type explosives.

Smokeless powders are characterized by morphology, physical dimensions, additives, and whether single- or double-base. To determine the base, a few grains of the powder are treated with acetone and the extract examined by thin-layer chromatography (TLC). A few milliliters are spotted on a silica gel plate and developed with a solvent such as 3:1 dichloroethane–carbon tetrachloride for ~10 cm. The plate is sprayed with ethanolic potassium hydroxide, heated at 100°C for 5–10 min, resprayed with Griess reagent, and reheated if needed. NC appears as a red spot at or close to the origin and NG, if present, at about  $R_F = 0.50$ . Alternatively, the grains may be extracted with methylene chloride and examined by GC. The use of GC–mass spectrometry (GC–MS) to distinguish single- and double-base powders has replaced TLC, in many laboratories, because of the unambiguous nature of the identification attained. If a few drops of acetone extract are added to water, a white film or threads of NC are observed. The identity of NC can be confirmed by infrared (IR) spectroscopy in preference to MS or GC–MS.

### Instrumental Analysis of Low Explosives

A variety of instrumental techniques is suitable for the characterization of low explosive propellants.

Scanning electron microscopy-energy dispersive X-ray analysis (SEM-EDX) provides a quick nondestructive determination of the elemental composition of the sample readily identifying barium, potassium, strontium, and chlorine. Inductively coupled plasma (ICP) emission spectroscopy or ICP-MS may be used to fully characterize the elemental composition of a material. X-ray powder diffraction (XRD) provides additional information. Data from the elemental profile are used in combination with a computer search of the powder diffraction file. In this way, compounds typical of a particular type of explosive can be identified, e.g.,  $\text{KNO}_3$  or  $\text{KClO}_4$ . Ion chromatography (IC) of a water extract detects soluble oxidizer cations and anions in an inorganic explosive, but for effective results, a very dilute solution must be used. Anions readily detected include nitrate, chlorate, and perchlorate. Cations such as potassium, sodium, and ammonium are also detected with excellent sensitivity. A complementary technique to IC, requiring an even smaller sample is capillary electrophoresis (CE), which has high sensitivity and short analysis time. For definitive identification of Pyrodex, liquid chromatography (LC) with ultraviolet (UV) detection is used to identify the characteristic additives sodium benzoate and dicyandiamide. Identification of NG may be made on an organic solvent extract by GC with either electrochemical or infrared chemiluminescence (also known as thermal energy analysis (TEA)) detection. GC–TEA systems are commercially available as ‘sniffer’ systems. Both LC and GC are also useful to identify additives or impurities such as dinitrotoluenes (DNTs) and stabilizers such as ethyl and methyl centralite or diphenylamine to characterize a sample by producer or brand.

### Identification of Residues Typical of Low Explosives

When postblast debris is received in the laboratory, visual examination may identify fragments of a potential container. Deflagrating explosives require a container such as a metal or plastic pipe or heavy cardboard tube to allow pressure to buildup. The presence of container fragments suggests low explosive use and examination of the fragments may provide additional information. In metal pipes, limited fragmentation suggests the use of a relatively low-energy filler such as black powder, Pyrodex, single-base smokeless powder, or an improvised mixture. Higher fragmentation and long thin strips of pipe suggest the use of double-base smokeless powder. With flash powder, residues such as silvery powder and bulking material, e.g., wood shavings, may

be found adhering to the inside surface of glue or epoxy sealed end plugs. A thin surface layer of silver grey aluminum may be seen on the inside surface of a cardboard tube. When sufficient residues are present, they may be removed from the surface manually or by rinsing with hot water. Surface aluminum may be difficult to remove but can be tested as described above directly on the surface. The hot water rinse is concentrated by slow evaporation to 1–2 ml and a few drops removed for anion testing using the spot or microcrystal tests described in **Table 3**.

For black powder, typical combustion/decomposition residues are: potassium sulfate ( $K_2SO_4$ ) and potassium nitrite ( $KNO_2$ ) (major), potassium carbonate ( $K_2CO_3$ ) or potassium hydrogen carbonate ( $KHCO_3$ ) and  $KNO_3$  (minor), and potassium thiocyanate ( $KSCN$ ) (trace). In some instances, an odor of hydrogen sulfide ( $H_2S$ ) may be noted when the test solution is acidified, indicative of sulfides, also seen as interference in the chloride test. Pyrodex produces similar products but a strong chloride test will be obtained and residual perchlorate may be detectable. To confirm Pyrodex as the filler, the extract should be tested for the two additives by LC. Burning of mixtures of black powder and chlorates may produce detectable levels of perchlorate but in tests of such mixtures, levels of residual chlorate are well in excess of those of perchlorate. Combustion of smokeless powders may produce significant levels of carbonates and some nitrites but, if no intact or partial grains are located, additional verification is needed, such as detection of NC/NG.

Both the original oxidizing anion and its decomposition products are often found in the case of improvised mixtures based on chlorates, perchlorates, or nitrates. One improvised explosive/incendiary is a mixture of sodium or calcium hypochlorite with a glycol. When this mixture is used, a strong odor of chlorine may be noted in the sample and chloride and hypochlorite are readily identified in a water extract. Another improvised mixture encountered in the UK is finely ground ammonium nitrate fertilizer and confectioner's sugar. The  $NH_4NO_3$  products are identifiable as described and LC is used for detection of the sugar. In residues from sugar-based improvised explosives, detection of 5-hydroxymethyl furfural by LC may be a useful indicator of the presence of sugar.

Cations are simply identified by acidifying the water extract of the sample with hydrochloric acid and conducting a flame test with a platinum wire loop. The lavender color of potassium is best observed through cobalt blue glass, which filters out the yellow sodium emission. High levels of strontium will also be seen through the cobalt glass but, because of its

intensity, the crimson is visible without the glass. A small spectroscope is useful for the observation of the characteristic spectral lines of barium or calcium if these are present. As indicated in **Table 3**, crystal tests are useful in confirming cation results but with the exception of ammonium ions are only rarely required.

When few or weak results are obtained with flame, spot, or crystal tests, IC or CE provides an excellent approach to the detection and/or confirmation of both anions and cations in residues from inorganic explosives. IC is especially valuable in the detection of carbonate, not easily identified by chemical tests, and thiocyanate, usually present only in very low amounts. Although it has been claimed that perchlorate is undetectable after an explosion, it is readily identified by IC. A major advantage of IC is that it detects a number of ions in a single run but because of its high sensitivity, the water used for extraction of the residue must be of high purity and handling procedures should avoid contamination from containers. In addition, the presence of ions detected at extremely low levels must be interpreted with care. As with the intact materials, SEM-EDX, ICP, and XRD provide useful information with XRD permitting identification of compounds characteristic of the particular explosive used.

When smokeless powder is suspected as having been used in a pipe bomb, the inside surface of the pipe is swabbed, rinsed, or, more effectively, the pipe fragments are sonicated with methylene chloride. The extract is filtered and concentrated by slow evaporation. The extract is examined by GC-TEA or LC-TEA or using an electrochemical detector to detect NG indicative of double-base powder. Similarly, GC-MS may be used for detection of NG with good sensitivity and excellent selectivity. The use of supercritical fluid extraction (SFE) has been used to extract smokeless powders and residues but the equipment cannot accommodate bulky fragments such as pipe fragments.

Pipe fragments may be extracted by sonication with acetone prior to TEA analysis. The concentrated acetone extract may be subjected to TLC as described above. Pipe samples without visible residue may be extracted with methylene chloride and examined by GC-TEA for the detection of NG. These pipe fragments may be re-extracted with acetone and the extract examined by TLC to detect the presence of both NG and NC. This provides good confirmation of smokeless powder as the explosive in the absence of even partial powder grains. Although examination of residues of low explosives by IR spectroscopy has been reported, the results are equivocal and difficult to interpret.

## Identification of Intact High Explosives

High explosives encountered in the forensic laboratory may be either pure or nearly pure compounds: nitro aromatics, nitrate esters, nitramines, or mixtures of these with or without other ingredients. A second category of high explosives with limited commercial application is the binary explosive. These involve the mixing of two, normally nonexplosive components, just prior to use to produce a cap-sensitive high explosive. A third, and more recent, type of high explosive is the gel or emulsion. This type, which is not normally sensitive to initiation by a number 8 detonator (a requirement for classification of an explosive as being cap-sensitive), incorporates one or more added sensitizers to permit initiation by a cap.

Aromatic nitro compounds, primarily 2,4,6-TNT, have long use as military explosives and commercial application in cast boosters to initiate insensitive blasting agents. In the laboratory, TNT maybe encountered as prills, flake, or a cast solid ranging in color from light yellow to brown. Acetone solutions of TNT slowly develop a pink color and red is produced with ethanolic KOH. Numerous TLC systems have been described for the identification of TNT, usually with visualization by alcoholic KOH. TLC or GC-MS identification of minor products, other isomers of TNT or DNT, provides information to establish a potential common source. The use of isotope ratio MS is now being explored for this purpose. GC or LC-TEA readily identifies TNT. Combinations of TNT and  $\text{NH}_4\text{NO}_3$ , known as 'Amatols', have been used as military explosives to conserve TNT and could be encountered in old military ordnance.

Nitrate esters such as NG and ethylene glycol dinitrate (EGDN) are the major explosive components of commercial dynamite. The reported 'nitroglycerin content' of dynamite represents a mixture of NG and EGDN, with EGDN being the major component. In some dynamites, only EGDN may be present. NG and EGDN are soluble in acetone and identifiable by TLC on silica gel or by GC with electron capture detection, GC-TEA or GC-MS. In recent 'nitroglycerin-free dynamites' the NG and EGDN have been replaced with nitro esters such as diethylene glycol dinitrate (DEGDN) or metriol trinitrate (MTN), which are identified using the same techniques. Other ingredients of dynamite that should be identified for characterization include  $\text{NH}_4\text{NO}_3$  or  $\text{NaNO}_3$ , which replace a portion of the 'NG', combustible fillers such as sawdust, walnut, or oat hulls on which the explosive oil is adsorbed, and

stabilizers such as calcium carbonate. A nitro ester commonly used as the filler in detonating cord and in sheet or plastic explosives such as Semtex is pentaerythritol tetranitrate (PETN). Pure PETN is a white powder identifiable by IR spectroscopy, TLC, or GC-MS. When examined by common 'dynamite systems' for TLC, PETN is poorly separated from NG and frequently exhibits streaking on the plate. During screening with a TLC system on silica gel, observation of significant streaking should be followed by reexamination on a 'PETN-specific' TLC system.

Cyclic nitramines such as RDX or cyclotetramethylenetetranitramine (HMX) are widely used in military composites such as Composition B (TNT and RDX) and Composition C-4 (US) or PE-4 (British) and in commercial blasting explosives such as Semtex (a Czech-made mining explosive). HMX is present as a by-product in RDX made by the Bachmann process and has applications in explosives to be used in high-temperature environments. Chemical tests for RDX include the 'J-Acid' and thymol tests. A number of TLC systems for RDX and HMX have been reported. With adequate sample, IR identification of the pure material in a micro-potassium bromide pellet is simple. If a diamond anvil sample holder or microscope attachment is available, excellent spectra of pure samples of milligram size or even of single crystals are easily obtained. When HMX is observed in RDX-based explosives, its concentration may suggest the national origin of the explosive.

Binary explosives are used primarily where only limited quantities are required or when a need for explosives cannot be anticipated. Because they are not considered as explosives until mixed, they can be safely stored and transported. A liquid-liquid type, PLX, was developed for minefield clearing and is composed of nitromethane sensitized with ethylene diamine. Commercial binary explosive products include Astrolite, based on hydrazine and  $\text{NH}_4\text{NO}_3$ . Astrolite was offered in both aluminized and non-aluminized formulations, the former advertised as the most powerful chemical explosive available. A more widely sold binary, Kinestik, combines finely pulverized  $\text{NH}_4\text{NO}_3$  and nitromethane. This combination, offered by several vendors, is the only binary explosive currently available commercially. To ensure complete migration of the liquid into the solid phase, a dye is added to the liquid. When the intact mixed explosive is examined, dye can be extracted from the solid with acetone, examined by TLC or GC-MS, and compared to the dye used by producers of the product. Identification of the  $\text{NH}_4\text{NO}_3$  is straightforward and nitromethane can be identified by low-temperature GC.

Among the newer commercial explosives are cap-sensitive water gel slurries and emulsions. For reasons of safety and economy, they are rapidly replacing dynamite in commercial practice and as a result of this wider availability, are increasingly being encountered in criminal activity. Although both are based on aqueous solutions of  $\text{NH}_4\text{NO}_3$ , in the water gel types a continuous phase of gel is produced with a polysaccharide such as guar gum, which serves both as fuel and gelling agent. Additional oxidizer or fuel, typically flake aluminum, may be dispersed in the gel to produce a slurry and most commercial products are, in fact, slurries. For improved storage stability or density control, co-oxidizers such as calcium nitrate ( $\text{Ca}(\text{NO}_3)_2$ ) may be used. To ensure cap sensitivity, voids or chemical sensitizers are added. The chemical sensitizers are generally nitrate salts of amines such as methylamine nitrate and hexamine nitrate with the sensitizer compound indicative of the producer. Voids used for sensitization are either entrained air or hollow glass microspheres  $<100\text{ }\mu\text{m}$  in diameter.

A suspected water gel is examined on a microscope slide to identify the microspheres used in slurry and emulsion explosives. The gel is directly extracted with methanol, which dissolves the amine salt, and a small amount of  $\text{NH}_4\text{NO}_3$ . TLC on a cellulose plate using a chloroform-methanol-water system separates the sensitizers. The plate is sprayed with ninhydrin and heated to visualize the amine salt. A second spray with diphenylamine in ethanol followed by long-wave UV irradiation visualizes the  $\text{NH}_4\text{NO}_3$  if desired. Alternatively, the methanol extract may be evaporated to near dryness, redissolved in water, and analyzed by IC to identify the sensitizer. An intact sample can also be extracted with water using either a small homogenizer or ultrasonic agitation to disrupt the gel structure. Microspheres float on the surface and are removed for examination by SEM-EDX to characterize the manufacturer. Spot tests and IC identify ammonium, calcium, and nitrate ions in the water extract. Flake aluminum, if present, is identified as described above.

Emulsions are extracted with pentane or hexane, which separates the emulsion and, after concentrating the solution, high-temperature GC is used to identify oil and wax. The residue from the organic extraction is dried and extracted with hot water to dissolve oxidizer(s). Testing of the water extract by spot tests or IC identifies ammonium, potassium, nitrate, and perchlorate ions when potassium perchlorate is used for increased sensitivity. Identification of the emulsifying system has potential for further characterization of emulsions. The residue from the water extract is primarily aluminum

granules, rather than the flake material used in gel/slurry types.

Improvised high explosives vary from the relatively simple to prepare but highly shock-sensitive-triacetone triperoxide or hexamethylene triperoxidediamine to obsolete explosives such as tetranitroaniline. In general, if an adequate collection of reference spectra is available, the pure material is probably best identified by IR spectroscopy as a micro-KBr pellet or by mass spectral techniques.

### Identification of High Explosives in Postblast Debris

Early work on detection of high explosives in debris samples focused on NG and EGDN from dynamite or gelignite-type explosives. Vapor headspace sampling using GC with electron capture detection is highly effective as are ion mobility spectrometry (IMS), GC-MS, or GC-MS/MS. Dynamite has also been identified by purge-and-trap collection on charcoal sorbent. With dynamite, particulates consisting of NC, binder material, adsorbed NG and EGDN, MTN, DEGDN, and ammonium or sodium nitrates often can be observed microscopically in the debris. Their removal permits identification of the components by GC-TEA, IR, GC-MS, LC-MS, TLC, IC, ICP, XRD, or chemical tests.

For plastic explosives, extraction of the debris with an organic solvent dissolves TNT, RDX, HMX, or PETN, which are then identifiable by TLC or a suitable instrumental method. Combinations of two or more high explosives and/or other ingredients such as oil, polyisobutylene, etc., suggest the original type of explosive.

Postblast identification of water gel/slurry explosives is usually based on detection of the characteristic sensitizer compound by swabbing of debris fragments with methanol or methanol-water extract. Purge-and-trap collection of ethylene glycol mononitrate has also been reported. TLC or LC of derivatives with UV absorption or photolysis and electrochemical detection are effective. IC of a water extract for cations may also be used to detect the amine salt. While considerable success has been obtained in postblast detection of water gel/slurry explosives, only limited work has been reported for the detection of emulsion explosives. Initial work indicates some success in postblast recovery of microspheres from detonation of emulsions. Because of their simplicity, and the common materials in their formulations, these explosives pose a significant challenge to postblast detection.

While not widely used, the binaries will similarly be difficult to characterize following an explosion.



An improvised chlorate–nitrobenzene binary explosive formerly used by terrorists in the UK was abandoned because its presence was suggested by the persistent odor of nitrobenzene (NB) both on the clothing of the handlers of the explosive and in the postblast debris. Relatively low levels of NB are easily detectable by vapor sampling and GC with electron capture detection. Commercial nitromethane (NM) binary types will probably exhibit little residual NM postblast, but container fragments may permit detection of the liquid absorbed into the plastic.

### Detection of Taggants

In the aftermath of the mid-air terrorist bombing over Lockerbie, Scotland, in 1988, in which 270 people were killed, it was recognized that the bomb detection equipment available at that time was inadequate. In terms of the trace detection of explosives, it was thought that the most common high explosives had vapor pressures that were so low as to preclude the realistic possibility of vapor detection. A range of more volatile ‘taggant’ materials was therefore proposed as additives to commercially available explosive formulations to ameliorate this situation. EGDN, *ortho*-, and *para*-mononitrotoluene (*o*-MNT and *p*-MNT) and 2,3-dimethyl-2,3-dinitrobutane (DMNB) were proposed as suitable contenders. These could be used both to enhance the detection of explosives prior to detonation and to trace explosive materials to their source postblast. In 1993, 34 countries ratified a convention recommended by the United Nations’ Council of the International Civil Aviation Organization (ICAO) requiring the marking of commercial explosives. This was to come into force in 1997, after which participating countries were given a grace period of 3 years in which unmarked stocks could be used up. There has been opposition to the use of such

taggants on a number of grounds, such as safety, environmental, and utility problems, as well as cost. It should also be recognized that an unknown proportion of terrorist bombings will involve commercially produced explosives. Neither the less taggants, in particular DMNB, are now widely incorporated into commercially produced explosives.

*See also:* **Extraction:** Supercritical Fluid Extraction. **Forensic Sciences:** Gunshot Residues; Thin-Layer Chromatography. **Fourier Transform Techniques. Gas Chromatography:** Mass Spectrometry; Forensic Applications. **Microscopy Applications:** Forensic. **X-Ray Absorption and Diffraction:** X-Ray Absorption; X-Ray Diffraction – Powder.

### Further Reading

- Advances in Forensic Applications of Mass Spectrometry, Ed J Yinon, CRC press 2003, ISBN 0849315220.
- Chemistry of Technology of Explosives, Volume 1, Tadeusz Urbanski, Pergamon press 1963, ISBN 0-08-010238-7.
- Chemistry and Technology of Explosives, Volume 2, Tadeusz Urbanski, Pergamon press 1963, ISBN 0-08-010239-5.
- Chemistry and Technology of Explosives, Volume 3, Tadeusz Urbanski, Pergamon press 1963, ISBN 0-08-010401-0.
- Chemistry and Technology of Explosives, Volume 4, Tadeusz Urbanski, Pergamon press 1963, ISBN 0-08-026206-6.
- Forensic and Environmental Detection of Explosives, Yinon J, John Wiley and Sons Ltd. 1999, ISBN 0471983713.
- Modern Methods and Applications in Analysis of Explosives, Yinon J and Zitrin S, John Wiley and Sons Ltd. 1996, ISBN 0471965626.
- The Chemistry of Explosives 2nd Edition, Akhavan J, Royal Society of Chemistry 2004, ISBN 0854046402.

## Fibers

**N Watson**, Strathclyde University, Glasgow, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Shinichi Suzuki, Yoshiteru Marumo, and Sueshige Seta, pp. 1612–1617, © 1995, Elsevier Ltd.

### Introduction

Fibers provide one of the most common and important type of trace evidence retrieved by crime scene

investigation. They are transferred from one surface to another in any case where contact between two individuals has occurred and have quite high evidential values to link the suspects to the crime scene. In the examination of fibers, there are two distinct areas: fiber identification and fiber comparison. The former gives information even when a suspect’s fiber is not available; when it is available, it is regarded a part of fiber comparison. The latter elucidates the identity of the suspect’s fiber and the fiber from known origin that is transferred and/or cross-transferred between each other at the crime scene. These

An improvised chlorate–nitrobenzene binary explosive formerly used by terrorists in the UK was abandoned because its presence was suggested by the persistent odor of nitrobenzene (NB) both on the clothing of the handlers of the explosive and in the postblast debris. Relatively low levels of NB are easily detectable by vapor sampling and GC with electron capture detection. Commercial nitromethane (NM) binary types will probably exhibit little residual NM postblast, but container fragments may permit detection of the liquid absorbed into the plastic.

### Detection of Taggants

In the aftermath of the mid-air terrorist bombing over Lockerbie, Scotland, in 1988, in which 270 people were killed, it was recognized that the bomb detection equipment available at that time was inadequate. In terms of the trace detection of explosives, it was thought that the most common high explosives had vapor pressures that were so low as to preclude the realistic possibility of vapor detection. A range of more volatile ‘taggant’ materials was therefore proposed as additives to commercially available explosive formulations to ameliorate this situation. EGDN, *ortho*-, and *para*-mononitrotoluene (*o*-MNT and *p*-MNT) and 2,3-dimethyl-2,3-dinitrobutane (DMNB) were proposed as suitable contenders. These could be used both to enhance the detection of explosives prior to detonation and to trace explosive materials to their source postblast. In 1993, 34 countries ratified a convention recommended by the United Nations’ Council of the International Civil Aviation Organization (ICAO) requiring the marking of commercial explosives. This was to come into force in 1997, after which participating countries were given a grace period of 3 years in which unmarked stocks could be used up. There has been opposition to the use of such

taggants on a number of grounds, such as safety, environmental, and utility problems, as well as cost. It should also be recognized that an unknown proportion of terrorist bombings will involve commercially produced explosives. Neither the less taggants, in particular DMNB, are now widely incorporated into commercially produced explosives.

*See also:* **Extraction:** Supercritical Fluid Extraction. **Forensic Sciences:** Gunshot Residues; Thin-Layer Chromatography. **Fourier Transform Techniques. Gas Chromatography:** Mass Spectrometry; Forensic Applications. **Microscopy Applications:** Forensic. **X-Ray Absorption and Diffraction:** X-Ray Absorption; X-Ray Diffraction – Powder.

### Further Reading

- Advances in Forensic Applications of Mass Spectrometry, Ed J Yinon, CRC press 2003, ISBN 0849315220.
- Chemistry of Technology of Explosives, Volume 1, Tadeusz Urbanski, Pergamon press 1963, ISBN 0-08-010238-7.
- Chemistry and Technology of Explosives, Volume 2, Tadeusz Urbanski, Pergamon press 1963, ISBN 0-08-010239-5.
- Chemistry and Technology of Explosives, Volume 3, Tadeusz Urbanski, Pergamon press 1963, ISBN 0-08-010401-0.
- Chemistry and Technology of Explosives, Volume 4, Tadeusz Urbanski, Pergamon press 1963, ISBN 0-08-026206-6.
- Forensic and Environmental Detection of Explosives, Yinon J, John Wiley and Sons Ltd. 1999, ISBN 0471983713.
- Modern Methods and Applications in Analysis of Explosives, Yinon J and Zitrin S, John Wiley and Sons Ltd. 1996, ISBN 0471965626.
- The Chemistry of Explosives 2nd Edition, Akhavan J, Royal Society of Chemistry 2004, ISBN 0854046402.

## Fibers

**N Watson**, Strathclyde University, Glasgow, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Shinichi Suzuki, Yoshiteru Marumo, and Sueshige Seta, pp. 1612–1617, © 1995, Elsevier Ltd.

### Introduction

Fibers provide one of the most common and important type of trace evidence retrieved by crime scene

investigation. They are transferred from one surface to another in any case where contact between two individuals has occurred and have quite high evidential values to link the suspects to the crime scene. In the examination of fibers, there are two distinct areas: fiber identification and fiber comparison. The former gives information even when a suspect’s fiber is not available; when it is available, it is regarded a part of fiber comparison. The latter elucidates the identity of the suspect’s fiber and the fiber from known origin that is transferred and/or cross-transferred between each other at the crime scene. These



examinations include several kinds of analyses such as morphological and fluorescence observation by several kinds of microscopes and investigation of optical and physical properties. Some instrumental analyses for color and fluorescence are evaluated by microspectrophotometer (MSP) and microspectrofluorimeter; inorganic elements in fibers are investigated by scanning electron microscopy with energy dispersive X-ray microanalysis (SEM-EDX); Fourier transform infrared spectrometry (FTIR)/microscopy and pyrolysis gas chromatography–mass spectrometry (Pyr-GC–MS) are widely utilized for the analysis of the polymer composition of man-made fibers.

## Fibers Encountered at the Scene of the Crime

Fibers encountered at the scene of the crime are retrieved directly or by using adhesive tapes from the surface of several kinds of materials such as clothes, floor, carpets, bed, and inside of a car. They are classified into two groups: natural fibers and man-made fibers. In natural fibers, there are animal fibers, from wool (sheep), goat hair (common goat, mohair, and cashmere), llama (Alpaca), camel, and rabbit hair, and various kinds of fibers obtained from other animals; they are frequently encountered in criminal laboratories. There also are vegetable fibers in this category. The most common is cotton, and others are hemp, linen, and jute. Another natural fiber is the mineral fiber asbestos. Man-made fibers may be subdivided into two groups: one is regenerated fiber, the other is synthetic fiber. In regenerated fibers, there are three generic classes: acetate, triacetate, and rayon. In synthetic fibers, the major generic classes are nylon, acrylic, and polyester, and there are various other generic classes, such as aramide, nitril, and olefin. Most of these generic classes have a number of subtypes, for example, common forms of nylon are nylon 6, nylon 6,6, and nylon 1,1. The classification of fibers and generic classes of man-made fibers is shown in Table 1.

## Fiber Recovery

Adhesive tape can be used to collect a sample of the surface debris from clothing or other surfaces. The tape is pressed onto a surface and then placed onto a clear plastic sheet. Debris from the sampled surface is thus collected and then securely packaged. A low-power microscope is used to locate fibers that are of interest. Such fibers can be removed by forceps from the tape lift and transferred to a microscope slide mounting for further examination. Some laboratories have adopted one-to-one, or 1:1, tape lifting. This

**Table 1** Classification of fibers

<i>Natural fiber</i>	<i>Man-made fibers</i>
<i>Vegetable fibers</i>	Acetate
Seed fibers	Acrylic
Cotton, kapok, etc.	Acrylonitrile
Bast fibers	Acrylonitrile/methylacrylate/ methyl vinyl pyridine
Flax, hemp, jute, ramie, etc.	Acrylonitrile/vinyl acetate
Leaf fibers	Acrylonitrile/vinyl acetate/ methyl vinyl pyridine
Sisal, Abaca, etc.	Acrylonitrile/methyl methacrylate
Fruit fibers	Acrylonitrile/any other polymers
Coir, etc.	Anidex
<i>Animal fibers</i>	Aramide
Wool	Nomex, Kevlar
Other animal fibers	Azlon
Alpaca, Mohair, Cashmere, Angola, camel, etc.	Carbon
Silk	Chlorofiber
Cultivated	Fluorofiber
Tussah	Glass
<i>Mineral fibers</i>	Modacrylic
Asbestos	Novoloid
	Nylon
	Nylon 6, nylon 6,6, nylon 1,1, etc.
	Nytril
	Olefin
	Polyethylene, polypropylene, etc.
	Polyester
	PCDT, PET, PBT, etc.
	Rayon
	Viscose, etc.
	Poly(vinyl alcohol)
	Poly(vinyl chloride)
	Poly(vinylidene chloride)
	Triacetate
	Vinyl

consists of taking single tape lifts from each part of the entire surface area of a suspected contact. The positions are recorded so that there is a one-to-one mapping of the surface area. This has the advantage of preserving precise information of the place where recovered fibers were positioned, a more rigorous sampling of the surface debris, and ensuring that the tapes are not overloaded and causing difficulties in the searching and recovery steps in the fiber examination process.

The disadvantage of the 1:1 tape lifting process of surface debris recovery is that a large number of tapes is generated. The searching of tape lifts is a protracted exercise demanding a great deal of examiner's time and slowing the entire examination process. A potential solution is the use of automated devices to screen tape lifts and search for certain specified target fiber types.

There have been five such devices described. They are the Foster and Freeman Fiber Finder Fx5 made by Foster and Freeman Ltd, the Fiber Finder and Maxcan devices made by Cox Analytical Systems of Sweden, the Q550ffi by Leica Vertrieb GmbH, and the Lucia Fiber Finder from Laboratory Imaging, Czech Republic. They all operate by scanning tapings carrying the fibers using a charge coupled device camera in the case of the Fiber Finder, the Q550ffi, and the Lucia fiber Finder and a Line Scanner in the case of the remaining two. Their operation is in two stages. First, a tape containing reference fibers is prepared and scanned. These reference fibers having been selected as a potential source of transferred fibers to be found among the tape lifts of the evidence. The color of the reference fibers is recorded and used as the matching criterion. The tapes containing the fibers to be searched are then scanned and the positions of those fibers matching the reference fibers are recorded. The operator can inspect images of the matching fibers to eliminate obvious mismatches based on their morphology.

In operation the devices are effective in the hands of an experienced operator but the preliminary step of scanning the reference fibers means that they are most efficiently used where a large number of tapes have to be searched for the same set of fibers. They would therefore seem ideally suited for use in conjunction with 1:1 taping. It seems likely that this is one area of fiber examinations that will continue to be developed.

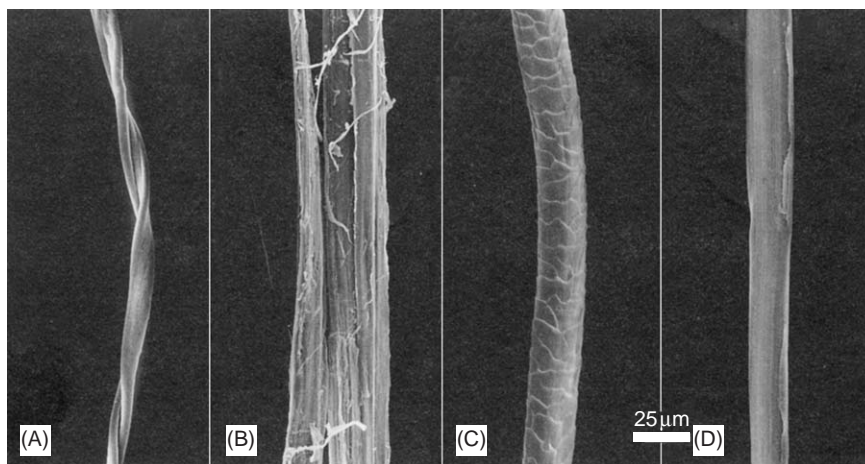
## Identification of Types

There are two purposes in forensic fiber identification: one is the determination of fiber type, the other

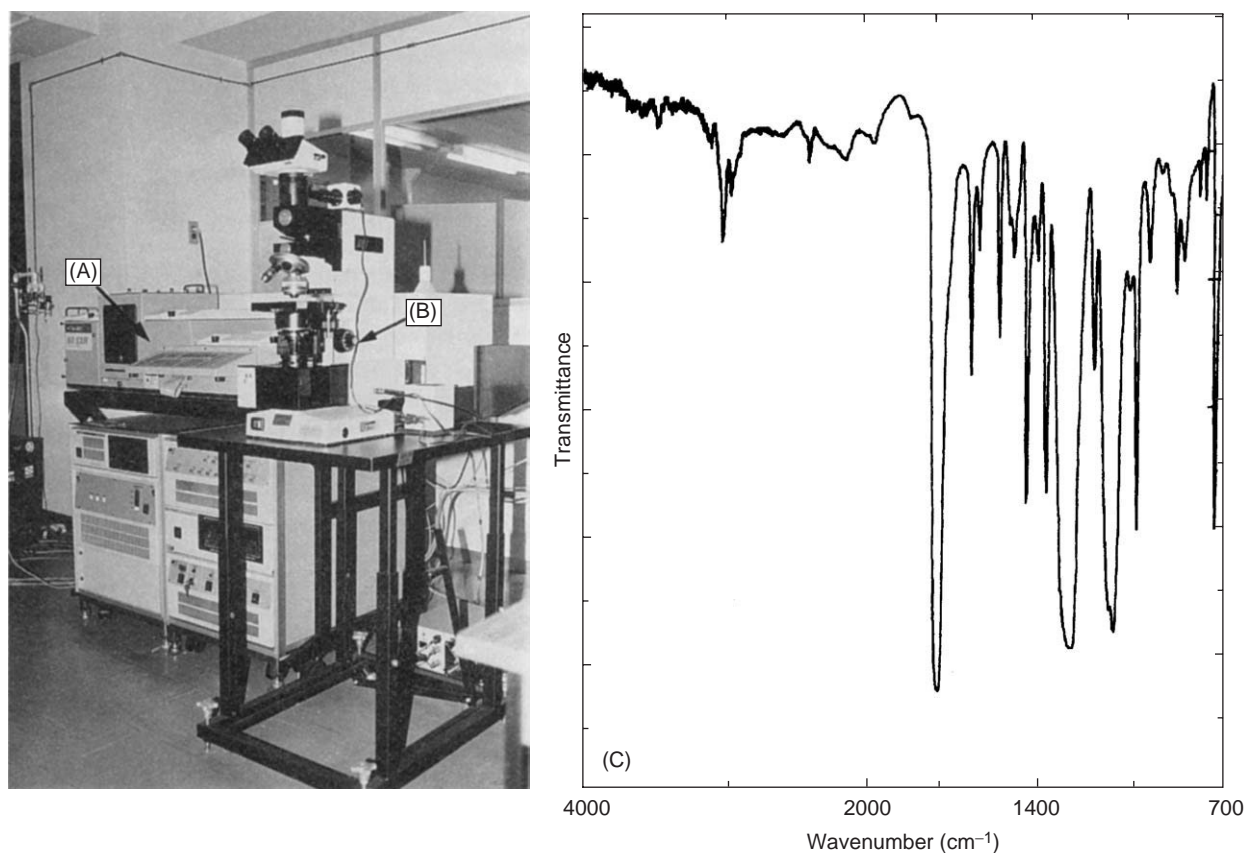
is determination of other characteristic properties of the fibers. In the first step of the fiber identification, a microscopic examination and investigation of optical properties are conducted. The microscopic appearance of the typical natural fibers cotton, hemp, wool, and silk are shown in **Figure 1**.

The diameter of the fibers, generic classes of natural fibers, the presence of delustrants and their distribution, birefringence, and refractive indices are also examined microscopically. Fluorescence and colors are measured by MSPs and microspectrofluorimeters under nondestructive conditions and color coordinates are calculated by an attached computer to evaluate the color objectively. In addition, if sufficient sample material is available, infrared (IR) spectroscopy (to clarify the polymer components), a staining test, hot stage microscopy for measuring melting points of fibers, and a solubility test for identifying generic classes may be carried out. But these tests, apart from IR spectrometry, are destructive and require large amounts of samples, so they are used only occasionally, and recently they have been replaced by other instrumental analyses. Furthermore, dyes used in fibers are identified, after their extraction from the fibers, by thin-layer chromatography (TLC) or liquid chromatography (LC).

In natural fiber identification for animal fibers, careful microscopic observation to assess the following components is important: one is the medulla, its structure, continuity, cuticular thickness, and the distribution and color of pigments; other important features are those of scales, such as its pattern, shape, margins, and changes. For vegetable fiber identification, polarizing microscopic analysis is also useful; for example, fibers that have a Z-type spiral are differentiated from those with S-type spirals.



**Figure 1** Scanning electron photomicrographs of (A) cotton, (B) hemp, (C) wool, (D) silk.



**Figure 2** (A) Fourier transform IR spectrometer with (B) Spectratech IR plan microscope attachment. (C) Infrared spectrum from a single polyester fiber measured by using the Spectratech IR plan microscope.

For synthetic fibers, identification only by microscopic observation and measurements of optical properties is insufficient. In addition to these examinations, analyses by FTIR/microscopy and Pyr-GC-MS are indispensable. The combination of FTIR spectrometry with a microscope makes it possible to measure IR spectra of a single fiber nondestructively and to identify the polymer subtypes, thus making this technique very valuable for fiber identification. An FTIR spectrometer with microscope attachment and the IR spectrum of a single fiber of polyester (polyethylene terephthalate (PET)) obtained by using this device are shown in **Figure 2**. Another application of FTIR microspectroscopy is to discriminate between fibers where chemical modifications have been used to give different properties. For example, acrylic fibers are defined as being composed of not less than 85% acrylonitrile polymer. The balance of the mass of the fiber can be composed of a copolymer used to enhance some attribute of the finished fiber for its end use. Acrylic fibers possessing less than 85% of acrylonitrile polymer are termed modacrylic fibers. Some examples of modacrylic fibers cannot be distinguished from acrylic fibers by microscopic examination (**Table 2**).

**Table 2** Acrylic fibers and variations

<i>Fiber type</i>	<i>Mass of acrylonitrile (%)</i>
Acrylic fiber	> 85
Modacrylic (BISFA definition)	50 > 85
Modacrylic (US definition)	35 < 85

Modacrylic fibers are composed of copolymers of acrylonitrile and a comonomer containing a halogen. The comonomer is usually either vinyl chloride ( $\text{CH}_2=\text{CHCl}$ ) or vinylidene chloride ( $\text{CH}_2=\text{CCl}_2$ ) but vinyl bromide and vinylidene bromide have also been reported to be used in fibers. The halogen groups give flame-retardant qualities to the fiber and the acrylonitrile components give thermal stability, light fastness, and strength. Also, other additives can be used to extend dyeability to different types of dye. The proportions of the monomer types can be varied by the manufacturer, even within the same trade name, according to the application intended for the fiber.

The presence of such additives and modifications can be detected in specimens of fiber fragments by making a detailed observation of the FTIR spectra. Such analyses can be important in forensic analyses

to determine the fiber type. Different formulations of the modacrylic can still carry the same trade name so that discrimination between fibers sold under the same trade name may be possible in some circumstances. The fiber environment is therefore a complex one, as illustrated by the modacrylics, because there are many additives and chemical modifications that may be used. In addition, many fibers that are no longer produced remain in use and are therefore potential sources of trace evidence.

## Dye Extraction and Analysis

For TLC and LC, dyes in fibers should be extracted by suitable solvents. It is possible to extract dyes from a single fiber. As an extracting solvent, a mixture of pyridine and water (4:3 by volume) is preferred for the majority of samples, but some modification may be required, and some dyes such as sulfur dyes on cotton are difficult to extract. A microcapillary glass tube is used to contain a single sample fiber; suitable extracting solvent is added, the tube is sealed at both ends, and is incubated usually at 100°C for a length of time dependent on the solvent and the dyes. The required minimum lengths of sample fiber for subsequent analysis by TLC or LC depends on many factors, such as the dyed fiber properties, the chemical class of the dyes, and the intensity of the colors. Extraction solvents and conditions for various fiber types are summarized in Table 3.

TLC is used for qualitative analysis; its reproducibility is relatively poor. For the analysis of dyes by TLC, extracts are spotted on areas as small as possible on the silica gel precoated plate and developed by a suitable developing solvent. Selection of developing systems depends on the chemical properties of

the dyes. Discrimination of the dyes is carried out by the comparison of  $R_F$  values of the components of the dyestuffs and the color of the separated spots on the plate, but such color comparisons are subjective. Extraction of dyes from the known fiber and suspect's fiber and their analysis by TLC should be conducted simultaneously under exactly the same conditions.

For LC, reversed-phase conditions are widely used for the separation of dyes, but in the case of basic dyes (cationic dyes used in acrylics), a silica column is utilized. Mobile phases also depend on the chemical properties of the dyes and ion-pair chromatography is widely used. Detection is achieved by measuring the absorbance in the visible region as well as at 254 nm. A photodiode array (PDA) detector makes possible simultaneous detection at several wavelengths and spectral analysis of the chromatographic peaks. If the absorption spectrum for the peak for each separated dye component is available, the color of the dyes can be assessed objectively and the combination of retention times and peak ratios obtained from the chromatogram; the spectral data make it possible to conduct an accurate and detailed comparison of the dyes.

## Color Matching

The color of fibers is one of the most significant characteristics in fiber comparison; color matching including fluorescence is inevitable for forensic fiber identification. Measurement of the color of fibers is conducted initially by microspectrophotometry. The MSP provides a great increase of discriminating power. Without this instrument, many false positive errors occur in fiber comparison. It is a convenient, nondestructive method that requires only a small amount of sample. Microspectrophotometry enables

**Table 3** Solvents and conditions for fiber dye extraction<sup>a</sup>

Fiber type	1	2	3	4
Cotton	Pyridine:water (4:3)	1.5% aq. NaOH	DMF:water (3:11)	DMF
Wool	Pyridine:water (4:3)	2% aq. oxalic acid (15 min at 90°C followed by pyridine: water (4:3))	1,2-Diaminoethane (90°C, 1 h)	
Acetate	Pyridine:water (4:3)	Methylene chloride		
Acrylic	Pyridine:water (4:3)	Formic acid:water (1:1)	Formic acid	DMF:water (3:1)
Nylon	Pyridine:water (4:3)	Formamide (135°C)	DMF:water (3:1)	
Polyester	Chlorobenzene:acetic acid (1:1)	Chlorobenzene (130°C)	Pyridine:water (4:3)	DMF:formic acid (1:1)
Rayon	Pyridine:water (4:3)	1.5% aq. NaOH	DMF:water (3:1)	DMF
Triacetate	Pyridine:water (4:3)	Methyl propyl ketone		

<sup>a</sup>Extractions are conducted at 100°C in sealed tubes unless otherwise mentioned. Solvent ratios are by volume.

DMF=dimethyl formamide. (From Barry D and Gaudette BS (1988) The forensic aspects of textile fiber examination. *Forensic Science Handbook*, vol. 2, pp. 209–272.)



an evaluation of colors to be made objectively using UV-visible transmittance profiles. It is the 'only' technique available for comparison of colors when the sample fiber is too short or its color is too pale for TLC or LC. Furthermore, the attached computer can convert the color to numerical values of complementary chromaticity coordinates. Variations of the absorption spectra or complementary chromaticity coordinates are observed in the comparison of fibers, especially the natural fibers cotton and wool; the variation is due to the uneven uptake of dyes or the irregular cross-sectional shape of these fibers. It is essential, therefore, for forensic fiber comparison to conduct an adequate number of repeat measurements of control fibers to cover the range of variation on the absorption spectra or complementary chromaticity coordinates, and then to clarify whether the results of the measurement of suspect's fibers are within this range. The measured complementary chromaticity coordinates of control and suspect's fibers should be compared mathematically and statistically. An ellipse model for the variation of these complementary chromaticity coordinate values is proposed; the ellipse is larger for natural fibers, such as wool and cotton, than for man-made fibers. This arises from the greater variability in dye uptake.

The introduction of improved MSPs capable of reading down to 240 nm, in the UV spectral range, is enabling improved discrimination among indigo and black dyed fibers. A further development of the MSP instrumentation has been the introduction of PDA devices, which conduct scanning much faster than the older instruments. Indeed, the enhanced usefulness of MSP has led to the concept of a fiber examination workstation that employs a PDA mounted on a microscope incorporating both polarized light and fluorescence capabilities as well.

Blue and black denim clothing is one of the most common types of clothing. With so many potential sources of such fiber fragments their significance as evidence is often regarded as being too small to be useful. In studies using reference collections of identified fibers it has been demonstrated that UV-MSP is capable of resolving different groups within this category of fibers. Dyestuffs commonly used on cotton fibers are the reactive dyes, used for bright colors and which possess good color fastness, and sulfur dyes used for gray, brown, and olive-green colors. The reactive dye type is the most important class of dye used with cotton and its use is probably increasing because of environmental concerns over sulfur dyes. Other dye types include direct dyes, which can be used for many colors and are inexpensive but have poor color fastness, and the vat dyes. Vat dyes are often used for industrial applications due to their

resilience. Vat dyed fibers are usually dark blue-gray and sometimes yellow-gray.

The different dye classes can be determined by their spectra in most cases. The sulfur dyes are the most easily determined and the reactive dyes have characteristic spectra in the UV region, below 300 nm. It has been possible to identify several subgroups within the reactive dyes. The direct dyes and vat dyes produce spectra that can be readily distinguished from those of sulfur dyes and almost all of the reactive dyes.

A common type of cotton garment is blue denim clothing composed of indigo dye cotton textiles and the significance of the occurrence of indigo cotton fibers is considerably reduced due to the widespread prevalence of this material. However, in experimental comparisons of indigo dye and seven indigo derivatives, used on cotton fibers, a close examination of the spectra produced distinguished the eight different dyes by spectra produced from a single fiber specimen. Thus, even within blue dyed cotton fibers, usually regarded as a homogenous group, UV-MSP is a potentially discriminating measure. It is particularly significant that the dyes commonly used in cotton fibers used in denim textiles are reactive dyes. These dyes are less easily extracted by solvents and therefore less readily analyzed by dye separation techniques such as TLC. This further emphasizes the advantage of the nondestructive approach of microspectrophotometry.

If the sample amounts are adequate, dyestuffs may be compared by TLC or LC. The techniques, however, are destructive. For TLC of the dye components, the developing solvents depend on the chemical composition of the dyes. **Table 4** shows the TLC conditions recommended for separation of various dye classes.

Recently, LC techniques were introduced for the evaluation of dyestuffs. They have much higher sensitivity than TLC, the choice of conditions, mobile phase, and columns again depends on the chemical properties of the dyes, and an extraction procedure is also required. The results obtained by LC can be used for quantitative data processing and, by comparison of retention times, peak height or area ratios, and trace impurity detection by its higher sensitivity than TLC, more detailed discrimination of fiber samples is possible. Multi-wavelength detection using a PDA detector is also useful, as described above.

## Analysis for Metals

Metals analysis is usually conducted by SEM-EDX. Fibers sometimes contain inorganic elements as

tagging elements added by manufacturers to differentiate their product from others (see below). Metals from the dyes are also sometimes detected. For example, when the fiber, such as an amide fiber, wool, is dyed by acid mordant dyes, chromium is used in the coloration process and it is detected when this kind of fiber is analyzed. The results of energy dispersive X-ray microanalysis of wool fibers dyed by an acid dye and an acid-mordant dye are shown in Figure 3.

Other metals such as iron, copper, cobalt, and nickel are reported to be detected in metal-containing reactive dyes. In some cases, the fiber itself contains

several specific elements; for example, the presence of chlorine and bromine in modacrylic fibers can help to discriminate these fibers from the acrylic fibers.

## Additives and Contaminants

There are various types of additives in fiber samples. They are fiber dyes, fluorescent brighteners, delustrants, flammability retardants, and tagging metals.

There are about 10 000 dyes and pigments and they are used to produce characteristic colorations of fibers and textiles. They are divided into some generic classes and there is a wide range of colors in each generic class. These dyes are used as a single component or as a mixture; for example, black is produced by a mixture of yellow and blue components. The dyes are distinguished by class generic name and color index (CI) name. Color fibers are elucidated by MSP and dyes are also analyzed by TLC and LC for the purpose of fiber comparison by variety and complex composition of dye especially black dyes, as described above. Fluorescent brighteners are utilized to prevent fibers or textiles from bleaching even after allowed for a long period under sunlight irradiation. They absorb UV radiation from the sunlight and give off a blue fluorescence emission. Some of their main structures involve stilbene and coumarin.

Delustrants are added to some synthetic fibers (rayon, nylon, polyesters, and acrylics), which have a strong luster in order to control the opacity and luster. Titanium dioxide particles are widely used for this purpose; the delustrant properties are controlled by the concentration and particular size of the titanium dioxide. These particles can be observed microscopically and are useful for fiber comparison.

There are also some specialized fibers, such as flame-retardant fibers. They are utilized, for example, in hotel blankets, subway seats, and in aircraft. There are two methods of introducing flame retardancy. One is to blend in flame-retardant agents during the making of the fibers. Most flame-retardant agents are organophosphorus compounds. The other is to make inherently flame-retardant synthetic fibers. Some of these fibers are composed of poly(benzimidazole) and an aromatic copolyamide.

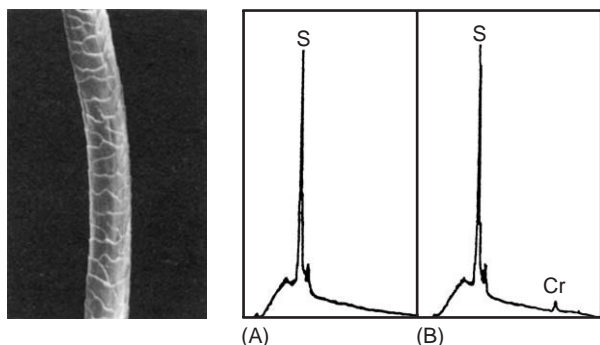
Trace amounts of inorganic elements, such as indium and samarium, are added to fibers by their manufacturers to distinguish their products from others; these elements are useful for forensic fiber comparison. The determination of these elements is carried out by SEM-EDX.

Contamination, which interferes in forensic fiber identification, is very frequent. Contaminants, such as blood and oil, interfere in the color evaluation of

**Table 4** Thin-layer chromatography developing solvents for dye analysis

<i>Wool-acid or metallized dyes</i>	
n-Butanol, acetone, water, ammonia	5:5:1:2 v/v
Pyridine, amyl alcohol, 10% ammonia	4:4:3
<i>Cotton and viscose-direct or reactive dyes</i>	
n-Butanol, acetone, water, ammonia	5:5:1:2
Methanol, amyl alcohol, water	5:5:2
n-Butanol, ethanol, ammonia, pyridine, water	8:3:4:4:3
<i>Acrylic-basic dyes</i>	
Chloroform, methyl ethyl ketone, acetic acid, formic acid	8:6:1:1
n-Butanol, acetic acid, water (upper phase)	4:1:5
n-Butanol, acetone, water, ammonia	5:5:1:2
<i>Polyester-disperse dyes</i>	
Chloroform, ethyl acetate, ethanol	7:2:1
n-Hexane, ethyl acetate, acetone	5:4:1
Toluene, methanol, acetone	20:2:1
Toluene, pyridine	4:1
<i>Polyamide-acid dyes</i>	
n-Butanol, acetic acid, water (upper phase)	2:1:5
n-Butanol, acetic acid, ammonia, pyridine	4:1:3:2

From Grieve MC (1990) *Fibers and their examination in forensic science. Forensic Science Progress* 4: 41–125.



**Figure 3** Scanning electron photomicrograph of a wool single fiber and the energy dispersive X-ray spectra of a wool dyed (A) with acid and (B) with metallized acid dye.



fibers by microspectrophotometry; laundering also decreases the fluorescence properties of fibers and bleaching agents may cause false results. Oils and other organic substances also interfere in the measurement of IR spectra and change the chromatographic patterns obtained by Pyr-GC and cause the incorrect interpretation of the polymer composition.

## Other Methods

Microspectrofluorimetry is used to measure the fluorescence of fibers and to code it using the modification of the CIE (Commission International d'Eclairage) system. Fluorescence emission spectra are useful for the differentiation of fluorescent fibers. For excitation four spectral regions (UV, violet, blue, and green) are recommended for fiber comparison. The combination of Pyr-GC-MS makes it possible to obtain chromatographic and mass spectral data simultaneously. The Curie-point pyrolyzer enables the pyrolysis temperature of the fiber to be obtained in a very short time (200 ms) and this temperature is strictly controlled by the specific composition of the foil. For these reasons, high reproducibility of the pyrolysis of the fiber is obtained. Capillary gas chromatography is used for the separation of the volatile compounds produced by pyrolysis. The use of a computerized data processing system makes the analysis of the polymer components much easier and a detailed comparison of the components of man-made fibers is possible.

X-ray diffraction and X-ray fluorescence (XRF) spectrometry have been utilized in the differentiation of man-made fibers, but the sample amounts required are large, so that some modifications for single fiber analysis are required. Total reflection XRF has many advantages in this respect.

Though methods for identification of man-made fibers utilizing differential scanning calorimetry, differential thermal analysis, and thermogravimetric analysis have been reported, the sample amounts required are too large for actual case studies, so these methods are rarely used in forensic identification and are only adopted in very specific cases.

Another potential development in fiber analysis techniques has been the use of Raman spectroscopy for dye analysis. Electromagnetic radiation can interact with molecules so that it can be reflected, absorbed, or scattered. The light that is scattered can either be more energetic than the incident light, in which case it is called Stokes scattering, or of a lower energy, i.e., anti-Stokes scattering. This effect is called the Raman effect and is named after C.V. Raman, who was the first to observe this effect.

This type of spectroscopy offers a potentially highly discriminating and nondestructive method of analyzing fiber dyes. Its use has been curtailed by the lack of sensitivity that makes it unsuitable for application to recovered fiber fragments. However, certain refinements have been reported that offer the increased sensitivity required for this specialist application. These are resonance Raman scattering (RR) and surface enhanced Raman scattering (SERS). A further refinement is a combination of the RR and SERS developments, known as surface enhanced resonance Raman scattering. Studies using Raman spectroscopy have demonstrated the possibility of obtaining spectral information from dye pigments in polyacrylic and cellulose viscose fibers. The use of this approach for fiber dye analysis offers an alternative spectroscopic technique that can be applied to the fiber dyes *in situ* and so brings all the advantages of sequential testing using nondestructive techniques.

## The Significance of Fiber Evidence

The reporting of forensic fiber examinations has been contrasted with the numerical evaluations of the likelihood of a coincidental match commonly used for DNA analyses. One difficulty is that the total numbers of items of clothing that are being currently used and worn that possess any given fiber and dye combination are difficult to determine so that the 'population' of potential sources is unknown. Some workers therefore feel that the significance of fiber matches is consistently underestimated and as a result the overall value of fiber examinations has not been sufficiently recognized.

These impressions of underestimated significance are not supported by experimental evidence produced by exercises that set out to look for coincidental matches. In one such exercise the surface debris from a single evidence item from each of 20 different cases was compared. The cases were unconnected, except for their submission to the same forensic examiner, and the evidence items consisted of items of clothing. A total of 2083 fibers were collected from 20 evidence items and were compared using brightfield microscopy, polarized light microscopy, fluorescence microscopy, and MSP. Comparisons between fibers coming from the same garment that were indistinguishable were not counted and nor were colorless fibers and denim fibers. In the resulting comparisons no two fibers were found that could not be distinguished by the microscopy or analyses used.

The value of fiber examinations in forensic science in comparison to DNA analysis has been the subject

of comment and debate. An improvement in the instrumentation available, and its wider usage, is leading to greater discrimination potentials. The sheer complexity and diversity of the fiber environment poses a challenge to the statistical evaluation of fiber evidence but empirical studies of fiber frequencies consistently demonstrate that fiber matches can be highly significant, especially if more than one fiber type is involved. The speed of examinations, and hence cost-effectiveness, of fiber examinations can be greatly enhanced by the use of the improved instrumentation and its integration into the examination procedure.

*See also:* **Asbestos.** **Color Measurement.** **Forensic Sciences:** Thin-Layer Chromatography. **Gas Chromatography:** Pyrolysis; Mass Spectrometry; Fourier Transform Infrared Spectroscopy. **Microscopy Applications:** Forensic. **Spectrophotometry:** Diode Array. **Textiles:** Natural; Synthetic. **X-Ray Absorption and Diffraction:** X-Ray Diffraction – Powder. **X-Ray Fluorescence and Emission:** X-Ray Fluorescence Theory; Energy Dispersive X-Ray Fluorescence; Total Reflection X-Ray Fluorescence.

## Further Reading

- Biermann TW (1999). Fibre finder systems. In: Robertson J (ed.) *Forensic Examination of Fibres*. London: Taylor and Francis.
- BISFA; The International Bureau for the Standardisation of Man-Made Fibres (2000) *Terminology of Man-Made Fibres* (available on the internet at <http://www.bisfa.org/termino/termino.htm>).
- De Forest PR, Gaensslen RE, and Lee HC (1983) *Forensic Science, An Introduction to Criminalistics*. New York: McGraw-Hill.
- Grieve MC (1990) Fibres and their examination in forensic science. *Forensic Science Progress* 4: 44–125.
- Grieve MC (2000) 40 Years of fibre examinations in forensic science. *Science & Justice* 40(2): 93–99.
- Grieve MC (2000) A survey on the evidential value of fibres and on the interpretation of the findings in fibre cases. Part 1. *Science & Justice* 40(3): 189–200.
- Houck MM (1999) Statistics and trace evidence: The tyranny of numbers. *Forensic Science Communications* 1(1) (available on the internet at [www.fbi.gov](http://www.fbi.gov)).
- Houck MM (2003) Inter-comparison of unrelated fibre evidence. *Forensic Science International* 135: 146–149.
- Jochem G and Lehnert RJ (2002) On the potential of Raman microscopy for the forensic analysis of coloured textile fibres. *Science & Justice* 42(4): 215–221.
- Jochem G and Lehnert RJB (2002) On the potential of Raman microscopy for the forensic analysis of coloured textile fibres. *Science & Justice* 42(4): 215–221.
- Langdon S, Maynard P, Robertson J, and Roux C (2003) An evaluation of the Maxcam fibre finder version 3.3 on cotton fibres. *Forensic Science International* 135: 137–145.
- Moenssens AA, Inbau FE, and Starrs JE (1986) *Scientific Evidence in Criminal Cases*. New York: The Foundation Press.
- Saferstain R (1988) *Forensic Science Handbook*. New Jersey: Prentice Hall.
- Saferstain R (1990) *Criminalistics, An Introduction to Forensic Science*. New Jersey: Prentice Hall.
- Suzuki S, Suzuki Y, Ohta H, Sugita R, and Marumo Y (2001) Microspectrophotometric discrimination of single fibres dyed by indigo and its derivatives using ultraviolet–visible transmittance spectra. *Science & Justice* 41(2): 107–111.

## Fingerprint Techniques

**C Lennard**, Australian Federal Police, Canberra, NSW, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Physical evidence at a crime scene may be present in an infinite variety of forms and materials. Some of the trace materials and marks left by the offender may be of limited evidential value, while others may be of significant benefit to the investigation. Several forms of evidence can permit the identification of an individual (e.g., fingermarks, dental structure, and

genetic code), but the patterns formed by the friction ridges of the fingers and palms are arguably the most useful and general proof of identity. These patterns are unique, immutable, universal, easy to classify, and leave marks on any smooth object handled with the bare hands.

Smooth objects that have been touched with the bare hands will retain latent impressions, composed of natural secretions and contaminants, of the corresponding ridge patterns on the fingers and palms. In criminal investigations, the detection and recording of a fingermark on a particular item can serve as proof that the item was touched by a specific person. It is for this reason that physical and chemical techniques, capable of developing latent fingermarks on

of comment and debate. An improvement in the instrumentation available, and its wider usage, is leading to greater discrimination potentials. The sheer complexity and diversity of the fiber environment poses a challenge to the statistical evaluation of fiber evidence but empirical studies of fiber frequencies consistently demonstrate that fiber matches can be highly significant, especially if more than one fiber type is involved. The speed of examinations, and hence cost-effectiveness, of fiber examinations can be greatly enhanced by the use of the improved instrumentation and its integration into the examination procedure.

*See also:* **Asbestos.** **Color Measurement.** **Forensic Sciences:** Thin-Layer Chromatography. **Gas Chromatography:** Pyrolysis; Mass Spectrometry; Fourier Transform Infrared Spectroscopy. **Microscopy Applications:** Forensic. **Spectrophotometry:** Diode Array. **Textiles:** Natural; Synthetic. **X-Ray Absorption and Diffraction:** X-Ray Diffraction – Powder. **X-Ray Fluorescence and Emission:** X-Ray Fluorescence Theory; Energy Dispersive X-Ray Fluorescence; Total Reflection X-Ray Fluorescence.

## Further Reading

- Biermann TW (1999). Fibre finder systems. In: Robertson J (ed.) *Forensic Examination of Fibres*. London: Taylor and Francis.
- BISFA; The International Bureau for the Standardisation of Man-Made Fibres (2000) *Terminology of Man-Made Fibres* (available on the internet at <http://www.bisfa.org/termino/termino.htm>).
- De Forest PR, Gaensslen RE, and Lee HC (1983) *Forensic Science, An Introduction to Criminalistics*. New York: McGraw-Hill.
- Grieve MC (1990) Fibres and their examination in forensic science. *Forensic Science Progress* 4: 44–125.
- Grieve MC (2000) 40 Years of fibre examinations in forensic science. *Science & Justice* 40(2): 93–99.
- Grieve MC (2000) A survey on the evidential value of fibres and on the interpretation of the findings in fibre cases. Part 1. *Science & Justice* 40(3): 189–200.
- Houck MM (1999) Statistics and trace evidence: The tyranny of numbers. *Forensic Science Communications* 1(1) (available on the internet at [www.fbi.gov](http://www.fbi.gov)).
- Houck MM (2003) Inter-comparison of unrelated fibre evidence. *Forensic Science International* 135: 146–149.
- Jochem G and Lehnert RJ (2002) On the potential of Raman microscopy for the forensic analysis of coloured textile fibres. *Science & Justice* 42(4): 215–221.
- Jochem G and Lehnert RJB (2002) On the potential of Raman microscopy for the forensic analysis of coloured textile fibres. *Science & Justice* 42(4): 215–221.
- Langdon S, Maynard P, Robertson J, and Roux C (2003) An evaluation of the Maxcam fibre finder version 3.3 on cotton fibres. *Forensic Science International* 135: 137–145.
- Moenssens AA, Inbau FE, and Starrs JE (1986) *Scientific Evidence in Criminal Cases*. New York: The Foundation Press.
- Saferstain R (1988) *Forensic Science Handbook*. New Jersey: Prentice Hall.
- Saferstain R (1990) *Criminalistics, An Introduction to Forensic Science*. New Jersey: Prentice Hall.
- Suzuki S, Suzuki Y, Ohta H, Sugita R, and Marumo Y (2001) Microspectrophotometric discrimination of single fibres dyed by indigo and its derivatives using ultraviolet–visible transmittance spectra. *Science & Justice* 41(2): 107–111.

## Fingerprint Techniques

**C Lennard**, Australian Federal Police, Canberra, NSW, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Physical evidence at a crime scene may be present in an infinite variety of forms and materials. Some of the trace materials and marks left by the offender may be of limited evidential value, while others may be of significant benefit to the investigation. Several forms of evidence can permit the identification of an individual (e.g., fingermarks, dental structure, and

genetic code), but the patterns formed by the friction ridges of the fingers and palms are arguably the most useful and general proof of identity. These patterns are unique, immutable, universal, easy to classify, and leave marks on any smooth object handled with the bare hands.

Smooth objects that have been touched with the bare hands will retain latent impressions, composed of natural secretions and contaminants, of the corresponding ridge patterns on the fingers and palms. In criminal investigations, the detection and recording of a fingermark on a particular item can serve as proof that the item was touched by a specific person. It is for this reason that physical and chemical techniques, capable of developing latent fingermarks on

a wide variety of surfaces, are essential tools for the criminalist. Considerable research into these techniques has been conducted over the last 25 years, and this has led to a dramatic improvement in the sensitivity of fingerprint detection. Latent fingermarks that would once go undetected can now be exploited using routine processing methods.

The general term *lophoscopy* is used to describe the development, classification, and identification of the marks left by the papillary ridges of the skin. These friction ridges are only found on the inner surfaces of the hands (fingers and palms) and the soles of the feet. The study of fingerprints, or *dactylocopy*, forms the section of *lophoscopy* that is the most widely used in practice (even though marks from the palm of the hand or the sole of the foot may also be used for identification purposes).

The classification of fingerprints into distinct groups, based on general pattern similarities, allows the fingerprint examiner to limit the search for an unidentified fingermark to a specific section of the fingerprint reference collection. The various fingerprint classification systems in use throughout the world are based on three fundamental ridge formations: the arch, the loop, and the whorl (**Figure 1**). However, computer systems that entirely automate the storage, searching, and retrieval of fingerprints (automated fingerprint identification systems, or AFIS) have tended to render these traditional classification systems obsolete.

The individualization of fingerprints is achieved through the 'accidental' characteristics formed along the papillary ridges. These characteristics, or *minutiae*, may be summarized as three basic structures: the ridge ending, the bifurcation, and the dot (or island) (**Figure 1**). All other denominations, used for convenience or for statistical purposes, result from a

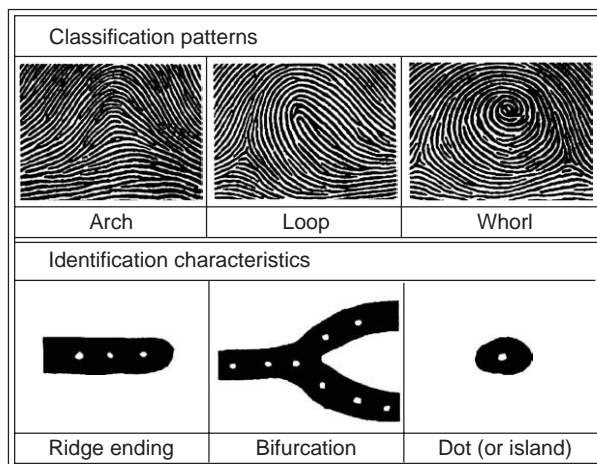
combination of these three ridge features. The study of friction ridge detail (ridgeology) may be further subdivided into the study of ridge contours (edgeoscopy) and pores (poroscopy). The majority of fingerprint identification experts, however, only exploit these latter subdivisions to a limited degree.

Two fingerprints may be considered to be from the same finger if a sufficient number of matching ridge characteristics is present and no unexplainable differences exist between the two prints. The actual number of minutiae required for a formal identification is, by tradition, somewhere between 8 and 18, with '12 points' being a common informal benchmark. However, there is no scientific basis for requiring a minimum number of matching ridge characteristics to reach a conclusion of identity.

## Latent Fingerprints

Three categories of fingermarks may be encountered as evidence at a crime scene or on an object collected for examination. Indented (or plastic) finger impressions are three-dimensional prints pressed into a malleable substance such as putty or candle wax. Visible fingermarks may be either negative, where the fingerprint ridges remove surface material such as dust or soot, or positive, where an image is formed by fingerprint ridges contaminated with a colored substance such as blood, ink, or paint. Latent marks are the most frequently encountered and cause the most difficulty for the criminalist; they are invisible and require treatment for their exploitation. The correct selection of a detection technique (or sequence of techniques) requires a detailed knowledge of natural secretions, together with a consideration of surface characteristics and environmental factors.

A latent fingermark, as deposited onto a surface by the friction ridges of the skin, is a complex mixture of natural secretions (**Table 1**) and contaminants picked up from the environment. Three types of glands are responsible for the natural secretions of the skin: the eccrine glands (found all over the body, but particularly numerous on the palms of the hands and soles of the feet), the apocrine glands (concentrated in the groin, in the arm pits, and in the perianal regions), and the sebaceous glands (principally located on the chest, the back, and the forehead). As the ridges of the hands are covered exclusively by eccrine glands, eccrine gland secretions are present to some degree in every latent fingermark at the moment of deposition. In addition, latent fingermarks often contain lipid material from the sebaceous glands due to contamination of the fingers from activities such as touching the face, combing the hair, etc. Apocrine gland



**Figure 1** The three primary fingerprint classification patterns and identification characteristics (friction ridge minutiae).



**Table 1** Major chemical constituents of the glandular secretions

Source	Inorganic constituents	Organic constituents
Eccrine glands	Chloride	Amino acids
	Metal ions (Na <sup>+</sup> , K <sup>+</sup> , Ca <sup>2+</sup> )	Urea
	Sulfate	Uric acid
	Phosphate	Lactic acid
	Ammonia	Sugars
	Water	Creatinine
		Choline
Apocrine glands	Iron	Proteins
	Water	Carbohydrates
		Sterols
Sebaceous glands –		Fatty acids
		Glycerides
		Hydrocarbons
		Alcohols

From Knowles AM (1978) Aspects of physicochemical methods for the detection of latent fingerprints. *Journal of Physics E: Scientific Instruments* 11: 713–721.

secretions are unlikely to be found in a latent fingerprint deposit except in particular cases such as sexual assault.

The latent fingerprint is invisible and hence, for its exploitation, it must be given characteristics that differentiate it from the surface on which it is found. This differentiation is normally achieved by inducing a coloration or luminescence in the mark by way of a physical or chemical treatment. In some cases, latent fingerprints may be detected using nondestructive optical techniques.

A good-quality latent fingerprint will generally contain ~1000 ng of material, sufficient for detection using a simple powdering technique. Chemical techniques such as ninhydrin treatment, which rely on the formation of a colored reaction product, will detect marks containing as little as 100 ng of material. Techniques that induce a strong luminescence (diazfluorenone (DFO), for example) have detection limits as low as 10 ng. When the quantity of natural secretions in the latent fingerprint falls below such levels, the developed print will tend to be weak and/or partial – generally unsuitable for identification purposes.

## Optical Detection Methods

### Inherent Luminescence

A small percentage of latent fingerprints exhibit a weak inherent luminescence that may be exploited for the detection of ridge detail. The luminescence is generally excited in the blue-green region of the

spectrum using a suitable high-intensity light source such as an argon ion laser or a filtered arc lamp. The origin of this luminescence is unclear but it is most probably the result of contamination of the fingerprint deposit rather than from natural secretions.

In some cases, the luminescence of the surface itself may be used to reveal fingerprint detail. For example, on smooth surfaces that are luminescent under ultraviolet (UV) light (as is the case with white glossy paper), observation under a UV lamp may reveal dark ridge detail against a light background. This is due to the absorption of UV light by the latent fingerprint deposit and the luminescence emission from the substrate.

### Ultraviolet Imaging

The optical detection of latent fingerprints by short-wave UV reflection requires the use of a UV-sensitive charge-coupled device (CCD) camera (equipped with a quartz lens) and a source of UV light. The technique is based on the contrast between the surface, which may absorb or reflect UV light, and the fingerprint deposit that absorbs some UV radiation and diffusely reflects the remainder. The technique may therefore give either light ridges on a dark background or dark ridges on a light background, depending on the nature of the surface and the composition of the fingerprint deposit. The type of UV lamp employed and the angle of incidence of the UV illumination are both critical factors for obtaining satisfactory results. A setup for the application of this technique is known as a reflected UV imaging system (RUVIS).

Another UV imaging technique involves the use of short-wave UV excitation (200–300 nm) with the observation and recording of the corresponding long-wave UV emission (300–400 nm). This form of UV emission has been observed with latent fingerprints on certain surfaces and with body fluids including blood, semen, and saliva.

### Diffuse Reflectance

This technique relies on the light being diffusely reflected by the fingerprint ridges. Latent fingerprints on glass, for example, can sometimes be revealed by varying the angle of incidence of the light source until an angle is found where the ridges diffusely reflect the light while the surface itself remains dark.

Episcopic coaxial illumination is a relatively simple optical setup that exploits diffuse reflectance for revealing fingerprints on smooth, shiny surfaces such as glass, plastic, and polished metal (e.g., firearms). A semitransparent mirror is employed to observe the reflection of white light perpendicular to the surface under examination. The light is diffused by the fingerprint

deposit but specularly reflected by the shiny surface. A dark image of the print will be observed against a light background. This method can also be very effective for the enhancement of fingermarks developed by cyanoacrylate fuming (see below).

## Fingerprint Detection on Nonporous Surfaces

### Powders

Powdering is the traditional fingerprint detection technique used by law enforcement agencies worldwide. Fingerprint powders may be readily applied at the crime scene on all nontransportable objects that have smooth, nonporous surfaces (e.g., gloss painted surfaces, glass, and metal). The powder particles adhere to the humid, sticky, or greasy substances in the latent fingerprint deposit. The technique is economical, relatively easy to apply, and the results are immediately apparent. An additional advantage is that a developed fingermark can be lifted off the surface using an adhesive film.

Many substances and formulations have been investigated as fingerprint powders (e.g., carbon black, graphite, metal oxides, sulfates, and carbonates), but the most sensitive powder for general use is considered to be aluminum flake. Aluminum flake powder, because of the flat plate-like structure of each particle and a surface layer of stearic acid, adheres particularly well to a latent fingerprint deposit. In addition, its hardness and reflectivity aid in the photography and transfer of developed prints.

Fingerprint powder is commonly applied to a surface using a brush, which may have bristles that are synthetic (e.g., fiberglass filaments) or natural (e.g., squirrel or camel hair). A developed fingermark may be subsequently lifted, using adhesive tape or gelatin film, and the lift transported to the laboratory.

Magnetic fingerprint powder, produced by mixing iron grit with aluminum flake, may be applied to a surface using a magnetic wand. This technique is less destructive than conventional powdering as there are no bristles coming into contact with the fragile fingerprint deposit. Colored or luminescent powders have been developed for special cases where conventional powder formulations give unsatisfactory results. A UV-luminescent powder, for example, may be useful for the detection of latent marks on a nonporous, multicolored surface.

Powdering is a physical process that is not particularly sensitive. As latent fingermarks age, they dry out, lose their stickiness, and become less receptive to fingerprint powders. This process will therefore detect only relatively fresh, good-quality latent

fingermarks. As a result, fingerprint powders are typically reserved for the treatment of large or fixed surfaces at the crime scene. More sensitive fingerprint detection methods are applied to objects that can be conveniently transported to the laboratory.

### Small Particle Reagent

Latent fingermarks may be detected on wet, nonporous surfaces by treatment with an aqueous suspension of an insoluble powder followed by rinsing with water. The small particle reagent (SPR), as it is commonly known, is typically prepared from powdered molybdenum disulfide and a detergent mixed with tap water. The powder suspension can be conveniently applied to the surface using a garden spray. As with conventional fingerprint powders, the molybdenum disulfide particles adhere to the greasy, water-insoluble components of the latent fingerprint deposit. The treated object is then rinsed with water to remove excess reagent.

As SPR is a wet powdering technique that is not particularly sensitive, it is reserved for the treatment of wet objects at the crime scene that cannot be readily transported to the laboratory for conventional processing. For example, it may be desirable to treat a wet stolen motor vehicle with SPR rather than wait for the car to be dried before processing by another method.

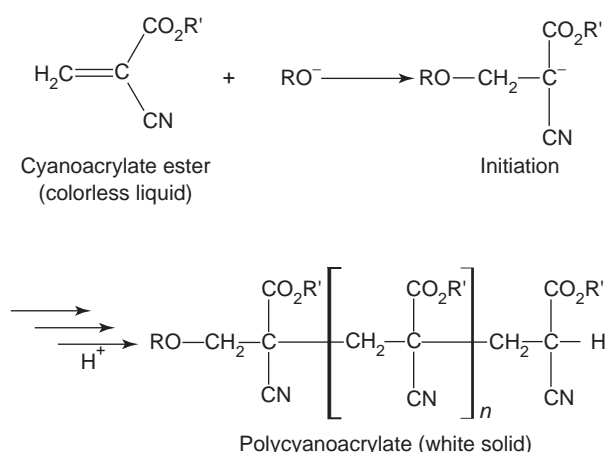
### Cyanoacrylate Fuming

Cyanoacrylate esters (generally the methyl or ethyl ester) are colorless, monomeric liquids sold commercially as quick-setting, high-strength adhesives (e.g., Superglue<sup>®</sup>). These esters polymerize rapidly at ambient temperature via an anionic mechanism in the presence of weak bases such as alcohol and water (Figure 2). The reaction product, polycyanoacrylate, is a hard white solid.

Cyanoacrylate vapor, formed by evaporating the adhesive at room temperature or by heating (80–120°C), selectively polymerizes on latent fingerprint ridges to form a solid white image of the fingermark. The moisture and alcohols present in the latent deposit catalyze the polymerization. Excellent fingerprint detail can be obtained on nonporous surfaces such as glass, metal and plastic, even with relatively old latent marks (up to several years of age).

Typically, objects to be treated are left in an enclosed chamber in the presence of a small quantity of cyanoacrylate ester that is generally heated to aid evaporation (Figure 3). The objects are removed when sufficient fingerprint detail is observed. More recently, vacuum cyanoacrylate systems have been developed that are purported to give better fingerprint detail within a shorter period of time.





**Figure 2** The polymerization reaction of cyanoacrylate ester in the presence of an anionic catalyst, leading to the formation of polycyanoacrylate, a hard white solid.



**Figure 3** Example of a commercially available cyanoacrylate fuming chamber.

Cyanoacrylate developed marks on smooth, flat surfaces such as credit cards may be efficiently recorded using diffuse reflectance (episcopic coaxial illumination) as described previously. In general, the contrast in cyanoacrylate-developed marks may be significantly enhanced by the application of a colored

or luminescent stain. The choice of a particular staining technique will depend on the color and the luminescence properties of the surface under examination. Examples of commonly employed staining procedures include gentian violet, rhodamine 6G, Ardrex<sup>®</sup>, Basic Yellow 40, and safranin.

Cyanoacrylate fuming may also be followed by vacuum metal deposition (VMD) in cases where background printing or background luminescence interferes with normal enhancement procedures (e.g., some credit cards, polymer banknotes, etc.).

### Vacuum Metal Deposition

Fingerprint contamination on a smooth, nonporous (or semiporous) surface can hinder the deposition of metallic films following metal evaporation under vacuum. This phenomenon has been known for a significant period of time but its application to fingerprint development is relatively recent. It is now accepted that VMD is one of the most sensitive techniques for the detection of latent fingerprints on surfaces such as glass and plastic.

A commercial VMD chamber is required for the treatment of articles by this method. After a vacuum is achieved, gold is first evaporated to form a very thin layer of metal on the surface under examination. A layer of zinc is then deposited in a similar manner. (Cadmium may also be used but this metal is generally avoided due to its toxicity.) The gold film is uniformly deposited across the surface of the sample and penetrates the latent fingerprint deposit. The zinc then deposits preferentially on the exposed gold but does not penetrate the fingerprint deposit. This results in transparent fingerprint ridges against a zinc-plated background. On some polymer substrates, depending on the amount of gold deposited, reverse development may also occur (i.e., zinc-plated ridges against a transparent background).

VMD processing can give excellent fingerprint detail on surfaces such as glass and plastic, even for very old or degraded marks. The technique may also develop fresh latent fingerprints (less than 48 h old) on fine-weave synthetic fabric. On some surfaces, such as polymer banknotes, there may be an advantage in using the sequence:

Cyanoacrylate fuming → VMD → Luminescent stain

## Fingerprint Detection on Porous Surfaces

### Ninhydrin and Ninhydrin Analogs

Ninhydrin is the most widely used chemical reagent for the detection of latent fingerprints on porous

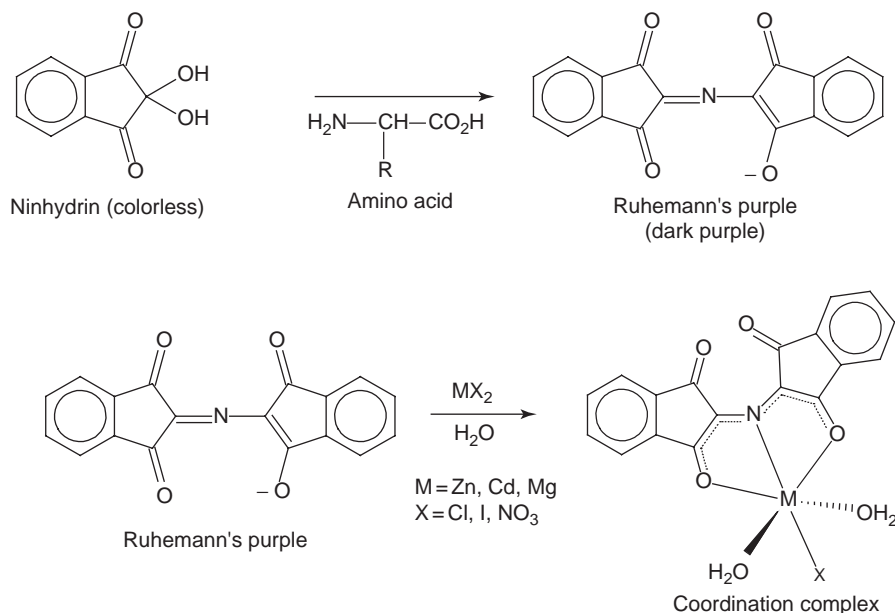
surfaces such as paper and cardboard. The compound reacts with the amino acid (eccrine) component of the fingerprint deposit to give a dark purple product known as Ruhemann's purple (Figure 4). The chemical processes involved are quite complex and development conditions, such as temperature, acidity (pH), and humidity, must be controlled if optimum results are to be obtained.

Paper items are typically treated by brief immersion in a nonaqueous solution of ninhydrin. (For example, ninhydrin may be dissolved in a limited quantity of polar solvent and acetic acid, and then made up to volume with a nonpolar carrier solvent such as HFE-7100 or HFC-4310. A final concentration of ~0.5% w/v is usually employed.) The solution may also be applied using a brush or spray. After drying, fingerprint development is generally achieved at room temperature over 24–48 h, with 50–80% relative humidity being required for optimal development. Developed marks, visible as dark purple images, are generally photographed under white light. Although the technique shows good overall sensitivity (fingerprint marks over 40 years of age have been detected using ninhydrin), weak marks, or marks on dark or multicolored surfaces, may suffer poor contrast.

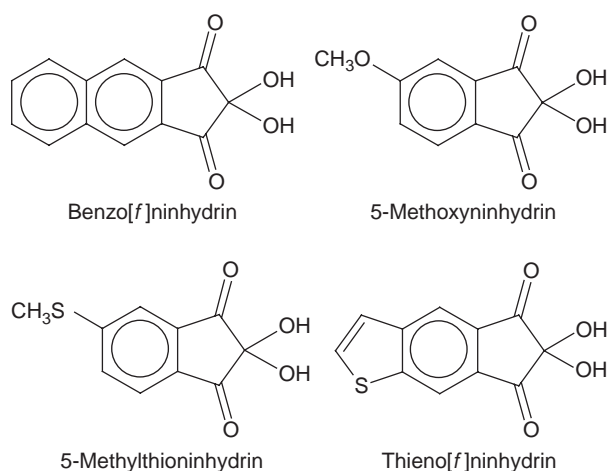
The dark purple coloration in ninhydrin-developed fingerprints may be modified by treatment with a

metal salt solution. For example, treatment with a zinc(II) salt gives an orange color while that produced by a cadmium(II) salt is red. These color changes are due to the formation of a complex between the Ruhemann's purple (the product from the ninhydrin reaction) and the metal ion (Figure 4). The zinc and cadmium complexes show good luminescence properties that may be exploited for the enhancement of ninhydrin-developed fingerprints. For example, a weak ninhydrin-developed mark may be enhanced by treatment with a zinc salt solution (to form a colored and luminescent complex), followed by visualization under conditions favoring the luminescence emission (in this instance, excitation at 490 nm, emission at 550 nm, with the sample cooled to 77 K using liquid nitrogen to increase the emission intensity).

Several ninhydrin analogs have been synthesized and evaluated as alternatives to the conventional ninhydrin reagent for the development of latent fingerprints on paper surfaces. A number of examples are indicated in Figure 5. Some ninhydrin analogs display increased sensitivity for fingerprint detection, particularly after secondary metal salt treatment. More intense fingerprint luminescence can be achieved in comparison with ninhydrin itself. Unfortunately, the high cost and/or lack of availability of these reagents limits their use for routine work.



**Figure 4** The reaction between ninhydrin and an amino acid, leading to the production of a dark purple compound known as Ruhemann's purple. Further treatment with a metal salt (zinc(II), cadmium(II), or mercury(II)) results in the formation of a luminescent coordination complex. (Reprinted with permission from Lennard CJ, Margot PA, Sterns M, and Marrener RN (1987) Photoluminescent enhancement of ninhydrin developed fingerprints by metal complexation: structural studies of complexes formed between Ruhemann's Purple and Group 11b metal salts. *Journal of Forensic Sciences* 32: 597–605.)



**Figure 5** Several ninhydrin analogs that have been successfully applied to the detection of latent fingerprints on paper surfaces.

### Diazafluorenone

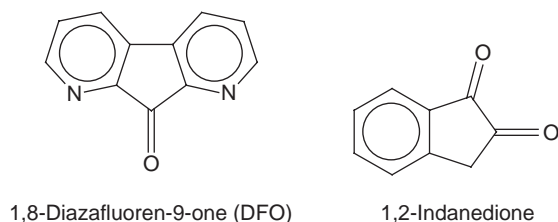
Latent fingerprints on paper may be developed by treatment with a solution of the amino acid sensitive reagent DFO (1,8-diazafluoren-9-one, **Figure 6**). Unlike ninhydrin, heat must be applied to accelerate the reaction. This can be achieved by heating at 100°C for 20 min or by the use of a heat press (180°C for 10 s). Developed marks are pale pinkish-purple in color and exhibit strong room-temperature luminescence (excitation at 470–550 nm, emission at 570–620 nm). The chemical reactions involved are similar to those proposed for ninhydrin and related analogs. At present, DFO is considered to be the most sensitive commercially available chemical reagent for the routine detection of latent fingerprints on paper surfaces.

### Indanedione

A relatively new addition to the range of techniques available for fingerprint detection on paper is the reagent 1,2-indanedione (**Figure 6**). This reagent is chemically similar to ninhydrin and is believed to react with amino acids in a similar fashion. However, unlike ninhydrin, developed marks are luminescent at room temperature without any further treatment. The results from studies conducted in a number of countries suggest that indanedione has a sensitivity comparable to that of DFO on a range of paper substrates. Indanedione is now commercially available and a number of reagent formulations have been proposed.

### Physical Developer

Physical developer (PD) is a fingerprint processing technique for porous surfaces that is sensitive to the



**Figure 6** DFO and indanedione – reagents that react with amino acids to give a luminescent product.

water-insoluble (sebaceous) components of the latent fingerprint deposit. PD can therefore be effective even if the surface has been wet, contrary to amino acid reagents such as ninhydrin and DFO. Fingermarks developed by this procedure are visible as dark gray to black images due to the deposition of silver metal along the print ridges.

PD is an aqueous solution containing silver ions, a ferrous/ferric redox (reduction/oxidation) system, a buffer (citric acid), and a cationic surfactant (generally *n*-dodecylamine acetate). The ferrous (Fe<sup>2+</sup>) ions in solution reduce the silver (Ag<sup>+</sup>) ions to silver metal (Ag<sup>0</sup>), with ferric (Fe<sup>3+</sup>) ions being present to hold back the reaction (eqn [1]):



Citric acid is required to complex with the ferric ions and to maintain a low pH. The action of the surfactant is to inhibit the premature deposition of silver metal by trapping randomly generated silver particles, as they are formed, within positively charged spheres of surfactant molecules known as micelles. The micelles repel the positive silver ions in solution, thereby blocking further silver formation. The PD solution is therefore a delicate balance of ferrous, ferric, and silver ions, stabilized by the presence of citric acid and a surfactant. On contact with the sebaceous material in the fingerprint deposit, the micelles break down and metallic silver is deposited from the solution. As well as detecting fingerprints on items that have been wet, PD complements amino acid sensitive methods on dry paper surfaces in that it can sometimes detect latent marks where other reagents have failed.

The PD technique presents some inconveniences: it is delicate to prepare, the solution is unstable and cannot be stored for longer than about 2 weeks, the method is destructive, some paper surfaces react strongly with the reagent, and no other fingerprint development technique is effective after PD treatment (with the exception of some radioactive detection methods).

## Miscellaneous Techniques

### Multimetal Deposition

Fingerprint development by multimetal deposition (MMD) is a two-step process, the first being immersion of the object in a solution of tetrachloroauric acid (gold(III) chloride) where the active constituent is colloidal gold. Colloidal gold, at a specific pH (between pH 2.5 and 3.9), binds to the peptides and proteins in the latent fingerprint deposit to give a metallic gold image of the fingerprint ridges. (Colloidal gold reagents are commonly employed in biochemistry to detect peptides and proteins.) The second step involves treatment with a modified PD solution that serves to enhance the metallic gold image that is generally quite weak (low contrast). The bound colloidal gold provides a nucleation site around which silver precipitates from the PD solution. A dark-gray to black coloration results that can greatly improve the contrast between the developed print and the substrate.

Many types of surfaces, porous or non-porous, wet or dry, can be treated by the MMD process (e.g., fingerprints may be developed on paper, plastic, expanded polystyrene, and adhesive tape). On some difficult semiporous surfaces such as wax paper and latex gloves, MMD may develop latent fingerprints not detected by other methods.

### Iodine; Iodine-Benzoflavone

Latent fingerprints on a wide variety of surfaces (porous and nonporous) may be detected by fuming with iodine vapor. The iodine is absorbed by the greasy sebaceous material in the fingerprint deposit to give a brown image of the mark. While the technique is simple to apply, it suffers from some major disadvantages:

- iodine vapor is toxic and corrosive;
- contrast in developed prints is generally poor;
- developed marks quickly fade due to reevaporation of absorbed iodine; and
- the technique has limited sensitivity and will generally only detect marks up to several days old.

The contrast and stability of marks fumed with iodine can be improved by secondary treatment with a solution of 7,8-benzoflavone ( $\alpha$ -naphthoflavone). This reagent forms a dark-blue charge transfer complex with the absorbed iodine in the fingerprint deposit.

A combined iodine-benzoflavone reagent can be prepared from two solutions: iodine in a suitable carrier solvent (e.g., cyclohexane or HFC-4310) and

benzoflavone in dichloromethane. The working solution can be used as a spray reagent for the treatment of fixed surfaces at the crime scene. The iodine-benzoflavone spray reagent is effective on a range of different surfaces (e.g., painted surfaces and wallpaper) for prints up to 2 weeks of age, with treated marks becoming visible after a few seconds as dark-blue images.

Fingermarks treated with iodine vapor on human skin or leather may be transferred to a highly polished silver plate by firm contact over several seconds. The transferred fingerprint is then visualized by exposure of the silver plate to a strong UV light source (or sunlight) to produce a dark image of the print. Human skin is one of the most difficult surfaces on which to detect latent fingerprints and very few casework successes have been reported.

### Gentian Violet

A staining solution containing gentian violet (crystal violet) may be employed for the detection of the fatty components in the latent fingerprint deposit. In addition to the stain itself, the solution generally contains phenol to aid the absorption of the stain by the sebaceous material in the fingerprint. An intense purple coloration results. Although the technique may give results on many different surfaces (particularly on nonporous supports contaminated with oils and grease), it is particularly efficient for the detection of latent marks on adhesive surfaces such as pressure-sensitive tapes and films. These surfaces are generally treated by immersion in the reagent, followed by rinsing with water to remove excess stain. In addition, marks developed on dark adhesive tape may be readily transferred onto fixed and washed photographic paper to improve contrast. Gentian violet stain is relatively simple to employ but must be used with caution due to the toxicity of phenol.

### Sticky-Side Powder

Sticky-side powder, a suspension of black fingerprint powder in a detergent solution, is an extremely simple but effective technique for the detection of latent fingerprints on the sticky side of adhesive tape. The technique was developed in Japan in the early 1990s and is now widely employed throughout the world. To prepare the suspension, a detergent solution is mixed with a small amount of black powder until a consistency similar to that of thin paint is achieved. This suspension is then painted onto the adhesive surface using a soft brush (e.g., camel-hair fingerprint brush). After 10–15 s, the adhesive surface is gently rinsed under running tap water. Treated prints are



dark gray to black in color. The process can be repeated if development is weak.

## Fingermarks in Blood

### Optical Detection

Dried blood exhibits a strong absorption maximum at  $\sim 415$  nm. As a result, fingermarks in blood (deposited by fingertips contaminated with blood) can often be enhanced by illumination with a light source filtered to provide violet light at around this wavelength. Dry blood observed under these conditions will appear black. If the surface is reflective or luminescent at this wavelength, good fingerprint contrast will be obtained.

### Amido Black

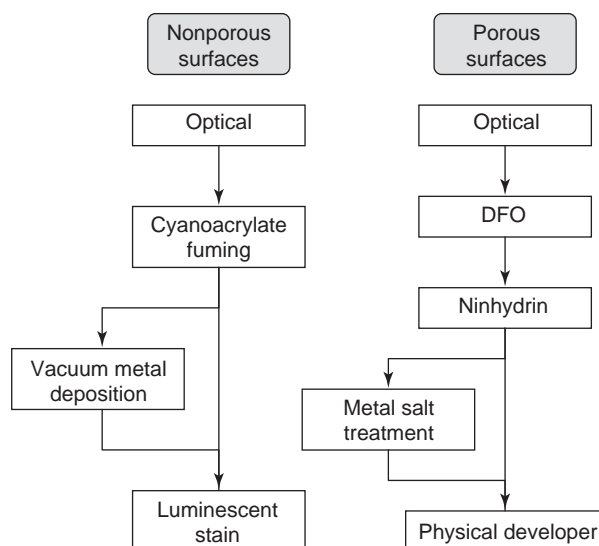
Amido black is a protein stain that may be applied to fingermarks in blood to improve contrast. Blood marks, on either porous or nonporous surfaces, are treated by immersion in a methanolic solution of the stain. The surface is then washed in three destaining solutions to reveal enhanced marks that are dark blue in color. Although the technique is quite effective on nonporous objects, treated marks on porous surfaces such as paper and wood tend to show less contrast due to background staining.

### Diaminobenzidine

The reagent 3,3'-diaminobenzidine (DAB) is a sensitive alternative to protein stains for the enhancement of fingermarks in blood. Blood marks are first fixed with a solution of 5-sulfosalicylic acid. A buffered solution of DAB and hydrogen peroxide is then applied and the reaction allowed to proceed over  $\sim 4$  min. The reaction between hydrogen peroxide and DAB, catalyzed by the peroxidase activity of blood, gives a dark brown insoluble product. Dark brown enhanced marks therefore result. The treatment gives little background coloration, even on porous items, and no destaining procedure is required.

## Fingerprint Detection Sequences

The application of a sequence of fingerprint detection techniques provides the best chance of developing any latent marks that may be present on a particular surface. A correctly chosen combination of methods can increase the total number of fingermarks detected and optimize the amount of detail contained in each mark. However, the techniques must be applied in a logical, predetermined order as the incorrect choice or application of a particular method may preclude the later use of another technique (or, at



**Figure 7** Examples of fingerprint detection sequences for nonporous surfaces, such as glass and plastic, and porous surfaces, such as paper and cardboard. The application of such sequences provides the best chance of detecting any latent fingermarks that may be present.

least, lessen its effectiveness). The choice of a particular sequence will depend on a number of factors including the nature of the support (whether it is porous or nonporous, for example), the environmental conditions to which the object has been exposed (e.g., whether or not the surface has been wet), the presence of any contaminants (e.g., blood), and the likely age of the fingermarks of interest.

Detailed fingerprint detection sequences have been proposed in the literature for various surfaces. Some simple detection sequences are presented in **Figure 7** for general porous surfaces, such as paper and cardboard, and nonporous surfaces, such as glass and plastic. For surfaces that are not commonly encountered, it is recommended that tests be performed on a similar surface to determine what techniques, and what sequence of techniques, should be applied on the evidential item.

See also: **Forensic Sciences:** Overview; Glass; Gunshot Residues; Questioned Documents.

## Further Reading

- Ashbaugh DR (1999) *Qualitative-Quantitative Friction Ridge Analysis – An Introduction to Basic and Advanced Ridgeology*. Boca Raton: CRC Press.
- Cantu AA (2001) Silver physical developers for the visualization of latent prints on paper. *Forensic Science Review* 13: 29–64.

- Champod C, Lennard C, Margot P, and Stoilovic M (2004) *Fingerprints and Other Ridge Skin Impressions*. Boca Raton: CRC Press.
- Cowger JF (1983) *Friction Ridge Skin: Comparison and Identification of Fingerprints*. New York: Elsevier Science Publishing.
- Hardwick SW, Kent T, and Sears VG (1990) *Fingerprint Detection by Fluorescence Examination: A Guide to Operational Implementation*. Sandridge: Home Office, Police Scientific Development Branch.
- Joullié MM, Thompson TR, and Nemeroff NH (1991) Ninhydrin and ninhydrin analogs. Synthesis and applications. *Tetrahedron* 47: 8791–8830.
- Kent T (ed.) (1998) *Manual of Fingerprint Development Techniques*, 2nd edn (Revised January 2001). Sandridge: Home Office, Police Scientific Development Branch.
- Knowles AM (1978) Aspects of physiochemical methods for the detection of latent fingerprints. *Journal of Physics E: Scientific Instruments* 11: 713–721.
- Lee HC and Gaensslen RE (2001) *Advances in Fingerprint Technology*, 2nd edn. New York: Elsevier.
- Lennard CJ and Margot PA (1988) Sequencing of reagents for the improved visualisation of latent fingerprints. *Journal of Forensic Identification* 38: 197–210.
- Lennard CJ and Patterson T (1999) Dactyloscopy: the science of fingerprinting. In: Freckelton I and Selby H (eds.) *Expert Evidence in Criminal Law*, pp. 45–119. Pyrmont: LBC Information Services.
- Menzel ER (1989) Pretreatment of latent prints for laser development. *Forensic Science Review* 1: 43–66.
- Menzel ER (1999) *Fingerprint Detection with Lasers*, 2nd edn. New York: Dekker.
- Olsen RD (1978) *Scott's Fingerprint Mechanics*. Springfield: CC Thomas.
- Pounds CA (1988p) Developments in fingerprint visualisation. In: Maehly A and Williams RL (eds.) *Forensic Science Progress*, vol. 3, pp. 91–119. Berlin: Springer.

## Glass

**J A Lambert**, Forensic Science Service, Wetherby Laboratory, Wetherby, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1635–1642, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by Sarah Cresswell.

## Introduction

Glass is one of the major evidence types encountered in forensic science. In a typical case, small glass fragments recovered from the clothing of a suspect are compared with broken glass from the scene of a crime. Characterization of glass requires a combination of physical and chemical techniques that can be applied to small fragments. Refractive index (RI) is the most informative physical property. A wide range of chemical techniques is available to determine the elemental composition of the glass. Interpretation of results depends on many factors, and although assisted by statistical methods, relies largely on the knowledge and experience of the scientist. The making and forming of glass by melting is an art that originated in Egypt and spread slowly westward. In the sixteenth century, Venice became an important center of progress and glass making flourished. The first methods for the continuous production of window glass were introduced early in the twentieth century, and a further major development was the invention of float glass in 1959.

For the purposes of forensic science, glass can be regarded as a hard brittle amorphous solid, breaking with a conchoidal fracture and formed from the fusion of inorganic oxides. A glass can also be regarded as a supercooled liquid, with the random nature of the liquid frozen into the solid structure. Crystal nucleation and growth is prevented by rapid cooling through the critical region.

Silica ( $\text{SiO}_2$ ) is the most important of the glass-forming oxides although the chemical composition of glasses can vary greatly. The main components of commercially important glasses are shown in **Table 1**.

In addition, a number of trace elements will also be present. These may be deliberately added to give a color or tint to the glass, or may represent impurities in the major components. Elements such as lithium, fluorine, sulfur, chlorine, scandium, titanium, chromium, manganese, cobalt, nickel, zinc, arsenic, selenium, rubidium, strontium, cesium, lanthanum, cerium, tantalum, and uranium fall into this category, and some have considerable importance in forensic science either in identifying the source of a glass fragment or in discriminating between glasses from different sources.

## Glass as Forensic Evidence

The main types of glass dealt with in forensic science casework are listed in **Table 2**. Although total production of container glass considerably exceeds that of other types, glass from buildings and vehicles



- Champod C, Lennard C, Margot P, and Stoilovic M (2004) *Fingerprints and Other Ridge Skin Impressions*. Boca Raton: CRC Press.
- Cowger JF (1983) *Friction Ridge Skin: Comparison and Identification of Fingerprints*. New York: Elsevier Science Publishing.
- Hardwick SW, Kent T, and Sears VG (1990) *Fingerprint Detection by Fluorescence Examination: A Guide to Operational Implementation*. Sandridge: Home Office, Police Scientific Development Branch.
- Joullié MM, Thompson TR, and Nemeroff NH (1991) Ninhydrin and ninhydrin analogs. Synthesis and applications. *Tetrahedron* 47: 8791–8830.
- Kent T (ed.) (1998) *Manual of Fingerprint Development Techniques*, 2nd edn (Revised January 2001). Sandridge: Home Office, Police Scientific Development Branch.
- Knowles AM (1978) Aspects of physiochemical methods for the detection of latent fingerprints. *Journal of Physics E: Scientific Instruments* 11: 713–721.
- Lee HC and Gaensslen RE (2001) *Advances in Fingerprint Technology*, 2nd edn. New York: Elsevier.
- Lennard CJ and Margot PA (1988) Sequencing of reagents for the improved visualisation of latent fingerprints. *Journal of Forensic Identification* 38: 197–210.
- Lennard CJ and Patterson T (1999) Dactyloscopy: the science of fingerprinting. In: Freckelton I and Selby H (eds.) *Expert Evidence in Criminal Law*, pp. 45–119. Pyrmont: LBC Information Services.
- Menzel ER (1989) Pretreatment of latent prints for laser development. *Forensic Science Review* 1: 43–66.
- Menzel ER (1999) *Fingerprint Detection with Lasers*, 2nd edn. New York: Dekker.
- Olsen RD (1978) *Scott's Fingerprint Mechanics*. Springfield: CC Thomas.
- Pounds CA (1988p) Developments in fingerprint visualisation. In: Maehly A and Williams RL (eds.) *Forensic Science Progress*, vol. 3, pp. 91–119. Berlin: Springer.

## Glass

**J A Lambert**, Forensic Science Service, Wetherby Laboratory, Wetherby, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1635–1642, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by Sarah Cresswell.

## Introduction

Glass is one of the major evidence types encountered in forensic science. In a typical case, small glass fragments recovered from the clothing of a suspect are compared with broken glass from the scene of a crime. Characterization of glass requires a combination of physical and chemical techniques that can be applied to small fragments. Refractive index (RI) is the most informative physical property. A wide range of chemical techniques is available to determine the elemental composition of the glass. Interpretation of results depends on many factors, and although assisted by statistical methods, relies largely on the knowledge and experience of the scientist. The making and forming of glass by melting is an art that originated in Egypt and spread slowly westward. In the sixteenth century, Venice became an important center of progress and glass making flourished. The first methods for the continuous production of window glass were introduced early in the twentieth century, and a further major development was the invention of float glass in 1959.

For the purposes of forensic science, glass can be regarded as a hard brittle amorphous solid, breaking with a conchoidal fracture and formed from the fusion of inorganic oxides. A glass can also be regarded as a supercooled liquid, with the random nature of the liquid frozen into the solid structure. Crystal nucleation and growth is prevented by rapid cooling through the critical region.

Silica (SiO<sub>2</sub>) is the most important of the glass-forming oxides although the chemical composition of glasses can vary greatly. The main components of commercially important glasses are shown in **Table 1**.

In addition, a number of trace elements will also be present. These may be deliberately added to give a color or tint to the glass, or may represent impurities in the major components. Elements such as lithium, fluorine, sulfur, chlorine, scandium, titanium, chromium, manganese, cobalt, nickel, zinc, arsenic, selenium, rubidium, strontium, cesium, lanthanum, cerium, tantalum, and uranium fall into this category, and some have considerable importance in forensic science either in identifying the source of a glass fragment or in discriminating between glasses from different sources.

## Glass as Forensic Evidence

The main types of glass dealt with in forensic science casework are listed in **Table 2**. Although total production of container glass considerably exceeds that of other types, glass from buildings and vehicles

**Table 1** Main components of glass encountered in forensic science laboratories

<i>Component</i>	<i>Typical source of raw material</i>	<i>Function in the glass</i>
SiO <sub>2</sub>	High-quality sand or sandstone	Glass-forming oxide, sometimes partially replaced by B <sub>2</sub> O <sub>3</sub>
Na <sub>2</sub> O	Na <sub>2</sub> CO <sub>3</sub> – as soda ash	Reduces melting point and lowers viscosity
K <sub>2</sub> O	Depends on sources of other raw materials – not deliberately added	Similar to Na <sub>2</sub> O
MgO	Dolomite	Permits slower setting and reduces crack formation
CaO	Limestone	Decreases solubility of sodium silicates and hence increases chemical durability; increases fluidity when hot and allows for more rapid cooling
Al <sub>2</sub> O <sub>3</sub>	Feldspars, blast furnace slag	Increases mechanical strength and inhibits devitrification; raises the viscosity of the melt
Fe <sub>2</sub> O <sub>3</sub>	Impurity associated with other raw materials, especially the sand	Discolors glass – used to introduce a tint
BaO	Barytes	Increases durability, and also increases refractive index
PbO	Lead oxide, lead sulfide	Increases refractive index, enhances appearance, e.g., lead crystal wine glasses
B <sub>2</sub> O <sub>3</sub>	Borax	Increases scratch resistance and durability

**Table 2** Main types of glass encountered in forensic science casework

<i>Type</i>	<i>Manufacturing process</i>	<i>End use</i>
Float	Glass flows from the furnace onto a layer of molten tin in a controlled reducing atmosphere	Wide – especially for buildings, vehicles and mirrors. Has largely replaced other types of flat glass
Sheet	Drawn vertically as a continuous ribbon	Horticultural, domestic where the appearance is not critical
Plate	Initially cast into trays, later a continuous process. Glass is ground and polished to produce a flat surface	Buildings, mirrors. Superseded by float glass
Patterned	Partly cooled glass ribbon passes through metal rollers, one of which bears the pattern	Buildings
Wire reinforced	Wire mesh is sandwiched between two layers of partly cooled glass. Can be clear or patterned	Buildings – increases security and fire resistance – as compared to standard glass
Toughened	Rapid cooling of glass to introduce a high level of stress. Usually float glass	Motor vehicles, safety glass in buildings
Laminated	Thin sheets of glass are bonded to clear plastic	Motor vehicle windscreens, security glass
Containers and tableware	Blow/molding process	Packaging industry, especially food and drinks, drinking glasses

forms the majority of glass examined in forensic science laboratories. Float glass is used almost exclusively for clear windows although, as glass is a very durable material, many older sources of glass are also routinely encountered. In common with other brittle materials, when a glass object such as a window or container breaks, many small sharp-edged glass fragments are produced. In particular, when a window is broken by an applied force such as bodily pressure, or by striking with a solid object, the glass will bend, causing the surface on the side of the applied force to be compressed and the opposite surface to be in a state of tension. Stress is relieved by the formation of cracks that rapidly propagate through the thickness of the glass from the remote side to the side bearing the breaking force. Small fragments flake off along the edges of the cracks, and can be

projected for a distance of at least 3 m from the window. A considerable proportion of these fragments will include part of the original surface of the glass, and the great majority will be less than 1 mm in size. Such fragments are readily transferred to the clothing of a person who is nearby. Glass fragments can also be transferred by contact with sharp edges of freshly broken glass. The clothing is searched and recovered glass fragments are compared with a known control sample of glass from the scene of the crime.

## Measurement of Physical Properties

Typical glass fragments recovered from clothing will have a largest dimension in the range 0.2–1.0 mm and a mass of 100 µg or less. A considerable

proportion of such fragments will have a surface that represents part of the original surface of the unbroken glass object. Although larger pieces of glass are available as control samples from the known source of broken glass, it follows that all the most significant comparison techniques have been developed to deal with small fragments.

The two most widely measured physical properties are density and RI. These two properties are highly correlated in glass, consequently if one is measured with sufficient accuracy and precision, little additional information is to be gained by measuring the other. Optical dispersion and physical stress are the other properties that are measured in small glass fragments. Examination of glass surfaces also provides useful information.

### Density

Until about 1980, density was the most important physical property used for glass comparisons in forensic science laboratories. Although largely replaced by RI measurement, it is still currently used, and is particularly valuable as a screening technique for large numbers of glass fragments.

Two main methods have been developed for density measurement. Both methods are based on a mixture of two liquid components, one less dense and the other denser than glass. Bromohydrocarbons are typically used. The first is a comparative method in which the components are combined to give a liquid of uniform density that is equal to that of the control glass sample. Fragments of control glass will suspend in the liquid. Recovered fragments will also suspend if they have the same densities as the control glass, otherwise they will either float or sink. The density of the liquid mixture can be altered by either changing the temperature or by adding more of one of the liquid components.

In the second method, a stable density gradient is formed in a tall container by introducing a mixture of the two liquid components. As the container is filled, the density of the liquid mixture is decreased by altering the proportions of the components. Fragments of glass are added to the column of liquid and will sink to a position at which the liquid has the same density as the glass. Standards of known density are added to the column and the densities of the glass under test can be obtained by interpolation. An accuracy of  $0.5 \text{ mg ml}^{-1}$  is typically achieved.

The precision of density measurement is affected by the shape and size of the glass fragments and also by the presence of surface deposits or cracks. The liquids used are noxious and recovery of small glass fragments is not easy. These disadvantages are overcome by measurement of RI.

### Refractive Index

Refractive index is the most useful physical property in forensic glass characterization. It can be measured to a high degree of accuracy and precision, which is essentially independent of fragment size. The RI of a glass depends upon the combination of raw materials used, together with the nature of the manufacturing process.

The RI of glasses examined in forensic science laboratories generally falls in the narrow range of 1.51–1.54, with the exception of borosilicate glass from motor vehicle headlamp lenses, which typically has an RI of 1.475–1.480. As RI measurement can discriminate glasses with RI differences of 0.0001, glasses can be divided into a large number of groups by this technique.

The inherent variation in RI within the bulk of a single sheet of glass is of the order of 0.00005–0.0001, so greater accuracy and precision of measurement is unlikely to increase the discriminating power of this technique.

Current techniques of RI measurement are based on the fact that the RI of a liquid changes with temperature at a much greater rate than that of a solid. Typical RI/temperature coefficients are  $4 \times 10^{-4} \text{ C}^{-1}$  for a liquid, and  $1 \times 10^{-6} \text{ C}^{-1}$  for a glass.

A fragment of glass is immersed in a colorless oil and placed on a hotstage that is programmed to change temperature with time at a linear rate. The oil is calibrated by glass standards of known RI. This enables the RI of the oil to be calculated for any particular hotstage temperature.

The glass fragment is viewed by transmitted light through a phase contrast microscope. Monochromatic light from a sodium lamp or filtered light of a narrow range of wavelength produces a bright line around the edge of the fragment. The temperature of the hotstage is varied until the oil has the same RI as the glass. At this point, known as the match point temperature, the bright line around the fragment disappears. The match point of a glass fragment can be detected automatically, although the image of the fragment is displayed on a screen throughout the measuring process.

### Dispersion

The RI of a material varies with wavelength and this wavelength dependence, known as dispersion, is particularly useful for identifying minerals. The dispersion of a glass can be estimated by measuring the RI at three different wavelengths. Although dispersion is quite widely measured in the USA, recent studies have suggested that compared with RI measurement at a single wavelength, limited additional

discrimination is achieved when dealing with glass of modern production.

### Physical Stress

Thermal and structural stresses are introduced into glass during the manufacturing process. Above a temperature known as the annealing point (typically about 600°C for float glass), the structure of the glass readily changes with any increase in temperature and therefore any stress will be rapidly relieved. Below the annealing point, the time taken for the glass to reach structural equilibrium increases rapidly as the temperature falls, and below a temperature known as the strain point (typically ~450°C), the structure of the glass is essentially fixed and does not change with temperature or time. One of the final stages in glass manufacture is to remove residual stress by annealing the glass; that is, maintaining it at a fixed temperature and then allowing it to cool at a slow controlled rate. Toughened glass is produced by rapidly cooling below the annealing point in order to introduce a high degree of stress. The interior of the glass is in tensile stress and the two surfaces are in compressive stress. This makes the glass more resistant to low velocity blunt impacts, hence its use in motor vehicles.

The high degree of stress in a toughened glass causes it to have a lower RI than ordinary glass of the same elemental composition. Some of the stress is relieved when the glass breaks, but the residual stress in a fragment of toughened glass is considerably greater than that remaining in a fragment from a nontoughened glass source. This high stress level can be revealed by reannealing the glass. Relieving the stress increases the RI. The increase depends on the reannealing schedule. A typical increase in RI for a toughened glass would be 0.0015, whereas the increase for most nontoughened glasses would be less than 0.0005. In glasses with a very low stress level, reannealing may increase the stress, causing the RI to decrease. Reannealing, therefore, provides a technique for distinguishing between fragments of toughened and nontoughened glasses.

### Surface Characteristics

Examination of any original surface present on a recovered glass fragment can provide information on its probable origin. Interference microscopy is one of the techniques used to visualize the surface features. Monochromatic light is split into two beams, one reflected from the surface of the fragment under test and the other from the surface of a reference mirror. Recombination of the beams produces an interferogram, interpretation of which can assist in determining the topography of the surface. Parallel lines

(or fringes) indicate a flat original surface. Curved lined or irregular patterns are more difficult to interpret and require comparison with the surface of the control glass sample, or with fringe patterns produced by known glass surfaces.

Differential interference contrast microscopy is also used to enhance surface features. Plane polarized light is split by a prism into two beams that travel in slightly different directions. The beams are reflected from the surface of the glass fragment, and then recombined to produce an interference pattern that gives surface features a three-dimensional appearance. A useful application of this technique is in examining the scratches in glass fragments from the outer surface of a vehicle windscreen, where the wiper blades have been in contact with the glass surface.

The surface of float glass that has been in contact with molten tin gives a yellow fluorescence under ultraviolet radiation at 254 nm. This fluorescence can be seen in small glass fragments. The presence of tin in the surface layer of float glass produces a higher RI than that of the bulk glass. The difference between the surface and bulk RI is difficult to quantify, but is typically of the order of 0.003 when measured by the oil immersion method.

### Elemental Analysis

Measurement of physical properties enables glass fragments to be divided into many different groups, but relatively little information is obtained as to the source of the broken glass object from which the fragments have originated. Elemental analysis can assist in:

1. classifying the type of the glass, e.g., window, container, or tableware;
2. identifying the probable country of origin;
3. discriminating between samples that are indistinguishable in physical properties;
4. assessing the evidential significance of matching glass.

Elements that are useful for classification and discrimination include lithium, aluminum, magnesium, potassium, iron, manganese, arsenic, barium, rubidium, and strontium (Table 3).

An ideal method of elemental analysis for glass should:

1. be nondestructive; the sample should be available for further analysis;
2. be capable of analyzing the smallest glass fragments recovered in casework;

**Table 3** Typical glass compositions

Glass type	Composition (as % oxides by wt.)							
	SiO <sub>2</sub>	Na <sub>2</sub> O	K <sub>2</sub> O	MgO	CaO	Al <sub>2</sub> O <sub>3</sub>	Fe <sub>2</sub> O <sub>3</sub>	BaO
Float	72.6	13.0	0.6	3.9	8.4	1.0	0.1	
Plate	72.5	13.2	0.7	3.8	8.1	1.3	0.2	
Container	73.0	14.0		0.3	10.9	1.4		
Drinking glass	73.8	15.3		3.8	8.1	1.3		
Bulb	70.3	16.4	1.0	3.4	5.5	1.4		2.0
Television tube	64.0	9.6	6.5	1.2	1.7	3.3	0.1	12.6

**Table 4** Current methods of elemental analysis for glass in forensic science

Probable year of introduction	Method	Destructive technique	Full range of elements detected	Further information
1970	Neutron activation analysis (NAA)	No	Yes	Pitts and Kratochvil (1991)
1973	Energy dispersive X-ray analysis using scanning electron microscopy (SEM-EDX)	No	No	Andrasko and Maehly (1978)
1977	Energy dispersive X-ray fluorescence spectrometry (EDXRF)	No	No	Koons et al. (1991)
1979	Inductively coupled plasma atomic emission spectroscopy (ICP-AES)	Yes	Yes	Catterick and Hickman (1979)
1989	Inductively coupled plasma mass spectrometry (ICP-MS)	Yes	Yes	Zurhaar and Mullings (1990)

3. have high absolute and relative sensitivity;
4. be able to detect and quantify all relevant elements;
5. be cost-effective, with a rapid throughput of samples.

A wide range of techniques has been used for the analysis of glass fragments. Several of these techniques, including atomic absorption spectrometry, DC arc emission spectrography, and spark source mass spectrometry are no longer in routine use. Historically, much of the early analytical effort was directed toward classification of the glass, but at the present time there is a greater emphasis on discrimination and on improving the assessment of evidential significance. Current techniques are listed in Table 4 and are briefly described below.

#### Neutron Activation Analysis (NAA)

Glass samples are placed in plastic vials and irradiated in the neutron flux of a nuclear reactor. Short irradiation times produce low levels of residual radiation, enabling the samples to be handled safely within a few days of analysis. Data are acquired by a germanium detector and multichannel analyzer. Most

elements of interest possess several radioactive isotopes, enabling some degree of choice between detecting either a short-lived or long-lived radionuclide.

Analysis of variance techniques can be used to study differences between glass samples, without reference to quantitative elemental standards.

#### Scanning Electron Microscopy-Energy Dispersive X-Ray Microanalysis (SEM-EDX)

Glass samples are presented to the scanning electron microscope on a stub. The preferred method is to encapsulate the glass fragment in a resin, then grind and polish to produce a flat surface. A quicker alternative is to mount the fragment directly on a carbon stub, but this method tends to give lower analytical precision. In either method, the fragment is coated with a thin layer of carbon to make the surface electrically conductive. A typical stub will bear at least one glass standard of known composition, together with control and recovered fragments. The prepared stub is mounted on a movable stage in the microscope. A flat area of the sample is selected for analysis by viewing the SEM video display. Bombarding the sample with an electron beam produces



an energy spectrum that is composed of peaks due to characteristic X-rays, superimposed on a relatively featureless background. Standard data processing techniques are used to compensate for factors that affect peak intensities, enabling reliable quantitative results to be obtained.

### **Energy Dispersive X-Ray Fluorescence (EDXRF) Spectrometry**

Glass fragments are cleaned, mounted on a plastic probe, and placed in the XRF spectrometer. The resultant energy spectrum is collected by a multichannel analyzer. Secondary X-rays from the detectable elements within the glass produce characteristic peaks in the energy spectrum. The detected peak intensity for a particular element in a glass fragment depends upon the amount present, and also on other factors including the geometry of the fragment. When fragments are to be compared they should be closely matched in shape and size. Available instruments are not able to analyze the smallest glass fragments encountered in casework, and it is not easy to obtain reliable quantitative results by this technique.

### **Inductively Coupled Plasma-Atomic Emission Spectrometry (ICP-AES)**

Glass fragments are cleaned, and then dissolved in a mixture of hydrofluoric and hydrochloric acids. A suitable element is added as an internal standard. The solution is nebulized and the resultant aerosol is introduced into the plasma by argon carrier gas. The intensities of the emission lines of the chosen elements are measured sequentially. Quantitative results are achieved by comparing intensities against those obtained from analysis of multielement standard solutions.

### **Inductively Coupled Plasma-Mass Spectrometry (ICP-MS)**

Sample preparation is similar to that required for ICP-AES, involving nebulizing a solution of the dissolved glass fragment into the ICP-MS system. Data are accumulated sequentially on all measured isotope intensities over the selected mass range. Application of appropriate correction factors allows the element concentrations to be calculated from the mass of the fragment and the volume of the solution. The instrumentation is checked by repeated analyses of standard glasses with known elemental compositions.

### **Advantages and Disadvantages**

Not one of the above methods is ideal in all respects. For a nondestructive analysis, the SEM-EDX and EDXRF techniques complement each other well, but

not all useful elements can be detected. Current developments in EDXRF spectrometry, including altering the geometry of the instrument with respect to the excitation source and the sample, have the potential to increase the power of this technique considerably.

The most generally used destructive technique is ICP-AES, although it cannot effectively deal with the smallest casework samples. ICP-MS is a newer technique that has superseded spark source mass spectrometry. A limitation of NAA is the requirement of a nuclear reactor as an energy source so it is unlikely that this technique will become widely used.

### **Surface Analysis**

As a consequence of the manufacturing process, the surface of a sheet of glass is likely to have a slightly different chemical composition from that of the bulk. In float glass, the two surfaces will be very different in composition, since one has been in contact with a layer of molten tin whereas the other has not. Some reusable containers may receive a surface coating during manufacture, to increase scratch resistance and strength. Atmospheric weathering may change the surface composition of a glass by leaching out the more soluble components.

A wide range of analytical techniques has been used by the glass industry to examine glass surfaces, but in general these are not applicable to small fragments and consequently have not been developed for use in forensic science. However, SEM-EDX does not suffer from this size limitation and can provide valuable information on the surface composition of glass fragments.

### **Interpretation of Results**

A typical forensic case concerning glass involves a search for and recovery of glass fragments, surface examination, measurement of RI, and elemental analysis. A two-stage approach is widely used to assess the results of the physical and chemical comparisons. The first stage is to determine how many types of recovered glass are present, and whether or not any match the control with respect to the measured properties. Graphical displays and statistical tests are used to assist in this process. The second stage is to assess the significance of any matching glass, usually by reference to a data collection. A type of glass that is rarely found on clothing is likely to be of greater evidential value than a type that is frequently encountered. The two-stage approach works well for simple cases, but it is difficult to apply in complex cases involving more than one control sample and several types of recovered glass.



An alternative approach, which encapsulates the matching and significance aspects of a comparison in a single stage, is to express the scientific evidence as a likelihood ratio of the form:

$$\frac{\text{probability of scientific findings} \\ \text{if suspect involved in crime}}{\text{probability of scientific findings} \\ \text{if suspect not involved in crime}}$$

Calculation of the numerator involves assessing the probability that glass fragments of a particular type will be transferred to and retained on the clothing of a suspect, given the circumstances of the case. Calculation of the denominator involves assessing the probability of finding the glass on the clothing of the suspect, given that he or she is a randomly selected member of an appropriate population.

Regardless of which method of interpretation is used, the evidence may have to be evaluated in the light of several different possible explanations. Combining all the relevant information to reach a final conclusion depends largely upon the experience of the scientist.

Knowledge-based computer systems can be used to assist the scientist by providing an inferential framework and access to relevant databases, including results of previous cases. It would be very difficult to construct a comprehensive statistical model to deal with all cases involving glass evidence. A compromise approach that combines statistical methods with expert experience and judgment is likely to be the most effective.

## Casework Examples

### Burglary at a Public House

The premises were illegally entered by smashing a ground floor window. Money was stolen from several gaming and video machines. A few minutes later a suspect was arrested nearby. He had a pair of gloves in the pocket of his coat. Several hundred glass fragments were recovered from each glove, and several hundred more from the surface and pockets of his coat. The RI of 10 of the fragments was measured and all were found to match the control float glass from the broken window of the public house. Interference microscopy showed that two of the 10 fragments possessed an original flat surface. The control glass and one of the recovered fragments also gave a similar result when analyzed by SEM-EDX. Measurement of these physical and chemical properties showed that the glass was a type that is fairly frequently encountered in casework.

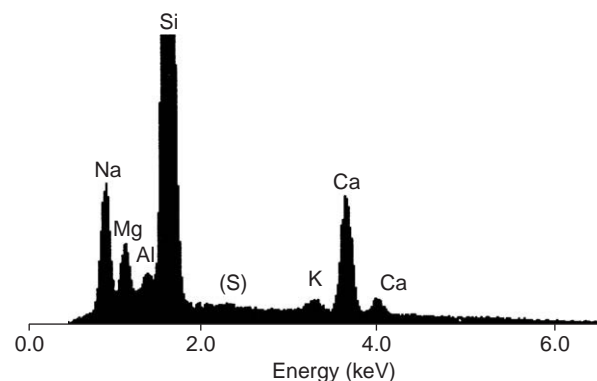
It is very unusual for such a large number of glass fragments to be found on items of clothing. The conclusion reached was that the wearer of the gloves and coat had been in recent contact with breaking or broken glass and that the findings provided very strong support for the allegation that he was involved in the incident in question.

### Criminal Damage at a Craft Shop

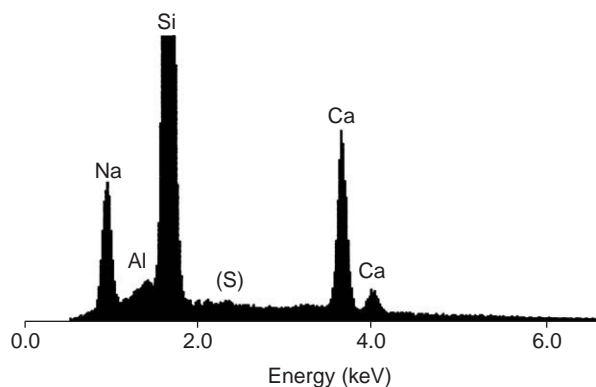
The shop window was broken and a Ford Escort motor car was witnessed leaving the scene. The car was stopped ~15 miles from the scene and the occupant was arrested. A baseball bat was found on the rear seat of the vehicle. On examination the bat was found to have numerous recent cuts in the surface of the wood. Many of these cuts contained embedded and crushed glass fragments. Fragments from five of the cuts were examined and found to match the control sample of float glass from the shop in RI and elemental composition. The results, as shown in **Figure 1**, were typical of a modern mass-produced float glass. The forensic scientist concluded that with regard to the alleged circumstances of the incident, the scientific evidence strongly supported the view that the bat had been in forcible contact with the broken window from the shop.

### Burglary at a Post Office

The sound of breaking glass was heard at a Post Office, and two men were seen running away. Two men were later arrested and items of clothing were obtained from them. A control glass sample was obtained from the broken window. Five types of glass were recovered from the clothing of the first man and six types from the clothing of the second man. RI measurement showed that one of the types of recovered glass from each man matched the glass



**Figure 1** Analysis by SEM-EDX of glass from a case involving criminal damage to a craft shop. The RI of the glass was 1.5175; glass recovered from cuts in a baseball bat gave similar results to glass from the broken window.



**Figure 2** Analysis by SEM-EDX of glass from a post office burglary; glass recovered from the clothing of the two suspects gave similar results to control glass from the broken window.

from the Post Office. Analysis of the control and recovered glass by SEM-EDX indicated that the glass was of an old type that is rarely encountered in casework. The results of the analysis are shown in **Figure 2**. Unlike modern window glass, the elements potassium and magnesium were not detected. The conclusion reached was that with regard to the alleged circumstances, the matching glass recovered from the clothing provided very strong support for the view that the two men were involved in the incident in question.

See also: **Activation Analysis:** Neutron Activation. **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Chemometrics and Statistics:** Expert Systems. **Glasses. Microscopy Applications:** Forensic. **Optical Spectroscopy:** Refractometry and Reflectometry. **X-Ray Fluorescence and Emission:** Energy Dispersive X-Ray Fluorescence.

## Further Reading

- Caddy B (ed.) (2001) *Forensic Examination of Glass and Paint, Analysis and Interpretation*. London: Taylor and Francis.
- Curran JM, Hicks TN, and Buckleton JS (2000) *Forensic interpretation of glass evidence*. Boca Raton, FL: CRC Press.
- Curran JM, Triggs CM, Almirall JR, Buckleton JS, and Walsh KAJ (1997) The interpretation of elemental composition measurements from forensic glass evidence. *Science and Justice* 37(4): 241–244.
- Evvett IW, Lambert JA, and Buckleton JS (1995) Further observations on glass evidence interpretation. *Science and Justice* 35(4): 283–289.
- Hicks T, Vanina R, and Margot P (1996) Transfer and persistence of glass fragments on garments. *Science and Justice* 36: 101–107.
- Koons RD, Peters CA, and Rebbert PS (1991) Comparison of refractive index, energy dispersive X-ray fluorescence and inductively coupled plasma atomic emission spectrometry for forensic characterization of sheet glass fragments. *Journal of Analytical Atomic Spectrometry* 6: 451–456.
- Lambert JA, Satterwhite MJ, and Harrison PH (1995) A survey of glass fragments recovered from clothing of persons suspected of involvement in crime. *Science and Justice* 35(4): 273–281.
- McJunkins SP and Thornton JI (1973) Glass fracture analysis: a review. *Forensic Science* 2: 1–27.
- McQuillan EK (1992) A survey of the distribution of glass on clothing. *Journal of Forensic Sciences* 32(4): 333–348.
- Pitts SJ and Kratochvil B (1991) Statistical discrimination of flat glass fragments by instrumental neutron activation analysis methods for forensic science applications. *Journal of Forensic Sciences* 36: 122–137.
- Zurhaar A and Mullings L (1990) Characterisation of forensic glass samples using inductively coupled plasma mass spectrometry. *Journal of Analytical Atomic Spectrometry* 5: 611–617.

## Gunshot Residues

**S W Lewis and K M Agg**, Deakin University, Geelong, VIC, Australia

**S J Gutowski and P Ross**, Victoria Police Forensic Services Centre, Macleod, VIC, Australia

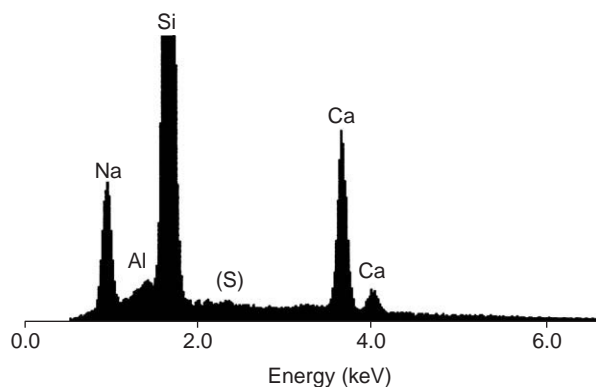
© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The identification of a substance as gunshot residue (GSR) can be vitally important in cases where

firearms have been discharged. Not only can GSR be used to identify bullet holes, it can also yield useful information as to firing distance, identification of a suspect as having recently used a firearm, and in some circumstances may help distinguish suicide from suspicious death.

When the trigger of a firearm is depressed, a hammer or firing pin strikes the base of a cartridge that sets off a series of chemical reactions that rapidly results in the projectile (or projectiles in the case of shot guns) being propelled from the cartridge case and out of the weapon.



**Figure 2** Analysis by SEM-EDX of glass from a post office burglary; glass recovered from the clothing of the two suspects gave similar results to control glass from the broken window.

from the Post Office. Analysis of the control and recovered glass by SEM-EDX indicated that the glass was of an old type that is rarely encountered in casework. The results of the analysis are shown in **Figure 2**. Unlike modern window glass, the elements potassium and magnesium were not detected. The conclusion reached was that with regard to the alleged circumstances, the matching glass recovered from the clothing provided very strong support for the view that the two men were involved in the incident in question.

See also: **Activation Analysis:** Neutron Activation. **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Chemometrics and Statistics:** Expert Systems. **Glasses. Microscopy Applications:** Forensic. **Optical Spectroscopy:** Refractometry and Reflectometry. **X-Ray Fluorescence and Emission:** Energy Dispersive X-Ray Fluorescence.

## Further Reading

- Caddy B (ed.) (2001) *Forensic Examination of Glass and Paint, Analysis and Interpretation*. London: Taylor and Francis.
- Curran JM, Hicks TN, and Buckleton JS (2000) *Forensic interpretation of glass evidence*. Boca Raton, FL: CRC Press.
- Curran JM, Triggs CM, Almirall JR, Buckleton JS, and Walsh KAJ (1997) The interpretation of elemental composition measurements from forensic glass evidence. *Science and Justice* 37(4): 241–244.
- Evvett IW, Lambert JA, and Buckleton JS (1995) Further observations on glass evidence interpretation. *Science and Justice* 35(4): 283–289.
- Hicks T, Vanina R, and Margot P (1996) Transfer and persistence of glass fragments on garments. *Science and Justice* 36: 101–107.
- Koons RD, Peters CA, and Rebbert PS (1991) Comparison of refractive index, energy dispersive X-ray fluorescence and inductively coupled plasma atomic emission spectrometry for forensic characterization of sheet glass fragments. *Journal of Analytical Atomic Spectrometry* 6: 451–456.
- Lambert JA, Satterwhite MJ, and Harrison PH (1995) A survey of glass fragments recovered from clothing of persons suspected of involvement in crime. *Science and Justice* 35(4): 273–281.
- McJunkins SP and Thornton JI (1973) Glass fracture analysis: a review. *Forensic Science* 2: 1–27.
- McQuillan EK (1992) A survey of the distribution of glass on clothing. *Journal of Forensic Sciences* 32(4): 333–348.
- Pitts SJ and Kratochvil B (1991) Statistical discrimination of flat glass fragments by instrumental neutron activation analysis methods for forensic science applications. *Journal of Forensic Sciences* 36: 122–137.
- Zurhaar A and Mullings L (1990) Characterisation of forensic glass samples using inductively coupled plasma mass spectrometry. *Journal of Analytical Atomic Spectrometry* 5: 611–617.

## Gunshot Residues

**S W Lewis and K M Agg**, Deakin University, Geelong, VIC, Australia

**S J Gutowski and P Ross**, Victoria Police Forensic Services Centre, Macleod, VIC, Australia

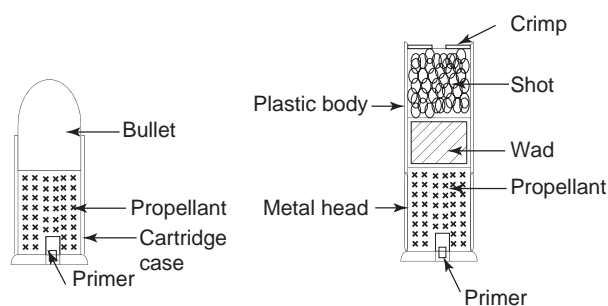
© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The identification of a substance as gunshot residue (GSR) can be vitally important in cases where

firearms have been discharged. Not only can GSR be used to identify bullet holes, it can also yield useful information as to firing distance, identification of a suspect as having recently used a firearm, and in some circumstances may help distinguish suicide from suspicious death.

When the trigger of a firearm is depressed, a hammer or firing pin strikes the base of a cartridge that sets off a series of chemical reactions that rapidly results in the projectile (or projectiles in the case of shot guns) being propelled from the cartridge case and out of the weapon.

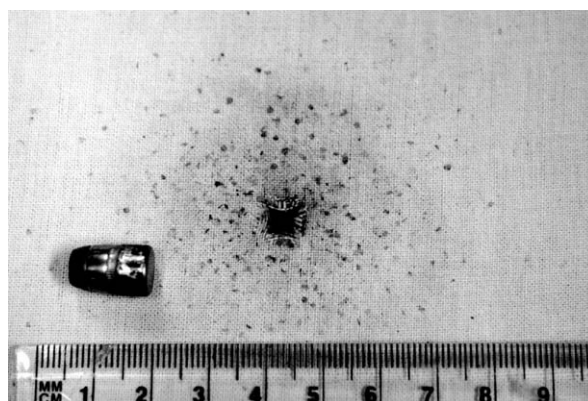


**Figure 1** Schematic diagram of firearm cartridges.

Cartridges, whether they are based on a single projectile (bullet) or many (shot), have the same basic architecture (see **Figure 1**). At the base of the cartridge is a small amount of shock-sensitive chemical mixture; this is termed the primer. Above the primer is an explosive mixture, the propellant, with the projectile (bullet or shot) clamped at the top of the cartridge. In the case of shot there may also be wadding material to hold the multiple projectiles separate from the propellant.

On depression of the trigger, the firing pin or hammer strikes the base of the cartridge case initiating the explosion of the shock-sensitive primer. The energy generated by this reaction in turn causes the rapid combustion of the propellant mixture and the generation of gases at high temperature and pressure. The pressure of gas causes the projectile(s) to be ejected from the cartridge and down the barrel of the firearm. The propellant continues to burn behind the projectile all the way down the barrel, thus resulting in acceleration of the projectile(s) due to the increased volume of gas that is produced.

GSR arises from the cooling and condensation of the gases produced by the combustion reactions within the firearm. Most of the gases are expelled down the barrel of the weapon; however, significant quantities also escape from the many openings present within the weapon such as the breech, ejection ports, and trigger. As the gases cool and condense, they form the particles that are referred to as GSR. The residues may be associated with the primer cap, propellant, bullet (slug/shot), bullet jacket and lubricant, cartridge case, and firearm barrel. Escaping GSR can be deposited on anyone or anything within close enough range of the discharging weapon. For example, if a victim is shot at short range, there is likelihood that any bullet hole in their clothing (or body) will be surrounded by deposits of GSR (see **Figure 2**). Besides identifying a hole as being caused by a bullet, the size of the deposits of residue can be used in estimations of shooting distance. If a suspect has discharged a firearm, GSR will



**Figure 2** Cloth sample showing spread of GSR around a bullet hole. (Photograph Kent Agg, Deakin University.)

be found on the hand that held the weapon and also on their clothing.

Key to forensic investigations such as these is the positive identification of suspect materials as GSR. The analytical methods that have been utilized over the years for the identification of GSR can be divided into two classes, based on determination of

- inorganic species such as, lead, barium, and antimony, which are mainly present in the primer and the bullet (but which also includes nitrate and nitrite ions from the propellant); and
- organic compounds from the propellant, which serve various roles in the discharge of the firearm.

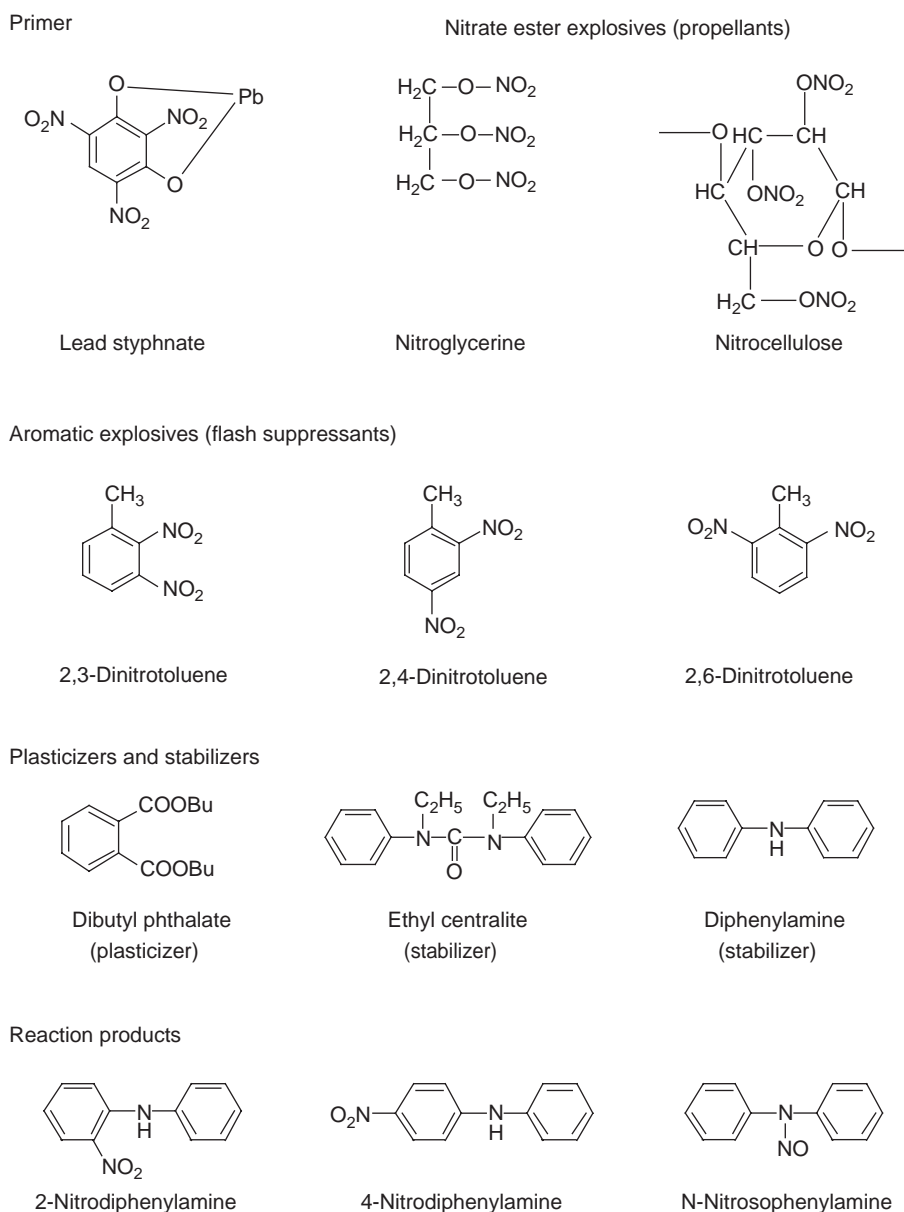
Identification of these residues, particularly when combined with consideration of the location and pattern of any residues detected, can give invaluable information to investigators and courts of law.

## Chemical Species Contributing to GSR

The various compounds that can contribute to GSR come from a variety of sources including the firearm itself as well as the cartridge and its contents. This wide range of chemical species described below is classified by their source; **Figure 3** gives the chemical structures for some of the more common compounds.

### Propellants

The first propellant employed for firearms was black powder (also colloquially known as gunpowder), which is a mixture of charcoal (carbon), sulfur, and potassium nitrate, with typical proportions of 15% carbon, 10% sulfur, and 75% potassium nitrate.



**Figure 3** Some examples of compounds contributing to GSR.

Smokeless propellants were developed toward the end of the nineteenth century and are based on nitrocellulose, which was produced by the nitration of cotton linters or wood pulp using a mixture of nitric and sulfuric acids. These propellants are classified as either being single base with nitrocellulose as the sole propellant, double base containing nitrocellulose and nitroglycerine, and triple base also containing nitroguanidine (picrite). The latter kind of propellant is generally restricted to military uses.

In addition to the chemical composition, the grain size and shape of the propellant is also of importance as it can have a marked influence on the rate of burning.

Single- and double-base propellant grains can have various morphologies, including long-, short-, or multiperforated tubes, short rods (macaroni), rings, cubes, flakes, strips, disks, balls, and flattened balls.

The combustion properties of the propellant can also be altered by using various surface coatings and additives such as stabilizers, oxidizers, plasticizers or gelatinizers, and deterrents or retardants.

Graphite, centralites (substituted phenylureas), dibutyl phthalate, camphor, and dinitrotoluene isomers are the commonly used retardants. Phthalates (such as dibutyl and diamyl) and camphor can also act as plasticizers during the manufacturing process.



Stabilizers that have been utilized in propellant formulations include diphenylamine, Vaseline (substituted diphenylureas), Arkadite (diphenylurea), centralites, and various substituted urethanes. Although rare, carbene, a polyacetylene, can be added to some fast-burning double-base propellants as an accelerator. This type of propellant is likely to only be encountered in large caliber military ammunition.

### Primers

Primers are designed to ensure that the propellant is ignited by rapid combustion instead of explosive detonation. In addition to a shock-sensitive detonator or initiator, primers also contain oxidizing agents, fuel compounds, sensitizers, fillers, and binding agents.

The earliest primers were based on mixtures of potassium chlorate, sulfur, and carbon (charcoal); these were replaced in the 1830s by mixtures utilizing mercury fulminate and lead azide as initiators. However, due to the corrosive effect of mercury fulminate, these primers have been superseded by formulations based on lead styphnate.

The majority of primers today are based on the 'synoxyd' primer, first introduced in 1928. These utilize lead styphnate as the initiator, antimony sulfide as the fuel and frictionator (a substance that helps to efficiently pass on the energy from compression of the primer cap), and barium nitrate as the oxidizer. Other compounds that may also be present include calcium silicide, acting as a fuel and frictionator; gum arabic, a fuel and binder; oxidizers such as lead dioxide, lead nitrate, and lead hypophosphite; ground glass or aluminum powder as frictionators; tetrazene, a sensitizer; and various other gums and resins as binders.

In more recent years there has been the development of lead-free primers based on the organic shock-sensitive initiator dinol (diazodinitrophenol) in place of the lead styphnate. Commercial formulations of dinol-based primer have been developed which include mixtures such as dinol, zinc peroxide, metallic titanium, and tetrazene; dinol, manganese(IV) oxide, and metallic aluminum; and dinol, antimony sulfide, and barium nitrate.

### Bullets, Cartridge Case, and Primer Cap

Bullets and shot are usually lead or lead-antimony alloys of between 0.5% and 3% antimony to harden the lead (in rare cases tin can be the hardener). Bullets can also be semijacketed, tipped, or fully jacketed, with copper and copper alloys with zinc and/or tin being the most common materials. Bullet jackets

can also be coated with chrome, nickel, cupro-nickel, and zinc.

Cartridge cases are generally brass (~70% copper and 30% zinc) and can be plated with nickel, aluminum, or steel. Steel coated with zinc or copper has been used but rarely. Shotgun cartridges are usually fabricated from cardboard or polyethylene with a brass base.

The primer cap is usually brass or copper that is often nickel plated; galvanized steel caps have also been employed.

### Other Sources of Residues

In addition to the cartridge and its contents, other substances present within the firearm such as oil and rust can also be sources of residue.

## Sampling and Sample Preparation Techniques

The process of sampling is a vital step in GSR analysis. Contamination is a significant issue and the location and distribution of particles, not only their identification as GSR, is often of great importance. Consequently, the method of sampling will depend on the nature of the investigation being carried out and the subsequent analysis method. The most commonly applied method of collection is swabbing using fabric swabs moistened with dilute (usually nitric) acid solutions, water, and a variety of solvents, depending on the analytical technique to be used. This technique partially solubilizes the GSR. It is of vital importance that the swabs and solvent used are free of contaminants and it is essential that a blank swab with solvent is collected at the time of sampling in order to detect any background contamination.

Vacuum lifting by collection of GSR from surfaces by vacuuming through a series of filters of different pore size has also been used. This method is useful particularly when recovering residues from clothing. However, its very efficiency can be a disadvantage as all particulates, GSR and otherwise, will be collected, with GSR being present only in very small proportions. Additionally, due to vacuuming being a relatively vigorous technique, any attempts to distinguish more recently deposited particles on the surface from older particulate matter within the weave of the cloth will be foiled.

Tape lifting, where samples are simply, quickly, and directly collected onto adhesive tape, is the main method of collection for analysis by scanning electron microscopy (see below). This technique with subsequent dissolution of the residues in a suitable solvent has also been used to collect GSR prior to



analysis by separation methods such as capillary electrophoresis.

## Analytical Methods for the Detection and Identification of GSR

A wide variety of analytical methods have been applied to the detection and identification of GSR. In common with other forensic analyses they can be divided into color (chemical) spot tests for initial screening and instrumental methods. The situation with GSR is complicated by the issues of contamination and by the fact that many of the chemical components seen in GSR may have other innocent or environmental sources.

### Color (Chemical) Tests

Chemical reactions that provide a characteristic color on reaction with components of GSR have widely been used. They have the advantage that they can often be carried out in the field with little or no scientific instrumentation and yield a rapid result. However, this result requires careful interpretation as the reactions have been shown to be nonspecific and have poor sensitivity. Consequently, they can only be considered a preliminary screening test, and are not suitable as a confirmation of identification of GSR.

The earliest method of detection of GSR on the hands and skin of a suspected shooter was the dermal nitrate or paraffin test. This is based upon the color reaction of nitrites and nitrates with diphenylamine in sulfuric acid. This reaction is not specific to nitrates and nitrites and will give positive results with other oxidizing agents such as chlorates, dichromates, iodates, bromates, permanganates, and higher metal oxides. This test can also give false negative results in certain situations where it is known that a person has fired a weapon and has not yet washed his or her hands. Consequently, this procedure is no longer recommended.

The Griess test similarly involves a reagent ( $\alpha$ -naphthylamine, and sulfanilic acid in acetic acid) targeting nitrites. This test is based on the formation of a diazonium salt when nitrite in acid solution reacts with a primary aromatic amine. The nitrites react with sulfanilic acid in acetic acid to form a diazonium ion that is then coupled to  $\alpha$ -naphthylamine to form a red azo dye. The reported carcinogen  $\alpha$ -naphthylamine can be replaced with Marshall's reagent *N*-(1-naphthyl)-ethylenediamine or  $\alpha$ -naphthol. A permanent reproduction of a GSR pattern can be produced by utilizing the reagent immobilized on desensitized photographic paper, which is then

placed against the bullet hole and surrounding residue and subjected to a heat treatment. Dark red spots are observed corresponding to the partially burnt GSR grains being transferred to the photographic paper. Other color tests for inorganic compounds include the sodium rhodizonate test where the reagent reacts with lead to form a violet-blue color and the rubeanic acid test where copper or nickel react to produce colored precipitates.

Due to the potential of environmental contaminants these tests are generally not used to detect GSR particles in the first instance, rather they are used to visualize GSR distribution around bullet holes.

### Instrumental Analysis

Several instrumental methods have been applied to the determination of the various components of GSR. To date most of the procedures used operationally have concentrated on the inorganic fraction of GSR. In particular, the methods of elemental analysis targeting lead, barium, and antimony have widely been used, as this elemental combination is characteristic of most GSR. More recently research interest has focused on the organic components of GSR.

### Neutron Activation Analysis

Neutron activation analysis (NAA) is a nondestructive method based upon the conversion of stable isotopes of chemical elements to unstable radioactive isotopes by irradiation with thermal neutrons within a nuclear reactor. If the ensuing decay of irradiated nuclides occurs via  $\gamma$ -radiation, a  $\gamma$ -ray spectrum can be obtained on a  $\gamma$ -spectrometer, which is characteristic of a specific element.

NAA can be used for identification of antimony and barium in bulk samples (as opposed to single particles) of GSR. It is highly sensitive but is not applicable to lead and requires an intense neutron source such as a suitable nuclear reactor. Despite these limitations NAA has found application to GSR analysis in forensic investigations.

### Atomic Absorption Spectroscopy and Related Techniques

Atomic absorption spectroscopy (AAS) has found wide application as a bulk analysis technique for GSR. This method of analysis is capable of sequentially determining lead, antimony, and barium; however, in order to detect the low concentration of these species that may be present in GSR, the more sensitive flameless AAS techniques such as electrothermal atomization are required.

More recently, inductively coupled plasma atomic emission spectroscopy has also been applied to the analysis of GSR; however, it has not found widespread use, presumably due to the high cost of the instrumentation.

### Scanning Electron Microscopy

Scanning electron microscopy in combination with X-ray spectroscopy, particularly in the energy-dispersive mode (SEM-EDX), has become the method of choice of most forensic facilities for the analysis of GSR. Instead of the traditional light microscope, a beam of electrons is aimed onto the specimen, which interacts with the atoms of the sample. Due to the large atomic mass of lead, antimony, and barium, particles containing these elements give strong, bright images due to backscatter of primary electrons. In addition, low-energy electrons are released from the surface of the specimen; these secondary electrons are important for imaging as they provide useful information concerning the morphology of the sample; in the case of GSR this is highly characteristic (see **Figure 4**). The rapid cooling of gases from high temperatures and pressures results in particles in a range of sizes and shapes from small spheres to irregular flakes, and in trained hands SEM can usually distinguish between environmental contaminants and actual GSR using each particle's size, shape, and morphology.

A major advantage of this method comes from the ability to couple the SEM instrument to an X-ray detector: in the case of SEM-EDX a silicon–lithium semiconductor detector. X-rays are produced when the electrons of the scanning electron microscope strike each GSR particle. Analysis of the X-rays

generated from that particle allows identification and quantitation of individual elements as each element emits a characteristic X-ray spectrum. Fine spheroidal grain particles abundant in antimony, barium, and lead make a strong case that the particle is in fact a GSR. The lead–antimony–barium ratio helps confirm that these elements are detected as a result of the discharge of a firearm rather than as a result of some other environmental means.

This combination of morphological and chemical information for single particles in a nondestructive method makes this approach to GSR analysis unrivalled in its specificity. Due to the obvious power of the technique, it has become widely established in larger forensic facilities and modern instruments have been developed to enable automated GSR particle searches.

### Chromatographic and Other Separation Techniques

Separation techniques such as gas chromatography (GC), high-performance liquid chromatography (HPLC), and capillary electrophoresis (CE) underpin attempts to characterize the organic species in GSR, which arise from partial combustion of the propellant. HPLC has found most application, with electrochemical detection being the most versatile method for the analysis of organic propellant residues. Other detection methods that have been employed for HPLC include mass spectrometry, MS/MS, and fluorimetry (after derivatization with a suitable fluorophore). As the organic constituents in GSR are uncharged, applications of CE to GSR analysis have been in general restricted to micellar electrokinetic capillary chromatography, of which there are a number of reports in the literature.

The use of GC for the analysis of GSR is limited due to the incompatibility of many of the analytes, such as the nitrate esters, and the lack of volatility of nitrocellulose, the major component of smokeless propellant. Despite this, GSR has been successfully analyzed using GC employing a variety of detectors including flame ionization, electron capture, and thermal energy analysis. Thermal energy analysis has been found to be particularly useful as it is selective for the nitro-containing compounds present in GSR. Recent attempts have been made to individualize GSR to a particular class of firearm and/or ammunition using chemometric techniques.



**Figure 4** Typical electron micrograph of a GSR particle. (Photograph Harald Wrobel, Victoria Police Forensic Services Department.)

See also: **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Capillary Electrophoresis:** Overview. **Forensic Sciences:** Overview; Explosives; Thin-Layer Chromatography. **Gas Chromatography:**

Overview. **Microscopy Techniques:** Scanning Electron Microscopy.

## Further Reading

Meng H and Caddy B (1997) Gunshot residue analysis – a review. *Journal of Forensic Sciences* 42: 553.  
Romolo FS and Margot P (2001) Identification of gunshot residue: a critical review. *Forensic Science International* 119: 195.

Ross P (1993) Firearm discharge residues. In: Freckleton I and Selby H (eds.) *Expert Evidence*, vol. 4, ch. 87. Sydney: The Law Book Company.  
Rowe WF (2000) Firearms: residues. In: Siegel JA (ed.) *Encyclopedia of Forensic Sciences*, vol. 2. San Diego: Academic Press.  
Saferstein R (2001) *Criminalistics – An Introduction to Forensic Science*. Upper Saddle River, NJ: Prentice-Hall.  
Schwoeble AJ and Exline DL (2000) *Current Methods in Forensic Gunshot Residue Analysis*. Boca Raton, FL: CRC Press.

## Hair

**N Watson**, Strathclyde University, Glasgow, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Yoshihiko Fujita, pp. 1655–1661, © 1995, Elsevier Ltd.

## Introduction

Recently, crime has become more vicious and skilful than ever before, so minute substances (trace evidence), which are difficult to hide, are important for the detection of the suspect, especially in cases where he is in contact with the victim or the crime scene. These substances are hair, fibers, soil, etc. Among them hair is noteworthy and can be retrieved easily from the crime scene, because there are ~100 000 hairs on an adult human head and they may be shed one by one every 15 min. The suspect therefore leaves his or her own head or pubic hairs at the crime scene unwittingly. Therefore, hair provides important physical evidence for solving of crime.

Nowadays in the forensic science laboratory, large pieces of human (and animal) hairs collected from the crime scene are examined by morphological methods using macroscope, microscope, and electron microscope, and may be compared with sample hairs of humans associated with the crime. Finally, their origins are determined.

However, it is well known that identification of hair by morphological methods is not always reliable and should be done by a well-trained analyst for maximum reliability. It is difficult to distinguish the suspect's hair from many other hairs left at the crime scene if there is no finding of particular characteristics in the morphology. Therefore, in addition to morphological examination, ABO blood typing from hair can be made, which is helpful in the

comparison of hairs. Furthermore, various instrumental methods for hair analysis have been investigated to obtain more objective information and enhance the degree of certainty in forensic hair comparison. These methods, including a new powerful method (deoxyribonucleic acid (DNA) typing analysis), are introduced in this article together with older ones.

One consequence of the application of DNA testing techniques to hair occurring as evidence has been the reexamination of old cases as the new technologies have become available. In the several instances that have been reported, DNA evidence from hairs has contradicted the interpretations based upon the microscopic examination of crucial hairs used in the original trials. In addition, an American Supreme Court decision in 1993, concerning the grounds of the admissibility of scientific evidence, has resulted in a review of many categories of scientific evidence and their use in the American courts. The reliability of hair evidence has been a cause of particular concern and a number of studies to establish the scientific credentials of such examinations have been published. While these issues are primarily of concern to American practitioners the forensic community in general has benefited from these studies.

Another issue that is continually being addressed is the speed and cost-effectiveness of hair comparisons. To this end the systematic codification of microscopic features in order to generate coded categories of variant forms to improve the consistency of hair examination reporting continues to be developed. Potentially, a count can be made of the different codes reported in a laboratory or system of laboratories to enable a numerical evaluation of the commonness of any given variant form to be estimated. This value would assist the court to assess the significance of evidence.

Overview. **Microscopy Techniques:** Scanning Electron Microscopy.

## Further Reading

Meng H and Caddy B (1997) Gunshot residue analysis – a review. *Journal of Forensic Sciences* 42: 553.  
Romolo FS and Margot P (2001) Identification of gunshot residue: a critical review. *Forensic Science International* 119: 195.

Ross P (1993) Firearm discharge residues. In: Freckleton I and Selby H (eds.) *Expert Evidence*, vol. 4, ch. 87. Sydney: The Law Book Company.  
Rowe WF (2000) Firearms: residues. In: Siegel JA (ed.) *Encyclopedia of Forensic Sciences*, vol. 2. San Diego: Academic Press.  
Saferstein R (2001) *Criminalistics – An Introduction to Forensic Science*. Upper Saddle River, NJ: Prentice-Hall.  
Schwoeble AJ and Exline DL (2000) *Current Methods in Forensic Gunshot Residue Analysis*. Boca Raton, FL: CRC Press.

## Hair

**N Watson**, Strathclyde University, Glasgow, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Yoshihiko Fujita, pp. 1655–1661, © 1995, Elsevier Ltd.

## Introduction

Recently, crime has become more vicious and skilful than ever before, so minute substances (trace evidence), which are difficult to hide, are important for the detection of the suspect, especially in cases where he is in contact with the victim or the crime scene. These substances are hair, fibers, soil, etc. Among them hair is noteworthy and can be retrieved easily from the crime scene, because there are ~100 000 hairs on an adult human head and they may be shed one by one every 15 min. The suspect therefore leaves his or her own head or pubic hairs at the crime scene unwittingly. Therefore, hair provides important physical evidence for solving of crime.

Nowadays in the forensic science laboratory, large pieces of human (and animal) hairs collected from the crime scene are examined by morphological methods using macroscope, microscope, and electron microscope, and may be compared with sample hairs of humans associated with the crime. Finally, their origins are determined.

However, it is well known that identification of hair by morphological methods is not always reliable and should be done by a well-trained analyst for maximum reliability. It is difficult to distinguish the suspect's hair from many other hairs left at the crime scene if there is no finding of particular characteristics in the morphology. Therefore, in addition to morphological examination, ABO blood typing from hair can be made, which is helpful in the

comparison of hairs. Furthermore, various instrumental methods for hair analysis have been investigated to obtain more objective information and enhance the degree of certainty in forensic hair comparison. These methods, including a new powerful method (deoxyribonucleic acid (DNA) typing analysis), are introduced in this article together with older ones.

One consequence of the application of DNA testing techniques to hair occurring as evidence has been the reexamination of old cases as the new technologies have become available. In the several instances that have been reported, DNA evidence from hairs has contradicted the interpretations based upon the microscopic examination of crucial hairs used in the original trials. In addition, an American Supreme Court decision in 1993, concerning the grounds of the admissibility of scientific evidence, has resulted in a review of many categories of scientific evidence and their use in the American courts. The reliability of hair evidence has been a cause of particular concern and a number of studies to establish the scientific credentials of such examinations have been published. While these issues are primarily of concern to American practitioners the forensic community in general has benefited from these studies.

Another issue that is continually being addressed is the speed and cost-effectiveness of hair comparisons. To this end the systematic codification of microscopic features in order to generate coded categories of variant forms to improve the consistency of hair examination reporting continues to be developed. Potentially, a count can be made of the different codes reported in a laboratory or system of laboratories to enable a numerical evaluation of the commonness of any given variant form to be estimated. This value would assist the court to assess the significance of evidence.



## Structure and Composition of Hair

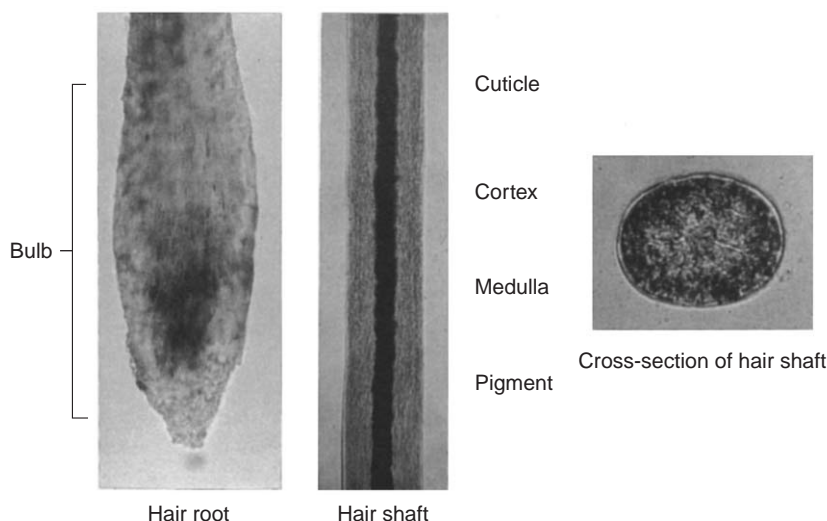
Hair is an appendage of the mammalian skin produced by an organ known as the hair follicle in the epidermal epithelium. The portion of the hair that lies in the follicle is called the root, whereas the portion above the skin surface is called the shaft. In a cross-sectional view of hair shaft, there is an outer layer of cuticular scales that surrounds the cortex shaft in the center of which a canal-like structure called the medulla is found (Figure 1). Pigment granules (melanin) are dispersed usually in the cortex, produced by melanocytes in the follicle. The hair follicle does not produce hair continuously. Hair grows cyclically, with alternate periods of growth and rest. This cycle is divided into three stages (anagen, catagen, and telogen). In the anagen stage, hair grows on, and in the subsequent catagen stage club hair is formed in the follicle. In the next resting telogen stage, hair ceases to grow. After the inactive telogen stage, the follicle goes into another anagen stage and resumes activity. In man, each follicle has its own growth cycle that is independent of the others. During the growth stage, hair grows at a rate of about one centimeter per month. An adult man has ~100 000 scalp follicles, 90% of which are anagen and 10% telogen.

Chemically, hair is mainly composed of particular proteins (keratins), hard and stable to usual physical and chemical treatment. This characteristic of keratin protein is a result of the high degree of disulfide bonding between polypeptide chains in the molecules. Human hair keratin is composed of 21 amino acids. When these amino acids are grouped together according to functional group, there are high content of hydroxyl, primary amide, and basic amino acid

functions. Human hair keratin is also characterized by high sulfur content due to high cystine content (17–19% residues as half cystine). The remaining constituents are water (*c.* 13% in 65% relative humidity), lipids (originating from the sebaceous gland in the follicle), pigment (principally eumelanins), and trace elements (calcium, magnesium, potassium, zinc, sodium, chlorine, etc., from sweat deposits and extraneous sources).

## Morphological Examination of Hair

At the crime scene, various fibers and hairs are collected by policemen. They are immediately examined at the forensic laboratory in order to identify the suspect and help solve the crime. The examination needs to establish: (1) whether it is a hair or a fiber (synthetic or plant); (2) its species if it is a hair; (3) its somatic origin if it is a human hair; (4) whether it is similar to the suspect or the concerned persons. Microscopic examination will resolve the first point. It will also allow discrimination between human and animal hair, as the medulla of human hair does not develop well and shows less than 0.30 of medullary index (ratio of medulla diameter to hair diameter). But determining the species of animal requires experience in animal hair examination. Almost every animal hair shows specific shape of cuticle and medulla under microscopical observation. With regard to the somatic origin of human hair, there are some basic features that usually indicate the bodily origin of the hair. For example, head hair has long length, 30–120  $\mu\text{m}$  diameter, tapered tip, little diameter variation, various medullation, sometimes cut tips, and treatment. Pubic hair has a twisted



**Figure 1** Microscopic pictures of human hair shed naturally.

form, shorter length, coarse diameter, prominent diameter variation, broad medulla. The morphological association of human hair is carried out by comparing every characteristic of both a suspect's and unknown hair macroscopically and microscopically. The association cannot be done by only one or two characteristics. There are variations within one person from natural growth, but the amount and kind of variation differs between persons. Generally, effective items of macroscopical and microscopical identification of human hair are color, length, diameter, contour, distribution of medulla, and cosmetic treatment. Concerning individuality, there is a limit to the possibility of identifying individual hair among large pieces of hair from many persons by morphological examination.

For some time many forensic laboratories have adopted some form of systematic comparison strategy to improve the consistency of the examinations. One example of this was demonstrated in a study of the efficacy of hair comparisons by microscopic examinations. Fourteen different characteristics were defined as primary and secondary characteristics for comparison. The primary characteristics were the color, cosmetic treatment, texture, pigment distribution, medulla, medullary index (i.e., the width of the medulla as a proportion of the width of the hair), maximum diameter, and presence of cortical fusi (microscopic structures observable within the cortex of the hair). The secondary characteristics were length, cuticular margin, pigment density, pigment size, and shaft diameter variation. By a systematic point-by-point comparison of these characteristics by two examiners working on the same samples in parallel but independently of one another both examiners were able to make correct assignments in a proportion of trials but more importantly did not make any false matches. In other blind trials forensic laboratories could achieve very high rates of correct associations, with no false associations, although some laboratories achieved much poorer success rates. The poorer rates of success have been attributed to the standards of training undergone by the examiners but the studies have provided evidence that, with experienced and well-trained examiners, microscopic hair examination can be a rigorous, truly scientific method of comparison. It is crucial that the hair examiner must be careful that an unwarranted degree of weight is not put on the results.

There are continuing developments of schemes to codify the characteristics by assigning a code to each different alternative form of any given characteristic. Thus, an individual hair specimen may be characterized by a numerical score for each feature defined by the scheme and these can be compiled into an

ordered series. In this way hairs can be categorized and a count made of the number of occurrences of a category in a population. Such a count can be made to generate frequency data so that an examiner could draw upon observational data to give a numerical estimate of the rarity or commonness of any of the defined hair types within a population. An important aspect of such scoring systems is the consistency with which the categories are assigned. However, by the use of photographic standards, termed archetypes, published as an atlas of microscopic features, the form, or variate, of each characteristic observed in the hair considered can be categorized consistently by reference to the archetype.

## Chemical Analysis of Hair Components

### Elements

Chemical features determined by analysis of elements in human hair are helpful for hair comparisons. Analysis of elements in the hair has been accomplished by instrumental methods such as atomic absorption spectrometry, neutron activation analysis, energy dispersive X-ray microanalysis (EDX). Among them, EDX equipped with scanning electron microscopy is widely used in forensic hair examination because it facilitates easy analysis of elements whilst observing the hair structure.

Twenty-seven elements are found in human hairs, and the majority of them are due to an accumulation of products from sweat. These elements can be grouped into two classes: physiological and environmental. The former includes sodium, phosphorus, potassium, and sulfur, which are present at high levels. The latter includes arsenic, cadmium, mercury, and lead, which are called toxic elements. As these elements may indicate some particular environment for an individual, and their amounts vary considerably across individual hairs, they can be specific and effective indications for forensic hair comparison. If a particular element was determined in a suspect's hair, it can be easily compared in sample hairs.

Content of elements in hair is closely concerned with hair color, disease, treatment, sex, and weathering. For instance, the amount of chlorine in male hairs detected by EDX is higher than in female ones among 38 adult persons (Japanese) aged from 20 to 50 years (males:  $0.67 \pm 0.20\%$  w/w; females:  $0.45 \pm 0.12\%$  w/w). Chlorine in hair is considered to come from sweat. Electrothermal atomic absorption spectrometric analysis also shows that calcium is more abundant in permanent waved hair than in untreated hair (waved hair:  $2000\text{--}4500 \mu\text{g g}^{-1}$ ).



Copper content is low in the hair of infantile pellagra (kwashiorkor) and grey hair of an old man, but is high in trichoschisis and black hair. Content of zinc is low in dwarf and blond hair and also in the hair of pregnant women, but is high in ill-nourished children and black hair. Some elements in hair show significant differences of value according to gender and age, as well as various hair treatments. These variations also occur in different scalp areas and different portions of single hair of a man. If there are some patterns of variation in sample hairs found, it may be helpful to compare them with those of suspect's hair.

### Instrumental Analysis

New methods in hair analysis using the many remarkable instruments that have now been developed are described in this section. Human hair (40–80  $\mu\text{g}$ ;  $\sim 1\text{ cm}$ ) can be analyzed by pyrolysis capillary column gas chromatography (Pyr-GC). Some main peaks are observed in the pyrograms of hair samples. Components of these peaks are identified by pyrolysis capillary column gas chromatography-mass spectrometry (PyrGC-MS). Toluene and styrene as pyrolysis products are considered to arise from phenylalanine, phenol and cresol from tyrosine, methanethiol from cystine and/or methionine, and carbon disulfide from cleaving free carboxyl groups and disulfide bridges. It is thought that variation in amino acid composition causes differences in pyrograms of the individual hairs. If reproducibility is established, this method is likely to be used in forensic hair comparison. Hair treatments, especially hair bleaching, were reported to have influenced the chemical and physical nature of hair. Infrared (IR) spectra from a single hair can be measured by Fourier transform infrared analysis for detecting the oxidation of hair. There is an absorbance at  $1042\text{ cm}^{-1}$  not only in bleached hair, but also in permanent waved hair and dyed hair. This absorbance arises from sulfonic groups that result from the cleavage of disulfide bridges in hair. The oxidation occurs more strongly in the hair tip than in the root. This phenomenon also arises from weathering and ultraviolet irradiation. Though it is usually difficult to establish whether a suspect's hair has been treated or weathered from IR spectra, IR analyses can provide evidential clues for hair comparison without consuming the sample. It is possible to investigate differences between chemical oxidation of a disulfide bond and weather-beaten oxidation in a single hair by electron spectroscopy for chemical analysis, because the bonding energy of sulfur atoms differs under the various conditions. A method of dating samples by electron spin resonance (ESR) spectroscopy has been

developed and applied to a wide variety of materials. Red melanin (phaeomelanin), which does not have unpaired electrons, and black melanin (eumelanin), which has unpaired electrons, show different spectra; thus, the intensity of the ESR signal of brown hair is weaker than that of black hair. The ESR signal of cut hair reduces according to the time since cutting. As the radical in the hair of 1 cm long can be detected by ESR analysis, this method will be applied to the field of forensic science. Derivatives such as amino-hydroxyphenylalanine and pyrrole-2,3,5-tricarboxylic acid, respectively, produced from phaeomelanin and eumelanin in human hair ( $\sim 4\text{ mg}$ ) are determined by liquid chromatography (LC). The contents of both melanins can be determined using LC and provide a clue for hair identification.

## Biochemical Analysis of Hair

### Intracellular Enzymes of Hair Root

It has become clear that many intracellular enzymes exist in the root sheath of human hair. They are alcohol dehydrogenase (ADH), esterase D (EsD),  $\alpha$ -L-fucosidase (FUC), glucose-6-phosphate dehydrogenase (G6PD), 6-phosphogluconate dehydrogenase (6PGD), phosphoglucomutase (PGM), etc. The biochemical properties of these enzymes in hair root are very similar to those in other tissue cells. Some of these enzymes have different proteins with similar enzymatic activity. They are called isoenzymes. These isoenzymes show 3–10 phenotypes. PGM<sub>1</sub> isoenzyme is frequently utilized for forensic hair identification, because it shows more phenotypes than other isoenzymes (Table 1). Many isoenzymes are detected by means of isoelectric focusing with polyacrylamide gel having a high separating ability. Therefore, polymorphism of these enzymes can be one item of information like ABO blood groups for hair identification if the activity of the hair root enzyme is observed.

### Electrophoresis of Hair Keratin Protein

Forensic hair identification has been investigated using the electrophoretic patterns of hair keratin

**Table 1** Representative isoenzymes in human hair root

<i>Isoenzyme system</i>	<i>Numbers of phenotype</i>	<i>Methods of electrophoresis</i>
ADH	3	Isoelectric focusing
EsD	3	Isoelectric focusing
FUC	3	Isoelectric focusing
G6PD	4	Starch gel or polyacrylamide gel
6PGD	3	Starch gel
PGM <sub>1</sub>	10	Isoelectric focusing

protein, which have a variety of composed peptides and a genetic polymorphism. Hair keratin, a polymer of amino acids, has a firm structure and is not dissolved in water, common buffers, or organic solvent. In order to solubilize hair keratin, the bridge linkage of disulfide bond is cleaved by reductant (e.g., 2-mercaptoethanol), and sulfhydryl groups produced by the reduction are alkylated by iodoacetate for the purpose of preventing reoxidation. As a result of that, a soluble derivative of *S*-carboxymethylkeratin corresponding to native protein in hair is obtained, and becomes a more soluble substance suited for electrophoresis by adding urea (protein-denaturing reagent).

Keratin proteins in human hair are composed of low-sulfur proteins (mol. wt.  $\sim 50\,000$ – $76\,000$ ) of cortical fibrils and high-sulfur proteins (mol. wt.  $\sim 25\,000$ – $43\,000$ ) of matrix among fibrils. Phenotype frequency of low-sulfur protein variant is  $\sim 4\%$  in Caucasian hairs detected by 10% polyacrylamide gel electrophoresis, but its variants have not been detected in Mongoloid (Japanese) hair. On the other hand, high-sulfur protein variants have been detected in  $\sim 30\%$  of Japanese. These variants are considered to be inherited in autosomal dominant fashion. Furthermore, a two-dimensional electrophoresis system has been developed for obtaining detailed information of these hair proteins. More different patterns can be observed by this analysis than by the one-dimensional system, and are said to arise from nutritional, physical, and environmental factors and disease in individuals. For example, a two-dimensional system using acidic (pH 3) polyacrylamide gel–sodium dodecyl sulfate provides eight variable protein groups or spots, in which some additional spots or their absence are related to those factors. These electrophoretic analyses of hair proteins are also effective in the identification of animal species.

### Amino Acids

Remarkable amino acid analyzers have been developed and can analyze one piece of hair. Variation of amino acid composition in human hair is caused by nutrition, disease, and treatments (permanent waving, bleaching etc.). The effect of cosmetic treatments such as bleaching, dyeing, or permanent waving on the amino acid composition is to decrease the content of cystine and increase the competent of cystic acid depending on the extent of the treatment. Thioacetylated lysine, carboxymethylthiol cystine, dithiodiglycolic acid, and methionine sulfone have been found in permanent waved hair. This information on variation in amino acid analysis of a single hair may be helpful for forensic hair comparison.

### DNA Analysis

DNA fingerprinting was developed for individualization system in 1985. Analyses using DNA profiling with variable numbers of tandem repeat polymorphism has been carried out for identification of forensic samples such as bloodstains and sectional stains. Concerning forensic hairs, several DNA analyses of hair root sheath cells have been reported. However, DNA analysis of a hair shaft has not succeeded to date, because DNA recovery from a hair shaft is in order of tens of picograms and only low molecular weight DNA (below 200 base pairs (bp)) is left in the hair shaft. Recently, microsatellite DNA polymorphism has been detected by the polymerase chain reaction technique. This polymorphism can be applied to DNA analysis of hair shaft. If DNA analysis is combined with microscopic and instrumental analysis in forensic examination, hair individualization can be more accurate.

### Mitochondrial DNA Analysis

Human identification by testing the nuclear DNA within cells has become a powerful and sensitive tool. It can be used with those hairs that possess some tissue, usually in the form of cells from the hair root or from sheath cells that adheres to forcibly removed anagen hairs. The presence of these tissues often requires microscopy to find them and in some instances a result can be obtained from the hair shaft. However, in many instances such tissue is not present, such as the case of telogen hairs that have been shed naturally, and there is insufficient material for tests of nuclear DNA to be productive. An alternative strategy is to test the DNA content of cell organelles called mitochondria. These are present in the cytoplasm of the cells and take part in the cellular respiratory mechanisms to provide energy for cellular processes. They are present in high numbers within the cells in tissues that are dividing rapidly. Each particle contains multiple copies of the mitochondrial DNA, or mtDNA, with the result that the copy number of the mtDNA can be several thousand-fold greater than the nuclear DNA.

The dermal material in the hair root where the hair shaft is produced is an example of such tissue. As the shaft is created and subsequently keratinized the mitochondrial particles are trapped in the shaft and persist in the hair. The mtDNA thus presents a more promising target for DNA testing and indeed the exploitation of this feature of hair has become an important part of forensic hair analysis.

### Haplotypes and Databases of mtDNA

The mtDNA genome occurs as a single, double-stranded, DNA loop. The bases making the loop are

numbered one to 16569, representing the entire loop. Two parts of the loop have been found to possess greater variety in the DNA base sequence than the rest of the loop. These two regions are called hypervariable regions I and II or HVRI and HVRII, respectively. The mtDNA sequence has been entirely determined and a widely recognized reference sequence, such the Cambridge Reference Sequence (CRS), can be used as a standard against which the sequence of any portion of the mtDNA of a sample is compared. There is no opportunity for recombination of the base sequences present on the mtDNA loop from generation to generation but mutations do occur within the mtDNA, especially within the hypervariable regions, so that the base sequence changes at certain sites. These changes can be in the form of transitions, transversions, or insertions or deletions. A transition is a change of one purine to another purine, for example, an adenine base changed to a guanine, or a pyrimidine changed to another pyrimidine, e.g., a cytosine to a thymidine. A transversion is a change of a purine to a pyrimidine, or a pyrimidine to purine. Insertions or deletions of base sequences are called indels. The mtDNA loop hypervariable regions HVRI and HVRII are rich these variations. Many forensic laboratories typically sequence positions corresponding to the HVRI and HVRII regions.

A nomenclature system has been developed for mtDNA typing. It has been adopted as a convention within various bodies of the forensic science community. Simply put the rules are that the bases determined from a sample of mtDNA are compared to that of a reference sequence, for example, the CRS, at corresponding base positions. Where transitions or transversions have occurred, the change is denoted using the mtDNA base position number followed by the initial letter of the base found in the sample at that position. Only those positions that differ from the reference sequence are stated. The CRS is based upon a human mtDNA sequence that has been found subsequently to possess uncommon base types at certain positions. For example, a specific position of the CRS may be an A while most people possess a G at that position. Therefore, many sample mtDNA determinations, should they include this position, will include the recorded variation of a G at that position because it differs from the reference sequence. Where there is concordance between the sample bases at a given position with those of the reference sequence then there will be no specific designation on the assumption that the base type of the sample is the same as reference sequence.

Indels are denoted by citing the base position at the 5' side of the indel and using a decimal point to

indicate an insertion; for example, specifying the base type as a capital letter of the base name following the position number, 85.1C or 85.1A. Where there is a mixture or an ambiguity then the letter N is used; however, the International Union of Pure and Applied Chemistry letter convention can be used. In the case of deletions again the 5' base position number is used, followed by the letter d; so, for example, a deletion might be denoted '85d'.

### Mitochondrial DNA Sequence Alignment

The comparison of sequence data with a reference sequence requires careful attention to sequence alignment due to the possibility of discrepancies between sequences of the samples and that of the reference sequence arising from indels. The convention for aligning sequences for comparison is given here in general terms but there are a number of special circumstances that are dealt with more thoroughly in the literature. The conventions propose that sequences should be aligned so as to minimize the number of differences between the sample sequence and the reference sequence. In cases where there are equally good alternative means of aligning the sequences the priorities should be to align by the indels first so that the displacement of portions of the sequence arising from deletions or insertions are allowed for, after which come translations and then transversions.

Once the alignment has been made the combination of variations occurring at different places on the mtDNA strand can be identified and sequences can be compared with the reference sequence and cross-comparisons between different hair samples made. In Caucasians, there appear to be an average of around eight positions within the HVRs where the nucleotides differ between unrelated people and around 15 differences between people of African descent. As there is no recombination of the mtDNA the combination of variations in the HVRs is inherited as a single entity called a haplotype. The mitochondrial databases therefore count the frequency of the different haplotypes.

The HVR regions possess the greatest number of variable sites in the mtDNA and the base sequence of the variable regions, or the major parts of them, of the questioned hair are compared to the corresponding sequences determined from the reference hairs. Although there may be potentially many different haplotypes the maternal inheritance of the mitochondrial genome will restrict the diversity observed in a population, as many individuals will possess co-ancestry by the maternal line. Nonetheless, in studies carried out of the haplotypes observed in different populations the number cooccurring in different

individuals is small. As databases grow and more mtDNA sequences are determined there will be continual adjustments to the frequencies observed for different haplotypes. Population databases consisting of the sequences from the mtDNA from unrelated individuals of different populations have been established.

Certain haplotypes share a high proportion of the types of nucleotides found at the variable sites. Phylogenetic studies can be used to identify groups of haplotypes that show a high degree of commonality. Such groupings are termed haplogroups and in Caucasians it has been found that 10 haplogroups account for 99% of the population and 90% of this proportion of the population are represented by five haplogroups.

### Heteroplasmy

The sperm head possesses no mitochondria and upon fertilization only the contents of the sperm head enter the egg cell. A consequence of this is that, in principal, at least, no paternal mtDNA is passed to the fertilized cell and only the maternal mtDNA is transmitted to the next generation so that in theory all the mtDNA sequences in a person should be the same, i.e., homoplasmy. However, in practice different populations of mtDNA may coexist within one person, i.e., heteroplasmy. In the course of embryonic development the number of mtDNA molecules is greatly reduced, perhaps to less than 10, so that the chance of a variant mtDNA haplotype, present in the original pool in only a small proportion of the total mtDNA, being passed on is reduced. This therefore acts as a brake, or bottleneck, on the accumulation of heteroplasmic variations in the pool of mtDNA types. As the reductive process does not select for particular mtDNA haplotypes heteroplasmic variants may be missing from some cells after the bottleneck. If different variants were present in the original pool then some cells may possess one type of variant and others a different variant. In most tissues there will be a mixture of mtDNA from different cells so that an analysis of the mtDNA haplotypes will reflect the heteroplasmic mixture if any is present. It is possible that one haplotype predominates to the extent that a different haplotype present may not be detectable but this is a separate issue. However, the analysis of the mtDNA from the shaft of a single hair in a forensic context may not represent a sampling of the different heteroplasmic haplotypes present in the donor.

The hair shaft is produced from a collection of rapidly dividing cells in the hair bulb matrix of germinal cells. These cells divide rapidly and contain many mitochondria. The new cells produced do not

divide again but swiftly differentiate into cuticle cells, cortical cells, and medullary cells. One member population of the germinal cell matrix is the melanocytes. These are pigmented cells that produce melanin and store it in melanosomes. These cells produce the pigment granules that give the color to the hair and have large energy requirements.

The melanocytes come from the neural crest, a group of cells in the neural tube of embryo. These cells move out to different places in the embryo to develop into a variety of types of adult cells. The melanocytes are distributed throughout the skin at random and constitute a different population of cells from the rest. They become a part of the germinal cell matrix.

The melanocytes divide slowly, unlike the other cells of the germinal matrix of the root bulb, but they produce dendritic arms that are assimilated by endocytosis into the precortical cells that are moving out of the germinal matrix. These precortical cells are undergoing keratinization as a part of the process of the formation of the cortex of the hair and that involves the intensive production of keratin protein filaments. Therefore, there are two sources of the mtDNA molecules present in the cortex of the hair, one from the germinal cells of the matrix of the root bulb and the other from the melanocytes. The proportions of each contribution of mtDNA molecules are not known; however, there will be very little opportunity for daughter mtDNA molecules to be produced in the differentiating cortical cells while the melanocytes are continuously producing more pigment and mitochondria. Consequently, there can be genetic drift in the proportions of mtDNA molecules from the two sources so that haplotypes of mtDNA observed along the shaft of the hair can be different in different places. Also, different hairs that have come from the same follicle can potentially have different haplotypes. Therefore, while mtDNA analysis is undoubtedly a powerful tool for the forensic analyst there are special considerations that have to be taken into account in any interpretation made.

Opinions have been expressed that microscopic examination of hairs should be regarded as a presumptive step to act as a screening procedure prior to mtDNA analysis. However, many reports also emphasize the importance of the combination of the two approaches because the contribution the microscopic comparison can make. The microscopic examination gives an independent comparison of different characteristics. The analysis of mtDNA remains an expensive procedure, and not all hair specimens yield a usable result. Microscopic examination can eliminate obvious mismatches such as many nonhuman hairs, it can be used on fragments



possibly too small for DNA testing, and it can yield other information such as an association with racial, Caucasian/European, African, or Asian, ancestries. It is necessary to use microscopic examination to obtain information about the possible body areas and it can also yield other information such as cosmetic treatment. The microscopist can also identify if cellular tissue is adhering to the hair. This is a potential source of material for DNA testing using the same system of combined tests of independent short tandem repeats, or multiplex STR analysis, as used with other biological materials. This test can produce very powerful evidence of association between a hair and a donor if there is a match of DNA types.

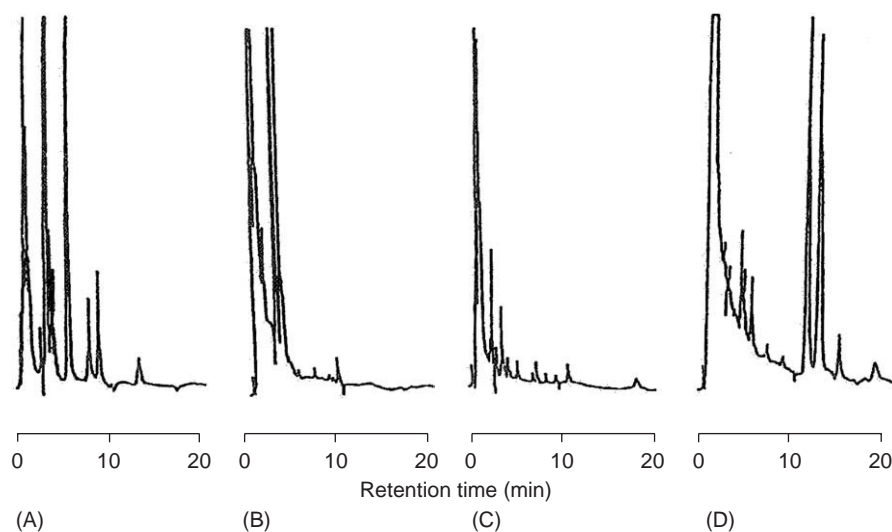
### Immunological Analysis of Hair

Keratin proteins and blood group substances such as antigens are identified in human hair, whereas immunoglobulin has not been detected in it. Blood group substances (A, B, and H) are contained mainly in the hair cortex. These substances have the chemical property of glycolipids. An absorption-elution technique for determination of ABO blood group from forensic hair sample has been developed. Using this technique it is possible to determine the ABO blood group from one single hair ~6 cm in length. Furthermore, the unlabelled antibody immunoperoxidase (peroxidase-antiperoxidase) method is utilized for determining the blood group of more

minute hair. The ABO blood grouping is used for discrimination in forensic hair comparison, especially when the analyst cannot distinguish the suspect's hair from the control sample hair with similar morphological features.

### Chemical Analysis of Components Remaining on or in Hair: Hair Shampoo, Hair Care Products, and Oxidative Dyes

People use hair shampoo, hair rinse, hair care products, and coloring preparations regularly for the purpose of hair beauty, and seldom change the brands of these products. Therefore, forensic hair identification becomes easier if these components remaining on or in human hair can be detected. The main components of hair shampoo are anionic (for instance, lauryl sulfate triethanolamine, sodium lauryl sulfate) or nonionic substances. Other components are amphoteric detergent, conditioner, agents increasing viscosity or foaming, dye, and perfume. Shampoo components remaining in 5–10 cm long hair are extracted with a solvent of acetonitrile and water, and these extracts are analyzed by LC. The liquid chromatograms for four individuals shown in **Figure 2** are different from each other. Because quantitative ratios of various components of shampoo products differ from ones of the hair's extracts, it is difficult to identify the brand of shampoo used. But LC

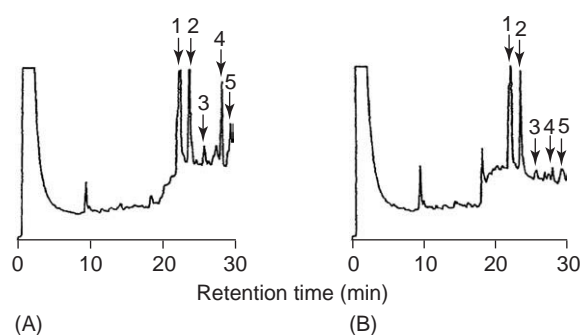


**Figure 2** Liquid chromatograms of hair extracts from four different persons. Extracting solvent: acetonitrile and water (70:30, v:v); equipment Varian model 5000 liquid chromatograph with Waters Associates fixed-wavelength absorbance detector (model 440) operating at 254 nm; column: 20 cm, 5  $\mu$ m Nucleosil C-18, 4.6 mm i.d. (Skandinaviska Gene Tec AB); mobile phase: acetonitrile and water (50:50); flow rate: 1.5 ml min<sup>-1</sup>. (Reproduced with permission from Andrasko J and Stocklassa B (1987). HPLC analysis of shampoo residues in human head hair. *Proceedings of the Eleventh Meeting of the International Association of Forensic Science*, vol. 20, pp. 249–255.)

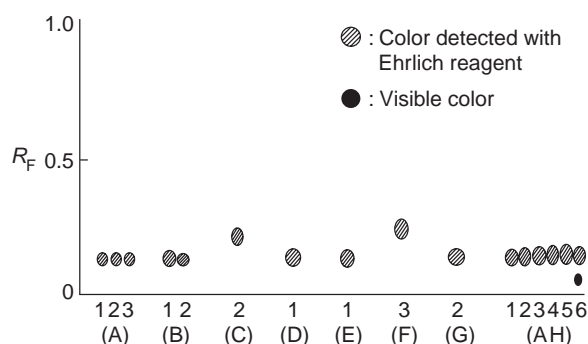
nondestructive analysis can give more objective and helpful information in individualizing of target hair among forensic hair samples. Identification of the brand of hair care products (men's and women's hair care products, oxidative dyes, hair sprays, and hair growth promoters) by components remaining on or in human hair is investigated. These components are identified mainly by GC, GC-MS, and thin-layer chromatography (TLC) analysis after extraction with ether or methanol. The main components of men's hair care products (hair liquids) remaining on human hair are identified as 2-hexyldecanol (emollient), isopropyl myristate (emollient), diethyl phthalate (perfume reserver), benzyl benzoate (perfume), methyl dihydrojasmonate (perfume), pentapropylene glycol (moisturizer), oxybenzone (sunscreen), etc. In the experiments, hair liquids of 57 brands produced by 17 companies are grouped into 21 classes on the basis of main components remaining on human hair, and among them hair liquids of 11 brands are identified by particular components (docosamethyldecasiloxane (polisher), isostearyl alcohol (emollient), diisopropyl sebacate (emollient), benzyl salicylate (sunscreen, perfume reserver), etc.). The major components of hair liquids remaining on human hair can be detected as long as 5 days after treatment. Other hair care products are also grouped similarly.

In the application of this method to an actual crime involving injury, components of hair care products remaining on hair adhering to a suspect's clothing were analyzed by GC and GC-MS, since morphological investigation had shown that some of these hairs were different from the victim's head hairs. The five main components remaining on hairs from both the victim and the suspect's clothing was found to be exactly the same. These gas chromatograms are shown in Figure 3. Two of these components were identified as dibutyl phthalate (perfume reserver or emollient) of peak 1 and the unidentified component of peak 4 originated from hair sprays which the victim used.

Hair coloring preparations are classified into four types: temporary dye, metallic (progressive) dye, ionic (acidic and basic) dye, and permanent (oxidative) dye. The oxidative dye is the most popular hair dye because of its color durability and easy application. Oxidative dyes are usually composed of dye compounds (A) solution and a hydrogen peroxide (B) solution. Both solutions are creamy. The A solution also contains strong alkaline ammonia to open hair cuticles. Dye components polymerize into indophenol-like trimers, etc., in the presence of hydrogen peroxide, primarily in the cortex of the hair. Hairs 3 cm long dyed by oxidative dyes of 34 brands are dissolved in the solution of methanol plus  $3 \text{ mol l}^{-1}$



**Figure 3** Gas chromatograms of components remaining on human hair adhering to suspect's clothing (A) and victim's head hair (B). Equipment: Shimadzu GC-4CM with a flame ionization detector column: 3 mm i.d.  $\times$  2 m, 1% Silicone OV-17 on Chromosorb-W (AW-DMCS, 80–100 mesh); column temperature: program from 100 to 280°C at 5°C min<sup>-1</sup>; carrier gas: nitrogen, 40 ml min<sup>-1</sup>.



**Figure 4** Thin layer chromatograms of oxidative dyes remaining in human hair. Capital letters (A–H) are assigned to indicate the companies of dye products. Numbers (1–6) indicate brands of products. TLC plate: Silica-gel 60 (20  $\times$  20 cm, 0.25 mm thickness, purchased from Merck Co.); developing solvent chloroform–ethyl acetate–methanol (6:3.5:0.5, v/v).

sodium hydroxide and sodium hydrosulfite dithionite), and these methanol extracts are analyzed by TLC. As shown in Figure 4, detectable amounts are present in hair dyed by 16 brands. These brands are classified into three groups from the identified components; i.e., in the first group only *p*-phenylenediamine (PPDA) is detected, the second group only 2,5-diaminotoluene sulfate and the third group PPDA and an unidentified compound that may be formed in the dyeing process are detected. Furthermore, after neutralization of the alkaline solution with  $3 \text{ mol l}^{-1}$  acetic acid, it is extracted with ether. From the extracts, other dye components, *o*-, *m*-, and *p*-aminophenols are identified by GC-MS after trifluoroacetylation. The identification of brands of oxidative dyes used has been no more successful than that of brands of shampoos used. But it is possible in

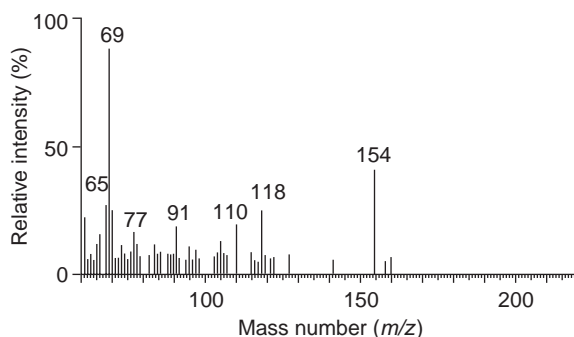


some cases to ascertain whether target hair is dyed or not by analysis of remaining components of hair coloring preparation.

## Drugs

If amphetamines are detected in a suspect's head hairs, the suspect may be arrested under stimulant control law and the hairs can be used as evidence in a trial. Furthermore, various drugs contained in human hairs not only provide evidence of crime under drug control laws but also important clues to hair comparison for other criminal investigations. Many cases of drug abuse have been established by using human hair. These drugs are stimulants (amphetamines), heroin, phencyclidine, cocaine, barbiturates, etc.

It is generally thought that blood capillaries break in bulb of hair root and drugs are taken into the hair while matrix cells produced in hair papilla grow and are keratinized. The drugs are contained mainly in hair cortex and amount to 50–100 ng mg<sup>-1</sup> single hair (equivalent to ~10 cm long). The drugs can be detected by radioimmunoassay and GC–MS. In the GC–MS analysis of amphetamines in human hairs, sample hairs of 10 mg are washed with 0.1% sodium lauryl sulfate and water three times under supersonication, and are dried. These hairs are dipped in the solvent mixture (2 ml) of methanol and 5 mol l<sup>-1</sup> hydrochloric acid (20:1) under supersonication. Then it is left to stand overnight at ambient temperature. The solvent mixture is separated, and is dried under a stream of nitrogen. The resulting residues are dissolved in the solvent mixture (200 µl) of trifluoroacetic anhydride and ethyl acetate (1:1), and are heated at 55°C for 20 min. The reaction mixture is dried under a stream of nitrogen. The residue



**Figure 5** Mass spectrum of trifluoroacetyl derivative of methamphetamine extracted from suspect's head hairs. Sample hairs: 100 pieces of 1 cm long hair; extracting solvent of hairs: methanol–5 mol l<sup>-1</sup> hydrochloric acid (20:1); equipment Shimadzu GCMS-OP 1100EX; column: Shimadzu capillary column HiCap-CBP10-W25-100, 0.53 mm i.d. × 25 m cyanopropyl (thickness 1.0 µm); column temperature: 140°C; carrier gas: helium 20 ml min<sup>-1</sup>; ion source temperature: 270°C; electron energy: 20/70 eV.

dissolved in ethyl acetate is subjected to GC–MS analysis. A mass spectrum of trifluoroacetyl derivative of methamphetamine extracted with methanol from suspect's hairs in a stimulant offences case is shown in Figure 5.

Compared with urine, hair samples have many advantages as an evidential material, because of easy sampling and handling, long stability time (approximately 3 days–1 year after intake), and high sensitivity (detectable at the nanogram per milligram level). Moreover, an individual history of drug intake can be worked out by analyzing each segment along the hair shaft from hair root to hair tip.

**See also:** **Activation Analysis:** Neutron Activation. **Amino Acids.** **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Cosmetics and Toiletries.** **Electron Spin Resonance Spectroscopy:** Biological Applications. **Electrophoresis:** Isoelectric Focusing; Polyacrylamide Gels; Two-Dimensional Gels. **Elemental Speciation:** Overview. **Forensic Sciences:** DNA Profiling. **Gas Chromatography:** Pyrolysis; Mass Spectrometry; Forensic Applications. **Immunoassays, Applications:** Forensic. **Polymerase Chain Reaction.** **Proteins:** Physiological Samples. **Surfactants and Detergents.** **X-Ray Fluorescence and Emission:** Energy Dispersive X-Ray Fluorescence.

## Further Reading

- Budowle B, Allard MW, Wilson MR, and Charkraborty R (2003) Forensics and mitochondrial DNA: applications, debates and foundations. *Annual Review of Genomics and Human Genetics* 4: 119–141.
- Clarence RR (1979) *Chemical and Physical Behavior of Human Hair*. New York: Van Nostrand Reinhold.
- Deedrick DW (2000) Hairs, fibers, crime, and evidence. *Forensic Science Communications* 2(3). (available online at [www.fbi.gov](http://www.fbi.gov)).
- Fujita Y, Nakayama M, Kanbara K, *et al.* (1987) Forensic chemical study on human hair I. Identification of brand of hair care products by components remaining on human hairs (1). *Eisei Kagaku* 33: 321–327.
- Houck MM (2002) Hair bibliography for the forensic scientist. *Forensic Science Communications* 4(1). (available online at [www.fbi.gov](http://www.fbi.gov)).
- Houck MM and Budowle B (2002) Correlation of microscopic and mitochondrial DNA hair comparisons. *Journal of Forensic Science* 47.
- Linch CA, Whiting DA, and Holland MM (2001) Human hair histogenesis for the mitochondrial DNA forensic scientist. *Journal of Forensic Science* 46(4): 844–853.
- Monson KL, Miller KWP, DiZinno JA, and Budowle B (2002) The mtDNA population database: an integrated software and database resource for forensic comparison. *Forensic Science Communications* 4(2). (available online at [www.fbi.gov](http://www.fbi.gov)).

- Ogle RR and Fox MJ (1999) *Atlas of Human Hair Microscopic Characteristics*. Boca Raton, FL: CRC Press.
- Prieto L, *et al.* (2003) The 2000–2001 GEP-ISFG collaborative exercise on MTDNA: assessing the cause of unsuccessful mtDNA PCR amplification of hair shaft samples. *Forensic Science International* 134: 46–53.
- Rowe WF (2001) The current status of microscopical hair comparisons. *The Scientific World* 1: 868–878.
- Saferstein R (ed.) (1982) *Forensic Science Handbook*. New Jersey: Prentice-Hall.
- Seta S, Sato H, and Miyake B (1988p) In: Maehly A and Williams RL (eds.) *Forensic Science Progress*, vol. II, pp. 47–166. Berlin: Springer.
- Tully G, Bar W, Brinkmann B, *et al.* (2001) Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretations of mitochondrial DNA profiles. *Forensic Science International* 124: 83–91.

## Illicit Drugs

N N Daéid, University of Strathclyde, Glasgow, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction – Types of Drugs

Drugs of abuse can be divided into different types according to their method of production or isolation. Natural drugs are active constituents and/or secondary metabolic products, of plants and other living systems, which are isolated by some method of extraction. Morphine is one such example. Semisynthetic drugs are products from natural sources, which have undergone some chemical refinement such as diamorphine, which is the diacetylated reaction product of morphine. Synthetic drugs are artificially produced substances for the illicit market, which are almost wholly manufactured from chemical compounds in illicit laboratories (for example amphetamine). Some forms of synthetic drugs are also called designer drugs and are substances whose molecular structure has been modified from the original in order to optimize their effect on the one hand and bypass laws and regulations governing the control of substances on the other.

### Natural and Semisynthetic Drugs

#### Opiates

Opium is a natural product obtained from unripe poppy capsules. There are over 100 species of the genus *Papaver* (or poppy) known, however only two, *Papaver somniferum* and *Papaver setigerum* D.C., are the main plants cultivated in the production of opium.

An incision is made in the poppy capsule and the latex, which oozes from the incision, is scraped off and collected to produce a raw opium gum. When fresh, the material is a sticky tar like brown substance

with a liquorice like odor which becomes brittle as it dries. Raw opium is also obtained from poppy straw. Opium is a complex mixture containing sugars, proteins, lipids, and water as well as the active alkaloid compounds (10–20%). About 40 alkaloids are known and fall broadly into two categories:

- Phenanthrene alkaloids such as morphine, codeine, and thebaine.
- Isoquinoline alkaloids such as papaverine and noscapine.

The relative amounts of these alkaloids varies and is dependent on factors such as climate, soil fertility, altitude, available moisture, and age of the plant. Raw opium can be used for smoking but must first be extracted most commonly with water. Morphine can still be detected in the left over dross that remains in the opium pipe.

**Morphine production** Several purification processes can be used in the preparation of crude morphine from raw opium, two of which are highlighted below. Not all production laboratories may use these methods.

1. The lime method involves the addition of calcium hydroxide (lime) to opium in water. This completely extracts morphine and codeine and partially extracts thebaine and papaverine. Ammonium chloride, alcohol, and ether are added and the solution filtered. Morphine precipitates while thebaine and papaverine remain in the ether layer. The crude morphine is purified by refluxing in dilute sulfuric acid in the presence of charcoal. The final morphine base is precipitated as an off white material by addition of ammonium hydroxide.

2. Raw opium is dissolved in water and the pH adjusted to 6.5 to precipitate noscapine and papaverine, which is removed. The solution is then made alkaline to precipitate morphine and extracted

- Ogle RR and Fox MJ (1999) *Atlas of Human Hair Microscopic Characteristics*. Boca Raton, FL: CRC Press.
- Prieto L, *et al.* (2003) The 2000–2001 GEP-ISFG collaborative exercise on MTDNA: assessing the cause of unsuccessful mtDNA PCR amplification of hair shaft samples. *Forensic Science International* 134: 46–53.
- Rowe WF (2001) The current status of microscopical hair comparisons. *The Scientific World* 1: 868–878.
- Saferstein R (ed.) (1982) *Forensic Science Handbook*. New Jersey: Prentice-Hall.
- Seta S, Sato H, and Miyake B (1988p) In: Maehly A and Williams RL (eds.) *Forensic Science Progress*, vol. II, pp. 47–166. Berlin: Springer.
- Tully G, Bar W, Brinkmann B, *et al.* (2001) Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretations of mitochondrial DNA profiles. *Forensic Science International* 124: 83–91.

## Illicit Drugs

N N Daéid, University of Strathclyde, Glasgow, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction – Types of Drugs

Drugs of abuse can be divided into different types according to their method of production or isolation. Natural drugs are active constituents and/or secondary metabolic products, of plants and other living systems, which are isolated by some method of extraction. Morphine is one such example. Semisynthetic drugs are products from natural sources, which have undergone some chemical refinement such as diamorphine, which is the diacetylated reaction product of morphine. Synthetic drugs are artificially produced substances for the illicit market, which are almost wholly manufactured from chemical compounds in illicit laboratories (for example amphetamine). Some forms of synthetic drugs are also called designer drugs and are substances whose molecular structure has been modified from the original in order to optimize their effect on the one hand and bypass laws and regulations governing the control of substances on the other.

### Natural and Semisynthetic Drugs

#### Opiates

Opium is a natural product obtained from unripe poppy capsules. There are over 100 species of the genus *Papaver* (or poppy) known, however only two, *Papaver somniferum* and *Papaver setigerum* D.C., are the main plants cultivated in the production of opium.

An incision is made in the poppy capsule and the latex, which oozes from the incision, is scraped off and collected to produce a raw opium gum. When fresh, the material is a sticky tar like brown substance

with a liquorice like odor which becomes brittle as it dries. Raw opium is also obtained from poppy straw. Opium is a complex mixture containing sugars, proteins, lipids, and water as well as the active alkaloid compounds (10–20%). About 40 alkaloids are known and fall broadly into two categories:

- Phenanthrene alkaloids such as morphine, codeine, and thebaine.
- Isoquinoline alkaloids such as papaverine and noscapine.

The relative amounts of these alkaloids varies and is dependent on factors such as climate, soil fertility, altitude, available moisture, and age of the plant. Raw opium can be used for smoking but must first be extracted most commonly with water. Morphine can still be detected in the left over dross that remains in the opium pipe.

**Morphine production** Several purification processes can be used in the preparation of crude morphine from raw opium, two of which are highlighted below. Not all production laboratories may use these methods.

1. The lime method involves the addition of calcium hydroxide (lime) to opium in water. This completely extracts morphine and codeine and partially extracts thebaine and papaverine. Ammonium chloride, alcohol, and ether are added and the solution filtered. Morphine precipitates while thebaine and papaverine remain in the ether layer. The crude morphine is purified by refluxing in dilute sulfuric acid in the presence of charcoal. The final morphine base is precipitated as an off white material by addition of ammonium hydroxide.

2. Raw opium is dissolved in water and the pH adjusted to 6.5 to precipitate noscapine and papaverine, which is removed. The solution is then made alkaline to precipitate morphine and extracted

with an organic solvent (such as ether) to remove thebaine and codeine.

**Heroin production** Morphine free base is refluxed in acetic anhydride for ~5 h. The mixture is allowed to cool and is neutralized with sodium carbonate causing the precipitation of crude diacetylmorphine free base, the illicit drug in heroin samples. Monoacetylation of morphine and acetylation of any codeine present also occurs. The crude product is purified by dissolving in acidified water containing charcoal and the free base precipitated by the addition of sodium carbonate. If the hydrochloride salt is required the free base is dissolved in acetone and the product precipitated by the addition of hydrochloride acid.

Because of the nature of the raw opium starting material, the production process, and the addition of excipients to produce 'street' heroin samples, there can be a considerable intersample variation (Figure 1).

**Worldwide heroin production** Southwest and Southeast Asia are the principal regions of illicit poppy cultivation, despite eradication and crop substitution programs carried out in these regions. In Southwest Asia, the main area of production is along the borders of Afghanistan and Pakistan in an area known as the Golden Crescent. In Southeast Asia, Myanmar is the major opium producing country. This area is known as the Golden Triangle (where Myanmar, Thailand, and Laos join). In the western hemisphere, Mexico and Colombia are the principal producers of opium. Enormous quantities of morphine base, produced in laboratories in the Helmand province of Afghanistan, are also transported overland through Iran to Turkey for conversion into diamorphine and heroin.

**Traffickers** Heroin is frequently trafficked through Asian airports like Bangkok, Mumbai, Delhi, Tehran, Karachi, Islamabad, and Lahore to smuggle heroin in body cavities or household goods. European nationals are recruited as couriers to bring heroin by

air from Southeast Asia to Europe. Moscow has developed into a hub of activity and from there trafficking is organized in Ukraine, Romania, and some Central Asian Republics. Heroin is also trafficked from Pakistan and Thailand by mailing parcels usually by express mail services to various destinations in North America and Europe. The drug is also smuggled from Turkey to Greece and into other European countries.

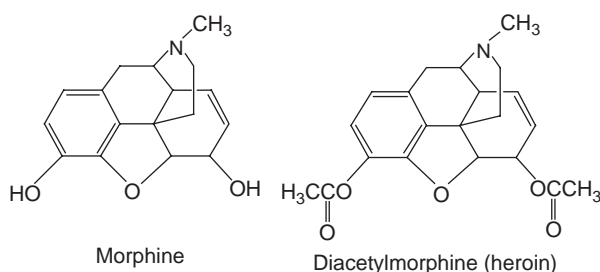
**Heroin trafficking in the USA and Australia** For a number of years, heroin in the USA arrived mainly from Southeast Asian countries, which had well-knit networks in the USA. Australia and North America are identified as two of the major international heroin markets for Golden Triangle heroin as well as the East and Southeast Asian countries. In Australia, the drug is mainly transported by boat from Ho Chi Minh city in Vietnam. Most of the heroin is imported into Sydney for distribution to the rest of Australia by road.

**Heroin trafficking in the European Union** The Central Asian Republics (Kazakhstan, Uzbekistan, Kyrgyzstan, Turkmenistan, and Tajikistan) are becoming a focal point for opium and heroin traffic from Afghanistan bound for Russia and Western Europe (Balkan route or Silk route). More than 65% of the Afghan opiates are exported via these countries. Small quantities of heroin are smuggled frequently by land vehicles principally train, bus, and truck. Since 1998, airlines are also being exploited by the trafficking organizations. Female couriers are frequently recruited for trafficking, concealing heroin in baggage or within their bodies. Moscow and other Russian cities are the usual rail and air destinations of this traffic. Heroin is further trafficked to the West European countries from Russia.

The organization of heroin traffic in Europe is controlled by Turkish and Albanian criminal syndicates. The latter were particularly active in Switzerland, Norway, Sweden, and Balkan countries. From Eastern Europe heroin is mainly distributed to Belgium, the Netherlands, Iberia, and the UK in routes controlled by Turkish criminal organizations, or to Denmark, Germany, Austria, and Scandinavia in routes controlled by Albanian and Yugoslav criminal organizations.

### Cocaine

Cocaine is derived from ecgonine alkaloids in the leaf of the Coca (*Erythroxylon* species) plant. There are various varieties of plant, which produce leaves of different size and appearance. In all species the upper



**Figure 1** Morphine and heroin.



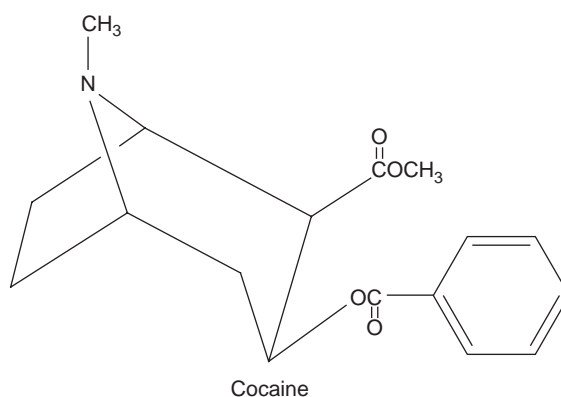
side of the leaf is darker in color. The underside of the leaf has two lines running parallel to the midrib of the leaf and are considered characteristic of the Coca leaf.

**Coca paste production** Coca leaves are mixed with calcium hydroxide (lime) and water. The mixture is crushed and stirred in a hydrocarbon solvent, usually kerosene. The extracted coca leaf residue is removed and the kerosene extracted with acidified water. The cocaine alkaloids are extracted into the aqueous layer and coca paste is precipitated by the addition of base. This paste contains crude cocaine as well as a mixture of inorganic salts. Alternatively, the leaves can be crushed in dilute sulfuric acid, extracted with kerosene, and the aqueous layer made alkaline with ammonia or a similar method to precipitate the alkaloids.

**Cocaine production** Coca paste is dissolved in sulfuric acid to hydrolyze the alkaloids and produce ecognine. This is treated with 10% boron trichloride to produce ecognine methyl ester. Potassium permanganate is also added until the solution turns and remains pink. This removes cinnamoylcocaine isomers, which may be present in the mixture. The mixture is filtered, and the addition of base to the filtrate precipitates cocaine free base known as 'crack' cocaine. The free base is dissolved in ether and precipitated as the hydrochloride salt through the addition of hydrochloric acid and is called 'snow' cocaine.

No two illicit samples of cocaine will be exactly the same though the variation from sample to sample is much less than for heroin. In most cases the drug appears as a fine white powder. Most cocaine, which is trafficked, contains ~30% cocaine hydrochloride with other substances such as lidocaine, procaine, or benzocaine added and the sample bulked with lactose, sucrose, or manose. In some circumstances cocaine is present as large white crystals and known as rock cocaine (Figure 2).

**Worldwide cocaine production** The major producers of cocaine are Bolivia, Columbia, and Peru, which produce ~700 tons of cocaine per annum. Of this, approximately 40 tons is destined for the European market and the remainder trafficked to the USA. The results of eradication programs and aggressive antinarcotics policies in Colombia, Peru, and Bolivia have caused growers and traffickers to search for new cultivation areas and transportation routes in the region. Traffickers are also using advanced technology to obtain better yields with each harvest; studying new methods to side step the governmental



**Figure 2** Cocaine.

measures such as neutralizing the effects of aerial spraying, and recycling chemical precursors which are becoming more difficult to obtain due to the stricter importation/exportation controls.

**Trafficking** Criminal groups operating from South America smuggle cocaine and heroin into North America, principally using land routes through Mexico, sea routes through the Caribbean corridor, and international flights. Although traffickers operating from Colombia control wholesale cocaine distribution throughout the heavily populated Northeastern United States, traffickers operating from Mexico control wholesale cocaine distribution throughout the West and Midwest of the country.

Africa continues to be an important transit point for cocaine from South America destined for domestic consumption and to European countries, mainly by use of the postal services, but also increasingly by sea containers. Nigerian criminal organizations operate widely in the African continent and they control the Sub-Saharan drug markets. Northern Africa is also used as a transshipment point for cocaine entering Europe and some evidence shows direct connections between local European drug traffickers and South American cocaine gangs. Cocaine has not yet made great inroads in Southwest and Southeast Asia markets. But increases in tourism and international trade, and the corresponding growth of air and shipping traffic contribute to the smuggling of small quantities of cocaine for domestic consumption.

**Cocaine trafficking in the European Union** Trafficking organizations target countries in Eastern and Southern Europe as intermediary stopovers for both sea and air freight consignments filled with large quantities of cocaine, which were ultimately destined to Western Europe. Analysis suggests an increasing presence of transnational criminal groups controlling

each step in the cocaine trafficking, such as Nigerian and Albanian trafficking organizations, which are becoming increasingly active in the importation and distribution of cocaine in Europe. Cocaine is smuggled both by sea and air in cruise liners, yachts, fishing boats, recreational vessels, and small planes specially adapted for drug smuggling. Most cocaine is transported to the European Union (EU) from Columbia via the Caribbean, Suriname, and into Belgium, France, the Netherlands and the Iberian Peninsula often along old historical trading routes. Other routes include importation through the Balkans and Eastern Europe. Chemicals used in the processing of cocaine are transported by European crime groups back to Columbia via these same routes.

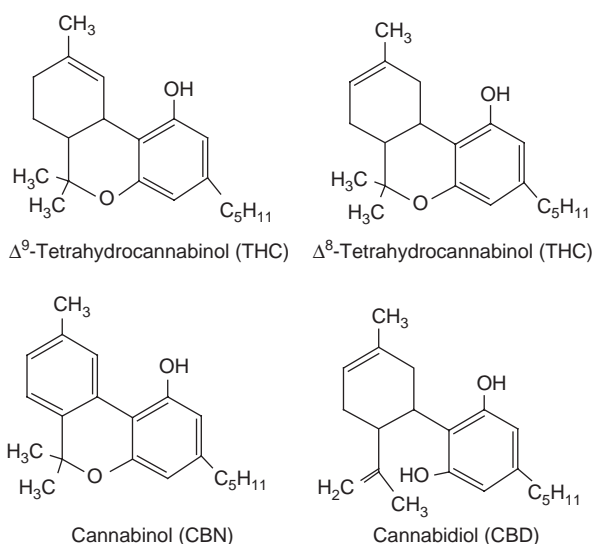
### Cannabis

The botanical name of cannabis is *Cannabis sativa* and it is often referred to as marijuana or grass (herbal cannabis) and hashish or hash (cannabis resin), terms associated with cannabis grown for its illicit use, or as hemp, a term usually associated with cannabis plants grown for their fiber content.

Cannabis is native to the mountainous areas of Central and South Asia and the plant has been used by man for over 6000 years. Cannabis grows over a wide variety of geographic terrains, altitudes, and latitudes. It is grown in many countries and on all continents. Although it prefers the higher temperatures and longer growing seasons of the equatorial areas, it has been cultivated as far north as 60° latitude.

Cannabis and its related products comprise over three-quarters of the drugs material submitted to Forensic Science Laboratories in the UK for analysis. A number of forms of the drug may be encountered, including plant material, resin, and 'hash oil'. The active ingredient in all of these formulations is  $\Delta^9$ -tetrahydrocannabinol (THC). Also found is  $\Delta^9$ -tetrahydrocannabinolic acid which is converted to THC through smoking and the compounds  $\Delta^8$ -tetrahydrocannabinol (an isomer of the active ingredient), cannabidiol (the precursor), and cannabinol (the degradation product) (Figure 3).

**Fresh plant material** Cannabis plant material can occur as live plants, or as dried plant material. It is important when examining plant material that it should be in the dried state as water content will effect any weight measurements, a crucial factor in determining the seriousness of any charges. Male and female cannabis plants are separate with the male plant flowering before the female to ensure cross-pollination. The plant grows between 30 cm and 6 m in height and grows from seed to maturity in



**Figure 3** Main cannabinoids.

~3 months, though harvesting can occur after 2 months. The leaves are palmate with serrated edges and are opposite and alternate around a hollow four-cornered stem. The leaves are coated with upward pointing unicellular hairs called trichomes. Warty walled cystolithic trichomes are found on the stems and leaves together with a few glandular trichomes, which contain the THC. The greatest concentration of glandular trichomes is in the flowering tops of the plants and those of the female plant have a greater concentration of THC. Material prepared from the flowering tops or leaves usually contains 0.5–5% THC.

**Dried plant material** This may occur in various ways. Low quality products contain stalks, seeds, leaves, and flowering tops. This may be compressed into blocks (West African and Caribbean material) or as loose material or rolled up and wrapped in vegetable leaves (Central and Southern Africa). High quality material contains mostly flowering tops and can be found rolled up or wrapped around bamboo sticks ('Buddha' or 'Thai' sticks) or sieved ('Kif') and generally originates from Southeast Asia or parts of Africa.

**Resin (hashish)** This is also produced directly from the plant material. The glandular trichomes produce a resin, which is scraped from the surface of the plant and pressed into blocks. Hashish contains ~2–10% by weight of the active constituent,  $\Delta^9$ -THC. Most resin is produced either in the Indian subcontinent or parts of the Mediterranean. Usually between 100 and 400 mg of resin are used in a joint, though this varies from user to user. Mediterranean resin is made by threshing the herbal material, sieving to remove seeds



and stems, and the remainder compressed into slabs. The final material is light in color and quite brittle. Indian subcontinent resin is quite sticky and is removed from the plant by rubbing the plant and then molded into slabs. The final material is dark brown and sticky. When examining resin several facts are recorded, including whether or not the blocks seized fit together, any striation marks (cutting marks), which may be present and whether these can then be linked to a cutting instrument. The color and different layers within the resin block should also be examined and recorded.

**Hash oil** This is a manufactured product of cannabis and is produced by extracting the whole plant material using an organic solvent (usually alcohol or ether). On extraction the resulting oil can contain between 10 and 30% by weight of  $\Delta^9$ -THC. The oil is used in various ways including smoking with tobacco.

**Worldwide cannabis production and trafficking** It is estimated that 20 000–30 000 tons of cannabis are produced worldwide per annum. Cannabis trafficking has been increasing in response to a high and rising demand. Southern Africa (South Africa, Lesotho, Malawi, and Swaziland) is one of the largest producing regions exceeding that of Morocco, which used to be the world leader in production. Most of the cannabis grown in this part of the world is intended for local consumption, large shipments are being sent to Europe and North America. Mexico, the largest cannabis producing country in the Americas is the primary supplier of herbal cannabis to the USA, and Jamaica supplies large quantities of cannabis oil to Canada, either directly or through the USA.

**Cannabis trafficking in the European Union** Approximately 40 million Europeans use cannabis each year with an estimated 1 in 4 15–34-year olds having smoked cannabis either as herbal material or cannabis resin. Most resin is brought in through Spain from Morocco with lesser amounts transported from Pakistan, Afghanistan, and Central Asia via Eastern Europe. Herbal cannabis originates mainly from Columbia, Jamaica, and Africa (South Africa, Nigeria, and Ghana) and is transported by sea into Spain and Portugal for distribution throughout Europe. The Central Asian Republics, where vast fields of cannabis cover several hundreds of thousands of hectares, remain for the time being a major source of supply for the illicit Russian market. Much herbal cannabis is also home grown in Europe particularly using hydroponic techniques, especially in the Netherlands.

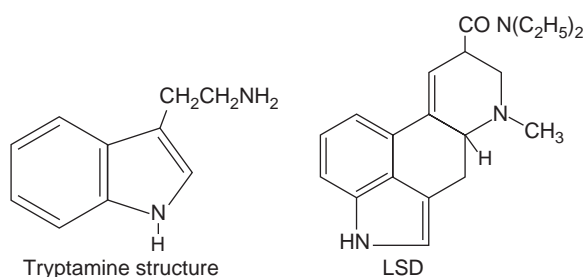
## Tryptamines and Lysergic Acid Diethylamide

Tryptamines are naturally occurring alkaloids found in a variety of plants and life forms around the world and exist in more than 1500 natural varieties. The basic element of tryptamine is the indol structure and tryptamine itself is an endogenous amine found in the human brain. Serotonin and melatonin are two essential tryptamines present as neurotransmitters in the brain. Tryptamines can be also produced either completely synthetically or semisynthetically (Figure 4).

Dimethyltryptamine (DMT), which is an active principle of various South American snuff, such as 'COHOBA' and 'YOPO'. It has been produced synthetically for a number of years, but its abuse has been restricted to a small number of dedicated users. Bufotin is very similar to DMT from a chemical point of view. It can be found either in the skin secretions of toads or in combinations with DMT in different trees in South America (for example, YOPO tree).

Psilocybin and psilocin are found in at least 15 species of mushrooms – so called 'magic mushrooms' – belonging to the genera *Psilocybe*, *Panaeolus*, and *Conocybe*. If a seizure contains whole or parts of mushrooms, the examination should include a description of the shape, size, and color of the fungus, as well as information regarding the gills, spores, cap, and stalk. Identification of fungi is difficult and should be performed by experts in that field. Illicit trafficking of cacti, which contain the psychoactive agent, mescaline, is also known. This drug occurs in at least three species of cacti. The most common illicit form of mescaline is produced as dried disk shaped cuttings taken from the tops of the cacti and called 'mescal buttons'.

Lysergic acid diethylamide (LSD) is a semisynthetic drug, which was first synthesized by Albert Hofmann in 1938. The starting materials are the lysergic acid compounds of the ergot, which is a spore capsule of a parasite mushroom. On the illicit drug market LSD has been sold in the form of impregnated paper (blotters/trips), microdots,



**Figure 4** Tryptamine and LSD.

thin squares of gelatine (window panes), or impregnated on sugar cubes. Stamps or blotters are the common dose form. They are made by impregnating paper with a solution of LSD in alcohol. These papers are 'trade marked' with various designs.

Analysis of such drug seizures should be carefully carried out and extreme caution should be taken as this drug readily penetrates the skin. Protective clothing should be worn at all times. If the presented drug sample is composed of sheets of absorbent paper the size and number of dose units, depth of perforations, and color and style of design should all be measured and recorded. If the drug is contained in powder or microdot form then it should be examined microscopically to identify any adulterants which may be present and can be subjected to further analysis. These usually include, dextrose, lactose, starch, talc potato starch, or maize. Dosage units can vary between 20 and 500 µg of LSD depending on the substrate used.

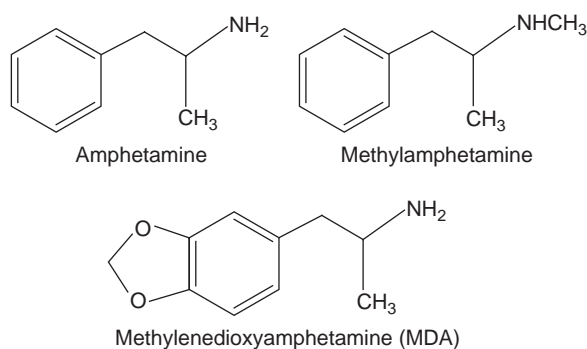
## Synthetic Drugs

### Amphetamine and Amphetamine-Type Stimulants

From a chemical point of view all amphetamine-type stimulants (ATS) are related to  $\beta$ -phenethylamine, which is the basic element of the body neurotransmitters (such as dopamine and adrenaline) that convey the neuronal information of the central and vegetative nervous system. Amphetamine was first synthesized in 1887 but was not used for medical purposes until the early 1930s, when it was found that it increased blood pressure, stimulated the central nervous system, was effective against asthma, and was useful in treating an epileptic seizure disorder. On the illicit drug market amphetamines have been sold in the form of powders, liquids, crystals, tablets, and capsules.

The most common amphetamine derivatives in the illicit drug market include amphetamine sulfate, methylamphetamine, methylenedioxy amphetamine (MDA), methylenedioxy methyl amphetamine (MDMA), and methylenedioxy ethyl amphetamine (MDEA). Dosages in powders and tablets vary considerably but are generally in the range of 50–250 µg of drug in a single dose. Most, if not all, amphetamine and related compounds are synthesized in clandestine laboratories by various synthetic routes (Figure 5).

**Ecstasy** When the term Ecstasy was first used in the early 1970s, it was an American street name for preparations containing the active agent MDMA. Now the term describes tablets or capsules predom-



**Figure 5** Amphetamines.

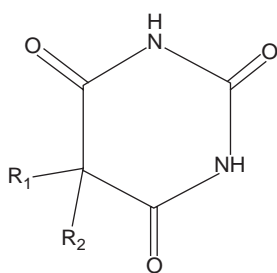
inantly containing one or more (or combinations) of psychotropic active agents derived from the  $\beta$ -phenethylamine group. It is becoming more and more common to compress amphetamine/methamphetamine into tablets, which are then also marketed as Ecstasy.

### Trafficking

*Amphetamine and Ecstasy use and trafficking in the European Union* Synthetic drugs particularly amphetamine and Ecstasy are the second largest type of drug consumed in Europe. Unlike with other drugs of abuse the EU is the major producer of these compounds and has a huge export market to other parts of the world particularly the USA, Canada, Australia, Asia, and Africa. Illicit amphetamine and Ecstasy laboratories have been found in 10 member states of the EU, particularly Belgium, the Netherlands, and the UK. Many laboratories are also known to exist in Eastern European countries such as Poland, Bulgaria, the Czech republic, and the Baltic states. These laboratories can vary from simple kitchens to professional laboratories with consequent variability in production capacity. Such laboratories also produce large quantities of chemical waste with associated disposal problems.

### Barbiturates

Barbiturates are therapeutically used as sedatives, hypnotics, anesthetics, and anticonvulsants. Virtually all barbiturates on the illicit drug market come from licit sources. There are 12 barbiturates recognized and scheduled by the United Nations and occur mainly as capsules and tablets and in some cases as injectable solutions. Illicit compounds occur as mixtures of barbiturates, or mixed with other drugs such as caffeine, aspirin, codeine, or in some cases heroin (Figure 6).



**Figure 6** Barbiturate structure.

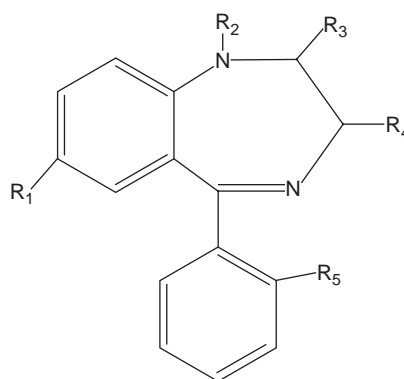
### Methaqualone and Meclaqualone

Methaqualone was first synthesized in 1951 and introduced as a new drug that produced sedation and sleep in 1956. Methaqualone has been initially designed to counter the nervous damages caused by long-term consumption and to reduce the risk of the dependency potential of barbiturates. Interest in methaqualone rose dramatically in recent years. Its popularity was due to its undeserved reputation as an 'aphrodisiac' often in combination with diphenhydramine. Because of its strong habit-forming properties the drug was placed in the list of controlled substances and the legal manufacturer stopped its production and removed it from the market in 1984.

Methaqualone and meclaqualone were prepared as nonbarbiturate sleeping tablets, though they have also legally been used as hypnotics in some European countries. They appear on the illicit drug market either through diversion from the legitimate pharmaceutical trade or through illicit synthesis. The illicit samples are usually brown or gray powders with varying degrees of purity. Methaqualone is also used sometimes as a cutting agent for heroin.

### Benzodiazepines

Benzodiazepines, therapeutically used as tranquilizers, hypnotics, anticonvulsants, and centrally acting muscle relaxants, rank among the most frequently prescribed drugs. In 1960, the first benzodiazepine, chlordiazepoxide, was introduced. To date, more than 50 benzodiazepines have been marketed in over 100 different preparations. They appear mainly as capsules and tablets, however, some are marketed in other forms such injectable solutions or powders. Benzodiazepines were introduced to replace barbiturates and methaqualone as tranquilizers, hypnotics, anticonvulsants, and muscle relaxants. Currently there are 33 benzodiazepines on the control list all of which appear as tablets or capsules, though some also appear as vials or powders for preparation of injection (Figure 7).



**Figure 7** Benzodiazepine structure.

On the illicit drug market, diazepam ('Valium'), temazepam (often referred to as 'jellies'), and flunitrazepam ('Rohypnol') are used as drug substitutes, as additives in drug-preparations, or in combination with alcohol, as methods of incapacitating individuals ('Rohypnol' is sometimes known as a 'date rape drug'). Virtually all of the benzodiazepines in the illicit market result from diversion from legitimate sources and there is no evidence of clandestine manufacture.

### Phencyclidine and Analogs

Phencyclidine (PCP) was synthesized and tested in the early 1950s and recommended for clinical trials as an anesthetic in humans in 1957. In 1965, further human clinical investigation of PCP was discontinued and the compound was marketed commercially as a veterinary anesthetic. PCP became available through the drug culture in the late 1960s, referred as 'PeaCePill', commonly sold as 'angel dust', 'crystal', or 'hog', on the illicit market in powder, tablet, leaf mixture, and 1 g 'rock' crystal forms, usually taken orally, by smoking, snorting, or intravenous injection. The laboratory synthesis of PCP and approximately 120 related substances such as Eticyclidin (PCE) or Tenocyclidin (TCP) is cheap but also work-intensive and time-consuming.

A structurally related anesthetic, ketamine ('Special K'), has been developed and has recently gained popularity as it is not controlled by legislation in most countries.

### Gamma-Hydroxybutyric Acid

Gamma-hydroxybutyric acid (GHB) is a drug that is very similar to a natural chemical in the human brain called gamma-aminobutyric acid (GABA). GHB is a simple sodium (or potassium) salt of 4-gamma-hydroxybutyric acid. The street names are 'liquid

E', 'liquid ecstasy', or 'fantasy'. The forms of use are tablets, white powder, or solutions (dissolved in water or other liquids). GHB (like Rohypnol) is sometimes added to alcoholic drinks as methods of incapacitating individuals and is also known as a 'date rape drug'.

Originally GHB was marketed as a supplement replacement for L-tryptophan, an amino acid. Since the early 1990s, it is being used illicitly by athletes and bodybuilders, believing that its growth hormone releasing effects contribute to anabolism and lipolysis, or misused as a sleep aid and for weight control. GHB can be very easily prepared from gamma-butyrolactone (GBL) – a common industrial chemical for paint removers – by alkaline hydrolysis. The clandestine manufacture of GHB requires no prior chemical expertise as evidenced by the simplistic

instructions given in the underground publications or on the Internet.

*See also:* **Forensic Sciences:** Drug Screening in Sport; Systematic Drug Identification; Thin-Layer Chromatography.

## Further Reading

- Cole MD (2003) *The Analysis of Controlled Substances*. Chichester: Wiley.
- Gough TA (1991) *The Analysis of Drugs of Abuse*. Chichester: Wiley.
- Laing R (2003) *Hallucinogens, a Forensic Drug Handbook*. New York: Academic Press.
- Moffat A, Osselson MD, and Widdop B (2003) *Clarke's analysis of drugs and poisons*, 3rd edn. London: The Pharmaceutical Press.

## Paints, Varnishes, and Lacquers

**P Buzzini**, Ecole de Sciences Criminelles, Lausanne, Switzerland

**W Stoecklein**, Kriminaltechnisches Institut, Wiesbaden, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

### Paint Evidence and Its Properties

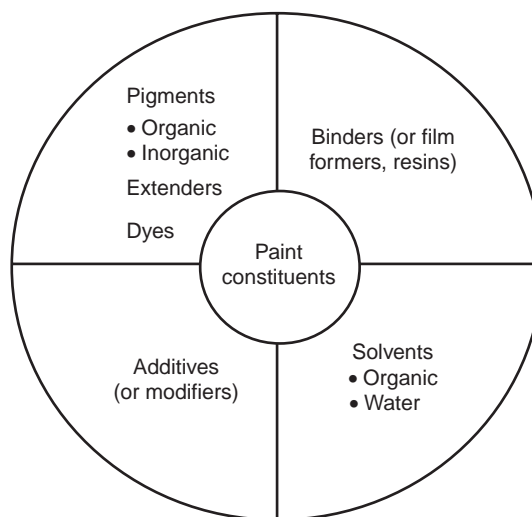
When two objects come in contact according to a given intensity, material will be transferred, persist, and can be recovered. If an object is coated, it is susceptible to leave its own paint and/or to import on its surface the one coming from the other object as traces in the form of multilayered flakes, chips, smears, or powder. The composition, form, quantity, and location of these traces can provide important information about their origin, the criminal action occurred, or the course of events in connection with crimes against people (e.g., hit-and-run accidents), property (e.g., vehicles road accidents, breaking and entering, vandalism), capital crimes (e.g., terrorist attacks, murder, etc.), and other offences involving the use of painted objects.

Paints are extremely complex in their composition. Depending on their intended use and the prescribed method of application, the paint formulations are composed of a multitude of individual components. These can be assigned to four main groups according to their functions showed in **Figure 1**.

The term paint is known to indicate a suspension of pigments contained in a nonvolatile vehicle liquid

used as a binder; on the other hand, the term coating designates a liquid or a liquefiable material converted to an opaque solid film applied to a surface. Pigment-free paint formulations based on drying oils and resins in organic solvents are referred to as varnishes. Lacquers are thermoplastic solution paints or varnishes which physically dry through ordinary evaporation of solvents rather than by oxidation or polymerization.

The primary purpose of paint is to hide the surfaces of the substrates and to protect them from corrosion and degradation. Therefore, important



**Figure 1** Main constituents characterizing paint materials formulation.

E', 'liquid ecstasy', or 'fantasy'. The forms of use are tablets, white powder, or solutions (dissolved in water or other liquids). GHB (like Rohypnol) is sometimes added to alcoholic drinks as methods of incapacitating individuals and is also known as a 'date rape drug'.

Originally GHB was marketed as a supplement replacement for L-tryptophan, an amino acid. Since the early 1990s, it is being used illicitly by athletes and bodybuilders, believing that its growth hormone releasing effects contribute to anabolism and lipolysis, or misused as a sleep aid and for weight control. GHB can be very easily prepared from gamma-butyrolactone (GBL) – a common industrial chemical for paint removers – by alkaline hydrolysis. The clandestine manufacture of GHB requires no prior chemical expertise as evidenced by the simplistic

instructions given in the underground publications or on the Internet.

*See also:* **Forensic Sciences:** Drug Screening in Sport; Systematic Drug Identification; Thin-Layer Chromatography.

## Further Reading

- Cole MD (2003) *The Analysis of Controlled Substances*. Chichester: Wiley.
- Gough TA (1991) *The Analysis of Drugs of Abuse*. Chichester: Wiley.
- Laing R (2003) *Hallucinogens, a Forensic Drug Handbook*. New York: Academic Press.
- Moffat A, Osselson MD, and Widdop B (2003) *Clarke's analysis of drugs and poisons*, 3rd edn. London: The Pharmaceutical Press.

## Paints, Varnishes, and Lacquers

**P Buzzini**, Ecole de Sciences Criminelles, Lausanne, Switzerland

**W Stoecklein**, Kriminaltechnisches Institut, Wiesbaden, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

### Paint Evidence and Its Properties

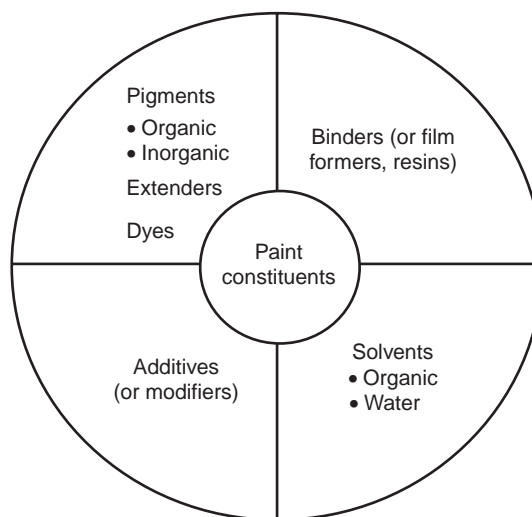
When two objects come in contact according to a given intensity, material will be transferred, persist, and can be recovered. If an object is coated, it is susceptible to leave its own paint and/or to import on its surface the one coming from the other object as traces in the form of multilayered flakes, chips, smears, or powder. The composition, form, quantity, and location of these traces can provide important information about their origin, the criminal action occurred, or the course of events in connection with crimes against people (e.g., hit-and-run accidents), property (e.g., vehicles road accidents, breaking and entering, vandalism), capital crimes (e.g., terrorist attacks, murder, etc.), and other offences involving the use of painted objects.

Paints are extremely complex in their composition. Depending on their intended use and the prescribed method of application, the paint formulations are composed of a multitude of individual components. These can be assigned to four main groups according to their functions showed in **Figure 1**.

The term paint is known to indicate a suspension of pigments contained in a nonvolatile vehicle liquid

used as a binder; on the other hand, the term coating designates a liquid or a liquefiable material converted to an opaque solid film applied to a surface. Pigment-free paint formulations based on drying oils and resins in organic solvents are referred to as varnishes. Lacquers are thermoplastic solution paints or varnishes which physically dry through ordinary evaporation of solvents rather than by oxidation or polymerization.

The primary purpose of paint is to hide the surfaces of the substrates and to protect them from corrosion and degradation. Therefore, important



**Figure 1** Main constituents characterizing paint materials formulation.

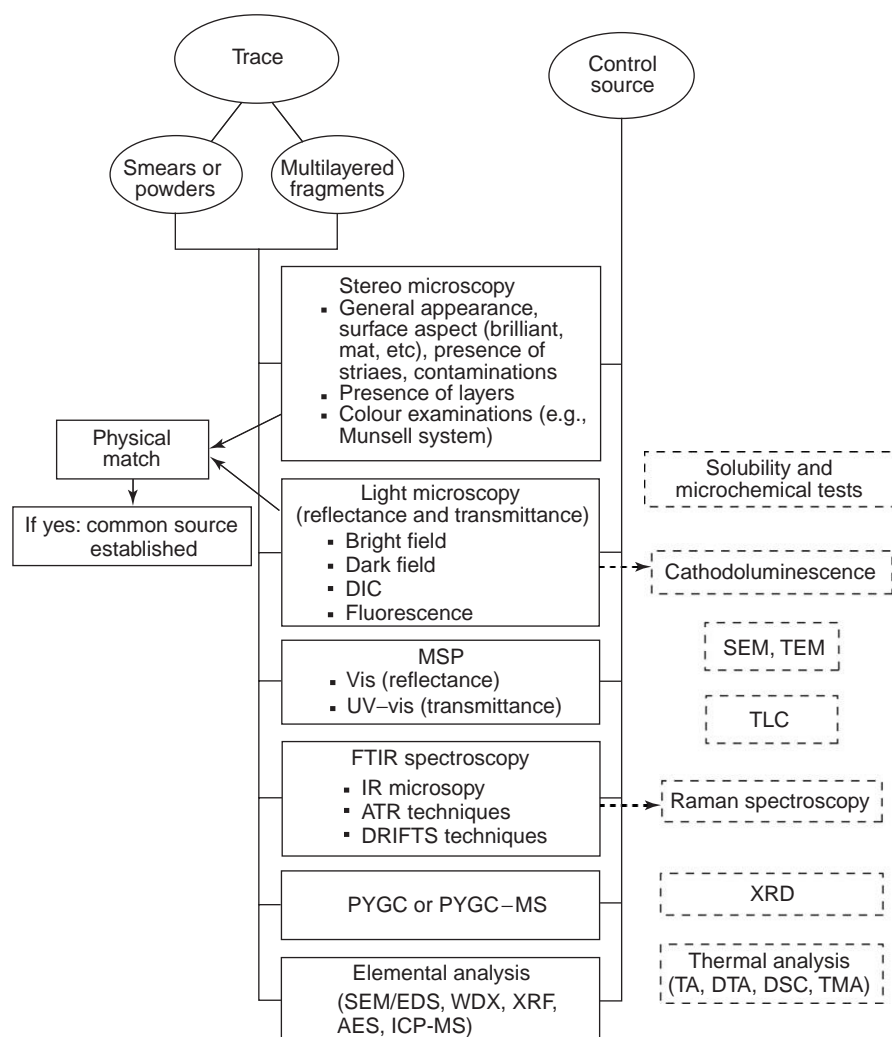


properties of such material are resistance to chemicals, abrasion, and weather conditions. The appearance of the surface with such features as gloss, color, or completely smooth or textured surfaces is almost as important. In order to guarantee lasting protection and decorative properties, the surfaces are often pre-treated in several stages and then layered in several additional stages with compatible paint materials (sealer, primer, primer-surfacer, topcoat). Thus, the final product in the sequence of operations is generally a multilayered coating system. During the period of service, additional layers of paint and surfacer may be applied on top of such systems during repairs (e.g., a painted vehicle body).

The main emphasis in forensic paint analysis lies in the comparison of materials: the physical and

chemical characteristics of a paint trace (suspect or unknown or questioned sample) and a control paint sample (or reference or known sample) are compared to establish if they have a common origin. A typical analytical sequence for paint is illustrated in Figure 2.

One of the main problems in comparative material analysis is how to interpret matching analytical data. Paints are mass products which are often produced in considerable annual tonnages: to come to the correct conclusions about the evidential value of the results, the forensic scientist must be able to distinguish between paints from batch to batch, to know information on the influence on analytical results of material ageing, and to find dealing with errors in application and defects in materials. Statistical data on the



**Figure 2** Comparative analytical sequence for paint evidence with its control paint known source. Such sequence depends on the quality and quantity of paint evidential material recovered, on the equipment of the laboratory and on the personal choice of the forensic scientist. At every stage of the sequence, if the questioned and the known samples result undifferentiated, one continues with the following examinations. If the trace and its control are differentiated with a given technique and if the differentiation cannot be explained, a common origin can be excluded.



frequency of occurrence expressing the rarity of certain combinations of paint components in a given relevant population are also included. The availability of a broad spectrum of microscopic techniques combined with methods of instrumental microanalysis is indispensable for forensic investigations with the objective of classification, identification, discrimination, or evaluation of a common source or moreover of a contact that occurred between painted surfaces.

A further important function is the determination of the origin of unknown coating systems (e.g., the make, model, and year of a vehicle from a fragment of paint recovered from the scene of a hit-and-run accident) in the aim to supply useful investigative information to the police. Other than classical vehicle ones, product coatings (e.g., tools or spray paints) or architectural (or household) paints can be typically submitted to a forensic laboratory for analysis. Such investigations are also necessary in the field of artistic paints in connection with art forgeries (e.g., certification of authenticity), for the determination of the age of coatings (e.g., anachronisms), or to supply useful information for the conservation or restoration of such paints.

## Sample Collection and Pretreatment Prior to Analysis

The detection and collection of paint traces (unknown or questioned samples) is fundamental because every mistake or forgetfulness at this stage will be irreparable. The form, the quantity and the quality of paint material collected will affect the number and the capacity of the available analytical methods.

For paint traces recovered on a solid surface as a vehicle, a doorway, a window frame, or a wall, paint fragments are isolated with tweezers; smears or powders are recuperated by scratching by means of a knife or a razor blade (recuperating the paint in a sheet of paper). Tape lifts should only be used if other methods fail.

Where the object is transportable (e.g., tools), it is suitable to bring it to the laboratory (covering the paint trace with papers) for a research under a stereo microscope and the following collection will be undertaken by means of tweezers or a metal point; if circumstances allow it, it would be useful to cut vehicle body or to collect a piece of the support as a wooden forced doorway or a tagged wall. If possible, in searching for paint traces, the scientist must consider the possibility of a cross-transfer. Paint fragments on textiles are searched with lens and under the stereo microscope using tweezers, and then

shaking recuperating fragments on a cleaned surface; finally a vacuum cleaner can be used.

Control paints (known samples) collection must be undertaken close (but not within) to the zone of contact where the foreign paint traces are sticking; this is important because of the changes of the paint system within big surfaces as those of vehicles.

Concerning the paint analysis, reproducible and correct measurements can only be achieved if the sample material is optimally prepared. In the preparation of paint traces it is essential that characteristic features such as common broken edges, striation marks, or texture effects on the upper or lower surfaces are not destroyed.

For multilayered paint fragments, in order to determine precisely the morphological features (number and thickness of layers, their color, luminescence, etc.), it is necessary to prepare cross-sections: the fragments are embedded in a suitable resin and either ground and polished in one step with the aid of an ultramilling cutter or processed in a grinding and polishing device. A resin based on mono- and difunctional methacrylates Technovit 2000 LC (Heraeus Kulzer, GmbH; hardening process 20 min by using blue light) and the epoxy resin Devcon 5 (hardening time: 15 min) are proven cold embedding materials. These materials dry quickly and do not react with paints. Microtome sections should be obtained for optical and scanning electron microscopic (SEM) examinations, for microspectrophotometry (MSP) and infrared (IR) spectroscopy.

## Optical Examination and Light Microscopy

Stereo microscope (magnification of 5–40 $\times$ ) is first used to observe the general appearance of the sample (shape, color, presence of layers, etc.) in order to establish if it can be a paint material. Color (intended as hue) is the first feature observed, being the most eye-catching one: this stage is subjective but it is useful for a first description and evaluation of that feature. There exist systems based on the classification in a 3D color space supplying labels or coordinates: some examples are the most used Munsell color coordinate system, the Standard color chart DIN 6164, or the Natural color system and the Muthen Handbook of Colours. Concerning automotive paints, Color collections distributed by the main paint producers (Du Pont, ICI, Nippon Paint, Glasurit BASF) have proved successful, including touch-up paints of standard car colors and nuances caused by weathering. The most useful today are the Akzo Colorscala (Akzo Coatings by, The

Netherlands) the Euro Colour Catalogue (Dupont Colour Book), and the Glasurit (Color-Profi-System, BASF, Germany). If the unknown and the control samples permit it, a physical match, even if very rare, must be searched by observing the edges and irregular contours of fragments and their surfaces striae (e.g., scratches).

Then, microscopic examinations follow: optical research microscopes allow to determine the number, thickness, and color sequence of layers in paint fragments, and to recognize the textures as well as fundamental features of pigment and extender mixtures. Bright field and dark field illuminations, polarized light microscopy (incident and transmitted), particularly the differential interference contrast (DIC) procedure, and fluorescence microscopy are necessary for paint examinations (see **Figure 3(A)–3(E)**).

Paint fragments are first placed on a microscope slide without previous preparation to determine whether there are significant imperfections (e.g., striations marks, 'fish eye' depressions, etc.) and other textures on the surfaces: DIC procedure is the most suitable. The determination of particle form, size class distribution, and the number of effect pigments (metal and pearl luster pigments) can also be accomplished quickly with plane-polarized light in incident light/bright field without prior sample preparation. All analyses for the determination of paint structures and for the most complete description of the morphological and physicochemical features of the colored pigments and extenders used in paints (color, size, crystal habit, crystallinity, and refractive index) must be carried out at the highest magnification possible (250–1000 $\times$ ) in transmitted light with bright field, dark field, and double polarization techniques on 3  $\mu\text{m}$  microtome sections preparations (embedded parallel or perpendicular to the surface). Fluorescence microscopy at an excitation wavelength of 365 nm must also be used for the characterization of pigments, additives, and film-forming components.

## Microspectrophotometry

The spectrophotometric technique is very useful for comparative purposes for a color classification system in a database and for the identification of individual organic or pearlescent (or interference) pigments or additives such as ultraviolet (UV)-absorbers; batch-to-batch discrimination is also possible for analyzing the basecoats. MSP permits to measure absorption spectra in a given spectral range comprising typically 250–780 nm and to calculate

the tristimulus CIE color coordinates. For paint analysis two main techniques are used: visible reflectance for the objective color measurement and UV–visible transmittance particularly for the discrimination of samples of similar shades of color and for the detection of additives.

Color is one of the most important as well as the most eye-catching characteristics of microscopically opaque objects such as paint. Color measurements are carried out using a single-beam technique (previously measuring a standard on a perfectly white reflective object as a pure barium sulfate tablet or a polished opal glass) with microspectrophotometers fitted with image-side monochromators with geometry 45°/0° (like J&M Tidas diode array spectrometers and the Zeiss MPM 800-system).

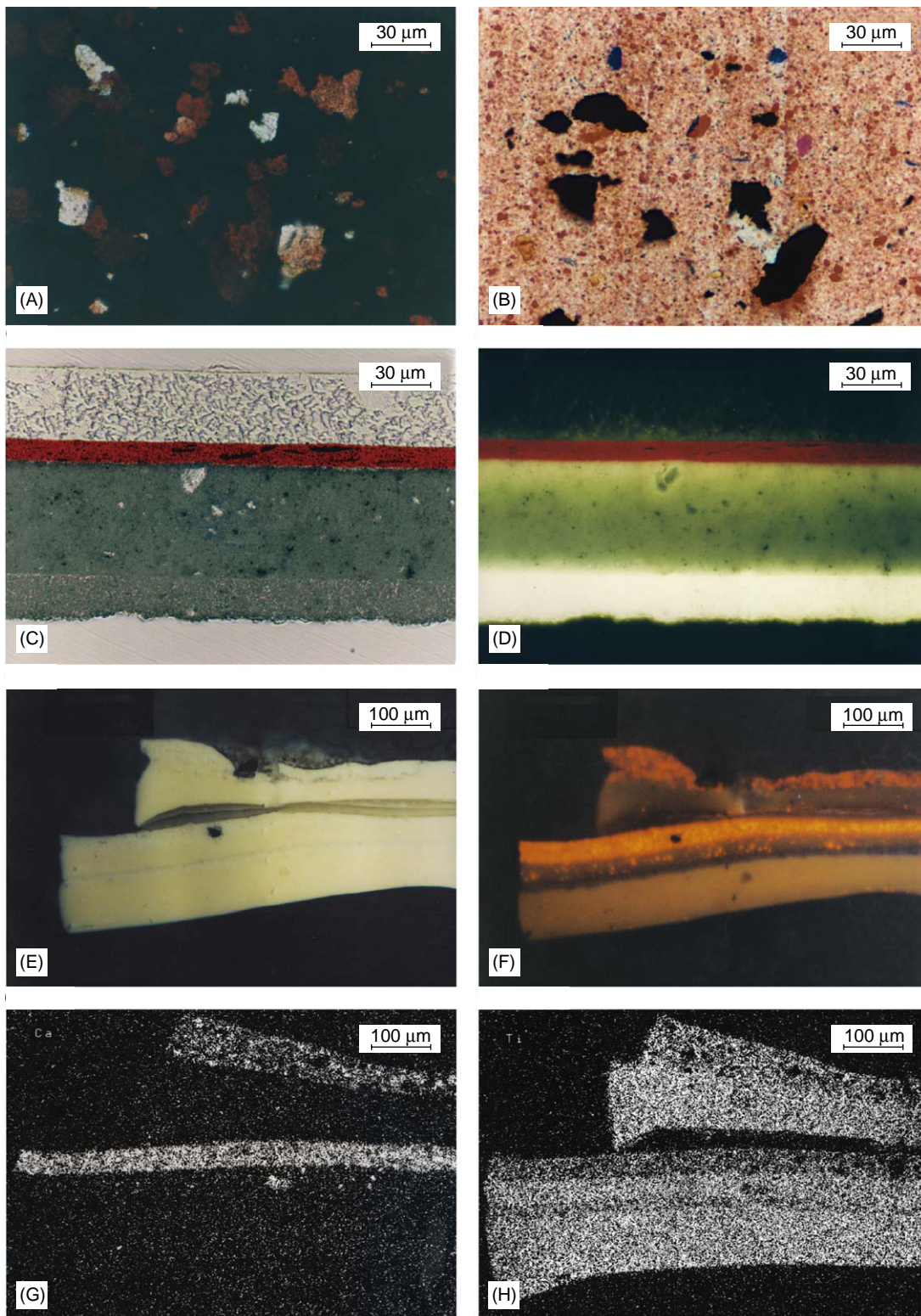
Visible reflectance spectra can be obtained from upper surfaces or from individual layers in a cross-section. Areas ranging from 6  $\times$  150 to 20  $\times$  80  $\mu\text{m}^2$ , respectively are measured. For microscopic color measurements only the measuring parameters of 45°/0° are realizable.

Since many paint layers already exhibit inhomogeneities in such dimensions for the calculation of color data (tristimulus values and chromaticity coordinates), it is necessary to use mean reflectance values derived from three to five individual measurements of different sample areas. For individual measurements, the sample surfaces, especially in the case of paints containing plate-like pigments (light interference and/or color variable pigments), should have varying orientations to the adjustable measuring aperture.

UV–visible spectra are accessible by means of analysis of thin sections of paint with microspectrophotometers equipped with quartz optics; quantifiable results are obtained from 3- $\mu\text{m}$ -thin sections fixed on quartz specimen slides. The reproducibility of absorbance spectra is very high for solid as well as for metallic paints even when a small measuring aperture (10  $\times$  30  $\mu\text{m}$ ) is used. Batch-to-batch discrimination is therefore possible. Individual organic pigments are also identifiable using UV–visible spectra when, for example, they are present in the form of aggregates and reagglomerates in the base coats of two-coat metallic paints.

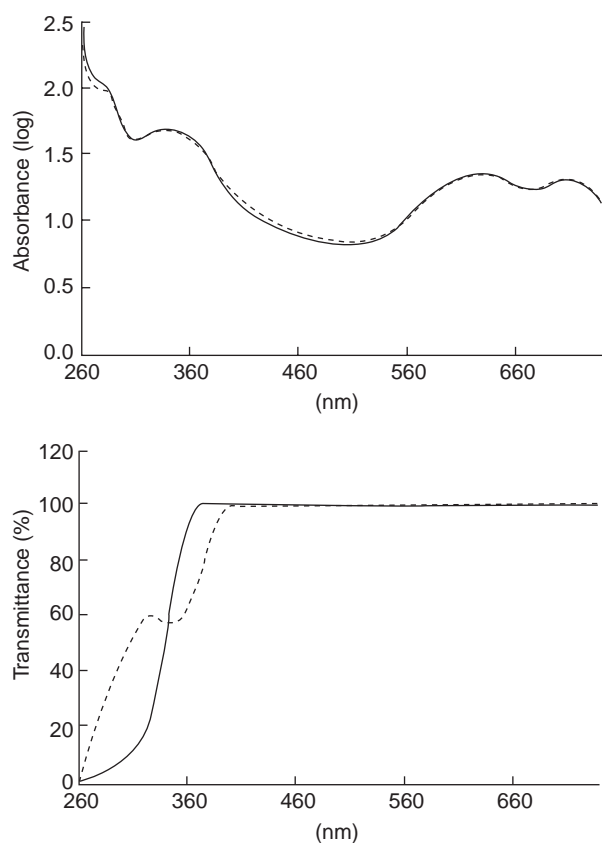
UV-absorbers (e.g., hydroxyphenylbenzotriazole) can also be identified in 20- $\mu\text{m}$ -thin sections of the clear coats of two-coat metallic paints (**Figure 4**).

Discrimination between different vehicles which were originally painted in an identical manner is based on the varying decrease of the UV-absorber concentration with time, depending on the substance class of light stabilizer (e.g., oxalanilides, benzophenones) and the type of environmental conditions.



**Figure 3** A multilayered paint sample taken from a vehicle with two-coat metallic top coat: burgundy red (BMW 199). (A) Surface as viewed under polarized/reflected light (500 $\times$ ; scale bar: 30  $\mu$ m). (B) Section (3  $\mu$ m) sliced parallel to the surface of the metallic base coat; transmitted light (500 $\times$ ; scale bar: 30  $\mu$ m). (C) Cross-section (3  $\mu$ m) with the layer sequence (from the bottom): primer, surfacer, base coat, clear coat; transmitted light (500 $\times$ ; scale bar: 30  $\mu$ m). (D) Sample as (C): fluorescence with 365 nm excitation wavelength (500 $\times$ ; scale bar: 30  $\mu$ m). Cross-section of a fragment of a seven-layered household paint containing different amounts of  $\text{CaCO}_3$  phases (scale bar: 100  $\mu$ m). (E) Photomicrograph (darkfield illumination) of the polished cross-section. (F) Cathodoluminescence-photomicrograph of the cross-section. (G) X-ray map (calcium; scanning electron microscopy-wavelength-dispersive X-ray fluorescence spectrometry (SEM-WDX) of the same area). (H) X-ray map (titanium; SEM-WDX) of the same area. (Reproduced with permission from Stöcklein W, Bundeskriminalamt, Wiesbaden, Germany.)





**Figure 4** UV-visible spectra of the base coats and clear coats from a suspect and a control sample. (Upper) Absorbance spectra of the base coats; transmission measurements made on a 3- $\mu\text{m}$  thin section. (Lower) Transmittance spectra of the clear coats; transmission measurements made on a 20- $\mu\text{m}$ -thick section.

## Vibrational Spectroscopy

### Infrared Spectroscopy

IR spectroscopy is particularly used for the identification of the main components like binder types (see Table 1), pigments and extenders (see Table 2), as well as for their classification.

Discriminations within a binder type are possible by means of identification of modifiers, for example. Further discriminations are based on the comparison of the relative intensity of absorption bands. The capacity for monitoring the chemical changes which occur from UV degradation (weathering) can also be used for the differentiation of paints.

Thanks to its time and signal-to-noise advantages as well as greater wavelength accuracy, Fourier transform IR (FTIR) spectroscopy has become a technique of choice for the analysis of transmissive or reflective forensic paint materials, especially if coupled to a microscope. By using FTIR microscopy good spectra can be obtained from paint layers as

**Table 1** IR characteristic absorptions peaks or bands of some common binder types; characteristic bands in bold (consider an incertitude of  $\pm 10 \text{ cm}^{-1}$ )

Binder	IR absorption peaks or bands ( $\text{cm}^{-1}$ )
Acrylic	1730, 1480, 1370, <b>1250–1050</b> (1180–1160, 1090–1080)
Alkyd	
Orthophthalic	1730, 1580, 1470, <b>1270, 1120, 1080</b> , 1040, 740, 700
Isophthalic	1730, 1610, 1475, <b>1305, 1240</b> , 1095, 1075, 730
Terephthalic	<b>1270</b> , 1250, <b>1120, 1105</b> , 1020, 730
Epoxy	<b>1510</b> , 1240, 1180, <b>830</b>
Melamine	<b>1550</b> , 815
Nitrocellulose	1730, <b>1660, 1280</b> , 1060, <b>840</b> , 750
Styrene	<b>1490</b> , 1450, 760, 700
Urea	<b>1650</b> , 1540, 1250–1270
Urethane	1730, <b>1690</b> , 1540, 1460, 1380, 1250, 1150, 770

**Table 2** IR characteristic absorptions peaks or bands of some common pigments and extenders; characteristic bands in bold (consider an incertitude of  $\pm 10 \text{ cm}^{-1}$ )

Pigment/extender	IR absorption peaks or bands ( $\text{cm}^{-1}$ )
Calcium carbonate (calcite)	<b>1440</b> , 880, 720, 315
Titanium dioxide (rutile)	<b>600</b> , 410, 340
Magnesium silicate (talc)	<b>3680, 1020</b> , 670, 465, 450, 420, 390, 345
Barium chromate	935, 895, <b>860</b>
Aluminium silicate (kaolinite)	<b>3700, 3620</b> , 1100, <b>1030, 1010</b> , 920, 700, 540, 470, 430, 350, 280
Iron oxide	<b>560</b>
Barium sulfate	<b>1180, 1120, 1080, 980, 630</b> , 610

small as  $10 \times 20 \mu\text{m}^2$ ; for this purpose microtome sections of 3–5  $\mu\text{m}$  are particularly suitable.

Diamond anvil cells (low-pressure or high-pressure) have long been used in the successful measurement of transmission spectra of small paint samples: these latter are squeezed with light pressure between two type II diamonds into a transparent film to obtain an appropriate thickness. For transmittance analysis, potassium bromide, KBr (or cesium iodide, CsI) micropellet constitutes a simple solution.

Different sampling techniques are available for the handling of micropaint samples. Among these techniques are diffuse reflectance IR Fourier transform spectroscopy (DRIFTS) and attenuated total reflection (ATR), using accessories like 'SplitPea' or 'Golden Gate' or a microATR objective for IR

microscopes. Using DRIFTS, single-layer metallic paint fragments can be analyzed directly without sample preparation; the use of abrasive disks (3 mm diameter) coated with silicon carbide (SiC) is an alternative sampling technique for DRIFTS. With this technique the sample is simply rubbed on the SiC paper and then measured *in situ* on the surface of the paper.

ATR accessory can also be employed for reflectance analysis, particularly for the surfaces of smears sticking on a given support: with, e.g., 'SplitPea' accessory (Harrick Scientific Corporation) very small paint traces can be measured ( $<0.3$  mm) without sample preparation, on the basis of internal reflection, external reflection, or in-line diffuse reflection spectroscopy.

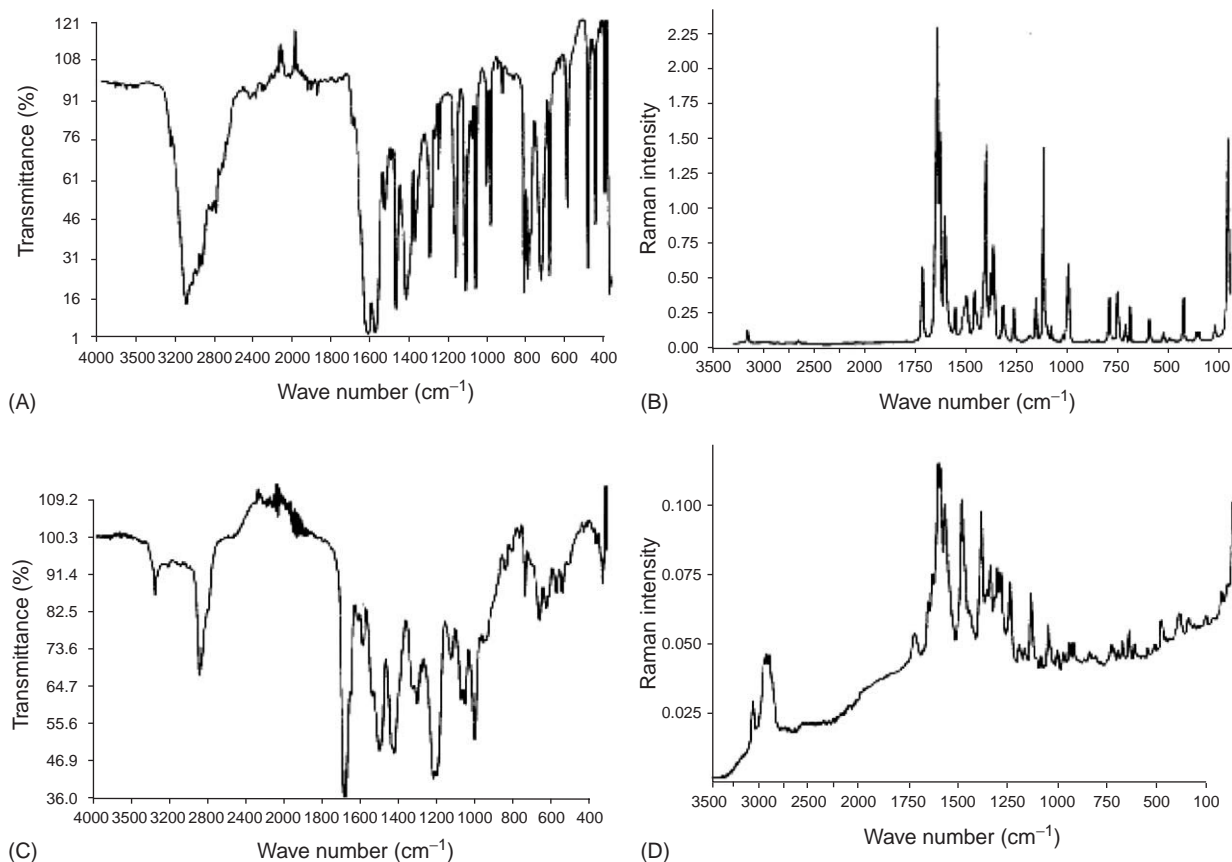
For paint analysis, a spectral range comprising between 4000 and 600  $\text{cm}^{-1}$  is usually used, but for the identification of some inorganic pigments and extenders (e.g., titanium dioxide), a lower frequency

domain is necessary: using deuterated triglycine sulfate (DTGS) as detector and CsI for the interferometer system instead of mercury cadmium telluride (MCT) and potassium bromide (KBr), spectral data until 200  $\text{cm}^{-1}$  can be obtained.

IR spectra also constitute the starting point for the research of vehicles involved in hit-and-run accidents: databases are built containing IR information of every layer characterizing manufactured vehicles. Some important examples are the European Collection of Automotive Paints (EUCAP) Database and the FBI/RCMP Paint Data Query (PDQ) Automotive Paint Database.

### Raman Spectroscopy

Raman spectroscopy is complementary to IR spectroscopy (Figure 5): for paint, while IR spectroscopy is used to mainly characterize binder type and some inorganic pigments, Raman techniques are more



**Figure 5** FTIR spectra and Raman spectra obtained from the pure C.I. Pigment Red 254 and from microtraces (3  $\mu\text{g}$ ) of an alkyd-melamine paint containing 4% of the C.I. Pigment Red 254. This example clearly reveals the complementary information obtainable from using both IR and Raman spectroscopy for paint analysis. In the IR spectrum of the paint sample nearly all the absorption bands of the pigment are masked by those of the resin; however, the fingerprint of the pigment can be seen in the Raman spectrum. (A) Diamond anvil cell spectrum of C.I. Pigment Red 254 (1,4-diketo-3,6-diaryl-pyrrolo(3,4-c)pyrrole); (B) NIR-FT Raman spectrum of C.I. Pigment Red 254; (C) Diamond anvil cell spectrum of the vehicle paint Tornado red (Volkswagen) containing 4% C.I. Pigment Red 254; and (D) NIR-FT Raman spectrum of the vehicle paint Tornado red (Volkswagen).

suitable for the identification of the main organic or inorganic pigments present. Extenders can be detected with both techniques.

Depending on instrumentation, different laser wavelengths are available: near IR lasers (e.g., Nd-YAG, diode array) are useful principally for red, orange, yellow, brown, and violet pigments. Visible lasers (e.g., argon, helium, and neon) are useful for the detection of blue and green pigments (e.g., phthalocyanines).

Fluorescence constitutes the greatest disadvantage of the Raman technique hiding spectral information. In the 1990s, Fourier transform Raman spectroscopy was developed on the basis of a near IR laser source being less energetic to produce fluorescence. However, with the introduction of the Peltier-cooled charge couple device (CCD) detector being more sensible and polyvalent for several laser sources, dispersive Raman spectrometers were reevaluated. The possibility to benefit from several laser sources going from the UV, visible, and to the near IR mounted on the same spectrometer constitutes an interesting tool for paint analysis. Moreover, research microscopes coupled to spectrometers permit to focus a microdetail of the sample (spatial resolution until  $\sim 1\ \mu\text{m}$ ).

Raman spectroscopy is above all applied in the field of artistic paints detecting inorganic pigments (e.g., vermilion, ochres) but some applications were also carried out for automotive paints allowing principally the detection of organic (e.g., diketo pyrrolo pyrrol, quinacridones) and inorganic (e.g., molybdate orange, chrome yellow) industrial pigments and some extenders (e.g., anatase) (see **Table 3**).

## Pyrolysis Gas Chromatography and Mass Spectrometry

Pyrolysis gas chromatography (PyGC), particularly when combined with MS, PyGC and mass spectrometry (PyGC-MS), is a powerful, fast, and elegant analytical method for the characterization of polymers. In addition to qualitative and quantitative detection of monomer units and the identification of comonomers, it also provides additional information about additives and impurities. In addition, the chromatogram, especially the reconstructed total ion chromatogram of the pyrolysis products, the pyrogram, can serve as a 'fingerprint' of the paint samples. The most frequently used pyrolyzers are the Curie-point type, the furnace type, and the pulsed filament type. In terms of reproducibility, an indirect heating system using a platinum coil (CDS system) has proved very successful. Among the columns, fused capillarity columns with nonpolar stationary phases or with intermediate polarity are generally accepted. As a rule, flame ionization detectors (FID) or mass spectrometers (MSs) (quadrupoles and ion traps) with dedicated data systems are used as detectors or identification systems.

Information about the composition of the sample material can be significantly increased if the paints have been derivatized during pyrolysis. Tetramethylammonium hydroxide (TMAH) is used as a derivatizing reagent for structure elucidation of alkyd, unsaturated polyester, epoxy, and phenol-formaldehyde resins. The derivatization of paints, whose pyrolysis products elute at very low retention times (such as formulations based on polyvinylacetate

**Table 3** Raman scattering bands of some common pigments and extenders

<i>Pigments or extenders</i>	<i>Laser source (nm)</i>	<i>Raman scattering bands (<math>\text{cm}^{-1}</math>)</i>
Diketo Pyrrolo Pyrrol (C.I. PR 254)	1064	1660, <b>1590, 1575</b> , 1550, 1495, 1440, 1400, <b>1345</b> , 1315, 1305, 1255, 1200, 1090, <b>1050</b> , 925, 725, 685, 620, 350, <b>125</b>
Titanium dioxide		
Rutile	785	1450, 1000, 980, <b>615, 450</b>
Anatase	785	<b>640, 520, 400</b>
Chrome Yellow	785	<b>845</b> , 405, <b>365</b>
Lead Molybdate Orange	785	880, <b>825, 360, 345</b> , 145
Iron Ferrocyanide (Prussian Blue)	785	<b>2150</b> , 2090, 275
C.I. Pigment Green 36 (Phthalocyanine)	632.8	<b>1530, 1430</b> , 1375, <b>1320, 1265, 1195, 960, 765, 745</b> , 710, <b>660</b> , 535, 330, 320, 275, 260, 220, 190
C.I. Pigment Green 7 (Phthalocyanine)	632.8	<b>1530, 1435</b> , 1385, <b>1330, 1275, 1205, 1075, 975</b> , <b>815, 770, 735, 700, 680</b> , 640, 555, 505, 290, 195
C.I. Pigment Green 7 (Phthalocyanine)	514.5	<b>1545, 1495</b> , 1380, <b>1270, 1200, 1075</b> , 970, 950, <b>815</b> , 770, 740, <b>680</b> , 640, 525, 505, 365, 345, 235, 160
C.I. Pigment Violet 23	514.5	1640, 1590, <b>1430, 1390, 1345, 1255, 1205, 1165</b> , 1130, 1105, 985, 930, 920, 720, 670, 615, 590, 525, 485, 415, 315



copolymer or on cellulose acetate butyrate), is accomplished with tetrabutylammonium hydroxide (TBAH).

The detection and identification of polar components is also possible using simultaneous hydrolysis and alkylation. These are normally difficult or impossible to determine because they tend to remain on the column, cause peak tailing, show poor reproducibility, or lose their identity due to the formation of low molecular weight fragments. Thus, for example, methyl esters of polybasic acids and long-chain fatty acids and methyl ethers of polyhydric alcohols are formed from alkyd resins with TMAH, whereas in conventional PyGC, aldehydes are formed from polyols, and alkanes and alkenes from fatty acids (Figure 6).

However, some problems have been encountered using TMAH on-line methylation in a Curie-point pyrolysis unit: the reaction between TMAH and fatty (di)acids (azelaic, palmitic, and stearic) showed that different quantities of subproducts are formed varying the quantity of TMAH reagent: deepening of such method is suitable to understand the formation of such subproducts and the way to avoid it.

In addition to the identification of the main components, the presence of modifiers such as wood resin and epoxy resins in alkyds can also be proved. It is also possible to draw conclusions about the oil length and the degree of cure or the age of the alkyd resins.

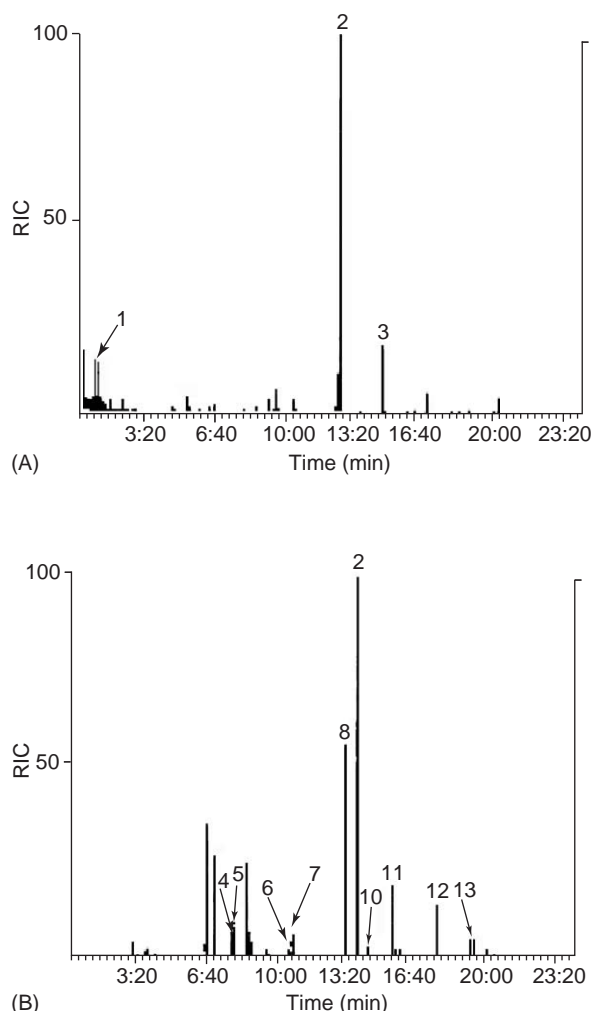
By means of high temperature reaction with TBAH the identity of comonomers in polyacrylates can also be determined. Among the identifiable components are in particular hydroxy-functional and carboxy-functional monomers (e.g., hydroxyethylacrylate, acrylic acid, and methacrylic acid) as well as additives (e.g., cellulose acetate butyrate). These partially strongly polar compounds are either not identifiable at all with conventional PyGC or their presence leads to strong tailing. Since butyl esters are formed from the acids in the reaction with TBAH, which may already be present in the product as comonomers, the analysis results are not always unambiguous. To determine the compositions of acrylates, it is therefore advisable to compare pyrograms which have been obtained with and without the use of a derivatizing reagent.

With the introduction of the simultaneous pyrolysis and alkylation techniques, the capacities for identification, classification, and differentiation have improved significantly.

## Elemental Analysis Techniques

### Scanning Electron Microscopy and Energy Dispersive X-Ray Spectrometry

Scanning electron microscopy and energy dispersive X-ray spectrometry (SEM/EDS or SEM/EDX) is the



**Figure 6** Conventional pyrolysis (A) and simultaneous pyrolysis-methylation (B) of an orthophthalic alkyl melamine resin; capillary GC-MS-pyrograms (RIC). 1 = methacrolein; 2 = phthalic anhydride; 3 = phthalimide; 4 = pentaerythritol tetramethyl ether; 5 = octanoic acid methyl ester; 6 = hexanedioic acid dimethyl ester; 7 = decanoic acid methyl ester; 8 = dodecanoic acid methyl ester; 9 = orthophthalic acid dimethyl ester; 10 = azelaic acid dimethyl ester; 11 = tetradecanoic acid methylester; 12 = hexadecanoic acid methyl ester; 13 = octadecanoic acid methyl ester. (Instrument conditions: Pyrolyzer: Pyrola (Pyrolab, Lund/Sweden); pyrolysis temperature: 750°C; GC: Finnigan MAT 9610 fitted with a fused-silica column (DB 5; 0.25  $\mu$ m phase thickness; 0.25 mm internal diameter; 30 m in length); injection mode: splitless (20 s)/split 1:20; temperature program settings: 40°C, hold 2 min; increase at 10°C min<sup>-1</sup> to 280°C.) Mass spectrometer: Finnigan MAT 4500.

method of choice for the characterization of the morphology and the elemental composition of paint samples. A wide range of elements present on the periodic chart (in quantities  $\sim 0.1$  wt.%) can be detected for the identification of inorganic components as pigments and extenders (see Table 4).

**Table 4** Elements indicating the presence of some inorganic and organic pigments and extenders

<i>Pigments or extenders</i>	<i>Element(s)</i>
Yellow lead chromate	Pb, Cr, S
Chromium oxide	Cr
Iron ferrocyanide (Prussian Blue)	Fe
Ochres	Fe, Si
Aluminum silicate	
Kaolinite	Al, Si
Mica	K, Si, Al
Magnesium silicate (talc)	Mg, Si
Barium sulfate	Ba, S
Calcium carbonate	Ca
Monoarylide yellow (Azo)	Cl
Phthalocyanine Green	Cu, Cl, or Br

Organic pigments cannot be identified, but their presence is often indicated by the presence of some elements as chlorine and bromine or nickel, manganese, and copper (e.g., for phthalocyanines). In lake pigments this can also be seen in combination with barium or calcium and sulfur.

Scanning electron microscope (SEM) systems are ideal for a pictorial representation of morphological features of single layers of multilayered paint samples (particularly those not well contrasted by optical microscopy as the white ones) obtained by the contrast of secondary (SE) and backscattered electrons (BSE). An important advantage of SEM is the ability to analyze even the smallest smears: for this purpose the smears are lifted from the underlying material using double-sided adhesive tape transferred to the sample holder (e.g., slide). Twenty micrometers microtome sections of paint cross-sections are recommended. In order to avoid electrical charging the samples have to be carbon coated or transferred into the chamber of a low vacuum system.

The main emphasis of paint analysis with SEM concerns EDS or EDX. To compensate for inhomogeneities, the scanned sample areas should be as large as possible. At least two to three areas must be analyzed under standardized measurement conditions. Mapping procedures are recommended for the analysis of cross-sections. In this manner the thinnest layers on the upper and lower surfaces (e.g., zinc phosphate layers on the lower surface of vehicle paint fragments) can be described and analyzed. Metallic paints should be examined at higher magnification (10 000  $\times$ ) in aluminum-free areas.

For quantitative X-ray analysis, electron probe microanalyzers (EPMA, EPA, or EMMA) are able to determine the elemental concentration by X-ray emission from the microvolume of paint samples where a static electron beam interacts. However, due to inhomogeneity of paint samples nonquantitative

analysis based on the comparison of spectra or peak rationing is preferred.

### Wavelength Dispersive X-Ray Fluorescence

Wavelength dispersive X-ray fluorescence (WDX) detects X-rays according to their wavelength rather than their energy (see **Figure 3(G)** and **3(H)**). Because of the limited resolution capacity of lithium-drifted silicon detectors (Si(Li)-diodes), the peak areas from a series of elements as PbM-SK-MoL, TiK-BaL, or CaK-SbL lines, being important for pigment analysis, cannot be separated using EDS. In these cases correct analyses can only be achieved with a wavelength dispersive spectrometer.

### X-Ray Fluorescence Spectrometry

For X-ray fluorescence spectrometry (XRF) technique an X-ray excitation source is used and, compared to SEM/EDS, it is more sensitive to higher atomic weight elements and less sensitive to the lower atomic weight elements commonly found in extenders.

Successful X-ray fluorescence spectrometric attachments are those which allow observation of the sample during analysis (microscope or image sensor) and which have an adjustable sample stage. Such devices are generally equipped with capillary optics, a low power X-ray tube (150–350 W) and a Si(Li)-detector.

### Laser-Excited Atomic Emission Spectroscopy

Atomic emission from the laser-excited plume is monitored by an emission spectrograph; with Q-switch lasers even individual layers in cross-sections of paint fragments can be analyzed. Material consumption lies in the nanogram range. The sample is fixed on the edge of a glass microscope slide with the aid of double-sided adhesive tape. An advantage of this multielement system is the very high sensitivity; a disadvantage is the time-consuming evaluation of the results via a photographic plate or film.

### Inductively Coupled Plasma Techniques

Inductively coupled plasmas either combined with atomic emission spectrometers (ICP-AES) or mass spectrometers (ICP-MS) where samples are excited using a high-temperature gaseous plasma can be used for elemental analysis. Since the development of ICPs, most applications have required digestion of solid samples with heat and/or strong acids. The coupling of laser-ablation (LA) with ICP-MS has resulted in the development of extremely sensitive microprobes capable of determining most elements of the periodic table. Another advantage of the

technique is that LA-ICP-MS is point specific, that is, only the area ablated is subject to analysis. As a result, LA-ICP-MS has become one of the most exciting new fields of research in materials science in general.

## Solubility and Microchemical Tests

In some laboratories, comparative analyses begin with solubility tests. The difference between the suspect and the control samples can often be determined quickly through the reaction and the different solubility of the paint binders with a variety of chemical reagents (e.g., chloroform, acetone, concentrated sulfuric acid, etc.): only when differences cannot be detected with these tests is it necessary to employ instrumental methods.

Microchemical testing is not only based on dissolution of binders but also on pigment and binder color reactions with oxidizing, dehydrating, or reducing agents. These tests are destructive: diphenylamine (in concentrated sulfuric acid and glacial acetic acid) and LeRosen (formaldehyde in concentrated sulfuric acid) tests are typically used.

For both solubility and microchemical tests, reactions and color changes must be observed under a stereo microscope concurrently for unknown and control samples at different time intervals, in order to detect differences.

## Other Techniques

### Cathodoluminescence

Cathodoluminescence (CL) permits to give prominence to the layer sequence particularly white and off-white (e.g., pale yellow, cream colors) multilayered paint samples (see **Figure 3(F)**). It is the excitation of visible light or of radiation of adjacent wavelengths in semiconductors and insulators (e.g., inorganic pigments and extenders) when these are bombarded by electrons emitted from a cathode and accelerated in an electric field.

CL is well observable using optical microscopy as well as SEM. Worthwhile results can be obtained when a CL detector (photomultiplier) is coupled with a monochromator. The recording of CL-spectra allows differentiation of inorganic pigments and extenders which otherwise vary only in the type or number of lattice defects (e.g., vacancies, trace activators on interstitial lattice sites, etc.).

### Thin-Layer Chromatography

If solubility tests reveal that microtraces of paint contain organic pigments, these latter can be

identified by thin-layer chromatography (TLC) or by high-performance thin-layer chromatography (HPTLC). Silica gel is used as the solid phase and chloroform as the eluent in a mixture with aliphatic or aromatic hydrocarbons. These methods, however, can only be applied to a few classes of pigments being soluble. For light-fast polycyclic pigments (phthalocyanines, quinacridones, perylenes, perinones, dioxazines, indanthrones, isoindolinones, benzimidazolones) trifluoroacetic acid/water is used as extractor and eluent. Pigments are converted into mobilizable onium compounds and are separated onto C<sub>18</sub>-reversed phase plates. The identification is achieved by means of the  $R_F$  values but microspectrophotometric, IR, and Raman spectroscopic examinations can be added for this purpose.

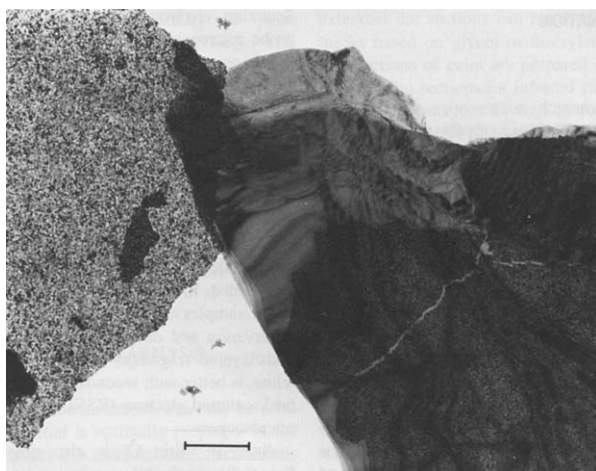
### X-Ray Diffraction Spectroscopy

X-ray diffraction spectroscopy (XRD) techniques have been used for the analysis of inorganic pigments and extenders by analyzing the crystalline structure of the material rather than its elemental content: e.g., it is typically able to differentiate the crystalline forms of titanium dioxide, rutile, and anatase. The method is less suitable for the identification of organic pigments as they do not generally scatter X-rays well and they are often present in paints only in comparatively small concentrations. Additional difficulties occur in the interpretation of the results because the diffraction pattern sometimes exhibits only the most subtle differences within a class of organic pigments.

### Transmission Electron Microscopy

Methods of transmission electron microscopy (TEM) can also be applied to the pigments and extenders analysis. Morphological studies of shadowgraphs have proven reliable. During preparation, the paint microtraces (>100 µg) are dispersed in an ultrasonic bath after ashing (450–550°C) and sedimented onto appropriate sample holders (copper grid with pioloformfilm). The morphological structures of such powder preparations observed at 12 000–65 000 × magnification display a broad range of variation. They can be used together with EDS and electron diffraction for the identification of the material. Especially interesting results are obtained from the examination of paints with effect pigments (e.g., pearl luster pigments) (**Figure 7**).

In one analytical step it is possible to classify and identify the various pigments and extenders used and to describe the morphological features and anomalies which are phase-typical or which are the result of manufacturing. Proof that paints contain



**Figure 7** Transmission electron microscopy image of a pearl luster pigment; during preparation the microcrystalline nanometer thick layers of rutile have become separated from the undersurface and partially from the upper surface of the mica surface, so that the layered structure of the effect pigment is clearly visible (scale bar: 1  $\mu\text{m}$ ).

microtitanium dioxide (grain size 0.01–0.05  $\mu\text{m}$ ) or amorphous silicon dioxide, can be obtained by using TEM-examination.

### Thermal Analysis

Of the many methods of thermal analysis (TA), only a few have been used in forensic paint analysis: among them are differential thermal analysis (DTA), differential scanning calorimetry (DSC), and thermomechanical analysis (TMA). For example, DSC can provide information of the rate of ageing of oil-based paint media. The determination of the age of oxidative drying paint films is also possible with TMA. In this procedure the glass transition temperature ( $T_g$ ) is determined. With  $T_g$  determination it is also possible to describe the conditions under which the paint films were hardened (e.g., the stoving temperatures of vehicle paints). However, samples in sufficient amounts must be available in order to obtain reproducible results (sample size: disks with radius >2.5 mm and thickness of the order of 0.05 mm). Such large samples are seldom available. Therefore, for the purposes of forensic science TMA is not a suitable technique for the routine examinations of small paint fragments.

An interesting procedure in forensic paint analysis, however, is thermogravimetry–mass spectrometry (TG–MS), determining simultaneously TG curve and its first derivative, total ion current versus the TG temperature, single ion currents for selected molecular fragments versus TG temperature, and determination of the pigment/volume concentration of

inorganic components residue analysis for trace elements.

**See also:** **Atomic Emission Spectrometry:** Principles and Instrumentation. **Color Measurement.** **Forensic Sciences:** Thin-Layer Chromatography. **Gas Chromatography:** Pyrolysis; Fourier Transform Infrared Spectroscopy. **Infrared Spectroscopy:** Sample Presentation. **Laser-Based Techniques.** **Liquid Chromatography:** Fourier-Transform Infrared Spectroscopy. **Microscopy Applications:** Forensic. **Microscopy Techniques:** Electron Microscopy. **Paints:** Water-Based; Organic Solvent-Based. **Raman Spectroscopy:** Near-Infrared. **Spectrophotometry:** Organic Compounds. **Thermal Analysis:** Overview; Temperature-Modulated Techniques; Coupled Techniques; Sample-Controlled Techniques. **X-Ray Absorption and Diffraction:** Overview; X-Ray Absorption; X-Ray Diffraction – Powder. **X-Ray Fluorescence and Emission:** X-Ray Diffraction – Single Crystal; X-Ray Fluorescence Theory; Wavelength Dispersive X-Ray Fluorescence; Energy Dispersive X-Ray Fluorescence; Total Reflection X-Ray Fluorescence; Particle-Induced X-Ray Emission.

### Further Reading

- ASTM (1996) *Terminology Relating to Paint, Varnish, Lacquer, and Related Products*. Standard D16.
- Caddy B (ed.) (2001) *Forensic Examination of Glass and Paint – Analysis and Interpretation*. New York: Taylor and Francis.
- Scientific Working Group on Materials Analysis (SWG-MAT) (1999) Forensic paint analysis and comparison guidelines. *Forensic Science Communications*, U.S. Department of Justice, FBI, July 1999, vol. 1(2). Online: [www.fbi.gov/programs/lab/fsc/backissu/july1999/painta.htm](http://www.fbi.gov/programs/lab/fsc/backissu/july1999/painta.htm).
- Scientific Working Group on Materials Analysis (SWG-MAT) (2002) Standard guide for using scanning electron microscopy/X-ray spectrometry in forensic paint examinations. *Forensic Science Communications*, U.S. Department of Justice, FBI, October 2002 vol. 4(4). Online: <http://www.fbi.gov/hq/lab/fsc/backissu/oct2002/bottr-ell.htm>.
- Stöecklein W, Franke M, and Goebel R (2001) Cathodoluminescence in forensic science. *Problems of Forensic Sciences XLVII*: 122–136.
- Stöecklein W (2001) The analysis of new plate-like pigments. *Paint & Coating Industry XVII*(9): 48–65. Online: <http://www.pcimag.com>.
- Thornton JI (2002) Forensic paint examination. In: Saferstein R (ed.) *Forensic Science Handbook*, 2nd edn., vol. I, ch. 8, pp. 429–478. Upper Saddle River, NJ: Prentice Hall, Pearson Education, Inc.
- Van den Berg JDJ and Boon JJ (2001) Unwanted alkylation during direct methylation of fatty (di)acids using tetramethyl ammonium hydroxide reagent in a Curie-point pyrolysis unit. *Journal of Analytical and Applied Pyrolysis* 61: 45–63.



## Questioned Documents

**C Neumann and W D Mazzella**, University of Lausanne, Lausanne, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

A document – from the Latin *Documentum*, ‘which serves to educate’ – is conveniently defined as something that holds information. Although there are multiple ways to share information in our modern society, such as magnetic tapes, electric appliances (telephone), and electronic devices (computers), questioned document experts have traditionally focused on the examination of documents normally made of paper, which bear written or printed messages.

### Documents as a Support for Traces

In the forensic science field, the examination of questioned documents is one of the most commonly carried out tasks. Its primary aims are to provide information about the material (to describe the different materials used and identify their origin) and the history of a document (the presence of traces of erased or modified entries in order to arrive at the truth regarding their integrity).

Therefore, the techniques described in this article can be used to compare trace materials (papers, inks, or toners of the questioned document) with reference materials from known sources. This process is typically carried out during checking of forgeries of passports, ID cards, and bank notes. The authenticity of the paper and ink materials and the identification of the various printing technologies generally indicate if the questioned document is a counterfeit or genuine. However, these techniques allow for studying the homogeneity of the material used across the whole questioned documents and thus to detect potential falsifications.

Beside these aspects, several other types of trace can also be found on documents. Among them, finger marks, if present, can be easily revealed on almost all types of paper; DNA can often be extracted from saliva that could be present on the back of stamps; and the use of electrostatic detection devices allows examiners to reveal latent marks on documents.

## Paper

### Introduction

The forensic examination of paper is aimed primarily to determine its origin, to detect any alteration of the

content of the documents, and, more generally, to look for genuineness of these documents.

### History and Composition

The need for a writing medium has constantly increased alongside the development of writing instruments. Early materials such as stone or animal skins were replaced in Ancient Egypt by a media made by crisscrossing and pressing together sliced strips of stem of Papyrus. In AD 105, paper was invented by Ts'ai Lun in China and completely substituted Papyrus around the tenth century AD.

Paper is made from the pulp of fibers from several origins, such as wood, cotton, linen, or rag. The raw material is first chemically or mechanically treated to break it down to pulp. Then, large quantities of water and other materials are added to the fiber pulp, such as resins, minerals, and dyes or whiteners. This mixture is then spread evenly over a surface and dried. At that stage, the frame of the surface and potential watermarks create identifiable patterns on the paper. Finally, the paper is cut to the required dimension.

### Analytical Methods for Paper Analysis

**Nondestructive tests** The operations conducted during the manufacturing process of paper generate features that can help easily distinguish one paper from another, either by direct observation or with the help of basic measuring tools or optical techniques.

The shape, size, thickness, and the weight of the sheets can be measured. Watermarks, frame patterns, and acquired defects, produced by the manufacturing process, can be observed in transmission and reflected light with or without magnification. These patterns can be mathematically compared by Fourier transform algorithm to discriminate papers.

The optical behavior of papers under filtered light examination and ultraviolet (UV)–visible (or visible–infrared (IR)) luminescence is also an important characteristic that helps to differentiate them. Absorption spectroscopy can be used to measure and compare the color of papers.

**Destructive tests** When possible, other tests can take advantage of removing a small amount of the paper to determine the fiber type, the method of pulping, the dyes, and the inorganic elements in the paper.

In the determination of the fiber type, a portion of the paper is broken using water or dilute acid. This

enables the fibers to be examined individually under a microscope. Thus, different varieties of fibers can be identified.

The elemental composition of the paper can be measured by X-ray fluorescence and by inductively coupled plasma-mass spectrometry (ICP/MS).

Finally, dyes and optical brighteners added in the manufacturing process can be analyzed using thin-layer chromatography, e.g., inks.

## Inks

### Introduction

The examination of ink material may provide evidence of considerable value. The detection of an erased ink may indicate a modification of the original message, and drawing comparisons between different inks may indicate a common origin; and with the help of databases, the brand of the writing instrument of ink entries can be identified. Furthermore, under certain circumstances, the date when inks were placed on documents can be determined.

### History and Composition

Inks were first used thousands of years ago in China as well as in Egypt, Greece, and the Roman Empire. These first inks were based, for centuries, on carbon particles in aqueous solutions. During the Middle Ages, iron-tannin, silver, and gold inks were developed and used, up until the nineteenth century, when the addition of indigo dye gave a blue-black color to the writing lines. Since then synthetic dyes from several chemical families have replaced natural dyes and multiple varieties of inks, of all colors, are used in different kinds of writing instruments. **Table 1** lists

the main components of contemporary writing instruments.

### Analytical Methods for Ink Analysis

**Optical examinations** Nearly every examination of a questioned document begins with nondestructive optical methods. Such indicative examinations are performed under normal lighting conditions with magnification. They allow determining the type of ink that is present in the document – and thus the writing instrument that was used – and help to decide which techniques are best suited for the subsequent analyses. The examinations are usually conducted under filtered lights, in the UV, visible, and IR ranges, using reflection and luminescence techniques, in order to reveal differences between inks, expose chemical erasures, and decipher obliterations.

### Spectroscopic techniques

**UV-visible spectroscopy** UV-visible absorbance microspectrophotometry and microspectrofluorimetry are used as nondestructive methods for analyzing inks directly on paper. They measure, respectively, the absorption and emission spectrum of inks and allow the discrimination of similarly colored inks. Therefore, they are a convenient way to render the comparison of ink colors more objectively than the naked eye.

**Raman spectroscopy** Raman and resonance Raman spectroscopy have also been applied to the analysis of inks on paper. These techniques allow the chemical characterization of the ink but were reported to be lacking in sensitivity and to be highly influenced by fluorescence effects caused by papers. Surface enhanced resonance Raman scattering spectroscopy

**Table 1** Ingredients commonly used in modern writing instruments

Writing instrument	First appearance	Coloring material (%)	Solvent	Others
Fountain pen	Reed: ~2500 BC Goose: ~100 AD	Acid dyes	0.6 –	Dextrin Poly(ethylene) glycol
	Modern form: 1809		3	Phenol
Ballpoint pen	Invention: 1888	Basic dyes	Up to 1950: vegetal or mineral oil	Binders: natural resins, polymers
	Modern form: 1939 Commercial: ~1950	Solvent dyes Phthalocyanine	~50 1950 and later: glycol	Fatty acids
Fiber pen	1962	Acid dyes	2 –	Glycol
Rollerball pen	1968		10	Formamide
Gel pen	1984	Pigments	~8	Glycol



helps avoid these deficiencies and shows promising results for the discrimination of gel pens and black ballpoint pen inks containing crystalline dyes.

**Mass spectroscopy** Field desorption mass spectrometry and secondary ion mass spectrometry allow the analysis of the inks' nonvolatile compounds, such as heavy dyes, which cannot be separated by gas chromatography (GC).

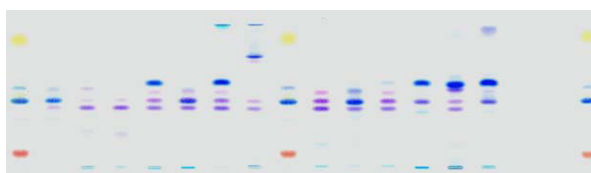
**Infrared spectroscopy** Fourier transform infrared (FTIR) and diffuse reflectance infrared Fourier transform spectroscopy have been reported for analyzing organic components of inks, such as resins.

**X-ray spectroscopy** Finally, X-ray fluorescence can be used to determine the elemental profile of inks and papers. Some techniques based on X-ray properties (particle-induced X-ray emission and inductively coupled plasma atomic emission spectroscopy) have been applied to writing inks. Nevertheless, it was found that there were not enough variation between the different particles, present in ink formula, to help distinguish from each other.

**Chromatographic techniques** Contrary to other areas of analytical chemistry, the determination of the absolute chemical composition of inks is not the overarching goal of forensic science. It is rather to determine if two samples are, or not, similar and thus if they are of a potential common origin. Therefore, chromatography is used to separate the components of ink mixtures in order to compare them one by one.

**Planar chromatography** The introduction of chromatographic methods for ink comparisons has been slow, due largely to the reluctance of the courts to accept 'destructive' techniques. Thanks to the improving sensitivity of planar chromatography techniques, starting with paper chromatography, followed by thin-layer chromatography (TLC) and, currently, with the use of high-performance thin-layer chromatography (HPTLC), the removal of small amounts of ink material has become accepted in practice.

HPTLC remains one of the most powerful techniques for the comparison and identification of inks in forensic science. Its advantages have been clearly established: it is cheap, user friendly, expedient, well reported, and tested. Overall, it is a screening technique that is only slightly dependent on the different chemical families of the several dyes included in the composition of each ink. Therefore, it remains the most widely used chromatographic technique for this purpose (see Figure 1).



**Figure 1** Separation of ink's colored material by high-performance thin-layer chromatography. Tracks 1, 9, and 18 are reference dyes' ladders; tracks 2–8 and 10–15 are ink samples; track 16 is a blank paper extraction, track 17 is an extractor blank.

A typical procedure for TLC/HPTLC analysis of inks is as follows. A sample of an ink line, of ~10 mm, is scraped from the document and dissolved in 10 µl of an appropriate solvent in a glass capillary (the extraction procedure can be accelerated by heating the capillary at 100°C for 15 min). The sample is spotted on a preconditioned or pre-washed silica plate and eluted with a solvent. Several mobile phases can be used, depending on the kind of writing instrument examined.

Control samples from alleged sources, or reference samples from databases, are usually run simultaneously with the questioned sample. The compositions of the samples are then compared visually, directly on the silica plate.

The position and color of the inks' various dyes on the silica plates can be measured with a densitometer, while their chemical composition can be determined using a Raman or FTIR spectrometer. The resulting data provide a more rigorous comparison of the inks and allow the storage of the ink profiles in databases.

Sometimes, differences between the obtained analytical results of two inks can occur, even if the two considered inks have a common origin. These differences can be explained under certain circumstances by the storage condition of the document, prior to the analysis. For example, the age of the samples has been shown to decrease the extraction efficiency and the exposure to sunlight has been shown to degrade certain dyes. The interpretation of such differences is often left to the judgment of an experienced forensic scientist, although some computer expert systems have actually been developed.

Finally, it should be noted that these properties, regarding extraction efficiency and dye degradation, are used in some cases for determining the date of the entry's creation.

**Liquid chromatography and capillary electrophoresis** Despite the advantages of HPTLC, many scientists have felt the need for an analytical technique that would offer more sensitivity, with a greater resolution, and a better reproducibility of the analyses.

Researches, done with modern liquid chromatography (LC) and capillary electrophoresis (CE), have demonstrated several advantages over planar chromatography techniques. First, there is the possibility of adding several types of detectors to characterize the components, separated by the analytical technique, such as diode array detector or mass spectrometry. Second, the use of these techniques allows overcoming one of the major disadvantages of planar chromatography, by improving the reproducibility.

However, the choice of the LC method that should be used for the analysis of a questioned dye has shown to be very dependent on the chemical families of dyes that are effectively present in a particular ink (for example, it appears that phthalocyanine dyes cannot be analyzed with the same solvent system and column as solvent dyes). LC analyses are also very time consuming and require a large quantity of solvent.

Despite the ecological advantages over LC (use of aqueous buffer instead of organic solvents), the efficiency of CE compared to LC is still a matter of discussion at this time.

*Gas chromatography* Heavy colored compounds of inks cannot be separated by GC. Therefore, this technique has mostly been used for the quantitative analysis of volatile solvents in inks. Such works are described in paper dealing with the determination of the age of ink entries, based on the relative speed of solvent evaporation.

## Toner Analysis

### Introduction

Due to the widespread availability of business machines over the past 30 years, the forensic document examiner is more than ever faced with the obligation of analyzing toner printed documents. The increasing popularity and quality of black and white and color copiers, laser printers, and fax (facsimile) machines that use toners, has led to a parallel expansion of the use of such systems in criminal activities. The purpose of this article is to point out how the analysis and differentiation of toner, employed in an office machine, can help establish the origin of a printed document or can be used to distinguish printed documents from different sources.

This article will cover only the analysis of the toner itself, in particular the analysis of the black toner. The examination of manufactured identifying characteristics, such as grip marks, roller marks, and acquired identifying characteristics, such as trash

marks or repetitive defects, will be not discussed here.

In this article the term document refers to a document produced by an electrophotographic process (for example, a photocopier, a laser printer, or a Fax machine employing a toner).

### Why do Toner Analysis?

1. To determine if two or more documents were printed with the same toner.
2. To determine if a certain photocopied document is a fraudulent alteration of an original document.
3. To determine the possible make and model of a photocopier that was used to produce a given photocopied document.

### General Theory

A toner can be defined as a material capable of accepting an electrical charge that is used in the general process called electrophotography to create an image (i.e., the printed document). The detailed explanation of the electrophotography process can be found in the specialized literature. It is not the purpose of this article to delve into this process.

### Types of Toners Encountered

Toner can be either dry or liquid. Dry toners are composed of microscopic dry particles ( $\sim 6\text{--}20\text{ }\mu\text{m}$ ) of polymer, additives, and carbon black (or other pigments or dyes), used to create the copy images. A dry toner has a raised, glossy appearance on paper and can be scraped off with a scalpel. Dry toners can either be of single component or a dual component. Single-component toners can be subdivided into magnetic and nonmagnetic toners. Magnetic toners are particularly common for single-color printing with dark colors. Nonmagnetic single-component toners are mostly used in systems functioning at a low operating speed. Magnetic single-component toners contain a magnetic carrier material (magnetite and ferrite are common) that can be detected on the final document. In dual-component toners the carrier (glass or plastic beads) is added separately and recovered afterward; therefore, it cannot be detected on the final copy.

Liquid toners are characterized by carbon black particles (size  $\sim <2\text{ }\mu\text{m}$ ) or dyes, and sometimes polymers, which are suspended in a liquid carrier used to create the images. Liquid toners appear to dye the individual paper fibers. Liquid toners are often encountered in the so-called Indigo process.

Table 2 lists some common ingredients found in modern dry toners.

**Table 2** Ingredients commonly used in single- and dual-component dry toners

Single component (%)	Ingredient	Example	Two component (%)
45–60	Binding agent, polymer	Styrene/acrylate copolymer, polymethacrylate resin, polyester, etc.	2–5
	Pigment	Carbon black, Prussian blue, various organic pigments	
	Additive	Fused silica	
40–55	Carrier	Spherical ferrite	95–98
	Magnetic pigment	Magnetite, ferrite, or maghemite	

### Analytical Methods for Toners

**Optical microscopy** Dry and liquid toners can be distinguished by observation under low power magnification ( $\sim 20\times$ ) using a binocular stereomicroscope. If distinct particles of the toner can be observed, then it is a dry toner. Dry toners have a raised, often glossy appearance, and can be scraped off the page with a scalpel blade. In some instances, the morphological aspect of the dry toner can be easily distinguished, by observing the document with an angled or a transmitted illumination.

Liquid toners penetrate into paper fibers and appear as a thin coating through which paper fibers can be seen quite unmistakably. Liquid toner copies can sometimes be mistaken for documents produced by other forms of printing (for example, an ink jet printer or an inked stamp). Microscopic examinations, at a magnification of  $100\times$ , can also discriminate the toner's methods of fusion with the paper substrate.

While this is a highly discriminatory technique, it is also difficult to apply, and great prudence should be exercised in its employment. The differences between the fusion methods are slight and it can be difficult to appreciate them unless the observer has great familiarity with toner examination. The best way of acquiring the right know-how is to examine numerous documents produced on machines utilizing well-known fusion methods.

**Optoelectronics systems** Single-component toner powders, which contain magnetic material, incorporated into the toner particles when fixed onto the paper, exhibit magnetic properties similar to other forms of magnetic printing. The magnetic properties can be elegantly detected using an optoelectronic system and allows differentiating magnetic single-component toners from dual-component toners.

**Scanning electron microscopy (SEM)** It can be applied to acquire comprehensive information on the surface morphology of toner samples, which can be

used as a resource to differentiate toners. Very small samples (1 mm diameter) of toner bearing paper are isolated from a document using a proper cutting instrument. Samples are then fixed on an SEM holder with a double-sided adhesive tape and then coated with carbon, gold, platinum, palladium, or silver; the choice of the coating depends on the type of analysis conducted. Three photographs of each sample are recorded, using an enlargement of  $500\times$ , which is found to be the most valuable single magnification for observing the surface morphology. The basic method of toner fusion (liquid toner, dry toner with heat fusing, dry toner with cold pressure fusing, and dry toner with combined heat and pressure fusing) on plain paper may be determined by SEM.

**Scanning electron microscopy coupled with energy dispersive X-ray (EDX) microanalysis** X-ray microprobe analysis provides information on the inorganic content of the toner and allows to characterize elemental data for elements of atomic numbers from boron ( $Z = 5$ ) to uranium ( $Z = 92$ ). Samples are prepared and examined by the scanning electron microscope, using the same method as that of surface morphology examination. This method may confirm the detection of magnetic material and easily distinguishes single-component toners from dual-component toners.

Single-component toners are further divided according to the type of magnetic material present (i.e., maghemite, magnetite, or ferrite). Additional differentiations can be carried out according to the minor constituents present and their relative proportions. Dual-component toners do not contain significant quantities of mineral compounds, although some may show traces of iron. The latter may be explained as an incomplete separation of the two components allowing some traces of the carrier material to be found on the paper.

In general, SEM–EDX toner analysis can highlight the presence of the following mineral elements: Ca, Cl, Cr, Cu, Fe, Mn, Ni, S, Si, Sr, and Zn. It must be stressed that a paper blank must always be

examined. Some papers have a high inorganic content, due to the fillers used, and often the inorganic content of the paper substrate itself, is reflected in a spectrum thought to derive from the toner. Due to the heterogeneity of the paper and toner, in combination with the minute area of analysis, multiple sampling and analysis are strongly recommended. This can be facilitated using models of environmental scanning electron microscopes that have a larger specimen chamber where it is possible to examine, in a nondestructive manner, an A4-size document.

### Spectroscopic examinations

**Color measurement** In the case of colored toners, UV and visible (UV-Vis) transmission spectroscopy is successfully applied. If an isolated toner particle of cyan, magenta, yellow, and black is available on a document, the problem of color separation is irrelevant. However, in routine cases, color separation must be practiced under a microscope. The various colored toner particles, composing a colored image, are isolated, embedded in glycerin, and placed on a quartz slide. The UV-Vis spectra are measured using a microspectrophotometer that provides a more objective measure of colors, compared to the subjective evaluation of the naked eye or the use of an optical microscope alone.

**Fourier transform infrared analysis** FTIR spectroscopy is a suitable method for the differentiation of the organic binders present in toners. Many different techniques can be applied for analyzing toners using IR spectrometry. The analysis of toners may be effectively performed by microscopic attenuated total reflectance, with an internal reflection element (germanium crystal, for example) or by diffuse reflectance (DR) microscopic reflection-absorption techniques that have a diversity of reflecting media including mirrored slides, low sensitivity glass, metal disks, and aluminum foil, and transmittance. The choice of an analytical technique can be made based on equipment accessibility, ease of sample preparation, speed of analysis, and the kind of damage that may be done to the original document.

### Chromatographic examinations

**Thin-layer chromatography** Dye as well as binder components of photocopy toners may be analyzed with the help of TLC.

**Pyrolysis gas chromatography mass spectrometry** Pyrolysis gas chromatography, combined predominantly with mass spectrometry, is a powerful analytical method for the categorization of polymers. Moreover, it is useful for the characterization of

monomers, comonomers, additives, and impurities. During pyrolysis, which is usually performed at  $\sim 700^{\circ}\text{C}$ , large molecules are broken up into smaller ones which are then separated by GC and detected by MS. Pyrolysis GC-MS is a highly discriminatory method of analysis, which is quite time consuming and has a limited reproducibility.

**Laser microprobe gas chromatography-mass spectrometry** A laser microprobe is used to target selectively microscopic samples of toner, for subsequently GC-MS detection. Fused toners are directly analyzed *in situ* avoiding the traditional separation of toner from the paper substrate. This method offers a high spatial resolution and selectivity but further studies may be necessary to improve its reproducibility.

### Miscellaneous examinations

**Differential scanning calorimetry** A brief study using differential scanning calorimetry showed potential usefulness for thermal analysis but was not pursued due to the time-consuming nature of the technique.

### Interpretation of Toners Comparison

Forensic laboratories have established databases of toner composition used by electrophotographic devices. Maintenance of such a database requires tremendous labor. There are few manufacturers of toner but there are many formulations composed by each manufacturer. To complicate matters further, toners and gray toners manufactured with nonoriginal equipments do not always have the same chemical composition as original ones. For these reasons, the analysis of toner is most useful to determine whether or not two or more documents were printed with the same toner.

**See also:** **Archaeometry and Antique Analysis:** Dating of Artifacts. **Electrophoresis:** Overview. **Fourier Transform Techniques.** **Gas Chromatography:** Pyrolysis. **Mass Spectrometry:** Overview. **Microscopy Techniques:** Light Microscopy; Scanning Electron Microscopy. **Thin-Layer Chromatography:** Overview. **X-Ray Fluorescence and Emission:** Energy Dispersive X-Ray Fluorescence.

### Further Reading

Armitage S, Saywell S, Roux C, Lennard C, and Greenwood P (2001) The analysis of forensic samples using laser micro-pyrolysis gas chromatography mass spectrometry. *Journal of Forensic Sciences* 46(5): 1043–1052.



- Brandi J, James B, and Gutowski SJ (1997) Differentiation and classification of photocopier toners. *International Journal of Forensic Document Examiners* 3(4): 324–343.
- Cantu AA (1995) A sketch of analytical methods for document dating part I. The static approach: determining age independent analytical profiles. *International Journal of Forensic Document Examiners* 1(1): 40–51.
- Cantu AA (1996) A sketch of analytical methods for document dating part II. The dynamic approach: determining age dependent analytical profiles. *International Journal of Forensic Document Examiners* 2(3): 192–208.
- Claybourn M and Ansell M (2000) Using Raman spectroscopy to solve crime: inks, questioned documents and fraud. *Science & Justice* 40(4): 261–271.
- Ellen DM (1997) *The Scientific Examination of Documents*, 2nd edn. London: Taylor & Francis.
- Lennard CJ and Mazzella WD (1991) A simple combined technique for the analysis of toners and adhesives. *Journal of Forensic Science Society* 31(3): 365–371.
- Merrill R, Bartick E, and Mazzella WD (1996) Studies of techniques for analyzing photocopy toners by IR. *Journal of Forensic Sciences* 41(2): 264–271.
- Mizrachi N, Aizenshtat Z, Levy S, and Elkayam R (1998) Classification and identification of color photocopiers by FT-IR and GC/MS. *Journal of Forensic Sciences* 43(2): 353–361.
- Spence LD, Francis RB, and Tinggi U (2002) Comparison of the elemental composition of office document paper: evidence in a homicide case. *Journal of Forensic Sciences* 47(3): 648–651.
- Tebbett IR (1991) Chromatographic analysis of inks for forensic science applications. *Forensic Science Review* 3(2): 71–82.
- Totty RN (1990) Analysis and differentiation of photocopy toners. *Forensic Science Review* 2(1): 1–23.
- Zlotnick JA and Smith FP (1999) Chromatographic and electrophoretic approaches in ink analysis (Review). *Journal of Chromatography B* 733(1–2): 265–272.

## Systematic Drug Identification

N N Daíid, University of Strathclyde, Glasgow, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

One of the main activities in a forensic chemical laboratory is the examination of powders, pills, and plant materials thought to be illicit or controlled substances. The analysis of suspected drugs of abuse occurs over a number of stages. These involve the preparation of the analytical space within the laboratory, a physical examination of the item including its packaging, the use of a sampling protocol depending on the nature of the samples presented and the analysis required and the analysis of the sample. There are various analytical techniques commonly used in the forensic examination of suspected drugs and these include presumptive testing (color/spot tests), thin layer chromatography (TLC), liquid chromatography (LC), gas chromatography (GC) with either flame ionization detection (FID) or mass spectrometry (MS), Fourier transform infrared spectroscopy (FTIR), and ultraviolet (UV) spectroscopy. In many situations the data obtained from these analyses would then be compared with laboratory reference standards and/or a standard reference text such as Clarke's analysis of drugs and poisons.

The choice of technique which is used, depends upon the requirements of the analysis and specifically

the questions and hypothesis to be addressed. Specifically this commonly includes:

- **dmac\_aq>p0015**Determination of the identification or confirmation of the presence of a controlled substance in a sample;
- **qmac\_aq>p0020**Quantification of any controlled substance (i.e., determining the amount of the controlled substance present); and
- determination of links between samples or synthetic route used in sample production through chemical profiling or characterization of synthetic by-products and/or reaction impurities within the sample.

### Legal Aspects of Illicit Drug Analysis

The United Nations through their office on drugs and crime (UNODC) has produced over the years three major international drug control treaties, which are mutually supportive and complementary. In addition to including general provisions on illicit trafficking and drug abuse, they seek to ensure the availability of narcotic drugs and psychotropic substances for medical and scientific purposes and to prevent their diversion into illicit channels.

The treaties are:

- Single Convention on Narcotic Drugs, 1961 as amended by the 1971 Protocol concerning psychotropic compounds.

- Brandi J, James B, and Gutowski SJ (1997) Differentiation and classification of photocopier toners. *International Journal of Forensic Document Examiners* 3(4): 324–343.
- Cantu AA (1995) A sketch of analytical methods for document dating part I. The static approach: determining age independent analytical profiles. *International Journal of Forensic Document Examiners* 1(1): 40–51.
- Cantu AA (1996) A sketch of analytical methods for document dating part II. The dynamic approach: determining age dependent analytical profiles. *International Journal of Forensic Document Examiners* 2(3): 192–208.
- Claybourn M and Ansell M (2000) Using Raman spectroscopy to solve crime: inks, questioned documents and fraud. *Science & Justice* 40(4): 261–271.
- Ellen DM (1997) *The Scientific Examination of Documents*, 2nd edn. London: Taylor & Francis.
- Lennard CJ and Mazzella WD (1991) A simple combined technique for the analysis of toners and adhesives. *Journal of Forensic Science Society* 31(3): 365–371.
- Merrill R, Bartick E, and Mazzella WD (1996) Studies of techniques for analyzing photocopy toners by IR. *Journal of Forensic Sciences* 41(2): 264–271.
- Mizrachi N, Aizenshtat Z, Levy S, and Elkayam R (1998) Classification and identification of color photocopiers by FT-IR and GC/MS. *Journal of Forensic Sciences* 43(2): 353–361.
- Spence LD, Francis RB, and Tinggi U (2002) Comparison of the elemental composition of office document paper: evidence in a homicide case. *Journal of Forensic Sciences* 47(3): 648–651.
- Tebbett IR (1991) Chromatographic analysis of inks for forensic science applications. *Forensic Science Review* 3(2): 71–82.
- Totty RN (1990) Analysis and differentiation of photocopy toners. *Forensic Science Review* 2(1): 1–23.
- Zlotnick JA and Smith FP (1999) Chromatographic and electrophoretic approaches in ink analysis (Review). *Journal of Chromatography B* 733(1–2): 265–272.

## Systematic Drug Identification

N N Daíid, University of Strathclyde, Glasgow, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

One of the main activities in a forensic chemical laboratory is the examination of powders, pills, and plant materials thought to be illicit or controlled substances. The analysis of suspected drugs of abuse occurs over a number of stages. These involve the preparation of the analytical space within the laboratory, a physical examination of the item including its packaging, the use of a sampling protocol depending on the nature of the samples presented and the analysis required and the analysis of the sample. There are various analytical techniques commonly used in the forensic examination of suspected drugs and these include presumptive testing (color/spot tests), thin layer chromatography (TLC), liquid chromatography (LC), gas chromatography (GC) with either flame ionization detection (FID) or mass spectrometry (MS), Fourier transform infrared spectroscopy (FTIR), and ultraviolet (UV) spectroscopy. In many situations the data obtained from these analyses would then be compared with laboratory reference standards and/or a standard reference text such as Clarke's analysis of drugs and poisons.

The choice of technique which is used, depends upon the requirements of the analysis and specifically

the questions and hypothesis to be addressed. Specifically this commonly includes:

- **dmac\_aq>p0015**Determination of the identification or confirmation of the presence of a controlled substance in a sample;
- **qmac\_aq>p0020**Quantification of any controlled substance (i.e., determining the amount of the controlled substance present); and
- determination of links between samples or synthetic route used in sample production through chemical profiling or characterization of synthetic by-products and/or reaction impurities within the sample.

### Legal Aspects of Illicit Drug Analysis

The United Nations through their office on drugs and crime (UNODC) has produced over the years three major international drug control treaties, which are mutually supportive and complementary. In addition to including general provisions on illicit trafficking and drug abuse, they seek to ensure the availability of narcotic drugs and psychotropic substances for medical and scientific purposes and to prevent their diversion into illicit channels.

The treaties are:

- Single Convention on Narcotic Drugs, 1961 as amended by the 1971 Protocol concerning psychotropic compounds.



- United Nations Convention against the Illicit Traffic in Narcotic Drugs and Psychotropic Substances, 1988.

The UNODC's Legal Advisory Programme promotes adoption and practical implementation of these conventions. It addresses the needs of States that have outdated drug control legislation, are most vulnerable to criminal activity, or lack success in major prosecution and asset forfeiture casework. The program advises states on the drafting, adoption, and application of all necessary legislation.

Each country generally has its own drug legislation, which in general has undergone some evolution over the years. A comprehensive list of the different legislative controls are to be found on the United Nations web site at the following URL: [http://www.unodc.org/unodc/legal\\_library/index-countries.html](http://www.unodc.org/unodc/legal_library/index-countries.html).

## Laboratory Examination

### Preparation of the Analytical Space

One of the most important (and fundamental) parts of the examination of samples suspected of containing controlled substances is the initial preparation of the space where the analysis is to take place. Benches should be washed down with detergent, dried, and then swabbed with methanol or similar solvent. The swabs should be retained for analysis if required. Fresh bench covering should be laid on the bench. Sample packages should be opened individually and the bench should be cleaned as described, in between subsequent packages being opened. There should be frequent changes of gloves and meticulous notes require to be kept on all procedures undertaken.

## Physical Examination

### Packaging

Once drug samples are presented for examination, their packaging should be checked to ensure that it is intact. Any breach of the packaging should be reported and a decision taken as to whether the analysis of that item should be continued. Packaging materials for drug samples come in a variety of forms, from wrapping materials such as cling film (often around cannabis resin), tin, or aluminum foil (powder samples and some tablets) to larger polyethylene bags ranging from bank bags to bigger ziplock type bags. In some cases linkages can be made from sample to sample based upon the packaging materials present and it is well worth spending time on their

examination. For example, cling film can be physically linked using polarized light techniques and chemically linked using FTIR.

### Samples

The physical description of an item, as well as the types of analysis undertaken, will depend upon whether that item is considered a bulk or trace sample. Each item should be described fully (with diagrams if appropriate) including a description of color, smell, and any packaging materials, which may be present. If there are logos (ecstasy tablets) or marks on the items (blocks of resin or packages of drugs), these should be fully described and possibly photographed. Microscopic examination may also be used to examine the morphological characteristics of the material. This includes descriptions of different types of crystals or other solids, which may be present, the structure of tablets, the presence of different botanical features in seizures of plant materials.

**Trace versus bulk samples** A trace sample is considered to be one, which is barely visible to the naked eye. Samples such as this will be present on the insides of reaction vessels, scales, knives, and other drug paraphernalia. In many cases, trace samples are swabbed from these items and are analyzed using confirmatory techniques first. Such samples are very easily contaminated and significant care should be taken in their analysis. Bulk samples require a sampling methodology and are generally subjected to presumptive tests and TLC followed by confirmatory tests.

## Sampling Protocols

Many laboratories have developed their own sampling protocol for drug samples. In all cases, samples should be taken from items, which have the same morphology (also called sampling by attributes). This means that if a seizure is found to contain different types of materials, then each type should be sampled separately. This is particularly true for samples containing tablets with different logos or of different colors, or items, which have been packaged separately from each other. Where blocks of resin are being sampled, this should be away from the edges of the block as such edges can potentially be used to make physical fits between the blocks. In cases where samples of powder are seized, it is necessary to ensure that the powder is homogenous which is generally accomplished using standard techniques.

Two things generally dictate sampling protocols:

- A legal obligation which may require that all items in a seizure are described and sampled.

- A policy left to the expert, which will include a description of the items and the selection of a sample from the items.

The criteria for selecting the type of sampling protocol undertaken should include a balance between the loss of completeness versus time saving.

In actual fact most sampling protocols are based upon a mathematical distribution called the hypergeometric distribution (eqn [1]). Some laboratories have adopted this as their sampling policy and will only analyze a maximum number of samples or packages from any one seizure regardless of the total seizure size. The number analyzed varies depending on the level of accuracy (99%, 95%, etc.) which is acceptable.

$$P(m/n, M, N) = \frac{{}^M C_m ({}^{N-M} C_{n-m})}{{}^N C_n} \quad [1]$$

where

$${}^a C_b = \frac{a!}{(a-b)!(b!)}$$

and  $n$  is the size of the subpopulation analyzed,  $m$  a desired outcome,  $N$  the size of the sample, and  $M$  the number of positive outcomes in the sample.

An international system of sampling, which has been adopted and promoted by the UNDCP is as follows.

#### Powders and Tablets and Packages

- Single package of material – The material should be removed from all packaging and weighed to a constant dry weight. The sample is then homogenized and a sample removed.
- More than one package – The material in each package should be examined by eye for color differences. The contents of each package should be weighed (cleaning balance in between each weighing) and each package should be tested using a color spot test or TLC. If it is assumed that all packages contain the same material then;

if there are less than 10 packages all should be tested

if there are between 10 and 100 packages 10 should be tested at random

if there are more than 100 packages the square root of the total should be tested at random.

#### Liquids

If all of one phase then a sample of the liquid is removed for testing. If there are more than one phase

present then a sample of each phase should be removed for testing.

## Identification and Quantification Techniques

### Presumptive Tests – Identification of Drug Class

The first tests carried out on a sample are presumptive tests to give an indication of the class of drug which may be present in the sample. These are generally performed on clean porcelain tiles, in solution or on adsorbent substrates. The most common type of presumptive test involves the addition of various reagents to the sample to produce a color. Other tests involve the use of microscopy to identify trichomes in suspected cannabis samples or microcrystalline tests where the formation of specific crystals is indicative of the class of drug present. In all cases appropriate positive and negative control samples must be used. Table 1 lists some of the results obtained with different spot test reagents for common drugs of abuse.

**Advantages and disadvantages of presumptive tests** Most presumptive tests are carried out on porcelain tiles. There are a number of clear advantages to tests of this nature: they are cheap, quick, and easy to use. They also have the advantage of being portable if required and various ‘road side’ test kits have been developed on the basis of color tests.

The main disadvantages are that the colors formed are subjective, may change over time, and may be produced by more than one drug compound or by compounds other than controlled substances. These tests are also of limited sensitivity (few micrograms) and so are often not suitable for the analysis of trace samples.

### Thin Layer Chromatography – Identification of Drugs within a Class

Having tentatively identified the class or classes to which the drug belongs to using presumptive testing, the next stage in drug identification is to examine which types of drugs from the class are present in the seizure. This can be achieved using TLC, a technique which has the advantages of being both rapid and cheap. The separation process depends on the relative strength of interaction of the sample components with a stationary phase, and a solvent moving through this material, the mobile phase. The solvent in which the sample is dissolved must be one that is suitable for the analysis of the suspected compound.

**Table 1** Expected color reactions for the most common drug compounds

Test	Drug									
	Opiates	Amphet.	Cocaine	Cannabis	LSD	Barb.	Benzo.	Mescaline	Methaqualone mecoqualone	Psilocin psilocybn
Marquis	Purple/violet	Brown/orange/green								
Mandelin	Red/brown	Green	Orange			Slight green	Pink/yellow	Orange/red		Green/brown
Duquenois-Levine				Violet						
Dille-Koppanyi						Red/purple				
Zimmermann							Red/purple/yellow			
Vitali-Morin							Yellow			
Cobalt thiocyanate							Blue (slow)			
Ehrlich			Blue		Blue/violet			Blue		Grey/brown
Simon		Red/brown								
Fast Blue B				Red/purple						

**Suitable solvents** Suitable solvents for drug analysis are ones where the drug of interest is freely soluble and the solvent does not react with the drug or catalyze drug breakdown. The solvent should also be volatile and water free to allow for easy concentration of the sample. The sample to be analyzed together with positive and negative (solvent only) controls should all be spotted onto the same TLC plate and developed in the usual manner.

After development the plate is viewed under UV light and any spots present are marked. Other visualization techniques are commonly used and different classes of drugs require different techniques usually involving spraying the plates with various reagents (Table 2).  $R_f$  values are determined and compared with the controls run on the same plate and may be compared with literature values (though these should be interpreted with care). In the case of trace samples, the samples are swabbed using a buccal swab. The swab is first extracted (rinsed) with a solvent suitable for the analysis. The solvent is concentrated under nitrogen and analyzed for the presence of drugs. The swab is then rubbed over the surface of the item thought to contain the trace drug, the head cut off and extracted with solvent, and the extract concentrated under nitrogen and analyzed. The swab head and concentrated extract should be refrigerated and retained.

Despite being relatively cheap, rapid, and easy to interpret, TLC has a number of disadvantages which include:

- lack of resolution;
- lack of specificity of spray reagents giving similar color reactions for some compounds of the same class;
- edge effects on the TLC plate;
- variable  $R_f$  values; and
- relatively low sensitivity.

### Chromatographic Methods used as Confirmatory Techniques

Having identified the drug class and perhaps the drug within that class, the next stage of the analytical procedure is to confirm the identity of any drug present. Instrumental techniques are used to accomplish this and these can be either chromatographic or spectroscopic. Such techniques provide much stronger evidence than noninstrumental techniques.

Chromatographic techniques identify drugs through their physical and chemical properties of the molecule as a whole such as structure and size, whereas spectroscopic tend to identify compounds on the basis of certain parts present within the molecule. Instrumental techniques are used not only in

**Table 2** Some suggested TLC systems and visualization reagents for different drug compounds

<i>Drug</i>	<i>TLC mobile phase</i>	<i>Visualization reagent</i>	<i>Result</i>
LSD	Chloroform/methanol (9:1 v/v)	1. UV at 254 nm 2. UV at 360 nm 3. Ehrlich's reagent	Absorbs Fluorescence Blue/purple
Opiates	1. Ethylacetate/methanol/ammonia (85:10:5 v/v) 2. Chloroform/methanol (9:1 v/v)	1. UV at 254 nm	Absorbs
Amphetamines	Methanol/ammonia (100:1.5 v/v)	2. Acidified iodoplatinate 3. Dragendorff solution 1. Ninhydrin reagent and heat at 115°C 2. Result of 1 + acidified iodoplatinate 3. 0.5 mol l <sup>-1</sup> NaOH + 0.5% Fast black K	Blue/purple Orange Violet/pink Gray/violet/brown Brown/pink/red
Barbiturates	1. Ethylacetate/methanol/25% ammonia (85:10:5 v/v) 2. Chloroform/acetone (80:20 v/v)	1. UV at 254 nm	Absorbs
Cocaine	Methanol/25% ammonia (100:1.5 v/v)	2. Mercury(II) chloride–diphenylcarbazon	Blue/pink
Mescaline/ psilocin	Methanol/ammonia (100:1.5 v/v)	Acidified iodoplatinate 1. UV at 254 nm  2. Fluorescamine reagent 3. Ninhydrin reagent and heat at 120°C	Absorbs  Bright at 365 nm Violet

confirming the identity of a drug, which may be present but also in quantification (determining the concentration) of the drug which may be present.

### Liquid Chromatography of Drugs of Abuse

LC is a technique used commonly to identify and quantify drugs of abuse. The technique has a number of advantages and disadvantages specific to drug analysis.

Advantages:

- it is nondestructive and samples can be recovered if required;
- the analyte does not need to be volatile;
- the sample generally does not require pretreatment such as chemical derivatization;
- the analysis can be automated; and
- quantification can be achieved without the necessity of an internal standard.

Disadvantages:

- the analyte needs to have properties which can be detected in a liquid stream;
- in most cases a UV or diode array detector and so the analyte needs to possess chromophores;
- the sample needs to be soluble in a wide range of solvents;
- quantification can be slow; and
- large volume of solvents are used.

In general a mixture of reverse-phase and straight-phase systems are used depending on the drug under

test. LC analysis is particularly used where quantification of the target drug is required as the system can be run without the use of an internal standard. Difficulties arise in complex mixtures of street drugs where reaction by-products and impurities as well as additives can cause complications within the chromatogram. As in any forensic examination, the injection sequence for samples is of importance. Calibration standards and samples should always be interspersed by blank injections to ensure cleanliness of the instrument.

### Gas Chromatography

GC has advantages and disadvantages as a confirmatory technique when compared to LC. The GC system has a greater resolving power than LC systems and does not have the same problems associated with mobile phase choice or require large volumes of solvents. However, GC does require the compound to be thermally stable, volatile, and exhibit good chromatographic qualities. This often requires derivatization of the compounds.

### Chemical Derivatization of Some Drugs

Many drugs can be chromatographed using GC directly; however, a number of compounds may give rise to problems such as thermal decomposition (cannabinoids and cocaine alkaloids), reactions within the instrument between compounds in the injected mixture (morphine), adsorption, or coelution. In order to improve chromatographic results some drug compounds require derivatization where  $-OH$ ,

–NH<sub>2</sub>, and –COOH groups are modified. The choice of derivatization reagent depends upon the types of functional groups present. Derivatization alters the chemical structure of the compound to a certain degree, which will in turn influence the results of any subsequent mass spectra produced. There are some common derivatization reagents in use:

1. *N,O*-bistrimethylsilylic acid (BSA) is commonly used to derivatize cannabinoids, opiates and cocaine analogs (ecognine, etc., cocaine shows good chromatography).
2. *N,O*-bistrimethylsilyltrifluoroacetamide (BSTFA) is commonly used to derivatize LSD and psilocybin.
3. Trifluoroacetic anhydride (TFAA) and heptabutylr anhydride (HFBA) are commonly used to derivatize amphetamines.

GC and LC provide a means of separating the components of a complex mixture but neither technique can definitively identify any component. More than one component may have the same retention time for a given system or may have chromophores, which are difficult to differentiate. For this reason it is generally accepted that two (preferably three) independent techniques are used to identify the compounds present. The generation of UV absorption spectra using diode array detection (DAD) can help accomplish this in LC analysis though many compounds have similar UV spectra. In GC analysis, the system can be coupled to a mass spectrometer and a mass fragmentation pattern produced for each compound, which can be used to identify the compound together with the chromatographic data.

GC–FID is often used in drug analysis but increasingly routine is the use of GC–MS. In this case the GC is interfaced with a mass spectrometer detector (MSD). This can be either in place of a conventional detection system such as an FID or in tandem with an existing detector. When using an MSD the carrier gas is generally helium, which has low molecular weight (does not contribute to observed ion signal) and which is oxygen and water free.

The information derived from GC–MS includes retention time and the mass spectrum of each eluted peak. Principal peaks are used in the identification of unknown compounds where these are the peaks of the most intense ions (usually 6–8 peaks). These can be compared to reference data in an attempt to identify the compounds.

**Selective ion monitoring** If a drug sample being analyzed is complex, it may be difficult to obtain baseline resolution for the different components of

the mixture. Selective ion monitoring essentially means setting the mass spectrometer to examine the mass spectra produced for a particular ion peak or set of peaks. The total ion chromatogram (TIC) can then be scrutinized for the presence of these peaks. This method can be problematic. The selected ions only form a small part of the total ion count. If the sample is weak then the selected ions may not be visible against the background noise. Secondly ions chosen must be characteristic of the compound in question.

**General comments** Quantification can be carried out using a GC–MS (or a GC–MS with FID detector) either by using the internal standard technique or using a deuterated homolog (compound of similar structure) as the compound under test. Generally the methyl group attached to a nitrogen moiety or functional groups attached to the main structure are those which are deuterated. The replacement of hydrogen atoms in this manner alters their mass from 1 to 2, thus causing different fragmentation patterns. This removes the necessity for baseline resolution of the target peak and its deuterated internal standard. Quantification is determined using the ratio of the peak heights (or areas) of the specific ions within the mass spectrometer.

### Spectroscopic Techniques

**Fourier transform infrared spectroscopy** FTIR is an extremely useful technique for confirming the identity of pure compounds, but has limited value if used for mixtures of compounds. The technique is based upon the identification of functional groups within molecules where such groups vibrate (either through stretching or bending in various ways) when irradiated with specific wavelengths of light. These vibrations and their intensity (% transmission) are plotted against the frequency of light (cm<sup>–1</sup>) to which the sample is exposed to produce an FTIR spectrum. Portions of the FTIR spectrum are unique to the compound under test (this is called the fingerprint region). Unfortunately, because the majority of seized samples are mixtures of compounds, FTIR has limited practical use in the analysis of street samples of drugs of abuse. However, it does have the advantages of being nondestructive and not requiring derivatization.

**Ultraviolet–visible (UV–Vis) spectroscopy** UV–Vis spectroscopy, like FTIR, is a technique which is useful in the identification of pure drug compounds. Different compounds contain chromophores, which will absorb specific wavelengths of UV or visible



light. The technique obeys the Beer–Lambert law and as such the absorption of spectra generated at given wavelengths have the added advantage of being directly related to the concentration of the sample (this is the basis of the diode array detector used in LC). Normally UV and UV–Vis spectra are recorded at high and low pH and the results of both for the sample under question compared with known standards.

UV–Vis is a cheap and easy technique, which allows sample recovery and good discrimination between pure compounds without the need for derivatization. It has less application for street samples involving complex mixtures.

## Which Instrumental Technique to Use?

The choice of instrumental technique used in any analysis depends upon the motive behind the analysis and on the nature of the sample. Normal procedure when dealing with bulk samples is to perform presumptive color tests to identify the drug class followed by TLC to identify the specific member of the drug class. When dealing with trace samples, it is often the case that these nonconfirmatory tests are circumvented and confirmatory tests only are employed. The choice of confirmatory test now depends on the motive behind the analysis as indicated below. The choice of technique will also depend upon the nature of the analytes under investigation.

For example, large compounds such as the cannabinoids are not particularly thermally labile and benefit for derivatization, however, this has implications in drug profiling as derivatization results in altering the chemical structure. LC analysis of cannabinoids will effect their identification but difficulties in resolution may hamper quantification and profiling. In practice many laboratories now identify cannabis on the basis of microscopy rather than instrumental techniques.

Samples thought to contain heroin require derivatization to help resolve some of the components though diamorphine itself chromatographs well without derivatization. Introduction into the GC system causes acetylation of morphine and monoacetyl morphine thus inflating the concentration of diamorphine above the actual value if GC is used as the quantifying technique. Heroin samples are generally profiled using a GC/MS system.

Amphetamine compounds are generally easily analyzed using either GC or LC and are profiled using a GC/MS system. LC is useful for quantification though a common diluent, caffeine has a significantly different molar extinction coefficient than

the amphetamines causing a large peak on the chromatogram, which may cause coelution problems. There are also coelution problems with some of the structurally similar compounds.

Barbiturates and benzodiazepines are difficult to analyze both by GC and LC. In GC analysis, barbiturates are derivatized on column by the addition of  $0.2 \text{ mol l}^{-1}$  trimethylanilium hydroxide in methanol and concentrating the solution using  $\text{N}_2$ . The sample is injected into GC where and alkylation reaction takes place. The sample retention is governed by the length of carbon chain at the C-5 position and the degree of polarity of the molecule. In general as the carbon chain length increases, so does the lipophilicity and consequently the retention time. Where chain lengths are the same, the degree of polarity becomes the deciding factor with more polar compounds eluting earlier.

## Data Interpretation

There are a number of mathematical techniques commonly used with data obtained from instrumental techniques. These enable the verification of the instrumental analysis with respect to reliability and reproducibility of response, comparison of retention time data, and the determination of the concentration of active component (to greater or lesser degrees of accuracy and with slight variation depending on the type of analytical technique used).

### Relative Retention Time

The relative retention time (RRT) is often calculated for peaks in chromatograms particularly when its identity is unknown. This can be compared to other sample data or data generated using standards or from a database generated under the same chromatographic conditions. An internal standard is added to the test solution and the following equation is used. RRT values can be calculated for both GC and HPLC analysis:

$$\begin{aligned} \text{Relative retention time} \\ &= \frac{\text{Retention time for analyte}}{\text{Retention time for internal standard}} \quad [2] \end{aligned}$$

It is important to be able to measure the reliability and reproducibility of analysis. This is usually conducted on a daily or more often weekly basis by analyzing standard samples. Both repeat injections of the same standard (showing instrumental variation) and different samples of the same standard (showing extraction and sample variation) and apply an appropriate statistical analysis to the results.



### Relative Standard Deviation (Coefficient of Variation)

This is a measure of the standard deviation expressed as a percentage over a number as of analytical results. The relative standard deviation (RSD) value (%) should ideally be as low as possible showing a high analytical reproducibility.

$$RSD = 100 \left( \frac{s}{\bar{x}} \right) \quad [3]$$

## Quantification Techniques

### Single Point Estimates

This method is one of the quickest that can be used, however, an assumption is made that the instrumental response is proportional to the concentration of the analyte under test and that the concentration of the sample is within the linear dynamic range of the detector. One standard solution is analyzed along with the sample and eqn [4] is used to relate their concentrations to each other. Equation [4] illustrates the application for LC analysis; if GC analysis were being carried out each solution would also include an internal standard.

$$\frac{\text{Peak area sample}}{\text{Peak area standard}} = \frac{\text{Concentration of sample}}{\text{Concentration of standard}} \quad [4]$$

$$\frac{\text{Peak area sample}}{\text{Peak area standard}} \times \text{Concentration of standard} = \text{Concentration of sample}$$

### Two-Point Estimates

In this case two standards are analyzed, one of a concentration higher than is expected in the sample and one of a concentration lower than expected.

Again the assumption is made that all measurements are made within the linear range of the detector. The relationships between the different standards and their respective responses are solved as simultaneous equations to determine an equation of the straight line between the points. The concentration of the sample is determined using this equation.

### Regression Analysis

This is by far the most accurate method for quantifying samples. It does not rest on the assumption that the concentrations determined are within the linear range of the detector but rather verifies whether this is correct. The method also involves running repetitive samples of the same standard and as such will take into account variation of detector response. However, the method does require analyzing many samples (at least 2 and preferably 3 data points at five different concentrations) and as such can be time consuming.

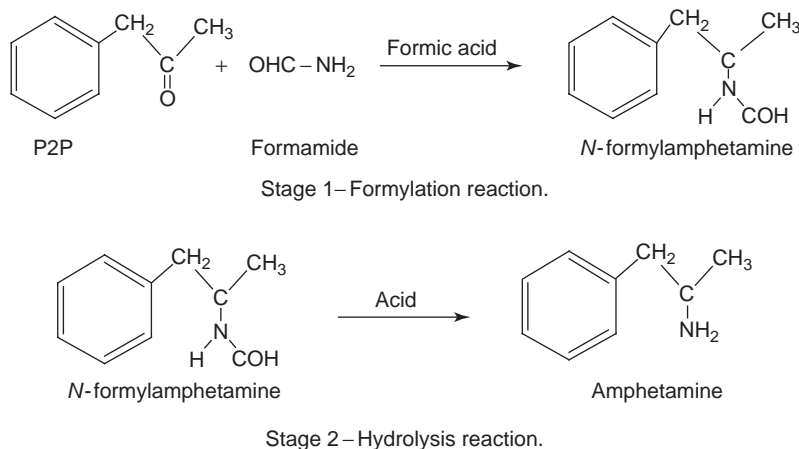
## Techniques in Drug Profiling

In the clandestine preparation of illicit drugs, specifically amphetamine and its analogs a number of common synthetic routes are utilized (Figure 1).

As a consequence of this route, specific impurities can be identified in GC analytical traces obtained from submitted samples and these can be used to determine which manufacturing route was utilized as well as in investigating links between illicit samples through impurity profiling.

### Organic Impurity Profiling

At the present time, chemical profiling methods are mainly based on determining and quantifying



**Figure 1** Leuckart synthesis of amphetamine.

organic impurities present in the seizure. In the case of synthetic drugs, some of these impurities and intermediates may be route specific. Techniques such as GC-FID, LC, and FTIR may be employed but perhaps the most beneficial is GC-MS, whereby, the identity of eluted components can be determined from the corresponding mass spectrum. Ultimately, determination of the organic impurities aims to identify common batch links among illicit seizures, while quantification has ramifications in subsequent criminal proceedings. Illegally synthesized drugs may eventually be distributed in more than one country and hence, the need for a universal method to establish links between seizures is apparent. Recent work has focused on the development of computer-based programs to improve the accumulation and appropriate accessibility of drug intelligence data.

Early chemical profiling attempted to establish batch links among illicit drug seizures by comparative studies, based on both qualitative and quantitative information, although success in achieving this has been varied. Such studies have been applied to link heroin and cocaine samples to common origins, based on the concentration of the organic impurities present. Isotopic analysis of impurities and intermediates has been employed in the comparison of seized heroin, methamphetamine, and ecstasy samples. X-ray diffraction and NMR spectroscopy (both  $^1\text{H}$  and  $^{13}\text{C}$ ) have been used to

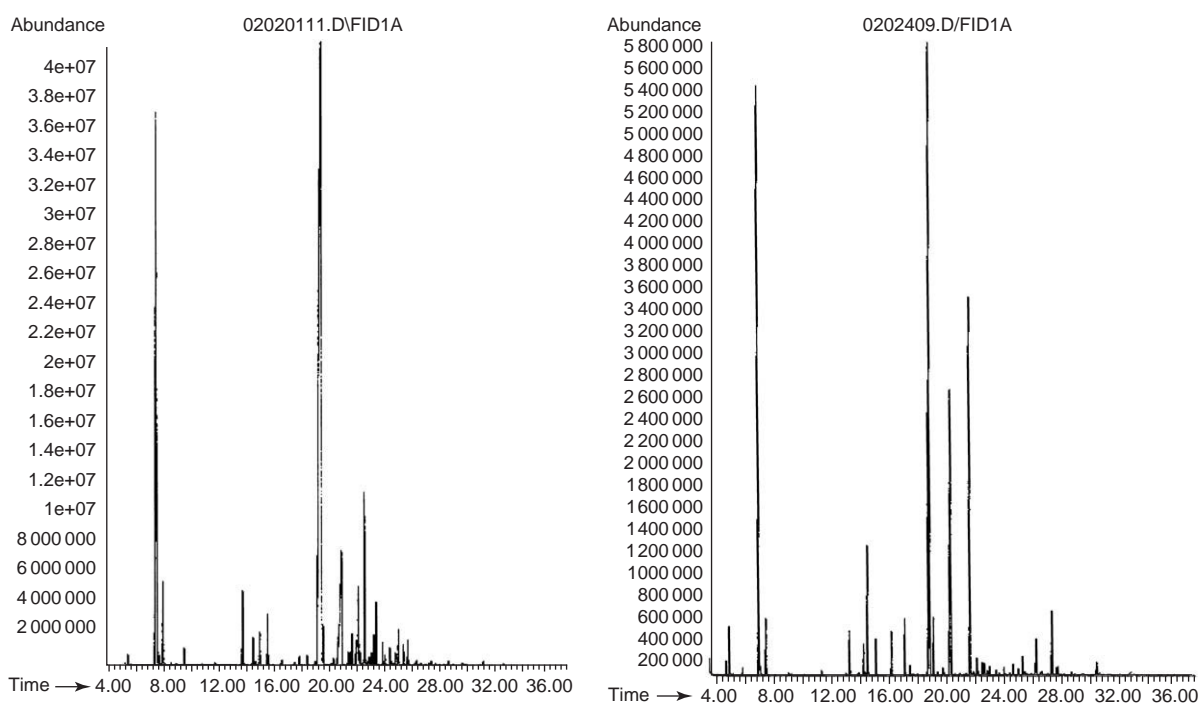
differentiate the origin of illicit ecstasy seizures (Figure 2).

### Inorganic Impurity Profiling

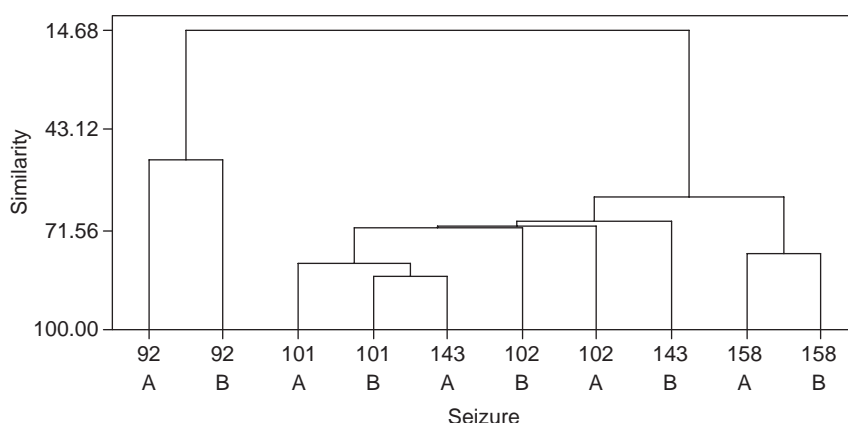
Since 1980, ICP-MS has emerged as a major and powerful technique in the area of elemental analysis. It offers extremely low detection limits which range from sub parts per billion (ppb) to sub parts per trillion. In addition, these detection limits are broadly achieved for almost all the elements across the periodic table. Also, the simple nature of the mass spectra of the elements makes this technique a quick tool for automated qualitative, semiquantitative, and quantitative elemental analysis. Metal content of bulk drug substances may originate from different sources:

- from the plant containing the drug which is then extracted (Poppy, Cannabis, or coca leaf);
- from catalysts, reagents, and solvents used in the extraction and/or syntheses;
- from exposure to air-borne particles;
- from container/closure systems, etc. used in synthesis/extraction.

There exists, therefore, a potential for use of this technique in relation to drug profiling. Moreover, easy sample preparation (powders directly dissolved in dilute nitric acid, with or without help of



**Figure 2** GC-TIC of impurity profiles of two unrelated (street) amphetamine samples.



**Figure 3** HCA of ecstasy tablets indicating linkages between samples.

microwave heating) and quick analysis time (about 2 min per sample) make this technique very attractive for drug profiling.

### Isotope Ratios

Stable isotope analysis is based on the fact that the major elements of organic compounds exist in their naturally occurring isotopic form. The corresponding stable isotope ratios ( $^{18}\text{O}/^{16}\text{O}$ ,  $^{15}\text{N}/^{14}\text{N}$ ,  $^{13}\text{C}/^{12}\text{C}$ , and  $^2\text{H}/^1\text{H}$ ) in natural products depend not only on the biosynthetic pathway but also on the environmental conditions such as humidity or temperature. Consequently, it is possible to use such analyses to determine the geographical origin of natural products. They are already routinely used by many food control laboratories to detect, for instance, addition of foreign sugar in pure fruit juices. For synthetic substances, the isotope ratios are linked to the synthetic conditions (reagents, time, catalysis, etc.) used and to the purification processes, e.g., distillation and it is possible to discriminate between batches of substances. For instance, analgesic drugs such as acetaminophen and aspirin, even manufactured following precise protocols, are significantly heterogeneous in isotopic composition. One advantage of stable isotope analysis is that isotope ratios are intrinsic parameters of the drug substance; therefore, they are not modified by partial degradation or addition of foreign materials during the distribution process.

### Data Analysis in Drug Profiling

The power of chemical profiling has been further enhanced through the application of chemometric procedures to the analytical data generated.

Amphetamine samples have been linked, according to the impurities present, using the Quotient method and a supervised learning method known as SIMCA (soft independent modeling of class analogy), both of which were applied to develop a harmonized method for heroin profiling. Canonical variate analysis has also been used to classify methylamphetamine hydrochloride according to synthetic route, again based on the chemical impurities present and multivariate methods of hierarchical clustering, principal component analysis, and  $k$ -nearest neighbors were utilized to establish batch links in illicit heroin samples (Figure 3).

**See also:** **Chemometrics and Statistics:** Multivariate Classification Techniques. **Forensic Sciences:** Drug Screening in Sport; Illicit Drugs; Thin-Layer Chromatography. **Gas Chromatography:** Overview; Mass Spectrometry; Forensic Applications. **Liquid Chromatography:** Clinical Applications. **Microscopy Applications:** Forensic. **Spot Tests.**

### Further Reading

- Cole MD (2003) *The Analysis of Controlled Substances*. Chichester: Wiley.
- Gough TA (1991) *The Analysis of Drugs of Abuse*. Chichester: Wiley.
- Laing R (2003) *Hallucinogens, a Forensic Drug Handbook*. New York: Academic press.
- Moffat A, Osselson MD, and Widdop B (2003) *Clarke's Analysis of Drugs and Poisons*. London: The Pharmaceutical Press.
- Tebbett I (1992) *Gas Chromatography in Forensic Science*. Chichester: Ellis Horwood.
- Yinon J (1995) *Forensic Applications of Mass Spectrometry*. Boca Raton, FL.

## Thin-Layer Chromatography

**R E Ardrey**, University of Huddersfield, Huddersfield, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 5217–5222, © 1995, Elsevier Ltd.

### Introduction

The extent to which thin-layer chromatography (TLC) is used in forensic investigations varies considerably from laboratory to laboratory and is now very dependent on the alternative analytical techniques to be found within any particular laboratory. Invariably TLC is but one of a number of techniques that could be used for a particular analysis. On many occasions the alternatives are likely to be more precise and/or specific than TLC and, since these criteria are usually of paramount importance, are likely to be used preferentially. For example, the combination of radioimmunoassay, to carry out the rapid screening of samples for the presence of a member of a particular class of drug, and gas chromatography–mass spectrometry (GC–MS) for the identification or confirmation of the presence of a particular drug, is likely to provide a more precise and sensitive analysis than will the use of TLC.

The major advantages of TLC are its simplicity, low cost, and the selectivity of detection that may be obtained by the use of a wide range of location procedures, and these account for its use in many laboratories. Its major disadvantage can be its lack of sensitivity. This has been highlighted by the tendency of modern drugs to be more potent than in the past and consequently to be prescribed in much smaller doses. The amount of active material that is encountered during certain investigations is therefore often below the detection limits afforded by TLC, and in these cases another, more sensitive analytical methodology is obligatory.

### Illicit Drugs and Toxicological Studies

In forensic science, ‘drugs’ are encountered in two specific sets of circumstances, each of which requires quite different investigative procedures.

The first, the drugs investigation, is that in which the possession, illicit manufacture, and/or distribution of drugs of abuse, such as cannabis, heroin, amphetamine, and LSD, are involved. This type of

investigation is usually characterized by a large amount, in analytical terms, of material that is available for analysis; this is usually in the milligram range, but often kilograms of sample are accessible. The sample may consist of a powder and/or tablets found in the possession of an individual, it may originate from a much larger seizure of material, for example, a consignment being smuggled across international borders, or may be from a number of samples taken from the site of manufacture of the drug, i.e., an illicit laboratory.

For drugs analyses the major applications of TLC are concerned with the identification of cannabis and the screening of drug powders and tablets for their active constituents.

Two TLC systems are in use for cannabis investigations. The first consists of a silica gel plate sprayed with diethylamine immediately before use and a mobile phase of xylene–hexane–diethylamine (25:10:1 by vol). The second consists of a mobile phase of toluene in an open tank and a silica gel plate sprayed with a 10% solution of silver nitrate and dried before use. A general location reagent is Duquenois reagent (vanillin (2 g) and acetaldehyde (0.3 ml) in ethanol (100 ml)), which, after overspraying with hydrochloric acid, gives blue to violet colors with cannabinoids. A more specific reagent is Fast blue B solution (a freshly prepared 1% solution of Fast blue B salt (diazotized *o*-dianisidine)), with which cannabidiol gives an orange color, cannabinol a violet color, and  $\Delta^9$ -tetrahydrocannabinol a red color. The colors obtained with the Fast blue B reagent may be intensified by overspraying with sodium hydroxide or by exposing the plate to ammonia fumes.

The majority of the active ingredients of drug powders that are encountered in forensic examinations are basic in nature and may therefore be examined using one of the general systems described in the article on Toxicological Analysis. Visualization is carried out using a general reagent, such as acidified iodoplatinate solution (hydrochloric acid (5 ml) added to a solution of platinum(IV) chloride (0.25 g) and potassium iodide (5 g) made up to 100 ml), which, for example, gives violet colorations with amines. Since the number of individual drugs encountered in these types of investigation is relatively small, a better indication of the drug involved may be obtained by use of a more specific location reagent, such as Marquis reagent (a mixture of formaldehyde solution (1 ml) with sulfuric acid (9 ml)) for alkaloids related to morphine,

and/or running control substances on the same plate as the unknown sample.

A number of the tablets that are encountered during the examination of 'illicit' drugs contain the same active ingredients as the drug powders described above. Many, however, do not, and in these circumstances the scope of the investigation needs to be widened to include the large number of other substances that may be legitimately prescribed but which are nevertheless subject to abuse in some way. In this case a general screen for acidic and neutral drugs, as described in the article on TLC in analytical toxicology, would be carried out.

In the second set of circumstances – the toxicological investigation – 'drugs' are encountered after ingestion by an individual. These may include drugs of abuse, as well as those prescribed for medicinal use and a number of other miscellaneous poisons. Identification of a drug may be required to ascertain the cause of a suspicious death, to determine the cause of impairment of function, for example, in a driving case, or determining the involvement of drugs as a contributory factor in a crime such as murder or rape. The major differences between cases involving illicit drugs and toxicological investigations are concerned with the presentation of the sample, the amount of sample present, and the presence of other materials that may complicate the analysis. In drugs cases the compounds of interest may be separated relatively easily from the few other materials present, such as excipients. In toxicological cases the compounds of interest are invariably part of a complex biological matrix, such as urine, blood, or tissue, and often bound to proteins from which they must be released before analysis. In drugs cases the compound(s) of interest may usually be defined readily; toxicology cases may be complicated by the presence of the many naturally occurring compounds in the body and the presence of drug metabolites, which in some cases are found to the total exclusion of the parent drug. The amount of active material in a toxicology case, in contrast to the milligrams or kilograms available when illicit drugs are involved, is invariably at the nanogram per milliliter level or below. TLC may be used for samples, such as stomach contents and urine, that contain the greatest amount of drugs and their metabolites, but the levels found in blood are usually low and preclude the use of TLC. The methodology for this type of analysis is, again, covered in the article on TLC in analytical toxicology.

## Explosives Residues

Detection of many commercially available explosives found at the site of explosions may be achieved using

TLC with visualization using a modified Griess reagent. The ability of this method to provide a definitive identification has been questioned because of the false positives that may be observed from totally unrelated compounds. When TLC is employed for the examination of explosives residues it is therefore recommended that further confirmatory analyses should be carried out before the presence of explosives is reported.

Extraction of an explosive may be achieved using acetone or toluene in which nitroglycerin (NG), ethylene glycol dinitrate (EGDN), RDX, HMX, pentaerythritol tetranitrate (PETN), tetryl, and the nitrotoluenes are sufficiently soluble to allow detection at the levels that are usually encountered. A combination of mobile phases must be used for their separation by TLC because RDX and HMX are unresolved using ethyl acetate–petroleum ether, while chloroform–methanol will separate these two explosives but will not separate NG and PETN. Visualization is with a modified Griess reagent after hydrolysis with sodium hydroxide at 105°C. The Griess reagent used consists of sulfanilic acid (1 g) in 30% acetic acid (100 ml) mixed with glacial acetic acid (100 ml) added to the supernatant liquid obtained after dissolving  $\alpha$ -naphthylamine (1 g) in boiling water (230 ml) and allowing to cool. The reagent is highly sensitive to nitrates and to most nitro-containing organic explosives. Organic nitrites give pink to red colors; inorganic nitrites give red colors that immediately fade to yellow. If no color is obtained initially, the addition of zinc dust will bring about the rapid development of pink to red colors if inorganic nitrates or organic nitro compounds are present.

Confirmation of any positives is advisable by using techniques such as liquid chromatography (LC) in conjunction with MS or by infrared (IR) spectroscopy and GC in combination with MS.

## Gunshot Residues

A similar system to that used for explosives may be used for confirming the presence of, and identifying, smokeless powder flakes taken from hand swabs. In this case the Saltzman–Griess reagent is used as a test for nitrite ions released from NG and nitrocellulose (NC) by alkaline hydrolysis. NC exists in two forms, propellant grade and nonpropellant grade, the latter being used as a base for some lacquers, inks, resins, and films. These two forms differ in the lengths of their cellulose polymer chains and it is necessary to differentiate between them in order to confirm conclusively the presence of smokeless powder. This is accomplished by developing the TLC plate



consecutively with two mobile phases, methylcellosolve–95% ethanol (15:85 v/v) and toluene–petroleum ether–ethyl acetate (12:12:1 by vol). In the first system propellant-grade NC remains at the origin while nonpropellant NC and NG move with the solvent front. In the second system NG has an  $R_F$  of 0.3 while the NC remains stationary. A pink color is obtained from NC and NG with the visualization reagent that consists of 0.1% by weight of *N*-1-naphthylethylenediamine dihydrochloride in distilled water, 80 ml of which is used to dilute glacial acetic acid (14 ml). Sulfanilic acid (0.5 g) is then dissolved in the resulting solution, which is made up to 100 ml with water.

To overcome the problem of false positives it is now more usual, whenever possible, to use a scanning electron microscope to detect the metals, such as barium, associated with discharging of a firearm while the organic constituents of smokeless powder are preferably analyzed as for explosive residues above, i.e., by using LC followed by MS confirmation, GC–MS, etc.

## Inks

One of the major differences between forensic science and other areas of analytical chemistry is that in a large number of cases the determination of the absolute chemical composition of a material, for example, a fiber, an ink sample, or a paint flake, is not the major requirement. It is rather determining that two samples are, or are not, identical, i.e., are so similar that there is an incontrovertible link between a suspect and a crime or, alternatively, that they are so dissimilar that it can be shown incontrovertibly that there is no link between a suspect and a crime. For this to be accomplished successfully, therefore, any particular class of evidential material must possess a particular discriminating feature and there must be an appropriate analytical technique available to allow meaningful differences in this feature to be recorded.

TLC is used extensively for the study of inks in many types of case, particularly those involving forgery of cheques. The comparison and discrimination that may be obtained is in terms of both the visible dyestuffs and the ultraviolet (UV)-fluorescent components that are present. A number of methods are available for the comparison of the color of ink and these include microscopy, IR luminescence and reflectance, and also TLC. Ink colors can change rapidly when the writing is exposed to light and this tends to complicate the results obtained by microscopic examination, while

IR luminescence may be affected by the exposure of the ink to moisture and many common household chemicals.

Ballpoint ink samples may be removed from the written line using a blunt hypodermic needle or a scalpel and then dissolved in pyridine. A mobile phase of ethyl acetate–ethanol–water (70:35:30 by vol) has been used to separate the constituent dyes sufficiently to permit the discrimination of a number of examples, although on occasions it proves necessary to employ a second mobile phase, *n*-butanol–ethanol–water (50:10:15 by vol) to achieve sufficient discrimination.

This second mobile phase has been used for the examination of black ballpoint inks, with a densitometer being used to scan the TLC plate and locate the positions of the individual dyes after UV irradiation. The wavelength chosen, 580, 425, or 420 nm, was on the basis of the color of the ink and the dyes present. The resulting trace allows a more easy rigorous comparison of the results from different ink samples and allows any differences that there may be to be more easily discerned.

An examination of a number of blue and black fibertip pens and fluid pen inks, applied to both card and photographic papers, used a mobile phase similar to that used for the examination of ballpoint inks, namely *n*-butanol–ethanol–water (4:1:1 by vol), and showed a general agreement in the number, color, and  $R_F$  values of the dyes present in each ink despite there being a visual difference between samples of the same ink on paper and card. The effect of age of the ink sample was ably demonstrated when a similar investigation was carried out on 12-month-old dried ink samples, and four pairs of sample were erroneously concluded to be different. Good discrimination of inks from fibertip pens has also been reported using ethanol to extract the dyes from the ink and mobile phases of *n*-butanol–acetic acid–water (6:1:2 by vol) and *n*-butanol–acetic acid–water–dioxan (6:2:2:1 by vol) to effect separation of the components prior to both visual and UV examination of the plates.

The differences in analytical results obtained as inks age has been attributed to a decrease in the extraction efficiency of the solvent toward the constituents of the ink; the longer the ink has been on the paper the harder they are to extract. This has been attributed to a number of factors either in isolation or in combination. These include evaporation of the solvent from the ink, chemical reactions such as oxidation and polymerization of the constituents of the ink, and ink–paper interactions. The method used involves comparing the concentration of the components of the ink from a sample of unknown age with



those obtained from samples of known age. Because of the complicated factors involved, the method is only applicable if the ink used to provide the control results is of the same composition as the aged sample and has been used on the same type of paper. UV-visible spectrophotometry, TLC with densitometry, LC, and fluorescence were used to obtain the concentration measurements, but it was concluded that the TLC method was one of the more tractable since it did not reflect the solvent used for extraction. Extraction of the ink was effected with toluene or benzyl alcohol and the dyes were separated using a mobile phase of ethyl acetate-ethanol-water (70:35:30 by vol).

Although many investigations involve handwritten documents, a significant number are encountered that have been produced using typewriters. The comparison and discrimination of typewriter ribbons may be carried out in a similar way to that described for inks and a meaningful comparison may be made from a dye extracted from a single character. Pyridine is used to extract the ink from the ribbon. Using a combination of mobile phases of ethyl acetate-ethanol-water (70:35:30 by vol) and *n*-butanol-ethanol-water (50:10:15 by vol) it proved possible to distinguish all of the ribbons that were studied that were ostensibly of the same color but produced by different manufacturers. It was also possible to differentiate between different dye batches, and it was therefore concluded that if inks were indistinguishable using these two TLC systems, they originated either from the same ribbon or from one with the same ink formulation.

## Dyes

A major use of TLC in forensic science is the analysis of the dyes to be found in fibers.

There are a number of different types of fiber that may be encountered and equally there are a number of different classes of dye, although some of these may only be associated with particular types of fiber. The type of fiber and the class of dye both affect the detailed scheme employed for the extraction of the dye from the fiber and the subsequent chromatographic analysis of the extract. A detailed discussion of the extraction schemes employed is beyond the scope of this text but, for example, acid dyes are best extracted using a mixture of pyridine and water (4:3 v/v), while reactive dyes require extraction with 1,2-diaminoethane at 90°C for 1 h. Schemes have been developed for the extraction of: (1) acid-chrome dyes from wool; (2) direct and reactive dyes from cellulosic fibers; (3) azo dyes from

cotton viscose fibers; (4) acid and disperse dyes from polypropylene; and (5) disperse, acid, and basic dyes from polyamides, polyesters, and polyacrylonitriles.

The TLC systems then employed for the discrimination of fibers is dependent upon the type of dye involved and the practical approach will involve the use of as many, or as few, as are necessary to provide discrimination. The systems used are listed below.

### *Acid dyes*

- chloroform-water-methanol-880 ammonia (11:1:7:1 by vol);
- butan-2-ol-acetone-water-ammonia (4% v/v) (5:5:1:2 by vol);
- pyridine-amyl alcohol-ammonia (10% v/v) (4:3:3 by vol);
- butan-1-ol-glacial acetic acid-water (upper phase) (20:10:50 by vol);
- butan-1-ol-ethanol-water-glacial acetic acid (60:10:20:5 by vol);
- *n*-butyl acetate-pyridine-water (2:2:1 by vol); and
- butan-1-ol-880 ammonia-pyridine (4:1:3:2 by vol).

### *Basic dyes*

- butan-1-ol-glacial acetic acid-water (upper phase) (4:1:5 by vol);
- chloroform-propan-2-ol-pyridine-glacial acetic acid-water (6:8:3:1:1 by vol); and
- chloroform-glacial acetic acid-formic acid-water (8:6:6:1 by vol).

### *Direct dyes*

- butan-1-ol-ethanol-880 ammonia-pyridine-water (8:3:4:4:3 by vol);
- pyridine-butan-2-ol-880 ammonia (9:6:6 by vol); and
- pyridine-amyl alcohol-880 ammonia (4:3:3 by vol).

### *Disperse dyes*

- toluene-pyridine (8:2 v/v);
- hexane-methyl acetate-acetone (5:4:1 by vol); and
- toluene-methanol-acetone (20:2:1 by vol).

### *Reactive dyes*

- methanol-amyl alcohol-water (5:5:2 by vol) and
- propan-1-ol-methanol-water-880 ammonia (8:6:2:1 by vol).

An alternative parameter that has been used for the discrimination of fibers is an examination for the presence of optical brighteners that originate from the detergent regime to which the fibers have been exposed. These brighteners are mainly coumarin, benzazole, or stilbene derivatives and may be extracted, for all fibers except acrylics, using dimethyl-formamide in subdued light. In the case of acrylics, pyridine must be used for extraction. Four mobile phases have been used for separation of the individual brighteners: chloroform–isopropanol (100:1.5 v/v); acetone–5 mol l<sup>-1</sup> aqueous ammonia–n-heptane (45:5:10 by vol), methyl ethyl ketone–5 mol l<sup>-1</sup> aqueous ammonia–methanol (40:15:10 by vol), and pyridine–n-butanol–ethanol–880 ammonia (30:40:10:20 by vol). Their visualization is achieved using UV irradiation at 366 nm.

A TLC method for the analysis of the dyes extracted from car sidelights has been used to distinguish similarly colored lenses.

## Other Evidential Materials

Paint is encountered in a number of types of crime, particularly burglary where a door or window has been forced to gain entry and in a range of car-related crimes such as hit-and-run incidents. TLC has been shown to provide good discrimination between visually similar colored gloss paints on the basis of the organic pigments present. Red, orange, yellow, dark-blue, purple, and brown paints are best examined by making an extract of a paint flake with dichloromethane and using a TLC mobile phase of chlorobenzene–toluene–1,2-dichloromethane (1:1:1 by vol). This system is not as effective for green, blue–green, and light-blue paints, which are best studied by extraction using 1,2-dichlorobenzene and a TLC mobile phase of 1,2-dichlorobenzene–toluene–1,2-dichloroethane (2:1:1 by vol), which also improves the discrimination of dark-blue, purple, and brown paints.

The analysis of a pharmaceutical preparation can yield much useful information in addition to the identity of the active ingredient, the determination of which is usually a trivial analytical task. Illicit drug preparations seized in different locations are studied to determine whether they originate from the same source. Law-enforcement agencies may thus be provided, at any particular time, with an indication of the number of primary sources of a drug and are consequently able to utilize their resources more

effectively. Large sums of money are spent developing and testing drugs to ensure their safety before they reach the market, and this investment is protected by means of patents. Pharmaceuticals may therefore be synthesized and distributed with a high financial return by those who are not themselves prepared to go through this lengthy, and costly, procedure but who prefer to utilize procedures patented by others. Monitoring of patent infringements is an essential part of the analytical work of a pharmaceutical company. One parameter that may be used during both of these types of investigation is the characterization of the food dyes that are often used in tablets to give them a characteristic color. TLC forms an important part of the procedure used for their analysis. These dyes may be extracted using methanol or water, the stability and solubility of the dyes depending on the acidity of the extraction medium. The extract is screened using TLC with a mobile phase of trisodium citrate–water–ammonia (2 g:100 ml:5 ml) and this is often sufficient to be able to show the similarity of the extracts from different tablets. Should identification of the dyes be necessary this may be carried out using visible spectrophotometry at three different pH values, to give tentative identifications, followed by LC for a positive identification.

TLC may also be used for the comparison of cosmetic dyes; for example, those that occur in lipsticks, and of the rubber in tyres and for the analysis of bitumens.

**See also:** Cosmetics and Toiletries. **Forensic Sciences:** Explosives; Fibers; Gunshot Residues; Paints, Var-nishes, and Lacquers; Systematic Drug Identification. **Thin-Layer Chromatography:** Overview; Principles; Plate Technology; Method Development.

## Further Reading

- De Zeeuw RA, Francke JP, Degel F, *et al.* (1992) *Thin-layer Chromatographic R<sub>F</sub> Values of Toxicologically Relevant Substances on Standardized Systems*. Weinheim: VCH.
- Kunjappu JT (2001) *Essays in Ink Chemistry (For Paints and Coatings Too)*. Huntingdon, NY: Nova Science Publishers.
- Laing DK, Boughey L, and Hartshorne AW (1991) The standardisation of thin-layer chromatographic systems for comparison of fibre dyes. *Journal of the Forensic Science Society* 30: 299.
- Venkataraman K (1977) *Analytical Chemistry of Synthetic Dyes*. New York: Wiley.

## Volatile Substances

**R J Flanagan**, Guy's and St. Thomas' Hospital Trust, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Acute poisoning with volatile substances usually follows the deliberate inhalation of vapor in order to become intoxicated ['glue sniffing', inhalant abuse, solvent abuse, volatile substance abuse (VSA)]. Solvents from adhesives, notably toluene, some correcting fluids and thinners, hydrocarbons such as those found in cigarette lighter refills [usually liquefied petroleum gas (LPG)], aerosol propellants, and anesthetic gases such as nitrous oxide are amongst the compounds/products that may be abused in this way (Tables 1 and 2). Nail varnish/varnish remover (acetone and esters) and felt-tip marker pen fluids, although strong smelling, are probably too water soluble to be intoxicants. Methanol, ethanol, 2-propanol, and ethylene glycol, and also diesel fuel, aviation fuel (kerosene, Avgas), white spirit, turpentine (or substitute), and paraffin are not sufficiently volatile to be abused by inhalation. Petrol (gasoline), on the other hand, is often abused, especially in less well-off communities. Isobutyl and isopentyl ('amyl') nitrites are also inhaled in order to experience their vasodilator properties, sometimes by male homosexuals. In addition, those who ingest, or even more rarely inject, solvents or solvent-containing products, either accidentally or deliberately, and the victims of clinical, industrial, and domestic accidents may be poisoned by the compounds under consideration. Finally, chloroform, diethyl ether, and other volatiles are still used occasionally during crimes such as rape and murder, whilst a further volatile compound, chlorobutanol (1,1,1-trichloro-2-methyl-2-propanol), sometimes employed as a sedative and a preservative, has been used in doping racing greyhounds.

### Diagnosis of Acute Poisoning with Volatile Substances

VSA should be suspected in children and adolescents with 'drunken' behavior, unexplained listlessness, anorexia, and moodiness. The hair, breath, and clothing may smell of solvent, and empty adhesive tubes or other containers, potato crisp bags, cigarette lighter refills, or aerosol spray cans are often found.

The smell of solvent on the breath is related to the dose and duration of exposure and may last for many hours. The so-called 'glue-sniffer's rash' (perioral eczema) is probably caused by repeated contact with glue in a plastic or other bag held to the face. Although primarily a phenomenon of adolescence, it must be remembered that adults, especially those with ready occupational access to abusable volatiles, may also indulge in VSA. In the late 1970s, for example, it was estimated that some 1–1.6% of US dentists were abusing nitrous oxide.

The analytical toxicology laboratory may be asked to perform analyses for solvents and other volatile compounds in biological samples and related specimens to: (1) assist in the diagnosis of acute poisoning; (2) confirm a suspicion of chronic VSA in the face of denial from the patient and/or a caretaker; (3) aid the investigation of deaths where poisoning by volatile compounds is a possibility, including deaths associated with anesthesia; (4) aid investigation of rape or other assault, or other offences such as driving a motor vehicle or operating machinery, which may have been committed under the influence of volatile substances; (5) aid investigation of incidents such as rape or other assault in which volatile substances may have been administered to the victim; (6) help investigate fire or explosion where VSA might have been a contributory factor; and (7) assess occupational or environmental exposure to anesthetic or solvent vapor. However, other techniques such as ambient air monitoring or, in a few instances, the measurement of urinary metabolite excretion may be more appropriate in this latter context.

The analysis of volatiles presents particular problems. First, many of the compounds of interest occur commonly in laboratories and this necessitates special precautions against contamination and interference. Second, collection, storage, and transport of biological samples must be controlled as far as practicable in order to minimize loss of analyte – quantitative work is futile if very volatile compounds such as propane are encountered unless special precautions are taken to prevent the loss of analyte from the sample prior to the analysis. Third, many compounds of interest are excreted unchanged via the lungs and thus blood (and/or other tissues in fatalities), and not urine, is usually the sample of choice. Finally, the interpretation of results can be difficult, especially if legitimate exposure to solvent vapor is a possibility.

A diagnosis of VSA should be based on a combination of circumstantial, clinical, and analytical

**Table 1** Some volatile substances that may be abused by inhalation

Hydrocarbons	Oxygenated compounds and others
Aliphatic	Acetylene Butane <sup>a</sup> Isobutane (2-methylpropane) <sup>a</sup> Hexane <sup>b</sup> Propane <sup>a</sup>
Alicyclic/ aromatic	Cyclopropane (trimethylene) Toluene (toluol, methylbenzene, phenylmethane) Xylene (xylol, dimethylbenzene) <sup>c</sup>
Mixed	Petrol (gasoline) <sup>d</sup> Petroleum ethers <sup>e</sup>
Halogenated	Bromochlorodifluoromethane (BCF, FC 12B1) Carbon tetrachloride (tetrachloromethane) Chlorodifluoromethane (FC 22, Freon 22) Chloroform (trichloromethane) Dichlorodifluoromethane (FC 12, Freon 12) 1,1-Dichloro-1-fluoroethane (FC 141b, Genetron 141b) Dichloromethane (methylene chloride) 1,2-Dichloropropane (propylene dichloride) 1,1-Difluoroethane (FC 152a) Difluoromethane (FC 32) Ethyl chloride (monochloroethane) Halothane [( <i>R,S</i> )-2-bromo-2-chloro-1,1,1-trifluoroethane] Pentafluoroethane (FC 125) Perfluoropropane (octafluoropropane, FC 218) Tetrachloroethylene (perchloroethylene) 1,1,1,2-Tetrafluoroethane (FC 134a) 1,1,1-Trichloroethane (methylchloroform, Genklene) 1,1,1-Trifluoroethane (FC 143a) 1,1,2-Trichlorotrifluoroethane (FC 113) Trichloroethylene ('trike', Trilene) Trichlorofluoromethane (FC 11, Freon 11) Oxygenated compounds and others Butanone (2-butanone, methyl ethyl ketone, MEK) Butyl nitrite <sup>f</sup> Cyclohexyl nitrite <sup>f</sup> Enflurane [( <i>R,S</i> )-2-chloro-1,1,2-trifluoroethyl difluoromethyl ether] Ethyl acetate Desflurane [( <i>R,S</i> )-difluoromethyl 1,2,2,2-tetrafluoroethyl ether] Diethyl ether (ethoxyethane) Dimethyl ether (DME, methoxymethane) Isobutyl nitrite ('butyl nitrite') <sup>f</sup> Isoflurane [( <i>R,S</i> )-1-chloro-2,2,2-trifluoroethyl difluoromethyl ether] Isopentyl nitrite (3-methylbutan-1-ol, isoamyl nitrite, 'amyl nitrite') <sup>f,g</sup> Methoxyflurane (2,2-dichloro-1,1-difluoroethyl methyl ether) Methyl acetate Methyl isobutyl ketone (MIBK, isopropyl acetone, 4-methyl-2-pentanone) Methyl tert.-butyl ether (MTBE) Nitrous oxide (dinitrogen monoxide, 'laughing gas')

**Table 1** Continued

Sevoflurane [fluoromethyl 2,2,2-trifluoro-1-(trifluoromethyl)ethyl ether]
Xenon

<sup>a</sup>Principal components of purified LPG; some unpurified LPGs can contain up to 40% (v/v) unsaturated compounds (butenes and propenes).

<sup>b</sup>Commercial 'hexane' mixture of hexane and heptane with small amounts of higher aliphatic hydrocarbons.

<sup>c</sup>Mainly *m*-xylene (1,3-dimethylbenzene).

<sup>d</sup>Mixture of aliphatic and aromatic hydrocarbons with boiling range 40–200°C.

<sup>e</sup>Mixtures of pentanes, hexanes, etc., with specified boiling ranges (e.g., 40–60°C).

<sup>f</sup>Abused primarily for its vasodilator properties.

<sup>g</sup>Commercial 'amyl nitrite', mainly isopentyl nitrite but other nitrites also present.

evidence rather than on any one factor alone. It is especially important to consider all circumstantial evidence in cases of possible VSA-related sudden death since suicide or even homicide cannot be excluded simply on the basis of the toxicological examination. There have been a number of reports of the use of inhalational anesthetics for suicidal purposes and there has been one example in the UK of a serial murderer whose victims were thought initially to have died as a result of VSA.

## Analytical Methods

The analysis of biological samples for solvents and other volatiles that may be abused by inhalation has similarities to the analysis of methanol, ethanol, and 2-propanol. However, poisoning with these latter compounds is normally the result of ingestion or occupational exposure to vapor. Headspace gas chromatography (GC) with flame ionization and/or electron capture detection (FID and/or ECD) is widely used in the analysis of volatiles in blood and other biological specimens that may be obtained without using special apparatus such as breath-collection tubes. Nitrous oxide and most halogenated compounds respond on the ECD, although the thermal conductivity detector (TCD) may be used as an alternative if nitrous oxide poisoning is suspected. Although not reported as yet, abuse of xenon is a possibility. GC methods for this element would require a TCD or mass spectrometry (MS). Direct MS of expired air can also detect many compounds several days postexposure. However, at present the use of this technique is limited by the need to take breath directly from the patient. Vapor-phase infrared (IR) spectrophotometry may be useful in the analysis of abused products or ambient atmospheres.

**Table 2** Some products that may be abused by inhalation<sup>a</sup>

<i>Product</i>	<i>Major volatile components</i>
Adhesives	
Balsa wood cement	Ethyl acetate
Contact adhesives	Butanone, hexane, toluene, and esters
Cycle tyre repair cement	Toluene and xylenes
Poly(vinyl chloride) cement	Acetone, butanone, cyclohexanone, trichloroethylene
Woodworking adhesives	Xylenes
Aerosols	
Air freshener	LPG, DME, and/or fluorocarbons <sup>b</sup>
Deodorants, antiperspirants	LPG, DME, and/or fluorocarbons <sup>b</sup>
Fly spray	LPG, DME, and/or fluorocarbons <sup>b</sup>
Hair lacquer	LPG, DME, and/or fluorocarbons <sup>b</sup>
Paint	LPG, DME, and/or fluorocarbons <sup>b</sup> and esters
Anesthetics/analgesics	
Inhalational	Nitrous oxide, cyclopropane, diethyl ether, halothane, enflurane, desflurane, isoflurane, methoxyflurane, sevoflurane, xenon
Topical	Ethyl chloride, fluorocarbons <sup>b</sup>
Dust removers ('air brushes')	DME, fluorocarbons <sup>b</sup>
Commercial dry cleaning and degreasing agents	Dichloromethane, FC 113, FC 141b, methanol, 1,1,1-trichloroethane, tetrachloroethylene, toluene, trichloroethylene (now very rarely carbon tetrachloride, 1,2-dichloropropane)
Domestic spot removers and dry cleaners	Dichloromethane, 1,1,1-trichloroethane, tetrachloroethylene, trichloroethylene
Fuel gases	
Cigarette lighter refills	LPG
'Butane'	LPG
'Propane'	Propane and butanes
Halocarbon fire extinguishers	BCF, FC 11, FC 12
Paints/paint thinners	Acetone, butanone, esters, hexane, toluene, trichloroethylene, xylenes
Paint stripper	Dichloromethane, methanol, toluene
Racing fuel super-charge	Nitrous oxide
'Room odorizer'	Cyclohexyl nitrite, isobutyl nitrite
Surgical plaster/chewing gum remover	1,1,1-Trichloroethane, trichloroethylene
Typewriter correction fluids/thinners (some)	1,1,1-Trichloroethane
Whipped cream dispensers	Nitrous oxide

<sup>a</sup> see **Table 1** for full chemical names of some compounds; the composition of some products varies with time and country of origin.

<sup>b</sup> Nowadays often 1,1,1,2-tetrafluoroethane (FC 134a), but chlorodifluoromethane (FC 22), 1,1-difluoroethane (FC 152a), difluoromethane (FC 32), pentafluoroethane (FC 125), perfluoropropane (FC 218), and 1,1,1-trifluoroethane (FC 143a) might also be encountered.

If the analyte is very volatile (e.g., propane or butane) and a quantitative analysis is required, a blood sample should be collected directly into the headspace vial in which the analysis will be carried out. Many other volatile compounds are relatively stable in blood and other tissues if simple precautions are taken. In the case of blood, the container used for the sample should be glass, preferably with a cap lined with metal foil; greater losses may occur if plastic containers are used. The tube should be as full as possible and should only be opened when required for analysis and then only when cold (4°C). If the sample volume is limited, it is advisable to select the container to match the volume of blood so that there is minimal headspace. An anticoagulant [sodium ethylenediaminetetraacetate (EDTA) or lithium heparin] should be used. Specimen storage between -5 and 4°C is recommended and 1% (w/v)

sodium fluoride should be added to minimize esterase and other enzymic activity. Tissues (~10 g each of brain, lung, liver, kidney, and subcutaneous fat) should also be obtained if a necropsy is to be performed in addition to standard toxicological specimens if available (femoral blood, urine, stomach contents, vitreous humor). Tissues should be stored before analysis in the same way as blood. No preservative should be added. Products thought to have been abused or otherwise implicated in the incident (and stomach contents if ingestion is suspected) should be packed, transported, and stored entirely separately from (other) biological specimens to avoid cross-contamination. Investigation of deaths occurring during or shortly after anesthesia should include the analysis of the inhalation anesthetic(s) used in order to exclude an administration error.



## Headspace Gas Chromatography

Packed columns, for example, 2 m  $\times$  2 mm i.d. 0.3% (w/w) Carbowax 20M on Carbowax C programmed from 35°C to 175°C, have been used extensively in conjunction with headspace sample preparation for the GC analysis of volatile compounds in biological specimens. On-column septum injections of up to 400  $\mu$ l headspace can be performed and good sensitivity (of the order of 0.1 mg l<sup>-1</sup> or better using 200  $\mu$ l of sample) can be obtained. Disadvantages include the poor resolution of some very volatile substances, a long total analysis time, and variation in the peakshape given by alcohols between different batches of column packing. Porous layer open tubular (PLOT) columns give good retention and thus resolution of compounds with similar relative formula mass, but peakshapes of polar compounds are poor and it is difficult to screen for compounds of widely different volatility in one analysis.

Bonded-phase wide-bore capillary columns permit relatively large-volume septum injections and can offer advantages of improved efficiency, reproducibility, and reliability. A 60 m  $\times$  0.53 mm i.d. fused silica capillary coated with the dimethylpolysiloxane SPB-1 (5  $\mu$ m film thickness) programmed from 40°C to 200°C offers many advantages over packed column and PLOT systems. Improved resolution of very volatile compounds is obtained and, even with an initial temperature of 40°C, the total analysis time can be reduced to 26 min, and good peakshapes can be obtained even for alcohols (Figure 1). Moreover, splitless septum injections of up to 300  $\mu$ l headspace can be performed with no noticeable effect on column efficiency; hence, sensitivity is at least as good as that attainable with a packed column. Indeed, nowadays packed and PLOT columns are only used in specific applications or if a capillary GC is not available.

The use of a capillary column together with two different detectors (FID and ECD) confers a high degree of selectivity, particularly for low formula mass compounds where there are very few alternative structures. If more rigorous identification is required, GC combined with MS or Fourier transform IR (FTIR) spectrometry may be used. However, GC-MS cannot be efficient when the fragments produced are less than  $m/z$  40, particularly if the instrument is used for other purposes as well as solvent analyses. In particular, the available sensitivity and spectra of the low molecular weight alkanes renders them very difficult to confirm by GC-MS. Inertial spray MS allows introduction of biological fluids directly into the MS without prior chromatographic

analysis and has been used in the analysis of halothane in blood during anesthesia.

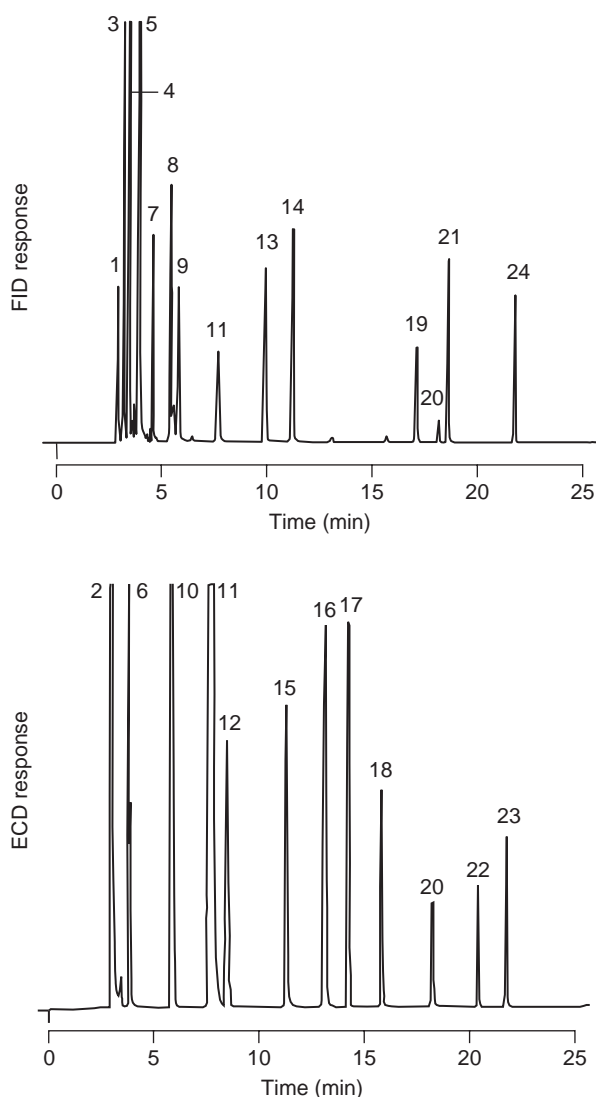
GC-FTIR is generally more appropriate than GC-MS in the analysis of volatiles, but sensitivity is poor particularly when compared with ECD. In addition, interference, particularly from water and carbon dioxide in the case of biological specimens, can be troublesome. 'Purge and trap' and multiple headspace extraction offer ways of increasing sensitivity and, although not needed for most clinical and forensic applications, have been used either in conjunction with GC-FTIR or in occupational/environmental monitoring. Pulse heating has also been employed in the analysis of volatiles in biological specimens. Advantages of this latter technique include the use of a small sample volume (0.5–5  $\mu$ l), short extraction time, and lack of matrix effects.

**Use of the SPB-1 column system** A qualitative standard mixture is first analyzed (see Figure 1). A reagent 'blank' analysis is then performed to monitor interference from the laboratory atmosphere. A sealed headspace vial containing internal standard solution (1,1,2-trichloroethane + ethylbenzene in analyte-free diluted whole blood) is incubated (65°C, 15 min) and a portion (100–300  $\mu$ l) of the headspace is taken for analysis using a warmed (40°C) gas-tight glass syringe. Subsequently, the sample (whole blood, plasma, serum, or urine) (200  $\mu$ l) is added to the same vial and, after reincubation (65°C, 15 min), a further portion of headspace is taken for analysis.

Tissue samples can be analyzed after incubation with a proteolytic enzyme (subtilisin A) in a sealed headspace vial. The 'blank' analysis should be performed in a separate vial. Adhesives and similar products should be introduced into a glass vial. The vial is sealed and, after 1–15 min, a portion (50–100  $\mu$ l) of the headspace is analyzed. Liquids can be analyzed in the same way, except that it is often possible to withdraw a portion (5–50  $\mu$ l) of the headspace directly from the container. Aerosols and fuel gases can be analyzed after releasing a portion of the product into a headspace vial and sealing the vial.

Quantitative analyses can be performed either isothermally or with a temperature program. For analytes that are liquids at room temperature, assay calibration is by analysis of standard solutions prepared in analyte-free human blood. The same calibration solutions are used in the analysis of tissue digests. Analyte concentrations in the range 0.1–10 or 0.5–50 mg l<sup>-1</sup> are usually adequate in acute poisoning. Portions of the standards are transferred to headspace vials for analysis as described above, and a





**Figure 1** Analysis of a standard volatile mixture. Column: 80 m  $\times$  0.53 mm i.d. SPB-1 (5  $\mu$ m film). Carrier gas: helium (flow rate 8.6 ml min<sup>-1</sup>). Oven temperature: 40°C (6 min), then to 80°C at 5°C min<sup>-1</sup>, then to 200°C at 10°C min<sup>-1</sup> (run time 26 min). Injection: 10  $\mu$ l vapor. Detector sensitivities (fsd): FID, 3.2 nA; ECD, 64 kHz. Peaks: 1, propane; 2, FC 12; 3, dimethyl ether; 4, isobutane; 5, butane; 6, BCF; 7, ethanol; 8, acetone; 9, 2-propanol; 10, FC 11; 11, FC 113; 12, halothane; 13, butanone; 14, hexane; 15, chloroform; 16, 1,1,1-trichloroethane; 17, carbon tetrachloride; 18, trichloroethylene; 19, methyl isobutyl ketone; 20, 1,1,2-trichloroethane (internal standard); 21, toluene; 22, tetrachloroethylene; 23, 2,2,2-trichloroethanol; 24, ethyl benzene (internal standard). (Reproduced with permission from Streete PJ, Ruprah M, Ramsey JD, and Flanagan RJ (1992) Detection and identification of volatile substances by head-space capillary gas chromatography to aid the diagnosis of acute poisoning. *Analyst* 117: 1111–1127. The Royal Society of Chemistry.)

calibration graph of peak height ratio of analyte to internal standard against analyte concentration is prepared. Often, either 1,1,2-trichloroethane or ethylbenzene can be used as the internal standard.

For analytes that are gases at room temperature, assay calibration is much more complicated and involves preparation of dilutions of gaseous calibration mixtures.

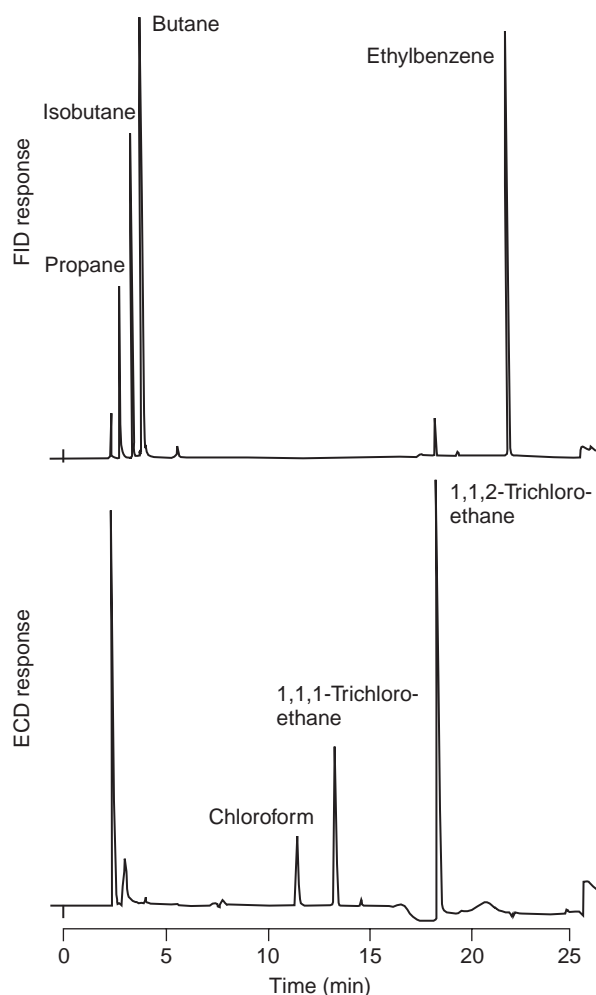
Compounds that do not elute during the program generally have boiling points of 170°C or above and Kovats retention indices on SE-30/OV-1/OV-101 packed columns of 1000 or more. ECD-responding compounds are primarily halogenated substances, but many compounds containing nitro- or keto-moieties also respond. Other substances such as allyl isothiocyanate and nitrous oxide show a good response. In contrast, the ECD response to compounds such as 1,1,1,2-tetrafluoroethane is relatively poor and some ketones, for example, the heptanones, show no response.

The retention data are highly reproducible in routine use and should be applicable to other dimethylpolysiloxane-coated capillaries (OV-1, SE-30, etc.) of similar dimensions and film thickness. Retention times are more convenient than retention indices when identifying unknowns. The analysis of a blood specimen from an adolescent who died after abusing 'butane' gas and vapor from a typewriter correcting fluid is shown in **Figure 2**. The analysis of blood from a patient who died after inhaling vapor from an electrical component cleaner is illustrated in **Figure 3**. Note that the concentrations of the components of interest were well above the limit of detection of the system. Indeed, although no formal studies have been performed, the limit attainable appears to be of the order of 0.01 mg l<sup>-1</sup> for ECD-responding compounds and 0.1 mg l<sup>-1</sup> for the remainder.

## Pharmacokinetics and the Interpretation of Results

Knowledge of the pharmacokinetics of volatile compounds is important in understanding the rate of onset, the intensity, and the duration of intoxication with these substances. The UK Maximum Exposure Limit (MEL) or Occupational Exposure Standard (OES) (**Table 3**) provide information on the relative toxicities of different compounds after chronic exposure to relatively low concentrations of vapor.

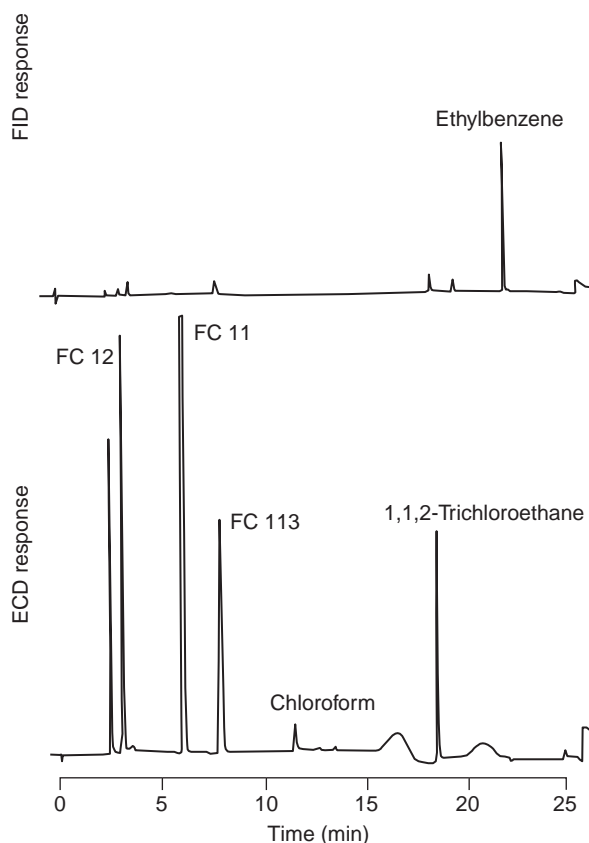
Inhaled compounds may rapidly attain high concentrations in well-perfused organs (brain, heart), while concentrations in muscle and adipose tissue may be very low. Should death occur, this situation is 'frozen' to an extent, but if exposure continues the compound will accumulate in less accessible (poorly perfused) tissues, only to be slowly released once exposure ceases. Thus, the plasma concentrations of



**Figure 2** Analysis of a whole blood specimen (200  $\mu$ l) from a patient who died after abusing a cigarette lighter refill and a typewriter correcting fluid containing 1,1,1-trichloroethane. GC conditions as for **Figure 1**. Injection: 300  $\mu$ l headspace. Detector sensitivities (fsd): FID 80 pA, ECD 2 kHz. Whole blood 1,1,1-trichloroethane concentration 1.2 mg l<sup>-1</sup>. (Reproduced with permission from Streete PJ, Ruprah M, Ramsey JD, and Flanagan RJ (1992) Detection and identification of volatile substances by head-space capillary gas chromatography to aid the diagnosis of acute poisoning. *Analyst* 117: 1111–1127. The Royal Society of Chemistry.)

some compounds may fall monoexponentially, while others may exhibit two (or more) separate rates of decline (half-lives).

The solubility of a volatile compound in blood is an important influence on the rate of absorption, tissue distribution, and elimination of the compound. The partition coefficients of a number of compounds between air, blood, and various tissues have been measured *in vitro* using animal tissues, and some *in vivo* distribution data have been obtained from postmortem tissue measurements in humans (**Table 3**). However, these latter data must be used with caution



**Figure 3** Analysis of a whole blood specimen (200  $\mu$ l) from a patient who died after abusing an aerosol designed for cleaning electrical components that contained FCs 11, 12, and 113. GC conditions as for **Figure 1**. Injection: 150  $\mu$ l headspace. Detector sensitivities (fsd): FID 160 pA, ECD 4 kHz. Whole blood FC 11 and FC 113 concentrations 3.2 and 1.0 mg l<sup>-1</sup>, respectively. (Reproduced with permission from Streete PJ, Ruprah M, Ramsey JD, and Flanagan RJ (1992) Detection and identification of volatile substances by head-space capillary gas chromatography to aid the diagnosis of acute poisoning. *Analyst* 117: 1111–1127. The Royal Society of Chemistry.)

since there are many difficulties inherent in such measurements (sampling variations, analyte stability, external calibration, etc.). Published data on the elimination half-lives of volatile substances (**Table 3**) are not easily comparable, either because too few samples were taken or the analytical methods used did not have sufficient sensitivity to measure the final half-life accurately.

### Metabolism of Volatile Substances

Exogenous compounds may be metabolized in a number of ways, a frequent result being the production of metabolites of greater polarity (water solubility) and thus lower volatility than the parent compound. The pharmacological activity and

**Table 3** Physical properties and pharmacokinetic data of some volatile compounds

Compound	MEL/OES <sup>a</sup> (mg m <sup>-3</sup> )	Vapor pressure <sup>b</sup> (20°C, mmHg)	Inhaled dose absorbed (%)	Proportion absorbed dose (%)		Half-life <sup>c</sup> (h)	Brain–blood distribution ratio (deaths)	Partition coefficient (blood–gas) (37°C)
				Eliminated unchanged	Metabolized			
Acetone	1210	183	–	–	–	3–5 <sup>d</sup>	–	243–300
Benzene	16	75	46	12	80	9–24	3–6	6–9
Butane	1450	(1554)	30–45	–	–	–	–	–
Isobutane	1750 <sup>e</sup>	(2282)	–	–	–	–	–	–
Butanone	600	75	70	99 +	0.1	0.5	–	116
Carbon disulfide	32	294	40	<30	50–90	<1	–	2.4
Carbon tetrachloride	13	90	–	50?	50?	48	–	1.6
Chlorodifluoromethane	3590	(6701)	–	–	–	–	1.9	–
Chloroform	9.9	157	–	20–70 (8 h)	>30	–	4	8
Cyclopropane	–	(4701)	–	99	0.5	–	1.5–3.6	0.55
Desflurane	–	669	–	–	0.02	–	1.29 <sup>f</sup>	0.42
Dichlorodifluoromethane	5030	(3639)	35	99	<0.2	–	1.4	0.15
Dichloromethane	350	350	–	50?	<40	0.7	0.5–1	5–10
Diethyl ether	310	438	–	>90	–	–	1.1	12
Enflurane	383	172	90 +	>80 (5 d)	2.5	36	1.4 <sup>f</sup>	1.9
Ethyl acetate	1460	72	–	–	–	–	–	–
Halothane	82	244	90 +	60–80 (24 h)	<20	2–3	2–3	2.57
Hexane	72	122	–	–	–	–	–	–
Isoflurane	383	240	–	–	0.2	–	1.57 <sup>f</sup>	1.38
Methoxyflurane	–	23	–	19 (10 d)	>44	–	2–3	13
Methyl isobutyl ketone	208	15	–	–	–	–	–	–
Nitrous oxide	183	(39 800)	–	>99	–	–	1.1	0.47
Propane	1750 <sup>e</sup>	(6269)	–	–	–	–	–	–
Sevoflurane	–	157	–	–	3	20	1.7 <sup>f</sup>	0.68
Styrene	430	4	–	1–2	>95	13	–	32
Tetrachloroethylene	345	14	60 +	>90	1–2	72	9–15	9–19
Toluene	191	22	53	<20	80	7.5	1–2	8–16
1,1,1-Trichloroethane	555	98	–	60–80 (1 w)	2	10–12	2	1–3
Trichloroethylene	550	58	50–65	16	>80	30–38	2	9.0
Trichlorofluoromethane	5710	667	92	89	<0.2	1.5	2.5	0.87
‘Xylene’	220	6	64	5	>90	20–30	–	42.1

<sup>a</sup>UK MEL/OES (8 h time weighted average) (Health and Safety Executive. EH40/2002. *Occupational Exposure Limits 2002*. Norwich, HMSO).<sup>b</sup>Figures in parentheses indicate compound gas at 20°C.<sup>c</sup>Terminal phase plasma half-life.<sup>d</sup>Longer after high doses.<sup>e</sup>As components of LPG.<sup>f</sup>Experimental: 37°C.

Data summarized from Baselt (2002), Fiserova-Bergerova (1983) and Pihlainen and Ojanperä (1998).

pharmacokinetics of any metabolite(s) often differ from those of the parent compound(s). After ingestion, extensive hepatic metabolism can reduce systemic availability ('first-pass' metabolism). Many volatile substances, including butane, dimethyl ether, most fluorocarbon refrigerants/aerosol propellants, isobutane, nitrous oxide, propane, tetrachloroethylene, and 1,1,1-trichloroethane, are largely eliminated unchanged in exhaled air. Others are partly eliminated in exhaled air and also metabolized in the liver and elsewhere, the metabolites being eliminated in exhaled air or in urine (Table 4), or incorporated into intermediary metabolism.

### Interpretation of Qualitative Results

The likelihood of detecting exposure to volatile substances by headspace GC of blood is influenced by the dose and duration of exposure, the time of sampling in relation to the time elapsed since exposure, and the precautions taken in the collection and storage of the specimen. In a suspected VSA- or anesthetic-related fatality, analysis of tissues (especially fatty tissues such as brain) may prove useful since high concentrations of volatile compounds may be present even if very little is detectable in blood.

Analysis for metabolites in urine may extend the time in which exposure may be detected but, of the compounds commonly abused, only toluene, the xylenes, and some chlorinated solvents, notably trichloroethylene, have suitable metabolites (see Table 4). Chronic petrol sniffing has been diagnosed by the measurement of blood lead concentrations, although with the move to lead-free petrol in many countries this is likely to be less useful in future. Petrol sniffing has also been diagnosed by detection of aromatic components such as toluene and ethylbenzene in blood. One consequence of the move to lead-free petrol has been that the aromatic content (including benzene content) of some kinds of petrol has increased – ironically benzene might present a greater danger in 'sniffers' than did lead. Abuse of the fluorinated anesthetic methoxyflurane has been detected by measuring serum and urine fluoride ion concentrations. With some kinds of petrol and other complex mixtures such as petroleum ethers (Table 1), however, the blood concentrations of the individual components are often below the limit of detection of headspace GC methods even after significant inhalational exposure.

Detection of a volatile compound in blood does not always indicate VSA or occupational/environmental exposure to solvent vapor. Acetone and some of its homologs may occur in high concentrations in ketotic patients. Large amounts of acetone and

butanone may also occur in blood and urine from children with acetoacetylcoenzyme A thiolase deficiency, for example, and may indicate the diagnosis. In addition, acetone is the major metabolite of 2-propanol in humans (Table 4). Conversely, 2-propanol has been found in blood from ketotic patients. Other ketones may also give rise to alcohols *in vivo*. Cyclohexanol, for example, is the principal metabolite of cyclohexanone in man (Table 4). Other volatile compounds such as halothane or chlorobutanol may be used in therapy or inadvertently added to the sample as a preservative. When interpreting the results of qualitative analyses it is important to remember that some compounds often occur in association – one with another (Table 5).

The use of aerosol disinfectant preparations when collecting specimens may contaminate the sample if an aerosol propellant is used. Contamination of blood samples with ethanol or 2-propanol may also occur if an alcohol-soaked swab is used to cleanse skin prior to venepuncture. Gross contamination with technical xylene (a mixture of *o*-, *m*-, and *p*-xylene together with ethylbenzene) has been found in blood collected into Sarstedt Monovette Serum Gel blood collection tubes; contamination with toluene (up to  $22 \text{ mg l}^{-1}$ ), 1-butanol, ethylbenzene, and xylene has been found in batches of these same tubes. Contamination with 1-butanol or 2-methyl-2-propanol occurs commonly in blood collected into tubes coated with EDTA. Care should be taken when handling frozen tissue prior to analysis as any compounds present in ambient air may condense on the cold surface and give rise to false positives. Processing blank frozen tissue can control for this possibility.

The interpretation of case data involving chloroform is particularly difficult, especially since this compound is still sometimes used in the course of crimes such as rape and murder. In addition to sometimes being present in drinking water at low concentrations, chloroform is found in a variety of medicinal preparations, in cigarette smoke, soft drinks, margarines, and in swimming pools if a chlorination plant is in operation. A further possible source of chloroform on headspace GC is from thermal decomposition of trichloroacetic acid. Trichloroacetic acid is a metabolite of several compounds including the solvent trichloroethylene (Table 5) and the drugs chloral hydrate, dichloralphenazone, and triclofos. Trichloroacetic acid has a half-life in blood of 3–5 days and thus may be detected for a relatively long time after exposure to, or ingestion of, a precursor. Trichloroacetic acid plasma concentrations of up to  $40 \text{ mg l}^{-1}$  have been reported after occupational exposure to trichloroethylene vapor.

**Table 4** Metabolites of some solvents and other volatile substances that may be measured to assess exposure<sup>a</sup>

<i>Compound</i>	<i>FW<sub>t</sub></i>	<i>Parent compound</i>	<i>Fluid</i>	<i>Normal<sup>b</sup></i>	<i>High<sup>c</sup></i>	<i>Comment</i>
Acetaldehyde	44.1	Ethanol	Blood	0.2 mg l <sup>-1</sup>	Not known	Blood/urine acetone concentrations can rise to 2 g l <sup>-1</sup> in ketosis. 2-Propanol also acetone metabolite
Acetone	58.1	2-Propanol	Blood	10 mg l <sup>-1</sup>	–	
			Urine	10 mg l <sup>-1</sup>	80 mg l <sup>-1</sup>	
4-Aminophenol	109.1	Aniline	Urine	–	10 mg l <sup>-1</sup>	The bromide concentrations associated with toxicity are lower after exposure to organobromines than when bromide salts given orally
Bromide ion	79.9	Bromomethane, other organobromines	Serum	10 mg l <sup>-1</sup>	40 mg l <sup>-1</sup>	
Butoxyacetate	132.2	2-Butoxyethanol	Urine	–	100 mg l <sup>-1</sup>	CO blood half-life 13 h breathing air, atmospheric pressure (CO half-life 5 h after inhalation of CO). Blood HbCO useful indicator of chronic exposure
Carbon monoxide	28.0	Dichloromethane	Blood	<5% HbCO	>20% HbCO	
Cyanide ion	26.0	Acetonitrile, acrylonitrile, other organonitriles	Blood	0.2 mg l <sup>-1</sup> (nonsmokers)	2 mg l <sup>-1</sup>	Cyanide metabolized to thiocyanate; both compounds may accumulate during chronic exposure
Cyclohexanol	100.2	Cyclohexane, cyclohexanone	Urine	–	5 mg l <sup>-1</sup>	Additional cyclohexane metabolites: <i>trans</i> -1,2-cyclohexanediol, <i>trans</i> -1,4-cyclohexanediol (alcohol metabolites excreted mainly as glucuronides in adults)
Cyclohexanone	98.2	Cyclohexane, cyclohexanol	Urine	–	0.5 mg l <sup>-1</sup>	See entry for cyclohexanol above
2,5-Dichlorophenol	163.0	1,4-Dichlorobenzene	Urine	–	100 mg l <sup>-1</sup>	
Dimethylsulfone	94.1	Dimethylsulfoxide	Urine	–	Not known	0.8 g l <sup>-1</sup> legal UK driving limit
Ethanol	46.1	Ethyl acetate	Blood	0.1 g l <sup>-1</sup>	0.8 g l <sup>-1</sup>	
Ethoxyacetate	104.1	Ethoxyethanol, 2-ethoxyethyl acetate	Urine	–	50 mg l <sup>-1</sup>	
β-Hydroxyethoxyacetate	120.1	1,4-Dioxane	Urine	–	0.5 g l <sup>-1</sup>	Dioxane-2-one also found in urine
Fluoride	19.0	Methoxyflurane and some other organofluorines	Serum	0.2 mg l <sup>-1</sup>	2 mg l <sup>-1</sup>	
			Urine	1 mg l <sup>-1</sup>	2 mg l <sup>-1</sup>	
Formate	46.0	Formaldehyde, methanol, methyl formate	Urine	20 mg l <sup>-1</sup>	30 mg l <sup>-1</sup>	
2,5-Hexanedione + 4,5-dihydroxy-2-hexanone	–	Hexane, 2-hexanone (MBK)	Urine	–	5 mg l <sup>-1</sup>	2-Hexanol and 2-hexanone additional hexane metabolites
Hippurate	179.2	Toluene	Urine	0.2 g l <sup>-1</sup>	2 g l <sup>-1</sup>	Not ideal indicator of toluene exposure as there are other (dietary, pharmaceuticals) sources of benzyl alcohol/benzoate hence hippurate
Mandelate	152.1	Ethylbenzene, styrene	Urine	0.005 g l <sup>-1</sup>	2 g l <sup>-1</sup>	Phenylglyoxylic acid also urinary metabolite
Methanol	32.0	Methyl acetate, methyl formate	Urine	–	30 mg l <sup>-1</sup>	
N-Methylacetamide	73.1	N,N-Dimethylacetamide	Urine	–	65 mg l <sup>-1</sup>	

<i>N</i> -Methylformamide	59.1	<i>N,N</i> -Dimethylformamide	Urine	–	15 mg l <sup>-1</sup>	Hydroxymethylformamide and formamide additional metabolites
4-Methyl-2-pentanone	100.2	Methyl isobutyl ketone	Urine	–	3.5 mg l <sup>-1</sup>	Hippurate and other methylphenols additional toluene metabolites
2-Methylphenol ( <i>o</i> -cresol)	108.1	Toluene	Urine	–	3 mg l <sup>-1</sup>	
<i>Trans,trans</i> -Muconate ( <i>t,t</i> -MA)	142.1	Benzene	Urine	–	2 mg l <sup>-1</sup>	Phenol and <i>S</i> -phenylmercapturic acid additional benzene metabolites
Nitrite ion	46.0	Butyl nitrite, isopentyl nitrite, other organonitrites	Plasma	2.5 mg l <sup>-1</sup>	Not known	Glycolate and glyoxylate also plasma and urinary ethylene glycol metabolites
			Urine	–	10 mg l <sup>-1</sup>	
Oxalate	90.0	Ethylene glycol	Urine	2.5 mg l <sup>-1</sup>	4 mg l <sup>-1</sup>	
Phenol	94.1	Benzene	Urine	0.1 mg l <sup>-1</sup>	50 mg l <sup>-1</sup>	Excreted as sulfate and glucuronide conjugates. Other metabolites include catechol and hydroquinone. Urinary phenol excretion not a reliable indicator of benzene exposure. <i>trans,trans</i> -MA and <i>S</i> -phenyl mercapturic acid excretion have been used to assess exposure
<i>S</i> -Phenyl-mercapturate	239.2	Benzene	Urine	–	45 µg l <sup>-1</sup>	<i>trans, trans</i> -MA and phenol additional benzene metabolites
2-Propanol	60.1	Acetone	Blood	–	2.5 g l <sup>-1</sup>	Acetone also 2-propanol metabolite
Thiocyanate ion	58.1	Acetonitrile, acrylonitrile, other organonitriles	Plasma	4 mg l <sup>-1</sup> (non smokers) 20 mg l <sup>-1</sup> (heavy smokers)	120 mg l <sup>-1</sup>	Cyanide/thiocyanate may accumulate during chronic exposure.
2-Thiothiazolidine-4-carboxylate (TTCA)	163.2	Carbon disulfide	Urine	–	8 mg l <sup>-1</sup>	TTCA glutathione conjugate of carbon disulfide. Urinary TTCA excretion reliable indicator of exposure
Tolurates (methylhippurates)	193.2	Xylenes	Urine	0.01 mg l <sup>-1</sup>	1.5 g l <sup>-1</sup>	Metabolite of 2,2,2-trichloroethanol Also metabolite of chloral hydrate, dichloralphenazone, and triclofos
Trifluoroacetate	114.0	Halothane and some other fluorinated anesthetics	Urine	–	2.5 mg l <sup>-1</sup>	
Trichloroacetate	163.4	Trichloroethylene	Urine	–	100 mg l <sup>-1</sup>	
2,2,2-Trichloroethanol	149.4	Trichloroethylene	Plasma	10 mg l <sup>-1</sup>	50 mg l <sup>-1</sup>	

<sup>a</sup>Urinary excretion often expressed as a ratio to creatinine.

<sup>b</sup>Upper limit of normally expected or 'nontoxic' concentration.

<sup>c</sup>Lower limit of concentration associated with toxicity/occupational exposure action limit.



**Table 5** Associated volatile compounds<sup>b</sup>

Compound	Associated compound(s)
Acetone	Butanone and higher ketones in ketoacidosis, 2-propanol (metabolite <sup>a</sup> )
BCF	FC 11
Butane	Butanone (metabolite <sup>a</sup> ), isobutane, 2-butanol (metabolite <sup>a</sup> ), propane
Cyclohexanone	Cyclohexanol (metabolite)
Dimethyl ether	FC 22 or other fluorocarbon
Ethanol	Propanols and higher alcohols if bacterial fermentation has occurred; methanol or other volatile poisons if denatured/synthetic alcohol has been consumed
Ethyl acetate <sup>a</sup>	Ethanol (metabolite)
Ethylbenzene	See xylenes below
FC 11	BCF, FC 12
FC 12	FC 11
FC 22	Dimethyl ether
Halothane	2-Chloro-1,1-difluoroethylene, 2-chloro-1,1,1-trifluoroethane (metabolites <sup>a</sup> )
Isobutane	Butane, 2-methyl-2-propanol (metabolite <sup>a</sup> ), propane
Isobutyl nitrite <sup>a</sup>	2-Methyl-1-propanol (degradation product)
Isopentyl nitrite <sup>a</sup>	3-Methyl-1-butanol (degradation product)
Methyl acetate <sup>a</sup>	Methanol (metabolite)
Propane	Butane, isobutane, 2-propanol (metabolite <sup>a</sup> )
2-Propanol	Acetone (metabolite)
2,2,2-Trichloroethanol	Trichloroethylene (also metabolite of chloral hydrate, dichloralphenazone, and triclofos)
Trichloroethylene	2,2,2-Trichloroethanol (metabolite), chloroform [possibly from thermal degradation of trichloroacetic acid (metabolite) <i>in vitro</i> ]
Xylenes	<i>ortho</i> -, <i>meta</i> -, and <i>para</i> -Xylene occur together in technical xylene, <i>m</i> -xylene predominating. Ethylbenzene also contaminant in technical xylene

<sup>a</sup> Rarely found.<sup>b</sup> See **Table 1** for full chemical names of certain compounds.

In 25 Caucasian adult women in Florida, USA, over a period of 6 months average plasma chloroform concentrations were generally less than  $25 \mu\text{g l}^{-1}$ , but in two subjects plasma chloroform concentrations of  $2.9 \text{ mg l}^{-1}$  and  $4.0 \text{ mg l}^{-1}$ , respectively, were found during routine sampling. All subjects were carefully screened to exclude occupational and recreational exposure to chloroform and other compounds that could give rise to chloroform on headspace GC. At the other extreme, postmortem blood chloroform concentrations in fatalities involving this agent have been reported as  $10\text{--}50 \text{ mg l}^{-1}$ .

It is well known that ethanol may be both produced and metabolized by microbial action in biological specimens. Small amounts of hexanal may arise from degradation of fatty acids in blood on long-term storage even at  $-5$  to  $-20^\circ\text{C}$ . Hexanal is resolved from toluene on the SPB-1 capillary GC system discussed above, but resolution may be lost if an isothermal quantitative analysis is performed. Interference from hexanal is only likely to be important, however, if very low concentrations of toluene ( $0.1 \text{ mg l}^{-1}$  or less) are to be measured.

In some deaths attributed to the abuse of LPG, only butane, isobutane, and propane are detected on headspace GC of postmortem samples. In other cas-

es, these three compounds are present, but in addition 2-propanol, acetone, 2-methyl-2-propanol, 2-butanol, and/or butanone are present (**Table 5**). By analogy with the metabolism of hexane (**Table 4**), these latter compounds probably arise from the metabolism of the butanes and propane.

The alkyl nitrites that can be abused by inhalation (cyclohexyl nitrite, isobutyl nitrite, isopentyl nitrite) are a special case in that (1) they are extremely unstable and break down rapidly *in vivo* to the corresponding alcohols and (2) usually also contain other isomers (butyl nitrite, pentyl nitrite). Any products submitted for analysis will usually contain the corresponding alcohols as well as the nitrites.

### Interpretation of Quantitative Results

Conversion factors for ambient air conditions (ppm to  $\text{mg m}^{-3}$ ) and for concentrations in blood and other fluids (SI mass to molar units) for a number of volatile compounds are given in **Table 6**. There may be a big overlap in the blood concentrations of volatile compounds attained after workplace exposure and as a result of deliberate inhalation of vapor. For example, blood toluene concentrations after exposure to up to 127 ppm toluene (UK OEL at the time was 100 ppm) for 8 h ranged between 0.4 and

**Table 6** Some volatile compounds: conversion factors

Compound	Formula mass	Ambient air concentrations <sup>a</sup>		Blood/urine concentrations <sup>b</sup>	
		1 ppm to mg m <sup>-3</sup>	1 mg m <sup>-3</sup> to ppm	1 mg l <sup>-1</sup> to µmol l <sup>-1</sup>	1 µmol l <sup>-1</sup> to mg l <sup>-1</sup>
Acetone	58.1	2.37	0.422	17.21	0.058
Acetonitrile	41.1	1.68	0.595	24.33	0.041
Benzene	78.1	3.20	0.313	12.80	0.078
Bromochlorodifluoromethane	165.5	6.77	0.147	6.04	0.166
Bromomethane	95.0	3.89	0.257	—	—
Butane	58.1	2.38	0.422	17.21	0.058
Butanone	72.1	2.95	0.339	13.87	0.072
Butyl acetate	116.2	4.74	0.211	—	—
Carbon disulfide	76.1	3.11	0.322	13.14	0.076
Carbon tetrachloride	153.8	6.29	0.159	6.50	0.154
Chlorodifluoromethane	86.5	3.54	0.283	11.56	0.087
Chloroform	119.4	4.89	0.205	8.38	0.119
Cyclohexane	84.2	3.44	0.291	11.88	0.084
Cyclopropane	42.1	1.72	0.582	23.75	0.042
Dichlorodifluoromethane	120.9	4.95	0.202	8.27	0.121
1,2-Dichloroethane	99.0	4.05	0.247	10.10	0.099
Dichloromethane	84.9	3.48	0.288	11.78	0.085
Diethyl ether	74.1	3.03	0.330	13.50	0.074
<i>N,N</i> -Dimethylformamide	73.1	2.99	0.335	13.68	0.073
Dimethylsulfoxide	78.1	3.19	0.313	12.80	0.078
1,4-Dioxane	88.1	3.60	0.278	11.35	0.088
Enflurane	184.5	7.57	0.133	5.42	0.185
Ethanol	46.1	1.88	0.532	21.7	0.046
Ethyl acetate	88.1	3.60	0.278	—	—
Ethylbenzene	106.2	4.34	0.231	9.42	0.106
Fluorotrichloromethane	137.4	5.62	0.178	7.28	0.137
Halothane	197.4	8.08	0.124	5.07	0.197
Hexane	86.2	3.53	0.283	11.60	0.086
Isobutane	58.1	2.37	0.422	17.21	0.058
Isoflurane	184.5	7.55	0.133	5.42	0.185
Methanol	32.0	1.31	0.764	31.25	0.032
Methoxyflurane	165.0	6.75	0.148	6.06	0.165
4-Methyl-2-pentanone	100.2	4.10	0.244	9.98	0.100
Methyl <i>tert</i> .-butyl ether	88.2	3.60	0.278	11.34	0.088
Nitrous oxide	44.0	1.80	0.556	22.73	0.044
Propane	44.1	1.80	0.556	22.68	0.044
2-Propanol	60.1	2.55	0.408	16.67	0.060
Styrene	104.1	4.25	0.235	9.61	0.104
Tetrachloroethylene	165.9	6.79	0.147	6.02	0.166
Tetrahydrofuran	72.1	2.94	0.340	13.87	0.072
Toluene	92.1	3.76	0.266	10.86	0.092
1,1,1-Trichloroethane	133.4	5.46	0.183	7.50	0.133
Trichloroethylene	131.4	5.38	0.186	7.61	0.131
Xylene	106.2	4.35	0.231	9.42	0.106

<sup>a</sup> 25°C, 760 mm Hg (data from Bingham E, Cohns B, and Powell CH (eds.) (2001) *Patty's Toxicology*, 5th edn., vol. 9. New York: Wiley). To convert 1 ppm to mg m<sup>-3</sup> multiply by atomic or formula mass and divide by 24.45; to convert 1 mg m<sup>-3</sup> to ppm multiply by 24.45 and divide by atomic or formula mass.

<sup>b</sup> To convert mg l<sup>-1</sup> to µmol l<sup>-1</sup> divide by atomic or formula mass and multiply by 1000; to convert µmol l<sup>-1</sup> to mg l<sup>-1</sup>, multiply by atomic or formula mass and divide by 1000.

6.7 mg l<sup>-1</sup>. After brief exposure only signs of moderate intoxication (e.g., slurred speech, unsteady movements) have been associated with blood toluene concentrations as high as 30 mg l<sup>-1</sup>. Blood toluene concentrations in samples from 132 patients who were thought to have engaged in VSA ranged from 0.2 to 70 mg l<sup>-1</sup>, and were above 5 mg l<sup>-1</sup> in

22 of 25 deaths. On the other hand, 13 patients with blood toluene concentrations greater than 10 mg l<sup>-1</sup> were either asymptomatic or only mildly intoxicated (headache, nausea, vomiting, and/or drowsiness), although these manifestations of toxicity can lead to 'indirect' acute VSA-related death as discussed above. Apart from individual differences in tolerance

and possible loss of toluene from the sample prior to analysis, sample contamination, etc., the lack of a strong correlation between blood concentrations and clinical features of poisoning is probably due to rapid initial tissue distribution and elimination.

About 80% of a dose of toluene is converted to hippuric acid (Table 4). Similarly, more than 90% of a dose of xylene is metabolized to methylhippuric (toluric) acids. The principal isomer found in urine is 3-methylhippurate, since *m*-xylene is the principal component of technical grade xylene (Table 1). Methylhippurates are not normal urinary constituents, but hippuric acid may arise from the metabolism of benzoates in foods and medicines, and thus caution is needed in the interpretation of results. Hippurate and methylhippurate excretion is often expressed as a ratio to creatinine since this obviates the need for 24 h urine collections. Occupational exposure to toluene can give rise to ratios of up to 1 g of hippurate per gram creatinine or more; in patients suspected of VSA, a ratio of more than 1 g of hippurate per gram creatinine strongly suggests, but does not prove, toluene exposure. Measurement of urinary *o*-cresol has been proposed as an alternative means of monitoring toluene exposure selectively, particularly in occupational circumstances, but the assay procedure is relatively complex and is thus not widely used.

See also: **Forensic Sciences:** Overview.

## Further Reading

Baselt RC (2002) *Disposition of Toxic Drugs and Chemicals in Man*, 6th edn. Foster City, CA: Chemical Toxicology Institute.

Broussard LA (2000) The role of the laboratory in detecting inhalant abuse. *Clinical Laboratory Sciences* 13: 205–209.

Brouette T and Anton R (2001) Clinical review of inhalants. *American Journal on Addictions* 10: 79–94.

Field-Smith ME, Bland JM, Taylor JC, et al. (2002) Trends in deaths associated with abuse of volatile substances 1972–2002. Report 15. London: St George's Hospital Medical School (<http://www.sghms.ac.uk>)

Fiserova-Bergerova V (1983) *Modelling of Inhalation Exposure to Vapours: Uptake, Distribution and Elimination*, vols. I and II. Boca Raton: CRC Press.

Flanagan RJ (2003) Volatile substances. In: Moffatt AC, Osselton D, and Widdop B (eds.) *Clarke's Isolation and Identification of Drugs*, 3rd edn., pp. 227–238. London: Pharmaceutical Press.

Flanagan RJ, Streete PJ, and Ramsey JD (1997) *Volatile Substance Abuse – Practical Guidelines for Analytical Investigation of Suspected Cases and Interpretation of Results*. UNDCP Technical Series No. 5. Vienna: UN International Drug Control Programme (<http://www.undcp.org>).

Haverkos HW and Dougherty JA (eds.) (1988) *Health Hazards of Nitrite Inhalants*. NIDA Research Monograph 83. Rockville: National Institute on Drug Abuse.

Pihlainen K and Ojanperä I (1998) Analytical toxicology of fluorinated inhalation anaesthetics. *Forensic Science International* 97: 117–133.

Watanabe-Suzuki K, Nozawa H, Suzuki O, and Ishii A (2003) Sensitive analysis of alkyl alcohols as decomposition products of alkyl nitrites in human whole blood and urine by headspace capillary GC with cryogenic oven trapping. *Journal of Chromatographic Science* 41: 63–66.

de Zeeuw RA, Franke JP, Machata G, et al. (1992) *Gas Chromatographic Retention Indices of Solvents and Other Volatile Substances for Use in Toxicological Analysis*. Report XIX of the DFG Commission for Clinical-Toxicological Analysis. Weinheim: VCH.

# FOURIER TRANSFORM TECHNIQUES

**P F Bernath**, University of Waterloo, Waterloo, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The Fourier transform (FT) is ubiquitous in science and engineering. For example, it finds application in the solution of equations for the flow of heat, for the diffraction of electromagnetic radiation, and for the analysis of electrical circuits. The concept of the FT lies at the core of modern

electrical engineering, and is a unifying concept that connects seemingly different fields. The availability of user-friendly commercial computer programs such as Maple<sup>™</sup>, Mathematica<sup>™</sup>, and Matlab<sup>™</sup> allows the FT to be part of every technical person's toolbox. The FT can be used to interpolate functions and to smooth signals. For example, in the processing of pixelated images, the high spatial frequency edges of pixels can easily be removed with the aid of a two-dimensional FT. This article, however, is not about the use of the FT as a tool in applied mathematics, but as the basis for techniques of analytical measurement.

and possible loss of toluene from the sample prior to analysis, sample contamination, etc., the lack of a strong correlation between blood concentrations and clinical features of poisoning is probably due to rapid initial tissue distribution and elimination.

About 80% of a dose of toluene is converted to hippuric acid (Table 4). Similarly, more than 90% of a dose of xylene is metabolized to methylhippuric (toluric) acids. The principal isomer found in urine is 3-methylhippurate, since *m*-xylene is the principal component of technical grade xylene (Table 1). Methylhippurates are not normal urinary constituents, but hippuric acid may arise from the metabolism of benzoates in foods and medicines, and thus caution is needed in the interpretation of results. Hippurate and methylhippurate excretion is often expressed as a ratio to creatinine since this obviates the need for 24 h urine collections. Occupational exposure to toluene can give rise to ratios of up to 1 g of hippurate per gram creatinine or more; in patients suspected of VSA, a ratio of more than 1 g of hippurate per gram creatinine strongly suggests, but does not prove, toluene exposure. Measurement of urinary *o*-cresol has been proposed as an alternative means of monitoring toluene exposure selectively, particularly in occupational circumstances, but the assay procedure is relatively complex and is thus not widely used.

See also: **Forensic Sciences:** Overview.

## Further Reading

Baselt RC (2002) *Disposition of Toxic Drugs and Chemicals in Man*, 6th edn. Foster City, CA: Chemical Toxicology Institute.

Broussard LA (2000) The role of the laboratory in detecting inhalant abuse. *Clinical Laboratory Sciences* 13: 205–209.

Brouette T and Anton R (2001) Clinical review of inhalants. *American Journal on Addictions* 10: 79–94.

Field-Smith ME, Bland JM, Taylor JC, et al. (2002) Trends in deaths associated with abuse of volatile substances 1972–2002. Report 15. London: St George's Hospital Medical School (<http://www.sghms.ac.uk>)

Fiserova-Bergerova V (1983) *Modelling of Inhalation Exposure to Vapours: Uptake, Distribution and Elimination*, vols. I and II. Boca Raton: CRC Press.

Flanagan RJ (2003) Volatile substances. In: Moffatt AC, Osselton D, and Widdop B (eds.) *Clarke's Isolation and Identification of Drugs*, 3rd edn., pp. 227–238. London: Pharmaceutical Press.

Flanagan RJ, Streete PJ, and Ramsey JD (1997) *Volatile Substance Abuse – Practical Guidelines for Analytical Investigation of Suspected Cases and Interpretation of Results*. UNDCP Technical Series No. 5. Vienna: UN International Drug Control Programme (<http://www.undcp.org>).

Haverkos HW and Dougherty JA (eds.) (1988) *Health Hazards of Nitrite Inhalants*. NIDA Research Monograph 83. Rockville: National Institute on Drug Abuse.

Pihlainen K and Ojanperä I (1998) Analytical toxicology of fluorinated inhalation anaesthetics. *Forensic Science International* 97: 117–133.

Watanabe-Suzuki K, Nozawa H, Suzuki O, and Ishii A (2003) Sensitive analysis of alkyl alcohols as decomposition products of alkyl nitrites in human whole blood and urine by headspace capillary GC with cryogenic oven trapping. *Journal of Chromatographic Science* 41: 63–66.

de Zeeuw RA, Franke JP, Machata G, et al. (1992) *Gas Chromatographic Retention Indices of Solvents and Other Volatile Substances for Use in Toxicological Analysis*. Report XIX of the DFG Commission for Clinical-Toxicological Analysis. Weinheim: VCH.

# FOURIER TRANSFORM TECHNIQUES

**P F Bernath**, University of Waterloo, Waterloo, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The Fourier transform (FT) is ubiquitous in science and engineering. For example, it finds application in the solution of equations for the flow of heat, for the diffraction of electromagnetic radiation, and for the analysis of electrical circuits. The concept of the FT lies at the core of modern

electrical engineering, and is a unifying concept that connects seemingly different fields. The availability of user-friendly commercial computer programs such as Maple<sup>TM</sup>, Mathematica<sup>TM</sup>, and Matlab<sup>TM</sup> allows the FT to be part of every technical person's toolbox. The FT can be used to interpolate functions and to smooth signals. For example, in the processing of pixelated images, the high spatial frequency edges of pixels can easily be removed with the aid of a two-dimensional FT. This article, however, is not about the use of the FT as a tool in applied mathematics, but as the basis for techniques of analytical measurement.

## Fourier's Integral Theorem

Fourier's integral theorem is a remarkable result:

$$F(v) = \int_{-\infty}^{\infty} f(t) e^{-i2\pi vt} dt \quad [1]$$

and

$$f(t) = \int_{-\infty}^{\infty} F(v) e^{+i2\pi vt} dv \quad [2]$$

where  $F(v)$  is the FT of an arbitrary time-varying function  $f(t)$  and  $i = \sqrt{-1}$ . We adopt the notation of lower case letters for a function,  $f(t)$ , in the time domain and upper case letters for the corresponding Fourier-transformed function,  $F(v)$ , in the frequency domain. The variable  $v$  is the frequency in units of hertz ( $s^{-1}$ ). The second eqn [2] defines the inverse FT that yields the original function,  $f(t)$ . Equations [1] and [2] can be written in many ways, but the version above is convenient for practical work because of the absence of factors (e.g.,  $1/2\pi$ ) in front of the integrals. These factors can be an annoying source of errors in the computation of FTs. The variables  $t$  and  $v$  can be replaced by any reciprocal pair (e.g.,  $x$ , in cm, for optical path difference and  $\tilde{\nu}$ , in  $cm^{-1}$ , for wavenumber in an infrared FT spectrometer) as long as their product is dimensionless.

The exponential in eqn [2] can be expanded as

$$e^{i2\pi vt} = \cos(2\pi vt) + i \sin(2\pi vt) \quad [3]$$

to give

$$f(t) = \int_{-\infty}^{\infty} F(v) \cos(2\pi vt) dv + i \int_{-\infty}^{\infty} F(v) \sin(2\pi vt) dv \quad [4]$$

The simple physical interpretation of eqn [4] is that any arbitrary (not necessarily periodic!) function  $f(t)$  can be expanded as an integral ('sums') of sine and cosine functions, with  $F(v)$  interpreted as the amplitudes of the 'waves.' The necessary amplitudes  $F(v)$  can be obtained from eqn [1], which thus represents the frequency analysis of the arbitrary function,  $f(t)$ . In other words, eqn [1] analyses the function  $f(t)$  in terms of its frequency components and eqn [2] puts the components back together again to recreate the function. Notice that if  $f(t)$  is an even function (i.e.,  $f(-t) = f(t)$ ), then the cosine transform suffices (i.e., only first term on the right hand side of eqn [4] need be retained), but this is rarely the case in practice.

## Fourier Transforms

The FT can be applied to a number of simple functions as presented in pictorial form in **Figure 1**. The FT of a Gaussian is another Gaussian, the decaying exponential (double-sided) gives a Lorentzian, and the boxcar function gives a sinc(=  $\sin(x)/x$ ) function.

The FTs of a number of elementary functions require the use of the delta function,  $\delta(v - v_0)$ . The  $\delta(v - v_0)$  function has the 'sifting' property,

$$f(v_0) = \int_{-\infty}^{\infty} \delta(v - v_0) f(v) dv \quad [5]$$

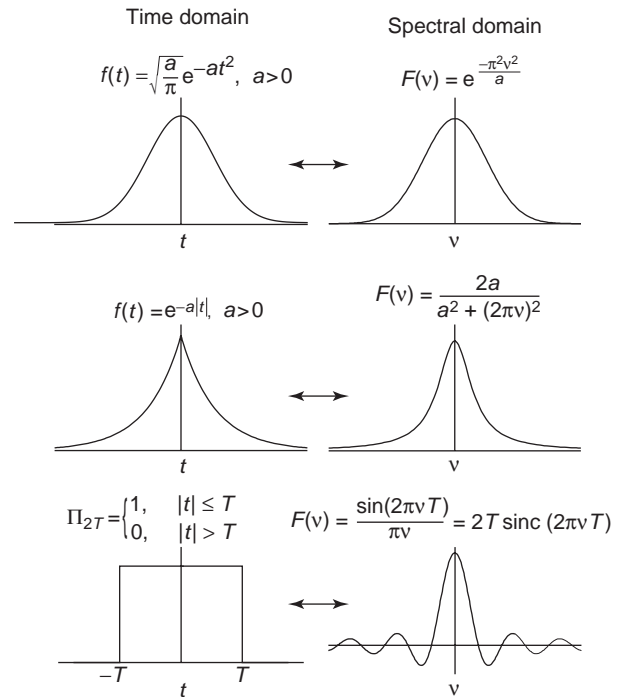
that implies unit area,

$$1 = \int_{-\infty}^{\infty} 1 \cdot \delta(v - v_0) dv \quad [6]$$

and a value of 0 for  $v \neq v_0$ .

The FT of the infinitely long wave,  $\cos(2\pi v_0 t)$ , is thus,  $(\delta(v - v_0) + \delta(v + v_0))/2$ , which has a value of  $\infty$  when  $v = v_0$  and  $v = -v_0$ . The appearance of negative frequencies is at first surprising, but is required by the mathematics of complex numbers. The identity

$$\delta(v - v_0) = \int_{-\infty}^{\infty} e^{\pm i2\pi(v-v_0)t} dt \quad [7]$$



**Figure 1** The Fourier transforms of Gaussian, double-sided exponential, and boxcar functions.



can be interpreted as the infinite sum of waves all phased to add up at  $\nu = \nu_0$  and to cancel for  $\nu \neq \nu_0$ . No function in the usual sense can be infinitely high, infinitely narrow and still have unit area. In mathematics, the delta function is thus defined as the limit of a series of peaked functions such as Gaussians.

While the FT of the even cosine function is real, the result for  $\sin(2\pi\nu_0 t)$  is imaginary:  $(\delta(\nu - \nu_0) + \delta(\nu + \nu_0))/(2i)$ . Since the sine function is  $90^\circ$  out of phase to the corresponding cosine, it is clear that the imaginary axis is used to keep track of phase shifts, consistent with the polar representation of a complex number:  $x + iy = re^{i\phi}$  with  $r = \sqrt{x^2 + y^2}$  and  $\phi = \tan^{-1}(y/x)$ . In this phasor picture, positive and negative frequencies can be interpreted as clockwise and counterclockwise rotations in the complex plane.

The FT has many useful mathematical properties including linearity, and using eqn [2], the derivative  $df(t)/dt$  has the transform  $(i2\pi\nu)F(\nu)$ . The convolution theorem is particularly useful because it relates the product of two functions,  $F(\nu) \cdot G(\nu)$ , in the frequency domain to the convolution integral

$$f * g(t) \equiv \int_{-\infty}^{\infty} f(\tau)g(t - \tau) d\tau \quad [8]$$

in the time domain (or vice versa), using the upper case/lower case FT notation for the  $(F(\nu), f(t))$ ,  $(G(\nu), g(t))$  pairs. For example, a finite piece of a cosine wave represented by the product,  $\cos(2\pi\nu_0 t)\Pi_{2T}(t)$ , leads to the convolution (Figure 1) of two delta functions with a sinc function in the frequency domain. The result is,  $T\text{sinc}(2\pi T(\nu - \nu_0)) + T\text{sinc}(2\pi T(\nu + \nu_0))$ , i.e., the infinitely narrow  $\delta$ -functions at  $\pm \nu_0$  produced by the FT of the infinite cosine have been broadened into sinc functions for the more realistic case of a finite length cosine wave. Similarly, the FT of the double-sided decaying exponential wave,  $\cos(2\pi\nu_0 t)\exp(-a|t|)$ , is two Lorentzians centered at  $\pm \nu_0$  (Figure 1) in the frequency domain.

## Discrete Fourier Transform

The trouble with practical applications of Fourier's integral theorem is that it requires continuous functions for an infinite length of time. These conditions are clearly impossible, so the case of a finite number of observations must be considered. Consider sampling the data every  $\Delta t$  for  $2N$  equally spaced points from  $-N$  to  $N-1$  with  $t_j = j\Delta t$ ;  $j = -N, -N+1, \dots, 0, \dots, N-1$ . The FT, eqn [1], then becomes the discrete FT:

$$F(\nu) = \Delta t \sum_{j=-N}^{N-1} f(j\Delta t)e^{-i2\pi\nu j\Delta t} \quad [9]$$

The question immediately arises as to the number of points required to sample the signal  $f(t)$ . If not enough points are taken the signal will be distorted, and if too many points are used there is a waste of resources. The answer is a remarkable result due to Nyquist: given a signal with no frequency components above  $\nu_{\max}$ , the signal can be completely recovered if it is sampled at a frequency of  $2\nu_{\max}$  (or greater). The Nyquist (or critical) sampling at  $2\nu_{\max}$  corresponds to two data points per wavelength for the frequency component at  $\nu_{\max}$ . It seems almost magical that such sparse, minimal sampling allows exact recovery of the original signal by interpolation. The connection between the time and the frequency domains for critically sampled data is illustrated in Figure 2. In the time domain, the interferogram is recorded from  $-T$  to  $T$  with a point spacing  $\Delta t = 1/(2\nu_{\max})$ , while in the spectral domain the data is present from  $-\nu_{\max}$  to  $\nu_{\max}$  with a frequency point spacing of  $\Delta\nu = 1/(2T)$ . In the particular case of an even function, although  $2N$  points ( $2N = 2T/\Delta t = 2\nu_{\max}/\Delta\nu$ ) are displayed in Figure 2, only  $N$  of them are independent. The discrete inverse FT is

$$f(t) = \Delta\nu \sum_{k=-N}^{N-1} F(k\Delta\nu)e^{i2\pi kt\Delta\nu} \quad [10]$$

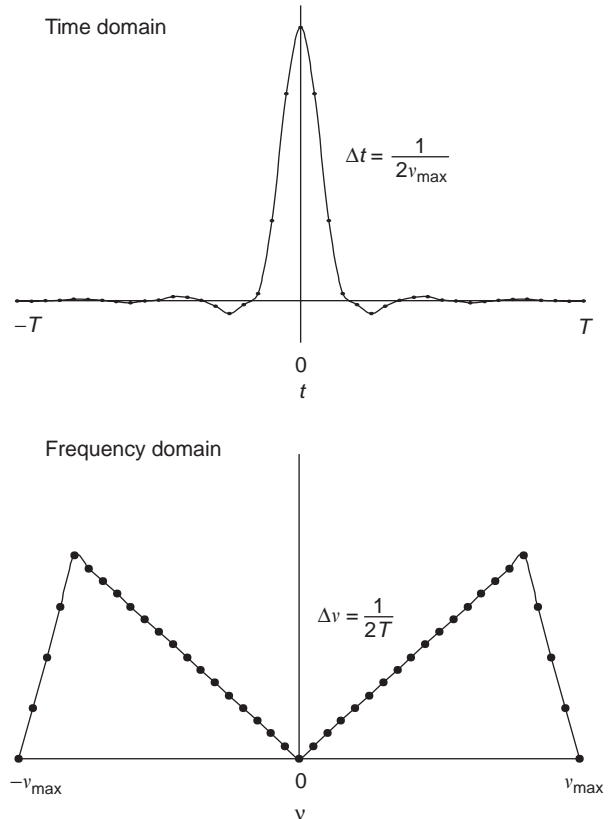


Figure 2 Nyquist sampling in the spectral and time domains.



with  $t$  only given at  $t_j = j\Delta t$  in eqn [10] and  $v$  given at  $v_k = k\Delta v$  in eqn [9].

Undersampling a signal is unfortunately a common occurrence and causes aliasing. A simple example of aliasing is the observation on television of a wheel on a wagon or a car rotating ‘backwards’ as the vehicle moves forward. For television in the USA, motion picture cameras sample the image at 30 times per second (the frame rate) and the image of the rotating wheel is thus often undersampled. When the sampling frequency  $v_s$  is somewhat less than the required  $2v_{\max}$ , then the spectrum between  $v_s/2$  and  $v_{\max}$  is ‘folded back’  $\sim v_s/2$  and appears as a reversed artifact between  $-v_s/2$  and  $v_s/2$ . If the signal is very undersampled, then it can be folded (aliased) many times (like fan-folded printer paper) between  $-v_s/2$  and  $v_s/2$ . This folding of the spectrum  $\sim v_s/2$  is called aliasing.

Aliasing can sometimes be used to advantage. For example, if a spectrum has no signal between 0 and  $v_{\max}/2$ , then every second point can be deleted (‘decimation’) because this undersampling by a factor of 2 will fold the signal from  $v_{\max}/2$  to  $v_{\max}$  backwards into the empty region from 0 to  $v_{\max}/2$ .

The discrete FT, eqn [9], can be evaluated in a brute force fashion on a computer using the available sine and cosine functions, eqn [3], but this method is very slow for a large number of points. The FT algorithm of Cooley and Tukey is much faster. The derivation of the Cooley–Tukey algorithm (‘fast Fourier transform’) starts by rewriting the exponent in eqn [10] as

$$i2\pi tk\Delta v = i2\pi j\Delta t k\Delta v = i\pi jk/N \quad [11]$$

using

$$t_j = j\Delta t; \quad j = -N, \dots, N-1 \quad \text{and} \quad \Delta t\Delta v = \Delta t/(2T) = 1/(2N) \quad [12]$$

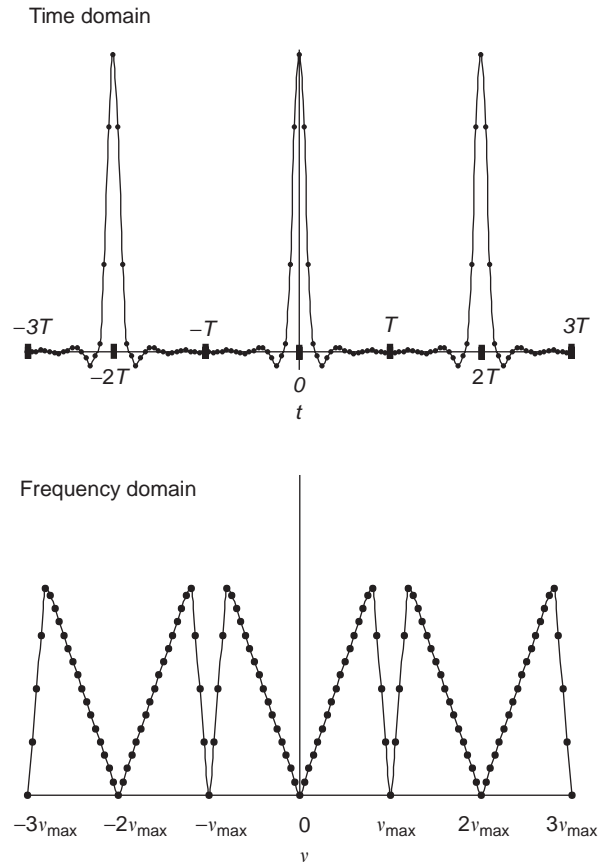
The discrete FT and inverse FT, eqns [9] and [10], thus become:

$$\begin{aligned} F(v_k) &= \Delta t \sum_{j=-N}^{N-1} f(j\Delta t) e^{-i\pi jk/N} \\ &= \Delta t \sum_{j=0}^{2N-1} f(j\Delta t) e^{-i\pi jk/N} \end{aligned} \quad [13]$$

$$\begin{aligned} f(t_j) &= \Delta v \sum_{k=-N}^{N-1} F(k\Delta v) e^{i\pi jk/N} \\ &= \Delta v \sum_{k=0}^{2N-1} F(k\Delta v) e^{i\pi jk/N} \end{aligned} \quad [14]$$

and the limits of the summation are shifted from  $-N$  to  $N-1$  to 0 to  $2N-1$  by considering the  $f(t)$  from  $-T$  to  $T$  and  $F(v)$  from  $-v_{\max}$  to  $v_{\max}$  as periodic functions (Figure 3). The fast FT requires  $2N$  to be a power of two, i.e.,  $2^m$  so the data are padded with zeros up to the next highest power of 2. The algorithm works by repeatedly ( $m$  times) dividing a  $2N$ -point transform into two smaller  $N$ -point transforms. The resulting final transform has  $2N$  points in the frequency domain, with the first  $N$  covering 0 to  $v_{\max}$ . The second  $N$  points from  $v_{\max}$  to  $2v_{\max}$  are the aliased points from  $-v_{\max}$  to 0 (see Figure 3).

The fast FT allows optimal interpolation of data. The original  $N$  points are folded  $\sim 0$  to make  $2N$  points and are then shifted by  $+N$  (Figure 3). The fast FT then creates  $2N$  points in the frequency domain, which are padded by the desired number of extra zeros in the appropriate location in the middle (e.g.,  $6N$  zeros in total for fourfold interpolation) and transformed back. (The extra zeros are added in the middle because of the aliasing of points from  $-v_{\max}$  to 0 into  $v_{\max}$  to  $2v_{\max}$  as shown in Figure 3.) This procedure creates interpolated points between the original data points.



**Figure 3** Frequency and time domains with the signals considered to be periodic.

## Fourier Transform Spectroscopy

Spectra are traditionally recorded by dispersing the radiation and measuring the absorption or emission, one point at a time. In some regions of the spectrum, for which tunable radiation sources are available, one can imagine stepping the frequency of the source from  $\nu_n$  to  $\nu_{n+1}$  by  $\Delta\nu$  (Figure 2) and recording the absorption. The primary attraction of FT techniques as compared to the 'traditional' approach is that all frequencies in the spectrum are detected at once. This property is the so-called multiplex or Fellgett advantage of FT spectroscopy.

Most FT measurements at long wavelengths (e.g., Fourier transform nuclear magnetic resonance (FT-NMR)) are made by irradiating the system with a short broadband pulse capable of exciting all the frequency components of the system, and then monitoring the free induction decay response. Such an approach presumes the availability of a coherent, high-power source of radiation that covers the entire spectral region of interest. A simple free induction decay has

$$f(t) = e^{-at} \cos(2\pi\nu_0 t), \quad t \geq 0, \quad a > 0$$

and

$$f(t) = 0, \quad t < 0 \quad [15]$$

with the corresponding spectrum

$$F(\nu) = \frac{1}{2(a + i2\pi(\nu - \nu_0))} + \frac{1}{2(a + i2\pi(\nu + \nu_0))} \quad [16]$$

If  $\nu \approx \nu_0$  and  $a \ll \nu_0$ , then the second term of eqn [16] can be dropped to give

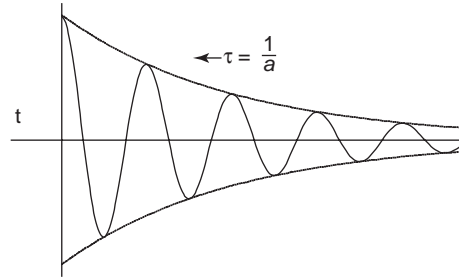
$$F(\nu) = \frac{a}{2(a^2 + 4\pi^2(\nu - \nu_0)^2)} - \frac{i2\pi(\nu - \nu_0)}{2(a^2 + 4\pi^2(\nu - \nu_0)^2)} \quad [17]$$

The real part of  $F(\nu)$  is a Lorentzian centered at  $\nu_0$  (first term on the right of eqn [17]), while the imaginary part (second term on the right of eqn [17]) is the corresponding dispersion curve (Figure 4). The constant  $1/a$  is the lifetime  $\tau$  of the decay and the full width at half maximum of the Lorentzian is  $a/\pi$ .

As compared to a double-sided even function, it is the abrupt turn-on of the free induction decay at  $t = 0$  that causes the large imaginary signal. All causal signals (defined to have  $f(t) = 0$  for  $t < 0$ ) have this property. The physical interpretation is that the abrupt start of the signal excites an in-phase response (the real part) to the cosine part of the original excitation wave as well as a response  $90^\circ$  out of phase (the imaginary part). The antisymmetric imaginary part is needed to cancel the symmetric real

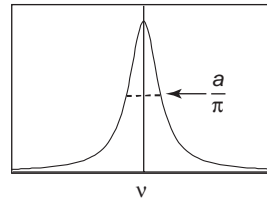
Time domain

$$f(t) = \begin{cases} e^{-at} \cos(2\pi\nu_0 t), & t \geq 0, \quad a > 0 \\ 0, & t < 0 \end{cases}$$



Frequency domain

$$\text{Re}\{F(\nu)\} = \frac{a/2}{a^2 + 4\pi^2(\nu - \nu_0)^2}$$



$$\text{Im}\{F(\nu)\} = \frac{-\pi(\nu - \nu_0)}{a^2 + 4\pi^2(\nu - \nu_0)^2}$$

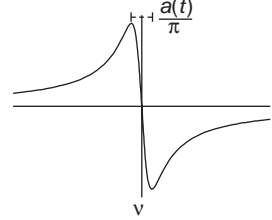


Figure 4 Fourier transform of a free induction decay.

part for  $t < 0$ . The large imaginary frequency component means that typically the magnitude spectrum is computed as

$$|F(\nu)| = \sqrt{(\text{Re}\{F(\nu)\})^2 + (\text{Im}\{F(\nu)\})^2} \quad [18]$$

for practical applications in, for example, FT mass spectrometry.

As usual there is a reciprocal relationship between the time domain and the frequency domain. The more rapid is the damping (decay) of the signal (i.e., larger  $a$  and shorter lifetime  $\tau = 1/a$ ), the wider the Lorentzian and dispersion lineshape functions become in the spectral domain (Figure 4).

At higher frequencies, FT spectroscopy is generally carried out with a Michelson interferometer rather than by detection of a coherent transient decay. The Michelson interferometer divides the input radiation into two parts with a beamsplitter and then recombines them. As the optical path difference of the two parts is varied, the interference of the recombined beams produces an interferogram. If the optical path difference,  $x$ , changes at a constant rate,  $\nu$ , then the interferogram becomes a function of time,  $f(x) = f(\nu t)$ , and the FT yields the desired spectrum,  $F(\nu)$ . In general, double-sided interferograms are

recorded from  $-L(=-\nu T)$  to  $+L(=\nu T)$  in optical path difference, as shown in **Figure 2**. For practical reasons, to save resources sometimes, only a short double-sided interferogram is recorded to determine the phase. This low-resolution-phase function is then interpolated to generate a high-resolution-phase function and applied to a high-resolution single-sided spectrum recorded only from 0 to  $L$ , rather than from  $-L$  to  $+L$ .

The main barrier to the use of free induction decay for FT spectroscopy in the infrared region is the lack of a convenient powerful source of broadband coherent radiation. In addition, the decay times for the coherently excited polarization in the system tend to be very short at higher frequencies. Coherent terahertz spectrometers operating in the far infrared region are now practical because of the success of ultrafast laser technology in generating broadband terahertz pulses.

FT techniques are inherently digital and were not practical before the advent of the digital computer. The digital data can thus be filtered and manipulated with ease. In particular, the instrumental resolution can be varied at will and chosen to match the inherent resolution of the physical system. As shown in **Figure 2**, recording the interferogram for a longer time  $T$  (longer optical path difference for a Michelson interferometer) results in a closer point spacing  $\Delta\nu = 1/(2T)$  and higher instrumental resolution in the frequency domain. There are practical limits because when the interferometric signal is no longer visible because it is buried in noise, no further increase in resolution is possible. The FT spectrometer can thus easily trade decreased spectral resolution for increased signal-to-noise ratio. Very high instrumental resolutions are possible and resolving powers,  $R \equiv \nu/\Delta\nu$ , in excess of  $10^6$  can be achieved in the visible and infrared spectral regions, and for mass spectrometry.

When FT spectroscopy is implemented with a Michelson interferometer, there is an additional advantage over a grating or prism spectrograph that uses rectangular slits. The circular entrance aperture of the Michelson interferometer can be opened until the subtended solid angle,  $\Omega_{\max}$ , (in steradians, as measured using the distance to the collimator) is given by  $R\Omega_{\max} = 2\pi$ , with resolving power,  $R$ , computed for the highest measured frequency  $\nu_{\max}$ . For the same resolving power, a grating spectrograph typically has 50–100 times smaller  $\Omega_{\max}$  as subtended by the slits. Much more radiation can thus enter the larger entrance aperture of a FT spectrometer than a grating spectrometer, as was first pointed out by Jacquinot. The throughput or Jacquinot advantage is always operative for a Michelson interferometer.

Surprisingly, the multiplex ‘advantage’ of FT spectroscopy is not always available and depends on the principal noise source in the system. At low frequencies, the principal noise source is often detector or background noise. In this case, the noise is independent of the signal level and is constant. Compared to a single element detector used to detect a single frequency interval,  $\Delta\nu$ , the signal level for the FT spectrometer that detects  $N$  channels of width  $\Delta\nu$  is  $N$  times greater. The noise in these  $N$  channels also now appears at the detector, but because of partial cancellation due to the random nature of noise (mean value of zero), the noise level increases by only  $\sqrt{N}$ . Note that true noise always partially cancels in this manner but that signal artifacts (‘periodic noise’) will not generally do so. Overall then the signal-to-noise ratio has improved by  $\sqrt{N}$  for a FT spectrometer.

To higher frequencies, in the visible region, photons carry more energy and can be detected, for example, by counting. The noise in such a system is typically due to ‘shot’ noise arising from the Poisson statistics of random fluctuations in the arrival times of photons at the detector. Shot noise is proportional to the square root of the number of photons arriving at the detector. If the primary noise source is shot noise, then the noise is proportional to the square root of the signal level. In this case, for the  $N$ -channel FT spectrometer the signal level increases by  $N$  and the noise increases by  $\sqrt{N}$  for the increased number of channels and another factor of  $\sqrt{N}$  for the increased intensity. Overall there is no change in the signal-to-noise ratio as compared to a single channel spectrometer. Surprisingly then, in the visible region the multiplex advantage usually does not apply, although the throughput and ‘digital’ advantages remain.

Unfortunately, there is a third possibility for the noise. In remote sensing applications in the earth’s atmosphere or in astronomy, the noise may be directly proportional to the signal level. Transmission through the atmosphere has random fluctuations (‘scintillations’) that are analogous to ‘ $1/f$ ’ noise that appears in electronic circuits at low frequency. In the laboratory, emission from a sample may be excited by a laser or other source that is plagued by amplitude fluctuations. There may also be a weak molecular emission of interest in a spectrum dominated by strong fluctuating lines from an extraneous atom or molecule. In all these cases, the  $N$ -channel FT spectrometer increases the signal by  $N$  over a single channel spectrometer, but the noise also increases by another factor of  $\sqrt{N}$  for the increased number of channels and by  $N$  because of the increased intensity level. Incredibly, the signal-to-noise ratio has therefore decreased by  $\sqrt{N}$  and there is a multiplex

‘disadvantage’ for a FT spectrometer. Because the FT spectrometer detects the entire spectrum (including the noise) at one time, the noise from all the channels is distributed throughout the spectrum. In a simple single channel spectrometer, only the noise from that single channel appears, even though nearby channels may have strongly fluctuating signals. It is this redistribution of noise from all channels that causes the Fourier disadvantage.

The presence of a multiplex disadvantage accounts for the enormous success of small multichannel visible spectrometers based on array detectors, as compared to low resolution visible FT spectrometers. In any case, for a particular application a careful analysis of the noise is indispensable for an optimum instrument.

## Fourier Transform Applications

The FT technique has been applied in a multitude of different areas. Starting at low frequencies, FT methods have been used for dielectric response spectroscopy of solids (sometimes called time domain reflectometry). A short picosecond voltage pulse is applied to a dielectric and the current response is measured. Fourier transformation of the current gives the dielectric response function,  $\epsilon(\nu)$ , which is typically interpreted as the Debye relaxation of dipoles.

The main application at low frequencies is, of course, NMR. In FT-NMR, a magnetized sample is irradiated by radio frequency (RF) radiation to manipulate the bulk magnetization. The basic signal is the free induction decay, but hundreds of sophisticated RF pulse sequences have been invented for specific purposes, including medical imaging.

A related RF technique to NMR is nuclear quadrupole resonance (NQR). In NQR, transitions between nuclear quadrupole levels of nuclei in a solid material are induced by the applied radiation. The electric field gradients in the solid orient the quadrupolar nuclei ( $I > 1/2$ ) and give rise to quantized energy levels that yield transitions in the MHz range. FT-NQR spectroscopy measures these splittings and the relaxation times by free induction decay or various pulse echo experiments. FT-NQR spectroscopy provides information about the local environment around the quadrupolar nucleus in a crystal.

Electron spin resonance (ESR) operates at somewhat higher frequencies in the GHz range and is sometimes called electron paramagnetic resonance (EPR). ESR is like NMR but uses electron spins rather than nuclear spins. By definition, FT-ESR studies free radicals, and it is more sensitive than FT-NMR but of less general applicability. Related to

ESR is muon spin resonance ( $\mu$ SR), carried out at high-energy accelerator sources such as TRIUMF (Vancouver, Canada) that provide muons.

In the gigahertz region, FT-microwave experiments were pioneered by the late W.H. Flygare. FT-microwave experiments use a short pulse of microwave radiation to polarize a gaseous sample in a waveguide or, more commonly, in a Fabry–Perot cavity. Free induction decay is detected and Fourier transformed into a spectrum. FT-microwave spectroscopy of cold molecules in pulsed jet expansion is particularly popular because of the increased sensitivity associated with low temperatures, and the possibility of studying large molecules and van der Waals complexes.

In the infrared, visible, and near ultraviolet regions, FT methods generally use the Michelson interferometer rather than detecting a free induction decay. The practical short wavelength limit for the Michelson interferometer is  $\sim 170$  nm. There is no hard long wavelength limit, and measurements down to a few  $\text{cm}^{-1}$  are possible. Recently, coherent time domain terahertz spectroscopy has become possible in the far infrared region. In this case, the system is polarized with an ultrashort pulse of terahertz radiation and the coherent decay is detected using ultrafast laser techniques.

FT techniques have also been applied with great success to mass spectrometry. The ion cyclotron resonance (ICR) spectrometer traps ions in a magnetic field. The ions travel in circles about the applied magnetic field (cyclotron motion) and are trapped in the direction along the magnetic field by small voltages applied to the end caps of the trapping cell. A short pulse of RF radiation coherently phases the ions in the trap and increases their orbits. The phase-coherent orbiting ions induce small image currents on the two opposite metal walls of the trapping cell. The FT of these image currents yields the mass spectrum. In this case, the decay time of the free induction decay signal can be very long because the dephasing of the cyclotron motion of the ions in a homogenous magnetic field is controlled by collisions with residual gas. The FT-ICR technique can thus have ultrahigh mass resolution and great sensitivity.

*See also:* **Gas Chromatography:** Fourier Transform Infrared Spectroscopy. **Liquid Chromatography:** Fourier Transform Infrared Spectroscopy.

## Further Reading

Beard MC, Turner GM, and Schmuttenmaer CA (2002) Terahertz spectroscopy. *Journal of Physical Chemistry B* 106: 7146.

- Bell RJ (1972) *Introductory Fourier Transform Spectroscopy*. New York: Academic Press.
- Bracewell RN (2000) *The Fourier Transform and its Applications*, 3rd edn. New York: McGraw-Hill.
- Brigham E (1988) *Fast Fourier Transform and its Applications*. Englewood Cliffs, NJ: Prentice Hall.
- Chamberlain J (1979) *The Principles of Interferometric Spectroscopy*. Chichester: Wiley-Interscience.
- Davis S, Abrams MC, and Brault JM (2001) *Fourier Transform Spectrometry*. San Diego: Academic Press.
- Kauppinen J and Partanen J (2001) *Fourier Transforms in Spectroscopy*. Berlin: Wiley-VCH.
- Lathi BP (1968) *Communication Systems*. New York: Wiley.
- Marshall AG (ed.) (1982) *Fourier, Hadamard, and Hilbert Transforms in Chemistry*. New York: Plenum.
- Marshall AG and Verdun FR (1990) *Fourier Transforms in NMR, Optical, and Mass Spectrometry: A User's Handbook*. Amsterdam: Elsevier.
- Schweiger A and Ieschke G (2001) *Principles of Pulse Electron Paramagnetic Resonance*. New York: Oxford University Press.
- Slichter CP (1992) *Principles of Magnetic Resonance*, 3rd edn. Berlin: Springer.
- Thorne A, Litzen U, and Johansson S (1999) *Spectrophysics*. Berlin: Springer.

## FRUITS

See **FOOD AND NUTRITIONAL ANALYSIS: Fruits**

## FT TECHNIQUES

See **FOURIER TRANSFORM TECHNIQUES**

## FUELS

Contents

**Gaseous**

**Oil-Based**

### Gaseous

**H Müller**, Malsch, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 1667–1672,  
© 1995, Elsevier Ltd., with revisions made by the Editor.

### Introduction

Analysis of gaseous fuels is necessary for quality control. The most important quality criteria are the calorific value (CV) and the sulfur (S) content. The CV is related to the quantity of heat energy, which is liberated when the fuels are burned; the sulfur

content gives information concerning the potential for sulfur dioxide (SO<sub>2</sub>) formation, which must be limited for environmental reasons. The major types of gaseous fuels are natural gas, processed gas, e.g., coal gas, and blends of both. The CV and the total sulfur content can be determined as sum parameters (i.e., without knowledge of individual compounds) or by analyzing the fuel and calculating the CV and the sulfur content from the composition. This article provides a survey of the methods of gaseous fuel analysis with regard to online and offline analysis and special techniques.

### General Methods

Fuel gases consist of combustible compounds like hydrocarbons, hydrogen (H<sub>2</sub>), and carbon monoxide



- Bell RJ (1972) *Introductory Fourier Transform Spectroscopy*. New York: Academic Press.
- Bracewell RN (2000) *The Fourier Transform and its Applications*, 3rd edn. New York: McGraw-Hill.
- Brigham E (1988) *Fast Fourier Transform and its Applications*. Englewood Cliffs, NJ: Prentice Hall.
- Chamberlain J (1979) *The Principles of Interferometric Spectroscopy*. Chichester: Wiley-Interscience.
- Davis S, Abrams MC, and Brault JM (2001) *Fourier Transform Spectrometry*. San Diego: Academic Press.
- Kauppinen J and Partanen J (2001) *Fourier Transforms in Spectroscopy*. Berlin: Wiley-VCH.
- Lathi BP (1968) *Communication Systems*. New York: Wiley.
- Marshall AG (ed.) (1982) *Fourier, Hadamard, and Hilbert Transforms in Chemistry*. New York: Plenum.
- Marshall AG and Verdun FR (1990) *Fourier Transforms in NMR, Optical, and Mass Spectrometry: A User's Handbook*. Amsterdam: Elsevier.
- Schweiger A and Ieschke G (2001) *Principles of Pulse Electron Paramagnetic Resonance*. New York: Oxford University Press.
- Slichter CP (1992) *Principles of Magnetic Resonance*, 3rd edn. Berlin: Springer.
- Thorne A, Litzen U, and Johansson S (1999) *Spectrophysics*. Berlin: Springer.

## FRUITS

See **FOOD AND NUTRITIONAL ANALYSIS: Fruits**

## FT TECHNIQUES

See **FOURIER TRANSFORM TECHNIQUES**

## FUELS

Contents

**Gaseous**

**Oil-Based**

### Gaseous

**H Müller**, Malsch, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 1667–1672,  
© 1995, Elsevier Ltd., with revisions made by the Editor.

### Introduction

Analysis of gaseous fuels is necessary for quality control. The most important quality criteria are the calorific value (CV) and the sulfur (S) content. The CV is related to the quantity of heat energy, which is liberated when the fuels are burned; the sulfur

content gives information concerning the potential for sulfur dioxide (SO<sub>2</sub>) formation, which must be limited for environmental reasons. The major types of gaseous fuels are natural gas, processed gas, e.g., coal gas, and blends of both. The CV and the total sulfur content can be determined as sum parameters (i.e., without knowledge of individual compounds) or by analyzing the fuel and calculating the CV and the sulfur content from the composition. This article provides a survey of the methods of gaseous fuel analysis with regard to online and offline analysis and special techniques.

### General Methods

Fuel gases consist of combustible compounds like hydrocarbons, hydrogen (H<sub>2</sub>), and carbon monoxide



(CO), which are the source of energy, and of incombustible compounds like helium (He), nitrogen (N<sub>2</sub>), and carbon dioxide (CO<sub>2</sub>), which decrease the specific energy content. Sulfur-containing compounds are in the gas as natural contaminants and odorants. In a refined fuel gas their concentrations are normally in the µl per l range and do not make a significant contribution to the CV. Traditionally CV and sulfur content are determined as sum parameters. Therefore a known quantity of the gas is burned in an excess oxygen atmosphere under exactly defined conditions. To obtain the CV, this burning takes place in a combustion calorimeter: the reaction enthalpy is absorbed in a known amount of fluid and the temperature rise of this fluid is measured. So it is generally possible to determine the CV. Operating a combustion calorimeter is difficult, because the parameters temperature, pressure, gas, and fluid volume have to be measured very accurately. Minor changes in ambient conditions can dramatically affect the result. To prevent this, many technical resources are necessary. Moreover, combustion calorimeters are very expensive and incur high costs during operation. To determine the total sulfur content, the sample is burned in oxyhydrogen in a *Wickbold* apparatus by Wickbold method to determine sulfur in fuels by oxidation to form sulfur dioxide according to DIN 51409. Sulfur dioxide is formed and hydrogen peroxide is added to create sulfate anions; their concentration is determined by titrimetric methods. This analysis is very difficult and time consuming, with poor precision and reproducibility. There are other methods for total sulfur determination, based on the principle of burning the fuel sample and measuring the concentration of the SO<sub>2</sub> formed by gas chromatography or luminescence detection.

However, none of the burning methods gives any information about individual compounds. To overcome these problems and following the development and proven efficacy of gas chromatography in many other fields, this technique was introduced in fuel analysis. Now it has been possible to determine individual compounds and to calculate the integral parameters of density, compressibility, *Wobbe*-index, the volumetric basis superior calorific value at specified reference conditions divided by the square root of the relative density at the same reference conditions. etc. from physical properties of the pure compounds. So such analysis became a powerful tool for gas quality control to ensure that gas distributors deliver the contracted heat energy to their customers. The gas chromatographs (GCs) can be operated online using process GCs or offline using laboratory GCs. Online operation means installation of the GCs

near the 'production process', i.e. near the drilling hole, at the pipeline in the delivery point or in the refinery plant and implies fully automatic operation 24 hours a day even under extreme ambient conditions. Offline operation means that a qualified person operates the instrument in any control station or in an analytical laboratory.

## Instrumentation

### Gas Chromatographs

Basic components of GCs are the sample injector, the separation column(s), switching valves, detector(s), and a computer system for evaluating and reporting the results. Process GCs are equipped and tuned to analyze the same type(s) of fuel in a very short analysis time with high accuracy and reliability and they must be able to calibrate themselves. Modern process GCs are operated using a personal computer (PC) connected directly to the control unit or via a telephone link. The PC may be connected directly to the control unit of the GC or via modem and telephone lines. If necessary, the operator can carry out operation, communication, and diagnosis from a remote location such as a control laboratory. It is also possible to retrieve all data from the controller and generate reports and chromatograms on a printer/plotter connected to the PC or GC.

The requirements of a laboratory GC are different from those of a process GC. The need for accuracy and reliability is the same but ambient conditions, analysis time, and remote control are less important. More crucial are high analytical performance, flexibility, and smooth operation because laboratory GCs are used to analyze a wide spectrum of gaseous fuels like natural gases of different origin and their substitutes.

Obviously, the chromatographic principles are the same in process and laboratory GCs and they are built up in a very similar way. Standard detectors are in each case the thermal conductivity detector (TCD), which is a universal detector for all components, and the flame ionization detector (FID), which is a specific detector for hydrocarbons. To detect sulfur gases selective detectors like an electrochemical detector, chemiluminescence detector and, most important, flame photometric detector (FPD) are used. Gaseous fuels like natural gas, synthetic gases, and blends are complex mixtures that cannot completely be separated in a single column. Two or more different columns must be combined. To monitor the fuel quality a quasi-continuous analysis is necessary; this means that very short cycle times must be realized. To do so, high-boiling components are removed

from the system by backflushing. To distribute the components to different columns and to backflush any column, special switching techniques which must correspond with the chromatographic system are necessary. Multiport valves are normally used together with packed columns whereas 'valveless' techniques like *Deans*- and *Live*-switching are preferably used together with capillary columns. Advantages of the 'valveless' techniques are low dead volume, no movable parts in the chromatographic path, and no polymer and adsorptive surfaces to cause memory effects and peak broadening. These properties are essential to achieve high resolution with capillary columns.

## Separation Columns

Traditional gaseous fuel analysis is based on packed columns. Multidimensional chromatography, i.e. the use of columns with different retention characteristics coupled in series in one GC system, improves resolution of complex samples. But packed-column systems are not normally able to separate all compounds. To obtain more precise information about the composition of a gas sample, capillary columns were introduced in fuel analysis. The first were wall-coated open tubular (WCOT) columns with a nonpolar methylsilicone as stationary phase for hydrocarbon separation. However, the permanent gases were still separated in packed columns containing cross-linked polystyrene carbosieve, and molecular sieve as separation phases.

With the introduction of the so-called porous layer open tubular (PLOT) columns a new epoch of gas analysis began. PLOT columns are capillary columns with a very high resolution capability. The dimensions are ~10 to 50 m in length (standard length: 25 or 30 m) and 0.32 or 0.53 mm in internal diameter. The increased resolution of these columns can be quantified by the theoretical plate number, which is in the range of 500 to 1000 (packed columns: 100 to 300) theoretical plates per meter of column. The separation mechanism is based on gas-solid chromatography, which is different from gas-liquid chromatography in silicone-oil coated columns. The high activity of the solid adsorbent in PLOT columns causes an intensive and selective retention force.

Typical adsorbents in PLOT columns are, similar to packed columns, aluminum oxide, molecular sieve, and modified cross-linked porous polymers. Aluminum oxide is a very selective stationary phase and it enables the separation of all C<sub>1</sub>- to C<sub>4</sub>-hydrocarbons according to C-number, structure, and degree of saturation. The high surface activity is

combined with a very strong adsorption of polar substances like sulfur gases, water, etc., which change the polarity and thereby the separation characteristics of the column. Molecular sieve separates the noble gases, hydrogen, oxygen, nitrogen, methane (CH<sub>4</sub>) and carbon monoxide without cooling the column, but the retention is affected even by low concentrations of polar gases, which are adsorbed on the surface. Porous polymers are much more resistant. They are ideal for separating carbon dioxide from the permanent gases, light hydrocarbons, and water. The separation behavior can be influenced by generating polar groups in the polymer material. The favorable properties of the porous polymer, molecular sieve, and methyl-silicone columns and the excellent resolution, especially if combined with a temperature program, has made them standard materials in gas analysis. To exemplify the relevance of the instrumentation and column characteristics for gaseous fuel analysis, representative analytical methods with packed and capillary columns are described below.

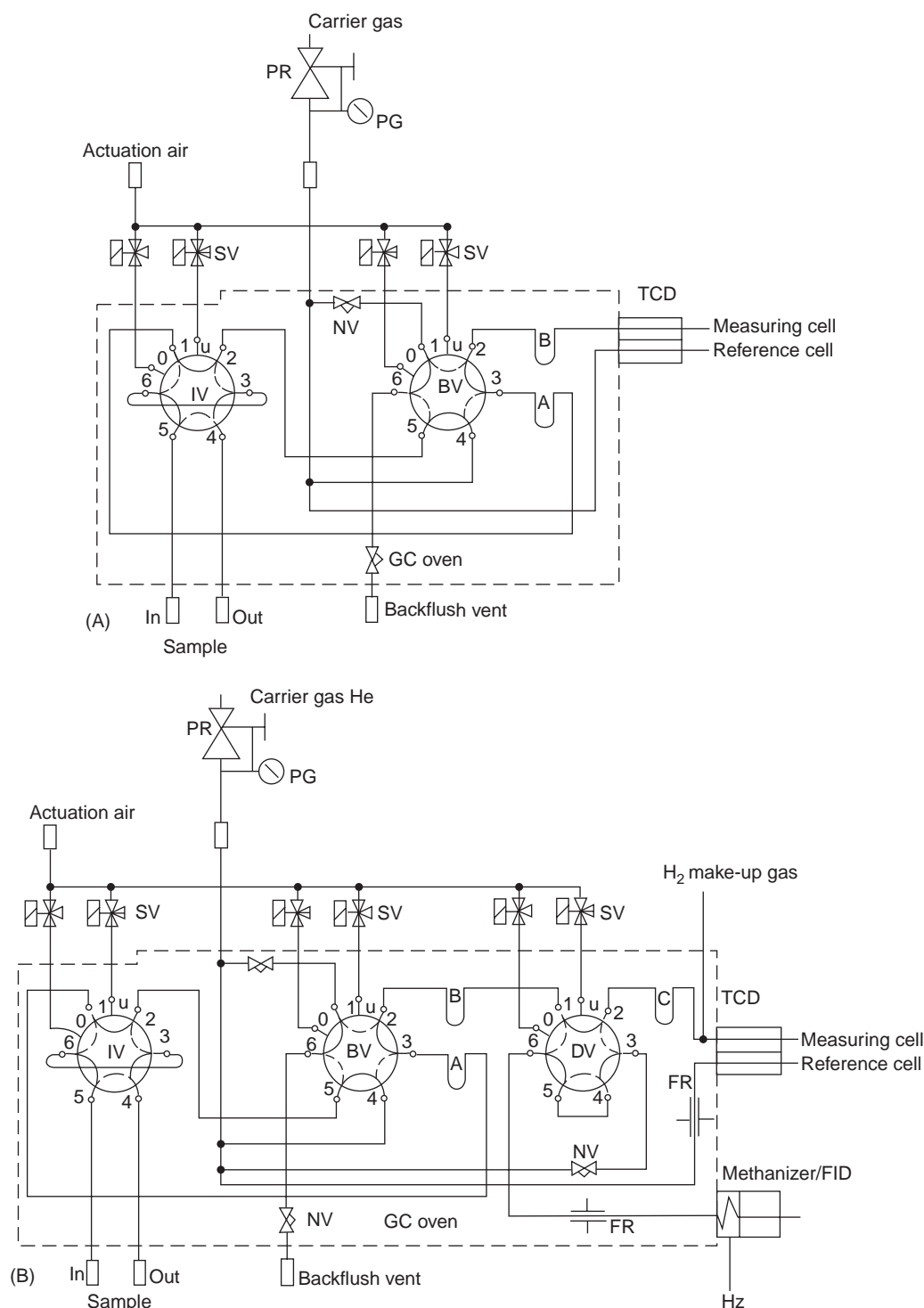
## CV Determination in Gaseous Fuel

### Analysis with Packed Columns

Gaseous fuels may be quite different in composition and therefore no universal analytical method exists. Particularly in process chromatography, only the most interesting compounds are individually determined, to obtain short cycle times. Less important compounds are neglected or determined as a sum peak. There are a number of different analytical methods.

International Standard ISO 6569-1981 describes a rapid natural gas analysis of samples during exploration consisting of a gas sampling valve, two packed columns which are filled with Porapak Q and Porapak T material as stationary phase. Porapak Q and Porapak T are porous polymers of different polarity. Detection is by a TCD and helium is used as carrier gas. The measured compounds N<sub>2</sub>, CO<sub>2</sub>, C<sub>1</sub>- to C<sub>4</sub>-hydrocarbons, and hydrogen sulfide (H<sub>2</sub>S) can be detected sensitively because of the big differences between their thermal conductivities and of the carrier gas helium. A disadvantage is that oxygen is not separated from nitrogen and contamination of the sample by air will give a value for the nitrogen content which is too high.

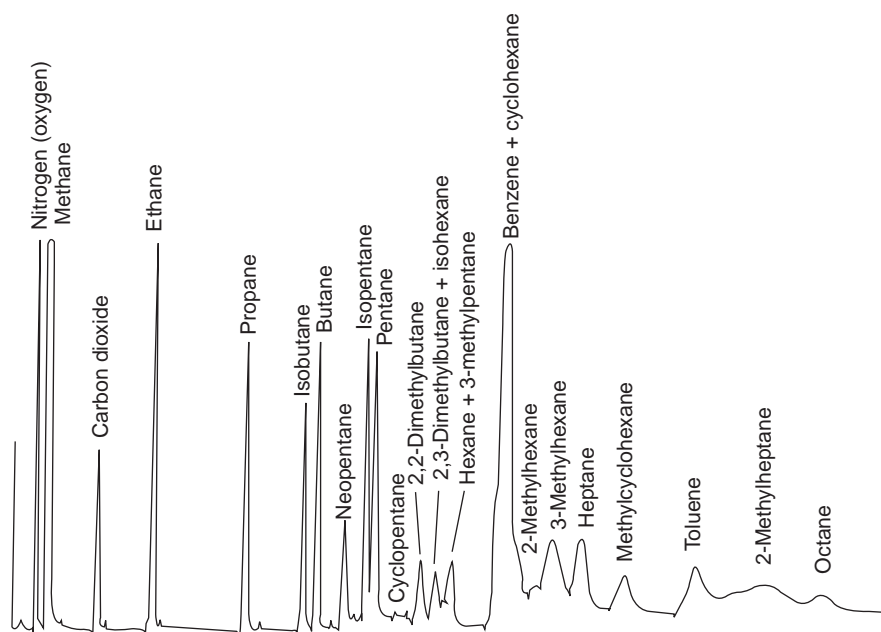
The GC system is shown schematically in **Figure 1A**. The backflush function (valve BV) is typical and essential for process analysis. It is necessary for conditioning the GC system and removing high-boiling hydrocarbons backwards from the first column (A)



**Figure 1** Packed column systems with six-port valves: (A) system with backflush capability; (B) system with backflush and column switching capability. PR, pressure regulator; PG, pressure gauge; SV, solenoid valve; NV, needle valve; IV, injection valve; BV, backflush valve; DV, distribution valve; A/B/C, column A/B/C; FR, flow restrictor.

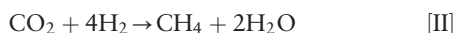
within analysis time. In laboratory GCs the backflush function is not always installed; in this case high-boiling compounds are eluted by a temperature program. A typical chromatogram is shown in

**Figure 2.** If a natural gas is blended with a coal gas for instance, the GC system described above is not sufficient, because hydrogen and carbon monoxide are present. For this analysis an extended system is



**Figure 2** Thermal conductivity detector chromatogram from a packed column system, temperature programmed according to ISO 6974–1984.

necessary, shown in **Figure 1B**: the additional molecular sieve column (C) allows the separation of hydrogen, oxygen, nitrogen, methane, and carbon monoxide. To overcome the anomaly in thermal conductivity of helium/hydrogen and to obtain a linear signal/concentration relation for hydrogen it is necessary to use hydrogen as make-up gas for the TCD. A further improvement can be effected by installing an FID as a second detector. In this procedure only the permanent gases and methane are detected by the TCD; carbon monoxide and carbon dioxide are detected after passing through a methanizer to the FID, like the separated and backflushed hydrocarbons. The methanizer is a reactor that carries out reactions [I] and [II] in the presence of hydrogen and a suitable catalyst.

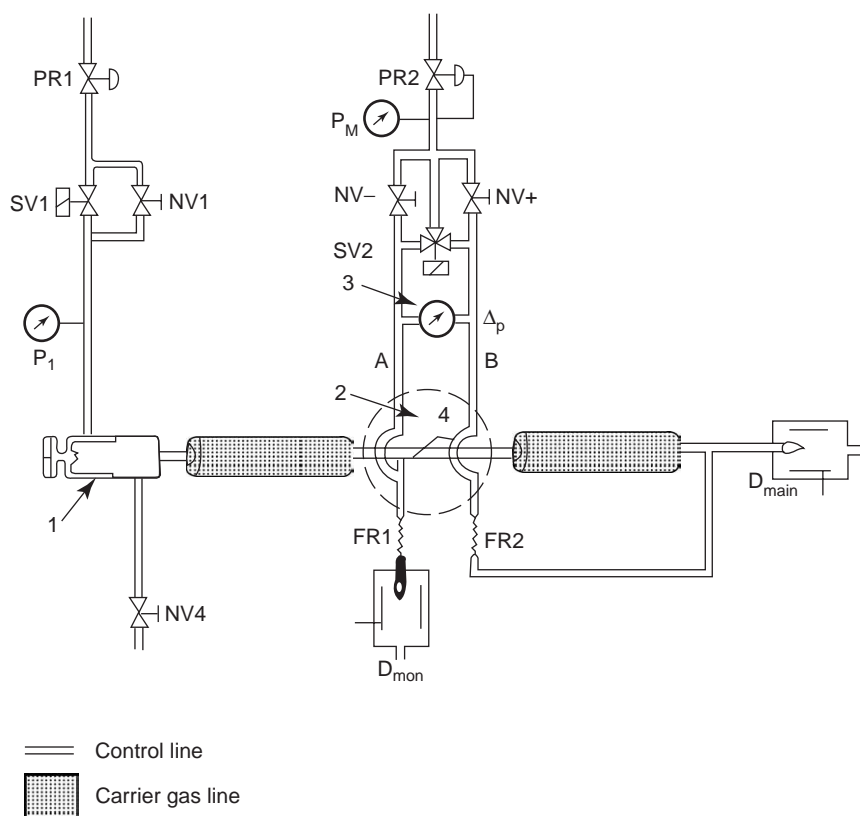


### Analysis with Capillary Columns

Capillary columns need a temperature program for maximum performance, which is why the main sphere of application is still the analytical laboratory. Process GCs are normally running under isothermal conditions because of the stability of retention times; nevertheless, more and more process GCs are equipped with capillary columns to benefit from the high resolution. A representative analysis

procedure is presented below (German Standard DIN 51872). By this method a large number of individual components in different kinds of gas fuels can be determined: permanent gases: helium, hydrogen, oxygen, nitrogen, carbon monoxide, carbon dioxide; hydrocarbons: methane, ethene, ethyne, ethane, propene, propane, *i*-butane, *n*-butane, 2,2-dimethylpropane, 2-methylbutane, *n*-pentane, cyclopentane, 2,2-dimethylbutane, 2,3-dimethylbutane, 3-methylpentane, *n*-hexane, benzene and cyclohexane, heptane, methylcyclohexane, toluene, octane, and higher hydrocarbons.

Three capillary columns are necessary. Two are connected in series in a two-column GC system (e.g. by Live-switching equipment as shown in **Figure 3**), and the third runs in a single-column system. After injection the gas sample passes through the first PLOT column, which contains porous polymer as the stationary phase. The  $\text{C}_2$ -hydrocarbons and carbon dioxide are separated, and the permanent gases including methane are flushed as sum peak to the second PLOT column containing molecular sieve. Here they are totally resolved. The higher hydrocarbons are removed from the first column by backflushing. The hydrocarbons eluting from the first column are detected by an FID. Carbon dioxide, which also elutes from this column, can be detected as methane after conversion in a methanizer. The components eluting from the molecular sieve column are detected by the TCD. If the TCD and



**Figure 3** Live equipment for valveless column switching. 1, injector with split valve. 2, live T-piece. 3, differential pressure manometer. 4, ring slot. NV, needle valve; SV, solenoid valve; PR, pressure regulator; FR, flow restrictor;  $D_{mon}$ , monitor detector;  $D_{main}$ , main detector;  $P_1$ , inlet pressure;  $P_M$  mean pressure;  $\Delta_p$ , differential pressure; A, control line A; B, control line B.

methanizer/FID are connected, methane and carbon monoxide can also be detected by the FID.  $C_3$ - and higher hydrocarbons are separated in a WCOT-column with a very thick film of methyl-silicone as stationary phase and detected by an FID. A representative chromatogram is shown in **Figure 4**.

A double oven GC equipped with two or three detectors is an ideal instrument for this analysis because each column can work under optimal temperature conditions. In the two-column system argon is used as the carrier gas to give good sensitivity for helium and hydrogen. Carbon monoxide and carbon dioxide are detected by the FID. The TCD response is very poor because the thermal conductivities are similar to that of argon. The same analysis is possible using a single oven GC with only one FID. In this case very precise tuning of the chromatographic system is necessary to avoid simultaneous elution of peaks from two columns to one detector.

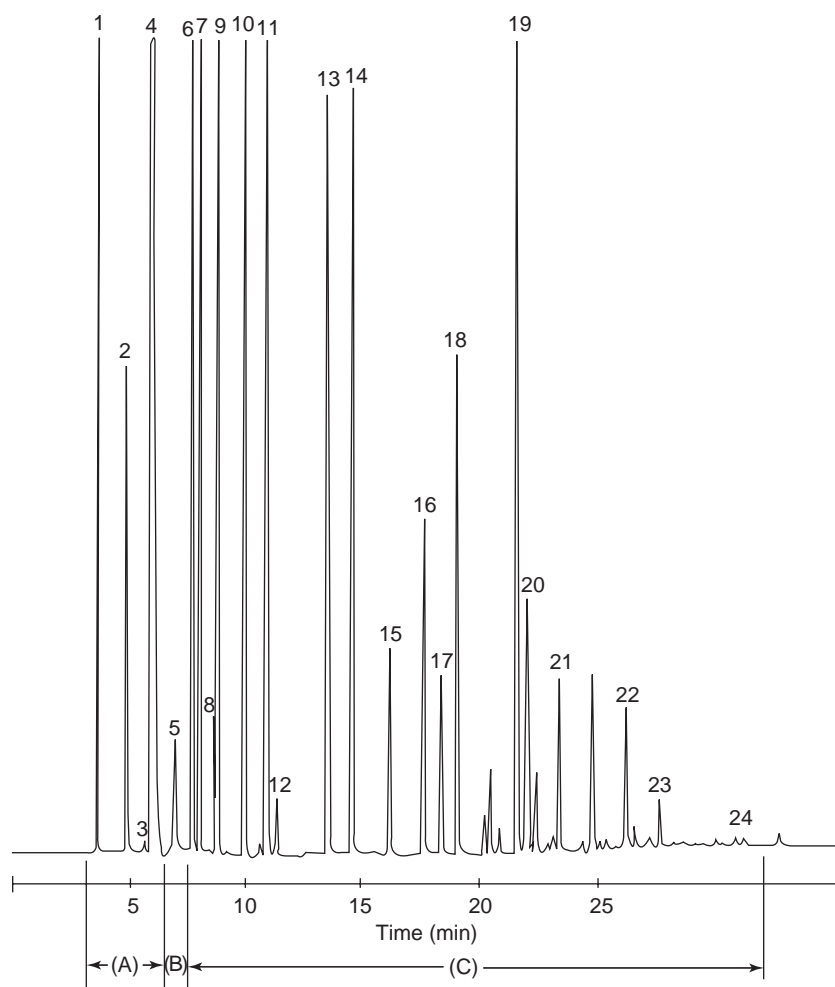
## Determination of Sulfur Compounds

The sulfur gases hydrogen sulfide, carbonyl sulfide, mercaptans, thiols, sulfides and tetrahydrothiophen,

can be determined using a separation column containing porous polymer as stationary phase and a sulfur-selective detector, such as an electrochemical, potentiometric or flame photometric detector. In spite of the selectivity, high concentrations of hydrocarbons may influence the signal. For this reason it is necessary to separate the sulfur compounds from the hydrocarbon matrix. This is possible for example using the heart-cut technique of the Live-switching system. Different methods are available based on temperature-programmed packed columns and capillary columns; packed columns are still very common but capillary columns are becoming more and more popular because of their high resolution capability and the lower tendency of the sulfur compounds to adsorb on the surface. Adsorption is a general problem in sulfur gas analysis and much work is devoted to avoiding or to reducing mistakes caused by this phenomenon.

## On-Site Analysis

Historically, high-quality gas analyses in fuels required sophisticated laboratory instrumentation. However, fast, portable, micro-gas chromatographs



**Figure 4** Flame ionization detector chromatogram a capillary column system with Live-switching equipment. (A) Separation by porous polymer column: 1, carbon dioxide; 2, ethene; 3, ethyne; 4, ethane. (B) Separation by molecular sieve column: 5, carbon monoxide. (C) Separation by WCOT column: 6, methane; 7, ethane (ethene, ethyne); 8, propene; 9, propane; 10, *i*-butane; 11, *n*-butane; 12, neopentane; 13, *i*-pentane; 14, *n*-pentane; 15, 2,2-dimethylbutane and cyclopentane; 16, 2,3-dimethylbutane; 17, 3-methylpentane; 18, *n*-hexane; 19, benzene; 20, cyclohexane; 21, *n*-heptane; 22, toluene; 23, *n*-octane; 24, xylene.

now allow the on-site analysis of complex gas streams, such as the natural gas used in gas turbines. This has the potential to save time and reduce costs. Typical sample throughput time is 1–3 min and hence, even with replicate injections, results are available within 10 min. Precision is often better than with laboratory methods due to the higher level of automation. Micro-gas chromatographs also have a wide linear dynamic range, typically from 10 ppm to 100% for many analytes.

See also: **Gas Chromatography:** Gas–Solid Chromatography; Detectors; Petrochemical Applications. **Sulfur.**

## Further Reading

- De Zeeuw J (1991) Analysis of natural gas using a porous polymer-coated capillary column and temperature programming. *Liquid Chromatography – Gas Chromatography International* 4: 36–38.
- German Standard DIN 51872, Part 5 (1992) Testing of gaseous fuels and other gases; determination of the components; capillary gas chromatographic procedure.
- International Standard ISO 6568 (1981) (E) Natural gas – Simple analysis by gas chromatography.
- International Standard ISO 6974 (1984) (E) Natural gas – Determination of hydrogen, inert gases and hydrocarbons up to C<sub>8</sub> – Gas chromatographic method.



## Oil-Based

**M Rawdon**, Texaco Inc, Beacon, NY, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 1672–1679, © 1995, Elsevier Ltd., with revisions made by David Stiles.

### Introduction

In order to understand the chemistry of liquid fuels derived from oil (petroleum) it is first necessary to know a little about the crude oil from which these fuels originate. Petroleum is a highly variable, complex mixture of thousands of compounds. It can range from thin and watery to waxy solid, from brown to green to black, from ‘sweet’ low-sulfur oil to corrosive, foul-smelling ‘sour’ crude. Predominant among the compounds present are the hydrocarbons, compounds comprised solely of carbon and hydrogen. But a significant percentage of constituents in petroleum contain heteroatoms, most frequently sulfur and nitrogen. The amount and nature of these vary considerably with the type of petroleum; these will be mentioned specifically throughout this article where needed. The term ‘hydrocarbons’ is frequently used in the oil (and related) industry to refer to the compounds found in petroleum fractions whether heteroatomic compounds are present or not. Hydrocarbon fuels can also be obtained by refining non-petroleum crudes such as shale oil or liquid from tar, sand, or coal. Products derived from these materials, while important in a few localized markets and distinct in some ways, will not be included in the following discussion.

### Petroleum Refining

The compounds found in petroleum deposits cover a wide range of relative molecular mass, and the process of refining petroleum to produce usable fuels primarily serves to separate this complex mixture into narrower fractions (defined by boiling point, as described later) with properties suited to the end-use application. The lightest fraction is typically material that, while easily liquefied by pressure, is gaseous at ambient conditions. Called liquefied petroleum gas it will not be described in this section, but rather included in the discussion of gaseous fuels.

The next lightest fraction isolated at the refinery consists of the naphtha streams. These generally find their way into gasoline (petrol), either directly or after further treatment within the refinery to remove

impurities or to boost octane quality. Gasoline is comprised primarily of molecules containing four to 12 carbon atoms. Sulfur is the only significant heteroatom, and it is present as sulfide or mercaptan.

The gas oils are next, and they form the fuel for diesel engines. The carbon number is in the 10–20 range, and the sulfur level may be as high as 1 or 2% (m/m). In the gas oil boiling range sulfur compounds usually include thiophenes. Nitrogen and oxygen compounds are present in small amounts as well, and it is these polar compounds that give diesel fuel (gas oil) its color.

The heaviest fraction of petroleum that remains (the residuum or resid) may be treated in a number of ways, depending on the configuration of the particular refinery. Lubricant stocks may be distilled off under reduced pressure, and/or the residual material may be cracked to produce additional volume of lighter fuel components. The resid may be sold as heavy fuel oil, usually with the addition of a light ‘cutter’ stock (distillate) to improve handling, or it may be converted to petroleum coke for use as a solid fuel. The residual fraction contains virtually all of the organometallic compounds present in the starting crude. As these may determine the suitability of the residuum for further treatment or as a fuel, they are of significant analytical interest.

These products are discussed below in order of increasing boiling range (and thus carbon number), along with the analytical techniques used in their characterization. Most of the key properties that determine fuel performance are measured using physical parameters that often relate only indirectly to a particular fuel’s chemical composition. Since these physical properties comprise fuel specifications in virtually all industrialized countries, they are included in the discussion and the reader is referred to the appropriate standard-setting body for more details. Analytical testing generally serves to supplement the more routine characterization, and is typically needed to differentiate different batches of similar fuels.

### Gasoline (Petrol)

A fundamental and important property of gasoline is volatility. Volatility is described by vapor pressure and distillation, and these specifications may vary through the year and across geographical areas in order to match the volatility of the fuel to ambient conditions. Gasoline is typically used exclusively as a fuel for spark ignited internal combustion engines.

Since it is the fuel vapor that burns rather than the liquid, it is important that the fuel vaporizes easily. Ease of starting a cold engine depends primarily on the vapor pressure of the fuel. This is measured by the Reid vapour pressure (RVP) method and is primarily a function of the amount of butanes and pentanes present in the gasoline. At higher temperatures, as the engine is warming up, it is the mid-point of the distillation, the T50, which relates to smooth engine performance. Other distillation points, e.g., T10 and T90, may be controlled as well to ensure satisfactory vehicle operation. An elevated distillation endpoint may indicate contamination of the gasoline with heavier material. The distillation itself is performed using a procedure that, like all specification tests for petroleum products, is highly standardized by industry organizations such as the Institute of Petroleum and the American Society for Testing and Materials. These organizations should be consulted for details and further information on testing procedures discussed throughout this article. The distillation of petroleum products, while adequately repeatable, lacks the resolving power (theoretical plates) of fractional distillation, e.g., using a spinning band or packed distillation column. Nevertheless, it forms the basis for definition of the boiling range of the various petroleum products, as listed in Table 1.

A second key measure of a gasoline's performance is octane rating. As the gasoline vapor-air mixture is compressed within the cylinder by the rising piston, it heats up (from compression as well as by absorbing heat from the engine surfaces). If it gets hot enough it will spontaneously ignite. In a normally operating, spark-ignited engine the fuel burns as a result of the spark plug igniting the mixture. Spontaneous ignition ahead of the normal flame causes a metallic rapping noise called knock. Such abnormal combustion generates much higher than normal pressure within the cylinder and, in severe cases, can damage the engine. Octane rating is nothing more than an expression of a particular fuel's ability to withstand the tendency to

preignite. It is measured in a laboratory engine that has a movable cylinder, and thus variable compression ratio. A test fuel is bracketed by reference fuels of known octane values and the knock intensity is compared to determine the octane rating of the sample. Octane rating is one of the key parameters in determining the value and use of refinery streams.

Until the last few years, most gasoline contained organolead compounds to improve the fuel's octane rating. These antiknock additives are typically mixtures of tetramethyl lead, tetraethyl lead, and the mixed analogs thereof. While these components are highly toxic in neat form, their low cost and high effectiveness made them the additive of choice for boosting octane ratings. They can be determined individually by gas chromatography (GC) using a lead-selective detector (atomic absorption or emission). More commonly, however, total lead is determined by X-ray fluorescence and the results are expressed as tetraethyl lead equivalent.

As lead compounds are being phased down to reduce environmental lead pollution, other compounds are being called upon to replace the lost octane value. Aromatics have the highest octane ratings of naturally occurring petroleum constituents, and their use has increased significantly in recent years. They can be determined in gasoline by a number of techniques. The most widely used is the fluorescent indicator adsorption (FIA) technique. A desorption (dry column) separation is carried out on a long, slender silica gel column and the hydrocarbon types (saturates or paraffins followed by olefins followed by aromatics) are quantified by measuring the length of the column wetted by each. A mixture of fluorescent dyes is added to enable detection.

The method is rapid and fairly simple, but subject to variation in the quality of the silica gel employed. The lightest components in gasoline need to be removed from each sample in order to obtain the best results, and certain compound types can elute with the wrong group, for example, cyclic di-olefins elute with the aromatics. The method gives results in

**Table 1** Typical properties of some representative refined products

Product	Principal application	Boiling range (°C)	Vapor pressure <sup>a</sup>	Yield per barrel <sup>b</sup> (%)
Gasoline	Motor fuel	0–205	0.5–1 bar	50
Light distillate (kerosene)	Motor fuel, stoves, lighting	180–290	Less than 0.1 bar	5
Gas oil (diesel fuel)	Motor and boiler fuel	180–350	Less than 0.1 bar	20
Lubricating oil	Lubricant	> 400	Nil	5
Residual fuel	Motor and boiler fuel	> 600	Nil	10
Other products	Chemicals, waxes, etc.	Varies	Varies	10

<sup>a</sup> Reid vapor pressure.

<sup>b</sup> Average figures. May vary considerably for any particular refinery.

liquid volume per cent. Mass spectrometry can be used to determine aromatics in gasoline, giving information about what types of aromatic ring (benzene, indane, naphthalene) are present in addition to total aromatic content in weight per cent. GC with an aromatic selective detector (photoionization, mass spectrometer) can also be used once reasonable correction is made for detector response. A universal GC detector can be used with very high resolution columns, although this is the most tedious approach, requiring complete separation of the sample and extensive calibration. Alternatively, a polar column will retain aromatics for elution after the bulk of the paraffins. In this manner benzene can be made to elute after *n*-decane.

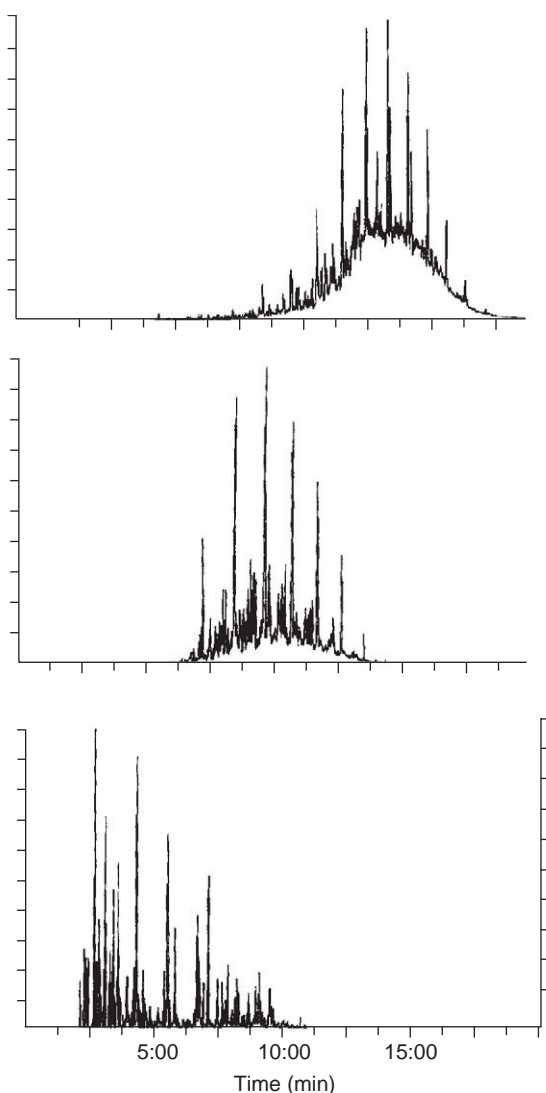
In addition to the octane available from the refinery streams, many gasoline blenders today rely on oxygenated compounds for providing the required octane rating. The lower relative molecular mass alcohols have high blending octane values and have been widely used. As a group they tend to degrade the gasoline's tolerance for water. Once the solubility limit is exceeded the alcohol can separate from the gasoline, with the resulting water-alcohol mixture being fed to the engine. Stalling or inability to start can result. Also, being small molecules, methanol and ethanol raise the RVP of the gasoline-alcohol blend. This is undesirable from an environmental viewpoint. Alcohols also have a higher latent heat of vaporization than hydrocarbons, and this can adversely affect the warming up performance of the engine. The other class of gasoline oxygenates is ethers. By far the most common is methyl tertiary-butyl ether. This is much more like gasoline in its water tolerance and vapor pressure behavior, which accounts for its widespread acceptance by the industry. Other ethers can be expected to emerge as oxygenates become more important contributors to the gasoline pool, e.g., ethyl tertiary-butyl ether and tertiary-amyl methyl ether. All these oxygenates can be determined using GC. A multiple column arrangement employing switching valves is the preferred technique. Use of an oxygen-selective detector such as the atomic emission or oxygen-flame ionization detector provides another option. An alternative approach, particularly for rapid screening of samples, is to use mid-infrared (IR) spectroscopy. A number of instruments are commercially available with preprogrammed calibrations for the common oxygenates.

Sulfur is generally present in gasoline at levels up to 0.3% (m/m). The preferred technique for sulfur determination is X-ray fluorescence, although other procedures may be employed. The X-ray procedure is rapid, sensitive down to  $10 \text{ mg kg}^{-1}$ , and calibration is straightforward. The sulfur K-alpha line at

53.73 nm is used, and no sample preparation is needed. Alternatively, the sample may be combusted and the sulfur oxidized to sulfate, which is then determined gravimetrically. While tedious, the technique does not require sophisticated instrumentation. The method is potentially less sensitive than X-ray fluorescence, and high aromatic levels can cause smoke during the combustion step, necessitating sample dilution. For trace sulfur levels (below  $100 \text{ mg kg}^{-1}$ ) oxidative microcoulometry may be used. Within the past few years governments have begun to introduce legislation that will have the effect of reducing the amount of sulfur in gasoline to less than 30 ppm. Most governments intend to reach this goal by 2005. At these reduced levels of sulfur content X-ray fluorescence is still a satisfactory method of sulfur analysis, but less costly and quicker analytical methods based on ultraviolet (UV) fluorescence are also standard.

Occasionally, it is important to characterize a gasoline (or a refinery stream used to blend gasoline) more completely than the typical properties described above. Examples might be: calculating physical properties from the individual components present, correlating the fuel composition with unburned hydrocarbons present in the engine exhaust, or predicting petrochemical end products from refinery conversion processes. In most cases GC is the technique of choice. The volatility and thermal stability of the hydrocarbons facilitate vapor-phase analysis, while few other techniques possess the resolving power needed to separate the hundreds of components found in a typical gasoline. As such, the modern petroleum research laboratory will be well equipped with GC instrumentation. Long, nonpolar capillary columns serve most purposes adequately, and provide a separation based primarily on boiling point. To accommodate the wide boiling range of most products, extensive use is made of temperature programming. **Figure 1** illustrates a number of different fuel products separated by GC according to boiling point. Notice how resolution of the individual components is lost in the heavier products. Selective detectors may be used to decrease interference from saturated hydrocarbons when functionalized or heteroatomic (generally sulfur) species are the target.

In many cases analysis by GC involves quite significant peak calibration and identification, only to be followed by extensive peak summing and averaging. Thus, GC is not the technique of choice for such tasks as predicting octane numbers. More often, spectroscopic techniques are correlated through sophisticated statistical treatment known broadly as chemometrics. Applications of near- and mid-IR



**Figure 1** Gas chromatographic separation by boiling point (solute volatility) of various fuel products: gasoline (bottom), kerosene (middle), diesel fuel (top). Note that heavier (higher boiling range) products elute at greater retention time. Prominent components are seen as discrete peaks while a complex mixture of unresolved isomers yields an elevated baseline.

spectroscopy as well as nuclear magnetic resonance spectroscopy are being established, which can outperform traditional testing methods. This can be expected to become more important in online process control, where the real-time results can be used to improve product quality and reduce follow-up specification testing.

## Diesel Fuel (Gas Oil)

Fuel for diesel engines is different from gasoline in several important respects. While gasoline is introduced to the combustion chamber of the engine as a

premixed fuel–air charge to be ignited by the spark plug, diesel fuel is injected into the hot, compressed air at roughly the time the piston is at the top of its stroke. With the exception of cold starting in those vehicles equipped with a glow plug, it is the heat of this compressed air that ignites the fuel. Vaporization, fuel–air mixing, and combustion all take place simultaneously. In contrast to gasoline, where the fuel's ability to resist ignition by heat alone (as indicated by octane number) is desired, the best diesel fuels ignite readily after injection. The term used to describe this is cetane number and, like octane number, it is determined in a special laboratory engine. Octane and cetane ratings are inversely related. Aromatic hydrocarbons have a high octane but a low cetane number. Straight chain paraffins are just the opposite. In addition to engine determination, cetane rating can be estimated using chemometrics or by a formula using density and distillation values. This latter procedure gives a value known as the cetane index.

Having a higher relative molecular mass than gasoline, diesel fuel can present a problem in cold weather. Straight chain paraffins, desirable from a cetane rating viewpoint, can precipitate at low temperature as wax crystals. These can plug fuel filters, starving the engine of fuel. They can also interlock in a rigid lattice causing the entire fuel volume to solidify. The temperature at which this occurs is called the pour point. Additives are often used which prevent the formation of the rigid lattice (but not the wax crystals themselves); these are called pour point depressants. They are typically ethylene–vinyl acetate copolymers, and can be determined using IR spectroscopy. Another low-temperature property is the cloud point. This is the temperature at which the first wax crystals can be seen forming. Somewhere between the cloud and pour points there is usually enough wax to plug a vehicle filter. Several test rigs have been developed to determine low-temperature operability of diesel fuels: cold filter plugging point, low temperature flow test, and so on. No test can completely predict field performance, as each vehicle has design features that affect its response to fuel wax.

The sulfur present in diesel fuel depends primarily on the nature of the crude from which it was refined and how extensively it was treated during the refining process. High sulfur crude, of course, yield higher sulfur distillate products. Cracking of heavy material will likewise tend to contribute to higher sulfur levels in the middle distillates, as the heavier portion of the crude contains proportionally more sulfur. Hydrotreating, the refining process by which hydrogen is added to the fuel and sulfur (or nitrogen) is removed, can be used to a greater or lesser degree



depending on the specifications of the desired end product. Sulfur content is a concern to the end user as a portion of the sulfur is converted to sulfuric acid in the crankcase lubricant (engine oil). Engine oils contain detergent and other material that neutralize this acid, but a low sulfur fuel places less stress on the lubricant and can lead to longer engine life. Additionally, a portion of the sulfur is converted to sulfate in the exhaust gas, which contributes to exhaust particulate levels. Since particulates are one of the regulated pollutants for diesel engines, newer low-emission engines require low sulfur fuel, typically 0.05% (m/m). Judicious crude selection combined with hydrotreating is used to produce such fuel.

Sulfur in middle distillates is determined using much the same methodology as for gasoline. X-ray fluorescence is the technique of choice. For best results standards should be prepared in a matrix that matches the sample type. A variety of combustion-based approaches are also available, though their applicability to low sulfur distillates may be limited owing to poor sensitivity.

Most middle distillate fuels tend to be comprised primarily of straight run distillate streams, as opposed to cracked stocks. Olefinic unsaturation is thus low. Aromatics, however, are most often in the 20–50% (v/v) range. Being dense, aromatics possess high-energy density measured in British thermal units per unit volume. This is an advantage to the consumer. However, aromatics also have an inherently low cetane value, a distinct disadvantage, particularly for ease of starting a cold engine. More recently, aromatics have been implicated as contributing to the generation of nitrogen oxides ( $\text{NO}_x$ ) in diesel engines. Surprisingly, only California has enacted legislation limiting the amount of aromatics allowed in diesel fuel, although most other western legislatures have stated their intention to follow suit. Since 1993, all diesel fuel sold for on-road use in California has been restricted to no more than 10% (v/v) aromatics. This level is far below current US averages, and one that requires extensive hydrotreating to reach. Given the investment required for the high-pressure processing units involved, some refiners have opted to abandon this market rather than comply. This establishes a new landmark in the role of base fuel parameters in diesel fuel marketing. Historically, diesel fuel has been bought and sold as a commodity with little attention to its composition.

As with gasoline, aromatics are most frequently determined using the FIA test, although this method was not designed for middle distillates (these cause slow development of the separation with attendant large variability). Nevertheless, it is readily available in most petroleum laboratories and has been widely

used. A newly developed technique is emerging that offers an alternative to FIA for aromatics determination. Supercritical fluid chromatography (SFC) is a hybrid of gas and liquid chromatography that uses high-pressure carbon dioxide to separate the fuel into hydrocarbon classes. Aromatics are eluted as a class (as in liquid chromatography) and detected with a flame ionization detector (as in GC). The method is more rapid than FIA and offers better precision as well. Mass spectrometry can be used to characterize the types of aromatics present. GC by itself is not as useful as for gasoline; the individual components are not well resolved making identification and calibration difficult. Analysis by liquid chromatography most often employs UV absorbance detection. This requires careful calibration before total aromatics can be measured. The separation is easily achieved using normal phase (silica) columns, however.

Other parameters of importance in diesel fuels are physical properties. Viscosity of these fuels is higher than gasoline, and influences the injection of the fuel into the combustion chamber. Lubricity, a somewhat related property, is a concern since the fuel also serves to lubricate the moving parts of the fuel pump and injectors. Inadequate lubricity can lead to scoring and wear. This is currently a subject of debate, as the naturally occurring sulfur compounds are thought to provide the film needed to keep highly loaded metal parts from coming into direct contact. The move to lower sulfur levels threatens to allow more pump wear.

## **Turbine Fuels**

Fuels for turbine engines are a special class of distillate fuels. By far the most common turbine engines are those of jet aircraft. Other applications can be found in power plants and aboard ships. These engines require a clean burning, economical fuel. Light distillate (kerosene) is almost exclusively used, and there can be much exchange and rebranding of products between aviation turbine fuel, light diesel fuel (typically used during cold weather and for two-cycle urban bus operation), and other distillate markets. The turbine blades are of special metallurgy, so residual material that could contain vanadium is to be avoided at all costs. Apart from special cleanliness and fuel handling considerations for aircraft fuel, analytical considerations of turbine fuel are the same as for other distillates.

## **Residual Fuels**

The heavy material that remains after refinery processing to remove the lighter material is frequently

sold as a fuel. Since it is essentially nonvolatile it is unsuitable for use in spark-ignited engines. It may be used in boilers, in which case it is pumped (sprayed) in continuously. Boiler fuel application is rather tolerant of fuel properties, with sulfur content and viscosity being the prime criteria. Stationary boilers, e.g., power plants, may be fitted with scrubbing equipment to remove sulfur gases from the stack gas. Pour point and viscosity of residual fuel is typically very high, with some fuels being a solid at ambient temperature. External heat may need to be supplied to keep the fuel sufficiently fluid for handling.

Alternatively, the residue may serve as fuel for medium- and low-speed diesel engines. These large engines are typically found in vessels and electric generating plants. In this application the fuel is injected intermittently to coincide with the start of the power stroke of the engine cycle. Straight resid (called bunker fuel) is often too heavy or viscous for optimum operation in these engines, so most fuel is blended with a distillate stock to a viscosity specification. One potential problem arising from the addition of such cutter stocks is the precipitation of asphaltenes. Several spot and screening tests have been developed to identify heavy fraction-cutter combinations that produce unstable fuel. Unfortunately, two stable fuels can become unstable when mixed together, depending on their asphaltene and solubility characteristics. The problem is most pronounced for vessel operation, where fuel may be obtained at ports located throughout the world and segregation of the fuel batches is not feasible. Instability leads to greater sludge burden on shipboard fuel handling systems (centrifuges, filters, and homogenizers). As these large engines are either kept running continually or are started with a lighter, distillate fuel, cold starting is generally not a major concern. As such, residual fuel is not routinely rated for ignition quality (cetane number). The distillate cutter stock typically provides for adequate auto-ignition of the fuel.

The trace elements present in the original crude are also concentrated in the residue, giving rise to several concerns. Residual fuel sulfur level is the highest of all oil-based fuels, and can be 3–5% (m/m). This has a direct effect on the acid loading of the crankcase oil, as discussed previously, and marine diesel lubricants often have very high alkalinity to control wear and corrosion. Vanadium is a common trace metal in petroleum, and its presence in the fuel can cause problems due to the formation of corrosive combustion products such as sodium vanadate and vanadium pentoxide. Vanadium is typically determined by ashing the sample followed by digestion of the residue in acid and quantification by atomic absorption or emission. Sodium also can contribute to corrosive attack; it may enter the fuel through contamination with seawater. It can often be determined at the same time as vanadium. Aluminum is a metal of concern in heavy fuel, as it indicates the presence of catalyst fines from refinery cracking operations. It is the abrasive nature of these small (less than 0.02 mm) particles that is the problem. Different catalysts contain widely different amounts of aluminum, so this is an imprecise as well as indirect measure of catalyst fines. Aluminum is determined in much the same way as vanadium, except that the ash is fluxed with lithium tetraborate prior to acid dissolution.

## Performance Additives

Gasoline and diesel fuel often contain additives to improve engine performance. Some examples of these are shown in Table 2. The most important class of performance additives is detergents/dispersants. While these perform the indicated task of minimizing engine deposits, they also serve as the primary means by which marketers differentiate their products. Very often products are distributed and exchanged as commodities between the point of

**Table 2** Some common gasoline additives

<i>Description</i>	<i>Function</i>	<i>Composition (typical)</i>
Antiknock	Boost octane value	Tetraalkyl lead, methyl cyclopentadienyl manganese tricarbonyl
Antioxidant	Minimize gum (residue)	Aromatic diamines, hindered alkyl phenol
Metal deactivator	Chelate trace metals	<i>N,N'</i> -Disalicylidene-1,2-propanediamine
Dye	Visual identification	Anthraquinones, azo compounds
Corrosion inhibitor	Prevent rust	Organic acids and salts
Detergent	Keep carburetor and injectors clean	Alkyl amines, amine phosphates
Dispersant	Keep entire intake system <sup>a</sup> clean	Polymeric <sup>b</sup> amines and succinimides
Dehazer	Coalesce suspended water droplets	Ethoxylated/propoxylated surfactants
Biocide	Prevent biological growth in storage tanks	Quaternary ammonium salts, boron compounds

<sup>a</sup> Carburetor, injectors, and intake valves.

<sup>b</sup> Poly(isobutenyl) or poly(butoxy).



manufacture and final sale to the consumer. As a result, the properties of the base fuel may be the same across several marketers. Addition of additives at the fuel terminal thus provides a simple and effective means of adding value to the product. A high-performance additive package can confer a competitive advantage upon a particular marketer; therefore, composition of these materials is often a closely kept secret.

Owing to their chemical dissimilarity to the matrix, the low dose at which they are used, and their strategic importance for brand identification, additives present one of the best opportunities for sophisticated analysis in the oil industry. When the analyst knows the chemical nature of the target compound, e.g., when he/she is developing a method to monitor an additive, the technique of choice may be obvious. When he/she is seeking to detect or identify an unknown material, as in a competitor's fuel, the situation is more difficult. In either case one will most likely want to remove the bulk of the gasoline, generally by evaporation. In fact, in some cases the weight of the residue after evaporation (the gum value) may be sufficient to indicate the presence of an additive. More likely spectral techniques can be applied to the residue to locate characteristic functional group absorbances. The polymeric backbone of most additives generally covers a range of relative molecular mass. This limits the usefulness of mass spectrometry, which otherwise would be valuable owing to the high sensitivity afforded by that tool. More often than not it is IR spectroscopy that provides the greatest amount of information on performance additives.

Separation of the evaporation residue by liquid chromatography provides fractions for subsequent analysis that are typically less complex than the whole residue. A typical scheme might employ silica gel or a similar polar adsorbent in a 10–20 mm diameter column. Elution of the sample with solvents of increasing strength thus serves to separate nonpolar components (mineral oil or hydrocarbon polymers such as polyisobutylene) from more highly functionalized components of the additive package. The range of relative molecular mass of many common additive classes makes their collection in a single fraction difficult, and trial and error is often needed to achieve the best separation. There is no standard practice for such analysis, and there is considerable 'art' involved. As with many techniques,

experience is the best teacher, and each laboratory probably has a preferred way of doing things, along with a library of spectra from previous work. Nevertheless, there is abundant opportunity for novel application of instrumental and microscale techniques, e.g., gel permeation and SFC followed by nuclear magnetic resonance spectroscopy or mass spectrometry. Finally, aspiring analysts should remember that additives, being commercial products, are rarely as simple as the patent or synthetic pathway would suggest. By-products are common, and these are often of higher relative molecular mass, e.g., dimers or further reaction products. The presence of these materials in minor amounts can often complicate an interpretation and skew calibration factors.

Additives in diesel fuel are generally similar to gasoline, being nitrogenous, hydrocarbon-soluble detergents and dispersants. The less volatile nature of the matrix can complicate the concentration step, but otherwise the general approach is similar to that for gasoline additives.

*See also:* **Distillation.** **Elemental Speciation:** Waters, Sediments, and Soils. **Fuels:** Gaseous. **Gas Chromatography:** Petrochemical Applications. **Infrared Spectroscopy:** Near-Infrared; Industrial Applications. **Nuclear Magnetic Resonance Spectroscopy:** Overview. **Sulfur.** **Supercritical Fluid Chromatography:** Applications. **X-Ray Fluorescence and Emission:** Wavelength Dispersive X-Ray Fluorescence; Energy Dispersive X-Ray Fluorescence.

## Further Reading

- American Society for Testing and Materials (1992) *Annual Book of ASTM Standards*, Section 5. Philadelphia PA: ASTM.
- Berger BD and Anderson KE (1981) *Modern Petroleum*. Tulsa, OK: Penn Well.
- Clark GH (1988) *Industrial and Marine Fuels Reference Book*. London: Butterworth.
- Institute of Petroleum (1990) *Standard Methods for Analysis and Testing of Petroleum and Related Products*. Chichester: Wiley.
- McManus TR (1991) Application review – petroleum and coal. *Analytical Chemistry* 63(12): 48R–64R.
- Owen K and Coley T (1990) *Automotive Fuels Handbook*. Warrendale, PA: Society of Automotive Engineers.
- Speight JG (1990) *Fuel Science and Technology Handbook*. New York: Dekker.

# FULVIC SUBSTANCES

*See* HUMIC AND FULVIC COMPOUNDS

# FUNCTIONAL GROUP ANALYSIS

**K Nakashima**, Nagasaki University, Nagasaki, Japan

**K Imai**, Musashino University, Tokyo, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

For functional group analysis, chemical reactions that are selective and specific to the analytes have been utilized. The products obtained after the reactions have properties that are desirable (sensitive and selective) for analysis. The reagents used for functional group analysis are required to have the two structural features: (1) a reactive moiety towards the target functional group; (2) a signaling moiety with a high detector response. To date, many reagents used for functional group analysis have been developed and used in combination with analytical instruments (e.g., gas chromatography, GC; liquid chromatography, LC; capillary electrophoresis, CE, etc.).

In this article, the representative analytical methods for several functional groups such as amino, hydroxy, aldehyde and ketone, carboxyl, thiol, and halogen groups are described as well as their application.

## Amino Group

Amines such as amino acids, polyamines, catecholamines, and peptides exist in most of the biological samples. In general, amino groups are so reactive that chemical reactions with the analytical reagents proceed under fairly mild conditions, and the numerous reagents for amines have been developed and applied to the highly sensitive determination of a large variety of analytes.

Primary amines and amino acids selectively react with carbonyl compounds such as benzaldehyde, furfural, and pentafluorobenzaldehyde in aqueous medium at room temperature, and give Schiff's bases in good yields. This reaction is used for the determination of aliphatic diamines and aromatic amines by GC with a flame photometric detector or an electron capture detector. However, a cleanup step for sample preparation is required to remove the excess reagent.

*o*-Phthalaldehyde (OPA) in the presence of certain reducing reagents, such as 2-mercaptoethanol or 3-mercaptopropionic acid, reacts with primary amines and amino acids to form highly fluorescent isoindole derivatives at room temperature. The detection limits for amino acids by an LC-ultra violet (UV) detection

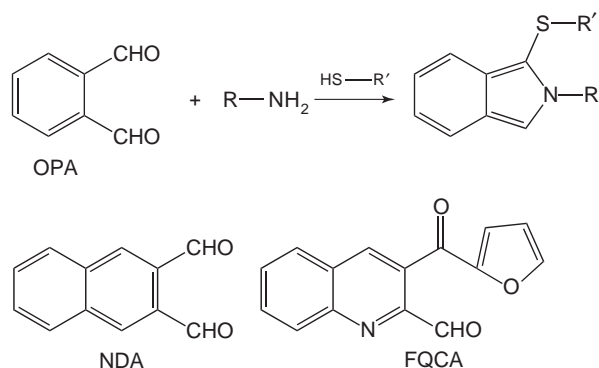
method are  $\sim 1$  pmol on-column at a signal-to-noise (S/N) ratio of 3. The reaction of OPA using a chiral thiol compound such as *N*-acetyl-L-cysteine as a reducing agent is important for the separation of chiral amines, especially amino acid enantiomers. OPA is to be applied to both pre- and postcolumn LC, because OPA itself is nonfluorescent. The resultant isoindole derivatives are also electroactive, and thus nanogram levels of amino acids can be determined with an electrochemical detector (ECD). Moreover, after postcolumn derivatization with OPA and 2-mercaptopropionic acid, primary amino acids are determined by liquid chromatography (LC)-electrospray ionization (ESI)-mass spectrometry (MS). The detection limits (20–250 fmol) are 100 times higher than those with a direct determination of amino acids by LC-ESI-MS. OPA derivatives of amino acids are rapidly separated by capillary zone electrophoresis (CZE) or micellar electrokinetic capillary electrophoresis (MEKC) and sensitively detected by laser-induced fluorescence (LIF) detection. 2,3-Naphthalenedialdehyde (NDA), an OPA analog, reacts with primary amines in the presence of cyanide ion to produce highly fluorescent cyanobenz[*f*]isoindole (CBI) derivatives. The CBI derivatives have also been found to be electroactive. The detection limits of CBI derivatives of amino acids with Fluorescence (FL) and ECD are 15–30 and 100 fmol, respectively. After derivatization with NDA, the amphetamine related compounds are determined by LC-peroxyoxalate chemiluminescence (PO-CL) detection with the detection limit of 0.2 fmol. OPA analogs, NDA and 3-(2-furoyl)quinoline-2-carbaldehyde (FQCA), are shown in **Figure 1**.

Fluorescamine under the alkaline conditions rapidly reacts with primary amines and amino acids to give fluorescent derivatives at room temperature. The advantageous features of this reaction are as follows: (1) fluorescamine is nonfluorescent; (2) fluorescamine can be hydrolyzed to the nonfluorescent product; (3) the reaction with secondary amines can form nonfluorescent derivatives, which allows selectivity to primary amines. For those reasons, fluorescamine can be applied to pre- and postcolumn derivatization of primary amino compounds with LC-FL or CZE-LIF detection.

2,4-Dinitrofluorobenzene is utilized for the analysis of primary and secondary amines and amino acids. The derivatives are determined by measuring absorbance at 330–360 nm. *N,N*-diethyl-2,4-dinitro-5-fluoroaniline is used in a similar manner and

applied to the LC–UV detection of these compounds. A precolumn derivatization with 2,4-dinitrofluorobenzene is adopted to improve the detection properties of amino alcohols and amino acids in LC–ECD. The dinitrophenyl group of the derivatives is appropriate for ECDs operated in the reduction mode.

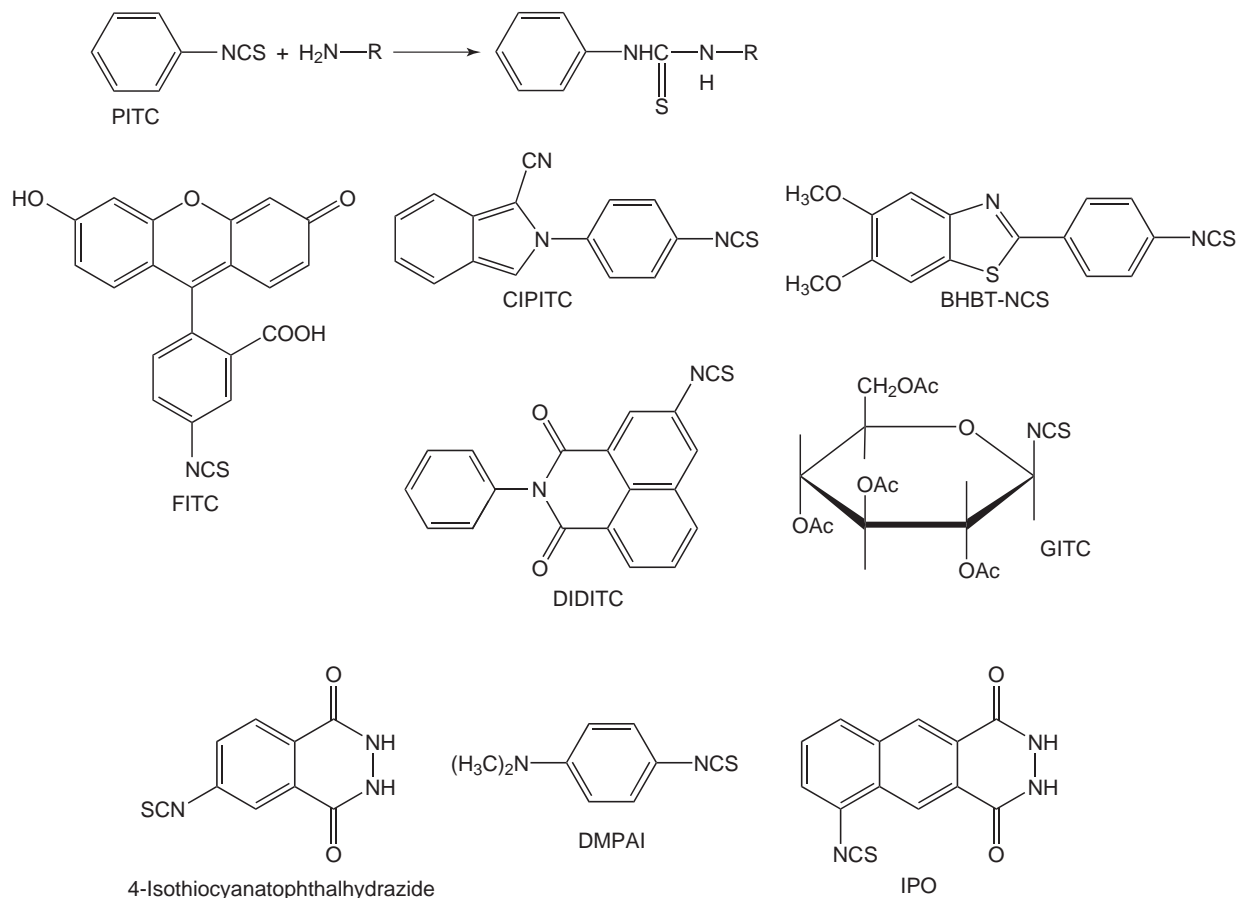
Isocyanates and isothiocyanates react with primary and secondary amines to form urea and thiourea



**Figure 1** The reaction scheme for primary amine with OPA and OPA analogs.

derivatives (**Figure 2**). Phenyl isothiocyanate (PITC) and naphthyl isothiocyanate are reported to be used as reagents for spectrophotometric detection. 2,3,4,6-Tetra-*O*-acetyl- $\beta$ -D-glucopyranosyl isothiocyanate (GITC) is used for enantioselective analysis of amino acid. These reagents can be applied to the precolumn derivatization of a variety of amines in LC and CE. PITC, as one of the Edman reagents, has been used for the sequential analysis of *N*-terminal amino acid. Usually, the derivatives are once cyclized and rearranged under the acidic condition to form 3-phenyl-2-thiohydantoin derivatives. The derivatives separated are detected at 254 nm by LC. Similarly, fluorescein isothiocyanate (FITC), 4-(*S'*,6'-dimethoxybenzothiazolyl)phenyl isothiocyanate (BHBT-NCS), 4-(2-cyanoisindolyl)phenyl isothiocyanate (CIPITC), and 5-isothiocyanato-1,3-dioxo-2-*p*-tolyl-2,3-dihydro-1*H*-benzo[*d,e*]isoquinoline (DTDITC) have been developed and applied to LC–FL methods (**Figure 2**).

Luminol derivatives, 4-isothiocyanatophthalhydrazide and 6-isothiocyanobenzo[*g*]phthalazine-1,4-(2*H*,3*H*)-dione (IPO) are synthesized as highly sensitive chemiluminescent reagents (**Figure 2**). Reactions with



**Figure 2** The reaction scheme for primary amine with PITC and analytical reagents containing isothiocyanate group.

primary and secondary amines proceed at 80°C for 10 min in the presence of triethylamine as a catalyst. The detection limits for primary and secondary amines are in the range of 30–120 and 0.8–3 fmol on-column ( $S/N = 3$ ), respectively. *p*-*N,N*-Dimethylaminophenylisothiocyanate (DMAPI) reacts with amino acids to give phenylhydantoin derivatives. After oxidation at pH 2, the derivatives are determined by LC–ECD with the detection limits of 0.5–1.0 ng.

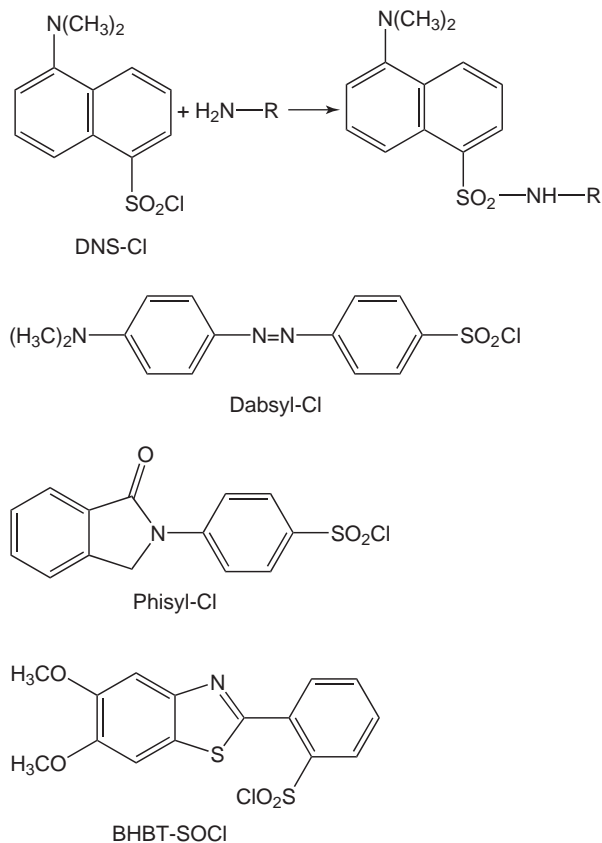
The analytical reagents having sulfonyl chloride as a functional group can react with secondary and primary amines and amino acids under the weak alkaline condition to form sulfonamide derivatives. 1-Dimethylaminonaphthalene-5-sulfonyl chloride (DNS-Cl), 4-*N,N*-dimethylaminoazobenzene-4'-sulfonyl chloride (Dabsyl-Cl), 4-(*N*-phthalimidyl)benzenesulfonyl chloride (Phisyl-Cl), and 2-(5',6'-dimethoxybenzothiazolyl)benzenesulfonyl chloride (BHBT-SOCl) react in a similar manner and can be applied to the LC–FL or CE–LIF method (Figure 3). As these reagents and their hydrolyzates are fluorescent, they should be used as precolumn reagents. The analyses of amines and amino acids with DNS-Cl

have been widely performed in combination with various instruments. For example, DNS-amino acids are separated by CZE followed by detection of absorbance at 254 nm. Although DNS-amino acids are fluorescent by nature, an aqueous condition in CE results in a marked decrease of their quantum yields. A highly sensitive determination of DNS-amino acids with detection limit at the subfemtomole level is achieved by LC–PO–CL with a microbore column.

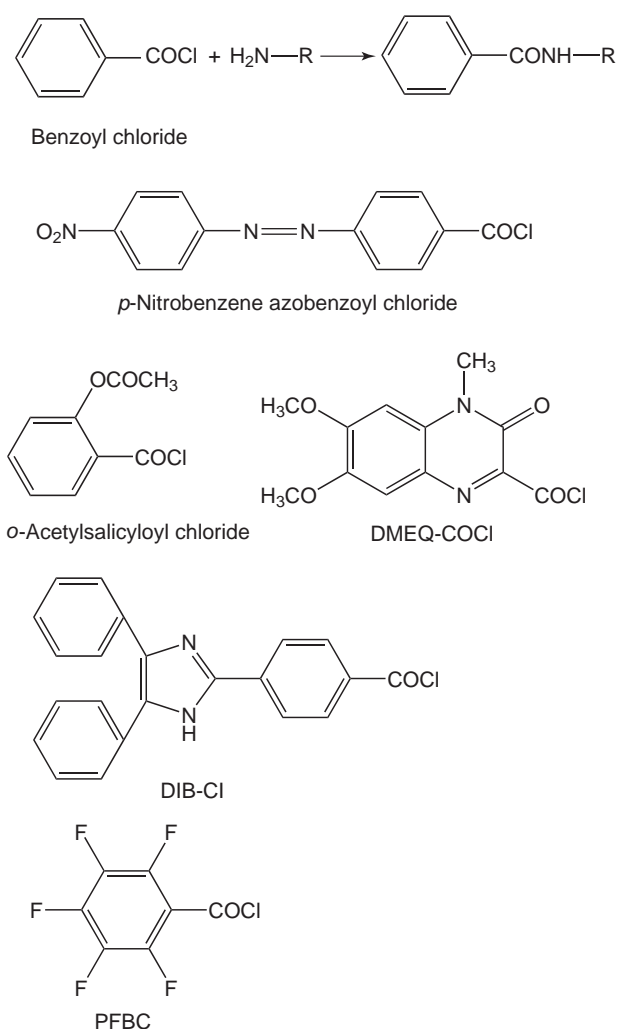
An acyl chloride such as benzoyl chloride can react with secondary and primary amines and amino acids to form *N*-substituted benzamide derivatives. Analytical reagents used for precolumn derivatizations are *p*-nitrobenzene azobenzoyl chloride for LC–UV, *o*-acetylsalicyloyl chloride for LC–ECD, and 3,4-dihydro-6,7-dimethoxy-4-methyl-3-oxoquinoxaline-2-carbonyl chloride (DMEQ–COCl) and 4-(4,5-diphenyl-1*H*-imidazol-2-yl)benzoyl chloride (DIB-Cl) for LC–FL, respectively. Pentafluorobenzoyl chloride (PFBC) is used as an acylating reagent for GC with electron-capture detection (Figure 4). Many of those described above can also react with hydroxyl and phenol groups.

The analytical reagents having an *N*-succinimidyl group as a reactive functional group can easily react with primary and secondary amines. 6-Aminoquinolyl-*N*-hydroxysuccinimidylcarbamate (AQC) can rapidly react with amines to form stable fluorescence derivatives. *N*-Succinimidyl-1-naphthylcarbamate (SINC) and *N*-succinimidyl-1-fluorenylcarbamate (SIPC) immediately react with analytes in a few seconds, and thus are suitable for LC–FL with an automated precolumn derivatization after removing the excess reagent by hydrolysis. Luminarin 1, having quinolizino coumarin structure, is used for precolumn labeling of amino acids in LC–PO–CL, and the labeled compounds can be detected as low as ~100 fmol on-column. *N*-(4-Aminobutyl)-*N*-ethylisoluminol (ABEI) is known as a chemiluminescent reagent for amines. The derivatization of amines with ABEI is performed as follows: initially, ABEI reacts with *N,N'*-disuccinimidyl carbonate (DSC) to form the intermediate, ABEI–DSC, which reacts with amines to give the final derivatives in the presence of triethylamine as a catalyst. The succinimidyl types of analytical reagents are shown in Figure 5.

Halogenobenzofurazans such as 4-fluoro-7-nitro-2,1,3-benzoxadiazole (NBD-F) and 4-chloro-7-nitro-2,1,3-benzoxadiazole (NBD-Cl) work on amines rapidly under alkaline conditions. As the hydrolyzed products of these reagents also fluoresce intensely, they are used only for precolumn derivatization. 4-(*N,N*-Dimethylaminosulfonyl)-7-fluoro-2,1,3-benzoxadiazole (DBD-F) is a similar sensitive reagent



**Figure 3** The reaction scheme for DNS-Cl with amine and analytical reagents containing sulfonyl chloride group.



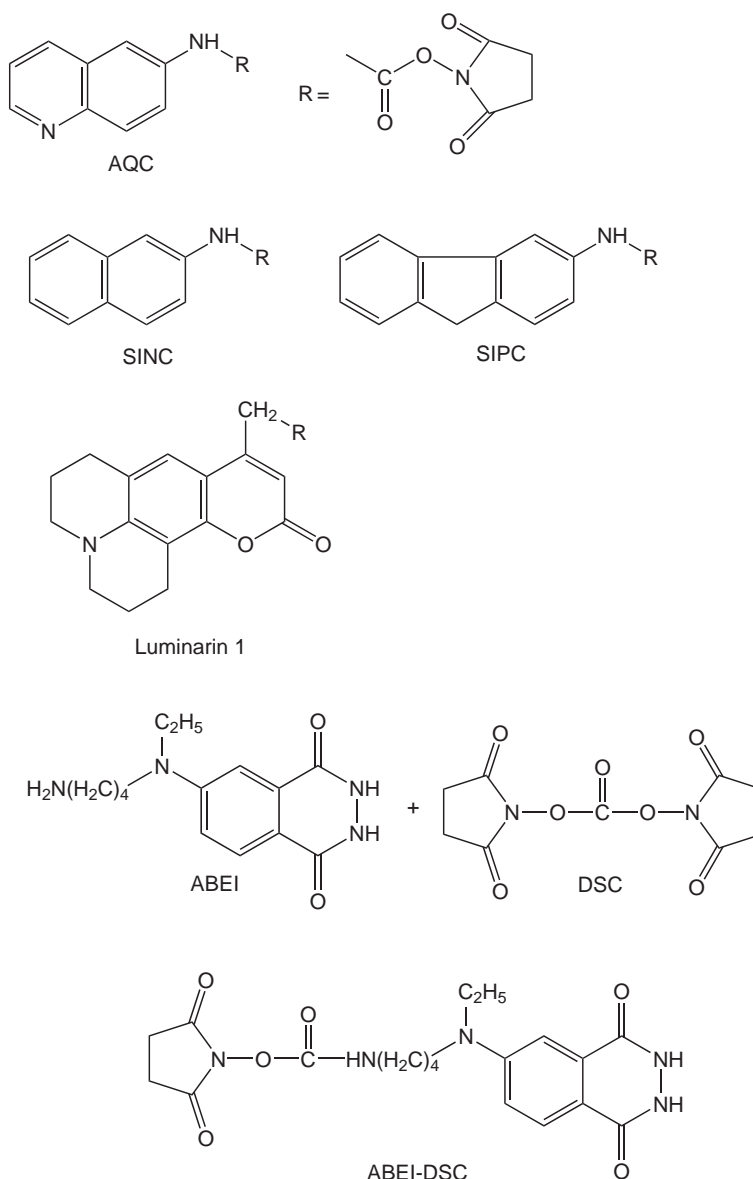
**Figure 4** Analytical reagents with acyl chloride for amines.

to NBD-F. DBD-labeled amino acids were separated on a reversed phase column and detected at femtomole levels by PO-CL (**Figure 6**). Halogenobenzofurazans can also react with hydroxyl and phenolic moieties.

Acyl amides (**Figure 7**) such as *N,O*-bis(trimethylsilyl)acetamide (BSA) and *N,O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA) have been used as silylating reagents for GC. Amino compounds can be converted to silyl derivatives to increase volatility. The reactivity of these reagents to primary amine is higher than that of secondary amine. On the other hand, *N*-methyl bis(trifluoroacetamide) (MBTFA) has been used as an acylating reagent. This reagent is very volatile and the by-product accompanied, *N*-methyltrifluoroacetamide, does not cause column damage. The reagent reacts not only with amines, but also with alcohols, phenols, and thiols.

Chloroformates are useful analytical reagents for primary, secondary, and tertiary amines. The reaction of these reagents with amines can be performed easily, and the resultant carbamate derivatives are suitable for GC analysis. The carbamate derivatives obtained from primary and secondary amines are determined selectively and sensitively by GC-MS and -ECD. After dealkylation, tertiary amines are derivatized to carbamates in the same manner for primary and secondary amines. The carbamate derivatives obtained with pentafluorobenzyl chloroformate are determined with high sensitivity by GC-ECD. Chloroformates also react with phenols, thiols, imidazoles, as well as amines.

Fluorenylmethyl chloroformate and its analogs (2-naphthyl chloroformate (NT-COCl), 9-fluorenylmethyl chloroformate (FMOC-Cl), and 2-(1-pyrenyl)ethyl chloroformate (PE-COCl)) are used as precolumn derivatization reagents for HPLC.



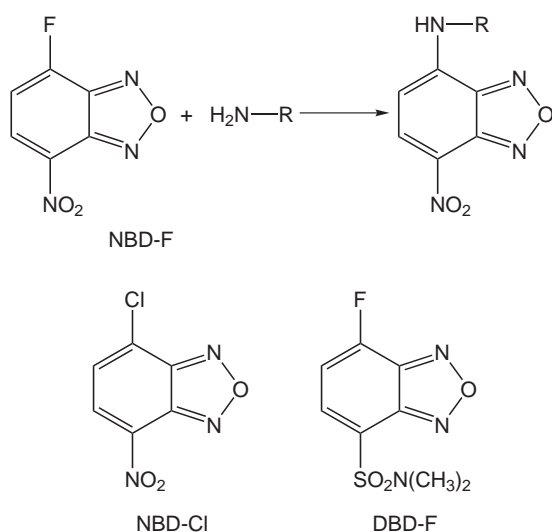
**Figure 5** Succinimidyl analytical reagents for amines.

Tris(2,2'-bipyridyl)ruthenium(II) ( $\text{Ru}(\text{byp})_3^{2+}$ ) is used as the electrogenerated CL (ECL) reagent for the amino compounds in the LC-CL detection system. The CL reaction is supposed to proceed as follows: initially, the oxidation of  $\text{Ru}(\text{byp})_3^{2+}$  to  $\text{Ru}(\text{byp})_3^{3+}$  is performed at the electrode surface, and then  $\text{Ru}(\text{byp})_3^{3+}$  reacts with amines to emit light at 620 nm. The ECL intensity is proportional to the amount of amines, and the order of increasing CL intensities is tertiary > secondary > primary amines. Dansyl derivatization of primary, secondary, and tertiary amines are used to increase the sensitivity. DNS-Glu can be determined by ECL with the detection limit of 2 pmol.

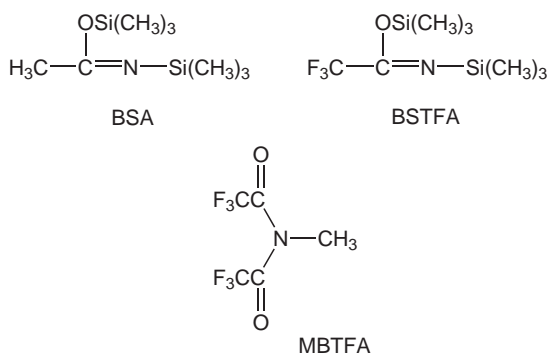
## Hydroxy Group

Hydroxy compounds such as alcohols, carbohydrates, steroids, and phenols are known to play important physiological roles. Although many analytical reagents have been developed for determination of hydroxy compounds, some of them can be also applied to amines as described in the previous section (e.g., isocyanates, sulfonyl chlorides, acyl chlorides, halogenobenzofurazans, acyl amides, and chloroformates). The derivatization conditions for hydroxy compounds are relatively more drastic than those for amines. By using fairly different conditions from those for amino compounds,





**Figure 6** Reaction scheme for halogenobenzofurazan and its analogs.



**Figure 7** Acyl amide reagents for amines.

selective analyses for hydroxy compounds can be achieved.

Carbonyl azides react with alcoholic hydroxy compounds to form the corresponding urethanes. In the same manner, ferrocenoyl azide, and 3-ferrocenylpropionyl azide can be derivatized, hydroxysteroids and the resultant derivatives are determined electrochemically by LC with the detection limits at subpicomole levels. 2-[2-(Azidocarbonyl)ethyl]-3-methyl-1,4-naphthoquinone (AMQ) reacts with primary and secondary alcohols to produce carbamic acid derivatives and is used in LC with ECD and FL detections. 7-Methoxycoumarin-4-carbonyl azide (7-MC-4-CON<sub>3</sub>) and 3,4-dihydro-6,7-dimethoxy-4-methyl-3-oxoquinoline-2-carbonyl azide (DMEQ-CON<sub>3</sub>) are also applied to LC with FL detection as precolumn reagents.

The reaction of carbonyl nitrites with primary or secondary alcohols in the presence of basic catalyst is

utilized for the analysis of alcohols. 1-Anthrionitrile (1-AN), pyrene-1-carbonylnitrile (PCN), and 4-dimethylamino-1-naphthoylnitrile (DNN) have been reported on their applications in LC-FL analysis.

Phenols react with 4-aminoantipyrine to give indophenols in the presence of hydrogen peroxide and peroxidase under weak alkaline conditions. The resultant indophenols are determined by absorptiometry. Potassium hexacyanoferrate (III) can also be used as an oxidant. As the indophenols have electrochemical activity, this reaction is utilized for precolumn derivatization for LC-ECD.

## Aldehyde and Ketone Groups

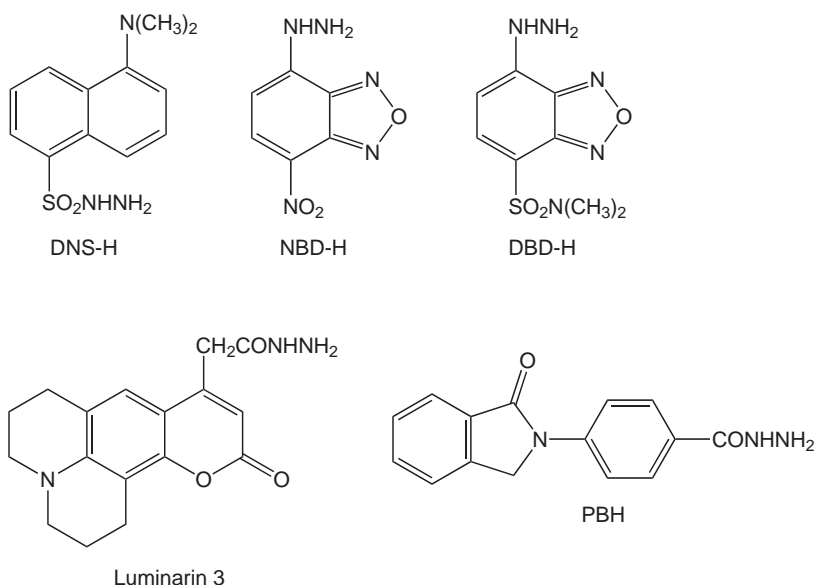
Compounds containing aldehyde and ketone groups in their molecules are major constituents of food aromas, and are widely distributed in the environment. Especially carbonyl compounds in the air are considered to exhibit toxicity.

In this reaction, hydrazones are efficiently formed between carbonyl compounds and hydrazino compounds under acidic conditions (Figure 8). For instance, 2,4-dinitrophenyl hydrazine is used as a colorimetric reagent for carbonyl compounds. Hydrazone produced from the reaction of 2,5-dihydroxybenzohydrazide (DBH) with dehydroepiandrosterone can be electrochemically detected at a low oxidative potential with detection limit at subpicomole levels. 5-*N,N'*-Dimethylaminonaphthalene-1-sulfonylhydrazide (DNS-H), 4-hydrazino-7-nitro-2,1,3-benzoxadiazole (NBD-H), 4-(*N,N*-dimethylaminosulfonyl)-7-hydrazino-2,1,3-benzoxadiazole (DBD-H), 1*H*,5*H*,11*H*-[1]benzopyrano[6,7,8-*ij*]quinolizine-9-acetic acid, 2,3,6,7-tetrahydro-11-oxohydrazide (Luminarin 3), and 4-(2-phthalimidyl)-benzohydrazide (PBH) are used as fluorescence labeling reagents and have been successfully applied to LC determination of carbonyl compounds in biological samples.

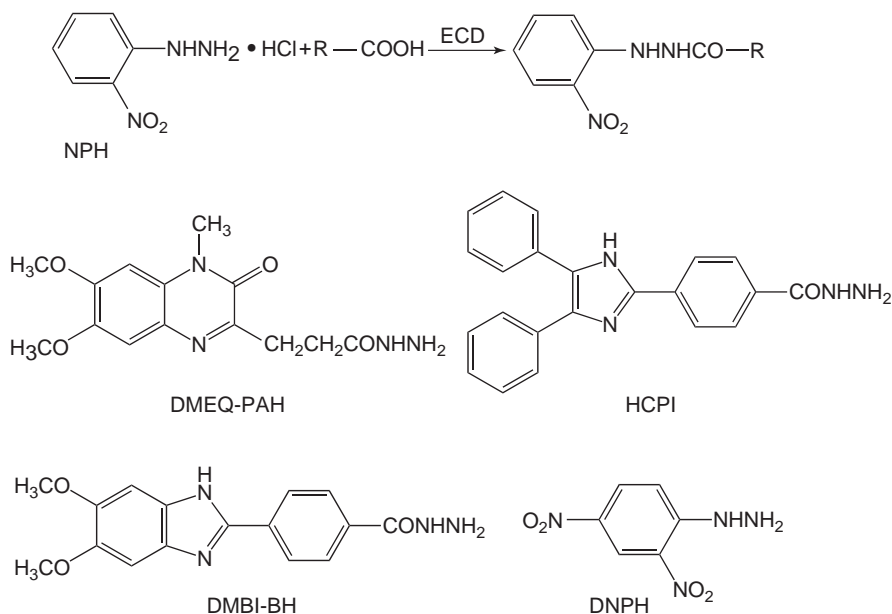
DNS-H can be used in combination with PO-CL, and its application study on the determination of oxo-steroids and oxo-bile acid ethyl ether by LC has been reported. The detection limits for hydrazone derivatives are at a level of a few femtomoles.

## Carboxy Group

There are many pharmaceutical preparations and biological compounds containing carboxylic group, i.e., fatty acids, prostaglandins, steroids, bile acids,  $\alpha$ -keto acids, and glucuronic acid. These compounds play important physiological roles to achieve homeostasis.



**Figure 8** Analytical reagents with hydrazine for aldehyde and ketones.



**Figure 9** Reaction scheme for NPH with carboxylic acid and hydrazine analytical reagents.

Hydrazino compounds can react with carboxy compounds in aqueous medium to form hydrazide. For the reaction to proceed under mild conditions, *N*-ethyl-*N'*-(3-dimethylaminopropyl) carbodiimide (EDC) is required as a condensing reagent. Fatty acids are determined with 2-nitrophenylhydrazine hydrochloride (NPH) by LC-UV detection. The detection limits for long chain fatty acids are 2.5–5 pmol per injection. As FL derivatization reagents, 6,7-dimethoxy-1-memyl-2(1*H*)-quinoxalinone-3-propionyl carboxylic

acid hydrazide (DMEQ-PAH), 2-(hydrazinocarbonylphenyl)-4,5-diphenylimidazole (HCPI), and 4-(5,6-dimethoxy-2-benzimidazol)benzohydrazide (DMBI-BH) are developed (Figure 9). Hydrazides can be detected at the femtomole level by LC-FL detection. DMEQ-PHA and HCPI are applied for the determination of fatty acids by PO-CL detection with the detection limits in the range of subfemtomole to femtomole. 2,4-Dinitrophenylhydrazine (DNPH) has been used as a precolumn reagent for LC-ECD.

Reagents having bromomethyl group reacts with carboxy compounds in acetonitrile or acetone to form the corresponding esters. Crown ether and potassium ion, or organic bases such as triethylamine and ethylamine, catalyze the reaction. 4-Bromomethyl-7-methoxycoumarin (Br-MMC) is a typical reagent. Br-MMC has been widely used for the derivatization of fatty acids, dicarboxylic acids, prostaglandins, bile acids, and glucuronic acids. Detection limits reported are in the range of picomole by UV detection to femtomole level by LIF detection. Other analytical reagents having a bromomethyl moiety are 4-bromomethyl-6,7-dimethoxy-1-methyl-2(1*H*)quinoxalinone (Br-DMEQ), 3-(bromoacetyl)-7-diethylaminocoumarin (Br-DAC), and 4-bromomethyl-6,7-dimethoxycoumarin (Br-MAC).

Reagents having diazoalkyl group react with carboxylic acids to form the corresponding ethers (Figure 10). 9-Anthryldiazomethane (ADAM) has been utilized for the LC–FL determination of carboxylic acids in biological samples. In spite of the instability even at  $-10^{\circ}\text{C}$ , ADAM has been applied for the determination of many biological carboxylic acids such as fatty acids, prostaglandins, oxalic acids, amino acids, and carnitines. Picomole level detection can be achieved with fluorescence detection. As similar type of reagents, 1-pyrenyldiazoraethane (PDAM) and 4-diazomethyl-7-methoxycoumarin (DAM-MC) have also been synthesized.

*p*-Methoxyaniline and 1-naphthylamine (1-NA) are used for the determination of fatty acids. The resulting derivatives are separated by RP-LC, and detected with UV absorption. Fluorescence derivatization reagents *N*-(1-naphthyl)ethylene diamine (NEDA) and 2-[*p*-(5,6-methylenedioxy-2*H*-benzotriazol-2-yl)]phenethylamine (MBPA) have also been developed.

The reagents containing hydroxy group have been developed for FL derivatization of carboxylic acid or other activated acids. 1-Pyrenemethanol (1-PM), 4-hydroxymethyl-7-methoxycoumarin (HMC), and 5-(4-pyridyl)-2-thiophenemethanol (PTM) have been applied to carboxylic acids with LC–FL detection.

Trifluoromethanesulfonate derivatives such as 2-(2,3-naphthalimido)ethyl trifluoromethanesulfonate (NE-OTf) and 2-(2,3-antrathenedicarboxyimide)ethyl trifluoromethanesulfonate (AE-OTf) can be used for LC determination of carboxylic acids. Thermolabile carboxylic acids in mouse brain are determined by NE-OTf with the detection limits of hundreds of femtomoles by UV detection, and a few femtomoles by FL detection.

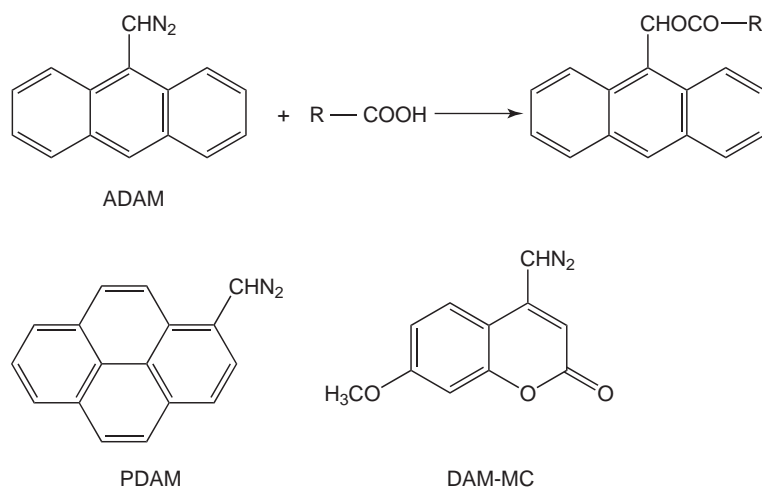
Many benzofurazan reagents have been synthesized for labeling of carboxylic acids, e.g., 4-(aminosulfonyl)-7-*N*-(2-aminoethylamino)-2,1,3-benzoxadiazole (ABD-AE), 4-(*N,N*-dimethylaminosulfonyl)-7-carverino-2,1,3-benzoxadiazole (DBD-CD), 4-nitro-7-(2-carbazoylpyrrolidine-1-yl)-2,1,3-benzoxadiazole (NBD-ProCZ), etc.

## Thiol Group

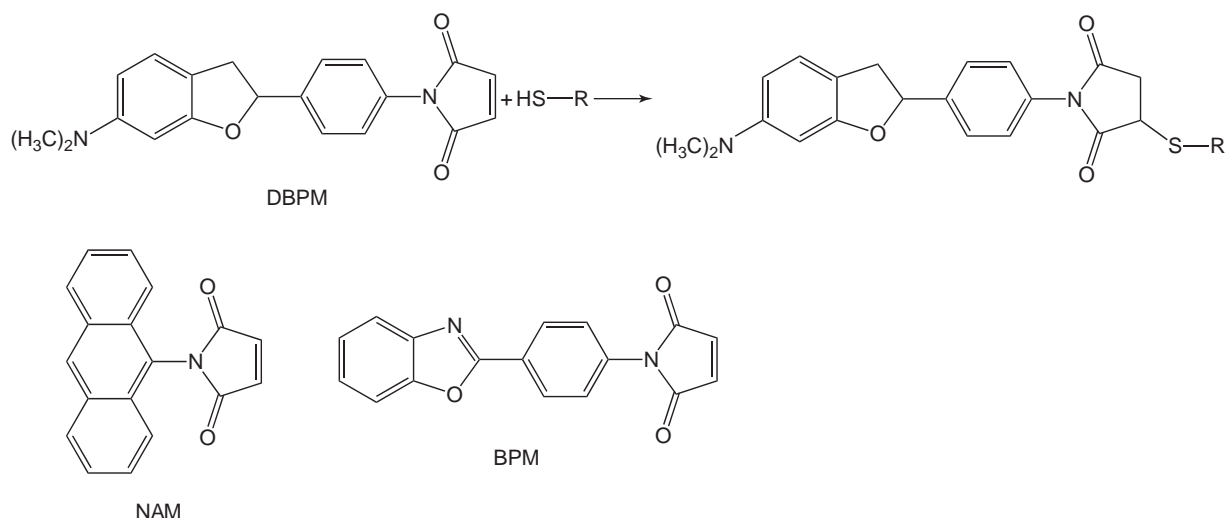
Thiol group-containing compounds are widely distributed in industrial, urban waste, and biological materials. The analysis of thiol contents in biological molecules gives valuable information in pharmacodynamic studies of drugs and in diagnosis of some diseases.

As selective reagents for thiol group, *N*-substituted maleimides, active halogen, and aziridine have been reported.

Maleimides easily react with thiols at pH 5–8 as shown in Figure 11. The typical analytical



**Figure 10** Reaction scheme for ADAM with carboxylic acid and diazoalkyl analytical reagents.



**Figure 11** Reaction scheme for DBPM with thiol compound and maleimides analytical reagents.

maleimide-type reagents *N*-[*p*-(2-benzoxazolyl)phenyl]maleimide (BPM), *N*-[4-(6-dimethylamino-2-benzofuranyl)phenyl]maleimide (DBPM), and *N*-(9-acridinyl)maleimide (NAM) have been developed. These reagents are mainly used as precolumn reagents for LC–FL detection. The advantageous features of these reagents are that they do not have significant fluorescence properties themselves, but after reaction with thiols the derivatives show high fluorescence. The sensitivities of these reagents are in the range of picomole levels for thiols. However, as the products produced are not stable and easily convert to fluorescent ring-cleaved compounds, two peaks sometimes appear in the chromatogram. DBPM can be applicable for the determination of six kinds of thiol-containing compounds in biological fluids. The derivatives are detected sensitively by LC with FL and PO–CL detections. *N*-(4-Anilinophenyl)maleimide (APM) and *N*-(ferrocenyl)maleimide, which have favorable electrochemical properties, are synthesized as analytical reagents for LC–ECD.

Halogenobenzofurazan, e.g., 7-fluoro-2,1,3-benzoxadiazole-4-sulfonate (SBD-F), 4-(aminosulfonyl)-7-fluoro-2,1,3-benzoxadiazole (ABD-F), and 4-(*N*,*N*-dimethylaminosulfonyl)-7-fluoro-2,1,3-benzoxadiazole (DBD-F), having reactive fluorine groups are synthesized. The application studies of these reagents on the determination of cysteine, GSH, and other thiol compounds by LC–FL detection have been reported with detection limits at the subpicomole level.

Bimane derivatives containing active halogen react with thiol compounds to form fluorescence thioethers. This reaction proceeds under relatively mild conditions, i.e., pH 8 at room temperature in 5 min. Monobromobimane and monobromotrimethylaminobimane as precolumn derivatization reagents

for biological thiols are used in LC with FL detection. The resultant derivatives from CoA, acyl-CoA, and *N*-acetylcysteine can be detected at the picomole level.

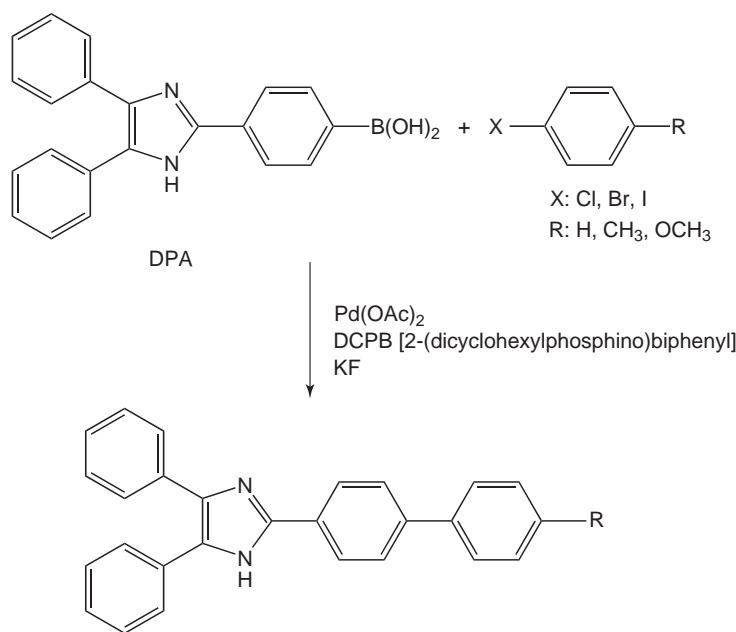
Reagents having aziridine group selectively reacts with thiol compounds to produce thioethers, which accompany with the cleavage of the aziridine ring. 5-Dimethylaminonaphthalene-1-sulphonylaziridine is used as a precolumn derivatization reagent. The determination of amino acids containing thiol group and penicillamine has been achieved with detection limits at the picomole level.

OPA reacts with thiols and amines to form highly fluorescent isoindole compounds as described in the section on ‘Amino group’. The analytical methods using OPA have been reported for pre- and post-column derivatization of amino acids. These methods are very sensitive and afford detection as low as 25 fmol.

Ellman’s reagent, 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB), produces 2-nitro-5-sulphydrylbenzoic acid by the reaction with thiols, which shows visible absorption at ~412 nm. Acetyl-CoA thioester is separated by ion-pair LC followed by conversion to thio-CoA with a postcolumn immobilized enzyme column reactor of phosphotransacetylase. Thio-CoA thus liberated can be determined spectrophotometrically after reaction with Ellman’s reagent with the detection limit of 0.05 nmol.

## Halogen Group

Organic compounds containing aryl halides are widely spread in the environment, foods, and biological materials, and is a serious health concern.



**Figure 12** The reaction for aryl halides with DPA.

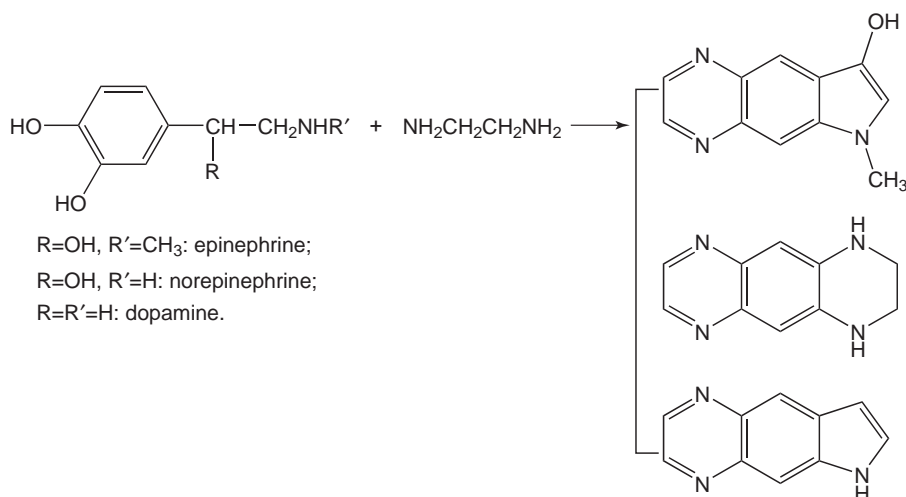
In general, aryl halides can be directly analyzed by GC-ECD or LC-UV detection. GC methods have higher performance in separation and detection than that achieved by LC methods. Many practical applications are attained by GC methods. However, there are few reagents for selective and sensitive determination of aryl halides. Recently, an anticancer drug, prosopidin, was determined in a biological sample by CZE after labeling with diethyldithiocarbamic acid (DDTC). The derivative is detected at 254 nm with the detection limit of 5 fmol on-column. However, disadvantages of this method are that  $\sim 10$  times excess of reagent and  $\sim 1.5$  h of reaction time are required to complete the derivatization. More recently, the derivatization method for aryl halides with a fluorescent arylboronic acid, 4-(4,5-diphenyl-1H-imidazol-2-yl)phenylboronic acid (DPA), based on Suzuki's coupling reaction, was developed. This reaction proceeds in the presence of palladium(II) acetate, 2-(dicyclohexylphosphino)biphenyl, and potassium fluoride in dioxane (**Figure 12**). The derivatives are separated by LC and detected fluorometrically at 410 nm with excitation at 320 nm. The detection limits for aryl halides are at sub- and low-picomole levels.

## Applications

### Determination of Catecholamines and Their 3-O-Methyl Metabolites

Catecholamines (CAs) are known to be important biogenic amines and are present in extremely small

amounts in human body as neurotransmitters or hormones. A highly selective and sensitive reagent is required for the determination of CAs in biological samples, where many of their precursors and metabolites coexist. An LC-ECD method is often used for the determination of CAs. Although the sensitivity of this method is enough to determine CAs in biological samples, its selectivity and stability are not satisfactory. On the other hand, the ethylene diamine (ED) method has been reported as highly selective for CAs (**Figure 13**). The reaction is suited for the simultaneous detection of CAs, because all the CAs are reactive with ED to yield strong fluorescent compounds. The derivatization method for CAs with ED followed by FL detection can be used as a good alternative for the direct amperometric detection of the catecholic compounds. Recently, a highly sensitive method based on the reaction described above for simultaneous determination of CAs and their 3-O-methyl metabolites was developed by utilizing the PO-CL reaction and semi-microcolumn LC. The automated LC system consisted of four LC pumps, an autosampler, a rotary six-way valve, a precolumn, an analytical column C18 ( $250 \times 1.5$  mm, ID), a coulometric cell, a reaction coil ( $0.25$  mm (ID)  $\times 6.5$  m) in a thermostatically controlled bath ( $80^\circ\text{C}$ ), and a chemiluminescence detector. ED as a fluorogenic reagent and a mixture of TDPO,  $H_2O_2$ , and TFA as a chemiluminescence reagent are used. By this method, CAs and their 3-O-methylmetabolites in mouse plasma of a very low volume of  $15\ \mu\text{L}$  can be determined. The quantification limits for CAs (norepinephrine, epinephrine, and dopamine) and



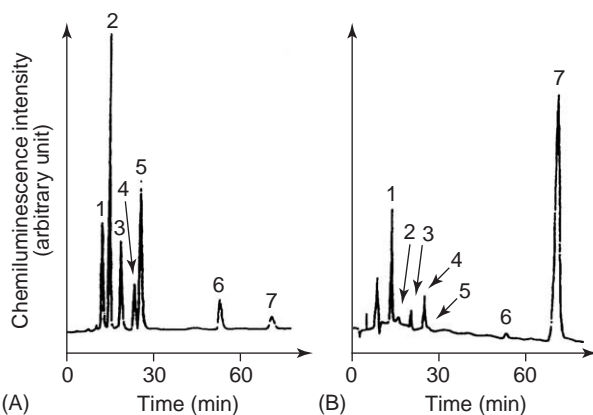
**Figure 13** Reaction scheme for catecholamines with ethylene diamine.

their metabolites (normetanephrine, metanephrine, and 3-methoxytyramine) are in the range of subpicomole levels.

#### Determination of Polyamines Based on Intramolecular Excimer-Forming Fluorescence Derivatization

Polyamines having two to four amino groups occur in all eukaryotic cells and play important roles in cell growth and differentiation. These compounds can be determined with analytical reagents for amines described above. A unique approach of fluorescence derivatization method for polyamine has been reported. This method is based on an intramolecular excimer-forming fluorescence derivatization with a pyrene reagent, 4-(1-pyrene)butyric acid *N*-hydroxysuccinimide ester (PSE). All primary and secondary amines in polyamine molecule are labeled with PSE to form the corresponding dipyrene- to tetrapyrene-labeled derivatives (Figure 14). The derivatives afford intramolecular excimer to fluoresce, which can clearly be discriminated from the monomer fluorescence emitted by PSE, its hydrolysate, and monopyrene-labeled derivatives of monoamines (Figure 15). The PSE derivatives of putrescine, cadaverine, spermidine, and spermine can be separated by a reversed-phase LC on a  $\text{C}_8$  column with a linear gradient elution. The detection limits for the polyamines are at 1–8 fmol on-column.

The determination of basic amino acids such as lysine and ornithine, and triethylenetetramine, a therapeutic drug for Wilson's disease, in biological samples has been performed in the same manner.

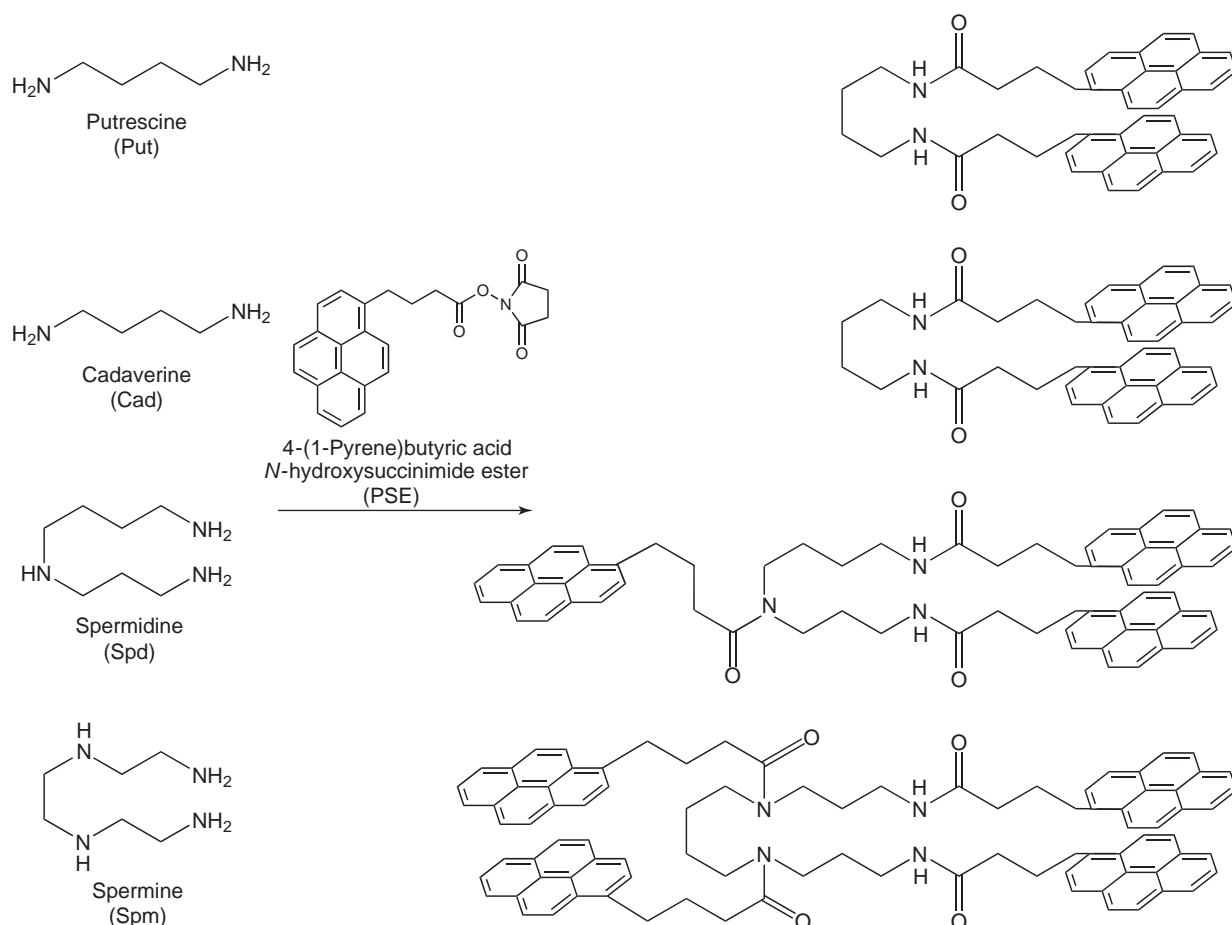


**Figure 14** Inter-molecular excimer-forming fluorescence derivatization of polyamines with PSE. (Reprinted with permission from Chromatogram (2000). *Analytical Chemistry* 72: 4203; © American Chemical Society.)

#### Determination of Plasma Saturated Fatty Acids Using Pentafluorophenyldimethylsilyl Derivatives by GC-MS

The common fatty acids (FAs) of plant and animal origin contain several carbon atoms in their straight chains with a terminal carboxyl group. Some of the saturated FAs in tissues or body fluids are used as the qualitative biomarkers for fatty acids intake and peroxisomal disorders. GC-MS has been widely used for the determination of FAs in biological mixtures including their methyl, benzyl, and trimethylsilyl esters. The appropriate choice of derivatization in GC-MS with electron impact ionization is primarily concerned not only for GC properties, but also for the selection of efficient quantitative ions in the





**Figure 15** Chromatograms obtained with a standard mixture of the polyamines. (Reprinted with permission from Intramolecular excimer-forming. (2000). *Analytical Chemistry* 72: 4200; © American Chemical Society.)

selected ion-monitoring (SIM) mode. The GC–SIM–MS for saturated fatty acids (SFAs) as their flophemesyl ester derivatives has been developed for the detection of 11 SFAs including C12:0–C26:0 (even numbers only), C17:0, C19:0, and C23:0 in human plasma with stable isotope  $d_3$ -stearic acid as an internal standard. FAs derivatized with flophemesyl chloride at room temperature for 15 min are separated with a HP Ultra-2 (25 m  $\times$  0.2 mm, ID) fused-silica capillary column. The detection limits are in the range of 1–10 pg. This method can determine saturated FAs in plasma sample of normal subjects and patients with X-linked adrenoleukodystrophy, which is one of the hereditary peroxisomal disorders.

**See also:** **Analytical Reagents:** Specification. **Derivatization of Analytes.** **Fluorescence:** Derivatization; Fluorescence Labeling; Quantitative Analysis. **Lipids:** Fatty Acids. **Liquid Chromatography:** Liquid Chromatography–Mass Spectrometry; Pharmaceutical Applications. **Mass Spectrometry:** Forensic Applications. **Spectrophotometry:** Derivative Techniques.

## Further Reading

- Bardelmeijer HA, Lingeman H, Ruiter C, and Underberg WJM (1998) Derivatization in capillary electrophoresis. *Journal of Chromatography A* 807: 3–26.
- Kataoka H (1996) Derivatization reactions for the determination of amines by gas chromatography and their applications in environmental analysis. *Journal of Chromatography A* 733: 19–34.
- Molnar-Perl I (2000) Role of chromatography in the analysis of sugars, carboxylic acids and amino acids in food. *Journal of Chromatography A* 891: 1–32.
- Mukherjee PS and Karnes HT (1996) Ultraviolet and fluorescence derivatization reagents for carboxylic acids suitable for high performance liquid chromatography. *Biomedical Chromatography* 10: 193–204.
- Ohkura Y, Kai M, and Nohta H (1994) Fluorogenic reactions for biomedical chromatography. *Journal of Chromatography B* 659: 85–107.
- Shimada K and Mitamura K (1994) Derivatization of thiol-containing compounds. *Journal of Chromatography B* 659: 227–241.
- Toyo'oka T (ed.) (1999) *Modern Derivatization Methods for Separation Sciences*. Chichester: Wiley.



## GAMMA-RAY SPECTROSCOPY

See RADIOCHEMICAL METHODS: Gamma-Ray Spectrometry

## GAS CHROMATOGRAPHY

Contents

**Overview**

**Principles**

**Column Technology**

**Gas–Solid Chromatography**

**Multidimensional Techniques**

**High-Temperature Techniques**

**High-Speed Techniques**

**Instrumentation**

**Online Coupled LC–GC**

**Pyrolysis**

**Detectors**

**Mass Spectrometry**

**Fourier Transform Infrared Spectroscopy**

**Physicochemical Measurements**

**Environmental Applications**

**Forensic Applications**

**Petrochemical Applications**

**Chiral Separations**

### Overview

**K Robards**, Charles Sturt University, Wagga Wagga, NSW, Australia

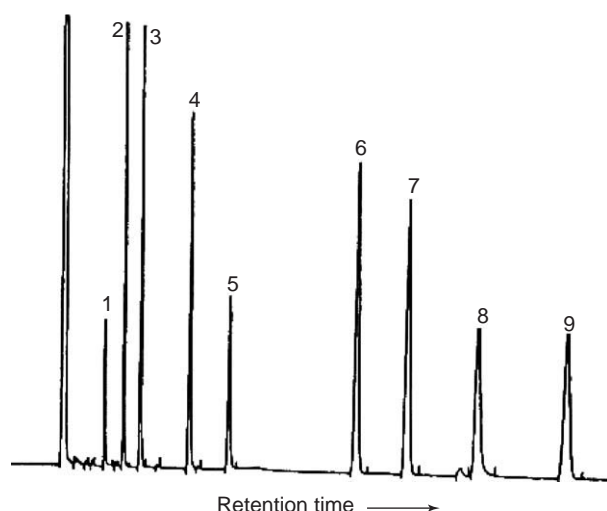
© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Gas chromatography (GC) is a dynamic method of separation and detection of volatile compounds. It was first introduced in the 1950s and rapidly

established itself as a routine analytical technique in most industrial and academic laboratories. From its introduction and until the advent of high-performance liquid chromatography, it dominated separation methods. This can be attributed to the capability for high resolution, selectivity, and sensitivity.

Separation in GC is achieved by partitioning of gaseous solutes between a typically inert gaseous mobile phase and a stationary liquid or solid phase retained in a column. These variants are described as gas–liquid chromatography (GLC) and gas–solid chromatography (GSC), respectively. With the exception of some specialized areas such as the analysis for inorganic gases, it is GLC which is used.

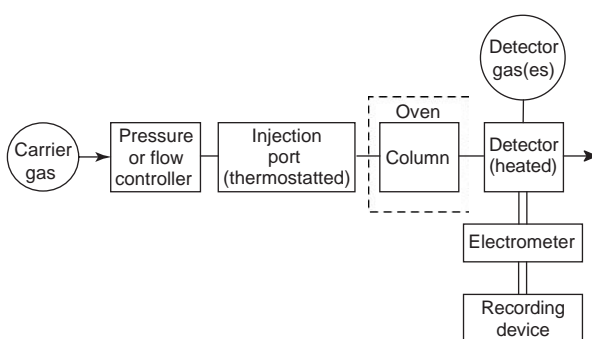


**Figure 1** A gas chromatogram showing the separation of a nine-component mixture. Peaks due to individual compounds are labeled 1 through 9. The unlabeled peak is attributed to the sample solvent. The retention time for individual components can be read from the chromatogram although it is generally provided in a separate report.

Nevertheless, the instrumentation is virtually identical for the two techniques.

In the most common approach (elution development), the sample is introduced into the chromatograph via the sample inlet into a continuous flow of mobile phase, which is referred to as the carrier gas. The sample is vaporized in the inlet system and transported by the carrier gas to the thermostatted column where separation occurs. The individual components give rise to an electrical signal in the detector that may have provision for the inlet of additional make-up gas. This is necessary to permit separate optimization of gas flow through the column and detector. After suitable amplification the detector signal is conducted to a recording device. The detector output is produced as a chromatogram (Figure 1). This is a plot of detector signal versus time in which individual peaks represent the separated components of the sample. Sample components can be identified from their characteristic retention times. With proper calibration, the amounts of the components of a mixture can be measured accurately also.

Figure 2 shows the essential components of a gas chromatograph as a block diagram. These parts can be identified as the carrier gas, sample introduction system, the column, the detector, and data acquisition system comprising an electrometer and integrator/recording device. Although not apparent from the figure, there are three separately controlled heated zones for the inlet, column, and detector in the typical instrument. An overview of some of these aspects



**Figure 2** Block diagram of a gas chromatograph.

is given in the following sections. The reader is also referred to the more detailed discussion of detectors and specific techniques such as pyrolysis GC, high-speed GC, and GSC. Extensive resources for training, method development, and support services are provided by a number of manufacturers of gas chromatographic instrumentation.

## Sample

It follows from the preceding discussion that the basic requirement with respect to sample is that it has an appreciable vapor pressure at the column temperature. As usually practiced, the sample must also be thermally stable. This allows the sample components to vaporize in, and move with, the gaseous mobile phase. This requirement is not as severe a restriction as it appears since column temperatures as high as 450°C (300°C is more common) are used in GC. Thus, GC can be applied to all permanent gases, most nonionized small- or medium-sized organic molecules (typically up to C30), and many organometallic compounds but it cannot be used for macromolecules or salts. In some instances, nonvolatile compounds can be converted into more volatile and stable derivatives before chromatography. In a typical sample containing a mixture of volatile and nonvolatile components, care must be taken that the nonvolatile solutes are not deposited in the system where they can interfere with subsequent analyses.

## Mobile Phases

Substances capable of interacting with the analyte and influencing selectivity have been used as carrier gases in rare instances. However, the ideal mobile phase for GC is usually nonreactive toward the analyte(s), nonflammable, cheap, and environmentally friendly since it is vented at the end of the instrument. Hence, the choice of a mobile phase or carrier

**Table 1** Physical properties (at 273 K and 101 kPa) of carrier gases used in gas chromatography

Gas	Thermal conductivity ( $10^8 \text{ W m}^{-1} \text{ K}^{-1}$ )	Viscosity ( $10^{-7} \text{ Pa s}$ )	Density ( $\text{kg m}^{-3}$ )
Hydrogen	16.75	84	0.0899
Helium	14.07	186	0.1785
Nitrogen	2.39	166	1.2505
Argon	1.67	212	1.7839
Neon	4.56	298	0.8999

gas is determined by practical constraints of cost, availability, inertness, and detector compatibility rather than its ability to effect a particular separation. The usual mobile phase in GC is therefore a noninteractive gas that does not influence selectivity. However, the carrier gas can influence resolution through its effect on column efficiency because of differences in solute diffusion rates for various gases. Moreover, it can effect analysis time and plays a role in pressure-limiting situations because of differences in gas viscosities (see **Table 1**).

Taking these considerations into account, hydrogen, helium, and nitrogen are the most popular carrier gases in GC. Carrier gases are usually supplied from a compressed gas cylinder. Gas purity is a major consideration and, in general, the highest purity gas should be used to reduce deterioration of the stationary phase and lessen detector noise. Moreover, it is usual to include oxygen and moisture traps in the carrier gas lines. These traps are commercially available, containing activated carbon (to remove organic impurities) or molecular sieves (for moisture and oxygen). The traps must be monitored and periodically regenerated. When changing cylinders, it is important to ensure that all fittings are free of dust and dirt particles before connection to gas lines.

Typical compressed gas cylinders contain a pressure of 20 MPa whereas supply line pressures to the gas chromatograph are more commonly in the range 50–300 kPa. Thus, appropriate regulators and controllers are used to step down and control the pressure and flow rate to the column. With traditional instruments, the carrier gas is regulated by either a pressure regulator or flow controller. The choice between the two is dependent on the inlet system and column type. In recent years, instrument manufacturers have introduced completely electronic programmable pressure-controlled gas chromatographs.

When a capillary column is installed in an instrument it should be checked for carrier gas flow before connecting the detector end of the column to avoid the possibility of heating a column with no flow.

When the column connection has been completed, the system should be checked for gas leaks. Once any leaks have been eliminated and the column purged with carrier gas, the volumetric flow rate ( $\text{ml min}^{-1}$ ) or the theoretically more useful average linear gas velocity ( $\text{cm s}^{-1}$ ) can be measured.

## Sample Introduction Systems

The injection port or inlet system is the next major component of the gas chromatograph. It must receive the sample and deliver the correct amount of material to the column so as not to exceed the sample capacity of the column or the linear range of the detector in use. Several types of inlet and sample introduction techniques have been developed to accommodate the diversity of sample types and particularly the state of aggregation of the sample and the range of columns. Specialized techniques include equilibrium headspace sampling, purge and trap sampling, pyrolysis GC, and multidimensional chromatography in which the sample entering the column differs from the composition of the original sample. However, in the more usual case, the material entering the column must have the same composition as the original sample. Additionally, the sample has to be delivered to the column as a sharp band.

The most common analysis involves injection of 1–3  $\mu\text{l}$  of a liquid into a heated inlet. This is accomplished by means of a microsyringe through a septum made of elastomer or rubber, which seals the inlet system as the syringe needle is withdrawn. Septa have a limited lifetime dependent on the mode of injection (automatic versus operator injection), the injector temperature, and the septum quality. They are available from a number of manufacturers and a good operating principle is to perform a separation with a solvent blank, particularly with a new batch of septa. Syringe injection is also applicable to gases (50–1000  $\mu\text{l}$ ) but reproducibility is relatively poor and a sampling valve is more common. Syringes are available from a number of manufacturers in various configurations; needle point style, length of needle, fixed or replaceable needle. Most needles are constructed of stainless steel but specialty fused silica needles are available for on-column injection. An important consideration in choosing a syringe is the correct needle length to ensure delivery of the sample at the correct position in the injection zone.

With packed columns, the sample solution is introduced via the syringe into the sealed injection port that is heated to a higher temperature than the column in order to assist vaporization. Sample discrimination, which can be regarded as a measure of how well the detected peak areas reflect the original

sample composition, is not a problem. On the other hand, the much smaller sample capacity and lower carrier gas flow-rates associated with capillary columns magnify the extent of any problems and these are manifest as sample discrimination. Thus, more attention has been given to detailed investigation of various injection techniques when using capillary columns. These include the use of a vaporizing injector (i.e., heated injection port), cold syringe needle injection, hot needle injection, and solvent flush technique. The hot needle and solvent flush techniques are about equally effective in reducing discrimination and are preferred over other methods.

Traditional sample inlet systems were constructed of metal, thus providing metallic surfaces where sample decomposition was possible during sample evaporation. Interchangeable glass liners in the inlet, which are available in a range of configurations, are now standard in practically every sample injection system involving evaporation of the injected sample. Capillary columns have a very low sample capacity, and to avoid overloading the stationary phase specialized injection systems have evolved. The more important of these are split injection, splitless injection, cold on-column injection, and programmed temperature vaporizer split/splitless injection. These variants have evolved to meet the diverse needs of sample type and analyte concentration. For instance, splitless injection is more suited to trace/ultratrace analysis than is split injection.

## Columns

In GC where the mobile phase is noninteractive, the column alone determines the selectivity of the separation. From its inception, up to the 1980s, almost all separations in GC were performed on conventional packed columns despite the demonstration by Golay in 1957 of much greater efficiency obtainable with capillary columns. However, the obvious advantages of capillary columns in terms of higher resolution, greater sensitivity (despite injection

of less solute), reduced analysis time (to achieve equivalent resolution), and greater chemical inertness were gradually recognized. More recently, polymer-clad flexible fused silica capillary columns with chemically bonded and/or cross-linked immobilized stationary phases have become commercially available at reasonable cost and this has led to the current popularity of capillary columns. These columns now routinely provide high efficiency, inertness, and reproducibility. Alternatively, some separation efficiency can be sacrificed by using shorter columns to achieve very rapid analyses.

Capillary columns are available from several manufacturers in a wide range of column internal diameters (0.1–1.0 mm), column lengths (5–50 m), and stationary phase film thicknesses (0.1–5.0  $\mu\text{m}$ ). Generally, sample capacity increases but the efficiency decreases as the internal diameter or film thickness increases. The larger bore capillary columns with internal diameters between 0.53 and 1.00 mm are termed wide bore or megabore capillary columns and these have similar capacities, but greater efficiencies, than packed columns (see Table 2).

The largest variation in properties between conventional packed columns and capillary columns is associated with the column permeability. For this reason, capillary columns offer much less flow resistance and can be used in much longer lengths. Ultimately, the comparison of different column types is between the efficient use of column head pressure. Thus, a packed column containing 10  $\mu\text{m}$  particles can generate 50 000 theoretical plates per meter but requires a head pressure of 20  $\text{MPa m}^{-1}$ , whereas a 70 m capillary column of 50  $\mu\text{m}$  internal diameter can provide over one million theoretical plates with a column pressure drop of  $\sim 2.2$  MPa.

The stationary phase distinguishes GSC from GLC. In the former it is a solid adsorbent whereas in GLC it is a liquid either coated on a solid support (packed column) or deposited directly on the column walls. GSC preceded GLC but has never achieved the same prominence. Nonetheless, GSC has some

**Table 2** Comparison of packed and capillary columns

Parameter	Column type			
	Packed	Microbore capillary	Capillary	Megabore capillary
Internal diameter (mm)	1/4 in	100 $\mu\text{m}$	200 $\mu\text{m}$	530 $\mu\text{m}$
Length (m)	0.5–3	5–50	5–100	5–100
Permeability ( $10^{-7} \text{ cm}^2$ )	1–50	300–20 000		
Film thickness ( $\mu\text{m}$ )	1–10	0.1	0.2–2	1–5
Carrier gas average linear velocity ( $\text{cm s}^{-1}$ )	2–4	20–30	20–35	20–40
Flow rate ( $\text{ml min}^{-1}$ )	50–60	0.2–0.5	0.2–2.0	3–5
Sample capacity (ng)	20 000	<5	20–500	1000–15 000



important application areas such as the separation of inorganic gases and low molecular mass hydrocarbons for which GLC shows little selectivity. The main adsorbents for GSC are based on silica, charcoal, alumina, or molecular sieves although the development of new adsorbents is continuing.

The liquids used as stationary phases in packed and capillary columns are closely related. Nevertheless, liquid phases in capillary columns are usually cross-linked and bonded and may exhibit slight differences in selectivity to nominally equivalent packed column materials. The selection and comparison of stationary phases is confusing for newcomers as some 300 phases are available and in excess of 1000 have been described in the literature. Nevertheless, a fairly limited set of packed columns will suffice in most laboratories while an even more limited set of capillary columns will satisfy the needs of most laboratories. Moreover, two forces have combined to contain the proliferation of phases. Firstly, the high efficiency of capillary columns has reduced the necessity for many selective liquid phases and, secondly, theoretical studies have aided in phase selection.

There are several factors to consider in selecting a stationary phase. General considerations include temperature limits of the stationary phase, column efficiency, and lifetime and detector compatibility. Since nonpolar phases generally give more efficient columns that also exhibit superior lifetimes, it is wise to use the least polar phase that provides satisfactory separation. Phases containing the specific element corresponding with element-selective detectors (e.g., cyanopropyl phases with an NPD detector; trifluoropropyl phases with an ECD detector) should be avoided where possible. These selective detectors will be substantially more sensitive to 'normal' column bleed with such phases.

The most difficult factor to assess is the ability of a phase to effect the desired separation. From this perspective, the selection of a stationary phase and column is a daunting prospect. In theory, the selection is based upon maximizing the difference in selectivity between the solutes toward the phase. The separation is increased by exploiting solute-stationary phase interactions that retard the progress of some solutes relative to others so as to increase their retentions. The types of interactions to consider are:

- London or dispersion forces which are weak and nonspecific;
- dipole-dipole interactions or dipole-induced dipole interactions; and
- acid-base interactions or proton transferring (or sharing) tendencies of either the solute or stationary phase.

In practice, experience of similar separation problems, literature data relating to the target separation, and availability of the column phases are often the factors that determine the choice of a particular phase and column for a specific application.

The ideal liquid phase has a low vapor pressure, high thermal and chemical stability, low viscosity, nonreactivity toward sample components, and a wide temperature operating range, extending from  $-80^{\circ}\text{C}$  to  $450^{\circ}\text{C}$ . The phase must exhibit reasonable solvent properties (i.e., dissolving power) for the solutes in order to ensure symmetrical peaks. Stationary phases can be divided into nonpolar, polar, and specialty phases. These differ in their ability to interact with solutes of different structure, i.e., their selectivity. The nonpolar phases contain no functional groups capable of specific interaction (e.g., hydrogen bonding or dipole interactions) with the sample. Here, interaction between solute and stationary phase is limited to dispersive forces, and components therefore separate according to their volatility with the elution order following the boiling points. Compounds that cannot be differentiated on the basis of their boiling points (i.e., they have similar or equal boiling points) require a different stationary phase for separation. To obtain the differentiation of solutes by forces other than dispersion, a polar phase containing groups capable of specific interactions with sample components is required. The elution order now depends on a combination of volatility and specific polar-polar interactions. The relative magnitude of the various interactions (dispersive, dipole, hydrogen bonding, and acid/base) determines the selectivity of the phase toward particular solutes. The selectivity and resolution of a separation can be optimized by choosing a stationary phase that exploits the different interactions.

Nonpolar phases include a variety of hydrocarbons, such as squalane or Apolane C87, or mixtures of long-chain *n*-alkanes such as Apiezon L. Polymers based on a silicon-oxygen-silicon backbone form the basis of the most widely used group of stationary phases. These linear polysiloxanes differ in their average molecular mass, thermal stability, and viscosity. The chemical difference lies in the substituent and degree of substitution on the silicon backbone. Polar phases have been prepared by substituting polar trifluoropropyl or cyano groups for the methyl groups of the dimethylsilicones. By incorporating different proportions of the polar groups, stationary phases with a wide range of polarities can be produced. Other polar materials include polyethylene glycols or polyoxiranes with the structure  $-(\text{CH}_2\text{CH}_2-\text{O})_n-$ .

Specialty phases have been developed for use with particular analytical techniques such as GC-MS



where low bleed phases are essential, to meet the needs of particular groups (e.g., United States Environmental Protection Agency methods), or to separate particular classes of solutes. Included in the latter are chiral phases and Carbowax phases modified for separation of acids and bases.

### Column Temperature

Column temperature is an important variable that must be controlled in GC. Thus, the column is housed in a thermostatted oven. For simple samples containing relatively few peaks, an appropriate column temperature can be determined experimentally to achieve the separation and isothermal analysis is suitable. Nonetheless, many samples contain components with a wide range of volatility and more volatile components are eluted rapidly with no resolution when analyzed isothermally at a high temperature whilst the analysis time is unacceptably long and later eluting peaks are very broad and may be lost as baseline drift when analyzed isothermally at a low temperature. For such samples, temperature programming in which the column temperature is ramped during the analysis is essential.

### Detectors

Online detection is an integral part of a gas chromatograph. The detector monitors the column effluent and produces an electric signal that is proportional to the amount of analyte being eluted. The output signal is recorded as a continuous trace of signal intensity against time. In principle, any physical or physico-chemical property of the analyte that deviates from the properties of the carrier gas can serve as the basis for detection. Thus, over 100 detectors for GC have been described but relatively few are in common use.

The operation and applicability of different detectors can be compared against several performance criteria. These criteria include the sensitivity, noise, minimum detectable quantity or detection limit, detector time constant and response time, and the selectivity of the response. For purposes of screening a sample of unknown composition, a universal

detector has definite advantages whereas a selective detector may aid in the identification of an unknown compound or a given class of compounds. Selective detectors are particularly useful for the analysis of complex mixtures, where the selectivity may greatly simplify the chromatogram through suppression of the response of many potentially interfering compounds.

Detectors can also be classified as destructive or nondestructive. With nondestructive detectors, the original chemical form of the analyte persists throughout the detection process. This is an obvious advantage when the analyte is required for further analysis. In destructive detectors, the process of detection involves an irreversible chemical change in the analyte. A more useful classification distinguishes detectors on the basis of the transducer mechanism as ionization, spectroscopic, etc.

A consideration of the characteristics discussed above and the needs of a particular analytical problem will determine the most appropriate detector for a given problem. A detector with a wide linear dynamic range and low detection limit will be adopted for the determination of trace components in addition to main components in a sample. On the other hand, the use of a selective detector is convenient if the trace components belong to a particular class of substance or possess some common functional group.

Of the many available detectors, the most common (Table 3) are thermal conductivity detector (TCD), flame ionization detector (FID), electron-capture detector (ECD), alkali-flame ionization detector (AFID or NPD), flame photometric detector (FPD), and mass selective detector. The TCD and FID are usually considered universal detectors as they respond to most analytes whereas the ECD, AFID, and FPD are the most useful selective detectors and give differential responses to analytes containing different functional groups. Note that this does not imply that the magnitude of the response of the universal detectors is constant to all analytes. The mass selective detector has the advantage of operation in either universal or selective detection mode whilst an infrared detector is a powerful tool for distinguishing isomers.

**Table 3** Classification of the most common gas chromatographic detectors

<i>Detector</i>	<i>Response</i>	<i>Optimal detection limit</i>	<i>Destructive</i>
TCD	Organic and inorganic solutes	5–100 ng	No
FID	All organic solutes except formic acid and formaldehyde	10–100 pg	Yes
ECD	Halogenated and nitro compounds	0.05–1 pg	No
AFID	P- or N-containing solutes	0.1–10 pg	Yes
FPD	P- or S-containing solutes	10–100 pg	Yes
Mass selective	General all-purpose detector that is replacing FID in a number of situations	Dependent on mode of operation	Yes

The first detector commercially available for GC, the TCD or katharometer, remains a consideration for situations requiring universal detection. The TCD responds to any compound, irrespective of its structure, whose thermal conductivity differs from that of the carrier gas. Hence, it is the only choice for detection of compounds to which other more sensitive detectors give a poor or negligible response. In particular, it is the standard detector for determination of inorganic gases such as the permanent gases, hydrogen, oxygen, nitrogen, carbon dioxide, carbon monoxide, carbon disulfide, and water.

The FID is the standard workhorse detector in GC. It consists of a stainless steel jet constructed so that carrier gas exiting the column flows through the jet, mixes with hydrogen gas, and flows to a microburner tip that is swept by a high flow of air for combustion. Ions produced by the combustion are collected at a pair of polarized electrodes, constituting a small background current that is the signal. When solutes enter the detector, they are combusted and the signal increases. The current produced is then amplified and passed to a recording device. Unlike the TCD, the FID gives virtually no response to inorganic compounds. Most organic compounds, however, give similar responses, which is approximately proportional to the total mass of the carbon and hydrogen in the analyte. A reduced response is usually observed with the first members of a homologous series and compounds with a large proportion of oxygen.

The popularity of the ECD can be attributed to the high sensitivity to organohalogen compounds, which include many compounds of environmental interest, including polychlorinated biphenyls and pesticides. It is the least selective of the so-called selective detectors but has the highest sensitivity of any contemporary detector. The NPD or thermionic ionization or emission detector is a modified FID in which a constant supply of an alkali metal salt, such as rubidium chloride, is introduced into the flame. It is a detector of choice for analysis of organophosphorus pesticides and pharmaceuticals. The FPD detects specific luminescent emission originating from various excited state species produced in a flame by sulfur- and phosphorus-containing compounds.

## Dual Detection

The simultaneous use of two or more detectors, whose outputs complement each other, can aid in compound identification by generating substance-characteristic detector response ratios. In some instances, the detectors are operated sequentially or, alternatively, the column eluate is split and passed separately to the individual detectors. The combination of a selective with a universal detector can provide information on both the whole sample and, at the same time, greater quantitative sensitivity on specific components.

*See also:* **Gas Chromatography:** Principles; Column Technology; Instrumentation; Detectors; Mass Spectrometry.

## Further Reading

- Cazes J and Scott RPW (2002) *Chromatography Theory. Chromatographic Science Series*, vol. 88. New York: Dekker.
- Gehrke CW, Wixom RL, and Bayer E (eds.) (2001) Chromatography—a century of discovery 1900–2000. *Journal of Chromatography Library*, vol. 64.
- Grant DW and Grant RPW (1996) *Capillary Gas Chromatography. Separation Science Series*. New York: Wiley.
- Grob RL and Barry EF (1995) *Modern Practice of Gas Chromatography*, 3rd edn. New York: Wiley.
- Handley AJ and Adlard ER (2001) *Gas Chromatographic Techniques and Applications*. Sheffield: Academic Press/Blackwell Science.
- Issaq HJ (ed.) (2002) *A Century of Separation Science*. New York: Dekker.
- Jennings W, Mittlefehldt E, and Stremple P (1997) *Analytical Gas Chromatography*, 2nd edn. San Diego, CA: Academic Press.
- McNair HM and Miller JM (1998) *Basic Gas Chromatography*. New York: Wiley-Interscience.
- Moldoveanu SC and David V (2002) Sample preparation in chromatography. *Journal of Chromatography Library*, vol. 65.
- Niessen WMA (ed.) (2001) *Current Practice of Gas Chromatography–Mass Spectrometry. Chromatographic Science Series*, vol. 86. New York: Dekker.
- Robards K, Haddad PR, and Jackson PE (1994) *Principles and Practice of Modern Chromatographic Methods*. London: Academic Press.

## Principles

**P J Marriott**, RMIT University, Melbourne, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Gas chromatography (GC) is the premier chemical separation method for volatile compounds. It

The first detector commercially available for GC, the TCD or katharometer, remains a consideration for situations requiring universal detection. The TCD responds to any compound, irrespective of its structure, whose thermal conductivity differs from that of the carrier gas. Hence, it is the only choice for detection of compounds to which other more sensitive detectors give a poor or negligible response. In particular, it is the standard detector for determination of inorganic gases such as the permanent gases, hydrogen, oxygen, nitrogen, carbon dioxide, carbon monoxide, carbon disulfide, and water.

The FID is the standard workhorse detector in GC. It consists of a stainless steel jet constructed so that carrier gas exiting the column flows through the jet, mixes with hydrogen gas, and flows to a microburner tip that is swept by a high flow of air for combustion. Ions produced by the combustion are collected at a pair of polarized electrodes, constituting a small background current that is the signal. When solutes enter the detector, they are combusted and the signal increases. The current produced is then amplified and passed to a recording device. Unlike the TCD, the FID gives virtually no response to inorganic compounds. Most organic compounds, however, give similar responses, which is approximately proportional to the total mass of the carbon and hydrogen in the analyte. A reduced response is usually observed with the first members of a homologous series and compounds with a large proportion of oxygen.

The popularity of the ECD can be attributed to the high sensitivity to organohalogen compounds, which include many compounds of environmental interest, including polychlorinated biphenyls and pesticides. It is the least selective of the so-called selective detectors but has the highest sensitivity of any contemporary detector. The NPD or thermionic ionization or emission detector is a modified FID in which a constant supply of an alkali metal salt, such as rubidium chloride, is introduced into the flame. It is a detector of choice for analysis of organophosphorus pesticides and pharmaceuticals. The FPD detects specific luminescent emission originating from various excited state species produced in a flame by sulfur- and phosphorus-containing compounds.

## Dual Detection

The simultaneous use of two or more detectors, whose outputs complement each other, can aid in compound identification by generating substance-characteristic detector response ratios. In some instances, the detectors are operated sequentially or, alternatively, the column eluate is split and passed separately to the individual detectors. The combination of a selective with a universal detector can provide information on both the whole sample and, at the same time, greater quantitative sensitivity on specific components.

*See also:* **Gas Chromatography:** Principles; Column Technology; Instrumentation; Detectors; Mass Spectrometry.

## Further Reading

- Cazes J and Scott RPW (2002) *Chromatography Theory. Chromatographic Science Series*, vol. 88. New York: Dekker.
- Gehrke CW, Wixom RL, and Bayer E (eds.) (2001) Chromatography—a century of discovery 1900–2000. *Journal of Chromatography Library*, vol. 64.
- Grant DW and Grant RPW (1996) *Capillary Gas Chromatography. Separation Science Series*. New York: Wiley.
- Grob RL and Barry EF (1995) *Modern Practice of Gas Chromatography*, 3rd edn. New York: Wiley.
- Handley AJ and Adlard ER (2001) *Gas Chromatographic Techniques and Applications*. Sheffield: Academic Press/Blackwell Science.
- Issaq HJ (ed.) (2002) *A Century of Separation Science*. New York: Dekker.
- Jennings W, Mittlefehldt E, and Stremple P (1997) *Analytical Gas Chromatography*, 2nd edn. San Diego, CA: Academic Press.
- McNair HM and Miller JM (1998) *Basic Gas Chromatography*. New York: Wiley-Interscience.
- Moldoveanu SC and David V (2002) Sample preparation in chromatography. *Journal of Chromatography Library*, vol. 65.
- Niessen WMA (ed.) (2001) *Current Practice of Gas Chromatography–Mass Spectrometry. Chromatographic Science Series*, vol. 86. New York: Dekker.
- Robards K, Haddad PR, and Jackson PE (1994) *Principles and Practice of Modern Chromatographic Methods*. London: Academic Press.

## Principles

**P J Marriott**, RMIT University, Melbourne, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Gas chromatography (GC) is the premier chemical separation method for volatile compounds. It

involves the passage of a carrier gas over a supporting solid or liquid phase in a chromatographic column, with the carrier gas causing the compounds to move along the support material from the injector to the detector. These compounds must possess the physical property of having a given vapor pressure at the temperature of analysis, which in turn permits them to have the necessary concentration in the gas phase to allow them to travel along the GC column when a carrier gas flows along the column.

The GC experiment involves setting of conditions for carrier gas flow and the temperature of analysis, and selecting appropriate column properties. Column-based parameters include choice of packed or capillary format, column dimensions, and stationary phase and/or support. For all practical purposes, today analytical GC is conducted using capillary GC columns. There are only a few niche applications for which packed column GC is the preferred method. This includes preparative GC studies, selected catalyst studies, some inverse GC methods, and methods that might involve simple separations where larger volumes can easily be applied to the GC column. **Figure 1** shows the delineation of the two approaches, and **Table 1** compares various properties for packed and capillary columns for the specific case of analysis of fatty acid methyl esters, whilst **Table 2** reports selected properties of different modes of GC. Note that PLOT columns, which have an adsorbent phase fixed to the capillary wall, are primarily suited to permanent gases and other very volatile compounds. SCOT columns were originally developed to allow ease of coating thick film phases to the capillary, but this requirement is less important today as improved coating procedures are available. However, very polar phases may still present a challenge to capillary column manufactures, and so the range of phases is not as extensive as is available for packed column GC.

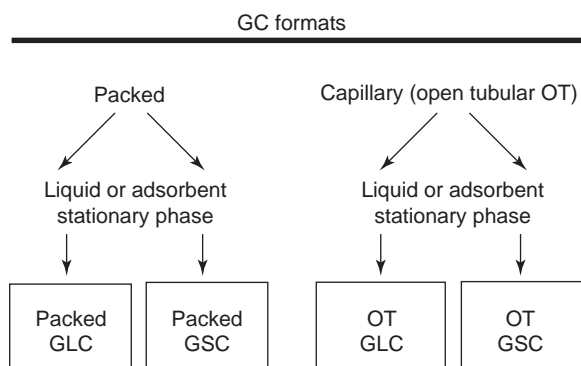
## Introductory Relationships of GC

The basic equation that determines the suitability of a compound to undergo gas chromatographic elution

is given by the relationship between distribution constant,  $K$ , and concentrations of solute in each phase of the GC column, at the temperature of the experiment:

$$K = C_S/C_M \quad [1]$$

The concentration terms,  $C_S$  and  $C_M$  (in stationary and mobile phases, respectively), here depend upon the specific type of phase employed in the experiment.



**Figure 1** Gas chromatography is often used in packed column or capillary column modes, with either a liquid or adsorbent stationary phase. GLC, gas–liquid chromatography; GSC, gas–solid chromatography.

**Table 1** Comparison of properties for packed and capillary columns for analysis of fatty acid methyl esters by using a diethylene glycol succinate (DEGS) phase, at 180°C

Column property	Packed column	Capillary column
Length $L$ (m)	2.4	100
Average linear carrier gas velocity $\bar{u}$ (cm s <sup>-1</sup> )	8	16
Unretained peak time $t_M$ (s)	30	625
Retention time $t_R$ (methyl oleate) (min)	29.8	38.5
Retention factor $k$ (methyl oleate)	28.5	2.7
Theoretical plates $N$ (methyl oleate)	3300	294 000
$R_s$ , methyl oleate and methyl stearate	1.5	10.6
Plate height $H$ (methyl oleate) (mm)	0.73	0.34

Adapted from Ettre LS (1979) *Introduction to Open Tubular Columns*. Norwalk, USA: Perkin-Elmer Corp.

**Table 2** Selected general properties of different modes of gas chromatography

Mode	$L$ (m)	ID (mm <sup>a</sup> )	$d_f$ (μm)	Phase ratio $\beta$	$H_{min}$ (mm)	$N$ (m)
Packed	2–5	2–4	5–20% w/w load	10–30	0.5	2000
SCOT	15–20	0.3–0.53	1.0+	20–100	0.5–1.0	1000–2000
PLOT	30	0.3–0.53	NA <sup>b</sup>	NA <sup>b</sup>		1000+
WCOT	10–60	0.05–1.0	0.05–5.0	15–250	0.05–1.0	2000–15 000

<sup>a</sup> Capillary columns traditionally have the following IDs: 0.05, 0.1, 0.15, 0.22, 0.25, 0.32, 0.53, 1.0 mm.

<sup>b</sup> PLOT columns have an adsorbent phase, so film thickness and normal  $\beta$  value cannot be determined.

SCOT = support-coated open tubular column. These columns are less popular today since thicker WCOT phases are now available.

PLOT = porous layer open tubular column.

WCOT = wall-coated open tubular column.



Clearly,  $C_M$  can be represented as the moles/volume concentration in the gas phase. Where the stationary phase is a liquid (or more likely, a polymer phase),  $C_S$  can be represented as moles/volume concentration in the liquid phase. However, for a gas–solid chromatography case, the stationary phase is an uncoated solid, and  $C_S$  will be related to the adsorption of solute onto the support, which is a surface sorption phenomenon. In this case,  $C_S$  will be moles/unit surface area on the solid phase.

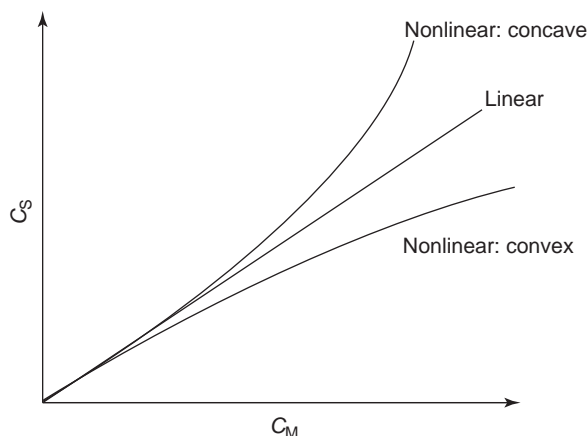
Hence, eqn [1] can be rewritten as

$$C_S/C_M = n_S/V_S(V_M/n_M) = (n_S/n_M)(V_M/V_S) = k\beta \quad [2]$$

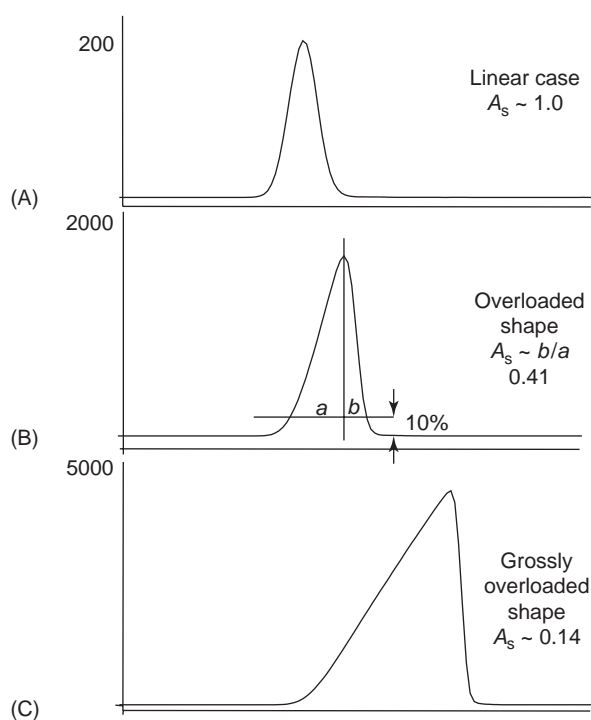
where  $n_S/n_M$  are number of solute moles in the two phases, with corresponding volumes  $V_S$  and  $V_M$ . Thus, a primary equation of chromatography is

$$K = k\beta \quad [3]$$

Since  $K$  is related to the concentration equilibrium of solute between the two phases, it follows that if more solute is introduced into the column, then for  $K$  to be constant, both  $C_S$  and  $C_M$  should increase proportionally. This means that  $k$  then must be a constant value (note that  $\beta$  must be constant, since it only depends on the column properties) and so retention time will also be constant. Since more sample is introduced, the chromatographic peak will be larger, but will be symmetric and located at the same time as for smaller injected amounts. This corresponds to linear chromatography conditions, and explains why retention time is a constant property of the solute provided linear conditions are maintained. **Figure 2** illustrates that linear chromatography arises when  $K = \text{constant}$ . Clearly there is a limit to how much solute can be introduced into a column and produce the same retention time – i.e., has a symmetric peak shape. Excessive injected solute will cause nonlinearity, and the peak shape will become nonsymmetric. Usually in gas–liquid chromatography, the peak shape will be a fronting shape, which becomes increasingly triangular with larger amounts injected. The reason for this effect is that the gas phase has limited ability to accept vapor of the solute and becomes saturated under those conditions (vapor phase overload). The equilibrium condition is no longer maintained, and the amount of solute in the stationary phase is more than it should be for an equilibrium case. This causes the peak maximum of the solute to be retarded compared with the linear case, so retention increases. **Figure 3** shows the progressive increase in peak asymmetry ( $A_s$ ; see **Figure 3B**) as more solute is introduced onto a capillary GC column, with the increase in response and total area confirming the greater amount of solute injected, and the increase in  $t_R$  arising from overloading. In **Figure 2**, the concave



**Figure 2** Isotherms in gas chromatography.



**Figure 3** (A) Symmetric peak shape and (B), (C) Progressively more overloaded peak shapes for a convex isotherm.

isotherm explains this effect ( $C_M$  limited). For gas–solid chromatography, the tendency for the solid phase is to be more readily saturated at higher injected amounts (there is limited surface area to accept solute under near to equilibrium conditions) and so the excess solute will tend to be forced into the gas phase, leading to a peak shape opposite to that for the overloaded gas–liquid case. Note that this can be seen for phases with some degree of adsorption, such as enantioselective phases with guest–host cyclodextrin additives.

## Efficiency of GC Columns

Efficiency is essentially the duration that a solute has been exposed to the chromatographic process compared with the narrowness of the solute band that results:

$$N = (L/\sigma)^2 \quad [4]$$

Since the chromatographic peak can be approximated by a normal distribution, the width of which can be expressed as its standard deviation, then the bandwidth is best reported as the peak variance,  $\sigma^2$ . Hence, the length measure is given as  $L^2$ . This can be measured in terms of distance units along the column, the time of the process, or obtained by directly measuring the recorded chromatogram. All these are equivalent measures, but the most useable and common method is to use the time basis, since these are given by the chromatographic report:

$$L = t_R \quad \text{and} \quad s = 4w_b \quad \text{or} \quad 2.35w_h \quad \text{or} \quad 2w_i \quad [5]$$

where  $w_b$ ,  $w_h$ , and  $w_i$  are peak widths at base, half-height, and the inflexion point, respectively. Hence,

$$\begin{aligned} N &= 16(t_R/w_b)^2 \quad \text{or} \quad N = 5.54(t_R/w_h)^2 \\ \text{or} \quad N &= 4(t_R/w_i)^2 \end{aligned} \quad [6]$$

This is shown schematically in Figure 4, with the other peak parameters defined in Figure 5. The inflexion lines to the peak intersect at  $1.213h$  and pass exactly through  $w_i$ , and this aids the correct lines to be drawn. The usual chromatographic data system will give retention time, total peak area, height, and peak-width at half-height ( $w_h$ ). Hence, calculating  $N$  based on  $t_R$  and  $w_h$  is easiest.

Theoretical plates calculation uses retention from the time of injection ( $t=0$ ). This means that even a peak eluting almost with the solvent or unretained peak will have quite good efficiency. But this is not really usable efficiency in terms of the ability to resolve peaks. It is often more realistic to represent efficiency in terms of effective plates, where  $t'_R$  is used in the respective formulas in eqn [6]. Thus, a peak eluting near the unretained peak has a low number of effective plates.

The  $N$  calculation above cannot be used for temperature programmed conditions. It is only relevant to isothermal operation. For temperature program analysis, it is more relevant to describe the quality of the chromatography column by resolution or Trennzahl (Tz). The latter is related to the separation number SN, which is calculated for the separation of successive pairs of a homologous series, such as

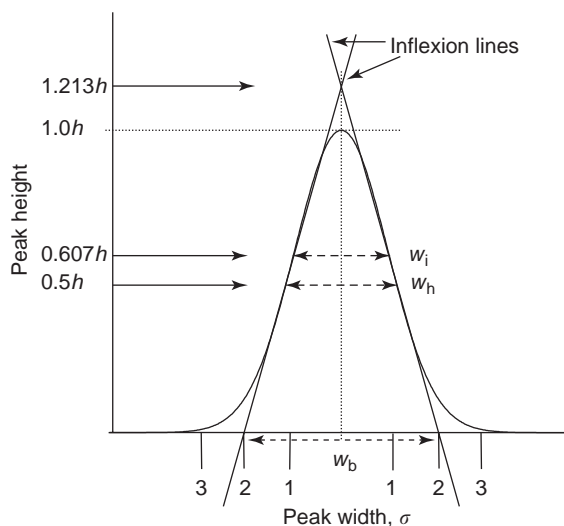


Figure 4 Parameters of a normal curve.

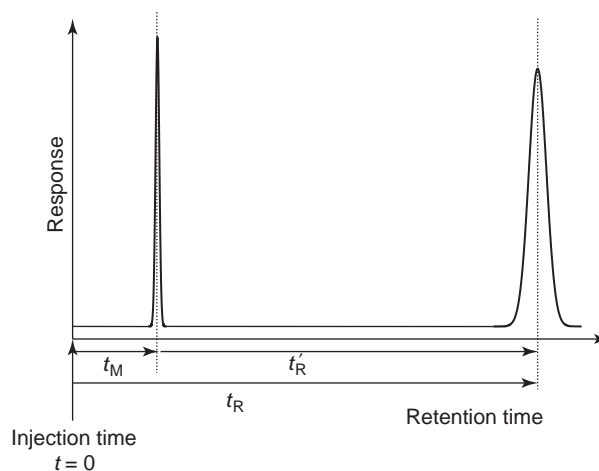


Figure 5 Typical chromatogram, showing retention terms.

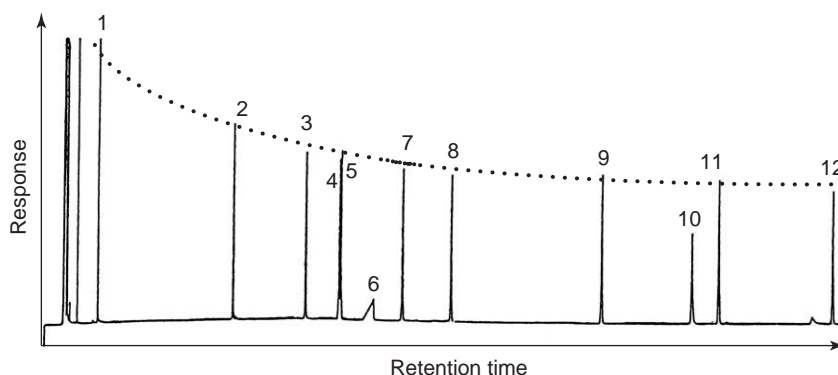
normal alkanes:

$$SN = [(t_{R(z+1)} - t_{R(z)})/(w_{h(z)} - w_{h(z+1)})] - 1 \quad [7]$$

This essentially describes how many solutes can be placed between the two series members. The better the efficiency of the column (i.e., narrower peaks), the larger will be the separation number. This can be used as part of a generalized quality measure of a column's quality.

For a capillary column, quality is also measured by the use of a standardized test mix, developed by Grob *et al.* (Figure 6). This mix is intended to provide qualitative information about the stationary phase coating, and in particular activity toward certain solutes. Thus, the mixture contains 2,3-butanediol, *n*-decane, 1-octanol, nonanal, 2,6-dimethylphenol, 2-ethylhexanoic acid, 2,6-dimethylaniline, *n*-dodecane,





**Figure 6** Typical results of the standardized Grob test, for determination of column quality. Mixture contents (and mass in the split method-injected solution, in  $\text{mg ml}^{-1}$ ) are: (1) 2,3-butanediol (0.95); (2) *n*-decane (0.43); (3) 1-octanol (0.555); (4) nonanal (0.625); (5) 2,6-dimethylphenol (0.485); (6) 2-ethylhexanoic acid (0.605); (7) 2,6-dimethylaniline (0.513); (8) *n*-dodecane (0.44); (9) methyl decanoate (0.605); (10) dicyclohexyl amine (0.51); (11) methyl undecanoate (0.59); and (12) methyl dodecanoate (0.578). The dotted line represents the expected peak maxima when each solute exhibits complete recovery, and excellent chromatographic behavior. (From Jennings W (1987) *Analytical Gas Chromatography*, p. 234. Orlando: Academic Press.)

methyl decanoate, dicyclohexylamine, methyl undecanoate, and methyl dodecanoate, although manufacturers often use other specific test mixtures. In this case, SN can be calculated between the homologous methyl esters. The peak shapes of various polar molecules (acidic/basic) and in particular attenuation of peak response is related to 'activity' of the column and may indicate that it is unsuited to particular analyses. The 2-ethylhexanoic acid solute is a stringent and possibly inappropriate test since the tendency to vapor phase overload almost inevitably leads to asymmetric fronting peaks on many stationary phases. The solutes are normally prepared in the test solution in amounts that produce very similar peak height responses for all solutes on a relatively nonpolar phase column (i.e., taking into account response factors and different peak broadening extents), which make the identification of peak attenuation relatively easy to recognize. Clearly, a different phase polarity will cause peak retention changes to what is shown in Figure 6.

## The van Deemter Equation

The van Deemter equation is the basic interpretive equation of column efficiency, which provides an insight into the factors that lead to broadening of a compound as it travels along the column. The equation considers various contributions to diffusion broadening of the chromatographic band, and so incorporates various terms that include diffusion coefficients in each of the phases. In its most basic form, the van Deemter equation can be written as

$$H = A + B/\bar{u} + C\bar{u} \quad [8]$$

These terms are referred to as the multiflow path (or Eddy diffusion; *A*), longitudinal diffusion (*B*), and resistance to mass transfer (*C*) terms, respectively.

A short comment on each is appropriate here.

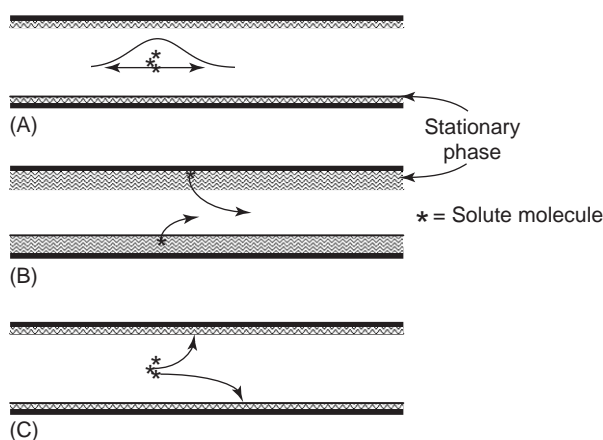
The *A* term is a constant, independent of flow rate, which accounts for the effects of multipath flows in a column, and also may include the effects of injection and detection (i.e., extra column effects), which serve to broaden a peak. Note that this term is taken as zero for a capillary column.  $A = 2\lambda d_p$ , where  $d_p$  is particle diameter and  $\lambda$  is a packing factor.

The *B* term accounts for the effect of a solute undergoing diffusional broadening within the mobile phase. Clearly a band of gaseous molecules must spread apart relatively readily within the low density carrier gas as shown in Figure 7A where the gas phase distribution is superimposed in the figure, and so the higher the diffusion coefficient the greater the spreading. Hence,

$$B = 2\gamma D_M \quad [9]$$

where  $\gamma$  is a constant for tortuosity, which obstructs the longitudinal spread of the solute molecules. In a capillary column this constant will be unity.

The *C* term is a resistance to mass transfer term. Whilst this is sometimes referred to as the preference for a solute molecule to stay in the phase it is in, it can also be considered the effect of ease of crossing the phase boundary, from gas to liquid, or vice versa. Ideally, equilibrium across the phase boundary is established very rapidly, in order for the molecular distribution of the chromatographic band in the liquid phase to match that in the gaseous phase. If the reemergence of solute from the stationary phase is



**Figure 7** Representation of the  $B$  and  $C$  terms in capillary GC. (A) The  $B$  term leads to longitudinal diffusion (along the column) as shown by the superimposed distribution. (B) Mass transfer in the liquid phase (solutes re-emerging from the stationary phase) may be slow if the phase film thickness is large. (C) Mass transfer in the gas phase is aided by high diffusion coefficients in the gas phase (i.e., by use of  $H_2$  carrier gas) so that solute can migrate to the stationary phase and undergo equilibration processes.

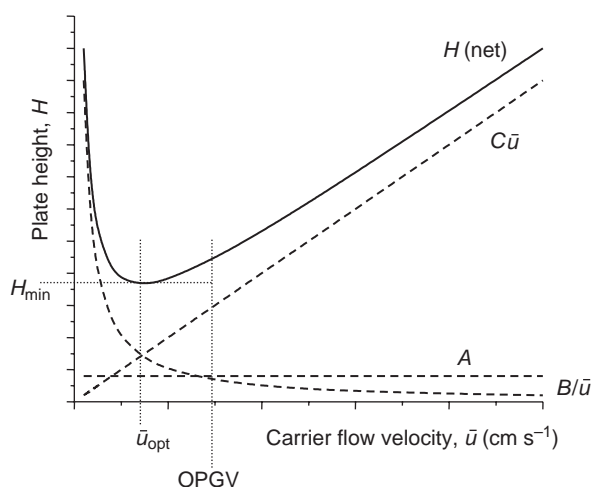
slow compared to the linear velocity of the solute molecules along the column in the mobile phase, then there will be a natural broadening effect arising from this mass transfer process. There are two primary processes to consider – transfer from gas to liquid, and also liquid to gas. Hence, there should be at least two terms comprising  $C$ . The first is  $C_S$ , stationary phase mass transfer. This is a function of retention factor  $k$ , stationary-phase diffusion coefficient  $D_S$ , and incorporates phase film thickness,  $d_f$ . The equation often reported is:

$$C_S = 2/3[k/(1+k)^2]d_f^2/D_S \quad [10]$$

It is clear from schematic in **Figure 7B** that if the film thickness is large, the tendency will be for a solute which diffuses deeply into the stationary phase to lag the solute distribution in the gas phase. In the capillary GC experiment, provided  $d_f$  is small, this term may be neglected in preference to the mass transfer in the mobile phase:

$$C_M = [(1 + 6k + 11k^2)/96(1+k)^2][d_c^2/D_M] \quad [11]$$

where  $d_c$  is the diameter of the column. The important conclusions from this expression are that  $C_M$ : depends on the retention factor ( $k$ ) of a solute; is smaller for narrower ID capillary columns; and is inversely proportional to the diffusion coefficient in the mobile phase. In this case, the larger the  $D_M$  value, the smaller the plate height contribution. This can be rationalized by reference to **Figure 7C**. The



**Figure 8** van Deemter plot ( $H$  versus  $\bar{u}$ ) showing effects of individual terms  $A$ ,  $B/\bar{u}$ , and  $C\bar{u}$ . The most efficient conditions giving minimum plate height ( $H_{\min}$ ) at optimum flow velocity ( $\bar{u}_{\text{opt}}$ ) occur at the minimum of the van Deemter curve. Optimum practical gas velocity (OPGV) is located at  $\sim 2 \times \bar{u}_{\text{opt}}$ .

most important feature of the chromatographic bed in providing separation and efficiency is its ability to provide equilibration of the solute between the two phases. Thus, this term indicates that the faster the movement of the solute through the carrier gas (higher diffusion) or shorter mean free path to the phase surface (provided by a narrow column ID) the more efficient the outcome. For a common column diameter of 0.32 mm, the  $C_M$  term dominates the two mass transfer terms up to film thicknesses of 1–2  $\mu\text{m}$   $d_f$ .

The relative contribution of each of the above terms can be seen in the  $H - \bar{u}$  curve generated at different carrier gas velocities (**Figure 8**). The  $B$  term is dominant at low carrier flows, because excessive spreading in the carrier gas arises, and the low flow will mean that solute molecules in the gas and stationary phases will not be displaced very much in space arising from mass transfer effects. The  $C$  term is dominant at high flows because at high carrier velocities the solute molecules residing in the carrier gas will be carried a long distance along the column and therefore will be well displaced from the molecules that are held in the stationary phase, before each of them undergoes their next mass transfer equilibration. It is interesting that  $D_M$  appears in both the  $B$  term and the  $C_M$  term, but its effect on each is the opposite. For  $B$ , a low diffusion constant is desirable, but for  $C_M$ , a high  $D_M$  is desirable. Since it is often preferred that a column is operated at higher flow than optimum, such as at the OPGV (optimum practical gas velocity) defined as  $2 \times \bar{u}_{\text{opt}}$ ,

where faster analysis time is achieved without too much loss of efficiency, then the  $B$  term becomes less important and the properties that decrease  $C$  should be considered. This therefore suggests use of the carrier gas that decreases  $C_M$ , i.e., that which has the higher  $D_M$  – hydrogen – is the obvious choice. Thus,  $H_2$  is often the carrier gas of choice in high-resolution capillary GC analysis; it is also cheaper than the traditional helium carrier gas. An individual laboratory must evaluate concern regarding safety issues if  $H_2$  is to be used as carrier gas.

In his book (see Further Reading), Jennings presents a compilation of  $H - \bar{u}$  curves for a range of capillary columns that vary from short (5 m) and very narrow bore (0.05 mm), to long (60 m) and wide bore (0.53 mm ID) columns, as an aid to understanding optimization of the operational conditions.

## Variations to the van Deemter Equation

Two major variations can be made to the basic van Deemter equation. The first is a modification that makes it specific to capillary columns. By deleting the  $A$  term (the capillary column has no multifold path nature), the Golay equation is defined as

$$H = B/\bar{u} + (C_M + C_S)\bar{u} = B/\bar{u} + C_M\bar{u} \quad [12]$$

Note that the  $C_S$  term may also be deleted provided the above assumption of thin film stationary phase applies. The Golay curve will be as in Figure 8, but with no  $A$  term.

This can now give a number of useful interpretations of the performance of the capillary column. From the minimum of the Golay curve, the minimum plate height can be derived:

$$H_{\min} = 2\sqrt{BC} \quad [13]$$

For the case of the unretained solute,  $k=0$ , this is evaluated as

$$H_{\min} \sim 1/4d_c \quad [14]$$

Hence, for a column of 0.25 mm ID, methane should give a peak with  $\sim 15\,000$  theoretical plates per meter.

For a retained solute,  $k \neq 0$ , and an expression in  $k$  is derived:

$$H_{\min} \sim 1/4d_c \left[ (1 + 6k + 11k^2)/(1 + k)^2 \right]^{0.5} \quad [15]$$

This allows the best possible performance of a capillary column to be determined for any solute, under optimum conditions (i.e., at best flow velocity). Again, this depends on column inner radius ( $r_c$ ) and

**Table 3** Minimum plate height and maximum theoretical plates for capillary columns of various inner diameters

$d_c$ (mm)	$H_{\min}$ (mm)	$N_{\max}$ (plates $m^{-1}$ )
0.05	0.0354	28 200
0.1	0.0709	14 100
0.2	0.1418	7 050
0.22	0.1560	6 400
0.25	0.1772	5 600
0.32	0.2268	4 400
0.54	0.3828	2 600
1.0	0.7089	1 400

Data calculated for  $k=4$ .

solute retention factor. Table 3 lists optimum efficiencies for various capillary column inner diameters, for a solute with  $k=4$ .

This is a useful expression, since it provides a comparison of experimental performance against what might be expected theoretically. This can provide a measure of coating quality, or utilization of theoretical efficiency, %UTE. For a uniform, thin film coating, %UTE might be expected to be high (e.g., 80%). For a difficult to coat phase, a lower coating efficiency may be accepted for preparation of a capillary column.

The second variation arises from the use of conditions that affect the primary assumptions of the basic derivation of the van Deemter equation. There is an increasing interest in use of conditions that lead to fast analysis. This can be achieved by short, narrow bore columns, very high temperature program rates, and high carrier gas inlet pressure. Often, a combination of two or more of these may be employed. For higher phase ratios  $\beta$  (i.e., for thinner film phase coatings), shorter columns, or higher carrier velocities, a higher temperature programming rate ( $r_T$ ) can (should) be used for optimum performance:

$$r_{T\text{opt}} \propto \bar{u}\beta/L \quad [16]$$

It is accepted as a rule of thumb that optimum programming rate performance is taken as  $10^\circ\text{C min}^{-1}$  per unretained peak time (e.g., if  $t_M = 0.6$  min, then  $r_{T\text{opt}} \sim 10/0.6 = 16^\circ\text{C min}^{-1}$ ).

Hinshaw provides a useful comparison of experimental results arising from experiments where different isothermal temperatures are used, the temperature programming rate is altered, column length is altered, carrier gas flow rate is altered, and where column ID is altered to achieve equivalent resolution performance using a combination of a narrower and shorter column.

With the trend toward use of narrow columns, it has been proposed that the high-pressure drops that will be associated with generating reasonable carrier

gas flows will require modification of the Golay equation as follows:

$$H = B/\bar{u}^2 + C_M\bar{u}^2 + C_S\bar{u} \quad [17]$$

## Resolution of Solutes

Resolution is defined as

$$R_s = 2[t_{R(2)} - t_{R(1)}]/[w_{b(1)} + w_{b(2)}] \\ \sim \Delta t_R/w_b \sim \Delta t_R/4\sigma \quad [18]$$

This is a simple expression, which relates the difference in peak retention to their respective widths. Since the peaks will naturally be located closely, their widths will be similar so the second expression will be approximately the same as the first.

The greater the relative response difference between the two peaks, the less well can the smaller peak be recognized as the resolution diminishes. Figure 9 is a schematic diagram that shows the extent of overlap for different peak resolutions, for an equally abundant mixture (i), and a 1:4 mixture ratio (ii).

The resolution can also be expressed in terms of how well the peaks are separated in respect of their purities. The better the resolution, the more pure will be the peak (for example, if we were to collect each peak exactly to their valley point). Thus, Table 4 reports the peak purity of solutes A and B for different degrees of resolution.

A further important expression is the one that relates resolution to the column efficiency, relative

separation and retention;

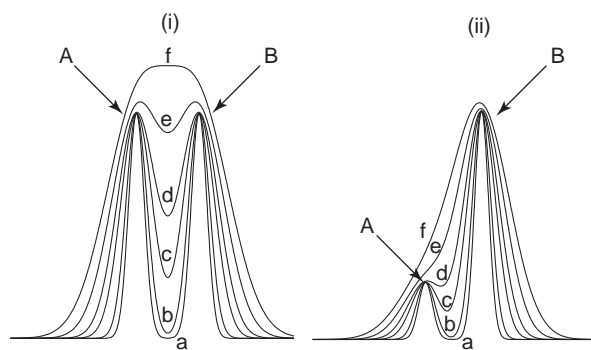
$$R_s = 1/4(N)^{0.5}[(\alpha - 1)/\alpha][k/(k + 1)] \quad [19]$$

(where  $\alpha$  is relative retention  $t'_{R2}/t'_{R1} = k_2/k_1$ ). It is clear then that resolution doubles for a fourfold increase in column efficiency. This relation also holds for column length.

It is resolution that effectively determines the quality of an analysis. It has been shown above that performance of the column with respect to the ability to fit in resolved peaks between homologous series members at a given resolution is called the separation number. Here, the calculation of resolution, and the meaning of compound measurement quality is defined. It is clear that as measurement accuracy increases ( $R_s$  between neighboring compounds increases) fewer total analytes can be measured in a given 'separation space'. There is a point where increased  $R_s$  is no longer desirable, since the maximum measurement accuracy for quantitative measurement of a peak area will be reached, and the total analysis uncertainty is determined by other factors, such as injection precision. The usual minimum  $R_s$  is taken as 1.5, but a greater minimum  $R_s$  might be chosen for improved ease of quantitation or ruggedness of the separation. This requirement is much more critical for nondiscriminating detectors, from which solute-specific response criteria cannot be chosen (these might be termed single channel detectors). Thus, the flame ionization detector will not permit 'deconvolution' of individual peak responses for grossly overlapping peaks (note that chemometric deconvolution is a separate consideration). The mass spectrometer, however, is a multichannel detector, and may permit choice of specific ions that allow each component peak of overlapping peaks to be uniquely selected and their areas quantitated.

The ultimate chromatography requirement, however, is for resolved peaks, and hence the continual search for or development of methods that increase component separation. The separation space can be interpreted in respect of the peak capacity of the analysis. Peak capacity represents the theoretical total number of compounds that can be placed, at a given resolution, in the separation space. For a single dimension analysis, this will be (total analysis time)/(width of each resolved peak). Or

$$n_c = (1.5\Delta t)/(R_s w_b) \quad [20]$$



**Figure 9** Peak resolution plots for  $R_s$  values of (a) 2.0; (b) 1.5; (c) 1.0; (d) 0.8; (e) 0.625; (f) 0.5, and for situations of (i) two peaks A and B of equal size; (ii) peak A = 0.25 peak height of B.

**Table 4** Peak purity for different degrees of peak resolution, where purity is defined as the purity of each component that could be collected from an equally abundant mixture where sampling is taken to the valley point between the peaks (see Figure 9(i))

$R_s$	0.5	0.6	0.7	0.8	1.0	1.25
Peak purity	84	88	92	95	98	99.4

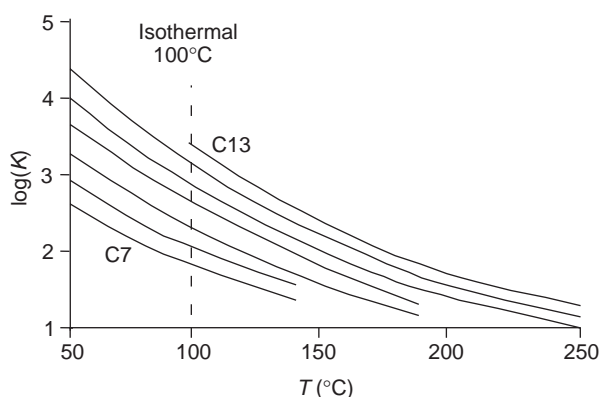
where  $n_c$  is the total peak capacity,  $\Delta t$  is the total analysis time, and  $w_b$  is the peak basewidth.

Since maximum use of separation power for a multicomponent sample is obtained in temperature programmed analysis, peak widths are often similar (although this is not exactly so). Taking, for example, a total analysis time of 60 min, and average peak width as 8 s at base (measured at  $4\sigma$ ), then  $(60 \times 60)/8 = 450$  peaks resolved at the  $4\sigma$  level can be fitted into the analysis. However, a sample that comprises 450 individual components will not have them uniformly distributed across the total analysis space, and so peak overlap will invariably occur. It has been deduced that when a sample contains more components than about 20% of the column peak capacity, randomly distributed throughout the separation space, then peak overlap will arise. Of course, most samples would not contain randomly distributed compounds, but they will be more likely clustered into a given region of the separating space, and this further reduces resolution.

Column efficiency is only one strategy for achieving peak separation, albeit an important one. Column stationary phase selectivity is another important parameter, and is best considered to be a primary parameter where there are only a few target solutes of interest, and where their elution properties are similar (i.e., they have similar retention times). Where there are a plethora of compounds, all of which are important, and they all show problem resolutions, it may be difficult to choose an optimum stationary phase that simultaneously achieves resolution of all these.

## Isothermal and Temperature Programming Operation

The temperature condition of analysis affects the vapor pressure or volatility of the solute in the experiment, and therefore the  $C_M$  value, and hence  $K$ . Since  $K = k\beta$ , then increased temperature gives reduced retention time. The  $\log(K)$  value decreases approximately linearly with increased  $T$ , and Figure 10 shows such a diagram. Note that at any given temperature, e.g., along the vertical line at 100°C, the homologous series of alkanes exhibit a constant increase in  $\log(K)$ , which is a general observation for homologous series in isothermal GC. The graph of  $\log(t'_R)$  versus carbon number is a straight line. Hence, the chromatographic result will show the alkanes or other homologues series members increase their retention times exponentially as the chain length increases. This observation is the basis for retention index calculations (see below).



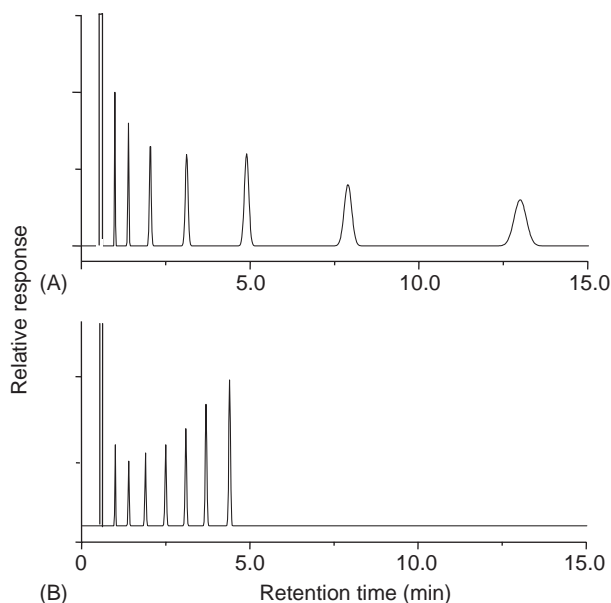
**Figure 10** Effect of temperature on distribution constant ( $K$ ; proportional to retention factor,  $k$ ), for a series of  $n$ -alkanes. At any one temperature (e.g., 100°C here), the homologous series of compounds have logarithmically increasing  $K$  values.

Given that the relatively small boiling point range of the alkanes in the above example ( $C7 = 98^\circ\text{C}$ ; tridecane  $= 235^\circ\text{C}$ ) leads to a large difference in  $K$  and therefore retention difference of about 50-fold, it is impractical to analyze a mixture that contains a range of compounds of widely varying boiling points by using isothermal means. The appropriate method is to employ temperature programming oven conditions. Thus, at low-temperature operation, the more volatile compounds (lower boiling point) should have sufficient transport properties to permit them to elute, and as the oven temperature is increased, the compounds of higher boiling point are progressively increased in vapor phase concentration to allow them to also elute. Figure 11 represents the difference in results that will be observed in these two approaches for a mixture of homologous compounds. Figure 11A is an isothermal analysis that illustrates the increase in retention times of the members of the series, and the increase in peak width that accompanies this. In contrast, temperature programming will give a result similar to that in Figure 11B, with now an almost linear increase in absolute retention time from one member to the next, a similar peak width for each component, and if the same sample is analyzed as in Figure 11A, the resultant relative peak heights will be greater for the later peaks as their widths are comparatively decreased by a greater extent.

## Retention Indices

A single peak eluting from a GC column possess much information, but if not adequately described, this information may not be effectively communicated. The peak expresses both qualitative and quantitative information; however, as a single isolated line



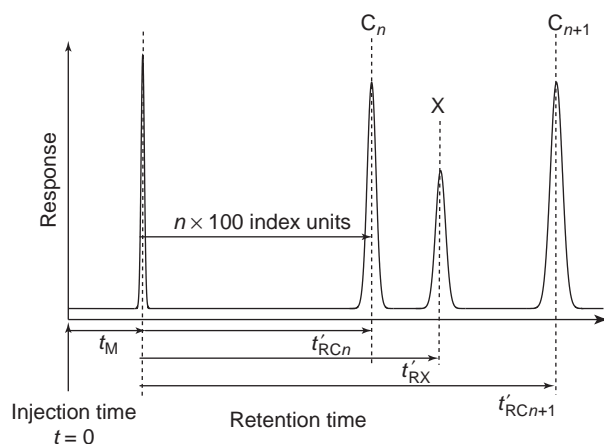


**Figure 11** (A) Isothermal temperature conditions lead to an exponential increase in retention times for successive members of a series of homologous compounds. Peaks get broader, and for approximately equal amounts injected, their peaks will get smaller (assuming similar response factors). (B) Under temperature programmed conditions, peak widths are now similar over the elution range for the sample, and total analysis time can be reduced significantly compared with isothermal operation. Peaks should be approximately similar heights if the same amounts are injected. Here, it would be suspected that more of the later eluting compounds are present.

on a chromatogram the peak itself is almost meaningless. For instance, the peak area is useful only if the relationship between peak area or response, and amount of compound is established. So, also for the retention of a peak, this contains the essence of the compound identity, but only if the retention is reproducible from run-to-run, and if the retention has been classified against an 'elution scale'. It is also important to recognize that the absolute matching of retention time is a necessary, but not sufficient, criterion for confirmation of a component's identity. One compound must have a single retention time, but two or more compounds can have the same retention time.

It is important to compare reference standards and unknown samples under exactly the same conditions, preferably on exactly the same column (and it must be using the same stationary phase).

The retention index system of Kovats is one way to enhance the quantitative elution information of a compound on a particular column. It essentially relates the elution of a compound to that of a reference series of compounds, in this case the  $n$ -alkanes. The retention index  $I$  is formally calculated under



**Figure 12** Procedure for calculation of isothermal retention indices. See text for explanation.

isothermal conditions, according to:

$$I = 100n + 100 \times \frac{(\log t'_{RX} - \log t'_{RCn})}{(\log t'_{RCn+1} - \log t'_{RCn})} \quad [21]$$

where these values are defined in Figure 12.  $C_n$  is the  $n$ -alkane eluting just prior to solute X, and  $C_{n+1}$  the alkane eluting just after X. The procedure is a logarithmic interpolation. It is apparent that  $n$ -alkanes have retention index values of  $n \times 100$ .  $I$  values may be calculated by the mathematical calculation of eqn [21], or by plotting a graph of  $I$  versus  $\log t'_R$  for the  $n$ -alkanes, and then using this to find  $I$  for solute X from its  $\log t'_R$  value.

Other retention bases may be used for this approach; thus, homologous series of alcohols, ethers, esters, etc., might be used in place of the  $n$ -alkanes, which might not be so suited for use with especially polar phases. The specific example of fatty acids is an important analysis, and by using saturated fatty acid methyl ester (FAME) standards as the retention series, other FAME may be quantified in terms of their 'effective chain lengths (ECL)'. In this case, unsaturation of the FAMEs can lead to either longer or shorter retention than their saturated counterparts, depending on the stationary phase used, and so they will have noninteger  $I$  (or ECL) values.

When the available standards are not successive carbon numbers, but, for example only the even carbon-number compounds, eqn [21] may be modified by using  $t'_{RCn}$  and  $t'_{RCn+2}$  data, with interpolation within 200 index units.

Since much analysis is conducted under temperature programmed conditions, eqn [21] is modified by recognizing that absolute elution times are now



almost linear, with carbon number. Thus,

$$I = 100n + 100 \times [(t'_{RX} - t'_{RCn}) / (t'_{RCn+1} - t'_{RCn})] \quad [22]$$

It is important in all cases to specify the analysis conditions, such as the isothermal conditions and phase for eqn [21], and temperature program rate and phase for eqn [22].

For eqn [22], also called linear temperature programmed retention indices (LTPRI), a single temperature program ramp is most appropriate. These LTPRI values find particular application in essential oil analysis, where standard conditions have been developed by Adams, requiring a program rate of  $3^{\circ}\text{C min}^{-1}$ , and either a 5% phenyl dimethyl polysiloxane column, or a polyethylene glycol phase (note that LTPRI values are tabulated for each of these columns). Since the retention indices are not definitive by themselves, these are usually augmented by mass spectral library matching, with the two independent identification data providing a more secure confirmation of compound identity.

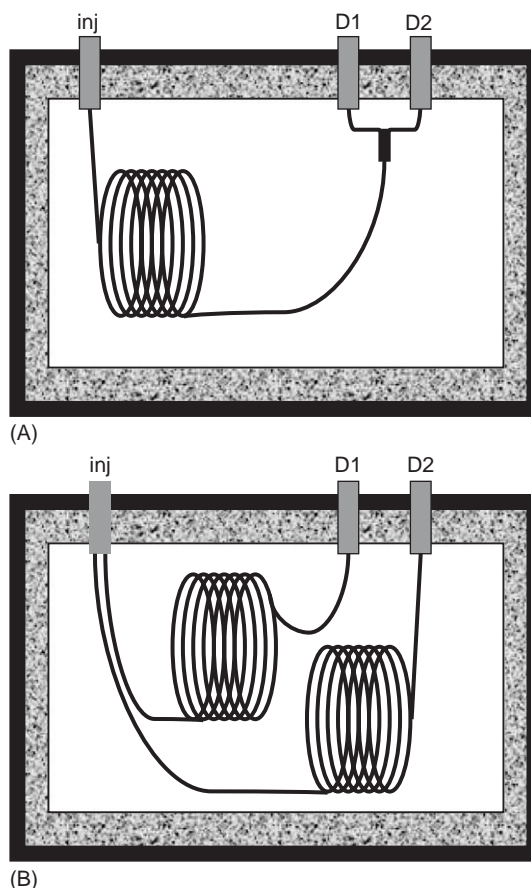
Retention indices find further application in characterization of stationary phases. Application of the concept of McReynolds uses a series of five solutes of differing chemical class (alcohol, aromatic, ketone, nitro, pyridine) to compare their retention indices on the test phase, against their retention indices on a stationary phase accepted to have very little 'polarity' (squalane). It can be argued that if a solute exhibits greater retention index on the test phase, then the intermolecular interactions between the solute and the phase are stronger than in the case when squalane is used, which means that the test phase should also exert a similar effect on similar solutes. Colloquially this is referred to as the 'phase polarity', and a scale may be set up according to the following:

$$\begin{aligned} \sum \Delta I = & \Delta I(\text{benzene}) + \Delta I(1 - \text{butanol}) + \Delta I(2 - \text{pentanone}) \\ & + \Delta I(1 - \text{nitropropane}) + \Delta I(\text{pyridine}) \end{aligned} \quad [23]$$

The larger the  $\sum \Delta I$  value, the larger the intermolecular interactions, the larger the phase polarity. This tends to be a rather imprecise measure of phase property, and how it is used in practice is not clear. Other descriptions are much more theoretically acceptable (refer to Poole's book, listed in the Further Reading) and are based on solute descriptors and allow more rigorous prediction of retention properties, but the above approach, being commonly used, relatively easily understood, and using the language of 'polarity' will make this concept difficult to supplant.

## Multiple Column Operation

As implied above, the use of two columns of disparate phases may provide useful supplementary results for identification. This is especially useful when the two phases have complementary resolution performance for compounds in a specific analysis. The expectation is that compounds not resolved on one phase may be resolved on the second phase, provided sufficiently different retention mechanisms are available on these phases for the compounds of interest. **Figure 13B** shows the possible instrument arrangement for a dual column–dual detector single injection analysis. Often the injectors will be the same if it is only the phase difference and therefore resolution differences that are of interest. It is necessary to have a well-characterized set of component overlap problems, so that routine data interpretation is possible. An alternative approach to obtaining two sets of retention data simultaneously is shown in **Figure 13A**, where outlet splitting of the effluent from the column



**Figure 13** (A) Dual detection analysis where the effluent from the single column is split into two different detectors. (B) Dual column analysis provides for different columns to be used to attempt to maximize resolution of components in complex sample analysis.

allows the response of two detectors to be acquired, where at least one of these will be a specific detector.

*See also:* **Chromatography:** Overview; Principles. **Derivatization of Analytes. Gas Chromatography:** Overview; Multidimensional Techniques; High-Speed Techniques; Instrumentation; Detectors.

## Further Reading

- Adams RP (1995) *Identification of Essential Oil Components by Gas Chromatography/Mass Spectrometry*. Carol Stream, IL: Allured Publishing Corp.
- Braithwaite A and Smith FJ (1985) *Chromatographic Methods*, 4th edn. London: Chapman and Hall.

- Handley AJ and Adlard ER (2001) *Gas Chromatographic Techniques and Applications*. Sheffield: Sheffield Academic Press.
- Heftmann E (2004) *Chromatography: Fundamentals and Applications of Chromatography and Related Differential Migration Methods*, 6th edn. Part A: Fundamentals and Techniques. Part B: Applications. Amsterdam: Elsevier.
- Hinshaw J (1995) GC in the fast lane. *LC GC-Mag. Sep. Sci.* 13: 944–949.
- Jennings W (1987) *Analytical Gas Chromatography*. Orlando: Academic Press.
- Mondello L, Lewis AC, and Bartle KD (2001) *Multidimensional Chromatography*. Chichester: Wiley.
- Poole CF (2003) *The Essence of Chromatography*. Amsterdam: Elsevier.

## Column Technology

**C F Poole**, Wayne State University, Detroit, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

One of the major changes in the practice of gas chromatography over the last decade or so was the development of columns from being a highly skilled laboratory preparation procedure to a general laboratory supply item. Few laboratories prepare their own columns today and the general skills to do so are largely lost outside of the manufacturing facilities of the main suppliers of columns to the laboratory market. This became possible because of a gradual evolution in the understanding of the unit operations of column preparation procedures and their simplification together with the benefits of automation that made available high-quality columns at a reasonable price. On the other hand, this means that further developments in column technology are dictated more by market forces or the possibility of reducing manufacturing costs than exploratory science.

## Column Types

Wall-coated open tubular columns (WCOT columns), or simply capillary columns, and classical packed columns dominate the practice of gas–liquid chromatography today. Porous layer open tubular columns (PLOT columns) and classical packed columns dominate the practice of gas–solid chromatography. WCOT columns are typically up to 100 m

long with internal diameters of capillary dimensions coated with a thin, and usually immobilized, film of stationary phase leaving an open interior passageway down the center of the column. PLOT columns are identical to WCOT columns with the liquid phase replaced by a layer of fine adsorbent particles. WCOT and PLOT columns are the first choice for analytical separations because of their superior peak capacity and chemical inertness. Classical packed columns are usually 0.5–3 m long with an internal diameter greater than 2 mm packed with adsorbent or liquid-coated support particles of 100–250  $\mu\text{m}$  diameter. Packed columns offer a lower-cost alternative for some applications, are easier to use, are relatively tolerant of labile and involatile sample components, and are better suited to isolating preparative-scale quantities of materials. Only a limited number of stationary phases have been immobilized successfully in WCOT columns compared to the larger number and variety of stationary phases available for use in packed columns. Packed columns are generally preferred for the measurement of physicochemical properties of liquids and solids suitable for use as a stationary phase.

Micropacked, packed capillary, and support coated open tubular (SCOT) columns are rarely used today, having been superseded by developments in WCOT and PLOT column technology. Micropacked columns have diameters less than 1 mm and a similar packing density to classical packed columns. Packed capillary columns have an internal diameter less than 0.6 mm and are packed with particles of 5–20  $\mu\text{m}$  diameter. SCOT columns are capillary columns containing a liquid phase coated on a surface covered

allows the response of two detectors to be acquired, where at least one of these will be a specific detector.

*See also:* **Chromatography:** Overview; Principles. **Derivatization of Analytes. Gas Chromatography:** Overview; Multidimensional Techniques; High-Speed Techniques; Instrumentation; Detectors.

## Further Reading

- Adams RP (1995) *Identification of Essential Oil Components by Gas Chromatography/Mass Spectrometry*. Carol Stream, IL: Allured Publishing Corp.
- Braithwaite A and Smith FJ (1985) *Chromatographic Methods*, 4th edn. London: Chapman and Hall.

- Handley AJ and Adlard ER (2001) *Gas Chromatographic Techniques and Applications*. Sheffield: Sheffield Academic Press.
- Heftmann E (2004) *Chromatography: Fundamentals and Applications of Chromatography and Related Differential Migration Methods*, 6th edn. Part A: Fundamentals and Techniques. Part B: Applications. Amsterdam: Elsevier.
- Hinshaw J (1995) GC in the fast lane. *LC GC-Mag. Sep. Sci.* 13: 944–949.
- Jennings W (1987) *Analytical Gas Chromatography*. Orlando: Academic Press.
- Mondello L, Lewis AC, and Bartle KD (2001) *Multidimensional Chromatography*. Chichester: Wiley.
- Poole CF (2003) *The Essence of Chromatography*. Amsterdam: Elsevier.

## Column Technology

**C F Poole**, Wayne State University, Detroit, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

One of the major changes in the practice of gas chromatography over the last decade or so was the development of columns from being a highly skilled laboratory preparation procedure to a general laboratory supply item. Few laboratories prepare their own columns today and the general skills to do so are largely lost outside of the manufacturing facilities of the main suppliers of columns to the laboratory market. This became possible because of a gradual evolution in the understanding of the unit operations of column preparation procedures and their simplification together with the benefits of automation that made available high-quality columns at a reasonable price. On the other hand, this means that further developments in column technology are dictated more by market forces or the possibility of reducing manufacturing costs than exploratory science.

## Column Types

Wall-coated open tubular columns (WCOT columns), or simply capillary columns, and classical packed columns dominate the practice of gas–liquid chromatography today. Porous layer open tubular columns (PLOT columns) and classical packed columns dominate the practice of gas–solid chromatography. WCOT columns are typically up to 100 m

long with internal diameters of capillary dimensions coated with a thin, and usually immobilized, film of stationary phase leaving an open interior passageway down the center of the column. PLOT columns are identical to WCOT columns with the liquid phase replaced by a layer of fine adsorbent particles. WCOT and PLOT columns are the first choice for analytical separations because of their superior peak capacity and chemical inertness. Classical packed columns are usually 0.5–3 m long with an internal diameter greater than 2 mm packed with adsorbent or liquid-coated support particles of 100–250  $\mu\text{m}$  diameter. Packed columns offer a lower-cost alternative for some applications, are easier to use, are relatively tolerant of labile and involatile sample components, and are better suited to isolating preparative-scale quantities of materials. Only a limited number of stationary phases have been immobilized successfully in WCOT columns compared to the larger number and variety of stationary phases available for use in packed columns. Packed columns are generally preferred for the measurement of physicochemical properties of liquids and solids suitable for use as a stationary phase.

Micropacked, packed capillary, and support coated open tubular (SCOT) columns are rarely used today, having been superseded by developments in WCOT and PLOT column technology. Micropacked columns have diameters less than 1 mm and a similar packing density to classical packed columns. Packed capillary columns have an internal diameter less than 0.6 mm and are packed with particles of 5–20  $\mu\text{m}$  diameter. SCOT columns are capillary columns containing a liquid phase coated on a surface covered

with a layer of solid support material, leaving an open passageway through the center of the column. Some characteristic properties of the different column types are summarized in **Table 1**. The most significant difference in kinetic properties is a result of differences in relative permeability. Open tubular columns provide much lower flow resistance and consequently can be used in longer lengths to obtain very high plate numbers compared with packed columns. The optimum mobile-phase velocity for open tubular columns is shifted to higher velocities compared with packed columns favoring faster separations.

### Column Properties

Narrow bore and thin-film WCOT columns afford the highest intrinsic efficiency per unit length and are selected for fast chromatography (**Table 2**). Since resolution increases only as the square root of the plate number, and also the column length, large values for the plate number are required for difficult separations. Such large numbers are available in gas chromatography, albeit at the expense of separation time, allowing separations to be achieved with only minimal differences in selectivity. In contrast to other chromatographic methods, separations in gas chromatography are often achieved through kinetic

optimization, allowing many separations to be obtained on a limited number of stationary phases.

At a given temperature, the partition coefficient is constant and the observed retention factor will depend on the phase ratio. The latter is defined as the ratio of the column volume accessible to the mobile phase and the volume of stationary phase. Columns with a large phase ratio provide small retention factors for volatile compounds and require inconveniently large plate numbers to provide adequate separations. Columns with a low phase ratio, that is, thick film columns, have a lower intrinsic efficiency than thin film columns, but provide better resolution of volatile compounds, because they provide more favorable retention factors. They also allow separations to be performed in a higher and more convenient temperature range than is possible with thin film columns. For high boiling compounds, columns with a low phase ratio are not useful because they lead to long separation times. Increasing the phase ratio by reducing the film thickness lowers the retention factors to a value within the optimum range so that there is little deterioration in resolution and faster separations are obtained. Packed columns have low phase ratios compared to WCOT columns. For compounds of moderate and low volatility separation times are longer.

**Table 1** Representative properties of different column types for gas chromatography

Column type	Phase ratio	$H_{\min}$ (mm)	$u_{\text{opt}}$ (cm s <sup>-1</sup> )	Relative permeability
Classical packed	4–200	0.5–2	5–15	1–50
Micropacked	50–200	0.02–1	5–10	1–100
Packed capillary	10–300	0.05–2	5–25	5–50
SCOT	20–300	0.5–1	10–100	200–1000
WCOT	15–500	0.03–0.8	10–100	300–20 000

$H_{\min}$  is the minimum plate height at the optimum mobile phase velocity  $u_{\text{opt}}$ .

**Table 2** Characteristic properties of some representative columns

Column type	Length (m)	Internal diameter (mm)	Film thickness (μm)	Phase ratio	Retention factor (k)	$H_{\min}$ (mm)	Column plate number	Plates per meter
Classical	2	2.16	10%(w/w)	12	10.4	0.55	3 640	1 820
Packed	2	2.16	5%(w/w)	26	4.8	0.50	4 000	2 000
WCOT	30	0.10	0.10	249	0.5	0.06	480 000	16 000
	30	0.10	0.25	99	1.3	0.08	368 550	12 285
	30	0.25	0.25	249	0.5	0.16	192 000	6 400
	30	0.32	0.32	249	0.5	0.20	150 000	5 000
	30	0.32	0.50	159	0.8	0.23	131 330	4 380
	30	0.32	1.00	79	1.6	0.29	102 080	3 400
	30	0.32	5.00	15	8.3	0.44	68 970	2 300
	30	0.53	1.00	132	0.9	0.43	70 420	2 340
	30	0.53	5.00	26	4.8	0.68	43 940	1 470

$H_{\min}$  is the minimum plate height and  $k$  is for undecane at 130°C.



## Column Preparation (Open Tubular Columns)

### Drawing Columns of Capillary Dimensions

Today, virtually all open tubular columns are prepared from fused silica and to a lesser extent borosilicate and soda-lime glasses. Fused silica is prepared by the oxidation of pure silicon tetrachloride in an oxygen plasma and is essentially pure silica, containing less than 1.00 ppm of metal impurities. The surface of fused silica glass is relatively inert, containing primarily silanol and siloxane groups. The presence of silanol groups is responsible for the residual acidic character of the glass. Because of its high melting point and ease of fracture, columns of capillary dimensions are drawn in a clean room facility at a rate of about  $1 \text{ ms}^{-1}$  from wide-bore silica tubes on a special apparatus, similar to those used for manufacturing optical fibers. The thin-walled tube must be protected immediately from moist air or dust particles that promote the growth of fissures or cracks. This is done by coating the outside of the capillary tube as it emerges from the softening furnace with a protective film of polyimide or aluminum. The drawn capillary column is inherently straight but sufficiently flexible to be coiled on a spool for collection purposes.

The thermal stability of the polyimide protective coating limits the application of fused silica capillaries for high-temperature applications. Typical polyimide coatings are stable to  $\sim 350^\circ\text{C}$  with some newer materials claimed to have a reasonable service lifetime at  $400^\circ\text{C}$ . Aluminum has excellent high-temperature stability but provides only modest protection of the fused silica. Fused silica capillaries are also limited for use in portable instruments and other applications where higher mechanical strength is desirable. Stainless steel capillary columns with an integral lining of silicon or fused silica provide the benefits of high-temperature stability and unsurpassed mechanical strength while approaching the low adsorptive and catalytic activity associated with fused silica. The fused silica layer of  $\sim 0.4 \mu\text{m}$  is prepared by the decomposition of an inorganic polymer in air and is intimately bonded to the oxidized metal surface.

Soda-lime and borosilicate glass capillary columns can be made in the laboratory using a glass drawing machine. The glass tube is fed into an electrically heated softening furnace at a controlled speed and drawn out at the lower end by a pair of motor-driven rollers. Capillary columns of different dimensions are drawn by maintaining a fixed differential between the feed rate of the stock tube to the furnace and the

pull rate of the softened tube at the exit of the furnace. The capillary tubing emerging from the softening oven is coiled by passage through a heated bent pipe. Unlike fused silica columns, these columns are relatively inflexible and require end straightening before installation in the gas chromatograph. The latter requires some skill. Altogether they are more difficult to handle and use than fused silica columns and require more extensive chemical treatment prior to coating with stationary phases. On account of these difficulties they are not as widely used as fused silica columns.

### Surface Treatments

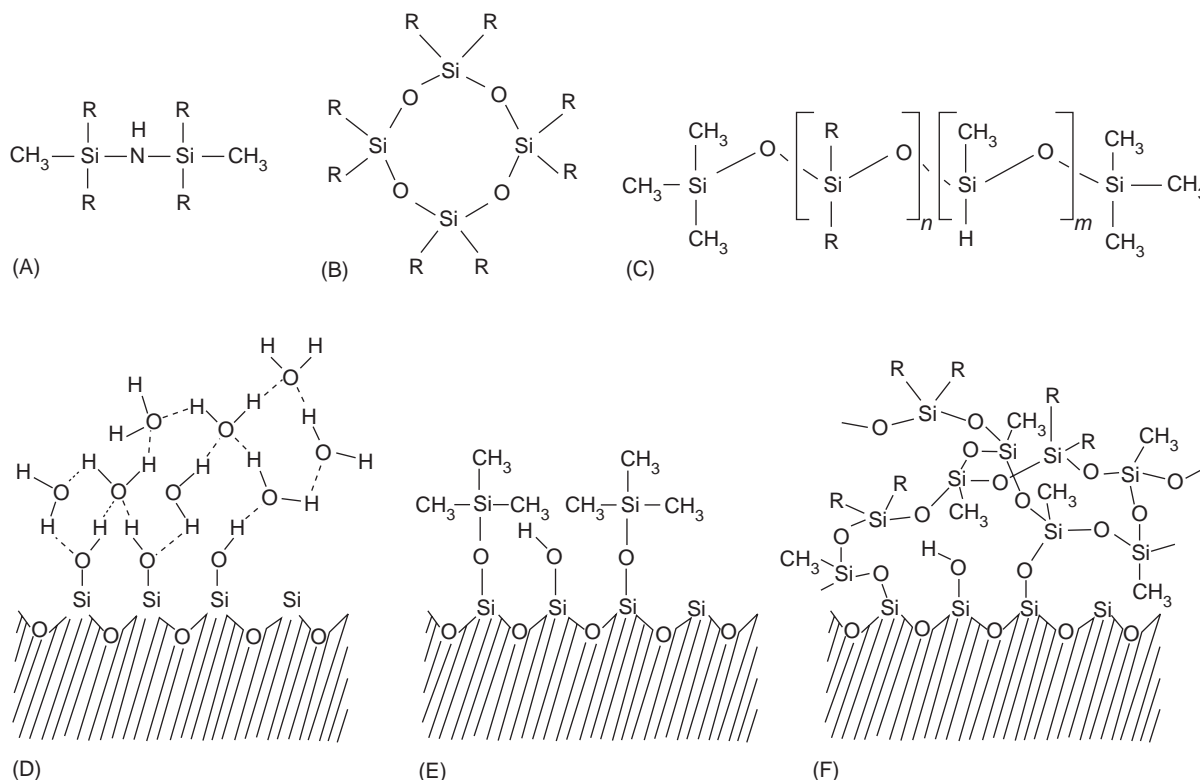
During the drawing process the fused silica is subject to a range of temperatures resulting in a surface with an undefined silanol concentration that is also likely contaminated by trapped acidic impurities. Conditioning and hydrothermal treatment (acid leaching) are generally required to remove acid impurities and provide a uniform and defined silanol concentration for column preparation. These procedures are usually performed in the laboratory and are essential to minimize batch-to-batch variation in column properties.

The energy of smooth glass surfaces can be increased by chemical modification or roughening. Chemical modification is virtually the only approach used for fused silica columns while roughening and/or chemical modification is used with other glasses. The most important reactions are leaching with aqueous hydrochloric acid, etching by gaseous hydrogen chloride, whisker formation by hydrogen fluoride etching, and solution deposition of a layer of solid particles. Acid leaching under relatively mild conditions is used to optimize the concentration of surface silanol groups for uniform deactivation and possible bonding of the stationary phase to the fused silica surface. More aggressive conditions are often used with other glasses to form a silica-rich surface denuded of metal ions. This greatly diminishes the effects of glass variety and composition on subsequent treatments. Etching, a different process to leaching, is performed with dry hydrogen chloride gas and almost exclusively with soda-lime glass capillaries. This results in the formation of a regularly spaced array of sodium chloride crystals on the inner surface of the column. Whisker surfaces of different density are prepared by increasingly more aggressive reactions between hydrogen fluoride and the glass surface. A light etch is achieved by coating the column with a thin film of potassium or ammonium bifluoride solution and subsequent thermal treatment. For a high-density etch, higher temperatures and reaction with hydrogen fluoride gas, usually generated by the

*in situ* thermal decomposition of a film of a fluorine-containing ether is generally used. The area of whisker modified surfaces is about a 1000-fold higher than the smooth glass surface and consists of a densely packed layer of filamentary projections whose size and length depend on the experimental conditions. Whisker surfaces can be coated with almost any stationary phase but are very active, difficult to deactivate, and are intrinsically less efficient than columns prepared with less radical surface modifications. Particle deposition methods for surface modification rely upon the production of a uniform layer of fine particles obtained by coating with a stable sol, in the case of sodium chloride, or reaction between carbon dioxide and a thin film of barium hydroxide solution, in the case of barium carbonate. General problems with particle deposition layers are the difficulty of further deactivation and the lower intrinsic column efficiency compared with other methods.

Surface modification reactions used to improve the wettability of glasses by the stationary phase and to deionize some glasses result in an increase in the activity of the glass surface. Without subsequent deactivation, columns coated with low-polarity

stationary phases exhibit poor chromatographic properties manifested as tailing and even incomplete elution or decomposition of polar solutes. No universal deactivation method exists, but some techniques have emerged as more useful than others. The most widely used methods include high-temperature silylation using disilazanes or cyclosiloxanes, reaction with poly(alkylhydrosiloxanes), or reaction with the thermal degradation products of poly(siloxane) or poly(ethylene glycol) stationary phases (**Figure 1**). The R groups of the deactivating reagent, or a fraction of them, are generally chosen to match the substituents attached to the backbone of the stationary phase. Not all aspects of the deactivation mechanisms are understood. Although, there is probably a high commonality in the way the different reagents react. Above  $\sim 400^{\circ}\text{C}$  several alternative reactions compete with simple silylation. These include formation of siloxane bonds by condensation of neighboring silanol groups; nucleophilic displacement of the organic group from the surface-bonded reagent by a neighboring silanol group; and reactions with surface adsorbed water, forming short-chain polymers bonded to the surface. These reactions result in



**Figure 1** Reagents used for the deactivation of silanol groups on glass surfaces. A, disilazane; B, cyclic siloxane; and C, poly(alkylhydrosiloxane) in which R is usually a complementary group to the stationary phase (e.g., methyl, cyanopropyl, etc.). The lower portion of the figure provides a view of the surface of fused silica with adsorbed water (D), after deactivation with a trimethylsilylating reagent (E), and after treatment with a poly(alkylhydrosiloxane) reagent (F). (From Poole CF and Poole SK (1991) *Chromatography Today*. Amsterdam: Elsevier. With permission.)



a decrease in the number of silanol groups on the deactivated surface and the formation of a protective umbrella-type film, restricting access to the remaining unreacted silanol groups. Deactivation with poly(alkylhydrosiloxanes) occurs primarily by condensation with silanol groups on the fused silica surface with formation of surface-to-polymer bonds. In addition, physically adsorbed water hydrolyzes silane bonds to silanols, which react further with silane groups or other silanol groups to form a highly cross-linked resin film.

### Coating Techniques

The object of all coating techniques is the distribution of the stationary phase as a thin, even film that completely covers the glass surface. Dynamic coating and static coating methods, with several variations, are used for this purpose. Dynamic coating is commonly used for applying reagents for surface modification reactions, but since it does not allow control of the film thickness, it is not as widely used for coating the stationary phase. Dynamic coating is the only possibility for coating stationary phases of low viscosity. For dynamic coating, a suitable reservoir is charged with a solution of the stationary phase (or reagent) in a volatile solvent, from which a plug of solution is forced by gas pressure into the capillary, occupying perhaps 25% of the capillary length. The plug is then pushed through the column by gas pressure at a constant velocity aided by a buffer capillary attached to the outlet end of the column to be coated. When the plug has exited the column to be coated, the buffer capillary is disconnected, and the gas pressure increased to evaporate the solvent.

The static coating method yields efficient columns of predictable film thickness with gum or solid stationary phases. The column is filled entirely with a dilute solution of the stationary phase in a volatile solvent, which is then evaporated by sealing one end of the column and attaching the other to a vacuum source. The solvent is evaporated under quiescent conditions, leaving behind a thin film of stationary phase with a thickness that can be calculated from the column radius and concentration of the coating solution. The principal practical problem with the static coating method is the breakthrough of gas bubbles when the column is placed under vacuum or due to fluctuations of the column temperature during coating. The small cross-section of capillary columns means that the static coating process is relatively slow. Optimization of the coating solvent and the use of elevated temperatures are important for accelerating the process.

### Sol-Gel Techniques

Failure of any step in the multistep column preparation procedure discussed above is the main cause of poor column quality. Sol-gel technology affords a simpler and alternative approach for column preparation that is increasingly popular for self-prepared columns. The main advantage of the sol-gel approach is that it combines the individual steps associated with conventional column preparation into a single and reproducible process. Typically, the sol solution contains an alkoxide-based precursor, a hydroxy-terminated stationary phase, a surface deactivating reagent (if required), and a catalyst dissolved in a suitable solvent. The column is filled with the sol solution, and after an incubation period, simply expelled from the column, which is then dried, conditioned, and rinsed with solvent prior to use. The key sol-gel reactions involved in the coating procedure are: (1) catalytic hydrolysis of the alkoxide precursor; (2) condensation of the hydrolyzed products into a three-dimensional sol-gel network; (3) chemical bonding of the hydroxy-terminated stationary phase to the evolving sol-gel network; and (4) bonding of the evolving sol-gel polymer to the interior wall of the capillary. The stationary phase in this case differs from a conventionally coated liquid film column in two important aspects. The stationary phase is an inorganic-organic polymer with a structure that is different to a typical organic polymer and the stationary phase layer has a highly roughened surface of greater surface area than liquid film columns.

### Column Preparation (Packed Columns)

Packed columns, unlike WCOT columns, are easily prepared in the laboratory and many analysts prefer to prepare their own columns. Almost any liquid of low vapor pressure can be employed extending the scope of packed columns to stationary phases unavailable as WCOT columns, as well as allowing gas chromatography to be used to determine physicochemical properties of a wide range of materials evaluated as stationary phases.

### Supports

An ideal support would act like a sponge, holding the liquid phase as a thin film, without playing any further role in the separation process. Although no such ideal support exists, the diatomaceous earths provide a reasonable compromise of properties (Table 3). Diatomite (diatomaceous earth) is composed of the skeletons of diatoms, single-celled algae, which have accumulated in large beds in various parts of the

**Table 3** Characteristic properties of Chromosorb supports

Property	P	W	G	A	750
Color	Pink	White	Oyster	Pink	White
Apparent pH	6.5	8.5	8.5	7.1	8.0
Free fall density (g ml <sup>-1</sup> )	0.38	0.18	0.47	0.40	0.37
Packed density (g ml <sup>-1</sup> )	0.47	0.24	0.58	0.48	0.36
Surface area (m <sup>2</sup> g <sup>-1</sup> )	4.0	1.0	0.5	2.7	0.5–1.0
Maximum liquid loading (% w/w)	30	15	5	25	12

world. The skeletal material is essentially amorphous silica with small amounts of alumina and metallic oxide impurities. In their native form the diatomite skeletons are too small and fragile to be used as supports. Calcining at temperatures in excess of 900°C is used to agglomerate and strengthen the natural material. The presence of iron oxides gives this material its characteristic pink color. Calcining in the presence of a small amount of sodium carbonate results in the formation of a white material with most of the pore structure destroyed and metal impurities converted to colorless sodium silicates. Thus, the white material has a smaller surface area and is slightly basic when compared with the pink material. Pink supports are relatively hard, have a high packing density, large surface area, and high loading capacity. They are used in both analytical and preparative-scale gas chromatography. White supports are more friable, less dense, have a smaller surface area and loading capacity, and are used in analytical gas chromatography. For analytical separations a support with a mesh range of 80–100 (149–177 µm) or 100–120 (125–149 µm) is a reasonable compromise between column performance and permeability.

Undesirable chromatographic properties, such as tailing, adsorption, and decomposition, are associated with the presence of metallic impurities and silanol groups at the surface of the diatomaceous supports. Acid and/or base washing is used to remove metal impurities and silanization to cap surface silanol groups. Silanization changes the character of the support from hydrophilic to hydrophobic. Consequently, silanized supports are used primarily with low-polarity stationary phases. They are unsuitable for coating with polar stationary phases. The treatment undergone by the diatomaceous support is usually stated on the manufacturer's label, e.g., Chromosorb W-AW-DMCS indicates that the white support was acid-washed and silanized with dimethyldichlorosilane. Addition of small quantities of 'tailing reducers' is sometimes necessary for successful chromatography of strong acids and bases. Tailing reducers are coated onto the support in a similar manner to the stationary phase or with the stationary phase, and to be effective, they must be stronger

acids or bases than the compounds to be separated. Examples include potassium hydroxide and poly(ethyleneimine) for the separation of bases and phosphoric acid for the separation of acids.

Other support materials of minor importance include various fluorocarbon powders, glass beads, carbon, and dendrite salts. Fluorocarbon powders are used primarily for the separation of reactive compounds that would destroy or be destroyed by other materials. A low surface energy and electrostatic properties make fluorocarbon powders difficult to coat and pack into columns. Because of their controlled shape, glass beads are used mainly for theoretical studies but have a low loading capacity and an active surface.

### Coating and Packing Techniques

Supports are usually coated with the stationary phase by any of several evaporation methods. For example, the stationary phase in a suitable solvent is mixed with the support and the solvent gently evaporated while stirring in an open dish under a heat lamp or slowly evaporated under vacuum from a fluted flask attached to a rotary evaporator. The damp support is then air-dried, oven-dried, or dried in a fluidized bed dryer. The latter is recommended because it is faster and leads to more permeable and efficient packed columns by stripping away fines produced in the coating process. The dried packing is then added in small increments to an empty column with a glass wool plug at its exit and consolidated with the aid of suction and gentle vibration. By convention, packings are prepared on a weight per weight basis and quoted as percent liquid phase. Typical phase loadings vary from 0.5% to 30% w/w depending upon the intended use for the column.

Stainless steel, nickel, or glass tubes bent into various shapes to fit the column oven of the gas chromatograph provide the container for the packings. Nickel (after acid passivation) and glass are the most inert column materials, and are preferred for the separation of labile compounds. Teflon or plastic tubing is also used occasionally, often for the separation of reactive compounds for which fluorocarbon powder supports are also required.

## Stationary Phases

Over the years thousands of substances have been used as stationary phases. For several reasons most of these have been abandoned in favor of a small number of liquids and adsorbents with favorable thermal stability and kinetic properties, complementary selectivity, reasonably well-defined and reproducible chemical composition, and if used in WCOT columns, the possibility of immobilization. Practical considerations dictate that liquid stationary phases should be inert, of low vapor pressure, have good coating characteristics, and have reasonable solubility in some common volatile organic solvent. The desirability of a wide temperature operating range tends to dictate that most common stationary phases are polymeric materials, although polymers are more likely to show greater composition variation than stoichiometric compounds.

Common liquid phases belong to one or other of the categories: (1) hydrocarbon and perfluorocarbon liquids, (2) ether and ester liquids, (3) ionic liquids, and (4) poly(siloxanes). Of these, poly(siloxanes) and poly(ethylene glycol) stationary phases dominate the practice of WCOT columns, because unlike most liquids they can be immobilized by simple chemical reactions to prepare films of different thickness that are stable to temperature variation and solvent rinsing while retaining favorable kinetic properties.

Some representative examples of packed column stationary phases belonging to categories (1)–(3) are summarized in Table 4. Hydrocarbons are used as low-selectivity liquid phases for separations of (mainly) low-polarity compounds by differences in their vapor pressure. Highly fluorinated liquid phases

are used for the speciation of fluorocarbon compounds or the separation of reactive compounds such as metal fluorides, interhalogen compounds, and nonmetal halides that tend to destroy conventional phases. The meta-linked poly(phenyl ethers) with five or six rings are weakly selective stationary phases with a defined chemical structure of exceptionally low volatility for their molecular weight. The term poly(ester) is used to describe a wide range of resinous composites derived from the reaction of a polybasic acid with a polyhydric alcohol. The most frequently used poly(ester) phases are the succinate and adipate esters of ethylene glycol, diethylene glycol, and butanediol. These phases provide high selectivity for the separation of polar compounds but have been replaced in many of their former applications by the more stable poly(cyanoalkylsiloxanes). Thermal-induced changes in composition, susceptibility to degradation by oxygen and water at elevated temperatures, and reaction with some polar compounds have contributed to their diminished use in recent years. The poly(ethylene glycols) are popular general-purpose liquid phases for the separation of volatile polar compounds. The poly(ethylene glycol) Carbowax 20M is one of the most popular phases for packed column gas chromatography, and in a modified form, is widely used for separations by WCOT columns. Ionic liquids are novel solvents composed entirely of ions. Those used for gas chromatography are alkylammonium, alkylphosphonium, or 1,3-dialkylimidazolium salts with weak nucleophilic anions such as sulfonates or tetrafluoroborate. These anions promote high thermal stability affording adequate liquid ranges for separation

**Table 4** Characteristic properties of some packed column liquid phases

Name and structure	Temperature range (°C)	
	Minimum	Maximum
Squalane 2,6,10,15,19,23-hexamethyltetracosane	<20	120
Apolane-87 $(C_{18}H_{37})_2CH(CH_2)_4C(C_2H_5)_2(CH_2)_4CH(C_{18}H_{37})_2$	30	280
Fomblin YR $-[OCF(CF_3)CF_2]_n[OCF_2]_m-$	30	<255
PPE-5 $C_6H_5O(C_6H_5O)_3C_6H_5$	20	200
Diethyl phthalate $C_6H_4(COOC_2H_5)_2$	<20	160
EGS $HO(CH_2)_2[OOCCH_2CH_2COO(CH_2)_2]_nOH$	100	210
DEGS $HO(CH_2)_2O(CH_2)_2[OOCCH_2CH_2COO(CH_2)_2O(CH_2)_2]_nOH$	20	200
Carbowax 20M $HO(CH_2CH_2O)_nCH_2CH_2OH$	60	225
FFAP	50	250
1,2,3-Tris(2-cyanoethoxy)propane $(CH_2OCH_2CH_2CN)_3$	20	170
Tetrabutylammonium perfluorooctanesulfonate	<20	220
Tetrabutylammonium 4-toluenesulfonate	55	200
Tetrabutylammonium tetrafluoroborate	162	290
Ethylammonium 4-toluenesulfonate	121	220
Tetrabutylphosphonium chloride	83	230

PPE-5, poly(phenyl ether); EGS, poly(ethylene glycol succinate); DEGS, poly(diethylene glycol succinate); Carbowax 20M, poly(ethylene glycol); and FFAP, Carbowax 20M treated with 2-nitroterephthalic acid.

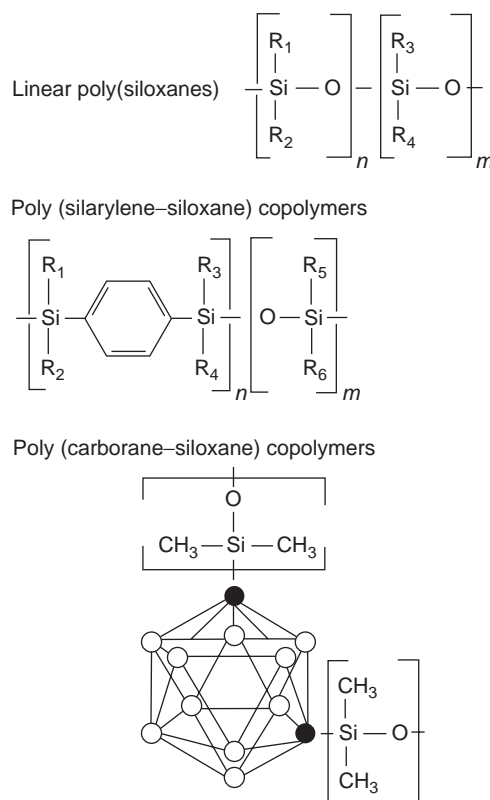
purposes. Their separation characteristics are complementary to polar nonionic liquid phases.

### Poly(siloxane) Liquid Phases

The poly(siloxanes) are the most popular stationary phases for both WCOT and packed column gas chromatography. Favorable features include: a wide temperature operating range; acceptable diffusion and solubility characteristics for different solute types; chemical inertness; good film forming properties; ease of synthesis with a wide range of chromatographic selectivity; and ease of immobilization for applications employing WCOT columns. The poly(siloxanes) used in gas chromatography are generally linear polymers of the type shown in **Figure 2**, in which the R group can be a number of different substituents such as alkyl, vinyl, phenyl, 3,3,3-trifluoropropyl, or 3-cyanopropyl. By varying the identity and relative amount of each R group as well as the ratio of monomer units, polymers with a wide range of solvent properties can be prepared.

The gradual evolution of column technology revealed that the low-to-medium viscosity poly(siloxane) oils preferred as packed column stationary phases (**Table 5**) were unsuitable for preparing WCOT columns of high thermal stability. Gum phases of generally higher molecular weight and viscosity are required for the preparation of WCOT columns with stable films resistant to disruption at elevated temperatures. The realization that cross-linking of gum phases to form a rubber affords a means of further stabilizing the poly(siloxane) films

without destroying their favorable diffusion characteristics was a further important breakthrough in WCOT column technology.



**Figure 2** General structures of poly(siloxane) stationary phases for gas chromatography. (Reprinted with permission from LCGC, vol. 15, no. 11, Nov. 1997, p. 1024. LCGC is a copyrighted publication of Advanstar Communications Inc. All rights reserved.)

**Table 5** Characteristic properties of some poly(siloxane) liquid phases for packed column gas chromatography

Name	Structure	Viscosity (cP)	Average molecular weight	Temperature operating range (°C)	
				Minimum	Maximum
OV-1	Dimethylsiloxane	Gum	> 10 <sup>6</sup>	100	350
OV-101	Dimethylsiloxane	1 500	30 000	< 20	350
OV-7	Phenylmethyldimethylsiloxane	500	10 000	< 20	350
OV-17	80% methyl and 20% phenyl	1 300	40 000	< 20	350
	Phenylmethylsiloxane				
OV-25	50% methyl and 50% phenyl	> 100 000	10 000	< 20	300
	Phenylmethyldiphenylsiloxane				
OV-210	25% methyl and 75% phenyl	10 000	200 000	< 20	275
	Trifluoropropylmethylsiloxane				
OV-225	50% methyl and 50% 3,3,3-trifluoropropyl	9 000	8 000	< 20	250
	Cyanopropylmethylphenylmethylsiloxane				
Silar 7CP	50% methyl, 25% phenyl, and 25% 3-cyanopropyl	20 000	5 000	50	250
	Cyanopropylphenylsiloxane				
OV-275	75% 3-cyanopropyl and 25% phenyl	20 000	5 000	50	250
	Di(cyanoalkyl)siloxane				
Silar 10CP	70% 3-cyanopropyl and 30% 2-cyanoethyl				
	Di(3-Cyanopropyl)siloxane			50	250

Two approaches are commonly used to immobilize poly(siloxanes) for the preparation of WCOT columns. These methods are: (1) thermal immobilization of silanol-terminated nonpolar and moderately polar poly(siloxanes) and (2) radical initiated cross-linking of endcapped poly(siloxanes) of a wider polarity range. Thermal induced reactions result in simultaneous bonding of the polymer to the column wall and the formation of cross-links between polymer chains. The use of silanol-terminated poly(siloxanes) was instrumental in the development of columns for high-temperature gas chromatography (Table 6). Free radical cross-linking of poly(siloxanes) requires the presence of a suitable catalyst (peroxide, azo-compound, or  $\gamma$ -radiation) as a free radical generator. In this case, cross-linking occurs through the formation of carbon-carbon bonds involving the organic substituents on neighboring poly(siloxane) chains. Only a low level of cross-linking (0.1–1%) is required for adequate immobilization. Increasing substitution of methyl groups by bulkier and/or polar substituents along the polymer backbone decreases the efficiency of radical cross-linking. For this reason, moderately polar poly(siloxane) stationary phases are synthesized with small amounts of vinyl, tolyl, or octyl groups attached to the siloxane backbone or as endcapping groups to increase the success of the cross-linking reaction. Even so, for poly(siloxanes) with a high percentage of polar monomers, immobilization may be incomplete and these phases are generally referred to as ‘stabilized’ in the trade literature. Stabilized columns usually have greater thermal stability than physically coated columns but are not resistant to solvent rinsing or as durable as columns containing immobilized stationary phases.

Another general advantage of immobilization is that it enables thick-film columns to be prepared to optimize the phase ratio for different applications. It is difficult to prepare conventionally coated columns with films thicker than  $\sim 0.5\ \mu\text{m}$ . Immobilized films of 1–8  $\mu\text{m}$  are easily prepared using cross-linking reactions.

The thermal stability of the poly(siloxanes) is enhanced by inserting phenyl groups (silarylene copolymers) or carborane groups into the polymer backbone (see Figure 2). Thermal degradation of poly(siloxanes) occurs by release of small ring cyclic siloxanes. These are formed in an intramolecular displacement reaction in which the silanol end group attacks a silicon atom in the same chain releasing the cyclic siloxane and forming a new silanol group. Incorporating phenyl or carborane groups in the polymer backbone inhibits this process by reducing the flexibility of the polymer chains. In addition, alternate and symmetrical substitution on the polymer backbone results in poly(siloxanes) of higher thermal stability.

## Classification of Liquid Phases

Various scales of solvent strength (polarity) and selectivity have been used to classify stationary phases. Classification based on polarity had to be abandoned because of the lack of a working definition. There is no substance that is uniquely polar and suitable to probe the polarity of other substances. Selectivity is defined as the relative capacity of a stationary phase for specific intermolecular interactions, such as dispersion, induction, orientation, and hydrogen-bond formation. Early attempts at a systematic definition of selectivity scales were based on the system of phase constants introduced by Rohrschneider and

**Table 6** Guide to the temperature operating range for bonded poly(siloxane) phases in WCOT columns

Type	Temperature range ( $^{\circ}\text{C}$ )		High-temperature version
	Minimum	Maximum	
Dimethylsiloxane	–60	325	420
Dimethyldiphenylsiloxane (5% diphenyl)	–60	325	420
Dimethyldiphenylsiloxane (35% diphenyl)	40	300	340
Dimethyldiphenylsiloxane (50% diphenyl)	40	325	390
Methylphenylsiloxane	0	280	
Dimethyldiphenylsiloxane (65% diphenyl)	50	260	370
3,3,3-Trifluoropropylmethylsiloxane (50% trifluoropropyl)	45	240	300
3-Cyanopropylphenyldimethylsiloxane (6% cyanopropylphenyl and 84% dimethyl)	20	280	
3-Cyanopropylphenyldimethylsiloxane (25% cyanopropylphenyl and 75% dimethyl)	40	240	
3-Cyanopropylphenyldimethylsiloxane (50% cyanopropylphenyl and 50% dimethyl)	40	230	
3-cyanopropyl-silphenylene co-polymer (equivalent to 70% dicyanopropyl)			290
Poly(ethylene glycol)	20	250	280
FFAP	40	250	



subsequently modified by McReynolds. McReynolds' phase constants are still indicated by many column manufacturers but are not a useful guide for column selection. Modern approaches to stationary phase classification by selectivity ranking are based on the cavity model of solvation. This model assumes that the transfer of a solute from the gas phase to solution in the stationary phase involves three steps. Initially, a cavity is formed in the stationary phase of the same size as the solute. The solute is then transferred to the cavity with reorganization of solvent molecules around the cavity and setup of solute-solvent interactions. In gas-liquid chromatography retention will depend on the cohesive energy of the stationary phase, represented by the free energy required for cavity formation, the formation of additional dispersion interactions of a solute-solvent type, and on selective solute-solvent polar interactions dependent on the complementary character of the polar properties of the solute and stationary phase. Quantitatively these interactions are described by the solvation parameter model set out below in the form suitable for gas-liquid chromatography:

$$SP = c + eE + sS + aA + bB + lL \quad [1]$$

where SP is the retention factor,  $\log k$ , or the gas-liquid partition coefficient,  $\log K_L$ . The remainder of the equations is made up of product terms called system constants ( $e$ ,  $s$ ,  $a$ ,  $b$ ,  $l$ ) and solute descriptors ( $E$ ,  $S$ ,  $A$ ,  $B$ ,  $L$ ). Each product term represents a contribution from a defined intermolecular interaction to the solute property. The solute descriptors are known for ~4000 compounds with others available by estimation or from experiment. They are not of immediate interest to us here. The system constants contain the information of the stationary phase properties and provide an unambiguous means of classification. The  $e$  system constant refers to the ability of the stationary phase to interact with solute  $n$ - or  $\pi$ -electron pairs. The  $s$  system constant to the ability of the stationary phase to take part in dipole-type interactions. The  $a$  system constant is a measure of stationary phase hydrogen-bond basicity and the  $b$  system constant a measure of stationary phase hydrogen-bond acidity. The  $l$  system constant (in part) describes the contribution of cavity formation and dispersion interactions to retention and more specifically the ability of the stationary phase to separate members of a homologous series. The phase constants for any stationary phase can be determined through the method of multiple linear regression analysis by measurement of a retention property for a series of varied solutes with known solute descriptors.

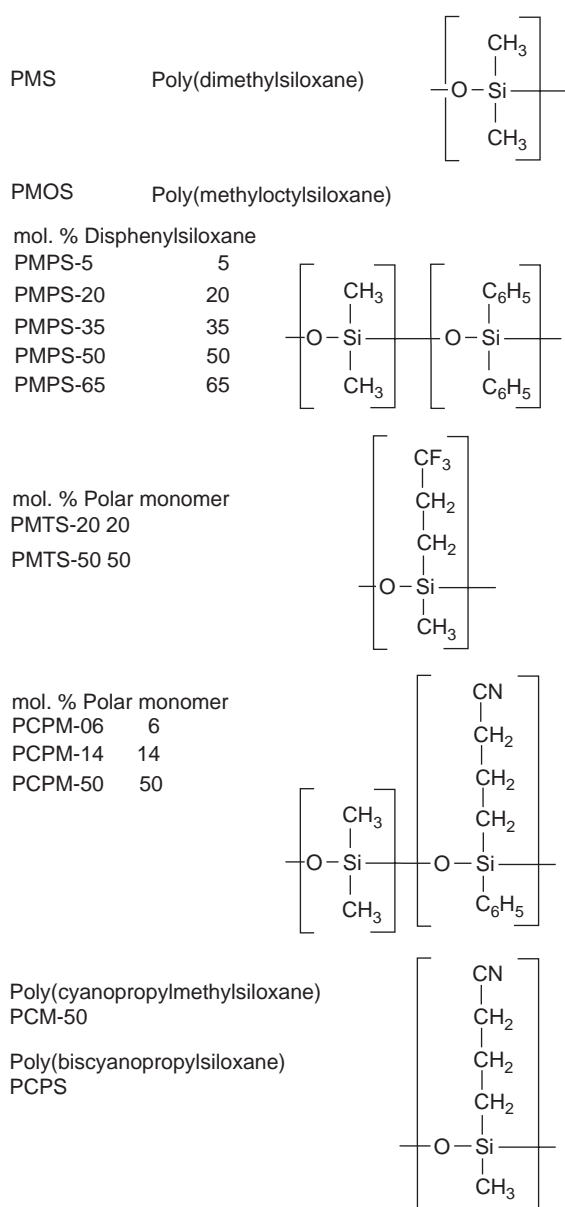
The system constants for a wide range of WCOT columns are summarized in Table 7 (this is part of a larger database containing systems constants at temperatures of 60°C, 80°C, 100°C, 120°C, and 140°C). The system constants are only loosely scaled to each other. Consequently, changes in magnitude in any column of Table 7 can be read directly, but changes in magnitude along rows must be interpreted more cautiously. Of immediate note is that none of the stationary phases are hydrogen-bond acids, and this interaction is, therefore, unimportant for classification of WCOT column stationary phases. The poly(siloxane) stationary phases (Figure 3) span a wide selectivity range. The poly(dialkylsiloxane) stationary phases are the least selective of the poly(siloxane) stationary phases, although compared to typical hydrocarbon phases, they possess a significant capacity for dipole-type and hydrogen-bond base interactions. Attempts to explain their separation characteristics based on nonselective interactions alone should be seen as an oversimplification. The incorporation of diphenylsiloxane monomer groups into the poly(dimethylsiloxane) backbone primarily increases the dipolarity/polarizability of the stationary phase with a small change in hydrogen-bond basicity and cohesion. The system constants change approximately linearly with composition up to ~50% diphenylsiloxane monomer and nonlinearly thereafter. The incorporation of 3,3,3-trifluoropropylmethylsiloxane monomer groups into the poly(dimethylsiloxane) backbone causes a significant change in the dipolarity/polarizability of the stationary phase with little alteration of the hydrogen-bond basicity. This results in a unique change in the  $s/a$  system constant ratio compared with other poly(siloxane) phases. These phases are also electron repulsive (negative  $e$  system constant) resulting from the presence of highly electronegative fluorine atoms. The incorporation of monomers containing cyanoalkyl groups into the poly(dimethylsiloxane) backbone increases the capacity of the poly(siloxanes) for dipole-type interactions as well as hydrogen-bond basicity. This results in a different range of  $s/a$  system constant ratios and a different selectivity to poly(siloxanes) containing 3,3,3-trifluoropropylmethylsiloxane monomer groups. For all poly(dimethylsiloxanes), incorporation of monomer groups containing polar functional groups results in an increase in the cohesion of the stationary phase, reflected in a smaller value for the  $l$  system constant. The poly(ethylene glycol) stationary phases complement the selectivity of poly(siloxane) stationary phases. They are the most hydrogen-bond basic of the stationary phases in Table 7 and have a significant capacity for dipole-type interactions. In



**Table 7** System constants for WCOT column stationary phases at 120°C (dependent variable log *k*)

General abbreviation	Column identity	System constants ( <i>b</i> = 0 in all cases)			
		<i>l</i>	<i>e</i>	<i>s</i>	<i>a</i>
<i>Poly(dimethylsiloxane)</i>					
PMS	DB-1	0.504	0	0.207	0.185
	SolGel-1	0.534	0	0.219	0.215
<i>Poly(methyloctylsiloxane)</i>					
PMOS	SPB-Octyl	0.615	0	0.232	0
<i>Poly(dimethyldiphenylsiloxane)</i>					
PMPS-5	DB-5	0.513	0	0.280	0.193
	HP-5	0.518	0	0.309	0.205
	OV-5	0.503	0	0.286	0.223
	SPB-5	0.504	0	0.293	0.212
	PTE-5	0.505	0	0.293	0.210
PMPS-20	Rt <sub>x</sub> -20	0.549	0	0.564	0.259
PMPS-35	DB-35	0.540	0	0.695	0.314
PMPS-50	HP-50 +	0.474	0.160	0.623	0.281
	Rt <sub>x</sub> -50	0.519	0.057	0.796	0.339
PMPS-65	Rt <sub>x</sub> -65	0.531	0.108	0.839	0.358
<i>Arylene-siloxane copolymer (nominally similar to HP-5)</i>					
AS-5	HP-5TA	0.595	0	0.350	0.284
<i>Poly(methyltrifluoropropylsiloxane)</i>					
PMTS-20	DB-200	0.464	− 0.340	1.010	0.203
PMTS-50	DB-210	0.439	− 0.343	1.278	0.077
<i>Poly(cyanopropylmethylsiloxane)</i>					
PCM-50	DB-23	0.438	0	1.537	1.468
<i>Poly(cyanopropylphenyldimethylsiloxane)</i>					
PCPM-06	DB-1301	0.547	− 0.057	0.498	0.450
PCPM-14	DB-1701	0.494	− 0.066	0.667	0.643
PCPM-50	DB-225	0.444	− 0.051	1.249	1.110
<i>Poly(cyanopropylsiloxane)</i>					
PCPS	SP-2340	0.418	0	1.993	1.960
<i>Poly(ethylene glycol)</i>					
PEG	HP-INNOWax	0.458	0.219	1.351	1.882
	HP-20M	0.452	0.209	1.335	2.014
	AT-Wax	0.440	0.225	1.318	1.889
	SolGel-Wax	0.495	0.201	1.440	2.031
NPEG	DB-FFAP	0.428	0.214	1.424	2.077
<i>Poly(siloxane) of unknown composition</i>					
VRX	DB-VRX	0.543	0	0.304	0.159
<i>Mixtures of poly(dimethylsiloxane) and poly(ethylene glycol)</i> <sup>a</sup>					
DX-1	DX-1	0.536	0	0.384	0.465
DX-3	DX-3	0.539	0.148	1.000	1.632
DX-4	DX-4	0.508	0.209	1.247	1.786
<i>Dissolved β-cyclodextrin derivatives</i> <sup>b</sup>					
Cyclodex B	Cyclodex B	0.544	− 0.067	0.759	1.030
Cyclosil-B	Cyclosil-B	0.545	− 0.065	0.682	1.082
Rt-βDEXsa	Rt-βDEXsa	0.484	− 0.195	1.071	1.070

<sup>a</sup>DX-1 = 10%, DX-3 = 50%, and DX-4 = 85% poly(ethylene glycol).<sup>b</sup>Cyclodex B (10.5% permethylated β-cyclodextrin in DB-1701); Cyclosil-B (30% heptakis[2,3-di-*O*-methyl-6-*O*-*t*-butyldimethylsilyl]-β-cyclodextrin in DB-1701); and Rt-βDEXsa (heptakis[2,3-di-*O*-acetoxy-6-*O*-*t*-butyldimethylsilyl]-β-cyclodextrin in a poly(cyanopropylphenyldimethylsiloxane) solvent containing 14% cyanopropylphenylsiloxane monomer).



**Figure 3** Structures and general abbreviations for WCOT column stationary phases.

addition, they also possess a significant capacity for electron lone pair interactions.

The system constants for packed column stationary phases are summarized in **Table 8**. Classification of their properties by cluster analysis results in the connection dendrogram shown in **Figure 4**. Stationary phases with similar solvation properties are located next to each other and connected close to the left-hand side of the dendrogram. Stationary phases with no paired descendents are singular phases with properties that cannot be duplicated by other phases from the data set (**Table 8**). Classification results in six groups with three phases behaving

independently. Group one contains squalane, Apolane-87, OV-3, OV-7, SE-30, and OV-105. These are phases of low cohesive energy with minimal capacity for polar interactions. The second group contains OV-22, OV-25, OV-11, OV-17, PPE-5, and DDP. These phases have low cohesive energy and are weakly dipolar and hydrogen-bond basic. QF-1 is loosely connected to this group but is significantly more dipolar and electron lone pair repulsive. The third group contains OV-330 and OV-225 with UH50B loosely connected to this group. Compared with the second group, these stationary phases are more dipolar and hydrogen-bond basic and slightly more cohesive. They represent an increase in the intensity of the same range of interactions as the larger group two of stationary phases. The fourth group contains H10 and PSF6 (**Figure 5**). These are strong hydrogen-bond acid stationary phases but in other respects quite different to each other. The fifth group contains the ionic liquids. Stationary phases in this group are significantly dipolar/polarizable and the strongest hydrogen-bond bases. The sixth group of stationary phases is divided into two subgroups. TCEP and OV-275 are strongly dipolar/polarizable, hydrogen-bond basic, and have high cohesive energy. EGAD, CW20M, and DEGS have a similar range of polar interactions, but not quite as intense, and have a lower cohesive energy than TCEP and OV-275. These phases are also weakly hydrogen-bond acidic but it is likely that this is an artifact due to alteration of the structure of the stationary phases during use.

## Liquid-Crystalline Stationary Phases

Liquid-crystalline stationary phases are used for the separation of positional and geometric isomers of rigid molecules. Liquid crystals exhibit the mechanical properties of a liquid while retaining some of the anisotropic properties of the solid state. This preservation of order permits shape selectivity, while the liquid properties result in acceptable chromatographic efficiency. Several hundred liquid-crystalline phases of different chemical types have been used in packed column gas chromatography. They all have in common a markedly elongated, rigid, rod-like structure, and generally have polar terminal groups. The most common types are Schiff bases, esters, azo, and azoxy compounds. Most of these materials provide limited opportunities for use in WCOT columns because of poor column efficiency, poor coating characteristics, poor thermal stability, high column bleed, and limited temperature operating ranges. These shortcomings were largely overcome by the development of side-chain liquid-crystalline poly

**Table 8** System constants for packed column stationary phases at 121°C

Stationary phase	System constant				
	<i>e</i>	<i>s</i>	<i>a</i>	<i>b</i>	<i>l</i>
<i>(i) Hydrocarbon phases</i>					
Squalane	0.129	0.011	0	0	0.583
Apolane-87	0.170	0	0	0	0.549
<i>(ii) Ether and ester phases</i>					
Poly(phenyl ether) 5 rings PPE-5	0.230	0.829	0.337	0	0.527
Carbowax 20M CW20M	0.317	1.256	1.883	0	0.447
Poly(ethylene glycol) Ucon 50 HB 660	0.372	0.632	1.277	0	0.499
1,2,3-Tris(2-cyanoethoxypropane) TCEP	0.116	2.088	2.095	0.261	0.370
Didecylphthalate DDP	0	0.748	0.765	0	0.560
Poly(ethylene glycol adipate) EGAD	0.132	1.394	1.820	0.206	0.429
Poly(diethylene glycol succinate) DEGS	0.230	1.572	2.105	0.171	0.407
<i>(iii) Ionic liquids</i>					
Tetrabutylammonium 4-toluenesulfonate QBAPTS	0.156	1.582	3.295	0	0.459
Tetrabutylammonium tris(hydroxymethyl)methyl-amino-2-hydroxy-1-propanesulfonate QBATAPSO	0.266	1.959	3.058	0	0.317
Tetrabutylammonium 4-morpholinepropanesulfonate QBAMPS	0	1.748	3.538	0	0.550
Tetrabutylammonium methanesulfonate QBAMES	0.334	1.454	3.762	0	0.435
<i>(iv) Poly(siloxane) phases</i>					
Poly(dimethylsiloxane) SE-30	0.024	0.190	0.125	0	0.498
Poly(dimethylmethylphenylsiloxane) (10 mol.% phenyl) OV-3	0.033	0.328	0.152	0	0.503
Poly(dimethylmethylphenylsiloxane) (20 mol.% phenyl) OV-7	0.056	0.433	0.165	0	0.510
Poly(dimethylmethylphenylsiloxane) (35 mol.% phenyl) OV-11	0.097	0.544	0.174	0	0.516
Poly(methylphenylsiloxane) OV-17	0.071	0.653	0.263	0	0.518
Poly(methylphenyldiphenylsiloxane) (65 mol.% phenyl) OV-22	0.201	0.664	0.190	0	0.482
Poly(methylphenyldiphenylsiloxane) (75 mol.% phenyl) OV-25	0.277	0.644	0.182	0	0.472
Poly(cyanopropylmethyltrimethylsiloxane) (10 mol.% cyanopropylmethylsiloxane) OV-105	0	0.364	0.407	0	0.496
Poly(cyanopropylmethylphenylmethylsiloxane) (50 mol.% cyanopropylmethylsiloxane) OV-225	0	1.226	1.065	0	0.466
Poly(dicyanoalkylsiloxane) (70 mol.% dicyanopropyl and 30 mol.% dicyanoethyl) OV-275	0.206	2.080	1.986	0	0.294
Poly(trifluoropropylmethylsiloxane) QF-1	−0.449	1.157	0.187	0	0.419
Poly(dimethylsiloxane)–poly(ethylene glycol) copolymer OV-330	0.104	1.056	1.419	0	0.481
PSF6	−0.360	0.820	0	1.110	0.540
<i>(v) Miscellaneous</i>					
Bis(3-allyl-4-hydroxyphenyl)sulfone H10	−0.051	1.323	1.266	1.457	0.418

(siloxane) phases. The biphenylcarboxylate ester poly(siloxane) (**Figure 6**) is the most widely used liquid-crystalline stationary phase for WCOT columns. It can be immobilized by standard methods, and has a useful temperature range from about 100°C to 250°C.

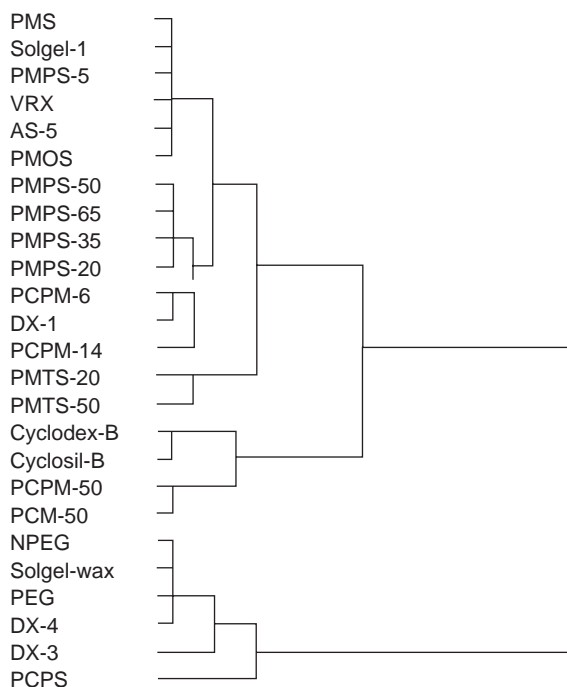
## Chiral Stationary Phases

The development of amino acid ester, dipeptide, and diamide chiral stationary phases provided the first indication that enantiomers could be separated by gas chromatography on stationary phases designed to promote multiple simultaneous interactions with an enantiomer. Real interest and progress in chiral separations resulted from the preparation of diamide

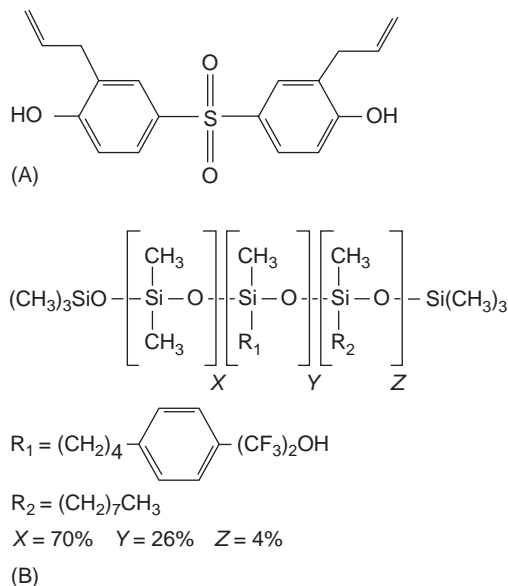
phases, grafted onto a poly(siloxane) backbone, suitable for coating on WCOT columns. The only commercially available example of this class of chiral selector is the poly(methylsiloxane) phase containing L-valine-*t*-butylamide (Chirasil-Val) (**Figure 7**), which permitted for the first time, a single separation of all the enantiomers of the common protein amino acids after derivatization.

## Cyclodextrin Derivatives

Initial attempts to use cyclodextrins and their derivatives as chiral stationary phases in gas chromatography met with limited success owing to their unfavorable melting points or decomposition

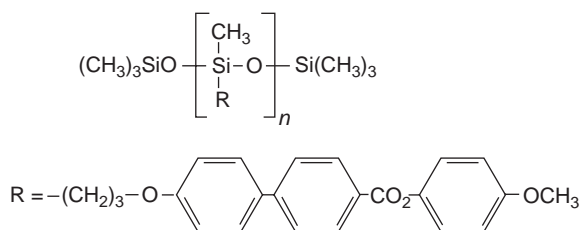


**Figure 4** Nearest neighbor complete link cluster dendrogram for the stationary phases in **Table 8**.

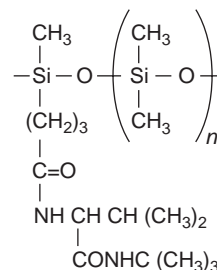


**Figure 5** Structures of the hydrogen-bond acid stationary phases H10 (A) and PSF6 (B).

temperatures. This changed with the discovery that some peralkylated derivatives (e.g., pentyl), partially alkylated derivatives, and mixed alkylated and acylated derivatives were viscous liquids that could be coated on deactivated glass surfaces (**Table 9**). Some 85% of enantiomers separated by gas



**Figure 6** Structure of the SB-smectic side-chain liquid-crystal line stationary phase (R is ~50% liquid crystal substituents).



**Figure 7** Structure of Chirasil-Val.

chromatography on derivatized cyclodextrins can be separated on one of three derivatized cyclodextrins: octakis(3-*O*-trifluoroacetyl-2,6-di-*O*-*n*-pentyl)- $\gamma$ -cyclodextrin; heptakis(2,6-di-*O*-methyl)- $\beta$ -cyclodextrin; and heptakis(2,6-di-*O*-*n*-pentyl)- $\beta$ -cyclodextrin. Octakis(3-*O*-butyryl-2,6-di-*O*-*n*-pentyl)- $\gamma$ -cyclodextrin also has a broad application range. It is generally accepted that enantioselectivity results from either inclusion formation and/or surface interactions controlled by the structure of the cyclodextrin derivatives. So far no simple method has emerged to identify the optimum stationary phase for a separation from enantiomer structure.

The general use of cyclodextrin derivatives as stationary phases has declined in recent years in favor of cyclodextrin derivatives either dissolved in, or chemically bonded to, a poly(siloxane) stationary phase. These phases possess higher thermal and mechanical stability, as well as higher efficiency. A wider range of cyclodextrin derivatives can be employed, since melting point and phase transition are no longer considerations for solutions (**Table 9**). The cyclodextrin derivatives are typically dissolved in moderately polar poly(siloxane) stationary phases at a concentration of 5–50% (w/w). The  $\alpha$ - and  $\beta$ -cyclodextrins with 6-*O*-*t*-butyldimethylsilyl ether (or 6-*O*-*t*-hexyldimethylsilyl ether) groups with methyl or acetyl groups at C-2 and C-3 are among the most effective cyclodextrin derivatives for the separation of different enantiomer types.

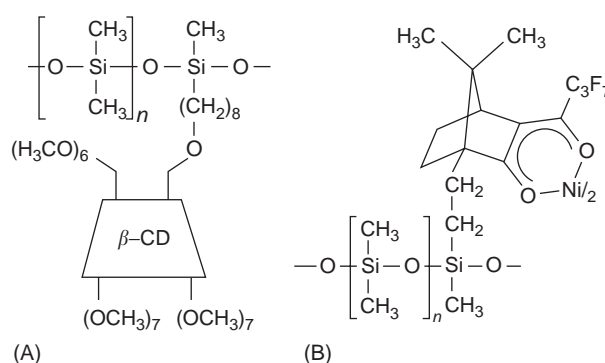
Improvements in column stability were achieved by grafting the cyclodextrin derivative onto the

**Table 9** Cyclodextrin derivatives used as chiral selectors in gas chromatography (stationary phases commercially available on open tubular columns)

Cyclodextrin derivative	MAOT <sup>a</sup>
<i>(i) Viscous liquids</i>	
Hexakis(2,3,6-tri- <i>O-n</i> -pentyl)- $\alpha$ -cyclodextrin	180
Hexakis(3- <i>O</i> -acetyl-2,6-di- <i>O-n</i> -pentyl)- $\alpha$ -cyclodextrin	200
Hexakis(3- <i>O</i> -trifluoroacetyl-2,6-di- <i>O-n</i> -pentyl)- $\alpha$ -cyclodextrin	180
Hexakis(2,6-di- <i>O-n</i> -pentyl)- $\alpha$ -cyclodextrin	200
Hexakis( <i>O</i> -( <i>S</i> -2-hydroxypropyl)-per- <i>O</i> -methyl)- $\alpha$ -cyclodextrin (mixture)	200
Heptakis(2,3,6-tri- <i>O-n</i> -pentyl)- $\beta$ -cyclodextrin	200
Heptakis(2,3,6-tri- <i>O</i> -methyl)- $\beta$ -cyclodextrin	230
Heptakis(3- <i>O</i> -acetyl-2,6-di- <i>O-n</i> -pentyl)- $\beta$ -cyclodextrin	200
Heptakis(3- <i>O</i> -trifluoroacetyl-2,6-di- <i>O-n</i> -pentyl)- $\beta$ -cyclodextrin	200
Heptakis(di- <i>O</i> -methyl)- $\beta$ -cyclodextrin (mixture)	230
Heptakis(2,6-di- <i>O-n</i> -pentyl)- $\beta$ -cyclodextrin	200
Heptakis( <i>O</i> -( <i>S</i> -2-hydroxypropyl)-per- <i>O</i> -methyl)- $\beta$ -cyclodextrin (mixture)	200
Octakis(3- <i>O</i> -trifluoroacetyl-2,6-di- <i>O-n</i> -pentyl)- $\gamma$ -cyclodextrin	180
Octakis(3- <i>O</i> -propyl-2,6-di- <i>O-n</i> -pentyl)- $\gamma$ -cyclodextrin	200
Octakis(3- <i>O</i> -butyl-2,6-di- <i>O-n</i> -pentyl)- $\gamma$ -cyclodextrin	200
Octakis(2,6-di- <i>O-n</i> -pentyl)- $\gamma$ -cyclodextrin	200
Octakis( <i>O</i> -( <i>S</i> -2-hydroxypropyl)-per- <i>O</i> -methyl)- $\gamma$ -cyclodextrin (mixture)	200
<i>(ii) Cyclodextrin derivatives dissolved in a poly(siloxane)<sup>b</sup></i>	
Hexakis(2,3,6-tri- <i>O</i> -methyl)- $\alpha$ -cyclodextrin	250
Hexakis(2,3- <i>O</i> -dimethyl-6- <i>O-t</i> -butyldimethylsilyl)- $\alpha$ -cyclodextrin	260
Hexakis(2,3- <i>O</i> -diacetyl-6- <i>O-t</i> -butyldimethylsilyl)- $\alpha$ -cyclodextrin	230
$\beta$ -Cyclodextrin	230
Heptakis(2,3,6-tri- <i>O</i> -methyl)- $\beta$ -cyclodextrin	250
Heptakis(3- <i>O-n</i> -pentyl- <i>O</i> -2,6-dimethyl)- $\beta$ -cyclodextrin	250
Heptakis(2,3- <i>O</i> -dimethyl-6- <i>O-t</i> -butyldimethylsilyl)- $\beta$ -cyclodextrin	260
Heptakis(2,3- <i>O</i> -diethyl-6- <i>O-t</i> -butyldimethylsilyl)- $\beta$ -cyclodextrin	230
Heptakis(2,3- <i>O</i> -dipropyl-6- <i>O-t</i> -butyldimethylsilyl)- $\beta$ -cyclodextrin	230
Heptakis(2,3- <i>O</i> -diacetyl-6- <i>O-t</i> -butyldimethylsilyl)- $\beta$ -cyclodextrin	230
Octakis(2,3,6-tri- <i>O</i> -methyl)- $\gamma$ -cyclodextrin	250
Octakis(2,3- <i>O</i> -dimethyl-6- <i>O-t</i> -butyldimethylsilyl)- $\gamma$ -cyclodextrin	260
Octakis(2,3- <i>O</i> -diacetyl-6- <i>O-t</i> -butyldimethylsilyl)- $\gamma$ -cyclodextrin	230

<sup>a</sup> Highest temperature (°C) for isothermal operation.<sup>b</sup> Most common poly(siloxane) solvents are poly(cyanopropylphenyldimethylsiloxane) with 14% cyanopropylphenylsiloxane monomer or poly(dimethyldiphenylsiloxane) with 20–35% diphenylsiloxane monomer. Mixing ratio for the cyclodextrin derivative is 5–50% (w/w).

backbone of a poly(dimethylsiloxane) polymer suitable for thermal immobilization (Figure 8). Several columns of this type are commercially available (e.g., heptakis(2,3,6-tri-*O*-methyl)- $\beta$ -cyclodextrin

**Figure 8** Structures of immobilized poly(siloxane) chiral stationary phases containing (A) a cyclodextrin derivative (Chirasil-Dex) and (B) a metal complex (Chirasil-Nickel).

(Chirasil-Dex), heptakis(3-*O*-trifluoroacetyl-2,6-di-*O-n*-pentyl)- $\beta$ -cyclodextrin, and octakis(3-*O*-butyl-2,6-di-*O-n*-pentyl)- $\gamma$ -cyclodextrin (Chirasil- $\gamma$ -Dex)), although columns containing dissolved selectors remain the most widely used.

Cyclodextrin derivatives have been spectacularly successful in separating a wide range of low molecular mass enantiomers of different types that are difficult or impossible to separate by other means.

### Metal Complexes

Metal(II) bis[3-(trifluoroacetyl)-(1*R*)-camphorate] and bis[3-(heptafluorobutanoyl)-(1*R*)-camphorate] of nickel, cobalt, and manganese dissolved in a noncoordinating solvent, such as poly(dimethylsiloxane), or grafted onto the backbone of a poly(dimethylsiloxane) (Figure 8) are highly selective stationary phases for the separation of a variety of stereoisomers that includes hydrocarbons and oxygen-, nitrogen-, and sulfur-containing electron-donor analytes by gas chromatography. Selectivity results from the difference in stability constants for the fast and reversible chemical equilibrium between the metal coordination compound and analytes. Thermal instability of the coordination complexes, and small stability constants at higher temperatures, restrict separations to volatile compounds that can be separated at temperatures below 125°C.

### Solid Stationary Phases

Gas-solid chromatography is used for a narrower range of separations than gas-liquid chromatography. Because of higher retention, typical applications are the separation of fixed gases, volatile hydrocarbons, halocarbons, organic solvents, and sulfur gases. The presence of immobilized active centers enhances the separation of isomers and isotopes.



**Table 10** Characteristic properties of porous polymer beads for packed column gas chromatography

Porous polymer	Monomers	Surface area ( $\text{m}^2 \text{g}^{-1}$ )	Average pore diameter (mm)	Temperature limit ( $^{\circ}\text{C}$ )
Chromosorb 101	STY-DVB	30–40	300–400	275
Chromosorb 102	STY-DVB	300–400	8.5–9.5	250
Chromosorb 103	STY	15–25	300–400	275
Chromosorb 104	ACN-DVB	100–200	60–80	250
Chromosorb 105	Polyaromatic	600–700	40–60	250
Chromosorb 106	STY	700–800	5	225
Chromosorb 107	Acrylic ester	400–500	8–9	225
Chromosorb 108	Acrylic ester	100–200	25	225
Porapak N	STY-DVB-VPO	225–500	9	200
Porapak P	STY-DVB-EVB	100–200	7.5–10	250
Porapak Q	EVB-DVB	500–850	7.5–10	250
Porapak R	STY-DVB-VPO	450–600	7.5–10	250
Porapak S	STY-DVB-VP	300–550	7–9	250
Porapak T	EGDMA	250–450	9	200
Porapak PS	Silanized P			
Porapak QS	Silanized Q			250
Tenax-GC	DPO	19–30	25–7500	375

Values for surface area vary widely in the literature.

STY, styrene; DVB, divinylbenzene; ACN, acrylonitrile; EVB, ethylvinylbenzene; EGDMA, ethylene glycol dimethacrylate; VPO, vinylpyrrolidone; VP, vinylpyridine; DPO, 2,6-diphenyl-*p*-phenylene oxide.

These separations are often difficult or impossible with liquid phases. A general guide to the selection of sorbents for particular applications is given in Table 10. PLOT columns provide higher efficiency, faster separations, and faster column regeneration compared with packed columns. PLOT columns, however, are generally less efficient than WCOT columns and have a limited sample capacity. The carrier gas competes with the analyte for interactions at the stationary phase and thus the choice and average column pressure of the carrier gas can affect selectivity.

Modern PLOT columns are prepared either by *in situ* polymerization or by addition of a chemical binder to the coating solution resulting in immobilization and bonding of particles to the inside column wall. PLOT columns containing immobilized layers of inorganic oxides, carbon, molecular sieves, cyclodextrins, and porous polymers are available in lengths up to 100 m with internal diameters from 0.25 to 0.53 mm and layers of 5–50  $\mu\text{m}$  thickness.

### Inorganic Oxides

The most important inorganic oxide adsorbents are silica gel and alumina. These adsorbents are available in a wide range of particle sizes and morphology. For packed columns, rigid spherical beads with surface areas in the range 5–500  $\text{m}^2 \text{g}^{-1}$  and average pore diameters from 8–150 nm are generally used. Similar porous materials of a smaller average particle size less than 10  $\mu\text{m}$  are used for the preparation of PLOT columns. Retention is a function of the specific

surface area, the degree of surface contamination (by water in particular), the prior thermal conditioning of the adsorbent, and the ability of the solute to participate in specific interactions, such as hydrogen bonding, with surface functional groups. The different surface functional groups (silanol groups in the case of silica and aluminum ions in the case of alumina) results in different selectivity for these adsorbents. Coating the adsorbent with a small quantity of an involatile liquid (usually a common stationary phase) or an inorganic salt is used to modify selectivity and improve efficiency. Alkali metal salts (potassium chloride and carbonate, sodium sulfate, etc) at loadings of 0.5–30% w/w are common salt modifiers.

### Carbon

Active carbons are rarely used in gas chromatography because of the difficulty in standardizing their activity and morphological properties. Graphitized carbon blacks and carbon molecular sieves have largely superseded them. Graphitized carbon blacks are prepared by heating ordinary carbon blacks to  $\sim 3000^{\circ}\text{C}$  in an inert atmosphere. This high-temperature treatment destroys functional groups originally present on the carbon surface and induces the growth of graphite crystallites. Graphitized carbon blacks used in gas chromatography are generally nonporous materials with surface areas from  $\sim 5$  to 100  $\text{m}^2 \text{g}^{-1}$ . Ideally, they should behave as nonspecific adsorbents with retention dominated by dispersion interactions. In reality, residual polar adsorption sites, albeit few



in number, establish strong specific interactions with polar solutes and affect selectivity and sample capacity. Modification of the surface by the addition of small quantities of polar liquids is used to adjust selectivity and to minimize peak tailing. PLOT columns prepared from carbon molecular sieves are more widely used today.

### Molecular Sieves

Carbon molecular sieves are prepared by the controlled pyrolysis of poly(vinylidene chloride) or sulfonated polymers (Carboxen<sup>TM</sup>). They consist of very small graphite crystallites cross-linked to yield a disordered cavity-aperture structure. Carbon molecular sieves are microporous and of high surface area, 200–1200 m<sup>2</sup> g<sup>-1</sup>. They are used primarily for the separation of inorganic gases, C<sub>1</sub>–C<sub>3</sub> hydrocarbons, and for the separation of small polar molecules such as water, formaldehyde, and hydrogen sulfide. Less volatile compounds cannot be desorbed efficiently at acceptable temperatures.

Molecular sieves (zeolites) are artificially prepared alkali metal aluminosilicates. Those used for gas chromatography are type 5A, a calcium aluminosilicate with an effective pore diameter of 0.5 nm, and type 13X, a sodium aluminosilicate with an effective pore diameter of 1 nm. The molecular sieves are microporous with a tunnel-like pore structure of similar dimensions to those of small molecules. Retention is primarily governed by the size of the analyte, which determines whether it can enter the pore structure of the molecular sieve, and the strength of adsorption interactions that take place on the internal pore surface. Molecular sieves are used primarily for the separation of permanent gases (hydrogen, oxygen, nitrogen, carbon monoxide, and the inert gases) and low molecular weight linear- and branched-chain hydrocarbons. Water, carbon dioxide, and other polar molecules are retained excessively and catalytic transformations of labile analytes are also common.

### Porous Polymers

Uniform porous polymer beads, sufficiently rigid to be dry packed into a column, are simply prepared from different monomers by suspension polymerization (Table 10). For the preparation of PLOT columns the polymerization is performed in the column followed by a second step in which the particles are immobilized onto the inside column wall. The properties of the polymers vary with the choice of

monomers, the pore structure, and the surface area. Porous polymers with average pore diameters less than 10 nm are used primarily for the separation of gases while those with larger pore diameters are used for the separation of low molecular weight organic compounds. Compared with other common sorbents their surfaces are relatively inert, and polar compounds such as water, formaldehyde, carboxylic acids, and alkylamines can be separated without appreciable difficulty.

Tenax-GC is a linear polymer of *p*-2,6-diphenylphenylene oxide with a molecular weight of about one million. It is a granulated powder of low surface area with remarkable thermal stability. This allows applications at isothermal temperatures up to 375°C.

*See also: Gas Chromatography: Overview; Gas–Solid Chromatography; High-Temperature Techniques; Chiral Separations.*

### Further Reading

- Abraham MH, Poole CF, and Poole SK (1999) Classification of stationary phases and other materials by gas chromatography. *Journal of Chromatography A* 842: 79–114.
- Jennings W, Mittlefehldt E, and Stremple P (1997) *Analytical Gas Chromatography*. San Diego: Academic Press.
- Ji Z, Majors RE, and Guthrie EJ (1999) Porous layer open-tubular capillary columns: preparations, applications and future directions. *Journal of Chromatography A* 842: 115–142.
- Kiridena W, Poole CF, Nawas MI, and Koziol WW (2003) Non-specific retention characteristics of dissolved  $\beta$ -cyclodextrin derivatives in open tubular column gas chromatography. *Journal of Separation Science* 26: 1111–1118.
- Poole CF (2003) *The Essence of Chromatography*. Amsterdam: Elsevier.
- Poole CF and Poole SK (2002) Column selectivity from the perspective of the solvation parameter model. *Journal of Chromatography A* 965: 263–399.
- Rotzsche H (1991) *Stationary Phases in Gas Chromatography*. Amsterdam: Elsevier.
- Schurig V (2002) Practice and theory of enantioselective complexation gas chromatography. *Journal of Chromatography A* 965: 315–356.
- Schurig V (2002) Separation of enantiomers by gas chromatography. *Journal of Chromatography A* 906: 275–299.
- Unger KK (ed.) (1990) *Packings and Stationary Phases in Chromatographic Techniques*. New York: Dekker.

## Gas-Solid Chromatography

**V G Berezkin**, Russian Academy of Sciences, Moscow, Russia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Gas chromatography is one of the major analytical methods; it is used in various branches of industry (petrochemical, chemical, gas industry, etc.), in pharmaceuticals, in the control of environmental pollution, in medicine, in agriculture, and in scientific research.

Chromatographic techniques are often classified according to the physical state of the two (moving and stationary) phases used. Accordingly, the following terms are in use for gas chromatography: gas-solid chromatography (GSC) and gas-liquid chromatography (GLC). Often, a classification is based on the mechanism of retention of compounds to be separated. In 'adsorption' chromatography separation is based mainly on differences between the adsorption affinities of the sample components for the surface of an active solid. Strictly speaking, gas adsorption chromatography (GAC) can also be performed in the framework of GLC, for example, if the chromatographed substances are practically insoluble in the stationary liquid phase (for example, water in poly(dimethylsiloxane) and hydrocarbons in polar liquid phases (for example, diglycerol,  $\beta\beta'$ -thiodipropionitrile)). In these examples, the retention is primarily due to adsorption of the sorbate on the gas-stationary liquid phase interface. Very often the mechanism of retention in gas chromatography is mixed (i.e., adsorption-absorption).

Schuftan was, probably, the first to suggest GAC. Important contributions to this variant of chromatography are due to Hesse.

The contributions of Cramer, Janak, Kiselev, and Yashin are notable in the initial development of GAC. Capillary gas-adsorption columns were widely used during the last 10–15 years.

### Special Features of GAC

The widespread use of GSC is explained by the following main advantages of this variant of chromatography:

1. Enhanced selectivity and sorption capacity of adsorbents, which make it possible to separate

organic or inorganic compounds having similar physicochemical properties. Examples include the separation of inorganic and organic compounds on molecular sieves, graphitized carbon black, or organic polymeric adsorbents.

2. The higher efficiency of separation compared with GLC, which is due to the fact that mass exchange processes occur more rapidly.
3. The high thermal stability of adsorbents.
4. The possibility of separating of gases and volatile compounds at 40–50°C using conventional gas chromatographs without special cooling.
5. The lower detector 'noise' due to the fact that the adsorbents are usually nonvolatile.

The unique features of GAC have also attracted the attention of Giddings: 'One immediate advantage of gas-solid chromatography resides in the fact that a surface coated with any reasonable degree of uniformity will exhibit a coefficient of resistance to mass transfer substantially smaller than the corresponding coefficient for a liquid adsorption layer in gas-liquid chromatography. The second immediate advantage of gas-solid chromatography resides in the great potential selectivity of the adsorption process. Surface adsorption is potentially capable of offering the most versatile and selective characteristics of any of the known retentive mechanisms. The rigidly fixed forces of a solid surface contrast sharply with the fluid forces of a liquid phase.'

Of course, GSC is not free of drawbacks. These include the following:

1. Poor reproducibility of sorption and chromatographic characteristics of adsorbents used (compared with those of liquid phases).
2. Asymmetry of chromatographic zones resulting from nonlinearity of adsorption isotherms.
3. A higher probability (compared with that in GLC) of irreversible adsorption and catalytic transformations of the compounds being analyzed on the surface of an adsorbent.
4. The restricted range of available adsorbents for GSC.

Of the above drawbacks the most important is the asymmetry of analytical zones, which usually results from the presence on the adsorbent surface of a number of active adsorption centers. The most efficient method of 'straightening' the adsorption isotherm lies in the modification of solid adsorbents with small amounts of polar nonvolatile liquids,

which predominantly 'deactivate' most active adsorption centers.

Comparing the advantages and drawbacks of GAC shows that the advantages of this variant are much more significant than its drawbacks. This is also supported by the fact that columns with such adsorbents as alumina, silica, graphitized carbon black, molecular sieves, and organic polymers are widely used.

The advantages of GAC can be fully utilized by using its capillary variant.

## Retention of Compounds in GSC

Retention in GSC is determined both by the nature of the surface of the adsorbent and by its structure. It has been shown that the net retention volume,  $V_N$ , is directly proportional to the surface area ( $S_{ad}$ ) of the adsorbent (it has been assumed that at relatively low coverages, the accessibility of the adsorbent surface does not change):

$$V_N = K_s S_{ad}$$

where  $K_s$  is the coefficient of distribution of the substance subjected to chromatography between the adsorbent surface and the gas phase.

Note an important specific feature of GSC, namely that the retention of compounds depends on the nature of the mobile phase. The replacement of light carrier gases by heavy gases leads to dynamic modification of the adsorbent surface, to a decrease in the coefficients of distribution of the compounds being analyzed in the 'solid adsorbent-gas' system, and, hence, to a decrease in retention times.

As in other variants of chromatography, the retention values used most in GSC are not absolute but relative (retention factor,  $k$ , relative retention,  $r$ , and Kovats retention index,  $I$ ), characterized by a higher reproducibility than that of absolute retention values.

As has been shown in the last ten years, relative retention values depend both on the nature of the carrier gas adsorbent (stationary phase) and chromatographed sorbate used and on its average pressure,  $P_{av}$ , in the column. Here nature is complex of properties, which characterize chemical substance or their mixture.

With relatively small pressures ( $P_{av}$  lower than 5–10 atm), relative retention values are linear functions of the average pressure,  $P_{av}$ , of the carrier gas in the column and its identity. For example, the following equation is valid for the retention index,  $I_i(P)$ :

$$I_i(P_{av}) = I_{0i}(0) + b_{li}P_{av}$$

where  $I_{0i}$  is the limit value of the retention index of the  $i$ th chromatographed compound when  $P_{av}$  tends

to zero, and  $b_{li}$  is the adsorption coefficient, depending on the adsorbent-carrier gas system.

## Zone Broadening in Capillary Gas Adsorption Columns

Zone broadening during migration along the column is influenced by the flow profile of the carrier gas in the column, by the diffusion of the substances being separated in mobile and stationary phases, by the rate of interfacial mass transfer, and by the characteristics of the adsorption layer used.

An equation relating the height equivalent to a theoretical plate (HETP) to the linear velocity of the carrier gas for GAC was proposed by Giddings. The Giddings equation can be represented as

$$H = \frac{2D_g}{u_o} + \left[ \frac{1 + 6k + 11k^2}{24(1 + k)^2} \right] \frac{r^2 u}{D_g} f_1 + \frac{8}{a_k u_m} \left( \frac{k}{k + 1} \right)^2 \frac{V_g}{S} f_n u_o f_2$$

where  $H$  is the HETP;  $D_g$  is the diffusion coefficient of the compound under analysis in the gas phase;  $u$  is the linear velocity of the carrier gas;  $k$  is the retention factor; and  $V_g$  is the specific retention volume.  $u_o$  is the linear velocity of the carrier gas at the outlet of the column:

$$u_o = \frac{L}{t_m f_2}$$

where  $L$  is the length of the column; and  $t_m$  is the 'dead time' of the column:

$$f_1 = \frac{9(P^4 - 1)(P^2 - 1)}{8(P^3 - 1)}$$

$$f_2 = \frac{3(P^2 - 1)}{2(P^3 - 1)}$$

where  $P = P_i/P_o$ ,  $P_i$  and  $P_o$  are the carrier-gas pressures at the inlet and outlet of the column, respectively;  $a_k$  is the accommodation coefficient;  $u_m$  is the average velocity of the sample in the gas phase;  $S$  is the overall surface area of the adsorbent layer in the column; and  $f_n$  is a factor reflecting the heterogeneity of the adsorbent.

## Preparation of Capillary Adsorption Columns

Only methods of preparation of capillary columns with a porous adsorbent layer are described in the

literature. This type of column is characterized by high capacity and stable performance.

Two methods are used for the preparation of capillary gas adsorption columns: the suspension method, in which the inner walls of the column are coated with a suspension of the adsorbent, and the chemical method, in which an adsorption layer is formed on the walls of the column through a process of synthesis of the adsorbent in the capillary column. Recently, a new type of porous polymer (poly(1-(trimethylsilyl)-1-propin) (PTMSP)) has been suggested as an organic adsorbent and has been actively studied as a promising material in membrane technology. This polymer dissolves well in some volatile solvents, and a layer of it can be formed in a capillary column using simple techniques for coating from a solution of a stationary phase in a volatile solvent.

## The Use of GSC

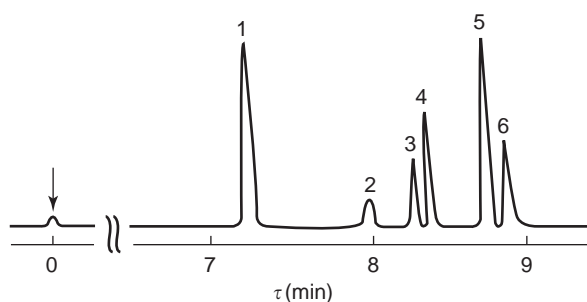
At present, GSC is widely used in analytical practice, and the capillary variant of GSC is used most effectively and most often. Considering the advantages of capillary columns over packed columns, only the higher efficiency of the former is often mentioned. At the same time, capillary columns are also characterized by faster separations, by more reproducible temperature regimes, by the small sizes of the apparatus needed, and by lower consumption of carrier gases and sorbents; they also extend applications to the lower temperatures.

At present, GAC is used most frequently for analysis of volatile compounds with various isotope compositions, of inorganic and organic gases, of volatile organic polar compounds, and of aqueous solutions of organic compounds.

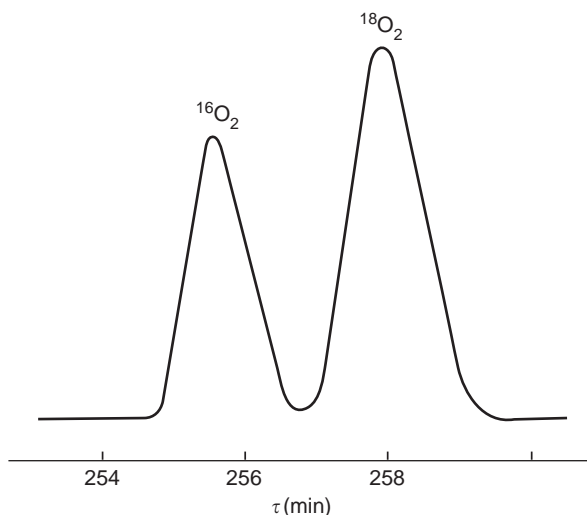
The separation of nuclear spin isomers and isotopes of hydrogen was the first prominent example of a practical application of capillary adsorption chromatography (Figure 1). Later, it was shown that the time required for the separation of protium and deuterium could be shortened by decreasing the column temperature to 47 K.

Oxygen isotopes ( $^{16}\text{O}_2$  and  $^{18}\text{O}_2$ ) were separated by capillary adsorption chromatography (Figure 2). The adsorption layer on the inner wall of a glass column was produced by etching with 20% NaOH for several hours at 100°C. The efficiency of the capillary columns was  $\sim 350\,000$  theoretical plates. A mixture of nitrogen (65%) and helium (35%) was used as the carrier gas.

The use of porous polymers as adsorbents offers a number of advantages, among them the high



**Figure 1** Chromatogram of the separation of nuclear spin isomers and isotopes of hydrogen in an open tubular adsorption glass capillary column. The 80 m long open tubular capillary column contained a 20 mm thick silica layer as an adsorbent;  $T = 77.4\text{ K}$ ; neon was used as the carrier gas. 1, helium; 2, *p*-protium; 3, *o*-protium; 4, DH; 5, *o*-deuterium; 6, *p*-deuterium.

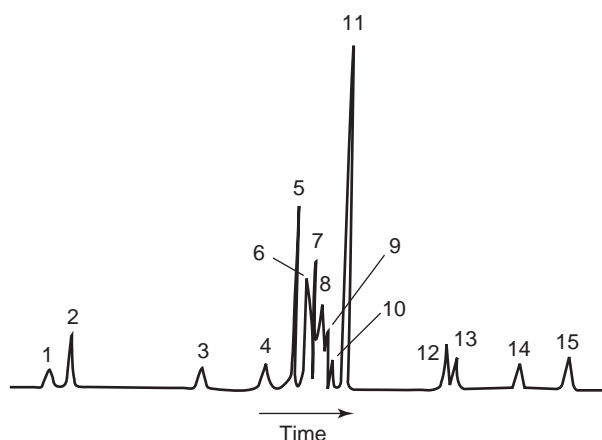


**Figure 2** Chromatogram of the separation of a mixture of the isotopes  $^{16}\text{O}_2$  and  $^{18}\text{O}_2$  in an adsorption glass capillary column (175 m  $\times$  0.3 mm) at 77 K, with a  $\text{N}_2(65\%)$ -He(35%) binary carrier gas.

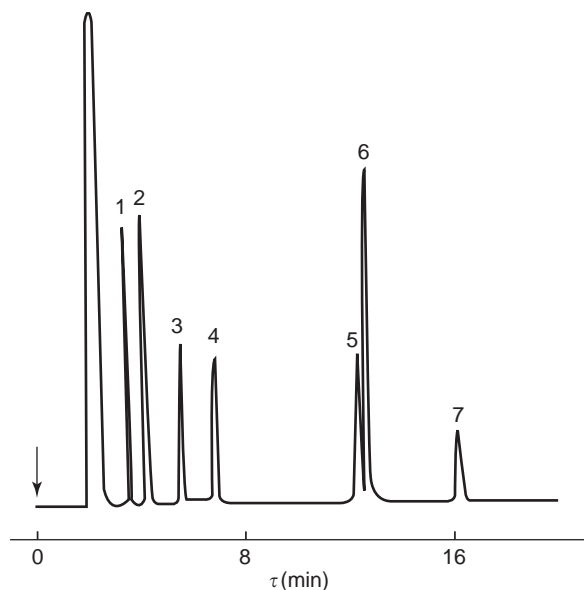
homogeneity of the surface with controllable (within certain limits) porosity and hydrophobicity.

Note that natural organic adsorbents can also be used in capillary columns. Figure 3 shows the results of the separation of organic acids of natural origin on a natural polymer, *Staphylococcus aureus* cells. A fairly satisfactory degree of separation was achieved. The use of natural adsorbents seems promising.

Columns with traditional adsorbents, for example, graphitized carbon black, also find practical application. 'Structural sensitivity', which is manifested in the separation of isomers, is a specific feature of graphitized carbon black. This feature is illustrated by the separation of hydrocarbons (Figure 4). All three isomers of xylene are separated.

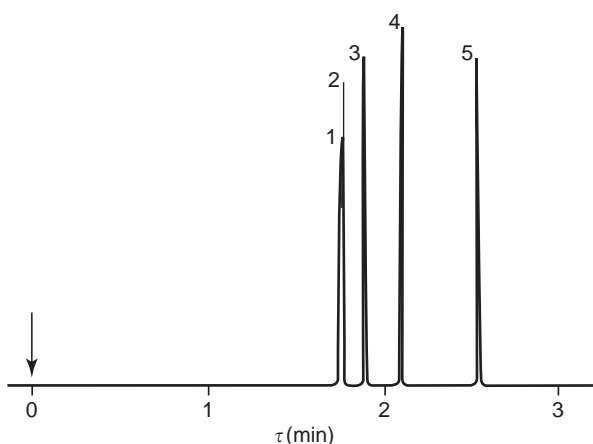


**Figure 3** Chromatogram of the separation of unsaturated acids in an open tubular capillary column in which *Staphylococcus aureus* cells were used as the adsorbent. Acids: 1, 14:1 tetradecenoic; 2, 14:0 myristic; 3, 15:0 pentadecanoic; 4,  $c\Delta^8$  16:1 *cis*-hexadec-8-enoic; 5,  $c\Delta^9$  16:1 *cis*-hexadec-9-enoic (palmitic); 6,  $c\Delta^{10}$  16:1 *cis*-hexadec-10-enoic; 7,  $c\Delta^{11}$  16:1 =  $t\Delta^9$  16:1 *cis*-hexadec-11-enoic and *trans*-hexadec-9-enoic; 8,  $t\Delta^{10}$  16:1 *trans*-hexadec-10-enoic; 9,  $t\Delta^{11}$  16:1 *trans*-hexadec-11-enoic; 10,  $t\Delta^{12}$  16:1 *trans*-hexadec-12-enoic; 11, 16:0 palmitic; 12,  $c\Delta$  17:0 *cis*-methylenehexadecanoic; 13,  $t\Delta$  17:0 *trans*-methylenehexadecanoic; 14, 18:0 octadecenoic; 15, 18:0 stearic.



**Figure 4** Chromatogram for the separation of a mixture of organic hydrocarbons in a capillary column with graphitized carbon black. 1, *n*-octane; 2, benzene; 3, *n*-nonane; 4, toluene; 5, *p*-xylene; 6, *m*-xylene; 7, *o*-xylene. A 30 m  $\times$  0.32 mm column with Carbowack B was used;  $T = 60^\circ\text{C}$ ; helium was used as the carrier gas.

The use of film-forming organic polymers, which are commonly used in membrane technology, as chromatographic adsorbents for packed and capillary columns has been suggested. It was shown that



**Figure 5** Chromatogram of a mixture of inorganic hydrides: 1,  $\text{SiH}_4$ ; 2,  $\text{H}_2\text{S}$ ; 3,  $\text{PH}_3$ ; 4,  $\text{GeH}_4$ ; 5,  $\text{AsH}_3$ . Experimental conditions: flame-photometric detector,  $\lambda = 394\text{ nm}$  (peak 2),  $\lambda = 525\text{ nm}$  (peaks 1, 3–5); quartz capillary column (25 m  $\times$  0.22 mm), adsorbent PTMSP (layer thickness 0.4  $\mu\text{m}$ ), column temperature  $100^\circ\text{C}$ , carrier gas helium (35  $\text{cm s}^{-1}$ ).

$\text{C}_1$ – $\text{C}_4$  hydrocarbon gases were separated on columns with a layer of PTMSP. This polymer is also promising for the separation of inorganic gases (Figure 5).

See also: **Gas Chromatography**: Overview; Principles.

## Further Reading

- Arnold G, Berezkin VG, and Ettre LS (1985) Paul Schuftan and the early development of gas-adsorption chromatography. *Journal of High Resolution Chromatography* 8: 651–658.
- Berezkin VG (1991) *Gas-Liquid-Solid Chromatography*. New York: Dekker.
- Berezkin VG (2004) *What Is Chromatography? A New Approach Defining Chromatography*. The Foundation: IOPMS, Belgium, Kortrijk.
- Berezkin VG and de Zeeuw J (1996) *Capillary Gas Adsorption Chromatography*. Heidelberg: Hüethig Verlag.
- Berezkin VG, Korolev AA, Malyukova IV, et al. (2002) Poly [1-(trimethylsilyl)-1-propyne] as chromatographic adsorbent and prospects of its application in packed and capillary columns. *Journal of Chromatography A* 960: 151–158.
- Ettre LS (1993) Nomenclature for chromatography (IUPAC recommendations 1993). *Pure and Applied Chemistry* 65(4): 819–872.
- Giddings JC (1964) The theory of gas-solid chromatography. *Analytical Chemistry* 36: 1170–1175.
- Kiselev AV and Yashin YaI (1962) *Gas Adsorption Chromatography*. New York: Plenum Press.
- Kiselev AV and Yashin YaI (1985) *Gas-und Flussigadsorption Chromatographie*. Heidelberg: Hüethig Verlag.
- Phillips CSG (1956) *Gas Chromatography*. London: Butterworth.



## Multidimensional Techniques

P J Marriott, RMIT University, Melbourne, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The gas chromatographic (GC) technique dates from the 1950s. One of the earliest applications was the analysis of fatty acid methyl esters. It is recognized that in the decades since, almost every conceivable sample type containing volatile compounds has been subjected to GC analysis. Not only is the general application of GC to a vast array of samples of historical interest, but the modification of the method to provide better or optimum analysis of a given sample type has been a feature in the development of the GC method since its first demonstration. In this respect, it is the effective resolution (separation) of components of a sample that is the desired outcome. Quite simply, this will invariably mean that qualitative and quantitative analysis of samples is more reliable. Thus, in terms of better analysis we will almost always demand improved separation, because this is the essence of any chromatographic technique – to provide separation. This technological push is precisely why capillary GC, with its significantly greater peak capacity, has effectively replaced packed column GC for most routine analysis. Whilst there are other advantages pertaining to capillary GC, such as improved inertness, the consideration still is that even these factors are related to separation power and/or improved quantification.

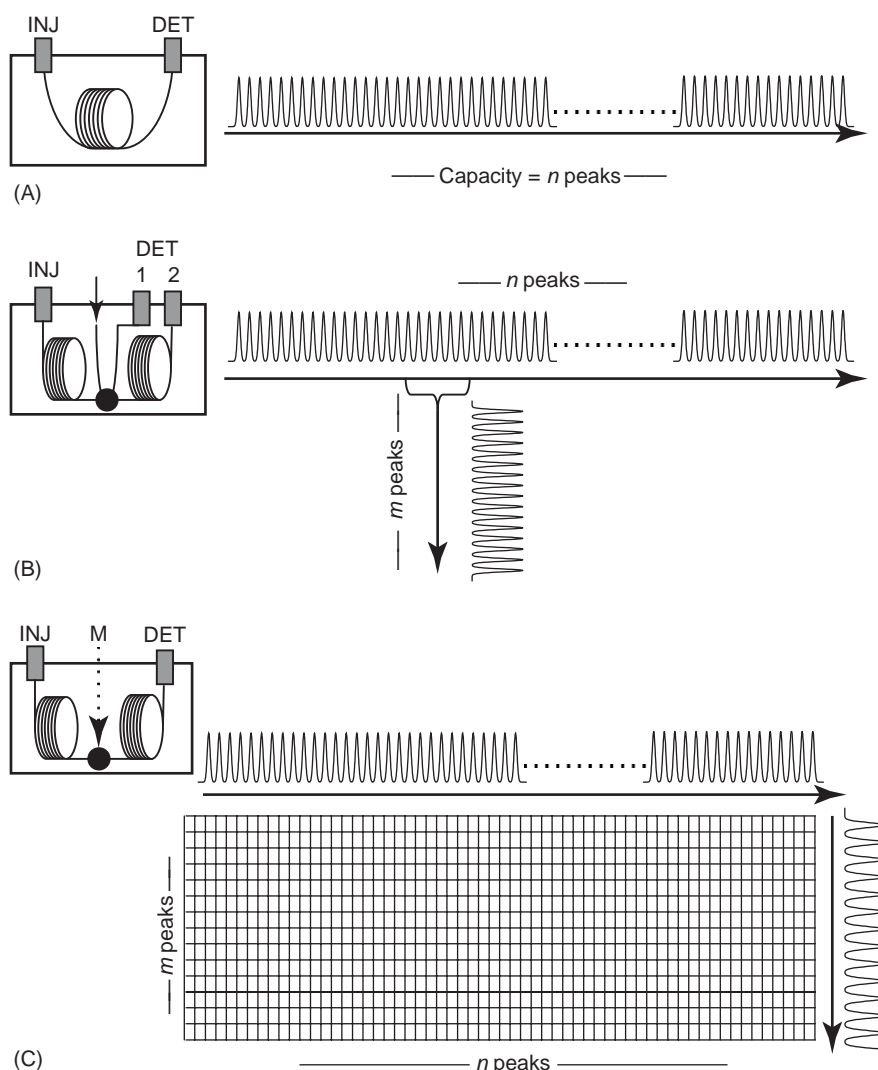
Against this backdrop, it is recognized that even the highest separation power single capillary column method cannot achieve the analysis performance demanded to effectively separate to baseline all components of many complex samples. Statistically, once the number of compounds exceeds a certain level (e.g., ~20%) of the separation capacity of a column for randomly distributed components, then peak overlaps and reduced resolution of individual components arises. The separation capacity may be defined as the number of baseline-resolved components that can be fitted into the analysis separation space (i.e., how many peaks can be placed side-by-side in the chromatographic space; **Figure 1A**). For some multicomponent samples, such as a petroleum oil or product (e.g., kerosene), the vast majority of compounds will be unresolved (i.e., they will not be pure peaks). Major components such as *n*-alkanes may be readily recognized, but underlying these peaks will be other minor components. Depending on the degree

of accuracy required of the analysis, the co-eluting impurity may not be too critical, but for certainty of analysis, the presence of an unaccounted impurity leads to unacceptable results.

Presented with this scenario, and the need to provide greater analytical measurement certainty, there are a number of strategies that can be taken to improve analysis of individual components, and provide greater separation power. Accepting the limited capacity of the single column, these strategies will involve either making the sample mixture less complex, or making the GC step more efficient – which can only be achieved by using a suitable arrangement comprising more than one column. Since these involve more than one sample separation step, they can be collectively called multidimensional separations. Multidimensional separation may be considered to encompass a broader scope than multidimensional gas chromatography (MDGC). The former can incorporate a preseparation of a sample into fractions, each of which can then be separately analyzed by GC. If two components that would normally co-elute are now present in different analyzed fractions, then by definition they will be no longer unresolved. This approach has been widely employed in petroleum analysis, where such methods as thin layer chromatography or liquid chromatography are used to fractionate a crude sample, and each zone or class of compounds that are isolated into separate extracts are then analyzed by GC. High-performance liquid chromatography can be used in a like manner, whether offline or online coupled to the GC.

With respect to post-GC column procedures, there is limited scope for increasing the resolving power of the total analysis. It is essentially only possible to follow a GC step (dimension) with another GC dimension. This now defines the MDGC method (**Figure 1B**). Considering this to be described as using a GC step to further analyze or differentiate the effluent from a first GC column, then this can be contrasted to using a spectroscopic detection step to analyze GC effluent. This is important, because today it is widely believed that for unresolved GC peaks the only way to effectively, uniquely, measure the individual compounds is by spectroscopic means (e.g., mass spectrometry (MS)). Note that however it is more appropriate to term the latter multidimensional analysis, as opposed to multidimensional chromatographic separation. The conceptual difference between GC-MS and MDGC is illustrated in **Figure 2**. MDGC aims to separate peaks in two





**Figure 1** (A) Peak capacity ( $n_c$ ) expressed for a single GC column. (B) Peak capacity expressed for a multidimensional GC analysis, with a first-dimension GC column ( $n_c$ ) and one heartcut event to a second-dimension GC column capacity ( $m_c$ ). Total capacity  $\sim n_c + m_c$ . (C) Peak capacity expressed for a comprehensive two-dimensional GC analysis where each dimension has capacity  $n_c$  and  $m_c$ , respectively. Total capacity  $\sim n_c \times m_c$ .

dimensions, whereas GC–MS attempts to provide unique mass spectral ions to measure quantitatively and qualitatively peaks that might overlap.

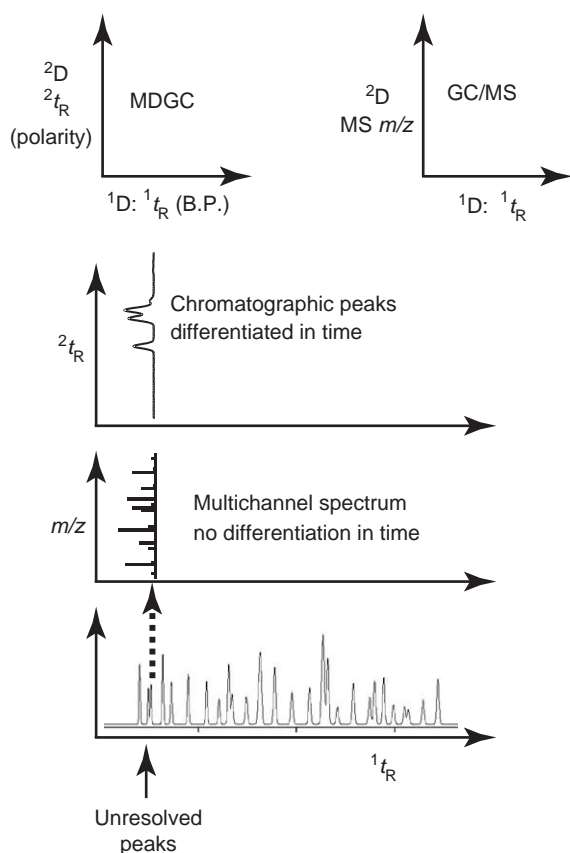
Multidimensional gas chromatography may be defined as follows: ‘the process of selecting a (limited) region or zone of eluted compounds issuing from the end of one GC column, and subsequently subjecting the zone to a further GC displacement’.

Figure 3 shows how different regions (here two regions) from a first column (Figure 3A) can be selected and passed to a second column, where better separation occurs (Figure 3B). Mass spectrometric data will be a series of contiguous mass spectra (Figure 3C) acquired for the total GC run.

This process can be repeated as many times as required or as physically permitted by the technical

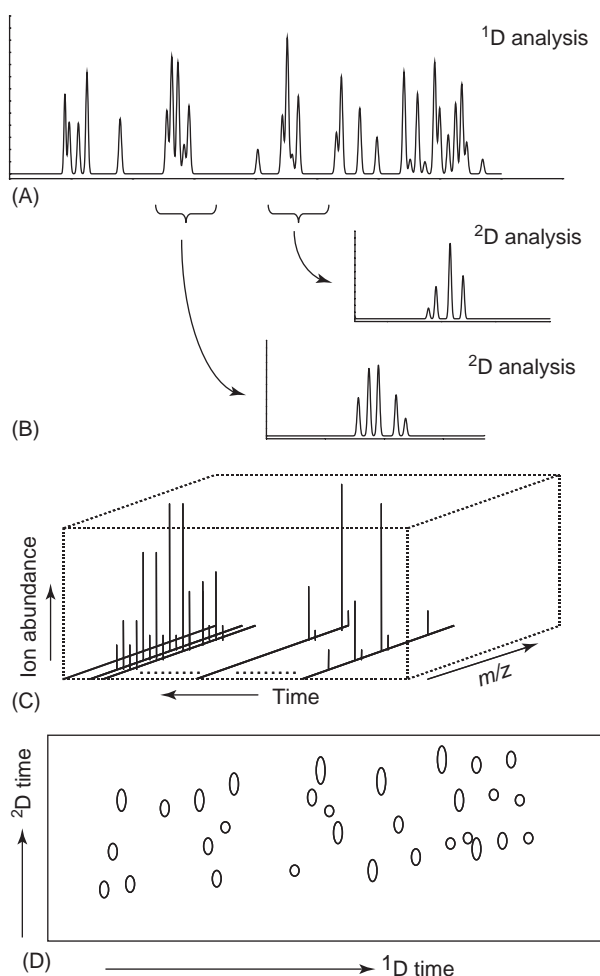
system employed. The selected zone may be chosen to enhance the separation of a particularly complex zone, which may be minutes broad, or it may be used to isolate just a few peaks of importance. The limiting example of this will be to select just ‘one’ peak from  $^1D$  and transfer it to  $^2D$ . This may seem a trivial example, but if the target peak is enantiomeric, and  $^2D$  is required to provide enantioseparation, then an enantioselective phase in  $^2D$  will achieve this: the goal is to isolate the ‘single peak’ and subsequently to resolve its enantiomers. If a single enantiomeric column is used for the total sample, then it may be that the resolved enantiomers will co-elute with other matrix peaks and so compromise the analysis.

Various generic arrangements of coupled column GC are shown in Figure 4. Some are trivial, such as



**Figure 2** Illustration of the difference between GC/MS and MDGC. Peaks elute from  $1^D$  at their characteristic  $1t_R$ , either into a second column (MDGC) where their resolution is improved through the use of a second column which provides differentiation in time (i.e.,  $2t_R$  values for each component) or into the MS source (GC–MS), where there is no differentiation in time, but characteristic  $m/z$  ions are produced for all peaks eluting from  $1^D$ . For unresolved peaks, either unique ions or spectral deconvolution is required for providing reliable analysis.

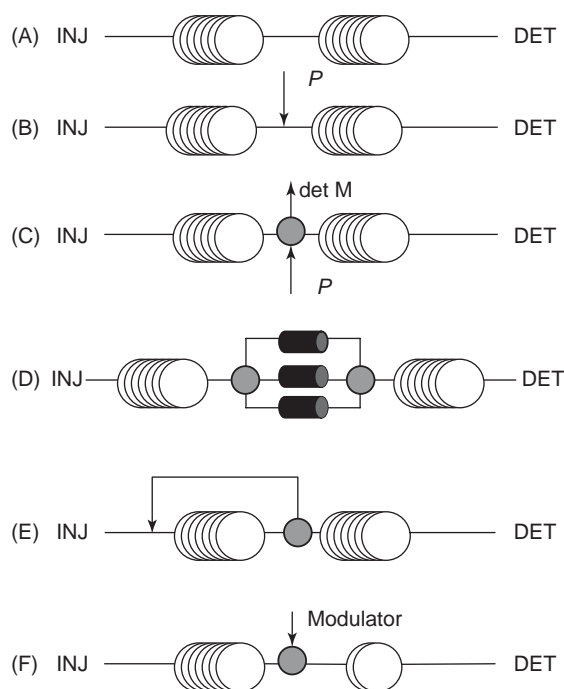
**Figure 4A**, which shows directly connected columns. Adding a midpoint pressure (**Figure 4B**) allows tuning of the overall phase selectivity, and whilst **Figure 4A** and **4B** may give better resolution, they are not strictly multidimensional. **Figure 4C** is the simplest representation of an MDGC arrangement, with a monitor detector and switching device to heartcut solute from  $1^D$  to  $2^D$ . Often a cryotrap will collect/focus selected heartcuts at the head of  $2^D$ . A multiple trap design (**Figure 4D**) permits collections of fractions of the  $1^D$  effluent independent of the need to analyze them on the second column. Using a flow switch device between the two columns, it is possible to perform a range of advanced modes, such as back-flushing from the first column, or recycling selected components through the analytical column (**Figure 4E**). Finally, the comprehensive two-dimensional GC procedure is shown in **Figure 4F**. This method is



**Figure 3** (A) Single-column GC analysis result. (B) Selected incompletely resolved heartcuts from A are directed to a second column where improved separation is achieved. Here two heartcuts are performed. (C) Illustration of multidimensional analysis GC/MS, where the second dimension is a mass spectrum recorded at the mass spectral data acquisition rate. The success of this approach depends on the uniqueness of the mass spectrum at each point in the chromatogram. (D) The comprehensive two-dimensional GC method ideally spreads out the components in a two-dimensional space, according to the characteristics of each of the columns. Each component has a characteristic  $1^D$  and  $2^D$  retention time in the plane.

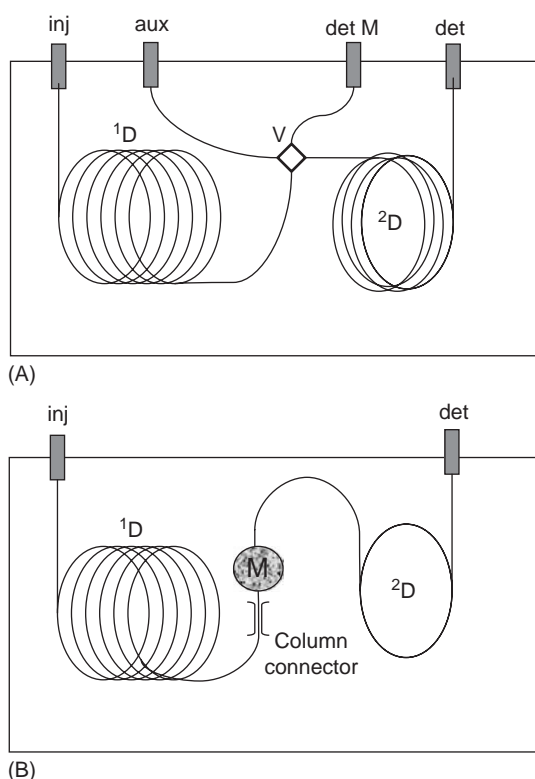
distinguished from the above by using a rapid modulation method to quickly alternate between focus and expel steps for solute eluting from  $1^D$ .

**Figure 5A** is a more detailed although still simplified diagram of the MDGC instrument. The heartcut valve is shown as V, although the different ways of performing the heartcut is not apparent. Thus, the Deans switch that relies upon pressure to shunt column 1 effluent to either the midpoint detector or to column 2 is one way to implement the heartcut. Other systems variously use valves or a combination of flow and shutting off stream flows to the monitor detector.



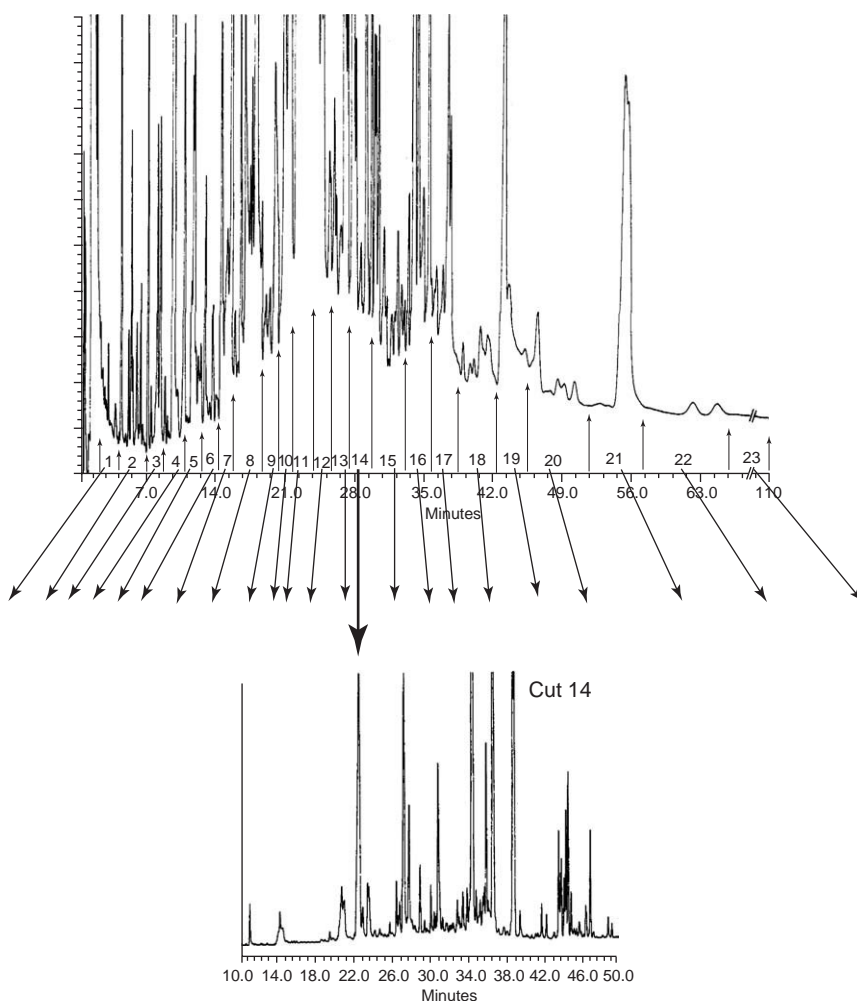
**Figure 4** Various arrangements for different coupled column techniques. (A) Multichromatography method – two columns are simply joined together. (B) Pressure-tuning. Supplementary flow ( $\Delta P$ ) at the mid-point alters the relative contribution of each column to the separation, so alters overall selectivity. (C) Normal MDGC. A switching valve for heartcut transfer, and monitor detector (det M) to show the  $^1D$  column result. (D) Multiple traps between  $^1D$  and  $^2D$  allow sequential heartcuts to be trapped and then separately eluted into the second column. (E) Recycle chromatography. The eluate from the first column can be recycled through the separation column a number of times to allow increased separation. (F) Comprehensive two-dimensional gas chromatography.

Typically, an MDGC method will focus on a limited number of heartcutting regions; however, more extensive analysis has been described where most of the sample is analyzed on both columns. The study of Gordon (24 heartcuts; **Figure 6**), each of 1–3 min duration of  $^1D$ , exemplifies this, where each heartcut had to be analyzed over a conventional timeframe, and gave a total analysis of about 2 days. In this case, the heartcuts were performed sequentially on separate sample injections, so each heartcut analysis involved one sample injection, one heartcut event, and one  $^2D$  analysis. The process was then repeated for each successive heartcut. Alternatively, it is possible to perform multiple heartcuts from  $^1D$  to  $^2D$ , with cryotrapping at the head of  $^2D$ . After these heartcuts are all co-collected for the single  $^1D$  analysis, the oven is cooled, the cryofluid turned off, and then all heartcuts are eluted as one run on the second dimension. Wilkins employed a series of heartcuts with parallel cryotrap, and at the conclusion of the first dimension separation,



**Figure 5** (A) Schematic description of a multidimensional gas chromatography arrangement, where a switching system or valve (V) is located between the two columns. The process of switching the flow between  $^1D$  and  $^2D$  or to the monitor detector (det M) is not shown. The auxiliary flow (aux) provides flow to the system to assist in the switching process, and/or to provide make-up flow into the column, which is not receiving flow from the  $^1D$  pre-column.  $^2D$  is normally a column of regular dimensions, and may incorporate a cryofocusing step at the head of the column. (B) The comprehensive two-dimensional gas chromatography arrangement essentially only requires a mechanism for modulation between the two columns, which provides a series of narrow peaks (at least four normally) to  $^2D$  for each  $^1D$  peak. The modulator (M) is shown near the column connection.  $^2D$  is normally a short, fast elution column.

each cryotrap was individually eluted into the second column. Suitable valve switching is required. Again, this involves multiple, full chromatographic development on  $^2D$ . In this case, parallel Fourier transform infrared (FTIR) and MS detection was used with the outlet flow of  $^2D$  unequally split to the detectors. Note that in all these cases,  $^2D$  is a conventional analysis in terms of column dimensions and conditions. Recently, a heartcutting/cryotrapping/fast-GC analysis was described, where the heartcut zone was cryofocused and then rapidly remobilized into a short, narrow bore capillary column for very fast GC discrimination of components. This allows the MDGC procedure to be completed within the same time as the first-dimension separation.



**Figure 6** Example of MDGC with multiple heartcuts (each heartcut required a separate injection) and separate  $^2D$  analysis of each heartcut.  $^2D$  result for heartcut event 14 is shown. (Reproduced with permission from Gordon BM, Uhrig MS, Borgerding MF, *et al.* (1988) Analysis of flue cured tobacco essential oil by hyphenated analytical techniques. *Journal of Chromatographic Science* 26: 174–180; © Preston Industries, Inc.)

## Modes of MDGC

### Packed-to-Capillary

The situation of requiring analysis of a target compound that is severely overlapped by a major component is not isolated. If we analyze such a sample by using capillary GC then there is a risk that the major component will obscure the minor component, either because of the widening base of the major peak, or by the common observation of peak overloading that causes band asymmetry and broadening. Also, to increase the detectability of the small peak, it may be desirable to increase the sampled amount. This exacerbates the above problems. If a packed column is used in  $^1D$ , then more sample (both total, and target compound) can be introduced to the first column without too much overloading concern, and then heartcutting the target compound will lead to

most of the interfering compound being rejected from the  $^2D$  column. When the  $^2D$  separation is developed, the interferent will be much less likely to overload the capillary column than it normally would, and the target compound has an enhanced response because more of it enters  $^2D$ .

### Back-flushing

MDGC systems employ a switching valve at the confluence of the two columns. If the component(s) of interest are in the more volatile fraction of the injected sample then it is only necessary to pass this early eluting segment into  $^2D$ . Once this has been effected, the flows can be switched to cause the carrier gas to reverse-flow out the primary column, whilst the carrier still flows into  $^2D$  to allow the target compounds to be analyzed. Since the heavy compounds will not

have eluted far into the  $^1D$  column, they are relatively quickly eliminated from the inlet to  $^1D$ , and out the split vent. **Figure 4E** demonstrates this action. This allows faster turnaround of analysis, since it is never necessary to completely elute even the heaviest compounds completely through the analytical column.

### Multiple Heartcutting

A usual MDGC procedure may require multiple heartcuts. If these are transferred to  $^2D$  then each of the peaks in these zones will have an input peakwidth equal to the dispersion they have experienced on  $^1D$ . Also, peaks that swap positions on  $^2D$  compared with their  $^1D$  elution order may suffer worsened resolution on the two-column set, as they try to establish their reversed elution order as they travel along  $^2D$ . It is normally preferred that a cryofocusing event is used at the head of  $^2D$  to provide essentially a sharp band injection into  $^2D$ .

### Multiple $^2D$ Columns

The use of multiple (parallel) cryotrap or second dimension columns has been reported. The most common method in this category will employ two  $^2D$  columns, since this gives an improved performance without the added complexity of having more than two second-dimension columns. It is possible to divert the  $^1D$  effluent to either of the  $^2D$  columns.

### Multiple Parallel Traps

In the above instance, the discrete columns in  $^2D$  mean that effluent for  $^1D$  can be transferred to either of the columns, and a decision must be made as to which  $^2D$  column to use.

In conventional MDGC, it has been stated that a limited number of heartcuts can be performed. If the total sample must be subjected to the two-column analysis, then we face a dilemma: how can we study more of the effluent for  $^1D$ , but not run the risk of deteriorating the high performance of  $^2D$  by taking too large a heartcut into  $^2D$ . This can be solved in two ways. The first is to repeat the injection onto  $^1D$  and choose a different heartcut region for each injection. Eventually, it should be possible to analyze the whole sample on both dimensions – at the risk of significantly increasing the total analysis time.

The second is to heartcut successive or desired fractions into a range of parallel traps (**Figure 4D**) over the course of the development of the  $^1D$  separation, and then elute the contents of each of the traps into the  $^2D$  column as discrete events. Note that the above two methods involve discrete or sequential analysis of the selected heartcuts under essentially

standard conditions. Again, both approaches impose time penalties on the total analysis.

### Multiple Oven Operation

The use of multiple ovens (normally two) allows isolation of the optimum or desired conditions of analysis on  $^1D$  from the conditions used for analysis of the heartcut fraction on  $^2D$ . Once the heartcut has been delivered to  $^2D$ , it is then possible to independently complete the analysis on  $^1D$  without compromising the temperature conditions for analysis on  $^2D$ .

### MDGC with Selective Detection

Even though MDGC gives improved separation power, there is still an important requirement to achieve either solute identification through the use of MS, or using the specificity available with selective detection. Thus, two-column operation with a range of detection options is possible. Apart from the universally used flame ionization detector, the most popular detector for MDGC will be the mass spectrometer. Structural/identity information with full scan acquisition, and the selectivity and sensitivity available with selected ion monitoring makes the mass spectrometer a most attractive option. The use of other structural identification techniques such as FTIR has been reported by select groups. The complementarity of MS and FTIR is recognized when analyzing isomers of alkyl benzenes and alkanes. MS gives poor specificity toward the former, since they have similar MS fragmentation and the same molecular ion, whereas FTIR is able to distinguish positional isomer differences, due to different vibrational modes. However, for alkanes, FTIR is information-poor in its ability to distinguish homologs of an alkane class, whereas MS provides molecular ion information (and an indication of the chemical class) and so is able to correctly identify these species.

The use of simultaneous FTIR/MS using stream splitting to the parallel detectors was reported by Wilkins for petrochemical samples in an MDGC system. Asymmetric splitting was necessary because of the different sensitivities of the detection methods – most of the flow is diverted to the FTIR channel.

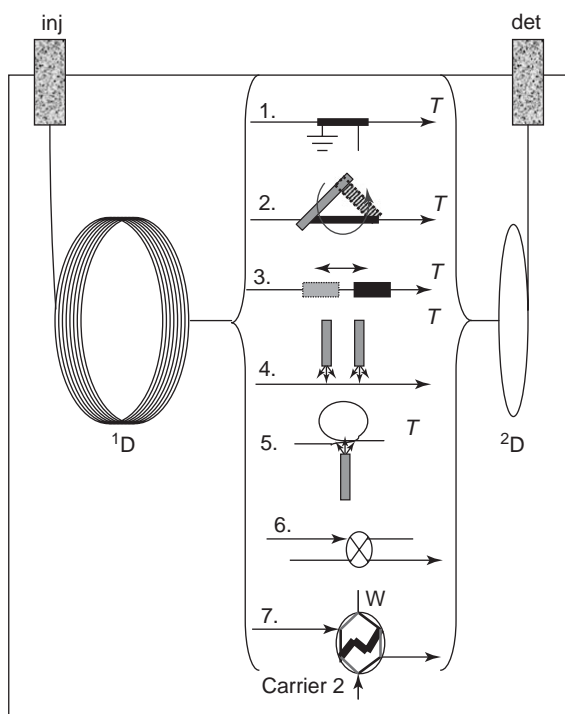
Any selective GC detector that provides the necessary molecular specificity for the analysis task may be used with MDGC methods.

### Comprehensive Two-Dimensional Gas Chromatography (GC $\times$ GC)

According to the previous definition, it is possible to increase the speed (frequency) of the heartcutting



operation to the point where each heartcut is of a similar duration as a single peakwidth, and to also operate these heartcuts contiguously. There is a valuable rationale for doing this, although taking the conventional wisdom of MDGC, it would not be logical to do MDGC in this manner. Conventionally, only selected target zones would be chosen for enhanced separation – this fast, continuous heartcutting process used in  $GC \times GC$  appears to be nondiscriminative and so is difficult to rationalize. Second, if  $^2D$  is to provide analysis of these heartcuts effectively, it is impossible to use the methods or approaches described above – discrete trapping events for each heartcut, or parallel traps, and so forth. If the procedure involves direct transfer from a single  $^1D$  to a single  $^2D$  column, then  $^2D$  will have to be capable of analyzing each heartcut within the time duration of the heartcut. This will then allow  $^2D$  to complete the analysis of each heartcut before the next heartcut is delivered to it. This essentially describes the comprehensive ( $GC \times GC$ ) technique. The multiplicative operation ( $\times$ ) is deliberately intended to indicate that the separation power of  $^1D$  is combined with the separation power of  $^2D$  over the total sample, through multiplying their individual peak capacities. Figure 1C is a diagram that puts this process into context. Giddings considered this expansion in separation space some time ago, but did not indicate how it could be achieved. The only effective way to do this is to ensure that each  $^1D$  peak is sampled a multiple of times (at least three to four times) into  $^2D$ . Clearly this technique demands a novel approach to implementing the coupling of the columns, and this is done through use of a ‘modulator’. Figure 5B is a schematic diagram of the  $GC \times GC$  instrument, with the modulator located near the junction of the two columns. Both mass-conservative and peak subsampling modulators have been described. Figure 7 includes seven different types of modulators for  $GC \times GC$ . Those that use either elevated or reduced temperature (1–5) are denoted  $\Delta T$ , and some cryogenic types also have supplemental heating to aid expelling of the trapped bands to  $^2D$ . For the mass-conservative modulator, the first and second dimension columns may be directly joined, without requiring valves between the columns. The process therefore requires that the part or segment of the peak entering the modulation region should be compressed in space along the column into a concentrated packet, and then rapidly introduced into dimension 2. The initial way this was done was by application of heating to accelerate the chromatographic band in an accumulation region (often a thicker film stationary phase column). Today, most modulators of the mass conservative type



**Figure 7** Schematic representations of various modulator types used in comprehensive two-dimensional gas chromatography. (1 and 2) Use alternating high temperature at a thick segment of column to collect then expel solute from  $^1D$  into  $^2D$ . (3–5) Rely on cold-trapping and then rapid heating of the column segment for the collect/expel process. The heating may be provided by a secondary heated gas flow, or just the stirred oven. (6 and 7) Are valve-based sampling methods which subsample the peak from  $^1D$  into  $^2D$ .

employ cryogenic cooling. This cryofocuses the selected part (or all) of the peak into a very narrow band, and when the cooling is removed from that column segment, the capillary column heats up and allows the trapped band to pass rapidly to  $^2D$  to permit its fast analysis (and separation of co-trapped compounds). This necessitates repetitive sampling or modulation across the total chromatogram, which generates a series of contiguous second-dimension chromatograms. To adequately display these  $GC \times GC$  data, a two-dimensional plot of the data is prepared, with the two axes being the total analysis time (or first-dimension elution time) and the second-dimension time (or modulation period, e.g., 4 s). This might be implicit in the nature of the separation space shown in Figure 1C and Figure 3D is a more realistic illustration of the  $GC \times GC$  result. Peaks are now spots in a two-dimensional plan, much like those spots found in thin-layer chromatography. The power of the technique resides in (1) its very high peak capacity which permits a much greater separation of components, (2) the increase in peak



- Hirschfeld T (1980) The hyphenated methods. *Analytical Chemistry* 52: 287A–312A.
- Marriott P and Shellie R (2002) Principles and applications of comprehensive two-dimensional gas chromatography. *Trends in Analytical Chemistry* 21(9–10): 573–583.
- Marriott P, Ong R, Shellie R, *et al.* (2003) Ultra-high resolution capillary gas chromatography by using cryogenic modulation. *Australian Journal of Chemistry* 56: 187–191.
- Mondello L, Lewis AC, and Bartle KD (2001) *Multidimensional Chromatography*. Chichester: John Wiley.
- Phillips JB and Beens J (1999) Comprehensive two-dimensional gas chromatography: A hyphenated method with strong coupling between two dimensions. *Journal of Chromatography A* 856: 331–347.
- Schomburg G (1985) Multidimensional gas chromatography as sampling technique. In: Sandra P (ed.) *Sample Introduction in Capillary Gas Chromatography*, pp. 235–261. Heidelberg: Dr Alfred Huthig Verlag.
- Smith H and Sacks R (1997) Pressure-tunable GC columns with electronic pressure control. *Analytical Chemistry* 69(24): 5159–5164.
- Wilkins CL (1994) Multidimensional GC for qualitative IR and MS of mixtures. *Analytical Chemistry* 66: 295A–301A.

## High-Temperature Techniques

**W Blum and R Aichholz**, Novartis Pharma AG, Basel, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

High-temperature gas chromatography (HTGC) is an extension of conventional gas chromatography (GC) to higher working temperatures. It is generally accepted to use the term 'high temperature' for GC operations starting at the working limit of conventional GC ( $\sim 350^{\circ}\text{C}$ ). The limit for HTGC is not clearly defined but is  $\sim 450^{\circ}\text{C}$ . At a first glance, this extension of the working temperature appears only marginal, yet in practice it is significant. By increasing the maximum working temperature, the application range, in terms of mass units of compounds that can be analyzed, can be extended from  $\sim 600$  to above 1000 Da, depending on the nature and volatility of the compounds. This opens the opportunity to use HTGC for applications that are handled, by general consensus, by alternative chromatographic methods, such as the omnipresent high-pressure liquid chromatography (HPLC).

However, particularly in cases of complex mixtures of more or less apolar compounds, which can be only poorly resolved by HPLC, HTGC is the method of choice for both quantitative and qualitative analysis, because both the separation efficiency and the signal-to-noise ratio of HTGC are significantly higher compared to alternative chromatographic techniques.

Gas chromatographic systems designed for conventional GC cannot be used easily for HTGC work. In practice, HTGC needs special instrumentation. The injector, column oven, detectors, carrier gas,

and, last but not least, the capillary columns should be appropriate for high-temperature operation.

### Instrumental Considerations

#### Injection Techniques

For sample introduction in HTGC, split, splitless, on-column injection, or combinations of the latter, the programmable temperature vaporizing (PTV) injector are suitable. However, with the exception of cold on-column injection, none of these introduction systems fulfill the requirements of an adequate HTGC injection device and guarantees reliable results.

All heated injector systems are well-known sources of error, due to the possibility of artifact formation, and in many cases, split, splitless, or PTV injection provides discrimination of high-boiling compounds by thermal degradation. Only cold on-column injection ensures quantitative sample transfer from the injection port into the capillary column without thermal stress of the samples or discrimination of high-boiling compounds.

In addition, by cold on-column injection of crude samples into a deactivated, but uncoated, precolumn (retention gap), installed in a separate GC oven, the analytical column can be protected from nonvolatile by-products, and the injected sample volume can be significantly enlarged. The large retention gap and the analytical column are linked by a heated transfer line (GC–GC). The retention gap is heated separately and programmed from  $\sim 70^{\circ}\text{C}$  to a temperature  $50^{\circ}\text{C}$  below the maximum working temperature of the analytical column. The temperature program of the analytical column is not started until the temperature of the retention gap exceeds  $120^{\circ}\text{C}$ . A representative example for the direct injection of a

- Hirschfeld T (1980) The hyphenated methods. *Analytical Chemistry* 52: 287A–312A.
- Marriott P and Shellie R (2002) Principles and applications of comprehensive two-dimensional gas chromatography. *Trends in Analytical Chemistry* 21(9–10): 573–583.
- Marriott P, Ong R, Shellie R, *et al.* (2003) Ultra-high resolution capillary gas chromatography by using cryogenic modulation. *Australian Journal of Chemistry* 56: 187–191.
- Mondello L, Lewis AC, and Bartle KD (2001) *Multidimensional Chromatography*. Chichester: John Wiley.
- Phillips JB and Beens J (1999) Comprehensive two-dimensional gas chromatography: A hyphenated method with strong coupling between two dimensions. *Journal of Chromatography A* 856: 331–347.
- Schomburg G (1985) Multidimensional gas chromatography as sampling technique. In: Sandra P (ed.) *Sample Introduction in Capillary Gas Chromatography*, pp. 235–261. Heidelberg: Dr Alfred Huthig Verlag.
- Smith H and Sacks R (1997) Pressure-tunable GC columns with electronic pressure control. *Analytical Chemistry* 69(24): 5159–5164.
- Wilkins CL (1994) Multidimensional GC for qualitative IR and MS of mixtures. *Analytical Chemistry* 66: 295A–301A.

## High-Temperature Techniques

**W Blum and R Aichholz**, Novartis Pharma AG, Basel, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

High-temperature gas chromatography (HTGC) is an extension of conventional gas chromatography (GC) to higher working temperatures. It is generally accepted to use the term 'high temperature' for GC operations starting at the working limit of conventional GC ( $\sim 350^{\circ}\text{C}$ ). The limit for HTGC is not clearly defined but is  $\sim 450^{\circ}\text{C}$ . At a first glance, this extension of the working temperature appears only marginal, yet in practice it is significant. By increasing the maximum working temperature, the application range, in terms of mass units of compounds that can be analyzed, can be extended from  $\sim 600$  to above  $1000\text{ Da}$ , depending on the nature and volatility of the compounds. This opens the opportunity to use HTGC for applications that are handled, by general consensus, by alternative chromatographic methods, such as the omnipresent high-pressure liquid chromatography (HPLC).

However, particularly in cases of complex mixtures of more or less apolar compounds, which can be only poorly resolved by HPLC, HTGC is the method of choice for both quantitative and qualitative analysis, because both the separation efficiency and the signal-to-noise ratio of HTGC are significantly higher compared to alternative chromatographic techniques.

Gas chromatographic systems designed for conventional GC cannot be used easily for HTGC work. In practice, HTGC needs special instrumentation. The injector, column oven, detectors, carrier gas,

and, last but not least, the capillary columns should be appropriate for high-temperature operation.

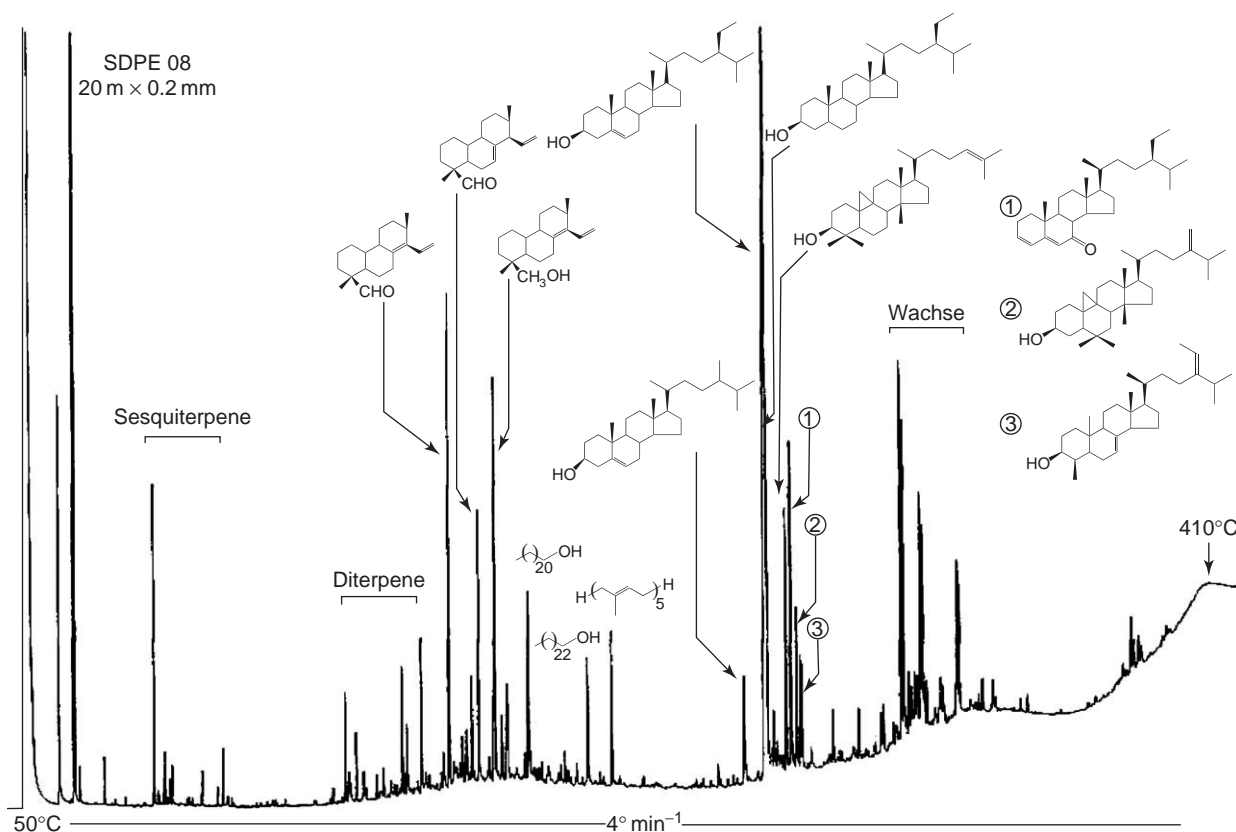
### Instrumental Considerations

#### Injection Techniques

For sample introduction in HTGC, split, splitless, on-column injection, or combinations of the latter, the programmable temperature vaporizing (PTV) injector are suitable. However, with the exception of cold on-column injection, none of these introduction systems fulfill the requirements of an adequate HTGC injection device and guarantees reliable results.

All heated injector systems are well-known sources of error, due to the possibility of artifact formation, and in many cases, split, splitless, or PTV injection provides discrimination of high-boiling compounds by thermal degradation. Only cold on-column injection ensures quantitative sample transfer from the injection port into the capillary column without thermal stress of the samples or discrimination of high-boiling compounds.

In addition, by cold on-column injection of crude samples into a deactivated, but uncoated, precolumn (retention gap), installed in a separate GC oven, the analytical column can be protected from nonvolatile by-products, and the injected sample volume can be significantly enlarged. The large retention gap and the analytical column are linked by a heated transfer line (GC–GC). The retention gap is heated separately and programmed from  $\sim 70^{\circ}\text{C}$  to a temperature  $50^{\circ}\text{C}$  below the maximum working temperature of the analytical column. The temperature program of the analytical column is not started until the temperature of the retention gap exceeds  $120^{\circ}\text{C}$ . A representative example for the direct injection of a



**Figure 1** FID chromatogram of the direct injection of the neutral fraction of Polish Tall oil.

crude extract via a large, separately heated retention gap is shown in **Figure 1**.

### Carrier Gas, Carrier Gas Control

In GC, the most common carrier gas is helium, followed to an increasing degree by hydrogen. When expressed in plate numbers, the separation efficiency of a column neither seems to be significantly influenced by the nature of the carrier gas, nor by the column temperature, and, at a first glance, both gases are equally well suited. Looking at it more closely, it becomes obvious that the separation power of a capillary column strongly depends on both parameters. When the efficiency of a column is measured by *TZ* values (*TZ* = Trennzahl, separation number), it becomes clear that the separation efficiency decreases at higher temperatures. In other words, the higher the retention temperature at which a given compound elutes, the lower the resolution. In order to decrease the retention temperatures and to extend the application range of HTGC, carrier gas flow control or flow programming instead of pressure control is recommended. This in turn will lead to nonoptimal flow conditions, which has negative impact on separation efficiency. With helium as carrier gas this effect is

significant; with lower-viscosity hydrogen this effect is negligible.

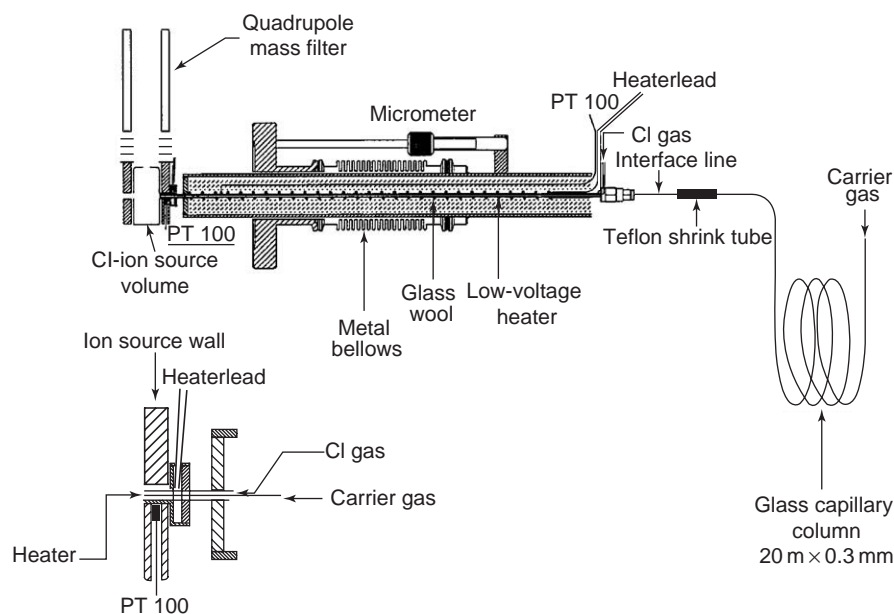
Therefore, most of the proprietary HTGC applications are carried out with hydrogen as a carrier gas.

### Detectors

The detectors used in HTGC are, apart from the highly versatile flame ionization detector (FID), the phosphorus/nitrogen-selective alkali-flame ionization detector (AFID), the atomic emission detector, the inductively coupled plasma (ICP)-mass spectrometer, and, last but not least, mass spectrometers with electron impact and chemical ionization ion sources (EI/CI-MS).

The FID and the AFID have to be equipped with high-temperature resistant flame tips, temperature-protected electronic devices, and reinforced detector heating with detector blocks from bronze or stainless steel. In order to avoid 'cold spots' at the very end of the separation column, it is important that heat transfer is efficient.

The AFID is identical to the FID, except that an alkali metal salt source is placed between the burner tip and the collector. The response of the AFID strongly depends on the concentration of hydrogen in



**Figure 2** Diagram of a high-temperature glass capillary interface linked to quadrupole mass spectrometer. The enlarged presentation shows the entrance into the ion chamber with an additional electrical heater and CI reagent gas entrance coaxial to the interface capillary. (Reproduced from Blum W, Ramstein P, and Eglinton G (1998) Simultaneous coupling of supercritical fluid chromatography and high temperature glass capillary gas chromatography to a mass spectrometer. *Journal of High Resolution Chromatography and Chromatography Communications* 11: 441–448.)

the gaseous boundary of the alkali metal bead. Accurate control of the hydrogen flow, and, if hydrogen is used as mobile phase, of the carrier gas, is required and constant carrier gas flow regulation is mandatory.

Apart from the FID, the mass spectrometer is the most widely used detection device in capillary GC. However, most of the commercially available GC–mass spectrometry (MS) systems do not fulfill the criteria for use with HTGC. The most serious problems encountered are the coupling of the HT-capillary column to the ion source, the choice of carrier gas, and, if CI is used, the reagent gas, and the limited scan range of most GC–MS instruments.

An HTGC–MS system should have a scan range up to  $\geq 1500$  Da. It should be equipped with turbo molecular pumps resistant to corrosive gases, with sufficient pumping capacity to deal with a continuous flow of gas mixtures such as ammonia (CI reagent gas) and hydrogen (carrier gas). The interface temperature should be kept at  $\sim 400^\circ\text{C}$  and as constant as possible along the transfer line up to the very end, in order to avoid ‘cold spots’ and to run the ion source at significantly lower temperatures as the column or the interface line.

Simply inserting the exit of a flexible fused-silica capillary directly into the ionizing chamber, and setting the ion source temperature to the maximum, as is common practice, can create serious background from degradation products of the polyamide outer

coating, and will deteriorate the quality of the mass spectra, in particular of the CI spectra. If aggressive reagent gases, such as ammonia, are introduced in CI mode, the background can rise to intolerable levels. A highly inert, coaxial HTGC–MS glass interface system can be used in both electron impact (EI) and CI modes at moderate ion source temperatures, which takes all requirements mentioned above into account (see Figure 2).

### High-Temperature Capillary Columns

The heart and the soul of any GC system is the separation column. Historically, three support materials were used in capillary GC. In chronological order of column development this was metal, borosilicate glass, and fused silica.

Modern GC is based almost exclusively on fused-silica capillaries, which has consequences for HTGC. Despite the fact that the thermal stability of the common polyamide clad columns could be improved to allow temperatures up to  $400^\circ\text{C}$ , one inherent problem of fused-silica remains. It was reported that already at ambient temperatures, low molecular gases like helium or hydrogen permeate through the fused-silica capillary wall and its outer coating. This will inevitably limit the application range for fused silica, in particular for HTGC applications under pressure regulation.

Although a systematic comparison among the various high-temperature resistant supports has never been undertaken, for most of the few groups dealing with fundamental applications in HTGC, borosilicate glass is still the favorite support material.

Nearly all stationary phases used in HTGC are polysiloxanes, which, apart from a few exceptions, bear reactive termini such as OH or CH<sub>3</sub>O groups. Compared to coatings based on traditional end-capped polysiloxanes, OH or CH<sub>3</sub>O terminated polymers showed a significantly improved thermal stability when coated on silanol-rich surfaces. Pronounced covalent bonding of the coatings leads to increased inertness and less catalytic degradation.

The key mechanism of the thermal degradation of OH-terminated polysiloxane coatings in an inert environment is given in **Scheme 1**. Polysiloxane chains with terminal silanol groups can undergo nucleophilic substitution, which leads to the formation of low molecular weight cyclic silyloxy compounds (bleeding). The thermal depolymerization is governed mainly by structural and kinetic factors.

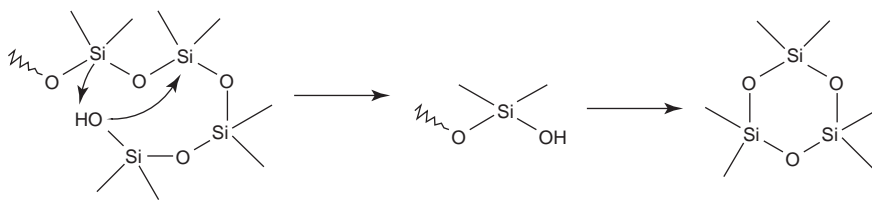
Any measure that avoids or suppresses this 'unzipping' mechanism results in improved thermal stability of the coating.

The degree of thermal degradation can be decreased by lowering the density of the silanol group. This can be achieved by further polymerization of the stationary phase by addition of so-called polyfunctional cross-linkers, e.g., trimethoxysilane, as shown in **Scheme 2**.

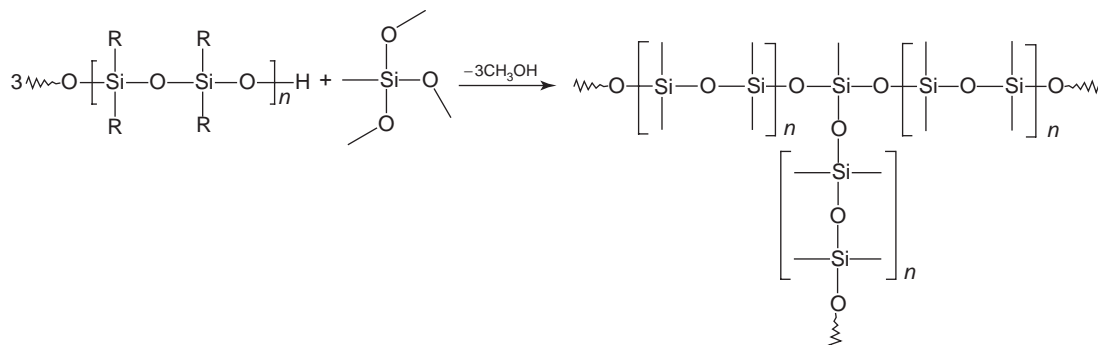
Moreover, the covalent chemical bonding between the mobile phase and the column wall stabilizes the coating. The cross-linking takes place even at room temperature by partial condensation of terminal methoxy groups with surface silanol groups, or vice versa. At elevated temperatures, i.e., during thermal conditioning of the column, it proceeds by intramolecular nucleophilic substitution and release of volatile polymer substituents like benzene. The stabilization of the coating at elevated temperatures by autocross-linking is characteristic for OH-terminated polysiloxane phases. In conventional GC, comparable reactions are of minor importance.

The thermal stability of polysiloxane stationary phases is also influenced by structural factors. Chemical modifications of the polymer backbone, e.g., by introducing planar groups such as phenyl, diphenyl, diphenylether, etc., reduce the flexibility of the polysiloxane backbone and its ability to undergo the 'unzipping' reaction shown in **Scheme 1**. Consequently, the maximum working temperatures of the respective polymer increases (see **Tables 3–5**). The most common polymer modifications are shown in **Scheme 3**.

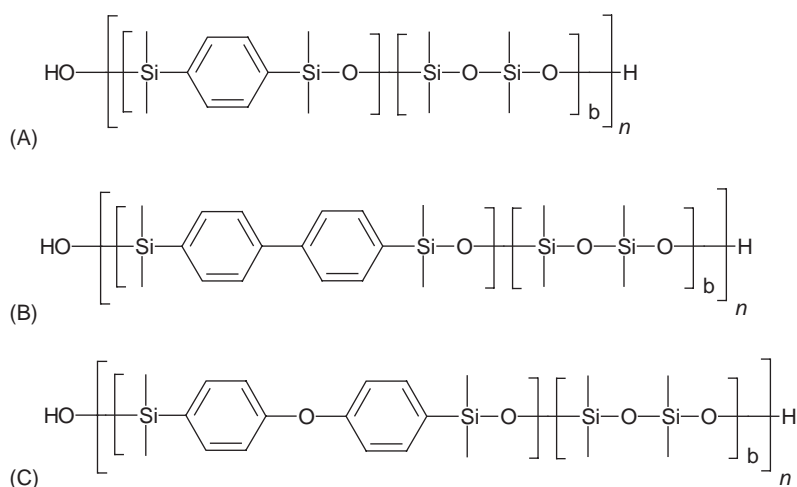
Besides the modifications of the silicone backbone, the order of the substituents along the polymer chain is of fundamental importance for thermal stability. If a polymer bears different polar or polarizable functional groups, as is common for most of the medium-polarity phases, the thermal stability of the coating will be improved if the polymer



**Scheme 1** Thermal depolymerization of an OH-terminated polysiloxane in inert environment.



**Scheme 2** Mechanism of *in situ* cross-linking of OH-terminated polysiloxanes by methyltrimethoxy silane.



**Scheme 3** Structures of high-temperature stable siloxane copolymers used as stationary phases in HTGC: (A) silphenylene/siloxane, (B) sildiphenyl/siloxane, (C) sildiphenylether/siloxane.

**Table 1** Technical polysiloxanes suited for HTGC, symmetrically substituted, OH- (CH<sub>3</sub>O-) terminated polysiloxanes phases, maximum working temperature (on glass) > 400°C

$\text{HO}-\left[ \left[ \text{Si}-\text{O}-\text{Si}-\text{O} \right] \left[ \text{Si}-\text{O}-\underset{\text{R}}{\underset{\text{R}}{\text{Si}}}-\text{O} \right] \right]_n-\text{H}$ <p>R = phenyl</p>			
PS-347.5	100% Dimethyl		OH-terminal
PS-349.5	100% Dimethyl		OH-terminal
PS-089	95% Dimethyl	5% Diphenyl	OH-terminal
OV-1-OH	100% Dimethyl		OH-terminal
PS-086	85% Dimethyl	15% Diphenyl	OH-terminal
PS-090	80% Dimethyl	20% Diphenyl	CH <sub>3</sub> O-terminal

**Table 2** Commercially available asymmetrically substituted, OH-terminated polysiloxane phases, designed for gas chromatography, working temperatures (on glass) 350–370°C

$\text{HO}-\left[ \left[ \text{Si}-\text{O}-\text{Si}-\text{O} \right] \left[ \text{Si}-\text{O}-\underset{\text{R}}{\text{Si}}-\text{O} \right] \right]_n-\text{H}$ <p>R = phenyl, cyanopropyl</p>			
OV-1701-OH	88% Dimethyl	5% Phenyl, 7% cyanopropyl	OH-terminal
OV-31-OH	83% Dimethyl	17% Cyanopropyl	OH-terminal
OV-240-OH	67% Dimethyl	33% Cyanopropyl	OH-terminal
OV-61-OH	67% Dimethyl	33% Phenyl	OH-terminal
OV-17-OH	50% Dimethyl	50% Phenyl I	OH-terminal
OV-225-OH	50% Dimethyl	25% Phenyl, 25% cyanopropyl	OH-terminal

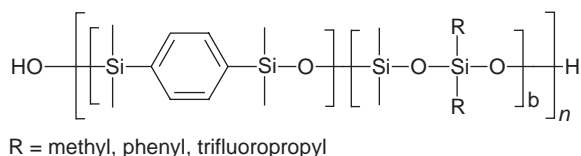
is symmetrically substituted. Asymmetric substitution leads to polarization at the respective silicon atoms, resulting in an easier elimination of functional groups at elevated temperatures, induced by intramolecular nucleophilic attacks. The inadequate thermal stability of polysiloxane phases listed

in Table 2 results, above all, from asymmetrical substitution.

Table 1 shows symmetrically substituted, technical polysiloxanes suitable for HTGC.

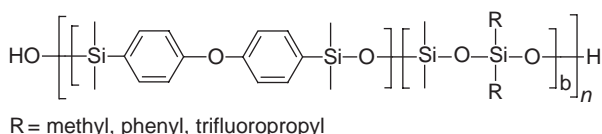
Table 2 lists commercially available, asymmetrically substituted OH-terminated GC phases. These



**Table 3** OH-terminated silphenylene/siloxane copolymer phases, specially designed for high-temperature work

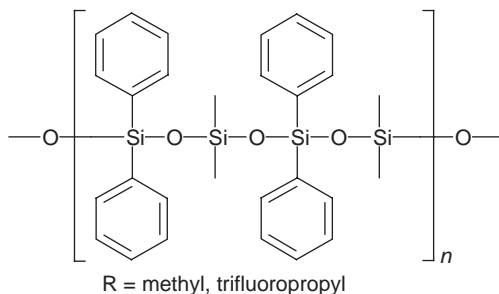
SAP-0/1	50% Dimethylsiloxane	50% Tetramethylsilphenylene	OH-terminal
SAP-0/2	60% Dimethylsiloxane	40% Tetramethylsilphenylene	OH-terminal
SAP-15/1	35% Dimethyl, 15% diphenylsiloxane	50% Tetramethylsilphenylene	OH-terminal
SAP-15/2	45% Dimethyl, 15% diphenylsiloxane	40% Tetramethylsilphenylene	OH-terminal
SA-TF	60% Trifluoropropylmethylsiloxane	40% Tetramethylsilphenylene	OH-terminal

Maximum working temperature (on glass) > 400°C.

**Table 4** OH-terminated silphenylether/siloxane copolymer phases, specially designed for high-temperature work

SDPE-0/1	60% Dimethylsiloxane	40% Tetramethylsilphenylether	OH-terminal
SDPE-08/1	45% Dimethylsiloxane, 15% diphenylsiloxane	40% Tetramethylsilphenylether	OH-terminal
SDPE-TF	60% Trifluoropropylmethylsiloxane	40% Tetramethylsilphenylether	OH-terminal

Maximum working temperature (on glass) > 400°C.

**Table 5** Symmetrically substituted polysiloxane phases, specially designed for high-temperature work

SOP-50	50% Dimethylsiloxane	50% Diphenylsiloxane	CH <sub>3</sub> O-terminal
SOP-50-TF	50% Diphenylsiloxane	50% Trifluoropropylmethylsiloxane	CH <sub>3</sub> O-terminal
SOP-50-PFD	50% Diphenylsiloxane	50% Perfluorodecylmethylsiloxane	CH <sub>3</sub> O-terminal

Maximum working temperature (on glass) > 400°C.

phases are not designed for HTGC, and can be used up to 350–370°C only.

The stationary phases listed in **Tables 3–5** are specially designed for high-temperature work. These phases can easily be synthesized on a laboratory scale. The underlying methods are reported and are easy to reproduce.

## Applications

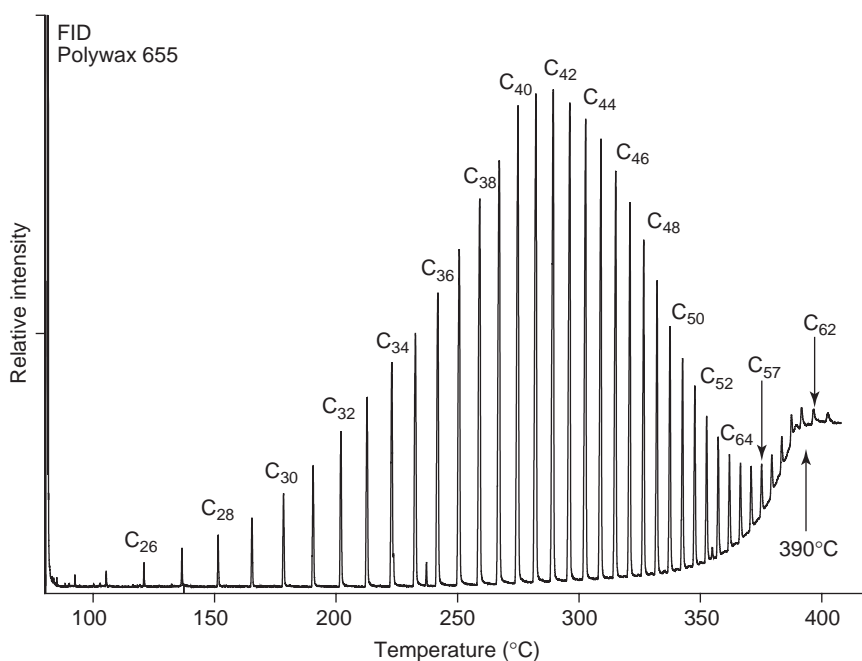
### Organic Geochemistry

In organic geochemistry HTGC is an established tool. The applications are mainly focused on aspects

such as simulated distillation of crude oils and petroleum residues, and the analysis of high molecular weight waxes and biomarkers and in geochemical petroleum prospecting (**Figure 3**).

### Simulated Distillation

Distillation is a standard procedure to classify crude oils into a manageable number of compounds. HTGC techniques were successfully introduced to reduce the analysis time significantly. Since the first report on simulated distillation of heavy petroleum fractions by HTGC up to a true boiling point of 800°C, the usefulness of simulated distillation, in



**Figure 3** FID chromatogram of a homologous series of aliphatic hydrocarbons C<sub>30</sub>–C<sub>60</sub> separated on a SOP-50-PFD (unpublished results).

order to classify fossil materials with a wide carbon number distribution, is generally accepted in the petrol industry. Usually simulated distillation is performed on short 5–10 m, thin film, apolar coated capillary columns.

### Waxes and Biomarkers

The higher molecular weight hydrocarbons (>C<sub>40</sub>) and asphaltenes are important constituents of crude oils. Despite their relatively low concentration, they can cause serious problems related to crystallization and disposition of paraffin waxes during oil production and transportation, and in formation of tar mats. Thus, it is important to have a method, such as HTGC or HTGC–MS, for accurate assessment of the weight of wax fractions in crude oil, as well as an accurate assessment of the associated asphaltene fraction.

Biomarkers are compounds whose carbon skeletons can be unambiguously linked with known natural precursors. They have been used to obtain information about the source, disposition, biodegradation, migration, and other information concerning the biochemistry of crude oils. Possible sources of higher molecular weight hydrocarbons include waxes and wax esters from precursors in living organism such as higher-plant cuticular waxes, or aliphatic material from algaenans, which can provide an important source of *n*-alkanes in marine oils, etc.

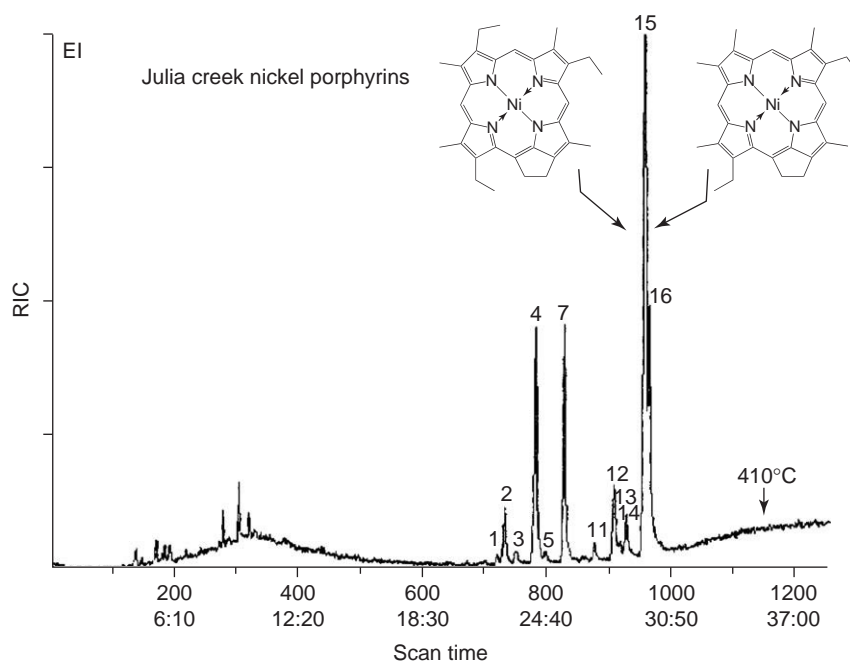
The investigation of other classes of biomarkers such as chlorines and porphyrins has shown a similar trend toward the discovery of novel high molecular mass compounds. One reason for the rapid increase in the number of identified biomarkers is the availability of HTGC–MS (Figure 4).

### Natural Products

Many studies have shown that HTGC, in particular in combination with MS, can extend the application range for the rapid analysis of complex natural product samples. Although complete structure elucidation of the separated compounds solely by MS has, in most cases, a poor prognosis, modern high-resolution MS/MS techniques in combination with HTGC permit fast characterization of several classes of natural products such as *n*-alkanes and alkenes, steroids, terpenes, alcohols, aldehydes, carboxylic acids, esters, triglycerides, amino acids, vitamins, and oligosaccharides.

### Plant Materials, Lipids

The separation and identification of the volatile ingredients of plant extracts is a demanding task. Most of the extracts cover a wide boiling range, in which the high-boiling compounds remain unidentified. Representative applications are the identification of long-chain triterpenoidylesters in smoke from biomass combustion and the characterization of



**Figure 4** HTGC-MS EI-total ion current chromatogram (TIC) of a Julia creek oil shale extract. Column: OV-61-OH, 10 m  $\times$  0.2 mm, glass. Carrier gas: H<sub>2</sub>. (Reproduced from Blum W, Ramstein P, and Eglinton G (1988) Simultaneous coupling of supercritical fluid chromatography and high temperature glass capillary gas chromatography to a mass spectrometer. *Journal of High Resolution Chromatography and Chromatography Communications* 11: 441–448.)

triacylglycerols containing oxygenated fatty acid acyl groups (Figure 5).

### Triglycerides and Wax Esters

Triglycerides, fatty acids, and fatty acid-esters are the main ingredients of vegetable oils and animal fats. The separation and characterization of the complex mixtures of saturated and unsaturated, long-chain fatty acids, mono-, di-, and triglyceride, and fatty acid esters is of interest in food industry in order to classify and differentiate the various crude fats and oils. HTGC and HTGC-MS are particularly powerful tools for the separation and characterization of complex mixtures of high-boiling, long-chain fatty acid esters (wax esters).

As mentioned before, in GC low retention temperatures are generally advantageous, because the higher the chosen working temperature, the lower is the separation efficiency. Therefore, in practice of HTGC not only the maximum working temperature of a column is of importance, but also the retention power of its coating. In this context it was reported that the system-inherent low retention power of perfluoroalkyl-polysiloxane coatings significantly reduces the retention temperatures, compared to nonfluoroalkyl-substituted polysiloxanes, without any loss of column efficiency.

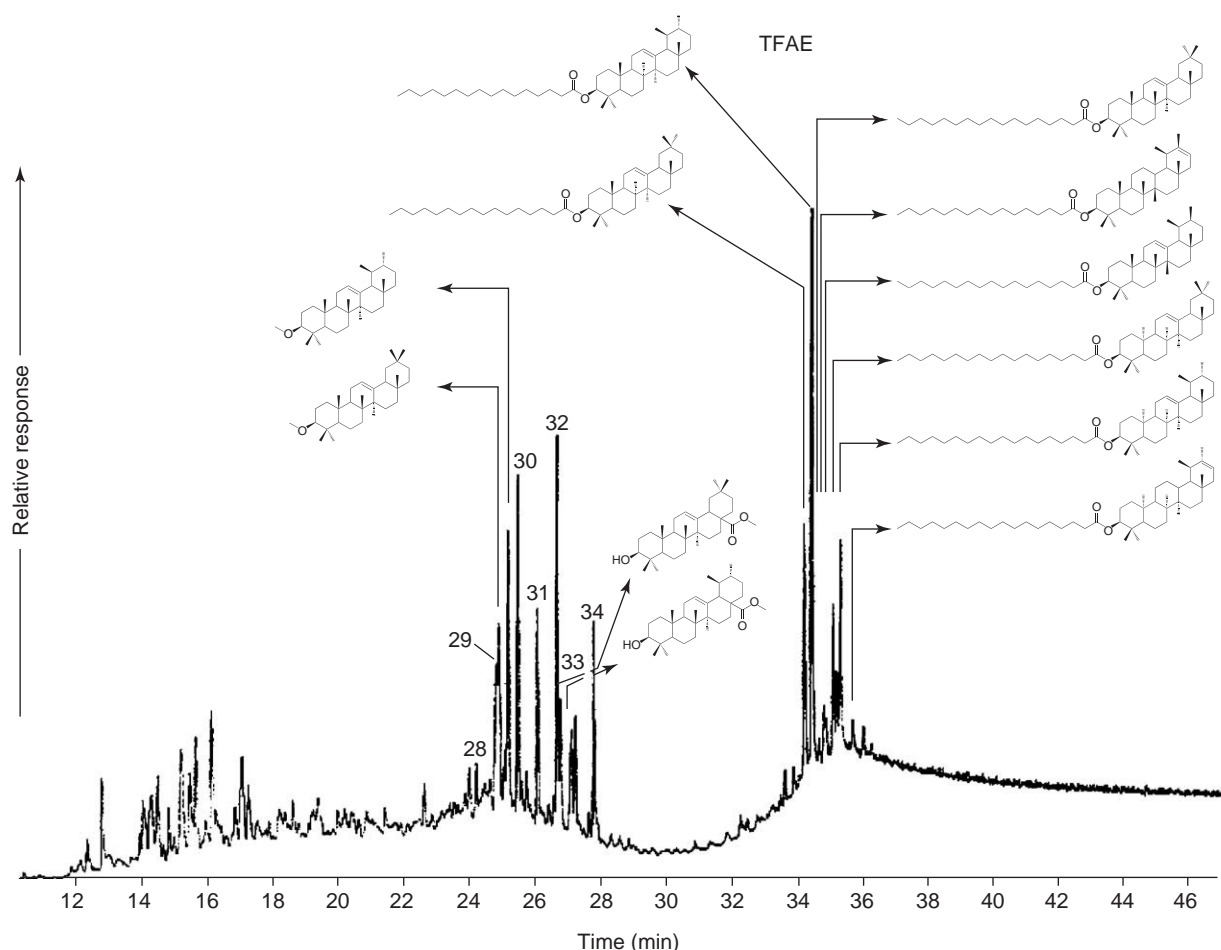
The extension of the working range of HTGC on a poly(diphenyl-perfluorodecylmethyl) siloxane phase (SOP-50-PFD, see Table 5) is demonstrated in HTGC and HTGC-MS/MS investigations of crude comb-waxes of various honey bee species. By pattern recognition of the resulting wax-ester chromatograms, a chemotaxonomic distinction among the honeybee species was feasible (Figure 6).

### Oligosaccharides

Free or covalently bound oligosaccharides (glycoproteins, glycolipids) are widely distributed in nature. For example, it has been suggested that most common proteins are glycosylated and that glycosylation plays an important role in biological function, dynamics, and physiochemical properties. ‘Sugar mapping’, e.g., by HTGC-MS/MS has proven to be a useful tool in order to characterize the complex oligosaccharide mixtures released from glycoproteins. Representative for this application is the analysis of the complex mixture of O-linked oligosaccharides after premethylation, released from muscin glycopeptides from porcine small intestine, as shown in Figure 7.

### Industrial Products

Most of the reports dealing with HTGC analysis of industrial products made on a ton scale are



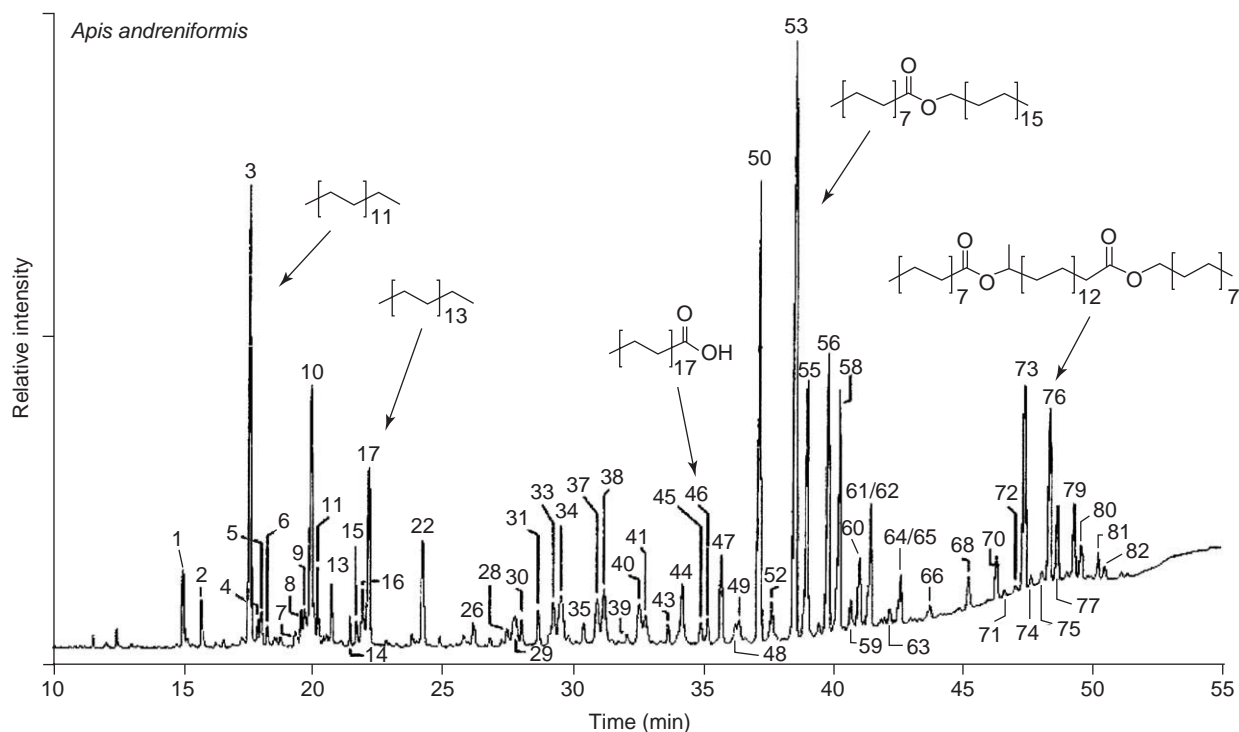
**Figure 5** Representative HTGC–MS total ion current trace of the ester fraction in smoke extract from Castanha-do-Para. Column: OV-1701-OH, 20 m × 0.3 mm, glass. Carrier gas: He. (Reproduced with permission from Vladimir OE, Bernd RTS, Alberto SP, and Jari NC (1998) High temperature gas chromatography with a glass capillary column for the analysis of high molecular weight tracers in smoke samples from biomass burning. *Journal of High Resolution Chromatography* 21(2): 87–93; © Wiley-VCH.)

investigations of synthetic organic chemicals after release into the aquatic environment. Many of them are toxic at low concentrations and accumulate in sediments or organisms. In most of the cases pre-concentration of the sample involves liquid–liquid extraction, e.g., with dichloromethane, or solid-phase extraction.

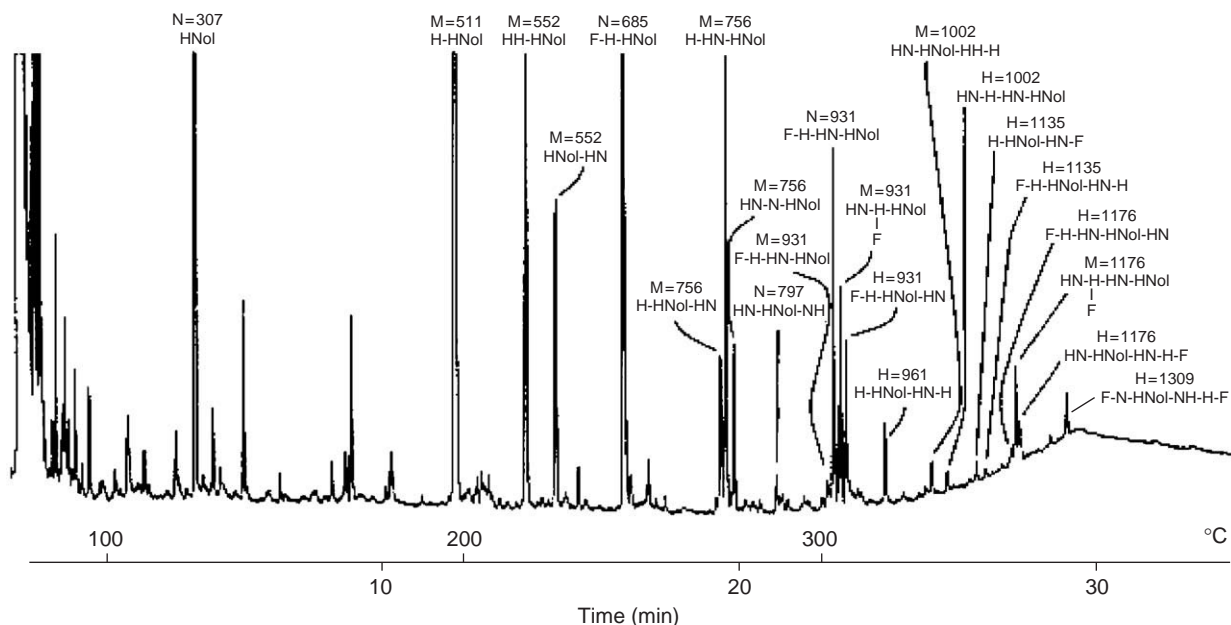
The compound classes that can be tackled successfully by HTGC cover a wide range of organic chemicals such as polychlorinated biphenyls, antioxidants, and ultraviolet (UV) stabilizers, non-ionic surfactants, commonly used as wetting agents or emulsifiers, polymer degradation products of bisphenol-A epoxy resins, rolling mill oils, etc. Representative for the HTGC analysis of industrial products is the separation of an artificial mixture of omnipresent antioxidants and UV stabilizers, up to molecular weights above 1000 Da (Figure 8).

## Concluding Remarks

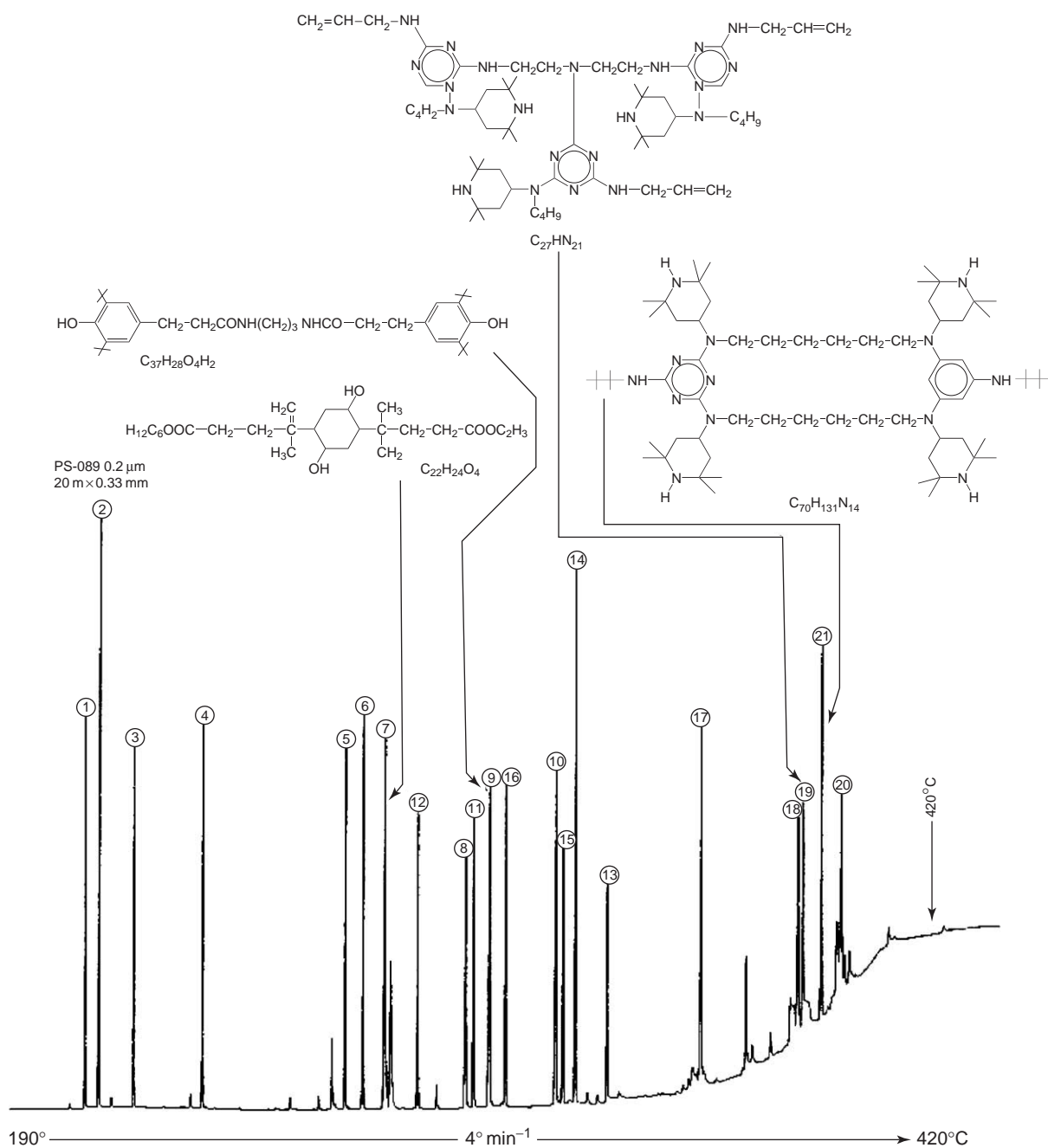
Gas chromatography, in particular HTGC, is a mature analytical technique. The underlying HTGC column technology was developed more than 15 years ago. Since then only little innovation was added. What has changed over the years is the attitude of the succeeding generation of chromatographers, from column makers to column consumers. Today, column preparation is fully commercialized. The secret behind outstanding HTGC applications, the deeper knowledge of the ‘art and the science’ of column technology, and the craft of column preparation is no longer a subject of education of analytical chemists, and faded, more or less, into oblivion. Many of the outstanding applications, published in the late 1980s and the early 1990s, can hardly be reproduced on the basis of today’s commercially available columns.



**Figure 6** FID chromatogram of TBDMS/TMS-derivatized crude combwax of *Apis andreniformis*. Column: SOP-50-PFD, 10 m  $\times$  0.2 mm, glass. Carrier gas: H<sub>2</sub>. (Reproduced with permission from Aicholz R and Lorbeer E (1999) Investigation of comb wax of honeybees with high-temperature gas chromatography-chemical ionization mass spectrometry. II High-temperature gas chromatography-chemical ionization mass spectrometry. *Journal of Chromatography B* 883: 75–88; © Elsevier.)



**Figure 7** FID chromatogram of a permethylated oligosaccharides from mucin-glycopeptide of small intestine of the pig. H = hexose; F = fucose; HN = acetylhexosamine; HNol = N-acetylhexaminitol. Column: SE-54, 10 m  $\times$  0.33 mm, glass. Carrier gas: H<sub>2</sub>. (Reproduced with permission from Karlson H, Carstedt I, and Hanson GC (1989) The use of gas chromatography-mass spectrometry for the characterization of permethylated oligosaccharides with molecular mass up to 2300. *Analytical Biochemistry* 182: 438–446; © Elsevier.)



**Figure 8** FID chromatogram of antioxidants and UV stabilizers. Column: PS-089, 20 m  $\times$  0.3 mm, glass. Carrier gas:  $\text{H}_2$ . (Reproduced from Blum W and Damasceno L (1987) High temperature glass capillary gas chromatography using OH-terminated polysiloxane stationary phases. Separation of antioxidants and UV stabilizers. *Journal of High Resolution Chromatography and Chromatography Communications* 10: 472–476.)

The most important detector in capillary column chromatography is the mass spectrometer. Experience has shown that any chromatographic method that cannot be coupled with a mass spectrometer remains of little practical relevance in the long term. In this context, it must be taken into consideration that high-temperature GC–MS was developed during a transitional phase, at a time when HPLC–MS was

still in its infancy and of limited practical importance. This has changed over the years. The sensitivity and simplicity of modern HPLC–MS systems, equipped with electrospray ionization sources (ESI), can easily compete with GC–MS instruments. Therefore, nowadays most of the separation problems in the medium and high molecular mass range are tackled by HPLC–MS.



Although in some reports the future of HTGC and HTGC–MS is pointed out as promising, some doubts remain. HTGC was and is a demanding borderline technique, requiring experienced specialists to successfully put into action. The necessary equipment is only partially commercially available and has to be modified and adapted to specific tasks. Most of the outstanding applications are performed in a handful of key laboratories, active in this field for many years, where the art of HTGC column making is not completely forgotten. In their hands the method is a mass powerful tool and may have a future. Many complex apolar and medium molecular mass or artificial mixtures, such as lipids and waxes, wax esters, polysaccharides, additives, fashionably analyzed nowadays by HPLC or HPLC–MS, could be tackled much more successfully by HTGC and HTGC–MS, due to the significantly higher separation efficiency of GC. In these application fields HTGC is difficult to beat, and there the method may find a niche.

See also: **Gas Chromatography:** Column Technology; Instrumentation; Online Coupled LC–GC; Pyrolysis; Detectors; Mass Spectrometry.

## Further Reading

- Aichholz R and Lorbeer E (1999) Investigation of comb wax of honeybees with high-temperature gas chromatography and high-temperature gas chromatography–chemical ionization mass spectrometry. I. High-temperature gas chromatography. *Journal of Chromatography B* 855: 601–615.
- Blum W (1985) Preparation of inert and high temperature stable apolar and medium polar glass capillary columns using OH-terminated polysiloxane stationary phases. *Journal of High Resolution Chromatography and Chromatography Communications* 8: 719–726.
- Blum W, Richter WJ, and Eglinton G (1988) Glass capillary chromatography–mass spectrometry at high temperatures. Direct analysis of free base porphyrins and metal porphyrin complexes extracted from the serpiano oil shale. *Journal of High Resolution Chromatography and Chromatography Communications* 11: 148–155.
- Blum W and Aichholz R (1991) *Hochtemperatur Gas Chromatographie*. Heidelberg: Hüthig Verlag.
- Blum W and Aichholz R (1993) High temperature stable CH<sub>3</sub>O-terminated poly(diphenyl/dimethyl)- and poly(diphenyl/3,3,3-trifluoropropylmethyl)siloxane co-polymer stationary phases for capillary gas chromatography. *Journal of Microcolumn Separation* 5: 297–302.
- Hsieh M, Philp RP, and del Rio JC (2000) Characterization of high molecular weight biomarkers in crude oil. *Organic Geochemistry* 1581: 1581–1588.
- Pereira A, Pinto AC, Cardoso JN, *et al.* (1998) Application of high temperature high resolution gas chromatography to crude extracts of propolis. *Journal of High Resolution Chromatography* 21: 396–400.
- Pereira A and Radler de Aquino Neto F (1999) High temperature high-resolution gas chromatography: Breaching the barrier to the analysis of polar and high molecular weight compounds. *Trends in Analytical Chemistry* 18: 126–136.
- Pereira A, Carbonell SA, Radler de Aquino Neto F, Fernandes do Amaral AC, and Barnes RA (2002) High-temperature gas chromatography–mass spectrometry with glass capillary column for screening of natural products. *Journal of Chromatography A* 947: 255–265.

## High-Speed Techniques

**P J Marriott and R A Shellie**, RMIT University, Melbourne, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

There are many definitions of fast gas chromatography (GC). For some, the definition of fast may be ‘that we achieved the result in less time than we did yesterday’. Others adhere strictly to rules such as ‘the inherent speed of a given separation is best defined by the width of peaks’ and then seek to minimize this width for all peaks. In practical GC, the best definition of fast will be dependent on the analysis task at hand. The approach that an analyst accepts as a ‘fast’

method will be determined by the sample, and also by the type (or amount) of information required from the sample. Typical approaches can be exemplified by various ways and information can be derived from GC as follows:

- For the analysis of complex samples, for example, detailed hydrocarbon analysis of naphtha and gasoline, or flavor and fragrance analysis, often the only suitable approach to speeding up the analysis is to implement a method that provides a short analysis time but still achieves a very high peak capacity.
- Where the sample is less complicated, or where only a small number of components in the sample are of interest, the requirement for very high peak capacity may be relaxed. Here, the main criterion for selecting the appropriate conditions is to ensure

Although in some reports the future of HTGC and HTGC–MS is pointed out as promising, some doubts remain. HTGC was and is a demanding borderline technique, requiring experienced specialists to successfully put into action. The necessary equipment is only partially commercially available and has to be modified and adapted to specific tasks. Most of the outstanding applications are performed in a handful of key laboratories, active in this field for many years, where the art of HTGC column making is not completely forgotten. In their hands the method is a mass powerful tool and may have a future. Many complex apolar and medium molecular mass or artificial mixtures, such as lipids and waxes, wax esters, polysaccharides, additives, fashionably analyzed nowadays by HPLC or HPLC–MS, could be tackled much more successfully by HTGC and HTGC–MS, due to the significantly higher separation efficiency of GC. In these application fields HTGC is difficult to beat, and there the method may find a niche.

**See also:** Gas Chromatography: Column Technology; Instrumentation; Online Coupled LC–GC; Pyrolysis; Detectors; Mass Spectrometry.

## Further Reading

- Aichholz R and Lorbeer E (1999) Investigation of comb wax of honeybees with high-temperature gas chromatography and high-temperature gas chromatography–chemical ionization mass spectrometry. I. High-temperature gas chromatography. *Journal of Chromatography B* 855: 601–615.
- Blum W (1985) Preparation of inert and high temperature stable apolar and medium polar glass capillary columns using OH-terminated polysiloxane stationary phases. *Journal of High Resolution Chromatography and Chromatography Communications* 8: 719–726.
- Blum W, Richter WJ, and Eglinton G (1988) Glass capillary chromatography–mass spectrometry at high temperatures. Direct analysis of free base porphyrins and metal porphyrin complexes extracted from the serpiano oil shale. *Journal of High Resolution Chromatography and Chromatography Communications* 11: 148–155.
- Blum W and Aichholz R (1991) *Hochtemperatur Gas Chromatographie*. Heidelberg: Hüthig Verlag.
- Blum W and Aichholz R (1993) High temperature stable CH<sub>3</sub>O-terminated poly(diphenyl/dimethyl)- and poly(diphenyl/3,3,3-trifluoropropylmethyl)siloxane co-polymer stationary phases for capillary gas chromatography. *Journal of Microcolumn Separation* 5: 297–302.
- Hsieh M, Philp RP, and del Rio JC (2000) Characterization of high molecular weight biomarkers in crude oil. *Organic Geochemistry* 1581: 1581–1588.
- Pereira A, Pinto AC, Cardoso JN, *et al.* (1998) Application of high temperature high resolution gas chromatography to crude extracts of propolis. *Journal of High Resolution Chromatography* 21: 396–400.
- Pereira A and Radler de Aquino Neto F (1999) High temperature high-resolution gas chromatography: Breaching the barrier to the analysis of polar and high molecular weight compounds. *Trends in Analytical Chemistry* 18: 126–136.
- Pereira A, Carbonell SA, Radler de Aquino Neto F, Fernandes do Amaral AC, and Barnes RA (2002) High-temperature gas chromatography–mass spectrometry with glass capillary column for screening of natural products. *Journal of Chromatography A* 947: 255–265.

## High-Speed Techniques

**P J Marriott and R A Shellie**, RMIT University, Melbourne, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

There are many definitions of fast gas chromatography (GC). For some, the definition of fast may be ‘that we achieved the result in less time than we did yesterday’. Others adhere strictly to rules such as ‘the inherent speed of a given separation is best defined by the width of peaks’ and then seek to minimize this width for all peaks. In practical GC, the best definition of fast will be dependent on the analysis task at hand. The approach that an analyst accepts as a ‘fast’

method will be determined by the sample, and also by the type (or amount) of information required from the sample. Typical approaches can be exemplified by various ways and information can be derived from GC as follows:

- For the analysis of complex samples, for example, detailed hydrocarbon analysis of naphtha and gasoline, or flavor and fragrance analysis, often the only suitable approach to speeding up the analysis is to implement a method that provides a short analysis time but still achieves a very high peak capacity.
- Where the sample is less complicated, or where only a small number of components in the sample are of interest, the requirement for very high peak capacity may be relaxed. Here, the main criterion for selecting the appropriate conditions is to ensure

that only the target analytes are separated, and only with the minimum resolution required to provide a sufficiently accurate result.

- Maximizing the selectivity of the chromatographic system may allow a reduction of analysis time. This type of approach is especially useful for providing improved resolution for samples of low to moderate complexity.
- A further powerful approach is to use a selective detector. This can simplify method optimization for specialized analyses because the selectivity of the detector toward the target analyte may reduce the requirement for chromatographic resolution of the target analyte from interferences.

For many applications only the peakwidth definition will be appropriate; however, the 'less time than yesterday' definition will also be satisfactory for many analyses. For the purpose of clarity, in this article the authors have chosen to use the following terms:

1. High-speed analysis: this description shall describe GC analyses that have short analysis times, but do not necessarily satisfy the requirements of the peakwidth description. These analyses will generally be achieved by sacrificing resolution, by increasing selectivity toward a target analyte, by using selective detection, or a combination of any of these.
2. Fast gas chromatography: this term will be used here to describe analyses that maintain high efficiency. Fast-GC, very fast GC, and ultrafast-GC – GC can be distinguished by the relative widths of peaks in the chromatogram. A list of typical characteristic peakwidths defining each category of GC is given in Table 1.

Regardless from which approach the most appropriate conditions for the required analysis are drawn, it is important to consider some fundamental GC theory.

**Table 1** Comparison of calculated plate number, retention time, and peak-width at half-height using standard, fast, very fast, and ultrafast GC

Type of analysis	$d_c$ ( $\mu\text{m}$ )	$L$ (m)	$N$	$t_R$ (s)	$\omega_n$ (s)
Standard	320	25	90 000	160	1
Fast	50	10	260 000	60	0.2
Very fast	50	1	25 000	2.0	0.03
Ultrafast	50	0.3	7 000	0.40	0.01

Conditions for these calculations are:  $\beta = 62.5$ , compound = hexane,  $T = 330$  K, helium carrier gas. See the above reference for details of the calculation program.

Reprinted with permission from van Deursen MM, Beens J, Janssen H-G, Leclercq PA, and Cramers CA (2000) Evaluation of time-of-flight mass spectrometric detection for fast gas chromatography. *Journal of Chromatography A* 878: 205–213; © Elsevier.

## Some Fundamental GC Theory and Routes toward High-Speed Analysis

In practical GC, it is important that the number of theoretical plates required ( $N_{\text{req}}$ ) in order to provide a desired resolution ( $R_s$ ) for a given pair of components (defined by  $\alpha$  and  $k$ ) can be calculated. This can be achieved by using the Purnell equation (eqn [1]):

$$N_{\text{req}} = 16R_s^2 \left( \frac{\alpha}{\alpha - 1} \right)^2 \left( \frac{k + 1}{k} \right)^2 \quad [1]$$

where the separation factor  $\alpha = t'_{R2}/t'_{R1}$ , and retention factor  $k$  refers to the last eluting peak in the pair.  $R_s$  is measured experimentally as

$$R_s = \frac{\Delta t_R}{4\sigma} \quad [2]$$

Equation [1] can be applied to assist in the selection of the appropriate chromatographic conditions to provide the minimum resolution required for the analysis task. Baseline resolution is achieved at  $R_s > 1.5$ , and will be sufficient to provide accurate quantitative data in most applications. In many cases solutes tend to be over-separated ( $R_s \gg 2$ ), and excess resolution equates to excessive analysis time.

By extending this theory it is possible to determine the minimum time in which sufficient resolution of a critical pair of peaks can be achieved (eqn [3]):

$$t_R = 16R^2 \left[ \frac{H}{\bar{u}} \right] \left[ \left( \frac{(1+k)^3}{k^2} \right) \left( \frac{\alpha}{\alpha - 1} \right)^2 \right] \quad [3]$$

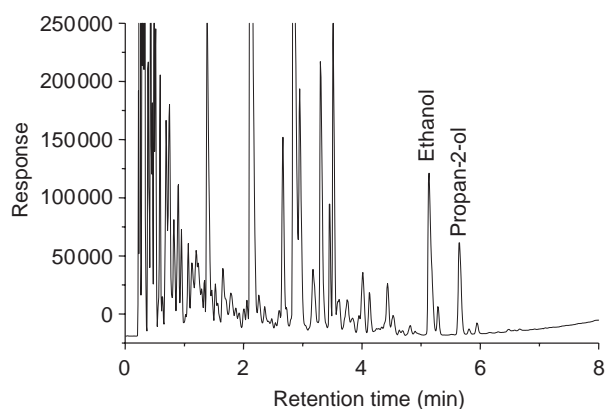
Equation [3] reveals that the column temperature (because this influences  $k$ ), carrier gas linear velocity (which will change the plate duration (the Purnell criterion)  $H/\bar{u}$ ), and column selectivity (affects  $k$ ;  $\alpha/\alpha - 1$ ) are important considerations for reducing analysis time.

Since it is likely that the critical pair of components will not be the last eluting peaks, they do not provide the best indication of the total analysis time; however, this can be given by eqn [4]:

$$t_R = (L/\bar{u})(k_n + 1) \quad [4]$$

where  $L$  is the column length,  $\bar{u}$  the average linear velocity of the carrier gas, and  $k_n$  the retention factor of the last eluting compound in the sample.

Equation [4] is very useful for deducing the general operating requirements of high-speed analysis. The use of a short column, a higher than usual carrier gas velocity, and relatively small retention factors (which can be easily achieved by using high temperature and/or thin film columns), can reduce analysis times



**Figure 1** The use of a selective stationary phase column for the high-speed analysis of ethanol in unleaded petrol.

significantly. There will, however, be a penalty associated with reducing the analysis time by using these straightforward steps, namely a reduction in the resolving power due to reduced peak capacity and increased band broadening. Equation [4] does not implicitly allow the effective resolution of the system to be evaluated.

Column selection and stationary phase selectivity can play a major role in reducing the requirement for very high plate numbers. For example, a relatively new PLOT capillary column, which uses a polar adsorbent as the stationary phase, has unique selectivity toward oxygenated compounds. The capability of such a stationary phase column to provide excellent separation of oxygenates in the presence of hydrocarbons is illustrated in **Figure 1**. Stationary phase selectivity can be manipulated to achieve better separation of target compounds in a shorter time than would normally be required using a nonselective column, but stationary phase optimization will be limited to the analysis of low-complexity samples, i.e., those containing fewer than say 50 components, or to mixtures containing a small number of target analytes. This is supported by statistical overlap theory, in which the extent of peak overlap in gas chromatograms of complex mixtures can be predicted.

Other specialized high-speed GC alternatives include the use of multicapillary columns, packed columns, and flash-GC. Multicapillary columns are made by combining some 900 capillaries, each with a  $d_c$  of 40  $\mu\text{m}$  into a bundle of about 1 m length. These multicapillary columns allow the use of high flow rates and are characterized by a high sample capacity. The 1980s saw some promising work in the area of fast-GC using micropacked columns; for example, the separation of a mixture of alkanes was performed using a short 320  $\mu\text{m}$  i.d. column, packed with 10  $\mu\text{m}$  particles. Flash-GC devices are based

upon resistively heated metal capillary columns, allowing extremely high heating rates and very fast cooling times. These latter approaches cannot be considered mainstream technologies, and no further comment shall be made in the present article.

## Routes toward Fast Gas Chromatography

The basic Golay equation (eqn [5]) describes height equivalent to one theoretical plate ( $H$ ) in terms of the average linear carrier gas velocity for a capillary column:

$$H = \frac{B}{\bar{u}} + C_G \bar{u} + C_L \bar{u} \quad [5]$$

where  $B$  represents the longitudinal gaseous diffusion;  $C_G$  and  $C_L$  are the resistance to mass transfer in the gas phase and liquid phase, respectively. The hyperbolic van Deemter plot is obtained by plotting  $H$  against  $\bar{u}$  from eqn [5]. This hyperbola has a minimum at

$$H_{\min} = 2\sqrt{B(C_G + C_L)} \quad [6]$$

The Golay equation can be expanded (see eqn [7]) to outline some fundamental relationships that allow the analyst to set guidelines for developing a high-resolution fast-GC approach; some of these are discussed below:

$$H = \frac{2D_G}{u} + \frac{(1 + 6k + 11k^2)r^2}{24(1 + k)^2 D_G} u + \frac{2kd_f^2}{3(1 + k)^2 D_L} u \quad [7]$$

## New Method Development

**Column dimensions** Smaller-diameter columns often provide a direct route to speeding up analyses. In fact, the separation of 15 peaks in  $\sim 2$  s using a 1.2 m  $\times$  34.5  $\mu\text{m}$  capillary column was demonstrated as early as 1962 (see **Figure 2**). By reducing the column diameter  $d_c$ , a higher efficiency per unit length ( $L$ ) is achieved. This can be demonstrated by examination of eqn [7], where it is clear that two different situations can occur. For large  $\beta$  (phase ratio) values the  $C_L$  term can be neglected, for small  $\beta$  values the  $C_L$  term determines efficiency.

Where resistance to mass transfer in the gas phase controls  $H$ , the  $C_L$  term may be ignored; thus,

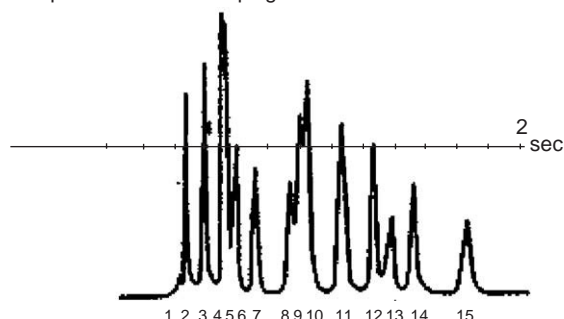
$$H_{\min} = 2(BC_G)^{1/2} = r\sqrt{\frac{1 + 6k + 11k^2}{3(1 + k)^2}} \quad [8]$$

For high  $k$  values it has been shown experimentally that  $H_{\min}$  approaches  $d_c$ . The minimum  $L$  required to



$C_5$ ,  $C_6$ , and  $C_7$  Paraffins

Column length 120 cm  
 Column diameter 34.5  $\mu$   
 Stationary phase Squalane  
 Carrier gas Hydrogen  
 Inlet pressure 200 psig



- |                       |                          |
|-----------------------|--------------------------|
| 1 2-Methylbutane      | 9 2,4-Dimethylpentane    |
| 2 <i>n</i> -Pentane   | 10 2,2,3-Trimethylbutane |
| 3 2,2-Dimethylbutane  | 11 3,3-Dimethylpentane   |
| 4 2,3-Dimethylbutane  | 12 2-Methylhexane,       |
| 5 2-Methylpentane     | 2,3-dimethylpentane      |
| 6 3-Methylpentane     | 13 3-Methylhexane        |
| 7 <i>n</i> -Hexane    | 14 3-Ethylpentane        |
| 8 2,2-Dimethylpentane | 15 <i>n</i> -Heptane     |

**Figure 2** A fast-GC chromatogram of a mixture of  $C_5$ ,  $C_6$ , and  $C_7$  paraffins from the pioneering work of Desty in the early 1960s using a  $1.2\text{ m} \times 34.5\text{ }\mu\text{m}$  capillary column. (Reprinted with permission from Desty DH, Goldup A, and Swanton WT (1962) In: Brenner N, Callen JE, and Weiss MD (eds.) *Gas Chromatography*, pp. 105–135. New York: Academic Press.)

produce 100 000 theoretical plates has been calculated for the most common commercially available capillary column dimensions. These respective column lengths are reported in Table 2, where it is apparent that equivalent plate numbers are achieved using significantly shorter smaller-diameter columns. Also, as  $d_c$  is reduced, the optimal average linear carrier gas velocity  $\bar{u}$  is also higher. Both of these factors lead to faster GC analysis (see eqn [4]).

Whilst the advantages of smaller-diameter columns are apparent, typically there has been limited implementation of these columns for fast-GC. This can be attributed to a number of factors, such as the need for high column head pressure, which were historically seen as distinct limitations; however, most of these can be overcome by the use of modern instrumentation. Narrow bore capillary columns also have a low sample loading capacity, so small injection sizes (of the order of a few nanograms) are required.

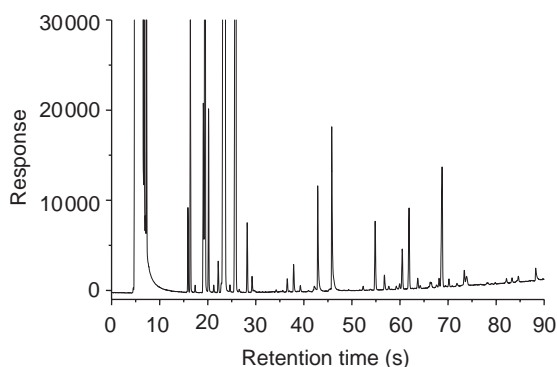
Smaller-diameter columns generally contain thin film stationary phase coatings, and will require care in manufacture to reduce column activity. Recent advancements in capillary column technology have allowed more routine application of smaller-diameter

**Table 2** Calculated number of plates per meter for a range of commonly available capillary column dimensions, and the minimum length required to produce 100 000 theoretical plates

$d_c$ ( $\mu\text{m}$ )	$h_{\text{min}}^a$	$N/m$	$L$ for $N = 100\,000$
50 <sup>b</sup>	0.05	20 000	5 m
100	0.10	10 000	10 m
250	0.25	4 000	25 m
320	0.32	3 125	32 m
530	0.53	1 887	53 m

<sup>a</sup>These calculations are true for columns with large  $\beta$  values.

<sup>b</sup>50  $\mu\text{m}$  i.d. capillary columns are not commonly available at present.



**Figure 3** Fast-GC analysis of lime essential oil using a  $5\text{ m} \times 50\text{ }\mu\text{m}$  capillary column with  $0.05\text{ }\mu\text{m}$   $d_f$  with hydrogen as the carrier gas. One microliter of a 1% essential oil in *n*-hexane solution was injected using a split ratio of 750:1. The column head pressure was 880 kPa and the GC oven was temperature programmed from 50 to  $150^\circ\text{C}$  at  $80^\circ\text{C min}^{-1}$ , to  $200^\circ\text{C}$  at  $70^\circ\text{C min}^{-1}$ , and finally to  $250^\circ\text{C}$  at  $55^\circ\text{C min}^{-1}$ .

columns. Columns of 100  $\mu\text{m}$  i.d. with various stationary phase coatings are common in all column manufacturer catalogs. With ever-improving instrument capabilities, it is expected that demand for even smaller-diameter columns will rise. The potential of ultra-narrow bore columns for high-resolution fast-GC analysis is illustrated in Figure 3 by the analysis of lime essential oil.

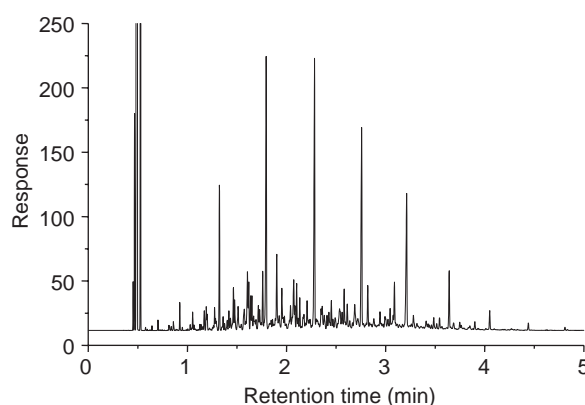
**Carrier gas selection and carrier gas velocity** The correct choice of carrier gas has a dramatic effect on analysis speed. For columns with large  $\beta$ ,  $h_{\text{min}}$  is essentially the same for nitrogen, helium, and hydrogen. However, hydrogen is the best choice because its  $u_{\text{opt}}$  is higher than that of the other gases. Optimum carrier gas velocity is proportional to the diffusivity in the gas phase, and since  $D_G$  is approximately inversely proportional to the square root of the molecular weight of the carrier gas,  $u_{\text{opt}}(\text{H}_2) = 1.4 \times u_{\text{opt}}(\text{He})$ , and  $= 3.7 \times u_{\text{opt}}(\text{N}_2)$ . The use of hydrogen carrier gas reduces analysis times by a factor of 1.4 and 3.7 compared to helium and nitrogen, respectively.

The carrier gas linear velocity through the column is critical in determining how quickly an analysis will proceed, and also how efficiently the column will resolve the peaks. The choice of carrier gas linear velocity will often represent a compromise between speed and efficiency. Selection of a low  $u$  will deliver a relatively poor result and take a long time to do so; for this reason low velocities should be avoided. Operating at high velocities produces a chromatogram more rapidly but sacrifices some column efficiency. The optimum practical linear carrier gas velocity (OPGV) depends upon the carrier gas type, the column characteristics, the temperature, and the sample (and the column pressure drop – see below). Typical OPGVs are of the order of  $20\text{--}60\text{ cm s}^{-1}$ .

The optimum linear velocity for a capillary column depends on the pressure in the column because  $u_{\text{opt}}$  is proportional to the average diffusion coefficient, which varies inversely with pressure. Operation of a short wide-bore column at vacuum outlet conditions results in a significantly faster analysis than would occur if the same column was used under atmospheric outlet pressures. Mass spectrometry (MS) has made vacuum GC very easy to implement, since the mass spectrometer provides both detection and a source of vacuum. Vacuum GC can be achieved practically by incorporating a restriction at the inlet end of a wide-bore capillary column, and interfacing the terminal end of the column directly into the MS. The function of the restriction is to deliver an optimal helium flow for the mass spectrometer, and it can be as simple as a short section of  $20\text{ }\mu\text{m}$  i.d. capillary (or a longer section of  $100\text{--}150\text{ }\mu\text{m}$  i.d. capillary). An optimal carrier gas velocity of  $\sim 90\text{--}100\text{ cm s}^{-1}$  can be expected for a  $10\text{ m} \times 50\text{ }\mu\text{m}$  column with a restriction at the inlet, and a speed gain of a factor of 3–5 times can easily be obtained.

**Temperature programming requirements for fast-GC** An optimal temperature program rate for fast-GC will deliver the best separation in the shortest possible time. Method development and optimization can be achieved using a chemometric approach such as central composite design. This type of approach is suitable for the development of a method for the analysis of a sample of limited complexity, but optimization procedures for temperature programs become more complex, more time consuming, and may offer limited success for samples containing a greater number of components (e.g.,  $>20$ ).

A recent study determined that a generally applicable optimum temperature program rate offered a suitable compromise between peak capacity, separation power, and analysis time. This concept uses  $t_{\text{M}}$  as a fundamental time unit, and showed that a temperature

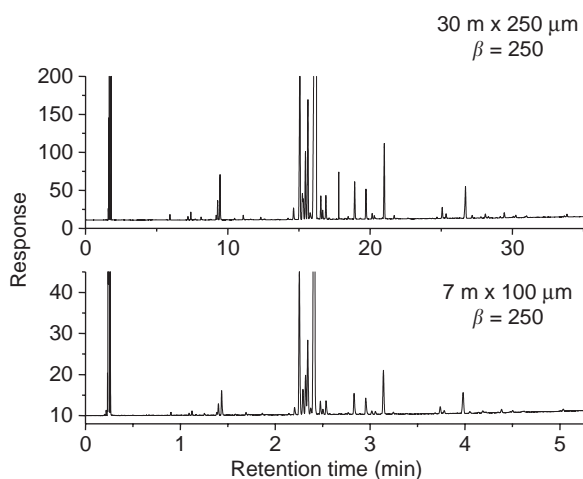


**Figure 4** Fast-GC chromatogram of kerosene using a method developed using the  $10^\circ\text{C}/t_{\text{M}}$  criterion.

program rate of  $10^\circ\text{C}/t_{\text{M}}$  was generally applicable. A typical fast-GC chromatogram of kerosene using a  $10^\circ\text{C}/t_{\text{M}}$  criterion for new method development is given in **Figure 4**. Ramp rates significantly faster than  $10^\circ\text{C}/t_{\text{M}}$  will not allow for sufficient partitioning time in the stationary phase for the effective use of the column. This optimization approach arose from the concept of method translation described next.

**Method translation** Although there will often be a requirement to create entirely new fast-GC methods, it will also be desirable in many circumstances to translate optimized methods for conventional columns to narrow bore columns. The aim of method translation is to vary components (columns, carrier gases, detectors, etc.) and parameters (pressures, temperature programs, etc.) of a method in a way that maintains the peak elution pattern. In other words, method translation yields a scaled version of the original chromatogram. A common mistake in attempting to gain analysis speed is to only apply a more rapid temperature program. Since relative retention (and elution order) of solutes depends on temperature, the application of a more rapid temperature program can cause peaks to shift relative to one another, and valuable well-known fingerprint information can be lost. However, changes in heating rates, and durations of temperature plateaus, column dimensions, carrier gas type, and pneumatic conditions are translatable changes. Changes in stationary phase type and phase ratio are not translatable. A fast-GC analysis using a smaller diameter column is given in **Figure 5**, where the slower GC analysis using a more conventional diameter column is overlaid to illustrate the effectiveness of method translation in providing a scaled version of the original chromatogram. Method translation software is available at no cost (GC Method Translation Freeware).





**Figure 5** Illustration of the method translation technique for the analysis of peppermint essential oil. Top – GC analysis using a 30 m × 250 μm column. Bottom – fast-GC analysis using a 7 m × 100 μm column of the same stationary phase coating and phase ratio ( $\beta = 250$ ).

Agilent Technologies, Wilmington, DE, 1998; <http://www.agilent.com>).

## Instrument Requirements for High-Speed GC

### Carrier Gas

The important role that the carrier gas linear velocity plays in determining column efficiency has been discussed above. This is distinctly different to the column flow rate, which is important in determining the suitability of the column with the GC system inlets and detectors. Peak distortion can result if the column flow rate is too low or too high. Long, smaller-diameter columns require high column head-pressures to provide an appropriate carrier gas flow rate. Conversely, short columns call for a low, but very precise head pressure, which can cause difficulties in maintaining stable system operation. The low flow rates in capillary GC are suitable for use with mass spectrometric detection, which operates most effectively up to a maximum rate (typically up to a few milliliters per minute), after which the detector performance begins to deteriorate. Detectors such as electron capture detector (ECD) or flame ionization detector (FID) will require a make-up flow for proper functioning. Modern instrumentation design generally addresses the typical problems; for example, a new chromatograph can be equipped with electronic pressure control units operating up to 10–12 bar, providing sufficient pressure to efficiently operate a 10 m × 50 μm or 20 m × 100 μm capillary column.

Whilst hydrogen is the most appropriate carrier gas choice in most cases for fast-GC, safety concerns are a deterrent to some laboratories. However, modern chromatographs incorporate safety features to reduce the risk of H<sub>2</sub> leaks and the risk of explosion is inherently very low.

### Sample Introduction

The general requirements for sample introduction in fast-GC include the accurate delivery of a small injection size, typically using a fast injection device. The effectiveness of the sample introduction system will be measured by its ability to deliver a sufficiently narrow injection band. In 1966, Sternberg described how column performance is degraded by the design of the injection port, dead volume, and surface activity in the flow lines, detector cell design, and electronic time constants. These effects are summarized in eqn [9], where the subscript 'o' denotes all contributions other than the injector and the column:

$$\sigma^2 = \sigma_{\text{col}}^2 + \sigma_{\text{inj}}^2 + \sigma_o^2 \quad [9]$$

Since on-column peak variance is generally low in fast-GC, the contribution of the injector to the total peak variance can become significant, so the injector must be able to deliver a suitably narrow injection bandwidth. Split injection is the most common method of sample introduction for fast-GC. This technique is the only sample introduction method that allows injection at any column temperature. This can be contrasted with on-column injection that requires the column to be below the solvent boiling point at the time of sample introduction, and splitless injection that usually requires cooling to achieve cold trapping or solvent effects. Using very fast split flow rates can produce very sharp initial bands; split flows of 21 min<sup>-1</sup> have been shown to provide injection bandwidths in the millisecond range. However, since the typical flow rate through narrow bore columns is low, it results in a very high split ratio, which can cause irregular sample transfer into the column, and irreproducibility with samples containing analytes at widely differing concentration. The use of a very high split ratio also increases the limit of detection. The requirement for very high split ratio can be relaxed by using temperature programmed GC because thermal focusing at the initial oven temperature may refocus the input band.

In circumstances where injection cannot produce sharp enough initial bands, they will need to be focused following the injection and immediately prior to the commencement of the analysis. This can be achieved by the use of specialized cold-traps, using liquid nitrogen or carbon dioxide to cause cryogenic

focusing of the initial band. An on-column thermal desorption modulator, such as that developed by Phillips and co-workers in the late 1980s, also offers a technically simple and reliable method for sample introduction into smaller internal diameter columns. Another contemporary approach for sample introduction in fast-GC includes the use of fast rotating microinjection valves (providing injection times  $<100$  ms), which perform very fast injections to minimize injection band broadening. A method described by Sacks employed cryotrapping of injected sample, which was then very rapidly thermally desorbed at  $1000^{\circ}\text{C s}^{-1}$  and back-flushed into the capillary column. An injection bandwidth of a few milliseconds was reported.

### Oven

The main requirement of the oven for fast temperature programmed GC will be its ability to provide rapid linear temperature gradients. The earlier section on temperature programming requirements for fast-GC described the concept of using  $t_M$  as time unit. Here, we shall put this concept into practice to illustrate the stresses that this imparts on the instrument. As a first example, consider a  $10\text{ m} \times 100\text{ }\mu\text{m}$  column provided with 200 kPa head pressure ( $\text{H}_2$ ) at  $60^{\circ}\text{C}$ . This column will have an average linear carrier gas velocity of  $\sim 55\text{ cm s}^{-1}$ , giving  $t_M$  of  $\sim 0.303$  min. Most instruments should easily deliver a temperature program of  $33^{\circ}\text{C min}^{-1}$  (i.e.,  $10^{\circ}\text{C}/t_M$ ; see above). However, if a 2 m column of the same internal diameter was used,  $t_M$  (for  $55\text{ cm s}^{-1}$ ) would be  $\sim 0.0606$  min. Whilst temperature programming at  $33^{\circ}\text{C min}^{-1}$  would still provide a faster chromatogram than the previous experiment, the  $10^{\circ}\text{C}/t_M$  approach would suggest that a temperature program rate of  $165^{\circ}\text{C min}^{-1}$  would still render a chromatogram with sufficient peak capacity and separation power. Clearly, this chromatogram would be generated in a greatly reduced time, but the inability of most instruments to achieve such a high temperature program rate would limit this application.

Another important consideration of the GC oven in fast-GC analysis is the cool-down period. Indeed, the total analysis time should take into consideration the GC runtime, the oven cool-down period, and also the sample preparation step. Automated sample preparation, often in large batches, is becoming increasingly popular, reducing this part of the total analysis time. Thus, maximum time efficiency will require the shortest possible oven cool-down and equilibration time. Modern GC equipment is constructed of lightweight material to assist in heat transfer from the oven walls to the oven air, which is

expelled by an efficient ventilation system. Some chromatographs are fitted with a cryogenic cooling system to rapidly decrease the oven temperature. The use of a device commonly referred to as a 'pillow', which can be inserted into the oven cavity to reduce the size of the oven, is also effective for fast-GC. Ultimately, a series of 2 min analyses with 5 min cool-down periods will always take less time than a series of 30 min analyses with 5 min cool-down periods, so the cool-down period should generally not be considered the rate determining step.

### Considerations for Detection

Peaks generated by fast GC are no different to measure than other chromatographic peaks, except that they are narrower, and a high detector sampling frequency will be required to measure very narrow peaks accurately. Data points from 10 to 20 per  $\omega_b$  is generally considered a sufficient number; however, it has been suggested that this number should be far greater to reduce measurement uncertainty for asymmetric peaks.

FID detection can be routinely performed using data acquisition rates in excess of 100 Hz, and other detectors such as the  $\mu$ -ECD offer suitably rapid data acquisition for fast-GC.

### Mass Spectrometry

Scan speed, sensitivity, and useful dynamic range affect the compatibility of a mass spectrometer with a high-resolution chromatograph. Improvements in GC, particularly in the direction of high-speed GC have led to a strain in the relationship between residence time of the sample in the ion source and mass spectral scan time. Quadrupole and ion-trap mass spectrometers are most commonly used for GC-MS. The physical laws governing mass selection processes limit the ability of these detectors to provide suitable detection in fast-GC. For example: a quadrupole MS passes ions of various  $m/z$  (one  $m/z$  at a time) to the detector by simultaneously ramping direct current (DC) and radio frequency (RF) voltages carried by four quadrupole rods. To ensure maximum transmission of the selected ions, the RF and DC voltages should not change during the transit time of an ion through the quadrupole. Contemporary MS instruments are capable of mass scanning rates of around  $4000\text{--}7000\text{ u s}^{-1}$ . Thus, for a full scan mass range of  $50\text{--}350\text{ u}$ , the maximum spectral acquisition rate would be expected to be less than 10–20 spectra per second (including an interscan period).

Time-of-flight (TOF) MS does not suffer the same speed limitations of the previously mentioned instrument types, because the spectral acquisition rate is essentially determined by the transit time of the

heaviest ion in the mass spectrum. The major advantage TOF-MS has is that a complete spectrum is produced serially at the detector for each pulse of ions from the source. Considering an ion of 800 u and 2 keV acceleration potential will take  $\sim 90 \mu\text{s}$  to travel from the source to the detector in a 2 m flight tube, it is possible to acquire a full spectrum from 0 to 800 u every  $90 \mu\text{s}$ . This equates to a spectral acquisition rate of  $\sim 11\,000$  spectra per second. Fast-GC-TOF-MS instruments will become more widespread in the coming years. Whilst fundamental principles of mass analysis govern the scan speed of quadrupole and ion trap instruments, the usable spectral acquisition rate of TOF analyzers are only determined by the method of data collection and storage.

See also: **Gas Chromatography**: Principles; Instrumentation; Mass Spectrometry. **Mass Spectrometry**: Time-of-Flight.

## Further Reading

Cramers CA and Leclercq PA (1988) Consideration on speed of separation, detection, and identification limits in

capillary GC and GC/MS. *CRC Critical Reviews in Analytical Chemistry* 20(2): 117–147.

van Es A (1992) *High Speed Narrow Bore Capillary Gas Chromatography*. Heidelberg: Hüthig.

Grob K (2001) *Split and Splitless Injection for Quantitative Gas Chromatography*, 4th edn. Weinheim: Wiley.

Holland JF, Enke CG, Allison J, *et al.* (1983) Mass spectrometry on the chromatographic timescale: Realistic expectations. *Analytical Chemistry* 55(9): 997A–1012A.

Klee MS and Blumberg LM (2002) Theoretical and practical aspects of fast gas chromatography and method translation. *Journal of Chromatographic Science* 40: 234–247.

Korytár P, Janssen H-G, Matisová E, and Brinkman UATH (2002) Practical fast gas chromatography: Methods, instrumentation, and applications. *Trends in Analytical Chemistry* 21: 558–572.

Sacks R, Patterson DG, and Dimandja J-MD (2001) *High Speed Gas Chromatography*. Chichester: Wiley.

de Zeeuw Y, Peene J, Janssen H-G, and Lou X (2000) A simple way to speed up separations by GC-MS using short 0.53 mm columns and vacuum outlet conditions. *Journal of High Resolution Chromatography* 23(12): 677–680.

## Instrumentation

C F Poole, Wayne State University, Detroit, MI, USA

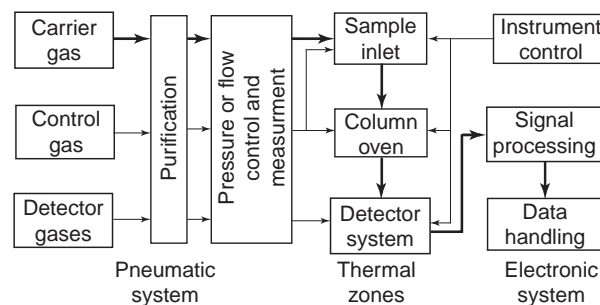
© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The principal function of the gas chromatograph is to provide those conditions required by the column for achieving a separation without adversely affecting its performance. Operation of the column requires a regulated flow of carrier gas; an inlet system to vaporize and mix the sample with the carrier gas; a thermostatted oven to optimize the temperature for the separation; an online detector to monitor the separation; and associated electronic components to control and monitor instrument conditions, and to record, manipulate, and format the chromatographic data. Individual instruments differ mainly in their control functions (whether software-based or mechanical), ease of portability (designed for laboratory, field, or process applications), level of automation, and flexibility (number of supported options). The primary support functions of a typical gas chromatograph can be broken down into pneumatic, thermal, and electronic system components as illustrated in Figure 1.

## Pneumatic Systems

Gas supplies are required for the carrier gas, and depending on the instrument configuration, perhaps also for the detector, for operating pneumatic controls such as switching valves, and for providing automatic cool down by opening the oven door. Gases are normally provided from pressurized cylinders, although compact generators capable of providing air or nitrogen filtered from the atmosphere



**Figure 1** Schematic diagram of the principal components of a gas chromatograph. Bold lines indicate the path taken by sample and carrier gas resulting in the production of the chromatogram. The thin lines represent support and control functions.

heaviest ion in the mass spectrum. The major advantage TOF-MS has is that a complete spectrum is produced serially at the detector for each pulse of ions from the source. Considering an ion of 800 u and 2 keV acceleration potential will take  $\sim 90 \mu\text{s}$  to travel from the source to the detector in a 2 m flight tube, it is possible to acquire a full spectrum from 0 to 800 u every  $90 \mu\text{s}$ . This equates to a spectral acquisition rate of  $\sim 11\,000$  spectra per second. Fast-GC-TOF-MS instruments will become more widespread in the coming years. Whilst fundamental principles of mass analysis govern the scan speed of quadrupole and ion trap instruments, the usable spectral acquisition rate of TOF analyzers are only determined by the method of data collection and storage.

See also: **Gas Chromatography**: Principles; Instrumentation; Mass Spectrometry. **Mass Spectrometry**: Time-of-Flight.

## Further Reading

Cramers CA and Leclercq PA (1988) Consideration on speed of separation, detection, and identification limits in

capillary GC and GC/MS. *CRC Critical Reviews in Analytical Chemistry* 20(2): 117–147.

van Es A (1992) *High Speed Narrow Bore Capillary Gas Chromatography*. Heidelberg: Hüthig.

Grob K (2001) *Split and Splitless Injection for Quantitative Gas Chromatography*, 4th edn. Weinheim: Wiley.

Holland JF, Enke CG, Allison J, *et al.* (1983) Mass spectrometry on the chromatographic timescale: Realistic expectations. *Analytical Chemistry* 55(9): 997A–1012A.

Klee MS and Blumberg LM (2002) Theoretical and practical aspects of fast gas chromatography and method translation. *Journal of Chromatographic Science* 40: 234–247.

Korytár P, Janssen H-G, Matisová E, and Brinkman UATH (2002) Practical fast gas chromatography: Methods, instrumentation, and applications. *Trends in Analytical Chemistry* 21: 558–572.

Sacks R, Patterson DG, and Dimandja J-MD (2001) *High Speed Gas Chromatography*. Chichester: Wiley.

de Zeeuw Y, Peene J, Janssen H-G, and Lou X (2000) A simple way to speed up separations by GC-MS using short 0.53 mm columns and vacuum outlet conditions. *Journal of High Resolution Chromatography* 23(12): 677–680.

## Instrumentation

C F Poole, Wayne State University, Detroit, MI, USA

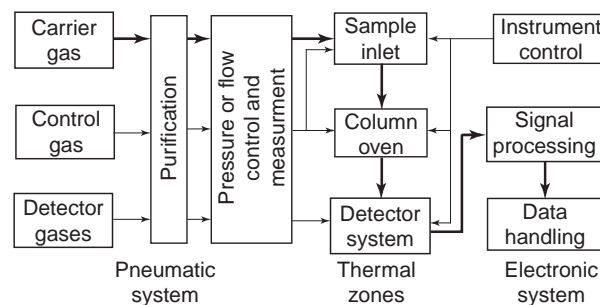
© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The principal function of the gas chromatograph is to provide those conditions required by the column for achieving a separation without adversely affecting its performance. Operation of the column requires a regulated flow of carrier gas; an inlet system to vaporize and mix the sample with the carrier gas; a thermostatted oven to optimize the temperature for the separation; an online detector to monitor the separation; and associated electronic components to control and monitor instrument conditions, and to record, manipulate, and format the chromatographic data. Individual instruments differ mainly in their control functions (whether software-based or mechanical), ease of portability (designed for laboratory, field, or process applications), level of automation, and flexibility (number of supported options). The primary support functions of a typical gas chromatograph can be broken down into pneumatic, thermal, and electronic system components as illustrated in **Figure 1**.

## Pneumatic Systems

Gas supplies are required for the carrier gas, and depending on the instrument configuration, perhaps also for the detector, for operating pneumatic controls such as switching valves, and for providing automatic cool down by opening the oven door. Gases are normally provided from pressurized cylinders, although compact generators capable of providing air or nitrogen filtered from the atmosphere



**Figure 1** Schematic diagram of the principal components of a gas chromatograph. Bold lines indicate the path taken by sample and carrier gas resulting in the production of the chromatogram. The thin lines represent support and control functions.



and hydrogen electrolytically from water are becoming increasingly popular. Each gas cylinder is fitted with a two-stage regulator for coarse pressure and flow control. To minimize contamination of the chromatograph high-purity gases are used together with additional chemical and/or catalytic gas purifying devices. The carrier gas flow is directed through a particle filter and charcoal and molecular sieve traps to remove volatile organic compounds and moisture, and then through an additional chemical trap to remove oxygen. Oxygen causes degradation of some stationary phases, shortens filament lifetime for thermal conductivity detectors, and yields unstable baselines with the electron-capture detector. Water shares some of the same properties as well as being a strong deactivating agent that can cause poor reproducibility of retention times in gas-solid chromatography. Support gases for detectors and valve operation, etc., usually require less intensive purification depending on how they are used.

The carrier gas and support gases then enter the pneumatic section of the gas chromatograph, in which pressure regulators, flow controllers, and, perhaps, additional gas purifying traps and particle filters are housed in a thermostatted box. Fuel gases required by flame-based detectors need only be coarsely controlled by, for example, a calibrated restrictor or needle valve. In most early instruments carrier gas flow control was widely used for packed columns while pressure control was more common for open tubular columns. Since the early 1990s, electronic pneumatic control devices have become increasingly popular and are common on all modern keyboard controlled instruments. Electronic pressure control uses a solid-state pressure sensor and an electronically controlled proportional valve to achieve precise closed loop control of the column inlet pressure. The use of sensors and valves allows operation at constant mass flow independent of the column flow resistance, constant pressure, and pressure program modes under full software control. Typical electronic flow controllers with a range of  $1\text{--}500\text{ ml min}^{-1}$  are capable of an accuracy of  $\pm 1\text{--}2\%$  full scale and a repeatability of the set point of  $\pm 0.2\%$  or better.

## Thermal Zones

The column oven is essentially an insulated box of sufficient size to allow comfortable installation of the longest columns and any accessory equipment normally used. The oven is heated by electrical heating elements arranged around a powerful circulatory fan. Microprocessors, sensors, and proportional heating

networks are used to maintain a stable isothermal temperature and to control the initial temperature lag, the linearity of the program rate, and the final temperature overshoot in temperature programmed operation. A temperature stability of  $\pm 0.2^\circ\text{C}$  in time and  $\pm 1.0^\circ\text{C}$  in space are typical specifications for modern instruments. Poor column temperature stability is a source of peak distortion or splitting and retention time variation with open tubular columns.

Separations in gas chromatography are carried out within the temperature limits from about  $-100^\circ\text{C}$  to  $450^\circ\text{C}$ . Purpose-built instruments are usually required for high-temperature operation between  $375^\circ\text{C}$  and  $450^\circ\text{C}$ . Subambient temperature operation using the boil over vapors from liquid nitrogen or carbon dioxide for cooling is available as an option for standard instruments. The oven temperature is adjusted using an electrically controlled solenoid valve to pass coolant into the oven where it is mixed with air and then circulated at high velocity.

Typical linear temperature program rates are  $0.1\text{--}50^\circ\text{C min}^{-1}$ , selectable in incremental steps. Rapid cool down of the column oven is achieved by automated opening of the oven door or flaps to allow heat dissipation. Fast gas chromatography requires heating and cooling rates much faster than those specified for standard operation. Resistive heating using a tube within a tube or a wire adjacent to the column configuration are the favored options.

Other thermal zones, which should be thermostatted separately from the column oven, include the injector and detector modules. These are generally insulated metal blocks fitted with cartridge heaters and controlled by sensors located in a feedback loop with the power supply. Detector blocks are usually maintained at a constant temperature while requirements for injectors may be different, and may include provision for temperature programmed operation.

## Sample Handling Devices

Gas chromatography is used to separate gases, vapors of volatile compounds supported in a gas matrix, liquids, and solids. Solid samples are usually dissolved in a suitable solvent and handled as for liquid samples. Insoluble solids can be analyzed indirectly by separation of the decomposition products generated by controlled pyrolysis. Compounds in the vapor phase are often separated after concentration on solid sorbents or cryotrap followed by rapid heating to release the organic vapors into the column. Thermal desorption interfaces are widely used for the analysis of trace organic volatiles in air or generated by headspace and gas purging devices.

## Microsyringes

Microsyringes are commonly used for introducing liquid samples into a gas chromatograph. A typical microsyringe consists of a calibrated glass barrel with a close fitting metal plunger, which is used to dispense a chosen volume of sample by displacement through the syringe needle. Gas-tight syringes, some with a valve mechanism to close the passageway to the needle, are available for injecting gases and vapors.

Although microsyringes are considered indispensable for injection of liquids, they do present certain difficulties. The accuracy of injection depends on the rate of sample introduction, syringe dead volume, heating of the syringe needle by the injector, and sample handling techniques. Packed columns accommodate relatively large sample volumes (e.g., 1–5  $\mu\text{l}$ ) and are forgiving of poor injection techniques. The injection of comparatively small volumes, typically 0.1–2.0  $\mu\text{l}$ , is more common for open tubular columns. In this volume range needle dead volumes and backflushing of sample past the plunger can represent significant contributions to poor injection precision. Discrimination of sample components is possible with hot vaporizing injectors with open tubular columns. The sample leaves the syringe and enters the vaporizer as a stream of droplets, formed by the movement of the plunger and by evaporation of the remaining sample from the syringe needle. It is at this evaporation stage that discrimination is most likely to occur; the solvent and more volatile sample components distill from the syringe needle at a greater rate than the less volatile components.

## Gas Sampling Valves

Rotary valves with six, eight, or more connection ports, operated manually or automated through electronic or pneumatic controllers, are commonly used for introducing gases and volatile compounds into all column types as well as for column switching applications in multidimensional gas chromatography. Rotary valves are also used for large volume sample introduction in on-column injection and as an interface for coupling liquid chromatography to gas chromatography.

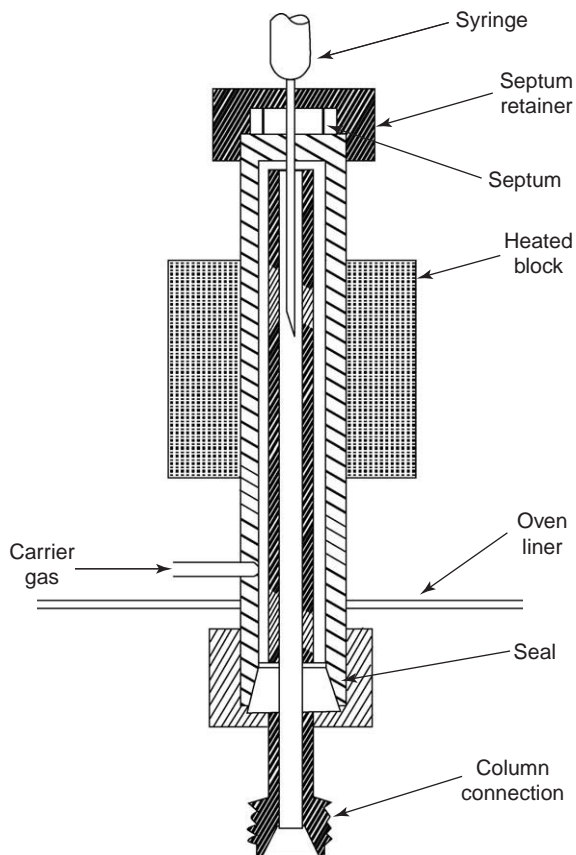
The gas sample to be injected is isolated in a sample loop of defined volume at a known temperature and pressure (usually ambient). Subsequent rotation of the valve body causes the carrier gas to sweep the entire contents of the sample loop onto the column. External sample loops with defined volumes in the range 0.5  $\mu\text{l}$  to 5 ml are commonly used. Valves with internal sample loops are available for handling small sample volumes of 0.2–5  $\mu\text{l}$ . Multiport valves with two or more sample loops can be used for

simultaneous sampling of several sources or for continuous sampling of a single source by storing individual samples in different loops for sequential analysis at a later time. Gas sampling valves are usually mounted in the column oven, in a separately heated adjacent oven, or external to the column oven and connected to the column by a short length of capillary tubing. For the most accurate work it is recommended that the injection valve is thermostated in its own oven.

## Liquid Sample Inlets

### Packed Column

Liquid samples are introduced by a microsyringe through a silicone rubber septum into a glass liner or the front portion of the column, which are heated and continuously swept by the carrier gas, **Figure 2**. When injection is made in the on-column mode, the column is pushed right up to the septum area and the column end within the injector is packed with glass wool. Whichever technique is used, the injector must



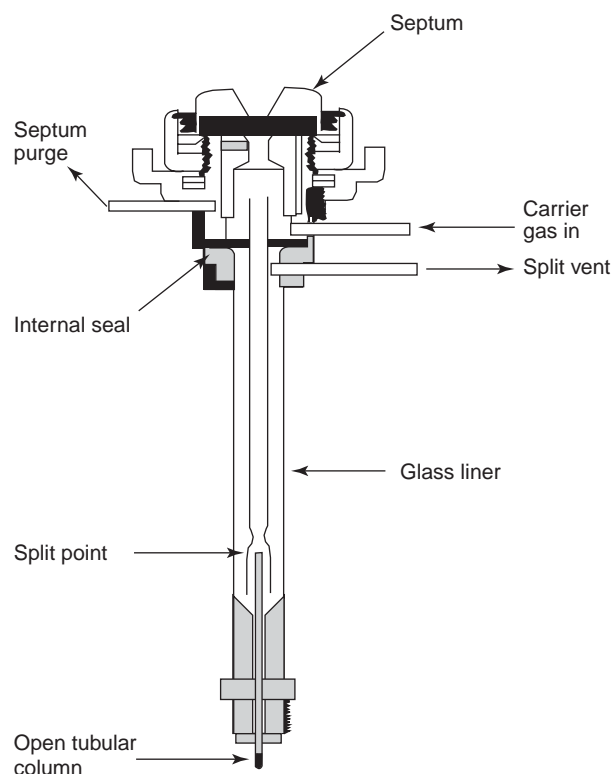
**Figure 2** A flash vaporization inlet for packed column gas chromatography. (Reproduced with permission from LCGC, vol. 15, no. 11, Nov. 1997, 1024p.)



have sufficient thermal mass to rapidly vaporize the sample. The incoming carrier gas is usually preheated by directing its flow through a section of the injector block to avoid possible condensation of the vaporized sample upon mixing with the cool carrier gas.

### Hot Split/Splitless

The classical hot split injector for open tubular columns, **Figure 3**, is an isothermal vaporizing injector, in which the evaporated sample is mixed with the carrier gas and divided between two streams of different flow, one entering the column (carrier gas flow) and the other vented to the outside (split flow). The vaporization chamber is typically a stainless steel tube lined with a removable glass liner to minimize sample contact with hot reactive metal surfaces. In the forward-pressure configuration the inlet pressure is controlled by a pressure regulator in front of the inlet and the split flow by an on/off solenoid valve and a fixed restrictor or needle valve at the split vent exit from the inlet. The backpressure configuration uses a flow controller to establish the total flow of carrier gas to the inlet and a backpressure regulator in the split vent line to control the inlet pressure.



**Figure 3** A hot split/splitless injector for open tubular column gas chromatography. (From Tipler A and Ettre LS (1997) *The Prevent System and its Application in Open-Tubular Column Gas Chromatography*. Norwalk: Perkin-Elmer Corporation.)

Liquid samples are introduced into the vaporization chamber by syringe through a septum or airlock. An auxiliary flow of gas is used to purge septum bleed products and contaminants away from the vaporization chamber. Appropriate column loads are usually achieved by injecting sample volumes of 0.2–2.0  $\mu\text{l}$  with split ratios between 1:10 and 1:1000.

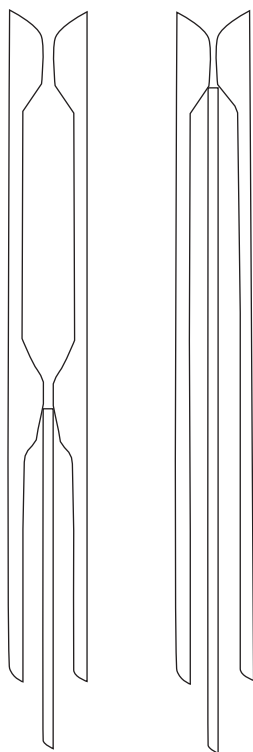
For mixtures containing sample components of unequal volatility split injection discriminates against the less volatile sample components due to selective evaporation from the syringe needle and from incomplete sample vaporization and inhomogeneous mixing of sample vapors with the carrier gas in the vaporization chamber. Sample evaporation generates an instantaneous pressure pulse and rapid change in the viscosity of the carrier gas sample mixture, altering the flow of gas between the column and split line in an irreproducible manner. Condensation of sample vapors in the cooler column inlet creates a zone of reduced pressure that sucks in further amounts of sample vapor. In most cases the sample arrives at the split point only partially evaporated as a mixture of vapor and droplets of various sizes. The sample components are unlikely to be evenly distributed between the vapor and liquid phases, the latter being split to different extents and resulting in discrimination. Discrimination is the principal cause of difficulties in quantitative analysis.

Splitless injection uses a similar injector design to split injection but is more suitable for quantitative analysis of trace components in dirty samples, such as biological and environmental extracts. Conversion of a split to a splitless injector usually requires no more than the installation of a different liner and the interruption of the split flow at the start of the injection using a solenoid valve located in the vent line. The split flow is restarted only at the end of the sampling period. Since the flow of gas through the vaporization chamber is normally the same as the carrier gas flow for the column the transport of sample vapors to the column is relatively slow. During the initial rapid evaporation of the sample there is minimal transfer of vapors into the column. Consequently, the volume of the injection liner must be large enough to hold the entire volume of vapor produced by the evaporated sample. A liner volume of  $\sim 1$  ml is sufficient to hold up to about  $\sim 0.5$ – $2$   $\mu\text{l}$  of vaporized solvent. Splitless inlets equipped with electronic pneumatic control allow high inlet flows at the start of injection, followed by a rapid reduction of pressure to a value required to provide the desired column flow rate. This minimizes the backflusing of sample vapors out of the injection liner during injection. It also allows larger sample volumes, up to 5  $\mu\text{l}$ , to be introduced with a standard liner.

Sample transfer to the column by splitless injection requires a comparatively long transfer time, from several seconds up to a few minutes, relying on cold trapping and/or solvent effects to refocus the sample at the column inlet. The sample transfer time is roughly equivalent to about twice the time required by the carrier gas to sweep out the volume of the vaporization chamber. Since the sample vapors are continually diluted with carrier gas, and some sample vapors accumulate in areas poorly swept by the carrier gas, complete sample transfer is difficult to achieve. At the end of the sample transfer period, the split flow is re-established to purge the inlet of remaining solvent vapors.

### Direct

Direct injection is used with wide bore capillary columns ( $\geq 0.32$  mm) operated at relatively high flow rates of  $5\text{--}15\text{ ml min}^{-1}$ . The injector consists of a thermostatted hot vaporization chamber with a direct connection between the liner and the column, **Figure 4**. Typically, a double gooseneck liner is used with the column fixed to the constriction at the bottom end (the seal is usually made between the polyimide outer coating of the column and the inner glass surface of the liner). Using a liner with a single

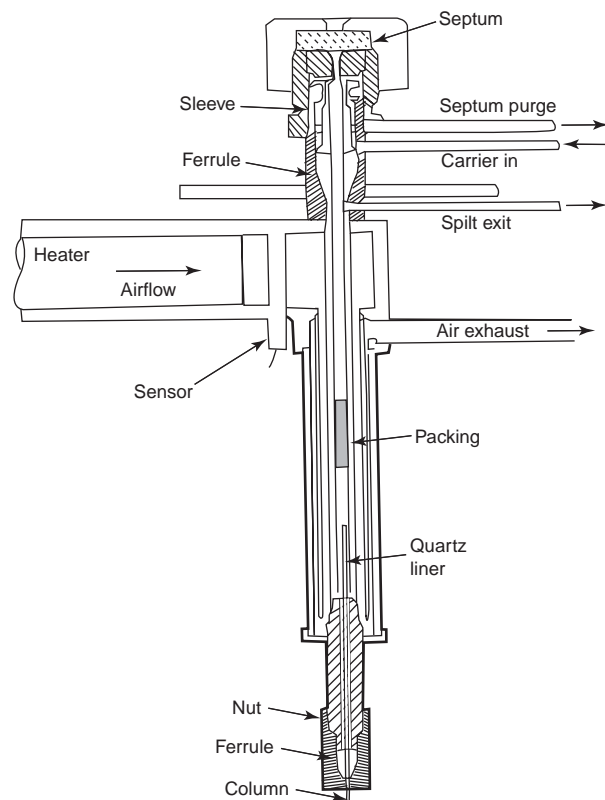


**Figure 4** Suitable liners for direct splitless (left) and on-column (right) injection onto wide bore open tubular columns.

restriction at the top end where the column seal is made allows the injector to be converted into a hot on-column injector. In this case, the restriction acts as a base for the column seal and a needle guide to direct the syringe needle into the column or a retention gap connected to the column. Injection volumes are typically in the range  $2\text{--}8\text{ }\mu\text{l}$ . A hot split/splitless injector can be converted to a direct injector by changing the liner and disabling the split vent.

### Programmed Temperature Vaporizer

The programmed temperature vaporizer (PTV) injector, **Figure 5**, can be used as a split/splitless injector for samples of normal volume or as a large volume injector. The PTV vaporization chamber is about one-tenth the size and of lower thermal mass than chambers employed for classical hot split/splitless injection. This allows the vaporization chamber to be rapidly heated by either direct or indirect resistive heating, using cartridge heaters or circulated hot air. Cooling can be performed using cold air, a Peltier element, or expanding carbon dioxide or



**Figure 5** Programmed temperature vaporizer injector. (From Hinshaw JV and Seferovic W (1986) Programmed-temperature split-splitless injection of triglycerides: comparison to cold on-column injection. *Journal of High Resolution Chromatography* 9: 69–77.)

liquid nitrogen vapors. Compared with hot split/splitless injection, the PTV injector greatly reduces discrimination of less volatile compounds, minimizes thermal degradation because of the shorter sample residence time at elevated temperatures, and is adaptable to the introduction of a wide range of sample volumes. Large volume injections are of particular interest since they allow an increase in detected peak signals, important for trace analysis. The PTV injector is also more tolerant of involatile matrix components and affords good quantitative accuracy. The PTV is a useful interface for coupling a number of sample preparation methods to gas chromatography such as high-performance liquid chromatography, supercritical fluid extraction, solid-phase extraction, and thermal desorption.

For normal split injection, the sample is introduced at a temperature below the solvent boiling point with the split exit open. Shortly after withdrawal of the syringe needle the injector is rapidly heated to a temperature high enough to ensure rapid evaporation of the least volatile sample components. Discrimination effects, which are common for hot vaporizing injectors, are virtually eliminated and the sample split ratio and split flow ratio are similar.

For cold splitless injection, the sample is introduced into the vaporizing chamber at a temperature close to the solvent boiling point with the split vent closed. Shortly after sample introduction, the injector is rapidly heated to the temperature required to transfer the sample into the column, and the vaporization chamber purged of solvent residues by opening the split vent. Similar to hot splitless injection, cold trapping and solvent effects are employed as refocusing mechanisms.

For large volume injection, the PTV injector can be operated as a cold split injector with solvent elimination or as a hot splitless injector with vapor overflow. Splitless injection with vapor discharge is not as popular as the above methods. In this case, solvent elimination is comparatively slow and the maximum volume of sample that can be introduced is only 20–30  $\mu\text{L}$ .

In cold split injection, the sample is introduced into the injector at a temperature below the solvent boiling point with the split vent open. The solvent is concurrently evaporated in the glass insert and swept through the split vent by the carrier gas. When evaporation of the solvent is complete, the split vent is closed and the injector rapidly heated to the temperature required to vaporize the sample and transfer it to the column. Large sample volumes can be introduced in several increments, or more elegantly, with a speed-controlled injector. Maximum sample volumes are  $\sim 20 \mu\text{L}$  (1.2 mm ID liner) or  $150 \mu\text{L}$  (3.4 mm ID liner)

using a single injection or  $\sim 1 \text{ mL}$  using speed-controlled injection. During solvent elimination, cold trapping and solvent effects are responsible for retention of the analytes within the liner.

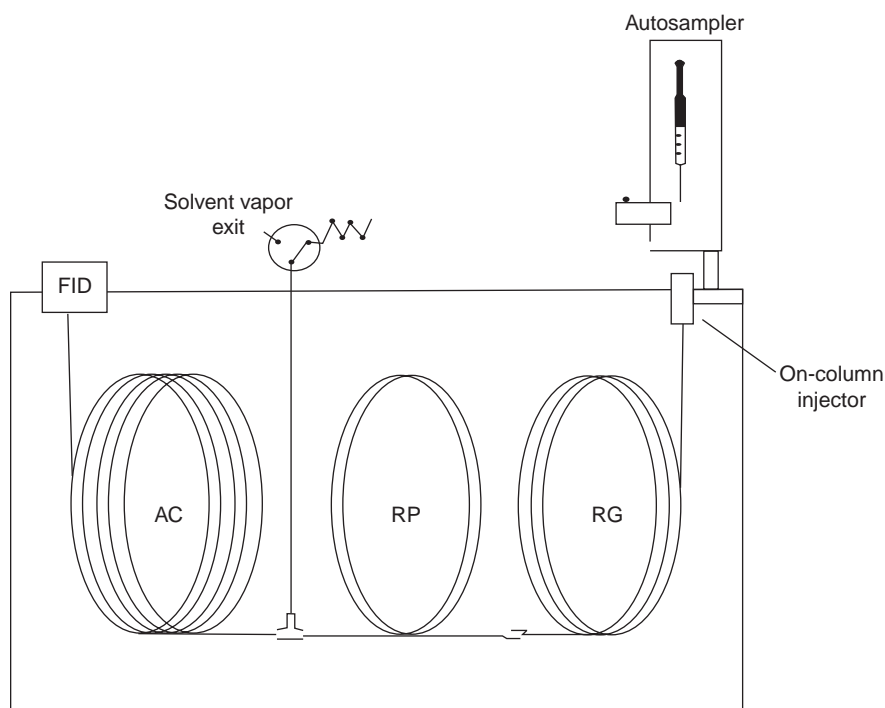
For hot splitless injection with vapor overflow, the sample is injected rapidly into the lower part of a packed liner at a temperature well above the solvent boiling point. Violent evaporation causes most of the expanding solvent vapor and volatile sample components to escape through the septum purge outlet. Solutes of low volatility are retained in a low-temperature zone created by the evaporating solvent. After solvent elimination the temperature of the retaining zone rapidly returns to the injector temperature transferring the analytes to the column. Tenax or silanized glass wool is commonly used as packing material to increase the retention power of the liner. Even so, semivolatile analytes are difficult to trap quantitatively.

### **Cold On-Column**

Cold on-column injection employs the direct introduction of the liquid sample into the oven-thermostatted column inlet or retention gap (precolumn) without prior vaporization in a heated external chamber. The sample is subsequently vaporized from the liquid film formed in the column inlet or retention gap. Discrimination is virtually eliminated and quantification of components of different volatility is facilitated. On-column injection is easily automated using a wide bore retention gap connected to the separation column and is easily adapted to the introduction of large sample volumes. On-column injection is not suitable for the introduction of samples containing significant amounts of involatile matrix components that accumulate at the column inlet increasing its retentive power and activity.

On-column injection requires special syringes with small external diameter needles (since at the point of sample release the needle must reside within the column) or a wide bore retention gap if standard size needles are used. The injector scaffold is designed to afford a mechanism to guide the syringe needle into the column, for positioning the needle at the correct height within the column, and for sealing the needle passageway, or at least restricting the flow of carrier gas out through the syringe entry port. Circulating air is used in some designs to control the temperature of the needle passageway to avoid solvent evaporation.

Sample volumes of  $\sim 0.2\text{--}1.5 \mu\text{L}$  can be injected rapidly into the column. For larger volumes of up to  $\sim 50 \mu\text{L}$  a retention gap is required to avoid peak distortion by band broadening in space. A retention



**Figure 6** A typical arrangement for large volume on-column injection with an early solvent vapor exit and retaining precolumn. AC = separation column; RP = retaining precolumn; and RG = retention gap.

gap combined with an early solvent vapor exit and optional retaining precolumn, **Figure 6**, is suitable for injections up to  $\sim 100\ \mu\text{l}$  with larger volumes introduced at a controlled speed using partial concurrent evaporation. To avoid loss of volatile analytes the solvent vapor exit must be closed just before the last drop of solvent evaporates. A number of methods are available for this critical step, such as monitoring the effluent leaving the vapor exit with a flame or by observing the restoration of the carrier gas flow at the completion of the evaporation step. Quantitative accuracy and precision ( $<1\%$  RSD for automated injectors) is excellent using internal standards to correct for differences in volume delivery.

### Retention Gaps

Retention gaps are column inlets with a reduced retention power compared with the separation column. They function as guard columns to protect the separation column from contamination by nonvolatile residues; wide bore retention gaps permit the use of autoinjectors with standard syringe needles; and retention gaps facilitate the injection of large sample volumes. Most retention gaps are simply different lengths of uncoated fused-silica tubing, deactivated by the same techniques as those used for column preparation, as required, to modify their solvent

compatibility or surface activity. It is important that the solvent wet the retention gap in order to produce a film of liquid over the column wall to facilitate solvent evaporation. Solutes migrate through the retention gap at temperatures well below those required to cause elution from the separation column. Thus, they arrive at the entrance to the separation column at a temperature too low for significant migration and accumulate there until a temperature is reached at which migration starts. Retention gaps are essential when sample volumes greater than  $\sim 5\ \mu\text{l}$  and up to several hundred microliters are injected.

### Vapor Sample Inlets

A number of important sample preparation techniques rely upon gas extraction or the analysis of samples in the gas phase. These samples usually contain low concentrations of volatile analytes and higher concentrations of water vapor in a low molecular weight gas or air. Direct sample introduction by syringe or valves is only suitable for small volumes of relatively concentrated samples. More common sample introduction methods involve analyte accumulation by sorbent or cryogenic trapping followed by vaporization in the presence of a flow of gas to transport the analytes to the column. Important

automated sample introduction methods based on sorbent or cryogenic trapping include static and dynamic headspace (purge-and-trap) and solid-phase microextraction. In addition, rapidly heated cryogenic traps provide suitable inlets for fast gas chromatography and interfaces for coupled-column gas chromatography.

### Thermal Desorption

Typical sorbents used for analyte accumulation are porous polymers (e.g., Tenax), graphitized carbon blacks, and carbon molecular sieves, and in special cases inorganic oxides (e.g., silica gel). These are packed into stainless steel or glass tubes of different dimensions. Stainless steel tubes of 89 mm  $\times$  5 mm ID and 6 mm OD (volume  $\sim$ 3 ml containing 0.2–1.0 g of sorbent) or glass tubes 89 mm  $\times$  4 mm ID and 6 mm OD (volume  $\sim$ 2 ml) are considered standard sizes for compatibility with commercial thermal desorption devices. The trapped analytes are transferred to the column by thermal desorption at the chosen temperature using resistive heating or cartridge heaters housed in a heating block around the sampling tube. The time required for complete desorption is  $\sim$ 5–15 min with a purge gas flow of 30–50 ml min<sup>-1</sup> in the opposite direction to the sample flow employed for trapping. The slow desorption kinetics and high purge gas flow rates are compatible with packed columns allowing direct transfer with cold trapping by the stationary phase providing band focusing. For open tubular columns, a two-stage desorption process is generally required. The second accumulation trap is either a small-scale sorbent trap or cryogenic trap. Small-scale sorbent traps contain less than 100 mg of sorbent and have a low thermal mass allowing rapid heating ( $>35^\circ\text{C s}^{-1}$ ) to improve the band shape of transferred analytes. Desorption from the small-scale trap occurs at a flow rate of  $\sim$ 3 ml min<sup>-1</sup> with a desorption time of 1–5 min. A split flow between the trap and the column can be used to further adjust the sample concentration to the sample capacity of the column. Fully automated thermal desorption analyzers using a sequential tube sampler are available. Thermal desorption devices are usually designed as accessories to interface to a standard gas chromatograph with minimal modification. A thermostatted short length of deactivated fused silica or fused silica-lined stainless steel capillary tube acts as a transfer line terminated in a standard hot split/splitless injector.

### Cryogenic Traps

Volatile compounds can be retained by condensation if the temperature of the trap is below the dew point

for the compound. The trap can be an empty tube, a length of coated capillary column, or a packed tube containing a nonretaining material (e.g., glass beads). A nonretaining packing enhances condensation of volatile compounds while allowing rapid thermal desorption for their transfer to the column. Retaining materials afford higher trapping efficiency but slow desorption kinetics may result in poor chromatographic performance and low recovery of less volatile compounds.

Cold traps are often based on a tube within a tube design. The inner tube is a length of metal or fused-silica capillary surrounded by a wider metal or polymer tube through which cold gas is circulated around the annular space between tubes. The cold trap is usually located in the oven of the gas chromatograph. In most cases the fused-silica capillary inner tube is simply a portion of the column inlet or retention gap. The cold gas is obtained by passing gas through a tube contained in a cryostat or generated from the expanding vapors above liquid nitrogen or liquid carbon dioxide. The lowest temperatures that can be obtained are  $-196^\circ\text{C}$  to  $-180^\circ\text{C}$  for liquid nitrogen and  $-65^\circ\text{C}$  to  $-60^\circ\text{C}$  with liquid carbon dioxide. The low thermal mass of fused-silica capillary tubes allows rapid heating by the circulating air in the chromatographic oven, by circulating hot air from an external source, or by resistive heating of the outer tube surrounding the capillary. A feedback mechanism is used to control the flow of cryogen or hot air to allow accurate setting and cycling between set point temperatures with total automation.

Cryogenic traps are convenient accumulation and injection devices for fast gas chromatography and interfaces for coupled-column gas chromatography, where a heartcut sample is collected and focused from the first column, and reinjected into the second column. The main requirement for a cryogenic trap used in these applications is efficient accumulation over time with rapid injection of the collected analytes as a narrow pulse in both time and space. Commercially available systems using a capacitance discharge for heating provide injection bandwidths of 5–20 ms.

### Data Handling

Dedicated electronic integrators and personal computers with appropriate software for integration are routinely used for recording chromatograms. Computer-based systems are increasingly used since they can combine instrument control functions with chromatogram recording as well as providing electronic data storage. Computer-based systems also provide



flexible approaches for reporting results and for performing advanced data analysis techniques.

The continuous voltage output from chromatographic detectors is not a suitable signal for data handling by computer. The conversion of the detector signal to a computer-readable form requires an interface usually resident as an expansion card in the computer. The interface scales the detector output to an appropriate range, digitizes the signal, and then transfers the data to a known location in the computer. The original input signal is transformed into a series of voltages on a binary counter and is stored as a series of binary words suitable for data processing.

The important characteristics of the analog-to-digital conversion device are its sampling frequency, resolution, and range. The accurate recording of chromatographic peaks requires that at least 10 data points are collected over the peak width. For typical sampling frequencies of  $\sim 5$  Hz undersampling is usually only a problem in high-speed chromatography. Analog-to-digital converters used for chromatographic applications are usually autoranging, meaning that the detector signal is automatically divided into ranges so as to provide sufficient resolution near the baseline for peak detection while accurately registering the peak maximum for large signals. The resolution of the analog-to-digital converter is the smallest change in the analog signal that can be seen in the digital output (specified as the number of bits).

Converted data are usually averaged (bunched) to minimize storage space. Using the local peakwidth to determine the frequency of bunching creates uniform sampling density throughout the chromatogram. Long stretches of stable baseline are stored in one bunch represented by a single datum and the number of times it recurs. Stored data are initially smoothed and peak locations identified by a slope sensitivity function. Small peaks and baseline artifacts are removed using a threshold function or later by a minimum area reject function. Standard procedures may fail in the case of complex baselines, tailing peaks, or excessive peak overlap.

## Preparative Gas Chromatography

The most common types of preparative-scale gas chromatographs are based on packed-column technology. For simple separations, a short, wide packed column is generally used (e.g.,  $1\text{--}3\text{ m} \times 6\text{--}10\text{ cm ID}$ ). For difficult separations, higher column efficiencies are required, and long, narrow packed columns (e.g.,  $10\text{--}30\text{ m} \times 0.5\text{--}1.5\text{ cm ID}$ ) are used. Wide-bore packed columns cannot be coiled, and a purpose designed gas chromatograph is required for their use.

Typical sample volumes are  $0.1\text{--}10.0\text{ ml}$ , which are too large for injection by conventional techniques developed for analytical packed columns. In this case, an automated injector employing pneumatic transfer from a reservoir through a capillary restrictor, a pneumatic piston pump, or syringe pump is used. The injection process is controlled by time; a necessity owing to the limited capacity of injection block heaters to flash vaporize large sample volumes. To assist solvent vaporization, an evaporation device between the injector and column is often installed. This is often a tube, heated separately from the column oven, and packed with glass beads to increase its thermal capacity. Separations are usually performed isothermally since wide-bore packed columns are slow to respond to temperature changes. Separated components are detected by a thermal conductivity or flame ionization detector connected to the column via a splitter, so that only a few per cent of the total column flow is diverted to the detector. Collection of sample vapors in the carrier gas effluent is usually performed automatically, initiated by the detector signal. The exit from the detector splitter is led out through a side wall of the column oven and thermostatted to avoid condensation of the sample. Several methods are used to trap samples, including packed and unpacked cold traps, solution and entrainment traps, total effluent and adsorption traps, and electrostatic precipitators.

Open-tubular columns have a lower sample capacity than packed columns, but are useful for isolating a few milligrams of components from complex mixtures under high-resolution conditions. Automated sequencing of injection and fraction collection to accumulate sufficient sample is generally used. Short lengths of wide-bore capillary tubes, coated with a thick film of stationary phase, are efficient traps for room-temperature sample collection.

*See also:* **Air Analysis:** Sampling; Workplace Air. **Gas Chromatography:** Multidimensional Techniques; High-Temperature Techniques; High-Speed Techniques; Online Coupled LC–GC; Pyrolysis; Detectors. **Head-space Analysis:** Static; Purge and Trap.

## Further Reading

- Dyson N (1998) *Chromatographic Integration Methods*. Cambridge: Royal Society of Chemistry.
- Felinger A (1998) *Data Analysis and Signal Processing in Chromatography*. Amsterdam: Elsevier.
- Grob K (1987) *On-Column Injection in Capillary Gas Chromatography*. Heidelberg: Huthig.



- Grob K (1993) *Classical Split and Splitless Injection in Capillary Gas Chromatography*. Heidelberg: Huthig.
- Grob RJ (ed.) (1995) *Modern Practice of Gas Chromatography*. New York: Wiley.
- Guiochon G and Guillemin CL (1988) *Quantitative Gas Chromatography for Laboratory Analysis and On-Line Process Control*. Amsterdam: Elsevier.

- Jennings W, Mittlefehldt E, and Stremple P (1997) *Analytical Gas Chromatography*. San Diego: Academic Press.
- Poole CF (2003) *The Essence of Chromatography*. Amsterdam: Elsevier.
- Rood D (1995) *A Practical Guide to the Care, Maintenance and Troubleshooting of Capillary Gas Chromatographic Systems*. Heidelberg: Huthig.

## Online Coupled LC–GC

**K Grob**, Official Food Control Authority of the Canton Zurich, Zurich, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Coupled high-performance liquid chromatography (HPLC)–gas chromatography (GC) is a two-step chromatography method: the sample is first separated by LC and then by GC. Usually LC serves as a sample preparation step for GC, but GC (particularly GC–MS) may also serve for the identification of LC peaks.

The development of online coupling of LC to GC was initiated in the early 1980s in the ‘mineral oil’ and ‘food analysis’ sectors. In 1989, a fully automated instrument was brought into the market by Fisons (Milan, Italy; today Thermo Electron), but was later discontinued.

### Reasons for Using LC–GC

The use of LC for sample preparation or preseparation has two main advantages: separation efficiency is much higher than with alternative chromatographic systems, such as conventional columns or solid-phase extraction (SPE), and control by standard LC detectors enables accurate control for cutting of the fractions to be transferred to GC. Hence, the technique offers outstanding selectivity for the analysis of trace components in complex samples or for group-type analysis of mixtures containing many similar components that are not separated by GC alone (e.g., mineral oil). The gain in efficiency for the preseparation is no longer substantial if broad mixtures of components are to be analyzed, which cannot be eluted from the HPLC column as sharp bands, e.g., pesticides.

### Offline versus Online LC–GC

The gain in separation efficiency is obtained irrespective of whether LC and GC are combined offline

or online. Offline LC–GC means collection of the LC fractions at the outlet of the column or detector, e.g., using autosampler vials that can directly be transferred to a GC autosampler. Introduction into GC usually requires a large-volume injection technique (100–500  $\mu$ l), because the amounts that can be injected into LC are usually not sufficient to obtain adequate sensitivity when only 1–2  $\mu$ l is injected into GC. The same techniques that are used for online transfer can be used for large-volume injection.

A routinely used example of offline LC–GC is the preseparation of complex migrates from food packaging materials (e.g., can coatings) by NPLC and GC–MS of the LC peaks. LC does not provide sufficient separation efficiency. GC–MS offers efficient additional separation, the ease of MS coupled to GC, and access to the large MS libraries. As identification is performed on only a few selected examples, online coupling offers little advantage.

When LC is coupled online to GC, far-reaching automation (sample preparation is integrated into the final analysis) enables the analysis of large numbers of samples with a minimum of manual interference. Highly accurate and reliable results can be expected, because losses are largely ruled out and the process is under almost complete control. For some types of analysis, exclusion of sample contamination from external sources is an additional important advantage.

### Typical Applications

Numerous applications for online LC–GC have been described for all important fields of analysis (summarized in Further Reading), but more than half of the methods applied concern mineral oil and edible oil or fat. A few are briefly outlined to illustrate typical strategies and conditions.

#### Aliphatic and Aromatic Hydrocarbons in Mineral Oil

As one of the first applications of online LC–GC in the early 1980s, mineral oil products were analyzed

- Grob K (1993) *Classical Split and Splitless Injection in Capillary Gas Chromatography*. Heidelberg: Huthig.
- Grob RJ (ed.) (1995) *Modern Practice of Gas Chromatography*. New York: Wiley.
- Guiochon G and Guillemin CL (1988) *Quantitative Gas Chromatography for Laboratory Analysis and On-Line Process Control*. Amsterdam: Elsevier.

- Jennings W, Mittlefehldt E, and Stremple P (1997) *Analytical Gas Chromatography*. San Diego: Academic Press.
- Poole CF (2003) *The Essence of Chromatography*. Amsterdam: Elsevier.
- Rood D (1995) *A Practical Guide to the Care, Maintenance and Troubleshooting of Capillary Gas Chromatographic Systems*. Heidelberg: Huthig.

## Online Coupled LC–GC

**K Grob**, Official Food Control Authority of the Canton Zurich, Zurich, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Coupled high-performance liquid chromatography (HPLC)–gas chromatography (GC) is a two-step chromatography method: the sample is first separated by LC and then by GC. Usually LC serves as a sample preparation step for GC, but GC (particularly GC–MS) may also serve for the identification of LC peaks.

The development of online coupling of LC to GC was initiated in the early 1980s in the ‘mineral oil’ and ‘food analysis’ sectors. In 1989, a fully automated instrument was brought into the market by Fisons (Milan, Italy; today Thermo Electron), but was later discontinued.

### Reasons for Using LC–GC

The use of LC for sample preparation or preseparation has two main advantages: separation efficiency is much higher than with alternative chromatographic systems, such as conventional columns or solid-phase extraction (SPE), and control by standard LC detectors enables accurate control for cutting of the fractions to be transferred to GC. Hence, the technique offers outstanding selectivity for the analysis of trace components in complex samples or for group-type analysis of mixtures containing many similar components that are not separated by GC alone (e.g., mineral oil). The gain in efficiency for the preseparation is no longer substantial if broad mixtures of components are to be analyzed, which cannot be eluted from the HPLC column as sharp bands, e.g., pesticides.

### Offline versus Online LC–GC

The gain in separation efficiency is obtained irrespective of whether LC and GC are combined offline

or online. Offline LC–GC means collection of the LC fractions at the outlet of the column or detector, e.g., using autosampler vials that can directly be transferred to a GC autosampler. Introduction into GC usually requires a large-volume injection technique (100–500 µl), because the amounts that can be injected into LC are usually not sufficient to obtain adequate sensitivity when only 1–2 µl is injected into GC. The same techniques that are used for online transfer can be used for large-volume injection.

A routinely used example of offline LC–GC is the preseparation of complex migrates from food packaging materials (e.g., can coatings) by NPLC and GC–MS of the LC peaks. LC does not provide sufficient separation efficiency. GC–MS offers efficient additional separation, the ease of MS coupled to GC, and access to the large MS libraries. As identification is performed on only a few selected examples, online coupling offers little advantage.

When LC is coupled online to GC, far-reaching automation (sample preparation is integrated into the final analysis) enables the analysis of large numbers of samples with a minimum of manual interference. Highly accurate and reliable results can be expected, because losses are largely ruled out and the process is under almost complete control. For some types of analysis, exclusion of sample contamination from external sources is an additional important advantage.

### Typical Applications

Numerous applications for online LC–GC have been described for all important fields of analysis (summarized in Further Reading), but more than half of the methods applied concern mineral oil and edible oil or fat. A few are briefly outlined to illustrate typical strategies and conditions.

#### Aliphatic and Aromatic Hydrocarbons in Mineral Oil

As one of the first applications of online LC–GC in the early 1980s, mineral oil products were analyzed

for aromatic and aliphatic compounds. As this was carried out mostly by mineral oil companies, hardly any details are public. A typical method (e.g., used for gasoline) uses a 2 mm i.d. normal-phase LC (NPLC) column packed with silica gel and pentane as the mobile phase. The paraffins are eluted at breakthrough and transferred to GC with about 200  $\mu$ l of eluent. Then the column is backflushed to recombine the aromatics to a narrow band, which is easily transferable to GC. Transfer into GC involves the retention gap technique (see below) to minimize the loss of volatile components: benzene and *n*-heptane are the limit.

### Mineral Hydrocarbons in Food

Many foods contain mineral oil material, either because of contamination from environmental sources or from packaging materials, lubricating oils, release agents, or as components in meat or eggs owing to the presence of paraffin oil or waste oils in the animal feed.

The final analysis occurs by GC-FID, because GC enables one to distinguish paraffins from natural and mineral origin and FID is the only detector with a practically constant response for all hydrocarbons. However, FID is not sensitive, and as the mineral paraffins form broad humps of unresolved components rather than a sharp signal, sensitivity tends to be low. The lack of selectivity is an additional problem, and the proof that the humps represent paraffins lies in a preseparation that allows only paraffins to reach GC. The final difficulty concerns blanks: laboratory glassware, fingers, etc. tend to release mineral paraffins, which renders manual sample preparation tedious. Online procedures eliminate most sources of sample contamination.

Paraffins are isolated from the sample matrix (edible oils or food extracts often containing much fat or oil) and separated from aromatics and olefins (e.g., sterenes from the raffination of edible oils) by NPLC. As it is difficult to maintain a high activity of silica gel to separate paraffins from olefins when large amounts of matrix material are injected, two columns are used: up to 30 mg edible oil or fat is injected onto a first column of 25 cm  $\times$  2 mm i.d. The hydrocarbons are transferred to a second column of the same dimension. The first column is backflushed with dichloromethane to remove the oil and other materials of the food extract. The second column only comes in contact with pentane as the mobile phase and nonpolar sample components, i.e. it maintains its high retention power for unsaturated components. Transfer of the fraction of about 400  $\mu$ l volume to GC mostly involves the retention gap

technique, using a 5–10 m  $\times$  0.53 mm i.d. uncoated precolumn fitted to a Y-piece branching to the solvent vapor exit and the separation column (see below). The detection limit of mineral paraffins in edible oils is 1–5 mg kg<sup>-1</sup>.

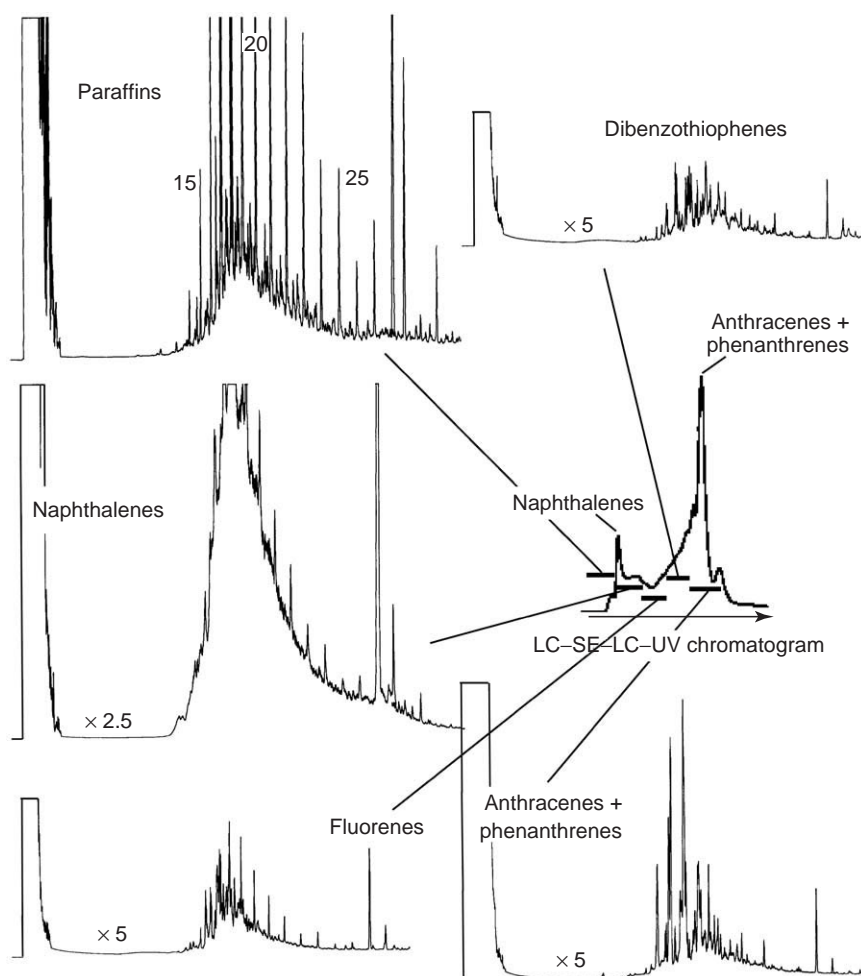
An online system for the analysis of aromatic hydrocarbons of mineral origin in foods, preseparating the aromatics by ring number, has been described. About 100 mg of oil or fatty extract is injected onto a 25 cm  $\times$  5 mm i.d. silica gel column. The resulting fraction of the hydrocarbons is too large (some 6 ml) and the mobile phase (pentane containing dichloromethane) is not suitable for the separation of the aromatics by ring number. Therefore, an online solvent evaporator consisting of a 40  $\times$  1 mm i.d. chamber packed with silica gel and connected to vacuum eliminates the solvent from the first LC step. The aromatics are transferred to an amino-derivatized silica gel column for separation by ring number before being introduced into GC. Mineral paraffins and aromatics in linseed oil contaminated from the jute bags used for transporting the linseeds are shown in Figure 1.

### Edible Oil Analysis

Some high-priced edible oils are adulterated by ever-more sophisticated methods, making it impossible to detect the added cheap oil by common control analysis (composition of fatty acids, triglycerides, or sterols). This compels the control to apply new and usually more complex analytical approaches. Since hundreds of samples have to be checked, a high degree of automation is a prerequisite for efficient control.

A first method simplified the classical analysis of the sum of free and esterified sterol: instead of saponification and tedious extraction from a soap solution, the sample is transesterified at ambient temperature and the mixture (containing the fatty acids as methyl esters) is preseparated by NPLC. The LC-UV chromatogram enables one to select the type of sterols ( $\Delta$ 5,  $\Delta$ 7, methyl-, dimethyl sterols) to be analyzed by GC (after online transfer through the loop-type interface).

Thousands of edible oils were checked for authenticity through the fingerprint of their 'minor components' (free sterols, sterol esters, wax esters, alcohols, tocopherols). The method primarily aimed at efficiency and automation. Some droplets of oil are silylated, diluted, and injected into NPLC. Through silylation, all the components of widely differing polarity are combined in a fairly narrow NPLC fraction eluted clearly before the triglycerides. This fraction is transferred to GC by the loop-type



**Figure 1** Perhaps the most complex LC-GC application used so far. LC-UV chromatogram from the second LC column (after isolation of the hydrocarbons on a large first LC column and eluent evaporation in a small online solvent evaporator): paraffins (with some carbon numbers indicated) and alkylated aromatics grouped by ring number. (Reproduced with permission from Moret *et al.* 1996.)

interface (all components are eluted from GC above 200°C).

Sterenes are dehydroxylated sterols, i.e., olefins, and result from raffination (bleaching, deodorization) or other strong heat treatment of edible oils. Their presence is used as an indicator of refined oil and their composition is fairly specific for a type of oil. For instance, the addition of refined vegetable oil to, e.g., nonrefined (extra virgin) olive oil is easily detectable in this way. The analysis presupposes high efficiency in the preseparation in order to isolate the sterenes from the (largely natural) paraffins and squalene with its many isomerization products. Even more detailed preseparation is required for a satisfactory resolution of the various types of sterenes, as needed for the identification of the nature of the oil the sterenes originate from. Two-step NPLC is applied, as described for the mineral paraffins in food, because only the second step provides reliable high

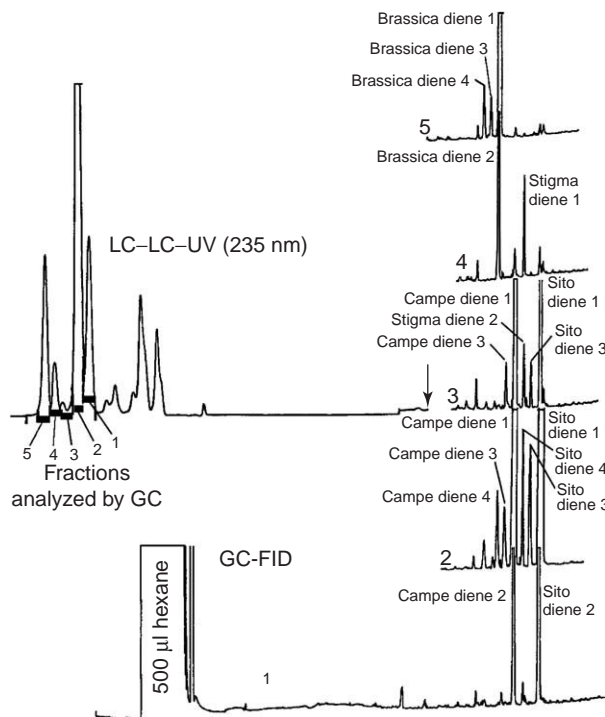
selectivity. As all compounds are rather high boiling, transfer involves the loop-type interface or other concurrent evaporation techniques. **Figure 2** shows the LC-LC-UV chromatogram of the sterenes from a refined rapeseed oil as well as the GC analysis of the fractions marked and numbered in the LC trace.

## LC for Coupled LC-GC

LC for coupling to GC is nearly exclusively in normal phase:

1. Transfer of normal-phase eluents to GC up to about 1.5 ml is easy.
2. At the concentrations required for the analysis, many sample matrices (mineral oil, edible oil, fat extracted from foods) are not soluble in reversed-phase media.
3. The solutes typically analyzed by GC are of rather low polarity and easy for NPLC.

4. Many samples are derivatized before injection into LC-GC, and such derivatives may not be compatible with water-containing eluents.



**Figure 2** LC–LC–GC analysis of the sterenes (dehydroxylated sterols) from refined rapeseed oil: LC–LC–UV chromatogram (center) with the fractions transferred to GC as numbered in the LC–LC–UV chromatogram. Designation of the sterenes: abbreviated name of the sterol (Sito, sitosterol; Stigma, stigmasterol; Campe, campesterol; Brassica, brassicasterol), number of double bonds (e.g., diene), and molecular structure in numbers referring to the original paper. (Reproduced with permission from Grob K, Biedermann M, Artho A, and Schmid JP (1994) LC, GC, and MS of sterol dehydration products. *Riv. Ital. Sostanze Grasse* 71: 533–538.)

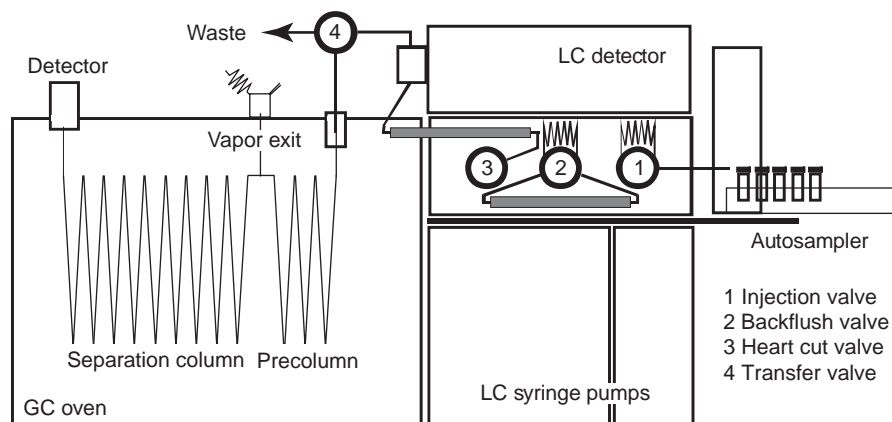
LC columns are usually packed with silica gel or silica derivatized with amino or cyano functions. Online coupling mostly involves 2 mm i.d. columns, compromising between the capacity for matrix material (e.g., edible oil) and the fraction volumes to be transferred. Furthermore, the commonly used flow rates (200–400  $\mu\text{L min}^{-1}$ ) well suit the requirements of transfer into GC, as evaporation rates through an early vapor exit are in the same range. Initially, some applications involved packed fused silica capillary LC columns with fraction volumes in the order of 5–10  $\mu\text{L}$ . However, most applications presuppose higher capacities for the sample matrix.

The mobile phase is usually based on pentane or hexane, containing modifiers, such as dichloromethane, methyl *tert.* methyl ether, or alcohols. Pentane is important because of its easy purification, the absence of oxidation products, and the ease of transfer to GC. Also, polar solvents can be used as long as they are free of water.

Size exclusion chromatography (SEC) with columns of  $\sim 3$  mm i.d. was used for the isolation of pesticides from edible oils or fatty food extracts. Reversed-phase LC with water-containing eluents was used with various (special) transfer techniques, but has probably never been used for routine applications. Water causes several problems, attack of deactivated surfaces being the most severe.

## LC Equipment

**Figure 3** schematically shows an LC–GC instrument as marketed by Fisons/Thermo Electron in the 1990s (Dualchrom 3000). LC equipment optimized for LC–GC differs from conventional LC instruments in several respects. The frequent use of pentane (bp 36°C)



**Figure 3** LC–GC instrument: autosampler, two syringe pumps (main pump and slave pump), LC-switching box with injection valve, valve with first LC column and backflush loop (filled from pressurized bottle, not shown), second LC column on heart cutting valve, LC detector (mostly UV), transfer valve directing to waste or into GC, on-column injector or other inlet device, precolumn (coated or uncoated) ending on T-piece directing to solvent vapor exit or separation column.



mixed with polar modifiers (often with even lower azeotropic boiling points) and the need for accurate flow rates in the order of  $100\text{--}500\ \mu\text{l min}^{-1}$  are the main reasons why syringe pumps are preferable. Often two pumps are used.

For most applications, the LC column (or the first of two columns) is backflushed with a stronger solvent after every sample in order to remove the matrix material. This presupposes a backflush valve of eight ports and a pressurized flask supplying flushing solvent to the backflush loop.

Preseparation with two LC columns presupposes at least one more switching valve enabling heart cutting. This valve is also used for multitransfer applications (several fractions of an LC run transferred to GC), switching the LC column to two plugs while waiting for the end of the GC analysis. Hence LC usually involves three valves.

Dedicated software supports rather sophisticated techniques. For instance, ‘peak detection’ enables adjustment of the time for transfer to GC to signals recorded by the LC detector, thus compensating for shifts in retention times.

## Techniques for Online Transfer to GC

The methods for transferring LC fractions to GC were developed hand in hand with large-volume GC injection. They mostly involved on-column techniques, since these show best performance and follow rather simple rules. The main drawback, the sensitivity to nonevaporating by-products, is not important, since efficient preseparation by LC is also efficient in removing the ‘contaminants’. The principal alternative, programmed temperature vaporizing (PTV) injection in solvent split mode, has rarely been applied.

Techniques enabling the transfer of typically  $200\text{--}1000\ \mu\text{l}$  fractions into GC primarily deal with selective solvent evaporation and the reconcentration of the solutes of interest to sharp initial bands.

### Band Broadening in Time/Space

In GC, two types of broadening of initial bands are distinguished, because they fundamentally differ in their characteristics and the techniques used for their reconcentration.

Band broadening in time is the result of slow introduction of a gaseous sample or of sample vapors into the column: for all solutes, bands have the same width in terms of time, because the first material has the same lead over the last in terms of chromatographic migration time. A band is considered to be sharp

if it is narrow compared to the width of the final peak. If a peak is, e.g., 5 s wide, the initial band should be no broader than 2 s, i.e., the solute material must start the separation process during not more than 2 s. If the column flow rate is, e.g.,  $2\ \text{ml min}^{-1}$ ,  $\sim 50\ \mu\text{l}$  of gas or vapor can enter the column during this time (a  $500\ \mu\text{l}$  fraction in pentane forms some  $150\,000\ \mu\text{l}$  of vapor, dilution with carrier gas not being considered!). Band broadening in time is the problem concerning the volatile components, i.e., those eluted between the column temperature during sample introduction and up to several tens of degrees above it.

Band broadening in space is the result of sample entering the GC system in the liquid phase: the liquid flows into the column until it is spread as a more or less stable layer. All solutes are dispersed over the whole length of the so-called ‘flooded zone’. The initial band should be narrow (short) compared to the terminal bandwidth (length), i.e., the length of the column over which a solute is distributed when leaving into the detector. For 15 and 50 m separation columns, initial band lengths of about 15 and 30 cm, respectively, can be tolerated, but flooded zones often reach 10 m. Band broadening in space is a problem of components eluted at least a few tens of degrees above the column temperature during transfer.

### Reconcentration of Bands Broadened in Time: Cold Trapping, Solvent Effects, Packed Beds (Traps)

Reconcentration of bands broadened in time requires that the first solute material entering the system is hindered in its gas chromatographic migration until the last one entered, i.e., a temporary increase of the retention power.

To achieve cold trapping, the whole column, or just an inlet section, is cooled below the elution temperature of the component(s) of interest as long as sample introduction lasts. Afterwards the temperature is increased to start the chromatographic process. Cooling the column below the elution temperature by  $100^\circ\text{C}$  results in reconcentration by a factor of about 100. Cooling is, however, limited by the dew point of the solvent/carrier gas mixture: lower temperatures cause flooding, and also solvent trapping.

A temporary increase of the retention power can also be achieved by condensed sample solvent acting as a stationary phase. There are two solvent effects: solvent trapping takes place in the flooded column inlet and describes the fact that the solutes are usually released only at the end of solvent evaporation. The elegance of this method is the automatic decay of the retention power at the end of solvent



evaporation. Phase soaking describes the result of co-chromatography of the solute material with the strongly overloading solvent peak in the coated column beyond the flooded zone.

A bed of highly retentive packing material can be used, e.g., in a PTV injector. After completion of sample introduction, this retentive power must be overcome by heating the bed to a high temperature.

### Reconcentration of Bands Broadened in Space: Retention Gap

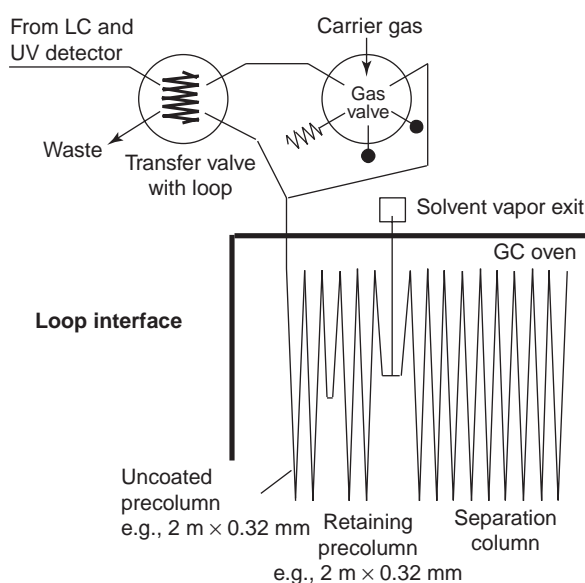
Solute material spread into the GC separation column by the flow of liquid can be reconcentrated by placing the flooded zone into an uncoated (but deactivated) precolumn. This precolumn represents a zone of low retention power ('retention gap'): the solute material moves through it at a low temperature and is accumulated in the inlet of the separation column because the temperature is still too low to enable noticeable chromatographic migration. The uncoated precolumn must be at least as long as the flooded zone. The retention gap technique shortens initial bands by a factor of 100–1000, enabling the acceptance of flooded zones as long as 50 m.

### Concurrent Eluent Evaporation

(Fully) concurrent eluent evaporation means the liquid (i.e., the solvent) is evaporated during its introduction, i.e., no sample liquid accumulates in the inlet. Basically unlimited volumes of liquid can be introduced in this way (although bands are increasingly broadened in time as introduction lasts longer). However, concurrent solvent evaporation also means that there is no solvent trapping to retain volatile components, i.e., these are lost through the vapor exit together with the solvent vapors or form broad peaks. Hence, the application of concurrent eluent evaporation is restricted to the analysis of components eluted at column temperatures above 120–180°C, depending on the solvent, the volume transferred, and the GC conditions.

To achieve solvent evaporation rates suitable for LC at flow rates of 200–400  $\mu\text{L min}^{-1}$ , vapors must be discharged at a rate of 50–300  $\text{mL min}^{-1}$ . This presupposes a solvent vapor exit (SVE) between the precolumn (sometimes consisting of two pieces, mostly of 0.53 mm i.d.) and the separation column. The SVE is fully open during solvent evaporation and switched to a strong restriction to achieve a small purge flow during analysis.

Concurrent eluent evaporation is most commonly performed with the loop-type interface (Figure 4): the liquid is pushed into the oven-thermostatted precolumn by the carrier gas and stopped there by its



**Figure 4** Design of the loop-type interface. Easy to use for large fraction volumes, but not suitable for components eluted less than 50–80°C above the column temperature during transfer. (Reproduced with permission from Grob K (2000) Efficiency through combining HPLC and HRGC: progress 1995–1999. *Journal of Chromatography A* 892: 407–420.)

own vapor pressure. This requires a GC oven temperature above the solvent boiling point at the inlet pressure. The sample is loaded into a loop with a volume corresponding to the fraction to be transferred, mounted in the transfer switching valve. The carrier gas is supplied from a separate valve (mounted on the same actuator), either pushing the sample liquid into GC or directly entering GC and backflushing the transfer valve. The solvent vapors are usually released through an early vapor exit installed after an uncoated and a retaining (coated) precolumn of 2–3 m each. The uncoated precolumn is used for the reconcentration of solutes spread by the moving liquid, and the retaining precolumn is used for the retention of solutes while solvent vapors are released.

Alternatively, the fraction is pushed by the LC pump into a 5–10 cm fused silica capillary containing a piece of wire and heated to 250–350°C ('wire interface'). The vapors escape through a retaining precolumn of some 3 m  $\times$  0.32 mm i.d. The oven temperature must again be above the solvent boiling point.

The loop-type and the wire interface stop the carrier gas flow during sample introduction; the solvent vapors are discharged through their expansion ('overflow system'). Alternatively, the fraction is introduced into a gas stream ('gas discharge'), usually by a transfer line entering the column through an on-column injector (on-column interface, see below).

Since the vapors are diluted with carrier gas, this permits lower oven temperatures. Performance for volatile solutes is improved, but the suitable oven temperature is more difficult to predict. The vapors are released through a 50 cm  $\times$  0.53 mm i.d. retaining precolumn.

### Retention Gap Technique

If more volatile solutes are to be analyzed (components eluted after the solvent peak), solvent trapping in a flooded precolumn is needed to retain these, i.e., not all solvent can be evaporated concurrently with the introduction. This means that an increasing amount of solvent accumulates during transfer and produces an extending flooded zone. An uncoated precolumn of sufficient capacity to retain the liquid, typically of 5–10 m  $\times$  0.53 mm i.d., is used upstream of the solvent vapor exit (see Figure 5). With the high gas flow rates obtained through an open vapor exit, a 10 m  $\times$  0.53 mm i.d. precolumn safely retains 250  $\mu$ l of (wetting) liquid. A corresponding volume of eluent can be transferred if concurrent solvent evaporation is negligible. If 80–90% of the solvent evaporates during introduction (partially concurrent eluent evaporation), 5–10 times larger fractions are amenable to GC (1250–2500  $\mu$ l).

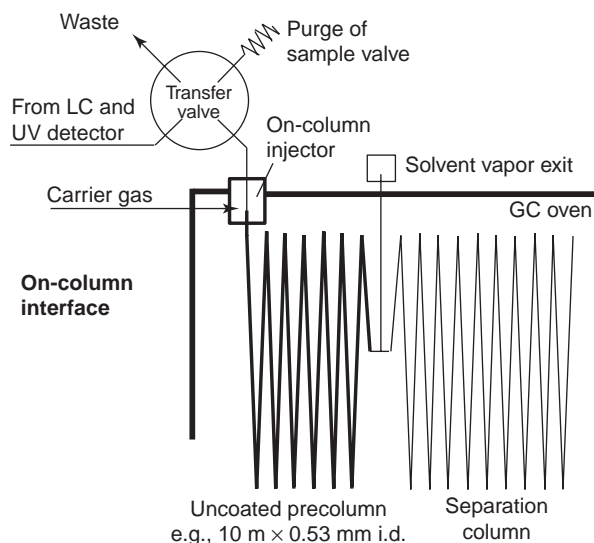
The on-column interface centers on an on-column injector. Instead of a syringe needle, a thin fused

silica transfer line is permanently installed. The LC fraction is driven by the LC pump and passes through a switching valve (transfer valve), directing the eluent either to waste or into the inlet of the uncoated precolumn. The carrier gas enters from the side (on-column injector) and discharges the solvent vapors, which is the prerequisite for solvent evaporation at the rear of the flooded zone and solvent trapping. The vapor exit is closed shortly before solvent evaporation is completed in order to retain volatile solutes within the system.

*See also:* **Gas Chromatography:** Multidimensional Techniques; High-Temperature Techniques; Instrumentation. **Liquid Chromatography:** Multidimensional; Instrumentation.

### Further Reading

- Beens J and Tijssen R (1995) An on-line coupled HPLC–HRGC system for the quantitative characterization of oil fractions in the middle distillate range. *Journal of Microcolumn Separations* 7: 345–354.
- Blomberg J, Mes EPC, Schoenmakers PJ, and van der Does JJB (1997) Characterization of complex hydrocarbon mixtures using on-line coupling of SEC and NPLC to GC. *Journal of High Resolution Chromatography* 20: 125–130.
- Goossens EC, de Jong D, de Jong GJ, and Brinkman UATH (1998) On-line sample treatment – capillary GC. *Chromatographia* 47: 313–345.
- Grob K (1991) *On-Line Coupled LC–GC*. Heidelberg: Hüthig.
- Grob K (1995) Development of the transfer techniques for on-line HPLC–capillary GC (review). *Journal of Chromatography* 703: 265–276.
- Grob K (2000) Efficiency through combining HPLC and HRGC: progress 1995–1999. *Journal of Chromatography A* 892: 407–420.
- Grob K and Biedermann M (1996) Vaporising systems for large volume injection or on-line transfer into gas chromatography: classification, critical remarks, and suggestions (review). *Journal of Chromatography* 750: 11–23.
- Mondello L, Dugo G, and Bartle K (1996) On-line microbore HPLC–capillary GC for food and water analysis. A review. *Journal of Microcolumn Separations* 8: 275–310.
- Moret S, Grob K, and Conte L (1997) Mineral oil polyaromatic hydrocarbons in foods, e.g. from jute bags, by on-line LC–solvent evaporation (SE)–LC–GC–FID. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung*, 204: 241–246.
- Vreuls JJ, Louter AJH, and Brinkman UATH (1999) On-line combination of aqueous sample preparation and capillary GC. *Journal of Chromatography A* 856: 279–314.



**Figure 5** On-column interface, here for the transfer of LC fractions containing volatile components by the retention gap technique. (Reproduced with permission from Grob K (2000) Efficiency through combining HPLC and HRGC: progress 1995–1999. *Journal of Chromatography A* 892: 407–420.)

## Pyrolysis

**T P Wampler**, CDS Analytical, Ltd., Oxford, PA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by W J Irwin, pp. 1820–1833, © 1995, Elsevier Ltd

### Introduction

The direct analysis of nonvolatile and polymeric organic samples, which may be complex mixtures of natural or synthetic origin, is compromised by virtue of their intractability. Various degradation methods have been developed and pyrolysis, in which a sample is subjected to a short burst of intense heat that initiates thermal fragmentation and the production of a range of smaller molecules, was one of the earliest of these. The pyrolysis products may be separated, identified, and quantified by chromatographic or spectroscopic means (analytical pyrolysis) to provide a characteristic profile of the sample. When the analysis is undertaken by gas chromatography (GC) directly integrated with the pyrolysis system, the technique is known as pyrolysis-gas chromatography (Py-GC). Typically, the sample – which may be a pure or formulated synthetic polymer, a biological matrix such as whole microorganisms, soil or organic geochemical components, or specimens from various origins for forensic analysis – is loaded into the pyrolyzer. Ideally, very small (microgram) amounts of sample are loaded onto a wire or filament, which is then rapidly heated to a temperature range that fragments the macromolecular components into smaller, volatile compounds – usually, 600–800°C. These pyrolysis products are then immediately swept onto the column of a gas chromatograph for separation, detection, and quantification to yield the pyrolysis chromatogram (pyrogram). The GC eluate may be further analyzed by mass spectrometry (MS) to provide reliable identification of the components – a technique known as Py-GC–MS. The pyrogram may be used to provide qualitative information on the identity or composition of the sample or quantitative data on its constitution, or else it may allow mechanistic and kinetic studies of thermal decomposition and fragmentation to be undertaken. A typical polymer pyrogram is shown in **Figure 1**.

The analysis of rubber, which yielded isoprene and dipentene, was the first reported application of analytical pyrolysis (1860). Accessible MS (1948) widened the utility of the technique, which developed into integrated pyrolysis-mass spectrometry (Py-MS, 1953). Pyrolysis with gas chromatographic analysis

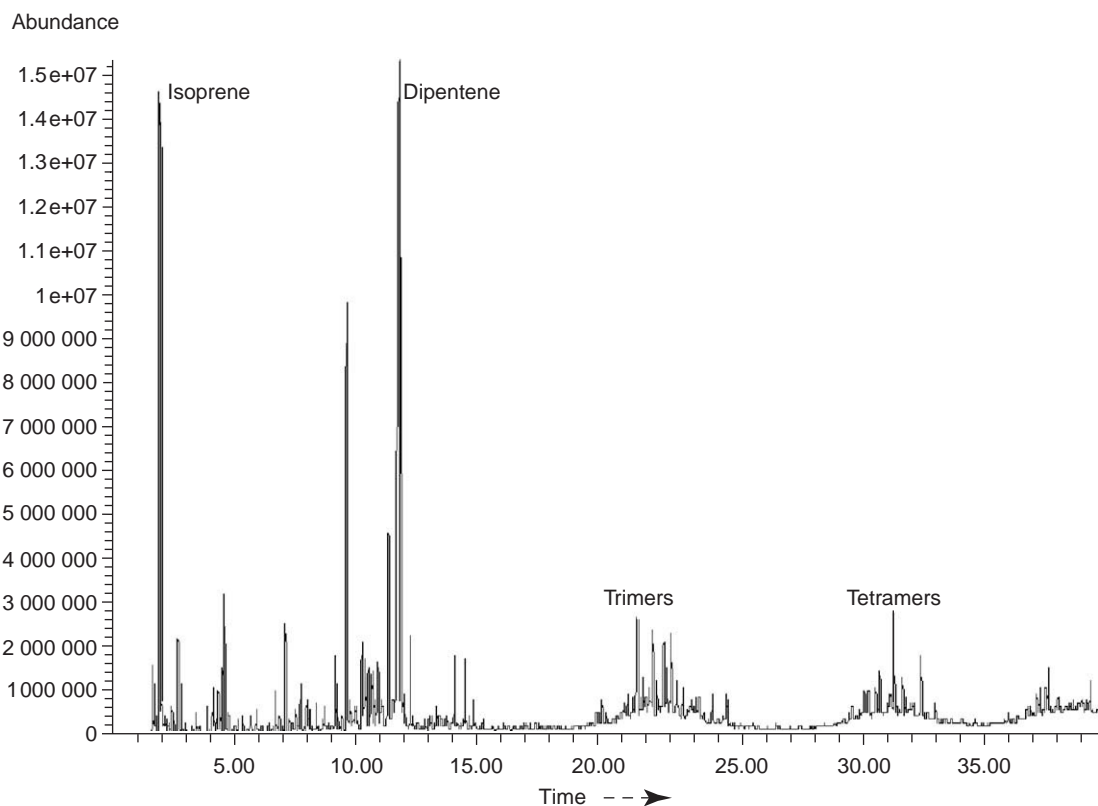
(1953) sought to overcome the deficiencies of contemporary MS by using the newly described GC technique (1952) and was rapidly succeeded by fully integrated Py-GC (1959). Py-GC with offline MS soon followed (1961) and combined Py-GC–MS was introduced in 1966. This was further extended (1976) to yield a fully automated system that was so compact and efficient that it could operate on the surface of Mars as part of the Viking program in the search for life on that planet.

### Principles of Pyrolysis-GC

Py-GC is a relatively simple procedure, but there are many processes that must be controlled to yield an effective analysis:

- Sample handling: selection of a representative portion, preparation, and loading into the pyrolyzer in a reproducible way.
- Pyrolysis: delivery of a precisely controlled thermal profile to the sample to initiate reproducible degradation into smaller fragments.
- Transference: migration of pyrolysis products from pyrolyzer to GC without loss or further reaction.
- Separation: efficient separation and elution of the pyrolysis products to provide peaks comprising single components.
- Detection: detection of all significant components in the mixture to allow a reproducible profile to be obtained.
- Quantification: measurement of the peak areas, and perhaps the use of calibration data, to allow effective analysis of the data.
- Identification: reliable confirmation of the structure of the pyrolysis fragments to enable rational assignment of the origins of the products.
- Data handling: mathematical or statistical analysis of quantitative data to provide reliable comparisons and conclusions.

For productive Py-GC analyses to be undertaken, consideration and control of these stages, particularly sample preparation, pyrolysis, and transfer, is necessary and much of the development of the technique has been aimed at defining and minimizing the sources of variation. Although much has been made of the potential for variability in Py-GC, the technique is simple to install and, in many cases, reliable results may be obtained without problems. As sample complexity increases and results are to be exchanged



**Figure 1** Pyrogram of natural rubber (polyisoprene) at 750°C, showing the monomer (isoprene), dimer (dipentene), trimers, and tetramers.

between various centers, control becomes more important. This control is now discussed in more detail.

## Pyrolysis

### Pyrolysis Temperature

When sufficient heat is supplied to a sample, a series of primary bond fissions result that yield the primary pyrolysis fragments. These reactions are likely to have different activation energies and, thus, the product distribution (and hence the pyrogram) is dependent upon the temperature of pyrolysis. Smaller, less characteristic fragments are frequently favored at higher pyrolysis temperatures. To ensure reproducible pyrograms, it is essential that a precisely controlled pyrolysis temperature is used. This is often described as the final pyrolysis temperature but is more properly referred to as the equilibrium temperature ( $T_{eq}$ ). The control of this parameter alone, however, is not sufficient to guarantee reproducible pyrolysis. The fact that the product distribution is highly sensitive to temperature means that the heating rate, too, may exert a profound effect on the appearance of a pyrogram. This is because pyrolysis reactions occur very rapidly, so that substantial, or total, degradation may occur while the

sample is being heated to  $T_{eq}$ . Reactions may thus occur nonisothermally during heating, with or without a subsequent isothermal phase. The temperature dependence of pyrolysis reactions ensures that analyses undertaken using heating regimes that accentuate such nonisothermal conditions are likely to be subject to much variation. A precisely defined temperature–time profile (TTP) is an essential component of pyrolyzer design; the temperature rise time (TRT) should be small and values of 8 ms or less are now possible.

Additionally, variability in the total heating time (THT) may result in variable volatilization of fragments from the pyrolyzer to the analyzer and high, uncontrolled temperatures may also cause secondary pyrolysis of the primary fragments. Again, such processes endanger reproducibility. For similar reasons, the cooling of the pyrolyzer unit after removal of the energy source should be as rapid as possible. Units with low thermal mass minimize secondary pyrolysis and evaporation during the cooling stage. A half-square-wave heating profile (rapid heating to  $T_{eq}$ —precise duration of heat—rapid cooling) is ideal for pyrolysis performance, with heating and cooling as rapid as possible so that the time during which the sample is subjected to  $T_{eq}$  is close to the THT. The pyrolyzer must also be interfaced with the analytical

device so that the loss of volatile samples is reduced to a minimum with no opportunities for adsorptive loss, condensation, or secondary reaction. In Py-GC, pyrolysis should be as close to the column as possible.

### Sample Handling

The temperature dependence of pyrolysis pathways implies that reproducible fragmentation requires that the entire sample experiences the same heating profile. Ideally, the temperature of the sample should follow that of the pyrolyzer probe, a situation that is best achieved with intimate contact between the sample and the filament. Very thin samples, near the center of the pyrolysis filament (to avoid cooler end regions), are advised. This also minimizes temperature gradients throughout the depth of the sample. In contrast, thicker films allow the primary pyrolysis products, which are produced close to the surface of the pyrolysis wire, to diffuse through non-degraded sample before escaping to the analyzer. At the high temperatures involved, uncontrolled, competitive secondary reactions that will compromise pyrogram reproducibility may occur. Samples should preferably be within the 5–50  $\mu\text{g}$  range for optimum results and deposited directly on to the heated wire or conductor. For soluble compounds, a known volume of solution in a volatile solvent may be used for quantitative transfer to the filament as a thin, uniform film. Greater care is necessary with insoluble samples, firstly to ensure that a thin uniform film is obtained – perhaps by deposition on to a rotating wire – and to ensure that the sample is homogeneous or representative of the bulk. Solvent must be totally removed prior to pyrolysis. Heated-filament systems allow gentle preheating for this purpose. The pyrolysis filament should also be scrupulously clean: heating in a moist hydrogen atmosphere for ferromagnetic conductors or prefiring filament systems have been recommended.

### Instrumentation

Instrument development has concentrated on the control of TRT,  $T_{\text{eq}}$ , and THT while enabling small samples to be handled easily. Although some specialized pyrolysis units, such as high- or low-energy laser, dielectric discharge, infrared, and photolytic systems and those suitable for liquids, have been reported, most workers have used pyrolyzers that are based upon furnace heating, inductive heating, or a resistively heated filament; each may be designed to ensure adequate performance, so that final temperatures with relative standard deviations of <1% are obtainable.

### Furnace Pyrolyzers

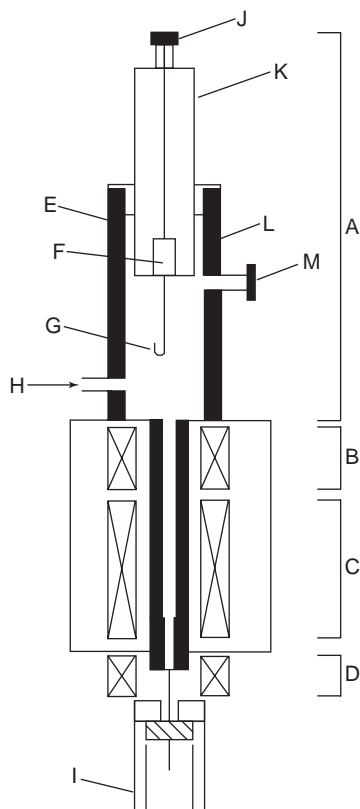
Furnace systems were the first to be used in pyrolysis and usually involve the sample being dropped or pushed into the preheated pyrolysis zone. Such units have been widely described but have a low reputation for reproducibility. This is largely due to the poor control of the heating profile and the large sample sizes that have been used. The pyrolyzer design has often favored secondary reactions, although in some instances (e.g., the mechanism of tobacco pyrolysis) this may be of advantage. A vertical design, however, is outstanding in its performance. The sample ( $\sim 50 \mu\text{g}$ ) is held in a platinum bucket, which is suspended above the heating zone so that no thermal reactions can occur. When the bucket is released, the sample falls into the heated zone to initiate a rapid pyrolysis. Volatile products are rapidly swept away from the heated zone by the carrier gas; to minimize secondary products the low-dead-volume unit progressively narrows to increase carrier gas flow rates. This system may also be modified to provide two independent heating zones (Figure 2). At first the sample is held in an upper chamber at 300°C to desorb volatiles such as plasticizers, giving a desorption chromatogram, followed by pyrolysis at 550°C in the lower chamber to generate the pyrogram.

### Curie-Point Pyrolyzers

A ferromagnetic sample-holder is placed in a radio-frequency (RF) oscillator ( $\sim 500 \text{ kHz}$  to 1.2 MHz). This induces an alternating magnetic flux in the conductor and hysteresis losses and eddy currents cause a rapid increase in temperature. Heating continues until the Curie point of the material is reached, when a transition from ferromagnetism to paramagnetism occurs, energy input to the sample holder falls, and the temperature is held constant. This Curie-point temperature is a function of the specific alloy used, so the pyrolysis temperature is not continuously variable but a range of values may be obtained by adjusting the composition of the conductor:

Temperature (°C)	Composition	
	%Ni/%Fe	%Ni/%Co
400	45/55	
510	49.4/50.6	
590	60/40	
700		67/33
770	0/100	
800		55/45
900		40/60
1128		0/100

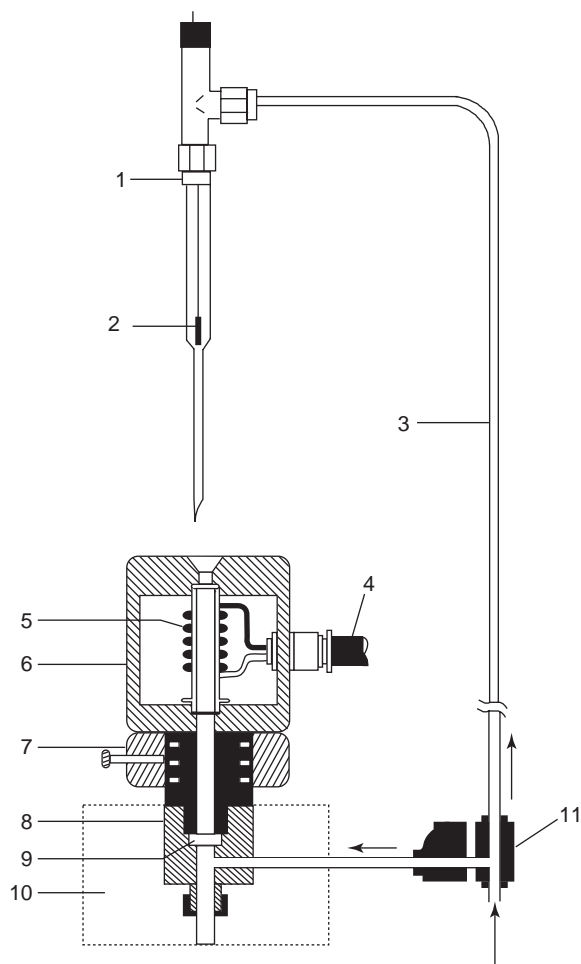




**Figure 2** Two-stage high-performance vertical microfurnace pyrolyzer: A, sample holder unit; B, thermal desorption oven, 50–400°C; C, pyrolyzer oven, 100–800°C; D, Py-GC interface, 50–400°C; E, O-ring; F, sample holder chuck; G, sample cup; H, carrier gas; I, GC column; K, pole; L, cylinder; M, air-purging nut. (Diagram kindly supplied by Professor Tsuge S, Nagoya University, Japan as described in Watanabe C, Teraishi K, Tsuge S, Ohtani H, and Hashimoto K (1991) *Journal of High Resolution Chromatography* 14: 269–272.)

This system offers the choice of sample holder shape (wire, filament, boat, tube, folded foil) and the lack of electrical contact between the power supply and the conductor, which facilitates automation. Conductors are frequently enclosed within quartz tubes, which collect nonvolatile products (Figure 3).

In practice, the temperature control of Curie-point pyrolyzers is dependent upon features such as the power of the RF generator and the geometry of the sample holder. With low-power units (~30 W), TRTs of 500 ms to 600°C and 1.35 s to 700°C are typical with the Curie point never being reached, while high-power units (2 kW at 1.1 MHz) give better heating (20–30 ms to 700°C) but may overshoot the Curie point. Different RF frequencies require different-diameter wires for optimal performance while the shape of the conductor, and its position within the induction coil, may significantly affect the heating profile; even small variations in alloy composition cause deviations in pyrolysis temperature.



**Figure 3** High-powered Curie-point pyrolyzer: 1, glass pyrolysis injector with stainless steel needle; 2, ferromagnetic wire; 3, Teflon carrier gas tubing; 4, impulse cable; 5, induction coil; 6, aluminum housing; 7, adaptor for fastening; 8, GC inlet; 9, GC septum; 10, GC column; 11, carrier gas changeover valve. (Reproduced from Fischer Labor- und Verfahrenstechnik, Meckenheim bei Bonn, Germany.)

Further, it is not possible to vary the pyrolysis temperature readily. Stepped pyrolyses, in which the same sample is repetitively subjected to a series of increasing temperatures, are not possible, and a good description of the thermal profile of the sample holder is generally not available. However, they are used frequently; they are robust and convenient to operate; and they allow a series of samples to be loaded onto the conductors and queued for analysis.

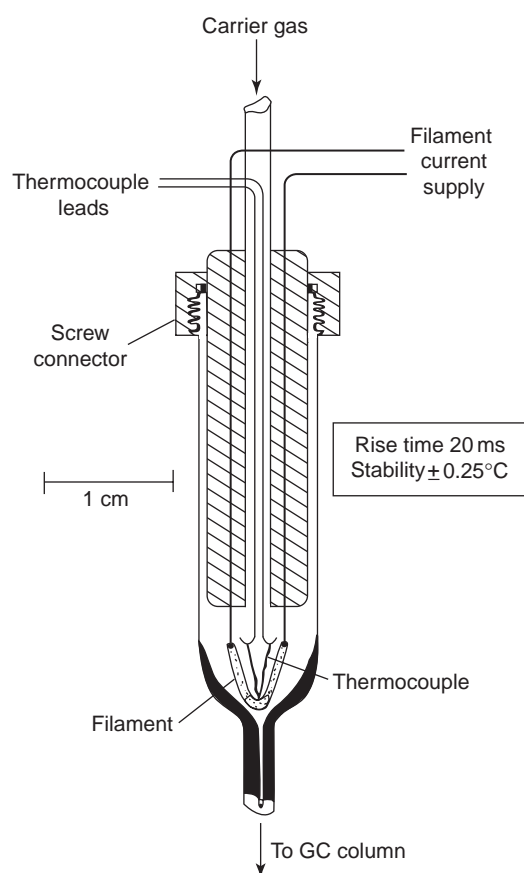
### Heated-Filament Pyrolyzers

With resistively heated-filament pyrolyzers, the sample is placed on a ribbon that is heated by the passage of an electric current. Fixed voltages provide poor control. With low voltages very long TRTs (10–30 s) result, whereas at higher voltages TRTs are decreased



( $\sim 1$  s) but very high temperatures that might melt the conductor may be generated. Better performance is provided by boosted heating systems in which an initial surge of power, provided by a high current or a capacitor-discharge system, heats the filament rapidly. Power to the conductor is reduced to a maintenance current when  $T_{eq}$  is reached. A highly refined system that uses a capacitor discharge is the Pyroprobe<sup>®</sup> (CDS analytical, Oxford, PA, USA) in which a platinum pyrolysis filament may operate as the sample holder, the heating element, and the temperature sensor. The increase in resistance of the filament with temperature is monitored to provide feedback to a computer that controls the filament voltage. The system is capable of controlled variable heating rates ( $20^{\circ}\text{C ms}^{-1}$  to  $0.01^{\circ}\text{C min}^{-1}$ ) giving declared TRTs of 8 ms to  $600^{\circ}\text{C}$  and 17 ms to  $1000^{\circ}\text{C}$ . The pyrolysis temperature may be set from ambient up to  $1400^{\circ}\text{C}$  at  $1^{\circ}\text{C}$  intervals and variation in pyrolysis time from 0.01 s to 1000 min is also possible. Various probes are available to facilitate sampling handling and include the pyrolysis of samples held in quartz boats. In contrast, the Pyrola<sup>®</sup> system (Pyrolab, Lund, Sweden) uses two sequential half-square-wave voltage pulses. The first is high enough to heat the Pt-filament to the expected pyrolysis temperature (up to  $1400^{\circ}\text{C}$ ) and short enough ( $\text{TRT} \leq 8$  ms) to minimize secondary effects. The second current pulse compensates for cooling effects. An optic cable transmits the light from the middle of the filament to a photodiode in the control unit. In this way the TTP can be displayed on the PC. This system is unique in that the TTP can be measured and printed. The pyrolysis temperature is also measured by the resistance of the Pt-filament, which is needed, since the photodiode is not sensitive at temperatures lower than about  $600^{\circ}\text{C}$ . A removable glass cell is surrounding the Pt-filament in order to let the light pass to the optic cable and to let the nonvolatile pyrolysis products condense to protect the GC-MS.

It is essential in mechanistic work to exert precise control over the pyrolysis temperature. However, quoted temperatures may be average values over the whole filament. A pyrolysis unit that uses a refined feedback system was developed at the University of Birmingham to overcome this problem (Figure 4). The filament temperature is measured near its central point by means of a very fine thermocouple spot-welded to the rear of the filament. The thermocouple output is compared to that of a reference voltage preset to correspond to the required  $T_{eq}$ . A large current ( $\sim 40$  A) drives a temperature increase and, as the two voltages converge, the filament current is reduced rapidly to a maintenance level. TRTs of 50 ms to  $800 \pm 0.5^{\circ}\text{C}$  were obtained with 15 ms to



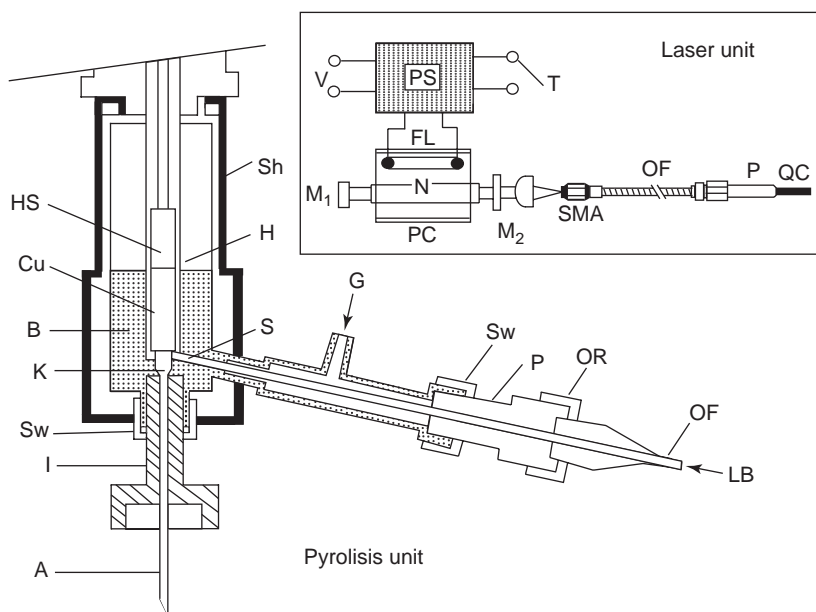
**Figure 4** Advanced boosted heated-filament pyrolyzer with thermocouple feedback. The chromel–alumel thermocouple wires ( $d$ ,  $25\ \mu\text{m}$ ) attached to the nichrome ribbon have a small thermal capacity with a fast response time. When mounted on the GC, the chamber is surrounded by a heater jacket. (Reproduced with permission from Lehrle RS, Robb JC, and Suggate R (1982) *European Polymer Journal* 18: 443–461; © Elsevier.)

$500^{\circ}\text{C}$  achievable with less-robust filaments. Cooling periods are the limiting factor in producing the ideal square-wave thermal profile.

Resistively heated-filament pyrolyzers offer the most versatility of the available units. They allow a wide range of programmed temperature and time profiles including stepped pyrolysis. This allows the elucidation of the thermal stability profile of the sample; it provides data to allow the kinetic analysis of polymer degradation and may facilitate the identification of unknown samples.

### Laser Pyrolyzers

Pyrolysis by laser has involved both high- and low-powered systems, with the former, in general, giving smaller, less-characteristic molecular fragments. The specialized nature of the technique has limited the applications of this pyrolysis mode. However, a relatively cheap system has now been designed that



**Figure 5** Convenient low-power laser pyrolysis system: A, injection needle; B, stainless steel body; Cu, copper cylinder; FL, flash lamp; G, gas; H, heater; HS, heated sensor; I, injector; K, pyrolysis chamber; LB, laser beam; M<sub>1</sub>, HR mirror; M<sub>2</sub>, output mirror; N, Nd-GGG medium; OF, optical fiber; OR, O-ring; P, probe; PC, pumping chamber; PS, power supply; QC, quartz capillary; S, sample; Sh, shield; Sw, Swagelok; T, trigger; V, voltage supply. (Reproduced with permission from Cecchetti W, Polloni R, Bergamasco G, *et al.* (1992) *Journal of Analytical and Applied Pyrolysis* 23: 165–173; © Elsevier.)

simplifies and improves access to laser pyrolysis. This system uses an Nd–Cr–GGG laser that delivers 600 mJ pulses of 500  $\mu$ s with repeats at 40 s intervals. No cooling is necessary due to the slow repeat rate and the laser energy is delivered to the pyrolysis chamber via an optical fiber (Figure 5).

## Pyrolysis-GC

### The Integrated System

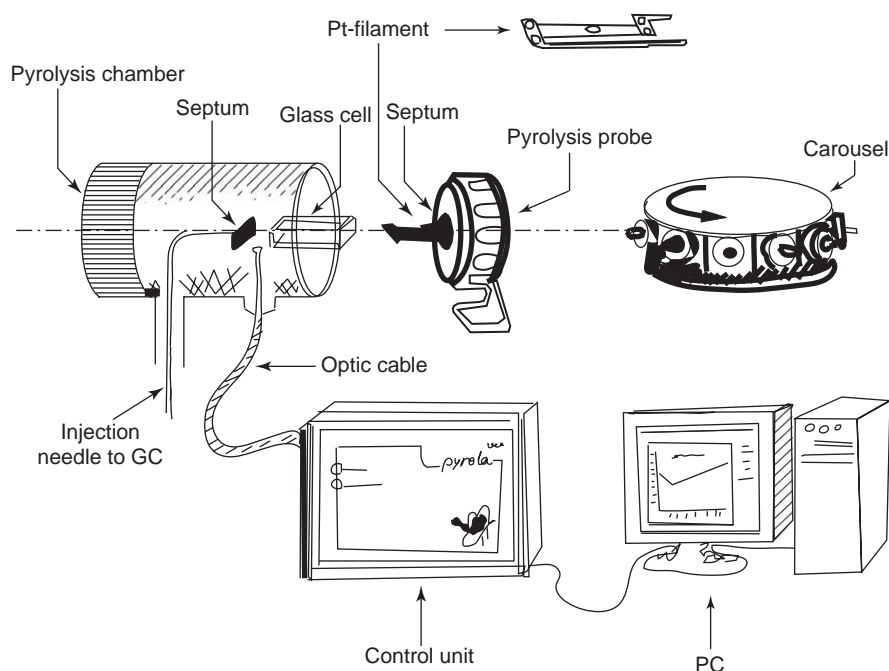
Pyrolyzers are interfaced with standard gas chromatographs by inserting the pyrolysis unit into the GC carrier gas upstream of the column inlet, and the products are swept directly onto the column for analysis. Connections should minimize dead volume and the presence of cold spots that may result in component loss, peak broadening, lack of resolution, and insensitivity (Figure 6). Capillary columns can separate small amounts of product and they provide high resolution, which allows separation of many of the products from complex samples, and are used routinely, generally with split ratios of 50–100:1. To compensate for dead volume while allowing the analysis of small samples, cryogenic focusing has been employed. This allows on-column trapping of the pyrolysate, using a liquid nitrogen trap, prior to analysis and allows better resolution and more sensitive detection. The development of tarry deposits on the top of a capillary column reduces efficiency;

these should be eliminated by removing the affected inlet of the column or by back-flushing.

Early Py-GC analyses used flame ionization detectors. These are sensitive and respond quantitatively over a wide range of sample loadings, but for some components this may prove insensitive. For such compounds, and to provide reliable identification of eluted components, MS is generally employed (Py-GC–MS), with electron impact and chemical ionization being readily available. Py-GC–MS also provides a powerful data-handling ability where stored spectra may be reconstituted to provide mass pyrograms, mass spectra of individual components or removal of background contamination, or to monitor uniquely specific fragments.

### Automation

Pyrolyzers have been adapted to provide automatic, unattended control of Py-GC. An early system used precoated pyrolysis wires held in quartz tubes on a turntable. These were sequentially loaded, accurately positioned in the induction coil, pyrolyzed, analyzed by capillary GC, and ejected. An alternative has used an automatic solids injector for samples enclosed in iron foil, and a furnace system has enabled sampling of the Martian surface. Autosampling systems based upon conventional pyrolyzers are now commercially available for resistively heated filaments, microfurnaces, and Curie-point pyrolyzers. One such system



**Figure 6** Heated interface for Py-GC. Sample material is placed directly onto the platinum filament, which is inserted into the glass cell of the interface. The injection needle passes through the injection port of the GC to the column inlet. (Diagram kindly supplied by Dr I. Ericsson, Pyrolab, Lund, Sweden.)

(CDS Analytical, USA) uses an autosampler with a carousel to hold 36 quartz tubes for sequential analysis (Figure 7). The tubes are dropped into a stabilized platinum coil for pyrolysis, where air is purged from the system before connection to the GC. Another automated filament system (Pyrolab, Sweden) uses either 6 or 14 probes (Figure 6). Different pyrolysis methods like isothermal, sequential, fractionated pyrolysis, and pyrolysis can be chosen individually for each probe. In another design (Japan Analytical Industry Co., Japan) samples enclosed in inductive foil are loaded into a 20-compartment magazine. This is flushed with carrier gas and then a sample is dropped into the induction coil for pyrolysis. Exhausted samples are retrieved magnetically and are retained in a cool storage compartment. An auto-sampling microfurnace system (Frontier Laboratories LTD, Japan) uses a tray that holds up to 48 deactivated stainless steel sample cups.

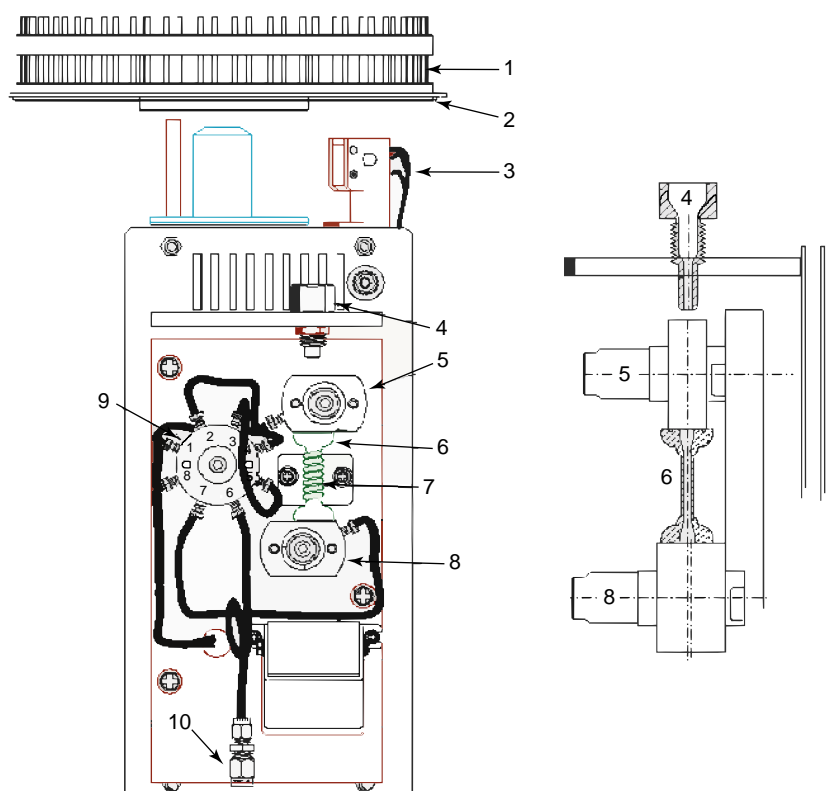
## Data Handling

Appropriate data handling is essential to the successful application of Py-GC techniques. This is particularly so when fingerprint comparisons are to be made using standard pyrograms to identify unknown samples. In forensic applications, taxonomy or other areas of classification, where complex pyrograms are obtained, a detailed statistical analysis, using

multivariate methods such as cluster analysis, non-linear mapping, and factor analysis may be required. In contrast, in studies concerned with the mechanistic and kinetic aspects of polymer degradation, relatively few primary products may be observed and it is sufficient to be able to identify and quantify these, perhaps by the use of Py-GC-MS. In general, the same integration routines employed in the analysis of peaks found in chromatograms of routine GC applications may be applied in the analysis of pyrolysis chromatograms.

## Scaling of Pyrograms

The use of Py-GC to characterize unknown specimens by comparing a pyrogram from an unidentified sample with that of a standard is a common application. If such a comparison is to be made on the basis of peak area intensities rather than by the presence of unique components, care must be taken to ensure that variations in sample loading do not compromise the analysis. Rescaling the peak areas may be achieved simply by normalizing all peaks to the intensity of the largest peak in the pyrogram, as indicated in eqn [1] in Table 1. However, some peaks in the pyrogram are intrinsically more variable than others and this simple procedure may threaten reliable transformation if normalization is applied to such a peak. To minimize intensity differences, the



**Figure 7** Automated resistively heated filament pyrolysis using quartz sample tubes, front view, and side view of the pyrolysis chamber. 1, Quartz tubes; 2, sample carousel; 3, optical sensor for locating tubes; 4, funnel to guide tubes into inlet valve; 5, inlet valve; 6, quartz chamber; 7, heating filament; 8, outlet valve; 9, GC online valve; 10, connection to GC injector.

use of the fractional total area (intensity) for each peak, as indicated by eqn [2] (Table 1), is preferred.

Problems may be also encountered in obtaining precise retention times over extended periods of analysis. This may lead to difficulties in the assignment of corresponding peaks within different pyrograms, particularly if many components are present. When the intensities of such peaks are to be compared for identification purposes, normalization of retention times may be undertaken. This procedure uses standard components, which may be present naturally in the pyrogram and are readily identified as such, or are added as internal or external markers. This allows the calculation of a time-rescaled pyrogram, from the retention times of the standard components, according to eqn [3] (Table 1).

### Fingerprint Comparisons

Except in the simplest of cases, it is unusual for all peaks within a pyrogram to be used in the analysis, and hence the matter of peak selection must be addressed. It is sensible to select the most intense peaks within the pyrogram for this purpose, but two other aspects should be considered. These are the within

(intra)-sample and the between (inter)-sample variations. The within-sample variation measures the reproducibility of the pyrogram and highly variable peaks could be eliminated from the analysis. However, those peaks that have a high between-sample variation are most useful in assigning identity. Care needs to be taken that weak or variable peaks, which may be highly diagnostic, are not eliminated from the dataset. Typically, three peaks may be sufficient for simple synthetic polymer samples, while for taxonomic applications 13–24 or even up to 100 may be used. Various techniques have been used to assess the degree of correspondence between pyrograms; these are summarized in Table 1 (eqns [4]–[10]). To compare pyrograms that have only subtle differences, more complex chemometric methods should be chosen. The nonlinear mapping technique, which uses a two-dimensional representation of differences between pyrograms, has shown particular promise in taxonomic applications using pyrolysis data.

### Applications

Py-GC is widely used in the analysis of complex, insoluble, and nonvolatile samples and applications are

many and varied. Some of these are illustrated with examples from the major fields of application.

### Synthetic Polymers

Typical pyrolysis reactions of addition polymers are illustrated in Figure 8 and involve:

- Depolymerization or unzipping (Figure 8A), which results in successive free-radical eliminations of the monomer, often accompanied by intramolecular rearrangement to generate dimers and trimers. This behavior is typical of polystyrene and polymethacrylates.
- Side-group elimination (Figure 8A), which involves loss of pendant groups on the polymer backbone to form a polyene. Further reaction involves scission, aromatization, and char formation. Typical is the case of PVC, which eliminates HCl followed by backbiting to form benzene, toluene, naphthalene, and other aromatics.
- Random chain scission (Figure 8C), which occurs in polymers such as polyethylene with the formation of a free radical. This initiates C–C bond scission with the formation of an olefin and a terminal free radical. The latter may abstract hydrogen from the chain to produce a methyl terminus. Random repetition of these processes gives a homologous series of alkane, -alkene, and  $\alpha,\omega$ -diene.

In addition to empirical fingerprint comparisons, for which library data are available and which are more usual in taxonomic and forensic work, these reactions may be used to study various aspects of polymer structure and properties.

### Analysis of Composition

Mixtures, formulated blends, or copolymers usually provide distinctive pyrolysis fragments that enable qualitative and quantitative analysis of the components to be undertaken, e.g., natural rubber (isoprene, dipentene), butadiene rubber (butadiene, vinylcyclohexene), styrene–butadiene rubber (butadiene, vinylcyclohexene, styrene). Pyrolyses are performed at a temperature that maximizes the production of a characteristic fragment, perhaps following stepped pyrolysis for unknown samples, and components are quantified by comparison with a calibration graph from pure standards. Different yields of products from mixed homopolymers and from copolymers of similar constitution may be found owing to different thermal stabilities. Appropriate copolymers should thus be used as standards and mass balance should be assessed to allow for nonvolatile additives. The amount of polymer within a matrix (e.g., ~0.5%

styrene–butadiene on filter paper) may also be readily quantified without extraction.

### Determination of Structure

Detailed information concerning the microstructure of polymers may be obtained using high-resolution Py-GC; block and random copolymers may be distinguished and sequence lengths may be investigated. Poly(chloropropylene) undergoes degradation by side-chain stripping (Figure 8B) and the preferred head-to-tail component produces 1,3,5-trimethylbenzene. Small amounts of head-to-head sequences generate the 1,2,4-isomer. The random chain scission of polyolefins, particularly when combined with hydrogenation, provides information on the stereoregularity.

For example, the reduced  $C_{11}$ ,  $C_{12}$ , and  $C_{13}$  fragments of polypropylene appear as doublets due to two geometrical isomers that originate from the syndiotactic or isotactic disposition of the pendant methyl groups. Similar analyses provide details on branching due to the clusters of isomeric peaks. Product distribution is also dependent upon molecular mass. Figure 9 records pyrograms from poly(2,6-dimethyl-1,4-phenylene ether), which yields a range of phenolic products due to Fries rearrangements. End-group analysis using specific fragments correlated well with  $^{13}\text{C}$  NMR values.

### Mechanisms of Thermal Degradation

Product distribution, variable or stepped pyrolysis and kinetic studies have all been used to expose mechanistic events during polymer degradation. For example, the thermal degradation of polystyrene has been elucidated using the block copolymer poly(styrene-*b*-styrene- $d_8$ ). Hybrid and homo monomers (styrene) and dimers (2,4-diphenylbut-1-ene) were detected but without hybrid trimers (2,4,6-triphenylhex-1-ene) (Figure 10). The amount of hybrid dimer far exceeds that which might arise from adjacent residues, while similar proportions of dimers were obtained when the two homopolymers were pyrolyzed together. Data indicate intermolecular reaction rather than the previously proposed 1,3-transfer.

Kinetic parameters may be determined by pyrolyzing samples for various periods of time and measuring the extent of degradation. For example, if a pyrolysis time  $t$  yields a component peak of area  $A$ , and an extended pyrolysis period, allowing complete degradation, gives an area  $A_\infty$ , the rate constant for a first-order degradation ( $k_{\text{obs}}$ ) may be obtained from eqn [11] using a series of pyrolysis times:

$$-\ln(1 - A_t/A_\infty) = k_{\text{obs}}t \quad [11]$$



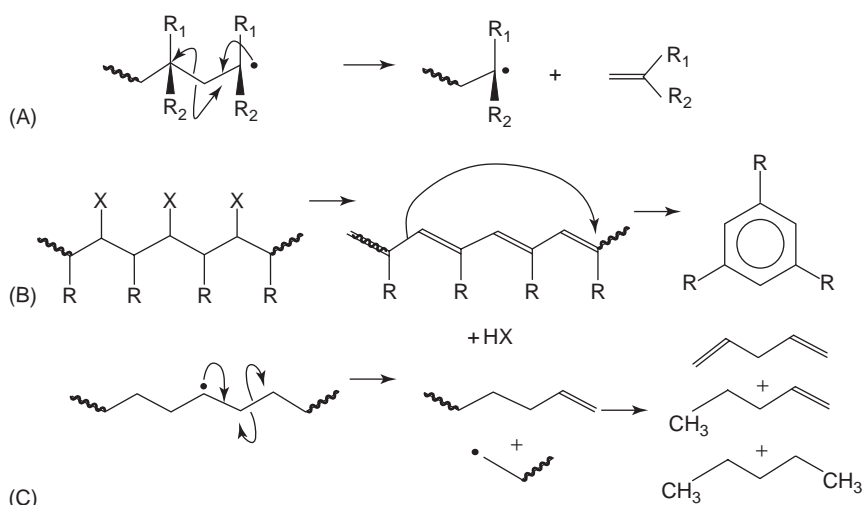
**Table 1** Expressions used in the quantitative description of pyrolysis data

Value	Equation	Parameters	Comments
Normalized intensity	$I_i^n = \frac{100I_i}{I_{100\%}}$	[1] $I_i^n$ = intensity (%) of the normalized peak $I_i$ = intensity of peak before normalization $I_{100\%}$ = intensity of the largest peak in the pyrogram	Re-scales peak intensities to largest in pyrogram: ideally, this should be of low variability
Fractional intensity	$I_i^n = \frac{I_i}{\sum_{j=1}^N I_j}$	[2] $I_i^n$ = fractional intensity of a normalized peak $i$ of initial intensity $I_i$ $I_j$ = intensities of the $N$ selected peaks in the pyrogram	Re-scales intensities as a fraction of total intensity: reduces problem of highly variable peaks used for normalization
Corrected retention time	$t_B^s = t_A^s + \frac{(t_B^r - t_A^r)(t_C^s - t_A^s)}{t_C^r - t_A^r}$	[3] $t_B^s$ = rescaled retention time of component B, occurring between the two standards A and C Subscripts A, B, C indicate three species (A, B, C); superscripts r ( $t^r$ ) indicate retention times in the original program; superscripts s ( $t^s$ ) indicate retention times in the standardized pyrogram	Standardizes retention times to those of standard components eluting before and after the peak of interest. Compensates for variable retention times in pyrogram
Similarity value	$S_{i,j}^v = \frac{100N_c}{N_c + N_u}$	[4] $S_{i,j}^v$ = similarity value between pyrogram ( $i$ ) and pyrogram ( $j$ ) $N_c$ = number of peaks common to both pyrograms $N_u$ = number of unique peaks in both pyrograms	Simple comparison of two pyrograms. A value of $S_{i,j}^v = 100$ indicates a perfect match
Dissimilarity index	$d_{i,j} = 100 \sum_{k=1}^N \frac{ R_i^k - R_j^k }{(N^2 - I)/2}$	[5] $d_{i,j}$ = dissimilarity index between pyrogram ( $i$ ) and pyrogram ( $j$ ) $R_i^k$ = rank order of the $k$ th peak in pyrogram ( $i$ ) $R_j^k$ = rank order of the $k$ th peak in pyrogram ( $j$ ) $N$ = number of peaks in each pyrogram $I$ = 0 when $N$ is even; 1 when $N$ is odd	Compares the intensity rank order of corresponding peaks in each pyrogram. A value of $d_{i,j} = 0$ indicates a perfect match
Similarity coefficient	$S_{i,j} = \frac{\sum_{k=1}^N \left( \frac{I_i^k}{I_j^k} \right)}{N}$	[6] $S_{i,j}$ = similarity coefficient between pyrogram ( $i$ ) and pyrogram ( $j$ ) $I_i^k$ = normalized intensity of the $k$ th peak in pyrogram ( $i$ ) $I_j^k$ = corresponding intensity in pyrogram ( $j$ ) The quotient is arranged such that the more intense peak is the denominator ( $I_j^k > I_i^k$ ) $N$ = number of unique peaks in the programs	A value of 1.0 signifies a perfect match. Values in excess of 0.84, based upon an analysis of 13–15 peaks selected from the overall pyrogram, have provided an acceptable degree of similarity for microorganisms

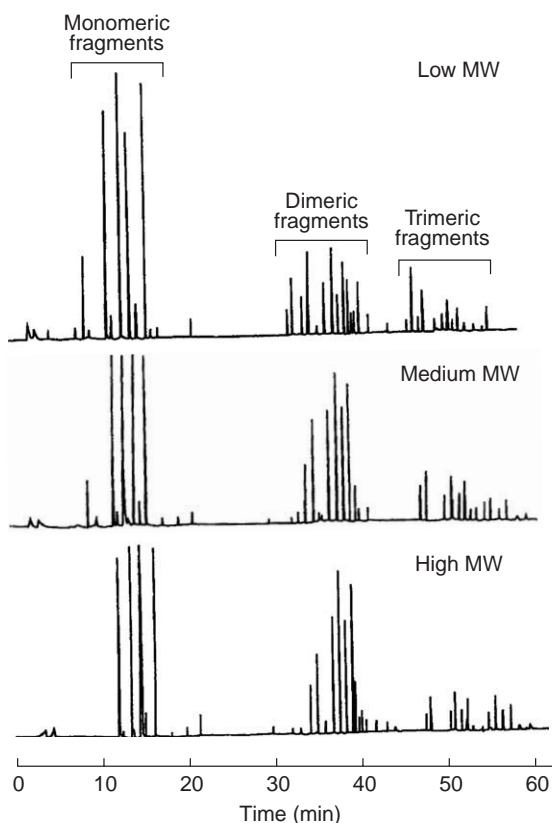


Euclidean distance	$D_{i,j} = \left( \sum_{k=1}^N (I_i^k - I_j^k)^2 \right)^{1/2} \quad [7]$	<p><math>D_{i,j}</math> = the Euclidean distance between pyrogram (<math>i</math>) and pyrogram (<math>j</math>)</p> <p><math>I_i^k</math> = intensity of the <math>k</math>th peak in pyrogram (<math>i</math>)</p> <p><math>I_j^k</math> = intensity of the <math>k</math>th peak in pyrogram (<math>j</math>)</p> <p><math>N</math> = number of peaks in each pyrogram (dimensionality)</p>	<p>Calculates the dissimilarity between two pyrograms by assessing the difference as a distance in <math>N</math>-dimensional space</p> <p>Weighting functions enhance the utility of this approach</p>
FIT factor	$F_{i,j} = 10001 - \frac{\sum_{k=1}^N (I_i^k - I_j^k)^2}{\sum_{k=1}^N ((I_i^k)^2 + (I_j^k)^2)} \quad [8]$	<p><math>F_{i,j}</math> is the FIT factor between pyrograms (<math>i</math>) and (<math>j</math>)</p> <p><math>I_i^k</math> = intensity of the <math>k</math>th peak in pyrogram (<math>i</math>)</p> <p><math>I_j^k</math> = corresponding intensity in pyrogram (<math>j</math>)</p> <p><math>N</math> = number of peaks selected from each pyrogram</p>	<p>A perfect match is indicated by <math>F_{i,j} = 1000</math></p>
Matching coefficient	$M_{i,j} = 100 \sum_{j=1}^N \left( 1 - \frac{I_i^L - I_j^U}{I_i^L + I_j^U} \right) \quad [9]$	<p><math>M_{i,j}</math> = matching coefficient between pyrograms (<math>i</math>) and (<math>j</math>)</p> <p><math>I_i^L</math> = intensities of the peaks in the library spectra</p> <p><math>I_j^U</math> = intensities of the corresponding peaks in the unknown spectrum</p> <p><math>N</math> = number of peaks in the library pyrogram</p>	<p>Proposed for the fingerprint comparison of unknown samples with pyrograms held in a library</p> <p>A perfect match is indicated by <math>M_{i,j} = 100</math></p>
Probability of error	$P_{n,m}^N = \frac{(n - N)!}{n!m!} \quad [10]$	<p><math>P_{n,m}^N</math> = probability of two samples being considered identical when they are not</p> <p><math>n</math> = number of strips along abscissa (temperature)</p> <p><math>m</math> = number of strips up ordinate (intensity %)</p> <p><math>N</math> = number of peaks compared and matching in two pyrograms</p>	<p><math>m</math> and <math>n</math> are determined by the reproducibility of intensity and temperature measurements and are <math>\pm 3\sigma</math> (standard deviation <math>\sigma_1, \sigma_T</math>) wide</p> <p><math>m = 100/6\sigma_1 (\approx 5)</math></p> <p><math>n = (T_2 - T_1)/6\sigma_T (\approx 25-30)</math></p>

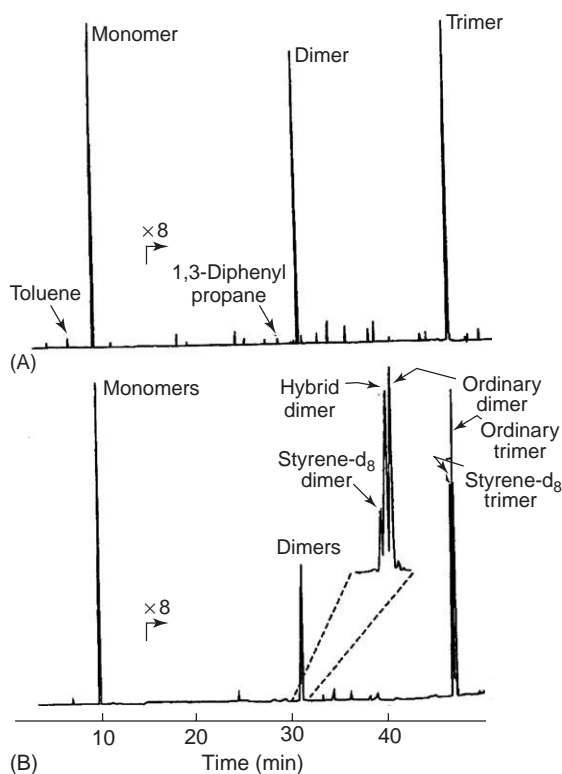
---



**Figure 8** Typical pyrolytic fragmentation reactions of synthetic polymers: (A) depolymerization; (B) side-chain elimination; (C) random scission.



**Figure 9** Pyrolysis of different molecular mass (MW) fractions of poly(2,6-dimethyl-1,4-phenylene ether). (Reproduced with permission from Usami T, Keitoku F, Ohtani H, and Tsuge S (1992) *Polymer* 33: 3024–3030; © Elsevier.)



**Figure 10** Pyrograms of (A) polystyrene and (B) poly(styrene-*b*-styrene- $d_8$ ) at 500°C, dimer intensities 14:45:41%. (Reproduced with permission from Ohtani H, Yuyama T, Tsuge S, Plage B, and Schulten H-R (1990) *European Polymer Journal* 26: 893–899; © Elsevier.)

The areas may be obtained by subjecting a single sample to a series of pyrolyses at a fixed temperature. This enables the cumulative degradation versus time to be obtained, but aggregated errors and thermal

conditioning are possible problems. A better method involves heating each of a series of samples twice only. The initial pyrolysis, for a period  $t$  that yields  $A_t$ , is followed by a second pyrolysis that completes

the degradation and generates an area  $A$  such that  $A_{\infty} = A_t + A$ . Alternatively, a series of replicate samples may be subjected to increasing pyrolysis times sufficient to give a series of  $A_t$  values and  $A_{\infty}$ .

Kinetic schemes for the degradation of various polymers may be derived. For example, depolymerization involves chain end ( $k_i$ ) or random scission ( $k'_i$ ) initiation; monomer depropagation ( $k_d$ ); and complete unzipping, first- ( $k'_t$ ) or second- ( $k_t$ ) order termination, with the observed rate constant being related to the degree of polymerization ( $D_0$ ). Thus, for a first-order termination reaction, the observed rate constant is as follows:

Chain-end initiation:  $k_{\text{obs}}$  versus  $1/D_0$  is linear through the origin (eqn [12]):

$$k_{\text{obs}} = \frac{k_d k_i}{k'_t D_0} \quad [12]$$

Random scission initiation:  $k_{\text{obs}}$  versus  $D_0$  is constant (eqn [13]):

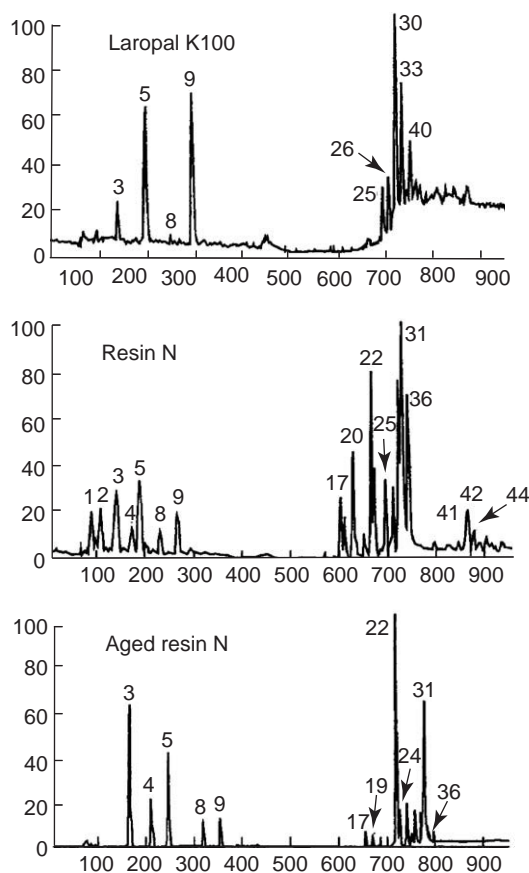
$$k_{\text{obs}} = \frac{2k'_i k_d}{k'_t} \quad [13]$$

A mixture of chain-end and random scission initiation:  $k_{\text{obs}}$  versus  $1/D_0$  is linear but not through the origin (eqn [14]):

$$k_{\text{obs}} = \frac{k_i k_d}{D_0 k'_t} + \frac{2k'_i k_d}{k'_t} \quad [14]$$

## Forensic Science

Forensic aspects are largely concerned with the fingerprint identification of samples (paint, fibers, adhesives) that may connect a suspect with the scene of a crime. Emphasis has been placed upon the reliability of comparisons and considerable data are available. Paint samples are especially important in automobile accidents and may be thermosetting alkyds (e.g., polyesters and glycerol), thermoplastic acrylics (e.g., poly(methyl methacrylate)), and thermosetting acrylics (e.g., styrene-acrylate copolymers). Each may be distinguished, and paints from different years, manufacturers, and even models may often be identified. For example, a characteristic component of alkyd resins is phthalic anhydride; other peaks include styrene modifications (styrene), toluene modifications (vinyl toluene), urethane modifications (toluene diisocyanate), epoxy esters (phenol), together with aliphatic residues (acrolein, methacrolein) and propenol (from glycerol or pentaerythritol). Related interests are concerned with the identification of paints and lacquers used in art, particularly to identify forgeries. For example, Figure 11 records pyrograms from cyclohexanone-formaldehyde resins used as varnishes.



**Figure 11** Pyrolysis of cyclohexanone-formaldehyde resins at 650°C for 5 s showing differences due to source and ageing. Identified peaks are: 1 = cyclohexene; 3 = cyclohexanone; 4 = 2-methylcyclohexanone; 5 = 2-methylenecyclohexanone; 8 = 2-methyl-6-methylenecyclohexanone; 9 = 2,6-dimethylenecyclohexanone; peaks 22–44 correspond to products with two or three cyclohexane residues. (From Mestdagh H, Rolando C, and Sablier M (1992) *Analytical Chemistry* 64: 2221–2226.)

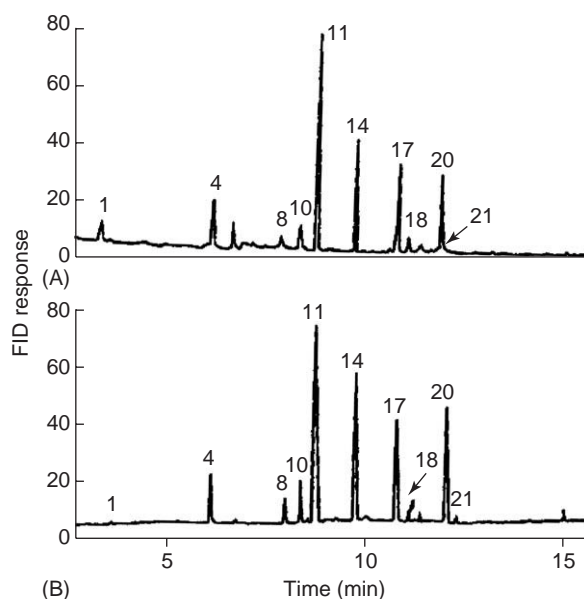
Characteristic pyrograms are obtained from Laropal K100, Resin N, and an aged sample of Resin N sufficient to discriminate between them. Documents may be assayed for the identity of toner materials in photocopies, inks, coatings, and the adhesives used on labels. Other forensic investigations involve tapes used in the fabrication of bombs or to bind victims.

The technique is also proving useful in archaeology where, for example, charred food remains have been characterized, showing clear chemical differences in their origin; and it has been used to show that external encrustation (patina) on supposedly authentic ancient Egyptian vessels (Thutmosis III, 1475–1425 BC) was identical (i.e., artificially added) and was similar to that produced by fatty acid salts. Samples of amber have been studied, both to discriminate between authentic and forged materials and to help locate the geographic origin of a particular sample.

### Taxonomy

Pyrolysis has found widespread application in the identification of microorganisms, although detailed data handling is often necessary. In particular, the technique offers a standard analytical procedure for a wide range of samples that can be completed in a much shorter time than a battery of morphological, biochemical, and serological tests. Pyrograms from microorganisms reflect the pyrolysis of the major groups of biochemicals – proteins (e.g., characterized by elimination, yielding nitriles, and side-chain stripping to give benzene, pyrrole, and indole derivatives), carbohydrates (e.g., dehydration to furan and pyran derivatives), and lipids (e.g., elimination, scission, and decarboxylation to fatty acids, esters, acrolein, and homologous olefins and paraffins). In addition to pyrolysis control, culture of the microorganisms and the collection of a representative sample, free from medium, also require standardization. Over 200 components are usually produced and high-resolution Py-GC is essential.

Pyrograms usually display the same products in different relative yields but, in some cases, unique fragments occur. These may readily be investigated using mass fragmentography. For example, streptococci have been distinguished by the presence of a unique component (possibly derived from glucitol phosphate), monitored exclusively at  $m/z$  86, due to a specific antigen. Additionally, some microorganisms produce poly( $\beta$ -hydroxy-butyrate), a biodegradable polymer now gaining acceptance in drug delivery. This polymer yields *trans*-2-butenic acid, together with smaller amounts of the *cis*-isomer and 3-butenic acid, which serves as a reliable marker for Legionellaceae and Bacilli on pyrolysis. In contrast, Enterobacteriaceae showed low levels of this compound, which, coincidentally, also appears at  $m/z$  86, but at much shorter retention times. More typical is the characterization of *Ascomycetes* (yeasts) where 28 strains from 13 genera could be distinguished, on the basis of many qualitative and quantitative differences, using 14 peaks, within 49 h. This compares favorably with conventional analyses, which require 7–10 days. An alternative procedure for the identification of microorganisms has used the profile of fatty acids volatilized as the methyl esters. An extension of Py-GC involves on-column methylation of whole organisms and GC analysis of the resultant fatty acid methyl esters. The utility of this simple and rapid procedure is compared, in Figure 12, with standard extraction and derivatization. Py-GC has also been used to identify more complex samples. For example, winter wheat cultivars (*Triticum aestivum*) may be distinguished by pattern-recognition techniques, but seasonal growth conditions were as important as variety in determining pyrogram variability.



**Figure 12** Fatty acid methyl ester profiles from *Escherichia coli* by (A) Py-GC with trimethylanilinium hydroxide at 500°C for 10 s and (B) extraction and derivatization. Major peaks are: 4,  $C_{14:0}$ ; 11,  $C_{16:0}$ ; 14, cyclopropyl- $C_{17}$ ; 17,  $C_{18:1}$ ; 20, isopropyl- $C_{19}$ . (Reproduced with permission from Holzer G, Bourn TF, and Bertsch W (1989) *Journal of Chromatography* 468: 181–190; © Elsevier.)

### Organic Geopolymers

It is estimated that there are some  $6 \times 10^{14}$  tons of organic matter in the earth's crust, the bulk of which is in Phanerozoic (<570 million years old) sedimentary rocks. Terrestrial organic polymers are of interest in the characterization, assessment, and maturation processes that result in the conversion of vegetation into humic substances and fossil fuels. Intractable, condensed aromatic structures, which may be present in low amounts, contribute to the structure of these substances and Py-GC dramatically extends the analytical options. The polysaccharide and lignified components of cells, boosted by the products of microbial activity, contribute strongly to the composition of the organic matter, with lignins providing abundant phenolic derivatives in the pyrograms of partially decayed humus. Various soil horizons, separated into fractions, provide characteristic oxygen- and nitrogen-containing components derived from biomacromolecules and over 300 have been described. Products such as homologous fatty acids, residues from proteins, polysaccharides, and lignin allow samples to be characterized and classified. The analysis of kerogen and petroleum source rock may be undertaken in a similar way and pyrograms from such samples are characterized by high yields of hydrocarbons. Samples range from hydrogen-rich types that yield low-mass linear paraffins and olefins, to the hydrogen-depleted fractions based on

condensed hydrocarbons that yield many aromatic products. Sequential pyrolysis also provides information on distribution and maturation. Almost 150 compounds have been identified from kerogen samples, including those from lignin, lipid, and condensed origin. Principal components analysis on 37 of these components allows prediction of the source rock hydrogen index (HI<sub>0</sub>) from a single Py-GC analysis. The rank of coals may also be assessed by Py-GC, for example, by studying the oxygen functionalization of monocyclic derivatives, while the formation of a series of long-chain *n*-alkanes is indicative of a thermoplastic coal. Pyrograms from coal macerals (which are petrological unit of a geological sample) consist mainly of alkylbenzenes, alkyl-naphthalenes, alkylphenols, alk-1-enes, and alkanes with quantitative differences depending upon precursor material and maturation. The development of life-detection systems for extraterrestrial exploration has also depended upon organic geochemical analysis. Even desert sands show abundant evidence of biomacromolecules and pyrograms are characterized by olefins, nitriles, benzenes, phenols, indoles, and furans. The lack of any such compounds from Lunar and Martian samples indicates their lifelessness.

## Limitations

Although the applications described here have concerned diverse macromolecular samples, pyrolysis is also of utility in the analysis of smaller molecules, particularly quaternary ammonium compounds, which undergo quantitative thermal fragmentation to volatile products. For example, the specificity and sensitivity afforded by Py-GC/MS with mass fragmentography renders the technique suitable for the analysis of endogenous neurotransmitters such as acetylcholine at the 1–2 pmol level. However, in all cases, analytical throughput is limited by the retention times of the pyrolysis products, a condition that is particularly severe in taxonomic applications. Automation allows extended use, but direct Py-MS

offers substantial reductions in analytical time, although a single pyrolysis product may give rise to several ions, a single ion may be derived from several pyrolysis products, and identical  $M^{+•}$  values result from isomers that may be resolved by Py-GC. Direct Py-MS also allows detection of larger or more polar pyrolysis products that are unlikely to be efficiently chromatographed. This may provide a substantial increase in structural information for complex samples of biological origin.

**See also:** **Coal and Coke.** **Forensic Sciences:** Paints, Varnishes, and Lacquers. **Gas Chromatography:** Multidimensional Techniques; Instrumentation; Detectors; Mass Spectrometry. **Geochemistry:** Inorganic; Soil, Major Inorganic Components; Soil, Minor Inorganic Components; Soil, Organic Components. **Laser-Based Techniques.** **Polymers:** Synthetic. **Quality Assurance:** Internal Standards.

## Further Reading

- The *Journal of Analytical and Applied Pyrolysis* (Elsevier) is dedicated to this field and, additionally, it publishes the proceedings of the biennial International Conference on Pyrolysis (from 1984). Formerly, proceedings were published as separate volumes as included below.
- Hammond T and Lehrle RS (1989) Pyrolysis-gas chromatography. In: Allen O and Bevington JC (eds.) *Pyrolysis-Gas Chromatography*, pp. 589–601. Oxford: Pergamon.
- Irwin WJ (1982) *Analytical Pyrolysis: A Comprehensive Guide*. New York: Dekker.
- Jones CER and Cramers CA (eds.) (1977) *Analytical Pyrolysis. Proceedings of the Third International Symposium on Analytical Pyrolysis*. Amsterdam: Elsevier.
- Liebman SA and Levy EJ (1985) *Pyrolysis and GC in Polymer Analysis*. New York: Dekker.
- May RW, Pearson EF, and Scothern D (1977) *Pyrolysis-Gas Chromatography*. London: Chemical Society.
- Moldoveanu SC (1998) *Analytical Pyrolysis of Natural Organic Polymers*. Amsterdam: Elsevier.
- Voorhees KJ (1984) *Analytical Pyrolysis: Techniques and Applications. Proceedings of the Fifth International Symposium on Analytical Pyrolysis*. London: Butterworths.
- Wampler TP (ed.) (1995) *Applied Pyrolysis Handbook*. New York: Dekker.

## Detectors

**C F Poole**, Wayne State University, Detroit, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Many techniques have been described for the detection of compounds separated by gas chromatography.

Space limitations, however, allow only those detectors that are widely used and have reached commercial maturity to be discussed in this article. The important spectroscopic detectors (e.g., mass and infrared spectrometry) used for structure elucidation as well as detection are discussed in separate articles.

The principal methods of detection can be categorized as ionization, bulk physical property, optical, and



condensed hydrocarbons that yield many aromatic products. Sequential pyrolysis also provides information on distribution and maturation. Almost 150 compounds have been identified from kerogen samples, including those from lignin, lipid, and condensed origin. Principal components analysis on 37 of these components allows prediction of the source rock hydrogen index (HI<sub>0</sub>) from a single Py-GC analysis. The rank of coals may also be assessed by Py-GC, for example, by studying the oxygen functionalization of monocyclic derivatives, while the formation of a series of long-chain *n*-alkanes is indicative of a thermoplastic coal. Pyrograms from coal macerals (which are petrological unit of a geological sample) consist mainly of alkylbenzenes, alkyl-naphthalenes, alkylphenols, alk-1-enes, and alkanes with quantitative differences depending upon precursor material and maturation. The development of life-detection systems for extraterrestrial exploration has also depended upon organic geochemical analysis. Even desert sands show abundant evidence of biomacromolecules and pyrograms are characterized by olefins, nitriles, benzenes, phenols, indoles, and furans. The lack of any such compounds from Lunar and Martian samples indicates their lifelessness.

## Limitations

Although the applications described here have concerned diverse macromolecular samples, pyrolysis is also of utility in the analysis of smaller molecules, particularly quaternary ammonium compounds, which undergo quantitative thermal fragmentation to volatile products. For example, the specificity and sensitivity afforded by Py-GC/MS with mass fragmentography renders the technique suitable for the analysis of endogenous neurotransmitters such as acetylcholine at the 1–2 pmol level. However, in all cases, analytical throughput is limited by the retention times of the pyrolysis products, a condition that is particularly severe in taxonomic applications. Automation allows extended use, but direct Py-MS

offers substantial reductions in analytical time, although a single pyrolysis product may give rise to several ions, a single ion may be derived from several pyrolysis products, and identical  $M^{+•}$  values result from isomers that may be resolved by Py-GC. Direct Py-MS also allows detection of larger or more polar pyrolysis products that are unlikely to be efficiently chromatographed. This may provide a substantial increase in structural information for complex samples of biological origin.

**See also:** **Coal and Coke. Forensic Sciences:** Paints, Varnishes, and Lacquers. **Gas Chromatography:** Multidimensional Techniques; Instrumentation; Detectors; Mass Spectrometry. **Geochemistry:** Inorganic; Soil, Major Inorganic Components; Soil, Minor Inorganic Components; Soil, Organic Components. **Laser-Based Techniques. Polymers:** Synthetic. **Quality Assurance:** Internal Standards.

## Further Reading

- The *Journal of Analytical and Applied Pyrolysis* (Elsevier) is dedicated to this field and, additionally, it publishes the proceedings of the biennial International Conference on Pyrolysis (from 1984). Formerly, proceedings were published as separate volumes as included below.
- Hammond T and Lehrle RS (1989) Pyrolysis-gas chromatography. In: Allen O and Bevington JC (eds.) *Pyrolysis-Gas Chromatography*, pp. 589–601. Oxford: Pergamon.
- Irwin WJ (1982) *Analytical Pyrolysis: A Comprehensive Guide*. New York: Dekker.
- Jones CER and Cramers CA (eds.) (1977) *Analytical Pyrolysis. Proceedings of the Third International Symposium on Analytical Pyrolysis*. Amsterdam: Elsevier.
- Liebman SA and Levy EJ (1985) *Pyrolysis and GC in Polymer Analysis*. New York: Dekker.
- May RW, Pearson EF, and Scothern D (1977) *Pyrolysis-Gas Chromatography*. London: Chemical Society.
- Moldoveanu SC (1998) *Analytical Pyrolysis of Natural Organic Polymers*. Amsterdam: Elsevier.
- Voorhees KJ (1984) *Analytical Pyrolysis: Techniques and Applications. Proceedings of the Fifth International Symposium on Analytical Pyrolysis*. London: Butterworths.
- Wampler TP (ed.) (1995) *Applied Pyrolysis Handbook*. New York: Dekker.

## Detectors

**C F Poole**, Wayne State University, Detroit, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Many techniques have been described for the detection of compounds separated by gas chromatography.

Space limitations, however, allow only those detectors that are widely used and have reached commercial maturity to be discussed in this article. The important spectroscopic detectors (e.g., mass and infrared spectrometry) used for structure elucidation as well as detection are discussed in separate articles.

The principal methods of detection can be categorized as ionization, bulk physical property, optical, and



electrochemical based on the principle of the detection mechanism. Further division into universal, element-selective, and structure-selective detectors is possible based on the characteristics of the detector response. The flame ionization and thermal conductivity detectors (TCDs) respond to nearly all organic compounds and are examples of general or (near) universal detectors. Other detectors respond only to a particular heteroatom (e.g., flame photometric, thermionic ionization, or atomic emission detectors) and are element-selective. Or they respond only to a structural feature related to the bond energy of two or several atoms in a compound (e.g., electron capture and photoionization detectors) and are structure-selective. Element-selective and structure-selective detectors are used for target compound analysis, for example, the analysis of a specific drug in biological fluids or pesticides in an environmental extract, where the matrix is often complex and masks the presence of the target compounds when a universal detector is used.

Detectors can also be compared by figures of merit that define their signal characteristics in quantitative terms. The fundamental properties of general interest are sensitivity, limit of detection, operational response range, response time, and noise characteristics. Sensitivity is defined as the detector response per unit mass or concentration of a substance and is determined as the slope of the calibration curve for detectors with a linear response. A detector with a high sensitivity, corresponding to a larger slope, is better able to discriminate between small differences in sample amounts. Sensitivity is often confused with the limit of detection, the latter defined as the concentration or mass flow of a substance that produces a detector response equal to some multiple (typically three times) of the average detector noise. The dynamic range is defined as the range of sample amount for which a change in sample size induces a discernible change in the detector signal. For many, although not all detectors, the relationship between response and sample amount is linear over a wide range. It is the extent of this range that is generally of most interest for comparing detector performance.

## Ionization Detectors

Common carrier gases behave as near perfect insulators. In the absence of conduction by the gas molecules themselves, the increased conductivity due to the presence of very few charged species is easily measured, providing the low sample detection limits characteristic of ionization-based detectors. Common ionization detectors include the flame ionization detector (FID), thermionic ionization detector (TID), photoionization detector (PID), the electron-capture

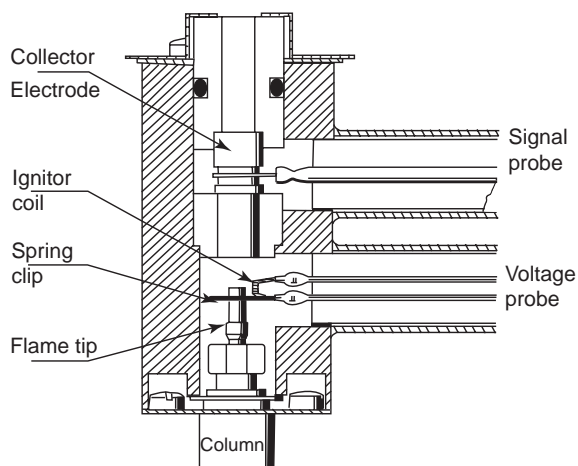
detector (ECD), and the helium ionization detector (HID). A different method of ion production is employed by each detector, but in all cases, the quantitative basis of the detector response is due to the change in ion current in the presence of sample vapors.

### Flame Ionization Detector

The popularity of the FID is explained by its near universal response to organic compounds, low detection limits, simple construction, fast response, and exceptional linear range. Only the fixed gases (e.g., He, Xe, H<sub>2</sub>, N<sub>2</sub>), certain nitrogen oxides (e.g., N<sub>2</sub>O, NO, etc.), compounds containing a single carbon atom bonded to oxygen or sulfur (e.g., CO<sub>2</sub>, CS<sub>2</sub>, etc.), inorganic gases (e.g., NH<sub>3</sub>, SO<sub>2</sub>, etc.), water, and formic acid fail to provide a significant response. Typical detection limits correspond to about  $10^{-12}$  g s<sup>-1</sup> with a linear response range of  $10^6$ .

The FID signal results from the combustion of organic compounds in a small hydrogen–air diffusion flame (Figure 1). The column carrier gas is mixed with hydrogen and burned at a narrow orifice in a chamber through which excess air is flowing. The ions produced in the flame are collected by application of a small voltage between the jet tip and a cylindrical electrode located a few millimeters above the flame. The small ion currents ( $10^{-14}$ – $10^{-5}$  A) are amplified by a precision electrometer. Detector performance is influenced primarily by the ratio of air-to-hydrogen-to-carrier (and make-up) gas flow rates, the choice of carrier gas, and detector geometry. The optimum response plateau for detector gas ratios is fairly broad, however, permitting operation over a rather wide range of gas flow rates without incurring a large penalty in diminished response.

On account of the low thermal energy of flames it is believed that ionization results from chemical



**Figure 1** Cross-sectional view of a flame ionization detector.

**Table 1** Contributions of structure to the response of the flame ionization detector

Atom	Type	Effective carbon number
C	Aliphatic	1.0
C	Aromatic	1.0
C	Olefinic	0.95
C	Acetylenic	1.30
C	Carbonyl	0
C	Carboxyl	0
C	Nitrile	0.3
O	Ether	−1.0
O	Primary alcohol	−0.5
O	Secondary alcohol	−0.75
O	Tertiary alcohol	−0.25
N	In amines	Similar to O in alcohols
Cl	On olefinic C	0.05
Cl	Two or more on aliphatic C	−0.12 per Cl

ionization process, such as the one shown below:



Degradation of organic compounds in the flame likely results from attack by hydrogen atoms causing fission of carbon–carbon bonds. Fission of unsaturated bonds proceeds after hydrogenation. The hydrocarbon radicals produced by fission are unstable and undergo a cascade of fast fractionation and hydrogenation reactions such that all carbon atoms are eventually converted to methane. Consequently, each carbon atom capable of hydrogenation yields the same signal, and the overall FID response to each substance is proportional to the sum of these ‘effective’ carbon atoms. The FID response is highest for hydrocarbons, being proportional to the number of carbon atoms, while substances containing oxygen, nitrogen, sulfur, or halogens yield lower responses, depending on the characteristics of the carbon–heteroatom bond and the electron affinity of the combustion products. The lower response is due to competition between hydrogenation of the carbon–heteroatom bond and hydrogen abstraction with formation of neutral species (e.g., CO, HCN), which are poorly ionized in the flame. The effective carbon number for a particular compound can be estimated by summation of the various carbon and heteroatom contributions (Table 1), which in turn can be used to predict relative response factors with reasonable accuracy.

### Thermionic Ionization Detector

All modern versions of the TID, also known as the nitrogen–phosphorus or NPD detector, employ a ceramic or glass bead or cylinder doped with an alkali metal salt, molded onto an electrical heater wire, heated to 400–800°C, as the thermionic source. Carrier gas is combined with hydrogen at the detector base and flows through a jet where it is mixed

with air. The flow of detector gases is insufficient to establish a flame, but sufficient to maintain a plasma localized around the thermionic source. The thermionic source is located immediately above the jet tip and the cylindrical collector electrode either surrounds the source or is located immediately above it, depending on the detector design. A voltage set between the collector electrode and jet tip allows collection of (usually) negative ions and the ion current is measured by an electrometer.

Two different mechanisms have been proposed to explain the element-selective response of the TID to nitrogen- and phosphorus-containing compounds. These mechanisms differ principally in whether the interaction between the alkali metal atoms and organic fragments occurs as a homogeneous reaction in the gas phase or is purely a surface phenomenon. According to the gas-phase ionization theory, alkali metal atoms are vaporized from the hot source after acceptance of an electron from the heating wire or plasma. While in the boundary layer near the surface of the thermionic source, the alkali metal atoms are excited and ionized by collision with plasma particles. The negatively charged source rapidly recaptures the positively charged metal ions. This cyclic process results in a steady-state population of metal atoms in the boundary layer. If a process that results in ionization of metal atoms disturbs this equilibrium, then more metal atoms will leave the source to restore the equilibrium accompanied by an increase in the ion current. The selectivity of the detector results from the fact that only those radicals with electron affinities equal to or greater than the ionization potential of the metal atoms will contribute to the ion current. Among the many fragments generated by the decomposition of organic compounds in the plasma, only the  $\text{CN}^\bullet$ ,  $\text{PO}^\bullet$ , and  $\text{PO}_2^\bullet$  radicals meet this criterion for the alkali metals.

The surface ionization model assumes that the principal role of the alkali metal in the source is to lower the work function of the surface. Electro-negative decomposition products from phosphorus- or nitrogen-containing compounds are then selectively ionized by extracting an electron from the surface of the thermionic source. These negative ions are responsible for the increase in observed ion current measured at the collector electrode. To account for the influence of experimental variables on the detector response, it is assumed that a combination of the heat from the source and the reaction of sample molecules within the boundary layer are responsible for sample decomposition. The active decomposition products must be similar and largely independent of molecular structure for nitrogen- and phosphorus-containing compounds to account for the narrow range of response factors.

The response of the detector to compounds containing elements other than nitrogen and phosphorus depends on the thermionic source temperature, the work function of the source surface, and the chemical composition of the gas environment surrounding the source. With nitrogen as the plasma gas and a thermionic source with a low work function (high cesium content), the detector can be made selective to compounds containing functional groups of high electron affinity (e.g., nitro and thiol groups). With a relatively low source temperature (400–500°C) and air or oxygen as the plasma gas enhanced selectivity toward halogen-containing compounds is obtained.

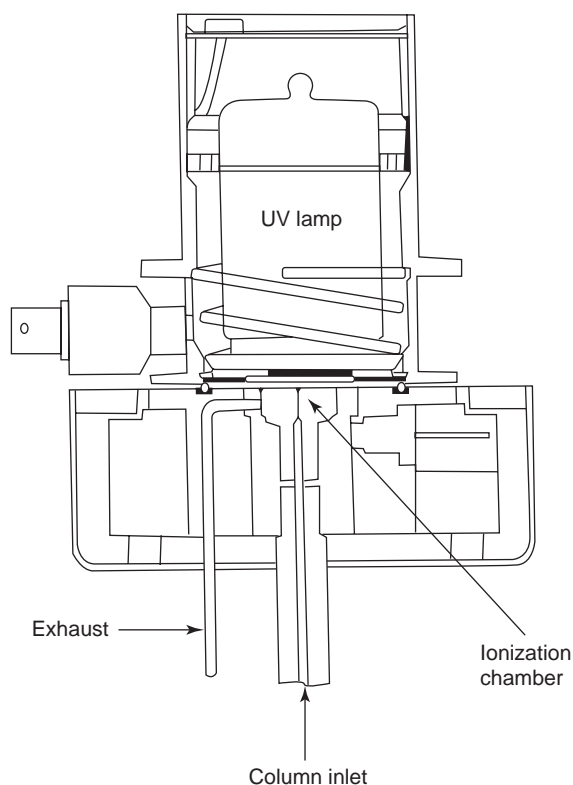
Although details of the detector response mechanism are not certain, the TID is not a particularly difficult detector to use and affords reliable results. Typical detection limits for nitrogen-containing compounds are about  $5 \times 10^{-14}$  to  $2 \times 10^{-13}$  g N/s and for phosphorus-containing compounds about  $1 \times 10^{-14}$  to  $2 \times 10^{-13}$  g P/s. The selectivity against carbon is about  $10^3$  to  $10^5$  g C/g N and  $10^4$  to  $5 \times 10^5$  g C/g P. Discrimination against phosphorus in the nitrogen mode is poor at 0.1–0.5 g P/g N. The linear range is about  $10^4$ – $10^5$ . The TID is widely used in environmental and biomedical research for determining pesticides residues and drugs as well as for obtaining element-selective profiles, where its high sensitivity and selectivity are useful in minimizing sample preparation requirements.

### Photoionization Detector

Selective ionization of most organic compounds in the gas phase is possible by absorption of photons with energy either close to or greater than the ionization potential of the compounds. For typical organic compounds this requires photons in the far ultraviolet of about 5–20 eV. A cross-sectional view of a typical PID is shown in **Figure 2**. The photon source is a compact discharge lamp, containing an inert gas or gas mixture at low pressure, which emits monochromatic light of a specific energy, depending on the choice of fill gases and window material. Sources of different nominal energies (8.3, 9.5, 10.2, 10.9, and 11.7 eV) provide

for the possibility of selective ionization of organic compounds with the 10.2 eV source used for general applications (**Table 2**). An optically transparent window made of a metal fluoride separates the discharge compartment from the ionization chamber. The carrier gas passes through the thermostatted ionization chamber and between two electrodes, positioned at opposite ends of the chamber. An electric field is applied between the electrodes to collect the ions formed (or electrons, if preferred) and the current amplified by an electrometer.

An alternative approach is based on an open-source design. Photons are generated by an atmospheric pressure helium discharge that can be doped with other inert gases to provide a range of photon energies. The detector is constructed from a quartz



**Figure 2** Cross-sectional view of a photoionization detector.

**Table 2** Selectivity of the photoionization detector to different compound classes

Photon energy (eV)		Fill gas	Compounds ionized
Nominal	Actual		
8.3	8.44	Xe	Compounds with low ionization potential such as polycyclic aromatic compounds Simple aromatic compounds in the presence of alkanes, mercaptans in H <sub>2</sub> S, amines in ammonia
9.5	8.44 (97.6%)	Xe	
	9.57 (2.1%)		
	10.4 (0.18%)		
10.2	10.03 (82.9%)	Kr	Most organic compounds except permanent gases, C <sub>1</sub> –C <sub>4</sub> alkanes, methanol, acetonitrile, and chloromethanes
	10.64 (17.1%)		
11.7	11.62 (71.8%)	Ar	Low molecular weight compounds with high ionization potentials (e.g., formaldehyde, ethane, chloromethanes, acetylene, etc.)
	11.82 (26.2%)		

tube divided into two zones with electrodes located at the end flanges and middle section. A high-voltage pulse is used to initiate the discharge in the top section and a bias potential on the middle electrode to direct electrons to the collector electrode housed in the bottom section. The carrier gas is introduced at a point close to the middle electrode with photon ionization occurring in the region between the middle electrode and the collector electrode. Since there is no barrier between the source and ionization chamber the pulse discharge PID can be operated at lower wavelengths (13.5–17.8 eV) than the conventional PID design functioning as a near universal detector.

The response of the PID can be described by the relationship

$$i = IF\eta\sigma NL[AB]$$

where  $i$  is the detector ion current,  $I$  the initial photon flux,  $F$  the Faraday constant,  $\eta$  the photoionization efficiency,  $\sigma$  the absorption cross-section,  $N$  Avogadro's number,  $L$  the path length, and  $[AB]$  the concentration of an ionizable substance. Thus, for a particular detector and source, the PID signal is proportional to the ionization yield, absorption cross-section, and molar concentration of the analyte. The product ( $\eta\sigma$ ) is the photoionization cross-section, which expresses both the probability that a molecule will absorb a photon and the probability that the excited state will ionize. A number of secondary processes involving collisions with carrier gas molecules and electron-capturing impurities compete with the ionization mechanism reducing the detector signal. Detector design and control of the operating conditions minimize these effects.

The PID is inexpensive, of rugged construction, and easy to operate. Due to the low efficiency of photoionization (<0.1%) the detector is classified as nondestructive. Since no combustion gases are required the PID can be used in environments where combustion gases are considered hazardous or in portable instruments, where the additional weight of several gas bottles is undesirable. For compounds with favorable ionization properties the PID is 5–50 times more sensitive than the FID with a comparable linear range ( $\approx 10^7$ ). For individual compounds detector response factors vary over a wide range allowing the PID to be used as a selective detector for some applications. Major uses include the analysis of volatile organic compounds from environmental samples and in field-portable gas chromatographs.

### Electron-Capture Detector

The ECD owes its popularity to its unsurpassed sensitivity to a wide range of environmentally important and biologically active compounds. Examples of

general applications include the determination of pesticides and industrial chemicals in the environment, assessment of the fate of ozone-depleting chemicals in the upper atmosphere, and the determination of drugs and hormones in biological fluids. To maintain optimum detector performance some experience with the operation of the detector is required as well as familiarity with the causes and effect of detector contamination.

The ECD is available in several different designs dominated by two alternative methods for generating thermal electrons responsible for the detector's operating characteristics. From inception, high-energy  $\beta$  electrons generated by the decay of a radioisotope have been used as the primary source of ionizing radiation. Either  $^{63}\text{Ni}$  or  $^3\text{H}$  radioisotope sources supported by a metal foil are used in commercial detectors. These  $\beta$ -particle emitters produce a large number of secondary electrons through multiple collisions with carrier gas molecules forming a plasma of thermal electrons (mean energies 0.02–0.05 eV), radicals, and positive ions. Radioisotope-based detectors require periodic wipe tests to ensure safety as well as compliance with regulations concerning storage, use, and transport of radioactive materials. Regulatory constraints provided the impetus for the development of a non-radioactive detector. The non-radioactive ECD uses high-energy photons from a pulsed discharge in pure helium to ionize a support gas added downstream of the discharge. A plasma of thermal electrons and positive ions is formed in the ionization chamber with properties similar to the plasma generated by the radioisotope-source ECD.

Application of a fixed or pulsed potential to the ionization chamber allows collection of the thermal electrons establishing the standing (or baseline) current. When an electron-capturing compound enters the ionization chamber, thermal electrons are removed by formation of negative ions. The increased rate of neutralization of these ions by positive ions, or their reduced drift velocity during collection of the thermal electrons, is responsible for the detector signal. In contrast to the other ionization detectors a decrease in the detector standing current proportional to solute concentration is measured rather than an increase in the number of ions or electrons produced.

The majority of detectors with radioisotope sources in current use are based on either the coaxial cylinder or asymmetric configurations, as shown in **Figure 3**. The low specific activity of the  $^{63}\text{Ni}$  source requires a relatively larger source area to provide a suitable background current that is easier to accommodate in these designs. Virtually all contemporary detectors employ pulse-sampling techniques to collect the thermal electrons based on the variable frequency constant



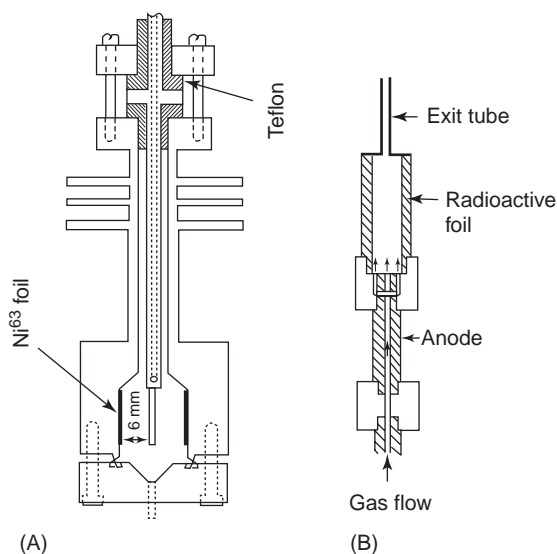
current mode. In this case, the pulse frequency is varied throughout the separation to maintain the cell current at a fixed reference value. The detector signal is a voltage proportional to the pulse frequency. The two principal advantages of this method are an increased linear response range of  $10^4$ – $10^5$  and reduced detector disturbance from column contamination except for a few compounds with ultrafast electron attachment rate constants, which exhibit a nonlinear response, and are better measured with a constant pulse frequency.

Oxygen-free nitrogen is the most common carrier gas for packed columns and hydrogen or helium for open tubular columns. Argon–methane or nitrogen is then added as make-up gas at the column exit to minimize extra column peak broadening due to the relatively large effective detector volume. Pure argon and helium are unsuitable make-up gases as they readily form metastable species, which can transfer

their excitation energy by collision with sample vapors, resulting in undesirable ionization effects (Penning reaction). The addition of 5–10% of methane to argon removes these ions by deactivating collisions as quickly as they are formed. Oxygen and water vapor traps should be used to purify all gases.

The ECD is a structure-selective detector with a response range that covers about seven orders of magnitude (Table 3). The greatest response is observed for compounds containing halogen atoms ( $I > Br > Cl \gg F$  increasing synergistically for multiple substitution on the same carbon atom) or nitro groups, to organometallic compounds, and to compounds containing conjugated electrophores. This latter group is the least well defined, and is comprised of compounds containing two or more weakly electron-capturing groups, connected by some specific bridge that promotes a synergistic interaction between the two electron-capturing groups. Examples of compounds containing conjugated electrophores include conjugated carbonyl compounds (benzophenones, quinones, phthalate esters, coumarins), some polycyclic aromatic hydrocarbons, some sulfonamides, and certain steroids.

The effect of temperature on the detector response is frequently overlooked. The detector response can vary by as much as 100–1000-fold for a  $100^\circ\text{C}$  change in detector temperature. In most cases electron capture occurs by either of two mechanisms. By non-dissociative electron capture resulting in the formation of a stable molecular ion, or by dissociative electron capture with formation of a molecular ion in an excited state that instantaneously decomposes with elimination of a negative ion fragment. Nondissociative electron capture is generally associated with compounds containing a conjugated electrophore and dissociative electron capture with halogen-containing compounds. An increase in detector temperature favors the populating of vibrationally excited states



**Figure 3** Cross-sectional view of a coaxial cylinder (A) and asymmetric (B) electron-capture detector.

**Table 3** Relative response of the electron-capture detector to various organic compounds

General organic compounds	Relative response	Fluorocarbon compounds	Relative response
Benzene	0.06	$\text{CF}_3\text{CF}_2\text{CF}_3$	1.0
Acetone	0.50	$\text{CF}_3\text{Cl}$	3.3
Di- <i>n</i> -butyl ether	0.60	$\text{CF}_2=\text{CFCl}$	$1.0 \times 10^2$
Methylbutyrate	0.90	$\text{CF}_3\text{CF}_2\text{Cl}$	$1.7 \times 10^2$
1-Butanol	1.00	$\text{CF}_2=\text{CCl}_2$	$6.7 \times 10^2$
1-Chlorobutane	1.00	$\text{CF}_2\text{Cl}_2$	$3.0 \times 10^4$
1,4-Dichlorobutane	15.00	$\text{CHCl}_3$	$3.3 \times 10^4$
Chlorobenzene	75.00	$\text{CHCl}=\text{CCl}_2$	$6.7 \times 10^4$
1,1-Dichlorobutane	$1.1 \times 10^2$	$\text{CF}_3\text{Br}$	$8.7 \times 10^4$
1-Bromobutane	$2.8 \times 10^2$	$\text{CF}_2\text{ClCFCl}_2$	$1.6 \times 10^5$
Bromobenzene	$4.5 \times 10^2$	$\text{CF}_3\text{CHClBr}$	$4.0 \times 10^5$
Chloroform	$6.0 \times 10^4$	$\text{CF}_3\text{CF}_2\text{CF}_2\text{I}$	$6.0 \times 10^5$
1-Iodobutane	$9.0 \times 10^4$	$\text{CF}_2\text{BrCF}_2\text{Br}$	$7.7 \times 10^5$
Carbon tetrachloride	$4.0 \times 10^5$	$\text{CFCl}_3$	$1.2 \times 10^6$

and thus the dissociative mechanism. Conversely, nondissociative electron capture may be destroyed if the rate constant for thermal detachment is too fast, and a higher detector response is usually observed at low detector temperatures. The optimum detector response for practical work is usually obtained at either the maximum temperature recommended for the radioisotope source for compounds with a dissociative mechanism for electron capture, or at the lowest practical operating temperature for compounds with a nondissociative mechanism of electron capture.

### Helium Ionization Detector

The HID is a universal and ultrasensitive detector used primarily for the analysis of permanent gases and some volatile organic compounds that have a poor response to the FID and are present in too low a concentration for detection with a TCD. Typical detection limits are around  $10^{-13} \text{ g s}^{-1}$  with a linear range of about  $10^4$ . Many features of the design of the HID are similar to those of the pulsed discharge ECD. Metastable helium ions thought to be primarily responsible for ionization of organic compounds are produced by bombardment of helium atoms with high-energy  $\beta$ -particles from a radioisotope source (e.g.,  $^{63}\text{Ni}$ ) or produced directly in a pulsed discharge. The HID has a reputation as a difficult detector to use and must be operated under stringent conditions with respect to contamination from carrier gas impurities and the ingress of air or column bleed. Because applications have remained focused on inorganic gases and simple volatile organic compounds, gas-solid columns are often used for separations to minimize contamination problems and drifting baselines with contaminated liquid stationary phases.

### Bulk Physical Property Detectors

Bulk physical property detectors respond to some difference in a carrier gas property due to the presence of the sample. Usually, a large signal for the carrier gas property is desirable to provide a reasonable working range, but for low sample concentrations the detector signal corresponds to a very small change in a large signal, and is noise limited. The sensitivity of the bulk physical property detectors tends to be poor compared with ionization detectors. The most important of the bulk physical property detectors is the TCD. Other examples include the gas density balance and ultrasonic detector.

### Thermal Conductivity Detector

The TCD is a universal, nondestructive, concentration-sensitive detector that responds to the difference in thermal conductivity of the carrier gas and the

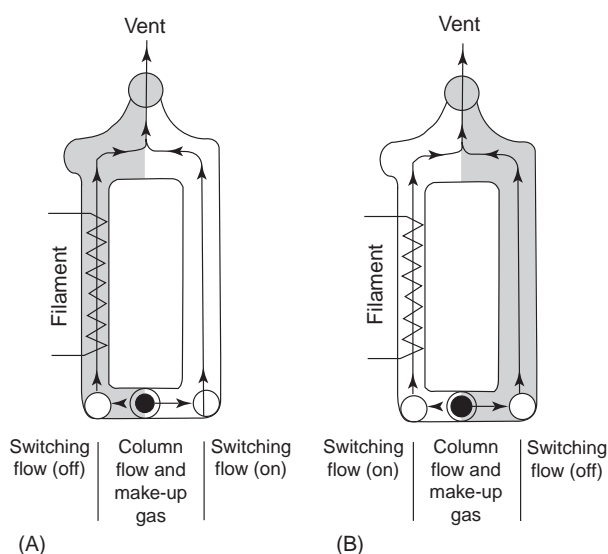
carrier gas containing sample. It is generally used to detect permanent gases, light hydrocarbons, and compounds that respond poorly to the FID. Typical detection limits are  $10^{-6}$  to  $10^{-8} \text{ g}$  per peak with a linear range of about  $10^4$ . If lower detection limits are required for compounds with a poor response to the FID, then the HID is a suitable alternative.

In a typical TCD, the carrier gas flows through a heated thermostatted cavity that contains the sensing element, either a heated metal wire or thermistor. With carrier gas flowing through the cavity the heat loss from the sensor is a function of the temperature difference between the sensor and cavity and the thermal conductivity of the carrier gas. When carrier gas containing sample vapors enters the cavity, there is a change in the thermal conductivity of the carrier gas mixture resulting in a change in the sensor temperature. The sensor may be operated in a constant current, constant voltage, or constant mean temperature mode as part of a Wheatstone bridge circuit. A temperature change in the sensor results in an out-of-balance signal proportional to the concentration of sample vapor in the sensor cavity.

The TCD has appeared in several different designs, some of which have advantages for particular applications. They usually represent some variation of the three basic geometries: the flow-through, semidiffusion, and diffusion cells. In the flow-through cell, carrier gas passes over the sensor and in the diffusion cell the sensor is located in a recess into which a portion of the carrier gas stream enters by diffusion. The diffusion and semidiffusion cells have a slow response and are relatively insensitive. They are used mainly for packed column analytical and preparative gas chromatography. Flow-through cells with volumes of 1–100  $\mu\text{l}$  are easily fabricated for use with open tubular columns. To improve thermal stability two or four cells are mounted in the detector oven block, half of which are used as reference cells through which only carrier gas flows, generating a difference signal proportional to sample concentration. In an alternative design, flow modulation is used to switch the carrier gas between two channels, one of which contains a single filament (**Figure 4**). Every 100 ms a switching valve fills the filament channel alternately with carrier gas and column effluent. With an effective detector volume of 3.5  $\mu\text{l}$ , it can be used with even narrow bore open tubular columns.

Carrier gases of low molecular weight and high thermal conductivity (e.g.,  $\text{H}_2$  and  $\text{He}$ ) are required to maximize the detector response and to maintain a large linear response range. Heavier carrier gases such as nitrogen, as well as influencing sensitivity and linearity, may give rise to negative or split top sample peaks. A number of response factor compilations for





**Figure 4** Schematic diagram of a single-filament thermal conductivity detector with flow modulation. In (A) the carrier gas is directed to flow over the filament and in (B) bypasses the filament.

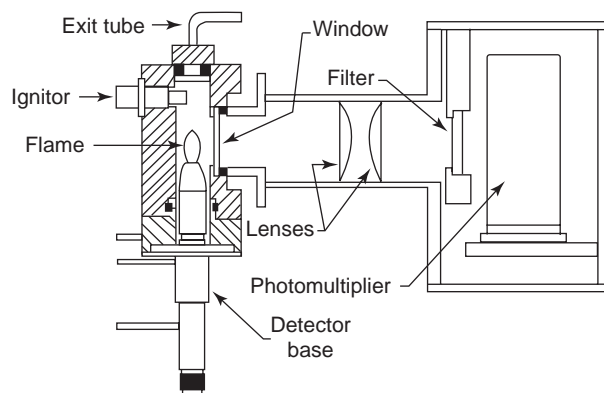
different carrier gases are available, usually expressed on a weight or molar response basis relative to benzene. These values are generally sufficiently accurate for estimating sample concentrations. For more accurate results it is necessary to calibrate the detector for each substance individually.

## Optical Detectors

The use of flames as atom reservoirs is a well-established technique in atomic spectroscopy. The principal emission lines for most nonmetallic elements of interest for gas chromatography occur in the ultraviolet region, where flame background contributions are troublesome. About 28 elements, including phosphorus and sulfur, can be determined by their chemiluminescence emission in hydrogen-diffusion flames with a flame photometric detector (FPD). For direct atomic emission detection, microwave induced and inductively coupled plasmas provide more appropriate atom sources for organic compounds. A number of chemiluminescence reaction detectors are also used in gas chromatography for the specific detection of sulfur- and nitrogen-containing compounds and nitrosamines.

### Flame Photometric Detector

The FPD is an element-selective detector commonly used for the determination of sulfur- and phosphorus-containing compounds. The FPD uses a hydrogen-diffusion flame to first decompose and then excite to a higher electronic state the fragments generated by the combustion of sulfur- and phosphorus-containing



**Figure 5** Cross-sectional view of a flame photometric detector. (Reproduced with permission from Patterson PL, Howe RL, and Abushumays A (1978) Dual-flame photometric detector for sulfur and phosphorus compounds in gas chromatographic effluents. *Analytical Chemistry* 50: 339–344.)

compounds in the effluent from a gas chromatograph. These excited state species subsequently return to the ground state, emitting characteristic band spectra. This emission is isolated by a filter and monitored by a photomultiplier detector.

In the single-flame detector the carrier gas and air are mixed, conveyed to the flame tip, and combusted in an atmosphere of hydrogen. With this burner and flow configuration interfering emissions from hydrocarbons occur mainly in the oxygen-rich flame regions close to the burner orifice, whereas sulfur and phosphorus emissions occur in the diffuse hydrogen-rich upper portions of the flame. Problems with solvent flameout, hydrocarbon quenching, and structure–response variations for different sulfur- and phosphorus-containing compounds with the single-flame detector can be partially overcome using a dual-flame design (Figure 5). The lower flame is hydrogen rich and functions as a matrix-normalization reactor in which all compounds are decomposed to a highly reduced state (e.g.,  $\text{H}_2\text{S}$ ,  $\text{S}_2$ ,  $\text{H}_2\text{O}$ ,  $\text{CH}_4$ , etc.). The combustion products from the first flame are swept into a second longitudinally separated flame where the desired optical emission is generated under optimized flame conditions. The pulsed-flame detector is based on a flame source and flame gas flow rates that cannot sustain a continuous flame. The combustion gases ( $\text{H}_2$  and air) are mixed together in a small chamber and flow to a continuously heated wire igniter. The ignited flame then propagates back to the gas source and is self-terminated once all of the combustible gas mixture present in the combustion path is consumed. The continuous gas flow removes the combustion products and creates additional ignition in a periodic fashion. The pulsed flame emission provides enhanced detection sensitivity and selectivity by time resolution of the various

flame luminescent species. The luminescence from hydrocarbon and flame combustion products (e.g.,  $\text{OH}^*$ ,  $\text{CH}^*$ ,  $\text{C}_2^*$ , etc.) is limited to the time duration for the flame front to pass across the photomultiplier viewing area. Because of lower bond energies the sulfur- and phosphorus-containing species continue to emit in the cooler, yet reactive, postpulse flame conditions. By gating the detector the heteroatom emission can be time resolved from the hydrocarbon and flame background emission.

In the relatively low-temperature and hydrogen-rich flame, sulfur-containing compounds are decomposed to species such as  $\text{H}_2\text{S}$ ,  $\text{HS}$ ,  $\text{S}$ ,  $\text{S}_2$ ,  $\text{SO}$ , and  $\text{SO}_2$  in relative proportions that depend on the flame chemistry. Excited state  $\text{S}_2^*$  species are formed from these primary species in low yield by several two- or three-body collision reactions. The relaxation of  $\text{S}_2^*$  results in broadband emission over the wavelength range from 320 to 460 nm with a maximum emission at 394 nm. The response for sulfur is inherently nonlinear and proportional to the concentration of sulfur atoms entering the flame ( $[\text{S}]^n$ ). The theoretical value for  $n$  is 2, but in practice, values between 1.6 and 2.2 are frequently observed for the single-flame detector. Nonoptimized flame conditions, compound-dependent decomposition, hydrocarbon quenching, and competing flame reactions that lead to de-excitation all contribute to the deviation of  $n$  from its theoretical value. Decoupling the compound decomposition process from the excitation process in the dual-flame and pulsed-flame detectors results in a more truly quadratic response.

In the case of phosphorus, phosphorus-containing compounds are first decomposed to  $\text{PO}$  molecules and finally to electronically excited  $\text{HPO}^*$  species in three-body collisions. A linear relationship between detector response and the concentration of phosphorus atoms entering the flame is expected for phosphorus-containing compounds, at least for low sample concentrations.

Typical detection limits are  $5 \times 10^{-13}$  to  $1 \times 10^{-14}$  g P/s (all detector types) and  $5\text{--}50 \times 10^{-12}$  g S/s (single flame),  $1 \times 10^{-11}$  g S/s (dual flame), and  $2 \times 10^{-13}$  g S/s (pulsed flame). The linear range for phosphorus usually exceeds  $10^3$  while the selectivity is more than  $5 \times 10^5$  g C/g P. Sulfur selectivity varies in the range  $10^4\text{--}10^6$  g C/g S (single flame),  $10^3\text{--}10^4$  g C/g S (dual flame), and  $>10^7$  g C/g S (pulsed flame).

### Chemiluminescence Detectors

Nitrosamines can be detected after thermal cleavage and nitrogen-containing compounds after oxidation to nitric oxide, which reacts with ozone to form nitrogen dioxide in an excited state with photon emission in the near-infrared around 1200 nm.

Sulfur-containing compounds can be decomposed by thermal oxidation to sulfur monoxide and subsequently reacted with ozone to form sulfur dioxide in an excited state with photon emission centered  $\sim 360$  nm.

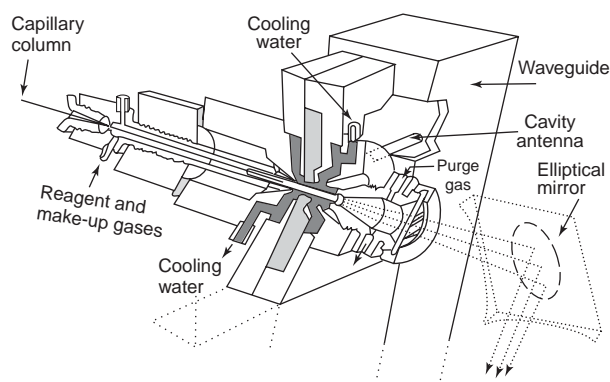
A typical chemiluminescence detector consists of a series-coupled thermal decomposition and ozone reaction chambers. The selective detection of nitrosamines is based on their facile low-temperature ( $275\text{--}300^\circ\text{C}$ ) catalytic pyrolysis to release nitric oxide. Thermal decomposition in the presence of oxygen at about  $1000^\circ\text{C}$  affords a mechanism for conversion of nitrogen-containing compounds to nitric oxide (catalytic oxidation at lower temperatures is also possible). Decomposition in a hydrogen-diffusion flame or thermal oxidation in a ceramic furnace is used to produce sulfur monoxide from sulfur-containing compounds.

The products of the conversion reaction flow into the ozone reaction chamber, which is maintained at a pressure of 10–30 mmHg. Reduced pressure operation has three advantages: it improves sensitivity by diminishing collisional deactivation of the excited state reaction products; it prevents condensation of water in the reaction chamber; and it reduces the effective detector volume maintaining compatibility with open tubular columns. Ozone is generated by a high-voltage discharge in air or oxygen and enters the reaction chamber by a separate connection. The chemiluminescence emission produced is isolated by an optical filter and detected by a photomultiplier. If one of the reactants in the chamber is maintained in large excess (e.g., ozone), the reaction becomes pseudofirst order for the other reactant, resulting in a linear response.

The most important applications of the chemiluminescence detector are the determination of volatile nitrosamines in food, environmental, and industrial products (detection limits  $\sim 0.5$  ng per peak with a linear range of  $10^3$ ) and sulfur-containing compounds in complex matrices, for example, in the petroleum and gas industries for process control and regulatory compliance, and in the beverage industry to characterize flavor compounds. The response of the chemiluminescence detector to sulfur-containing compounds is almost equal on a per gram of sulfur basis. This allows the use of a single calibration curve for the determination of all sulfur-containing compounds in many cases. The detection limit for the flame-based detector is about  $10^{-12}$  g S/s, with a linear range of  $10^4\text{--}10^5$ , and selectivity  $>10^6$  g C/g S. For the flameless detector the detection limit is about  $10^{-13}$  g S/s, with a linear range of  $10^4\text{--}10^5$ , and selectivity  $>10^7$  g C/g S. In the nitrogen-selective mode detection limits are about  $10^{-12}$  g N/s, with a linear range of  $10^4\text{--}10^5$ , and selectivity  $10^7$  g C/g N.

### Atomic Emission Detector

The atomic emission detector (AED) is a relatively complex instrument and is described below in outline only. Coupling of the separation column to the thermostatted microwave cavity is made through a heated transfer line. The plasma is produced in a thin-walled silica discharge tube within a microwave 'reentrant' cavity (Figure 6). Power is supplied by a magnetron, and coupled to the plasma through a waveguide. The exit of the cavity is closed with a fused silica window and purged with helium to prevent back-diffusion of air into the cavity, and to allow flow reversal so that the solvent peak can be vented in front of the cavity to minimize carbon build-up in the cavity. The plasma is generated in an atmospheric pressure flow of helium made up of the column flow and additional make-up flow, as required. Depending on the elements being determined, low concentrations of various scavenger gases (e.g., O<sub>2</sub>, H<sub>2</sub>, and CH<sub>4</sub>) are also



**Figure 6** Cross-sectional view of the cavity block of an atomic emission detector. (Reproduced with permission from Sullivan JJ and Quimby BD (1989) Detection of C, H, N and O in capillary gas chromatography by atomic emission. *Journal of High Resolution Chromatography* 12: 282–286.)

added. An elliptical mirror collects emissions from 2 mm inside the end of the discharge tube. The emission sensor consists of a flat focal-plane spectrometer with a movable photodiode array detector with a range of about 25 nm. This determines which element combinations can be measured simultaneously. In general, up to four elements can be detected and displayed as element-specific chromatograms at the same time. Typical figures of merit are summarized in Table 4. Obviously, the AED is capable of replacing most of the element-selective detectors described in this article. It is more complex and expensive to operate, however, and has made only modest inroads into the detector market place. On the other hand, it is almost uniquely suited to the detection of organometallic compounds.

In theory, it should be possible to determine the empirical formula for each compound in the chromatogram from the ratio of the AED response to the individual elements. A lack of plasma stability, incomplete compound destruction, and deviations from linearity of the individual element responses, however, limits the accuracy of such measurements. Accurate formula values for oxygen and nitrogen can be particularly difficult to determine due to entrainment of atmospheric gases into the plasma. In addition, it is not always possible to use a single compound internal standard for quantification with the desired accuracy for the same reasons that affect the accuracy of empirical formula determinations.

### Electrochemical Detectors

There are two general problems associated with electrochemical detection in the gas phase. First, few electrochemical detectors are gas-phase sensing devices, and therefore the separated sample components must be transferred into solution for detection.

**Table 4** Response characteristics of the atomic emission detector to different elements

Element (X)	Wavelength (nm)	Minimum detectable amount (pg s <sup>-1</sup> )	Selectivity (g C/g X)	Linear range
C	193.1	2.6		2 × 10 <sup>4</sup>
H	486.1	2.2		6 × 10 <sup>3</sup>
Cl	479.5	39	2.5 × 10 <sup>4</sup>	2 × 10 <sup>4</sup>
Br	470.5	10	1.1 × 10 <sup>4</sup>	1 × 10 <sup>3</sup>
F	685.6	40	3.0 × 10 <sup>4</sup>	2 × 10 <sup>3</sup>
S	180.7	1	3.5 × 10 <sup>4</sup>	1 × 10 <sup>4</sup>
P	177.5	1	5.0 × 10 <sup>3</sup>	1 × 10 <sup>3</sup>
N	174.2	15	2.0 × 10 <sup>3</sup>	4 × 10 <sup>3</sup>
N	388	15	8.0 × 10 <sup>5</sup>	1 × 10 <sup>4</sup>
O	777.2	50	3.0 × 10 <sup>4</sup>	3 × 10 <sup>3</sup>
Sn	303.1	0.5	3.0 × 10 <sup>4</sup>	1 × 10 <sup>3</sup>
Se	196.1	4	5.0 × 10 <sup>4</sup>	1 × 10 <sup>3</sup>
Hg	253.7	0.1	3.0 × 10 <sup>6</sup>	1 × 10 <sup>3</sup>

Second, the majority of organic compounds separated by gas chromatography are neither electrochemically active nor highly conducting. The electrolytic conductivity detector (ELCD) solves both of these problems by decomposing the gas-phase sample into small inorganic molecules, which are detected by their conductivity in a support solvent. The ELCD is used primarily as an element-selective detector for halogen-, sulfur-, and nitrogen-containing compounds.

The carrier gas is mixed with a reaction gas and then passed through a small diameter nickel tube at 850–1000°C. The nickel tube acts as catalyst for the decomposition reaction. With hydrogen as the reaction gas, halogen-containing compounds are converted to hydrogen halide (HCl, HBr), and nitrogen-containing compounds are converted to ammonia. Sulfur-containing compounds are reacted with air to produce sulfur dioxide with a small amount of sulfur trioxide. A chemical scrubber mounted at the exit of the reaction tube is used as needed to enhance the specificity of the detection process. For example, silver wires are used to remove hydrogen halides or hydrogen sulfide, potassium hydroxide supported on quartz fibers to remove acidic species, and aluminum silicate to remove sulfur oxides.

The reaction products from the furnace are swept into a gas–liquid contactor where they are mixed with an appropriate solvent (Figure 7). The support solvent is selected to promote ionization of the reaction species over ionization of interfering compounds. This solvent is usually circulated through a closed system containing beds of ion exchange resins to purify and condition the solvent for reuse. From the gas–liquid contactor the support solvent flows to the conductivity cell, where detection takes place, either after separation of the liquid from insoluble

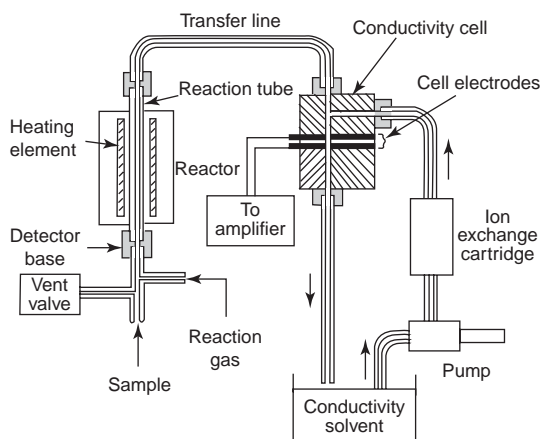
gases, or as a mixed phase, depending on the detector design.

The ELCD has the reputation of being a demanding detector to operate and its popularity has declined as other element-selective detectors with simpler operating features became available. Common practical problems include the loss of response, excessive noise, poor linearity, and poor peak shape. The ELCD is capable of high sensitivity and selectivity with detection limits of about  $10^{-12}$  g N/s,  $10^{-12}$  g S/s, and  $5 \times 10^{-13}$  g Cl/s with a linear range of  $10^3$ – $10^5$ . Selectivity varies with the heteroatom detected as well as detector operating conditions. Values of  $10^4$ – $10^9$  g C/g N, S, or Cl are possible. The ELCD is used primarily for the determination of chlorine- and nitrogen-containing compounds in environmental samples and for the determination of sulfur-containing compounds in petroleum products.

**See also:** Atomic Emission Spectrometry: Microwave-Induced Plasma. Chemiluminescence: Gas-Phase. Derivatization of Analytes. Gas Chromatography: Mass Spectrometry; Fourier Transform Infrared Spectroscopy.

## Further Reading

- Dojahn JG, Wentworth WE, Deming SN, and Stearns SD (2001) Determination of percent composition of a mixture analyzed by gas chromatography. Comparison of a helium pulsed-discharge photoionization detector with a flame ionization detector. *Journal of Chromatography A* 917: 187–204.
- Dressler M (1986) *Selective Gas Chromatographic Detectors*. Amsterdam: Elsevier.
- Grob RJ (ed.) (1995) *Modern Practice of Gas Chromatography*. New York: Wiley.
- Handley AJ and Adlard ER (eds.) (2001) *Gas Chromatographic Techniques and Applications*. Sheffield: Sheffield Academic Press.
- Holm T (1999) Aspects of the mechanism of the flame ionization detector. *Journal of Chromatography A* 842: 221–228.
- McMinn DG and Hill HH (eds.) (1992) *Detectors for Capillary Chromatography*. New York: Wiley.
- Poole CF (2003) *The Essence of Chromatography*. Amsterdam: Elsevier.
- Scott RPW (1996) *Chromatographic Detectors Design, Function and Operation*. New York: Marcel Dekker.
- Wentworth WE, Huang J, Sun K, et al. (1999) Non-radioactive electron-capture detector. *Journal of Chromatography A* 842: 229–266.
- Yan X (1999) Detection by ozone-induced chemiluminescence in chromatography. *Journal of Chromatography A* 842: 267–308.
- Zlatkis A and Poole CF (eds.) (1981) *Electron Capture. Theory and Practice in Chromatography*. Amsterdam: Elsevier.



**Figure 7** Cross-sectional view of an electrolytic conductivity detector. (From McMinn DG and Hill HH (eds.) (1992) *Detectors for Capillary Chromatography*. New York: Wiley.)



## Mass Spectrometry

D J Harvey, University of Oxford, Oxford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The coupling of gas chromatography (GC) with mass spectrometry (MS), techniques that both employ the sample in the gas phase, was first achieved in 1957. Since then, GC-MS has developed into one of the most sensitive and selective analytical methods for the separation, identification, and quantification of components of complex organic mixtures. It gives a two-dimensional identification consisting of both a GC retention time and a mass spectrum for each component of the mixture. This dual approach is particularly useful for isomer differentiation where good GC separation frequently compensates for difficulties encountered as a result of indeterminate mass spectra. Conversely, the mass spectrum allows structural features to be assigned to the compound producing the GC peak and often provides unambiguous structural assignment. Although the method is limited to the analysis of those compounds that can be made volatile, without thermal decomposition, many compounds that initially fail this requirement can be successfully handled after chemical derivatization.

The early success in the development of this so-called hyphenated technique owes much to the general compatibility between the two methods in terms of sample size and volatility. The only major difference is that the exit from the gas chromatographic column is at low ( $10^{-6}$  Torr), rather than at atmospheric pressure. Nevertheless, modern systems provide excellent GC and MS performance without significant compromise to the behavior of either technique.

### Instrumentation

The instrumentation consists essentially of three components: the gas chromatograph, the mass spectrometer, and a data system (Figure 1). The latter component can be used both for data acquisition and processing, and for instrument control. In systems where packed GC columns are used, an interface is included between the column and the spectrometer to reduce the carrier gas flow from  $\sim 30 \text{ ml min}^{-1}$  to  $\sim 1.5\text{--}2 \text{ ml min}^{-1}$  so that it can be handled by the mass spectrometer's vacuum system. However,

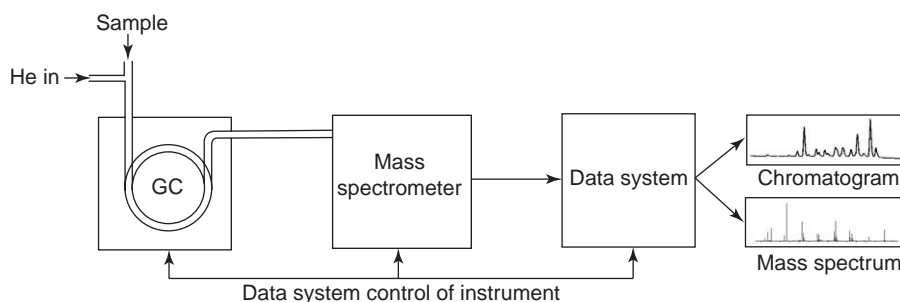
modern systems usually employ capillary columns made up of polyimide-coated fused silica that are interfaced directly with the mass spectrometer. In addition to the above equipment, GC-MS instruments may also be provided with additional features such as an autosampler, an infrared spectrometer, or a flame-ionization detector.

### The Gas Chromatograph

Almost any commercial gas chromatograph can be interfaced to a mass spectrometer and, although its operation is similar to that when used in the stand-alone mode, there are certain features that are specific to successful GC-MS operation.

Column dimensions and carrier gas flows for packed columns tend to be similar to those normally used for stand-alone GC. However, with capillary columns that are interfaced without a separator, the carrier gas flow needs to be fairly low and, consequently, narrow-bore columns with internal diameters of 0.35 mm or less are generally used. Typical columns are 5–30 m long with internal diameters of 0.25 or 0.32 mm. Although columns with internal diameters of 0.53 mm can be used, these usually require a splitter to reduce the flow into the mass spectrometer. Care must be taken with the metal-coated columns, used for high-temperature work, as the metal can short out the ion source voltage. This problem can be avoided by removing the metal coating from the terminal 2–3 cm of the column.

Both gas/solid adsorption and gas/liquid partition chromatography can be used for GC-MS, but GC is by far the most common. Because, in GC, the stationary phase is a liquid, usually a polymer, its vapor pressure will cause a continual low flow, or bleed into the ion source of the mass spectrometer. This bleed, which usually consists of decomposed stationary phase, will produce a spectrum whose intensity increases with column temperature. Stationary phases should therefore be of the high-boiling, low-bleed type. Most currently used stationary phases for routine GC-MS are based on alkyl-polysiloxanes or alkyl-phenyl-polysiloxanes that are chemically bonded to the column wall to increase stability. Columns containing such phases can, in some cases, be used at temperatures of up to 400°C. One advantage, however, to the presence of bleed peaks in the spectrum is that they enable a continual check to be made on the mass spectrometer calibration. For the alkyl siloxanes, ion peaks are present, in decreasing relative abundance, at  $m/z$  73, 207, 281, 355, 429,



**Figure 1** Block diagram of a typical GC–MS system.

503, and 577. The 74 mass unit intervals correspond to O–Si(Me)<sub>2</sub>.

### The GC–MS Interface

Much of the earlier literature on GC–MS instrumentation was concerned with the construction and operation of devices for removing much of the carrier gas from packed column effluents while, at the same time, achieving maximum sample transfer to the mass spectrometer. However, as the use of fused silica capillary columns is now nearly universal, separators for interfacing packed columns retain little more than historical interest.

**Interfacing packed columns** Removal of the excess of carrier gas is most frequently achieved by utilizing the relative diffusion differences between the carrier and the sample molecules. The most widely used device is the jet molecular separator in which the column effluent is allowed to pass through a small jet into an evacuated chamber where it expands. A skimmer placed ~1–2 mm away from the inlet jet captures the center of the effluent stream containing much of the analyte, while allowing most of the carrier gas, which has undergone more rapid lateral diffusion because of its lower molecular weight, to be removed by the pumping system. Helium is invariably used as the carrier gas to maximize the effect, and enrichment factors of greater than 20 are frequently achieved. A similar device is the frit-type or Watson–Biemann separator. This works in an analogous manner in that the carrier gas is removed by lateral diffusion through a fritted glass or ceramic tube through which the column effluent passes. In a third method, the column effluent passes over a silicone rubber membrane through which the organic molecules diffuse into the mass spectrometer. This system, however, has never been widely used, as it tends to cause a reduction in GC resolution.

**Interfacing capillary columns** Direct introduction of the capillary column into the ion source of the

mass spectrometer is now the most generally used method for interfacing the two techniques. However, the presence of the mass spectrometer vacuum results in a much greater pressure drop across the column than would be encountered in normal gas chromatographic operation, and this has led to concern, usually unjustified, about the maintenance of GC resolution. To overcome this difficulty, an open-split interface can be employed. Here, a small gap separates the column from the transfer line that leads the effluent into the ion source. This gap is bathed with helium and, by suitable adjustment of the helium pressure, the exit of the column can be maintained at atmospheric pressure with good sample transfer into the mass spectrometer.

### The Mass Spectrometer

Most types of mass spectrometer can be adapted to take the effluent from a GC column, provided the ion source pumping is sufficient to remove the carrier gas. Quadrupole instruments, in which ion separation is achieved by allowing the ions to drift between four concentric rods carrying DC and AC potentials, are probably best adapted to the technique because of their fast scanning properties and the absence of high ion source potentials. The ion trap, which can be regarded as a development of the quadrupole ion filter, provides a cheap but, at present, limited mass range detector, but can give varying spectra depending on the residence time of the ions in the trap. Ions that remain for longer periods tend to undergo reactions by collision with gas molecules, with the result that electron impact (EI) spectra can acquire some chemical ionization (CI) character (see below). Magnetic sector instruments, in which ion separation is achieved by passage through a magnetic field, allow high-resolution spectra to be acquired for the calculation of elemental compositions. Although maximum scan speeds are lower than with quadrupole-based systems, as a result of magnetic hysteresis, they are sufficient for most applications. More recently, time-of-flight (TOF) instruments have been



adapted for monitoring GC column effluents. In order to accommodate the requirement for a pulsed source of ions, a voltage applied to a 'pusher plate' pulses successive bursts of ions orthogonally into the analyzer from the continuous ion beam leaving the ion source. One of the main advantages of this TOF instrumentation is high sensitivity as most ions that are pulsed into the analyzer eventually reach the detector. Other advantages are the ability to achieve very rapid data acquisition and to overcome spectral bias (see below), essential features of instruments producing GC peak widths of the order of milliseconds.

## **Instrument Operation**

One of the most important aspects of setting up a GC-MS system is proper temperature control of the various areas through which the sample passes. The injector and column are operated as with conventional gas chromatography but, then, care must be taken to preserve chromatographic integrity until the sample reaches the mass spectrometer. For most samples, the transfer lines (and separator, if used) should ideally be set to a temperature slightly higher than that of the upper column temperature limit to avoid condensation problems. The ion source should be similarly heated. In practice, however, it is found that a somewhat lower (20–30°C) transfer line temperature can be tolerated although this is, to some extent, sample dependent. The optimum temperature may need to be determined experimentally for a particular analysis and may be a compromise between the need to avoid condensation and to prevent thermal decomposition. Also of great importance in transfer line design is the need to avoid dead volumes where remixing of the separated sample constituents can occur. However, in most modern systems, the separation column can simply be inserted directly into the mass spectrometer ion source, thus overcoming problems associated with separate connections.

Once the sample has reached the ion source, it may be ionized using either EI or CI. EI gives extensive fragmentation and is the most appropriate method for structure determination and for producing spectra suitable for library searching. CI, which is a softer ionization technique, produces little fragmentation and is normally used when molecular ion assignment is in doubt. The presence of residual carrier gas does not significantly affect the spectrum under EI conditions, but must be taken into account with the higher ion source pressures encountered under CI conditions. The helium signal can be avoided simply by using a mass greater than 4 as the lowest scanned

mass in the spectrum. In practice, the spectrometer is usually operated with a lower mass value of  $\sim 50$  to avoid the additional abundant ion peaks from atmospheric gases present as the result of small leaks.

The scan speed of the mass spectrometer should be adjusted to give at least 5–6 scans over a typical GC peak in order to provide a usable peak profile. However, more scans will lead to a better definition of the peak profile and will be necessary for quantitative work where better peak shape definition is required. However, a compromise should be sought in order to avoid excessively large data files. For temperature-programmed work with capillary columns, scan speeds in the 0.1–1.0 s range are appropriate. If isothermal conditions are employed, the scan speed can be decreased during later stages of an analysis to accommodate the resulting broader peaks. However, for mixture analysis, temperature programming is recommended. For packed columns, slower scan speeds (typically 1–2 s) are appropriate. It must be recalled that, as the concentration of sample entering the ion source varies with time, single spectra will contain a concentration bias depending on which side of the GC peak the spectrum was acquired and in which direction (high to low mass or vice versa) the scan was taken. Spectra with the least bias are acquired at the peak top. Alternatively, bias can be minimized by averaging, with the aid of the data system, the several spectra acquired over the GC peak, assuming, of course, that only one component is present.

## **Processing GC-MS Data**

When first introduced, GC-MS equipment was not supplied with a data system as suitable computers had not been developed. Spectra were recorded on UV chart paper with the operator initiating the scan in response to a peak appearing on the GC trace. Now, all spectral acquisition and processing are data-system-dependent. The major function of the GC-MS data system is to digitize the analog output from the mass spectrometer detector and to perform a time-to-mass conversion to produce the mass spectrum. The digitization rate should be set to a value that gives  $\sim 10$ –20 sample periods over each ion peak in the spectrum. This sampling rate will depend on the mass spectrometer resolution and is usually set automatically once the resolution is specified. Various filters can then be set to remove noise and accommodate multiplet ion peaks. The most common filters are the position of the baseline and the requirement that a certain number of successive samples exceed the baseline value before the signal is

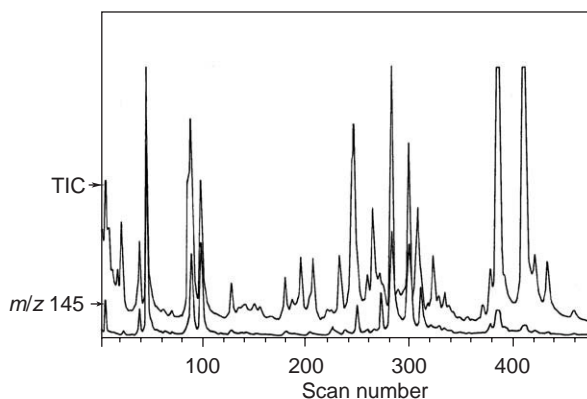
determined to be from a GC peak. The computer then calculates the peak centroid to produce the  $m/z$  value and stores this on disk together with the peak intensity. These data can then be further processed by the data system to produce spectral plots, various reconstructed chromatograms, or library searches.

### Ion Chromatograms

A plot of total ion current (TIC), the sum of all ion peak intensities in a spectrum, against time (or scan number) gives a chromatogram comparable to that obtained by GC using a conventional detector. However, chromatographic peak abundances will vary for compounds of different structure because of their different ionization efficiencies in the spectrometer. This difference in response factors applies to both EI and CI data. For example, under EI conditions, the signal recorded for hydrocarbons will be much weaker in a TIC than that recorded with a flame-ionization detector, relative to other compounds, because these compounds lack a heteroatom or unsaturation and are, thus, difficult to ionize.

The plot of the intensity of the  $m/z$  value of a single ion with time provides an accurate method for locating specific compounds in a chromatogram, particularly when they occur as components of an unresolved GC peak. Thus, chromatograms of the mass of a selected molecular ion can be used to identify candidate chromatographic peaks for a particular molecule, and chromatograms of the mass of a specific fragment ion can indicate which peaks in the TIC are produced by compounds with a given structural feature (Figure 2). It must, however, be borne in mind that false positives are possible. These can occur, for example, from the presence of an ion at the selected mass formed from a different fragmentation pathway or an ion present as an isotopic peak from an ion at a lower  $m/z$  value. This method, however, although giving a result similar to the peak profiles obtained by selected ion recording (see below), does not have the sensitivity advantage of the latter technique. It does allow, however, for retrospective ion plotting.

Chromatograms derived from combinations of several ions may also be plotted. These ions may be from a single compound, thus adding confidence to the assignment of structure, or they can be common to a group of compounds with similar structures. A variation on this theme is to use ions whose mass separation and relative abundance are characteristic of the presence of a given isotopic profile such as those given by chlorine-containing compounds. Another technique is to use only ions above a given mass for compounds that produce ions mainly in the



**Figure 2** (Top trace) TIC plot of the TMS derivatives of metabolites of  $\Delta^9$ -tetrahydrocannabinol extracted from mouse liver. Metabolites appear in scans 160–360, fatty acids in scans 0–160, and bile acids in scans 360–480. (Lower trace) Single ion plot of the ion at  $m/z$  145, diagnostic for hydroxylation at a specific site in the drug molecule. The metabolites producing peaks in the region of scans 250–320 contain this structural feature, while the other metabolite-related peaks do not. However, the ion is also present in the spectra of the fatty acids but not in those of the bile acids. The separation used a 2 m SE-30 packed column with a temperature-programmed run ( $2^\circ \text{ min}^{-1}$ ).

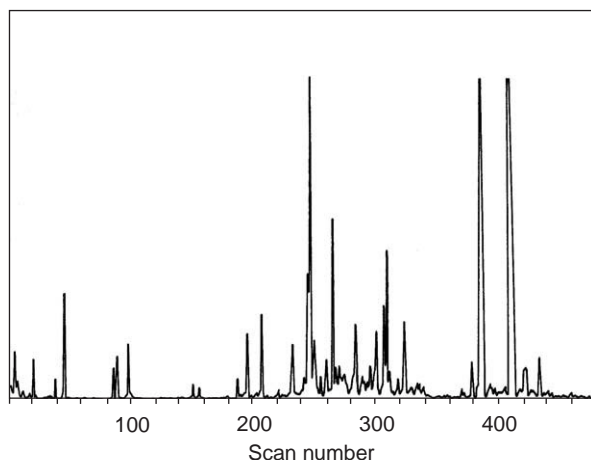
upper-mass range. In this way, many of the undiagnostic, low-mass ions leading to high backgrounds in the chromatograms may be removed.

### Deconvolution of Unresolved GC Peaks

Methods for peak deconvolution usually rely on detecting changes in mass spectral ion abundances from the individual components as they elute across the peak. The Biller–Biemann and similar algorithms, for example, look at each spectrum and set a flag against all ions whose intensity is greater than that in adjacent spectra. This essentially marks the top of the GC peak corresponding to the compound containing the flagged ions. By plotting chromatograms from only flagged ions, a much-improved chromatogram can be obtained (Figure 3). Spectra of individual components can also be obtained by using only the flagged ions. A disadvantage of these algorithms is that, for successful deconvolution, the spectra must possess ions of different mass and these must pass through an intensity maximum in different scans. Recently, more sophisticated deconvolution techniques such as those based on maximum entropy algorithms have been introduced.

### Spectrum Addition and Subtraction

Spectral addition enables all the spectra acquired from a given GC peak to be averaged, with the result that a greater signal-to-noise ratio is obtained and



**Figure 3** The TIC trace from **Figure 2** processed by flagging ions in the scans in which their intensity maximizes and ignoring ions below  $m/z$  300. Several of the peaks eluting in the region of scans 250 and 310 have been resolved into several components. One of the components of two of these resolved peaks, which can be seen as a shoulder in **Figure 2**, contains the ion at  $m/z$  145.

much less bias is present. Spectral subtraction can be used, either to remove background or to subtract ions from a given compound from a mixed spectrum. Most data systems now allow an almost unlimited selection of peaks to be either added to or subtracted from each other.

### Library Searching

Library searching is invaluable for the rapid identification of sample constituents and can often give clues to the identity of unknown compounds. Several commercial libraries are available containing many tens of thousands of spectra. Among the largest are the Wiley Registry of Mass Spectral Data 7th Edition (390 000 spectra) and the NIST/EPA/NIH Mass Spectral Library (129 000 spectra). Several search algorithms are available, of which the so-called reverse search involving finding the peaks from the library spectrum in the spectrum of interest is probably the most useful. To avoid searching the entire library for each compound, it is common practice to limit searches to, for example, compounds below a certain mass or to compounds with specific structures. Best-fit spectra are usually presented to the analyst, together with a goodness-of-fit factor for evaluation. However, it is important to realize that the best library match may not be the correct molecular identity. More confidence in a correct match can be gained by making use of the compound's retention index, another parameter that can be incorporated into databases.

### Calculation of Elemental Composition

Where spectra have been acquired under high-resolution conditions and ionic masses measured to a few parts per million, the data system can be used to calculate possible elemental compositions that fit the observed mass to within a specified mass window. Again, upper limits to the elements present must be provided to avoid the production of an excess of redundant data.

### Quantitative Measurements

#### Selected Ion Monitoring

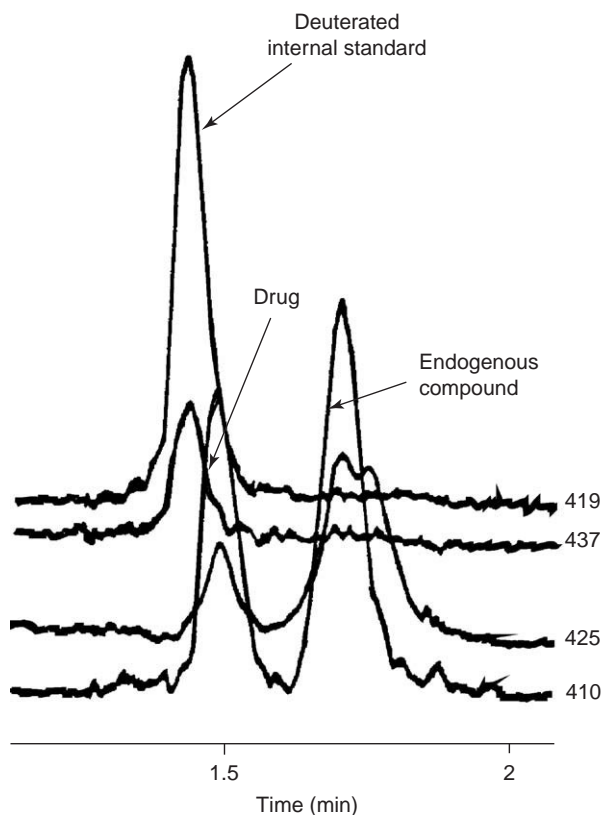
Selected ion monitoring (SIM) is a major application for GC-MS and allows the instrument to be used in its most sensitive mode for quantitative measurements. SIM was first introduced in the 1960s by Hammar and co-workers under the name mass fragmentography and has acquired a number of other names such as multiple ion monitoring (or recording) over the years. It owes its high sensitivity to the method of focusing the mass spectrometer. The instrument is focused on only a few ions that are recorded continuously rather than operated in the normal scanning mode. In its most sensitive mode, one ion is monitored for the duration of the experiment. Consequently, all the current from this ion is captured, whereas if the mass spectrometer were scanned only a small fraction of the ions at this  $m/z$  value would be acquired. Obviously, this method has reduced selectivity compared with that of scanning the whole spectrum, but selectivity can be partially restored by continually jumping between several significant ions, recording their ion current for a few milliseconds and then reconstructing the ion profiles.

The technique is most conveniently implemented with a quadrupole instrument as mass changes can be accomplished very rapidly. Channel switching by changing the magnetic field of a magnetic sector instrument is slower because of hysteresis effects. More rapid switching is usually achieved with sector instruments by altering the accelerating voltage, but at the expense of a reduction in sensitivity for ions with higher  $m/z$  values (focused at lower acceleration potentials) and some defocusing of the ion source.

Sensitivity in mass spectrometry is determined by the signal-to-noise ratio, which, in turn, has components of instrumental and chemical noise, the latter arising from the sample. Instrumental noise is generally a fixed parameter, but chemical noise can often be reduced by the user, not only by more efficient sample cleanup but by the way in which the instrument is operated. One such method is the technique of selected reaction monitoring, implemented on

sector or tandem mass spectrometers, whereby both parent and fragment ions are selected, thus ensuring that the monitored fragment ion arises from a specific precursor. Such an experiment could be set up with a triple quadrupole mass spectrometer by using the first quadrupole to select the parent ion and then monitoring the product or products of fragmentation with the analyzing quadrupole. Although the ion current of the monitored ion(s) is lower than that of the parent, this can be more than compensated for by a great reduction in the chemical noise.

For quantitative measurements, the 'ion-jumping' techniques allow variants of the analyte labeled with stable isotopes to be used as internal standards. The instrument can be set to record one or more ions from the analyte together with the corresponding ions from the internal standard (Figure 4). As the



**Figure 4** Selected ion monitoring for the quantification of a drug in human plasma using an internal standard labeled with 12 deuterium atoms. The molecular ions are recorded with the channels labeled  $m/z$  425 and 437 for the drug and internal standard, respectively, and the  $[M - CH_3]^+$  ions are recorded at  $m/z$  410 and 419. Note that three deuterium atoms are missing from the  $[M - CH_3]^+$  ion. The second set of peaks is from endogenous cytidine, which contains ions at the same mass as the analyte and emphasizes the importance of ensuring separation of analyte from contaminating compounds and of using both retention time and peak intensity in these measurements.

standard is, to a first approximation, chemically equivalent to the analyte, extraction and derivatization losses during sample preparation are compensated for and high accuracy and precision are obtained. Care must however, be taken to ensure that the isotopically labeled standard has a sufficient mass difference from the analyte to avoid interference by the latter's natural isotopes, and that no unlabeled material is present in the standard. Where interference is experienced, nonlinear calibration curves are obtained, which require special treatment to restore accuracy. Several such methods have been described in the literature but none appears to be universally valid.

Quantification by SIM can also be accomplished using nonlabeled standards. If a chromatographic separation exists between the standard and analyte, monitoring can take place by recording a common ion; this allows the technique to be used in its most sensitive mode. The most common nonlabeled standards are homologs of the analyte, but other chemically similar compounds can also be used, although sometimes at the expense of some accuracy.

#### Use of GC-MS for Trace Analysis

Trace analysis requires the highest sensitivity and is usually accomplished either by increasing the mass spectrometer resolution (a double focusing magnetic instrument is required) or by using MS/MS techniques. Both methods increase the selectivity of detection and improve the signal-to-noise ratio rather than by increasing the absolute signal amplitude. The first method makes use of the fact that ions appearing at a given nominal  $m/z$  value will frequently have different compositions and, thus, different absolute masses. By increasing the instrument resolution to such a value that these ion peaks are resolved, and monitoring only the ion of the specific composition required, most of the interfering compounds can be rejected. A good example of this is in dioxin analysis, where several other chlorinated aromatic compounds have the same nominal but different accurate masses. By monitoring only the absolute mass of the dioxin molecule ( $m/z$  321.8936), the interfering compounds can be rejected.

In the MS/MS methods, an ion from the analyte is selected early in its path through the instrument, fragmented in a collision cell, and selected product ions are monitored. This procedure is most easily accomplished with a tandem, hybrid, or triple quadrupole instrument and ensures that the ions monitored have originated from a specific precursor ion rather than from any other compound that has the same GC retention time.



## Sample Preparation

For successful GC–MS analysis, the sample must be both volatile and thermally stable. Few organic compounds fulfill these criteria as they possess polar or reactive functional groups that have a detrimental effect on these properties. Protecting these functional groups by derivatization is the normal method for conferring good GC–MS behavior. Suitable derivatives are listed in **Table 1**.

Although the derivatives used for GC analysis are generally suitable for conferring volatility, not all derivatives give ideal mass spectra. A good example of this is given by the methyl and trimethylsilyl (TMS) esters of fatty acids. The mass spectra of these derivatives give very little information on the alkyl chain structure, as fragmentation is dominated by cleavages around the carboxylic ester group (**Figure 5**).

A better derivative is the picolinyl ester. This derivative, under EI, acquires a positive charge on the nitrogen atom rather than on the ester oxygen atoms. The stability of the pyridine ring prevents fragmentation near the charge site, but causes the nitrogen atom to act as a proton acceptor and to abstract a proton from various positions of the alkyl chain to

leave a radical site. The chain then fragments by radical-induced cleavage in modes that reveal the chain structure (**Figure 6**).

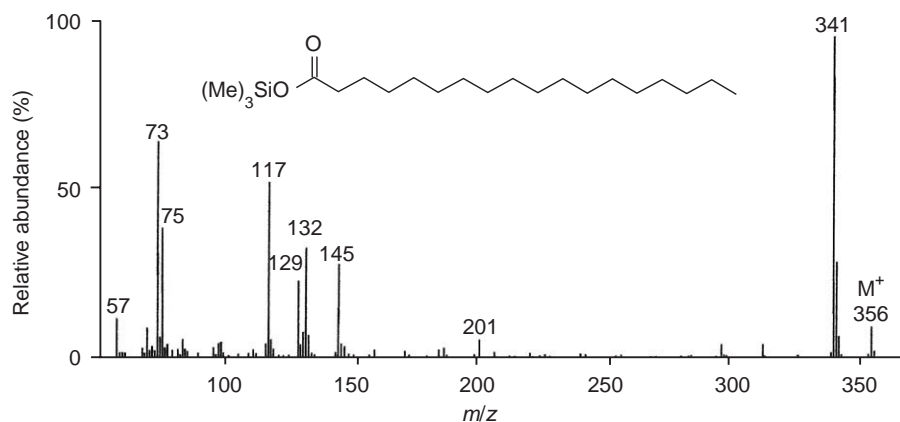
Another use for derivatization in quantitative GC–MS work is to produce spectra in which much of the ion current is diverted into a single ion in order to give high sensitivity for selected ion monitoring studies. Good derivatives for this purpose are those that contain the *tert*-butyldimethylsilyl (TBDMS) group. This derivative fragments preferentially by loss of the *tert*-butyl radical to give a very abundant ion at  $[M - 57]^+$  (**Figure 7**).

Derivatives used for GC–MS should preferably be simple to prepare and form in quantitative yield. Most of those listed in **Table 1** fulfill these criteria. In some cases, such as with the TMS ethers, preparation consists of little more than dissolving the sample in the reagent, possibly with heating, and direct injection of the mixture into the chromatograph.

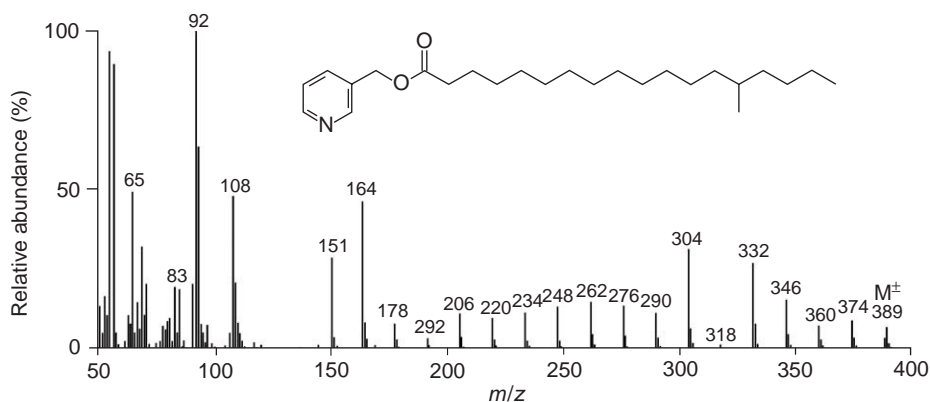
Sample volumes should be in the 0.1–2  $\mu$ l range, with  $\sim 10$ –1000 ng being optimum for the acquisition of scanned spectra. Although intuitively obvious, it should be emphasized that samples should be prepared in sample vials of the appropriate size for holding 10–100  $\mu$ l of solution, of which many can be

**Table 1** Derivatives suitable for GC–MS

Functional group	Derivative	Comments
Hydroxyl	TMS ether	Good GC and MS properties
	TBDMS ether	$[M - 57]^+$ ion for high sensitivity, longer retention time than TMS
	Methyl ether	Good GC and MS properties. Not as easy to prepare as TMS
	Higher alkyl	Used as GC shift reagents (reagents used to alter the retention time of selected compounds).
	Acetyl	Easy to prepare. Sometimes relatively large increases in GC retention time
<i>cis</i> -Diol	Cyclic alkane boronate	Low mass increment
	Cyclic acetal	
Carboxylic acid	Methyl ester	Good GC performance. Relatively poor MS fragmentation
	TMS ester	Slightly acid labile
	Nitrogen-containing, e.g., pyridine	Good for structural studies of long-chain compounds
Amine	Acyl	Good
	Perfuroacyl	Good electron-capturing properties for high sensitivity negative ion detection
	Schiff bases	Less stable than TMS ethers of alcohols
	TMS	
Carbonyl	Alkyloxime	Can give two GC peaks ( <i>syn</i> and <i>anti</i> )
	Schiff bases	
Phosphate	TMS	Spectra contain many rearrangement ions, e.g., $m/z$ 299
	Methyl ester	More stable than TMS derivatives
Sulfate	Unstable	Usually detected by replacement following derivatization of other functional groups to give a mixed derivative such as TMS/acetyl



**Figure 5** EI (70 eV) mass spectrum of the TMS derivative of octadecanoic (stearic) acid. The four ions at  $m/z$  117, 129, 132, and 145 are a characteristic feature of the TMS derivatives of fatty acids, but there are few ions that provide information on the structure of the aliphatic chain.



**Figure 6** EI (70 eV) mass spectrum of the picolinyl ester of 14-methyloctadecanoic acid. Abstraction of a proton from different positions of the aliphatic chain produces radical sites that induce chain cleavage. The position of the methyl group is revealed by the absence of the even-mass ion at  $m/z$  318 and the relatively high abundance of the ions at  $m/z$  304 and 332.

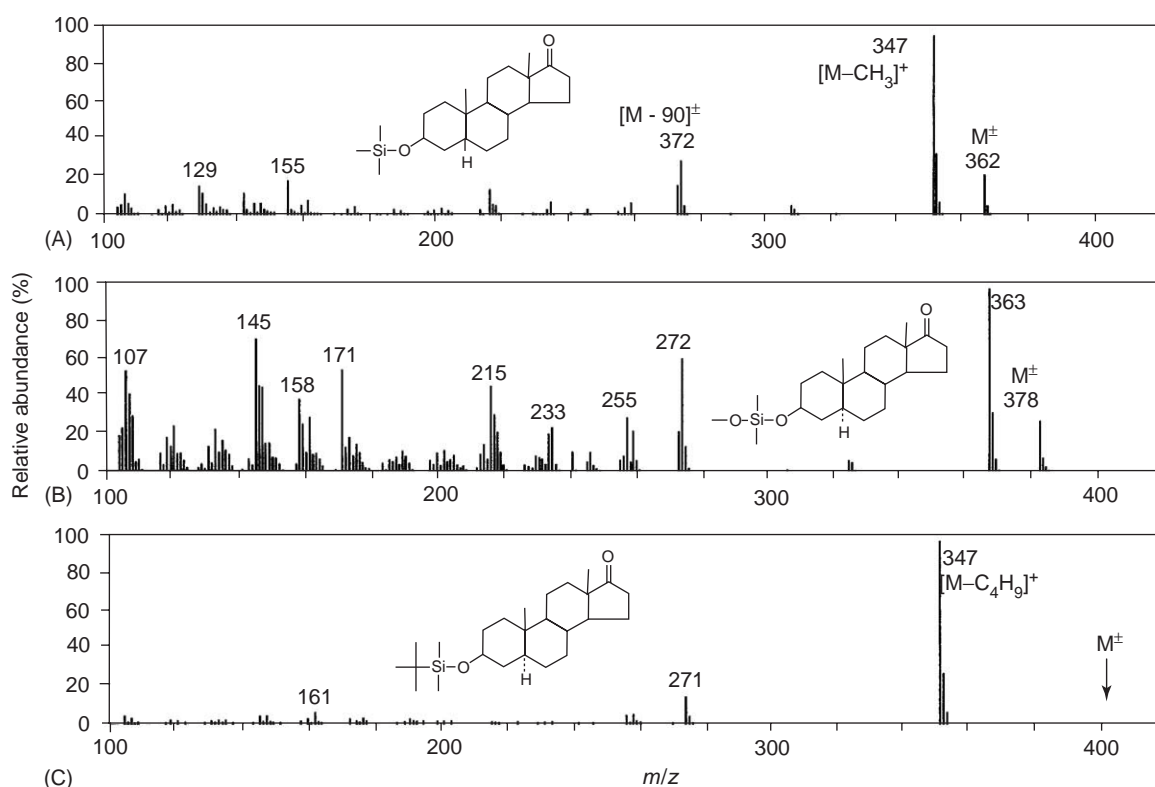
obtained commercially. Sample vials made for autosamplers containing an insert with a small capacity are ideal. The large (10 ml) sample tubes that are frequently used for sample preparation should be avoided as containers for the final sample solution because the small volume of solvent appropriate for the sample will evaporate into the large container volume, causing loss of control over the sample concentration. Where larger volumes of solvent have been used during work-up and subsequently evaporated, the sample should be transferred to a small sample vial before analysis.

Plastic apparatus is also to be avoided when organic solvents are involved as plasticizers are easily leached out of such materials by the solvent and, after concentration of the sample, yield spectra that frequently mask that from the analyte. Dialkyl phthalates are the plasticizers most commonly encountered and can be recognized by their production

of major ions at  $m/z$  149 and 167. Teflon is free from this problem and is frequently used as a liner for the closures of sample vials. Polypropylene microcentrifuge tubes also appear to be relatively free from phthalates and are finding increased use in GC-MS work. Commercial solvents very often contain traces of these and related compounds that can become problematical after sample concentration. Thus, all solvents used for extraction should be redistilled before use.

With very low concentrations of analyte, adsorption onto glass surfaces becomes a problem. This problem is best avoided by silanizing the glassware before use. A suitable method is to soak the glass in a 10% solution of dichlorodimethylsilane in toluene for  $\sim 1$  h, rinse with dry toluene and then with methanol. The silane forms dimethylsilyl bridges between adjacent hydroxyl groups on the glass surface, and the methanol converts the chlorine atom of





**Figure 7** Effect of derivatization on the 25 eV EI mass spectral fragmentation of 3 $\beta$ -hydroxy-5 $\alpha$ -androstan-17-one: (top) TMS derivative; (centre) methoxydimethylsilyl derivative showing more extensive fragmentation; (bottom) *tert*-butyldimethylsilyl derivative with a large  $[M - 57]^+$  ion suitable for high-sensitivity monitoring.

dimethylsilane groups that are bonded to only one oxygen atom into a methoxy group. Deactivated glass sample vials are now available commercially.

## Use of Stable Isotopes

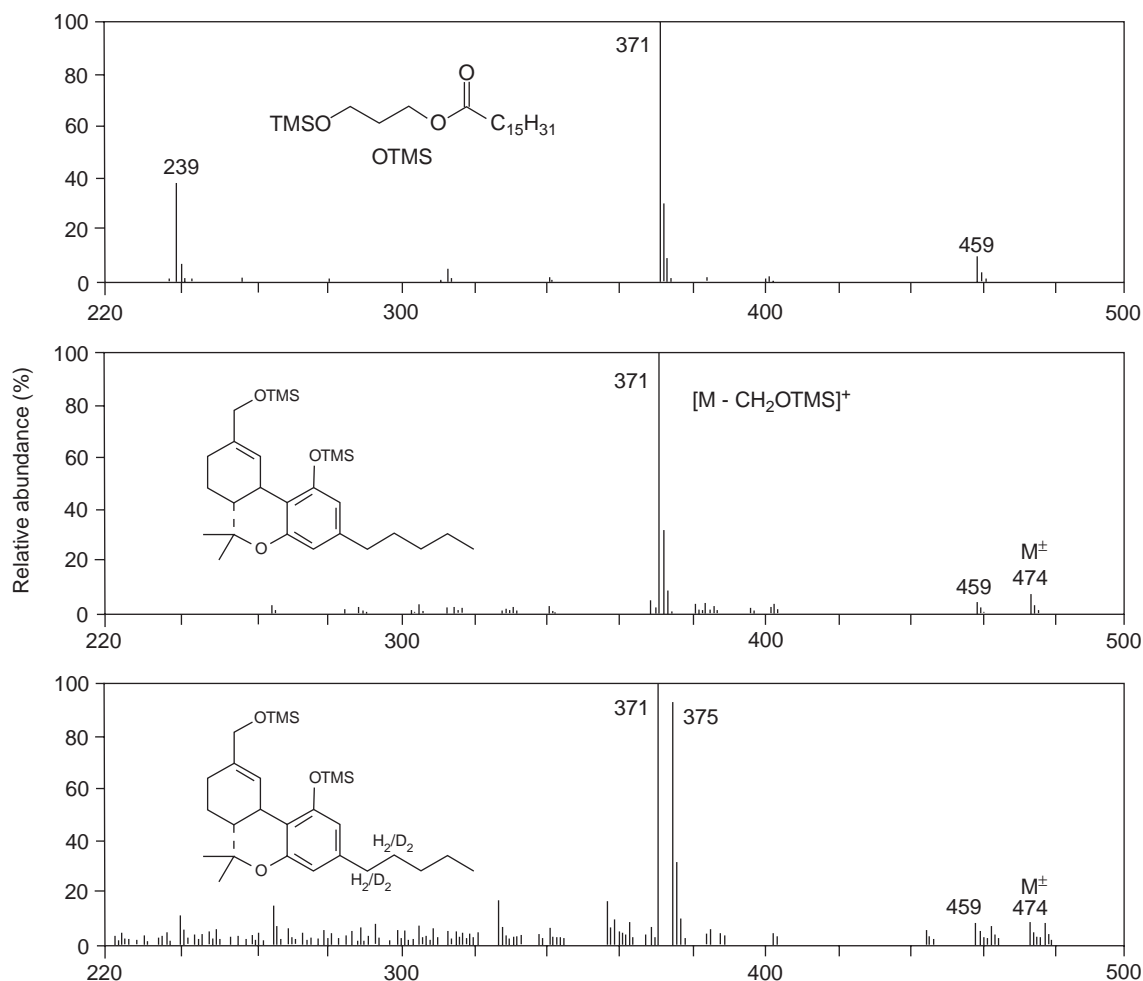
In addition to their use as internal standards, compounds labeled with stable isotopes have numerous other uses in GC-MS. Thus, the numbers of functional groups in a molecule can be determined by measuring the mass difference between two portions of the sample that have been derivatized with an unlabeled and a labeled reagent. TMS and [ $^2\text{H}_9$ ]TMS derivatives are particularly useful for hydroxyl and carboxyl groups. Where a derivative such as TMS, which reacts with different functional groups (i.e., alcohols, carboxylic acids, etc.), has been used, the preparation of group-specific derivatives such as methyl esters can subsequently be used to differentiate them. Another example is use of the isotope doublet technique in cases where it is not easy to identify sample (e.g., drug)-related peaks in the TIC from peaks produced by endogenous compounds. This technique involves the use of a 1:1 mixture of the

drug and a labeled variant as the test substance and to search the resulting GC profile for spectra containing doublet peaks. An example is shown in Figure 8.

Finally, stable isotopes are frequently used in clinical studies, where their main advantage over radioactive isotopes is that they present no hazard to the patient and can be used quite safely in children. Many applications have been published such as their use to monitor the pharmacokinetics of a single dose of drug during chronic therapy or to measure the bioavailability of a drug from various formulations.

## Future Trends

GC-MS has reached the state of a mature technique and one that the use of modern data systems has made relatively easy to use. Further increases in sensitivity are always welcome and are currently being addressed by the use of TOF instruments. High sensitivity, in turn, allows narrower GC columns with higher resolution and shorter run times to be used. It is probable that the recent introduction of TOF instrumentation into this field will lead to the



**Figure 8** Use of the isotope doublet technique. (Top) EI (25 eV) mass spectrum of the *bis*-TMS derivative of monopalmitin. (Centre) EI (25 eV) mass spectrum of the *bis*-TMS derivative of the major hydroxy metabolite of  $\Delta^9$ -tetrahydrocannabinol. The two compounds have the same molecular weight and their spectra contain the same base peak. (Bottom) The spectrum of the metabolite obtained following the use of a 1:1 molar ratio of the drug and its  $[^2\text{H}_4]$ -analog.

advancement of the technique for the analysis of even smaller amounts of material in shorter times than is currently possible.

See also: **Derivatization of Analytes. Gas Chromatography:** Column Technology; High-Temperature Techniques; High-Speed Techniques; Detectors; Environmental Applications; Forensic Applications; Petrochemical Applications. **Mass Spectrometry:** Ionization Methods Overview; Mass Separation; Selected Ion Monitoring.

## Further Reading

- Blau K and Halket J (eds.) (1993) *Handbook of Derivatives for Chromatography*. New York: Wiley.
- Chapman JR (1985) *Practical Organic Mass Spectrometry*. Chichester: Wiley.
- De Hoffmann E, Charette J, and Stroobant V (1996) *Mass Spectrometry: Principles and Applications*. Chichester: Wiley.
- Gudzinicwicz BJ, Gudzinicwicz MJ, and Martin HF (1976) *Fundamentals of Integrated Gas Chromatography–Mass Spectrometry Part II, Mass Spectrometry* (vol. 7, Chromatographic Science Series). New York: Dekker.
- Hübschmann H-J (2001) *Handbook of GC/MS: Fundamentals and Applications*. Chichester: Wiley.
- Karasek FW and Clement RE (1991) *Basic Gas Chromatography/Mass Spectrometry – Principles and Techniques*. New York: Elsevier.
- Kitson FG, Larsen BS, and McEwen CN (1996) *Gas Chromatography and Mass Spectrometry: A Practical Guide*. London: Academic Press.
- Knapp DR (1979) *Handbook of Analytical Derivatization Reactions*. New York: Wiley.
- Linskens HF and Jackson JF (eds.) (1986) *Gas Chromatography/Mass Spectrometry*. New York: Springer.

- McFadden W (1973) *Techniques of Combined Gas Chromatography/Mass Spectrometry: Applications in Organic Analysis*. New York: Wiley.
- Mastovská K and Lehotay SJ (2003) Practical approaches to fast gas chromatography–mass spectrometry. *Journal of Chromatography A* 1000: 153–180.
- Message GM (1984) *Practical Aspects of Gas Chromatography/Mass Spectrometry*. New York: Wiley.
- Millard BJ (1978) *Quantitative Mass Spectrometry*. London: Heyden.
- Pfleger K, Maurer HH, and Weber A (1985) *Mass Spectral and GC Data of Drugs, Poisons, Pesticides, Pollutants and Their Metabolites*, 2nd edn. Weinheim: Wiley-VCH.
- Ragunathan N, Krock KA, Klawun C, Sasaki TA, and Wilkins CL (1999) Gas chromatography with spectroscopic detectors. *Journal of Chromatography A* 856: 349–397.

## Glossary

Chemical ionization	A method for producing ions in which a gas such as methane or ammonia is first ionized by electron impact and the resulting ions are used to ionize the sample by collision.
Double focusing	The use of both electric and magnetic fields to focus ions with high resolution.
Electron impact	A method for producing ions by bombarding the sample molecules with electrons.
Maximum entropy	A computer algorithm that can be used to enhance resolution given knowledge of the expected half-height peak width.

## Fourier Transform Infrared Spectroscopy

P R Griffiths, University of Idaho, Moscow, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1865–1871, © 1995, Elsevier Ltd.

Knowing the identity of each component in a mixture is necessary for many analytical-scale separations, and simply measuring retention data for this purpose is often too ambiguous for the identification of molecules eluting from a capillary gas chromatography (GC) column, which has the capability of resolving several hundred components. Prior knowledge about the chemical structure of the components and spiking of the mixture with one or more reference standards may aid the identification process; however, a less ambiguous identification can be accomplished by interfacing the chromatograph to a sensitive, rapid-scanning spectrometer to obtain unique signatures of each component. This instrument should allow each component to be detected in real time without any loss in chromatographic resolution. Mass spectrometry (MS) is the most commonly applied technique for this purpose, but it has certain limitations, in particular for distinguishing between structural isomers, such as *ortho*-, *meta*-, and *para*-xylene, whose electron-impact and chemical-ionization mass spectra are identical. For such molecules a complementary technique to MS is desired. Fourier transform infrared (FTIR) spectrometry, which yields unique spectra for most structural isomers, has frequently been used as an alternative technique for this purpose.

## Light-Pipe-Based GC–IR Instruments

### Measurement of the Spectrum

The coupling of gas chromatographs and FTIR spectrometers (GC–FTIR) has been accomplished by three approaches. In the first, and by far the simplest, the GC column is connected directly to a heated flow-through cell. For capillary GC, this cell is usually fabricated from a 10-cm length of heated glass tubing with an internal diameter of  $\sim 1$  mm. The inside bore of this tube is coated with a thick enough film of gold to be highly reflective to infrared radiation. Infrared-transparent windows (such as potassium bromide) are attached to both ends of the tube. Infrared radiation entering one window is multiply reflected down the gold-coated interior bore before emerging from the other window, giving rise to the name *light-pipe* for this device. The effluent from the GC column is passed into one end of the tube and out of the other via heated fused-silica transfer lines. The entire unit is held at a temperature between 250 and 300°C to preclude the condensation of semivolatile materials.

Infrared radiation from an incandescent source, such as an SiC Globar, is collimated and passed through a rapid-scanning interferometer so that each wavelength in the spectrum is modulated at a different frequency. The beam of radiation is then focused onto the first window of the light-pipe and the infrared beam emerging from the second window is refocused onto a sensitive detector (typically a liquid-nitrogen-cooled mercury cadmium telluride (MCT) photoconductive detector). A typical system is

- McFadden W (1973) *Techniques of Combined Gas Chromatography/Mass Spectrometry: Applications in Organic Analysis*. New York: Wiley.
- Mastovská K and Lehotay SJ (2003) Practical approaches to fast gas chromatography-mass spectrometry. *Journal of Chromatography A* 1000: 153–180.
- Message GM (1984) *Practical Aspects of Gas Chromatography/Mass Spectrometry*. New York: Wiley.
- Millard BJ (1978) *Quantitative Mass Spectrometry*. London: Heyden.
- Pfleger K, Maurer HH, and Weber A (1985) *Mass Spectral and GC Data of Drugs, Poisons, Pesticides, Pollutants and Their Metabolites*, 2nd edn. Weinheim: Wiley-VCH.
- Ragunathan N, Krock KA, Klawun C, Sasaki TA, and Wilkins CL (1999) Gas chromatography with spectroscopic detectors. *Journal of Chromatography A* 856: 349–397.

## Glossary

Chemical ionization	A method for producing ions in which a gas such as methane or ammonia is first ionized by electron impact and the resulting ions are used to ionize the sample by collision.
Double focusing	The use of both electric and magnetic fields to focus ions with high resolution.
Electron impact	A method for producing ions by bombarding the sample molecules with electrons.
Maximum entropy	A computer algorithm that can be used to enhance resolution given knowledge of the expected half-height peak width.

## Fourier Transform Infrared Spectroscopy

P R Griffiths, University of Idaho, Moscow, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1865–1871, © 1995, Elsevier Ltd.

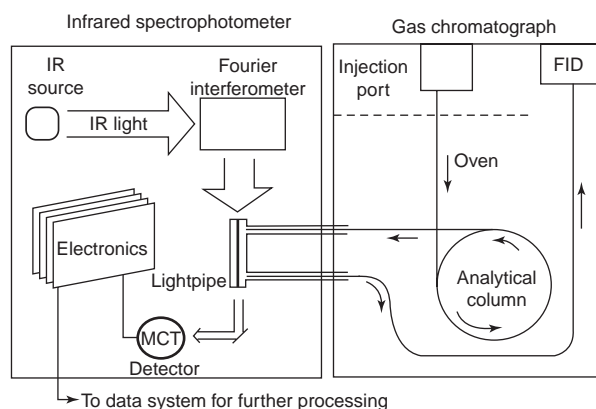
Knowing the identity of each component in a mixture is necessary for many analytical-scale separations, and simply measuring retention data for this purpose is often too ambiguous for the identification of molecules eluting from a capillary gas chromatography (GC) column, which has the capability of resolving several hundred components. Prior knowledge about the chemical structure of the components and spiking of the mixture with one or more reference standards may aid the identification process; however, a less ambiguous identification can be accomplished by interfacing the chromatograph to a sensitive, rapid-scanning spectrometer to obtain unique signatures of each component. This instrument should allow each component to be detected in real time without any loss in chromatographic resolution. Mass spectrometry (MS) is the most commonly applied technique for this purpose, but it has certain limitations, in particular for distinguishing between structural isomers, such as *ortho*-, *meta*-, and *para*-xylene, whose electron-impact and chemical-ionization mass spectra are identical. For such molecules a complementary technique to MS is desired. Fourier transform infrared (FTIR) spectrometry, which yields unique spectra for most structural isomers, has frequently been used as an alternative technique for this purpose.

## Light-Pipe-Based GC-IR Instruments

### Measurement of the Spectrum

The coupling of gas chromatographs and FTIR spectrometers (GC-FTIR) has been accomplished by three approaches. In the first, and by far the simplest, the GC column is connected directly to a heated flow-through cell. For capillary GC, this cell is usually fabricated from a 10-cm length of heated glass tubing with an internal diameter of ~1 mm. The inside bore of this tube is coated with a thick enough film of gold to be highly reflective to infrared radiation. Infrared-transparent windows (such as potassium bromide) are attached to both ends of the tube. Infrared radiation entering one window is multiply reflected down the gold-coated interior bore before emerging from the other window, giving rise to the name *light-pipe* for this device. The effluent from the GC column is passed into one end of the tube and out of the other via heated fused-silica transfer lines. The entire unit is held at a temperature between 250 and 300°C to preclude the condensation of semivolatile materials.

Infrared radiation from an incandescent source, such as an SiC Globar, is collimated and passed through a rapid-scanning interferometer so that each wavelength in the spectrum is modulated at a different frequency. The beam of radiation is then focused onto the first window of the light-pipe and the infrared beam emerging from the second window is refocused onto a sensitive detector (typically a liquid-nitrogen-cooled mercury cadmium telluride (MCT) photoconductive detector). A typical system is



**Figure 1** Schematic of typical light-pipe-based GC-FTIR interface (based on Hewlett Packard IRD).

illustrated schematically in **Figure 1**. The signal measured in this way is known as an *interferogram* and the Fourier transform of the interferogram yields a single-beam spectrum. By calculating the ratio of a single-beam spectrum measured when a component is present in the light-pipe to one measured when only the helium carrier gas is present, the transmittance spectrum,  $T(\nu)$ , of the component is obtained. The transmittance spectrum is usually converted to absorbance,  $A(\nu)$ , immediately by the standard Beer's law operation,  $A(\nu) = -\log_{10} T(\nu)$ , as the relative intensities of bands in absorbance spectra are independent of the concentration of the analyte, thereby allowing spectral library searching (*vide infra*) to be performed. For light-pipe-based GC-FTIR systems, it is rarely necessary to measure spectra at high resolution, as the spectral bands are quite broad. Since most bands in the spectra of molecules in the vapour phase have a width of at least  $10 \text{ cm}^{-1}$ , the typical resolution at which GC-FTIR spectra are measured is  $8 \text{ cm}^{-1}$ .

When operated at their highest scan speeds, FTIR spectrometers can measure between 5 and 20 interferograms per second that would yield spectra of this resolution. During a chromatographic analysis, interferograms are measured continuously. Thus for a 30-minute-long chromatogram, it would be possible to measure tens of thousands of interferograms, giving rise to an amount of data that would be too great to store on a disk on a typical personal computer (PC). Fortunately, most GC peaks have a full-width at half-height (FWHH) of several seconds. Thus it is common practice to average blocks of interferograms for a period of time that is slightly less than the FWHH of the narrowest peak in the chromatogram (usually 1–2 s). The single-beam spectrum is then computed from this signal-averaged

interferogram and ratioed against an appropriate background spectrum; finally, the resulting transmittance spectrum is converted to a linear absorbance format. On contemporary PCs this entire sequence of operations is performed while the next block of interferograms is being acquired.

### Reconstruction of Chromatograms

The end result of this process is that over 1000 absorbance spectra, corresponding to the contents of the light-pipe measured at  $\sim 1$ -second intervals throughout the entire chromatogram, are stored at the end of the run. Many of these spectra contain no useful information as they were measured when no component was present in the light-pipe; thus, the next step in a GC-FTIR analysis is to determine which of the stored spectra contain useful information. To achieve this, a chromatogram must be reconstructed from the spectroscopic data. This is usually achieved in two ways, the first of which is known as the *Gram-Schmidt* (GS) vector orthogonalization method. Here, short, information-rich regions of the interferograms are treated as vectors, and the vector distance between this part of each interferogram measured during the chromatographic run and several interferograms that were acquired when nothing except the helium carrier gas was flowing through the light-pipe (known as the basis set) is calculated. When an analyte elutes from the column, the magnitude of the vector difference is approximately proportional to the quantity of this material in the light-pipe. Because only a short region of the interferogram is examined, calculation of the GS 'signal' can be achieved in a few milliseconds. Furthermore, since all compounds besides monatomic and homonuclear diatomic molecules have at least one band in their infrared spectrum, GS chromatograms are very nonselective. Some compounds yield much stronger infrared spectra than others, however. For example, the spectra of most nonpolar compounds are rather weak, whereas the spectra of very polar compounds are usually much stronger. As an example, the detection limits for GS chromatograms of polycyclic aromatic hydrocarbons (which have very low absorptivities over most of their infrared spectra) are  $\sim 20$  times greater than the corresponding values for the barbiturates (which are very polar and have several strong infrared absorption bands in their spectra).

The other commonly used algorithm by which chromatograms are constructed from the infrared data involves calculating the integrated absorbance in one or more specified spectral regions. These regions are usually chosen to correspond to the



characteristic absorption frequencies of the functional group present in the class(es) of molecules of interest. The chromatograms generated by this approach have been called by a variety of names, including Chemigrams<sup>TM</sup>, functional group (FG) chromatograms and selective wavelength (SW) chromatograms. FG chromatograms are, of course, far more selective than GS chromatograms, but are rarely completely selective as many molecules have weak overtone and combination bands over much of the fingerprint region of the infrared spectrum. For compounds with functional groups giving rise to intense absorption bands, such as the C=O stretching mode of carbonyl compounds, the limits of detection of FG chromatograms may be less than those of the corresponding GS chromatograms, but the two algorithms often have comparable sensitivity. A useful way to detect the presence of a particular functional group is to compare the relative heights of peaks in the GS and FG chromatograms. If the ratio of the peak heights in the FG and GS chromatograms is large, the presence of that functional group in that component is indicated; if the ratio is small, there is a much smaller probability that the analyte contains that functional group.

### Spectral Searching

Once the chromatography has been completed, the spectra of those components of interest can be displayed. (In fact, several GC-FTIR software packages allow the spectra to be displayed even while data acquisition is still in progress.) Each component generating a peak in the GS chromatogram with a signal-to-noise ratio greater than  $\sim 10$  can usually be identified by comparing its spectrum to a library of vapor-phase reference spectra. The unknown and reference spectra are first scaled so that the most intense band in each spectrum has the same absorbance (usually 1.0). By treating the spectra as vectors, the Euclidean distance between the unknown and each reference spectrum is calculated. This distance is usually called the *hit quality index* (HQI); the smaller the HQI, the better the spectral match. The highest probability for the identification of the unknown is that of the compound in the reference library yielding the smallest HQI. However, unequivocal identifications cannot be made on this basis alone, for several reasons. The reference spectrum of the authentic analyte may not be present in the spectral library. If the spectrum of the unknown is noisy, the value of the HQI may be determined more by noise than by the true absorption spectrum. The reference spectrum may have been measured with the sample at a different temperature from the light-pipe, measured

at a slightly different resolution, or computed with a different apodization function. Finally, some members of homologous series can have very similar spectra, so it is not uncommon for compounds of the same type (e.g., methacrylate esters) to give similar HQI values.

It is always recommended that the user should make a side-by-side comparison of the GC-IR spectrum and the reference spectra of the top few 'hits' to get a good idea of the probability that the structure of the analyte is best represented by the structure of the top hit, or by one of the other close hits, or whether there is enough similarity between the GC-IR spectrum and all of the closely matching reference spectra that unequivocal identification is impossible. In this case, the simultaneous application of mass spectrometry may be necessary to yield an unequivocal identification.

### Limits of Detection and Identification

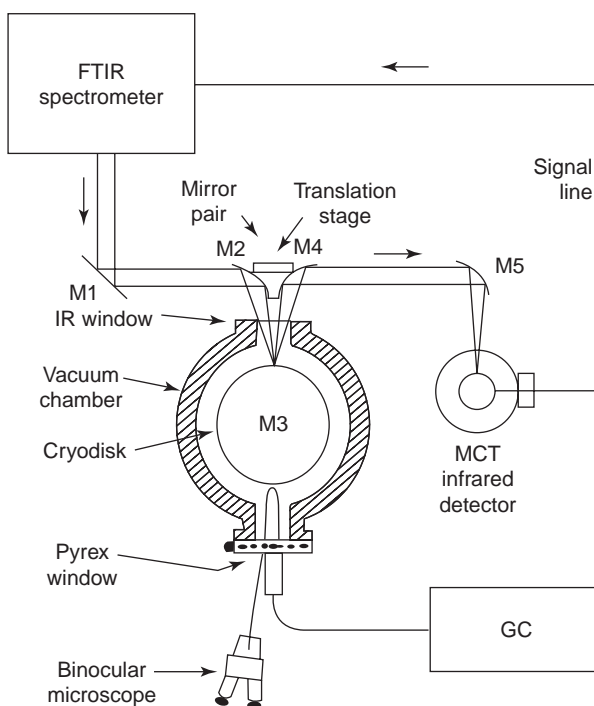
The limit of detection (LOD) for an acceptable GC-IR response for most compounds is between about 1 and 20 ng (injected) per component, the actual value depending on the chemical nature of the analyte. The LOD is often defined with respect to the strongest band in the spectrum. Most bands in the infrared spectrum of nonpolar compounds are fairly weak and so these compounds tend to have the highest LODs, but even these compounds usually have at least one band in the spectrum with a high absorptivity. Examples include the C-H stretching bands of alkanes and the aromatic C-H out-of-plane deformation bands of polycyclic aromatic hydrocarbons. Detection limits also depend on the width of the GC peak; the wider the peak, the more dilute the analyte and the higher the LOD.

The amount of a given component that must be injected into the chromatograph to yield an identifiable spectrum, often known as the minimum identifiable quantity (MIQ), depends not on the *strongest* band, but on the signal-to-noise ratio of the *most characteristic bands* in the spectrum. For an analyte with a spectrum that is very different from any other spectrum in the reference database, the MIQ may be only slightly higher than the LOD. On the other hand, there are often only very subtle differences between the spectra of members of this class of compounds. If the analyte is a member of a homologous series and several reference spectra of members of this series are contained in the library, the signal-to-noise ratio of the spectrum must be high, and hence the MIQ will be much greater than the LOD if the analyte is to be correctly identified.

If a minor peak is present in a chromatogram measured with a conventional GC detector such as a flame ionization detector (FID), but is not observable in the GS or FG chromatogram, it may be possible simply to inject a greater volume of the sample into the chromatograph. Even if the major components overload the GC column in this case, the minor components will not. However, sometimes the major peaks will broaden to the point that they start to overlap a neighboring minor peak. In this case, it may become necessary to subtract the spectrum of the major peak (linear in absorbance) from the spectrum measured in the region of the minor peak, to identify the minor component. This procedure is needed because of the relatively low sensitivity of light-pipe-based GC-IR instruments. Two other approaches have been described that have led to increased sensitivity for GC-IR measurements, and are described below.

### Matrix-Isolation GC-IR

In the first of these, argon is mixed with the helium mobile phase, either as a minor ( $\sim 1\%$ ) component in the carrier gas or by addition at the end of the GC column. The column effluent is then sprayed from a heated fused-silica transfer line onto a rotating gold-plated disk that is maintained at a temperature of less than 15 K. Helium does not condense at this temperature but argon does. By locating the end of the transfer line an appropriate distance from the cooled disk, argon is deposited as a track  $\sim 300\ \mu\text{m}$  in width. Any component emerging from the transfer line at the same time is trapped in the argon matrix. After the separation has been completed, the disk is rotated to a position where the focused beam from an FTIR spectrometer is transmitted through the track of argon, reflects from the gold-coated disk, passes again through the argon and then is collected and focused on to an MCT detector, as shown in Figure 2. In principle, if the concentration of any analyte in the argon matrix is low enough, each analyte molecule will be isolated from like molecules by the argon matrix. Despite the fact that the concentration is usually a little too high for true matrix isolation to be achieved in GC-IR measurements, this technique none the less is known as *matrix isolation GC-IR*. By rotating the disk slowly, a series of spectra can be measured that is analogous to the series of spectra that is measured in real time during a light-pipe-based GC-IR run and either GS or FG chromatograms can be constructed from these data. Each component may be identified by spectral library searching, but a special library of spectra of matrix-isolated standards is required.



**Figure 2** Schematic of matrix isolation GC-FTIR interface (based on Mattson Instruments Cryolect).

The advantages of matrix isolation are based on the following considerations. First, the width of the track is  $\sim 300\ \mu\text{m}$ , compared with 1 mm for the diameter of a light-pipe. Thus, the sample is more concentrated over the cross-sectional area of the IR beam and a given amount of sample will yield a spectrum with more intense absorption bands. Second, because each component is trapped on the disk, it is common practice when minor components are to be identified by matrix isolation GC-IR to signal-average interferograms for several minutes with the disk stationary, enabling a significant increase in sensitivity to be achieved over real-time measurements. A final advantage that has been claimed for matrix-isolation GC-IR measurements is the increase in the absorptivity at the peak of each band in the spectrum because of the decrease in bandwidth that occurs on matrix isolation (the band area remaining approximately constant). This is true for small molecules, but large molecules disrupt the crystal structure of the argon to such an extent that a certain amount of molecular motion is possible. As a result, the widths of many bands in the spectra of large asymmetric molecules prepared in this way are surprisingly similar to widths of corresponding bands in the spectra of the corresponding molecules prepared as KBr disks.

The exception to this behavior is observed in the spectra of compounds that contain O-H or N-H

groups. In the crystalline form of such compounds, the O–H and N–H groups are strongly intermolecularly hydrogen bonded. As a consequence, the O–H and N–H stretching bands in their KBr-disk spectra are exceptionally broad, often having a width of several hundred wavenumbers. When these molecules are isolated in an argon matrix, however, no intermolecular hydrogen bonding takes place, and the O–H and N–H stretching bands appear as very narrow spectral features. Thus, when the spectra of matrix-isolated species such as alcohols, phenols, or amines are measured at high resolution, excellent specificity is often gained by matrix isolation GC–FTIR.

The major problem with this approach to GC–IR (which can to a certain extent be shared with vapor-phase measurements) is the lack of extensive libraries of appropriate reference spectra. This disadvantage has largely been overcome by the final type of GC–IR interface to be described below.

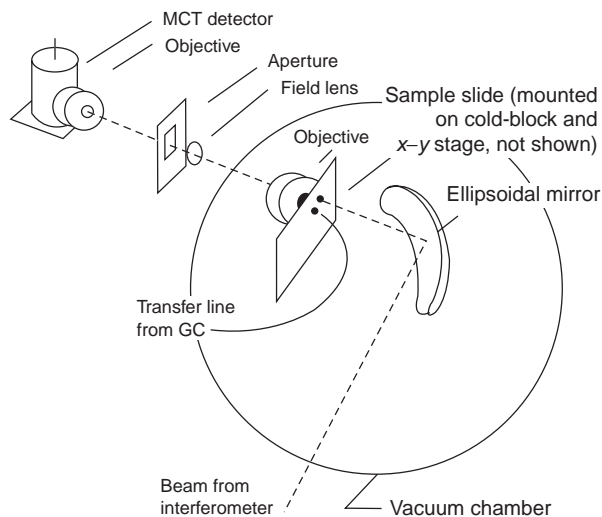
### Direct Deposition GC–IR

In the remaining approach to GC–FTIR, the effluent from the column is directed at a slowly moving, cooled window mounted on a computer-controlled  $x$ – $y$  stage. Zinc selenide cooled to the temperature of liquid nitrogen is the most commonly used substrate. Each eluting component is deposited on the window as a very narrow spot. In the commercially available form of this interface, shown in Figure 3, the typical width of each spot is  $\sim 100\text{ }\mu\text{m}$ . The stage moves so that each deposited component passes through the beam focus of an infrared microspectrometer shortly after deposition. As for light-pipe-based GC–IR systems, spectra are measured continuously throughout

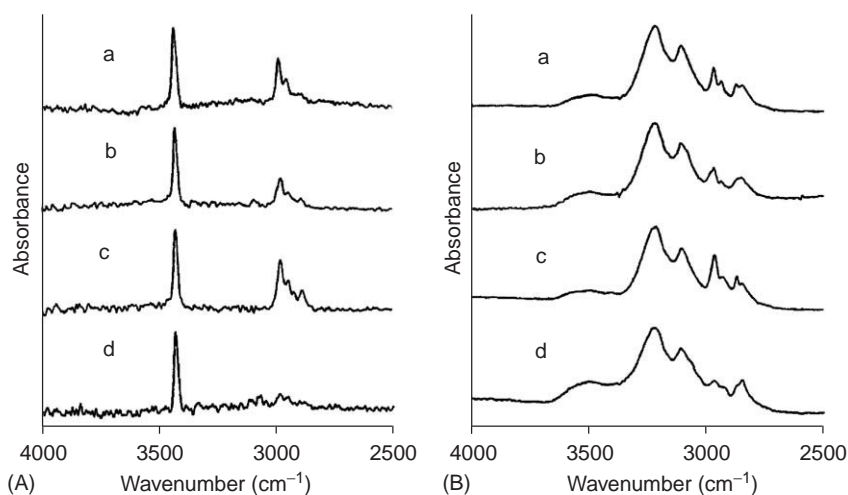
the chromatographic run and GS and/or FG chromatograms can be output in real time. This *direct deposition* approach for GC–IR has two important advantages over light-pipe or matrix-isolation GC–IR systems – it yields higher sensitivity and the measured spectra are very similar to reference spectra of standards prepared as KBr disks. Let us first recognize the reason for the increased sensitivity of direct deposition GC–IR measurements.

As seen in the previous section, the smaller the cross-sectional area of the sample, the greater the absorbance of all bands in the spectrum. Because the sample is contained in a  $100\text{-}\mu\text{m}$  diameter spot rather than a 1-mm diameter light-pipe, its cross-sectional area is 100 times less, so that bands will be  $\sim 100$  times more intense. To attain the optimal sensitivity, the diameter of the IR beam should be approximately equal to the width of the spot, i.e.,  $\sim 100\text{ }\mu\text{m}$ , and a detector of the same size should also be used. Several other optical factors should be included in the comparison, but in general it is found that the signal-to-noise ratio of GC–IR spectra measured online using the direct deposition technique is  $\sim 50$  times greater than the corresponding measurement made using a light-pipe system. The sensitivity advantage of direct-deposition GC–FTIR systems can be further increased by post-run signal averaging in a manner analogous to the matrix-isolation GC–IR system described above. If each real-time spectrum is measured over 1-second blocks, post-run averaging for just 1 min will yield an improvement in sensitivity of almost a factor of 8.

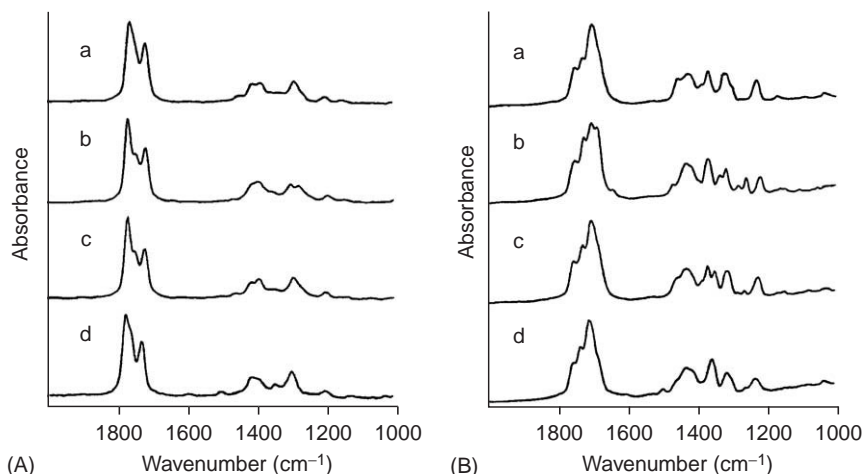
As noted above, the MIQ for direct deposition GC–IR measurements varies with the polarity of the analyte. The LOD for real-time measurements of several analytes by this technique was  $\sim 50\text{ pg}$ . When very polar analytes are injected, this number can be further reduced. For example, the LOD for several barbiturates was found to be  $\sim 13\text{ pg}$ . Spectra of these barbiturates in the high-wavenumber region measured by a light-pipe-based GC–FTIR instrument (Hewlett Packard IRD) and a direct-deposition system (Bio-Rad/Digilab Tracer) are shown in Figure 4. Differences between the vapor-phase and condensed-phase spectra of molecules that can exhibit strong intermolecular hydrogen bonding are readily apparent in this figure. For example, the sharp bands absorbing near  $3430\text{ cm}^{-1}$  in the vapor-phase spectra are due to the N–H stretching vibration of isolated (non-hydrogen-bonded) molecules. In the corresponding condensed-phase spectra measured by direct deposition GC–FTIR, the N–H stretching modes of the intermolecularly hydrogen-bonded barbiturates are seen as broad bands near  $3220$  and  $3110\text{ cm}^{-1}$ . Similar differences between vapor-phase



**Figure 3** Schematic of direct deposition GC–FTIR interface (based on Bio-Rad/Digilab Tracer).



**Figure 4** (A) Flow-cell and (B) direct-deposition GC-FTIR spectra of (a) barbital, (b) aprobarbital, (c) butabarbital, and (d) phenobarbital from 4000 to 2500  $\text{cm}^{-1}$ ; 12.5 ng and 375  $\mu\text{g}$  of each component were injected for the light-pipe and direct-deposition spectra, respectively.



**Figure 5** Low-wavenumber region of the spectra shown in Figure 4.

and condensed-phase spectra of barbiturates are also seen in the spectral region between 2000 and 1000  $\text{cm}^{-1}$  (see Figure 5). The difference between the sensitivity of the light-pipe and direct-deposition GC-FTIR measurements can be seen by comparing the noise levels of the spectra shown in Figure 4 and recognizing the fact that it required 30 times more of each barbiturate to be injected for the spectra measured using a light-pipe than by direct deposition.

On deposition, the molecules of a given analyte form randomly oriented crystallites on the zinc selenide window. These crystallites are similar to the crystallites that are formed on grinding of solid samples during the preparation of KBr disks or mineral-oil mulls. Not surprisingly, therefore, the spectra of compounds obtained by direct deposition GC-IR are very similar to the KBr-disk spectra of the

corresponding compounds. Extensive libraries (>150 000 entries) of reference spectra of standards prepared in this way are available commercially. The only compounds that cannot be readily identified in this manner are molecules with very strongly hydrogen-bonding groups or for analytes exhibiting polymorphism. For trace analytes containing O-H or N-H groups, the best results on library searching are usually found by examining only the spectral region below 2000  $\text{cm}^{-1}$  and eliminating the region containing the strong, broad O-H and N-H stretching modes from the search.

## Prognostication

On-line infrared spectrometry is proving to be an important way of identifying molecules eluting from

a gas chromatograph. Light-pipe-based systems are the simplest, least expensive and most reliable, but often prove to have inadequate sensitivity for the identification of minor components. Of the two deposition-based techniques, the direct deposition approach has limits of detection that rival those of benchtop GC–MS systems and has the great advantage of producing spectra that are directly comparable to KBr-disk reference spectra, of which there are over 150 000 available in digital form (i.e. suitable for computerized library searching). Thus one can forecast an increasing use of systems based on this principle in the future. It is also noteworthy that interfaces between FTIR spectrometers and both supercritical-fluid and liquid chromatographs based on the same principle have been described.

See also: **Gas Chromatography: Mass Spectrometry.**

## Further Reading

- Bourne S, Haefner AM, Norton KL, and Griffiths PR (1990) Performance characteristics of a real-time direct deposition GC/FT-IR system. *Analytical Chemistry* 62: 2448–2452.
- Griffiths PR and de Haseth JA (1986) In: *Fourier Transform Infrared Spectrometry*, chap. 18, New York: Wiley-Interscience.
- Herres W (1987) *HRGC-FTIR: Capillary Gas Chromatography – Fourier Transform Infrared Spectroscopy; Theory and Application*. Heidelberg: Huethig.
- Reedy GT, Ettinger DG, Schneider JF, and Bourne S (1985) High resolution gas chromatography/matrix isolation infrared spectrometry. *Analytical Chemistry* 57: 1602–1609.
- White R (1990) *Chromatography/Fourier Transform Infrared Spectroscopy and its Applications*. New York: Dekker.

## Physicochemical Measurements

**T Paryczak, J Rynkowski, and M I Szynkowska,**  
Technical University of Lodz, Lodz, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Nonanalytical applications of gas chromatography (GC) in the study of various phenomena in gases, solutions, and on solid surfaces are set out in **Figure 1**. They allow investigation of several phenomena in the same relatively simple apparatus, whereas other much more costly methods usually only permit the study of one such process.

The other main advantages of GC over static methods in investigating physicochemical phenomena are:

- GC is mainly used for examining very small samples at near-infinite dilution. Such concentrations can be applied directly to investigations of thermodynamic equilibrium.
- The accuracy of measurement of physicochemical phenomena by chromatography is at least as high as that of other methods, and the rate at which measurements can be made is much higher.
- Physicochemical properties may be measured over a wide range of temperature.

The accuracy of chromatographic results depends on whether a specific or relative retention volume,

retention time, or peak width or area is used for estimation. The accuracy is greatest for the retention time, lower for the peak width, and lowest of all for the peak area.

The precision of physicochemical measurements by GC relies on the ability of the instrument to control and measure all parameters relating to the chromatographic process. Sources of inaccuracy include the determination of stationary phase mass, column temperature and its gradients, flow rate and stability, pressure drop along the column, ‘dead space’ in the system, etc. Sophisticated computerized equipment allows precise measurement and control of all these operating parameters.

High purity of the carrier gas is essential, since even trace amounts of a contaminant that may be adsorbed will, as a result of accumulation, badly distort the results.

In analytical GC, particular emphasis is laid on the separation of the mixture into its components, which are recorded on the chromatogram as a series of peaks or steps. On the other hand, where GC is used as a physicochemical research tool, we are usually interested only in a single peak and how this alters with changes in the experimental conditions. The task is therefore to determine the effect of elementary processes taking place in the column on the shape of the output curve, and conversely, the possibility of identifying the course of elementary processes in the column on the basis of a given output curve. The initial purity of the sample is not essential in most



a gas chromatograph. Light-pipe-based systems are the simplest, least expensive and most reliable, but often prove to have inadequate sensitivity for the identification of minor components. Of the two deposition-based techniques, the direct deposition approach has limits of detection that rival those of benchtop GC–MS systems and has the great advantage of producing spectra that are directly comparable to KBr-disk reference spectra, of which there are over 150 000 available in digital form (i.e. suitable for computerized library searching). Thus one can forecast an increasing use of systems based on this principle in the future. It is also noteworthy that interfaces between FTIR spectrometers and both supercritical-fluid and liquid chromatographs based on the same principle have been described.

See also: **Gas Chromatography: Mass Spectrometry.**

## Further Reading

- Bourne S, Haefner AM, Norton KL, and Griffiths PR (1990) Performance characteristics of a real-time direct deposition GC/FT-IR system. *Analytical Chemistry* 62: 2448–2452.
- Griffiths PR and de Haseth JA (1986) In: *Fourier Transform Infrared Spectrometry*, chap. 18, New York: Wiley-Interscience.
- Herres W (1987) *HRGC-FTIR: Capillary Gas Chromatography – Fourier Transform Infrared Spectroscopy; Theory and Application*. Heidelberg: Huethig.
- Reedy GT, Ettinger DG, Schneider JF, and Bourne S (1985) High resolution gas chromatography/matrix isolation infrared spectrometry. *Analytical Chemistry* 57: 1602–1609.
- White R (1990) *Chromatography/Fourier Transform Infrared Spectroscopy and its Applications*. New York: Dekker.

## Physicochemical Measurements

**T Paryczak, J Rynkowski, and M I Szynkowska,**  
Technical University of Lodz, Lodz, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Nonanalytical applications of gas chromatography (GC) in the study of various phenomena in gases, solutions, and on solid surfaces are set out in **Figure 1**. They allow investigation of several phenomena in the same relatively simple apparatus, whereas other much more costly methods usually only permit the study of one such process.

The other main advantages of GC over static methods in investigating physicochemical phenomena are:

- GC is mainly used for examining very small samples at near-infinite dilution. Such concentrations can be applied directly to investigations of thermodynamic equilibrium.
- The accuracy of measurement of physicochemical phenomena by chromatography is at least as high as that of other methods, and the rate at which measurements can be made is much higher.
- Physicochemical properties may be measured over a wide range of temperature.

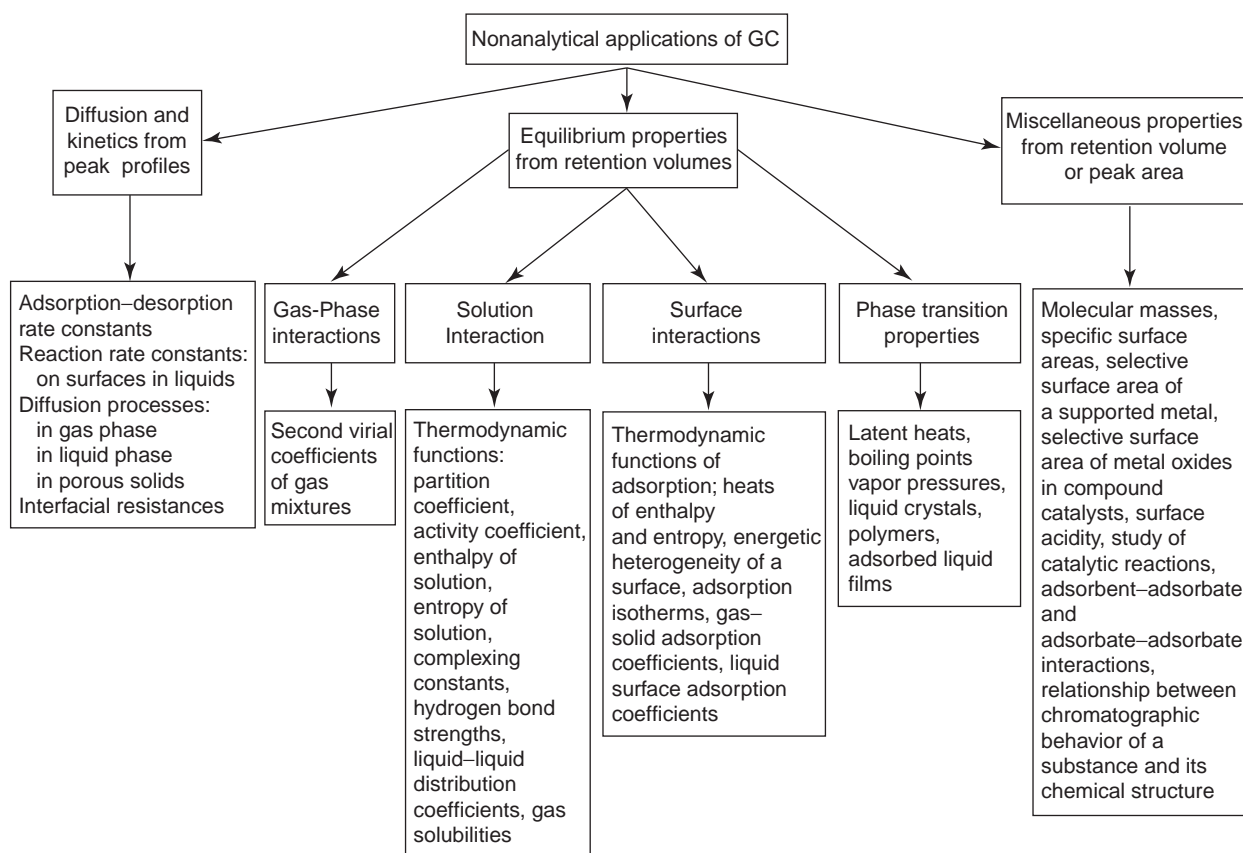
The accuracy of chromatographic results depends on whether a specific or relative retention volume,

retention time, or peak width or area is used for estimation. The accuracy is greatest for the retention time, lower for the peak width, and lowest of all for the peak area.

The precision of physicochemical measurements by GC relies on the ability of the instrument to control and measure all parameters relating to the chromatographic process. Sources of inaccuracy include the determination of stationary phase mass, column temperature and its gradients, flow rate and stability, pressure drop along the column, ‘dead space’ in the system, etc. Sophisticated computerized equipment allows precise measurement and control of all these operating parameters.

High purity of the carrier gas is essential, since even trace amounts of a contaminant that may be adsorbed will, as a result of accumulation, badly distort the results.

In analytical GC, particular emphasis is laid on the separation of the mixture into its components, which are recorded on the chromatogram as a series of peaks or steps. On the other hand, where GC is used as a physicochemical research tool, we are usually interested only in a single peak and how this alters with changes in the experimental conditions. The task is therefore to determine the effect of elementary processes taking place in the column on the shape of the output curve, and conversely, the possibility of identifying the course of elementary processes in the column on the basis of a given output curve. The initial purity of the sample is not essential in most



**Figure 1** Nonanalytical applications of GC. (Reprinted with permission from Valko *et al.* (1984) *Journal of Chromatography*, 301: 355–364; © Elsevier.)

cases and, with appropriate experimental design several physicochemical measurements can be made simultaneously on the different components in a mixture.

The appropriate sample size depends on whether the investigations are carried out in the range of 'infinite dilution' or of 'finite-concentration'. At infinite dilution, the concentrations of the components in the carrier gas may be neglected and thus the gaseous phase may be considered ideal. At finite concentration, a correction for retention calculations should be introduced, which makes them more complicated. Certain limitations of GC methods for physicochemical determinations should also be mentioned. These are mainly restricted to the study of interactions that occur on solids, in liquids, in the mobile gas phase, and at their interfaces. While measurements can be made simultaneously, it is possible that interference from other physicochemical properties of the materials may introduce inaccuracy. Another important limitation is the volatility of the stationary phase in gas-liquid chromatography.

It is impossible in this brief article to describe all the nonanalytical applications of GC presented in

**Figure 1.** They have been described in full detail in numerous monographs. Thus, the use of GC in determining some selected representative properties and parameters of particular importance in this wide field of chromatographic applications is discussed.

## Surface Characteristics

### Adsorption Isotherm

The adsorption isotherm (a plot of the concentration distribution of the solute in the mobile gas and stationary solid phases at a given temperature) is the foundation upon which the surface characteristics of adsorbents are defined. From this one can determine the specific surface area, capillary distribution (from the desorption curve), porosity, and other properties of a solid. Using adsorbates with diverse physical and chemical characteristics, it is possible to define the type of adsorbate-adsorbent interactions involved and the nature of adsorption in the system being examined.

Retention of the solute is determined by the overall equilibrium distribution between the stationary and

mobile phases. The linear isotherm is reflected in the distribution constant remaining unchanged. It may be obtained over a limited concentration range and in the absence of selective interactions between adsorbate and adsorbent. The result is a Gaussian distribution of the solute in the column eluate. If such interactions take place or if higher concentrations are employed, the sorption isotherm becomes curved, leading to an asymmetric, broadened peak shape (Figure 2). There are many methods of determining adsorption isotherms. They have been elaborated both for the case of ideal GC (elution at infinite dilution) and nonlinear nonideal GC (finite concentration), taking into consideration sorption effects, pressure gradients, gas imperfections, and other factors. Regardless of which chromatographic technique is used to determine the adsorption isotherm, the problem is to find the amount of adsorbed solute and its corresponding equilibrium concentration (pressure) in the mobile phase for each point of the isotherm. The desired data may be obtained either from a series

of adsorbate samples of varying size injected into the column (the peak maxima method) or from a single chromatogram (elution by characteristic point) and frontal analysis by characteristic point methods.

### Heat and Other Thermodynamic Functions of Adsorption

The use of temperature as a variable yields information on the thermodynamics of the adsorption-desorption process. A thermodynamic parameter of great importance measurable by GC techniques is the heat of adsorption. It provides a quantitative measure of the interactions occurring between the adsorbate and the adsorbent. The solute heat of adsorption at zero surface coverage ( $\Delta H^0$ ) is related to the retention data via

$$\ln V_g^0 = \frac{\Delta H^0}{RT} + C \quad [1]$$

where  $V_g^0$  is the solute-specific retention volume corrected to 0°C,  $R$  is the gas constant, and  $T$  is the column temperature. Plots of  $\ln V_g^0$  versus  $1/T$  yield  $\Delta H^0$  from the slope. At finite solute surface concentration (nonlinear isotherms), an accurate calculation of retention volumes from peak maxima is difficult. For such a case, therefore, initially the adsorption isotherms are estimated at various temperatures. From these, the adsorption isosters can be found, whose slope gives the isosteric heat of adsorption ( $\Delta H_{st}^0$ ):

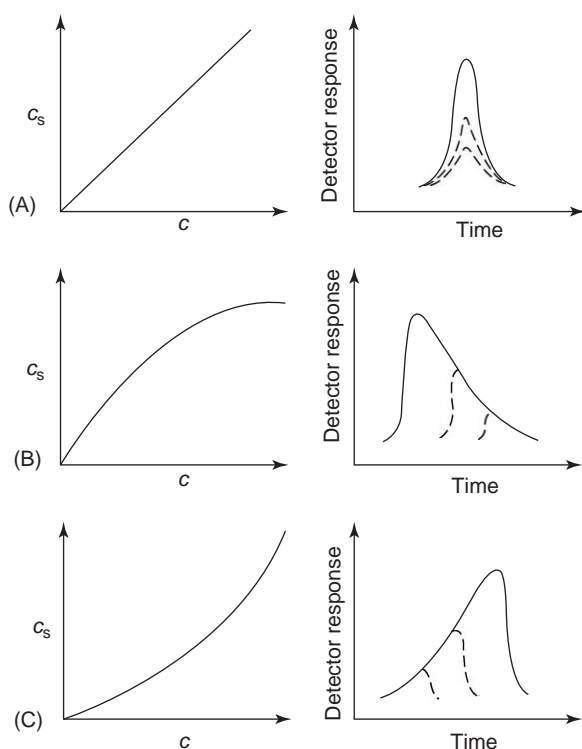
$$\ln p = \frac{\Delta H_{st}^0}{RT} + C \quad [2]$$

where  $p$  is the solute partial pressure.

The heat determined in this way depends on the coverage ( $a$ ); thus, the coverage dependence of  $\Delta H_{st}^0$  can be examined by calculating it for different values of  $a$ . Such dependence gives valuable information about features of surface interactions, the magnitude of the direct interaction of adsorbates, and the surface heterogeneity. Other thermodynamic functions of adsorption may be calculated from the temperature and coverage dependence of  $\Delta H_{st}^0(a, T)$ . The equation relating the standard free energy of adsorption  $\Delta G^0(a, T)$  to  $\Delta H^0(a, T)$  is

$$\Delta G(a, T) = \Delta G^0(a, T_0) \frac{T}{T_0} - T \int_{T_0}^T \frac{\Delta H^0(a, T)}{T^2} dT \quad [3]$$

where  $T_0$  is the temperature for which the standard free energy is defined.



**Figure 2** Shapes of adsorption isotherms and their corresponding elution peaks: (A) linear isotherm, symmetrical peak with position of peak maximum constant; (B) convex isotherm, distorted peak, tailing line spread, position of peak maximum dependent on sample size; (C) concave isotherm, distorted peak, leading profile spread, position of peak maximum dependent on sample size. (Reproduced with permission from Paryjczak T (1986) *Gas Chromatography in Adsorption and Catalysis*, p. 31. Warsaw: PWN-Polish Scientific Publishers. Chichester: Ellis Horwood.)

Having determined  $\Delta H^0(a, T)$  and  $\Delta G^0(a, T)$ , the entropy of adsorption can be calculated directly:

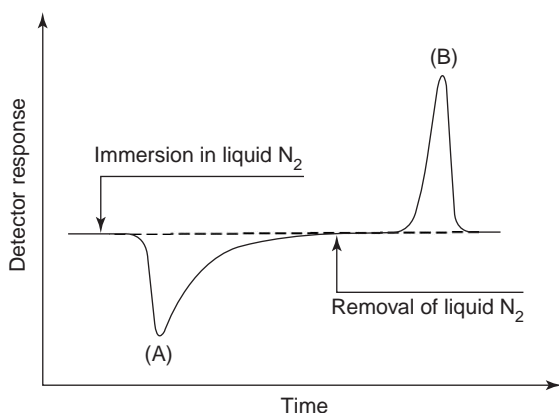
$$\Delta S^0(a, T) = \frac{\Delta H^0(a, T) - \Delta G^0(a, T)}{T} \quad [4]$$

### Specific Surface Area of Solids

Of all existing GC techniques for determining the specific area of a solid, the heat desorption method is the one most often used. This method was developed by Nelson and Eggertsen and modified by a number of workers. In principle, the heat desorption method is based on the traditional Brunauer, Emmett, Teller (BET) technique in which the quantity of adsorbed gas (usually nitrogen) at a temperature near its boiling point is determined. By determining the adsorption at various pressures, it is possible, using the BET equation, to calculate the amount of adsorbate required for the formation of a monolayer.

In the classic example of this method, the principle of measurement is based on the adsorption of nitrogen from a stream of nitrogen–helium or nitrogen–hydrogen mixture of a given composition at the temperature of liquid nitrogen. The sample comes to equilibrium at this temperature, indicated by a settling of the detector response. Removal of the liquid nitrogen bath results in desorption of the previously adsorbed nitrogen (Figure 3).

The quantity of adsorbed nitrogen at an appropriate relative pressure is calculated from the desorption peak since this is more symmetrical. Each experiment performed for a given concentration of adsorbate in the mixture gives one point of the adsorption isotherm.



**Figure 3** Schematic detector response in the heat desorption method: (A) adsorption peak; (B) desorption peak. (Reproduced with permission from Paryjczak T (1986) *Gas Chromatography in Adsorption and Catalysis*, p. 130. Warsaw: PWN-Polish Scientific Publishers. Chichester: Ellis Horwood.)

Easy and rapid estimation of the specific surface area of a solid is achieved using a comparative method, in which one uses the proportional relationship between the peak area and the specific surface area of the adsorbent. In this case the specific surface area of a similar standard species must be known.

The heat desorption method is widely used because of its simplicity in measurement and its great sensitivity, which means that specific surface areas can be determined over a broad range, from  $0\text{--}07\text{ m}^2\text{ g}^{-1}$  up to  $1000\text{ m}^2\text{ g}^{-1}$ .

Serpinet *et al.* have developed a GC method independent of the BET adsorption model for determining the specific areas of solids with hydroxyl groups on their surface (e.g., silica, alumina, diatomite).

### Solution Thermodynamics

According to chromatographic theory, the solute net (adjusted) retention volume  $V'_R$  is related to the solute distribution constant (partition coefficient)  $K$  by the equation

$$V'_R = KV_L \quad [5]$$

where  $V_L$  is the stationary phase volume. This relationship is strictly true when contributions from interfacial adsorption are negligible. Adsorption at the gas–liquid interface for solutes of low solubility in the stationary phase and at the liquid–solid interface for compounds with a strong affinity for the support are significant sources of error for the determination of partition coefficients under some experimental conditions. Equation [5] is appropriate when the experimental partition coefficient is independent of changes in the stationary phase volume, providing a simple test for the reliability of eqn [5]. In addition, by the purposeful variation of the stationary phase volume and extrapolation of experimental  $V'_R$  values to an infinite stationary phase volume, the partition coefficient can be estimated for conditions where interfacial adsorption contributes to the retention mechanism:

$$K = \frac{\gamma_S}{\gamma_M} \frac{C_S}{C_M} = \frac{a_S}{a_M} \quad [6]$$

where  $a$ ,  $\gamma$ , and  $C$  are the solute activity, activity coefficients, and concentrations in the stationary (S) and mobile (M) phases, respectively.

In ideal conditions, the solute in the gas phase at temperature  $T$  and pressure  $101.325\text{ kPa}$  ideally transfers to solution in the liquid phase at temperature  $T$ , unit molar concentration, and molecular interactions characteristic of an infinitely dilute solution; thus, one

can obtain a set of equations directly relating the thermodynamic functions of distribution equilibrium to the specific retention volume, corrected to 0°C ( $V_g^0$ ):

$$K = \frac{MV_g^0}{273R} \quad [7]$$

$$\Delta G_T^0 = RT \ln \frac{MV_g^0}{273R} \quad [8]$$

$$\Delta H_T^0 = RT^2 \frac{d \ln V_g}{dT} \quad [9]$$

$$\Delta S_T^0 = \frac{\Delta H_T^0 - \Delta G_T^0}{T} \quad [10]$$

where  $\Delta G_T^0$ ,  $\Delta H_T^0$ , and  $\Delta S_T^0$  are the standard partial molar Gibbs free energy, enthalpy (heat), and entropy, respectively;  $M$  is the stationary phase molecular mass.

Another pivotal thermodynamic parameter that accounts for deviations from ideal Raoult's law is the activity coefficient at infinite dilution  $\gamma^\infty$ :

$$\gamma^\infty = \frac{273R}{V_g^0 p^0 M} \quad [11]$$

where  $p^0$  is the vapor pressure of the pure solute.

The difference between the thermodynamic properties of solution, based on  $K$ , and the analogous properties of concentration from the 101–325 kPa ideal gas state may be evaluated directly from the solute activity coefficient and can be identified with the excess thermodynamic properties of solution:

$$\Delta G_E^0 = RT \ln \gamma^\infty \quad [12]$$

$$\Delta H_E^0 = -RT^2 \frac{d \ln \gamma^\infty}{dT} \quad [13]$$

$$\Delta S_E^0 = \frac{\Delta H_E^0 - \Delta G_E^0}{T} \quad [14]$$

The assumption of ideal gas-phase behavior is an approximation. For greater accuracy, fugacity and virial corrections should be applied. The corrections are small when helium, hydrogen, or nitrogen are used as mobile phases at low column pressures but may become appreciable at higher pressures and when other carrier gases are used.

In the studies interfacial adsorption should be taken into consideration as in most cases this results in the largest measurements error in partition and activity coefficients, particularly for solutes of different polarity to the stationary phase.

## Quantitative Structure–Retention Relationships

Quantitative structure–retention relationships (QSRR) are helpful in elucidation retention mechanisms, for predicting retention indices and estimating some physicochemical properties. Gas chromatographic retention is a phenomenon that is mainly dependent on molecule–stationary phase interactions. Thus, each molecule, at least in theory, will exhibit unique retention characteristics based on its chemical, structural, and electronic properties.

Numerous structural descriptors – numerical representations of molecular structure – can be used to develop statistical correlations between a compound's structure and its corresponding retention index ( $I$ ) (Kovats retention indices are the most frequently used parameters in this connection). Predictive models can be generated through multiple linear regression analysis of calculated descriptors, which are divided into four general categories: topological, geometric, physical, and electronic.

A general advantage of QSRR is that the equations derived allow the prediction of retention indices for compounds structurally similar to those used to develop the model but not represented in the initial data set. An essential condition for derivation of significant quantitative relationships between retention data and molecular structure descriptors is the precision and reliability of the chromatographic data used. Although there are several potential sources of errors in routinely determined GC retention indices, the correlations obtained are often very promising and meaningful.

Precise control of the chromatographic conditions by means of computers in modern instruments and further development of computing methods of calculation and subsequent analysis of descriptors should produce significant improvements of QSRR derived from GC, which will be especially useful for characterization of intermolecular interactions.

Abraham *et al.* presented the possibility of using GC data to correlate biological phenomena involving gaseous solutes. GC data on a given stationary phase might be suitable for the modeling of a biological process that involves the transfer of a solute from the gas phase to a biological phase.

## Catalysis

GC has been used widely to study properties of catalysts, such as diffusivity, dispersion, and selective surface area of supported catalytic systems, surface acidity, thermodynamics of adsorption, and various aspects of catalytic reactions.



In general, GC can be applied in two ways to study catalytic reactions. In the first, the catalyst is placed inside the chromatographic column. Substrates are usually injected pulsewise into the stream of carrier gas. Packed with catalyst, the column acts as both reactor and separating column. A chemical reaction taking place under the so-called chromatographic conditions differs from one occurring normally in that the intermediate and final products are generally separated from the substrates, thus preventing any reaction between them. Obviously, the kinetics of catalytic reactions taking place on a chromatographic column differs fundamentally from the kinetics of reactions occurring in flow or flow-circulation apparatus or in a microreactor. This method is particularly suitable for studying the mechanism of contact processes.

In the second method, a microreactor, in which the catalytic reaction proceeds, is placed immediately before the chromatographic column. The substrates, injected pulsewise into the carrier gas, pass into the microreactor and from there into the chromatographic column for postreaction analysis. This is a very promising technique, used to analyze such physicochemical properties of catalysts as their activity, selectivity, and catalytic reaction mechanisms. The microreactor technique provides advantages of simplicity and rapidity of measurement; high sensitivity; the capacity to operate with very small quantities of reactants and catalysts; isothermicity, since local overheating is avoided; and analysis of the initial periods of operation of a catalyst and its poisoning. Finally, the coefficients of adsorption of the reactants can be determined, and thus also the surface concentration of the reactants, which enables the real reaction rate constant to be established. However, pulse techniques do not guarantee constant conditions of the process, and data obtained using them do not define the rate of reaction, which is measured in flow and flow-circulation methods. This results from the fact that the reactant concentrations vary, as does the state of the catalytically active surface, its chemical composition, degree of oxidation or reduction, etc.

When using the pulse technique to study catalytic processes in the microreactor, it is important to select correctly the input volume of the pulse ( $V_p$ ) and the volume of the catalyst ( $V_{cat}$ ). If the ratio  $V_p/V_{cat} \gg 1$ , the results obtained with this method basically correspond to those from the traditional flow method; thus, the pulse microreactor can be regarded as a flow reactor in which the movement of the pulse takes place under ideal piston displacement conditions.

If the volumetric ratio of the pulse and catalyst is inverted, i.e., when  $V_p/V_{cat} \ll 1$ , in a number of cases the catalytic reaction will take place under

chromatographic conditions. The considerable concentration gradient of the reactants interferes with the working up of the results, but nevertheless these are very valuable in analyzing reaction mechanisms.

The surface of the catalyst during a reaction proceeding under pulse conditions is constantly renewed by the carrier gas flowing over it at the reaction temperature. Since very small amounts of the reactant are introduced into a microreactor working under chromatographic conditions, one may assume that the results refer to the properties of that section of the catalyst's surface that was not affected by the reaction mixture. On the other hand, if pulses are introduced rapidly one after another, the change in the catalyst's activity during the chemical reaction can be assessed. The catalytic activity of the first pulses is often greater than that estimated under stationary conditions by a few orders of magnitude.

The processes involved in the flow of a pulse of reactant through a layer of catalyst, such as diffusion, adsorption, and desorption, exert a significant influence on the reaction kinetics. The study of catalytic reactions under chromatographic conditions is especially suitable when the equilibrium constant of the reaction is small but the rate of reaction sufficiently fast. It is the separation of products and not the rate of reaction that then limits the extent of the process.

For irreversible reactions of first order, the rate constant  $k$  may easily be obtained from Bassett and Habgood's equation, often used in the literature:

$$kK = \frac{F_c^0}{273Rm} \ln \frac{1}{1-\alpha} \quad [15]$$

where  $k$  and  $K$  are the rate constant and Henry's constant, respectively,  $F_c^0$  is the volumetric flow rate under normal conditions,  $m$  is the mass of the catalyst,  $R$  is the gas constant, and  $\alpha$  is the degree of conversion.

Practical applications of GC methods for assessing catalytic activity and studying heterogeneous reaction are outrunning the development of the theory of nonstationary processes under pulse conditions. Although numerous models describing different types of both reversible and irreversible reactions have been elaborated, many theoretical problems are still far from solution.

## Surface Acidity

The temperature programmed desorption method originating from GC is most commonly used in determining the surface acidity of catalysts.

The catalytic activity of a large group of catalysts used in isomerization, polymerization of alkenes, cracking, hydrocracking, dehydration of alcohols, manufacturing of alkyl ethers, reforming on bifunctional catalysts, and many other processes is due to the acidity of their surfaces. The knowledge of the number and strength of acid sites and also type of acidity (Brönsted and Lewis) involved is very important as it helps to explain the mechanism of reactions occurring at the surface of these catalysts.

Mostly ammonia but also pyridine, aniline, *N*-methyl tetrahydropyrrole, diethylamine, isopropylamine, and cyclohexylamine are used as probe particles, whose desorption is the source of information on acid centers.

The basis of temperature programmed desorption (TPD) method is chemisorption of base vapor, ammonia being most often used, on the surface of an acidic catalyst at a given temperature, followed by its desorption as a result of a temperature rise. The area of the obtained desorption peak gives the total acid site density and the maximum of the peak characterizes the activation energy of desorption and may be considered as a measure of the acid strength.

Figure 4 shows TPD ammonia spectra where the h-peak is the desorption peak of ammonia that was adsorbed on the acid sites, whereas the l-peak was assigned to ammonia weakly held or physically adsorbed on the zeolite.

Niwa *et al.* derived an equation describing TPD of ammonia for zeolites including its readsorption. The equation allows for calculating acidity strength expressed as a change of desorption enthalpy  $\Delta H$ :

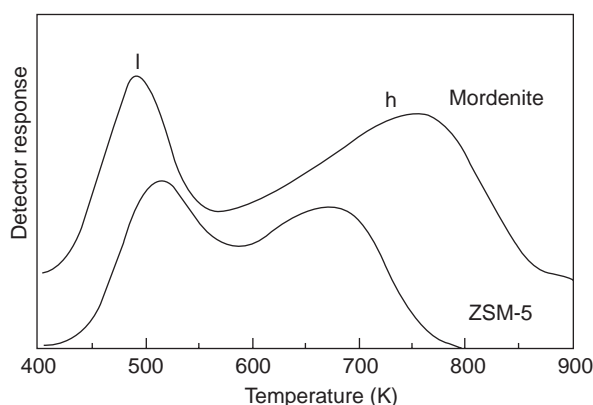
$$\ln T_m - \ln \frac{A_0}{F} = \frac{\Delta H}{RT_m} + \ln \frac{\beta(1 - \Theta_m)^2(\Delta H - RT_m)}{P^0 \exp(\Delta S/R)} \quad [16]$$

where  $T_m$  is the maximum peak temperature,  $A_0$  the number of acid sites,  $W$  the mass sample,  $F$  the flow rate of carrier gas,  $\Delta S$  the change of desorption entropy, and  $P^0$ ,  $R$ ,  $\beta$ ,  $\Theta$ , the pressure (standard conditions), gas constant, heat rate, degree of surface coverage at maximum peak temperature, respectively. This equation shows the dependence of  $T_m$  on the parameters of TPD.

## Other Applications

### Complexation Equilibrium

Metal ions can act as electron-pair acceptors, reacting with electron donors to form coordination compounds or complexes with either the stationary phase



**Figure 4** TPD spectra on H-mordenite ( $\text{SiO}_2/\text{Al}_2\text{O}_3 = 15.0$ ) and H-ZSM-5 ( $\text{SiO}_2/\text{Al}_2\text{O}_3 = 23.8$ ), l- and h-peaks due to low and high temperatures of desorption. (Reproduced with permission from Niwa M and Katada N (1997) Measurement of acidic property of zeolites by temperature programmed desorption of ammonia. *Catalysis Surveys from Japan* 1: 215–226.)

or the mobile phase. Such complexation imparts a high degree of selectivity to the GC system, and the chromatographic data can be used to calculate the stability constants and related parameters of the complexes.

### Polymer Studies

Physical properties of polymers have frequently been studied using the polymers as stationary phases. Interactions between solvents and polymers such as the adsorption propensity of the polymeric stationary phases or solubility in polymer films can readily be determined by GC.

Polymers can also be identified and studied by pyrolysis followed by GC analysis of the volatile breakdown products.

### Virial Coefficients

Development of the theory of virial corrections to partition coefficient data has led to the use of GC for the direct measurement of virial coefficients. In general, the method of determination of virial coefficients by GC consists of measuring the retention volumes at various carrier gas pressures and extrapolating to zero pressures.

Gas–solid virial coefficients, characterizing the interaction of the adsorbate molecule with the surface of the solid, can be used to estimate surface area and for quantitative studies of the energetic heterogeneity of a surface.

### Liquid Crystals

GC on liquid-crystalline stationary phases can be used for investigation of the physicochemical

properties of the liquid crystals, particularly the phenomena occurring in the transition zones. Using GC methods, phase diagrams of liquid crystals, the degree of ordering of the liquid-crystal molecules, and the relationship between the liquid-crystal molecular orientation and the polarity of the substrate surface can also be determined.

See also: **Gas Chromatography**: Overview; Principles; Gas–Solid Chromatography; Instrumentation.

## Further Reading

Abraham MH, Gola JMR, Kumarsingh R, Cometto-Muniz JE, and Cain WS (2000) Connection between chromatographic data and biological data. *Journal of Chromatography B* 745: 103–115.

- Conder IR and Young CL (1979) *Physicochemical Measurement by Gas Chromatography*. New York: Wiley.
- Grob RL (ed.) (1985) *Modern Practice of Gas Chromatography*. New York: Wiley.
- Kaliszan R (1987) *Quantitative Structure-Chromatographic Retention Relationships*. New York: Wiley.
- Katsanos NA and Roubani-Kalantzopoulou F (2000) *Advances in Chromatography* 40: 231.
- Laub RJ and Pecsok PL (1978) *Physicochemical Applications of Gas Chromatography*. New York: Wiley.
- Locke DC (1976) Physicochemical measurements using chromatography. *Advances in Chromatography* 14: 87–198.
- Poole CF (2003) *The Essence of Chromatography*. Amsterdam: Elsevier.
- Szynkowska MI, Lesniewska E, and Paryjczak T (2003) Temperature-programmed techniques in catalysis. *Polish Journal of Chemistry* 77: 657–673.

## Environmental Applications

**M T Galceran and F J Santos**, University of Barcelona, Barcelona, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Environmental Analysis

Gas chromatography (GC) is one of the most frequently used chromatographic separation techniques for the determination of the presence of organic contaminants in the environment. The popularity of GC is due to a favorable combination of very high sensitivity and resolution, wide dynamic concentration range, and high selectivity. The constant advance in the development of new and robust stationary phases and detectors has made GC a highly reliable tool for routine trace analysis. Moreover, the introduction of fused-capillary columns and the coupling of GC with mass spectrometry (MS) have enabled the most advanced analytical technique currently available for volatile solutes. GC is also a valuable reference technique for environmental analysis. The vitality of GC is also reflected in the development of new technologies such as high-speed GC (HSGC) and comprehensive two-dimensional GC (GC × GC). Such technologies greatly increase the separation capabilities of GC.

A wide range of organic environmental pollutants are highly suitable for GC analysis as they are relatively nonpolar compounds with relatively low molecular mass (<500) and are stable at the normal range of temperatures used for GC (50–350°C). Due

to the low concentration of organic pollutants present in the environmental samples, ranging from micrograms per gram to femtograms per gram, an extensive cleanup and fractionation of the analytes is needed before GC analysis. The complexity of the cleanup steps depends on the sample matrix and the concentration of the analytes. Thus, for relatively clean water samples, a simple procedure is often used, whereas for sediments or biota, complex cleanup procedures are needed. Quality assurance of the complete method including GC determination is required to obtain reliable results in environmental analysis.

## Columns

The introduction of capillary columns for GC analysis produced a breakthrough in the analysis of environmental pollutants due to their high separation efficiency. For environmental applications, fused-silica wall-coated open-tubular columns with internal diameters from 0.1 to 0.32 mm and film thickness of 0.1–0.2 µm, and lengths from 25 to 60 m are currently used. The wide range of stationary phases commercially available with different polarities and high thermal stability provides the tool required to maintain the prominent position of GC in environmental analysis. In addition, the availability of chiral stationary phases gives GC the capability to perform GC enantiomer separations. **Table 1** gives the recommended columns used in routine analysis of some selected pollutants.

properties of the liquid crystals, particularly the phenomena occurring in the transition zones. Using GC methods, phase diagrams of liquid crystals, the degree of ordering of the liquid-crystal molecules, and the relationship between the liquid-crystal molecular orientation and the polarity of the substrate surface can also be determined.

See also: **Gas Chromatography**: Overview; Principles; Gas–Solid Chromatography; Instrumentation.

## Further Reading

Abraham MH, Gola JMR, Kumarsingh R, Cometto-Muniz JE, and Cain WS (2000) Connection between chromatographic data and biological data. *Journal of Chromatography B* 745: 103–115.

Conder IR and Young CL (1979) *Physicochemical Measurement by Gas Chromatography*. New York: Wiley.

Grob RL (ed.) (1985) *Modern Practice of Gas Chromatography*. New York: Wiley.

Kaliszan R (1987) *Quantitative Structure-Chromatographic Retention Relationships*. New York: Wiley.

Katsanos NA and Roubani-Kalantzopoulou F (2000) *Advances in Chromatography* 40: 231.

Laub RJ and Pecsok PL (1978) *Physicochemical Applications of Gas Chromatography*. New York: Wiley.

Locke DC (1976) Physicochemical measurements using chromatography. *Advances in Chromatography* 14: 87–198.

Poole CF (2003) *The Essence of Chromatography*. Amsterdam: Elsevier.

Szynkowska MI, Lesniewska E, and Paryjczak T (2003) Temperature-programmed techniques in catalysis. *Polish Journal of Chemistry* 77: 657–673.

## Environmental Applications

**M T Galceran and F J Santos**, University of Barcelona, Barcelona, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Environmental Analysis

Gas chromatography (GC) is one of the most frequently used chromatographic separation techniques for the determination of the presence of organic contaminants in the environment. The popularity of GC is due to a favorable combination of very high sensitivity and resolution, wide dynamic concentration range, and high selectivity. The constant advance in the development of new and robust stationary phases and detectors has made GC a highly reliable tool for routine trace analysis. Moreover, the introduction of fused-capillary columns and the coupling of GC with mass spectrometry (MS) have enabled the most advanced analytical technique currently available for volatile solutes. GC is also a valuable reference technique for environmental analysis. The vitality of GC is also reflected in the development of new technologies such as high-speed GC (HSGC) and comprehensive two-dimensional GC (GC × GC). Such technologies greatly increase the separation capabilities of GC.

A wide range of organic environmental pollutants are highly suitable for GC analysis as they are relatively nonpolar compounds with relatively low molecular mass (<500) and are stable at the normal range of temperatures used for GC (50–350°C). Due

to the low concentration of organic pollutants present in the environmental samples, ranging from micrograms per gram to femtograms per gram, an extensive cleanup and fractionation of the analytes is needed before GC analysis. The complexity of the cleanup steps depends on the sample matrix and the concentration of the analytes. Thus, for relatively clean water samples, a simple procedure is often used, whereas for sediments or biota, complex cleanup procedures are needed. Quality assurance of the complete method including GC determination is required to obtain reliable results in environmental analysis.

## Columns

The introduction of capillary columns for GC analysis produced a breakthrough in the analysis of environmental pollutants due to their high separation efficiency. For environmental applications, fused-silica wall-coated open-tubular columns with internal diameters from 0.1 to 0.32 mm and film thickness of 0.1–0.2 µm, and lengths from 25 to 60 m are currently used. The wide range of stationary phases commercially available with different polarities and high thermal stability provides the tool required to maintain the prominent position of GC in environmental analysis. In addition, the availability of chiral stationary phases gives GC the capability to perform GC enantiomer separations. **Table 1** gives the recommended columns used in routine analysis of some selected pollutants.



**Table 1** Recommended capillary columns for the determination of selected pollutants

<i>Stationary phase</i>	<i>Compound</i>
Methylpolysiloxane	Polycyclic aromatic hydrocarbons, disinfection by-products, organochlorine pesticides, chlorinated paraffins
5% Phenyl methylpolysiloxane	Polycyclic aromatic hydrocarbons, disinfection by-products, polychlorinated biphenyls, polychlorinated dibenzo- <i>p</i> -dioxins and dibenzofurans, organochlorine pesticides, polybrominated biphenyls, polybrominated diphenyl ethers, chlorinated paraffins, methyl- <i>tert</i> -butyl-ether
14% Cyanopropylphenyl methylpolysiloxane	Disinfection by-products (derivatized aldehydes), polychlorinated biphenyls, polychlorinated dibenzo- <i>p</i> -dioxins and dibenzofurans, pesticides, polybrominated diphenyl ethers
50% Phenyl methylpolysiloxane	Polychlorinated biphenyls, pesticides
Specific columns	Polychlorinated biphenyls (DB-XLB, HT-8); polychlorinated dibenzo- <i>p</i> -dioxins and dibenzofurans (DB-Dioxin)

Recently, HSGC has been applied to the analysis of some environmental pollutants such as pesticides, polychlorinated biphenyls (PCBs), and polycyclic aromatic hydrocarbons (PAHs). Using this technique, fast separations have been obtained using short capillary columns (<10 m) with internal diameters of 50  $\mu$ m. An important reduction in analysis time can be achieved so that separations by conventional capillary GC that are obtained in 20 min take no longer than 20 s using this technique.

An important improvement in the separation capacity of GC can be obtained using comprehensive GC  $\times$  GC. This involves the use of two capillary columns with different separation mechanisms by which the second column separates unresolved compounds that elute from the first. The instrument design uses a modulator interface that couples the two columns. The first column is generally a conventional nonpolar GC column, the second one being a short polar column that allows very fast separation. This technique has proved to be a promising tool for the unambiguous separation of several contaminants such as PAHs, PCBs, polychlorinated dibenzo-*p*-dioxins and dibenzofurans (PCDD/Fs), and pesticides in environmental samples.

### Sample Introduction

The low concentration levels of the compounds in environmental samples impose specific requirements in terms of sample injection for GC analysis. In addition to the common injection techniques of capillary GC (split, splitless, on-column, and programmed temperature vaporized (PTV) injection), some other sample introduction methods coupled to GC such as solid-phase microextraction (SPME), headspace, etc., have favored the versatility of GC and reduced the time required for sample preparation. These techniques have an advantage over the conventional injection methods, which is that a preconcentration step prior to GC

analysis can be performed; therefore, an enhancement on the sensitivity of the GC method is obtainable.

Splitless and on-column are the most useful injection techniques, although in recent years the use of the PTV injector has increased as a result of its capacity to reduce discrimination effects in the injector of analytes with high boiling points. Moreover, PTV and on-column injectors allow direct GC analysis of water samples using large-volume injection. PTV and on-column injectors have been successfully applied to pesticide residue analysis in water. In ambient air analysis, the sampling and preconcentration procedures most widely used are sorbent sampling and the cryogenic method. Different solid sorbents have been used for the analysis of volatile organic compounds (VOCs) and semi-VOCs, the desorption of which is performed thermally or by means of an organic solvent. Other techniques often used for the analysis of VOCs, such as headspace, purge and trap, and closed-loop stripping analysis can be easily combined off- or online with the GC system. Solid-phase extraction (SPE) methods coupled to GC have also been used for the analysis of environmental contaminants, mainly for pesticides. SPE technique allows an important preconcentration of the analytes and can be coupled directly to the GC injector if it is combined with large-volume injection techniques. SPME is an inexpensive and easily automated technique based on sorptive extraction that has shown itself to be a good alternative to liquid extraction. This technique, combined with GC, has been applied to the analysis of a great number of organic contaminants, e.g., BTEX (benzene, toluene, ethylbenzene, and xylenes), polar solvents, PAHs, pesticides, PCBs, and disinfection by-products (DBPs) in different environmental samples such as water and soils. Another solventless sample preparation technique recently developed is stir bar sorptive extraction which, followed by thermal desorption, has been proposed for the analysis of pollutants in water.



## Detection

One important feature of GC is the availability of selective and sensitive detectors. The flame ionization detector (FID) is the most popular system for environmental analysis because it responds to nearly all organic compounds. However, FID has a detection limit in the low nanogram range, reducing its applicability to the determination of major contaminants. Selective detectors such as the electron-capture detector (ECD), thermionic ionization detector (TID), flame-photometric detector (FPD), and atomic emission detector (AED) are currently used for the determination of a specific group of contaminants in environmental samples. ECD has a high sensitivity (picogram to femtogram range) and selectivity for detecting halogen-containing compounds. This detector has been widely applied for the analysis of PCBs, halogenated pesticides, PCNs, and brominated organic compounds. In comparison with the FID system, nitrogen and phosphorus detectors have the ability to enhance the ionization of N and P compounds, allowing low detection limits and good selectivity over (interfering) carbon compounds. FPD and the more recently developed pulsed flame photometric detector have an important field of applications in the detection of phosphorus and sulfur organic compounds. The element-selective detector for GC, the AED, has a high selectivity and has led to numerous applications, mainly for pesticide analysis.

GC coupled to mass spectrometry (GC-MS) has proved to be an advantageous and powerful technique for the analysis of a wide range of environmental contaminants, mainly because of the high sensitivity, selectivity, and versatility. In fact, selective detectors have progressively been replaced by GC-MS. The very high number of applications is the result of the efficiency of separation by GC and the good qualitative information and sensitivity provided by mass spectrometry. The MS fragmentation pattern can often provide unambiguous component identification by comparison with library spectra in electron ionization (EI). In addition, the use of chemical ionization (CI) provides better selectivity and sensitivity than EI and, in some cases, this ionization mode allows to achieve limits of detection similar to those obtained with selective detectors. Several analyzers are currently used in GC-MS. The most popular is the single quadrupole analyzer, but the ion-trap analyzer has achieved a prominent position in environmental applications. The reason for this lies in the enhancement selectivity obtainable in ion-trap tandem MS (MS/MS) mode, which carries with it the additional advantage of confirmation capabilities using the full-scan spectra of product ions.

## Applications

### Polycyclic Aromatic Hydrocarbons

PAHs represent an important class of environmental pollutants that have gained special attention because some of them are strong mutagens and carcinogens. PAHs are natural constituents of crude oil and many other petrochemical products, but they can also be formed by incomplete combustion of fossil fuels and other organic matter. Natural sources of PAH emission such as forest fires and volcanic eruptions have always existed but there has been an increase in the loading of the environment by anthropogenic sources that include incomplete combustion or the spillage of fossil fuel. On the basis of frequency of occurrence in the environment and their proved mutagenicity and carcinogenicity, 16 PAHs were selected in 1997 by the US Environmental Protection Agency (USEPA) as priority pollutants. Identification of these compounds in environmental samples is important in order to determine their specific emission sources, a step that would help the authorities to establish strategies to reduce PAH emissions into the environment. Pyrolytically produced PAHs from incomplete combustion of organic fuels such as oil, wood, or petrol can be distinguished from those of direct fossil fuel origin by the strong dominance of the unsubstituted parent PAH.

Capillary GC, first used to separate PAHs in 1964, is now one of the standard methods for the determination of these compounds in environmental matrices. Speed of analysis, high power of resolution, and low detection limits are the advantages of GC procedures. Since the PAHs most likely to be found in the environment are those that are volatile and semivolatile (up to 24 carbon atoms), they are analyzable by GC. In general, nonpolar stationary phases such as methylpolysiloxane or 5% phenyl methylpolysiloxane are the most adequate for the separation of these compounds in routine applications. Cold on-column injection is preferred as it improves both the resolution obtained for the first-eluting compounds and discrimination against high molecular weight compounds, which is difficult to avoid completely using splitless injection. The FID is normally adequate for sensitive detection, but coupling GC with MS affords greater selectivity through the application of selected ion monitoring. EI has been successfully applied to the analysis of these compounds but PAH isomers are almost always indistinguishable. To differentiate between PAH isomers, positive CI with methane, a mixture of methane/argon, or dimethyl ether as reagent gas is used. Nowadays, GC continues to be a popular technique for the analysis of PAHs because of its proven capabilities. Moreover, the technique is well established

and GC instrumentation is a common feature in routine analytical laboratories.

### Disinfection By-Products

Disinfection of drinking water is one of the major public health triumphs of the twentieth century. Today, most of the pathogenic microorganisms in water are killed when oxidizing chemicals such as chlorine, ozone, chlorine dioxide, and chloramines are added to the water in municipal treatment plants. Nevertheless, chemical DBPs are produced when disinfectants react with natural organic matter present in water. Trihalomethanes (THMs) are the main DBPs formed when chlorine is used but other compounds such as haloacetic acids (HAAs), haloacetonitriles (HANs), and haloketones (HKs) are also found in drinking water. The presence of DBPs in drinking water is a health hazard and may also cause unpleasant taste and odor. For these reasons, the occurrence of DBPs in drinking water has been evaluated and regulatory legislation has been established.

Purge-and-trap, liquid-liquid extraction, SPE, and headspace extraction are the most common preconcentration techniques used as a first step for control of DBPs in drinking water. Moreover, in the last few years SPME has been proposed for the analysis of several DBPs, with very good results. Table 2 provides a summary of the USEPA methods proposed for the analysis of DBPs in drinking water. For quantification of the halogenated high-priority compounds (THMs, HAAs, HANs, and HKs), GC is the technique of choice. Several detection systems have been used but ECD is the most popular due to its low cost and the high selectivity provided by this detector for halogenated compounds. Nevertheless, single quadrupole GC-MS instruments in SIM mode and ion-trap analyzers in full scan or MS/MS are also frequently used. Some DBPs are not amenable to GC analysis and have to be derivatized. For instance, HAAs are mainly deprotonated, so they cannot be extracted from water and injected into a GC column. Methylation with diazomethane or acidic methanol

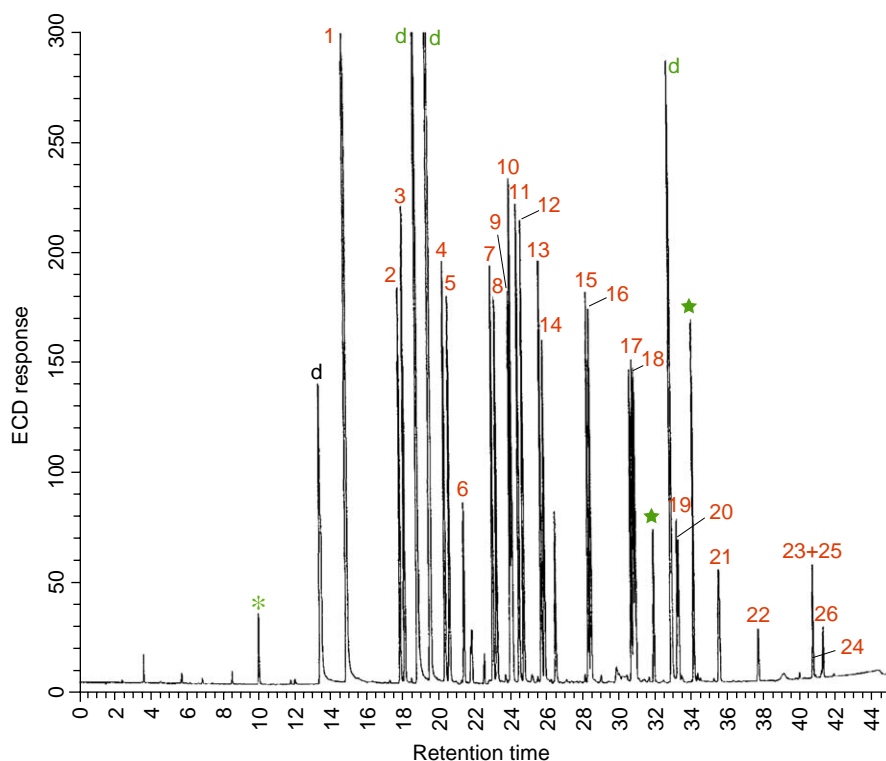
is the approach proposed for the analysis of these compounds (Table 2). Aldehydes, which are mainly formed when ozone is used as disinfection reagent, are polar compounds, making it difficult to separate DBPs from the water matrix. Therefore, derivatizing agents that convert the polar groups into less polar ones are used. For instance, these compounds are analyzed by GC after oximation with pentafluorobenzylhydroxylamine (PFBHA). In general, capillary columns with stationary phases such as the nonpolar methylpolysiloxane (DB-1), the slightly polar ones with 5% of phenyl groups such as DB-5 and the DB-624 specially designed for the analysis of volatile priority pollutants, are employed to control DBPs in water. These columns are used for THM, nitriles, ketones, and methylated HAAs. The more polar stationary phase, 14% cyanopropylphenyl methylpolysiloxane, such as DB-1701 and CPSil19 CB, is currently used for analysis of the PFBHA derivatized aldehydes. Figure 1 shows as an example of the chromatogram of a water sample spiked with aldehydes.

### Polychlorinated Biphenyls

PCBs have been used extensively in electronic equipment as dielectric in capacitors and transformers, hydraulic fluids, and as plasticizers. These compounds have been withdrawn from use or have remained in very restricted use since 1977. Nevertheless, the widespread use and persistence of PCBs in the environment are the cause of their ubiquitous presence around the world. Recently, these compounds have received special attention due to their properties as endocrine disruptors and environmental estrogens. PCBs are found as complex congener mixtures, consisting of a large number of major and minor components. From the total number of possible congeners (209 compounds), ~100 can be identified in environmental samples, but analysis of PCBs is at present almost exclusively carried out on a congener-specific basis. Seven selected individual congeners (PCBs 28, 52, 101, 118, 138, 153, and 180), which are considered as indicators of

**Table 2** EPA methods for the analysis of disinfection by-products (DBPs) in drinking water

Compounds	EPA method	Procedure
Trihalomethanes	502.2	Purge and trap, GC-photoionization-electrolytic conductivity detection
Trihalomethanes	524.2	Purge and trap, GC-MS
Trihalomethanes	551.1	Liquid-liquid extraction, GC-ECD
Haloacetic acids	552	Liquid-liquid extraction, diazomethane methylation, GC-ECD
Haloacetic acids	552.1	Anion exchange resin, acidic methanol methylation, GC-ECD
Haloacetic acids	552.2	Liquid-liquid extraction, acidic methanol methylation, GC-ECD
Haloacetic acids	552.3	Liquid-liquid microextraction, acidic methanol methylation, GC-ECD
Aldehydes, ketones	556	Oximation, liquid-liquid extraction, GC-ECD



**Figure 1** PFBHA-headspace-SPME-GC-ECD chromatogram of a water sample spiked with aldehydes at  $5 \mu\text{g l}^{-1}$ . Peak identification (*E* and *Z* isomers): (1) formaldehyde; (2, 3) acetaldehyde; (4, 5) propanal; (6) 2-methylpropanal; (7, 8) butanal; (9, 10) 2-methylbutanal; (11, 12) 3-methylbutanal; (13, 14) pentanal; (15, 16) hexanal; (17, 18) heptanal; (19, 20) octanal; (21) nonanal; (22) decanal; (23, 24) glyoxal; (25, 26) methyl glyoxal, (\*) 1,2-dibromopropane (IS); (★) 2,4,5-trifluoroacetophenone; (d) artifacts. Chromatographic conditions: GC column: DB-1701 fused-silica column (J&W Scientific) 30 m  $\times$  0.32 mm ID, 1  $\mu\text{m}$  film thickness. (Reprinted with permission from Cancho B, Ventura F, and Galceran MT (2001) Determination of aldehydes in drinking water using pentafluorobenzylhydroxylamine derivatization and solid-phase microextraction. *Journal of Chromatography A* 943: 1–13.)

PCB contamination, are currently analyzed by capillary GC. In addition to these compounds, other PCB congeners such as PCBs 105, 128, 149, 156, and 170 are also monitored due to their toxicity and their large contribution to total PCB contamination. Complete separation of all congeners is impossible on a single capillary column, and the use of at least two columns with different stationary phases is recommended. Thus, nonpolar stationary phases such as 5% phenyl methylpolysiloxane (for instance, DB-5 or CPSil 8CB) and medium polar columns such as 14% cyanopropylphenyl methylpolysiloxane (for instance, DB-1701) or 50% phenyl methylpolysiloxane (for instance, DB-17), or a polar stationary phase such as HP-FFAP (nitroterephthalic acid modified-polyethylene glycol) are currently used. Specially developed stationary phases for PCB analysis such as DB-XLB or HT-8 (carborane backbone phase) show a low number of congener co-elutions that in some cases can be resolved by MA. By using a good selection of columns, an unequivocal separation and identification of the individual congeners can be obtained.

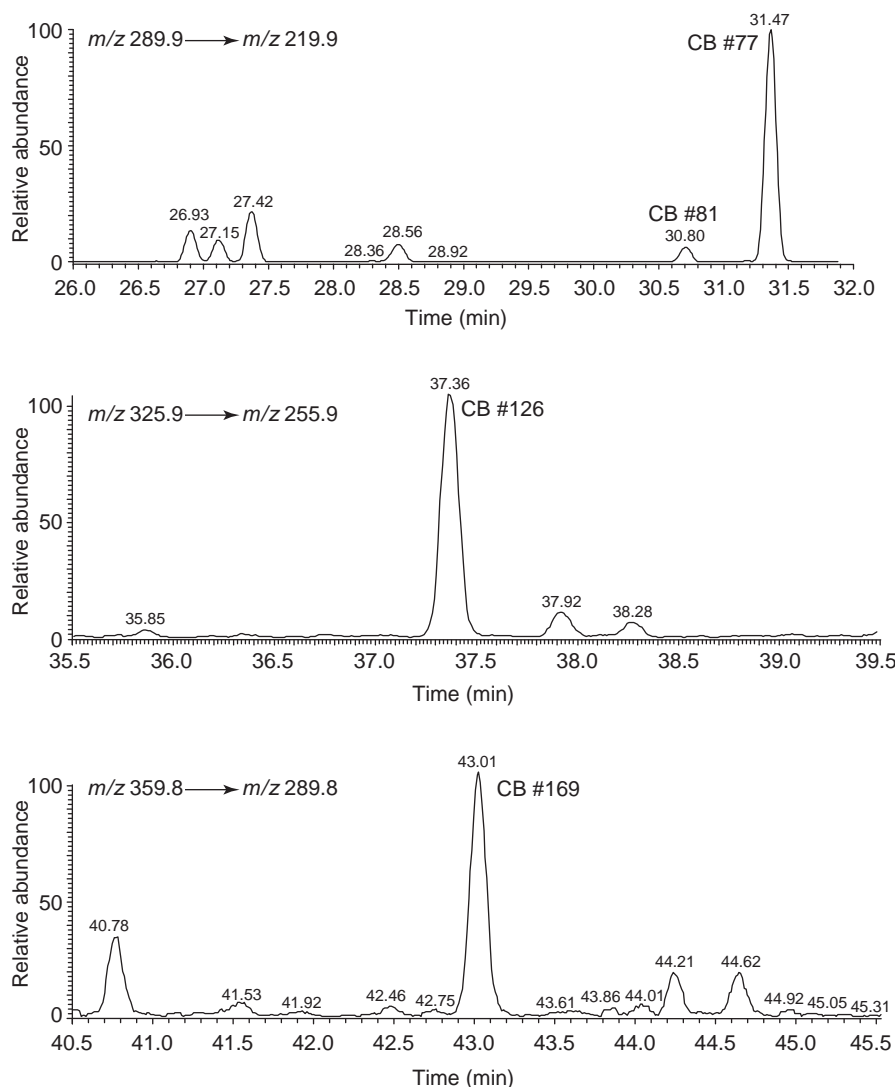
Over the course of the last few years, increasing attention has been focused on the study of non-*ortho* substituted (PCBs 77, 81, 126, and 169) and mono-*ortho* substituted PCBs (PCBs 105, 114, 118, 123, 156, 157, 167, and 189). These are also called dioxin-like PCBs. These substances show the same type of toxicity as polychlorinated dibenzo-*p*-dioxins and dibenzofurans. The main problem in the analysis of the non-*ortho* PCBs is their extremely low concentration compared with other PCBs. In most environmental samples the concentration difference between these compounds and the major PCBs ranges between 100 and 3000 times. Therefore, a separation of these compounds from other PCBs before GC analysis is required.

GC coupled with low-resolution MS (LRMS) provides the required sensitivity and selectivity for the determination of PCBs. Nevertheless, the use of high-resolution mass spectrometry (HRMS) at a resolution between 6000 and 10 000 in EI mode allows complete removal of the contribution of matrix interfering compounds, achieving low limits of detection (20–100 fg). Therefore, this is the technique of

choice when the dioxin-like PCBs are analyzed in environmental samples. The negative chemical ionization (NCI) mode can be used as an alternative to the EI mode to achieve lower limits of detection (6–100 times), but problems with the reproducibility and instrument-to-instrument variability makes its routine use difficult. Recently, however, GC ion-trap MS working in tandem mode (GC-ITMS/MS) has been successfully applied to the analysis of these compounds in environmental samples. This technique provides very high selectivity, although limits of detection are slightly higher than HRMS (60–300 fg). As an example, the GC-ITMS/MS chromatograms corresponding to the non-*ortho*-PCBs from a fish sample (chub) are given in Figure 2.

### Polychlorinated Dibenzo-*p*-Dioxins and Dibenzofurans

Polychlorinated dibenzo-*p*-dioxins (PCDDs) and dibenzofurans (PCDFs) are the most hazardous environmental pollutants that have received prolonged attention by the scientific community and by environmental regulation. These compounds are not commercially produced but can be formed during combustion processes, such as municipal waste incineration, in which halogenated compounds are exposed to very high temperatures. Other primary sources of formation of these compounds are related to various chemical processes that result in contaminated products and waste. Although there are theoretically 75 PCDFs and 135 PCDDs, only 17 of them

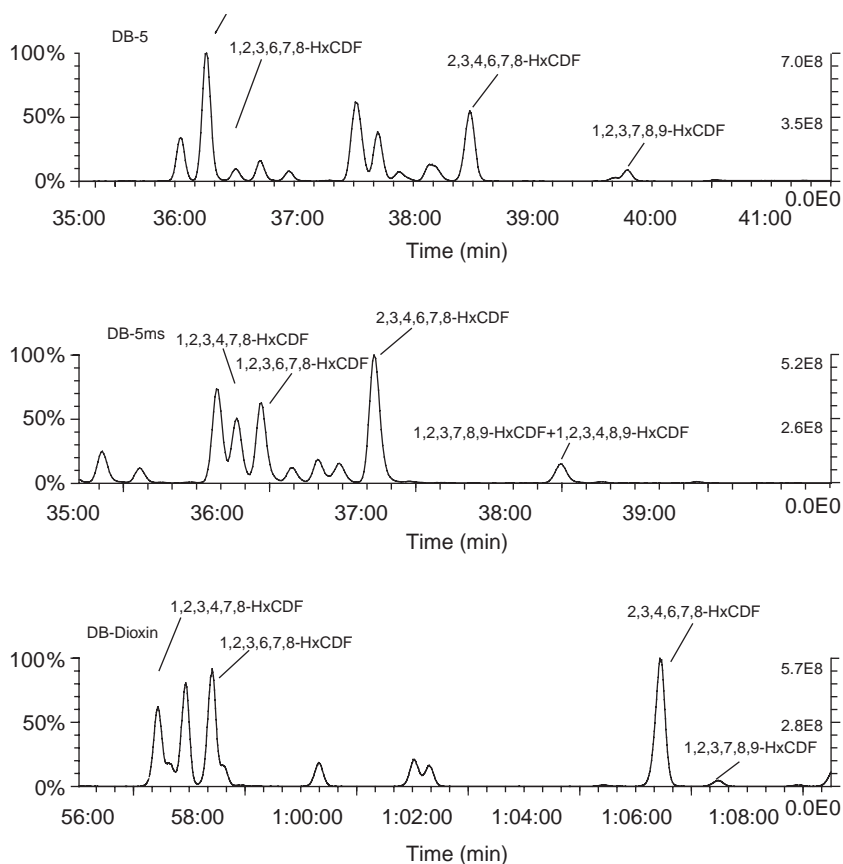


**Figure 2** GC-ion trap MS/MS chromatograms of the non-*ortho*-PCBs (PCBs 77, 81, 126, and 169) in a fish sample. Experimental conditions: GC column: DB-5 (J&W Scientific) 60 m  $\times$  0.25 mm ID, 0.25  $\mu$ m film thickness. MS analyzer: ion-trap MS, working in full-scan product mode. (Reproduced with permission from Malavia J, Santos FJ, and Galceran MT (2002) Ion trap MS/MS as alternative to HRMS for the analysis of non-*ortho*-PCBs in biota samples. *Organohalogen Compounds* 55: 103–106.)

(seven 2,3,7,8-substituted PCDDs and 10 PCDFs) are regularly analyzed and found in environmental samples. Of these, 12 are considered to be more toxic than the others, especially 2,3,7,8-tetrachloro-dibenzodioxin, which has the highest toxicity. Owing to the very lipophilic character of PCDD/Fs, they accumulate in biota and are considered a potential health risk for humans as well as for animals.

GC with fused-silica capillary columns with different stationary phases such as 5% phenyl methylpolysiloxane (DB-5, DB-5ms), 44% methyl-28% phenyl-20% cyanopropylpolysiloxane (DB-Dioxin) or cyanopropylpolysiloxane (SP-2330 or CPSil-88) are currently used, although problems related to the co-elution of some congeners have been described. Column lengths of ~50–60 m have to be used. Generally, a complete separation of the toxic PCDD/Fs from other congeners cannot be accomplished using only a single column. For instance, nonpolar stationary phases enable the separation between homolog groups

and between 2,3,7,8-chloro-substituted congeners, but some co-elutions between toxic and less toxic PCDD/F congeners are produced. An enhancement in the resolution can be obtained using more polar stationary phases such as CPSil-88 and SP2330, but the separation of 2,3,7,8-TCDD, 1,2,3,7,8-PeCDF, and 1,2,3,4,7,8-HxCDF remains incomplete. In the same way, the polar stationary phase DB-Dioxin, which was especially developed for the analysis dioxins and furans, allows the successful separation of all 2,3,7,8-chloro-substituted dioxins and furans, but 1,2,3,7,8-PeCDF, 2,3,4,7,8-PeCDF, and 1,2,3,6,7,8-HxCDF co-elute with other less toxic furans. **Figure 3** shows an example of the separation obtained of the 2,3,7,8-chloro-substituted HxCDFs from other less toxic congeners using three different capillary columns, DB-5, DB-5ms, and DB-Dioxin. The choice of the columns depends on the sample analyzed, but the use of at least two columns of different polarity is recommended for the complete isomer separation.



**Figure 3** GC-HRMS selected ion monitoring of hexachlorodibenzo-*p*-dioxins ( $m/z$  373.8207) of a municipal waste incineration emission on DB-5, DB-5ms, and DB-Dioxin GC columns. Experimental conditions: GC columns: DB-5, DB-5ms, and DB-Dioxin (J&W Scientific) 60 m  $\times$  0.25 mm ID, 0.25  $\mu$ m film thickness. MS conditions: EI + at a resolution of 10 000 (10% valley definition). SIM mode. (Reprinted with permission from Abad E, Caixach J, and Rivera J (1997) Application of DB-5 ms gas chromatography column for the complete assignment of 2,3,7,8-substituted polychlorodibenzo-*p*-dioxins and polychlorodibenzofurans in samples from municipal waste incinerator emissions. *Journal of Chromatography A* 786: 125–134.)



The analysis of PCDD/Fs is difficult due to the low concentration levels (part per trillion) found in the environmental samples. GC–HRMS is the only technique capable of providing the required selectivity and sensitivity for the analysis of these compounds, and it is considered the reference technique for such analysis. Working in selected ion monitoring (SIM) mode at a mass resolution of 10 000, the presence of matrix components in the extracts does not interfere and detection at a high level of mass accuracy can be obtained. Quantification is generally performed by using stable isotope-labeled  $^{13}\text{C}_{12}$  analogs of the target compounds as internal standards (isotope dilution method). EI is the most popular ionization technique, although NCI, with methane as reagent gas, can be used for improving molecular mass determination and increasing sensitivity. Limits of detection for GC–EI–HRMS down to 10–200 fg have been obtained with the newest generation of HRMS systems. In recent years, ion-trap tandem mass spectrometry has been proposed as an alternative technique to HRMS for the analysis of these compounds. However, further studies related to the sensitivity and the selectivity of this technique must be performed in order to assure the quality of the results.

### Pesticides

The extensive use of pesticides to improve agricultural productivity played an important role in the twentieth century. These compounds have been applied for years to control the effect of pests and to increase food production. They were especially widely used after World War II. Nevertheless, this widespread use and the additional environmental pollution due to industrial emissions during their production have resulted in the occurrence of residues of these chemicals and their metabolites in the environment and the food chain. Since some of these compounds are important carcinogens and mutagens, regulation and environmental monitoring programs have been adopted in order to determine the risks to human health. At present, over 500 compounds are registered worldwide as pesticides or metabolites of pesticides. From these, organochlorine pesticides (OCPs) have been the most intensely used for agricultural purposes due to their insecticidal properties. OCPs such as those of the DDT group, HCH and its isomers, and cyclodienes are often detected in environmental samples, and they are ubiquitous across the globe, even in remote Polar Regions. Although most OCPs have been banned since the 1970s, they and their metabolites are still frequently found in different environmental samples due to their high persistence and lipophilicity, and they remain an

important source of contamination throughout the food chain. Organophosphorus (OPPs) and organonitrogen pesticides (ONPs) have been used to replace OCPs in many agricultural applications due to their lower toxicity and low residence time in the environment.

Capillary GC is the technique of widest application in pesticide analysis. Nowadays, more than 60% of the total pesticide and/or metabolites registered are amenable to GC. Only polar or semipolar pesticides are analyzed by using other techniques such as liquid chromatography. The selection of the GC stationary phase depends on the nature of the pesticides to be separated. For instance, for organochlorine and pyrethroid pesticides, nonpolar stationary phases such as methylpolysiloxane or 5% phenyl methylpolysiloxane are currently used. On the other hand, for the separation of more polar pesticides such as organophosphorus compounds, semipolar stationary phases such as 14% cyanopropylphenyl methylpolysiloxane or 50% phenyl methylpolysiloxane are often preferred, while polar stationary phases, e.g., DB-wax, are suitable for more polar compounds such as methamidofos. Splitless injection is currently used for GC analysis of pesticides due to its robustness, but on-column and PTV injectors have also been used. Moreover, direct injection of large volumes combined with PTV is an interesting approach for the analysis of pesticides in water samples with an important reduction of the detection limits. Due to the different characteristics of the pesticides, selective detectors based on the measure of some specific properties are currently used in their quantification so as to achieve enough selectivity and sensitivity. For instance, ECD is used for the analysis of halogenated pesticides, TID for triazines and organophosphorous pesticides, and FPD for organotin and organosulfur compounds. Nevertheless, the use of these selective detectors is being progressively replaced by GC–MS for pesticide residue analysis, mainly due to the proved high sensitivity of CI and the high selectivity of MS/MS. Moreover, ion-trap MS working in MS/MS mode has shown itself to be a powerful technique for pesticide residue analysis at low concentration levels.

### Emerging Contaminants

A new and important field of research in environment analysis is the study of the so-called emerging contaminants. These compounds are mainly unregulated pollutants that were previously undetected or considered as risk free. Nevertheless, the development of highly sensitive analytical techniques has allowed the establishment of methodologies for the

determination of these compounds. Among the emerging contaminants, brominated flame retardants, chlorinated paraffins, petrol additives, new DBPs, and short-chain alkylphenol ethoxylates can be analyzed by GC.

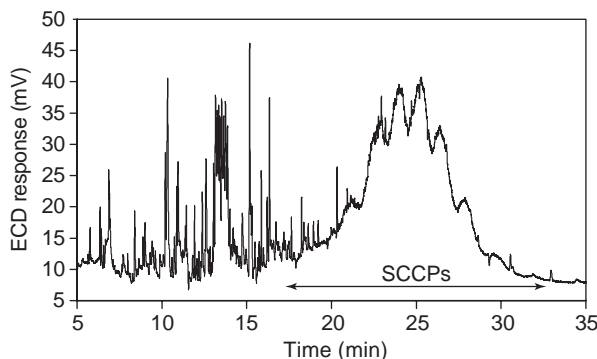
Brominated flame retardants are widely used as polymer additives in plastics, electric appliances, and textile and polyurethane foams. Some of the technical flame retardant products contain brominated organic compounds, including polybrominated biphenyls (PBBs) and polybrominated diphenyl ethers. These compounds are structurally similar to PCBs and, therefore, their chemical properties, persistence, bioaccumulation, and distribution in the environment follow similar patterns. Moreover, they have potential endocrine disrupting properties and there are concerns over their exposure health effects.

The analysis of PBBs is more difficult than that of PCBs mainly because the higher brominated compounds can decompose during GC analysis at high temperatures. Generally, GC with nonpolar stationary phases such as 5% phenyl methylpolysiloxane is used for congener-specific determination of PBB in environmental samples. Semiquantitative total-PBB analyses are usually performed because the number of available individual PBB standards is very limited. As highly chlorinated PCBs and PCTs may interfere with low brominated PBBs during GC analysis, ECD is not recommended for detection. MS, especially in NCI, which offers high sensitivity for compounds with four or more bromine atoms, is the technique currently used for the analysis of these compounds. For PBDE analysis long capillary columns (30–50 m) are used with stationary phases such as 5% phenyl methylpolysiloxane and/or 14% cyanopropylphenyl methylpolysiloxane in order to have enough resolution among congeners. Nevertheless, short columns (10–15 m) are proposed for the higher brominated PBDEs, such as BDE209, which shows long retention times and are thermally labile. Both ECD and MS have been proposed as detectors, but co-elution of PBDE congeners with organochlorine pesticides and PCBs which elute together in the same fraction when silica is used for cleanup presents problems when ECD is used. These problems can be solved by MS detection. For LRMS, both EI and NCI have been proposed but NCI provides better sensitivity for the higher brominated compounds (hepta- to decabDEs). A drawback of NCI is the lack of selectivity, since for most of the compounds only the bromine cluster ions ( $m/z$  79 and 81) are monitored. In contrast, EI offers more selectivity and confirmation from the mass spectrum but has a much lower sensitivity. Electron ionization high-resolution MS (EI-HRMS) has the advantage of increased selectivity

and sensitivity, although it requires experienced personnel and is more expensive and labor-intensive than LRMS.

Chlorinated paraffins (CPs) are complex industrial formulations of polychlorinated *n*-alkanes (PACs) with carbon chain lengths between  $C_{10}$  and  $C_{30}$  and a degree of chlorination between 30% and 70%. These compounds are used as additives in cutting fluids, paints, plasticizers, and sealants as well as flame retardants. The analysis of CPs is difficult due to the large number of congeners (a minimum of several thousands) present in the technical mixtures. Generally, nonpolar GC capillary columns such as methylpolysiloxane and 5% phenyl methylpolysiloxane are used and the chromatograms show a characteristic broad profile of unresolved peaks. This profile is used for quantitative analysis, as can be seen in **Figure 4**, where the chromatogram of a water sample containing short-chain CPs is given. CPs are usually detected by ECD or by electron-capture negative ion MS at both low and high resolutions.

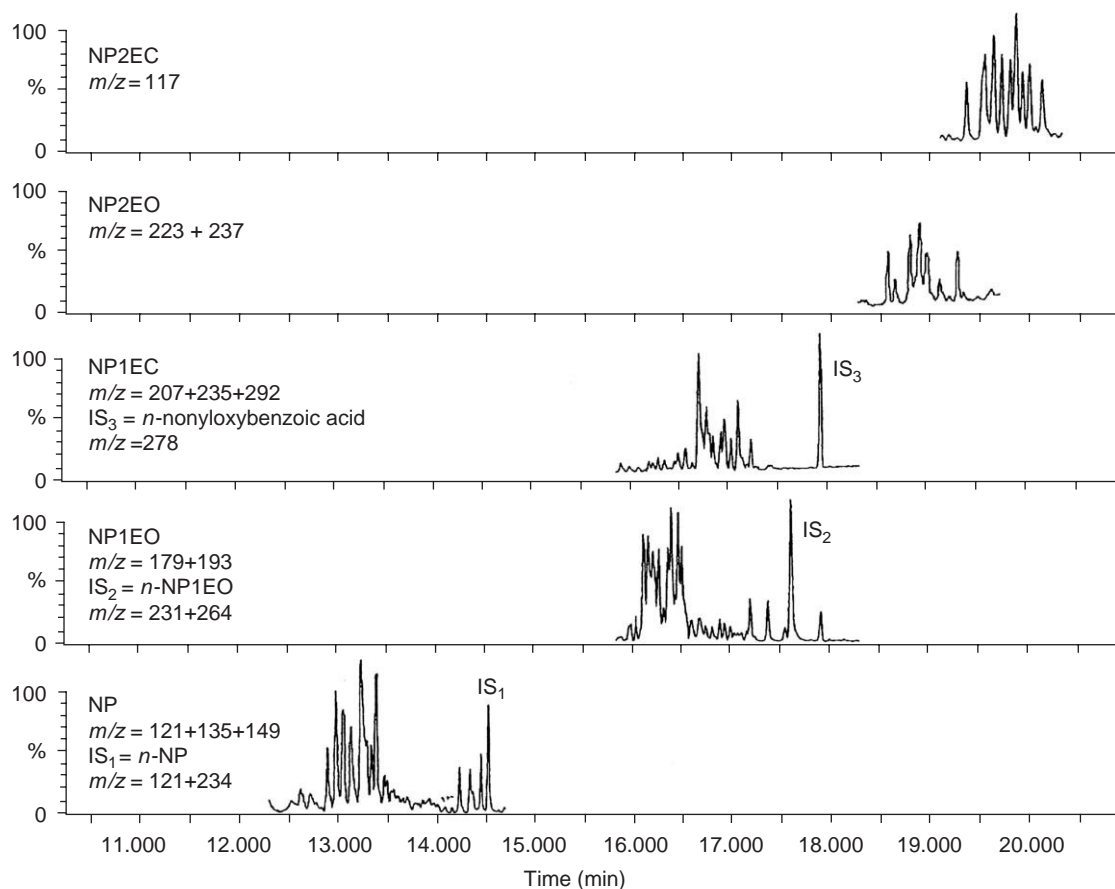
Methods including headspace, purge and trap, or SPME combined with GC have been used to measure methyl *tert*-butyl ether (MTBE) in ground water and surface water. MTBE has been used as a petrol additive since its introduction in 1979, and water contamination is due to leaking underground storage gasoline tanks and discharges from fuel from boats. This compound is responsible for odor and taste problems in drinking water. GC with nonpolar stationary phases and FID, photoionization, and MS detection systems are the most frequently used methods for the analysis of this compound.



**Figure 4** SPME-GC-ECD chromatogram of a water river sample. Elution time of short-chain chlorinated paraffins is indicated in the chromatogram. Experimental conditions: GC column: DB-5 (J&W Scientific) 30 m  $\times$  0.25 mm ID, 0.25  $\mu$ m film thickness. (Reprinted with permission from Castells P, Santos FJ, and Galceran MT (2003) Solid-phase microextraction for the analysis of short-chain chlorinated paraffins in water samples. *Journal of Chromatography A* 984: 1–8.)

DPBs beyond those that are currently regulated (Table 2) are becoming important. For instance, brominated DBPs are now being recognized as toxicologically important because they are proving to be much more carcinogenic than their chlorinated analogs. Moreover, preliminary studies suggest that iodinated compounds may be more toxic than their brominated analogs. These compounds are formed when relatively high concentrations of bromide and/or iodide are present in source waters, as frequently happens in coastal cities. Specific DBPs of current interest are bromo- and iodo-THM, cyanogen bromide, bromonitromethanes, and brominated forms of 3-chloro-4-(dichloromethyl)-5-hydroxy-2-(5H) furanone (MX). The methods of analysis of these compounds are similar to those used for the regulated DBPs but some of them require special GC analytical conditions because they are thermally unstable and decompose under the commonly used injection temperatures in GC analysis.

Other compounds that have received considerable attention in the last few years are alkylphenol polyethoxylate (AP $n$ EOs,  $n$ =number of ethoxy units) degradation products such as octyl- and nonylphenols and the corresponding mono- and diethoxylates. These compounds have been included in the list of priority substances in the field of water policy and amending of the EU because of their widespread use and potential endocrine disruption capacity. Short ethoxylated chain AP $n$ EOs, octyl and nonylphenols, their carboxylate metabolites, and also the halogenated derivatives that are formed during chlorine disinfection can be analyzed using GC after derivatization. Capillary GC with nonpolar stationary phases, e.g., 5% phenyl methylpolysiloxane (DB-5), coupled to MS can be used for the analysis of these compounds. As an example, Figure 5 shows a GC–MS total-ion chromatogram of methylated nonylphenol, nonylphenol ethoxylates, and their acidic metabolites in a river water sample.



**Figure 5** Headspace-SPME–GC–MS single-ion chromatograms of derivatized compounds from river water entering water treatment plant. NP: nonylphenol, NP1EO: nonylphenol monoethoxylate, NP2EO: nonylphenol diethoxylate, NP1EC: nonylphenoxyacetic acid, NP2EC: nonylphenoxyethoxyacetic acid. Experimental conditions: GC column: DB-5 MS, 30 m  $\times$  0.25 mm ID, 0.25  $\mu$ m film thickness. MS analyzer: quadrupole, operating in SIM mode. (Reprinted with permission from Díaz A, Ventura F, and Galceran MT (2002) Simultaneous determination of estrogenic short ethoxy chain nonylphenols and their acidic metabolites in water by an insamples derivatization/solid-phase microextraction method. *Analytical Chemistry* 74: 3869–3876.)

See also: **Dioxins. Gas Chromatography:** Column Technology; Multidimensional Techniques; High-Speed Techniques; Instrumentation; Detectors; Mass Spectrometry. **Lead. Mass Spectrometry:** Electron Impact and Chemical Ionization; Ion Traps; Selected Ion Monitoring. **Mercury. Pesticides. Polychlorinated Biphenyls. Polycyclic Aromatic Hydrocarbons:** Determination. **Tin.**

## Further Reading

De Boer J (1999) Capillary gas chromatography for the determination of halogenated micro-contaminants. *Journal of Chromatography A* 843: 179–198.

Kleiböhmer W (ed.) (2001) *Handbook of Analytical Separations – Vol. 3: Environmental Analysis*. Amsterdam: Elsevier Science B.V.

Niessen WMA (eds.) (2001) *Current Practice of Gas Chromatography-Mass Spectrometry*. Chromatographic Science Series, vol. 86. New York: Dekker.

Santos FJ and Galceran MT (2002) The application of gas chromatography to environmental analysis. *Trends in Analytical Chemistry* 21: 672–685.

Santos FJ and Galceran MT (2003) Modern developments in gas chromatography–mass spectrometry-based environmental analysis. *Journal of Chromatography A* 1000: 125–151.

## Forensic Applications

**H Neumann**, Bundeskriminalamt Kriminaltechnisches Institut, Wiesbaden, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Following ultraviolet–visible spectrophotometry and infrared (IR) spectroscopy, gas chromatography (GC) was one of the first instrumental techniques to help in solving forensic science problems. The early very successful applications included the determination of blood alcohol by direct injection of blood or serum, and the detection and identification of petroleum products in debris from arson cases in 1958/59. The breakthrough of GC in these areas and in drug analysis was an event of the 1960s and the 1970s.

The routine use of capillary columns for much better separation compared with packed columns was achieved during the 1980s and the 1990s, thus allowing, for example, much better discrimination for the forensic comparison of samples. Other practical improvements during that time included the automation of GC instruments, from the injection of samples using autosamplers to the overall instrument control and evaluation of results by modern personal computers. Recent advances in GC have been achieved in combining the very successful technique of headspace solid-phase microextraction (SPME) as sample preparation method with the GC analysis. Especially in arson investigations, this solvent-free technique, with the ability for field sampling, has fundamentally changed the workup and detection of ignitable solvents. Other new improvements include the introduction of the Fast GC technique, which means the reduction of separation times for

multicomponent mixtures from hours to minutes by using smaller diameter, shorter capillary columns with the necessary adjustments of the carrier gas flow and using electronic pressure control.

Today, GC is still one of the most widely used instrumental techniques in forensic science laboratories around the world, now mostly in combination with a mass spectrometer as the substance-specific detector. The types of specimens analyzed range from gases (e.g., cyclopropane) to polymers (e.g., plastics), which can be made amenable to GC through pyrolysis.

## Alcohol in Blood

The gas chromatographic determination of ethanol in blood has been a well-established method since the 1960s. It is also today one of the classical examples for the application of GC in the forensic sciences. The early introduction of headspace sampling has simplified greatly sample preparation, injection, and analysis (no contamination of the column and the injector by blood residues). The automation of the whole procedure was also achieved very early and today with coupled automated data-handling systems the requirements of standard quality control (QC) procedures can be fulfilled in a very good manner. Flame ionization detection is the favorite detection technique. The headspace gas chromatographic procedure has been proven as a highly specific method with good precision and reproducibility. Typical results of the reproducibility for within-run and between-run coefficients of variation are <1.5 and <2.5%, respectively.

Scientific developments in gas chromatographic analysis in the 1980s and the 1990s concentrated on the detection and determination of congeners from

See also: **Dioxins. Gas Chromatography:** Column Technology; Multidimensional Techniques; High-Speed Techniques; Instrumentation; Detectors; Mass Spectrometry. **Lead. Mass Spectrometry:** Electron Impact and Chemical Ionization; Ion Traps; Selected Ion Monitoring. **Mercury. Pesticides. Polychlorinated Biphenyls. Polycyclic Aromatic Hydrocarbons:** Determination. **Tin.**

## Further Reading

De Boer J (1999) Capillary gas chromatography for the determination of halogenated micro-contaminants. *Journal of Chromatography A* 843: 179–198.

Kleiböhmer W (ed.) (2001) *Handbook of Analytical Separations – Vol. 3: Environmental Analysis*. Amsterdam: Elsevier Science B.V.

Niessen WMA (eds.) (2001) *Current Practice of Gas Chromatography-Mass Spectrometry*. Chromatographic Science Series, vol. 86. New York: Dekker.

Santos FJ and Galceran MT (2002) The application of gas chromatography to environmental analysis. *Trends in Analytical Chemistry* 21: 672–685.

Santos FJ and Galceran MT (2003) Modern developments in gas chromatography–mass spectrometry-based environmental analysis. *Journal of Chromatography A* 1000: 125–151.

## Forensic Applications

**H Neumann**, Bundeskriminalamt Kriminaltechnisches Institut, Wiesbaden, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Following ultraviolet–visible spectrophotometry and infrared (IR) spectroscopy, gas chromatography (GC) was one of the first instrumental techniques to help in solving forensic science problems. The early very successful applications included the determination of blood alcohol by direct injection of blood or serum, and the detection and identification of petroleum products in debris from arson cases in 1958/59. The breakthrough of GC in these areas and in drug analysis was an event of the 1960s and the 1970s.

The routine use of capillary columns for much better separation compared with packed columns was achieved during the 1980s and the 1990s, thus allowing, for example, much better discrimination for the forensic comparison of samples. Other practical improvements during that time included the automation of GC instruments, from the injection of samples using autosamplers to the overall instrument control and evaluation of results by modern personal computers. Recent advances in GC have been achieved in combining the very successful technique of headspace solid-phase microextraction (SPME) as sample preparation method with the GC analysis. Especially in arson investigations, this solvent-free technique, with the ability for field sampling, has fundamentally changed the workup and detection of ignitable solvents. Other new improvements include the introduction of the Fast GC technique, which means the reduction of separation times for

multicomponent mixtures from hours to minutes by using smaller diameter, shorter capillary columns with the necessary adjustments of the carrier gas flow and using electronic pressure control.

Today, GC is still one of the most widely used instrumental techniques in forensic science laboratories around the world, now mostly in combination with a mass spectrometer as the substance-specific detector. The types of specimens analyzed range from gases (e.g., cyclopropane) to polymers (e.g., plastics), which can be made amenable to GC through pyrolysis.

## Alcohol in Blood

The gas chromatographic determination of ethanol in blood has been a well-established method since the 1960s. It is also today one of the classical examples for the application of GC in the forensic sciences. The early introduction of headspace sampling has simplified greatly sample preparation, injection, and analysis (no contamination of the column and the injector by blood residues). The automation of the whole procedure was also achieved very early and today with coupled automated data-handling systems the requirements of standard quality control (QC) procedures can be fulfilled in a very good manner. Flame ionization detection is the favorite detection technique. The headspace gas chromatographic procedure has been proven as a highly specific method with good precision and reproducibility. Typical results of the reproducibility for within-run and between-run coefficients of variation are <1.5 and <2.5%, respectively.

Scientific developments in gas chromatographic analysis in the 1980s and the 1990s concentrated on the detection and determination of congeners from



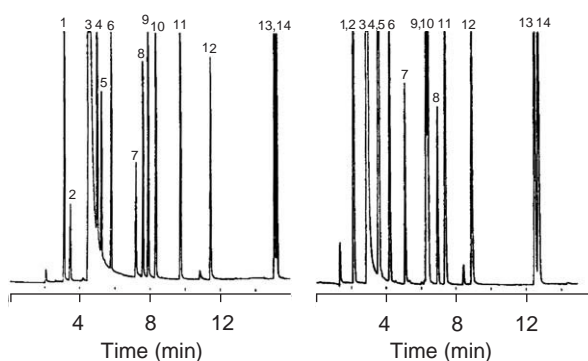
alcoholic beverages in blood, serum, and urine. After thorough studies in this complex area, the assessment of the type and amount of consumed alcoholic beverage is possible with some limitations, particularly in cases of drinking after a forensically relevant event. As always in the forensic sciences, much background information concerning the drinking event, the type of the suspected consumed alcoholic beverage (for example, the original sample found at a suspect's home), and metabolism of special congener compounds is necessary for applications of proper expertise in this field. Differences in the individual absorption rates of alcohol also have to be taken into account. 1-Propanol and isobutanol are the most important congener alcohols and need to be quantified in each analysis. It has to be considered that some beverages (like vodka and grain spirit) contain no forensically relevant concentrations of congeners and that through nonaseptic handling 1-propanol and isobutanol can be produced during the preparation and storage of blood samples, taken for blood alcohol analysis. Capillary GC with microheadspace sampling and the use of different internal standards for compensation of matrix effects has proved effective for good quantification of important congeners. The use of the flame ionization detector (FID) is part of this basic technique, but today as a second indispensable detection mode the use of a mass spectrometer is highly recommended to identify additional compounds that may be used as markers. **Figure 1** shows, as an example, chromatograms of congener compounds analyzed on two different capillary columns with an FID being used for detection. A mass spectrometer is recommended for the detection of additional compounds. These compounds could then be added to a retention index list.

The well-established GC methods for blood ethanol determination are today the Gold Standard in cases, where, for instance, results from the quantitative breath-alcohol analyzer used for evidential purpose are challenged. The basis for this importance is an overwhelming amount of data available from the GC tests and the rigorous QC scheme for precision and accuracy testing.

## Drug Determinations

A major work for all forensic science laboratories is directed to the analysis of drugs of abuse. GC plays a prominent role in analyzing these exhibits, both for identification of the particular drug (qualitative analysis) and for the determination of the amount of pure drug in the sample (quantitative analysis).

One of the advantages of using capillary GC in the determination of drugs is that a wide variety of



**Figure 1** Gas chromatograms of an aqueous calibration mixture for analysis of congeners of alcoholic beverages in blood after splitless headspace sampling. Left: capillary column 60 m  $\times$  0.32 mm i.d. coated with DB 1701, film thickness 1  $\mu$ m. Right: capillary column 60 m  $\times$  0.32 mm i.d. coated with CP-Sil 8-CB, film thickness 2.8  $\mu$ m. Carrier gas 1.5 bar (150 kPa) hydrogen; manual injection of 0.5 ml, splitless 6 s, then 50 ml min<sup>-1</sup> split flow. Concentration of congeners  $\sim$ 0.8 mg l<sup>-1</sup>, methanol  $\sim$ 5 mg l<sup>-1</sup>, ethanol  $\sim$ 1.6 g l<sup>-1</sup>. Peaks: 1, acetaldehyde; 2, methanol; 3, ethanol; 4, acetone; 5, 2-propanol; 6, *t*-butyl alcohol; 7, 1-propanol; 8, ethyl acetate; 9, ethyl methyl ketone (2-butanone); 10, 2-butanone; 11, isobutyl alcohol (2-methyl-1-propanol); 12, 1-butanol; 13, 3-methylbutanol; 14, 2-methylbutanol. (Reproduced with permission from Wolf M, Weiler JP, Urban R, and Tröger HD (1987) *Zur Begleitstoffanalytik 2. Mitteilung: Aspekte der quantitativen Bestimmung. Blutalkohol* 24: 378–390; © Bund gegen Alkohol und Drogen.)

chemically different compounds can be chromatographed in the same system. This has led, for instance, to the very comprehensive and successful approach for the identification of drugs and poisons using capillary GC and mass selective detection by Pfleger, Maurer, and Weber resulting in their famous collection of mass spectra and gas chromatographic data. The unique separation efficiency of this technique is also essential for the detailed analysis of illicit drugs as required for the comparison of different samples and seizures. Today, most GC applications in this field use capillary columns with nonpolar or slightly polar phases and mass selective detection and/or FID. The FID is well suited for quantification of the separated compounds because it has a linear response over seven orders of magnitude, it is very stable, and the response is reproducible. Normally, quantification is done using the internal standard method. Nitrogen-selective detection is sometimes used in forensic toxicology. The new types of mass spectrometers like the ion-trap quadrupole instruments are also used as detectors because they enable controlled experiments of further fragmentation of some key ions in order to obtain an even more specific and sensitive method. The electron-capture detector (ECD), as an example of not only a selective and sensitive but also a more complicated detector to

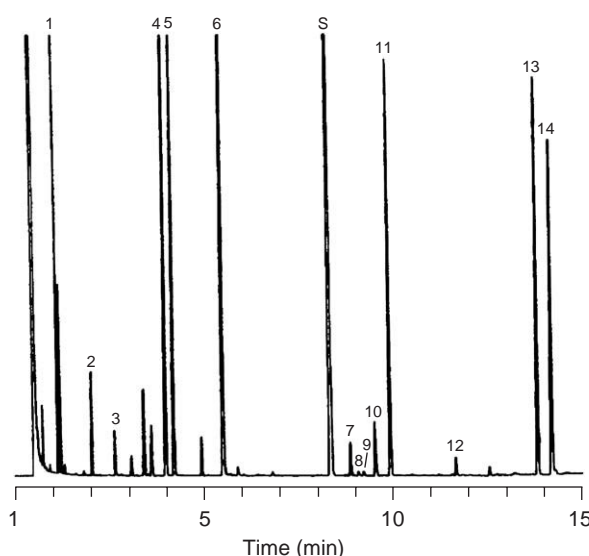
operate, is only applied for special tasks, such as the detection of trace impurities in illicit drug samples after derivatization with heptafluorobutyric anhydride. Also, only a few applications were reported for the coupling of GC with an IR spectroscopic detector. This combination was used for the differentiation of isomers of amphetamine derivatives because here mass spectrometry (MS) alone has some limitations to unequivocally identify the specific isomer.

In spite of the fact that the basic GC techniques used are much the same for analysis of solid forms and for the determination of drugs in biological specimens like body fluids and hair, the special needs for adequate sample preparation and adjustments to the different matrices have led to distinct developments in the two areas and are discussed separately here.

### Drugs in Solid Form

Large amounts of illicit drugs seized by police and customs are normally in solid form, such as powders, tablets, capsules, and those impregnated on paper. The main requirements for law enforcement and the courts are to identify correctly the material and in many cases to report a percentage purity of the active and controlled compound, e.g., diacetylmorphine in illicit heroin. For most drugs of abuse GC is well suited to fulfill both these tasks, but other chromatographic techniques also have their role in the forensic analysis of drugs. Thin-layer chromatography (TLC), for instance, is generally cheaper and easier to operate, whereas liquid chromatography (LC) is suitable in cases where the compound in question can thermally decompose during GC. Another modern alternative to GC and LC is the use of capillary electrophoresis, for instance, in cases to determine the enantiomeric composition of certain drugs.

Derivatization is mostly used if the drug does not chromatograph well without chemical modification and in cases where better separation of the derivatized substances can be expected. Derivatization is also a good approach in cases where a complex mixture of original drug ingredients, added pharmacologically active compounds (adulterants), and inactive substances (diluent) such as sugars have to be separated. **Figure 2** shows an example from our laboratory where silylation with *N*-methyl-*N*-trimethylsilyltrifluoroacetamide (MSTFA) was used to identify and quantify the opium alkaloids and derivatives and the 'cutting' substances in an illicit heroin sample in one analytical run. The silylation conditions used with heroin, crude morphine, and opium samples help to avoid many possible problems such as transacetylation, adsorption, and different responses for salt and base. The alkaloid derivatives from acetylcodeine



**Figure 2** Gas chromatogram of an illicit heroin sample. Conditions: fused-silica capillary column 25 m  $\times$  0.32 mm i.d. coated with OV-1 cb; carrier gas 0.7 bar (70 kPa) hydrogen; split  $\sim$  1:15, make-up gas argon at 18 ml min<sup>-1</sup>; oven temperature program 150–280°C at 9°C min<sup>-1</sup>, then 0.5 min isothermal; injector/detector (FID) temperatures 250/280°C. Sample silylated with MSTFA. Peaks: 1, nicotinamide-TMS; 2, meconin; 3, caffeine; 4, glucose-TMS; 5, phenobarbital-TMS; 6, methaqualone; 7, acetylcodeine; 8, acetylthebaol; 9, morphine-TMS; 10, 6-O-acetylmorphine-TMS; 11, diacetylmorphine; 12, papaverine; 13, phenolphthalein-TMS; 14, narcotine; S, tetracosane (internal standard) (TMS, trimethylsilyl derivative).

(peak 7) to diacetylmorphine (peak 11) in **Figure 2** are better separated with derivatization than without using this sample preparation step.

Cocaine can also be analyzed well by GC. Here again silylation can help to achieve good separation of major, minor, and trace components of the illicit cocaine samples and was therefore proposed as early as in 1973 by Moore. The sample preparation procedure used in our laboratory for the analysis of heroin and cocaine starts by dissolving the sample in chloroform. Then a mixture of MSTFA and pyridine is added and allowed to react at 80°C for 10 min. After that the solution is chromatographed directly.

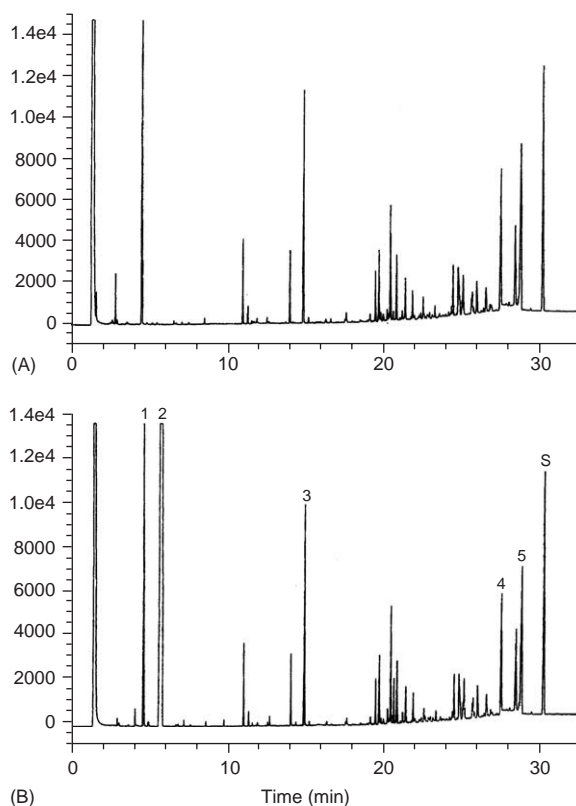
GC is also the predominant technique to analyze cannabis products like marijuana and hashish. Here, the determination of the 'total tetrahydrocannabinol (THC) content' or the 'available THC content' is useful, because THC-acid is present in these samples as well as THC. The application of direct GC is the technique of choice for this determination, after extraction from the plant material with a suitable solvent such as methanol or chloroform. Methaqualone and phencyclidine and their analogs are other drug compounds well suited for gas chromatographic analysis. For some benzodiazepines, thermal

decomposition during GC has been reported. For barbiturates and psilocybin, derivatization has to be applied. Lysergic acid diethylamide also cannot be easily analyzed by GC.

The detailed analysis of drug samples, named 'profiling', for the purpose of sample comparison and/or identification of key impurities, characteristics of certain routes of synthesis, or used in origin-related studies ('markers') was much improved by the application of capillary columns in GC. The state-of-the-art applications of this high-resolution technique can be divided into two approaches. In the first, the sample is analyzed by capillary GC without selective extraction or special pretreatment. In the second, the classic liquid/liquid extraction or similar steps like solid-phase extraction (SPE) are used to pre-separate the characteristic impurities (congeners, by-products of synthesis, etc.) from the bulk drug matrix to allow for their detection in trace amounts. **Figure 3** shows the trace impurity profiles (drug signatures) of two illicit heroin samples after toluene extraction from  $0.5 \text{ mol l}^{-1}$  sulfuric acid. The gas chromatographic profiles are stored in a database for retrospective comparisons, using 14 characteristic impurity compounds as descriptors. By searching the database with the help of a statistical method these two samples could be grouped together.

In general, the optimization of an appropriate extraction method (e.g., with respect to solvent type, pH value, volumes) in combination with the selection of the characteristic impurity compounds is critical for a good reproducibility of the whole procedure. This proven reproducibility is the basis of the successful application of this approach. After groundbreaking work on amphetamine produced by the Leuckart synthesis, heroin, and cocaine in the 1970s and the 1980s, the last few years have seen a further refinement of the developed profiling GC procedures for these drugs and an application to the investigation of amphetamine derivatives like methamphetamine and 3,4-methylenedioxymethamphetamine (MDMA), the main ingredient of the so-called 'ecstasy tablets'. The quantitation of MDMA in illicit preparations is normally done by LC. But for detailed studies directed to the identification of key impurities for identification of the route of synthesis and for the comparison of samples the application of capillary GC is preferred. **Figure 4** shows chromatograms of minor and trace components from three typical MDMA tablets. The peak areas of the selected impurity substances form the basis for the comparison of samples.

In the most recent developments SPME in combination with capillary GC and GC-MS is used for the characterization of volatile impurities in illicit

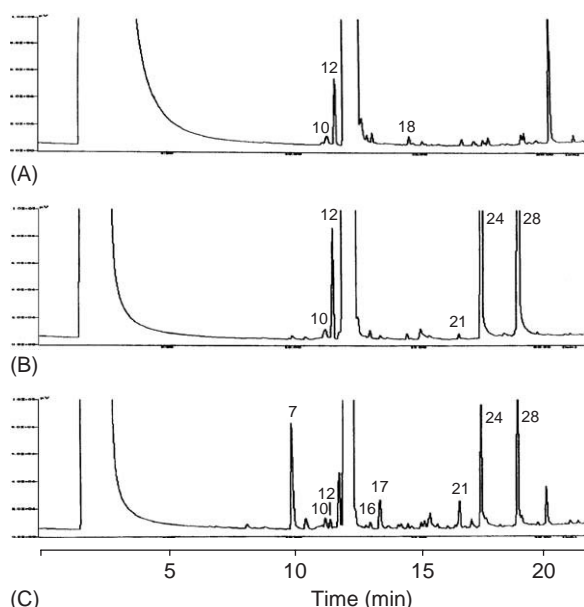


**Figure 3** Sample comparison via capillary gas chromatographic impurity profiles of two seizures of Southwest Asian heroin base. Note the dilution of the sample at the bottom with caffeine (peak no. 2). Peaks: 1, meconin; 2, caffeine; 3, acetylthebaol; 4, (*E*)-*N*-acetylanhydronarceine; 5, (*Z*)-*N*-acetylanhydronarceine; S, tetracontane (internal standard). (Reproduced with permission from Strömberg L, Lundberg L, Neumann H, Bobon B, Huizer H, and Van der Stelt NW (2000) Heroin impurity profiling. A harmonization study for retrospective comparisons. *Forensic Science International* 114: 67–88; © Elsevier Ireland Ltd.)

amphetamine, methamphetamine, 4-methoxyamphetamine, and MDMA samples. In this solvent-free sample preparation technique, a fiber coated, for instance, with polydimethylsiloxane (PDMS) or polyacrylate is used to concentrate the analyte. Normally, the fiber is then put into the injector of the gas chromatograph and the compounds are thermally desorbed to the column. In general, these profiling methods have demonstrated their practical value for the forensic comparison of illicit drug samples.

### Drugs in Biological Specimens

The most prominent biological samples used for testing for drugs of abuse are urine, blood, hair, and saliva. Drugs can also be detected in many other biological specimens, such as sweat, vitreous humor, and autopsy material such as liver in general



**Figure 4** Typical gas chromatographic impurity profiles of different MDMA samples. The big peak at retention time 12.2 min (after peak no. 12) corresponds to MDMA. Peaks: 7, 3,4-methylenedioxy-*N*-methylbenzylamine; 10, 3,4-methylenedioxyphenyl-2-propanone; 12, 3,4-methylenedioxy-phenyl-2-propanol; 16, *N*-ethyl-3,4-methylenedioxyamphetamine; 17, *N,N*-dimethyl-3,4-methylenedioxyamphetamine; 18, 1-(1,2-dimethyl-1-azacyclopentyl)methyl-3,4-methylenedioxybenzene; 21, caffeine; 24, palmitic acid; 28, stearic acid. (Reproduced with permission from Palhol F, Boyer S, Naulet N, and Chabrilat M (2002) Impurity profiling of seized MDMA tablets by capillary gas chromatography. *Analytical and Bioanalytical Chemistry* 374: 274–281; © Springer-Verlag.)

toxicological analysis for the detection of poisons and other unusual substances.

The key element that sets analysis of drugs in biological specimens like urine and plasma apart from the analysis in solid form is the necessity for the appropriate sample preparation. This involves isolation and, if necessary, cleavage of conjugates and/or derivatization. Traditionally, liquid–liquid extraction is used for isolation. However, in the last 25 years SPE has been introduced as a very efficient technique. Its main advantages after careful selection and testing with various drugs are that higher selectivity is achieved, cleaner extracts can be obtained, and this type of sample preparation can be easily automated. The last development in this area was the application of SPME to the investigation of biological samples such as whole blood, plasma, urine, and hair for illicit drugs and poisons. The analytes included are, for instance, amphetamine and derivatives like methamphetamine and MDMA, cocaine, THC, and benzodiazepines.

The state-of-the-art technique for the analysis of illicit drugs in body fluids is GC–MS, mostly using

mass selective detectors. Recommended guidelines for forensic GC–MS procedures in toxicology laboratories analyzing biological samples have been published to establish a high standard of practice. A good-quality, full-scan mass spectrum and the analyte's expected retention time in the gas chromatogram provide the best criteria for a sure forensic identification of drugs of abuse, e.g., in the context of urine testing. Duplicate specimens should allow an independent analysis when the results of testing of the first sample are challenged. For support of the systematic toxicological analysis of drugs, poisons, pesticides, and their metabolites, the excellent collection of mass spectra and GC data from Pfleger, Maurer, and Weber is very helpful.

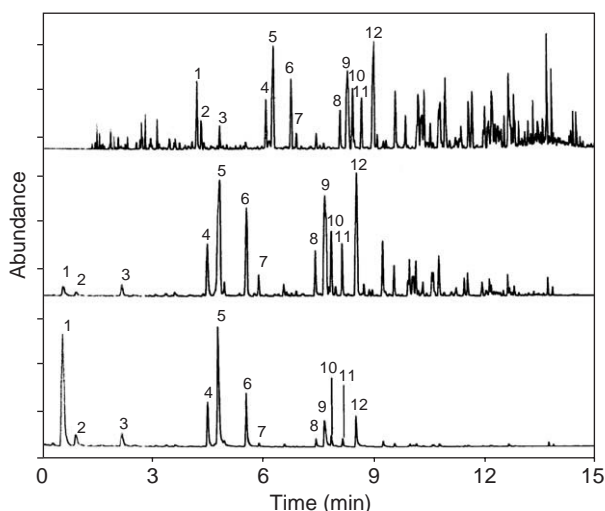
While biological fluids allow the detection of drugs after a recent misuse (normally a few days), hair is an interesting substrate for the investigation of chronic drug abuse (over periods up to 6 months and 1 year). For detection of drugs in human hair again GC–MS is the most important confirmation technique in connection with immunoassay screening. Careful consideration with hair must be given to the possibility of external contamination by drugs of abuse, because not all wash procedures are effective in removing the contaminating substances. The detection of heroin and its first biotransformation product, 6-acetylmorphine, in hair by GC–MS was used to differentiate heroin exposure from other sources of opiates, for example, licit codeine preparations or poppy seeds.

## Arson, Explosives, and Propellant Residues

After an investigation of a scene of fire or explosion, debris samples are frequently collected and sent to the forensic science laboratory for analysis of the presence of an accelerant or residues of explosives.

GC was introduced very early as the technique of choice for the detection and identification of accelerants in debris from arson cases because of its high selectivity and sensitivity. But to use the full potential of the technique the methods of recovery of traces of common accelerants from fire debris had to be developed and adjusted. The used methods include solvent extraction, direct headspace analysis, and enrichment by adsorbent-based techniques. In the past, the most common concentration steps prior to the analysis have been (heated) headspace direct injection using a gastight syringe for analyte collection and GC injection or headspace adsorption techniques, mostly using charcoal followed by carbon disulfide (CS<sub>2</sub>) elution. Some of these procedures have been quite effective and are standardized by





**Figure 5** Gas chromatographic comparison of gasoline component recovery by headspace SPME sampling using a Carboxen/PDMS fiber. Top: results obtained by direct injection of gasoline. Middle: headspace sampling by SPME over neat gasoline. Bottom: headspace sampling by SPME of gasoline on carpet. Peaks: 1, toluene; 2, 3-methylheptane; 3, octane; 4, ethylbenzene; 5, 1,3- and 1,4-dimethylbenzene; 6, 1,2-dimethylbenzene; 7, nonane; 8, propylbenzene; 9, 1-ethyl-(2 or 3)-methylbenzene; 10, 1,2,5-trimethylbenzene; 11, 1-ethyl-2-methylbenzene; 12, 1,2,4-trimethylbenzene. (Reprinted with permission from Lloyd JA and Edmiston PL (2003) Preferential extraction of hydrocarbons from fire debris samples by solid phase microextraction. *Journal of Forensic Sciences* 48: 130–134; © ASTM International.)

American Society for Testing and Materials. A new and very successful development of the last 10 years was the introduction of headspace SPME for sampling of arson accelerants. The method uses a fiber coated, for instance, with Carboxen/PDMS for concentration of ignitable liquids from fire debris samples. SPME has a lower detection limit for some analytes, it is a fast technique, and is easy to use. So, for instance, the older procedure with 16 h of sampling using charcoal strips can be replaced by an SPME method that lasts only 10 min. **Figure 5** shows an example where the chromatogram from direct injection of gasoline is compared with the headspace sampling by SPME over neat gasoline and of gasoline on carpet. The result was that the Carboxen/PDMS fiber preferentially adsorbs the aromatic hydrocarbons quite effectively. In SPME, the thermal analyte desorption mode, used normally to introduce the substances from the SPME fiber to the gas chromatograph, can be replaced by solvent-based desorption. Here again, elution with  $\text{CS}_2$  is most effective and the extracts can be archived and used for repeated automated analysis.

As column coatings for the capillary columns, used in the gas chromatographic analysis of arson accelerants, nonpolar phases of the methyl silicone type

are preferred, because hydrocarbons are eluted in boiling-point order. Capillary GC, mostly coupled now to a mass spectrometer, is an excellent technique for the comparison of unchanged automotive gasoline. By comparing the relative concentrations of hydrocarbons it was found that most of the discriminating information is contained in the fraction with a volatility ranging from *n*-pentane to *n*-octane. When comparing gasoline recovered from a fire scene with a possible source, several factors must be considered, including the variability introduced by the recovery method, evaporative losses from the gasoline, possible contamination by pyrolysis products (mainly aromatic hydrocarbons), and other alterations of the recovered gasoline by chemical and physical interactions. Typical examples of interfering organic compounds include lanolin from carpet underlay, terpenes from wood samples, and styrene, toluene, and benzene from polystyrene or other polymeric components of construction and furnishing materials. GC–MS can help to rapidly identify background pyrolysis products and help to distinguish these matrix interferences from common accelerants. As a result of a detailed study in arson cases it was concluded that a comparison by capillary GC is extremely useful in eliminating the possibility of a common origin of two samples, but it is very difficult to determine conclusively that two samples have a common origin.

In conclusion, good discrimination can be achieved of unchanged and weathered (altered) gasoline, gas oils (diesel), and crude oils by using GC with capillary columns. For the comparison of unchanged gasoline samples the use of the extracted ion chromatogram technique in GC–MS from the elution region between *n*-pentane and *n*-octane is very useful. The sampling for this part of the gas chromatogram can easily be achieved with the SPME technique using a fiber coated with PDMS. The intensities of characteristic gasoline compounds in the ion chromatograms can be compared visually and evaluated with statements like ‘smaller’, ‘bigger’, etc. These data can be stored in a database, together with other information concerning the gasoline sample, and are used for retrospective comparisons. The prerequisites for the successful application of this method using the extracted ion chromatograms is that the whole analytical procedure be regularly checked by ‘calibration gasoline’ and the comparison investigations be performed under the exact same conditions.

The detection of explosives and propellants in postblast residues and in hand swabs from persons suspected of firing a gun or handling explosives is extremely important for the success of a prosecution. The role of GC for the analysis of these polar and



sometimes thermally unstable compounds is limited, and therefore other chromatographic techniques such as LC, TLC, and ion chromatography, and spectroscopic techniques like IR are also often used. However, modern fused-silica column capillary GC using very sensitive special detectors such as the ECD and the thermal energy analyzer enables nanogram to picogram sensitivity to be achieved for many explosive substances. Capillary GC-MS is the other very useful analytical combination for the unequivocal identification of explosives. Trinitrotoluene and dinitrotoluene isomers, cyclotrimethylene trinitramine, pentaerythritoltetranitrate, tetra, cyclotetramethylene tetranitramine, nitroglycerine, and ethylene glycol dinitrate have been detected successfully in this respect. Modern SPE is one of the proven techniques for efficient and selective sample clean-up and concentration. SPME was evaluated for the recovery of explosive residues from aqueous samples and real postexplosion solid debris samples, also as a field sampling device.

Pyrolysis GC (Py-GC) has been used for the identification and comparison of propellants. Smokeless powder propellants were analyzed by pyrolysis capillary GC-MS, and main organic propellant constituents together with additives could be detected and used as criteria for comparison. However, the standard technique for the examination of gunshot residues is scanning electron microscopy.

## Other Applications

GC in a forensic context has been used for the analysis of many other specimens besides the major groups reported above. The prerequisite is that the analytes have an organic chemical nature. Closely related to drug misuse is the abuse of organic solvents for sniffing among adolescents. Toluene, contained in paint thinner and glue for example, is the most commonly misused organic solvent. Other misused volatile substances include aerosols such as propane and butane and the group of halogenated hydrocarbons, alkyl nitrites, lighter fuels, and gasoline. Volatile solvents are central nervous system depressants and therefore dangerous in connection with road traffic for instance.

Generally, GC is a very suitable technique for the analysis of these volatile substances and the definitive proof of exposure to them is their detection in biological fluids and tissues. All the commonly abused solvents can be detected in the headspace from 200  $\mu$ l of blood. Toluene could be detected by MS in the breath of known glue sniffers up to 4 days after the last episode. GC and GC-MS were used to determine toluene in various tissues and blood in a

fatal case of oral ingestion of toluene. The toluene levels determined in postmortem blood, lung, liver, and brain from a toluene poisoning case that happened during painting were not high enough to be definitely lethal. But they were high enough to anesthetize the central nervous system. This is in agreement with general findings that death normally results from a dangerous situation existing while the person is intoxicated. Even cyclopropane gas taken from a cylinder of cyclopropane in a university laboratory was identified by a headspace GC-MS procedure as the abused substance causing the death of a student. Alkyl nitrites have been hydrolyzed with methanol to their respective alcohols, which can be identified easily by GC. Other toxic chemicals, such as carbon monoxide, hydrocyanic acid, chloroform, and tetrahydrofuran in blood and body material, have been identified and determined by headspace GC and GC-MS.

Nonvolatile organic substances such as polymers, car paints, hair, etc., have to be subjected to pyrolysis before GC separation of the products formed can be achieved. Normally, a suspect polymeric specimen is compared with a control sample in order to establish common origin. The high resolution and inertness of capillary columns offers enhanced discrimination and much greater column-to-column and instrument-to-instrument reproducibility. The combination of Py-GC with MS (Py-GC-MS) added the most valuable mass selective detection to the retention time parameter and brought a major step forward in generating objective data for sample comparisons. Databases for retrospective searching can be generated based on the criteria of certain characteristic masses and the retention time combined with peak area. Py-GC-MS is now routinely used for the examination of the organic components of paint, especially for automobile paint, plastics, and synthetic fibers. Other published applications of Py-GC and Py-GC-MS include the characterization of adhesives (e.g., as used in the construction of explosive devices), photocopier toner type, trace rubber residues (e.g., from shoe soles and tyres), plastic automobile bumper bars and light lenses, phenolic resins, polymer additives, polyurethane foams, tapes, and clingfilms. Fourier-transform IR spectroscopy is the other well-established technique for the analysis of these polymeric products.

The lachrymators 2-chloroacetophenone and *o*-chlorobenzylmalonitrile in tear gas, bank security devices, etc., have been identified by GC-MS. Capillary GC has been utilized to discriminate waxes and greases in cosmetic materials such as lipsticks submitted in forensic casework. For applications of this sort the recent improvements in high-temperature GC are very useful.

A dramatic decrease of the time of analysis can be achieved by using narrower columns combined with an adjusted improved electronic pressure control system. As today in most forensic science laboratories the procedures developed with the normal capillary GC instrumentation are part of a quality management system, a major change using the possibilities of 'Fast GC' requires a new validation. This, together with the high level of performance reached during the last 15 years in capillary GC, may to some extent slow down the process of innovation in this area.

*See also: Forensic Sciences:* Alcohol in Body Fluids; Explosives; Fibers; Gunshot Residues; Hair; Paints, Varishes, and Lacquers; Systematic Drug Identification; Thin-Layer Chromatography. **Fuels:** Oil-Based. **Gas Chromatography:** Pyrolysis; Detectors; Mass Spectrometry. **Polymers:** Natural Rubber; Synthetic; Polyurethanes.

## Further Reading

- Bertsch W and Ren Q (1999) Contemporary sample preparation methods for the detection of ignitable liquids in suspect arson cases. *Forensic Science Review* 11: 141–156.
- Beveridge AD (1992) Developments in the detection and identification of explosive residues. *Forensic Science Review* 4: 17–49.
- Blackledge RD (1992) Application of pyrolysis gas chromatography in forensic science. *Forensic Science Review* 4: 1–16.
- Bogusz MJ (ed.) (2000) *Handbook of Analytical Separations. Volume 2: Forensic Science*. New York: Elsevier.
- Furton KG, Wang J, Hsu YL, Walton J, and Almirall JR (2000) The use of solid-phase microextraction-gas chromatography in forensic analysis. *Journal of Chromatographic Science* 38: 297–306.
- Huizer H (1991) The use of gas chromatography for the detection and quantitation of abused drugs. In: Gough TA (ed.) *The Analysis of Drugs of Abuse*, pp. 23–92. Chichester: Wiley.
- Junting L, Peng C, and Suzuki O (1998) Solid-phase microextraction (SPME) of drugs and poisons from biological samples. *Forensic Science International* 97: 93–100.
- Lloyd JA and Edmiston PL (2003) Preferential extraction of hydrocarbons from fire debris samples by solid phase microextraction. *Journal of Forensic Sciences* 48: 130–134.
- Maurer HH (1992) Systematic toxicological analysis of drugs and their metabolites by gas chromatography-mass spectrometry. *Journal of Chromatography* 580: 3–41.
- Moeller MR (1992) Drug detection in hair by chromatographic procedures. *Journal of Chromatography* 580: 125–134.
- Moore JM (1990) The application of chemical derivatization in forensic drug chemistry for gas and high-performance liquid chromatographic methods of analysis. *Forensic Science Review* 2: 79–124.
- Palhol F, Boyer S, Naulet N, and Chabrilat M (2002) Impurity profiling of seized MDMA tablets by capillary gas chromatography. *Analytical and Bioanalytical Chemistry* 374: 274–281.
- Pfleger K, Maurer HH, and Weber A (2000) *Mass Spectral and GC Data of Drugs, Poisons, Pesticides, Pollutants and their Metabolites*, Part 4. Weinheim: Wiley-VCH.
- Strömberg L, Lundberg L, Neumann H, Bobon B, Huizer H, and Van der Stelt NW (2000) Heroin impurity profiling. A harmonization study for retrospective comparisons. *Forensic Science International* 114: 67–88.
- Tagliaro F, Lubli G, Ghielmi S, Franchi D, and Marigo M (1992) Chromatographic methods for blood alcohol determination. *Journal of Chromatography* 580: 161–190.

## Petrochemical Applications

**E R Adlard**, Burton, South Wirral, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1907–1913, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

Any account of gas chromatography (GC) should include the analysis of petroleum and petroleum products, since the petroleum industry has made a major contribution to the development of the technique. The two main reasons are that the mixtures encountered are frequently of complex composition and that hydrocarbon mixtures are extremely

difficult to separate by any other analytical technique. These statements are generalizations and there are numerous exceptions, but they do not alter the main premise. The mixtures encountered range from permanent gases and low-boiling-point hydrocarbons to intractable polymeric materials containing compounds with a relative molecular mass of well above 1000. In addition to the wide range in molecular mass, there is also a broad range of molecular types and sulfur, nitrogen and organometallic compounds are present in significant amounts in crude oils. Ethers and alcohols are present in unleaded gasoline; nitrogen, phosphorus, sulfur and halogen compounds are used as lubricating oil additives, and

A dramatic decrease of the time of analysis can be achieved by using narrower columns combined with an adjusted improved electronic pressure control system. As today in most forensic science laboratories the procedures developed with the normal capillary GC instrumentation are part of a quality management system, a major change using the possibilities of 'Fast GC' requires a new validation. This, together with the high level of performance reached during the last 15 years in capillary GC, may to some extent slow down the process of innovation in this area.

*See also:* **Forensic Sciences:** Alcohol in Body Fluids; Explosives; Fibers; Gunshot Residues; Hair; Paints, Varishes, and Lacquers; Systematic Drug Identification; Thin-Layer Chromatography. **Fuels:** Oil-Based. **Gas Chromatography:** Pyrolysis; Detectors; Mass Spectrometry. **Polymers:** Natural Rubber; Synthetic; Polyurethanes.

## Further Reading

- Bertsch W and Ren Q (1999) Contemporary sample preparation methods for the detection of ignitable liquids in suspect arson cases. *Forensic Science Review* 11: 141–156.
- Beveridge AD (1992) Developments in the detection and identification of explosive residues. *Forensic Science Review* 4: 17–49.
- Blackledge RD (1992) Application of pyrolysis gas chromatography in forensic science. *Forensic Science Review* 4: 1–16.
- Bogusz MJ (ed.) (2000) *Handbook of Analytical Separations. Volume 2: Forensic Science*. New York: Elsevier.
- Furton KG, Wang J, Hsu YL, Walton J, and Almirall JR (2000) The use of solid-phase microextraction-gas chromatography in forensic analysis. *Journal of Chromatographic Science* 38: 297–306.
- Huizer H (1991) The use of gas chromatography for the detection and quantitation of abused drugs. In: Gough TA (ed.) *The Analysis of Drugs of Abuse*, pp. 23–92. Chichester: Wiley.
- Junting L, Peng C, and Suzuki O (1998) Solid-phase microextraction (SPME) of drugs and poisons from biological samples. *Forensic Science International* 97: 93–100.
- Lloyd JA and Edmiston PL (2003) Preferential extraction of hydrocarbons from fire debris samples by solid phase microextraction. *Journal of Forensic Sciences* 48: 130–134.
- Maurer HH (1992) Systematic toxicological analysis of drugs and their metabolites by gas chromatography-mass spectrometry. *Journal of Chromatography* 580: 3–41.
- Moeller MR (1992) Drug detection in hair by chromatographic procedures. *Journal of Chromatography* 580: 125–134.
- Moore JM (1990) The application of chemical derivatization in forensic drug chemistry for gas and high-performance liquid chromatographic methods of analysis. *Forensic Science Review* 2: 79–124.
- Palhol F, Boyer S, Naulet N, and Chabrilat M (2002) Impurity profiling of seized MDMA tablets by capillary gas chromatography. *Analytical and Bioanalytical Chemistry* 374: 274–281.
- Pfleger K, Maurer HH, and Weber A (2000) *Mass Spectral and GC Data of Drugs, Poisons, Pesticides, Pollutants and their Metabolites*, Part 4. Weinheim: Wiley-VCH.
- Strömberg L, Lundberg L, Neumann H, Bobon B, Huizer H, and Van der Stelt NW (2000) Heroin impurity profiling. A harmonization study for retrospective comparisons. *Forensic Science International* 114: 67–88.
- Tagliaro F, Lubli G, Ghielmi S, Franchi D, and Marigo M (1992) Chromatographic methods for blood alcohol determination. *Journal of Chromatography* 580: 161–190.

## Petrochemical Applications

**E R Adlard**, Burton, South Wirral, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1907–1913, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

Any account of gas chromatography (GC) should include the analysis of petroleum and petroleum products, since the petroleum industry has made a major contribution to the development of the technique. The two main reasons are that the mixtures encountered are frequently of complex composition and that hydrocarbon mixtures are extremely

difficult to separate by any other analytical technique. These statements are generalizations and there are numerous exceptions, but they do not alter the main premise. The mixtures encountered range from permanent gases and low-boiling-point hydrocarbons to intractable polymeric materials containing compounds with a relative molecular mass of well above 1000. In addition to the wide range in molecular mass, there is also a broad range of molecular types and sulfur, nitrogen and organometallic compounds are present in significant amounts in crude oils. Ethers and alcohols are present in unleaded gasoline; nitrogen, phosphorus, sulfur and halogen compounds are used as lubricating oil additives, and

C<sub>16</sub>–C<sub>20</sub> fatty acids are essential components of greases. The topics discussed below have been chosen as illustrations and the article cannot be considered to be comprehensive.

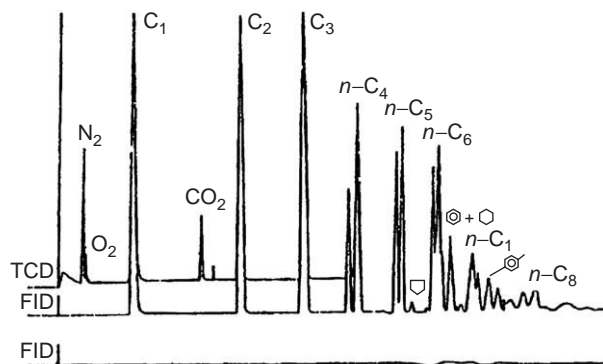
## Gases and Light Hydrocarbons

The analysis of mixtures of nitrogen, oxygen, carbon dioxide, and hydrocarbon gases up to C<sub>8</sub> is one that the industry is required to carry out for fiscal reasons, which means that the pertinent Institute of Petroleum method (IP 345/80) has been the subject of stringent trials. This type of sample may seem, at first sight, to be a relatively simple mixture to analyze, until it is examined more closely. The first problem is that the separation of oxygen and nitrogen often requires operation of the column at subambient temperatures. The analysis is specifically concerned with so-called 'associated gas', i.e., the gas in equilibrium with crude oil at around atmospheric pressure. Under these conditions the concentrations of the C<sub>6</sub>–C<sub>8</sub> hydrocarbons are below the lower limit of detection of the thermal conductivity detector (TCD) and it is necessary to employ a TCD followed by a flame ionization detector (FID). The TCD detects the permanent gases and the C<sub>1</sub>–C<sub>4</sub> hydrocarbons and the FID detects all the hydrocarbons up to C<sub>8</sub> but not the permanent gases. The two sets of signals are linked through the ethane peak, which gives a large response with both detectors (Figure 1).

Apart from the use of two detectors in series, this analysis is unusual in that it is effected by gas–solid chromatography (GSC) on a porous polymer packing and the column is temperature programmed from –50°C to +240°C at 15°C min<sup>–1</sup> with an initial isothermal period of 2 min and a final isothermal

period of 10 min. The analysis is of the headspace gases above the crude oil and, therefore, the sample does not contain high-boiling-point compounds. On account of this, and the fact that the separation is by GSC rather than gas–liquid chromatography, the method is extremely robust and can be carried out in very isolated locations provided that liquid nitrogen is available for the subambient starting temperature that is required for the separation of nitrogen from oxygen. There is no oxygen in the crude oil and the separation is required only if there has been an ingress of air during sampling. If this possibility can be eliminated, the temperature programme can be carried out from 20°C, which can be readily attained with carbon dioxide coolant rather than liquid nitrogen. The analysis time can also be shortened from about 32 min to 25 min, with a bigger saving in the turn-round time (the time between each analysis).

The alternative to this method is to use column switching so that the permanent gases pass through a column containing a polymer bead stationary phase that does not separate them and then on to a column containing a molecular sieve that does. After the permanent gases and methane and ethane have left the first column, the exit of this column is switched to an FID to monitor the separation of the hydrocarbons. The advantages of this approach are that subambient programming is not required and that the analysis turn-round time can be substantially reduced. The disadvantage is that it depends on the reliability of the switching valves over many operating cycles, so that it is potentially less robust than the first method. Under normal laboratory conditions, column switching is a well-known and commonly used technique. Another example of column switching is described more fully later.



**Figure 1** Typical chromatogram of associated natural gas using thermal conductivity (TCD; upper curve) and flame ionization (FID; lower curve) detection. (From IP 345/80, courtesy of the Institute of Petroleum.)

## Gasoline Analysis

There are few analyses more impressive than that of a finished gasoline by capillary GC. This may be carried out on a stainless steel column 100 m long by 0.25 mm internal diameter coated with squalane and temperature-programmed from 0°C to 90°C at 5°C min<sup>–1</sup>. Pressure programming with helium carrier gas is used from an inlet pressure of 5 psi to 40 psi since squalane is too volatile to be taken much above 100°C. Analysis under these conditions is capable of yielding over 280 peaks in about 100 min with a turn-round time of 120 min. This means that only about six samples can be analyzed in duplicate every 24 h, which in turn means that automation of the analysis is necessary.



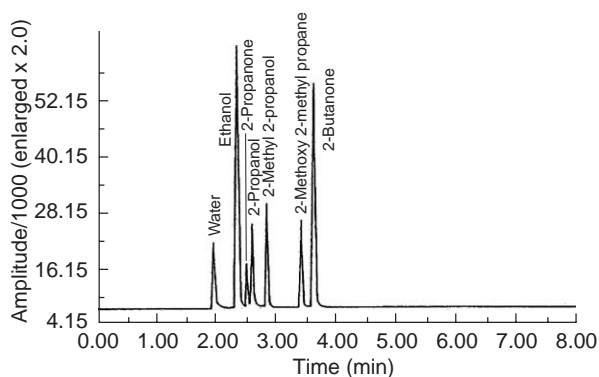
The biggest handicap in increasing the speed of the analysis is the limited thermal stability of squalane. Silicone oils can be used to much higher temperatures, but so far no other stationary phase has been found to be as good as squalane from the point of view of resolution. Another reason for using squalane is that hydrocarbons emerge from such a column mainly in the order of their boiling points, although the effect of molecular shape can be observed in the elution order of multibranched and cyclic molecules as compared with straight-chain molecules. This effect can give a reversal of elution order of up to 5°C but 1–2°C is more common. There is also a large amount of information in the literature on the identification of the emerging peaks by gas chromatography–mass spectrometry (GC–MS).

Hydrogen in place of helium as carrier gas effects a substantial reduction in analysis time, and shortening the column also shortens the time, although there is, naturally, some loss in separation. However, by these means the turn-round time can be reduced to less than an hour, which is more acceptable.

The reason for the considerable effort expended by the petroleum industry on gasoline analysis is the possibility of being able to calculate octane ratings from the full chemical composition determined by GC instead of the standard engine test. By the mid-1970s most of the problems of this approach had been solved except the intractable one of the large effect on octane rating of small concentrations of lead alkyls. The widespread and increasing use of lead-free gasolines, which have their octane rating increased by the addition of relatively large amounts (percentage quantities) of alcohols and ethers, has raised the possibility that this approach is now more likely to be successful.

## Oxygenates in Gasoline

As indicated above, lead-free gasolines currently contain substantial amounts of compounds such as methyl *t*-butyl ether (MTBE) as octane rating improvers, and various methods for the analysis of these compounds in a complex hydrocarbon matrix have been proposed. The method currently favored is the use of an oxygen-specific detector. This is a modified FID first suggested by Schneider and colleagues in 1982. The effluent from the capillary column in the carrier gas (nitrogen with a small amount of hydrogen) passes into a small capillary coil of platinum/iridium electrically heated to 1150°C. Under these conditions, hydrocarbons are cracked, i.e., completely broken down to carbon and

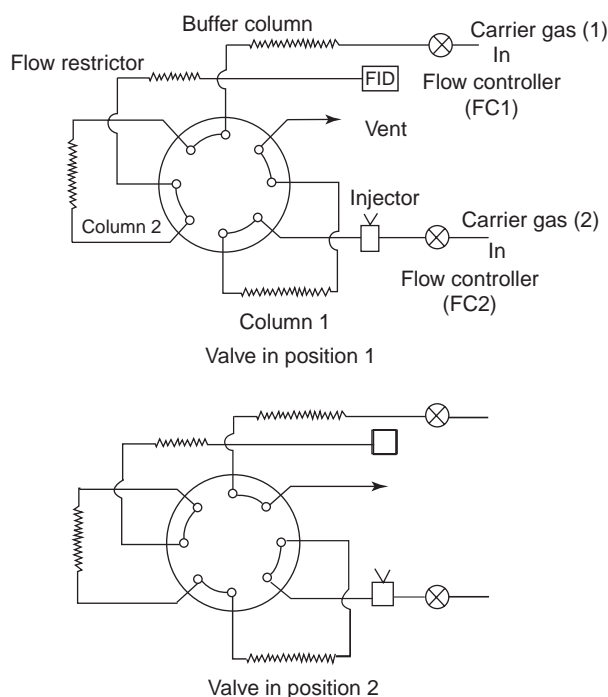


**Figure 2** Oxygenated compounds in an experimental gasoline. O-FID chromatogram. Analysis performed using a 15 m × 0.25 mm i.d. fused-silica capillary column coated with 0.25 μm DB-1, temperature programmed from 40°C to 50°C at 2°C min<sup>-1</sup> then 50°C to 290°C at 30°C min<sup>-1</sup>. (Courtesy of Shell Research Ltd.)

hydrogen. Compounds containing oxygen give carbon monoxide under these conditions. The carbon monoxide passes on to a methanizer, which is a small glass capillary coated with aluminum oxide mixed with nickel catalyst and heated to about 350°C. The methanizer is situated within the jet of the flame detector and the normal hydrogen supply for the flame also serves for the hydrogenation of the carbon monoxide to methane. In spite of this apparently rather complicated approach, the detector (now commercially available and known as the O-FID) works remarkably well with little loss of resolution due to the extra dead volume introduced (Figure 2). The main disadvantage is the relatively poor limit of detection of about 0.1% for oxygenated compounds, since each oxygen atom is equivalent to one molecule of methane; methane has a relatively low sensitivity in the FID, which is a mass detector. Under favorable conditions such as the analysis of small amounts of oxygenated compounds in an essentially pure solvent, this limit may be lowered by one or two orders of magnitude if solvent venting is employed (*see Gas chromatography, specialized detectors*).

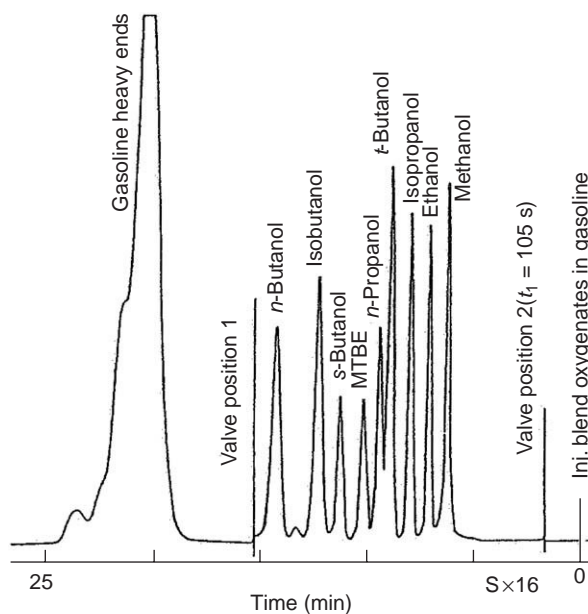
This section has been included because it is a good example of the use of selective detectors in petroleum analysis, where the need often arises to analyze for small amounts of heterocompounds in a complex hydrocarbon matrix. Another good example is the use of the flame photometric detector for the determination of small concentrations of sulfur compounds in various plant streams. The analysis of oxygenated compounds can, however, be carried out in a totally different manner using column switching instead of the selective detector.





**Figure 3** Column-switching arrangement for the determination of oxygenated compounds in gasoline. (From IP AX/87, courtesy of the Institute of Petroleum.)

In this method (IP Proposed Method AX/87) the sample is injected, with a suitable internal standard, on to a polar polyester packed column that allows the lower-boiling-point hydrocarbons to elute rapidly. After hydrocarbons up to *n*-heptane have emerged from this first column, the flow through the column is reversed and the effluent is taken to a second column containing Porapak P (a porous polymer bead material), which separates the oxygenates from one another. The oxygenates emerge from this second column before the higher-boiling-point hydrocarbons, which remained in the sample after passage through the first column. To speed up the analysis, the system is then switched back to the original configuration so that the higher boiling hydrocarbons are eluted as an unresolved composite peak. **Figure 3** shows the flow through the system with the 8-port rotary valve in the two positions, and **Figure 4** shows a chromatogram of the oxygenates in a gasoline sample. The buffer column in **Figure 3** is identical to column 1 and is included to minimize flow disturbances on switching the valve. Both columns and switching valve are installed in an oven operating at 150°C. The timing for the valve switching must be determined by the use of calibration blends of *n*-heptane and butanol in *n*-pentane. As in the column-switching method described briefly in the section on gas analysis, these



**Figure 4** Chromatogram obtained with the column-switching system shown in **Figure 3**. (From IP AX/87, courtesy of the Institute of Petroleum.)

are advantages and disadvantages in the column-switching technique. The advantages in this case are the use of packed columns, as opposed to the capillary columns required with the O-FID, which makes quantification simpler and gives a better lower limit of detection. The disadvantages are that the apparatus is complex to set up, and having been set up is completely dedicated to this particular determination. The demands on the switching valve are considerable and operating conditions (column temperature and carrier gas flow rate) must be carefully controlled since otherwise the valve switching times will alter.

## Dilution of Engine Lubricants by Fuels

In petrol and diesel engines for cars, buses, and trucks, the lubricating oil inevitably becomes contaminated to a greater or lesser extent with the engine fuel. The concentration of the fuel in the oil may range from less than 0.1% in a well-maintained engine to greater than 10% in a badly maintained one. Clearly, by the time this sort of concentration has been reached, the lubricant is not performing as it should. By analyzing a sample of the used oil on a short silicone packed column, a bimodal distribution chromatogram is obtained – the first part from the fuel residues and the second part from the lubricating oil. If the used oil is from a gasoline engine, it is very easy to differentiate between the two contributions

to the chromatogram and to quantify the individual contributions. However, if the used oil is from a diesel engine this is much more difficult because the high-boiling-point residues of the diesel fuel overlap the low-boiling-point components of the lubricant. This problem can be solved by running a sample of uncontaminated oil and a sample of 'topped' fuel, that is fuel that has had the low-boiling-point components distilled off. These chromatograms are stored in a computer, which compares them with the chromatogram of the used oil and decides on the allocation of the respective areas due to the fuel and the lubricant. This computer technique has been found to give significantly more accurate results than the arbitrary choice of a split point between fuel and lubricant.

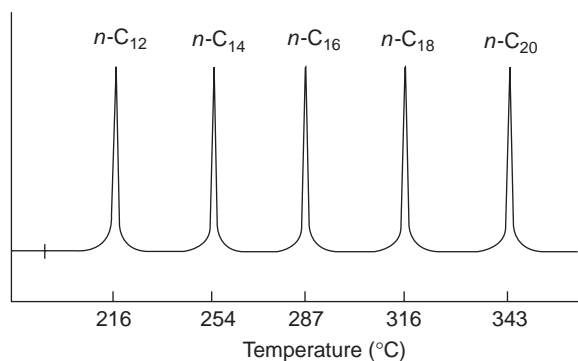
## Crude Oil Analysis

Crude oil analysis is difficult: (1) because of the very wide range of molecular masses involved and (2) because it is impossible to obtain a complete recovery of the high-boiling-point components of the oil. A light North Sea crude will give a recovery of about 85% up to  $C_{42}$ , but this figure could be as low as 50% for a heavy Middle East crude. This problem is tackled by the addition of known amounts of markers to the crude oil sample. If, say, 2% (m/m) of a pure compound that is of low enough boiling point to ensure its complete recovery from the chromatographic analysis is added to the sample, then, from the areas of the peaks obtained, the overall recovery can be calculated. This is standard practice, but in the case of crude oils there are no convenient gaps in the chromatogram where readily available marker compounds will emerge. The solution to the problem is to use known amounts of several (commonly about three) normal paraffins, say 2% each of  $n\text{-}C_{12}$ ,  $n\text{-}C_{14}$ , and  $n\text{-}C_{16}$ . These are present in all crude oils, but they may be used as markers to calculate recovery by measuring the areas due to the known amounts added by subtracting the relevant areas obtained from a sample of the original crude oil. Another convenient well-resolved peak is used as a marker to correct for variations in sample size. Once the recovery has been measured, the analytical results can be expressed in a variety of ways – the amount of an individual compound such as benzene, the amounts of various carbon number groups, e.g.,  $C_5$ ,  $C_6$ ,  $C_7$  present, or as a simulated distillation curve (see below). The average molecular mass of the oil can also be determined as accurately as by other methods and with considerably less effort.

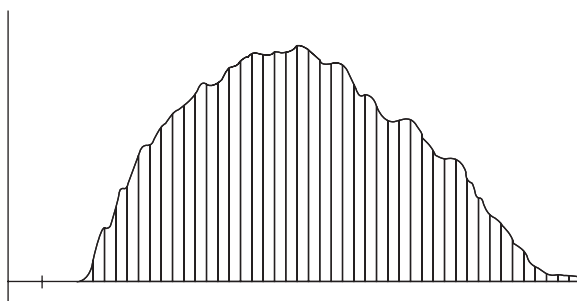
## Simulated Distillation

This technique, which is almost exclusive to the petroleum industry, depends on the fact already stated that, with a nonpolar stationary phase, hydrocarbons emerge more or less in the order of their boiling points. By calibration of the GC with compounds of known boiling point (normal paraffins), the usual time axis of a chromatogram may be converted into a boiling point axis. When an unknown sample is analyzed under identical conditions to those for the  $n$ -paraffin calibration mixture (same column, same carrier gas flow rate, same temperature programme), increments of area at regular short time intervals are measured and plotted against boiling point, i.e., retention time (see Figures 5–7 for a pictorial explanation of the technique).

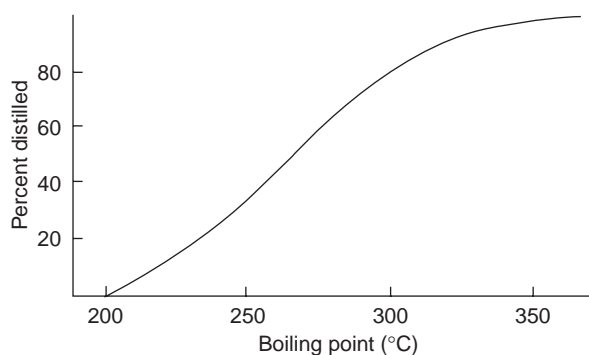
With modern data-handling systems, the time slices may be as small as 0.1 s but 0.5 s is adequate for packed-column work. The plot of cumulative area vs temperature is a simulated distillation graph. This method has a number of advantages over a true analytical distillation in that it is quicker, requires much less sample, and gives much more accurate values for the initial and final boiling points. In order to minimize any molecular-type effects on retention, such as an aromatic emerging earlier than a paraffin of the



**Figure 5** Temperature scale calibration of the gas chromatogram.



**Figure 6** Chromatogram of a light oil split into equal time slices.



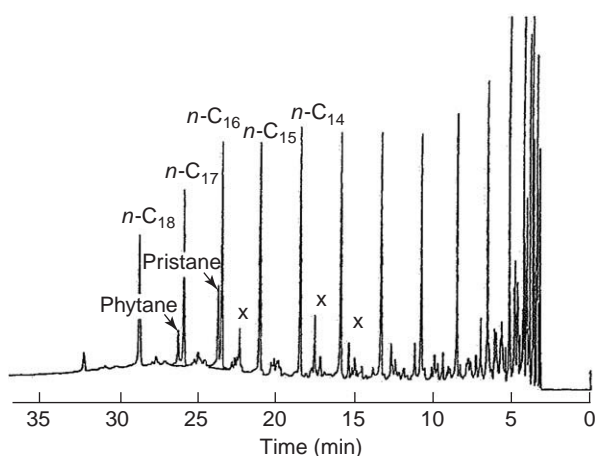
**Figure 7** Simulated distillation curve for the oil.

same boiling point, this method is unique in specifying a column of low efficiency that must not exceed a certain value.

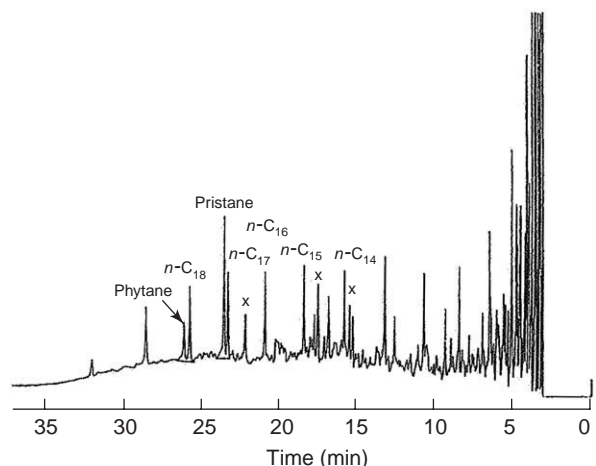
## Oil Spill Identification

There are, fortunately, very few incidents where there is a major spill of crude oil, although lesser incidents involving 1–100 ton quantities are more common. Sometimes the source of the spill is beyond any reasonable doubt, but more often proof of the origin is required for the purpose of compensation claims. There are numerous analytical methods for the identification of crude oils and heavy fuel oils, but in practice they all have limitations. Tentative identification may be made by GC on a low-resolution packed column to give a fingerprint of the oil and high-resolution GC with a capillary column on the material boiling below 343°C. The high-resolution method allows the determination of the ratio of pristane to phytane, which is characteristic for many oils (Figures 8 and 9). Further tentative identification may be made by employing a flame photometric sulfur detector to give a sulfur fingerprint of the oil. This method is especially useful for biodegraded samples, which may give an unrecognizable FID hydrocarbon fingerprint but still give a characteristic sulfur fingerprint since the sulfur-containing compounds in the oil are more resistant to biodegradation than many of the hydrocarbons. In its ultimate form this type of identification is carried out by GC–high-resolution mass spectrometry, when it is possible to identify crude oils from closely related sources by measuring the ratios of individual benzothiophenes and dibenzothiophenes.

It should be noted that all the methods identify the original geographical and geological source of the oil and not necessarily the immediate source such as a particular ship or storage tank. This requires confirmatory circumstantial evidence and the availability of a valid reference sample.



**Figure 8** Chromatogram of <343°C fraction of Sarir crude oil. (From IP 318/75, courtesy of the Institute of Petroleum.)



**Figure 9** Chromatogram of <343°C fraction of Light Nigerian crude oil. (From IP 318/75, courtesy of the Institute of Petroleum.)

Another, as yet unsolved, problem is how to pinpoint the time of a particular spill, since the effects of weathering are extremely variable and depend on the microclimatic conditions to which a sample has been subjected. The use of two radioactive isotopes of iodine in known ratios to label cargoes has been proposed to resolve this problem, but this solution, although elegant from a purely scientific point of view, is as bad, environmentally, as the original problem and the method has never been adopted on a large scale.

## Plant Control

GC can be used to give automatic feedback control of plant and operating systems such as pipelines but is less frequently used on a direct basis than might be expected. Most plant control by GC is effected by

manually transporting a sample from the plant to a standard laboratory instrument housed in a safe area within the refinery. The main reason is that it takes a considerable time for any altered conditions such as temperature to have an effect on large plants. The second reason, which follows from the first, is that if rapid analysis near the plant is unnecessary then there is little point in using a plant instrument and exposing it to a hostile environment. Another factor is the difficulty of sampling on a continuous basis directly from a plant stream. It is extremely difficult to ensure a completely clean sample and the presence of a small rust particle is sufficient to destroy the highly polished surfaces of liquid sampling valves. Fluid logic devices have been employed experimentally to overcome this problem but they do not seem to have achieved wide acceptance.

## Other Techniques

The methods outlined above are just a few of the analyses carried out by GC in the petroleum industry. No examples of the use of headspace GC, or GC coupled to other chromatographic methods such as thin-layer chromatography or LC, have been given. GC coupled to mass spectrometry, which has been briefly mentioned, is probably the most important technique

available today for the qualitative analysis of petroleum mixtures and has been applied to topics ranging from organic geochemistry to occupational hygiene.

**See also:** **Fuels:** Gaseous; Oil-Based. **Gas Chromatography:** Mass Spectrometry. **Process Analysis:** Chromatography.

## Further Reading

- Adlard EE (ed.) (1995) *Chromatography in the Petroleum Industry*. Amsterdam: Elsevier.
- Altgelt KH and Gouw TH (1979) *Chromatography in Petroleum Analysis*. New York: Dekker.
- ASTM International (2002) *Annual Book of ASTM Standards*. Conshohocken: ASTM International.
- Institute of Petroleum. *Standard Methods for Analysis and Testing of Petroleum and Related Products*. Published annually by J Wiley on behalf of the Institute of Petroleum, London. Method IP 345/80: Determination of composition of associated natural gas – gas chromatography method. Method AX/87: Determination of oxygenates in gasoline – gas chromatography method.
- Pauls RE (1995) Chromatographic characterization of gasolines. *Advances in Chromatography* 35: 259–342.
- Wang ZD and Fingas M (1997) Developments in the analysis of petroleum products and oil-spill-related environmental samples by gas chromatography. *Journal of Chromatography A* 774: 51–78.

## Chiral Separations

**V Schurig**, Institute für Organische Chemie der Universität, Tübingen, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 1851–1860, © 1995, Elsevier Ltd., with revisions made by the Editor.

The unambiguous determination of enantiomeric compositions and absolute configurations is an important analytical task in the synthesis, characterization, and use of nonracemic compounds (optical isomers, enantiomers) such as chiral research chemicals, auxiliaries, drugs, pesticides, herbicides, pheromones, flavors, and fragrance. As the insight into chirality–activity relationships steadily improve and, as a consequence, legislation concerning chiral compounds becomes more and more stringent, the development of precise methods for the determination of enantiomeric purities up to, and in some cases higher than, ee = 99% is of great importance (ee, enantiomeric excess, is the percentage by which one enantiomer, *R*, is in excess in the (racemic) mixture of *R* and *S*: %ee = 100(*R* – *S*)/(*R* + *S*)).

The difficulty in determining ee arises from the fact that enantiomers have (apart from their chiroptical properties) *identical* properties in an *achiral* environment (assuming ideal conditions). Methods of distinguishing enantiomers must, therefore, rely on their chiroptical properties (polarimetry, circular dichroism, optical rotary dispersion) or must employ a *chiral* auxiliary in the spirit of Pasteur's resolution principles via diastereomer formation or interaction (e.g., in nuclear magnetic resonance (NMR) spectroscopy or chromatography). Because of high efficiency, sensitivity, and speed, chiral separation by gas chromatography represents a highly versatile technique for enantiomer analysis in the entire range of enantiomeric excess. Yet the prerequisite of the method is the volatility, thermal stability, and resolvability of the chiral analyte, which restricts its general use.

The separation of enantiomers by gas chromatography (GC) can be performed in two modes.

(a) *Indirect method*. Off-column conversion of enantiomers into diastereomeric derivatives by chemical



manually transporting a sample from the plant to a standard laboratory instrument housed in a safe area within the refinery. The main reason is that it takes a considerable time for any altered conditions such as temperature to have an effect on large plants. The second reason, which follows from the first, is that if rapid analysis near the plant is unnecessary then there is little point in using a plant instrument and exposing it to a hostile environment. Another factor is the difficulty of sampling on a continuous basis directly from a plant stream. It is extremely difficult to ensure a completely clean sample and the presence of a small rust particle is sufficient to destroy the highly polished surfaces of liquid sampling valves. Fluid logic devices have been employed experimentally to overcome this problem but they do not seem to have achieved wide acceptance.

## Other Techniques

The methods outlined above are just a few of the analyses carried out by GC in the petroleum industry. No examples of the use of headspace GC, or GC coupled to other chromatographic methods such as thin-layer chromatography or LC, have been given. GC coupled to mass spectrometry, which has been briefly mentioned, is probably the most important technique

available today for the qualitative analysis of petroleum mixtures and has been applied to topics ranging from organic geochemistry to occupational hygiene.

**See also:** **Fuels:** Gaseous; Oil-Based. **Gas Chromatography:** Mass Spectrometry. **Process Analysis:** Chromatography.

## Further Reading

- Adlard EE (ed.) (1995) *Chromatography in the Petroleum Industry*. Amsterdam: Elsevier.
- Altgelt KH and Gouw TH (1979) *Chromatography in Petroleum Analysis*. New York: Dekker.
- ASTM International (2002) *Annual Book of ASTM Standards*. Conshohocken: ASTM International.
- Institute of Petroleum. *Standard Methods for Analysis and Testing of Petroleum and Related Products*. Published annually by J Wiley on behalf of the Institute of Petroleum, London. Method IP 345/80: Determination of composition of associated natural gas – gas chromatography method. Method AX/87: Determination of oxygenates in gasoline – gas chromatography method.
- Pauls RE (1995) Chromatographic characterization of gasolines. *Advances in Chromatography* 35: 259–342.
- Wang ZD and Fingas M (1997) Developments in the analysis of petroleum products and oil-spill-related environmental samples by gas chromatography. *Journal of Chromatography A* 774: 51–78.

## Chiral Separations

**V Schurig**, Institute für Organische Chemie der Universität, Tübingen, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 1851–1860, © 1995, Elsevier Ltd., with revisions made by the Editor.

The unambiguous determination of enantiomeric compositions and absolute configurations is an important analytical task in the synthesis, characterization, and use of nonracemic compounds (optical isomers, enantiomers) such as chiral research chemicals, auxiliaries, drugs, pesticides, herbicides, pheromones, flavors, and fragrance. As the insight into chirality–activity relationships steadily improve and, as a consequence, legislation concerning chiral compounds becomes more and more stringent, the development of precise methods for the determination of enantiomeric purities up to, and in some cases higher than, ee = 99% is of great importance (ee, enantiomeric excess, is the percentage by which one enantiomer, *R*, is in excess in the (racemic) mixture of *R* and *S*: %ee = 100(*R* – *S*)/(*R* + *S*)).

The difficulty in determining ee arises from the fact that enantiomers have (apart from their chiroptical properties) *identical* properties in an *achiral* environment (assuming ideal conditions). Methods of distinguishing enantiomers must, therefore, rely on their chiroptical properties (polarimetry, circular dichroism, optical rotary dispersion) or must employ a *chiral* auxiliary in the spirit of Pasteur's resolution principles via diastereomer formation or interaction (e.g., in nuclear magnetic resonance (NMR) spectroscopy or chromatography). Because of high efficiency, sensitivity, and speed, chiral separation by gas chromatography represents a highly versatile technique for enantiomer analysis in the entire range of enantiomeric excess. Yet the prerequisite of the method is the volatility, thermal stability, and resolvability of the chiral analyte, which restricts its general use.

The separation of enantiomers by gas chromatography (GC) can be performed in two modes.

(a) *Indirect method*. Off-column conversion of enantiomers into diastereomeric derivatives by chemical



reaction with an *enantiomerically pure* chiral auxiliary and subsequent gas chromatographic separation of the diastereomers on a conventional *achiral* stationary phase.

(b) *Direct method*. Gas chromatographic separation of the enantiomers on a *chiral* stationary phase containing an auxiliary resolving agent of high (but not necessarily complete) enantiomeric purity.

While method (a) involves the formation of *diastereomers* before separation, method (b) involves the rapid and reversible *diastereomeric interaction* between the nonracemic stationary phase (selector) and the racemic (or enriched) analyte (selectand). The conversion of enantiomers into diastereomers and the subsequent chromatographic analysis of the diastereomeric ratio by method (a) requires the following conditions:

- The analyte must contain a suitable reactive group (preferably only one functionality in close proximity to the chiral center) undergoing a quantitative transformation with the chiral auxiliary compound with little formation of side-products.
- No kinetic resolution must result from the transformation of the mixture of enantiomers into diastereomers. The reaction should ideally therefore go to completion.
- There must be no racemization of either the analyte or of the auxiliary compound in the reaction that transforms the mixture of enantiomers into diastereomers.
- The auxiliary reagent must be enantiomerically pure for the sake of convenient and accurate analysis. Otherwise, an error is introduced in the determination of enantiomeric excess which, however, can be corrected for to a first approximation if the ee of the auxiliary is known.
- There must be no accidental change in the ratio of the diastereomers formed during sample manipulations. Since diastereomers display different physical properties, fractionation by incomplete recovery, decomposition, and losses may occur during work-up, isolation, and sample handling. In chromatography, fractionation may occur in split injection techniques and the original enantiomeric composition may be seriously falsified by differing detector responses to diastereomers.

Consequently, method (b) is preferred whenever possible for the determination of enantiomeric excess because it lacks the above drawbacks. This approach requires an efficient selector–selectand system displaying chiral recognition. Fortunately, with

high-resolution capillary columns, efficiency is high enough to resolve racemates having a difference of free enthalpy (Gibbs free energy) of diastereomeric association as little as  $0.08 \text{ kJ mol}^{-1}$ . With method (b) no fractionation of the enantiomers is expected at any time during the analytical procedure (cf. sources of error, below) and enantiomeric excess determined for the analyte by method (b) is independent of the enantiomeric purity of the chiral stationary phase employed. A low enantiomeric purity of the chiral auxiliary, however, results in small separation factors  $\alpha = k_R/k_S$  ( $k$  = retention factor, formerly capacity factor). The success of a chiral separation can be verified by control experiments. Unequivocal criteria of enantiomer separation are the occurrence of *peak coalescence (first kind)* when a racemic stationary phase is used and of *peak inversion (first kind)* of clearly resolved peaks of an enriched analyte when stationary phases of opposite configuration are applied. Method (b) is especially useful for determination of enantiomeric excess when no sample derivatization is required, e.g. in headspace analysis where the volatile enantiomers can be analysed directly, in the absence of enantioselective associations in the liquid phase, with minute amounts of gas. Owing to the enormous separating power of gas chromatography, notably in the high-resolution capillary column mode of operation, contaminants and impurities are separated from the sample and the simultaneous analysis of the enantiomers of different compounds (e.g., all amino acids) is feasible in one analytical run. Moreover, established ancillary techniques such as multidimensional chromatography (multiple-column operation), use of interfacing and coupling methods (GC–mass spectrometry (MS)) can also readily be adapted in chiral separations. Detection limits can be extended down to the picogram level by GC–MS or GC electron capture detection. GC–MS selected-ion monitoring (SIM) can detect enantiomers in complex matrices.

In gas chromatographic chiral separation, useful information may be obtained from the peak parameters depicted in Table 1. Enantioselectivity  $-\Delta_{R,S}(\Delta G^\circ)$  (cf. (2) in Table 1) arises from the diastereomeric interaction between the enantiomers of the selectand and the selector. Assuming a fast and reversible equilibration, enantiomer separation by gas chromatography is governed by thermodynamics according to the Gibbs–Helmholtz equation (the subscripts refer to enantiomers  $R$  and  $S$ ):

$$\begin{aligned} RT \ln(K_R/K_S) &= -\Delta_{R,S}(\Delta G^\circ) \\ &= -\Delta_{R,S}(\Delta H^\circ) + T\Delta_{R,S}(\Delta S^\circ) \quad [1] \end{aligned}$$

**Table 1** Enantiomeric peak parameters and derived properties

Peak parameter	Derived property	Expression
Peak retention	Chemoselectivity <sup>a</sup>	$K, -\Delta G^\circ$
Peak separation	Enantioselectivity	$-\Delta_{R,S}(\Delta G^\circ) = RT \ln \alpha^b$
Peak coalescence (third kind)	Enantiomerization barrier	$\Delta G^\ddagger$
Peak ratio	Enantiomer ratio	ee
Peak assignment	Enantiomer configuration	$R$ or $S$

<sup>a</sup> Involving dilute metal coordination compounds permitting the measurement of association constants from retention data.

<sup>b</sup> If the stationary phase is a binary system with an achiral solvent the chiral separation factor  $\alpha = k_R/k_S$  ( $k$ =retention factor) also includes retention on the achiral component.

From Schurig V and Bürkle W (1982) Extending the scope of enantiomer resolution by complexation gas chromatography. *Journal of the American Chemical Society* 104: 7573–7580.

For an association or coordination process, the quantities  $\Delta_{R,S}(\Delta S^\circ)$  and  $-\Delta_{R,S}(\Delta H^\circ)$  usually have opposite signs, thus having opposing effects in determining  $-\Delta_{R,S}(\Delta G^\circ)$ . When  $\Delta_{R,S}(\Delta G^\circ) = 0$  ( $K_R = K_S$ ; no enantiomer separation), the *isoenantioselective* temperature  $T_{\text{isoenant}}$

$$T_{\text{isoenant}} = \Delta_{R,S}(\Delta H^\circ) / \Delta_{R,S}(\Delta S^\circ) \quad [2]$$

is reached, where *peak coalescence (second kind)* occurs. At  $T_{\text{isoenant}}$  the sign of enantioselectivity changes (*peak inversion (second kind)*). Below the coalescence temperature, the sign of enantioselectivity  $\Delta_{R,S}(\Delta G^\circ)$  is governed by  $-\Delta_{R,S}(\Delta H^\circ)$ , and above it by  $\Delta_{R,S}(\Delta S^\circ)$ . Evidence for the reversal of enantioselectivity has indeed been found in gas chromatographic enantiomer separation. In practice, gas chromatography can rarely be carried out at or above  $T_{\text{isoenant}}$ ; hence, in general, enantioselectivity is enthalpy-controlled and separation factors  $\alpha$  increase with decreasing temperature.

The configurational integrity of the enantiomers during the gas chromatographic process of separation is essential to the correct enantiomer analysis. When enantiomers are labile to inversion of configuration, characteristic peak profiles are obtained that are recognized by the appearance of a plateau between the terminal peaks of the enantiomers. Using peak form analysis, activation parameters of enantiomerization ( $\Delta G^\ddagger$ ) (cf. (3) in Table 1) can be determined by *dynamic* gas chromatography. If enantiomerization is fast within the chromatographic timescale, *peak coalescence (third kind)* will arise.

Because an (achiral) detection device will respond equally to enantiomers irrespective of their molecular configuration, the comparison of relative peak areas provides an unambiguous measure of the enantiomeric ratio from which enantiomeric excess can be

**Table 2** The development of chiral stationary phases in gas chromatography

Amino acid derivatives	Metal chelates	Carbohydrates
D or L	$R$ or $S$	D only
Diamide-phase	Complexation GC	Cyclodextrins
Hydrogen-bonding	Coordination	e.g., Inclusion
1966 Gil-Av, Charles, Feibush	1977 Schurig, Gil-Av	Koscielski <i>et al.</i> , 1983
		Juvancz <i>et al.</i> , 1987
		Schurig <i>et al.</i> , 1988
		König <i>et al.</i> , 1988
		Armstrong <i>et al.</i> , 1989
Chirasil-Val	Chirasil-Metal	Chirasil-Dex
Frank, Nicholson, Bayer, 1977	Schurig <i>et al.</i> , 1991	Schurig <i>et al.</i> , 1990

See Schurig V (1994) Enantiomer separation by gas chromatography on chiral stationary phases. *Journal of Chromatography* 666: 111–129.

calculated, provided the detector response is strictly linear.

The assignment of absolute configurations is another important application in enantiomer analysis. Unequivocal proof of absolute configuration can be obtained by coinjection of a sample with established stereochemistry. The correlation of absolute configuration with the elution order of structurally related enantiomers, e.g., members of homologous series of compounds, may lead to serious pitfalls, since many causes of peak inversions are feasible.

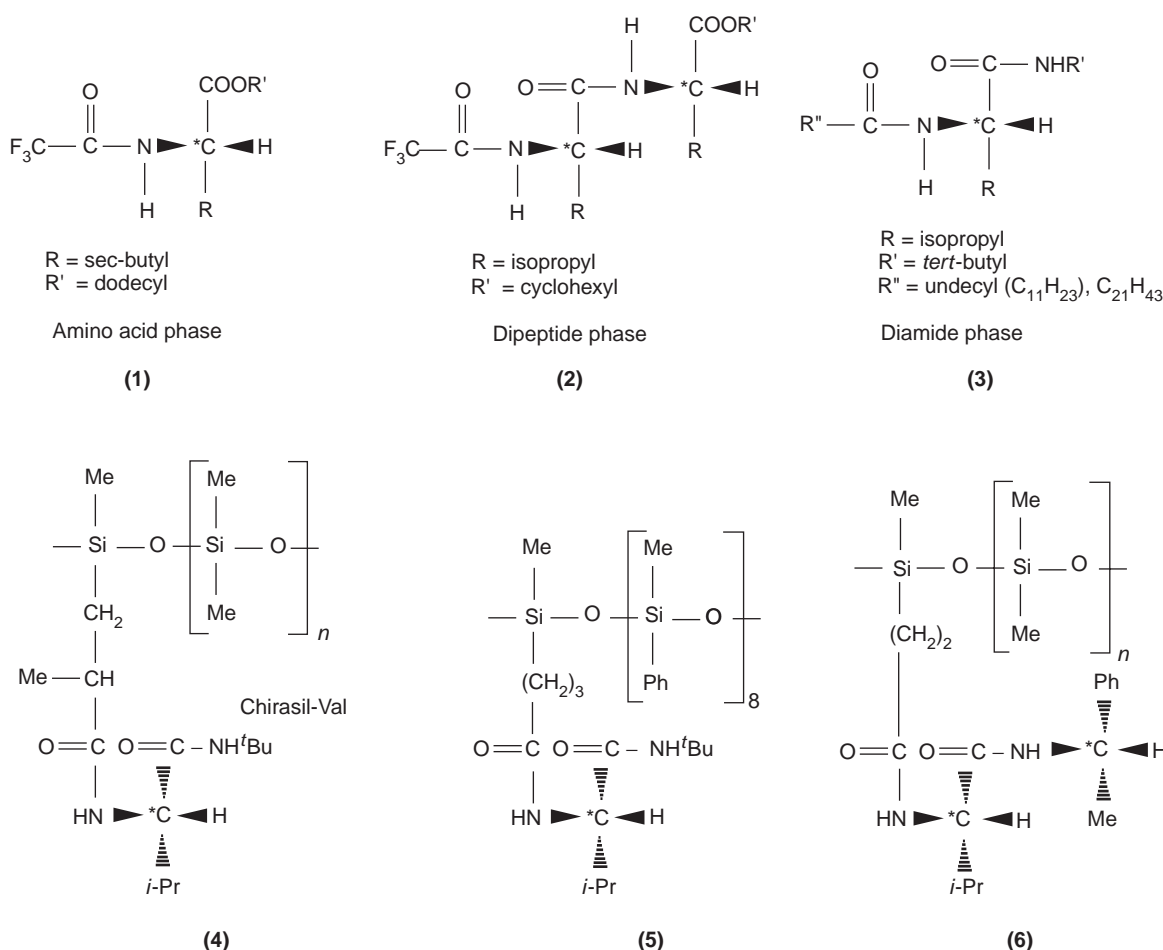
## Classification of Chiral Stationary Phases

The classification of the principal chiral stationary phases for gas chromatography in its historical context is depicted in Table 2.

Enantiomer separation by gas chromatography has essentially been carried out by three principal methods:

- separation on nonracemic amino acid derivatives via hydrogen bonding;
- separation on nonracemic metal coordination compounds via complexation; and
- separation on cyclodextrin derivatives via (*inter alia*) inclusion

Initially, all selectors were used as involatile liquids or as solutions in squalane or polysiloxane. Later, the selectors were chemically linked to polysiloxanes (Chirasil type stationary phases). This strategy was first exemplified by the synthesis of Chirasil-Val (Scheme 1). This approach has recently been extended



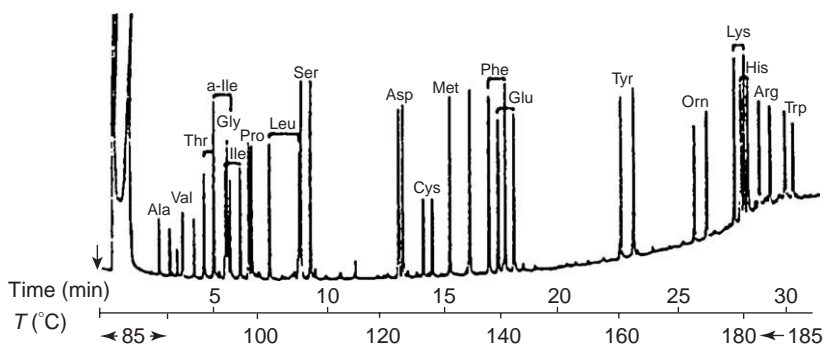
Scheme 1

to complexation gas chromatography by the synthesis of Chirasil-Metal and to inclusion gas chromatography by the synthesis of Chirasil-Dex. The chemically bonded chiral polymers combine the chemical affinity of the selector with the unique coating properties and thermal stability of polysiloxanes, affording capillary columns with high efficiency and extended range of operating temperatures. Moreover, Chirasil stationary phases can be cross-linked and immobilized on the inner wall of fused-silica capillary columns. Immobilization of Chirasil stationary phases is a prerequisite for their use in chiral supercritical fluid chromatography (SFC) and electrochromatography for enantiomer analysis of involatile and thermally labile racemates.

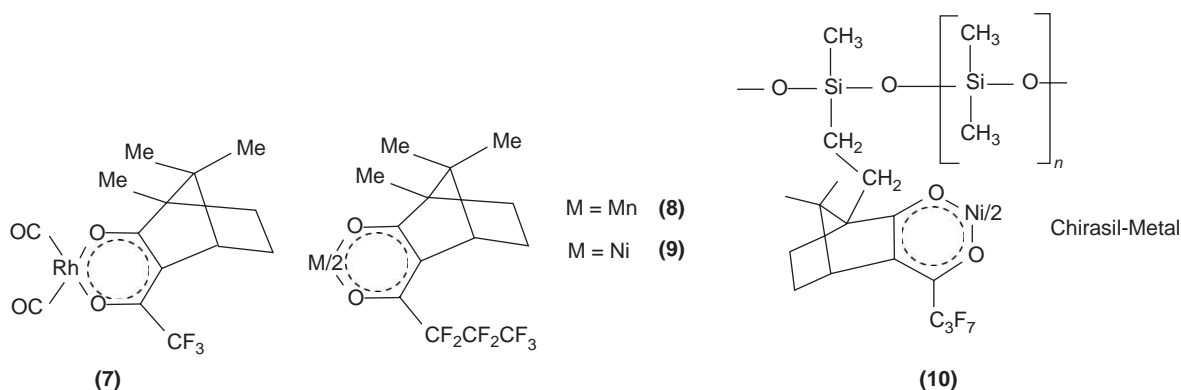
### Chiral Gas Chromatography Using Nonracemic Amino Acid Derivatives

The first successful separation of racemic *N*-trifluoroacetyl amino acid esters on glass capillary columns

coated with *N*-trifluoroacetyl-L-isoleucine lauryl ester ((1), **Scheme 1**) was achieved by Gil-Av, Feibush, and Charles-Sigler in 1966, and the great potential of this fundamental approach was clearly demonstrated. A number of improved hydrogen-bonding chiral phases were investigated subsequently, such as the dipeptide phase (2). It was recognized that the C-terminal amino acid was not essential to chiral recognition, while the additional amide function proved important for hydrogen bonding. Therefore, the second chiral center was removed by preparing the diamide phase (3), e.g., derived from valine. This selector was subsequently coupled by Frank, Nicholson and Bayer via the amino function to a statistical copolymer of dimethylsiloxane and (2-carboxypropyl)methylsiloxane of appropriate viscosity to yield Chirasil-Val (4). Chirasil-Val exhibits excellent efficiencies for the enantiomer separation of a variety of classes of compounds prone to hydrogen bonding over a broad temperature range. Chirasil-Val is commercially available in both enantiomeric forms. The simultaneous enantiomer separation of all proteinogenic



**Figure 1** Simultaneous enantiomer separation of 20 amino acids as *N,O,S*-pentafluoropropanoate isopropyl ester (histidine as *N*<sup>im</sup>-ethoxycarbonyl) derivatives by gas chromatography on Chirasil-Val [IV] between 85 and 185 °C at 0.35 bar (gauge) hydrogen. Column: 50 m × 0.27 mm (i.d.) glass capillary. D-enantiomers are eluted before L-enantiomers. (From Bayer E (1983) *Chirale Erkennung von Naturstoffen an optisch aktiven Polysiloxanen*. *Zeitschrift für Naturforschung* 38b: 1281–1291.)



**Scheme 2**

amino acids in less than 25 minutes is illustrated in Figure 1.

Another direct approach to chiral polymeric stationary phases is the modification of commercially available polysiloxanes containing reactive side-groups. Thus, the diamide phase (3) was linked to a modified XE-60 polysiloxane to give (5) and (6). In (6) another center of chirality (*R*- or *S*-configuration) has been introduced in the amide group (XE-60-L-Val-(*R* or *S*)- $\alpha$ -pea) by König *et al.*

Enantiomer separation by hydrogen-bonding stationary phases generally requires substrate derivatization.

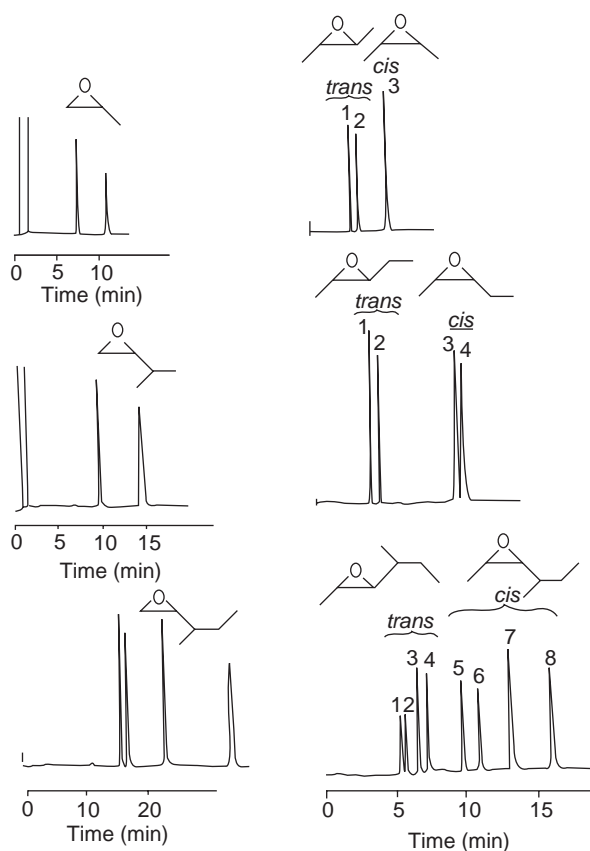
## Chiral Gas Chromatography Using Nonracemic Metal Coordination Compounds

Coordination interactions represent another strategy for producing chiral recognition in gas chromatography. Thus, the chiral metal coordination compound dicarbonylrhodium(I)-3-trifluoroacetyl-(1*R*)-campho-

rate ((7), Scheme 2) was employed to separate racemic 3-methylcyclopentene. The scope of enantiomer separation by complexation gas chromatography was later extended to oxygen-, nitrogen-, and sulfur-functionalized racemates using nonracemic ketoenolate *bis* chelates of divalent transition metal ions derived from terpene ketones such as camphor, menthone, carvone, and pulegone. Thus, manganese(II)-, and nickel(II)-[3-(heptafluorobutanoyl)-(1*R*)-camphorate] (8), (9) proved to be versatile stationary phases for the separation of underivatized cyclic ethers, esters, acetals, aldehydes, ketones, and alcohols. In Figure 2 the enantiomer separation of oxiranes by complexation gas chromatography is illustrated.

A limiting factor of Chirametal stationary phases is the low temperature range (25–120 °C). Improved thermo-stable polymeric stationary phases (Chirasil-Metal) (10) have also been prepared. Figure 3 shows the enantiomer separation of cyclic ethers in less than 20 s.

The immobilized Chirasil-Metal stationary phase can also be used in supercritical fluid chromatography

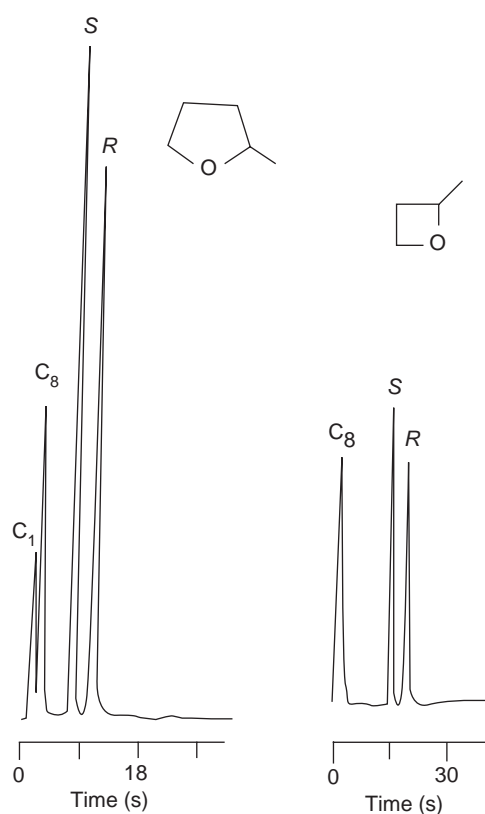


**Figure 2** Enantiomer separation of aliphatic oxiranes by complexation gas chromatography on nickel(II) bis[3-(heptafluorobutanoyl)-(1*S*)-19-ethylidene-camphorate] ( $0.125 \text{ mol kg}^{-1}$  in OV-101) between 70°C and 90°C. Column:  $25 \text{ m} \times 0.25 \text{ mm}$  (i.d.) glass capillary. (From Schurig V and Betschinger F (1992) Metal-mediated enantioselective access to unfunctionalized aliphatic oxiranes: prochiral and chiral recognition. *Chemical Reviews* 92: 873–888.)

(SFC) with supercritical carbon dioxide as mobile phase.

## Chiral Gas Chromatography Using Cyclodextrin Derivatives

The first gas chromatographic enantiomer separation on a cyclodextrin-based stationary phase was that of the apolar racemic hydrocarbons  $\alpha$ - and  $\beta$ -pinene and *cis*- and *trans*-pinane on *packed* columns coated with native  $\alpha$ -cyclodextrin dissolved in formamide. Very soon it was recognized that alkylated cyclodextrins could be employed in capillary columns for high-resolution enantiomer analysis. Thus, molten permethylated  $\beta$ -cyclodextrin ( $\beta$ -CD, heptakis (2,3,6-tri-*O*-methyl)- $\beta$ -cyclodextrin) (11) was used at temperatures above 100°C. Permethylated  $\beta$ -cyclodextrin (11) has also been used by Schurig



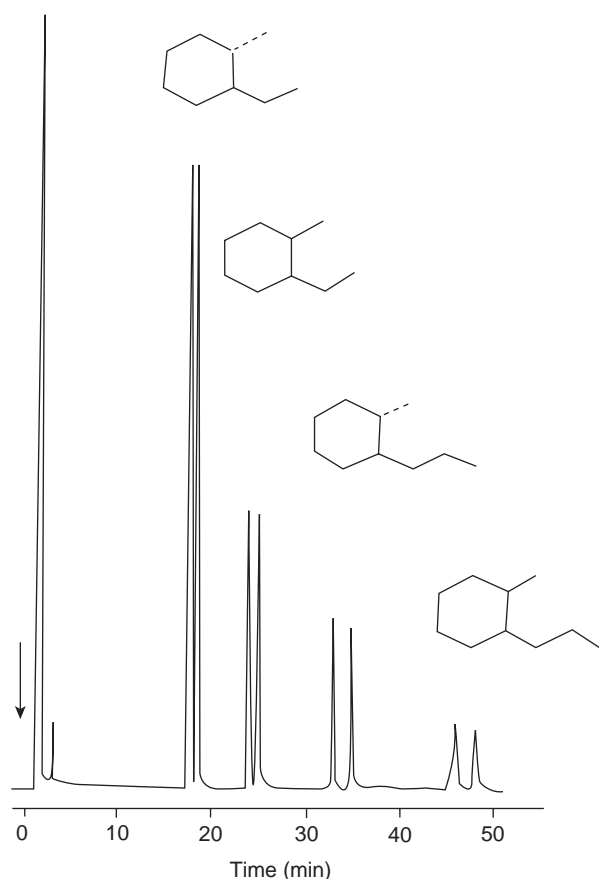
**Figure 3** Enantiomer separation of 2-methyltetrahydrofuran and 2-methyloxetane by complexation gas chromatography on immobilized Chirasil-Nickel (10) at 115°C and 2.0 bar (gauge) nitrogen (left) and 140°C and 2.0 bar (gauge) nitrogen (right). Column:  $1.5 \times 0.05 \text{ mm}$  (i.d.) fused-silica capillary. (From Schurig V, Schmalzing D, and Schleimer M (1991) Enantiomeric separation on immobilized Chirasil-Metal and Chirasil-Dex by gas chromatography and supercritical fluid chromatography. *Angewandte Chemie, International Edition* 30: 987–989.)

and Nowotny as a solution in moderately polar polysiloxanes (e.g., OV-1701). By diluting the chiral selector, the inherent enantioselectivity of cyclodextrins is combined with the unique chromatographic properties of polysiloxanes. Fused-silica capillary columns coated with (11) in OV-1701 are commercially available and show very broad applicability in enantiomer separation of various classes of compounds. The resolution of racemic unfunctionalized saturated hydrocarbons is illustrated in **Figure 4**.

A chiral test mixture has been devised to probe the efficiency and performance of permethylated  $\beta$ -cyclodextrin/OV-1701 capillary columns. This so-called *Schurig test mixture* contains both highly polar racemates (a free acid, a free base, a diol, an alcohol, and an amine) and apolar racemates (hydrocarbons), as shown in **Figure 5**.

Chemically linking the cyclodextrin derivatives to the polysiloxane backbone affording Chirasil-Dex (12)

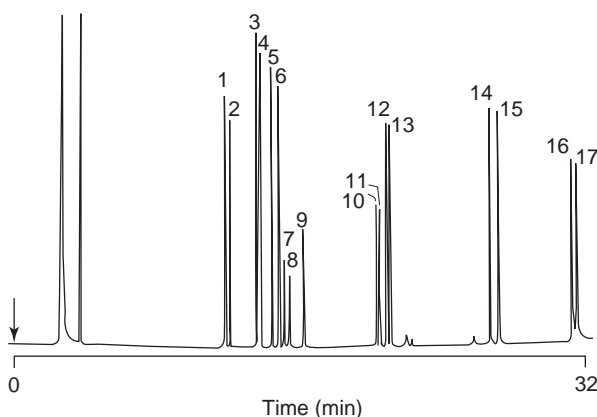




**Figure 4** Enantiomer separation of *trans*- and *cis*-1-ethyl-2-methylcyclohexane and of *trans*- and *cis*-1-methyl-2-*n*-propylcyclohexane by gas chromatography on heptakis(2,3,6-tri-*O*-methyl)- $\beta$ -cyclodextrin (11) (10%, 0.08 mol kg<sup>-1</sup> in OV 1701) at 50 °C and 0.7 bar (gauge) helium. Column: 25 m  $\times$  0.25 mm (i.d.) fused-silica capillary. (From Schurig V, Nowotny H-P, and Schmalzing D (1989) Gas-chromatographic enantiomer separation of unfunctionalized cycloalkanes on permethylated  $\beta$ -cyclodextrin. *Angewandte Chemie, International Edition* 28: 736–737.)

proved to be a useful extension of this approach. The immobilization of the resulting Chirasil-Dex stationary phases on the fused-silica surface by thermal treatment represents another refinement of the methodology. Immobilized Chirasil-Dex phases are quite resistant to stationary-phase bleeding, they are compatible with solvent intake and are insensitive to temperature shock, and have an extended working temperature range (–10 °C to 250 °C). The immobilization of Chirasil-Dex (19) was the prerequisite for enantiomer separation of involatile racemates by SFC.

Per-*n*-pentylated and 3-acyl-2,6-*n*-pentylated cyclodextrins are viscous liquids even below room temperature. The cyclodextrin derivatives (13)–(18) (Lipodex) (Scheme 3) have been used in the *undiluted* form for the separation of enantiomers of many classes of compounds on deactivated glass capillary



**Figure 5** Enantiomer separation of  $\alpha$ -pinene (1,2), *trans*-pinene (3,4), *cis*-pinene (5,6), 2,3-butanediol (rac) (7,8), 2,3-butanediol (meso) (9),  $\gamma$ -valerolactone (10,11), 1-phenylethylamine (12,13), 1-phenylethanol (14,15), and 2-ethylhexanoic acid (16,17) (Schurig test mixture) by gas chromatography on heptakis (2,3,6-tri-*O*-methyl)- $\beta$ -cyclodextrin (10%, 0.08 mol kg<sup>-1</sup> in OV 1701) at 50 °C and 0.7 bar (gauge) helium. Column: 50 m  $\times$  0.25 mm (i.d.) fused-silica capillary. (Chrompack International, Middelburg, The Netherlands.)

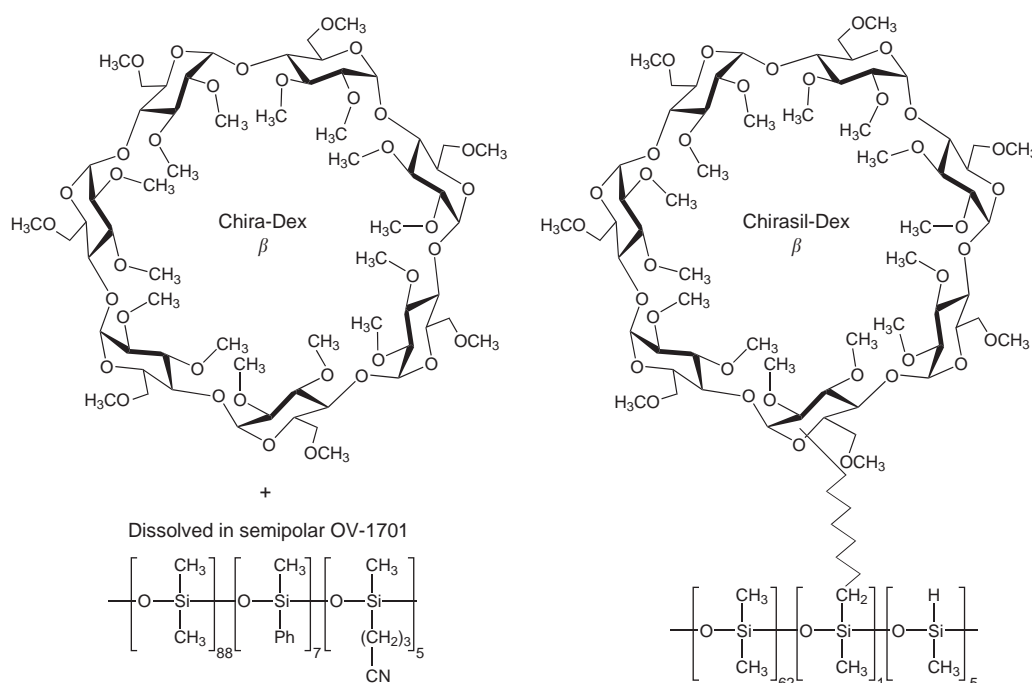
columns by König *et al.* Addition of polysiloxanes also allows the efficient coating of fused-silica columns that are commercially available. The more polar cyclodextrin derivatives containing hydroxypropyl, free hydroxy groups or trifluoroacetyl groups, (19)–(23) were coated on nondeactivated fused-silica capillary columns by Armstrong *et al.*

The 6-*O*-*tert*-butyldimethylsilyl (TBDMS)-protected cyclodextrin derivatives (24) and (25) employed by Mosandl *et al.* proved to be versatile selectors for enantiomer separation by gas chromatography.

In Figures 6 and 7 representative gas chromatograms employing different cyclodextrin derivatives are shown. The presence of three hydroxyl groups of different reactivities offers an enormous number of possible cyclodextrin ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) derivatives. As yet, no universally applicable cyclodextrin phase is available and selection is still a matter of trial and error. Separation factors  $\alpha$  for enantiomers are generally low. This feature is compensated by the high efficiency of capillary columns. Of disadvantage is the fact that cyclodextrins are available only in the all-*D* form. Hence, peak switching in the analysis of enantiomeric impurities eluted as the second peak, and therefore difficult to detect, is not possible with cyclodextrin stationary phases.

## Applications

Enantioselective gas chromatography is applied for enantiomeric analysis in many different fields such as



**Scheme 3** Cyclodextrin-derived stationary phases. **Lipodex-stationary phases.** Hexakis(2,3,6-tri-*O*-*n*-pentyl)- $\alpha$ -cyclodextrin (**13**). Hexakis(3-*O*-acetyl-2,6-di-*O*-*n*-pentyl)- $\alpha$ -cyclodextrin (**14**). Heptakis(2,3,6-tri-*O*-*n*-pentyl)- $\beta$ -cyclodextrin (**15**). Heptakis(3-*O*-acetyl-2,6-di-*O*-*n*-pentyl)- $\beta$ -cyclodextrin (**16**). Octakis(2,3,6-tri-*O*-*n*-pentyl)- $\gamma$ -cyclodextrin (**17**). Octakis(3-*O*-butanoyl-2,6-di-*O*-*n*-pentyl)- $\gamma$ -cyclodextrin (**18**). **Polar cyclodextrin stationary phases.** (*O*-(*S*)-2-hydroxypropyl)-per-*O*-methyl)- $\alpha$ -cyclodextrin (PMHP- $\alpha$ -CD) (**19**). (*O*-(*S*)-2-hydroxypropyl)-per-*O*-methyl)- $\beta$ -cyclodextrin (PMHP- $\beta$ -CD) (**20**). Hexakis(2,6-di-*O*-*n*-pentyl)- $\alpha$ -cyclodextrin (dipentyl- $\alpha$ -CD) (**21**). Heptakis(2,6-di-*O*-*n*-pentyl)- $\beta$ -cyclodextrin (dipentyl- $\beta$ -CD) (**22**). Heptakis(3-*O*-trifluoroacetyl-2,6-di-*O*-*n*-pentyl)- $\beta$ -cyclodextrin (DPTFA- $\beta$ -CD) (**23**). **6-TBDMS-modified cyclodextrin stationary phases.** Heptakis(2,3-di-*O*-acetyl-6-*O*-*tert*-butyldimethylsilyl)- $\beta$ -cyclodextrin (**24**). Heptakis(2,3-di-*O*-methyl-6-*O*-*tert*-butyldimethylsilyl)- $\beta$ -cyclodextrin (**25**).

authenticity control of essential oils, flavors, and fragrances, in clinical chemistry, in the characterization of chiral terpenes, pheromones and organochlorine pesticides, and in the search for extraterrestrial homochirality. Asymmetric and enzymatic reactions, kinetic resolutions require the correct determination of high ee values. Many hyphenated approaches in enantioselective gas chromatography are known such as GC–MS (SIM = selected ion monitoring), two-dimensional approaches GC–GC and LC–GC and GC–GC–MS (IR = isotopic ratio). Derivatized amino acid analysis is preferentially performed on Chirasil-Val stationary phases.

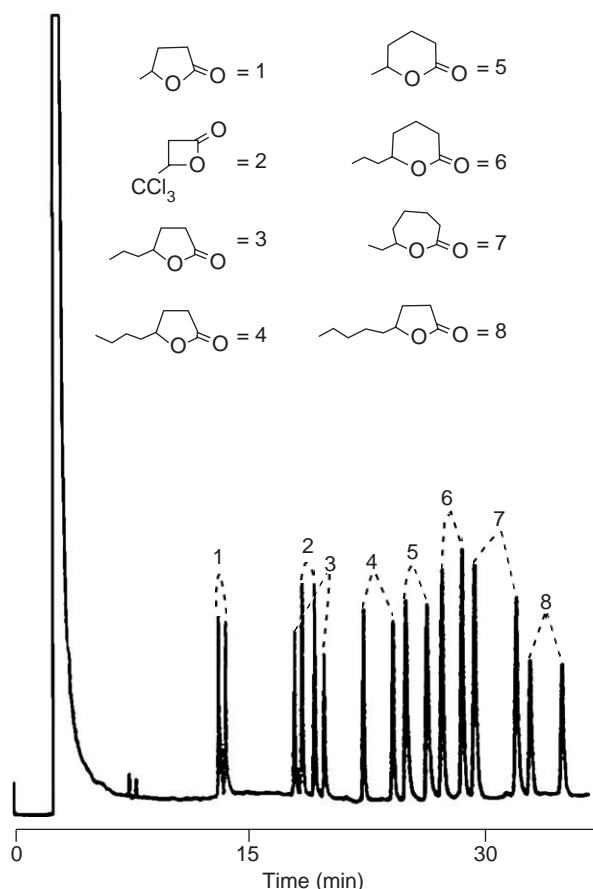
It is recommended to use short columns (2–5 m  $\times$  250  $\mu$ m) for the analysis of less volatile analytes. The loss of efficiency is usually outweighed by the increase of selectivity at reduced elution temperatures.

## Preparative Enantiomer Separations by Gas Chromatography

Preparative enantiomer separations are restricted to volatile racemic compounds. However, in contrast to liquid chromatography, the removal of the gaseous

mobile phase from volatiles is straightforward. In the development of enantioselective gas chromatography, packed columns containing chiral selectors coated on solid supports were initially involved (e.g., involatile amino acid derivatives, metal coordination compounds, and native cyclodextrins).

In complexation gas chromatography, semipreparative separations of spiroketals (among them pheromones) have been reported. The preparative invertomer separation of 1-chloro-2,2-dimethylaziridine permitted the determination of chiroptical data, absolute configuration and inversion barrier. Very large separation factors  $\alpha$  were observed for saturated hydrocarbons (*cis*- and *trans*-pinane, camphene) on a mixture of  $\alpha$ -cyclodextrin in formamide impregnated on celite. The preparative enantiomer separation of  $\alpha$ -ionone on (**18**) and of all-*trans*-perhydrotriphenylene on (**25**) has been described. The large separation factors observed for the inhalational anesthetics enflurane, isoflurane, and desflurane on (**18**) allowed their enantiomer separation for subsequent biomedical trials and acquisition of chiroptical data. The continuous preparative enantiomer separation of enflurane and



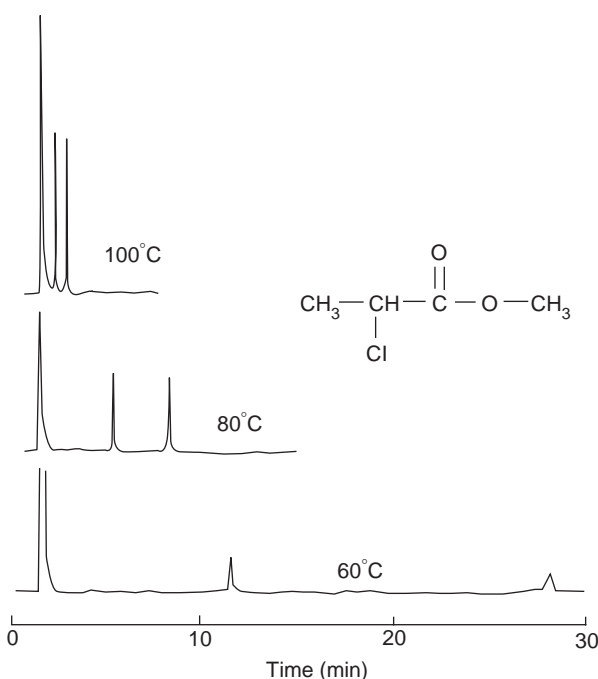
**Figure 6** Enantiomer separation of lactones by gas chromatography on octakis(3-*O*-butanoyl-2,6-di-*O*-*n*-pentyl)- $\gamma$ -cyclodextrin (**18**) (undiluted) at 150 °C and 1 bar (gage) hydrogen. Column: 50 m  $\times$  0.25 mm (i.d.) glass capillary. (From König WA, Krebber R, and Mischnick P (1989) Cyclodextrins as chiral stationary phases in capillary gas chromatography. *Journal of High Resolution Chromatography* 12: 732–738.)

isoflurane was also performed by an enantioselective gas-chromatographic simulated moving bed unit (enantioselective GC-SMB).

For semi-preparative separations in the 0.5–2 mg range analytical thick-film-wide-bore capillary columns coated with modified cyclodextrins can be used.

## Data Retrieval for Enantioselective Gas Chromatography

The Chirbase/GC and Chirbase/Flavor data banks document separations of enantiomers performed by gas chromatography. It contains bibliographical, structural, and chromatographic (separation factor  $\alpha$ , retention factor  $k$  and resolution  $R_s$ ) information based on standard database software.



**Figure 7** Enantiomer separation of methyl 2-chloropropanoate by gas chromatography on heptakis(3-*O*-trifluoroacetyl-2,6-di-*O*-*n*-pentyl)- $\beta$ -cyclodextrin (**23**) (undiluted) at three temperatures. Column: 20 m  $\times$  0.25 mm (i.d.) nondeactivated fused-silica capillary. (From Berthod A, Li W, and Armstrong DW (1992) Multiple enantioselective retention mechanisms on derivatized cyclodextrin gas chromatography chiral stationary phases. *Analytical Chemistry* 64: 873–879.)

See also: **Gas Chromatography:** Mass Spectrometry. **Liquid Chromatography:** Amino Acids.

## Further Reading

- Gil-Av E (1975) Present status of enantiomeric analysis by gas chromatography. *Journal of Molecular Evolution* 6: 131–144.
- Kallenborn R and Hühnerfuss H (2001) *Chiral Environmental Pollutants – Trace Analysis and Ecotoxicology*. Berlin: Springer.
- König WA (1987) *The Practice of Enantiomer Separation by Capillary Gas Chromatography*. Heidelberg: Hüthig.
- König WA (1992) *Gas Chromatographic Enantiomer Separation with Modified Cyclodextrins*. Heidelberg: Hüthig.
- König WA (1998) Enantioselective capillary gas chromatography in the investigation of stereochemical correlations of terpenoids. *Chirality* 10: 499–504.
- Koppenhoefer B, Graf R, Holzschuh H, et al. (1994) CHIRBASE, a molecular database for the separation of enantiomers by chromatography. *Journal of Chromatography A* 666: 557–563.
- Mosandl A (1997) Progress in the authenticity assessment of wines and spirits. *Analusis* 25: M31–M38.

- Schurig V (2001) Separation of enantiomers by gas chromatography – review. *Journal of Chromatography A* 906: 275–299.
- Schurig V (2002) Practice and theory of enantioselective complexation gas chromatography. *Journal of Chromatography A* 965: 315–356.
- Schurig V (2002) Chiral separations using gas chromatography. *TRAC, Trends of Analytical Chemistry* 21: 647–661.
- Schurig V and Schmidt R (2003) Extraordinary chiral discrimination in inclusion gas chromatography. Thermodynamics of enantioselectivity between a racemic perfluorodiether and a modified  $\gamma$ -cyclodextrin. *Journal of Chromatography A* 1000: 311–324.
- Trapp O, Schoetz G, and Schurig V (2001) Determination of enantiomerization barriers by dynamic and stopped-flow chromatographic methods – review article. *Chirality* 13: 403–414.
- Vetter W and Schurig V (1997) Enantioselective determination of chiral organochlorine compounds in biota by gas chromatography on modified cyclodextrins. *Journal of Chromatography A* 774: 143–175.

## GAS SENSING PROBES

See ION-SELECTIVE ELECTRODES: Gas Sensing Probes

## GASEOUS FUELS

See FUELS: Gaseous

## GASTRIC JUICES

**T J Gana and J Koo**, St. Michael's Hospital, Toronto, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1915–1924, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

Gastric juice is a composite secretion of the different cells that line the glands of the gastric mucosa of the stomach. It consists of parietal and nonparietal components. The parietal component includes hydrochloric acid (HCl) and intrinsic factor (IF). The nonparietal component, itself a composite alkaline secretion, includes mucus,  $\text{HCO}_3^-$ ,  $\text{Na}^+$ ,  $\text{Cl}^-$ ,  $\text{K}^+$ , and proteins that include the proteolytic enzymes pepsinogens I and II secreted by the chief cells. Despite knowledge of the composition of gastric juice and the availability of techniques to measure the various components, conventional gastric analysis as performed in man has involved mainly the measurement of gastric acid secretion. This is largely because the clinical significance of abnormal gastric acid secretion is well established and is also due in part to the fact that it is generally agreed that gastric acid secretion parallels pepsinogen and IF secretion in

infants, children, and adults in both the basal and stimulated states. It may also be partly because the determination of most of the other components of gastric juice is still restricted to research investigations with no proven clinical significance.

In clinical practice, gastric juice analysis is routinely used to assist in the diagnosis of diseases associated with hypersecretion, hyposecretion, or no secretion of acid. The major clinical indications for gastric juice analysis include (1) peptic ulcer patients not responding to medical treatment, (2) Zollinger-Ellison syndrome (ZES), (3) the evaluation of post-vagotomy patients to test completeness of vagotomy, and (4) the pre- and postoperative assessment of acid-reducing surgery in patients with peptic ulcers. Other clinical indications include atrophic gastritis, pernicious anaemia, Ménétrier's disease, and antral G-cell hyperplasia. In the laboratory, gastric juice analysis is used to monitor the onset, duration, and peak effect of drugs that inhibit gastric acid secretion, and consequently in the determination and scheduling of therapeutic dosage regimens. Also, gastric secretory tests are used in research for investigating the physiology and pathophysiology of gastric function.

- Schurig V (2001) Separation of enantiomers by gas chromatography – review. *Journal of Chromatography A* 906: 275–299.
- Schurig V (2002) Practice and theory of enantioselective complexation gas chromatography. *Journal of Chromatography A* 965: 315–356.
- Schurig V (2002) Chiral separations using gas chromatography. *TRAC, Trends of Analytical Chemistry* 21: 647–661.
- Schurig V and Schmidt R (2003) Extraordinary chiral discrimination in inclusion gas chromatography. Thermodynamics of enantioselectivity between a racemic perfluorodiether and a modified  $\gamma$ -cyclodextrin. *Journal of Chromatography A* 1000: 311–324.
- Trapp O, Schoetz G, and Schurig V (2001) Determination of enantiomerization barriers by dynamic and stopped-flow chromatographic methods – review article. *Chirality* 13: 403–414.
- Vetter W and Schurig V (1997) Enantioselective determination of chiral organochlorine compounds in biota by gas chromatography on modified cyclodextrins. *Journal of Chromatography A* 774: 143–175.

## GAS SENSING PROBES

See **ION-SELECTIVE ELECTRODES: Gas Sensing Probes**

## GASEOUS FUELS

See **FUELS: Gaseous**

## GASTRIC JUICES

**T J Gana and J Koo**, St. Michael's Hospital, Toronto, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 1915–1924, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

Gastric juice is a composite secretion of the different cells that line the glands of the gastric mucosa of the stomach. It consists of parietal and nonparietal components. The parietal component includes hydrochloric acid (HCl) and intrinsic factor (IF). The nonparietal component, itself a composite alkaline secretion, includes mucus,  $\text{HCO}_3^-$ ,  $\text{Na}^+$ ,  $\text{Cl}^-$ ,  $\text{K}^+$ , and proteins that include the proteolytic enzymes pepsinogens I and II secreted by the chief cells. Despite knowledge of the composition of gastric juice and the availability of techniques to measure the various components, conventional gastric analysis as performed in man has involved mainly the measurement of gastric acid secretion. This is largely because the clinical significance of abnormal gastric acid secretion is well established and is also due in part to the fact that it is generally agreed that gastric acid secretion parallels pepsinogen and IF secretion in

infants, children, and adults in both the basal and stimulated states. It may also be partly because the determination of most of the other components of gastric juice is still restricted to research investigations with no proven clinical significance.

In clinical practice, gastric juice analysis is routinely used to assist in the diagnosis of diseases associated with hypersecretion, hyposecretion, or no secretion of acid. The major clinical indications for gastric juice analysis include (1) peptic ulcer patients not responding to medical treatment, (2) Zollinger-Ellison syndrome (ZES), (3) the evaluation of post-vagotomy patients to test completeness of vagotomy, and (4) the pre- and postoperative assessment of acid-reducing surgery in patients with peptic ulcers. Other clinical indications include atrophic gastritis, pernicious anaemia, Ménétrier's disease, and antral G-cell hyperplasia. In the laboratory, gastric juice analysis is used to monitor the onset, duration, and peak effect of drugs that inhibit gastric acid secretion, and consequently in the determination and scheduling of therapeutic dosage regimens. Also, gastric secretory tests are used in research for investigating the physiology and pathophysiology of gastric function.



This article gives an overview of the indications, method of sampling, and the methods and techniques used for the analyses of the various constituents of gastric juice; the interpretations and clinical importance of the results with particular emphasis on acid secretion are also discussed. The methods are limited to those employed in man but relevant techniques and results obtained in animal studies are mentioned.

## Sampling of Gastric Juice

### Nasogastric Intubation

Samples of gastric juice for analyses are obtained through nasogastric intubation into the most dependent part of the stomach, the antrum. The tube should be stiff to avoid bending or coiling, its walls strong enough to prevent collapse during suction and preferably plastic, which is less irritating than rubber. Double-lumen nasogastric tubes with the tips of the main and smaller tubes in the gastric antrum and upper body, respectively, are preferable. The smaller tube is used for the infusion of nonabsorbable markers. A variety of nasogastric tubes are available commercially, although it is common practice among investigators to have them custom made; generally, sump tubes are the most popular. The side holes should be made close to the tip of the tube to ensure a localized aspiration of gastric juice. Confirmation of accurate nasogastric tube placement in the stomach is extremely important because tubes may pass into the duodenum or coil in the stomach or oesophagus. Nasogastric tube placement is most accurately performed under fluoroscopic guidance. Nevertheless, nonradiologic methods of confirming 'blind' tube placement are also in use and include the water recovery test, epigastric auscultation of insufflated air, measurement of the pH of aspirated fluid, and passing the nasogastric tube so that its tip is 550–600 mm beyond the nares. The water recovery test has been shown in several studies to be as accurate as fluoroscopy in nasogastric tube placements; hence it is the most widely used nonradiologic method. It involves drinking 20 ml of water after the nasogastric tube is passed into the stomach and attempting immediate syringe aspiration through the tube. The tube is then withdrawn 25 mm at a time and the test repeated until the shortest distance of the tube where 90% or more (18–20 ml) of the amount of water drunk was recovered, which signifies a 'pass'. In patients with previous gastric resection or drainage operations, fluoroscopically guided tube placement is indicated to ensure the most optimal position.

### Collection

With the patient sitting or lying in the left lateral decubitus position, the throat is sprayed with a local anesthetic to reduce any discomfort and the tube is passed into the gastric antrum under fluoroscopic guidance. Gastric juice is aspirated by constant hand or continuous mechanical suction and put into separate, consecutive 10- or 15-min sample bottles. It is recommended that the suction be stopped periodically when using either method to prevent blockage of the side holes of the nasogastric tube by mucus, mucosa, or food. Accessory small air bleed tubes can also be attached to the nasogastric tube to prevent blockage.

### Nonabsorbable Volume Markers

In the presence of an accurate nasogastric tube placement, most studies using nonabsorbable volume markers have reported about a 90% recovery, indicating that an inevitable loss of about 10% of the gastric juice volume occurs through the pylorus. The infusion of nonabsorbable volume markers is used to improve the accuracy of the recovered gastric volumes. Table 1 summarizes marker substances that have been used and those that have potential for human use and their methods of estimation. Of these markers, poly(ethylene glycol) (PEG) has been the most widely used. However, the methods used for its estimation are still laborious, nonspecific, insensitive, and are associated with analytical errors. In addition,

**Table 1** Summary of nonabsorbable gastric volume markers and their methods of analysis

Markers	Methods of estimation
Poly(ethylene glycol) (PEG)	Turbidimetric analysis
<sup>14</sup> C-PEG	Liquid scintillation counting
<sup>3</sup> H-PEG	Liquid scintillation counting
<sup>a</sup> Radioiodinated ( <sup>131</sup> I-) serum albumin	Scintillation counting
Phenol red	<sup>b</sup> Spectrophotometry
<sup>51</sup> Chromium chloride ( <sup>51</sup> CrCl <sub>3</sub> )	Scintillation counting
<sup>51</sup> Cr-EDTA	Scintillation counting
<sup>c</sup> Cobalt-EDTA (Co-EDTA)	<sup>d</sup> Atomic absorption spectrometry Instrumental neutron activation analysis
<sup>c</sup> Chromium-EDTA (Cr-EDTA)	<sup>d</sup> Atomic absorption spectrometry Instrumental neutron activation analysis

<sup>a</sup>No longer in general use.

<sup>b</sup>Spectrophotometry is performed at 560 nm.

<sup>c</sup>Used so far in animals. Potential for routine use in human subjects.

<sup>d</sup>Cheaper, easier to use, and more readily available.

there is the risk of radioactive contamination associated with the use of  $^{14}\text{C}$ -PEG. Recently, in animals, nonradioactive cobalt-EDTA and chromium-EDTA have been introduced as gastrointestinal tract liquid-phase markers and have more accurate, sensitive, and specific methods of estimation, thus obviating the problems associated with the use of radioactive markers. Preliminary validation studies show good and comparable gastrointestinal recovery rates for both markers in animals and, therefore, great potential for human application. Gastric volume corrected for pyloric losses can be calculated from the following equation:

$$\frac{\text{Corrected gastric volume (ml)}}{\text{Recovered volume (ml)}} = \frac{\text{Recovered volume (ml)} \times [\text{Marker}] \text{ in instilled solution}}{[\text{Marker}] \text{ in recovered volume}}$$

where [Marker] is the concentration of marker.

## Methods and Techniques

The subject is intubated with a nasogastric tube after an 8–12 h overnight fast, usually, early in the morning. In patients with gastric outlet obstruction or decreased gastric motility, a clear liquid diet should be advised for at least 1 day prior to the study. Also, acid inhibitory drug treatment should be stopped 24 h before the commencement of a study or 48 h in the case of subjects on  $\text{H}^+/\text{K}^+$  ATPase inhibitors. Acid output in millimoles (mmol) is the product of the volume secreted in milliliters (ml) and the  $\text{H}^+$  concentration in millimoles per milliliter ( $\text{mmol ml}^{-1}$ ). The secretion of gastric acid is measured over a period of time, usually an hour, and is expressed as the acid output per hour ( $\text{mmol h}^{-1}$ ). Furthermore, acid output is measured in the resting basal state and during stimulation of acid secretion by pharmacologic agents and/or food. The results are reported respectively as the basal acid output (BAO), maximal acid output (MAO), and the peak acid output (PAO).

### Measurement of Gastric Acid Secretion

**Gastric aspiration** Simple aspiration of gastric juice through a nasogastric tube is the most widely used method of measuring gastric acid secretion in humans. Aspiration of gastric juice can also be performed during fiberoptic gastroscopy. The residual gastric contents are aspirated. The residual gastric juice sample and the samples of gastric juice aspirated thereafter, during the specific tests of acid secretion, should have the following gross

observation and analytic measurements performed on them: color, volume, pH,  $\text{H}^+$  concentration, and acid output.

Blood-staining or ‘coffee ground’ gastric juice indicates bleeding which could arise from the stomach, lower oesophagus, or duodenum. Bile-staining of gastric juice indicates duodenogastric reflux. In both these cases, the study should be repeated the next day and the cause of the bleeding should be sought. In subjects with previous gastric resection, it may be impossible to obtain bile-free samples.

The pH of an aliquot of the gastric aspirate is measured to the nearest 0.01 pH unit by using a glass electrode and pH meter. It is vital to accurately calibrate the glass electrode with standard buffer solutions of pH 1.09, 4.01, and 7.0 before and halfway through each batch of hourly samples of gastric juice. Gastric juice pH that stays above 6.0 after stimulation is defined as achlorhydria.

Hydrogen ion concentration in the aspirate can be determined either from the pH obtained by the glass pH electrode or by determining the titratable acidity. According to the method of Moore and Scarlata, the pH values determined by the glass pH electrode represent the negative logarithm of the hydrogen ion activity  $a_{\text{H}^+}$ . The hydrogen ion concentration ( $C_{\text{H}^+}$ ) is then calculated from the equation  $C_{\text{H}^+} = a_{\text{H}^+}/\gamma_{\text{H}^+}$ , where  $\gamma_{\text{H}^+}$ , a correction factor, is the activity coefficient of hydrogen ion in gastric juice which is assumed to be equal to the hydrochloric acid activity,  $\gamma_{\text{HCl}}$ , since it is not possible to determine the former for a single ion species. Tables of activity coefficients for determining hydrogen ion concentration at different concentrations of  $\text{Na}^+$  and  $\text{K}^+$  in gastric juice have been published.

Titratable acidity is determined by titration of aliquots of the gastric juice sample with  $0.1 \text{ mol l}^{-1}$  NaOH to neutrality. Although neutrality has been traditionally set at pH 7.0, the physicochemical point of neutrality, others have recommended titration to pH 7.4, the pH of plasma. This issue remains unsettled. The amount of NaOH (in  $\text{mmol l}^{-1}$ ) used for titration to neutrality (pH 7.0) is the titratable acidity.

Measurements of gastric acid secretion performed by gastric aspiration include basal, maximal, and peak acid output. Acid output cannot be accurately measured by gastric aspiration in the presence of food in the stomach; hence, it is not used to measure meal-stimulated acid output.

**Basal acid output (BAO)** An estimate of the resting, unstimulated acid secretion is defined as the sum of the acid output from the gastric aspirate obtained over four consecutive 15-min periods. Table 2 shows

**Table 2** Upper and lower limits of normal acid outputs in healthy adult men and women<sup>a</sup>

	Sex	Lower limit	Upper limit
BAO (mmol h <sup>-1</sup> )	Men	–	10.5
	Women	–	5.6
MAO (mmol h <sup>-1</sup> )	Men	6.9	47.8
	Women	5.0	30.2
PAO (mmol h <sup>-1</sup> )	Men	11.6	60.6
	Women	8.0	40.1
BAO/PAO ratio	Men	–	0.29
	Women	–	0.23

<sup>a</sup>Upper and lower limits of normal (95% confidence limits) were derived from data obtained from 100 healthy men and 50 healthy women. There were no lower limits of normal for BAO or for the ratio of BAO to PAO because many normal subjects have a BAO of zero. BAO, basal acid output; MAO, maximal acid output; PAO, peak acid output. (Reprinted with permission from Feldman M (1989) Gastric secretion in health and disease. In: Sleisenger MH and Fordtran JS (eds.) *Gastrointestinal Disease: Pathophysiology, Diagnosis, Management*, vol. I, pp. 713–734. Philadelphia: W.B. Saunders Company.)

the normal range of basal acid outputs for healthy adult men and women. For unknown reasons, normal BAO is higher in men than in women. The emotional state of the individual affects BAO. Repeated measurements of BAO in the same individual show a coefficient of variation of about 50% and vary from hour to hour in a circadian pattern.

**Maximal acid output (MAO) and peak acid output (PAO)** After measuring the BAO, the maximal acid secretory capacity expressed as the MAO and PAO are determined. A supra-physiological one-time dose of pentagastrin is administered subcutaneously (6 µg kg<sup>-1</sup>), intramuscularly (6 µg kg<sup>-1</sup>), or by continuous intravenous infusion (6 µg kg<sup>-1</sup> h<sup>-1</sup>). Histamine or the synthetic histamine analog, beta-zole, can also be used. The gastric contents are aspirated every 15-min period for 1 h and analyzed. The PAO is calculated by multiplying the sum of the two highest 15-min acid outputs by 2 and is expressed in mmol h<sup>-1</sup>. The MAO, also in mmol h<sup>-1</sup>, is the sum of the acid output of the four consecutive 15-min periods. The normal values of PAO and MAO for healthy adult men and women are shown in Table 2. The PAO and MAO are a reflection of the total parietal cell mass and vary with age, sex, body weight, and lean body mass. They are lower in women than in men. Studies have shown that this difference is not only due to the lower parietal cell mass but is also due in part to a decreased sensitivity of the parietal cell mass to gastrin and other stimulants of acid secretion in women. The reason for the latter observation is not known. Maximal

secretory capacity is lower in children than in adults but comparable when expressed as a function of body weight. Both PAO and MAO are reproducible in the same subject day after day and over a long period of time. However, in the elderly there is a decline in gastric acid secretion.

**Continuous *in vivo* intragastric titration** This technique is used to measure the gastric acid secretory response induced by a meal. The gastric pH is raised to 5.5 by the infusion of sodium hydrogencarbonate or hydroxide. A homogenized test meal, also with a pH adjusted to 5.5, is ingested or instilled through a nasogastric tube. For a period of 4 h thereafter, frequent aspirations of 2–3 ml samples of gastric fluid are obtained, the pH is quickly determined and then returned to the stomach. Whenever the pH is observed to drop, sodium hydrogencarbonate or hydroxide is infused into the stomach to maintain the pH at 5.5. The amount of sodium hydrogencarbonate or hydroxide (in mmol h<sup>-1</sup>) required to keep the pH at 5.5 is equal to the rate of postprandial gastric acid secretion. The method is reproducible and is the simplest of the methods used to measure meal-stimulated gastric acid secretion. A major criticism of this method is that the pH is maintained at an unphysiological level which is associated with rising levels of serum gastrin and which could cause alterations in gastric acid secretion.

**Serial dilution indicator technique** This technique, also used to measure meal-stimulated acid output, utilizes a dye dilution and double sampling technique to calculate gastric volume and the rate of gastric emptying. The principle is based on ascertaining the volume of fluid in the stomach by determining the increase in the concentration of a dye (phenol red) produced by the addition of a small concentrated measure of the same dye. By determining the concentration of the solution before ( $c_1$ ) and after ( $c_3$ ) addition of a known concentration ( $c_2$ ) and volume ( $V_2$ ) of the dye, the original gastric volume ( $V_1$ ), can be calculated from the equation

$$V_1 = V_2 \frac{c_2 - c_3}{c_3 - c_1}$$

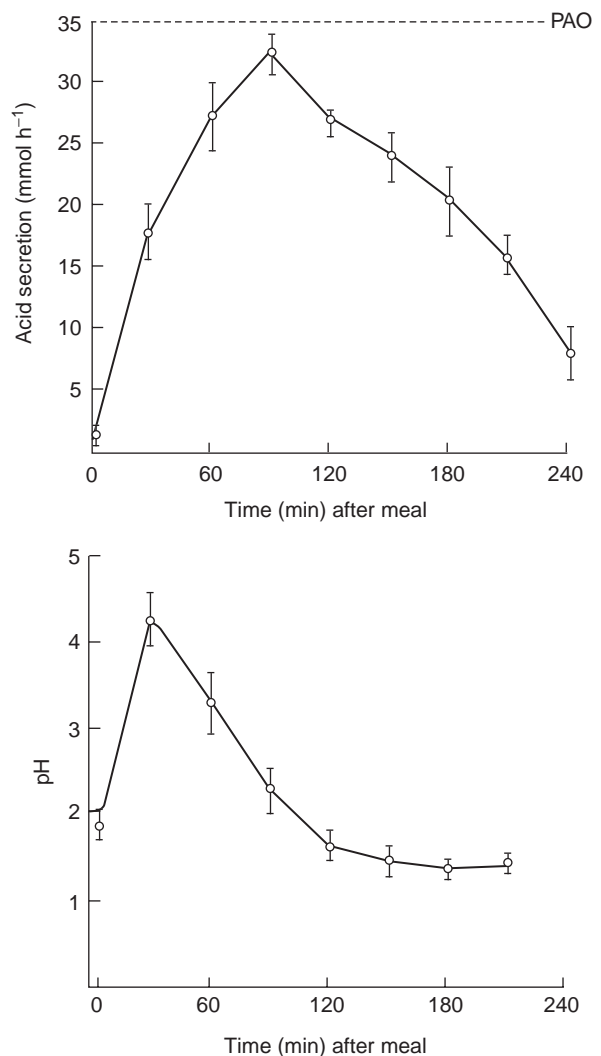
By measuring the H<sup>+</sup> concentration of the gastric juice and from the gastric volume ( $V_1$ ) obtained, the acid secretory rate can be calculated. The results obtained by this method are similar to those obtained by the continuous *in vivo* intragastric titration method. The method is easy to perform, convenient, does not alter intragastric pH, and the volume calculations are independent of any action of the

dye in the stomach; it also allows multiple measurements of total gastric volume and simultaneous measurement of the rate of gastric emptying during the study. Disadvantages of the method are that it does not consider contributions from swallowed saliva and possible duodenogastric reflux. It is also a cumbersome method and its use has been limited to research investigations.

**Meal-stimulated acid output** The rate of gastric acid secretion following a meal (food is the most important physiological stimulus of gastric acid secretion) is measured by *in vivo* intragastric titration and by the serial indicator dilution techniques. Meal-stimulated gastric acid secretion reaches a peak in about 90 min (**Figure 1**) after the meal. It is not significantly affected by the temperature of the meal or postprandial exercise. In contrast, postprandial intragastric pH increases above 3.0 within 60 min after the meal, an effect attributed to the buffering effect of the proteins in the food. The pH decreases to below basal levels (pH 2.0) in about 2 h after the meal when the acid secretion rate is still relatively high. The latter effect is thought to be the result of complete utilization of the protein food buffers and/or its gastric emptying.

**Measurement of 24-h intragastric pH, acidity, and secretion** Using the techniques of gastric aspiration combined with *in vivo* intragastric titration (only in the postprandial periods), 24-h profiles of the intragastric pH and acidity, and total acid output have been obtained. **Figure 2** shows the results obtained from one such study conducted in healthy adult men and duodenal ulcer patients. Average 24-h acid secretion was 208.30 mmol in the normal subjects and 408.30 mmol in the duodenal ulcer patients. About two-thirds of the acid secreted in the normal subjects was during the day (9 a.m. to 9 p.m.) and about one-third was during the night (9 p.m. to 9 a.m.). Gastric aspiration has been criticized as unphysiological because it may increase acid secretion and stimulate duodenogastric reflux. Also, the repeated hourly aspirations of gastric samples is laborious, confines the subject, and can be uncomfortable especially at night.

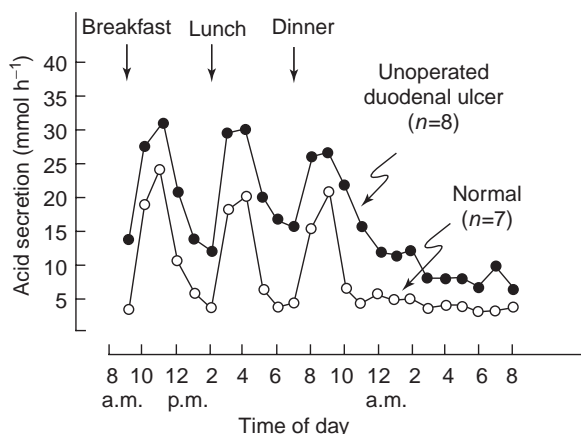
Another method of assessing 24-h intragastric acidity but which cannot measure volume and hence acid output, is the continuous monitoring of intragastric pH (pH-metry) with an electrode attached to the tip of the nasogastric tube. The bipolar glass pH electrode is the most widely used type. It is small, 2 to 4 mm in diameter, has a linear pH range (0–12), good response time (<2 s), and an electrode error of less than 0.05 pH. Other electrodes that are in use



**Figure 1** Mean ( $\pm$  SE) acid secretion (top) and intragastric pH (bottom) following an eaten sirloin steak meal. On one day (top), acid secretion was measured by *in vivo* intragastric titration to pH 5.5 in six subjects. On a second day (bottom), intragastric pH was allowed to seek its natural level in ten subjects. The mean ( $\pm$  SE) basal acid secretion rate (top) and basal pH (bottom) prior to the meal are shown at 0 min. Peak acid output (PAO) is also indicated. (Reprinted with permission from Feldman M (1989) Gastric secretion in health and disease. In: Sleisenger MH and Fordtran JS (eds.) *Gastrointestinal Disease: Pathophysiology, Diagnosis, Management*, vol. 1, pp. 713–734. Philadelphia: W.B. Saunders Company.) SE, standard error.

include the monocrystalline antimony electrode (unipolar, hence requires a skin reference electrode), the ion-sensitive field effect transistor electrode (ISFET), and the polymeric membrane electrode. Disadvantages of this pH monitoring technique are: it only reflects the  $H^+$  activity, maintaining stable electrode placement can be difficult and may affect the results, and the equipment may develop technical problems. The advantages are: it is not





**Figure 2** Mean hourly gastric acid secretion during a 24 h period in eight unoperated duodenal ulcer patients and in seven normal men. Breakfast, lunch, and dinner were infused at 9 a.m., 2 p.m., and 7 p.m., respectively. Hourly acid secretion rates are plotted at the end of each hour. (Reprinted with permission from Feldman M and Richardson CT (1986) Total 24-hour gastric acid secretion in patients with duodenal ulcer: comparison with normal subjects and effects of cimetidine and parietal cell vagotomy. *Gastroenterology* 90: 540–544; © W.B. Saunders Company.)

labour-intensive and the study can be conducted with the subjects ambulatory and in their normal daily environment.

**Measurement of other constituents of gastric juice** The methods used to measure the other components of gastric juice, which include hydrogencarbonate, intrinsic factor, pepsin, proteins, other electrolytes ( $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Cl}^-$ ), and mucus, are summarized in Table 3. The principles or procedures, disadvantages and, in some cases, the range of normal values are also listed.

## Significance and Interpretation of Results

### Disease-Associated Changes in Gastric Acid Secretion

The disorders associated with abnormal gastric secretion rates and the frequency with which the abnormality occurs are listed in Table 4. In the Zollinger-Ellison syndrome (ZES), basal acid secretion rates are 60% of maximal or greater ( $\text{BAO}/\text{PAO} \geq 0.6$ ). In addition, decreased gastric acid secretion is rarely seen in patients with the somatostatinoma syndrome, a condition associated with an islet cell tumor that secretes excessive amounts of somatostatin, a hormone that inhibits acid secretion. There are recent reports of decreased gastric acid, pepsin and intrinsic factor outputs as part of a gastric

secretory failure in patients with the acquired immunodeficiency syndrome. Disorders in which there is oversecretion of histamine, for example, systemic mastocytosis, foregut carcinoid tumors, and basophilic leukemia are rare causes of acid hypersecretion. Raised intracranial pressure from head injury, although uncommon, is also a cause of acid hypersecretion.

### Testing Completeness of Vagotomy

Absolute values for BAO and PAO that indicate a complete vagotomy and/or antrectomy are not available. Nevertheless, two tests that assess the integrity of the vagal innervation and efferent vagal pathways of the gastric parietal cell and acid secretion have been used in assessing the completeness of vagotomy. These are the insulin hypoglycemia test and the sham feeding test. The use of pepsin secretion in response to sham feeding in testing for the completeness of vagotomy has been recently introduced with good results. However, further clinical evaluation of the test is required to establish its clinical usefulness.

**Insulin-induced hypoglycemia test** The insulin-induced hypoglycemia test (Hollander test) is based on the assumption that hypoglycemia induced by administration of insulin stimulates the vagal centers in the hypothalamus, leading to an increase in gastric acid secretion mediated primarily through the vagal efferent supply of the stomach. Vagotomy will abolish or markedly reduce insulin hypoglycemia-induced gastric acid secretion. Tolbutamide has also been used for performing this test.

The results of this test are difficult to interpret and insulin-induced hypoglycemia stimulates acid secretion through nonvagal mechanisms as well. More importantly, insulin (and tolbutamide) may induce cerebral hypoglycemia, resulting in seizures, coma, and death. Myocardial infarctions, dysrhythmias induced by hypokalemia and release of catecholamines leading to death have also been reported during this test. Because of these problems, the test has been virtually abandoned in most centers around the world. It is being replaced by the sham feeding test.

**Sham feeding test** This is a recently introduced test in which the subject sees, smells, tastes, and actually chews appetizing food but does not swallow it. The accompanying sham feeding-stimulated gastric acid secretion is believed to be mediated through efferent vagal pathways. Studies have shown that sham feeding stimulates as much acid secretion as  $0.1 \text{ U kg}^{-1}$  of intravenous regular insulin and that an adequate



**Table 3** Summary of methods, their principles and/or procedures, and disadvantages for the measurement of other constituents of gastric juice

Constituents	Methods	Principle/procedure	Disadvantages	Normal values
Hydrogencarbonate	Titrateable alkalinity	Back-titration of alkaline gastric juice with HCl to neutrality (pH 7.0)	Requires the absence of acid	$236 \pm 48 \mu\text{mol h}^{-1}$
	pH and $\text{pCO}_2$ method	Measurement of pH and $\text{pCO}_2$ of gastric juice and calculation of hydrogencarbonate concentration from Henderson-Hasselbalch equation	May be affected by loss of $\text{CO}_2$	$366 \pm 23 \mu\text{mol h}^{-1}$
	Two-component method	Assumes fixed ratio of osmolality of plasma to osmolality of nonparietal secretion	Requires the presence of acid	$2600 \pm 600 \mu\text{mol h}^{-1}$
Intrinsic factor (IF)	Coated charcoal method	Charcoal separation of free IF from IF-bound radioactive vitamin $\text{B}_{12}$ ( $^{57}\text{CoB}_{12}$ )	Result affected by temperature, period of incubation, pH, and concentration of reagents	Basal output: $100\text{--}8000 \text{ ng h}^{-1}$  Stimulated output: $1000\text{--}30\,000 \text{ ng h}^{-1}$
Pepsin	Modified Anson and Mirsky method	Protein digestion of substrates and estimation of the soluble or dialyzable digestion products	Result affected by temperature, type of substrate used, pH, and duration of incubation	
Proteins				
Total proteins	Biuret reaction	Precipitation of neutralized gastric juice with tungstic acid and biuret reaction		
Protein components	Biuret reaction and cellulose acetate electrophoresis	Precipitation of neutralized gastric juice with tungstic acid and biuret reaction; separation of protein fractions		Gastric albumin clearance: $14\text{--}39 \text{ ml } 24 \text{ h}^{-1}$
Other electrolytes				
$\text{Na}^+$	<sup>a</sup> AES (Flame) <sup>b</sup> AAS			
$\text{K}^+$	AES (Flame) AAS			
$\text{Cl}^-$	Coulometry Potentiometry with ion-selective electrodes			
Mucus				
Adherent gel layer	<sup>c</sup> Histochemical techniques	Evaluates quantity of acid and neutral mucoproteins in stained tissue sections	Qualitative	
	<sup>c</sup> Incorporation of radiolabeled components of mucoproteins	Evaluates rate of synthesis of the glycoproteins in mucus	Complex; gives no information on protective ability of mucus	
	<sup>c</sup> Gel filtration methods	Analyzes glycoprotein fractions	Complex	
	Estimation of gel layer thickness	Microscopic examination of mounted mucosal sections; slit lamp and pachymeter	Not applicable <i>in vivo</i>	$190\text{--}600 \mu\text{m}$
	<sup>d</sup> Measurement of spinability of mucus (ability to form threads)	Relates spinability to viscoelastic property of the mucus	New technique; requires further evaluation	

Continued

Table 3 Continued

Constituents	Methods	Principle/procedure	Disadvantages	Normal values
Soluble	Measurement of sialic acid in gastric juice	Uses sialic acid as an index of mucus secretion	Insight about mucus viscosity and protective property limited	
	Chromatographic techniques	Analyses of separate mucus fractions	Results difficult to interpret	
	Measurement of single mucus components	Measures the concentrations of the mucus components	Results difficult to interpret; affected by presence of bile	
	Estimation of acid and neutral mucoproteins	Evaluates prevalence of acid and neutral mucoproteins in gastric juice	Results unreliable in the presence of duodenogastric bile reflux	
	Estimation of the mucoprotective index: ratio of neutral to total (acid + neutral) mucoproteins $\times 100$	Assumes that neutral mucoprotein content is an index of the viscosity and protective properties of the mucus	Qualitative character of mucus indirectly related to its protective capacity	$67 \pm 7\%$
	Measurement of mucus viscosity	Relates viscosity to protective capacity	Accuracy of results uncertain	

<sup>a</sup> AES = atomic emission spectrometry.

<sup>b</sup> AAS = atomic absorption spectrometry.

<sup>c</sup> Can only be used in resected human gastric specimens.

<sup>d</sup> Can be applied in humans *in vivo*.

vagotomy abolishes sham feeding-stimulated gastric acid secretion.

The appetizing meal consists of sirloin steak, French-fried potatoes and water, but varies between laboratories. Subjects are advised not to swallow the food as it invalidates the test and has to be repeated if it happens. The subject is asked to start chewing the food after 1 h of BAO measurements. During sham feeding (lasts about 30 min) and for 30 min thereafter, serial 15-min samples of gastric contents are aspirated. The sham feeding-stimulated acid output (SAO) is the sum of the four consecutive 15-min period outputs during sham feeding. At the end of the sham feeding hour, a subcutaneous injection of pentagastrin is given to start the third hour of PAO measurement. In patients with vagotomy, it has been determined that  $12 \mu\text{g kg}^{-1}$  of pentagastrin, rather than the normal dose of  $6 \mu\text{g kg}^{-1}$ , is needed to stimulate a maximal acid secretory response. Values for BAO and PAO are obtained as described above.

In nonvagotomized subjects, the SAO/PAO ratio is approximately 0.40 with a lower limit of 0.10 (95% confidence limits). Approximately 70% of vagotomized subjects with no clinical evidence of ulcer recurrence have SAO/PAO ratios of less than 0.10 indicating an adequate vagotomy. Most vagotomized subjects have SAO/PAO ratios of zero. A false increase in SAO can occur from accidentally swallowing food (which may increase acid secretion) during a test and a false decrease in SAO may occur from the

buffering effect (neutralization of acid) of swallowed food. In cases where PAO values are extremely low, close to zero, the results should be interpreted cautiously as the SAO/PAO ratio will be misleadingly higher than 0.10.

### Other Constituents of Gastric Juice

The clinical importance of abnormal gastric  $\text{HCO}_3^-$  secretion is not well understood. More studies are required to define precisely the role of secreted hydrogencarbonate in ulcer disease. Intrinsic factor secretion is increased in duodenal ulcer disease and in other gastric hypersecretory states. Markedly reduced or absent IF secretion occurs in atrophic gastritis with or without pernicious anemia, juvenile pernicious anemia, gastric cancer, large gastric ulcers in the corpus and in partial gastrectomy with complete disappearance of pernicious anemia. In Ménétrier's disease (giant hypertrophic gastropathy), gastric albumin clearance and the protein concentration of gastric juice are markedly increased. Hypertrophic hypersecretory gastropathy is a condition also associated with an increased loss of proteins in the gastric juice but can be differentiated from Ménétrier's disease by a finding of hyperchlorhydria instead of the hypochlorhydria or achlorhydria associated with the latter. Normal ranges of values for the other electrolytes present in gastric juice are not available because of the significant intersubject variations of

**Table 4** Disorders associated with abnormal rates of gastric acid secretion

Disorder	<sup>a</sup> Frequency (%)
<i>With abnormally low acid secretion rates</i>	
Chronic atrophic gastritis	100
Pernicious anemia	100
Vitiligo	20–25
Alopecia areata	6
Rheumatoid arthritis	10–20
Thyrotoxicosis	10
Gastric ulcer	Common <sup>b</sup>
Gastric carcinoma	50
Ménétrier's disease	75
Chronic renal failure	13 <sup>c</sup>
Iatrogenic <sup>d</sup>	
Postvagotomy, postantrectomy	> 90
Medical – e.g. substituted benzimidazoles, potent H <sub>2</sub> -receptor antagonists	> 80
<i>With abnormally high acid secretion rates</i>	
Duodenal ulcer	40–45
Zollinger-Ellison syndrome (ZES)	100
Hyperplasia or hyperfunction of antral gastrin cells	> 90
Hypertrophic hypersecretory gastropathy	100
Massive resection of the small intestine	50 <sup>e</sup>
Systemic mastocytosis	Rare <sup>f</sup>

<sup>a</sup> Some figures shown in this column are extrapolations based on estimates in the literature.

<sup>b</sup> Precise figures are not available; rates of gastric acid secretion tend to be lower than normal, particularly when ulcers are located in the corpus.

<sup>c</sup> Rates of acid secretion are usually normal and can be elevated in some cases.

<sup>d</sup> The purpose of surgical and medical therapy is to lower the rate of acid secretion. Vagotomy is incomplete in ~10% of cases, and the effects of antisecretory medication are dose-dependent.

<sup>e</sup> Hypersecretion of gastric acid is generally transient, and rates return to normal within weeks of the operation.

<sup>f</sup> Hyperchlorhydria is rarely observed, and most often acid secretion is normal or even decreased.

Reprinted with permission from Wolfe MM and Soll AH (1988) The physiology of gastric secretion. *New England Journal of Medicine* 319: 1707–1715.

the values obtained in reported studies. Furthermore, since the clinical significance of the Na<sup>+</sup>, Cl<sup>–</sup>, and K<sup>+</sup> content of gastric juice are not known, their measurement has been limited to specific research investigations. Despite the numerous techniques listed for the measurement of mucus in gastric juice, their complexity, uncertain value, and technical problems still hinder successful application in man. Hence, there remains the need to develop a method that is accurate, reliable, easy to use, and that directly assesses the quantity and protective role of gastric mucus.

## Specific Problems

### Contamination and Neutralization

Contamination of gastric juice and loss of H<sup>+</sup> can result from dilution and neutralization by alkaline gastric secretion, hydrogencarbonate-containing saliva, esophageal secretion, and duodenal and pancreatic secretions through duodenogastric reflux. Hydrogen ions can also back-diffuse and be lost into the gastric mucosa. These losses of H<sup>+</sup>, although leading to an underestimation of the rate of gastric acid secretion, are believed to be small relative to the rate of acid secretion. The losses are generally ignored in routine clinical tests of gastric acid secretion.

**Saliva** Contamination of gastric juice by saliva can be prevented by continuous suction using a dental suction set during gastric secretory tests. Inserting dental cotton wool packs between the gums and cheek to absorb any secreted saliva, and instructing study subjects to spit out and avoid swallowing saliva are other methods that have also been used to prevent salivary contamination of gastric juice.

**Esophageal secretions** Contamination of gastric juice samples by esophageal secretions has not been investigated. However, recent studies show that the esophagus secretes basal hydrogencarbonate.

**Duodenogastric reflux** Visible bile staining of gastric juice signifies duodenogastric reflux with the possibility of contamination with pancreatic and duodenal secretions. Such samples should be discarded.

### Isolation and Stability of Gastric Juice

Gastric juice collected for analyses should be stored on ice until the analyses can be performed. If analysis is delayed for more than 4 h, the samples should be stored in a refrigerator. Samples collected for IF and protein determination should be stored on ice until they can be depepsinized and the assay performed or stored in a refrigerator. The samples can be stored for 5 days with only 7% loss of IF activity. Samples for the estimation of pepsin should be stored in a refrigerator for not more than 1 week to ensure retention of significant peptic activity. For periods greater than 1 week, gastric juice mixed with glycerol can be stored at –20°C for many months without loss of peptic activity.

## Acid Secretory Tests

Over the past four decades, measurement of gastric acid secretion has been markedly improved by

advances in electrochemistry and technology, particularly the development of precision pH and titration equipment, and computer hardware and software. Despite these improvements, there has been a decline in the requests for tests of gastric acid secretion by physicians. The reason for the decline in the use of acid secretion tests is mostly due to the availability of more accurate and easier diagnostic tests. These include serum measurements, e.g. of gastrin and vitamin B<sub>12</sub> in ZES and pernicious anemia respectively, radiology and particularly endoscopy and biopsy. It is also due in part to the observed decline in the incidence of peptic ulcer disease and perhaps to the inconvenience and cost of acid secretion tests. Nevertheless, acid secretory tests may still be used as important adjuncts to the more specific tests in the following situations:

(1) In patients with peptic ulcer disease (duodenal ulcers) on medical treatment whose ulcer fails to heal after 12 weeks of treatment. The tests may establish the reason for the failure, e.g. a missed gastrinoma, reveal the need for a higher therapeutic drug dosage, or even establish the need for a switch to a more potent antisecretory drug.

(2) In routine pre- and postoperative assessment of patients undergoing surgery for peptic ulcer and in the assessment of recurrent ulcers in postgastrectomy and postvagotomy patients. In this case, the sham feeding test is a most valuable tool for evaluating the surgical procedure used and the skills of the operating surgeon.

Acid secretory tests remain the best research tools for investigating normal and abnormal gastric function and the best tests for evaluating the effective suppression of meal-stimulated gastric acid secretion, a mandatory test of the therapeutic effectiveness of antisecretory drugs. Measurement of 24-h intragastric pH, a new trend, has become increasingly easier

to perform, and is more convenient than gastric aspiration. However, for clinical usefulness, further improvement in the analysis of the results is needed.

*See also: Clinical Analysis: Sample Handling.*

## Further Reading

- Asano S, Morii M, and Takeguchi T (2004) Molecular and cellular regulation of the gastric proton pump. *Biological and Pharmaceutical Bulletin* 27: 1–12.
- Baron JH (1978) *Clinical Tests of Gastric Secretion: History, Methodology, and Interpretation*. London: Macmillan.
- Brooks FP (1985) Tests related to the stomach. In: Berk JE, Haubrich WS, Kalser MH, Roth JLA and Schaffner F (eds.) *Gastroenterology*, vol. 1, *Clinical Evaluation*, pp. 367–377. Philadelphia: W.B. Saunders.
- Feldman M (1989) Gastric secretion in health and disease. In: Sleisenger MH and Fordtran JS (eds.) *Gastrointestinal Disease: Pathophysiology, Diagnosis, Management*, vol. 1, pp. 713–734. Philadelphia: W.B. Saunders.
- Gana TJ, MacPherson BR, Ng D, and Koo J (1990) Cobalt as a gastric juice volume marker: Comparison of two methods of estimation. *Journal of Surgical Research* 48: 78–83.
- Londong W, Pounder RE and Scarpignato C (eds.) (1990) Long-term intragastric pH measurement in man: Basic and methodological aspects. *Digestive Diseases* 8: Suppl 1.
- Scarpignato C and Bianchi Porro G (eds.) (1990) *Clinical Investigation of Gastric Function*. Frontiers of Gastrointestinal Research, vol. 17. Basel: Karger.
- Schubert ML (2003) Gastric secretion. *Current Opinion in Gastroenterology* 19: 519–525.
- Schubert ML (1998) Regulation of gastric acid secretion. *Current Opinion in Gastroenterology* 14: 425–432.
- Van Herwaarden MA, Samsom M, and Smout AJPM (1999) 24-h recording of intragastric pH: Technical aspects and clinical relevance. *Scandinavian Journal of Gastroenterology* 34: 9–316.

# GEL PERMEATION CHROMATOGRAPHY

*See LIQUID CHROMATOGRAPHY: Size-Exclusion*

# GEOCHEMISTRY

## Contents

### Inorganic

### Sediment

### Soil, Major Inorganic Components

### Soil, Minor Inorganic Components

### Soil, Organic Components

## Inorganic

**C Trueman**, University of Portsmouth, Portsmouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Geochemistry is the study of the chemistry of natural earth materials and the chemical processes operating within and upon the Earth, both now and in the past. Geochemical analyses are carried out on any natural sample such as air, volcanic gas, water, dust, soil, rock or biological hard tissues (especially ancient biological tissues) and also on anthropogenic materials such as industrial effluent and sewage sludge. Geochemical analyses therefore involve a wide range of materials and analytes of interest and may be performed for industrial, environmental, or academic reasons. All of the naturally occurring elements in the periodic table are important for one geochemical investigation or another. Element concentrations may range from major or minor constituents ( $>0.1\%$ ) in one material to trace levels (usually expressed in  $\mu\text{g g}^{-1}$  (parts per million) or  $\mu\text{g kg}^{-1}$  (parts per billion)) in another. Geochemical tracers are commonly employed to investigate geological processes that are necessarily inaccessible to direct observation. Such tracers frequently demand high precision analysis of trace element abundance or isotopic ratio, and the demands of inorganic geochemists have played a significant role in the development of high precision multielement analytical techniques. Geological materials are commonly refractory substances and may require aggressive chemical preparation prior to analyses. Several articles in this volume deal specifically with the chemical analysis of soils, waters, and biological tissues, and organic geochemistry is also covered separately. This article concentrates on analysis of the inorganic constituents of the materials most closely associated with geology – rocks.

## The Nature of Geological Materials

Rocks of the upper portion of the Earth's crust are divided into three main types: igneous, sedimentary, and metamorphic rocks. Igneous rocks such as basalt, granite, gabbro, and rhyolite form from crystallization of mainly silicate liquids and are therefore primary rocks. Igneous rocks make up around 95% of the crustal volume, but only around 25% of the crustal area. Metamorphic rocks such as gneiss, schist, and slate are formed when existing rocks are subjected to high temperatures and/or pressures, and are commonly exposed in ancient cratonic regions of the continents. Sedimentary rocks such as sandstone and limestone are formed from the consolidation of weathered and eroded components of other rocks, or from chemical or biological precipitation of minerals from solution at the Earth's surface. Sedimentary rocks account for only 5% of the crustal volume, but around 70% of its area.

Rocks are composed of grains or crystals of one or more minerals. Although there are many thousands of naturally occurring minerals, most rocks are composed from a restricted suite, collectively known as the rock forming minerals. The composition of 29 of the most common rock forming minerals is shown in **Table 1**. The Earth is chemically differentiated, with dense ferromagnesian elements concentrated in the core and mantle, lighter rock forming elements such as Al and Si concentrated in the crust, and the lightest elements forming the atmosphere. The bulk composition of the different layers of the Earth are shown in **Table 2** and the average composition of the main rock types within the crust are shown in **Table 3**. The most abundant elements in the crust are Si, Al, and O, which together comprise around 73% of the crust. Most minerals are therefore silicates, and chemical analysis of rocks is focused on determination of elements in a silicate matrix.

## Applications of Geochemical Analyses

Given the broad definition of geochemical analyses it is unsurprising that the applications to which



**Table 1** Composition and classification of major rock forming minerals

<i>Classification</i>	<i>Group</i>	<i>Mineral</i>	<i>Composition</i>
Native elements	Metals	Gold	Au
		Silver	Ag
		Diamond, graphite	C
		Sulfur	S
Sulfides		Pyrite	FeS <sub>2</sub>
		Galena	PbS
Oxides	Hematite	Hematite	Fe <sub>2</sub> O <sub>3</sub>
		Ilmenite	FeTiO <sub>3</sub>
	Rutile	Rutile	TiO <sub>2</sub>
	Spinel	Spinel	MgAlO <sub>4</sub>
		Magnetite	Fe <sub>3</sub> O <sub>4</sub>
Halides		Halite	NaCl
Carbonates	Calite	Calcite	CaCO <sub>3</sub>
	Dolomite	Dolomite	CaMg(CO <sub>3</sub> ) <sub>2</sub>
Sulfates		Barite	BaSO <sub>4</sub>
		Gypsum	CaSO <sub>4</sub> · 2H <sub>2</sub> O
Phosphates	Apatite	Apatite	Ca <sub>5</sub> (PO <sub>4</sub> ) <sub>3</sub> (F,Cl,OH)
Neosilicates	Olivine	Forsterite	Mg <sub>2</sub> SiO <sub>4</sub>
	Garnet	Pyrope	Mg <sub>3</sub> Al <sub>2</sub> (SiO <sub>4</sub> ) <sub>3</sub>
	Zircon	Zircon	ZrSiO <sub>4</sub>
Inosilicates	Pyroxene	Augite	(Ca,Na)(Mg,Fe,Al,Ti)(Si,Al) <sub>2</sub> O <sub>6</sub>
	Amphiboles	Ferrohornblende	Ca <sub>2</sub> (Fe,Mg) <sub>4</sub> Al(Si <sub>2</sub> Al)O <sub>22</sub> (OH,F) <sub>2</sub>
Phyllosilicates	Clay minerals	Kaolinite	Al <sub>2</sub> Si <sub>2</sub> O <sub>5</sub> (OH) <sub>4</sub>
		Montmorillonite	(Na,Ca) <sub>0.3</sub> (Al,Mg) <sub>2</sub> Si <sub>4</sub> O <sub>10</sub> (OH) <sub>2</sub> · nH <sub>2</sub> O
	Mica	Biotite	K(Mg,Fe) <sub>3</sub> (AlSi <sub>3</sub> O <sub>10</sub> )(OH) <sub>2</sub>
Tectosilicates	Silica	Quartz	SiO <sub>2</sub>
	Feldspar	Orthoclase, sanidine	KAlSi <sub>3</sub> O <sub>8</sub>
		Albite	NaAlSi <sub>3</sub> O <sub>8</sub>
	Feldspathoid	Anorthite	CaAl <sub>2</sub> Si <sub>2</sub> O <sub>8</sub>

geochemical analyses are put are equally wide. Some examples are shown below:

- ‘Identifying unknown natural materials’. Mineral identification (petrology) may be performed by a combination of microscopy, X-ray diffraction and chemical analyses. Electron-probe microanalysis (EPMA) is particularly valuable in mineral identification as the technique combines spatially resolved quantitative determination of major elements with the imaging capabilities of the scanning electron microscope.
- ‘Assessing the quality of a product’. Geochemical analyses are a vital tool in the metal extraction industry where a few micrograms per gram may represent an ore for some elements. Ores may also contain significant concentrations of toxic or polluting trace elements. Phosphate ore, for instance, may

contain 50 mg kg<sup>-1</sup> Cd, which will be concentrated and potentially released during processing of the ore.

- ‘Reconstructing the history of a rock’. Many important geological processes are either extremely slow and/or occur at great depths within the crust where they cannot be observed directly. Rocks and the minerals from which they are composed form under specific temperature and pressure conditions, in contact with fluids of particular composition. The trace element and stable isotopic composition of the mineral components in a rock provide a record of these conditions and can be used to reconstruct fundamental geological processes occurring within the Earth. Rocks and minerals (especially biominerals) that form at the Earth’s surface provide a geochemical record of conditions at the Earth’s surface through time.

**Table 2** Bulk composition of the whole Earth, and major geological divisions of the Earth

Element	Earth	Core	Upper mantle	Continental crust
K	162	0	24	15 772
Ca	1.71	0	4.57	2.4
Sc	10.93	0	15.39	22
Ti	813	0	928	4197
V	94.4	150	82	131
Cr	3720	9000	2625	119
Mn	1680	3000	1045	852
Fe	30.3	85	6.26	5.13
Co	838	2000	105	25
Ni	17 600	52 000	1960	51
Cu	60.9	125	29.1	24
Zn	37.1	0	55	73
Ga	2.7	0	3.8	16
Ge	7.24	20	1.1	
As	1.669	5	0.01	
Se	2.65	8	0.071	
Br	3.57	0.034	0.005	
Kr				
Rb	0.405	0	0.041	58
Sr	13.4	0	12.935	325
Y	2.9	0	3.655	20
Zr	7.1	0	6.195	123
Nb	0.444	0	0.112	12
Mo	1660	5000	30	
Ru	1310	4000	4.97	
Rh	242	740	0.91	
Pd	1015	3100	3.85	
Ag	54.4	150	7.84	
Cd	76	150	39.2	
In	6.8	0	8.5	
Sn	251.1	500	97.5	
Sb	46.2	130	1.1	
Te	285.7	850	11.76	
I	6.8	0.13	1	
Xe				
Cs	0.014	0.065	0.504	2.6
Ba	4.445	0	0.449	390
La	0.437	0	0.08	18
Ce	1.131	0	0.538	42
Pr	0.171	0	0.114	5
Nd	0.844	0	0.738	20
Sm	0.274	0	0.305	3.9
Eu	0.104	0	0.119	1.2
Gd	0.367	0	0.43	3.6
Tb	0.067	0	0.08	0.056
Dy	0.455	0	0.559	3.5
Ho	0.101	0	0.127	0.76
Er	0.296	0	0.381	2.2
Tm	0.046	0	0.06	0.32
Yb	0.297	0	0.392	2
Lu	0.046	0	0.061	0.33
Hf	0.191	0	0.167	3.7
Ta	0.025	0	0.006	1.1
W	0.173	0.47	0.002	
Re	75.3	0.23	0.27	
Os	900	2750	3.43	
Ir	455	835	3.185	
Pt	1866	5700	7.07	
Au	164	500	0.96	
Hg	23.1	50	9.8	

**Table 2** Continued

Element	Earth	Core	Upper mantle	Continental crust
Tl	12.2	30	0.35	
Pb	0.232	0.4	0.018	12.6
Bi	9.853	25	0.5	
Th	0.054	0	0.006	5.6
U	0.014	0	0.002	1.42

Concentrations presented as % element, parts per million (ppm), or parts per billion (ppb).

Data taken from compilation in Albarède F (2003) *Geochemistry, An Introduction*. Cambridge: Cambridge University Press.

- ‘Dating rocks’. The absolute age of rocks may be established through radiometric dating. The ratio of radiogenic and stable isotopes of a single element in a mineral is related to the original concentration of the radiogenic isotope at the point when the mineral formed, and the time elapsed since formation. Several geochronometers are currently used, including those based on Rb–Sr, Nd–Sm, Lu–Hf, K–Ar, Ar–Ar, and the U–Pb decay series. Reliable dating requires accurate and precise measurements of isotopic ratios (for instance the  $^{87}\text{Sr}/^{86}\text{Sr}$  ratio of a clay mineral with an initial  $^{87}\text{Rb}/^{86}\text{Sr}$  ratio of 5 will increase by only  $\sim 0.00007$  in a million years). Clearly, accurate dating requires extremely accurate determination of isotope ratios.

- ‘Temporal variation in concentrations of elements and isotopic ratios’. Geological materials often form over long timescales and therefore provide time-resolved measurements of the natural (or anthropogenic) distribution of elements. Much attention has recently focused on assessing geochemical variation in materials that grow more or less constantly – so that spatial variation in analytes across the sample corresponds to temporal variation in the distribution of those elements or isotopes over time. This application has been used widely to determine paleoclimate fluctuations from ice cores, speleothems (e.g., stalactites and stalagmites), corals, and other incrementally grown records.

- ‘Spatial distribution of elements’. Mapping the spatial distribution of elements across land surfaces provides important information on soil chemistry, pollution and transport of nutrient and toxic elements through the soil. In this instance large numbers of analyses are frequently required so analytical techniques must be rapid and protocols standardized to cope with a wide range of sample matrices.

- ‘Determining source or provenance of natural materials’. The trace element and isotopic composition of minerals is controlled by the environment of formation and can therefore be used to discriminate

**Table 3** Working values for elemental composition in commonly used reference materials. Note that minor and trace element composition may vary greatly within major rock types

	<i>Granite GA</i>	<i>Basalt JB-1</i>	<i>Shale SGR-1</i>	<i>Limestone NIST 1c</i>
SiO <sub>2</sub>	69.9	52.17	28.24	
Al <sub>2</sub> O <sub>3</sub>	14.5	14.53	6.52	
Fe	<b>1.885</b>	<b>6.21</b>	<b>2.028</b>	<b>0.385</b>
MnO	0.09	0.16	0.034	
MgO	0.95	7.73	4.44	
CaO	2.45	<b>6.8</b>	<b>5.82</b>	<b>35.8</b>
Na	<b>2.6</b>	<b>2.03</b>	<b>2.25</b>	<b>0.0188</b>
K	<b>3.4</b>	<b>1</b>	<b>1.24</b>	<b>0.24</b>
TiO <sub>2</sub>	0.38	1.34	0.264	
P <sub>2</sub> O <sub>5</sub>	0.12	0.26	0.328	
H <sub>2</sub> O +	0.87	1.02		
H <sub>2</sub> O –	0.09	0.95		
CO <sub>2</sub>	0.11		11.58	
Ag		41	0.25	
As	1.6	2.1	66.2	1.16
Au	<b>&lt;3 ppb</b>	<b>0.7</b>	<b>8.9</b>	<b>1.3</b>
B	26	12.4	54	
Ba	<b>816</b>	<b>488</b>	<b>270</b>	<b>78</b>
Be	3.6	1.5	1.06	
Bi		31	940	
Br	<b>1.9</b>	<b>0.39</b>	<b>0.49</b>	<b>0.5</b>
C		470	31 600	
Cd	130	103	930	
Ce	<b>78</b>	<b>65.1</b>	<b>34.4</b>	<b>7.27</b>
Cl	250	172	32	
Co	<b>4.71</b>	<b>37.6</b>	<b>11.73</b>	<b>1.709</b>
Cr	<b>7.3</b>	<b>419</b>	<b>31.2</b>	<b>17.4</b>
Cs	<b>6.29</b>	<b>1.23</b>	<b>5.19</b>	<b>0.545</b>
Cu	16	56.3	66	
Dy	<b>3.7</b>	<b>4.2</b>	<b>1.8</b>	<b>0.71</b>
Er	<b>2.1</b>	<b>2.3</b>	<b>1</b>	<b>0.41</b>
Eu	<b>1.033</b>	<b>1.46</b>	<b>0.466</b>	<b>0.156</b>
F	500	393	1960	
Ga	16	18.1	11	
Gd	<b>3.7</b>	<b>4.2</b>	<b>1.9</b>	<b>0.7</b>
Ge		0.8	1.6	
Hf	<b>4.5</b>	<b>3.56</b>	<b>1.371</b>	<b>0.77</b>
Hg		28	313	
Ho	<b>0.79</b>	<b>0.87</b>	<b>0.38</b>	<b>0.15</b>
I		31		
In		52.3	96	
Ir	<b>&lt;1 ppb</b>			
La	<b>40.3</b>	<b>37.1</b>	<b>18.4</b>	<b>4.51</b>
Li	90	11.5	147	
Lu	<b>0.304</b>	<b>0.296</b>	<b>0.146</b>	<b>0.0585</b>
Mo	1	34	35.1	
Nb	12	34.5	5.2	
Nd	<b>27</b>	<b>25</b>	<b>13.8</b>	<b>4.3</b>
Ni	<b>&lt;30</b>	<b>145</b>	<b>26</b>	<b>7.7</b>
Os		1.9		
Pb	30	7.1	38	
Pd			5.2	
Pr	<b>7.9</b>	<b>6.8</b>	<b>3.7</b>	<b>1</b>
Pt			3	
Rb	<b>169</b>	<b>39</b>	<b>79</b>	<b>11.6</b>
Re		4.9		
Rh				
Ru				

**Table 3** Continued

	<i>Granite GA</i>	<i>Basalt JB-1</i>	<i>Shale SGR-1</i>	<i>Limestone NIST 1c</i>
S	80	17.9	15 300	
Sb	<b>0.124</b>	<b>0.24</b>	<b>3.29</b>	<b>1.55</b>
Sc	7.08	27.4	4.94	1.203
Se	<b>&lt;0.5</b>	<b>&lt;1</b>	<b>4</b>	<b>0.06</b>
Sm	<b>5.18</b>	<b>5.06</b>	<b>2.6</b>	<b>0.815</b>
Sn	4	1.8	1.9	
Sr	<b>312</b>	<b>459</b>	<b>393</b>	<b>269</b>
Ta	<b>1.26</b>	<b>1.66</b>	<b>0.402</b>	<b>0.087</b>
Tb	<b>0.6</b>	<b>0.7</b>	<b>0.297</b>	<b>0.116</b>
Te			248	
Th	<b>16.5</b>	<b>8.72</b>	<b>4.48</b>	<b>0.949</b>
Tl	0.008	0.11	0.33	
Tm	<b>0.31</b>	<b>0.32</b>	<b>0.15</b>	<b>0.06</b>
U	<b>5.5</b>	<b>1.7</b>	<b>5.31</b>	<b>1.42</b>
V	38	212	128	
W	<b>1.9</b>	<b>2</b>	<b>2.3</b>	<b>0.13</b>
Y	<b>21</b>	<b>23</b>	<b>10.3</b>	<b>4.1</b>
Yb	<b>2.05</b>	<b>2.04</b>	<b>0.966</b>	<b>0.388</b>
Zn	80	83	74	29
Zr	<b>152</b>	<b>129</b>	<b>45</b>	<b>29</b>

Data in bold represent best values from replicate INAA analyses. Values in light type represent working values. Major element and major element oxide composition present as wt% oxide, other elements presented as parts per million element (w/w) unless indicated.

INAA data taken from Korotev RL (1996) A self-consistent compilation of elemental concentration data for 93 geochemical reference samples. *Geostandards Newsletter* 20: 217–245. Working values data taken from Govindaraju K (1989) Compilation of working values and sample description for 272 geostandards. *Geostandards Newsletter* 13: 1–113.

between rocks or minerals formed in different places. Geochemical provenance analysis has many applications including establishing prehistoric trade routes for obsidian artifacts, identifying products of large historical volcanic eruptions, and tracking the illegal trade in fossil bones.

## Analytes of Geological Importance

As mentioned above, virtually the entire periodic table is used in some geological applications. The growth of multielement techniques such as XRF and latterly inductively coupled plasma-atomic emission spectrometry (ICP-AES) and inductively coupled plasma-mass spectrometry (ICP-MS) has led to an increasingly wide suite of elements being determined routinely in geochemical analyses. Certain groups of elements are particularly important analytes in geochemical analyses.

### Group 2 Elements

The group 2 elements are major constituents in many rock forming minerals, particularly in low temperature

carbonate systems. Substitutions of group 2 elements for calcium are at least partially temperature dependant, and as biomineralized tissues such as shells and coral form in equilibrium with seawater, Ba/Ca, Sr/Ca, and Mg/Ca ratios in historic and ancient biominerals are frequently employed as paleotemperature proxies and as records of ocean chemistry.

### Transition Elements

The transition elements are useful tracers in many geological systems. They are industrially important and form economic ores, especially in hydrothermal systems where they are often present as sulfide minerals. Cd, Hg, Zn, and Pb are persistent industrial pollutants and determination of low levels of these elements in ores and fossil fuels is critical as processing of ores or burning of fuels may concentrate and release toxic elements. The concentration of such toxic trace elements may affect the economic value of an ore or fossil fuel deposit significantly.

### C, N, O, S, H

The isotopic composition of light elements in geological materials provides abundant evidence of formation processes in rocks, including temperature of crystallization, salinity, redox conditions, evidence of bacteriological processes, and dietary composition.

### Halogens

Halogens are relatively infrequently determined in geological materials, but may be important constituents of some rock forming minerals (e.g., apatite Table 1). Increasingly, the connection between the geological distribution of elements and endemic nutrient deficiency is being recognized. For instance, goiter is the expression of I deficiency in humans, and is endemic in areas where bedrock is deficient in I. Regional geochemical mapping may help to identify areas where dietary mineral supplements could alleviate endemic mineral deficiencies.

### Rare Earth Elements (REE)

The lanthanides or rare earth elements are particularly useful in geological analyses, as their similar chemical properties but smoothly decreasing ion radius make them ideal tracers for many geological and environmental processes. The REE have been used as tracers of magma mixing, to establish provenance of archaeologically traded artifacts, to examine hydrological mixing processes and to reconstruct redox conditions in ancient oceans. The isotopic composition of Nd provides a valuable tracer for past and present oceanic mixing – a key to understanding

global climate dynamics. The REE also form natural analogues for radioactive products of the nuclear industry, and provide two radiogenic dating schemes (Lu/Hf and Sm/Nd). Separation and analysis of the REE is difficult due to their similar chemistry and relatively low abundances in many natural materials. Analysis of the REE in geological materials was instrumental in development of ICP-MS technology.

### Platinum Group Elements (PGE)

The platinum group elements are among the most valuable naturally occurring elements largely due to their scarcity. In addition to their economic importance, the PGE provide useful tracers of certain geological processes. The presence of high levels of iridium in globally distributed sediment layers, for instance, has been taken as evidence of meteoric impact or global volcanism. The PGE are relatively difficult to analyze due to their extreme low abundance, and the resistance of PGE-bearing minerals to acid digestion.

### Sampling

The goal of all sampling is to produce a subpopulation of samples drawn from a natural population that accurately represents the true abundance, distribution, and variation of the analyte(s) in question within the accepted limits of accuracy of the study. Clearly, sample design requires prior knowledge or informed assumptions regarding the distribution of the analyte in question. When analyzing geological materials, sample strategy and design is frequently compromised by the complexity and inaccessibility of suitable geological materials. In addition to general sampling guidelines, specific points should be borne in mind when sampling geological materials:

1. 'Variability of sample type'. Many different types of rock or sediment may be encountered within a single geological survey. These materials may present contrasting challenges for preparation and analysis. Comparisons of analyte concentration in different matrices have lower confidence than analyses of analyte concentrations in similar matrices.
2. 'Spatial variability'. Spatial variation or heterogeneity in analyte concentration is found at all spatial scales. Most rocks are composed of different minerals, and elemental concentration will therefore vary widely within each rock. Further spatial variations occur at outcrop and regional scales. Homogenization of the sample may be required to limit errors associated with variation in single samples, and repeated analyses across an outcrop or across a geographic region may be necessary to

control for variation on larger spatial scales. In rocks, the larger the grain size, the more material must be homogenized to ensure a valid bulk analysis. The investigator must choose between the merits of a bulk analysis and an analysis targeted on a specific mineral where background variation will be less affected by mineralogical and therefore crystal chemical factors.

3. 'Temporal variability'. Geological materials take time to form and thus each sample represents a discrete period of time. As geological processes occur over an almost infinite range of timescales, the amount of time represented by a sample also varies. If time is an important consideration of a study, one must factor in some assessment of time resolution into the sample design. It is impossible, for instance, to use ablation-based analytical techniques with a laser spot size of  $10\mu\text{m}$ , to determine the effect of annual fluctuations in climate or rainfall in a speleothem that grows by less than  $10\mu\text{m year}^{-1}$ . Furthermore, rates of geological processes are generally calculated by simply dividing change in mass/length by time. This implies constant growth rates whereas most geological processes (including mineral growth) occur stochastically.

### **Sample Collection**

Collection of samples in the field may involve simple grab samples of unconsolidated sediments, removal of chips of rock by hammering or mechanical coring or mechanical concentration of samples by panning. If sampling unconsolidated sediment, a simple polythene scoop is sufficient. Depth of sampling will be to some extent determined by the requirements of the study as many elements are actively cycled within sediments particularly at redox boundaries. Choice of sample location may be truly random, or may be manipulated to avoid or seek anthropogenic disturbance.

Collection of rock samples is most easily achieved by removing a chip from the rock surface with a hammer. Representative sampling is difficult to achieve in this manner as accessible surfaces are by definition surfaces that are exposed to weathering. Weathering may significantly alter the elemental composition of a rock, as mobile elements or elements contained in unstable minerals are lost to soil and pore waters, thereby concentrating immobile elements or elements contained in relatively stable minerals. The depth to which an exposed rock is affected by weathering varies with length of exposure, climate, vegetation, mineralogy and porosity of the rock. Sampling at depth within the body of the rock with mechanical rock-coring devices may alleviate

some of these problems, particularly if recovered cores are split and central portions removed for analysis, thus limiting contamination from the metal-coring device.

Storage of rock or rock powder samples presents few problems. However, care should be taken to ensure samples remain dry. Rocks containing minerals that formed under reducing conditions may be susceptible to oxidative decomposition. Many sulfide minerals in particular break down readily if left exposed to moist air.

### **Sample Preparation**

A wide range of analytical techniques are routinely employed in the geochemistry laboratory, and the sample preparation techniques required vary depending on the nature of the material, the type of analysis to be performed and the purpose of the study. In most cases some form of mechanical processing is required. A typical process for preparing solid rock samples for analysis is given below:

- Removal of weathered or contaminated surfaces using a hammer or hydraulic rock breaker.
- Crushing of the sample to gravel or small chips in a reciprocating steel crusher.
- Grinding of the chips to a coarse or fine powder according to the application. A coarse powder, usually sieved to a specific particle size range, is generally required for mineral separation. For bulk analyses, the sample is generally crushed to a fine powder either in a ball mill, or for small sample concentrations, hand ground in a ceramic, agate or corundum mortar.
- Splitting of powder into subsamples for replicate analyses or storage.

### **Bulk Rock or Targeted Analyses?**

Most rocks are aggregates of different minerals. The separate minerals within a rock each have differing major and trace element compositions, providing a complex, heterogeneous sample. Bulk rock analyses provide a picture of the rock as a whole, and are suitable for comparing different examples of particular rock types – for instance contrasting basalts formed at midocean ridges with those formed in island arcs. Whole rock analyses are also used to compare fine-grained sedimentary rocks, as the bulk rock composition reflects an integrated record of all sediment sources contributing to deposition of the final rock.

In many cases, bulk rock analyses are inappropriate and analyses must be targeted. Zircon, for example, is an extremely resistant mineral that can



survive multiple cycles of weathering and transport. Zircon crystals present in sedimentary rocks may have formed in granitic melts billions of years before deposition of the sandstone. Clearly analysis of the chemical composition of zircon crystals within a sandstone will say nothing about the depositional environment of the sand, but may be extremely informative as to the land surfaces eroding to produce that sand. The oldest land surfaces on Earth have been identified and studied through targeted analyses of such refractory minerals redeposited in younger sedimentary rocks.

Elemental partitioning from melt or solution into a crystal is specific to the particular mineral lattice, and monomineralic samples are required for most trace element analyses. Monomineralic analyses may be performed by mechanical separation of mineral grains, or by targeted *in situ* analyses. *In situ* analyses are clearly preferable as sample preparation is much simpler, however, the analytical accuracy and precision currently available via *in situ* analysis methods is not sufficient for all applications, particularly dating of mineral grains. In such cases physical separation of mineral grains is necessary. Mineral separation may be performed by hand picking, especially for isotopic work where relatively small sample sizes are adequate. Coarse crushed and sieved powder is spread in a glass dish or on a clean paper sheet and the required minerals are identified under a binocular microscope and removed with fine tweezers.

### Density Separation

Minerals are frequently separated according to their specific gravity (SG). Minerals with a density greater

than 2.87 are referred to as heavy minerals and can be separated from commoner minerals such as quartz, feldspar, and calcite using a 'heavy liquid' separation. A liquid with a SG of 2.87 or above is placed in a separating funnel, and the sample introduced. Dense minerals settle out where they can be retrieved, lighter minerals can be filtered from the heavy liquid. Bromoform (SG = 2.87) was commonly employed as the heavy liquid, but as this liquid is carcinogenic, most laboratories now employ aqueous sodium polytungstate solution (SG = 2.96–3.06 at 20°C).

These physical sample preparation steps may be followed by one or more stages of chemical preparation. Table 4 outlines some analytical techniques commonly employed by geochemists and the sample preparation techniques required. More details are given in technique-specific articles. While some techniques such as EPMA, X-ray fluorescence (XRF), ion microprobe and laser ablation mass spectrometry require little preparation other than presentation of a clean, relatively flat surface, most techniques require either complete or partial digestion.

### Sample Decomposition

All geological materials can be dissolved quantitatively, and there is an extensive literature on the various processes available. The choice of dissolution procedure will vary according to the nature of the sample, the number and size of the samples, the requirements of the analytical technique in terms of analyte concentration and total dissolved solid concentration, volatility of the analytes of interest, precision and accuracy required, contamination from

**Table 4** Summary of analytical techniques commonly employed for specific analytical tasks within geoscience, and preparation that may be required

Analysis	Analytical techniques available	Preparation required
Major-minor elements	EPMA ICP-OES, Flame AA	Polished solid sample Solution
Minor-trace elements	XRF ICP-MS, ICP-OES, Flame, GF AA	Clean surface, pressed powder pellet Solution
Trace elements	XRF ICP-MS INAA	Clean surface, pressed powder pellet Solution
Spatially resolved analysis	EPMA LA-ICP Ion probe PIXE	Polished sample Flat sample Flat/polished sample Flat/polished sample
Isotope analysis	MC-ICP-MS SIMS, PIMS SHRIMP IRMS	Solution (may be coupled with laser ablation) Solution, elemental separation (preconcentration) Varied Chemical gas evolution, laser fluorination, thermal combustion

Note that this is a summary of some of the more commonly applied techniques, and does not reflect the total range of techniques employed, or the full range of applications for a particular analytical instrument.

introduced chemicals, the cost and time associated with the procedure and safety concerns. Choice of an appropriate decomposition method is critical for a reliable, accurate and precise method. All dissolution methods must be evaluated through digestion and analysis of certified standard reference materials that as far as possible match the composition of the sample and reagent blanks should always be prepared along with samples. The dissolution method employed must be clearly and entirely described when reporting the results of geochemical analyses as there are few standardized dissolution protocols and the analyses obtained can only be as reliable as the dissolution method employed. Following decomposition, the sample is generally brought up to volume in a suitable matrix (usually dilute nitric acid ( $\text{HNO}_3$ )). The volumes of sample and reagent used and the final dilution will be determined by the concentration of the analytes of interest in the sample and the requirements of the analytical technique. Sample dilution is often a compromise between maximizing the signal for low-level analytes and reducing the total dissolved solid burden.

### **Acid Decomposition**

**Carbonates** Carbonate minerals such as calcite and aragonite, the principle components of limestone and most phosphate and sulfate minerals will dissolve readily in 1 N HCl at room temperature. Chloride ions may be avoided by digestion in  $\text{HNO}_3$ , however, heating may be necessary for some phosphate minerals. Many carbonate and phosphate samples are not pure, and also containing clay minerals which will not be decomposed by HCl or  $\text{HNO}_3$ . Partial digestion is therefore a chemical purification of the target carbonate or phosphate mineral. Trace elements adsorbed onto clay mineral surfaces may, however, be released during HCl/ $\text{HNO}_3$  decomposition of carbonates and phosphates. Reference materials should be subjected to the same partial extraction techniques.

**Sulfides** Sulfides and salts of the hydride-forming elements (Se, As, Te, Sb) are decomposed by  $\text{HNO}_3$ .

**Silicates** Most geological materials are silicates and as these are chemically resistant materials, powerful reagents are required to produce complete digestions. Typically silicate minerals and resistant oxides are decomposed by heating with hydrofluoric acid (HF) and another strong mineral acid with a higher boiling point, usually  $\text{HNO}_3$  or perchloric acid ( $\text{HClO}_4$ ). HF digestion is performed on a hotplate, the solution evaporated to dryness, and the resultant cake

redissolved in a weaker acid (generally 1–2%  $\text{HNO}_3$ ). HF attacks the silicate matrix and silicon is quantitatively removed as volatile  $\text{SiF}_4$ . The presence of the second acid with a higher boiling point than HF ensures that sparingly soluble metal fluorides are converted to more soluble salts on completion of the evaporation.  $\text{HClO}_4$  is preferable to  $\text{HNO}_3$  as it has a higher boiling point and is more effective at dissociating fluorides.  $\text{HClO}_4$  is, however, especially hazardous if samples contain organic materials, is more expensive than  $\text{HNO}_3$  and introduces chloride ions which may cause interferences for mass spectrometric analysis.

Open vessel HF digestion is the most commonly employed decomposition technique for silicate materials but it is not without disadvantages. Some refractory minerals fail to dissolve fully in HF– $\text{HClO}_4$ / $\text{HNO}_3$  including zircon, rutile, cassiterite, chromite and tourmaline. In addition to Si, other volatile metals such as Se, Hg, As, S, Sb and possibly Pb and Cd will be lost during the evaporation stage. Open vessel digestions are also susceptible to airborne contamination. Closed vessel digestions or microwave digestions provide an alternative to open vessel digestions; however, while these techniques may limit loss of volatiles and airborne contamination and may provide a safer working environment, there is no real improvement in dissolution efficiency unless reaction temperatures exceed 200°C. This cannot be achieved with conventional microwave digestions.

### **Partial Digestions (Acid Leaching)**

Complete digestions of rock samples may not be necessary or desirable in some situations. Environmental analyses are typically concerned with bioavailable metals and nutrients. Bioavailable metals are defined by their ease of acid extraction, and are leached from a sample using strong mineral acids such as aqua regia (3:1 mixture of HCl and  $\text{HNO}_3$ ). Consequently there is no advantage to performing a complete digestion. Acid leaching has several advantages over complete mineral digestions. Loss of volatile elements is reduced as no evaporation stage is required, and the total dissolved solid concentration is reduced. This is important in determination of low-level analytes by ICP-MS where a high dissolved solid burden may require further dilution of the sample.

HCl should be avoided if the final analysis is to be performed by conventional ICP-MS or AA, particularly if the analytes of interest are transition elements, as the presence of the chloride ion in the analyte solution produces several persistent analytical interferences. Use of  $\text{HNO}_3$  alone in acid

leaching, however, results in low recovery of available elements. These problems may be circumvented by using cell technology ICP-MS systems where interference problems caused by chloride ions are dramatically reduced.

All partial digestions suffer from a lack of standardized methods and recognized reference materials that can be used to validate decomposition and analytical protocols. Where standards are available, care must be taken to follow the precise leach protocol described in the reference certificate.

### Fusion Techniques

Fusion decomposition provides an alternative digestion method. This is a high-temperature technique where powdered samples are heated with a suitable flux to produce a residue that may be readily dissolved. Fusion decompositions are the most rigorous digestions available and all silicate materials can be brought into a complete solution when fused with an appropriate flux. The principal disadvantage of the fusion technique is the introduction of extra salts into the final solution from the flux (thereby increasing total dissolved solids). Fusion decomposition remains the preferred technique for quantitative analysis of silicon, and may be the only practical method for complete decomposition of refractory minerals such as zircon, rutile, and cassiterite. Many fluxes have been used, but perhaps the most commonly used is lithium metaborate ( $\text{LiBO}_3$ ). This flux introduces only Li and B into the final solution, and is used at a relatively low flux:sample ratio of 3:1.

### Spatially Resolved Analyses

There is currently an explosion in the use of *in situ* analysis techniques, particularly techniques that decompose a small portion of sample by firing a pulse of laser light at the sample surface and transporting ablated particles directly into the analytical machinery (laser ablation). Such *in situ* techniques have the considerable advantage that there is no sample preparation required other than cutting, grinding and polishing the sample, and analyses can be targeted accurately at particular mineral grains or regions of a sample. The sensitivity available using laser ablation techniques is, however, considerably lower than that obtained using the equivalent analytical equipment with solution-based sample introduction. A second problem arises from the difficulty of standardizing the analyses. The amount of material ablated by the laser is a function of its power, the duration of the laser pulse, and the interaction of the laser with the sample surface. As most samples are heterogeneous, the quantity of sample introduced into the analytical

device may vary throughout an analysis. Some of these problems may be minimized through the choice of a suitable laser system. Lasers operating at wavelengths  $<200\text{ nm}$  induce ablation by breaking chemical bonds rather than heating, and this 'cold' ablation results in a reduction of interelement fractionation. The recovered signal is also relatively insensitive to differences in sample matrix allowing more reliable cross-calibration between sample types. Nonetheless, natural materials are generally heterogeneous over distances of a few micrometers, and the amount of sample delivered to the analytical device will vary in an unpredictable way. Effects of uneven ablation may be minimized by standardizing counts for the analyte of interest against those of an element with known and constant concentration throughout the sample. For instance, analyses of trace metal concentrations in ancient shells are often normalized to Ca concentrations within the shell calcite or aragonite. Variation in Ca concentration within the samples may be independently assessed through electron microprobe analysis. Clearly this is only possible in monomineralic samples. It is also difficult to achieve representative calibrations in LA work as calibrations should ideally be performed against matrix matched calibration standards. Few spatially homogeneous calibration standards exist, so calibration is often performed against synthetic glasses or pressed or fused powders, whose composition and ablation characteristics may be very different from the sample. Finally for the same reasons of sample heterogeneity, few standard reference materials are available to assess accuracy of laser ablation analyses. Despite these difficulties, the advantages offered by *in situ* analyses of trace level analytes in geological materials are numerous and LA-ICP-MS is rapidly becoming a mainstream technique in geochemical analysis.

### Analytical Techniques

A huge range of analytical techniques are used by geochemists and applied to geological problems. Some questions may be addressed using more than one piece of analytical equipment, whereas other applications demand specific machine capabilities. A list of the most commonly used analytical techniques is provided in Table 4.

#### Mass Spectrometric Techniques

Inductively coupled plasma mass spectrometry ICP-MS is arguably the work-horse in the modern geochemist's stable, and geological applications have driven many developments in ICP-MS

technology. ICP-MS has largely taken over from atomic absorption and atomic emission spectrometric techniques. Modern ICP-MS systems are capable of analyzing most elements in the periodic table simultaneously at levels of  $10 \text{ pg ml}^{-1}$  with a mass resolution of  $<1 \text{ amu}$ . The major disadvantage of ICP-MS is the formation of polyatomic interferences through interactions of the carrier gas, atmospheric gases and matrix ions. Polyatomic interferences are most problematic over the mass range 31–80, and because of this the main use of ICP-MS in geochemistry has been determination of low-level, high mass ( $>70$ ) elements. Recent development of collision and reaction cell technologies where matrix interferences are reduced or removed prior to detection, and high-resolution ICP-MS capable of resolving mass differences  $\ll 1 \text{ amu}$  have vastly increased the potential for ICP-MS in analysis of lighter, particularly transition, elements and their isotopes in complex geological and environmental matrices.

ICP-MS is easily coupled with laser ablation sample introduction systems giving spatially resolved analyses (smallest spot sizes around  $10 \mu\text{m}$ ) with working detection limits of  $\sim 1\text{--}10 \text{ mg kg}^{-1}$ . High-resolution ICP-MS may be coupled with LA introduction systems to provide high spectral resolution with rapid scanning across the entire mass range. This is particularly important when analyzing small sample volumes such as fluid inclusions, where the entire sample is rapidly consumed.

**Thermal ionization mass spectrometry (TIMS) and multicollector ICP-MS** The dating of geological materials requires highly accurate determination of isotopic abundances. This is generally achieved using a magnetic sector mass analyzer and faraday cup detection system. TIMS is commonly used to for radiometric dating and for investigations of mixing of water masses using (for instance) Sr, Nd, and U-series isotope systems.

Recently magnetic sector, multicollector ICP-MS has been used to determine isotope ratios with similar precision to TIMS. As production of the ion beam in ICP-MS is not dependent on ionization efficiency, a wider range of isotope systems may be analyzed. The development of multicollector ICP-MS has demonstrated the existence of significant isotopic variations among a wide range of materials, and many potential new geochemical and biogeochemical isotopic tracers are under investigation. The isotopic composition of Fe, Cu, Zn, and Ca in particular offer great potential as geochemical and biogeochemical tracers. Despite the problems associated with *in situ* analyses, multicollector ICP-MS may be coupled to a laser ablation system, as instabilities in analyte

supply caused by uneven ablation are offset by the rapid simultaneous detection in multiple faraday cups. Unlike high-resolution quadrupole ICP-MS, however, MC-ICP-MS systems are inherently incapable of rapidly scanning over a wide mass range. The application of MC-ICP-MS to the analysis of geological materials is in its infancy, and the field is expanding rapidly.

### X-Ray Fluorescence

XRF provides high precision analyses of major and minor elements in a wide variety of samples including rocks. XRF analyses are usually performed on pressed powder pellets of the ground sample, mixed with an organic binder. One distinct advantage of XRF analysis, therefore, is that samples do not need to be decomposed prior to analysis. This may compensate for the relatively high limits of detection of XRF ( $>1 \text{ mg kg}^{-1}$ ) for many elements, as samples need not be diluted. Quantitative determination of trace elements in rock samples by XRF requires calibration against matrix-matched geological reference materials. Fortunately a great variety of geological reference materials have been developed and a suitable reference material may be obtained for most rock types.

Synchrotron radiation also has many applications in geological sciences, particularly for determination of the local environment, geometry, and coordination of a specific element. Synchrotron sources can be used in XRF analysis, providing high-resolution elemental maps. Synchrotron-based X-ray investigations are frequently used to investigate the sorption of elements onto mineral or particle surfaces.

### Electron Probe Microanalysis

EPMA combines the imaging capability of a focused electron beam with the analytical potential afforded by induced X-rays to produce spatially resolved analyses of a wide range of elements with a limit of detection of  $\sim 100 \text{ mg kg}^{-1}$ . EPMA has many uses in geological science, particularly in mapping the spatial distribution of major and minor elements within solid samples. Samples are prepared by embedding in a resin or polymer matrix and polishing to yield a flat surface. Calibration is performed against mineral standards. The development of alternative spatially resolved analytical techniques with lower limits of detection has to some extent superseded the use of EPMA in geochemical analyses, but the combination of elemental mapping with high-resolution imaging within an electron microscope remains useful.



## Miscellaneous Techniques

Mineral identification may be performed using polarized light microscopy, but where possible single-crystal or powder X-ray diffraction is the technique of choice for determining mineral structure. Fourier transform infrared spectrometry (FTIR) is also commonly used to identify mineral structure and determine crystal perfection in geological materials.

## Standards and Standardization in Geochemical Analysis

The performance of any analytical procedure must be assessed through analysis of certified reference materials. As geological materials are varied, often with complex heterogeneous matrices, artificially produced standards are seldom sufficient. Homogenized, certified standard reference materials are available for a wide range of rock, soil, mineral, and biomaterial types and must be used to validate analytical protocols. Interlaboratory comparisons of measurements of the elemental and isotopic composition of reference materials are frequently published in dedicated journals such as the *Geostandards Newsletter*.

See also: **Archaeometry and Antique Analysis:** Dating of Artifacts; Metallic and Ceramic Objects. **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Inductively Coupled

Plasma. **Atomic Mass Spectrometry:** Laser Microprobe. **Environmental Analysis. Geochemistry:** Sediment; Soil, Organic Components. **Infrared Spectroscopy:** Overview. **Mass Spectrometry:** Stable Isotope Ratio; Archaeological Applications. **Microscopy Techniques:** X-Ray Microscopy. **Quality Assurance:** Reference Materials; Method Validation. **Sample Dissolution for Elemental Analysis:** Wet Digestion; Microwave Digestion. **Sampling:** Theory; Practice. **X-Ray Absorption and Diffraction:** X-Ray Diffraction – Powder. **X-Ray Fluorescence and Emission:** X-Ray Diffraction – Single Crystal; Wavelength Dispersive X-Ray Fluorescence; Energy Dispersive X-Ray Fluorescence; Synchrotron X-Ray Fluorescence.

## Further Reading

- Deer WA, Howie RA, and Zussman J (1966) *An Introduction to the Rock Forming Minerals*. Harlow, Essex: Longman.
- Faure G (1986) *Principles of Isotope Geology*. New York: Wiley.
- Fenter PA, Rivers ML, Strurchio NC, and Sutton SR (eds.) (2002) *Applications of Synchrotron Radiation in Low-Temperature Geochemistry and Environmental Science*, Reviews in Mineralogy and Geochemistry vol. 49. Washington, DC: Mineralogical Society of America.
- Gill R (ed.) (1997) *Modern Analytical Geochemistry*, Geochemistry Series. Harlow, Essex: Longman.
- Potts PJ, Bowles JFW, Reed SJB, and Cave MR (1995) *Microprobe Techniques in the Earth Sciences*, The Mineralogical Society Series, vol. 6. Dordrecht: Kluwer.

# Sediment

**G E Millward**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

This article covers general methods applicable to the physical and chemical analysis of sediments from lakes, rivers, estuaries, and coastal seas. The material fluxes from the land to the sea are dominated by the sediment loads delivered by the global rivers. Although the river sediment flux cannot be predicted accurately, it is thought to be in a range from 10 to 20 Gt a<sup>-1</sup>. Contributions by individual rivers are conditional upon the geology, topography, and climate of their drainage areas. Much of the sediment transported by rivers is temporarily stored

in the lower part of the river basin and often only reaches the ocean by slow migration through its estuary. Although there are compositional differences between sediments from these environments they generally comprise highly variable and complex mixtures of aluminosilicate particles, quartz grains, terrestrial organic detritus, planktonic material, and ferromanganese oxides. The surfaces of mineral phases, such as clays, are modified by coatings of organic matter and iron and manganese oxides. The various particle types in a sedimentary assemblage have unique surface properties that influence their chemical reactivity with dissolved constituents. Thus, particle–water interactions, which are accentuated in estuaries because of the sudden change in reaction conditions across the freshwater–seawater interface, give rise to complex array of adsorbed and absorbed species of inorganic and organic



## Miscellaneous Techniques

Mineral identification may be performed using polarized light microscopy, but where possible single-crystal or powder X-ray diffraction is the technique of choice for determining mineral structure. Fourier transform infrared spectrometry (FTIR) is also commonly used to identify mineral structure and determine crystal perfection in geological materials.

## Standards and Standardization in Geochemical Analysis

The performance of any analytical procedure must be assessed through analysis of certified reference materials. As geological materials are varied, often with complex heterogeneous matrices, artificially produced standards are seldom sufficient. Homogenized, certified standard reference materials are available for a wide range of rock, soil, mineral, and biomaterial types and must be used to validate analytical protocols. Interlaboratory comparisons of measurements of the elemental and isotopic composition of reference materials are frequently published in dedicated journals such as the *Geostandards Newsletter*.

See also: **Archaeometry and Antique Analysis:** Dating of Artifacts; Metallic and Ceramic Objects. **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Inductively Coupled

Plasma. **Atomic Mass Spectrometry:** Laser Microprobe. **Environmental Analysis. Geochemistry:** Sediment; Soil, Organic Components. **Infrared Spectroscopy:** Overview. **Mass Spectrometry:** Stable Isotope Ratio; Archaeological Applications. **Microscopy Techniques:** X-Ray Microscopy. **Quality Assurance:** Reference Materials; Method Validation. **Sample Dissolution for Elemental Analysis:** Wet Digestion; Microwave Digestion. **Sampling:** Theory; Practice. **X-Ray Absorption and Diffraction:** X-Ray Diffraction – Powder. **X-Ray Fluorescence and Emission:** X-Ray Diffraction – Single Crystal; Wavelength Dispersive X-Ray Fluorescence; Energy Dispersive X-Ray Fluorescence; Synchrotron X-Ray Fluorescence.

## Further Reading

- Deer WA, Howie RA, and Zussman J (1966) *An Introduction to the Rock Forming Minerals*. Harlow, Essex: Longman.
- Faure G (1986) *Principles of Isotope Geology*. New York: Wiley.
- Fenter PA, Rivers ML, Strurchio NC, and Sutton SR (eds.) (2002) *Applications of Synchrotron Radiation in Low-Temperature Geochemistry and Environmental Science*, Reviews in Mineralogy and Geochemistry vol. 49. Washington, DC: Mineralogical Society of America.
- Gill R (ed.) (1997) *Modern Analytical Geochemistry*, Geochemistry Series. Harlow, Essex: Longman.
- Potts PJ, Bowles JFW, Reed SJB, and Cave MR (1995) *Microprobe Techniques in the Earth Sciences*, The Mineralogical Society Series, vol. 6. Dordrecht: Kluwer.

# Sediment

**G E Millward**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

This article covers general methods applicable to the physical and chemical analysis of sediments from lakes, rivers, estuaries, and coastal seas. The material fluxes from the land to the sea are dominated by the sediment loads delivered by the global rivers. Although the river sediment flux cannot be predicted accurately, it is thought to be in a range from 10 to 20 Gt a<sup>-1</sup>. Contributions by individual rivers are conditional upon the geology, topography, and climate of their drainage areas. Much of the sediment transported by rivers is temporarily stored

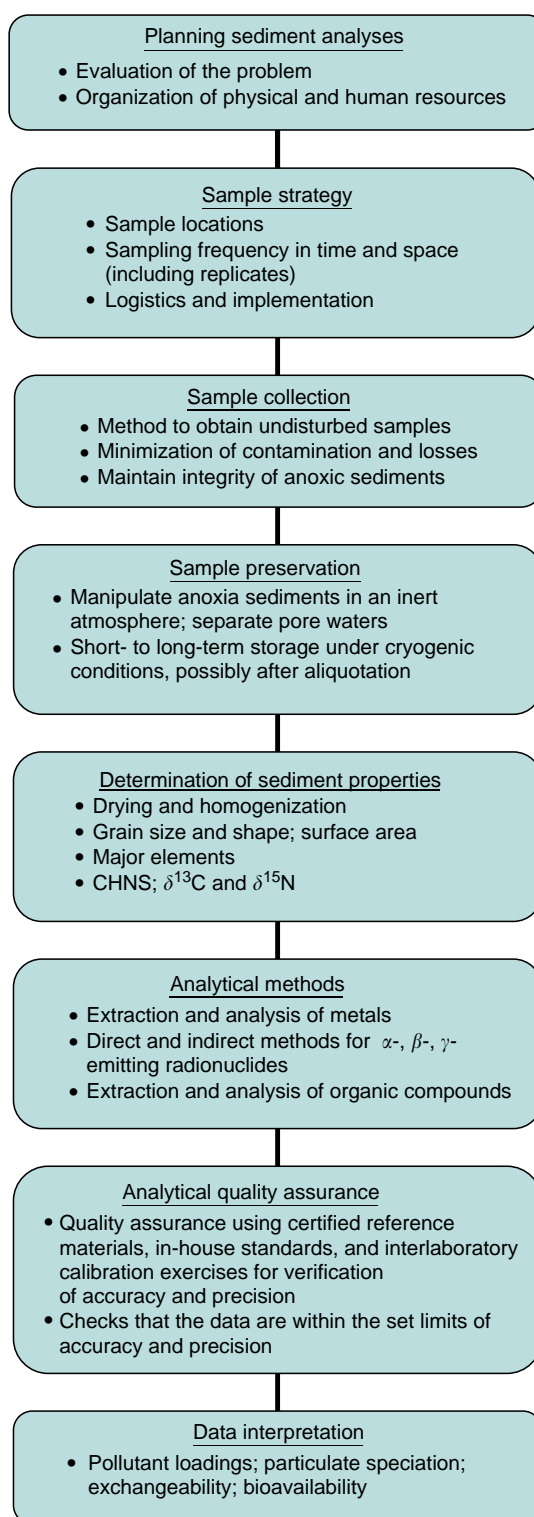
in the lower part of the river basin and often only reaches the ocean by slow migration through its estuary. Although there are compositional differences between sediments from these environments they generally comprise highly variable and complex mixtures of aluminosilicate particles, quartz grains, terrestrial organic detritus, planktonic material, and ferromanganese oxides. The surfaces of mineral phases, such as clays, are modified by coatings of organic matter and iron and manganese oxides. The various particle types in a sedimentary assemblage have unique surface properties that influence their chemical reactivity with dissolved constituents. Thus, particle–water interactions, which are accentuated in estuaries because of the sudden change in reaction conditions across the freshwater–seawater interface, give rise to complex array of adsorbed and absorbed species of inorganic and organic

constituents. Depending on their grain size, carbon content, and the length of time deposited at the site, consolidation of sediments may occur and anaerobic conditions may also develop in the sediment column, as a consequence of the bacterially moderated decomposition of particulate organic matter. Consequently, the focus of analytical methods for metals, radioisotopes (including metals, nonmetals, and metalloids) and persistent organic pollutants have become highly oriented to preserving, detecting, and quantifying the concentrations of particulate species in sediments and to establishing their relationship with basic particle properties.

From the applications viewpoint, chemical analyses of aquatic sediments are essential in the assessment of their contaminant loadings, in reconstructing the history of contaminant deposition, and in quantifying the rate at which sediment–water exchange processes contribute to recovery from contamination. The bioavailability of toxins in contaminated sediments has also become a priority for managers of aquatic systems. In an attempt to assess the impact of sediment chemistry and toxicity on the health of bottom-dwelling organisms a Sediment Quality Triad has been developed.

## Sediment Sampling and Preservation

There is little point in undertaking complex chemical analyses of sediment samples if they have been significantly modified during collection or contaminated during subsequent laboratory manipulation. Thus, sediment collection, preservation, and analysis require careful preplanning and execution, as shown in the hierarchy of operations in **Figure 1**. Sediments should be obtained in such a way as to retain their natural integrity, although some disturbance and losses of the fine fraction may occur because of the application of relatively crude sampling methods. These include heavy mechanical grabs and corers that are deployed from a vessel and whose operation requires them to impact with the seabed, causing shock waves and sediment disturbance. Large volume sediment samples, typically of 1 m depth, can be obtained by a spade-box corer from which undisturbed sediment subcores can be taken by carefully inserting plastic tubes. Multicorers are used to obtain several undisturbed samples at the same site, allowing assessment of small-scale variations in sediment composition to be made, especially in areas of bioturbation or variable sediment texture. In shallow waters, *in situ* collection



**Figure 1** Hierarchy of operations when conducting a sediment survey, sample preparation, and analysis.

by divers may involve less sample alteration, although care should be taken to ensure that movements near the seabed do not cause sediment resuspension.

Oxic surface sediment samples should be transferred to a suitable container and stored in a freezer as soon as possible after collection. Suitable containers will depend on the analyte, such that sediments for metal and/or radiochemical analyses are stored in acid-washed, plastic containers. Glass or metal containers should be used when storing sediments for the analysis of organic molecules, since plastic may contain similar organic compounds that may leach into the sample. Prior to use, glass or metal containers are normally washed repeatedly in an organic solvents, such as high-performance liquid chromatography (HPLC)-grade acetone and hexane and then baked at high temperature ( $\sim 450^{\circ}\text{C}$ ) overnight.

The preservation of the integrity of subsurface anoxic sediments, normally retained in a plastic core tube, is more taxing. Immediately on retrieval, the top of the sediment core tube is enclosed within a plastic glove bag filled with an inert gas, e.g., 'white spot' nitrogen and the core is extruded from the tube by applying a piston to its base. The sample can be sectioned, into slices of appropriate thickness, as the sediment core is extruded into the nitrogen-filled glove bag. The separation of pore waters from the sediments is essential since they may contain significant concentrations of redox-sensitive elements, which, if exposed to the atmosphere, may precipitate on the sediment particles thereby modifying their composition. Isolation of the pore waters is achieved by centrifugation using sealed, plastic centrifuge tubes flushed with an inert gas. After decanting the pore water under nitrogen the centrifuge tubes containing the anoxic sediments are stored in a freezer.

## Determination of Sediment Properties

The concentrations of adsorbed sedimentary constituents tend to increase with decreases in particle diameter owing to an enhancement of the surface area per unit weight. To account for regional diversity of sediment grain size distributions, separation of a fine fraction is undertaken by sieving through a mesh (grain size commonly  $< 63\ \mu\text{m}$ ) of plastic or metal, appropriate to the analyte. However, the fine fraction may form only a small part of the total sediment and given that sieving is time-consuming and does not always disperse aggregates, it may be preferable to carry out chemical analyses of the bulk sediments. Concentrations of the constituents can be normalized with respect to a property of the fine fraction, such as percentage weight, specific surface area (via BET nitrogen adsorption on a freeze-dried sample), or by reference to the total concentration of a major

element, such as Al or Li. The grain sizes of cohesive muds can be obtained from sedimentation tubes, generally known as Owen Tubes, or by laser granulometry (operating in the range  $2\text{--}1000\ \mu\text{m}$ ). High-resolution scanning and transmission electron microscopy yield details of particle morphology, including size and shape. If the instrument is fitted with an element dispersive X-ray analyzer, major elements (Al, Ca, Si, Fe) can be determined, thereby identifying differences in the detrital (i.e., the underlying solid phase) composition of individual particles as a function of particle size. Where mineral phases have been separated on the basis of density or magnetic properties, their mineralogy can be examined by X-ray diffraction methods.

The surfaces of sediment particles normally have deposits of natural organic macromolecules, such as humic and fulvic acids, and ferromanganese oxides. The organic macromolecules have functional groups that are effective sorbents for persistent organic pollutants and, together with the oxides, they are sites where metals undergo surface complexation reactions. The proton and cation exchange capacities of natural sediments have been obtained from acid-base titrations. Extraction of surface-bound humic (molecular weight  $\sim 10\,000\ \text{Da}$ ) and fulvic ( $< 5\,000\ \text{Da}$ ) compounds can be made using sodium hydroxide. The extracts can be concentrated on a resin column, separated by liquid chromatography, and detected by ultraviolet (UV) spectrophotometry or Fourier transform infrared spectrophotometry or by gas chromatography-mass spectrometry (GC-MS). Total carbon, hydrogen, nitrogen, and sulfur analyses are undertaken by using a commercially available combustometric technique and particulate organic carbon can be similarly determined after removal of inorganic carbon by acidification. Isotope ratio-mass spectrometry (IRMS) has been used to determine the mean  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  in sediments:

$$\delta^{13}\text{C} = [((^{13}\text{C}/^{12}\text{C})_{\text{sample}}/(^{13}\text{C}/^{12}\text{C})_{\text{PDB}}) - 1] \times 1000$$

$$\delta^{15}\text{N} = [((^{15}\text{N}/^{14}\text{N})_{\text{sample}}/(^{15}\text{N}/^{14}\text{N})_{\text{air}}) - 1] \times 1000$$

The equations give  $\delta^{13}\text{C}$  values (per mil or per thousand) using Chicago PDB carbonate as the standard and  $\delta^{15}\text{N}$  values (per mil or per thousand) with atmospheric nitrogen as the standard value. The  $\delta^{13}\text{C}$  values for individual compounds, albeit with some loss of sensitivity, can be determined by coupling IRMS with GC separation of sediment extracts. The combination of C:N ratios and  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values give vital information on the provenance of sediments since the values are significantly different for various particle types (Table 1). The determination of

**Table 1** Examples of C:N ratios and  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values for various particle types

Particle type	C:N ratio	$\delta^{13}\text{C}$ (per mil)	$\delta^{15}\text{N}$ (per mil)
Soil organic matter <sup>a</sup>	12–20	– 23 to – 29	+ 5
Sediments from stratified lakes <sup>a</sup>	6–14	– 20 to – 33	– 2 to + 20
Sediments from Forth Estuary, UK <sup>b</sup>	21.7 ± 6.6	– 24.0 ± 0.2	+ 5.6 ± 0.6
Marine sediment <sup>b</sup>	11.3	– 23.2	+ 5.9
Sewage <sup>b</sup>	12.6	– 26.7	+ 2.3
Micro-phytoplankton <sup>b</sup>	6.3	– 17.2 to – 26.5	+ 0.5 to + 8.2

<sup>a</sup>Last WM and Smol JP (eds.) (2002) *Tracking Environmental Change Using Lake Sediments: Volume 2, Physical and Chemical Methods*. Dordrecht: Kluwer Academic.

<sup>b</sup>Graham MC, Eaves MA, Farmer JG, Dobson J, and Fallick AE (2001) A study of carbon and nitrogen stable isotopes and elemental ratios as potential indicators of source and fate of organic matter in sediments of the Forth Estuary, Scotland. *Estuarine, Coastal and Shelf Science* 52: 375–380.

$^{18}\text{O}/^{16}\text{O}$  and deuterium/hydrogen ratios in sedimentary organic matter could be of advantage to paleolimnological studies concerned with the reconstruction of paleoclimate. However, there are considerable practical difficulties in the analysis of these stable isotopes and few laboratories currently have the required level of analytical skills.

## Chemical Analyses of Sediments

The determination of trace constituents in sediments requires that contamination be minimized during sample manipulation, which should be conducted in a clean room, or, at least, in a laminar flow hood. The laboratory-ware should be of good-quality borosilicate glass, polyethylene, or Teflon and high-purity reagents be used throughout to reduce the blanks. Normally, the analysis of sedimentary constituents involves a freeze-dried or air-dried sample that can be accurately weighed. If appropriate, samples may be homogenized by mortar and pestle and then sieved. Drying and chemical treatment of anoxic sediments should be undertaken in an inert atmosphere and with reagents that are oxygen-free.

## Metals

The analysis of metals by X-ray fluorescence has been widely used on geological and sediment samples, either deposited on filters or as thin films. The method can be made quantitative by using geological standards and transition metals can be determined in the 1–5  $\mu\text{g/g}$  range. The surfaces of sediment particles can be examined by the direct use of electron microprobe X-ray emission spectrometry and Auger electron spectroscopy. Although these methods are not particularly sensitive, they can allow the determination of a depth-profile of trace metals within a sediment particle.

Determinations of total metals in sediments can be achieved with electrothermal atomic absorption

spectrometry or multielement inductively coupled plasma-mass spectrometry (ICP-MS), following microwave assisted extraction (MAE) using concentrated  $\text{HNO}_3$  and HF in sealed Teflon vessels. The accuracy of the method can be verified by certified reference sediments, such as MESS-3 and PACS-2 available from the National Research Council of Canada (NRCC). High-resolution ICP-MS is also used for determinations of isotopes, for example,  $^{204}\text{Pb}$ ,  $^{206}\text{Pb}$ ,  $^{207}\text{Pb}$ , and  $^{208}\text{Pb}$  in studies of source apportionment. Direct analysis of solid samples may be possible and has a number of advantages, including reducing the preparation time, minimizing contamination, retention of volatile elements, and determination of highly refractory matrix-bound elements. Methods involve slurry nebulization ICP-MS using finely ground, fluidized sediments and laser ablation ICP-MS where solid samples are mixed with a binder and pelletized. Determinations of total metals allow geochemical normalization using concentrations of Al, since it is a major component of aluminosilicate minerals and it is usually unaffected by anthropogenic inputs. However, Li is a potentially superior normalizer because, unlike Al, it is not usually present in feldspars. An added advantage of geochemical normalization is that it allows estimation of enrichment factors.

Metals associated with various binding sites on sediments have been assessed using extraction procedures applied as single digests or as a set of sequential steps. Selective dissolution of trace metals from the particle surfaces is followed by determination using atomic absorption spectrometry (AAS), ICP-MS, or total reflection X-ray fluorescence. The use of sequential extraction schemes for the operational definition of metal species in sediments has proved contentious. They have been criticized on the basis that the reactions are not sufficiently phase selective and labile phases could be transformed during sample preparation, causing a marked reduction



**Table 2** A sequential extraction scheme for trace metal speciation in sediments. The first three steps relate to the standardized sequential extraction method applicable to BCR CRM 601

Fraction	Extractant	Extracted component
Exchangeable and carbonatic	0.11 mol l <sup>-1</sup> HOAc	Exchangeable ions and carbonates
Moderately reducible	0.1 mol l <sup>-1</sup> NH <sub>2</sub> OH–HCl acidify with HNO <sub>3</sub> to pH 2	Amorphous iron and manganese oxides
Sulfidic and organic	8.8 mol l <sup>-1</sup> H <sub>2</sub> O <sub>2</sub> and 0.02 mol l <sup>-1</sup> HNO <sub>3</sub> at pH 2, extracted with 1 mol l <sup>-1</sup> NH <sub>4</sub> OAc at pH 2 with HNO <sub>3</sub>	Sulfides and organic matter
Residual	Concentrated HNO <sub>3</sub> /HF	Detrital matrix

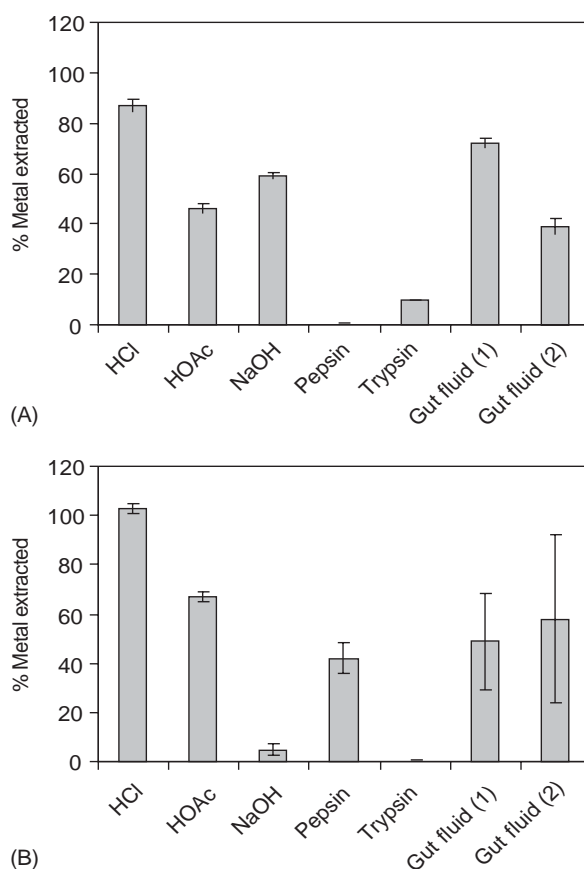
HOAc, acetic acid.

From Caruso J, Sutton KL, and Ackley KL (2000) *Elemental Speciation: New Approaches for Trace Element Analysis*. In: Barcelo D (ed.) *Comprehensive Analytical Chemistry*, vol. 33. Amsterdam: Elsevier.

in selectivity, particularly for anoxic sediments. Moreover, until recently there was no appropriate certified reference material (CRM), allowing international comparison of data. The Community Bureau of Reference (BCR, now renamed as the Standards, Measurements and Testing Programme) of the European Union has produced sediment reference material CRM 601 which can be used in the validation of a three-stage sequential extraction procedure (Table 2). In the case of anoxic sediments, methods have been developed for determination of acid volatile sulfides simultaneously with the extraction of metals, under oxygen-free conditions. This combined approach is useful in interpreting the data from bioassay tests of metal toxicants on aquatic sediments.

Innovative methods to determine the bioavailability of trace metal in estuarine sediments have been developed recently using digests that are more representative of the digestive fluids of aquatic organisms. Studies with trypsin and pepsin (two of the most important enzymes involved in the digestion of food by marine organisms) and the intestinal fluids extracted from codfish have shown significant variation from mineral reagents used in sediment extractions. Figure 2 shows the proportions of Cu and Zn released by single extractions by various reagents, relative to the concentration of total metal on estuarine sediments. Compared to Zn, Cu is associated with particulate organic material as shown by the amount released by the NaOH digest. Pepsin removed more Zn than Cu, whereas the gut fluids (1 and 2) solubilized more Cu and Zn than the enzymes pepsin and trypsin. It appears that acetic acid represents the best reagent with which to define the metals that are available under gastrointestinal conditions. However, more research is required on the extraction efficiency of other enzymes, such as proteinase-K, which targets metals associated with proteinaceous material on sediments.

Butylated Sn compounds have been determined in sediments by exploiting microwave technology to simultaneously extract and derivatize Sn compounds



**Figure 2** Percentage metal extracted, relative to the total concentration, from marine sediments by different reagents, enzymes, and the gut fluid from cod. HCl = 1 mol l<sup>-1</sup> HCl; HOAc = 25% glacial acetic acid; NaOH = 0.1 mol l<sup>-1</sup> NaOH; Pepsin = Pepsin a pH 2; Trypsin = Trypsin II-S pH 7.6; Gut fluid (1) = natural filtered intestinal fluid, pH 7; Gut fluid (2) = natural filtered intestinal fluid, pH 2. (A) Cu and (B) Zn. (Adapted from Turner A and Olsen YS (2000) Chemical versus enzymatic digestion of contaminated estuarine sediment: Relative importance of iron and manganese oxides in controlling trace metal bioavailability. *Estuarine, Coastal and Shelf Science* 51: 717–728.)

(usually by ethylation) to give volatile tetra-substituted species, suitable for separation by GC. Methyl mercury species have been released from sediments by using KOH/methanol mixtures in sealed Teflon



reactors with an open focused MAE system. Subsequently, the methylated Hg species are extracted into a suitable organic solvent, separated by HPLC and the individual methylated compounds decomposed by UV irradiation as they come off the column. The pulses of inorganic Hg are then detected by cold vapor-atomic fluorescence spectrometry. The accuracy of analytical methods for Sn and Hg can be verified with certified reference sediments for mono-, di-, and tributyl tin (BCR CRM 646) and for methyl mercury (BCR CRM 580). Arsenic species on sediments (including the methylated forms) have been determined by coupled HPLC-ICP-MS following their extraction in phosphoric acid using a low-power MAE system.

### Radionuclides

Radionuclides such as  $^{238}\text{Pu}$ ,  $^{234}\text{U}$ , and  $^{232}\text{Th}$  emit  $\alpha$ -particles that have a limited penetration in matter, requiring a complex separation and preconcentration of the individual elements from sediments. Alpha-emitting radionuclides can be separated from sediments by a primary digestion with concentrated mineral acids (e.g., aqua regia) followed by several steps involving precipitation and ion exchange. Chemical yields for the various elements can be obtained using internal spikes that are not present in natural sediments (e.g.,  $^{236}\text{Pu}$ ) and the accuracy of the analyses can be verified with CRMs, such as International Atomic Energy Agency (IAEA)-367 sediment for  $^{239(240)}\text{Pu}$ . Thin sources are prepared, for example, by electrodeposition or by microprecipitation with  $\text{NdF}_3$  onto polyethylene filters (pore size  $0.1\ \mu\text{m}$ ), prior to  $\alpha$ -spectrometry, using silicon surface barrier detectors coupled with pulse height analysis by a multichannel analyzer. More progress has been made in the use of nonradiometric methods for  $\alpha$ -emitting radionuclides at natural concentrations (e.g.,  $^{232}\text{Th}$ ,  $^{235}\text{U}$ ,  $^{239}\text{Pu}$ ,  $^{240}\text{Pu}$ ) using acid digests of sediments followed by ICP-MS determination.

Pure  $\beta$ -emitters such as tritium ( $^3\text{H}$ ),  $^{14}\text{C}$  (radio-carbon widely used in geochronology),  $^{35}\text{S}$ ,  $^{63}\text{Ni}$ , and

$^{90}\text{Sr}$  are chemically separated from the sediments before liquid scintillation counting. Tritiated molecules and  $^{14}\text{C}$  compounds have been separated from freeze-dried sediment samples using a computer-controlled, combustion technique with a  $\text{CuO}$  catalyst. The radioactive gases,  $^{14}\text{CO}_2$  and  $^3\text{HHO}$ , are separated by selective trapping and each isotope is determined by liquid scintillation counting. Limits of detection for a 20 g sample are  $\sim 20\ \text{mBq g}^{-1}$  for both isotopes. Activity concentrations for  $\beta$ -emitters can be determined, using thin sources with a  $^{40}\text{K}$  standard, or for solutions via a liquid scintillation counter or for low levels using a gas-flow multicounter.

Gamma rays from some radionuclides, such as  $^{137}\text{Cs}$ ,  $^{241}\text{Am}$ , and  $^{106}\text{Ru}$ , have a specific energy that can be detected by nondestructive gamma spectrometry using sensitive hyper-pure germanium detectors, with relative efficiencies of 20–25%. The homogenized, freeze-dried sediment is normally placed in a Marinelli beaker, shaped to surround the detector, and of 0.8 l capacity. Sediment geochronology may be determined from the gamma emission of  $^{210}\text{Pb}$  (half-life 22 years) and  $^{226}\text{Ra}$ , from which the excess  $^{210}\text{Pb}$  can be calculated and the sediments dated. Dating of sediment cores is also assisted by peaks in the activity concentration of  $^{137}\text{Cs}$ , e.g., the peak generated by deposition from the Chernobyl accident in 1986. The gamma-emitters  $^7\text{Be}$  (half-life 53 days) and  $^{234}\text{Th}$  (half-life 24 days) are both particle-reactive and, since they have different source functions, they can be applied to the tracing of sediment migration in estuaries.

Activity concentrations for various gamma-emitting radionuclides in coastal marine sediments from the Irish sea are given in Table 3. A major source of radionuclides to the Irish Sea is the Sellafield nuclear reprocessing plant and the distribution of radionuclides shows that higher concentrations may be found away from the source, thereby emphasizing the point concerning particle properties because the higher activity concentrations are associated with the finer sediments either side of the source.

**Table 3** Mean activity concentrations<sup>a</sup> of gamma-emitting radionuclides in sediments from the coastal region of Cumbria, UK

Location	Sediment type	Activity concentrations ( $\text{Bq kg}^{-1}$ )			
		$^{241}\text{Am}$	$^{60}\text{Co}$	$^{137}\text{Cs}$	$^{106}\text{Ru}$
St. Bees	Sand	180	4.5	87	<4.0
Sellafield	Sand	230	5.2	90	<4.6
Ravenglass	Mud	1500	64	610	190
Ravenglass	Mud and sand	500	26	210	51

<sup>a</sup>Becquerels per kilogram dry weight.

From Centre for Environment, Fisheries, and Aquaculture Science (2002) *Radioactivity in Food and the Environment, 2001*. London: Food Standards Agency and Scottish Environment Protection Agency.

## Organic Compounds

**Polychlorinated organic compounds** Polychlorinated biphenyls (PCBs) comprise a family of 209 individual compounds, referred to as congeners, and to make identification easier systematic numbering was introduced by the International Union for Pure and Applied Chemistry (IUPAC). For example, toxicologically important PCB congeners notably 3,3,4,4'-tetrachlorobiphenyl, 2,3,3',4,4'-pentachlorobiphenyl, and 3,3',4,4',5-pentachlorobiphenyl have the IUPAC numbers 77, 105, and 126, respectively. Polychlorinated dibenzodioxins (PCDDs) have 75 congeners and polychlorinated dibenzofurans (PCDFs) have 135 congeners, although there is no numbering system. PCBs and PCDDs/PCDFs are found at low concentrations in sediments (in the range ng per g to pg per g) and co-contaminants have to be eliminated from the samples prior to analysis, which requires considerable analytical expertise to identify the compounds on a congener-specific basis.

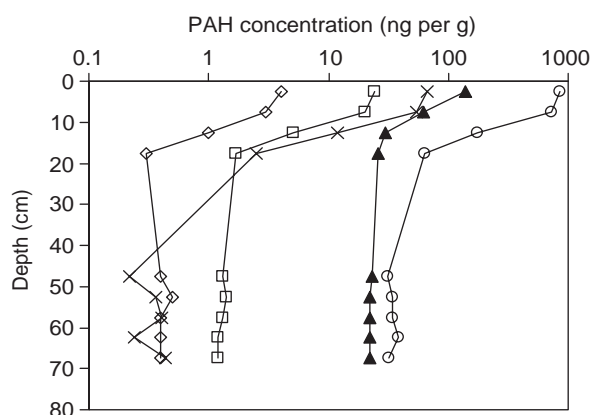
Due to the wide range of organic compounds in sediments, a three-stage analytical protocol is generally adopted:

1. Extraction of the organic compounds from a dried sediment sample: Prior to extraction an internal standard mixture of  $^{13}\text{C}$ -labeled compounds is added to the sediments. The extraction is normally carried out in a Soxhlet apparatus with an azeotropic mixture of organic solvents such as hexane and acetone or benzene and toluene. The extract is then concentrated using rotary evaporation and dried if appropriate.
2. Cleanup of the extract: This step involves the removal of other organic components that have been coextracted in stage (1) and would interfere with subsequent analyses. This stage is usually carried out by using column chromatography, thereby allowing the various homologs to be eluted and separated. This process may need to be repeated on an iterative basis to ensure complete removal of interfering compounds.
3. Separation, identification, and quantification: Using GC-MS with internal quantification standards that are added as spikes at the extraction stage. If the subsequent GC-MS analyses reveal that accurate determination of PCDDs and PCDFs is not possible owing to the presence of interfering peaks then liquid chromatography separation may be employed.

Developments in high-resolution GC and/or MS allow the detection of ultralow concentrations,  $\sim 10^{-15}$  g, on-column. However, any increased sensitivity brings additional problems and emphasizes the need to control contamination in reagents and glassware. Considerable improvements have been made in the precise quantification of many

compounds by isotope dilution-MS because of the availability of a wider range of  $^{13}\text{C}$  and deuterated compounds of organic molecules. Advances are being made in the coupling of instrumental techniques such as HPLC with MS, which will yield a highly sensitive method for the detection and quantification of more polar and water-soluble compounds. A novel approach to the determination of broad classes of chlorine-, bromine-, and iodine-containing organic compounds in sediments, extracted into a mixture of methylene chloride and hexane, involves monitoring the gamma-ray emissions of  $^{38}\text{Cl}$ ,  $^{80}\text{Br}$ , and  $^{128}\text{I}$ , following neutron activation of the extracts.

Determinations of polyaromatic hydrocarbons (PAHs) in freeze-dried sediments (sample mass of the order 10–20 g) has been carried out where hexamethylbenzene added as a standard to the sediment prior to extraction. The sediments were extracted sequentially with methanol followed by methylene chloride and substituted and unsubstituted were determined by GC-MS. The accuracy of the method was verified using the marine reference material HS6 (NRCC). The technique allowed the determination of 17 unsubstituted and 33 substituted PAHs with detection limits in the range 50–200 ng per g. An example of the results from the sediment core from a Canadian fjord contaminated with PAHs derived from an aluminum smelter is shown in Figure 3. The PAH concentration illustrates the wide concentration range available to this analytical method and also show the variation of PAH concentrations with depth. Each PAH has an increase in concentrations



**Figure 3** Concentrations of PAHs in a sediment core from Kitimat Arm, Douglas Channel, British Columbia.  $\circ$  – total unsubstituted PAHs;  $\blacktriangle$  – unsubstituted perylene;  $\square$  – phenanthrene;  $\times$  – fluoranthrene;  $\diamond$  – 2-methylphenanthrene. (Adapted from Simpson CD, Harrington CF, Cullen WR, Bright DA, and Reimer KJ (1998) Polycyclic aromatic hydrocarbon contamination in marine sediments near Kitimat, British Columbia. *Environmental Science and Technology* 32: 3266–3272.)

between 20 and 45 cm depth, indicating the beginning of Al smelting.

Surfactant compounds that sorb onto sediments after discharge from sewage works are thought to be endocrine disrupters and are, therefore, of considerable interest to environmental managers. One of the largest classes of commercially available nonionic surfactants are the nonyl phenol-*n*-ethoxylates defined by the abbreviation Np*n*EO, where *n* is the number of ethoxylate groups that normally reach a maximum of 9, giving rise to a large number of isomers. Biodegradation of Np*n*EOs takes place during sewage treatment and the more lipophilic and persistent alkylphenols are formed. Current methods involve the extraction of freeze-dried sediments, ground to a free-flowing powder, with hexane/acetone mixtures, using an accelerated solvent extraction system where the sample is held at 100°C and 1500 psi. After the extract is reduced to dryness, it is reconstituted with hexane and the extract cleaned up on a solid-phase extraction cartridge. The final extracts can be analyzed by coupled normal liquid phase chromatography–electrospray mass spectrometry. Detection limits are in the low ng per g range. Shang *et al.* in 1999 found that the average concentrations for total NpEO in the sediments of the Straits of Georgia, Canada were 1500 ± 130 ng per g; whereas in the Venice Lagoon the range was 500–6700 ng per g (total of Np, Np1EO, and Np2EO), and in marine sediments near a Barcelona sewage outfall the range was 100–6600 ng per g. Concentrations of the breakdown product octylphenol were typically <20 ng per g.

See also: **Analytical Reagents:** Purification. **Arsenic.** **Atomic Absorption Spectrometry:** Electrothermal. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Dioxins.** **Elemental Speciation:** Overview. **Endocrine Disrupting Chemicals.** **Geochemistry:** Sediment. **Humic and Fulvic Compounds.** **Isotope Dilution Analysis.** **Mass**

**Spectrometry:** Overview. **Mercury.** **Microscopy Techniques:** Scanning Electron Microscopy; X-Ray Microscopy. **Particle Size Analysis.** **Polychlorinated Biphenyls.** **Polycyclic Aromatic Hydrocarbons:** Environmental Applications. **Radiochemical Methods:** Overview. **Sample Handling:** Sample Preservation. **Sampling:** Theory. **Surface Analysis:** Auger Electron Spectroscopy. **Tin.** **X-Ray Absorption and Diffraction:** Overview. **X-Ray Fluorescence and Emission:** Energy Dispersive X-Ray Fluorescence; Particle-Induced X-Ray Emission.

## Further Reading

- Chapman PM and Wang FY (2001) Assessing sediment contamination in estuaries. *Environmental Toxicology and Chemistry* 20: 3–22.
- Evans EH, Dawson JB, Fisher A, *et al.* (2001) Atomic spectrometry update. Advances in atomic emission, absorption, and fluorescence spectrometry, and related techniques. *Journal of Analytical Atomic Spectrometry* 16: 672–711.
- Gill R (ed.) (1997) *Modern Analytical Geochemistry*. Harlow, UK: Pearson Education.
- Jickells TD and Rae JE (eds.) (1997) *Biogeochemistry of Intertidal Sediments*. Cambridge: Cambridge University Press.
- Jones KC and de Voogt P (1999) Persistent organic pollutants (POPs): State of the science. *Environmental Pollution* 100: 209–211.
- Morrow RW and Crain JS (eds.) (1998) *Applications of Inductively Coupled Plasma-Mass Spectrometry to Radionuclide Determinations*. Chelsea, MI: American Society for Testing and Materials.
- Pawliszyn J (2002) *Sampling and Sample Preparation for Field and Laboratory*. In: Barcelo D (ed.) *Comprehensive Analytical Chemistry*, vol. 37. Amsterdam: Elsevier.
- Shang DY, MacDonald RW, and Ikononou MG (1999) Persistence of nonylphenol ethoxylate surfactants and their primary degradation products in sediments from near a municipal outfall in the Strait of Georgia, British Columbia, Canada. *Environmental Science and Technology* 33: 1366–1372.
- Turner A and Millward GE (2002) Suspended particles: their role in biogeochemical cycles. *Estuarine, Coastal and Shelf Science* 55: 857–883.

## Soil, Major Inorganic Components

**H van der Jagt**, KIWA Water Research, Nieuwegein, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Major inorganic compounds, such as trace elements, are of concern as contaminants of terrestrial and

aquatic systems because of their persistence and toxicity at low concentrations. There is considerable evidence that the bioavailability and toxicity of such trace elements are markedly influenced by the physicochemical forms in which they are present in waters, in sediments, and in soils. Specifically, concerns about trace elements relate to utilization, disposal, and discharge of sewage and wastes. With regard to aquatic systems, complexation reduces the toxicity of

between 20 and 45 cm depth, indicating the beginning of Al smelting.

Surfactant compounds that sorb onto sediments after discharge from sewage works are thought to be endocrine disrupters and are, therefore, of considerable interest to environmental managers. One of the largest classes of commercially available nonionic surfactants are the nonyl phenol-*n*-ethoxylates defined by the abbreviation Np*n*EO, where *n* is the number of ethoxylate groups that normally reach a maximum of 9, giving rise to a large number of isomers. Biodegradation of Np*n*EOs takes place during sewage treatment and the more lipophilic and persistent alkylphenols are formed. Current methods involve the extraction of freeze-dried sediments, ground to a free-flowing powder, with hexane/acetone mixtures, using an accelerated solvent extraction system where the sample is held at 100°C and 1500 psi. After the extract is reduced to dryness, it is reconstituted with hexane and the extract cleaned up on a solid-phase extraction cartridge. The final extracts can be analyzed by coupled normal liquid phase chromatography–electrospray mass spectrometry. Detection limits are in the low ng per g range. Shang *et al.* in 1999 found that the average concentrations for total NpEO in the sediments of the Straits of Georgia, Canada were 1500 ± 130 ng per g; whereas in the Venice Lagoon the range was 500–6700 ng per g (total of Np, Np1EO, and Np2EO), and in marine sediments near a Barcelona sewage outfall the range was 100–6600 ng per g. Concentrations of the breakdown product octylphenol were typically <20 ng per g.

See also: **Analytical Reagents:** Purification. **Arsenic.** **Atomic Absorption Spectrometry:** Electrothermal. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Dioxins.** **Elemental Speciation:** Overview. **Endocrine Disrupting Chemicals.** **Geochemistry:** Sediment. **Humic and Fulvic Compounds.** **Isotope Dilution Analysis.** **Mass**

**Spectrometry:** Overview. **Mercury.** **Microscopy Techniques:** Scanning Electron Microscopy; X-Ray Microscopy. **Particle Size Analysis.** **Polychlorinated Biphenyls.** **Polycyclic Aromatic Hydrocarbons:** Environmental Applications. **Radiochemical Methods:** Overview. **Sample Handling:** Sample Preservation. **Sampling:** Theory. **Surface Analysis:** Auger Electron Spectroscopy. **Tin.** **X-Ray Absorption and Diffraction:** Overview. **X-Ray Fluorescence and Emission:** Energy Dispersive X-Ray Fluorescence; Particle-Induced X-Ray Emission.

## Further Reading

- Chapman PM and Wang FY (2001) Assessing sediment contamination in estuaries. *Environmental Toxicology and Chemistry* 20: 3–22.
- Evans EH, Dawson JB, Fisher A, *et al.* (2001) Atomic spectrometry update. Advances in atomic emission, absorption, and fluorescence spectrometry, and related techniques. *Journal of Analytical Atomic Spectrometry* 16: 672–711.
- Gill R (ed.) (1997) *Modern Analytical Geochemistry*. Harlow, UK: Pearson Education.
- Jickells TD and Rae JE (eds.) (1997) *Biogeochemistry of Intertidal Sediments*. Cambridge: Cambridge University Press.
- Jones KC and de Voogt P (1999) Persistent organic pollutants (POPs): State of the science. *Environmental Pollution* 100: 209–211.
- Morrow RW and Crain JS (eds.) (1998) *Applications of Inductively Coupled Plasma-Mass Spectrometry to Radionuclide Determinations*. Chelsea, MI: American Society for Testing and Materials.
- Pawliszyn J (2002) *Sampling and Sample Preparation for Field and Laboratory*. In: Barcelo D (ed.) *Comprehensive Analytical Chemistry*, vol. 37. Amsterdam: Elsevier.
- Shang DY, MacDonald RW, and Ikononou MG (1999) Persistence of nonylphenol ethoxylate surfactants and their primary degradation products in sediments from near a municipal outfall in the Strait of Georgia, British Columbia, Canada. *Environmental Science and Technology* 33: 1366–1372.
- Turner A and Millward GE (2002) Suspended particles: their role in biogeochemical cycles. *Estuarine, Coastal and Shelf Science* 55: 857–883.

## Soil, Major Inorganic Components

**H van der Jagt**, KIWA Water Research, Nieuwegein, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Major inorganic compounds, such as trace elements, are of concern as contaminants of terrestrial and

aquatic systems because of their persistence and toxicity at low concentrations. There is considerable evidence that the bioavailability and toxicity of such trace elements are markedly influenced by the physicochemical forms in which they are present in waters, in sediments, and in soils. Specifically, concerns about trace elements relate to utilization, disposal, and discharge of sewage and wastes. With regard to aquatic systems, complexation reduces the toxicity of



dissolved trace elements to a range of aquatic organisms and trace element uptake and toxicity often relate best to the free metal ion activity. In general, systematic differences in the geochemistry of groundwater evaluation in aquifers are traceable to local solid variability. For instance, variations in Sr isotopes and Mg/Ca and Sr/Ca ratios offer insight into the influence of soils on groundwater geochemistry, sources of dissolved constituents in groundwater, water-rock interaction pathways, and groundwater residence time.

The chemical composition of water in relation to the Earth's crust constitutes the focus of the science called geohydrochemistry or simply geochemistry. Water quality is one of the primordial environmental factors determining which flora and fauna will thrive and which material will dissolve or precipitate. Environmental research helps gain insight into aspects such as economics, sensitivity, and relevance in pollution studies: economics for the possibility of cost-effective direct measurement without complicated extractions prior to analysis, sensitivity because of an analytically more agreeable result when compared to the solid phases in soil, and relevant on account of a direct selection of the mobile phases.

## Geochemistry of Soil

The main objectives of the geochemistry of soil are based on the aids required in the solution of specific hydrological problems, e.g., the determination of the origin of groundwater by recognition through specific natural tracers and visualization of actual patterns by mapping groundwater of various origins and characteristics. Geological formations, soil classes, or flow systems are discerned on the basis of their genesis.

### Scientific Issues

Experiments during artificial recharge are ideal for comparison between laboratory experiments and field studies. They unify the advantages of both: most experimental conditions are well defined, the scale of operation is large enough to contain the natural heterogeneity, and the timescale is also sufficient for the detection of specific trends. The scientific issues involved are generally studying the kinetics of denitrification, pyrite oxidation, the oxidation of organic matter, and the quantification of the reactive fraction of organic matter and of the trace element content of pyrite.

The main processes in soils demonstrate oxic influents (rainwater, surface water) and anoxic soil containing pyrite resulting in mean order of increasing duration: displacement of native groundwater, cation exchange, pyrite oxidation, acid buffering by

calcite dissolution, and oxidation of organic matter. Differences in quality changes between different sites relate to their deviating aquifer geochemistry and influent quality.

## Geochemical Monitoring

Water quality monitoring consists of frequent analysis of the main constituents. The required data input consists of: (1) mean composition of the influent; (2) mean composition of native groundwater in each layer of the target aquifer; (3) native geochemistry of each layer of the target aquifer; (4) the cumulative frequency curve of detention times in each model layer or flow path as derived from either separately run hydrological model or tracer breakthrough data; and (5) specific information derived from the mass balance of the water phase (the reactions that are needed, how  $O_2$  and  $NO_3^-$  distribute over the various redox reactions, etc.).

The results demonstrate the following effects: (1) displacement of the native groundwater by the influent, including effects of dispersion; (2) leaching of reactive aquifer constituents: exchangeable cations, calcite, organic matter iron sulfides and  $MnO_2$ ; and (3) breakdown of organic micropollutants.

With regard to chemical reactions, target aquifer samples have been determined to give information on grain size distribution and porosity, and quantitative chemical analysis for total element content (by X-ray fluorescence (XRF)), iron sulfides, calcium carbonate, exchangeable cations, organic matter, and organic carbon.

## Geochemical Composition

The main chemical composition of an aquifer is deep anoxic, mainly quartz sand, poor in organic carbon, containing a small amount of pyrite and calcite. Cation-exchange capacity (CEC) is low, dominated by  $Ca^{2+}$  and  $Mg^{2+}$ . The top of the layer contains more organic carbon, calcite, and pyrite.

### Organic Compounds

Due to variations in the soil, such as humic constituents, sand fractions, clay minerals, peat minerals, wood rests, algal toxins, and deposits of shell lime, different reactions take place. The natural organic matter (NOM) in an aquifer is a complex mixture of substances, such as humic acids, fulvic acids, amino acids, carbohydrates, lipids, lignin, and waxes.

Dissolved organic carbon and color become progressively reduced upon aquifer passage probably by degradation in presence of oxygen, nitrate,



and sulfate (organic compounds react in these cases like a redox buffer). Oxidation of organic compounds leads to dissolving of organic compounds (fulvic acid, dissolved organic carbon, inorganic carbon and nutrients, iodide, nitrogen, hydrogen sulfide, and methane). The reactions are important with regard to strong absorbent working formations.

### **Carbonate Minerals**

Calcium carbonate (such as calcite, siderite, dolomite, and aragonite) is typically present in marine deposits and Pleistocene sediments deposited in riverbanks and practically absent in deposits from ancient rivers and deposits of Miocene age. Calcium is generally the dominant exchangeable cation (56–82%) followed by magnesium (6–26%). An important parameter for behavior during aquifer transport is the calcite saturation index ( $SI_c = \log[Ca^{2+}][CO_3^{2-}]/K_{CaCO_3}$ ).

### **Sulfide Minerals**

The geochemical heterogeneity of the aquifer mainly consists of a higher reactivity of pyrite and elementary sulfur in the lower zone and a higher reactivity of organic matter in the upper parts. Pyrite oxidation is relatively fast in the presence of oxygen and nitrate (reaction nitrate in nitrogen). During the oxidation of pyrite elemental sulfur, sulfate, iron, and trace elements (heavy metals) will be mobilized.

### **Hydroxide Minerals**

The presence of iron hydroxides, manganese oxides, and aluminum hydroxides (gibbsite and jurbanite) in soils will be mobilized under anoxic or acid ( $pH < 5$ ) environmental conditions.

### **Clay Minerals**

Illite is the most dominant mineral present in soil samples. Apart from this mineral even smectite, kaolite, and chlorite will be present in the soil. Illite absorbs, in practice, some potassium, ammonium, cesium, and rubidium species. The solubility of clay mineral is quite high under acidic ( $pH < 4$ ) environmental conditions.

## **General Overview of Analytical Techniques**

In this overview, techniques highlighted are those that, based on our experience, represent current trends and state-of-the-art technologies in elementary analysis, and specific extractions in real environmental samples. Also reviewed are relatively new analyses of interest and detections that are made possible by recently developed analytical methods.

### **Elemental Analysis**

This technique comprises a group of quantitative analytical methods. Iron, carbon, and calcium will be determined. It is necessary to use strong techniques such as microwave to digest the soil samples completely. The whole soil sample will not be completely reactive in all cases, only a part will be reactive.

### **Selective Extractions**

The main reason for determining a specific constituent like calcite, organic carbon, and pyrite or a selective bounding form of elements is the necessity to perform selective extractions. Solid-phase extraction of (in)organic compounds continues to be the leading technology for the exhaustive extraction of selective compounds from soil samples.

### **Sequential Extractions**

The solubility of the specified constituent can be determined by sequential (step by step) extractions. An advantage of these step-by-step performances is the availability of information on individual groups of oxides, carbonates, sulfides, and silicates. One disadvantage is the high cost involved.

### **Other Techniques (Mostly Semiquantitative)**

Nuclear magnetic resonance spectroscopy has emerged as a powerful technique to explore analyte–matrix interactions. Another powerful technique is scanning electron microscopy, which provides details on morphology and mineral variations of the soil. The resolution is at a microlevel ( $\pm 5 \mu m$ ).

Complex reactive transport models are integrated with geophysical information systems applied to radionuclides. The behavior of  $^{90}Sr$ ,  $^{60}Co$ ,  $^{106}Ru$ , and  $^{137}Cs$  can be used for breakthrough purposes. Short pulses (of 1–28 days) with 100% activity can be chosen as input signal in order to simulate a calamity at the recharge facilities.

## **Geochemical Analyses**

Most studies were inspired by questions regarding the fate of pollutants upon detection in spreading basins and upon passage of aquifer systems, before recollection, post-treatment, and distribution as drinking water.

Since there already exist several review articles, highlights of the most important methods and comments on the different principles will be described. Variations in performing analytical methods especially for calcite, organic carbon, pyrite, sulfides, and CEC reaction have resulted in the development of standard methods. In **Figures 1–4**, the most common analytical

methods are described with linking pins to their comprehensive approaches based on numerous trends.

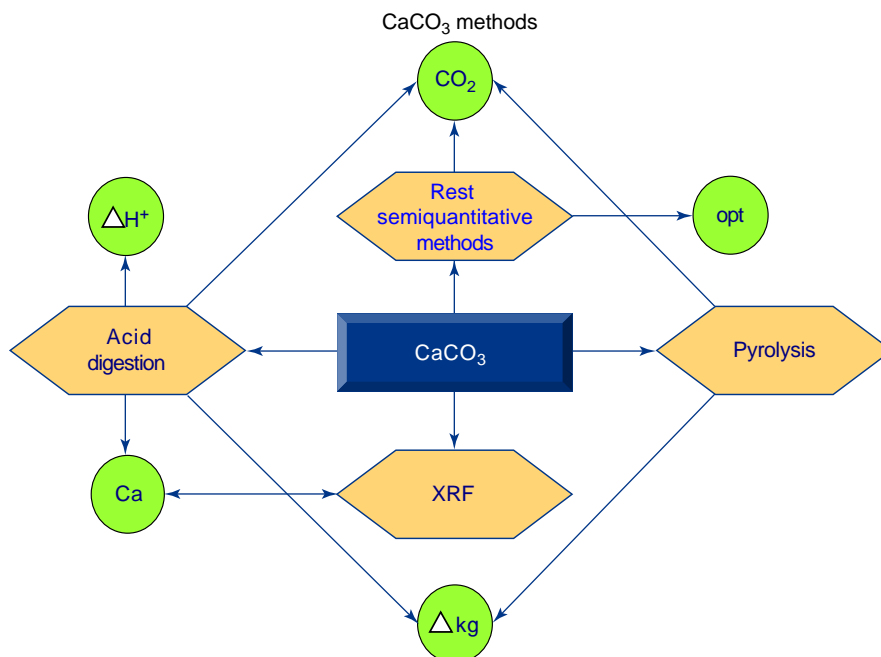
### Calcium Carbonate (Calcite)

Quantitation provides the highest accuracy level achievable by  $\text{CaCO}_3$  methods. The digestion of calcium carbonate can be determined in four different ways: (1) digestion with acid; (2) thermal gravimetric analysis (TGA at  $105^\circ\text{C}$ ,  $450^\circ\text{C}$ ,  $550^\circ\text{C}$ , and  $1000^\circ\text{C}$

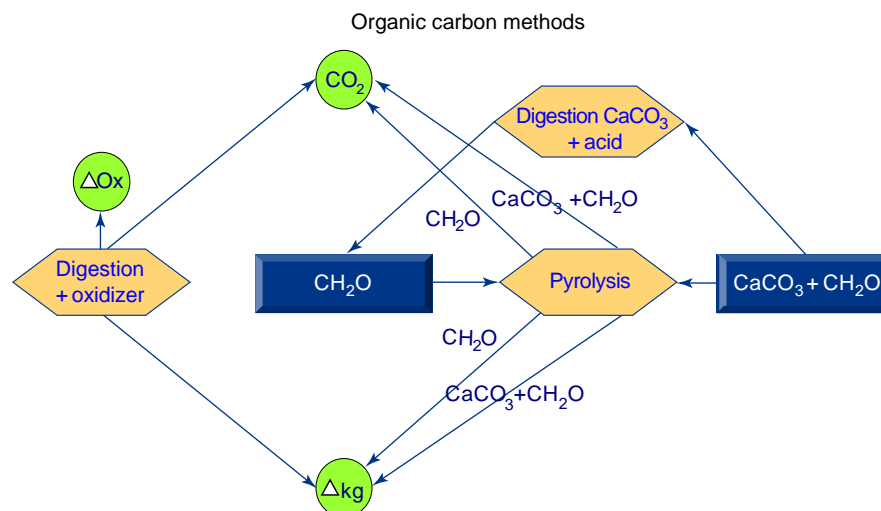
on bulk organic material); (3) XRF; and (4) semiquantitative methods (granular, or ion-exchange complex and CEC, or  $\text{CO}_2$  pyrolysis with acid addition). In some cases, postanalysis calculations correct overestimation for the presence of exchangeable calcium.

### Organic Carbon

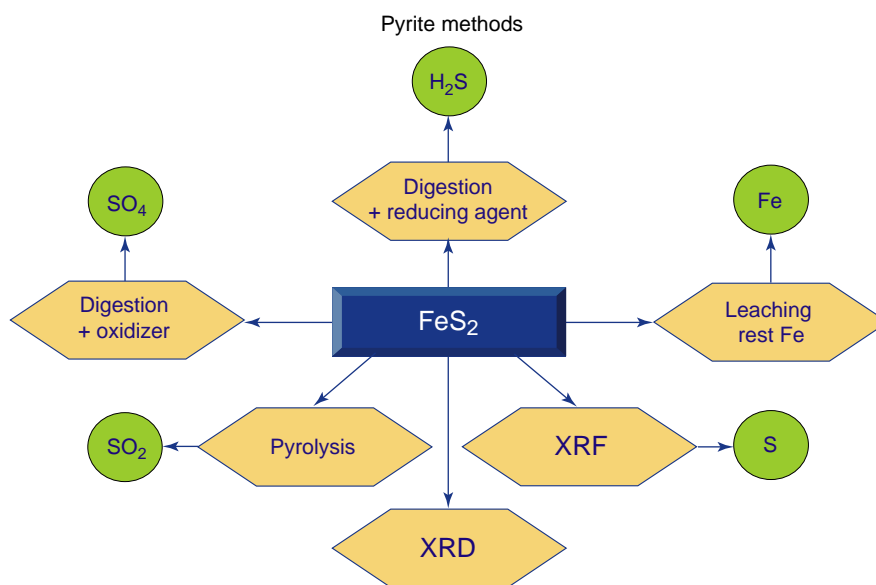
Sorbing and oxidizing solutes are retarded during aquifer passage and similarly desorbing and dissolving



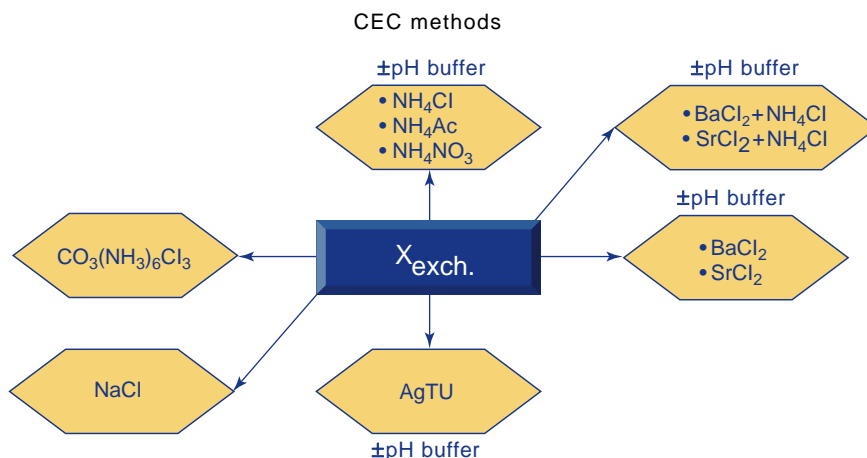
**Figure 1** Five analytical methods for the determination of calcite in soil with specific types of digestion. In the circle is a declaration of the analytical parameter ( $\Delta\text{H}^+$  = concentration of used acid,  $\Delta\text{kg}$  = decrease of weight). The hexagons contain the method of digestion. XRF = X-ray fluorescence.



**Figure 2** Three analytical methods for the determination of organic carbon in soil with specific types of digestion. In the circle is a declaration of the analytical parameter ( $\Delta\text{Ox}$  = concentration of used oxidation, e.g., hydrogen peroxide or  $\text{K}_2\text{Cr}_2\text{O}_7$ ,  $\Delta\text{kg}$  = decrease of weight). The hexagons contain the method of digestion.



**Figure 3** Six analytical methods for the determination of pyrite ( $\text{FeS}_2$ ) in soil with specific types of digestion. In the circle is a declaration of the analytical parameter. The hexagons contain the method of digestion. XRD = Roentgen diffraction, XRF = X-ray fluorescence.



**Figure 4** Five analytical methods for the determination of exchange cation salts for the determination of exchange cations and the CEC in soil. The hexagons contain the method of cation exchange.

compounds are delayed by leaching. The leaching of organic matter is generally the result of oxidation by  $\text{O}_2$  and/or  $\text{NO}_3^-$ . After leaching of the easy degradable (labile) fraction of organic matter (sometimes 2%), the tough fraction is to be leached. About 20% is considered inert then. The highest accuracy level of organic carbon dating methods can be approved by quantitation. The digestion of organic carbon can be determined in three different ways: (1) TGA ( $105^\circ\text{C}$ ,  $450^\circ\text{C}$ ,  $550^\circ\text{C}$ , and  $1000^\circ\text{C}$  on bulk organic material); (2) digestion of  $\text{CaCO}_3$  with acid after which pyrolysis is conducted; and (3) digestion with an oxidizer. In some cases, postanalysis

calculations correct the overestimation for the presence of lutum (particles  $<2\ \mu\text{m}$ ) and free iron with bounded water crystals.

### Sulfur and Iron Sulfides (Pyrite)

Pyrite oxidation can be considered as the most important water-sediment reaction. Using kinetic methods, this reaction proved to be relatively a slow process, as the oxidation of pyrite takes  $\sim 10$ – $100$  days to deplete  $\text{O}_2$  and  $\text{NO}_3$  into the water phase. The progress of the reaction is the observation of an increase of sulfate ( $\text{SO}_4$ ). It appears to be that

oxygen is reacting faster with pyrite than with labile NOM and that the reverse holds for nitrate. The oxidation of pyrite leads to some mobilization of As, Co, and Ni.

Multivariate techniques can define structure in providing the highest accuracy level achievable by analytical methods for mono- and disulfides (pyrite). Pyrite ( $\text{FeS}_2$ ) can be determined in eight different ways: (1) digestion of sulfate with an oxidizer (hydrogen peroxide, aqua regia, or concentrated  $\text{HNO}_3$ ); (2)  $\text{SO}_2$  pyrolysis according to Strohlein CS-5500; (3)  $\text{SO}_2$  pyrolysis, correction on carbonates with acid; (4) digestion with reducing agent; (5)  $\text{SO}_4$  in silver thio-ureum (AgTU) extract (with and without ammonium acetate) by CEC determination; (6)  $\text{SO}_4$  sum by sequential extractions, differentiation of sulfur constituents; (7) XRF; and (8) Röntgen diffraction (XRD) semiquantitative methods (granular, or ion-exchange complex and CEC, or  $\text{CO}_2$ -pyrolysis with acid addition). Postanalysis of geochemical data utilizing robust statistical information accounts for the sum of elemental sulfur, monosulfides, and disulfides as a result of the two-way sequential extractions. In most cases, inductively coupled plasma mass spectrometry (ICP-MS) and ion chromatography (IC) are used because of their low detection limits and analytical behavior in complex salt matrices.

### Cation-Exchange Capacity

Generalized evaluation of quality changes of an oxic aquifer evolves in a specific way with time due to sequential leaching of reactive components of the aquifer matrix. The evaluation reveals the following phases: displacement, exchange reactions, redox reactions,  $\text{NO}_2$  breaking through, leaching of pyrite and calcite. Exchange and sorption reactions between influent and aquifer matrix dominate during the quality changes. While most methods reviewed are instrument based, wet chemical techniques are also used.

Some multivariate techniques (Figure 4) can be used to define structure in providing the highest accuracy level achievable by analytical methods for CEC: exchange of cations in strontium chloride ( $\text{SrCl}_2$ ) or a combination of  $\text{SrCl}_2$  + ammonium chloride ( $\text{NH}_4\text{Cl}$ ), pH 7.6; determination of CEC with silver thio-ureum (AgTU) extract (with and without ammonium acetate) by CEC determination; and exchange of cations with salt solutions ( $\text{NaCl}$ ).

Postanalysis of geochemical data involves robust calculations to account for the total exchange capacity for cations as a result of the cation exchanges. In most cases, ICP-MS and IC are performed in the final extracts due to its low detection limit capacity and their analytical behavior in complex salt matrices.

## Analytical Techniques

This article surveys brief information according to the analytical techniques used for determining terrestrial samples.

### Laser Particle Sizer

This technique claims a working range of 0.16–1250  $\mu\text{m}$ . All particles give diffraction in all directions. Thus, light scattered by particles outside the measured range affects the results over that measured range. The suspension is pumped through a sample cell placed in the convergent laser beam and the forward-scattered light falls on 31 photosensitive sensor rings. The particle size distributions from the diffraction patterns measured by the laser are calculated using the Chahine inversion scheme.

### Atomic Spectrometry

This technique comprises a group of quantitative instrumental analytical methods based on the capacity of free atoms of both emitting and absorbing radiation at a specific wavelength. The radiation lies within the range for ultraviolet and visible light. A distinction is made between atomic emission spectrometry (AES), atomic absorption spectrometry (AAS), and atomic fluorescence spectrometry. The most commonly applied techniques are flame-AAS, graphite furnace-AAS, and ICP-AES. With ICP, excitation takes place in a plasma at a temperature of  $\sim 7000\text{ K}$ .

The best-known technique based on a combination of methods is ICP-MS. Here, the excited atoms are introduced upon their return to a lower energy level, through an interface into the ion source of a quadrupole of a mass spectrometer. The ICP thus acts as an ion source and the mass spectrometer as the ion detector. The latest development in atomic spectrometry is the electrothermal evaporation-ICP-MS technique, where a graphite furnace is coupled to an ICP-MS. In this case, use is made of the most remarkable property of a graphite furnace (elimination of matrix interferences) by a graphite tube atomizer and subsequent transport of the atomic phase into the plasma and quadrupole.

### Instrumental Neutron Activation Analysis

In instrumental neutron activation analysis (INAA), following the activation of a soil sample, preconcentrated or otherwise, by radiation with neutrons, photons (gamma radiation), or charged particles, the radioactivity generated is measured by sensing the entire gamma spectrum with a semiconductor

detector. The most commonly applied variant is epithermal neutron activation analysis.

### **X-Ray Fluorescence**

Soil is bombarded with X-rays when electrons are ejected from the innermost shells of the atoms. When the hole made in the electron shell is refilled, X-rays are emitted at a longer wavelength than the radiation previously absorbed. This wavelength is characteristic for each kind of atom so that measurement of the whole spectrum provides an analysis of the chemical proposition. XRF is a well-established analytical method for determination of major and minor elements in geological materials.

### **Ion Chromatography**

A major development is the determination of trace levels of anions in digested soil samples in a closed system using microwave-assisted heating. A specific conductometric method combines the quantification of anions by IC with specific dilution steps, which mainly comprises a selective de-enrichment of all water solution constituents.

### **Combined Methods**

Several recent studies have used quadrupole ICP-MS in combination with laser ablation (LA-ICP-MS), liquid chromatography-ICP-MS, and/or high-resolution and multi-collector ICP-MS. Despite the fact that the advantages of high sensitivity of mass spectrometric techniques are mostly applicable to the analysis of terrestrial samples simultaneously, scanning electron microscopy demonstrates the performance characteristics of inclusions in minerals and generally agrees well with published values.

## **The Mass Balance Approach**

The mass balance approach demonstrates both reactive solutes in the water recharged and solid reactive phases in the aquifer system. Some empirical rules regarding the sequence of reactions, reaction kinetics, and mobility of elements completes the dynamic quality changes of an infiltrating solute, which leaches the aquifer. Chemometric data analysis techniques allow geoanalytical chemists to control the quality of results from routine analysis by discovering trends within large multivariate datasets, and assessing and even improving data quality. Therefore, the use of multivariate quality control techniques is favored to assess data quality in routine analyses by bulk analytical techniques such as ICP-MS, XRF, and INAA. It has also been suggested that a combination of mass balance approach, principal component analysis, and

statistics could provide tools for error recognition and diagnosis. Easy-Leacher is an analytical two-dimensional spreadsheet model simulating the dynamic quality changes of an infiltrating solute, which leaches the aquifer. This mass balance approach for both reactive solutes in the water recharged and solid reactive phases in the aquifer system is combined with some empirical rules regarding the sequence of reactions, reaction kinetics, and mobility of elements. This model simulates for as many strata as desired: e.g., displacement of the native groundwater by influent, including effects of dispersion, and leaching of reactive aquifer constituents (exchangeable cations, calcite, organic matter, iron sulfides, and  $\text{MnO}_2$ ).

Information with regard to the breakthrough of solutes, the sorption, and the breakdown of organic micropollutants will be available.

## **Challenges**

Analysis of samples on soil continues to present new challenges to the analytical geochemist. In fact, the analytical trend in geochemical research is getting more compositional information from ever-smaller samples. ICP-MS with all of its hyphenated permutations continues to be the analytical method used most widely. Other trace analysis techniques are of lesser importance particularly for noteworthy studies. A second trend in geochemical research that continues to broaden during the next decade is the application of focus more specific routes in analyzing to characterize a terrestrial sample or suite of samples. This approach has long been a staple in geochemistry where samples are rare and usually small and where bulk host material is not generally plentiful. The major appeal of the multitechnique approach is that sets generally prove synergistic.

Previous hybrid hyphenated techniques and dissolution procedures have been developed and will have benefits for the future in geochemical analysis.

**See also:** **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Geochemistry:** Inorganic; Sediment; Soil, Minor Inorganic Components; Soil, Organic Components.

## **Further Reading**

- Konert M and van den Berghe J (1997) Comparison of laser grain size analysis with pipette and sieve analysis: a solution for the underestimation of the clay fraction. *Sedimentology* 44: 523–535.
- Lipschutz ME, Wolf SF, Hanchar JM, *et al.* (2003) Geochemical and cosmochemical materials. *Analytical Chemistry* 75: 2797–2811.



- Peters JH, Stuyfzand PJ, Saaltink MW, *et al.* (1998) *Artificial Recharge of Groundwater*. Nieuwegein, Netherlands: Kiwa Research and Consultancy.
- Stuyfzand PJ (in preparation) Manual of Easy-Leacher, a simple model for reactive transport of main constituents and pollutants, especially for artificial recharge and bank filtration.
- Stuyfzand PJ and Van der Jagt H (1996) Development and optimisation of chemical soil research for the prognosis on deep filtration and monitoring ground water recharge. Kiwa – Report KOA 96.109, 12pp.
- Stuyfzand PJ and Van der Jagt H (1997) Explanation on inorganic analytical techniques for soil. Kiwa – Report KOA 96.116, version 3.
- Stuyfzand PJ, Griffioen J, and Broers HP (1997) Geochemical developments on soil samples in aquifer. *H<sub>2</sub>O* 30(1): 8–13.
- Stuyfzand PJ, Van der Jagt H, and Visser HJ (1998) Optimisation of geo-chemical methods for aquifer sands, calcite, organic carbon, sulphur and exchangeable cations. Kiwa – Report KOA 98.172.
- Van Beek CGEM and van der Jagt H (1999) Mobilisation and speciation of trace elements in ground water. *Proceedings ISWA International Workshop*, Vienna, May 6–8, 1996.
- Van der Jagt H and Stuyfzand PJ (1996) Methods for trace element analysis in surface water, atomic spectrometry in particular. *Fresenius Journal of Analytical Chemistry* 354: 32–40.

## Soil, Minor Inorganic Components

A C Edwards, Peterhead, Aberdeenshire, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Soil presents the analyst with various technical problems with respect to the complex and variable chemical nature of its matrix (aluminosilicate – organic) and the wide range in concentration of the minor constituents that may be present. The heterogeneous distribution of minor components means that the appropriateness of collection and preparation procedures in providing a representative soil sample must be considered. Many minor components commonly present in soil are essential to biological life processes but become toxic when present in excess. Soil receives minor components from a range of anthropogenic sources and their accumulation is of direct concern for both human and general ecosystem health. The quantification of a small but highly dynamic bioavailable fraction of the total minor component present in soil represents one important topic for analytical chemistry.

Soil is composed of a heterogeneous mixture of solid, gaseous, and aqueous phases, the relative proportions of which can vary both spatially and temporally. The solid matrix consists of an intimate mixture of inorganic (primarily weathering products) and organic (products of biological decomposition) components. The proportion of total soil porosity occupied by water may vary from very low values up to 100% in a fully saturated soil. Soils typically develop a distinctive vertical zonation as a consequence of selective mineral dissolution coupled with

transport and reprecipitation of inorganic components and a decline in organic material.

A precise distinction between what constitutes a ‘major’ or ‘minor’ soil component is difficult to make. For present purposes the term ‘minor component’ is applied to situations where total concentrations are generally <100 µg per g. This may encompass more than 50 minor components many of which could be present at concentrations below 0.1% (Figure 1).

Although most of the minor components are associated with the solid phase a small but functionally significant fraction is present in the soil solution and represents the immediately bioavailable and mobile fraction. The aqueous phase is the primary medium for redistribution and transportation of soluble and particulate bound chemical species within soil. A constant state of flux, resulting in a dynamic equilibrium, exists between solid and solution phases as both respond to local situations of depletion (e.g., plant uptake) or resupply (e.g., application of sewage sludge). Apportioning between solid and solution is sensitive to the prevailing local physicochemical environment through influences of factors such as pH, redox, and electrolyte concentrations.

### Origin of Minor Components

Total amounts of minor components vary widely between soils and are often reflective of a strong geochemical influence (Table 1). Here, soil developed from Serpentine (an ultrabasic, igneous rock) is typically high for Co, Cr, and Ni with the result that even in this ‘natural state’ concentrations are greater than the EC recommended limits for soils receiving metal-containing waste materials (see Table 2). Natural background concentrations have become enhanced

- Peters JH, Stuyzand PJ, Saaltink MW, *et al.* (1998) *Artificial Recharge of Groundwater*. Nieuwegein, Netherlands: Kiwa Research and Consultancy.
- Stuyfzand PJ (in preparation) Manual of Easy-Leacher, a simple model for reactive transport of main constituents and pollutants, especially for artificial recharge and bank filtration.
- Stuyfzand PJ and Van der Jagt H (1996) Development and optimisation of chemical soil research for the prognosis on deep filtration and monitoring ground water recharge. Kiwa – Report KOA 96.109, 12pp.
- Stuyfzand PJ and Van der Jagt H (1997) Explanation on inorganic analytical techniques for soil. Kiwa – Report KOA 96.116, version 3.
- Stuyfzand PJ, Griffioen J, and Broers HP (1997) Geochemical developments on soil samples in aquifer. *H<sub>2</sub>O* 30(1): 8–13.
- Stuyfzand PJ, Van der Jagt H, and Visser HJ (1998) Optimisation of geo-chemical methods for aquifer sands, calcite, organic carbon, sulphur and exchangeable cations. Kiwa – Report KOA 98.172.
- Van Beek CGEM and van der Jagt H (1999) Mobilisation and speciation of trace elements in ground water. *Proceedings ISWA International Workshop*, Vienna, May 6–8, 1996.
- Van der Jagt H and Stuyfzand PJ (1996) Methods for trace element analysis in surface water, atomic spectrometry in particular. *Fresenius Journal of Analytical Chemistry* 354: 32–40.

## Soil, Minor Inorganic Components

A C Edwards, Peterhead, Aberdeenshire, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Soil presents the analyst with various technical problems with respect to the complex and variable chemical nature of its matrix (aluminosilicate – organic) and the wide range in concentration of the minor constituents that may be present. The heterogeneous distribution of minor components means that the appropriateness of collection and preparation procedures in providing a representative soil sample must be considered. Many minor components commonly present in soil are essential to biological life processes but become toxic when present in excess. Soil receives minor components from a range of anthropogenic sources and their accumulation is of direct concern for both human and general ecosystem health. The quantification of a small but highly dynamic bioavailable fraction of the total minor component present in soil represents one important topic for analytical chemistry.

Soil is composed of a heterogeneous mixture of solid, gaseous, and aqueous phases, the relative proportions of which can vary both spatially and temporally. The solid matrix consists of an intimate mixture of inorganic (primarily weathering products) and organic (products of biological decomposition) components. The proportion of total soil porosity occupied by water may vary from very low values up to 100% in a fully saturated soil. Soils typically develop a distinctive vertical zonation as a consequence of selective mineral dissolution coupled with

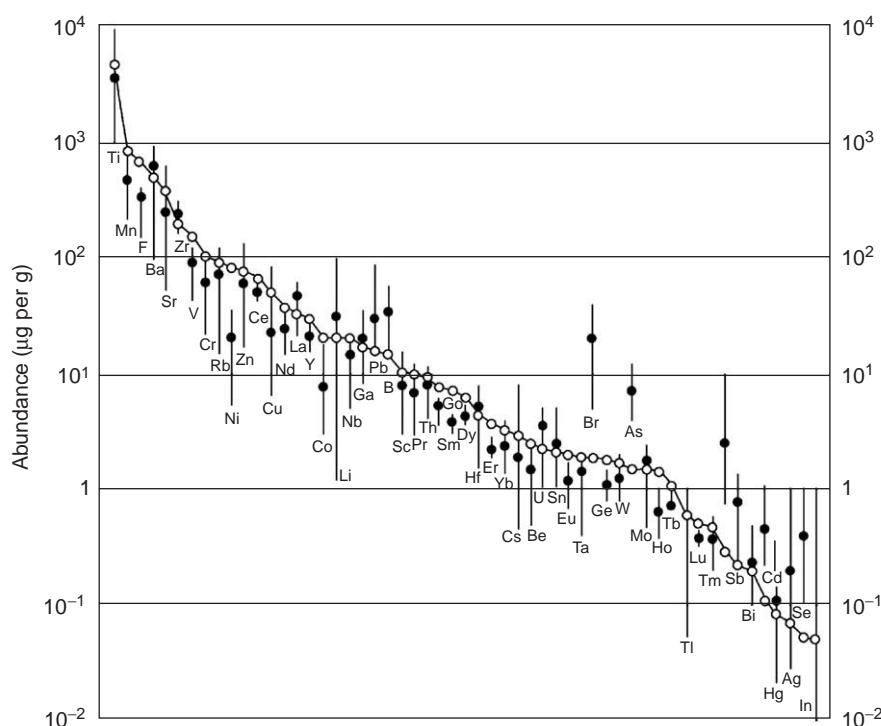
transport and reprecipitation of inorganic components and a decline in organic material.

A precise distinction between what constitutes a ‘major’ or ‘minor’ soil component is difficult to make. For present purposes the term ‘minor component’ is applied to situations where total concentrations are generally <100 µg per g. This may encompass more than 50 minor components many of which could be present at concentrations below 0.1% (Figure 1).

Although most of the minor components are associated with the solid phase a small but functionally significant fraction is present in the soil solution and represents the immediately bioavailable and mobile fraction. The aqueous phase is the primary medium for redistribution and transportation of soluble and particulate bound chemical species within soil. A constant state of flux, resulting in a dynamic equilibrium, exists between solid and solution phases as both respond to local situations of depletion (e.g., plant uptake) or resupply (e.g., application of sewage sludge). Apportioning between solid and solution is sensitive to the prevailing local physicochemical environment through influences of factors such as pH, redox, and electrolyte concentrations.

### Origin of Minor Components

Total amounts of minor components vary widely between soils and are often reflective of a strong geochemical influence (Table 1). Here, soil developed from Serpentine (an ultrabasic, igneous rock) is typically high for Co, Cr, and Ni with the result that even in this ‘natural state’ concentrations are greater than the EC recommended limits for soils receiving metal-containing waste materials (see Table 2). Natural background concentrations have become enhanced



**Figure 1** Comparison of minor components in soils and their abundance in the lithosphere. Open circles mean content in the lithosphere; black circles mean content in topsoils; vertical lines mean values commonly found in top soils. (Reprinted with permission from Kabata-Pendias A and Pendias H (1984) *Trace Elements in Soils and Plants*. Boca Raton, FL: CRC Press; © CRC Press, Boca Raton, FL.)

**Table 1** Comparison of selected minor constituents for three soils developed from contrasting parent material and the range in total concentrations in topsoils of England and Wales (all mg per kg oven-dry soil)

Element	Parent material <sup>a</sup>			Total <sup>b</sup>		
	Serpentine	Granite	Sandstone	Mean	Min	Max
Co	70	1–2	4.0	10.6	0.2	322
Cr	2300	10	68.0	—	—	—
Cu	15	7.0	8.5	23.1	1.2	1508
Mn	1600	450	440	760.9	3.0	42 603
Ni	670	5.4	17.0	24.5	0.8	440
Zn	143	36	5.7	97.1	5.0	3648

<sup>a</sup>Data from Ure AM, Bacon JR, Berrow ML, and Watt JJ (1979) The total trace element content of some Scottish soils by spark source mass spectrometry. *Geoderma* 22: 1–23.

<sup>b</sup>Data from McGrath SP and Loveland PJ (1992) *The Soil Geochemical Atlas of England and Wales*. London: Blackie Academic and Professional.

to varying degrees through ‘anthropogenic’ additions. Sources of human derived contamination may reach soil through a combination of ‘passive’ or ‘active’ mechanisms. Examples of the former include atmospheric deposition while the latter may involve the regulated application of ‘contaminated’ materials such as industrial or domestic wastes. The accumulation of Pb in remotely located soils serves to demonstrate that long-distance atmospheric transport occurs. Isotopic Pb signatures allow individual sources of contamination to be identified. Repeated application

of phosphatic fertilizers manufactured from sedimentary deposits containing a Cd impurity provides a further example that has resulted in the widespread contamination of agricultural soils with the potential for direct transfer into the human food chain.

The composition of anthropogenic sources and individual chemical species present is highly variable. Repeated applications of contaminated materials such as sewage sludge have resulted in situations where the long-term accumulation of potentially toxic species represents a potential health risk.

**Table 2** Influence of soil pH on the maximum permissible concentrations (mg per kg per dm) of minor components in soil (CEC, 2000 Commission of the European Communities Working document on sludge: 3rd draft. ENV.E.3/LM.CEC, Brussels)

Elements	Current values proposed values			
	Directive 86/278/EEC	pH 5–6	pH 6–7	pH > 7
Cd	1–3	0.5	1	1.5
Cr	–	30	60	100
Cu	50–140	20	50	100
Hg	1–1.5	0.1	0.5	1
Ni	30–75	15	50	70
Pb	50–300	70	70	100
Zn	150–300	60	150	200

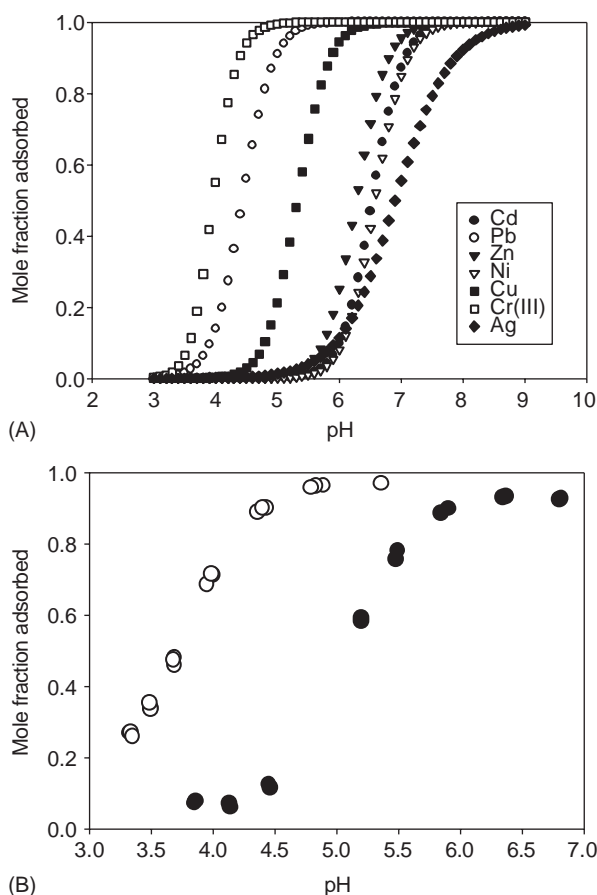
### Wider Significance of Minor Components and Current Regulations

Most minor components while being essential to the biochemical, life processes, also happen to be potentially extremely toxic when present in excess. Defining what constitutes a 'safe' soil concentration is difficult. Guideline concentration limits for soil that is likely to receive contaminated waste are continually under review. Many of these guidelines tend to be based upon a total concentration in spite of the often poor relationship between total and biologically significant fractions. This situation probably reflects the relative ease of determining 'total' concentrations compared to some biologically labile and highly dynamic component. The incorporation of factors that might be influential in determining bioavailability, such as a weighting factor for soil pH, have been attempted (Table 2).

Potential toxicity is sensitive to the prevailing local environment through selective influences on chemical speciation or activity. Toxicity can result from either 'chronic' or more 'acute' type exposures. Overlapping ranges in sensitivities between individual organisms makes the setting of threshold concentrations for individual ecosystems difficult.

### Developments in the Determination of Minor Constituents

A significant early requirement for soil analysis arose from a need to define optimum conditions for crop growth and appropriate dietary intake from feed for animal health purposes and identify soil deficiencies. Recently, research emphasis has changed to consider situations where bioaccumulation and potential toxicity is the primary concern. Bioassays have incorporated a range of test species including plants, earthworms, and, recently, toxicity fingerprinting with a metabolic lux-marked bacterial biosensor.



**Figure 2** (A) Calculated adsorption by a hydrous ferric oxide of several metal cations at a total added metal concentration of  $10^{-6} \text{ mol l}^{-1}$  using the diffuse double layer surface complexation model. (From Dzombak DA and Morel FMM (1990) *Surface Complexation Modeling: Hydrous Ferric Oxide*. New York: Wiley.) (B) Experimentally measured cadmium ( $[\text{Cd}]_{\text{total}} = 0.3 \text{ mmol l}^{-1}$ ) adsorption by O (open circles) and B (filled circles) horizons ( $18.5 \text{ g dm}^{-3}$ ) in  $0.01 \text{ mol l}^{-1} \text{ NaNO}_3$ . (Data from Lumsdon DG (2004) Partitioning of organic carbon, aluminium and cadmium between solid and solution in soils: application of a mineral-humic particle additivity model. *European Journal of Soil Science*, in press.)

Predicting likely bioavailability and the potential for transfer into the human food chain presents a continuing challenge. This requires an understanding of the various factors that interact to influence ion speciation and partitioning between soil and solution phases. Figure 2A demonstrates a modeled response of the influence of pH on adsorption of various species by a hydrous ferric oxide. The important influence that the type (composition) of solid phase has on partitioning is shown in Figure 2B where experimental data for organic (O) and mineral (B) horizons are compared. The two soils behave very differently; at a given pH the proportion of the Cd adsorbed by the organic soil is greater. The need to increase

understanding of the factors affecting solid–solution dynamics is especially apparent for toxicity related studies.

## Distribution of Minor Components in Soil

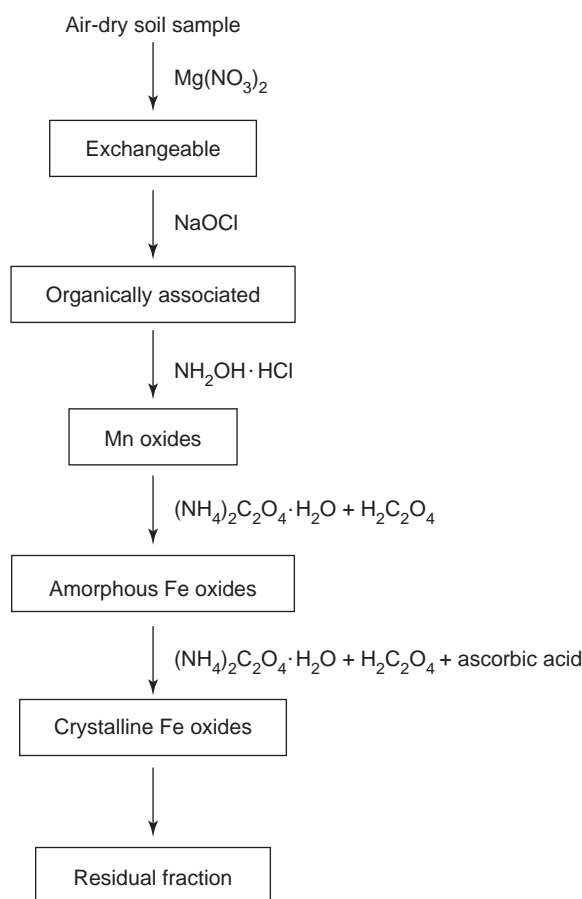
Soil particles differ widely in chemical composition, which introduces a degree of heterogeneity into the spatial distribution of minor components. Chemically reactive species become concentrated within smaller sized particles or associated with secondary surface coatings and must be considered with respect to the collection and preparation of truly ‘representative’ soil samples.

Soil analysis has followed two broad but complementary approaches. The first and probably still most widely performed analysis relies upon measurements made on a solution after an initial solubilization or extraction operation and utilizes readily available techniques and instrumentation. The second involves some form of direct measurement on soil that requires specialist equipment and interpretation. Extraction-based approaches provide intrinsic information capable of describing field-scale spatial and temporal variability. Direct analysis tends to yield detailed compositional information at individual aggregate and subparticle scales. Overlap exists between approaches with both offering multielemental capabilities at the required limits of detection.

### Extraction and Subsequent Analysis

Minor components form a wide range of chemical associations with soil constituents including surface exchange sites, carbonates, metallic oxyhydroxides, and organic matter as well as the ‘residual’ crystalline structures of silicate minerals. This realization underpinned the development of selective procedures including sequential extractions that are designed to preferentially solubilize these individual fractions (Figure 3). Although widely adopted, concern remains as to the chemical rigor of these operationally based procedures. A lack of methodological consistency coupled with issues that arise from soils having different chemical properties makes interlaboratory quality control and standardization difficult. The introduction of internationally agreed standard soil samples is helping to resolve this important issue.

The increased quality of analytical information and greatly improved detection limits has had a positive feedback enabling less aggressive extractants to be employed, which provide an improved quantification of the bioavailable fraction. Parallel developments in the extraction conditions employed have



**Figure 3** Sequential extraction scheme for various soil zinc fractions. (From Perveen Z *et al.* (1994) Redistribution of zinc from sewage-sludge applied to a range of contrasting soils. *Science of the Total Environment* 155: 161–171.)

enabled greater definition of the kinetic parameters regulating depletion and resupply (e.g., diffusive gradients in thin films).

### Direct Analysis

Analysis performed directly on soil has continued to benefit from instrumental improvements, which now readily allow analysis at subparticle scale. Many of the techniques that provide a direct analysis also offer the potential for excellent spatial resolution with possibilities of even some depth resolution whilst also providing information on the chemical associations between individual minor components.

## Soil Sampling and Storage

### Collection

Soil represents a difficult matrix for which to obtain a representative and homogeneous sample. In the field, properties change over relatively short vertical



and also horizontal distances that have been described using intensive sampling and appropriate geostatistical techniques. Variability also occurs within a single well-mixed soil sample and exists across individual soil particle sizes. Separating variability due to natural factors from that arising due to anthropogenic sources will be an increasing requirement.

The collection and subsequent analysis of soil samples can follow one of two broad strategies. Either a number of cores may be taken and analyzed individually, or samples can be combined and thoroughly mixed to provide a single bulked sample. Details of the spatial variation in concentrations are sacrificed in the latter case. Choice of the most appropriate analytical technique requires consideration of detection limits, range of chemical species to be determined, and concentration range to be determined.

### Contamination

It is possible to minimize potential sources of contamination considerably through careful thought and prudent choice of materials likely to contact the soil sample. Soil sieving and milling stages can be particularly prone to introducing contamination; the use of carbon steel, plastic, or aluminum can minimize this.

### Storage

Soil samples stored in an air-dry condition are relatively stable particularly when total contents are to be determined. Samples must be well mixed after long-term storage to minimize the affect of size separation. Storage of soil extracts is much more problematical due to precipitation from solution and sorption processes on the container walls and should be avoided when possible. This is particularly the case for soil solutions or when weak electrolytes are used as extractants. The common method of preserving samples is acidification, but is only suitable when speciation is not required. Biological activity can be minimized by freezing samples or by adding small amounts of toluene.

## Preparation Procedures

There are few situations where pretreatment of the solid sample is not desirable or indeed essential. Invariably the aim is to achieve one or more of the following objectives: (1) sample homogeneity, (2) decomposition, (3) elimination of possible matrix interferences, and (4) preconcentration.

The extent to which any treatment is required varies considerably and depends upon the broad

chemical characteristics of the sample and the analytical method used.

### Sieving and Subsampling

Once collected, soil samples are generally air-dried and passed through a 2 mm sieve. It may be necessary to take into account the weight of discarded material (which is  $> 2$  mm). A more homogeneous sample can be produced using additional milling (e.g., agate ball mill) and is especially recommended for many digestion and direct analysis procedures. Subsampling can be achieved by either coning and quartering or the use of a mechanical riffler.

### Decomposition

**Dissolution** For the majority of soil samples a total dissolution requires the use of an acidic cocktail, which includes hydrofluoric acid and an oxidant to remove the organic component. Less chemically rigorous mixtures include aqua regia ( $\text{HCl}/\text{HNO}_3$  3:1 v/v) and  $\text{HNO}_3/\text{HClO}_4$ , which are often adequate for the nonsilicate bound components. Various digestion and reflux procedures are in common use although care must be taken when either calcareous or organic-rich samples (due to the release of carbon dioxide) are being digested. Flow-through microwave-assisted digestion apparatus permits soil slurries to be continuously processed and analyzed directly. These online approaches offer many potential advantages.

**Fusion** Various fusion melts can be used, but probably the most common uses alkaline borax. Fusion disks can be analyzed directly or after the solidified melt has been dissolved in dilute acid.

**Selective extraction and concentration** Soil digest and extraction solutions often contain excessive concentrations of potentially interfering elements, originating directly from the sample or from the chemicals used during the pretreatment. This can cause instrumental problems, which include for atomic absorption spectrometry (AAS), chemical (ionization suppression), spectral (line overlap), and physical (such as nebulizer blocking caused by high salt levels) interferences. These matrix effects are often more serious in the present context, because of the low analyte concentrations involved. Various separation techniques can help to minimize these effects and include precipitation, preferential complexation, and separation through solvent extraction, or selective chelation onto ion-exchange material. There are also certain selective procedures such as hydride generation for metals such as arsenic and selenium.

Several of these procedures have an additional advantage of introducing a preconcentration step. Complexation and selectivity using chemically active compounds such as ethylenediaminetetraacetic acid or synthetic ion-exchange materials have been used with considerable success for many years. After extraction there are a number of possible analytical options: the resin material may be either analyzed directly by X-ray fluorescence spectrometry or as a solution after digestion or elution of the resin.

**Filtering** It is often important to distinguish between what is present in soil solution as soluble or colloidal/particulate material. An operationally defined filtering stage (0.45  $\mu\text{m}$ ) is commonly employed.

## **Analytical Techniques**

This section is subdivided on the basis of whether the analysis is made directly on the 'solid' soil sample or on a 'solution' after a pretreatment stage. A further division is made on whether the analysis technique determines the total amount present or is capable of distinguishing between chemical species.

### **Direct Analysis of Soil Material**

The majority of techniques capable of direct analysis provides an analysis of the whole sample and has multielement capabilities. Some spatial resolution can be achieved by using, for example, X-ray microprobe attachment in conjunction with electron microscopy.

**Instrumental neutron activation analysis** This is an extremely sensitive technique being especially suited for rare-earth species and requires from 50 to 200 mg of finely ground sample. All trace elements, other than boron, that are essential to plants can be determined. There are few matrix interference effects but specialized irradiation facilities are required making its routine use limited. The high aluminum content of many soil samples produces a highly radioactive matrix but fortunately it has a relatively short half-life (150 s). Digestion of samples after irradiation allows the use of a chemical separation step, resulting in lower detection limits and an extended range of determinable chemical species. One recent analysis of an international standard soil sample reported concentrations of most species were within 10% of the certified values and precision (six replicates) was found within 15%.

**X-ray analysis** Two techniques, X-ray fluorescence (XRF) spectrometry and electron microprobe, are

widely used in soil science. The data produced are complementary. The former technique involves detection by either energy or wavelength dispersive means while the latter is more commonly energy dispersive. XRF can be automated and is capable of providing a rapid and comprehensive preliminary analysis (elements with atomic number  $>11$ ). Samples are commonly presented either as pressed disks of finely milled soil (which may need binding agent) or fused discs (e.g., borate flux). It is possible to produce suitable calibration standards using artificial mixes based upon iron, aluminum, and silicon oxides. Certain of the procedures already discussed can be combined with XRF. An example would be complexation of metals (with ammonium pyrrolidine dithiocarbamate), followed by extraction into chloroform and evaporation on a filter paper.

Use of the microanalysis attachment in conjunction with electron microscopy can provide a visual representation of concentrations. When linked to nondestructive preparation techniques, such as critical point drying, this offers the potential of assessing specific soil processes. Adamo *et al.* used scanning electron microscopy in back-scattered electron mode to examine the soil-root interface of plants grown at a range of soil pH and soil phosphorus concentrations. Excellent spatial resolution is, however, counterbalanced by relatively poor detection limits due to poor spectral resolution resulting in severe peak overlap.

**Spark source mass spectrometry** This technique has considerable potential for soil analysis, but continues to prove to be unpopular. This has been variously attributed to the high cost, major interelement effects, and rather specialized nature of both the equipment and sample preparation procedure. The technique does have particular advantages in that it has a high sensitivity and a wide range of chemical species are detectable.

### **Direct Analysis of Extract Solution**

A wide range of the available atomic spectroscopic techniques are routinely used for soil analysis. Some of the more relevant information for individual instrumental techniques is summarized in **Table 3**.

**Flame and electrothermal techniques** Both atomic absorption and emission have been used, with the former most widely applied to the analysis of minor soil components. Detection limits can often be very similar to the concentration found in extraction solutions of 'natural' soils. While the sampling procedure is easily automated, AAS, it is inefficient (both time and sample volume) for routine multielement

**Table 3** Comparison of some attributes of instrumental techniques commonly employed for soil analysis

	Detection limit (mg per kg or mg l <sup>-1</sup> )	Multielement	Selectivity	Matrix interference	Precision (%)	Accuracy
GF-AAS	10 <sup>-2</sup> –1	No	Good	Small	0.5–5	Good
F-AAS	1–10 <sup>3</sup>	No	Good	Small	0.1–1	Good
ICP-AES	10 <sup>-2</sup> –10	Yes	Moderate	Moderate	0.3–2	Moderate
ICP-MS	10 <sup>-3</sup> –1	Yes	Good <sup>a</sup>	Moderate	0.5–2	Good
TR-XRF	10 <sup>-2</sup> –1	Yes	Good	Moderate	2–5	Good
LIBS	10 <sup>2</sup> –10 <sup>3</sup>	Yes	Moderate	Moderate	1–10	Moderate
GF-LEAFS	10 <sup>-5</sup> –10	No	Excellent	Small	3–5	Good
Hydride generation	10 <sup>-3</sup> –1	Yes	Excellent	Moderate	–	Good

<sup>a</sup> Isotope selectivity.

From Sturgeon RE (2000) Current practice and recent developments in analytical methodology for trace element analysis of soils, plants, and water. *Communications in Soil Science and Plant Analysis* 31(11–14): 1479–1512.

analysis. It remains useful in cases when only individual elements are required or where complex matrices are involved and can handle total dissolved solids (TDS) up to 10%. Analysis of soil extracts using either flame technique is prone to chemical and physical interferences. Clogging of the burner and nebulizer as a consequence of the high salt contents of many soil extractants is a particular problem. The presence of high concentrations of calcium, magnesium, potassium, sodium, and aluminum in many soil extracts can often cause chemical and background interferences.

Relatively poor detection limits are the main drawback of flame AAS, changing the method of sample introduction by using furnace atomizers can improve detection limits by two orders of magnitude or more. An additional advantage of electrothermal-AAS is the low sample volume requirements of the technique and a reduction of sampling-related interferences (viscosity and burner clogging). It remains an extremely popular technique.

Hydride generation for minor components such as arsenic, selenium, and tin continues to have advantages that include separation of the analyte from the matrix, increased transport efficiency to the point of atomization, and the capability for organometallic speciation.

**Inductively coupled plasmas** The commercial availability of instruments using inductively coupled plasmas (ICP) means they are now in routine use. High initial and running costs are offset against the multi-element capability of the techniques, some being capable of determining 30 elements simultaneously. ICP offers numerous advantages with respect to analysis of soil extracts when compared to conventional flame techniques. A wide linear range, often covering more than three orders of magnitude, is especially suited to analysis of soil. It is suitable for all elements other than halides, inert gases, and

atmospheric gases. Many of the interferences commonly associated with flame techniques are less significant. Some spectral (overlap of emission lines) and matrix interferences (affecting excitation conditions within the plasma and nebulization efficiency) do occur. Various 'online' cleanup or concentration procedures have been reported, such as hydride generation for the simultaneous determination of arsenic, antimony, bismuth, selenium, and tellurium.

The increasing availability and superior performance of mass spectrometric (MS) detection systems means these are replacing optical (emission) based systems for routine use. Each has its own particular attributes; the MS system does allow higher sensitivities to be achieved and can provide additional isotopic information that may be particularly useful where pollution sources have a unique isotope fingerprint. The main disadvantage of MS for soil extracts is related to matrix interferences particularly regarding the use of organic solvents. Improved nebulizer design means it is possible to analyze soil slurries (up to 10% TDS) directly.

## Speciation

The increasing emphasis for information on speciation and bioavailability assessments has been discussed earlier; the following techniques allow direct speciation on soil extraction solutions.

### Electrochemical Methods

Anodic stripping voltammetry, potentiometric stripping voltammetry, and differential pulse polarography are used for the simultaneous determination of up to 10 analytes at extremely low concentrations (detection limits <0.01 µg l<sup>-1</sup>). Electroanalytical techniques are applicable for ~30 elements. Stripping analysis allows differentiation between chemical forms but is subject to interference from adsorption

and poisoning by organic compounds onto the mercury electrode. Additional interferences include peak overlapping due to similarity of redox potentials and the formation of intermetallic compounds between the deposited metals.

### Spectrophotometric Procedures

A wide range of spectrophotometric and fluorometric methods have been developed that are extremely sensitive and are easily adaptable for automated analysis. The usefulness of these methods can be reduced because of a nonspecific nature and interferences caused by soluble organic material.

### Liquid Chromatography

Numerous specialized liquid chromatography (LC) systems and ion-exchange chromatography in particular have been used to determine a wide selection of minor components that include selenium, arsenic, molybdenum, chromium, and boron. Various detection systems have been employed, including AAS, ICP-atomic emission spectroscopy, electrochemical, and conductivity. A concentrator column can also be included inline.

### Isotopic Analysis

Isotopic investigations are widely used in soil/plant nutritional and pollution studies. Introduction of an isotope by design or accident enables specific fractions or pools to be labeled, making it possible to follow rates of exchange between soil matrices: soil solution–plant uptake and drainage water. The usefulness of this technique ultimately depends upon the availability of an isotope, with a suitable half-life. A number of recent pollution incidents have involved both stable (lead petroleum additions) and radioisotopes (including  $^{137}\text{Cs}$  and  $^{125}\text{I}$ ).

## Examples of Applications for Analysis of Soil Minor Components

### Monitoring of Contaminated Sites

Onsite total analysis by energy dispersive XRF spectrometry is increasingly popular. When linked with various contour-mapping software packages, the technique can give a rapid assessment of a suspected pollution incident.

### Estimation of Soil Erosion Rates and Geographical Sourcing of Material

A recent example using the artificial radionuclide  $^{137}\text{Cs}$  was used to estimate the net ( $\sim 35$  year) soil

flux (erosion and deposition) in a  $5 \times 5 \text{ km}^2$  area of East Anglia (UK). The continuous spatial variation of  $^{137}\text{Cs}$  over this area was assessed and the redistribution of soil by wind investigated. Similar approaches have been used to successfully quantify soil erosion by water.

### Analysis of Road Dusts and Roadside Soils

Elevated concentrations of platinum group species (Pt, Pd, and Rh) have been detected in roadside media by ICP-MS after microwave digestion and cation exchange. The various ratios in road dusts and soils are consistent with known automobile catalytic converter compositions although no direct relationship between abundance and traffic volume was evident. The significance of other local factors such as driving style, topography, road drainage, and climate were suggested.

*See also:* **Activation Analysis:** Neutron Activation. **Atomic Absorption Spectrometry:** Flame; Electrothermal. **Atomic Emission Spectrometry:** Flame Photometry; Inductively Coupled Plasma. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Extraction:** Solvent Extraction Principles. **Ion Exchange:** Overview. **Isotope Dilution Analysis.** **Polarography:** Inorganic Applications. **Radiochemical Methods:** Radiotracers. **Sample Dissolution for Elemental Analysis:** Wet Digestion. **Sample Handling:** Comminution of Samples. **Voltammetry:** Anodic Stripping; Cathodic Stripping; Inorganic Compounds. **X-Ray Fluorescence and Emission:** Wavelength Dispersive X-Ray Fluorescence; Energy Dispersive X-Ray Fluorescence.

### Further Reading

- Adamo P, McHardy WJ, and Edwards AC (1998) SEM observations in the back-scattered mode of the soil-root zone of *Brassica napus* (cv. Rafal) plants grown at a range of soil pH values. *Geoderma* 85(4): 357–370.
- Batley GE (1989) *Trace Element Speciation: Analytical Methods and Problems*. Boca Raton, FL: CRC Press.
- Chappell A and Warren A (2003) Spatial scales of Cs-137-derived soil flux by wind in a  $25 \text{ km}^2$  arable area of eastern England. *Catena* 52(3–4): 209–234.
- Hooda PS, Zhang H, Davison W, and Edwards AC (1999) Measuring bioavailable trace metals by diffusive gradients in thin films (DGT): Soil moisture effects on its performance in soils. *European Journal of Soil Science* 50(2): 285–294.
- Nolan AL, Lombi E, and McLaughlin MJ (2003) Metal bioaccumulation and toxicity in soils – Why bother with speciation?. *Australian Journal of Chemistry* 56(2–3): 77–91.
- Potts PJ (1992) *A Handbook of Silicate Rock Analysis*. Dordrecht: Kluwer Academic.
- Smith KA and Cresser MS (2003) *Soil and Environmental Analysis, Modern Instrumental Techniques*, 3rd edn. New York: Dekker.



- Turner NL, Horsburgh A, Paton GI, *et al.* (2001) A novel toxicity fingerprinting method for pollutant identification with lux-marked biosensors. *Environmental Toxicology and Chemistry* 20(11): 2456–2461.
- Ure AM (1994) Methods of analysis for heavy metals in soils. In: Alloway BJ (ed.) *Heavy Metals in Soils*. Dordrecht: Kluwer Academic.

- Whitely JD and Murray F (2003) Anthropogenic platinum group element (Pt, Pd and Rh) concentrations in road dusts and roadside soils from Perth, Western Australia. *The Science of the Total Environment* 317: 212–235.
- Zhang R, Rahman S, Vance GF, and Munn LC (1995) Geostatistical analyses of trace-elements in soils and plants. *Soil Science* 159(6): 383–390.

## Soil, Organic Components

**M Bernoux**, Institut de Recherche pour le Développement, Piracicaba, Brazil

**C E P Cerri**, Universidade de Sao Paulo, Piracicaba, Brazil

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Soil organic matter (SOM) is commonly defined as the organic fraction of the soil exclusive of undecayed plant and animal residues. Sometimes, in a broader definition, SOM encompasses the totality of organic material (living and nonliving) present in soils, including thus, living microorganisms and undecayed residues. In cultivated soil, synthetic or natural organic compounds may also be introduced in the soil (sewages, pesticides, etc.). The terms SOM and humus have often been used synonymously.

SOM plays a primordial role in the productivity of soils, being a major reservoir and source of N, P, S, and other nutrients indispensable for the plant growing, and an energy source for heterotrophic soil fauna. Also, SOM has environmental importance on issues related with erosion, pollution, and global climate change.

By convention, soil samples used in the analysis have to be air-dried, crushed, and sieved to 2 mm. By convention, chemical, physical, and mineralogical analysis are made on the <2 mm material. Therefore, SOM includes only those organic materials that accompany soil particles through a 2 mm sieve.

SOM is either analysis in terms of quantity, or of quality depending on the final objectives.

### Quantitative Analyses

SOM content is highly variable in soils (variations with depth and soil type). Major elements are C, H, O, N, P, and S. Organic carbon (C) is the major component of SOM. It is difficult to quantitatively

estimate the amount of OM present in a soil. Procedures used in the past involved determination of the change in weight of a soil sample resulting from destruction of organic compounds by H<sub>2</sub>O<sub>2</sub> treatment or by ignition at high temperature. Both techniques are subject to error. The H<sub>2</sub>O<sub>2</sub> method does not quantitatively remove organic matter, and the ignition method gives an overestimate because both inorganic and organic constituents lose weight during heating. Alternatively, SOM quantity is assessed through organic C content: SOM = organic C × 1.724. This last coefficient, called the van Bemmelen factor, is used by assuming that SOM contains 58% organic C. It is clear that it is only an approximation, as the proportion of organic C in SOM is highly variable. Published organic C–organic matter conversion factors for surface soils have varied from 1.5 to 2.5. The appropriate factor must be experimentally determined for each soil. Although neither the direct determination of organic matter nor the calculation of organic matter content is completely accurate, the best procedures currently available are described here. Because of the problem associated with determining the organic matter content of a soil, it is strongly suggested that investigators determine and report the organic C concentration as a measure of the organic matter in a soil.

### Total Carbon in Soil

Total C in soils is the sum of both organic and inorganic C. Most organic C is present in the soil organic matter fraction, whereas inorganic C is largely found in carbonate minerals. Not all soils contain inorganic C because of dissolution during soil formation of carbonate minerals originally present in parent material. However, organic C is present in all agricultural soils. In soils formed from calcareous parent material under arid conditions, it is not unusual for the inorganic C concentration to exceed the amount of organic C present. Organic C is mainly in the soil organic fraction, which consists of the cells of microorganisms, plant, and animal



- Turner NL, Horsburgh A, Paton GI, *et al.* (2001) A novel toxicity fingerprinting method for pollutant identification with lux-marked biosensors. *Environmental Toxicology and Chemistry* 20(11): 2456–2461.
- Ure AM (1994) Methods of analysis for heavy metals in soils. In: Alloway BJ (ed.) *Heavy Metals in Soils*. Dordrecht: Kluwer Academic.

- Whitely JD and Murray F (2003) Anthropogenic platinum group element (Pt, Pd and Rh) concentrations in road dusts and roadside soils from Perth, Western Australia. *The Science of the Total Environment* 317: 212–235.
- Zhang R, Rahman S, Vance GF, and Munn LC (1995) Geostatistical analyses of trace-elements in soils and plants. *Soil Science* 159(6): 383–390.

## Soil, Organic Components

**M Bernoux**, Institut de Recherche pour le Développement, Piracicaba, Brazil

**C E P Cerri**, Universidade de Sao Paulo, Piracicaba, Brazil

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Soil organic matter (SOM) is commonly defined as the organic fraction of the soil exclusive of undecayed plant and animal residues. Sometimes, in a broader definition, SOM encompasses the totality of organic material (living and nonliving) present in soils, including thus, living microorganisms and undecayed residues. In cultivated soil, synthetic or natural organic compounds may also be introduced in the soil (sewages, pesticides, etc.). The terms SOM and humus have often been used synonymously.

SOM plays a primordial role in the productivity of soils, being a major reservoir and source of N, P, S, and other nutrients indispensable for the plant growing, and an energy source for heterotrophic soil fauna. Also, SOM has environmental importance on issues related with erosion, pollution, and global climate change.

By convention, soil samples used in the analysis have to be air-dried, crushed, and sieved to 2 mm. By convention, chemical, physical, and mineralogical analysis are made on the <2 mm material. Therefore, SOM includes only those organic materials that accompany soil particles through a 2 mm sieve.

SOM is either analysis in terms of quantity, or of quality depending on the final objectives.

### Quantitative Analyses

SOM content is highly variable in soils (variations with depth and soil type). Major elements are C, H, O, N, P, and S. Organic carbon (C) is the major component of SOM. It is difficult to quantitatively

estimate the amount of OM present in a soil. Procedures used in the past involved determination of the change in weight of a soil sample resulting from destruction of organic compounds by H<sub>2</sub>O<sub>2</sub> treatment or by ignition at high temperature. Both techniques are subject to error. The H<sub>2</sub>O<sub>2</sub> method does not quantitatively remove organic matter, and the ignition method gives an overestimate because both inorganic and organic constituents lose weight during heating. Alternatively, SOM quantity is assessed through organic C content: SOM = organic C × 1.724. This last coefficient, called the van Bemmelen factor, is used by assuming that SOM contains 58% organic C. It is clear that it is only an approximation, as the proportion of organic C in SOM is highly variable. Published organic C–organic matter conversion factors for surface soils have varied from 1.5 to 2.5. The appropriate factor must be experimentally determined for each soil. Although neither the direct determination of organic matter nor the calculation of organic matter content is completely accurate, the best procedures currently available are described here. Because of the problem associated with determining the organic matter content of a soil, it is strongly suggested that investigators determine and report the organic C concentration as a measure of the organic matter in a soil.

### Total Carbon in Soil

Total C in soils is the sum of both organic and inorganic C. Most organic C is present in the soil organic matter fraction, whereas inorganic C is largely found in carbonate minerals. Not all soils contain inorganic C because of dissolution during soil formation of carbonate minerals originally present in parent material. However, organic C is present in all agricultural soils. In soils formed from calcareous parent material under arid conditions, it is not unusual for the inorganic C concentration to exceed the amount of organic C present. Organic C is mainly in the soil organic fraction, which consists of the cells of microorganisms, plant, and animal

residues at various stages of decomposition, stable 'humus' synthesized from residues, and highly carbonized compounds such as charcoal, graphite, and coal (elemental forms of C). Organic C in soil may be estimated as the difference between total C and inorganic C. Organic C can be determined directly by total C procedures after removal of inorganic C or by rapid dichromate oxidation–titration techniques. Procedures involving a total C analysis generally recover all forms of organic C in soils, whereas dichromate oxidation procedures recover variable proportions of elemental C (e.g., charcoal) and, in some procedures, variable amounts of organic C contained in 'humus'. Calcite and dolomite are the principal carbonate minerals present in soil, and most inorganic C is associated with these compounds. However, in some alkaline soils, significant amounts of inorganic C may be present as soluble carbonate –  $\text{CO}_3^{2-}$  and bicarbonate –  $\text{HCO}_3^-$  salts.

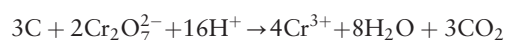
Total C determinations for soil involve conversion of all forms of C in soils to  $\text{CO}_2$ , by wet or dry combustion and subsequent quantification of evolved  $\text{CO}_2$ , by gravimetric, titrimetric, volumetric, spectrophotometric, or gas chromatographic techniques. Dry combustion is carried out by heating a soil–catalyst mixture in a resistance furnace or induction furnace in a stream of  $\text{O}_2$ , or  $\text{CO}_2$ -free air, followed by quantification of the evolved  $\text{CO}_2$ . Wet combustion is normally carried out by boiling a soil sample with a mixture of  $\text{K}_2\text{Cr}_2\text{O}_7$ ,  $\text{H}_2\text{SO}_4$ , and  $\text{H}_3\text{PO}_4$  in a closed system flushed with a stream of  $\text{CO}_2$ -free air and absorption of the evolved  $\text{CO}_2$ , in a tarred weighing bulb filled with Ascarite. Alternatively, wet combustion may be carried out in a van Slyke–Neil apparatus and evolved  $\text{CO}_2$ , estimated by manometric procedures.

Analytical procedures used for determining total C in soils must quantify both inorganic and organic forms. In humid regions where extensive leaching of the soil profile has occurred, organic C will be the predominant form present. In arid and semiarid regions, carbonate minerals (e.g., calcite, dolomite) along with soluble carbonate salts will constitute a significant percentage of the total C. Two basic approaches are used to quantify total C in soils, namely, dry combustion and wet combustion. In both instances, the  $\text{CO}_2$  liberated from organic and inorganic C is determined through volumetric, titrimetric, gravimetric, and conductimetric techniques. An apparatus for performing total C analysis by dry combustion can be fabricated from conventional laboratory glassware and a medium-temperature ( $\sim 1000^\circ\text{C}$ ) resistance furnace. Dry combustion procedures using either high-temperature ( $\sim 1500^\circ\text{C}$ ) or induction furnaces are most commonly found in

commercially available, automated total C analyzers. The majority of dry combustion methods employ gravimetric determination of  $\text{CO}_2$ . Wet combustion methods for total C employ a strong oxidant ( $\text{K}_2\text{Cr}_2\text{O}_7$ ) in an acidic digestion mixture for quantitative oxidation of organic C and dissolution of carbonate minerals.

### **Wet Oxidation**

Undoubtedly, the most widely used method worldwide, for historical and cost reasons, to determine organic C content is a wet oxidation of organic C in the presence of dichromate ( $\text{Cr}_2\text{O}_7^{2-}$ ) in highly acidic medium, and back-titration with ferrous sulfate of excess dichromate:



This method was proposed by Shollenberger in 1927 and then by Walkley and Black in 1934, who gave their names to this method. The oxidation occurs at a temperature of *c.*  $120^\circ\text{C}$  and unfortunately the recovery is incomplete, an average factor of 76% proposed by Walkley and Black is used. Other studies have showed that the recovery may vary between 60 and 86%. Therefore, it must be highlighted that this method gives only an approximate or semiquantitative estimate of C. Several authors have proposed modifications, mainly heating the reaction and most laboratories nowadays use  $150^\circ\text{C}$ .

In any case, all these modifications give low recovery (2–36%) of carbonized material (e.g., charcoal, soot) that may exist in soil due following combustion of the aboveground biomass.

### **Dry Combustion**

In combustion, all forms of C are converted to  $\text{CO}_2$  followed by a quantification of the evolved  $\text{CO}_2$ . In soils containing inorganic C, generally in the form of carbonates, an acidic pretreatment must be effected. Soil sample must be prepared in a fine powder ( $< 80$ – $100$  mesh). The sample is heated in an oven at  $1000$ – $1500^\circ\text{C}$  for quantitative recovery of C. Modern instruments perform simultaneous determination of C, N, and H and sometimes O and S. Analyses used microsamples of about 50–100 mg.

### **Spectroscopic Techniques**

Improving estimates of C in soil is currently hindered by lack of rapid, accurate, and precise analysis method. Two novel approaches based on spectroscopy are being developed and proposed even for *in situ* soil analysis.

Near-infrared reflectance spectroscopy (NIRS) is an analytical technique that correlates diffusely

reflected near-infrared radiation with the chemical and physical properties of materials. Near-infrared spectra (wavelengths in the 400–2498 nm range) are dominated by weak overtones and combinations of fundamental vibrational bands for H–C, H–N, and H–O bonds from the mid-infrared region. Soil minerals, organic matter, and moisture are the major components of soils, with distinct spectral features in the visible and near-infrared regions.

Analytical applications of NIRS have been explored since the early 1950s. However, owing to instrumental difficulties, its use as a quantitative approach started only 30 years later, together with the development and the widespread use of computers. As the coupling between equipment and microcomputers became easier and faster, obtaining and storing large amounts of data became possible.

The NIRS technique is rapid, effective, convenient, accurate, nondestructive, does not consume the sample, and does not present danger to the operator. Moreover, it provides a reduction in both reagent consumption, and the time needed for performing the analysis, in addition to permitting *in situ* analysis. These advantages confer particular characteristics on the technique, making it useful for routine soil analysis and online control in industrial processes.

In the NIRS technique, the spectra need to be processed by using either principal component regression (PCR), multiple linear regression (MLR), stepwise multiple linear regression (SMLR), maximum  $r^2$  improvement (MAXR) model, Fourier transform filter, multiplicative scattering correction (MSC), and partial least-squares regression (PLSR).

PLSR nowadays is a reference method for multivariate calibration and its utilization has overcome limitations in the use of multiple linear regressions. In the PLSR approach, the full spectrum is used to establish a linear regression model, where the significant information contained in the near-infrared spectra is concentrated in a few latent variables that are optimized to produce the best correlation with the desired property to be determined.

The second approach named laser-induced breakdown spectrometry (LIBS) is based on atomic emission spectroscopy. In this method, a laser is focused on a solid sample and forms a microplasma that emits light characteristics of the elemental composition of the sample. The emitted light is collected, spectrally resolved, and detected to monitor concentrations of elements via their unique spectral signatures. When calibrated, the method can also provide quantitative measurements.

The main advantage of LIBS is that it simplifies the conventional methodology by avoiding laborious chemical steps, e.g., the preparation and dissolution

of the soil sample. Further, this technique allows the direct chemical analysis of the solid sample by its vaporization achieved by interaction with a laser beam of adequate frequency and intensity. Because of the lack of pretreatment of the material, as well as the speed of analysis, not mentioning the possibility of *in situ* analysis, this technique offers an attractive solution for a wide range of applications.

This promising new method, however, has potential drawbacks, namely interference from Fe at approximately 248 nm. Additional testing of LIBS is required to understand the effects of soil properties such as texture, moisture content, and mineralogical composition (e.g., silicon content) on LIBS measurements.

## Qualitative Analyses

### Historical Review

The first attempt to isolate organic substance from soil appears to have been made by F.K. Achard in 1786, and T. de Saussure is usually credited for introducing the term ‘humus’ (Latin equivalent of soil) to describe the dark-colored organic material in soil. Several researchers tried to find out the chemical nature of humus. By the end of the nineteenth century, it had been firmly established that humus was a complex mixture of organic substances that were mostly colloidal in nature and which had weakly acidic properties.

During the first half of the twentieth century renewed efforts were deployed to classify organic substances and to determine their chemical nature and structure. Most of the approaches proposed classification based on solubility characteristics: humic acid (soluble in alkali, insoluble in acid), fulvic acid (soluble in alkali and acid), hyalomelanin acid (alcohol-soluble part of humic acid), humin (insoluble in alkali).

The modern-day concept recognizes that soil organic matter includes a broad spectrum of organic constituents with two major types: on one hand, compounds belonging to well-known classes of organic chemistry such as amino acids, carbohydrates, and lipids, and on the other hand, a series of high-molecular-weight compounds formed by secondary synthesis.

The International Humic Substance Society was founded in 1981 to bring together scientists in the coal, soil, and water sciences with interests in humic substances. Recommended methods are available at [www.ihss.gatech.edu](http://www.ihss.gatech.edu).

More recently, soil scientists proposed the separation of SOM in physical fractions according to different sizes and/or densities. These methods intend to

fractionate SOM into pools of different turnover times and bioavailability as they split SOM into free particulate OM and OM associated with soils minerals. Most physical fractionation procedures include different types of ultrasonic treatments for dispersion. Although there is no single fractionation procedure applicable for all soil types, the fraction sizes separated are commonly 200, 50, 2, and 0.2  $\mu\text{m}$ . However, a major concern remains the lack of standardization of ultrasonic energies and time of application.

The advantage of this approach resides in its concept that modeled SOM into compartments. This concept has been and is still actively being used to develop and calibrate mechanistic models of SOM dynamics.

The methods that have been used for obtaining information on the chemical structure can be divided into two groups, degradative and nondegradative methods.

### **Degradative Methods**

Degradative methods include oxidative and reductive degradation, hydrolysis, various types of irradiation, thermal analysis, and biological degradation. Of these, the oxidative method is the most informative. Several reagents have been used for the oxidation of humic substances, such as potassium permanganate, copper(II) oxide, aqueous chloride (under alkaline conditions); nitric acid, peracetic acid, hydrogen peroxide, ultraviolet radiation (under acidic conditions). Premethylation is important for obtaining good yields. Alkaline permanganate oxidation seems to be the best method, from the standpoints of producing relatively large amounts of identifiable digestion products and providing significant information on the chemical structures of humic substances. All oxidative degradations of soil substances essentially produce aliphatic substances (alkanes, fatty acids, and aliphatic carboxylic acids), phenolic acids, and benzenecarboxylic acids. Soil humic acids and soil fulvic acids yield similar quantities of aliphatic compounds on oxidation, whereas the oxidation of soil fulvic acids yields more phenolic acids but less benzenecarboxylic acids in comparison to soil humic acids. Results obtained from degradative techniques must be treated with care in order to avoid overinterpretation. Several of the compounds identified in many digests could be artifacts of the reaction processes and conditions.

### **Nondegradative Methods**

Among the nondegradative methods employed are infrared and UV-visible spectroscopy, spectrofluorimetry, electron spin resonance spectroscopy, and

nuclear magnetic resonance (NMR) spectroscopy. Very little information is obtained from any of the other spectroscopic methods. The study of humic substances is the study of complicated, ill-defined mixtures. Most of the physical and chemical methods of investigation are confined to simple systems and depend on the analyst's skill and ability to interpret the data. In applying any spectroscopic method to humic substances, measurement is taken of the summation of the signals of the numerous components in the mixture which respond to that particular frequency range, with all signals superimposed upon each other. What is measured is the net or average response of these particular components in the assemblage. In the case of humic substances, deciphering this garbled message into chemically meaningful information is, at best, a formidable task.

### **Infrared Spectroscopy**

With a few exceptions, all infrared spectra of humic substances have been measured in dry solid samples, and the pressed-pellet method has been used almost exclusively. The spectrum of humic acid consists of relatively few bands that are very broad. The broadness of the bands reveals that humic substances are complex mixtures where a particular type of functional group can exist in a wide variety of chemical environments, each characterized by slightly different force constants for its bonds. Fourier-transform infrared (FTIR) spectroscopy allows the samples to be observed in their native wet state and avoids the shifts in chemical equilibrium which must necessarily accompany the drying process.

### **Raman Spectroscopy**

Attempts to measure the Raman spectra of humic substances have been unsuccessful to date: the relatively intense fluorescence exhibited by the humic substances obscures any Raman signal that may be present.

### **Ultraviolet-Visible Spectroscopy**

The UV-visible spectra of humic substances are generally featureless and the technique has little value for studying functionality of humic substances and cannot be used for the direct determination of functional groups in these materials. Nevertheless, UV-visible spectroscopy does have other useful applications such as the estimation of the degree of humification using  $E_4/E_6$  ratios (the ratio of the absorbance at 465 nm to that at 665 nm) and for determining the concentration of dissolved humic substances based on Beer's law plots.



### Spectrofluorimetry

Humic substance groups emit in a spectral region relatively free from interferences by other groups. Measurement of the fluorescence is a relatively selective method of analysis. Fluorescence cannot be used for the direct determination of functionality.

### Electron Spin Resonance Spectroscopy

The electron spin resonance (ESR) spectrum of humic and fulvic acids consists of a single line identified by its position and width. In general, the ESR spectra are devoid of hyperfine splitting. The ESR spectra of humic substances contain relatively little data and they result from only a small fraction of the total number of molecules that comprise the humic and fulvic acid. However, ESR yields some information on the nature of free radicals in humic substances. Comparison of the ESR properties among humic fractions reveals that fulvic acids contain greater quantities of free radicals than humic acids.

### Nuclear Magnetic Resonance Spectroscopy

This is the most promising nondegradative technique for humic structural analysis. This is primarily due to two fundamental instrumental advances: the Fourier transform technique and the cross-polarization/magic-angle spinning (CP/MAS) technique which allows the acquisition of  $^{13}\text{C}$  NMR spectra from solid samples. The  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of all humic substances are highly resolved and exhibit several peaks.

Although SOM has been studied extensively by spectroscopic methods, only recently have investigators turned to NMR techniques to understand its structure. This is due largely to the fact that in the late 1960s and early 1970s problems associated with the low sensitivity of  $^{13}\text{C}$  NMR spectroscopy were overcome. The technique of pulse Fourier transform spectroscopy uses multichannel excitation to allow many scans to be recorded and averaged in a comparatively short time. Much higher signal-to-noise ratios may be obtained than with conventional methods and this has allowed otherwise unobtainable spectra of complex mixtures to be recorded.

NMR is now useful in measuring a range of structural groups in soils and soil extracts, both in solution and the solid state. Studies are not solely confined to the  $^{13}\text{C}$  nucleus.  $^1\text{H}$ ,  $^{29}\text{Si}$ ,  $^{27}\text{Al}$ ,  $^{15}\text{N}$ , and  $^{31}\text{P}$  can also be studied in soils and information on functional group content can be obtained readily. However, operation of a new-generation NMR spectrometer requires a high degree of skill and training. The soil chemist does not normally have this background and will require the NMR specialist to produce a spectrum.

Theoretically, it should be possible to study any element present in a soil by NMR. However, there are a number of limitations. Some nuclei are intrinsically insensitive and if present in only trace concentrations, cannot be detected. This is particularly troublesome for elements for which rare nuclides have to be studied. A further problem is that the width of the lines in the NMR spectra can be greater than the differences between individual resonance frequencies so that whatever the structure, only one broad line is obtained. As already noted, to date only a few nuclei ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{27}\text{Al}$ ,  $^{29}\text{Si}$ ,  $^{15}\text{N}$ ,  $^{31}\text{P}$ ) have been studied successfully in soils, but others are almost certain to be exploited in the future.

### Isotopic Ratio

The isotopic ratio provides information on the origin of humic substances. During biochemical and chemical reactions, the isotopes of a given element undergo fractionation, which depends not only on their mass, but also on the nature of the reaction. Isotopic ratios, in particular  $^{13}\text{C}/^{12}\text{C}$ , depend on the organic matter evolution and enable the origin of the organic matter to be specified to some degree. The humic substances (primarily as fulvic acids) in deep ocean waters have been established to be 3000–6000 years BP with a  $\delta^{13}\text{C}$  of  $-22$  to  $-23\text{‰}$ . All stream humic substances analyzed to date have  $^{14}\text{C}$  ages near zero (all modern carbon) with  $\delta^{13}\text{C}$  values within  $-25$  and  $-28\text{‰}$ . The  $^{14}\text{C}$  and  $\delta^{13}\text{C}$  values for soil humic substances vary considerably. The  $^{14}\text{C}$  age of humic acids in soils varies from a few hundred to several thousand years;  $\delta^{13}\text{C}$  values range from  $-25$  to  $-31\text{‰}$ . Such range for soils may reflect cropping practices and residue management, the methods used for isolation of fulvic and humic acids and sampling techniques rather than true pedogenic differences.

### Structural Models

Structural representations of humic and fulvic acids are purely hypothetical and should just be considered as an attempt to rationalize chemical behavior. Several structural models have been proposed based on data gathered from degradative and nondegradative models. The different proposed models should be considered as complementary since each emphasizes certain particular properties. Briefly, humic acids are macromolecular and complex, and composed of substituted aromatic and aliphatic hydrocarbon core materials. It is probable that some aliphatic and aromatic ring compounds are heterocyclic with N, O, or S as the heteroatom. Segments or some side-chain components in the structures are substantially hydrophobic. In general, however, the



primary structures contain significant amounts of polar substituents, incorporating oxygen in a variety of functional groups. Peptide and saccharide materials can be linked covalently to the core. The following observations can be made:

1. Fulvic acids are generally characterized by their lower levels of total and aromatic carbon.
2. The oxygenated sites, especially carboxylates, are more numerous in fulvic than in humic acids. Studies using NMR seem to indicate that this excess of carboxylates is mostly aliphatic in nature.
3. The free radical content is appreciably lower for fulvic acids than for humic acids.
4. The degree of aromaticity of soil humic and fulvic acids has been estimated to be 70% on the basis of the analysis of the products of their oxidative degradation. There is little information on the degree of aromaticity of water fulvic acids, the autochthonous fraction being more aliphatic.

**See also:** **Geochemistry:** Inorganic; Sediment; Soil, Major Inorganic Components; Soil, Minor Inorganic Components. **Infrared Spectroscopy:** Overview. **Nuclear Magnetic Resonance Spectroscopy:** Overview.

## Further Reading

Creaser CS, and Davies AMC (eds.) (1998) Section 1: Near infrared spectroscopy. In: *Analytical Applications of*

*Spectroscopy*, pp. 1–172 London: Royal Society of Chemistry.

Essington ME (2003) *Soil and Chemistry: An Integrative Approach*. Boca Raton, FL: CRC Press.

Ghabbour EA and Davies G (2001) *Humic Substances: Structures, Models and Functions*. Cambridge: Royal Society of Chemistry.

Kögel-Knaber I (2000) Analytical approaches for characterizing soil organic matter. *Organic Geochemistry* 31: 609–625.

Pansu M and Gautheyrou J (2003) *L'analyse du sol: minéralogie, organique et minérale*. Paris: Springer (in French).

Piccolo A (1996) *Humic Substances in Terrestrial Ecosystems*. Amsterdam: Elsevier.

Reeves J III, McCarty G, and Meisenger JJ (1999) Near infrared reflectance spectroscopy for the analysis of agricultural soils. *Journal of Near Infrared Spectroscopy* 7: 179–193.

Sparks DL, Page AL, Helmke PA, et al. (1996) *Methods of Soil Analysis: Part 3 – Chemical Methods*. Madison, WI: Soil Science Society of Agronomy and American Society of Agronomy.

Stevenson FJ (1994) *Humus Chemistry: Genesis, Composition, Reactions*, 2nd edn. New York: Wiley.

Tan KH (1998) *Principles of Soil Chemistry*, 3rd edn. New York: Dekker.

Tognoni E, Palleschi V, Corsi M, and Cristoforetti G (2002) Quantitative micro-analysis by laser-induced breakdown spectroscopy: a review of the experimental approaches. *Spectrochimica Acta Part B* 57: 1115–1130.

# GLASSES

**M P Syomov and O V Rakhimova**, IV Grebenshchikov Institute of Silicate Chemistry, St. Petersburg, Russia

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by M Palaček, pp. 1943–1947, © 1995, Elsevier Ltd.

## Introduction

Glass is a product of cooling a melt that becomes solid without crystallization. This process results in the formation of a continuous polymeric network that includes structural units of the network-forming component(s) (network formers). This network is irregular and unsymmetrical. Among network formers, the so-called glass formers are marked out. These compounds are able to form glasses without other components being added. Other important constituents of glasses are the so-called modifiers. These

substances do not build into the glass network. They either facilitate the transition of the main component to the glassy state (vitreous state) or modify significantly the properties of glasses.

Glasses are an instance of amorphous solids. Most glasses are transparent or translucent.

## Typical Constituents of Glasses

Glasses can be divided into various types according to their chemical composition. Both inorganic and organic substances can be used to produce glasses.

Organic glasses are polymers with an irregular sequence of monomers (as in polystyrene) or polymers with such complex molecules that they cannot form a compact regular network under cooling. Organic glasses are also known as optical polymers.

Inorganic glasses can be divided into two main groups. Simple glasses such as metallic, halide,

primary structures contain significant amounts of polar substituents, incorporating oxygen in a variety of functional groups. Peptide and saccharide materials can be linked covalently to the core. The following observations can be made:

1. Fulvic acids are generally characterized by their lower levels of total and aromatic carbon.
2. The oxygenated sites, especially carboxylates, are more numerous in fulvic than in humic acids. Studies using NMR seem to indicate that this excess of carboxylates is mostly aliphatic in nature.
3. The free radical content is appreciably lower for fulvic acids than for humic acids.
4. The degree of aromaticity of soil humic and fulvic acids has been estimated to be 70% on the basis of the analysis of the products of their oxidative degradation. There is little information on the degree of aromaticity of water fulvic acids, the autochthonous fraction being more aliphatic.

**See also:** **Geochemistry:** Inorganic; Sediment; Soil, Major Inorganic Components; Soil, Minor Inorganic Components. **Infrared Spectroscopy:** Overview. **Nuclear Magnetic Resonance Spectroscopy:** Overview.

## Further Reading

Creaser CS, and Davies AMC (eds.) (1998) Section 1: Near infrared spectroscopy. In: *Analytical Applications of*

*Spectroscopy*, pp. 1–172 London: Royal Society of Chemistry.

Essington ME (2003) *Soil and Chemistry: An Integrative Approach*. Boca Raton, FL: CRC Press.

Ghabbour EA and Davies G (2001) *Humic Substances: Structures, Models and Functions*. Cambridge: Royal Society of Chemistry.

Kögel-Knaber I (2000) Analytical approaches for characterizing soil organic matter. *Organic Geochemistry* 31: 609–625.

Pansu M and Gautheyrou J (2003) *L'analyse du sol: minéralogie, organique et minérale*. Paris: Springer (in French).

Piccolo A (1996) *Humic Substances in Terrestrial Ecosystems*. Amsterdam: Elsevier.

Reeves J III, McCarty G, and Meisenger JJ (1999) Near infrared reflectance spectroscopy for the analysis of agricultural soils. *Journal of Near Infrared Spectroscopy* 7: 179–193.

Sparks DL, Page AL, Helmke PA, et al. (1996) *Methods of Soil Analysis: Part 3 – Chemical Methods*. Madison, WI: Soil Science Society of Agronomy and American Society of Agronomy.

Stevenson FJ (1994) *Humus Chemistry: Genesis, Composition, Reactions*, 2nd edn. New York: Wiley.

Tan KH (1998) *Principles of Soil Chemistry*, 3rd edn. New York: Dekker.

Tognoni E, Palleschi V, Corsi M, and Cristoforetti G (2002) Quantitative micro-analysis by laser-induced breakdown spectroscopy: a review of the experimental approaches. *Spectrochimica Acta Part B* 57: 1115–1130.

# GLASSES

**M P Syomov and O V Rakhimova**, IV Grebenshchikov Institute of Silicate Chemistry, St. Petersburg, Russia

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by M Palaček, pp. 1943–1947, © 1995, Elsevier Ltd.

## Introduction

Glass is a product of cooling a melt that becomes solid without crystallization. This process results in the formation of a continuous polymeric network that includes structural units of the network-forming component(s) (network formers). This network is irregular and unsymmetrical. Among network formers, the so-called glass formers are marked out. These compounds are able to form glasses without other components being added. Other important constituents of glasses are the so-called modifiers. These

substances do not build into the glass network. They either facilitate the transition of the main component to the glassy state (vitreous state) or modify significantly the properties of glasses.

Glasses are an instance of amorphous solids. Most glasses are transparent or translucent.

## Typical Constituents of Glasses

Glasses can be divided into various types according to their chemical composition. Both inorganic and organic substances can be used to produce glasses.

Organic glasses are polymers with an irregular sequence of monomers (as in polystyrene) or polymers with such complex molecules that they cannot form a compact regular network under cooling. Organic glasses are also known as optical polymers.

Inorganic glasses can be divided into two main groups. Simple glasses such as metallic, halide,

chalcogenide, and oxide glasses are formed by components of the same chemical origin. Most common silicate glasses are examples of oxide glasses. Complex glasses such as oxychalcogenide or metal-oxide glasses are formed by components belonging to different types of chemical compounds.

As, B, C, Ge, P, S, Se, Si, and Te can form elemental glasses. Among such glasses, the most important are noncrystalline silicon glass, used as the working substance of solar batteries, vitreous carbon glass, used to produce crucibles and protective layers, and selenium glass, used as a photosensitive material in copying techniques.

Metallic glasses can be produced under very rapid quenching of the melt; the quenching rate can be up to  $100\text{ K s}^{-1}$ . Metallic glasses are also known as amorphous metals. Al, Cu, Pb, and Sn form metallic glasses most easily. To make glasses on a base of Ni, Fe, Co, Mn, or Cr, it is necessary to add modifiers such as B, C, P, S, Si, or others. The structure of metallic glasses is characterized by a dense random packing of atoms. Amorphous metals are used in practice mostly as high-quality soft-magnetic materials.

Fluorides of some metals are practically important for halide glasses. Fluoroaluminate and fluorozirconate glasses are examples of oxyhalide glasses. They are transparent in the infrared (IR) range and are candidate materials for producing optical waveguides with an extremely low level of light losses.

Chalcogenide glasses are formed by sulfides, selenides, and tellurides of some elements, for example  $\text{B}_2\text{S}_3$ ,  $\text{SiS}_2$ ,  $\text{Sb}_2\text{S}_3$ ,  $\text{P}_4\text{Se}_4$ ,  $\text{B}_2\text{Se}_3$ , and  $\text{As}_2\text{Te}_3$ . They are used as IR-transparent media, semiconductors, and materials for ion-selective electrodes.

Oxide glasses are formed from the oxides of the type  $\text{R}_2\text{O}_3$  ( $\text{B}_2\text{O}_3$ ,  $\text{As}_2\text{O}_3$ ,  $\text{Al}_2\text{O}_3$ ,  $\text{Sb}_2\text{O}_3$ ),  $\text{RO}_2$  ( $\text{SiO}_2$ ,  $\text{GeO}_2$ ,  $\text{TiO}_2$ ), and  $\text{R}_2\text{O}_5$  ( $\text{P}_2\text{O}_5$ ). Of special importance is silica glass.  $\text{SiO}_2$  forms structural units of  $\text{SiO}_4$  in which the silicon atom is placed in the center of a tetrahedron and oxygen atoms occupy the vertices. One oxygen bond is directed to the central atom, and the second may be directed to another tetrahedron with a central silicon atom. Silica glass has some unique properties. It cannot be deformed up to  $1000^\circ\text{C}$ , it is thermal shock resistant, it displays high bulk and surface electrical resistivity, and it is ultraviolet (UV)-visible transparent, though it has a high melting and can hardly be transformed into glassware. The high cost of silica glass limits its application to special glassware such as in optical components operating at high temperatures, in fiber-optics communication, and in chemical glassware.

Silica is the major component of silicate glasses, which constitute more than 98% of commercial glasses. Silicate glasses include silica together with

other network-forming oxides. They are used widely in building, transport, instrument making, decorative art, etc.

Silicate glasses contain modifiers such as Na, K, Li, Ca, Mg, Ba, and Zn ions that affect the properties of glasses. In addition to the constituents added to glasses to influence their properties, other constituents are added for melting and refining ( $\text{F}^-$ ,  $\text{As}_2\text{O}_3$ ,  $\text{Sb}_2\text{O}_3$ ,  $\text{SO}_3$ ) or for coloring ( $\text{Cr}_2\text{O}_3$ ) or decolorizing ( $\text{CoO}$ ,  $\text{NiO}$ , Se). Glasses always include constituents that are transferred as a result of corrosion of the refractory materials of the pots or ovens in which the glass is melted ( $\text{Al}_2\text{O}_3$ ,  $\text{ZrO}_2$ ) or are presented as impurities in the raw materials used to make the glass ( $\text{Fe}_2\text{O}_3$ ).

The oldest and most common are soda-lime-silica glasses, which contain 12–16% alkali (mainly  $\text{Na}_2\text{O}$ ), 10–12% alkaline-earth oxides ( $\text{CaO}$ ,  $\text{MgO}$ ), and  $\sim 71\%$   $\text{SiO}_2$ . This type includes window glasses, container glasses, glass bricks, and certain types of glass fibers.

Borosilicate glasses are characterized by the presence of a greater amount ( $>80\%$ ) of  $\text{SiO}_2$  (ISO 33853). These glasses have good hydrolytic and thermal shock resistance and are used for large-scale technical applications in the chemical, pharmaceutical, and food industries and for laboratory apparatus. Alkali lead silicate glasses usually contain  $\text{PbO}$  ( $>10\%$ ),  $\text{SiO}_2$  (54–58%) and alkali (14%, mainly  $\text{K}_2\text{O}$ ). Glasses of similar composition are used for making lead crystal as, for instance, in jewelry and expensive drinking glasses (ISO 7086). Lead alkali silica glasses with  $>30\%$   $\text{PbO}$  have good electrical resistance properties and are therefore used in electro-technology, in lamp stems, in cathode-ray tubes, etc. and are also useful in components used to provide protection from X-rays.

In crystal glasses,  $\text{BaO}$  may substitute part of the  $\text{PbO}$ . Optical glasses are a special subgroup of lead glass.

## Recommendations and Standards

Various recommended procedures have been published by, for example, The International Commission on Glass (in *Glass Technology*) and the American Society of Testing Materials, and in British Standards and German Standards (DIN).

## Analysis of Oxide Glasses

### Samples and Sampling

The hazards of sampling commercial glasses are greatly reduced by the fact that glass as a product of

large-scale melts is generally quite homogeneous. However, care must be taken that the glass sample taken for analysis does not contain faults (striae and stones). Even if the sample seems to be homogeneous, it is possible, particularly for products made from a glass containing constituents that volatilize at higher temperatures ( $B_2O_3$ ,  $PbO$ ), that significant differences in composition are observed in different parts of the product.

Samples are crushed to grains of about 5 mm, which are then combined, ground further, and quartered to provide an analytical sample of about 20 g that is then ground to the fineness required for subsequent decomposition. To minimize pick-up of water and carbon dioxide as well as oxidation of some constituents such as iron(II), coarse grinding (60–80 mesh) is recommended, particularly if acid decomposition is to be used. If fusion is planned, a much finer grinding (200–400) mesh is to be preferred.

In order to get correct data on the quantitative content of the glass components in the further course of the analysis, it is first required to determine the humidity of the sample and its weight loss after calcination. The glass sample is placed in the platinum crucible and calcined in the muffle kiln at  $110^\circ\text{C}$  to determine the humidity and then at  $1100$ – $1200^\circ\text{C}$  to measure the weight loss after calcination.

## Decomposition

**Dissolution with hydrofluoric acid** All silicate glasses are decomposed under attack with a mixture of hydrofluoric and other mineral acids. Hydrofluoric acid is very effective because, even at room temperature, hexafluorosilicic acid is formed, which is decomposed at higher temperatures into silicon tetrafluoride and hydrofluoric acid. Both these products are removed from the reaction mixture by evaporation. Some other fluorides ( $BF_3$ ,  $AsF_3$ ,  $SeF_4$ ) are also volatilized.

Glass network modifiers form fluorides and complex fluorides. Heating with a mineral acid such as sulfuric or perchloric acid before analysis will decompose the fluorides formed. Perchloric acid has an advantage because perchlorates are very water soluble, but in the presence of some constituents ( $TiO_2$ ,  $ZrO_2$ ,  $Sb_2O_3$ ), sulfuric acid is preferred. Sulfuric acid cannot be used in the presence of constituents forming insoluble sulfates ( $PbSO_4$ ,  $BaSO_4$ ).

For decomposition of a number of fluorides, it is essential that the fuming step, commonly accompanied by rinsing of the vessel's walls with water, must be thorough and repeated not less than three times. The greater the content of metal(III) and metal(IV)

oxides, the more tenaciously is fluoride likely to be retained in the sample.

The decomposition of glasses with hydrofluoric and perchloric or sulfuric acids is carried out customarily in platinum dishes of  $\sim 100$  ml capacity. Unless hydrogen sulfide, free chlorine, or nitrous oxide is evolved during the decomposition, the corrosion of platinum is negligible.

The sample solution obtained this way serves for the determination of alumina ( $Al_2O_3$ ), iron oxide ( $Fe_2O_3$ ), titania ( $TiO_2$ ), lime ( $CaO$ ), magnesia ( $MgO$ ), and alkali metal oxides.

**Decomposition by alkaline fusion** Fusion with an alkali metal carbonate, usually anhydrous sodium or potassium carbonate or their mixture, decomposes glass samples very effectively. This occurs above the melting point of the carbonate(s):  $Na_2CO_3$ ,  $853^\circ\text{C}$ ;  $K_2CO_3$ ,  $903^\circ\text{C}$ ;  $KNaCO_3$ ,  $712^\circ\text{C}$ . If it is known that the glass does not contain boron, the fusion sample is fused with a mixture of sodium carbonate and borax to decrease the fusion temperature. The complex silicate network is broken down, and monosilicates are formed.

Other anions present are released, and network modifiers form salts corresponding to the acid used finally to dissolve the cold fusion mixture. This reaction is accompanied by evolution of carbon dioxide.

The fusion is carried out in platinum dishes or crucibles. The amounts of the sample and flux depend on the composition of the glass. Customarily, for one part of glass, two to four parts of anhydrous sodium carbonate are used. After dissolution in mineral acid, the solution can be analyzed using various analytical procedures.

**Decomposition for determination of constituents using atomic absorption spectrometry** Weigh 0.1 g of the sample into a small platinum dish. Moisten with a few milliliters of water. Add 2–3 ml of perchloric acid and 5–10 ml of hydrofluoric acid. Transfer the dish to an electrically heated and regulated hot plate, starting at a low temperature. Raise the temperature gradually and evaporate to a slightly moist residue, taking care to achieve a slow rate of decomposition of fluorides because at higher temperatures the decomposition of fluorides is violent and spraying losses may occur. When the perchloric acid has evaporated and the residue is slightly moist, remove the dish, cool, add 1 ml of perchloric acid, rinse the dish walls, and evaporate again.

Rinsing and evaporating with perchloric acid is repeated once or twice for proper removal of fluorides. Moisten the residue with 0.5 ml of perchloric acid, add 25 ml of water, and digest on a hot plate at



~80°C for 10–15 min. Transfer the dissolved residue to a 100 ml volumetric flask, cool, and dilute to the mark. With this procedure a solution of  $1000\times$  dilution is prepared. Other dilutions are prepared from this solution.

### Atomic Spectrometry and Other Instrumental Methods

Modern instrumental methods are used widely in analysis of glasses. First of all, atomic spectrometry and mass spectrometry (MS) should be mentioned.

Qualitative and semiquantitative analysis of glasses having complex composition is performed by atomic emission spectrometry (AES), with the excitation of the spectrum in the direct-current arc. This implementation of AES does not require dissolving the sample and enables one to detect 10–15 elements simultaneously with a detection limit of  $10^{-2}$ – $10^{-4}\%$ . Unfortunately, because of its low precision (10–20%), this method is not suitable for determination of the content of the basic components of glass (glass formers and modifiers). A more universal technique is AES with an inductively coupled plasma (ICP) as a source of atomization and excitation. ICP–AES provides higher detection limits ( $10^{-2}$ – $10^{-8}\%$  for different elements) and precision (1–5%), as well as a wider range of the determined concentrations.

X-ray fluorescence spectrometry (XRF) can be used to determine the content of the basic components of glasses. This multielement analytical technique also does not require dissolving the sample. For a content of 1–100% the precision of the method is not worse than 5%.

Flame atomic absorption spectrometry (AAS) and flame AES are the most commonly used analytical instrumental techniques in analysis of glasses.

The present applications of flame AAS for glass analysis are usually limited to the minor and trace constituents (see Table 1). Although the usual

precision of AAS is 1–2%, by taking special care, a better precision can be achieved as is required for glass analysis. There are some problems with the storage of samples and calibration solutions. They must be stored in the refrigerator below 5°C in polypropylene flasks. After 2–4 months the solutions must be replaced. Great attention must be paid to preparation of the solutions and the quality of the volumetric flasks. Volumes less than 5 ml should not be measured with a pipette; microburettes are recommended. Glass volumetric flasks should not be used as storage flasks because of problems arising from leaching from the glass surface. After measurement the solutions must be transferred to polypropylene flasks or discarded.

Chemical interference occurs, and to suppress such interferences a spectrochemical buffer solution equivalent in amount to 10% of the total sample solution volume is added. Such a buffer contains 1500 µg of lanthanum and 100 µg of cesium in a 1 ml volume.

For evaluation, a direct calibration or a bracketing technique is used. The direct calibration technique is a method for determining the concentration of analytes in the sample solution by reference to a calibration graph produced from a set of calibration solutions having different analyte concentrations (single element calibration solutions) or from a set of solutions having different concentrations of analytes and other elements (multielement calibration solutions). Special care must be given to dilution of the sample and calibration solutions. For normal work a dilution of 0.1 g in 100 ml (1000 times dilution) is used. Calibrating solutions must contain the same concentration of acid as used for dissolving evaporated residues.

Flame AES is used commonly to determine alkali and alkali earth elements in oxide glasses, using the parameters shown in Table 2.

**Table 1** Conditions for determination of elements in glasses by atomic absorption spectrometry

Oxide	Wavelength (nm)	Flame <sup>a</sup>	Maximum linear response range (µg ml <sup>-1</sup> )	Concentration range in glasses (mass%)	Concentration in solution analyzed after 1000-fold dilution (µg ml <sup>-1</sup> )
Al <sub>2</sub> O <sub>3</sub>	309.3	2-red	60	0.1–20	1–200
BaO	553.6	2-red	10	0.1–10	1–150
Fe <sub>2</sub> O <sub>3</sub>	248.3	1-ox	5	0.01–5	0.1–50
K <sub>2</sub> O	766.5	1-ox	2	0.05–15	0.05–150
MgO	285.2	1-stoich	0.4	0.01–5	0.1–50
MnO	279.5	1-ox	3	0.05–5	0.5–50
Na <sub>2</sub> O	589.0	1-ox	1	0.05–20	0.5–200
PbO	217.0	1-ox	15	0.1–30	1–300
ZnO	213.9	1-ox	1	0.05–10	0.5–100

<sup>a</sup>Flames: 1, acetylene–air; 2, nitrous oxide–acetylene; ox, oxidizing; red, reducing; stoich, stoichiometric.



**Table 2** Conditions for determination of alkali metals in glasses using flame atomic emission spectrometry

	$\text{Na}_2\text{O}$	$\text{K}_2\text{O}$	$\text{Li}_2\text{O}$
Range ( $\text{mg ml}^{-1}$ )	0.025–5	0.025–5	0.05–2
Wavelength (nm)	589.0	766.5	670.8
Flame		Acetylene–air or propane–air	

When analyzing elemental glasses, it is often necessary to determine ultratrace amounts of the elements since practical use of these glasses requires high purity of the material. For example, high purity silica glass used in optical waveguides must include no more than  $10^{-4}$  mass% of most elements. In the analysis of such glasses standard preconcentration techniques are used with subsequent trace analysis of the elements in the concentrate using instrumental methods such as AES, MS, AAS, and XRF. Flame AAS is used most commonly, but techniques such as ICP–MS offer the potential to provide multielement capability and lower detection limits.

The distinctive feature of the above instrumental methods is that they are indirect analytical techniques. They require preliminary calibration using standard samples. The standards must be very close to the analyzed samples in both total chemical composition and physical properties. In this regard these methods are mostly useful for controlling glass production or for analysis of large series of similar samples.

### Determination of Basic Components

**Silica** Silica is the main constituent of the majority of glasses. Successful determination of silica is essential for analysis of glass. This is usually achieved by fusion with sodium carbonate or a mixture of sodium carbonate and borax (2:1) with subsequent dissolution of the cooled melt in 0.1 N hydrochloric acid to avoid polymerization of silica.

In the classical method, the solution is evaporated to dryness, soluble salts are dissolved in hydrochloric acid and hot water, and the system is filtered. The precipitate is transferred to a platinum dish, which is heated to 1100–1200°C and then weighed. The residue is then treated with hydrofluoric and sulfuric acids and evaporated to fumes of  $\text{SO}_3$  and then to dryness and weighed. This process volatilizes silicon as  $\text{SiF}_4$ . The difference between weighings represents the amount of silica.

In modern analytical technique, it is considered that more simple and reliable methods are those using silicomolybdic heteropoly acids. Silica is determined using either a gravimetric or spectrophotometric method. In the first case silica is determined as

silicomolybdic acid quinolate  $((\text{C}_9\text{H}_7\text{ON})_4\text{H}_4\text{SiMo}_{12}\text{O}_{40})$ . The prepared solution ( $\text{SiO}_2 \leq 10$  mg) is mixed under agitation with 10 ml of 10% aqueous solution of ammonium molybdate  $((\text{NH}_4)_2\text{MoO}_4)$  and then with 15 ml of a 2% aqueous solution of quinoline. The precipitate is coagulated in a water bath at 60–80°C, filtered through a Schott filter (grade 4), and dried at 140°C to get the constant weight.

Spectrophotometric determination is performed using UV–visible spectrophotometry with formation of yellow saturated silicomolybdate. The method is used normally for determination of 0.002–1 mg of  $\text{SiO}_2$ . For greater sensitivity (for 0.0001–0.25 mg of  $\text{SiO}_2$ ), the yellow silicomolybdate may be reduced with sodium sulfite or a mixture of citric and ascorbic acids (the solution must contain 5% citric acid and 1% ascorbic acid) to the corresponding silicomolybdenum blue.

Boric oxide increases the apparent amount of  $\text{SiO}_2$  found, but it can be removed before the evaporation is finished by addition of methanol to form volatile trimethylborate. The presence of fluoride decreases the amount of silica found. It can be masked by addition of aluminum nitrate hexahydrate; however, if the fluoride content is greater than 5%, a special procedure must be followed.

**Boric oxide** Boric oxide has a significant effect on the properties of glass and is therefore an important and useful glass constituent. Many other components of glass interfere in its determination; it is therefore necessary to use a separation procedure before the measurement stage, for example, by precipitation of certain metal hydroxides and carbonates.

**Titrimetric determination** The mannitol or glycerol acid–base titration procedure is applied to the dissolved sodium carbonate melt for the determination of milligram amounts of  $\text{B}_2\text{O}_3$ . A blank prepared from a glass that does not contain boric oxide should be subjected to the same procedure. There are certain interferences: the effect of fluoride is eliminated by adding calcium carbonate to the fusion mixture; lead and zinc can be filtered off after dissolution of the melt; phosphate is precipitated by addition of iron(III) chloride. The influence of some interference from such elements as aluminum or titanium can be eliminated by adding ethylenediaminetetraacetic acid (EDTA).

**Alumina** Alumina is almost always present in glasses in amounts varying from tenths of a percent to several percent; in some special glasses the concentration exceeds 10%. Aluminum may be determined gravimetrically as the hydrated oxide or by EDTA

titration. There is the disadvantage that in both methods other constituents ( $\text{Fe}_2\text{O}_3$ ,  $\text{TiO}_2$ ,  $\text{ZrO}_2$ ,  $\text{La}_2\text{O}_3$ ) are determined together with  $\text{Al}_2\text{O}_3$ ; the determined sum of these constituents is reported as sesquioxides,  $\text{R}_2\text{O}_3$ . The individual oxides must be determined separately and the amounts found subtracted from the total  $\text{R}_2\text{O}_3$ . The influence of  $\text{TiO}_2$  can be eliminated by adding  $\text{H}_2\text{O}_2$  to retain titanium in the ternary complex titania-EDTA- $\text{H}_2\text{O}_2$ .

**Lead oxide** Lead oxide is an important constituent of glasses. Other constituents, especially zinc oxide, barium oxide, and calcium oxide, influence the determination of lead oxide in various glasses. Separation of lead oxide using electrolysis to form lead dioxide on the anode and subsequent EDTA titration is a popular method, applied to 50–100 mg of  $\text{PbO}$  after nitric acid-hydrofluoric acid treatment of the glass in a platinum dish.

The solution after electrolytic separation can be used for determination of barium and zinc.

**Alkaline-earth metal oxides** Calcium and magnesium oxides are constituents of nearly all glasses and may be determined using EDTA titration. Neither determination is free of interference.

If the glass contains lead and/or barium oxide, the glass should be treated with a mixture of hydrofluoric and sulfuric acids and the insoluble lead and barium sulfates filtered off before the titration. Lead may also be precipitated with hydrogen sulfide. Small amounts of aluminum and iron may be masked with triethanolamine. Greater amounts of aluminum and iron must be pre-separated by precipitation as their hydrated oxides in the procedure for gravimetric determination of alumina. If the content of iron oxide, alumina, and titania is large, they can be separated using a 25% solution of urotropin.

**Alkali metal oxides** Precise and rapid determination of the alkali metals is essential for controlling glass production because these metal ions have a considerable influence on the properties of glass.

Along with the commonly used flame AES, sodium (50–100 mg  $\text{Na}_2\text{O}$ ) can also be determined gravimetrically through precipitation with zinc uranyl acetate. Potassium may be determined gravimetrically by precipitation with sodium tetraphenylborate. Separation of sodium and potassium in their combined presence can be performed by precipitation of  $\text{KClO}_4$  using perchloric acid. It is possible to determine  $\text{Li}_2\text{O}$  microchemically as hexamethylenetetramine hexacyanoferrate(II) and (III).

**Iron oxide** Iron oxide causes coloration in glass depending on its oxidation state. Iron(II) causes a bluish-green coloration, iron(III) a yellowish coloration. For determination of total iron expressed as  $\text{Fe}_2\text{O}_3$ , a UV-visible spectrophotometric method based on 1,10-phenanthroline is used. Hydroxylamine is used to ensure all the iron is present as iron(II).

## Peculiarities of Analysis of Other Types of Glasses

### Sampling

Standard analytical procedures are used for decomposition of prepared samples of elemental, metallic, halide, and chalcogenide glasses with subsequent determination of the main elements.

For example, sampling of chalcogenide glasses includes treatment of the probe with an aqueous solution of  $\text{NaOH}$  (0.1–17 N) under heating up to  $75^\circ\text{C}$ . The  $\text{NaOH}$  concentration and temperature are selected depending on the particular composition.

### Analysis

The analytical techniques used for metallic and polymer glasses are the same as those used to analyze metallic alloys and plastics. Analysis of chalcogenide glasses implies determination of the elements that form these glasses. The peculiarity of the analysis of halide glasses is the necessity to determine chlorine and fluorine.

**Determination of chlorine** Chlorine can be determined in two ways, titrimetrically and spectrophotometrically.

Titrimetric analysis of chlorine is based on the formation of  $\text{HgCl}_2$ . In order to exclude the influence of iron, chromium, and vanadium, 1–2 ml of a 5% solution of hydroxylamine sulfate is added. An ethanol solution of  $0.005 \text{ g ml}^{-1}$  diphenyl carbazone and bromophenol blue is used as an indicator.

Spectrophotometric determination of chlorine is performed at 340 nm as a complex, with a mixed solvent containing  $0.2 \text{ g ml}^{-1}$   $\text{Pb}(\text{NO}_3)_2$  and  $0.1 \text{ g ml}^{-1}$   $\text{AgNO}_3$ . The detection range is from  $0.06 \times 10^{-2}$  to  $0.14 \times 10^{-2} \text{ g ml}^{-1}$  chlorine.

**Determination of fluorine** After dissolution of the prepared glass, the sample fluorine is determined gravimetrically, by precipitating as  $\text{CaF}_2$ , or volumetrically, for example by argentometry after precipitation as  $\text{PbClF}$ , or by titration with a solution of a thorium salt in the presence of alizarin as an indicator. The aliquot of the analyzed solution must

contain not less than  $1.5 \text{ mg ml}^{-1} \text{ F}^-$ . The interference is from ions that either form stable fluoride complexes or precipitate  $\text{F}^-$ , such as Al, Be, Fe, Zr, Ba, and others. To eliminate their influence, it is required to pre-top fluorine from the analyzed solution as fluosilicic acid.

See also: **Atomic Absorption Spectrometry:** Principles and Instrumentation; Flame. **Atomic Emission Spectrometry:** Principles and Instrumentation.

## Further Reading

- ASTM. (2000) ASTM C169-92. *Standard Test Methods for Chemical Analysis of Soda-Lime and Borosilicate Glass*.
- BS(1955) BS2649 Part 1: *Recommended Procedure for the Analysis of Glass of Soda-Lime-Magnesia Type*.
- BS (1957) BS2649 Part 2: *Recommended Procedures for the Analysis of Soda-Boric Oxide-Silica Glass with High Silica and Boric Acid Content*.
- BS (1958) BS2649 Part 3: *Recommended Procedure for the Analysis of Potash Oxide-Lead Oxide-Silica Glass*.
- BS (1961) BS2649 Part 4: *Recommended Procedure for the Analysis of Fluoride Opal Glasses*.
- EEC (1969) Council Directive 69/493 on the approximation of the laws of the Member States to crystal glass. *Official Journal of the European Communities* 1: 326–336.
- German Standard DIN 52340. Testing of glass. Chemical analysis of colorless soda-lime glass.
- International Commission on Glass Technical Committee 2(1980) A report by the International Commission on Glass Technical Committee 2 Durability and Analysis. Parts 1–6. *Glass Technology* 21(1–6): 65–278.
- Palaček M and Peters A (1984) A report by the International Commission on Glass Technical Committee 2 Durability and Analysis Recommended methods for the determination of titania, alumina, lime, and magnesia in chrome green soda-lime-magnesia-silica glasses. *Glass Technology* 25(2): 55–61.
- Peters A, Fletcher WW, and Sales RM (1987) Analysis of glasses using atomic absorption and atomic emission spectrometry. Part 1. General considerations and results of collaborative investigations. *Glastechnische Berichte* 60(1): 11–20.

Peters A, Fletcher WW, and Sales RM (1987) Analysis of glasses using atomic absorption and atomic emission spectrometry. Part 2. Recommended procedures using the single buffer method with single- or multiple-element calibrations. *Glastechnische Berichte* 60(5): 163–173.

Zil'bershtein KhI (ed.) (1977) *Spectrochemical Analysis of Pure Substances*. Bristol: Adam Hilger Ltd.

## Glossary

Chalcogenide glasses	Glasses formed by sulfides, selenides, and/or tellurides.
Complex glasses (glasses of the mixed type)	Glasses formed by components belonging to different types of chemical compounds.
Elemental glasses	Glasses formed by chemical elements.
Glass	Product of cooling a melt that becomes solid without crystallization.
Glass former	Network former that can form glasses without adding other components.
Halide glasses	Glasses formed by halides (mostly fluorides and chlorides).
Metallic glasses (amorphous metals)	Glasses whose main components are metallic elements.
Modifier	Glass component that facilitates the transition of the main component to the glassy state or significantly modifies the properties of the glass.
Network former	Glass component whose structural units form the continuous glass network.
Organic glasses (optical polymers)	Glasses formed by organic polymers.
Simple glasses	Glasses formed by components belonging to the same type of chemical compound.

# GRAPHITE FURNACE

See **ATOMIC ABSORPTION SPECTROMETRY: Electrothermal**

# GRAVIMETRY

**R T Sane**, Therapeutic Drug Monitoring Laboratory, Mumbai, India

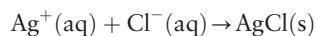
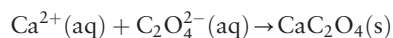
© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Gravimetric analysis is the quantitative estimate of a particular species by measurement of its weight. This is usually achieved by converting the analyte into an almost insoluble compound (precipitate reaction), which can then be weighed directly or can be converted into a stable form and then weighed. Other gravimetric procedures include measurements based on loss in weight of a sample (e.g., determination of moisture content) or gain in weight (e.g., increase in weight due to absorption of gases). Two other types of gravimetric analyses that deserve mention are electrogravimetry and thermogravimetry.

## Precipitation Methods

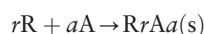
Many metallic elements in their ionic forms react with negative counterions to produce stable precipitates. Silver ions form stable and highly insoluble salts with chloride, bromide, and iodide. Calcium precipitates quantitatively with oxalate and can be measured reproducibly at any of the three temperature-dependent plateaus as the oxalate, the carbonate, and the oxide. Barium precipitates quantitatively as the sulfate. The reactions often follow the same pattern:



Positive and negative ions in an aqueous solution, otherwise soluble with the counterions in their environment, produce highly insoluble precipitates with certain added reagents.

## Gravimetry Based on Precipitation

Gravimetric analysis by precipitate formation involves a reaction of the following type:



where  $a$  is the number of moles of analyte A and  $r$  is the number of moles of reactant R. The product  $\text{RrAa}$  should be only sparingly soluble in the mother

liquor. This precipitate is filtered off and either dried and weighed or converted into a compound of definite stable composition and then weighed. The reagent may be an organic or inorganic species. Most metals can be precipitated with organic reagents as water-insoluble complexes and can be weighed after filtration and drying. Various techniques are used to ensure that precipitation is complete and to avoid unwanted interferences.

The procedure used for gravimetric analysis based on precipitation thus involves a number of steps:

- preparation of sample solution,
- precipitation,
- digestion,
- filtration and washing,
- drying, ignition, and weighing, and
- computation of results.

### Preparation of Sample Solution

In gravimetric analysis, the solution-preparation step has its own special significance. The analyte must be separated from interfering species, or the interferences must be masked. For example, if iron(III) is to be estimated as its hydrated oxide in the presence of chromium(III), then the mixture is initially treated with perchloric acid so as to oxidize chromium(III) to chromate (chromium(VI),  $\text{CrO}_4^{2-}$ ), followed by addition of ammonia to precipitate the hydrated iron oxide. Sometimes it is necessary to remove interferences, for example, when calcium is to be estimated as calcium sulfate in the presence of barium. The barium is removed as its chromate and the calcium is precipitated quantitatively as its sulfate.

### Precipitation

Precipitation is a critical step in gravimetric analysis because the manner in which precipitation is carried out will decide the particle size and purity of the precipitate. For gravimetric analysis, it is desirable that the precipitate has the following properties:

1. The solubility of the precipitate should be low so that losses from this source are negligible.
2. The precipitate should be able to be filtered and washed easily.
3. The precipitate should be stable and of definite composition.
4. A small amount of analyte should yield a relatively large amount of precipitate, since this

increases the sensitivity of the method. The use of high relative – molecular – mass precipitants (usually organic molecules) is advantageous.

The solubility product,  $K_s$ , of an ionic compound is defined as the product of the concentration of the ions that exist in equilibrium with the solid compound in a saturated solution. Under nonideal conditions activities are used in the place of concentrations. Thus, for an ionic compound  $A_m^{n+}B_n^{m-}$  the solubility product can be expressed in terms of the activities  $a$  of the ions as follows:

$$K_s = (a_{A^{n+}})^m \times (a_{B^{m-}})^n \quad [1]$$

However, when no solid is present, as is the case at the start of a precipitation reaction, the ionic product on the right-hand side of eqn [1] can greatly exceed  $K_s$ . The solution is therefore supersaturated, but it is only when the concentration of solute exceeds the critical supersaturation ratio value that precipitation begins. The critical supersaturation ratio,  $S_{\text{crit}}$ , is defined as the ratio of the smallest concentration needed to bring about the energetically most difficult stage of precipitation, nucleation, to the equilibrium solubility. Nucleation is the formation of a nucleus, the smallest particle of a precipitate that is capable of spontaneous growth.

**Kinetics of nucleation** Nucleation cannot strictly be described as a chemical reaction of a definite order, so that in a nucleation the rate of a reaction is not proportional to a power of concentration *per se*. However, in a restricted concentration interval it is a fairly good approximation to calculate the number of nuclei formed in terms of the supersaturation ratio  $S$ :

$$\ln N = \log K_N - B \left[ \frac{\sigma_{\text{CL}}^3 \nu^2}{k^3 T^3 \ln^2 S} \right] \quad [2]$$

where  $K_N$  is the kinetic constant,  $B$  is a factor depending upon the geometry of the nucleus and the type of crystal growth,  $\sigma_{\text{CL}}$  is the interfacial energy per unit area,  $\nu$  is the molecular volume,  $k$  is the Boltzmann constant,  $T$  is the absolute temperature, and  $S$  is the supersaturation ratio, the ratio of actual concentration to the solubility.

Two types of nucleation have been recognized, heterogeneous and homogenous. Heterogeneous nucleation is a low-energy process that takes place at lower supersaturation and involves impurity particles such as dust, glass fragments. Homogeneous nucleation only involves ions of the precipitate and occurs at higher supersaturation. For example, for barium sulfate  $S_{\text{crit}}$  is 20–30 for heterogeneous nucleation and  $\sim 1000$  for homogeneous nucleation.

**Table 1** Variation of morphological structure of barium sulfate precipitate as a function of supersaturation

Relative supersaturation	Property of precipitate
$7-3 \times 10^5$	Gelatinous
$3-0.75 \times 10^5$	Amorphous and partly colloidal
$5 \times 10^3$	Tiny needles and star-shaped crystals formed immediately
100	Precipitation after 2 h, microcrystalline form
20	Crystalline only after considerable time (1 month)

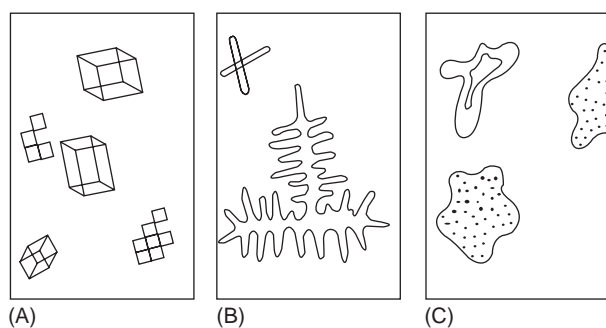
At lower supersaturations, where heterogeneous nucleation takes place, the concentration of the heteronuclei formed is dependent on the concentration of the impurity particles and is not greatly influenced by the ionic concentration, so that particle size increases as the concentration of precipitate ions increases. However, when the supersaturation ratio exceeds that required for homogeneous nucleation to take place, the concentration of nuclei increases rapidly with increasing ionic concentration. At very high supersaturation ratios the increase is so rapid that the ultimate particle size decreases with increasing ionic concentration. **Table 1** shows the variation of the form of barium sulfate precipitate as a function of supersaturation of the solution.

For gravimetric analysis it is desirable that the precipitates obtained can readily be filtered, that is, the particle size must not be too small. Suitable particle sizes are obtained by restricting nucleation to the heterogeneous type. This can often be achieved by using a dilute solution of precipitant and adding it dropwise with stirring. For some precipitates, however, neither of these precautions is successful because homogeneous nucleation occurs even at very low concentrations. This is the case with hydrated iron(III) oxide, which forms a gelatinous precipitate in which colloidal particles are loosely bound by adsorbed ions and much water is held in the structure. Such a precipitate is difficult to filter because it clogs the pores of the filter paper. In such instances, restriction to heterogeneous nucleation can only be achieved by the process known as precipitation from homogeneous solution (see below).

**Particle growth** The rate at which particles grow from nuclei to the final particle size increases with increasing concentration. The slower the growth, the more regular the crystal form and, as described below, the less contaminated is the precipitate.

**Figure 1** shows characteristic crystals of three different forms of barium sulfate. Regular, well-developed crystals are shown in **Figure 1A**, these have

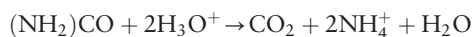




**Figure 1** Characteristic crystals of three different forms of barium sulfate: (A) regular, well-developed crystals formed in a strongly acidic medium at low relative supersaturation; (B) dendritic crystals formed at pH 2–3 and higher relative supersaturation; and (C) gelatinous precipitate obtained from a neutral medium.

been precipitated from a strongly acidic medium with a small relative supersaturation. If the precipitation is carried out at pH 2–3 at a higher relative supersaturation, the rate of accumulation is more rapid than the subsequent redistribution of material in the solid and the structure is dendritic (see **Figure 1B**). The precipitate obtained from neutral solution is barely crystalline because the rate of nucleation and accumulation is so fast (see **Figure 1C**).

**Precipitation from homogeneous solution** Precipitation from homogeneous solution (PFHS), or kinetically controlled precipitation, is a means of avoiding the buildup of high supersaturation by generating the precipitant *in situ* by a chemical reaction. For example, urea may be used as a homogenous precipitant for iron(III) or aluminum ions as hydrated oxides. Urea itself is such a weak base that it does not precipitate these species. It hydrolyzes only extremely slowly at room temperature, but does so quite rapidly at 80–90°C. If present in the acidic solution, this results in an increase in pH:



A typical procedure involves adding a suitable amount of urea to an acidified solution of, e.g., aluminum ions. A weak acid, such as formic or acetic acid, is also added to provide buffering action so that the pH does not rise too rapidly as the urea hydrolyzes. The solution is then heated until the pH rises to the desired value, which is  $\sim 4$  for quantitative precipitation of hydrated aluminum oxide. Homogeneously precipitated hydrated aluminum oxide is superior to that precipitated by addition of ammonia in that it is purer, denser, more compact, and more easily filtered. Some reagents used in PFHS are listed in **Table 2**.

**Table 2** Reagents used for PFHS

Precipitant	Reagent	Elements precipitated
$\text{OH}^-$	Urea	Al, Fe, Sn, Ga, Th, Bi
$\text{PO}_4^{3-}$	Trimethyl phosphate	Zr, Hf
$\text{C}_2\text{O}_4^{2-}$	Ethyl oxalate	Mg, Zn, Ca
$\text{SO}_4^{2-}$	Dimethyl sulfate	Ba, Sr, Ca, Pb
$\text{CO}_3^{2-}$	Trichloroacetic acid	La, Ba, Ra
$\text{H}_2\text{S}$	Thioacetamide	Sb, Mo, Cu, Cd
Dimethylglyoxime	Biacetyl + hydroxylamine	Ni, Pd

**Contamination of precipitates** Precipitates usually contain impurities. Such contamination mostly occurs during precipitation and the phenomenon is known as co-precipitation. Some contamination may occur after precipitation, i.e., on the surface of the final particles.

Co-precipitation causes contamination throughout the body of the precipitate if the particles or impurity ions or molecules are present during precipitation and are incorporated into the crystal structure of the precipitate particles throughout their growth. If precipitation is slow, the amount of co-precipitation is governed by the relative solubilities of the compound concerned, but with an increase in speed of precipitation the relative amount of co-precipitation also increases. For this reason, PFHS is advantageous for producing purer precipitates. Digestion of the precipitate, i.e., heating it for a time in the precipitation medium, causes some expulsion of foreign species, if there is any temporary dissolution of the precipitate, but some contamination will remain. Some of the mother liquor may also be occluded during rapid precipitation.

Co-precipitation on the surface of the precipitate is particularly important for precipitates of large surface area, e.g., gelatinous precipitates. It is essentially an adsorption effect, because the amount of contamination is the same whether the contaminating ions are present during precipitation or added afterwards. Digestion of the precipitate does not decrease the amount of adsorption because the surface area of gelatinous precipitates is not greatly affected by such treatment. Moreover, such contamination is essential for the precipitate to retain its gelatinous structure and not disperse as a colloidal suspension. However, a purer precipitate can sometimes be obtained by appropriate treatment, e.g., when silver is precipitated as its chloride, washing the precipitate with dilute acid will replace all the adsorbed potassium ions with hydrogen ions, which then volatilize as hydrochloric acid when the precipitate is dried.

### Digestion

The process of digestion is usually carried out by allowing the precipitate to remain in contact with the mother liquor, generally at a high temperature, for some time before filtration. By doing so complete precipitation is ensured and the smaller particles, which have higher solubility, tend to dissolve and make the solution supersaturated with respect to large particles. To restore the solution equilibrium, the dissolved material deposits on the larger particles, a process known as Ostwald ripening. A similar process applies to irregularities in the crystal structure. Protrusions are more soluble than cavities, and so are dissolved and precipitated in cavities, making the precipitate more regular. This is known as internal ripening. The net result of digestion is to improve the filterability of the precipitate, and also its purity since impurities may be released from the crystal structure on temporary dissolution.

### Filtration and Washing

Depending upon the nature of the precipitate and its subsequent treatment, filtration is carried out in different ways. If the precipitate is to be ignited, then it is generally filtered through a special type of ashless filter paper having a specified pore size chosen to retain the precipitate completely. If the precipitate is to be weighed after drying without any other treatment, then it is filtered on a special type of sintered bed. The precipitate is retained on the bed, which is dried and weighed.

Most precipitates are contaminated with one or more soluble compounds. The washing process is used to remove these as completely as possible. The composition of the washing solution chosen will depend upon:

1. the solubility and chemical properties of the precipitate;
2. the tendency of the precipitate to form a colloidal suspension (peptization);
3. the nature of the impurity to be removed; and

4. the effect of residual traces of the washing solution on the precipitate in the subsequent stages of drying/ignition.

Water often cannot be used because of the possibility of peptization of the precipitate, and also because of small losses arising from slight solubility of the precipitate. A solution of a suitable electrolyte is therefore generally used. The electrolyte chosen should possess an ion in common with the precipitate to suppress solubility and it should be easily volatilized when the precipitate is heated prior to weighing.

### Drying, Ignition, and Weighing

After a precipitate has been filtered and washed, it must be dried or ignited (heated to a high temperature) to obtain a compound of constant and known composition that can be weighed. The term drying is used when the chosen temperature is below 250°C and the term ignition used when the temperature range is 250–1200°C. Drying implies removal of solvent, while ignition indicates decomposition of the precipitate to produce a stable weighing form.

**Organic precipitating agents** There are a number of functional groups present in organic reagents that precipitate with metal ions by one of two routes: (1) chelating agents are organic compounds that ‘wrap around’ a metal ion due to cationic side chains that form coordinate bonds with the ion, and (2) a straightforward ion–ion bond that produces a new species that excludes water of solvation and thus the product precipitates. Good examples of chelating agents include oxalic acid, 8-hydroxyquinoline, and dimethylglyoxime.

Some common organic precipitating agents are given in Table 3.

**Inorganic precipitating reagents** Most inorganic ions have yielded to gravimetric analytical techniques, but one finds many interfering ions. Table 4 illustrates both the abundance of reagents available

**Table 3** Organic precipitants used in gravimetric analysis

Compound	Ions precipitated
Dimethylglyoxime	$\text{Ni}^{2+}$ , $\text{Pd}^{2+}$ , $\text{Pt}^{2+}$
Cupferron	$\text{Fe}^{3+}$ , $\text{VO}_2^+$ , $\text{Ti}^{4+}$ , $\text{Zr}^{4+}$ , $\text{Ce}^{4+}$ , $\text{Ga}^{3+}$ , $\text{Sn}^{4+}$
8-Hydroxyquinoline	$\text{Fe}^{3+}$ , $\text{Al}^{3+}$ , $\text{Mg}^{2+}$ , $\text{Zn}^{2+}$ , $\text{Cu}^{2+}$ , $\text{Cd}^{2+}$ , $\text{Pb}^{2+}$ , $\text{Bi}^{3+}$ , $\text{Ga}^{3+}$ , $\text{Th}^{4+}$ , $\text{Zr}^{4+}$ , $\text{TiO}^{2+}$ , $\text{UO}_2^{2+}$
Salicylaldehyde	$\text{Bi}^{3+}$ , $\text{Ni}^{2+}$ , $\text{Pd}^{2+}$ , $\text{Zn}^{2+}$ , $\text{Cu}^{2+}$ , $\text{Pb}^{2+}$
1-Nitroso-2-naphthol	$\text{Fe}^{3+}$ , $\text{Co}^{2+}$ , $\text{Pd}^{2+}$ , $\text{Zr}^{4+}$
Nitron	$\text{NO}_3^-$ , $\text{ClO}_4^-$ , $\text{BF}_4^-$ , $\text{WO}_4^{2-}$
Sodium tetraphenylborate	$\text{NH}_4^+$ , organic ammonium, $\text{Ag}^+$ , $\text{Cs}^+$ , $\text{Rb}^+$ , $\text{K}^+$
Tetraphenylarsonium chloride	$\text{Cr}_2\text{O}_7^{2-}$ , $\text{MnO}_4^-$ , $\text{ReO}_4^-$ , $\text{MoO}_4^{2-}$ , $\text{WO}_4^{2-}$ , $\text{ClO}_4^-$

**Table 4** Inorganic precipitants used in gravimetric analysis

Analyte	Precipitate	Measured from	Interferences
K <sup>+</sup>	KB(C <sub>6</sub> H <sub>5</sub> ) <sub>4</sub>	KB(C <sub>6</sub> H <sub>5</sub> ) <sub>4</sub>	NH <sub>4</sub> <sup>+</sup> , Ag <sup>+</sup> , Hg <sup>2+</sup> , Ti <sup>+</sup> , Rb <sup>+</sup> , Cs <sup>+</sup>
Mg <sup>2+</sup>	Mg(NH <sub>4</sub> )PO <sub>4</sub> · 6H <sub>2</sub> O	Mg <sub>2</sub> P <sub>2</sub> O <sub>7</sub>	Many metals (none from Na <sup>+</sup> and K <sup>+</sup> )
Ca <sup>2+</sup>	CaC <sub>2</sub> O <sub>4</sub> · H <sub>2</sub> O	CaCO <sub>3</sub> or CaO	Many metals (none from Mg <sup>2+</sup> , Na <sup>+</sup> , and K <sup>+</sup> )
Ba <sup>2+</sup>	BaSO <sub>4</sub>	BaSO <sub>4</sub>	Na <sup>+</sup> , K <sup>+</sup> , Li <sup>+</sup> , Ca <sup>2+</sup> , Al <sup>3+</sup> , Cr <sup>3+</sup> , Fe <sup>3+</sup> , Sr <sup>2+</sup> , Pb <sup>2+</sup>
TiO <sup>2+</sup>	TiO(5,7-dibromo-8-hydroxyquinolate) <sub>2</sub>	TiO(5,7-dibromo-8-hydroxyquinolate) <sub>2</sub>	Fe <sup>3+</sup> , Zr <sup>4+</sup> , Cu <sup>2+</sup> , C <sub>2</sub> O <sub>4</sub> <sup>2-</sup> , citrate, HF
VO <sub>4</sub> <sup>3-</sup>	Hg <sub>3</sub> VO <sub>4</sub>	V <sub>2</sub> O <sub>5</sub>	Cl <sup>-</sup> , Br <sup>-</sup> , I <sup>-</sup> , SO <sub>4</sub> <sup>2-</sup> , CrO <sub>4</sub> <sup>2-</sup> , AsO <sub>4</sub> <sup>3-</sup> , PO <sub>4</sub> <sup>3-</sup>
Cr <sup>3+</sup>	PbCrO <sub>4</sub>	PbCrO <sub>4</sub>	NH <sub>4</sub> <sup>+</sup> , Ag <sup>+</sup>
Mn <sup>2+</sup>	Mn(NH <sub>4</sub> )PO <sub>4</sub> · H <sub>2</sub> O	Mn <sub>2</sub> P <sub>2</sub> O <sub>7</sub>	Interferences from numerous metals
Fe <sup>3+</sup>	Fe(HCO <sub>2</sub> ) <sub>3</sub>	Fe <sub>2</sub> O <sub>3</sub>	Interferences from numerous metals
Co <sup>2+</sup>	Co(1-nitroso-2-naphtholate) <sub>3</sub>	CoSO <sub>4</sub> (by reaction with H <sub>2</sub> SO <sub>4</sub> )	Fe <sup>3+</sup> , Zr <sup>4+</sup> , Pd <sup>2+</sup>
Ni <sup>2+</sup>	Ni(dimethylglyoximate) <sub>2</sub>	Ni(dimethylglyoximate) <sub>2</sub>	Pd <sup>2+</sup> , Pt <sup>2+</sup> , Bi <sup>3+</sup> , Au <sup>3+</sup>
Cu <sup>2+</sup>	CuSCN	CuSCN	NH <sub>4</sub> <sup>+</sup> , Pb <sup>2+</sup> , Hg <sup>2+</sup> , Ag <sup>+</sup>
Zn <sup>2+</sup>	Zn(NH <sub>4</sub> )PO <sub>4</sub> · H <sub>2</sub> O	Zn <sub>2</sub> P <sub>2</sub> O <sub>7</sub>	Interferences from numerous metals
Ce <sup>4+</sup>	Ce(IO <sub>3</sub> ) <sub>4</sub>	CeO <sub>2</sub>	Th <sup>4+</sup> , Ti <sup>4+</sup> , Zr <sup>4+</sup>
Al <sup>3+</sup>	Al(8-hydroxyquinolate) <sub>3</sub>	Al(8-hydroxyquinolate) <sub>3</sub>	Interferences from numerous metals
Sn <sup>4+</sup>	Sn(cupferron) <sub>4</sub>	SnO <sub>2</sub>	Cu <sup>2+</sup> , Pb <sup>2+</sup> , As(III)
Pb <sup>2+</sup>	PbSO <sub>4</sub>	PbSO <sub>4</sub>	Ca <sup>2+</sup> , Sr <sup>2+</sup> , Ba <sup>2+</sup> , Hg <sup>2+</sup> , Ag <sup>+</sup> , HCl, HNO <sub>3</sub>
NH <sub>4</sub> <sup>+</sup>	NH <sub>4</sub> B(C <sub>6</sub> H <sub>5</sub> ) <sub>4</sub>	NH <sub>4</sub> B(C <sub>6</sub> H <sub>5</sub> ) <sub>4</sub>	K <sup>+</sup> , Rb <sup>+</sup> , Cs <sup>+</sup>
Cl <sup>-</sup>	AgCl	AgCl	Br <sup>-</sup> , I <sup>-</sup> , SCN <sup>-</sup> , S <sup>2-</sup> , S <sub>2</sub> O <sub>3</sub> <sup>2-</sup> , CN <sup>-</sup>
Br <sup>-</sup>	AgBr	AgBr	Cl <sup>-</sup> , I <sup>-</sup> , SCN <sup>-</sup> , S <sup>2-</sup> , S <sub>2</sub> O <sub>3</sub> <sup>2-</sup> , CN <sup>-</sup>
I <sup>-</sup>	AgI	AgI	Br <sup>-</sup> , Cl <sup>-</sup> , SCN <sup>-</sup> , S <sup>2-</sup> , S <sub>2</sub> O <sub>3</sub> <sup>2-</sup> , CN <sup>-</sup>
SCN <sup>-</sup>	CuSCN	CuSCN	NH <sub>4</sub> <sup>+</sup> , Pb <sup>2+</sup> , Hg <sup>2+</sup> , Ag <sup>+</sup>
CN <sup>-</sup>	AgCN	AgCN	Cl <sup>-</sup> , Br <sup>-</sup> , I <sup>-</sup> , SCN <sup>-</sup> , S <sup>2-</sup> , S <sub>2</sub> O <sub>3</sub> <sup>2-</sup>
F <sup>-</sup>	(C <sub>6</sub> H <sub>5</sub> ) <sub>3</sub> SnF	(C <sub>6</sub> H <sub>5</sub> ) <sub>3</sub> SnF	Many interferences, except alkali metals and SiO <sub>4</sub> <sup>4-</sup> , CO <sub>3</sub> <sup>2-</sup>
ClO <sub>4</sub> <sup>-</sup>	KClO <sub>4</sub>	KClO <sub>4</sub>	
SO <sub>4</sub> <sup>2-</sup>	BaSO <sub>4</sub>	BaSO <sub>4</sub>	Na <sup>+</sup> , K <sup>+</sup> , Li <sup>+</sup> , Ca <sup>2+</sup> , Al <sup>3+</sup> , Cr <sup>3+</sup> , Fe <sup>3+</sup> , Sr <sup>2+</sup> , Pb <sup>2+</sup>
PO <sub>4</sub> <sup>3-</sup>	Mg(NH <sub>4</sub> )PO <sub>4</sub> · 6H <sub>2</sub> O	Mg <sub>2</sub> P <sub>2</sub> O <sub>7</sub>	Many interferences except Na <sup>+</sup> , K <sup>+</sup>
NO <sub>3</sub> <sup>-</sup>	Nitron nitrate	Nitron nitrate	ClO <sub>4</sub> <sup>-</sup> , I <sup>-</sup> , SCN <sup>-</sup> , CrO <sub>4</sub> <sup>2-</sup> , ClO <sub>3</sub> <sup>-</sup> , NO <sub>2</sub> <sup>-</sup> , Br <sup>-</sup> , C <sub>2</sub> O <sub>4</sub> <sup>2-</sup>
CO <sub>3</sub> <sup>2-</sup>	CO <sub>2</sub> (by addition of acid)	CO <sub>2</sub>	CO <sub>2</sub> is trapped as Na <sub>2</sub> CO <sub>3</sub> on Ascarite

for use as well as the problems that can be encountered from interfering ions.

### Gravimetric Determination by Weight Loss

Another type of gravimetric determination involves measurement of loss in weight on heating or ignition of a known amount of sample. For example, the sample (analyte) can be hydrogencarbonate ion as shown or a mixture of carbonate and hydrogencarbonate. The total amount of carbonate in whatever form is found by placing the analyte in a solution containing an excess of H<sub>2</sub>SO<sub>4</sub>. This solution is in a flask connected to incoming nitrogen gas gently bubbled through the solution and an exit tube first to a drying agent to absorb aerosolized water and water vapor, and then to a mixture of NaOH and drying agent to absorb the CO<sub>2</sub> and water subsequently produced by the absorption by NaOH:



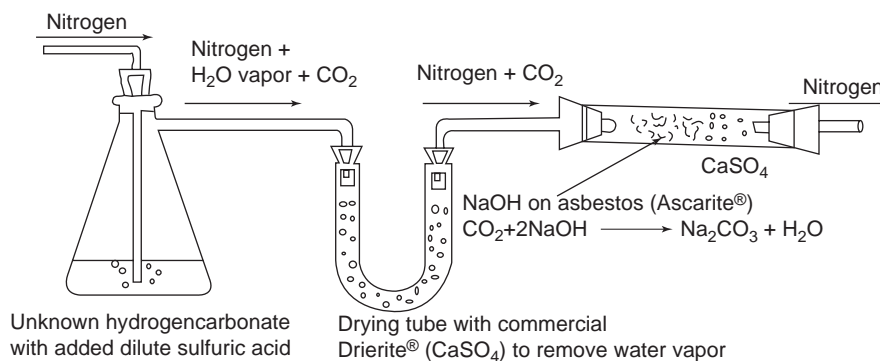
The apparatus is shown in Figure 2. The tube containing NaOH on asbestos and CaSO<sub>4</sub> to absorb

the final water product is pre- and post-weighed to give the total amount of carbonate in the sample. Note that the nitrogen gas acts only as a carrier and does not take part in any reaction.

### Gravimetric Determination by Weight Gain

A further type of gravimetric determination involves measurement of weight gain. For example, it would be possible to measure the amount of carbon dioxide and water evolved in the reaction described above for conversion of hydrogencarbonate to carbonate by passing the evolved gases through appropriate absorption tubes (a tube filled with anhydrous calcium chloride for absorption of water and a tube filled with potassium hydroxide for absorption of carbon dioxide). The gain in weight is measured and the amount of evolved gas is then correlated directly.

**Electrogravimetry** In electrogravimetry, the analyte is deposited (electroplated) on an electrode under controlled conditions; the increase in the weight is a



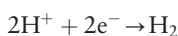
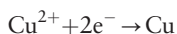
**Figure 2** Gravimetric determination by weight loss. (Modified from [www.csudh.edu](http://www.csudh.edu).)

direct measure of the amount of analyte in the sample. The procedure has the advantage that filtration is not required and, if conditions are carefully controlled, co-deposition of other metals can often be avoided. When a current is passed through a solution containing two or more electroactive species, the electrochemical process with the most positive reduction potential will occur first at the cathode. For example, if the solution contains copper, hydrogen, and cadmium ions, copper will deposit first. As the copper deposits the electrode potential becomes more negative until the potential is reached when hydrogen ions are reduced and hydrogen gas is formed at the cathode. The potential at the cathode will remain virtually constant as long as hydrogen gas is evolved, which is usually until all the water is electrolyzed, so that the potential of the cathode cannot become sufficiently negative to cause deposition of cadmium ions. Thus, metal ions with a less negative electrode potential can be separated from those having a more negative electrode potential using the electrodeposition technique.

The electrolytic separation of copper and nickel from cupronickel alloy can be achieved by controlling the pH and electrolytic conditions. Copper is determined in strongly acidic solution at a potential not exceeding 4 V (generally 2–3 V) as above this potential nickel may also plate out. The solution should contain a mixture of sulfuric and nitric acids for complete deposition. However, a high concentration of acid is not recommended as it may lead to incomplete deposition of copper or the deposit may not adhere satisfactorily to the cathode.

The reactions taking place at the respective electrodes are:

At the cathode,



At the anode,



Once all the copper is electroplated, concentrated ammonia is added to the remaining solution (iron if present is precipitated as its hydrated oxide and removed by filtration) and nickel is plated on another cathode at 4 A and 3–4 V. The test for completeness of deposition is carried out either by adding water and continuing the electrolysis or by testing with dimethylglyoxime.

**Thermogravimetric analysis** This technique monitors the change in mass of a substance as a function of temperature or time while the sample is subjected to a controlled temperature program.

## Examples of Applications

### Determination of Aluminum in Deodorants

**Sample preparation** For liquid samples, a fixed aliquot is diluted to a known volume and a fraction containing about ~200 mg of aluminum is pipetted out for analysis. For solids or creams a fixed amount is weighed and dissolved in dilute hydrochloric acid and then it is filtered through a fine medium to give a fixed volume.

**Determination** To a fixed volume of a diluted solution add one or two drops of phenolphthalein indicator and dropwise  $2 \text{ mol dm}^{-3}$  ammonia until turbidity occurs. To this solution add  $5 \text{ cm}^3$  of 10% acetic acid and  $10 \text{ cm}^3$  of 8-hydroxyquinoline solution to get a permanent precipitate of the metal complex. Digest and filter through tared Gooch crucible. Wash the precipitate with hot water and weigh after drying at  $120^\circ\text{C}$ . Calculate the amount of aluminum by using the standard molar relationship.

### Determination of Sodium from Dried Plant Powder

Moisten 1–10 g of the sample with sulfuric acid, dry it in an oven, and combust it in a furnace at 500–550°C to destroy the organic matter. Heat the residue on a steam bath with 2–5 cm<sup>3</sup> of concentrated hydrochloric acid, then add ~40 cm<sup>3</sup> of water, and heat till boiling. Add enough 5% CaCl<sub>2</sub> solution to ensure the precipitation of all the phosphates. Check the completeness of the precipitation by making the solution slightly alkaline using ammonia solution. Filter and evaporate to ~5 cm<sup>3</sup>, cool, and add 100 cm<sup>3</sup> of magnesium uranyl acetate solution, place the mixture in a water bath at 20°C, and either stir vigorously or let it stand for 24 h. Filter under suction into a previously weighed sintered bed crucible and wash the precipitate with ethanol saturated with magnesium uranyl acetate solution. Dry at 110°C for 30 min and weigh the precipitate as Na–Mg–uranyl acetate. The amount of sodium can be calculated as

$$\text{Weight of Na} = \text{weight of precipitate} \times 0.0153$$

### Determination of Benzaldehyde in Almond Oil

Weigh a sample containing ~10–50 mg of benzaldehyde in a distillation flask. Add enough ethanol and distil out nearly 100 cm<sup>3</sup> of distillate into a volumetric flask kept in an ice bath. To this distillate add 25 cm<sup>3</sup> of sulfuric acid, mix well, and immediately add 2,4-dinitrophenylhydrazine solution. Heat on water bath for 30 min at 75°C. Then, cool and filter through a previously weighed Gooch crucible. Wash with 25 cm<sup>3</sup> of water and dry at 100°C for 2 h. The amount of benzaldehyde can be calculated from:

$$\text{Weight of benzaldehyde} = \text{weight of precipitate} \times 0.3707$$

### Determination of Copper in Copper Ore

A common example of application of electrogravimetry is the determination of copper. Copper is reduced at potentials more positive than the reduction of hydrogen ion, so it can readily be deposited on a platinum cathode from acidic solutions (see above).

In this experiment, the copper ore sample is first oxidized to dissolve all the analyte. Then the analyte solution is electrolyzed. When the copper analyte is

electrochemically reduced, it deposits on the cathode. The cathode has a large area to allow the reduction of solution phase analyte to take place rapidly and quantitatively. The weight of analyte in the ore is determined by weighing the cathode prior to and after electrolysis.

See also: **Thermal Analysis:** Overview.

### Further Reading

- Belcher R, Nutton AJ, and Macdonald AMG (1970) *Quantitative Inorganic Analysis*, 3rd edn. London: Butterworths.
- Erdey L (1965) *Gravimetric Analysis*, Parts I–III: Oxford: Pergamon Press.
- Gordon L, Sautsky ML, and Willard HH (1959) *Precipitation from Homogenous Solution*. New York: Wiley.
- Kline DH and Gordon L (1958) Nucleation in analytical chemistry. *Talanta* 1: 342–344.
- Kolthoff IM (ed.) (1979) *Treatise on Analytical Chemistry*, Part I, vol. II. New York: Wiley Interscience.
- Laitinen HA and Harris WE (1975) *Chemical Analysis*, 2nd edn. New York: McGraw-Hill.
- Nielsen AE (1964) *Kinetics of Precipitation*. Oxford: Pergamon Press.
- Townshend A, et al. (1993) *Dictionary of Analytical Reagents*. London: Chapman & Hall.
- Walton AG (1967) *The Formation and Properties of Precipitates*. New York: Interscience.

### Glossary

- |  |  |
|--|--|
| Critical supersaturation ratio ( $S_{\text{crit}}$ ) | The critical supersaturation ratio $S_{\text{crit}}$ is defined as the ratio of the smallest concentration needed to bring about the energetically most difficult stage of precipitation, nucleation, to the equilibrium solubility. Nucleation is the formation of a nucleus, the smallest particle of a precipitate that is capable of spontaneous growth. |
| Co-precipitation                                     | Precipitates usually contain impurities. Such contamination mostly occurs during precipitation and the phenomenon is known as co-precipitation.  |
| Supersaturation ratio ( $S$ )                        | The supersaturation ratio is defined as the actual concentration to the equilibrium solubility.  |

## GUNSHOT RESIDUES

See **FORENSIC SCIENCES: Gunshot Residues**



# H

## HAEMOGLOBINS (HEMOGLOBINS)

**R A Sherwood**, King's College Hospital, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Iron

Iron has diverse functions in biology and medicine, with the result that both deficiency and excess have severe consequences in humans. Indeed, iron deficiency is the most prevalent nutritional problem in the world today. Most of the iron in the body (on average  $50 \text{ mg kg}^{-1}$  in the adult male and  $40 \text{ mg kg}^{-1}$  in the female) is contained in hemoglobin, which is essential for the transport of oxygen to the tissues. About 15% of the iron is found in myoglobin, which acts as a reservoir of oxygen in muscles. Iron not required immediately is stored in the form of ferritin or hemosiderin, both of which can bind large numbers of iron atoms. A very small proportion of body iron (0.2%) is bound to transferrin, which transports iron between sites of need. The measurement of iron, hemoglobin, and a combination of the proteins described above is necessary for the diagnosis and monitoring of iron deficiency or excess as in hemochromatosis.

Inherited variations in hemoglobin structure exist, some of which have a significantly reduced oxygen-carrying capacity (e.g., sickle-cell anemia and thalassemia). Identification of these hemoglobins is essential for correct treatment. Additionally, the hemoglobin molecule is glycosylated by circulating glucose and the measurement of the proportion of glycosylated hemoglobin in blood has proved invaluable in the monitoring of diabetes mellitus.

### Measurement of Plasma, Urine, and Tissue Iron

Plasma iron is derived mainly from the breakdown of hemoglobin in the reticuloendothelial system. Transferrin reversibly binds iron very firmly – there are no free iron ions present in plasma as they would be very toxic. Molecules of the transferrin–iron complex combine with receptor sites on erythroblasts and are then internalized, releasing iron into the cytosol. Consequently, the first stage in the measurement of iron in plasma (or serum) is dissociation of

the iron from transferrin, either with or without protein precipitation. A number of chemical agents have been successfully used for this process: trichloroacetic acid (TCA), hydrochloric acid, thorium(IV), acetate buffer and heat, and chloroform.

The iron released may be measured using a variety of methods based on two main techniques: atomic absorption spectrometry (AAS) (flame or furnace) and ultraviolet (UV)–visible spectrophotometry. Flame AAS provides the reference method for determination of plasma iron. Protein precipitation with TCA is followed by centrifugation and measurement of iron in the supernatant by the absorption at 248.3 nm in an air–acetylene flame. While atomic absorption methods are routinely used for urine iron measurements, the need to remove protein and any hemoglobin contamination restricts the use of this technique in routine clinical chemistry for plasma iron. Electrothermal atomization AAS methods are typically used for determination of iron in tissues although inductively coupled plasma (ICP) is becoming more widely available.

A number of spectrophotometric/colorimetric methods have been developed for use on the automated analyzers commonly available in clinical chemistry laboratories. These methods typically involve several steps: dissociation of iron from transferrin; reduction of iron(III) to iron(II); formation of colored complexes with iron-chelating chromogens.

Early methods utilized protein precipitation prior to complex formation, while today direct methods predominate. The method recommended by the International Committee for Standardization in Hematology requires protein precipitation with TCA, reduction with thioglycolic acid, followed by complex formation with ferrozine or ferene. Direct methods have significant benefits in analytical rapidity and automation. Dissociation of the transferrin–iron complex may be achieved by means of strong inorganic acids – typically hydrochloric acid – with solubilization of plasma proteins by the use of detergents. The dissociation is facilitated by the presence of a reducing agent, with the added benefit of reducing iron(III) to iron(II) at the same time. A variety of reducing agents have been used including hydroquinone, hydrazine,

**Table 1** Colorimetric reagents for the determination of serum iron, with their wavelength of maximum absorption and their molar absorptivity ( $\epsilon$ ) at that wavelength

Reagent	$\lambda_{\text{max}}$ (nm)	$\epsilon$ ( $\text{mol l}^{-1} \text{cm}^{-1}$ )
2,2'-Bipyridyl	522	8650
1,10-Phenanthroline	510	11 100
2,2',2''-Tripyridyl	552	11 000
Nitroso R salt	720	20 000
Bathophenanthroline	535	22 140
Tripyridyl-s-triazine (TPTZ)	595	22 600
Ferrozine	562	27 900
Ferene-S	593	35 500

sulfite, ascorbic acid, dithionite, hydroxylamine, thioglycolic acid, and mercaptoacetate.

Ascorbic acid is the most commonly used reducing agent. The free iron can then be reacted with a number of iron-chelating agents that form colored complexes. The more common chelating chromogens available are listed in **Table 1** together with the wavelength of maximum absorption and the molar absorptivity. It can be seen that since the discovery of 2,2'-bipyridyl in 1936 there has been a fourfold increase in the sensitivity of the available chromogens. TPTZ and Ferene-S produce blue complexes that have the advantage of being free from interference from bilirubins or the carotenoids, thus obviating the need for a sample blank measurement. These chromogens, however, are not entirely specific for iron as there is some binding and complex formation with copper, which can be removed by incorporating agents such as neocuprine, thiosemicarbazide, or thiourea that preferentially bind copper.

As all the methods mentioned above rely on the measurement of free iron after its release from proteins, iron released from hemoglobin would also be measured. For this reason any hemolysis will lead to falsely elevated plasma iron concentrations.

The reference range for plasma iron in the adult male is  $9\text{--}29 \mu\text{mol l}^{-1}$  with a slightly lower range for females ( $7\text{--}27 \mu\text{mol l}^{-1}$ ). A circadian rhythm has been described for plasma iron, with peak concentrations in the morning and a trough in late afternoon. Plasma iron is typically higher in the neonate, falling rapidly to adult levels, and typically results for the elderly are lower than the adult reference range. Various drugs, e.g., aspirin, allopurinol, and cholestyramine, can cause a decrease in plasma iron concentrations.

### Urine Iron Assays

Urinary iron excretion in normal individuals is relatively low ( $0.08 \text{ mg day}^{-1}$ ), producing concentrations too low to be measured by the methods available for

plasma iron. AAS is the technique of choice for urine iron measurements. Isolated measurements of urine iron concentration are of little clinical value. Clinically, the most valuable measurement is assay of the urine iron excretion after administration of desferrioxamine, an iron chelator. In normal subjects urine iron is elevated up to 10 times the baseline level, while in patients with genetic hemochromatosis elevations of up to 50 times baseline can be seen. Intermediate increases of 10–20-fold may be found in patients with hemosiderosis or following treatment of secondary iron overload from blood transfusions, e.g., in thalassemia patients.

### Tissue Iron

Tissue iron includes iron from heme-containing enzymes of cellular metabolism, accounting for  $\sim 300 \text{ mg}$ , and storage iron of  $800\text{--}1000 \text{ mg}$  in men, with lower amounts in women. The storage iron is predominantly in hepatocytes and in reticuloendothelial cells as ferritin and hemosiderin. Ferritin is located intracellularly and can sequester up to 4500 atoms of iron within its central core. Hemosiderin appears to be an iron-dense material consisting of multiple aggregates of ferritin molecules in which part of the protein shell has been degraded so that the iron cores coalesce. Ferritin is water soluble, whereas hemosiderin is not. Ferritin is widely distributed in liver, spleen, bone marrow, placenta, heart, pancreas, skeletal muscle, and intestinal mucosa. A variety of isoforms exist specific to particular locations, ranging from acidic to basic in nature.

Tissue iron can be visualized and localized on histological sections by the use of Perl's Prussian blue stain, which is based on the reaction of acidic hexacyanoferrate(II) principally with hemosiderin, but also with ferritin. A semiquantitative scoring system has been devised by estimating the iron load in parenchyma cells or Kupffer cells. Quantitative methods utilize both chemical and physical techniques. The tissue is first digested with TCA-HCl at  $65^\circ\text{C}$  for 20 h and the iron in the supernatant is reacted with bathophenanthroline sulfate and thioglycolic acid as in the plasma method. Although suitable for crude assessment of iron status, the precision of the method is less than satisfactory. The most frequently used alternative is AAS after nitric acid digestion. ICP atomic emission spectrometry has been used for assay of liver iron following nitric acid – perchloric acid digestion. ICP-mass spectrometry (ICP-MS) has been used as an online system for the detection of  $^{57}\text{Fe}$  originally incorporated into transferrin. The liver iron was fractionated by liquid chromatography (LC) and the eluate was passed to the ICP-MS system operating with an argon plasma.

Even more exotic techniques such as Mossbauer spectroscopy or electron diffraction measurements have been used in the research environment for detection, identification, and quantification of different forms of hemosiderin.

### Transferrin and Iron-Binding Capacity

Transferrin has two high-affinity binding sites for iron and exists in plasma as a mixture of apotransferrin and the mono- and diferric forms. Transferrin may be measured either directly by immunological techniques or indirectly by its functional capacity to bind iron. The most widely used technique for the measurement of total iron-binding capacity (TIBC) is based on the saturation of transferrin with a solution of iron(III) ions in excess. The excess of iron is removed with an iron adsorbent, usually magnesium carbonate or an anion-exchange resin. The iron concentration in the supernatant is then measured using standard methods for plasma iron. An alternative approach is to measure the unsaturated iron-binding capacity (UIBC). An excess, but known amount, of iron is added and the unbound fraction is measured using a reagent with specificity to unbound iron. The difference between the known and measured amounts of iron is that taken up by the protein. Measurement of UIBC is becoming the more commonly used technique as it does not require manual steps and is more suitable for automation on the large clinical chemistry analyzers.

The reference range for TIBC is typically  $50\text{--}70\ \mu\text{mol l}^{-1}$ . The combination of plasma iron and TIBC measurements can be used to derive the 'transferrin saturation' ( $(\text{plasma iron}/\text{TIBC}) \times 100$ ). In a normal individual transferrin is 30–40% saturated with iron, whereas in anemic subjects the degree of saturation can fall as low as 5%. In patients with hemochromatosis transferrin can be 100% saturated and 'free iron' can be detected in plasma. Pregnancy and oral contraception can increase the TIBC to  $90\ \mu\text{mol l}^{-1}$ .

Most of the common immunological techniques (nephelometry, turbidimetry, etc.) can be used to determine the concentration of transferrin directly. Generally, enhanced turbidimetric assays are now used for transferrin measurements on automated analyzers. Transferrin is decreased in disorders involving reduced or altered protein synthesis (e.g., liver disease) and increased in iron deficiency anemia. A rare condition of near complete absence of plasma transferrin (atransferrinemia) has been reported.

Transferrin has two bi-antennary carbohydrate side chains that can contain up to six sialic acid

residues and the sialylation process appears to be directly affected by excess alcohol intake. Examination of serum from chronic alcohol abusers has revealed that in such subjects there is a greater proportion of asialo-, mono-, and disialo transferrin compared to normal subjects where the tri- and tetrasialo forms predominate. The desialylated forms have been termed 'carbohydrate-deficient transferrin' (CDT) and form the basis of a test for chronic alcohol misuse. The effect of alcohol on transferrin appears to be a direct consequence of alcohol interfering with the sialylation process rather than as a result of any liver damage affecting synthesis as abstinence in alcoholics with liver disease results in the percentage of CDT in plasma returning to normal ( $<2.6\%$ ) within 10–14 days. CDT can be measured using separation techniques such as isoelectric focusing (IEF) or LC where the exact composition of the isoforms can be determined. For routine use in detecting or monitoring subjects misusing alcohol, separation of CDT is usually achieved using microcolumn ion-exchange chromatography with measurement of total transferrin and the carbohydrate-deficient forms allowing calculation of the percentage CDT. This has benefits over absolute measurements as it avoids false negative or positive results in cases of decreased or increased transferrin synthesis (e.g., protein malnutrition or advanced liver disease causing decreased synthesis and iron deficiency or pregnancy causing increased synthesis). A family of inherited diseases is now known in which proteins are deficient in carbohydrate side chains – the carbohydrate-deficient glycoprotein syndromes.

### Serum Transferrin Receptor

The mechanism for the transfer of iron from transferrin to the intracellular compartment of cells is now better understood. A specific transferrin receptor on cell membranes has been identified to which circulating diferric transferrin binds, after which endocytosis of the ligand–receptor complex occurs releasing iron into the cell. In iron deficiency there appears to be upregulation of the transferrin receptors to enable more efficient transfer of iron. The extracellular domain of the transferrin receptor (sTfr) appears to be released into the circulation by proteolytic action during cell turnover and may be detected in plasma (serum) by immunological techniques. Its concentration is increased in iron deficiency anemia ( $18\text{--}35\ \text{mg l}^{-1}$  compared with a normal range of  $3\text{--}9\ \text{mg l}^{-1}$ ). Measurement of sTfr is beneficial for distinguishing iron-deficiency anemia from the anemia of chronic disease.

### Ferritin

Ferritin consists of a large spherical shell of 24 single protein subunits surrounding an inner core of insoluble iron(III) oxide hydroxide with a portion of iron(III) phosphate. The iron-free protein, apo-ferritin, has a relative molecular mass of  $\sim 450\,000$  that can double when fully saturated with iron. Ferritin is measured immunologically with isotopic or nonisotopic (enzyme or chemiluminescence) immunoassays. In normal subjects there is a wide range of serum ferritin concentrations ( $15\text{--}300\,\mu\text{g l}^{-1}$ ), predominantly due to a highly negative skew to the ferritin results owing to the lower iron stores in women. Ferritin is decreased in iron-deficiency anemia but is an acute-phase protein and may be normal or increased if there is coexisting infection or tissue injury. Ferritin may also be produced by a number of malignant tumors. A ferritin greater than  $750\,\mu\text{g l}^{-1}$  in the absence of an acute phase is indicative of iron overload and if secondary causes can be excluded, genetic hemochromatosis should be considered. The disease causing mutation for hemochromatosis (C282Y mutation in the HFE gene) is now known and can be detected using PCR and restriction enzyme digest methods.

### RBC Zinc Protoporphyrin

In iron deficiency where iron is unavailable for incorporation into heme in the developing reticulocytes or where incorporation is inhibited as in lead toxicity, protoporphyrin IX accumulates in the red cells. This can be measured by extraction into ethyl acetate–acetic acid and then back-extracted into HCl. After neutralization, an ether extraction followed by differential extraction into HCl allows separation of protoporphyrin and coproprophyrin.

### Hemoglobin

Iron is incorporated into the heme produced by the porphyrin synthetic pathway by linkage to the four nitrogen atoms of protoporphyrin. In hemoglobin, four heme units are linked to four polypeptide chains, consisting of two pairs of like chains. There are four major types of hemoglobin (A, A<sub>2</sub>, F, and G) with chain structures:

- HbA 97% of total Hb in an adult =  $\alpha_2\beta_2$ ,
- HbA<sub>2</sub> 3% of total Hb in an adult =  $\alpha_2\delta_2$ ,
- HbF fetal Hb =  $\alpha_2\gamma_2$ , and
- HbG intrauterine Hb = G1  $\xi_2\varepsilon_2$  and G2  $\alpha_2\varepsilon_2$ .

Disorders of hemoglobin resulting from structural abnormalities of a globin chain are termed

hemoglobinopathies, with over 400 variants known, the majority of which are clinically benign. The separation of these variants will be discussed later.

The measurement of total hemoglobin concentration in blood is one of the most frequently requested tests in medicine. Automated methods predominate but the principle of the method is little changed from the original manual techniques. The Fe(II) of hemoglobin is oxidized to Fe(III) by the addition of hexacyanoferrate(III) and then converted into stable cyanmethemoglobin by the addition of potassium cyanide. The absorbance of cyanmethemoglobin is measured at 540 nm. The normal range for adult males is  $14\text{--}18\,\text{g dl}^{-1}$  and for females  $12\text{--}16\,\text{g dl}^{-1}$ .

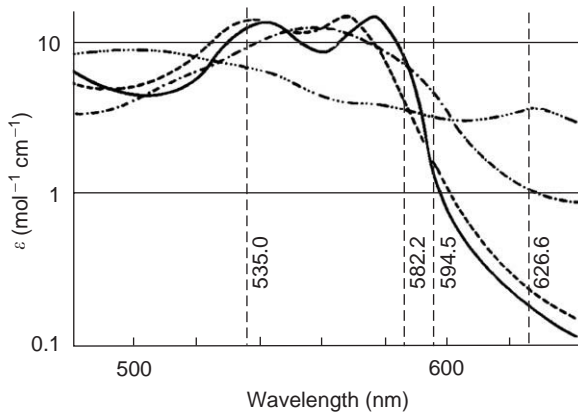
### Hemoglobin Derivatives

The function of hemoglobin is the transport of oxygen to the tissues from the lungs. When oxygen is associated with the molecule it is termed oxyhemoglobin (OHb), whilst in the absence of oxygen it is termed deoxyhemoglobin or reduced hemoglobin (RHb). In these forms iron is present as iron(II). When the iron is present as iron(III), a brown pigment is formed – methemoglobin. If carbon monoxide is present, i.e., because of smoke inhalation, a hemoglobin–carbon monoxide complex is formed that is referred to as carboxyhemoglobin (COHb). A final derivative of clinical interest is the degradation product sulfhemoglobin (SHb) that contains one more sulfur atom than normal hemoglobin. The iron is in the iron(II) state but the binding of oxygen is inhibited.

For the assessment of the oxygenation state of a patient, measurement of the proportion of oxyhemoglobin can be of value and a variety of commercial instruments exist for the measurement of the hemoglobin derivatives (oximeters, co-oximeters, and hemoglobinometers). The hemoglobin derivatives have differing absorption maxima (Figure 1) and these instruments are based on the differential spectroscopic measurement of the main derivatives: OHb, RHb, and COHb, plus in some cases SHb. The absorption maxima are known to vary with temperature and most commercial instruments are fitted with a temperature-controlled cuvette. COHb normally constitutes less than 5% of total hemoglobin, although this may be slightly higher in heavy smokers. COHb concentrations greater than 20% from smoke or carbon monoxide inhalation are usually fatal unless treated by the use of a hyperbaric oxygen chamber.

The normal concentration of methemoglobin is 1.5% of total hemoglobin, although in congenital methemoglobinemia due to methemoglobin reductase deficiency this may increase to 10–20%. Methemoglobin





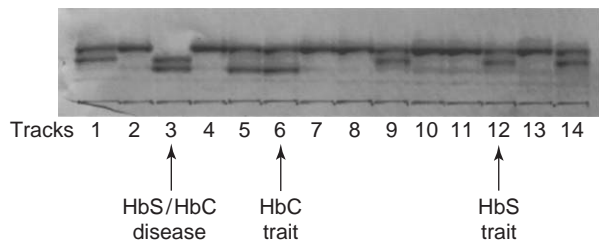
**Figure 1** Absorption spectra in the region 500–650 nm of the main hemoglobin derivatives: reduced hemoglobin (---), oxyhemoglobin (—), carboxyhemoglobin (-.-), and methemoglobin (.....). (Reproduced with permission from Instrumentation Laboratories 482 co-oximeter manual.)

can be measured by its absorbance in the region of 630–635 nm, which is eliminated by the addition of cyanide. Sulfhemoglobin is not a normal constituent in blood but may be present owing to reaction to certain drugs, e.g., phenacetin. Sulfhemoglobin has a broad absorbance band in the region of 600–620 nm that is not destroyed by cyanide.

### Separation of Variant Hemoglobins

The best-known hemoglobinopathies are sickle-cell anemia and thalassemia, which manifest clinically as a chronic anemia, hemolytic in the former and microcytic in the latter. Sickle cell disease is found predominantly in people of African descent and is characterized by sickle-shaped erythrocytes due to the homozygous inheritance of HbS in which valine is substituted for glutamic acid at position 6 in the  $\beta$ -chain. When deoxygenated HbS is much less soluble than deoxygenated HbA and forms rod-like polymers (tactoids), giving the characteristic sickle shape to the red blood cells. Sickling also occurs with HbC, in which lysine is substituted for the same glutamic acid. Thalassemia occurs in subjects of Mediterranean origin or Asian origin and is characterized by defective Hb synthesis. Both  $\alpha$ - and  $\beta$ -thalassemia exist where  $\alpha$ -chain or  $\beta$ -chain synthesis is decreased.

The identification of the variant hemoglobins is based on their separation by electrophoresis or chromatography utilizing the different overall charge on the molecule caused by the amino acid substitutions. Cellulose acetate electrophoresis is normally used for screening and is carried out at pH 8.6, where HbS has two more and HbC four more charges per



**Figure 2** Typical separation of hemoglobin variants using cellulose acetate electrophoresis. Track 1 is a standard mixture of HbA and HbS, tracks 2, 4, 7, 8, 10, 11, 13 are normals; tracks 5 and 6 are HbC traits; tracks 9, 12, 14 are HbS traits; and track 3 is HbS/HbC disease. (Reproduced with permission from the Department of Haematology, King's College Hospital, London.)

molecule than native HbA (Figure 2). Citrate gel electrophoresis at pH 6.2 relies on the interactions of the substituted amino acids with agaroseptin. Those variant hemoglobins with substitutions near the surface, e.g., HbS and HbC, are retained by these interactions while those with deeper substitutions, such as HbD or HbG, are not. Chromatographic techniques use ion-exchange columns (diethylaminoethyl-DEAE cellulose) to separate the variants on the basis of the aforementioned charge differences.

Preliminary screening with cellulose acetate electrophoresis will permit the rapid identification of many variant types of hemoglobin, but for some variants, such as HbD, coelution occurs with HbS. Globin chain electrophoresis can be used to separate these variants. The heme group is dissociated from the globin chain with dithiothreitol and urea followed by cellulose acetate electrophoresis using Tris-EDTA-borate buffer at pH 8.9 and pH 6.2. Excellent resolution of hemoglobin variants can be achieved by IEF. More recently, capillary electrophoresis has been shown to be an effective technique for the separation of hemoglobin variants. Similarly, there is considerable interest in the application of tandem MS to hemoglobin separations.

### Glycohemoglobin and Its Measurement

All the hemoglobin types previously discussed will react with hexose sugars, the product being referred to as glycohemoglobin (GHb). The predominant adult hemoglobin HbA consists of two  $\alpha$ - and two  $\beta$ -chains in a tetramer. Condensation with hexoses gives rise to five GHb derivatives, designated collectively HbA<sub>1</sub>, HbA<sub>1a1</sub>, HbA<sub>1a2</sub>, HbA<sub>1b</sub>, HbA<sub>1c</sub>, and HbA<sub>1d</sub>. *In vivo* condensation between glucose and hemoglobin occurs throughout the lifespan of erythrocytes in proportion to the circulating blood glucose concentration. The GHb concentration in a



particular blood sample therefore reflects the mean blood glucose concentration of the preceding weeks to months. Measurement of GHb has been shown to be a valuable means of monitoring long-term glycemic control in diabetics and is widely used in clinical practice. Since the report of the Diabetes Control and Complications Trial (DCCT) clearly demonstrated a reduction in long-term complications (i.e., nephropathy, retinopathy, and neuropathy) if HbA<sub>1c</sub> is kept below 7.5%, measurements of GHb have been steadily standardized to HbA<sub>1c</sub> DCCT aligned methods.

### Methods for Measurement of GHb

A wide variety of techniques, particularly separation techniques, have been used to measure GHb (Table 2).

**Ion-exchange chromatography** Cation-exchange chromatography, generally involving the use of Biorex 70, has enjoyed widespread popularity for GHb measurements. 'Minicolumns' separate glycated HbA (all fractions) from nonglycated HbA with the hemoglobin concentration in the eluate being measured at either 414 or 552 nm. Although commonly used, these methods suffer the drawback of being sensitive to pH, ionic strength of the buffers used, and the temperature of the column. The results are expressed as HbA<sub>1</sub> with a target range of 6–8% for good glycemic control. Many of these methods are now being standardized to DCCT aligned HbA<sub>1c</sub> values.

**Liquid chromatography** The LC methods are also based on the use of cation-exchange columns. Separation of HbA<sub>1c</sub> from HbA<sub>1a</sub> and HbA<sub>1b</sub> can be achieved using LC and dedicated automated systems are commercially available. A significant disadvantage occurs in populations with a high percentage of abnormal hemoglobins, as to achieve an acceptable throughput a separation time of 5–8 min is used, which is too short to resolve the abnormal hemoglobins, leading to potentially misleading results. Typically, results are expressed as HbA<sub>1c</sub> with a range of 4–6% for nondiabetic subjects.

**Affinity chromatography** Unlike the techniques based on ion exchange, which rely on the small charge differences between the glycated hemoglobins, affinity chromatography utilizes the complex formation that occurs between boronate and the *cis*-diol groups of sugars, particularly fructose. The boronic acid is attached to a support matrix such as agarose and will bind all the species of GHbs. These are then eluted with sorbitol and the hemoglobin quantified colorimetrically. The major advantage of this method is its insensitivity to abnormal hemoglobins and automated versions using LC are available commercially with short cycle times (2–3 min). This method can be used in the clinic setting using capillary blood samples prior to the patient being seen by the diabetologist. Although the affinity chromatography based methods measure all GHb fractions, most are now calibrated to DCCT aligned HbA<sub>1c</sub> values.

**Agar gel electrophoresis** Electrophoresis on agar slabs separates glycated HbA (all forms) from nonglycated HbA as in ion-exchange chromatography but does not suffer from sensitivity to pH, ionic strength, or temperature. The method will not, however, resolve HbF from the glycated fraction, leading to gross overestimation in individuals with persistent HbF. A disadvantage is the analysis time, which even in automated versions, is usually too long for use in a clinic setting.

**Isoelectric focusing** IEF has been used to achieve separation of HbA<sub>0</sub>, HbA<sub>1c</sub>, HbS, and HbC. HbF is also separated but acetylated HbF comigrates with HbA<sub>1c</sub>. Although capable of excellent separation, IEF is too cumbersome for routine use in clinical chemistry.

**Immunoassays** Developments in the production of monoclonal antibodies have led to the production of monoclonal antibodies specific to HbA<sub>1c</sub>. These antibodies have been used with a variety of immunoassay platforms to produce automated methods. Care must be taken to determine if the epitope of the antibody is

**Table 2** Interferences with GHb measurements in the most commonly used techniques

	<i>Ion exchange</i>	<i>Boronate affinity</i>	<i>Agar gel</i>	<i>Immunoassay</i>
Negatively charged hemoglobins, e.g., HbF	Positive	None	Positive	Negative
Positively charged hemoglobins, e.g., HbC, HbS	Negative or none	None	Negative or none	Negative
Lipemia	Positive	None	None	None
Renal failure (carbamylated Hb)	Positive	None	None	None
Aspirin (hemoglobin–acetylsalicylic acid adduct)	Positive	None	Positive	None

within a region of hemoglobin that can be altered in the hemoglobinopathies.

**Reference method** A reference method has been established that involves cleavage of hemoglobin into peptides by the endoproteinase Glu-C, followed by separation and quantitation of the hexapeptides by LC–electrospray ionization mass spectrometry or LC–capillary electrophoresis.

*See also:* **Blood and Plasma. Clinical Analysis:** Overview. **Immunoassays, Applications:** Clinical.

## Further Reading

- Bry L, Chen PC, and Sacks DB (2001) Effects of hemoglobin variants and chemically modified derivatives on assays for glycohemoglobin. *Clinical Chemistry* 47: 153–163.
- Brugnara C (2003) Iron deficiency and erythropoiesis: New diagnostic approaches. *Clinical Chemistry* 49: 1573–1578.
- Cook JD, Flowers CH, and Skikne BS (2003) The quantitative assessment of body iron. *Blood* 101: 3359–3363.
- International Committee for Standardization in Haematology (1990) Recommended method for measurement of iron in serum. *British Journal of Haematology* 75: 615–616.
- Jeppsson JO, Kobold U, Barr J, *et al.* (2002) Approved IFCC reference method for the measurement of HbA1c in human blood. *Clinical Chemistry and Laboratory Medicine* 40: 78–89.
- Park CH, Valore EV, Waring AJ, and Ganz T (2001) Hepcidin, a urinary antimicrobial peptide synthesized in the liver. *Journal of Biological Chemistry* 276: 7806–7810.
- Scouller K, Conigrave KM, Macaskill P, Irwig L, and Whitfield JB (2000) Should we use carbohydrate-deficient transferrin instead of  $\gamma$ -glutamyltransferase for detecting problem drinkers? A systematic review and metaanalysis. *Clinical Chemistry* 46: 1894–1902.
- Sherwood RA, Pippard MJ, and Peters TJ (1998) Iron homeostasis and the assessment of iron status. *Annals of Clinical Biochemistry* 35: 693–708.
- Thomas C and Thomas L (2002) Biochemical markers and hematologic indices in the diagnosis of functional iron deficiency. *Clinical Chemistry* 48: 1066–1076.
- Worwood M (2002) Serum transferrin receptor assays and their application. *Annals of Clinical Biochemistry* 39: 221–230.

# HAIR

*See* **FORENSIC SCIENCES: Hair**

# HEADSPACE ANALYSIS

Contents

**Static**

**Purge and Trap**

## Static

**J D Green**, Beverley, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Headspace analysis is a technique for sampling and examining the volatiles associated with a solid or liquid sample. The actual headspace itself is the

volume of vapor or gas above the sample. For most headspace analysis purposes the sample and its associated headspace are held within an enclosed container (**Figure 1**). Often, it is the volatiles that are of primary importance when, for example, the chemical composition of the mixture giving rise to the aroma of the sample is to be determined. In other circumstances, examination of the volatiles rather than the whole sample greatly simplifies the analysis and still gives information relating to the whole sample. In effect, the analytical complications that are often apparent as a result of a complex matrix are considerably decreased.

within a region of hemoglobin that can be altered in the hemoglobinopathies.

**Reference method** A reference method has been established that involves cleavage of hemoglobin into peptides by the endoproteinase Glu-C, followed by separation and quantitation of the hexapeptides by LC–electrospray ionization mass spectrometry or LC–capillary electrophoresis.

*See also:* **Blood and Plasma. Clinical Analysis:** Overview. **Immunoassays, Applications:** Clinical.

## Further Reading

- Bry L, Chen PC, and Sacks DB (2001) Effects of hemoglobin variants and chemically modified derivatives on assays for glycohemoglobin. *Clinical Chemistry* 47: 153–163.
- Brugnara C (2003) Iron deficiency and erythropoiesis: New diagnostic approaches. *Clinical Chemistry* 49: 1573–1578.
- Cook JD, Flowers CH, and Skikne BS (2003) The quantitative assessment of body iron. *Blood* 101: 3359–3363.
- International Committee for Standardization in Haematology (1990) Recommended method for measurement of iron in serum. *British Journal of Haematology* 75: 615–616.
- Jeppsson JO, Kobold U, Barr J, *et al.* (2002) Approved IFCC reference method for the measurement of HbA1c in human blood. *Clinical Chemistry and Laboratory Medicine* 40: 78–89.
- Park CH, Valore EV, Waring AJ, and Ganz T (2001) Hepcidin, a urinary antimicrobial peptide synthesized in the liver. *Journal of Biological Chemistry* 276: 7806–7810.
- Scouller K, Conigrave KM, Macaskill P, Irwig L, and Whitfield JB (2000) Should we use carbohydrate-deficient transferrin instead of  $\gamma$ -glutamyltransferase for detecting problem drinkers? A systematic review and metaanalysis. *Clinical Chemistry* 46: 1894–1902.
- Sherwood RA, Pippard MJ, and Peters TJ (1998) Iron homeostasis and the assessment of iron status. *Annals of Clinical Biochemistry* 35: 693–708.
- Thomas C and Thomas L (2002) Biochemical markers and hematologic indices in the diagnosis of functional iron deficiency. *Clinical Chemistry* 48: 1066–1076.
- Worwood M (2002) Serum transferrin receptor assays and their application. *Annals of Clinical Biochemistry* 39: 221–230.

# HAIR

*See* **FORENSIC SCIENCES: Hair**

# HEADSPACE ANALYSIS

Contents

**Static**

**Purge and Trap**

## Static

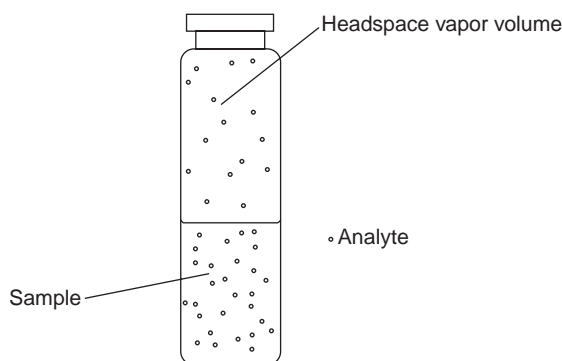
**J D Green**, Beverley, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Headspace analysis is a technique for sampling and examining the volatiles associated with a solid or liquid sample. The actual headspace itself is the

volume of vapor or gas above the sample. For most headspace analysis purposes the sample and its associated headspace are held within an enclosed container (**Figure 1**). Often, it is the volatiles that are of primary importance when, for example, the chemical composition of the mixture giving rise to the aroma of the sample is to be determined. In other circumstances, examination of the volatiles rather than the whole sample greatly simplifies the analysis and still gives information relating to the whole sample. In effect, the analytical complications that are often apparent as a result of a complex matrix are considerably decreased.



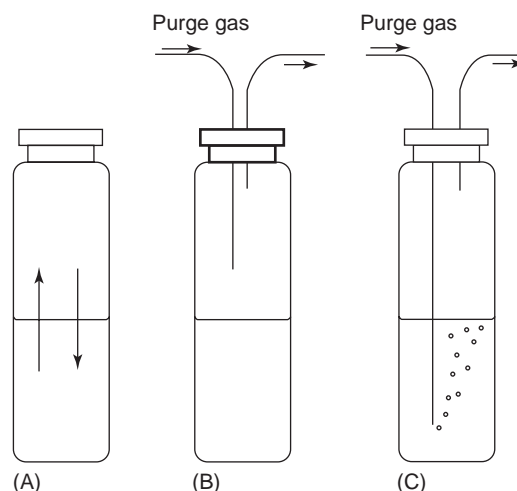
**Figure 1** The fundamental concept of headspace analysis.

## Principles

Headspace volatiles can be examined by a number of analytical methods and, although the technique is most often associated with gas chromatography (GC), it is perfectly feasible for headspace samples to be analyzed by other analytical methods. Such methods as mass spectrometry (MS), molecular spectrometric techniques, and voltammetry have been used, although infrequently and, if the component of interest exists in a volatile form, there is no reason why atomic spectrometry should not be used.

Sampling of headspace vapors at its simplest can be effected by using a gas-tight syringe. In this case a sample of the headspace contained in a closed volume above the sample is drawn into the syringe and subsequently introduced into the analytical instrumentation. Alternatively, an automated system can be used in which the sample can be conditioned in various ways according to the capabilities of the instrumentation prior to introduction into the analytical system using autosampler syringe technology or a sample transfer line. The results obtained depend upon the analytical technique chosen to supplement the headspace sampling procedure.

Static and dynamic headspace sampling techniques may be distinguished; this is shown diagrammatically in **Figure 2** together with the closely related technique of 'purge and trap'. Static methods involve the removal of a small fraction of the headspace in equilibrium with the liquid. Dynamic methods involve the continuous removal of the headspace sample from above the sample as an enrichment step followed by a concentration of the vapors either cryogenically or using a porous adsorbent. The concentrated sample from the dynamic sampling may then be revaporized (from a cryogenic trap) or desorbed (from a solid adsorbent) and analyzed by an appropriate technique. Less commonly, a headspace sampling technique known as 'continuous flow thin



**Figure 2** Concepts of 'static' headspace (A), 'dynamic' headspace (B), and purge and trap (C) sampling systems.

layer headspace' can be used to concentrate headspace volatiles in a thin layer of a flowing liquid. The concentrate can then be analyzed by the appropriate technique, including those such as voltammetry that are not usually associated with headspace methods.

Qualitative headspace analysis is used in many applications to determine differences between selected samples and to assess the importance of specific compounds in defining the aroma character of certain substances. This approach is often termed 'fingerprinting' as in many instances the composition remains chemically unknown. For additional qualitative data, headspace may be used in conjunction with knowledge of GC retention data or with GC-MS techniques for a more definitive chemical identification.

Quantitative relationships exist between the concentrations of volatiles in equilibrium with liquid samples and these form the basis of quantitative theory for headspace techniques. The partition coefficient,  $K$ , is the parameter of fundamental importance and it is defined as the ratio of concentration of analyte  $a$  in the sample phase  $s$ ,  $C_a^s$  and that in the vapor phase  $v$ ,  $C_a^v$ , at equilibrium (eqn [1]):

$$K = \frac{C_a^s}{C_a^v} \quad [1]$$

The sensitivity of headspace analysis is influenced principally by two sample-related parameters: the partition coefficient of the analyte between the phases and the phase ratio of sample to vapor. Consideration of the mass balance of an analyte in a sample under headspace conditions of equilibrium allows a simple expression to be developed relating sensitivity

to these two parameters (eqn [2]):

$$M_a^o = M_a^s + M_a^v \quad [2]$$

where  $M_a^o$  is the mass of analyte in the original sample,  $M_a^s$  is the mass of analyte in the sample phase under headspace conditions, and  $M_a^v$  is the mass of analyte in the vapor phase under the headspace conditions. This develops to

$$C_a^o V^s = C_a^s V^s + C_a^v V^v \quad [3]$$

where  $C_a^o$  is the concentration of analyte in the original sample,  $V^s$  is the volume of the sample phase, and  $V^v$  is the volume of the vapor phase under headspace conditions.

Incorporating  $K = C_a^s/C_a^v$  from eqn [1] gives

$$C_a^o V^s = K C_a^v V^s + C_a^v V^v \quad [4]$$

and therefore

$$\begin{aligned} C_a^o &= K C_a^v + C_a^v \left( \frac{V^v}{V^s} \right) \\ &= C_a^v \left[ K + \frac{V^v}{V^s} \right] \end{aligned} \quad [5]$$

Now,  $V^v/V^s$  is the phase ratio  $\beta$ , and therefore

$$C_a^o = C_a^v [K + \beta] \quad [6]$$

It is  $C_a^v$  that is proportional to the analytical response, therefore

$$C_a^v = \frac{C_a^o}{K + \beta} \quad [7]$$

Equation [7] shows that for a given concentration of analyte in the original sample,  $C_a^o$ , the analytical response, which is proportional to  $C_a^v$ , will depend upon  $K$  and  $\beta$ . With a large value of  $K$  (e.g., 1000 for ethanol in water), the phase ratio has no influence. Conversely, for a small value of  $K$  (e.g.,  $K \approx 0$  for halogenated species in water), the phase ratio has a significant bearing upon the sensitivity. Chemical methods for modifying the partition characteristics of the system are appropriate in some cases, perhaps the most well known being the addition of salts to the liquid phase to induce a higher concentration of organics in the vapor phase. In addition, the sensitivity of the detector used and the capacity of the analytical system for the sample will influence the sensitivity achievable as it does in any analytical method.

Quantitative analysis using headspace techniques is dependent upon the partition of the analyte between the sample and vapor phases. The concentration of the analyte in the sample phase is usually

the value sought, although it is the vapor phase that is sampled and analyzed. Four different approaches to headspace analysis quantification are discussed below. These are total vaporization, matrix-matched standards of analyte, multiple headspace extraction, and standards addition.

### Total Vaporization

Total vaporization of the analyte into the vapor phase can be realized for a limited number of samples and this provides the simplest case for quantification. As an example, a polymeric matrix containing residual water could be heated to 150°C to effectively drive all the water into the vapor. Headspace analysis of the vapor phase in conjunction with external calibration will give the concentration of water in the vapor phase. From this and knowledge of the vapor phase volume, the total amount of water present in the sample can be calculated. This is expressed in the trivial relationship:

$$M_a^s = M_a^v \quad [8]$$

which becomes

$$M_a^s = C_a^v V^v \quad [9]$$

when the concentration term is introduced, where the terms have the same meanings as given earlier. With a known mass or volume of sample the concentration of analyte in the sample may be calculated.

### Matrix-Matched Analyte Standards

In most practical cases the analyte will be distributed between the sample and vapor phases. Calibration involves preparing standards in a matrix identical to that of the sample with known concentrations of the analyte. The headspaces of these standards are then analyzed under identical conditions to the sample and thereby a calibration graph can be constructed. This procedure is practical for liquids and gives eqn [10]:

$$C_a^s = C_a^{\text{std}} \frac{A_a}{A_{\text{std}}} \quad [10]$$

In eqn [10],  $C_a^{\text{std}}$  is the concentration of analyte in the prepared standard,  $A_a$  is the analytical response to the analyte in the sample, and  $A_{\text{std}}$  is the analytical response to the analyte in the prepared standard. For solid samples, preparation of valid calibration standards is more difficult because the analyte in the real sample may be bound in a different form from that in the standard. These so-called matrix effects are difficult to overcome; however, the method described next provides a viable approach.



### Multiple Headspace Extraction

Multiple headspace extraction is a simple concept. As the headspace is extracted several times from one sample, the analyte concentration is progressively decreased. If the procedure were carried out a sufficient number of times then no analyte would be detected in later extracts. Summing all the values obtained would allow the analyte concentration in the original sample to be calculated. While this procedure would be tedious and time consuming, simpler practical approaches using this concept are available. Multiple headspace extraction is a variant of methods known as stepwise gas extraction and discontinuous gas extraction. Such procedures can be applied to a variety of samples and can be used to quantify analytes in samples for which it is difficult or impossible to set up calibration standards as a result of the matrix complications or difficulties in preparing standards. These methods are therefore particularly valuable for quantification of volatiles in solid samples or for quantifying highly volatile analytes for which the preparation of accurate standards is experimentally difficult.

A series of headspace analyses is carried out on the same sample, which is conditioned to equilibrium between each analysis step. From the progressive, measured decrease in concentrations determined, the analyte concentration in the original sample can be calculated. The theory is outlined below.

With a continuous gas extraction, the decrease in concentration of analyte with time will be proportional to the remaining concentration of analyte (eqn [11]):

$$\frac{-dC_a}{dt} = kC \quad [11]$$

This integrates to give an expression for the concentration of analyte present at time  $t$ :

$$C = C_a^0 e^{-kt} \quad [12]$$

in which  $C_a^0$  is the initial concentration and  $k$  is an exponent that includes the partition coefficient  $K$ . Where concentrations are proportional to the analytical responses of the techniques used, the expression may be transposed to

$$A_n = A_1 e^{(1-n)k'} \quad [13]$$

where concentrations become analytical responses ( $A_n$  is the analytical response of the  $n$ th extraction, etc.) and  $t$  is replaced by the number of extraction steps ( $n$ ) in a multiple sequence having equal time intervals. The new exponent  $k'$  includes contributions from both  $K$  and some instrumental parameters.

From eqn [13], using the mathematics of geometric progression, an expression may be developed to give the

sum of the analytical responses for a series of analyses:

$$\sum A_n = \frac{A_1}{1 - e^{-k'}} \quad [14]$$

in which  $k' = \ln(A_1/A_2)$ . For practical purposes this can be simplified to a procedure involving two extractions, and the sum of analytical responses becomes

$$\sum A_n = \frac{A_1^2}{A_1 - A_2} \quad [15]$$

Further details of the mathematical derivation of these relationships can be found in the texts cited in the 'Further Reading'.

### Standard Addition

Standard addition is also a method commonly used to quantify headspace analysis and can be used in cases where a standard solution of the analyte can be added to the sample in such a way that it exists in an identical form in relation to the matrix as the analyte itself. Analysis of the original sample and the sample with added standard analyte can then provide a value for the concentration of analyte in the original sample.

The analytical response  $A_a$  to the analyte from the sample and the concentration of analyte in the sample,  $C_a^s$ , are related by

$$A_a = K' C_a^s \quad [16]$$

where  $K'$  is a function of the distribution coefficient and the detector response. In a similar way, when a standard is added to the sample to be analyzed the analytical response,  $A_a^*$ , may be expressed as

$$A_a^* = K' \left[ C_a^s \left( \frac{V_a}{V_a + v} \right) + C_{std} \right] \quad [17]$$

in which  $C_a^s$  is the concentration of the analyte in the sample that is to be determined,  $C_{std}$  is the concentration of the standard added,  $V_a$  is the volume of the solution (liquid phase) in which the analyte is dispersed, and  $v$  is the volume of standard solution added (of known concentration).

This expression may be developed to give

$$\begin{aligned} A_a^* &= \frac{A_a}{C_a^s} \left[ C_a^s \left( \frac{V_a}{V_a + v} \right) + C_{std} \right] \\ &= A_a \left[ \left( \frac{V_a}{V_a + v} \right) + \frac{C_{std}}{C_a^s} \right] \end{aligned} \quad [18]$$

$$\frac{A_a C_{std}}{C_a^s} = A_a^* - A_a \left( \frac{V_a}{V_a + v} \right) \quad [19]$$

$$C_a^s = \frac{C_{\text{std}} A_a}{(A_a^* - A_a [V_a / (V_a + v)])}$$

$$= \frac{C_{\text{std}} A_a (V_a + v)}{(A_a^* (V_a + v) - A_a V_a)} \quad [20]$$

This simplifies to

$$C_a^s = \frac{A_a C_{\text{std}}}{A_a^* - A_a} \quad [21]$$

when the volume added as a result of the standard addition is small.

The choice of quantification in headspace analysis therefore depends on the sample and the analyte. Procedures include total vaporization, the use of matrix-matched standards, multiple extractions, and standard additions. A method using an internal standard may be used provided knowledge of the equilibrium constants for the standard and the analyte can be determined.

## Apparatus

The equipment used for headspace analysis in its simplest form may comprise a vial of suitable volume with a septum turn closure through which the sample may be drawn using a gas-tight syringe. With the use of a means of temperature control, this method of sampling headspace volatiles can be effective for simple applications where a qualitative or semiquantitative analytical screening procedure is required.

For more demanding applications, a range of apparatus has been designed for the sampling and

collection of headspace vapors. Commercially available systems exist from several of the major instrument manufacturers (Table 1). Modern equipment is designed for flexibility and can be used for static headspace methods, dynamic sampling, sometimes includes the option of multiple headspace extraction together and a capability to do solid-phase microextraction (SPME) injections. Some equipment is aimed at the electronic nose market for aroma profiling and the identification of odor components. Most manufacturers provide the means for automatic conditioning of samples and provide features to allow vial shaking, constant equilibration times for the sample vials, and method development options whereby the parameters can be progressively changed over a number of samples. Methods for introducing samples into the analytical instrumentation differ from automated syringe technology through sample valve and loop techniques to direct coupling using pressure-balanced sampling.

A wider range of equipment that has particular features for special applications is described in the literature. For example, special apparatus has been designed and reported for sampling the headspace of canned foods and whole cheeses, and from single cigarettes.

## Applications

Headspace techniques have been used for examining a wide variety of samples including environmental samples of soil, water, and air, biological fluids,

**Table 1** Commercially available headspace sampling and injection equipment

<i>Manufacturer (web address)</i>	<i>Models available</i>	<i>Selected features</i>
Agilent (www.agilent.com)	7694 Headspace sampler	Flexible sample conditioning for 44 samples with inert interface options to Agilent and other gas chromatographs
Alpha MOS (www.presearch.co.uk)	HS 100 – Odorscanner	Sampler aimed at providing headspace samples to 'electronic nose' equipment
Gerstel (www.gerstel.com)	Headspace Chemsensor 4440	Provides an odor sample to a MS (quadrupole) or GC/MS analyzer for identification or comparative profiling
Perkin Elmer (http://uk.instruments.perkinelmer.com)	Turbomatrix Headspace Samplers HS-110/-16/-40	Pressure balanced sampling with flexible sample numbers and conditioning
Shimadzu (www.shimadzu.co.uk)	Liquid, Headspace & SPME GC Injection system	Combined injection capability including headspace
Teledyne Tekmar (www.teledynetekmar.com)	7000HT Headspace Autosampler (also 7050 with 50 sample vial positions)	Sample equilibration temperatures of up to 300°C, inert sample path and method optimization mode
Thermo Finnigan (www.thermo.com)	HS 2000 Headspace Sampler	Robotic system to control sample equilibration, allow method optimization
Varian (www.varianinc.com)	Combi PAL autosampler injection system	Combines headspace with liquid and SPME injections. Flexible sample conditioning and headspace injection modes

foodstuffs and related materials, forensic samples, pharmaceuticals, natural products, and polymers. Such samples are examined to ensure environmental protection, to diagnose medical conditions, to establish compliance with regulations in the food, pharmaceutical, and packaging industries, and for the direct determination of aroma volatiles with their associations to taste and smell. All these applications of the headspace analysis benefit either from the simplification of the analysis by removing some of the matrix effects or as a result of the technique being directly related to the examination of aroma volatiles. As mentioned above, headspace analysis is closely associated with techniques of purge and trap, which also direct attention to the volatiles 'extracted' from a sample by the purge gas.

Environmental samples including soils, water, and air have been examined using various headspace techniques. Soils, estuarine sediments, and industrial wastes have been examined for volatile organics including chlorinated hydrocarbons, benzene, and other aromatics. Groundwater samples have been examined for a wide range of volatile substances including chlorinated hydrocarbons, carbon disulfide, components of gasoline spillages, and methylmercury. The latter analysis was carried out using GC with atomic emission detection. Drinking water has been analyzed for volatile constituents including halocarbons, aromatics, and chloral hydrate using headspace sampling techniques. Dissolved gases and light hydrocarbons have been determined in seawater. In some cases headspace methods give improved limits of detection compared to direct injection procedures, and the problems of column overloading and detector memory effects that can occur with aqueous samples are eliminated. Headspace sampling is also used for determining solvents in waste gases and for a variety of substances in workplace and domestic environments including vulcanization volatiles in the rubber industry, and alkylated benzenes and formaldehyde from particle board products.

Trace components in alcoholic drinks can be determined following headspace concentration procedures, although great care is required to establish that the major components present in the sample are not influencing the partition of the analyte of interest between the liquid and headspace phases.

Volatile components of packaging materials in foods can be detected and quantified using headspace techniques. As an example, vinyl chloride as a component of poly(vinyl chloride) (PVC) has been determined in olive oil after storage in PVC containers.

Analysis of gases associated with fresh food packaging is carried out using a headspace technique, with

instruments being available to determine oxygen and carbon dioxide concentrations using a zirconia oxygen probe and an infrared detector for carbon dioxide.

The requirements for the determination of the impurities present in gases and especially in air are frequently so demanding that direct determination (usually by GC and therefore direct injection) is not feasible. The technique of reverse headspace analysis has been applied to such samples. The gas, the minor impurities of which are to be determined, is equilibrated with a liquid, usually in a special flow cell. The liquid may then be directly injected into a gas chromatograph or, in the case of a nonvolatile liquid being used to trap the impurities, thermal elution of the volatiles can be effective prior to analysis. Such methods have been used to determine sulfur-containing compounds in natural gas and air.

Monomeric residues in polymeric formulations or products can be determined using headspace techniques. Thus, monomeric impurities in aqueous suspensions of polymers for subsequent processing can be determined as a means of reducing possible off-odors in the processed materials. In commercial polymers, monomeric species can be identified and quantified using headspace sampling techniques with the appropriate analytical method. Such monomers may be undesirable with respect to subsequent usage, especially as packaging for food that may then become tainted by the volatile species present.

Clinical samples of urine, blood, expired air, and tissue have been examined using headspace sampling approaches. Thus, chlorinated organic compounds, methanol, acetone, methyl ethyl ketone, and phenols have been determined in urine. Volatile substances in urine have also been used as a guide to acute poisoning, and the determination of stimulants in urine has been proposed as screening test for 'field' use. The determination of the concentration of blood alcohol is the most well-known application of headspace techniques to biological samples. Blood has also been examined for cyanide, methyl sulfide, and formaldehyde levels, the last as a measure of methanol intoxication. The headspace approach for blood samples overcomes the difficulties associated with the alternative direct injection of two-phase samples.

Milk, being a two-phase substance, also creates some analytical difficulties for direct injection methods and therefore headspace techniques have been used to overcome these and to determine the odor components that have a bearing upon milk quality in processing and storage.

Volatiles associated with solids such as polymers and tablet formulations of pharmaceuticals can be

**Table 2** Standard methods using headspace techniques

<i>Application</i>	<i>Standard(s)<sup>a</sup></i>
<i>Determination of residual volatiles</i>	
Residual acrylonitrile content of styrene-acrylonitrile copolymers and nitrile rubber	ASTM D4322
Residual vinyl chloride monomer (ng g <sup>-1</sup> range) in vinyl chloride homopolymers and copolymers	ASTM D4443
Residual vinyl chloride monomer in poly(vinyl chloride) resins	ASTM D3749
Acetaldehyde content of freshly blown poly(ethylene terephthalate) bottles	ASTM D4509
<i>Fire debris samples</i>	
Separation/concentration of flammable and combustible residues from fire debris samples by dynamic headspace analysis	ASTM E1413
Separation/concentration of flammable and combustible residues from fire debris samples by passive headspace analysis	ASTM E1412
<i>Recommended methods and apparatus</i>	
Practice of sampling of headspace vapors from fire debris samples	ASTM E1388
Small-scale environmental chamber determinations of organic emissions from indoor materials/products	ASTM D5116
Volatile matter content in polymers by headspace chromatography	ASTM D4526
<i>Water samples</i>	
Examination of water, wastewater and sludge for individual constituents	DIN 38413
<i>Foodstuffs</i>	
Tests for vegetables, fruits, and their products including headspace techniques	SASO 440

<sup>a</sup> Sources: ASTM, American Society for Testing and Materials; DIN, Deutsche Institute für Normung; SASO, Saudi Arabian Standards Organization.

determined using headspace techniques. Those samples that dissolve completely can be dealt with as solutions, when the headspace approach allows the volatiles to be analyzed separately from the nonvolatile solids. For insoluble solids the volatiles can be identified but quantification, which has been discussed above, is less straightforward.

Numerous applications of headspace analysis to the determination of the volatile aroma compounds of natural products, their essential oils and extracts have been reported. Thus, fruits, tobacco, hops, mushrooms, and various plants have been examined to determine the volatiles present. Such investigations provide information relating to the important chemistry of aroma, taste, and flavor. This has advanced the understanding of the development of odor in fresh, processed, and preserved food products as well as in flowers, from which fragrances are extracted and developed. Consequently, criteria of quality have been established and an understanding has developed of the changes in volatile composition occurring as a result of aging, storage, and processing procedures. As diverse examples, headspace techniques have been used to examine truffle aroma, volatiles in canned salmon, and from kiwi fruit flowers, and garlic volatiles. Synthesis of nature-identical flavors has been significantly advanced by the results of headspace analysis in this area of aroma and flavor research.

Household and personal hygiene products comprise another group of samples that can be screened and examined for volatile components. Thus, 1,4-dioxane in shampoos, ethylene oxide in cosmetics, and fragrances in soaps and cleaning fluids are examples of typical applications of the technique.

In a small number of selected examples it is possible to use headspace sampling techniques as part of an analytical procedure to determine involatile components in a sample. The headspace sampling usually follows a chemical treatment to produce a volatile compound from the analyte of interest. It is thus possible to determine fluoride in fluorinated milk following initial treatment to produce volatile trifluorosilane.

Headspace techniques are used in a number of standard methods for the examination of a wide variety of samples. Details are given in **Table 2**.

Headspace analysis has largely been neglected in the area of process analysis, although, with the inherent matrix complications that can slow down and complicate such analyses, significant benefits may be realized from such an approach.

Headspace techniques can be used to determine a number of thermodynamically important properties relating to vapor-liquid equilibria. The technique has been used to develop distillation procedures specifically where the selection of solvents in extractive distillation is required. Vapor pressures of pure

substances and activity coefficients may also be determined using headspace techniques.

*See also:* **Blood and Plasma. Cosmetics and Toiletries. Essential Oils. Food and Nutritional Analysis:** Pesticide Residues; Alcoholic Beverages; Dairy Products. **Polymers:** Natural Rubber; Synthetic; Polyurethanes. **Process Analysis:** Chromatography. **Quality Assurance:** Primary Standards; Spectroscopic Standards;

**Internal Standards. Sensory Evaluation. Sulfur. Water Analysis:** Organic Compounds.

## Further Reading

Kolb B and Ettre L (1997) *Static Headspace–Gas Chromatography: Theory and Practice*. New York: Wiley.  
Manufacturers' Websites (see Table 1) and additionally [www.atas.co.uk](http://www.atas.co.uk).

## Purge and Trap

**D Kou**, Hoffmann-La Roche Inc., Nutley, NJ, USA  
**S Mitra**, New Jersey Institute of Technology, Newark, NJ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Purge and trap refers to the technique in which volatile organic compounds (VOCs) are purged out of the sample matrix by an inert gas and carried on to a sorbent trap, where they are concentrated and later introduced into an instrument (gas chromatography (GC) or GC/mass spectrometry (MS)) for analysis. After its introduction in the 1970s, purge and trap quickly gained widespread acceptance, and was adopted by regulatory agencies as a standard method. Today, it is still the workhorse for the analysis of low-concentration VOCs in solids and liquids. The instrumentation and procedures are slightly different for liquid and solid samples. This article discusses its basic theory, instrumentation, operational procedures, quantification methods, and selected applications.

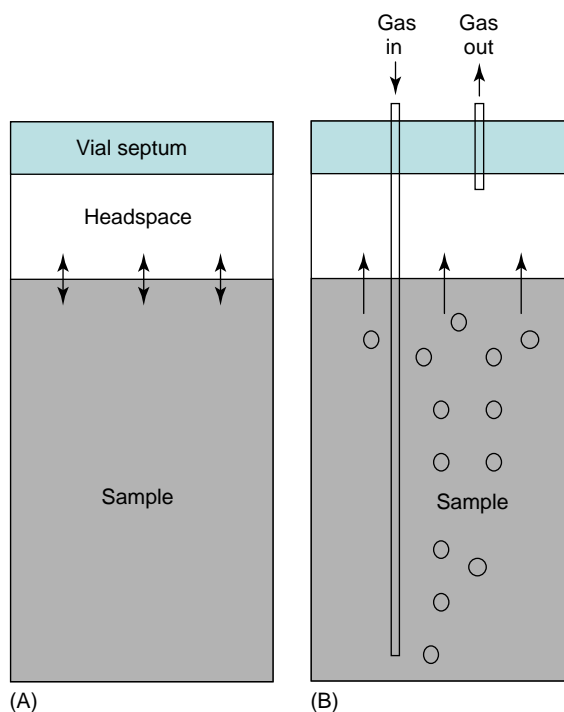
### Theoretical Considerations

Purge and trap is also known as dynamic headspace extraction, which is closely related to static headspace extraction. In both techniques, organic compounds migrate from the sample matrix into its headspace. This is illustrated in Figure 1. In static headspace, equilibrium is reached between the sample and its headspace. In purge and trap, an inert gas is used to continuously remove the analytes out of the sample, thus creating a higher concentration gradient. This enhances the mass transfer of the analytes. Consequently, purge and trap is more sensitive than static headspace analysis. It is suitable for trace analysis with detection limits at parts per billion to

parts per trillion levels. On the other hand, static headspace is often used as a complementarily screening method for high-concentration samples.

There are three main processes in purge and trap: extraction (purging), simultaneous adsorption (trapping), and subsequent desorption (heating). The overall purge-and-trap performance depends on the efficiency of each individual process.

Purging efficiency (the percentage of analytes removed from the sample) depends upon the properties of the analytes, purge temperature, flow rate, duration of the purging, and sample matrix. Purging is most efficient for VOCs that are insoluble or



**Figure 1** Comparison of purge and trap with static headspace extraction. (A) Static head space and (B) purge and trap.



substances and activity coefficients may also be determined using headspace techniques.

*See also:* **Blood and Plasma. Cosmetics and Toiletries. Essential Oils. Food and Nutritional Analysis:** Pesticide Residues; Alcoholic Beverages; Dairy Products. **Polymers:** Natural Rubber; Synthetic; Polyurethanes. **Process Analysis:** Chromatography. **Quality Assurance:** Primary Standards; Spectroscopic Standards;

**Internal Standards. Sensory Evaluation. Sulfur. Water Analysis:** Organic Compounds.

## Further Reading

Kolb B and Ettre L (1997) *Static Headspace–Gas Chromatography: Theory and Practice*. New York: Wiley.  
Manufacturers' Websites (see Table 1) and additionally [www.atas.co.uk](http://www.atas.co.uk).

## Purge and Trap

**D Kou**, Hoffmann-La Roche Inc., Nutley, NJ, USA  
**S Mitra**, New Jersey Institute of Technology, Newark, NJ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Purge and trap refers to the technique in which volatile organic compounds (VOCs) are purged out of the sample matrix by an inert gas and carried on to a sorbent trap, where they are concentrated and later introduced into an instrument (gas chromatography (GC) or GC/mass spectrometry (MS)) for analysis. After its introduction in the 1970s, purge and trap quickly gained widespread acceptance, and was adopted by regulatory agencies as a standard method. Today, it is still the workhorse for the analysis of low-concentration VOCs in solids and liquids. The instrumentation and procedures are slightly different for liquid and solid samples. This article discusses its basic theory, instrumentation, operational procedures, quantification methods, and selected applications.

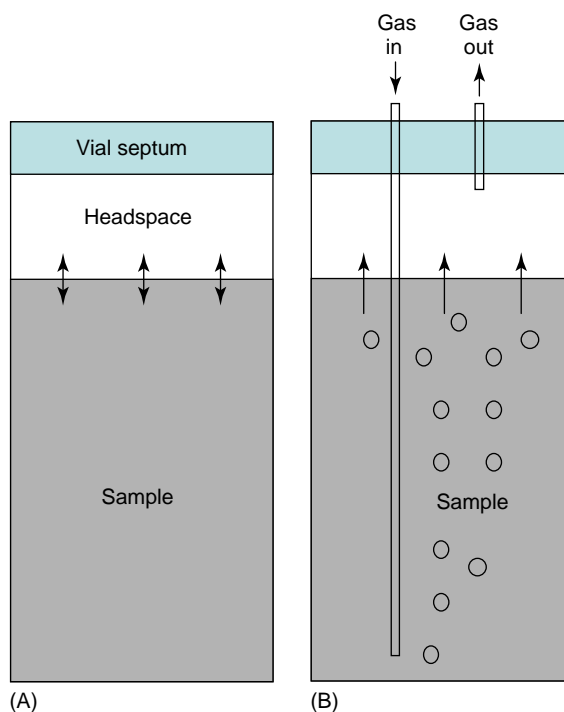
### Theoretical Considerations

Purge and trap is also known as dynamic headspace extraction, which is closely related to static headspace extraction. In both techniques, organic compounds migrate from the sample matrix into its headspace. This is illustrated in Figure 1. In static headspace, equilibrium is reached between the sample and its headspace. In purge and trap, an inert gas is used to continuously remove the analytes out of the sample, thus creating a higher concentration gradient. This enhances the mass transfer of the analytes. Consequently, purge and trap is more sensitive than static headspace analysis. It is suitable for trace analysis with detection limits at parts per billion to

parts per trillion levels. On the other hand, static headspace is often used as a complementarily screening method for high-concentration samples.

There are three main processes in purge and trap: extraction (purging), simultaneous adsorption (trapping), and subsequent desorption (heating). The overall purge-and-trap performance depends on the efficiency of each individual process.

Purging efficiency (the percentage of analytes removed from the sample) depends upon the properties of the analytes, purge temperature, flow rate, duration of the purging, and sample matrix. Purging is most efficient for VOCs that are insoluble or



**Figure 1** Comparison of purge and trap with static headspace extraction. (A) Static head space and (B) purge and trap.

slightly soluble in water. Purging at ambient temperature is adequate for these compounds. Water-soluble VOCs (e.g., ketones, alcohols, aldehydes, and some ethers) are more difficult to purge, and detection limits can be an order of magnitude higher if purged under the same conditions as those for the nonpolar VOCs. Purging efficiency can also be improved by using elevated temperatures. Acceptable recovery can be obtained for many water-soluble compounds when purged at 80°C. Purge time has traditionally been set as 11 min and is rarely used as a variable. For most compounds, the optimum flow rate is in the range of 20–40 ml min<sup>-1</sup>. High flow rates can improve the recovery of compounds that are hard to purge. Very low flow rates should be avoided for such compounds (e.g., bromoform). At high flow rates, however, gaseous compounds (boiling point lower than 35°C, e.g., chloromethane) can be lost from the trap due to breakthrough. Thus, excessively high flow rates should be avoided. Although matrix effects are generally not an issue for relatively clean samples such as drinking water and groundwater, they are not uncommon for complex samples such as soils, sludges, foods, and polymers. Appropriate quantification methods that can compensate for the matrix effects should be used accordingly.

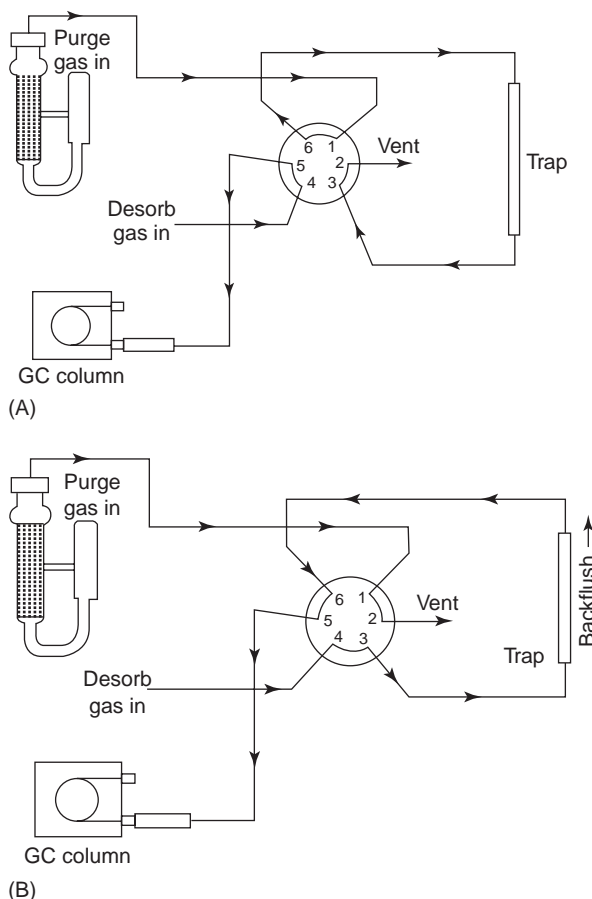
Trapping efficiency is dependent on the properties of the target compounds and the sorbent materials used in the trap. Generally, compounds with lower boiling points are more difficult to trap, and require the use of stronger sorbents. Trapping is also more efficient at lower temperatures. For most compounds, ambient temperature is adequate.

Desorption efficiency is affected by temperature, trap heating rate, and desorb-gas flow rate. Analytes are released faster at a higher temperature, which is limited by the thermal stability of the sorbent materials. Also, the faster the trap is heated, the shorter the desorption time. In addition, desorption is more efficient at higher flow rate. The actual flow rate depends upon the GC column used.

## Instrumentation

Figure 2 shows a schematic diagram of a typical purge-and-trap system. It consists of two major components: a purging device and a sorbent trap. The two parts are connected with each other and to the analytical instrument through transfer lines, with a six-port switching valve controlling the flow path.

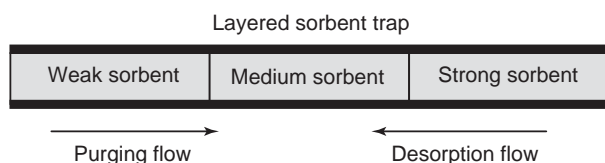
Three types of purging vessels are commonly used: frit spargers, fritless spargers, and needle spargers. The frit creates fine gas bubbles with large surface areas, which facilitate mass transfer of analytes into



**Figure 2** Schematic diagram of purge and trap: (A) purge mode – valve positioned to purge sample onto trap; (B) desorb mode – valve position rotated to desorb sample from trap and onto the GC column. (Courtesy of Restek Corp., used by permission.)

the gas. Frit spargers (typically 5 or 25 ml) are suitable for relatively clean water samples. The frit can get clogged if sample contains particles, and some sample tends to foam when purged through a frit. Fritless spargers are to be used for these samples. Needle spargers (typically 40 ml) are good for solid samples, which are usually mixed with a liquid before purging.

A trap is typically a 25 cm long tube with an internal diameter of 0.267 cm, and contains one or more layers of sorbents whose trapping abilities range from the weakest at the inlet to the strongest at the outlet. This is shown in Figure 3. During purging, the analyte-laden gas reaches the weak sorbent first, which traps the less volatile compounds. More volatile compounds break through the weak sorbent and subsequently are trapped by strong sorbent. During desorption, the trap is heated and backflushed with the GC carrier gas. In this way, the less volatile compounds never come in contact with



**Figure 3** Schematic diagram of a multilayer sorbent trap, used with backflushed desorption.

strong sorbents, and thus irreversible adsorption is avoided.

Conventional traps typically contain Tenax<sup>®</sup>, silica gel, and charcoal in series, each material filling one-third of the trap. Tenax is a 2,6-diphenylene oxide polymer. It is hydrophobic and has relatively weak trapping ability. It can only trap compounds with boiling points higher than 35°C, and more volatile species and polar compounds are retained poorly. It also has limited thermal stability and is subject to decomposition if heated over 200°C. Silica gel is a hydrophilic sorbent with trapping capacity stronger than Tenax. It is used for trapping compounds that are gases at room temperature (boiling point lower than 35°C). However, water is also retained by silica gel. Charcoal is a very strong sorbent used for trapping the highly volatile species (e.g., dichlorodifluoromethane, boiling point -29°C) that break through Tenax and silica gel. Charcoal is hydrophobic and therefore does not retain much water. The selection of the trap should be based on the target analytes. If only compounds with boiling temperature above 35°C are to be analyzed, silica gel and charcoal can be eliminated and the trap is filled with Tenax.

Alternative sorbent materials that are commonly used include granulated carbon black (GCB or Carbo-pack<sup>®</sup>), carbon molecular sieves (Carbosieve<sup>®</sup>), and Carboxen<sup>®</sup>. These carbon-based materials are hydrophobic. They also have higher thermal stability than Tenax, and can be used at ~245°C. GCB has about the same trapping capacities as Tenax. It is often used in series with Carbosieve or Carboxen, which are strong adsorbents serving as an alternative to silica gel and charcoal.

The transfer lines are usually made of inert materials such as deactivated fused silica, or silica-lined stainless steel tubing. These do not have active sites for adsorption. The lines are heated at ~120°C to avoid analyte condensation. The switching valve is also maintained above 100°C.

Newer instruments usually have a water management system to reduce the amount of water entering the GC. The gas exiting the purge vessel is saturated with water, which is then accumulated on the trap. If not removed, water can enter the GC when the trap is heated. Even at ambient purging temperature, as

much as 11 µl of water can be condensed on the trap and later transferred into the GC during desorption. The amount is significantly more with heated purging. Although water has little effect on analyte desorption from the trap, at a certain level it can cause the deterioration of GC resolution and interfere with the detector (especially photo-ionization detector (PID) and MS). It can also plug the column when a cryogenic interface is used. Therefore, water removal is critical in purge and trap.

There are two approaches for water removal: permeation and condensation. Permaselective membrane (e.g., Nafion<sup>®</sup> tubing) is effective in removing water while retaining the VOCs. However, some polar compounds such as alcohols also go through the membrane along with water vapors. Thus, this method is not suitable for polar analytes. A condensation device is more commonly used. It has less effect on the recovery of polar, water-soluble analytes, although there is always the possibility of losing such compounds. The device is a piece of tubing made of inert material (e.g., nickel). It is placed between the trap and the GC, and is maintained at ~30°C during trap heating, serving as a cold spot in the heated transfer lines. Water is condensed on it and removed from the GC carrier gas. After sample desorption, the device is heated and water vapor is vented. With a state-of-the-art water removal system, the amount of water entering the GC can be reduced to as low as 0.25 µl.

## Operational Procedures

Since purge and trap is hyphenated to GC or GC-MS, its operation must be synchronized with that of the analytical instrument, and the operational parameters of the two units must be compatible. The processing of each sample is a multistep cycle, which includes purge, dry purge, desorb preheat, desorb, and trap recondition/baking. The flow path during each step is controlled by the switching valve, which has two modes/positions: purge and desorb. **Figure 2** shows these two positions and the corresponding flow paths. During the purge-and-trap cycle, the valve is set in the purge position except during the desorb step.

### Purge

An aliquot of sample (5 or 25 ml) is introduced into the purging vessel either using an autosampler or manually using a syringe. The six-port valve is in the purge mode. The purge gas (nitrogen or helium) passes through the sample, purges the VOCs out of the matrix, and carries them to the trap. The VOCs are retained by the sorbents while the gas is vented.

Meanwhile, the desorb gas directly enters the GC. Purge time is typically 11 min at a flow rate of 20–40 ml min<sup>-1</sup>.

Solid samples (typically 5 g) are collected in a vial that contains a magnetic stir bar. Then the vial is sealed hermetically. At the time of analysis, a certain volume (typically 5 ml) of reagent water is added into the sealed vial, and the sample is purged with heating and agitation.

### Dry Purge

After purging, the purge gas bypasses the purging vessel and directly goes to the trap. This is called 'dry purge'. Its purpose is to reduce the water that has condensed on the trap during wet purging. Dry purge typically takes 1–4 min. During this step, the flow path of desorb gas is unchanged. It should be noted that traps containing silica gel are not compatible with dry purge.

### Desorb Preheat

After dry purge, the purge gas flow is stopped. The trap is heated to 5°C below the desorb temperature without gas flow. Preheat makes the subsequent desorption of analytes faster and helps prevent peak tailing.

### Desorb

Following preheat, the valve is switched to the desorb position and the trap is heated to the desorb temperature, which depends upon the type of sorbents. The trap is backflushed with the desorb (carrier) gas at a flow rate compatible with the GC column. The GC run and data acquisition are initiated simultaneously at the onset of desorption (except when a secondary trap is used, see below). Desorption time is typically 2–4 min.

Desorption is fast at a flow rate of ~40 ml min<sup>-1</sup> and this produces a sharp injection for the GC. Packed columns can directly accept such a high flow. Capillary columns offer superior separation capacity than packed columns, but they require lower flow rates. The optimum flow rate for wide-bore capillary columns (0.53 mm ID or larger) is 5–10 ml min<sup>-1</sup> at which desorption is slower. Long columns (60 m or longer) with a thick stationary phase (3 µm or more) can provide adequate retention for highly volatile compounds. For shorter columns, subambient cooling (e.g., 10°C) is often necessary at the beginning of the GC run to retain early eluting compounds. If GC–MS is used, an open split interface or a jet separator is needed to reduce the flow from the GC column into the MS detector.

Narrow-bore capillary columns (0.32 mm ID or smaller) are operated at a flow rate of 2–4 ml min<sup>-1</sup>.

This flow rate is much lower than the optimum flow rate for sample desorption. To solve this problem, two configurations are commonly used. One is to use a high desorb flow rate, which is then split into the GC. However, this also decreases the amount of analytes that enter the GC, leading to decreased sensitivity. The other solution is to desorb at a low flow rate (typically 4 ml min<sup>-1</sup>) for 5 min. Desorption is slow under such conditions and the desorbed analytes need to be refocused. A secondary cryogenic trap is often used for this purpose. It is a piece of uncoated silica tubing cooled to -150°C by liquid nitrogen. Once desorption from the primary trap is finished, the cryogenic trap is rapidly heated to 250°C, and the chromatographic run and data acquisition are started at the same time.

### Trap Recondition

After desorption, the switching valve is returned to the purge position. The trap is baked/reconditioned at (or 5–20°C above) the desorb temperature, and is flushed with purge gas for 8–10 min. The purpose of trap baking is to reduce sample carryover and possible trap contamination. After reconditioning, the trap heater is turned off and purge flow is stopped. Once the trap has cooled down, it is ready for the next sample.

## Methods of Quantification

The most commonly used quantification methods in purge and trap are external standard calibration and internal standard calibration. External standards refer to target compounds of known concentrations that are prepared and analyzed separately from the samples. A calibration curve or calibration factor (CF) is obtained by plotting the peak area count (or height) of the analyte versus its concentration in the standards. The analyte concentration in the sample is equal to its peak area divided by the CF. External standard calibration works well for relatively clean samples where matrix has no or little effects on analyte recovery.

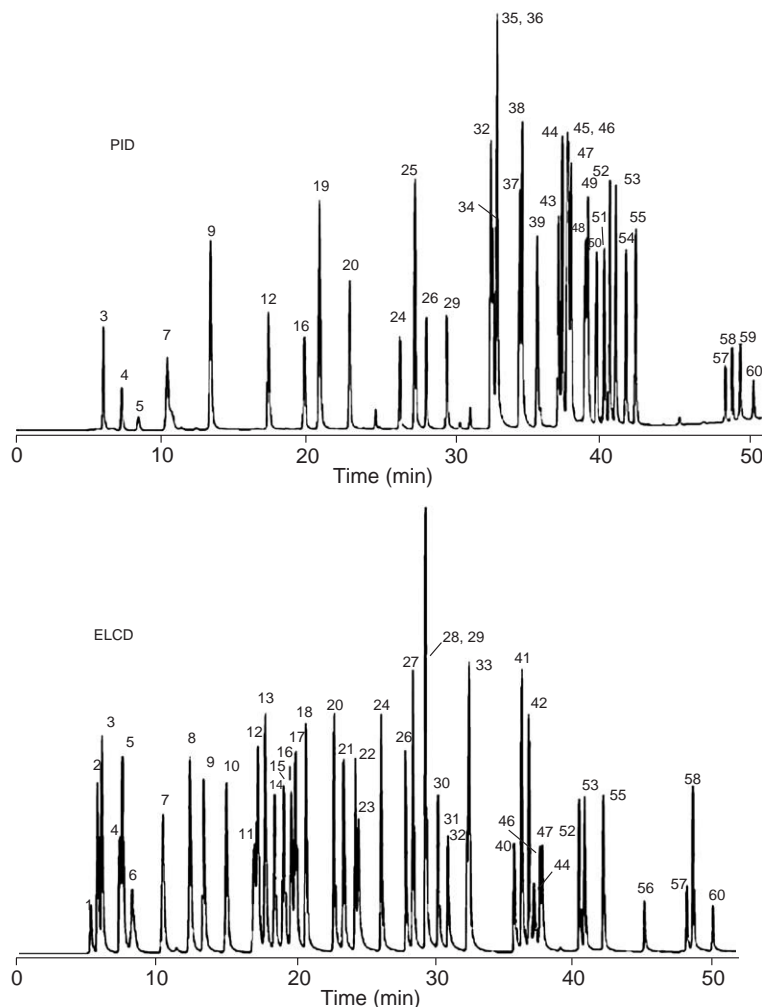
An internal standard is a compound that is not a sample analyte but is similar to the analytes in analytical behavior. It is spiked into each sample and calibration solution at a fixed concentration. The calibration curve is a plot of area ratio of analyte to internal standard versus the analyte concentration in the calibration solutions. The slope of the plot is called the response factor (RF). When a sample is analyzed, the peak areas of the analytes and the internal standard are obtained. The analyte concentration in the sample can be calculated from the area

## Conditions:

Sample: 5 ml of water, 5 ppb each analyte  
 Instrument: O. I. Analytical Model 4560 Purge and Trap  
 Trap: VOCARB® 3000 (Supelco), 10 cm Carboxen® B/6 cm Carboxen 1000/1 cm Carboxen® 1001  
 Purge: 11 min, 25°C, 37 ml min<sup>-1</sup>  
 Desorb: 0.5 min at 220°C  
 Column: VOCOL® (Supelco), 105 m × 0.53 mm ID, 3.0 µm film thickness  
 Oven temperature: 35°C (10 min) to 200°C at 4°C min<sup>-1</sup>, hold  
 Carrier gas: Helium, 8.5 ml min<sup>-1</sup>  
 Detectors: PID (O. I. Analytical Model 5230) and  
 ELCD (O. I. Analytical Model 5220) in series

## Analytes:

- |                                       |                                       |                               |                                 |
|---------------------------------------|---------------------------------------|-------------------------------|---------------------------------|
| 1. Dichlorodifluoromethane            | 16. 1,1-Dichloropropane               | 31. 1,2-Dibromoethane         | 46. 2-Chlorotoluene             |
| 2. Chloromethane                      | 17. Carbon tetrachloride              | 32. Chlorobenzene             | 47. 4-Chlorotoluene             |
| 3. Vinyl chloride                     | 18. 1,2-Dichloroethane                | 33. 1,1,1,2-Tetrachloroethane | 48. <i>tert</i> -Butylbenzene   |
| 4. Bromomethane                       | 19. Benzene                           | 34. Ethylbenzene              | 49. 1,2,4-Trimethylbenzene      |
| 5. Chloroethane                       | 20. Trichloroethylene                 | 35. <i>m</i> -Xylene          | 50. <i>sec</i> -Butylbenzene    |
| 6. Trichlorofluoromethane             | 21. 1,2-Dichloropropane               | 36. <i>p</i> -Xylene          | 51. <i>p</i> -Isopropyltoluene  |
| 7. 1,1-Dichloroethylene               | 22. Bromodichloromethane              | 37. <i>o</i> -Xylene          | 52. 1,3-Dichlorobenzene         |
| 8. Methylene chloride                 | 23. Dibromomethane                    | 38. Styrene                   | 53. 1,4-Dichlorobenzene         |
| 9. <i>trans</i> -1,2-Dichloroethylene | 24. <i>cis</i> -1,3-Dichloropropene   | 39. Isopropylbenzene          | 54. <i>n</i> -Butylbenzene      |
| 10. 1,1-Dichloroethane                | 25. Toluene                           | 40. Bromoform                 | 55. 1,2-Dichlorobenzene         |
| 11. 2,2-Dichloropropane               | 26. <i>trans</i> -1,3-Dichloropropene | 41. 1,1,2,2-Tetrachloroethane | 56. 1,2-Dibromo-3-chloropropane |
| 12. <i>cis</i> -1,2-Dichloroethylene  | 27. 1,1,2-Trichloroethane             | 42. 1,2,3-Trichloropropane    | 57. 1,2,4-Trichlorobenzene      |
| 13. Chloroform                        | 28. 1,3-Dichloropropane               | 43. <i>n</i> -Propylbenzene   | 58. Hexachlorobutadiene         |
| 14. Bromochloromethane                | 29. Tetrachloroethylene               | 44. Bromobenzene              | 59. Naphthalene                 |
| 15. 1,1,1-Trichloroethane             | 30. Chlorodibromomethane              | 45. 1,3,5-Trimethylbenzene    | 60. 1,2,3-Trichlorobenzene      |



**Figure 4** Chromatograms of volatile compounds by purge and trap with CG-PID/ELCD. (Courtesy of Supelco Inc., used by permission of © Sigma-Aldrich Co.)

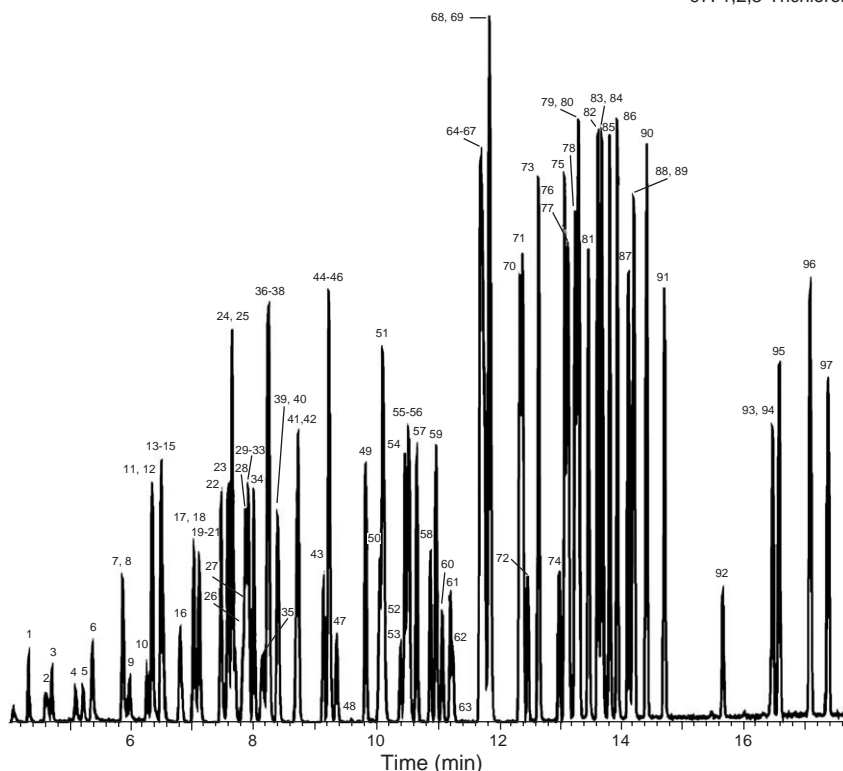


## Conditions:

Sample: 5 ml of water, each analytes at 10 ng ml<sup>-1</sup> (25 ng ml<sup>-1</sup> for ketones)  
 Instrument: LSC-3100 (Tekmar) Purge and Trap  
 Trap: Vocab® 3000 (Supelco), 10 cm Carboxen® B/6 cm Carboxen® 1000/1 cm Carboxen® 1001  
 Purge: 11 min at 40 ml min<sup>-1</sup> at ambient temperature  
 Dry purge: 1 min at 40 ml min<sup>-1</sup>  
 Desorb preheat: 245°C  
 Desorb: 250°C for 2 min, flow 26 ml min<sup>-1</sup>  
 Bake: 260°C for 8 min  
 Interface: Transfer line 0.53 mm ID Silcosteel® tubing 1:20 split at injection port. 1 mm ID sleeve  
 Column: Rtx®-VMS (Restek), 60 m, 0.25 mm ID, 1.40 µm film thickness  
 Oven temperature: 60°C (hold 2 min) to 180°C at 12°C min<sup>-1</sup> (hold 0 min) to 225°C at 45°C/min (hold 6 min)  
 Carrier gas: Helium at ~1.3 ml min<sup>-1</sup> constant flow. Adjust dichlorodifluoromethane to a retention time of 4.03 min at 60°C  
 Detector: HP 5973 MSD, scan range: 35–300 amu

## Analytes:

- |   |  |   |   |
|---|--|---|---|
| 1. Dichlorodifluoromethane                                | 25. Chloroform                                   | 49. <i>cis</i> -1,3-Dichloropropene       | 73. Isopropylbenzene                        |
| 2. Chloromethane  | 26. Ethyl acetate                                | 50. Toluene-d8 (SMC)                      | 74. 4-Bromo-1-fluorobenzene (SMC)           |
| 3. Vinyl chloride   | 27. Methyl acrylate                              | 51. Toluene                               | 75. <i>n</i> -Propylbenzene                 |
| 4. Bromomethane   | 28. Propargyl alcohol (500 ng ml <sup>-1</sup> ) | 52. 4-Methyl-2-pentanone                  | 76. 1,1,2,2-Tetrachloroethane               |
| 5. Chloroethane   | 29. Dibromofluoromethane (SMC)                   | 53. Pyridine (250 ng ml <sup>-1</sup> )   | 77. Bromobenzene                            |
| 6. Trichlorofluoromethane                                 | 30. Tetrahydrofuran                              | 54. <i>trans</i> -1,3-Dichloropropene     | 78. 1,3,5-Trimethylbenzene                  |
| 7. Ethanol (2500 ng ml <sup>-1</sup> )                    | 31. Carbon tetrachloride                         | 55. Ethyl methacrylate                    | 79. 2-Chlorotoluene                         |
| 8. 1,1-Dichloroethene                                     | 32. 2-Butanone                                   | 56. Tetrachloroethene                     | 80. 1,2,3-Trichloropropane                  |
| 9. Carbon disulfide (40 ng ml <sup>-1</sup> )             | 33. 1,1,1-trichloroethane                        | 57. 1,1,2-Trichloroethane                 | 81. 4-Chlorotoluene                         |
| 10. Allyl chloride  | 34. 1,1-dichloropropene                          | 58. Dibromochloromethane                  | 82. <i>tert</i> -Butylbenzene               |
| 11. Methylene chloride                                    | 35. Pentafluorobenzene (IS)                      | 59. 1,3-Dichloropropane                   | 83. 1,2,4-Trimethylbenzene                  |
| 12. Acetone   | 36. <i>tert</i> -Amyl methyl ether               | 60. <i>n</i> -Butyl acetate               | 84. Pentachloroethane                       |
| 13. <i>trans</i> -1,2-Dichloroethene                      | 37. Benzene                                      | 61. 1,2-Dibromoethane                     | 85. <i>sec</i> -Butylbenzene                |
| 14. <i>tert</i> -Butyl alcohol (100 ng ml <sup>-1</sup> ) | 38. Isobutyl alcohol (500 ng ml <sup>-1</sup> )  | 62. 2-Hexanone                            | 86. <i>p</i> -Isopropyltoluene              |
| 15. Methyl <i>tert</i> -butyl ether                       | 39. 1,2-Dichloroethane                           | 63. 2-Picoline (250 ng ml <sup>-1</sup> ) | 87. 1,3-Dichlorobenzene                     |
| 16. Diisopropyl ether                                     | 40. Isopropyl acetate                            | 64. Ethylbenzene                          | 88. 1,4-Dichlorobenzene-d4 (IS)             |
| 17. 1,1-Dichloroethane                                    | 41. 1,4-Difluorobenzene (SMC)                    | 65. Chlorobenzene-D5 (IS)                 | 89. 1,4-Dichlorobenzene                     |
| 18. Acrylonitrile   | 42. Trichloroethene                              | 66. Chlorobenzene                         | 90. <i>n</i> -Butylbenzene                  |
| 19. Vinyl acetate*  | 43. Dibromomethane                               | 67. 1,1,1,2-Tetrachloroethane             | 91. 1,2-Dichlorobenzene                     |
| 20. Allyl alcohol (250 ng ml <sup>-1</sup> )              | 44. Bromodichloromethane                         | 68. <i>m</i> -Xylene                      | 92. 1,2-Dibromo-3-chloropropane             |
| 21. Ethyl- <i>tert</i> -butyl ether*                      | 45. 1,2-Dichloropropane                          | 69. <i>p</i> -Xylene                      | 93. Nitrobenzene (250 ng ml <sup>-1</sup> ) |
| 22. <i>cis</i> -1,2-Dichloroethene                        | 46. Methyl methacrylate                          | 70. <i>o</i> -Xylene                      | 94. Hexachlorobutadiene                     |
| 23. 2,2-Dichloropropane                                   | 47. <i>n</i> -Propyl acetate                     | 71. Styrene                               | 95. 1,2,4-Trichlorobenzene                  |
| 24. Bromochloromethane                                    | 48. 2-Chloroethanol (2500 ng ml <sup>-1</sup> )  | 72. Bromoform                             | 96. Naphthalene                             |
|   |  |   | 97. 1,2,3-Trichlorobenzene                  |



**Figure 5** Chromatograms of volatile compounds by purge and trap with GC-MS. (Courtesy of Restek Corp., used by permission of © Sigma-Aldrich Co.) IS=Internal Standard

ratio divided by the RF. The advantage of using internal standards is that it compensates for the variations in experimental conditions such as sample volume and analyte recovery. It is especially desirable for complex samples where purging efficiency varies with matrices. The main limitations of internal standards are that they must not exist in the original sample, and that they must not interfere with the analysis of the target analytes. For environmental samples, fluorinated compounds can be chosen as internal standards since their presence in the environment is uncommon. The use of MS detectors has made the use of internal standard in purge and trap more feasible because internal standards that may coelute with the analytes in a chromatogram can be differentiated by their difference in mass. Stable isotopically labeled compounds are widely used as internal standards in purge and trap with GC-MS. In general, they offer superior precision and accuracy, because these isotope analogs are almost identical to the target compounds in physical and chemical properties, mimicking the behavior of the analytes in every step of the analysis.

## Applications and Trends

Purge and trap has been widely used in environmental, biological, pharmaceutical, food, and other types of analyses. As the determination of volatile pollutants in water and soil remains its most common application, the following offers a description of such an example.

Both water and soil samples can be collected into standard 40 ml volatile organic analysis vials. Commercial autosamplers are available to accommodate multiple (water and soil) samples in the same sequential run. An aliquot of water sample (5–25 ml) is drawn from the vial, and transferred into the sparge vessel. A certain volume of reagent water is added into the soil vial through a syringe without disturbing the hermetic seal. The soil sample is heated and stirred during purging. The selection of the trap largely depends on the target analytes. Each standard and sample goes through the same purge-and-trap cycle described before. Quantitation can be done by using either external standards or internal standards.

Figure 4 shows the experimental conditions and the chromatograms of  $\text{ng ml}^{-1}$  level priority pollutants in drinking water using purge and trap with GC-PID/electrolytic conductivity detector (ELCD). Figure 5 shows a chromatogram of  $\text{ng ml}^{-1}$  level

volatile organics in solid waste by purge and trap with GC-MS.

Purge and trap has undergone significant developments and refinements since its invention nearly three decades ago. Many improvements have come along with the advances in capillary column technology and the widespread use of GC-MS, which have resulted in faster separation, better resolution, and detection. Faster GC analysis calls for faster purge and trap. To that end, commercial automated dual-unit systems have been introduced to increase sample throughput. Although often regarded as a mature technique, advances in purge and trap will continue to reduce cycle time, improve productivity, and find new applications.

**See also:** **Gas Chromatography:** Environmental Applications. **Headspace Analysis:** Static. **Water Analysis:** Overview; Organic Compounds.

## Further Reading

- Abeel SM, Vickers AK, and Decker D (1994) Trends in purge and trap. *Journal of Chromatographic Science* 32(8): 328–338.
- Bellar TA and Lichtenberg JJ (1974) Determining volatile organics at microgram-per-liter levels by gas chromatography. *Journal of American Water Works Association* 66(12): 739–744.
- EPA Method 502.2 (1995) *Methods for the Determination of Organic Compounds in Drinking Water, Supplement III*. Washington DC: US Environmental Protection Agency, Office of Research and Development.
- EPA Method 524.2 (1995) *Methods for the Determination of Organic Compounds in Drinking Water, Supplement III*. Washington DC: US Environmental Protection Agency, Office of Research and Development.
- EPA method 5030B (1996) *Purge and Trap for Aqueous Samples, Methods for Evaluating Solid Waste, Physical/Chemical Methods* (<http://www.epa.gov/epaoswer/haz-waste/test/sw846.htm>).
- EPA method 5035 (1996) *Closed-system Purge-and-trap and Extraction for Volatile Organics in Soil and Waste Samples, Methods for Evaluating Solid Waste, Physical/Chemical Methods* (<http://www.epa.gov/epaoswer/haz-waste/test/sw846.htm>).
- Snow N, Slack G, and Kou D (2003) Extraction of volatile organic compounds from solids and liquids. In: Mitra S (ed.) *Sample Preparation Techniques in Analytical Chemistry*. New Jersey: Wiley.
- USP Method 467 (1993) *Organic Volatile Impurities*. United States Pharmacopeia, Rockville, MD.

# HERBICIDES

**M N Vasilescu**, Institute of Public Health, Bucharest, Romania

**A V Medvedovici**, University of Bucharest, Bucharest, Romania

© 2005, Elsevier Ltd. All Rights Reserved.

## Definition

A herbicide is any individual substance or mixture that is intended to prevent, control, destroy, repel, or mitigate growth of weeds within a crop. Herbicides belong to the pesticide class.

## General Considerations

### Classification

There are many criteria involved in the classification of herbicides, as presented in **Figure 1**. Classifications according to the way of action (**Figure 2**) and chemical structure (**Figure 3**) are to some extent complementary and provide the best comprehensive basis on which the extensive variety of herbicides can be grouped.

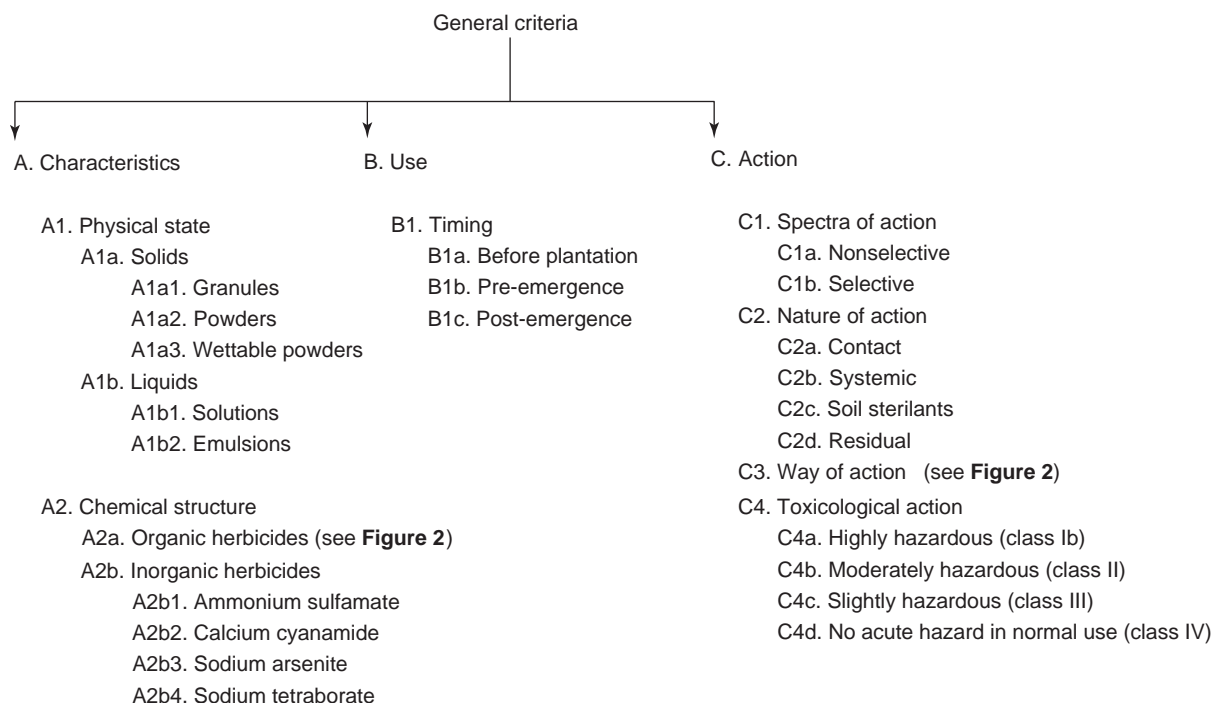
### Herbicide Sources

Herbicide sources are either nonpoint sources or point sources. The nonpoint sources are represented by (1) accidental release on storage; (2) accidental drift during application; (3) surface runoff from land at application points; and (4) leaching through soil. The point sources are represented by (1) direct application points; (2) wastes from manufacturing sites; and (3) spills during mixing, loading, and transportation.

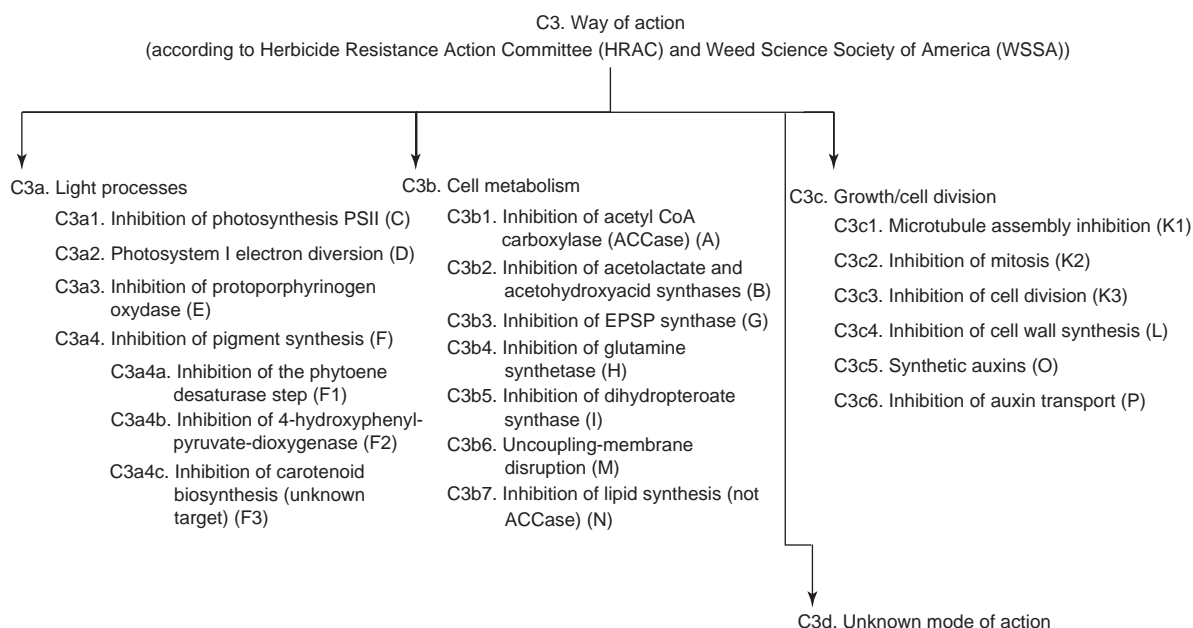
### Herbicide Transport

The transport of herbicides in the environment can be achieved according to the following processes: (1) direct fallout from application points; (2) adsorption on carriers; (3) overland and/or subsurface flow (erosion); and (4) atmospheric deposition (dry or wet).

The migration of herbicides depends on (1) application factors (amount, place, period, frequency, type of formulation, and method of application); (2) persistence and mobility (related to molecular structure and intrinsic physical/chemical properties); (3) soil composition (geomorphology of the application site); (4) geography of the application place; and (5) climate (especially temperature profiles, humidity, rain/storm frequency, and duration).



**Figure 1** General criteria for classification of herbicides.



**Figure 2** Classification of herbicides according to the way of action.



**Figure 3** Classification of herbicides according to the chemical structure.

## Herbicide Effects

Apart from their intended use, herbicides mainly impact on the environment and on human health. Environmental effects should be discussed according to the herbicide toxicity to mammals, birds, fishes, aquatic/soil plants, and microorganisms. The action of the herbicides on human health should be discussed in terms of irritant action, effects on metabolism (uncoupling oxidative phosphorylation, generation of methemoglobin, inhibition of cholinesterase), as well as carcinogenic/mutagenic effects. A rough index of the potential lethal toxicity of a given herbicide is given by the rat acute oral LD<sub>50</sub> dose.

## Herbicide Fate

Most of the lipophilic herbicides tend to undergo enzymatic reactions that make them more water soluble and reactive by the attachment of polar functional groups such as -OH. Most of these reactions are microsomal mixed-function oxidase reactions catalyzed by the cytochrome P-450 enzyme system associated with the endoplasmic reticulum of the cell and occurring most abundantly in the liver of vertebrates. The final products are then involved in conjugation reactions with endogenous conjugating agents (glucuronide, glutathione, sulfate, acetyl) resulting in conjugated species, which have a higher polarity, a greater water solubility, and thus are more easily eliminated.

## Herbicide Persistence

The period of time during which a herbicide remains active in soil at the application area describes its persistence. Its carryover effect describes the presence of the herbicide at the application area once its weed control mission has been accomplished. Transformation processes acting on an herbicide structure are briefly discussed in **Figure 4**. A specific decomposition pattern is illustrated for atrazine in **Figure 5**. Data about degradability and transport of some classes of herbicides are given in **Table 1**. The half-life represents a measure of herbicide persistence and is quantified as the amount of time taken by a herbicide to decompose by 50% from the applied chemical to a herbicidal inactive form. Decomposition products may exhibit higher, reduced, or no herbicidal activity compared to the parent compound, and may exhibit higher, reduced, or no risk effects on environment and human health.

## Herbicide Resistance

The resistance developed by the target weeds to commonly used herbicides over a period of years by

means of natural selection of the occurring biotypes is known as herbicide resistance. Mechanisms for resistance depend on the herbicide's mode of action (e.g., photosynthesis inhibition induced by triazines can be avoided by some weed biotypes by developing slight changes in the chloroplast protein structure). To date, it is known that ~53 species of weeds exhibit resistance to at least five different herbicide classes (e.g., goosegrass and green foxtail to the dinitroaniline congener trifluralin, pigweed and witchgrass to triazines, Russian thistle to the sulfonylurea congener chlorsulfuron).

## Herbicide Market

At present, herbicides are the most widely marketed group of pesticides, including ~267 active compounds formulated in more than 3000 commercial products. The herbicide market in 2001 expressed as a function of the target crop consists of 20% for soybeans and oilseeds, 19% for corn, 17% for cereals, 8% for rice, 6% for fruits and nuts, and 5% for cotton. From the geographic point of view, 40% of the herbicide market is in North America, 25% in Europe, Middle East, and Africa, 18% in Pacific Asia, and 17% in Latin America.

According to the information given in this section, a general characterization file of an herbicide should be presented as given for some examples in **Table 2**.

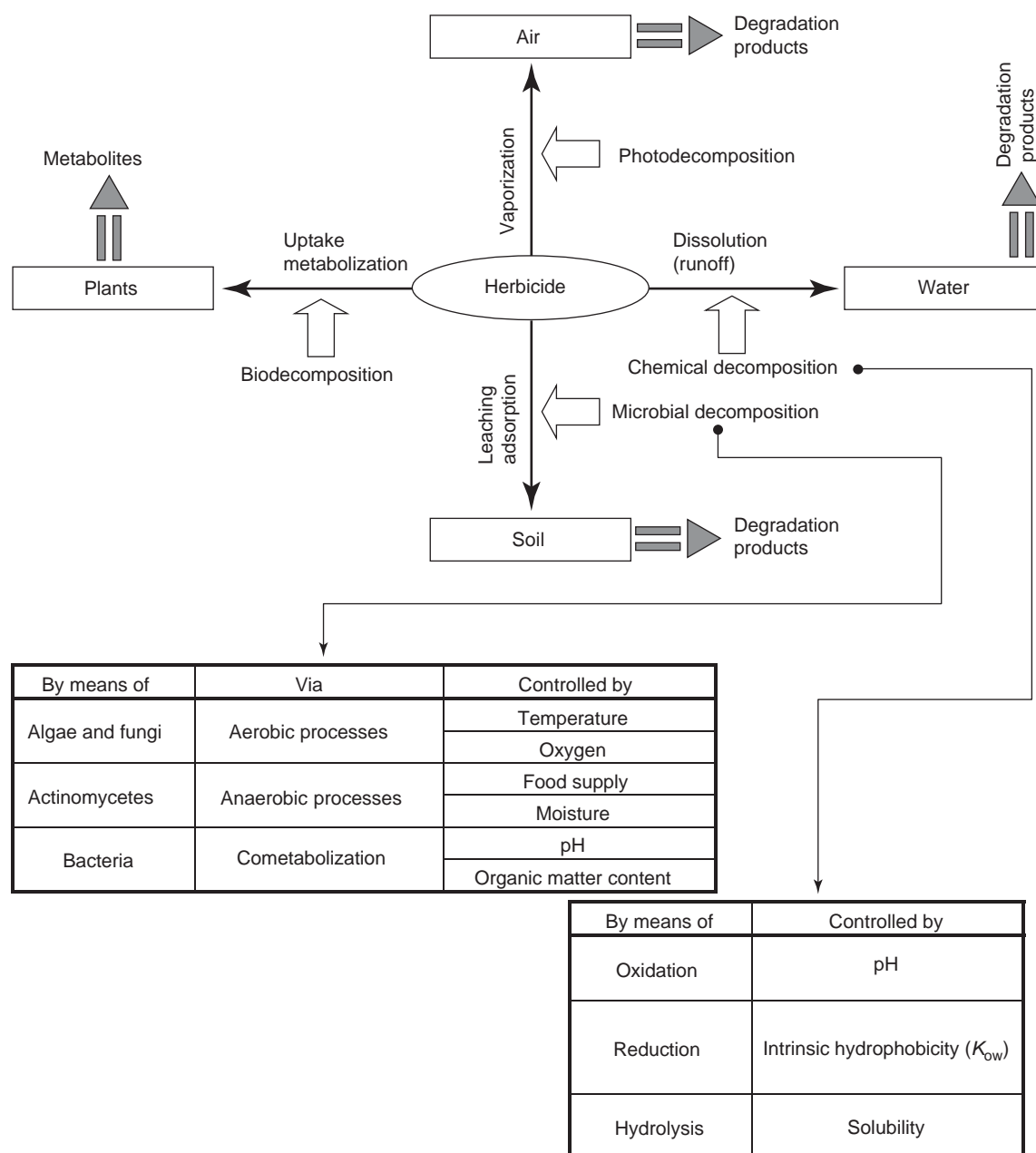
## Herbicide Analysis

The increasing complexity of the chemicals used as herbicides determines the development of more selective and accurate methods of analysis. Basically, analytical aspects related to herbicides should focus on five main directions: (1) quality control of herbicide formulations during production and marketing; (2) research and development for new herbicides; (3) impact on the environment (air, water, soil, and organisms other than humans); (4) impact on food safety; and (5) impact on human health, including forensic aspects.

The targets of the analytical processes are not related only to the determination of the parent compounds. Intermediate synthesis products, breakdown products, and metabolites should also be monitored in more or less complex matrices, sometimes at very low concentrations.

The complexity of the analytical objectives requires not only sensitive and accurate measurements, but, at the same time, correct sampling and sample preservation as well as selective sample preparation methods.





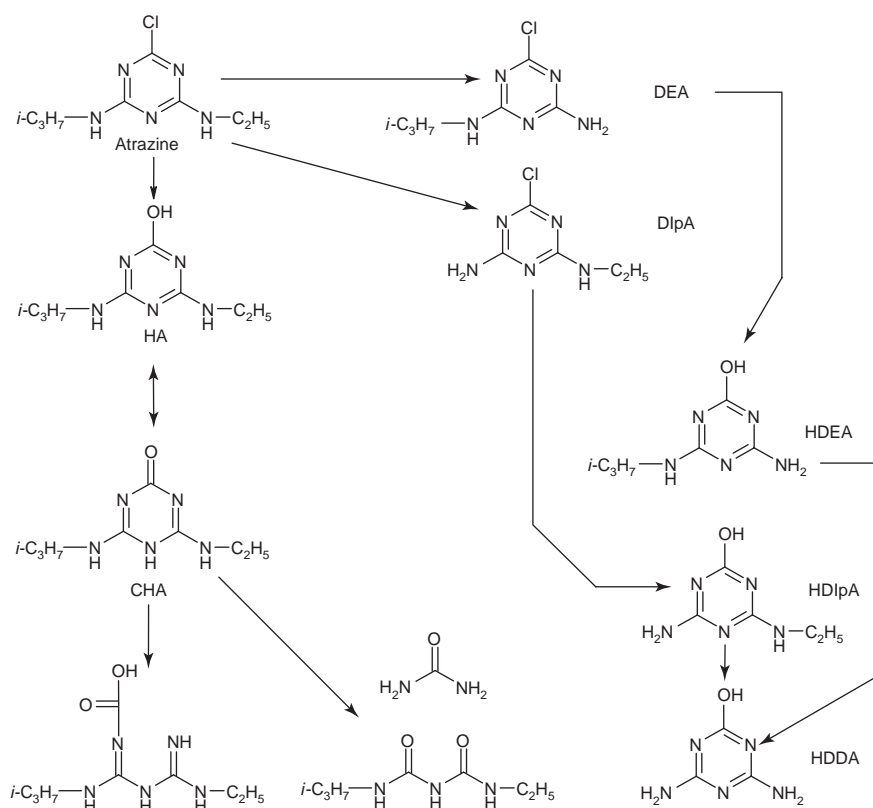
**Figure 4** Transformation processes acting on an herbicide.

## Sampling and Sample Preparation

### Sampling Herbicides

The sampling action should be oriented toward representativity and homogeneity. Sampled quantities should be in accordance with the analytical method further considered and the inherent sensitivity required. Usually, enough material is taken for at least the application of a second full analysis procedure. Traditionally, sampling containers are made of glass. Plastic containers (polyethylene,

polypropylene, or polycarbonate) are often preferred for their robustness. Laboratory tests should be performed prior to the collecting action, in order to demonstrate the suitability of the material of the sampling container. For some particular compounds, adsorption on polymeric materials or even glass may be significant (captan adsorbs readily on glass, while organophosphorus compounds, but not glyphosate, are adsorbed on polyethylene). Mainly for water quality monitoring (environmental analysis) samples may be subjected to concentration at the point of



**Figure 5** Decomposition pattern of atrazine.

**Table 1** Degradability and transport of some classes of herbicides

Class	Degradability			Transport	
	Microbial	Chemical	Photochemical	Vaporization	Leachability
Acetamides	H	L	L	L-M	L-M
Alkyl and aryl ureas	H	L	L	L	M
Bipyridyl derivatives	L	L	M	L	L
Dinitro anilines	L-M	M-H	M-H	L-M	L
Diphenyl ethers	H	L	M-H	L	L-M
Imidazolinones	H	L	L	L	M-H
Nitriles	-	-	-	L	L
Phenoxy-carboxylic acids	H	L	L	L-M-H	L-M-H
Sulfonyl ureas	L-M	M-H	L	L	M-H
Thiocarbamates	H	L	L	H	L-M
Triazines	M-H	L-M-H	L	L	M-H
Ureas	H	L	L	L	M

L, low; M, medium; H, high.

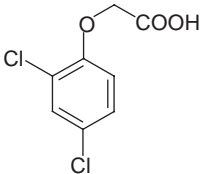
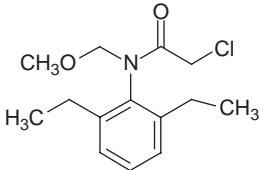
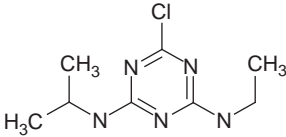
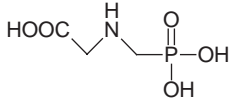
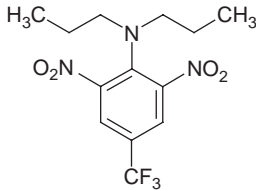
collection. Adsorption cartridges or discs are obtained on the collection site and transported to the laboratory for analysis.

### Sample Preservation and Storage

Preserving sample integrity depends on the stability of the target analytes in a specific matrix. Freezing

samples immediately after collection should provide preservation for thermally unstable extremely volatile compounds. Specific chemical addition of stability enhancing compounds is used occasionally (controlling the pH of the media or blocking active sites of the target analytes by means of derivatization). Contamination and cross-contamination risks should be attentively considered (conditions for transportation

**Table 2** General characterization files for some common herbicides

#	1	2	3	4	5
Name	2,4-D	Alachlor	Atrazine	Glyphosate	Trifluralin
Chemical name	2,4-Dichlorophenoxyacetic acid	2-Chloro-2',6'-diethyl-N-(methoxymethyl)acetanilide	6-Chloro-N-ethyl-N'-(1-methylethyl)-1,3,5-triazine-2,4-diamine	N-phosphono methyl glycine	2,6-dinitro-N,N'-dipropyl-4-trifluoromethyl phenylamine
Classes	A2a3d B1c C3c5 C4b	A2a1d B1c C3c3 C4a	A2a7b4a B1b, B1c C1b C3a1 C4b	A2a10b B1c C3b3 C1a, C4d	A2a9b B1b C2b C3c1, C4b
Structure					
Molecular formula	C <sub>8</sub> H <sub>6</sub> Cl <sub>2</sub> O <sub>3</sub>	C <sub>14</sub> H <sub>20</sub> ClNO <sub>2</sub>	C <sub>8</sub> H <sub>14</sub> ClN <sub>5</sub>	C <sub>3</sub> H <sub>8</sub> NO <sub>5</sub> P	C <sub>13</sub> H <sub>16</sub> F <sub>3</sub> N <sub>3</sub> O <sub>4</sub>
Molecular weight	221.04	269.77	215.69	169.07	335.28
Related compounds	Sodium salt Isopropyl ester Butyl ester Amine salts	None	None	Isopropyl ester Amine salt	None
Physical properties	m.p. 138°C b.p. 160°C	m.p. 40–41°C	m.p. 171–174°C	m.p. 230°C	m.p. 46–47°C m.p. 130–140°C
Water solubility (μg ml <sup>-1</sup> ) at 25°C	620	140	70	12 000	24
Solubility in organics	Alcohols Acetone Benzene Diethyl ether	Alcohols Acetone Benzene Diethyl ether	Chloroform Diethyl ether Methanol	Insoluble	Acetone Xylene
MCL (mg l <sup>-1</sup> )	0.07	0.002	0.003	0.7	0.005
LD <sub>50</sub> (mg kg <sup>-1</sup> )	375	1200	1750	4873	500
Persistence	Low	Moderate	Moderate	Moderate	Moderate–high
Adsorption	Low	Low	Low	High	Moderate–high
Volatility	Low	High	Low	Low	High
Mobility	High	Moderate–high	Moderate–high	High	Low–moderate
Degradation pattern	1. Biodegradation 2. Photodegradation	Aerobic biodegradation	1. Biodegradation 2. Chem. hydrolysis	Aerobic and anaerobic biodegradation	1. Biodegradation 2. UV photodegradation

and manipulation of empty sampling containers and collected samples, plasticizers contamination from container walls, etc.).

### Sample Preparation

Sample preparation methods used for herbicide isolation from different matrices are summarized in Figure 6. Some other details are further discussed.

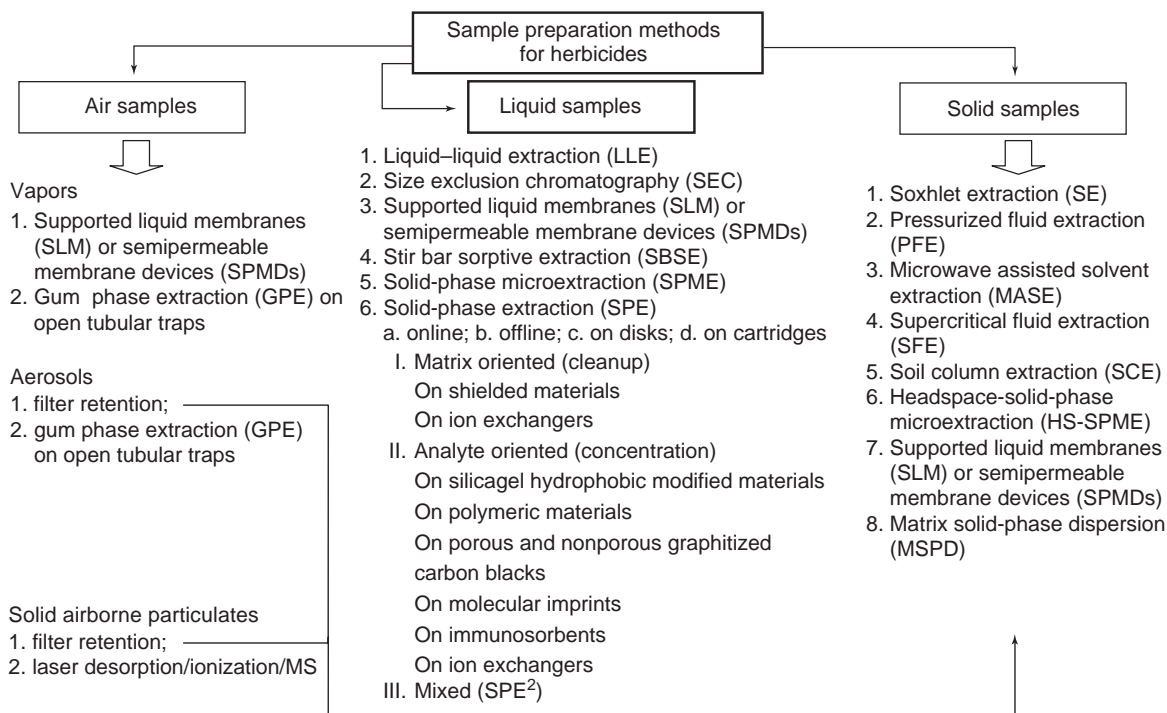
**Liquid-liquid extraction (LLE)** LLE is the classical method used for herbicide isolation, especially from water and biological fluid samples. Ethyl acetate, dichloromethane, and their mixtures are among the preferred extraction solvents for phenylureas, triazoles, amides, carbamates, benzimidazoles, and chlorotriazines. The extraction efficiency is modified by adjustment of pH and ionic strength in the aqueous phase. *In situ* derivatization of the target analytes is also used as an effective tool (e.g., chlorophenoxy acidic herbicides are derivatized with dimethyl sulfate prior to their extraction by n-hexane). The classical way of performing LLE is the separation funnel extraction. Some continuous LLE extractors or steam distillers are also available.

**Solid-phase extraction (SPE)** SPE is used for isolation of the target compounds from liquid media by means of adsorption on a granular solid bed

(cartridge) or on a porous solid membrane (disk). After matrix removal, analytes are desorbed from the solid material using a specific solvent or mixture of solvents. The extract is concentrated by solvent thermal vaporization or gas flush. Direct desorption of the analytes into the liquid chromatography (LC) or supercritical fluid chromatography (SFC) columns by the mobile phase is achievable using an online setup. When using desorption solvents nonmiscible with the solvent of the initial matrix, special attention should be paid to cartridge/disk drying (this operational step is required between sample loading/cleanup and analyte desorption). When using online procedures, it is important to estimate the desorption kinetics of the trapped analytes. Contrarily, serious problems related to analyte focusing and chromatographic efficiency loss may arise.

Four types of adsorbents are usually shared by SPE applications to herbicides. Hydrophobic modified silica materials (C18, C8, C2, C1) are extensively used for a large variety of samples (biological samples such as serum and urine for atrazine, simazine, prometryne, ametryn, sulfonyl ureas, or environmental samples such as different types of waters for alachlor, aldicarb, methiocarb – with concomitant hydrolysis –, medium polar, neutral, and alkaline herbicides, phenyl ureas).

Polymeric materials (styrene-divinyl benzene based copolymers such as PLRP and Lichrolut EN)



**Figure 6** Sample preparation methods for herbicides.

as well as vinyl pyrrolidone–divinyl benzene copolymers (commercially named OASIS) are mainly in current use. These materials behave better than the silica modified adsorbents. They have higher adsorptive properties, allow high volume sample loadings, tolerate alkaline media and larger particles better without any loss of recovery, and are suitable for repetitive use. Special attention should be paid on the desorption step, due to stronger interactions realized between the adsorbent and the trapped analytes. Triazines, ureas, and amidic herbicides are better isolated on the more polar OASIS materials while other compounds including carbamates, thiocarbamates, aryloxypropionic acids, aryloxyphenoxypropionic acids, and related metabolites are well isolated on styrene–divinyl benzene (SDVB) hydrophobic materials.

Nonporous and porous graphitized carbon black solid beds (NPGC and PGC, respectively) are also used for herbicide isolation from water samples. Chlorophenoxy herbicides (requiring online derivatization with tetraalkylammonium salts), acetanilides, neutral diphenyl ethers, triazines, phenyl ureas, triazoles, benzimidazoles, phenoxyalkanoic acids, and quats (paraquat, diquat, difenzoquat) are readily isolated on graphitized carbon beds.

Few applications on herbicide isolation relate to the use of ion exchanging resins. The extraction of chlorophenoxy acidic herbicides residues from green bean samples (2,4-D, MCPA, 2,4-DP, 2,4-DB, MCPB, 2,4,5-TP, and benzoic acid derivative of dicamba) was achieved on hydroxymethylmethacrylate MFE polymer containing quaternary ammonium functional groups. Strong cation exchangers are moreover used for sample cleanup (matrix elimination). Glufosinate and related metabolites in hard water samples and chlormequat in pear, juices, and cereals require matrix cleanup on strong cation exchangers.

Two new directions evolved in the last few years related to the SPE of herbicides: the synthesis of immunosorbents and polymeric molecular imprints. Extraction on immunoaffinity sorbents is based upon the molecular recognition using antibodies. Extraction and cleanup of complex biological and environmental aqueous samples are achieved in a single step even for large volumes. Entrapment of esterases in a ceramic SiO<sub>2</sub> sol–gel matrix leads to isolation of organophosphates and carbamates. Antiatrazine monoclonal antibodies and antidinitrophenyl polyclonal antiserum were also immobilized on the same matrix. Molecular imprints are synthetic polymeric phases on which selective receptors complementary to the target analytes have been generated during preparation. Ethylene glycol dimethylacrylate–methacrylic acid copolymers can be imprinted with

trialkylmelamines or dibuthylmelamine in order to generate templates selective to atrazine. Surface functionalization with molecular imprinted polymers was obtained also for porous polypropylene films. 2-Acrylamido-2-methylpropane sulfonic acid as monomer, *N,N'*-methylbis(acrylamide) as cross-linker and benzophenone as photoinitiator were used in aqueous media for depositing thin films on porous polypropylene. Desmetryn served for producing its own template.

Owing to the increasing complexity of the sample matrices, sequential coupling of two SPE processes (SPE<sup>2</sup>) is sometimes required. Generally, one SPE procedure is addressed to the isolation of the target compounds, while the other acts as a matrix remover (cleanup). A dynamic ion exchange–SPE sequence was applied for the simultaneous isolation of the acidic herbicides and removal of the interfering matrix from environmental waters. The first SPE step is performed on a C18 modified silica gel while the second step is made on an ion exchanging adsorbent. A selective desorption step strongly reduces the interferences of humic acids. Recovery >70% was obtained for chloramben, MCPB, and 2,4-DB using a 0.5 l sample. The same design was applied for analysis of atrazine degradation ozonolysis products. The tandem SPE (on C18 material)/SPE (graphitized carbon deposited on zirconia ZrO<sub>2</sub>) was used for isolation of triazine and carbamates from water samples, while phenolic herbicides were more often isolated on a SPE<sup>2</sup> setup using a SDVB adsorbent in the first step followed by a C1-Hysphere material in the second one.

SPE processes may be subjected to computer assisted optimization. As an example, an orthogonal array design was employed for the optimization of an SPE process applied to atrazine, diazinon, ametryn, and fenthion in surface water. Seven parameters (type of desorption solvent, type of sorbent, flow rate of the elution solvent, sample pH, sample volume, elution volume, organic modifier addition, and flow rate of the water sample) were studied and optimized.

**Solid-phase microextraction (SPME)** SPME consists in the adsorption of the target compounds on a thin polymeric film deposited on the surface of a capillary fiber. The mass transfer can be achieved from liquid media in direct contact with the extracting coated fiber as well as from gaseous environments. Volatile or semivolatile herbicides existing in solid samples can be easily transferred in the gas phase on heating in closed vials, followed by trapping of the resulting vapors in the coated fiber (procedure is known as 'Headspace' HD/SPME).



The nature of the polymeric film as well as its thickness should be related to the polarity characteristics of the extracted compounds. Chloroanilines, sulfamides, phthalimides, and oxazolidones (dichloran, chlorothalonil, vinclozolin, dichlofluanid, captan, folpet, and captofol) were extracted on polyacrylate, polydimethylsiloxane (PDMS), carbowax-divinylbenzene, and PDMS-divinylbenzene films having thicknesses between 30 and 100  $\mu\text{m}$ . Recoveries of the analytes ranged between 70% and 124% when determined in sea water samples. Parameters such as sample pH, ionic strength, organic additives, stirring rates, and contact duration have to be individually considered. Organochlorine herbicides were monitored in landfill leachates using PDMS-coated fibers while ureas (chlorsulfuron, fluometuron, isoproturon, linuron, metobromuron, and monouron) are better isolated on polyacrylate fibers. Trapped compounds can be further thermally desorbed directly into GC inlets, solvent desorbed directly to LC columns, or online/offline desorbed by the supercritical state carbon dioxide for SFC separations.

**Pressure fluid extraction (PFE)** PFE mainly consists of a static or dynamic pressure and temperature assisted liquid–solid extraction. Phenyl ureas from soils and sulfonyl ureas from maize samples were extracted under pressure with methanol at 50°C, using a dynamic setup at 1 ml min<sup>-1</sup> and using a total extractant volume of 25 ml. 2,4-D, 2,4,5-T, dicamba, trichlorpyr, and bentazone have been *in situ* derivatized during the PFE procedure. The variables temperature, pressure, static extraction time, and derivatization reagent amount should be subjected to optimization in order to increase recoveries. Addition of sodium EDTA in the extraction chamber strongly increases 2,4-D recovery.

Water is the most effective modifier of PFE quantitative recovery of chloroacetanilides and triazines in clayey soils. The method requires a pretreatment with 37.6% water after two static extraction cycles with a total of 32 ml of acetone at 1500 psi and 100°C. Extraction time decreases to 22 min (compared to 18 h when using classical Soxhlet extraction) and organic solvent consumption is <80 ml (compared to 300 ml for the same Soxhlet procedure).

An increase of the extraction yield can be obtained by the simultaneous application of a microwave field to the extraction chamber. The resulting procedure is known as microwave-assisted solvent extraction (MASE). Only static procedures are necessary. Atrazine, simazine, terbutometon, terbuthylazine, terbutryn, molinate, and bromacil were extracted from soils by using MASE with methanol, resulting in recoveries >80% for the 5–200 ng per g

concentration interval. Similar results are cited for phenyl ureas.

**Supercritical fluid extraction (SFE)** SFE generally uses supercritical state CO<sub>2</sub> for extracting the target analytes from solid or liquids samples (if liquids are analyzed, adsorption on inert solids is required). The versatility and the enhanced selectivity of the technique relate to the following aspects: (1) the use of the modifiers added to the supercritical state extractant in order to control its polarity; (2) a fully controlled relationship between temperature–pressure–density–solvating power allowing close correlation with the characteristics features of the analytes; (3) the choice of the solid material used for trapping analytes after the supercritical fluid extraction; (4) the choice of the solvent used for desorbing analytes from the solid trapping material; (5) eventual addition of selective adsorbents added to the extraction cartridges acting selectively on the coextracted matrix; (6) either a static or dynamic setup. SFE is used for herbicide analysis in soils and sediments, vegetal materials, foodstuffs, organs, and tissues. Triazine SFE from soils and foodstuffs was extensively studied. The influence of 18 organic modifiers and different solid collectors (florisil produced the best results) were evaluated. A coupled procedure of SPME and SFE for triazines is also available. Interesting applications refers to subcritical water extraction of chlorinated acid herbicides and related esters from soil at 150°C with the *in situ* hydrolysis and online derivatization with trimethylsilyl trifluoroacetamide. The main drawback of SFE lies in terms of the high costs of the equipment and the complexity of the optimization/validation procedures.

**Soil column extraction (SCE)** SCE consists on a specific solvent percolation to a column packed with the sampled material (soil, sediment) containing herbicides. About 100 mmol l<sup>-1</sup> potassium phosphate buffer adjusted to pH 8 at ambient temperature readily extracts imidazolinones, diphenyl ethers, sulfonyl ureas, aryloxyphenoxypionic acids, and triazolopyrimidines from soil samples. The same SCE procedure was applied to 0.2–0.4% carbon-containing sediments for isolation of sulfonamides and triazines with recoveries ranging from 63% to 99%. Further cleanup (e.g., on Carbograph cartridges) or concentration (SPME) may sometimes be necessary.

**Supported liquid membranes (SLMs) or semipermeable membrane devices (SPMDs)** SLMs are passive *in situ* partitioning systems monitoring dissolved organic contaminants (in air/water/sediments). A porous polymeric

membrane (polyethylene or polytetrafluoroethylene, usually) is impregnated with an organic solvent, forming a barrier between two aqueous phases. The target analyte is extracted from the donor phase into the hydrophobic membrane and then back extracted into a second solution, the acceptor. The acceptor choice should be made according to the degree of lipophilicity of the target compounds. Di-*n*-hexyl ether, *n*-undecane, and triolein were successfully used as stagnant acceptors for 17 triazines, DDTs, DDDs, DDEs, diphenylethers, and chloronitrobenzenes. More polar herbicides, such as glyphosate, require cationic carriers incorporated to the membrane (Aliquat 336). The pH of the donor phase strongly influences extraction yields. Correlations between the log of the partition coefficients octanol-water ( $\log P$ ) characterizing target compounds and extraction yields have been made. Herbicide  $\log P$  values were determined by correlating their retention behavior in reverse-phase liquid chromatography (LC) with the water percentage in the mobile phase.

**Matrix solid-phase dispersion (MSPD)** MSPD is a sample preparation process dating back to 1989, realizing simultaneous disruption and extraction of solid or semisolid samples. The method is applied for the isolation of herbicides in tissues (animal or human origin), food, fruits, and vegetables. As an example, atrazine, cyanazine, metribuzin, and simazine were studied to determine their toxicity in aquatic media (including correlations with environmental temperatures and dissolved oxygen content) by monitoring their concentrations in catfish muscle and liver using MSPD as selective extraction step.

**Gum-phase extraction (GPE)** GPE is an application of the sorption phenomena on polymeric materials used during the sampling process at temperatures above their glass transition point ( $T_g$ ). In such conditions, polymeric materials no longer behave as pure solids but enter a gum-like or even liquid-like state with properties similar to those of organic solvents (considering diffusion and distribution constants). The commonly used sorbent is PDMS. Open tubular traps are more commonly coated in 1–3 m lengths of fused silica columns with 0.3–0.5 mm inner diameters and 10–15  $\mu\text{m}$  film thickness. Air samples should be sucked through the coated tubing. Trapped analytes are thermally desorbed into the gas chromatographic column. Extra column cryofocusing may be necessary. Stir bar sorptive extraction is based on the same principle but the polymer covers the external surface of a magnetic bar stirrer. The rod is inserted within the liquid sample and stirred for a defined period of time. The same effect is obtained if PDMS gum is packed into a cartridge and the liquid

sample is pushed through using LC pumps. Analytes are thermally desorbed directly in the gas chromatographic column or back extracted into an appropriate solvent, with subsequent solvent removal for concentration, redissolving, and injection to an LC column. Good recoveries were obtained for triazines analyzed in water samples when isolation was made as stated previously ( $\sim 75\%$  for simazine, atrazine, propazine, and sebuthylazine,  $\sim 80\%$  for terbutryn and prometryn, and  $\sim 110\%$  for terbuthylazine). Cyanazine, metribuzin, desisopropyl-atrazine, and desethylatrazine were poorly recovered ( $< 7\%$ ). Gas chromatography–mass spectrometry (GC–MS) was used for separation and detection.

**Size exclusion chromatography (SEC)** SEC is generally required as an isolation procedure for herbicides in fat matrices. High molecular mass compounds (e.g., triglycerides, sterols) are excluded from the stationary phase material, while the light fraction is eluted later and is collected for subsequent clean-up, separation, or analysis. A practical example is given by the determination of thiofensulfuron methyl and tribenuron methyl in cottonseeds and cotton gin trash. Restricted access (RA) or shielded materials combine the size exclusion principles with reverse-phase interactions and are also successfully used during sample preparation processes. Basically, RA materials have calibrated pores with strongly hydrophobic inner surfaces while the external particle surface is hydrophilic. Apolar analytes with low molecular masses are retained inside the pores while the medium to polar large molecule fraction is excluded. Applications are focused on biological samples (for protein exclusion and for high-speed immunoassays) (the detection principle is based on the chromatographic separation of the immunocomplex that was formed and the free remaining antigen). Immunoassays based on fluorescence detection using RA materials exist for atrazine and 2,4-D.

**Derivatization** Derivatization represents a chemical transformation induced to the target analyte in order to enhance its: (1) selective isolation; (2) chromatographic separation; and (3) sensitive detection. Derivatization is applied to nonvolatile herbicides for making gas chromatographic separations possible. Ampropylfos, glyphosate, glufosinate, and bialafos are reacted with trifluoroacetic anhydride (TFA) in the presence of trifluoroethanol. Phosphoro herbicides containing amino groups are esterified and acetylated simultaneously using acetic acid and trimethylacetate. Bis-trimethylsilyl trifluoroacetamide is often used for reacting hydroxyl or carboxyl active functions. The carboxylic group present in related

phenoxyacetic herbicides can be esterified at the same time or after their extraction. Chlorophenoxy-acidic herbicides may also be derivatized with pentafluorobenzoyl bromide in the presence of 18-crown-6 ether. Urea herbicides are derivatized with heptafluorobutyric anhydride (HPB) in toluene with pyridine addition, for sensitive electron capture detection (ECD). Under such conditions, difluorobenzuron is the only herbicide which decomposes. TFA or pentafluoropropionic anhydride is most reactive and forms more volatile compounds, but the increase in the ECD response is lower to HPB derivatives. 2,4-D, MCPA, 2,4,5-T, MCPB, and CCPB are derivatized through ion pairing prior to isolation using 1-cyano(2-trimethylammonium) ethylbenz(f)isoindole. Chlorophenoxy acids are derivatized for fluorescent detection in LC with 9-anthryldiazomethane. Sulfonyl urea, glibenclamide, and chlorpropamide are ultraviolet (UV) labeled with 2,4 dinitrofluorobenzene while uracils require 4-bromomethyl-6,7-dimethoxycoumarin.

## Separation Methods

### Thin Layer Chromatography

Phenoxyacetic herbicides are generally separated on Silicagel 60, Silicagel G/CaSO<sub>4</sub>, CaSO<sub>4</sub>, or Al<sub>2</sub>O<sub>3</sub>/CaSO<sub>4</sub>. Triazines are separated on silica plates. Preferred solvents for elution are acetone, benzene, carbon tetrachloride, chloroform, ethyl acetate, dioxane, and propanol. For urea herbicides, reversed phase C-18 plates were used for separation. Ion pair-based separation mechanisms are achieved on paraffin or methyl silicone-impregnated silica, using cetrimide or alkylammonium bromide as counterion. Eluents are based mainly on acetone and propanol mixtures. For increasing selectivity, sequential modes or bidimensional techniques are often used. Overpressure high performance thin layer chromatography (TLC) methods, driven at flow rates of 0.7–1 ml min<sup>-1</sup> and pressures of 1–2 MPa also increase separation power and the quality of the resulting plates. Identification is realized with acid–base indicators (e.g., bromophenol blue) for acidic herbicides, with chloroplasts homogenate for triazines, and ninhydrin for herbicides containing primary amino groups. Fluorimetric detection is also used. Densitometric UV detection made in the reflectance mode increases sensitivity and accuracy.

### Gas chromatography

GC is a powerful technique for herbicide separation. Due to the reactive organic functions contained by

the molecules of most herbicides, precolumn derivatization processes are often required.

**Injectors** Hot splitless injection (HSI) should be considered a rugged technique suitable for most herbicides. Flash vaporization is carried out at ~220°C for most analytes. Some phenyl ureas, carbamates, and organophosphorus compounds are susceptible to thermal degradation with HSI. On-column injection or programmed temperature vaporization (PTV) are offering powerful alternatives. Silanized glass wool inserts also favor decomposition of labile compounds. Some applications of large volume injection on PTV are also cited in the literature for phenoxyacetic, trichlorophenoxy acetic, and phenoxypropionic herbicides. Electronic pressure control with pulse programming reduces decomposition of carbamates on injection.

**Columns** Of the GC methods for herbicides reported in literature, more than 90% use capillary columns (25–30 m length, 0.25–0.32 mm inner diameter, and 0.15–0.30 µm film thickness). Medium polarity stationary phases are widely used (SE 52, 54, OV17, DB 5, DB7, or equivalent). Triazines and some phenyl ureas are separated also on apolar PDMSs (OV 1 or equivalent). Organophosphorus herbicides and chlorotriazines are separated with increased selectivity on SPB-35 or polyethylene glycols (Carbowax 20 M, Supelcowax RSL 300, or equivalent).

**Temperature programming** The temperature gradient is the key for tuning selectivity in GC. Starting program temperatures for triazines and phenyl ureas are lower (40–60°C) while higher values are used for phenoxy-carboxylic esters (80–85°C) and organophosphorus derivatives (100°C). Temperature gradients of 15–30°C min<sup>-1</sup> are generally used for less complex samples, up to 260–290°C, and a single ramp. Two different steps are required for complex mixtures or crowded matrices (first gradient range in the interval 10–25°C min<sup>-1</sup>, the second one between 3°C and 6°C min<sup>-1</sup>).

**Detection systems** Halogen-containing herbicides are detected at the picogram level with the ECD (<sup>63</sup>Ni). Low limits are also obtained in some cases with selective nitrogen–phosphorus detector (NPD) or electrolytic conductivity (Hall) detector (ELCD) detection (1 ng). Mass spectrometric detection is widely used due to its universal character as well as for its intrinsic sensitivity and selectivity. Both EI and CI (using methane or isobutane as reagent gases) are used for ionization. Specific ions monitored in the

**Table 3** Specific ions of some herbicides usually monitored in mass spectra

#	Name	Derivative	<i>m/e</i> signals (in the decreasing order of their relative abundance)		
1	Ametryn		227	212	170
2	Atrazine		200	215	202
3	Bentazon	PFB ester <sup>a</sup>	378	420	341
4	2,4-D	PFB ester	177	400	402
5	Dicamba	PFB ester	203	400	402
6	Dichlobenil		171	173	136
7	Dichlorprop	PFB ester	162	414	416
8	Fenoprop	PFB ester	196	448	450
9	Fluoazifop	PFB ester	282	254	507
10	Haloxifop	PFB ester	316	288	541
11	MCPA	PFB ester	380	141	382
12	MCPB	PFB ester	267	142	408
13	Mecoprop	PFB ester	142	169	394
14	Metabuzine		198	199	214
15	Metolachlor		162	238	240
16	Prometryn		241	184	226
17	Propazine		214	229	216
18	Sebuthylazine		200	202	229
19	Simazine		201	186	203
20	Terbuthylazine		214	173	229
21	Terbutryn		226	185	241
22	Triclopyr	PFB ester	210	435	437

<sup>a</sup>Perfluorobutyl ester.

mass spectra for some of the herbicides or herbicide derivatives separated in GC are given in **Table 3**.

### Liquid Chromatography

LC still represents the most extensively used separation technique for herbicides.

**Stationary phases** Octadecyl-modified silica gel (ODS) either in monomeric or polymeric forms, of irregular or spherical shapes, with different end capping levels, is widely used for selective herbicide separation, according to the reversed phase mechanism (RP). Although the classical analytical column dimensions (25 cm length, 4.6 mm inner diameter) are still preferred, a lot of applications deal with the use of narrow bore (25 cm length, 1–2.1 mm inner diameters) or microbore packed (at least 30 cm length, 0.25 mm inner diameter) columns. The monolithic ODS columns introduced recently on the market are offering the advantages of higher elution speeds due to their increased porosity. Only few separations on cyano-modified silica gel were reported for triazines, using the normal phase (NP) separation mechanism. More often, chiral separations (for aryloxypropionic acids, for example) require optically active stationary phases and NP separation mechanisms. Capillary electrically driven chromatography (CEC) seems to represent a powerful alternative to LC. Octadecyldimethyl(3-trimethoxysilylpropyl)ammonium silica is

a stationary phase specially designed for CEC application on herbicide separation. The quaternary amine function generates the anodic electroosmotic flow while octadecyl functions are responsible for solute retention under a RP mechanism.

**Mobile phases** For RP-LC, mixtures of an aqueous component and organic solvents are invariably used. The aqueous component is eventually buffered (e.g., phosphate buffers 1–10 mmol l<sup>-1</sup> for triazines and phenylureas) or acid additivated (e.g., 0.1% phosphoric acid for chlorophenoxy acidic herbicides or 0.1–0.4% for phenoxyacetic herbicides). The organic solvents are mainly methanol, acetonitrile, or their mixtures. For NP-LC, hexane, dichloromethane, ethanol, and *iso*-propanol are frequently used.

**Detection systems** UV detection (including diode array detection, DAD, for spectral confirmation) is extensively used with LC. Triazines are detected at 220, 222, or 230 nm, phenoxyacidic herbicides at 240 nm, chlorophenoxyacidic herbicides at 220, 230, or 280 nm, carbamates at 245 nm, and substituted anilides at 220 nm. The higher value of the analytical wavelength, the poorer is the resulting detection limit and the less is the matrix interference. Fluorescence detection allows higher sensitivity (at least one order of magnitude) compared to UV, but also requires pre- or postcolumn derivatization (fluorescent labeling). Chlorophenoxy herbicides react with



9-anthryldiazomethane to generate fluorescent derivatives (ex. 365 nm, em. 412 nm), phenoxyacidic herbicides are derivatized with monodansyl cadaverine (ex. 366 nm, em. 450 nm), primary amino groups-containing herbicides yield products (ex. 230 or 340 nm, em. 420 or 455 nm) with *o*-phthalaldehyde and 2-mercaptoethanol. Some dinitroaniline derivatives undergo photoreduction in the presence of anthraquinone-2,6-disulfate to yield fluorescent products while benfluralin, trifluralin, isopropalin, and oryzalin are UV decomposed to active fluorescent degradation products. Electrochemical detection (including electrochemical array detection) can be considered as a sensitive alternative although it lacks robustness. MS undoubtedly offers a solution for a sensitive and selective detection. Mild ionization techniques such as electrospray interface (ESI) and atmospheric pressure chemical ionization (APCI) are rapidly covering a large selection of applications. Their flexibility and the fast interchangeability characteristics are strongly reducing restrictions on LC conditions in terms of flow rates, mobile phase

additives, diversity of analytes, etc. As a basic rule, polar compounds are better ionized with ESI while apolar ones with APCI. A few applications are reviewed in Table 4. The major drawback affecting LC/MS coupling is related to reduced possibilities of generating library searchable mass spectra. This inconvenience is overwhelmed by the recent creation of ESI–CID libraries for sulfonyl ureas (reproducibility of the resulting spectra is controlled over the collisional induced dissociation (CID) process realized in the electro spray transport region).

### Supercritical Fluid Chromatography

Some reports exist about separation of herbicides using capillary columns C-SFC (e.g., triazoles on a cross-linked CN with supercritical CO<sub>2</sub>/methanol mobile phase and UV detection). Owing to problems related to injection reproducibility and the overall method robustness, C-SFC is now obsolete. Interesting applications were obtained on packed columns (P-SFC).

**Table 4** Applications of mass spectrometric detection of herbicides in LC techniques

#	Analytes (class of compounds or compound)	Technique	Interface/ionization	Detection limits	Matrix
1	Triazines	LC/MS	APCI (+/-)	50 ng ml <sup>-1</sup>	Food
2	Sulfonylureas, diphenyl ethers, sulfonanilides	LC/MS/MS	TISP (+) (MRM)	10 ng ml <sup>-1</sup>	Soil
3	Benzoylureas	LC/MS	ESI (+/-); APCI (+/-)	10 ng ml <sup>-1</sup>	Ground water
4	Imidazolinones	LC/MS	ISP (+/-); APCI (+/-)	ng ml <sup>-1</sup> level	Tissues
5	Triazines	UHPLC/MS (TOF)	ESI (+/-)	ng ml <sup>-1</sup> level	Water
6	Alachlor	LC/MS	PBI (+ Cl; + EI)	ng ml <sup>-1</sup> level	Environmental water
7	Paraquat, diquat	LC/MS	ESI (+)	100 ng ml <sup>-1</sup>	Water
8	Sulfonylureas	LC/MS	ESI (+)	2–10 ng ml <sup>-1</sup>	Serum
9	Carbamates	LC/MS with buffer suppression	ESI (+)	ng ml <sup>-1</sup> level	Water
10	Imidazolinones, diphenyl ethers, sulfonylureas, aryloxyphenoxypropionic acids, triazolopyrimidines, sulfonanilides	LC/MS/MS	TISP (+) (MRM)	ng ml <sup>-1</sup> level	Environmental water
11	Diphenyl ethers	LC/MS/MS	TISP (+) v (SRM)	ng ml <sup>-1</sup> level	Environmental water
12	5-Trifluoromethylpyridone and metabolites	LC/MS	ESI	ng ml <sup>-1</sup> level	Hydroponically cultivated maize plants
13	Phenylureas	LC/LC/MS	APCI (+)	50 pg ml <sup>-1</sup>	Environmental water
14	Triazines, amides, phenylureas, triazoles, triazinones, benzimidazoles, phenoxyalkanoic herbicides	LC/MS and LC/MS/MS	ESI (-) for acidic compounds APCI (+)	20–100 pg ml <sup>-1</sup> 2–6 pg ml <sup>-1</sup>	Environmental water
15	Chlormequat	LC/MS/MS	ESI	7 ng ml <sup>-1</sup>	Fruits
16	Sulfonylureas, imidazolinones, sulfonamides	LC/MS	ESI (+)	10 pg ml <sup>-1</sup>	Water

TISP, turbo ion spray; MRM, multiple reaction monitoring; UHPLC, ultra-high-pressure liquid chromatography; TOF, time-of-flight mass analyzer; ISP, ion spray; SRM, selected reaction monitoring; ESI, electrospray interface; APCI, Atmospheric pressure chemical ionization.



**Columns** Pure silica, diol, or nitrile modified silica as well as ODS materials are used in P-SFC for herbicide separation. It is to be noted that in SFC all the cited phases behave as normal phases. This means that the stationary phase is always more polar than the mobile phase and, accordingly, the solutes are separated in order of increasing polarity. Reproducibility of retention data is very good, contrarily to NP-LC. Serially coupling two or more of the same type of columns will result in a fair increase of the efficiency while serially coupling of different columns can be a key in controlling selectivity.

**Mobile phases** SF-CO<sub>2</sub>/methanol mixtures are more often used in P-SFC. Gradient elution is offering a gain in terms of selectivity. Usually, methanol mobile phase content starts at 5% and increases up to 20% with 1–3% min<sup>-1</sup>. Flow rates are placed in the 2–3 ml min<sup>-1</sup> interval.

**Other conditions** Temperatures ~50°C are used in P-SFC for separation of herbicides, making this technique valuable for thermally labile compounds. Pressure programming is often used, in order to control the chromatographic run time. Generally, pressure gradients starting from 100 bar at 30 bar min<sup>-1</sup> up to 300 bar allows baseline separation of 8–12 compounds within 20 min. The high desorbing power of SF-CO<sub>2</sub> makes the online coupling of SPE and P-SFC very attractive. Limits in the low ng ml<sup>-1</sup> range are attainable when desorption takes place from the SPE cartridge directly to the P-SFC column. The cartridge drying period after sample loading seems to play a key role in the success of the online coupling.

**Detection** For P-SFC, UV detection (including DAD for spectral confirmation) is mainly used. Short wavelength detection (210 nm) generally increases sensitivity. Good results were obtained for separation of phenyl ureas and carbamates. The use of the APCI interfaces with practically no modifications compared to commercially available LC models increases the potential of the P-SFC technique. Triazines, carbamates, and sulfonyl ureas were separated in P-SFC and detected using APCI-MS. The APCI mass spectra mainly consist of the protonated molecular ion at low cone voltages. Increasing the CID voltage will induce greater fragmentation.

#### Micellar Electrokinetic Chromatography

Micellar electrokinetic chromatography (MEKC) applications are related to phenyl ureas, their corresponding metabolites, and methyl carbamates. Sodium dodecyl sulfate (SDS) and cetyltrimethylammonium

chloride are the commonly used surfactants. Reversed migration (RM)-MEKC using SDS and  $\gamma$ -cyclodextrin was also used for phenyl ureas. Stacking phenomena with RM micelles is often used as additional sample cleanup prior to separation. Addition of 10–20% organic solvent to the migrating media is controlling eventually arising solubility related problems. UV, fluorescence, or photochemical induced fluorescence is used for detection. Z-shaped flow cells in UV detection drastically increase sensitivity.

#### Capillary Zone Electrophoresis

Phenoxy acid herbicides, sulfonyl ureas, quaternary ammonium derivatives (quats), and aryloxy propanoic acids are the main classes of compounds subjected to capillary zone electrophoresis (CZE). Triazines are also separated using nonaqueous CZE, while low pK<sub>a</sub> characterized chlorotriazines require an ion-pair-like solubilization using cationic surfactants (tetradecylammonium bromide, dodecyltrimethylammonium bromide). Chiral selectors are added in CZE for obtaining enantioselectivity. Chiral selectors used for herbicide enantiomeric discrimination are vancomycin,  $\gamma$ -cyclodextrin, ethyl carbonate  $\beta$ -cyclodextrin, cyclohexyl-alkyl- $\beta$ -D-maltoside, sulpropyl ether  $\alpha$ -cyclodextrin, and hexakis(2,3-di-O-methyl)- $\alpha$ -cyclodextrin.

### Analytical Methods

#### Qualitative Reactions

These qualitative methods of analysis are dependent upon characteristic chemical reactions of herbicides to produce colored compounds or metal complexes (see Table 5). The detection limit of these spot tests ranges from micrograms to milligrams.

#### Titrimetric Methods

Acid–base titrations with perchloric or hydrochloric acid characterize the herbicides containing basic moieties such as aliphatic or aromatic primary or secondary amines (atratone, ametryn, prometryn, prometone). Titrations, generally with standard sodium hydroxide solutions, are used for herbicides containing acidic moieties such as carboxyl or phenolic hydroxyl groups (3-amino-*s*-triazole, sesone, maleic hydrazide). Redox titrimetry is specific for herbicides containing a reducing site such as phosphite residues (oxidizable to phosphate) or phenolic structures (oxidizable to quinones). As an example, falone is determined by titration with alcoholic iodine solution. Precipitation titrimetry could also be used for all herbicides for which the organically bound chlorine can be converted to chloride

**Table 5** Color reactions for qualitative identification of herbicides

<i>Herbicide class</i>	<i>Compound</i>	<i>Reagent</i>	<i>Color</i>
Organic acids	TCA	Pyridine + NaOH	Red-violet, after heating
	2,4-D	Formaldehyde + H <sub>2</sub> SO <sub>4</sub>	Pink, after heating
	Endothal	FeCl <sub>3</sub>	Violet-purple
Amides	CDAA	CS <sub>2</sub> , Cu <sup>2+</sup>	Yellow
	Dicryl	NaOH, NaNO <sub>2</sub> + HCl, N-1-naphthylethylenediamine	Purple
Arylcarbamates	Propham	NaOH, NaClO	Blue
Aminotriazoles	Amitrole	NaOH, sodium nitroprusside (Na <sub>2</sub> [Fe(CN) <sub>5</sub> (NO)] · 2H <sub>2</sub> O)	Green

(e.g., sample refluxing with sodium in ethanol). Silver nitrate is used as titrant. Radox and vegadex are quantified in such a manner. Titrimetric methods are exclusively used as quantitation tools for individual compounds in herbicide formulations. The main advantages of chemical methods are their simplicity and low analysis costs. However, selectivity and sensitivity are poor.

### Spectrometric Methods

Spectrometric methods are commonly used for structural characterization. A linear relationship between absorbance and concentration of the chemical species makes them also suitable for quantification of small concentrations of analyte.

**Ultraviolet-visible spectrometry** UV activity is related to the existence in the herbicide structure of aromatic rings and unsaturated heterocycles. For accurate quantification of herbicides, elimination of matrix effects is an essential process. In some cases, hydrolysis is used in order to displace the analyte signal to a spectral region unaffected by matrix interferences. Table 6 shows the absorption maxima and molar absorptivities of some widely used herbicides. There are only a few herbicides having spectral activity in the visible region of the spectrum. The intensities of the bands are low, and consequently are unsuitable for trace analysis. Sensitive detection can be achieved, however, by derivatization. For increasing selectivity, derivative UV spectrometric methods can be applied for more complex mixtures (e.g., zero-crossing technique). Simazine, propazine, hexazinone, bromacil, and metoxuron were analyzed accordingly. On using derivative spectrometry, relative standard deviations (RSDs) for the determined values are normally below 4% and recoveries ranged in the interval 95–110%.

**Infrared spectrometry** This technique has powerful applications in the structure confirmation of

**Table 6** UV absorption data for some herbicides

<i>Compound</i>	<i>Absorption maximum (nm)</i>	<i>Molar absorptivity (l cm<sup>-1</sup> mol<sup>-1</sup>)</i>
2,3,6-TBA	206	35 000
	277	320
	286	270
Fenac	204	4 300
	278	340
	284	280
Amiben	214	35 000
	298	2 400
2,4-D	201	3 800
	230	7 500
	283	1 900
	291	1 600
2,4,5-T	207	61 000
	298	1 200

herbicides. It can be used as a detection system in GC separations, if the functioning principle of the detector is Fourier transform based. The existence of a great number of spectral bands allows the analytical scientist to choose one that is less affected by matrix interferences. Sensitivity of IR spectrometry for quantitative methods is usually limited to microgram amounts. For GC-FTIR detection, multiple reflection detection cells or the matrix isolation setup will increase sensitivity.

**Nuclear magnetic resonance (NMR) spectrometry** NMR is mainly used for structural studies. Its intrinsic lack of sensitivity is compensated by the deep structural information provided (including conformation, chirality, inclusion phenomena, etc.). Full characterization of the major metabolites of 5-trifluoro-methylpyridone has been achieved by means of NMR detection in LC. The combination of NMR and MS data allowed identification of the N-glucoside and O-malonylglucoside conjugates of the parent pyridone. NMR (<sup>31</sup>P) was also used for the determination of the stability constants of the

complexes between some transition cations and nitriolo-tris(methylenephosphonato) herbicides.

**Fluorimetry and chemiluminometry** Only a few herbicides are fluorescent (e.g., indole 3-acetic acid, 1-naphthylacetic acid, and 2-naphthoxyacetic acid). As a general rule, the presence of the chlorine atoms in the molecule quenches fluorescence. Nonfluorescent herbicides are transformed to fluorescent ones by derivatization. Quantification of herbicides is also possible by measuring the radiative emission generated during a specific chemiluminescent reaction of the analyte. Phenolic herbicides are easily determined in soil samples by the peroxyoxalate chemiluminescent reaction. The first step of the process is the oxalate-herbicide ester formation in anhydrous media, using oxalyl chloride as reagent. The ester is then dissolved in tetrahydrofuran containing 9,10-diphenylanthracene. Chemiluminescence is generated by direct addition into the system of a concentrated hydrogen peroxide solution. Determination is made 1 and 3 s after peroxide addition. Bromofenoxim, bromoxynil, and *p*-cyanophenol were determined in such a way at the  $1 \text{ pmol l}^{-1}$  level.

**Mass spectrometry** Some herbicides can be determined directly in unextracted solid samples by thermolysis-APCI-MS/MS (a single MS process is strongly affected by matrix interferences). The condition for satisfactory results is the thermal stability of the analytes under the analytical conditions, yielding mass spectra with intense molecular ions at higher *m/e* values. Determination of atrazine, ametryn, and propazine fall in the  $10\text{--}100 \text{ pg ml}^{-1}$  range when using MS/MS and multiple reaction monitoring (MRM). Aerosol mass spectrometry with laser-desorption/ionization has also been investigated as a possible tool for real-time monitoring of the presence of herbicides on the surface of airborne soil particles. This technique applied for paraquat produces only singly charged ions, the most abundant species being  $\text{M}^+$ ,  $[\text{M} - \text{H}]^+$ , and  $[\text{M} - \text{CH}_3]^+$ .

## Electrometric Methods

**Amperometric biosensors** Biosensors are devices coupling biological components (providing high selectivity and/or sensitivity) to a physicochemical transducer recording and amplifying the signal. A typical amperometric microcell contains a working and counter electrodes of platinum and an SCE reference electrode. A neutral buffer containing thylakoid membranes isolated from plant leaves immobilized on albumin, introduced in the amperometric cell, under irradiation of a halogen lamp at

$109 \text{ W cm}^{-2}$ , produces a photocurrent. Triazine herbicides induce a decrease of the photocurrent proportional to the concentration of the sensed compound. Class-selective amperometric microcells (e.g., triazine and phenylurea type selective against phenolic herbicides) were developed using photosynthesis complexes (PSII) isolated from the thermophilic cyanobacteria *Synechococcus elongatus*. The biosensing element is immobilized on the surface of a screen printed sensor composed of a graphite working electrode and an Ag/AgCl reference electrode deposited on a polymeric substrate. The microcell is illuminated. The principle of detection is based on the fact that herbicides are selectively blocking PSII electron transport activity in a concentration-dependent manner. Changes of the activity were registered as the rate of the photoreduction of an artificial electron acceptor. Detection limits fall in the  $10^{-9} \text{ mol l}^{-1}$  range.

Pseudohomogenous amperometric displacement immunosensors have also been developed. The displacement assay is performed on a disposable screen printed carbon electrode and takes advantages of the cross-reactivity of some monoclonal antibodies. As an example, the monoclonal anti-2,4-D antibody exhibits a relative cross-reactivity toward immobilized MCPA. In the presence of the target analyte (2,4-D) a displacement effect is observed. Consequently, the peroxidase label of the remaining anti-2,4-D antibodies is detected at the carbon electrode, when adding hydrogen peroxide and potassium iodide as redox mediator, at 0 V versus Ag/AgCl.

**Polarographic techniques** Polarographic processes can be generated by reduction or oxidation reactions involving electrochemically active functional groups (e.g., nitro, carbonyl, carboxyl, carboimide groups). The polarographic signals of different herbicides present together in the sample may exhibit great overlap (e.g., atrazine against simazine, terbutryn against prometryn). For this reason, different multi-component approaches such as partial least squares and artificial neural networks could be used for enhancing resolution. Square wave and differential pulse voltammetric methods are proposed for the determination of paraquat and diquat, based on their oxidation peak at the glassy carbon electrode, following their degradation in strongly alkaline media. Differential pulse polarographic determination of aziprotryne and desmetryn based on their reduction in biphasic media (as oil in water emulsions) was also reported. Sulfonyl ureas are also subject to differential pulse polarography, in homogenous aqueous methanolic potassium chloride media, acidified with perchloric acid. Cyclic, differential pulse, and square wave voltammetry were also applied for

determination of molinate, bensulfuron-methyl, mefenacet, and thiobencarb used to treat rice crops. Attention should be focused on the adsorption and consequent poisoning of the electrode surface by the products of the electrochemical reaction. Undoubtedly, polarographic techniques are the most sensitive electrometric methods for herbicide determination, but are not selective enough for most of environmental samples.

### Enzyme Methods

The enzymatic assay of herbicides is methodologically based on their common pharmacological property to inhibit the activity of a group of enzymes involved in the hydrolysis of choline esters. When an enzyme activity is measured under controlled conditions, the degree of inhibition induced by the sample is a measure of the quantity of the target herbicide. The enzymatic analytical process involves competitive interactions between the enzyme and the substrate, on the one hand, and between the enzyme and the target analyte, on the other.

### Immunoassay

Chemical measurements based on immune reactions have proved to be versatile tools for herbicide analysis. Crude extracts of plant tissues, soil, water, and biological specimens such as blood, urine, or bovine rumen fluids can be assayed for herbicide contamination using enzyme multiplied immunoassay technique or enzyme-linked immunosorbent assay (ELISA). Both techniques are designed to measure microgram or sub-microgram levels of herbicides in test samples, combining the convenience, speed, and reproducibility of the enzyme measurement with the selectivity and the sensitivity of immunoassays. Immunoassays can also be developed for the herbicide metabolites. As an example, hydroxypropazine, hydroxyatrazine, and hydroxysimazine assays are based on a covalent immobilization of antigen in combination with an enzyme-labeled anti-hydroxy-s-triazine monoclonal antibody. Sometimes, even such selective systems do not differentiate the parent structure from its metabolites. The competitive indirect ELISA method developed for quantitation of glyphosate using glyphosate polyclonal antisera did not cross-react with a large number of herbicides, but did cross-react with the metabolite aminomethylphosphonic acid and a structurally related herbicide glyphosine. Contrarily, polyclonal antibodies developed in rabbits against the haptenic analog *N*-(2,6-dinitro-4-trifluoromethylphenyl)-*N*-n-propyl-6-amino-hexanoic acid exhibit increased selectivity for the herbicide trifluralin, even in complex matrices such

as vegetable juices. Some interference from ethalfluralin and benfluralin are observed, while isopropalin and pendimethalin did not show cross-reactivity. Because herbicides are often extracted from different matrices in organic solvents, a lot of work was directed toward the ability of immunosystems to act in mixed water/organic or even pure organic media. Finally, the existence of affinity interactions in pure organic solvent has been proved (e.g., 2,4-D and DCB modified with a polystyrene moiety functioning as a mass label interact in pure toluene, triazine, and phenylureas could be detected in pure ethanol and hexane using covalent immobilization of the respective monoclonal antibodies K4E7 and B76-BF5).

### Biological Assay

Biological assay is based on a measurable growth response from a test organism, as a result of the biological action of any compound. In many instances, this growth response will be an inhibition of elongation or a curvature of an organ such as a stem. The validity of biological assay procedures is based on two assumptions: (1) the plant response increases with the dose of the herbicide; and (2) within the limits of sample variation, these responses are reproducible when the plant material and the environmental conditions are the same. Under controlled conditions these techniques are very sensitive and accurate for solutions of pure chemicals. Their use for determination of herbicides in a matrix of plants or soils is seriously affected by errors. Biological assay is suitable for the phenoxyacetic acids (2,4-D, MCPA, 2,4,5-T) and triazinic herbicides.

*See also:* **Environmental Analysis.** Extraction: Microwave-Assisted Solvent Extraction; Pressurized Fluid Extraction; Supercritical Fluid Extraction; Solid-Phase Extraction; Solid-Phase Microextraction. **Gas Chromatography:** Overview, Mass Spectrometry; Environmental Applications. **Immunoassays:** Overview. **Liquid Chromatography:** Overview; Reversed Phase; Size-Exclusion; Liquid Chromatography–Mass Spectrometry. **Pesticides.** **Supercritical Fluid Chromatography:** Overview; Applications. **Thin-Layer Chromatography:** Overview. **Water Analysis:** Organic Compounds.

### Further Reading

- Anton K and Berger C (eds.) (1997) *Supercritical Fluid Chromatography with Packed Columns*, pp. 161–194, 369–401. New York: Marcel Dekker.
- Barr DB (2002) Analytical methods for biological monitoring of exposure to pesticides: a review. *Journal of Chromatography A* 778: 5–29.



- Byrne ME, Park K, and Peppas NA (2002) Molecular imprinting within hydrogels. *Advanced Drug Delivery Reviews* 54: 149–161.
- Camel V (2000) Microwave-assisted solvent extraction of environmental samples. *Trends in Analytical Chemistry* 194: 229–248.
- Dascenzo C, Curini R, Gentili A, Bruno F, Marchese S, and Perret D (2000) Determination of herbicides in water using HPLC–MS techniques. *Advances in Chromatography* 40: 567–598.
- Krutz LJ, Senseman SA, and Sciumbato AS (2003) Solid-phase microextraction for herbicide determination in environmental samples. *Journal of Chromatography A* 999: 103–121.
- Patel PD (2002) (Bio)sensors for measurement of analytes implicated in food safety: a review. *Trends in Analytical Chemistry* 212: 96–115.
- Picó Y, Rodriguez R, and Mañes J (2003) Capillary electrophoresis for the determination of pesticide residues. *Trends in Analytical Chemistry* 223: 133–151.
- Ramos L, Kriestenson EM, and Brinkman UATH (2002) Current use of pressurised liquid extraction and subcritical water extraction in environmental analysis. *Journal of Chromatography A* 975: 3–29.
- Schobel U, Barzen C, and Gauglitz S (2000) Immunoanalytical techniques for pesticide monitoring based on fluorescence detection. *Fresenius Journal of Analytical Chemistry* 366: 646–658.
- Stan HJ (ed.) (1995) *Analysis of Pesticides in Ground and Surface Water*. Berlin: Springer Verlag.
- Tadeo JL, Sanchezbrunete C, Perez RA, and Fernandez MD (2000) Analysis of herbicides in cereals, fruits and vegetables. *Journal of Chromatography A* 882: 175–191.

## HEROIN

**M L Merves and B A Goldberger**, University of Florida College of Medicine, Gainesville, FL, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Narcotic analgesic agents, including heroin, morphine, codeine, hydrocodone, oxycodone, and methadone, account for a significant number of drug-related illnesses, injuries, and deaths. According to the 2002 data collected by the US Department of Health and Human Services Drug Abuse Warning Network (DAWN), heroin was the third most frequently mentioned illicit drug in emergency-room episodes within the US. Current trends suggest an increase in heroin use among younger populations. In addition, the preferred route of administration seems to be changing from injecting to smoking or snorting. This change may be due to an increased awareness of HIV and other infectious diseases, as well as the increase in purity of heroin.

Heroin (3,6-O-diacetylmorphine, diamorphine) was first synthesized from morphine by an English chemist, CR Wright, in 1874, and was introduced commercially as a morphine-like analgesic by the Bayer Company of Germany in 1898. Heroin was initially used as a morphine and codeine substitute for the treatment of tuberculosis and other respiratory illnesses. However, it quickly became apparent that this drug was more toxic than morphine and exhibited equivalent abuse liability and tolerance-

producing properties. As a result, during the early 1900s, laws were enacted in many countries to reduce the demand for opiates. For example, legislation was passed in 1914 in the US to restrict the distribution of narcotics, but use in medical practice was still permitted. Several years later, federal law prohibited any use, including medical, of heroin. An epidemic of heroin use in the US occurred during the post-World War I period and more recently, during the early 1970s. Despite continuing attempts at restricting the supply and demand for heroin, illicit use continues unabated and is a major problem for most law-enforcement agencies. Since heroin has a substantial potential for abuse and has no currently accepted medical use in the US, it is presently classified as a Schedule I substance according to the US Code of Federal Regulations. In accordance with measures recommended by international organizations such as the World Health Organization, other nations classify heroin similarly in order to control availability, and if applicable, ensure that heroin is used exclusively for medical and scientific reasons. There are various United Nations (UN) agreements that address the use and misuse of opium alkaloids. The Single Convention on Narcotic Drugs of 1961 as amended by the 1972 Protocol, the Convention on Psychotropic Substances of 1971, and the United Nations Convention against Illicit Traffic in Narcotic Drugs and Psychotropic Substances of 1988 help to establish the guidelines for the international manufacture, distribution, and consumption of narcotic drugs. Many countries adhere to the policies set forth by these treaties and the International Narcotics



- Byrne ME, Park K, and Peppas NA (2002) Molecular imprinting within hydrogels. *Advanced Drug Delivery Reviews* 54: 149–161.
- Camel V (2000) Microwave-assisted solvent extraction of environmental samples. *Trends in Analytical Chemistry* 194: 229–248.
- Dascenzo C, Curini R, Gentili A, Bruno F, Marchese S, and Perret D (2000) Determination of herbicides in water using HPLC–MS techniques. *Advances in Chromatography* 40: 567–598.
- Krutz LJ, Senseman SA, and Sciumbato AS (2003) Solid-phase microextraction for herbicide determination in environmental samples. *Journal of Chromatography A* 999: 103–121.
- Patel PD (2002) (Bio)sensors for measurement of analytes implicated in food safety: a review. *Trends in Analytical Chemistry* 212: 96–115.
- Picó Y, Rodriguez R, and Mañes J (2003) Capillary electrophoresis for the determination of pesticide residues. *Trends in Analytical Chemistry* 223: 133–151.
- Ramos L, Kriestenson EM, and Brinkman UATH (2002) Current use of pressurised liquid extraction and subcritical water extraction in environmental analysis. *Journal of Chromatography A* 975: 3–29.
- Schobel U, Barzen C, and Gauglitz S (2000) Immunoanalytical techniques for pesticide monitoring based on fluorescence detection. *Fresenius Journal of Analytical Chemistry* 366: 646–658.
- Stan HJ (ed.) (1995) *Analysis of Pesticides in Ground and Surface Water*. Berlin: Springer Verlag.
- Tadeo JL, Sanchezbrunete C, Perez RA, and Fernandez MD (2000) Analysis of herbicides in cereals, fruits and vegetables. *Journal of Chromatography A* 882: 175–191.

## HEROIN

**M L Merves and B A Goldberger**, University of Florida College of Medicine, Gainesville, FL, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Narcotic analgesic agents, including heroin, morphine, codeine, hydrocodone, oxycodone, and methadone, account for a significant number of drug-related illnesses, injuries, and deaths. According to the 2002 data collected by the US Department of Health and Human Services Drug Abuse Warning Network (DAWN), heroin was the third most frequently mentioned illicit drug in emergency-room episodes within the US. Current trends suggest an increase in heroin use among younger populations. In addition, the preferred route of administration seems to be changing from injecting to smoking or snorting. This change may be due to an increased awareness of HIV and other infectious diseases, as well as the increase in purity of heroin.

Heroin (3,6-O-diacetylmorphine, diamorphine) was first synthesized from morphine by an English chemist, CR Wright, in 1874, and was introduced commercially as a morphine-like analgesic by the Bayer Company of Germany in 1898. Heroin was initially used as a morphine and codeine substitute for the treatment of tuberculosis and other respiratory illnesses. However, it quickly became apparent that this drug was more toxic than morphine and exhibited equivalent abuse liability and tolerance-

producing properties. As a result, during the early 1900s, laws were enacted in many countries to reduce the demand for opiates. For example, legislation was passed in 1914 in the US to restrict the distribution of narcotics, but use in medical practice was still permitted. Several years later, federal law prohibited any use, including medical, of heroin. An epidemic of heroin use in the US occurred during the post-World War I period and more recently, during the early 1970s. Despite continuing attempts at restricting the supply and demand for heroin, illicit use continues unabated and is a major problem for most law-enforcement agencies. Since heroin has a substantial potential for abuse and has no currently accepted medical use in the US, it is presently classified as a Schedule I substance according to the US Code of Federal Regulations. In accordance with measures recommended by international organizations such as the World Health Organization, other nations classify heroin similarly in order to control availability, and if applicable, ensure that heroin is used exclusively for medical and scientific reasons. There are various United Nations (UN) agreements that address the use and misuse of opium alkaloids. The Single Convention on Narcotic Drugs of 1961 as amended by the 1972 Protocol, the Convention on Psychotropic Substances of 1971, and the United Nations Convention against Illicit Traffic in Narcotic Drugs and Psychotropic Substances of 1988 help to establish the guidelines for the international manufacture, distribution, and consumption of narcotic drugs. Many countries adhere to the policies set forth by these treaties and the International Narcotics

Control Board plays a role in monitoring adherence to these policies.

## Chemistry

Heroin is produced by the direct acetylation of morphine, an alkaloid obtained from the dried latex of incised unripe flower pods of the opium poppy *Papaver somniferum* L. Opium is cultivated in Southeast Asia, Turkey, Pakistan, Afghanistan, India, Iran, Nigeria, and Mexico. The only chemicals needed to produce heroin clandestinely from opium are calcium hydroxide, a pH modifier, and an acetylating agent. As shown in **Figure 1**, in the presence of excess acetic anhydride and heat, the 3- and 6-hydroxyl groups of morphine are acetylated, yielding heroin. After acetylation, the solution is usually neutralized with sodium carbonate and treated with concentrated hydrochloric acid in order to produce heroin hydrochloride. ‘Street-grade’ heroin is usually sold as the hydrochloride salt.

The predominant constituents of raw opium are alkaloids, fatty acids, sterols, alcohols, resins, sugars, and plant fragments. At least 30 alkaloids have been identified and have been classified into two categories, phenanthrenes and benzyloquinolines. The phenanthrene alkaloids include morphine, codeine, and thebaine, and the benzyloquinoline alkaloids consist of papaverine and noscapine. If not removed during preparation, many of these compounds will be present in heroin as minor contaminants.

The chemical and physical appearance of illicit heroin varies depending upon its origin and synthesis. Pure heroin is a fine white powder, whereas illicit heroin varies in appearance from white to dark brown. Crudely processed heroin, known as black tar, is dark brown to black. The quantity and purity of heroin produced during illicit manufacture is a function of the alkaloidal content of the raw opium, the procedures used to extract morphine, and the techniques used during acetylation and the purification processes. Prior to distribution, illicit heroin is usually mixed with bulk diluents such as quinine,

mannitol, dextrose, lactose, and baking soda. Paracetamol, lidocaine, and procaine are sometimes added to enhance the effect of heroin or to relieve pain during injection. The purity of street-grade heroin in the US varies greatly from 1% to 60%. If sold undiluted, the purity of black tar heroin may range as high as 40–80%. **Table 1** summarizes compounds often found in illicit heroin. Using various separation techniques and statistical methods, these compounds can be separated from the heroin and characterized in order to track the origin of the heroin sample.

The physicochemical effect of the addition of two acetyler functional groups to morphine is to decrease molecular polarity, and thus increase lipid solubility and membrane permeability. Heroin is a strong base with a  $\text{pK}_a$  of 7.6 at  $23^\circ\text{C}$  and readily hydrolyzes to 6-acetylmorphine under various conditions. Heroin is especially susceptible to base-catalyzed hydrolysis, but will also hydrolyze in the presence of protic solutions including alcoholic and

**Table 1** Compounds commonly found in illicit heroin

### *Alkaloids found in opium*

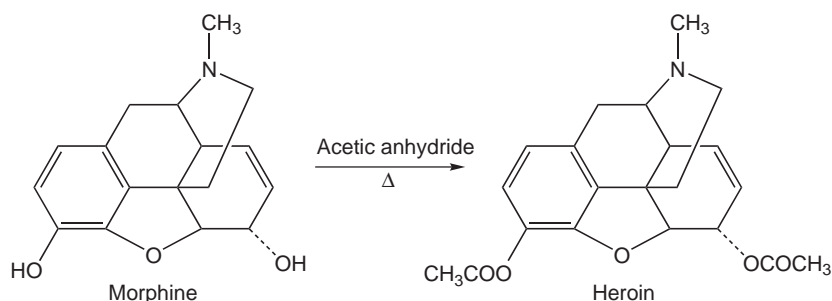
Codeine  
Morphine  
Noscapine  
Papaverine  
Thebaine

### Agents added to heroin

Baking soda  
Boric acid  
Cornstarch  
Dextrose  
Lactose  
Mannitol  
Talc  
Quinine  
Paracetamol  
Lidocaine  
Procaine

### Agents formed during synthesis of heroin

Acetylcodeine  
Acetylmorphine



**Figure 1** Synthesis of heroin from morphine.

aqueous media. Free base heroin is soluble in chloroform, alcohol, and ether and is insoluble in water. Heroin hydrochloride is soluble in chloroform, alcohol, and water. Heroin is increasingly unstable as the temperature of the solution increases. As a result of its instability, aqueous pharmaceutical preparations of heroin must be prepared shortly before use in order to ensure potency. A summary of the physical and

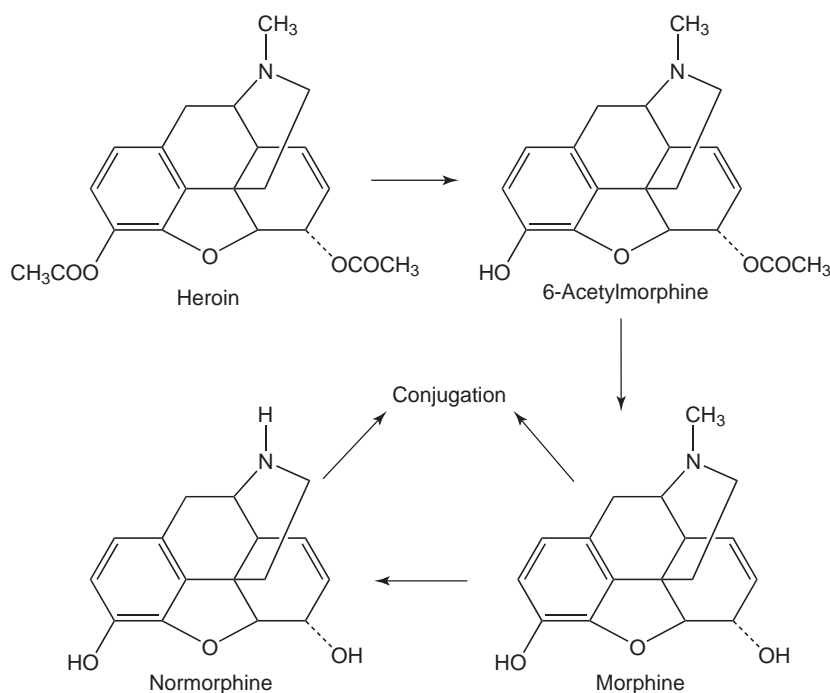
chemical properties of heroin is shown in Table 2. Heroin is rapidly and extensively metabolized in the human body. The metabolic pathway of heroin is illustrated in Figure 2.

## Forensic Identification

Opiate materials submitted to forensic chemistry laboratories for identification are subjected to an analytical scheme that includes a combination of initial screening tests leading to presumptive positives and confirmatory testing with more specific methods. Illicit heroin reacts with most alkaloidal precipitation reagents including the Liebermann's, Mandelin's, and Marquis color tests. In the presence of heroin, these reagents produce black, blue-gray, and violet colors, respectively. Heroin reacts with platinum chloride and mercury(II) chloride reagents forming small golden yellow clustered needles and rosettes, respectively, which can be viewed microscopically. The color or microcrystalline formation produced by the reaction of heroin with these reagents is not necessarily specific to heroin since other opium alkaloids including morphine and codeine produce similar results. Furthermore, the formation of color may be affected by the reaction of strong acids present in the reagents with impurities, such as sugars, found in illicit heroin. The microcrystalline precipitation

**Table 2** Physical and chemical properties of heroin and heroin hydrochloride

Molecular formula	
Heroin	$C_{21}H_{23}NO_5$
Heroin hydrochloride monohydrate	$C_{21}H_{24}NO_5Cl \cdot H_2O$
Relative molecular mass ( $M_r$ )	
Heroin	369.4
Heroin hydrochloride monohydrate	423.9
Melting point ( $^{\circ}C$ )	
Heroin	173
Heroin hydrochloride monohydrate	243–244
Solubility	
Heroin	Chloroform, alcohol, ether
Heroin hydrochloride monohydrate	Chloroform, alcohol, water
Dissociation constant ( $pK_a$ )	7.6 at $23^{\circ}C$
Partition coefficient, $\log P$ (ether/water, pH 7.0)	0.2



**Figure 2** Metabolic pathway of heroin. (Reproduced with permission from Goldberger BA, Darwin WD, Grant TM, *et al.* (1993) Measurement of heroin and its metabolites by isotope-dilution electron-impact mass spectrometry. *Clinical Chemistry* 39: 670–675.)

procedures can be affected by impurities in heroin, producing irregular and/or imperfect crystals.

Opiates can be identified by thin-layer chromatography (TLC) with many different combinations of mobile-phase solvent systems and detection reagents. A solvent system prepared with ethyl acetate–methanol–ammonium hydroxide (85:10:5) is commonly used to resolve heroin, codeine, acetylcodeine, morphine, and acetylmorphine. After development, opiates can be visualized by spraying or dipping the chromatogram in Dragendorff, iodoplatinate, or Marquis reagent.

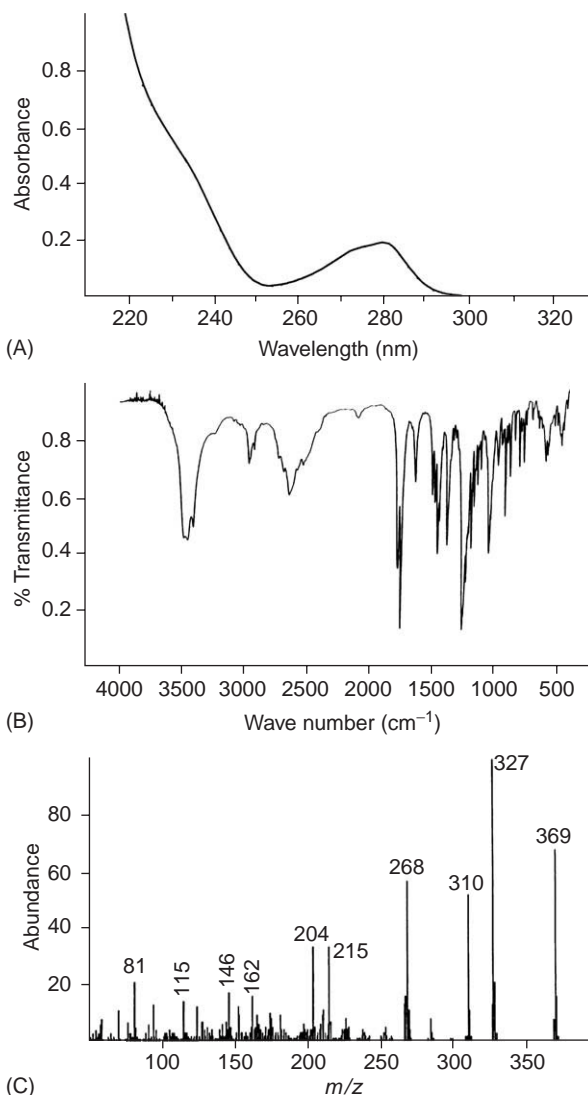
After screening tests have been completed, instrumental techniques such as ultraviolet (UV)–visible spectrophotometry, liquid chromatography (LC), and gas chromatography (GC) are used to confirm the presence of heroin. To analyze an unknown opiate by instrumental analysis, the sample is dissolved and extracted with a suitable solvent. The UV spectrum of pure heroin prepared in dilute acid is shown in Figure 3A. In dilute aqueous acid, the absorbance maximum is 279 nm.

Either LC with UV detection or GC with flame ionization detection can be used to identify and quantify heroin and related compounds. These techniques have been used extensively by many laboratories to assay heroin samples. A study of the chemical composition of illicit heroin samples also can reveal the geographical origin of the opiate material.

Infrared (IR) spectrophotometry and gas chromatography–mass spectrometry (GC–MS) are commonly utilized to confirm the presence of heroin. Both IR and MS of heroin are unique and are easily distinguishable from other related compounds, thus providing a high degree of confidence in the identification process. The Fourier transform (FT) IR and MS spectra of heroin are shown in Figures 3B and 3C. The major IR absorption peaks for heroin hydrochloride, prepared in potassium bromide, are 1753, 1732, 1365, and 1230  $\text{cm}^{-1}$ . The primary spectral features of heroin by electron-impact MS are ions at  $m/z$  369, 327, 310, and 268.

## Analysis of Biological Specimens

The analysis of heroin in biological specimens is difficult because heroin is rapidly metabolized *in vivo* and is subject to chemical and metabolic/enzymatic hydrolysis *in vitro*. In addition, the effective therapeutic dose of heroin is small owing to its high potency, and thus, biological fluid and tissue heroin concentrations are exceedingly low. Therefore, in order to quantify heroin accurately in tissues, adequate



**Figure 3** (A) UV spectrum of heroin in  $0.1 \text{ mol l}^{-1}$  sulfuric acid. (B) Fourier transform IR spectrum of heroin hydrochloride in potassium bromide. (C) Electron-impact mass spectrum of heroin.

precautions during specimen handling and extract preparation must be taken in order to stabilize heroin and 6-acetylmorphine. Heroin can be stabilized by rapid freezing of specimens following collection and by preventing *in vitro* hydrolysis by adding esterase inhibitors such as sodium fluoride.

Heroin or its metabolites have been identified in various biological matrices including: plasma, urine, and saliva collected from human subjects administered heroin; hair and sweat collected from heroin users; urine, vitreous humor, and cerebrospinal fluid collected from cadavers whose death was due to narcotic intoxication; and meconium collected from infants with prenatal exposure to heroin. Frequently, only morphine is assayed because of the instability of heroin and 6-acetylmorphine in blood, plasma, and

urine. Since most specimens do not contain intact heroin, biological specimens testing positive for morphine are usually presumed to have arisen as a result of heroin use. However, since codeine is also metabolized to morphine, it may be necessary to test for a specific metabolite of heroin, 6-acetylmorphine, to establish that heroin was used. In addition, studies have indicated that acetylcodeine, detected in urine and hair, is a potential marker for illicit heroin use.

## Extraction

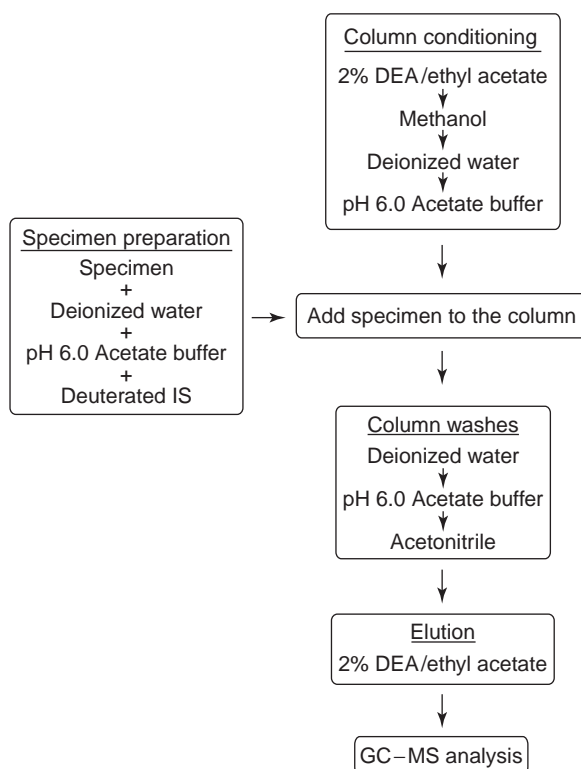
Heroin and its metabolites can be isolated from human and animal biological fluids and tissues by various procedures including liquid–liquid and solid-phase extraction techniques. In order to prevent overestimation of 6-acetylmorphine resulting from the hydrolysis of heroin, heroin and 6-acetylmorphine standard curves can be prepared and extracted separately. Final results must be corrected for the percentage conversion of heroin to 6-acetylmorphine.

Historically, heroin was first isolated by a procedure based upon alkaline extraction with 10% butanol in chloroform. 6-Acetylmorphine and morphine were separated from heroin by paper chromatography or countercurrent distribution. Countercurrent separation was accomplished with an eight-transfer system with ethylene dichloride and phosphate buffer.

The first practical method developed for the simultaneous determination of heroin and its metabolites utilized a complex multistep liquid–liquid extraction scheme. Under mild alkaline conditions, analytes from plasma are isolated with a mixture of toluene–butanol. The analytes are back-extracted into dilute sulfuric acid and the acid phase is washed with *n*-hexane. After adjusting the pH, the aqueous phase is extracted with toluene–butanol. Following separation and evaporation of the organic phase, the extract is reconstituted in methanol and analyzed by LC.

Heroin and metabolites have been isolated from the hair of heroin addicts with a single-step liquid–liquid extraction procedure. Under mild alkaline conditions, a mixture of toluene–heptane–isoamyl alcohol is added to the specimen and, after mixing, the organic phase is removed, evaporated, and derivatized with *N*-methyl-bis(trifluoroacetamide). The derivatized extract is subjected to GC–MS analysis.

Solid-phase extraction using co-polymeric bonded phases with hydrophobic and cation exchanged functionalities has also been used for the extraction of heroin and metabolites from biological fluids and tissues. Solid-phase extraction can yield clean extracts with efficient recovery of analytes. In



**Figure 4** Simplified solid-phase extraction scheme for heroin and metabolites. IS, internal standard; DEA, diethylamine; GC–MS, gas chromatography–mass spectrometry.

addition, since drug isolation and subsequent extraction are rapid, hydrolysis of heroin to 6-acetylmorphine can be reduced or eliminated. A simplified solid-phase extraction scheme is shown in Figure 4.

## Detection Techniques

### Immunoassay

Clinically, opiates are detected in biological specimens by a variety of commercial screening techniques including enzyme multiplied immunoassay, radioimmunoassay, and fluorescence polarization immunoassay. Opiate immunoassays are usually designed to detect morphine and morphine-3-glucuronide (conjugated morphine) as the target analytes and most are formulated to provide qualitative or semiquantitative results. Most opiate assays generally exhibit significant cross-reactivity with other opioid compounds such as codeine, hydromorphone, hydrocodone, and dihydrocodone; consequently, routine and forensic confirmatory analysis of presumptive positive specimens requires identification by a more specific technique such as GC–MS.



### Spectrophotometry, Paper Chromatography, and Thin-layer Chromatography

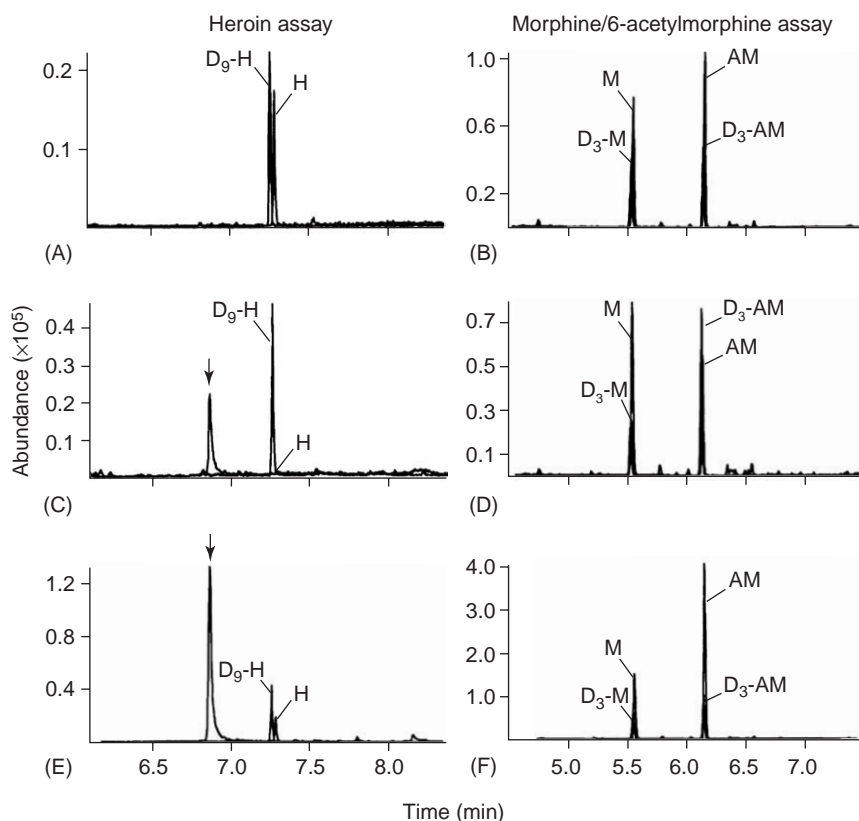
Early analytical methods developed to assay heroin and its metabolites in biological fluids and tissues utilized indicator-dye, paper chromatography (PC), and TLC techniques. Indicator-dye techniques, based upon the reaction of organic bases with colored organic acids such as methyl orange to form complexes soluble in organic solvents, were used to quantify heroin and metabolites. The colored complexes were subsequently measured by spectrophotometry.

Early studies of the disposition and metabolism of heroin in humans and animals were performed with PC and TLC. Heroin, 6-acetylmorphine, and morphine were separated by normal-phase and reversed-phase chromatographic systems. The analytes were visualized by spraying with iodoplatinate, Folin-Ciocalteu reagent, or Dragendorff's reagent.

### Liquid Chromatography

LC with either normal- or reversed-phase systems can be used to measure heroin and metabolites. In order to produce accurate measurements of heroin, liquid chromatographic systems must employ aqueous mobile phases of low hydrolytic activity (e.g., neutral pH and low alcohol content). Since reversed-phase chromatographic analysis of weak bases requires suppression of analyte ionization by increasing the pH of the mobile-phase, normal-phase chromatography is used for the analysis of heroin owing to heroin's instability at elevated pH.

For analysis of heroin metabolites, reversed-phase chromatography coupled with electrochemical detection is frequently used. Other methods used to improve sensitivity include oxidation to fluorescent pseudomorphine-like dimers using potassium hexacyanoferrate(III) and derivatization using a fluorescent chromophore such as dansyl chloride. Detection by MS or tandem MS after separation by LC is



**Figure 5** SIM recordings of standards and postmortem blood and urine specimens. Panels A and B illustrate 15 ng ml<sup>-1</sup> heroin standard and 50 ng ml<sup>-1</sup> 6-acetylmorphine and morphine standards, respectively. Panels C and E illustrate a postmortem blood specimen. Panels D and F illustrate a postmortem urine specimen diluted 1:10 with distilled water. Arrows in panels C and E denote presence of underivatized 6-acetylmorphine. Chromatograms were constructed with the following ions: heroin, *m/z* 327; d<sub>9</sub>-heroin, *m/z* 334; morphine, *m/z* 364; d<sub>3</sub>-morphine, *m/z* 367; 6-acetylmorphine, *m/z* 364; and d<sub>3</sub>-6-acetylmorphine, *m/z* 367. (Reproduced with permission from Goldberger BA, Cone EJ, Grant TM, *et al.* (1994) Disposition of heroin and its metabolites in heroin-related deaths. *Journal of Analytical Toxicology* 18: 22–28.)

becoming increasingly popular. Various ionization techniques like electrospray and atmospheric pressure ionspray are used for the MS detection of heroin, 6-acetylmorphine, morphine, morphine 3-glucuronide, and morphine-6-glucuronide.

### Gas Chromatography

Heroin is generally present in biological fluids and tissues at very low concentrations, whereas the metabolites 6-acetylmorphine and morphine may be present in much higher concentrations. Consequently, heroin metabolites are frequently measured by GC with flame ionization, nitrogen–phosphorus, electron-capture detection, or MS. Depending upon the chromatographic system, derivatization of 6-acetylmorphine and morphine to perfluorinated or silylated derivatives is performed to improve analyte volatility, aid chromatographic performance, and enhance detector response.

Because of its sensitivity and specificity, GC–MS is the most commonly used technique for forensic and clinical testing of heroin and metabolites. Analysis of heroin does not require derivatization; however, derivatization of morphine and 6-acetylmorphine to perfluorinated or silylated derivatives is necessary to improve chromatographic response. Coupled with solid-phase extraction, analysis of heroin by electron-impact mass spectrometry can produce accurate and precise measurement of heroin and metabolites. **Figure 5** panels A and B illustrate characteristic selected ion monitoring (SIM) recordings of extracts of a  $15 \text{ ng ml}^{-1}$  heroin standard and a  $50 \text{ ng ml}^{-1}$  6-acetylmorphine and morphine standard, respectively. Both standards had been prepared in a blood matrix. Panels C–F illustrate SIM tracings of postmortem blood and urine extracts. The arrow in panels C and E denotes the presence of underivatized 6-acetylmorphine.

**See also:** **Clinical Analysis:** Sample Handling. **Derivatization of Analytes. Forensic Sciences:** Systematic Drug Identification; Thin-Layer Chromatography. **Fourier Transform Techniques. Gas Chromatography:** Mass Spectrometry; Forensic Applications. **Immunoassays, Applications:** Clinical; Forensic. **Liquid Chromatography:** Normal Phase; Reversed Phase. **Spectrophotometry:** Pharmaceutical Applications.

### Further Reading

- Brenneisen R and Hasler F (2002) GC/MS determination of pyrolysis products from diacetylmorphine and adulterants of street heroin samples. *Journal of Forensic Sciences* 47: 1–4.
- Brenneisen R, Hasler F, and Wursch D (2002) Acetylcodeine as a urinary marker to differentiate the use of street heroin and pharmaceutical heroin. *Journal of Analytical Toxicology* 26: 561–566.
- Cone EJ and Darwin WD (1992) Rapid assay of cocaine, opiates and metabolites by gas chromatography–mass spectrometry. *Journal of Chromatography* 580: 43–61.
- Cone EJ, Welch P, Mitchell JM, and Paul BD (1991) Forensic drug testing for opiates: I. Detection of 6-acetylmorphine in urine as an indicator of recent heroin exposure; drug and assay considerations and detection times. *Journal of Analytical Toxicology* 15: 1–7.
- Cooper DA (1989) Clandestine production processes for cocaine and heroin. In: Klein M, Sapienza F, McClain H, and Kahn I (eds.) *Clandestinely Produced Drugs, Analogues and Precursors*, pp. 95–116. Washington, DC: US Department of Justice, Drug Enforcement Agency.
- Dams R, Benijts T, Lambert WE, *et al.* (2001) Heroin impurity profiling: trends throughout a decade of experimenting. *Forensic Science International* 123: 81–88.
- Gough TA (ed.) (1991) *The Analysis of Drugs of Abuse*. Chichester: Wiley.
- International Narcotics Control Board, [www.incb.org](http://www.incb.org) (accessed March 22, 2004).
- Moody DE and Cheever ML (2001) Evaluation of immunoassays for semiquantitative detection of cocaine and metabolites or heroin and metabolites in extracts of sweat patches. *Journal of Analytical Toxicology* 25: 190–197.
- Single Convention on Narcotic Drugs (1961), as amended by the 1972 Protocol. United Nations.
- Substance Abuse and Mental Health Services Administration, Office of Applied Studies (2003) *Emergency Department Trends from the Drug Abuse Warning Network, Final Estimates 1995–2002*, DAWN series: D-24, DHHS Publication No. (SMA) 03-3780. Rockville, MD: DHHS.
- Umans JG, Chiu TSK, Lipman RA, *et al.* (1982) Determination of heroin and its metabolites by high-performance liquid chromatography. *Journal of Chromatography* 233: 213–225.
- Zuccaro P, Ricciarello R, Pichini S, *et al.* (1997) Simultaneous determination of heroin, 6-monoacetylmorphine, morphine, and its glucuronides by liquid chromatography–atmospheric pressure ionspray-mass spectrometry. *Journal of Analytical Toxicology* 21: 268–277.

# HISTORY OF ANALYTICAL SCIENCE

**D Thorburn Burns**, The Queen's University of Belfast, Belfast, UK

**F Szabadváry**, National Museum for Science and Technology, Budapest, Hungary

© 2005, Elsevier Ltd. All Rights Reserved.

## Antiquity and the Middle Ages

The use of the senses of taste, touch, sight, and smell are 'methods', as old as mankind itself, for examining and ascribing characteristics to various materials. They may thus be considered to be the oldest analytical techniques. In many fields, such as food and drink, they have stood the test of time, some of them are now elegantly called 'organoleptic tests'. However, throughout history there have been important materials that could not be characterized by the senses and other methods were sought.

Among the first materials to be tested by such methods were the precious metals, gold and silver. Testing of gold by cupellation or fire assay was mentioned by many ancient writers including Biblical Old Testament authors. The assay was carried out in essentially the same way as it is today: melting with lead, then elimination of the lead by oxidation and adsorption in the cupel, followed by weighing of the noble metal nugget (Figure 1). Balances were available to the ancients; for example, in Sumerian cities 5000–6000 years ago.

The earliest recorded attempt in England at regulating the standard for gold and silver wares was the order in 1288, passed by King Henry III. In 1300, the method of assay by touchstone was prescribed. The touchstone test was known in early times and described by Pliny, ~AD 60. Similar controls

were put in place in other nations, for example, King Charles Robert of Hungary issued a decree in 1342 that ordered the establishment of gold-testing houses and described the method of testing using the fire assay method of Pliny. Gold and silver were separated by cupellation with lead and then parted by cementation with powdered brick, iron oxide, and salt. Later when mineral acids became available, in the Middle Ages, parting was by boiling with *aqua fortis* (nitric acid). Details were given by Biringuccio, in *Pirotechnica*, 1540, and by Agricola, in *De Re Metallica*, 1566; both are based on the earlier *Pro-bierbuchlien*, 1510.

Little has been recorded on the early examination of solutions. The oldest record comes from Pliny in *Naturalis historiae libre 33* in which mention is made of the detection of the adulteration of copper sulfate by iron sulfate by using a strip of papyrus soaked in an extract of gall-nuts. If the 'vitriol' solution was adulterated with iron, the papyrus turned black.

Methods of wet analysis were slow to develop as lack of acids made it difficult to bring many materials into solution. Mineral acids were not available until the Middle Ages, then first prepared by unknown alchemists. Nitric acid was first mentioned in the thirteenth century as *aqua fortis* (strong water) by which silver could be separated from gold. It was the availability of mineral acids that enabled wet methods of analysis to be developed for a wide range of materials.

## Qualitative Analysis

A number of qualitative analytical reactions were known and used for identifying substances in solution as early as the seventeenth century. It was known to Boyle that solutions of some metals gave precipitates with alkalis, that a solution of silver becomes turbid on addition of common salt, and that a solution of copper turns dark blue on addition of ammonia. Boyle was the first to mention that some plant extracts (violet, cornflower, damask-rose, black cherry, etc.) change color in the presence of acids or alkalis. This observation contributed greatly to the development of the concept of acids and bases. A great deal of knowledge on qualitative analytical reactions gradually accumulated from the examination of mineral and medicinal waters. Boyle was the first to examine the use of hydrogen sulfide for the detection of copper and arsenic. He also made



**Figure 1** Gold assay by cupellation. From Biringuccio: *Pirotechnica* 1540.

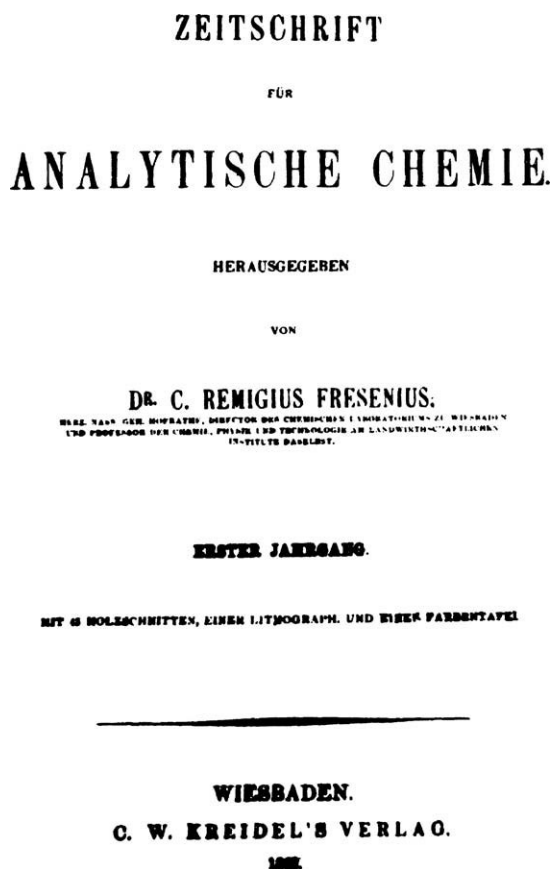
experiments to assess the sensitivity of reactions and estimate limits of detection.

In the eighteenth century, a new analytical device appeared for carrying out qualitative analysis by a fire process that was borrowed from the glass industry by unknown analysts. The device was the blow pipe; using charcoal the reduction processes of metallurgy were scaled down. It was with this device, now more or less forgotten, that a number of chemical elements, for example, tantalum, were discovered in the eighteenth century and at the beginning of the nineteenth century.

Fusion with soda/potash was first used for analytical purposes in the eighteenth century. Wet methods were also developed, new reactions were introduced for detecting newly discovered metals (e.g., precipitation with sulfate for barium, oxidation to the dioxide for manganese), and newly found compounds were used as reagents (e.g., iron(II) cyanide for iron, oxalic acid for calcium, and hydrogen fluoride for silicic acid). Bergman was the first to attempt a comprehensive summary and systemization of analytical procedures in his books on water analysis, ore analysis, and on the examination of precious stones and minerals. He used 20 reagents in water analysis; among them were litmus for testing for acidity, barium chloride for sulfate, lime water for carbon dioxide, mercury nitrate for alkaline compounds, silver nitrate for chloride and sulfur, and soap solution for 'water hardness'. He also noted the sensitivity, preparation, and purification of reagents. Surprisingly, he did not use hydrogen sulfide, the use of which seems to have been forgotten. Most probably it was Winterl who reintroduced hydrogen sulfide for the detection of metals. Bergman's work was followed up and extended by Kirwan.

Rose, in his book *Handbuch der analytischen Chemie* published in 1829, first classified qualitative analytical methods according to elements and not according to the natural substances to be analyzed. This concept was further developed by Fresenius into the now classical method of ion detection by separating metal ions into five groups, described in *Anleitung zur qualitative chemischen Analyse*, 1841. The first synthetic ion-specific organic reagent was the mixture of sulfanilic acid and 1-naphthylamine, the Griess-Ilosvay reagent for nitrite ion, introduced in 1889.

Fresenius founded the first journal devoted entirely to analytical chemistry, *Zeitschrift für analytische Chemie*, in 1862 (Figure 2). This was followed in 1876 by *The Analyst*. Rational explanations for reactions entered analytical chemistry from physical chemistry, an approach pioneered from the 1890s by Ostwald.



**Figure 2** Title page of volume 1 of the world's first journal solely devoted to analytical chemistry. It continues today after mergers with *Analysis* and with *Química Analítica* as *Analytical and Bio-analytical Chemistry*.

## Gravimetric Analysis

Initial data concerning gravimetry were based on the observation that equal amounts of an element produce the same amounts of precipitate with a given reagent, and are to be found in the literature of the eighteenth century. Much of these data refer to the determination of silver in the form of silver chloride. Bergman in *Docimasia humida*, 1780, gave a table of how many parts of product are obtained from 100 parts of metals when precipitated by various reagents. Some of his results were remarkably good. The results of gravimetric analysis led to the laws of combining weight ratios and later were used to determine atomic weights.

By the late 1800s, the technique of gravimetric analysis was more or less developed and has not changed very much since then. Precipitates were mainly ignited to constant weight; porcelain crucibles were used after 1800, platinum crucibles became readily available commercially from about 1820.



Previously results had to be corrected for the weight of the ash of the filter paper. 'Ash free' filter paper was first prepared by Austen in 1878, the same year as Gooch devised the version of the filter crucible that still bears his name. Sintered-glass crucibles came into use in the 1920s. The introduction of electric drying ovens made the controlled drying at lower temperatures of precipitates feasible. Measurement of the loss of weight of a precipitate with controlled rise in temperature, thermogravimetric analysis, owed much from 1940 onwards to Duval. Classical gravimetric analysis had a renaissance in the 1950s from the development of reagents to produce precipitant species at controlled rates in 'precipitation from homogeneous solution'.

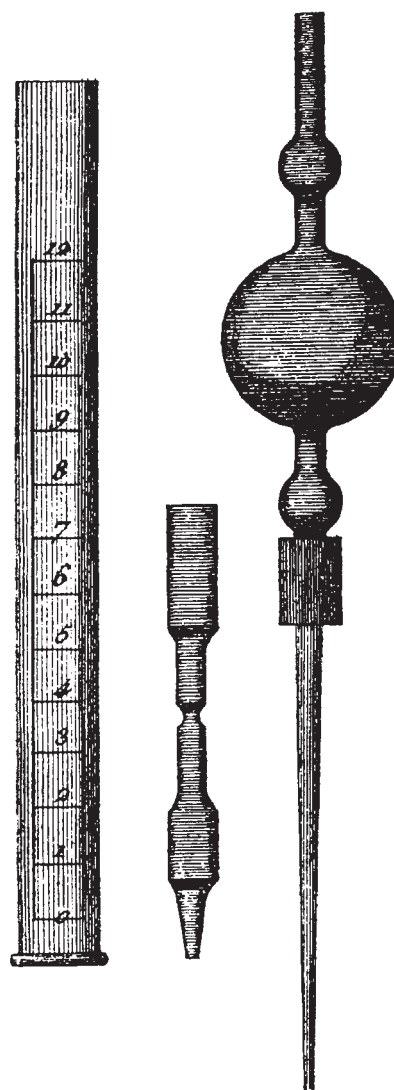
### Volumetric Analysis (Titrimetry)

Volumetric analysis (titrimetry) was developed as a control method in the textile industry in the eighteenth century for determining potash, sulfuric acid, and, later, hypochlorite, all solutions used in textile bleaching. The first methods developed were for practical purposes, to control the 'goodness' (in French, *titre*) of solutions rather than to determine accurate concentrations. The Scottish chemist Home in *Experiments on Bleaching*, 1756, described a method for testing the 'goodness' of potash, in which nitric acid of a given dilution (1–6) was added by teaspoonful to 1 drachm of potash until the effervescence ceased. If less than 12 teaspoonful of the acid mixture were required the potash was not considered suitable for bleaching. In 1767, Lewis used litmus paper instead of cessation of effervescence to indicate completion of reaction. He determined the amount of titrant by weighing. He determined the acid titer using pure potassium carbonate, thus making absolute determinations.

Guyton de Morveau was the first to carry out a precipitation titration, 1782. Hydrochloric acid was titrated with lead nitrate solution until precipitate formation ceased. The first redox titration was due to Descroizilles, around 1788, who used indigo solution to determine the strength of sodium hypochlorite solution used for textile bleaching. The end point was indicated by the persistence of color of the titrant, indigo. This method was not absolute but it was suitable for testing the 'goodness' of the bleaching solution. Descroizilles also worked on improvements for the titration of potash. In this work he recommended the use of specially designed laboratory glassware, describing what were probably the first forms of burette, pipette, and volumetric flask (Figure 3).

Gay-Lussac improved the design of these devices and was the first to use the names burette, pipette, and normal solution (Figure 4). The latter term did not mean what it means now but only a solution of a given concentration. At the time the concentration of titrants used in various determinations was chosen arbitrarily, so that the volume of titrant consumed by a certain amount of sample indicated directly the quality, expressed in arbitrary degrees. Some of the degrees used then still remain in use, for example, those used for 'water hardness'.

Ure suggested in 1844 that concentrations of titrant solutions should be related to the equivalent weights in order to produce a general 'acidimeter'. As he used British units his suggestion was not adopted



**Figure 3** Descroizille's volumetric flask, burette, called a 'bertholli-mètre', and pipette. (From Descroizille (1795) *Arts et Man.* 1 (an III), pp. 256–276.)



internationally. In 1855, Mohr introduced the concept of normal solutions that contain the equivalent weight of a substance in 1 l, a concept eventually accepted universally.

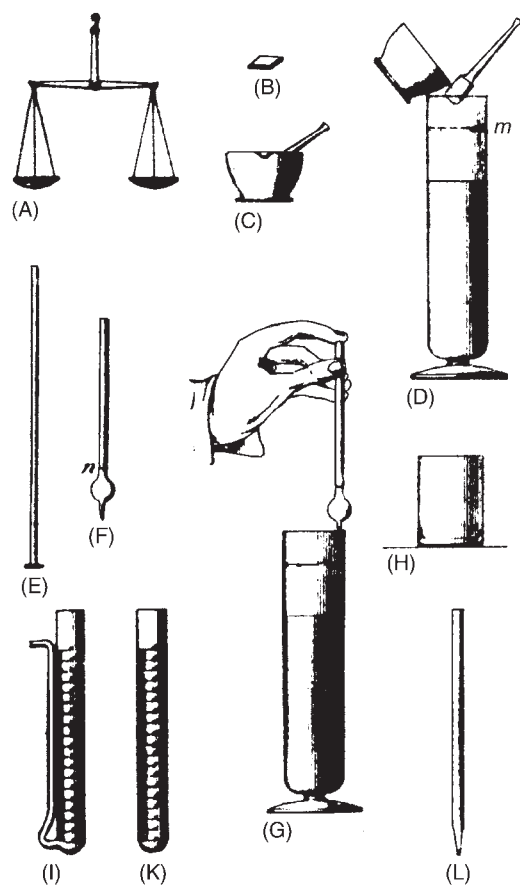
Titrimetric analysis became a precise and accurate method of scientific standard primarily due to the activities of Gay-Lussac and Mohr. Mohr critically re-examined the majority of analytical procedures known at the time. In his immensely successful book *Lehrbuch der chemisch-analytischen Titrimethode*, first published in 1855 and in numerous editions until 1936, he gave

ongoing summaries of developments up-to-date. Some of the important achievements in the various areas of titrimetry are briefly indicated below.

Potassium hydrogencarbonate, which is still used for standardization in acid–base titrations, was introduced by Than in 1860. The various redox titrants, their originators, and the dates are listed in Table 1.

In the field of redox titrations, most of the problems were connected with indicators. Hence, iodometry and permangometry, which are self-indicating methods, became the most widely used. The first redox indicator, diphenylamine, for the dichromate titration of iron(II), was introduced by Knop in 1924.

Titrimetric measurements based on the formation of soluble complexes had little importance for a long time. The first such method, the titration of chloride by mercury(II) using urea as indicator, was introduced by Liebig in 1851. The introduction of chelating agents and metallochromic indicators was the greatest advance in titrimetric analysis in the twentieth century and is linked with the name of Schwarzenbach (1946). Titration in nonaqueous media is important for functional group determinations of organic compounds. The first titration in this area was that by Vorländer, who, in 1903, determined aniline by titration with hydrochloric acid in benzene.



**Figure 4** Gay-Lussac's titrimetric equipment. (From Gay-Lussac (1824) *Annales de Chimie* 26: 162.)

## Organic Elemental Analysis

Organic substances have been distinguished from inorganic ones since the seventeenth century. Chemists in the eighteenth century, starting with Lavoisier, who recognized the role of oxygen in combustion, studied the combustion products of organic substances. At first it was believed that organic compounds were mainly composed of carbon, hydrogen, and oxygen. Lavoisier was the first to carry out experiments to determine the composition of organic substances by measuring the volume of gases produced by combustion and the subsequent absorption of carbon dioxide and measurement of the decrease in volume

**Table 1** Developments in the field of redox titration

Titrant	Originators	Year of first use
Iodine solution	Du Pasquier	1840
Iodometric determination based on the measurement of iodine liberated from iodide	Bunsen	1853
Potassium permanganate standard solution	Margueritte	1846
Potassium dichromate standard solution	Schabus and Penny, independently	1850
Sodium thiosulfate standard solution	Schwarz	1853
Potassium bromate standard solution	Györy	1893
Ascorbic acid standard solution	Erdey	1950

(1789). For methyl alcohol, he found the weight ratio of hydrogen to carbon to be 3.6:1. Considering the experimental difficulties for the first such determination, it was reasonably close to the correct value, 4:1.

Refinements to his ideas and improvements in the apparatus brought improved results. Gay-Lussac and Thenard used vertical combustion tubes. Berzelius used a horizontal tube and determined hydrogen by condensing the bulk of water and absorption of the residual in calcium chloride. Later, Leibig instead of measuring gas volumes used solid absorbents for water and for carbon dioxide (1831). His apparatus (Figure 5) can be considered as the prototype of the combustion devices still in use. Pregl succeeded in miniaturizing the apparatus in the first years of the twentieth century following Kuhlmann's production for him of a sufficiently sensitive microbalance. Pregl received the Nobel prize for chemistry in 1923. In the nineteenth century, one way of determining nitrogen was by gasometry after combustion. This is still done in essentially the same way as was introduced by Dumas in 1831, although on a microscale. The basis of the chemical determination of nitrogen is by its transformation into ammonia, a method first suggested by Dumas but whose practical realization was due to Varrentrap and Will (1841). After boiling with barium hydroxide solution the ammonia formed was precipitated with platinum chloride and weighed. This method was soon replaced by Kjeldahl in 1883, by digestion in sulfuric acid in the presence of a catalyst, making alkaline and distilling out the ammonia, absorbing in dilute sulfuric acid, and titration of the excess acid. The method has been improved and automated systems are in current routine use.

Elemental organic analysis together with knowledge of the reactions used was, till the advent of structure-sensitive spectroscopic techniques, the normal method of identification and determination of the products of organic synthesis.

However, elemental analysis alone was not sufficient to allow the identification and control of

natural and agricultural products, for which characteristic functional groups need to be determined. The search for suitable reactions and the systematization of these reactions was a long and complex task. The first book in this complex and heterogeneous area was that by Mullikan, *Identification of Pure Organic Compounds*, 1904.

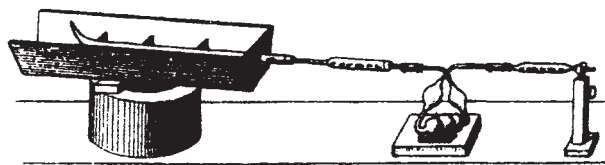
During the last few decades organic analysis has undergone fundamental changes. Chemical methods have reduced importance and automated physical methods are now regarded as essential to confirm structures.

## Optical Methods

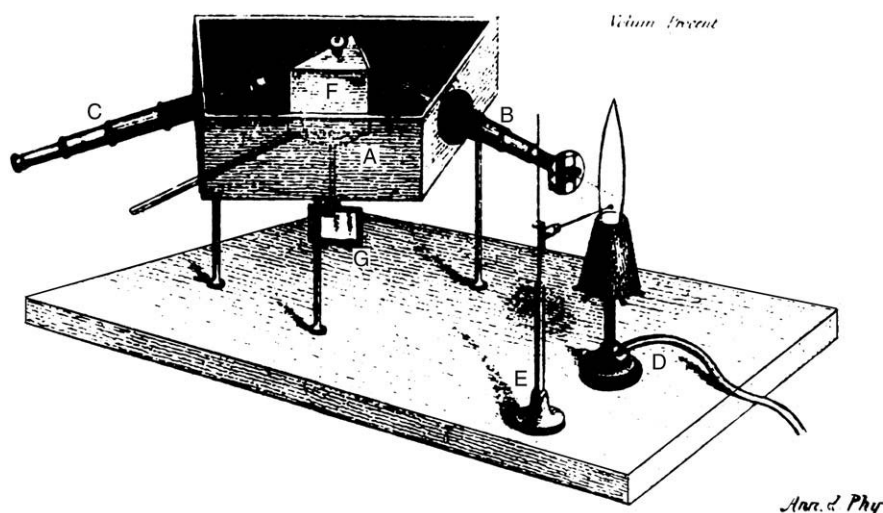
Boyle was the first to use flame coloration for analytical purposes, the detection of adulteration of silver by copper. Maggraff was a pioneer both in using optical devices and the characteristic optical behavior of materials for analytical purposes. He was probably the first to use a microscope for identifying substances when, in 1747, he showed that cane sugar was identical to beet sugar. He also made early use of the characteristic colors imparted to flames and distinguished potash from soda.

The development of emission spectroscopy was a long process initiated by Descartes in 1637 with his illustration that white light can be decomposed into its components with a prism. Several scientists later observed dark lines in the solar spectrum (e.g., Wollaston in 1802). Their constancy was noted and they were mapped in detail by Fraunhofer in 1814. He also noted the appearance of colored, changing lines from other sources. Talbot in 1826 correlated the appearance of some lines with the presence of certain elements. Similar observations were described later by other scientists without arousing much interest at the time. In this context the work of Alter is significant, he produced the first spectral atlas, a tabulation relating spectral lines to chemical elements, in 1854.

Emission spectroscopy achieved complete scientific recognition owing to the work of Bunsen and Kirchhoff, who, in their famous lecture in 1859, described their construction of the spectroscope and its application to the identification of elements (Figure 6). They built this by quite simply using two telescopes, one without the eyepiece lenses, which were replaced by a slit, a prism, and a mirror. Various flames, in which materials were suspended, were used for excitation. In later papers Kirchhoff explained the reasons for the appearance of dark and bright lines in the solar spectrum, and Bunsen proved the power of the method by reporting the discovery of two new elements, cesium and rubidium. In the following



**Figure 5** Leibig's combustion apparatus for organic elemental analysis. (From Liebig (1837) *Anleitung zur Analyse organischer Körper*.) The small triangular glass bubbler contained potassium hydroxide solution to absorb carbon dioxide. The horizontal tubes on each side of the bubbler contained anhydrous calcium chloride to absorb water.



**Figure 6** Kirchhoff and Bunsen's spectroscope: (A) a box, blackened interior; (B) and (C) small telescopes, (B) contains a metal sheet with slit, replacing the eye-piece; (D) light source in front of the slit, Bunsen burner and (E) platinum wire for holding material to be examined; (F) a hollow glass prism filled with carbon disulfide; (G) turnable mirror. (From Kirchhoff and Bunsen (1860) *Poggendorfs Annalen* 110: 160.)

decades mainly nonanalytical scientists were concerned with improvements to the technique. Astronomers made major contributions since spectroscopy was as important to their field as in chemistry. The most important and time-consuming task was to identify the wavelengths of the characteristic lines of all the elements.

Spectroscopy was for a long time only used as an efficient tool for qualitative analysis. It was only in the early twentieth century that it was used in quantitative analysis. Pioneering work was done in this field by Hartley, followed by Gerlach and Schweitzer, who developed the method of homologous line pairs based on the comparison of the line-blackness on a photographic plate.

Starting with the rather unreliable comparison by human eye and then by photography, emission spectrum evaluation made spectacular improvements as a result of electronics and instrumentation, via photoelectric line densitometers introduced by Lundegårdh in 1929, to the commercial manufacture of today's direct reading spectrometers.

The history of flame spectroscopy is difficult to separate from the other types of emission spectroscopy, but it does have some unique features. The analytical use of flame spectra has a long history since the time of Bunsen and Kirchhoff. De Gramont made a major contribution in 1923 by introducing the very hot oxygen-acetylene flame. Lundegårdh constructed the first flame photometer in 1928 with an atomizer (now called a nebulizer), a pressure control unit, and a photocell for detection. Nebulization was used as early as 1879, by Gouy when

investigating emission processes. The practical use of atomic absorption spectroscopy dates from Walsh's classic paper of 1955.

The first paper on X-ray emission spectroscopy with 'quantitative analysis' in the title was Heading's paper of 1922. In 1923, Coster and Hevesy discovered the element hafnium by this technique. However, owing to the experimental difficulties it was not widely used until improvements in radiation detection from radiochemical studies became available. With the development of scintillation counters it became a relatively simple and effective analytical technique, now known as X-ray fluorescence spectrometric analysis.

Electron microprobe analysis is based on excitation of an electron beam concentrated on a very small area (a few square micrometers) of a sample and examination of the X-rays produced. The method was pioneered by Castaing in 1949.

The history of light absorption methods is long and complicated. Without defining any law it was considered obvious that for colored compounds the intensity or depth of color was proportional to the concentration, and quantitative analyses were made by color comparison. The first such analysis was for iron in Tunbridge water, using tincture of galls, by Boyle in 1684. The method was rediscovered 150 years later by Lampadius for iron in 1838, and by Heine for bromide in 1845 using color comparison against standards in simple glass cylinders. Müller constructed the first colorimeter based on the use of a complementary colored glass filter at the base of a cylinder containing the solution, a smaller empty

cylinder was lowered into the solution and the depth to yield approximately white light was noted. Bouguer in 1729 and Lambert in a more detailed form in 1760 established the law describing how the intensity of light was reduced on passing through sheets of glass. In 1852, Beer extended the observations to colored solutions and, he also introduced the concept of the absorption coefficient. Vierordt was the first in 1870 to make a 'photometric' determination using the first instrument based on light intensity. His spectrometer slit had two halves: the light source was viewed through solution covering the lower part and empty cell the upper part. The upper slit was narrowed until the light intensity matched. From the relative slit widths the fraction absorbed by the sample was known. A simplified instrument with filters in place of the spectroscopy was built by Krüss. Measurements were also made by matching light in intensity by varying the depth through which the sample is viewed relative to the blank using Duboscq colorimeters. Berg introduced the first photoelectric colorimeter in 1911 although they did not come into common use till the mid-1920s.

Measurements in the ultraviolet range posed no particular difficulties but instruments only became widely available in the mid-1940s based on the design of Cary and Beckman in spite of the fact that the principles and practice of the method were fully developed at the start of the century.

Infrared radiation was discovered by Herschel in 1800. The bolometer detector (sensitive resistance thermometer) was invented by Langley and was used by him for studying the infrared spectrum of the sun in 1881. Coblentz in *Investigations of Infrared Spectra*, published in 1905, described the infrared spectrum of several hundred compounds, measured point by point using a custom-made instrument containing a rock-salt prism, a mirror spectrometer, and a radiometer detector. Each spectrum took 3–4 h to obtain. Infrared spectroscopy only began to develop in more routine use in the 1930s initiated by needs of the petroleum industry.

The polarization of light by certain crystals has been known for about a long time, as has the ability of some to rotate the plane of polarization of light. Biot in 1813 observed that some liquids and solutions behaved similarly. In 1828, Nichol constructed the prism that bears his name by cementing together with Canada balsam two halves of a split calcite spar crystal. It is probable that Ventzke was the first to use such prisms in 1842 to construct a polarimeter, used to determine the concentration of sugar solutions by measuring the angle of rotation of the plane of polarization. Polarimeters rapidly improved in design and were widely used in many countries where sugar

was subject to special taxation by revenue officers. The correlation of optical activity with the presence of an asymmetric carbon atom in dissymmetrical organic compounds was first made by Le Bel and van't Hoff in 1874. From the mid-1950s, Djerassi and his students studied the variation of rotation with wavelength in the vicinity of absorption bands, optical rotatory dispersion, for structural analyses.

## Electroanalytical Methods

Although chemical effects of an electric current were observed as early as 1800, immediately after the Volta column battery was invented, analytical applications of the phenomena took a long time to develop. Wolcott Gibbs weighed the amounts of metal deposited and, in 1864, he was the first to determine copper by electrogravimetry. Classen made many fundamental contributions to the study and development of the method and in 1882 published the first book on electrogravimetry, *Quantitative Analysis auf elektolytischen Wege*.

Coulometric titration is a direct application of Faraday's laws of electrolysis, it was first applied in this mode by Szebellédy and Somogyi in 1939.

The first potentiometric titration was carried out by Behrend in 1893 at Ostwald's Institute in Leipzig. He titrated mercury(I) nitrate with potassium chloride or with potassium bromide standard solutions using mercury electrode and a mercury/mercury(I) reference electrode. Behrend used a capillary electrometer to measure the electromotive force. The hydrogen electrode was also discovered in 1893 by Le Blanc. He did not, however, use it for analytical purposes but only for electrochemical investigations. Böttger, another student of Ostwald, was the first in 1897 to use a hydrogen electrode to follow an acid–base titration. Potentiometry was first applied to redox titrations by Crotonogino in 1900. He titrated iodide with potassium permanganate using a platinum indicator electrode. For many years potentiometry remained of theoretical interest only since it was difficult to carry out in practice. The first monograph on the topic, in 1923, was *Die elektrometrische Massanalyse* by Müller. It only became an accepted analytical technique after the appearance of commercial electronic instruments in the 1930s. In spite of this, several potentiometric procedures and end-point detection methods are much older than is commonly believed, their principles established early in the twentieth century.

Conductimetric titration techniques were developed as a result of studies by physicists, pioneered by Kohlrausch, who developed both the method and the



necessary instrumentation. Analytical chemists had only the task of finding suitable analytical applications. It is characteristic of analytical chemistry that it has frequently applied techniques originally developed in other areas of science for its own purposes. For conductivity this step was taken by Küster and Grütters in 1903 for the determination of acids and bases.

Oscillometric or high-frequency titration was invented simultaneously and independently by Foreman and Crisp and also by Jensen and Parrak in 1946. Inorganic ion-selective heterogeneous membrane electrodes were introduced for electrometric titrations by Pungor in 1961 and those with neutral organic complexing or ion-exchange compounds by Simon in 1966.

The first voltammetric method was polarography. Electrolysis at a dropping mercury electrode was first described by Heyrovsky in 1922, and the first polarograph constructed by Heyrovsky and Shikita in 1925. Heyrovsky was awarded the Nobel Prize in 1959. Further developments of the method are linked with the names of Matheson and Nichols (oscillopolarography), Heyrovsky (derivative polarography), and Barker (square wave and pulse polarography).

## Radiochemical Analysis

Radioactivity was a discovery that was soon brought into the arsenal of analytical chemistry. It is interesting to note that, as in some other fields, radiation from radioactive species was examined for analytical purposes in a variety of different ways at a time when they had little practical importance, because the radiation detectors were primitive and the field of potential applications limited owing to the small number of natural radionuclides. After the production of artificial radionuclides had started the importance of the earlier radiochemical methods increased rapidly and this field became one of the most powerful in trace analysis.

The analytical application of radionuclides, along with other applications in radioanalysis, results directly from Hevesy and Paneth's invention of radio-tracer and radio-indication techniques in 1912. They pointed out that the addition to a solution of an element of its radioactive isotope makes possible the identification and determination of the element. In their first application they labeled a solution of lead with radium-D, a natural radioactive lead isotope and determined the solubility of sparingly soluble lead salts.

Isotope dilution analysis is probably the most important radiochemical method. A radioactive isotope, whose specific activity is known, is added to a

sample. After a separation, which need not be quantitative, a pure sample is taken and the activity measured. The ratio of this activity to the initial activity (corrected for any decay), together with the weight of the sample, will give the amount of inactive element present in the initial sample. This concept was due to Hevesy. In the first application Hevesy and Hobbie determined the lead content of rocks, the separation was electrolytic, deposition of lead peroxide. The procedure has been simplified by the introduction of substoichiometric analysis by Starý and Ruzicka in 1961.

Radioactive reagents were pioneered by Ehrenberg in 1925 using reagents that contained a radionuclide, calculating the amount of sample from the precipitate or the filtrate after precipitation. The method was not used much until the ready availability of artificial radioisotopes. Radiometric titrations achieved greater popularity. In these methods either the titrand or the titrant may be labeled. The product of the reaction is separated from the solution, in simple cases as a precipitate or by extraction. Evaluation is carried out by a graphical method. The first example, a precipitation titration, was by Langer in 1941.

Radiochemical versions of many other techniques were soon developed, for example, paper chromatography, ion-exchange chromatography.

Activation analysis is the other field of radiochemical analysis that has become of major importance, particularly neutron activation analysis. In this method nuclear transformations are carried out by irradiation with neutrons. The nature and the intensity of the radiation emitted by the radionuclides formed are characteristic, respectively, of the nature and concentrations of the atoms irradiated. Activation analysis is one of the most sensitive methods, an important tool for the analysis of high-purity materials, and lends itself to automation. The technique was devised by Hevesy, who with Levi in 1936 determined dysprosium in yttrium by measuring the radiation of dysprosium after irradiation with neutrons from a Po-Be neutron source. At the time the nature of the radiation was characterized by half-life, and the only available neutron sources were Po-Be and Ra-Be, which were of low efficiency. Hevesy's paper was not followed up for many years. The importance of activation analysis increased dramatically after the emergence of accelerators and reactors in which almost all elements could be activated. Hevesy received the 1943 Nobel prize in chemistry for 'work on the use of isotopes as tracers in the study of chemical processes'.

For a long time only electroscopes were used for measuring radiation from radioactive sources. The first



counter was constructed in 1908 as a result of experiments of Rutherford and Geiger. The counter tube was improved by Geiger and Müller in 1928, and it was the most widely used detector in radiochemistry for decades. Scintillation detectors, which, with appropriate amplification and attenuation, enabled beta and gamma radiations to be recorded as functions of their energies, were developed soon after World War II. One of the best materials used for this purpose was a sodium iodide single crystal activated with thallium, first used by Hofstadter in 1948. With the rapid developments of electronics and transistor techniques, this detector became the basis of modern multichannel analyzers for beta and gamma spectrometry.

## Chromatographic Techniques

Chromatography was one of the most rapidly developing areas in the second half of the twentieth century that opened up new vistas, especially for the examination of complex organic materials. Although the method is in many ways new, its recent forms, its origin, and inventor cannot be identified with certainty as some experiments go back a long way. Methods that we now recognize as 'chromatographic' can be found in the literature of the nineteenth century. For example, Schönbein reported in 1861 when developing tests for ozone that the solvent precedes the dissolved substances and that different substances are drawn up the paper to varying degrees. His student Goppelsröder devoted his life to the elaboration of this so-called capillary analysis technique. He published many papers on the subject and summarized his studies in a monograph *Anregung zum Studium der auf Capillaritäts und Absorptionserscheinungen beruhenden Capillaranalyse* in 1906. Tswett came into the history of chromatography from botany. He used a column filled with inulin for separating plant pigments. In 1903, when passing leaf extracts through a column he found colored components situated in the form of discs at different parts of the column. He called the technique chromatography from the Greek word *chroma* for color. Although adsorption column chromatography proved to be effective for the separation of natural organic materials it only became of practical importance after it was used by Kuhn, Winterstein, and Lederer for the separation of carotenoids. This marked the beginning of the era of modern chromatography.

The amalgamation of the principle of countercurrent extraction and experience with adsorption chromatography led to the development of partition chromatography, an invention linked with the names

of Martin and Synge, who in 1941 used partition between chloroform and water retained in silica gel to separate components of protein hydrolysates. They jointly received the Nobel prize for chemistry in 1952. Paper chromatography, in which the support is paper, was later used by Gordon, Martin, and Synge (1943). For a while paper chromatography made much progress. Gas chromatography invented by Martin and James in 1952 developed more slowly at first due to difficulties with instrumentation. Paper chromatography gradually lost its importance, mainly due to the rapid development of thin-layer chromatography, following the advances made by Stahl that were published in his 1962 monograph.

Ion-exchange methods are also important in modern analytical chemistry. They now take the place of a number of complicated separations which in the past were achieved by series of precipitations. The ion-exchange properties of certain clay minerals were observed in the middle of the nineteenth century. The phenomenon was first utilized in industrial and agricultural processes, for example, for removing alkali from sugar beet liquors. Ion exchange was used for analytical purposes only at the end of World War I, when Folin and Bell (1917) bound ammonia in urine as ammonium in its separation from creatine using an artificial ion exchanger, synthetic aluminum silicate patented by Gans in 1905. Analytical applications of ion exchangers only really developed after ion-exchange resins were discovered. The first examples of such resins were made by Adams and Holmes in 1935. Since then, a wide range of ion-exchange resins have been produced commercially. The successful incorporation of ion exchangers into chromatography has enabled separations to be achieved that were not possible by other techniques, such as those for the rare earths and the radioactive materials in the Manhattan project. Possibly the first report of this method was made by Russel, Svartout, Hume, and Kettle in 1944, first published in the open literature in 1947. Ion-exchange chromatography received a boost and renewed interest in 1975 after Small, Stevens, and Bauman's incorporation of a suppressor column between the separator and a conductivity detector to increase sensitivity by reduction of the background.

Molecular sieves had been in use since the mid-1920s for chromatographic separations of small molecules. The application to water-soluble fragile materials of biological origin was the subject of intense activity after the discovery of cross-linked dextran gels by Porath and Flodin in 1957 in what became known as 'gel filtration' or size exclusion chromatography.

See also: **Activation Analysis:** Neutron Activation. **Atomic Emission Spectrometry:** Principles and Instrumentation. **Bleaches and Sterilants.** **Chiroptical Analysis.** **Chromatography:** Principles. **Conductimetry and Oscillometry.** **Coulometry.** **Fire Assay.** **Food and Nutritional Analysis:** Overview. **Gas Chromatography:** Principles. **Gravimetry.** **Indicators:** Redox. **Infrared Spectroscopy:** Overview. **Ion Exchange:** Overview. **Iso-  
tope Dilution Analysis.** **Lipids:** Fatty Acids. **Liquid Chromatography:** Size-Exclusion. **Radiochemical Methods:** Natural and Artificial Radioactivity; Radionuclide Monitoring; Radiotracers. **Spot Tests.** **Thin-Layer Chromatography:** Overview. **Titrimetry:** Overview; Potentiometric. **X-Ray Fluorescence and Emission:** X-Ray Fluorescence Theory.

## Further Reading

### General, Chemistry

- Child E (1940) *The Tools of the Chemist*. New York: Reinhold.
- Gillespie CC (ed.) (1970–1990) *Dictionary of Scientific Biography*, vols. 1–14. New York: Scribner's Sons.
- Laidler KJ (1993) *The World of Physical Chemistry*. Oxford: Oxford University Press.
- Partington JR (1961–1970) *A History of Chemistry*, Vols. 1–4. London: Macmillan.
- Russel CA (1985) *Recent Developments in the History of Chemistry*. London: Royal Society of Chemistry.

### General, Analytical Chemistry

- Laitinen HA and Ewing GW (1977) *A History of Analytical Chemistry*. Division of Analytical Chemistry of the American Chemical Society.
- Morris PJT (ed.) (2002) *From Classical to Modern Chemistry, The Instrumental Revolution*. London: RSC.
- Ostwald W (1894) *Die Wissenschaftlichen Grundlagen der Analytische Chemie*. Leipzig: Engelman. (English translation McGowan (1895). London: Macmillan.)
- Rosenfeld L (1999) *Four Centuries of Clinical Chemistry*. New York: Gordon and Breach.
- Szabadváry F (1966) *History of Analytical Chemistry*. Oxford: Pergamon Press.
- Szabadváry F and Robinson A (1980) The history of analytical chemistry. In: Svehla G (ed.) *Comprehensive Analytical Chemistry*, vol. X. Amsterdam: Elsevier.

### Antiquity and the Middle Ages

- Hoover HC and Hoover LC (trans. eds.) (1912). *Georgii Agricolae De Re Metallica*. London: The Mining Magazine. (Reprinted Dover, New York, 1950.)
- Mackenzie RC (1984) A history of thermal analysis. *Thermochimica Acta* 73: 249–367.
- Sisco AG and Smith CS (1949) *Bergwerk- und Probierrbüchlein*. New York: American Institute of Mining and Metallurgical Engineers.
- Thorburn Burns D (1981) A touch-stone for gold and silver wares. *Analytical Proceedings* 21: 146–151.

### Gravimetric Analysis

- Beck CM II (1994) Classical analysis: a look at the past, present and future. *Analytical Chemistry* 66: 224A–239A.

### Qualitative Analysis

- Burns DT, Townshend A, and Carter AH (1981) *Inorganic Reaction Chemistry. Reactions of the Elements and Their Compounds*, vols. 1 and 2. Chichester: Ellis Horwood.
- Thorburn Burns D (1979) Edward Jordan MD (1569–1632): early contributions to solution analysis. *Proceedings of the Analytical Division of the Chemical Society* 16: 219–221.

### Volumetric Analysis

- Bishop E (1972) *Indicators*. Oxford: Pergamon Press.
- Rancke-Madsen E (1958) *The development of Titrimetric Analysis till 1806*. Copenhagen: Gad.

### Organic Elemental Analysis

- Thorburn Burns D (1993) Precursors to and evolution of elemental organic tube combustion analysis over the last two hundred years. *Analytical Proceedings* 30: 272–275.
- Thorburn Burns D (1984) Kjeldahl, the man, the method and the Carlsberg Laboratory. *Analytical Proceedings: Proceedings of the Analytical Division of the Royal Society of Chemistry* 21: 210–214.

### Optical Methods

- Boumans PWJM (1980) Atomic absorption spectroscopy past, present, and future. *Spectrochimica Acta* 35B: 637–883.
- Hentschell K (2002) *Mapping the Spectrum: Techniques of Visual Representation in Research and Teaching*. Oxford: Oxford University Press.
- Thorburn Burns D (1986) Robert Boyle's determination of iron in Tunbridge Water: the earliest quantitative colorimetric reaction? *Analytical Proceedings: Proceedings of the Analytical Division of the Royal Society of Chemistry* 23: 75–77.
- Thorburn Burns D (1988) The development of apparatus for spectrochemical 1500–1870; European art, craft and science. In: Roth ME (ed.) *Euroanalysis VI, Reviews on Analytical Chemistry*. Paris: Les editions de physique.

### Electroanalytical Methods

- Stock JT and Orna MV (eds.) (1989) *Electrochemistry past and present*. Washington: American Chemical Society.

### Radiochemical Analysis

- Gibbons D and Lambie DA (1971) Radiochemical methods of analysis. In: Wilson CL and Wilson DW (eds.)

*Comprehensive Analytical Chemistry*, vol. IIC. Amsterdam: Elsevier.

### Separation and Chromatographic Techniques

Ettre LS (2002) *Milestones in the Evolution of Chromatography*. Franklin: ChromSource.

Ettre LS and Zlatkis A (eds.) (1979) *75 years of chromatography – a historical dialogue*. Amsterdam: Elsevier.

Gehrke CW, Wixom RL, and Bayer E (eds.) (2001) *Chromatography a Century of Discovery 1900–2000*. Amsterdam: Elsevier.

Berezkin VG (compiler), Masson MR (translator) (1990) *Chromatographic Adsorption Analysis: Selected Works*. M.S. Tswett. Chicester: Horwood.

Issaq HJ (2002) *A Century of Separation Science*. New York: Dekker.

# HORMONES

## Steroids

**S Görög**, Chemical Works of Gedeon Richter Ltd., Budapest, Hungary

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The analysis of steroid hormones and their semisynthetic analogs falls into a number of different areas, such as

1. Determination of steroid hormones, the intermediates of their biosynthesis and their metabolites, and the semisynthetic analogs and their metabolites in biological samples. This is of great importance in clinical biochemistry, diagnostics, therapeutic drug monitoring, doping control of athletes, and analytical control of meat products.
2. Structure elucidation of steroid hormone analogs (products and intermediates of synthetic drug research, impurities, degradation products, metabolites).
3. Assay of bulk hormone drugs and impurity determination.
4. Assay of hormone formulations, including stability assays.

Highly selective and sensitive methods are required to solve these problems; in the majority of cases the complex application of chromatographic, spectroscopic, and immunoassay methods is necessary.

### Main Types of Steroid Hormones and Their Semisynthetic Analogs

Steroid hormones can be classified into two main groups: sex hormones and corticosteroid hormones. Of the sex hormones, androgens are secreted mainly by the testes (main representative: testosterone), estrogens by the ovary (main representative: estradiol), and the gestogens (progestogens, progestins) by the corpus luteum (main representative: progesterone). Corticosteroid hormones are secreted by the adrenal cortex. One of the main representatives is cortisol (hydrocortisone). The structures of these hormones are given in **Figure 1**.

For drug therapy, native hormones have been replaced to a large extent by synthetic or semisynthetic derivatives that possess higher and more selective therapeutic potential. Some very important synthetic androgen analogs are the esters and the 17-methyl derivative of testosterone. Of the numerous related anabolic steroids, the esters of nandrolone (19-nortestosterone), methandienone (17 $\alpha$ -methyl-17-hydroxy-1,4-androstadien-3-one), stanozolol (17 $\alpha$ -methyl-2'H-5 $\alpha$ -androst-2-eno[3,2-c]pyrazol-17-ol, and trenbolone acetate (17 $\beta$ -hydroxy-4,9,11-estratrien-3-one acetate) are most frequently used.

Among the synthetic estrogens and gestogens, the 17 $\alpha$ -ethinyl-17-hydroxy derivatives merit special mention. The introduction of the ethinyl group greatly increases their hormonal potential, thus enabling their dose to be decreased (e.g., in contraceptive tablets) to the 10  $\mu$ g per tablet and 100  $\mu$ g per tablet level, respectively. Characteristic examples of this type are the estrogenic ethinylestradiol (17 $\alpha$ -ethinyl-1,3,5(10)-estratrien-3,17-diol) and the gestogenic norethisterone (17 $\alpha$ -ethinyl-17-hydroxy-4-estren-3-one), and norgestrel, the synthetic 13-ethyl analog of the latter.

The most important synthetic corticosteroid is prednisolone, which is the  $\Delta^1$  analog of cortisol.

*Comprehensive Analytical Chemistry*, vol. IIC. Amsterdam: Elsevier.

### Separation and Chromatographic Techniques

Ettre LS (2002) *Milestones in the Evolution of Chromatography*. Franklin: ChromSource.

Ettre LS and Zlatkis A (eds.) (1979) *75 years of chromatography – a historical dialogue*. Amsterdam: Elsevier.

Gehrke CW, Wixom RL, and Bayer E (eds.) (2001) *Chromatography a Century of Discovery 1900–2000*. Amsterdam: Elsevier.

Berezkin VG (compiler), Masson MR (translator) (1990) *Chromatographic Adsorption Analysis: Selected Works*. M.S. Tswett. Chicester: Horwood.

Issaq HJ (2002) *A Century of Separation Science*. New York: Dekker.

# HORMONES

## Steroids

**S Görög**, Chemical Works of Gedeon Richter Ltd., Budapest, Hungary

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The analysis of steroid hormones and their semisynthetic analogs falls into a number of different areas, such as

1. Determination of steroid hormones, the intermediates of their biosynthesis and their metabolites, and the semisynthetic analogs and their metabolites in biological samples. This is of great importance in clinical biochemistry, diagnostics, therapeutic drug monitoring, doping control of athletes, and analytical control of meat products.
2. Structure elucidation of steroid hormone analogs (products and intermediates of synthetic drug research, impurities, degradation products, metabolites).
3. Assay of bulk hormone drugs and impurity determination.
4. Assay of hormone formulations, including stability assays.

Highly selective and sensitive methods are required to solve these problems; in the majority of cases the complex application of chromatographic, spectroscopic, and immunoassay methods is necessary.

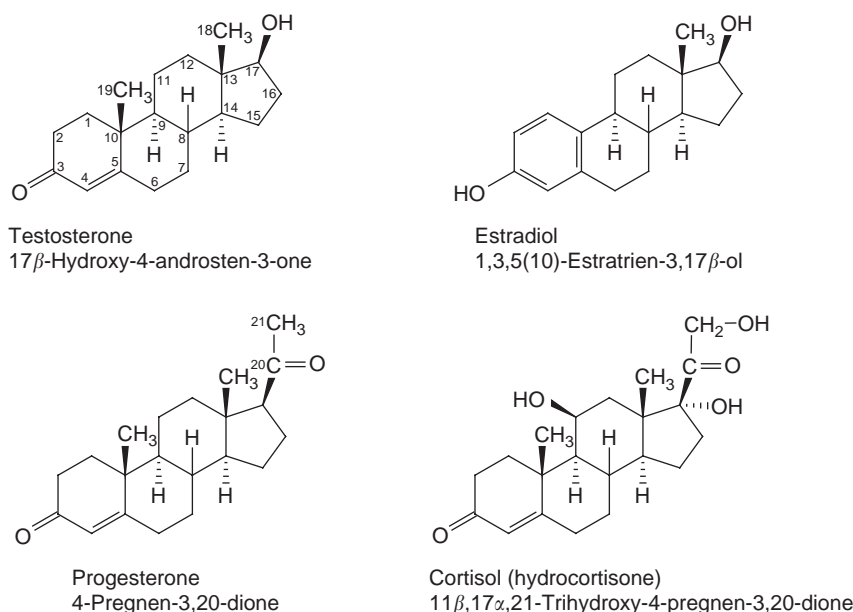
### Main Types of Steroid Hormones and Their Semisynthetic Analogs

Steroid hormones can be classified into two main groups: sex hormones and corticosteroid hormones. Of the sex hormones, androgens are secreted mainly by the testes (main representative: testosterone), estrogens by the ovary (main representative: estradiol), and the gestogens (progestogens, progestins) by the corpus luteum (main representative: progesterone). Corticosteroid hormones are secreted by the adrenal cortex. One of the main representatives is cortisol (hydrocortisone). The structures of these hormones are given in **Figure 1**.

For drug therapy, native hormones have been replaced to a large extent by synthetic or semisynthetic derivatives that possess higher and more selective therapeutic potential. Some very important synthetic androgen analogs are the esters and the 17-methyl derivative of testosterone. Of the numerous related anabolic steroids, the esters of nandrolone (19-nortestosterone), methandienone (17 $\alpha$ -methyl-17-hydroxy-1,4-androstadien-3-one), stanozolol (17 $\alpha$ -methyl-2'H-5 $\alpha$ -androst-2-eno[3,2-c]pyrazol-17-ol, and trenbolone acetate (17 $\beta$ -hydroxy-4,9,11-estratrien-3-one acetate) are most frequently used.

Among the synthetic estrogens and gestogens, the 17 $\alpha$ -ethinyl-17-hydroxy derivatives merit special mention. The introduction of the ethinyl group greatly increases their hormonal potential, thus enabling their dose to be decreased (e.g., in contraceptive tablets) to the 10  $\mu$ g per tablet and 100  $\mu$ g per tablet level, respectively. Characteristic examples of this type are the estrogenic ethinylestradiol (17 $\alpha$ -ethinyl-1,3,5(10)-estratrien-3,17-diol) and the gestogenic norethisterone (17 $\alpha$ -ethinyl-17-hydroxy-4-estren-3-one), and norgestrel, the synthetic 13-ethyl analog of the latter.

The most important synthetic corticosteroid is prednisolone, which is the  $\Delta^1$  analog of cortisol.



**Figure 1** Structures of several steroid hormones.

Highly active derivatives can be derived from prednisolone by introducing halogens, mainly fluorine, at positions 9 and 6, methyl groups at 6 and 16, a hydroxyl group at 16, etc. For instance, dexamethasone and betamethasone are the 16 $\alpha$ - and  $\beta$ -methyl derivatives of 9 $\alpha$ -fluoroprednisolone, respectively, and triamcinolone is the 16 $\alpha$ -hydroxy derivative of 9 $\alpha$ -fluoroprednisolone.

## Methods Used in the Analysis of Steroid Hormones

### Classical Methods

Due to the lack of suitable functional groups in steroid hormones, titrimetry has never played an important role in steroid hormone analysis. Of the several indirect methods, however, one is still widely used: all pharmacopoeias contain methods for the determination of ethinyl steroids on the basis of the exchange of the acetylenic hydrogen for silver ions and the titration of the liberated acid with standard sodium hydroxide.

$\Delta^4$ -3-oxo and  $\Delta^{1,4}$ -3-oxo groups can be reduced at the dropping mercury electrode in the range between  $-1.22$  and  $-1.37$  V in the concentration range of  $10^{-3}$  mol l $^{-1}$ , enabling their polarographic determination in drug formulations. Although with the aid of modern polarographic techniques, especially differential pulse polarography, the sensitivity can be increased 10- to 100-fold, even this sensitivity and

the poor selectivity preclude application of this technique to the analysis of biological materials.

### Spectroscopic Methods

**Ultraviolet-visible spectrophotometry** Due to the presence of either an unsaturated oxo group or a phenolic ring in their molecules, the majority of steroid hormones absorb strongly in the ultraviolet (UV) region (see Table 1). The UV spectra of steroid hormones are of limited importance in elucidating their structure, but enable their moderately selective quantitative determination in simple pharmaceutical formulations.

To enhance the applicability of UV spectrophotometry in pharmaceutical steroid analysis by increasing its selectivity and sensitivity, various methods have been developed based on chemical reactions leading to colored derivatives. Although their importance has decreased considerably, some of these methods are still in use in pharmacopoeias, mainly in the assay of formulations. For example, the  $\Delta^4$ -3-oxo and  $\Delta^{1,4}$ -3-oxo steroids can be determined as isonicotinoyl hydrazones ( $\lambda_{\max}$  380 and 410 nm;  $\epsilon \sim 11\,500$  and  $17\,000$ , respectively). The dihydroxyacetone side chain of corticosteroids at C17 reduces tetrazolium reagents to colored formazans ( $\lambda_{\max}$  485 and 525 nm with triphenyl tetrazolium chloride and Tetrazolium Blue reagents, respectively;  $\epsilon \sim 16\,000$  and  $24\,000$ ), thus creating the basis for a stability-indicating indirect colorimetric assay. Corticosteroids



**Table 1** UV spectral data of some important groups of steroid hormones and their synthetic analogs (solvent: ethanol)

Functional group	$\lambda_{\max}$ (nm)	$\epsilon$ ( $\text{l mol}^{-1} \text{cm}^{-1}$ )	Occurrence	Typical derivatives
$\Delta^4$ -3-Oxo	240	17 000	Androgens, anabolics, gestogens, corticosteroids	Testosterone, nortestosterone esters, progesterone, cortisol
$\Delta^{1,4}$ -3-Oxo	243	16 000	Anabolics, corticosteroids	Methandienone, prednisolone
$\Delta^{4,6}$ -3-Oxo	285	25 000	Gestogens	Megestrol acetate
$\Delta^{4,9,11}$ -3-Oxo	341	28 000	Anabolics	Trenbolone
1,3,5(10)-Triene-3-ol			Estrogens	Estrone, estradiol, ethinylestradiol
Neutral	280	2 000		
Alkaline	300	2 800		
1,3,5(10)-Triene-3-alkoxy	287	1 900	Estrogens	Mestranol
	278	1 750		
1,3,5(10)-Triene-3-acyloxy	268	800	Estrogens	Estradiol dipropionate

$\epsilon$ , molar absorptivity;  $\lambda_{\max}$ , wavelength of maximum absorbance.

are also determined with the phenylhydrazine/sulfuric acid reagent ( $\lambda_{\max}$  410 nm;  $\epsilon \sim 28\,000 \text{ l mol}^{-1} \text{ cm}^{-1}$ ).

Enzymatic methods making use of the redox reactions of steroids and the measurement of NADH at 340 nm or other suitable colorimetric indicators greatly enhance the utility of spectrophotometry and enable their use in bioassays.

**Fluorimetry** Of the steroid hormones, estrogens, having a phenolic ring, possess natural fluorescence (excitation  $\sim 284$  nm, emission in the range 310–327 nm). The intensity of this relatively weak fluorescence is sufficient for the determination of estrogens in contraceptive pills even at the pharmacopoeia level, especially in the liquid chromatography (LC) assay of the pills using fluorescence detector.

Visible fluorescence and much higher intensity can be obtained by using various chemical reactions. These may be characterized by the high concentration of mineral acids in the reagents and the complex mechanism of the reactions. It is mainly estrogens that can be determined in this way, but sensitive methods are available for the determination of other groups of steroid hormones also. The high sensitivity of these methods enables them to be used in the analysis of biological samples, but these methods are being replaced by the even more sensitive and selective chromatographic and immunological methods.

**Infrared (IR) spectroscopy** In the early period of steroid research, IR spectroscopy was the main method for the structural elucidation of steroids. Although the importance of this technique in the structural elucidation of steroids has decreased considerably, it still retains some of its importance.

IR spectroscopy is widely used for the identification of steroids: it has completely replaced the classical color reactions used in earlier editions of the pharmacopoeias. It is also useful for the identification and characterization of polymorphic modifications, whose occurrence is quite common among steroid derivatives.

In the majority of cases the potassium bromide disk technique is used (see Table 2), but many examples can also be found in the literature about the application of the Nujol mull method. The use of solution spectra is usually restricted to the less important quantitative analytical applications.

**Nuclear magnetic resonance (NMR) spectroscopy** Since its introduction into the structural analysis of steroids about half a century ago, NMR spectroscopy has become the most powerful technique in this field. Even the age-old 'bottleneck' of NMR, i.e., the need for relatively large sample quantities, is gradually disappearing: continuing improvements in magnetic field homogeneity and strength as well as in hardware (e.g., low-noise electronic elements, superconducting probes, or probes employing high-speed magic-angle sample spinning) enable the sample quantity to be reduced to the microgram level.

Even one-dimensional (1D)  $^1\text{H}$  NMR spectra are often adequate to deduce some delicate configurational and conformational characteristics of a molecule. Such features are inferred from the chemical shifts and coupling constants of the respective protons as well as from the position of the easily detectable singlets of the C18 and C19 angular methyl groups. The use of the nuclear Overhauser effect (NOE) and various multipulse NMR techniques enable the solving of even more difficult problems.

**Table 2** Some characteristic IR bands of steroids

Group	Character of the band	Average wavenumber ( $\text{cm}^{-1}$ )	Intensity <sup>a</sup>
Bare steroid skeleton	$\nu_{\text{as}}$ ( $\text{CH}_3$ , $\text{CH}_2$ )	2950	s
	$\nu_{\text{sy}}$ ( $\text{CH}_3$ , $\text{CH}_2$ )	2880	s
	$\nu_{\text{sy}}$ ( $\text{CH}_3$ )	1380	m
	$\nu_{\text{as}}$ ( $\text{CH}_2$ )	1460	m
Isolated double bond	$\nu$ ( $\text{C}=\text{C}$ )	1660	w
	$\nu$ ( $=\text{C}-\text{H}$ )	3100–3000	w
	$\gamma$ ( $=\text{C}-\text{H}$ )	880	m
	$\nu$ ( $\text{C}\equiv\text{C}$ )	2200	w
17-Ethynyl group	$\nu$ ( $\text{C}\equiv\text{C}$ )	3300	s
	$\nu$ ( $\text{C}=\text{C}$ )	1610, 1590, 1500	m
Phenolic ring	$\gamma$ ( $\text{C}-\text{H}$ )	880, 830	m
	$\nu$ ( $\text{O}-\text{H}$ )	3600–3400	s
	$\nu$ ( $\text{C}-\text{OH}$ )	1260	s
	$\nu$ ( $\text{O}-\text{H}$ )	3600–3300	s
Alcoholic hydroxyl	$\nu$ ( $\text{C}-\text{OH}$ )	1100	s
	$\nu$ ( $\text{C}-\text{O}-\text{C}$ ) <sub>as, sy</sub>	1230, 1050	s
Carboxylic esters of alcohols	$\nu$ ( $\text{C}=\text{O}$ )	1710	s
3-, 11-, 20-Oxo	$\nu$ ( $\text{C}=\text{O}$ )	1735	s
17-Oxo	$\nu$ ( $\text{C}=\text{O}$ )	1660	s
4-Ene-3-oxo	$\nu$ ( $\text{C}=\text{O}$ )	1600	m
	$\gamma$ ( $=\text{C}-\text{H}$ )	880	m
	$\nu$ ( $\text{C}=\text{O}$ )	1660	s
	$\nu$ ( $\text{C}=\text{C}$ )	1620, 1605	m, w
1,4-Diene-3-oxo	$\gamma$ ( $=\text{C}-\text{H}$ )	900	m

<sup>a</sup>s, strong; m, medium; w, weak; as, asymmetrical valence vibrations; sy, symmetrical valence vibrations.

This success of  $^1\text{H}$  NMR spectroscopy is all the more remarkable since the majority of the protons in steroids belong to skeletal methylene and methine groups, which give a crowded group of only partly resolved signals in the range of 1–2.5 ppm if traditional, low-, or medium-resolution instruments are used. (Using high-resolution, e.g., 900 MHz instruments even skeletal proton signals are distinguishable.) In contrast, the very wide range of chemical shifts in the  $^{13}\text{C}$  NMR spectra of steroids ( $\sim 220$  ppm, compared to the 10 ppm range in  $^1\text{H}$  NMR) enables the signals of all the 18–27 carbon atoms to be resolved and assigned. The assignment is greatly facilitated by the application of, for example, the DEPT pulse sequence technique, while for the assignment of both the  $^1\text{H}$  and  $^{13}\text{C}$  NMR signals  $^1\text{H}-^1\text{H}$  COSY,  $^1\text{H}-^{13}\text{C}$  COSY (HETCOR), and various long-range correlation techniques are successfully used. The application of modern 2D and 3D techniques employing pulsed field gradients and inverse detection enables the full assignment of spectra and thus aids rapid structure elucidation. Some of these techniques are g-DQF-COSY, g-HMQC, g-HMBC, g-HSQC-TOCSY, and DPFGE-NOE.

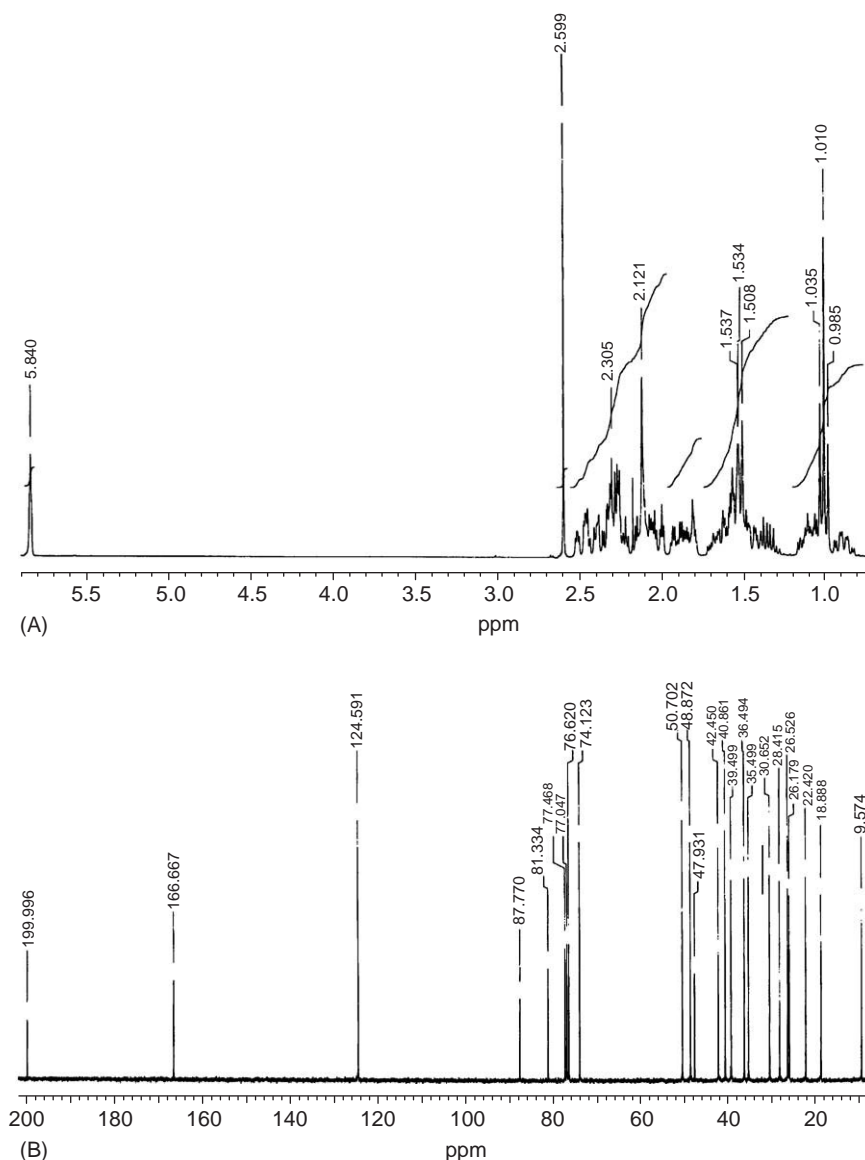
Figures 2 and 3 show the  $^1\text{H}$  and  $^{13}\text{C}$  spectra of norgestrel and the assignment of the signals.

**Mass spectrometry** Next to NMR spectroscopy in importance, mass spectrometry (MS) is the most

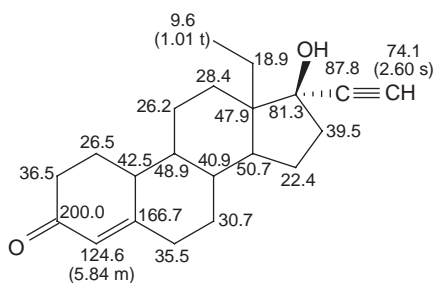
effective and most widely used technique for the structural elucidation of steroids.

The majority of steroid hormones and their synthetic analogs afford interpretable molecular ions and fairly characteristic fragmentation patterns, if the classical electron ionization technique in positive ion mode is used. Therefore, the latter is a very powerful method for structural elucidation. (A typical mass spectrum and fragmentation pattern are shown in Figures 4 and 5.) There are various possibilities for extending the utility of MS to less stable and less volatile steroids. With the aid of chemical derivatization (acetylation, trimethylsilylation, butyldimethylsilylation, etc.), their thermal stability and volatility can be increased. Chemical ionization affords molecular ions, even in the case of less stable steroids, but sacrifices at the same time the information afforded by fragmentation. With the aid of various field desorption and fast-atom bombardment techniques, highly polar steroids and their high relative molecular mass derivatives, such as conjugates (sulfates, glucuronides) and glycosides, can be analyzed.

Although generally, NMR spectroscopy affords more structural information, the great advantages of MS are its much higher sensitivity (enabling mass spectra to be obtained in the picogram range) and the possibility of coupling it with highly efficient chromatographic techniques (e.g., gas chromatography (GC)



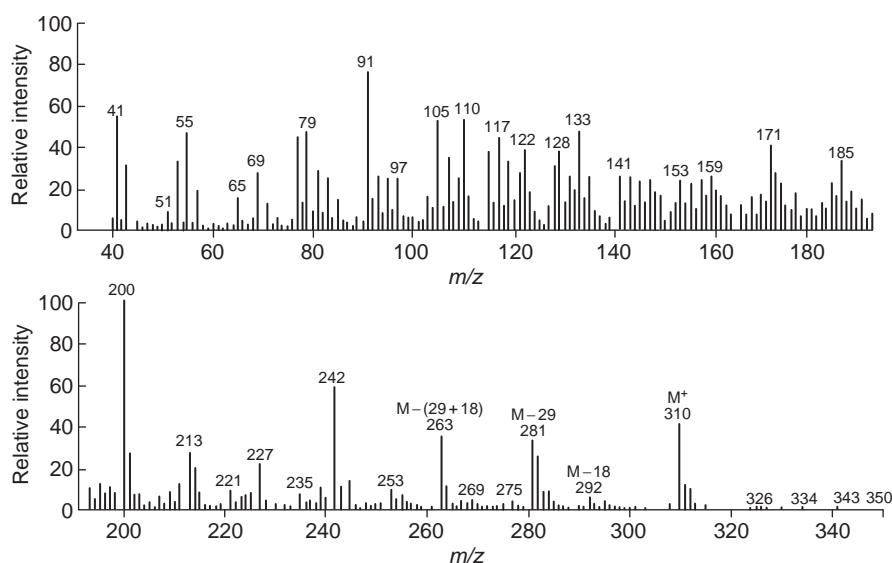
**Figure 2** NMR spectra of norgestrel. Solvent/reference:  $\text{CDCl}_3/\text{TMS}$ . (A)  $^1\text{H}$  NMR spectrum; (B)  $^{13}\text{C}$  NMR spectrum.



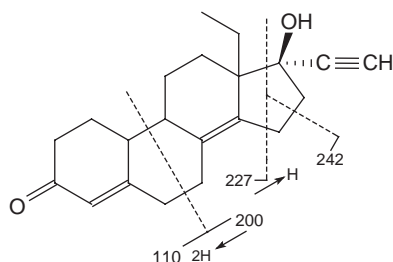
**Figure 3** Assignment of the NMR signals of norgestrel in ppm. The upper (or single) values are the chemical shifts of the respective carbon nuclei from the  $^{13}\text{C}$  spectrum, while the lower values (in parentheses) are the assigned proton shifts in the  $^1\text{H}$  spectrum. Solvent/reference:  $\text{CDCl}_3/\text{TMS}$ . (Courtesy of G. Balogh, Chemical Works of Gedeon Richter Ltd., Budapest.)

and HPLC) for both structural elucidation and quantification. GC-MS is a reference method in clinical steroid analysis; the potential of LC-MS, especially in the highly sensitive electrospray ionization mode, is still far from exhausted.

**Other methods for structure elucidation** Optical rotatory dispersion and circular dichroism spectroscopy are important methods for solving stereochemical problems in the steroid field. Moreover, thanks to the relatively rigid structure of the steroid skeleton and its chromophoric groups, steroids often serve as model compounds for such investigations. Specifically, the Cotton effect is easily observable in the steroid field.



**Figure 4** Electron impact mass spectrum of  $\Delta^{8(14)}$ -norgestrel (17 $\alpha$ -ethinyl-17-hydroxy-13-ethyl-gona-4,8(14)-dien-3-one).



**Figure 5** Fragmentation of  $\Delta^{8(14)}$ -norgestrel. Ionization voltage, 70 eV. (Courtesy of É. Csizér, Chemical Works, Gedeon Richter Ltd., Budapest.)

X-ray diffractometry, which requires only a single crystal of the analyte for molecular structure elucidation, is also used for the solution of the most difficult problems in the investigation of steroids. This is illustrated by atlases containing over 400 steroid structures. The potential of this technique in the investigation of receptor binding of steroid hormones merits special mention.

**An example of the application of spectroscopic techniques** The structure elucidation of an impurity in norgestrel detected by reversed-phase LC at a relative retention of 0.90 was carried out as follows.

The UV spectrum recorded by the liquid chromatograph's diode array UV detector shows an intense maximum at 236 nm. This indicates that the impurity contains the 4-ene-3-oxo moiety. However, the slightly hypsochromic shift (see Table 1), and especially the broadening of the peak as compared to that of norgestrel, indicate the presence of a double bond in the skeleton, presumably in the 8(14) position, which is not in conjugation with the

4-ene-3-oxo group but affects its conformation ('through-space conjugation'). The mass spectrum (obtained from GC-MS studies), with its molecular peak at  $m/z$  310, and the fragmentation as shown in Figure 5, indicate a norgestrel structure with a second double bond in ring A, B, or C. In order to obtain convincing evidence regarding the location of the double bond, a few milligrams of the impurity are isolated by preparative LC.

The IR spectrum of this sample, with its bands characteristic of the 4-ene-3-oxo group (1651 and 1622  $\text{cm}^{-1}$ ), 17-ethinyl group (3277 and 2100  $\text{cm}^{-1}$ ), and the 17-hydroxy group (3344 and 1089  $\text{cm}^{-1}$ ) (see Table 2), also indicates a 'norgestrel-like' structure but with no data with respect to the second double bond. The  $^1\text{H}$  NMR spectrum alone does not afford direct data for the location of the second double bond either: no protons are attached to carbon atoms of the second double bond. Therefore, its position could be 8(9), 8(14), or 9(10). The latter can be precluded on the basis of the above-mentioned UV spectrum. Convincing evidence for the location of the double bond in the 8(14) position is furnished by the joint application of  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy using 2D homonuclear and heteronuclear correlation experiments: (1) mapping the  $^1\text{H}$ - $^1\text{H}$ ,  $^1\text{H}$ - $^{13}\text{C}$ , and long-range  $^1\text{H}$ - $^{13}\text{C}$  coupling networks in the molecule ( $^1\text{H}$ - $^1\text{H}$  COSY, HETCOR, and FLOCK) and (2) comparing the  $^{13}\text{C}$  chemical shifts with those of norgestrel (see Figure 3). The most important  $^{13}\text{C}$  shifts in this regard are:

C8 129.3 ppm  
C9 42.3 ppm

C10 45.2 ppm  
 C13 49.2 ppm  
 C14 137.2 ppm  
 C15 24.4 ppm  
 C16 37.1 ppm

### Chromatographic Methods

**Thin-layer chromatography** Thin-layer chromatography (TLC) is very important in steroid analysis. The apolar skeletons and polar substituents of steroids make them good models for separation studies.

A variety of adsorbents have been successfully applied to the separation of steroids; in the majority of cases silica gel plates have been used. Although the separation power of these techniques does not approach that of the most developed forms of GC and LC, in practice most separation problems can be solved using TLC (especially by using the overpressure and high-performance variants of the method). Most of the steroid hormones (bearing either the 4-ene-3-oxo or phenyl group) strongly absorb UV radiation, thus enabling their detection at the ng level. A variety of spray reagents are available (mostly containing various mineral acids) to enhance detectability and to increase sensitivity.

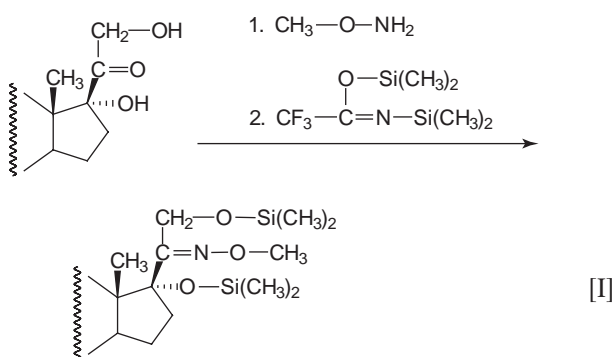
The main field of application of TLC is the rapid characterization of steroid mixtures and checking of the purity of bulk steroid hormones. For the latter purpose the pharmacopoeias widely recommend the semiquantitative estimation of impurities based on visual comparison of their spots with those of reference materials.

The early quantitative applications of TLC in steroid hormone analysis were based on spot elution followed by photometric measurement; this technique is still mentioned in some cases in pharmacopoeias. *In situ* densitometry, making use of the absorption of the colored zone (or fluorescence) that forms in the course of the reaction with spray reagents, is the basis of a great variety of methods. The linear range is usually in the region of 10 ng, but measurement of the fluorescence enables the limit of determination to be lowered to the subnanogram range. These methods are almost equivalent alternatives to the other chromatographic techniques in pharmaceutical steroid analysis, but they are only of secondary importance in the analysis of biological samples.

**Gas chromatography** The introduction of GC into steroid analysis in the beginning of the 1960s revolutionized the field. The high resolving power (especially of capillary GC) and the high sensitivity

attainable (especially with the electron capture and MS detectors) enabled hitherto unsolvable problems to be solved (e.g., the separation and quantification of minor steroid components in serum). The methodologies also found wide application in pharmaceutical steroid analysis.

The great dilemma of the GC analysis of steroids is whether to use derivatization or not. In the case of thermally unstable steroids, the answer is unequivocally yes. Steroids of this type are mainly the corticosteroids, containing the unstable  $\alpha$ -ketol side chain. Of several possibilities, double derivatization to form the methoxime trimethylsilyl derivatives is the most widely used (see reaction [I]):



Most of the other steroid hormones are sufficiently stable and volatile to be chromatographed without decomposition, and no derivatization is needed in the majority of pharmaceutical applications. However, in the analysis of biological specimens the hydroxy groups of these steroids are derivatized (mainly acetylated or trimethylsilylated) to avoid irreversible adsorption on the column when trace analyses are carried out. If the aim of the derivatization is to increase the sensitivity, heptafluorobutyrylation and the use of an electron-capture detector are recommended.

It should be mentioned that, with the exception of a few moderately important cases (e.g., GC of estrogen glucuronides after permethylation of both the steroid and the carbohydrate moieties), steroid conjugates (glucuronides or sulfates) are chromatographed as the free steroids after solvolytic or enzymatic hydrolysis.

In the early applications of GC to steroids packed columns were used, but capillary GC has gradually become the standard method in this field. The analyses are usually carried out at temperatures in the range 200–250°C, and hence the thermal stability of the stationary phases, which may be a consideration in some cases, is not problematic. Rather apolar or moderately polar stationary phases of the siloxane



type are most widely used. Fused-silica capillaries, coated with chemically bonded or cross-linked stationary phases, are the most modern separation columns. Relatively few columns are needed to cover all the problems encountered in steroid analysis. For example, a fused-silica capillary 25 m long with an i.d. of 0.2 mm and a layer thickness of 0.33  $\mu\text{m}$  of cross-linked silicone gum (5% diphenyl + 95% dimethyl) can be used to provide an apolar stationary phase. A similar column with a 50% trifluoropropyl + 50% methyl silicone phase can be used to provide a more polar phase. With the aid of two columns such as these, a variety of problems can usually be solved: impurity profiling of steroid drugs, their assay by the internal standard technique, investigation of reaction mixtures, analysis of fermentation liquors, etc.

An important feature of capillary GC is that it is easily combined with MS. The GC-MS technique is not only the most effective method for the identification of the steroid components of complex mixtures (profiling of biological samples, impurity profiling of industrial samples), but various kinds of mass fragmentographic techniques also enable the extremely selective and sensitive quantification of the separated components.

GC-(high resolution)MS after enzymatic hydrolysis of the glucuro- and sulphoconjugates and trimethylsilylation is the most widely used method for the determination of anabolic steroids and their metabolites in the urine of athletes. The most up-to-date technique for the solution of the difficult problem of differentiating between endogenous and exogenous testosterone is gas chromatography – combustion/isotope ratio ( $\text{C}^{13}/\text{C}^{12}$ ) mass spectrometry (GC-C-IRMS).

**Column liquid chromatography (CLC)** In the complex procedures aimed at determining individual steroids in biological samples, CLC plays a predominant role. This technique is mainly used for the separation of steroids from other compounds, for group fractionation of steroids, and for the separation of individual steroids prior to other (usually immunological) assays.

In addition to the classical column packing materials (silica gel, alumina, celite), other widely used columns are of the polydextran gel type: for the fractionation of the polar steroids, Sephadex LH-20, and for the less polar steroid compounds its hydroxyalkyl derivative, Lipidex. The use of ion exchange columns (e.g., DEAE-Sephadex A-25, Amberlite XAD-2) is also widespread. High selectivity is attainable with immunoaffinity chromatography, where antibodies are immobilized on chemically modified agarose gel and filled in columns. For the solid-phase

extraction of steroids from biological fluids  $\text{C}_{18}$ -silica cartridges are most widely used.

**Liquid chromatography** LC was introduced into the analysis of steroid hormones ~10 years after GC. This field of application is now shared by the two techniques, and the advantages and disadvantages of LC must be taken into account when an explanation is sought as to why LC has superseded GC in some areas while GC retained its position in others.

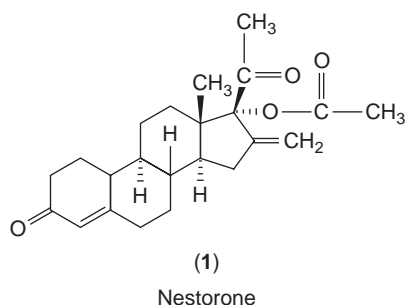
The advantages of LC in steroid hormone analysis are:

1. It is usually carried out at ambient temperature, and for this reason it is applicable quite generally and without derivatization or hydrolysis, even in those cases where thermal instability or low volatility precludes the direct application of GC (separation and quantification of corticosteroids, conjugates, etc.).
2. Although the resolving power of LC is generally lower than that of capillary GC, the greater versatility of LC may be exploited so well that suitable combinations of mobile and stationary phases can be found for the separation of practically all steroids, including 'critical pairs' (epimers, derivatives with various degrees of unsaturation, etc.).
3. LC enables even minor steroid components to be obtained on the semipreparative or preparative scale. This facilitates the identification of metabolites, impurities, and degradation products by subsequent spectroscopic studies.
4. Although coupling the mass spectrometer with the high-performance liquid chromatograph involves greater technical problems than is the case with GC-MS, which is more popular in steroid hormone analysis, LC-MS has already been successfully introduced into steroid profiling. A unique feature of LC is that it can be linked with radioimmunoassay (RIA). The combination of the high resolving power of the former and the high selectivity and sensitivity of the latter produces an extremely powerful method.

Of the various techniques, reversed-phase systems are most frequently used. Their use in biochemical analysis is almost exclusive. It should be mentioned that for the successful determination of the highly polar steroids fully end-capped  $\text{C}_{18}$  columns must be used. Although reversed-phase systems are far more widely used in pharmaceutical steroid analysis, the use of normal-phase systems enables very difficult separation problems to be solved.

Most steroid hormones possess a strong enough UV absorption for sensitive monitoring with the UV detector. For instance, 4-ene-3-oxo steroids can be measured at 240 nm in the nanogram range. This is similar to the sensitivity of the commonly used flame ionization detector in GC. Even the spectrophotometrically inactive steroid hormone derivatives contain weak chromophores (isolated double bonds, oxo groups, etc.), which enable their monitoring in the rather inconvenient short-wavelength range (200–210 nm). The selectivity and sensitivity of detection can be increased by precolumn or postcolumn derivatization. However, derivatization is of only moderate importance in steroid hormone analysis.

An advantageous feature of LC with a diode-array UV detector is that UV spectra are easily obtainable. This is illustrated in **Figures 6A** and **6B** where the diode-array UV spectra of Nestorone and its impurities obtained after reversed-phase LC separation are shown. As seen in **Figure 6B** not only the presence of additional double bond (5) but also an oxo group (4) in conjugation with the chromophoric 4-ene-3-oxo group of Nestorone (1) (see **Table 1**) is of diagnostic value but even the presence of hydroxyl groups at 6 $\alpha$ - (2) or 6 $\beta$ - (3) position causes small but characteristic shift of the absorption maxima.



## Immunoassay Methods

**Introduction** Steroids are not immunogenic, but as haptens covalently bound to a suitable protein they can be used to raise antisera. This is the basis of their determination by immunoanalytical methods. Antibodies raised in this way can more or less selectively bind any steroid, whose protein complex has been used for immunization of the animals producing the antiserum. The competition between the steroid to be determined and a suitably labeled analog forms the basis of immunoassay. In the overwhelming majority of cases polyclonal antibodies are used. The advantage of monoclonal antibodies is their well-defined specificity. A disadvantage, however, is their lower affinity to the antigen, causing difficulties in attaining

high sensitivity. In addition, they are more sensitive to variations in pH and ionic strength.

Depending on the manner of labeling, various kinds of immunoassays are used in the analysis of steroid hormones: RIA, enzyme immunoassay, fluorescence immunoassay, fluorescence polarization immunoassay, and luminescence immunoassay.

While the other quantitative analytical methods for the determination of steroid hormones are more or less generally used for the investigation of both pharmaceutical products and biological samples, the immunoassay methods are used exclusively for biological samples. Although the majority of the problems routinely investigated by these techniques in clinical and biochemical analysis could also be solved by other (mainly chromatographic) methods, immunoassays are the most commonly used methods for the determination of individual steroid hormones in biological samples. The reasons for this are the speed, simplicity, and relatively low cost of these techniques, as well as their high selectivity and extremely high sensitivity. The selectivity of the immunoassay methods can be further increased by preliminary chromatographic separation. Thus, they can serve as extremely sensitive offline detection systems in LC profiling of steroids.

**Radioimmunoassay** For the production of the antiserum, steroids are usually bound to bovine serum albumin (BSA). Usually 10–30 molecules of the steroid per mole of BSA are bound. The easiest binding method is via a hemisuccinate bridge between the 17- or 3-hydroxy groups of the steroid and the lysine moiety of the protein, or the O-carboxymethyloxime bridge between the 3- or 20-oxo group and the lysine moiety of the protein. The selectivity can be increased if the same bridges are built up at other locations in the steroid nucleus where originally no functional group was present. For example, an antiserum for testosterone prepared via the 3-O-carboxymethyloxime has a cross-reactivity of 65% to 5 $\alpha$ -dihydrotestosterone, while this value can be reduced to 0.8% if the 6 $\beta$ -carboxymethyl derivative is linked to the protein.

A variety of tritiated ( $^3\text{H}$ ) steroids are available as the radiolabeled compounds for use in the RIA. The relatively low sensitivity of  $^3\text{H}$  can be increased by multiple labeling. Another alternative is  $^{125}\text{I}$  labeling. Radioiodinated aromatic amines or amino acids, such as tyramine, tyrosine, or histamine, are linked to the steroid by the above-mentioned reactions to obtain the label.

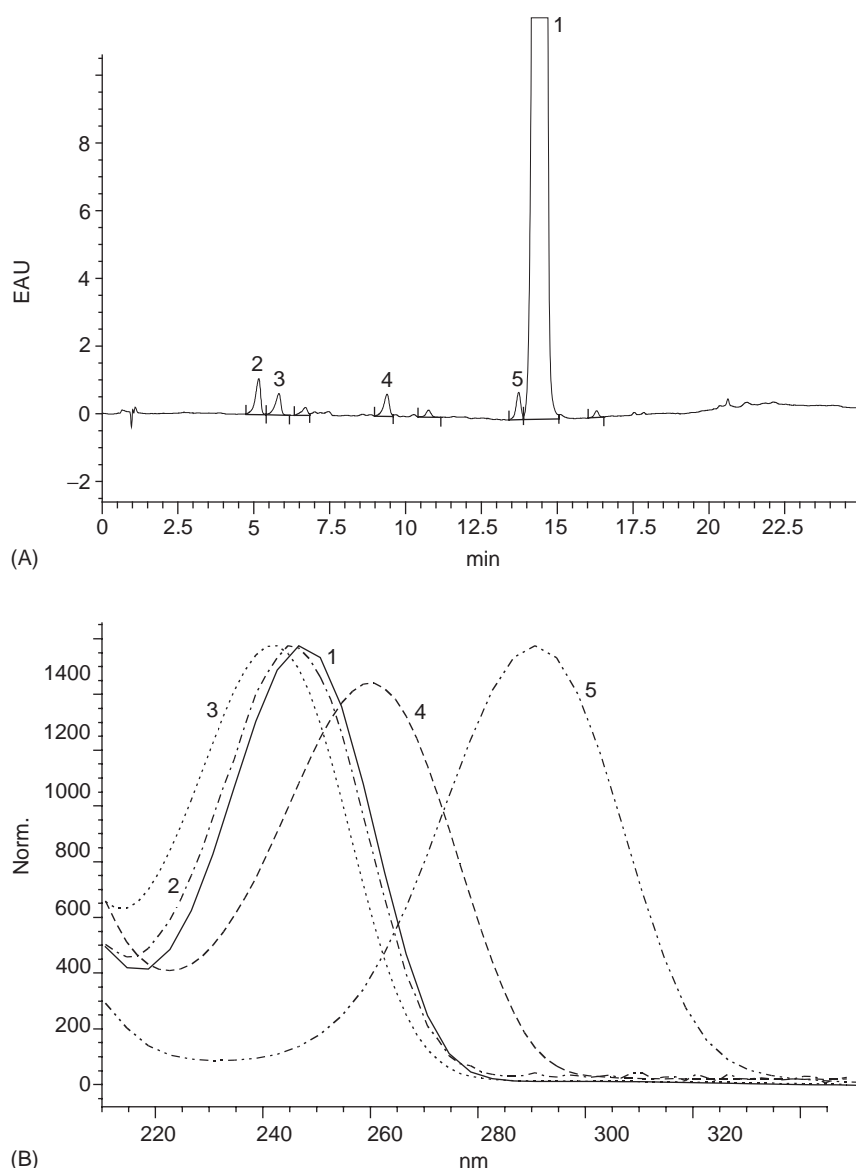
After the equilibration stage of the assay, the separation of the labeled steroid, bound to the antibody from the free steroid, can be achieved in several

ways. The classical, but still widely used method, is the dextran-coated charcoal (DCC) method. Here, the free steroid is adsorbed on DCC, removed by centrifugation, and the radioactivity present in the supernatant is determined by liquid scintillation counting (in the case of  $^3\text{H}$  labeling) or by gamma counting (in the case of  $^{125}\text{I}$  labeling). Other methods based on adsorption and/or precipitation (e.g., with polyethylene glycol) followed by centrifugation are also used in steroid RIAs. In the double-antibody technique, which merits special mention, the steroid bound to the antibody is precipitated by adding a

second antibody. In this case, the free steroid label is measured. Some RIA kits contain magnetizable particles for the bound-free separation.

In the most widely used solid-phase methods no centrifugation step is required. The commercially available kits for steroids contain the antibody (or the second antibody, in the double-antibody technique) immobilized on the walls of a tube, plastic beads, sticks, or membranes.

The sensitivity of the RIA of steroids depends on the affinity of the antibody for the steroid, the background effect of nonspecific binding, the temperature



**Figure 6** LC separation (A) and diode-array UV spectra (B) of Nestorone (1) and its impurities, 6 $\alpha$ -hydroxy (2), 6 $\beta$ -hydroxy (3), 6-oxo (4), 6-ene (5) derivatives. Column: Nova Pack C-18, 4  $\mu\text{m}$ , 150  $\times$  3.9 mm. Eluent: A: acetonitrile–water 1:9 v/v, B: acetonitrile. Linear gradient, 0 min 25% B, 14 min 45% B, 20 min 75% B. Flow rate 1 ml min $^{-1}$ . UV detector 254 nm. (Reproduced with permission from Görög S, Babják M, Balogh G, *et al.* (1998) Estimation of impurity profiles of drugs and related materials Part 19. Theme with variations. Identification of impurities in 3-oxosteroids. *Journal of Pharmaceutical and Biomedical Analysis* 18: 511–525.)

of incubation, and the nature and specific activity of the radiolabel. In typical, optimized methods the calibration graph covers the range 10–1000 pg.

**Enzyme immunoassay** The tendency to avoid the use of radioisotopes in immunoassays has led to the increasing use of nonisotopic labels in steroid hormone assays. Of these, enzyme immunoassays are most widely used.

The coupling of steroids to enzymes is carried out in a manner similar to that described in the previous section for their coupling to BSA. The most frequently used enzymes in steroid assays are horseradish peroxidase and alkaline phosphatase. Although the enzyme activity is conserved in the course of coupling, it decreases during the immunogenic reaction. The enzyme–steroid complex used in the competitive equilibrium may be the same as that used to raise the antiserum (homologous system) or different (heterologous system). The advantage of the former is higher specificity, while that of the latter is improved sensitivity.

Although unlike RIA, enzyme immunoassays can be carried out in homogeneous systems without a separation step (based on the change in enzyme activity during the immune reaction), in practice, heterogeneous (enzyme-linked immunosorbent assay) methods are much more frequently used. The antibody, or in the case of the double-antibody method the second antibody, is immobilized, either covalently or by coating enzyme multiplied immunoassay technique (EMIT). This can be done on the walls of microtiter plates. After the immunogenic reaction, the enzyme activity, which is the equivalent of radioactivity in RIA systems, can be measured by suitable photometric methods on the microtiter plates themselves.

A typical photometric reaction is, for example, the interaction between hydrogen peroxide and 2-phenylenediamine, leading to the colored quinonediimine with an absorption maximum at 492 nm. This reaction is catalyzed by the above-mentioned peroxidase enzyme and, hence, the absorbance is proportional to the concentration of the unbound enzyme steroid complex in the equilibrium mixture.

Generally, the specificities and sensitivities of the enzyme immunoassay methods for steroid hormones are comparable with those of the RIA methods. The sensitivity can be improved by using a fluorophoric reaction to replace the chromophoric reaction. For instance, the horseradish peroxidase discussed above also catalyzes the reaction between hydrogen peroxide and 4-hydroxyphenylacetic acid. The interaction of the oxidation product with glycine leads to a fluorophore (excitation 327 nm, emission 410 nm).

**Other immunoassays** No fluorogenic reaction is needed if the radiolabel or the enzyme label is replaced by a fluorogenic label, which can be measured after the bound-free separation. An example of such a label is 4-methylumbelliferone 3-acetic acid. The advantage of these fluoroimmunoassay methods (FIA) is their simplicity. However, a disadvantage is that, as a consequence of the strong background fluorescence, the sensitivity is poor.

This disadvantage can be overcome by making use of the excellent fluorescence properties of the lanthanide (mainly  $\text{Eu}^{3+}$ ) chelates. These are characterized by high quantum yield, extremely large Stokes' shift (excitation  $\sim 350$  nm, emission  $\sim 610$  nm), narrower emission peaks, and extremely long fluorescent half-life (5–6 orders of magnitude longer than that of the background fluorescence). This enables the two kinds of fluorescence to be separated in the time-resolved fluorimeter, which is widely used in steroid hormone analysis.

The basis of the measurement is that the lanthanide chelate (e.g., isothiocyanatophenyldiethylenetriamine pentaacetic acid- $\text{Eu}^{3+}$ ), conjugated with the antibody, is not fluorescent. In the solid-phase system, after the bound-free separation, effected by simple washing of the walls of the microtiter plates, the above complex of  $\text{Eu}^{3+}$  is caused to dissociate through the action of a stronger complex-forming agent (2-naphthoyltrifluoroacetone), which forms a chelate with the good fluorescence properties listed above. These 'dissociation-enhanced lanthanide fluoroimmunoassay' methods are very popular for the assay of a variety of steroids in biological specimens. In addition to its high sensitivity (comparable to that of RIA), this method is easy to automate.

Other variants of immunoassay methods used in the analysis of steroid hormones are the fluorescence polarization immunoassay, where the label is fluorescein, coupled to the steroid, and luminescence immunoassay, where the steroid is bound to aminoalkylethyl luminol. The basis of the luminescence assay is the luminescence generated by the action of hydrogen peroxide/peroxidase, which is measured by a luminometer (LIA, CELIA).

## Applications

It can be concluded from the outline of the tasks to be solved by steroid analysts and the methods available for these purposes that the gap between the methodologies of industrial/pharmaceutical/regulatory steroid analysis and that of the biological/clinical field is rather wide, and further widening is to be expected in the future as a consequence of the refinement of the methodologies. The main reason

for this is that in the first field the quantity of the sample available for the analysis is practically unlimited, the number of components is usually low, and matrix effects are easy to handle. For these reasons, sensitivity is not a major problem and the demands regarding specificity are also few. On this basis, UV spectrophotometric and colorimetric methods still play an important role and GC as well as LC (the most widely used technique in industrial/pharmaceutical/regulatory steroid analysis) can be used with their conventional detectors.

At the same time, in biological/clinical steroid analysis where extremely low concentrations of steroids (and at an increasing rate their individual conjugates) have to be determined in the presence of several structurally closely related derivatives and in complicated matrices, both the sensitivity and the selectivity of the methods are of great importance, but the demands regarding the accuracy and precision of the methods are naturally lower than in the other fields. For this reason immunological methods are most widespread either directly or after chromatographic separation. The use of colorimetry and fluorimetry is becoming restricted to various kinds of enzyme- and fluoroimmunoassays where these labels and reactions serve as the basis of the quantitative evaluation. GC plays an important role in this field too, mainly in conjunction with MS. This is not only a tool for the identification of steroids but also a reference method for their quantification in clinical samples. This is the predominant method in the doping analysis of athletes. The importance of LC-MS, too, has greatly increased due to the use of the new ionization techniques in mass spectrometry.

**See also:** **Chemiluminescence:** Overview. **Chiroptical Analysis. Derivatization of Analytes. Enzymes:** Enzyme-Based Assays. **Fluorescence:** Overview; Clinical and Drug Applications. **Gas Chromatography:** Column Technology; Mass Spectrometry. **Immunoassays:** Overview. **Immunoassays, Applications:** Clinical. **Immunoassays, Techniques:** Enzyme Immunoassays; Luminescence Immunoassays. **Infrared Spectroscopy:** Overview. **Liquid Chromatography:** Column Technology; Normal

Phase; Reversed Phase; Liquid Chromatography–Mass Spectrometry. **Polarography:** Organic Applications. **Spectrophotometry:** Pharmaceutical Applications. **Thin-Layer Chromatography:** Overview. **X-Ray Fluorescence and Emission:** X-Ray Diffraction – Single Crystal.

## Further Reading

- De Pablo F, Scanes CG, and Weintraub BD (eds.) (1993) *Handbook of Endocrine Research Techniques*. San Diego: Academic Press.
- De la Torre R and Krstulovic AM (eds.) (1996) *Doping Control* (Special issue of *Journal of Chromatography B*, vol. 687, no. 1).
- Duax WL and Griffin JF (1998) Steroid hormone structure, receptor binding and activity: empirical drug design. In: Codding PW (ed.) *Structure-Based Drug Design*, pp. 1–14. Dordrecht: Kluwer Academic Publishers.
- Gosling JP (ed.) (2000) *Immunoassays: A Practical Approach*. Oxford: Oxford University Press.
- Görög S (1983) *Quantitative Analysis of Steroids*. Amsterdam: Elsevier.
- Görög S (ed.) (1989) *Steroid Analysis in the Pharmaceutical Industry*. Chichester: Ellis Horwood.
- Görög S (1995) *Ultraviolet–Visible Spectrophotometry in Pharmaceutical Industry*, pp. 323–334. Boca Raton, FL: CRC Press.
- Görög S (ed.) (2000) *Identification and Determination of Impurities in Drugs*. pp. 712–731. Amsterdam: Elsevier.
- Görög S and Szász Gy (1978) *Analysis of Steroid Hormone Drugs*. Amsterdam: Elsevier.
- Kautsky MP (ed.) (1981) *Steroid Analysis by HPLC*. New York: Dekker.
- Makin HL, Gower DB, and Kirk DN (eds.) (1995) *Steroid Analysis*. Glasgow: Blackie.
- Makin HL, Trafford DJH, and Nolan J (1998) *Mass Spectra and GC Data of Steroids: Androgens and Gestagens*. Weinheim: Wiley-VCH.
- Schänzer W, Geyer H, Gotzmann A, and Mareck U (2002) *Recent Advances in Doping Analysis*. Köln: Sport und Buch Strauß.
- Szepesi G (1986) Assay of steroids and related materials. In: Treiber L (ed.) *Quantitative TLC and its Industrial Application*. New York: Dekker.
- Touchstone JC (1986) *Handbook of Chromatography, Steroids*. Boca Raton, FL: CRC Press.

# HUMIC AND FULVIC COMPOUNDS

**M Filella, J Buffle, and N Parthasarathy**, University of Geneva, Geneva, Switzerland

## Introduction

Humic substances refer to products resulting from the decomposition of plant and animal residues. They are omnipresent in soils, sediments, and water.



for this is that in the first field the quantity of the sample available for the analysis is practically unlimited, the number of components is usually low, and matrix effects are easy to handle. For these reasons, sensitivity is not a major problem and the demands regarding specificity are also few. On this basis, UV spectrophotometric and colorimetric methods still play an important role and GC as well as LC (the most widely used technique in industrial/pharmaceutical/regulatory steroid analysis) can be used with their conventional detectors.

At the same time, in biological/clinical steroid analysis where extremely low concentrations of steroids (and at an increasing rate their individual conjugates) have to be determined in the presence of several structurally closely related derivatives and in complicated matrices, both the sensitivity and the selectivity of the methods are of great importance, but the demands regarding the accuracy and precision of the methods are naturally lower than in the other fields. For this reason immunological methods are most widespread either directly or after chromatographic separation. The use of colorimetry and fluorimetry is becoming restricted to various kinds of enzyme- and fluoroimmunoassays where these labels and reactions serve as the basis of the quantitative evaluation. GC plays an important role in this field too, mainly in conjunction with MS. This is not only a tool for the identification of steroids but also a reference method for their quantification in clinical samples. This is the predominant method in the doping analysis of athletes. The importance of LC-MS, too, has greatly increased due to the use of the new ionization techniques in mass spectrometry.

**See also:** **Chemiluminescence:** Overview. **Chiroptical Analysis. Derivatization of Analytes. Enzymes:** Enzyme-Based Assays. **Fluorescence:** Overview; Clinical and Drug Applications. **Gas Chromatography:** Column Technology; Mass Spectrometry. **Immunoassays:** Overview. **Immunoassays, Applications:** Clinical. **Immunoassays, Techniques:** Enzyme Immunoassays; Luminescence Immunoassays. **Infrared Spectroscopy:** Overview. **Liquid Chromatography:** Column Technology; Normal

Phase; Reversed Phase; Liquid Chromatography–Mass Spectrometry. **Polarography:** Organic Applications. **Spectrophotometry:** Pharmaceutical Applications. **Thin-Layer Chromatography:** Overview. **X-Ray Fluorescence and Emission:** X-Ray Diffraction – Single Crystal.

## Further Reading

- De Pablo F, Scanes CG, and Weintraub BD (eds.) (1993) *Handbook of Endocrine Research Techniques*. San Diego: Academic Press.
- De la Torre R and Krstulovic AM (eds.) (1996) *Doping Control* (Special issue of *Journal of Chromatography B*, vol. 687, no. 1).
- Duax WL and Griffin JF (1998) Steroid hormone structure, receptor binding and activity: empirical drug design. In: Codding PW (ed.) *Structure-Based Drug Design*, pp. 1–14. Dordrecht: Kluwer Academic Publishers.
- Gosling JP (ed.) (2000) *Immunoassays: A Practical Approach*. Oxford: Oxford University Press.
- Görög S (1983) *Quantitative Analysis of Steroids*. Amsterdam: Elsevier.
- Görög S (ed.) (1989) *Steroid Analysis in the Pharmaceutical Industry*. Chichester: Ellis Horwood.
- Görög S (1995) *Ultraviolet–Visible Spectrophotometry in Pharmaceutical Industry*, pp. 323–334. Boca Raton, FL: CRC Press.
- Görög S (ed.) (2000) *Identification and Determination of Impurities in Drugs*. pp. 712–731. Amsterdam: Elsevier.
- Görög S and Szász Gy (1978) *Analysis of Steroid Hormone Drugs*. Amsterdam: Elsevier.
- Kautsky MP (ed.) (1981) *Steroid Analysis by HPLC*. New York: Dekker.
- Makin HL, Gower DB, and Kirk DN (eds.) (1995) *Steroid Analysis*. Glasgow: Blackie.
- Makin HL, Trafford DJH, and Nolan J (1998) *Mass Spectra and GC Data of Steroids: Androgens and Gestagens*. Weinheim: Wiley-VCH.
- Schänzer W, Geyer H, Gotzmann A, and Mareck U (2002) *Recent Advances in Doping Analysis*. Köln: Sport und Buch Strauß.
- Szepesi G (1986) Assay of steroids and related materials. In: Treiber L (ed.) *Quantitative TLC and its Industrial Application*. New York: Dekker.
- Touchstone JC (1986) *Handbook of Chromatography, Steroids*. Boca Raton, FL: CRC Press.

# HUMIC AND FULVIC COMPOUNDS

**M Filella, J Buffle, and N Parthasarathy**, University of Geneva, Geneva, Switzerland

## Introduction

Humic substances refer to products resulting from the decomposition of plant and animal residues. They are omnipresent in soils, sediments, and water.

They constitute  $\sim 75\%$  w/w of the organic matter in most soils and  $\sim 50\%$  of the organic carbon in surface waters. Humic substances are composed of complex heterogeneous mixtures of organic compounds and are characterized as being yellow to brown in color, of high relative molecular mass, and refractory. Unlike many other natural organic products, they cannot be described in terms of unique chemical structures and are operationally defined by the technique used for their extraction and fractionation.

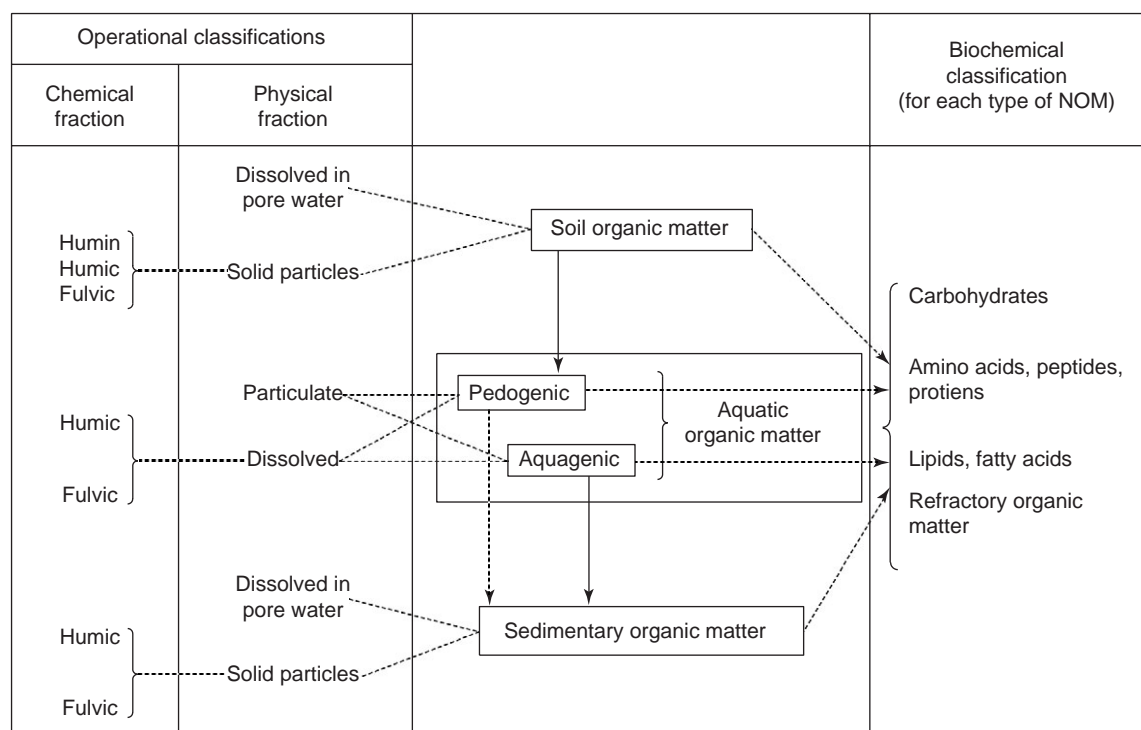
Soil humic substances have been studied since the eighteenth century because they are essential for holding water and nutrients and releasing it to plants. In recent years, the importance of humic substances in aquatic systems, especially in water quality, was realized and aquatic humic substances have been investigated by scientists in aquatic chemistry and limnology. The presence of these substances in natural water systems affects markedly the productivity of plant systems, influences the distribution of many other organic and chemical species in the environment, and produces toxic by-products when exposed to water disinfectants containing halogen.

## Definition and Classification

Natural organic matter (NOM), other than living organisms and anthropogenic compounds, possesses

a large variety of properties and is composed of an extremely complex mixture of compounds, most of which are not yet identified. Taking into account the diversity of the natural processes of synthesis and degradation, the number of constituents in this mixture can be considered as infinite and therefore there is little hope of completely separating them. Consequently, studies of NOM have nearly always been concerned not with pure compounds but rather with groups of compounds separated from the initial mixture by means of relatively arbitrary chosen techniques.

Figure 1 shows the classification of various types of NOM. The biochemical classification is the best justified from a theoretical perspective. It distinguishes between the three major classical groups: carbohydrates, proteins, and lipids. In water, these together represent only  $\sim 20\text{--}30\%$  of the NOM, the remaining components being materials resistant to degradation (refractory organic matter). The physical size classification serves to distinguish between particulate and dissolved NOM. The boundary is defined in a purely operational and arbitrary way, usually by use of  $0.45\text{ }\mu\text{m}$  pore filters. The chemical fractionation classification yields fractions that may include mixtures, in varying proportions, of the 'biochemical' compound mentioned above. The most widely used fractionation is separation into fulvic and humic compounds. Humic compounds are the fraction that



**Figure 1** Classification of various types of NOM. (From Buffle J (1984) Natural organic matter and metal-organic interactions in aquatic systems. In: Sigel H (ed.) *Metal Ions in Biological Systems*, p. 165. New York: Dekker.)

is soluble at alkaline pH but precipitates at acidic pH, and fulvic compounds the fraction water soluble at alkaline and acidic pH. The terms humic and fulvic compounds imply only specific acid–base solubility behavior, and these organic fractions isolated from diverse environments do not necessarily have similar chemical structures or result from similar source materials or formation processes.

The environment clearly influences humic substance formation not only by dictating the range of possible precursors but by influencing the type of possible processes. Thus, the principal source of seawater NOM is planktonic and bacterial excretion and degradation while, in contrast, terrestrial biomass and NOM production is principally provided by higher plants. A major problem in elucidating environmental influences is that humic substances are often isolated from different environments than those in which they were formed. For example, of all aquatic systems, lakes present one of the most complex mixtures of NOM because they accumulate in a limited space organic matter from several different sources: primary autochthonous productivity of the same type as that of oceanic pelagic zones, allochthonous inputs, and the release of sediment organic matter. A classification of humic materials based on the origin of the compounds (aquagenic, pedogenic (Figure 1)) would probably be more environmentally sound than the current separation schemes based on pure analytical considerations.

## Isolation and Fractionation

Humic acids do not occur alone in the environment. They occur in mixtures with amino acids, sugars, various aromatic and aliphatic acids, and other organic compounds, and/or associated with mineral components. In order to study the chemistry of humic substances, they must be first isolated. Because humic and fulvic acids are defined by the technique that is used for their extraction, and in order to get comparable results, it is particularly recommended to follow the isolation procedures used by the International Humic Substance Society (IHSS).

### Isolation from Soils

The classical method for isolation of humic substances from solid-phase source materials such as soils, peat, and leonardite is alkaline extraction with aqueous NaOH, followed by precipitation of humic acid at low pH, and a series of desalting steps involving cation exchange, dialysis, etc., to obtain fulvic acid. Through this procedure, samples extracted from solid-phase materials include both

hydrophobic and hydrophilic acids. A number of methods for the extraction of humic substances from soil using sodium hydroxide solution have been published. These methods are generally successful and yield comparable results. The one described below is the method currently used by the IHSS. It produces relatively high yields and can be used as a standard method for comparisons between and within laboratories.

After removing roots and sieving the dried soil sample to pass a 2.0 mm sieve, the sample is equilibrated at a pH value between 1 and 2 with HCl (final ratio 10 ml liquid–1 g dry sample). The supernatant is kept for the isolation of the fulvic acid. The soil sample is first neutralized to pH 7 with  $1 \text{ mol l}^{-1}$  NaOH then  $0.1 \text{ mol l}^{-1}$  NaOH is added under  $\text{N}_2$  atmosphere to give a final extractant-to-soil ratio of 10:1. After extraction and settling for some hours, the alkaline supernatant is collected and acidified with  $6 \text{ mol l}^{-1}$  HCl to pH 1. After standing for 10–12 h, centrifugation allows the separation of the humic acid (precipitate) from the fulvic acid (supernatant). The two supernatants containing the fulvic acid fraction are purified by using a XAD-8 resin. They are finally  $\text{H}^+$ -saturated by using a cation-exchange resin and freeze-dried. The humic acid fraction is redissolved by adding KOH under  $\text{N}_2$  and purified by repeated precipitation, treatment with HCl and HF, and dialysis. Humic acids are freeze-dried.

XAD-8 is a nonionic, macroporous (pore size  $25 \mu\text{m}$ ), methyl methacrylate ester resin. Because it is sometimes difficult to obtain, it may be necessary to use an alternative resin such as Polyclar, which is a cross-linked poly(vinylpyrrolidone). Extensive purification of the resins is always required before use.

If it is not possible to purify the fulvic acid using resin treatments, exhaustive dialysis against distilled water is an alternative method. However, it is less convenient because much of fulvic acid has proved to pass through commercially available dialysis membranes.

### Isolation from Water

Humic substances can be isolated from waters by using column adsorption techniques and concentrated by methods such as vacuum distillation, lyophilization, freezing concentration, ultrafiltration, reverse osmosis. The IHSS has adopted the XAD-8 resin adsorption method to isolate humic and fulvic acids from natural waters. In this method, after filtration through a  $0.45 \mu\text{m}$  silver or polymer membrane filter, the dissolved organic matter (DOM) is fractionated initially into hydrophobic and hydrophilic fractions through preferential adsorption of the hydrophobic fraction on the XAD-8 resin (water sample preacidified at pH 2

with HCl). Subsequently, the hydrophobic fraction is eluted from the resin by alkaline extraction with aqueous  $0.1 \text{ mol l}^{-1}$  NaOH, followed by precipitation of humic acid at low pH (pH 1, adjusted with HCl), and a desalting step involving cation exchange to obtain fulvic acid. Finally, fulvic and humic acids are freeze-dried. The humic acid and fulvic acid that are isolated from aqueous samples thus contain only hydrophobic organic acids. Starting in the late 1990s, IHSS has also employed reverse osmosis to concentrate NOM from selected water sources.

Standard and reference samples of humic substances (river, soil, peat, leonardite) are available from the IHSS. They can be used for quality control of analytical work on humic substances.

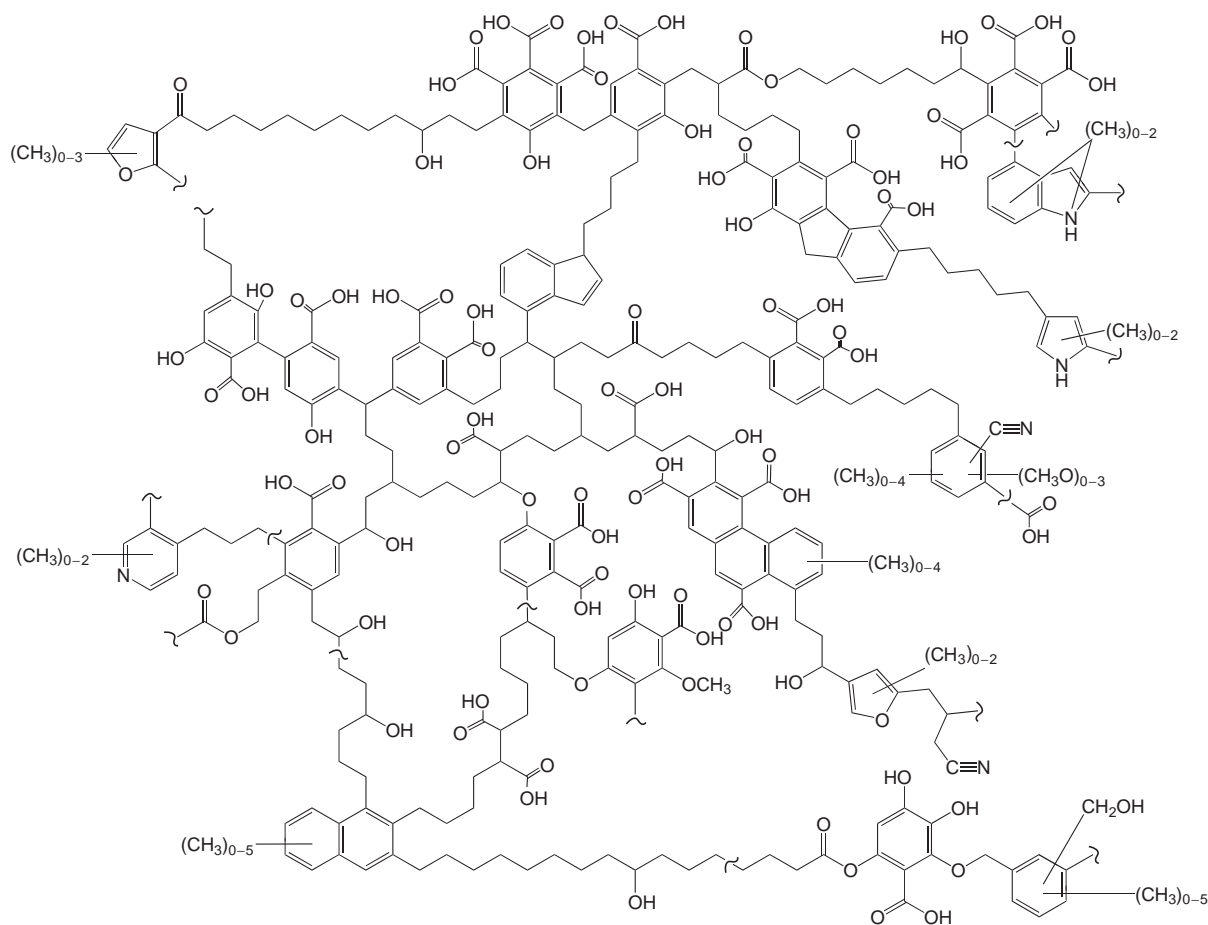
### Characterization: Physical Properties

Any isolate of soil and water humic substances is intrinsically polydisperse, and physical and chemical properties derived from these substances are average values (Figure 2). In some cases, it is preferable to

determine distribution curves (or spectra) that relate the changes in the population of the distribution to the value of the parameters used (e.g., number of particles versus particle size, number of complexing sites versus their complexation stability constants). When this approach is not taken, the polydispersity of the system should be addressed in order to guide the average values estimation.

### Charge

Humic substances invariably carry an electrostatic charge. They are therefore classified as polyelectrolytes. Although a variety of sites possessing the potential to carry a positive or negative charge can be identified on humic substances, the predominant charge is negative and at most naturally occurring pH values this charge is due to the dissociation of carboxylic, hydroxyl, and phenolic groups. Typically, soil humic acids have 3–4 charge types per 1000 Da, which is a relatively high level for a naturally occurring polyelectrolyte. The presence of



**Figure 2** Proposed model chemical structure for fulvic compounds. (Reproduced with permission from Schulten H-R and Schnitzer M (1993) *Naturwissenschaften* 80: 29–30.)



electric charge and of the necessary counterbalancing ions has a marked effect on any measurement of relative molecular mass, shape, or size in solution, as well as on the metal complexation properties of these substances. A typical example of this effect is exhibited when charge of the humic substances in soil samples is determined by *in situ* and laboratory measurements. The total acidity of the *in situ* soil organic matter is much lower than that of soil extracted humic and fulvic acids. This is probably due to the potential charge site being blocked by interaction with other soil components.

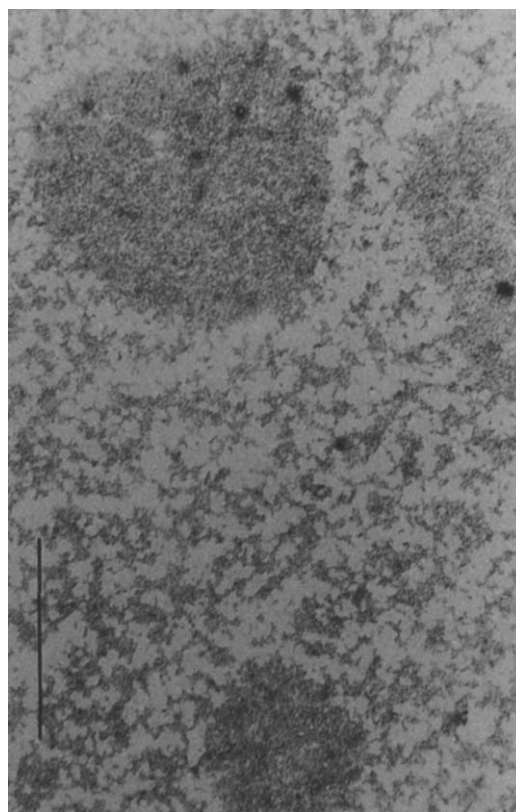
### Degree of Hydration

The degree of hydration of humic compounds is high. It can be estimated from the hydration number of the hydrophilic groups and from the composition of humic and fulvic acids. Calculations show that at least 0.8 and 0.9–1.5 g of water are retained per gram of soil humic acid and soil fulvic acid, respectively. A similar value for the degree of hydration has been found for dissolved water fulvic acid. Hydrogen-bridge formation is probably the basis for hydration of both fulvic and humic acids.

### Relative Molecular Mass, Size, and Shape

Numerous methods have been applied for the determination of the relative molecular mass, size, and shape of humic and fulvic acids: colligative properties measurements (vapor pressure osmometry, cryoscopy), viscometry, exclusion (gel) chromatography, field flow fractionation, ultracentrifugation, ultrafiltration, small-angle X-ray scattering, light scattering, electrophoresis, fluorescence correlation spectroscopy, transmission electron microscopy (TEM) (Figure 3), and atomic force microscopy (AFM) (Figure 4). Size determinations depend on the polydispersity and polyelectrolytic nature (charge effects) of the compounds as well as on their aggregation and hydration properties. Consequently, the results in the literature are relatively disparate. Data on the molecular radii of humic substances vary from fractions of nanometers to several hundreds of nanometers. In general, the aggregation phenomena can be minimized by using the lowest possible concentrations of organic matter and electrolytes and media not too acidic.

As a result of polydispersity of the compounds, the measured relative molecular masses are weighted averages of the relative molecular masses of all the individual components. Depending on the method used, the weighting can be based on the number of molecules containing the sample (osmotic pressure, vapor pressure osmometry, depression of freezing point), on their weight proportion with respect to the



**Figure 3** Transmission electron micrographs of a water fulvic acid; spheres (minimum visible size 2–3 nm) and spheroidal aggregates. Scale bar = 450 nm. (Reprinted with permission from Leppard GG, Buffle J, and Baudat R (1986) *Water Research* 20: 185–186; © Elsevier.)

total mass of the compounds (sedimentation velocity, light scattering), or on the  $z$ -value (ultracentrifugation). The differences among these values increase with polydispersity.

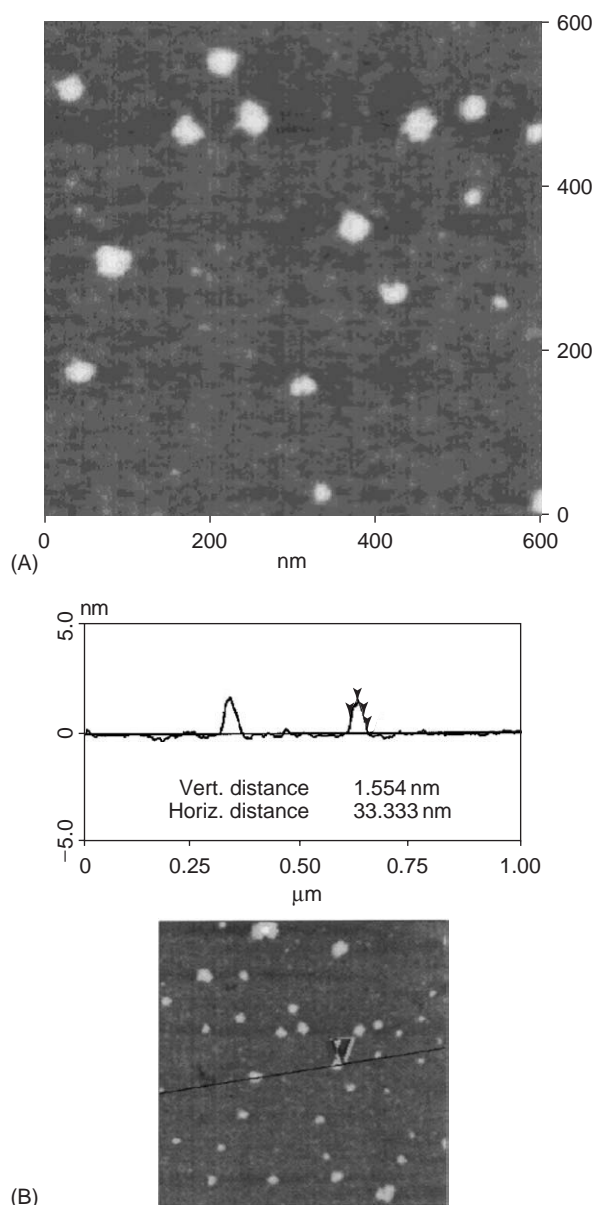
Humic acid colloids are mass fractals in aqueous solution. Their fractal dimension varies as a function of metal content, organic functional group content, and the separation technique applied in the original extractions. The range of fractal dimensions found – from 1.3 to 2.8 – suggests that their formation processes extend beyond usual aggregation descriptions (diffusion limited, reaction limited aggregation regimes).

## Characterization: Chemical Composition and Structure

### Elemental Analysis

Elemental analysis includes carbon, hydrogen, oxygen, nitrogen, phosphorus, sulfur, halogen, and elemental ratios C/H, O/C, and C/N. Typical elemental composition of water and soil humic substances are shown in Table 1. Carbon is the major element and it





**Figure 4** (A) AFM image of isolated Suwannee River humic acid macromolecules ( $10 \text{ mg l}^{-1}$ ;  $50 \text{ mmol l}^{-1}$  NaCl, pH 5.5) adsorbed on mica. Scan size is  $600 \text{ nm} \times 600 \text{ nm}$ . (B) Cross-sectional analysis of the image shown at the right-hand side. The height of the measured particles is  $1.6 \text{ nm}$ , and the lateral dimension at half-height (overestimated) is  $33 \text{ nm}$ . (Reprinted with permission from Balnois E, Wilkinson KJ, Lead JR, and Buffle J (1999) *Environmental Science and Technology* 33: 3911–3917; © American Chemical Society.)

varies between 50% and 60% for humic acids and between 40% and 50% for fulvic acids. There are very little differences in H, N, and S contents. The ash content of humic substances depends on the isolation procedure. Ash contents of less than 1% are found if the XAD resin method is used for isolation.

### Isotopic Ratio

The isotopic ratio provides information on the origin of humic substances. During biochemical and chemical reactions, the isotopes of a given element undergo fractionation, which depends not only on their mass, but also on the nature of the reaction. Isotopic ratios, in particular those of  $^{13}\text{C}/^{12}\text{C}$ , depend on the organic matter evolution and enable the origin of the organic matter to be specified to some degree. The humic substances (primarily as fulvic acids) in deep ocean waters have been established to be 3000–6000 years BP with a  $\delta^{13}\text{C}$  of  $-22$  to  $-23$  ppm. All stream humic substances analyzed to date have  $^{14}\text{C}$  ages near zero (all modern carbon) with  $\delta^{13}\text{C}$  values between  $-25$  and  $-28$  ppm. The  $^{14}\text{C}$  and  $\delta^{13}\text{C}$  values for soil humic substances vary considerably. The  $^{14}\text{C}$  age of humic acids in soils varies from a few hundred years to several thousand years;  $\delta^{13}\text{C}$  values range from  $-25$  to  $-31$  ppm. Such range for soils may reflect cropping practices and residue management, the methods used for isolation of fulvic and humic acids and sampling techniques rather than true pedogenic differences.

### Functional Groups

Table 1 shows the functional groups in water and soil humic and fulvic fractions. The major functional groups are carboxyl, phenolic hydroxyl, hydroxyl, and carbonyl. The sum of carboxyl and phenolic hydroxyl is total acidity. Total acidity of water fulvic acids is slightly higher compared to soil fulvic acids while total acidity of soil and water humic acids is comparable.

Carboxyl group content is determined either by potentiometric titration of humic material or by calcium acetate titration. Calcium acetate method overestimates carboxyl content by at least 10%. Potentiometric titration of humic substances (pH 2.8) in  $0.1 \text{ mol l}^{-1}$  KCl with a base up to pH 10 yields more precise values for carboxyl content.  $^{13}\text{C}$  nuclear magnetic resonance (NMR) (solid and liquid) can also be used for carboxyl determination. The results obtained are comparable to potentiometric titration.

Phenolic hydroxyl content is more difficult to determine by potentiometric titration owing to the weak acidic nature of phenolic hydroxyl. Phenolic hydroxyl content can, therefore, be only estimated. Again  $^{13}\text{C}$  NMR can be used for estimation of phenolic hydroxyl content. Comparable values are obtained using both these methods, e.g.,  $1\text{--}2 \text{ meq g}^{-1}$  for water humic substances. This is less than phenolic hydroxyl content of soil humic substances, which is  $3.0\text{--}3.9 \text{ meq g}^{-1}$ . Phenolic contents of water humic

**Table 1** Composition of soil and water humic and fulvic acids

<i>Parameter</i>	<i>Soil humic acid</i>	<i>Water humic acid</i>	<i>Soil fulvic acid</i>	<i>Water fulvic acid</i>
<i>Elemental composition (%)</i>				
C	53.8–58.7	53.4–59.4	40.7–53.1	41.6–51.1
O	32.7–38.3	44.8	43.1–49.8	43.5–51.6
H	3.2–6.2	3.1–5.1	3.2–7.0	2.7–4.2
N	0.8–5.5	1.9–2.0	0.9–3.3	1.0–2.2
S	0.1–1.5		0.1–3.6	<0.2
<i>Functional group content (meq g<sup>-1</sup>)</i>				
Total acidity	5.6–8.9	7.1–8.9	6.4–14.2	9.6–16.6
–COOH	1.5–5.7	4.5–5.9	6.1–11.2	6.5–10.7
Phenolic OH	2.1–5.7	2.2–3.7	1.2–5.7	1.0–4.7
Alcoholic OH	0.2–4.9		2.6–9.5	
>C=O quinones	1.4–2.6		0.3–2.0	
>C=O ketones	0.3–1.7	4.3–5.1	1.6–2.7	4.3–7.4
–O–CH <sub>3</sub>	0–0.8		0–1.2	
Free radicals (10 <sup>17</sup> spins g <sup>-1</sup> ) (5<pH<10)	8–37.0		1.5–3.0	
<i>Structural composition (%)</i>				
Aromaticity	69		71	
%H: aromatic	16–43		20	
%C: aromatic	21–35		41	20.8
Carboxylic	10–30			14.6
(Aromatic + carboxylic)	65–72	41	53	
<i>Fraction of total OH (%)</i>				
Aliphatic COOH	9			15
Aromatic COOH	20			24
Phenolic OH	14			12
(aliphatic + carbohydrates OH)	57			49

substances are considerably less than soil humic substances. This is an important difference between soil and aquatic humus.

### Analysis of the Chemical Structure

The methods that have been used for obtaining information on the chemical structure can be divided into two groups, viz., degradative and nondegradative methods.

**Degradative methods** Degradative methods include oxidative and reductive degradation, hydrolysis, various types of irradiations, pyrolysis, biological degradation. The oxidative degradation method is one of the most useful methods for obtaining information on chemical structure. Several reagents have been used for oxidation of humic substances: under alkaline conditions, potassium permanganate, copper(II) oxide, aqueous chlorine; under acidic conditions, nitric acid, peracetic acid, hydrogen peroxide; and ultraviolet (UV) radiation. Alkaline permanganate oxidation seems to be the best method, from the standpoints of producing relatively large amounts of identifiable digest products and providing significant information on the chemical

structures of humic substances. Premethylation is important for obtaining good yields. In all oxidative degradations, soil substances essentially produce carboxylic acids (mono-, di-, tri-, and tetra-), phenolic acids, and benzenecarboxylic acids. There are indications that whereas soil humic acids and soil fulvic acids yield similar quantities of aliphatic compounds on oxidation, the oxidation of soil fulvic acids yields more phenolic acids but fewer benzenecarboxylic acids than does that of soil humic acids.

The most widely used reductive degradation methods are zinc distillation and sodium amalgam reduction. Fused aromatic structures (methyl substituted naphthalene, anthracene, pyrene, and perylene) are the major digest products of this reduction of humic substances.

Nowadays, analytical pyrolysis is widely used. Recently, there have been significant improvements in the pyrolysis procedure and, by coupling pyrolysis to gas chromatography (GC) and mass spectrometry (MS) a significant body of structural information on humic substances may be obtained. However, quantitation of pyrolysis GC–MS is limited by matrix effects that distorts the signals of detected and the high temperatures used during pyrolysis produce

secondary reactions (rearrangement, cracking, hydrogenation, and polymerization). Therefore conclusions regarding structural make-up of the original samples must be drawn with caution. Another modern method is thermo-chemo analysis with tetramethyl ammonium hydroxide coupled to mass spectrometry. Its advantage is that numerous polar functionalities associated with humic substances, which are not detected by other methods, can be detected, e.g., propyl side-chains.

In the last decade there has been a fast progress in the field of mass spectrometry employing soft ionization techniques such as electrospray ionization mass spectrometry (ESI-MS), Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS), and matrix-assisted laser desorption ionization mass spectrometry (MALDI)-MS. These tools provide molecular level information on humic substances.

**Nondegradative methods** The nondegradative methods used for obtaining information on chemical structure and properties include infrared (IR) and UV-visible spectroscopy, spectrofluorimetry, electron spin resonance (ESR) spectroscopy, and NMR. The study of humic substances is the study of complicated, ill-defined mixtures. Most of the physical and chemical methods of investigation are confined to simple systems and the analyst's ability to interpret the data is extremely limited. In applying any spectroscopic method to humic substances, measurement is taken of the summation of the signals of the numerous components in the mixture that respond to that particular frequency range, with all signals superimposed upon each other. What is measured is the net or average response of these particular components in the assemblage. Therefore, in the case of humic substances, there is little point in seeking definite structure but instead have a knowledge of the types of structures in the mixtures and the types of association structure that these can form.

**Infrared spectroscopy** IR and Fourier transform infrared (FTIR) spectra of humic substances show bands at  $3400\text{ cm}^{-1}$  (H bonding OH),  $2990\text{ cm}^{-1}$  (aliphatic C-H),  $1725\text{ cm}^{-1}$  (C=O of  $\text{CO}_2\text{H}$ , C=O of ketone),  $1630\text{ cm}^{-1}$  (aromatic C=C, C=O of carbonyl,  $\text{COO}^-$  or quinone),  $1450\text{ cm}^{-1}$  (aliphatic C-H),  $1400\text{ cm}^{-1}$  ( $\text{COO}^-$ ) and  $1200\text{ cm}^{-1}$  (C-O or OH of  $\text{CO}_2\text{H}$ ). The bands are usually broad due to overlapping of individual absorbances. While IR and FTIR provide worthwhile information about functional groups, they reveal little about the chemical structure of humic substances. FTIR and diffuse reflectance infrared Fourier transform (DRIFT) are the techniques most widely used.

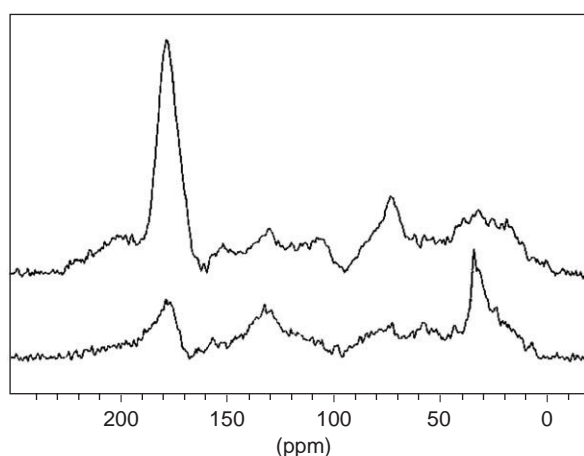
**Ultraviolet-visible spectroscopy** UV-visible spectra of humic substances are generally featureless. The technique has little value for studying functionality of humic substances and cannot be used for the direct determination of functional groups in these materials. Nevertheless, UV-visible spectroscopy does have other useful applications such as the estimation of the degree of humification using  $E4/E6$  ratios (the ratio of the absorbance at 465 nm to that at 665 nm) and for determining the concentration of dissolved humic substances based on Beer's law plots. The ratio UV absorption at 280 nm/dissolved organic carbon (DOC) allows to estimate the contribution of the two major groups of dissolved NOM (pedogenic versus aquagenic, see Figure 1) on the basis of their chemical structure.

**Spectrofluorimetry** Fluorescence measurements of humic substances can provide criteria for the differentiation and classification of substances of various origins. Humic substance groups are excited at  $\sim 350\text{ nm}$  and emit at  $\sim 450\text{ nm}$ , a spectral region relatively free from interference by other groups. Measurement of the fluorescence is a relatively selective method of analysis. Fluorescence cannot be used for the direct determination of functionality.

**Electron spin resonance spectroscopy** The ESR spectrum of humic and fulvic acids consists of a single line identified by its position and width. In general, the ESR spectra are devoid of hyperfine splitting. The ESR spectra of humic substances result from only a small fraction of the total number of molecules that comprise the humic and fulvic acids. The number of free radicals per unit weight, the  $g$  value and the line width can be calculated. The predominant radicals are semiquinones. Comparison of the ESR properties among humic fractions reveals that fulvic acids contain greater quantities of free radicals than humic acids.

**Nuclear magnetic resonance spectroscopy** Nuclear magnetic resonance spectroscopy has made a significant contribution to our knowledge of the structure of humic substances. Samples can be analyzed either by solid state or liquid state NMR. Typical  $^{13}\text{C}$  NMR spectra of solid state and liquid state of soil humic substances show that chemical shifts are mostly the same for liquid and solid state. The most popular  $^{13}\text{C}$  NMR technique for the study of humic materials and coal is the combination of cross-polarization (CP) and magic angle spinning (MAS) on solids. Scores of papers have been published on the application of CP-MAS  $^{13}\text{C}$  NMR to the study of humic samples these last years. The advantages of the solid state over the liquid state is that there is no concentration

limit and no solubility concerns, the solid state is more stable over time than solutions, and there is less sample handling involved. The CP-MAS technique also has the advantage of giving better signal in less accumulation time. It is now possible to obtain good relaxation parameters for humic materials within a reasonable spectral acquisition time (2–4 days). However, evidence has grown that indicated that a substantial fraction of the carbon content indicated by classic elemental analysis is missed in some samples by CP-MAS  $^{13}\text{C}$  methods. Typical CP-MAS  $^{13}\text{C}$  NMR spectra of Laurentian humic and fulvic acids are shown in Figure 5. Table 2 shows the corresponding chemical shift assignments.



**Figure 5** Ramp-CP-MAS  $^{13}\text{C}$  NMR spectra of Laurentian fulvic acid (top) and humic acid (bottom). (Reprinted with permission from Cook RL and Langford CH (1998) *Environmental Science and Technology* 32: 719–725; © American Chemical Society.)

Recently, liquid and solid state  $^{15}\text{N}$  NMR has become a powerful tool for determination of chemical structure of humic substances. The  $^{15}\text{N}$  NMR spectra of humic substances show evidences of the presence of quinones, ketones, and esters. Also it shows that most of the nitrogen occurs in amide forms.

Newly developing multidimensional NMR might become a powerful tool in the future for obtaining valuable information on the chemical structure of humic substances. Two-dimensional phase-sensitive  $^{13}\text{C}$ ,  $^1\text{H}$  correlation spectra has also been successfully applied for substructure elucidation in size-fractionated humic substances.

## Complexation Properties

The evaluation of interactions between trace metals and humic substances has received considerable attention. Evidence for strong complexation has been demonstrated under controlled experimental conditions. However, the complexity of fulvic and humic compounds is often at odds with the amount of information contained in the experimental data, thus making it very difficult to obtain a rigorous thermodynamic description of their complexing properties.

When the complexation reaction is studied by titration with a metal ion, the concentration of the free metal ion is monitored typically by an electrochemical method, such as potentiometry with an ion-selective electrode, differential pulse polarography, or anodic stripping voltammetry. Fluorescence spectroscopy is being increasingly used to study the reaction of humic material with paramagnetic metal

**Table 2** NMR chemical shifts of Laurentian fulvic acid (top) and humic acid (bottom)

Chemical shift assignments	Chemical shift regions (ppm)	% TOC	$T_{1\rho}$ $^1\text{H}$ (ms)	$T_2$ $^{13}\text{C}$ (ms)
<i>Laurentian fulvic acid</i>				
Ketonic	220–190	8.8	7.3	11.9
Carboxyl	190–162	33.8	4.9	13.1
Phenolic	162–145	2.2	6.0	22.7
Aromatic	145–108	12.0	5.5	11.4
O–C–O	108–96	3.6	3.6	10.4
Carbohydrate	96–50	17.8	2.6	12.3
Aliphatic	50–0	21.8	2.8	8.0
<i>Laurentian humic acid</i>				
Ketonic	220–190	6.5	2.4	7.8
Carboxyl	190–162	15.7	3.2	8.3
Phenolic	162–145	2.8	4.6	9.1
Aromatic	145–108	18.7	4.1	7.8
O–C–O	108–96	1.0	3.6	5.1
Carbohydrate 1	96–60	11.9	2.3	6.0
Carbohydrate 2	60–50	9.1	2.8	5.1
Aliphatic	50–0	34.3	3.0	6.9

Reprinted with permission from Cook RL and Langford CH (1998) *Environmental Science and Technology* 32: 719–725; © American Chemical Society.



ions that quench the fluorescence of humic matter. Indirect methods based on differences in size such as equilibrium dialysis, ultrafiltration, and size exclusion (gel) chromatography have also been applied. Some of the main experimental difficulties encountered are:

1. Determination of the molar concentration of total complexing sites. The molar complexation of total complexing sites is most often not measurable unambiguously because the relative molecular mass of the complexing entity is generally not clearly defined. A surrogate concentration expressed, for example, in grams of carbon per liter must be used.
2. Limited analytical window. The analytical windows of the different techniques available are smaller than that required to 'cover' all the complexing sites of a humic substance. Moreover, average equilibrium constants calculated by using different techniques will give noncomparable values.
3. Data need to be collected under conditions of metal concentration as close as possible to those existing in natural environments. Use of unrealistic high metal loading conditions and extrapolations of the results so obtained to natural conditions cannot give realistic complexation data.

Humic and fulvic compounds exhibit a wide range of free energy for metal complexation since they possess two major characteristic properties: (1) polyfunctionality: each molecule possesses many complexing sites of different chemical nature; (2) polyelectrolytic nature: all complexant molecules bear high electric charge densities due to the presence of a large number of dissociable functional groups per physical entity. Complexation properties have been represented by means of average parameters, usually one or more site concentrations and stability 'constant' couples corresponding to one or more site types, but the use of such average parameters should be avoided. They have no physicochemical meaning since (1) they are not constant over the entire titration as the free metal ion concentration is successively controlled by sites of different nature and (2) the measured values depend on the analytical window of the method used. They represent only average equilibrium quotients or overall site concentrations for a range of site types having widely different affinities. More sophisticated interpretation methods have been developed, e.g., application of semiempirical complexation isotherms or *a priori* affinity spectra. All these methods have in common a conceptual approach similar to that used in the study of metal complexation by simple ligands: a number

of assumptions are made concerning the nature of the reactions taking place followed by the fitting of the resulting mathematical model to the experimental data. For major sites such a 'classical' description of the complexing behavior may be applicable, provided the necessary corrections to take into account the influence of physical factors such as changes in electric charge density or particle conformation are made, but for minor sites, the complexity of humic compounds is much too great to be unambiguously represented by any model based on *a priori* assumptions that include all the necessary descriptive parameters. In order to overcome these difficulties, two interpretation models that tend to minimize the imposition of *a priori* assumptions have been proposed: the site affinity distribution function and the differential equilibrium function. None of these models is entirely satisfactory, but the conceptual approach that is embodied in them is definitely more consistent with the complexity of humic substances.

Interpretation of binding by noncompletely labile systems, such as it is often the case with humic and fulvic compounds, and assessment of their relationship to biogeochemical processes require combined equilibrium and kinetic approaches. In some cases, kinetics may even predominate over equilibrium considerations. The importance of kinetic studies is increasingly being realized and within the past few years many studies have been reported. Techniques employed include fluorescence quenching, ligand exchange with simple complexants, thermal lens spectrometry, isotope exchange, complexing resins, ion-exchange resins, and ion exchange coupled with voltammetric detection. Site heterogeneity will influence both the equilibrium and kinetic properties of the complexes.

**See also:** Dioxins. **Electrophoresis:** Proteins. **Ethanol.** **Flow Injection Analysis:** Instrumentation. **Geochemistry:** Inorganic. **Quality Assurance:** Traceability. **Voltammetry:** Inorganic Compounds. **X-Ray Absorption and Diffraction:** X-Ray Diffraction – Powder.

## Further Reading

- Aiken GR, McKnight DM, Wershaw RL, and MacCarthy P (eds.) (1985) *Humic Substances in Soil, Sediment, and Water. Geochemistry, Isolation, and Characterization*. New York: Wiley-Interscience.
- Buffle J (1989) *Complexation Reactions in Aquatic Systems. An Analytical Approach*. Chichester: Ellis Horwood.
- Clapp CE, Hayes MHB, Senesi N, Bloom PR, and Jardine PM (eds.) (2001) *Humic Substances and Chemical Contaminants*. Madison, WI: Soil Science Society of America.



- Frimmel FH, Abbt-Braun G, Heumann KG *et al.* (eds.) (2002) *Refractory Organic Substances in the Environment*. Weinheim: Wiley-VCH.
- Gaffney JS, Marley NA, and Clark SB (eds.) (1996) *Humic and Fulvic Acids* (ACS Symposium Series, vol. 651). Washington DC: American Chemical Society.
- Ghabbour EA and Davies G (eds.) (2001) *Humic Substances: Structures, Models, and Functions*. Cambridge: Royal Society of Chemistry.
- Hansell D and Carlson C (2002) *Biogeochemistry of Marine Dissolved Organic Matter*. New York: Academic Press.
- Hayes MHB, MacCarthy P, Malcolm RL, and Swift RS (eds.) (1989) *Humic Substances II. In Search of Structure*. Chichester: Wiley-Interscience.
- Perdue EM and Gjessing ET (eds.) (1990) *Organic Acids in Aquatic Ecosystems*. Chichester: Wiley-Interscience.
- Senesi N and Miano TM (eds.) (1994) *Humic Substances in the Global Environment and Implications for Human Health*. Amsterdam: Elsevier.
- Stevenson FJ (1994) *Humus Chemistry*, 2nd edn. New York: Wiley.
- Suffet IH and MacCarthy P (eds.) (1988) *Aquatic Humic Substances* (Advances in Chemistry Series, vol. 219). Washington, DC: American Chemical Society.
- Tan KH (2003) *Humic Matter in Soil and the Environment. Principles and Controversies*. New York: Dekker.

## HYDRIDE GENERATION

See **ATOMIC ABSORPTION SPECTROMETRY: Vapor Generation**



## ICP

See **ATOMIC EMISSION SPECTROMETRY: Inductively Coupled Plasma**

## ICP-MS

See **ATOMIC MASS SPECTROMETRY: Inductively Coupled Plasma**

## IMMOBILIZED ENZYMES

See **ENZYMES: Immobilized Enzymes**

## IMMUNOASSAYS

Contents

**Overview**

**Production of Antibodies**

### Overview

**R M Twyman**, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Immunoassays are analytical methods that achieve the detection and quantitation of analytes in biological or environmental samples through the formation of a stable complex between the analyte and a specific antibody. By labeling the analyte or the antibody, or another constituent of the assay, complex formation results in a measurable change in the distribution and/or activity of the label, allowing very

sensitive measurements to be obtained. Immunoassays are simple, rapid, inexpensive, and versatile, and they are very amenable to automation and multiplexing. Therefore, they play an important role in many technological areas, including analytical chemistry, pharmacology, molecular cell biology, and clinical biochemistry. They have become standard practice in many laboratories around the world, and a large number of companies market diagnostic or detection kits that use immunoassays to measure the levels of particular analytes of interest, such as pollutants, hormones, disease markers, and pathogens. In the last few years, both solution-phase and solid-phase immunoassays have become increasingly sensitive and sophisticated, culminating in the development of biochips with arrayed antibody microdots that can be used for the highly multiplexed analysis of hundreds or even thousands of individual analytes

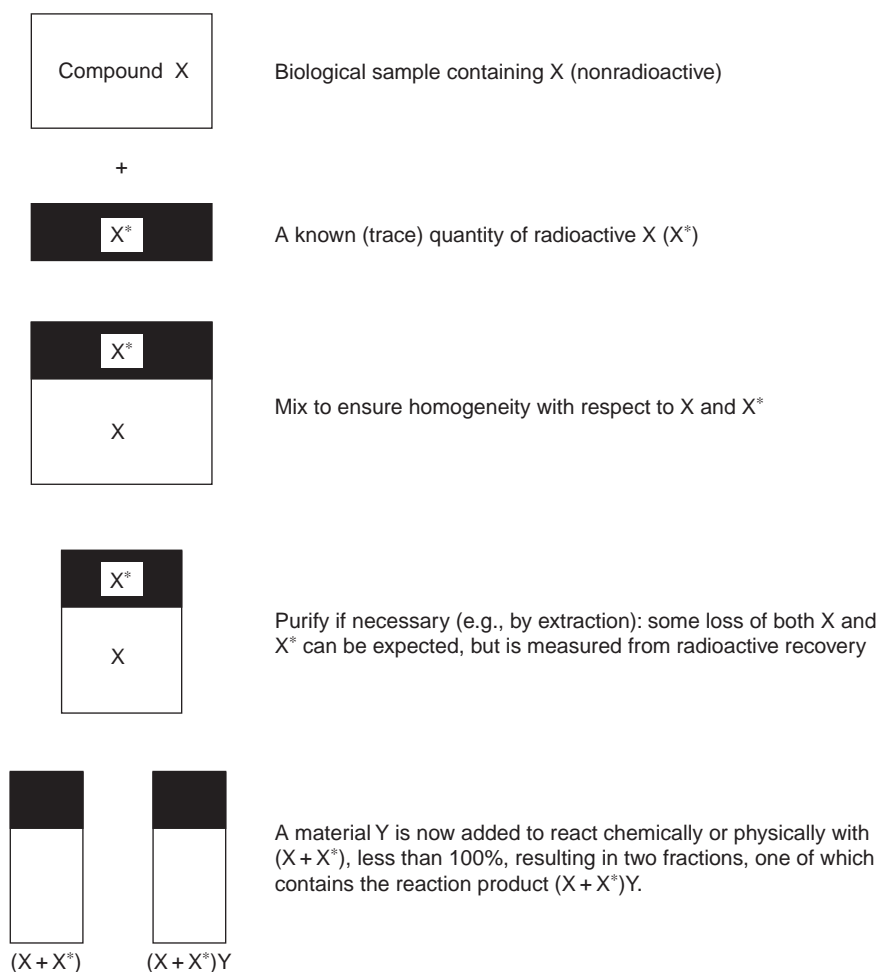
in a complex sample. This article presents a general overview of the principles of the immunoassay and describes some of the more widely used variations of the technique. Subsequent articles deal with methods for the production of antibodies, and specific applications.

## Radioimmunoassays

The first immunoassay technique was the isotopic dilution radioimmunoassay, developed by Rosalyn Yalow and Solomon Berson for the detection of insulin in human blood. This work gained Yalow a share of the 1977 Nobel Prize for Physiology and Medicine, but Berson was not honored because of Nobel rules stating that the prize cannot be awarded posthumously. The following passage, translated from the presentation speech, explains the technique in simple terms: "...As a result of mixing in a test tube a known quantity of radioactive insulin with a known quantity of antibodies against insulin, a

specific amount of the insulin becomes attached to these antibodies. Subsequently, if one adds to this mixture a small amount of blood which contains insulin, the insulin from the blood becomes similarly attached to the antibodies and a certain portion of the radioactive insulin is detached from the antibodies. The higher the concentration of insulin is in the blood sample, the larger is the amount of radioactive insulin that will be detached from the antibodies. The amount of radioactive insulin thus removed can easily be determined, providing an exact measure of the amount of insulin present in the blood sample..."

The isotopic dilution radioimmunoassay is based on the principle of saturation analysis, which is outlined in **Figure 1**. A test compound (in this case insulin, represented by X) is quantified by its ability to progressively saturate a suitable reagent (in this case an antibody specific for insulin, represented by Y). In the first step, a known quantity of radioactively labeled tracer insulin ( $X^*$ ) is added to the sample,



**Figure 1** Principles of saturation analysis. Y is typically an antibody or other protein, or charcoal; it alters the properties of X. The proportional distribution indicates the extent of reaction between X and Y, which is related to X in standards and samples.

which increases the total amount of insulin to  $(X + X^*)$ . Antibody Y is then added to the sample. The amount of Y added must be lower, in molar terms, than  $(X + X^*)$  so that the analyte and tracer are not saturated, and can compete for a limited number of Y sites. Also, it must be possible to separate  $(X + X^*)$  from  $(XY + X^*Y)$ , so that the extent of the complex formation can be determined by the proportional change in the distribution of radioactivity between  $(X + X^*)$  and  $(XY + X^*Y)$  at the beginning and end of the assay. Thus, the first radioimmunoassays were often classified according to how this separation was achieved (e.g., using polyethylene glycol or a second antibody) and were generally referred to as heterogeneous assays because separation into two phases was necessary. In general terms, the more analyte X present in the sample, the more  $X^*$  would be displaced from Y and the more of the radioactive label would remain in the unbound fraction  $(X + X^*)$ . A calibration graph is required, showing the distribution of radioactivity against known quantities of X in standards, allowing unknown quantities of X in experimental samples to be assessed in terms of the distribution of radioactivity.

In the above immunoassay format, the analyte is the labeled entity and the assay is carried out in solution. The principle is similar to that of the competitive protein binding assay, which has been used for many years to characterize receptor–ligand and enzyme–substrate interactions. The intrinsic advantages of the radioimmunoassay over other forms of competitive assay are the specificity of the antibody–analyte interaction and the general applicability of antibodies as opposed to other interacting reagents. Antibodies can be raised against any immunogenic molecule, and even small molecules that do not induce an immune response can be attached to carrier molecules as haptens to render them immunogenic. However, the necessity to label the analyte in each type of experiment and the need to use infinitesimally small concentrations of reagents often causes technical problems and it is impossible to measure more than one analyte simultaneously.

The next significant development was the immunoradiometric assay (IRMA), in which the isotopic label is attached to the antibody rather than the analyte, thus resolving problems associated with labeling diverse analytes. This was the first example of a sandwich immunoassay, in which two antibodies are used firstly to capture the analyte and then to detect and quantify it. It was also the first example of a solid-phase heterogeneous immunoassay, i.e., the first (capture) antibody is immobilized on a solid substrate such as the surface of a microtiter dish or membrane. The surface is then flooded with the

sample, resulting in the capture of the analyte and the formation of antibody–analyte complexes. After washing to remove unbound components of the sample, the surface is flooded with a probe antibody labeled, e.g., with  $^{125}\text{I}$ . A further washing step then removes the unbound probe antibody and the surface is dried and exposed to a film or placed in a gamma counter. Control samples spiked with different amounts of the analyte can be used to produce a calibration graph, allowing the concentration of the analyte in the experimental sample to be determined.

The great advantage of this method is that quantitative data are generated by measuring a stoichiometric interaction between labeled antibody (present in excess) and the captured analyte. The signal produced by the antibody is thus directly proportional to the concentration of the analyte, rather than being related to it in a complex manner as is the case for competition analysis. The limitation of this method is that two antibodies recognizing different epitopes on the target analyte are required. Therefore, while it is suitable for the detection of large molecules such as proteins, it cannot be used to detect small organic molecules, which often fit completely within the antigen-binding pocket of the antibody.

## Nonisotopic Immunoassays

Radioisotopes are convenient labels with high specific activity, and they continue to be used in many laboratories. However, due to potential health hazards, the expense of radioisotopes and the apparatus used to detect them, the short half-life of  $^{125}\text{I}$ , and issues surrounding the storage and disposal of low-level radioactive waste, there has been considerable interest in alternative labeling strategies. Another significant development in the history of immunoassays is therefore the increased use of nonisotopic labels, which include enzymes that catalyze colorimetric or chemiluminescent reactions, and fluorogenic compounds. Many of these labels incorporate intrinsic amplification strategies to increase signal intensities, or can be coupled to systems that provide extraneous signal amplification. Some of these immunoassay formats are discussed in more detail below.

### Homogeneous Competition Assays with Nonisotopic Labels

Although radioactive labels are unaffected by antibody–analyte interactions, the performance of other types of labels can be made to depend on either the formation or disruption of such a complex. This removes the necessity to separate the bound and unbound fractions in competition assays since the label

would only be active in one of the fractions anyway. Nonisotopic immunoassays that are carried out entirely in solution and do not require a separation stage are described as homogenous, and are advantageous because of the reduced number of steps required and hence the lower likelihood of operator error.

**Enzyme-multiplied immunoassay technique** Perhaps the best known homogeneous assay format is the enzyme-multiplied immunoassay technique (EMIT), in which the analyte is covalently attached to an enzyme, and the formation of an analyte–antibody complex blocks the active site and inhibits enzyme activity. When this blocked enzyme is mixed with the experimental sample, there is competition between the enzyme-linked analyte and the sample analyte for occupation of the antibody's antigen-binding site. The more of the analyte present in the sample, the more of the enzyme is released from inhibition, and the level of enzyme activity can thus be used to determine the quantity of the analyte.

A variation of the EMIT is the apoenzyme reconstruction immunoassay system, where the analyte is coupled not to the enzyme itself but to a cofactor or prosthetic group that is required for enzyme activity. For example, the analyte may be coupled to flavin-adenine dinucleotide (FAD), which is required for glucose oxidase activity. The antibody binds to the FAD–analyte conjugate and prevents FAD from interacting with the enzyme, whose activity is consequently reduced. The more analyte present in the sample, the more competition there is for the antibody and the more FAD–analyte conjugate remains free in solution and available to the enzyme. Again, the level of enzyme activity is related to the quantity of analyte in the sample.

**Fluorescence immunoassays** Fluorescent labels can also be used in homogeneous assay formats, and several different types of detection system can be employed. Direct fluorescence is a straightforward approach, which requires the formation or disruption of the analyte–antibody complex to generate the label. This is generally achieved by coupling the analyte to an inert fluorophore via an enzyme-cleavable bridge, which is rendered inactive by antibody binding. Perhaps the most common homogeneous, direct fluorescence-based immunoassay format is the substrate-labeled fluorescence immunoassay. The analyte is conjugated to a fluorogenic substrate of the enzyme  $\beta$ -galactosidase (4-methylumbelliferyl- $\beta$ -D-galactoside, more commonly known by its acronym MUG). The enzyme can cleave this substrate and release the fluorophore, methylumbelliferone, as long

as the analyte is not associated with its corresponding antibody. The more analyte present in the sample, the more MUG is free in solution and available to the enzyme, and the greater the direct fluorescent signal that is produced. When the fluorophore is excited by incident light at a particular excitation wavelength, energy is absorbed by electrons promoting them to an unstable higher energy state. When the electrons revert to their ground state, the energy is released as visible light with a characteristic emission wavelength that can be detected by a photomultiplier or charge-coupled device.

An alternative homogeneous fluorescence-based detection system is fluorescence resonance energy transfer (FRET). This phenomenon occurs when two fluorophores are in close proximity, and the donor fluorophore has an emission spectrum that overlaps the excitation spectrum of the acceptor fluorophore. When the donor fluorophore is excited, energy is transferred from donor to acceptor with the result that the intensity of emission from the donor is reduced (quenched). If both the analyte and antibody are coupled to fluorophores with overlapping spectra, then FRET will occur only when the complex forms. Thus, FRET has not been applied widely in conventional immunoassays.

A further fluorescence effect, fluorescence polarization (FP), has recently been employed in immunoassays. FP occurs when a fluorophore is excited by plane-polarized light and emits fluorescence that is also polarized. Complete FP occurs only when the fluorophore is stationary. The degree of observed FP is therefore dependent on how fast a molecule tumbles in solution, a property related to its molecular mass. Changes in molecular mass, for example, caused by antibody–analyte complex formation, can therefore be detected by changes in FP as long as the temperature, viscosity, and other conditions remain constant. This is the principle of the fluorescence polarization immunoassay, in which the analyte disrupts a tracer complex introduced into the sample, reducing the degree of FP as a function of analyte concentration.

**Other homogeneous immunoassays** Several further homogeneous assay formats have been developed that capitalize on the ability of antigen–antibody complexes to form detectable clusters or networks. Examples include the latex particle agglutination assay and the latex particle agglutination inhibition assay, in which the analyte or the antibody is conjugated to latex beads, and the analyte is quantified by virtue of its ability to promote or disrupt agglutination. Another example is the liposome immunoassay, in which analyte molecules are coupled to lipids,



**Table 1** Summary of ELISA assay formats and requirements

<i>ELISA protocol</i>	<i>Uses</i>	<i>Reagents required</i>	<i>Notes</i>
Indirect	Antibody screening	Antigen (pure or semipure)	Does not require preexisting specific antibodies
	Epitope mapping	Test solution containing antibody Enzyme conjugate that binds Ig of immunized species	Requires relatively large amounts of antigen
Direct competition	Antigen screening	Antigen (pure or semipure)	Rapid assay with only two steps
	Detect soluble antigen	Test solution containing antigen Enzyme–antibody conjugate specific for antigen	Excellent for measuring antigenic cross-reactivity
Antibody sandwich	Antigen screening	Capture antibody (purified or semipurified)	Most sensitive antigen assay
	Detect soluble antigen	Test solution containing antigen Enzyme–antibody conjugate specific for antigen	Requires relatively large amounts of pure or semipure specific capture antibody
Double sandwich	Antibody screening	Capture antibody (specific for Ig of immunized species)	Does not require purified antigen
	Epitope mapping	Test solution containing antigen Enzyme–antibody conjugate specific for antigen	Relatively long assay with five steps
Direct cellular	Screen cells for antigen	Cells expressing antigen of interest	Sensitive assay for bulk screening
	Measures antigen levels	Enzyme–antibody conjugate specific for cellular antigen	Insensitive to heterogeneity of expression in mixed cell population
Indirect cellular	Screen for antibodies	Cells for immunization	May not detect antibodies specific for cellular antigens expressed at a low density
	against cellular antigen	Test solution containing antibody Enzyme conjugate that binds Ig of immunized species	

allowing them to form an artificial plasma membrane or liposome. The liposome is used to compartmentalize an enzyme that is separated from its substrate. Addition of the antibody causes the liposome to be disrupted and the enzyme released. As in other competition assays, the more analyte present in the sample, the less of the tracer will be sequestered into antibody–analyte complexes, and the lower the resulting enzyme activity. Although many researchers have used this method, it suffers from the inherent disadvantage of liposome instability.

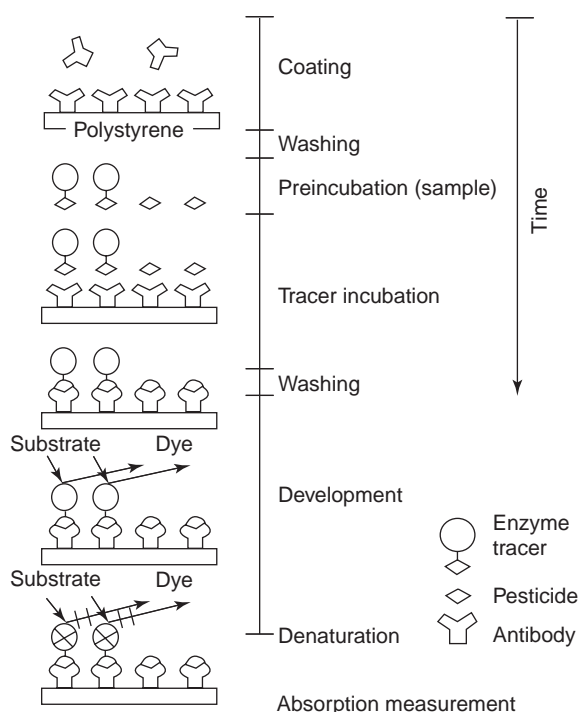
### Solid-Phase Assays with Nonisotopic Labels

Solid-Phase assays, which utilize a solid substrate to capture the reaction or reaction products, have also been adapted to work with nonisotopic labels. Either the antibody or the analyte can be immobilized on the solid phase, by adsorption or using a covalent bond. Both competition-type and immunometric-type assays can be performed using the solid-phase approach.

**Enzyme-linked immunosorbent assay (ELISA)** The most widely used heterogeneous solid-phase immunoassay format is the ELISA, which was first described in 1971. Usually, the antibody is immobilized on the solid phase, and is used to capture the unlabeled analyte from the sample. Six common variants of the technique are practiced, i.e., the direct competitive ELISA, the

sandwich ELISA, the double-sandwich ELISA, the indirect ELISA, the indirect cellular ELISA, and the direct cellular ELISA. The differences between these methods are summarized in **Table 1**.

The principle of the direct competitive ELISA is similar to that of the original radioimmunoassay except that the separation of bound and unbound labels is achieved by immobilizing the antibody (**Figure 2**). The substrate, often a microtiter dish, is first coated with analyte-binding antibody (in commercial kits the substrate usually comes ready-prepared). The sample is then added, allowing the analyte to bind to the antibody. A predefined amount of tracer is then added, which comprises a known concentration of enzyme-labeled analyte. Competition ensues between the sample analyte and the tracer because only limited numbers of antibodies are available. Following an incubation period, which may last several hours to allow the system to come to equilibrium, the excess reagents are washed away. The retained tracer population is inversely proportional to the amount of analyte in the sample, because the greater the amount of analyte, the fewer antibody sites will be occupied by the tracer. The amount of bound tracer is determined using the appropriate enzyme substrate reaction, which is usually colorimetric or chemiluminescent, and can therefore be quantified using a photometer or luminometer. More than half of ELISA tests employ horseradish peroxidase because of its high turnover number and the range of

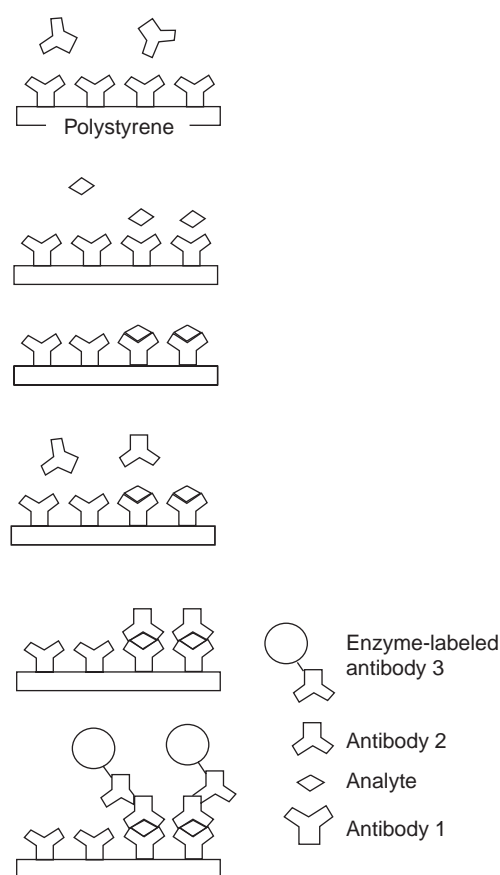


**Figure 2** Principle of the competitive ELISA. (Reproduced with permission from Wiley-VCH; © Wiley-VCH.)

sensitive detection systems available, while alkaline phosphatase is also very popular.

The sandwich immunoassay, which requires two analyte-specific antibodies, is the most popular ELISA variant used for the quantification of large molecules, such as proteins (**Figure 3**). This is a non-competitive assay similar in principle to the IRMA. Two antibodies are required that recognize different epitopes on the target analyte and they must be able to bind simultaneously. The first antibody is immobilized on the microtiter plate, which is flooded with the sample in order to capture the analyte. After washing, the second antibody is added, which binds to the captured analyte and provides a quantitative signal. In the IRMA, this second antibody emits a radioactive signal that provides a surrogate quantitation for the analyte. In the ELISA, the second antibody is usually unlabeled, but is detected by a third, enzyme-conjugated antibody, facilitating a further level of signal amplification.

A number of additional strategies for signal amplification in ELISAs have been developed. One example is the use of secondary antibodies conjugated to biotin, a molecule that has very strong affinity ( $>10^{15} \text{ l mol}^{-1}$ ) for the egg yolk protein avidin. After application of the secondary antibodies and washing to remove unbound reagents, the substrate is flooded with an excess of enzyme-conjugated avidin. Not only is the interaction between avidin and biotin

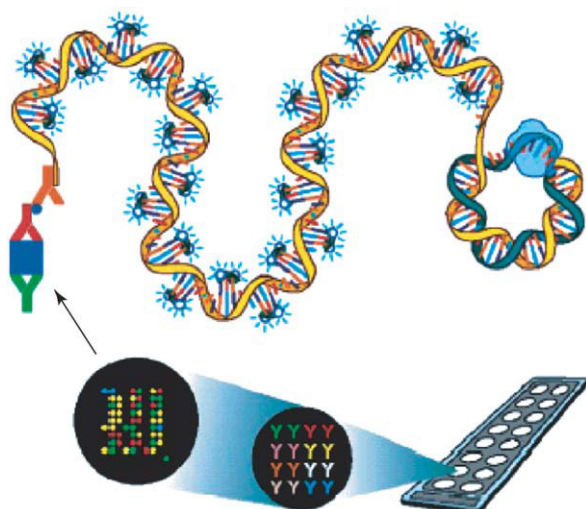


**Figure 3** Principle of the noncompetitive sandwich ELISA. (Reproduced with permission from Wiley-VCH; © Wiley-VCH.)

molecules highly specific and irreversible, but each avidin molecule has multiple biotin-binding sites. Therefore, further amplification can be achieved by adding enzyme-conjugated biotin to the assay.

Another variation of the sandwich assay is the immuno-RCA technique, which involves a tertiary level of detection by rolling circle amplification. The principle is that an analyte, captured by an immobilized antibody, is recognized by a second antibody in a sandwich assay as above, but the second antibody has an oligonucleotide covalently attached to it (**Figure 4**). In the presence of a circular DNA template, a strand-displacing DNA polymerase, and the four dNTPs, rolling circle amplification of the template occurs resulting in a long concatemer comprising hundreds of copies of the circle, which can be detected using a fluorescent or enzymatically labeled oligonucleotide probe. Solid-phase direct fluorescent-detection immunoassays are becoming increasingly popular with the development of analytical protein biochips (see below).

**Magnetic particle immunoassay** Another variant of the solid-phase immunoassay format depends on coupling antibodies to magnetic particles. The assay



**Figure 4** Sensitive protein detection using the RCA (rolling circle amplification) immunoassay chip and solid-phase direct fluorescence detection. The chip is divided into 16 teflon wells, each containing an array of 256 antibodies as probes. When a protein, represented by the blue square, is captured by one of the probes, it can be recognized using a second, biotinylated antibody (red), which is subsequently detected by a tertiary universal antibody connected to a circular oligonucleotide. A strand-displacing DNA polymerase can use this circular template, generating a long concatamer containing many fluorescent labels.

is carried out in solution, but the beads represent the solid phase. The sample is first mixed with a tracer, which comprises the analyte conjugated to horseradish peroxidase. After thorough mixing, the antibody-magnetic particle suspension is added and allowed to form complexes. The particles are then precipitated with a magnet, washed, and reprecipitated, and then resuspended in the enzyme substrate buffer. The depth of color, or the degree of chemiluminescence, is inversely proportional to the concentration of analyte in the sample, since the greater the concentration of the analyte, the less of the tracer remains attached to the antibody-magnetic particle complexes.

## Highly Multiplexed Immunoassays

Miniaturized, multiplexed, solid-phase immunoassays were developed in the 1980s using protein microdots spotted manually onto nitrocellulose sheets and other solid supports. It was shown quite clearly that this multianalyte immunoassay format was far more sensitive than standard immunoassays carried out in microtiter dishes because the sample volumes were so much smaller. Over the last few years, these devices have evolved into protein biochips, in which microdots of antibodies or protein analytes are arranged in a dense miniature grid on a solid support such as a microscope slide. Two types of

immunoassay-type protein biochip can be distinguished: antibody arrays, which are used predominantly for the quantitative analysis of proteins (but also other analytes), and antigen arrays, which are used to profile antibodies in serum and other fluids.

Developments in protein biochip technology have resulted in some novel detection methodologies for the quantification of analytes. There are three broad classes of detection methods. In the first method, the proteins in the sample are labeled universally, either with a radioisotope, or an enzymatic or fluorescent conjugate, and the signal is detected directly from the bound analyte molecules on the chip. In the second method, the proteins are not labeled, but a sandwich reaction is used to detect molecules bound to the array and the signal is produced by the labeled detection reagent (essentially the same as a standard sandwich ELISA or solid-phase fluorescent immunoassay, although highly multiplexed). Finally, several label-free detection methods can be used to detect and/or identify proteins bound to antibodies or protein analytes on the chip surface.

Although radiolabels and colorimetric assays can be used with low-density protein chips, fluorescent labels are safer, more convenient, and provide a better spatial resolution. The label can be incorporated directly into the analyte or into a secondary detection reagent that is applied to the chip after it has been washed to remove unbound proteins. There are advantages and disadvantages to both methods. The advantage of direct labeling is that protein detection and quantification can be carried out in a one-step reaction, and multiplex analysis is possible. A disadvantage of direct labeling is that not all proteins are labeled with the same efficiency, and the label itself can alter the structure of some proteins and interfere with their binding capabilities. These problems do not arise when a sandwich assay is used, so this approach may be preferable where accurate quantitation is needed. However, the main disadvantage of sandwich assays is the requirement for two antibodies recognizing different epitopes for each protein captured on the chip, a problem that increases in magnitude as the number of simultaneously analyzed proteins gets larger.

Label-free methods use the intrinsic properties of proteins to report binding events on protein chips. Ciphergen 'ProteinChips' can be incorporated directly into a matrix-assisted laser desorption/ionization (MALDI) source of a mass spectrometer, which can identify the proteins and provide quantitative analysis. The ionization of proteins bound to 'ProteinChips' is enhanced by the properties of the chip surface, leading to more uniform mass spectra than possible with standard MALDI mass spectrometric

analysis, a phenomenon described as surface-enhanced laser desorption/ionization. The quality of mass spectra is improved even further by incorporating the matrix compound into the chip surface, a technique called surface-enhanced neat desorption.

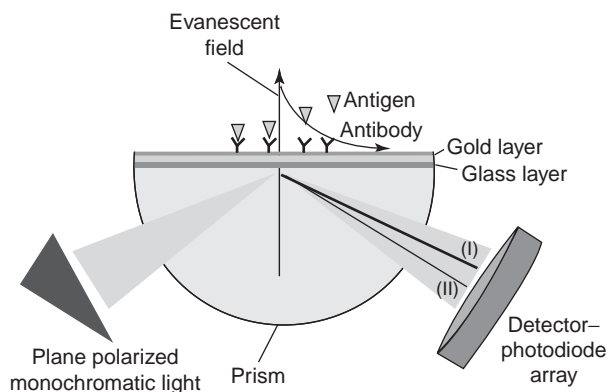
BIA-Core Inc., Uppsala, Sweden, produces a range of protein chips on which protein interactions can be detected by changes in surface plasmon resonance (SPR). This is an optical effect that occurs when monochromatic polarized light is reflected from thin metal films. Some of the incident light energy interacts with the plasmon (the delocalized electrons in the metal), which results in a slight reduction in reflected light intensity. The angle of incidence at which this shadowing effect, the SPR, occurs is determined by the material adsorbed onto the metal film, which in this case is one or more antibodies. There is a direct relationship between the mass of the immobilized molecules and the change in resonance energy at the metal

surface, which can be used to study interactions in real time. Thus, when light is shone on a gold-coated glass chip from underneath, the angle of incidence that induces SPR will change when analytes bind to antibodies on the chip surface, and the change will reflect the molecular mass of the analyte (Figure 5). Direct coupling of SPR spectroscopy and mass spectrometry allows interacting proteins to be characterized.

**See also:** **Immunoassays, Applications:** Clinical; Food; Forensic. **Immunoassays, Techniques:** Radioimmunoassays; Enzyme Immunoassays; Luminescence Immunoassays.

## Further Reading

- Bock JL (2000) The new era of automated immunoassay. *American Journal of Clinical Pathology* 113: 628–646.
- Borrebaeck CAK (2000) Antibodies in diagnostics – from immunoassays to protein chips. *Immunology Today* 21: 379–382.
- Engvall E and Periman P (1971) ELISA quantitative assay of immunoglobulin G. *Immunochemistry* 8: 871–874.
- Gosling JP (ed.) (2000) *Immunoassays: A Practical Approach*. Oxford: Oxford University Press.
- Gosling JP (1990) A decade of development in immunoassay methodology. *Clinical Chemistry* 36: 1408–1427.
- Hage DS (1999) Immunoassays. *Analytical Chemistry* 71: 294R–304R.
- Harlow E and Lane D (1999) *Using Antibodies: A Laboratory Manual*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
- Joos TO, Stoll D, and Templin MF (2002) Miniaturised multiplexed immunoassays. *Current Opinion in Chemical Biology* 6: 76–80.
- Ronald A and Stimson WH (1998) The evolution of immunoassay technology. *Parasitology* 117(Supplement): S13–S27.
- Wiese R, Belosludtsev Y, Powdrill T, Thompson P, and Hogan M (2001) Simultaneous multianalyte ELISA performed on a microarray platform. *Clinical Chemistry* 47: 1451–1457.
- Yalow RS (1978) Radioimmunoassay: A probe for the fine structure of biologic systems. *Science* 200: 1236–1245.



**Figure 5** Surface plasmon resonance spectroscopy for the detection of antigen–antibody interactions in highly multiplexed immunoassays. Plane polarized light is incident on a gold-coated glass chip containing immobilized antibodies. A change in mass at the surface, caused by antigen binding, causes a change in the refractive index and thus the resonance state. This is reported by a change in the angle of the reflected light (I to II), which can be detected using a photodiode array.

## Production of Antibodies

**S Stapleton, R O' Kennedy, and E Tully**, Dublin City University, Dublin, Republic of Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The aim of this article is to give a review of the current techniques used for the generation of antibodies

and antibody fragments. It describes the methodology involved in antibody production. This includes the choice of antigen and adjuvant and updated screening, purification, and characterization approaches.

### Antibody Structure

Although antibodies can vary in structure, they are generally typified by the immunoglobulin G (IgG)



analysis, a phenomenon described as surface-enhanced laser desorption/ionization. The quality of mass spectra is improved even further by incorporating the matrix compound into the chip surface, a technique called surface-enhanced neat desorption.

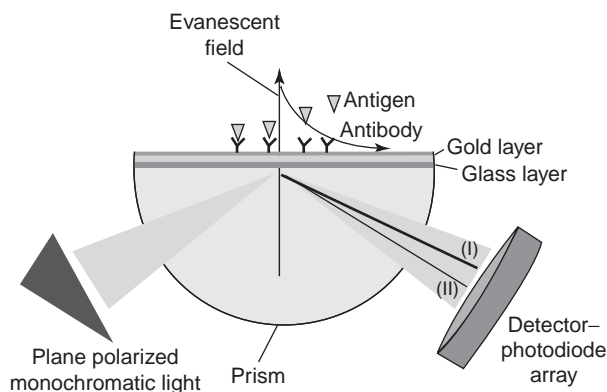
BIA-Core Inc., Uppsala, Sweden, produces a range of protein chips on which protein interactions can be detected by changes in surface plasmon resonance (SPR). This is an optical effect that occurs when monochromatic polarized light is reflected from thin metal films. Some of the incident light energy interacts with the plasmon (the delocalized electrons in the metal), which results in a slight reduction in reflected light intensity. The angle of incidence at which this shadowing effect, the SPR, occurs is determined by the material adsorbed onto the metal film, which in this case is one or more antibodies. There is a direct relationship between the mass of the immobilized molecules and the change in resonance energy at the metal

surface, which can be used to study interactions in real time. Thus, when light is shone on a gold-coated glass chip from underneath, the angle of incidence that induces SPR will change when analytes bind to antibodies on the chip surface, and the change will reflect the molecular mass of the analyte (Figure 5). Direct coupling of SPR spectroscopy and mass spectrometry allows interacting proteins to be characterized.

**See also:** **Immunoassays, Applications:** Clinical; Food; Forensic. **Immunoassays, Techniques:** Radioimmunoassays; Enzyme Immunoassays; Luminescence Immunoassays.

## Further Reading

- Bock JL (2000) The new era of automated immunoassay. *American Journal of Clinical Pathology* 113: 628–646.
- Borrebaeck CAK (2000) Antibodies in diagnostics – from immunoassays to protein chips. *Immunology Today* 21: 379–382.
- Engvall E and Periman P (1971) ELISA quantitative assay of immunoglobulin G. *Immunochemistry* 8: 871–874.
- Gosling JP (ed.) (2000) *Immunoassays: A Practical Approach*. Oxford: Oxford University Press.
- Gosling JP (1990) A decade of development in immunoassay methodology. *Clinical Chemistry* 36: 1408–1427.
- Hage DS (1999) Immunoassays. *Analytical Chemistry* 71: 294R–304R.
- Harlow E and Lane D (1999) *Using Antibodies: A Laboratory Manual*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
- Joos TO, Stoll D, and Templin MF (2002) Miniaturised multiplexed immunoassays. *Current Opinion in Chemical Biology* 6: 76–80.
- Ronald A and Stimson WH (1998) The evolution of immunoassay technology. *Parasitology* 117(Supplement): S13–S27.
- Wiese R, Belosludtsev Y, Powdrill T, Thompson P, and Hogan M (2001) Simultaneous multianalyte ELISA performed on a microarray platform. *Clinical Chemistry* 47: 1451–1457.
- Yalow RS (1978) Radioimmunoassay: A probe for the fine structure of biologic systems. *Science* 200: 1236–1245.



**Figure 5** Surface plasmon resonance spectroscopy for the detection of antigen–antibody interactions in highly multiplexed immunoassays. Plane polarized light is incident on a gold-coated glass chip containing immobilized antibodies. A change in mass at the surface, caused by antigen binding, causes a change in the refractive index and thus the resonance state. This is reported by a change in the angle of the reflected light (I to II), which can be detected using a photodiode array.

## Production of Antibodies

**S Stapleton, R O' Kennedy, and E Tully**, Dublin City University, Dublin, Republic of Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

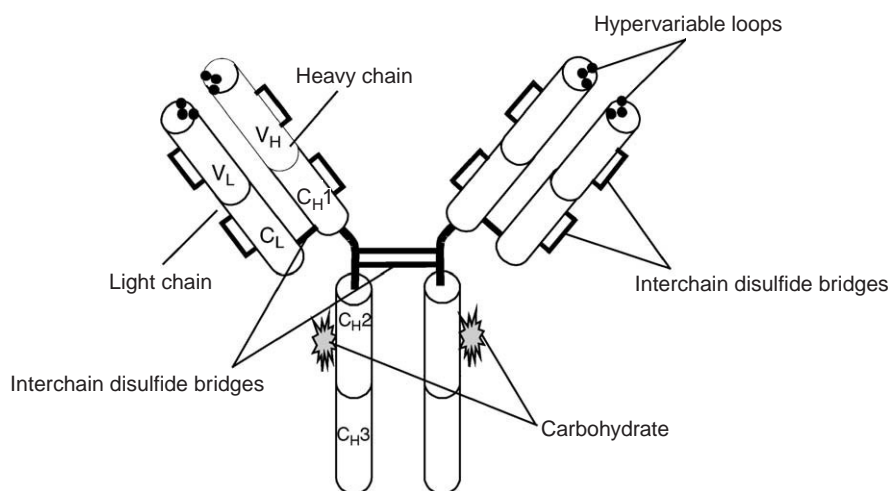
The aim of this article is to give a review of the current techniques used for the generation of antibodies

and antibody fragments. It describes the methodology involved in antibody production. This includes the choice of antigen and adjuvant and updated screening, purification, and characterization approaches.

### Antibody Structure

Although antibodies can vary in structure, they are generally typified by the immunoglobulin G (IgG)





**Figure 1** Basic structure of an antibody. The molecule consists of two identical heavy chains and two identical  $\kappa$  or  $\lambda$  light chains and has a molecular weight of 150–160 kDa. The antibody chains can be further divided into constant (C) and variable (V) regions based on their amino acid variability. The variable regions, located at the N-terminal portions of both the heavy and light chains, create the scaffolding for the antigen-binding site, which is formed by six hypervariable loops or CDRs. The Fc region of the molecule, which contains the C<sub>H2</sub> and C<sub>H3</sub> domains, plays a number of important physiological roles in immunological responses.

molecule, the most abundant subclass found in the serum of mammals. The basic four-polypeptide chain structure of an antibody is shown in **Figure 1**. It consists of two identical heavy chains and two identical light chains, which are held together by a number of disulfide bonds. The antibody chains can be further divided into constant (C) and variable (V) regions based on their amino acid variability. The variable regions, which are located at the N-terminal portions of both the heavy and light chains, form the antigen-binding fragment (Fab), one of the most important regions of the molecule. Within this region there are areas of hypervariability, referred to as the hypervariable loops or the complementarity determining regions (CDRs), which are primarily involved with binding antigen. It is the variation in the amino acid sequence of the CDRs that allows for the generation of a multitude of different antibodies with different binding specificities. The remainder of the variable heavy and light domains exhibit far less variation, and these stretches are known as the framework regions (FRs). The crystallizable fragment (Fc) of the antibody is involved in binding surface receptors and plays a number of important physiological roles in immunological responses. This Fc portion is linked to the two Fab domains by a hinge region, which provides flexibility, allowing binding sites on each Fab arm to work independently of each other.

## Antigens

Antigens can be defined as molecules that contain distinct sites or epitopes that are recognized and

interact with various components of the immune system. Immunization of an animal with a single antigen yields monospecific antiserum while mixtures of antigen, such as a cell or unpurified crude preparations, yield multispecific or polyspecific antiserum. However, not all antigens present can be presumed to elicit the same immune response, and the degree of response to a particular antigen may differ from one to another. Generally proteins and large molecules are capable of eliciting good immune responses with the aid of adjuvants (immune stimulants). Smaller molecules, such as drugs, with a molecular weight of less than 5000 Da, are often too small to cause an immune response on their own. They are referred to as 'haptens' and require conjugation to larger carrier molecules to render them immunogenic. Various carrier proteins, such as bovine serum albumin (BSA), ovalbumin (OVA), keyhole limpet hemocyanin (KLH), or thyroglobulin, are commonly used for this purpose. However, other carriers composed of synthetic or natural polymers (e.g., dextran, poly-L-lysine), lipid bilayers, or synthetic organic molecules have also been used. Regardless of the type of molecule, a carrier must be highly immunogenic, have the required solubility properties, be nontoxic *in vivo*, and possess suitable functional groups for coupling to the hapten.

There are a variety of coupling chemistries available for coupling haptens to carrier molecules. The choice of method is governed by the functional groups available on both the hapten and the carrier and the orientation of the hapten desired for presentation to the immune system. Conjugation procedures employing

the carbodiimide EDC (1-ethyl-3-(3-dimethyl-amino-propyl)carbodiimide hydrochloride) are commonly used for immunogen formation, as the method is both efficient and relatively simple. Another advantage of this cross-linking methodology is that no bridging molecule is introduced between the hapten and carrier, thus eliminating the potential of antibodies being generated against the coupling reagent, which could dilute the desired antibody response against the hapten. However in some cases the linker molecules between the hapten and the carrier may be necessary to increase exposure of the hapten and, thus, its immunogenicity.

Conjugation is a crucial step in antibody production as the specificity of the resultant antibodies is dependent on the hapten used to produce them. The one major problem associated with using conjugates for antibody production is that antibodies will be produced against both the hapten and the carrier molecule. To reduce the production of nonspecific antibodies and direct the production of antibodies to the hapten rather than the carrier, the number of haptens conjugated to each carrier molecule can be optimized and different proteins used in conjugates for immunization and screening.

## Adjuvants

Although carrier molecules increase immunogenicity, as discussed above, adjuvants are still generally required to stimulate initially an immune response in an animal. Adjuvants are nonspecific stimulators and are used in immunization methods to maximize an immune response. The most commonly used adjuvants are Freund's complete and incomplete adjuvants, FCA and FICA, respectively. FCA, which is only used for the initial immunization, consists of heat-killed *Mycobacterium tuberculosis*, nonbiodegradable mineral oils, and an emulsifier. The purpose of the mycobacteria in the adjuvant is to cause an inflammation, which attracts macrophages and other cells to the injection site. The antigen, which is emulsified in the oil, is released slowly to give a prolonged exposure to the immune system. To minimize side effects in the animal, FICA, which contains no bacteria, is used for subsequent boosts. Other adjuvants include Hunter's TiterMax, liposomes, the RIBI adjuvant system, and *Bordetella pertussis*. The RIBI adjuvant system (RAS) has been used as a successful alternative to Freund's adjuvant, because it is less viscous and, therefore, easier to handle. Antiserum can also be successfully produced without the use of adjuvant. This can be achieved by immunization of the antigen immobilized on a solid matrix (e.g., sepharose beads, nitrocellulose). It should be

noted that adjuvant is not required when cells are being used as immunogens.

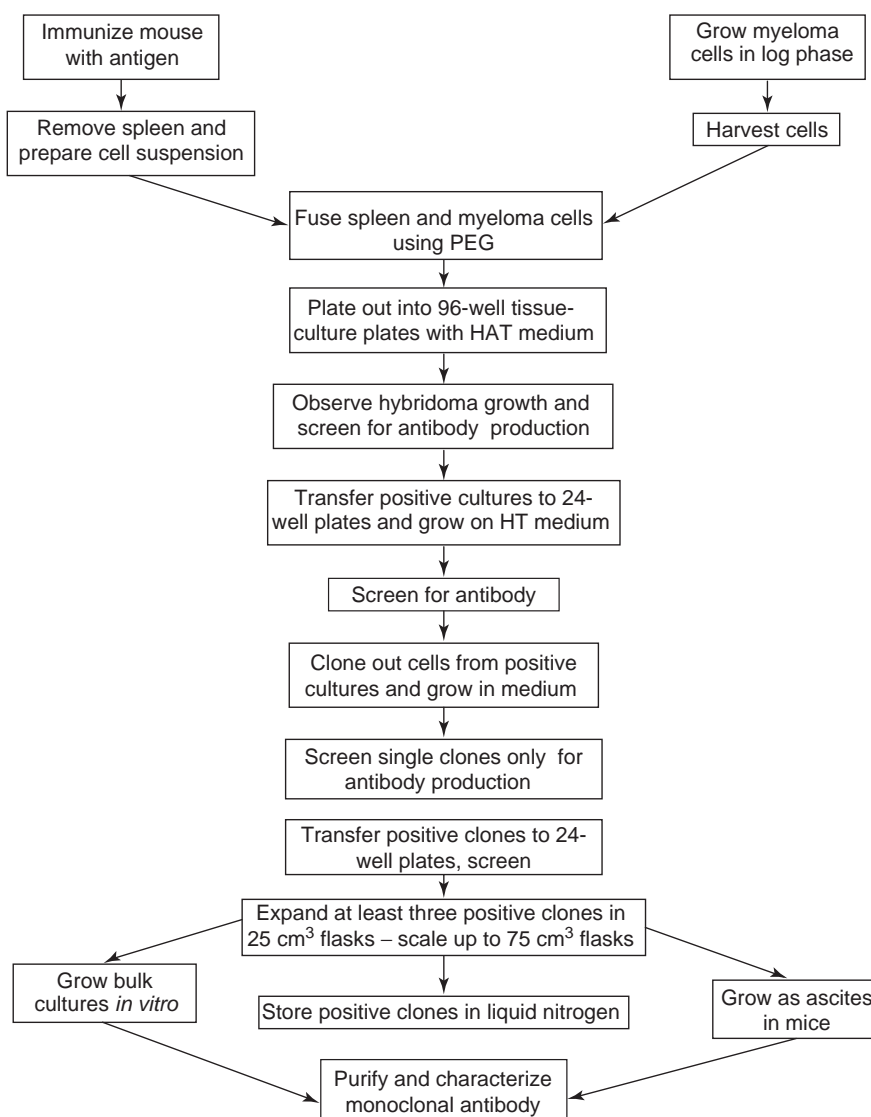
## Polyclonal Antibody Production

Polyclonal antibody production involves the repeated immunization of an animal with a desired antigen. Animals such as rabbits, goats, and sheep are generally used for polyclonal antibody production, as they are relatively easy to handle for immunization and bleeding purposes. When a sufficient titer (antibody concentration) is obtained, the animal is bled and the antibodies are purified from the serum. Polyclonal antibodies can be produced quickly and relatively cheaply and do not require the same amount of expertise or time as monoclonal antibody production. They can also be very specific and high concentrations can be purified from relatively small amounts of serum. However, polyclonal antiserum contains a heterogeneous population of antibodies, which can be hard to reproduce in subsequent immunizations. Unlike monoclonal antibody production, a consistent source of antibodies cannot be generated.

## Monoclonal Antibody Production

Antibodies secreted in response to immunization with a specific antigen are usually polyclonal, in that a mixed population of antigen reactive B-cells is stimulated, which recognize different epitopes on the immunogen. This system produces heterogeneous antibody populations, with a wide range of binding affinities. In monoclonal antibody production, however, the antibody molecule is derived from a single clone of B-cells and each antibody secreted by the clone has identical antigen-binding specificity. In other words, monoclonal antibodies provide single epitope specificity and potentially limitless amounts of identical antibody.

Monoclonal antibodies are most commonly produced by employing the hybridoma technique. **Figure 2** illustrates this technique. In this method, splenocytes (normal B-cells) from an immunized animal are fused with myeloma cells (tumor B-cells) using a fusion medium such as polyethylene glycol (PEG). Stock myeloma cell lines selected for the fusion process are specifically chosen because they are derived from a common ancestor MOPC-21 known to have lost the ability to produce IgG, so that the only antibody produced is derived from the antigen-sensitized splenocyte. The resultant hybrid cell is known as a hybridoma and inherits antigen specificity from the splenocytes and immortality from the myeloma cells, thus creating a



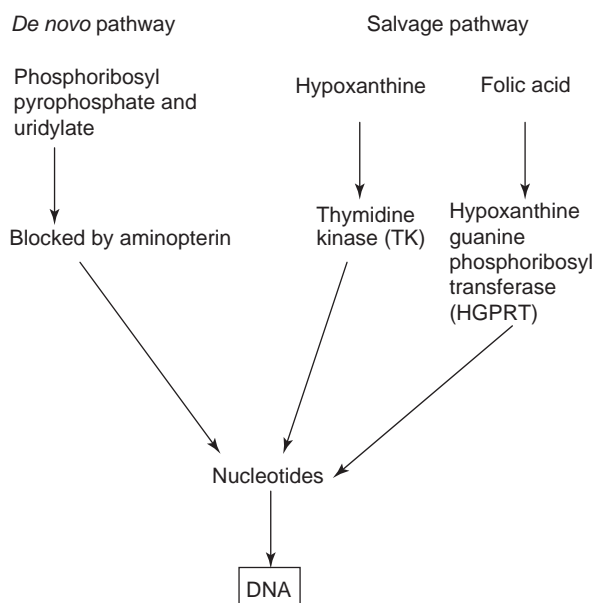
**Figure 2** Flow diagram illustrating the production of monoclonal antibodies using hybridoma technology.

permanent cell line, which secretes a homogeneous antibody of desired specificity.

Once the fusion has taken place, it is necessary to eliminate any unfused myeloma cells and to select only hybrid cells secreting antibody. This is primarily achieved by the use of hypoxanthine aminopterin thymidine (HAT) media and cells that are deficient in the enzyme responsible for incorporation of hypoxanthine into DNA. **Figure 3** illustrates this process. The unfused splenocytes are not immortal and naturally die off in culture. The elimination of the unfused myeloma cells is carried out by the initial use of mutant myeloma cells selected for a deficiency in the enzymes hypoxanthine guanine phosphoribosyl transferase (HGPRT) and thymidine kinase (TK), rendering them unable to use the salvage pathway for nucleic acid synthesis. The myelomas will die off

since the main pathway for purine and pyrimidine synthesis is blocked by aminopterin contained in the HAT media. Mammalian cells synthesize nucleotides by two different pathways. HAT medium contains aminopterin to block the *de novo* pathway and hypoxanthine and thymidine to allow growth by the salvage pathway. Cells that lack either HGPRT or TK will die in HAT medium, because they lack the ability to use the salvage pathway to acquire essential intermediates for the synthesis of nucleic acids.

Once selection is successful, it is necessary to perform initial screening to identify positively the hybridomas secreting the antibody of interest. Screening methods will be discussed in more detail later, but it is important to note that the results of initial screening are vital in determining what clones should be expanded. It is important to identify positive cultures as



**Figure 3** Schematic representation of HAT selection. HAT medium contains aminopterin to block the *de novo* pathway and hypoxanthine and thymidine to allow growth by the salvage pathway. Cells that lack either HGPRT or TK will die in HAT medium, because they lack the ability to use the salvage pathway to synthesize nucleic acids.

early as possible in order to concentrate on expanding the antibody-secreting hybrids, but it is equally important to eliminate nonsecreting clones at this stage to avoid overgrowth and wasting time and materials propagating them. When antibody activity has been determined, immediate cloning and retesting should take place. The type of assay used to detect the antibody should ideally be chosen to mimic the eventual application envisaged and should be standardized prior to hybridization.

Cloning is undertaken to ensure that all hybridomas are derived from a single parent that secretes the required antibody and that nonsecreting clones arising either in the original fusion wells or as spontaneous variants do not outgrow the antibody-secreting hybrids. This ensures that the hybridomas secrete antibody that is homogeneous and monospecific. Early cloning allows the selection of cells that have chromosomes for antibody production. Otherwise variants not producing IgG will overgrow. There are two common methods used to clone out hybridomas: soft agar and limiting dilution. In soft agar the cells are plated out on a semisolid medium such as low-agar plates. Single cells produce colonies that are further expanded. Limiting dilution involves seeding the cells at a known concentration in 96-well plates. It is important to seed at very low densities of less than one cell per well to ensure monoclonality. Growth conditions must be optimized to ensure that such low

cell densities can survive; this is achieved by the use of special media such as Briclone. However, it is necessary to carry out several rounds of cloning before a stable monoclonal hybrid can be assumed. Hybridomas can become unstable over time and hence stocks of original clone must be maintained by cryopreservation in liquid nitrogen.

Another method for producing hybridomas is as ascites tumors in mice. In this method, hybridomas are injected into animals. The antibody is contained in the rich ascites fluid that develops. Large doses of cells are required to initiate the tumor and it is important that the hybridoma and mouse are compatible. The productions of ascites tumors are facilitated by pretreatment of the mouse with, and intraperitoneal injection of, pristane. The ascites fluid is drained once sufficient swelling has been produced at the tumor site. The main concerns with this method are animal welfare issues, but the main advantage is that large amounts of antibody can be prepared.

## Problems Associated with the Production of Monoclonal Antibodies

There are several problems associated with monoclonal antibodies that can detrimentally affect production yields. It is vital to use fresh reliable cells because a low yield of hybrids can occur due to spontaneous drift in the myeloma cell line. A poor immunogen can also affect hybrid yield; however, a change in immunization schedule or strain of mouse can rectify this problem. The instability of hybrids can result in a complete loss of antibody secretion. This is usually due to overgrowth by nonsecreting cells, and therefore it is vital to plate out cells at high dilution postfusion. Many studies have been carried out to find optimal culture conditions for positive secreting clones. One such study indicated that in order to generate the largest possible number of positive hybrids, it is necessary to seed the cells in a concentration that will result in the growth of one hybrid clone per well. In a culture containing many clones, the number of hybrids actively secreting a specific antibody is relatively low; therefore, maximal yield of antibody is obtained by seeding at high dilution and low density.

Chromosome loss and gene segregation can have a significant effect on heavy and light chain synthesis in a cloned hybrid line, resulting in total loss of antibody production. The reason for this is that maximally producers (good antibody secretors) are terminally differentiated because they devote a higher percentage of protein synthetic activity to the antibody, rendering them more likely to die after a few generations in culture.

Mycoplasma contamination of cell culture systems continues to present major problems for monoclonal antibody production. Mycoplasma-positive cell cultures are themselves the major source of infection. It is recommended that all myeloma cell lines be tested for mycoplasma prior to fusion. If a hybridoma culture is considered irreplaceable, it is possible to eliminate effectively the mycoplasma contamination by injecting the hybrids into mice. The animal's immune system will destroy the infection and effectively clean up the cells. The mycoplasma-free cells are then recovered by draining the ascites fluid. Drug treatment is also an effective method to decontaminate hybrid cells from mycoplasma infection.

## Screening Methods

There are many principles to consider in selecting a suitable screening assay for the detection of hybridomas secreting antibodies of desired specificity. Speed and convenience are important factors to consider since hybrid clones mature and grow at different rates. Batch testing is possible, but not recommended as delays in screening can risk the loss of positive clones due to overcrowding or selective growth of nonsecreting cells. A standardized, reliable, sensitive assay is required that must be relevant to the format of the antibody application. There are three main types of screening method: antibody capture, antigen capture, or functional assays. Antibody and antigen capture assays are the most frequently used techniques but functional assays are becoming more common, whereby the antibodies in the hybridoma supernatants are used to block a reaction. Binding assays are suitable for monoclonal antibody screening as they detect all antibody activity against a particular antigen and can usually be modified for the use of isotype-specific IgG antibodies. Enzyme-linked immunosorbent assay (ELISA) techniques are well-established methods used to determine soluble or insolubilized antigens or antibodies. It is sometimes possible that sandwich assays are more suitable than direct assays for the initial screening of monoclonal antibodies, since the antigen is immobilized in a number of orientations using polyclonal antibodies, therefore allowing detection of the monoclonal with a wide range of specificities.

Live cell ELISA assays are used for the screening of monoclonal antibodies directed to cell surface antigens. This is advantageous for antigen detection on homogeneous cell lines or cell mixtures with a large number of positive secreting cells. Screening methods must be optimized and the chosen optimal conditions controlled. This is because the binding affinity of an

antibody can depend on physical parameters such as pH and temperature, thus varying the apparent specificity of the antibody if these conditions are altered. Current general screening methods include ELISA, radioimmunoassay, dot and Western blotting, flow cytometry, and arrays.

## Large-Scale Production

Once a monoclonal antibody with optimal properties has been selected and fully characterized, it is usual to scale up and attempt to produce the antibody-secreting hybridoma on a large scale. This can be achieved in a number of ways. Large amounts of monoclonal antibody can be produced either *in vivo* or *in vitro*. The concentration of monoclonal antibody produced in culture ranges from 1 to 20  $\mu\text{g ml}^{-1}$ , whereas *in vitro* production via ascites fluid yields between 1 and 20  $\text{mg ml}^{-1}$ . However, large-scale production involves a lot more than the simple culturing of large batches of cells or their injection into large numbers of mice.

A concern associated with culture technology is the problem of serum contamination. The presence of animal serum proteins can affect the purity of the monoclonal antibody. Recent studies, however, show that serum-free media can sustain certain hybridomas. The biggest issue in scaling up is the elimination of the water-soluble nongaseous products of cell metabolism. Two chamber devices are required to rectify this, but have proved difficult to fabricate and more complicated to operate than traditional methods. Novel large-scale culture systems such as the Tecnomouse, Berlin Tumbling Chamber, Bochum Glass Mouse, and miniPERM Bioreactor are now being used for monoclonal antibody production. These systems have semipermeable membranes that keep back the cells and the secreted antibody, but allow the passage of nutrients and metabolic products. In recent years, there have been a number of major advancements in large-scale production. There are variations in each design, but the principle is the same and most large-scale systems are agitated to enhance production. The principle of rotating cultures in scaled-up systems greatly increases the antibody yield, whereby eight times more IgG was produced in a roller bottle system when the cultures were continuously agitated.

The commercial production of monoclonal antibodies is carried out for two main reasons: antibodies are used in diagnostic test development and in the research of emerging antibody therapies. Glennie and co-workers have reported on recent advances in the commercial production of antibodies for therapeutic treatments. Currently, over 130 monoclonal



antibodies are undergoing clinical and commercial trials for approval by the US Food and Drug Administration (FDA). Reichert has reported on the approval of 11 monoclonal antibodies for use in the treatment of diseases and conditions such as cancer, Crohn's disease, and transplant rejection. The human immune system however tends to react to mouse-derived antibodies, which in turn leads to a decrease in the effectiveness of the therapy or therapeutic complications. Transgenic mice strains have been engineered to produce human gene repertoires in the absence of mouse antibodies, thus allowing the production of human antibodies to specific antigens while eliminating the problem of negative immune responses in humans. Jakobovits has described the production of a transgenic strain called xenomouse. This particular strain provides an excellent source for the manufacture of antigen-specific humanized monoclonal antibodies for therapeutic use.

## Antibody Engineering

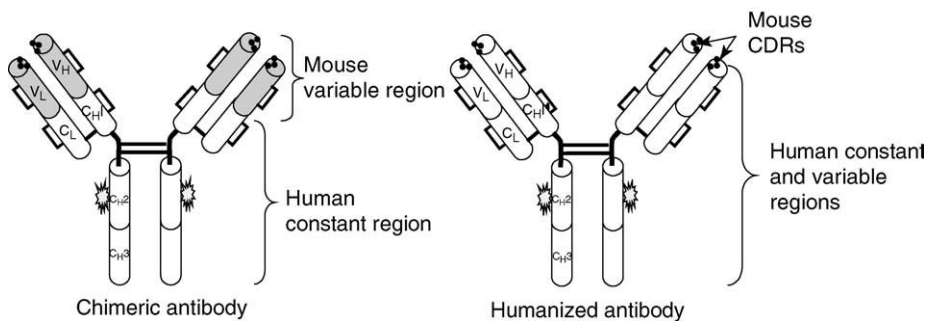
Hybridoma technology has facilitated the production of large amounts of homogeneous antigen-specific rodent-derived antibodies for use as diagnostic and therapeutic agents. Although this technique has indeed proven invaluable, there are a number of drawbacks. Hybridomas are sometimes difficult and expensive to maintain in culture, but probably the main disadvantage with this technology is the difficulty in producing large quantities of stable human monoclonal antibodies, due to the absence of suitable myeloma cell lines. These disadvantages prompted the development of recombinant antibody technology.

Genetic engineering allowed the first generation of humanized monoclonal antibodies, referred to as chimeric antibodies. These antibodies consist of variable regions of a murine monoclonal antibody linked to the constant regions of a human IgG

molecule. A second generation of humanized antibodies soon followed, in which the antigen-binding loops (CDRs) of the murine monoclonal antibodies were grafted onto a human IgG molecule (Figure 4).

The main reason for the generation of these constructs was to attempt to circumvent the problem of inducing human anti-mouse antibody (HAMA) responses when murine monoclonal antibodies are used as therapeutic agents in human patients. However, they did not fully overcome this problem and it was not till much later, with the development of combinatorial display libraries, that the production of fully human monoclonal genetically derived antibodies was possible. This technology has allowed the generation of monoclonal antibodies from several species, including human, chicken, rabbit, and sheep. Advances in molecular biology, facilitated by the introduction of techniques such as polymerase chain reaction (PCR), have allowed the creation of large repertoires of antibodies from antibody variable genes bypassing hybridoma technology and, sometimes, even immunization. These libraries, combined with phage display techniques, attempt to reproduce the diversity of the immune system by displaying millions or even billions of antibody fragments on the surface of phage. This technology has now made it possible to select an antibody to almost any antigen from an antibody repertoire with a large number (greater than  $10^8$ ) of clones.

The first step in the construction of a combinatorial antibody display library is to obtain the genes encoding for an antibody fragment. These genes can be isolated from B-cells of immunized animals (immune libraries) or from nonimmunized donors (naive libraries). Alternatively, antibodies may be constructed artificially by the *in vitro* assembly of V-gene segments and D/J segments (synthetic libraries). Conventional hybridomas can also be used as a source of variable genes for the construction of antibody fragments. The variable antibody genes, which

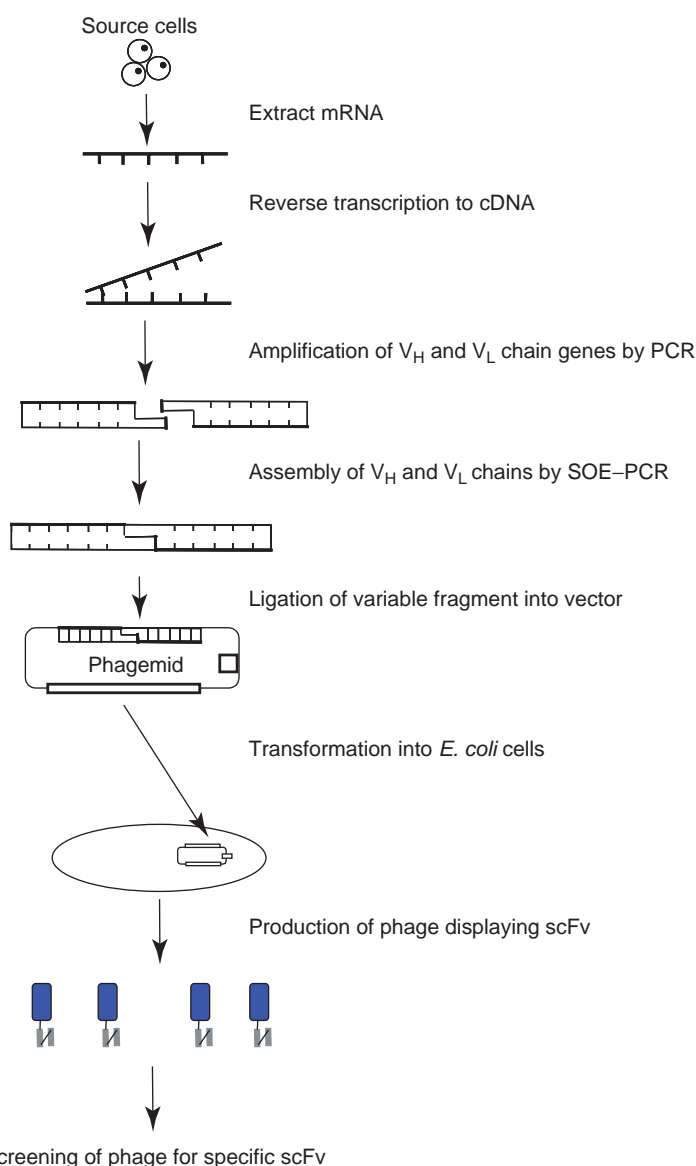


**Figure 4** Illustration of chimeric and humanized antibodies. Chimeric antibodies consist of variable regions of a mouse antibody linked to constant regions of a human IgG molecule, whereas humanized antibodies are almost entirely human except for murine CDRs.

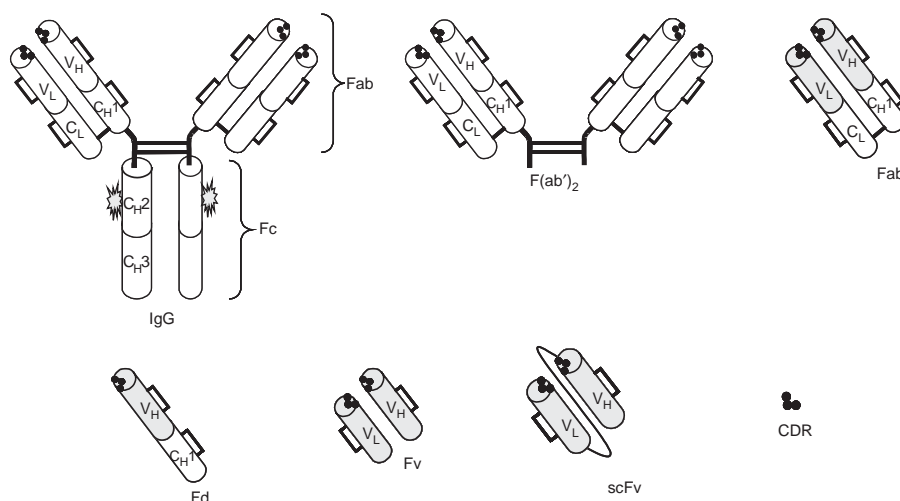
should contain a full repertoire of antigen specificities, are first amplified using suitable primers. They are then randomly combined by splice by overlap extension-PCR (SOE-PCR), inserted onto a suitable plasmid, and applied to one of a variety of expression systems, including yeast (e.g., *Pichia pastoris*), bacterial cells (e.g., *Escherichia coli*), plant cells (e.g., *Nicotiana tabacum*), mammalian cells (e.g., COS cells), and baculovirus-infected insect cells.

The most widely used methodology for recombinant antibody library production utilizes filamentous phage, a bacteriophage that infects *E. coli*. The phage infect *E. coli* by attaching to the tip of the F pilus translocating the phage genome into the bacterial cytoplasm. *E. coli* is one of the most popular bacterial

expression hosts used, as it is easily manipulated and its fast growth allows the large-scale production of proteins. The method (Figure 5) involves cloning the genes encoding the antibody fragment into a particular vector as a fusion to the gene encoding one of the phage coat proteins (pIII or pVIII). Originally complete phage vectors were used as the display vector, but now small plasmid vectors or phagemids, which contain the appropriate packaging signal and cloning sites, are more commonly used. The phagemid encoding the antibody fragment-pIII fusion is packaged into phage particles using a particular helper phage (e.g., M13K07 or VCS-M13) and then used to transform a suppressor strain of *E. coli* (e.g., XL-1 blue). The resulting antibody fragment is expressed on the



**Figure 5** Schematic representation of the production of a combinatorial antibody phage display library.



**Figure 6** Structure of an antibody molecule and its various fragments. The  $F(ab')_2$  fragment comprises two disulfide-linked antigen-binding fragments, while the Fab fragment is only one antigen-binding fragment. Fab can be broken down into an Fv fragment consisting of the variable domains of the antibody. Incorporation of a hydrophilic peptide linker into the fragment produces an scFv. Other fragments include the Fd fragment, comprising a  $V_H$  and  $C_{H1}$  domain, and the CDRs, the smallest fragment capable of binding antigen.

surface of the phage, while its genetic material resides within the phage particle, providing a direct link between antibody genotype and phenotype. After expression, the displayed antibody fragment can be screened for antigen specificity by a process known as biopanning. This involves incubating the phage library in an immunotube coated with a specific antigen. Unbound phage particles are washed away, and bound phage are recovered from the surface, reinfected into the bacteria, and regrown for further enrichment. An antibody fragment with the best affinity and specificity can then be isolated and analyzed.

Prior to the development of recombinant technologies, antibody fragments could only be generated by proteolytic cleavage. Pepsin digestion of the IgG molecule produced a  $(Fab')_2$  fragment while papain digestion resulted in the generation of single Fab fragments. Now recombinant antibody technology can be used for the generation of a number of antibody fragments, which include Fab, Fv, and scFv, shown in **Figure 6**. The Fv fragment is the smallest fragment of an antibody that still contains the complete antigen site. However, the fragment, consisting of the variable domains of the light and heavy chain, can be quite unstable, as it lacks the disulfide bond present in the Fab fragment. This problem has been overcome by protein engineering. Disulfide bonds can be formed by the introduction of cysteine residues into the variable domains producing a disulfide-stabilized Fv fragment (dsFv). Another means of overcoming stability problems is to incorporate a flexible peptide linker of  $\sim 15$ – $20$  amino acids into

the Fv fragment resulting in a more stable single-chain variable fragment (scFv), the most commonly used configuration (**Figure 6**). Antibody engineering techniques have also been employed to manipulate the fragments to create dimeric, trimeric, and multimeric scFvs. These multimeric molecules can be specific for one particular target antigen or have multiple specificities to different antigens, by associating different scFv molecules derived from different parental antibodies. The range of antibody formats achieved by combining antibody fragments is virtually unlimited. Smaller antibody fragments, which have the ability to bind antigen, can also be engineered; these include the Fd fragment, consisting of a  $V_H$  and  $C_{H1}$  domain, and the CDR.

Recombinant antibody technology has several advantages, including the speed of antibody production, the possibility of altering affinity and specificity, and the ability to generate novel functionalities. Once the system has been set up in the laboratory, it can take weeks rather than months to produce an antibody fragment of desired specificity. With the development of new expression vectors, it is now also possible to express antibody fragments on a large scale, link the fragment to an enzyme allowing direct detection, or engineer a histidine (His) tag into the antibody sequence to facilitate purification by immobilized metal affinity chromatography (IMAC).

The emergence of this technology has made it possible to generate high-binding-affinity molecules against any chosen target, resulting in potentially

limitless applications. Recombinant antibodies have already contributed greatly in the analytical and diagnostic fields. They are especially ideal for tumor targeting, where small rapidly penetrating but high-affinity molecules are essential. However, further work in developing the technology is required and there are many challenges ahead before their full potential is reached.

## Antibody Purification

Purification of antibodies is carried out to enhance the reactivity of the antibody by removing any interfering immunoglobulins or contaminants. There are a number of methods available for antibody purification, and the choice of technique is dependent on a number of factors. These include the level of purity required, antibody class, antibody source, and intended application. Antibodies may be separated according to charge or size, using traditional protein purification techniques. A precipitation technique such as saturated ammonium sulfate (SAS) precipitation isolates the IgG molecules from other serum proteins, thus stabilizing the antibody and reducing lipid content.

Purification by ion-exchange chromatography is based on the fact that  $\gamma$  globulins (including antibodies) are the least negatively charged of the serum proteins. This makes them easy to purify using positively charged ion-exchange matrices. Affinity chromatography isolates a specific protein or group of proteins with similar characteristics. Affinity purification techniques can also be used to purify antibodies by capturing them on a gel matrix containing a covalently coupled analyte. Purified antigen can be acquired by coupling antibodies to an affinity

column. The most commonly used method of affinity purification exploits the ability of protein A and protein G to bind specifically to certain portions of the IgG molecule. When protein A and protein G are immobilized onto sepharose, the purification of monoclonal and polyclonal IgG-type antibodies is possible. There are several useful methods for recombinant antibody purification, many of which are similar to those used for monoclonal and polyclonal purification. The major technique currently used for scFv purification is IMAC. The affinity column involved in IMAC purification consists of transition metals, chelated by various acids. It is possible to engineer a histidine tag into an antibody fragment sequence and it is this His tag protein that binds the metal chelate column, thus facilitating purification.

Some of the other main techniques that are available are outlined in **Table 1**, along with their specific advantages and disadvantages.

There are also a number of commercially available antibody purification kits that have emerged in recent years; details of such kits are available from the manufacturers.

## Antibody Characterization

Characterization of antibodies is useful in determining antibody class, binding kinetics, epitope specificity, cross-reactivity, and antibody affinity. **Table 2** illustrates some of the techniques used to assess these characteristics.

Determination of the class and subclass of the antibody is performed in a number of ways, including radioimmunoassay, ELISA, and immune precipitation, using subclass-specific antisera. These are commercially available in the form of antibody isotyping kits.

**Table 1** Antibody purification techniques

<i>Technique</i>	<i>Usefulness</i>	<i>Advantages</i>	<i>Disadvantages</i>
Ammonium sulfate precipitation	Use as a single step, or with other methods	Cheap, easy, convenient for large volumes, nearly pure antibody	Moderate yields, multiple steps
Protein G	IgG isolation	High yields	Expensive, antibodies with high affinity to antigen are sometimes difficult to elute
Protein A	Pure IgG but low yield of IgG <sub>1</sub>	Binds bovine IgG weakly, high yields	Expensive, does not bind IgG <sub>1</sub>
IMAC	His tag recognition; recombinant antibody isolation	High yields, milder elution conditions	Formation of inclusion bodies, difficulty in solubilization
Gel filtration and DEAE	Relatively pure IgM	Cheap, useful for large volumes, extremes of pH not necessary	Less useful for separation of IgG classes
Antigen affinity chromatography	Useful for all types of antibody isolation	Binds specific antibodies	Must make specific antigen-linked column/matrix

**Table 2** Characterization of antibodies

Characteristic	Technique
Antibody specificity and cross-reactivity	Screening test against different related antigens using ELISA, RIA, BIAcore <sup>®</sup> , Western, and dot blotting
Antibody affinity	Kinetic studies, BIAcore <sup>®</sup>
Stability on storage	Testing of titre on recovery after storage
Monoclonality	Cross-absorption studies, subcloning and analysis, flow cytometry
Antibody class/subclass	Immune precipitation, RIA, ELISA

BIAcore<sup>®</sup> is a major manufacturer of advanced bioanalytical systems that provide real-time quantitative data on binding interactions between biomolecules using surface plasmon resonance (SPR) technology. For more information, see [www.biocore.com](http://www.biocore.com)

In the case of monoclonal antibodies, determination of the antibody class does not prove that the antibody is truly monoclonal. However, if the antibody was produced from a hybridoma that has undergone a rigorous and technically satisfactory cloning procedure and the antibody shows evidence of monoclonal specificity as established by a predetermined classification, it is assumed that the antibody exhibits monoclonality. The antibody specificity is investigated by performing cross-reactivity studies with the antigen and other structurally similar antigens using ELISA, RIA, BIAcore, Western, and dot blotting. The affinity of the antibody for its specific antigen can be investigated using ELISA or biosensors. BIAcore is a real-time SPR-based biosensor that can be used to study biomolecular interactions without the need for labeling. The antibody–antigen interaction can be studied using BIAcore, thus providing valuable information on the kinetics of association and dissociation.

## Acknowledgment

The support of Enterprise Ireland and the Food Safety Authority of Ireland is gratefully acknowledged.

See also: **Immunoassays:** Overview. **Immunoassays, Applications:** Clinical; Food; Forensic. **Immunoassays, Techniques:** Radioimmunoassays; Enzyme Immunoassays; Luminescence Immunoassays.

## Further Reading

Brennan J, Dillon P, and O’Kennedy R (2003) Production, purification and characterisation of genetically derived scFv and bifunctional antibody fragments capable of detecting illicit drug residues. *Journal of Chromatography B* 786: 327–342.

Brennand DM, Danson MJ, and Hough DW (1986) A comparison of ELISA screening methods for the production of monoclonal antibodies against soluble protein antigens. *Journal of Immunological Methods* 93: 9–14.

Campbell AM (1984) *Monoclonal Antibody Technology*. New York: Elsevier.

Delves PJ (1995) *Antibody Applications: Essential Techniques*. Chichester: Wiley.

Glennie MJ and Johnson PWM (2002) Clinical trials of antibody therapy. *Immunology Today* 21: 403–410.

Hermanson GT (1996) *Bioconjugate Techniques*. London: Academic Press.

Hurrell JGR (1985) *Monoclonal Hybridoma Antibodies: Techniques and Applications*. Boca Raton, FL: CRC Press.

Jakobovits A (1998) Production and selection of antigen-specific fully human monoclonal antibodies from mice engineered with human IgG loci. *Advances in Drug Delivery Reviews* 31: 33–42.

Köhler G and Milstein C (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 256: 495–497.

Kontermann R and Dübel S (2001) *Antibody Engineering*. New York: Springer.

Krebber A, Bornhauser S, Burmester J, et al. (1997) Reliable cloning of functional antibody variable domains from hybridoma and spleen cell repertoires employing a reengineered phage display system. *Journal of Immunological Methods* 201: 35–55.

Mayer RJ and Walker JH (1987) *Immunochemical Methods in Cell and Molecular Biology*. London: Academic Press.

Nolan O and O’Kennedy R (1991) Bifunctional antibodies: concepts and applications. *International Journal of Clinical and Laboratory Research* 22: 21–27.

Reichert JM (2002) Therapeutic monoclonal antibodies: trends in development and approval in the US. *Current Opinions in Molecular Therapeutics* 4: 110–118.

Roitt IM (1994) *Essential Immunology*. Oxford: Blackwell Scientific Publications.



# IMMUNOASSAYS, APPLICATIONS

Contents

**Clinical**

**Food**

**Forensic**

## Clinical

**R M Twyman**, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by V Marks, pp. 2086–2093, © 1995, Elsevier Ltd.

## Introduction

Antibodies were used for the detection and quantitation of clinically important substances long before the development of immunoassays by Rosalyn Yalow and Solomon Berson in the 1960s and 1970s. However, since its first use for the measurement of insulin in blood, the immunoassay has become established as the most popular analytical technique applied in clinical medicine. Immunoassays are the first resort for the detection and measurement of a large number of hormones, blood products, enzymes, drugs, nutritional factors, and disease markers (**Table 1**). The popularity of the technique stems from its simplicity, rapidity, accuracy, and portability, the last allowing the technique to be adapted for ‘near-patient testing’. In most cases, immunoassays can be performed directly on untreated samples, such as plasma, serum, urine, saliva, and cerebrospinal fluid. While renowned for its ability to detect and quantify single analytes, more recent developments have seen the emergence of highly multiplexed immunoassays for the detection and quantification of hundreds or even thousands of analytes simultaneously. These novel platforms have permitted a shift in perspective, in which immunoassays are used to generate profiles of clinical samples that can facilitate accurate disease diagnoses and the prediction of drug responses.

## Types of Immunoassay Used in Clinical Applications

Immunoassays are analytical methods that achieve the detection and quantitation of analytes in clinical samples through the formation of a stable complex between the analyte and a specific antibody. The first

immunoassay techniques were based on the principle of competition between the analyte and a radio-labeled tracer with the same antigenic properties as the analyte. These two components would compete to occupy a limited number of antibody molecules. The amount of analyte present in the sample could be determined when the system reached equilibrium by separating the bound and unbound fractions of the tracer, and measuring the latter. The more analyte present in the sample, the more tracer would be displaced from the antibody and the greater the strength of the radioactive signal from the unbound fraction.

Such assays were described as heterogeneous because they depended on the separation of free and complexed tracer components, a process that was necessary because the label itself was unaffected by antibody binding. This is not necessarily true of non-isotopic labels such as enzymes and fluorophores, whose activities are often modified when the antigen to which they are attached interacts with its cognate antibody. Where antibody binding either inhibits or stimulates the signal generated by such a label, separation into free and complexed components is no longer necessary, and such assays are described as homogeneous. The main advantage of homogeneous assays is the absence of a separation step, which reduces the hands-on time and the likelihood of technical errors.

One of the disadvantages of the competition assays discussed above is that it is necessary to ensure that the amount of tracer added to the sample is not vastly in excess of the amount of analyte, and that the antibody is present at a lower molar concentration than the tracer. This means that very small amounts of reagents are used, which places high demands on the sensitivity of the assay. This problem was addressed by placing the label on the antibody rather than the analyte, as first demonstrated in the case of the immunoradiometric assay. In this format, quantitative data are generated by establishing a stoichiometric interaction between the analyte and the antibody. The signal produced by the antibody is thus directly proportional to the concentration of the analyte, rather than being related to it in a complex manner as is the case for competition analysis.

**Table 1** Representative list of immunoassays of major clinical interest

<i>Substance</i>	<i>Sample</i>	<i>Clinical indications</i>
<i>Polypeptide hormones</i>		
ACTH	Plasma	Distinguishes Cushing syndrome and Addison's disease
AVP	Plasma	Differential diagnosis of diabetes insipidus
Calcitonin	Plasma	Diagnosis of medullary carcinoma of the thyroid
C-peptide	Plasma	Differential diagnosis of hypoglycemia
Erythropoietin	Plasma	Investigation of anemia/polycythemia
FSH	Plasma	Investigation of infertility
Gastrin	Plasma	Diagnosis of Zollinger–Ellison syndrome
Glucagon	Plasma	Diagnosis of glaucoma
Growth hormone	Plasma	Diagnosis of acromegaly
	Urine	Diagnosis of growth hormone deficiency
hCG	Urine	Confirmation of pregnancy
Insulin	Plasma	Differential diagnosis of hypoglycemia
IGF-I	Plasma	Monitoring acromegaly, response to hCG therapy
IGF-II	Plasma	Differential diagnosis of hypoglycemia
LH	Plasma	Investigation of infertility
Neurotensin	Plasma	Investigation of diarrhea
Pancreatic polypeptide	Plasma	Diagnosis of APUDomas
Parathyroid hormone	Plasma	Differential diagnosis of hypo/hypercalcemia
Proinsulin	Plasma	Differential diagnosis of hypoglycemia
Prolactin	Plasma	Investigation of infertility, galactorrhea
Somatostatin	Plasma	Diagnosis of somatostatinoma
TSH	Plasma	Investigation of thyroid disease
VIP	Plasma	Differential diagnosis diarrhea
<i>Steroids</i>		
Aldosterone	Plasma	Investigation of hypertension, hypokalemia
Androstenedione	Plasma	Investigation of hirsutism, virilism, infertility
Bile-salts	Plasma	Investigation of liver disease
Cortisol	Plasma	Diagnosis of Cushing syndrome and Addison's disease
DHAS	Plasma	Investigation of hirsutism, virilism, infertility
11-Deoxycortisol	Plasma	Investigation of virilism
Dihydrotestosterone	Plasma	Investigation of virilism, impotence, feminization
17 $\alpha$ -Hydroxyprogesterone	Plasma	Investigation of congenital adrenal hyperplasia
Estradiol	Plasma	Investigation of infertility
Progesterone	Plasma	Investigation of infertility
Testosterone	Plasma	Investigation of virilism, impotence, infertility
<i>Amino/fatty acid derivatives</i>		
5-HIAA	Urine	Diagnosis of carcinoidosis
Melatonin	Plasma, urine	Investigation of sleep disorders
Reverse T <sub>3</sub>	Plasma	Investigation of thyroid disease
Thyroxine	Plasma	Investigation of thyroid disease
Triiodothyroxine	Plasma	Investigation of thyroid disease
<i>Proteins</i>		
Apolipoprotein A1	Plasma	Assessment of coronary heart disease risk
Apolipoprotein B	Plasma	Assessment of coronary heart disease risk
CK-MB	Plasma	Differential diagnosis of chest pain
C-reactive protein	Plasma	Monitoring response to anti-inflammatory therapy
Ferritin	Plasma	Investigation of iron storage capacity
Lipoprotein	Plasma	Assessment of coronary heart disease risk
Sex-hormone-binding globulin	Plasma	Investigation of hirsutism, virilism, infertility
Thyroxine-binding globulin	Plasma	Investigation of thyroid disease
Thyroglobulin	Plasma	Investigation of thyroid disease
Trypsin	Plasma	Investigation of thyroid disease
<i>Disease markers</i>		
$\alpha$ -Fetoprotein	Plasma	Investigation of liver cancer and other cancers
$\beta$ -2-Microglobulin	Plasma, CSF	Prognosis of myeloma and lymphoma, neuronal involvement
CA 15-3	Plasma	Monitoring breast cancer
CA 125	Plasma	Monitoring ovarian cancer

**Table 1** Continued

<i>Substance</i>	<i>Sample</i>	<i>Clinical indications</i>
CA 549	Plasma	Monitoring breast cancer
CEA	Plasma	Monitoring colorectal cancer
HCG	Plasma	Monitoring choriocarcinoma
MCA	Plasma	Monitoring breast cancer
NSE	Plasma	Monitoring neuroectodermal tumors
PSA	Plasma	Monitoring prostate cancer
SCC	Plasma	Monitoring squamous cancer of cervix, lung etc.
<i>Nutritional factors</i>		
Vitamin B <sub>12</sub>	Plasma	Investigation of anemia, B <sub>12</sub> status
Folic acid	Plasma	Investigation of anemia, folate status, malabsorption
Vitamin D	Plasma	Investigation of metabolic bone disease
25-Hydroxy-vitamin D	Plasma	Investigation of metabolic bone disease
1,25-Dihydroxy-vitamin D	Plasma	Investigation of hypercalcemia and bone disease

ACTH, adrenocorticotrophic hormone; APUD, amine precursor uptake and decarboxylation; AVP, arginine vasopressin; CEA, carcinoembryonic antigen; CK-MB, creatine kinase MB isoform; CSF, cerebrospinal fluid; DHAS, dehydroepiandrosterone sulfate; hCG, human chorionadotrophin; HIAA, hydroxyindoleacetic acid; FSH, follicle-stimulating hormone; IGF, insulin-like growth factor; LH, lutenizing hormone; NICTH, non-islet-cell tumor hypoglycemia; NSE, neuron-specific enolase; PSA, prostate-specific antigen; SCC, squamous cell carcinoma; TSH, thyroid-stimulating hormone; VIP, vasoactive internal polypeptide.

The immunometric-type assay has also been adapted for use with nonisotopic labels and is typically carried out in a heterogeneous format in which the antibody is immobilized on a solid support, such as a microtiter dish, membrane, or collection of beads. The canonical clinical immunoassay format in today's laboratories is the enzyme-linked immunosorbent sandwich assay, which employs two antibodies, one to capture the analyte and the other to detect and quantify it. More details of the principles of these and other immunoassay techniques are given elsewhere in this encyclopedia.

## Specific Applications

### Detection of Polypeptide Hormones

The first clinical immunoassays were developed for the detection of relatively large polypeptide hormones such as insulin and glucagon, since these are naturally immunogenic and can therefore stimulate the production of antibodies when injected into animals. Antibodies with binding characteristics suitable for competition assays were produced for a range of polypeptides with molecular masses in excess of 3 kDa, but this was not a suitable approach for the production of antibodies against smaller peptides, thyroid hormones, or small-molecule drugs. Solid-phase immunoassays have also been developed for a broad range of polypeptide hormones. The most widely used of these is probably the test for human chorionadotrophin (hCG) in urine, which is the underlying principle of home pregnancy-test kits.

### Detection of Steroid and Amino Acid/Fatty Acid-Derived Hormones

With the rediscovery of Landsteiner's observations that even very small molecules could be rendered immunogenic by linking them covalently to a protein, it became possible to develop immunoassays for almost every molecule of clinical interest. The main medical application of immunoassays remained in endocrinology, producing a revolution in clinical practice by facilitating the detection and quantitation of steroids (such as aldosterone and cortisol) and of hormones derived by the biotransformation of single amino acids and fatty acids (such as thyroxine, melatonin, and 5-hydroxytryptophan). Not only is it possible to measure these hormones at the infinitesimally low concentrations at which they naturally occur in tissues, plasma, and other body fluids, but the assays are sufficiently rapid and economical to be useful in day-to-day patient care.

### Detection of Therapeutic Agents and Drugs of Abuse

Once the raising of antibodies to small molecules ceased to be a technical problem, the application of immunoassays to the detection and measurement of drugs in body fluids as part of the process of therapeutic drug monitoring became commonplace. The number of drugs for which immunoassays have been developed is extensive, although comparatively few have found a place in regular clinical practice (Table 2). Similar techniques have been employed for the semi-quantitative detection of drugs of abuse in urine, blood, saliva, and body tissues. In this context,

**Table 2** Clinically useful drug immunoassays

<b>Therapeutic drug monitoring</b>		
<i>Anticonvulsants</i>		
Carbamazepine	Ethosuximide	Phenobarbitone
Phenytoin	Primidone	Valproic acid
<i>Antibiotics</i>		
Amikacin		Chloramphenicol
Gentamicin		
Netilmicin	Tobramycin	Vancomycin
<i>Cardioactive drugs</i>		
Digoxin	Diisopyramide	Lignocaine
Procaineamide	Quinidine	
<i>Psychoactive drugs</i>		
Amitriptyline	Desipramine	Imipramine
Nortriptyline		
<i>Others</i>		
Caffeine	Cyclosporin A	Methotrexate
Theophylline		
<b>Drugs of abuse screening</b>		
Amphetamines	Barbiturates	Benzodiazepines
Canabinoids	Cocaine	LSD
Opiates	Phencyclidine	

immunoassays are advantageous over most of the other available methods for drugs testing because they are sensitive, inexpensive, and simple to perform. This has led to them being used as screening techniques in sport, where drug abuse is common, as well as in more commonplace clinical settings.

### Blood Products, Enzymes, and Other Proteins

Increasingly, plasma proteins that were once measured exclusively by their catalytic activities and expressed in terms of enzyme units are now being measured by immunoassay techniques, and expressed in terms of concentration (e.g., micrograms per liter). When the results from these very different forms of analysis are compared, there is often little correlation. In many cases, this reflects the fact that proteins can exist in multiple structural forms, differing in conformational properties, in terms of proteolytic cleavage, or in terms of specific forms of post-translational modification (e.g., phosphorylation). While these forms may retain some conserved epitopes, and thus have common immunological properties, structural differences often have a profound effect on their activities. For example, it has been found that the blood protein factor VIII, which is one of the key proteins in the blood clotting cascade, loses its catalytic activity in hemophilia while

retaining many of its immunological properties. Consequently, while factor VIII may appear to be absent from the blood when measured in terms of its ability to convert fibrinogen into fibrin, is found to be present at normal levels when measured by immunoassay.

### Disease Markers

In the context of disease, immunoassays were first used for the management of patients suffering from choriocarcinoma, whose tumors produce excessive amounts of hCG. Until the 1970s, this disease was invariably fatal within a few months of diagnosis. However, with the advent of specific chemotherapeutic treatments for tumors of trophoblastic origin and the development of ultrasensitive immunoassays for hCG in plasma as a guide to how long therapy should be continued, the prognosis for complete recovery is now greater than 90%. By ensuring that treatment continues until hCG can no longer be detected, it becomes possible to eliminate every neoplastic cell in the body and avoid recurrence.

The application of immunoassay techniques to the detection and quantification of antigens specifically present or absent in disease has grown enormously since these early successes. Indeed, at the current time, it is estimated that over 1000 recombinant antibodies recognizing disease-specific targets are being developed by biopharmaceutical companies around the world. A large proportion of these antibodies recognize cancer antigens but others have been developed for the diagnosis (and treatment) of infectious diseases, autoimmune disorders, as well as blood, neurological, cardiovascular, skin, respiratory, and eye diseases.

Immunoassay techniques can be used not only for the detection and quantitation of disease-specific markers, but also for their discovery. This has been facilitated by the development of highly multiplexed immunoassay formats, otherwise known as antibody arrays. For example, an array containing 146 antibodies recognizing proteins involved in cell cycle regulation, stress response, and apoptosis has been produced for the screening of tumor samples. VoLo carcinoma cells were irradiated with a cobalt-60 source and cultured for 4 h before protein extracts were obtained and labeled with the fluorogenic molecule Cy3. Protein extracts from parallel cultures of nontreated cells were labeled with Cy5, which is similar to Cy3 but has a different emission wavelength, allowing protein levels in the two samples to be compared on the same array. These experiments identified 11 proteins that were upregulated in colon cancer, six of which were previously not known to be

involved. Most of these proteins had roles in apoptosis, and increased apoptosis of the cells was observed after radiation treatment. Another protein, the carcinoembryonic antigen (which is used in the immunodiagnosis of colorectal cancer), was shown to be downregulated.

Although there have been several reports of novel biomarkers identified using antibody arrays, it is rare to find a single diagnostic marker that is reliable in a clinical setting. However, the ability to determine the levels of hundreds of protein simultaneously allows the overall pattern of protein expression to be used as a diagnostic tool, a method known as pattern profiling (Figure 1). Disease diagnosis is achieved by feeding the antibody array into a mass spectrometer and looking at the mass spectra produced by surface-enhanced laser desorption/ionization (SELDI). This provides higher sensitivity than the analysis of single biomarkers, which are often expressed in multiple diseases making a precise diagnosis difficult. This is especially true in closely related diseases, such as different forms of cancer or dementia.

A useful example of SELDI pattern profiling is the early diagnosis of ovarian cancer, a disease that is usually detected at the late stage when cancer cells have already spread and the prognosis is poor. In the original study, mass spectra derived from the serum samples of women with ovarian cancer and from unaffected controls were used as a training set for a pattern-matching algorithm. A discriminatory pattern was identified, which was applied to another set of samples. This resulted in the correct diagnosis of all ovarian cancers (including 18 stage I cancers, where the prognosis is favorable because the neoplastic cells are still contained within the ovary) and a false positive rate of only 5%. Similar algorithms have been used to diagnose breast and prostate cancers. In each

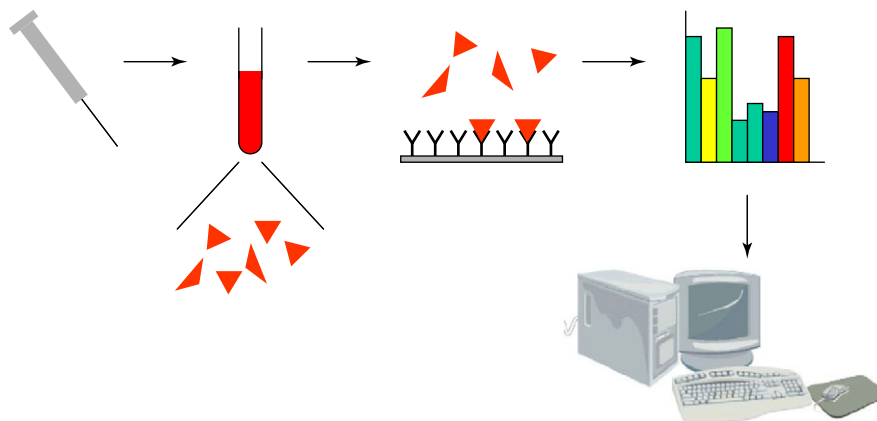
case, the sensitivity of the pattern-profiling method has been significantly higher than tests relying on the presence of single biomarkers, and in at least one study the sensitivity of the method has been high enough to achieve 100% correct diagnosis. This new approach to disease detection could revolutionize the clinical laboratory, providing the means to process more samples than before and detect diseases earlier and more accurately than is currently possible using standard immunoassay methods.

## Limitations of Immunoassays in Clinical Settings

### Specificity

The early claims for the almost absolute specificity of immunoassays for particular antigens were based more upon what was then understood about the specificity of antibody–antigen interactions than on empirical evidence. For example, antibodies raised against insulin did not react against any other proteins known at the time, or against any other known substance. It soon became clear, however, that ‘specificity’ in the context of an immunoassay is a relative term. Nowhere was that message clearer than in relation to glucagon, which was among the first few polypeptide hormones for which radioimmunoassays were developed.

Antisera raised against pancreatic glucagon, the only form of the hormone known at the time, reacted with substances that could be extracted from the gut, and which were also present in the circulation, but with sufficient differences from native glucagon to suggest their nonidentity. These substances differed from glucagon not only immunologically but also in their biological properties. Portions of the molecule



**Figure 1** Protein patterns in disease diagnosis. A blood sample contains many proteins, which can be captured and quantified on an antibody array. The relative abundances of the proteins provide a unique signature, or fingerprint, which can be detected by specialized algorithms and used to diagnose disease and classify different forms of tumor. (Reproduced with permission from Twyman RM (2004) *Principles of Proteomics*. Abingdon, UK: BIOS/Garland Publishers.)



that were essential for biological activity were not necessarily immunogenic, while portions that were extremely immunogenic were seemingly irrelevant in terms of biological activity.

Because many molecular variants of a hormone with shared epitopes but different biological properties may exist in the plasma at the same time, it may be very difficult to measure any one of them accurately and precisely without their prior separation by chromatography or a similar method. Investigations stemming from the problem of immunoassay specificity have undoubtedly led to a greatly increased understanding of the nature of immunological determinants or epitopes. The knowledge so gained has made it possible to develop improved immunoassays with more appropriate specificities. This has been particularly important, for example, in the development of immunoassays for parathyroid hormone in plasma, where an ability to differentiate between the active hormone and its inactive degradation products is essential if the results of the assay are to have any clinical validity. Ironically, while the situation for insulin is similar (in that insulin immunoassays pick up not only mature insulin, but also proinsulin and split insulin fragments that continually circulate in the plasma) the pan-specificity of the immunoassay has been clinically beneficial, as the following example shows.

Despite the enormous contribution that the ability to measure insulin has made to our understanding of disease (especially diabetes), the most reliable clinical indication for abnormally high insulin levels is hypoglycemia (low blood glucose). One of the most important of the uniformly rare diseases causing hypoglycemia is a small benign tumor of the pancreas often referred to as an insulinoma. Left untreated, such tumors can cause potentially fatal hypoglycemia, but they can be cured by surgical excision. The biochemical hallmark of insulinoma is the continued secretion of insulin in the presence of low blood glucose, which normally inhibits it. However, many such tumors produce large amounts of proinsulin, either alone or in addition to insulin. Therefore, the ability of insulin immunoassays to detect these tumors relies upon the antibody recognizing proinsulin as well as the mature form. If the clinical insulin assays were absolutely specific for mature insulin, their usefulness would have been reduced rather than enhanced! Unfortunately, other diseases, including insulin resistance, require accurate measurements of the levels of mature insulin in the blood. Attempts to investigate insulin resistance using comparatively nonspecific immunoassays that also recognize unprocessed forms of the protein are likely to yield unreliable results.

The problem of antibody specificity has also arisen in the context of multiplex immunoassays (antibody arrays). This reflects the fact that most antibody arrays are currently used to detect particular, restricted classes of proteins, often cytokines or other secreted factors that are released into the serum or culture medium. Such antibodies have been developed especially for serum profiling and in many cases have not been checked for broader cross-reactivity, e.g., in cell lysates. In the few studies that have addressed this issue, the data suggest that up to 50% of antibodies used on chips cross-react with nontarget antigens. The proteins in a typical clinical sample cover a broad dynamic range, so antibodies with high affinity for a scarce target analyte and low cross-affinity for an abundant nontarget analyte might bind both equally well. This would provide a completely false indication of the relative abundances of the two analytes in the sample. It is therefore likely that many of the antibodies currently used for single-target immunoassays will be unsuitable for antibody arrays.

### **Monoclonal, Polyclonal, and Recombinant Antibodies**

Although the lack of specificity described above was at first thought to be confined to assays using polyclonal antibodies and would not occur with assays constructed using monoclonal antibodies, it is now quite clear that this was based on a misconception. The problem of cross-reactivity is not so much to do with the multiple antibody types present in polyclonal sera but instead reflects the conservation of epitopes between target and nontarget proteins, such as the processed and unprocessed forms of insulin. The great advantage of monoclonal antibodies is that, because they are produced from an immortal strain of cells rather than a single (mortal) animal, they can be characterized much more thoroughly than a polyclonal antibody. Also, from the point of view of a kit manufacturer, continuity of supply can be guaranteed.

Cross-reactivity can be reduced using the sandwich assay approach, because two noncompeting antibodies (i.e., antibodies recognizing different epitopes of the antigen) are required for each target protein. However, even this does not guarantee specificity. More recent approaches to antibody generation, such as phage display, can be used to circumvent some of the limitations of conventional hybridoma technology and select antibodies that show specificity for particular variants of a given protein. However, maximum specificity and sensitivity requires a combination of immunoassay and physical separation techniques, such as electrophoresis or liquid chromatography.

**Table 3** Protein-bound analytes of major clinical significance

Analyte	Binding protein	Factors influencing abundance of binding protein
Cortisol	Transcortin	Increased by estrogens, pregnancy
IGF-I, IGF-II	IGF-binding proteins (IGF-BPs)	Growth hormone increases IGF-BP 1 and 3 Insulin decreases IGF-BP 1
Insulin	Autoantibodies	Autoimmune disease
Testosterone	Sex hormone binding globulin	Estrogen, thyroxine and pregnancy increase, testosterone decreases
Thyroxine and T <sub>3</sub>	Prealbumin	Malnutrition decreases
	Thyroid binding globulin	Estrogens, pregnancy increases Adrogens, malnutrition decreases
Vitamin A	Retinol-binding protein	Zinc deficiency decreases

This approach has been used, for example, to measure tetrahydrocannabinol (the active ingredient of cannabis) in blood plasma in the presence of much higher concentrations of its inactive metabolites, and insulin in the presence of proinsulin and its split products.

### Standardization

It is an absolute requirement of all validated immunoassays, whether competitive or immunometric, that for accurate quantitative analysis the standard material used to calibrate the reaction is identical to the analyte. This may be difficult for several reasons, including the chirality of the analyte and its heterogeneity in terms of post-translational modification.

Chirality is generally only a problem in clinical applications when the assay involves a comparatively small molecule, such as a drug or a small synthetic peptide. Antibodies, like most biological reagents, exhibit 'handedness' or chiral specificity. Therefore, where the analyte occurs as a mixture of enantiomeric forms, only one enantiomer may bind to the antibody, or different enantiomers may bind with different affinities. The relevance of the standard depends on whether it is presented as a pure enantiomeric form, or as a mixture of enantiomers that reflects the composition of the analyte.

Where the analyte is a protein, variations in the type and amount of post-translational modification are important. Most proteins produced in mammalian cells are modified either by phosphorylation, glycosylation, or some other form of chemical adduct. In the case of phosphorylation, a protein may have several target sites for the addition of phosphate groups and may exist as a complex mixture of phosphoforms, not all of which will bind to the antibody with equal affinity. The situation with glycoproteins is even more complex, since there may be hundreds of different variants. In many cases, the protein used as a standard in immunoassays is a

recombinant protein produced in a heterologous expression system. It should be noted that mammalian proteins synthesized in bacteria are neither phosphorylated (at least not on the typical serine, threonine, and tyrosine residues) nor glycosylated. Additionally, the glycan structures produced in different expression systems are very diverse, and even mouse cells produce subtly different glycans to human ones. A further consequence of this diversity is that the results obtained using one manufacturer's antiserum, label, and standards are generally incompatible with those obtained using another's.

### Interference from Binding Proteins

Most clinical immunoassays require neither prior extraction nor purification of the sample containing the analyte before they are added to the reaction containing the appropriate amounts of high-affinity specific antibody and label. This is not the case, however, for analytes such as thyroxine or cortisol, which are tightly bound to highly avid binding proteins that compete with the antibody for both the analyte and the label. Some of the more important protein-bound analytes are shown in **Table 3**.

Similar constraints apply when the patient has developed antibodies of their own to the analyte, either as a result of autoimmune disease (e.g., in the case of thyroxine) or prior immunization (e.g., in the case of insulin or growth hormone). Under these circumstances, extraction of the sample prior to analysis is essential if reliable and accurate results are to be obtained.

**See also:** Immunoassays: Overview; Production of Antibodies. **Immunoassays, Applications:** Forensic. **Immunoassays, Techniques:** Enzyme Immunoassays.

### Further Reading

Bock JL (2000) The new era of automated immunoassay. *American Journal of Clinical Pathology* 113: 628–646.

- Borrebaeck CAK (2000) Antibodies in diagnostics – from immunoassays to protein chips. *Immunology Today* 21: 379–382.
- Coakley WA, Andreas CM, and Jacobowitz SM (1996) Quality assurance indicators for immunoassay test kits. *ACS Symposium Series* 646: 254–264.
- Fare TL, Sandberg RG, and Herzog DP (1996) Considerations in immunoassay calibration. *ACS Symposium Series* 646: 240–253.
- Gosling JP (1990) A decade of development in immunoassay methodology. *Clinical Chemistry* 36: 1408–1427.
- Hubbard R and Marks V (1988) *Clinical Applications of Monoclonal Antibodies*. London: Plenum Press.
- Ng JH and Ilag LL (2002) Biomedical applications of protein chips. *Journal of Cellular and Molecular Medicine* 6: 329–340.
- Petricoin EF, Zoon KC, Kohn EC, Barrett JC, and Liotta LA (2002) Clinical proteomics: Translating benchside promise into bedside reality. *Nature Reviews in Drug Discovery* 1: 683–695.
- Shannon M, Cox MN, and Baum CR (1998) Toxicology reviews: Immunoassay in detecting drugs of abuse. *Pediatric Emergency Care* 14: 372–375.
- Sokoll LJ and Chan DW (1999) Clinical analyzers. Immunoassays. *Analytical Chemistry* 71: 356R–362R.

## Food

**S Iametti**, Università degli studi di Milano, Milano, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

National and transnational regulatory frames, improved awareness of health- and quality-related issues, and growing occurrence among the general population of adverse reactions toward some food components have increased the need for specific and sensitive methods capable of assessing specific food constituents, as well as additives, adulterants, and contaminants.

In many cases, only immunoassays are able to fulfill these requirements, and thus are gaining increased recognition as regular analytical methods in laboratories involved in food research or inspection. This trend is supported by a growing array of companies that supply the materials – and develop or perfect procedures – for this type of application. The increased diffusion of these techniques has also caused enzyme-based immunoassays to become the most popular methodology, in that it does not bring up all the regulatory issues and does not require the sizable investments associated with the use of radiochemicals. In this frame, enzyme-linked immunosorbent assays (ELISA) offer the advantage of being convenient, of being suitable for automated handling of a large number of samples, and of being relatively cheap.

From the standpoint of sample preparations, food-stuffs are much more complicated systems than clinical samples, which represent the other major area of application for the different techniques used in immunoassays. Food comes in many different physical forms, and very often it has undergone technological processes (physical, chemical, enzymatic,

alone or in combination) that may strongly affect the procedures to be developed for detection and/or recovery of a given analyte. This is especially true for proteins, whose structure (including epitopic regions or sequences) is highly sensitive to process-induced changes, which result in a modification of their immunoreactivity. In the case of proteins, changes may stem from process-induced cuts in the primary structure, from the introduction of nonphysiological amino acid through adduct formation, and from alteration in their high-order structures, which may affect their physical properties (such as solubility), as well as destroy conformational epitopes or hide sequential ones. In general terms, the processes used either by the food industry or in the preparation of meals at home may alter fundamental properties of the analytes, such as their solubility in a given extractant or their relationships to other food components.

Furthermore, analytes involved in food immunoassays are of extremely different chemical nature, and are present in a very broad range of concentrations. Whereas macromolecules (such as protein, enzymes, polysaccharides, and microbial cell-wall components) are generally present in the mg per kg range, environmental contaminants, residues from farming practices, antibiotics, toxins, and pathogens are often present at much lower concentration, as low as 0.1 µg per kg. Food analytes for which immunoassays have been developed and are commercially available are listed in Table 1.

### Identification and Quantification of Selected Proteins

Immunoassays are broadly used in food analysis for detection and quantification of small quantities of proteins in a given food for three main reasons.

- Borrebaeck CAK (2000) Antibodies in diagnostics – from immunoassays to protein chips. *Immunology Today* 21: 379–382.
- Coakley WA, Andreas CM, and Jacobowitz SM (1996) Quality assurance indicators for immunoassay test kits. *ACS Symposium Series* 646: 254–264.
- Fare TL, Sandberg RG, and Herzog DP (1996) Considerations in immunoassay calibration. *ACS Symposium Series* 646: 240–253.
- Gosling JP (1990) A decade of development in immunoassay methodology. *Clinical Chemistry* 36: 1408–1427.
- Hubbard R and Marks V (1988) *Clinical Applications of Monoclonal Antibodies*. London: Plenum Press.
- Ng JH and Ilag LL (2002) Biomedical applications of protein chips. *Journal of Cellular and Molecular Medicine* 6: 329–340.
- Petricoin EF, Zoon KC, Kohn EC, Barrett JC, and Liotta LA (2002) Clinical proteomics: Translating benchside promise into bedside reality. *Nature Reviews in Drug Discovery* 1: 683–695.
- Shannon M, Cox MN, and Baum CR (1998) Toxicology reviews: Immunoassay in detecting drugs of abuse. *Pediatric Emergency Care* 14: 372–375.
- Sokoll LJ and Chan DW (1999) Clinical analyzers. Immunoassays. *Analytical Chemistry* 71: 356R–362R.

## Food

**S Iametti**, Università degli studi di Milano, Milano, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

National and transnational regulatory frames, improved awareness of health- and quality-related issues, and growing occurrence among the general population of adverse reactions toward some food components have increased the need for specific and sensitive methods capable of assessing specific food constituents, as well as additives, adulterants, and contaminants.

In many cases, only immunoassays are able to fulfill these requirements, and thus are gaining increased recognition as regular analytical methods in laboratories involved in food research or inspection. This trend is supported by a growing array of companies that supply the materials – and develop or perfect procedures – for this type of application. The increased diffusion of these techniques has also caused enzyme-based immunoassays to become the most popular methodology, in that it does not bring up all the regulatory issues and does not require the sizable investments associated with the use of radiochemicals. In this frame, enzyme-linked immunosorbent assays (ELISA) offer the advantage of being convenient, of being suitable for automated handling of a large number of samples, and of being relatively cheap.

From the standpoint of sample preparations, food-stuffs are much more complicated systems than clinical samples, which represent the other major area of application for the different techniques used in immunoassays. Food comes in many different physical forms, and very often it has undergone technological processes (physical, chemical, enzymatic,

alone or in combination) that may strongly affect the procedures to be developed for detection and/or recovery of a given analyte. This is especially true for proteins, whose structure (including epitopic regions or sequences) is highly sensitive to process-induced changes, which result in a modification of their immunoreactivity. In the case of proteins, changes may stem from process-induced cuts in the primary structure, from the introduction of nonphysiological amino acid through adduct formation, and from alteration in their high-order structures, which may affect their physical properties (such as solubility), as well as destroy conformational epitopes or hide sequential ones. In general terms, the processes used either by the food industry or in the preparation of meals at home may alter fundamental properties of the analytes, such as their solubility in a given extractant or their relationships to other food components.

Furthermore, analytes involved in food immunoassays are of extremely different chemical nature, and are present in a very broad range of concentrations. Whereas macromolecules (such as protein, enzymes, polysaccharides, and microbial cell-wall components) are generally present in the mg per kg range, environmental contaminants, residues from farming practices, antibiotics, toxins, and pathogens are often present at much lower concentration, as low as 0.1 µg per kg. Food analytes for which immunoassays have been developed and are commercially available are listed in Table 1.

### Identification and Quantification of Selected Proteins

Immunoassays are broadly used in food analysis for detection and quantification of small quantities of proteins in a given food for three main reasons.



**Table 1** Analytes detected and quantified in food immunoassays

<i>Role</i>	<i>Chemical nature</i>	<i>Categories or examples</i>
Quality markers	Proteins Amino acid derivatives	Naturally occurring, species-specific; process sensitive Formed in processing
Allergens	Proteins	Naturally occurring in: nuts, fruits, gluten, eggs, milk, etc.
Additives	Proteins Polysaccharides	Enzymes used for processing purposes Gums
Adulterants	Proteins	Proteins from cheaper sources
Risk factors	Proteins	Spinal and CNS tissue
Residues	Small molecules	Antibiotics, hormones, anabolic agents, pesticides, drugs
Contaminants	Whole cells Small molecules	Pathogens Bacterial and fungal toxins

The first is to detect the possible use of cheap ingredients to supplement high-quality ones. Typical examples include: (1) the addition to beef and pork of cheaper meats (such as sheep, goat, horse, rabbit, or kangaroo), or of soyabean proteins; (2) the addition of cow's milk to sheep's and goat's milk and cheese; (3) the replacement of malted barley by rice or maize; and (4) the replacement of hazelnuts with various cheaper nuts.

The second is the detection and quantification of proteins that are involved in intolerance or in allergic response to some foods by sensitive individuals. Typical examples include the detection of: (1) gluten in food to be consumed by coeliac individuals; (2) milk allergens such as caseins or beta-lactoglobulin; (3) egg protein allergens; and (4) allergens of vegetable origin, such as those found in nuts, almonds, beans, and in many fruits.

The third is the detection of tissue-specific and/or species-specific antigens, as in the assays developed to detect cerebral or spinal tissues that may be present in food and feed and that may be linked to transmissible encephalopathies.

In all the cases, the main problem with protein assay in foodstuff is the different sensitivity of the various proteins to the large variety of processes used for food preparation/preservation. For example, all heat treatments affect the high-order structure in proteins (and therefore conformational epitopes), and quite often result in protein insolubilization, thus making it difficult to recover the target analyte from the processed food.

These difficulties have been circumvented by two different approaches. The first relies on the choice of heat-resistant protein antigens as the analyte. These antigens may be represented by carefully selected intact proteins (as exemplified by the choice of  $\omega$ -gliadins for gluten detection), or by antigenic peptide fragments that are not as sensitive to treatments as the intact protein they come from. The second approach to the detection/quantification of process-sensitive proteins calls for the solubilization of the

denatured protein analyte by suitable combinations of solubilizing agents (typically, by mixtures of chaotropes such as urea and guanidine, detergents, and disulfide-reducing agents). In this case, it is worth to remember that a protein denatured under similar conditions is best suited as an antigen when developing and isolating the antibodies to be used subsequently for analytical purposes.

Besides the analytical uses discussed above, protein immunoassays also may be useful for assessing the suitability of raw materials for production of a given food, and for controlling different stages of processing. For instance, epitope-specific monoclonal antibodies have been used for studying residual protein components in the course of brewing, or for distinguishing between normal and heat-damaged gums. Process-induced disappearance of milk proteins and formation of specific peptides during cheese ripening have also been assessed by immunochemical techniques.

Protein immunochemistry has been used also to verify the presence, the nature, and the origin of a number of enzymes currently used in the food industry, since immunochemical techniques allow detection of these additions even after the original activity of the enzyme is lost upon stabilization of the product (e.g., by heat treatment or by addition of enzyme-inactivating agents), or when measurement of the enzyme activity does not provide hints as for its origin (as is the case for many renneting enzymes used in the dairy industry).

In many cases, for the purpose of protein identification, immunochemical techniques have been combined with separation methods and with other methodologies. A good example of a combined procedure is the identification of the protease-resistant pathological form of the prion protein, in which the products of specific proteolysis of the whole tissue are first separated by electrophoresis, and identified by specific monoclonal antibodies after transfer (blotting) to a suitable membrane. Another common application of immunoblotting techniques is the identification of the food proteins (and of peptides



derived from their hydrolysis) that may be responsible of the allergic response in sensitive individuals. In this case, blotted proteins/peptides may be tested for immunoreactivity toward the patient's serum, thus providing information as for the nature of the allergens that have sensitized a given individual.

## Microorganisms and Microorganism-Derived Toxins

Modern, sensitive immunoassays based on an ELISA format are gradually replacing several of the serological methods (agglutination, precipitation) widely applied in food microbiology for the differentiation of bacterial and fungal strains. Most immunoassays are based on the occurrence of specific antigens in cell membranes or in the cytoplasm, although in some cases extracellular polysaccharides, toxins, or enzymes are detected. Some microbial species and strains for which immunoassays have been developed and for which analytical kits are commercially available are listed in **Table 2**, which also provides information as for the analysis format.

In the field of food microbiology, immunochemical methods are advantageous with respect to microbiological methods in terms of labor and time, in particular for those species and strains whose isolation and identification call for lengthy and cumbersome procedures. Immunochemical procedures allow cutting down substantially on these requirements, by following different strategies.

In some cases, as for *Campylobacter*, a one-broth, two-step enrichment protocol is immediately followed by a rapid immunoassay. In other cases, as for

*Listeria*, after a preliminary enrichment step, a quick immunoassay step is used to establish possible positives, in which the presence of *Listeria* will be confirmed through microbiological tests. In the case of *Salmonella*, a preliminary enrichment step is followed by growth of the bacteria in a medium capable of stimulating production of the somatic and flagellar proteins that represent the specific reactive antigens detectable in subsequent immunoassay.

When the food contaminant acts through the production of bacterial toxins, which may be heat-resistant and be present also when no viable cells are present, antibodies may be raised against the toxin itself, or against different subsets of toxin. Such approaches have been used for developing immunoassays for the verocytotoxin producing serotypes (VTEC) of the *Escherichia* genus that are well known causes of diarrhoea in humans, and include the serotype O157, currently the most prominent food-borne agent connected to enterohemorrhagic *Escherichia coli* (EHEC) infections. A similar case is the enterotoxins produced by certain strains of *Staphylococcus aureus*. Staphylococcal enterotoxins (SET) have almost total resistance to dehydration, proteolytic enzymes, and heat treatment. The traditional methods used to detect SET are costly and lack the level of sensitivity required for industrial use. Available tests employ monoclonal antibodies that react specifically with food-borne SET A, B, C, D, and E. Compared with the microslide double-diffusion reference method, which is laborious, slow, and inadequate in detecting low levels of contamination, ELISA-format kits are simple, rapid, and highly sensitive, with detection limits of less than 1 ng per g or ml<sup>-1</sup>. Results are available within 3 h, including sample preparation.

Often, commercial kits aimed at the detection of food pathogens are available in both 'card' and 'plate' formats, with the card format being suited for the purpose of rapid, preliminary screening, and the plate format – which typically requires skilled operators and a minimum of lab equipment – being more suited whenever reliable results are required.

As made evident in **Table 2**, immunoassays are also in routine use for analysis of mycotoxins, a group including the most dangerous and analytically elusive food-related toxins. In spite of the small structural differences within specific groups of mycotoxins (aflatoxins, ochratoxins, trichothecenes), assay specificity toward the other mycotoxin groups is always total. Determination of AFB<sub>1</sub>, toxin T-2, and ochratoxin A in a single extract from barley grains has been reported. Mycotoxins may be determined in crude extracts of various foodstuffs at concentration over 20 ng per kg. This detection limit can

**Table 2** Availability of commercial immunochemical kits for microorganism and toxins

Microorganism/toxin detected	Food analyzed	Assay format
<i>Salmonella</i>	Various foods	Card, ELISA
<i>Listeria</i>	Milk, cheese, meat	ELISA
<i>E. coli</i> O157	Meat products	Card, ELISA
<i>Clostridium tyrobutyricum</i>	Milk	ELISA
<i>Staphylococcus aureus</i>	Meat products, beans	Tube, ELISA
<i>Campylobacter</i>	Chicken products	ELISA
Aflatoxin M1	Milk, cheese	ELISA
Aflatoxin B1	Cereals, cocoa, peanuts	ELISA
Zearalenone	Maize, wheat, milk	ELISA
Ochratoxin A	Cereals, feed	ELISA
Deoxynivalenol	Cereals, beer	ELISA
Seaweed toxins	Fish	ELISA
Total aflatoxins	Cereals	Card

be further improved by fast clean-up of the extract on solid-phase extraction cartridges.

In the frame of mycotoxin analysis, it is worth remembering that immunoaffinity columns are used for enrichment of samples to be analyzed by solution chemistry or by liquid chromatography. The use of immunoaffinity in this case allows a one-step removal of possible interfering agents from the sample.

## Residues and Contaminants

The residues of many contaminant compounds in food are present at extremely low concentrations (0.01–100 µg per kg). Preliminary extraction and purification steps are often necessary before measuring such low concentrations, calling for specialized laboratories, expensive procedures, and high standards of expertise. Furthermore, traditional methods based on liquid or gas chromatography (coupled to mass spectrometry when necessary) do not easily accommodate a large number of samples.

The high sensitivity and specificity of immunoassays make it possible to minimize the sample size and the volume of extracting solvents, and often allows to eliminate time-consuming purification procedures altogether. However, in the particular case of small molecules with similar chemical properties, immunochemical cross-reactivity may impair to some extent the selectivity of the assay, making it most suitable for screening purposes rather than for quantification of a single compound, as will be discussed in what follows. On the other hand, this cross-selectivity makes immunoassays the best choice when applied to the detection of a family of analytes for semiquantitative screening of a large number of samples or for monitoring a process or a process ingredient in continuous-flow fashion.

Although immunoassays have been developed at the laboratory scale for a broad variety of possible food contaminants and of natural food components whose presence may affect the sensory or nutritional properties of a given food, we will discuss in what follows only those applications that have already gained acceptance in the analytical laboratory, and for which commercial kits are available, as reported in Table 3.

Antibiotics and other drugs have been applied to farm animals ranging from catfish to honeybees, for the prevention or treatment of diseases, elimination of parasites, or growth promotion. Transfer of these substances to humans through the food chain is a serious concern, and antibiotic levels in various foods are often the subject of strict regulation. Immunoassays developed for these molecules are highly

**Table 3** Availability of commercial ELISA kits for residues

<i>Residue</i>	<i>Analyzed material</i>
<i>Antibiotics</i>	
Chloramphenicol	Meat, eggs
Sulfamethazine	Meat, milk
Streptomycin	Meat, milk
Tetracycline	Meat, milk, eggs
Nitrofurantoin	Meat, milk
Sulphathiazole	Honey
<i>Pesticides</i>	
Triazine	Maize
Isoproturon	Barley, wheat
Alachlor	Maize
<i>Hormones and <math>\beta</math>-agonists</i>	
Clenbuterol	Meat, liver, powder milk
Acetylgestagene	Fat
Melengestrol acetate	Meat
17 $\beta$ -Estradiol	Plasma
Testosterone	Plasma
Ethinylestradiol	Plasma, urine
Trenbolone	Meat
Zeranol	Urine
Methyltestosterone	Urine
Nortestosterone	Meat
Dexamethasone	Meat, milk
Salbutamol	Meat, liver, powder milk

sensitive and specific, and often require only a minimal clean-up after the initial high-efficiency extraction from the tissue or food.

The cheap synthetic anabolic agents used to increase the efficiency of meat production are another category of contaminants that has harmful effects on the human consumer, and therefore are banned in almost all countries. Successful analytical procedures for hormones and  $\beta$ -agonists combine liquid chromatography separation with sensitive quantification by immunochemistry. Since often these analyses are carried out in large and dedicated institutional laboratories, this is one case in which the radioimmunological methods typical of the clinical lab (such as radioimmunoassay and several types of surface-based immunochemistry) are still enjoying great popularity. This attitude also reflects in the sampling procedure, which often involves taking samples of biological fluids from live animals prior to slaughtering.

In recent years, immunoassays for pesticides, phyto-pharmaceuticals, and industrial pollutants (polychlorinated biphenyls, dioxins, etc.) have become accepted as methods that complement traditional analytical procedures also in the field of food analysis (including drinking water). In the case of pesticides, both quantitative and semiquantitative kits are available. Cross-reactivity may occur – although with different efficiency – between an antibody raised against a protein conjugate of a given compound, and structurally

similar compounds in the same class. This can be exploited to provide a very fast estimate of total contaminants (within a given class) in a food sample.

### **Practical Considerations, Assessment of Assay Performance, and Comparison with Alternative Procedures**

Rapid screening tests such as the so-called 'test-tube', 'spot', and 'dipstick' tests available for semiquantitative analysis of food have placed immunochemical detection within the reach of nontrained operators aiming at a coarse first look at foods to be used for home consumption.

Although the procedures involved in performing immunoanalytical procedures seem to be simple, and the kits available for immunoassay in food have greatly increased their popularity, there are a number of caveats that must be considered to avoid mistakes, in particular when quantitative analysis is required. In this case, attention should be paid to product shelf-life, to running proper standards for each set of measurements, to strict adherence to the recommended procedures for sample preparation, and to proper dilution of the sample so that the analyte concentration falls within the range of the calibration curve. Given the typical logarithmic response of all ELISA tests, and the margin of error intrinsic to these determinations, performance of measurements at multiple dilutions of the same sample is highly recommended, also in routine work.

Issues related to sampling procedures, efficiency of extraction, cross-reactivity, and analyte recovery may often require the use of internal standards. The analyst should also be aware that interference – often unexpected – from other food components is also common in the complex matrices that characterize foods. For this reason, a protocol devised for a particular food (e.g., cow's milk) cannot be transferred to an apparently similar food (e.g., sheep's or goat's milk) without suitable experimentation and adequate validation. Removal of interfering materials – once their nature is known – may also be necessary in some cases, and may call for specific procedures, such as addition of precipitating agents, ultrafiltration, dialysis, etc. The easiest way to detect interference, if suspected, is to prepare analyte standards in buffer and in the food matrix at concentrations compatible with the assay. Assays performed with both sets of standards should produce identical curves in the absence of interference. The effects of these treatments on the analyte itself will have to be assessed by experiment.

Also, comparison of results obtained with different commercial kits is often difficult, in what they may

differ in the nature of the target antigen, in procedure used for sample preparation, and in the properties (specificity, titer) of the particular antibody used for the assay itself. Cross-reactivity is often an issue, in particular when the supplier of the antibodies does not specify whether it has been tested on a specific material.

In a still growing number of reports, immunochemical methods have been compared with conventional analytical methods, and extremely good agreement has been found, in particular when researchers were familiar with all the methods involved. In most cases, the advantages of immunochemical methods in food analysis (sensitivity, specificity, cost, ease of use, rapidity, ease of automation) outweigh the intrinsic disadvantages discussed above. In this framework it has to be noted that traditional analytical procedures suffer from the same limitations (matrix effects, interference, recovery difficulties, etc.) when applied to food systems, and that immunochemical methods often offer analytical possibilities that just are not attainable with conventional methods. The major limitation of immunochemical methods in food analysis, as of today, is that they are not readily applicable to simultaneous multianalyte analysis. On the other hand, they likely are the most appropriate choice whenever screening a large number of samples for one – or a few – known analytes.

When compared to other methods based on recognition of foreign biological matter in food, immunochemical methods still appear to have some important advantages even on high-sensitivity, polymerase chain reaction (PCR)-based analytical techniques, including quantitative PCR methods. First and foremost, immunochemical methods are aimed at the detection of a specific protein in a given food, rather than at the presence of residual DNA from a given organism. This is of paramount importance in all cases when the protein itself (and not other biological material from the same organism) is somewhat deleterious for the quality/safety of a given food. In other words, taking gluten as an example, food that tests gliadin-free by properly performed immunochemical analysis may be consumed safely by a gliadin-sensitive individual, even if it proves to contain residual wheat DNA. Also, the application of PCR-based techniques to food analysis suffers from recovery and interference problems not smaller than those encountered when immunochemical techniques are used.

**See also:** **Food and Nutritional Analysis:** Contaminants; Meat and Meat Products; Dairy Products. **Hormones:** Steroids. **Immunoassays, Techniques:** Radioimmunoassays; Enzyme Immunoassays. **Pesticides.** **Polymerase Chain Reaction.** **Toxins:** Mycotoxins; Neurotoxins.

## Further Reading

- Ishikawa E (1999) *Ultrasensitive and Rapid Enzyme Immunoassay*. Amsterdam: Elsevier.
- Lessof MH (1992) *Food Intolerance*. London: Chapman & Hall.
- Morgan MRA, Smith CJ, and Williams PA (1992) *Food Safety and Quality Assurance: Applications of Immunoassay Systems*. London: Elsevier Applied Science.

- Pierson MD and Stern NJ (1986) *Foodborne Microorganisms and their Toxins: Developing Methodology*. New York: Dekker.
- Rittenburg JH (1990) *Development and Application of Immunoassays for Food Analysis*. London: Elsevier Applied Science.
- Trevino RJ and Dixon HS (1997) *Food Allergy*. New York: Thieme.

## Forensic

**S Miyaishi**, Okayama University Graduate School of Medicine & Dentistry, Okayama, Japan

**F Moriya**, Kochi University, Kochi, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The terms forensic medicine, forensic science, and forensic pathology are often confused. However, forensic medicine or science in the broad sense can be classified into three main fields: forensic pathology, forensic toxicology, and forensic biology. Forensic pathology deals with the issues diagnosed at forensic (and administrative) autopsy. In forensic toxicology, all problems in relation to substances with pharmacological/poisonous effects on humans are investigated. Forensic biology is the field in which genetic markers or proteins are investigated in cases of stain analysis, autopsy of unidentified bodies, and paternity tests.

Immunoassay commonly means a method for quantitative analysis of a substance using immunoreactions. However, the original meaning of the word 'assay' is not restricted to quantitative analysis and enzyme-linked immunosorbent assay (ELISA) has been developed in the area of histology as a qualitative investigation method to identify the existence/localization of antigens (immunohistochemistry). Besides, forensic analysis frequently requires a qualitative judgment and it is often achieved by examination using immunoreactions. Therefore, in this article we deal with the word 'immunoassay' in the broad sense including both quantitative and qualitative/semiquantitative analyses using immunoreactions.

## Immunoassays in Forensic Pathology

### General Aspects

Forensic pathology involves the diagnosis of cause of death, determination of time of death, estimation of

degree and properties of injuries, etc. Analysis of morphological changes and pathogenesis induced by external factors or internal diseases are also important issues in this field. However, it is sometimes very difficult to solve such problems only by autopsy findings and attempts have been made to develop helpful examinations for forensic pathological diagnosis. For this purpose, proteins/hormones as clinically useful markers have also been utilized in this field; many authors have determined quantitatively such substances in autopsy materials using commercially available immunoassay kits and have discussed the results obtained.

We must pay attention, however, to the fact that the nature of autopsy materials is quite different from those used in clinical medicine. A blood sample obtained at autopsy is sometimes extremely hemolyzed and its supernatant is neither serum nor plasma, because postmortem intensive fibrinolysis following blood coagulation had already occurred in the corpse at autopsy. Body fluids other than blood/urine or extract of tissue-homogenate cannot normally be considered for clinical use. Most important is that many kinds of proteins are contained in extremely large amounts in autopsy material, which may be available for forensic examination and can by no means be expected to be used in clinical laboratory medicine. Such samples can cause very strong suppression of necessary immunoreactions in the assay system established for clinical use. This means that the values determined may not be comparable if different immunoassay kits are used, even if the same protein is assayed. Besides, the value for a postmortem sample may be false if its assay is completely carried out in a clinical laboratory where no one knows about the particular nature of the autopsy material. This is the probable reason that some conclusions obtained from different studies on the same substance are sometimes contradictory. In spite of such circumstances, the attempted quantitative analysis has been reported as useful in most of the earlier studies.



## Further Reading

- Ishikawa E (1999) *Ultrasensitive and Rapid Enzyme Immunoassay*. Amsterdam: Elsevier.
- Lessof MH (1992) *Food Intolerance*. London: Chapman & Hall.
- Morgan MRA, Smith CJ, and Williams PA (1992) *Food Safety and Quality Assurance: Applications of Immunoassay Systems*. London: Elsevier Applied Science.

- Pierson MD and Stern NJ (1986) *Foodborne Microorganisms and their Toxins: Developing Methodology*. New York: Dekker.
- Rittenburg JH (1990) *Development and Application of Immunoassays for Food Analysis*. London: Elsevier Applied Science.
- Trevino RJ and Dixon HS (1997) *Food Allergy*. New York: Thieme.

## Forensic

**S Miyaishi**, Okayama University Graduate School of Medicine & Dentistry, Okayama, Japan

**F Moriya**, Kochi University, Kochi, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The terms forensic medicine, forensic science, and forensic pathology are often confused. However, forensic medicine or science in the broad sense can be classified into three main fields: forensic pathology, forensic toxicology, and forensic biology. Forensic pathology deals with the issues diagnosed at forensic (and administrative) autopsy. In forensic toxicology, all problems in relation to substances with pharmacological/poisonous effects on humans are investigated. Forensic biology is the field in which genetic markers or proteins are investigated in cases of stain analysis, autopsy of unidentified bodies, and paternity tests.

Immunoassay commonly means a method for quantitative analysis of a substance using immunoreactions. However, the original meaning of the word 'assay' is not restricted to quantitative analysis and enzyme-linked immunosorbent assay (ELISA) has been developed in the area of histology as a qualitative investigation method to identify the existence/localization of antigens (immunohistochemistry). Besides, forensic analysis frequently requires a qualitative judgment and it is often achieved by examination using immunoreactions. Therefore, in this article we deal with the word 'immunoassay' in the broad sense including both quantitative and qualitative/semiquantitative analyses using immunoreactions.

## Immunoassays in Forensic Pathology

### General Aspects

Forensic pathology involves the diagnosis of cause of death, determination of time of death, estimation of

degree and properties of injuries, etc. Analysis of morphological changes and pathogenesis induced by external factors or internal diseases are also important issues in this field. However, it is sometimes very difficult to solve such problems only by autopsy findings and attempts have been made to develop helpful examinations for forensic pathological diagnosis. For this purpose, proteins/hormones as clinically useful markers have also been utilized in this field; many authors have determined quantitatively such substances in autopsy materials using commercially available immunoassay kits and have discussed the results obtained.

We must pay attention, however, to the fact that the nature of autopsy materials is quite different from those used in clinical medicine. A blood sample obtained at autopsy is sometimes extremely hemolyzed and its supernatant is neither serum nor plasma, because postmortem intensive fibrinolysis following blood coagulation had already occurred in the corpse at autopsy. Body fluids other than blood/urine or extract of tissue-homogenate cannot normally be considered for clinical use. Most important is that many kinds of proteins are contained in extremely large amounts in autopsy material, which may be available for forensic examination and can by no means be expected to be used in clinical laboratory medicine. Such samples can cause very strong suppression of necessary immunoreactions in the assay system established for clinical use. This means that the values determined may not be comparable if different immunoassay kits are used, even if the same protein is assayed. Besides, the value for a postmortem sample may be false if its assay is completely carried out in a clinical laboratory where no one knows about the particular nature of the autopsy material. This is the probable reason that some conclusions obtained from different studies on the same substance are sometimes contradictory. In spite of such circumstances, the attempted quantitative analysis has been reported as useful in most of the earlier studies.



Many morphological studies have also been reported in this field. In these studies, several antigens are stained by immunohistochemistry and the diagnostic value of the findings is evaluated according to issues such as cause of death, nature of injury, and postmortem interval. In these studies, commercially available antibodies have been used.

The situation stated above means that the immunoassays of convenient markers are helpful for forensic pathological diagnosis. However, almost all investigations ended with a particular approach or report, and unfortunately none of them has become a routine procedure in forensic autopsy in all institutes. In the following section, we introduce the hitherto reported approaches to forensic pathological diagnosis using immunoassays. Some of them are introduced in **Table 1**.

### Practical Applications

The main issue in which qualitative immunoassay is used in forensic pathology is diagnosing the cause of death. The most frequent approach is to diagnose sudden cardiac death (SCD), myocardial infarction, and ischemic damage of the myocardium by determining the proteins in the myocardium, such as myosin, myoglobin, troponins, in blood, and/or pericardial fluid by immunoassay. The determination of the content, not activity, of creatine kinase-MB isozyme by immunoassay has also been attempted. Other approaches analyzing the relationship between the concentration of proteins in blood and cause and/or pathogenesis of death were: asphyxia by neck compression and thyroglobulin, drowning and SP-A or -D (pulmonary surfactant), electric death and

myoglobin, and so on. Some authors took note not of a special cause of death but of a special substance, e.g., SP-A in blood, CRP in blood, myoglobin in urine, and analyzed the relationship between these substances and causes of death in general.

Quantitative/semiquantitative immunoassays are applied to the analysis of the nature of an injury. The estimation of the grade and analysis of pathogenesis of brain injury were attempted by determining neuron-specific enolase (NSE) and tumor necrosis factor (TNF)- $\alpha$  in cerebrospinal fluid. Concerning the skin wound, it has been said that the fibrin-fibrinogen degradation product (FDP)-D dimer can be used as a marker of vital reaction and that the content of interleukins (IL)-6, IL-1 $\beta$ , and TNF- $\alpha$  in the tissue is useful for estimating wound age. For the discrimination of subcutaneous hemorrhage from exuded hemoglobin due to putrefaction, determination of glycophorin-A has been recommended.

Some examples of other studies using quantitative immunoassays in forensic pathology are: vascular endothelial growth factor (VEGF) contents in brain and some other organs for the diagnosis of postmortem interval/time of death, melatonin concentrations in blood, urine, and pineal gland for the same purpose, and serum triptase concentration for the diagnosis of heroin-related death.

Immunohistochemistry is the most widely used technique for qualitative immunoassay in the field of forensic pathology. The frequent approach is to diagnose cause of death, especially SCD. C5b-9 (complement complex), fibronectin, cardiac troponin, and other proteins are stained for this purpose and the usefulness of their staining has been discussed. Many

**Table 1** Approaches to forensic pathological diagnosis using immunoassay

Analyzed/identified issue	Material	Type of assay	Representative antigen(s) noticed
Diagnosis of cause of death			
Sudden cardiac death	Blood, pericardial fluid	Qt	Myosin, troponin, myoglobin
	Heart	IH	C5b-9, fibronectin, troponin
Drowning	Blood	Qt	Pulmonary surfactant
	Lung	IH	Pulmonary surfactant
Asphyxia by neck compression	Blood	Qt	Thyroglobulin
	Soft tissue of the neck	IH	Myoglobin, fibronectin, glycophorin-A
Sudden infantile death	Lung or other organs	IH	Milk components, S-100, insulin
Diagnosis of nature of injury			
Wound age	Skin	IH	Collagen, fibronectine, cytokeratin, interleukins
	Brain	IH	GFAP, thrombomodulin
Vital reaction	Skin	Qt	FDP-D dimer
	Hematoma	IH	Glycophorin-A, CD62
Diagnosis of postmortem interval	Some organs	Qt	VEGF
	Pancreas	IH	Insulin, glucagon
	Thyroid gland	IH	Thyroglobulin
	Heart	IH	Myoglobin

Qt, quantitative analysis; IH, immunohistochemistry.

studies on sudden infant death syndrome and infantile sudden death using immunostaining have also been reported. In these studies, S-100 protein, hypothalamic hormones, insulin, glial fibrillary acidic protein (GFAP), calcitonin, CD 4, SP-A, and other proteins/hormones were stained in various organs/tissues and the significance of such staining was evaluated from the forensic pathological viewpoint. An interesting approach to diagnose the cause of infantile sudden death was the immunostaining of the components of human and cow milk in the lung. Other applications of immunohistochemistry to the diagnosis of cause of death were in the investigations of the relationship between SP-A in lung and drowning, adult respiratory distress syndrome, death due to fire; pulmonary mast cell triptase and amniotic fluid embolism; glycophorin-A, myoglobin, fibronectin in muscle and asphyxia by neck compression, and so on.

There have also been many immunohistochemical studies on injury. For the estimation of the age of a skin wound, numerous substances were stained and their usefulness discussed. The stained proteins were: fibronectin, collagen, laminin, cytokeratin, CD54, IL-6, IL-1 $\beta$ , and TNF- $\alpha$ . In order to confirm the vital reaction in a skin wound, immunostaining of  $\alpha$ 1-antichymotrypsin,  $\alpha$ 2-microglobulin, fibronectin, and some other substance was reported to be useful. In the analysis of brain injury using immunostaining, the following combinations of protein(s) have been described to be useful: NSE for demonstrating vital reaction and diffuse axonal injury; Collagen IV, thrombomodulin, GFAP, and some other proteins for the estimation of the age of the wound; and HSP (heat shock protein) 70, GFAP, and some other proteins for analysis of hypoxic/ischemic brain damage. For the discrimination of antemortem bleeding from postmortem hemoglobin diffusion, immunostaining of glycophorin-A and its quantitation are useful. In another report, it has been described that immunoelectronmicroscopical staining of CD 62 and other antigens on platelets could distinguish intravital hematoma from postmortem hematoma.

Other analyses using qualitative immunoassays in forensic pathology are as follows: insulin and glucagon in pancreas, myoglobin in heart, thyroglobulin in thyroid gland, S-100, cytokeratin, and other substances in sweat gland, and other attempts for diagnosing the postmortem interval/time of death, thyrosin-phosphorylated proteins in thymus for demonstrating child abuse,  $\beta$ -amyloid, GFAP, and Tau-protein in brain for postmortem diagnosis of Alzheimer's disease, leukotoxin in lung for analyzing pathogenesis of paraquat poisoning/hyperoxic condition.

There have been some approaches in the border area between pathology and biology or toxicology.

An example is the method for identifying the used weapon in the case of injury/homicide by detecting the organ-specific antigen. If the specific antigen of the injured organ is detected from the stain on the suspected weapon, it can be identified as the weapon used in the incident. Though identifying/presuming the used weapon is one of the essential issues in forensic pathological diagnosis, stain analysis belongs principally to forensic biology (see the section 'Immunoassays in forensic biology – practical applications'). Another example is the analysis of cause and/or pathogenesis of death in a case of poisoning or death under the effect of drug/poison. Though this kind of analysis is an issue of diagnosis required at forensic autopsy, it also depends on forensic toxicological analysis: drug-related pathological changes, including immunohistochemical detection of drug/poison distributed in the body, and the quantitation of drug/poison in autopsy materials.

## Immunoassays in Forensic Biology

### General Aspects

Forensic biology has anthropological and genetic aspects. Analysis of genetic markers is indispensable for a paternity test and for the personal identification of stains or unidentified bodies. In these circumstances, examination of genetic markers is the most frequent and important business in forensic biological practice.

Up to the middle or end of the 1990s, blood types were examined using immunoreactions as important genetic markers. Many immunoassays have been developed to determine their phenotypes. The status of immunoassays in this field, however, has changed greatly in recent years. With the development of DNA technology, blood cell/serum protein types as the tool for personal identification/paternity test were substituted by STRs (short tandem repeats polymorphisms)/SNPs (single nucleotide polymorphisms) and the necessity of examining classical genetic markers was remarkably reduced. Nevertheless, immunoassay will still be necessary in forensic biology in the future, because some examinations required in this field cannot be established in DNA analysis (e.g., identification of semen).

Some commercially available immunoassay kits are applied to the analysis in forensic biology. Most of them are for qualitative/semiquantitative analysis. Although they are manufactured for clinical purposes and materials, they are also useful for forensic biological analysis with different purposes and materials. For example, an immunoassay kit for detecting prostate-specific antigen (PSA) in serum as a marker of prostate cancer is also applicable for

identifying semen using the extract obtained from a suspected dried body fluid stain (see below).

Many in-house developed ELISAs for the investigation of special issues have also been reported. They are not distributed in the market, but are worth being referred to when the same or similar investigation is required.

### Practical Applications

In the field of forensic biology, many kinds of immunoassays have been used for the examination of blood groups. A hemagglutination test using antisera is the classical method for determining blood cell types. HLA types have been determined using complement-fixation test. Some serum protein polymorphism has been examined using immunoblotting or immunofixation following isoelectric focusing.

These examinations of blood types were substituted for DNA analyses, especially STRs and SNPs. Because STRs/SNPs are much more polymorphic than the classical blood types, they have a great advantage in forensic biological practice. Besides, the polymerase chain reaction used for DNA analysis is very sensitive. Blood group determination, however, may be required in cases other than personal identification/paternity test: for example, in the case of incompatible blood transfusion. An absorption test of saliva can contribute to identifying the blood group of the recipient. The ABO blood group of the recipient may be identified in a short time if the distinct mosaic pattern with both agglutinated and nonagglutinated red blood cells is observed by a hemagglutination test. Though immunostaining of ABO blood group antigen using autopsy material is also useful for this purpose, it is now being substituted by DNA analysis.

Absorption and absorption-elution tests are the conventional methods used for determining ABO blood groups from stains. Immunohistochemistry can be applied to some stains and autopsy material. An absorption test on a body fluid stain can identify a secretor/nonsecretor. Many ELISAs to detect or quantify ABO/Lewis blood group antigen have also been reported. Although ELISAs had satisfactory specificity and sensitivity, they were not substituted for the absorption/absorption-elution tests, and ABO blood grouping by DNA analysis is now becoming a routinely used method. ELISAs using monoclonal antibodies, however, may still be useful, because some of them can determine the ABO blood group of a specific component in a mixed stain, e.g., ABO blood grouping of vaginal secretion mixed with semen.

Immunoassays have also been applied to species identification of stains. A precipitation test of the extract of the suspected stain against anti-human

serum or anti-human IgG (immunoglobulin G) is a classical method used for this purpose. Immunological detection of human Hb (hemoglobin) is now a popular method for species identification of bloodstains. This method has the advantage of not only identifying human origin but also confirming blood at the same time, compared with the methods for detecting other human components. Recently, commercially available immunoassay kits for clinical laboratory investigation to detect occult blood in feces have been applied to forensic human identification of bloodstains, because the antibodies used can reveal excellent specificity for humans. Since the last decade a highly sensitive immunochromatography device has been available. It is useful for species identification of stains from body fluids other than blood if they have only a minute contamination of Hb. However, very highly sensitive methods using DNA analysis have also been developed for this purpose.

We must note that some kinds of examinations in forensic biological practice cannot be established in DNA analysis and require immunological analysis. Identification of semen is one of the typical and important issues from this viewpoint. PSA is a well-known semen-specific substance and its detection by ELISA is the most common method for identifying semen from body fluid stain/vaginal content. Since it is an important clinical marker for prostate carcinoma, devices for detecting serum PSA using immunochromatography are marketed worldwide. Although the prepared sample is not serum for forensic biological examination, such a device is useful for detecting/identifying semen in the case of sexual assault. Gamma-seminoprotein is also a good marker for this purpose and it is said to be the same substance as PSA. ELISAs for  $\beta$ -microseminoprotein, semenogelin, and other substances were developed for identification of semen.

In bloodstain analysis, there are also some problems for which DNA analysis can give no solution. For example, in the case of murder/injury of a woman in menstruation, the offender sometimes claims that the bloodstain at the scene has originated from the blood of menstruation. In such cases the immunological determination of FDP is a useful means, because it is profusely contained in menstruation blood. There are some more examples in which immunoassay can be a useful tool for analysis. By detection of human chorionic gonadotropin (hCG) the blood of a pregnant woman can be identified, and determination of myoglobin, which is very much contained in the blood of a corpse, is useful for discrimination of antemortem and postmortem blood.

Identification of injured organ/tissue from bloodstain also requires immunological analysis, and ELISAs for the following proteins have been developed: S-100

**Table 2** Forensic biological analysis requiring immunoassay

<i>Material</i>	<i>Analyzed/identified issue</i>	<i>Representative antigen(s) noticed</i>
Bloodstain	Blood of pregnant woman	HCG
	Menstrual blood	FDP
	Postmortem blood	Myoglobin
	Injury of skeletal muscle	$\beta$ -enolase
	Injury of brain	S-100 protein
Body fluid stain/ vaginal content	Injury of liver	Liver-specific antigen
	Semen	PSA, $\gamma$ -Sm, semenogelin
Body fluid stain/ tissue fragment	Human origin	Human Hb, human IgG
Tissue fragment	Brain	Myosin, neurofilament
	Skin	Squamous cell carcinoma-related antigen
	Skeletal muscle	Myoglobin

protein (specific for brain),  $\beta$ -enolase (specific for skeletal muscle), elastase III (specific for pancreas), Tamm–Horsfall protein (specific for kidney). Liver-specific antigen was originally isolated and an ELISA of this protein was developed for the investigation of liver injury. Identification of organ/tissue itself from the fragment left at the scene is also an important issue in forensic stain analysis and some ELISAs have been reported: myosin or neurofilament for identifying brain tissue, myoglobin for identifying skeletal muscle, and squamous cell carcinoma-related antigen for identifying skin. Some ELISAs for organ/tissue-specific substances reveal a good specificity for humans at the same time. Representative analyses requiring immunoassay in forensic biology are shown in Table 2.

## Immunoassays in Forensic Toxicology

### General Aspects

Forensic toxicology refers to the broad field in which immunoassays are routinely used in forensic medicine/science. Determination of substances with pharmacological or poisonous effects on humans in autopsy materials plays a very important role in forensic toxicological practice, and detecting/determining drugs of abuse in urine/blood of living subjects is frequently required in this field. In these circumstances, many kinds of immunoassays for drugs have been developed, and they are marketed for practical use in forensic and clinical toxicology. Some immunoassays are also used to detect common drugs of abuse in extracts from hair, which has a large window for monitoring past drug history.

When detailed qualitative and precise quantitative analyses for drugs are required, we must choose instrument-based testing without immunoreactions, because the latter cannot discriminate the parent drugs from its metabolites/derivatives and chemicals that have similar structures. For example, putrefactive amines, such as phenethylamine, frequently give false positive results for amphetamines; dihydrocodeine gives positive results for opiates. However, immunoassay kits for chemical compounds are still useful, especially in screening.

The status of immunoassays in forensic toxicology is different from that in forensic pathology and biology. In the latter fields, analyses by immunoassays are tools for diagnoses, and to determine the target substance is not the final purpose of analysis. For example, a precipitation test for human IgG is a method for forensic species identification, but the essential purpose is not to detect IgG itself, but to identify its human origin, and there may be another more useful immunoassay for a new substance such as human Hb (see the section 'Immunoassays in forensic biology' above). On the contrary, immunoassay is itself the purpose in forensic toxicology. Immunoassays have been developed and investigated as methodology in this field. In the following section, we describe the present situation of immunoassays distributed in the market for forensic or clinical toxicology, and comment on their characteristics and principles. Immunohistochemistry in toxicology is briefly described in the section 'Immunoassays in forensic pathology'.

### Practical Applications

Immunoassays are used in forensic toxicology for routine screening and quantifying compounds. They have excellent sensitivity and specificity, and can be completed in a short period. Many immunoassays are available for testing common poisons in both urine and serum. The most common immunoassays used in forensic toxicology are: RIA (radioimmunoassay), EMIT<sup>®</sup> (enzyme-multiplied immunoassay technique, Behring Diagnostics), FPIA (fluorescence polarization immunoassay), CEDIA<sup>®</sup> (cloned enzyme donor immunoassay, Roche Diagnostics/Boehringer Mannheim Corporation), KIMS<sup>®</sup> (kinetic interaction of microparticles in solution, Roche Diagnostic Systems), and ELISA. The most commonly measured substances are amphetamines, barbiturates, benzodiazepines, cannabinoids, cocaine, methadone, opiates, phencyclidine, propoxyphene, and tricyclic antidepressants.

RIAs are preferred for postmortem samples due to their resistance to matrix effects. Most other techniques are difficult to apply to these specimens.



Postmortem samples are homogenized in a deproteinizing medium such as acetonitrile or a mixture of zinc sulfate/sulfosalicylic acid/methanol. As a result of their sensitivity and low matrix effect, RIAs can detect low levels of lysergic acid diethylamide (LSD) in urine, ricin in tissues, and tetrahydrocannabinol (THC) in hair. Although RIAs have many analytical advantages and  $^{125}\text{I}$ , instead of  $^{131}\text{I}$ , has been used to label drugs in most current RIA methods because of its longer half-life, there are problems in disposing of radioactive waste and using an isotope with a short shelf-life (~60 days). These shortcomings have reduced the demand for RIA and manufacturers no longer make kits for common drugs.

EMIT<sup>®</sup> is a homogeneous assay that uses an enzyme (glucose-6-phosphate dehydrogenase) labeled drug to compete with unlabeled drug for a specific antibody. If the labeled drug is free from the antibody, the enzyme oxidizes glucose-6-phosphate and reduces  $\text{NAD}^+$  to NADH, which is measured at 340 nm with a spectrophotometer. Automated analyzers are available for these immunoassays, and hundreds of samples can be processed in an hour. While the EMIT<sup>®</sup> assay was developed for drugs in urine, it has also been applied to aqueous extracts from other biological samples (e.g., meconium). If a clear and nearly neutral solution can be obtained from a sample, then the sample can be assayed by EMIT<sup>®</sup>.

FPIA methods use a fluorescein-labeled drug to compete with the unlabeled drug for an antibody. If the fluorescein-labeled drug combines with the antibody, the emitted light remains polarized when the incoming light is polarized. As the concentration of drug in the sample increases, free fluorescein-labeled drug molecules also increase in number. Since the unbound fluorescein-labeled drug molecule rotates freely, the polarized light emitted is reduced. Automated analyzers, such as TDx<sup>®</sup>, ADx<sup>®</sup>, and Axysm<sup>®</sup> (Abbott Laboratories), are used to measure these reactions. Drugs are more accurately measured by FPIA than EMIT<sup>®</sup> because sample matrices have less effect on changes in fluorescein polarization.

CEDIA<sup>®</sup> is a relatively new technique that uses two genetically engineered fragments of *Escherichia coli*  $\beta$ -galactosidase as an enzyme label. The activity of the enzyme requires assembly of the two fragments, termed enzyme donor and enzyme acceptor fragments. The enzyme donor fragment is bound to a drug. The enzyme donor fragment-labeled drug competes with the drug present in urine samples for a specific antibody. When the enzyme donor fragment-labeled drug combines with the antibody, the fragment does not associate with the enzyme acceptor fragment and the enzyme activity does not

develop. When the enzyme donor fragment-labeled drug associates with the enzyme acceptor fragment, it becomes active and hydrolyzes chlorophenol red- $\beta$ -galactoside to chlorophenol red and galactose. Production of chlorophenol red is measured at 570 nm with a spectrophotometer. The concentration of the drug of interest is directly proportional to the change in absorbance and automated analyzers are commercially available.

KIMS<sup>®</sup> assay (Abuscreen Online by Roche Diagnostic Systems) uses microparticles labeled with several drug molecules. The drugs on the microparticles compete with the drug in urine for a specific antibody. Reaction of the drug molecules on the microparticle with several antibody molecules facilitates the formation of a large aggregate that increases absorbance. Thus, the concentration of the drug is inversely related to the change in absorbance. Since microparticle drug conjugates are more stable than enzyme drug conjugates, these kits have a long shelf-life.

ELISA is a sensitive and specific method that has been used for many years to detect body components and pathogenic microorganisms. However, its application to drug testing is relatively new. Since ELISA is a heterogeneous assay, it is not as sensitive to matrix effects. Thus, this technique is easily applied to postmortem fluid samples. A commercially available kit (MTA-8<sup>®</sup>, Venture Labs) can simultaneously test for eight drugs of abuse on one microtiter plate. MICROPLATE<sup>®</sup> Forensic EIA kits (OraSure Technologies) are also marketed.

Many immunoassay kits that do not require special instrumentation are commercially available. They do not require a formal laboratory and are used for on-site testing, such as driving-under-the-influence cases, workplace drug testing, and emergency room settings. These kits are also used in toxicology laboratories to test a relatively small number of samples at low cost. On-site drug testing devices use competitive immunoassay and the results are visualized as red bands of immunoreacted colloidal gold-labeled antibodies or drugs. The kits available for forensic on-site testing of urine are summarized in Table 3. Triage<sup>®</sup> DOA has been successfully applied to postmortem blood samples if the latter are properly pre-treated. For testing drugs-of-abuse in saliva, Cozart's RapiScan<sup>®</sup> (Cozart Biosciences) is the only device currently available.

Detecting agricultural chemicals in biological samples is important in forensic toxicology. A forensic toxicologist can use enzyme immunoassay kits (Rapid Assay<sup>®</sup>, Ohmicron Environmental Diagnostics) that were originally developed to monitor environmental chemicals, such as organophosphates, carbamates,



**Table 3** Descriptions of noninstrumental immunoassay kits for qualitative on-site drug testing

Test	Analytes	Testing time required (min)	Sample volume required
Abscreen ONTRAK	AMP, BAR, BZO, THC, COC, MTD, OPI, PCP	5–7	0.011 ml
Accusign	AMP, BAR, BZO, THC, COC, MTD, OPI, PCP, TCA	3–5	Three drops
EZ-SCREEN	AMP, BAR, THC, COC, OPI	3–5	Six drops
Frontline	AMP, BZO, THC, COC, OPI	3–10	Test strip immersed
InstraCheck	AMP, BAR, BZO, THC, COC, MTD, OPI, PCP	3–8	Four to five drops
Ontrak TesTcup	AMP, BAR, BZO, THC, COC, OPI, PCP	3–5	At least 30 ml
Ontrak TesTstik	AMP, BZO, THC, COC, OPI	3–5	Test strip immersed
Syva RapidTest	AMP, BAR, BZO, THC, COC, MTD, OPI, PCP, TCA	3–10	Three drops
Triage	AMP, BAR, BZO, THC, COC, MTD, OPI, PCP, TCA	11–15	0.14 ml
Visualine II	AMP, BZO, THC, COC, OPI	3–5	Three drops

AMP, amphetamines; BAR, barbiturates; BZO, benzodiazepines; THC, cannabinoids; COC, cocaine metabolite; MTD, methadone; OPI, opiates; PCP, phencyclidine; TCA, tricyclic antidepressants.

organochlorines, and paraquat, to measure these compounds in biological samples. These assays use an antibody bound to magnetic particles and the final color development is measured with a spectrophotometer. Since this method is a heterogeneous immunoassay, matrix effects are minimal.

See also: **Immunoassays: Overview. Immunoassays, Applications: Clinical. Immunoassays, Techniques: Radioimmunoassays; Enzyme Immunoassays. Polymerase Chain Reaction.**

## Further Reading

- Foley T (1995) Enzyme immunoassays. In: Adamovics JA (ed.) *Analysis of Addictive and Misused Drugs*, pp. 1–20. New York: Dekker.
- Gosling JP (ed.) (2000) *Immunoassays: A Practical Approach*. Oxford: Oxford University Press.
- Harlow E and Lane D (1999) *Using Antibodies: A Laboratory Manual*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
- Ho MH (ed.) (1990) *Analytical Methods in Forensic Chemistry*. Chichester: Ellis Horwood.

# IMMUNOASSAYS, TECHNIQUES

## Contents

### Radioimmunoassays

### Enzyme Immunoassays

### Luminescence Immunoassays

## Radioimmunoassays

**A Lacy and R O'Kennedy**, Dublin City University, Dublin, Republic of Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Few innovations have had as much impact on endocrinology, diagnostic, and clinical laboratory medicine as radioimmunoassay (RIA). It was the first technique to use radioisotopic technology to study the primary reaction of antigen with antibody. The

era of RIA can be said to have begun in 1959. Its development is attributed to the collaborative research developed by Rosalyn Yalow and Solomon Berson. By combining techniques from radioisotopic tracing and immunology they were able to investigate the metabolism of insulin labeled with radioactive iodine in diabetic adults. Over a considerable period RIA has proved to be a very sensitive and relatively simple means for measuring minute concentrations of biological and pharmacological substances in blood and other fluid samples. During the 1960s and 1970s, RIA emerged as an important tool in biological research. Knowledge in such diverse areas as endocrinology, cellular biology, and pharmacology was greatly expanded by the application of

**Table 3** Descriptions of noninstrumental immunoassay kits for qualitative on-site drug testing

Test	Analytes	Testing time required (min)	Sample volume required
Abscreen ONTRAK	AMP, BAR, BZO, THC, COC, MTD, OPI, PCP	5–7	0.011 ml
Accusign	AMP, BAR, BZO, THC, COC, MTD, OPI, PCP, TCA	3–5	Three drops
EZ-SCREEN	AMP, BAR, THC, COC, OPI	3–5	Six drops
Frontline	AMP, BZO, THC, COC, OPI	3–10	Test strip immersed
InstraCheck	AMP, BAR, BZO, THC, COC, MTD, OPI, PCP	3–8	Four to five drops
Ontrak TesTcup	AMP, BAR, BZO, THC, COC, OPI, PCP	3–5	At least 30 ml
Ontrak TesTstik	AMP, BZO, THC, COC, OPI	3–5	Test strip immersed
Syva RapidTest	AMP, BAR, BZO, THC, COC, MTD, OPI, PCP, TCA	3–10	Three drops
Triage	AMP, BAR, BZO, THC, COC, MTD, OPI, PCP, TCA	11–15	0.14 ml
Visualine II	AMP, BZO, THC, COC, OPI	3–5	Three drops

AMP, amphetamines; BAR, barbiturates; BZO, benzodiazepines; THC, cannabinoids; COC, cocaine metabolite; MTD, methadone; OPI, opiates; PCP, phencyclidine; TCA, tricyclic antidepressants.

organochlorines, and paraquat, to measure these compounds in biological samples. These assays use an antibody bound to magnetic particles and the final color development is measured with a spectrophotometer. Since this method is a heterogeneous immunoassay, matrix effects are minimal.

See also: **Immunoassays: Overview. Immunoassays, Applications: Clinical. Immunoassays, Techniques: Radioimmunoassays; Enzyme Immunoassays. Polymerase Chain Reaction.**

## Further Reading

- Foley T (1995) Enzyme immunoassays. In: Adamovics JA (ed.) *Analysis of Addictive and Misused Drugs*, pp. 1–20. New York: Dekker.
- Gosling JP (ed.) (2000) *Immunoassays: A Practical Approach*. Oxford: Oxford University Press.
- Harlow E and Lane D (1999) *Using Antibodies: A Laboratory Manual*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
- Ho MH (ed.) (1990) *Analytical Methods in Forensic Chemistry*. Chichester: Ellis Horwood.

# IMMUNOASSAYS, TECHNIQUES

## Contents

### Radioimmunoassays

### Enzyme Immunoassays

### Luminescence Immunoassays

## Radioimmunoassays

**A Lacy and R O'Kennedy**, Dublin City University, Dublin, Republic of Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Few innovations have had as much impact on endocrinology, diagnostic, and clinical laboratory medicine as radioimmunoassay (RIA). It was the first technique to use radioisotopic technology to study the primary reaction of antigen with antibody. The

era of RIA can be said to have begun in 1959. Its development is attributed to the collaborative research developed by Rosalyn Yalow and Solomon Berson. By combining techniques from radioisotopic tracing and immunology they were able to investigate the metabolism of insulin labeled with radioactive iodine in diabetic adults. Over a considerable period RIA has proved to be a very sensitive and relatively simple means for measuring minute concentrations of biological and pharmacological substances in blood and other fluid samples. During the 1960s and 1970s, RIA emerged as an important tool in biological research. Knowledge in such diverse areas as endocrinology, cellular biology, and pharmacology was greatly expanded by the application of

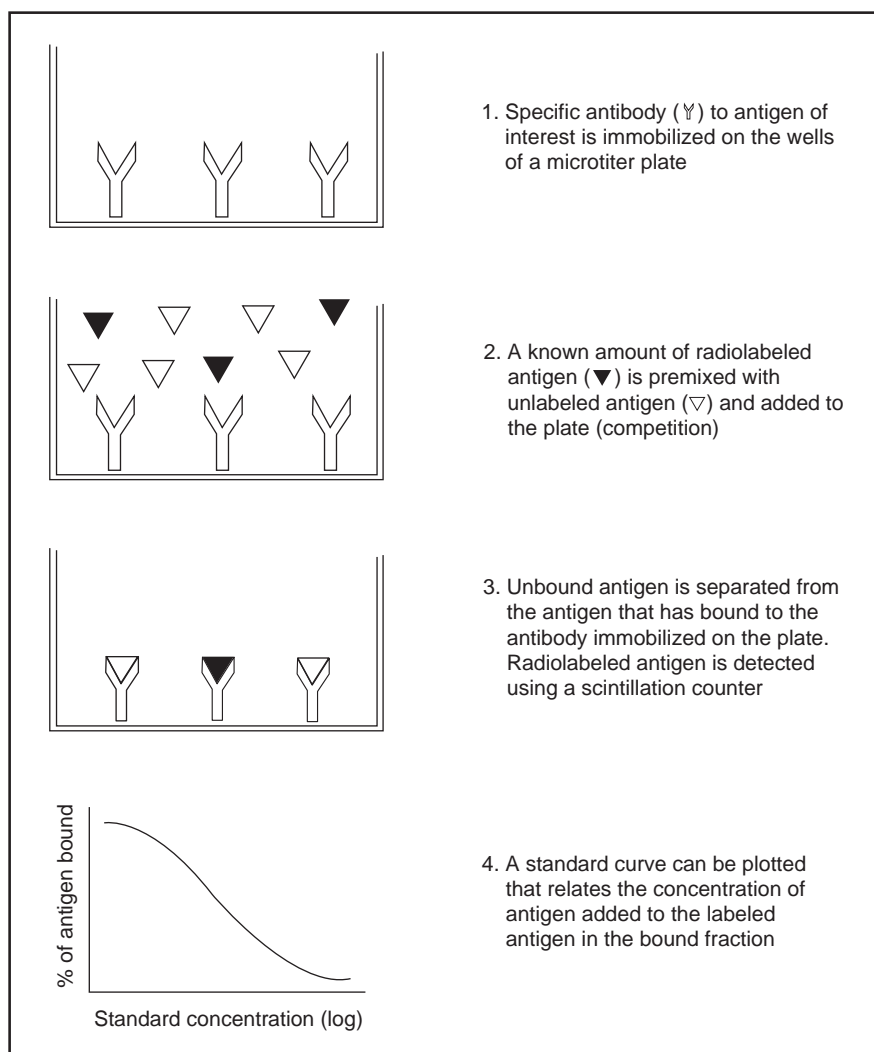
this technique. Over the years the principle of RIA was adapted and nonradioactive labels, such as linked enzymes and fluorescent markers, have been used in place of radioisotopes. In this article, the basic principles of RIA are described, followed by a discussion of its potential and its limitations. Finally, several more recent applications of RIA are presented.

## Principle of Radioimmunoassay

The determination of a substance in biological fluid usually consists of at least two steps, reaction and detection. RIA depends upon the ability of an antibody to bind an antigen such as a hormone (or other molecule of interest) in a reversible fashion. Before discussing the principles of RIA, it is necessary to briefly define some immunological terms. An antigen

is a foreign substance that will induce the formation of immunoglobulins (antibodies) or sensitized cells that react specifically to the antigen. An antibody is an immunoglobulin that is formed in the body in response to a foreign substance (an antigen) and is specific to that substance. Antibodies to compounds too small to be antigenic can be produced by chemically linking the compound (a hapten) to an antigenic carrier such as bovine serum albumin and using this conjugate for immunization.

RIA employs isotopically labeled molecules and permits measurements of extremely small amounts of antigen, antibody, or antigen–antibody complexes. It can be used to measure any substance that will serve as an antigen or hapten. The concentration of labeled molecules is determined by measuring their radioactivity. This highly sensitive analytical method has many applications in hormone assays as well as for



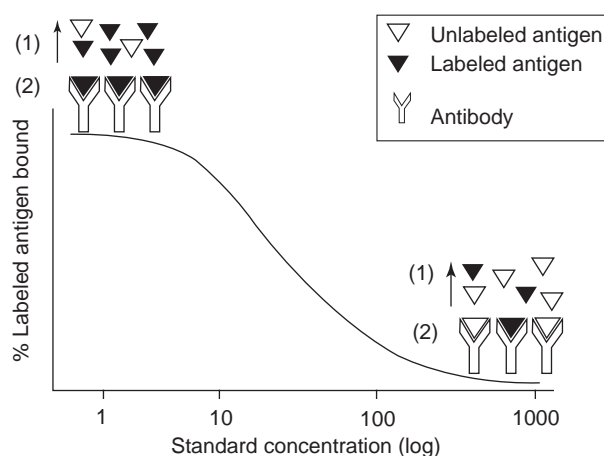
**Figure 1** Simple illustration of the principle of a competitive radioimmunoassay where an antibody is immobilized onto a plate in order to detect the antigen in the test sample.

assays of other substances found at low levels in biological fluids. However, it is not an absolute technique and requires standards. The three basic requirements of RIA are a pure preparation of antigen (as standard), a radiolabeled preparation of antigen, and specific antibody to the antigen in question. RIAs belong to a group of assays known as radio-label assays. These can be divided into RIAs and immunoradiometric assays. RIAs are competitive assays for the detection of antigen where the radiolabel is on the antigen. The slightly less-sensitive immunoradiometric assay involves antibodies that are radiolabeled. In the late 1960s, researchers Miles and Hayes applied the term immunoradiometric to this assay format, although strictly speaking all RIAs are immunoradiometric.

The principle of any RIA involves competition between radiolabeled antigen and nonradiolabeled antigen for a fixed number of antibody binding sites. The antigen (labeled or unlabeled) can bind to its specific antibody to form an antigen-antibody complex. The presence of the radiolabeled antigen-antibody complex can be detected using a scintillation counter. In a standard solid-phase RIA, unlabeled antibodies with binding sites for the antigen of interest are coupled to a solid support such as plastic scintillation tubes or, as is more common nowadays, a 96-well microtiter plate.

Serial dilutions of the unknown sample of antigen (unlabeled) are added as in **Figure 1**. The concentration of the antigen in the unknown sample can be estimated by competitive inhibition with a standard sample of radiolabeled antigen molecules added in a known concentration. The antibody-bound antigen is separated from the unbound antigen. The separation of antibody-bound to labeled antigen from free labeled antigen enables the quantitation of antigen (either standard or unknown). A standard curve can be plotted that relates the concentration of antigen added with the labeled antigen in the bound fraction (where antigen has bound to antibody to form a complex) by setting up a series of tubes or microtiter wells containing progressively more standard antigen. At zero concentration of unlabeled antigen, a large proportion of labeled antigen will bind to antibody. However, the higher the concentration of unlabeled antigen, the less labeled antigen will bind to the fixed amount of antibody because of competition with unlabeled antigen (**Figure 2**).

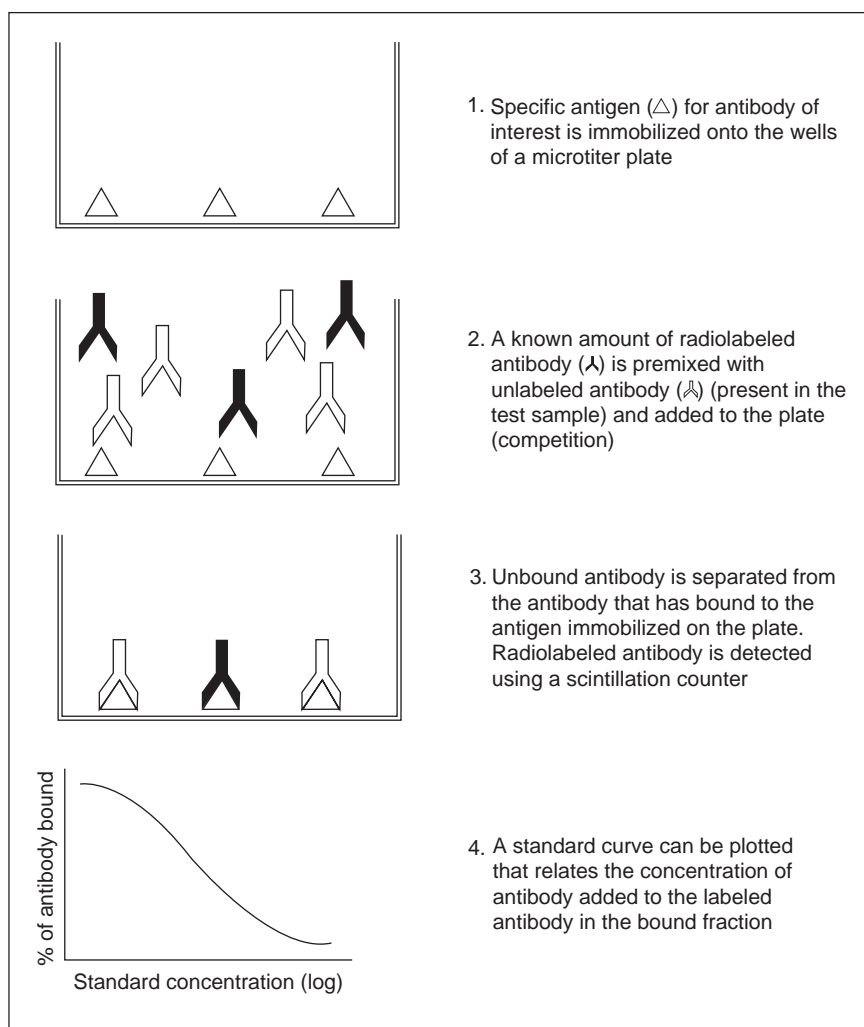
The standard curve has a linear scale of 'percentage of antigen bound' on the y-axis and a log scale of 'standard concentration' on the x-axis. The concentration of antigen in an unknown sample (e.g., hormone in a blood sample) can be calculated by



**Figure 2** Illustration of antigen-antibody interaction and how it is reflected in the resulting standard curve. (1) Both labeled and unlabeled antigens that have not bound to immobilized antibody are separated from the antigen that has bound to antibody. (2) Labeled antigen bound to immobilized antibody is detected using a scintillation counter and the values obtained are used to plot a standard curve.

observing the binding of radiolabeled antigen in that sample and reading it off the standard curve. **Figure 1** illustrates the principle of RIA.

So far the emphasis has been on the use of radiolabeled antigen rather than antibody. Use of radiolabeled antibody can be advantageous when detecting molecules such as IgE in patients suffering from allergies. Iodination of antibody rather than antigen may be favorable if the antigen is susceptible to damage during iodination, during storage, or if it lacks the necessary residues for iodination. In this situation, the antigen is fixed to the solid phase and is then reacted with unlabeled antibodies (e.g., from patients serum), in competition with labeled antibodies. Two modifications of the method for detecting antibodies have a major significance for allergy sufferers. These patients synthesize large amounts of IgE with binding sites for certain common, noninfectious environmental proteins, such as pollens. Patients need to be desensitized to the allergen, which first must be identified before treatment can begin. The first antibody detection method is known as radioallergosorbent test and involves the covalent linking of suspected allergens to cellulose discs, which are reacted with the patient's serum. Allergens that bind IgE from the patient's serum can be identified by indirect radiolabeling, using radiolabeled anti-human-IgE antibodies (**Figure 3**). The second method for detecting antibody is known as the radioimmunosorbent test. This assay measures the total concentration of IgE in patient's serum, regardless of its antigen specificity.



**Figure 3** Simple illustration of the principle of a competitive radiolabeled assay where an antigen is immobilized onto a plate in order to detect the antibody in the test sample.

## Radioimmunoassay Optimization

Several steps are involved in optimizing the development of an RIA and these are now discussed in sequence.

### Production of Radiolabeled Antigen or Antibody

There are four commonly favored isotopes for the labeling of antigen or antibody in RIA. These are  $^{14}\text{C}$ ,  $^3\text{H}$ ,  $^{125}\text{I}$ , and  $^{131}\text{I}$ , which vary greatly in their half-lives and their specific activities. A critical factor in the choice of a radioactive marker is the specific activity of the isotope. The importance of half-life in radioactive emission is evident when the relative specific activities of the different isotopes are calculated. A radioactive marker of low specific activity requires a relatively high concentration of this marker in

**Table 1** Details of specific activity and half-lives of four radioisotopes that are generally favored for the labeling of antigen or antibodies in RIA

Element	Radioisotope	Half-life ( $t_{1/2}$ )	Specific activity ( $\text{Ci mmol}^{-1}$ )
Carbon	$^{14}\text{C}$	5730 years	0.07
Hydrogen	$^3\text{H}$	12.3 years	33
Iodine	$^{125}\text{I}$	60 days	2560
Iodine	$^{131}\text{I}$	8.04 days	19250

order to obtain a practical level of radioactivity for use in the immunoassay. **Table 1** illustrates the half-lives and specific activities of the four most commonly used isotopes. With its combination of both long half-life and high specific activity,  $^{125}\text{I}$  is the most commonly used radioactive label for antigen binding in RIA. Compounds containing  $^{125}\text{I}$  tend to



be more stable than those with  $^{131}\text{I}$  because of the absence of beta radiation. The gamma radiation given off by  $^{125}\text{I}$  is less energetic than that of  $^{131}\text{I}$  and penetrates much less readily through glass and clothing.  $^{125}\text{I}$  is therefore much less of a radiation hazard than  $^{131}\text{I}$  for laboratory personnel.

For antigens that are unusually labile, it may be better to iodinate the antibody rather than the antigen. The antibody that contains numerous tyrosine residues can be substituted with several iodine molecules without adverse effects on antibody activity. The antibody is radiolabeled with  $^{125}\text{I}$ , usually after specific purification with antigen, to reduce or eliminate labeling of nonspecific gamma globulins.

There are a number of procedures for radiolabeling antigen or antibody. Radioiodine can be bound to tyrosine side chains via hydrogen substitution under alkaline conditions, or by conjugation as in the case of iodinated Bolton–Hunter reagent that reacts with lysine amino groups. This prevents the antibody being exposed to harsh chemical conditions, which may reduce the quality of the antigen or antibody.

### Production of Antibody

The performance of an RIA depends not only on the quality of the radioantigen, but also on the properties of the antibody. Traditionally, polyclonal antibodies were produced by immunization of an animal, such as a rabbit. The resultant antiserum contained many slightly different antibodies directed toward different antigenic sites on the antigen surface. Using hybridoma techniques developed in the 1970s it is possible to set up RIAs utilizing monoclonal antibodies from hybridoma cells secreting a single antibody type that bind specifically to one antigenic site on the antigen surface. Monoclonal antibodies are obtained by fusing mouse spleenocytes with myeloma cells to form hybridoma cells that secrete the antibody of interest.

### Optimization of Antiserum Dilution

The optimal working dilution of the antiserum is that which will bind 50–70% of the radioantigen in the absence of unlabeled antigen. A good antiserum with a high antibody titer will often be used in excess of a 1/10 000 dilution for RIA.

### Optimization of Antibody–Antigen Binding

In order to develop a sensitive RIA, the antibody must be capable of binding the antigen with high affinity. Antibody–antigen binding is reversible and utilizes noncovalent chemical bonding such as hydrogen bonding, electrostatic interactions, and van der Waal's forces. These types of chemical bonding

are sensitive to changes in pH and, therefore, it is important to determine the optimal pH that will give maximal antibody–antigen binding. Buffers can affect binding and so several buffer materials should be investigated at the optimal pH. Knowledge of the time taken for antibody–antigen binding to reach equilibrium is of importance when designing an RIA. This can be determined by assaying the proportional binding of a fixed amount of radioantigen to its antibody after various incubation times. Incubation times can vary from 1 to 72 h. Assays requiring greater than 4 h of incubation are carried out at 4°C to minimize possible degradation or denaturation of antibody and/or antigen.

### Separation of Antibody-Bound and Free Antigen

The final step in an RIA involves the measurement of bound or free antigen. Ideally, the separation procedure should be applicable to the simultaneous processing of a large number of immunoassay samples. In addition, separation should be achieved without undue dissociation of preformed antibody–antigen complexes or re-equilibration to form new complexes. Separation of antibody bound to antigen is a crucial step since incomplete separation would lead to a loss of assay precision and sensitivity. The choice of method for separating free and bound antigen depends on a number of factors, including the susceptibility of the antigen–antibody system to dissociation during separation, the solubility and absorption properties of the radioactive antigen, and the requirements of the assay in terms of speed and reproducibility.

If the antigen is a large polypeptide, a particularly useful method of separation is the so-called double-antibody technique. This method of separation utilizes a second antiserum that is raised to the gamma-globulin fraction of the animal species in which the primary antiserum was produced. Centrifugation will then leave the antibody-bound radiolabeled antigen in the supernatant. Either the supernatant or pellet containing radiolabeled antibody-bound antigen can be counted. The double-antibody method has the advantage of broad applicability, sensitivity, and reproducibility. If free antigen is of small molecular mass (i.e., a hapten), it is often possible to absorb it onto solid materials such as dextran-coated charcoal, silica gel, or hydroxyapatite. Salt or solvent precipitation or non-specific absorption methods permit more rapid measurements and, in general, utilize relatively inexpensive reagents. Specific immunoabsorbent systems may also lend themselves to rapid and sensitive measurements. Chromatoelectrophoresis or gel

**Table 2** Various separation methods and their application in RIA

<i>Separation method</i>	<i>Application</i>
Double-antibody method	Large polypeptide antigen Broad applicability, sensitivity, and reproducibility Can cross-react with human gamma globulins
Salt/solvent precipitation (e.g., NaCl/ethanol, chloroform)	Rapid, sensitive, and reproducible Inexpensive reagents Difficult to achieve complete separation
Nonspecific absorption method (e.g., dextran-coated charcoal, hydroxyapatite)	Small molecular mass antigen Rapid, sensitive and reproducible Careful separation needed to prevent erratic results
Specific immuno-absorbent system (e.g., sepharose matrix)	Rapid, simple, and sensitive Difficult to achieve reproducible levels of binding
Chromatoelectrophoresis	Detects incubation damage to radiolabeled antigen Laborious and expensive
Gel filtration	Requires careful monitoring of column/laborious Low sample capacity

filtration separation methods are more laborious and less desirable in general.

Another approach to separating the antibody-bound antigen is to immobilize the antibody. Many commercial RIA kits are now supplied where the antibody is attached to a particulate solid support such as Sephadex and can be pelleted by centrifugation. Another separation method that can be applied in RIA kits is magnetic separation. An example of this is an RIA kit for the detection of progesterone that uses  $^{125}\text{I}$  labeled progesterone and a progesterone-specific antibody bound to magnetic polymer particles. The progesterone bound to the antibody-magnetic polymer complex is separated from free progesterone by using a magnetic field. Table 2 refers to details regarding the various separation methods available.

## Measurement of Radioactivity

Scintillation counting devices exploit the property of certain chemical compounds to emit short light pulses after excitation by the passage of charged particles or by photons of high energy. Radioactivity is the property of an unstable isotope emitting energetic

particles and rays from its nucleus. The amount of radioactive material is measured by how many nuclei decay each second. This is called the activity and is measured in units of curies, abbreviated Ci. There are four basic types of radiation (alpha ( $\alpha$ ), beta ( $\beta$ ), gamma ( $\gamma$ ), and neutron (n)) emitted during decay. These types can be emitted alone or in combination with one another. Radiation ( $\gamma$ - or  $\beta$ -rays) emitting from radiolabels is detected by the combination of a scintillator and a photomultiplier tube, so that the antigen of interest can be quantitatively measured. Most radioactive isotopes such as  $^{75}\text{Se}$  emit  $\gamma$ -rays; however, there are some exceptions (such as  $^3\text{H}$ ,  $^{14}\text{C}$ ,  $^{32}\text{P}$ ,  $^{35}\text{S}$ , and  $^{45}\text{Ca}$ ) that are known as  $\beta$ -emitters. Different scintillation devices are used depending on the type of radiation emitted. Crystal scintillation counters are solid-state radiation detectors for radioactive isotopes that emit  $\gamma$ -rays. This device uses a scintillation crystal (phosphor) to detect radiation and produce light pulses. There are three classes of solid-state scintillation phosphors: organic crystals, inorganic crystals, and plastic phosphors.

The solid crystal scintillators (e.g., sodium iodide crystal) give off a flash of yellow light when they absorb  $\gamma$  radiation. A photomultiplier, which is optically coupled to the sodium iodide crystal, produces a pulse every time it 'sees' a scintillation. The pulse is then amplified and measured in order to determine whether it is of sufficient intensity and of a certain discrete energy level characteristic of the isotope chosen for measurement. The measurements from a crystal scintillation counter are in counts per minute.

Liquid scintillation counters are designed to detect isotopes that emit low levels of  $\beta$  particles, which are electrons that carry a single electrical charge. These devices are extremely sensitive and can detect radioactivity below the detection limits of traditional Geiger counters.  $\text{H}^3$  emits a very low energy  $\beta$  particle that can only travel about half a centimeter in air. Therefore, these low-energy  $\beta$  particles are measured using a liquid scintillation counter. The  $\text{H}^3$  sample can be suspended in a mixture that acts similarly to the sodium iodide crystal.  $\beta$  energy is absorbed into the mixture and produces flashes of blue light. This blue light is detected and measured by photomultiplier tubes in a similar manner to the crystal scintillator, and the pulse height represents the energy of the  $\beta$  particle.

Scintillation counting device advantages include their efficiency, high precision, and accurate counting rates. The latter two attributes are a consequence of the extremely short duration of the light flashes, from  $\sim 10^{-9}$  to  $10^{-6}$  s. The intensity of the light flash and the amplitude of the output voltage pulse are

proportional to the energy of the particle responsible for the flash. Consequently, scintillation counters can be used to determine the energy, as well as the number, of the exciting particles. Recently, the quantity of samples and the number of items to be measured for assays are rapidly increasing. Therefore, equipment for RIAs has been automated, for example, the well scintillation counter that utilizes crystal scintillators. Measurements are made by automatically inserting test tubes, which contain a mixture of antigens and antibodies including labels, into each hole in the scintillator. Each detector section including a scintillator is covered by lead shield to block extraneous radiation.

## Safety Aspects

The equipment and facilities normally available in a pathological laboratory or a hospital can be readily supplemented and used for RIA work. A license must be obtained and adequate training is required in order to permit the handling of radioactive material in any facility. If RIA kits are employed, there is generally no significant radiation exposure to the working personnel. However, precautions should be taken to safeguard against spread of contamination to the scintillation tubes and the scintillation device itself. This can be achieved to a large extent by good work practice. For example, any liquid waste arising from the RIA procedure should be disposed of in accordance with the relevant national and international safety regulations. The personnel engaged in the actual work should have adequate knowledge of the basic procedures of RIA and how to use the scintillation device. They should be aware of the necessary precautions taken when handling radioactive substances and all RIA facilities should provide their staff with access to an RIA safety-training course. The following list illustrates some important facility requirements when first establishing an RIA laboratory:

1. RIA scintillation device for measurements.
2. Benches/areas specifically designed or covered to ensure radioactive contamination cannot be absorbed.
3. Sink with smooth finish and elbow operated taps; foot operated waste bins.
4. Washable paint on all walls, doors, etc., and PVC/linoleum or other suitably covered floors.
5. Separate rooms for source storage and radioactive waste, waste receptacles for collection and storage of liquid and solid radioactive waste.
6. A portable contamination monitor.
7. Items for protection of users, where relevant, e.g., shields.

8. Clear and well designed standard operating procedures.

The documentation of the radiation dose received by personnel working with radioactive material is critical to minimizing such exposures. The most suitable approach to monitoring occupational radiation exposures in individuals is through the use of a personal dosimeter. The appropriate use of the radiation monitoring dosimeters is essential to minimize radiation exposure and they are minimally intrusive to the worker. The dosimeters are usually configured as lapel clips or finger rings. Wristband badges are utilized in circumstances where a ring badge is not suitable.

## Advantages and Limitations

RIAs offer many advantages in terms of specificity, sensitivity, and convenience, but they also have inherent limitations and disadvantages. Some of the limitations are unique or largely restricted to immunoassays. Other limitations represent more general problems in interpretation that arise when quantitative measurements in tissue, blood, or urine are used as an index of what the physiologic or pharmacologic response might be in an organism as a whole.

A major advantage of RIA is that it combines specificity of immunoassays with the sensitivity of radiochemical methods. Concentrations in the picomolar region can be measured accurately. Regardless of the exact RIA procedure used, the great advantage of all RIAs is their sensitivity. Commercially available kits may detect as little as 1 ng or pg of antigen. This is obviously of great significance in monitoring or determining blood levels of certain hormones or therapeutic agents that seldom exceed a few micrograms per milliliter. The purpose of RIA is similar to enzyme-linked immunosorbent assay (ELISA), i.e., to quantitate antigen or antibody. However, when a higher level of sensitivity is needed, plate-bound radioactivity is used for detection rather than color. RIAs used to be carried out in tubes rather than in microtiter plates (as is the case for ELISA) to facilitate radioactive measurements in a scintillation counter. Now, however, with the availability of suitably designed microtiter plates and scintillation counters, solid-phase RIAs using a microtiter format are possible.

RIA also has some notable disadvantages, which include the need for expensive reagents and equipment and the requirements for licensing and containment. Criticism of the use of radiolabels has focused on their relatively short half-life, and

**Table 3** Potential of RIA and its limitations

<i>Advantages</i>	<i>Disadvantages</i>
High sensitivity, good specificity	Expensive instrumentation
Good reproducibility (as far back as Yalow and Berson's work in 1950s)	Special training and license required Disposal of radioactive waste
Availability of RIA kits to a wide range of antigens	Reagents subject to decay (e.g., short half-life of some isotopes)
Well-established formats and labeling procedures	Laboratory safety issues

relatively long counting times required in order to achieve good statistical accuracy, their limited shelf-life, the expense of disposal and concern over general laboratory safety. **Table 3** illustrates both the advantages and disadvantages associated with the use of RIAs.

## Applications

Since the early 1960s, RIAs and other competitive binding assays have been developed to a remarkable array of biochemically or pharmacologically important substances and in many instances have virtually replaced other methods of measurement. RIA is commonly used to measure the low concentrations of peptide and protein hormones that occur in blood. It is widely used both in routine clinical chemistry laboratories and in research laboratories. In this section, important applications developed in the past will be surveyed and recent developments of RIA in areas such as endocrinology and pharmacology shall be reviewed in detail.

### Endocrinology

Due to the relatively long historical association of RIA with endocrinology there has been a vast amount of research carried out since Yalow and Berson performed their work on insulin. Among the early applications (during the 1960s) in this field were immunoassays for growth hormone, corticosteroids, bradykinin, angiotensin II, prolactin, and vasopressin. In subsequent years, RIAs were developed for other major hormones, including estrogens, progesterones, androgens, calcitonin, prostaglandins, and human chorionic gonadotrophin. Some of the important general applications of immunoassays to endocrinologic studies have been as follows:

1. Clinical laboratory measurements for differential diagnosis of endocrine disorders.
2. Determination of hormonal content of different tissues.
3. Studies on the effect of anti-hormone antibodies on hormone activity and metabolism.
4. Recognition and investigation of endocrine-secreting tumors.

'Ready-to-use' RIA kits have been utilized in many studies due to their ease of application and commercial availability for detection of a vast range of hormones. Evaluation of thyroid hormones using RIA kits has helped in the diagnosis of hypothyroidism and hyperthyroidism. Other kits detect human chorionic gonadotrophin for fetal monitoring and angiotensin-1 for differential diagnosis of hypertension. Another commercially available kit is utilized for the quantitative measurement of enzymatically and chemically derivatized noradrenalin and adrenalin in plasma and urine. Some RIA kits have been modified to suit the researcher's needs. Researchers at a behavioral endocrinology laboratory have recently developed a simple, reliable, and highly sensitive assay by modifying commercially available RIA kits for the measurement of estradiol in saliva and blood spot specimens. In 2002, a conventional  $^{125}\text{I}$  RIA was adapted to a microtiter plate format. This is of considerable importance when dealing with a large batch of samples. The  $12 \times 75$  mm scintillation tubes are still commonly used for routine experiments. However, if large batch samples are to be tested this method can be laborious. Commercial microtiter plates have been available for some time but they can be expensive. A novel microtiter format was developed in order to cut the cost and the time (reduced from 4 to 1.5 h per 384 assay tubes) it takes to run these assays. This mini-RIA contains a simple draining device for batch decantation of free antigen from 96-well 1.2-ml mini-tubes. This assay also showed increased sensitivity when compared to conventional RIAs. Since the format of this assay is flexible, most conventional assays can utilize this method for batch samples.

RIAs are used extensively in hormone detection in both serum samples and saliva samples. An example of this involves the development of a specific RIA to detect estrone sulfate in male plasma samples. Estrone sulfate is quantitatively a major circulating plasma estrogen in both female and male humans. It has been widely studied, especially its metabolism, in breast cancer tissues.

Concern has been raised about the adequacy of RIAs for the measurement of steroid sex hormones in population studies. Sex hormones play an important role in regulating growth, maturation, and reproduction and are believed to be of etiologic



importance in several chronic diseases including prostate and breast cancers, osteoporosis, and cardiovascular disease. A recent study was completed by Dorgan and co-workers in order to determine if RIA is an adequate assay system for measuring sex hormones. The investigation involved the comparison of steroid sex hormone measurements in serum by RIA with mass spectrometry. Samples were taken from both male and female pools. Estradiol, estrone, androstenedione, testosterone, and dehydroepiandrosterone sulfate were measured in female pools; testosterone, dihydrotestosterone, androstenedione, and dehydroepiandrosterone sulfate were measured in male pools. The findings of this study indicate that although absolute concentrations may differ for some hormones, RIA and mass spectrometry can yield similar estimates of steroid sex hormone concentrations in serum.

A study by De Niu and co-workers reported an RIA to detect human stanniocalcin in serum and tissue samples. Stanniocalcin is a polypeptide hormone first discovered in fish but only recently identified in humans and other mammals. Before this assay was developed there was no information on circulating levels of stanniocalcin in mammals or the regulation of its secretion. The RIA for detecting stanniocalcin was developed to help explore the possibility that stanniocalcin may have a short half-life in mammals. From the pharmacokinetic data obtained from this research it was determined that mammalian stanniocalcin does not normally circulate in the blood and functions instead as a local mediator of cell function. Another RIA was developed recently to detect vasoactive intestinal polypeptide (VIP). VIP belongs to the family of related peptides containing secretin, glucagons, and growth hormone-releasing hormone. The assay has helped to gather data regarding the comparative distribution of VIP in the central nervous system of various species. A separate study has used both RIA and real-time quantitative polymerase chain reaction to understand the regulation of neurotrophic peptide expression in sympathetic neurons.

RIA has found uses in other areas besides the medical and healthcare sector. Assays have been developed for the agricultural sector. For example, a blood spot direct RIA has become a useful tool for early pregnancy diagnosis in swine with an overall assay accuracy of 96.8%. The assay works by detecting progesterone in blood dried onto filter paper. This method is versatile as it can be used for monitoring progesterone levels during the estrous cycle so that the optimal time for artificial insemination can be estimated. RIAs are also helping to expand knowledge of insect endocrinology. Adipokinetic hormone concentrations in the hemolymph of *Schistocerca gregaria*

(locust) have been measured by an RIA developed in the University of Birmingham.

### Pharmacology

Among the early contributions in this field were immunoassays for digitoxin and morphine. RIAs have subsequently been applied to the measurements of prostaglandins, barbiturates, gentamycin, amphetamines, and vitamins A and D. Important general applications of immunoassays to pharmacology include the following:

1. Detection of drug abuse using rapid and convenient screening procedures.
2. Diagnosis of suspected drug poisoning.
3. Monitoring routine drug therapy with drugs subject to uncertain absorption or metabolism.
4. Studies on drug absorption, distribution, metabolism, and excretion.

Some of the recent RIA developments are quite diverse with regard to pharmacology and include vitamin D detection, illicit drug detection in hair samples, and antibiotic detection in livestock.

Analytical and clinical validation of an RIA for the measurement of 1,25-dihydroxyvitamin D was performed by Clive and co-workers. Vitamin D is a critical regulator of calcium and phosphorus homeostasis. This assay was specific for both 1,25-dihydroxyvitamin D<sub>2</sub> and D<sub>3</sub>. The data collected during this study demonstrated that the RIA validated was a robust, accurate, and precise tool for the assessment of vitamin D.

An area of pharmacology that holds immense interest and importance is illicit drug detection. A commercial RIA kit has been used in an Italian laboratory to detect morphine, cocaine, and ecstasy constituents in extracts of hair samples. This diagnostic strategy has been applied to determine whether former users of illicit drugs are fit enough to obtain a driving license.

A commercially available RIA developed as a screen for tetracycline antibiotics in serum, urine, milk, and tissue of livestock has been adapted by Meyer and co-workers to analyze water samples. The interest in pharmaceutical compounds in the environment is relatively new. Immunoassays for pharmaceutical compounds are commonly used in biological media where the concentrations are quite high. Therefore, in this RIA the lower limit of antibiotic detection had to be modified to enable quantification of antibiotic levels as low as one part per billion in water samples.



Other areas where RIA has found some application include the diagnosis of infectious diseases, oncology, cardiology, and hematology.

It seems that ELISA will replace RIA as it has many advantages over RIA. Despite the continual prediction of doom for the use of isotopic labels for analysis they continue to be widely utilized. This stems in part from well-established methods of preparation, optimization, a history of highly reproducible performance, and the wide availability of scintillation-counting equipment. In certain applications, such as imaging, there are still few if any real alternatives to the use of radiolabels.

**See also:** **Forensic Sciences:** Illicit Drugs. **Immunoassays:** Production of Antibodies. **Polymerase Chain Reaction.**

## Further Reading

- Abdul-Ahad W and Brett M (2000) Labels and Endpoints. In: Gosling J (ed.) *Immunoassays: A Practical Approach*, pp. 99–100. Oxford: Oxford University Press.
- Benjamini E, Coico R, and Sunshine G (2000) Antigen–antibody interactions, immune assays, and experimental systems. In: *Immunology: A Short Course*, pp. 98–100. New York: Wiley-Liss.
- Billington D, Jayson G, and Maltby PJ (1992) Radioimmunoassays and ligand binding assays. In: Graham J and Billington D (eds.) *Radioisotopes: Introduction to Biotechniques*, pp. 67–73. Oxford: Bios Scientific Publishers.
- Chadio S, Xylouri E, Kalogiannis D, *et al.* (2002) Early pregnancy diagnosis in swine by direct radioimmunoassay for progesterone in blood spotted on filter paper. *Animal Reproduction Science* 69: 65–72.
- Clive D, Sudhaker D, Giacherio D, *et al.* (2002) Analytical and clinical validation of a radioimmunoassay for the measurement of 1,25-dihydroxy vitamin D. *Clinical Biochemistry* 35: 517–521.
- Davey B (1992) Applications of antigen recognition. In: *Immunology: A Foundation Text*, pp. 101–102. New York: Wiley.
- De Niu P, Radman D, Jaworski E, *et al.* (2000) Development of a human stanniocalcin radioimmunoassay: serum and tissue hormone levels and pharmacokinetics in the rat. *Molecular and Cellular Endocrinology* 162: 131–144.
- Dorgan J, Fears T, McMahon R, *et al.* (2002) Measurement of steroid sex hormones in serum: a comparison of radioimmunoassay and mass spectrometry. *Steroids* 67: 151–158.
- English L (1994) Antigen–antibody reactions. In: Van Dam-Mieras M and Leach C (eds.) *Technological Applications of Immunochemicals*, pp. 90–94. Oxford: Butterworth-Heinemann.
- Niessner R and Knopp D. Enzymes, immunoassays, ch. 2 ‘Immunoassays in Analytical Chemistry’. In: Günzler H and Williams A (eds.) *Ullmann’s Encyclopedia of Industrial Chemistry*, 6th edn. on CD-ROM. Weinheim: Wiley-VCH.
- Shanley J, Biegel D, and Pachter J (1996) A rapid and sensitive radioimmunoassay for the detection of human cytomegalovirus binding and infection of human fibroblasts. *Journal of Virological Methods* 58: 121–129.

## Enzyme Immunoassays

**Y Seo**, University of Miyazaki, Miyazaki, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Immunoassays are used for the detection of a wide variety of biological and serological substances. They are generally classified into two groups, labeled and unlabeled immunoassays. Labeled immunoassays include antigen or antibody conjugated with radioisotopes, fluorescent materials, metals, enzymes, and other reagents. In enzyme immunoassays (EIAs), an enzyme conjugate can be used to produce a very sensitive colored, fluorescent, or luminescent output.

EIAs have been developed over the last 30 years, and in the last 10 years their use has become

widespread throughout the sciences, including medicine, pharmacology, toxicology, biology, ecology, and food chemistry. This popularity is due to the fact that they can be more sensitive, specific, and simple to use than other types of analytical techniques. The sensitivity of the assay can be manipulated at a variety of steps, including the specifics of the assay procedure, the selection of the enzyme for conjugation, the reaction conditions for the enzyme, the conjugation protocol, and the antibody preparation.

Rapid EIAs have been developed for the diagnosis of viral infections as well as for the identification of valuable markers for carcinomas. Using commercially available kits, these EIAs can be utilized even in laboratories with very limited facilities. In this article, we discuss the advances and improvements of EIAs in the last 10 years and supplement the basic

Other areas where RIA has found some application include the diagnosis of infectious diseases, oncology, cardiology, and hematology.

It seems that ELISA will replace RIA as it has many advantages over RIA. Despite the continual prediction of doom for the use of isotopic labels for analysis they continue to be widely utilized. This stems in part from well-established methods of preparation, optimization, a history of highly reproducible performance, and the wide availability of scintillation-counting equipment. In certain applications, such as imaging, there are still few if any real alternatives to the use of radiolabels.

**See also:** **Forensic Sciences:** Illicit Drugs. **Immunoassays:** Production of Antibodies. **Polymerase Chain Reaction.**

## Further Reading

- Abdul-Ahad W and Brett M (2000) Labels and Endpoints. In: Gosling J (ed.) *Immunoassays: A Practical Approach*, pp. 99–100. Oxford: Oxford University Press.
- Benjamini E, Coico R, and Sunshine G (2000) Antigen–antibody interactions, immune assays, and experimental systems. In: *Immunology: A Short Course*, pp. 98–100. New York: Wiley-Liss.
- Billington D, Jayson G, and Maltby PJ (1992) Radioimmunoassays and ligand binding assays. In: Graham J and Billington D (eds.) *Radioisotopes: Introduction to Biotechniques*, pp. 67–73. Oxford: Bios Scientific Publishers.
- Chadio S, Xylouri E, Kalogiannis D, *et al.* (2002) Early pregnancy diagnosis in swine by direct radioimmunoassay for progesterone in blood spotted on filter paper. *Animal Reproduction Science* 69: 65–72.
- Clive D, Sudhaker D, Giacherio D, *et al.* (2002) Analytical and clinical validation of a radioimmunoassay for the measurement of 1,25-dihydroxy vitamin D. *Clinical Biochemistry* 35: 517–521.
- Davey B (1992) Applications of antigen recognition. In: *Immunology: A Foundation Text*, pp. 101–102. New York: Wiley.
- De Niu P, Radman D, Jaworski E, *et al.* (2000) Development of a human stanniocalcin radioimmunoassay: serum and tissue hormone levels and pharmacokinetics in the rat. *Molecular and Cellular Endocrinology* 162: 131–144.
- Dorgan J, Fears T, McMahon R, *et al.* (2002) Measurement of steroid sex hormones in serum: a comparison of radioimmunoassay and mass spectrometry. *Steroids* 67: 151–158.
- English L (1994) Antigen–antibody reactions. In: Van Dam-Mieras M and Leach C (eds.) *Technological Applications of Immunochemicals*, pp. 90–94. Oxford: Butterworth-Heinemann.
- Niessner R and Knopp D. Enzymes, immunoassays, ch. 2 ‘Immunoassays in Analytical Chemistry’. In: Günzler H and Williams A (eds.) *Ullmann’s Encyclopedia of Industrial Chemistry*, 6th edn. on CD-ROM. Weinheim: Wiley-VCH.
- Shanley J, Biegel D, and Pachter J (1996) A rapid and sensitive radioimmunoassay for the detection of human cytomegalovirus binding and infection of human fibroblasts. *Journal of Virological Methods* 58: 121–129.

## Enzyme Immunoassays

**Y Seo**, University of Miyazaki, Miyazaki, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Immunoassays are used for the detection of a wide variety of biological and serological substances. They are generally classified into two groups, labeled and unlabeled immunoassays. Labeled immunoassays include antigen or antibody conjugated with radioisotopes, fluorescent materials, metals, enzymes, and other reagents. In enzyme immunoassays (EIAs), an enzyme conjugate can be used to produce a very sensitive colored, fluorescent, or luminescent output.

EIAs have been developed over the last 30 years, and in the last 10 years their use has become

widespread throughout the sciences, including medicine, pharmacology, toxicology, biology, ecology, and food chemistry. This popularity is due to the fact that they can be more sensitive, specific, and simple to use than other types of analytical techniques. The sensitivity of the assay can be manipulated at a variety of steps, including the specifics of the assay procedure, the selection of the enzyme for conjugation, the reaction conditions for the enzyme, the conjugation protocol, and the antibody preparation.

Rapid EIAs have been developed for the diagnosis of viral infections as well as for the identification of valuable markers for carcinomas. Using commercially available kits, these EIAs can be utilized even in laboratories with very limited facilities. In this article, we discuss the advances and improvements of EIAs in the last 10 years and supplement the basic

information presented in the section Enzyme Immunoassays in the first edition of the *Encyclopedia of Analytical Science*.

## Sensitivity of EIAs

The sensitivity of EIAs is mainly influenced by the assay procedures. Improvement of these variables has led to the development of extremely sensitive EIAs for antigens, antibodies, and haptens. For example, in some cases, the antigen detection limit is at the one zeptomol ( $1 \times 10^{-21}$  mol) per assay ( $33 \text{ zmol ml}^{-1}$ ) level.

## Basic Procedure

EIAs are classified as either homogeneous or heterogeneous. In homogeneous EIAs, the enzyme activity is altered by the interaction of labeled antigen and antibody. Because the degree of signal changes in proportion to the concentration or amount of antigen, separation of the bound and free (B/F) antigen is not required. Although this technique is rapid and can be applied in some limited cases, it is not generally useful for the development of highly sensitive EIAs. In contrast to the homogeneous assays, heterogeneous EIAs require B/F separation prior to the measurement of enzyme activity. Such a separation of B/F antigen is also used in radioimmunoassays. In general, the heterogeneous method is much more sensitive than the homogeneous method because interference from free ligand is eliminated.

In addition to the heterogeneous and homogeneous assays, EIAs can be classified as competitive or noncompetitive. In competitive EIAs, labeled antigen competes with unlabeled antigen in the test sample for a limited amount of bound antibody. In this case, the signal is due to the amount of bound, labeled antigen-antibody complex, and the decrease in enzyme activity corresponds with the concentration of unlabeled antigen in the test sample. In the noncompetitive method, the signal is due to the direct binding of the labeled antibody complex to the antigen, so that it essentially corresponds directly to the amount of antigen in the test sample.

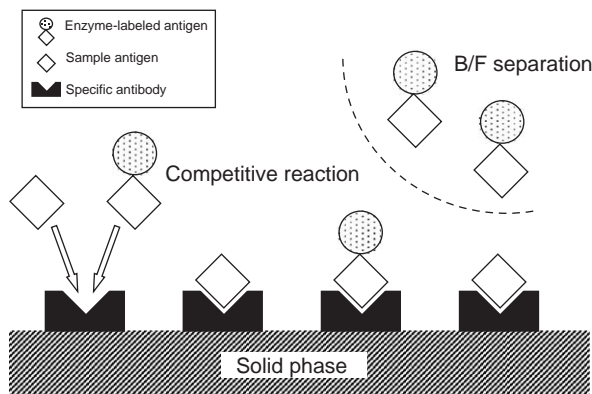
These four methods have been combined to develop a number of EIAs for quantifying antigens, antibodies, and haptens. For example, the enzyme multiplied immunoassay technique (EMIT<sup>®</sup>) is a commercially available homogeneous and competitive EIA for the determination of a variety of drugs. In addition, double antibody methods that utilize a heterogeneous and competitive EIA system are available. In this case, enzyme-labeled antigen and sample antigen are mixed

with an unlabeled primary antibody in solution, and once this first immunoreaction is over, a secondary antibody, typically anti- $\gamma$ -globulin, is added to bind and precipitate the primary antibody-antigen complex. Finally, B/F separation is performed and enzyme activity is measured. However, this assay procedure typically lacks sensitivity because it is a competitive assay wherein small amounts of antigen produce do not significantly decrease the signal.

## Enzyme-Linked Immunosorbent Assay

A significant improvement of these solution-based EIA methods is to utilize antibodies or antigens coupled to a solid support. These are called enzyme-linked immunosorbent assays (ELISAs), and are extremely sensitive. This system requires rapid binding of the antibody or antigen to the solid phase and stability of this complex through multiple steps of washing. The solid phase is typically a latex emulsion, plastic or glass beads, polystyrene balls, or microtiter plate wells. These ELISA techniques are principally heterogeneous EIAs, and there are both competitive and noncompetitive versions.

An example of a competitive ELISA is shown in **Figure 1**. This simple method can be used to measure a wide variety of antigens, antibodies, and haptens. The first step in this assay is the preparation of an antibody-coated solid phase. In many laboratories, this is accomplished by physical adsorption, for example, to the wells of a polystyrene microtiter plate. Next, a known amount of enzyme-labeled antigen and an unlabeled antigen-containing sample is added to the immobilized antibody. Competition of the labeled and unlabeled antigen for the antibody occurs on the solid phase. After this binding step, the unbound antigen is removed by washing. Finally, the bound enzyme-labeled antigen is assayed by addition of a substrate solution (**Figure 1**). In this system,



**Figure 1** A typical competitive ELISA for the determination of antigen. The antigen in the test sample competes with enzyme-labeled antigen for binding to the antibody.

because there is competition for the labeled antigen, the higher the level of antigen in the test sample, the lower the signal. This protocol can also be applied to an antigen capture immunoassay, which is used for the determination of the antibody concentration in a test sample. In this case, the antigen is coupled to the solid phase, and the competition is between a known amount of labeled antibody and the unlabeled antibody in the test sample. In most cases, these competitive ELISAs can achieve femtomole ( $1 \times 10^{-15}$  mol) per assay or higher sensitivities, which is equivalent to the sensitivity of radioimmunoassay. One potential problem with these assays is that the enzyme-conjugated antigen can be too large and obstruct the reaction between the labeled antigen and the antibody-coated solid phase.

### Sandwich ELISA

Two-site enzyme immunoassay or the so-called sandwich enzyme immunoassay (sandwich ELISA) is widely used in many laboratories as a more specific and sensitive assay than other EIAs. This method is a heterogeneous and noncompetitive type of EIA. The principle of this system is shown in Figure 2. This system is more versatile than other EIAs, especially for detection of protein antigens, and it can achieve attomole ( $1 \times 10^{-18}$  mol) per assay sensitivity. Sandwich ELISAs utilize both an antibody-coated solid phase and an enzyme-labeled antibody. In the first step, the antigen is bound to the antibody-coated solid phase, and the unbound antigen is removed by washing. Next, the enzyme-labeled antibody is added, and, after the incubation and immunoreaction are complete, the unbound enzyme-labeled antibody is removed by washing. Finally, the amount of bound enzyme-conjugated antibody is determined by assay

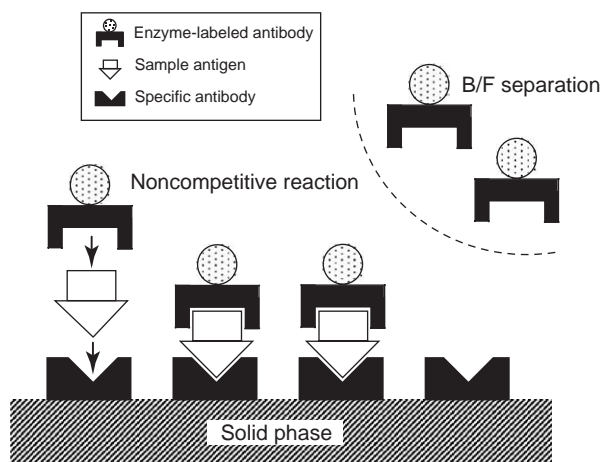
of the enzyme activity. In this case, the enzyme activity corresponds directly to the amount of antigen. Due to the washing, there is little nonspecifically bound labeled antibody. Therefore, background signals are very low, and the system is highly sensitive.

There are variety of sandwich ELISA systems. For example, in indirect sandwich ELISAs, the assay uses an unlabeled primary antibody for the sandwich step in conjunction with an enzyme-labeled anti-immunoglobulin secondary antibody. This system can also employ a biotinylated sandwich antibody and enzyme-labeled streptavidin for detection. The indirect sandwich ELISA can be used not only for detection of antigens but also for measurement of antibody concentration, a system more specifically referred to as an indirect antibody capture immunoassay. In this case, an antigen-coated solid phase is prepared and then incubated with a serum or plasma test sample. Next, a secondary enzyme-labeled anti-immunoglobulin antibody or enzyme-labeled antigen is added, unbound reagents are removed by washing, and enzyme activity is assayed.

Although these ELISA systems are usually carried out in microtiter plates and with colorimetric enzyme reactions, the assays may be even more sensitive using small polystyrene balls (of 3 mm diameter) as the solid phase along with a fluorometric enzyme assay. In fact, such fluorometric assay systems can achieve 0.3 attomol per assay ( $10 \text{ mol ml}^{-1}$ ) sensitivity. There are, however, some limitations to the sandwich ELISA. For example, the sandwich system requires two immunoreactions, first to bind antigen to immobilized antibody, and, second, to bind the sandwich antibody. Therefore, the antigen must have more than one antigenic epitope. As a result, it may not be possible to assay small haptens with this technique. Also, a single monoclonal antibody cannot be used both for coupling to the solid phase and as the sandwich antibody. However, a polyclonal antibody could be used both as the coupling and sandwich antibodies because it contains a heterogeneous mixture of antibodies against multiple antigen epitopes.

### Immune Complex Transfer Immunoassay

Currently, one of the most sensitive EIA methods is the immune complex transfer immunoassay, which is a heterogeneous and noncompetitive EIA. This system can measure zeptomole per assay levels of protein antigens and antibodies as well as attomole per assay levels of several haptens. In this method, the primary immune reaction occurs on a solid support, after which the immune complexes are specifically



**Figure 2** A 'sandwich' type enzyme immunoassay for the detection of an antigen.



eluted. Next, the eluted immune complexes are trapped on a second solid phase, enzyme-labeled reagents are added, and, finally, enzyme activity is assayed. This technique is extremely sensitive because it greatly reduces nonspecific binding of labeled reagents.

The immune complex transfer assay has allowed the development of very sensitive EIAs for the detection of antibodies. In one example, the assay utilizes a 2,4-dinitrophenyl (DNP)–biotin-conjugated antigen. This is incubated with the sample antibody to be measured, and, after the immune reaction, the immune complexes are trapped onto a primary solid phase coupled with anti-DNP antibody. After washing away unbound materials, immune complexes are eluted with DNP–lysine and transferred to a secondary streptavidin-coated solid phase. Next, an enzyme-labeled anti-immunoglobulin antibody is added, and, after washing, enzyme activity is assayed. In this case, the enzyme activity correlates with the amount of antibody in the sample. The sensitivity depends on the amount of nonspecific binding of the enzyme-labeled antibody on the second solid phase. Thus, several modified assay systems have been developed to reduce background.

## Enzyme and Enzyme Labeling

Generation of enzyme-conjugated antigens or antibodies is the most important aspect of EIAs. The enzyme selected for labeling must correspond to the

needs of the individual EIAs. In addition, the methods used for enzyme conjugation must be selected to maintain antigen-binding activity and to prevent nonspecific binding.

## Enzyme

Selection of the enzyme for conjugation is a principal factor in the development of EIAs. The enzyme selected must be of high purity, there must be a good assay for the enzyme, and there must be instrumentation for the detection of the products of the enzyme reaction. Also, the enzyme must be stable during storage, enzyme conjugation must be possible under neutral conditions, and conjugation must not damage the activity of the enzyme. There are only a few enzymes that satisfy these requirements. Therefore, emphasis has been placed on the development of highly sensitive assays and detection methods for these few enzymes.

Detection methods for the enzyme reactions include colorimetric, fluorometric, chemiluminometric, and bioluminometric assays, as well as electrochemical methods, and their detection ranges vary. The principal enzymes used for labeling, assay methods available, and levels of sensitivity are summarized in Table 1. Amongst these, horseradish peroxidase,  $\beta$ -D-galactosidase from *Escherichia coli* (*E. coli*), and alkaline phosphatase from calf intestine or *E. coli* are currently the most widely used for conjugating antigens or antibodies.

**Table 1** Assays and approximate detection ranges of enzymes used in EIAs

Enzyme	Assay	Detection range (mol/assay)			
		$10^{-12}$	$10^{-15}$	$10^{-18}$	$10^{-21}$
Peroxidase (E.C. 1.11.1.7)	Colorimetry				
	Fluorometry				
	Chemiluminometry				
$\beta$ -D-galactosidase (E.C.3.2.1.2.3)	Colorimetry				
	Fluorometry				
	Bioluminometry				
Alkaline phosphatase (E.C. 3.1.3.1)	Colorimetry				
	C. enzyme cycling <sup>a</sup>				
	Fluorometry				
	Chemiluminometry				
Glucose-6-phosphate dehydrogenase (E.C. 1.1.1.49)	Photometry <sup>b</sup>				
	Bioluminometry				
Glucose oxidase (E.C. 1.1.3.4)	Colorimetry				
	Chemiluminometry				
Malate dehydrogenase (E.C. 1.1.1.82)	Photometry <sup>b</sup>				

<sup>a</sup>Colorimetric assay of enzyme cycling using NADP as a substrate.

<sup>b</sup>Measurement of absorbance from NADH production.



## Enzyme Labeling

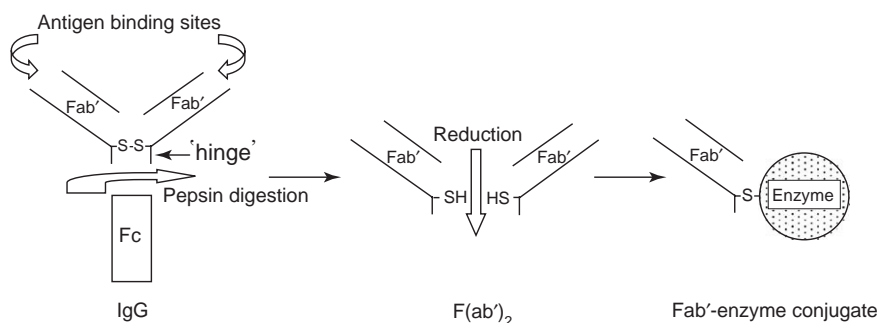
A variety of methods and reagents for coupling an enzyme to an antigen or antibody have been developed. Of these, glutaraldehyde is the most commonly used because it easily reacts with amino groups under mild conditions, and stable enzyme conjugation is accomplished within a few hours. While this one-step method is simple and applicable to almost all proteins, the labeling is nonselective, and heteropolymers are frequently formed, which can interfere with antigen-antibody binding. Two-step coupling methods generally avoid these problems, although they also may use glutaraldehyde as the coupling agent. In this process, one of the components is labeled with the coupling reagent, after which the excess reagent is removed by dialysis or gel filtration. Finally, the conjugated component is reacted with the second component to form the final product. Peroxidase- or alkaline phosphatase-labeled antibody is often prepared using this method, but the production of heteropolymeric products is not avoided, and the recovery of conjugated antibody is very low.

In addition to glutaraldehyde, periodate can be used to prepare antibody or antigen conjugates. Periodate reacts with glycoprotein carbohydrates to generate reactive aldehyde groups. These aldehyde groups can react with amino groups to form a stable covalent bond. This method is very effective for labeling antibodies with peroxidase because it contains ~18% by weight of carbohydrates, and the reaction does not interfere with enzyme activity.

A number of maleimide derivatives are used for coupling reagents to introduce maleimide groups into enzymes and antibodies. The maleimide groups in these reagents react efficiently with thiols to form stable covalent bonds. Free thiol groups can be produced in the hinge portion of antibodies by mild reduction, and some enzymes, such as  $\beta$ -D-galactosidase from *E. coli*, naturally contain free thiols. Pyridyldisulfide groups can also be used to

conjugate thiols, but the efficiency of coupling is lower than for maleimides. Although a variety of coupling reagents are synthesized and available, they require satisfying some important factors, such as solubility, stability, reactivity, and purity. Maleimide, thiol, and pyridyldisulfide derivatives are frequently less soluble in the buffer, but almost all of these may be soluble in *N,N*-dimethylformamide. Maleimide groups of some maleimide derivatives are stable at acidic pH value; however, their reactivities are obstructed in the presence of  $\text{NaN}_3$  as an antiseptic. The reactions between maleimide and thiol groups are usually completed within 5–10 h at 4°C or 1 h at 30°C subjected to the weakly acidic condition. The purity of coupling reagents is also essential to obtain sufficiently good results.

There are two categories of methods for conjugating antibodies with enzymes: hinge and non-hinge methods. Hinge methods label the antibody in the hinge region, while nonhinge methods label the antibody at other sites. The nonhinge method includes the use of glutaraldehyde or periodate because these reagents can label throughout the antibody. IgG molecules are liable to fragmentation by proteolytic attack by enzymes. Fab and  $\text{F(ab')}_2$  fragments containing the entire L chain and disulfide linked portion of the H chain can be produced by digestion with papain and pepsin, respectively. The  $\text{F(ab')}_2$  fragment is divalent and the Fab fragment is univalent, and Fab fragment binds to the antigen through non-cross linkage and precipitation. Hinge methods frequently use mild reduction of  $\text{F(ab')}_2$  to produce free thiols in the hinge region of  $\text{Fab'}$ , followed by the use of maleimide-based coupling reagents. **Figure 3** depicts enzyme conjugation using a maleimide-hinge method. Enzyme-labeled antibodies produced using the maleimide-hinge method, such as  $\text{Fab'}$ -peroxidase conjugates, have many advantages, including the absence of heteropolymers and full maintenance of enzyme and antigen-binding activity. In addition, in the case of  $\text{Fab'}$ -peroxidase conjugates, nonspecific binding is low because they lack the hydrophobic Fc



**Figure 3** Conjugation of  $\text{Fab'}$  with enzyme using the maleimide-hinge method.

portion of IgG, which frequently causes nonspecific binding.

## Antibody

The preparation of antibody is another important factor for the development of EIAs. High affinity and specificity antibodies are needed for the development of sensitive and specific EIAs. Fab'-enzyme conjugates prepared using maleimide-hinge methods may become indispensable in this regard. Furthermore, affinity purification of antibody and/or enzyme-labeled antibody as well as utilization of monoclonal antibodies may also be needed for the development of specific and sensitive EIAs. Although several bimolecular approaches that are producing recombinant antibody possessing enzyme activity have been attempted, sufficiently good results have not yet been acquired and practical usages have not been accomplished.

### Preparation of Fab'

Fab' is often used along with the maleimide-hinge method of enzyme conjugation. F(ab')<sub>2</sub> can be prepared from rabbit, goat, sheep, guinea pig, rat, mouse, and capybara IgG by digestion with pepsin. The IgGs in this case can be purified by precipitation with saturated sodium sulfate, followed by DEAE ion-exchange chromatography. Antibody can also be purified using proteins A or G. Fab' is generated from F(ab')<sub>2</sub> by reduction under mild conditions, often with 2-mercaptoethylamine because it does not damage antigen-antibody binding. This also generates free thiol groups in the hinge region of Fab', which can be conjugated with maleimide (Figure 3). Because  $\beta$ -D-galactosidase from *E. coli* naturally has many thiol groups, maleimides can easily be conjugated to the reduced Fab'.

### Affinity Purification of Antibodies

ELISA systems are typically 10- to 100-fold more sensitive when they utilize affinity-purified antibodies instead of unpurified antibodies. Coupling of affinity-purified antibodies to the solid phase can result in an increase in background because minute amounts of antigen-antibody complexes can be eluted with the affinity-purified antibody and because antibodies can be damaged during elution procedures. To avoid these problems, the enzyme-conjugated antibody should be affinity purified, or an affinity-purified antibody should be used for enzyme labeling.

Antibodies are affinity purified by binding them to immobilized antigen. The most common solid support for immobilizing antigen is cyanogen bromide-activated Sepharose. Almost all antigens containing amino

groups can be conjugated in this way to this resin. Ligands containing carboxyl, sulfhydryl, or aldehyde groups can also be attached to support materials.

### Monoclonal Antibodies

A large number of monoclonal antibodies with high specificity and affinity are commercially available for a variety of antigens. Although many EIAs using monoclonal antibodies have been developed, the antigen-binding activities are frequently reduced due to covalent attachment and/or the solid phase. For sandwich ELISAs, two distinct monoclonal antibodies that react with different sites on the same antigen are necessary. Alternatively, a monoclonal antibody can be used in conjunction with a polyclonal antibody to the same antigen.

### Preparation of Immobilized Antibodies

Antibodies are commonly immobilized to solid supports by either physical adsorption or covalent linkage. Although glutaraldehyde can be used to couple antibody or antigen to a solid phase containing amino groups, physical adsorption is sufficient for most cases. Polystyrene appears to be the most stable solid support for this purpose.

### Interaction between Antigens and Antibodies

The antigen-antibody interaction is performed in neutral pH buffers containing reagents for stabilizing enzyme activity as serum albumin, and the incubation time is normally in the range of 1-48 h at 4-37°C. In ELISAs, however, higher temperature and long periods of incubations cause reduction of enzyme activity and elevation of nonspecific binding of the labeled antibody on the solid phase. The interaction between the enzyme-labeled antibody and the antigen should be carefully controlled at a temperature at 20°C or 30°C for 2-4 h.

## Assays

Colorimetric detection is the most commonly used method for the detection of enzyme activity. Fluorometric or chemiluminometric assays are also utilized when high sensitivity is desired because they are often more than 10-fold more sensitive. In fact, fluorometric or chemiluminometric assays can be several thousand times more sensitive when using  $\beta$ -D-galactosidase or alkaline phosphatase as the conjugating enzyme (Table 1).

### Colorimetric Assays

In colorimetric assays, the substrate is converted by the enzyme into a soluble, colored reaction product.

This allows precise determination of the enzyme activity by optical density. For example, peroxidase activity is measured by the reduction of the substrate in the presence of hydrogen peroxide, producing a soluble chromogen that is measured with a spectrophotometer. There are a variety of chromogenic peroxidase substrates that can be used, including the commonly used 1,2-phenylene diamine, 2,2'-azino-di(3-ethyl)benzothiazoline-6-sulphonic acid, 3,3',5,5'-tetramethylbenzidine, and 5-aminosalicylic acid. These give rise to products with a variety of colors, and are therefore measured at various wavelengths. When  $\beta$ -D-galactosidase is used as the conjugating enzyme, 2-nitrophenyl- $\beta$ -D-galactoside is generally used as the substrate, and the colored product is measured at 420 nm. For the colorimetric assays of alkaline phosphatase, 4-nitrophenylphosphate is normally used, and the colored product is measured at 405 nm. In addition, enzymatic recycling of NADP makes the colorimetric assay for alkaline phosphatase as sensitive as fluorometric assays. These colorimetric assay methods are often used for microtiter plate-based ELISAs in combination with an automated spectrophotometer.

### Fluorometric Assays

In general, fluorometric assays are as much as 100- to 1000-fold more sensitive than colorimetric assays (Table 1). This method has, therefore, become essential for the recent development of highly sensitive EIA systems. Spectrofluorometers can detect a minute amount of product when the substrates are not themselves fluorescent or have very low intrinsic fluorescence.

There are number of such artificial substrates that are commercially available for the various enzymes used in EIAs. For example, the fluorometric assay for peroxidase utilizes homovanillic acid, tyramine, 4-hydroxyphenylacetic acid, or 3-(4-hydroxyphenyl)propionic acid (HPPA) as a substrate. While the fluorescence of the product from homovanillic acid is relatively unstable, the products from tyramine, 4-hydroxyphenylacetic acid, and HPPA are fairly stable and are more sensitive. Amongst these three substrates, HPPA is the most effective substrate when a highly sensitive ELISA is sought because it can be used to detect attomoles per assay or even lower levels of antigen or antibody (Table 1).

The activity of  $\beta$ -D-galactosidase can also be assayed fluorometrically, and this generally utilizes 4-methylumbelliferyl- $\beta$ -D-galactoside as the substrate. Enzyme activity produces the fluorescent product 4-methylumbelliferone and  $\beta$ -D-galactoside. The assay system is very stable and there is little interference from blood. In addition, this assay system is

even more sensitive than the fluorometric assays for peroxidase.

The fluorescent assay for alkaline phosphatase takes advantage of a related reagent, 4-methylumbelliferylphosphate. As in the case of  $\beta$ -D-galactosidase, enzyme activity produces the fluorescent product 4-methylumbelliferone. However, 4-methylumbelliferylphosphate has a relatively high intrinsic fluorescence. As a result, the fluorometric alkaline phosphatase assay is not as sensitive as the  $\beta$ -D-galactosidase assay.

### Chemiluminescent Assays

In addition to colorimetric and fluorometric enzyme assays, chemiluminescence can also be used to measure enzyme activities in EIAs. Closely related assays are used for chemiluminescent immunoblotting. For example, peroxidase activity can be measured in the presence of hydrogen peroxide using luminol as a substrate. In this case, the oxidized luminol is in an excited state and releases light, which can then be measured with a luminometer. Alkaline phosphatase can also be measured using the chemiluminescent substrate 3-(2'-spiroadamantane)-4-methoxy-4-(3''-phosphoryloxy)phenyl-1,2-dioxetan. Chemiluminescence permits extremely high sensitivity because the substrate alone produces no light (Table 1). Additional improved chemiluminescent substrates that maintain high stability and light output have been developed for various EIA systems.

### Bioluminescent Assays

Bioluminescent assays are the most sensitive methods for determining enzyme activity (Table 1). There are two types, those that use the firefly luciferin-luciferase system, and those that use a bacterial luciferase system. Of these methods, firefly luciferase has been more extensively studied, and has been utilized in a wide range of scientific fields. This highly sensitive method uses ATP as a substrate to produce light. The bacterial luciferase, usually obtained from *Photobacterium fischeri*, is in a complex with FMN reductase as NAD(P)H:FMN oxidoreductase. In this case, FMN or NADH is used by the enzyme to generate light. A highly sensitive bioluminescence assay has been developed using this system for glucose-6-phosphate dehydrogenase. In this system, glucose-6-phosphate dehydrogenase produces NADH, which then reacts with the bacterial luciferase to produce light. The bioluminescent assay for NADH can be amplified by using a NAD<sup>+</sup> recycling system. Finally,  $\beta$ -D-galactosidase can be measured by a bioluminescent method using 2-nitrophenyl- $\beta$ -D-galactoside as a substrate. In this case, the galactose produced is used by

**Table 2** Examples of commercially available EIAs

Diagnosis	Subject	System	Assay	Operations	
				Manual	Automatic
Drugs	Phenobarbital lidocaine	Homogeneous-competitive	Photometry <sup>a</sup>	—	○
		Homogeneous-competitive	Photometry <sup>a</sup>	—	○
Markers of carcinoma	CEA	Sandwich ELISA	Colorimetry	○	○
			Fluorometry	—	○
			Chemiluminometry	—	○
	CA 19-9	Sandwich ELISA	Colorimetry	○	○
			Chemiluminometry	—	○
	PSA	Sandwich ELISA	Colorimetry	○	○
Chemiluminometry			—	○	
Infectious diseases	HBc antibody	Sandwich ELISA	Colorimetry	—	○
			Chemiluminometry	—	○
		Competitive ELISA	Colorimetry	—	○
			Fluorometry	—	○
	Chemiluminometry		—	○	
	Chlamydia antigen	Sandwich ELISA	Colorimetry	○	○
			C. Enzyme cycling <sup>b</sup>	○	—
	Chlamydia antibody	Competitive ELISA	Colorimetry	○	—
Hormones	T <sub>4</sub>	Homogeneous-competitive	Colorimetry	—	○
		Homogeneous-competitive	Colorimetry	—	○
		Competitive ELISA	Colorimetry	—	○
			Chemiluminometry	—	○
	Insulin	Sandwich ELISA	Colorimetry	○	○
			Fluorometry	—	○
			Chemiluminometry	—	○
Environment	17β-Estradiol	Competitive ELISA	Colorimetry	○	—
Food chemistry	Staphylococcal enterotoxin	Sandwich ELISA	Colorimetry	○	—

<sup>a</sup>Monitoring of absorbance from NADH.<sup>b</sup>Colorimetric assay of alkaline phosphatase with NADP as a substrate and using enzyme cycling.

galactose dehydrogenase to produce NADH, which is then used to produce light by bacterial luciferase.

## Commercially Available EIAs

There are a variety of EIA-based kits for laboratory assays, clinical diagnoses, and environmental and food chemistry. These kits use a combination of homogeneous or heterogeneous methods with competitive or noncompetitive assays (Table 2). For example, determination of several drugs is measured by either homogeneous or competitive EIAs, and several serological markers for diagnosing carcinoma are assessed using ELISA techniques, such as sandwich ELISAs. Determination of antibodies by EIA is also used for the diagnosis of several infectious diseases. Increasingly, these commercially available EIAs are

automated. Because of the sensitivity and selectivity of more recent EIA systems, the use of these techniques will continue to grow throughout the sciences.

See also: **Bioluminescence**. **Chemiluminescence**: Liquid-Phase. **Electrophoresis**: Blotting Techniques. **Enzymes**: Immobilized Enzymes; Enzyme Assays. **Fluorescence**: Quantitative Analysis. **Forensic Sciences**: Blood Analysis. **Immunoassays**: Overview. **Immunoassays, Applications**: Clinical; Forensic. **Immunoassays, Techniques**: Radioimmunoassays.

## Further Reading

- Diamandis EP and Christopoulos TK (eds.) (1996) *Immunoassay*. San Diego: Academic Press.  
 Edwards R (ed.) (1996) *Immunoassays: Essential Data*. New York: Wiley.

- Gosling JP (ed.) (2000) *Immunoassays: A Practical Approach* (The Practical Approach Series, 228). North Carolina: Oxford University Press.
- Goers J (ed.) (1993) Enzyme-linked immunosorbent assay: ELISA. In: *Immunochemical Techniques Laboratory Manual*, pp. 119–135. San Diego: Academic Press.
- Hashida S, Hashinaka K, Ishikawa S, and Ishikawa E (1997) More reliable diagnosis of infection with human immunodeficiency virus type 1 (HIV-1) by detection of antibody IgGs to *pol* and *gag* proteins of HIV-1 and p24 antigen of HIV-1 in urine, saliva, and/or serum with highly sensitive and specific enzyme immunoassay (immune complex transfer enzyme immunoassay): A review. *Journal of Clinical and Laboratory Analysis* 11: 267–286.
- Ishikawa E, Hashida S, Kohno T, and Tanaka K (1988) Methods for enzyme labeling of antigens, antibodies and their fragments. In: Ngo TT (ed.) *Nonisotopic Immunoassay*, pp. 27–55. New York: Plenum.
- Ishikawa E (1999) *Ultrasensitive and Rapid Enzyme Immunoassay*. Vol. 27 – *Laboratory Techniques in Biochemistry and Molecular Biology*. Amsterdam: Elsevier.
- Miyai K (1981) Enzyme immunoassay system – schematic representation. In: Ishikawa E, Kawai T, and Miyai K (eds.) *Enzyme Immunoassay*, pp. 123–135. Tokyo: Igaku-Shoin.
- Ternyck TH and Avrameas S (eds.) (1990) Quantitative enzyme immunoassays. In: *Immunoenzymatic Techniques*. Vol. 1 – *Techniques in Immunology*, pp. 43–68. Amsterdam: Elsevier.
- Tsuji A (1981) Enzymatic assay by chemiluminescence and fluorescence reactions. In: Ishikawa E, Kawai T, and Miyai K (eds.) *Enzyme Immunoassay*, pp. 41–53. Tokyo: Igaku-Shoin.
- Wu J (2000) *Quantitative Immunoassay: A Practical Guide for Assay Establishment, Troubleshooting, and Clinical Application*. Washington DC: AACC Press.

## Luminescence Immunoassays

**G Gübitz and M G Schmid**, Karl-Franzens University, Graz, Austria

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Immunoassays represent very selective and sensitive techniques that have found application in several areas such as clinical chemistry, bioanalysis, pharmaceutical analysis, toxicological analysis, and environmental analysis. The first immunoassays developed were radioimmunoassays, which are very sensitive; however, one drawback is the need to work radiochemicals. To circumvent the drawbacks, other labels such as enzymes in combination with photometric measurement have been introduced, however, at the expense of sensitivity. With the development of fluoroimmunoassays (FIAs) an improvement in sensitivity was obtained and by introducing chemiluminescence in immunoassays, a sensitivity equivalent to radioimmunoassays was achieved. In this article, different variations and techniques of luminescence immunoassays are described.

### Theory of Luminescence

Luminescence can be created by photoirradiation, which results in fluorescence, delayed fluorescence, and phosphorescence or by a chemical or biochemical reaction, which produces chemiluminescence and

bioluminescence. Molecules can absorb energy, whereby electrons are excited from the ground state to a higher state. By returning of the electrons to the ground state, the absorbed energy can be turned to radiation as fluorescence. If the transition to the ground state occurs through a metastable triplet state, phosphorescence is observed. Some species, such as lanthanides, show delayed fluorescence, which is a result of two intersystem crossings, the first from the singlet to the triplet and the second from the triplet to the singlet state.

Chemiluminescence can be a result of a chemical or biochemical reaction, an electrochemical process, or a thermal treatment.

## Fluorescence Immunoassays

### Fluorescent Probes

Fluorescent probes used as labels should have high quantum yield, emission wavelengths longer than 500 nm, and Stokes' shifts of over 50 nm. They must possess suitable functional groups such as carboxylic acid, amino, hydroxyl, sulfonic acid residues, acid anhydride, acid chloride, diazonium, isocyanate, isothiocyanate, *N*-hydroxysuccinimide, maleimide.

Commonly used probes are fluorescein derivatives, rhodamine derivatives, polycyclic aromatic hydrocarbons, coumarines, amine reagents such as fluorescamine or NBD-Cl, phycobiliproteins, porphyrins, and metal chelates. The latter may show fluorescence,



- Gosling JP (ed.) (2000) *Immunoassays: A Practical Approach* (The Practical Approach Series, 228). North Carolina: Oxford University Press.
- Goers J (ed.) (1993) Enzyme-linked immunosorbent assay: ELISA. In: *Immunochemical Techniques Laboratory Manual*, pp. 119–135. San Diego: Academic Press.
- Hashida S, Hashinaka K, Ishikawa S, and Ishikawa E (1997) More reliable diagnosis of infection with human immunodeficiency virus type 1 (HIV-1) by detection of antibody IgGs to *pol* and *gag* proteins of HIV-1 and p24 antigen of HIV-1 in urine, saliva, and/or serum with highly sensitive and specific enzyme immunoassay (immune complex transfer enzyme immunoassay): A review. *Journal of Clinical and Laboratory Analysis* 11: 267–286.
- Ishikawa E, Hashida S, Kohno T, and Tanaka K (1988) Methods for enzyme labeling of antigens, antibodies and their fragments. In: Ngo TT (ed.) *Nonisotopic Immunoassay*, pp. 27–55. New York: Plenum.
- Ishikawa E (1999) *Ultrasensitive and Rapid Enzyme Immunoassay*. Vol. 27 – *Laboratory Techniques in Biochemistry and Molecular Biology*. Amsterdam: Elsevier.
- Miyai K (1981) Enzyme immunoassay system – schematic representation. In: Ishikawa E, Kawai T, and Miyai K (eds.) *Enzyme Immunoassay*, pp. 123–135. Tokyo: Igaku-Shoin.
- Ternyck TH and Avrameas S (eds.) (1990) Quantitative enzyme immunoassays. In: *Immunoenzymatic Techniques*. Vol. 1 – *Techniques in Immunology*, pp. 43–68. Amsterdam: Elsevier.
- Tsuji A (1981) Enzymatic assay by chemiluminescence and fluorescence reactions. In: Ishikawa E, Kawai T, and Miyai K (eds.) *Enzyme Immunoassay*, pp. 41–53. Tokyo: Igaku-Shoin.
- Wu J (2000) *Quantitative Immunoassay: A Practical Guide for Assay Establishment, Troubleshooting, and Clinical Application*. Washington DC: AACC Press.

## Luminescence Immunoassays

**G Gübitz and M G Schmid**, Karl-Franzens University, Graz, Austria

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Immunoassays represent very selective and sensitive techniques that have found application in several areas such as clinical chemistry, bioanalysis, pharmaceutical analysis, toxicological analysis, and environmental analysis. The first immunoassays developed were radioimmunoassays, which are very sensitive; however, one drawback is the need to work radiochemicals. To circumvent the drawbacks, other labels such as enzymes in combination with photometric measurement have been introduced, however, at the expense of sensitivity. With the development of fluoroimmunoassays (FIAs) an improvement in sensitivity was obtained and by introducing chemiluminescence in immunoassays, a sensitivity equivalent to radioimmunoassays was achieved. In this article, different variations and techniques of luminescence immunoassays are described.

### Theory of Luminescence

Luminescence can be created by photoirradiation, which results in fluorescence, delayed fluorescence, and phosphorescence or by a chemical or biochemical reaction, which produces chemiluminescence and

bioluminescence. Molecules can absorb energy, whereby electrons are excited from the ground state to a higher state. By returning of the electrons to the ground state, the absorbed energy can be turned to radiation as fluorescence. If the transition to the ground state occurs through a metastable triplet state, phosphorescence is observed. Some species, such as lanthanides, show delayed fluorescence, which is a result of two intersystem crossings, the first from the singlet to the triplet and the second from the triplet to the singlet state.

Chemiluminescence can be a result of a chemical or biochemical reaction, an electrochemical process, or a thermal treatment.

## Fluorescence Immunoassays

### Fluorescent Probes

Fluorescent probes used as labels should have high quantum yield, emission wavelengths longer than 500 nm, and Stokes' shifts of over 50 nm. They must possess suitable functional groups such as carboxylic acid, amino, hydroxyl, sulfonic acid residues, acid anhydride, acid chloride, diazonium, isocyanate, isothiocyanate, *N*-hydroxysuccinimide, maleimide.

Commonly used probes are fluorescein derivatives, rhodamine derivatives, polycyclic aromatic hydrocarbons, coumarines, amine reagents such as fluorescamine or NBD-Cl, phycobiliproteins, porphyrins, and metal chelates. The latter may show fluorescence,

delayed fluorescence, phosphorescence, or energy-transfer fluorescence.

Instead of directly labeling antigens or antibodies, the biotin-avidin (or streptavidin) binding system may be involved. Avidin or streptavidin shows extremely high binding affinity to biotin. One component of the immunoreagents is labeled with biotin and the other with avidin or streptavidin. Thereby direct labeling of proteins with dyes, which may cause inactivation or denaturation if a protein is labeled to a high degree, is circumvented. Streptavidin can bind up to 12 label molecules. This fact results in an enormous amplification of sensitivity. Activated biotinylation reagents are commercially available and there also exist several biotinylated antibodies. Furthermore, labeled streptavidin can be used as a universal reagent for different types of assays.

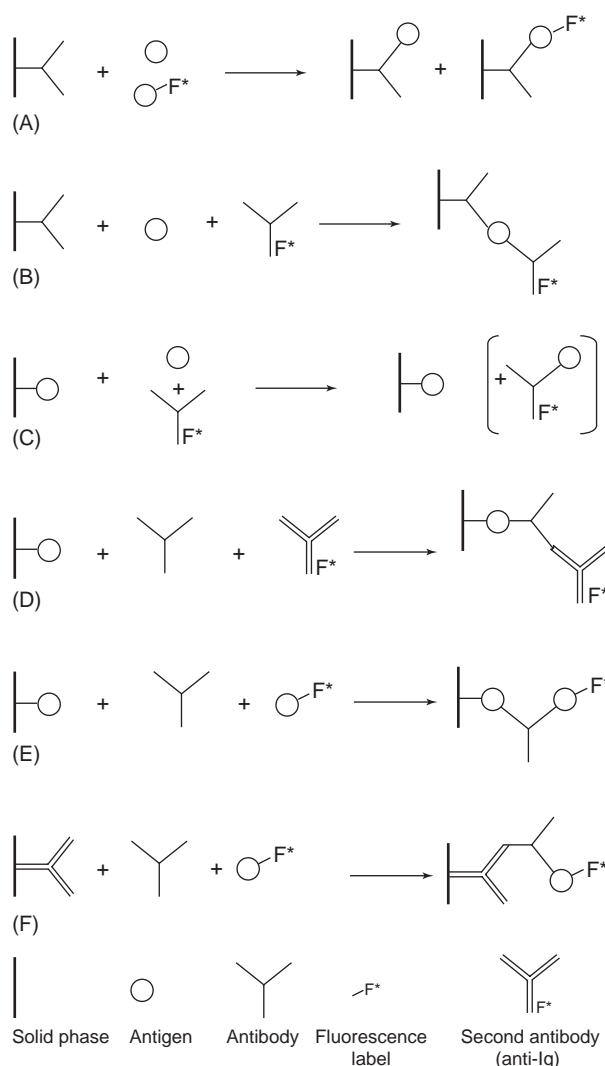
## Heterogeneous Fluorescence Immunoassays

Heterogeneous immunoassays involve a separation step of the immunocomplex formed between the labeled immunoreagent and the analyte from the free components. Antigens or antibodies are immobilized on a solid phase (columns, coated tubes, microtiter plates, or beads). Mostly polystyrene, polyacryl, agarose, glass, and more recently magnetic beads are used. Heterogeneous assays are usually more sensitive compared to homogeneous assays.

There are different formats of heterogeneous fluorescence immunoassays.

In the competitive format for the determination of haptens, sample haptens, and fluorescent-labeled haptens compete for the binding sites at the immobilized antibodies (in limited amount) (**Scheme 1A**). Fluorescence can be measured either at the solid phase or in solution.

In the two-site assay, the antigens bind to the immobilized antibodies and in the second step labeled antibodies bind to second binding sites of the antigen. This approach is applicable only to large molecules such as proteins (**Scheme 1B**). In another variation the hapten is immobilized and competes with the sample hapten for the reaction with the labeled antibodies (**Scheme 1C**). For the determination of antibodies a so-called sandwich or double antibody assay can be used. The hapten to which the antibodies bind is immobilized and in the second step second labeled antibodies (anti-immunoglobulin), which are directed against the antibodies to be determined, are added (**Scheme 1D**). In another variation, labeled haptens are added after binding of the sample antibody to the immobilized hapten (**Scheme 1E**).



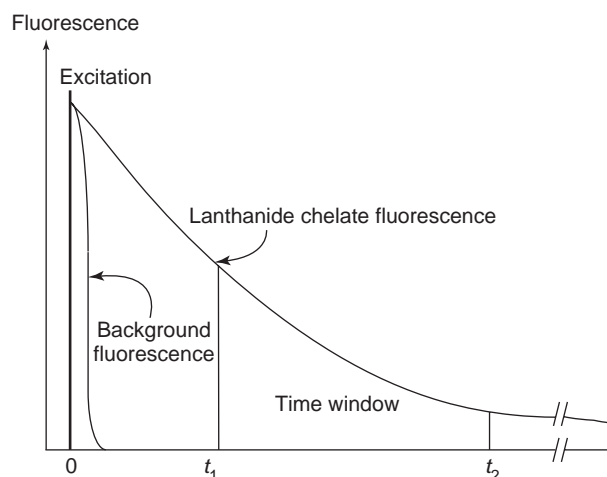
**Scheme 1**

Furthermore, the antibodies directed against the analyte antibodies can also be used in the solid phase (**Scheme 1F**). Instead of these immobilized antibodies, a protein support can also be used for binding the analyte antibodies.

Generally, immunoassays can be classified into limited reagent immunoassay systems (FIA) and excess-reagent immunometric methods (immunofluorometric assay, IFMA).

## Time-Resolved Fluoroimmunoassay

Time-resolved techniques (time-resolved fluoroimmunoassay (TRFIA)) make use of the fact that some compounds have long decay times resulting in phosphorescence or delayed fluorescence (e.g., lanthanide chelate complexes) (**Figure 1**). Apparatus for time-resolved measurements use normally a pulsed excitation source and electronically gated detectors. By this way background fluorescence can be eliminated.



**Figure 1** Principle of time-resolved fluorescence.

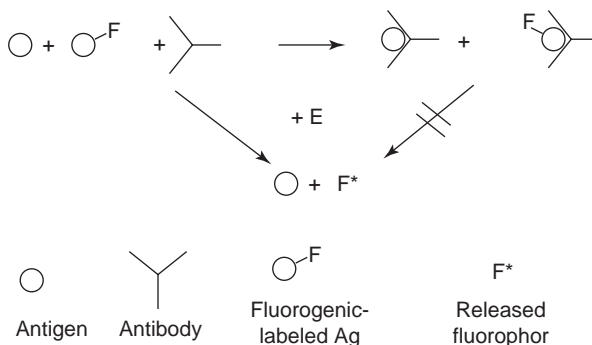
On this basis, the dissociation-enhanced lanthanide fluorescence immunoassay was developed. Mostly a two-site assay format is used whereby antibodies are immobilized at microtiter plates and after applying the sample antigen, europium-chelate-labeled antibodies bind to the sample. A commonly used chelating reagent is N1-(*p*-isothiocyanato-benzyl)-diethylene triamine. After the addition of a  $\beta$ -diketone as a dissociation enhancer, at acidic pH, europium is released and fluorescence of Eu is measured. Other TRFIAs use stabilized lanthanide chelates or stable chelates.

## Homogeneous Fluoroimmunoassays

In homogeneous immunoassay techniques no separation step is involved. A characteristic of the homogeneous immunoassay is the modulation of a signal after binding of the labeled component to the partner, forming the immunocomplex.

### Fluorescence Polarization Immunoassay

When a fluorescent dye is excited with polarized light, the fluorescence will also be polarized to a degree that is inversely related to the amount of Brownian motion occurring during the interval between absorption and emission of light. The degree of polarization of the emitted light depends on the lifetime of the excited state and the rotational motion of the molecule. The Brownian motion of larger molecules or molecular complexes (such as antigen-antibody complexes) is reduced. Small molecules show a fast rotation (rotational times of less than 1 ns) whereas large molecules have slow rotation (rotational times of 10–100 ns). Therefore, small fluorescent molecules show minimal emission polarization, while large molecules exhibit high polarization.



**Scheme 2**

The decay time of the fluorophore used as label should be longer than the rotational time of the hapten but shorter than the rotational time of the complex formed. Fluorescein, which has a fluorescence lifetime of 4.5 ns, or rhodamine are the most frequently used fluorophores. For larger antigens, fluorophores with somewhat longer decay times such as Lucifer yellow, dansyl, and umbelliferone derivatives have been tested. In a typical competitive format, the degree of polarization is inversely proportional to the concentration of the analyte.

### Release Fluoroimmunoassay

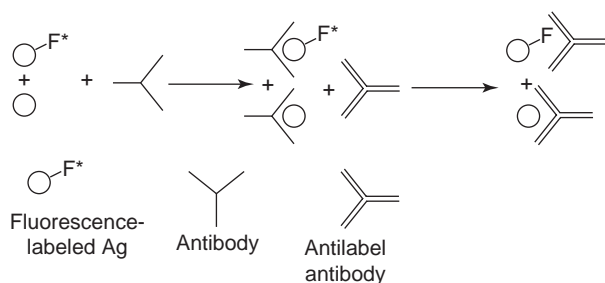
Release fluoroimmunoassay makes use of an antigen or a hapten, respectively, conjugated through ester or glycosidic bonds to a fluorophore (e.g., a coumarine derivative). The conjugate is nonfluorescent because of quenching effects of the substituent. Hydrolysis of the bond catalyzed by the binding of antibodies releases the fluorophore.

A special variation represents the enzyme-hydrolyzed antibody-protected FIA called substrate-labeled FIA. This approach, which is not to be mixed up with the enzyme-labeled fluoroimmunoassay (ELFIA), is based on cleavage of the glycosidic linkage by the addition of an enzyme, thus releasing the fluorophore (Scheme 2).

Contrary to the ELFIA, the antigen is labeled with a fluorophore and is present in limited concentration, whereas the enzyme is free and used in excess. The analyte and the substrate-labeled hapten compete for the binding sites at the antibody. The conjugate that is bound to the antibody is not accessible for the enzyme.

### Fluorescence Quenching Immunoassay

Indirect quenching FIA, also called fluorescence protection immunoassay, is based on the use of fluorescence-labeled antigens and antibodies directed against the label. When antilabel antibodies bind to the labeled probe, the fluorescence is almost completely quenched. This approach can be either

**Scheme 3**

applied to the determination of antigens (Scheme 3) or in another variation for antibodies.

### Fluorescence Enhancement Immunoassay

Fluorescence enhancement can be based on the change in environment, such as change of polarity, dielectric strength, pH, viscosity, formation of hydrogen bonds. One principle makes use of polarity sensitive fluorescent probes such as dansyl derivatives or anilinonaphthalene sulfonic acid, which show fluorescence enhancement properties upon binding to proteins, e.g., antibodies. The second variation is based on the fact that the fluorescence of a fluorescein-labeled analyte is quenched in the presence of heavy atoms, such as iodine. When the labeled probe binds to the antibody, this quenching is partially reversed.

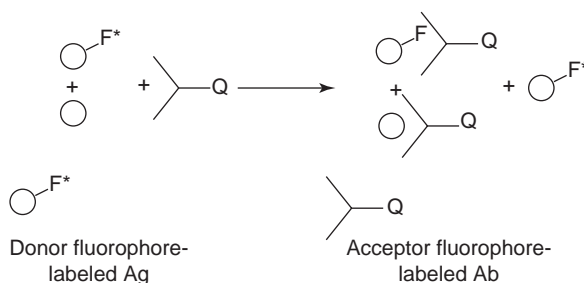
### Fluorescence Excitation Energy Transfer Immunoassay

This technique uses two different fluorescence labels. A requirement is that the emission spectrum of the donor label overlaps with the excitation spectrum of the acceptor. One fluorophore acting as energy donor is coupled to an antigen, whereas an energy accepting moiety is coupled to the antibody. Upon antigen-antibody binding, energy transfer takes place between the two groups resulting in quenching of the fluorescence of the donor (Scheme 4). The distance between the two molecules has to be less than 10 nm. In a competitive assay an increase in concentration of the analyte leads to a reduction in quenching and an increase in free labeled antigen. Thus, the measured fluorescence intensity is directly proportional to the concentration of the analyte.

As donor-acceptor label pairs, the following have been used: fluorescein-rhodamine, phycoerythrin-pyrene, phycoerythrin-Texas-Red, quinacrine-eosin, fluorescamine-fluorescein, squaric acid-galloycyanine.

### Time-Resolved Fluorescence Immunoassay

Time-resolved fluorescence also allows carrying out homogeneous immunoassays. For this purpose, lanthanide chelates, which are fluorescent in aqueous

**Scheme 4**

solution and do not require the addition of enhancement reagents, are used. For example, a fluorescent europium chelate is coupled to the hapten. After binding to the antibody, fluorescence is quenched. The fluorescence of the free conjugate measured in solution is indirectly proportional to the concentration of the analyte.

### Phase-Resolved Fluoroimmunoassay

Another technique recently applied for immunoassays is phase-resolved fluorometry (frequency-domain fluorometry). This technique is also based on different fluorescence decay times. The decay time can change upon antigen-antibody binding. Instead of pulsed excitation, in this technique the sample is excited with sinusoidally modulated light. With phase-resolved fluorometry, decay times and decay time differences within the range of subnanoseconds can be measured. The phase-resolved technique can be used also for elimination of background noise. This technique, however, has found only a few applications to immunoassays yet.

### Liposome Immunoassay

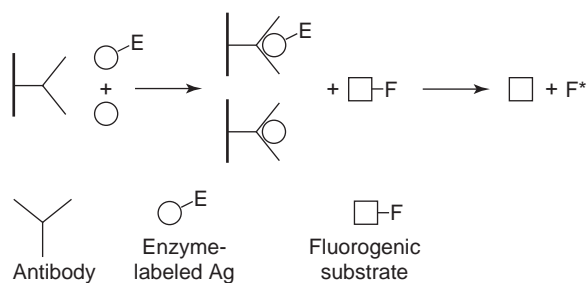
Liposomes are vesicles produced from phospholipids, cholesterol, fatty acids, etc. In typical liposome immunoassays, a fluorophore, mostly carboxy-fluorescein, is encapsulated in the liposomes. The fluorophore can be released by addition of a lysing component such as serum complement or a surfactant. In several liposome based immunoassay formats the hapten is linked to a cytolytic agent such as mellitin. After the immunoreaction the only free portion of the conjugate is capable of disrupting the liposomes releasing the fluorophore.

### Fluorescence Enzyme Immunoassay

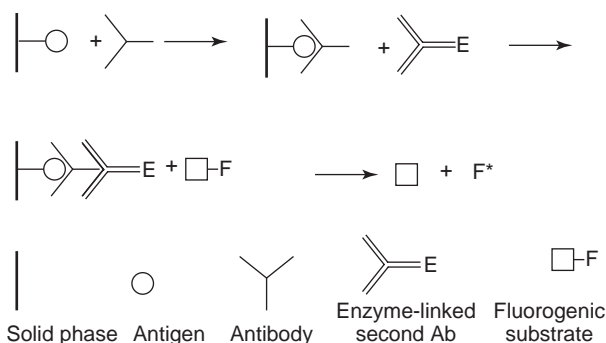
ELFIA make use of enzyme-labeled immunoreagents (antibodies or antigens) and a fluorogenic substrate. The most frequently used enzyme labels are horseradish peroxidase, alkaline phosphatase, and  $\beta$ -D-galactosidase. As coupling procedures mostly the

glutaraldehyde, carbodiimide, or periodate/sodium borohydride methods are used. Instead of direct labeling of the immunoreagents with enzymes, the biotin–streptavidin binding system may be involved. As substrates fluorogenic compounds, which exhibit fluorescence after enzymatic cleavage, are used. Most ELFIA systems are on solid-phase basis (enzyme-linked immunosorbent assay (ELISA)), although a few homogeneous assays (enzyme-multiplied immunoassay technique (EMIT)) have been described. As solid phases polystyrene balls, sepharose microbeads, nitro cellulose membranes, and magnetic particles are used. Homogeneous systems are mainly based on the fact that the antigen–antibody binding inhibits enzyme activity. Generally, heterogenic enzyme immunoassays can be classified into immunoreagent-excess or reagent-limited formats. For the determination of haptens, usually a reagent-limited assay in a competitive mode is applied (Scheme 5).

The substrate is cleaved by the enzyme-labeled hapten in the immunocomplex at the surface. The fluorescence can be measured in solution or at the surface. For the determination of high molecular mass antigens a two-site sandwich assay can be applied. A variation represents the double antibody assay, whereby the second antibody, which is directed against the hapten-specific antibody, is enzyme-labeled. Reagent-excess based assays are mainly used for the determination of antibodies in a noncompetitive format. In Scheme 6, a procedure for an indirect ELISA for the determination of antibodies is described.



Scheme 5



Scheme 6

## Phosphorescence Immunoassay

Erythrosin, eosin, and, more recently, metalloporphyrins have been checked as labels for phosphorescence immunoassays (PIAs). Usually phosphorescence can be measured only at low temperatures (77 K) to avoid molecule collisions. Since this requires a rather complicated instrumental setup, this technique found only rare application. Room temperature phosphorescence in solution is a rather rare phenomenon because of molecule collisions and quenching effects of oxygen. Some compounds such as biphenyl, biacetyl, and bromonaphthalene derivatives show room temperature phosphorescence, providing a deoxygenated solution. The addition of detergents for forming micelles or the inclusion into the cavities of cyclodextrins has been shown to support room temperature phosphorescence. Furthermore, the creation of a rigid state by adsorption or immobilization to a solid matrix has been found to enable room temperature measurements. Room temperature phosphorescence, although a promising approach, has only rarely been applied to immunoassays up to now.

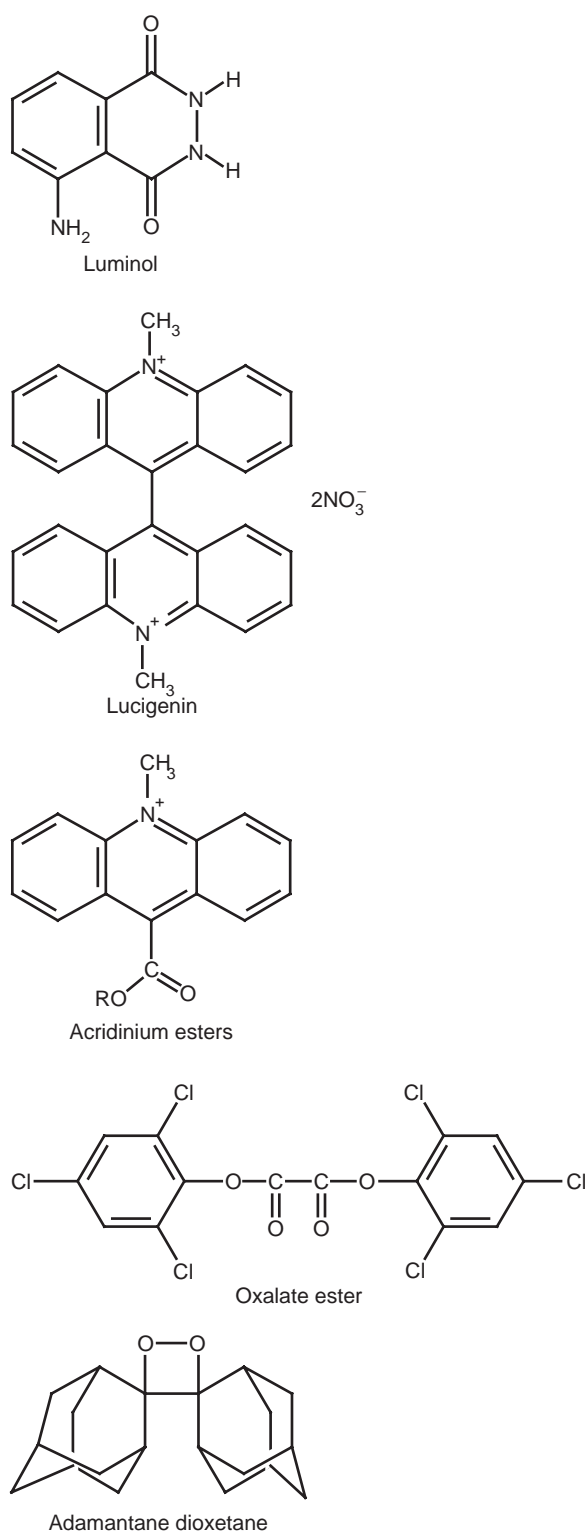
## Chemiluminescence Immunoassay

Chemiluminescence represents emitted light as a result of a certain chemical reaction. Some examples of chemiluminescent reagents are shown in Figure 2. Luminol, acridinium esters, and dioxetanes are the most frequently used labels for chemiluminescence immunoassays (CLIAs).

In several cases the chemiluminescence reaction is initiated by addition of an oxidant such as hydrogen peroxide. Certain stabilized dioxetanes can be chemically cleaved to produce chemiluminescence. Some dioxetane derivatives show thermoluminescence. When simply heated, they are transferred to an excited state and exhibit luminescence. The peroxyoxalate chemiluminescence reaction consists of different steps. The oxalate ester forms upon treatment with hydrogen peroxide an energy-rich dioxetane intermediate, which can transfer its energy to excite an acceptor fluorophore, which emits light. An increase in chemiluminescence emission in the luminol system was achieved by using certain catalysts as ‘enhancers’. Such enhanced chemiluminescence systems use, for example, horseradish peroxidase or certain phenolic molecules as enhancers.

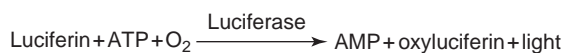
In analogy to FIA, CLIAs have also been developed using chemiluminogenic substrates. Dioxetane derivatives such as adamantyl-1,2-dioxetane phosphate are widely used substrates, which exhibit chemiluminescence upon cleavage by alkaline phosphatase.





**Figure 2** Chemiluminescence reagents.

As in FIAs, different formats such as competitive assays, two-site assays, double antibody assays, are possible in CLIAs also. Both heterogeneous and homogeneous CLIAs have been developed. However,



**Scheme 7**

the benefit of chemiluminescence is the high potential for carrying out homogeneous assays. The main advantage of CLIAs is a significantly higher sensitivity that is comparable to radioimmunoassay or even better. A further advantage is the high selectivity of chemiluminescence. Since the light is produced by a chemical reaction, there is no light scatter from a lamp and there is almost no background noise. According to the selectivity of chemiluminescence there are no interferences from the sample matrix. Moreover, the instrumental setup is very simple and inexpensive.

### Bioluminescence Immunoassay

If luminescence is a result of a biochemical reaction, the principle is called bioluminescence. The most frequently used bioluminescence system is that of the firefly. The enzyme luciferase catalyses the oxidation of luciferin as a substrate in the presence of adenosine triphosphate (ATP) (Scheme 7). Another bioluminescence system makes use of a luciferase from certain marine bacteria. A long-chain aldehyde is oxidized in the presence of luciferase, an oxidoreductase and NAD/NADH. Recently, a photoprotein isolated from the bioluminescent jellyfish *Aequorea victoria*, has been found to be an efficient bioluminescence label for immunoassays.

### Electroluminescence Immunoassay

Chemiluminescence can also be produced by electrochemical oxidation or reduction (electrogenerated chemiluminescence). Ruthenium and osmium chelates show the ability to emit light when undergoing an electrochemical process. Based on this phenomenon, an electroluminescence immunoassay has been developed using a ruthenium(II) tris(bipyridyl) chelate as a label.

## Technologies

### Immunoassays in Tubes, Cuvettes, and Microtiter Plates

For manual handling several test kits are commercially available. Special fluorometers dedicated for immunoassays have been introduced. For enhancement in sensitivity, microfluorometers with laser excitation have been developed. For special purposes fluorescence microscopes and flow-cytometers are available. Microtiter plates are very frequently used in combination

with commercially available kits and fluorometric plate readers are offered by several suppliers.

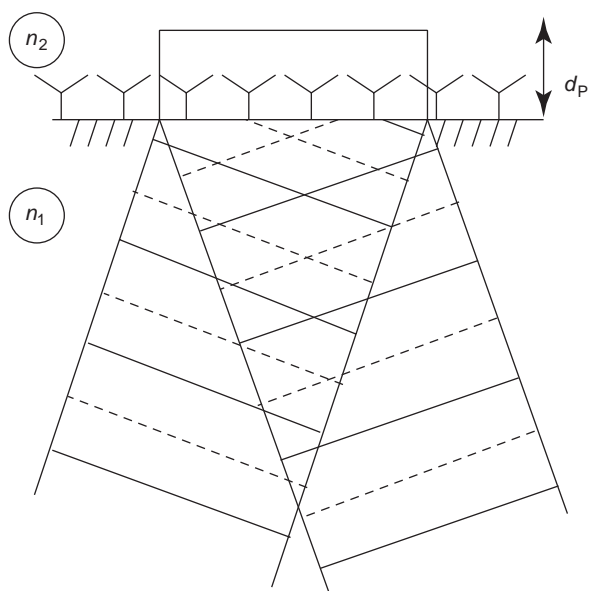
### Automated Random-Access Instrumentation

In addition to semiautomated systems, several fully automated random-access instrumentation have been developed, which enable a high sample throughput. Both for homogeneous and heterogeneous assays equipments have been developed, whereby for the latter approach frequently magnetic particles are used as carriers. Such random-access instrumentation allows the application of individual analyte and assay types. In addition to instrumentation for FIAs, automate instrumentation for CLIAs are also available. Of course, since such fully automated systems involve roboting operations, they are rather expensive.

### Immunosensors

Most of the fluoroimmunosensors are based on the principle of total internal reflection. When a light beam strikes the interface between two transparent media from the optically denser medium to the optically rarer medium ( $n_1 > n_2$ ), total internal reflection occurs. Upon reflection, some of the light (called evanescent wave) penetrates a fraction of a wavelength beyond the reflecting surface into the rarer medium of a refractive index  $n_2$  (Figure 3).

One type of immunosensors uses a planar waveguide with immobilized antibodies. Excitation is done by the evanescent wave. Alternatively, thin film cuvettes containing waveguides have been developed. A simple approach represents the use of fiber waveguides containing antibodies covalently immobilized



**Figure 3** Principle of total internal reflection:  $n_1$ , optically denser medium;  $n_2$ , optically rarer medium;  $d_P$ , penetration depth.

at the distal end (Figure 4). Such fiber waveguides are very suitable for remote measurements. Another variation consists of a fiber containing a microcuvette at the distal end provided with a dialysis membrane. This setup was used for a homogeneous fluoroimmunoassay applying the fluorescence excitation energy transfer principle involving phycoerythrin-labeled haptens and Texas-red-labeled avidin in combination with biotinylated antibodies.

Although immunosensors represent a promising approach, there are still some problems. The kinetics of antigen-antibody binding is rather slow requiring long incubations. The most important requirement on a sensor is the reversibility. The regeneration of immunosensors still remains a problem. Nonspecific binding and a relatively high background level are further drawbacks. Nevertheless, there is still research work being done to solve these problems.

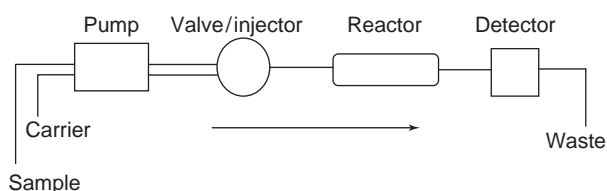
### Flow-Injection Immunoassay

Flow-injection immunoassay (FIIA) represent a relatively new technique, which involves a flow system for carrying out all steps of the immunoreaction and detection. FIIA systems can be regarded as flow-through sensors. Both homogeneous and heterogeneous FIAs have been developed. In homogeneous assays sample and reagents are mixed by the merging of two streams. A prerequisite is that binding of the antigen to the antibody modulates a standing signal produced by the labeled component. For example, fluorescence excitation energy transfer FIAs and enzyme FIAs have been developed. Heterogeneous FIAs are more frequently used. The essential piece of a heterogeneous FIIA setup is the immunoreactor. It consists of a small column or cell containing immobilized antibodies or antigens, depending on the mode. As carriers, porous glass, polymeric particles, membranes, or, more recently, magnetic particles are used. A typical instrumental setup of a heterogeneous FIIA system is shown in Figure 5.

The simplest setup consists of a peristaltic pump, an injection valve, the immunoreactor, and a detector. For special purposes multichannel pumps and valve switching systems have been developed. Sample and all reagents are injected into a flowing stream and transported to the immunoreactor where binding to the immobilized antibodies (or antigens, respectively) takes place. Detection is carried out either



**Figure 4** Fiber-optic waveguide immunosensor.



**Figure 5** Typical setup for heterogeneous FIIA.

downstream after the immunoreactor or directly at the immunoreactor in a flow-cell.

The main advantages of heterogeneous FIAs over homogeneous assays are the higher sensitivity and selectivity. Usually with heterogeneous FIAs no sample pretreatment is required, since the selective binding in the immunoreactor represents a cleanup and a preconcentration step. Washing of the unbound components is continuous. Furthermore, due to the dynamic system there is very low nonspecific binding. Heterogeneous FIAs show extremely high binding kinetics. There is a high surface to volume ratio in the immunoreactor. Contrary to static systems in FIAs the binding is not based on passive diffusion, the flowing stream actively transports the sample antigens to the binding sites of the immobilized antibodies. All immunoassay formats discussed in the previous paragraphs, including enzyme immunoassays, can be applied in FIAs. Heterogeneous FIAs have been called by some authors as immunochromatographic assays, since the immunoreactor can be regarded as a small affinity chromatographic column. However, this term should be used only if additional chromatographic processes are involved.

Since the instrumental setups for FIAs is very simple, they are predestinated for automation. Several fully automated setups have been developed. The immunoreactors are reusable. After an assay cycle the reactor can be regenerated by a washing step. Up to 500 assays with one immunoreactor have been described. Therefore, the consumption of immuno-reagents is very low. However, of course, the lifetime of the immunoreactor is limited and it has to be exchanged after a series of injections. To overcome this drawback automated membrane or particle exchange devices have been developed. Another alternative is to load the reactor column before every assay with fresh antibodies. Protein A columns, which bind immunoglobulins, have been used for this purpose. The main benefit of FIAs is the speed and flexibility for single assays. The assay times range between 1 and 15 min and the system can be adapted very easily to another type of assay. This is especially of relevance when rapid results are required for different analytes and varying problems. Regarding sample throughput, however, the existing FIIA systems cannot compete with random access carousels.

Besides fluorescence FIAs, chemiluminescence FIAs have also been developed, whereby mainly luminol and acridinium ester were used as chemiluminescence labels. High sensitivity was reported with systems using flow-cells packed with immobilized antibodies or antigens, whereby the luminescence is measured directly at the cell. The assays are carried out by stepwise injection of all components. Detection limits in the attomole range have been reported with such systems. Recently, the stereoselective recognition ability of antibodies raised against one enantiomer has been demonstrated. On this basis an enantioselective chemiluminescence FIIA for amino acids, which allows the detection of less than 0.01% of the unwanted enantiomer, has been developed.

### Sequential-Injection Immunoassay

A variation of flow-injection analysis, sequential-injection analysis (SIA), has recently been introduced. Contrary to flow-injection analysis, where the components are injected into a continuous flowing stream, in SIA injections are done sequentially directly into the reactor/detector system. The instrumental setup usually consists of a syringe pump in combination with an electronically controlled multiport valve. SIA systems have been shown to be more flexible regarding injection volumes and injection times. This technology has been demonstrated to be very efficient for carrying out immunoassays. Both fluorescence and chemiluminescence immunoassays have already been carried out with this technique.

**See also:** **Chemiluminescence:** Overview. **Fluorescence:** Time-Resolved Fluorescence. **Immunoassays, Techniques:** Enzyme Immunoassays. **Phosphorescence:** Room-Temperature. **Sequential Injection Analysis.**

### Further Reading

- Engvall E (1980) Enzyme immunoassay, ELISA and EMIT. *Methods Enzymologica* 70: 419–439.
- Gübitz G and Shellum C (1993) Flow-injection immunoassays. *Analytica Chimica Acta* 283: 421–428.
- Gübitz G, Schmid MG, Silviaeh H, and Aboul-Enein HY (2001) Chemiluminescence flow-injection immunoassays. *Critical Reviews in Analytical Chemistry* 31: 141–148.
- Hemmilä IA (1991) Applications of fluorescence in immunoassays. In: Winefordner JD (ed.) *Chemical Analysis*. Chichester: Wiley.
- Kricka LJ and Carter TJN (eds.) (1982) *Clinical and Biochemical Luminescence*. New York: Dekker.
- Lövgren T, Hemmilä I, Pettersson K, et al. (1985) Time-resolved fluorometry in immunoassay. In: Collins WP (ed.) *Alternative Immunoassays*, pp. 203–217. Chichester: Wiley.

- O'Donnell CM and Suffin SC (1979) Fluorescence immunoassays. *Analytical Chemistry* 51: 33–40.
- Price CP and Newman DJ (eds.) (1997) *Principles and Practice of Immunoassay*. Bath: The Bath Press.
- Rhys Williams AT (1988) Fluorescence polarisation immunoassay. In: Collins WP (ed.) *Complementary Immunoassays*, pp. 135–147. Chichester: Wiley.
- Ullmann EF, Ballet NF, Brinkley JM, *et al.* (1980) Homogeneous fluorescence immunoassays. In: Nakamura RM, Dito WR, and Tucker ES (eds.) *Immunoassays: Clinical Laboratory Techniques for the 1980s*, pp. 13–43. New York: Alan R. Liss.
- Weeks I (1992) Chemiluminescence immunoassay. In: Svehla G (ed.) *Comprehensive Analytical Chemistry*, vol. 29. Amsterdam: Elsevier.

## INDICATORS

Contents

**Acid-Base**

**Redox**

**Complexometric, Adsorption, and Luminescence Indicators**

### Acid-Base

**J Barbosa**, University of Barcelona, Barcelona, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

An acid–base titration is a quick and convenient method for the quantitative analysis of substances with acidic or basic properties. Many inorganic and organic acids and bases can be titrated in aqueous media, but others, mainly organic, are insoluble in water. Fortunately, most of them are soluble in organic solvents; hence they are conveniently determined by nonaqueous acid–base titrimetry. Although acid–base titrations can usually be followed potentiometrically, visual endpoint detection is quicker and can be very precise and accurate if the appropriate indicator is chosen.

In aqueous solutions the choice of a suitable indicator for an acid–base titration is usually quite simple. It is only necessary to choose an indicator with a good quality of color change and with a transition pH range that encompasses the pH at the equivalence point of the titration. The quality of color changes and transition pH ranges of many indicators has been determined for aqueous solutions. The use of chemical indicators in most organic solvents can be more complicated and less straightforward than it is in water, but fortunately in recent years the chemistry of nonaqueous solutions has been developed to such an extent that it is possible with a reasonable degree of certainty to predict the behavior of a substance in a given solvent.

To select the most suitable indicator in each solvent in order to titrate a particular acid or base

with good accuracy and precision, it is necessary to know the pK values (see below) or transition pH ranges of a series of indicators with good quality of color change over the whole useful acidity range in each of the selected solvents, but it is also necessary to know the equivalence range of each particular acid titration curve. The discussion of acid–base indicators for titration in nonaqueous solvents in this article is intended to be critical rather than exhaustive and has the aim of solving the main problem of the use of chemical indicators. It therefore focuses on the behavior and choice of indicators at visual titrations in solvents that show practical advantages for acid or base titrations. In this way the most widely used and advantageous solvents for the titration of bases and acids and their mixtures are considered, such as water, acetic acid, acetonitrile, dimethylformamide, isopropyl alcohol, and *t*-butyl alcohol. Obviously, many other solvents have been used in everyday work, but the selected solvents allow particular quantitative problems to be resolved. The use of chemical indicators in mixed solvents is not discussed because there are very few systematic studies of chemical equilibria in these mixed solvents, although some approaches are possible from data on indicator behavior in several pure solvents and taking into account the effect of preferential solution. At the present time, a series of indicators over the whole useful acidity range in each of the chosen solvents has been proposed. Chromatic parameters of transition ranges for each indicator and constants involved in color changes are known. From the solvent selection, from these sets of indicators, and from the equivalence range of each particular titration curve in the solvent chosen it should be possible to select the best fit for the problem at hand.

- O'Donnell CM and Suffin SC (1979) Fluorescence immunoassays. *Analytical Chemistry* 51: 33–40.
- Price CP and Newman DJ (eds.) (1997) *Principles and Practice of Immunoassay*. Bath: The Bath Press.
- Rhys Williams AT (1988) Fluorescence polarisation immunoassay. In: Collins WP (ed.) *Complementary Immunoassays*, pp. 135–147. Chichester: Wiley.
- Ullmann EF, Ballet NF, Brinkley JM, *et al.* (1980) Homogeneous fluorescence immunoassays. In: Nakamura RM, Dito WR, and Tucker ES (eds.) *Immunoassays: Clinical Laboratory Techniques for the 1980s*, pp. 13–43. New York: Alan R. Liss.
- Weeks I (1992) Chemiluminescence immunoassay. In: Svehla G (ed.) *Comprehensive Analytical Chemistry*, vol. 29. Amsterdam: Elsevier.

## INDICATORS

Contents

**Acid-Base**

**Redox**

**Complexometric, Adsorption, and Luminescence Indicators**

### Acid-Base

**J Barbosa**, University of Barcelona, Barcelona, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

An acid–base titration is a quick and convenient method for the quantitative analysis of substances with acidic or basic properties. Many inorganic and organic acids and bases can be titrated in aqueous media, but others, mainly organic, are insoluble in water. Fortunately, most of them are soluble in organic solvents; hence they are conveniently determined by nonaqueous acid–base titrimetry. Although acid–base titrations can usually be followed potentiometrically, visual endpoint detection is quicker and can be very precise and accurate if the appropriate indicator is chosen.

In aqueous solutions the choice of a suitable indicator for an acid–base titration is usually quite simple. It is only necessary to choose an indicator with a good quality of color change and with a transition pH range that encompasses the pH at the equivalence point of the titration. The quality of color changes and transition pH ranges of many indicators has been determined for aqueous solutions. The use of chemical indicators in most organic solvents can be more complicated and less straightforward than it is in water, but fortunately in recent years the chemistry of nonaqueous solutions has been developed to such an extent that it is possible with a reasonable degree of certainty to predict the behavior of a substance in a given solvent.

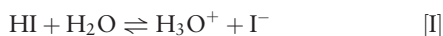
To select the most suitable indicator in each solvent in order to titrate a particular acid or base

with good accuracy and precision, it is necessary to know the pK values (see below) or transition pH ranges of a series of indicators with good quality of color change over the whole useful acidity range in each of the selected solvents, but it is also necessary to know the equivalence range of each particular acid titration curve. The discussion of acid–base indicators for titration in nonaqueous solvents in this article is intended to be critical rather than exhaustive and has the aim of solving the main problem of the use of chemical indicators. It therefore focuses on the behavior and choice of indicators at visual titrations in solvents that show practical advantages for acid or base titrations. In this way the most widely used and advantageous solvents for the titration of bases and acids and their mixtures are considered, such as water, acetic acid, acetonitrile, dimethylformamide, isopropyl alcohol, and *t*-butyl alcohol. Obviously, many other solvents have been used in everyday work, but the selected solvents allow particular quantitative problems to be resolved. The use of chemical indicators in mixed solvents is not discussed because there are very few systematic studies of chemical equilibria in these mixed solvents, although some approaches are possible from data on indicator behavior in several pure solvents and taking into account the effect of preferential solution. At the present time, a series of indicators over the whole useful acidity range in each of the chosen solvents has been proposed. Chromatic parameters of transition ranges for each indicator and constants involved in color changes are known. From the solvent selection, from these sets of indicators, and from the equivalence range of each particular titration curve in the solvent chosen it should be possible to select the best fit for the problem at hand.



## Indicator Behavior

Acid–base indicators are generally weak protolytes that change color in solution according to the pH. The acid–base equilibrium of a weak acid type of indicator (HI) in water can be represented as



The acid, HI, and the conjugate base,  $\text{I}^-$ , have different colors. The equilibrium expression for this process is

$$K = \frac{[\text{I}^-][\text{H}_3\text{O}^+]}{[\text{HI}]} \quad [1]$$

This can be expressed in logarithmic form:

$$\text{pH} = \text{p}K + \log \frac{[\text{I}^-]}{[\text{HI}]} \quad [2]$$

From eqn [2], it can be predicted whether the indicator is in its acid or base form, depending on the pH. The last term is the concentration ratio between the base and the acid forms of the indicator, and thus determines the color of the solution.

The human eye has limited sensitivity for distinguishing the components of a mixture of colors. It can be accepted as a working rule, however, that for a mixture of two complementary colors the eye can only detect a change in hue when the concentration

ratio is varied between 10 and 1/10. If these two limiting ratios are substituted into eqn [2], the well-known relation  $\text{pH} = \text{p}K \pm 1$  giving the color transition pH range of the acid–base indicator is obtained. This interval represents the pH region in which the acid–base indicator fully changes its color. Thus the color transition pH range is located in a pH region that depends on the protolysis constant of the indicator. A similar relationship is easily derived for an indicator of the basic type.

However, the transition pH range is often not symmetrically distributed with respect to the pK value of the indicator because the spectral sensitivity of the human eye is not equal throughout the whole visible spectrum. This is why the transition interval is best defined by experiment and often reported in the literature. In Table 1 the transition pH ranges of a wide series of indicators with good quality of the color change that cover the whole useful acidity range in water are presented, together with the pK values. From these values it is seen, as expected, that the interval is not symmetrical around the pK value and also that some indicators require smaller concentration ratio changes and others larger.

If an indicator is to be intelligently and scientifically applied, it is necessary to have quantitative information on the indicator pK values, transition pH range, optimum concentration for the titration arrangement and quality of color changes of the indicator. This is not in itself sufficient – quantitative

**Table 1** pK values, optimum concentrations, and transition pH ranges of indicators in water and in solvents used for the titration of weak acids and their mixtures

Indicator (dissociation model) <sup>b</sup>	Optimum concentration (mol l <sup>-1</sup> )	Water		Isopropyl alcohol		t-Butyl alcohol		Dimethylformamide	
		pK	Transition pH range	pK	Transition pH range	pK	Transition pH range	pK	Transition pH range
Thymol blue(I) (II)	$2 \times 10^{-4}$	1.65	1.2–2.8	4.95	3.8–6.7	6.87	6.8–8.1		
Methyl orange(I) (II)	Sat./3 <sup>a</sup>	3.46	3.1–4.4	2.40	1.8–3.7				
Cresol red(I) (II)	$8 \times 10^{-5}$		0.2–1.8	4.33	3.2–5.7	6.62	6.4–7.8		
Bromophenol blue(IV)	$1 \times 10^{-4}$	4.10	3.0–4.6	8.80	6.8–10.2	11.01	10.1–12.1		
Bromocresol green(IV)	$8 \times 10^{-5}$	4.90	4.0–5.6	9.66	8.0–10.7	12.12	10.5–13.2	9.06	7.2–10.0
Azo violet(II)	$2 \times 10^{-5}$							9.15	8.0–9.2
Methyl red(I) (II)	$6 \times 10^{-5}$	5.00	4.4–6.2	11.23	11.0–12.5	14.03	13.4–15.6		
Bromocresol purple(IV)	$1 \times 10^{-4}$	6.40	5.2–6.8	11.64	9.9–13.6	14.15	12.0–15.3		
Bromothymol blue(IV)	$1 \times 10^{-4}$	7.30	6.2–7.6	13.34	11.5–14.2	15.95	15.2–16.9	12.96	11.4–13.7
Neutral red(III)	$8 \times 10^{-5}$	7.40	6.8–8.0	7.24	6.2–9.0	9.30	8.5–11.4		
Phenol red(IV)	$3 \times 10^{-5}$	8.00	6.4–8.0					15.37	13.5–16.7
m-Cresol purple(IV)	$7 \times 10^{-5}$	8.32	7.4–9.0					15.16	14.1–15.8
Cresol red(IV)	$8 \times 10^{-5}$	8.46	7.2–8.8			16.93	14.6–18.5		
Phenolphthalein(IV)	$3 \times 10^{-5}$	9.60	8.0–10.0						
Thymol blue(IV)	$2 \times 10^{-4}$	9.20	8.0–9.6	15.0	14.2–16.7	18.75	16.2–18.5	15.24	13.5–15.9
Thymolphthalein(IV)	$2 \times 10^{-5}$	9.70	9.4–10.6						
Azo violet(IV)	$2 \times 10^{-5}$	12.00	11.0–13.0					13.52	13.0–13.6

<sup>a</sup>Sat. = saturated solution.

<sup>b</sup>See below.

information on the equilibria of the titration system is also required to supply the context. The most important part of the titration curve is in the vicinity of the equivalence point, the equivalence region of the titration curve. It is convenient to regard the equivalence region as the part of the curve corresponding to 99.9–100.1% titration. However, in some cases, the equivalence region is considerably shorter than 6 pH units. In such cases, which are not especially rare in analytical practice, it is convenient to accept a less stringent criterion for accuracy, for instance to accept an error of  $\pm 1\%$ . In this case we may conveniently refer to the  $\pm 1\%$  equivalence region.

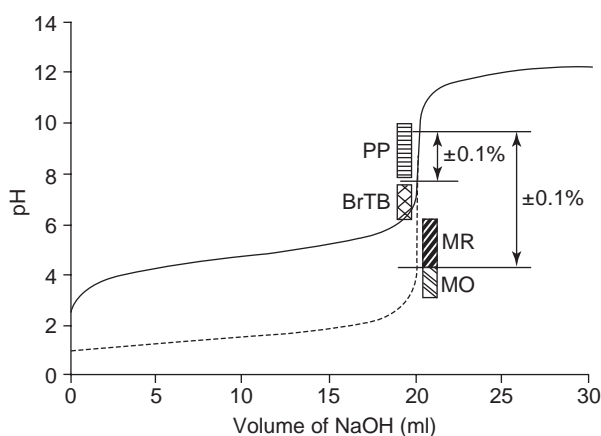
The equivalence point of the titration has been defined as the point at which the stoichiometric amount of reagent has been added, and the endpoint is the point at which the indicator signal changes. The pH of the equivalence point will be indicated by a suitable indicator. The titration error of the endpoint as visualized by means of indicators consists of three different parts.

(1) The chemical error is caused by the fact that the indicator does not change at exactly the equivalence region, i.e., the pH value of the transition pH range of the indicator differs from the pH values of the equivalence region. (2) The visual discrimination error corresponds to the deviation that originates from the limited capability of the eye in remembering or comparing colors. (3) The indicator error follows from the fact that a certain amount of the titrant will be consumed by the indicator itself.

Errors (2) and (3) are negligibly small in comparison to error (1) and consequently in the selection of a suitable indicator only the magnitude of the chemical error is of great importance. Thus the appropriate acid–base indicator must have its transition pH range within the equivalence region.

Taking into account the wide series of indicators established in water and given in Table 1, Figure 1 shows that the selection of an indicator for the titration of a strong acid with a strong base is not critical when the reagent concentration is  $\sim 0.1 \text{ mol l}^{-1}$ . Thus methyl red, bromothymol blue and phenolphthalein are suitable for detecting the endpoint, because they ensure an accurate analysis with an error less than  $\pm 0.1\%$ ; these limits are shown by horizontal lines in Figure 1 at pH values 4.00 and 10.00. The indicator methyl orange cannot be strongly recommended.

Figure 1 also indicates that the choice of indicator for the titration of a weak acid is more limited than that for a strong acid. The transition pH range of phenolphthalein, which changes color in the basic regions, coincides fairly well with the equivalence region, and hence this indicator ensures an accurate



**Figure 1** Titration curves of  $0.1 \text{ mol l}^{-1}$  acetic acid (—) and  $0.1 \text{ mol l}^{-1}$  hydrochloric acid (---) with  $0.1 \text{ mol l}^{-1}$  sodium hydroxide. Indicator transition pH ranges are indicated for methyl orange (MO), methyl red (MR), bromothymol blue (BrTB), and phenolphthalein (PP).

determination. The other indicators shown in the figure, bromothymol blue, and especially methyl red and methyl orange are not recommended for this determination.

Indicator selection for titration of a base with a strong acid is based upon the same considerations as noted for the titration of an acid with strong base. Clearly, indicators with acidic transition ranges must be employed for weak bases.

From these examples it is obvious that the weaker the protolyte the more difficult it is to determine. The  $pK$  value of the titrated protolyte ( $pK_a$  or  $pK_b$ ) should not be greater than 5–6. Protolytes weaker than this are determined with greater error in aqueous solutions, as the color transition of the indicators is not sharp enough. In such cases nonaqueous solvents should be used.

## Indicators for Titration of Acids in Nonaqueous Solvents

Acid–base indicators for titrations in nonaqueous solvents are normally weak protolytes. For dissociation model I as given in Table 1, the acid–base equilibrium of a weak acid type of indicator in the pure solvent can be represented in water by the following reaction:



HS denotes the solvent,  $\text{H}_2\text{S}^+$  the lyonium and  $\text{I}^-$  the conjugate base of the weak acid indicator HI: HI and  $\text{I}^-$  have different colors. The equilibrium constant of the acid–base interaction can be expressed in the following logarithmic form where  $\gamma_{\text{I}^-}$  represents

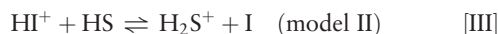
the activity coefficient of the chemical species  $I^-$ :

$$pK = pH - \log \frac{[I^-]}{[HI]} - \log \gamma_{I^-} \quad [3]$$

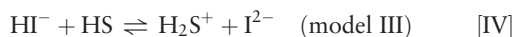
As shown in eqn [3], in nonaqueous solvents activity coefficients must be taken into account to relate the color of an indicator and the pH. Moreover, acid-base indicators may be not only neutral (model I), but also positively or negatively charged. The ionic strength of the solution modifies activity coefficients in a different way for each kind of dissociation model, according to the ionic charge of chemical species.

Many nonaqueous solvents have been proposed as suitable media for the determination of substances with weakly acidic properties or of their mixtures, but only some of them have found widespread use, because of their suitable characteristics and practical advantages. Thus, liquid ammonia and ethylene diamine (1,2-diaminoethane) show unsuitable physical characteristics and high basicity and absorb atmospheric carbon dioxide. Dimethyl sulfoxide and pyridine are unpleasant to use and dangerous in everyday work, although a series of chemical indicators have been established for use in these two solvents. On the other hand, *N,N*-dimethylformamide (DMF) is a very popular solvent for the titration of acids and mixtures of acids, and isopropyl and *t*-butyl alcohols show a large resolution of acid strength and allow to successfully solve acid mixtures.

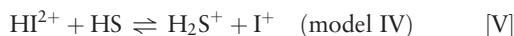
The series of indicators studied in the use of nonaqueous solvents for the titrations of acids, isopropyl and *t*-butyl alcohols, and dimethylformamide exhibit three more models of acid dissociation shown in the following reactions and equations:



$$pK = pH - \log \frac{[I]}{[HI^+]} + \log \gamma_{HI^+} \quad [4]$$



$$pK = pH - \log \frac{[I^{2-}]}{[HI^-]} - \log \frac{\gamma_{I^{2-}}}{\gamma_{HI^-}} \quad [5]$$



$$pK = pH - \log \frac{[I^+]}{[HI^{2+}]} + \log \frac{\gamma_{HI^{2+}}}{\gamma_{I^+}} \quad [6]$$

$pH_0$ , the pH value at zero ionic strength, is used to refer to the color changes.  $pH_0$  values of the transition pH ranges can be transformed into pH values with knowledge of the model of indicator dissociation and approximative ionic strength value. Table 2 shows the effect of ionic strength on activity coefficients in the solvents considered; owing to the high relative permittivity of DMF,  $\epsilon = 37$ , this effect is much less in DMF than in isopropyl alcohol and *t*-butyl alcohol.

Even at zero ionic strength the  $pK$  values of the indicators do not agree with the pH for the maximum change of color ( $pH_{mcc}$ ), as a result of the different sensitivity of human eye to different colors. The  $pH_{mcc}$  values are displaced, relative to the  $pK$  values, towards pH values at which the concentration of the less intensely colored form is higher.

A series of indicators with good quality of color change have been established that cover the whole useful acidity range in isopropyl alcohol, *t*-butyl alcohol and DMF and are shown in Table 1 together with the dissociation model of the indicator. Table 1 also shows the optimum concentration of the indicator, referred to 10 mm path length, obtained considering the sequence of color changes of the indicators described by the color spaces CIELAB ( $L^*$ ,  $a^*$ ,  $b^*$ ), and CIELUV ( $L^*$ ,  $u^*$ ,  $v^*$ ), recommended by CIE, along with the indicator dissociation constants and their transition pH ranges in each solvent (see below).

To compute the pH of the equivalence point in order to select the most appropriate indicator, a discussion of ionic equilibria in selected solvents is needed. In aqueous solution, calculation of the pH at the equivalence point is quite simple. In the same way, because of the relatively high relative permittivity of DMF, pH values can be calculated on the basis of simple dissociation of

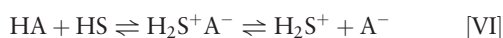
**Table 2** Influence of ionic strength ( $I$ ) on activity coefficient in isopropyl and *t*-butyl alcohols and in dimethylformamide (DMF)

$I \text{ (mol kg}^{-1}\text{)}$	$\log \gamma_{HI^+} = \log \gamma_{I^-}$			$\log(\gamma_{I^{2-}} / \gamma_{HI^-}) = \log(\gamma_{HI^{2+}} / \gamma_{I^+})$		
	<i>Isopropyl alcohol</i>	<i>t</i> -Butyl alcohol	DMF	<i>Isopropyl alcohol</i>	<i>t</i> -Butyl alcohol	DMF
$1 \times 10^{-6}$	-0.00	-0.01	0.00	-0.01	-0.03	-0.00
$1 \times 10^{-5}$	-0.01	-0.03	0.00	-0.04	-0.08	-0.01
$1 \times 10^{-4}$	-0.04	-0.07	-0.02	-0.11	-0.22	-0.05
$5 \times 10^{-4}$	-0.07	-0.14	-0.03	-0.21	-0.42	-0.10
$1 \times 10^{-3}$	-0.09	-0.18	-0.05	-0.28	-0.54	-0.14
$1 \times 10^{-2}$	-0.18	-0.33	-0.13	-0.55	-1.00	-0.39

the acid and complete dissociation of the salts, but taking into account the activity coefficients ( $\gamma$ ), which can be calculated by means of the Debye-Hückel equation  $\log \gamma = -A I^{1/2} / (1 + B \alpha I^{1/2})$  where  $A = 1.594$ ,  $B = 0.480 \times 10^{10}$ , and  $\alpha = 5 \times 10^{-10} \text{ m}$  and  $I$  is the ionic strength.

Ionic equilibria in solvents of low relative permittivity have been described in the literature and equations for calculating titration curves have also been proposed. The equations involve the partial dissociation of the acids, bases, and salts in these media, and the effect of the activity coefficients. The equations can be applied to pH and titration curve computation in solvents such as *t*-butyl and isopropyl alcohols, ethylene diamine, pyridine, or tetrahydrofuran.

In solvents of low relative permittivity, because of ion pair formation, the dissociation of an acid (HA) must be considered in two steps:



Usually an overall acidity constant,  $K_a$ , is defined by

$$K_a = \frac{[\text{H}_2\text{S}^+][\text{A}^-]\gamma^2}{[\text{HA}] + [\text{H}_2\text{S}^+\text{A}^-]} \quad \text{[7]}$$

For a base, expressions similar to those for acid can be written:



$$K_b = \frac{[\text{BH}^+][\text{S}^-]\gamma^2}{[\text{B}] + [\text{BH}^+\text{S}^-]} \quad \text{[8]}$$

At the equivalence point, solution equilibria of salts have to be considered. Because salts are constituted of ions, only the dissociation equilibrium must be taken into account. On the basis of equilibrium of



The constant ( $K_{\text{salt}}$ ) can be defined

$$K_{\text{salt}} = \frac{[\text{BH}^+][\text{A}^-]\gamma^2}{[\text{BH}^+\text{A}^-]} \quad \text{[9]}$$

The autoprotolysis constant for the solvent is as given in the following equation:

$$K_{\text{ap}} = [\text{H}_2\text{S}^+][\text{S}^-]\gamma^2 \quad \text{[10]}$$

For computing pH at the equivalence point it is assumed that a few ion pairs and ions can react with the solvent to give the undissociated acidic and basic forms. Total concentrations of acid and base, however, must be the same. The mass

balance is then

$$[\text{HA}] + [\text{H}_2\text{S}^+\text{A}^-] + [\text{A}^-] = [\text{B}] + [\text{BH}^+\text{S}^-] + [\text{BH}^+] \quad \text{[11]}$$

Taking into account that in solvents of low relative permittivity the dissociation constants have low values, the concentrations  $[\text{A}^-]$  and  $[\text{BH}^+]$  can be omitted (moreover, they are nearly the same). Substitution of eqns [7], [8], and [10] into eqn [11] gives

$$([\text{H}_2\text{S}^+]\gamma)^2 = \frac{K_{\text{ap}}K_a}{K_b} \quad \text{[12]}$$

Thus, the pH of the equivalence point of a titration is given by

$$\text{pH}_{\text{ep}} = \frac{1}{2}(\text{p}K_{\text{ap}} + \text{p}K_a - \text{p}K_b) \quad \text{[13]}$$

As can be observed from the practical simplifications presented above, eqn [13] is of general application to any solvent of low relative permittivity if conjugated species and ionic aggregates higher than ion pairs can be neglected, which is the case for isopropyl and *t*-butyl alcohols.

In eqn [13],  $\text{p}K_b$  corresponds to the titrant. The most common titrant solution for the titration of acids in non-aqueous solvents is tetrabutylammonium hydroxide, and stable commercial solutions use methyl or isopropyl alcohol as solvents. The addition of isopropyl alcohol to *t*-butyl alcohol can be neglected because their relative permittivity values are not very different. Methanol addition causes larger variations in  $\text{p}K_a$  values. Thus, the use of titrants without or with low contents of methanol is strongly recommended. For application of eqn [13],  $\text{p}K_b$  values of tetrabutylammonium hydroxide are 2.58 in isopropyl alcohol and 4.91 in *t*-butyl alcohol.

On the other hand, the effects on dissociation equilibria of the presence of small quantities of water in common solvents used in the ordinary analytical laboratory have to be considered. The presence of water decreases the pH scale of the solvent according to the water concentration and should be controlled.

The greatest practical difficulty in the use of the above equations for selection of the most suitable indicators for a particular titration is the lack of  $\text{p}K_a$  values for nonaqueous solvents, although the literature offers some sets of values corresponding to several chemical families of compounds. To determine the resolution of acid strength in nonaqueous solvents  $\text{p}K_a$  values in the organic solvent have been plotted vs  $\text{p}K_a$  in water.  $\text{p}K_a$  values obtained in isopropyl and *t*-butyl alcohols show slopes higher than unity in the families studied. However, it is possible to use these relationships to predict the approximate value of the  $\text{p}K_a$  of the acid to be determined if its  $\text{p}K_a$  value in water is known.

## Indicators for Titration of Bases in Nonaqueous Solvents

It is quite simple with the aid of Table 1 to select the most appropriate indicator for the titration of a basic substance in aqueous solution. However, if the basic substance is not soluble in water or has a weak basic character, anhydrous acetic acid is the solvent of choice for its titration in many cases.

However, as has been shown by Kolthoff and Bruckenstein, there is no direct relationship between pH and the color of an indicator in acetic acid medium. Thus, it is not feasible to specify transition ranges of indicators in terms of pH or millivolts in this solvent. Taking into account that in acetic acid medium the titrant is usually a solution of perchloric acid in acetic acid, since it is the strongest acid in this medium, and moreover is very stable and no changes are caused in the reaction medium, the behavior of acid-base indicators is best represented by the general reaction



The appropriate equilibrium constant is the formation constant of indicator perchlorate  $K_f^{\text{IHClO}_4}$ , which easily relates the color of an indicator in acetic acid medium to the concentration of free perchloric acid,  $c_{\text{HClO}_4}$ :

$$\frac{\sum[\text{IH}^+]}{[\text{I}]} = K_f^{\text{IHClO}_4} c_{\text{HClO}_4} \quad [\text{14}]$$

where  $\sum[\text{IH}^+]/[\text{I}]$  is the ratio between the concentration of the acid forms of the indicator, ionized and dissociated, and the concentration of the basic form of the indicator. As a fair approximation, activities can be substituted by concentrations in acetic acid.

Critical aspects concerning the use of indicators in acetic acid medium have been studied on the basis of color theory and the behavior of the indicator in a titration process of a base B. This behavior is defined by

$$\frac{\sum[\text{IH}^+]}{[\text{I}]} = \frac{K_f^{\text{IHClO}_4}}{K_f^{\text{BHClO}_4}} \cdot \frac{c_{\text{BHClO}_4}}{c_{\text{B}}} \quad [\text{15}]$$

$K_f^{\text{IHClO}_4}$  and  $K_f^{\text{BHClO}_4}$  are the formation constants of indicator perchlorate and base perchlorate, respectively, and  $c_{\text{BHClO}_4}$  and  $c_{\text{B}}$  the concentrations of the species indicated by the subscripts. Hence, the minimum  $K_f^{\text{BHClO}_4}$  value required to titrate a particular base with a maximum error of 1% using the indicator I is given by

$$K_f^{\text{BHClO}_4} \geq \frac{10^2 \times K_f^{\text{IHClO}_4}}{\left( \frac{\sum[\text{IH}^+]}{[\text{I}]} \right)_{\text{mcc}}} \quad [\text{16}]$$

where  $(\sum[\text{IH}^+]/[\text{I}])_{\text{mcc}}$  refers to the point of maximum color change (mcc) in the indicator transition range.  $(\sum[\text{IH}^+]/[\text{I}])_{\text{mcc}}$  can be obtained on the basis of the CIE chromatic theory. From  $K_f^{\text{IHClO}_4}$  and  $(\sum[\text{IH}^+]/[\text{I}])_{\text{mcc}}$  values, the minimum  $K_f^{\text{BHClO}_4}$  required for a particular base to be titrated in acetic acid medium with a particular indicator can be determined; values are shown in Table 3.

Therefore, in order to predict the usefulness of the indicator it is necessary to know the  $K_f^{\text{IHClO}_4}$  and  $(\sum[\text{IH}^+]/[\text{I}])_{\text{mcc}}$  values of an indicator series and to know the  $K_f^{\text{BHClO}_4}$  value of the base to be titrated. There are few  $K_f^{\text{BHClO}_4}$  values available in the literature, but the following equation can be used:

$$K_f^{\text{BHClO}_4} = \frac{K_{\text{HClO}_4} K_{\text{b}}}{K_{\text{BHClO}_4} K_{\text{ap}}} \quad [\text{17}]$$

$K_{\text{ap}} = 10^{-13.9}$  is the autoprotolysis constant of acetic acid and  $K_{\text{HClO}_4} = 10^{-4.87}$ ,  $K_{\text{b}}$  and  $K_f^{\text{BHClO}_4}$  are the dissociation constants of perchloric acid, base, and base perchlorate salt defined in eqns [7], [8], and [9], respectively.

$K_{\text{b}}$  and  $K_{\text{BHClO}_4}$  for a great number of significant bases are given in the literature. Moreover,  $K_{\text{b}}$  and  $K_{\text{BHClO}_4}$  values of nonreferenced bases can also be estimated from potentiometric curves that are commonly found in the literature. Furthermore, the majority of  $K_{\text{BHClO}_4}$  values determined for different bases fall between  $10^{-5}$  and  $10^{-6}$  and the decrease in the dissociation degree of bases in acetic acid medium parallels that in water. Bases stronger than about  $\text{p}K_{\text{a}} = 5$  in water are leveled in acetic acid with a  $\log K_f^{\text{BHClO}_4}$  value of  $\sim 9$ . So, their titration curves in acetic acid are very similar. Also  $\text{p}K_{\text{a}}$  values in water must be equal to or greater than 1.5 to titrate the base in acetic acid medium.

On the other hand, despite interesting studies of visual titrations in acetic acid medium, in actual practice, crystal violet and methyl violet, which exhibit identical color changes, are by far the most widely used indicators. This is probably because their two chromatic transitions in acetic acid cover an unusually large potential range of  $\sim 200$  or  $300$  mV. However, in titrations of bases with  $\text{p}K_{\text{b}} > 6.8$ , the endpoint of crystal violet is assigned to a color that is not in agreement with the complete chromatic transition. Consequently, the visual endpoint location is difficult and depends on the base to be titrated and also on its concentration.

The most appropriate indicator in each case for the titrations in acetic acid medium that guarantees accuracy and precision can be selected with the aid of Table 3. Thus, the concentration of perchloric acid at



**Table 3**  $K_f^{\text{HClO}_4}$ ,  $\sum([\text{IH}^+]/[\text{I}])_{\text{mcc}}$ ,  $(\log K_f^{\text{BHClO}_4})_{\text{min}}$  and transition  $\text{pO}_{\text{HClO}_4}$  range values in anhydrous acetic acid;  $\text{pK}$  values and transition  $\text{pH}$  range in acetonitrile; and optimum concentration of indicators used for the titration of weak bases and their mixtures

Indicator (dissociation model in acetonitrile)	Optimum concentration ( $\text{mol l}^{-1}$ )	$\log K_f^{\text{HClO}_4}$	Transition $\text{pO}_{\text{HClO}_4}$	Acetic acid		Acetonitrile	
				$(\sum[\text{IH}^+]/[\text{I}])_{\text{mcc}}$	$(\log K_f^{\text{BHClO}_4})_{\text{min}}$	$\text{pK}$	Transition $\text{pH}$ range
Bromocresol green(IV)	$1 \times 10^{-4}$					18.80	17.1–19.2
Neutral red(II) (III)	$8 \times 10^{-5}$					16.34	15.1–18.4
Quinaldine red(V)	$8 \times 10^{-5}$	7.59	6.3–8.4	2.03	9.28	9.41	7.6–10.0
4'-Dimethylaminobenzalrhodanine	$5 \times 10^{-5}$	6.71	5.4–7.4	1.50	8.53		
Tropaeolin OO(II)	$6 \times 10^{-5}$	6.45	5.7–6.9	2.33	8.08	8.96	7.5–10.5
Brilliant green(V)	$8 \times 10^{-5}$	6.00	4.3–6.8	4.26	7.37	6.64	4.2–7.9
Crystal violet	$6 \times 10^{-5}$	5.80	5.0–6.8	1.30	7.80		
Malachite green(V)	$7 \times 10^{-5}$	5.30	3.6–6.5	4.55	6.64	5.98	3.8–7.2
p-Naphtholbenzein(III)	$4 \times 10^{-4}$	5.00	4.7–7.0	0.54	7.25	7.18	6.8–10.5
Nile blue	$9 \times 10^{-5}$	4.00	2.8–5.6	0.66	6.18		
Neutral red(V)	$4 \times 10^{-5}$					4.67	3.4–6.4

the equivalence point of a titration in acetic acid medium is given by

$$c_{\text{HClO}_4} = \left( \frac{c_{\text{BHClO}_4}}{K_f^{\text{BHClO}_4}} \right)^{1/2} \quad [18]$$

$c_{\text{BHClO}_4}$  is the concentration of the base perchlorate at the equivalence point that corresponds to the initial concentration of the base and the indicator series in Table 3 also provides transition ranges in terms of  $\text{pO}_{\text{HClO}_4}$ .

Thus, for a titration of a base B, it becomes easy to estimate the  $c_{\text{HClO}_4}$  value of the equivalence point and according to this value to select the most suitable indicator of the series mentioned.

Table 3 shows optimum concentrations of the indicators referred to 10 mm path length of solution.  $(\sum[\text{IH}^+]/[\text{I}])_{\text{mcc}}$  refers to the point of maximum color change in the indicator range and transition  $\text{pO}_{\text{HClO}_4}$  ranges for a series of indicators in acetic acid medium that show a good quality of color change in this medium. All these values are obtained on the basis of chromatic theory of the CIE.

On the other hand, although for many years acetic acid has been the most popular solvent for titration of bases, acetonitrile – which is a considerably weaker base and much weaker acid than water – shows a greater  $\text{pH}$  jump in the titration curves of bases. Acetonitrile solvation of anions is less than that of cations and, from an analytical viewpoint, conjugation is negligible for bases. This fact makes acetonitrile very suitable for the determination of very weak bases and mixtures of bases. In this medium many workers have used a solution of perchloric acid in anhydrous acetic acid as a titrant. However, for the differentiating titration of two or more bases of different strengths, it is necessary to avoid the leveling effect of acetic acid, whose presence decreases the potential break in acetonitrile, especially in the basic side. A solution of perchloric acid monohydrate in nitromethane is stable and very suitable for titrations in acetonitrile.

The precise interpretation of potentiometric titration curves for base titrations in acetonitrile and the calculation of the  $\text{pH}$  changes near the equivalence point require only knowledge of the dissociation constants,  $K_{\text{HB}^+}$ , of the protonated bases to be titrated, because in solvents of relatively high relative permittivity (such as acetonitrile) perchloric acid and perchlorate salts can be considered to be completely dissociated. The well-known expression commonly used is

$$K_{\text{HB}^+} = \frac{a_{\text{H}^+}[\text{B}]\gamma_{\text{B}}}{[\text{HB}^+]\gamma_{\text{HB}^+}} \quad [19]$$

The neutralization curves of such bases or their mixtures and the change in pH near the equivalence point can then be calculated as they are in water.

The dissociation constants of the protonated forms of a large number of bases are known. A plot of  $pK_{HB^+}$  values of bases in acetonitrile versus the  $pK_{HB^+}$  values in water is roughly linear, with a slope of 1.06. This low value of resolution of acid strength is in agreement with the observation that aprotic solvents do not much improve the resolution of acid strength of cationic acids, in contrast to the case for uncharged acids.

Thus, it is possible to calculate the neutralization curves of bases in acetonitrile and it is also possible to predict which indicator will give a sharp color change at the equivalence range when the  $pK$  values or transition pH ranges of a series of indicator with good quality of color change are known (Table 3).

Table 3 shows a series of indicators in acetonitrile medium, with a pronounced color change, covering the more useful acidity range in acetonitrile. This indicator series takes into account different kinds of equilibria, as described for indicators used in titration of acids, and refers to the transition ranges of pH at zero ionic strength in the same way as those for alcoholic and DMF media. The influence of ionic strength on activity coefficients in acetonitrile is very similar to that in DMF given in Table 2. Because of the high relative permittivity of acetonitrile ( $\epsilon = 36$ ), the activity coefficient values have less influence than in other solvents.

In Table 3, optimum concentration, dissociation model,  $pK$  values and transition pH ranges for the indicator series in acetonitrile are given. From this table and constants of the base to be titrated, it is simple to select the best indicator for a specific titration in acetonitrile medium in the same way as in water, but taking into account the activity coefficients, which can be calculated by the following reduced Debye-Hückel equation, where  $z_i$  is the ionic charge,  $\epsilon$  is the relative permittivity and  $I$  is the ionic strength:

$$\log \gamma_{HB^+} = -355z_i^2 \epsilon^{3/2} I^{1/2} \quad [20]$$

## Quality of Color Change of Indicators

Another critical aspect that must be considered for selection of a suitable indicator for an acid-base titration in a given solvent is the quality of color change of the indicator in this solvent. The quality of color change is related to the sharpness of the chromatic transition and to the sensitivity of the color change.

The color change qualities of a series of commercially available indicators with pronounced color change have been studied in several nonaqueous solvents such as acetic acid, acetonitrile, dimethylformamide, and isopropyl and *t*-butyl alcohols.

To establish the color change quality of an indicator on an objective basis, the color changes have to be characterized by chromatic parameters. The sequence of color changes of indicators can be described for both CIE and complementary chromaticity systems from the experimental values obtained in indicator solutions, by continuous measurement of the absorbance values at 10 nm intervals between 380 and 770 nm over the pH range of the color change of the indicator, and simultaneously by potentiometric measurement.

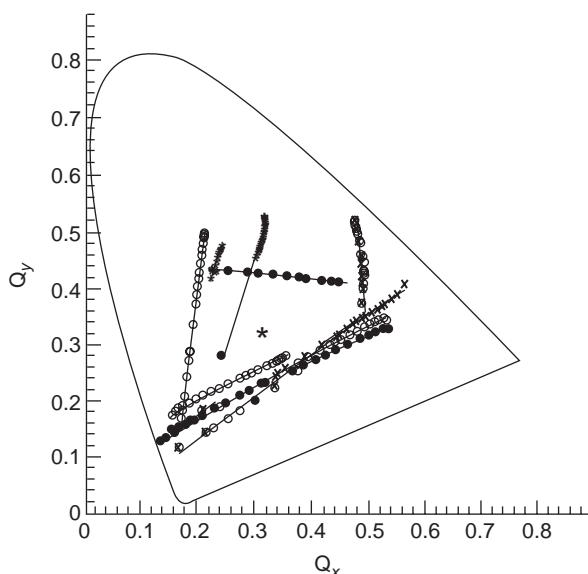
The color transitions of series of indicators in the solvents discussed above have been described by the traditional 1931 CIE system ( $x, y, Y$ ) or by the latest color spaces CIELAB ( $L^*, a^*, b^*$ ), and CIELUV ( $L^*, u^*, v^*$ ), recommended by the CIE.

The CIELAB and CIELUV systems can be used advantageously to calculate color differences  $\Delta E$  where  $L^*, a^*, b^*$  and  $L^*, u^*, v^*$  are the chromaticity coordinates of the color points:

$$\Delta E = [(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]^{1/2} \quad [21]$$

$$\Delta E = [(\Delta L^*)^2 + (\Delta u^*)^2 + (\Delta v^*)^2]^{1/2} \quad [22]$$

The complementary chromaticity coordinates ( $Q_x, Q_y$ ), which are color physical constants, have also been used to describe the color sequences of indicators (Figure 2). The plots of  $Q_x, Q_y$  coordinates of the color sequences of indicators are intersecting straight lines, denoting mixtures of two pure colored species in all instances. Thus, for example, the fact that crystal violet in acetic acid medium produces two linear



**Figure 2** Color transitions in acetic acid medium in the complementary chromaticity diagram: ★, gray point; +, quinaldine red; Δ, 4'-dimethylaminobenzalrhodanine; ◇, metanil yellow; ☆, tropaeolin 00; ●, brilliant green; ⊗, crystal violet; □, malachite green; ◇, sudan III; ○, *p*-naphtholbenzein; ×, Nile blue.

**Table 4** Color changes of indicators

Indicator	Isopropyl alcohol			<i>t</i> -Butyl alcohol			Dimethylformamide			Acetonitrile		
	SCD	$\Delta pH_{1/2SCD}$	$pH_{mcc}$	SCD	$\Delta pH_{1/2SCD}$	$pH_{mcc}$	SCD	$\Delta pH_{1/2SCD}$	$pH_{mcc}$	SCD	$\Delta pH_{1/2SCD}$	$pH_{mcc}$
Thymol blue(I)	86	1.7	5.1	150	0.7	7	73	1.2	14.8			
Methyl orange(I)	104	1.0	2.4									
Cresol red(I)	70	1.7	4.0	77	0.7	7						
Bromophenol blue(III)	92	2.1	9.0	110	1.0	11.1						
Bromocresol green(III)	98	1.3	9.3	105	1.4	11.6	74	1.4	8.4	83	1.1	18.2
Azo violet(I)							76	0.6	8.6			
Methyl red(I)	57	1.1	11.9	50	0.9	14						
Bromocresol purple(III)	80	2.2	11.8	77	1.6	14.3						
Bromothymol blue(III)	84	2.0	12.9	152	0.9	15.7	75	1.1	12.4			
Neutral red(II)	63	1.9	7.9	51	1.4	10.1				40	1.3	4.7
Phenol red(III)							86	1.6	14.8			
<i>m</i> -Cresol purple(III)							140	0.9	14.8			
Cresol red(III)				76	2.0	16.8						
Thymol blue(III)				122	1.1	17.4						
Azo violet(III)							112	0.3	13.5			
Quinaldine red(IV)										48	1.1	9.0
Tropaeolin OO(I)										70	1.5	9.5
Brilliant green(IV)										66	2.2	6.1
Malachite green(IV)										76	1.8	5.4
<i>p</i> -Naphtholbenzein(II)										60	2.0	8.2
Neutral red(IV)										50	1.9	16.9

segments shows the presence of two different equilibria involving a particular color change, each first from violet to blue and then to yellow, but only the first can be observed easily and permits results with the desired level of accuracy and precision to be obtained.

The color distribution in these chromaticity diagrams is not uniform, however, and it does not allow the direct evaluation of the quality of color changes. As a consequence, it is not possible to compare the qualities of color changes of the indicators, because the same distance has different physical meanings in different areas of the diagram. Further, complementary chromaticity data give no information about the ideal concentration for the use of the indicators.

The method of Kotrlý and Vytras has been used to obtain the optimum concentrations of the indicators for the titration arrangement used. They pointed out that the spatial transition curve of a color change can be divided into small segments related to a suitable variable, usually pH or the concentration of free perchloric acid,  $c_{HClO_4}$ , in acetic acid medium. The ratio  $\Delta pH/\Delta E$  or  $\Delta p_{HClO_4}/\Delta E$  was taken as an index of color change perceptibility.

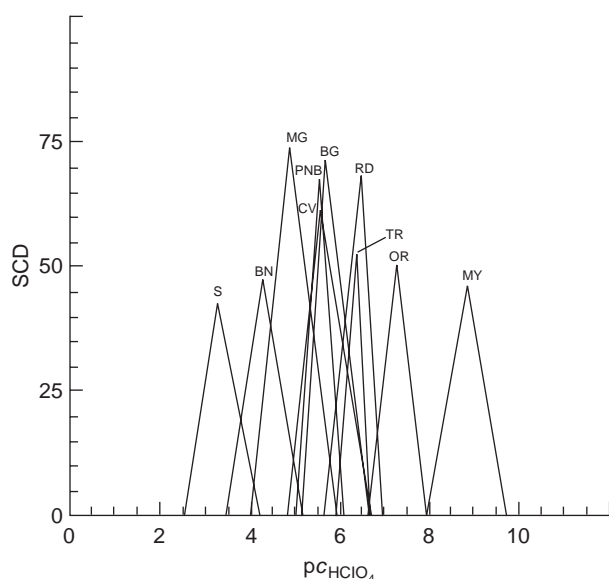
The minimum of the graph of this index versus  $\log(c/c_{ref})$  gives the optimum indicator concentration;  $c_{ref}$  is the reference experimental value of the indicator concentration,  $c$  is the indicator concentration and  $\Delta E$  is the total difference between two colors, calculated by using the CIELAB or CIELUV systems. The results obtained for the series of indicators studied are summarized in Tables 1 and 3.

To describe objectively the quality of color changes of the indicators, a modification of the method of Bhuchar and Agrawal, based on a plot of  $\Delta E/\Delta pH$  or  $\Delta E/\Delta p_{HClO_4}$  (the specific color discrimination (SCD) values) against a suitable variable, pH or  $p_{HClO_4}$ , has been used, using indicator solutions at the optimum concentration.

In this plot, the bandwidth of each peak is the transition range of the indicators in terms of pH or  $p_{HClO_4}$  in acetic acid medium, and the pH or  $p_{HClO_4}$  value at the maximum is the value for the maximum color change ( $(pH)_{mcc}$  or  $(p_{HClO_4})_{mcc}$ ). The half-bandwidth [ $(\Delta pH)_{1/2SCD}$  or  $(\Delta p_{HClO_4})_{1/2SCD}$ ] is also related to the sharpness of the chromatic transition, and the maximum of each peak (in SCD units) is taken as the sensitivity of the color change. The data obtained for the series of indicators studied in the different media are summarized in Tables 1, 3, and 4. The results obtained in acetic acid medium using CIELAB system are shown in Figure 3, which shows the usefulness of all the indicators studied in this medium and the great color change sensitivity of malachite green, brilliant green and *p*-naphtholbenzein. From Figure 3, and Tables 1, 3, and 4 it is easy to select the best indicator for a specific titration.

## Types of Acid–Base Indicators

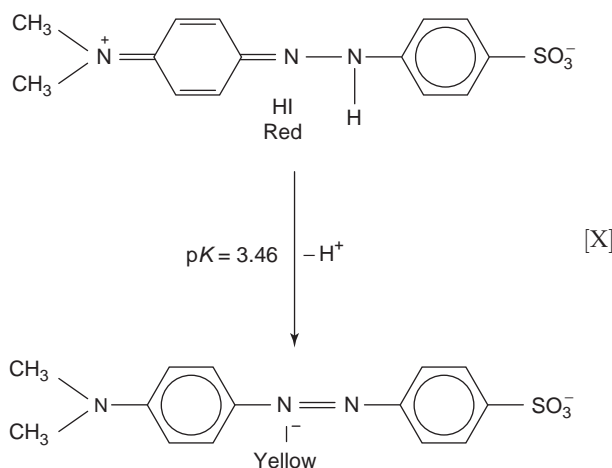
Most acid–base indicators are dyestuffs from four classes of organic compounds: azo dyes, phthaleins, sulfonphthaleins, and triphenylmethane dyes.



**Figure 3** Changes of SCD values with  $p\text{CHClO}_4$  in acetic acid medium. QR, quinaldine red; RD, 4'-dimethylaminobenzalrhodanine; MY, metanil yellow; TR, tropaeolin OO; BG, brilliant green; CV, crystal violet; MG, malachite green; S, sudan III; PNB, *p*-naphtholbenzein; BN, nile blue.

### Azo Dyes

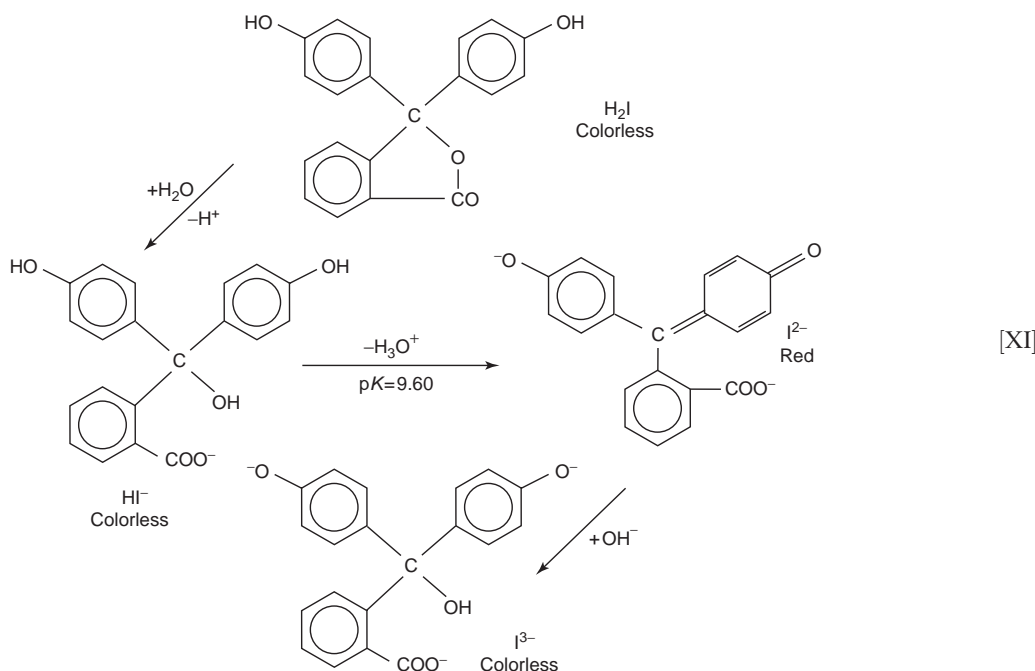
The classical azo indicators exhibit a color change from red to yellow with increasing basicity. The most commonly encountered examples are methyl orange and methyl red. The behavior of the former is described by reaction [X].



Methyl red is similar to methyl orange except that the sulfonic acid group is replaced by a carboxylic acid group. Other azo dyes commonly used in nonaqueous solvents are azo violet and tropaeolin OO.

### Phthaleins

Most phthalein indicators are colorless in moderately acidic solutions and exhibit a variety of colors in alkaline media. The structural changes causing the color changes can be illustrated with the example of phenolphthalein, which is widely used in analytical practice (reaction [XI]).



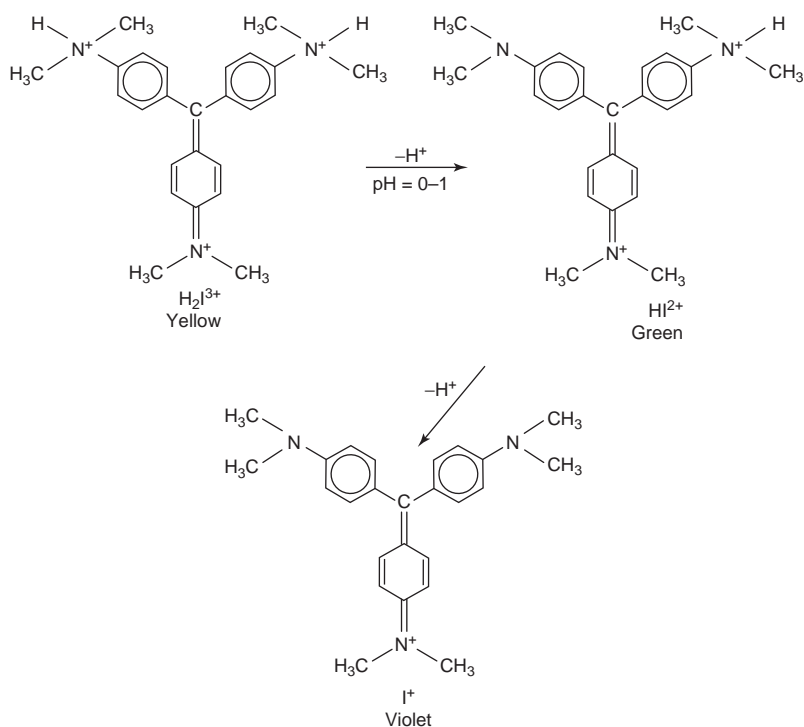
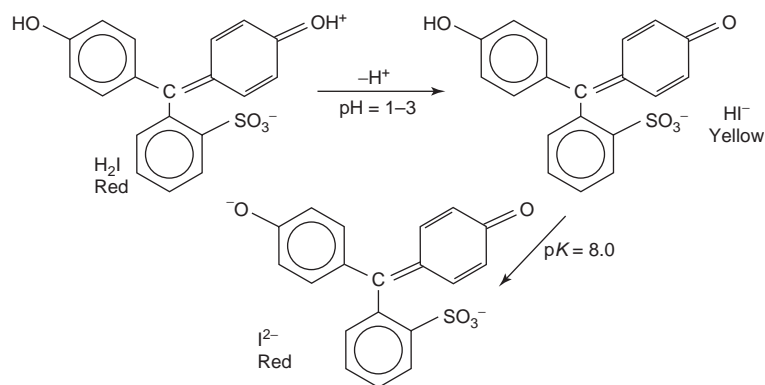
Loss of the first proton does not cause any great structural change in the molecule. With the removal of the second, however, a quinonoid structure is obtained, a structure that is often associated with color. At very high pH values phenolphthalein again changes to a colorless form. The other phthalein indicators differ only in that the phenolic rings contain additional functional groups: thymolphthalein, for example, has two alkyl groups on each ring. The basic structural alterations associated with the color changes of this indicator are similar to those of phenolphthalein.

### Sulfonophthaleins

The sulfonophthalein indicators are increasingly used in aqueous and nonaqueous solutions because

of their sharp color changes and great color intensity. Many of the sulfonophthaleins exhibit two useful color change ranges; one occurs in rather acidic solutions and the other in neutral or moderately basic media. The parent compound of this group is phenol red (phenol-sulfonophthalein). The principal equilibria of this compound are shown in reaction [XII].

Thus, phenol red is a bifunctional indicator with two transition pH ranges. However, only the second of the two color changes, occurring in the pH range between 6.4 and 8.2, is useful. Substitution of halogen or alkyl groups for the hydrogen in the phenolic rings of the parent compound yields sulfonophthaleins that differ in color and transition pH ranges. The most important are bromocresol green, bromocresol purple, bromothymol blue, cresol blue, phenol red, *m*-cresol purple, and thymol blue.





### Triphenylmethane Dyes

The most important indicators of this group are crystal violet, malachite green, and methyl violet. These compounds act like very weak bases and play only an inferior role as indicators in titrations in aqueous solution. However, these indicators are important for the endpoint indication of titrations in nonaqueous media. The equilibrium relations of crystal violet are shown in reaction [XIII].

See also: **Color Measurement.**

### Further Reading

- Barbosa J, Barrón D, and Bosch E (1987) Chromatic characteristics and applicability of neutralisation indicator series in anhydrous acetic acid. *Analyst* 112: 1717–1722.
- Barbosa J, Bosch C, and Sanz-Nebot V (1992) Effect of the solvent on the equilibria of acid–base indicators in aprotic and amphiprotic solvents. *Mikrochimica Acta* 106: 327–337.
- Barbosa J, Bosch E, Cortina JL, and Rosés M (1992) Ionic equilibria in amphiprotic solvents of low dielectric constants. Titration curves and buffer capacity of bases in anhydrous acetic acid. *Analytica Chimica Acta* 256: 211–220.
- Barbosa J, Barrón D, Beltrán JL, and Sanz-Nebot V (1995) PKPOT, a program for the study of ionic equilibria in aqueous and non-aqueous media. *Analytica Chimica Acta* 317: 75–81.
- Bishop E (1972) *Indicators*. Oxford: Pergamon.
- Escolar D, Haro MR, Saucedo A, *et al.* (1996) Determining acid equilibrium-constants by means of chromatic parameters. *Applied Spectroscopy* 50: 1290–1294.
- Fritz JS (1973) *Titrations in Non-Aqueous Solvents*. Boston: Allyn and Bacon.
- Kolthoff IM and Elving PJ (1979) *Treatise on Analytical Chemistry*, 2nd edn., part 1, vol. 2. New York: Wiley.
- Marquès I, Fonrodona G, Butí S, and Barbosa J (1999) Solvent effects on mobile phases used in LC by factor analysis applied to protonation equilibria and solvatochromic parameters. *Trends in Analytical Chemistry* 18: 472–479.
- Prasad KMMK, Raheem S, Vijayalekshmi P, and Sastri CK (1996) Basic aspects and applications of tristimulus colorimetry. *Talanta* 43: 1187–1206.
- Safarik L and Stránský Z (1986) In: Wilson CL and Wilson DW (eds.) *Comprehensive Analytical Chemistry*, vol. XXII. Amsterdam: Elsevier.
- Tucker SA, Bates HC, and Acree WE (1995) Acid–base indicators – transition colors and pH ranges determined in select aqueous-organic mixed-solvents. *Analyst* 120: 2277–2279.

## Redox

**J R Chipperfield**, Bury St. Edmunds, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The pH of an aqueous solution gives a measure of the concentration of solvated protons present. Its oxidation–reduction potential,  $E$  (redox potential), is a measure of its ability to act as an oxidizing or reducing agent. Acid–base indicators are used to give a colorimetric estimation of pH. In a similar way, redox indicators give a colorimetric estimation of the redox potential of a solution. Redox indicators are organic compounds, or metal chelate complexes, that can undergo redox reactions. The oxidized and reduced forms of the indicator have different colors. Whereas pH measures the concentration of solvated protons present but does not tell us about the intrinsic acidity of chemical species present in solution, electrode potentials inform us about the intrinsic oxidizing power of the species present but give no

information on their concentration. By the addition of a suitable amount of redox indicator to a solution, its redox potential can be measured as an alternative to using redox electrodes.

Redox indicators have been widely used to detect the endpoint of titrimetric redox analyses. Potentiometric detection of endpoints has now largely replaced the use of indicators, but redox indicators are still in use because of their simplicity. Redox indicators can be used to assess redox potentials in many redox systems where visual rather than electrical measurements can sometimes be more helpful. Recent applications of redox indicators include flow-injection analysis with colorimetric monitoring, or the measurement of electrode potentials of solutions using an immobilized redox indicator on the end of a fiber-optic probe. In studies of the metabolism of cells, redox indicators with their color or fluorescence changes are sometimes more convenient than potentiometry. Redox indicators suffer from their dependence on pH changes, and there is not yet a ‘universal redox indicator’ that can show the redox potential of a solution over a wide range of potentials

### Triphenylmethane Dyes

The most important indicators of this group are crystal violet, malachite green, and methyl violet. These compounds act like very weak bases and play only an inferior role as indicators in titrations in aqueous solution. However, these indicators are important for the endpoint indication of titrations in nonaqueous media. The equilibrium relations of crystal violet are shown in reaction [XIII].

See also: **Color Measurement.**

### Further Reading

- Barbosa J, Barrón D, and Bosch E (1987) Chromatic characteristics and applicability of neutralisation indicator series in anhydrous acetic acid. *Analyst* 112: 1717–1722.
- Barbosa J, Bosch C, and Sanz-Nebot V (1992) Effect of the solvent on the equilibria of acid–base indicators in aprotic and amphiprotic solvents. *Mikrochimica Acta* 106: 327–337.
- Barbosa J, Bosch E, Cortina JL, and Rosés M (1992) Ionic equilibria in amphiprotic solvents of low dielectric constants. Titration curves and buffer capacity of bases in anhydrous acetic acid. *Analytica Chimica Acta* 256: 211–220.
- Barbosa J, Barrón D, Beltrán JL, and Sanz-Nebot V (1995) PKPOT, a program for the study of ionic equilibria in aqueous and non-aqueous media. *Analytica Chimica Acta* 317: 75–81.
- Bishop E (1972) *Indicators*. Oxford: Pergamon.
- Escolar D, Haro MR, Saucedo A, *et al.* (1996) Determining acid equilibrium-constants by means of chromatic parameters. *Applied Spectroscopy* 50: 1290–1294.
- Fritz JS (1973) *Titrations in Non-Aqueous Solvents*. Boston: Allyn and Bacon.
- Kolthoff IM and Elving PJ (1979) *Treatise on Analytical Chemistry*, 2nd edn., part 1, vol. 2. New York: Wiley.
- Marquès I, Fonrodona G, Butí S, and Barbosa J (1999) Solvent effects on mobile phases used in LC by factor analysis applied to protonation equilibria and solvatochromic parameters. *Trends in Analytical Chemistry* 18: 472–479.
- Prasad KMMK, Raheem S, Vijayalekshmi P, and Sastri CK (1996) Basic aspects and applications of tristimulus colorimetry. *Talanta* 43: 1187–1206.
- Safarik L and Stránský Z (1986) In: Wilson CL and Wilson DW (eds.) *Comprehensive Analytical Chemistry*, vol. XXII. Amsterdam: Elsevier.
- Tucker SA, Bates HC, and Acree WE (1995) Acid–base indicators – transition colors and pH ranges determined in select aqueous-organic mixed-solvents. *Analyst* 120: 2277–2279.

## Redox

**J R Chipperfield**, Bury St. Edmunds, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The pH of an aqueous solution gives a measure of the concentration of solvated protons present. Its oxidation–reduction potential,  $E$  (redox potential), is a measure of its ability to act as an oxidizing or reducing agent. Acid–base indicators are used to give a colorimetric estimation of pH. In a similar way, redox indicators give a colorimetric estimation of the redox potential of a solution. Redox indicators are organic compounds, or metal chelate complexes, that can undergo redox reactions. The oxidized and reduced forms of the indicator have different colors. Whereas pH measures the concentration of solvated protons present but does not tell us about the intrinsic acidity of chemical species present in solution, electrode potentials inform us about the intrinsic oxidizing power of the species present but give no

information on their concentration. By the addition of a suitable amount of redox indicator to a solution, its redox potential can be measured as an alternative to using redox electrodes.

Redox indicators have been widely used to detect the endpoint of titrimetric redox analyses. Potentiometric detection of endpoints has now largely replaced the use of indicators, but redox indicators are still in use because of their simplicity. Redox indicators can be used to assess redox potentials in many redox systems where visual rather than electrical measurements can sometimes be more helpful. Recent applications of redox indicators include flow-injection analysis with colorimetric monitoring, or the measurement of electrode potentials of solutions using an immobilized redox indicator on the end of a fiber-optic probe. In studies of the metabolism of cells, redox indicators with their color or fluorescence changes are sometimes more convenient than potentiometry. Redox indicators suffer from their dependence on pH changes, and there is not yet a ‘universal redox indicator’ that can show the redox potential of a solution over a wide range of potentials

analogous to the way in which a 'universal pH indicator' covers a wide range of pH.

In this article, the physicochemical principles of redox reactions are outlined as a preliminary to an account of the function of redox indicators. Redox indicators are sensitive to the pH of a solution as well as to its redox potential, and the nature of this dependence is described. There then follows a discussion of the various types of redox indicators used, and Table 1 gives details of a range of these indicators.

## Oxidation Potentials in Solution

### Nernst Equation and Standard Hydrogen Electrode

The redox potential of a solution,  $E$ , can be written for each component undergoing a redox reaction [I]



using the Nernst equation [1]:

$$E = E^0 + \frac{RT}{nF} \ln \left( \frac{a_{\text{Ox}}}{a_{\text{Red}}} \right) \quad [1]$$

where  $a_{\text{Red}}$  and  $a_{\text{Ox}}$  are the activities of the oxidized and reduced species, respectively, and  $E^0$  is the standard electrode potential. As the activities are usually not known accurately, but concentrations of species are known, the modified Nernst equation

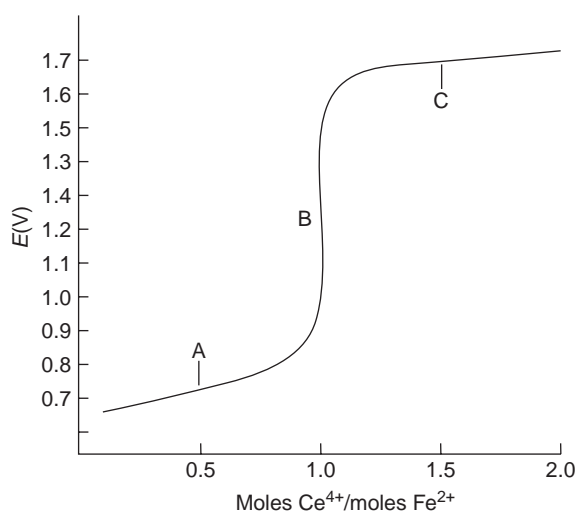
$$E = E'_0 + \frac{RT}{nF} \ln \left( \frac{[\text{Ox}]}{[\text{Red}]} \right) \quad [2]$$

is normally used, where the formal potential  $E'_0$  is the potential of a solution containing equal concentrations (not necessarily of unit molarity) of the oxidized and reduced species, under specified conditions

of other electrolytes present. In order to compare redox indicators easily, the values of formal potentials at pH 0 ( $1 \text{ mol l}^{-1}$  acid) are usually given. For accurate work, it is necessary to check which acid has been used for the measurement of  $E'_0$ . Formal potentials vary with pH and the symbol  $E'_m$  is used to denote the formal potential measured at pH  $a$ .

### Variation in Oxidation Potential as an Oxidizing Agent Reacts with a Reducing Agent

When a one-electron oxidant such as  $\text{Ce}^{4+}$  is added to a one-electron reductant such as  $\text{Fe}^{2+}$  at pH 0, the electrode potential,  $E$ , of the system changes during the addition as shown in Figure 1. When an amount of  $\text{Ce}^{4+}$  has been added equal to half the number of moles of  $\text{Fe}^{2+}$ , half the  $\text{Fe}^{2+}$  has been oxidized to



**Figure 1** Variation of electrode potential  $E$  as  $\text{Ce}^{4+}$  is added to  $\text{Fe}^{2+}$  at pH 0. For significance of A, B, and C, see text.

**Table 1** Selected redox indicators

Indicator	$E'_0$ (V)	$E'_m$ (V) <sup>a</sup>	Type	Color of oxidized form <sup>b</sup>	Color of reduced form <sup>b</sup>	Useful pH
Safranin T	0.24	−0.29	Azine	Magenta	Colorless	1–12
Phenosafranin	0.28	−0.25	Azine	Magenta	Colorless	1–11
Indigo disulfonic acid (indigo carmine)	0.29	−0.12	Indigo	Blue	Colorless	<9
Indigo tetrasulfonic acid	0.36	−0.05	Indigo	Blue	Colorless	<9
Methylene blue	0.53	0.01	Thiazine	Blue	Colorless	1–13
Brilliant cresyl blue	0.58	0.05	Phenoxazine	Blue	Colorless	0–11
2,6-Dichloro-indophenol	0.67	0.22	Indophenol	Blue	Colorless	6–11
Bindshedler's green	0.68	0.22	Indamine	Green	Orange	2–9.5
Phenol blue	0.68	0.22	Aminoindophenol	Blue	Colorless	1–12.5
Variamine blue	0.71	0.31	Variamine	Blue	Colorless	1.5–6.3
Diphenylamine sulfonic acid	0.85	–	Diphenylamine	Red-violet	Colorless	<1
Tris(bipyridyl)iron	1.03	1.10	Chelate	Violet	Green	<7
Ferriin	1.11	1.12	Chelate	Blue	Red	<7

<sup>a</sup>Where measured.

<sup>b</sup>Colors often depend on pH of solutions.

$\text{Fe}^{3+}$ , and the potential (A in Figure 1) is the  $E'_0$  of  $\text{Fe}^{3+}/\text{Fe}^{2+}$ . When all  $\text{Fe}^{2+}$  has been oxidized to  $\text{Fe}^{3+}$ , the potential changes rapidly (B in Figure 1). If a further amount of  $\text{Ce}^{4+}$  equal to half the number of moles of  $\text{Fe}^{2+}$  is added, the potential reaches the  $E'_0$  of  $\text{Ce}^{4+}/\text{Ce}^{3+}$  (C in Figure 1).

### What a Redox Indicator is Required to Do

For titrations of reductants with oxidants, a redox indicator is required to indicate when the potential of the solution has reached that at equivalence (B in Figure 1). For an observer to see that an indicator has fully changed color, it is generally taken that the ratio  $[\text{Ind}_{\text{red}}]/[\text{Ind}_{\text{ox}}]$  needs to change from 1:10 to 10:1. Application of the Nernst equation shows that this involves a change of potential of  $\sim 0.12/n_{\text{ind}}$  V, that is 120 mV for a one-electron indicator or 60 mV for a two-electron indicator at room temperature.

### Complications Arising out of pH Changes

If the redox equation for a reaction involves hydrogen ions, the potential of the solution will depend on pH as well as on the concentrations of the oxidized and reduced species.

For instance, for the reduction of permanganate (reaction [II])



the redox potential depends on the eighth power of the hydrogen ion concentration. The Nernst equation is then

$$E = E'_0 + \frac{RT}{nF} \ln \left( \frac{[\text{MnO}_4^-]}{[\text{Mn}^{2+}][\text{H}^+]^8} \right) \quad [3]$$

Changes in pH may also change the speciation of the reductants or oxidants as protonation changes.

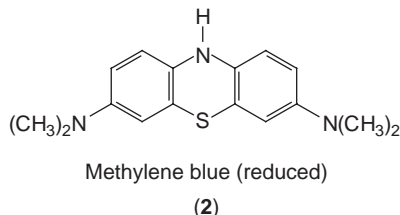
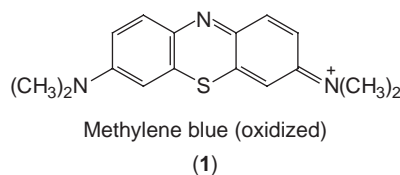
### Basic Features of Redox Indicators

A redox indicator is a substance that undergoes redox reactions as in reaction [III], where the oxidized and reduced forms have different colors:



#### How Reduced and Oxidized Forms Have Different Colors

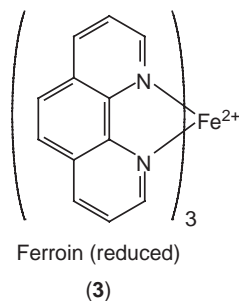
The chemical structures of the oxidized and reduced forms of an indicator are different. Many organic redox indicators contain aromatic rings, and there are large differences in ring conjugation between the oxidized and reduced forms, for instance, oxidized (blue color) and reduced (colorless) methylene blue ((1) and (2)).



Metal chelate redox indicators have the metal in different oxidation states in the two forms, and the colors consequently reflect the different numbers of d electrons involved in the metal–ligand interactions.

### Simplest Example – No Dependence on pH

The indicator tris(1,10-phenanthroline)iron(II), sometimes called ferroin (3), has an  $E'_0$  value of 1.06 V.



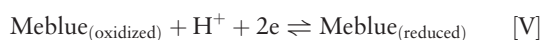
The redox reaction [IV]



does not involve hydrogen ions, and the formal potential,  $E_m$ , of the indicator is largely independent of pH from neutral solution up to  $5 \text{ mol l}^{-1} \text{ H}_2\text{SO}_4$ . The color change between reduced and oxidized forms will take place at an electrode potential largely independent of pH.

### Dependence on pH When the Redox Equilibrium Contains an $\text{H}^+$ Term

For the majority of redox indicators, the redox reaction involves hydrogen ions, and consequently the potential when  $[\text{Ind}_{\text{red}}]/[\text{Ind}_{\text{ox}}]$  is unity varies with pH. Thus for the indicator methylene blue, the redox reaction may be written as



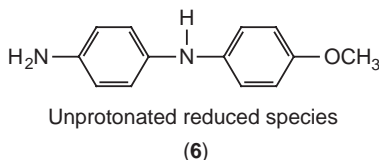
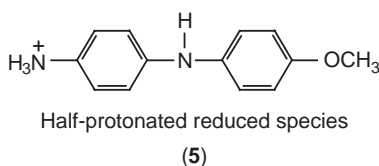
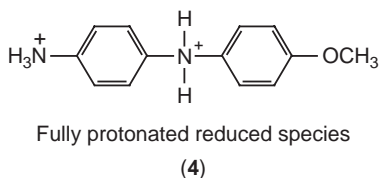
and the Nernst equation is given by

$$E = E'_0 + \frac{RT}{2F} \ln \frac{[\text{Meblue}_{\text{oxidized}}][\text{H}^+]}{[\text{Meblue}_{\text{reduced}}]} \quad [4]$$

Hydrogen ions will alter the potential at which  $[\text{Meblue}_{\text{(oxidized)}}] = [\text{Meblue}_{\text{(reduced)}}]$ , and so the indicator will need to be used under defined conditions of acid strength if it is being used to measure electrode potentials.

### Further pH Complications as Oxidized and Reduced Forms May Each Involve Protonation or Deprotonation of Suitable Groups

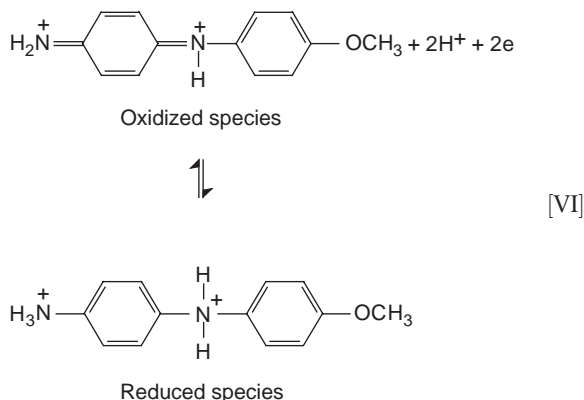
Most organic redox indicators contain groups that can ionize (e.g., OH, NH) or can be protonated (e.g., amine N, ketonic O). As pH is changed, the speciation of the oxidized and reduced form of the indicator can change, depending upon the  $\text{p}K_{\text{a}}$  values of these acid/base groups. Thus for variamine blue, there are two nitrogen atoms that can be protonated or not. For the reduced indicator, any of (4–6) can be present.



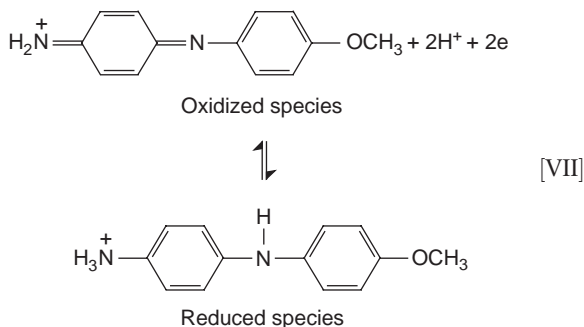
For the oxidized form, an analogous three species are possible. Their relative concentrations are determined by the pH of the medium and the  $\text{p}K_{\text{a}}$  values of the protonations.

**Combination of pH-related effects** The combination of pH effects described above results in a complex curve of  $E_{\text{m}}$  against pH. In  $1 \text{ mol l}^{-1}$  acid, the reduced and oxidized forms of the indicator both contain protonated nitrogen atoms. The redox equation for the

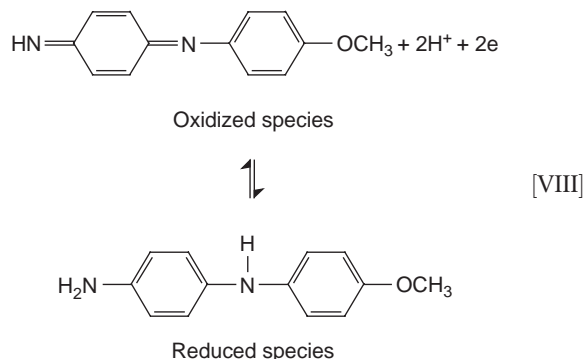
indicator in acidic medium may be written as



If no other effects were present, the presence of the hydrogen ions in the redox equation would change the formal redox potential,  $E_{\text{m}}$ , by 60 mV for each pH unit increase. However, as pH is raised, the secondary amine nitrogen deprotonates in both the oxidized and reduced forms, until at about pH 9 both oxidized and reduced forms have lost one proton, and the redox equilibrium is



Further basification causes the primary amine nitrogen to deprotonate and the redox equilibrium is



If the  $\text{p}K_{\text{a}}$  values for the secondary amine protonation are  $\text{p}K_{\text{red1}}$  and  $\text{p}K_{\text{ox1}}$  respectively for the reduced and oxidized forms, and the  $\text{p}K_{\text{a}}$  values for the



primary amine protonation are  $pK_{\text{red2}}$  and  $pK_{\text{ox2}}$  respectively, it can be shown that the electrode potential of a variamine blue solution is related to the formal potential,  $E'_0$ , by

$$E_m = E'_0 + \frac{RT}{2F} \ln \frac{[\text{H}^+]^4 + K_{\text{red1}}[\text{H}^+]^3 + K_{\text{red1}}K_{\text{red2}}[\text{H}^+]^2}{[\text{H}^+]^2 + K_{\text{ox1}}[\text{H}^+] + K_{\text{ox1}}K_{\text{ox2}}} \quad [5]$$

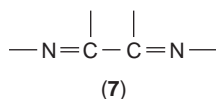
Complex expressions of this type can be deduced for all redox indicators that have ionizable protons. In practice, it is often easier to measure how  $E_m$  varies with pH than to calculate it.

**Reversibility of redox indicators** For a reversible redox indicator, the addition of a suitable oxidant oxidizes its reduced form to its oxidized form. Then the addition of a suitable reductant will reduce it back again. Many organic compounds undergo an irreversible color change on oxidation. The addition of reductant does not give the original reduced indicator. Such compounds have been used as redox indicators in titrimetry, but definition of their formal potential is impossible. However, the potential when they are oxidized can be determined roughly. Such indicators may prove useful in titrations if no reversible indicator with suitable  $E_m$  or suitable color change can be found.

## Classes of Redox Indicators

### Metal Chelates

Transition metal chelate complexes of ligands containing the  $\alpha,\alpha'$ -diimine grouping (7) are numerous, and those of iron, ruthenium, and osmium have successfully been used as redox indicators. The tris(1,10-phenanthroline)iron complex (ferroin) is the most familiar. The iron(II) derivative is orange-red ( $\epsilon = 11\,100 \text{ mol}^{-1} \text{ l cm}^{-1}$  at 510 nm) and the iron(III) derivative is pale blue ( $\epsilon = 900 \text{ mol}^{-1} \text{ l cm}^{-1}$  at 590 nm), so a large color change takes place when the indicator undergoes a redox reaction. The formal potential,  $E'_0$ , is 1.06 V in  $1 \text{ mol l}^{-1} \text{ H}_2\text{SO}_4$ . Changes in substituent on the phenanthroline rings alter  $E'_0$  a little. For example, the 5-nitro derivative has  $E'_0$  at 1.25 V whereas the 3,4,6,7-tetramethyl derivative has  $E'_0$  at 0.87 V. Potentials of ruthenium analogs are  $\sim 0.1 \text{ V}$  higher, and those of osmium analogs are 0.2 V higher. The ruthenium derivatives are fluorescent (see later).

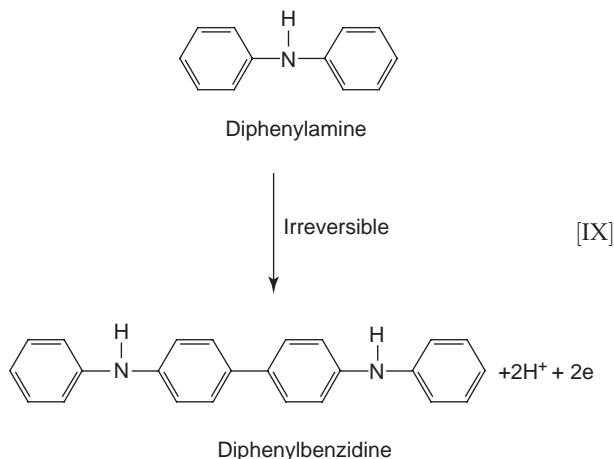


Bipyridyl complexes are less resistant to hydrolysis in acidic solution, and have lower  $E'_0$  values. For

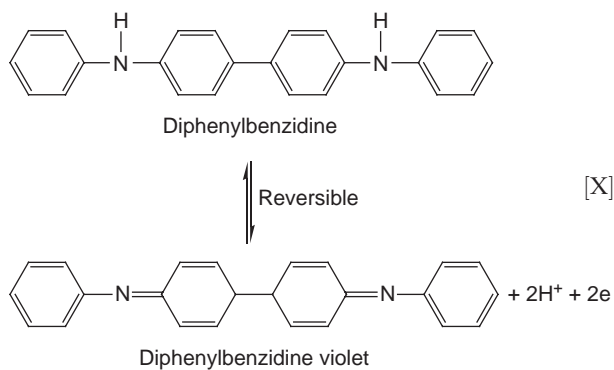
example, the tris(2,2'-bipyridyl) complex of iron has an  $E'_0$  of 1.06 V.

### Diphenylamine and Its Derivatives

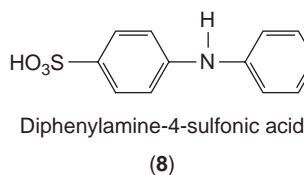
Diphenylamine and its derivatives have been successfully used as redox indicators for many years. The precise detail of how these indicators work is still under discussion. The conventional explanation is that colorless diphenylamine is irreversibly oxidized to colorless diphenylbenzidine in a two-electron reaction:



Diphenylbenzidine is then able to act as a two-electron reversible redox indicator to give purple-violet diphenylbenzidine violet:



Unfortunately diphenylamine is not very soluble in water, so more soluble derivatives are used such as the popular 4-sulfonic acid derivative (8).



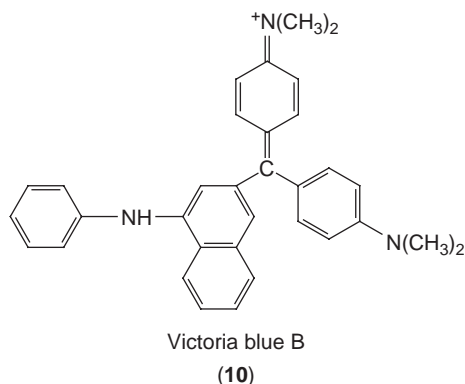
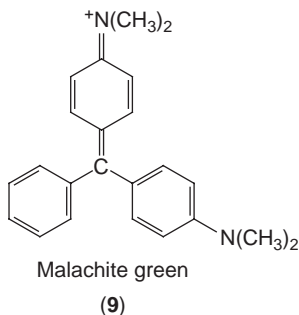
The formal potentials of diphenylamine derivatives are in the range 0.7–1.1 V.

### Triphenylmethane Dyes

These are one-electron redox indicators that are based on substituted triphenylmethane,  $\text{Ar}_3\text{CH}$ . The basic redox reaction is



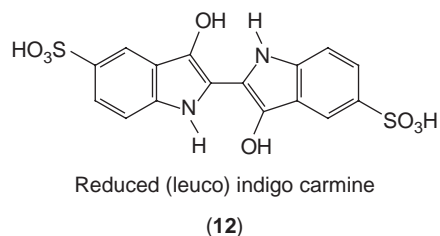
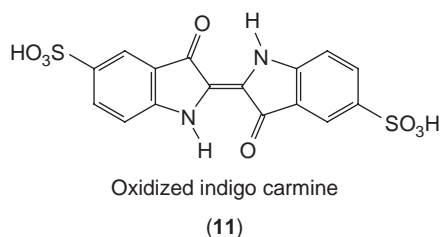
The oxidized forms,  $\text{Ar}_3\text{C}^\bullet$ , are radicals and their colors arise from electron delocalization over the three aromatic rings. The aromatic rings contain at least one (and often two or more) *para*-amino group or *para*-hydroxy group, which upon oxidation become double-bonded imine or ketone functions. Malachite green (9) and Victoria blue B (10) are typical examples.  $E'_0$  values for many triphenylmethyl dyes are  $\sim 1$  V in acidic solution. In their reduced forms, most are blue or purple in neutral solution, changing to green or yellow on acidification, and give intense orange-red colors on oxidation.



### Indigo Dyes

Indigo is the basis of intense dyestuffs, and can be made water soluble by the incorporation of sulfonic acid substituents. Indigo sulfonic acids are two-electron reversible redox indicators, with formal

potentials  $E'_0$  ranging from 0.26 to 0.41 V. The reduced forms are pale yellow or colorless, while the oxidized forms range from deep blue to reddish blue depending upon the number of sulfonic acid groups. With their low formal potentials, they have found most use in acidic conditions with strong reducing agents such as tin(II), titanium(III), and chromium(II). The most widely used is indigo-5,5'-disulfonic acid under the name indigo carmine ((11) and (12)).



### Indophenols

These are two-electron redox indicators with  $E'_0$  at 0.54–0.69 V. Quoted  $E'_0$  values are generally extrapolations, as these compounds decompose in strong acid. They are especially useful in alkaline solution.

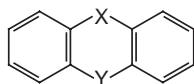
### Azo Dyes

Many diazo dyes have been examined as redox indicators. Although distinct color changes are observed for many on oxidation, the changes are nonreversible, and, in general, these compounds have no advantages over other classes of reversible indicators.

### Azines, Oxazines, and Thiazines

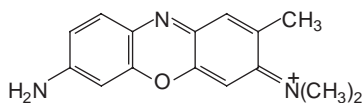
These are two-electron redox indicators based on phenazine, phenoxazine, or phenthiazine (13) structures. Their  $E'_0$  values range from 0.05 to 0.58 V. This means that they are principally used in the presence of such strong reducing agents as titanium(III), chromium(II), or vanadium(II). Typical examples of these compounds, which have been used as redox

indicators, are brilliant cresyl blue, methylene blue, and brilliant alizarin blue ((14)–(16)).



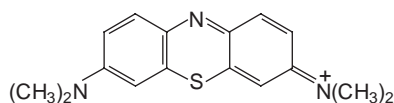
Phenazine X = Y = NH  
 Phenoxazine X = O, Y = NH  
 Phenothiazine X = S, Y = NH

(13)



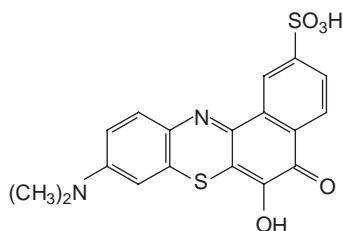
Brilliant cresyl blue

(14)



Methylene blue

(15)

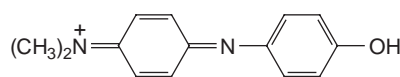


Brilliant alizarin blue

(16)

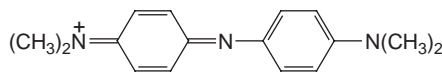
### Indamines and Indoanilines

These compounds have been used as redox indicators, but have a limited range of pH stability. The indoaniline phenol blue (17) has an  $E'_0$  value of 0.68 V, but decomposes in both acidic and alkaline solution, and so can only be used safely in solutions of pH 3–9. The related indamine, Bindschedler's green (18), with an  $E'_0$  of 0.68 V can be used over a pH range wider than phenol blue and related compounds, but decomposes in strong acid or alkali.



Phenol blue

(17)

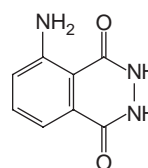


Bindschedler's green

(18)

### Chemiluminescent Redox Indicators

Chemiluminescent reactions usually involve oxidation of an organic molecule. The chemiluminescent oxidation of luminol (19) is well known.



Luminol

(19)

This chemiluminescence can be used to detect the presence of excess oxidant after a reductant has been oxidized. Chemiluminescent light levels are low, and darkened ambient conditions are needed to detect the chemiluminescence.

### Fluorescent Redox Indicators

These are used in a similar way to visual indicators but, instead of color change being detected, one form of the indicator is fluorescent, and its presence can be detected by a suitable fluoroscope. Useful fluorescent indicators are the ruthenium analogs of ferroin, and tris(2,2'-bipyridyl)iron. In these, the reduced ruthenium(II) complexes show an orange-red fluorescence that disappears on oxidation to ruthenium(III).

See also: **Chemiluminescence**: Overview; Liquid-Phase. **Indicators**: Acid–Base. **Titrimetry**: Overview.

### Further Reading

Bishop E (1972) *Indicators*. Oxford: Pergamon Press.  
*Dictionary of Analytical Reagents* (1993) London: Chapman and Hall.

## Complexometric, Adsorption, and Luminescence Indicators

**K Krishnankutty**, University of Calicut, Malappuram, Kerala, India

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Although numerous highly efficient instrumental methods are now available for the determination of metal ions and other inorganic species, complexometric titrations are still widely used in routine analysis. The versatility, sensitivity, and general convenience of complexometric titrations are dependent on the correct choice of indicators for endpoint detection. The endpoint detection in complexometric titration was revolutionized by Schwarzenbach in the 1940s with the discovery of visual color indicators sensitive to the concentration of metal ions in solution. A wide variety of such indicators, often called complexometric or metal ion indicators, are available for the selective determination of almost all metal and metalloid ions from a variety of matrices. In this article, the basic mode of action of these visual indicators, their classification, and relevant details on the use of typical indicators are discussed.

Compared to complexometric indicators, the use of adsorption and luminescence indicators is limited. However, they are important in many specified areas of chemical analysis. The essential theory of action of these two types of indicator systems in titrimetric analysis and important examples of practical value are also included in this article.

### Complexometric Indicators

Visual methods that can be used to indicate the endpoint in complexometric titrations almost always involve complex formation between the indicator and the metal ion whose pM ( $-\log_{10}[\text{metal ion}]$ ) changes sharply at the endpoint. In complexometric titrations in which a complexing agent (e.g., EDTA) is added to a solution of a metal ion, the pM changes slowly until at the equivalence point, when virtually all the metal ions are complexed, a sudden and rapid increase in pM occurs. Metal indicators detect the disappearance of the last traces of free metal ion by forming a colored complex with the metal ion, and showing a different color or no color at all when all the particular metal ions have been complexed by the titrant. That is, the color of the metal-indicator complex and the free indicator should differ significantly.

Numerous compounds are known to form colored products with metal ions, but only a few of these are

suitable as visual indicators. The essential requirements of such metal indicators can be summarized as follows: (1) The color must be sufficiently intense that a minimal amount of the indicator is needed. (2) The color contrast between the free indicator and the metal-indicator complex must be sufficient to permit easy visual observation. (3) The metal-indicator complex must be weaker than the metal-titrant complex. (4) The formation of the metal-indicator complex and its reaction with the titrant must be sufficiently rapid and preferably reversible in order to ensure an abrupt color change at the equivalence point. (5) The metal-indicator complex must have an appropriate stability; extended dissociation will cause the color change interval to be spread over several drops of titrant. (6) To ensure a sharp endpoint, the metal-to-indicator ratio in the complex should preferably be 1:1. (7) The color reaction between the metal and the indicator should be selective, with as little interference as possible from other metal ions. (8) The indicator must be stable in the titration medium.

### Mode of Action

Organic compounds, usually dyestuffs, with acid-base properties and sensitive to low concentrations of metal ions, are the most important complexometric indicators (also termed metallochromic indicators). In most cases the indicator and the metal-indicator complex are both colored, so that the endpoint of the titration is characterized by a change in color. Like most complexing titrants, these indicators are also capable of forming chelate rings with the metal ion. They are also usually polybasic acids that change color with the pH of the solution. Suitable buffer mixtures are then necessary during the titration. For indicators of high molar absorptivity ( $10^4$ – $10^5 \text{ mol}^{-1} \text{ l cm}^{-1}$ ), indicator concentrations of  $10^{-6}$ – $10^{-5} \text{ mol l}^{-1}$  are sufficient to give a clearly detectable color change, and therefore the indicator error is generally negligible. On the other hand, there are colorless or weakly colored indicators that form colored metal-indicator complexes. Because of the relatively low molar absorptivities ( $\sim 10^3 \text{ mol}^{-1} \text{ l cm}^{-1}$ ) of these complexes, indicator concentrations of  $10^{-5}$ – $10^{-4} \text{ mol l}^{-1}$  are necessary to produce a detectable color.

The use of a metal (M) indicator in complexometry may be represented in the following reactions:



The indicator  $H_nI$  and the titrant  $H_nY$  are considered to be in fully dissociated anionic forms  $I$  and  $Y$  (charges are omitted for convenience). The reaction will proceed if the metal-indicator complex  $MI$  is sufficiently less stable than the metal-titrant complex  $MY$ . Here it is assumed that the indicator acid forms only a 1:1 metal complex  $MI$ , and that it is not involved in any side reaction. The stability constant  $K_{MI}$  is

$$K_{MI} = \left( \frac{[MI]}{[M][I]} \right) \quad [1]$$

$MI$  and  $I$  should have different colors. The color change will occur when  $[MI] \approx [I]$ . The transition point is expressed as  $pM_{trans}$  from eqn [1]:

$$pM_{trans} = -\log[M_{trans}] = \log K_{MI} \quad [2]$$

$pM_{trans}$  is a measure of the sensitivity of the indicator, a large value indicating high sensitivity.

The above reactions were obtained by ignoring all the possible side reactions between the indicator and various species present in the medium. In acidic solution the indicator usually occurs as one or more protonated forms  $HI$ ,  $H_2I$ , ...,  $H_nI$ , and only a very small fraction exists as the ion  $I$ . Thus, instead of  $[I]$  in eqn [1], it may be more appropriate to use  $[I']$ , where

$$[I'] = [I] + [HI] + \dots + [H_nI] \quad [3]$$

and the color change occurs when  $[MI] = [I']$ . The relationship between  $[I]$  and  $[I']$  is denoted by the coefficient  $\alpha$ :

$$\alpha_{HI} = \frac{[I']}{[I]} \quad [4]$$

By combining eqns [1] and [4],

$$\frac{[MI]}{[M][I']} = \frac{K_{MI}}{\alpha_{HI}} = K_{MI'} \quad [5]$$

$K_{MI'}$  is the apparent, conditional, or effective stability constant. It is not a real constant since it depends on experimental conditions (here pH). From eqn [5],

$$pM_{trans} = \log K_{MI} - \log \alpha_{HI} = \log K_{MI'} \quad [6]$$

The effect of side reactions of the metal ion on the color change can similarly be estimated by introducing  $[M']$  and  $\alpha_M$ . The corresponding expressions are

$$K_{M'I'} = \frac{[MI]}{[M'][I']} = \frac{K_{MI}}{\alpha_M \alpha_1} \quad [7]$$

$$pM'_{trans} = \log K_{M'I'} + \log \left( \frac{[I']}{[MI]} \right) \quad [8]$$

Since  $[I'] = [MI]$ , at the transition point

$$pM'_{trans} = \log K_{M'I'} \quad [9]$$

Therefore, the sensitivity of a metal indicator expressed as  $pM'_{trans}$  is equal to the logarithm of the conditional stability constant of the metal-indicator complex. Although the formation of  $MI_2$ ,  $MI_3$ , ...,  $MHI$ ,  $MOHI$ , ...,  $M_2I$  or  $M_nI$  type species can also influence the color change of an indicator, their effect is seldom very great. An indicator error may arise owing to the residual binding of some of the metal ions to the indicator at the endpoint. For accurate results, a correction has to be applied.

As the values of the dissociation constants of a large number of indicators are known, the  $pM$  values at which a color change occurs at various pH values, and the values of  $pM'_{trans}$  ( $= pM_{trans} - \log \alpha_M$ ) can be determined. The  $\alpha_M$  values depend on the nature and concentrations of buffer, masking agents, etc. present in the medium. In this way, it is possible to compare the sensitivities of the various indicators under various experimental conditions and to choose the most suitable indicator for a particular titration. There are certain factors that are detrimental to the effective functioning of an otherwise good indicator, such as impurities present in the indicator and the presence of other, strongly binding metal ions that can block the color change.

### Classification of Metallochromic Indicators

In general, these compounds contain at least two auxochromic groups bound to the conjugated chromophoric system. In addition, metal-binding groups such as  $OH$ ,  $NH_2$ ,  $COOH$ , etc., at suitable positions are also necessary to form stable five- or six-membered chelate rings with the metal ion. The number of substances tried and recommended as metallochromic indicators is large and even a specialist has difficulty in proper selection. Many of the recently reported indicators are of no great practical value. Therefore, in Table 1, only selected typical and readily available indicators are given. The indicators can be classified according to their structural types as (a) diarylazo dyes including compounds containing heterocyclic rings (1–26 in Table 1), (b) triphenylmethane derivatives (27–37 in Table 1), (c) anthracene derivatives (38–40 in Table 1), and (d) miscellaneous structural types.

From an analytical point of view, the following information is provided in the table. (1) The trivial name rather than the polysyllabic systematic nomenclature (chemical reagent manufacturers' catalogs contain trivial names also). As many of the compounds are commercial dyestuffs, a single dye may be



**Table 1** Common complexometric indicators

No.	Trivial names/synonyms (CI number) <sup>a</sup> ; indicator preparation <sup>b</sup>	Metal ion titrated (pH, buffer <sup>c</sup> ); color change at endpoint <sup>d</sup> (charges omitted for simplicity)
1	Acid alizarin black SE, anthrol fast black SE; NaCl (2.0)	Ca (11–14, NaOH) R/B; Mn <sup>2+</sup> (11.5, A-NH <sub>2</sub> OH) P/B
2	Acid alizarin black SN, anthrol fast black SM (21725); W (0.1)	Ba, Ca (11.5, A) R/B; Cd (8.5, Bx) R/B; Mn <sup>2+</sup> , Ni (10, A) P/B; Th (4, Ac) R/O; Zn (11.5, A) P/B
3	Acid chrome blue K (16575); E (0.1)	Ca (12, A) R/B; Cd, Mg, Pb, Zn (10, A) R-V/B
4	Arsenazo I, neothorin, uranon; W (0.1–0.3)	Ca, Mg (10, A) V/R-O; Pu <sup>4+</sup> (0.1–2 mol l <sup>-1</sup> HCl) B-V/P; Ln, Y (5.5–6.5, Py) V/R-O; Th <sup>4+</sup> , U <sup>4+</sup> (1.7–3.0) V/O
5	Calgamite	Ca, Mg (10, A) R/B
6	Chrome Bordeaux B (14210); W (1.0)	Mg, Zn (10, A) Y/V
7	Chromotrope 2R (16570); W (0.2)	Th (~2, HNO <sub>3</sub> , 40°C) V/Pk
8	Eriochrome black PV (16500); W (10 <sup>-3</sup> mol l <sup>-1</sup> )	Ca (11.5, A) R/B; Cd, Pb, Zn (10, A) R/B; Fe <sup>2+</sup> (4, Ac-hot) O/R; Fe <sup>3+</sup> (4, Ac-hot) Y/R; Mg, Mn <sup>2+</sup> (10, A) R/B
9	Eriochrome blue black B (14640); E or W (0.5)	Ca, Cd (11.5, A); Mg, Zn (10, A) W-R/B; Mn (10–11.5, A-as), U <sup>4+</sup> (1–2, HCl, hot), Zr (0.1–0.5 mol l <sup>-1</sup> HCl, hot) B/R
10	Eriochrome blue black R (15705); M or E (0.5) used also as Zn salt, zinchrome R	Ca, Cd (11.5, A) Pk/B; Mg, Zn (10, A) Pk/B; Mn (10, A-as) R/B
11	Eriochrome black T (14645); E or NaCl (0.5)	Cd, Zn (6.8–11.5, A), In (8–10, A-tr, hot), Mg (10, A), Mn (8–10, A-as), Pb (10, A-tr), Ln (8–9, A-tr) R/B; Zr (0.5–2 mol l <sup>-1</sup> HCl, 100°C) B-V/Pk
12	Eriochrome fast gray RRS, solochrome fast gray RA (15690)	Bi (2–2.5, HNO <sub>3</sub> ) V/O; Fe <sup>3+</sup> (HNO <sub>3</sub> ) Y/O; Th (2.5, HNO <sub>3</sub> ) Rs/O; Zn (8, A) St/V
13	Eriochrome red PE, solochrome red B, acid alizarin red B; W (0.1)	Ca (13) Y/O-R
14	Eriochrome violet B, solochrome violet RS (15670); W (0.01)	Ca, Sr (12.5–13, NaOH) R-O/R-V; Mg (10, A-as) Pk/p
15	Fast navy 2R, eriochrome violet 5B (14855)	Ca, Mg (8.5, A) W-R/B; Cd, Mn <sup>2+</sup> , Ni, Pb, Zn (8.35, NH <sub>4</sub> Ac) W-R/B
16	Fast sulfon black F (26990); W (0.5)	Cu (10, A) P/Y or G
17	Gallion	Ga (2, hot) B/Pk
18	Magon; M (0.01)	Mg (10–10.5, Bx) O/B-V
19	Metaomega chrome blue 2 RL, magneson (mordant blue 12)	Ca, Cd (11.5, A) R/B; Ni (4, Ac, hot) R/O; Sr (12.5, HNEt <sub>2</sub> ) R/B
20	Naphthol violet; NaCl (1.0)	Bi (1–3, HNO <sub>3</sub> ) R-V/R-O; Cd, Co, Cu, Mg, Zn (10–11, A) R-V/B; Mn <sup>2+</sup> (10–11, A-NH <sub>2</sub> OH) R-V/B
21	Omega chrome garnet, mordant red no. 72; E (0.1)	Ni (4, Ac) P/Y
22	Ponceau 3R (17110); W (0.1)	Cu (9, A) G/R or Y/R
23	Pyridylazoresorcinol, PAR; W (0.05–0.2)	Bi (1–2, HNO <sub>3</sub> ); Cd (8–11, A, Ur); Cu (5–11, A, Ur), Hg (3–6, Ur); In (2.5, Ac), Mn (9, A-as), Ni (5, Ac), Pb (5–9, A, Ur), Ln (6, Ur), Th (2.3–2.8), Tl (1.7), Zn (6–11, A, Ur) R/Y
24	Pyridylazonaphthol, PAN; E (0.01–0.1)	Bi (1–3) R/Y; Cd (6, A) R/Y; Cu (2–5, Ac) R/Y, (10, A) V/Y; In (2–5 Ac) R/Y; Ni (4, Ac) Pk/Y; Th (2.3–5, HNO <sub>3</sub> ) R/Y; UO <sub>2</sub> (4.4, Ur) R/Y; Zn (5–7, Ac) Pk/Y
25	Sulfonaphthylazoresorcinol; W (0.01)	Zr (0.8–2.5; HCl) B-V/R-S
26	Thorin, thoron, thoronol, APANS; W (0.5)	Bi (2–3, HNO <sub>3</sub> ) R/Y; Th (1–3, HNO <sub>3</sub> ) V/Y; U <sup>4+</sup> (1–1.8, HClO <sub>4</sub> ) Rs/O-Y; Sc (4.5–6.5), Pk/Y; Y <sup>3+</sup> (6), Rs/Y
27	Aluminon, chrome violet (43810)	Al (4.4, Ac, hot) R/B-V; Ca, Mg (8.5–9.9, A) R-V/Y; Fe <sup>3+</sup> (1–2, Acetone) V/C
28	Bromopyrogallol red	Bi (2–3, HNO <sub>3</sub> ) W-R/Y; Cd, Ni (9.3, A) B/R; Mg, Mn (10, A), B/V; Pb (5–6, Ac) B-V/R; Ln (7, NH <sub>4</sub> Ac) B/R
29	Chromazurol S (43825)	Al (4, Ac) V/Y-O; Ca, Mg (10–11, A) R/Y; Cu (6–6.5, Ac) B-V/Y; Fe <sup>3+</sup> (2–3, Chl. Ac) B/O; Ni (8–11, A) B/Y; Ln (8, A) V/Y; Th (2–3, HNO <sub>3</sub> ) R-V/O; Th (4–8, A) V/O; V (4, Ac) B-V/O
30	Eriochrome cyanine R (43820)	Al (5–6, Ac) P/Y; Ca, Cu, Mg (10–11.5, A) V/Y; Fe <sup>3+</sup> (2.3, Chl. Ac) V/O; Th (2–2.5, HNO <sub>3</sub> ) P/Pk; Zr (1.3–1.5, HCl) Pk/C
31	Glycine thymol blue	Cu (5–6, Ur) B/Y (G)
32	Metalphthalein	Ba, Sr (10.5–11, A) R/Rs, Ca, Mg (10–11, A) R/Pk, Cd (10, A) Pk/C
33	Methyl thymol blue	Ba, Mg (10–11, A) B/Gy; Bi (1–3, HNO <sub>3</sub> ) B/Y; Ca (12, NaOH) B/Gy; Cd (5–6, B-V/Y; Cu (11.5, A) B/C; Fe <sup>2+</sup> (4.5–6, Ur) B/C; Hg (6, Ur), In (3–4, Ac), Mn (6–6.5, Ur), Pb, Ln (6, Ur), Sc (2.2, HNO <sub>3</sub> ), Sn <sup>2+</sup> (5–5.6, Py + Ac), Th (1–3.5, HNO <sub>3</sub> ), Zn (6–6.5, Ur), Zr (0–2.3, Chl. Ac) B/Y; Tl <sup>3+</sup> (7–10, tr) R/B
34	Pyrocatechol violet	Bi (2–3, HNO <sub>3</sub> ) B/Y; Cd, Mg, Zn (10, A) B/R-P; Cu (5–6.3, Ac; 6–7, Py, 9.3, A) B/Y or P; Fe <sup>2+</sup> (3.6, A) B/Y; Ga (3.8, Ac) B/Y; In (5–6, Ac) B/Y; Mn (9.3, A) B/R-P; Ni (8–9, A) B/R-V; Pb (5.5, Ur) B/Y; Th (10, A) B/R-V
35	Pyrogallol red	Bi (2–3, HNO <sub>3</sub> ) R/O-Y; Co, Ni (9.3, A) B/R; Pb (5–6) V/R

Table 1 Continued

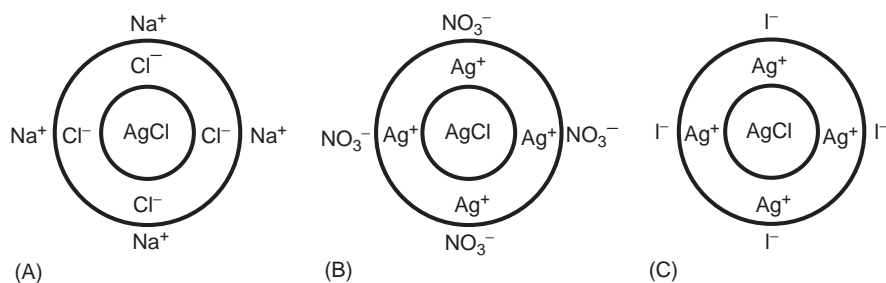
No.	Trivial names/synonyms (CI number) <sup>a</sup> ; indicator preparation <sup>b</sup>	Metal ion titrated (pH, buffer <sup>c</sup> ); color change at endpoint <sup>d</sup> (charges omitted for simplicity)
36	Thymolphthalexone	Ag (10–11, A) B/G; Ba, Sr (10–11, A/NaOH) B/G-Y; Ca (10.5–12, A/NaOH) B/C; Mn (10, A) B/Pk
37	Xylenol orange	Bi (1–3, HNO <sub>3</sub> ) R/Y; Cd (5.6, Ur) Pk/Y; Co (5–6, Ur) R-V/Y; Cu (4–6, Ur or Ac) R-V/Y; Fe <sup>3+</sup> (1–1.5, HNO <sub>3</sub> ) B-V/Y; Hg (5–6, Ur) P/Y; In (3.4–5, Ac) R-V/Y; Mn (10.5, A-as) V/Gy; Pb, Ln (4–6, HNO <sub>3</sub> ) R/Y; Sc (2.5, HNO <sub>3</sub> ) Pk/Y, Ti <sup>3+</sup> (4.5, Ac) R/Y; U <sup>4+</sup> (1.7–2.2) R/Y; V <sup>5+</sup> (1.8) R/Y; Y (4.5–6, Ur) R/Y; Zn (5–6, Ac) R/Y; Zr (1 mol l <sup>-1</sup> HCl hot) R/Y
38	Alizarin red S, alizarin S (58005); W (0.05–0.2)	Ln (4–4.5, Ac, hot) R/Y; Sc (2) R/G; Th (1.5–3.8, HNO <sub>3</sub> or Ac) Pk/Y; Y (5) Pk/Y
39	Alizarin complexan; W (0.5)	Ba, Ca, Cd, Sr (10, A) B/R; Cu, In, Pb, Zn (4.3, Ac) R/Y
40	Carminic acid (75470); W (0.3)	Ln, Ti (3.7, hot) V/Y; Zn (2N HCl) B-V/Pk
41	Diphenylcarbazide	Hg <sup>2+</sup> (5–6, Ur) V/C; VO <sup>2+</sup> (4–6, Ac) R/C
42	Diphenylcarbazone	Hg (1, HCl, 5–6, Ur) B-V/C; Pb (4.5–6.3, Ac) R/C; V <sup>5+</sup> (4.5–5.5, Ac) V/C
43	Dithizone	Bi (2.5–5, 20% Py or 60% E) R/G; Cd, In, Zn (4.5, 20% Py) R/Y-G
44	Murexide	Ca (> 10, NaOH) Pk/V; Co (8–10, A) Y/V; Cu (4, Ac) O/R, (7–8, A) Y/V; Mn (10, A) O/R; Ni (8.5–11.5, A) Y/P; Sc (2.6, HCl) Y/V; Th (2.5) Y/Pk; Zn (8–9, A) P/V

<sup>a</sup>Color index (CI) number, if known, is given in parentheses after the common name/synonyms of the indicator.

<sup>b</sup>The solvents used for preparing the indicator are given as E (ethanol), M (methanol), and W (water). This is followed by the weight percentage of the indicator.

<sup>c</sup>The buffer solutions are represented as A (NH<sub>3</sub>/NH<sub>4</sub>Cl), Ac (acetate), Bx (borax), as (ascorbic acid) Py (pyridine), t (tartrate), Ur (urotropine), and Chl. Ac (chloroacetic acid).

<sup>d</sup>The following abbreviations have been used to represent color: B = blue; Bn = brown; C = colorless; Cn = crimson; Cr = carmine; G = green; Gy = gray; O = orange; P = purple; Pk = pink; R = red; Rs = rose; WR = wine red; St = scarlet; V = violet; Y = yellow.



**Figure 1** (A) AgCl precipitated in the presence of excess Cl<sup>-</sup>. (B) AgCl precipitated in the presence of excess Ag<sup>+</sup>. (C) At the endpoint in the presence of indicator I<sup>-</sup>.

described by different trade names, so at least one such synonym, if available, is also given together with the color index number, if any. (2) The metal ions that can be titrated, the pH and buffer medium, and the color change at the equivalence point.

## Adsorption Indicators

Adsorption indicators were first introduced by Fajans for the detection of endpoints of precipitation titrations, particularly in argentimetry. Any substance that forms a deformable anion or cation exhibiting different colors in solution and when adsorbed may be capable of functioning as an adsorption indicator in a precipitation titration. Since any large ion is

deformable and all ionic precipitates show strong adsorptive properties, the field is extensive. In practice, colloidal precipitates that have a large surface-to-mass ratio, and organic dyestuffs that are weak acids or bases provide a fruitful field for the application of this technique. The classic example involves lyophobic silver halides and the anionic fluorescein family of dyes.

**Theory.** The action of these indicators is based on the properties of colloids and can readily be explained by considering the determination of chloride using silver nitrate solution as a titrant in the presence of fluorescein as indicator. The adsorption characteristics of AgCl precipitated during titration are shown diagrammatically in **Figure 1**.

At the equivalence point,  $\text{Ag}^+$  ions begin to be present in excess and the adsorbed layer corresponds to (B). However, the negatively charged fluorescein,  $\text{I}^-$ , when present in the solution, is adsorbed more strongly than  $\text{NO}_3^-$  and the precipitate appears pink. The color is due to the formation of a complex of silver and a distorted fluorescein ion on the surface at the appearance of the first trace of excess  $\text{Ag}^+$  ion. The origin of the color change is also attributed to a rearrangement of the structure of the indicator ion during the adsorption process. Thus, the indicator ion must be of opposite charge to the precipitating ion, and the indicator should only be adsorbed strongly immediately after the equivalence point. Typical adsorption indicators and their applications are listed in **Table 2**. The indicators have also been classified into surface precipitation, surface acid-base, surface redox, surface complexometric, and surface fluorescent indicators based on their mode of action.

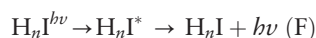
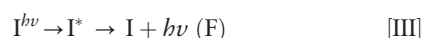
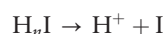
## Luminescent Indicators

Depending on the source of energy producing excited-state molecules, luminescent indicators available for titrimetric analysis can be conveniently divided into fluorescent (F) and chemiluminescent (CL) indicators. In the former incident radiation produces the excited state, and in the latter the chemical energy of a reaction raises a molecule or one of the reaction products to its excited state. The two types are considered separately below.

## Fluorescent Indicators

**Theory.** The F properties (such as color, intensity, or quenching of fluorescence) of many substances in solution are strongly dependent on the nature and concentration of other ions present in solution. The substances whose F characteristics are sensitive to pH (acid-base), pM (metallochromic), oxidation potential (redox), and adsorption effects (adsorption) have been used as indicators in appropriate titrimetric analyses.

The fluorescence of a good acid-base indicator should be in the visible region, and either the indicator  $\text{H}_n\text{I}$  or its ionized form  $\text{I}^-$  (charges are omitted) may be the fluorophore (see reaction [III]).



More than 75 F compounds have been recommended as acid-base indicators. These compounds conform to a few structural types. Typical compounds are listed in **Table 3A** along with their color change at the equivalence point, the pH range, and other relevant information.

In the case of F redox indicators, the oxidized state or the reduced state of the molecule is either F or non-fluorescent. Very few compounds have been proposed as F indicators for redox titrations. Rhodamine B and fluorescein are noted for the determination of  $\text{Sn(II)}$  and  $\text{As(III)}$  using  $\text{IO}_3^-$ ,  $\text{BrO}_3^-$ , and  $\text{MnO}_4^-$ .

**Table 2** Typical adsorption indicators

No.	Indicator	Ions titrable <sup>a</sup>	Color change at endpoint <sup>b</sup>
1	Alizarin sulfonic acid (Alizarin red)	$\text{SCN}^-$ , $\text{Fe(CN)}_6^{4-}$ , $\text{MoO}_4^{2-}$	Y/R
2	Bromophenol blue	$\text{Cl}^-$ , $\text{I}^-$	Y-G/B
3	Dichloro(P)fluorescein	$\text{Cl}^-$ , $\text{Br}^-$ , $\text{I}^-$	Y-G/Pk
4	Dichloro(P)tetraiodo(R)fluorescein (rose bengal)	$\text{I}^-$ in presence of $\text{Cl}^-$	R/B-R
5	Dichloro(R)fluorescein	$\text{Cl}^-$ , $\text{Br}^-$ , $\text{I}^-$ , $\text{BO}_3^-$	Y-G/R
6	Diiodo(R)fluorescein (erythrosin)	$\text{I}^-$	Y-R/R-V
7	Dimethyl(R)fluorescein	$\text{Cl}^-$ , $\text{Br}^-$ , $\text{I}^-$	G-Y/Pk
8	Diphenylamine	$\text{Cl}^-$ , $\text{Br}^-$ , $\text{I}^-$ , $\text{SCN}^-$	V/G
9	Fluorescein (resorcinol phthalein)	$\text{Cl}^-$ , $\text{Br}^-$ , $\text{I}^-$ , $\text{SCN}^-$	Y-G/Pk
10	Fluorescein complexone	$\text{Br}^-$ , $\text{I}^-$ , $\text{SCN}^-$	Y/Pk
11	Methyl red	$\text{Zn}^{2+}$ , $\text{Fe(CN)}_6^{4-}$	Pk/Y
12	Phenosafranine	$\text{Cl}^-$ , $\text{Br}^-$ , $\text{Ag}^+$ ( $\text{Br}^-$ )	R/B
13	Tartrazine	$\text{I}^- + \text{Cl}^-$ ; $\text{Ag}^+$ ( $\text{Cl}^-$ , $\text{Br}^-$ , $\text{I}^-$ , $\text{SCN}^-$ )	C/G
14	Tetrabromo(R)fluorescein (eosin)	$\text{Br}^-$ , $\text{I}^-$ , $\text{SCN}^-$	Pk/R-V
15	Tetrachloro(P)tetrabromo(R)fluorescein (phloxin)	$\text{Br}^-$ , $\text{I}^-$	Y-R/R-V
16	Tetraiodo(R)fluorescein (erythrosin-B)	$\text{I}^-$	R/B-R

<sup>a</sup>Titrimetric other than  $\text{Ag}^+$  are in parentheses.

<sup>b</sup>Abbreviations are as in **Table 1**.

as titrants. In cerimetry, Rhodamine 6G in the determination of U(IV), Fe(III), and V(IV) and 2,2'-bipyridyl for Ru(II) are some of the F redox indicators.

The function of metallofluorescent indicators is similar to metal-ion indicators in the sense that the endpoint of the titration is indicated by a change in F characteristics due to the destruction or formation of the metal-indicator complex. Although there are several organic compounds whose F characteristics are influenced by metal ions under specified conditions, only a

limited number have real use as indicators. A list of such indicators is given in Table 3B. Many of these compounds have fluorescent as well as chelating properties.

### Chemiluminescent Indicators

Although numerous CL reactions are known, relatively very few systems have proved useful as indicators in titrimetry. These are based largely on the CL reactions of luminol, lucigenin, lophine, and

**Table 3** Common fluorescent indicators  
**(A) Acid-base indicators**

No.	Compound	pH range	Color change at endpoint <sup>a</sup>
<i>Benzene derivatives</i>			
1	1-COOH, 2-NH <sub>2</sub>	1.5–3.0 4.5–6.0 12.5–14.0	nF/IB IB/dB dB/nF
2	1-COOH, 2-OH	2.5–4.0	nF/dB
3	1-CHO, 2-OCH <sub>3</sub>	3.1–4.4	nF/dB
4	1,4-(OH) <sub>2</sub> , 2,3-(CN) <sub>2</sub>	5.8–8.2	B/G
5	1-CH=CH-COOH, 2-OH	7.2–9.0	nF/G
6	1-(2-Benzothiazolyl), 2-OH	9.3	nF/B or g
<i>Naphthalene derivatives</i>			
7	2-OH	8.2–10.3	nF/Y or G
8	2-NH <sub>2</sub>	2.8–4.4	nF/V
9	1-OH, 2-SO <sub>3</sub> H	8.0–9.0	dB/IB
10	1-OH, 5-SO <sub>3</sub> H	6.5–7.5	nF/B
11	1-NH <sub>2</sub> , 4-SO <sub>2</sub> NH <sub>2</sub>	9.5–13.0	dB/W
12	2-NH <sub>2</sub> , 5-SO <sub>2</sub> NH <sub>2</sub>	12–13	B/V
13	2-NH <sub>2</sub> , 6,8-(SO <sub>3</sub> H) <sub>2</sub>	12–14	Y/Pk
14	2,8-(OH) <sub>2</sub> , 3,6-(SO <sub>3</sub> H) <sub>2</sub>	3.1–4.4	nF/B
15	1-COOH	2.5–3.5	nF/B
16	1-COOH, 3-NH <sub>2</sub>	1.5–3.0 4.0–6.0 11.6–13.0	nF/G G/B B/nF
<i>Heterocyclic compounds</i>			
17	Acridine	5.2–6.6	G/V
18	4-Ethoxy-9-acridone	1.2–3.2	G/B
19	3,6-Bis(dimethylamino)acridine (acridine orange)	8.4–10.4	nF/Y or G
20	2,7-Dimethyl-3,6-diamino-4-phenylacridine	0.3–1.7	Y/G
21	Coumarin	9.5–10.5	nF/G
22	7-Hydroxycoumarin (umbelliferone)	6.5–8.0	nF/G
23	6,7-Dihydroxycoumarin	1.5–2.0	IB/dB
24	4-Methyl-7-hydroxycoumarin	7.0	nF/B
25	4-Methyl-9,7-dihydroxycoumarin		nF/B
26	Morin	3.1–4.4 8–9.8	nF/G G/Y
27	Fluorescein	4.0–6	Pk/G
28	Dichloro(P)fluorescein	4.0–6.6	B-G/G
29	Tetrabromo(R)fluorescein (eosin)	0–2.4 6–8.0	B/G G/Y
30	Quinoline	6.2–7.2	B/nF
31	3-Naphthoquinoline	4.4–6.3	B/nF
32	6,7-Dimethoxyisoquinoline-1-carboxylic acid	9.5–11.0	Y/B
33	Quinine	3.0–5.0	B/V
34	Thioflavine S (CI No. 49010, direct yellow 7)	3.1–4.4	dB/IB

Continued

**Table 3** Continued  
**(B) Important metallofluorescent indicators<sup>b</sup>**

Compound	Ions titrated (pH), color change; other information <sup>b</sup>
Calcein or fluoxerone	Al <sup>3+</sup> , Mg <sup>2+</sup> , Ca <sup>2+</sup> , Cr <sup>3+</sup> , Mn <sup>2+</sup> , Co <sup>2+</sup> , Ni <sup>2+</sup> , Cu <sup>2+</sup> , Zn <sup>2+</sup> , and Ca <sup>2+</sup> in blood serum (5–10), nF/Y or G; used as 1% in KCl
Methyl calcein	Mn <sup>2+</sup> (9–5), Cu <sup>2+</sup> (5–5.5), using DCYT, nF/Y or G; used as 1% solution in H <sub>2</sub> O
Calcein blue	Cu <sup>2+</sup> (4.8) nF/B; Ca <sup>2+</sup> , Sr <sup>2+</sup> , Ba <sup>2+</sup> (~12, back titration) B/nF; Cr <sup>3+</sup> , Ni <sup>2+</sup> , Co <sup>2+</sup> , Fe <sup>2+</sup> , Ti <sup>4+</sup> , Al <sup>3+</sup> , Mn <sup>2+</sup> , Zn <sup>2+</sup> , Ag <sup>+</sup> (4.8 back titration), B/nF; used as 0.1% solution in H <sub>2</sub> O
8-Quinololinol	Ga <sup>3+</sup> (2.5–3.5) Y or G/nF; used as 0.09–0.5% solution in acetic acid;
8-Quinololinol-5-sulfonic acid	Zn <sup>2+</sup> (10), Zn <sup>2+</sup> , Cd <sup>2+</sup> , Mg <sup>2+</sup> (7), with diamino-cyclohexane tetraacetic acid, Y or G/nF; used as 0.2% solution in H <sub>2</sub> O
Morin	Ga <sup>3+</sup> (3–8), In <sup>3+</sup> (5.0) G/nF; used as 0.5% solution in methanol; Al <sup>3+</sup> (3) B/G

<sup>a</sup> F = fluorescent; nF = nonfluorescent; IB = light blue; dB = dark blue; G = green; O = orange; Pk = pink; V = violet; W = white; Y = yellow.

<sup>b</sup> Unless otherwise indicated, the titrant is Na<sub>2</sub>EDTA; the pH range is given in parentheses.

**Table 4** Typical chemiluminescent indicators  
**(A) Indicator systems for acid–base titrations**

Indicator system	Color of CL at endpoint	pH range at which CL occurs	Indicator preparation
Luminol-H <sub>2</sub> O <sub>2</sub> -catalyst	Blue	8–8.5	0.01% in 0.005 mol l <sup>-1</sup> aqueous NaOH
Luminol-fluorescein-H <sub>2</sub> O <sub>2</sub>	Green	8–9	Luminol (1%) + fluorescein (0.15%) in 0.01 mol l <sup>-1</sup> NaOH
Lucigenin-H <sub>2</sub> O <sub>2</sub>	Green	9–10	0.5% in water
Lophine-H <sub>2</sub> O <sub>2</sub> -catalyst	Greenish white	9–10	0.45% in ethanol

**(B) Indicators for redox titrations**

Indicator	Titrant	Species determined
Luminol	OCl <sup>-</sup>	As(III), N <sub>2</sub> H <sub>4</sub>
	OBr <sup>-</sup>	As(III), Sb(III), SCN <sup>-</sup> , CN <sup>-</sup> , S <sub>2</sub> O <sub>3</sub> <sup>2-</sup> , SO <sub>3</sub> <sup>2-</sup> , S <sup>2-</sup>
Lucigenin	H <sub>2</sub> O <sub>2</sub>	As(III), Fe(CN) <sub>6</sub> <sup>3-</sup>
Siloxene	Ce(IV)	Fe <sup>2+</sup>
	Cr <sub>2</sub> O <sub>7</sub> <sup>2-</sup>	Fe <sup>2+</sup> , Ti <sup>+</sup> , hydroquinone

siloxene systems. Their use as CL indicators is based on the fact that light emission is initiated only under definite conditions of pH and redox potential. If the conditions for the start or termination of the CL reaction are those obtained at the endpoint of a titration, the indicator will function properly to yield or quench chemiluminescence. The CL reaction of these systems is oxidative, involving oxygen, peroxide, or another strong oxidizing agent. CL indicators do not require an excitation light source or other special equipment, unlike F indicators. The CL emission can easily be observed in a semi-darkened room.

The indicator systems recommended for acid–base titrations are listed in Table 4A. The pH range and color of the CL emission and the indicator preparation are also included. In a typical direct titration, the indicator, an oxidant (usually 3% H<sub>2</sub>O<sub>2</sub> solution), and a catalyst (such as Fe(CN)<sub>6</sub><sup>3-</sup> or blood hemin in the case of luminol and lophine) are added to the acid

and titrated against the base. The equivalence point is indicated by the initiation of bright CL emission. These indicators were successfully used for the determination of the acidity of intensely colored and turbid solutions such as milk, wines, fruit juices and mustard, and the saponification number of fats and oils. Since H<sub>2</sub>O<sub>2</sub> is essential for the CL reaction, the indicator cannot be used in the presence of heavy metals, as H<sub>2</sub>O<sub>2</sub> is decomposed before the endpoint. Further, in most cases the indicator is irreversible.

In alkaline solutions when suitable reducing agents are titrated with oxidizing agents such as H<sub>2</sub>O<sub>2</sub> or a hypohalite in the presence of luminol and lucigenin, at the endpoint the redox potential of the solution suddenly moves towards a more positive value and the CL reaction occurs. In acidic solution (pH < 3.5), siloxene has been used as a CL redox indicator. An outline of the use of these indicators in redox titrations is presented in Table 4B.



See also: **Chemiluminescence:** Overview; Liquid-Phase. **Fluorescence:** Overview. **Indicators:** Acid-Base.

### Further Reading

- Bishop E (ed.) (1972) *Indicators*, pp. 209–468, 685–732. Oxford: Pergamon.
- Guilbault GG (1973) *Practical Fluorescence Theory, Methods and Techniques*, pp. 397–408, 597–607. New York: Dekker.
- Isacsson UI and Wettermark G (1974) Chemiluminescence in analytical chemistry. *Analytica Chimica Acta* 68: 339–362.

- Ringbom A (1958) Complexation reactions. In: Kolthoff IM and Elving PJ (eds.) *Treatise on Analytical Chemistry*, Part I, vol. I, ch. 14. New York: Interscience.
- Ringbom A (1963) *Complexation in Analytical Chemistry*. New York: Wiley-Interscience.
- Schwarzenbach G and Flaschka H (1969) *Complexometric Titrations*, 2nd edn. London: Methuen.
- Townshend A, Burns DT, Guilbault GG, *et al.* (eds.) (1993) *Dictionary of Analytical Reagents*. London: Chapman and Hall.
- Welcher FJ (1958) *The Analytical Uses of Ethylenediaminetetraacetic Acid*. Princeton, NJ: Van Nostrand.
- Wilson CL and Wilson DW (1960) (eds.) *Comprehensive Analytical Chemistry*, vol. 1B. Amsterdam: Elsevier.

## INDUCTIVELY COUPLED PLASMA

See **ATOMIC EMISSION SPECTROMETRY: Inductively Coupled Plasma**

## INDUCTIVELY COUPLED PLASMA-MASS SPECTROMETRY

See **ATOMIC MASS SPECTROMETRY: Inductively Coupled Plasma**

## INFRARED SPECTROSCOPY

Contents

**Overview**

**Sample Presentation**

**Near-Infrared**

**Photothermal**

**Industrial Applications**

### Overview

**P R Griffiths**, University of Idaho, Moscow, ID, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by B P Clark, pp. 2153–2170, © 1995, Elsevier Ltd.

### Introduction

The infrared (IR) region of the electromagnetic spectrum lies between  $\sim 10$  and  $12\,800\text{ cm}^{-1}$ . The

energy of IR photons is thus of the same order of magnitude as the energy differences between quantized molecular vibrational states. Transitions between these vibrational modes can be induced by IR radiation if there is a change in the molecular electric dipole moment in the course of the vibrational motion. IR spectroscopy is the study of the interaction of IR radiation with matter as a function of photon frequency. This interaction can take the form of absorption, emission, or reflection. IR spectroscopy is a fundamental analytical technique for obtaining quantitative and qualitative information about a substance in the solid, liquid, or vapor state.

See also: **Chemiluminescence:** Overview; Liquid-Phase. **Fluorescence:** Overview. **Indicators:** Acid-Base.

### Further Reading

- Bishop E (ed.) (1972) *Indicators*, pp. 209–468, 685–732. Oxford: Pergamon.
- Guilbault GG (1973) *Practical Fluorescence Theory, Methods and Techniques*, pp. 397–408, 597–607. New York: Dekker.
- Isacsson UI and Wettermark G (1974) Chemiluminescence in analytical chemistry. *Analytica Chimica Acta* 68: 339–362.

- Ringbom A (1958) Complexation reactions. In: Kolthoff IM and Elving PJ (eds.) *Treatise on Analytical Chemistry*, Part I, vol. I, ch. 14. New York: Interscience.
- Ringbom A (1963) *Complexation in Analytical Chemistry*. New York: Wiley-Interscience.
- Schwarzenbach G and Flaschka H (1969) *Complexometric Titrations*, 2nd edn. London: Methuen.
- Townshend A, Burns DT, Guilbault GG, *et al.* (eds.) (1993) *Dictionary of Analytical Reagents*. London: Chapman and Hall.
- Welcher FJ (1958) *The Analytical Uses of Ethylenediaminetetraacetic Acid*. Princeton, NJ: Van Nostrand.
- Wilson CL and Wilson DW (1960) (eds.) *Comprehensive Analytical Chemistry*, vol. 1B. Amsterdam: Elsevier.

## INDUCTIVELY COUPLED PLASMA

See **ATOMIC EMISSION SPECTROMETRY: Inductively Coupled Plasma**

## INDUCTIVELY COUPLED PLASMA-MASS SPECTROMETRY

See **ATOMIC MASS SPECTROMETRY: Inductively Coupled Plasma**

## INFRARED SPECTROSCOPY

Contents

**Overview**

**Sample Presentation**

**Near-Infrared**

**Photothermal**

**Industrial Applications**

### Overview

**P R Griffiths**, University of Idaho, Moscow, ID, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by B P Clark, pp. 2153–2170, © 1995, Elsevier Ltd.

### Introduction

The infrared (IR) region of the electromagnetic spectrum lies between  $\sim 10$  and  $12\,800\text{ cm}^{-1}$ . The

energy of IR photons is thus of the same order of magnitude as the energy differences between quantized molecular vibrational states. Transitions between these vibrational modes can be induced by IR radiation if there is a change in the molecular electric dipole moment in the course of the vibrational motion. IR spectroscopy is the study of the interaction of IR radiation with matter as a function of photon frequency. This interaction can take the form of absorption, emission, or reflection. IR spectroscopy is a fundamental analytical technique for obtaining quantitative and qualitative information about a substance in the solid, liquid, or vapor state.

**Table 1** The IR region of the spectrum

Region	$\lambda$ (cm)	$\bar{\nu}$ ( $\text{cm}^{-1}$ )	$\nu$ (Hz)
Near-IR	$2.5 \times 10^{-4}$ – $7.8 \times 10^{-5}$	4000–12 800	$1.2 \times 10^{14}$ – $3.8 \times 10^{14}$
Mid-IR	$5 \times 10^{-3}$ – $2.5 \times 10^{-4}$	200–4000	$6 \times 10^{12}$ – $1.2 \times 10^{14}$
Far-IR	$0.1$ – $5 \times 10^{-3}$	10–200	$3 \times 10^{11}$ – $6 \times 10^{12}$

It is convenient to divide the IR region into three parts (**Table 1**): the far-IR ( $10$ – $200 \text{ cm}^{-1}$ ); the mid-IR ( $200$ – $4000 \text{ cm}^{-1}$ ); and the near-IR ( $4000$ – $12\,800 \text{ cm}^{-1}$ ). (The regions are not exactly defined: slightly different boundaries for the IR regions are found in the literature. In particular, most Fourier transform infrared (FTIR) spectrometers operating in the mid-IR region have a low-wave-number limit of  $400 \text{ cm}^{-1}$ .) Each part of the spectrum plays a different role in analysis according to the different character of the transitions involved in each case.

Mid-IR radiation corresponds to fundamental transitions in which one vibrational mode is excited from its lowest energy state to its first excited state. For routine analysis a spectrum is normally taken from  $400$  to  $4000 \text{ cm}^{-1}$ . The mid-IR spectrum of a substance is effectively a unique fingerprint that can be used for the purpose of identification by comparison with a reference spectrum. When no reference spectrum is available, an IR spectrum can be used to identify the presence of certain structural units that, irrespective of their molecular environment, give rise to characteristic spectral features in a narrow frequency range.

Near-IR spectroscopy arises from transitions in which a photon excites a normal mode of vibration from the ground state to the second or higher excited vibrational state (overtones, *vide infra*) or transitions in which one photon simultaneously excites two or more vibrational modes (combinations bands, *vide infra*). The use of the near-IR, especially diffuse reflection spectroscopy, in both quantitative and qualitative analysis has increased significantly due to better instrumentation and the development of chemometrics to better handle the effect of seriously overlapping bands.

The far-IR region of the spectrum ( $<200 \text{ cm}^{-1}$ ) results from transitions involving low-frequency torsions and internal rotations in liquids and lattice vibrations in solids and is not commonly used for analysis, although recent developments in instrumentation for terahertz spectrometry may change the situation.

The scope and flexibility of IR spectroscopy in the mid-IR region have been greatly increased by the advent of FTIR spectroscopy. The multiplex and throughput advantages of this technique allow spectra to be run faster and with a greater signal-to-noise

ratio than dispersive spectroscopy, i.e., measurements made with prism or grating monochromators.

The following brief description of the principal IR techniques – using the near- or mid-IR – illustrates the range of sample handling possible with IR spectroscopy.

‘Diffuse reflection’ is the term used to describes the reflection of electromagnetic radiation from a sample after the radiation has undergone multiple scattering inside a powdered sample or at the surface of a matte substance. The radiation passes through the ‘micro-structural’ elements of the sample, e.g., the micro-crystallites of a powder or the surface fibers of a fabric, and is absorbed in the process before being scattered out of the sample to detector. The use of diffuse reflection spectroscopy in the ultraviolet (UV) region of the spectrum is a long-established technique. However, until FTIR was established, the weakness of the signal prevented the extension of the technique into the mid-IR. Diffuse reflection IR Fourier transform (DRIFT) spectroscopy has become a useful technique for obtaining IR spectra from powdered samples (or any matte material) with little or no sample preparation.

Photoacoustic IR spectroscopy has similar advantages to DRIFT spectroscopy in its ability to handle solids with the minimum of preparation. The principle of this technique is that when a modulated beam of IR radiation is absorbed by a sample, temperature oscillations set up thermal waves. If the sample is sealed in a cell and surrounded by gas, then a microphone can pick up the sound waves in the gas and an IR absorption spectrum generated.

‘Specular reflection’ is the term used to describe ‘mirror-like’ reflection, from the surface of a sample (angle of reflection equals angle of incidence). Specular reflected radiation ostensibly carries no information about the IR absorption of a sample and is a source of interference in diffuse reflection experiments when the sample is not completely matte, i.e., has an element of ‘shininess’ about it. However, if the reflected intensity from a sample is due ‘principally’ to reflection from the front surface of the sample, then an absorption index spectrum of the sample can be generated from the reflected intensity over the whole spectrum using the Kramers–Krönig transformation. (This complex transformation is an

integral part of the software packages driving most modern FTIR spectrometers.)

Aqueous solutions have traditionally posed a problem for IR spectroscopy due to the fact that water is a strong absorber of IR radiation. This difficulty, for aqueous solutions and other strongly absorbing liquid and solid samples, can be overcome by using attenuated total reflection spectroscopy. In this technique, the phenomenon of total internal reflection is used in such a way that it is only the evanescent wave associated with total internal reflection that enters the sample. The evanescent wave penetrates the sample very short distances only, hence the advantage for strongly absorbing species.

FTIR microscopy, in which IR spectra can be obtained from picogram quantities, is an invaluable nondestructive analytical tool in fields such as forensic science and pharmaceutical analysis.

The short data-capture times possible with the FTIR spectrometer means that time-resolved spectroscopy has become an important means of following the course of a chemical reaction in order to obtain information about kinetics, equilibria, and the nature of reaction intermediates.

Matrix isolation IR spectroscopy involves mixing trace amounts of solute into a rare gas matrix at low temperature. The advantage is that the solute molecules are isolated so that reactive species can be analyzed. Also, the absence of rotational structure and lattice modes increases resolution.

For chiral molecules a small difference in the magnitude of absorption of left- and right-circularly polarized IR radiation is observed. This is known as vibrational circular dichroism and, since the effect can be observed from each normal mode, absolute stereochemical information can be obtained from the entire molecule. This is different from the UV analog of this effect, where there may often be only one chromophore present.

## The Vibration of Diatomic Molecules

### The Classical Diatomic Rigid Vibrator

An understanding of the nature of vibrational motion is best obtained by first studying a simple system. To introduce some of the basic concepts involved it is useful to study the classical diatomic vibrator before going on to consider the quantum theory.

The simplest model for a diatomic molecule consists of two atoms of mass  $m_1$  and  $m_2$  connected by a rigid, massless spring of length  $r$ , which has the value  $r_0$  at equilibrium (Figure 1). If the  $z$ -axis is taken to lie along the internuclear line, then the Cartesian coordinates of the two atoms, referred to the center of

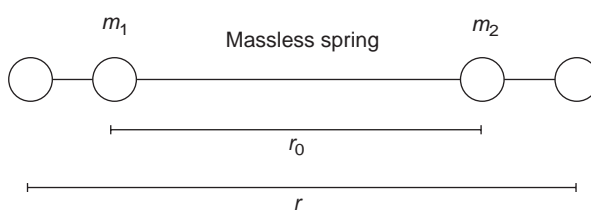


Figure 1 Model for a rigid diatomic vibrator.

mass, may be written as

$$z_1 = z_1^0 + \Delta z_1 \quad [1a]$$

$$z_2 = z_2^0 + \Delta z_2 \quad [1b]$$

where  $z_1^0$  and  $z_2^0$  are the equilibrium coordinates of atoms 1 and 2, respectively, and  $\Delta z_1$  and  $\Delta z_2$  are the 'Cartesian displacements coordinates'. An 'internal coordinate'  $R$  can be defined as the differences in the bond length from its equilibrium value

$$R = \Delta z_2 - \Delta z_1 = r - r_0 \quad [2]$$

The potential energy of the molecule  $V(R)$ , which increases as the bond is stretched or compressed, may be expressed as a power series in the internal coordinate

$$V(R) = V(0) + \left( \frac{dV(R)}{dR} \right)_0 R + \frac{1}{2} \left( \frac{d^2 V(R)}{dR^2} \right)_0 R^2 + \frac{1}{6} \left( \frac{d^3 V(R)}{dR^3} \right)_0 R^3 + \cdots \quad [3]$$

The subscripts (zero) indicate that the derivatives are to be taken at the equilibrium bond length. Only changes in the potential energy from the equilibrium value are important, so the energy scale may be chosen such that  $V(0)$  is zero. Also, the first derivative of the potential energy  $(dV(R)/dR)_0$  must be zero by definition at the equilibrium position since this corresponds to the energy minimum. This leaves only the quadratic and higher terms in the potential function

$$V(R) = \frac{1}{2} \left( \frac{d^2 V(R)}{dR^2} \right)_0 R^2 + \frac{1}{6} \left( \frac{d^3 V(R)}{dR^3} \right)_0 R^3 + \cdots \quad [4]$$

The cubic and higher terms are generally small for small departures from the equilibrium position. The effects of including these contributions in the potential energy expression will be considered later. It is, however, a good first approximation to set these higher-order terms to zero to give

$$V(R) = \frac{1}{2} \left( \frac{d^2 V(R)}{dR^2} \right)_0 R^2 = \frac{1}{2} K R^2 \quad [5]$$

A system with a potential function given by eqn [5] is said to be 'mechanically harmonic' and  $K$  is known

as the ‘force constant’ for the bond where

$$K = \left( \frac{d^2 V(R)}{dR^2} \right)_0 \quad [6]$$

The harmonic approximation corresponds to a restoring force  $F$  acting on the atoms that is proportional to the displacement of the bond length from its equilibrium value (Hooke’s law)

$$F = -KR \quad [7]$$

With the center of gravity as the coordinate origin, it is straightforward to show that the problem reduces to that of a single particle oscillating around the center of mass subject to a harmonic restoring force whose displacement is equal to the change in the internuclear distance of the molecule, i.e., the internal coordinate  $R$ . The mass  $\mu$  of this particle is known as the reduced mass of the molecule and is given by

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad [8]$$

Applying Newton’s second law gives

$$\frac{\mu}{2} \frac{d^2 R}{dt^2} = -KR \quad [9]$$

This equation describes the simple harmonic oscillator. The solution is

$$R = R^0 \sin(2\pi\nu t + \delta) \quad [10]$$

which corresponds to simple harmonic motion with maximum amplitude  $R^0$ , phase factor  $\delta$ , and frequency  $\nu$  given by

$$\nu = \frac{1}{2\pi} \left( \frac{K}{\mu} \right)^{1/2} \quad [11]$$

It can be shown that the maximum amplitude of vibration for each atom is inversely proportional to the atomic mass.

### The Quantum-Mechanical Harmonic Diatomic Vibrator

The harmonic potential function obtained in eqn [5]

$$V(R) = \frac{1}{2} KR^2$$

can be used in Schrödinger’s equation to yield the following wave function:

$$\Psi_v(R) = \frac{(\alpha/\pi)^{1/4}}{(2^v v!)^{1/2}} \exp\left(-\frac{\alpha R^2}{2}\right) H_v(\alpha^{1/2} R) \quad [12]$$

where  $\nu$  is the vibrational quantum number, and

$$\alpha = 2\pi \frac{(\mu K)^{1/2}}{h} = \frac{4\pi^2 \mu \nu}{h}$$

where  $H_v(\alpha^{1/2} R)$  are Hermite polynomials (Table 2).

**Table 2** The first few Hermite polynomials

$\nu$	$H_\nu(x)$
0	1
1	$2x$
2	$4x^2 - 2$
3	$8x^3 - 12x$
4	$16x^4 - 48x^2 + 12$
5	$32x^5 - 160x^3 + 120x$

The vibrational energy levels are given by

$$E_\nu = \left(\nu + \frac{1}{2}\right) h\nu, \quad \nu = 0, 1, 2, \dots \quad [13]$$

where

$$\nu = \frac{1}{2\pi} \left( \frac{K}{\mu} \right)^{1/2}$$

Note that this expression is exactly the same as eqn [11], the classical vibration frequency.

Vibrational term values  $G(\nu)$  in wave numbers can be defined from eqn [13]:

$$G(\nu) = \frac{E_\nu}{hc} = \left(\nu + \frac{1}{2}\right) \frac{h\nu}{c} = \left(\nu + \frac{1}{2}\right) \tilde{\nu} \quad [14]$$

Inspection of eqn [13] shows that the vibrational zero-point energy  $E_0$  is given by

$$E_0 = \frac{1}{2} h\nu \quad [15]$$

or

$$G(0) = \frac{1}{2} \tilde{\nu} \quad [16]$$

Figure 2 shows plots of the harmonic diatomic vibrator wave functions.

For harmonic wave function, the properties of the Hermite polynomials are such that the selection rules for vibrational transitions are

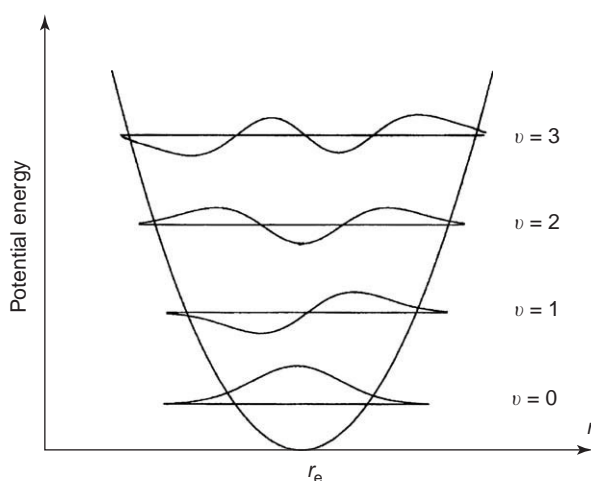
$$\Delta\nu = \pm 1 \quad [17]$$

There also exists the gross selection rule that, in order for electromagnetic radiation to be absorbed, the dipole moment of the molecule must change during the vibration, which means that a diatomic molecule must possess a permanent dipole moment in order to absorb IR radiation. (These selection rules will be discussed in detail in the section on the intensity of IR transitions.)

Since all energy levels are equally spaced for a harmonic diatomic vibrator, one line will appear in the IR absorption spectrum with wave number

$$\tilde{\nu} = \frac{1}{2\pi c} \left( \frac{K}{\mu} \right)^{1/2} \quad [18]$$





**Figure 2** Harmonic diatomic vibrator wave functions.

**Table 3** Vibrational frequencies and force constants for selected bonds

Bond	$\bar{\nu}$ ( $\text{cm}^{-1}$ )	$K$ ( $\times 10^2 \text{ Nm}^{-1}$ ) <sup>a</sup>
C–H	2960	4.7
=C–H	3020	5.0
≡C–H	3300	5.9
(OC)–H	2800	4.3
C–C	900	2.9
C=C	1650	9.6
C≡C	2050	15
C=O	1700	12
C≡N	2100	17
O–H	3600	7.2
N–H	3350	6.2
C–F	1100	5.3
C–Cl	650	2.2
C–Br	560	1.9
C–I	500	1.6

<sup>a</sup> 1 mdyne  $\text{\AA}^{-1} = 100 \text{ Nm}^{-1}$ .

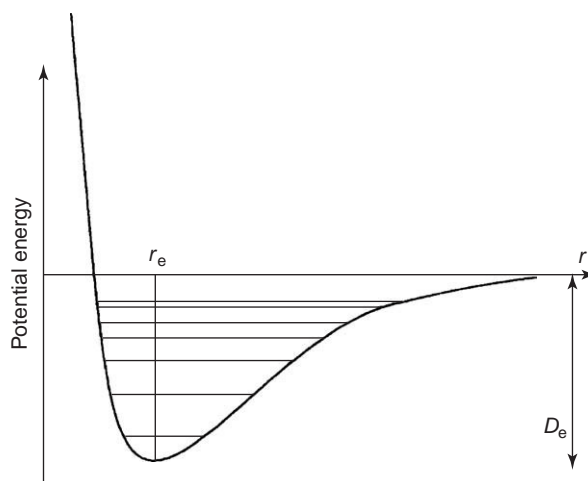
Equation [18] can be used to provide approximate values for the force constants of bonds (Table 3).

### Isotopic Substitution

If one of the atoms in the diatomic molecule is replaced by another isotopic species, then to a good approximation the electronic structure is unchanged. Hence, the force constant will be the same for both molecules. The change in the vibrational frequency will therefore be completely due to the change in the reduced mass. If  $\mu'$  and  $\tilde{\nu}'$  are the reduced mass and wave number of the isotopically substituted molecule, then

$$\frac{\tilde{\nu}'}{\tilde{\nu}} = \left(\frac{\mu'}{\mu}\right)^{1/2} \quad [19]$$

Similar relationships for frequency shifts induced by isotopic substitution in polyatomic molecules are given later.



**Figure 3** The Morse potential function.

### Anharmonic Vibrations of a Diatomic Molecule

The true potential function for a diatomic molecule departs from harmonicity, especially for large amplitude vibrations. It is useful to approximate an anharmonic potential using a Morse function (Figure 3)

$$E_{\text{anh}} = D_e \{1 - \exp[-a(r - r_e)]\}^2 \quad [20]$$

where  $D_e$  is the thermodynamic dissociation energy and

$$a = \omega \left(\frac{\mu}{2D_e}\right)^{1/2}, \quad \omega = 2\pi\nu \quad [21]$$

When this potential energy is used in the Schrödinger equation the vibrational levels become

$$G(\nu) = \left(\nu + \frac{1}{2}\right) \tilde{\nu}_e - \left(\nu + \frac{1}{2}\right)^2 x_e \tilde{\nu}_e, \dots, \quad x_e = \frac{a^2 \hbar}{2\mu\omega} \quad [22]$$

where  $\tilde{\nu}_e$  is the wave number corrected for anharmonicity, and  $x_e$  is the anharmonicity constant, a positive number  $\sim 10^{-2}$ . The negative term in eqn [22] results in the gap between successive vibrational energy levels decreasing as  $\nu$  increases. The point at which the levels merge into a continuum corresponds to the dissociation of the molecule.

Another effect of anharmonicity is to relax the selection rules to give

$$\Delta\nu = \pm 1, \pm 2, \pm 3, \dots \quad [23]$$

Along with the fundamental vibrations, 'harmonics' are now observed in the IR spectrum: the second harmonic (or first overtone) ( $\nu = 2 \leftarrow \nu = 0$ ), third harmonic (or second overtone) ( $\nu = 3 \leftarrow \nu = 0$ ) and so on occur with decreasing intensity. A knowledge of the wave numbers of the fundamental and first

**Table 4** Anharmonicity constants for some diatomic molecules

Bond	$\bar{\nu}_e$ ( $\text{cm}^{-1}$ )	$x_e \bar{\nu}_e$ ( $\text{cm}^{-1}$ )
H- <sup>12</sup> C	2861.6	64.3
H- <sup>19</sup> F	4138.5	90.07
H- <sup>16</sup> O	3735.2	82.81
H- <sup>35</sup> Cl	2991.0	52.85
H-Br	2649.7	45.21

overtone vibrations allows the anharmonicity constants to be calculated using eqn [22]. Table 4 shows values of anharmonicity constants for a selection of diatomic molecules.

## Normal Coordinates

Computational and interpretational aspects of vibrational spectroscopy are greatly simplified by the introduction of ‘normal coordinates’. First, mass-weighted Cartesian displacement coordinates for an  $N$ -atom molecule  $q_1, q_2, \dots, q_{3N}$  are defined according to

$$\begin{aligned} q_1 &= \sqrt{m_1} \Delta x_1, & q_2 &= \sqrt{m_1} \Delta y_1, \\ q_3 &= \sqrt{m_1} \Delta z_1, & q_4 &= \sqrt{m_2} \Delta x_2, \dots \end{aligned} \quad [24]$$

In these coordinates the kinetic energy  $T$  is given by

$$T = \frac{1}{2} \sum_{i=1}^{3N} \dot{q}_i^2 \quad [25]$$

where the dot indicates a time derivative. The potential energy is given by

$$\begin{aligned} V &= \frac{1}{2} \sum_{i,j=1}^{3N} \left( \frac{\partial V}{\partial q_i \partial q_j} \right)_0 q_i q_j \\ &= \frac{1}{2} \sum_{i,j=1}^{3N} b_{ij} q_i q_j \end{aligned} \quad [26]$$

where the  $b_{ij}$  denote the force constants in the Cartesian displacement coordinate system. (Note that there are cross-terms in the potential function involving two coordinates.)

The total energy is therefore given by

$$E = \frac{1}{2} \sum_{i=1}^{3N} \dot{q}_i^2 + \frac{1}{2} \sum_{i,j=1}^{3N} b_{ij} q_i q_j \quad [27]$$

If the potential energy term did not include any cross-terms, then Newton’s equation could be applied

$$\frac{d}{dt} \left( \frac{\partial T}{\partial \dot{q}_i} \right) + \left( \frac{\partial V}{\partial q_i} \right) = 0, \quad i = 1, 2, \dots, 3N \quad [28]$$

In this case the problem would reduce to the solution of  $3N$  independent equations. Therefore, in order to

enable the use of Newton’s equation, the mass-weighted Cartesian coordinates  $q_i$  are transformed into a set of new coordinates  $Q_i$  which results in no cross-terms in the potential function.

The  $Q_i$  are called the ‘normal coordinates’ of the system, and this transformation from Cartesian displacement to normal coordinates is the essence of the vibrational problem. Using normal coordinates, the kinetic and potential energy become

$$T = \frac{1}{2} \sum_{i=1}^{3N} \dot{Q}_i^2 \quad [29]$$

$$V = \frac{1}{2} \sum_{i=1}^{3N} \lambda_i Q_i^2 \quad [30]$$

Using eqns [29] and [30] in eqn [28] instead of the mass-weighted Cartesian coordinates gives

$$\ddot{Q}_i + \lambda_i Q_i = 0, \quad i = 1, 2, \dots, 3N \quad [31]$$

These are simply harmonic oscillator equations with solutions

$$Q_i = Q_i^0 \sin(\sqrt{\lambda_i} t + \delta_i) \quad [32]$$

with frequencies

$$\nu_i = \frac{1}{2\pi} \sqrt{\lambda_i} \quad [33]$$

Each vibration associated with a normal coordinate is known as a ‘normal vibration’. Each atom involved in a normal coordinate vibrates in phase with all the other atoms involved in the same vibration. Each atom passes through its equilibrium position at the same time and reaches each turning point at the same time.

An analysis of the stretching vibrational motion of a symmetrical linear triatomic molecule (Figure 4) reveals that in terms of mass-weighted Cartesian coordinates  $q_1, q_2$ , and  $q_3$  the normal vibrations (excluding a zero-frequency translational solution) are given by

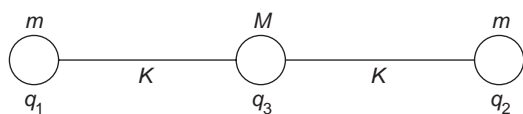
$$Q_+ = \frac{1}{2^{1/2}} q_1 - \frac{1}{2^{1/2}} q_2 \quad [34a]$$

$$Q_- = \left[ \frac{M}{2(M+2m)} \right]_1^q - \left[ \frac{2m}{M+2m} \right]_2^q + \left[ \frac{M}{2(M+2m)} \right]_3^q \quad [34b]$$

with frequencies

$$\nu_+ = \frac{1}{2\pi} \left( \frac{K}{m} \right)^{1/2} \quad [35a]$$

$$\nu_- = \frac{1}{2\pi} \left[ \frac{K(M+2m)}{Mm} \right]^{1/2} \quad [35b]$$



**Figure 4** The symmetrical linear diatomic molecule.

The + and – labels have been used to denote that the two normal vibrations are symmetric and anti-symmetric stretches, respectively.

The two bonds in this system are known as ‘coupled oscillators’. This type of system occurs often in the vibrations of polyatomic molecules in which two oscillators (which can be bonds or groups of bonds) couple to give symmetric and antisymmetric combinations. (In this case, where the system is exactly symmetrical, the symmetric combination will not be observed in the IR spectrum because the dipole moment of the molecule does not change.)

## High-Resolution IR Spectra of Linear Molecules

Each vibrational transition of vapor-phase molecules is accompanied by rotational transitions. In the liquid state the effect of molecular collisions is to broaden the rotational lines so that they cannot be resolved. This is why one observes IR bands as opposed to lines in solution spectra. In the vapor phase, however, it is possible to resolve the rotational structure in a vibrational transition.

### The Diatomic Vibrating-Rotator

It is normally a good approximation to express the total energy due to the motion of the nuclei  $E_{\text{nuc}}$  as the sum of the separate energies  $E_{\text{vib}}$  and  $E_{\text{rot}}$  (the Born–Oppenheimer approximation)

$$E_{\text{nuc}} = E_{\text{vib}} + E_{\text{rot}} \quad [36]$$

To a first approximation, wave numbers of the rotational levels are given by

$$\frac{E_J}{hc} = BJ(J+1) \text{ cm}^{-1}, \quad J = 0, 1, 2, \dots \quad [37]$$

when  $J$  is the rotational quantum number and  $B$  is the rotational constant given by

$$B = \frac{h}{8\pi^2 Ic} \text{ cm}^{-1} \quad [38]$$

where  $I$  is the moment of inertia of the molecule given by

$$I = \mu r_e^2 \quad [39]$$

where  $r_e$  is the equilibrium bond length.

Using eqns [22] and [37] the total vibrational and rotational energy is given by

$$\frac{E_{v,J}}{hc} = BJ(J+1) + \left(v + \frac{1}{2}\right)\tilde{\nu}_e - x_e\left(v + \frac{1}{2}\right)\tilde{\nu}_e \quad [40]$$

The selection rules for vibration–rotation transitions are the same as for separate transitions:

$$\Delta v = \pm 1, \pm 2, \pm 3, \dots$$

and

$$\Delta J = \pm 1 \quad [41]$$

Labeling the initial and final levels by double and single primes, respectively, and making the assumption that the rotational constants for the lower and upper vibrational states are the same, then the transition energies are given by

$$\tilde{\nu}_{J',J''} = \tilde{\nu}_0 + B(J' - J'')(J' + J'' + 1) \quad [42]$$

when  $\nu_0$  is the band center (or band origin) given by

$$\tilde{\nu}_0 = \tilde{\nu}_e(1 - 2x_e) \quad [43a]$$

$$\tilde{\nu}_0 = 2\tilde{\nu}_e(1 - 3x_e) \quad [43b]$$

$$\tilde{\nu}_0 = 3\tilde{\nu}_e(1 - 4x_e) \quad [43c]$$

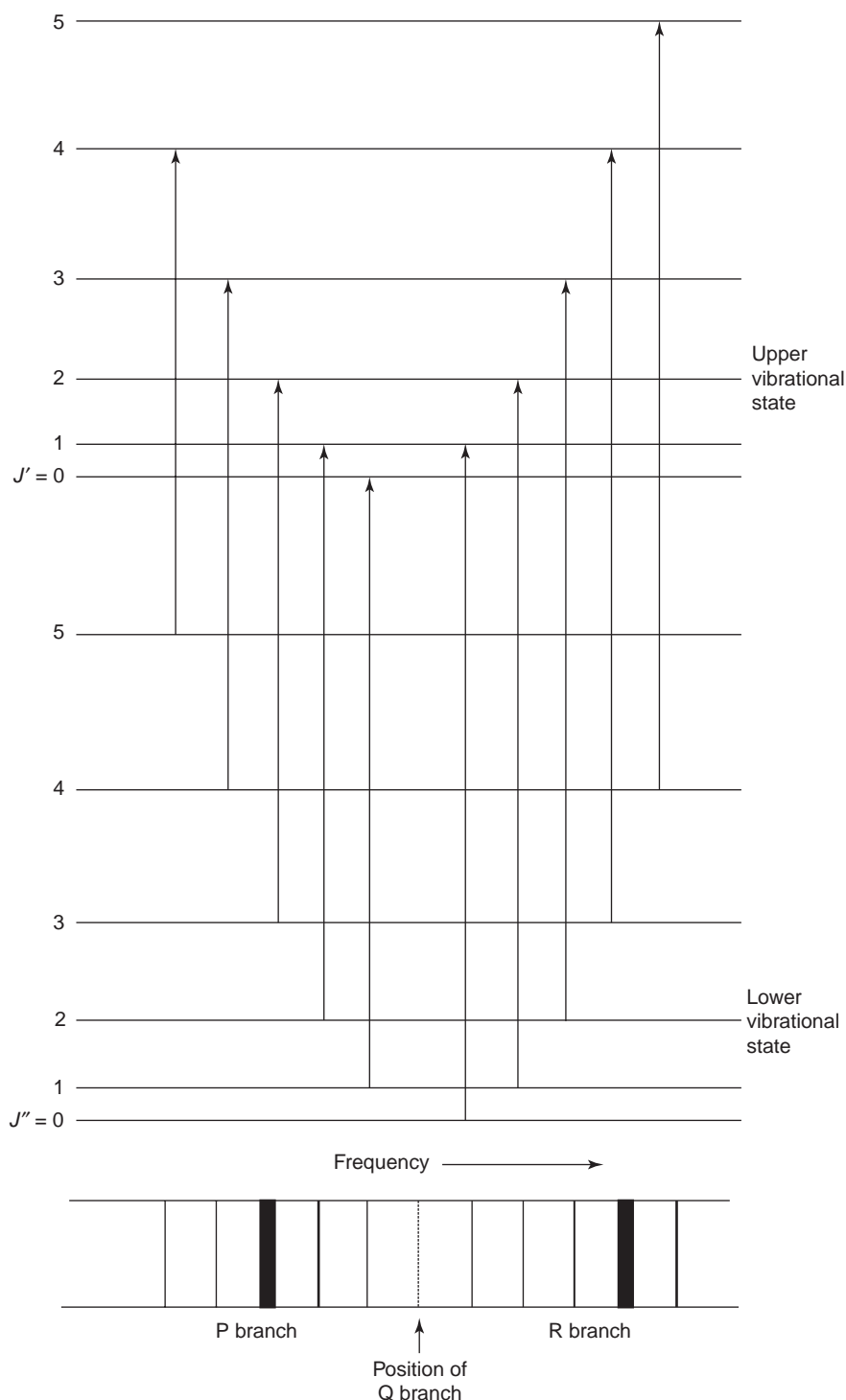
for the fundamental and first and second overtones, respectively. For  $\Delta J = +1$ ,

$$\tilde{\nu}_{J''+1,J''} = \tilde{\nu}_0 + 2B(J'' + 1), \quad J'' = 0, 1, 2, \dots \quad [44]$$

For  $\Delta J = -1$ ,

$$\tilde{\nu}_{J''-1,J''} = \tilde{\nu}_0 - 2B(J'' + 1), \quad J'' = 0, 1, 2, \dots \quad [45]$$

So it can be seen that the high-resolution spectrum will consist of two series of lines, one on either side of the band center. The series corresponding to  $\Delta J = -1$  is known as the P branch, and the series corresponding to  $\Delta J = +1$  is known as the R branch. Note that there is zero intensity at the line center. **Figure 5** shows a schematic diagram of the transitions involved and a stick representation of the line intensities which will be discussed below. **Figure 6** shows the high-resolution IR spectrum of the fundamental vibration of carbon monoxide and the same spectrum at lower resolution where only the envelope of the P and R branches can be seen.

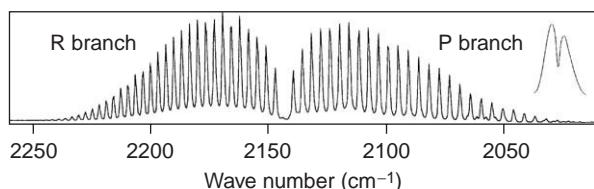


**Figure 5** Schematic diagram of P and R branch transitions for a diatomic molecule, where the line thickness compounds to the intensity of the transition.

It has been assumed that the rotational constant  $B$  is the same for the upper and lower vibrational states and that the vibrational terms will be unaffected by the rotational state (interaction between rotation and vibration, i.e., a breakdown of the Born–Oppenheimer approximation). When these assumptions

are not made, eqn [40] becomes

$$\frac{E_{v,J}}{hc} = B_v J(J+1) + \left(\nu + \frac{1}{2}\right) \tilde{\nu}_e + x_e \left(\nu + \frac{1}{2}\right)^2 \tilde{\nu}_e - D_v J^2 (J+1)^2 + \dots \quad [46]$$



**Figure 6** The high-resolution infrared spectrum of carbon monoxide showing P and R branches. (Inset on right is the same band at lower resolution.)

where  $B_v$  is the rotational constant associated with the vibrational level with quantum number  $v$  and  $D_v$  is the centrifugal distortion coefficient associated with the vibrational level  $v$ . Centrifugal distortion arises due to the fact that a bond will lengthen hence weaken as the molecule rotates. The resultant vibrational-rotational energy change is therefore

$$\tilde{\nu}_{v'J',v''J''} = \tilde{\nu}_0 + B_{v'}J'(J'+1) - D_{v'}J'^2(J'+1)^2 - B_{v''}J''(J''+1) + D_{v''}J''^2(J''+1)^2 \quad [47]$$

### Intensity of Lines in the P and R Branches

The intensity of each rotational line depends on the number of molecules occupying the initial rotational state. Using the Boltzmann distribution formula, at thermal equilibrium the ratio of the number of molecules  $N_J$  in rotational state  $J$  to the number  $N_0$  in the rotational ground state is given by

$$\frac{N_J}{N_0} = (2J+1) \exp\left[-\frac{BhcJ(J+1)}{kT}\right] \quad [48]$$

where the  $(2J+1)$ -fold degeneracy of each rotational state has been taken into account. By differentiating eqn [48] with respect to  $J$  and setting the derivative to zero it can be shown that the maximum population and hence the maximum intensity line occurs at

$$J_{\max} = \left(\frac{kT}{2Bhc}\right)^{1/2} - \frac{1}{2} \quad [49]$$

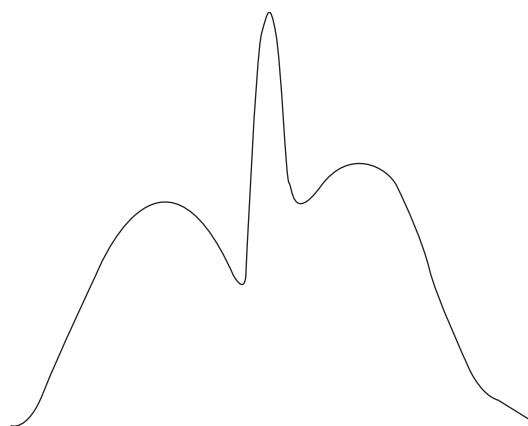
This corresponds to a maximum intensity at wave number  $\tilde{\nu}_{\max}$  given by

$$\tilde{\nu}_{\max} = \tilde{\nu}_0 \pm 2B \left[ \left(\frac{kT}{2Bhc}\right)^{1/2} + \frac{1}{2} \right] \quad [50]$$

where the  $+$  and  $-$  signs refer to the R and P branches, respectively.

### High-Resolution Vibrational Spectra of Linear Polyatomic Molecules

The normal vibrations of a linear polyatomic molecule which result in a change in the molecular dipole



**Figure 7** Typical PQR branch envelope for the perpendicular vibration of a linear polyatomic molecule.

moment (i.e., are IR allowed) can be classified into two types: parallel vibrations and perpendicular vibrations, for which the directions of the changes in the molecular dipole moment are parallel and perpendicular to the internuclear axis, respectively. The rotational selection rules for parallel vibrations are the same as for the vibration of a diatomic molecule and one observes P and Q branches as before. However, for perpendicular vibrations, rotational transitions are allowed in which the rotational quantum number does not change, i.e.,

$$\Delta J = 0 \quad [51]$$

If rotational constants in both vibrational states are equal then all Q branch transitions occur at the same wavelength, the band origin. For a fundamental vibration of a perpendicular vibration with unequal rotational constants in the two vibrational levels and introducing centrifugal distortion

$$\tilde{\nu}_{v'J',v''J''} = \tilde{\nu}_0 + (B_{v'} - B_{v''})J^2 + (B_{v'} - B_{v''})J + (D_{v'} - D_{v''})J^2 + (D_{v'} - D_{v''})J, \quad J = 0, 1, 2, \dots \quad [52]$$

A typical band contour for perpendicular bands is shown in Figure 7.

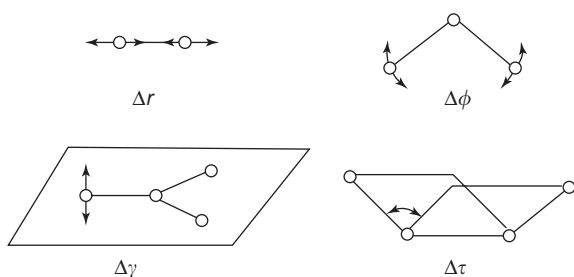
## Polyatomic Molecules

### Internal Coordinates

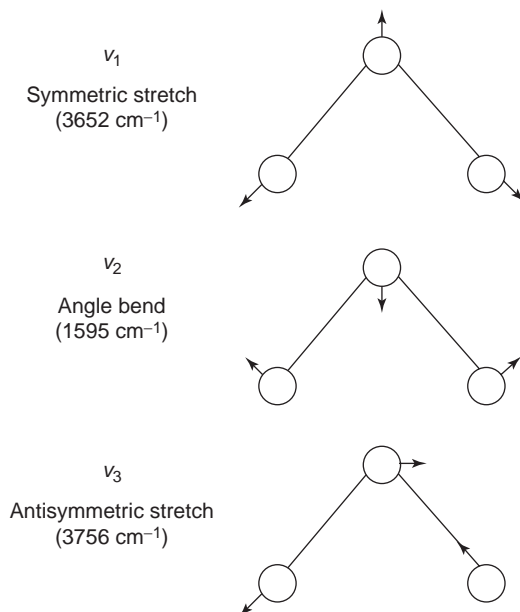
The set of internal coordinates required to describe the vibrational motion of a general polyatomic molecule consists of the bond stretch  $r$ , the bond angle bend  $\phi$ , the out-of-plane (o.o.p.) angle bend  $\gamma$ , and the bond torsion  $\tau$  (Figure 8).

In a polyatomic molecule with  $3N-6$  vibrational degrees of freedom ( $3N-5$  for a linear molecule) a set





**Figure 8** Internal coordinates for polyatomic molecules.



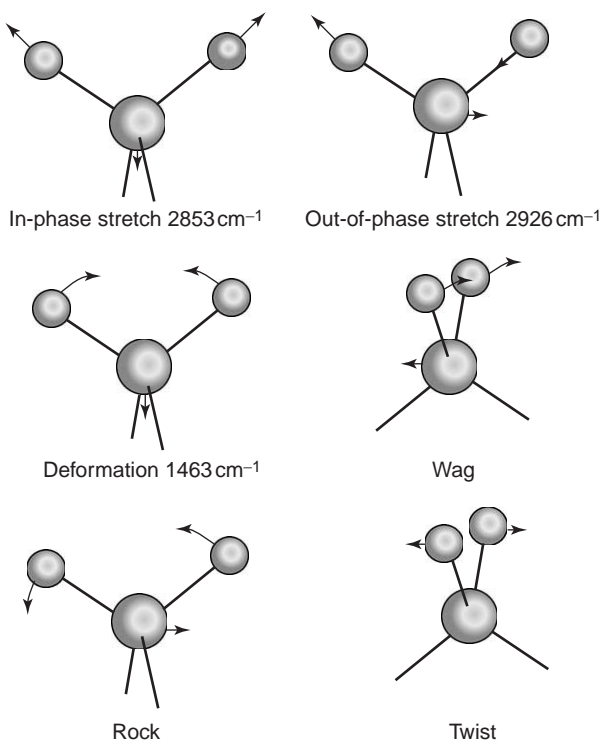
**Figure 9** The normal vibrations of  $\text{H}_2\text{O}$ .

of  $3N-6$  internal coordinates can be chosen to describe the molecular vibrations.

A normal coordinate for a polyatomic molecule can be expressed as a linear combination of the internal coordinates. The vibrational behavior of the atoms can be represented by attaching arrows to show their direction of motion. The lengths of the arrows are in proportion to the maximum amplitudes of each atom's normal coordinate excursion. The normal vibrations of water ( $\text{H}_2\text{O}$ ) are shown in Figure 9.

**Characteristic or group frequencies** With a knowledge of atomic masses, the molecular geometry and force constants, it is possible to calculate the internal coordinate composition of the normal vibrations of any molecule. Without this mathematical help, it is not possible to specify the origin of most of the bands in an IR spectrum which will, in general, contain major contributions from several internal coordinates.

However, it is found that certain structural units in a molecule give rise to bands that appear in the



**Figure 10** Characteristic vibrations of the methylene group.

spectra of different molecules within a sufficiently narrow range of frequencies for these bands to be used to identify the presence of the structural unit. These relatively constant bands are known as the characteristic frequencies or group frequencies of a molecule. The vibrations associated with methylene and methyl groups, which give rise to important group frequencies, are shown in Figures 10 and 11, respectively.

### Intensity of IR Transitions

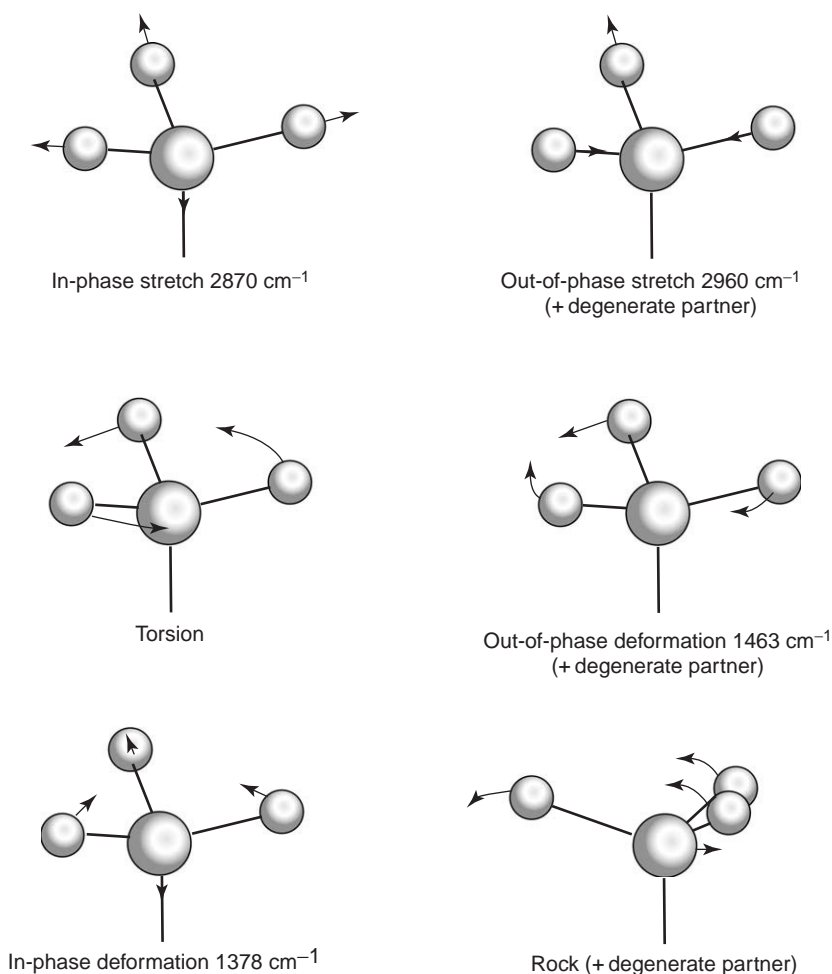
Let  $I_0$  and  $I$  be the incident and transmitted intensities, respectively, when infrared radiation passes through a sample of concentration  $C$  and cell length  $l$ . The transmittance  $T$  is defined as

$$T = \frac{I}{I_0} = 10^{-aCl} \quad [53]$$

where the quantity  $a$ , a function of wavelength, is called the absorptivity of the sample. Taking logarithms of the last equation, the absorbance of the sample is defined as

$$A = \log_{10} \left( \frac{I_0}{I} \right) = aCl \quad [54]$$

Equation [54] is known as the Beer-Lambert law: the absorbance of a sample is proportional to its concentration and the cell pathlength. Since  $A$  is a



**Figure 11** Characteristic vibrations of the methyl group.

dimensionless quantity, the units of  $a$  are the product of the units for reciprocal concentration and reciprocal pathlength. For example, with concentration in mol m<sup>-3</sup> and pathlength in meters, the units of  $a$  are m<sup>2</sup> mol<sup>-1</sup>. When concentration is expressed as molarity and the pathlength is either in meters or centimeters, the absorptivity  $a$  is known as the molar absorptivity and is given the symbol  $\epsilon$ .

The Beer-Lambert law is often used in quantitative IR analysis using peakheights in the absorbance spectrum as values for  $A$ , frequently after baseline correction. Peakheights, however, are strongly dependent on instrumental resolution and an alternative measure of the absorbance of a band is its integrated intensity, which is the intensity integrated over the whole IR band. The integrated intensity may be expressed using the integrated absorptivity  $\mathcal{A}$ , the absorptivity integrated over the whole IR band, given by

$$\mathcal{A} = \int_{\text{band}} a(\tilde{\nu}) d\tilde{\nu} \quad [55]$$

where  $a(\tilde{\nu})$  has been written to highlight the dependence of the absorptivity on wave number. Due to the integration with respect to the wave number the dimensions of the integrated absorptivity are that of the absorptivity divided by the dimension of length. Using eqn [54], the last equation may be expressed as

$$\mathcal{A} = \frac{1}{Cl} \int_{\text{band}} \log_{10} \left( \frac{I_0}{I} \right) d\tilde{\nu} \quad [56]$$

While the peakarea is a better measure of band intensity than peakheight in theory, the effect of absorption by neighboring bands leads to baseline errors that affect the calculation of area more adversely than peakheight. It is probably true to say that most contemporary quantitative determinations are made using peakheight.

The major interaction between the electromagnetic radiation and a molecule is due to the interaction of the electric field of the former,  $E$ , with the dipole moment of the latter,  $\mu$ . (Magnetic interactions are much smaller and generally are not important in

vibrational spectroscopy, although they are responsible for IR circular dichroism.) The interaction gives rise to a time-dependent perturbation of the quantum states of a molecule equal to  $-\mathbf{\mu} \cdot \mathbf{E}$ . Using the results of time-dependent perturbation theory (Fermi's Golden Rule) it can be shown that the integrated absorptivity of a transition between initial state  $\langle i|$  and final state  $\langle f|$  for an isotropic system is given by

$$\mathcal{A} = \frac{N_A 2\pi^2 \bar{\nu}_{if}}{3\epsilon_0 h c \ln_e 10} |\langle i|\mathbf{\mu}|f\rangle|^2 \quad [57]$$

where  $\langle i|$  and  $\langle f|$  are the initial and final states of the transition,  $\mathbf{\mu}$  is the molecular dipole moment operator,  $N_A$  is Avogadro's number,  $\bar{\nu}_{if}$  is the wave number corresponding to the band center,  $\epsilon_0$  is the permittivity of free space,  $h$  is Planck's constant, and  $c$  is the speed of light. The quantity

$$\langle i|\mathbf{\mu}|f\rangle \left[ \equiv \int \Psi_i^0 \mathbf{\mu} \Psi_f d\tau \right] \quad [58]$$

is known as the electric dipole transition moment and its magnitude determines the intensity of a transition.

The appearance of  $\ln_e 10$  in eqn [57] is to be consistent with the definition of absorbance using base-10 logs. Often the equation will be seen without divisor, in which case it should be noted that the absorptivity has been defined using the natural log scale.

It should be noted that eqn [57] strictly only applies to dilute gases. In condensed phases refractive index effects become important.

Equation [57] often appears in Gaussian units. The Gaussian version is obtained by replacing  $\epsilon_0$  by  $1/4\pi$  to give

$$\mathcal{A} = \frac{N_A 8\pi^3 \bar{\nu}_{if}}{3hc \ln_e 10} |\langle i|\mathbf{\mu}|f\rangle|^2 \quad [59]$$

The transition electric dipole moment in eqn [57] can be developed by invoking the Born–Oppenheimer approximation to express the total molecular wave function as a product of electronic and vibrational parts. (Rotational wave functions do not have to be included here since eqn [57] refers to an isotropic system. That is, the equation is a result of a rotational average which is equivalent to a summation over all the rotational states involved in the transition.) A general molecular state can now be expressed as the product of vibrational and electronic parts. Assuming that the initial and final electronic states are the ground state  $|e_g\rangle$ ,

$$|i\rangle = |i_{\text{vib}}\rangle |e_g\rangle \quad [60a]$$

$$|f\rangle = |f_{\text{vib}}\rangle |e_g\rangle \quad [60b]$$

and the transition moment becomes

$$\begin{aligned} \langle i|\mathbf{\mu}|f\rangle &= \langle v_i|\langle e_g|\mathbf{\mu}|e_g\rangle|v_f\rangle \\ &= \langle v_i|\mathbf{\mu}_e|v_f\rangle \end{aligned} \quad [61]$$

where  $|v_i\rangle$  and  $|v_f\rangle$  are the initial and final vibrational states, respectively, and  $\mathbf{\mu}_e$  is the permanent dipole moment of the molecule. The permanent dipole moment is now treated as a parametric function of the normal coordinates and expressed as a power series

$$\begin{aligned} \langle v_i|\mathbf{\mu}_0 + \sum_{p=1}^{3N-6} \left( \frac{\partial \mathbf{\mu}}{\partial Q_p} \right)_0 Q_p \\ + \frac{1}{2} \sum_{p=1}^{3N-6} \sum_{r=1}^{3N-6} \left( \frac{\partial^2 \mathbf{\mu}}{\partial Q_p \partial Q_r} \right)_0 Q_p Q_r + \dots |v_f\rangle \end{aligned} \quad [62]$$

Ignoring the quadratic terms (assuming the system is electrically harmonic) gives

$$\langle v_i|\mathbf{\mu}_e^0|v_f\rangle + \sum_{p=1}^{3N-6} \langle v_i|\left( \frac{\partial \mathbf{\mu}_e}{\partial Q_p} \right)_0 Q_p|v_f\rangle \quad [63]$$

The total vibrational wave function  $\langle v|$  is given by the product of the  $3N-6$  normal coordinate wave functions

$$\langle v| = \langle v_1|\langle v_2|\langle v_3|\dots\langle v_{3N-6}| \quad [64]$$

where  $\nu_p$  is the vibrational quantum number of the  $p$ th normal vibration.

The total vibrational energy is given by

$$\begin{aligned} G(\nu_1, \nu_2 \dots \nu_{3N-6}) &= \left( \nu_1 + \frac{1}{2} \right) \tilde{\nu}_1 + \left( \nu_2 + \frac{1}{2} \right) \tilde{\nu}_2 \\ &+ \dots + \left( \nu_{3N-6} + \frac{1}{2} \right) \tilde{\nu}_{3N-6} \end{aligned} \quad [65]$$

Let the vibrational ground state  $|0\rangle$  be represented by

$$|0\rangle = \langle 0_1|\langle 0_2|\langle 0_3|\dots\langle 0_{3N-6}| \quad [66]$$

and the state with the  $p$ th normal mode in the  $\nu_p = 1$  state be represented by

$$|1_p\rangle = \langle 0_1|\langle 0_2|\langle 0_3|\dots\langle 1_p|\dots\langle 0_{3N-6}| \quad [67]$$

Using the properties of harmonic oscillator wave functions (the Hermite polynomials) that

$$\langle 0|Q_p|1_r\rangle = \left( \frac{h}{8\pi^2\nu} \right)^{1/2} \delta_{pr} \quad [68]$$

where  $\delta_{pr}$  is the Kronecker delta (which is unity if  $p=r$  and zero otherwise), and

$$\langle 0|1_p\rangle = 0 \quad [69]$$

Eqn [63] becomes

$$\left(\frac{h}{8\pi^2\nu_p}\right)^{1/2}\left(\frac{\partial\mu}{\partial Q_p}\right)_0 \quad [70]$$

Substituting in eqn [57],

$$\mathcal{A} = \frac{N_A}{12\epsilon_0 c^2 \ln_e 10} \left| \left(\frac{\partial\mu}{\partial Q_p}\right)_0 \right|^2 \quad [71]$$

$$\left| \left(\frac{\partial\mu}{\partial Q_p}\right)_0 \right| = \left( \frac{12\epsilon_0 c^2 \ln_e 10}{N_A} \right)^{1/2} \mathcal{A}^{1/2} \quad [72]$$

Note that the sign of the derivative of the dipole moment cannot be determined directly by measuring the integrated absorbance of an IR band.

Equation [71] shows the origin of the gross selection rule that the dipole moment of a molecule must change in the course of a normal coordinate excursion for the vibration to absorb IR radiation. The transition moment in eqn [57] is only nonzero for the case where only one vibration is excited and for the situation in which the quantum number of the vibration involved changes by  $\pm 1$ . Hence the selection rule given earlier in eqn [17].

### Anharmonic Effects in Spectra of Polyatomic Molecules

**Combination and difference bands** Besides overtones, anharmonicity also leads to the appearance of combination bands and difference bands in the IR spectrum of a polyatomic molecule. In the harmonic case, only one vibration may be excited at a time (the transition dipole moment integral vanishes when the excited state is given by a product of more than one Hermite polynomial corresponding to different excited vibrations). This restriction is relaxed in the anharmonic case and one photon can simultaneously excite two different fundamentals. A weak band appears at a frequency approximately equal to the sum of the fundamentals involved. (Only approximately because the final state is a new one resulting from the anharmonic perturbation to the potential energy mixing the two excited state vibrational wave functions.)

A difference band is the result of a transition from an excited level of one normal vibration to a higher energy level of another vibration. The frequency of the difference band occurs at exactly the difference in the frequencies of direct transitions to the excited states involved from the vibrational ground state. ('Exactly' equal in this case because no new vibrational state is involved.)

Because difference bands originate from thermally populated excited states, they will be more frequently observed at lower frequencies and increase in intensity as the temperature is raised. (Transitions which occur from states other than the ground state are known as hot bands. They are generally weak in mid-IR spectra at room temperature due to vibrational energy gaps which are relatively large compared to  $kT$ ).

**Fermi resonance** If an overtone or combination transition occurs with nearly the same frequency as a fundamental transition of the same symmetry, then the anharmonic term in the potential function causes the two vibrations to interact or 'mix'. This is known as Fermi resonance. The extent of the mixing increase as the frequency difference decreases. The result is that the overtone or combination band acquires intensity through having some of the fundamental vibration mixed into it. Fermi resonance causes the two bands involved to split apart from the positions they would have occupied had no interaction occurred.

**Symmetry of molecular vibrations** Every normal coordinate of a molecule must transform according to an irreducible representation of the molecular point group. If the molecular geometry is known, then it is a routine matter to use the methods of group theory to deduce how many vibrations occur for each irreducible representation. The procedure is to assign three Cartesian displacement coordinates to each atom and to use the  $3N$  coordinates as basis functions for a  $3N \times 3N$  matrix representation of the point group. A reducible representation is then obtained by taking the trace of these matrices. This representation is then reduced to a sum of irreducible representations using

$$N_i = \frac{1}{h} \sum_{\hat{R}} \chi_i(\hat{R}) \chi_{\text{red}}(\hat{R}) \quad [73]$$

where  $N_i$  is the number of times that symmetry species  $i$  occurs,  $h$  is the order of the group,  $\chi_i(\hat{R})$  is the character associated with symmetry species  $i$  and symmetry operation  $\hat{R}$ , and  $\chi_{\text{red}}(\hat{R})$  is the character of the reducible representation associated with symmetry with symmetry operation  $\hat{R}$ .

One must then remove the irreducible representation which result from the three translational and three rotational degrees of freedom (two for a linear molecule). These can be identified from the character table of the molecular point group: the translational degrees of freedom transform as the functions  $x$ ,  $y$ , and  $z$  (denoted  $T_x$ ,  $T_y$ ,  $T_z$  or  $x$ ,  $y$ ,  $z$  in the character tables); and the rotational degrees of freedom transform as the components of an axial vector (denoted  $R_x$ ,  $R_y$ , and  $R_z$  in the character tables). The

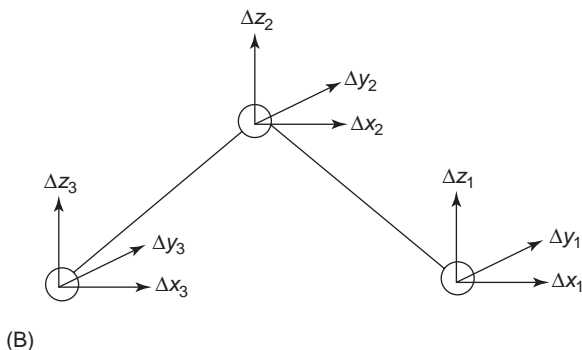
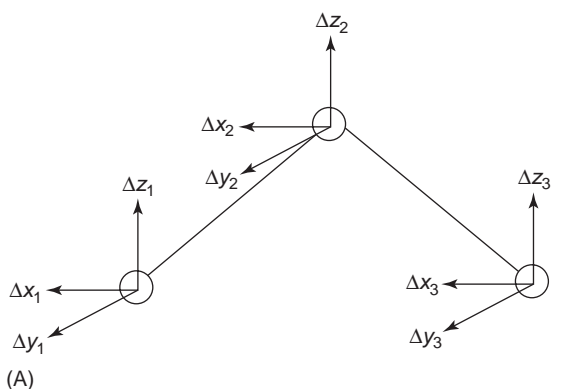
method is illustrated below for a bent triatomic molecule which belongs to the  $C_{2v}$  point group.

**Figure 12A** shows the molecule with three Cartesian coordinates associated with each atom. **Figure 12B** shows the effect of  $C_2$  rotation on the coordinates. This transformation can be described in matrix form as

$$\begin{array}{cccccccccccc}
 \Delta x_1 & -\Delta x_3 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & \Delta x_1 \\
 \Delta y_1 & -\Delta y_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & \Delta y_1 \\
 \Delta z_1 & \Delta z_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \Delta z_1 \\
 \Delta x_2 & -\Delta x_2 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & \Delta x_2 \\
 C_2 \Delta y_2 = -\Delta y_2 & = 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & \Delta y_2 \\
 \Delta z_2 & -\Delta z_2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \Delta z_2 \\
 \Delta x_3 & -\Delta x_1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \Delta x_3 \\
 \Delta y_3 & -\Delta y_1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \Delta y_3 \\
 \Delta z_3 & \Delta z_1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \Delta z_3
 \end{array} \quad [74]$$

From eqn [74] it can be seen that the trace of the  $C_2$  matrix  $\chi(C_2)$  is  $-1$ . Using the same procedure for the other symmetry operations, the complete reducible representation formed from the traces of the matrices corresponding to all the symmetry operations of the  $C_{2v}$  group is given by

$$\begin{array}{cccc}
 \chi_{\text{red}}(E) & \chi_{\text{red}}(C_2) & \chi_{\text{red}}(\sigma_v) & \chi_{\text{red}}(\sigma_{v'}) \\
 \Gamma_{\text{red}} & 9 & -1 & 1 & 3
 \end{array}$$



**Figure 12** (A) Cartesian displacement coordinates of symmetric triatomic molecules. (B) Cartesian coordinates after  $C_2$  rotation.

This representation can be reduced using eqn [73] and the information in  $C_{2v}$  character table. For example, the number of times  $N_{A_1}$  (the  $A_1$  irreducible representation) is contained in  $\Gamma_{\text{red}}$  is given by

$$N_{A_1} = \frac{1}{4}[(1)(9) + (1)(-1) + (1)(1) + (1)(3)] = 3$$

The same procedure can be carried out for each symmetry species of  $C_{2v}$  to give

$$\Gamma_{\text{red}} = 3A_1 + A_2 + 2B_1 + 3B_2 \quad [75]$$

From the  $C_{2v}$  character table, the three translations span  $A_1$ ,  $B_1$ , and  $B_2$ , and the three rotations span  $A_2$ ,  $B_1$ , and  $B_2$ . If these are taken away from eqn [75], this leaves  $2A_1$  and  $B_2$ . The three normal vibrations of the bent triatomic molecule, therefore, span these irreducible representations. (See **Figure 9**: the anti-symmetric stretch is  $B_2$ .)

This method may be simplified by noting that only Cartesian coordinates associated with atoms whose positions are unchanged by a symmetry operation may contribute to the trace of a matrix. If a symmetry operation  $\hat{R}$  leaves the position of  $U_{\hat{R}}$  atoms unchanged, then it can be shown that the character of the transformation matrix for not including transformations and rotations is given by

$$\chi(\hat{R}) = (U_{\hat{R}} - 2)(1 + 2 \cos \phi) \quad [76a]$$

for proper rotations, and

$$\chi(\hat{R}) = U_{\hat{R}}(-1 + 2 \cos \phi) \quad [76b]$$

for improper rotations (reflection, rotation–reflection, inversion) where  $\phi$  is the angle through which the molecule is rotated. (Inversion through the center of symmetry is equivalent to an improper rotation through  $180^\circ$ , a reflection in a plane of symmetry is an improper rotation through  $0^\circ$ , and the identity element is a proper rotation through  $0^\circ$ .)

For the molecule  $\text{ClCH}_3$ , which belongs to the point group  $C_{3v}$ , the number of atoms left unchanged



by the three symmetry operations  $I$ ,  $C_3$ , and  $\sigma_v$  is

$$U_I = 5, \quad U_{C_3} = 2, \quad U_{\sigma_v} = 3 \quad [77]$$

Using eqn [77] in eqn [76] gives:

$$\Gamma_{\text{red}} \begin{array}{ccc} \chi(I) & \chi(C_3) & \chi(\sigma_v) \\ 9 & 0 & 3 \end{array}$$

This representation can be reduced using eqn [73] to give  $3A_1 + 3E$ . Thus,  $\text{ClCH}_3$  ( $3N - 6 = 9$ ) has three totally symmetric vibrations and three doubly degenerate vibrations.

### Symmetry Selection Rules

Using eqn [61] for a fundamental vibration  $Q_p$  it can be seen that the band intensity is proportional to the following transition dipole moment integral:

$$\langle 0 | \mu_\alpha | 1_p \rangle, \quad \alpha = x, y, z \quad [78]$$

Unless the integrand is totally symmetrical the integral will be identically zero. Since the vibrational ground state is always totally symmetric, an IR fundamental will only be allowed when one or more components of the dipole moment operator span the same irreducible representation as the normal vibration. The components of the dipole moment operator  $\mu_x$ ,  $\mu_y$ , and  $\mu_z$  span the same irreducible representations as the functions  $x$ ,  $y$ , and  $z$ , respectively. Hence a fundamental vibration will only be allowed if it spans the same irreducible representations as  $x$ ,  $y$ , or  $z$  ( $T_x$ ,  $T_y$ , or  $T_z$ ).

It was shown before that a bent triatomic molecule undergoes two  $A_1$  and one  $B_2$  vibrations. Inspection of the  $C_{2v}$  character table reveals that  $z$  spans  $A_1$  and  $y$  spans  $B_2$ . Therefore, all the vibrations are allowed. This is not to say that all three bands will appear. The magnitude of the transitions dipole moments may be so small that a transition may not be observed.

If a molecule has several possible structures which belong to different molecular point groups then the methods above can be used for structure elucidation, especially in conjunction with the results of the analogous analysis for vibrational Raman bands. For example, if a molecule is known to have the molecular formula  $\text{AB}_4$ , a group theoretical analysis predicts that a tetrahedral molecule would have two active IR fundamentals, whereas a square planar molecule would have three.

### Infrared Spectroscopy of Crystals

In molecular crystals the molecules are held together by van der Waals forces, and since these bonds are

very much weaker than chemical bonds, the molecular vibrations are normally very similar to those of the free molecule. However, the crystal environment will generally lower the symmetry of the molecule, with the result that the degeneracy of vibrations may be lifted and vibrations, which were forbidden in the free molecule can become allowed in the crystal so that extra bands can appear. The formal treatment of the symmetry of vibrations in molecular crystals is obtained by considering the local symmetry in the crystal unit cell (site group analysis). A more complete theory, which includes lattice modes, is provided by factor group analysis.

Vibrations due to the crystal lattice occur in the far-IR from  $\sim 50$  to  $400 \text{ cm}^{-1}$ . It is possible to distinguish between some molecular and lattice vibrations using the fact that molecular vibrations are relatively insensitive to the effects of temperature and pressure while the frequencies of lattice vibrations generally increase with a decrease in temperature and with an increase in pressure.

### Infrared Linear Dichroism

For oriented single crystals there will generally be a difference in the absorption between two linearly polarized IR beams that are mutually orthogonal and orthogonal to the direction of propagations. The dichroic ratio is defined as

$$R = \frac{\int_{\text{band}} \varepsilon_{\parallel}(\tilde{\nu}) d\tilde{\nu}}{\int_{\text{band}} \varepsilon_{\perp}(\tilde{\nu}) d\tilde{\nu}} \quad [79]$$

where  $\varepsilon_{\parallel}$  and  $\varepsilon_{\perp}$  refer to polarization parallel and perpendicular to the crystal axis, respectively. If the symmetry of the crystal is known, then the dichroic ratio can give information about the symmetry of the vibration.

### Calculation of Normal Coordinates

Given the molecular geometry and a set of force constants for a polyatomic molecule, it is a routine matter to calculate the normal coordinates, a procedure known as normal coordinate analysis. Suites of computer programs are readily available that will calculate vibrational frequencies and the internal coordinate composition of each normal vibration. Most of the early calculation of vibration frequencies were made by Wilson's FG-matrix method, which is briefly summarized below. Today, a number of alternative techniques based on semiempirical methods, molecular mechanics, or density functional theory are also available, in convenient commercial software packages.

In the Wilson FG-matrix method, the problem is framed in internal coordinates rather than in

Cartesian displacement coordinates because the force constants involved are more meaningful in relation to the chemical structure of the molecule and are more readily transferred between similar molecules. Also, the theoretical procedure using internal coordinates are such that the translational and rotational motion of a molecule are automatically taken into account.

### The F Matrix

In the harmonic approximation the potential energy of a molecule can be expressed as

$$2V = \sum_{ij} F_{ij} R_i R_j = \tilde{\mathbf{R}} \mathbf{F} \mathbf{R} \quad [80]$$

where  $R_i$  are the internal coordinates,  $\mathbf{F}$  is the  $3N-6 \times 3N-6$  matrix formed by the force constants,  $\mathbf{R}$  is a column vector formed by the internal coordinates and  $\tilde{\mathbf{R}}$  is its transpose. (In order to have the same dimensions for all coordinates, and therefore all the force constants, the angle bending internal coordinates are sometimes scaled with a bond length, e.g., in water the angle bend coordinate  $\Delta\alpha$  would become  $r\Delta\alpha$  where  $r$  is the O-H bond length.)

Collectively, the  $F_{ij}$  are known as the force field of the molecule. These force constants are treated as empirical parameters whose values are optimized by obtaining the best fit of calculated to experimental results for vibrational frequencies, Coriolis coupling constants (which govern a type of coupling between rotational and vibrational motion), centrifugal distortion constants, and mean-square amplitudes of vibration. The simplest force field neglects all off-diagonal or interaction force constants. This valence force field (VFF), in general, gives poor results due to the poor number of adjustable parameters (the non-zero force constants).

In the generalized valence force field (GVFF) there is no neglect of the off-diagonal terms. However, for molecules of any appreciable size the number of force constants to be determined becomes too large to evaluate them with accuracy. Hence the simplified general valence force field (SGVFF) is frequently used in which all off diagonal force constants are set to zero except those involving two internal coordinates with common atoms.

The Urey-Bradley force field (UBFF) is also commonly used. This consists of diagonal stretch and bend force constants together with repulsive force constants representing nonbonded atom-atom interaction.

### The G Matrix

The kinetic energy part of the vibrational problem is expressed in the  $\mathbf{G}$  matrix whose elements depend

on atomic masses and molecular geometry. It can be shown that the vibrational kinetic energy  $T$  is given by

$$2T = \dot{\mathbf{R}} \mathbf{G}^{-1} \dot{\mathbf{R}} \quad [81]$$

where a dot denotes the time derivative.

The elements of the  $\mathbf{G}$  matrix are given in Figure 13.

As an example, the  $\mathbf{G}$  matrix for a nonlinear triatomic molecule (Figure 14) is given by

$$\mathbf{G} = \begin{bmatrix} \mu_1 + \mu_3 & \mu_3 \cos \phi & -\frac{\mu_3 \sin \phi}{r_2} \\ \mu_3 \cos \phi & \mu_2 + \mu_3 & -\frac{\mu_3 \sin \phi}{r_1} \\ -\frac{\mu_3 \sin \phi}{r_2} & -\frac{\mu_3 \sin \phi}{r_1} & \frac{\mu_1}{r_1^2} + \frac{\mu_2}{r_2^2} + \mu_3 \left( \frac{1}{r_1^2} + \frac{1}{r_2^2} - \frac{2 \cos \phi}{r_1 r_2} \right) \end{bmatrix} \quad [82]$$

(The exact form of the elements of  $\mathbf{G}$  depends on whether scaled or unscaled coordinates are being used. The above is for unscaled.)

### The Secular Equation

The relationship between internal coordinates and normal coordinates is defined as

$$\mathbf{R} = \mathbf{L} \mathbf{Q} \quad [83]$$

It can be shown that the matrix vibrational secular equation is given by

$$\mathbf{G} \mathbf{F} \mathbf{L} = \mathbf{L} \mathbf{A} \quad [84]$$

where  $\mathbf{A}$  is the diagonal eigenvalue matrix and  $\mathbf{L}$  is the matrix of eigenvectors of the matrix product  $\mathbf{G} \mathbf{F}$ . This last equation is solvable when

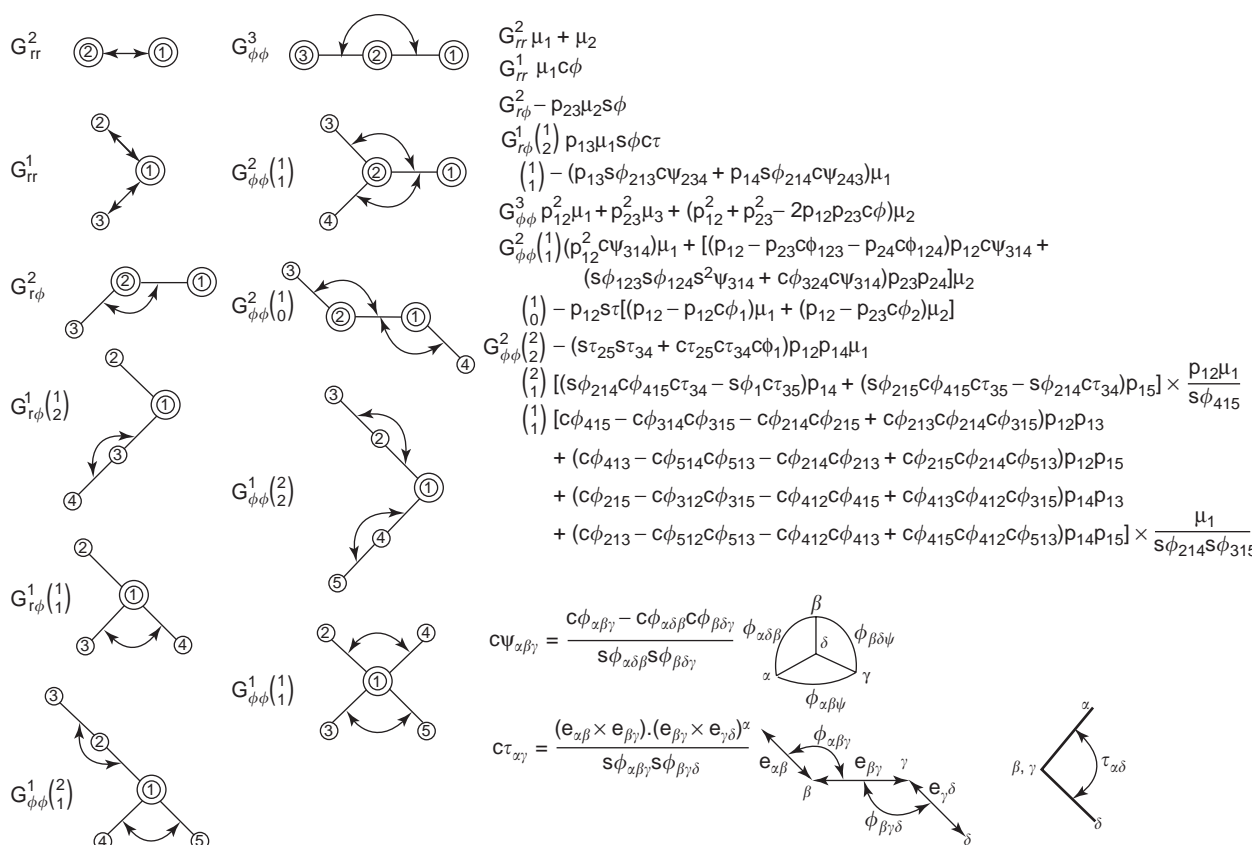
$$|\mathbf{G} \mathbf{F} - \mathbf{E} \lambda| = 0 \quad [85]$$

where  $\mathbf{E}$  is the unit matrix and  $\lambda$  is a root of the secular polynomial. There will be  $3N-6$  non-zero roots, which are equal to the squares of vibrational angular frequencies. So the problem is essentially the diagonalization of  $\mathbf{G} \mathbf{F}$ , a process which is easily carried out by computer using numerical methods.

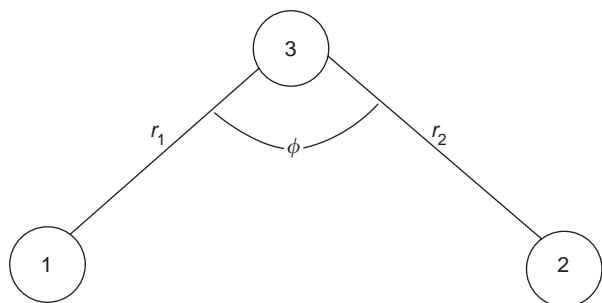
From eqn [83] we have

$$\mathbf{Q} = \mathbf{L}^{-1} \mathbf{R} \quad [86]$$

so that the elements of  $\mathbf{L}^{-1}$  can be used to obtain a picture of the normal vibration in terms of the internal coordinates. However, the reported results of a normal coordinate analysis often include the potential energy distribution (PED) for this information. The PED is the fraction of the potential energy



**Figure 13** Representation of common elements of the **G** matrix. Key: atoms common to both coordinates are double circles in horizontal line; number of common atoms as superscript; atoms above horizontal belong to first coordinate; those below belong second;  $\binom{n}{m}$  –  $n$  and  $m$  are numbers of (noncommon) atoms above horizontal on left and below horizontal on left, respectively.  $c$ ,  $\cos$ ;  $s$ ,  $\sin$ ;  $\rho_{\alpha\beta} = 1/r_{\alpha\beta}$ ;  $\mu_{\alpha} = 1/\text{mass } \alpha$ .



**Figure 14** Nonlinear triatomic molecule.

of a normal mode contributed by each force constant  $F_{ij}$ . The diagonal elements of this distribution for the major components of a normal vibration are quoted as a percentage. For vibration  $Q_p$  and internal coordinate  $R_i$

$$\text{PED}(R_i) = \left( \frac{100F_{ii}L_{ip}^2}{\sum_i F_{ii}L_{ip}^2} \right) \% \quad [87]$$

The sum of the diagonal elements in the PED can exceed 100% due to the neglect of the off-diagonal contributions.

### Use of Symmetry

When a molecule possesses symmetry the vibrational problem may be simplified by transforming the internal coordinates to symmetry coordinates. A vibration of a certain symmetry will be composed solely of symmetry coordinates belonging to the same symmetry species.

The transformation to symmetry coordinates  $\mathbf{R}^S$  is given by

$$\mathbf{R}^S = \mathbf{U}\mathbf{R} \quad [88]$$

The symmetry coordinates and hence the coefficients of the (unitary) symmetrization matrix are obtained by applying symmetry projection operators to the internal coordinates. For irreducible

representation  $i$

$$R_i^S = N \sum_{\hat{R}} \chi_i(\hat{R}) \hat{R} \times R \quad [89]$$

where  $N$  is a normalization factor.

A similarity transformation using  $U$  is carried out on the  $F$  and  $G$  matrices

$$F^S = UF\tilde{U} \quad [90a]$$

$$G^S = UG\tilde{U} \quad [90b]$$

The symmetrization process produces the block-factored matrices  $F^S$   $G^S$ . Hence the product  $G^S F^S$  will be block-factored and each block may be diagonalized separately.

### Isotopic Substitution in Polyatomic Molecules: The Teller–Redlich Product Rule

Observation of the changes in frequency that occur in a vibrational spectrum as a result of isotopic substitution of one or more atoms is an important method for assessing the accuracy of molecular force fields. Isotopic substitution is also important for making band assignments in large molecules: the only vibrations to be shifted will be those involving the isotopically substituted atoms.

For isotopic substitution in which the molecular point group is unchanged, the Teller–Redlich product rule links the two sets of vibrational frequencies. There is one product rule for each symmetry species of the molecule as follows

$$\frac{\Pi_V}{\Pi_{V'}} = \left\{ \left( \frac{M}{M'} \right)^t \left( \frac{I_x}{I'_x} \right)^{r_x} \left( \frac{I_y}{I'_y} \right)^{r_y} \left( \frac{I_z}{I'_z} \right)^{r_z} \Pi \left( \frac{m'}{m} \right)^a \right\}^{1/2} \quad [91]$$

where a prime is used to distinguish properties of the two molecules and  $\Pi$  denotes a product. The products on the left-hand side include all vibrations in the particular symmetry species to which the equation applies.  $M$  is the molecular mass;  $t$  is the number of

translations belonging to the symmetry species in question (which can be deduced using the methods described in the section on symmetry);  $I_x$ ,  $I_y$ , and  $I_z$  are moments of inertia about the three Cartesian axes;  $r_x$ ,  $r_y$ , and  $r_z$  are 1 if the respective rotation belongs to the symmetry species concerned and 0 otherwise;  $m$  is the mass of an atom which is a member of a set of symmetrically equivalent atoms and  $a$  is the number of external (rotational and translational) symmetry coordinates which these atoms give rise to. The product on the right-hand side involves all sets of symmetrically equivalent atoms.

*See also:* **Chemometrics and Statistics:** Optimization Strategies. **Chiroptical Analysis.** **Fourier Transform Techniques.** **Infrared Spectroscopy:** Sample Presentation; Near-Infrared. **Photoacoustic Spectroscopy.**

### Further Reading

- Bellamy LJ (1970) *The Infrared Spectra of Complex Molecules*, 3rd edn., vol. 1. London: Chapman and Hall.
- Bellamy LJ (1980) *The Infrared Spectra of Complex Molecules*, 2nd edn., vol. 2. London: Chapman and Hall.
- Bright Wilson E Jr, Decins JC, and Cross PC (1955) *Molecular Vibrations*. New York: McGraw-Hill.
- Chalmers JM and Griffiths PR (eds.) (2002) *Handbook of Vibrational Spectroscopy*. Chichester: Wiley.
- Colthup NB, Daley LH, and Wilberley SF (1990) *Introduction to Infrared and Raman Spectroscopy*, 3rd edn. London: Academic Press.
- Herzberg G (1945) *Infra-red and Raman Spectra of Polyatomic Molecules*. New York: Van Nostrand.
- Nakamoto N (1978) *Infrared and Raman Spectra of Inorganic and Coordination Compounds*, 3rd edn. New York: Wiley.
- Schrader B (ed.) (1995) *Infrared and Raman Spectroscopy*. New York: VCH.
- Woodward LA (1972) *Introduction to the Theory of Molecular Vibrations and Vibrational Spectroscopy*. Oxford: Oxford University Press.

## Sample Presentation

**J Chalmers**, VSConsulting, Stokesley, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Since an infrared spectrum can be recorded from almost any material, infrared spectroscopy is an

extremely important analytical technique. Mid-infrared spectroscopy is used extensively in applications involving qualitative analysis, providing either functional group or structural information about a sample or fingerprinting (identifying) a material. There is also widespread use of the technique for quantitative purposes, since the absorbance of a band is proportional to the concentration of the species that gives rise to the absorption band.

representation  $i$

$$R_i^S = N \sum_{\hat{R}} \chi_i(\hat{R}) \hat{R} \times R \quad [89]$$

where  $N$  is a normalization factor.

A similarity transformation using  $U$  is carried out on the  $F$  and  $G$  matrices

$$F^S = UF\tilde{U} \quad [90a]$$

$$G^S = UG\tilde{U} \quad [90b]$$

The symmetrization process produces the block-factored matrices  $F^S$   $G^S$ . Hence the product  $G^S F^S$  will be block-factored and each block may be diagonalized separately.

### Isotopic Substitution in Polyatomic Molecules: The Teller–Redlich Product Rule

Observation of the changes in frequency that occur in a vibrational spectrum as a result of isotopic substitution of one or more atoms is an important method for assessing the accuracy of molecular force fields. Isotopic substitution is also important for making band assignments in large molecules: the only vibrations to be shifted will be those involving the isotopically substituted atoms.

For isotopic substitution in which the molecular point group is unchanged, the Teller–Redlich product rule links the two sets of vibrational frequencies. There is one product rule for each symmetry species of the molecule as follows

$$\frac{\Pi_V}{\Pi_{V'}} = \left\{ \left( \frac{M}{M'} \right)^t \left( \frac{I_x}{I'_x} \right)^{r_x} \left( \frac{I_y}{I'_y} \right)^{r_y} \left( \frac{I_z}{I'_z} \right)^{r_z} \Pi \left( \frac{m'}{m} \right)^a \right\}^{1/2} \quad [91]$$

where a prime is used to distinguish properties of the two molecules and  $\Pi$  denotes a product. The products on the left-hand side include all vibrations in the particular symmetry species to which the equation applies.  $M$  is the molecular mass;  $t$  is the number of

translations belonging to the symmetry species in question (which can be deduced using the methods described in the section on symmetry);  $I_x$ ,  $I_y$ , and  $I_z$  are moments of inertia about the three Cartesian axes;  $r_x$ ,  $r_y$ , and  $r_z$  are 1 if the respective rotation belongs to the symmetry species concerned and 0 otherwise;  $m$  is the mass of an atom which is a member of a set of symmetrically equivalent atoms and  $a$  is the number of external (rotational and translational) symmetry coordinates which these atoms give rise to. The product on the right-hand side involves all sets of symmetrically equivalent atoms.

*See also:* **Chemometrics and Statistics:** Optimization Strategies. **Chiroptical Analysis.** **Fourier Transform Techniques.** **Infrared Spectroscopy:** Sample Presentation; Near-Infrared. **Photoacoustic Spectroscopy.**

### Further Reading

- Bellamy LJ (1970) *The Infrared Spectra of Complex Molecules*, 3rd edn., vol. 1. London: Chapman and Hall.
- Bellamy LJ (1980) *The Infrared Spectra of Complex Molecules*, 2nd edn., vol. 2. London: Chapman and Hall.
- Bright Wilson E Jr, Decins JC, and Cross PC (1955) *Molecular Vibrations*. New York: McGraw-Hill.
- Chalmers JM and Griffiths PR (eds.) (2002) *Handbook of Vibrational Spectroscopy*. Chichester: Wiley.
- Colthup NB, Daley LH, and Wilberley SF (1990) *Introduction to Infrared and Raman Spectroscopy*, 3rd edn. London: Academic Press.
- Herzberg G (1945) *Infra-red and Raman Spectra of Polyatomic Molecules*. New York: Van Nostrand.
- Nakamoto N (1978) *Infrared and Raman Spectra of Inorganic and Coordination Compounds*, 3rd edn. New York: Wiley.
- Schrader B (ed.) (1995) *Infrared and Raman Spectroscopy*. New York: VCH.
- Woodward LA (1972) *Introduction to the Theory of Molecular Vibrations and Vibrational Spectroscopy*. Oxford: Oxford University Press.

## Sample Presentation

**J Chalmers**, VSConsulting, Stokesley, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Since an infrared spectrum can be recorded from almost any material, infrared spectroscopy is an

extremely important analytical technique. Mid-infrared spectroscopy is used extensively in applications involving qualitative analysis, providing either functional group or structural information about a sample or fingerprinting (identifying) a material. There is also widespread use of the technique for quantitative purposes, since the absorbance of a band is proportional to the concentration of the species that gives rise to the absorption band.



This article covers the preparation and presentation methods, both traditional and the more recently developed, used in sample analyses by mid-infrared spectroscopy. It focuses more on the so-called macro-sampling techniques, and does not cover in detail some of the sample presentation methods, for example diamond-window compression cells (see example in Figure 7), more particularly used nowadays in studies made using an infrared microscope; many of the microsampling techniques are, however, merely adaptations of the macrosampling techniques.

A prime consideration in sample presentation has to be the purpose of the study. For instance in solids, one must decide whether the chemical structure of the solid is of singular importance, or whether its physical characteristics are of interest. If, for example, the requirement is to study polymorphism in drugs or minerals, or perhaps crystallinity or molecular orientation in polymers, then the integrity of these properties will need to be maintained throughout sample preparation, presentation, and measurement. Sample presentation methods that provide simple, rapid generic identification may prove inadequate for more detailed substructure investigations, such as compositional or conformational analyses.

The specimen or specimens analyzed must be representative of the sample in the context of the study purpose or problem definition, particularly if the sample is heterogeneous. Sampling techniques that are inherently surface sensitive may not yield spectra that are characteristic of the bulk of the sample. Quantitative precision is likely to impose greater restrictions on the choice of the sample preparation method than simple generic fingerprinting studies, since reproducibility will be an essential requirement for most product quality assurance applications.

## Sample Presentation and Preparation Techniques

The procedures, requirements, and some limitations of the various presentation methods as practiced in the laboratory are considered in detail in this section. Samples are categorized as gases, liquids, paste-like, or solids.

### Gases

In the laboratory, the mid-infrared absorption spectra of gases are usually measured in gas-tight transmission cells. Gas cells are essentially gas-tight containers fitted with: (1) infrared transparent windows to enable the radiation to enter and exit the container, and (2) a means for introducing, pressurizing, and

evacuating the gas. They vary from a cylinder of a few centimeters pathlength, typically 10 cm and constructed from Pyrex<sup>®</sup> or stainless steel, with windows at each end, to compact long pathlength cells which have internal gold-coated mirrors for multipassing the radiation over many meters pathlength through the gas sample (see Figure 1).

Beer's law for a gas may be represented as

$$A = kl\rho$$

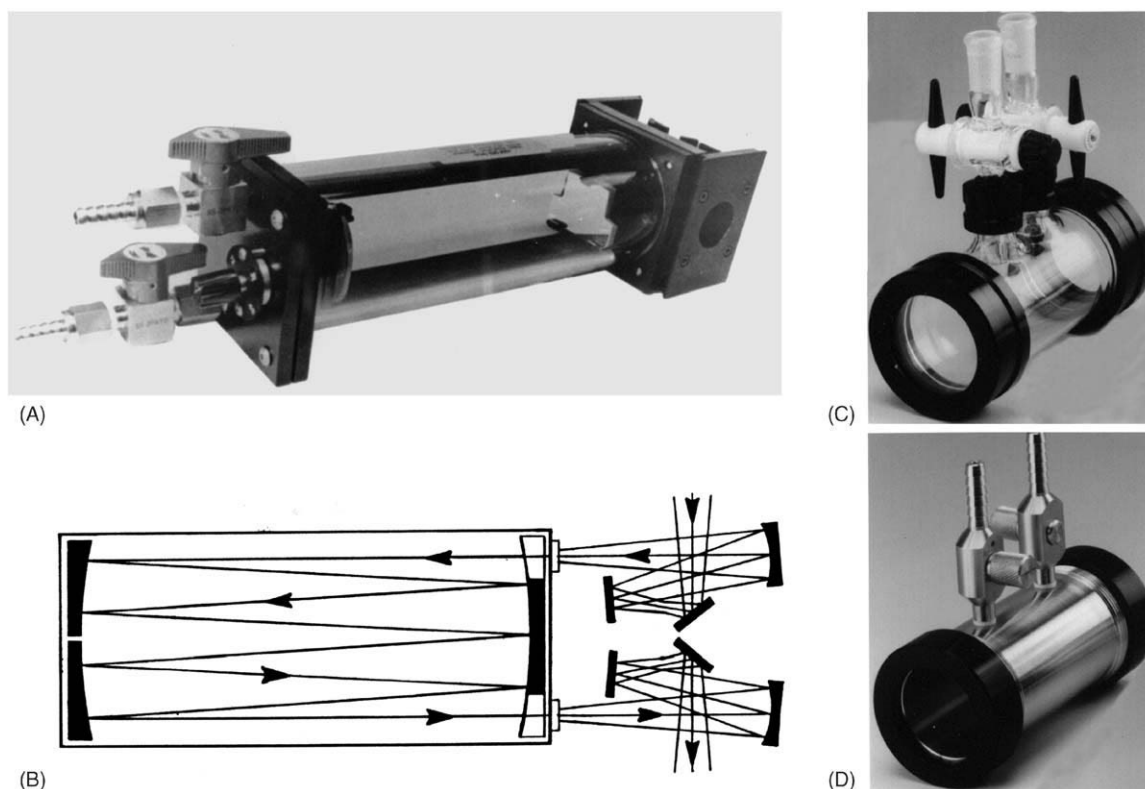
where,  $A$  is the absorbance,  $k$  the absorptivity,  $l$  the pathlength, and  $\rho$ , which replaces the concentration in the usual Beer's law notation, is the pressure, or partial pressure of the gas in a mixture of gases.

The absorbance of a component in a gas mixture depends not only on its partial pressure, but is also a function of the total pressure. Molecular collisions will broaden the rotational fine structure of the gas-phase bands, a phenomenon known as pressure broadening. So, although 50 mm Hg or less is a pressure at which many gases in a 10-cm pathlength cell yield useful spectra, a common practice is to keep the total pressure constant. This is achieved by adding a nonabsorbing gas such as nitrogen up to a standard pressure, such as 760 mm Hg.

### Liquids

Liquids may be sampled as neat liquids or in solution. A mid-infrared transmission spectrum sufficient for chemical identification may often be recorded from a capillary layer of a nonvolatile, pure liquid. This may be prepared simply from a drop of the liquid that has been sandwiched between a pair of mid-infrared transparent windows clamped together, which is also resistant to attack by the liquid. A more reproducible (and safer) practice, however, is to use an appropriate pathlength cell. Whichever method is selected, the specimen examined must be free from bubbles. For strongly absorbing liquids and some quantitative applications, a more efficient approach may be to use an appropriate infrared internal reflection technique accessory.

A liquid sample, either neat or as a solution, may be examined contained within a sealed or demountable cell. Cells are essential for transmission measurements of volatile liquids and samples in solution, and imperative for the majority of quantitative determinations. While contained gases are usually examined in cells of several centimeters to several meters pathlength, liquid and solid samples generally require a very much shorter thickness. Cells are available commercially in two types (see Figure 2). Fixed pathlength cells provide the optimal means of examining reproducibly a series of liquid samples



**Figure 1** Examples of gas cells for mid-infrared transmission measurements: (A) photograph of a multiple-pass gas cell ('Long Path Minicell'), with a high path-to-volume (530 ml) ratio. Allows paths from 1.2 m (eight passes) to 7.2 m (48 passes); (B) schematic of a multipass cell with transfer optics for use in a center-focus sample compartment; (C) and (D) photographs of Pyrex<sup>®</sup> and stainless steel bodied 10-cm pathlength cells, respectively. ((A and B) Reproduced by kind permission of Infrared Analysis, Inc., Anaheim CA, USA. (C and D) Reproduced by kind permission of Specac Ltd., Orpington, Kent, UK.)

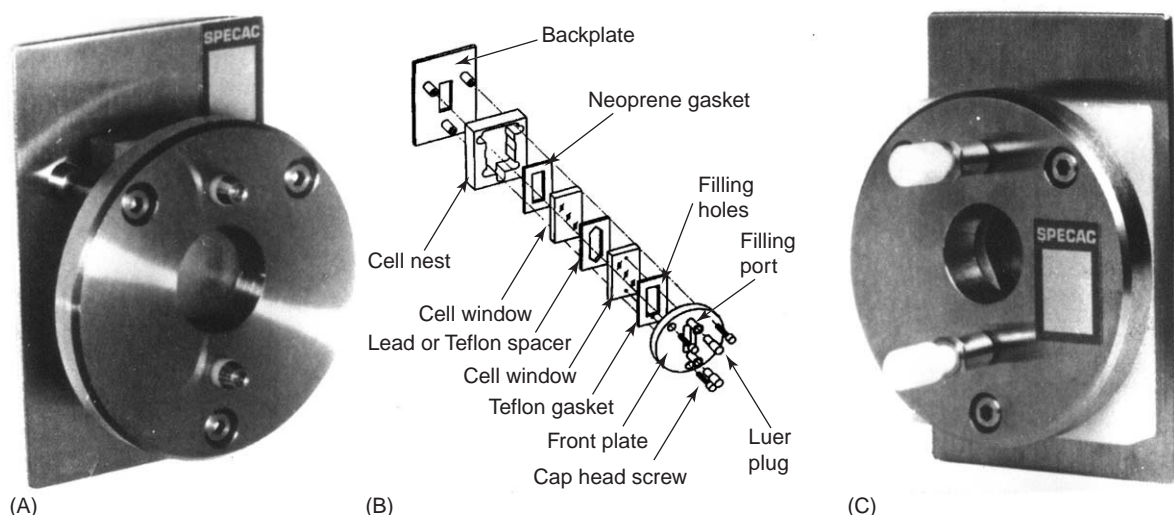
that are not too viscous, or corrosive or damaging to the cell windows or body. A pathlength of 0.1 mm will provide a sample thickness that is useful for observing the spectral detail of many neat nonpolar liquids. However, for polar molecules, if all the absorption bands are to be surveyed, it will usually be necessary to significantly reduce the pathlength to less than 0.025 mm, and in some instances to less than 0.01 mm. A typical macrosample cell of pathlength 0.1 mm will require about  $\sim 0.3$  ml to fill. Cavity cells of volumes as low as  $0.2 \mu\text{l}$  are available commercially for microsampling, although these are normally used in combination with a beam-condensing optical arrangement. A range of window materials is employed according to their infrared transmission characteristics and the properties of the liquid sample. Among the most common for mid-infrared are KBr, NaCl, CsI, and ZnSe, with BaF<sub>2</sub> and CaF<sub>2</sub> being frequently used because of their mechanical strength and insolubility in water, although these offer reduced spectral coverage. Polyethylene and TPX<sup>®</sup> (poly(4-methyl pent-1-ene)) are

frequently used window materials for the far-infrared region.

Maintaining and cleaning thin cells can be troublesome; in particular traces of viscous liquids may prove extremely difficult to remove. In these circumstances, a demountable cell (see Figure 2), which can be dismantled for cleaning, may be preferable. A range of spacers of differing uniform thickness is available to set the cell pathlength. For a cell that has flat, polished, and parallel windows its pathlength may be determined from a recording of the interference fringe pattern generated by the empty cell. Constructive and destructive interference between the primary transmitted beam and that which has been back reflected from within the cell results in a sinusoidal variation of transmitted intensity. A cell's pathlength (thickness)  $d$ , is calculated from

$$d = \frac{n}{2(v_1 - v_2)}$$

where,  $n$  is the number of complete fringes between wave numbers  $v_1$  and  $v_2$ .



**Figure 2** Examples of liquid cells for mid-infrared transmission measurements: (A) sealed vacuum tight liquid cell with an amalgam spacer for FT-IR, (B) exploded view of a liquid cell design with a rectangular aperture for use with a dispersive infrared spectrometer, and (C) semipermanent (dismountable) liquid cell for FT-IR with Teflon<sup>®</sup> spacer, and fitted with a Luer fitting and Teflon<sup>®</sup> plugs. (Reproduced by kind permission of Specac Ltd., Orpington, Kent, UK.)

Some of the problems associated with routine use of thin transmission cells may be circumvented through use of internal reflection accessories.

In mid-infrared internal reflection spectroscopy, often called attenuated total reflection (ATR) spectroscopy, the surface of a sample is placed in intimate contact with a higher refractive index, infrared transparent internal reflection element, and its spectrum recorded with infrared radiation that is incident through the reflection element at an angle greater than the critical angle. Common internal reflection elements are made from ZnSe, Ge, and diamond. The infrared radiation incident through the reflection element is totally internally reflected at the reflection element-sample boundary; the infrared beam may, however, be considered to penetrate a short distance into the sample and be attenuated by the absorption characteristics of the surface layer in contact with the reflection element; the penetration depth of the mid-infrared radiation into the sample is typically of the order of 0.3–3  $\mu\text{m}$ , which is wavelength dependent, and also dependent on the angle of incidence and the refractive indexes of both the reflection element material and the sample. Common designs for the analysis of liquids using Fourier transform infrared (FT-IR) spectroscopy come in essentially two types; those commonly referred to as horizontal ATR (H-ATR) units (see Figure 3), and those based on rod- or cylinder-like reflection elements, which are incorporated either into a dip-in probe or mounted through a chamber to contain the liquid. These units



**Figure 3** Photograph of horizontal (overhead) multiple internal reflection accessory. (Reproduced by kind permission of Specac Ltd., Orpington, Kent, UK.)

provide effective short constant pathlength sampling, typically  $\sim 25 \mu\text{m}$ , and are readily filled and easily cleaned; although, they may require more material than for an equivalent pathlength transmission cell.

Also available for qualitative transmission examination of nonvolatile liquids are microporous polymer film substrates mounted in a standard size sample-holder card; the polymer substrates are usually

either polytetrafluoroethylene (PTFE) or polyethylene into which the liquid is adsorbed. The aperture area of the card in which the substrate is mounted typically has a diameter of 19 mm. For solution samples, where appropriate, the solvent may be allowed to evaporate.

### Pastes, Dispersions, Lattices

Both, liquid and solid sample presentation techniques are used for studying these systems.

Many pastes are most easily directly sampled by internal reflection techniques.

More recent multiple internal reflection accessories, such as the H-ATR (see the section 'Liquids'), have enabled mid-infrared measurements to be made routinely and practically on a wide range of paste-like and latex-like samples and soft waxes. With dispersions, though, there may be a tendency for the solid phase to migrate preferentially to the surface of the internal reflection element. In some circumstances a capillary layer of the material between a pair of transparent infrared windows may suffice. The solute phase of a latex sample is most frequently presented for study as a thin film deposit on a transparent infrared support after elimination (evaporation) of the solvent (see the section 'Films cast from solution').

### Solids

Solid samples not only come in many diverse physical forms (powders, granules, fibers, films, sheet, biological tissue, etc.), but they may also be presented for examination in a wide variety of ways, and analyzed by a range of techniques that include both transmission and reflection methods. In the infrared spectroscopic examination of solids, due attention and consideration must be given to a number of factors.

Although, the infrared absorption characteristics of a solid material may be observed by many sampling techniques, the choice of the sample presentation method will likely be limited by both the study intent and by the sample's physical properties, complexity, and availability. For instance: Can the sample be readily ground to a fine-particle size?; Will the sample thermally degrade if heated?; Does the sample have a composite structure?; Is there a necessity for microsampling or information at high spatial resolution?; Is there a risk of inducing a change in polymorph? The consequences of any sample preparation technique on the sample's physical and chemical characteristics should be carefully considered prior to any examination. Similarly, the influences on the recorded spectra of the presentation method must be fully appreciated.

**Transmission methods** The traditional and most conventional solid sample presentation techniques for mid-infrared spectroscopy are those based on transmission measurements. Also, solids present the greatest choice and consideration in sample presentation methods for transmission measurements, since as discussed above the approach will vary with the nature of the solid and the analytical problem. Solids may be dissolved and examined in solution, or prepared as solid suspensions dispersed in a suitable fluid (the mull method) or a disk of a transparent medium (e.g., KBr). These suspension methods are best suited to crystalline, brittle, or easily ground solids, and are widely applicable to both monomeric organic and inorganic substances. Preparing a thin film from solution or solidification from the melt are methods well suited to examining amorphous materials, such as waxes and soft resins. Organic polymers will often yield self-supporting coherent films when compression molded from the melt, and this is a convenient procedure for many thermoplastic resins. Polymer samples suitable for presentation for transmission measurements may also be prepared by casting a thin film from solution, and in appropriate cases by microtoming sections, cold rolling, or pressing.

In many circumstances thin films or solutions provide the more direct, robust, and reproducible quantitative methods for single component analyses, which will operate over a wide dynamic range and at high sensitivity.

**Solutions** Solids may be examined in solution in transmission cells (see the section 'Liquids'), provided a suitable solvent is available. This can be an effective way of eliminating crystalline lattice effects. However, since no solvent is transparent throughout the entire infrared region, several solvents may be required for complete solute detail, or alternatively an analytically useful difference spectrum may be generated from a single solution, provided solvent-solute interactions are minimal. Solutions offer an important presentation method for quantitative applications, although many of the best solvents are toxic and flammable and must be used with appropriate care, safety, and caution.

**Films cast from solution** Thin films well suited to qualitative analyses, or quantitative measurements if an internal reference band for intensity normalization can be utilized, may often be prepared by casting a thin film from a solution of the sample and evaporating off the solvent. The film may be cast onto an infrared transparent window or onto a suitable support from which it can be readily peeled.



This technique of sample presentation is best suited to volatile solvents, and the sample preparation should be undertaken in a fume-cupboard or other appropriate containment environment. The window, for instance KBr, is placed on a hot-plate, which is then turned on and an appropriate temperature selected. This is normally quite hot to the touch, but not too hot, else the window may crack. Alternatively, the window may be warmed under an infrared lamp. A few drops of the solution are placed on the top surface of the window and gently spread over the entire surface with the tip of the pipette, and the solvent is allowed to evaporate. Successive additions of solution may be used to increase the film thickness. When all the solvent has evaporated and a film of the required thickness has been made, the film and window are allowed to cool to room temperature prior to infrared examination. Persistent solvent residues may need to be removed in a vacuum oven. Spectral contrast will be poor if the variation of the film thickness is large. The film should be uniform in thickness and cover the surface of the window, or be suitably masked, so that no stray-light reaches the infrared detector.

The solvent cast film can prove a particularly useful thin film presentation procedure for many thermoplastic polymers and clearly for solvent-based samples, such as paints, varnishes, and latexes. Though, if the sample crystallizes to any significant extent on evaporation of the solvent, then a highly scattering film may result, which proves unsuitable for transmission studies. AgCl and ZnS are useful supports for films cast from aqueous systems. The casting of thin films from some polyamides (e.g., nylons) dissolved in formic acid onto a glass microscope slide is a well-used example of casting onto a support other than an infrared transmitting window. After evaporation of the formic acid, the sample is water-washed, wiped dry, and then peeled from the glass slide and mounted in an appropriate holder for sample presentation.

Spin coating is a useful method for laying down onto an infrared transparent window a uniform thin layer of a few micrometers thickness of a polymer from a solution. A film can conveniently be cast from a polymer that is soluble in a convenient solvent at a concentration of  $\sim 1\text{--}2\%$ . An example of a spin-coater is shown in **Figure 4**; a clean 13 mm diameter window is attached to the sample stub, onto which several drops of the solution are laid down from a pipette, after which the rotor is spun for a few seconds. Thin films on gold-coated disks suitable for reflection-absorption measurements (see the section 'Thin films on metallic substrates'), may also be prepared by the spin-coating procedure.

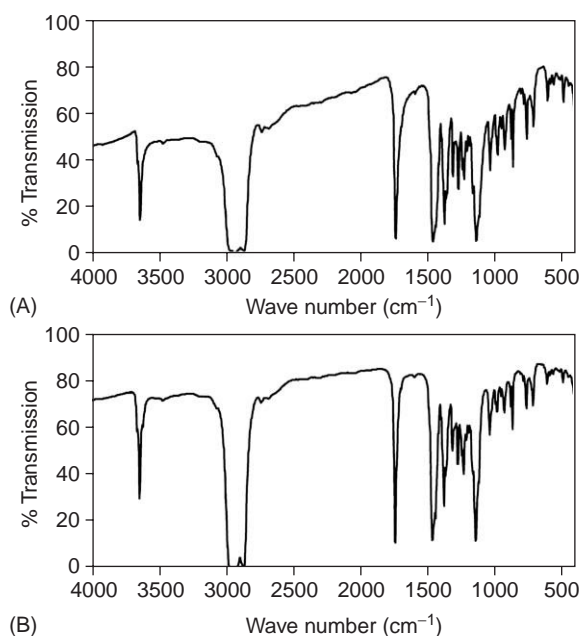


**Figure 4** Photograph of a spin-coater for preparing thin films suitable for mid-infrared studies. (The pen is included to give an indication of size.) (Reproduced by kind permission of Cordell Group, Research Engineering, Wilton Centre, Wilton, UK.)

*The mull technique* The basis of this approach is to fill the interstices between the particles of a finely powdered solid with an infrared transparent (or semitransparent) liquid medium. This mixture is then sandwiched into a thin film between a pair of infrared transparent windows. The mull, which should have the consistency of a paste, should be formed in such a manner as to minimize radiation scattering effects.

Ideally, the matrix medium (mulling agent) should have a refractive index matched or close to that of the solid dispersed in it. Also, the solid should be as finely ground as possible, but preferably to a particle size below that of the wavelength being used to study the sample in order to minimize spectral artifacts arising from scatter and anomalous dispersion. Scattering manifests essentially as two effects on the transmission spectrum: a 'sloping background' and absorption band distortion (see **Figure 5**). The spectrum of a mull prepared from a sample that is insufficiently ground will exhibit a gradual decrease in the background transmission from low to high wave numbers, since radiation scatter from particles increases toward shorter wavelengths, and is worse when there is greater mismatch between the refractive index of the dispersant and surrounding medium. In addition, because the refractive index of a sample changes in the vicinity of an absorption band, becoming less than the average to the high wavenumber side of the absorption maximum, and higher than the average to the low wavenumber side, then unsymmetrical scattering loss will be superimposed on the absorption band resulting in the appearance of a distorted band contour. This effect, a consequence of the





**Figure 5** Transmission mid-infrared spectra of liquid paraffin (Nujol<sup>®</sup>) mulls of a phenolic antioxidant for polymers, Irganox 1010 (Ciba-Geigy): (A) poorly ground sample and (B) well-ground sample.

anomalous dispersion in the refractive index, is known as the Christiansen effect (see **Figure 5**).

Mineral oil (liquid paraffin, Nujol<sup>®</sup>) is the most widely used mulling agent. However, this will most likely obscure detail in the C–H stretching ( $3000\text{--}2800\text{ cm}^{-1}$ ) and deformation ( $1500\text{--}1340\text{ cm}^{-1}$ ) regions. In this case, a second mull may need to be prepared, using a halogenated-oil mulling agent such as a perfluorohydrocarbon (Fluorolube<sup>®</sup>) or hexachlorobutadiene. The joint presentations then allow almost total unhindered observation of a sample's mid-infrared absorption bands.

The procedure to prepare a mull is typically to place 10–20 mg of the sample in an agate or mullite mortar, which is then ground to a fine-particle size powder using a vigorous rotary motion of the pestle. A small drop of the mulling agent is then added to the mortar, and rotary motion of the pestle is used to mix the components into a uniform paste. (A common fault is to add initially too much of the mulling agent.) The paste is then transferred to the center of a clean infrared transparent window (e.g., KBr, NaCl, CsI). A second window is then placed on top of the mull and the sandwich squeezed to form a thin film of the mull that is free from bubbles, which often looks translucent. The sandwich may be clamped together in a mull cell.

The mulling technique is not recommended as a general method for presenting samples for quantitative analysis, mainly due to problems of reproducible

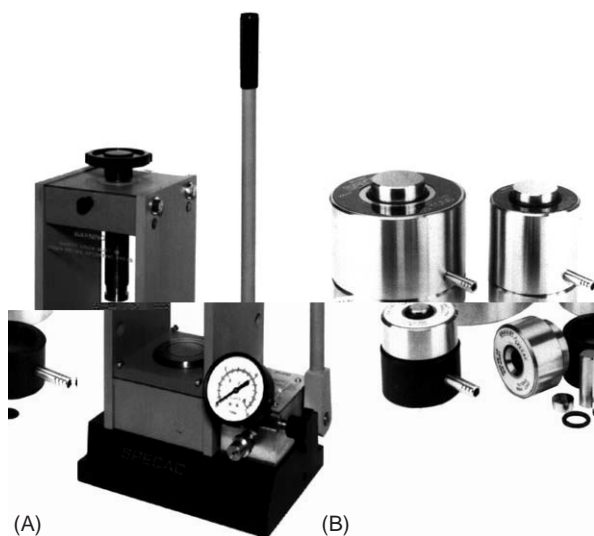
grinding; although, if special care is taken, successful semiquantitative analyses may be undertaken. These will necessitate use of a ratioing method to normalize band intensities, or the addition to the mull of an internal standard, although reproducible grinding and uniform mixing of this will require careful attention.

*The alkali-halide disk method* The alkali-halide disk technique is another traditional transmission sample presentation method, in which an intimate mixture of the finely ground solid sample and dry powdered alkali-halide, most commonly KBr, is pressed into a self-supporting disk. Finely powdered dry potassium bromide will coalesce to form a clear disk with high transmission when it is pressed under high pressure in an evacuated die.

The weight ratio of alkali-halide to analyte is large, typically of the order of 200–100:1. As with the mulling technique described in the section 'The mull technique' above, in order to reduce scattering loss and minimize absorption band distortions, the particle size of the sample must be reduced to well below that of the shortest infrared wavelength being used. Infrared spectroscopic-grade quality potassium bromide with a particle size of 100–200 mesh ( $\sim 100\text{ }\mu\text{m}$ ) may be obtained commercially. This should be dried overnight in a vacuum oven at  $\sim 105^\circ\text{C}$ , and then stored in a desiccator. It is important that the KBr is dry in order to eliminate or minimize the amount of adsorbed water in the KBr, which if present will reveal itself in the spectrum as a broad absorption centered  $\sim 3400\text{ cm}^{-1}$  together with a weaker band  $\sim 1640\text{ cm}^{-1}$ . The oven and desiccator should be kept free from other materials to avoid contamination of the KBr. Commercial dies in a range of diameters are available for the preparation of alkali-halide and similar disks, including minidisks (1 mm diameter), although the most commonly used is 13 mm (see **Figure 6**).

Although with experience, preferred local procedures will be developed and become standard practice, these should follow the general principles and the majority of steps outlined below, which is described for the preparation of a 13-mm diameter disk. The manufacturer's recommended procedures should be followed for the operation of the disk-making accessory and the press.

An optimal procedure, recommended by many spectroscopists, requires pregrinding of the sample, prior to mixing of the analyte and KBr to a homogenous mixture, which is subsequently used to make the disk. The pregrinding of the sample is usually achieved by one of the two methods. About 10–20 mg of a sample is ground in a mullite or agate



**Figure 6** Example of alkali-halide disk-making apparatus: (A) photograph of a manually operated hydraulic press for preparing alkali-halide disks and (B) photograph of a range of evacuable pellet dies. (Reproduced by kind permission of Specac Ltd., Orpington, Kent, UK.)

mortar to a fine-particle size, in the manner described for preparing the analyte for a mull in the section 'The mull technique'. An alternative approach, favored by some laboratories, is to use a vibrating agate or steel ball mill. A few milligrams of the sample is placed inside a clean vial together with the balls and ground for a set period of time recommended by the mill manufacturer, typically 30 s. One to three milligrams of the ground powder is then weighed and transferred to a clean mortar, or a clean vial. A weighed amount of dry KBr powder ( $300 \pm 5$  mg) is then added and mixed gently with the analyte to form a homogenous mixture. In a mortar this mixing can be achieved by initially adding  $\sim 15$  mg of the KBr and gently mixing with the pestle, followed by successively adding doubled quantities of the KBr with interleaved mixing. The pestle action should induce good mixing and minimal grinding, since further reduction of the KBr particle size will lead to increased water absorption. If the mixing is to be done in a vial, then the mixing time should consequently be as short as possible. The KBr plus analyte mixture should then be transferred completely to a clean 13-mm die, filling and assembling the apparatus according to the manufacturers' instructions. This die should then be connected to a vacuum system and evacuated for  $\sim 2$  min and placed in a hydraulic press and the pressure applied and increased slowly to the required level. The pressure is usually maintained for between 1 and 5 min, before disconnecting the vacuum, and then slowly releasing the pressure. A pressure corresponding to a dial reading

of  $\sim 10$ – $12$  tons ( $\sim 10\,000$ – $12\,000$  kg) is normally applied to the ram for a 13 mm disk. The disk is then removed from the die and mounted in a KBr disk holder, prior to placing it into the sample holder of the infrared spectrometer.

The KBr disk technique has advantages over the mull presentation method in that KBr exhibits no absorption bands above  $400\text{ cm}^{-1}$  (neglecting any absorbed water or impurities) and is better adapted to micropreparations. However, it may have certain disadvantages. The technique is more prone to introducing changes in sample polymorphism, hydration state, or crystallinity, thereby, tending to make it less reproducible and more unsuitable for studying these effects. In some circumstances ion exchange may occur between the analyte and alkali-halide. Notwithstanding, the presentation method has had widespread applications in the chemical characterization of solids. For less stable solids and polymer samples, wet grinding by adding a drop of ethanol or a solvent to the mortar sometimes proves advantageous; although, appropriate precautions must be taken and the solvent must be allowed to evaporate before adding the KBr. A fine powder may be able to be abraded from a bulk sample with the use of SiC paper or a diamond-dust-studded spatula. Some rubbers, fibers, and other polymer samples may be ground to a useable particle size in a vibrating mill that can be cooled with liquid nitrogen.

Other alkali-halide salts, which are commonly used instead of KBr, are KCl and, for extended low wavenumber observations, CsI, although this is much more hygroscopic and softer. PTFE, may also prove to be a useful matrix medium in some instances. PTFE powder and many of its copolymer powders will cold coalesce at ambient temperature under pressure and the alkali-halide die apparatus is very convenient for preparing analytically useful specimens for infrared transmission presentation. About 300 mg of neat powder is placed in the die and pressure is applied in a manner similar to that described above (without the need of vacuum) to form a translucent PTFE disk. For far-infrared presentation, the matrix disk is frequently made from finely powdered polyethylene.

*Thin films from the melt* Low-melting, thermally stable solids may be presented as thin films prepared by softening or melting a few milligrams of a sample between a pair of infrared transparent plates (windows). The assembly is then squeezed until the softened or molten material has covered the contact surfaces of the plates, and then the assembly is cooled, allowing the material to resolidify. The heating may be accomplished on a gradient temperature hot-plate or by radiation from a heating lamp.

This technique is suitable for amorphous materials, such as waxes, tarry solids, hot melt adhesives, and some low-melting solids. Rubbery materials often have a tendency to retract, since the applied pressure is normally only by hand, and the assembly may need to be clamped. For rubbers it is probably better to prepare a thin film by the melt method described in the section 'Hot compression molded films.' This melt approach is not recommended for materials which will crystallize on solidification, since scatter and molecular orientation effects will lead to both poor spectral contrast and irreproducibility.

*Hot compression molded films* Hot compression molding is one of the most effective and satisfactory means of preparing films from many thermoplastic resin samples for infrared presentation, particularly for quantitative compositional analyses. The polymer must of course be thermally stable at the pressing temperature.

A small laboratory hydraulic press, similar to that shown in **Figure 7**, is convenient for producing presentation samples. Such presses are capable of applying up to 20 tons ( $30 \text{ N m}^{-2}$ ) on the ram. The platens are typically  $\sim 15\text{--}20 \text{ cm} \times 15\text{--}20 \text{ cm}$ . They may be heated electrically to temperatures in excess of  $350^\circ\text{C}$ . Rapid cooling of the platens is provided by cooling water flowing through the platens, which is flushed out by air when at ambient or a cool enough temperature. A control box for programmed heating and cooling cycles may also be incorporated into the press system. The sample may be pressed between a pair of metal plates of the same size as the platens. Polished stainless steel plates, a few millimeters thick, are suitable for the purpose. However, these need to be conditioned to prepare film samples that are optimal for infrared examination. The sample contact surfaces should be grit-blasted (25–30 mesh) and then spray over-coated with a very thin layer of a PTFE, which acts as a mold release agent. The surface roughness introduced should be sufficient to minimize the appearance of an interference fringe pattern in the infrared spectrum recorded from the film sample, while the prepared film is still of uniform thickness such that its average thickness may be accurately determined using a dial-gauge micrometer and used, if necessary, to normalize quantitative band absorbance measurements.

According to the thickness required, which will be dependant on the study purpose and whether a thin film is required for a chemical structure survey or a thick specimen for a quantitative quality assurance application, suitable quantities of the sample may be molded into films with or without the aid of

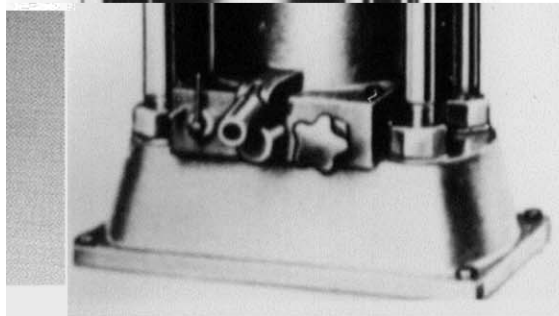
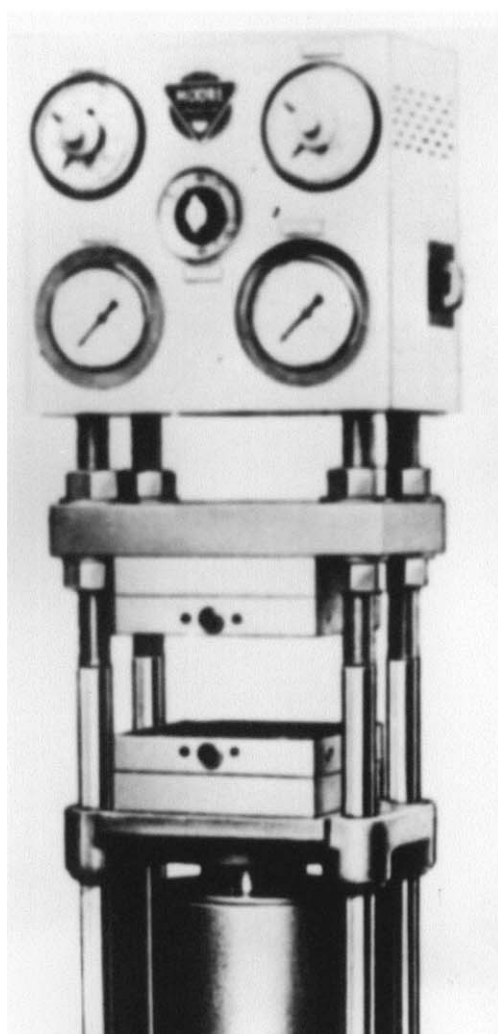
feeler-gauge blades, or for thicker samples a mold-template may be preferred. For fingerprinting purposes a film thickness of 0.1–0.2 mm is generally used for aliphatic hydrocarbon polymers, whereas for observation of all the peak maxima in the spectrum of an aromatic polyester the thickness may need to be less than  $10 \mu\text{m}$ . The optimum press conditions for a particular material will need to be found empirically, although the compression temperature is likely to be close to the melt temperature of the material. For a thin film,  $<25 \mu\text{m}$ , a few milligrams of the sample are placed between the press plates, and the assembly inserted between the press platens, which have been preheated to the press temperature. The press platens are closed onto the assembly, which is allowed to heat for  $\sim 1 \text{ min}$ , the pressure is then applied for about another minute, the heating is turned off and the water cooling is started following the press manufacturer's instructions. The sample is normally left to cool under pressure, after which the press plates are removed and separated to allow access to the film sample. A safe practice is to handle the press plates with heatproof gloves.

To prevent rubbery samples contracting after pressing, they may be hot pressed and then subsequently examined as a sandwich between two thin sheets of silver chloride. Thin specimens of brittle materials may need to be prepared in a similar manner. Films of high temperature thermoplastics will need to be prepared between aluminum foil, which for thicker specimens can be peeled away to reveal the sample but for thin samples can be dissolved away in caustic or dilute HCl. These should be pressed between uncoated plates, since the PTFE coat should not be heated to temperatures in excess of  $\sim 320^\circ\text{C}$ .

Samples will need to be dry and free from solvent residues before hot pressing, otherwise bubbles may form in the film and lead to the recording of erroneous spectral information. Pressing at too high a temperature or for too long may also give rise to bubbles in the sample, as well as thermal degradation, which if excessive will lead to discoloration. Too low a temperature or insufficient pressure will not cause the sample to fully coalesce, and if, for example, formed from a powder, may give rise to scatter in the spectrum, or, from granules, result in a sample that will separate readily at deformed granule boundaries, or, if formed from stacked layers of film, the sample may delaminate.

For certain analyses it may be desirable to ensure that the hot pressed compression molded film is frozen into an amorphous state. This may be achieved by melting a few milligrams between a pair of thin press plates or aluminum sheet, leaving part of the plates exposed from the press platens. This may be

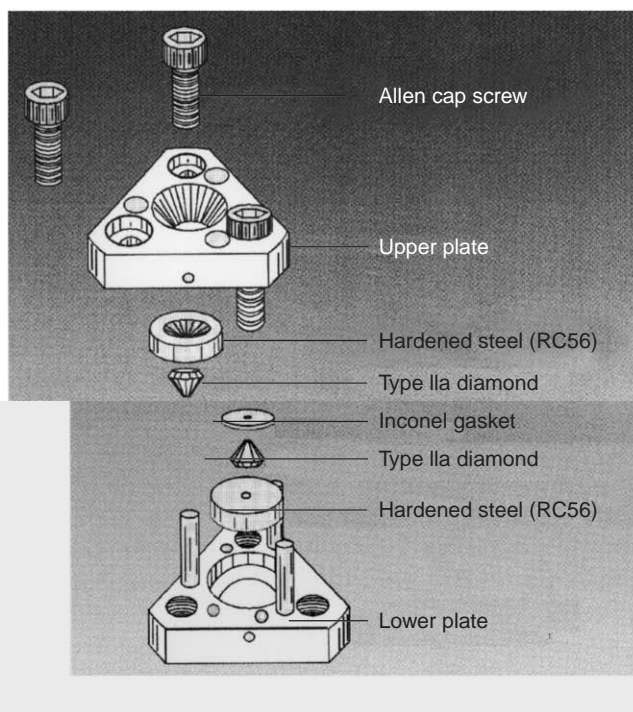




(A)



(B)



(C)

**Figure 7** Extreme examples of presses for preparing from solids films suitable for mid-infrared examination: (A) hand-operated, hydraulic, heated, molding press for preparing thin films from polymer samples for macro mid-infrared examinations; (B) diamond anvil minicell for solid sample compression and containment, particularly for FT-IR microscopy studies; and (C) schematic of diamond anvil minicell. ((A) Reproduced by kind permission of George E. Moore & Son Ltd., Birmingham, West Midlands, UK. (B and C) Reproduced by kind permission of High Pressure Diamond Optics, Inc., Tucson, AZ, USA.)

gripped carefully by a pair of pliers or similar tool held in a thermally insulated gloved hand, the pressure on the platens is then released and the sample-plates sandwich plunged (quenched) into ice-water or Drikold<sup>®</sup> (solid CO<sub>2</sub>)-methanol. At the other

extreme, after the pressure has been applied the electrical heating may simply be turned off and the press with the sample in it kept at pressure and allowed to cool slowly unaided. This annealing will yield a more crystalline morphology in crystalline

and semicrystalline polymers, Although, controlled morphologies are better introduced through the use of programmed temperature cycles, it is possible that metastable crystalline modifications may exist in the freshly pressed film, as occurs for instance with isotactic polybutene-1, which reverts slowly over time from type II (tetragonal crystalline phase) to type I (rhombohedral crystalline domains) with consequent changes in the spectrum, although this is not a particularly common occurrence.

If a prepared film exhibits interference fringes in its infrared spectrum, these may sometimes be suppressed in intensity or eliminated by lightly scouring the film surfaces with wire wool or an abrasive paper.

**Microtomed sections** Sections of appropriate thickness for transmission infrared measurements may be microtomed from continuous solid samples. This presentation method is particularly useful for FT-IR microscopy.

**Reflection methods** Internal reflectance (attenuated reflectance) spectroscopy is a valuable and widely used technique in both qualitative and quantitative examination of solids by mid-infrared spectroscopy. The external reflection technique of reflection-absorption is primarily used for observations on thin absorbing films on reflective substrates; while, although specular reflectance can be of value as a rapid, nondestructive generic fingerprinting tool, its applicability is limited to optically thick samples. Diffuse reflectance mid-infrared spectroscopy can provide a convenient means of examining many finely powdered or highly scattering solid samples. It has particular importance to a wide range of near-infrared (NIR), applications.

**Internal reflection spectroscopy of solids** Mid-infrared internal reflection accessories were developed initially largely for studying the surface layer characteristics of a continuous flat solid sample, e.g., a polymer film; although they may also be used as convenient methods of sample identification, when a surface layer spectrum may be taken as indicative of a sample's bulk characteristics. The sample surface is brought and held under low pressure into optical contact with the clean surface of an internal reflection element. Other common traditional uses have included the direct examination of fibers, foams, and malleable (soft) powders. It has, however, become common practice nowadays, with the sensitivity now available with FT-IR spectroscopy, to use single (or low-number) internal reflection accessories to conveniently and easily record a mid-infrared fingerprint

spectrum from almost any solid sample. An example of such an accessory is shown in Figure 8.

**Thin films on metallic substrates** An analytically useful fingerprint spectrum from a thin film ( $>1000 \text{ \AA}$ ) coated onto a flat reflective metal substrate, e.g., a gold mirror, may be recorded by the reflection-absorption technique. For such film thicknesses, the measurement is relatively straightforward and may be measured in a conventional specular reflectance accessory or in an FT-IR microscope operating in the reflection mode. The infrared radiation beam at near normal angle of incidence makes a double-pass through the film having been reflected back by the substrate. This spectroscopic measurement is often commonly referred to as transreflectance. It must be remembered that superimposed on the essentially double-pathlength transmission spectrum will be a weaker specular reflectance spectrum which will mostly distort the stronger absorption bands, and in some circumstances cause band inversion.

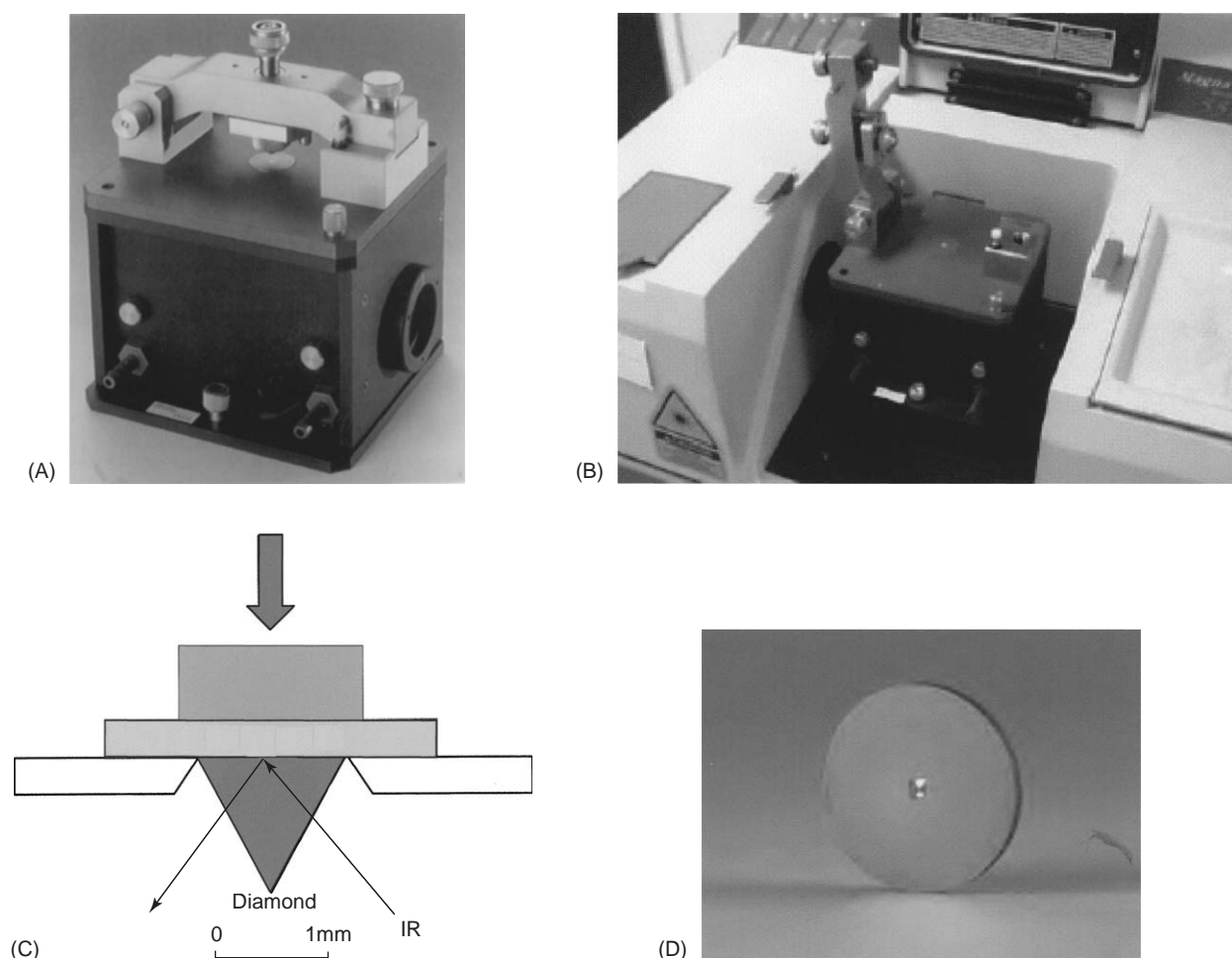
For much thinner films, such as monolayers, the more specialized technique of grazing-angle incidence reflection-absorption is used.

**Specular reflectance** Pure specular reflection spectra may be recorded directly from the surfaces of flat, nonscattering, optically thick (opaque) samples from which the absorption index spectrum may be extracted by application of the Kramers-Kronig algorithm. This is sometimes a useful approach for generically fingerprinting intractable or heavily filled polymer samples. An example is shown in Figure 9.

**Diffuse reflection** Diffuse reflection is a very convenient and useful method for examining by mid-infrared FT-IR spectroscopy highly scattering solids using minimal sample preparation. The diffusely scattered radiation is collected by some high-efficiency collection optics. The spectral contrast depends strongly on the physical and optical properties of the sample. In the mid-infrared region it is primarily a qualitative tool, although sometimes used for low precision semiquantitative measurements. Diffuse reflection is a commonly used sampling technique with NIR spectroscopy.

The majority of samples for mid-infrared investigations are presented as finely divided powders dispersed in an excess of a dry powdered nonabsorbing matrix, commonly KCl. This ensures that superimposed interferences from specular (front-surface) reflections are minimized in the recorded diffuse reflection spectrum. A useful approach for sampling intractable, composite, or gross objects is to abrade a fine powder from the article's surface with some SiC





**Figure 8** Single-reflection ATR accessory: (A) photograph of Golden-Gate™ single-reflection ATR accessory; (B) photograph of Golden Gate™ accessory sited in the sample compartment of an FT-IR spectrometer; the pressure bridge is raised; (C) schematic of a single-reflection ATR system; and (D) photograph of diamond ATR element bonded into its tungsten carbide mount in the accessory top-plate. (Reproduced by kind permission of Specac Ltd., Orpington, Kent, UK.)

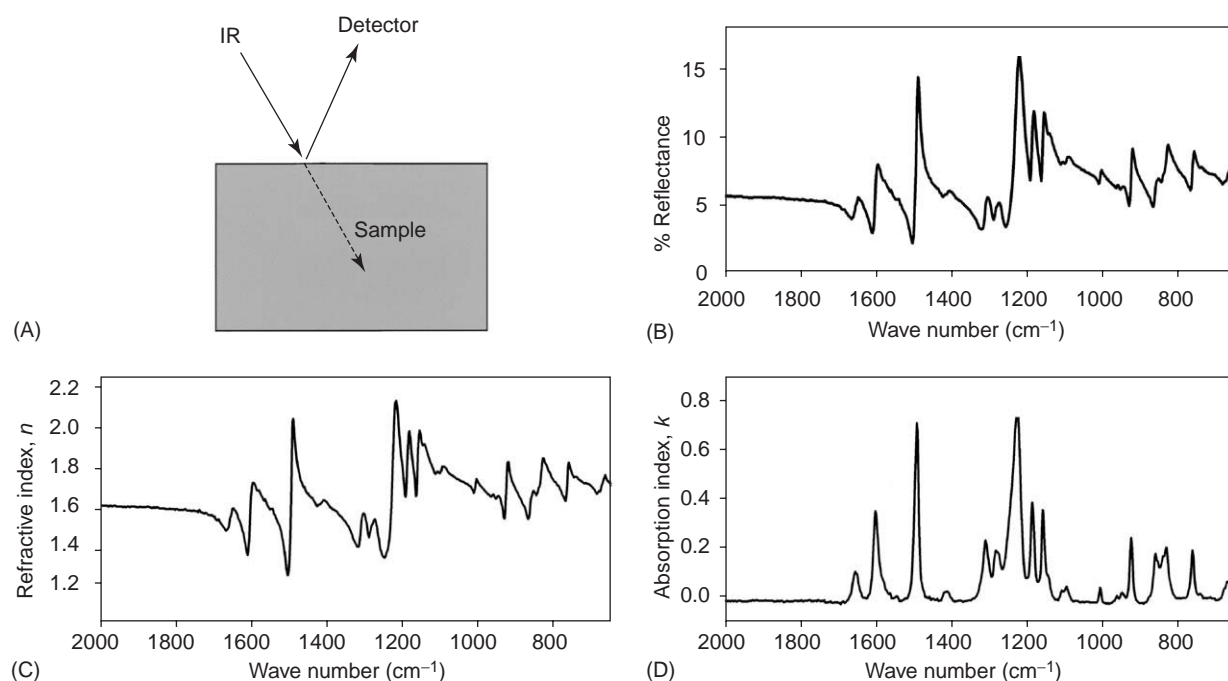
or diamond powder abrasive paper. The abraded powder may then be examined *in situ* on the abrasive paper or removed and dispersed in KCl powder. Diffuse reflectance can also be a particularly convenient method for examining directly foamed materials.

**Other solid sampling techniques** While in an article such as this, it is not possible to cover all mid-infrared sampling techniques; two other methods that have been used significantly in analytical spectroscopy applications are photoacoustic spectroscopy and emission spectroscopy.

**Photoacoustic spectroscopy** Mid-infrared spectra may be recorded by photoacoustic FT-IR spectroscopy from solids having a wide range of physical forms with minimal requirement for any sample preparation.

Photoacoustic spectroscopy provides a convenient qualitative sampling procedure for recording an absorbance spectrum from a wide range of solid materials regardless of their morphology. Essentially the only requirement is that the sample be made to fit into the photoacoustic cell sample holder, although sample form will affect spectral contrast and intensity. Consequently quantitative studies are usually restricted to measurements of a band ratio from a pair of weak to medium intensity bands.

**Emission spectroscopy** An alternative method of measuring the infrared spectrum of a thin film on a metallic substrate is to heat the sample and record its emission spectrum. A temperature of 50–100°C should suffice. This can be particularly useful for nonflat samples such as coatings on drinks or food cans; generic identification of such coatings is



**Figure 9** Specular reflection FT-IR and application of Kramers–Kronig algorithm: (A) schematic showing external (front-surface, specular) mid-infrared reflection measurement from an optically thick sample; (B) specular reflectance spectrum recorded from a 0.6-mm thick polymer molding; (C and D) refractive index and absorption index spectra derived by applying Kramers–Kronig algorithm to the recorded specular reflection spectrum (B), respectively.

probably now, however, more readily undertaken by transmittance under an FT-IR microscope.

## Trends in Sample Preparation

In the analytical laboratory, FT-IR spectrometers are now dominant, if not the almost exclusive mid-infrared tool. Many are also interfaced with FT-IR microscopes that enable high signal-to-noise ratio spectra to be recorded routinely from regions of samples with diameters as low as 20  $\mu\text{m}$ , in some cases 10  $\mu\text{m}$ . There is now a much greater emphasis on sampling techniques appropriate to biological and biomedical samples.

With the growth of FT-IR instrumentation in analytical laboratories throughout the 1980s and 1990s, there was an accompanying introduction and acceptance of many novel sample presentation procedures for mid-infrared spectroscopy. These included photoacoustic and many adaptations of the internal reflection spectroscopy technique, such as the H-ATR. In addition, the rarely used sampling technique of diffuse reflectance became routinely practicable in the mid-infrared region. As a consequence, there was a significant shift from the use of classical transmission measurements utilizing alkali-halide disks, mulls, and thin pathlength liquid cells, to less protracted preparation procedures, particularly

for qualitative purposes. Also, in the 1990s and early 2000s there were significant developments that made Raman spectroscopy available as a convenient, comparable cost, increased sensitivity technique, which with its much lower demands for sample preparation have meant that in niche areas, such as pharmaceutical polymorphic studies and art and archaeology forensics, its use has tended to not only complement but also supplant some mid-infrared applications.

Lately, the single-reflection mid-infrared ATR accessories have proven to be a very attractive tool within an industrial environment; they provide a simple to use means of quickly recording a mid-infrared spectrum from almost any solid material. With the acceptance of these and the increased number of installations of FT-IR microscope systems, there has been a general decline in the use of the diffuse reflection and photoacoustic techniques in the mid-infrared region, and emission measurements are now very rarely used.

Imaging has also become an important interrogation tool, and there is increasing use of FT-IR spectrometers, particularly FT-IR microscope systems, to generate spectroscopic images from samples. Consequently, there is a growing emphasis on methods of sample preparation that optimize specimens for such interrogations.

See also: **Infrared Spectroscopy: Near-Infrared. Photoacoustic Spectroscopy.**

## Further Reading

- Chalmers JM and Dent G (1997) *Industrial Analysis with Vibrational Spectroscopy*. Cambridge: Royal Society of Chemistry.
- Chalmers JM and Griffiths PR (eds.) (2002) *Handbook of Vibrational Spectroscopy*, vol. 2. Chichester: Wiley.
- Coleman PB (ed.) (1993) *Practical Sampling Techniques for Infrared Analysis*. Boca Raton, FL: CRC Press.
- Colthup NB, Daly LH, and Wiberley SE (1990) *Introduction to Infrared and Raman Spectroscopy*, 3rd edn. San Diego, CA: Academic Press.
- Harrick NJ (1967) *Internal Reflection Spectroscopy*. New York: Harrick Scientific Corporation.
- Messerschmidt RG and Harthcock MA (eds.) (1988) *Infrared Microspectroscopy: Theory and Applications*. New York: Dekker.
- Mirabella FM and Harrick NJ (1985) *Internal Reflection Spectroscopy: Review and Supplement*. New York: Harrick Scientific Corporation.
- Siesler HW and Holland-Moritz K (1980) *Infrared and Raman Spectroscopy of Polymers, Practical Spectroscopy Series*, vol. 4. New York: Dekker.
- Willis HA, van der Maas JH, and Miller RGJ (eds.) (1987) *Laboratory Methods in Vibrational Spectroscopy*, 3rd edn. Chichester: Wiley.

## Near-Infrared

**C N G Scotter**, Cherington, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Since the 1970s, near-infrared (NIR) spectroscopy has been applied as a qualitative and quantitative analytical tool in the industrial fields of agriculture, food, chemical and textile production, and pharmaceuticals. The technique is capable of making rapid, nondestructive multicomponent analyses of organic materials in a complex background matrix. In situations where inorganic materials are dissolved in polar solvents an indirect NIR measurement can be obtained. If the concentration of an analyte exceeds 0.1% then it is likely that an acceptable analytical error will be achieved. NIR spectroscopy is essentially an applied instrumental technique not often used in a pure research context, unlike the related mid-infrared (mid-IR) spectroscopy, the spectra from which can be interpreted to identify specific organic bond pairs, e.g., CO, OH, NH. This article discusses the physical theory underpinning the use of NIR, NIR calibration methodology, and types of instrumentation. Examples are given of the application of the technique in the industrial contexts noted above.

## NIR Theory – Molecular Spectroscopy

Two main areas of physical theory are relevant to the application and understanding of NIR: first, the chemical principles derived from quantum physics and molecular spectroscopy, and, second, the physical principles used to relate the spectra to transmission or diffuse reflectance of electromagnetic energy.

Mid-IR and NIR spectra are generated via appropriate instrumentation by electromagnetic energy absorption between 400 and 12 500 nm. Subsets of the wavelength range are employed by instruments with detectors and energy sources appropriate for obtaining spectra in their particular range. The mid-IR range is about 2500–25 000 nm, the NIR range about 1100–2500 nm, and the visible range about 400–800 nm. In the mid-IR range it is usual to refer to wavenumber or frequency in  $\text{cm}^{-1}$  according to the formula wave number ( $\text{cm}^{-1}$ ) =  $10^7/\text{wavelength (nm)}$ .

The mid-IR region provides fundamental absorption data, whereby for a specific bond pair the energy transfer from one level to another occurs at a unique frequency. It is for this reason that the mid-IR region has been used to analyze the bond pairs in organic molecules. It is the  $3600\text{--}1200\text{ cm}^{-1}$  region that gives rise to the overtones in the NIR part of the spectrum. The NIR region thus consists of overtones and combination bands that result in broad and/or overlapping absorption peaks. It is for this reason that NIR spectroscopy employs more statistical, mathematical approaches to the analysis of organic constituents, compared with the approaches used in mid-IR spectroscopy.

Thus, the two regions, although instrumentally and chemometrically distinct, are theoretically parts of the same vibrational and rotational spectroscopic model. Figure 1 illustrates the basic models for energy absorption by covalently bonded atoms of dissimilar mass. A dipole moment (produced when atoms of dissimilar mass are bonded) is a basic condition for absorption to occur. The H–H bond, for example, does not produce mid-IR or NIR absorbance. When molecules are subjected to an

See also: **Infrared Spectroscopy: Near-Infrared. Photoacoustic Spectroscopy.**

## Further Reading

- Chalmers JM and Dent G (1997) *Industrial Analysis with Vibrational Spectroscopy*. Cambridge: Royal Society of Chemistry.
- Chalmers JM and Griffiths PR (eds.) (2002) *Handbook of Vibrational Spectroscopy*, vol. 2. Chichester: Wiley.
- Coleman PB (ed.) (1993) *Practical Sampling Techniques for Infrared Analysis*. Boca Raton, FL: CRC Press.
- Colthup NB, Daly LH, and Wiberley SE (1990) *Introduction to Infrared and Raman Spectroscopy*, 3rd edn. San Diego, CA: Academic Press.
- Harrick NJ (1967) *Internal Reflection Spectroscopy*. New York: Harrick Scientific Corporation.
- Messerschmidt RG and Harthcock MA (eds.) (1988) *Infrared Microspectroscopy: Theory and Applications*. New York: Dekker.
- Mirabella FM and Harrick NJ (1985) *Internal Reflection Spectroscopy: Review and Supplement*. New York: Harrick Scientific Corporation.
- Siesler HW and Holland-Moritz K (1980) *Infrared and Raman Spectroscopy of Polymers, Practical Spectroscopy Series*, vol. 4. New York: Dekker.
- Willis HA, van der Maas JH, and Miller RGJ (eds.) (1987) *Laboratory Methods in Vibrational Spectroscopy*, 3rd edn. Chichester: Wiley.

## Near-Infrared

**C N G Scotter**, Cherington, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Since the 1970s, near-infrared (NIR) spectroscopy has been applied as a qualitative and quantitative analytical tool in the industrial fields of agriculture, food, chemical and textile production, and pharmaceuticals. The technique is capable of making rapid, nondestructive multicomponent analyses of organic materials in a complex background matrix. In situations where inorganic materials are dissolved in polar solvents an indirect NIR measurement can be obtained. If the concentration of an analyte exceeds 0.1% then it is likely that an acceptable analytical error will be achieved. NIR spectroscopy is essentially an applied instrumental technique not often used in a pure research context, unlike the related mid-infrared (mid-IR) spectroscopy, the spectra from which can be interpreted to identify specific organic bond pairs, e.g., CO, OH, NH. This article discusses the physical theory underpinning the use of NIR, NIR calibration methodology, and types of instrumentation. Examples are given of the application of the technique in the industrial contexts noted above.

## NIR Theory – Molecular Spectroscopy

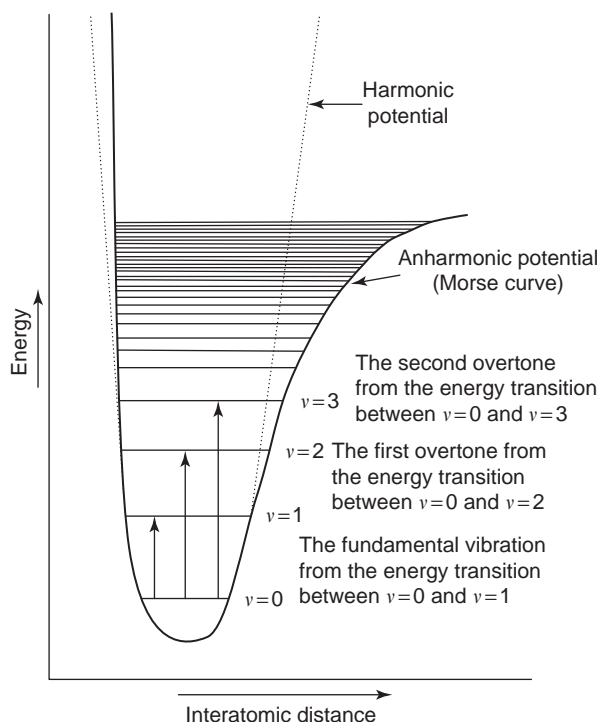
Two main areas of physical theory are relevant to the application and understanding of NIR: first, the chemical principles derived from quantum physics and molecular spectroscopy, and, second, the physical principles used to relate the spectra to transmission or diffuse reflectance of electromagnetic energy.

Mid-IR and NIR spectra are generated via appropriate instrumentation by electromagnetic energy absorption between 400 and 12 500 nm. Subsets of the wavelength range are employed by instruments with detectors and energy sources appropriate for obtaining spectra in their particular range. The mid-IR range is about 2500–25 000 nm, the NIR range about 1100–2500 nm, and the visible range about 400–800 nm. In the mid-IR range it is usual to refer to wavenumber or frequency in  $\text{cm}^{-1}$  according to the formula wave number ( $\text{cm}^{-1}$ ) =  $10^7/\text{wavelength (nm)}$ .

The mid-IR region provides fundamental absorption data, whereby for a specific bond pair the energy transfer from one level to another occurs at a unique frequency. It is for this reason that the mid-IR region has been used to analyze the bond pairs in organic molecules. It is the  $3600\text{--}1200\text{ cm}^{-1}$  region that gives rise to the overtones in the NIR part of the spectrum. The NIR region thus consists of overtones and combination bands that result in broad and/or overlapping absorption peaks. It is for this reason that NIR spectroscopy employs more statistical, mathematical approaches to the analysis of organic constituents, compared with the approaches used in mid-IR spectroscopy.

Thus, the two regions, although instrumentally and chemometrically distinct, are theoretically parts of the same vibrational and rotational spectroscopic model. Figure 1 illustrates the basic models for energy absorption by covalently bonded atoms of dissimilar mass. A dipole moment (produced when atoms of dissimilar mass are bonded) is a basic condition for absorption to occur. The H–H bond, for example, does not produce mid-IR or NIR absorbance. When molecules are subjected to an





**Figure 1** Hooke's law and Morse curves as functions of energy and interatomic distance.

external energy source they acquire the potential for energy changes and the amount of energy that can be imparted from a source is dependent upon the wavelength, according to the formula:

$$E = hc\bar{\nu}$$

where  $E$  is the energy,  $h$  is Planck's constant ( $6.626 \times 10^{-34}$  J s),  $c$  is the speed of light, and  $\bar{\nu}$  is the wave number in  $\text{cm}^{-1}$ .

When organic molecules absorb energy from an external source, e.g., a tungsten or halogen (while tight) lamp, some molecules will move from the ground state ( $v = 0$ ) to the next energy level ( $v = 1$ ) (see **Figure 1**). The combined vibrational and rotational energy changes involved in the energy level change constitute the absorption process.

**Figure 1** illustrates the simplest theoretical treatment for the energy absorption behavior of covalent bonds. Hooke's law curve corresponds to bonds vibrating in simple harmonic motion. In practice, covalent bond vibrations are anharmonic and follow the potential energies depicted in the Morse curve, also illustrated in **Figure 1**. It is this phenomenon of anharmonicity that allows the overtones or harmonics to occur in the NIR region. Thus, the frequencies of the overtones are slightly less than whole number

multiples of the fundamental frequency. Greater amounts of energy are required to displace molecules to higher overtones and the numbers of molecules displaced is in inverse proportion to the overtone number. Transitions occur most frequently from the ground state ( $v = 0$ , see **Figure 1**) to the first level, and are increasingly infrequent for the higher levels. Consequently, higher overtone absorptions are much weaker than the fundamental.

Combination bands are also encountered in NIR spectra. These result from energy transitions involving two or more different vibrational modes for the same functional group. For example,  $\text{CH}_2$  groups have two symmetrical vibrational modes, a symmetric stretching mode, and a corresponding in-plane bending, or scissoring. The combination band resulting from these two vibrational modes is seen at 2320 nm. A final factor that has been identified as complicating NIR spectra even further is that of Fermi resonance. In this condition, unlike sets of bonds, e.g., the O-C-O bending mode and the symmetrical C-O stretching mode, occur at approximately the same frequency and a repulsion occurs. One type of bond energy level is lowered and the other type gains energy. In the case of carbon dioxide this results in bands at 1388 and 1286  $\text{cm}^{-1}$  instead of one strong and one weak line at 1340 and 1330  $\text{cm}^{-1}$ , respectively. Although relatively obvious in the mid-IR region, Fermi resonance phenomena are usually hidden beneath the broad overlapping peaks in the NIR.

Although NIR absorbance peaks are generally broad, spectral assignment is possible. A number of authors have discussed bond assignment in the NIR region and tables and lists of assignments of particular bonds and vibrational modes have been published.

### Transmission and Diffuse Reflectance

When NIR radiation is passed through a non-scattering organic medium, the intensity of absorption can be described in terms of transmittance,  $T$ :

$$T = I/I_0$$

where  $I$  is the intensity of the emergent radiation and  $I_0$  is the incident energy. For absorption spectra,  $I$  can be expressed in terms of the Beer-Lambert relationship:

$$\log_{10}(I_0/I) = \log_{10}(1/T) = kcl = A$$

where  $A$  is the absorbance,  $k$  is the molar absorptivity,  $c$  is the concentration of the absorbing molecular species, and  $l$  is the pathlength of energy



through the sample. For a fixed pathlength the absorbance is thus directly proportional to the concentration of the absorbing species.

In the case of samples that produce scatter in transmission or diffuse reflectance spectra, a number of factors corrupt the linearity of the Beer–Lambert absorbance concentration relationship. Sample scattering of radiation results in an alteration of the proportion of absorbed and reflected radiation so that pathlength becomes another unknown in the Beer–Lambert relationship. Particle size, particle shape, crystalline form, bulk, density, and the nature of the pore space (filled with air, water, or oil) are all variables that dictate the effective pathlength of the radiation. Sample surfaces also reflect specular energy that has not interacted with molecular structures. This form of energy has an overall effect on spectra, contributing primarily to the curvilinearity of the spectral baseline.

A number of attempts have been made to describe both theoretically and practically diffuse reflectance and scattering functions to enable a linear relationship to be established between absorbance ( $A$ ), expressed as  $\log_{10}(1/R)$  where  $R$  is the reflectance, and molecular concentration. Perhaps the commonest relationship encountered is that ascribed to Kubelka and Munk, who established nine assumptions and 16 variables. These can be simplified to the Kubelka–Munk function, namely:

$$\frac{(1 - R_{\infty})^2}{2R_{\infty}} = \frac{K}{S} = \frac{kc}{S}$$

where  $R_{\infty}$  is the reflectance for a sample of effectively infinite depth,  $K$  is a constant proportional to the  $kc$  term in the Beer–Lambert relationship,  $S$  is a scattering coefficient usually taken as  $\delta^{-1}$  where  $\delta$  is the diameter of the particles in a densely packed scattering medium,  $k$  is the molar absorptivity, and  $c$  is the concentration of the absorbing analyte.

## NIR Calibration Development

The most important first step in the practical application of NIR is appropriate sample selection. To take advantage of the potential rapidity and accuracy of analysis that NIR spectroscopy can offer, the NIR instrument must be calibrated against an acceptable, accurate, and reproducible laboratory method for the analyte or analytes of interest. The selection of a sample set for NIR scanning and laboratory analysis should be, as far as practically possible, representative of the material that the calibration will encounter when it is used to predict the targeted analyte or analytes. Another critical aspect of NIR analysis

is appropriate and repeatable sample presentation. Instrument manufacturers provide a plethora of sampling devices that attempt to optimize presentation criteria for specific sample types.

One approach to establishing a calibration sample set is to scan a large number of samples that represent the potential population to be predicted and to submit the spectra to an analysis that establishes which combination of samples shows the largest spectral variability overall. This smaller subset of samples can then be analyzed by the laboratory method. In practice, it will seldom be possible to establish calibrations that will optimally predict all future samples. It is therefore helpful to include new samples in the calibration that are spectrally different from the original calibration set. The standard error of calibration (SEC) is used as an indication of the efficacy of a calibration.

It is equally important that a calibration is tested against a set of samples different from those of the calibration set. They should ideally be separately collected but they must embody the chemical and physical features encompassed by the calibration set. It is the standard error of prediction or performance (SEP), which defines the success of a calibration. It is important, however, to relate the SEP to the overall range of the analyte in question.

The factors listed below are commonly responsible for increasing the prediction error:

- faulty laboratory analysis,
- sample heterogeneity,
- lack of repeatability in the packing and presentation of the sample,
- instrumental malfunction.

The standard deviation for the residuals is due to differences between the laboratory analytical values and the NIR predicted values for samples in the calibration set (SEC) or in a test set (SEP).

## Mathematical Data Treatment

Although the theoretical models noted above attempt to satisfy well-established physical principles, in practice most NIR workers start the calibration equation development process with the orthodox  $\log_{10} 1/R$  or  $\log_{10} 1/T$  spectra. Although calibrations have been generated using the ‘draw’ log reciprocal data, it has been found in most cases that some form or forms of mathematical data treatment prior to regression analysis improves a calibration substantially.

Mathematical approaches that have been employed to obviate instrumental noise and other

physical effects in favor of chemical features in NIR spectra include:

**Smoothing functions** The simplest smoothing function employed is to average the absorbance ( $\log 1/R$  or  $\log 1/T$ ) values over a segment of the spectrum. Many other smoothing algorithms are available to smooth noisy spectra and reduce the numbers of wavelengths in spectra. It should be noted, however, that a high signal-to-noise ratio is characteristic of NIR spectra and smoothing algorithms are primarily used as an initial data reduction step.

**Derivative functions** Derivative functions (strictly termed differencing) provide a plot of the rate of change of the slope in the spectrum. The average slope is defined over a spectral segment that is followed by the omission of several wavelengths (a gap). The gap may be zero. The process is repeated along the spectrum to its final wavelength. This finite difference technique is usually combined with a smoothing algorithm. The size of the gap and segment may be varied. This technique has the effect of removing parallel baseline shift due to particle size differences in the samples, but it retains slope differences between spectra, and thus retains the basic peak height molecular species concentration relationship. Derivatized spectra contain lobes and valleys and inflected peaks and are therefore difficult to interpret. Up to four orders of derivative have been employed with NIR spectra. Trial and error and experience with gap and segment sizes provide optimal solutions for calibrations and predictions.

**Multiplicative scatter correction (MSC)** The model is based partly upon theoretical considerations of radiation scatter and partly upon empirical observations of spectra.

If  $x_1, x_2, \dots, x_k$  are the absorbance values of a sample and  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$  are the average absorbance values for the set of samples. The regression model used is:

$$x_k = a + b\bar{x}_k + e_k$$

The coefficients  $a$  and  $b$  are estimated by least squares over the wavelength range and the scatter-corrected spectrum is obtained by subtracting  $a$  and dividing by  $b$ , i.e.,

$$x_{k\text{new}} = (x_{k\text{old}} - a)/b$$

$e$  is a term accounting for 'error' caused by, for example, random measurement noise. MSC also requires a linearity transform to ensure a linear relationship between absorbance values at wavelengths absorbing for the analyte of interest.

**Standard normal variate (SNV) and detrending** For a given set of similar samples the reflectance spectra will vary in their slope and curvilinearity according to the particle size distribution between samples, differences in crystallinity, and differences in the chemically nonspecific scatter at the sample surfaces.

The SNV function for a 700 wavelength spectrum may be expressed as:

$$\text{SNV}_n = y_n - \bar{y} \left[ \sum_{i=1}^n (y_i - \bar{y})^2 \right]^{-1/2} \quad (n-1)$$

$$n = 1, \dots, 700$$

where  $\text{SNV}_n$  ( $n = 1, \dots, 700$ ) are the individual SNV terms for each of the 700 wavelengths,  $y_n$  are the 700  $\log 1/R$  values, and  $\bar{y}$  is the mean of the 700  $\log 1/R$  values.

A dominant feature of NIR diffuse reflectance spectra is the increase in absorbance values from 1100 to 2500 nm. This trend is curvilinear for densely packed samples. A second-degree polynomial function has proved to be an adequate model to linearize the spectral baseline.

The list of NIR spectral transformations is not exhaustive but it exemplifies the physical features that are perceived to interfere with the chemical absorbance features and the approaches used to overcome the perceived interference in diffuse reflectance NIR spectroscopy.

None of the above approaches optimizes the relationship between NIR absorbances and analyte for a range of sample types. Derivative transformations have been found to be generally useful when stepwise multiple linear regression (SMLR) techniques are used. When multidimensional statistics are employed, e.g., partial least-squares (PLS), principal component regression (PCR), or neural nets, it has been observed in some cases that the untransformed  $\log 1/R$  data can perform just as well in correlation coefficient and error terms as in any kind of transformation. It is considered that in some cases physical manifestations of the sample contained in the spectra provide valid and useful discriminant data.

## Fourier and Wavelet Transforms

In the past, the large amount of spectral data generated by NIR instruments challenged the ability of computers to provide computations within a reasonable time frame. Mathematical techniques, therefore, which offered a reduction of the raw data, but with minimum loss of information, were often employed. The decomposition of 1000 spectral data points to 100 Fourier transform coefficients provides a great saving in computational time with very little loss of

spectral information. Large memory computers have meant, however, that NIR workers can derive functional calibrations or discriminant functions using all the spectral data. With the advent of spectral imaging datasets, and very large sample databases, the need has reemerged for efficient data reduction techniques. One of the most recent, and promising techniques is wavelet transformation. This technique employs various wave functions that can be combined to represent a spectrum. Wavelet transforms have not been used widely so far, but there is hope that they may provide a powerful way of data reduction without loss.

### Developing the Regression Equation

Until recently, it was most common to use SMLR to develop a calibration model. A condition for the use of the technique is that the relationship between absorbance expressed as  $\log_{10} 1/R$  or  $\log_{10} 1/T$  and the laboratory measurement should be very close to linear. The specular reflectance and particle size effects seen in spectra discussed below also affect the optimal reduction of the SEC and SEP.

When the laboratory value is plotted against the NIR predicted value for the calibration sample set it may well be noted that some points lie well away from the computed regression line. This will, of course, reduce the correlation between laboratory and NIR data and increase the SEC or SEP. These samples may be outliers. The statistic  $h_i$  describes the 'leverage' or effect of an individual sample upon a regression. If a particular value of  $h_i$  is exceeded this may be used to determine an outlier sample. Evaluation criteria for selecting outliers, however, are somewhat subjective so there is a requirement for expertise in multivariate methods to make outlier selection effective.

Intercorrelation (or multicollinearity) phenomena also reduce the effectiveness of a calibration. First, there can be intercorrelation between the constituents in the samples. For example, water and fat in some meats are negatively correlated and it is possible to obtain the same calibration wavelengths for both components although the two components do not absorb in the same band.

Particle size differences between samples cause all calibration coefficients to change because of the baseline shift of the spectra. Furthermore, adjacent wavelengths are also highly correlated, so it is difficult for the SMLR algorithm to 'focus' on a single combination of wavelengths that will optimize the terms in the calibration. There are also correlations between areas in the spectrum that represent overtones of the same covalent bond.

It is often suggested that slope and bias corrections should be made to computed regression models, i.e., an adjustment be made to the regression line so that it can be extrapolated through the zero laboratory/NIR prediction values (bias or intercept). This procedure will also adjust the slope term in the equation. The most recent view of this technique is that it should be used with utmost caution, in that if such adjustments are apparently required to improve the statistical results then it probably indicates that the original model requires improvement or that laboratory or NIR measurements are at fault.

### Multidimensional Regression Techniques

There are a number of multidimensional, multivariate techniques that are capable of overcoming many of the effects that corrupt calibrations, without necessarily using data pretransformations. These techniques also involve wavelength data compression, resulting in a decrease of computational time. Equations are also derived from functions of the whole spectrum. PLS and PCR are the most commonly used of these techniques. Both techniques derive linear functions from the wavelength data (latent variables in PLS, principal components in PCR). Typically, a few of the functions in relationship to the sample number will provide sufficient data to generate coefficients for good calibrations equations.

In PCR, the components are derived by extracting the maximum amount of residual variation from the spectral data. Each principle component is orthogonal to the next (they have 0 correlation). In this way the wavelength intercorrelation problem noted above is alleviated. PLS functions are derived by interrelating the laboratory value variability and the NIR spectral variability and latent variables have low intercorrelation. Although a number of writers point to relative advantages for these techniques, in practice, used properly, they provide similar solutions. Fourier transform techniques are also used for calibration. Once again the wavelengths are compressed into a relatively few (Fourier) coefficients. In common with PLS and PCR, the adjacent Fourier functions (coefficients) have a low intercorrelation.

The three techniques of multidimensional calibration noted above also generate functions that are valuable for the chemical and physical interpretation of spectra.

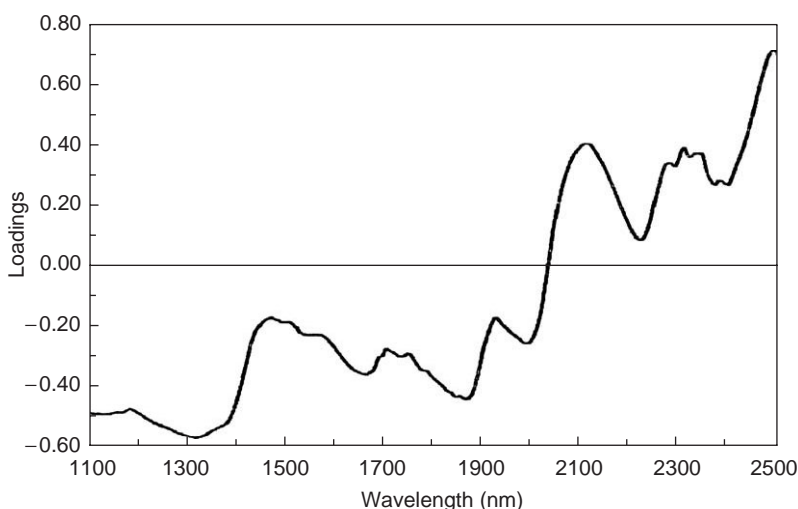
Principle components regression, for instance, generates 'loadings' for each principal component (PC) that may be plotted against wavelength, to illustrate the importance of specific wavelength regions to the regression analysis. The peaks and troughs of the plot correspond to high positive or

low negative loadings on the ordinate axis and are indicative of the degree of influence that the wavelength regions, encompassed by the peaks or troughs, have on the final model. Since the loadings are calculated from spectral variation in the whole sample set, the loadings against wavelength plot reflects spectral features in the dataset. **Figure 2** shows the loadings against wavelength plot for the second PC derived from a set of orange juice samples. In this plot, the peak at 2100 nm is homologous with a peak in the orange juice spectra at 2100 nm (see **Figure 3**), which in turn is a characteristic peak in the spectrum of sucrose, and sucrose is a major varying constituent of the orange juice samples from which the PC was derived. **Figure 3** illustrates an orange juice spectrum. In **Figure 4**, the first PC loadings are plotted against wavelength for the orange juice reflectance data referred to above. In this case, the plot shows a weak 'smoothed' orange juice spectrum (*cf.* **Figure 3**). This

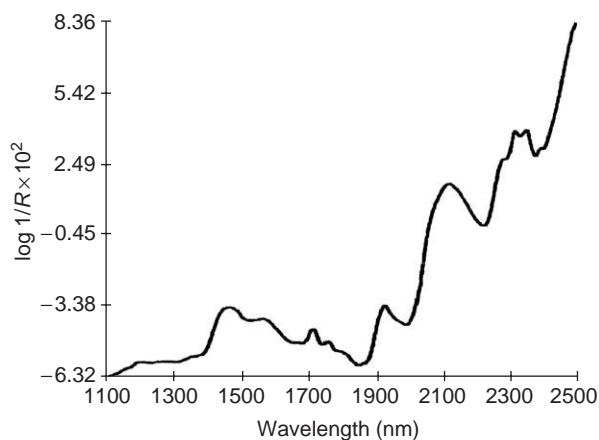
represents a spectral pattern that is generated by specular (mirror-like, no interaction with molecular bonds) reflection and particle size differences between samples. **Figure 5** shows a series of leaf tea reflectance spectra where the spectra are very similar in shape but the spectra are spaced up and down the ordinate axis. This is a typical spectral response to samples with particle size differences.

Fourier analysis offers the process of self-deconvolution that is capable of resolving underlying peaks from broad overlapping bands. Fourier self-deconvolution has the advantage over derivatized spectra that there is no shift or averaging effect and the self-deconvoluted spectrum is directly comparable with the log/reciprocal wavelength scale.

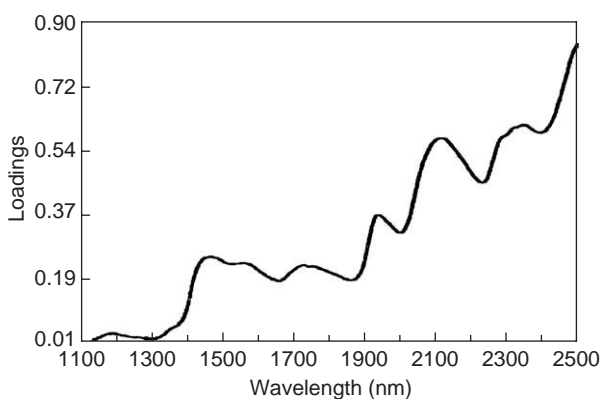
Two other techniques of calibration equation computation have recently been used by NIR workers, locally weighted regression (LWR) and artificial neural nets.



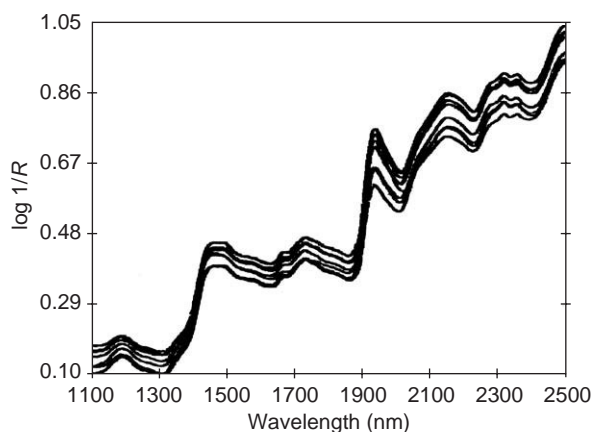
**Figure 2** Second principal component loadings versus wavelength plot of a set of 25 components used for canonical variates analysis of orange juice NIR reflectance spectra.



**Figure 3** log 1/R spectrum of orange juice dried onto a glass fiber disk.



**Figure 4** First principal component loadings versus wavelength plot for a set of orange juice reflectance spectra.



**Figure 5**  $\log 1/R$  spectra of tea leaf samples showing the 'stacking' of similarity shaped spectra along the ordinate axis.

LWR has been recommended to counteract the common inherent nonlinearity existing between laboratory values and the spectral absorbances. Essentially, segments of the NIR versus laboratory value regression computed are used to give a much better approximation to linearity than a regression line model for the whole dataset.

### Neural Networks

Neural networks have been used, most effectively, as a way of establishing robust NIR calibrations. It is the calibration method of choice for a whole cereal grain, transmission instrument, which is being used worldwide for many of the analyses required for estimating grain quality. An artificial neural network program performs many iterations in order to establish the optimum solution required from large, complex datasets.

### Discriminant Analysis

NIR instruments may not only be calibrated for quantitative analysis, but they may also be 'trained' for qualitative purposes. This process is usually termed discriminant analysis. The criteria noted previously for establishing quantitative calibrations with minimum prediction error are equally applicable to discriminant calibration sets.

Spectral matching algorithms are also employed using multidimensional mathematics. Such techniques are often used in quality assurance programs where a spectrum of a known material is compared with incoming raw materials to check its purity or authenticity. For example, samples of orange juice concentrate were collected to represent the types available from two countries. In addition, samples that were known to be adulterated were obtained from one of the countries. All other samples were

characterized as authentic. The question asked was 'Can NIR spectra be used to discriminate between the three categories of juice concentrate?' That is, country A, country B (authentic), and country B (adulterated).

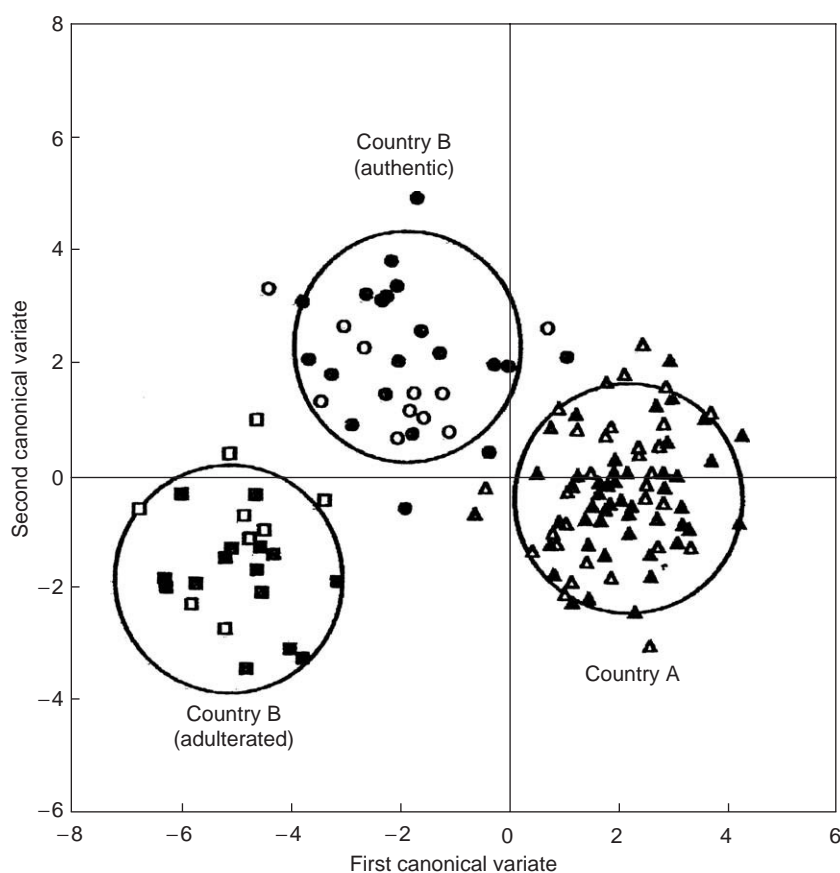
The technique chosen to perform the analysis was canonical variates analysis. Briefly, the method abstracts functions from combinations of PCs. It has been found with large sample sets ( $\geq 100$  or more) that up to 25 PCs give optimal predictive ability. It is unwieldy to select the predictively important analytes from 25 PC dimensions. For this reason the PC-derived canonical varieties are computed wherein the dimensions available are one less than the number of ascribed characteristics or groups. In the example noted above, therefore, the two CVs describe the 25 PC dimensions, since there are three groups in the analysis. In this example, a calibration discriminant analysis was calculated using 71 samples, and 23 prediction samples were used to test the calibration. **Figure 6** illustrates the two-dimensional CVs' calibration and prediction.

One problem that is evident from this analysis is that samples that lie outside the defined circles (95% of the samples sample points should lie within the group circles providing each group has the same variance), may (1) belong to one of the defined groups or (2) belong to a population not defined in the calibration. At present no satisfactory method exists to make an objective decision about such 'outlier' samples. One approach to help with this problem is to employ cluster analysis techniques. In this case sample spectra are allotted to specific groups in a hierarchical fashion, beginning with a division of the sample set into two groups, then four, and so on. This type of analysis requires decisions to be made about appropriate distance and similarity functions, which are required to perform the computations. There are many such functions and only experience of their use leads to valid analysis and samples that may be truly outside the group populations will at some level of the analysis be included in a group.

### A Nonregression-Based Technique

Comparison analysis using restructured near-infrared and constituent data (CARNAC) can be used to provide quantitative or discriminant analysis for large databases. In the current version of the technique, CARNAC-D, wavelet transforms are used to reduce the spectral information, and then the spectrum of the sample requiring a particular analyte measurement or discriminant comparison (transformed in the same way) is compared with the 'calibration' database. The sample giving the closest





**Figure 6** Canonical variates calibration and prediction sample plot of NIR diffuse reflectance spectra characterized as: ▲, △, country A; ●, ○ country B (authentic); and ■, □, country B (adulterated). Solid symbols, calibration samples; hollow symbols, prediction samples.

match gives the solution to the required analyte or discriminant identification.

## Types of NIR Instrumentation

Until the late 1980s there were only two basic types of NIR instrument, those employing discrete wavelength filters and wavelength scanning instruments using holographic gratings. Both types employ high-intensity white light sources and one or two types of detector depending upon the wavelength range being used. For the NIR region (1100–2500 nm) a lead sulfide detector provides appropriate sensitivity and in the higher-energy visible/far visible region from 400 to 1100 nm less sensitive silicon detectors are used.

Detector geometry is important to optimize the collection of diffuse reflectance energy in favor of specular reflectance energy that contains no chemical absorbance information. Two approaches are commonly used. One uses an integrating sphere that provides a double-beamlike instrument, whereby the arrangement of the detectors within the reflective sphere enables a reference and a sample beam to be employed. The reference spectrum is subtracted from

each sample spectrum obviating the inclusion of purely instrumental contributions to the sample spectrum. Single-beam instruments, using a detector array set to 45° to the sample surface, must have a reference material to scan, the spectrum of which is subtracted from the sample scan. The full-wavelength scanning instruments are more expensive than filter instruments and are in general used for calibration development and applied research.

In the late 1980s and into the new millennium instrumental approaches to the NIR region have increased substantially and some of these new approaches are noted briefly below.

### Calibration Transfer

NIR instruments of all kinds differ in their optical and operational properties. Instruments of the same kind and make have small differences, but these produce differences in spectra that are sufficient to erode the accuracy of a calibration if it is transformed from one instrument to another. There has been, and there continues to be, an interest in ways of transforming calibrations, such that, when they are

transferred from one instrument to another they retain their predictive accuracy in the new instrument.

### **Transmission through Solids**

Using the higher-energy visible region, instruments have been made to provide transmission spectra of solid samples that have been satisfactorily calibrated for kernel hardness of gain, meat components, and tenderness of peas, for example.

### **Photodiode Array Instruments**

Photodiode array (PDA) type of instrument is solid state (no moving parts) and because of the electronic switching of the diodes providing the NIR illumination, it is very rapid scanning, compared with other instrument types. Such a construction lends itself to miniaturization, and the result is physically robust, enabling it to be used in many situations where other instrument types could not be used, including situations where a hand-held analyzer is required. At present the wavelength range is limited between 500 and 1100 nm (other instruments can provide wavelengths up to 2400 nm). The spectra from a PDA instrument also tend to be noisier than spectra from other instruments. Currently, this type of NIR analyzer is being used as an integral measurement system in a cereal combine harvester. It is also being used experimentally to estimate quality parameters of whole fruit on the tree. The uses for this NIR technology are increasing rapidly.

### **NIR Image Analysis**

Developments are currently taking place in using the visible and NIR region for video imaging purposes. Experiments have been carried out to image meat surfaces and the textural appearance of bread crumbs.

A leading pharmaceutical manufacturer has developed an NIR-based image analysis technique to image the ingredients of capsules and tablets to check there even distribution in a formulated product.

Diffuse reflected NIR radiation from the surface of the sample is collected by a series of imaging optics and passes through an NIR tunable filter prior to forming an image on an IR focal plane array detector. The filter is continuously stepped through a predetermined spectral interval and an image at each wavelength is stored. The resulting data hypercube is analyzed by proprietary principal component-based software.

## **Industrial Uses of NIR Instrumentation**

NIR spectroscopy is an application-driven technique and perhaps one of the greatest motive influences in the development of the many kinds of NIR instrument

has been their application to at-, on-, or in-line analysis for factory processes. Once a calibration has been developed, the NIR technique does not require chemicals and is capable of rapid, multicomponent, noninvasive analysis. Two features of NIR instrumental development are most significant in this context. The first is the increasing use of optical fibers, which enables the instrument to be multiplexed and/or remote from the line. Optical fibers are also used within instruments to optimize the light path geometry. The second feature is the development of nonmoving part or solid-state spectrometers that are light, rugged, long-lasting, small, and up to 10 times less expensive than their equivalents with moving parts. As has been noted, this type of instrument can also be developed to be handheld.

## **Industrial Applications of NIR Technology**

### **Applications in Agriculture and the Animal Feed Industries**

In the agriculture and animal feed industries NIR instruments, and in particular discrete filter instruments, have found the widest application in the analysis of grain and flour. Protein, moisture, hardness, and baking quality are all examples of analysis that have been performed by NIR instrumentation, both off- and online. Several countries, including the USA and Canada, have adopted NIR protein analysis as national standards to gauge the level of payment for wheat. The protein method was approved by the AACC (American Association of Cereal Chemists) in 1983.

In the feed and forage industries NIR methods have been used extensively. NIR factory-based quality control schemes have been implemented which enhance the economy with which feed components are used, and the quality and consistency of the final product. Moisture, crude protein, and acid detergent fiber analyses of feeds are being steered through the AOAC (Association of Official Analytical Chemists) validation process. NIR is also used widely in the US tobacco industry for measuring nicotine and other process parameters both on- and offline.

### **Applications in the Food and Beverage Industry**

It is perhaps the food and beverage industry that presents the most difficult challenge to both off- and online NIR analysis. The materials are organically complex, and often need to be analyzed in the presence of a high percentage of moisture, the broad, highly absorbing bands of which, in the NIR region, tend to obscure the lesser absorbing constituents.

**Table 1** Matrix of food products and components and characteristics analyzed offline and online by NIR

Food product	Acids $\alpha$ , $\beta$	Alcohol	Amino acids	Ash	Caffeine	Capsaicin	Casein	Cellulose	Chlorophyll	Color	Dietary fiber	Egg	Fat/oil	Fungal spores	$\beta$ -Glucan	Gluten	Hardness	Heat treatment	Hot water extract	Insects	Iodine value	Lactose	Methionine	Moisture	Nicotine	Nitrogen (total)	Nylon	Original gravity	Pasteurization	Polypropylene	Polythene	Proportions in a mixture	Protein	Salt	Sedimentation volume	Soya bean flour	Starch	Starch damage	Sucrose	Sulfur	Sugars	Tenderometer	Total solids	Water absorption		
Barley	-	-	x	-	-	-	-	-	-	-	-	-	x	x	-	-	-	x	-	-	-	-	x	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-			
Beer	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	o	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Beet sugar	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	o	-	o	-	-	-	-			
Biscuit	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	o	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-			
Bread	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	x	-	-	x	-	-	-	-	-	-	-	-			
Breakfast cereal	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
Cake mixes	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-			
Cheese	-	-	-	-	-	-	-	-	-	-	-	o	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-		
Chocolate	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-			
Cocoa	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	x	-	-	x	-	-	-	-	-	-	-	-			
Coffee	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-		
Dough	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	x	-	-	-	-	x	-	-	-	-	-	-			
Edible oils	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Fish	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-		
Flour	-	-	-	x	-	-	x	-	x	-	-	-	-	-	o	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	o	-	-	-	x	-	-	-	-	-	-	-	x	-	
Hops	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Ice cream	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Liquors	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	
Maize	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Malt	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	
Margarine	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Meat	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	x	x	-	x	-	-	-	-	-	-	-	-	-	-	
Milk products	-	-	-	x	-	-	x	-	-	-	-	o	-	-	-	-	x	-	-	-	x	-	o	-	-	-	-	x	-	-	-	x	x	-	-	-	-	-	-	-	-	-	-	x	-	-
Nondairy creamer	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	
Oats	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	
Oilseed	-	-	-	-	-	-	-	x	-	-	-	x	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	
Packaging laminate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	x	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Pasta	-	-	-	-	-	-	-	-	-	-	x	x	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-
Peas	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	x	-	-	-	
Peppers	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Potato	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	
Pulses	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	x	-	-	x	-	-	-	-	-	-	-	-	-	
Rice	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	x	-	-	-	-	-	-	-	
Salad cream	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	
Sausages	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	
Spices	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	
Tea	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Tobacco	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	o	o	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Toffee	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	
Tomatoes	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Triticale (rye)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	
Wheat	-	-	x	-	-	-	-	-	-	-	-	x	-	-	-	x	-	-	-	x	-	-	x	-	-	-	-	-	-	-	-	-	x	-	x	-	-	-	-	-	-	-	-	-	-	-
Wine	x	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	
Yogurt	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-

—, data not known; x, offline analysis; o, online analysis.

**Table 2** Overall quality parameters measured by NIR in food and animal feed

<i>Product</i>	<i>Quality parameter</i>
Cheese	Ripeness
Forage	Digestibility
Fruit, vegetables	Maturity
Fruit juices	Authenticity
Meat	Heat treatment
Milk	Pasteurization
Peas	Tenderometer readings
Seaweed (edible)	Quality
Spices	Gamma irradiation
Tea	Taster-defined quality
Wheat	Seed viability

In spite of the challenging nature of NIR analysis of food, applications of the technique are most common and diverse in this industry. **Table 1** lists NIR food applications both off- and online. **Table 2** lists those applications where NIR has been used to analyze for broad, relatively indefinable quality characteristics in food.

### Whole Fruit and Vegetable Analysis

Following on from the success with whole cereal grain analysis using NIR transmission technology, many workers have been using a variety of instrument types to obtain spectra from whole fruit and vegetables, nondestructively. Some success has been gained, for example, in measuring dry matter and sugar concentration in this way.

### Chemicals and Textiles

Quantitative analysis by NIR of petrochemicals dates back to the 1930s. Since then a range of NIR calibrations have been developed, for example, for octane number, methyl group analysis, and methanol content in petroleum. NIR has been used to monitor water, detergent solids, and glycerol in shampoo, and to analyze moisture and lubricant levels on polymer films. Process NIR spectrometers have been used to monitor naphtha composition and NIR instrumentation has been used to monitor ethylene polymerization.

For over a decade NIR quality control has been used in the textile and fiber industries. Perhaps the widest use of NIR has been in the cotton industry. Cotton blending, mercerization, and fiber maturity measurement have been used offline for rapid process control. The wool industry has also employed NIR to measure the residual grease after scouring. Calibrations have been developed for measuring moisture and heat set temperature in nylon yarn. Online NIR analysis is being developed for the quality control of the dyeing procedure for carpet yarn and for measuring yarn diameter.

### Pharmaceuticals and Medical Applications

Discriminant analysis of NIR spectra was used in 1986 to assay the level of lincomycin in a pharmaceutical formulation. This was the first NIR analysis to be accepted by this US Food and Drug Administration (FDA). NIR is now used to measure the salicylic acid content of aspirin and pharmaceutical companies use discriminant NIR procedures to check incoming raw materials for drug production. Sample identification can be achieved using at-line NIR fiber optic systems.

Antibiotic production in fermentations can now be monitored by NIR. A highly significant development in the medical field is the use of a handheld non-invasive NIR-based filter instrument for the monitoring of blood glucose levels for diabetics. Research is currently being undertaken to explore the feasibility of measuring human body temperature using NIR. Another potential medical application being examined is the measurement of oxygen levels in brain blood using a fiber optic probe placed on the skull. In the medical laboratory context, user-friendly fecal fat analysis is performed using NIR.

### Other Industrial Applications of NIR

An NIR method has been developed for measuring the hardwood content of bleached hardwood and the lignin content of unbleached hardwood pulp. Work has also been carried out using NIR reflectance spectra of hardwoods to discriminate between the different species. Using a Fourier transform NIR instrument carbonate measurements have been made on soil samples. NIR analysis of forest humus samples has provided satisfactory calibrations for microbial basal respiration, based on the organic polymer content of the humus. NIR analysis of the lake sediments for pH has been used to construct the lake water history.

The examples noted above are only a small fraction of the industrial applications of NIR, and the advent of solid-state instrumentation, will further widen the way in which NIR measurements can be made in an industrial context.

**See also:** **Chemometrics and Statistics:** Statistical Techniques; Multivariate Classification Techniques; Multivariate Calibration Techniques. **Food and Nutritional Analysis:** Overview. **Fourier Transform Techniques.** **Fuels:** Oil-Based. **Infrared Spectroscopy:** Overview. **Pharmaceutical Analysis:** Drug Purity Determination. **Process Analysis:** Overview. **Proteins:** Foods. **Quality Assurance:** Quality Control. **Textiles:** Natural; Synthetic.

### Further Reading

Burns DA and Ciurczak EW (eds.) (1992) *Handbook of Near Infrared Analysis*. New York: Dekker.

Davies AMC (2002) Is the end of the NIR 'calibration problem' in sight? *NIR NEWS* 13(5): 10–12.

Davies AMC and Cho RK (2002) Near infrared spectroscopy. *Proceedings of the 10th International Conference*, Kyongju, Korea. Chichester: NIR Publications.

Fearn T and Davies AMC (2003) A comparison of Fourier and wavelet transforms in the processing of near infrared

spectroscopic data: Part 1. Data compression. *Journal of Near Spectroscopy* 11: 3–15.

Osborne BG, Fearn T, and Hindle PH (1993) *Practical NIR Spectroscopy with Applications in Food and Beverages Analysis*. Harlow: Longman Scientific & Technical.

Williams P and Norris K (eds.) (1987) *Near-Infrared Technology in the Agricultural and Food Industries*, pp. 1–15. St Paul, MN: American Association of Cereal Chemists, Inc.

## Photothermal

**S E Bialkowski**, Utah State University, Logan, UT, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by J F Power, pp. 2219–2225, © 1995, Elsevier Ltd.

## Introduction

Photothermal spectroscopy is a class of optical analysis methods that measures heat evolved as a consequence of light absorption in an irradiated sample. In conventional spectrometric methods information is obtained by measuring the intensities of light transmitted, reflected, or emitted by the sample. In photothermal spectrometry, spectroscopic information is obtained by measuring the heat accompanying non-radiative relaxation. Because of the universality of the photothermal effect (e.g., heat evolution accompanies essentially all optical absorption), photothermal spectroscopy has diverse applications in chemistry, physics, biology, and engineering. Some applications and measurements in the analysis of solids are reviewed here.

## Theory

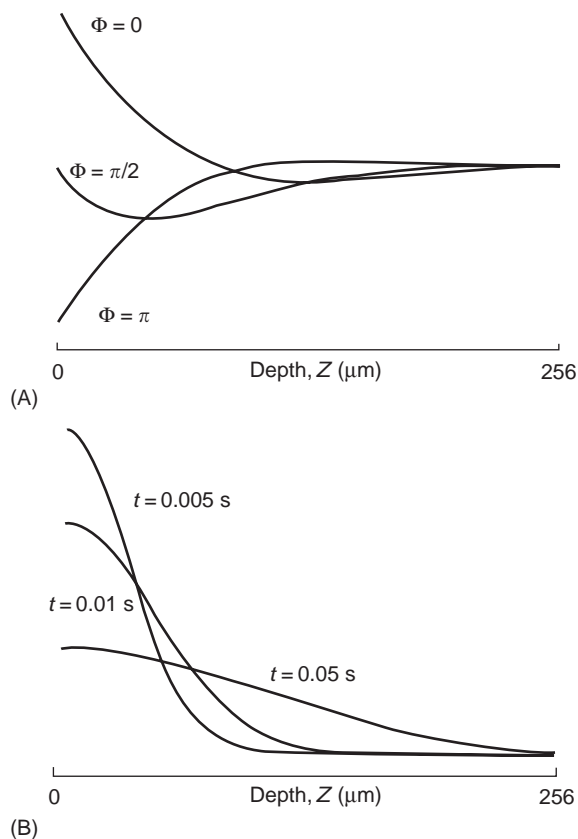
When a sample surface is heated with a periodically modulated beam, a thermal wave is generated which propagates away from the heated region. The thermal wave is a critically-damped temperature oscillation that, in a homogeneous material, decays exponentially with distance from the heated surface (Figure 1A). The thermal wave damping distance is the distance at which the temperature attenuates to 1/e of the value observed at the surface. This damping distance is controlled by varying the modulation frequency,  $\omega$  (rad/s), of the radiation source. For a sample of thermal diffusivity  $D$  ( $\text{m}^2/\text{s}$ ), the thermal wave damping distance,  $\mu$  (m), is given by

$$\mu = (2D/\omega)^{1/2} \quad [1]$$

With impulse heating, the time-dependent temperature profile below the heated surface has a Gaussian

dependence (Figure 1B) with a penetration distance,  $\mu_l$  (m), given by the time-dependent width of the Gaussian profile:

$$\mu_l = (2Dt)^{1/2} \quad [2]$$



**Figure 1** (A) Spatial dependence of a single-frequency (harmonically driven) thermal wave. Frequency 20.0 Hz, diffusivity  $1 \times 10^{-3} \text{ m}^2 \text{ s}^{-1}$ . (B) Spatial dependence of the thermal wave observed at various times after application of a heat pulse. Diffusivity  $1 \times 10^{-3} \text{ m}^2 \text{ s}^{-1}$ . The instantaneous phase  $\phi = \omega t$ . (Reprinted from Power JF (1993) Scanning probes III: Photoacoustic and photothermal imaging. In: Morris MD (ed.) *Microscopic and Spectroscopic Imaging of the Chemical State*, pp. 255–302. New York: Dekker, courtesy of Marcel Dekker, Inc.)



Davies AMC (2002) Is the end of the NIR 'calibration problem' in sight? *NIR NEWS* 13(5): 10–12.

Davies AMC and Cho RK (2002) Near infrared spectroscopy. *Proceedings of the 10th International Conference*, Kyongju, Korea. Chichester: NIR Publications.

Fearn T and Davies AMC (2003) A comparison of Fourier and wavelet transforms in the processing of near infrared

spectroscopic data: Part 1. Data compression. *Journal of Near Spectroscopy* 11: 3–15.

Osborne BG, Fearn T, and Hindle PH (1993) *Practical NIR Spectroscopy with Applications in Food and Beverages Analysis*. Harlow: Longman Scientific & Technical.

Williams P and Norris K (eds.) (1987) *Near-Infrared Technology in the Agricultural and Food Industries*, pp. 1–15. St Paul, MN: American Association of Cereal Chemists, Inc.

## Photothermal

**S E Bialkowski**, Utah State University, Logan, UT, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by J F Power, pp. 2219–2225, © 1995, Elsevier Ltd.

### Introduction

Photothermal spectroscopy is a class of optical analysis methods that measures heat evolved as a consequence of light absorption in an irradiated sample. In conventional spectrometric methods information is obtained by measuring the intensities of light transmitted, reflected, or emitted by the sample. In photothermal spectrometry, spectroscopic information is obtained by measuring the heat accompanying non-radiative relaxation. Because of the universality of the photothermal effect (e.g., heat evolution accompanies essentially all optical absorption), photothermal spectroscopy has diverse applications in chemistry, physics, biology, and engineering. Some applications and measurements in the analysis of solids are reviewed here.

### Theory

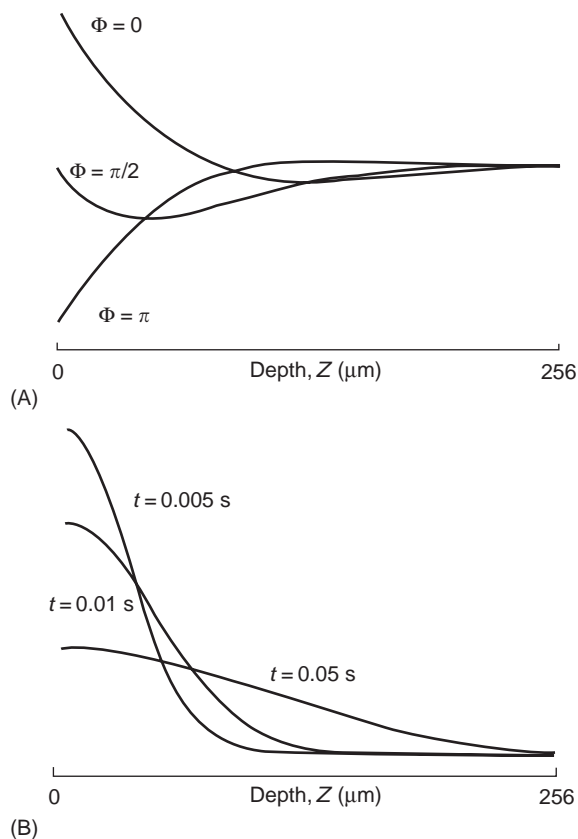
When a sample surface is heated with a periodically modulated beam, a thermal wave is generated which propagates away from the heated region. The thermal wave is a critically-damped temperature oscillation that, in a homogeneous material, decays exponentially with distance from the heated surface (Figure 1A). The thermal wave damping distance is the distance at which the temperature attenuates to 1/e of the value observed at the surface. This damping distance is controlled by varying the modulation frequency,  $\omega$  (rad/s), of the radiation source. For a sample of thermal diffusivity  $D$  ( $\text{m}^2/\text{s}$ ), the thermal wave damping distance,  $\mu$  (m), is given by

$$\mu = (2D/\omega)^{1/2} \quad [1]$$

With impulse heating, the time-dependent temperature profile below the heated surface has a Gaussian

dependence (Figure 1B) with a penetration distance,  $\mu_l$  (m), given by the time-dependent width of the Gaussian profile:

$$\mu_l = (2Dt)^{1/2} \quad [2]$$



**Figure 1** (A) Spatial dependence of a single-frequency (harmonically driven) thermal wave. Frequency 20.0 Hz, diffusivity  $1 \times 10^{-3} \text{ m}^2 \text{ s}^{-1}$ . (B) Spatial dependence of the thermal wave observed at various times after application of a heat pulse. Diffusivity  $1 \times 10^{-3} \text{ m}^2 \text{ s}^{-1}$ . The instantaneous phase  $\phi = \omega t$ . (Reprinted from Power JF (1993) Scanning probes III: Photoacoustic and photothermal imaging. In: Morris MD (ed.) *Microscopic and Spectroscopic Imaging of the Chemical State*, pp. 255–302. New York: Dekker, courtesy of Marcel Dekker, Inc.)

Photothermal methods may be used for depth profiling spectroscopic and/or thermal properties of condensed-phase samples because the sampling distance of the thermal waves may be controlled by varying the modulation frequency (as in eqn [1]) or the observation time (as in eqn [2]).

For photothermal methods that monitor surface temperature changes, heat generation must occur approximately within the thermal wave damping distance of the sample's surface, in order to contribute to the measured signal. As  $\mu$  increases in an optically transmitting sample, the measured absorption spectrum of the sample may change, as more deeply buried layers contribute to the measured signal.

Photothermal spectroscopy may be used to measure physical and chemical properties of sample as a function of depth. This 'depth profiling' may be thermal, spectroscopic, or a composite of the two. Thermal depth profiling is achieved by irradiating the sample at opaque wavelengths. The magnitude and phase of the surface temperature is measured at several modulation frequencies and subsurface properties are deduced by subsequent data analysis.

For surface heated inhomogeneous samples, thermal waves experience reflection or absorption when they encounter buried interfaces wherein there is a change in the thermal effusivity ratio:

$$b_{ij} = \frac{k_i}{k_j} \left( \frac{D_j}{D_i} \right)^{1/2} \quad [3]$$

Indices  $i$  and  $j$  refer to adjacent layers,  $k_i$  and  $k_j$  ( $\text{W m}^{-1}\text{s}^{-1}$ ) are thermal conductivities, and  $D_i$  and  $D_j$  are the thermal diffusivities. Thermal effusivity is analogous to a refractive index for thermal waves. Reflection and absorption may be evaluated by the thermal reflection coefficient:

$$s_{ij} = (b_{ij} - 1)/(b_{ij} + 1) \quad [4]$$

$s_{ij} = +1$  for total reflection and  $s_{ij} = -1$  for total absorption of (plane) thermal waves incident on an interface. Subsurface reflection or absorption may result in measurable surface temperature change. Surface temperature changes obtained by scanning an excitation beam across the sample surface may contain subsurface information limited to the depth of the thermal diffusion length. The subsurface thermal image may thus be obtained by analysis of the surface temperature change.

In photothermal spectroscopy, optical absorption coefficients are obtained from the depth dependence of the subsurface heat produced by light absorption. In high-absorbance samples, light attenuates rapidly at short distances below the surface. Heat generation

is thus localized within surface layers. Surface heating produces rapid arrival of the heat to that surface. In low-absorbance samples, more heat is generated below the surface and there is a delay in the arrival of the associated heat at the sample surface. Corresponding changes also occur in frequency-domain signals, especially the phase of the photothermal response, which, at constant frequency, lags increasingly as light absorption occurs at increasing depths.

In a pulsed measurement, a general relationship exists between the depth of an optical-absorbing feature and the time required for heat conduction to the surface. This principle is illustrated in Figure 2 for the case of pulsed excitation, with the temperature measured at the rear surface of a sample composed of three different absorbing layers.

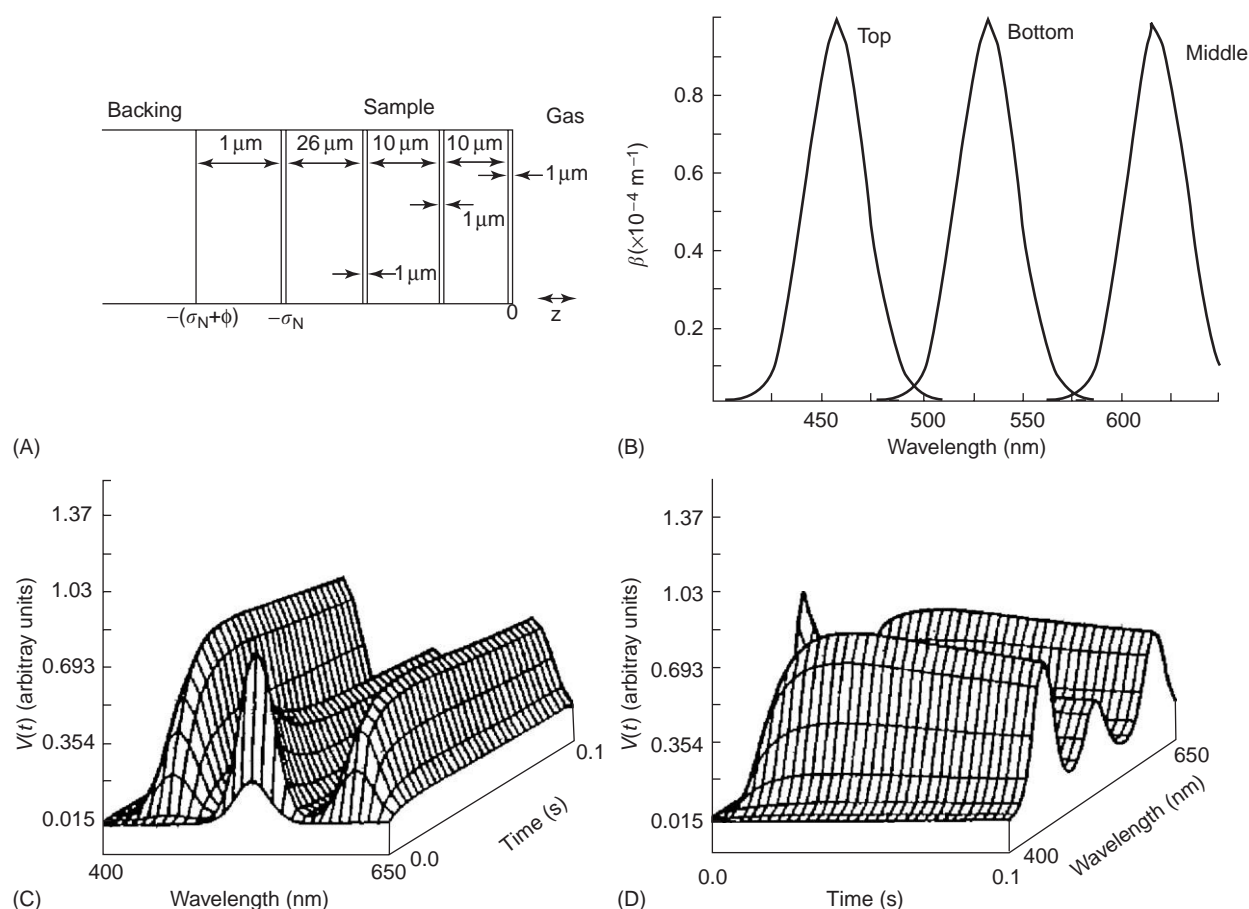
If the width of the irradiating beam is greater than the thicknesses of the layers probed and the variation of sample properties with distance is slow, the measured photothermal response data may be interpreted, in simple cases, by a theory derived by Rosencwaig and Gersho. In this case, the system is modeled as one-dimensional in optical absorbance and heat transport profiles. In this case, theoretical approaches for multiple layers are to be found in the literature. Continuous profiles of optical absorption or thermal diffusivity may be recovered from photothermal data by means of inversion methods.

## Instrumental Methods of Detection

Some processes that accompany the heat release from an irradiated condensed-phase sample are illustrated in Figure 3. The wavelength of this radiation may range from radiofrequency through to X-rays, although optical wavelength ranges are typically used. Heat may be detected directly by any one of a number of methods.

A direct detection strategy involves contacting a temperature sensor to the rear surface of the sample and monitoring the average temperature change in the sensor layer. Pyroelectric thin film sensor materials exhibit a polarization change that is approximately proportional to the average change in the temperature of the pyroelectric material. This change in polarization may be measured as a current or a voltage. This detection method, termed photopyroelectric spectrometry, has recently provided a rapid, sensitive method for the detection of thermal waves in very thin samples.

Heat conduction into the gas phase causes a modulated expansion of a layer of gas near the sample surface, producing a sound wave that can be detected using a microphone; the so-called photoacoustic effect. Thermally induced variations in the gas-phase



**Figure 2** Illustration of the theoretical photothermal response (C, D) at the rear surface of a multilayered sample (A). The optical absorption spectra of the various layers are given in (B). ( $V(t)$ , voltage response of the pyroelectric detector.) (Reprinted with permission from Power JF (1991) *Applied Spectroscopy* 45: 1240.)

density above the sample surface also produce refractive index gradients that may be probed through the deflection of an optical probe beam aligned parallel with the sample surface. This phenomenon is termed the mirage effect or photothermal deflection spectrometry.

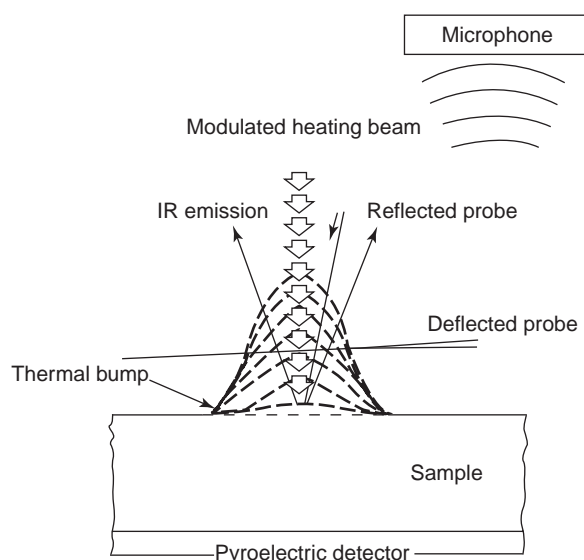
For reflective samples, temperature changes may cause variations in the sample's reflectance giving rise to photothermoreflectance spectrometry. Increased blackbody emission accompanies sample heating by the excitation beam. An infrared detector may be used to measure this blackbody emission signal and to detect temperature changes in the sample caused by optical absorption. This remote method of detection is termed photothermal radiometry.

Finally, heat generation in the sample gives rise to a thermoelastic effect that may be detected by aligning a probe beam at the inflection point of the 'thermal bump' that is generated using a focused excitation beam. The probe beam is reflected from the sample surface and its deflection is monitored using a position-sensing detector. The deflection signal is

directly proportional to the derivative of the surface deformation, which in turn is proportional to temperature.

Good depth profile resolution requires a method that exhibits high sensitivity in combination with a wide instrumental bandwidth. The bandwidth of available modulation frequencies, over which a particular method is capable of response, is determined both by the electrical bandwidth of the transducers used for detection and the optical excitation geometry. Optical energy in the wavelength range from the ultraviolet to the IR may be used for signal excitation. Methods such as the mirage effect and photoacoustic spectroscopy are readily adaptable to detection using Fourier transform infrared (FTIR) spectroscopy.

It is possible to perform depth-resolved spectroscopy using these FTIR photothermal methods. Two commonly available instrumentation designs are used to perform FTIR photothermal spectroscopy. In constant scan rate FTIR photothermal instruments, the uniform motion of the movable interferometer mirror



**Figure 3** Schematic diagram showing processes accompanying heating of a sample by a modulated beam of radiation. (Reprinted from Power JF (1993) *Scanning probes III: Photoacoustic and photothermal imaging*. In: Morris MD (ed.) *Microscopic and Spectroscopic Imaging of the Chemical State*, pp. 255–302. New York: Dekker, courtesy of Marcel Dekker, Inc.)

causes a modulation of all infrared frequencies, each at a different modulation frequency. This produces difficulties in interpreting the measured photothermal signals. In step-scan FTIR photothermal spectroscopy, independent fixed modulation is imposed on the IR beam at each stationary position of the interferometer moving mirror as it is stepped through a scan. This greatly improves the ease of signal interpretation but requires more sophisticated instrumentation designs to ensure stable, repeatable interferometry measurements. Both intensity and wavelength modulation are used.

## Applications

Photothermal methods are used in a variety of applications that involve measurements of the thermal and optical properties of thin specimens. As materials are subjected to processes such as annealing, cross-linking, extrusion, photodegradation, diffusion, etc., the thermal diffusivity of the sample may change significantly, especially in near-surface regions. In many cases,  $D$  may decay strongly with depth below the surface in a bulk sample, or spatially. Thermal conductivity is especially sensitive to these physical changes, although variations in density and heat capacity may also be involved. Variations in thermal conductivity generally give rise to primary changes in the thermal efflux ratio (eqn [3]), which

can be imaged spatially by photothermal techniques. Conventional thermoanalytical methods, in contrast, probe only the average properties of the bulk sample.

Anisotropy in thermal diffusivity provides the basis for a novel and interesting photothermal imaging method. Fibrous materials and crystals, for example, may have a preferred axis for heat conduction. In materials that have been subjected to tensile loading, an enhanced thermal conductivity may lie along the stretch axis.

Infrared photothermal radiometry can be used to measure thermal anisotropy, by displacing excitation and detection spots on the sample, and measuring the transverse heat flow between these spots. As the direction of the spot displacement is varied, directional variations are sampled. Entire images of the thermal anisotropy may be scanned using this method. Such images may be used to detect defects in aligned materials, where a defect destroys the local alignment of bulk fibers or molecular chains.

A number of photothermal applications exist in the remote inspection and evaluation of materials. Video radiometry has been used for rapid inspection of large area samples such as sheet metal panels, vehicle bodies, or engine components. Metallurgical samples give good examples of materials that are routinely subjected to a variety of processing stages that may include hardening, annealing, mechanical deformation, etc. Photothermal radiometry has been used to inspect welding seams and to estimate seam thicknesses using a single-ended measurement.

Microstructural changes produced in mechanically deformed metals produce significant variations in their thermal properties, which can be detected using photothermal imaging.

Modulated thermorefectance and thermoelastic effect imaging have found important applications in the nondestructive inspection of semiconductors.

Thermal wave imaging may be used to inspect ion implantations, to evaluate defects in silicon, and to acquire depth-dependent images of integrated circuit features, or as a general probe for the study of non-radiative processes in semiconductors. Devices may be inspected in active mode (e.g., with current flowing).

Photothermal imaging has also found applications in the study of practical industrial materials, including fibers, ceramics, and composites, where the depth dependence of the recovered image may be useful in identifying surface features due to mechanical or oxidative processes, or subsurface defects and delamination.

The depth profiling capabilities available with photothermal methods have found important applications in biology and biophysics. Photoacoustic and

photothermal spectroscopy have evolved in recent years as general tools for the study of photosynthesis and energy conversion by plant photosystems.

Depth profile analysis using photothermal methods has contributed a number of nondestructive methods for studying the spectroscopy of plant tissues. The phase rotation method of signal recovery has been used to examine a number of specimens of botanical interest, including leaves, tissues, and lichens.

Tissues such as skin have also been studied using photothermal methods, especially IR radiometry, which is capable of noncontact measurement, and is robust to alignment instabilities. In addition to the thermal and optical depth profiling of the various layers of the skin, photothermal spectroscopy has been used to study the time-dependent penetration of topically applied cosmetics and sunscreens below the skin surface. Photothermal studies have been used to monitor the time for which a topically applied film exists as a discrete phase on the outer surface of the skin.

Other examples of tissues that have been investigated with photothermal methods include ocular and arterial tissues. IR photothermal radiometry, in particular, shows promise as a general diagnostic tool in biomedical studies, in the nondestructive imaging of very thin layers.

See also: **Fourier Transform Techniques.**

## Further Reading

- Aamodt LC, Murphy JC, Royce BSH, and Machlachlan-Spicer JW (eds.) (1988) *Photoacoustic and Photothermal Phenomena II*. Berlin: Springer.
- Bertrand L, Cielo P, LeBJanc R, Monchalain JP, and Mongeau B (eds.) (1986) Proceedings of the Fifth International Topical Meeting on Photoacoustic, Thermal and Related Phenomena, Esterel, Quebec, Canada, 1985. *Canadian Journal of Physics* 64.
- Bialkowski SE (1996) *Photothermal Spectroscopy Methods for Chemical Analysis*. New York: Wiley.
- Hess P and Pelzl J (eds.) (1988) *Photoacoustic and Photothermal Phenomena*. Berlin: Springer.
- Hutchins DA and Tam AC (eds.) (1986) *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. UFFC-33, no. 5 (special issue devoted to photoacoustic and photothermal sciences).
- Mandelis A (ed.) (1987) *Photoacoustic and Thermal Wave Phenomena in Semi-Conductors*. New York: North-Holland.
- Mandelis A and Michaelian KH (1997) *Optical Engineering*, vol. 36, no. 2 (special issue devoted to Photoacoustic and photothermal science).
- Power JF (1989) Thermal wave imaging in the physicochemical analysis and evaluation of solid-phase samples. *Progress in Analytical Spectroscopy* 12: 453–506.
- Scudieri F and Bertolotti M (eds.) (1999) *Photoacoustic and Photothermal Phenomena AIP Conference Proceedings*, p. 463. New York: AIP Press.
- Sell JA (1989) *Photothermal Investigations of Solids and Fluids*. New York: Academic Press.

## Industrial Applications

**S F Parker**, Rutherford Appleton Laboratory, Chilton, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Infrared spectroscopy is widely used in industry. The applications range from quality control and quality assurance of raw materials to customer complaints (troubleshooting) to quantitative analysis and online process monitoring and control. Most of the instrumentation is based on a Fourier transform infrared (FTIR) spectrometer but specialist applications use instruments as diverse as a nondispersive infrared detector through to mid-infrared diode lasers.

In this section, the emphasis will be on applications that make use of the power of infrared spectroscopy to give information about the molecular

species present. The huge advantage in signal-to-noise ratio (per unit time) of FTIR spectrometers over dispersive instruments means that virtually all industrial applications use FTIR. This means that techniques that result in low signal levels, such as attenuated total internal reflectance or, especially, infrared microscopy, can generate spectra of sufficient signal-to-noise ratio in a reasonable time as to be viable analysis methods.

### Sample Types

The range of samples that are encountered in industry spans all three states of matter: solid, liquid, and gaseous. The key point is that, with sufficient effort, it is possible to obtain an infrared spectrum from most materials. For instance, while a metal does not exhibit an infrared spectrum, the oxide layer present at the surface can be characterized by infrared



photothermal spectroscopy have evolved in recent years as general tools for the study of photosynthesis and energy conversion by plant photosystems.

Depth profile analysis using photothermal methods has contributed a number of nondestructive methods for studying the spectroscopy of plant tissues. The phase rotation method of signal recovery has been used to examine a number of specimens of botanical interest, including leaves, tissues, and lichens.

Tissues such as skin have also been studied using photothermal methods, especially IR radiometry, which is capable of noncontact measurement, and is robust to alignment instabilities. In addition to the thermal and optical depth profiling of the various layers of the skin, photothermal spectroscopy has been used to study the time-dependent penetration of topically applied cosmetics and sunscreens below the skin surface. Photothermal studies have been used to monitor the time for which a topically applied film exists as a discrete phase on the outer surface of the skin.

Other examples of tissues that have been investigated with photothermal methods include ocular and arterial tissues. IR photothermal radiometry, in particular, shows promise as a general diagnostic tool in biomedical studies, in the nondestructive imaging of very thin layers.

See also: **Fourier Transform Techniques.**

## Further Reading

- Aamodt LC, Murphy JC, Royce BSH, and Machlachlan-Spicer JW (eds.) (1988) *Photoacoustic and Photothermal Phenomena II*. Berlin: Springer.
- Bertrand L, Cielo P, LeBlanc R, Monchalain JP, and Mongeau B (eds.) (1986) Proceedings of the Fifth International Topical Meeting on Photoacoustic, Thermal and Related Phenomena, Esterel, Quebec, Canada, 1985. *Canadian Journal of Physics* 64.
- Bialkowski SE (1996) *Photothermal Spectroscopy Methods for Chemical Analysis*. New York: Wiley.
- Hess P and Pelzl J (eds.) (1988) *Photoacoustic and Photothermal Phenomena*. Berlin: Springer.
- Hutchins DA and Tam AC (eds.) (1986) *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. UFFC-33, no. 5 (special issue devoted to photoacoustic and photothermal sciences).
- Mandelis A (ed.) (1987) *Photoacoustic and Thermal Wave Phenomena in Semi-Conductors*. New York: North-Holland.
- Mandelis A and Michaelian KH (1997) *Optical Engineering*, vol. 36, no. 2 (special issue devoted to Photoacoustic and photothermal science).
- Power JF (1989) Thermal wave imaging in the physicochemical analysis and evaluation of solid-phase samples. *Progress in Analytical Spectroscopy* 12: 453–506.
- Scudieri F and Bertolotti M (eds.) (1999) *Photoacoustic and Photothermal Phenomena AIP Conference Proceedings*, p. 463. New York: AIP Press.
- Sell JA (1989) *Photothermal Investigations of Solids and Fluids*. New York: Academic Press.

## Industrial Applications

**S F Parker**, Rutherford Appleton Laboratory, Chilton, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Infrared spectroscopy is widely used in industry. The applications range from quality control and quality assurance of raw materials to customer complaints (troubleshooting) to quantitative analysis and online process monitoring and control. Most of the instrumentation is based on a Fourier transform infrared (FTIR) spectrometer but specialist applications use instruments as diverse as a nondispersive infrared detector through to mid-infrared diode lasers.

In this section, the emphasis will be on applications that make use of the power of infrared spectroscopy to give information about the molecular

species present. The huge advantage in signal-to-noise ratio (per unit time) of FTIR spectrometers over dispersive instruments means that virtually all industrial applications use FTIR. This means that techniques that result in low signal levels, such as attenuated total internal reflectance or, especially, infrared microscopy, can generate spectra of sufficient signal-to-noise ratio in a reasonable time as to be viable analysis methods.

### Sample Types

The range of samples that are encountered in industry spans all three states of matter: solid, liquid, and gaseous. The key point is that, with sufficient effort, it is possible to obtain an infrared spectrum from most materials. For instance, while a metal does not exhibit an infrared spectrum, the oxide layer present at the surface can be characterized by infrared

reflectance methods; while hydrogen gas does not have an infrared spectrum at ambient pressure, it does when chemisorbed into a zeolite.

Solid samples can be a bulk material such as a polymer or a raw material, they can be a surface coating such as varnish buildup on a piston or a particle on a semiconductor wafer. Liquids can be pure or as a solution or mixture. Gases can be pure or a mixture, such as in stack gases; they may also be very dilute, down to  $\text{mg l}^{-1}$  (parts per million) level for trace gas analysis such as for atmospheric monitoring or breathing gas for divers. In addition to these possibilities, it may be necessary to carry out the analysis at temperature and pressure conditions well removed from ambient. A final complication is that pure samples are rarely encountered; much more common are mixtures, often with the material of interest present as the minor component.

## Sample Preparation

In industry, time is at premium so the emphasis is always to carry out the analysis in the most efficient manner possible. In practice, this means that the analyst employs the sampling method that requires the least possible sample preparation. Historically, the great majority of liquid samples were run as thin films between infrared transparent plates (usually KBr) or in fixed pathlength cells. Similarly, most solid samples were run as KBr discs, as mulls or as a thin film deposited from solution onto a suitable substrate. Polymers were usually hot-pressed to make a self-supporting film, with a thickness typically between 0.020 and 1.0 mm; the former being used for the identification of a bulk material and the latter for the identification of minor components, contaminants, or degradation products.

Sample preparation was greatly simplified in the late 1990s by the introduction of commercial diamond attenuated total internal reflection (ATR) systems. These use a small diamond as a single-pass ATR element held in a metal plate. The sample is placed on the diamond, a plate is then brought down onto it and clamped in position. This ensures excellent contact between the ATR element and the sample, so immediately eliminates the problem of poor contact that made ATR so problematical for use with solids, particularly powders. The accessory can be permanently installed in the purged sample compartment of the spectrometer, so there is only a very short pathlength that is in the open air. This means that there is no need to wait for the sample compartment to purge to reduce the atmospheric water and carbon dioxide absorption. The result is that the sample can be placed in the accessory and scanned immediately,

so that even unskilled operators can obtain reliable results. The method is readily applicable to all solid and semisolid materials including powders, pellets, sheets or film, pastes, and gels. Liquids are readily studied and the inertness of diamond also allows corrosive materials to be examined. The only disadvantage is the small number of reflections, and hence pathlength, which means that only the major components will be seen; thus, it is unsuitable for systems that are dilute in the material of interest.

However, there are cases where more involved methods have to be used. An example would be a contaminant in a piece of plastic piping. If the particle is large enough ( $>0.1$  mm) then it can be physically separated for subsequent analysis. Smaller particles are extremely difficult to manipulate and it is advantageous to examine them *in situ* if possible. It is then necessary to make a thin section, either by using a scalpel or, better, by microtoming the material followed by examination in an infrared microscope. This is an optical microscope that is interfaced to an FTIR spectrometer such that the visible and infrared radiation follow identical paths through the sample. The use of visible microscopy allows the area of interest to be located and then isolated by aperturing in an image plane. A mirror-flip then directs the infrared light along the same path. The instrument usually has a dedicated high sensitivity detector to improve the signal-to-noise ratio. An infrared microscope is now an essential part of the equipment in most industrial chemical laboratories.

Other methods that are used less frequently include diffuse reflectance and photacoustic spectroscopy. These are treated in detail elsewhere in this encyclopedia and each have their own sampling problems. In favorable cases it is possible that no sample preparation is required; more often, better results are obtained if the sample is ground to a fine powder. This often requires a mechanical grinder, sometimes at liquid nitrogen temperature.

## Applications

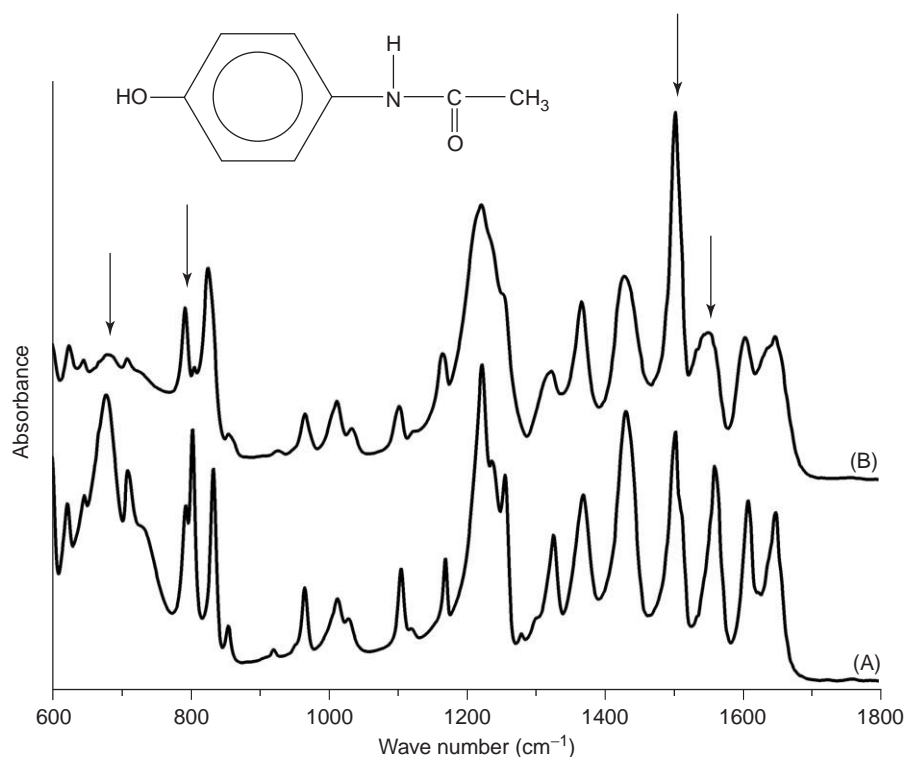
Infrared spectroscopy has two main uses in industry: qualitative and quantitative analysis. Qualitative analysis may range from the identification of contaminants in products, to the analysis of a competitor's product, to the study of reaction mechanisms. Quantitative analysis may be used in any physical state for components that may be the bulk of the sample to those present in parts per million concentrations. In the remainder of this section examples of both classes of analysis will be discussed; these are intended to give an appreciation for the approach that is used and the type and range of problems that can be addressed.

A simple type of qualitative analysis is the question; 'are these two materials the same?' This is a key question in the pharmaceutical industry where the occurrence of polymorphism is rife. Polymorphism is the existence of two or more crystal forms of the same compound but with distinct structures. To ensure patent protection of a new drug, all polymorphs need to be characterized. Polymorphs of the same compound may have different physical properties such as solubility and hence bioavailability. Paracetamol (4'-hydroxyacetanilide, see **Figure 1**) is a widely used drug for the relief of mild pain and reduction of fever. It has two important polymorphs, which are the commercially used Form I (monoclinic) and Form II (orthorhombic). The latter, unlike commercial paracetamol (Form I), undergoes plastic deformation and is suitable for direct compression. Consequently, Form II has attracted much interest because of the potential commercial benefits to be gained by not using binders during the manufacture of tablets. As **Figure 1** shows, the two polymorphs are readily distinguished by infrared spectroscopy and this is often the case with polymorphs.

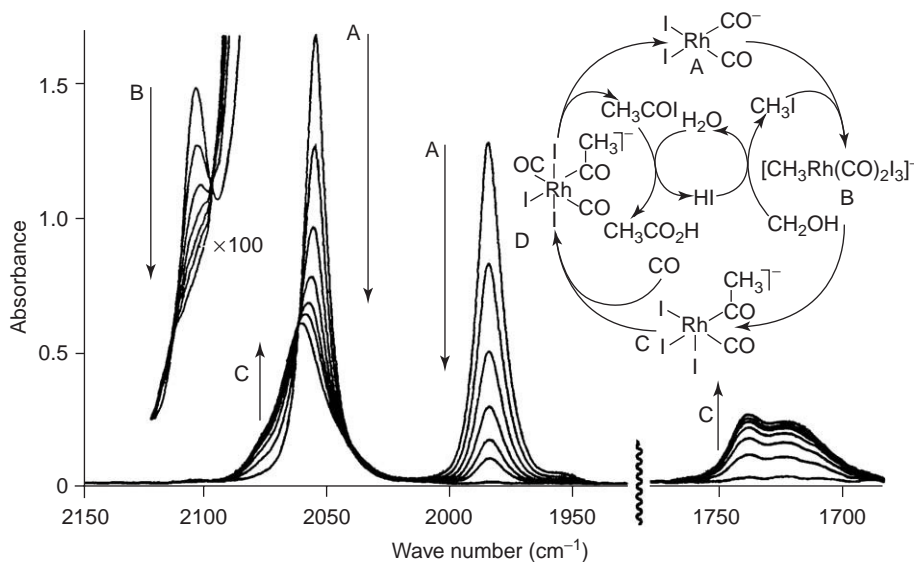
A second area where this type of question is of crucial importance is in forensic science. Infrared spectroscopy and, particularly, infrared microscopy are key techniques used for identification and comparison of fibers and paint types.

Infrared spectroscopy is also used to answer much more complex questions. It has long been used to investigate heterogeneous and homogeneous catalytic processes because it is often possible to study the reaction under industrially relevant conditions of temperature and pressure. The low pressure rhodium and iodide catalyzed carbonylation of methanol to acetic acid is a major industrial process with a large proportion of the annual production of 5 000 000 tons of acetic acid being made by this route. The overall mechanism is known (see **Figure 2**), but rates for all the individual steps were not. Infrared spectroscopy was able to show that the oxidative addition of  $\text{CH}_3\text{I}$  with the starting complex A was the rate determining step, since the resulting intermediate B decays at the same rate as A showing that the migratory insertion of CO into the  $\text{Rh}-\text{CH}_3$  bond is fast. Using a combination of *in situ* infrared and  $^{13}\text{C}$  nuclear magnetic resonance spectroscopies, it was possible to measure rate constants for all the steps and activation parameters for the migratory insertion.

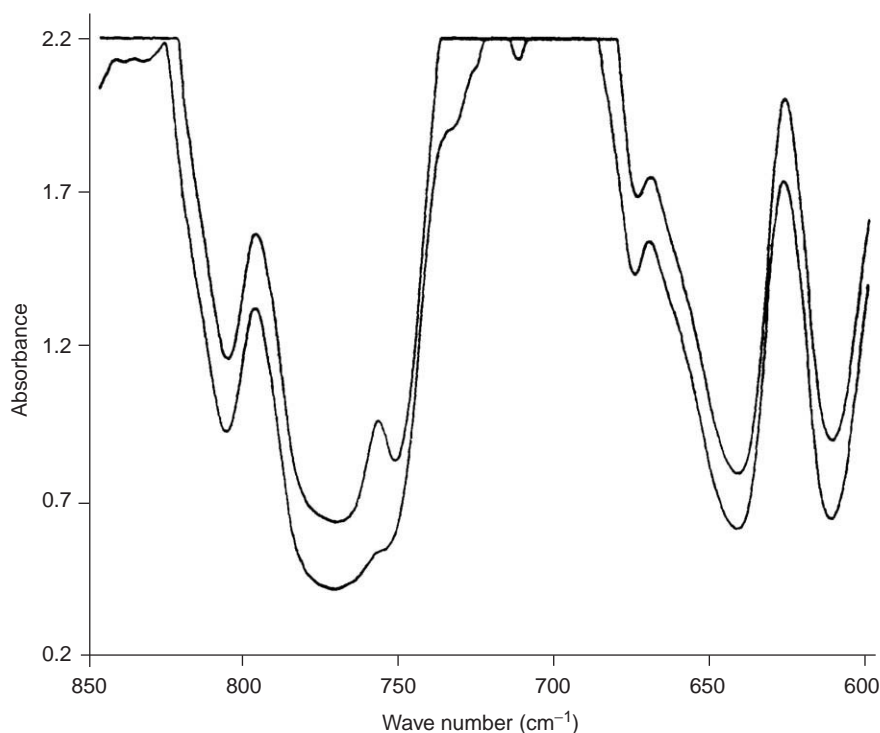
In the polymer industry a recurring problem is the presence of particles in the product. An example is shown in **Figure 3**. The particle was  $\sim 0.1$  mm in diameter and was embedded in a poly(aryl ether sulfone) film. The spectra are very similar, showing that the particle is not an inorganic contaminant. The decreased absorbance at  $760\text{ cm}^{-1}$ , shows that the



**Figure 1** Infrared spectra of (A) Form I and (B) Form II of paracetamol (4'-hydroxyacetanilide), bands indicated by arrows are significantly different in intensity or position in Form II as compared to Form I.



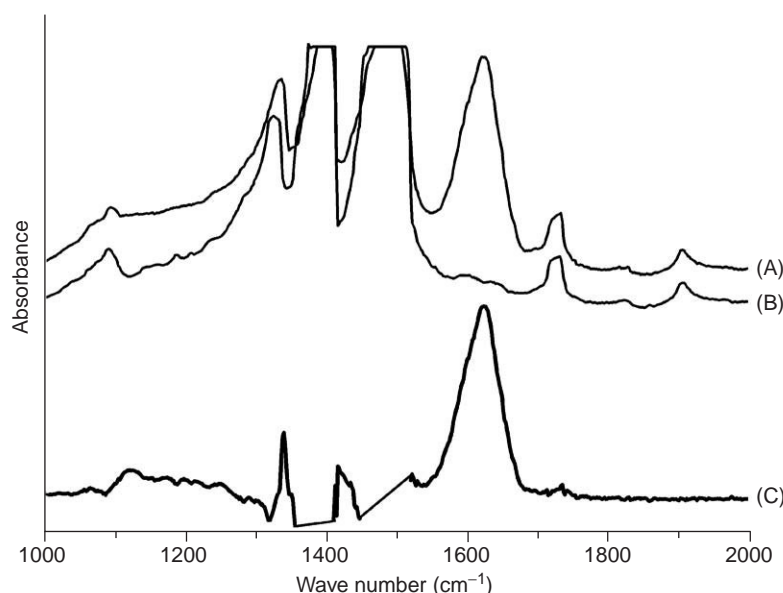
**Figure 2** Series of infrared spectra showing the reaction of  $[\text{Bu}_4\text{N}][\text{Rh}(\text{CO})_2\text{I}_2]$  in neat  $\text{CH}_3\text{I}$ . Note that the weak peak at  $2104\text{ cm}^{-1}$  assigned to the key intermediate B decays at the same rate as the starting complex A. The bands of C at  $2061$  and  $1740\text{ cm}^{-1}$  are complex because of a mixture of isomers. The inset shows the catalytic cycle. (Reprinted with permission from Haynes A, Mann BE, Morris GE, and Maitlis PM (1993) Mechanistic studies on rhodium-catalysed carbonylation reactions: Spectroscopic detection and reactivity of a key intermediate  $[\text{CH}_3\text{Rh}(\text{CO})_2\text{I}_3]^-$ . *Journal of the American Chemical Society* 115: 4093–4100; © American Chemical Society.)



**Figure 3** Offset infrared spectra of a microtomed section of poly(aryl ether sulfone) film. Top: normal film, bottom: particle in film. (Reproduced with permission from Chalmers JM and Everall N (1995) The role of vibrational spectroscopy-microscopy techniques in polymer characterisation. *Macromolecular Symposia* 9: 33–49; © Wiley-VCH.)

particle is deficient in the aryl-Cl end-group. The most likely explanation is that the particle is of higher molecular weight, which has a different rheology during processing.

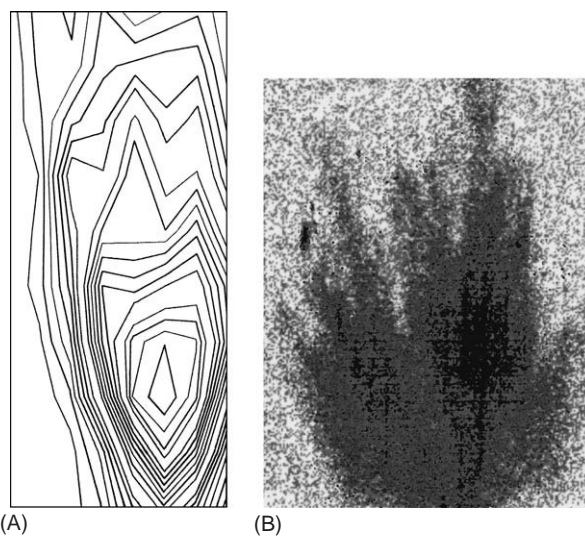
A better understanding of a system can often be gained by studying the spatial distribution of a component. Infrared microscopy is a very powerful tool for such investigations. In this case, the spectrum of



**Figure 4** FTIR microscopy of polyethylene cable insulation. (A) Water-tree, (B) undamaged area, and (C) the difference spectrum (A) – (B). (Parker SF (1995) Industrial applications of vibrational spectroscopy and the role of the computer. In: George WO and Steele D (eds.) *Computing Applications in Molecular Spectroscopy*, pp. 181–199. Cambridge: The Royal Society of Chemistry; reproduced by permission of The Royal Society of Chemistry.)

the sample is recorded at a large number of points and a map is generated from the spectra by plotting the intensity of a spectral feature as a function of its position. This process can be carried out manually; however, it is much better done automatically using an XY stage interfaced to the spectrometer's computer. This process is relatively slow and to map an area of  $1 \text{ mm}^2$  with a relatively coarse step size of  $0.1 \text{ mm}$  requires collection of 100 spectra, which will take at least an hour. Since most of the infrared beam is lost at the apertures that define the spot size, the signal is weak, so longer measurement times are needed and often the best solution is an overnight run.

An example of the utility of infrared mapping is shown in Figures 4 and 5. A failure mode of cross-linked polyethylene used for electrical cable insulation is water-treeling. Water-trees are damaged areas that when boiled in water and imaged with an optical microscope have a dendritic appearance. Figure 4 shows infrared spectra recorded from a microtomed section of a water-tree; Figure 4A is from the water-tree, Figure 4B from an undamaged area, and the difference spectrum is shown in Figure 4C. A strong band at  $\sim 1600 \text{ cm}^{-1}$  in Figure 4C suggests the presence of carboxylate. The region of the water-tree was then mapped and the ratio of the  $1600 \text{ cm}^{-1}$  band to that of a polyethylene band at  $2020 \text{ cm}^{-1}$  (to eliminate the effect of thickness variations in the section) plotted, as shown in Figure 5A. If the  $1600 \text{ cm}^{-1}$  band is associated with carboxylate formation, then there must be a counterion. Proton



**Figure 5** (A) Contour plot of the distribution of carboxylate ions in the water-tree by FTIR microscopy. (B) Proton-induced X-ray emission image of the sample showing the distribution of potassium. Note the similarity to (A). (Parker SF (1995) Industrial applications of vibrational spectroscopy and the role of the computer. In: George WO and Steele D (eds.) *Computing Applications in Molecular Spectroscopy*, pp. 181–199. Cambridge: The Royal Society of Chemistry; reproduced by permission of the The Royal Society of Chemistry.)

induced X-ray emission of the sample showed the presence of potassium with a distribution very similar to that of the carboxylate as shown in Figure 5B, providing a plausible counterion and confidence in the assignment to carboxylate. The results suggest



that water-treeing results from localized oxidation that changes the polymer from hydrophobic to hydrophilic. Water and ions can travel along and condense into these hydrophilic paths; repetition of the process propagates the water-tree as a myriad of such channels.

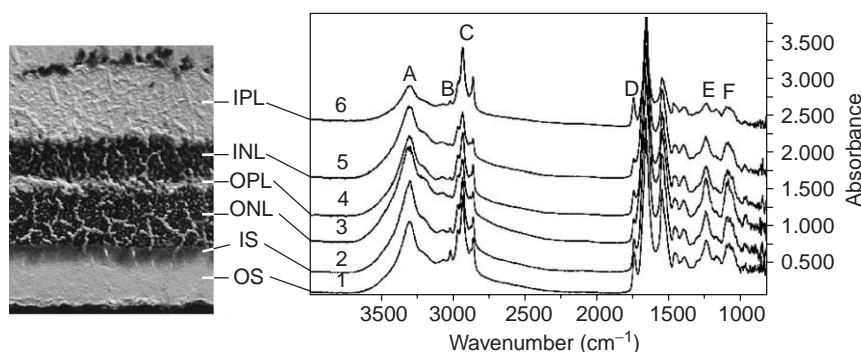
This example indicates two recurrent themes in industrial infrared analysis: the use of computer subtraction to look for differences between two (or more) samples is probably the most commonly used form of data manipulation in infrared spectroscopy. The second feature is that one technique in isolation is rarely capable of providing the entire answer, much more often data from a variety of sources must be combined in order to arrive at the solution.

The major area of growth in the use of vibrational spectroscopy in recent years has been in the application to biology. The identification of bacteria and other microbiological organisms is well established. Over 800 yeast strains can be identified with high certainty and bacteria can be classified and identified by infrared methods.

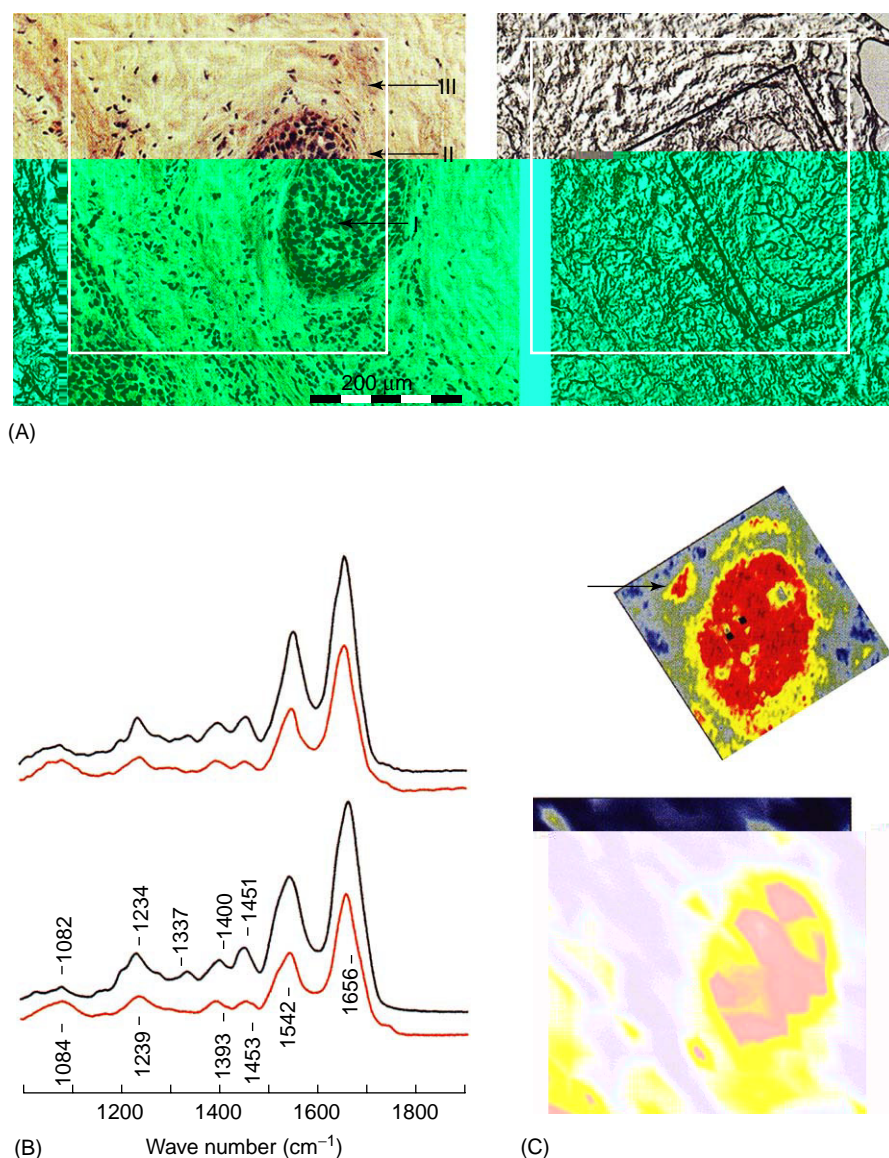
There is enormous interest in the use of infrared microscopy to distinguish tissue types. As an example, **Figure 6** shows a section through the retina of a pigmented rat and the corresponding spectra. The bands labeled A–F have different relative intensities in the spectra. The outer segments (OS, spectrum 1) of retinas from pigmented animals were found to have unusually strong absorption values for C=C–H unsaturation (band B) and carbonyl functional groups (band D). Docosahexaenoic acid ( $\text{CH}_3(\text{CH}_2\text{CH}=\text{CH})_6\text{CH}_2\text{CH}_2\text{CO}_2\text{H}$ ) is a major constituent of lipids in the outer segments which suggests that it is responsible for those enhanced absorption

values. Absorbance values for the unsaturation and carbonyl functional groups were substantially reduced in the outer segments of retinas from albino animals. This finding, together with data from other studies on light-induced oxidative events in the retina, indicates a loss of docosahexaenoic acid by a light-induced mechanism in albino animals. The outer nuclear layer (ONL, spectrum 3) had strong absorbance values for H–C–OH (band F) and P=O functional groups (band E), which are assigned to the sugar phosphate backbone of DNA. The outer (OPL, spectrum 4) and inner plexiform layers (IPL, spectrum 6) were found to contain greater concentrations of  $\text{CH}_2$  (band C) and C=O functional groups (band D) than the outer (ONL, spectrum 3) and inner nuclear layers (INL, spectrum 5), which is due to the high concentration of synaptic connections in the former layers. Thus, there is a unique chemical profile in the outer segments compared to other retinal layers, and this profile is different in albino animals.

The challenge is to be able to distinguish normal, precancerous, and cancerous cells. This has proven very difficult and there have been a number of false starts. Part of the problem is that cancerous and normal cells are of different size, so scanning a section with a fixed aperture does not result in spectra that are representative of the various types. To overcome this problem, attention has switched to infrared imaging rather than mapping. An imaging spectrometer uses a focal plane array of detectors, usually  $64 \times 64$  detectors, each  $0.055 \times 0.055$  mm. Each pixel in the acquired image corresponds to a region in the sample of  $0.005 \times 0.005$  mm, which is smaller than a single cell. It is also close to the diffraction limit, so caution in interpreting the spectra is needed.



**Figure 6** A series of spectra (right) from the various layers of a pigmented rat retina (left). Spectra were collected from photoreceptor outer segments (OS, spectrum 1), inner segments (IS, spectrum 2), outer nuclear layer (ONL, spectrum 3), outer plexiform layer (OPL, spectrum 4), inner nuclear layer (INL, spectrum 5), and inner plexiform layer (IPL, spectrum 6). Note the elevated absorbance values for C=C–H (B),  $\text{CH}_2$  (C):NH (A), C=O (D) in the outer segments and the elevated absorbance values for P=O (E) and H–C–OH (F) in the outer nuclear layer. The  $\text{CH}_2$ :NH ratio is also large in the inner plexiform layer. (Reprinted with permission from LeVine SM, Radcliff JD, Sweat JA, and Wetzel DL (1999) Microchemical analysis of retina layers in pigmented and albino rats by Fourier transform infrared microspectroscopy. *Biochimica et Biophysica Acta-General Subjects* 1473: 409–417; © Elsevier.)



**Figure 7** (A) A stained (left) and unstained (right) cryosection of breast tissue with ductal carcinoma. The unstained section was mapped using the point-by-point method (area in white square) and by the imaging method (area in black square). Spectra of the normal (upper spectrum of each pair) and tumor (lower spectrum of each pair) areas by the imaging method (upper pair) and by the point-by-point method (lower pair) are shown in 'B' and the corresponding images (upper: imaging, lower: point-by-point) in 'C'. The images are based on the peak intensity at  $1084\text{ cm}^{-1}$ , normalized to the integrated intensity between  $1000$  and  $1900\text{ cm}^{-1}$ . (Reproduced with permission from Fabian H, Lasch P, Boese M, and Haensch W (2002) Mid-IR microspectrometric imaging of breast tumour tissue sections. *Biopolymers* 67: 354–357; © John Wiley & Sons Ltd.)

Figure 7A shows a stained (left) and unstained (right) cryosection of breast tissue with ductal carcinoma. The unstained section was mapped in the point-by-point method (area in black square) and by the imaging method (area in white square). Spectra of the tumor and normal areas by both methods are shown in Figure 7B and the corresponding images in Figure 7C. The images are based on the peak intensity at  $1084\text{ cm}^{-1}$ . Both methods detect the presence of the tumor; however, additional fine structure is revealed in the focal plane array image. This is

because the aperture size used for the mapping method was  $0.030\text{ mm}$ , so many pixels in the image are a composite of the spectra of tumor + normal cells.

Quantitative analysis can take many forms but the basis of most analyses is the Beer–Lambert law:

$$A = \log_{10}(I_0/I) = \epsilon cl \quad [1]$$

where  $A$  is the absorbance of the band,  $I_0$  and  $I$  are the incident and transmitted intensities, respectively,  $\epsilon$  is the absorptivity,  $c$  is the concentration, and  $l$  is

the pathlength. The major difficulty is that the concentration of a species is related to its spectral intensity (at a given wavenumber) by the absorptivity, which is wavenumber dependent. While it is possible to calculate these by *ab initio* methods, in practice this is only possible for small molecules containing light atoms. This means that standards have to be generated by an independent method. This frequently causes considerable difficulties because the necessary reference materials are not available.

One area that quantitative infrared analysis is used routinely is in the semiconductor industry. Major applications include the determination of impurities, the quantitation of dopants, and epitaxial thickness measurements. Oxygen and carbon measurements in silicon are carried out routinely and at least two FTIR spectrometer manufacturers supply small dedicated instruments for this purpose. Oxygen atoms in interstitial sites absorb at  $1107\text{ cm}^{-1}$  and carbon atoms in substitutional sites in the silicon lattice at  $605\text{ cm}^{-1}$ . The bands are readily detected in a difference spectrum between the sample wafer and a high-purity reference wafer grown by the float-zone method. Considerable effort has gone into making these measurements quantitative with the result that they can be routinely determined down to  $5 \times 10^{16}\text{ atoms cm}^{-3}$ . At liquid helium temperatures the oxygen bands sharpen considerably, improving the detection limits but at a considerable cost in complexity.

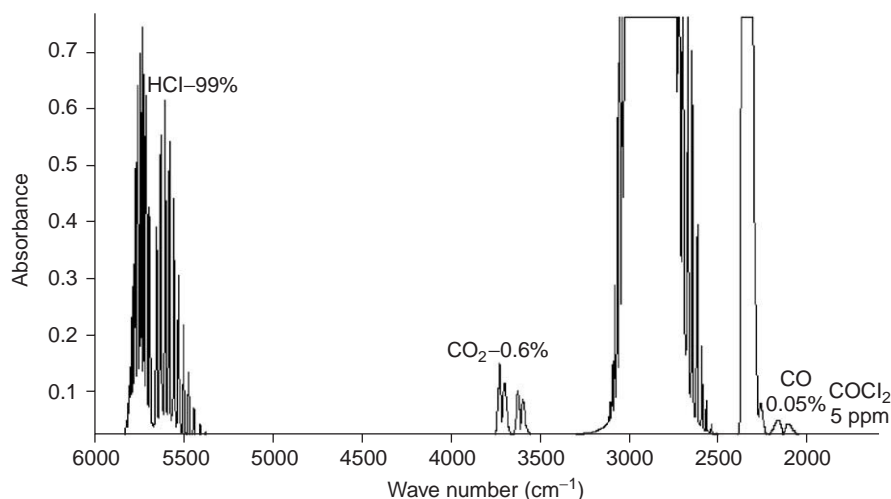
An area that will become increasingly important is the use of infrared spectroscopy for online and at-line analysis. Most of the novel applications of FTIR to industrial process analysis are proprietary. An example that illustrates the advantages of process FTIR is

to the analysis of impurities in anhydrous vapor-phase HCl. Typically, a large-scale isocyanate plant will produce  $>10^5$  tons of anhydrous HCl per annum and it is essential to the profitability of the plant that the HCl is utilized. One use for anhydrous HCl is production of 1,2-dichlorethane via the catalytic oxychlorination of ethylene. To be suitable as a raw material, the HCl must meet rigid quality specifications, as some impurities act as catalyst poisons. In addition to HCl, the components of interest include: phosgene, Freon 12, CO, CO<sub>2</sub>, and unknown organics such as process solvents. 'Ruggedized' FTIR spectrometers are commercially available and one was used for this application. The sample cell was a 20 cm gas cell with BaF<sub>2</sub> windows. The instrument was calibrated for five components and the precision achieved is shown in Table 1.

Figure 8 shows an online infrared spectrum of HCl. The wide dynamic range afforded by FTIR allows both intense bands such as the first overtone of the H–Cl stretch of the bulk sample and weakly absorbing bands such as the carbonyl stretch of phosgene to be used. It was recognized that variations in gas pressure will affect the analysis result because of gas law effects and spectral line shape changes. Pressure

**Table 1** Compounds monitored by FTIR in anhydrous gaseous HCl

Component	Range	Precision achieved
Hydrogen chloride	90–100%	0.56%
Carbon dioxide	0–5%	0.04%
Carbon monoxide	0–5000 ppm	3 ppm
Freon	0–500 ppm	2.2 ppm
Phosgene	0–500 ppm	1.7 ppm

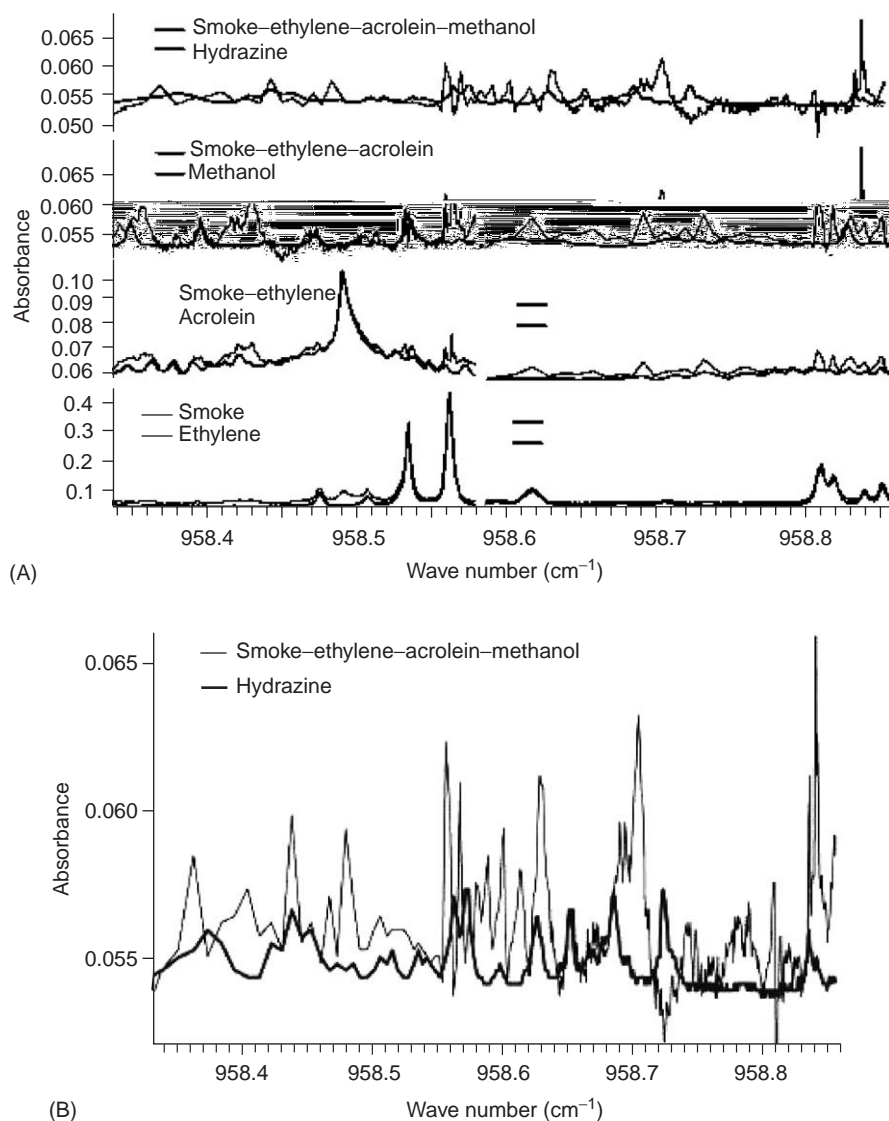


**Figure 8** An online spectrum of anhydrous HCl at 30 psig. (Reproduced with permission from Tate JD, Chauvel P, Guenard RD, and Harner R (2002) Process monitoring by mid- and near-infrared Fourier transform spectroscopy. In: Chalmers JM and Griffiths PR (eds.) *Handbook of Vibrational Spectroscopy*, vol. 4, pp. 2737–2769. Chichester: Wiley; © John Wiley & Sons Ltd.)

effects were compensated for in two ways: a carefully engineered sample system was used to regulate the sample pressure to  $3 \times 10^5$  Pa (35 psig) and the pressure in the sample cell was measured using a high-resolution pressure transducer. The calibration was developed by collecting spectra at 15 concentrations for each component at  $2.7 \times 10^5$  Pa (25 psig),  $3 \times 10^5$  Pa (30 psig), and  $3.3 \times 10^5$  Pa (30 psig). The calibration model itself was a set of partial least squares models, one for each component.

Infrared diode lasers offer high resolution,  $0.001 \text{ cm}^{-1}$  or better, ppb ( $10^{-6} \text{ g l}^{-1}$ ) detection limits and millisecond time resolution. These characteristics are ideally suited to low-pressure,

gas-phase measurements of light (five atoms or less) molecules. They find application in trace gas measurements of atmospheric species and also for measurements of isotopic species in human breath. An unusual application is to the determination of the components of inhaled cigarette smoke. **Figure 9A** shows the progressive removal (spectral stripping) of ethylene, acrolein, methanol, and comparison with a hydrazine spectrum. **Figure 9B** shows an expanded view of the final spectrum compared with that of hydrazine, which has been reported in cigarette smoke. It is apparent that there is at least one other component that has not been accounted for, but that this is not hydrazine.



**Figure 9** (A) Spectral stripping of components from the tunable diode laser infrared spectrum of cigarette smoke. From bottom to top: ethylene, acrolein, methanol, and comparison with hydrazine. The reference spectrum is the lower of the two traces in each case. (B) Expanded view of the residual spectrum compared to that of hydrazine. (Reprinted with permission from Plunkett S, Parrish ME, Shafer KH, Nelson D, Shorter J, and Zahniser M (2001) Time-resolved analysis of cigarette combustion gases using a dual infrared tunable diode infrared laser system. *Vibrational Spectroscopy* 27: 53–63; © Elsevier.)



## Trends in the Use of Infrared in Industry

There are three different types of infrared spectrometer available: dispersive, interferometric, and laser based. The present situation is that FTIR dominates the market. The reasons for this are that the combination of price, performance, and versatility offered by FTIR is unmatched by any competing technology. This trend is likely to continue for at least the next five years and probably the next decade. FTIR is a mature technology and most chemical laboratories have, at least, a simple bench-top unit as a piece of routine analytical equipment. There will be an increasing use of 'black-box' analyzers where the sample is inserted and an answer is produced. The fact that the box is actually an FTIR spectrometer will be largely irrelevant to the user; it will be a tool in the same way that a balance or a pH meter is. This is already happening; dedicated instruments for used oil analysis, semiconductor analysis, and on-bead analysis in combinatorial chemistry are commercially available.

The use of chemometrics will continue to increase and is almost mandatory for online applications. The spectral stripping of cigarette smoke could be done using a chemometrics approach rather than by sequential subtraction. The advantage is that all the components would be determined simultaneously without the laborious and somewhat subjective subtractions. The disadvantage is that the work to setup the method is somewhat greater; however, for a large number of samples this would be recovered in reduced time per analysis and the use of less skilled personnel. The trade-off between time to setup the method and time to carry out the analysis is the usual case.

Throughout the chemical industry there is an accelerating trend to obtaining data about a process online so that the process can be optimized in real time. Largely because fiber optic technology that can be used over hundreds of meters already exists, near-infrared and Raman spectroscopies are finding increasing application for this purpose; however, the mid-infrared is potentially much more informative. The major limitation is the length of the fibers that can be used before the losses become unacceptable. ZrF<sub>2</sub> fibers may be used up to 10 m or so but only allow the O-H and C-H stretching regions to be observed. Chalcogenide (AsSbSe) glasses allow the mid-infrared to 1200 cm<sup>-1</sup> to be measured but can only be used with a total length of a few meters. An alternative is that used for the HCl analysis: bring the sample to the spectrometer. This has engineering implications but results in much better spectra.

The sea change in infrared spectroscopy has been the explosion of applications in biology. One of the

major goals is diagnosis, particularly of cancer. The small size of cancer cells makes this difficult so two complementary approaches are being employed. Infrared synchrotron radiation is much brighter (photons cm<sup>-2</sup>) and better collimated than the thermal sources used in spectrometers. This allows spectra with high signal-to-noise ratio to be recorded down to the diffraction limit, ~0.005 mm. The second is the use of infrared microscopes with a focal plane array of detectors to collect images. These are then subjected to a variety of advanced data processing algorithms. The medical and commercial potential of automated diagnosis is so large that this is likely to be the application that dominates infrared spectroscopy for the next decade.

*See also:* **Infrared Spectroscopy:** Overview; Sample Presentation.

## Further Reading

- Bowers DI and Maddams WF (1989) *The Vibrational Spectroscopy of Polymers*. Cambridge: Cambridge University Press.
- Bugay DE (2001) Characterization of the solid-state: Spectroscopic techniques. *Advanced Drug Delivery Reviews* 48: 43–65.
- Diem M, Boydston-White S, and Chiriboga L (1999) Infrared spectroscopy of cells and tissues: shining light onto a novel subject. *Applied Spectroscopy* 53: 148A–161A.
- Koenig JL, Wang S-Q, and Bhargava R (2001) FTIR images. A new technique produces images worth a thousand spectra. *Analytical Chemistry* 73: 361A–369A.
- Naumann D (2001) FT-Infrared and FT-Raman spectroscopy in biomedical research. *Applied Spectroscopy Reviews* 36: 239–298.
- Sheppard N and de la Cruz C (1996) Vibrational spectra of hydrocarbons adsorbed on metals. Part I. Introductory principles, ethylene and the higher acyclic alkenes. In: Eley DD, Haag W, and Gates B (eds.) *Advances in Catalysis*, vol. 41, pp. 1–112. San Diego: Academic Press.
- Sheppard N and de la Cruz C (1998) Vibrational spectra of hydrocarbons adsorbed on metals. Part II. Adsorbed acyclic alkynes and alkanes, cyclic hydrocarbons including aromatics and surface hydrocarbon groups derived from decomposition of alkyl halides etc. In: Eley DD, Haag W, and Gates B (eds.) *Advances in Catalysis*, vol. 42, pp. 181–313. San Diego: Academic Press.
- Suzuki EM (1993) Forensic applications of infrared spectroscopy. In: Saferstrin R (ed.) *Forensic Science Handbook*, vol. 3, pp. 195–252. Englewood Cliffs: Prentice-Hall.
- Tate JD, Chauvel P, Guenard RD, and Harner R (2002) Process monitoring by mid- and near-infrared Fourier transform spectroscopy. In: Chalmers JM and Griffiths PR (eds.) *Handbook of Vibrational Spectroscopy*, vol. 4, pp. 2737–2769. New York: Wiley.



# ION CHROMATOGRAPHY

See **ION EXCHANGE: Ion Chromatography Instrumentation; Ion Chromatography Applications; Chelation Ion Chromatography; Isotope Separation**

## ION EXCHANGE

Contents

**Overview**

**Principles**

**Ion Chromatography Instrumentation**

**Ion Chromatography Applications**

**Chelation Ion Chromatography**

**Isolation of Biopolymers**

**Isotope Separation**

### Overview

**P R Haddad**, University of Tasmania, Hobart, TAS, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by J. Inczedy, pp. 2267–2273, © 1995, Elsevier Ltd.

### Introduction

Ion exchangers are solid materials or liquid solutions which are able to take up (or absorb) positively or negatively charged ions from aqueous electrolyte solutions and at the same time release other ions of equivalent amount into the aqueous solution. According to the electric charge of the ions taking part in the ion exchange process, one can speak about cation and anion exchangers. Ion exchangers which are able to interact with both types of ions are called amphoteric.

The ideal ion exchange process proceeds stoichiometrically, i.e., the ratio of the ions exchanged between the two phases is strictly determined by their charges. A clear distinction can be made between ion exchange and adsorption or liquid–liquid extraction processes, in which molecules are transferred from the aqueous phase into the solid, or organic solvent phase, without releasing any other species into the aqueous solution.

Ion-exchange processes are important in biological systems of living organisms. They are important not

only from the point of view of selective secretion of some biological substances, but also in the transport mechanism of certain ions crossing cell membranes, and also in the signal expansions in nerve systems.

Ion exchange processes have importance also in agriculture, because the ion-exchange properties of the natural silicates present in the soils highly influence the composition of the interstitial liquid available for the feeding of the plants.

Synthetic ion exchangers have proved to be very useful tools both in industry and in the laboratory, because by means of ion exchangers the concentration of certain ions in solution can be changed very easily without unwanted disturbances of the total composition of the electrolyte, and also selective separations, extractions, and enrichment procedures can be carried out with them.

### History

Key events in the development of ion exchangers and also that of their analytical applications are tabulated in chronological order in Table 1.

### Types of Ion Exchangers

The current synthetic ion exchange materials can be classified according to their form into the following main groups:

- solid beads or particles,
- solid membranes,
- solid sheets (papers or layers), and

**Table 1** Development of ion exchangers and of their analytical applications

1850	Observation of the ion-exchange properties of soils
1858	Zeolites: natural ion-exchangers in the soils
1917	Determination of ammonia in urine using a synthetic zeolite
1935	Condensation type ion-exchange resins
1942	Polymerization type cation-exchange resins
1947	Polymerization type anion-exchange resins
1947	Chromatographic separation of rare earths
1948	Specific ion-exchange resin for potassium
1949	Oxidation–reduction functional groups in resins
1948	Liquid anion exchangers
1950	Liquid cation exchangers
1950	Ion-exchange membranes
1951	Chromatographic separation of amino acids
1956	Cellulose-based ion exchangers
1957	Dextran-based ion exchangers
1967	Surface-coated silica ion exchangers for chromatography
1975	Conductivity detector system and surface-coated resins for 'ion-chromatography'
1975	Crown groups in resins
1997	Monolithic ion exchangers

- organic solvent solutions of ion exchange compounds (liquid ion exchangers).

The conventional ion exchangers belong to the first group. (The fibrous or powder solid ion exchangers are of much less importance.) These ion exchangers can be classified according to their chemical compositions (matrices), and to their functional groups as follows:

1. Ion-exchange resins with condensation type (phenol-formaldehyde) or with polymerization type (styrene-divinylbenzene copolymer or methacrylate-divinyl-benzene copolymer, etc.) matrix. The strongly acidic cation exchangers have sulfonic acid groups, the weakly acidic cation exchangers carboxylic or phosphonic acid groups. The strongly basic anion exchangers usually have quaternary ammonium groups, while the weak exchangers have primary, secondary, or tertiary amine groups. Chelating resins have chelate-forming functional groups containing O, N, or S donor atoms. Amphoteric ion-exchange resins have both acidic (cation exchanger) and basic (anion exchange) groups.
2. Cellulose-based ion exchangers. Their chemical structure is a hydrophilic cellulose network having acidic (carboxymethyl, sulfoethyl, etc.) or basic (amino-ethyl, diethylaminoethyl, etc.) groups.
3. Dextran- or agarose-based ion exchangers. The matrix is cross-linked hydrophilic dextran or agarose and the attached functional groups are acidic or basic, similar to those mentioned for the cellulose ion exchangers, for cation and anion exchangers, respectively.

4. Inorganic ion exchangers. Inorganic cation exchangers include the zeolites (crystalline hydrated aluminosilicates) and acid salts of polyvalent metal ions (like zirconium phosphate, titanium tungstate, nickel hexacyanoferrate(II), etc.), salts of some hetero-polyacids (ammonium molybdophosphate, etc.). The hydrous oxides of tri- and tetravalent metal ions show cation and anion exchange properties depending on the nature of the metal ion and on the pH of the solution.

5. Surface-functionalized ion exchangers. Surface-functionalized ion exchangers are produced for high-performance ion exchange chromatography as column packing materials. Usually they are spherical beads with an inert core (styrene divinylbenzene copolymer or silica) and a thin surface layer of ion exchanger properties, having acidic or basic functional groups serving as cation or anion exchangers, respectively. The functional groups can be chemically bonded directly onto the matrix of the ion-exchange material or can be attached physically to the matrix as very small, chemically functionalized latex particles. The composite ion exchangers have much lower exchange capacity than the ion-exchange resins, but their hydrodynamic and mass transfer properties are much more favorable for rapid, high-performance chromatographic separations.

6. Ion exchange papers and layers. For ion-exchange paper chromatography, papers are impregnated with ion-exchange reagents, or finely distributed ion-exchange resin powder is embedded into the paper, or the paper itself is prepared from cellulose ion exchanger. For ion-exchange thin-layer chromatography, the sheets are usually prepared from ion-exchange resin powder of very low particle size and fixed by an inert adhesive material.

7. Ion-exchange membranes. Ion-exchange membranes are films or thin (0.1–0.6 mm) ion-exchange sheets with appropriate mechanical properties, and with proper porosity and low electrical resistance.

8. Liquid ion exchangers. Liquid ion exchangers are solutions of high molecular mass organic acids (cation exchangers) or bases (anion exchangers) which are sparingly soluble in water but readily soluble in less polar organic solvents and by which ions can be extracted from aqueous solutions.

9. Monolithic ion exchangers. A monolithic ion exchanger is formed from a continuous bed of silica or polymer (rather than discrete particles) in which interconnecting and cross-linked strands of material form a uniform, porous structure. This structure is then functionalized to produce anion or cation exchangers. Monolithic materials have low flow resistance and can therefore be operated at high flow rates.

## The Properties of Solid Ion Exchangers

The physical and chemical properties of solid ion exchangers are crucial to their characterization. Of the physical properties, color and density are the most important. For the granular products, particle size, particle size distribution, the shape of the particles, the bulk density, porosity, compressibility, mechanical stability, and the volume change caused by acidification or alkalization are important. For high-performance chromatographic separations, uniformity, regular spherical form, low compressibility, and low volume change of the particles on solvation are important. For the characterization of the ion-exchange membranes, the thickness and resistance against pressure change are also guiding properties.

The most important data for chemical characteristics are as follows:

- the ion-exchange capacity (referred to unit volume or mass);
- the nature of the functional groups;
- the water uptake;
- the structural porosity and rigidity of the matrix; and
- the chemical stability (against acidic/alkaline solutions as well as against oxidants and reductants; against radiation induced decomposition, etc.).

The ‘capacity’ of the ion exchanger expresses the amount of exchangeable singly charged ions either per unit mass or per unit volume. The determination of the capacity is usually carried out by acidimetric titration. For the chelating resins, the capacity may be determined by the uptake of copper(II) ions referred to unit mass or volume.

The affinities of the exchangeable ions to the fixed sites depend highly on the ‘nature of the functional groups’. The ratio of the adsorption strength of similar ions gives the selectivity of the ion exchangers.

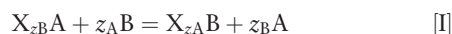
As the ion exchange occurring in aqueous solutions is always accompanied with the transport of water molecules, the ‘water uptake’ is an important feature of ion exchangers. The inner porosity, the number of available exchange sites for ions of higher size, and the exchange rate of ions at the gel type ion-exchange resins depend strongly on the water content and swelling properties of the resin. With the so-called macroporous resins, the extent of swelling is lower due to their more rigid structure.

Inorganic ion exchangers exhibit a rigid network in which the number of available sites depends strongly on the diameter of the inner channels and on the size of the hydrated or less hydrated counterions.

For the long-term use of ion exchangers, their ‘chemical stability’ is an important characteristic which means that during the repeated depletion–regeneration cycles they should preserve their exchange capacities, without degradation of the mechanical and physical properties and without releasing any soluble organic compounds into the contacting solutions. For nuclear purposes, stability against radiation is also an important characteristic.

## Ion-Exchange Equilibria

The exchange process of two cations (or anions) A and B taking place between two immiscible phases can be described by the following equation (the possible transfer of solvent molecules is neglected):



where  $z_A$  and  $z_B$  denote the charges on the two ions, respectively. The fraction of the ion exchanger equivalent to one singly charged counter ion is denoted by  $X$ . At equilibrium

$$K_{A/B}^T = \frac{(\bar{a}_B)^{z_A} (a_A)^{z_B}}{(\bar{a}_A)^{z_B} (a_B)^{z_A}} \quad [1]$$

where  $\bar{a}$  and  $a$  denote the activities in the ion exchanger and in the contacting liquid phase, respectively.  $K_{B/A}^T$  is the thermodynamic equilibrium constant. In the calculation and also in the prediction of thermodynamic equilibrium constant values, difficulties arise from the fact that the activities inside the ion exchanger phase are not measurable, and the values of the mass action ratio expressed in terms of concentrations ( $K_{B/A}^c$ , see eqn [2]) usually depend on the mole ratio of the ions inside the ion exchanger phase:

$$K_{B/A}^c = \frac{[B]_R^{z_A} [A]_{aq}^{z_B}}{[A]_R^{z_B} [B]_{aq}^{z_A}} \quad [2]$$

The mass action ratio can be expressed also in terms of equivalent fractions, denoted by  $\bar{x}$  and  $x$  in the ion exchanger and in the contacting liquid phase, respectively:

$$K_{B/A}^x = \frac{(\bar{x}_B)^{z_A} (x_A)^{z_B}}{(\bar{x}_A)^{z_B} (x_B)^{z_A}} \quad [3]$$

where

$$\bar{x}_B = \frac{z_B [B]_R}{z_A [A]_R + z_B [B]_R} \quad [4]$$

$$\bar{x}_A = \frac{z_A [A]_R}{z_A [A]_R + z_B [B]_R} \quad [5]$$

$$x_B = \frac{z_B[B]_{aq}}{z_A[A]_{aq} + z_B[B]_{aq}} \quad [6]$$

$$x_A = \frac{z_A[A]_{aq}}{z_A[A]_{aq} + z_B[B]_{aq}} \quad [7]$$

The 'selectivity ratio' (separation factor) of the two ions, independently of their charge, is as follows:

$$\alpha_{B/A} = \frac{[B]_R [A]_{aq}}{[B]_{aq} [A]_R} = \frac{D_B}{D_A} \quad [8]$$

where  $D_B$  and  $D_A$  are the distribution ratios of ion B and ion A (see later). The larger the value of  $\alpha_{B/A}$ , the larger the selectivity of the ion exchanger for the preference of B ions over A ions. The relationship between  $K^T$  and  $K^C$  is as follows:

$$K_{B/A}^C \frac{(\bar{f}_B)^{z_A}}{(\bar{f}_A)^{z_B}} \frac{(f_A)^{z_B}}{(f_B)^{z_A}} = K_{B/A}^T \quad [9]$$

$\bar{f}_A, \bar{f}_B$  and  $f_A, f_B$  are the activity coefficients for the A, B ions in the ion exchanger and in the liquid phase, respectively. The relation between  $K^x$  and  $K^C$  is

$$K_{B/A}^x = K_{B/A}^C \frac{Q^{z_B}}{Q^{z_A}} \frac{C^{z_A}}{C^{z_B}} \quad [10]$$

$Q (= z_A[A]_R + z_B[B]_R)$  is the total concentration in the ion exchanger,  $C (= z_A[A]_{aq} + z_B[B]_{aq})$  is the total concentration in the contacting liquid phase expressed in terms of equivalents per volume unit and assuming a binary mixture. The subscripts R and aq refer to the exchanger and aqueous phase, respectively:

$$K_{B/A}^x = \alpha_{B/A} \left( \frac{x_A}{\bar{x}_A} \right)^{z_B-1} \left( \frac{\bar{x}_B}{x_B} \right)^{z_A-1} \quad [11]$$

If  $z_A = z_B$ , the relation is very simple:

$$K_{B/A}^C = K_{B/A}^x = \alpha_{B/A} \quad [12]$$

and the thermodynamic equilibrium constant can be calculated formally from the following relation if  $K_{B/A}^C$  values are known at various  $\bar{x}_B$ :

$$\ln K_{B/A}^T = \int_0^1 \left[ \ln K_{B/A}^C + \ln \frac{(f_A)^{z_B}}{(\bar{f}_B)^{z_B}} \right] d\bar{x}_B \quad [13]$$

In analytical chemistry, mainly the  $K_{B/A}^C$  and  $\alpha_{B/A}$  values are used for informatory calculations. The hydrogen ion is normally used as a reference for cations and the chloride ion for anions. Both  $K^C$  and  $\alpha$  values depend not only on the nature of the exchanging ions, but also on the nature of the ion

exchanger. As reference resin, Dowex  $50 \times 8$  is usually proposed for cation and Dowex  $1 \times 8$  for anion exchangers. Many attempts have been made to assess the selectivity of ion exchangers, and to explain factors influencing it. The main factors include the charge of the ion being exchanged (the 'electroselectivity effect'), the size of the ion being exchanged, and the chemical nature and steric arrangement of the functional group on the resin.

**Distribution ratio.** The distribution of an ionic species between the ion exchanger and the contacting phase is an important term in ion-exchange chromatography, because the migration rate of the ion on the chromatographic column and hence the position of the peak of the component in the chromatogram is controlled by the distribution ratio.

For a metal ion,  $M^{2+}$ ,

$$D_M = \frac{[M^{2+}]_R}{[M^{2+}]_{aq}} \quad [14]$$

The distribution ratio depends on the concentration of the competing ion in the electrolyte,  $E^+ Y^-$  (e.g., NaCl or HCl):

$$K_{M/E}^C = \frac{[M^{2+}]_R [E^+]_{aq}^2}{[M^{2+}]_{aq} [E^+]_R^2} \quad [15]$$

$$D_M = \frac{[M^{2+}]_R}{[M^{2+}]_{aq}} = K_{M/E}^C \frac{[E^+]_R^2}{[E^+]_{aq}^2} \quad [16]$$

If the  $M^{2+}$  metal ion is present in low concentration and the  $E^+$  'eluent' ion in medium concentration, the resin is practically in E-ion form and  $[E^+]_R$  is very close to  $Q$  (capacity of the resin), and  $K_{M/E}^C$  can be regarded as constant:

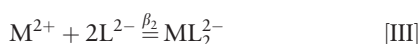
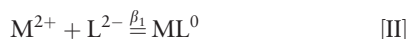
$$D_M = K_{M/E}^C Q^2 [E^+]_{aq}^2 \quad [17]$$

or in logarithmic form:

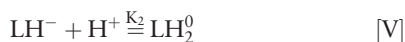
$$\log D_M = \log K_{M/E}^C + 2 \log Q - 2 \log [E^+] \quad [18]$$

There is a linear relation between the logarithm of the distribution ratio of the metal ion and the logarithm of the concentration of the competing eluent ion. The slope of the line corresponds to the ratio of the charges on the two ions (2/1). The distribution ratio of multicharged ions increases steeply with decreasing concentration of the singly charged eluent ion. It is interesting to note that the equation is valid not only in the case if  $Q$  is constant (i.e., the same ion exchanger is used) but also in the case if the concentration of the eluent is constant but the capacity of the ion exchanger is changing.

The  $D_M$  value of the metal ion can be influenced not only by the change of the eluent ion concentration but also by changing the actual value of the ion-exchange concentration constant. By using complex-forming agents, which form complexes of zero or of opposite charge with the metal ion, a conditional ion exchanger constant can be computed, which is of lower value than that of the original one. If a complex-forming agent,  $H_2L$ , is added to the solution, which forms complexes according to the equations



and the protonation of the ligand takes place according to the reactions



the conditional constant for the  $M^{2+}/E^+$  ion-exchange will be

$$K'_{M/E} = \frac{K_{M/E}^c}{\alpha_{M(L)}} \quad [19]$$

where

$$\alpha_{M(L)} = 1 + \frac{\beta_1 C_L}{\alpha_{L(H)}} + \frac{\beta_2 C_L^2}{\alpha_{L(H)}^2} \quad [20]$$

$$\alpha_{L(H)} = 1 + [H^+]K_1 + [H^+]^2K_1K_2 \quad [21]$$

If the pH, and the analytical concentration of the ligand,  $C_L$ , and  $[E^+]$  are known, the conditional constant value using eqn [19] and the conditional distribution ratio can be calculated too.

Similar calculations can be carried out for planning separations of metal ions, or of organic acids etc. using either solid or liquid ion exchangers, if the ion-exchange concentration constants, the corresponding protonation and complex formation constants are known.

Beside the ion-exchange process, 'absorption of electrolytes' also takes place if the ion exchanger is in contact with electrolyte solutions. The extent of the invasion of the ions depends on the concentration of the electrolyte in the solution and on the density of the ionic sites in the ion exchanger phase. The distribution of the electrolyte between the two phases is governed by Donnan equilibrium, and is significant only in solutions of higher than  $0.1 \text{ mol l}^{-1}$  concentration. At lower concentration the electrolyte is 'excluded' from the ion exchanger phase. For a single-single charge electrolyte of composition EY,

the concentration of  $Y^-$  in the ion exchanger phase can be estimated from the following equation:

$$[EY]_{aq}^2 = [Y^-]_R Q + [Y^-]_R^2 \quad [22]$$

The second term on the right-hand side can be neglected, and eqn [22] is simplified:

$$[Y^-]_R = \frac{[EY]_{aq}^2}{Q} \quad [23]$$

The ion-exclusion phenomenon is used in 'ion-exclusion chromatography' for separation of ions of different characters or charges.

Absorption of nonelectrolytes can take place from solution. The extent of the absorption depends on the matrix properties of the ion exchanger and on the electrolyte concentration of the liquid phase. The phenomenon is used for separation of nonelectrolytes and weak electrolytes by 'salting out chromatography'.

## Ion-Exchange Kinetics

As a general rule, the rate of ion-exchange processes is controlled by the diffusion of the exchanging ions. In a few cases the exchange rate is influenced by slow chemical reactions.

The rate controlling steps for solid ion exchangers are in general

- transport of the counterions across a thin liquid film (Nernst film) covering the surface of the ion exchanger and
- diffusion of the ions inside the solid ion exchanger phase.

In the mass transfer kinetics of the liquid-liquid ion exchanger processes, the contacting surface area and also the diffusion rate of the ions in the liquid phases are the controlling factors. The process can be highly accelerated by agitation.

For solid ion exchangers, considering spherical particles and for the simple ion exchange of isotopic species, if the process is controlled only by the diffusion of the ions in the thin film solution phase the fractional attainment of equilibrium as a function of time  $t$  can be expressed as follows:

$$F = 1 - e^{-(3dc/r\delta Q)t} \quad [24]$$

where  $d$  and  $c$  are the diffusion coefficient and concentration of the ion in the solution phase,  $r$  is the radius of the ion exchanger solid particle,  $\delta$  is the thickness of the liquid film, which can be influenced by agitation, and  $Q$  is the capacity of the ion exchanger. The half time (until  $F=0.5$ ) of the process is given by

$$t_{1/2} = 0.023r\delta d \frac{Q}{c} \quad [25]$$



For a reaction that is controlled solely by diffusion within the solid particle, the equation is

$$F = 1 - \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} e^{-(\bar{d}\pi^2 n^2 / r^2)t} \quad [26]$$

where  $\bar{d}$  is the diffusion coefficient of the ion in the ion exchanger phase. The half time

$$t_{1/2} = 0.03r^2 / \bar{d} \quad [27]$$

depends only on the particle size and diffusion coefficient but not on the concentration.

Since the diffusion coefficients of simple ions in aqueous solutions at room temperature are  $\sim 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ , while the diffusion coefficients inside the usual ion exchanger resins are in the range of  $10^{-9} - 10^{-6} \text{ cm}^2 \text{ s}^{-1}$ , film diffusion is rate controlling only at very small particle size ( $10 \mu\text{m}$ ) and at low concentration. High concentration, large particle size, compact ion exchanger phase (strongly cross-linked polymer), and less mobile, large ions favor inner diffusion control.

## Application Techniques

Ion-exchange processes using solid ion exchangers are usually carried out either by a batch or column technique.

Since in the batch procedure the ion exchange comes to equilibrium, quantitative exchange of ions from a solution can be obtained only in the case if the exchange reaction is highly favored ( $K_{B/A}^x$  is high) or multicharged ions are absorbed from singly charged ion-containing solution of low concentration, and the ion exchanger is in a fairly large excess.

Using fixed bed ion-exchange column techniques, the quantitative exchange of ions can also be attained in those cases where the selectivity conditions are not very favorable. In these cases the length of the column and flow rate must be chosen properly.

## Applications

Applications of the *solid ion exchangers* in the analytical laboratory can be grouped into the following classes:

1. procedures based on the total ion-exchange principle,
2. chromatographic separations, and
3. use of ion exchangers as carriers.

For the procedures belonging to class 1, mainly column techniques, but in a few cases batch techniques also, can be used.

In 'inorganic analysis' the most important applications are as follows:

- titrimetric determination of the salt content of solutions, by transforming them either into acid or base;
- preparation of standard solution of acids and bases (e.g., preparation of carbonate-free alkali hydroxide solution);
- dissolution of weakly soluble salts (like gypsum);
- removal of interfering ions: either cations, or anions, or both; and
- enrichment of ions of low concentration; collection of metal ion impurities from large volumes of drinks or from natural waters. The adsorbed metal ions can be eluted thereafter with a small volume of a suitable electrolyte solution for subsequent determination. For selective adsorption of certain metal ions, the presence of chelate-forming functional groups can also be useful.

In quantitative analysis based mainly on column chromatographic techniques, elution methods are preferred. The displacement chromatographic separations are useful for preparative purposes. The chromatographic procedures are useful for separation of metal ions from each other using cation exchange column or in their anionic complex ion form using an anion exchanger column. For the separation of ions of nonmetallic elements ( $\text{Cl}^-$ ,  $\text{Br}^-$ ,  $\text{HPO}_4^{2-}$ , etc.), an anion exchanger column is used.

For the detection of the separated ions, spectrophotometric and other spectroscopic detection methods are used.

If the distribution ratio of the analyte is fairly high ( $D > 100$ ), but that of the accompanying ions are low, separation with selective sorption (solid phase extraction) can be achieved.

Ion-exchange resins are useful as carriers for storing reagents or indicators, and to facilitate their use in laboratory work.

In 'organic analysis', the procedures based on total ion exchange are used for transformation of salts to acids or to bases; for removal of electrolytes from the solution of nonionic compounds; for isolation and separation of ionic or ionizable compounds from interfering substances; and group separation of components of a mixture containing acidic, basic, and less polar components.

Using chromatographic techniques, mixtures of acids, amines, aldehydes, ketones, amino acids, sugars etc. can be analyzed. If salting out and ion-exclusion chromatographic techniques are included, quite a wide range of organic compounds can be separated and determined. By the use of dextran-based packing

materials, ionizable bioactive compounds of higher molar mass can be analyzed chromatographically.

In organic analysis, the ion exchangers can be used as carriers for preparation of solid reagents (e.g., diazo reagents or enzymes).

### Liquid Ion Exchangers

Liquid ion exchangers are useful for extraction of metal ions, free acids, or bases from aqueous solutions. The collected ions can be back-extracted using an electrolyte of proper composition for the determination after the separation of the phases.

### Ion-Exchange Membranes

Ion-exchange membranes can be used as electrodes for electrometric titration of the salt content of water, or in electrodialysis cells for separation of ions of different charge or of different mobility.

The ion-exchange sheets or papers containing ion-exchange groups (or being impregnated) can be used for separation of inorganic and organic ions or ionizable compounds using a buffer solution as eluent and using established thin layer or paper chromatographic techniques.

Columns filled with resins containing functional groups with oxidizing–reducing properties can be

used as columns, for the quantitative oxidation or reduction of certain components of an aqueous solution (e.g., reduction of iron(III) traces to iron(II) using Variamine Blue reductor).

**See also:** Ion Exchange: Principles; Ion Chromatography Instrumentation; Ion Chromatography Applications. **Membrane Techniques:** Dialysis and Reverse Osmosis.

### Further Reading

- Dorfner K (1991) *Ion Exchangers*. Berlin: Walter de Gruyter.
- Greig JA (1996) *Ion Exchange Developments and Applications*. Cambridge: Royal Society of Chemistry.
- Harland CE (1994) *Ion Exchange: Theory and Practice*. Cambridge: Royal Society of Chemistry.
- Helfferich F (1962) *Ion Exchange*. New York: McGraw Hill.
- Inczédy J (1966) *Analytical Applications of Ion Exchangers*. Oxford: Pergamon Press.
- Inczédy J (1976) *Analytical Applications of Complex Equilibria*, chap. 3.9. Chichester: Ellis Horwood.
- Samuelson O (1952) *Ion Exchangers in Analytical Chemistry*. Stockholm: Almquist and Wiksell.
- Walton HF and Rocklin R (1990) *Ion Exchange in Analytical Chemistry*. New York: CRC Press.

## Principles

J Ståhlberg, AstraZeneca, AB, Södertälje, Sweden

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The first evidence for ion exchange was found in the middle of the 19th century in the field of agricultural chemistry. In the beginning most of the investigations were carried out on clays and minerals and this led later to the introduction of zeolites as an ion exchange material. Folin and Bell introduced the first analytical application of ion exchange in 1909 in a method for the determination of ammonia in urine. In the earlier studies of the ion exchange equilibrium, the obtained results were considered from a pure empirical point of view. In 1928, the Donnan model was introduced as a theoretical model by the Swedish agricultural chemist S. Mattsson. However, it was not until the end of the 1940s that this model was generally accepted and became a starting point for the more elaborate models that were developed in the 1950s.

It is important that the nomenclature is clear and in this article the following is used: the volume of the column that is filled up with the ion exchange material and the liquid phase is divided into the solid phase, the resin phase, and the electrolyte phase. The solid phase is the part of the column to which the liquid cannot penetrate. The resin phase is the part of the column to which the liquid can penetrate but is stagnant in a chromatographic process. For a porous ion exchange material this phase mainly consists of the pore volume. The external electrolyte phase is the part of the column that is filled by the electrolyte and at the same time can be associated with a flow velocity in a chromatographic process.

The charges that are bound to the resin phase are called the fixed (resin) charges and they are the reference point for the nomenclature of the other ions in the ion exchange system. The ions in the electrolyte solution, which have a sign of charge that is opposite to the fixed resin charges, are called the counterions. Analogously, the charges that have the same sign are called the co-ions.

materials, ionizable bioactive compounds of higher molar mass can be analyzed chromatographically.

In organic analysis, the ion exchangers can be used as carriers for preparation of solid reagents (e.g., diazo reagents or enzymes).

### Liquid Ion Exchangers

Liquid ion exchangers are useful for extraction of metal ions, free acids, or bases from aqueous solutions. The collected ions can be back-extracted using an electrolyte of proper composition for the determination after the separation of the phases.

### Ion-Exchange Membranes

Ion-exchange membranes can be used as electrodes for electrometric titration of the salt content of water, or in electrodialysis cells for separation of ions of different charge or of different mobility.

The ion-exchange sheets or papers containing ion-exchange groups (or being impregnated) can be used for separation of inorganic and organic ions or ionizable compounds using a buffer solution as eluent and using established thin layer or paper chromatographic techniques.

Columns filled with resins containing functional groups with oxidizing–reducing properties can be

used as columns, for the quantitative oxidation or reduction of certain components of an aqueous solution (e.g., reduction of iron(III) traces to iron(II) using Variamine Blue reductor).

**See also:** Ion Exchange: Principles; Ion Chromatography Instrumentation; Ion Chromatography Applications. **Membrane Techniques:** Dialysis and Reverse Osmosis.

### Further Reading

- Dorfner K (1991) *Ion Exchangers*. Berlin: Walter de Gruyter.
- Greig JA (1996) *Ion Exchange Developments and Applications*. Cambridge: Royal Society of Chemistry.
- Harland CE (1994) *Ion Exchange: Theory and Practice*. Cambridge: Royal Society of Chemistry.
- Helfferich F (1962) *Ion Exchange*. New York: McGraw Hill.
- Inczédy J (1966) *Analytical Applications of Ion Exchangers*. Oxford: Pergamon Press.
- Inczédy J (1976) *Analytical Applications of Complex Equilibria*, chap. 3.9. Chichester: Ellis Horwood.
- Samuelson O (1952) *Ion Exchangers in Analytical Chemistry*. Stockholm: Almquist and Wiksell.
- Walton HF and Rocklin R (1990) *Ion Exchange in Analytical Chemistry*. New York: CRC Press.

## Principles

J Ståhlberg, AstraZeneca, AB, Södertälje, Sweden

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The first evidence for ion exchange was found in the middle of the 19th century in the field of agricultural chemistry. In the beginning most of the investigations were carried out on clays and minerals and this led later to the introduction of zeolites as an ion exchange material. Folin and Bell introduced the first analytical application of ion exchange in 1909 in a method for the determination of ammonia in urine. In the earlier studies of the ion exchange equilibrium, the obtained results were considered from a pure empirical point of view. In 1928, the Donnan model was introduced as a theoretical model by the Swedish agricultural chemist S. Mattsson. However, it was not until the end of the 1940s that this model was generally accepted and became a starting point for the more elaborate models that were developed in the 1950s.

It is important that the nomenclature is clear and in this article the following is used: the volume of the column that is filled up with the ion exchange material and the liquid phase is divided into the solid phase, the resin phase, and the electrolyte phase. The solid phase is the part of the column to which the liquid cannot penetrate. The resin phase is the part of the column to which the liquid can penetrate but is stagnant in a chromatographic process. For a porous ion exchange material this phase mainly consists of the pore volume. The external electrolyte phase is the part of the column that is filled by the electrolyte and at the same time can be associated with a flow velocity in a chromatographic process.

The charges that are bound to the resin phase are called the fixed (resin) charges and they are the reference point for the nomenclature of the other ions in the ion exchange system. The ions in the electrolyte solution, which have a sign of charge that is opposite to the fixed resin charges, are called the counterions. Analogously, the charges that have the same sign are called the co-ions.

This article focuses on the theoretical principles for solid particular ion exchangers and their use in analytical chemistry but many of the principles are, however, of more general character. In all ion exchange systems the electrostatic or Coulombic interaction between the electrolyte ions in the resin phase and the fixed resin charges is of great importance. To maintain electroneutrality, the fixed resin charges have counterions associated to it and their distribution in the resin phase is mainly the result of a balance between three effects: (1) electrostatic attraction to the fixed resin charges; (2) the smearing out effect of the thermal motion (i.e., the entropy); and (3) the way the counterions that are close to the fixed resin charges shield each other. It is important to recognize that (1) due to the thermal motion the counterions are not connected to the fixed resin charges as a 1:1 complex but as a diffuse layer close to the fixed charges and (2) there are co-ions present in the diffuse layer. The fixed charges together with the diffuse layer of ions form the so-called diffuse double layer. Since the fixed charges and the ions in the diffuse layer are separated in space, a difference in the electrostatic potential between a point in the diffuse double layer and a point in the bulk electrolyte solution is created. An important aspect of the electrostatic interaction is that it is long ranged, which contrasts to, e.g., the van der Waals interaction, implying that the presence of an ion at a given point simultaneously affects the energy of many ions, not only those that are in its closest vicinity.

The electrostatic potential difference between a point in the resin phase and the external electrolyte phase mainly determines the general properties of an ion exchanger. There is a general theory that was developed by R.A. Marcus for polyelectrolytes that may be used to connect the potential difference to the activity coefficient of an ion, and is therefore discussed in more detail. This theory is a useful starting point with which the more specific theories, which have been put forward in ion exchange, can be compared. In traditional ion exchange theory, the Donnan potential model is generally used to derive useful relations between different quantities. A comparison between the Donnan model and the more general theory is made. In more recent years, several theories based on the electrical double layer concept have been presented to explain ion exchange data and these are also shortly discussed.

Differences in electrostatic potential between the external electrolyte phase and the resin phase cannot explain selectivity differences between counterions of the same net charge. Much work has been done in many areas of chemistry to investigate the cause of selectivity and no well-established theoretical model exists. Some general empirical results are presented here as well as a simple theoretical model.

## Theory

Consider a resin phase with fixed charged groups and which is in equilibrium with an external electrolyte solution containing an ion  $B_B^z$ . The thermodynamic condition for equilibrium for  $B_B^z$  between the external solution and the resin phase is determined by the equivalence of the chemical potential in the two phases. Although the chemical potential for a single charged species is not a measurable quantity, the following equation can be used to theoretically describe the equilibrium condition:

$$\begin{aligned}\mu_{B,E}^0 + RT \ln c_{B,E} \gamma_{B,E} \\ = \mu_{B,R}^0 + RT \ln c_{B,R} \gamma_{B,R}\end{aligned}\quad [1]$$

where  $\mu_{B,E}^0$  and  $\mu_{B,R}^0$  are the standard chemical potentials for  $B_B^z$  in the external solution and in the resin phase, respectively.  $c_{B,E}$  and  $c_{B,R}$  are the concentrations of  $B_B^z$  in the external solution and its mean concentration in the resin phase, respectively. The activity coefficients,  $\gamma_{B,E}$  and  $\gamma_{B,R}$  are introduced in order to take account of the, usually strong, deviations from classical laws that are exhibited in ion exchange systems.  $R$  is the Avogadro constant and  $T$  the temperature. Assuming that the standard chemical potential is equal in the two phases, i.e.,  $\mu_{B,E}^0 = \mu_{B,R}^0$ , eqn [1] can be rewritten as

$$\frac{\gamma_{B,R}}{\gamma_{B,E}} = \frac{c_{B,E}}{c_{B,R}}\quad [2]$$

Thus, the relative accumulation (depletion) of  $B_B^z$  in the resin phase, relative to the surrounding electrolyte solution, is determined by the corresponding ratio of activity coefficients. In eqn [2] the ratio of activity coefficients is introduced as a pure empirical quantity and to proceed from here a physical interpretation and its theoretical explanation is needed.

As explained above, the charges in the resin phase give rise to a difference in electrostatic potential between the external solution and the resin phase and also to electrostatic potential gradients within the resin phase. In the following discussion it is assumed that electrostatic effects are the only factors that contribute to changes in the activity coefficients.

From physics we know that the work needed to move a positive unit test charge between two points is proportional to the difference in electrostatic potential between these points. The expression for chemical equilibrium, eqn [1], is related to the energy for an ion in the external and resin phases, respectively. Therefore, a term for the electrostatic energy can be introduced in the expression for the chemical potential and the electrochemical potential is obtained. For a small ion  $B_B^z$ , the thermodynamic

condition for equilibrium between a point in the external solution and a point  $r$  in the resin phase is now determined by the equivalence of the electrochemical potential in the external solution and point  $r$ :

$$\mu_{B,E}^0 + RT \ln c_{B,E} = \mu_{B,R}^0(r) + RT \ln c_{B,R}(r) + z_B F \cdot \Delta\Psi(r) \quad [3]$$

where  $\Delta\Psi(r)$  is the difference in electrostatic potential between a point  $r$  in the resin phase and the external electrolyte phase and  $F$  is the Faraday constant. Assuming that the electrostatic potential is the only difference that  $B_B^z$  experiences between the two phases, the standard chemical potential is equal, i.e.,  $\mu_{B,E}^0 = \mu_{B,R}^0(r)$  and from eqn [3] the Boltzmann equation for the concentration of  $B_B^z$  in point  $r$  is obtained:

$$c_{B,R}(r) = c_{B,E} \cdot e^{\frac{-z_B F \cdot \Delta\Psi(r)}{RT}} \quad [4]$$

For a cation exchanger, which contains negatively charged surface groups  $\Delta\Psi(r)$  is negative and from eqn [4] it follows that when  $z_B$  is positive (negative) the concentration of  $B_B^z$  is higher (lower) in a point  $r$  in the resin phase than in the external electrolyte solution. In an anion exchanger the conditions are reversed, i.e., the electrostatic potential is positive giving higher (lower) concentrations of anions (cations) in the resin phase compared to the external solution.

The total amount of  $B_B^z$  in the resin phase is equal to its mean concentration,  $c_{B,R}$ , times the volume of the resin phase,  $V_R$ . According to eqn [4], this amount is equal to the integral of  $c_{B,R}(r)$  over the volume  $V_R$ , i.e.,

$$c_{B,R} \cdot V_R = \int_{V_R} c_{B,R}(r) dV_R = c_{B,E} \cdot \int_{V_R} e^{\frac{-z_B F \cdot \Delta\Psi(r)}{RT}} dV_R \quad [5]$$

By comparing eqns [2] and [5], a physical interpretation of the ratio of activity coefficients is obtained:

$$\frac{\gamma_{B,R}}{\gamma_{B,E}} = \frac{c_{B,E}}{c_{B,R}} = \frac{V_R}{\int_{V_R} e^{\frac{-z_B F \cdot \Delta\Psi(r)}{RT}} dV_R} \quad [6]$$

## Ion Exchange Equilibrium

### General Considerations

Consider a cation exchanger that is in equilibrium with an electrolyte solution containing two identically charged cations, A and B, and a common anion. The ion exchange equilibrium between cations A and

B is usually written as



where  $A_E$  and  $A_R$  are the ion A in the external solution and resin phase, respectively, and  $B_E$  and  $B_R$  those corresponding to ion B. Equilibrium (7) gives the following thermodynamic equilibrium constant:

$$K_{A,B}^T = \frac{c_{B,E} \gamma_{B,E} \cdot c_{A,R} \gamma_{A,R}}{c_{B,R} \gamma_{B,R} \cdot c_{A,E} \gamma_{A,E}} \quad [8]$$

The selectivity coefficient for ion exchange equilibrium is obtained when the mass action law without activity coefficients is used, i.e., for the equilibrium in eqn [7] the coefficient is

$$K_{A,B}^S = \frac{c_{B,E} \cdot c_{A,R}}{c_{B,R} \cdot c_{A,E}} \quad [9]$$

In ion exchange chromatography, the retention of an analyte ion A is related to its distribution ratio  $D_A$ , defined as

$$D_A = \frac{c_{A,R}}{c_{A,E}} \quad [10]$$

and by combining eqns [8] and [10], we find that

$$D_A = \frac{c_{B,R} \gamma_{B,R} \cdot \gamma_{A,E}}{c_{B,E} \gamma_{B,E} \cdot \gamma_{A,R}} \cdot K_{A,B}^T \quad [11]$$

Equation [11] shows that the distribution constant for A depends on the ratio of activity coefficients and the concentration of the counterion B in the resin and external phases, respectively. When  $c_{B,E}$  increases, the main effect is a decrease in the electrostatic interaction to the fixed charges in the resin phase, which in turn affects the individual activity coefficients. It is therefore interesting to evaluate the ratio of activity coefficients in eqn [11]. Since eqn [6] applies individually to all ions in the electrolyte solution that are in equilibrium with the resin phase, a relation between  $K_{A,B}^T$ ,  $K_{A,B}^S$ , and  $D_A$  can be found.

For the equilibrium in eqn [7], we have that  $z_A = z_B$  and we find that the integrals on the right-hand side of eqn [6] are equal for the two ions and we obtain

$$K_{A,B}^T = \frac{c_{B,E} \cdot c_{A,R}}{c_{B,R} \cdot c_{A,E}} = K_{A,B}^S = 1 \quad \text{or} \quad D_A = \frac{c_{B,E}}{c_{B,R}} \quad [12]$$

Equation [12] shows that when only electrostatic effects are operating, the relative concentration of equally charged ions in the resin phase is the same as in the external electrolyte solution. In this particular case, the selectivity coefficient and the thermodynamic equilibrium constant are equal.

A more complicated case arises when the equilibrium simultaneously involves a monocharged cation



( $z_B = 1$ ) and a dicharged cation ( $z_A = 2$ ). The ion exchange equilibrium is now written as



and the corresponding equilibrium constant is

$$K_{A,B}^T = \frac{c_{B,E}^2 \gamma_{B,E}^2 \cdot c_{A,R} \gamma_{A,R}}{c_{B,R}^2 \gamma_{B,R}^2 \cdot c_{A,E} \gamma_{A,E}} \quad [14]$$

By using eqn [6], eqn [14] can be written as

$$K_{A,B}^T = \frac{c_{B,E}^2 \left( \int_{V_R} e^{\frac{-F \cdot \Delta \Psi(r)}{RT}} dV_R \right)^2 \cdot c_{A,R} V_R}{c_{B,R}^2 V_R^2 \cdot c_{A,E} \int_{V_R} e^{\frac{-2F \cdot \Delta \Psi(r)}{RT}} dV_R} \quad [15]$$

The exact numerical value for  $K_{A,B}^T$ , thus, depends on the ratio of the integrals in eqn [15], a value that is not easily accessible. However, in mathematics we have Schwarz's inequality, which for this particular case gives the following relation:

$$\frac{\gamma_{B,E}^2 \cdot \gamma_{A,R}}{\gamma_{B,R}^2 \cdot \gamma_{A,E}} = \frac{\left( \int_{V_R} e^{\frac{-F \cdot \Delta \Psi(r)}{RT}} dV_R \right)^2 \cdot V_R}{V_R^2 \cdot \int_{V_R} e^{\frac{-2F \cdot \Delta \Psi(r)}{RT}} dV_R} \leq 1 \quad [16]$$

and that

$$\begin{aligned} K_{A,B}^T &= \frac{c_{B,E}^2 \gamma_{B,E}^2 \cdot c_{A,R} \gamma_{A,R}}{c_{B,R}^2 \gamma_{B,R}^2 \cdot c_{A,E} \gamma_{A,E}} \\ &\leq \frac{c_{B,E}^2 \cdot c_{A,R}}{c_{B,R}^2 \cdot c_{A,E}} = K_{A,B}^s \end{aligned} \quad [17]$$

Equation [17] shows that the selectivity coefficient in this case is always greater than the thermodynamic equilibrium constant. This implies that there will be an accumulation of dicharged ions in the resin phase compared to what is expected from a simple mass action approach.

A literal interpretation of the equilibrium in [13] would be that, in order to keep electroneutrality in the resin phase, the sorption of a dicharged ion must be accompanied by desorption of two monocharged ions. At a first glance, eqn [17], therefore, seems to violate the principle of electroneutrality in the resin phase. However, the concentration for all types of ions in the external electrolyte is in equilibrium with the resin phase. In this particular case, this implies that the sorption of a dicharged ion will to a certain extent be accompanied by a sorption of the co-ion to the resin phase so that electroneutrality in the resin phase is maintained.

The strength of the above result is that the ratio of activity coefficients, eqn [6], is based on a minimum

of assumptions. It is independent of the geometry of the resin phase and also applicable to all ions in an ion exchange system. However, a numerical evaluation of the integrals requires models for  $\Delta \Psi(r)$  at various compositions and for the geometry under consideration. In the following paragraphs two models will be discussed.

### The Donnan Model

In the Donnan model, it is assumed that the difference in electrostatic potential between the external electrolyte solution and the resin phase is constant and the same at all points in the resin phase, i.e.,  $\Delta \Psi(r) = \Psi_D$  for all  $r$ . When this assumption is applied to eqn [6], eqn [18] is obtained:

$$\frac{\gamma_{B,R}}{\gamma_{B,E}} = \frac{c_{B,E}}{c_{B,R}} = \frac{V_R}{\int_{V_R} e^{\frac{-z_B F \cdot \Delta \Psi_D}{RT}} dV_R} = \frac{1}{e^{\frac{-z_B F \cdot \Delta \Psi_D}{RT}}} \quad [18]$$

It is obvious that the Donnan model, when applied to an ion exchange equilibrium between equally charged ions, will give the same result as previously, eqn [12]. For the more complicated equilibrium in [13], the following is obtained:

$$\begin{aligned} K_{A,B}^T &= \frac{c_{B,E}^2 \left( \int_{V_R} e^{\frac{-F \cdot \Psi_D}{RT}} dV_R \right)^2 \cdot c_{A,R} V_R}{c_{B,R}^2 V_R^2 \cdot c_{A,E} \int_{V_R} e^{\frac{-2F \cdot \Psi_D}{RT}} dV_R} \\ &= \frac{c_{B,E}^2 \left( e^{\frac{-F \cdot \Psi_D}{RT}} \right)^2 \cdot c_{A,R}}{c_{B,R}^2 \cdot c_{A,E} \cdot e^{\frac{-2F \cdot \Psi_D}{RT}}} \\ &= \frac{c_{B,E}^2 \cdot c_{A,R}}{c_{B,R}^2 \cdot c_{A,E}} = K_{A,B}^s \end{aligned} \quad [19]$$

Thus, the assumptions behind the Donnan model give as a result an equilibrium distribution that is the same as from a straightforward use of the mass action law. This is the reason why the mass action law frequently is used to describe an ion exchange process. It is clear that the result from the Donnan model is limited to the case where the electrostatic potential within the resin phase is constant, or nearly so.

### Double Layer Models

In the general case the electrostatic potential is not constant but varies within the resin phase. To be able to calculate the electrostatic potential at a point as a function of the ionic strength, the geometry and the charge density of the system must be specified. The electrostatic potential at a point is usually obtained by solving the Poisson–Boltzmann equation, which is a combination of the Boltzmann equation, eqn [4], with the Poisson equation from electrostatics. The

latter equation relates the change in electric field strength, i.e., the derivative of the electrostatic potential at a point, to the charge density at that point:

$$\nabla^2 \Psi = -\frac{\rho}{\epsilon_0 \epsilon_r} \quad (\text{Poisson equation}) \quad [20]$$

where  $\rho$  is the charge density,  $\epsilon_0$  the permittivity of vacuum, and  $\epsilon_r$  the dielectric constant of the medium. The charge density at a point is the difference between the number of positive and negative charges (ions) at that point, i.e.,

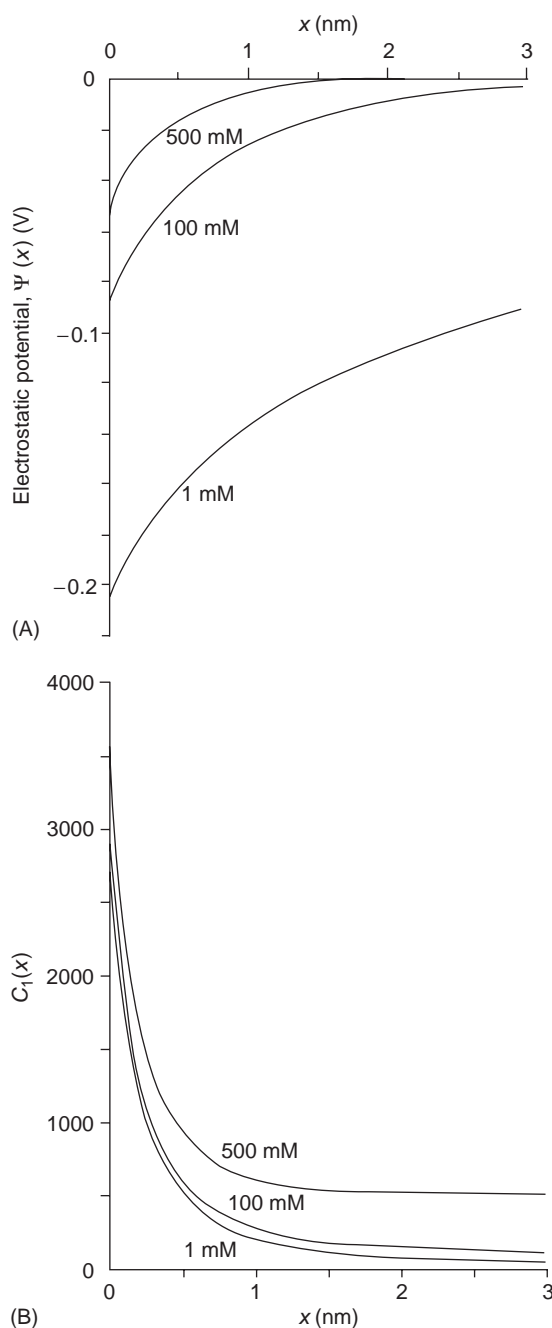
$$\rho = F \cdot \sum_i z_i c_i \quad [21]$$

where the summation is taken over all ions  $i$  present. By combining eqns [20] and [21], the Poisson–Boltzmann equation is obtained. This equation is the basis for our understanding of, e.g., the behavior of electrolyte solutions in contact with charged surfaces as well as many other physical phenomena. When solved for a particular geometry it gives the electrostatic potential, the concentration of the different ions, and the electric field at any point. Then, in principle, it is possible to calculate the ratio of activity coefficients in eqn [6]. However, for complicated geometries this is rather difficult to perform and has not been applied to resins used in ion exchange. For some simple cases exact solutions to the Poisson–Boltzmann equation exist, one such case is a planar charged surface in contact with an electrolyte solution containing a  $z$ - $z$  electrolyte, the Gouy–Chapman equation. This equation expresses the electrostatic potential at the surface and at any distance from the surface, as a function of the surface charge density and the ionic strength of the electrolyte solution. The exact equations can be found in most textbooks in surface and colloid chemistry.

Figure 1A shows the electrostatic potential calculated from the Gouy–Chapman equation for three different concentrations of monovalent salt in a water solution, as a function of the distance from a negatively charged planar surface. Figure 1B shows the concentration profile of the monovalent counterion corresponding to the three salt concentrations in Figure 1A. It is interesting to observe that the concentration at the surface is only  $\sim 20\%$  higher when going from 1 to 500  $\text{mol m}^{-3}$  in the electrolyte solution. The counterion concentration at the surface is therefore quite insensitive to changes in the salt concentration in the bulk electrolyte.

### Specific Interactions

In the Gouy–Chapman equation, the electrolyte ions are considered to be point charges that interact with



**Figure 1** (A) The electrostatic potential and (B) the counterion concentration as a function of distance from a negatively charged surface (surface charge density  $-0.1 \text{ C m}^{-2}$ ) according to the Gouy–Chapman theory, for three different concentrations of a 1:1 electrolyte (1, 100, and 500  $\text{mol m}^{-3}$ ). (Reprinted with permission from Ståhlberg J (1994) Retention models for ions in chromatography (review). *Analytical Chemistry* 66: 440; © American Chemical Society.)

smeared out surface charges by electrostatic interactions only. Although the pattern that emerges from this simple theory has been verified in numerous experiments, some modifications of it may be necessary

to make. When real ions interact with a charged surface, one must consider the possibility of specific interactions between the ion and the resin with its fixed charges. These interactions are usually referred to as the chemical part of the interaction and are short ranged, in contrast to the electrostatic interaction that is a long ranged interaction. The short ranged nature of the interaction implies that only ions that are close to the surface may interact chemically with the surface, while ions that are further away in the double layer do not. For an ion B that interacts chemically with the surface a mass action law is often used for describing the association

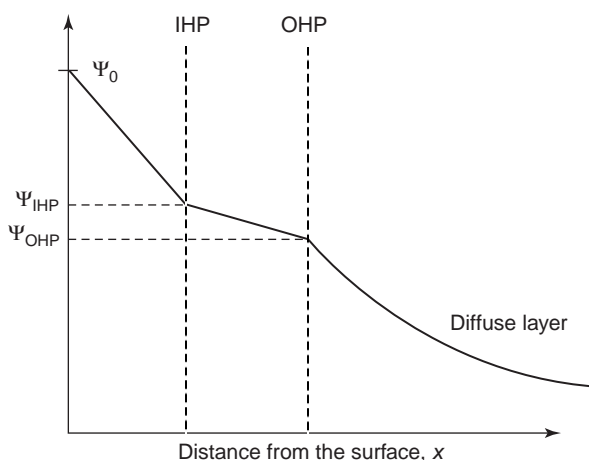
$$K_{\text{ch}} = \frac{[\text{B-S}]}{c_{\text{B}}(s) \cdot [\text{S}]} \quad [22]$$

where  $K_{\text{ch}}$  is the association constant for forming the B-S complex, (S) is the concentration of free surface groups to which B can bind.  $c_{\text{B}}(s)$  is the local concentration of B at the surface and can be calculated from eqn [4], i.e.,

$$c_{\text{B}}(s) = c_{\text{B,E}} \cdot e^{\frac{-z_{\text{B}}F \cdot \Psi(s)}{RT}} \quad [23]$$

where  $\Psi(s)$  is the electrostatic potential at the surface. In the physical chemistry of colloids and surfaces, much work has been done to find models that describe the adsorption of different ions to different kinds of charged surfaces. One problem is where to locate the plane of interaction between the counterions and the charged surface. Different models suggest different positions of the plane, depending on the geometrical properties of the surface. In the model above, it is assumed that the interaction takes place in the same plane as the surface charges.

An example of a system that has been extensively studied is the interaction between electrolyte ions and the molecularly smooth and impenetrable charged mercury surface. The theory developed for this system is the well-known Stern–Gouy–Chapman model that divides the interfacial region into three layers, see Figure 2. In this model, it is assumed that the size of the counterions hinders them from coming closer to the metal surface than a certain distance. The associated counterions are therefore located in a plane, the inner Helmholtz plane (IHP), separated from the solid surface and exposed to a potential  $\Psi_{\text{IHP}}$ . This creates a charge-free region confined between the surface and the IHP and a corresponding drop in the electrostatic potential from  $\Psi_0$  to  $\Psi_{\text{IHP}}$ . Ions situated at IHP are close enough to bind chemically and it is also assumed that there is room for a limited number of ions in this region, described as a



**Figure 2** A schematic representation of the electrostatic potential as a function of the distance from the surface according to the Stern–Gouy–Chapman model. (Reprinted with permission from Ståhlberg J (1999) Retention models for ions in chromatography (review). *Journal of Chromatography A* 855: 3; © Elsevier.)

Langmuir isotherm. In the model it is furthermore assumed that the Gouy–Chapman model holds for large distances from the surface and up to the outer Helmholtz plane (OHP) with the potential  $\Psi_{\text{OHP}}$ .

### Retention Models for Ion Exchange Chromatography of Small Ions

In analytical applications of ion exchange chromatography two aspects of retention are of main interest: (1) How does the retention factor vary with the concentration of eluent salt? (2) What factors determine the selectivity between different analyte ions? From a theoretical point of view, it is clear that with increasing concentration of electrolyte in the external solution, the magnitude of the electrostatic potential in the resin phase decreases. The reason is that the added electrolyte ions partly shield the charges bound to the resin phase. For analyte counterions this implies that their sorption to the resin phase decreases giving a lower retention factor.

In chromatography, the retention factor, or, equivalently, the capacity factor, is used to characterize the chromatographic equilibrium properties of an analyte. The retention factor of an analyte,  $k$ , is usually derived to be the product of the column phase ratio,  $V_{\text{R}}/V_{\text{E}}$ , and the distribution coefficient of the analyte between the stationary phase and eluent phase,  $c_{\text{A,R}}/c_{\text{A,E}}$ , eqn [24]:

$$k = \frac{V_{\text{R}}}{V_{\text{E}}} \cdot \frac{c_{\text{A,R}}}{c_{\text{A,E}}} = \frac{t_{\text{R}} - t_0}{t_0} \quad [24]$$

The retention factor is experimentally measured from the retention time (or volume) of the analyte,

$t_R$ , and the retention time (or volume) for a marker,  $t_0$ . A physical interpretation of the experimentally determined retention factor shows that it depends on a number of different assumptions. In the derivation of eqn [24] it is assumed that the marker used to determine the column dead volume,  $V_E$ , does not penetrate into the resin phase so that any part of  $V_R$  is not included into the value of  $V_E$ .

For equilibrium [11], the most used quantitative relation of  $k$  as a function of electrolyte salt concentration is based on the assumption of a Donnan potential in the resin phase, i.e., eqn [19] is assumed to be valid. It is also assumed that the concentration of co-ions to the surface charges is negligible in the resin phase so that  $c_{B,R}$  is equal to the concentration of surface charges in the resin phase,  $C_R$ . This value is therefore constant and independent of the concentration of  $c_{B,E}$  in the eluent phase. Combining eqn [19] with eqn [24] gives the following expression for the retention factor as a function of  $c_{B,E}$ :

$$k = \frac{V_R}{V_E} \cdot \frac{c_{A,R}}{c_{A,E}} = K_{A,B}^s \cdot \frac{c_{B,R}^2}{c_{B,E}^2} \cdot \frac{V_R}{V_E} \\ = K_{A,B}^s \cdot \frac{C_R^2}{c_{B,E}^2} \cdot \frac{V_R}{V_E} \quad [25]$$

In practice, ion exchange data are usually represented in a  $\log k$  versus  $\log c_{B,E}$  plot which for a divalent analyte eluted with a monovalent counterion have the form

$$\log k = \log K_{A,B}^s \cdot C_R^2 \cdot \frac{V_R}{V_E} - 2 \log c_{B,E} \quad [26]$$

The same type of arguments is applied to the more general case and the following expression is often used:

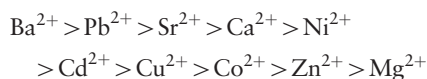
$$\log k = \log K - \frac{z_A}{z_B} \log c_{B,E} \quad [27]$$

where  $K$  is a system constant,  $z_A$  and  $z_B$  the charges of the analyte and the electrolyte counterion, respectively. Experimental data for mono- and discharged ions usually agree well with this equation, for higher charged ions the agreement is generally not very good.

### Selectivity

The discussion so far has been concerned with the role of the charge of the analyte and the electrolyte counterion on the equilibrium distribution of the analyte between the resin and external phase, respectively. It is clear that the distribution to the resin phase of a multivalent ion is higher than for a mono-charged ion, a phenomenon that is easy to understand from the previous discussion. There is therefore

selectivity for ions of higher charge, usually called electroselectivity. Selectivity differences are also found within the class of equally charged ions, which is the basis for their chromatographic separation. For a strong cation exchanger the following affinity sequence to the resin phase is usually found:



For a strong anion exchanger the sequence is often found to be



These sequences are not absolute but depend on the nature of the fixed resin charge and the type of resin. In general, the affinity increases with decreasing hydrated radius and increasing polarizability of the ion. The physical explanation to selectivity probably involves several mechanisms, many models have been proposed but no one model is generally accepted. The situation is analogous to the situation in colloid chemistry where the above series is similar to the Hofmeister series for coagulation. Both affinity series supports the idea that it is the ions that are located close to the surface and the fixed resin charges that is responsible for the selectivity. A simple approach, which has been used in colloid chemistry, is therefore to use the equilibrium constant in eqn [22] and ascribe a higher affinity to a higher value of this constant.

**See also:** Ion Exchange: Overview; Ion Chromatography Instrumentation; Ion Chromatography Applications; Isolation of Biopolymers; Isolation Separation.

### Further Reading

- Evans DF and Wennerström H (1994) *The Colloidal Domain*. Weinheim: VCH Publishers.
- Haddad PR and Jackson PE (1990) *Ion Chromatography*. Amsterdam: Elsevier.
- Harland CE (1994) *Ion Exchange: Theory and Practice*. Cambridge: The Royal Society of Chemistry.
- Helfferich F (1962) *Ion Exchange*. New York: McGraw-Hill.
- Marcus RA (1955) Calculation of thermodynamic properties of polyelectrolytes. *The Journal of Chemical Physics* 23: 1057–1068.
- Overbeek JThG (1956) The Donnan equilibrium. *Progress in Biophysics* 6: 58–84.
- Ståhlberg J (1994) *Analytical Chemistry* 66: 440.
- Ståhlberg J (1999) Retention models for ions in chromatography (review). *Journal of Chromatography* 855: 3–55.

## Ion Chromatography Instrumentation

**Ya I Yashin and A Ya Yashin**, Russia Scientific and Technical Center 'Chromatographia', Moscow, Russia  
**H F Walton**, University of Colorado, Boulder, CO, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Ion chromatography (IC) is an analytical variant of traditional ion-exchange chromatography providing the advantages of better separation, efficiency of columns, and higher sensitivity. Besides ion-exchange chromatography, variants such as ion-interaction (or 'ion pair') and ion-exclusion chromatography are also associated with IC.

It is generally accepted that the beginning of IC was in 1975, when the principle of reduction or suppression of the background conductivity of a strong eluent in a second sequential suppressed column was offered. Simultaneously, ion exchangers of low capacity were proposed for better separation and detection in such dual-column modes. In suppressed IC, a significant improvement in the sensitivity of conductivity detection was achieved. Later, in 1979, nonsuppressed IC or single-column IC using weak eluents of low-conductivity and low-capacity ion-exchange columns was offered. IC developed intensively in the past decades and is now one of the basic methods of analytical chromatography, along with high-performance liquid chromatography and gas chromatography. According to data of the *Journal of Chromatography* (Further Reading section), more than 5400 papers on different aspects of IC were published during 1975–2002. Many of them were devoted to improvement of IC instrumentation.

### General Questions of Ion Chromatography Instrumentation

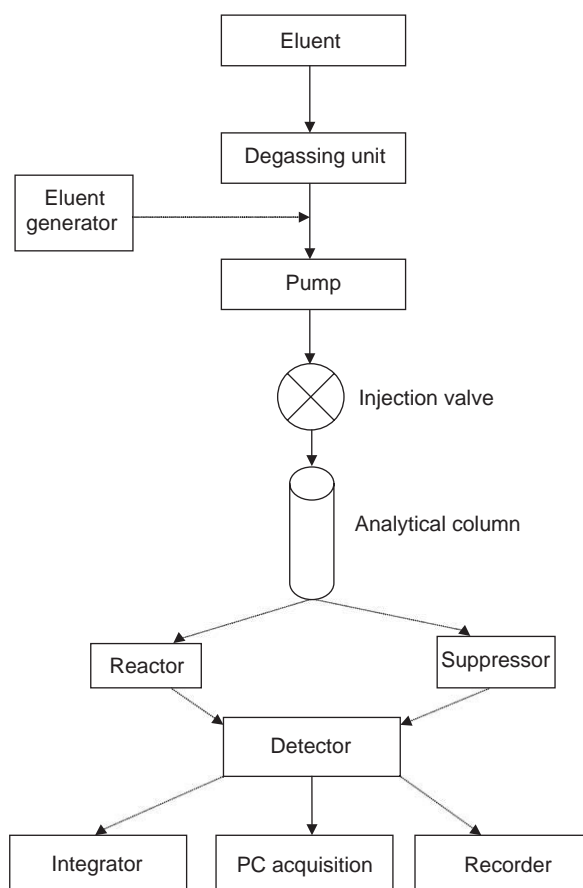
The first IC instruments appeared soon after the first publication about IC in 1977. At present, ~16–18 companies produce and sell laboratory equipment for IC totaling a sum of ~200 million dollars per year and automatic IC systems for industrial purposes totaling a sum of more than 10 million dollars. The wide application of IC for environmental control, foodstuff and drinks, in medicine, in energetics, in agrochemistry, and other areas is connected with the appearance of versatile instrumentation in the last few years. Different types of instruments ranging from small, portable, and personal systems to unattended industrial analytical systems have been developed.

**Scheme of an ion chromatograph** The scheme of a modern ion chromatograph is shown in Figure 1.

The typical ion chromatograph consists of an eluent generator or an eluent reservoir with degassing facilities; an eluent delivery system including an isocratic or a gradient high-pressure pump, sample pretreatment and injection system, which is either a manual valve or an autosampler; a separation column; a detection system, which includes modification of the effluent composition to achieve better selectivity or sensitivity; and a data acquisition system, which is mainly based on personal computers. The main feature of typical IC instrumentation is the so-called suppression system allowing the decrease of background conductivity of the effluent and providing better sensitivity for conductimetric detection.

### Evolution of Ion Chromatography Instrumentation

From the establishment of IC as an analytical technique, the evolution of IC instrumentation was directed toward improvements in separation efficiency



**Figure 1** Scheme of ion chromatograph.



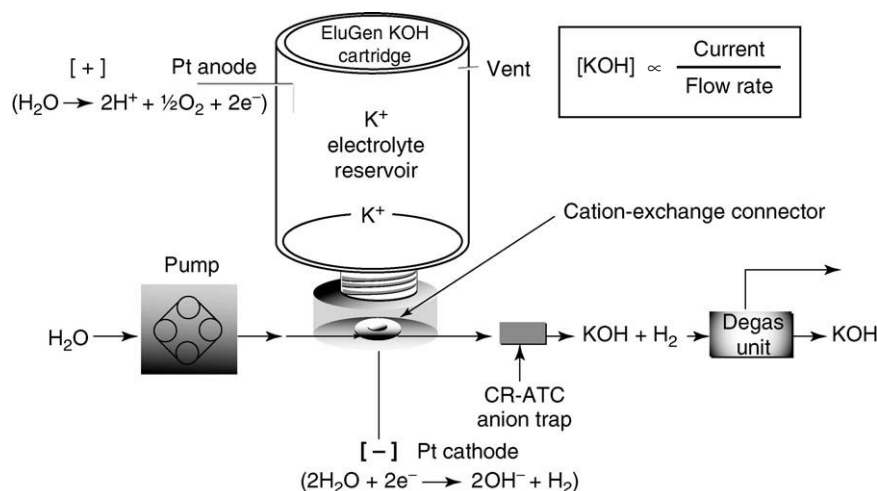
associated with getting more chromatographic peaks in a shorter time, increase of sensitivity of conductivity detection and introduction of new detectors, improvement of accuracy in IC analysis, and miniaturization of instruments. The main steps in the evolution of IC systems are presented in Table 1.

## Detection by Electrical Conductivity

All ions carry electric charges; therefore, all solutions of ions conduct electricity. Electrical conductivity provides a general method for detecting ions in solution. All ions do not conduct equally well; some migrate faster in an electric field than others.

**Table 1** Some evolution steps of IC instrumentation

Evolution step	Effect
Gradient elution in addition to isocratic elution	Increased number of separated ions in one chromatographic run, decrease of analysis time
Introduction of metal-free pumps and liquid paths from nonmetallic materials (ceramics, plastic PEEK, sapphire, PTFE e.a.) in ion chromatograph	Improved accuracy of determination of metal ions and related ions, decreased corrosion of instruments
Use of multiple column switching valves, especially for sample pretreatment and in simultaneous separation of anions and cations	More versatile IC analysis of complex samples
Development of dual-flow IC systems	Simultaneous separation of anions and cations
Automated in-line filtration of sample to prevent fouling of chromatographic columns	Longer lifetime of chromatographic columns
Programmable loop selection based on 10-port injection valve	Optimization of analytical procedure for analysis of samples with widely ranging concentration of ions
Development of PC databases for control of chromatographs and treatment of results	Significant simplification in data treatment
Miniaturization of systems based on micro-packed and capillary columns	Possibility of analysis of microsamples, better compatibility with MS detectors
Development of reliable chip ion chromatographs	
Introduction of new types of detectors including PAD, ICP-MS, etc.	Expansion of application area of IC, in particular, in analysis of sugars, amino acids, biogenic amines, etc.
Simultaneous separation and detection of anions and cations	Possibility to obtain full information of ion composition of the sample
Constant improvement of suppressor systems	Increase of sensitivity of conductimetric detection as the basic detection method
Development of reagent-free IC system with electrochemical generator of eluent (see Figure 2)	Possibility to manipulate with concentration of the eluent through electronics
Introduction of monolithic columns for	Ultra-fast separation, in particular, separation of common anions for 15 s ( $\text{H}_2\text{PO}_4^-$ , $\text{Cl}^-$ , $\text{NO}_2^-$ , $\text{Br}^-$ , $\text{NO}_3^-$ , $\text{ClO}_3^-$ , $\text{I}^-$ , $\text{SO}_4^{2-}$ )



**Figure 2** Schematic diagram of an electrochemical online KOH eluent generator. Based on the Dionex Eluent Generation Model EG40. Courtesy of Dionex Corporation.

**Table 2** Limiting equivalent ionic conductances ( $\text{S cm}^2 \text{equiv}^{-1}$ ) of ions in aqueous solutions

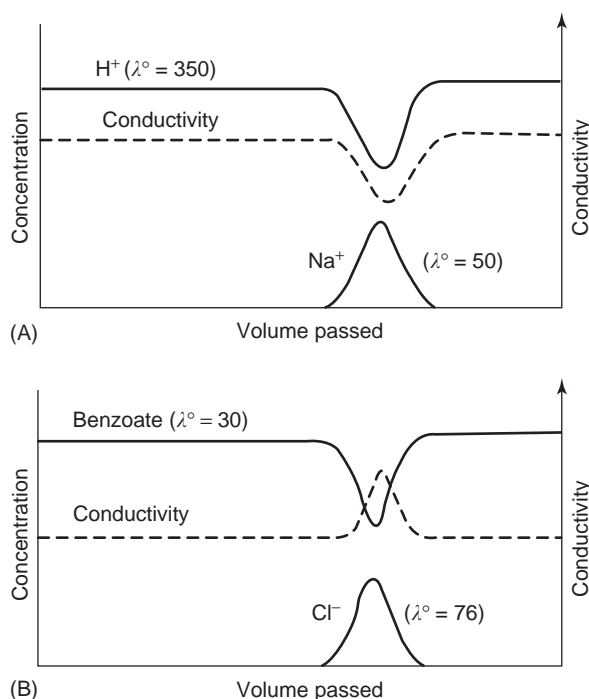
$\text{H}^+$	350	$\text{OH}^-$	198
$\text{Li}^+$	39	$\text{F}^-$	54
$\text{Na}^+$	50	$\text{Cl}^-$	76
$\text{K}^+$	74	$\text{NO}_3^-$	71
$\text{Mg}^{2+}$	53	$\text{HCO}_3^-$	45
$\text{Ca}^{2+}$	60	Benzoate	32
$\text{N}(\text{C}_2\text{H}_5)_4^+$	33	$\text{SO}_4^{2-}$	80
$\text{NH}_4^+$	73	$\text{CN}^-$	82
$\text{Ba}^{2+}$	64	$\text{Br}^-$	78
$\text{Sr}^{2+}$	59	$\text{I}^-$	77
$\text{Cu}^{2+}$	55	$\text{CO}_3^{2-}$	72
$\text{Zn}^{2+}$	53	$\text{PO}_4^{2-}$	69
$\text{Ni}^{2+}$	50	$\text{ClO}_4^-$	67

The conductance of a salt, AX, in dilute solution is the sum of the conductivities of the two ions:

$$\lambda_{\text{AX}} = \lambda_{\text{A}} + \lambda_{\text{X}}$$

The ions conduct independently and are free for electrostatic interaction, which can be neglected for our purposes. Table 2 lists some equivalent ionic conductances. These are the currents in amperes that would flow if a potential difference of 1 V were applied to a cell having two metallic plates, each  $1 \text{ cm}^2$  in area and placed 1 cm apart, with a solution between them that contained 1 g equiv of the ions in question in  $1 \text{ cm}^3$ . The unit of  $\lambda$ , therefore, is  $\Omega^{-1} \text{ cm}^2 \text{g-equiv}^{-1}$ . The reciprocal ohm,  $\Omega^{-1}$ , is commonly called the Siemens (S). The conductances of the dilute solutions used in IC are commonly in the range of  $\mu\text{S}$ .

When a solution of a salt, AX, is introduced at the head of a cation-exchange column into a flowing solution of BX, ions A enter the exchanger, move back and forth between the exchanger and the solution, and eventually come out in the effluent, taking the place of ions B. The concentration of ions X in the flowing solution remains constant, and therefore the sum of the concentrations of ions A and B must remain constant. A plot of these concentrations at the column exit, versus time, resembles the one shown in Figure 3. If the equivalent conductivities of ions A and B are equal, there is no change in the conductivity of the column effluent. If, however, ions B conduct better than ions A, the conductivity decreases as ions A emerge. This is the case when sodium ions are introduced into a flowing stream of hydrogen ions in a dilute acid. If ions A conduct better than ions B, the conductivity increases as ions A emerge. This is the case when calcium ions are displaced by an eluent containing the doubly charged ethylenediammonium cation, or, in anion exchange, when chloride or



**Figure 3** Concentrations and conductivities at column exit. (Reproduced from Walton HF (1987) *Ion-Exchange Chromatography*, ACS Audio Course C-93, p. 49. Washington, DC: American Chemical Society; © American Chemical Society.)

nitrate ions are displaced by an eluent of sodium benzoate.

The changes in concentration and conductivity shown schematically in Figure 3 take place against a high background conductivity of the eluent, BX, which is always present in excess. When high-capacity, gel-type resins are used the eluent concentrations are so high, between  $0.1$  and  $1 \text{ mol l}^{-1}$  or more, that the changes in conductivity caused by the analyte ions are practically undetectable. The only hope of detecting and measuring the conductivity changes lies in using very dilute eluent solutions, which, in turn, require ion exchangers of low capacity. This was the challenge that was faced by the developers of IC.

Even with the dilute eluents used in IC, the background conductivity is at least 10 times, and usually 1000 or more times the conductivity change that is in the analytical signal. The ratio of signal to background depends on the width of the chromatographic peak, the quantity of analyte injected, and the concentration of the eluent, as well as on the equivalent conductivities of the ions taking part.

There are two ways to overcome this difficulty. One is to accept the background conductivity and make it as small as possible by keeping the eluent concentration low. The signal is made as large as

possible by the correct choice of the eluent, and the baseline conductivity is offset electrically. As conductivity rises strongly with temperature, a sensitive and stable conductivity detector is essential, along with good temperature control. The second way to handle the background conductivity is to remove the excess eluent, or convert it to a low-conducting form, before the effluent from the column enters the detector. This operation requires a device called a 'suppressor'. The first method is called 'non-suppressed' or 'single-column' IC. The second is called 'suppressed' IC.

Contemporary conductivity detectors have full output ranges of 0.01–300  $\mu\text{S}$ , which is more than five orders of magnitude, electronic noise of  $\sim 0.2$  ns, cell volume of  $\sim 1$   $\mu\text{L}$ , and temperature stability of oven cell of  $\pm 0.01^\circ\text{C}$ . These characteristics allow one to reach high sensitivity and stability.

## Nonsuppressed Ion Chromatography

The eluent concentrations used in nonsuppressed detection are usually below  $0.01 \text{ mol l}^{-1}$ , and are often  $\sim 1 \text{ mmol l}^{-1}$ . They must, of course, match the capacity of the exchanger, which is commonly  $\sim 10 \text{ mequiv g}^{-1}$ . The typical gel-type cation-exchange resins have capacities of  $\sim 5 \text{ mequiv g}^{-1}$ . Eluents for cations include nitric acid for singly charged alkali metal ions and ethylenediammonium salts for doubly charged alkaline earth metal ions. Because the hydrogen ion has a weak affinity for sulfonic acid-type exchangers, relatively high concentrations must be used, up to  $0.01 \text{ mol l}^{-1}$ . The conductivity decreases when the alkali metal ions emerge.

In anionic IC, the common eluents are benzoate, phthalate, and salicylate. Their equivalent conductances are quite low; therefore, the conductivity of the effluent rises as simple inorganic anions like chloride and nitrate emerge. Because of their aromatic ring structure, these ions have a high affinity for polystyrene-based exchangers, and the eluents can be made dilute without unduly increasing the elution time. Phthalate is better than benzoate in that the fully deprotonated ions are doubly charged; doubly charged ions are more strongly bound than singly charged ions. The ionization constants ( $\text{pK}_a$  values) of phthalic acid ( $\text{H}_2\text{P}$ ) are 2.9 and 5.5. Above pH 5.5 the doubly charged phthalate ions predominate; below pH 5.5 the dominant ions are singly charged  $\text{HP}^{-1}$ , which have a smaller affinity for the resin. Thus, the displacing power of a phthalate eluent can be adjusted over a wide range by adjusting the pH.

Another suitable eluent for anion chromatography is a solution containing boric acid and a polyol, gluconic acid. These substances combine to give acids that are stronger than boric acid. At pH 8.5, the usual pH for borate–gluconate eluents, most of the acids are ionized and the eluting power is adequate. The conductivity of this eluent is low, and rises when other anions are eluted.

Sodium or potassium hydroxide is occasionally used as eluent in nonsuppressed IC. The hydroxide ion has a high conductivity, so the conductivity falls when other anions are eluted. The reason that this eluent is not much used is that the affinity of the hydroxide ion for the exchanger is low. High concentrations of hydroxide must be used, or long elution times must be accepted. Hydroxide eluents are used for the chromatography of anions of very weak acids: borate, carbonate, cyanide, silicate, sulfide. These anions cannot be detected by suppressed chromatography.

## Suppressed Ion Chromatography

In suppressed IC, the excess of eluent ions is removed or converted to a nonconducting form before the column effluent reaches the detector. Generally, the removal is accomplished by an ion-exchange resin of opposite charge to the exchanger used in the separator column. Consider the separation of sodium and potassium ions on a column of cation-exchange resin, with dilute nitric acid as the eluent. These ions come out of the column as their nitrates, accompanied by a considerable excess of nitric acid. If this solution is passed into a column of anion-exchange resin in its hydroxide form, that is, carrying exchangeable hydroxide ions, the nitrates of sodium and potassium are converted to their hydroxides, while the excess nitric acid is converted to water, the nitrate ions being held back in the exchanger. A conductivity detector placed after this column 'sees' peaks caused by the highly conducting sodium and potassium hydroxides, rising from a background conductivity that is nearly zero.

Two anions, say chloride and bromide, could be separated on a column of anion-exchange resin with sodium hydroxide solution as the eluent. They would emerge from the column as their sodium salts, accompanied by a large excess of sodium hydroxide. If this solution is passed through a column of cation-exchange resin that carries hydrogen ions, sodium chloride and sodium bromide form hydrochloric acid and hydrobromic acid, while the excess sodium hydroxide forms water, the sodium ions being retained by the cation-exchange resin.

As already noted, the hydroxide ion has weak displacing power. The retention of bromide, for instance, in a sodium hydroxide eluent would be excessively long, and the peak, when it appears, would be broad and shallow. Another objection to sodium hydroxide is that it readily absorbs carbon dioxide from the atmosphere. Carbonate ions and hydrocarbonate ions are much more strongly held by anion exchangers than are hydroxide ions; carbonate is more strongly held than hydrocarbonate. Thus, a favored eluent for anion chromatography is a mixture of sodium carbonate and sodium hydrogencarbonate, concentrations  $2.2$  and  $2.8 \text{ mmol l}^{-1}$ , respectively. Of course, these concentrations and their ratio can be varied; the higher the proportion of doubly charged carbonate ions, the greater the displacing power. Mixtures of carbonate and hydroxide ions are also used.

When carbonate or hydrocarbonate ions are passed into a column of cation-exchange resin in its hydrogen form, they are converted to carbonic acid, which has a low conductivity. Ions of strong acids, like  $\text{Cl}^-$  and  $\text{Br}^-$ , produce their respective acids, which are highly conducting.

In the early form of IC, columns of high-capacity, gel-type cation- and anion-exchanger resins were indeed used to remove the excess eluent; they were placed immediately after the separator columns, which were packed with low-capacity, surface-functional exchangers. These high-capacity columns were called suppressor columns. They were clumsy, because they needed regeneration from time to time, and they caused additional band spreading. Today, their place has been taken by ion-exchange membranes. The evolution of suppressors is presented in Table 3.

## The Membrane Suppressor

An ion-exchange membrane is simply an ion-exchange polymer in the form of a thin sheet or film. It swells in water, and its internal structure is like that of the gel-type exchanger beads. The

internal concentration of fixed ions should be as high as possible. Cation-exchanging membranes have fixed negative charges and anion-exchanging membranes have fixed positive charges. The counterions can exchange with other ions of like sign in the surrounding solution, and if a suitable force is applied, such as an electric field or a concentration difference, counter-ions can flow across the membrane, always maintaining electrical neutrality. Co-ions, on the other hand, are excluded from the membrane in accordance with the Donnan equilibrium. Exclusion is not total, but nearly so. The passage of co-ions across the membrane is determined not only by their (very small) internal concentration but also by their speed of diffusion.

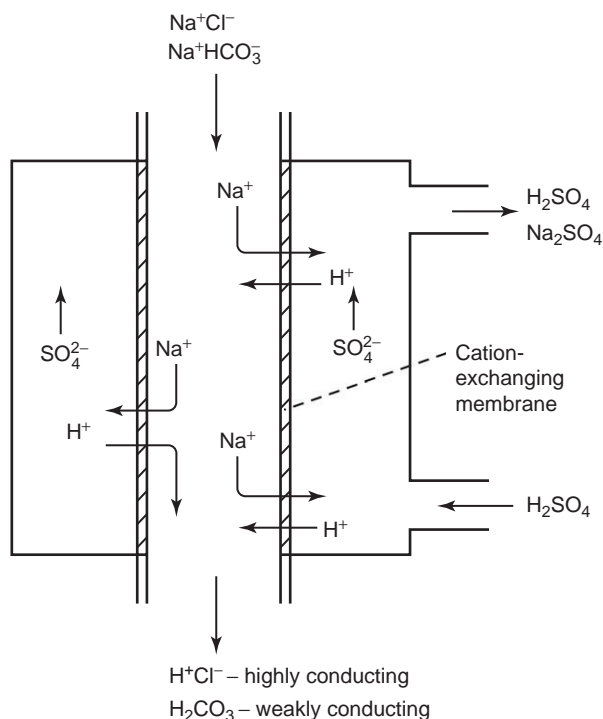
Suppose that the effluent from a cation-exchange column, containing sodium chloride plus excess hydrochloric acid, is made to flow along one side of an anion-exchanging membrane, while on the other side of the membrane a solution of a quaternary ammonium hydroxide is flowing. (The quaternary ammonium ion is chosen for its slow diffusion rate.) The cations,  $\text{Na}^+$ , cannot enter the membrane; chloride ions enter the membrane and change places with hydroxide ions. Hydroxide ions combine with hydrogen ions in the effluent to form water; sodium hydroxide, in equivalent amount to the sodium chloride originally present, flows on. The effect is just like that of the suppressor column previously described, except that regeneration of the suppressor device is continuous and there is less band broadening.

In IC, a device is used (Figure 4) that has two membranes mounted very close together – dimensions of  $\sim 1 \text{ cm}$  by  $15 \text{ cm}$ , spaced some  $50 \mu\text{m}$  apart, with a mesh of ion-exchange fibers in this space to deflect the flowing stream on to the membrane walls. On the outsides of the membranes are channels through which the regenerating solution, tetramethylammonium hydroxide in our example, flows in the opposite direction to the eluent.

For the chromatography of anions, cation-exchange membranes are used, and the regenerant is

**Table 3** Evolution of suppression devices

	<i>Suppression systems</i>	<i>Year</i>	<i>Note</i>
1	Packed column suppression	1975	(Small e.a.) only isocratic operation
2	Hollow – fiber suppressor	1981	External regeneration solution
3	Micromembrane suppressor	1985	Expanded to gradient operation
4	Autoregeneration	1987	Online, continuous regeneration
5	Micromembrane	1991	Electrolytic membrane-based suppressor, 2 mm
6	SRS-1 autosuppression	1992	Self-regenerating suppression, no external regeneration solution; high capacity
7	SRS-ULTRA	1997–2000	Ultimate in reliability and performance



**Figure 4** Membrane suppressor (not to scale). (Reproduced with permission from Walton and Rocklin (1990) *Ion Exchange in Analytical Chemistry*. Boca Raton, Florida: CRC Press; © CRC Press.)

commonly a solution of sulfuric acid. Sodium chloride is now converted to hydrochloric acid, and sodium hydrogencarbonate or carbonate eluent is converted to carbonic acid. (The concentration is too low to form carbon dioxide gas.) Some of the carbon dioxide diffuses as neutral molecules through the suppressor membranes.

The eluents used in suppressed IC must obviously be compatible with the membrane suppressor. An interesting eluent for cation chromatography is 2,3-diaminopropionic acid (DAP).

The ionization constants are ( $\text{pK}_a$  values) 1.3, 6.7, and 9.4. In strongly acid solution, DAP is a doubly charged cation and a strong eluent for chromatography, suitable for eluting doubly charged cations. At pH 4.0, the singly charged cation predominates. Above pH 6.7, the electrically neutral species predominates. The membrane suppressor converts DAP to the electrically neutral form, which does not conduct electricity.

A great advantage of the membrane suppressor is that it can suppress fairly concentrated effluents, up to  $0.1 \text{ mol l}^{-1}$ , a useful property in analyzing complex samples where one is looking for minor components in the presence of several other materials. There is less chance of overloading the column.

Another significant advantage is that gradient elution becomes possible. It is impossible to carry out gradient elution if the background conductivity is changing steeply during the run.

## Other Detection Methods

Any of the methods of detection used in liquid chromatography can be used in IC, though some are more useful than others. If the eluent does not affect the detector the need for a suppressor disappears. Common means of detection in IC are: ultraviolet (UV) absorption, including indirect absorption; electrochemical, especially amperometric and pulsed amperometric; and postcolumn derivatization. Detectors: atomic absorption spectrometry, chemiluminescence, fluorescence, atomic spectroscopic, refractive index, electrochemical (besides conductivity) including amperometric, coulometric, potentiometric, polarographic, pulsed amperometric, inductively coupled plasma emission spectrometry, ion-selective electrode, inductively coupled plasma mass spectrometry, bulk acoustic wave sensor, and evaporative light-scattering detection.

### Ultraviolet Absorption

Most simple inorganic cations and anions are transparent in the UV, but the larger, more polarizable anions absorb UV radiation, absorbing more strongly at shorter wavelengths. Absorbances at 200–210 nm are sufficiently strong to detect bromide, iodide, nitrate, nitrite, thiocyanate, thiosulfate, and chromate. Carboxylate ions absorb satisfactorily below 215 nm.

In indirect UV detection, an eluent is used that absorbs in the UV region, for example, benzoate or salicylate solution. When an ion emerges from the column that is transparent in the UV, like chloride, the concentration of eluent ions drops (see Figure 3): the absorbance decreases, and a negative peak is produced by the detector. In cation-exchange chromatography, an eluent containing cerium(III) ions has been used. The main difficulty with indirect UV detection is that the drop in absorbance is observed against a large background.

### Electrochemical Methods

**The basic advantages and features of amperometric detection** Amperometric detection is based on measuring of an electrical current arising during oxidation (reduction) of a decomposed substance on the surface of a working electrode, under a certain potential. The amperometric detector has series of



advantages: (1) a low limit of detection ( $10^{-9}$ – $10^{-13}$  g); (2) high selectivity of microimpurity in objects of an environment, in foodstuff, in biological liquids is detected in the analyte with a low limit of detection, thus the basic substances (matrix) are not detected by the detector; (3) small volume of the cell (0.1–5  $\mu$ l); (4) the simplicity of design; and (5) low cost (in comparison with optical detectors the amperometric detector is cheaper).

Halide, sulfide, and cyanide ions have been detected by passing the column effluent along a surface of metallic silver that is maintained at a positive potential. When one of these ions passes, a silver atom releases an electron to form a silver ion, which is precipitated or complexed; simultaneously, at the counter-electrode, a hydrogen ion accepts an electron to form a hydrogen atom, and current flows.

Few simple inorganic ions can be reduced or oxidized electrochemically, but many organic anions can be oxidized at an electrode of gold or platinum. The electrode can be fouled by partly oxidized decomposition products, which is a problem for reproducibility of results. Fouling can be avoided by the 'pulsed amperometric' technique. The electrode is first held for a few hundredths of a second at the working potential, a potential that is sufficiently anodic to oxidize the analyte eluting from the column, and the current is measured and recorded. Then, the potential is raised to a high value to convert any solid oxidation products to carbon dioxide. During this step the metal surface is also oxidized, so that a third step is necessary in which the electrode is held at a negative potential to reduce the oxide back to the free metal. Then the working potential is re-established and the cycle is repeated. The current at the working potential is proportional to the concentration of oxidizable material in the effluent.

Various electrodes are suitable for detection (see Table 4).

Pulsed amperometric detection (PAD) is used with good success in the chromatography of sugars. Sugars are very weak acids with  $pK_a$  values  $\sim 12$ – $14$ , and they can be separated by anion exchange with  $0.01$ – $0.1$  mol  $l^{-1}$  sodium hydroxide as eluent. As can be seen by writing the half-reaction for the oxidation of a sugar, a high pH is necessary for this oxidation to take place. PAD in alkaline solution detects many organic compounds, including amino acids and ascorbic acid.

Sugars are more commonly analyzed on a 'carbohydrate column'. This is packed with a gel-type, high-capacity cation-exchange resin loaded with calcium ions. The eluent is pure water, and detection is usually by refractive index. Sugars and polyhydric alcohols are retained, primarily, by coordination of hydroxyl groups with the calcium ion. The stability of the coordination complex, and hence the retention time, depends on the orientation of these hydroxyl groups, leading to the separation of closely related sugars.

### Postcolumn Derivatization

In this technique, a chemical reagent is pumped into the effluent stream as it comes out of the column. The mixed stream flows through a delay coil to allow time for the reaction to become complete, after which it passes into the detector. A good example of a postcolumn reaction is the ion-exchange chromatography of amino acids. These compounds do not absorb light themselves, but they combine with certain reagents to form strongly colored or fluorescent products. Ninhydrin forms red or purple products; o-phthalaldehyde plus a thiol yields products that are strongly fluorescent and allow very sensitive

**Table 4** The list of types of working electrodes and substances determined on them

<i>Material of a working electrode</i>	<i>Determined substances</i>
Glassy carbon	General purpose, but is most suitable for the analysis of: nitrite, sulfide, bromide, iodide, cyanide, sulfide, disulfides, etc.
Gold	Aliphatic alcohols, monosaccharides, disaccharides, oligopolysaccharides, aliphatic amines, aminoalcohols, amino sugars, amino acids
Platinum	Hypochlorite, arsenite, hydrazine
Silver	Cyanide, sulfide, thiosulfate, thiocyanide, bromide, iodide, hydrosulfide
Copper	Carbohydrates, amino acid, peptides, polypeptides, protein
Nickel	Carbohydrates, alcohols, amino acid
Palladium	Aromatic hydrocarbons

Other electrodes: electrodes modified by enzymes; electrodes modified by current-carrying polymers; electrodes modified by phthalocyanines, ferrocene, etc.

Admixture and alloys: Ni–Ti, Ni–glassy carbon, Ni–Cr, Ni–Cu, Ni–Cr–Fe, Pt–glassy carbon, Au–Cr, Co, etc.

detection. In inorganic analysis, postcolumn derivatization is used to detect and measure ions of transition metals and lanthanides. Reagents used for this purpose are 4-(2-pyridylazo)resorcinol, and the 'Arsenazo' dyes. They give products that absorb strongly in the visible region.

See also: **Amperometry**. **Carbohydrates**: Sugars – Chromatographic Methods. **Derivatization of Analytes**. **Electrophoresis**: Principles. **Flow Injection Analysis**: Principles. **Ion Exchange**: Principles. **Liquid Chromatography**: Column Technology; Chiral Analysis of Amino Acids. **Sensors**: Amperometric Oxygen Sensors.

## Further Reading

- Buchberger WW and Haddad PR (1997) Advances in detection techniques for ion chromatography. *Journal of Chromatography* 789: 67–83.
- Fritz JS and Gjerde DT (2000) *Ion Chromatography*, 3rd edn. Heidelberg and New York: Huethig.
- Haddad PR (2001) Ion chromatography retrospectivity. *Analytical Chemistry* 73: 266A–273A.
- Haddad PR, Jackson PE, and Shaw MJ (2003) Development in suppressor technology for inorganic ion analysis by ion chromatography using conductivity detection. *Journal of Chromatography A* 1000: 725–742.
- Lee HJ and Girault HH (1998) Amperometric ion detector for ion chromatography. *Analytical Chemistry* 70: 4280–4285.
- Lopez-Ruiz B (2000) Advances in the determination of inorganic anions by ion chromatography. *Journal of Chromatography A* 881: 607–627.
- Lucy CA (2003) Evolution of ion-exchange: From Moses to the Manhattan project to Modern Times. *Journal of Chromatography A* 1000: 711–724.
- Nesterenko PN (2001) Simultaneous separation and detection of anions and cations in ion chromatography. *Trends in Analytical Chemistry* 20: 311–319.
- Pietrzyk DJ (1998) Ion chromatography by HPLC. *Chromatographic Science Series* 78: 413–462.
- Saari-Nordhaus R and Anderson JM (2002) Recent advances in ion chromatography suppressor improve anion separation and detection. *Journal of Chromatography A* 956: 15–22.
- Sarzanini C (2002) Recent developments in ion chromatography (review). *Journal of Chromatography A* 956: 3–14.
- Stevens TS (2002) The membrane suppressor: A historical perspective (review). *Journal of Chromatography A* 956: 43–46.
- Viehweger KH (ed.) (2000) *Practical Ion Chromatography: An Introduction*. Metrohm: Herisau, Switzerland.
- Weiss J (2001) *Ion Chromatography*, 3rd edn. New York: Wiley-VCH.
- Yuan JP and Chen F (2001) Indirect photometric ion chromatographic analysis. *Biotechnology Letters* 23: 757–760.

## Ion Chromatography Applications

**B Paull**, Dublin City University, Dublin, Republic of Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

When first developed, ion chromatography (IC) was based on the use of a low-capacity anion exchange separator column used with a basic eluent and a suppressor column, with conductimetric detection. This allowed the sensitive detection of a limited number of inorganic anions in aqueous samples in reasonably short analysis times. Later work saw the introduction of nonsuppressed IC, which utilized low-conductivity eluents, mainly organic aromatic weak acids, that could be used without a suppressor module, thus simplifying the chromatographic system. Nonsuppressed IC, being less sensitive than suppressed IC, was generally more applied to

samples containing higher solute concentrations. Improvements in stationary-phase technology led to more efficient simultaneous separations of inorganic and organic anions, and combined with the development of improved suppressor systems, such as self-regenerating membrane suppressors, this saw IC establish itself as the method of choice for anion analysis of aqueous samples. In the application of IC to the determination of organic and inorganic cations, there has also been much progress. Again, both suppressed and nonsuppressed IC systems have been used, predominantly for the determination of alkali and alkaline earth metal ions and organic amines, again mostly in aqueous-based sample matrices. The following is a simple review of some of the more interesting applications of IC, focusing on those applications based on the use of ion exchange stationary phases, although alternative approaches to the separation of ionic species, such as ion interaction liquid chromatography, will also be included.

detection. In inorganic analysis, postcolumn derivatization is used to detect and measure ions of transition metals and lanthanides. Reagents used for this purpose are 4-(2-pyridylazo)resorcinol, and the 'Arsenazo' dyes. They give products that absorb strongly in the visible region.

See also: **Amperometry**. **Carbohydrates**: Sugars – Chromatographic Methods. **Derivatization of Analytes**. **Electrophoresis**: Principles. **Flow Injection Analysis**: Principles. **Ion Exchange**: Principles. **Liquid Chromatography**: Column Technology; Chiral Analysis of Amino Acids. **Sensors**: Amperometric Oxygen Sensors.

## Further Reading

- Buchberger WW and Haddad PR (1997) Advances in detection techniques for ion chromatography. *Journal of Chromatography* 789: 67–83.
- Fritz JS and Gjerde DT (2000) *Ion Chromatography*, 3rd edn. Heidelberg and New York: Huethig.
- Haddad PR (2001) Ion chromatography retrospectivity. *Analytical Chemistry* 73: 266A–273A.
- Haddad PR, Jackson PE, and Shaw MJ (2003) Development in suppressor technology for inorganic ion analysis by ion chromatography using conductivity detection. *Journal of Chromatography A* 1000: 725–742.
- Lee HJ and Girault HH (1998) Amperometric ion detector for ion chromatography. *Analytical Chemistry* 70: 4280–4285.
- Lopez-Ruiz B (2000) Advances in the determination of inorganic anions by ion chromatography. *Journal of Chromatography A* 881: 607–627.
- Lucy CA (2003) Evolution of ion-exchange: From Moses to the Manhattan project to Modern Times. *Journal of Chromatography A* 1000: 711–724.
- Nesterenko PN (2001) Simultaneous separation and detection of anions and cations in ion chromatography. *Trends in Analytical Chemistry* 20: 311–319.
- Pietrzyk DJ (1998) Ion chromatography by HPLC. *Chromatographic Science Series* 78: 413–462.
- Saari-Nordhaus R and Anderson JM (2002) Recent advances in ion chromatography suppressor improve anion separation and detection. *Journal of Chromatography A* 956: 15–22.
- Sarzanini C (2002) Recent developments in ion chromatography (review). *Journal of Chromatography A* 956: 3–14.
- Stevens TS (2002) The membrane suppressor: A historical perspective (review). *Journal of Chromatography A* 956: 43–46.
- Viehweiger KH (ed.) (2000) *Practical Ion Chromatography: An Introduction*. Metrohm: Herisau, Switzerland.
- Weiss J (2001) *Ion Chromatography*, 3rd edn. New York: Wiley-VCH.
- Yuan JP and Chen F (2001) Indirect photometric ion chromatographic analysis. *Biotechnology Letters* 23: 757–760.

## Ion Chromatography Applications

**B Paull**, Dublin City University, Dublin, Republic of Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

When first developed, ion chromatography (IC) was based on the use of a low-capacity anion exchange separator column used with a basic eluent and a suppressor column, with conductimetric detection. This allowed the sensitive detection of a limited number of inorganic anions in aqueous samples in reasonably short analysis times. Later work saw the introduction of nonsuppressed IC, which utilized low-conductivity eluents, mainly organic aromatic weak acids, that could be used without a suppressor module, thus simplifying the chromatographic system. Nonsuppressed IC, being less sensitive than suppressed IC, was generally more applied to

samples containing higher solute concentrations. Improvements in stationary-phase technology led to more efficient simultaneous separations of inorganic and organic anions, and combined with the development of improved suppressor systems, such as self-regenerating membrane suppressors, this saw IC establish itself as the method of choice for anion analysis of aqueous samples. In the application of IC to the determination of organic and inorganic cations, there has also been much progress. Again, both suppressed and nonsuppressed IC systems have been used, predominantly for the determination of alkali and alkaline earth metal ions and organic amines, again mostly in aqueous-based sample matrices. The following is a simple review of some of the more interesting applications of IC, focusing on those applications based on the use of ion exchange stationary phases, although alternative approaches to the separation of ionic species, such as ion interaction liquid chromatography, will also be included.

## Environmental Applications

### Natural Waters

**Inorganic anions** One of the commonest applications of IC is the analysis of natural nonsaline water samples, such as subsurface waters, spring waters, streams, river waters, and lakes. Particular interest lies in the use of IC as a laboratory-based technique for the routine monitoring of the above sample types for nutrient anion concentrations, such as nitrates, nitrites, sulfates, and phosphates. In most cases suppressed IC is used, due to its ability to quantify the above species, which are often present at sub-mmol l<sup>-1</sup> concentrations. The US EPA Method 300 (Determination of fluoride, chloride, nitrite, nitrate, phosphate, and sulfate in water samples by IC) describes suitable conditions for this particular application, based on the use of a Dionex IonPac AS4A anion exchange column, with a carbonate/bicarbonate eluent (or a similar stationary-phase/eluent combination that produces similar or better selectivity and efficiency) and suppressed conductivity detection. Utilizing these conditions, resolution of the above anions from each other and matrix anions, typically chloride, is possible. Approximate linear ranges quoted for the above anions, using a similar analytical setup to that described above, are between 0.01 and 5 mmol l<sup>-1</sup>, with detection limits in the order of 0.1–1 µmol l<sup>-1</sup>. Similar detection limits for nitrate and nitrite can be achieved using IC combined with direct UV absorbance detection at 225 nm. In this case large matrix peaks resulting from excess sulfate and chloride are essentially eliminated due to their UV transparency at this wavelength. With IC applications utilizing direct UV detection, the eluting anion within the mobile phase must also be UV transparent if sensitive detection is to be obtained.

Nonsuppressed IC methods for natural water samples are generally 1–2 orders of magnitude less sensitive for common inorganic anions than suppressed conductivity methods, but still find useful application for the determination of higher-concentration matrix anions, such as chloride and sulfate. Typical eluents used include *p*-hydroxybenzoic acid or phthalic acid, often used with small amounts of organic solvent to improve peak shapes. Such eluents can also be employed for indirect UV detection, although again sensitivity is somewhat less than for direct UV detection.

The determination of the above nutrient anions in natural saline samples, such as coastal seawaters, is also of interest to environmental scientists, and several IC methods have been developed that can tolerate the high salt content of such samples. Column switching techniques have been used for this type of application.

Saline samples are injected onto short high-capacity anion exchange guard columns, which are separated from the main analytical anion exchange column by a switching valve. Correct timing of the switching valve allows the early eluting excess chloride to be directed to waste before the more retained anions of interest, such as bromide, phosphate, sulfate, nitrite, nitrate, and residual chloride, are eluted. Redirecting the later eluting anions onto the analytical column allows separation and detection to take place and eliminates large-matrix chloride peaks.

An alternative approach is to include the matrix anion in the eluent itself, so-called 'matrix elimination IC'. Here a sodium chloride eluent is used with a strong anion exchange column, thus eliminating any 'self-elution' problems when analyzing saline samples. The method is most suitable for strongly retained anions such as iodide, and has indeed been used for the determination of this particular anion in seawater samples, with trace level detection made possible through postcolumn reaction and visible absorbance detection.

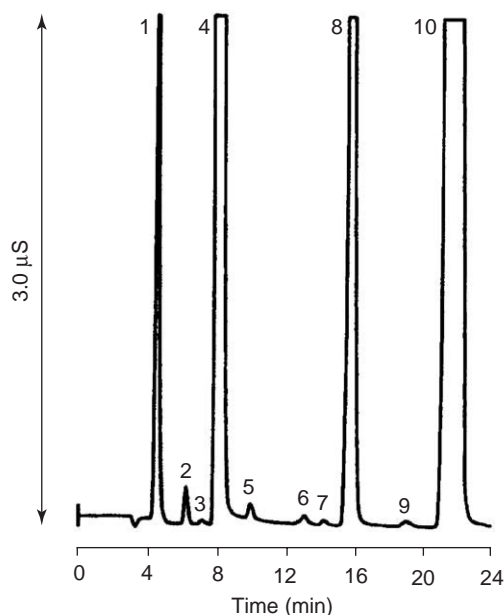
**Inorganic cations** For the determination of common inorganic cations in nonsaline natural waters, IC competes with atomic spectroscopy as the method of choice. Despite this, IC is used by many monitoring agencies for the determination of alkali and alkaline earth metal cations and, to a lesser extent, selected transition metal cations. Two approaches are commonly taken. Firstly, for the determination of alkali metal ions, a strong cation exchanger (sulfonated) is generally used with a strong acid eluent and combined with suppressed conductivity detection. For the simultaneous determination of alkali and alkaline earth metal ions, a weaker cation exchanger (carboxylated or carboxylated and phosphonated, e.g., a Dionex IonPac CS12A column) is more suitable, and used with a weaker eluent and either suppressed conductivity detection or indirect conductivity detection. Such a system allows the simultaneous determination of lithium, sodium, ammonia, potassium, magnesium, and calcium in water samples (EPA Method 300.7).

### Treated Waters

Treated waters for domestic use are routinely analyzed using IC for both naturally present common inorganic anions and trace anionic contaminants, several classes of which actually originate as by-products of the treatment processes themselves. Oxyhalides, which originate from various drinking water disinfection processes, such as chlorination and ozonation, can be found present in finished drinking



waters and require monitoring at sub-micromolar concentrations. For example, chlorate and chlorite can result from treatment with chlorine dioxide, and bromate and iodate can be formed from the treatment of bromide- and iodide-containing drinking waters with ozonation. The 1998 European Drinking Water Directive set a mandatory standard limit of  $10 \mu\text{g l}^{-1}$  for bromate in drinking water to be in place by 2008, necessitating sensitive analytical methods for this particular analyte to be developed. When using IC for the determination of bromate in drinking water, a high-capacity anion exchange column is generally used to allow the injection of larger sample volumes, thus improving detection limits. Detection of ultra-trace levels of bromate is achieved via postcolumn reaction, followed by visible detection (450 nm), using *o*-dianisidine dihydrochloride (as in EPA Method 317.0). Alternative detection can be achieved via postcolumn reaction with potassium iodide–ammonium heptamolybdate (triiodide method), although both methods result in a similar sensitivity for bromate with detection limits of  $\sim 0.5 \mu\text{g l}^{-1}$ . **Figure 1** shows an ion chromatogram of a mixture of common inorganic anions and trace oxyhalides, here using only suppressed conductivity detection (obtained using EPA Method 300.1, Part B).



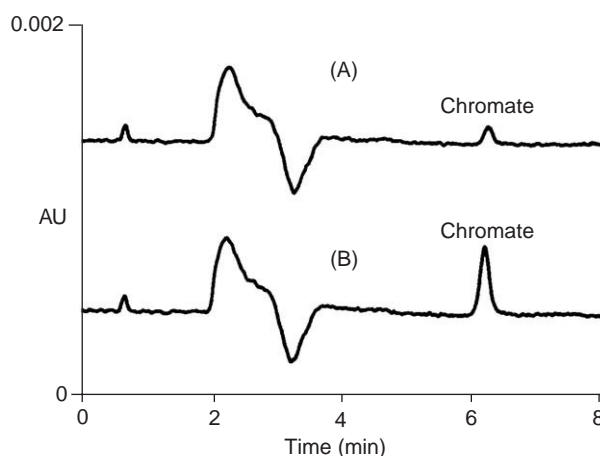
**Figure 1** Chromatogram of 10 inorganic anions and oxyhalides in a water sample using suppressed ion chromatography. Peak identification: 1 = fluoride, 2 = chlorite, 3 = bromate, 4 = chloride, 5 = nitrite, 6 = bromide, 7 = chlorate, 8 = nitrate, 9 = phosphate, 10 = sulfate. (Reprinted with permission from Saari-Nordhaus R and Anderson JM Jr. (2002) Recent advances in ion chromatography suppressor improve anion separation and detection. *Journal of Chromatography A* 956: 15–22; © Elsevier.)

Two compounds that are currently generating much concern are azide and perchlorate, salts of which are currently used in the explosives and pyrotechnics industry, with perchlorate also used as a primary oxidant in solid rocket fuel. IC can be used for the monitoring of both analytes in natural and treated waters. For perchlorate, EPA Method 314.0 (Determination of perchlorate in drinking water by ion chromatography) has been developed, which suggests the use of a Dionex IonPac AS16 anion exchange column (or equivalent) with a 35 mM NaOH eluent and suppressed conductivity detection.

Finally, there has also been much interest in recent years in the contamination of groundwater, well water, and drinking water supplies with hexavalent chromium. An EPA method has been developed utilizing IC for hexavalent chromium determinations (EPA Method 218.6). This method specifies the use of a high-capacity Dionex IonPac AS7 anion exchange column and UV/Vis detection following post-column reaction with diphenylcarbazine. **Figure 2** shows an ion chromatogram of a spiked and unspiked drinking water sample obtained using a modified version of EPA Method 218.6.

**Table 1** lists some of the IC methods prescribed by the US EPA for the analysis of drinking water including several oxyhalide disinfection by-products (DBPs).

Note how inductively coupled plasma mass spectrometry (ICP-MS) is also used for bromate detection in Method 321.8.



**Figure 2** Ion chromatograms of chromate in drinking waters: (A) unspiked sample, (B) sample spiked with  $0.2 \mu\text{g l}^{-1}$  Cr(VI). (Reprinted with permission from Thomas DH, Rohrer JS, Jackson PE, Pak T, and Scott JN (2002) Determination of hexavalent chromium at the level of California public health goal by ion chromatography. *Journal of Chromatography A* 956: 255–259; © Elsevier.)



**Table 1** EPA IC methods for drinking water analysis

<i>EPA method</i>	<i>Analysis</i>
Method 300.1	Determination of inorganic anions and oxyhalides in drinking water by IC
Method 300.7	Determination of inorganic cations in drinking water by IC
Method 314.0	Determination of perchlorate in drinking water by IC
Method 317.0	Determination of inorganic oxyhalide disinfection by-products in drinking water using IC with the addition of a postcolumn reagent for trace bromate analysis
Rev. 2.0	
Method 218.6	Determination of dissolved hexavalent chromium in drinking water, groundwater, and industrial wastewater effluents by IC
EPA Method 321.8	Determination of bromate in drinking waters by IC with ICP-MS detection
EPA Method 326.0	Determination of inorganic oxyhalide disinfection by-products in drinking water using IC incorporating the addition of a suppressor acidified postcolumn reagent for trace bromate analysis

### Soil Analysis

**Inorganic anions** IC is used extensively for the determination of common inorganic anions in soil extracts. The key to valuable data when carrying out such analyses is the use of correct and reproducible extraction methods. For the determination of inorganic anions, often the soil is simply extracted using water. Such an application of IC is used to provide information on the fate of nutrient anions resulting from agricultural practices and to determine soil nutrient retention and leaching rates.

**Organic anions** Both anion exchange chromatography and ion exclusion chromatography have been used extensively for the determination of low-molecular-weight organic acids in soil extracts. Malic acid, malonic acid, maleic acid, succinic acid, fumaric acid, ascorbic acid, citric acid, isocitric acid, succinic acid, tartaric acid, oxalic acid, and glycolic acid can all be determined using these techniques.

**Inorganic cations** For the extraction of cations from soils, a number of approaches are used, dependent on whether the analyst wishes to determine labile or nonlabile cation concentrations. Extractions can be carried out using simply water or electrolyte solutions, or strong acid solutions for nonlabile cations. In this way, IC can also be used to determine the ion exchange capacity of the soils in question.

Recently, IC has also found increased application in the field of metal speciation. Short anion exchange columns have been used for the rapid separation of anionic species of arsenic, selenium, and chromium,

extracted from contaminated soils, and followed by elemental specific detection.

### Atmospheric Samples

The determination of ambient concentrations of gaseous nitrogen- and sulfur-containing species (predominantly nitrate and sulfate) in the atmosphere has also been carried out using IC after collection using passive samplers. Wet denuder systems have also been developed for the collection of soluble ionogenic trace gases and soluble ionic species adsorbed onto atmospheric particles, used on-line with IC for continuous monitoring purposes.

The use of large-volume filter-based samplers for the collection of atmospheric particulates has also been combined with IC for the determination of absorbed metal species. Collected samples are extracted from the filters and desorbed from the particulate matter using acidic solutions prior to analysis by IC combined with elemental selective detection such as ICP-MS. Hexavalent chromium, platinum, and palladium have been determined in this way, as have many other transition and heavy metal ions.

**Rainwater** Automated rainwater collectors have been used in combination with IC for the determination of low-level concentrations of dissolved inorganic anions or cations. Due to the 'clean' nature of the sample, online preconcentration of the analytes can be readily achieved using short ion exchange preconcentrator cartridges onto which large sample volumes can be loaded, prior to elution onto the appropriate separator ion exchange column. This approach is used for the ultratrace analysis of rainwater samples.

**Table 2** Examples of foodstuffs to which IC has been applied for anion determinations following appropriate extraction methods

<i>Anions</i>	<i>Sample matrix</i>
Nitrate, nitrite, sulfate, phosphate, chloride	Milk products, fruits and fruit juices, beverages and alcoholic products, meat products, bakery products, vegetables, cereals
Fluoride	As above plus citrus fruits and leaves and spinach
Iodide	Seafood, food colorings, and iodized table salt
Sulfite	Beer, lemon juice, potatoes, seafood, fruits (grapes)
Bromide	Milk, food colorings, rice products, bakery products
Chlorite and chlorate	Vegetables
Bromate	Bakery products
Iodate	Iodized table salt
Chromate	Orange juice, potato products
Selenite and selenate	Vegetables, cereals, orange juice
Arsenite and arsenate	Food supplements, cereals, vegetables
Cyanide	Fruits and fruit juices

An alternative approach has been developed based on a weakly acidic cation exchange column, used with a dilute tartaric acid/crown ether eluent, which resulted in the ability to separate both inorganic anions (chloride, nitrate, and sulfate) and inorganic cations (sodium, potassium, ammonium, calcium, and magnesium) simultaneously in real rainwater samples. The anions were retained through an ion exclusion mechanism and the cations were retained through a simple cation exchange mechanism. Sensitive detection was achieved using direct conductivity.

## Industrial Applications

### Food and Beverages

As with most chromatographic methods applied to solid samples, sample digestion and analyte extraction methods are all important. IC is finding application in the analysis of foodstuffs following sample preparation using such techniques as microwave digestion, supercritical fluid extraction, accelerated solvent extraction, and pyrohydrolysis.

**Inorganic anions** The predominant anionic species determined in foodstuffs are once again the nitrogen-, sulfur-, and phosphorus-containing species, as well as the halide ions. Table 2 lists some inorganic anions and some of the foodstuffs that have been analyzed for these anions using IC.

Some of the more important applications in food analysis include nitrates and nitrites in baby food products, excess of which can lead to induce methemoglobinemia (blue baby syndrome), and the monitoring of sulfite, which is added to many foodstuffs as a preservative and to bleach food starches, and is only recently being linked to serious health effects. Also, residual bromate can be monitored in bakery products from the continuing use of bromate salts as dough conditioners.

**Organic acids** In beverages such as wines, beers, and fruit juices, IC has also been widely applied in the determination of various organic acids, although in many cases ion exclusion chromatography is often used in preference to anion exchange.

**Sugars** In the brewing industry, IC is used for the determination and monitoring of fermentable sugars, such as glucose, fructose, isomaltose, sucrose, maltose, maltotriose, and numerous others. For sensitive detection, pulsed amperometric detection is often preferred.

**Inorganic and organic cations** In the analysis of foodstuff extracts and digests, IC has been predominantly applied to the determination of alkali and alkaline earth metal ions and, to a lesser extent, selected organic amines. Alkali and alkaline earth metal ions are naturally present in most foodstuffs, although accurate monitoring is still necessary to evaluate nutritional values, e.g., the sodium or calcium content of foodstuffs. As mentioned previously, ammonium content can also be determined using IC simultaneously with alkali metal ions, and is often used as an indicator of food quality.

Transition and heavy metal ions have also been determined in foodstuffs using IC, particularly seafood, where heavy metal contamination with metals such as cadmium and lead is often a problem. After separation using cation exchange or ion interaction chromatography, sensitive detection is generally achieved using postcolumn reaction detection with a suitable color-forming ligand, such as 4-(2-pyridylazo) resorcinol (PAR). Other metals such as zinc, copper, iron, cobalt, nickel, chromium, and manganese can also be detected in this way.

In an interesting recent application, the determination of acrylamide in foodstuffs has been shown using accelerated solvent extraction followed by IC

with either UV or MS detection. Extracted samples can be analyzed directly using IC, with limits of determination of 50 ng per g acrylamide in foodstuffs possible using MS detection with single ion monitoring (SIM) at  $m/z$  72.

### Pharmaceuticals

In the preparation of pharmaceutical products, the purity of reagents is of utmost importance. IC is often used for trace anion and cation determinations in starting materials, the simplest of which is reagent grade water. However, actual pharmaceutical preparations are often complex mixtures and IC provides alternative column selectivity to standard reversed-phase HPLC, and is often more suited for the separation of very polar organics commonly used in pharmaceutical products. Table 3 lists some of the organic species that have been determined using IC. Often for such analytes UV absorbance is the preferred mode of detection, provided the analyte contains a suitable chromophore.

### Nuclear and Fossil Fuel Power Generation Industry

As with the pharmaceutical products above, there is also great concern over contamination problems in the power generation industry. IC is extensively used to determine ultratrace concentrations of ionic species in process waters, coolant waters, wastewater, and other waste materials. The popularity of the technique is due to IC being one of a few analytical technologies that is able to measure ng per g concentrations of potentially corrosive ions such as

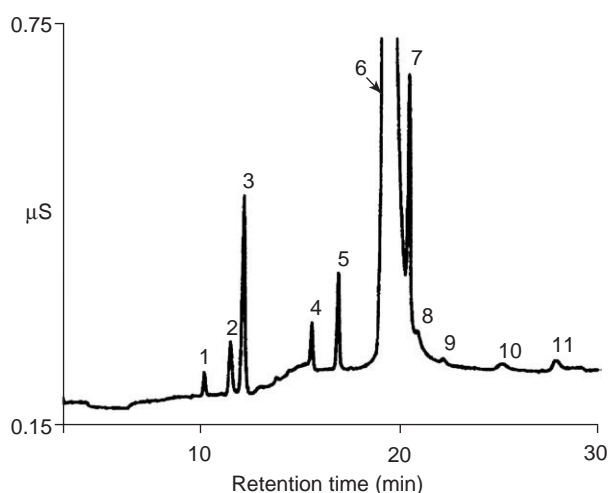
sodium, chloride, and sulfate on a near-real-time basis, and can also differentiate among different oxidation states of ions such as nitrate, nitrite, sulfate, and thiosulfate. These oxidation states respond to the general oxidizing tendency of coolant waters, which is an important variable in controlling localized corrosion. Some impressive recent applications of IC within this industry include: the determination of mono-, di-, and tributyl phosphates, the latter of which is an important compound used in the area of nuclear fuel reprocessing; the determination of long-lived artificial radionuclides resulting from fission reactions using IC coupled to ICP-MS detection; the use of IC to determine amines in amine-dosed feedwater used for erosion control in boiler tubes; and the application of IC to the determination of transition metal cations in the primary coolants of light water reactors. Figure 3 shows an ion chromatograph of a condensate discharge water sample from a fossil fuel power station containing trace amounts of inorganic anions and organic acids.

### Semiconductor Industry

The semiconductor industry utilizes IC to monitor and identify the possible contamination of products during manufacture. Much of the application of IC is to trace inorganic anion analysis and, to a lesser extent, trace alkali and alkaline earth cation determinations. The sample matrices that are monitored using IC in the semiconductor industry include the large variety of chemicals used for cleaning,

**Table 3** Examples of pharmaceutical preparations to which IC has been applied

<i>Analytes</i>	<i>Sample matrix</i>	<i>Column</i>	<i>Detection method</i>
Citrate	Liquid and tablet formulations	Anion exchange	Indirect UV absorbance
Saccharin aspartane, acesulfame-k, benzoate, sorbate, caffeine, theobromine, theophylline	Tablet preparations	Anion exchange	Tunable UV absorbance
Trifluoroacetate	Cell-based products	Anion exchange	Suppressed conductivity
Alenalol, metaprolol, alprenolol, oxprenolol, acebutolol, propanolol	$\beta$ -Blocker tablet preparations	Anion exchange	Tunable UV absorbance
Paracetamol, salicylate	Tablet preparations	Anion exchange	UV absorbance
Alkyl sulfonic acids	Sulfonated sugar preparations	Anion exchange	Suppressed conductivity
Alendronate	Dosage formulations	Anion exchange	Electrospray mass spectrometry
Catecholamines (norepinephrine, epinephrine, dopamine)	Injection formulations	Cation exchange	Nonsuppressed conductivity
Tetracyclines	Antibiotic preparations	Cation exchange	UV absorbance
Amylamine, <i>t</i> -butylamine	Hypocholesterolemic agent preparations	Cation exchange	Nonsuppressed conductivity



**Figure 3** Ion chromatogram of a condensate discharge water sample. Peak identification: 1 = fluoride, 2 = acetate, 3 = formate, 4 = chloride, 5 = nitrite, 6 = carbonate, 7 = sulfate, 8 = unknown, 9 = nitrate, 10 = unknown, 11 = phosphate. (Reprinted with permission from Lu Z, Liu Y, Barreto V, *et al.* (2002) Determination of anions at trace levels in power plant water samples by ion chromatography with electrolytic eluent generation and suppression. *Journal of Chromatography A* 956: 129–138; © Elsevier.)

washing, polishing, and treating surfaces, such as reagent water, solvents, strong acids and bases, and oxidizing agents. Process gases and clean room gases, and filtered air are also monitored.

## Biological Applications

### Analysis of Blood, Plasma, and Serum

The ability to determine quantitatively certain ionic analytes in blood, plasma, and serum samples can be of substantial benefit to those attempting the diagnosis of certain diseases, particularly where concentrations of these analytes are known to be directly related to specific physiological disorders.

**Anions** IC has established itself as the method of choice for the determination of common anions such as chloride, sulfate, and phosphate in blood and serum samples. For example, a common clinical application of IC is the study of sulfa drug metabolism through the monitoring of blood sulfate levels. Samples are generally pretreated using ultrafiltration or acidification and centrifugation, or both. Both IC and ion exclusion chromatography have been applied to the determination of bicarbonate in blood plasma. Ion exclusion has also been extensively applied to the determination of certain organic acids in blood plasma, such as pyruvate and lactate, the latter of which is used to

help in the diagnosis of lactic acidosis in diabetic patients.

**Cations** There have been many applications of IC to the determination of sodium, potassium, and ammonium in blood serum. In most cases, serum samples were simply ultrafiltered and injected directly. In other cases only dilution was necessary. Total and free concentrations of calcium and magnesium in blood serum can also be determined using IC. Total concentrations can be determined using sample acidification followed by centrifugation. Free concentrations of the cations can be determined after passage of the untreated sample through a cation exchange solid-phase extraction (SPE) cartridge to trap the cations and isolate them from those ions bound to serum proteins.

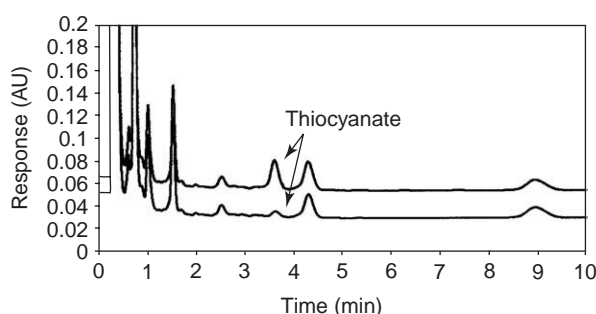
### Analysis of Urine

As with blood, urine analysis is used in clinical studies, and the relative concentrations of various ionic species are of great importance in both disease diagnostic and drug metabolism studies. For example, IC is used to determine urine oxalate concentrations. Urinary oxalate levels are an important parameter in urolithiasis research (kidney stones). Other anions that can be determined in urine using IC include phosphate, sulfate, bromide, citrate, nitrate, nitrite, and thiosulfate. As with blood and serum samples, both ultrafiltration and centrifugation are often used as sample cleanup steps.

Interesting applications include the use of IC coupled with ICP-MS detection for the determination of anionic arsenic species in urine resulting from occupational and dietary exposure. Species include urinary arsenate, arsenite, dimethylarsinic acid, and methylarsonic acid, and are separated using a hydrophilic anion exchange resin with a weak acid eluent. Studies have also been carried out using IC and ion interaction liquid chromatography to determine human urinary thiocyanate concentrations and relate concentrations found to levels of smoking. Thiocyanate is the main metabolic product of cyanide inhaled with cigarette smoke. **Figure 4** shows overlaid ion chromatograms of a smoker's urine sample and the same sample spiked with thiocyanate. The chromatograms shown were obtained using ion interaction chromatography, utilizing a short reversed-phase column and a tetrabutylammonium chloride and methanol eluent. Detection was carried out using direct UV absorbance at 230 nm.

### Analysis of Saliva and Sweat

Finally, IC methods have been applied to the determination of both inorganic anions and alkali and



**Figure 4** Analysis of urine for thiocyanate using ion interaction liquid chromatography. Lower trace: sample (heavy smoker) diluted 1:20. Upper trace (heavy smoker) diluted 1:20 and spiked with thiocyanate.

alkaline earth cations in saliva. The sample matrix is relatively simple compared to other biological fluids and can be analyzed directly or simply diluted prior to injection.

IC methods have also been developed for the determination of sweat samples for ionic analytes. Methods looking at sulfate levels and also concentrations of sodium and potassium in sweat samples have been developed, with relative levels of the latter metal ions being useful indicators of several important diseases, one of which is cystic fibrosis.

See also: **Ion Exchange:** Overview; Principles; Ion Chromatography Instrumentation; Chelation Ion Chromatography; Isolation of Biopolymers; Isotope Separation.

## Further Reading

- Betti M (1997) Use of ion chromatography for the determination of fission products and actinides in nuclear applications. *Journal of Chromatography A* 789: 369–379.
- Buldini PL, Cavalli S, and Trifiro A (1997) State-of-the-art ion chromatographic determination of inorganic ions in food. *Journal of Chromatography A* 789: 529–548.
- Haddad PR and Jackson PE (1990) *Ion Chromatography. Principles and Applications*. Amsterdam: Elsevier.
- Lopez-Ruiz B (2000) Advances in the determination of inorganic anions by ion chromatography. *Journal of Chromatography A* 881: 607–627.
- Singh RP, Smesko SA, and Abbas NM (1997) Ion chromatographic characterisation of toxic solutions: analysis and ion chemistry of biological liquids. *Journal of Chromatography A* 774: 21–35.
- Small H (1990) *Ion Chromatography*. New York: Plenum.
- Smith RE (1987) *Ion Chromatography Applications*. Boca Raton, FL: CRC Press.
- Vanatta LE (2001) Application of ion chromatography in the semiconductor industry. *Trends in Analytical Chemistry* 20: 336–345.

## Chelation Ion Chromatography

**P Jones**, University of Plymouth, Devon, UK

**P N Nesterenko**, Lomonosov Moscow State University, Moscow, Russia

© 2005, Elsevier Ltd. All Rights Reserved.

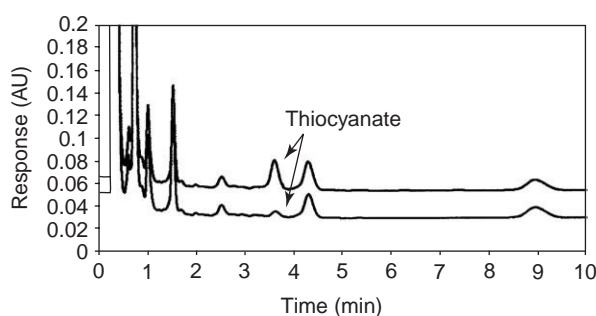
## Introduction

It is now over 60 years since the development of column ion-exchange chromatography using polymeric resins. Interestingly, little has changed since then in terms of the basic separation processes for inorganic anions and cations. The renaissance in the 1970s, where ion-exchange was the principal process in a group of techniques now known as ion chromatography (IC), was mainly associated with the improvement in efficiency of stationary phases and detection systems rather than new types of ion-exchange groups or elution systems. In essence, the IC separation of metal ions principally involves the use of eluents containing complexing organic acids,

such as tartaric, citric, oxalic, combined with high efficiency stationary phases. The complexing strength of acid and concentration chosen depended on whether the ion-exchange substrate was cationic, anionic, or mixed. Although poly(styrene-divinylbenzene) (PS-DVB) based resins are still the most common substrate, silica-based materials are increasingly being used.

There are a number of problems associated with present IC methods involving high efficiency ion-exchange separations of metal ions, the two principal ones being sensitivity to ionic strength and limited selectivity. The influence of ionic strength is particularly serious as a relatively large salt concentration in the sample can drastically affect the chromatography, in many cases making it impossible to resolve the analytes. Limited selectivity is also a restriction, as once the substrate is chosen for conventional ion-exchange separations there are only a small number of cases where changes in eluent composition can significantly alter separation order.





**Figure 4** Analysis of urine for thiocyanate using ion interaction liquid chromatography. Lower trace: sample (heavy smoker) diluted 1:20. Upper trace (heavy smoker) diluted 1:20 and spiked with thiocyanate.

alkaline earth cations in saliva. The sample matrix is relatively simple compared to other biological fluids and can be analyzed directly or simply diluted prior to injection.

IC methods have also been developed for the determination of sweat samples for ionic analytes. Methods looking at sulfate levels and also concentrations of sodium and potassium in sweat samples have been developed, with relative levels of the latter metal ions being useful indicators of several important diseases, one of which is cystic fibrosis.

See also: **Ion Exchange:** Overview; Principles; Ion Chromatography Instrumentation; Chelation Ion Chromatography; Isolation of Biopolymers; Isotope Separation.

## Further Reading

- Betti M (1997) Use of ion chromatography for the determination of fission products and actinides in nuclear applications. *Journal of Chromatography A* 789: 369–379.
- Buldini PL, Cavalli S, and Trifiro A (1997) State-of-the-art ion chromatographic determination of inorganic ions in food. *Journal of Chromatography A* 789: 529–548.
- Haddad PR and Jackson PE (1990) *Ion Chromatography. Principles and Applications*. Amsterdam: Elsevier.
- Lopez-Ruiz B (2000) Advances in the determination of inorganic anions by ion chromatography. *Journal of Chromatography A* 881: 607–627.
- Singh RP, Smesko SA, and Abbas NM (1997) Ion chromatographic characterisation of toxic solutions: analysis and ion chemistry of biological liquids. *Journal of Chromatography A* 774: 21–35.
- Small H (1990) *Ion Chromatography*. New York: Plenum.
- Smith RE (1987) *Ion Chromatography Applications*. Boca Raton, FL: CRC Press.
- Vanatta LE (2001) Application of ion chromatography in the semiconductor industry. *Trends in Analytical Chemistry* 20: 336–345.

## Chelation Ion Chromatography

**P Jones**, University of Plymouth, Devon, UK

**P N Nesterenko**, Lomonosov Moscow State University, Moscow, Russia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

It is now over 60 years since the development of column ion-exchange chromatography using polymeric resins. Interestingly, little has changed since then in terms of the basic separation processes for inorganic anions and cations. The renaissance in the 1970s, where ion-exchange was the principal process in a group of techniques now known as ion chromatography (IC), was mainly associated with the improvement in efficiency of stationary phases and detection systems rather than new types of ion-exchange groups or elution systems. In essence, the IC separation of metal ions principally involves the use of eluents containing complexing organic acids,

such as tartaric, citric, oxalic, combined with high efficiency stationary phases. The complexing strength of acid and concentration chosen depended on whether the ion-exchange substrate was cationic, anionic, or mixed. Although poly(styrene-divinylbenzene) (PS-DVB) based resins are still the most common substrate, silica-based materials are increasingly being used.

There are a number of problems associated with present IC methods involving high efficiency ion-exchange separations of metal ions, the two principal ones being sensitivity to ionic strength and limited selectivity. The influence of ionic strength is particularly serious as a relatively large salt concentration in the sample can drastically affect the chromatography, in many cases making it impossible to resolve the analytes. Limited selectivity is also a restriction, as once the substrate is chosen for conventional ion-exchange separations there are only a small number of cases where changes in eluent composition can significantly alter separation order.

There is an alternative to ion-exchange, known generally as chelation ion-exchange, giving greater control over selectivity and relatively unaffected by changes in ionic strength. Although related, chelation ion-exchange differs fundamentally from ion-exchange in that it involves the formation of coordinate bonds between metal ions and a complexing ligand immobilized on the surface of the substrate. Thus, the sorption process in chelation ion-exchange depends mainly upon the values of the stability constants of the metal complexes formed on the stationary phase. Inevitably, because ion-exchange and chelation ion-exchange arise from the same functional group, they can occur together. Nevertheless, conditions can be chosen where chelation ion-exchange is the dominant mechanism controlling separation, as explained in the next section.

It can be seen that the greater versatility of chelation ion-exchange arises from the wide range of complexing ligands available, where selectivity can be altered to suit the trace metal composition of a particular interest and sample type. Chelation ion-exchange is not a recently exploited technique, being used for nearly as long as ion-exchange itself. However, until recently, chelation ion-exchange columns have principally been used for matrix isolation and preconcentration of suites of metals rather than chromatographic separation. The isolation of trace metals from seawater using Chelex 100 (iminodiacetic acid functionalized PS-DVB) is a classic example. Later, chelating ion-exchangers such as 8-hydroxyquinolinol attached to silica (Table 1) have been extensively used in flow injection analysis for determination of metals.

In the last 10 years or so, chelation ion-exchange columns have become available for use as precolumns in IC systems where matrix isolation and preconcentration occur online before the suite of metals is swept onto an ion-exchange column for separation (see later). This approach became known as chelation ion chromatography (CIC). Another use of chelation ion-exchange columns, the main focus of this article, is for the high-efficiency separation of metal ions on small particle size substrates. It is a relatively new approach and shows great potential for the IC determination of metals in a wide variety of complex samples, including those with very high ionic strengths. Although the term CIC can be used to describe both uses of chelation ion-exchange substrates, it should be noted that some workers have preferred to use the term high-performance chelation ion chromatography (HPCIC) to differentiate the high-efficiency separation technique from CIC methods where chelation ion-exchange is only

used for preconcentration or matrix isolation purposes.

The principal purpose of this article is to describe the latest work involving high-efficiency chelation ion-exchange columns for determining trace metals in a wide variety of complex matrices. The basic principles relating to the sorption of metal ions will be described first, followed by illustrations of the different separation modes available, and, finally, key examples will be given of the wide range of samples analyzed by the technique.

## Chelation Ion-Exchange Equilibria

The efficient separation of metal cations can be achieved with three main methods:

1. using conventional cation-exchange based on electrostatic interactions;
2. by preconcentration of metal cations on a short column packed with chelating substrate with further separation by cation-exchange ion chromatography;
3. using the ability of chelating ion-exchangers to achieve analytical separations of metal ions with a single column.

The last two methods can both be described as CIC as discussed earlier. However, it is only with approach (3), the principal concern of this article, is it important to fully understand the factors affecting efficiency and selectivity to obtain good chromatography. Therefore, the basic principles described below will focus on these aspects.

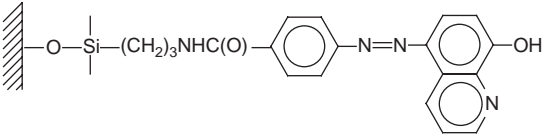
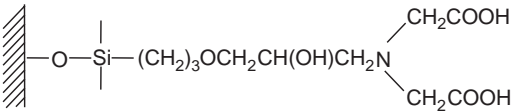
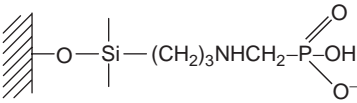
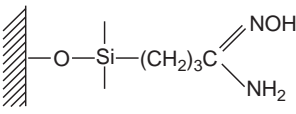
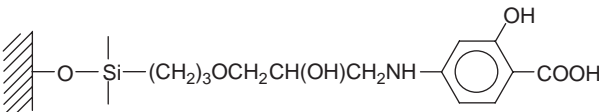
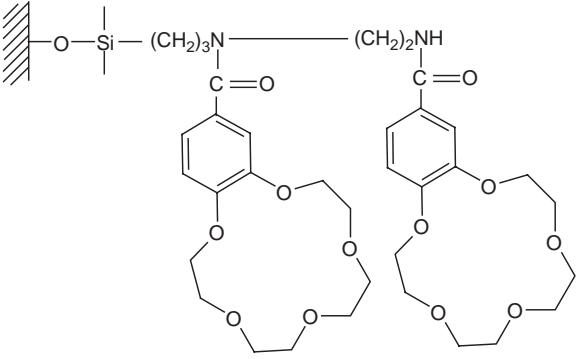
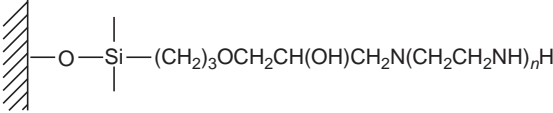
### Distribution Ratio

Pure sorption mechanisms are very rare in chromatography and chelation ion-exchange is no exception. As the chelating groups are usually charged, the chromatographic retention of alkaline-earth, transition, and heavy metal ions on such a column, in a noncomplexing eluent, occurs through the combination of electrostatic and coordination interactions between cations and chelating functional groups on the substrate. The distribution ratio  $D_M$  of metal cation  $M^{y+}$  between chelating ion-exchanger and the contacting mobile phase can be expressed by the following equation:

$$D_M = ([M_R^{y+}]^I + [M_R^{y+}]^C) / [M_{aq}^{y+}] \quad [1]$$

where  $[M_R^{y+}]^I$  and  $[M_R^{y+}]^C$  are equilibrium concentrations of cation retained by the stationary phase due to electrostatic interactions (conventional ion-exchange) and chelation ion-exchange, respectively.

**Table 1** Selectivity of chelating ion-exchangers for HPCIC

Functional groups	Selectivity
<b>(A) Chemically bonded</b> <b>8-Hydroxyquinolinol</b> 	$\text{Ni}^{2+} > \text{Co}^{2+} > \text{Zn}^{2+} > \text{Pb}^{2+} > \text{Cd}^{2+} > \text{Mn}^{2+}$ $\text{Yb}^{3+} > \text{Gd}^{3+} > \text{La}^{3+}$
<b>Iminodiacetic acid</b> 	(a) $\text{Mg}^{2+} > \text{Ca}^{2+} > \text{Sr}^{2+} > \text{Ba}^{2+}$ (b) $\text{Fe}^{3+} > \text{Cu}^{2+} > \text{In}^{3+} > \text{UO}_2^{2+} > \text{Pb}^{2+} > \text{Ni}^{2+} > \text{Zn}^{2+} > \text{Cd}^{2+} > \text{Co}^{2+} > \text{Be}^{2+} > \text{Fe}^{2+} > \text{Mn}^{2+}$ (c) $\text{Lu}^{3+} > \text{Yb}^{3+} > \text{Tm}^{3+} > \text{Er}^{3+} > \text{Ho}^{3+} > \text{Tb}^{3+} > \text{Eu}^{3+} > \text{Gd}^{3+} > \text{Dy}^{3+} > \text{Sm}^{3+} > \text{Y}^{3+} > \text{Nd}^{3+} > \text{Pr}^{3+} > \text{Ce}^{3+} > \text{La}^{3+}$
<b>Aminophosphonic acid</b> 	(a) $\text{Mg}^{2+} > \text{Ca}^{2+} > \text{Sr}^{2+} > \text{Ba}^{2+}$ (b) $\text{Mn}^{2+} > \text{Cd}^{2+} > \text{Pb}^{2+} > \text{Cu}^{2+} > \text{Zn}^{2+} > \text{Co}^{2+} > \text{Ni}^{2+}$ (c) $\text{Lu}^{3+} > \text{La}^{3+} > \text{Be}^{2+} > \text{Al}^{3+}$
<b>Amidoxime</b> 	$\text{UO}_2^{2+} > \text{Cu}^{2+} > \text{Cd}^{2+} > \text{Co}^{2+}$
<b>p-Aminosulfosalicylic acid</b> 	$\text{Fe}^{3+} > \text{Cu}^{2+} > \text{Zn}^{2+}$
<b>Bis(benzo-15-crown-5)</b> 	$\text{K}^+ > \text{Rb}^+ > \text{Cs}^+ > \text{Na}^+ > \text{Li}^+$ (as $\text{Cl}^-$ salts) $\text{Ba}^{2+} > \text{Sr}^{2+} > \text{Ca}^{2+} > \text{Mg}^{2+}$ (as $\text{Cl}^-$ salts) $\text{I}^- > \text{Br}^- > \text{Cl}^-$ (as $\text{K}^+$ salts)
<b>Poly(amines)</b>  <p style="text-align: center;"><math>n = 0-4</math></p>	$\text{Cu}^{2+} > \text{Co}^{2+} > \text{Zn}^{2+} > \text{Cd}^{2+} > \text{Fe}^{2+} > \text{Mg}^{2+}$

Continued

Table 1 Continued

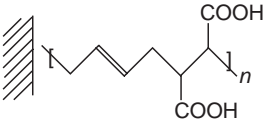
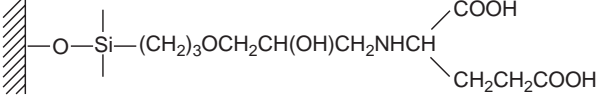
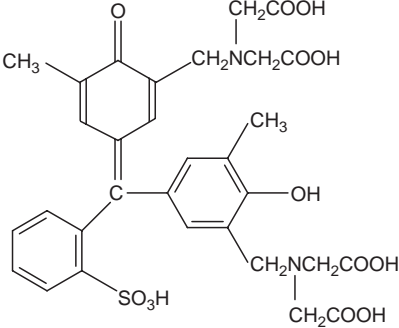
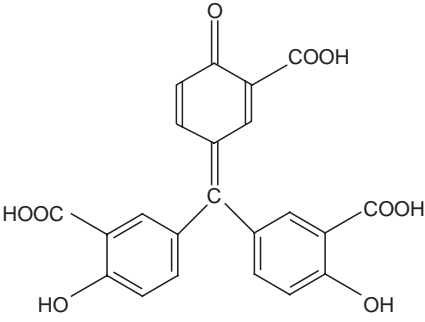
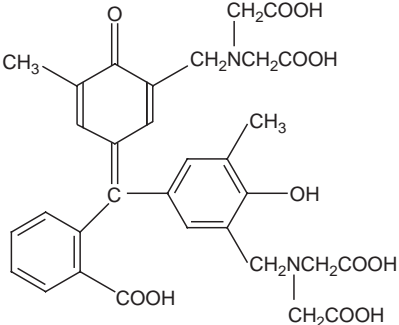
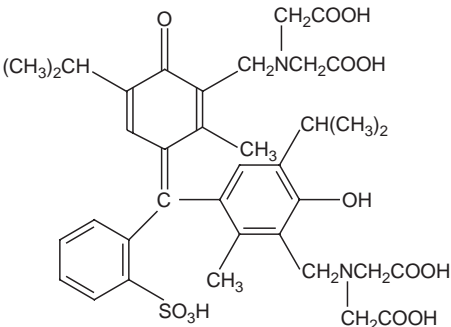
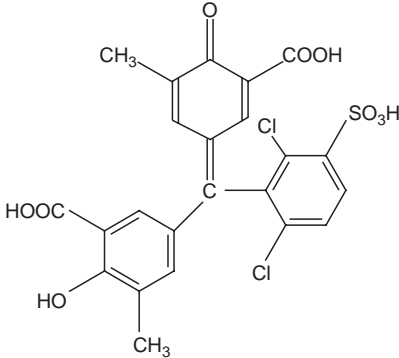
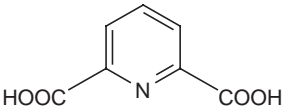
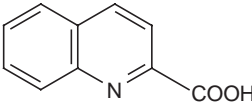
Functional groups	Selectivity
Poly(butadienemaleic acid) PBDMA	$\text{Cu}^{2+} > \text{In}^{3+} > \text{UO}_2^{2+} > \text{Pb}^{2+} > \text{Ni}^{2+} > \text{Co}^{2+} > \text{Zn}^{2+} > \text{Cd}^{2+} > \text{Mn}^{2+} > \text{Mg}^{2+}$
	
Aspartic acid	$\text{Cu}^{2+} > \text{Pb}^{2+} > \text{Zn}^{2+} > \text{Ni}^{2+} > \text{Mn}^{2+} > \text{Cd}^{2+} > \text{Co}^{2+} > \text{Mg}^{2+} > \text{Ca}^{2+} > \text{Ba}^{2+}$
	
(B) Precoated or impregnated substrates	
Xylenol Orange	<p>(a) <math>\text{Mg}^{2+} &gt; \text{Ca}^{2+} &gt; \text{Sr}^{2+} &gt; \text{Ba}^{2+}</math></p> <p>(b) <math>\text{Cu}^{2+} &gt; \text{Ni}^{2+} &gt; \text{Zn}^{2+} &gt; \text{Cd}^{2+} &gt; \text{Mn}^{2+}</math></p>
	
Aurin tricarboxylic acid	<p>(a) <math>\text{In}^{3+} &gt; \text{Ga}^{3+} &gt; \text{Al}^{3+}</math></p> <p>(b) <math>\text{Cu}^{2+} &gt; \text{Pb}^{2+} &gt; \text{Cd}^{2+} &gt; \text{Ni}^{2+} &gt; \text{Zn}^{2+} &gt; \text{Co}^{2+} &gt; \text{Mn}^{2+}</math></p>
	
$\alpha$ -Cresolphthalein complexone	$\text{Ca}^{2+} > \text{Mg}^{2+} > \text{Sr}^{2+} > \text{Ba}^{2+}$
	

Table 1 Continued

Functional groups	Selectivity
<p>Methyl thymol blue</p> 	$\text{Ca}^{2+} > \text{Mg}^{2+} > \text{Sr}^{2+} > \text{Ba}^{2+}$
<p>Chrome Azurol S</p> 	$\text{Fe}^{3+} > \text{Al}^{3+}$
(C) Dynamically coated	
Methyl thymol blue (see structure in section 1b)	<p>(a) <math>\text{Cu}^{2+} &gt; \text{UO}_2^{2+} &gt; \text{Mg}^{2+} &gt; \text{Ca}^{2+}</math></p> <p>(b) <math>\text{Lu}^{3+} &gt; \text{Tm}^{3+} &gt; \text{Tb}^{3+} &gt; \text{Sm}^{3+} &gt; \text{Nd}^{3+} &gt; \text{Ce}^{3+} &gt; \text{La}^{3+} &gt; \text{Cu}^{2+} &gt; \text{Pb}^{2+} &gt; \text{Cd}^{2+} &gt; \text{Zn}^{2+} &gt; \text{Co}^{2+} &gt; \text{Mn}^{2+} &gt; \text{Mg}^{2+}</math></p>
Dipicolinic acid	<p>(a) <math>\text{Cd}^{2+} &gt; \text{Pb}^{2+} &gt; \text{Zn}^{2+} &gt; \text{Ni}^{2+} &gt; \text{Cu}^{2+} &gt; \text{Mn}^{2+}</math></p> <p>(b) <math>\text{Zr(IV)} &gt; \text{Hf(IV)} &gt; \text{U(VI)} &gt; \text{Bi(III)} &gt; \text{V(V)} &gt; \text{Th(IV)} &gt; \text{Fe(III)}</math></p>
<p>4-Chlorodipicolinic acid</p> 	<p>(a) <math>\text{Cd}^{2+} &gt; \text{Pb}^{2+} &gt; \text{Cu}^{2+} &gt; \text{Zn}^{2+} &gt; \text{Ni}^{2+} &gt; \text{Co}^{2+} &gt; \text{Mn}^{2+}</math></p> <p>(b) <math>\text{UO}_2^{2+} &gt; \text{Fe}^{3+} &gt; \text{Lu}^{3+} &gt; \text{La}^{3+} &gt; \text{Al}^{3+}</math></p>
<p>Quinaldic acid</p> 	<p>(a) <math>\text{Cu}^{2+} &gt; \text{Ni}^{2+} &gt; \text{Co}^{2+} &gt; \text{Zn}^{2+} &gt; \text{Cd}^{2+} &gt; \text{Pb}^{2+} &gt; \text{Mn}^{2+}</math></p> <p>(b) <math>\text{Lu}^{3+} &gt; \text{La}^{3+}</math></p>



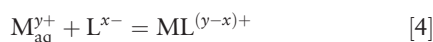
$[M_{aq}^{y+}]$  is the concentration of cation in the mobile phase. The retention factor  $k$  can be expressed as

$$k = (t - t_0)/t_0 = D_M \varphi \quad [2]$$

where  $t$  is retention time,  $t_0$  the void time, and  $\varphi$  is a characteristic constant for a given chromatographic column expressed as the ratio of the volumes of mobile  $V_{aq}$  and stationary  $V_R$  phases:

$$\varphi = V_{aq}/V_R \quad [3]$$

As suitable chromatographic separation can be achieved at values of  $k < 10$ –15 and the formation of complexes at the surface with more than one ligand is for simplification assumed not to take place due to thermodynamic and steric restrictions; then,



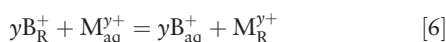
Thus,  $[M_R^{(y-x)+}]^C$  can be expressed as follows:

$$[M_R^{(y-x)+}]^C = \beta_1 [M_{aq}^{y+}] [L^{x-}] \quad [5]$$

where  $[L^{x-}]$  is the concentration of functional groups and  $\beta_1$  is the stability constant of complex  $ML^{(y-x)+}$  formed at the surface in accordance with (4).

### Selectivity Ratio

The ion-exchange process of cation  $M^{y+}$  and alkali metal cation  $B^+$  as competing cation at constant pH of the eluent on the chelating ion-exchanger can be expressed as follows:



where the subscript 'aq' denotes the mobile phase and 'R' denotes the stationary phase. The following selectivity ratio is given by

$$K_M^B = \frac{[M_R^{y+}][B_{aq}^+]^y}{[M_{aq}^{y+}][B_R^+]^y} \quad [7]$$

Taking into account the two possible types of interactions in accordance with [1], the retention factor  $k$  of metal cation  $M^{y+}$  can be expressed as

$$k = \left( K_M^B \frac{[B_R^+]^y}{[B_{aq}^+]^y} + \beta_1 [L^{x-}] \right) \varphi \quad [8]$$

The first member of the sum in brackets expresses the impact of conventional ion-exchange interactions on retention, and the second member expresses the impact of chelation on retention. Several conclusions can be formulated from eqn [8].

Assuming little change in  $\beta_1$  with high ionic strength ( $>0.1$ ), the retention of a metal ion due to

chelation should not depend upon the concentration of alkali metal cations  $[B_{aq}^+]$ , but solely on the concentration of functional groups in the chelating ion-exchanger. In practice, it means electrostatic interactions need to be suppressed for chelation to be the dominant sorption mechanism.

The retention of  $M$  due to conventional ion-exchange can be suppressed by increase of  $[B_{aq}^+]$ .

Thus, in the case of domination by chelation ( $\beta_1 [L] \gg K_M^B ([B_R^+]^y / [B_{aq}^+]^y)$  at high concentration level of  $B_{aq}^+$  in the eluent) the separation selectivity  $\alpha$  is defined by ratio of corresponding stability constants of metal chelates formed at the surface:

$$\alpha = k_2/k_1 = \beta_1^{M2}/\beta_1^{M1} \quad [9]$$

Strictly, it should be the ratio of conditional stability constants, but for a given ligand with no hydrolysis of the metal ions, the ratio is the same as the thermodynamic stability constants.

### Temperature Effects

As well as varying the ionic strength of the eluent, the separation selectivity  $K_M^B$  in CIC can be changed by variation of column temperature. Sorption heat relates the retention factor  $k$  with the column temperature by the van't Hoff equation:

$$\ln k = -\Delta H/RT + \Delta S/R + \ln \varphi \quad [10]$$

where  $\Delta H$  and  $\Delta S$  are sorption enthalpy and entropy, respectively, and  $\varphi$  is the phase volume ratio. The temperature effects are exothermic (negative values of  $\Delta H$ ) in the case of conventional ion-exchange. However, the heat values do not exceed 8–13 kJ mol<sup>-1</sup> and therefore have little influence on retention. In contrast, for stationary phases where metal ion complexation is present, the sorption process will normally exhibit much higher values of  $\Delta H$  in both exothermic and endothermic conditions. Thus, temperature effects will be significant and some regulation of selectivity is possible.

### Factors Affecting the Efficiency of Chelation Ion-Exchange

It is very important to understand the factors controlling the kinetics of the sorption process as they will have a major bearing on separation efficiency. There are three factors affecting the kinetics of chelating ion-exchange:

1. transport of metal ion from the mobile phase to the surface of chelating ion-exchanger;

2. diffusion of the ions inside of the stationary phase; and
3. kinetics of chelation of metal ion with functional groups.

The first two factors are similar to those occurring in simple conventional ion-exchange so they will not be discussed further.

The resulting kinetics of chelation depends upon the chemical nature of the functional groups and the structure of the chelating ion-exchanger, with the former being the most crucial for good separations.

### Chemical Nature of Functional Groups

**Charge on the functional groups** It was empirically established that better kinetics of CIC arise when negatively charged functional groups are present. Obviously, negatively charged groups at the surface, attracting oppositely charged cations from the mobile phase, will provide better transport of metal ions from the mobile phase than neutral or positively charged chelating surfaces.

**Dentation of the ligand** It would be expected that the higher the denticity of the ligand, the slower the kinetics as more groups need to be dissociated and more coordinate bonds formed for each metal exchange reaction. In practice, polydentate ligands are usually necessary to achieve adequate retention and selectivity. Fortunately, for most ligands with a denticity not greater than 3, the slower kinetic effect in most cases appears to be relatively minor, as long as the eluent conditions are chosen to ensure small conditional constants.

**Magnitude of the conditional constants** Of all the ligand properties affecting chelation ion-exchange kinetics, the value of the metal/ligand conditional constant has the largest influence. In simple terms, the magnitude of a stability or conditional constant is governed by the ratio of the velocity of the forward reaction (association) to the velocity of the back reaction (dissociation). Except for a small number of metal ion/ligating atom combinations (see later), the velocity of the forward reaction is very fast. Thus, the kinetics of chelation ion-exchange will be controlled by the value of the velocity of the back reaction (dissociation), which will increase as the conditional constant decreases. However, the retention factor  $k$  is directly proportional to the value of the conditional constant (eqn [8]). Therefore, to achieve optimum efficiency of separation, the conditional constants should be as low as possible, but not so low as to cause serious overlap of peaks close to

the solvent front. As most ligands arise from dissociation of an acid group, change in pH is the simplest way of changing the conditional constant and hence  $k$ . It should be noted that this property is very important for the chromatographic separation of metal cations, but not so critical when used for preconcentration or matrix isolation of groups of metal ions.

**Type of ligating atom** Nitrogen and oxygen are the most common ligating atoms in the chelating groups used in CIC. Sulfur-containing ones have been investigated, but are inherently unstable due to air oxidation. Nitrogen-ligating atoms can give slower kinetics than oxygen atoms with certain common metals, nickel being most notable. Even N, O, O chelators like iminodiacetic acid (IDA) give broad nickel peaks unless the conditional constant is very low and the nickel is close to the solvent front.

### Structural Properties of the Chelating Exchanger

**Conformational mobility of the bonded ligand** Chelating ion-exchangers having functional groups chemically attached as a monolayer to the surface through flexible long linkers will provide better selectivity and separation efficiency as they are more easily accessible to metal ions. On the other hand, chelating substrates formed by immobilizing ligands in micropores will be more sterically hidden, subsequently decreasing the kinetics of metal ion association.

**The distribution density of functional groups at the surface** A very high concentration (capacity) of functional groups can lead to nonuniform complexation at the surface of the chelating ion-exchanger. Perhaps a more important consideration is that close proximity of ligands allows the theoretical possibility of formation of  $ML_2$  and even  $ML_3$  complexes with much less favorable kinetics of dissociation, causing poor (broad and tailing) peak shapes in CIC. It should be noted that for some chelating substrates such as silica bonded IDA, due to thermodynamic considerations,  $ML_2$  and  $ML_3$  complexes are not formed at the typical low pHs used for separations.

It is interesting to note that, depending on the distribution and density of the functional groups, weak cation-exchangers with carboxylic and phosphonic acid functional groups could serve as chelating ion-exchangers as two or more groups in close proximity may be able to form chelate-related structures with metal ions, though this may have a bearing on peak broadening as discussed above.

**Pore size** The structure of the substrate can also affect the stoichiometry of complexes at the surface. In narrow pores the coordination of two attached ligands is much more probable.

## Main Types of Chelation Ion-Exchangers

Rigidity and mechanical stability of the substrate under conditions of high-performance liquid chromatography are important and microspherical particles of macroporous highly crosslinked PS-DVB or poly(metacrylate) and silica are well suited for CIC. There are three main methods of producing a chelating ion-exchange surface.

### Chemically Bonded

This is the most common approach to producing a chelating surface as can be seen from the selection shown in Table 1A. Although they are all capable of producing separations, only a small number appear to have sufficiently good kinetics for chromatographic separations, the most popular being IDA and aminophosphonic acid.

### Impregnation

Relatively large chelating dyestuff molecules can form an essentially permanent coating by becoming 'trapped' in the pores of the substrate. PS-DVB resins are the most successful for this mode as  $\pi$ - $\pi$  interactions between the aromatic rings of the dye and the resin help to hold the chelating molecule in place. Table 1B shows some representative examples of those used for chromatographic separations. Those containing the IDA group appear to give the best efficiency, though not as good as the chemically bonded type, presumably due to the chelating groups being more sterically hindered in the pore structure.

### Dynamic Coating

Dynamic coating or modification involves creating a chelating surface by having a ligand continuously present in the eluent. Conditions are chosen so that no permanent impregnation occurs, but rather a dynamic equilibrium is established between a sorbed layer of ligand on the surface of the substrate and the concentration of the ligand in the mobile phase. Due to the preferred adsorption of the ligand, its concentration at the surface is much higher than in the eluent, thus producing a chelation ion-exchange mechanism. Altering such parameters as the concentration of ligand in the eluent, mobile phase pH, and the nature of the stationary phase allows the dynamic

equilibrium to change, thus affecting metal selectivity. This is a relatively new approach and some representative examples are given in Table 1C. Although large molecules such as chelating dyestuffs have been successfully used in the dynamic mode for chromatographic separations, most work has concentrated on small molecules with aromatic groups such as dipicolinic acid.

## Practical Aspects of Chelation Ion-Chromatography

### Precolumn Systems with Common Chromatographic Equipment

The original and most widespread application of chelating ion-exchangers is in the selective adsorption of one or more metal ions of interest from diluted solutions followed by sensitive determination. The short columns or cartridges are not designed for the efficient chromatographic separation of metal ions, but rather for collection on the column before being swept off as a concentrated 'plug'. By this means matrix isolation and/or preconcentration can be carried out before further analysis online or offline. For CIC, one approach, as explained previously, is to use a chelating precolumn as part of an online IC system and can be considered as one constituent of the overall separation/determination process. Conventional ion-exchange columns, connected in series with the chelating column, are used for the analytical separation of metal ions before postcolumn reaction detection. The most common chelating columns are based on IDA or 8-hydroxyquinoline, bonded to controlled pore glass, silica, or PS-DVB resins. Some are available commercially, though many 'home made' columns are described in the literature. These systems have been used successfully to determine trace metals in a number of complex samples, though of necessity the set-up can be rather complex. One commercially available system, for example, involves three columns (the first one being the chelating ion-exchange column) several pumps, eluents, and switching valves.

### High-Performance Chelation Ion Chromatography

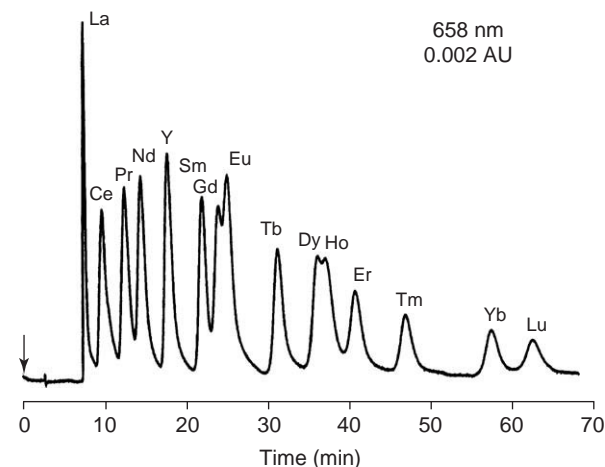
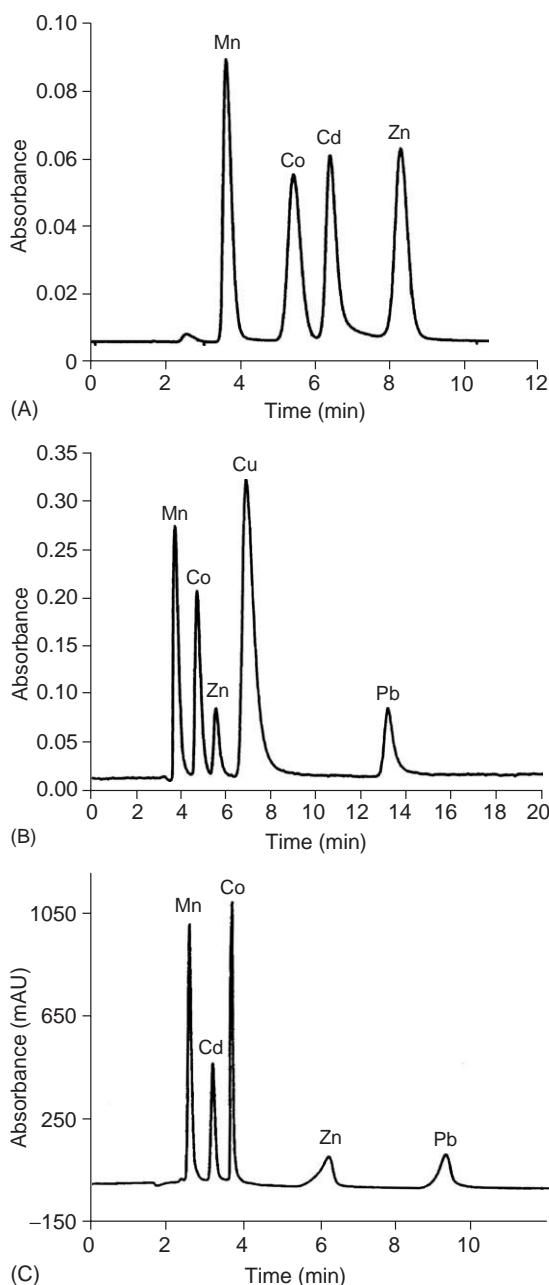
A relatively recent approach in CIC involves more efficient, small particle size substrates as the chelating surface. The use of these high-performance stationary phases means that reasonably fast analytical separations of groups of metal ions can be obtained similar to conventional high-performance ion-exchange. Matrix isolation is not normally required as the chelating column is relatively insensitive to concentrated salt solutions. Preconcentration, if necessary, can also

be carried out using a step pH gradient. Thus, matrix isolation, preconcentration, and analytical separation can be achieved using a single high-efficiency chelating column, removing the need for a complex multicolumn system as described in the previous section.

The sorption process controlling separation can be almost pure chelation or a combination of chelation and simple ion-exchange as discussed in detail in the section on chelation ion-exchange equilibria. A straightforward way of controlling the relative contribution of chelation and simple ion-exchange is

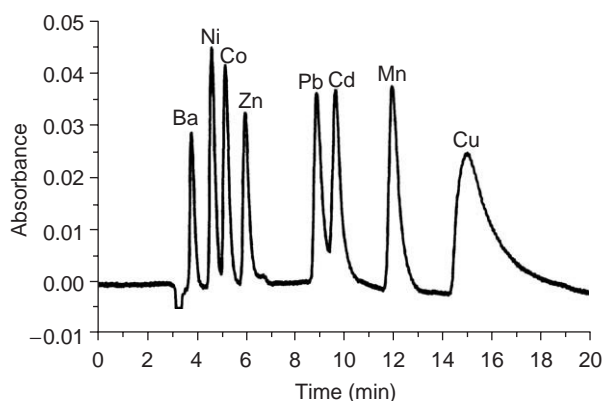
through the ionic strength of the eluent. Increasing the concentration of an alkali metal salt such as potassium nitrate in the eluent will suppress simple ion-exchange interactions. Concentrations as high as one molar have been used to achieve almost complete suppression of ion-exchange for many metal ions.

The chromatograms in Figures 1–6 were selected to show the kind of analytical separations and selectivities available for groups of metal ions for the three types of chelating surface: bonded,

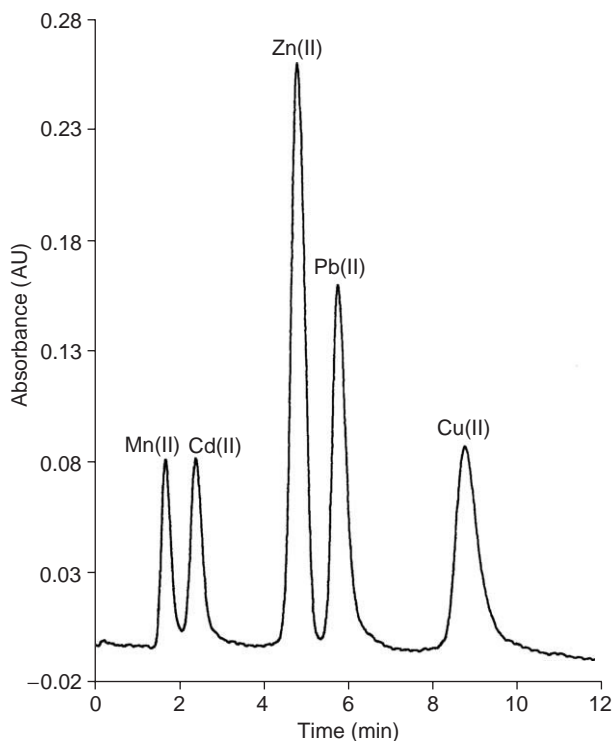


**Figure 2** Isocratic separation of standard mixture of 14 lanthanides and yttrium on a 250 mm × 4 mm column, packed with 5  $\mu\text{m}$  silica IDA. Eluent:  $1.6 \times 10^{-2} \text{ mol l}^{-1} \text{ HNO}_3$  with  $0.5 \text{ mol l}^{-1} \text{ KNO}_3$ ; flow rate  $1.0 \text{ ml min}^{-1}$ ; column temperature  $65^\circ\text{C}$ ; sample volume  $20 \mu\text{l}$ , sample concentration of each metal was 4 ppm in 0.2%  $\text{HNO}_3$ . Detection, Arsenazo III postcolumn reaction at 658 nm. (Reprinted with permission from Nesterenko PN and Jones P (1998) Isocratic separation of lanthanides and yttrium by high performance chelation ion chromatography. *Journal of Chromatography A* 804: 223–231; © Elsevier.)

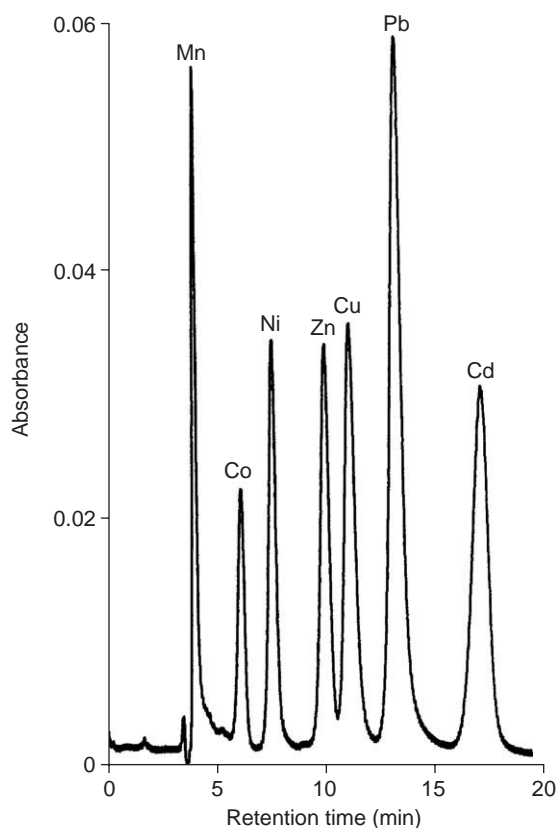
**Figure 1** Separations on silica IDA columns with different elution protocols. (A) Isocratic separation of four metal ions on a 100 mm × 4 mm column packed with 5  $\mu\text{m}$  IDA silica. Eluent,  $10 \text{ mmol l}^{-1}$  nitric acid. Detection, PAR postcolumn reaction at 510 nm. (Unpublished work, Nesterenko PN and Jones P.) (B) Isocratic separation of five metal ions on a 250 mm × 4 mm column packed with 5  $\mu\text{m}$  IDA silica. Eluent,  $0.5 \text{ mol l}^{-1} \text{ KCl}$ ,  $20 \text{ mmol l}^{-1}$  picolinic acid, and  $12.5 \text{ mmol l}^{-1}$  nitric acid. Detection, PAR postcolumn reaction at 510 nm. (Unpublished work, Nesterenko PN and Jones P.) (C) Step gradient separation of Mn(II), Cd(II), Co(II), Zn(II), and Pb(II). Eluent conditions:  $0.1 \text{ mol l}^{-1} \text{ NaCl}$  (pH 2.6) switched to  $0.1 \text{ mol l}^{-1} \text{ NaCl}$  (pH 1.6) at time = 3 min prior to standard injection. Column, 250 mm × 4 mm, packed with 8  $\mu\text{m}$  silica IDA. Detection, PAR postcolumn reaction at 495 nm. (Reprinted with permission from Bashir W and Paull B (2002) Ionic strength, pH and temperature effects upon selectivity for transition and heavy metal ions when using chelation ion chromatography with an iminodiacetic acid bonded silica gel column and simple eluents. *Journal of Chromatography* 942: 73–82; © Elsevier.)



**Figure 3** The separation of Ba(II) 1.5 ppm, Ni(II) 1 ppm, Co(II) 2 ppm, Zn(II) 1 ppm, Pb(II) 5 ppm, Cd(II) 2.5 ppm, Mn(II) 1.5 ppm, and Cu(II) 5 ppm on the aminophosphonic acid functionalized silica column, 250 mm  $\times$  4.6 mm. Eluent 1 mol l<sup>-1</sup> KNO<sub>3</sub>, 5 mmol l<sup>-1</sup> HNO<sub>3</sub>. Detection, PAR/ZnEDTA postcolumn reaction at 495 nm. (Reprinted with permission from Nesterenko PN, Shaw MJ, Hill S, and Jones P (1999) Aminophosphonate-functionalised silica: A versatile chromatographic stationary phase for high performance chelation ion chromatography. *Microchemical Journal* 62: 58–69; © Elsevier.)



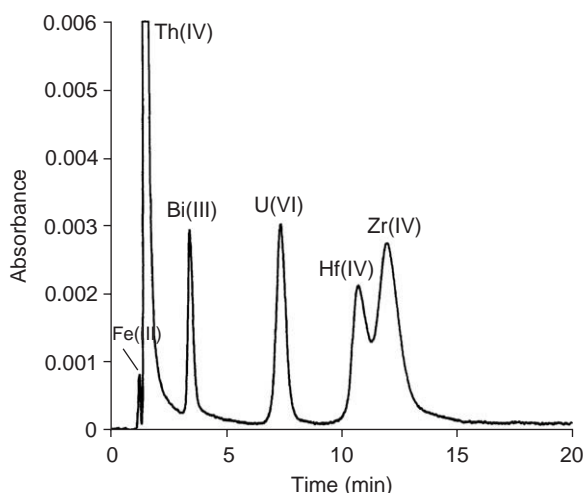
**Figure 4** Gradient elution of a 0.5 mg l<sup>-1</sup> Mn(II), 20 mg l<sup>-1</sup> Cd(II), 10 mg l<sup>-1</sup> Zn(II), 10 mg l<sup>-1</sup> Pb(II), and 1 mg l<sup>-1</sup> Cu(II) mixture on the 100 mm  $\times$  4.6 mm Xylenol Orange impregnated Hamilton resin column. Injection volume used was 100  $\mu$ l with detection at 520 nm with PAR postcolumn reaction. (Reprinted with permission from the Ph.D. dissertation of James Cowan (2002) The development and study of chelating substrates for the separation of metal ions in complex matrices. University of Plymouth, Figure 2.9, p. 84.)



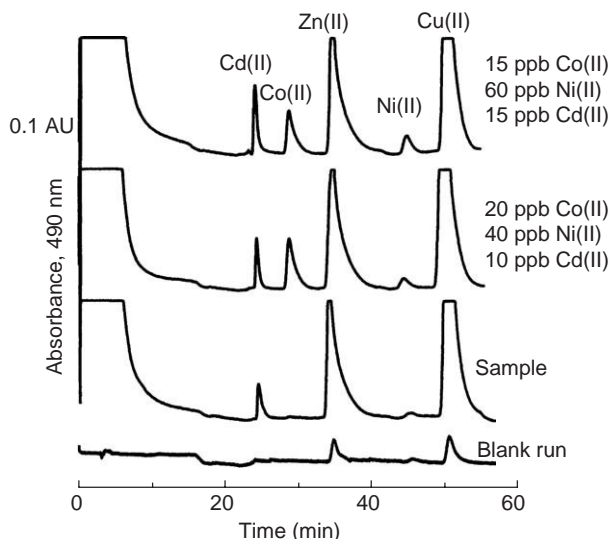
**Figure 5** The separation of Mn(II) 0.5 mg l<sup>-1</sup>, Co(II) 0.5 mg l<sup>-1</sup>, Ni(II) 0.5 mg l<sup>-1</sup>, Zn(II) 2 mg l<sup>-1</sup>, Cu(II) 1 mg l<sup>-1</sup>, Pb(II) 10 mg l<sup>-1</sup>, and Cd(II) 20 mg l<sup>-1</sup> on a 300  $\times$  4.6 mm PRP-1 7  $\mu$ m PS-DVB dynamically coated column. Eluent: 1 mol l<sup>-1</sup> potassium nitrate, 0.25 mmol l<sup>-1</sup> chlorodipicolinic acid, and 6.25 mmol l<sup>-1</sup> nitric acid (pH 2.2). Detection: PAR at 520 nm. (Reprinted with permission from Shaw MJ, Jones P, and Nesterenko PN (2002) Dynamic chelation ion chromatography of transition and heavy metals using a mobile phase containing 4-chlorodipicolinic acid. *Journal of Chromatography A* 953: 141–150; © Elsevier.)

impregnated, and dynamically loaded. For many chelating substrates the selectivity factors for metal ions in simple eluents can vary over a large range. This means that to separate four or more metal ions isocratically in a reasonable time can be a problem. **Figure 1** illustrates this situation quite clearly for silica IDA columns, where a simple dilute nitric acid eluent can separate four relatively weakly retained metal ions, but the more strongly held metal ions such as copper or lead require complexing eluents and/or gradients to elute them more quickly. Although the application of gradients is more complex, gradients have the advantage of sharpening later eluting peaks and allowing on-column preconcentration if required. Two exceptions to this general situation have been found so far. The first involves aminophosphonate bonded silica (**Figure 3**) where eight common metal ions can be separated isocratically in

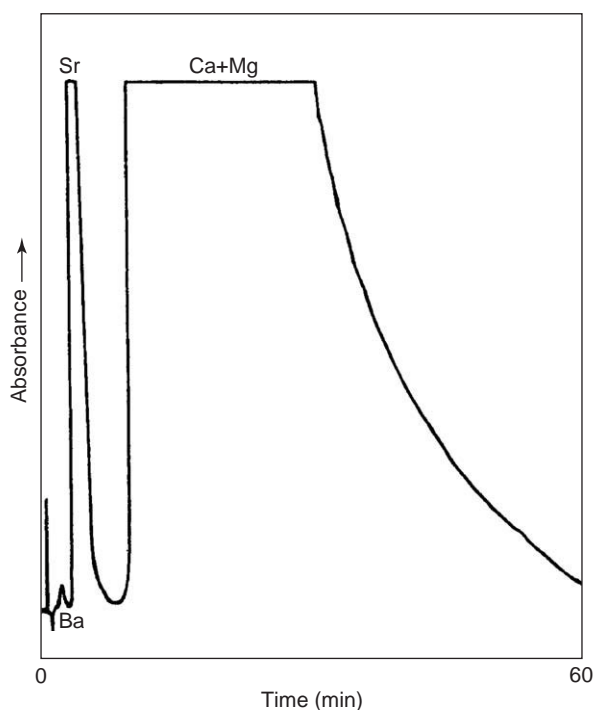




**Figure 6** Chromatogram showing the separation of six acid hydrolyzing metal ions on a dynamically coated column ( $100 \times 4.6$  mm ID), packed with  $5 \mu\text{m}$  PLRPS polystyrene divinylbenzene resin. Eluent,  $1 \text{ mol l}^{-1}$  potassium nitrate and  $0.1 \text{ mol l}^{-1}$  dipicolinic acid in  $0.5 \text{ mol l}^{-1}$  nitric acid. Detection, Arsenazo III postcolumn reaction monitored at  $654 \text{ nm}$ . Sample injection,  $100 \mu\text{l}$  of a mixture of  $20 \text{ ppm}$  Fe(III),  $0.5 \text{ ppm}$  Th(IV),  $20 \text{ ppm}$  Bi(III),  $0.5 \text{ ppm}$  U(VI),  $25 \text{ ppm}$  Hf(IV), and  $5 \text{ ppm}$  Zr(IV). (The ion chromatographic separation of high valence metal cations using a neutral polystyrene resin dynamically modified with dipicolinic acid. Cowan J, Shaw MJ, Achterberg EP, Nesterenko PN, and Jones P (2000) The analyst communication. *Analyst* 125: 2157–2159; reproduced by permission of The Royal Society of Chemistry.)



**Figure 7** Determination of trace metals in seawater showing two standard additions. Column,  $250 \times 4.0$  mm,  $6.5 \mu\text{m}$  silica IDA. Step gradient: 0–10 min,  $0.5 \text{ mol l}^{-1}$  nitric acid– $0.5 \text{ mol l}^{-1}$  KCl; 10–30 min,  $80 \text{ mmol l}^{-1}$  tartaric acid; 30–50 min,  $10 \text{ mmol l}^{-1}$  picolinic acid. Flow rate,  $0.8 \text{ ml min}^{-1}$ . Detection,  $490 \text{ nm}$ , postcolumn reaction with PAR/ $\text{NH}_3/\text{HNO}_3$  reagent. Sample,  $6 \text{ ml}$  Carnon estuary water. (From Nesterenko PN and Jones P, unpublished data.)



**Figure 8** Chromatogram showing the separation of barium and strontium from magnesium and calcium in an oil-well brine (sample 1, diluted 1:1 v/v with double distilled water) using a Methylthymol Blue impregnated column ( $\text{pH } 9.2$ ). (Determination of alkaline earth metals in offshore oil-well brines using high performance chelation ion chromatography. Paull B, Foulkes M, and Jones P (1994) *Analytical Proceedings* 31: 209–211; reproduced by permission of The Royal Society of Chemistry.)

16 min using dilute nitric acid eluent, though iron(III) is strongly retained under these conditions. The second exception concerns the separation of the lanthanides and yttrium on IDA bonded silica (Figure 2). The separation of such a large number of metals isocratically is possible because of the relatively small differences in stability constants and hence  $k$  values between the individual lanthanides. In contrast, this very difficult lanthanide separation is impossible to achieve isocratically by conventional ion-exchange in a realistic time frame and all published methods use gradients with complexing eluents.

### Detection in Chelation Ion Chromatography

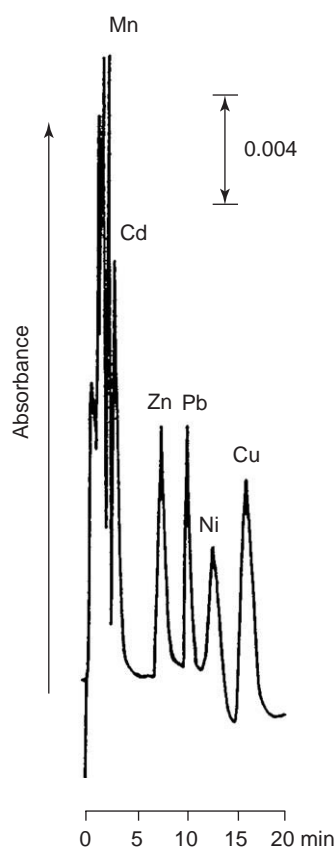
The main mode of detection of eluted metal species is the same as that used in conventional high-performance ion-exchange chromatography, so the reader is referred to the Further Reading section for more detailed information. Essentially, postcolumn reactions involving colorimetric reagents are employed to achieve the highest sensitivity detection. 4-(2-pyridylazo)resorcinol (PAR) is the most commonly used colorimetric reagent, reacting sensitively with a wide range of metals. However, it should be pointed

**Table 2** Practical applications of HPCIC for the determination of trace metals

Functional groups	Sorbent (particle size, pore diameter, surface area, and capacity)	Column (mm)	Mobile phase	Metal ions separated (in order of retention)	Photometric detection	Application
<b>(A) Chemically bonded ligands</b>						
Iminodiacetic acid	8 $\mu\text{m}$ silica, 13 nm, 350 $\text{m}^2\text{g}^{-1}$ , 320 $\mu\text{mol g}^{-1}$	250 $\times$ 4.0	Gradient elution 0.035 $\text{mol l}^{-1}$ KCl– 0.065 $\text{mol l}^{-1}$ $\text{KNO}_3$ , pH 2.5 1 $\text{mol l}^{-1}$ $\text{KNO}_3$ , pH 4.9	Mg, Ca, Mn, Cd, Co, Zn, Ni, Pb, Cu Mg, Ca, Mn, Cd, Co, Zn, Pb, Ni, Cu Mg, Ba, Sr, Ca, Be	PAR, 550 nm PAR, 550 nm $\alpha$ -CPC, 572 nm	Mg, Ca, Mn, Cd, Co, Zn, Ni, Cu in seawater Mn, Cd, Co, Zn in freshwater Mg, Ca in NaCl and KCl brine solutions; eyewash saline solution
			0.4 $\text{mol l}^{-1}$ $\text{KNO}_3$ , pH 2.5	Mg, Ba, Sr, Ca, Be	CAS, 590 nm	Be in tap water, waste water, seawater
			1 $\text{mmol l}^{-1}$ dipicolinic acid, pH 3.0	Cu, Zn, Pb, Co, Cd	Direct UV, 290 nm	Zn, Pb in waste waters of galvanic bath
	10 $\mu\text{m}$ polymer, 100 nm, 20 $\mu\text{mol ml}^{-1}$	75 $\times$ 7.5	0.2 $\text{mol l}^{-1}$ KCl, 50 $\text{mmol l}^{-1}$ phosphate buffer, 0.1 $\text{mmol l}^{-1}$ CPC, pH 5.3	Mg, Sr, Ca	Direct with buffering, 575 nm	Mg, Ca in coastal seawater
Aminomethylphosphonic acid	5 $\mu\text{m}$ silica, 13 nm, 350 $\text{m}^2\text{g}^{-1}$ , 100 $\mu\text{mol g}^{-1}$	50 $\times$ 4.6	1 $\text{mol l}^{-1}$ $\text{KNO}_3$ , 0.5 $\text{mol l}^{-1}$ $\text{HNO}_3$ , 0.08 $\text{mol l}^{-1}$ ascorbic acid	Ni, Zn, Cu, Cd, Mn, Al, Be, La, Lu	CAS, 560 nm	Be in stream sediment
2-Pyridinecarboxyaldehyde phenylhydrazone	5 $\mu\text{m}$ silica, 12 nm, 400 $\text{m}^2\text{g}^{-1}$ , 18 $\mu\text{mol g}^{-1}$	250 $\times$ 4.6	Oxalate buffer, pH 4.5	Mn, Fe, Cd, Zn, Co, Pb, Cu	PAR, 550 nm	Mn, Fe, Zn, Cu in tomato leaves and vitamin tablets
Amidoxime	5 $\mu\text{m}$ silica, 13 nm, 350 $\text{m}^2\text{g}^{-1}$ , 130 $\mu\text{mol g}^{-1}$	50 $\times$ 3	5 $\text{mmol l}^{-1}$ oxalic acid, pH 2.2	Cd, Pb, Co, Zn, Ni, Cu	PAR, 540 nm	Transition metals in seawater
<b>(B) Precoated or impregnated substrates</b>						
Xylenol Orange	8 $\mu\text{m}$ PS-DVB, 10 nm, 414 $\text{m}^2\text{g}^{-1}$ , 31 $\mu\text{mol g}^{-1}$	100 $\times$ 4.6	1 $\text{mol l}^{-1}$ $\text{KNO}_3$ , pH 7.7	Ba, Sr, Mg, Ca	PAR-ZnEDTA, 490 nm	Alkaline-earth metals in KCl or NaCl brines
			1 $\text{mol l}^{-1}$ $\text{KNO}_3$ , pH gradient	Ba, Sr, Mg, Ca, Mn, Cd, Zn, Ni, Cu, Fe	PAR-ZnEDTA, 490 nm	Transition metals in coastal or estuarine seawater, inorganic chemicals
Methyl thymol Blue	8.8 $\mu\text{m}$ PS-DVB, 12 nm, 470 $\text{m}^2\text{g}^{-1}$ , 48 $\mu\text{mol g}^{-1}$	100 $\times$ 4.6	1 $\text{mol l}^{-1}$ $\text{KNO}_3$ , 0.05 $\text{mol l}^{-1}$ lactic acid, pH 8.5	Ba, Sr, Mg, Ca	PAR-ZnEDTA, 490 nm	Ba, Sr in mineral water
			1 $\text{mol l}^{-1}$ $\text{KNO}_3$ , pH 7.9	Ba, Sr, Mg, Ca	PAR-ZnEDTA, 490 nm	Ba, Sr, Mg, Ca in oil-well brines
	5 $\mu\text{m}$ PS-DVB, 10 nm	150 $\times$ 4.1	0.5 $\text{mol l}^{-1}$ $\text{KNO}_3$ , pH 1.2	Ca, Mg, U, Cu	Arsenazo III, 600 nm	U in saline lake water

<i>o</i> -Cresolphthalein complexone	8.8 $\mu\text{m}$ PS-DVB, 12 nm, 470 $\text{m}^2\text{g}^{-1}$ , 48 $\mu\text{mol g}^{-1}$	100 $\times$ 4.6	1 $\text{mol l}^{-1}$ $\text{KNO}_3$ , 0.05 $\text{mol l}^{-1}$ lactic acid, pH 9.8	Ba, Sr, Mg, Ca	PAR-ZnEDTA, 490 nm	Sr in milk powder
Chrome Azurol S	8.8 $\mu\text{m}$ PS-DVB, 12 nm, 470 $\text{m}^2\text{g}^{-1}$ , 35 $\mu\text{mol g}^{-1}$	100 $\times$ 4.6	1 $\text{mol l}^{-1}$ $\text{KNO}_3$ , pH step gradient from 4.0 to 1.1	Al, Fe	PCV, 580 nm	Al in seawater
Aurin tricarboxylic acid (ATA)	7 $\mu\text{m}$ PS-DVB, 10 nm, 150 $\mu\text{mol g}^{-1}$	250 $\times$ 4.6	1 $\text{mol l}^{-1}$ $\text{KNO}_3$ , 50 $\text{mmol l}^{-1}$ acetic acid, pH step gradient from 4.5 to 1.0	Mn, Co, Ni, Cd, Pb, Cu, Ga, In	PAR, 520 nm	Pb, Cd, Cu in highly mineralized water
<i>(C) Dynamically coated</i>						
Dipicolinic acid	5 $\mu\text{m}$ PS-DVB	100 $\times$ 4.6	1 $\text{mol l}^{-1}$ $\text{KNO}_3$ , 0.5 $\text{mol l}^{-1}$ $\text{HNO}_3$ , 0.1 $\text{mmol l}^{-1}$ dipicolinic acid	Fe, Th, V, Bi, U, Hf, Zr	Arsenazo III, 654 nm or PCV, 585 nm	U, Bi, Zr in sediment. U in mineral water, seawater
4-Chlorodipicolinic acid	7 $\mu\text{m}$ PS-DVB, 10 nm	300 $\times$ 4.6	1 $\text{mol l}^{-1}$ $\text{KNO}_3$ , 0.25 $\text{mmol l}^{-1}$ chlorodipicolinic acid, pH 1.5	Mn, Co, Ni, Zn, Cu, Pb, Cd, Al, La, Lu, Fe, U	PAR, 520 nm	Pb, Cd, Cu in rice flour
<i>o</i> -Cresolphthalein complexone (CPC)	5 $\mu\text{m}$ porous graphitic carbon, 25 nm, 120 $\text{m}^2\text{g}^{-1}$	100 $\times$ 4.6	45–58% MeOH, 0.4 $\text{mmol l}^{-1}$ CPC, pH 10.0–10.5	Mg, Ca	Direct detection, 600 nm	Mg, Ca in seawater, saline lake water
	5 $\mu\text{m}$ PS-DVB, 10 nm	150 $\times$ 4.1	0.5 $\text{mol l}^{-1}$ $\text{KNO}_3$ , 0.2 $\text{mmol l}^{-1}$ CPC, 20 $\text{mmol l}^{-1}$ borate, pH 9.5	Ba, Sr, Ca, Mg	CPC, 575 nm	Sr in Antarctic saline lake water
Methyl thymol Blue (MTB)	5 $\mu\text{m}$ PS-DVB, 10 nm	150 $\times$ 4.1	0.5 $\text{mol l}^{-1}$ $\text{KNO}_3$ , 0.2 $\text{mmol l}^{-1}$ MTB, pH 1.2	Mg, Mn, Zn, Cd, Pb	Buffering at pH 5.9 with $\text{NH}_4\text{OAc}$ , 600 nm	Zn in industrial gypsum
<i>N</i> - <i>n</i> -dodecyliminodiacetic acid	5 $\mu\text{m}$ ODS-silica, 14 nm, 300 $\text{m}^2\text{g}^{-1}$	150 $\times$ 4.0	75 $\text{mmol l}^{-1}$ tartrate buffer, pH 5.5	Ca, Sr, Mg, Ba	CPC, 575 nm or HQS-Mg-EDTA, fluorimetric 405/525 nm	Mg, Ca in coastal seawater

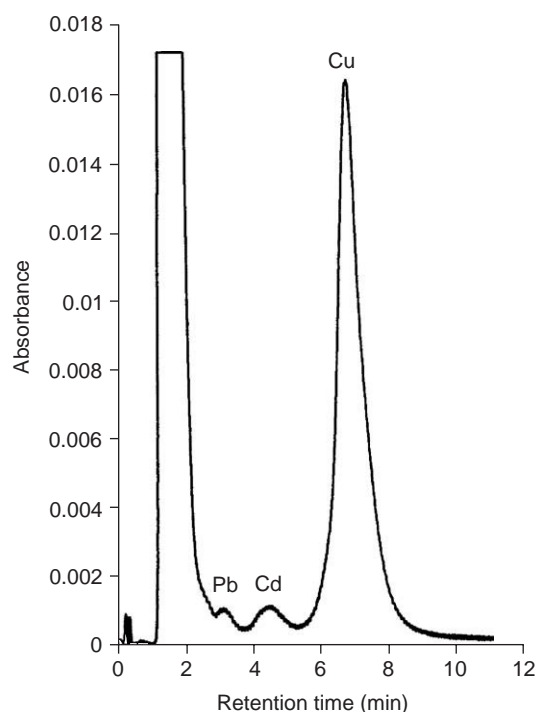
Postcolumn reagents: PAR – 4-(2-pyridylazo)resorcinol; *o*-CPC – *o*-cresolphthalein complexone; CAS – chrome azurol S; PCV – pyrocatechol violet; HQS – 8-hydroxyquinoline-5-sulfonic acid.



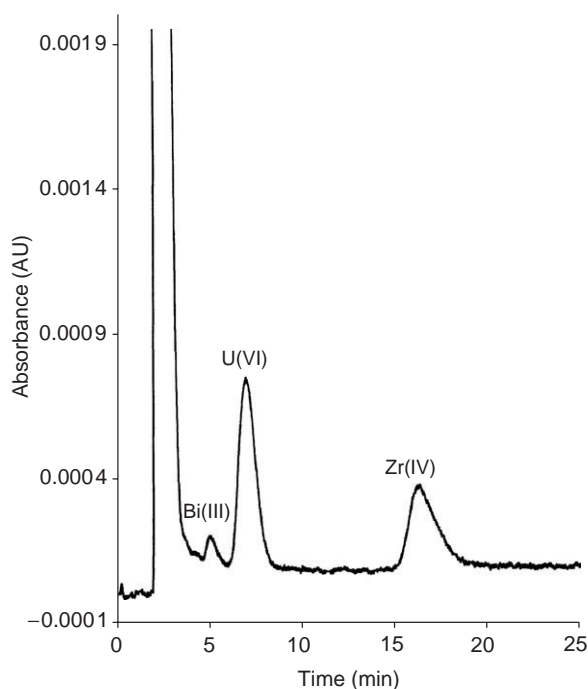
**Figure 9** Chromatogram showing the preconcentration and separation of trace metals in  $1 \text{ mol l}^{-1} \text{ Na}_2\text{SO}_4$  sample spiked with  $10 \mu\text{g l}^{-1} \text{ Mn}^{2+}$ ,  $25 \mu\text{g l}^{-1} \text{ Zn}^{2+}$ , and  $0.1 \text{ mg l}^{-1} \text{ Cd}^{2+}$ ,  $\text{Pb}^{2+}$ ,  $\text{Ni}^{2+}$ , and  $\text{Cu}^{2+}$ . Sample volume was 10 ml adjusted to pH 6. Column,  $100 \text{ mm} \times 4.6 \text{ mm}$ , packed with Xylenol Orange impregnated  $8 \mu\text{m}$  PLRPS resin. (Reprinted with permission from Challenger OJ, Hill SJ, and Jones P (1993) Separation and determination of trace metals in concentrated salt solutions using chelation ion chromatography. *Journal of Chromatography* 639: 197–205; © Elsevier.)

out that lower limits of detection can be achieved using CIC compared to conventional IC separations. This is because many IC methods use strong complexing acids at relatively high concentrations in the eluents, which compete with colorimetric reagents such as PAR, reducing optimum sensitivity. In contrast, most CIC methods use either noncomplexing eluents or eluents containing relatively weak complexing acids.

**Figure 11** Isocratic separation of  $\text{Bi(III)}$ ,  $\text{U(VI)}$ , and  $\text{Zr(IV)}$  in GBW07311 sediment sample at pH 0 on the  $15 \text{ cm}$  PLRP-S column dynamically modified with  $0.1 \text{ mmol l}^{-1}$  dipicolinic acid. Injection volume used was  $500 \mu\text{l}$  with detection at  $654 \text{ nm}$  with Arsenazo III postcolumn reaction. (Reproduced with permission from the Ph.D. dissertation of James Cowan (2002) The development and study of chelating substrates for the separation of metal ions in complex sample matrices. University of Plymouth, Figure 5.15, p. 218.)



**Figure 10** The separation of  $\text{Cd(II)}$ ,  $\text{Pb(II)}$ , and  $\text{Cu(II)}$  from matrix interferences in the certified rice flour GBW08502 on the  $100 \times 4.6 \text{ mm}$  PRP-1  $7 \mu\text{m}$  PS-DVB column. Eluent:  $1 \text{ mol l}^{-1}$  potassium nitrate,  $30 \text{ mmol l}^{-1}$  nitric acid, and  $0.25 \text{ mmol l}^{-1}$  chlorodipicolinic acid. Detection: PAR at  $520 \text{ nm}$ . (Reprinted with permission from Shaw MJ, Jones P, and Nesterenko PN (2002) Dynamic chelation ion chromatography of transition and heavy metals using a mobile phase containing 4-chlorodipicolinic acid. *Journal of Chromatography A* 953: 141–150; © Elsevier.)



## Applications

In this section, examples of applications involving high-efficiency CIC columns have been selected to demonstrate the wide range of complex samples analyzed using this technique (Table 2). The table is divided into three sections in terms of how the chelating substrate is formed, namely, bonded, impregnated, and dynamically coated. The insensitivity to ionic strength and special selectivity of these columns can be especially seen in the ability to analyze highly saline samples and digested sediment and mineral samples containing relatively enormous amounts of matrix metals. Figures 7–11 have been selected to illustrate some of the actual chromatographic separations obtained under these conditions.

See also: **Ion Exchange:** Overview; Principles; Ion Chromatography Instrumentation; Ion Chromatography Applications; Isotope Separation. **Liquid Chromatography:** Ion Pair.

## Further Reading

Fritz JS and Gjerde DT (2000) *Ion Chromatography*, 3rd edn. Heidelberg and New York: Huethig.

Haddad PR and Jackson PE (1990) *Ion Chromatography, Principles and Applications*. Amsterdam: Elsevier.

Jones P and Nesterenko PN (1997) High performance chelation ion chromatography: a new dimension in the separation and determination of trace metals. *Journal of Chromatography* 789: 413–435.

Paull B and Haddad PR (1999) Chelation ion chromatography of trace metal ions using metallochromic ligands. *Trends in Analytical Chemistry* 18: 107–114.

Paull B and Jones P (1996) A comparative study of the metal selective properties of chelating dye impregnated resins for the IC separation of trace metals. *Chromatographia* 42: 528–538.

Sarzanini C (1999) Liquid chromatography: a tool for the analysis of metal species. *Journal of Chromatography A* 850: 213–228.

Sarzanini C and Mentasi E (1997) Determination and speciation of metals by liquid chromatography. *Journal of Chromatography A* 789: 301–321.

Timerbaev AR and Bonn GK (1997) Complexation ion chromatography – an overview of developments and trends in trace metal analysis. *Journal of Chromatography* 640: 195–206.

Weiss J (1995) *Ion Chromatography*. Weinheim: VCH.

## Isolation of Biopolymers

**P R Levison**, Pall Europe Ltd., Portsmouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Biopolymers, the so-called ‘building blocks of nature’, are found in all living matter be it of animal, microbial, or vegetable origin. Biopolymers include proteins (polymers of amino acids), genetic material (polymers of nucleic acids), glycoforms (carbohydrates and glycosylated molecules), metabolites, and other structural molecules. By their very sequence and chemical composition many biopolymers have an electrical charge and can therefore be fractionated by ion-exchange processes. This article briefly reviews the principles underlying biopolymer purification by ion exchange and addresses some of the process issues associated with their purification.

## Ion-Exchange Chemistries

Ion-exchange chromatography is routinely used for the separation of biopolymers at laboratory scale

through to process scale. Ion exchangers are available from a number of manufacturers each produced using their proprietary synthesis methods. For biopolymer separations there are typically four ionic functionalities available. In the case of anion exchange, there are so-called weak anion exchangers (WAX) often based on weak amines including aminoethyl, diethylaminoethyl, guanidoethyl, *p*-aminobenzyl, ECTEOLA, and polyethyleneimine, and strong anion exchangers often based on a quaternary amine. Examples of each functionality are represented in Figure 1.

The choice of anion exchanger is process dependent and for high pH applications, i.e., pH > 8.5 the more ionized grade, e.g., *N,N,N*-trimethyl-2-hydroxypropyl amine, would be recommended.

In the case of cation exchange there are so-called weak cation exchangers often based on a carboxylic acid and strong cation exchangers often based on a sulfonic acid. Examples of each functionality are represented in Figure 2.

Choice of cation exchanger is also process dependent but for low pH applications, i.e., pH < 4.5, the more ionized grades, e.g., sufoxyethyl or sulfo-propyl, would be recommended.



## Applications

In this section, examples of applications involving high-efficiency CIC columns have been selected to demonstrate the wide range of complex samples analyzed using this technique (Table 2). The table is divided into three sections in terms of how the chelating substrate is formed, namely, bonded, impregnated, and dynamically coated. The insensitivity to ionic strength and special selectivity of these columns can be especially seen in the ability to analyze highly saline samples and digested sediment and mineral samples containing relatively enormous amounts of matrix metals. Figures 7–11 have been selected to illustrate some of the actual chromatographic separations obtained under these conditions.

See also: **Ion Exchange:** Overview; Principles; Ion Chromatography Instrumentation; Ion Chromatography Applications; Isotope Separation. **Liquid Chromatography:** Ion Pair.

## Further Reading

Fritz JS and Gjerde DT (2000) *Ion Chromatography*, 3rd edn. Heidelberg and New York: Huethig.

Haddad PR and Jackson PE (1990) *Ion Chromatography, Principles and Applications*. Amsterdam: Elsevier.

Jones P and Nesterenko PN (1997) High performance chelation ion chromatography: a new dimension in the separation and determination of trace metals. *Journal of Chromatography* 789: 413–435.

Paull B and Haddad PR (1999) Chelation ion chromatography of trace metal ions using metallochromic ligands. *Trends in Analytical Chemistry* 18: 107–114.

Paull B and Jones P (1996) A comparative study of the metal selective properties of chelating dye impregnated resins for the IC separation of trace metals. *Chromatographia* 42: 528–538.

Sarzanini C (1999) Liquid chromatography: a tool for the analysis of metal species. *Journal of Chromatography A* 850: 213–228.

Sarzanini C and Mentasi E (1997) Determination and speciation of metals by liquid chromatography. *Journal of Chromatography A* 789: 301–321.

Timerbaev AR and Bonn GK (1997) Complexation ion chromatography – an overview of developments and trends in trace metal analysis. *Journal of Chromatography* 640: 195–206.

Weiss J (1995) *Ion Chromatography*. Weinheim: VCH.

## Isolation of Biopolymers

**P R Levison**, Pall Europe Ltd., Portsmouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Biopolymers, the so-called ‘building blocks of nature’, are found in all living matter be it of animal, microbial, or vegetable origin. Biopolymers include proteins (polymers of amino acids), genetic material (polymers of nucleic acids), glycoforms (carbohydrates and glycosylated molecules), metabolites, and other structural molecules. By their very sequence and chemical composition many biopolymers have an electrical charge and can therefore be fractionated by ion-exchange processes. This article briefly reviews the principles underlying biopolymer purification by ion exchange and addresses some of the process issues associated with their purification.

## Ion-Exchange Chemistries

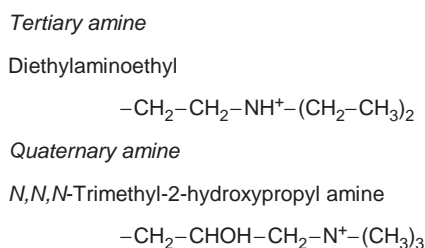
Ion-exchange chromatography is routinely used for the separation of biopolymers at laboratory scale

through to process scale. Ion exchangers are available from a number of manufacturers each produced using their proprietary synthesis methods. For biopolymer separations there are typically four ionic functionalities available. In the case of anion exchange, there are so-called weak anion exchangers (WAX) often based on weak amines including aminoethyl, diethylaminoethyl, guanidoethyl, *p*-aminobenzyl, ECTEOLA, and polyethyleneimine, and strong anion exchangers often based on a quaternary amine. Examples of each functionality are represented in Figure 1.

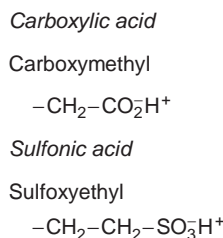
The choice of anion exchanger is process dependent and for high pH applications, i.e., pH > 8.5 the more ionized grade, e.g., *N,N,N*-trimethyl-2-hydroxypropyl amine, would be recommended.

In the case of cation exchange there are so-called weak cation exchangers often based on a carboxylic acid and strong cation exchangers often based on a sulfonic acid. Examples of each functionality are represented in Figure 2.

Choice of cation exchanger is also process dependent but for low pH applications, i.e., pH < 4.5, the more ionized grades, e.g., sufoxyethyl or sulfo-propyl, would be recommended.



**Figure 1** Chemical formulae for anion-exchange functional groups.



**Figure 2** Chemical formulae for cation-exchange functional groups.

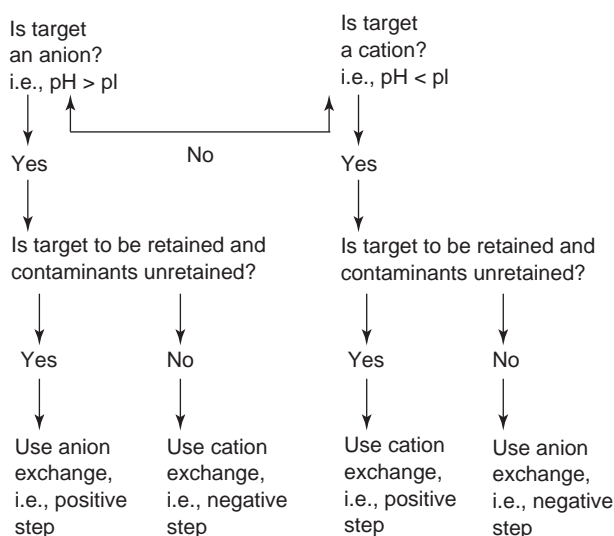
The charged functional groups are typically covalently bonded to an inert and insoluble support matrix often through an ether linkage. A number of support matrices are used traditionally based on polysaccharides such as agarose, cellulose, or dextran. More recently, polymeric and composite matrices have appeared, the selection often dictated by the nature of the chromatographic step.

## Media Selection

Media selection is perhaps the most challenging aspect of the development of the chromatographic separation. Media screening exercises are a useful avenue to pursue and there is enabling hardware available to facilitate parametric screening in an automated mode. However, rather than be faced with the task of screening in excess of 100 ion exchangers a preliminary desktop screen may be carried out to reduce the number of potential candidates. Factors such as the nature of the target molecule in terms of molecular weight and net surface charge, the nature of the contaminants, the conductivity of the feedstock, and both the upstream process and subsequent downstream process may impact on the selection of ion exchanger.

Ion exchangers can be used in either positive or negative chromatographic modes, i.e., target binds, impurities pass or target passes, impurities bind. This approach is summarized in **Figure 3**.

In order to effect purification the biopolymer-containing feedstream needs to be contacted with the ion



**Figure 3** Approaches to development of an ion-exchange chromatographic process.

exchanger. This process referred to as adsorption, which is kinetically limited, is an ionic process between adsorbent and adsorbate. This process operation must be carried out in a suitable contacting system. These are typically either column or batch stirred tanks and various designs exist. Several system configurations are available. Examples include use of axial or radial flow columns, fixed or variable volume, manual or automated packing/unpacking, use in packed bed, expanded bed, fluidized bed, or suspended bed, and magnetically coupled separations. The requirements of the chromatographic process itself and the adsorption kinetics that directly correlate with contact time may dictate the mode of contacting adsorbent and adsorbate. Following biopolymer adsorption, the adsorbent is typically washed to remove nonbound material. Biopolymers are next desorbed and eluted from the ion exchanger using either a change in pH or ionic strength. For similar reasons this desorption process, which again relies on an ionic interaction between adsorbate, counterion, and adsorbent, may be carried out under step or gradient conditions with the selection dictated by the process itself and the subsequent downstream operation.

Selection of anion or cation exchange will be dictated by many of these issues together with physicochemical parameters such as pH, pI, relative molecular mass, temperature, viscosity. The complex nature of naturally occurring biopolymer systems and the ability of biopolymers to adsorb to surfaces through several modes of interaction often results in incomplete desorption of adsorbate from the ion exchanger. It is well known that up to  $2 \text{ mol l}^{-1}$  NaOH is a suitable regenerant for ion-exchange media and will remove most, if not all, of these strongly

adsorbed species. It is important during media selection to develop a suitable regeneration strategy that will meet the quality requirements but without adverse impact on the chemical integrity of the adsorbent itself.

## Biopolymer Separations

While there are numerous biopolymer separations including proteins, nucleic acids, peptides, sugars, metabolites, etc., that have been published using ion-exchange cellulose at the laboratory scale, few process-scale separations are reported. By way of examples, the large-scale separations of hen egg-white proteins and the separation of immunoglobulin G from goat serum with anion-exchange celluloses have been reported and demonstrated to readily scale by up to 1000-fold.

These examples mentioned above used clarified feedstreams. Such separations are typically carried out using batch or column contactors and are well established. However, as the need for intensified and integrated processes has grown there has been a drive to position the ion-exchange steps earlier in the process and thereby reduce the complexity of upstream clarification-related processes, together with the cycle time and the overall operating costs.

One approach to achieve these objectives is the direct adsorption of the target from a particulate-containing feedstream. Over recent years the technique of expanded bed chromatography has emerged to address this very issue. By expanding an adsorbent bed with upward flow of the feedstream, bioparticulates (cells, cell debris, organelles, etc.) can pass relatively unimpeded through the enhanced bed voidage without seriously constraining the adsorption of target products to the stationary phase. This approach has been well documented for various biopolymer purifications including monoclonal antibodies, enzymes including xylanase, plasmid DNA, and other bioactive molecules such as endostatin. The technique demands the use of customized dense adsorbents in a specialized contactor requiring relatively sophisticated operating protocols.

An alternative approach is the technique of suspended bed chromatography. Here, ion exchangers and columns designed for packed bed operations may be used in alternative protocols. Briefly, adsorption is carried out in batch mode, and the resulting adsorbent suspension is filter collected/clarified in a conventional fixed bed contactor for washing and elution. This approach is enabled by the availability of pump-packed column chromatography systems. The technique was first demonstrated with a clarified

hen egg-white feedstock using anion-exchange cellulose and then used to directly isolate the intracellular enzyme glyceraldehyde-3-phosphate dehydrogenase from wet-milled Bakers' yeast also using anion-exchange cellulose. Although essentially a batch process, suspended bed chromatography could in principle be beneficially applied to the primary capture of biopolymers from unclarified fermentation broths, cell disruptates, and biological extracts provided that the inherent concentration of suspended solids did not compromise the characteristic adsorption, washing, and desorption procedures associated with this new technology.

Membrane-based adsorbers and monolithic phases are gaining a presence in the ion-exchange purification of biopolymers. These formats exploit the rapid binding kinetics and enable high volumetric flow rates to be achieved during operation. For the separation of small organic molecules the use of either liquid chromatography or polymeric resins is often favored. These techniques are preferred for reasons of speed, robustness, solvent compatibility, etc., but for charged organic molecules there is no reason, in principle, why ion exchange should not be used.

Peptides synthetically prepared or originating from protein hydrolysates can be separated on either cation or anion exchangers. For example, in the food and flavor industry fermentation of microorganisms has a role since several organisms produce natural organic acids that have commercial use. The organism *Propionibacterium acidipropionici* secretes propionic acid and *Clostridium acetobutyricum* secretes butyric acid, specialized products with added value and these have been shown to adsorb to anion exchangers. Such approaches may have application in the biotechnology industry particularly using a continuous removal approach such as expanded or suspended bed so feedback inhibition arising from high concentrations of bioproducts is minimized.

Nucleic acid purification is a growing area, for example, in applications of genomics and also the purification of plasmid DNA for gene therapy. Anion-exchange chromatography using packed beds, expanded beds, and magnetically coupled separations is established in this field often using the WAX products to aid desorption due to the high surface charge of the nucleic acid. The use of viral products for both gene therapeutics and vaccines offers scope for anion-exchange chromatography. These large particles typically have a negative charge and although generally excluded from the pores of the exchanger can be effectively purified by this means and industrial processes using such techniques are established.

Detailed examples of these applications are described in the articles cited in the Further Reading section.

See also: **Ion Exchange:** Overview; Principles; Ion Chromatography Applications. **Liquid Chromatography:** Overview.

## Further Reading

- Ganetsos G and Barker PE (eds.) (1993) *Preparative and Production Scale Chromatography*. New York: Dekker.
- Heilweil E, Butts ET, Clark FM, and Schwartz WE (1990) Analysis of butyric acid in a model system. In: Touchstone JC (ed.) *Planar Chromatography in the Life Sciences*, pp. 49–57. Chichester: Wiley.
- Knudsen HL, Fahrner RL, Xu Y, Norling LA, and Blank GS (2001) Membrane ion-exchange chromatography for process-scale antibody purification. *Journal of Chromatography A* 907: 145–154.
- Levison PR, Badger SE, Hathi P, *et al.* (1998) New approaches to the isolation of DNA by ion-exchange chromatography. *Journal of Chromatography A* 827: 337–344.
- Levison PR, Mumford C, Streater M, *et al.* (1997) Performance comparison of low-pressure ion-exchange chromatography media for protein separation. *Journal of Chromatography A* 760: 151–158.
- Levison PR, Streater M, and Dennis JW (1999) A scale-down study into the chromatography of a peptide using the cation-exchange cellulose Express-Ion C. *Journal of Chemical Technology and Biotechnology* 74: 204–207.
- Ling TC, Lyddiatt A, Carmichael I, *et al.* (2003) Direct enzyme adsorption from an unclarified microbial feedstock using suspended bed chromatography. *Journal of Chromatography A* 989: 109–118.
- Mattiasson B (ed.) (1999) Expanded bed chromatography. In: *Bioseparation* 8: 1–271.
- Rathore AS and Velayudhan A (eds.) (2002) *Scale-Up and Optimization in Preparative Chromatography, Principles and Biopharmaceutical Applications*. New York: Dekker.
- Roe S (ed.) (2001) *Protein Purification Applications*, 2nd edn. Oxford: Oxford University Press.
- Verrall MS (ed.) (1996) *Downstream Processing of Natural Products: A Practical Handbook*. Chichester: Wiley.

## Isotope Separation

**Y Marcus**, The Hebrew University of Jerusalem, Jerusalem, Israel

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Ion-exchange chromatography is able to combine the effects of many single-stage separations due to the small height of the equivalent theoretical plates, 0.1–10 mm. Hence, it can be used for the (partial) separation of isotopes for which the separation factors are of the order of 1.001 only. This method has been used for many elements, and in some cases, e.g., boron and uranium, has led to industrial isotope enrichment.

It is extremely difficult to separate isotopes by chemical means, although some of their properties are mass dependent. The differences in property are reflected in the equilibrium constants pertaining to the different isotopic species, corresponding to 0.1–10 J mol<sup>−1</sup> in Gibbs free energy. In order to achieve useful isotope separations it is necessary to multiply the single equilibrium effect many-fold and ion-exchange chromatography and related techniques are excellent methods for this purpose. In contrast to the equilibrium effect, the kinetic isotope effect and the difference in the diffusion coefficients between species in solution containing different isotopes have not found much use in isotope separation by ion

exchange. In solution the ions are solvated and the mass difference between the different isotopes of the ions becomes ‘diluted’ by the masses of the solvent molecules, so that this effect is much smaller than for the ‘bare’ ions.

Taylor and Urey were the first, in 1937, to separate isotopes (<sup>6</sup>Li and <sup>7</sup>Li as Li<sup>+</sup>, and <sup>14</sup>N and <sup>15</sup>N as NH<sub>4</sub><sup>+</sup>) by ion-exchange chromatography, using a zeolite cation exchanger. They employed a breakthrough as well as an elution process and found for lithium that <sup>6</sup>Li was enriched in the tailing edge of the band sorbed on the zeolite. A year later they reported that for potassium (<sup>39</sup>K and <sup>40</sup>K) it was the leading edge of the band in which the lighter isotope was enriched. The next significant report was that of Glueckauf, Barker, and Kitt, in 1949, who pointed out that an essential condition for the successful separation was the maintenance of sharp boundaries of the band sorbed on the ion-exchange column. They employed a polyfunctional cation-exchange resin, and again <sup>6</sup>Li was enriched in the rear edge of the band. Spedding, Powell, and Svec, in 1955, returned to the separation of the isotopes of nitrogen, this time on a synthetic polystyrene sulfonate (PSS) resin. The ion-exchange reaction was accompanied by a chemical equilibrium between ammonium ions and aqueous ammonia. The lighter isotope was enriched in the rear of the sharp band that was displaced a great many lengths along the resin column.



See also: **Ion Exchange:** Overview; Principles; Ion Chromatography Applications. **Liquid Chromatography:** Overview.

## Further Reading

- Ganetsos G and Barker PE (eds.) (1993) *Preparative and Production Scale Chromatography*. New York: Dekker.
- Heilweil E, Butts ET, Clark FM, and Schwartz WE (1990) Analysis of butyric acid in a model system. In: Touchstone JC (ed.) *Planar Chromatography in the Life Sciences*, pp. 49–57. Chichester: Wiley.
- Knudsen HL, Fahrner RL, Xu Y, Norling LA, and Blank GS (2001) Membrane ion-exchange chromatography for process-scale antibody purification. *Journal of Chromatography A* 907: 145–154.
- Levison PR, Badger SE, Hathi P, *et al.* (1998) New approaches to the isolation of DNA by ion-exchange chromatography. *Journal of Chromatography A* 827: 337–344.
- Levison PR, Mumford C, Streater M, *et al.* (1997) Performance comparison of low-pressure ion-exchange chromatography media for protein separation. *Journal of Chromatography A* 760: 151–158.
- Levison PR, Streater M, and Dennis JW (1999) A scale-down study into the chromatography of a peptide using the cation-exchange cellulose Express-Ion C. *Journal of Chemical Technology and Biotechnology* 74: 204–207.
- Ling TC, Lyddiatt A, Carmichael I, *et al.* (2003) Direct enzyme adsorption from an unclarified microbial feedstock using suspended bed chromatography. *Journal of Chromatography A* 989: 109–118.
- Mattiasson B (ed.) (1999) Expanded bed chromatography. In: *Bioseparation* 8: 1–271.
- Rathore AS and Velayudhan A (eds.) (2002) *Scale-Up and Optimization in Preparative Chromatography, Principles and Biopharmaceutical Applications*. New York: Dekker.
- Roe S (ed.) (2001) *Protein Purification Applications*, 2nd edn. Oxford: Oxford University Press.
- Verrall MS (ed.) (1996) *Downstream Processing of Natural Products: A Practical Handbook*. Chichester: Wiley.

## Isotope Separation

**Y Marcus**, The Hebrew University of Jerusalem, Jerusalem, Israel

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Ion-exchange chromatography is able to combine the effects of many single-stage separations due to the small height of the equivalent theoretical plates, 0.1–10 mm. Hence, it can be used for the (partial) separation of isotopes for which the separation factors are of the order of 1.001 only. This method has been used for many elements, and in some cases, e.g., boron and uranium, has led to industrial isotope enrichment.

It is extremely difficult to separate isotopes by chemical means, although some of their properties are mass dependent. The differences in property are reflected in the equilibrium constants pertaining to the different isotopic species, corresponding to 0.1–10 J mol<sup>−1</sup> in Gibbs free energy. In order to achieve useful isotope separations it is necessary to multiply the single equilibrium effect many-fold and ion-exchange chromatography and related techniques are excellent methods for this purpose. In contrast to the equilibrium effect, the kinetic isotope effect and the difference in the diffusion coefficients between species in solution containing different isotopes have not found much use in isotope separation by ion

exchange. In solution the ions are solvated and the mass difference between the different isotopes of the ions becomes ‘diluted’ by the masses of the solvent molecules, so that this effect is much smaller than for the ‘bare’ ions.

Taylor and Urey were the first, in 1937, to separate isotopes (<sup>6</sup>Li and <sup>7</sup>Li as Li<sup>+</sup>, and <sup>14</sup>N and <sup>15</sup>N as NH<sub>4</sub><sup>+</sup>) by ion-exchange chromatography, using a zeolite cation exchanger. They employed a breakthrough as well as an elution process and found for lithium that <sup>6</sup>Li was enriched in the tailing edge of the band sorbed on the zeolite. A year later they reported that for potassium (<sup>39</sup>K and <sup>40</sup>K) it was the leading edge of the band in which the lighter isotope was enriched. The next significant report was that of Glueckauf, Barker, and Kitt, in 1949, who pointed out that an essential condition for the successful separation was the maintenance of sharp boundaries of the band sorbed on the ion-exchange column. They employed a polyfunctional cation-exchange resin, and again <sup>6</sup>Li was enriched in the rear edge of the band. Spedding, Powell, and Svec, in 1955, returned to the separation of the isotopes of nitrogen, this time on a synthetic polystyrene sulfonate (PSS) resin. The ion-exchange reaction was accompanied by a chemical equilibrium between ammonium ions and aqueous ammonia. The lighter isotope was enriched in the rear of the sharp band that was displaced a great many lengths along the resin column.

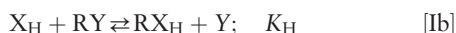
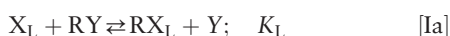


The sustained Japanese effort in this field started with Kakihana and his co-workers in 1959, again with the separation of sodium isotopes. This group at the Tokyo Institute of Technology (TIT) and a few other Japanese groups have devoted considerable efforts to the separation of the isotopes of many elements by ion exchangers, contributing to the theory and the laboratory practice. Preparative-scale isotope separation by ion exchange has been reported mainly for the isotopes of boron,  $^{10}\text{B}$  and  $^{11}\text{B}$ , and uranium,  $^{235}\text{U}$  and  $^{238}\text{U}$ , including industrial implementation. The work at TIT continues, led by Okamoto and now by Fujii. More recently, a South Korean group led by Kim has also undertaken intensive studies of the ion-exchange separation of isotopes.

## Processes for Isotope Separation

### Breakthrough Process

The breakthrough process is the simplest for the separation of two solutes on an ion-exchange column. As the absorbed band travels down the column its front is depleted in the less well sorbed solute, whereas the other solute is enriched in this part of the band. The concentration profiles of the solutes in the effluent are S-shaped. Only the slight difference in the equilibrium constants  $K$  for the ion-exchange reactions of the light (subscript L) and heavy (subscript H) isotope ions (X) with the counterion of the resin (Y) is employed as in reaction [I]:



where R represents an equivalent of the fixed ion of the exchanger. The separation factor, i.e., the ratio  $\alpha = K_\text{L}/K_\text{H}$  is very near unity, and the elementary separation effect (see below)

$$\varepsilon = \alpha - 1 = (K_\text{L}/K_\text{H}) - 1 \quad [1]$$

is very small ( $\sim 10^{-5}$ ). Some of the early studies of the separation of isotopes by ion-exchange chromatography employed this process, e.g., the 1949 study of Gluekauf, Barker, and Kitt on the separation of lithium isotopes, where  $^7\text{Li}$  broke through first.

### Elution Process

In an elution process, a definite amount of the two solutes is sorbed as a band at the top of the ion-exchange column and a solution of an eluent is passed through the column to develop the chromatogram. In ideal cases of fast diffusion in the grains of the ion exchanger and uniform particle size, a symmetrical

bell-shaped (Gaussian) plot of the total concentration of the solute against the volume of eluent results. When the solutes are isotopes of a given ion their individual elution curves are slightly displaced along the abscissa, but a single overall elution curve is obtained, with enrichment of the one isotope in the front of the band (the less strongly sorbed isotope) and of the other in its rear (see Figure 1). For nonideal elution conditions the elution curve becomes skewed, the front being steeper and the rear showing trailing.

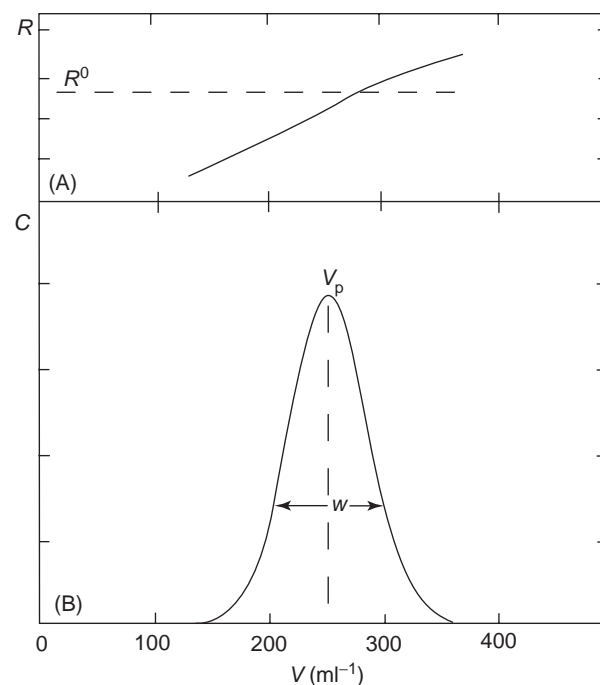
A special feature of the elution process (reaction [II]) is that it permits the superimposition of a complexing reaction with ligands L on the ion-exchange reactions [Ia] and [Ib]:



The separation factor is now

$$\alpha = (K_\text{L}/K_\text{H})(K_{n\text{L}}/K_{n\text{H}})^{\pm 1} \quad [2]$$

where the + sign in the exponent is used in the favorable case where the complexation works in the same direction regarding the isotopes as the ion-exchange reaction. This process has been used mainly for the determination of the separation factor for a given exchanger/eluent/isotope-pair system.



**Figure 1** Ideal elution curve from an ion-exchange column of a di-isotopic single solute ion. (A) The isotopic ratio  $R$  as a function of the elution volume  $V$ . (B) Total concentration  $C$  of the ion in the eluate as a function of  $V$ .

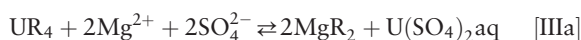
Since it is a batch process, and the band broadens due to longitudinal diffusion as it travels down the column, it is not suitable for preparative separations. An example of its use is the work of Aaltonen in 1972 on the separation of, say,  $^{88}\text{Sr}$  from  $^{87}\text{Sr}$  on a PSS resin of small grain size (400 mesh) with elution by  $\alpha$ -hydroxyisobutyrate solutions. The heavier isotope formed the more stable complex (the less well sorbed cation) and was enriched at the front of the band. A mathematical model for the elution separation of  $^{15}\text{N}$  from  $^{14}\text{N}$  by NaOH (converting  $\text{NH}_4^+$  in the cation exchanger to  $\text{NH}_3$  in the solution), based on the experimental results of Spedding and co-workers (1955), was recently presented by Fujii and co-workers (1997).

### Band Displacement Process

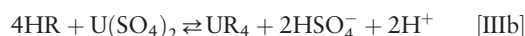
The band displacement process differs from the elution process (in which, too, the sorbed band is displaced) by providing for the self-sharpening of both edges of the band, so that the concentration

profile remains rectangular (see Figure 2). This is achieved by ensuring that the column ahead of the band has a counterion that is much less strongly sorbed than the isotopic ions and that the rear end of the band is displaced by ions that are much more strongly held. The first condition is met, for example, by the use of a cation exchanger in the hydrogen form and elution with a salt of a weak acid that removes the hydrogen ions effectively. The second condition is met by elution with a complexing agent combined with a counterion that is only weakly complexed, so that the strongly complexed isotopic ions travel down the column. Both conditions can, of course, be met with the same combination of reagents.

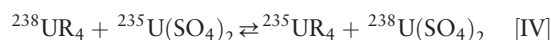
The work of Sakuma, Okamoto, and Kakihana is an example of the use of band displacement chromatography for the cation exchange separation of uranium isotopes, employing uranium(IV) sulfate and displacement with  $0.40 \text{ mol l}^{-1} \text{ MgSO}_4 + 0.02 \text{ mol l}^{-1} \text{ H}_2\text{SO}_4$ . At the rear end of the band, magnesium ions displace the uranium(IV) ions that form much stronger sulfate complexes as in reaction [IIIa]:



At the front edge of the band the uranium ions displace hydrogen ions that are bound as hydrogen sulfate anions as in reaction [IIIb]:



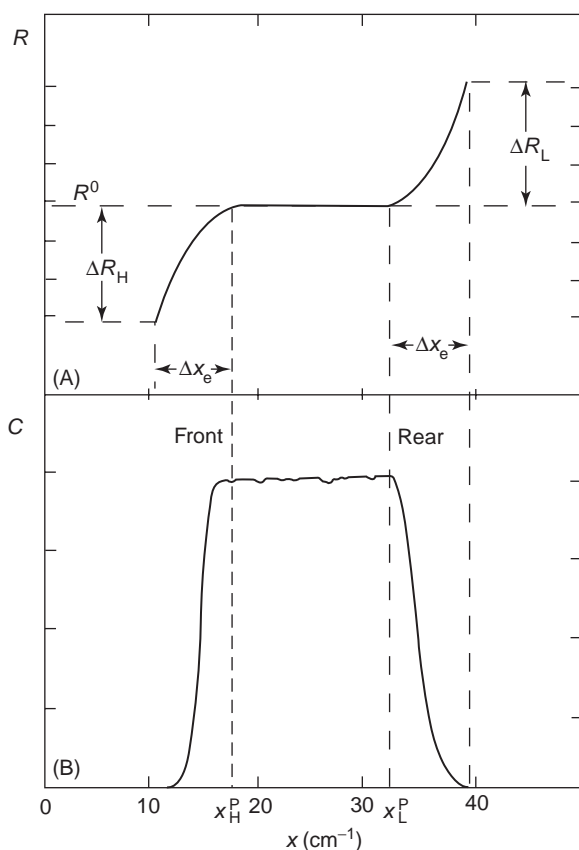
In both cases, the chemical association reactions work in the same direction for the isotopes as the cation-exchange reactions, resulting in an enhancement of the effect. The net isotope equilibrium is as in reaction [IV]:



and the rarer lighter isotope,  $^{235}\text{U}$ , is enriched at the rear edge of the band.

### Electromigration in a Column

The principle used for the counter-current ion-exchange isotope separation by electromigration in a column is the balancing of the hydraulic flow of the ions X, the isotopes of which are to be separated, by the opposite flow of ions Y in an electric field. The X ions should be the more strongly sorbed on the ion exchanger and the Y ions should be the more mobile ones in the field. A sharp boundary between the two kinds of ions is attained at some place in the column and is maintained there by the balancing of the flows. The steady passage of electric current is equivalent to movement of the sorbed band of X (which is actually



**Figure 2** Elution curve for self-sharpening band displacement of a di-isotopic solute ion on an ion exchange column. (A) The isotopic ratio  $R$  as a function of the distance  $x$  that the band was displaced. (B) Total concentration  $C$  of the ion as a function of  $x$ . For the other descriptors see the section on production of separated isotopes.

stationary) along the column, and this results in the separation of the isotopes.

The method was tested by Kobayashi, Fujii, and Okamoto in 1982 for the separation of X = calcium isotopes on a cation-exchange column with Y = lithium ions. The heavier calcium isotopes,  $^{42}\text{Ca}$ ,  $^{43}\text{Ca}$ ,  $^{44}\text{Ca}$ , and  $^{48}\text{Ca}$ , were enriched at the front of the sorbed band, and the lighter one,  $^{40}\text{Ca}$ , at the rear, when the band 'migrated' several meters.

### Electromigration in a Membrane

Electromigration experiments in a membrane did not use the counter-current technique; hence, the band moved along the membrane strip, and only short runs could be made. A longer membrane strip (up to 1 m), stored in the cathode compartment and slowly drawn out through the tetrachloromethane cooling bath and anode compartment, so as to keep the band stationary, constituted an improvement. As current is passed, the isotopes migrate at different rates that are proportional to the square root of the mass of the migrating particle: the isotopic ion together with its hydration shell. This shell is smaller in the ion-exchange membrane than in a homogeneous aqueous solution, yielding a small advantage to electromigration in the membrane. The preferred practice, which produces a sharp leading edge of the band, is to have a highly mobile ion in the membrane ahead of the band (hydrogen ions) and a slower ion in its rear.

Fujii and co-workers (1985) used such a membrane configuration to separate calcium isotopes with a migration length of 0.70 m for a  $0.5 \text{ mol l}^{-1}$   $\text{CaCl}_2$  feed solution, resulting in enrichment factors smaller than those achieved by electromigration on the column.

### The Elementary Separation Factor

In the following, the elementary separation factor,  $\varepsilon = \alpha - 1$ , pertains to the ratio of the lighter to the heavier isotopes, and is positive if the lighter one is enriched in the ion exchanger with respect to the feed solution, and negative when it is depleted. For the sake of comparison, values of  $10^4 \varepsilon$  are generally quoted, irrespective of the mass difference between the heavier and lighter isotopes. The values obtained for the absolute values,  $|10^4 \varepsilon|$ , range from 0.1 to 300 but are generally between 1 and 10. In the following,  $C^0$  is the total concentration of the isotopes in the feed solution,  $q$  is the amount sorbed per unit mass of exchanger,  $R$  is the isotopic ratio ( $R^0$  initially), the enrichment is  $R/R^0 = (C_L/C_H)/(C_L^0/C_H^0)$  for a volume  $V$  of eluting or displacing solution passed through the ion-exchange column of length  $L$ , having  $N$  equivalent theoretical plates of height  $H$  ( $H = \text{HETP} = L/N$ ).

### From Breakthrough or Elution Curves

Glueckauf and coworkers (1949) showed that for breakthrough experiments, the important quantities are the relative depletion  $\eta$  of the slower moving, more strongly sorbed, isotope (subscript s):

$$\eta = (C_s^0 - C_s)/C_s^0 \quad [3]$$

and its slope with respect to the elution volume  $V$ ,  $\eta' = (\partial\eta/\partial V)V_t$ , with  $V_t$  being the threshold (breakthrough) volume. The final expression is

$$\varepsilon = (\pi\eta^2/4V\eta')[1 + (\pi\eta/4) + (5\pi^2\eta^2/64)] \quad [4]$$

For elution experiments (Figure 1) Glueckauf showed (1958) that  $\varepsilon$  is obtained from the enrichment, if the number of equivalent theoretical plates  $N$  is large and known:

$$\ln(R/R^0) = N\varepsilon(1 - V/V_p) \quad [5]$$

where  $V_p$  is the mean (for the two isotopes) of the peak elution volume. If  $w$  is the width of the elution peak at  $1/e$  of its height, a good approximation of  $N$  is  $N \approx 8(V_p/w)^2$ . This method has been frequently employed in recent years.

### From Band Displacement Data

The elementary separation factor is obtained in band displacement experiments from data on fractions  $i$  in the leading and tailing parts of the displaced band when it is finally eluted (see Figure 2). Let  $f_i$  be the total amount of the isotopic ions in the  $i$ th fraction,  $R_i$  be the isotopic ratio in this fraction, and  $Q$  the total exchange capacity of the exchanger column. It is assumed that in the region of the sorbed band on the ion exchanger the total concentration of the ion in the exchanger and in the solution in equilibrium with it are constant, the only net effect being the counter-current movement of the isotopes. Then,

$$\varepsilon = \sum f_i(R_i - R_0)/QR_0(1 - R_0) \quad [6]$$

The value of  $\varepsilon$  should be independent of the number of fractions included in the sum in eqn [6], and whether the leading or tailing parts of the eluate are analyzed. This independence was found in practice.

### Variables Affecting the Separation Factor

The only way that isotopes of a given ion can be separated on an ion exchanger is by ensuring that the different isotopes have different properties in the exchanger phase and in the external-solution phase. The way to do this is by (1) causing the environments

in the exchanger and the solution to be very different, and even better by (2) causing different species of the ions to be present in the two phases.

The cross-linked ion-exchange resin has generally lower water content than the external solution, and the water is less structured in the exchanger. If the external solution is dilute in electrolytes, these differences are enhanced with increasing cross-linking of the resin. Further enhancement of the differences can be achieved by the use of mixed solvents. The effects of cross-linking (e.g.,  $10^4\epsilon$  rising from 6 to 38 for lithium as the cross-linking X of the PSS resin changed from 2% to 24% divinylbenzene (DVB)) and solvent composition indeed confirm that the isotopes are hydrated slightly differently.

It is noteworthy that under suitable conditions, if the electrolyte concentration in the external (eluting) solution is raised, reversals in the sign of  $\epsilon$  are observed. Such a reversal was observed for calcium isotopes,  $\epsilon$  changing from negative to positive as the concentration of the eluting acid, HCl or HClO<sub>4</sub>, increased from 2.5 to 9.0 mol l<sup>-1</sup> with a 12% DVB PSS resin. A reversal of the sign of  $\epsilon$  can also be achieved by suitable choice of the displacing counterion. For chlorine isotopes,  $\epsilon$  was positive for elution with nitrate and negative for elution with fluoride.

Less drastic effects on the magnitude of  $\epsilon$  than the reversal of its sign were observed in many cases on changes of the concentration of the eluent or the nature of the counterion. An increase in the temperature was generally accompanied with a decrease of  $|\epsilon|$ . This was explained by the lowered structure of the water in the external solution, making it more similar to the internal water in the exchanger. However, in some cases it is advantageous to work at an elevated temperature, in order to assure rapid diffusion and attainment of equilibrium, decreasing the HETP, even at the expense of the separation factor.

The dependence of  $|\epsilon|$  on the mass difference  $\Delta m$  was studied for several multi-isotopic elements and in general a proportionality of  $|\epsilon|$  to  $\Delta m$  was found. For instance,  $10^4\epsilon$  is nearly twice as large for <sup>40</sup>Ca/<sup>48</sup>Ca separation (8.7) than for <sup>40</sup>Ca/<sup>44</sup>Ca (4.7) and about four times larger for <sup>84</sup>Sr/<sup>88</sup>Sr separation ( $0.97 \pm 0.07$ ) than for <sup>87</sup>Sr/<sup>88</sup>Sr ( $0.21 \pm 0.04$ ).

## Production of Separated Isotopes

It is clear that the larger the elementary separation factor,  $\epsilon$ , and the number of equivalent theoretical plates,  $N$ , of the ion-exchange column are, the higher

is the isotope enrichment that can be achieved. As Kakihana had shown in 1980, there are some further factors that have to be considered for preparative-scale isotope separation (enrichment). Band displacement chromatography has proved to be the most effective method for isotope enrichment by ion exchange, hence this is considered in the following (Figure 2). Suppose that a sharp-edged band consisting of the two isotopes is formed on the column, with the front of the band enriched in the heavier isotope after some displacement. When its mole fraction at the front of the band reaches the value desired for the product,  $R_H^P$ , material is withdrawn from the band. Let the start-up time  $t_s$  be the time between the start of the displacement and the production stage, given approximately by

$$t_s = [1 + \epsilon(1 - R_H^0)]\Delta R_H / \eta \epsilon v_b k_e \quad [7]$$

where  $\Delta R_H = R_H^P - R_H^0$ ,  $\eta = R_H^0(1 - R_H^P)/(1 + \alpha_v/d)$ ,  $\alpha_v$  is the void fraction of the column,  $d = (1 - \alpha_v) C_{\text{resin}}/C_{\text{solution}}$ . Also,  $v_b$  is the rate of band displacement: if the linear velocity of the eluting solution is  $v_e$ , then  $v_b = v_e/(\alpha_v + d)$ . Finally,  $k_e$  is the 'slope coefficient', which is inversely proportional to the HETP, and equals the difference in the rate of displacement of the two isotopes divided by the difference in their diffusion coefficients:  $(v_H - v_L)/(D_H - D_L)$ . The slope coefficient should be maximized. The width of the enriched zone  $\Delta x_e$  (Figure 2) is approximately

$$\Delta x_e = (l/k_e) \ln(\Delta R_H/A_e) \quad [8]$$

where the 'enrichment coefficient'  $A_e$  is proportional to  $\exp(k_e \kappa_e t) \exp(k_e x_p)$ , and where  $\kappa_e = (v_H D_H - v_L D_L)/(D_H - D_L)$ , and  $x_p$  is the distance from the origin at which  $\Delta R_H$  starts to deviate from 0. The rate of growth of the enriched zone is given approximately by  $d\Delta x_e/dt = \eta \epsilon v_b / \Delta R_H$ . Finally, the maximal rate of production per unit time per unit band volume, if the product is removed continuously, is approximately

$$U^P = C^0 (d\Delta x_e/dt) 2\Delta x_e \quad [9]$$

As an example, the enrichment of <sup>10</sup>B from its initial  $R_L^0 = 0.198$  to a product mole fraction of  $R_L^P = 0.5$  for a feed concentration of  $C^0 = 1 \text{ mol l}^{-1}$  and  $10^4\epsilon = 10$  may be considered. For a band displacement rate of  $v_b = 10 \text{ cm h}^{-1}$  and a typical slope coefficient  $k_e = 0.1 \text{ cm}^{-1}$  the start-up time would be  $t_s \sim 10^3 \text{ h}$ , the width of the enriched zone would be  $\Delta x_e \sim 20 \text{ cm}$ , and its rate of growth  $d\Delta x_e/dt \sim 0.005 \text{ cm h}^{-1}$ . The maximal rate of production of the 50% enriched <sup>10</sup>B would be  $U^P \sim 1 \text{ kg m}^{-3} \text{ h}^{-1}$ .

In another analysis of the band displacement method by Obermoller and White in 1991, applied to the anion-exchange redox method for uranium isotope separation (see below), the optimal length of the displaced band was determined to be  $L = 1.313HN^{1/2}$ . The relative concentration of the enriched isotope in the product is then  $1 + 0.551\epsilon N^{1/2}$ .

## Isotopes of Individual Elements

### Overview

Isotope separation by ion exchange has been applied to many elements, as is seen in Table 1. Obviously, mainly elements that have at least two naturally occurring isotopes have been studied, although in a few cases radioactive isotopes (e.g., for sodium:  $^{22}\text{Na}$  and  $^{24}\text{Na}$ ) have been included in the studies. Most of the results are summarized in Table 2, with more detailed description being accorded to some of the more intensely studied elements.

### Alkali Metals

**Lithium** Many of the earlier studies dealt with lithium, reasoning that the large relative mass difference  $\Delta m/m$  of 1/7 should help the separation. It was later realized that  $\Delta m/m$  is much smaller, since it is the hydrated ions (with  $m \gg 7$ ) that play their roles in the solution and the exchanger according to reaction [I]. In all the cases studied,  $^6\text{Li}$  was enriched in the ion exchanger. With PSS cation exchangers and elution with strong acids (HCl) or their salts, only relatively low values of  $10^4\epsilon$ , up to 40, were achieved, with  $d\epsilon/dX > 0$  ( $X$  denoting the degree of cross-linking) and  $d\epsilon/dT < 0$ . When a zeolite or a weakly acid exchanger were used and the base LiOH or a salt of a weak acid, such as acetic acid, were employed for elution, making use of reaction [III], higher values of  $10^4\epsilon$  were achieved, up to 220 in favorable cases (e.g., presence of 20% acetone). With

the inorganic exchanger  $\text{HNbO}_3$  a value of  $10^4\epsilon = 430$  and with an azacrown resin even  $10^4\epsilon = 680$  were recently reported, but the elution was slow. Electromigration in a cation-exchange membrane led to  $d\epsilon/dT > 0$  and  $d\epsilon/di > 0$  ( $i$  denoting the current density),  $10^4\epsilon \sim 200$  being achieved at high  $i$  (Table 2). (Table 2 also provides information on sodium, potassium, and rubidium.)

### Alkaline Earth Metals

**Magnesium** Recent work by Kim, using either hydrous  $\text{MnO}_2$  or a Merrifield peptide anchored azacrown resin exchanger provided separation between  $^{24}\text{Mg}$ ,  $^{25}\text{Mg}$ , and  $^{26}\text{Mg}$  with  $10^4\epsilon/\Delta m$  of  $\sim 110$  (Table 2).

**Calcium** Many studies have been devoted to this element, since here again  $\Delta m/m$  values up to  $8/40 = 1/5$  were expected. The values of  $\epsilon$  were, indeed, proportional to  $\Delta m$ , but not so large. This is mainly because on elution from a PSS cation exchanger, where only equilibria [I] operate,  $d\epsilon/dC_{\text{eluent}} < 0$ , so that  $\epsilon = 0$  is passed when the concentration of the eluent is increased. Typical values of  $10^4\epsilon/\Delta m$  for dilute eluents ( $1.5 \text{ mol l}^{-1}$  HCl,  $\text{HNO}_3$ ,  $\text{HClO}_4$ , LiCl) are 0–0.6, for concentrated eluents ( $\sim 9$ – $12 \text{ mol l}^{-1}$ ) these values are  $-2.1$  to  $-3.0$ . Elution with a complexing eluent,  $\alpha$ -hydroxy isobutyrate, from a PSS resin, adding equilibria [III], did not produce remarkably large separations,  $10^4\epsilon/\Delta m = 1.2$ . Some use of a chelating cation exchange resin was made, yielding the highly negative  $10^4\epsilon/\Delta m = -10.2$ . Electromigration in a PSS cation-exchange column led to positive  $\epsilon$  values, increasing with the concentration of the  $\text{CaCl}_2$  feed solution, with  $10^4\epsilon/\Delta m = 1.8$  at a  $0.5 \text{ mol l}^{-1}$  feed. However, electromigration in a PSS membrane led to negative  $\epsilon$  values,  $10^4\epsilon/\Delta m = -1.26$  at  $20^\circ\text{C}$  with  $d\epsilon/dT < 0$  (Table 2). (Table 2 also provides information on beryllium and strontium.)

**Table 1** Periodic chart of the ion-forming elements, showing in bold the elements for which their natural isotopes were separated by ion exchange, in *italics* the elements for which radioactive isotopes were so separated, and in underline the elements that have only a single natural isotope. For many elements naturally occurring isotopes could still be separated by means of ion exchange as is evident

IA	IIA	IIIA	IVA	VA	VIA	VIIA	VIII			IB	IIB	IIIB	IVB	VB	VIB	VIIB
H																
Li	<i>Be</i>	<b>B</b>											<b>C</b>	<b>N</b>	<b>O</b>	<b>F</b>
Na	<b>Mg</b>	<u>Al</u>											Si	<u>P</u>	<b>S</b>	<b>Cl</b>
K	<b>Ca</b>	<u>Sc</u>	Ti	V	Cr	<u>Mn</u>	Fe	Co	Ni	<b>Cu</b>	Zn	Ga	Ge	<u>As</u>	Se	Br
Rb	<b>Sr</b>	<u>Y</u>	<b>Zr</b>	<u>Nb</u>	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	<b>Sn</b>	Sb	Te	<u>I</u>
Cs	Ba	La														
			Ce	<u>Pr</u>	<b>Nd</b>	Pm	<b>Sm</b>	<b>Eu</b>	<b>Gd</b>	<u>Tb</u>	Dy	<u>Ho</u>	Er	<u>Tm</u>	Yb	Lu
			Hf	<u>Ta</u>	W	Re	Os	Ir	Pt	<u>Au</u>	Hg	<u>Tl</u>	Pb	<u>Bi</u>	Po	At
Fr	Ra	Ac	<u>Th</u>	Pa	<i>U</i>	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr



**Table 2** Selected isotope separations with ion exchangers

Isotopes	Exchanger, method, conditions	$10^4 \varepsilon$	Remarks	Author (year)
$^6\text{Li}/^7\text{Li}$	Zeolite, bd	220		Taylor (1938)
	PSS, el, HCl aq acetone	>0		Davis (1956)
	PSS, X = 2–24, el, dil HCl	38 (X = 24)	$d\varepsilon/dX > 0$	Lee (1959)
	PSS, X = 24, el, acetate	26		Powell (1962)
	PSS, el, LiOH aq EtOH, acetone	>0	$d\varepsilon/dX > 0$	Kakahana (1962)
	WAC, Zr phosphate, LiOH	$\leq 220$		Kakahana (1962)
	PSS, X = 25, el, LiCl	120		Dickel (1964)
	PSS, X = 12, WAC, el, HCl, HOAc	>0		Lee (1965)
	PSS, membrane, em	100		Kakahana (1968)
	PSS, membrane, em	$\leq 200$	$d\varepsilon/di > 0$	Okamoto (1970)
	PSS, X = 16, bd, acetate, 20°C	30		Fujine (1983)
	PSS, membrane, em, 20°C	38	$d\varepsilon/dT < 0$	Klemm (1988)
	PSS, X = 8, el, dilute $\text{MCl}_n$	22–39	$d\varepsilon/dn > 0$	Kim (1991)
	PSS, bd, aq DMSO or acetone	15–25		Oi (1993)
	PSS, porous, X = 30, bt, 80°C	125		Araki (1998)
	Ti phosphate, el $0.2 \text{ mol l}^{-1} (\text{NH}_4)_2\text{CO}_3$	70		Ooi (1999)
	Inorganic exchangers, batch	430 ( $\text{HNbO}_3$ )		Kanzaki (2002)
	Azacrown-Merrifield peptide, el	680		Kim (1997, 2002)
$^{22}\text{Na}/^{24}\text{Na}$	PSS, X = 10, el, HCl, 25°C	1.4	$d\varepsilon/dT < 0$	Betts (1956)
	PSS, X = 1–12, bt	$\sim 0$		Ohtaki (1959)
$^{39}\text{K}/^{41}\text{K}$	Zeolite, el	<0		Taylor (1937)
	PSS, membrane, em	>0		Schmidhalter (1982)
	PSS, X = 8, el, crown ether/aq MeOH	–7.4	10°C	
$^{85}\text{Rb}/^{87}\text{Rb}$	PSS, membrane, em, 25°C	–9	$d\varepsilon/dT < 0$	Hosoe (1985)
	PSS, bd, 25°C	–5.2		Hosoe (1988)
	PSS, X = 12, el, HCl, $\text{HNO}_3$	18 (HCl)		Lee (1976)
$^7\text{Be}/^9\text{Be}$	PSS, X = 8, el, lactate	–1.6		Aaltonen (1971)
$^{25}\text{Mg}/^{26}\text{Mg}$	Azacrown-Merrifield peptide, el	120–220		Kim (2002)
$^{24}\text{Mg}/^{25,26}\text{Mg}$	Hydrous $\text{MnO}_2$ , el, $1 \text{ mol l}^{-1} \text{ NaCH}_3\text{CO}_2$	110–210		Kim (2002)
$^{40}\text{Ca}/^{45}\text{Ca}$	PSS, el, citrate, 95°C	<0	Retracted	Heumann (1972)
$^{40}\text{Ca}/^{44}\text{Ca}$	PSS, X = 8, el, $\alpha$ -hydroxyisobutyrate	–4.7		Aaltonen (1971)
$^{40}\text{Ca}/^{47}\text{Ca}$	PSS, X = 12, el, HCl	–2.6		Lee (1976)
$^{40}\text{Ca}/^{>40}\text{Ca}$	Chelating, el, $> 0.01 \text{ mol l}^{-1} \text{ HCl}$	>0		Heumann (1972)
$^{40}\text{Ca}/^{44,48}\text{Ca}$	PSS, el, HCl, $\text{HClO}_4$	>0, <0	$d\varepsilon/dC < 0$	Heumann (1979)
$^{40}\text{Ca}/^{44}\text{Ca}$	PSS, X = 8, el, $1.5 \text{ mol l}^{-1} \text{ HCl}$	2.1		Russell (1978)
$^{40}\text{Ca}/^{44,48}\text{Ca}$	PSS, X = 12, el, $1.5\text{--}9 \text{ mol l}^{-1} \text{ HNO}_3$	–12 ( $12 \text{ mol l}^{-1}$ )	$d\varepsilon/dC < 0$	Heumann (1979)
$^{40}\text{Ca}/^{48}\text{Ca}$	PSS, X = 12, el, $0.1\text{--}1.2 \text{ mol l}^{-1} \text{ BaCl}_2$	12.7 ( $0.1 \text{ mol l}^{-1}$ )	$d\varepsilon/dC < 0$	Heumann (1981)
$^{40}\text{Ca}/^{44}\text{Ca}$	Chelating, el	–41		Gutsykov (1981)
$^{40}\text{Ca}/^{48}\text{Ca}$	PSS, X = 12, bd, $3\text{--}12 \text{ mol l}^{-1} \text{ LiCl}$	–16.7 ( $12 \text{ mol l}^{-1}$ )	$d\varepsilon/dC < 0$	Heumann (1982)
$^{40}\text{Ca}/^{>40}\text{Ca}$	PSS, column, em, LiCl, 80°C	1.8 $\Delta m$	$d\varepsilon/dC > 0$	Kobayashi (1982)
$^{40}\text{Ca}/^{>40}\text{Ca}$	PSS, membrane, em, $\text{MgCl}_2$ , 20°C	–1.26 $\Delta m$	$d\varepsilon/dI$	Fujii (1985)
$^{84,87}\text{Sr}/^{88}\text{Sr}$	PSS, X = 8, el, $\alpha$ -hydroxyisobutyrate	0.97 ( $^{84}\text{Sr}$ )		Aaltonen (1972)
$^{88}\text{Y}/^{91}\text{Y}$	PSS, X = 8, el, $\alpha$ -hydroxyisobutyrate	2.3		
$^{142}\text{Nd}/^{150}\text{Nd}$	PSS, X = 8, el, $\alpha$ -hydroxyisobutyrate	2.7		Aaltonen (1967, 1971, 1974)
$^{144}\text{Sm}/^{154}\text{Sm}$	PSS, X = 8, el, $\alpha$ -hydroxyisobutyrate	1.9		
$^{151}\text{Eu}/^{153}\text{Eu}$	PSS, eex	–1.1 to –2.4		Ismail (1998)
	Not specified, eex	–3		Fujii (2002)
$^{154}\text{Gd}/^{160}\text{Gd}$	PSS, X = 8, el, $\alpha$ -hydroxyisobutyrate	4.2		Aaltonen (1967)
$^{154\text{--}158}\text{Gd}/^{160}\text{Gd}$	PSS, Cu-form, el, EDTA	0.25–0.49	$d\varepsilon/d\Delta m = 0.13$	Chen (1992)
$^{90}\text{Zr}/^{96}\text{Zr}$	PSS, bt, $\text{NH}_4$ citrate	$\leq 0.01$		Kogure (1983, 1985)
$^{56}\text{Fe}/^{57}\text{Fe}$	Not specified, eex	–2.5		Fujii (2002)
$^{56}\text{Co}/^{60}\text{Co}$	PSS, X = 8, el, $\alpha$ -hydroxyisobutyrate	2.75		Aaltonen (1967, 1971)
$^{59}\text{Co}/^{60}\text{Co}$	PSS, X = 12, el, dil. HCl	0.5		Lee (1976)
	SBA, X = 10, el, $4 \text{ mol l}^{-1} \text{ HCl}$	1.7		Lee (1976)
	PSS, bd, Mg-EDTA	8		Aaltonen (1961)
$^{63}\text{Cu}/^{65}\text{Cu}$	PSS, membrane, em	–2		Kakahana (1982)
	SBA, eex	3.8		Fujii (2002)
	SBA, eex, bd	15–22	$d\varepsilon/d\Delta m = 3.8$	Nakanishi (1996)
$^{232,234}\text{U}/^{238}\text{U}$	For studies < 1980 see Further Reading section			
$^{235}\text{U}/^{238}\text{U}$	PSS, bd, U(IV), $\text{H}_2\text{SO}_4$ , $\text{MgSO}_4$	0.5		Sakuma (1981)
	PSS, bd, U(IV), malate, lactate	0.7	$d\varepsilon/dT < 0$	Oi (1982)

Table 2 Continued

Isotopes	Exchanger, method, conditions	$10^4\epsilon$	Remarks	Author (year)
$^{112}\text{Sn}/^{112}\text{Sn}$	PSS, bd, U(VI), $\text{F}^-$ , acetate	-1.2 ( $\text{F}^-$ )		Tanaka (1981, 1982)
	PSS, bd, U(VI), carboxylates	-1	$d\epsilon/dT \sim 0$	Nakagawa (1983)
	PSS, bd, U(VI), carbonate	-3.4		Aoyama (1989)
	SBA, eex, HCl	58-251	$d\epsilon/d\Delta m = 25$	Najjar (1993)
$^1\text{H}/^2\text{H}$	SBA, batch, water	300		Fukutomi (1963)
$^{10}\text{B}/^{11}\text{B}$	SBA, WBA, bt, water, $25^\circ\text{C}$	110	$d\epsilon/dT \sim 0$	Kakahana (1977)
	WBA, bd, water, $40^\circ\text{C}$	100	256 m path	Sakuma (1980)
	WBA, el, water, $25^\circ\text{C}$	130	$d\epsilon/dC_B < 0$	Itoh (1985)
	SBA, $\text{F}^-$ , $\text{Cl}^-$ , $\text{Br}^-$ forms, el, water	100 ( $\text{F}^-$ )		Oi (1988)
$^{12}\text{C}/^{13}\text{C}$	SBA, $\text{F}^-$ and $\text{Cl}^-$ forms, bt	93-121 ( $\text{F}^-$ )	$d\epsilon/dT < 0$	Tsukamoto (1991)
	SBA, $\text{F}^-$ form			Kanzaki (2002)
	WBA, el, $\text{CO}_2$ aq, $10-60^\circ\text{C}$	-20		Bairanov (1981)
	$^{14}\text{N}/^{15}\text{N}$			
$^{32}\text{S}/^{34}\text{S}$	Zeolite, $\text{NH}_3$ aq, $\text{BaCl}_2$	<0		
	PSS, X = 12, el, $\text{NH}_3$ aq, NaCl	-257		Spedding (1955)
	PSS, X = 4-55, el, $\text{NH}_3$ aq/EtOH	-300		Kakahana (1963)
	WAC, el, $\text{NH}_3$ aq/acetone	-290		
	PSS, X = 4-12, $\text{Ni}^{2+}$ -form, bt, $\text{NH}_3$ aq	-70(X = 12)	$d\epsilon/dX < 0$	Gupta (1967)
	PSS, X = 9, el, $\text{NH}_3$ aq	<0		Bairanov (1971)
	PSS, $\text{Mg}^{2+}$ -form, el, $\text{NH}_3$ aq	-90		Kotaka (1978)
	PSS, X = 30, el, $0.11-1.14 \text{ mol l}^{-1}$ LiOH	250		Ohtsuka (1995)
	PSS, X = 8, porous, bd, $1 \text{ mol l}^{-1}$ NaOH	Simulated moving bed		Kruglov (1996)
	PSS, X = 8, porous, bd, $1 \text{ mol l}^{-1}$ NaOH	200	60 m path	Ohwaki (1998)
	PSS, X = 8, bd, $0.7 \text{ mol l}^{-1}$ NaOH	228	$d\epsilon/dT < 0$	Aguilera (2002)
	SBA, $\text{HSO}_3^-$	-100		Forberg (1958)
	$^{35}\text{Cl}/^{37}\text{Cl}$			
	SBA, X = 10, el, $0.01-1 \text{ mol l}^{-1}$ $\text{NaNO}_3$	5.4 ( $20^\circ\text{C}$ )	$0-60^\circ\text{C}$	Heumann (1977, 1979, 1980)
$^{35}\text{Cl}/^{37}\text{Cl}$	SBA, X = 10, el, $2.5-10 \text{ mol l}^{-1}$ KF	8.2 ( $2.5 \text{ mol l}^{-1}$ )	$d\epsilon/dC < 0$	
	SBA, X = 2-10, bd, $\text{NaNO}_3$ , $0^\circ\text{C}$	-6.3 (X = 10)	$d\epsilon/dX < 0$	
	$\text{ZrO}_2$ , el, $\text{NaNO}_3$	-6.1 ( $0.5 \text{ mol l}^{-1}$ )	$d\epsilon/dC < 0$	
	SBA, X = 10, bd, $0.1 \text{ mol l}^{-1}$ $\text{NaBF}_4$ , $\text{NaClO}_4$	-4.3 ( $\text{BF}_4^-$ )		Heumann (1980)

PSS, polystyrene sulfonate cation exchanger; WAC, weakly acid cation exchanger; SBA, strongly basic anion exchanger; WBA, weakly basic anion exchanger; bd, band displacement; el, band elution; eex, electron exchange (redox) process; em, electromigration; bt, breakthrough.

## Other Metals

**Yttrium, neodymium, samarium, europium, and gadolinium** See Table 2.

**Copper and iron** Electron exchange on an anion exchange resin between Cu(I) and Cu(II) for  $^{63}\text{Cu}/^{65}\text{Cu}$  separation and between Fe(II) and Fe(III) for  $^{56}\text{Fe}/^{57}\text{Fe}$  separation in chloride media were recently reported and yielded  $10^4\epsilon$  values of -3.8 and -2.5, respectively, at room temperature with  $|10^4\epsilon| > 0$ .

**Uranium** A great deal of effort has been accorded to the separation of the isotopes of uranium with ion exchangers. In 1983, Marcus summarized the results, and little further work has been done since then. Uranium(IV) showed very small enrichment of  $^{235}\text{U}$  in the resin, with  $10^4\epsilon < 1.0$  for both cation and anion exchangers and a variety of eluents. Uranium(VI), as the uranyl ion  $\text{UO}_2^{2+}$ , showed somewhat larger effects,  $^{235}\text{U}$  enriching in the solution. A good correlation was found between the enrichment factor and the asymmetric stretching frequency  $\nu_3$  of the

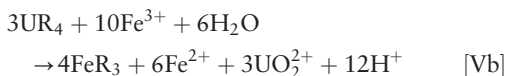
$\text{O}=\text{U}=\text{O}$  bonds:  $10^4\epsilon = -0.046[960 - (\nu_3/\text{cm}^{-1})]$ , leading to as high a value as  $10^4\epsilon = -3.4$  for carbonate as eluent, see Table 2.

The most promising procedure for the separation of the isotopes of uranium is to superimpose a redox reaction between U(IV) and U(VI) on the ion exchange and complexation reactions. This generally requires a catalyst for its implementation on a cation-exchange resin (with dilute HCl), but the medium used for the anion-exchange mode of operation,  $>4 \text{ mol l}^{-1}$  HCl, is itself the catalyst. In a cation exchanger, a band of green U(IV) is sorbed on the purple Ti(III)-form of the resin, and is displaced by the yellow Fe(III). At the front of the band the uranyl cation is reduced and the quadruply charged U(IV) displaces the doubly charged Ti(IV) from the resin fixed ion,  $\text{R}^-$  as in reaction [Va]:

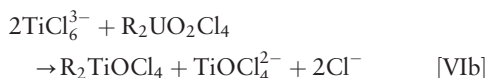
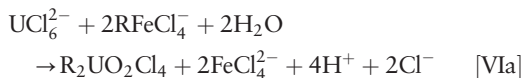


At the back of the band the U(IV) is re-oxidized to the uranyl ion (reaction [Vb]) that is displaced by the

better sorbed  $\text{Fe}^{3+}$ :



In an anion exchanger similar reactions take place, but with chloro-complexes of the titanium, iron, and uranium at the respective oxidation states participating: at the front reaction [VIa] and at the rear reaction [VIb]



With the cation exchanger  $10^4\epsilon = -7$  could be achieved in elution experiments, with the anion exchanger  $10^4\epsilon = 7$  in band displacement experiments. More recently this technique was employed for the separation of  $^{232}\text{U}$  and  $^{234}\text{U}$  from  $^{238}\text{U}$  with  $10^4\epsilon/\Delta m$  of  $\sim 4$ . See Table 2 for further information.

Electromigration experiments yielded disappointing results: with a cation-exchange membrane  $10^4\epsilon = -0.8$  was the maximal effect obtained. (Table 2 also provides information on zirconium, cobalt, and tin.)

### Nonmetals

**Boron** Boric acid is sorbed on an anion-exchange resin, whether weakly basic (in free base form) or strongly basic (in halide, X, preferably fluoride, form), with  $^{10}\text{B}$  enriched in the sorbed  $\text{B}(\text{OH})_4^-$  or  $\text{BX}(\text{OH})_3^-$  and  $^{11}\text{B}$  in the aqueous  $\text{H}_3\text{BO}_3$  species. Elution can take place with water, so that no regeneration is required. The separation factor depends only slightly on the resin basicity ( $10^4\epsilon = 130$  for a weak base resin, 100 for F-form strong base resin at room temperature for a  $0.1 \text{ mol l}^{-1}$  feed solution) and the temperature, but  $d\epsilon/dC_{\text{boron}} < 0$ . No isotope enrichment was found with a cation-exchange resin.

**Nitrogen** Experiments by Spedding in 1955 showed that for an X-12 PSS cation exchanger the equilibrium between aqueous ammonia and ammonium ions yielded  $10^4\epsilon = -257$  for  $^{14}\text{N}$  relative to  $^{15}\text{N}$ . Continued interest in this separation has not provided for substantially larger  $|10^4\epsilon|$  values in more recent years, although an X=30 porous PSS resin was employed. Large negative values were also obtained when the solution contained acetone or ethanol, and a small positive value of  $d\epsilon/dX$  was observed. Smaller values of  $10^4\epsilon$  ( $-70$  and  $-90$ ) were obtained when the ammonia was equilibrated not with the proton of the strongly acidic exchanger (to form ammonium ions) but with doubly charged metal ions,  $\text{Ni}^{2+}$  and  $\text{Mg}^{2+}$  (to form the hexamine complexes); see Table 2. (Table 2 also provides information on hydrogen, carbon, and sulfur and chlorine.)

See also: **Drug Metabolism: Isotope Studies. Isotope Dilution Analysis. Isotope Ratio Measurements. Liquid Chromatography: Isotope Separations.**

### Further Reading

- Aoki E, Kai T, and Fujii Y (1997) Theoretical analysis of separating nitrogen isotopes by ion-exchange. *Journal of Nuclear Science and Technology* 34: 277–282.
- Araki H, Umeda M, Enokida Y, and Yamamoto I (1998) Effect of development speed on separative performance of displacement chromatography using porous ion exchange resin for Li isotope separation. *Fusion Engineering and Design* 39–40: 1009–1013.
- Fujii Y, Aida M, Okamoto M, and Oi T (1985) A theoretical study of isotope separation by displacement chromatography. *Separation Science and Technology* 20: 377–392.
- Kakihana H (1980) Estimation of isotope separation power of chromatography. *Separation Science and Technology* 15: 567–585.
- Marcus Y (1983) Uranium isotope separation by ion exchange. *Gmelin Handbook of Inorganic Chemistry. Uranium Supplement*, vol. D4. Berlin: Springer.
- Obermoller HR and White DA (1991) Isotope separation by chemical exchange: A computer simulation. *Nuclear Technology* 96: 337–345.

# ION-SELECTIVE ELECTRODES

Contents

**Overview**

**Glass**

**Solid-State**

**Liquid Membrane**

**Gas Sensing Probes**

**Enzyme Electrodes**

**Clinical Applications**

**Food Applications**

**Water Applications**

## Overview

**K Štulík**, Charles University, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Ion-selective electrodes (ISEs) are electroanalytical sensors whose signals depend on the activities of ions in solution and exhibit a certain degree of selectivity for particular ionic species. The operation of classical ISEs is based on direct measurement of a single membrane potential at zero net current. At present, ISEs also include electrodes containing two or more membranes and/or highly selective biochemical systems, and semiconductor sensors. The range of chemical species sensed is thus greatly broadened, involving various nonionic compounds and some gases. The selectivity is also improved.

The history of ISEs is quite long – almost one century – and involves three distinct stages: 50 years of gradual development of the glass pH electrode; the period of rapid development of various types of ISE starting in the late 1950s and ending in the beginning of the 1980s; the present period during which the field of ISEs is being perfected technically with emphasis on miniaturization and the range of highly specialized ISEs is being extended (primarily those selective for biologically, medically, and environmentally important analytes). However, no qualitatively new principle has recently been proposed. The theoretical background has been developed gradually (by, e.g., Nikolskii and Eisenman) and its present state was completed during the 1970s. At present, efforts are primarily directed toward gathering of highly scattered information and unification and standardization

of experimental approaches and routine analytical procedures, as reflected, e.g., in the activities of the International Union of Pure and Applied Chemistry.

The oldest and still the best and most widely used ISE is the glass pH electrode invented as early as in 1906 by Cremer and in 1909 by Haber and Klemensiewicz. It became a standard laboratory tool during the 1930s. Early attempts at constructing glass electrodes sensitive to ions other than hydroxonium ion and electrodes with membranes of materials other than glass were unsuccessful. Only in the late 1950s did practically useful glass electrodes appear that were sensitive to the alkali metal, ammonium, and silver ions; however, they are no longer used. Around 1960, Pungor and co-workers developed electrodes with membranes of silver halide dispersed in a silicone rubber matrix and sensitive to the corresponding halides. Frant and Ross described the highly successful fluoride ISE with a lanthanum fluoride single-crystal membrane in 1966. The first ISE with a liquid membrane appeared in 1967 (Ross).

The versatility of ISEs was enhanced considerably by the introduction of membranes containing neutral ion carriers (ionophores). The first ISE of this type, with a membrane containing valinomycin and selective for potassium ions, was described by Štefanac and Simon in 1966. There are many liquid chemical systems that interact highly selectively with ions through, e.g., ion exchange, ion association, or solvent extraction. Practically useful ISEs based on these systems and on neutral ionophores have been obtained due to the gradual perfection of the technology of plasticized poly(vinyl chloride) (PVC) matrix membranes.

If a substance that is not sensed by any ISE undergoes a selective (bio)chemical reaction producing or consuming suitable ions, then it can be determined indirectly by monitoring these ions with an ISE. The selective (bio)chemical system is usually directly

attached to the ISE membrane. In this way, sensors have been obtained for acidic gases. A great variety of enzymatic and immunochemical sensors for many important organic compounds have been developed on this principle.

The membrane potentials can also be measured indirectly, if the membrane is fixed on the surface of a field-effect transistor (FET) from which the metallic gate has been removed. The electric current passing between the source and drain of the FET is then controlled by the membrane potential. The first sensor of this type [ion-selective FET (ISFET) or chemically sensitive FET (CHEMFET, CSFET)] was described by Bergveld in 1970 and the field has been developed primarily by Janata and co-workers and Japanese researchers.

## Principles

At the interface of two phases that contain electrically charged species (ions, electrons, holes, or dipoles), a potential difference develops. In equilibrium, the electrochemical potentials,  $\tilde{\mu}_i$ , of phases 1 and 2 are equal for all the charged species present. Hence, for the  $i$ th species,

$$\tilde{\mu}_i(1) = \tilde{\mu}_i(2) \quad [1]$$

Assuming that the chemical properties of the phases are independent of the electrical charges present, which is justified in work with ISEs, eqn [1] can be rewritten in the form:

$$\begin{aligned} \mu_i^\circ(1) + RT \ln a_i(1) + z_i F \varphi(1) \\ = \mu_i^\circ(2) + RT \ln a_i(2) + z_i F \varphi(2) \end{aligned} \quad [2]$$

where  $\mu^\circ$  are standard chemical potentials,  $a_i$  is the activity of the  $i$ th species that carries  $z_i$  elementary charges,  $\varphi$  is the internal potential of the phase, and the other symbols have their usual significance. The potential difference at the interface is thus given by

$$\begin{aligned} \Delta\varphi = \varphi(2) - \varphi(1) &= \frac{\mu^\circ(1) - \mu^\circ(2)}{z_i F} + \frac{RT}{z_i F} \ln \frac{a_i(1)}{a_i(2)} \\ &= \text{const.} + \frac{RT}{z_i F} \ln \frac{a_i(1)}{a_i(2)} \end{aligned} \quad [3]$$

Therefore, when the activity of the  $i$ th component in one phase is known or is, at least, constant, then the activity of the  $i$ th component in the other phase can be found from the  $\Delta\varphi$  value.

The  $\Delta\varphi$  value cannot be measured directly. However, a phase that selectively exchanges charged species with solution on the basis of ion exchange, ion association, complexation, solvent extraction, adsorption, etc., can be placed in the form of an electrochemical membrane between the sample solution and a standard solution.

Two reference electrodes can then be immersed in the two solutions and the electromotive force (e.m.f.) measured in this electrochemical cell. The measured value consists of liquid junction potential(s), which can be maintained constant within a certain experimental error by maintaining a constant, relatively high ionic strength and a suitable composition of the solutions, and the interfacial potential differences at the sides of the membrane adjacent to the sample and standard solutions. As the potential difference between the membrane and the standard solution is constant, the overall change in the e.m.f. corresponds to the change in the composition of the sample solution.

The ISE membrane thus behaves as a permselective membrane and the signal measured is given by the Donnan potential, which depends on the selectivity of the interfacial interaction, given by the equilibrium constants of the interactions of the membrane with the analyte(s) and interferents. Moreover, a diffusion potential develops across the membrane, as charge is passed on the passage of ions between the solution and the membrane.

Therefore, the voltage value measured in a cell containing an ISE can be expressed in terms of the Nikolskii equation:

$$E = E_0 \pm \frac{RT}{z_i F} \ln(a_1 + K_{12}^{\text{Pot}} a_2) \quad [4]$$

where  $E_0$  comprises all the constant terms in the cell voltage,  $a_1$  is the activity of the analyte, and  $z_i$  its charge,  $a_2$  is the activity of an interferent and  $K_{12}^{\text{Pot}}$  is the coefficient of selectivity for determination of analyte 1 in the presence of interferent 2.

The coefficient of selectivity involves the effect of the equilibrium constant,  $K_{\text{eq}}$ , of the selective interaction of the analyte with the membrane, and of the mobilities,  $U$ , of the analyte and the interferent within the membrane. There are two important classes of membrane: (1) those with fixed exchange sites, i.e., solid ion exchangers or solids with fixed adsorption sites; (2) membranes with mobile exchange sites, i.e., liquids containing ion-exchange sites (liquid ion exchangers), liquids with neutral ion carriers, solvent extraction systems, liquids containing ion association agents, etc. For group 1, the selectivity coefficient is

$$K_{12}^{\text{Pot}} = \frac{U_2}{U_1} K_{\text{eq}} \quad [5]$$

For group 2, the situation is more complicated and two limiting cases can be distinguished: (a) with completely dissociated systems of mobile exchange sites,

$$K_{12}^{\text{Pot}} = \frac{U_2 D_2}{U_1 D_1} \quad [6]$$



where  $D$  is the distribution coefficient between the solution and the membrane; (b) with strongly associated systems the equation

$$K_{12}^{\text{Pot}} = \frac{U_{2S}D_2}{U_{1S}D_1} \quad [7]$$

where  $U_{1S}$  and  $U_{2S}$  are the mobilities of the associates, holds for highly mobile sites and the equation

$$K_{12}^{\text{Pot}} = \frac{(U_2 + U_S)D_2}{(U_1 + U_S)D_1} \quad [8]$$

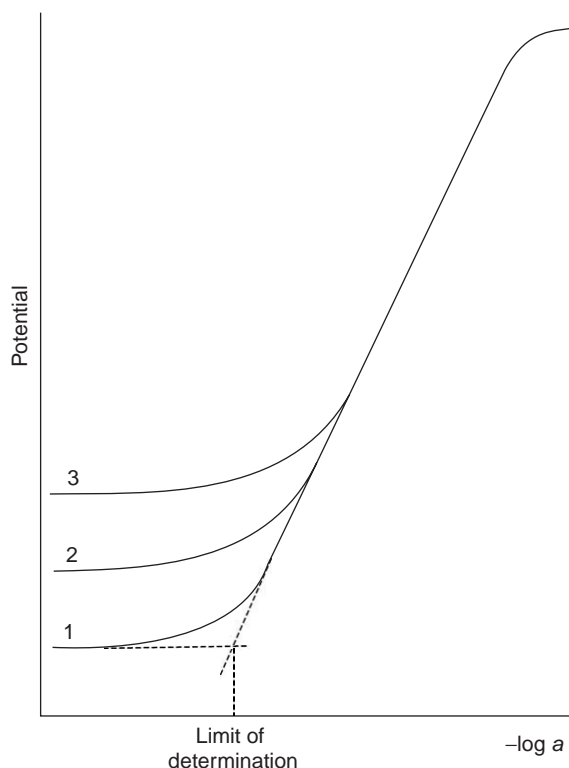
where  $U_S$  is the solvent mobility, holds for exchange sites of low mobility.

With neutral ion carriers, the coefficient of selectivity is given by

$$K_{12}^{\text{Pot}} = \frac{U_{2S^+}D_{2S^+}K_{2S^+}}{U_{1S^+}D_{1S^+}K_{1S^+}} \quad [9]$$

where  $U$  are the mobilities of the complexes,  $D$  their distribution coefficients, and  $K$  their stability constants.

The dependence of the ISE potential on the analyte activity thus has the shape depicted in Figure 1. Some typical values of the selectivity coefficient are given in Table 1.



**Figure 1** Typical ISE calibration curves: 1, curve in the absence of the interferents; 2 and 3, curves in the presence of the interferents (curve 3 corresponds to a higher interferent concentration).

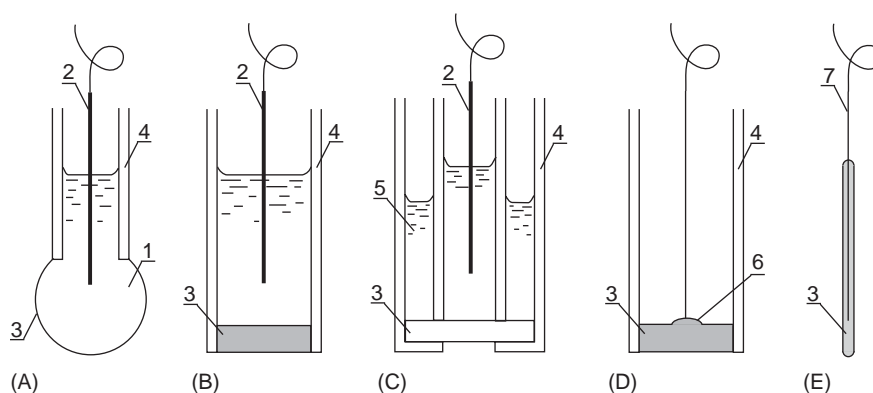
## Basic Electrode Types

As indicated in the previous section, there are two large groups of ISEs, those with fixed ion-exchange sites and the other with mobile exchange sites, irrespective of the kind of interaction underlying the exchange of ions.

The first group includes ISEs with solid membranes, consisting of glass, single crystals, polycrystalline materials, or a matrix of polymer containing particles of a polycrystalline active materials

**Table 1** Some important values of the selectivity coefficient

Analyte	ISE membrane type	Interferents	$\log K_{12}^{\text{Pot}}$
$\text{H}^+$	Glass	$\text{Na}^+$	$\sim -12$
$\text{Cl}^-$	$\text{AgCl}$ precipitate	$\text{CrO}_4^{2-}$	$-4.3$
		$\text{AsO}_4^{3-}$	$-3.5$
		$\text{CrO}_4^{3-}$	$-4.3$
		$\text{CO}_3^{2-}$	$-4.3$
		$\text{PO}_4^{3-}$	$-4.3$
$\text{Br}^-$	$\text{AgBr}/\text{Ag}_2\text{S}$	$\text{Cl}^-$	$-3.0$
		$\text{I}^-$	$+3.0$
		$\text{SCN}^-$	$-3.0$
		$\text{S}^{2-}$	$+10.0$
$\text{I}^-$	$\text{AgI}$ precipitate	$\text{Br}^-$	$-4.0$
		$\text{Cl}^-$	$-5.4$
		$\text{F}^-$	$-5.3$
		$\text{SCN}^-$	$0.0$
		$\text{SO}_4^{2-}$	$-5.4$
		$\text{NO}_3^-$	$-5.0$
$\text{S}^{2-}$	$\text{Ag}_2\text{S}$	$\text{CN}^-$	$-3.2$
		$\text{I}^-$	$-7.8$
		Other	$-10.7$
$\text{F}^-$	$\text{LaF}_3$ single crystal	$\text{OH}^-$	$-1.0$
		Other	$-3.0$
$\text{NO}_3^-$	$\alpha$ -Phenanthroline (substituted), complex with $\text{Ni}^{2+}$	$\text{ClO}_4^-$	$+3.0$
		$\text{I}^-$	$+1.3$
		$\text{ClO}_3^-$	$+0.3$
		$\text{Br}^-$	$-0.9$
		$\text{SH}^-$	$-1.4$
		$\text{NO}_2^-$	$-1.4$
		$\text{CN}^-$	$-2.0$
		$\text{HCO}_3^-$	$-2.0$
$\text{NH}_4^+$	Nonactin/monactin	$\text{Li}^+$	$-2.4$
		$\text{Na}^+$	$-2.7$
		$\text{K}^+$	$-0.9$
		$\text{Rb}^+$	$-2.4$
		$\text{Cs}^+$	$-2.3$
		$\text{Ca}^{2+}$	$-3.8$
		$\text{H}^+$	$-1.8$
$\text{K}^+$	Valinomycin + quaternary ammonium salt in dioctylphthalate	$\text{Li}^+$	$-4.0$
		$\text{Na}^+$	$-3.5$
		$\text{Rb}^+$	$+0.5$
		$\text{Cs}^+$	$-0.4$
		$\text{Mg}^{2+}$	$-5.1$
		$\text{Ca}^{2+}$	$-4.4$
		$\text{H}^+$	$-4.4$
$\text{Ca}^{2+}$	An acyclic, lipophilic neutral ionophore	$\text{Na}^+$	$-5.0$
		$\text{K}^+$	$-5.2$
		$\text{Mg}^{2+}$	$-5.1$
		$\text{H}^+$	$0.0$



**Figure 2** Basic types of ISE: (A) glass electrode; (B) electrode with a solid homogeneous or heterogeneous membrane; (C) classical liquid membrane electrode; (D) electrode without internal solution (all-solid-state electrode); (E) coated-wire electrode. 1, Internal standard solution; 2, internal reference electrode (Ag/AgCl); 3, membrane; 4, glass or plastic body of the electrode; 5, reservoir of the electroactive substance solution; 6, solid-state contact; and 7, metal wire.

(Figures 2A and 2B). The single-crystal membranes (e.g.,  $\text{LaF}_3$  doped with europium(II) in the fluoride ISE) tend to be expensive. Moreover, there are not many substances whose single crystals are easy to prepare. Pressed polycrystalline materials, sometimes containing two or more salts (e.g., a mixture of  $\text{Ag}_2\text{S}$  and  $\text{AgI}$  in an iodide ISE) and possibly also a binder, are most common. Membranes with polycrystalline particles dispersed in a polymeric matrix (such as Pungor's halide electrodes with silver halide precipitates in a silicone rubber matrix) have virtually disappeared now, primarily because they exhibit poorer dynamic properties than the former types.

The other group comprises ISEs with liquid membranes. The classical type (Figure 2C) is no longer used because of technical limitations (imperfect filling of the pores of the diaphragm with the liquid membrane material, leading to short-circuiting of the sample and internal solution, bleeding of the membrane material from the diaphragm). At present, PVC membranes are used, containing the active substance and usually a suitable plasticizer that may play an active role in the ISE response.

In both groups of ISE, the internal standard solution can be replaced by a solid-state contact, e.g., a simple metal contact (solid-state ISEs) (Figures 2D and 2E). This has certain practical advantages, as the electrodes are more robust, can be used in any position (e.g., upside down when analyzing a single drop of a sample solution placed over the ISE membrane), and, in the case of coated-wire ISEs (Figure 2E), are easy to prepare, just by dipping a wire in the membrane solution and allowing the solvent to evaporate. On the other hand, the potential difference between the membrane and the metal is poorly defined and the absolute ISE potentials are often poorly reproducible. These electrodes can readily be used for end-point

**Table 2** Some practically important ISEs

Analyte	Membrane
$\text{H}^+$	Solid, homogeneous, glass
$\text{Cl}^-$	Solid, AgCl precipitate
$\text{Br}^-$	Solid, AgBr/ $\text{Ag}_2\text{S}$ precipitate
$\text{I}^-$	Solid, AgI precipitate
$\text{S}^{2-}$	Solid, $\text{Ag}_2\text{S}$ precipitate
$\text{F}^-$	Solid, homogeneous, $\text{LaF}_3$ single-crystal doped with Eu(II)
$\text{Cd}^{2+}$	Solid, $\text{CdS}/\text{Ag}_2\text{S}$ precipitate
$\text{Pb}^{2+}$	Solid, $\text{PbS}/\text{Ag}_2\text{S}$ precipitate
$\text{Cu}^{2+}$	Solid, $\text{CuS}/\text{Ag}_2\text{S}$ precipitate
$\text{NO}_3^-$	Liquid, plasticized PVC matrix, complex substituted $\alpha$ -phenanthroline with $\text{Ni}^{2+}$
$\text{NH}_4^+$	Liquid, plasticized PVC matrix, nonactin/monactin
$\text{K}^+$	Liquid, plasticized PVC matrix, valinomycin and quaternary ammonium salt in dioctylphthalate
$\text{Ca}^{2+}$	Liquid, plasticized PVC matrix, an acyclic, lipophilic neutral ionophore

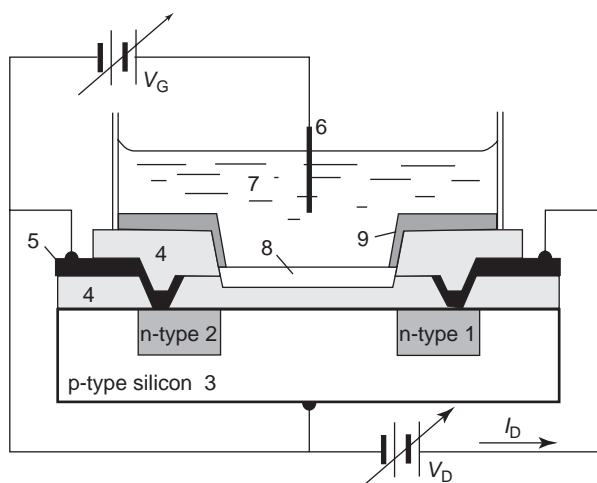
detection in titrations; for direct measurements, it is necessary to calibrate them frequently.

The most common classical ISEs are summarized in Table 2.

## Gas Sensors, Biosensors, and ISFETs

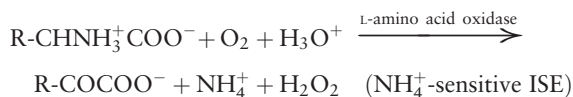
Gas sensors and biosensors are obtained by fixing an auxiliary chemical or biochemical system over the ISE membrane. The analyte reacts with the auxiliary system with production or consumption of the ion that is sensed by the ISE. Two basic types of gas sensor, the Severinghaus electrode and the 'air-gap' electrode, are described elsewhere in this encyclopedia.

Biosensors consist of a common ISE, on the membrane of which an enzyme, a thin layer of tissue containing a suitable enzyme, a culture of

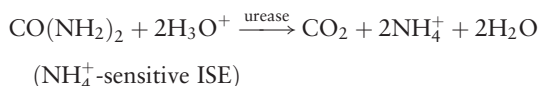


**Figure 3** Schematic diagram of an ISFET: 1, drain; 2, source; 3, substrate; 4, insulator; 5, metal lead; 6, reference electrode; 7, sample solution; 8, ion-selective membrane; and 9, encapsulant. 6, 7 and 8 replace the original metal gate of the FET. (Reprinted with permission from Janata J and Huber RJ (1979) *Ion-selective field effect transistors*. *Ion-selective Electrode Reviews* 1: 31–79; © Elsevier.)

microorganisms, cells, or organelles, is suitably fixed (e.g., in a plastic, porous matrix). Immunobiological systems (antigens or antibodies) can also be used. Examples of such sensors include:



which represents a sensor for L-amino acids, or



which is a urea sensor.

The principle of the ISFET (CHEMFET, CSFET) follows from **Figure 3**. The ion-selective membrane, replacing the metal gate, controls the current between the source and drain, which is measured. The main advantages of ISFETs involve small size and the possibility of placing several selective sensors (membranes) on a single chip. Furthermore, the response is rapid (in milliseconds, compared with classical ISEs whose response amounts to tens of seconds, and gas sensors and biosensors responding within a few minutes) and the impedance is low (compared with classical ISEs whose impedances are usually greater than 1 MΩ). On the other hand, the signal is not sufficiently stable (frequent recalibration measurements are necessary) and technical problems may be involved in producing a selective membrane on the chip surface and encapsulating the chip in a suitable insulator.

## Applications

The application range of ISEs is governed by two major features:

1. the apparatus is cheap and simple; measurements are simple and rapid and
2. the activities of ions, not their overall analytical concentrations, are measured.

Hence, ISEs find use primarily in simple determinations, in which a precision corresponding to a relative standard deviation of a few percent and a typical limit of determination of  $\sim 10^{-5} \text{ mol l}^{-1}$  are satisfactory and the sensor selectivity is sufficiently high, so that no sophisticated separation step is necessary. The above analytical parameters can be improved when ISEs are employed for end-point detection in titrations.

ISEs are often used to advantage for continuous monitoring and are very suitable as detection devices in flow injection analysis. However, they are not particularly convenient for detection in liquid chromatography or capillary electrophoresis, because of rather sluggish response, especially at low analyte concentrations, and a limited sensitivity of measurement.

The fact that the ion activities are measured rather than concentrations is ambivalent from the point of view of practical measurements. It is a great advantage for speciation purposes and for study of acid–base complexation and precipitation equilibria in solution. On the other hand, it is a drawback when the total content of an analyte in a sample is to be found, as great care must be taken to compensate or correct for interactions of the analyte with the sample matrix and the ambient atmosphere during calibration and measurement itself (cf. the problem of complexation of fluoride with iron(III) and aluminum(III) ions in analyses of natural waters, or oxidation of sulfide by atmospheric oxygen).

When an ISE is to be calibrated in terms of activity, conventional activity scales must be employed, which places great demands on the correct selection of the initial assumptions, the accuracy and precision of measurement, and introduces another element of uncertainty to the analytical results.

A large number of ISEs for a variety of analytes has been described in the literature. Only a few of them have found real routine use (**Table 2**). The glass pH ISE is unrivalled; in fact, the overwhelming majority of pH determinations in all fields of human activity is done potentiometrically with the glass electrode. Moreover, the pH ISE is very often used in more complex sensors, such as the Severinghaus-type gas probes and enzyme sensors.

The fluoride ISE is very popular, because it is highly selective and reliable, the determination of

fluoride is very important in water treatment and in environmental protection, and there are not many alternative analytical methods available. Potassium and calcium ISEs are very valuable in clinical analysis, as they determine the free ions in physiological samples, while spectral methods provide the total concentrations of the metals.

The sulfide ISE is sensitive, reliable, and useful especially in analyses of the atmosphere and waters. The nitrate ISE does not exhibit particularly good analytical properties, but it enables very fast and simple orientative determinations of nitrate in, e.g., waters, vegetables, and foodstuffs, which is welcome with regard to public hygiene. ISEs for various inorganic anions have somewhat lost their importance in competition with ion chromatography and those for inorganic cations often cannot compete with spectral methods; nevertheless, ISEs for copper(II), lead(II), and cadmium(II) ions are sometimes useful for end-point detection in complexometric titrations.

There are many biosensors based on ISEs, e.g., for urea, amino acids, and a wide range of biologically important substances, but they are mostly useful in highly specialized determinations. Ion-selective FETs and similar sensors should be very useful owing to their general advantages (see the previous section), but their wider routine use has so far been hampered by their drawbacks, which are rather poor signal stability and technical problems with the chip insulation. In competition with voltammetric sensors, ISFETs often lose because of their smaller versatility. On the other hand, the Severinghaus-type gas sensors have been practically very useful, especially in monitoring environmentally important gases, e.g., sulfur and nitrogen oxides, ammonia, and other gases.

**See also:** **Elemental Speciation:** Overview. **Ion-Selective Electrodes:** Liquid Membrane; Gas Sensing Probes. **pH. Sensors:** Overview; Amperometric Oxygen Sensors;

Calorimetric/Enthalpimetric; Chemically Modified Electrodes; Microorganism-Based; Photometric; Piezoelectric Resonators; Tissue-Based.

## Further Reading

- Bailey PL (1980) *Analysis with Ion-selective Electrodes*, 2nd edn. London: Heyden.
- Cammann K (1979) *Working with Ion-selective Electrodes*. Berlin: Springer.
- Freiser H (ed.) (1978) *Ion-selective Electrodes in Analytical Chemistry*, vol. 1. New York: Plenum.
- Freiser H (ed.) (1980) *Ion-selective Electrodes in Analytical Chemistry*, vol. 2. New York: Plenum.
- Janata J (1989) *Principles of Chemical Sensors*. New York: Plenum.
- Koryta J (ed.) (1980) *Medical and Biological Applications of Electrochemical Devices*. Chichester: Wiley.
- Koryta J and Štulík K (1983) *Ion-selective Electrodes*. Cambridge: Cambridge University Press.
- Morf WE (1981) *The Principles of Ion-selective Electrodes and of Membrane Transport*. Amsterdam: Elsevier.
- Oehme F (1991) *Liquid Electrolyte Sensors: Potentiometry, Amperometry, and Conductometry*. In: Göpel W, Hesse J, and Zemel JN (eds.) *Sensors. A Comprehensive Survey*, vol. 2, ch. 7. Weinheim: Wiley-VCH.
- Umezawa Y (ed.) (1990) *Handbook of Ion-selective Electrodes: Selectivity Coefficients*. Boca Raton, FL: CRC Press.
- Umezawa Y, Bühlmann P, Umezawa K, Tohda K, and Amemiya S (2000) Potentiometric selectivity coefficients of ion-selective electrodes. Part I. Inorganic cations. *IUPAC, Pure and Applied Chemistry* 72: 1851–2082.
- Umezawa Y, Umezawa K, Bühlmann P, Hamada N, Aoki H, Nakanishi J, Moritoshi S, Xia KP, and Nishimura Y (2002) Potentiometric selectivity coefficients of ion-selective electrodes. Part II. Inorganic anions. *IUPAC, Pure and Applied Chemistry* 74: 923–994.
- Umezawa Y, Bühlmann P, Umezawa K, and Hamada N (2002) Potentiometric selectivity coefficients of ion-selective electrodes. Part III. Organic ions. *IUPAC, Pure and Applied Chemistry* 74: 995–1099.

## Glass

**S Głąb and A Hulanicki**, University of Warsaw, Warsaw, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Glass electrodes sensitive to hydrogen ions are the most commonly used sensors in chemistry and related disciplines. They belong to the group of potentiometric

membrane sensors and are constructed in various configurations, depending on the application. Their basic design and properties are described and discussed in this article. Among their important properties is the selectivity, which depends on the composition of the glass. Glass electrodes are used mainly for pH measurements, but they may compose a part of more sophisticated systems in gas sensors or enzymatic sensors. Recently glasses based on a nonsilica structure have been used in potentiometric measurements.

fluoride is very important in water treatment and in environmental protection, and there are not many alternative analytical methods available. Potassium and calcium ISEs are very valuable in clinical analysis, as they determine the free ions in physiological samples, while spectral methods provide the total concentrations of the metals.

The sulfide ISE is sensitive, reliable, and useful especially in analyses of the atmosphere and waters. The nitrate ISE does not exhibit particularly good analytical properties, but it enables very fast and simple orientative determinations of nitrate in, e.g., waters, vegetables, and foodstuffs, which is welcome with regard to public hygiene. ISEs for various inorganic anions have somewhat lost their importance in competition with ion chromatography and those for inorganic cations often cannot compete with spectral methods; nevertheless, ISEs for copper(II), lead(II), and cadmium(II) ions are sometimes useful for end-point detection in complexometric titrations.

There are many biosensors based on ISEs, e.g., for urea, amino acids, and a wide range of biologically important substances, but they are mostly useful in highly specialized determinations. Ion-selective FETs and similar sensors should be very useful owing to their general advantages (see the previous section), but their wider routine use has so far been hampered by their drawbacks, which are rather poor signal stability and technical problems with the chip insulation. In competition with voltammetric sensors, ISFETs often lose because of their smaller versatility. On the other hand, the Severinghaus-type gas sensors have been practically very useful, especially in monitoring environmentally important gases, e.g., sulfur and nitrogen oxides, ammonia, and other gases.

**See also:** **Elemental Speciation:** Overview. **Ion-Selective Electrodes:** Liquid Membrane; Gas Sensing Probes. **pH. Sensors:** Overview; Amperometric Oxygen Sensors;

Calorimetric/Enthalpimetric; Chemically Modified Electrodes; Microorganism-Based; Photometric; Piezoelectric Resonators; Tissue-Based.

## Further Reading

- Bailey PL (1980) *Analysis with Ion-selective Electrodes*, 2nd edn. London: Heyden.
- Cammann K (1979) *Working with Ion-selective Electrodes*. Berlin: Springer.
- Freiser H (ed.) (1978) *Ion-selective Electrodes in Analytical Chemistry*, vol. 1. New York: Plenum.
- Freiser H (ed.) (1980) *Ion-selective Electrodes in Analytical Chemistry*, vol. 2. New York: Plenum.
- Janata J (1989) *Principles of Chemical Sensors*. New York: Plenum.
- Koryta J (ed.) (1980) *Medical and Biological Applications of Electrochemical Devices*. Chichester: Wiley.
- Koryta J and Štulík K (1983) *Ion-selective Electrodes*. Cambridge: Cambridge University Press.
- Morf WE (1981) *The Principles of Ion-selective Electrodes and of Membrane Transport*. Amsterdam: Elsevier.
- Oehme F (1991) *Liquid Electrolyte Sensors: Potentiometry, Amperometry, and Conductometry*. In: Göpel W, Hese J, and Zemel JN (eds.) *Sensors. A Comprehensive Survey*, vol. 2, ch. 7. Weinheim: Wiley-VCH.
- Umezawa Y (ed.) (1990) *Handbook of Ion-selective Electrodes: Selectivity Coefficients*. Boca Raton, FL: CRC Press.
- Umezawa Y, Bühlmann P, Umezawa K, Tohda K, and Amemiya S (2000) Potentiometric selectivity coefficients of ion-selective electrodes. Part I. Inorganic cations. *IUPAC, Pure and Applied Chemistry* 72: 1851–2082.
- Umezawa Y, Umezawa K, Bühlmann P, Hamada N, Aoki H, Nakanishi J, Moritoshi S, Xia KP, and Nishimura Y (2002) Potentiometric selectivity coefficients of ion-selective electrodes. Part II. Inorganic anions. *IUPAC, Pure and Applied Chemistry* 74: 923–994.
- Umezawa Y, Bühlmann P, Umezawa K, and Hamada N (2002) Potentiometric selectivity coefficients of ion-selective electrodes. Part III. Organic ions. *IUPAC, Pure and Applied Chemistry* 74: 995–1099.

## Glass

**S Głąb and A Hulanicki**, University of Warsaw, Warsaw, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Glass electrodes sensitive to hydrogen ions are the most commonly used sensors in chemistry and related disciplines. They belong to the group of potentiometric

membrane sensors and are constructed in various configurations, depending on the application. Their basic design and properties are described and discussed in this article. Among their important properties is the selectivity, which depends on the composition of the glass. Glass electrodes are used mainly for pH measurements, but they may compose a part of more sophisticated systems in gas sensors or enzymatic sensors. Recently glasses based on a nonsilica structure have been used in potentiometric measurements.



## Origin of Glass Electrodes

Glass electrodes are ion-selective electrodes that belong to the group of electrodes with a noncrystalline solid membrane. Typical glasses are supercooled liquids with a network composed most commonly of silicon oxide; however other compositions may have similar properties and can be used as sensors.

The formation of an electric potential difference on a thin glass membrane in contact with solutions was observed by Cramer in 1906 and used for construction of a device for measuring the acidity of solutions by Haber and Klemensiewicz in 1909. Due to the high resistance of the glass membrane, practical application in routine measurement of acidity, and of course solution pH, was postponed until the invention of vacuum tube amplifiers, and later semiconductors. One of the first types of glass used in manufacturing electrodes was introduced in 1928–1929 by Corning under the designation 015. The research devoted to the mechanism of this electrode action was a stimulus that helped to develop and introduce other types of ion-selective electrodes.

## Electrode Design

A typical glass electrode (Figure 1) consists of a thin glass membrane (thickness 50–300  $\mu\text{m}$ ) at the end of a glass tube acting as (a) an insulator and (b) as a container of the internal solution in which the

internal (reference) electrode is immersed. This internal electrode is usually a silver/silver chloride electrode, and therefore the internal solution should contain a fixed concentration of chloride ions. Thus the glass electrode is in fact a half-cell with a membrane forming a contact to the other half-cell:



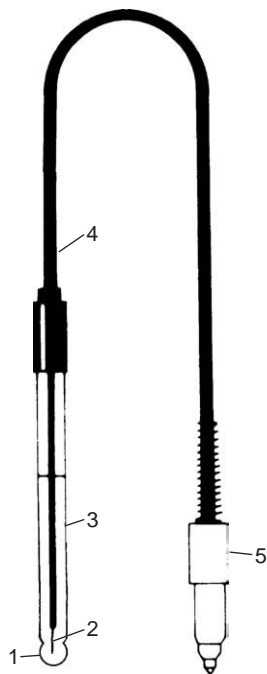
The resistance of the whole electrode, depending on its composition and dimensions, is in a broad range of 1–1000 M $\Omega$ . Due to the high resistance the cable connecting it to the measuring device (voltmeter, pH meter) must be protected from electrostatic interferences. The potential difference between the two sides of the glass membrane is proportional to the difference between the pH values on both sides, according to the Nernst equation

$$\Delta E = 2.303 RT/F(\text{pH}_1 - \text{pH}_2) \quad [1]$$

if the electrode for hydrogen ions is considered. To measure an unknown pH value it is therefore necessary to know the pH on one side of the membrane and measure  $\Delta E$ . In practice, for determination of the unknown pH value there is no need to know the exact pH of the internal solution but only to keep it constant. The cell with glass and reference electrodes is calibrated with buffer solutions used as hydrogen activity standards. The same principle is the basis of measurements with cation-selective electrodes that measure the pM rather than pH.

## Composition of Glass Membranes

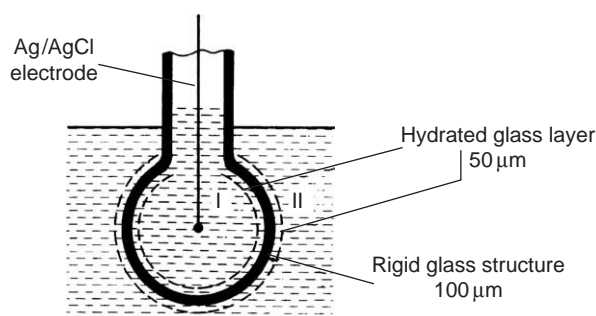
A glass membrane is composed of oxides such as  $\text{SiO}_2$ ,  $\text{B}_2\text{O}_3$ ,  $\text{P}_2\text{O}_3$ ,  $\text{As}_2\text{O}_3$ , and  $\text{GeO}_2$  and oxides of monovalent or divalent metals e.g.,  $\text{Na}_2\text{O}$ ,  $\text{K}_2\text{O}$ ,  $\text{Li}_2\text{O}$ ,  $\text{CaO}$ . The most common glass membranes are, however, based on the silicon skeleton, containing oxygen atoms around the silicon tetrahedrons. The oxygen may form bridges between the silicon atoms or be nonbridging. The latter can be associated with single charged cations. The alkali metal cations may be exchanged with ions in solution and generate the potential response of the glass electrode. The analyte ions on the surface sites tend to diffuse inside the hydrated gel layer. Such diffusion probably takes place by means of a defect and jump mechanism in which the interstitial cations play a major role. The tendency to interact more selectively with a given ion depends on the composition of the glass. The typical glass composition for a hydrogen ion-sensitive electrode (Corning 015) contains 22%  $\text{Na}_2\text{O}$ , 6%  $\text{CaO}$ , and 72%  $\text{SiO}_2$  and responds according to the Nernst equation (eqn [1]) up to pH 11–12. Replacement of sodium for lithium



**Figure 1** Typical commercial glass membrane electrode. 1, sensitive glass membrane; 2, Ag/AgCl internal reference electrode; 3, internal solution with constant pH; 4, cable; 5, plug.

**Table 1** Composition of some typical glasses used as membranes

Increased selectivity for	Membrane type	Chemical composition (mol%)
H <sup>+</sup>	Corning 015	22% Na <sub>2</sub> O, 6% CaO, 72% SiO <sub>2</sub>
Na <sup>+</sup>	NAS 11-18	11% Na <sub>2</sub> O, 18% Al <sub>2</sub> O <sub>3</sub> , 71% SiO <sub>2</sub>
K <sup>+</sup>	NAS 27-04	27% Na <sub>2</sub> O, 4% Al <sub>2</sub> O <sub>3</sub> , 69% SiO <sub>2</sub>
Li <sup>+</sup>	LAS 15-25	15% Li <sub>2</sub> O, 25% Al <sub>2</sub> O <sub>3</sub> , 60% SiO <sub>2</sub>
NH <sub>4</sub> <sup>+</sup>	NAS 27-03	27% Na <sub>2</sub> O, 3% Al <sub>2</sub> O <sub>3</sub> , 67% SiO <sub>2</sub> , 3% ZnO

**Figure 2** Schematic representation of the hydrated layer of a pH-sensitive glass membrane electrode. I, internal reference solution; II, external (sample) solution of unknown pH.

extends this range up to pH 13. The partial replacement of silicon dioxide by trivalent element oxides favors the selectivity for alkali metal cations as the excess of negative charge must be neutralized by the alkali metal cations. An appropriate composition of the glass membrane enables us to use them as sensors for several metal cations. The preferred composition of some glasses used in ion sensors is presented in Table 1.

Major changes in the composition of glasses cannot be made arbitrarily because excessive amounts of alkali metal oxides lower the membrane durability in solutions. On the other hand, too high a proportion of silica decreases the hydrophilicity of the membrane, which is indispensable for proper functioning of the membrane. The optimum water uptake by the hydrated layer of the electrode membrane varies from 50 to 100 mg cm<sup>-3</sup>. The thickness of the hydrated layer on both sides of the membrane, which is responsible for the electrode response, varies from 5 to 100 nm (Figure 2). Therefore drying out of the membrane is destructive to the electrode performance. The lifetime of glass electrodes depends on the glass composition, aggressivity of solutions, and temperature. An alkaline solution affects the silicon-oxygen network, and acid attack causes the replacement of metal ions by hydrogen ions, in the hydrated silica layer. Special electrodes can work up to 80°C,

but an increase in temperature promotes solution attack on the membrane and changes the parameters of the electrode. Changes in the temperature influence not only the prelogarithmic term of the Nernst-Nikolsky equation but also the equilibria in the solution and at the phase boundaries. A glass electrode with the internal redox system Pt|I<sub>2</sub>/I<sup>-</sup> replacing the usual silver/silver chloride system is much less dependent on the temperature changes (Ross electrode).

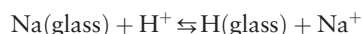
### Selectivity

The Nernstian response, i.e., range of the linear dependence of potential versus pH, in the absence of interfering ions extends usually from pH 1 up to 11; however, many contemporary electrodes, depending on the glass composition, have made possible pH measurements up to pH 13. The response of the glass electrodes to metal cations extends down to 10<sup>-5</sup> mol l<sup>-1</sup>, and this is caused mainly by the presence of interferents (hydrogen ions for cation electrodes).

A glass pH electrode exhibits the so-called alkaline error, which means that in the alkaline pH range the potential response is influenced by the presence of sodium ions. The study of this behavior gave rise to the Nikolsky equation,

$$E = E_0 + \frac{RT}{z_i F} \ln \left( a_i + \sum K_{ij}^{\text{pot}} a_j^{z_i/z_j} \right) \quad [2]$$

which is an expanded form of the Nernst equation valid in cases when apart from the main ion,  $a_i$ , with charge  $z_i$ , other ions,  $a_j$ , with charge  $z_j$ , influence the response. In the case of the sodium ion error  $a_j = a_{\text{Na}}$ . The Nikolsky equation indicates the importance of the electrode selectivity, which is quantitatively expressed by the selectivity coefficient,  $K_{ij}^{\text{pot}}$ . The values of selectivity coefficients are important when characterizing the behavior of electrodes sensitive to metal ions. However, these values are not real constants characteristic for a given type of electrode as they depend on the method of measurement and the history of the electrode. Their limiting values are expressed by the product of the constant for the exchange reaction, e.g.,



and the ratio of the ion mobilities,

$$K_{\text{Na,H}}^{\text{pot}} = K_{\text{exch}} \frac{U_{\text{Na}}}{U_{\text{H}}} \quad [3]$$

The values of the selectivity coefficients for some glass electrodes are given in Table 2.

The response times (95% of the equilibrium potential change) given in the literature often correspond to the response time of the whole cell. They are of the

order of a few seconds but also depend on the glass composition, the presence of interferents, the concentration, the temperature, and the electrode ageing.

The measured potential may also be influenced by the so-called asymmetry potential. This is a residual potential between the two glass membrane surfaces when the solutions on both sides are identical. The asymmetry potential is due to differences in structural stresses at the two surfaces, differences in the surface history (ageing), and interaction with different solutions. The asymmetry potential may change with time.

### Applications

Commercial glass electrodes, mostly for pH measurements, differ in construction because they have to serve different purposes. The electrode tips may have different shapes, including tubular flow-through electrodes (Figure 3). Special electrodes are used under conditions of high pressures or high temperatures or

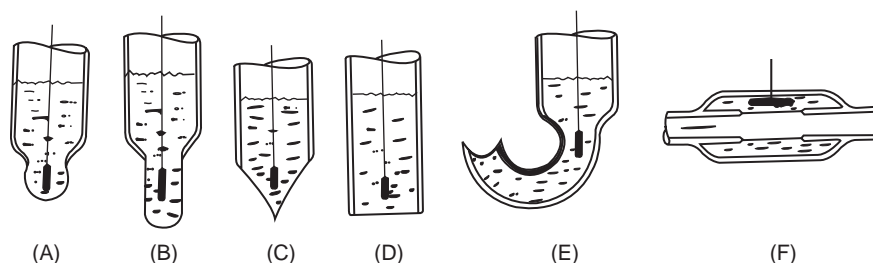
in shock conditions. Electrodes for medical applications, in particular for measurements *in vivo* and control of biotechnological processes, should not be destroyed during sterilization at 120°C. Special construction is necessary for microelectrodes for biochemical purposes. They can be as small as 0.5  $\mu\text{m}$  in diameter to enable measurements within single cells. In many cases it is convenient to use combination electrodes that contain the indicator glass membrane electrode and the reference electrode in one rigid body (Figure 4A).

Glass membrane electrodes, mainly used for pH measurements but also for the  $\text{NH}_4^+$  ion are used for construction of gas sensors (Figure 4B). Such sensors, based on a principle presented by Severinghaus (1958), consist of glass electrodes covered by a thin film made of porous hydrophobic plastic, so that the solution cannot penetrate into the pores. A thin layer of an electrolyte solution is located between the surface of the film and the glass surface. The pH value of that solution changes under the influence of a gas (e.g.,  $\text{NH}_3$ ,  $\text{CO}_2$ ,  $\text{SO}_2$ ), diffusing through the film and is the measure of the gas concentration in the sample solution.

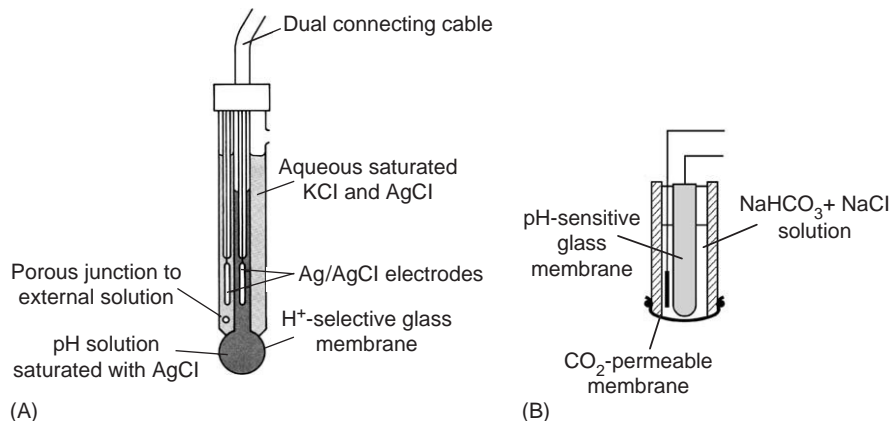
Another type of multiple membrane sensor consists of a glass electrode in contact with a thin layer of solution, being covered with an immobilized enzyme layer. The enzyme (e.g., urease) catalyzes the

**Table 2** Selectivity coefficients,  $K_{ij}^{\text{pot}}$ , for some glass membrane electrodes

Membrane type	Selectivity coefficients for interfering ions				
	$\text{H}^+$	$\text{Na}^+$	$\text{K}^+$	$\text{NH}_4^+$	$\text{Ag}^+$
NAS 11-18	$10^3$	1	$10^{-3}$	$3 \times 10^{-5}$	$4 \times 10^2$
NAS 27-04		$10^{-1}$	1	$3 \times 10^{-1}$	



**Figure 3** (A–E) Shapes of glass electrode tips and (F) tubular flow-through electrode.



**Figure 4** (A) Combination pH-sensitive glass electrode with the reference electrode (Ag/AgCl). (B) Carbon dioxide gas electrode based on pH-sensitive glass electrode.

reaction of the analyte (e.g., urea) and the products of the reaction  $\text{CO}(\text{NH}_2)_2 + \text{H}_2\text{O} \rightarrow \text{CO}_2 + 2\text{NH}_3$ . The resulting change in the pH of the solution in contact with the glass electrode allows quantitative determination of the analyte (in this case urea).

## Chalcogenide Glass Electrodes

The general term 'glass electrode' is also used in connection with electrodes whose membrane is composed of glasses of different composition and different construction from conventional glass electrodes. The glass membrane is composed of amorphous solids that are binary, ternary, or multicomponent compounds of the elements of group III, IV, or V (boron, aluminum, gallium, germanium, phosphorus, arsenic, antimony) of the periodic table with sulfur, selenium, or tellurium. They can be both stoichiometric (e.g.,  $\text{As}_2\text{S}_3$ ,  $\text{GeS}_2$ ) and nonstoichiometric (e.g.,  $\text{As}_x\text{Se}_{1-x}$ ). Often the membranes are doped with transition metal compounds. Typical glass membranes are obtained by melting the components and forming the membrane proper from the homogeneous melt. A chalcogenide glass membrane is obtained as a disk cut from the solidified melt, and after polishing is pasted into the plastic body in a way similar to that for crystalline ion-selective electrodes.

The chalcogenide glass electrodes may be selective e.g., for iron(III) (glass containing  $\text{Fe}_n\text{Se}_{60}\text{Ge}_{28}\text{Sb}_{12}$ ), copper(II) ( $\text{Cu}_6\text{As}_4\text{S}_9$  – sinnerite), silver(I), lead(II), and cadmium(II). The way it functions is not yet clear: it may be sensitive to powerful redox systems, its characteristics may be influenced by doping, and its behavior is often non-Nernstian. Such a glass may exhibit a mixed ionic or electronic conductance.

*See also:* **Enzymes:** Enzyme-Based Electrodes. **Glasses. Ion-Selective Electrodes:** Overview; Gas Sensing Probes; Enzyme Electrodes. **pH. Sensors:** Overview.

## Further Reading

- Bates RG (1973) *Determination of pH: Theory and Practice*, 2nd edn. New York: Wiley.
- Degner R and Leibl S (1995) *pH Messen*. Weinheim: VCH.
- Eisenman G (ed.) (1967) *Glass Electrodes for Hydrogen and Other Cations: Principles and Applications*. New York: Dekker.
- Lakshminarayanaiah N (1976) *Membrane Electrodes*. New York: Academic Press.
- Lavalle M, Schanne OF, and Hebert NC (eds.) (1969) *Glass Microelectrodes*. New York: Wiley.
- Schwabe K (1980) *pH Messung*. Berlin: Akademie Verlag.
- Vlasov YuG and Bychov EA (1987) Ion-selective chalcogenide glass electrodes. *Ion-Selective Electrode Review* 9: 5–93.

## Solid-State

**V Horváth and G Horvai**, Budapest University of Technology and Economics, Budapest, Hungary

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The classification of ion-selective electrodes (ISEs) into glass, solid-state, and liquid electrodes follows historical tradition and it also reflects differences in electrode fabrication procedures. As far as glass and liquid electrodes are concerned, the classification is more or less in agreement with usage of the terms glassy and liquid state in other areas of chemistry. But the term 'solid electrodes' defines a variety of potentiometric sensors. Membrane electrodes from crystalline solid materials belong to this group. These can be single crystals, pellets pressed from crystalline powder, or solidified melts. The crystalline material may be a single compound or a homogeneous mixture of compounds (e.g., AgI and  $\text{Ag}_2\text{S}$ ). If the mechanical

properties of a pressed pellet are unsatisfactory it may be advantageous to embed the crystals into an inert polymer matrix, such as silicone rubber. These are called heterogeneous membranes. Amorphous or glassy solids have also been used as membrane materials in ISEs and with the exception of silicate-type glasses these are usually regarded as solid electrodes.

Semiconductor processing technologies have often been used to produce ISEs, particularly as field-effect transistors (FETs) with ion-selective layers like silicon oxide over the gate region. Such ion-selective FETs (ISFETs) are, in principle, solid ISEs, although sometimes the dielectric over the gate is covered with a second, liquid membrane-type layer to achieve different selectivities.

Polymeric materials may also be used as electrodes. Very thin polymeric films may be produced at the surface of metal electrodes like platinum. Some of these electrodes may also be useful as ion-selective potentiometric sensors.



reaction of the analyte (e.g., urea) and the products of the reaction  $\text{CO}(\text{NH}_2)_2 + \text{H}_2\text{O} \rightarrow \text{CO}_2 + 2\text{NH}_3$ . The resulting change in the pH of the solution in contact with the glass electrode allows quantitative determination of the analyte (in this case urea).

## Chalcogenide Glass Electrodes

The general term 'glass electrode' is also used in connection with electrodes whose membrane is composed of glasses of different composition and different construction from conventional glass electrodes. The glass membrane is composed of amorphous solids that are binary, ternary, or multicomponent compounds of the elements of group III, IV, or V (boron, aluminum, gallium, germanium, phosphorus, arsenic, antimony) of the periodic table with sulfur, selenium, or tellurium. They can be both stoichiometric (e.g.,  $\text{As}_2\text{S}_3$ ,  $\text{GeS}_2$ ) and nonstoichiometric (e.g.,  $\text{As}_x\text{Se}_{1-x}$ ). Often the membranes are doped with transition metal compounds. Typical glass membranes are obtained by melting the components and forming the membrane proper from the homogeneous melt. A chalcogenide glass membrane is obtained as a disk cut from the solidified melt, and after polishing is pasted into the plastic body in a way similar to that for crystalline ion-selective electrodes.

The chalcogenide glass electrodes may be selective e.g., for iron(III) (glass containing  $\text{Fe}_n\text{Se}_{60}\text{Ge}_{28}\text{Sb}_{12}$ ), copper(II) ( $\text{Cu}_6\text{As}_4\text{S}_9$  – sinnerite), silver(I), lead(II), and cadmium(II). The way it functions is not yet clear: it may be sensitive to powerful redox systems, its characteristics may be influenced by doping, and its behavior is often non-Nernstian. Such a glass may exhibit a mixed ionic or electronic conductance.

*See also:* **Enzymes:** Enzyme-Based Electrodes. **Glasses. Ion-Selective Electrodes:** Overview; Gas Sensing Probes; Enzyme Electrodes. **pH. Sensors:** Overview.

## Further Reading

- Bates RG (1973) *Determination of pH: Theory and Practice*, 2nd edn. New York: Wiley.
- Degner R and Leibl S (1995) *pH Messen*. Weinheim: VCH.
- Eisenman G (ed.) (1967) *Glass Electrodes for Hydrogen and Other Cations: Principles and Applications*. New York: Dekker.
- Lakshminarayanaiah N (1976) *Membrane Electrodes*. New York: Academic Press.
- Lavalle M, Schanne OF, and Hebert NC (eds.) (1969) *Glass Microelectrodes*. New York: Wiley.
- Schwabe K (1980) *pH Messung*. Berlin: Akademie Verlag.
- Vlasov YuG and Bychov EA (1987) Ion-selective chalcogenide glass electrodes. *Ion-Selective Electrode Review* 9: 5–93.

## Solid-State

**V Horváth and G Horvai**, Budapest University of Technology and Economics, Budapest, Hungary

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The classification of ion-selective electrodes (ISEs) into glass, solid-state, and liquid electrodes follows historical tradition and it also reflects differences in electrode fabrication procedures. As far as glass and liquid electrodes are concerned, the classification is more or less in agreement with usage of the terms glassy and liquid state in other areas of chemistry. But the term 'solid electrodes' defines a variety of potentiometric sensors. Membrane electrodes from crystalline solid materials belong to this group. These can be single crystals, pellets pressed from crystalline powder, or solidified melts. The crystalline material may be a single compound or a homogeneous mixture of compounds (e.g., AgI and  $\text{Ag}_2\text{S}$ ). If the mechanical

properties of a pressed pellet are unsatisfactory it may be advantageous to embed the crystals into an inert polymer matrix, such as silicone rubber. These are called heterogeneous membranes. Amorphous or glassy solids have also been used as membrane materials in ISEs and with the exception of silicate-type glasses these are usually regarded as solid electrodes.

Semiconductor processing technologies have often been used to produce ISEs, particularly as field-effect transistors (FETs) with ion-selective layers like silicon oxide over the gate region. Such ion-selective FETs (ISFETs) are, in principle, solid ISEs, although sometimes the dielectric over the gate is covered with a second, liquid membrane-type layer to achieve different selectivities.

Polymeric materials may also be used as electrodes. Very thin polymeric films may be produced at the surface of metal electrodes like platinum. Some of these electrodes may also be useful as ion-selective potentiometric sensors.



Thin layers of solids like chalcogenides may also be deposited on a metal surface by electrolysis or by vacuum deposition techniques. These electrodes are closer to solid ISEs than to traditional electrodes of the second kind.

This article focuses on self-supporting disk-type membranes with liquid or solid internal contact, which is the typical configuration of most commercial ISEs. Other solid ISEs are briefly discussed.

All-solid-state electrodes refer to where there are no free-flowing liquids used in the electrode construction. The practical advantages that result include insensitivity to electrode orientation and smaller size. The difficulty with this structure is often the instability of the potential developing at the back contact of the ISE membrane. Sometimes such an ISE is called all-solid-state even if the sensitive membrane is a liquid membrane.

## History

Glass electrodes have been available since the early 1900s and solid-state electrodes are even older. The first membrane-type solid electrodes were made in the 1930s without arousing much interest. The potential of these devices was realized only when a variety of good and sufficiently selective electrodes were developed by Pungor and co-workers and subsequently by Frant and Ross. The former group embedded crystal-line precipitates in a silicone rubber matrix whilst Frant and Ross made single crystal fluoride electrodes and pressed pellet electrodes for some other ions. Both groups succeeded in transferring their inventions into industrial production. Since then many workers have contributed to the analytical utilization and understanding of the operation of solid ISEs. The work on ISFETs was initiated by Bergveld.

## Principles

Ion-selective electrode theories need to explain why a stable electromotive force (e.m.f.) develops in the measurement cell, how the concentration of the primary ion and interfering ions affect the e.m.f., and how fast the e.m.f. responds to changes in sample composition. The theories are usually based on a model description of the measurement cell and the process occurring in it. Before going into details it should be noted that different, sometimes mutually contradicting models, may equally well explain the same potentiometric observation. In fact, different real systems may behave potentiometrically in the same way. In other words, potentiometric observations are usually not very useful as positive proof of a model; additional information, e.g., from surface spectroscopies, impedance measurements, radiotracer studies, is necessary.

In the following discussion an electrode membrane consisting of a sparingly soluble salt  $I^+A^-$  and separating the aqueous sample from the inner reference solution is considered. Suitable reference electrodes immersed in the two solutions complete the cell used for e.m.f. measurement.

### The Nernst Equation

The relationship between an e.m.f. ( $E$ ) and primary ion activity is ideally given by the Nernst equation:

$$E = E^\circ + \frac{RT}{F} \ln a_1 \quad [1]$$

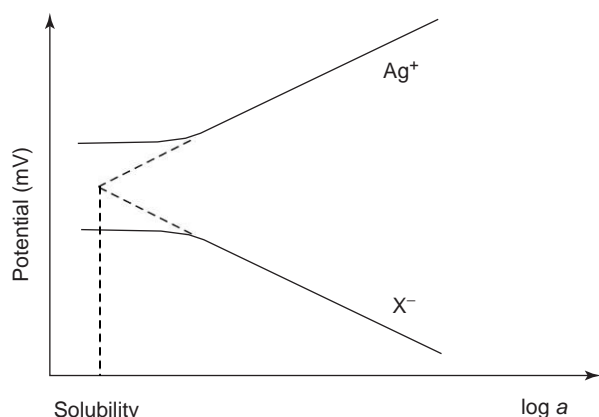
where  $E^\circ$  (standard e.m.f.) is a constant,  $R$ ,  $T$ , and  $F$  have their usual meaning, and  $a_1$  is the activity of the primary ion  $I^+$  in the sample. This equation may be derived by observing that the e.m.f. adds up from several potential drops in the cell but sample composition should influence only the potential drop across the membrane and at the sample/reference electrode liquid junction (in cells with transference). Neglecting the changes of the latter we need to pay attention only to the potential drop between sample and internal reference solution, i.e., across the membrane. In the absence of interferences this potential difference can be calculated from thermodynamic considerations:

$$\Delta\phi_{1,2} = \frac{RT}{F} \ln \frac{a_{1,1}}{a_{1,2}} = \frac{RT}{F} \ln \frac{a_{A,2}}{a_{A,1}} \quad [2]$$

where  $a_{1,1}$  is the activity of the primary cation in the sample,  $a_{1,2}$  is the same in the internal reference solution, and the  $a_{A,s}$  are the respective quantities for the primary anion. The composition of the internal reference solution is usually independent of the sample, so that eqn [1] or a similar equation for the primary ion is valid.

When writing eqn [2] it has been assumed that the internal reference solution is a soluble salt of  $I^+$  when this ion is measured, and a soluble salt of  $A^-$  when this is the analyte. In practice, there is no need to change the internal electrolyte according to the measured ion. If the internal solution is an  $I^+$  salt but the analyte is an  $A^-$  salt the Nernst equation is still valid albeit with a different  $E^\circ$ . **Figure 1** shows typical calibration lines with an unchanged internal solution. Apparent deviations from the Nernstian straight line are observed at low analyte concentration due to membrane solubility (see the section on detection limit).

The  $I^+A^-$  membrane may be contacted on the inner side with a solid, electronically conducting material to form an all-solid-state electrode. This is generally regarded to be a sound practice only if the two contacting solid phases can equilibrate, i.e., there is at least one charged species (ion or electron)



**Figure 1** Typical calibration lines of a silver halide, AgX electrode for  $\text{Ag}^+$  and  $\text{X}^-$ .

that can easily transgress the phase boundary in both directions. One example is metallic silver (or silver containing adhesive) in contact with a silver salt membrane. Thermodynamic reasoning yields in this case again the Nernst equation. Solid contacts that do not appear to fulfill the above requirement have occasionally been used successfully.

### Detection Limits

The analytically useful concentration range of a particular electrode type may be limited both from the higher and the lower ends. At the high end the membrane material may rapidly deteriorate, e.g., due to dissolution in complex form, e.g., AgCl-based chloride electrodes in concentrated chloride solutions, or AgI-based cyanide electrodes in strong cyanide solutions. The lower detection limit is more crucial for practical applications, and typically is governed by the solubility of the membrane material. This statement is quite natural if the sample becomes saturated with the membrane material during the time of measurement. In most practical cases the sample volume is large, the membrane solubility is low, so that the rate of equilibration is rather slow, because it is determined by the rate of membrane dissolution. One can assume, however, that at the membrane/solution interface local equilibrium exists, so that the concentration detected by the electrode in a zero concentration solution is indeed the saturation concentration. Near the detection limit the calibration line will be curved; useful equations have been derived to describe this range. The lower detection limit can be extended, however, either by applying a small current to the electrode, or by using hydrodynamic flow (rotating disk electrode or continuous flow analysis). The first method removes dissolved ions from the electrode surface, the second one diminishes the diffusion layer at the electrode surface,

thereby enhancing the flux of dissolved ions into the bulk solution.

Sometimes the practical detection limit is well above the solubility value. This may be due to interferences, secondary equilibria, and/or the small exchange current density at the interface.

### Interferences

Different types of interferences may be encountered in practice; some important ones will be mentioned here. Corrosive media may destroy the membrane or the electrode housing.

Ions with the same charge sign as the analyte may be sensed by the electrode. In this case, the Nicolsky equation is often approximately valid:

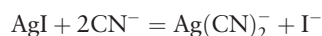
$$E = E^\circ + \frac{RT}{z_1 F} \ln(a_1 + K_{IJ}^{\text{pot}} a_J^{z_1/z_I}) \quad [3]$$

where  $a_J$  is the activity of the interfering ion,  $z_I$  and  $z_J$  are the respective charges of ions I and J, and  $K_{IJ}$  is the selectivity coefficient. As a simple case one can consider a membrane of  $\text{I}^+ \text{A}^-$  and an interfering ion  $\text{J}^+$ , which forms a more soluble precipitate  $\text{J}^+ \text{A}^-$  than  $\text{I}^+ \text{A}^-$ . If the electrode is in solubility equilibrium with a solution of  $\text{I}^+$  and subsequently more and more  $\text{J}^+$  is added to the solution, no precipitate  $\text{J}^+ \text{A}^-$  should form until

$$a_J \leq \frac{L_{JA}}{L_{IA}} \cdot a_I \quad [4]$$

(at least close to the electrode surface), where the  $L$ s are the respective solubility products of the precipitates. Above this limit  $\text{J}^+ \text{A}^-$  can be formed at the electrode surface or the surface is converted into a mixed crystal  $\text{IJA}_2$ . Exact theoretical treatment of these processes is not always possible. Yet for some cases it can be derived and in many cases it can be found experimentally that the selectivity coefficient is approximately  $L_{IA}/L_{JA}$ .

The interference may also be caused by agents forming a soluble complex with either component of the membrane salt and/or with the analyte. A prime example is the effect of  $\text{CN}^-$  on AgI electrodes.  $\text{CN}^-$  corrodes the AgI membrane by the reaction:



The cell voltage measured in a solution of  $\text{CN}^-$  (with or without the presence of  $\text{I}^-$ ) is determined by the stationary iodide concentration at the membrane/solution interface. This concentration in turn is determined by the rate of dissolution of AgI and the rate of  $\text{I}^-$  transport from the electrode surface to the solution bulk. Thus, the cell voltage (which is not an

e.m.f. because there is no equilibrium) depends on rate processes. Nevertheless, calculations show that the Nicolsky equation is approximately valid for the  $\text{CN}^-$  interference at the AgI electrode, with a selectivity coefficient of 0.5.

### Response Time

The time response of solid ISEs to sudden concentration changes of the primary ion is usually very fast, less than 100 ms for 90% of the total e.m.f. change, except in the vicinity of the detection limit. Such values can be observed, however, only in equipment of special design. In practical cells the time course of concentration change at the electrode surface usually cannot be made sufficiently abrupt and the measured time response is slower than that of the ISE. Many valuable results have been obtained with suitable equipment and approximate equations were given to describe the time course of response.

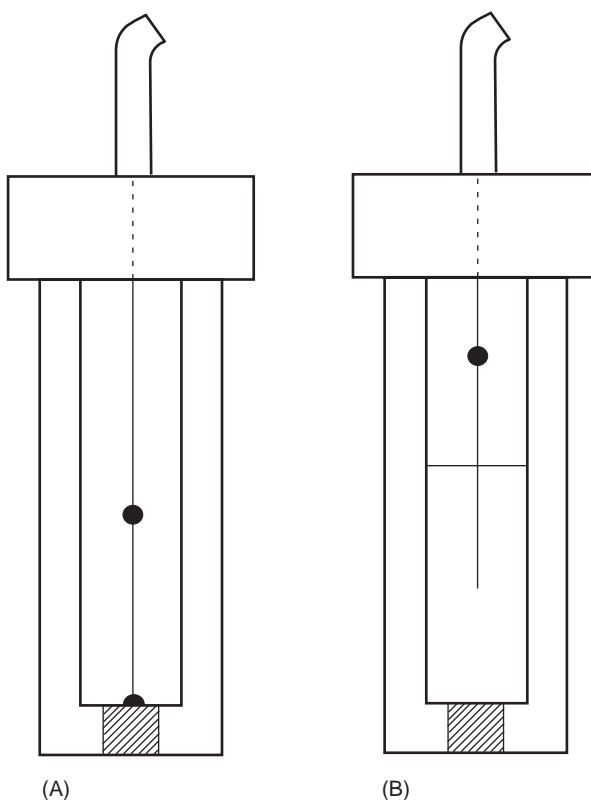
### Potential Stability and Lifetime

The short time repeatability of the e.m.f. measurement with solid ISEs is usually good, typically within 0.1 mV. The long-term potential stability in the same solution is an important parameter in monitoring applications. Literature data appear to be scarce in this respect. A value of 1 mV per week can be regarded as satisfactory.

## Practical Electrodes of Conventional Design

### Electrode Construction and Compensation of Temperature Effects

In commercial electrodes, the solid membrane is pressed or glued into a plastic electrode shaft. The internal contact is either solid, typically silver, or an internal filling solution is used with a silver-silver chloride reference element dipping into this electrolyte (Figure 2). The type and composition of the internal contact is important for the temperature dependence of the e.m.f. In laboratory work temperature compensation is rarely required, but in industrial or environmental monitoring it is necessary. The best practice is to use a variable internal filling solution and adjust its composition to the expected analyte concentration range. The external and internal reference elements should also be as similar as possible. In this way a nearly symmetrical cell is obtained with close to zero e.m.f. in the monitored concentration range. Under such conditions temperature effects are small and correction should be easier (provided that there are no temperature gradients within the measurement cell).



**Figure 2** Solid electrodes with (A) solid and (B) liquid internal contacts.

### Lanthanum(III) Fluoride Electrodes

Until very recently successful  $\text{LaF}_3$  electrodes could only be made from lanthanum(III) fluoride ( $\text{LaF}_3$ ) single crystals. The hexagonal lattice of this crystal consists of layers of  $\text{LaF}_2^+$  with a fluoride layer on either side. The fluoride ions have appreciable mobility in the lattice, which results in a specific conductivity of  $\sim 3 \times 10^{-7} \text{ S cm}^{-1}$ . This value can be raised 10-fold by doping with 1%  $\text{EuF}_3$ . The electrode response is linear from about  $10^{-6}$  to  $1 \text{ mol l}^{-1} \text{ F}^-$ , the detection limit is  $\sim 10^{-7} \text{ mol l}^{-1}$ . These data are valid for nearly neutral pH values; at higher pH  $\text{OH}^-$  ions interfere and at low pH values HF and  $\text{HF}_2^-$  begin to form. Apart from this, interferences are scarce and caused by the complexation of the analyte (fluoride) with metal ions such as iron(III) and aluminum(III) in the sample. The interferences can mostly be avoided by using a suitable sample adjustment buffer solution (total ionic strength adjustment buffer). The response time of the electrode is fast except near the detection limit.

The fluoride ISE has been the most successful of all ISEs. This is so partly because it has uniquely good characteristics including a long service lifetime and partly because the potentiometric fluoride determination is so much more convenient than any

alternative method. The applications extend from analysis of minerals and teeth to other biomedical, environmental, and industrial determinations.

### Silver Sulfide Electrodes

Silver sulfide is an excellent material for a solid-state electrode. It has very low solubility and consequently there are only few potentiometric interferences ( $\text{CN}^-$ ,  $\text{Hg}^{2+}$ ). It can be easily compressed into mechanically stable pellets and it is also accessible as single crystal for electrodes. At ambient conditions (below  $176^\circ\text{C}$ ) its monoclinic  $\beta$ -modification, acanthite, is stable, and shows high ionic conductivity due to silver ions. Silver sulfide electrodes respond to both silver and sulfide ions with good practical detection limits ( $10^{-8}$ – $10^{-7} \text{ mol l}^{-1}$ ) in Nernstian manner. Sulfide analysis is performed in strongly alkaline solutions.

Important applications of this electrode are the direct determination of sulfide (or other sulfur-containing species converted to sulfide) in aqueous solutions and as an indicator electrode in argentometric titrations.

### Silver Halides for Electrodes

The first ISEs were made from silver halide melt or from precipitate with inert silicone rubber binding material. This construction has been replaced in modern commercial electrodes by mixtures of silver halides with silver sulfide. The pure silver halides are difficult to compress into mechanically stable pellets, and this problem is overcome by the silver sulfide admixture. The conductivity is also considerably increased and light sensitivity diminished. The mixture is usually obtained by coprecipitation. In the compressed pellet the presence of some  $\text{Ag}_3\text{SBr}$  and  $\text{Ag}_2\text{SI}$  has been claimed. The detection limits for the halide ions and silver is close to the solubilities ( $10^{-5}$ ,  $10^{-6}$ , and  $10^{-8} \text{ mol l}^{-1}$  for  $\text{AgCl}$ ,  $\text{AgBr}$ , and  $\text{AgI}$ , respectively). For the analysis of chloride below this limit  $\text{Hg}_2\text{Cl}_2/\text{Ag}_2\text{S}$  electrodes have been useful. Important interferences in halide assays are from the less soluble halides, sulfide, cyanide, and thiocyanate. The slow corrosion process of  $\text{AgI}$  in dilute cyanide solutions is the basis for the potentiometric assay of this ion with an iodide selective electrode in alkaline solutions. For the measurement of silver ion the  $\text{Ag}_2\text{S}$  electrode is preferred to the halides. The greatest practical utility among the halide electrodes has been attained by the chloride ISE due to the general importance of chloride assay in analytical chemistry.

### Chalcogenides of Doubly Charged Metal Ions

The most important of this group are copper, lead, and cadmium compounds. With the exception of some nonstoichiometric copper chalcogenides, these

materials are insulators or semiconductors and not suitable for ISE preparation. Coprecipitation with silver sulfide has led, however, to useful electrode materials either as pressed (sintered) pellets or dispersed in an inert binder. The materials are sensitive to the precipitation procedure applied and to the temperature and pressure during compression when pellets are formed. The resulting materials are intimate mixtures or solid solutions of  $\text{PbS}$  or  $\text{CdS}$  with  $\text{Ag}_2\text{S}$ , but in the case of the copper preparation a ternary compound  $\text{Ag}_{1.5}\text{Cu}_{0.5}\text{S}$  (jalpaite) is also formed. Chalcogenide ISEs can be prepared by electrochemical deposition, also with optimized composition, and these electrodes are reproducible and have better performance in flow systems.

The lead electrode is sensitive to surface oxidation and requires periodical repolishing. The linear response for the primary cation usually extends from  $10^{-1}$  to  $10^{-6} \text{ mol l}^{-1}$ ; the detection limit is  $\sim 10^{-7} \text{ mol l}^{-1}$ . These statements hold only in the optimal pH range of  $\sim 4$ – $7$ . At higher pH values the formation of the hydroxides of the divalent metals reduces the upper limit of quantitation. At lower pH the solubility of the sulfides increases and this raises the lower detection limit. The selectivities over alkali and alkaline earth metal ions are very good.

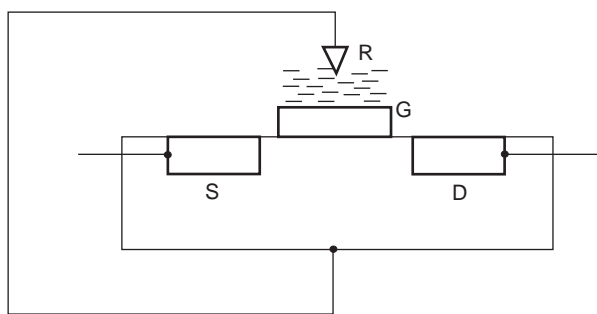
Mercury(II) and silver(I) strongly interfere in every case and copper(II) is an important interferent in sensing lead and cadmium. Copper and lead assays are much more disturbed by the presence of iron(III) than the cadmium assay. A large excess of anions like chloride may interfere with cadmium and lead quantitation but the most notable problem with chloride is found with the copper ISE. The response of these ISEs to complexing agents and complexes of their divalent metal components has been thoroughly studied. The electrode response in the presence of complexes does not always coincide with thermodynamic expectation.

The copper ISE is important for complexation studies in environmental samples. Potentiometric lead and cadmium electrodes are less popular but the lead electrode has been quite useful for indirect sulfate measurements. The electrodes require frequent recalibration.

## Ion-Selective Field-Effect Transistors

Ion-selective electrodes are mostly used in conjunction with a high impedance voltmeter. This means practically that the e.m.f. is measured by charging the gate capacitor of a FET to the cell voltage. The field in the capacitor regulates the source–drain current of the FET, which is measured as the output





**Figure 3** Schematic representation of an ISFET. S, source; D, drain; G, gate; R, reference electrode.

signal. Ion-selective field-effect transistors (ISFETs) are specially designed FETs where the gate material is the ion-selective membrane (Figure 3). For example, the gate can be an oxide insulator, like silicon oxide, which behaves as a pH-sensitive electrode membrane. The proton buffer capacity of this oxide layer will determine the sensitivity of the device. Advantages of ISFETs include the small size, the possibility of mass production, the possibility of integration of sensor and electronics, etc. Their disadvantage is that practical devices usually drift much more than similar electrodes in conventional design. This is a practical problem that can be overcome by compensating for the predetermined drift by a programmed amplifier. Another way to get around the drift problem is to carry out dynamic measurements, e.g., titration of weak acids or bases by integrating the ISFET with a pH actuator (a noble metal electrode that generates protons or hydroxyl ions coulometrically).

## Miscellaneous Designs

Metal chalcogenide solid ISE membranes have been made from single crystals and microcrystalline materials. Several researchers have found that chalcogenide glasses can also be used as ISEs. The characteristics of these materials may surpass those of crystalline materials with respect to chemical stability, selectivity, and sensitivity to redox agents. Sensor arrays of chalcogenide glass electrodes can be constructed by pulsed laser deposition. Artificial neural networks are used for signal processing to detect different heavy metal ions from environmental water samples.

Salts of certain organic radical ions have been found to be suitable for the preparation of solid ISEs, but they appear to have been little used in practice.

Ceramics, like NASICON (sodium super ionic conductor) prepared by the sol-gel method, were found to give selective potentiometric response toward various cations. Selectivity is provided by the

size of the channels, which conduct in the ceramic framework.

Liquid crystals representing an intermediate state between the liquid and crystalline states can also be used to prepare ISEs. A liquid-crystal-type ionophore is dissolved in a liquid crystal solvent to obtain a cation-selective electrode, the selectivity of which is highly temperature dependent.

In recent years, the modification of electrode surfaces has been intensively studied by electrochemists. The trend of surface modification has already reached the ISE field.

Inorganic films, like transition metal hexacyanoferrates, can be chemically or electrochemically deposited onto a conductive surface, and used as ISEs, e.g.,  $K^+$ ,  $Tl^+$ ,  $Cs^+$ ,  $NH_4^+$ , and  $Rb^+$  ions.

Another class of inorganic materials used in the preparation of modified electrodes are zeolite type ion-exchangers possessing shape, size, and charge selectivity. The zeolite is embedded into a polymer matrix and deposited onto a metal surface.

Sol-gel derived glasses, encapsulating neutral carriers were prepared as ion-sensing membranes in ISFETs or as supported membranes in macroelectrodes. Thromobogenicity of the membranes proved to be much better than that of plasticized poly(vinyl chloride) membranes, which makes them successful candidates in the analysis of biological samples.

Successful attempts were made to deposit a molecularly imprinted layer of  $TiO_2$  on the gate of an ISFET by the sol-gel technique. Binding of the analyte to the imprinted inorganic network changed its charge, which in turn controlled the gate potential, giving selective potentiometric response, e.g., to chloroaromatic acids.

Conducting polymers, e.g., polypyrrole, polythiophene, and polyaniline, deposited on electrode surfaces have both ion-exchange and redox properties. In most cases the ionic response is poorly selective, but by carefully choosing the polymerization conditions they can show appreciable selectivity. This is attributed to the micropores formed in the polymer having size and charge complementary to the ion to be measured. Selectivity of the conducting polymers can be greatly enhanced by incorporating ionophores. Conducting polymers can be either electropolymerized or solvent cast onto a substrate like a glassy carbon electrode resulting in an all-solid-state sensor.

The use of methacrylic-acrylic copolymers as a plasticizer-free matrix into which ionophores can be incorporated may speed up the development of miniaturized solid-state ion-sensors. These polymer films can be readily prepared by photocuring, have excellent adhesion properties, and satisfactory long-term stability.



**Table 1** Properties and selected applications of solid ISEs<sup>a</sup>

Electrode type	Concentration range <sup>b</sup> (mol l <sup>-1</sup> )	Approximate slope (mV per decade)	Interferences <sup>c</sup>	Selected applications
Fluoride	10 <sup>-6</sup> –1	–57	OH <sup>-</sup>	Control of fluoridation of drinking water. Dental care products and tooth enamel. Minerals, soil
Chloride	10 <sup>-5</sup> –1	–57	S <sup>2-</sup> , CN <sup>-</sup> , I <sup>-</sup> , Br <sup>-</sup> , S <sub>2</sub> O <sub>3</sub> <sup>2-</sup> , SCN <sup>-</sup>	Etchants containing H <sub>2</sub> F <sub>2</sub> Water analysis, soil extracts Chloride in sweat (cystic fibrosis)
Cyanide (AgI/Ag <sub>2</sub> S)	10 <sup>-6</sup> –10 <sup>-2</sup>	–57	S <sup>2-</sup> , I <sup>-</sup>	Food Analysis Industrial wastewater
Sulfide (for Ag <sup>+</sup> and S <sup>2-</sup> )	10 <sup>-7</sup> –1	57 (Ag <sup>+</sup> ) –28 (S <sup>2-</sup> )	Hg <sup>2+</sup>	Galvanic baths Wastewater (S <sup>2-</sup> analyte sensitive to air oxidation)
Bromide	10 <sup>-6</sup> –1	–57	S <sup>2-</sup> , CN <sup>-</sup> , I <sup>-</sup> , SCN <sup>-</sup>	Food and drinks Argentimetric titrations
Copper	10 <sup>-7</sup> –1	28	Hg <sup>2+</sup> , Ag <sup>+</sup> , Cu <sup>+</sup>	Bromine in organic material after combustion Water analysis
Lead	10 <sup>-6</sup> –10 <sup>-1</sup>	26	Hg <sup>2+</sup> , Ag <sup>+</sup> , Cu <sup>2+</sup>	Compleximetric titrations Water analysis Titration of sulfate

<sup>a</sup> The electrode properties given here have been extracted from several manufacturers' literature. Similar, more detailed tables can be found in the references by Cammann and van der Linden.

<sup>b</sup> The sources usually do not specify linear range and detection limit; only a useful concentration range is given.

<sup>c</sup> Only the strongest interferants are given here. These must be absent or present at a much lower concentration than the analyte.

## Applications

Table 1 shows some data for commercially available electrodes and a selection of interesting analytical uses of solid ISEs. The majority of applications has been the determination of fluoride in many kinds of samples. Chloride determinations are next in order of importance. The precision of direct determinations is usually limited to 1–10% rsd. The ruggedness of solid ISEs makes them suitable for online monitoring applications. If monitoring is based on direct measurement frequent recalibrations may be necessary and the temperature compensation may not always be straightforward. Monitors can also be based on titration reactions; in this case less problem may be encountered but more sophisticated hardware will be necessary. Solid ISEs can also be used in different solvents. This may be useful, e.g., in precipitation titrations.

See also: **Ion-Selective Electrodes:** Overview; Liquid Membrane.

## Further Reading

Bailey PL (1980) *Analysis with Ion-Selective Electrodes*, 2nd edn. London: Heyden.

Bergveld P (2003) Thirty years of ISFETOLOGY. What happened in the past 30 years and what may happen in the next 30 years. *Sensors and Actuators B* 88: 1–20.

Bergveld P and Sibbald A (1988) Analytical and biomedical applications of ion-selective field-effect transistors. In: Svehla G (ed.) *Comprehensive Analytical Chemistry*, vol. XXIII. Amsterdam: Elsevier.

Cammann K (1979) *Working with Ion-Selective Electrodes*. Berlin: Springer.

Karyakin AA (2001) Prussian Blue and its analogues: electrochemistry and analytical applications. *Electroanalysis* 13: 813–819.

Koryta J and Stulik K (1983) *Ion-Selective Electrodes*, 2nd edn. Cambridge: Cambridge University Press.

van der Linden WE (1981) Ion-selective electrodes. In: Svehla G (ed.) *Comprehensive Analytical Chemistry*, vol. XI, ch. 3. Amsterdam: Elsevier.

Lindner E, Tóth K, and Pungor E (1988) *Dynamic Characteristics of Ion-Selective Electrodes*. Boca Raton, FL: CRC Press.

Piletsky SA and Turner APF (2002) Electrochemical sensors based on molecularly imprinted polymers. *Electroanalysis* 14: 317–323.

Pungor E and Tóth K (1970) Ion-selective membrane electrodes. *Analyst* 95: 625.

Walcarius A (1999) Zeolite-modified electrodes in electroanalytical chemistry. *Analytica Chimica Acta* 384: 1–16.

## Liquid Membrane

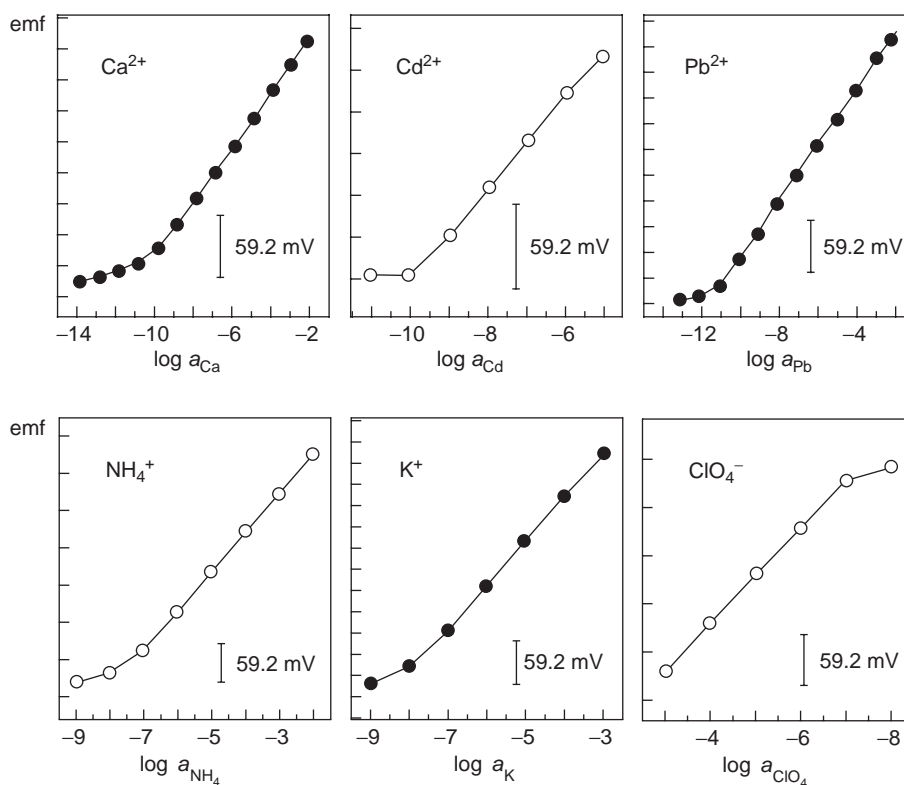
**E Bakker**, Auburn University, Auburn, AL, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Liquid membrane electrodes are a very versatile group of ion-selective electrodes (ISEs) because their response can be rationalized on the basis of liquid-liquid partitioning and complexation principles. Today, liquid membrane ISEs are key components of clinical analyzers in nearly every hospital in the world, with billions of assays performed each year. Their application is steadily broadening with the advent of more selective membrane materials, advances in miniaturization, and the availability of more rugged sensor materials. Recently, numerous liquid membrane electrodes with extremely low detection limits have become available (see **Figure 1**), opening up very important new avenues for this class of sensors. The ion-selective membrane was historically composed of a simple organic liquid, but today either plasticized hydrophobic polymers or plasticizer-free

polymers with low glass transition temperatures are used because of their improved ruggedness. The ion-selective membrane is doped with a number of active ingredients that are either freely dissolved or covalently anchored onto the polymer to avoid leaching. These membrane components are the key ingredients of such sensors: they mediate the selective extraction of the analyte and also make sure that the ISE membrane exhibits ion-exchanger properties. So far, ISEs for more than 60 different analytes have been described. ISEs for polyionic analytes such as heparin, protamine, and DNA have also been developed, demonstrating that they form a truly versatile class of sensors. Fabrication of polymeric ISE membranes is normally done by solvent casting from an organic solution, and can be accomplished with relative ease. The membranes are then either mechanically mounted into an electrode body or glued onto a polymer tubing. The fabrication of microelectrodes is also well established, and numerous microfabrication processes for liquid membrane ISEs have been reported. In traditional setups, an aqueous inner electrolyte solution is placed at the backside of the membrane, but the so-called solid-contact ISEs are



**Figure 1** Response functions of recently described liquid membrane electrodes with very low detection limits in unbuffered samples (Adapted from Bakker E and Pretsch E (2002) The new wave of ion-selective electrodes. *Analytical Chemistry* 74: 420A).

also used, although they may suffer from potential drifts. Two principal classes of liquid membrane ISEs can generally be distinguished: the first contains an ion-exchanger without any molecular receptor properties; the second is based on highly selective ionophores. Modern chemical research mainly deals with the improvement of the second class but numerous commercial ISEs are often still based on the first.

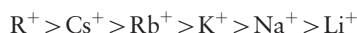
Recommended reviews on this topic are given in Further Reading, and include the biannual reviews in *Analytical Chemistry* on electrochemical sensors, a series of three comprehensive reviews on ionophore-based electrodes, a review on low detection limit ISEs, and recommendations for reporting selectivities including recent listings of ISE selectivities.

### **Ion-Exchanger-Based ISEs**

Liquid or solvent polymeric membranes for ISEs must exhibit ion-exchanger properties. It is known that the concentration of analyte in the membrane phase must remain approximately constant as the sample concentration is changed for a Nernstian response slope to be observed. This can be achieved by doping the liquid membrane with a lipophilic ion-exchanger, for example, tridodecylmethylammonium chloride for anion-selective electrodes or potassium tetrakis(4-chlorophenyl)borate for cation-selective electrodes. Numerous other ion-exchangers have been reported, but their main function is always the same. Before initial use, the liquid membrane is normally conditioned in a solution containing a high concentration of the cation or anion to be measured. During this conditioning period, the hydrophilic counterion of the ion-exchanger in the membrane is replaced with the analyte ion of interest. Because no chemical recognition/complexation takes place, the selectivity of such ISEs is a function of the free energy of solvation of the measured ions in both phases. For anion-selective ISEs (containing an anion-exchanger salt), the observed selectivity sequence always follows the so-called Hofmeister sequence:



where  $\text{R}^-$  is an organic anion. The corresponding selectivity sequence for ion-selective membranes incorporating a cation-exchanger is



Such ion-exchanger-based ISEs have been mainly used for the detection of lipophilic ions such as perchlorate and nitrate, as well as certain electrically

charged drugs and lipophilic quaternary ammonium ions. Note that the thermodynamic selectivity sequences listed above are not always observed in practice. If a strongly interfering but dilute ion is present in the sample (typically at less than  $10^{-4} \text{ mol l}^{-1}$  levels), it does not alter the cell potential significantly. Indeed, low concentrations of these ions are rapidly depleted at the membrane surface and can only insufficiently displace the analyte of interest from the membrane. For only short exposure times, the electrode can often still reliably be used to assess the analyte. Of course, prolonged exposure to samples containing such lipophilic interferents will eventually alter the response characteristics of such ISEs. These characteristics have often been used for sensing applications in samples for which no thermodynamic selectivity was available. This explains why some manufacturers offer ion-exchanger-based ISEs for a variety of ions, even though the basic membrane composition is essentially identical. Clearly, it is much more advisable to use ISEs with better ion-exchange selectivities, which may be achieved with membranes containing suitable ionophores.

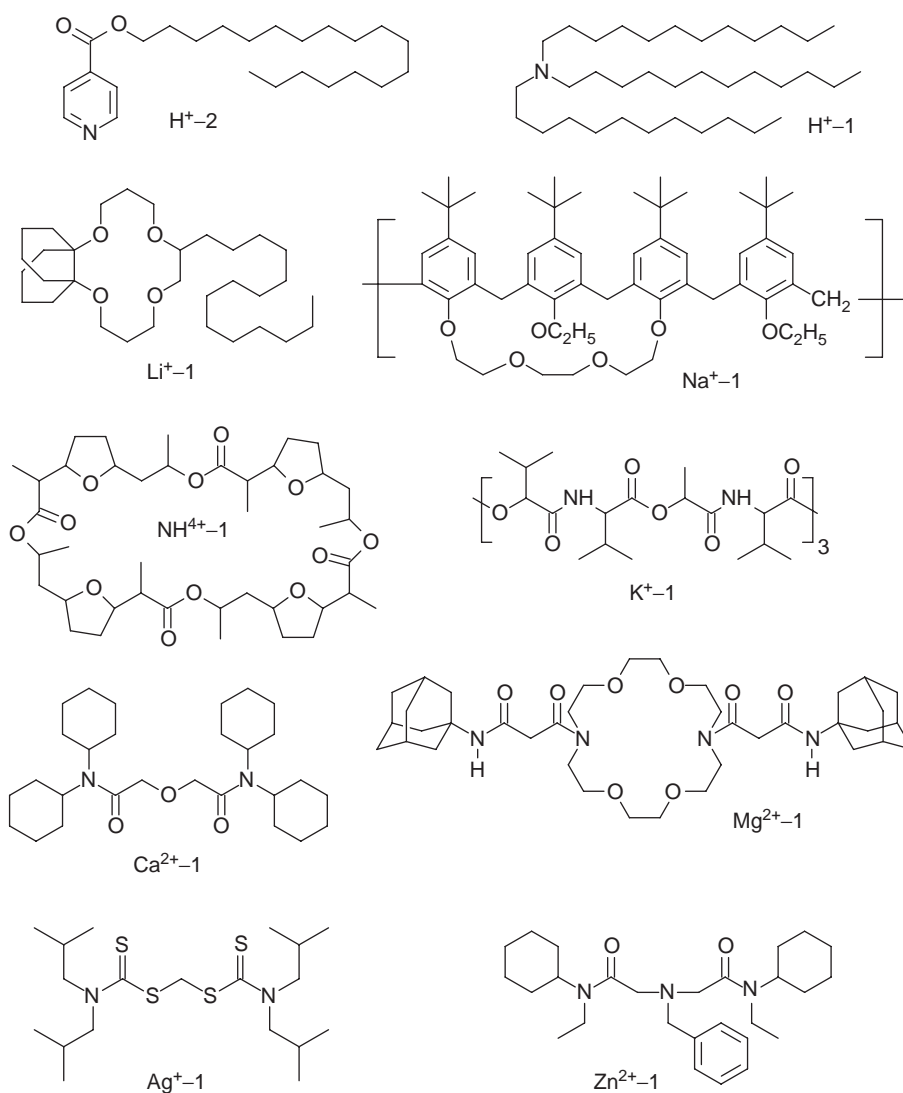
### **Ionophore-Based ISEs**

Lipophilic ionophores are essential to achieving a high sensing selectivity with liquid membrane ISEs. As explained above, ion-exchanger-based membranes always show the same selectivity pattern that follows the solvation energies of the ions, but ionophore-based membranes may show very different selectivities. This is achieved by the formation of strong complexes between the extracted analyte ion and the ionophore in the membrane. Complex formation constants have been recently determined in the membrane phase in the range of  $\sim 10^8$  for monovalent to  $\sim 10^{25}$  for divalent ions. With ionophore-based ISEs, the selectivity is now dictated by the free energy of transfer of the ions from the aqueous to the membrane phase, the complex formation constants between the extracted ions and the ionophore, and the concentrations of the active membrane components in the membrane. Because solvation energies are still an important component to the overall selectivity since an ion extraction process is involved, it is normally more difficult to design ISEs for hydrophilic ions than it is for hydrophobic ones. On the other hand, it is often a challenge to design selective receptors for large, bulky ions. Consequently, ionophore-based ISEs for potassium and calcium have been realized early on, while selective sensors for magnesium, lithium, sodium, chloride, and phosphate have only been developed recently or

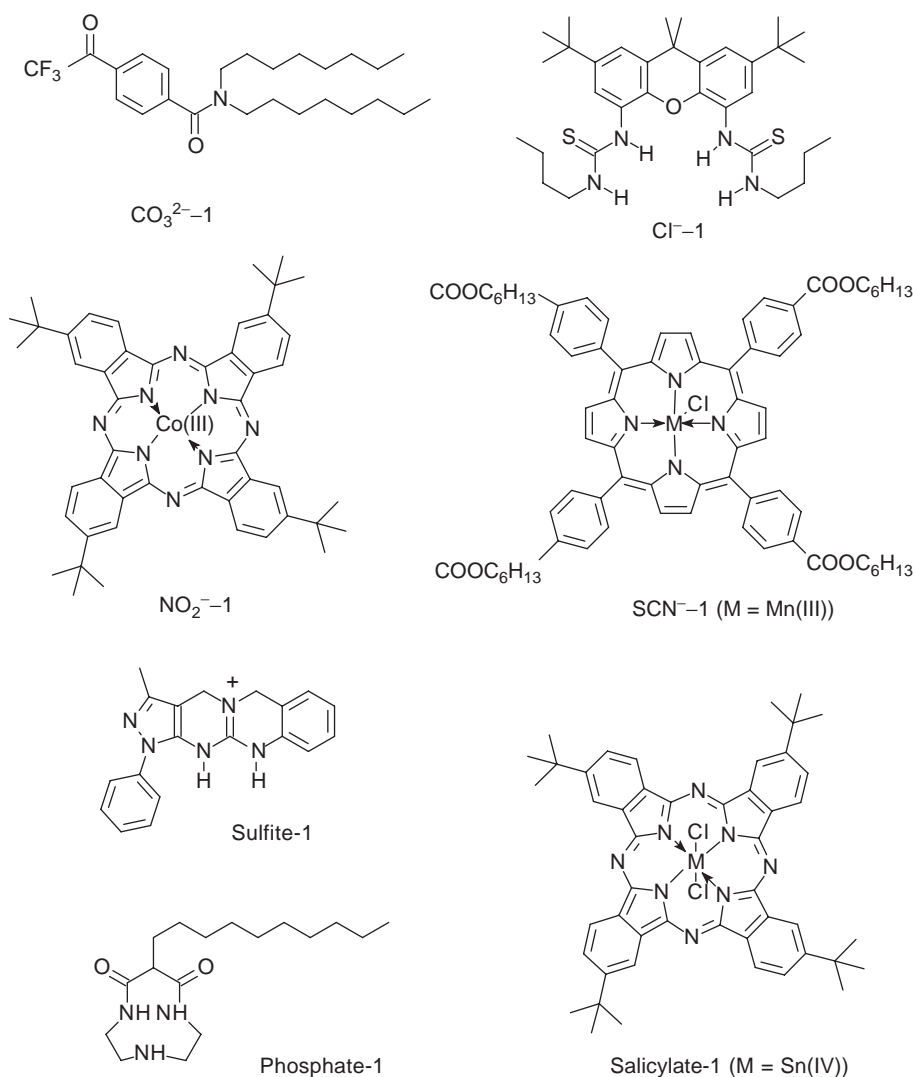
are still topics of current research. Ionophore-based ISEs for bulky anions such as perchlorate are essentially unknown, but note that ion-exchanger-based ISEs possess a good intrinsic selectivity for such anions (see above).

Figures 2 and 3 show a limited selection of successful ionophores used in liquid membrane ISEs. A much exhaustive list of ionophores has been published in the recent literature. Table 1 summarizes typical membrane compositions of the corresponding ISEs and lists the observed selectivity coefficients. Recall that a small selectivity coefficient indicates a high membrane selectivity: for an interfering ion possessing the same charge as the so-called primary ion, the selectivity coefficient is a weighting factor for the interfering ion activity. In ideal cases, a selectivity

coefficient of  $10^{-4}$ , for example, indicates that a 10 000-fold excess of interfering ion will give the same signal as the primary ion. The relationship is more complicated for ions of different charge or close to the detection limit of the ISE. As shown in Table 1, membrane selectivities may be very high, and explain why such ISEs have been very successful in a number of important applications. Ionophores can be developed on the basis of a variety of recognition principles: bis-crown ionophores, crowns with bulky side groups to prohibit intermolecular sandwich formation, noncyclic amide and thioamide ionophores, basket-shaped calix[4]arene and calix[6]arene ionophores, calixarenes with crown bridges, thiocarbamates, lipophilic bases as  $H^+$ -ionophores, guanidinium derivatives, multitopic



**Figure 2** A selection of cation-selective ionophores used in liquid membrane ISEs (Adapted from Bühlmann P, Pretsch E, and Bakker E (1998) *Chemical Reviews* 98: 1593).



**Figure 3** A selection of anion-selective ionophores used in liquid membrane ISEs (Adapted from Bühlmann P, Pretsch E, and Bakker E (1998) *Chemical Reviews* 98: 1593).

hydrogen bond forming ionophores, metalloporphyrins, corrins, and phthalocyanines with different metal centers and a variety of axial ligands, and aromatic trifluoroacetyl derivatives for the recognition of hydrophilic nucleophiles have been utilized in liquid membrane ISEs. Some of these ionophores are electrically charged in their uncomplexed form, but most are the so-called neutral ionophores.

Ionophores used successfully in ISEs normally share a number of important characteristics. They are highly lipophilic, that is, they are strongly retained in the hydrophobic membrane phase to ensure a long lifetime of the sensor. This is most often achieved by adding long alkyl chains, cyclohexyl, or adamantyl substituents to the structure. They possess a polar moiety or a set of polar functional groups responsible for ion recognition. The rest of the molecule contains hydrophobic regions that are

compatible with the surrounding membrane solvent. The historical argument that ionophore molecules must also exhibit a high mobility has been largely disproved by the availability of a number of membrane materials where the ionophore is covalently anchored onto the polymeric backbone. The covalent anchoring still retains certain ion mobility in the membrane that appears to be sufficient for ISE applications.

The membrane selectivity is not only a function of the ionophore. Membranes of relatively high polarity (such as poly(vinyl chloride) overplasticized with *o*-nitrophenylether) have historically been preferred for the development of divalent-ion-selective electrodes and many anion-selective electrodes. Nonpolar membranes, such as poly(vinyl chloride) overplasticized with bis(2-ethylhexyl)sebacate, are generally more suited for monovalent cations. Note, however, that



**Table 1** Membrane compositions and selectivity coefficients of selected liquid membrane ion-selective electrodes<sup>a</sup>

Analyte	Membrane composition	$\log K_{IJ}^{pot}$
H <sup>+</sup>	Tri- <i>n</i> -dodecylamine, KTpClPB, PVC, <i>bis</i> -2-ethylhexylsebacate	Na <sup>+</sup> : -10.4 K <sup>+</sup> : -9.8 Ca <sup>2+</sup> : < -11.1
H <sup>+</sup>	Octadecyl isonicotinate, KTpClPB, PVC, <i>ortho</i> -nitrophenyloctylether	Li <sup>+</sup> : -6.9 Na <sup>+</sup> : -5.6 K <sup>+</sup> : -4.4
Li <sup>+</sup>	7-Tetradecyl-2,6,9,13-tetraoxatricyclo[12.4.4.0 <sup>1,14</sup> ]docosane, KTpClPB, PVC, bis(benzylphenyl)adipate	Na <sup>+</sup> : -3.1 K <sup>+</sup> : -3.6 NH <sub>4</sub> <sup>+</sup> : -3.8 Ca <sup>2+</sup> : < -5.0
Na <sup>+</sup>	Calix[4]arene crown-4 ionophore, KTpClPB, PVC, <i>ortho</i> -nitrophenyloctylether	Li <sup>+</sup> : -2.8 K <sup>+</sup> : -5.0 NH <sub>4</sub> <sup>+</sup> : -4.4 Mg <sup>2+</sup> : -4.5 Ca <sup>2+</sup> : -4.4
K <sup>+</sup>	Valinomycin, NaTFPB, PVC, <i>bis</i> -2-ethylhexylsebacate	Na <sup>+</sup> : -4.5 Mg <sup>2+</sup> : -7.5 Ca <sup>2+</sup> : -6.9
NH <sub>4</sub> <sup>+</sup>	Nonactin/Monactin, KTpClPB, PVC, <i>ortho</i> -nitrophenyloctylether	Li <sup>+</sup> : -2.9 Na <sup>+</sup> : -2.3 K <sup>+</sup> : -1.1 Mg <sup>2+</sup> : -4.0 Ca <sup>2+</sup> : -4.0
Mg <sup>2+</sup>	Double armed dizazacrown ether ionophore (see Mg <sup>2+</sup> - 1 in <b>Figure 1</b> ), KTpClPB, PVC, <i>ortho</i> -nitrophenyloctylether	Li <sup>+</sup> : -3.7 Na <sup>+</sup> : -3.2 K <sup>+</sup> : -1.4 NH <sub>4</sub> <sup>+</sup> : -2.0 Ca <sup>2+</sup> : -2.5
Ca <sup>2+</sup>	<i>N,N',N'</i> -tetracyclohexyl-3-oxapentanediamide (ETH 129), KTpClPB, PVC, <i>ortho</i> -nitrophenyloctylether	Na <sup>+</sup> : -8.3 K <sup>+</sup> : -10.1 Mg <sup>2+</sup> : -9.3
Ag <sup>+</sup>	Methylenebis(diisobutylthiocarbamate), NaTFPB, PVC, <i>bis</i> -2-ethylhexylsebacate	Na <sup>+</sup> : -8.7 K <sup>+</sup> : -8.2 Ca <sup>2+</sup> : -11.0 Cu <sup>2+</sup> : -10.5 Pb <sup>2+</sup> : -10.3
Zn <sup>2+</sup>	<i>N</i> -Benzyliminodiacetic acid bis( <i>N</i> -ethyl- <i>N</i> -cyclohexylamide), KTpClPB, PVC, <i>ortho</i> -nitrophenyloctylether	Li <sup>+</sup> , Na <sup>+</sup> : -2.7 K <sup>+</sup> : -2.5 NH <sub>4</sub> <sup>+</sup> : -3.1 Ca <sup>2+</sup> : -2.8 Cd <sup>2+</sup> : -3.6 Pb <sup>2+</sup> : -2.1
Cl <sup>-</sup>	2,7-Di- <i>tert</i> -butyl-9,9-dimethyl-4,5-xanthenediamine, TDDMACl, PVC, <i>ortho</i> -nitrophenyloctylether	Sal <sup>-</sup> : +1.8 SCN <sup>-</sup> : +1.6 NO <sub>3</sub> <sup>-</sup> : +0.7 HCO <sub>3</sub> <sup>-</sup> : -2.6
HSO <sub>3</sub> <sup>-</sup>	Guanidinium derivative ionophore, PVC, <i>ortho</i> -nitrophenyloctylether	ClO <sub>4</sub> <sup>-</sup> : -2.2 Cl <sup>-</sup> : < -3.0 Sal <sup>-</sup> : -2.3
SCN <sup>-</sup>	Chloro[5,10,15,20-tetrakis[4-(hexyloxycarbonyl)phenyl]porphyrinato]manganese(III), PVC, ETH 469	ClO <sub>4</sub> <sup>-</sup> : -2.0 I <sup>-</sup> : -2.3 NO <sub>3</sub> <sup>-</sup> : -3.6 NO <sub>2</sub> <sup>-</sup> : -3.0 Cl <sup>-</sup> : -3.4 HCO <sub>3</sub> <sup>-</sup> : -5.1
CO <sub>3</sub> <sup>2-</sup>	<i>N,N</i> -Dioctyl-4-trifluoroacetylbenzamide, MTDDACl, PVC, <i>bis</i> -2-ethylhexylsebacate	SCN <sup>-</sup> : +1.0 NO <sub>3</sub> <sup>-</sup> : -1.6 Br <sup>-</sup> : -3.5

Continued

**Table 1** Continued

Analyte	Membrane composition	$\log K_{IJ}^{pot}$
$\text{NO}_2^-$	2,9,16,23-Tetra- <i>tert</i> -butylphthalocyanine)cobalt(III), hexadecyltrioctylammonium iodide, PVC, dibutylphthalate	$\text{Cl}^-$ : -5.0
		$\text{Sal}^-$ : +3.3
		$\text{SCN}^-$ : -1.0
		$\text{I}^-$ : -1.6
		$\text{NO}_3^-$ : -3.1
		$\text{Cl}^-$ : -3.5
Phosphate	3-Decyl-1,5,8-triazacyclodecane-2,4-dione, PVC, dibutylsebacate	$\text{Br}^-$ : -2.9
		$\text{SCN}^-$ : -2.3
		$\text{NO}_3^-$ : -2.8
		$\text{Cl}^-$ : -2.3
Salicylate	(2,9,16,23-Tetra- <i>tert</i> -butylphthalocyanine)tin(IV), PVC, dinonyl sebacate	$\text{OAc}^-$ : -3.2
		$\text{ClO}_4^-$ : -3.3
		$\text{SCN}^-$ : -2.9
		$\text{Cl}^-$ : -4.8
		$\text{OAc}^-$ : -3.4

<sup>a</sup> See Bühlmann P, Pretsch E, and Bakker E (1998) Carrier-based ion-selective electrodes and bulk optodes: 2. Ionophores applied in potentiometric and optical sensors. *Chemical Reviews* 98: 1593 and references cited therein.

**Table 2** Optimum concentration ratio of lipophilic anionic sites to neutral ionophore giving optimal selectivity for the analyte ion over an interfering ion<sup>a</sup>

Charge of cation		Stoichiometry of ligand-ion complex		Optimum ratio of sites to ionophore
<i>I</i> (analyte), $z_i$	<i>J</i> (interfering ion), $z_j$	<i>I</i> (analyte), $n_i$	<i>J</i> (interfering ion), $n_j$	
2	2	1	2	1.41
2	2	2	3	0.77
2	2	3	4	0.54
2	1	1	1	1.62
2	1	2	2	0.73
2	1	3	3	0.46
1	1	1	2	0.71

<sup>a</sup> The optimum ratios are dependent on the charges of both ions and the respective complex stoichiometries (number of ligands per ion) in the membrane.

Eugster R, Gehrig PM, Morf WE, Spichiger UE, and Simon W (1991) *Analytical Chemistry* 63: 2285.

the role of membrane polarity on the membrane selectivity has been overrated in the past as there appears to be no direct correlation when a large number of plasticizers are compared. Clearly, other parameters appear to influence selectivity as well, including the tendency to form ion pairs, the availability of functional groups on the plasticizer that can compete with the ionophore, and indirect variations of complex stoichiometries of the ionophore in different solvent environments. For these reasons, the optimization of ISE selectivity is, therefore, to a large part still an empirical science. On the other hand, if the complex stoichiometries are known, optimum membrane concentrations of ionophore and lipophilic ion-exchanger can be predicted. For example, Table 2 lists optimum concentration ratios of ionophore to ion-exchanger for a select number of assumed complex stoichiometries and charges of the two compared ions. Since some ionophores may be capable of forming mixed complex stoichiometries,

these values are best viewed as first guesses to guide the experimentalist.

## Low Detection Limits

Recent research has focused on improving the detection limit of liquid membrane ISEs in samples containing a low total concentration of analyte ion (low detection limits in samples containing a metal ion buffer could always be achieved with ISEs). It was found that the detection limit of highly selective ISEs is often dictated by an elevated concentration of primary ion at the membrane surface that is generated from a zero current ion flux from the membrane into the sample. This flux often originates from the asymmetrical setup of the electrode, where a relatively concentrated inner solution is placed at the backside of the membrane. Even extremely small concentration gradients across the membrane can lead to a

significant ion flux that may concentration polarize the sample solution at the membrane surface. Today, the detection limit for numerous ISEs has been improved to nanomolar and sometimes even picomolar concentrations, placing this class of electrodes for the first time among the most sensitive electrochemical detection systems available.

See also: **Ion-Selective Electrodes:** Overview; Glass; Solid-State; Gas Sensing Probes; Enzyme Electrodes; Clinical Applications; Food Applications; Water Applications.

## Further Reading

- Bakker E and Telting-Diaz M (2002) Electrochemical sensors. *Analytical Chemistry* 74: 2781.
- Bakker E, Bühlmann P, and Pretsch E (1997) Carrier-based ion-selective electrodes and bulk optodes: 1. General characteristics. *Chemical Review* 97: 3083.
- Bakker E, Bühlmann P, and Pretsch E (1999) Polymer membrane ion-selective electrodes—What are the limits? *Electroanalysis* 11: 915.
- Bakker E, Pretsch E, and Bühlmann P (2000) Selectivity of potentiometric ion sensors. *Analytical Chemistry* 72: 1127.
- Bobacka J, Ivaska A, and Lewenstam A (2003) Potentiometric ion sensors based on conducting polymers. *Electroanalysis* 15: 366.
- Buck RP and Lindner E (1995) Recommendations for nomenclature of ion-selective electrodes (IUPAC recommendations 1994). *Pure and Applied Chemistry* 66: 2527.
- Frant MS (1994) History of the early commercialization of ion-selective electrodes. *Analyst* 119: 2293.
- Umezawa Y, Bühlmann P, Umezawa K, Tohda K, and Amemiya S (2000) Potentiometric selectivity coefficients of ion-selective electrodes: Part I. Inorganic cations—technical report. *Pure and Applied Chemistry* 72: 1851.
- Umezawa Y, Umezawa K, Bühlmann P, Hamada N, Aoki H, Nakanishi J, Sato M, Xiao KP, and Nishimura Y (2002) Potentiometric selectivity coefficients of ion-selective electrodes: Part II. Inorganic anions. *Pure and Applied Chemistry* 74: 923.
- Umezawa Y, Bühlmann P, Umezawa K, and Hamada N (2002) Potentiometric selectivity coefficients of ion-selective electrodes: Part III. Organic ions. *Pure and Applied Chemistry* 74: 995.

## Gas Sensing Probes

**F Opekar and K Štulík**, Charles University, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A number of very important pollutants of the environment are gases, the most prominent of them being the oxides of nitrogen and sulfur; moreover, the content of carbon dioxide in the atmosphere is a highly significant factor affecting the global climate. Furthermore, determination or continuous monitoring of various gases is indispensable in industry, medicine, and many other fields. Consequently, simple, rapid, and reliable methods must be available for monitoring of gases. One of the choices is potentiometry using probes based on ion-selective electrodes (ISEs) and related sensors.

In general, any ion-selective potentiometric sensor can be used to monitor gases, provided that it is immersed in an electrolyte solution containing the ion sensed and whose activity changes on dissolution of the gas to be determined. Such a potentiometric cell, containing an indicating ion-selective sensor and a reference electrode immersed in an electrolyte, and

exchanging only gases with the environment to be analyzed, is termed the gas-sensing probe. In addition to these potentiometric probes with liquid electrolytes, solid-state gas sensors can be used that employ either solid electrolytes or are based on measurement of changes in the electrical parameters of semiconductors in the presence of a test gas.

Gas probes are used not only for monitoring of substances that are gaseous under normal ambient conditions, but also for determining solid compounds and compounds contained in solutions that are converted into gaseous products by a preceding (bio)chemical reaction; the latter determinations are advantageous in their high selectivity. Therefore, potentiometric gas probes are often used as sensing systems in biosensors.

## Principles

There are several experimental arrangements in potentiometric determination of gases.

### Determination in Gas-Stripping Solutions

This is the simplest system. The gaseous medium is passed through an absorption solution of an

significant ion flux that may concentration polarize the sample solution at the membrane surface. Today, the detection limit for numerous ISEs has been improved to nanomolar and sometimes even picomolar concentrations, placing this class of electrodes for the first time among the most sensitive electrochemical detection systems available.

See also: **Ion-Selective Electrodes:** Overview; Glass; Solid-State; Gas Sensing Probes; Enzyme Electrodes; Clinical Applications; Food Applications; Water Applications.

## Further Reading

- Bakker E and Telting-Diaz M (2002) Electrochemical sensors. *Analytical Chemistry* 74: 2781.
- Bakker E, Bühlmann P, and Pretsch E (1997) Carrier-based ion-selective electrodes and bulk optodes: 1. General characteristics. *Chemical Review* 97: 3083.
- Bakker E, Bühlmann P, and Pretsch E (1999) Polymer membrane ion-selective electrodes—What are the limits? *Electroanalysis* 11: 915.
- Bakker E, Pretsch E, and Bühlmann P (2000) Selectivity of potentiometric ion sensors. *Analytical Chemistry* 72: 1127.
- Bobacka J, Ivaska A, and Lewenstam A (2003) Potentiometric ion sensors based on conducting polymers. *Electroanalysis* 15: 366.
- Buck RP and Lindner E (1995) Recommendations for nomenclature of ion-selective electrodes (IUPAC recommendations 1994). *Pure and Applied Chemistry* 66: 2527.
- Frant MS (1994) History of the early commercialization of ion-selective electrodes. *Analyst* 119: 2293.
- Umezawa Y, Bühlmann P, Umezawa K, Tohda K, and Amemiya S (2000) Potentiometric selectivity coefficients of ion-selective electrodes: Part I. Inorganic cations—technical report. *Pure and Applied Chemistry* 72: 1851.
- Umezawa Y, Umezawa K, Bühlmann P, Hamada N, Aoki H, Nakanishi J, Sato M, Xiao KP, and Nishimura Y (2002) Potentiometric selectivity coefficients of ion-selective electrodes: Part II. Inorganic anions. *Pure and Applied Chemistry* 74: 923.
- Umezawa Y, Bühlmann P, Umezawa K, and Hamada N (2002) Potentiometric selectivity coefficients of ion-selective electrodes: Part III. Organic ions. *Pure and Applied Chemistry* 74: 995.

## Gas Sensing Probes

**F Opekar and K Štulík**, Charles University, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A number of very important pollutants of the environment are gases, the most prominent of them being the oxides of nitrogen and sulfur; moreover, the content of carbon dioxide in the atmosphere is a highly significant factor affecting the global climate. Furthermore, determination or continuous monitoring of various gases is indispensable in industry, medicine, and many other fields. Consequently, simple, rapid, and reliable methods must be available for monitoring of gases. One of the choices is potentiometry using probes based on ion-selective electrodes (ISEs) and related sensors.

In general, any ion-selective potentiometric sensor can be used to monitor gases, provided that it is immersed in an electrolyte solution containing the ion sensed and whose activity changes on dissolution of the gas to be determined. Such a potentiometric cell, containing an indicating ion-selective sensor and a reference electrode immersed in an electrolyte, and

exchanging only gases with the environment to be analyzed, is termed the gas-sensing probe. In addition to these potentiometric probes with liquid electrolytes, solid-state gas sensors can be used that employ either solid electrolytes or are based on measurement of changes in the electrical parameters of semiconductors in the presence of a test gas.

Gas probes are used not only for monitoring of substances that are gaseous under normal ambient conditions, but also for determining solid compounds and compounds contained in solutions that are converted into gaseous products by a preceding (bio)chemical reaction; the latter determinations are advantageous in their high selectivity. Therefore, potentiometric gas probes are often used as sensing systems in biosensors.

## Principles

There are several experimental arrangements in potentiometric determination of gases.

### Determination in Gas-Stripping Solutions

This is the simplest system. The gaseous medium is passed through an absorption solution of an

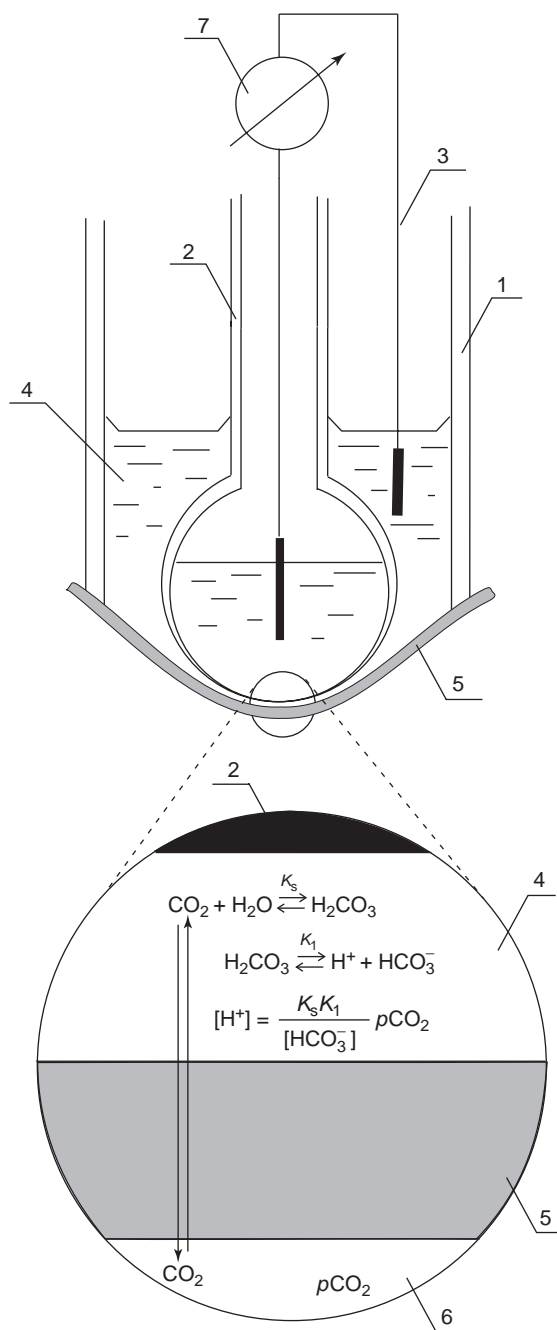
electrolyte and the change in the solution composition is monitored with an ion-selective sensor, which is either directly immersed in the absorption vessel or is placed in a flow cell to which the absorption solution is fed by a pump. The advantages of this approach are the possibilities of preconcentration of the analyte in the absorption solution and the obtaining of integral concentration values of the analyte over a certain time period. On the other hand, the response is sluggish and the selectivity of determination may sometimes be unsatisfactory.

#### Determination with Sensors Separated from the Test Medium

This is the most common situation and three basic approaches are outlined below.

**Permeation units** The gaseous test substance is absorbed from the atmosphere (or generated from a nongaseous analyte by a (bio)chemical reaction) in a donor solution that is fed to a permeation unit containing a gas-permeable membrane separating the donor solution from an acceptor solution. The composition of the donor solution is chemically modified to liberate the gas again, the gas then permeates from the donor solution into the acceptor solution, and the latter, whose composition is suitable for potentiometric detection, is transported to the indicating sensor system. The gas-transfer efficiency depends on the gas partition coefficient between the two solutions, the gas mass-transfer rate through the membrane (i.e., on the kind of gas, membrane material, and membrane thickness), the relative velocities of the donor and acceptor solutions, and the membrane surface area. Hence, the solution composition can be independently optimized from the point of view of the gas absorption and the potentiometric detection; analyte preconcentration can be attained through optimization of the solution flow rates and selectivity of determination improved by combining the selectivities of the gas absorption (or generation), the gas transport through the membrane, and the potentiometric detection itself. These permeation units are especially useful in flow analyses (continuous monitoring, flow injection analysis, and continuous-flow analysis).

**Compact gas probes** In compact gas probes (combination electrodes), the ion-selective sensor is covered with a gas-permeable membrane enclosing a thin film of electrolyte containing the ion sensed (the Severinghaus-type probe, **Figure 1**). A gas from the test medium passes through the membrane, dissolves in the electrolyte film, and changes the activity of the ion sensed. The type of ion-selective sensor,



**Figure 1** A simplified scheme of the  $\text{CO}_2$  gas probe (Severinghaus type). 1, Detector body; 2, indicator electrode (pH glass electrode); 3, reference electrode; 4, internal electrolyte; 5, gas-permeable membrane; 6, medium analyzed; and 7, voltmeter (pH-meter). (The components of the probe are not drawn to a real scale.)

the electrolyte composition, and the kind of gas-permeable membrane are selected according to the test gas (e.g., some gases undergo proteolytic reactions in the electrolyte film resulting in a change in the pH which is sensed by a pH sensor). The analytical parameters of these probes are complex functions of many factors. The response time, selectivity,



sensitivity, and the limit of detection depend on the properties of the membrane (i.e., the permeability coefficient, which is the product of the gas diffusion coefficient in the membrane and the gas partition coefficient between the test medium and the membrane material, and the membrane thickness), on the composition, concentration, and thickness of the electrolyte film, on the sample matrix, and also on the analyte concentration. Therefore, the probe design and experimental conditions must always be optimized for given analytical problems.

As many analytically important gases are acidic or basic, the most common sensor is the glass pH ISE. The shape of the ISE sensing surface is important from the point of view of the quality of the electrolyte film, which is usually obtained by pressing the sensing part of the glass ISE immersed in an electrolyte solution on the gas-permeable membrane. The film should be as thin as possible, to decrease the response time – a mildly convex ISE sensing surface is most convenient for obtaining thin electrolyte films without dead spaces (Figure 1).

Two principal kinds of gas-permeable membrane are used, namely, homogeneous polymers through which the gas is transported by molecular diffusion after dissolution in the membrane material (typical diffusion coefficients are given in Table 1) and microporous membranes, where the pores are larger than the gas molecules and thus the test gas is transported by diffusion in the gaseous phase within the membrane pores. Homogeneous membranes may contribute to the measurement selectivity, but the mass transport through them is slow; from this point of view, silicone rubber membranes are advantageous, as the diffusion coefficients are relatively high (see Table 1).

Microporous membranes create virtually no diffusion barrier, since the gas diffusion coefficients in the gaseous phase are of the order of  $0.1 \text{ cm}^2 \text{ s}^{-1}$ . Hence, the response of probes with these membranes is

faster, but the membrane transport provides no selectivity and the diffusion geometry is not particularly stable. Microporous membranes must not be wetted by the solutions present, as the flooding of the membrane pores causes a drastic suppression of the probe response. Microporous membranes also permit passage of water vapor. If the water vapor pressures in the sample and the probe electrolyte film, and/or the sample and electrolyte temperatures are not equal, then water is transported through the membrane, the electrolyte film thickness and its concentration vary, and the signal drifts; when gases are monitored in the gaseous phase, the electrolyte film can dry up completely. The thinner the electrolyte film, the more pronounced are these effects.

Finally, the mechanical strength of the membrane must be considered, as the electrolyte film thickness and its reproducibility depend on the constancy and reproducibility of the pressure of the ISE on the membrane. Polytetrafluoroethylene, polypropylene, polyethylene, and silicone rubber are the most common membrane materials.

The composition of the electrolyte film depends on the kind of equilibrium reaction in which the gaseous analyte participates. The limits of detection (LODs) are commonly  $\sim 10^{-5}$ – $10^{-6} \text{ mol l}^{-1}$ . Decreasing the electrolyte concentration can decrease the LOD value, but then the dynamic range of the response is limited at higher analyte concentrations. Useful electrolyte concentrations range from  $10^{-2}$  to  $10^{-5} \text{ mol l}^{-1}$ .

The LOD value further depends on the kind and concentrations of the interferents present. The effect of interferents depends on the length of the measuring interval. During short-time measurements, the selectivity is primarily determined by the ratio of the permeability coefficients of the analyte and the interferents in the membrane material, while after a certain time, in steady state, it is given by the ratio of the dissociation constants of the analyte and the interferents in the electrolyte solution.

The response rate depends on the magnitude of the change in the analyte concentration and also on its direction. It increases with increasing magnitude of the analyte concentration change and is greater and more reproducible when the analyte concentration increases. In general, the time constants are rather high, ranging from tens of seconds to several minutes.

Replacement of glass pH electrodes by pH-sensitive and other ion-selective polymer membranes can be considered as an attempt at a fundamental change in the approach to the construction of membrane gas probes. The use of these membranes allows innovative structural designs leading to miniaturization and to mass production of low-cost

**Table 1** Diffusion coefficients (in  $10^7 \text{ cm}^2 \text{ s}^{-1}$ ) of selected gases in various membrane materials

Gas	Membrane material		
	Polytetrafluoroethylene	Polyethylene	Silicone rubber
O <sub>2</sub>	1.4	9.0	160.0
H <sub>2</sub>	2.6		
CO <sub>2</sub>	1.9	8.0	110.0
NO <sub>2</sub>	0.7		
SO <sub>2</sub>	2.1		
NO	1.6		
CO	0.9		

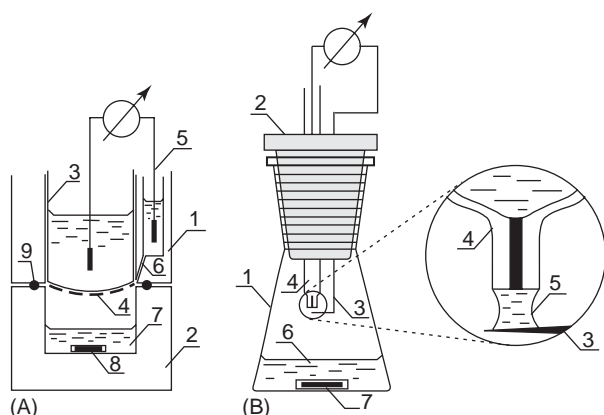
Data taken from Opekar F (1985) Application of metallized membrane electrodes for the electrochemical analysis of gases and solutions. *Chemické Listy* 79: 703–718. More data can be found in the *Polymer Handbook*.

disposable detectors especially appreciated in clinical use. However, the parameters of the membranes tested have not met the requirements for routine use.

**Air-gap detectors** In air-gap detectors, the sensing electrode is separated by air from the test medium (Figure 2). These detectors are useful for analyses of substances in solution, after converting the analytes to gaseous products. The gaseous product passes from the solution into the air, diffuses to the sensing electrode, and dissolves in the film of an electrolyte solution covering the electrode surface. The electrolyte film on the sensor is obtained by touching the sensor surface, e.g., by a sponge soaked with the required electrolyte solution (Figure 2A), or a drop of an electrolyte is placed between a sensing and a

reference electrode (e.g., in the determination of cyanide, in which the sensing electrode – a silver wire, is connected with the tip of a reference electrode by a drop of a buffered solution of  $\text{KAg}(\text{CN})_2$  – Figure 2B).

**Detectors based on field-effect transistor** Ion-selective field effect transistor (ISFET) or chemically sensitive field effect transistor (CSFET) may replace classical ISE in Severinghaus-type gas-sensing probes; the use of pH-sensitive ISFETs is most common. Because of their small dimensions, ISFET-based gas-sensing systems have been tested for, e.g., in *in vivo* monitoring of the contents of a clinically important gas,  $\text{CO}_2$ , in blood. Similar to gas probes with polymeric ion-selective membranes, these systems have not yet found extensive use in practice. Direct detection of gases by FET sensors (e.g., using palladium-gate devices or electron work-function sensors) is advantageous and practically important, but the principles of these techniques are outside the field of ion-selective sensing.



**Figure 2** Air-gap detectors. (A) Detector with an electrolyte film. 1, detector body; 2, sample vessel; 3, indicator (pH) electrode; 4, electrolyte film; 5, reference electrode; 6, salt bridge; 7, sample; 8, stirring bar; and 9, seal. (From Růžicka J and Hansen EH (1974) A new potentiometric gas sensor – the air-gap electrode. *Analytica Chimica Acta* 69: 129–141.) (B) Detector with an electrolyte drop. 1, Conical flask; 2, stopper with electrodes; 3, indicator electrode (Ag wire); 4, reference electrode (SCE); 5, electrolyte drop; 6, sample; and 7, stirring bar. (After Fligier J, Czichon P, and Gregorowicz Z (1980) A very simple air-gap cyanide sensor. *Analytica Chimica Acta* 118: 145–148.)

## Determination of some Important Gases

The most common applications of membrane gas-sensing probes are summarized in Table 2.

### Nitrogen Oxides

The content of nitrogen oxides ( $\text{NO}_x$ ) in the air usually varies from 1 to over  $100 \mu\text{g m}^{-3}$ . The nitrogen dioxide ( $\text{NO}_2$ ) gas-sensing probe suffers from interferences and thus the nitrate ISE is used more often for determination, after suitable trapping and conversion of  $\text{NO}_x$ .

Nitrogen oxides can be chemisorbed on lead dioxide heated to  $190^\circ\text{C}$ , with formation of lead(II) nitrate, which is extracted into hot water and then measured with a nitrate ISE. Carbon dioxide, water

**Table 2** A survey of the most common membrane gas-probes (the probes for gases printed in bold letters are commercially available)

Gas	Principal component of the internal electrolyte	Equilibrium reaction	Indicator electrode
<b><math>\text{NO}_2</math></b>	$\text{NaNO}_3$	$2\text{NO}_2 + \text{H}_2\text{O} \rightleftharpoons 2\text{H}^+ + \text{NO}_3^- + \text{NO}_2^-$	Glass pH
<b><math>\text{SO}_2</math></b>	$\text{NaHSO}_3$	$\text{SO}_2 + \text{H}_2\text{O} \rightleftharpoons \text{H}^+ + \text{HSO}_3^-$	Glass pH, antimony
<b><math>\text{CO}_2</math></b>	$\text{NaHCO}_3$	$\text{CO}_2 + \text{H}_2\text{O} \rightleftharpoons \text{H}^+ + \text{HCO}_3^-$	Glass pH, antimony, $\text{CO}_3^{2-}$ ISE
<b><math>\text{NH}_3</math></b>	$\text{NH}_4\text{Cl}$	$\text{NH}_3 + \text{H}_2\text{O} \rightleftharpoons \text{NH}_4^+ + \text{OH}^-$	
	Me (Ag, Cu, Hg)	$n\text{NH}_3 + \text{Me}^{n+} \rightleftharpoons \text{Me}(\text{NH}_3)_n^+$	
$\text{H}_2\text{S}$	Buffer (pH 5)	$\text{H}_2\text{S} \rightleftharpoons \text{HS}^- + \text{H}^+$	
$\text{Cl}_2$	$\text{HSO}_3^-$ buffer	$\text{Cl}_2 + \text{H}_2\text{O} \rightleftharpoons \text{Cl}^- + 2\text{H}^+ + \text{ClO}^-$	
$\text{HCN}$	$\text{KAg}(\text{CN})_2$	$\text{Ag}(\text{CN})_2^- \rightleftharpoons \text{Ag}^+ + 2\text{CN}^-$	
$\text{HF}$	Acid	$\text{HF} \rightleftharpoons \text{H}^+ + \text{F}^-$	
	$\text{Fe}^{3+}/\text{Fe}^{2+}$	$\text{FeF}_x^{2-x} \rightleftharpoons \text{FeF}_y^{3-y} + (x-y)\text{F}^-$	
$\text{N}_3\text{H}$	$\text{NaN}_3 + \text{NaCl}$	$\text{N}_3\text{H} \rightleftharpoons \text{H}^+ + \text{N}_3^-$	

vapor, and nitrogen are not sorbed and possible interference from hydrogen chloride is eliminated by the addition of lead(II) fluoride; however, sulfur dioxide causes a positive error of determination when present at a more than 10-fold excess over  $\text{NO}_x$ .

Another way of selectively trapping  $\text{NO}_x$  is oxidative absorption, e.g., by passing air through a 2% hydrogen peroxide solution for 24 h at a flow rate of  $2 \text{ l min}^{-1}$ . Excess hydrogen peroxide is removed by adding manganese dioxide and the nitrate formed in the absorption solution is determined with a nitrate ISE; sulfur dioxide does not interfere up to a 40-fold excess. These methods are, of course, not quite adequate for continuous sensing.

A personal  $\text{NO}_x$  dosimeter employs trapping of the gases on a glass filter impregnated with sodium dichromate and sulfuric acid ( $\text{NO}$  is oxidized to  $\text{NO}_2$ ), followed by passage of the  $\text{NO}_2$  through a silicone rubber membrane into an acidic solution of hydrogen peroxide in which nitrate is formed and measured by a nitrate ISE.

Nitrogen dioxide can also be selectively detected by an electrode with a chalcogenide ( $\text{Se}_{60}\text{Ge}_{28}\text{Sb}_{12}$ ) membrane, without interference from nitric oxide, sulfur dioxide, carbon monoxide, methane, and other gases. Another solid-state sensor for  $\text{NO}_2$  employs an alkaline nitrate electrolyte at a temperature of  $800^\circ\text{C}$ .

### Sulfur Dioxide

The content of sulfur dioxide ( $\text{SO}_2$ ) in the atmosphere varies widely and may reach values as high as  $1\text{--}5 \text{ mg m}^{-3}$  in heavily polluted air. Direct monitoring with an  $\text{SO}_2$  gas probe is often insufficiently sensitive and thus  $\text{SO}_2$  must be preconcentrated by absorption in a suitable solution. The most common procedure is based on absorption of  $\text{SO}_2$  in a tetrachloromercurate(II) solution (TCM), in which  $\text{Hg}(\text{SO}_3)_2^{2-}$  is formed. The pH of the absorption solution is maintained at a value of 6.9 that ensures virtually complete absorption of  $\text{SO}_2$ . Prior to analysis, amidosulfonic acid is added to remove nitrite and the pH is decreased to about unity, so that most of the sulfite present is converted into free sulfur dioxide, which is determined by an  $\text{SO}_2$  probe. With very low  $\text{SO}_2$  contents, a filter is soaked with a TCM solution and air is passed through the filter at a high flow rate and for a rather long time. These methods provide discontinuous values of the  $\text{SO}_2$  concentration in regular intervals.

Other possibilities include absorption of  $\text{SO}_2$  in a 3% solution of hydrogen peroxide followed by titration of the solution with a lead(II) standard solution (indication – lead(II) ISE), or the reduction of iodine by  $\text{SO}_2$  with determination of the iodide formed with an iodide ISE.

A solid-state sensor employs a solid alkali sulfate electrolyte at an elevated temperature ( $800^\circ\text{C}$ ).

### Carbon Dioxide

The gas-sensing probe for carbon dioxide ( $\text{CO}_2$ ) is the original Severinghaus sensor. The dynamic range of this determination is rather limited and the limit of detection is  $\sim 10^{-4} \text{ mol l}^{-1}$ , as the concentration of atmospheric  $\text{CO}_2$  dissolved in water is  $1\text{--}3 \times 10^{-5} \text{ mol l}^{-1}$ .

### Other Important Gases

Ammonia is best determined with a gas-sensitive probe (see Table 2), which is highly selective and exhibits a limit of detection of above  $5 \times 10^{-7} \text{ mol l}^{-1}$ . Volatile amines may interfere in the determination. When determining ammoniacal nitrogen in solution, the test solution is made alkaline ( $\text{pH} \geq 11.2$ ), so that the  $[\text{NH}_4^+]/[\text{NH}_3]$  ratio is  $< 100$  and acidic gases (such as carbon dioxide, sulfur dioxide, and hydrogen sulfide) are converted into the corresponding anions and cannot interfere by diffusion through the probe membrane. The use of an air-gap sensor for ammonia prevents certain interferences, e.g., from surfactants when determining ammoniacal nitrogen in waste waters.

Hydrogen sulfide can be absorbed in an alkaline aqueous solution and then determined with a high sensitivity by a sulfide ISE (LOD of  $\mu\text{g m}^{-3}$  in the atmosphere). Sensors based on chalcogenide membranes have been also tested for the detection of  $\text{H}_2\text{S}$ . The solutions containing hydrogen sulfide must be protected against atmospheric oxidation.

Chlorine can be determined using a gas-sensitive probe (Table 2) or by absorption in an  $\text{Ag}(\text{CN})_2^-$  solution and measurement with an iodide ISE or, better, a sulfide ISE.

Hydrogen fluoride is determined either by a gas probe (Table 2), or, more often, by a fluoride ISE (lanthanum fluoride based) after absorption in a suitable medium. Hydrogen fluoride emissions can be measured down to a content of  $0.1 \mu\text{g m}^{-3}$ .

**See also:** **Enzymes:** Enzyme-Based Electrodes. **Flow Injection Analysis:** Detection Techniques. **Ion-Selective Electrodes:** Glass. **Nitrogen.** **Sulfur.**

### Further Reading

- Bailey PL and Riley M (1975) Performance characteristics of gas-sensing probes. *Analyst* 100: 145–156.
- Janata J (1989) *Principles of Chemical Sensors*. New York: Plenum.
- Morf WE (1981) *The Principles of Ion-selective Electrodes and Membrane Transport*. Amsterdam: Elsevier.

- Oehme F (1991) Liquid electrolyte sensors: potentiometry, amperometry, and conductometry. In: Göpel W, Hese J, and Zemel JN (eds.) *Sensors. A Comprehensive Survey*, vol. 2, ch. 7. Weinheim: VCH.
- Riley M (1979) Gas-sensing probes. In: Covington AK (ed.) *Ion-selective Electrode Methodology*, vol. 2, ch. 1. Boca Raton, FL: CRC Press.
- Ross JW, Riseman JH, and Kruger JA (1973) Potentiometric gas sensing electrodes. *Pure and Applied Chemistry* 36: 473–487.
- Veselý J and Štulík K (1987) Potentiometry with ion-selective electrodes. In: Kalvoda R (ed.) *Electroanalytical Methods in Chemical and Environmental Analysis*, ch. 6. New York: Plenum.

## Enzyme Electrodes

**M Mascini and I Palchetti**, Università degli Studi di Firenze, Firenze, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Potentiometric enzyme electrodes form a large group of catalytic biosensors, resulting from a thin layer of enzyme integrated within or intimately associated with a potentiometric transducer. The enzyme generates a reaction that enables determination of the substrate (the target analyte), providing a highly selective and sensitive method for the determination of a given, ionic or nonionic, substrate. The determination of the electrode potential gives a direct indication of the concentration of the analyte. The signal obtained is proportional to the logarithm of the concentration, following the Nernst law.

### General Features of Enzyme Electrodes

#### Potentiometric Transducers

Guilbault and Montalvo were the first, in 1969, to detail a potentiometric enzyme electrode. They described a urea biosensor based on urease immobilized at an ammonium-selective liquid membrane electrode. Since then, over hundreds of different applications have appeared in the literature, due to the significant development of ion-selective electrodes (ISEs) observed during the last 30 years. The electrodes used to assemble a potentiometric biosensor include glass electrodes for the measurement of pH or monovalent ions, ISEs sensitive to anions or cations, gas electrodes such as the CO<sub>2</sub> or the NH<sub>3</sub> probes, and metal electrodes able to detect redox species; some of these electrodes useful in the construction of potentiometric enzyme electrodes are listed in Table 1.

In the 1980s, ion-selective membranes were used for the first time as the molecular recognition element

of sensors based on field-effect transistor (so called ion-selective field-effect transistor, ISFET).

Nowadays, ion-selective sensors are constructed using various advanced methods such as thick- and thin-film technology. A thick-film sensor comprises layers of special pastes (thickness 10–50 µm) deposited onto an insulating substrate; the method of film deposition is the screen-printing technique. Thin-film electrochemical transducers are built up by the successive deposition and patterning of dielectric and conductive materials, on top of an optically flat and polished substrate. The deposition of the thin metallic films (thickness 10–200 nm) is generally carried out by classical evaporation or sputtering of a solid metal source; then, photolithography and other techniques can be envisaged to pattern the electrode material. Using these technologies, microfabricated arrays of sensors and biosensors have been developed.

Interesting features of potentiometric transducers are simplicity of instrumentation (only a potentiometer is needed), low cost, selectivity, and the non-destructive nature of the analytical procedure.

#### Biosensor Assembly

In order to prepare the biosensor, the enzyme has to be immobilized in the form of a thin layer at the surface of the transducer. Among the several methods for immobilization available, the following are the most generally employed: (1) physical entrapment within an inert polymeric membrane (in this case the enzyme is mixed with a monomer solution, which is then polymerized to a gel – polyacrylamide gel, starch, agar gel, etc., thus trapping the enzyme) or behind a membrane (in this case the enzyme solution is simply confined by an analyte permeable membrane, such as a dialysis membrane, as a thin film covering the indicator electrode; this method is also called micro-encapsulation); (2) covalent bonding of the enzymes on membranes or surfaces activated by means of bi-functional groups or spacers, such as glutaraldehyde, carbodiimide (when the enzyme is bound to the solid support using a bi- or multifunctional agent, such as



- Oehme F (1991) Liquid electrolyte sensors: potentiometry, amperometry, and conductometry. In: Göpel W, Hese J, and Zemel JN (eds.) *Sensors. A Comprehensive Survey*, vol. 2, ch. 7. Weinheim: VCH.
- Riley M (1979) Gas-sensing probes. In: Covington AK (ed.) *Ion-selective Electrode Methodology*, vol. 2, ch. 1. Boca Raton, FL: CRC Press.
- Ross JW, Riseman JH, and Kruger JA (1973) Potentiometric gas sensing electrodes. *Pure and Applied Chemistry* 36: 473–487.
- Veselý J and Štulík K (1987) Potentiometry with ion-selective electrodes. In: Kalvoda R (ed.) *Electroanalytical Methods in Chemical and Environmental Analysis*, ch. 6. New York: Plenum.

## Enzyme Electrodes

**M Mascini and I Palchetti**, Università degli Studi di Firenze, Firenze, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Potentiometric enzyme electrodes form a large group of catalytic biosensors, resulting from a thin layer of enzyme integrated within or intimately associated with a potentiometric transducer. The enzyme generates a reaction that enables determination of the substrate (the target analyte), providing a highly selective and sensitive method for the determination of a given, ionic or nonionic, substrate. The determination of the electrode potential gives a direct indication of the concentration of the analyte. The signal obtained is proportional to the logarithm of the concentration, following the Nernst law.

### General Features of Enzyme Electrodes

#### Potentiometric Transducers

Guilbault and Montalvo were the first, in 1969, to detail a potentiometric enzyme electrode. They described a urea biosensor based on urease immobilized at an ammonium-selective liquid membrane electrode. Since then, over hundreds of different applications have appeared in the literature, due to the significant development of ion-selective electrodes (ISEs) observed during the last 30 years. The electrodes used to assemble a potentiometric biosensor include glass electrodes for the measurement of pH or monovalent ions, ISEs sensitive to anions or cations, gas electrodes such as the CO<sub>2</sub> or the NH<sub>3</sub> probes, and metal electrodes able to detect redox species; some of these electrodes useful in the construction of potentiometric enzyme electrodes are listed in Table 1.

In the 1980s, ion-selective membranes were used for the first time as the molecular recognition element

of sensors based on field-effect transistor (so called ion-selective field-effect transistor, ISFET).

Nowadays, ion-selective sensors are constructed using various advanced methods such as thick- and thin-film technology. A thick-film sensor comprises layers of special pastes (thickness 10–50 µm) deposited onto an insulating substrate; the method of film deposition is the screen-printing technique. Thin-film electrochemical transducers are built up by the successive deposition and patterning of dielectric and conductive materials, on top of an optically flat and polished substrate. The deposition of the thin metallic films (thickness 10–200 nm) is generally carried out by classical evaporation or sputtering of a solid metal source; then, photolithography and other techniques can be envisaged to pattern the electrode material. Using these technologies, microfabricated arrays of sensors and biosensors have been developed.

Interesting features of potentiometric transducers are simplicity of instrumentation (only a potentiometer is needed), low cost, selectivity, and the non-destructive nature of the analytical procedure.

#### Biosensor Assembly

In order to prepare the biosensor, the enzyme has to be immobilized in the form of a thin layer at the surface of the transducer. Among the several methods for immobilization available, the following are the most generally employed: (1) physical entrapment within an inert polymeric membrane (in this case the enzyme is mixed with a monomer solution, which is then polymerized to a gel – polyacrylamide gel, starch, agar gel, etc., thus trapping the enzyme) or behind a membrane (in this case the enzyme solution is simply confined by an analyte permeable membrane, such as a dialysis membrane, as a thin film covering the indicator electrode; this method is also called micro-encapsulation); (2) covalent bonding of the enzymes on membranes or surfaces activated by means of bi-functional groups or spacers, such as glutaraldehyde, carbodiimide (when the enzyme is bound to the solid support using a bi- or multifunctional agent, such as



**Table 1** Potentiometric enzyme electrodes

<i>Electrochemical sensor</i>	<i>Enzyme</i>	<i>Metabolite</i>	<i>Measurement range (mmol l<sup>-1</sup>)</i>
Glass, pH	Urease	Urea	0.05–5
	Glucose oxidase	Glucose	0.1–1
	Penicillinase	Penicillin	0.01–3
	Acetylcholinesterase	Acetylcholine	0.01–10
Antimony, pH	Urease	Urea	0.1–10
ISFET, pH	Urease	Urea	1–1000
	Glucose oxidase	Glucose	1–50
	Penicillinase	Penicillin	1–20
NH <sub>3</sub> probe	Urease	Urea	0.05–50
	Glutaminase	Glutamine	0.05–5
	Phenylalanine ammonia lyase	Phenylalanine	0.05–1
	Asparaginase	Asparagine	0.1–10
	AMP deaminase	5'-AMP	0.1–10
	Phosphodiesterase + AMP deaminase	3',5'-cyclic-AMP	0.01–10
	Creatinine deaminase	Creatinine	0.1–10
	Urease	Urea	0.1–10
CO <sub>2</sub> probe	Uricase	Uric acid	0.1–2.5
	Decarboxylating enzymes	Tyrosine	0.1–2.5
		Asparagine	0.1–10
		Lysine	0.1–30
		Glutamate	0.5–5
	Urease	Urea	0.01–10
Ammonium (neutral carrier based)	Creatininase	Creatinine	0.01–5
	Asparaginase	Asparagine	0.01–1
	Glucose oxidase	Glucose	0.01–10
Iodide	Amino acid oxidase	Amino acid	0.01–10
	Cholesterol oxidase	Cholesterol	0.01–1
Redox	Laccate dehydrogenase	Lactate	0.1–1
Cyanide	$\beta$ -Glucosidase	Amygdalin	0.1–100

gluteraldehyde, which forms a bridge between biocatalytic species or proteins, the technique is also called cross-linking or reticulation); and (3) immobilization on commercially available activated membranes such as Immudodyne produced by Pall Industries, or Ultra-Bind produced by Gelman Sciences.

Several enzymes can be immobilized within the same reaction layer, in order to increase the range of possible biosensor analytes, provide efficient regeneration of enzyme cosubstrates, or to improve the biosensor selectivity by decreasing the local concentration of electrochemical interfering substances.

Generally, the technique of chemical immobilization (covalent bonding) is preferred. This procedure improves the enzyme stability and defines a diffusion layer where the product, formed by the catalytic action of the enzyme, diffuses partly to the electrode surface and partly back to the bulk of the solution; therefore, by defining the steady state one obtains a reproducible potential value. However, the real advantage of the enzyme immobilization through chemical bonding is the lifetime of the final probe, which can reach 6–12 months.

An enzyme electrode operates via a five-step process: (1) transport of the substrate to the surface of the

electrode; (2) diffusion of the substrate through the membrane to the enzyme-active site; (3) a reaction occurring at the active site; (4) product formation in the enzymatic reaction and its transport through the membrane to the surface of the electrode; and (5) measurement of the product at the electrode surface.

The first step, transport of the substrate, is most critically dependent on the stirring rate of the solution, so that rapid stirring will bring the substrate very rapidly to the electrode surface. If the membrane is kept very thin using highly active enzymes, then steps (2) and (4) are eliminated or minimized; since step (3) is very fast, the response of an enzyme electrode should theoretically approach the response time of the base sensor. Many researchers have shown experimentally that one can approach this behavior using a thin membrane and rapid stirring.

## Analytical Features of Enzyme Electrodes

### Concentration Range

The maximum concentration of substrate that can be measured with an enzyme electrode is governed by  $K_m$  (the Michaelis–Menten constant). If the

**Table 2**  $K_m$  values reported for soluble enzymes and maximum concentration values of the linear concentration range reported for each biosensor probe

Enzyme	Substrate	$K_m$ (mmol l <sup>-1</sup> )	Concentration value (mmol l <sup>-1</sup> )
Urease	Urea	10	10
Glucose oxidase	Glucose	7.7	35
L-Amino acid oxidase	Leucine	1.0	10
Invertase	Sucrose	0.45	1
Alkaline phosphatase	<i>p</i> -Nitrophenylphosphate	0.1	1
Lactate oxidase	Lactate	0.7	20
Pyruvate oxidase	Pyruvate	1.7	2
Creatininase	Creatinine	0.33	1
Acetylcholinesterase	Acetylcholine	0.09	1
Uricase	Uric acid	0.017	1

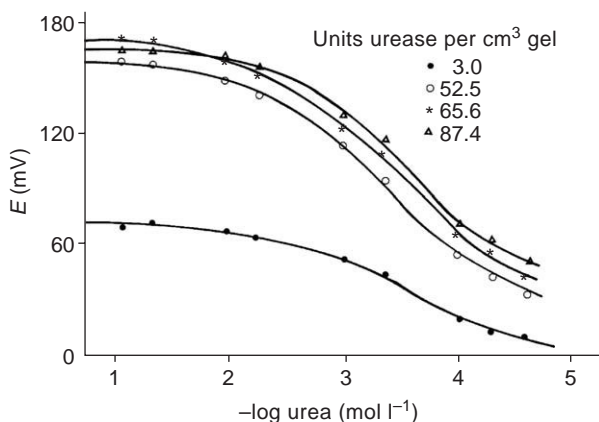
concentration is near to  $K_m$ , the linear relation between substrate concentration and the extent of enzyme reaction fails and a limiting value is reached. However, the picture is not always as straightforward as this. Table 2 lists the maximum concentration values of substrates measured with enzyme probes and compares them with  $K_m$  data for soluble enzymes. In general, the maximum concentration of a substrate exceeds the  $K_m$  value by approximately one order of magnitude.

The difficulty of measuring high concentration is usually overcome by diluting the sample. This procedure has the advantage of the control of pH and ionic strength by the use of suitable buffers, so that samples and standards are measured under identical conditions.

Immobilization of the enzyme is thought to increase the  $K_m$  value, the increase being related to change of substrate/carrier, diffusion effects, or changes in enzyme structure.

The minimum concentration is often related to the minimum amount of electroactive substance that can be measured with the electrochemical probe. Generally, with ISE or gas potentiometric probes the limit of concentration measurable is  $10^{-5}$  mol l<sup>-1</sup>. The calibration graphs are therefore often S-shaped, leveling off at high concentration due to the  $K_m$  or  $K_m$ -related maximum concentration measurable and at low values by the performances of the potentiometric base probe (Figure 1).

One of the most important parameters concerning the concentration range is the activity of the enzyme in the membrane layer. As a general rule, the slope of the sensor will increase with increasing amount of enzyme up to a limiting value. The activity of the enzyme determines the percentage of reaction of the substrate, and hence the amount of electroactive species. However, with increasing amounts of enzyme loading, the percentage of substrate reacting reaches a limit that matches its complete transformation.

**Figure 1** Calibration of a potentiometric urease electrode for urea. Curves are related to different loadings.

The amount of enzyme to be used depends on the immobilization procedure and on the enzyme purity. As a rule of thumb,  $\sim 10$ – $20$  IU per membrane is generally sufficient to give an excellent response. For the less stable, soluble, and physically entrapped enzyme electrodes, an excess of enzyme should be used, that is,  $50$  IU cm<sup>-2</sup> of electrode surface, so that a loss of enzyme does not affect the observed responses. Likewise, purified enzymes should be used so as to promote fast response rates.

### Response Time

Mathematical models describing response times have been considered in the literature. This is the essential feature of an assembled probe, and depends on the procedure of assembling the probe, specifically on the method of immobilization, the electrochemical probe employed, the geometry of the measuring cell, and the kinetics of the enzymatic reaction.

### Diffusion of Substrate

The difference in response between an unstirred solution and a stirred one can decrease the response time from 10 to 1–2 min or less. Also, the potential is

a function of stirring rate related to the amount of substance brought to the electrode surface and to its reactivity. Hence, for fast response times and steady states, a fast but constant rate of stirring of the substrate solution is recommended.

### Substrate Concentration

The rate of enzymatic reaction increases with substrate concentration; for example, in the case of a cyanide electrode coupled with immobilized  $\beta$ -glucosidase enzyme, a response time of 20 s for  $10^{-1} \text{ mol l}^{-1}$  amygdalin and 1 min for  $10^{-4} \text{ mol l}^{-1}$  amygdalin is obtained. Rather than waiting until an equilibrium potential is reached, the rate of potential change ( $\Delta E/\Delta t$ ) can be measured, the result being proportional to substrate concentration.

### Enzyme Concentration

There is a twofold effect on increasing the enzyme activity in the layer that is in close proximity to the electrode surface. First, a complete conversion of the substrate into products will be ensured; second, the response time of the electrode will be affected. The increase in the amount of enzyme affects the thickness of the membrane. This results in an increase in the time required for the substrate to diffuse through the membrane. Hence, for the best results, it is recommended that an enzyme of as high an activity as possible is used in order to ensure rapid kinetics by achieving the thinnest possible membrane.

### pH Effect

Every enzyme has an optimum pH at which it is most active and a certain pH range within which it does not exhibit reactivity. The pH ranges of immobilized enzymes are different from those of soluble enzymes because of the microenvironment. Highly negatively charged carriers create a lower pH at the boundary layer between the carrier and the bulk solution. Thus, the enzyme is in a more acidic environment than the bulk solution. The reverse occurs with positively charged supports. Even for uncharged supports, the overall charges of the enzyme may be sufficient to cause changes in the apparent pH optimum. However, in general, the greater the charge on the support, the greater is the effect, particularly with charged substrates. The optimum pH should be used for obtaining the fastest response, but this is not always possible because the electrode sensor may not itself give the optimum response at the pH of the enzyme reaction. Therefore, a compromise is usually necessary. This is especially important for the ammonia and carbon dioxide electrodes. For ammonia electrochemical sensors, samples should have a pH

>9 in order to obtain complete conversion of  $\text{NH}_4^+$  into  $\text{NH}_3$ , and to obtain a shorter response time of the electrochemical probe. However, few enzymes can function at such high pH, so a compromise is necessary. With the carbon dioxide sensor, the samples should be kept at  $\text{pH} < 5$  to obtain a complete conversion to carbon dioxide of hydrogencarbonate; however, these low pH values often interfere with the enzyme reaction, and a compromise is necessary between optimum pH for enzyme activity and optimum pH range of response of the probe.

### Temperature Effect

The effect of increasing temperature is twofold: an increase in the rate of reaction gives faster response time and a shift in the equilibrium value due to variations in the equilibrium constant. Electrochemical sensors of the gas-permeable membrane type (ammonia and carbon dioxide) lead to an additional effect, since the gas membrane and the features of the diffusion are sensitive to temperature variations. Despite the above considerations, a classical bell-shaped curve is almost always obtained when recording the response of the probe as function of temperature. Room temperature, or  $25^\circ\text{C}$  (controlled to  $\pm 0.2^\circ\text{C}$  is recommended), is often employed when using an electrochemical probe, although when using a gas permeable membrane control to  $\pm 0.1^\circ\text{C}$  is required.

### Thickness of the Enzyme Membrane

The time required to reach steady-state potential reading is dependent on the enzyme layer thickness because of the diffusion parameter for the substrate to reach the active sites of the enzyme and of the electroactive species to diffuse through the membrane to the sensor. A mathematical model relating the thickness of the membrane,  $d$ , the diffusion coefficient,  $D$ , the Michaelis constant,  $K_m$ , and the maximum velocity of the enzyme reaction,  $V_{\max}$ , has been developed:

$$V_{\max}d^2/DK_m = V$$

where  $V$  compares the rate of chemical reaction in the membrane with the rate of diffusion through the membrane. Larger  $V$  indicates faster enzyme catalysis relative to the diffusion process.

Membranes as thin as possible are recommended for the best results, and this is helped by the use of a highly active enzyme.

### Effect of Additional Membranes

Besides the enzymatic reacting layer, many biosensors, especially designed for biological or clinical

applications, incorporate one or more membranes, which serve mainly two important functions: (1) Protective barrier: to prevent the interference of molecules, such as proteins or cells of biological samples, with the reaction layer or to decrease the influence of electrochemical species detected by the transducer. It also reduces leakage of the reacting layer components into the sample solution. (2) Diffusional outer barrier for the substrate: the linear dynamic ranges may be large if the sensor response is controlled by the substrate diffusion through the membrane and not by the enzyme kinetics. This is achieved by placing a thin outer membrane over a highly active enzyme layer.

Generally, the thinner the membranes, the shorter is the biosensor response time.

### Electrode Probe

A limiting factor on the speed of response of an enzyme-based electrode is the response time of the base electrode. Electrodes based on gas-permeable membranes (ammonia or carbon dioxide) have longer response times than those without such membranes (ISEs, pH glass electrodes, ISFETs). In both cases, the response time of the electrode is the limiting factor at low substrate concentration. Response times of the order of 10–30 s can easily be obtained for concentrations in the range  $0.1\text{--}1\text{ mmol l}^{-1}$ .

### Washing Time of Probe

Because of a build-up of product in the enzyme membrane, enzyme electrodes require washing prior to contact with the next sample. The washing time varies from just 20 s for urease in conjunction with an ammonia electrode to as long as 10 min for urease with a pH electrode. The washing time increases with enzyme membrane thickness, as also observed for additional membranes discussed above. It will also depend on the enzyme used and on the characteristic of the base sensor itself, and will be affected by diffusion and kinetic effects. The use of flow injection analysis simplifies the procedures, since the carrier stream serves to wash out between samples.

Deterioration of the enzyme electrode can be seen in three changes in the response characteristics: (1) with age the upper limit will decrease from  $\sim 10^{-1}$  to  $10^{-2}\text{ mol l}^{-1}$ ; (2) the slope of the calibration curve of potential versus  $\log(\text{concentration})$ , 59 mV per decade, Nernstian, will drop to 50, 40, or perhaps 30 mV per decade or lower; (3) the response time of the electrode, originally 30 s to 2 min (approximately the same as that of the base sensor), will become longer as the enzyme ages.

### Interference Effects

For each enzyme-based sensor interferences fall into two categories, namely interference of the electrochemical probe and interference with the enzyme reaction.

#### Interference of the Electrochemical Probe

In the development of a urea electrode, the first approach was the use of a cation glass electrode, but the resulting probe could not be used for assaying urea in blood and urine because the sensor responded to sodium and potassium ions.

The sensor was improved by using the non-actin-based ammonium ISE. This ionophore has selectivity constants  $K_{\text{NH}_4, \text{K}}$  of 0.15 and  $K_{\text{NH}_4, \text{Na}}$  of  $1.3 \times 10^{-3}$ , thus partially eliminating the response of these ions by the sensor. Several articles have been published on this principle, taking into account the residual effect of sodium and potassium.

By using an ammonia electrode with a gas-permeable membrane that effectively excludes ions, it is possible to obtain the desired selectivity. Several assemblies have been studied to obtain rapid and reliable sensors.

The carbon dioxide probe can be used as a transducer for urea (Table 1); it is selective because the gas-permeable membrane excludes ions and other chemicals. However, many samples, especially in biological or clinical chemistry, have large levels of bicarbonate, which will interfere with this probe.

ISEs display different problems of interference when used as enzyme-based probes. In the cyanide sensor for amygdalin, sulfide can interfere, while an iodide electrode used for measuring glucose is susceptible to the presence of thiocyanate, sulfide, cyanide, and silver ions. Moreover, oxidizable compounds present in blood (e.g., uric acid, tyrosine, ascorbic acid, and iron(II)) compete for hydrogen peroxide. These compounds need to be removed by sample pretreatment. Finally, with regard to pH electrodes coupled to enzymes (Table 1), any acidic or basic component will interfere, as will a variable buffer capacity as found in clinical and biological samples. However, by adjusting the pH before initiating the enzyme reaction, and assuming that only the enzyme reaction will give a pH change, the effects can be minimized.

#### Interferences with the Enzyme Reaction

These fall into two classes, namely those where other substrates are present in the sample and those involving base enzyme activation or inhibition. With some enzymes, such as urease, the only substrate that reacts at a reasonable rate is urea; hence, a urease electrode is almost specific for urea. Likewise, uricase is almost specific for uric acid.

Others, such as penicillinase, will promote the reaction of many substrates. Thus, ampicillin, nafcillin, penicillin G, penicillin V, cyclibillin, and dicloxacillin can be determined with a penicillinase electrode. D- and L-amino acid oxidases are even less selective. The former, when coupled to an electrode, responds to D-phenylalanine, D-alanine, D-valine, D-methionine, D-leucine, D-norleucine, and D-isoleucine, while the latter responds to L-leucine, L-tyrosine, L-phenylalanine, L-tryptophan, and methionine.

For samples containing several substrates, a preliminary separation should be considered or the total must be determined. In the case of L-amino acid assay the use of decarboxylating enzymes acting selectively on different amino acids is an attractive possibility, and enzyme electrodes of this type are known for L-tyrosine, L-phenylalanine, and L-tryptophan. These sensors will be coupled with a carbon dioxide base sensor.

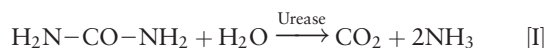
Enzyme inhibitors also need to be taken into account. These include heavy metal ions, such as silver, mercury, and sulfhydryl-reacting organic compounds, such as *p*-chloromercuribenzoate (due to their reaction with the free S-H groups at the active site of many enzymes, especially the oxidases).

## Examples of Enzyme Electrodes Based on ISEs

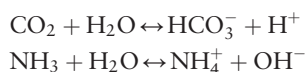
Table 1 lists enzyme electrodes that have been prepared for analysis of common substrates together with the enzyme used, the sensor, and the range of concentrations determinable. The most important features of some of these are described here.

### Urea

Urea is a diagnostic indicator for kidney function. Urea can be potentiometrically determined following the enzymatic reaction of the enzyme urease (reaction [I]) coupled with a variety of potentiometric transducers such as the  $p\text{CO}_2$  electrode, the  $p\text{NH}_3$  electrode, the pH electrode, and the  $p\text{NH}_4$  electrode:



The products of the enzymatic reaction given above also dissociate:



The first urea electrode was prepared in 1969 by immobilizing urease in a polyacrylamide matrix on nylon or Dacron nets. These nets were then placed onto a glass cation-selective electrode. To improve the selectivity, an ammonium-ion selective electrode with

nonactin, a neutral carrier, as the base sensor, together with immobilized urease in a polyacrylic gel was used.

### Creatinine

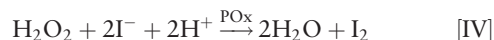
Creatinine is another diagnostic indicator of kidney function. Creatinine can be measured immobilizing purified creatininase (reaction [II]) onto nylon net and coupling it with an ammonia gas electrode:



The residual ammonia of blood serum can be removed using an enzyme reactor system, consisting of soluble NADPH and  $\alpha$ -ketoglutarate, with glutamate dehydrogenase immobilized on nylon tubes; this system is effective in removing 98% of the ammonia present in human blood and urine samples in 50 s. The entire process was carried out in a single flow stream.

### Glucose and Sugar Electrodes

An iodide-selective electrode can be used to measure glucose concentration. The measurement is based on reactions catalyzed by glucose oxidase (GOx) (reaction [III]) and peroxidase (POx) (reaction [IV]):



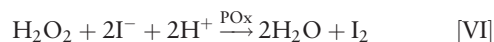
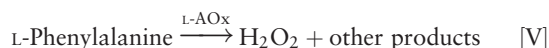
The highly sensitive iodide sensor monitors the decrease in the iodide activity at the electrode surface. The assay of glucose was performed both in a stream and at a stationary electrode. Pretreatment of the blood sample was required to remove interfering reducing agents such as ascorbic acid, tyrosine, and uric acid.

Moreover, pH electrodes can be used to detect the gluconic acid formed by the GOx reaction.

### Amino Acids

Enzyme electrodes have been widely used for the assay of amino acids in clinical analysis since several amino acids (tyrosine, phenylalanine, tryptophan, methionine) are important diagnostic health indicators.

The transducer can be an ammonium ISE ionophore (nonactin-based), a gas sensor for ammonia or carbon dioxide, and an iodide-selective sensor. The latter can be coupled with the enzymes L-aminooxidase (L-AOx) and peroxidase (e.g., coimmobilized in a polyacrylamide gel), which catalyses reactions [V] and [VI]:



Specific sensors for phenylalanine, tyrosine, glutamine, lysine, and methionine have been described in the literature.



## Penicillin

Penicillin is determined by cleaving the amide bond of the  $\beta$ -lactame ring with penicillase to produce penicilloic acid. This acid is detectable by a pH electrode covered with a thin film of penicillase and the biosensor produced was used to monitor penicillin concentration in many different fermentation broths.

## Neurotransmitters

Neurotransmitters may be determined using the respective hydrolases. Acetylcholine is hydrolyzed by acetylcholinesterase (AChE) according to reaction [VII]:



which can be monitored by immobilizing AChE on a pH electrode.

## Inhibitor-Sensitive Potentiometric Enzyme Electrodes

Some examples of potentiometric enzyme electrodes have been developed for indirect monitoring of organic pesticides or inorganic (heavy metals, fluoride, cyanide, etc.) substances that reduce the activity of the immobilized enzyme on the transducer. Parameters such as enzyme concentration, substrate concentration, pH, incubation time can affect the biosensor response, and have to be optimized. A classical example is the monitoring of organophosphorous and carbamate pesticides (whose mechanism of action is the inhibition of AChE, the enzyme implicated in the neurotransmission) using a glass electrode with immobilized AChE.

Heavy metals and fluoride ions have been reported to inhibit the activity of the enzyme urease.

**See also:** **Clinical Analysis:** Glucose. **Enzymes:** Immobilized Enzymes. **Ion-Selective Electrodes:** Overview. **Liquid Chromatography:** Amino Acids.

## Further Reading

- Cass AEG (ed.) (1990) *Biosensors. A Practical Approach*. Oxford: Oxford University Press.
- Eggins BR (2002) *Chemical Sensors and Biosensors*. London: Wiley.
- Guilbault GG (1984) *Analytical Uses of Immobilized Enzymes*. New York: Dekker.
- Guilbault GG and Mascini M (eds.) (1987) *Analytical Uses of Immobilized Biological Compounds for Detection. Medical and Industrial Uses*. Dordrecht: Riedel.
- Mascini M and Palleschi G (1989) Design and applications of enzyme electrode probes. *Selective Electrode Reviews* 11: 191.
- Mosbach K (ed.) (1988) Immobilized enzymes and cells, Part D. In: *Methods in Enzymology*, vol. 137. London: Academic Press.
- Ramsay G (ed.) (1998) *Commercial Biosensors, Application to Clinical, Bioprocess, and Environmental Samples*. New York: Wiley.
- Schmid RD and Scheller F (eds.) (1989) *Biosensors. Application in Medicine, Environmental Protection and Process Control*. Weinheim: VCH.
- Thevenot DR, Toth K, Durst RA, and Wilson GS (1999) Electrochemical biosensors: recommended definitions and classifications. *Pure and Applied Chemistry* 71(12): 2333–2348.
- Tran Minh C (1993) *Biosensor*. Paris: Chapman and Hall and Masson.
- Turner APF (ed.) (1992) *Advances in Biosensors*, vols. 1 and 2. London: JAI Press.
- Turner APF, Karube I, and Wilson G (1986) *Biosensors, Fundamentals and Applications*. Oxford: Oxford Science.
- Wise DL (ed.) (1990) *Bioinstrumentation Research, Development and Applications*. London: Butterworths.

## Clinical Applications

**U E Spichiger-Keller**, Centre for Chemical Sensors and Chemical Information Technology, Zurich, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 2341–2353, © 1995, Elsevier Ltd., with revisions made by the Editor.

## Introduction

An ion-selective electrode (ISE) is a chemical sensor that can quantitatively and reversibly measure an ion activity based on a potentiometric transduction of

the recognition process. Fiber optical sensors have been developed in response to the need for selective, reliable sensors for medical use, especially for patient-near offline or continuous *in vivo* monitoring, related to the problems inherent in electrochemical methods. Also, optical test kits for blood ions are commercially available. However, electrodes offer a broad field of applications in patient-near-monitoring as well as in large automated biochemistry analyzers.

Ion-selective electrodes introduced in chemical instruments are most frequently based on liquid

## Penicillin

Penicillin is determined by cleaving the amide bond of the  $\beta$ -lactame ring with penicillase to produce penicilloic acid. This acid is detectable by a pH electrode covered with a thin film of penicillase and the biosensor produced was used to monitor penicillin concentration in many different fermentation broths.

## Neurotransmitters

Neurotransmitters may be determined using the respective hydrolases. Acetylcholine is hydrolyzed by acetylcholinesterase (AChE) according to reaction [VII]:



which can be monitored by immobilizing AChE on a pH electrode.

## Inhibitor-Sensitive Potentiometric Enzyme Electrodes

Some examples of potentiometric enzyme electrodes have been developed for indirect monitoring of organic pesticides or inorganic (heavy metals, fluoride, cyanide, etc.) substances that reduce the activity of the immobilized enzyme on the transducer. Parameters such as enzyme concentration, substrate concentration, pH, incubation time can affect the biosensor response, and have to be optimized. A classical example is the monitoring of organophosphorous and carbamate pesticides (whose mechanism of action is the inhibition of AChE, the enzyme implicated in the neurotransmission) using a glass electrode with immobilized AChE.

Heavy metals and fluoride ions have been reported to inhibit the activity of the enzyme urease.

**See also:** **Clinical Analysis:** Glucose. **Enzymes:** Immobilized Enzymes. **Ion-Selective Electrodes:** Overview. **Liquid Chromatography:** Amino Acids.

## Further Reading

- Cass AEG (ed.) (1990) *Biosensors. A Practical Approach*. Oxford: Oxford University Press.
- Eggins BR (2002) *Chemical Sensors and Biosensors*. London: Wiley.
- Guilbault GG (1984) *Analytical Uses of Immobilized Enzymes*. New York: Dekker.
- Guilbault GG and Mascini M (eds.) (1987) *Analytical Uses of Immobilized Biological Compounds for Detection. Medical and Industrial Uses*. Dordrecht: Riedel.
- Mascini M and Palleschi G (1989) Design and applications of enzyme electrode probes. *Selective Electrode Reviews* 11: 191.
- Mosbach K (ed.) (1988) Immobilized enzymes and cells, Part D. In: *Methods in Enzymology*, vol. 137. London: Academic Press.
- Ramsay G (ed.) (1998) *Commercial Biosensors, Application to Clinical, Bioprocess, and Environmental Samples*. New York: Wiley.
- Schmid RD and Scheller F (eds.) (1989) *Biosensors. Application in Medicine, Environmental Protection and Process Control*. Weinheim: VCH.
- Thevenot DR, Toth K, Durst RA, and Wilson GS (1999) Electrochemical biosensors: recommended definitions and classifications. *Pure and Applied Chemistry* 71(12): 2333–2348.
- Tran Minh C (1993) *Biosensor*. Paris: Chapman and Hall and Masson.
- Turner APF (ed.) (1992) *Advances in Biosensors*, vols. 1 and 2. London: JAI Press.
- Turner APF, Karube I, and Wilson G (1986) *Biosensors, Fundamentals and Applications*. Oxford: Oxford Science.
- Wise DL (ed.) (1990) *Bioinstrumentation Research, Development and Applications*. London: Butterworths.

## Clinical Applications

**U E Spichiger-Keller**, Centre for Chemical Sensors and Chemical Information Technology, Zurich, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 2341–2353, © 1995, Elsevier Ltd., with revisions made by the Editor.

## Introduction

An ion-selective electrode (ISE) is a chemical sensor that can quantitatively and reversibly measure an ion activity based on a potentiometric transduction of

the recognition process. Fiber optical sensors have been developed in response to the need for selective, reliable sensors for medical use, especially for patient-near offline or continuous *in vivo* monitoring, related to the problems inherent in electrochemical methods. Also, optical test kits for blood ions are commercially available. However, electrodes offer a broad field of applications in patient-near-monitoring as well as in large automated biochemistry analyzers.

Ion-selective electrodes introduced in chemical instruments are most frequently based on liquid

membranes incorporating natural or specifically tailored synthetic carriers, which accomplish the recognition process, as well as sodium glass membranes. Liquid membranes with a polymer matrix stabilizing the membrane phase were introduced in clinical chemistry in 1972 by Simon, where the first valinomycin electrode was tested in the University Hospital, Zurich, Switzerland. Solid-state sensors were introduced in 1980 for sodium chloride measurements in sweat preferably for the diagnosis of cystic fibrosis. Slide techniques for electrolyte assays were introduced subsequently for analyzers as well as for patient-near-monitoring.

Liquid-membrane electrodes are classified in two groups:

1. Liquid-membrane electrodes based on charged sites (classical ion exchanger or charged ion-selective complexing agent (charged ion carrier, charged ionophores)). The membrane consists of an organic, water-immiscible liquid phase incorporating sites.

2. Neutral carrier liquid- or solvent-polymeric membrane electrodes, in which the membrane is usually formed from an organic solution of an electrically neutral, ion-selective complexing agent (neutral ion carriers, neutral ionophores) held in an inert polymer matrix. These membrane electrodes make use of the outstanding inherent ion selectivity of certain natural and synthetic ionophores.

Electrolyte solutions contain at least two classes of mobile ions, namely cations and anions, and the membranes are not ideally permselective. Thus, a perfect exclusion of interfering species cannot be realized in practice. If considerations are restricted to permselective membranes for the ion I with relative activity  $a_i'$  in the sample phase, charge number  $z_i$ , and interfering ions J with charge number  $z_j$ , the description of the e.m.f. function of a membrane cell is ideally consistent with the classical Nikolsky-Eisenman equation:

$$\text{e.m.f.} = E = E_i^0 + \frac{RT}{z_i F} \ln \left[ a_i' + \sum_{j \neq i} K_{ij}^{\text{Pot}} a_j^{z_i/z_j} \right] \quad [1]$$

where  $RT/z_i F$  is the slope  $s$  of the electrode function denoted by the gas constant  $R$ , the Faraday constant  $F$ , the absolute temperature  $T$ , and the charge number of the primary ion  $z_i$ .  $E_i^0$  is a constant contribution to the e.m.f. The selectivity factor  $K_{ij}^{\text{Pot}}$  denotes a factor for the rejection of each interfering ion J within the sample solution in terms of relative ion activities  $a_j$  with charge number  $z_j$ . The relative activity  $a_i$  denotes the active molality of the ion I,  $m_i$  divided by unit molality:  $m_i/1 \text{ mol kg}^{-1}$ .

## Electrode Characteristics

### Electrode Standard Potential and Liquid Junction Potential

The electrode standard potential  $E_i^0$  corresponds to the intercept of the linear response function of the electrode, which is the e.m.f. change as a function of the logarithm of the relative ion activity ( $\Delta \text{e.m.f.} = f(\log a_i)$ ) and should be constant. The electrode standard potential  $E_i^0$  represents the sum of all contributions to the e.m.f. that are not immediately dependent on the activity of the primary ion I or the interfering ion J of the sample:

$$E_i^0 = E_{\text{Ref}} + E_J + E_{\text{el}} + E_{\text{as}} \quad [2]$$

For biological samples four different contributions to the standard potential  $E_i^0$  can be observed: the contribution of the internal standard potential of the reference electrode  $E_{\text{Ref}}$ ; the diffusion potential over the liquid junction  $E_J$  generated between the sample solution and the reference electrode; a potential difference (electrical asymmetry) of the ion-selective membrane after preparation and conditioning  $E_{\text{el}}$ ; and a sample-induced asymmetry of the membrane  $E_{\text{as}}$ . For measurements in human blood samples directly the adsorption of sample components at the membrane surface creates a drift associated with the affinity of the membrane to lipids as well as proteins.

Measuring techniques recommended by the Working Group on Sensors and Electrodes as well as membrane technological operations intend to compensate or eliminate these asymmetries responsible for inaccuracies by permanent recalibration. The extent of the sample-induced membrane asymmetry depends on the composition of the membrane and should be checked in a symmetric cell configuration for different applications. In a first approximation it is assumed that the potential of the reference electrode ( $E_{\text{Ref}} + E_J$ ), including the liquid junction potential, is constant and  $E_J$  is independent of the sample composition. Changes are minimized by the use of highly concentrated equitransferent reference solutions. However, the ideal behavior is usually not attained for various reasons. The liquid junction potential does in fact change considerably, e.g., due to the different ion background of calibrating solutions as compared to real samples (cations added as chlorides). An e.m.f. shift corresponding to 2–8% ion activity units may be observed. The electrode function may be corrected by calculation of  $E_J$  based on the Henderson equation and the background ion activities. The response of the liquid junction potential of the reference electrode versus proteins and lipids is very much influenced by the construction of the reference electrode.

**Selectivity Factor  $K_{ij}^{\text{Pot}}$** 

The selectivity of an electrode has to be discussed and checked within the dynamic range of activity of the analyte and the interfering ions relevant for medical assays.

A lack of discrimination of ions is measured as inaccuracy, which amounts to the difference between zero activity of the interfering ion and the mean actual activity, and as imprecision of the calibrated setting point due to the varying background ion activity. The variation around the calibrated setting point has to be restricted to less than the total allowable analytical error with respect to a reliable medical interpretation of the results for the primary ion. The allowable analytical error is defined by the biological variation of the analyte (see **Tables 1** and **2**) assuming absence of any inaccuracy. The same rules may be applied to the calculation of the required selectivity coefficient with regard to the necessary discrimination of interfering species. With first priority the required selectivity

coefficient is calculated based on the biological variation ( $cv_a \leq 0.5 cv_b$ ) or maximum 1% in second priority:

$$K_{ij}^{\text{Pot}} \leq 0.5 cv_b \cdot (a_i/a_j^{z_i/z_j}) = cv_a \cdot (a_i/a_j^{z_i/z_j}) \quad [3]$$

where  $cv_b$  is the biological coefficient of variation and  $cv_a$  is the analytical coefficient of variation. Unfortunately,  $cv_b$  has not yet been evaluated for ion activities. The required selectivity coefficients for a large physiological concentration range are presented in **Table 3** based on the intraindividual  $cv_b$  for total ion concentrations. The selectivity coefficients reported in the literature may differ considerably since they might be determined by different methods such as the separate solution method (SSM), fixed interfering ion method, fixed primary ion method, and specific application method. The selectivity coefficients discussed here are based on results of the SSM method (see **Table 3**). Restrictions may be overcome by using sensor arrays for the determination of the

**Table 1** Biological setting points for electrolyte total ion concentrations and active molalities. Estimations of  $cv_b\%$  are based on physiological ranges given by the intra- and interindividual variation referred to the mean biological setting point

Primary ion <i>I</i>	Reference interval <sup>a,b,c</sup> adults, plasma total ion concentration (mmol l <sup>-1</sup> ), <i>c<sub>i</sub></i>	Mean setting point of <sup>a,b,c</sup> active molality, <i>m<sub>i</sub></i> (mmol kg <sup>-1</sup> )	Molal activity coefficient for <i>I</i> = 0.160 (37°C), $\gamma_i$	Interindividual variation <sup>a</sup> ( $cv_{\text{inter}}/\%$ of mean)	Intraindividual variation <sup>d</sup> ( $cv_{\text{intra}}/\%$ of mean)
Sodium, Na <sup>+</sup>	134–143	108	0.742	0.6	0.7
Potassium, K <sup>+</sup>	3.3–4.8	3.1	0.728	4.4	5.1
Calcium, Ca <sup>2+</sup>	2.03–2.70			2.2	1.7
Total ionized	1.12–1.45	0.42	0.325	4.1	?
Magnesium, Mg <sup>2+</sup>	0.66–0.95			5.9	2.2
Total ionized	0.46–0.665	0.19	0.336	?	?
Ammonium, NH <sub>4</sub> <sup>+</sup>	11–60 × 10 <sup>-3</sup>	28 × 10 <sup>-3</sup>	0.724	28	?
Hydrogen activity H <sub>3</sub> O <sup>+</sup> (pH) <sup>e</sup>	PH 7.35–7.45	10 <sup>-7.51</sup>	0.770	5.6	?
Lithium, Li <sup>+</sup> , therapeutic in 14 U ml <sup>-1</sup> heparinate	0.3–1.3	0.767			
Chloride Cl <sup>-1</sup>	Max. 0.9 99–111	0.68 83.6	0.760 0.741		
Total hydrogencarbonate				1.4	1.7
HCO <sub>3</sub> <sup>-</sup>	22–26	18.0	0.701	3.0	4.6
Total phosphate	0.56–1.28			9.4	7.8
H <sub>2</sub> PO <sub>4</sub> <sup>-</sup>	0.14–0.32	0.17	0.703		
HPO <sub>4</sub> <sup>2-</sup>	0.42–0.96	0.25	0.330		
Organic acids R-COO <sup>-</sup> , as acetate <sup>c</sup>	12 (10–17)	9.8	0.763	?	?

<sup>a</sup>Paschen K, Lammers M, and Müller-Plathe O (1987) Wasser- und Elektrolytstoffwechsel. In: Greiling H und Gressner AM (eds.) *Lehrbuch der Klinischen Chemie und Pathobiochemie*, pp. 368–390. Stuttgart: Schattauer.

<sup>b</sup>Spichiger UE (1989) *A Selfconsistent Set of Reference Values*. Thesis no. 8830, Swiss Federal Institute of Technology (ETH), Zürich, Switzerland.

<sup>c</sup>Preuss HG, Podlasek SJ, and Henry JB (1991) Evaluation of renal function and water, electrolyte and acid–base balance. In: Henry JB (ed.) *Clinical Diagnosis and Management*, pp. 119–139. Philadelphia: W.B. Saunders.

<sup>d</sup>Fraser CG (ed.) (1986) *Interpretation of Clinical Laboratory Data*. Oxford: Blackwell Scientific Publications.

<sup>e</sup>Total buffer capacity of blood: 46–52 mmol l<sup>-1</sup>; a pH value of 7.4 is equivalent to 39.8 nmol l<sup>-1</sup> hydrogen ions.

**Table 2** Allowable analytical errors in terms of concentrations,  $c$ , as required by the CAP (College of American Pathologists) and DGKC (Deutsche Gesellschaft für Klinische Chemie) and 1% maximum allowable analytical error

Analyte ion	1% total allowable analytical error, $\Delta c$ ( $\text{mmol l}^{-1}$ ); $cv_a$ (1%)	Total allowable analytical error CAP <sup>a</sup> $\Delta c$ ( $\text{mmol l}^{-1}$ ); $cv_a$ (%)	Total allowable analytical error DGKC <sup>b</sup> $\Delta c$ ( $\text{mmol l}^{-1}$ ); $cv_a$ (%)	Total allowable error CAP e.m.f. (mV) 37°C; ideal slope assumed (< 1%)
Sodium	1.40; 1.0	$\pm 0.42$ ; 0.30	$\pm 0.75$ ; 0.54	0.080
Potassium	$\pm 0.041$ ; 1.0	$\pm 0.090$ ; 2.2	$\pm 0.125$ ; 3.1	0.58
Ammonium	$\pm 0.36 \times 10^{-3}$ ; 10	$\pm 0.004$ ; 11.5	$\pm 0.045$ ; 1.96	(< 0.26)
Hydrogen activity		2.8		
PH			$\pm 0.008$ ; 1.9	(< 0.26)
Lithium				
Activity due to 14 U ml <sup>-1</sup>				
Li-heparinate	< 0.009; 1.0			
Therapeutic level	0.013; 1.0		$\pm 0.083$ ; 10.4	(< 0.26)
Calcium				
Total concentration	0.025; 1.0	0.0191; 0.85	$\pm 0.0558$ ; 2.5	0.11
Free ion concentration	0.0122; 1.0		$\pm 0.0275$ ; 2.3	(< 0.135)
Magnesium				
Total concentration	$\pm 0.008$ ; 1.0	$\pm 0.0088$ ; 1.1	$\pm 0.024$ ; 3.05	0.146
Free ion concentration	$\pm 0.0056$ ; 1.0			(< 0.135)
Chloride	1.05; 1.0	$\pm 0.74$ ; 0.7	$\pm 1.0$ ; 0.95	0.19
Hydrogencarbonate as total CO <sub>2</sub>	$\pm 0.24$ ; 1.0	$\pm 0.37$ ; 1.5	$\pm 0.33$ ; 1.4	0.39
Total inorganic phosphate	$\pm 0.0094$ ; 1.0	$\pm 0.037$ ; 3.9	$\pm 0.060$ ; 6.5	(< 0.16)
H <sub>2</sub> PO <sub>3</sub> <sup>-</sup> /HPO <sub>3</sub> <sup>2-</sup> = 1/4				
Organic anions, anion gap	$\pm 0.135$ ; 1.0			

<sup>a</sup> See Further reading section.<sup>b</sup> DGKC: Deutsche Gesellschaft für Klinische Chemie: the allowable analytical error is limited to 1/12 of the reference range, calculated in % of the mean level.

Last column, total allowable analytical error in millivolts related to activity measurements. Ideal slope (61.54 mV for singly charged and 30.77 mV for doubly charged ion (37°C) is assumed.

activity of the primary and the interfering ions combined with a chemometric correction. However, each sensor needs a high quality of performance. The dominating presence of elements with high abundance prevents the use of ISEs for assays of trace element activities in biological specimens or diluted samples.

### Long-Term Stability of Electrode and Lipophilicity of the Electrode Components

The lipophilicity,  $\log P_{\text{TLC}}$ , of membrane components as determined by thin-layer chromatography is of the utmost relevance to the lifetime of liquid-membrane sensors. Since blood serum and whole blood are much more lipophilic than a purely aqueous solution they are much more efficient extractants for membrane components. The lifetime decreases by a factor of  $10^{-2.5}$  for direct measurements. Hence, much higher lipophilicities of host molecules and additives are required when membranes have to be contacted with these media. The required lipophilicity,  $\log P_{\text{TLC}}$ , of neutral carriers incorporated in ion-selective liquid membranes with a thickness of

$\sim 200 \mu\text{m}$  and a minimum lifetime of 720 h of contact to whole blood or plasma is  $> 11.3$ . For plasticizers the required  $\log P_{\text{TLC}}$  is  $> 14.3$ . Covalent binding of membrane components increasing the lifetime of sensors can be used only in a few cases.

### Calibration, Standardization, and Comparisons with Definitive or Reference Procedures

Ion-selective electrodes used in whole blood or undiluted plasma or serum respond to the electrochemical activity of the ion. At the high ionic strengths of  $I \approx 0.16 \text{ mol l}^{-1}$  encountered in these biological samples, the molality  $m_i$  can no longer be substituted for active molality  $\tilde{m}_i$ . Free ion molality  $m_i$  may be converted to the active molality  $\tilde{m}_i$  by multiplying with the molal activity coefficient  $\gamma_i$ :

$$\tilde{m}_i = m_i \cdot \gamma_i \quad \text{with } \gamma_i \neq 1 \quad [4]$$

By definition the active molality is related to the relative molal activity  $a_i$  by dividing  $\tilde{m}_i$  by the unit



**Table 3** Required logarithmic selectivity coefficients for the potentiometric assay of a primary analyte ion I at the lower edge of the reference interval and discrimination of an interfering ion J at the upper level of the reference interval calculated based on a minimum allowable error according to Table 2

Primary ion I  Mean physiological setting point of active molality for background calibration (minimum active molality) Mean active molality for calibration Allowable error in %	Log required selectivity coefficient				
	Undiluted specimen <sup>a</sup>			Diluted sample <sup>b</sup> (1 + 19)	
	Secondary interfering ions J	Without calibration <sup>c</sup>	With calibration <sup>d</sup>	Without calibration <sup>c</sup>	With calibration <sup>d</sup>
<b>Sodium</b> Na <sup>+</sup> (0.1005) 0.108 ± 0.3%	Na <sup>+</sup>	—	—	—	—
	K <sup>+</sup>	− 1.1	− 0.33	− 1.1	− 0.33
	Ca <sup>2+</sup>	− 1.9	− 0.78	− 2.65	− 1.6
	Mg <sup>2+</sup>	− 1.7	− 0.45	− 2.4	− 1.2
	NH <sub>4</sub> <sup>+</sup>	+ 0.84	+ 1.2	+ 0.84	+ 1.2
	H <sub>3</sub> O <sup>+</sup> (pH)	+ 3.8	+ 4.8	—	—
	Li (therap.)	− 0.52	+ 0.18	− 0.52	+ 0.18
<b>Potassium</b> K <sup>+</sup> (2.40 × 10 <sup>−3</sup> ) 3.1 × 10 <sup>−3</sup> ± 1%	Na <sup>+</sup>	− 3.7	− 2.2	− 3.7	− 2.2
	K <sup>+</sup>	—	—	—	—
	Ca <sup>2+</sup>	− 3.0	− 1.9	− 3.8	− 2.7
	Mg <sup>2+</sup>	− 2.8	− 1.6	− 3.5	− 2.3
	NH <sub>4</sub> <sup>+</sup>	− 0.25	+ 0.2	− 0.25	+ 0.2
	H <sub>3</sub> O <sup>+</sup> (pH)	+ 2.7	+ 3.7	—	—
	Li <sup>+</sup>	− 16	− 0.91	− 1.6	− 0.91
<b>Calcium</b> Ca <sup>2+</sup> (0.690 × 10 <sup>−3</sup> ; ion. 0.381 × 10 <sup>−3</sup> ) 0.765 × 10 <sup>−3</sup> (total) ± 0.85% 0.42 × 10 <sup>−3</sup> (ionic) ± 0.85%	Na <sup>+</sup>	− 3.6	− 2.4	− 2.0	− 0.77
	K <sup>+</sup>	− 0.7	− 0.16	+ 0.97	+ 1.5
	Ca <sup>2+</sup>	—	—	—	—
	Mg <sup>2+</sup>	− 1.9	− 0.9	− 1.9	− 0.9
	NH <sub>4</sub> <sup>+</sup>	+ 3.2	+ 3.4	+ 4.8	+ 5.0
	H <sub>3</sub> O <sup>+</sup> (pH)	+ 9.2	+ 9.9	—	—
	Li <sup>+</sup>	+ 0.3	+ 0.9	+ 2.1	+ 2.5
<b>Magnesium</b> Mg <sup>2+</sup> (0.231 × 10 <sup>−3</sup> ; ion. 0.162 × 10 <sup>−3</sup> ) 0.2975 × 10 <sup>−3</sup> (tot.) ± 1% 0.208 × 10 <sup>−3</sup> (ion.) ± 1%	Na <sup>+</sup>	− 3.9	− 2.7	− 2.4	− 2.1
	K <sup>+</sup>	− 1.0	− 0.5	+ 0.5	+ 1.1
	Ca <sup>2+</sup>	− 2.5	− 1.7	− 2.6	− 2.7
	Mg <sup>2+</sup>	—	—	—	—
	NH <sub>4</sub> <sup>+</sup>	+ 2.9	+ 3.1	+ 4.4	—
	H <sub>3</sub> O <sup>+</sup> (pH)	+ 8.9	+ 9.6	—	—
	Li <sup>+</sup>	+ 0.2	+ 0.64	+ 1.7	—
<b>Ammonium</b> NH <sub>4</sub> <sup>+</sup> (0.80 × 10 <sup>−5</sup> ) 2.8 × 10 <sup>−5</sup> ± 1%	Na <sup>+</sup>	− 6.1	− 4.6	− 6.1	− 4.6
	K <sup>+</sup>	− 4.7	− 3.8	− 4.6	− 3.8
	Ca <sup>2+</sup>	− 5.4	− 4.4	− 6.2	− 5.2
	Mg <sup>2+</sup>	− 5.3	− 4.0	− 6.0	− 4.7
	NH <sub>4</sub> <sup>+</sup>	—	—	—	—
	H <sub>3</sub> O <sup>+</sup> (pH)	+ 0.25	+ 1.2	—	—
	Li <sup>+</sup>	− 4.1	− 3.4	—	− 3.4
<b>Hydronium ions</b> H <sub>3</sub> O <sup>+</sup> (2.5 × 10 <sup>−5</sup> ; ion pH 3.4 × 10 <sup>−8</sup> ) 4.9 × 10 <sup>−5</sup> 3.98 × 10 <sup>−8</sup> (pH 7.4) ± 0.13% (ΔpH ± 0.0006) active molality	Na <sup>+</sup>	− 9.4	− 7.9	—	—
	K <sup>+</sup>	− 7.9	− 7.1	—	—
	Ca <sup>2+</sup>	− 8.7	− 7.6	—	—
	Mg <sup>2+</sup>	− 8.5	− 7.3	—	—
	NH <sub>4</sub> <sup>+</sup>	− 6.0	− 5.6	—	—
	H <sub>3</sub> O <sup>+</sup> (pH)	—	—	—	—
	Li <sup>+</sup>	− 7.3	− 6.6	—	—
<b>Lithium</b> Li <sup>+</sup> (0.230 × 10 <sup>−3</sup> ) Mean therapeutic level: 0.767 × 10 <sup>−3</sup> ± 1%	Na <sup>+</sup>	− 4.7	− 3.2	− 4.7	− 3.1
	K <sup>+</sup>	− 3.2	− 2.5	− 3.2	− 2.5
	Ca <sup>2+</sup>	− 4.0	− 2.9	− 4.8	− 3.7
	Mg <sup>2+</sup>	− 3.8	− 2.6	− 4.6	− 3.3
	NH <sub>4</sub> <sup>+</sup>	− 1.3	− 0.89	− 1.3	− 0.88

Table 3 Continued

Primary ion I  Mean physiological setting point of active molality for background calibration (minimum active molality) Mean active molality for calibration Allowable error in %	Log required selectivity coefficient				
	Undiluted specimen <sup>a</sup>			Diluted sample <sup>b</sup> (1 + 19)	
	Secondary interfering ions J	Without calibration <sup>c</sup>	With calibration <sup>d</sup>	Without calibration <sup>c</sup>	With calibration <sup>d</sup>
	H <sub>3</sub> O <sup>+</sup> (pH)	+1.7	+2.7	—	—
	Li <sup>+</sup>	—	—	—	—
Chloride	HCO <sub>3</sub> <sup>−</sup>	−1.6	−0.3	−1.6	−0.3
Cl <sup>−</sup> (73.4 × 10 <sup>−3</sup> )	RCOO <sup>−</sup>	−1.4	−0.7	−1.4	−0.7
83.6 × 10 <sup>−3</sup> ± 0.7%	<sup>e</sup> SCN <sup>−</sup>	+0.8		+0.8	
	H <sub>2</sub> PO <sub>4</sub> <sup>−</sup>	+0.4	+1.05	+0.4	+1.05
	HPO <sub>4</sub> <sup>2−</sup>	−1.54	−0.57	−2.04	−1.2
Hydrogencarbonate	Cl <sup>−</sup>	−2.8	−1.5	−2.8	−1.5
HCO <sub>3</sub> <sup>−</sup> (15.7 × 10 <sup>−3</sup> )	RCOO <sup>−</sup>	−1.9	−1.4	−1.9	−1.4
18 × 10 <sup>−3</sup> ± 1.0%	<sup>e</sup> SCN <sup>−</sup>	+0.13		+0.13	
	H <sub>2</sub> PO <sub>4</sub> <sup>−</sup>	−0.17	+0.37	−0.17	+0.37
	HPO <sub>4</sub> <sup>−</sup>	−2.1	−1.2	−2.7	−1.9
Phosphate	Cl <sup>−</sup>	−3.7	−2.7	−2.4	−1.4
HPO <sub>4</sub> <sup>2−</sup> (0.14 × 10 <sup>−3</sup> )	HCO <sub>3</sub> <sup>−</sup>	−2.4	−1.5	−1.1	−0.24
0.25 × 10 <sup>−3</sup> ± 1.0%	RCOO <sup>−</sup>	−2.1	−1.8	−0.78	−0.5
	<sup>e</sup> SCN <sup>−</sup>	+2.0		+3.3	
	H <sub>2</sub> PO <sub>4</sub> <sup>−</sup>	+1.4	+1.7	+2.7	+3.0
Phosphate	Cl <sup>−</sup>	−5.0	−3.6	−5.0	−3.6
H <sub>3</sub> PO <sub>4</sub> <sup>−</sup> (0.098 × 10 <sup>−3</sup> )	HCO <sub>3</sub> <sup>−</sup>	−4.3	−3.2	−4.3	−3.2
0.17 × 10 <sup>−3</sup> ± 1.0%	RCOO <sup>−</sup>	−4.1	−3.6	−4.1	−3.6
	<sup>e</sup> SCN <sup>−</sup>	−2.1		−2.1	
	H <sub>2</sub> PO <sub>4</sub> <sup>−</sup>	−4.3	−3.4	−4.9	−4.1

<sup>a</sup> Calculations are based on active molality of ions as far as available (sodium, potassium, and chloride, especially).

<sup>b</sup> Calculations are based on active molalities derived from total substance concentrations, assuming an ionic strength of 0.16 mol l<sup>−1</sup>.

<sup>c</sup> Without: activity of the interfering ion = upper limit of reference range; activity of the primary ion = lower limit of reference range.

<sup>d</sup> With: activity difference of the interfering ion between mean value and the limit (upper or lower) of the reference range which results in the higher difference; activity of the primary ion = lower limit of reference range.

<sup>e</sup> SCN<sup>−</sup> as interfering ion is supposed to be at the maximum level for a smoker which is 0.15 mmol l<sup>−1</sup>.

Measurements in the specimen directly are differentiated from measurements in a 1 + 19 diluted sample. Data are calculated for an assay without calibration of the activities of the ion background as well as with calibration of a mean physiological background (2% complexation for Na<sup>+</sup>, but no complexation of further singly charged ions is assumed).

molality (1 mol kg<sup>−1</sup>):

$$a_i = \tilde{m}_i / (\text{mol kg}^{-1}) \quad [5]$$

Concomitantly the molality of free ions  $m_i$  differs from the total molality of the ion I,  $m_{i,\text{tot}}$ , by the association of the ions 'i' to organic and inorganic anions and cations ( $L_j$ ), respectively, and their molal concentration  $m_{Lj}$ . The degree of complexation is related to the chemical specification of an analyte ion and is higher for multiply charged ions and transition metals. It is denoted by the association constant  $K_{\text{ass}}$ :

$$m_{i,\text{tot}} = m_i + m_i \cdot \sum_{j=1}^n (K_{\text{ass},j} \cdot m_{Lj}) \quad [6]$$

Analytical systems that measure substance concentrations in moles per liter of plasma, e.g., atomic absorption spectrometry (AAS) and flame atomic emission spectrometry (FAES), coexist with determinations of relative ion activity in undiluted and diluted samples. In consequence a system of computing and reporting results of sodium and potassium ions was proposed by the International Federation of Clinical Chemistry (IFCC) to convert the results of direct ISE assays to FAES/AAS and vice versa.

The system reports on an adjustment factor by which the measured relative molal activity and the molality of the analyte ion I is related to the total substance concentration  $c_{i,\text{tot}}$ . The system is based on

the assumption of a standard plasma specimen defined as having a mass concentration of water in plasma of  $0.93 \text{ kg l}^{-1}$ , a hydrogencarbonate concentration of  $24 \text{ mmol l}^{-1}$ , and pH 7.4. For sodium ions the complexed fraction ( $\sum_{j=1}^n K_{\text{ass}j} \cdot m_{\text{L}j}$ ) is assumed to be 2%.

Based on these assumptions the total substance concentration of the ion I may be converted to the total molality of the analyte ion by dividing by the mass concentration of plasma water:

$$m_{i,\text{tot}} = c_{i,\text{tot}} / \rho_{\text{H}_2\text{O};\text{P}} \quad (\text{mol kg}^{-1}) \quad [7]$$

Hence, by the following completed equation the total substance concentration of I in moles per liter of plasma is related to the relative activity measured by ISEs:

$$c_{i,\text{tot}} = a_i \cdot \rho_{\text{H}_2\text{O};\text{P}} \cdot \left( 1 + \sum_{j=1}^n K_{\text{ass}j} \cdot m_{\text{L}j} \right) \cdot (1 \text{ mol kg}^{-1}) / \gamma_i \quad [8]$$

Although the system is based on molal units by definition (International Union of Pure and Applied Chemistry and the IFCC), the results are strongly related to the temperature since the slope of the ISE response function increases with increasing temperature. Routinely all measurements are made at  $37^\circ\text{C}$ .

To estimate the mass concentration of water in plasma a volume displacement effect by proteins and lipids has to be taken into account and is integrated in the mass concentration factor for standard plasma specimens. The correction factor for the mean specific volume of proteins is assumed to be 0.73 and 1.03 for lipids, respectively:

$$\rho_{\text{H}_2\text{O};\text{P}} = 0.991 - 0.73c_{\text{protein}} - 1.03c_{\text{lipid}} \quad (\text{kg l}^{-1}) \quad [9]$$

A theoretically predicted positive bias of 6.7% for direct potentiometry (total ion molality in plasma water) versus indirect assay (total ion concentration in the whole sample volume) is generally accepted for normal situations and is compensated by calibration.

Ion-selective electrodes offer an observable value for the biologically relevant parameter: the ion's active molality in the aqueous phase of the biological specimen directly. This might be a challenge for elements that are associated by 30–50% to different anions, e.g., magnesium and calcium. For alkaline and alkaline-earth metal cations the Debye–Hückel formalism allows a fair approximation of the mean activity coefficient  $\gamma_{i\pm}$  as measured for an ion and its counterion in a calibration solution. The notation of

the basic theory has varied during history and was adapted for higher accuracy of biological measurements and for estimations of single ion activities. The aim of these efforts was an optimal approximation of the true molal activity coefficients in plasma or whole blood for the calculation of ion concentrations based on the ion's active molality measurements. Factors for standard plasma estimated by the Debye–Hückel formalism at  $37^\circ\text{C}$  are given in Table 1. However, particularly in the most relevant cases when hyperhydration, dehydration, and strong changes in osmolality, ionic strength, and in the ion background activities occur, ion concentration levels do not correlate with measured active ion molality and computations based on fixed factors are useless. Hence, the outstanding information content of the molal ion activity parameter for biological measurements is lost through correlation with total ion concentrations.

No standard reference material and procedure is actually available for molal ion activity assays. Calibration of the electrode is required. The calibration solution should match the mean sample composition as closely as possible. Addition of albumin to the calibration standard does not guarantee high accuracy of the results due to a shift of the assay standard potential that is different for plasma, serum, and albumin. Reports on comparisons of clinical analyzers show that interinstrumental deviations may be considerable. Recommendations for calibration and standardization by the IFCC Committee on Electrolytes and Blood Gases are in preparation.

Electrolyte total concentrations are generally evaluated in samples diluted by an ionic strength and pH buffer in the so-called indirect method. Due to the known standardized background, calibration and calculation of the active ion molality based on the Debye–Hückel formalism are allowable. Differences between indirect ISE assays and FAES or AAS are obvious and amount to  $\sim 3\text{--}4\%$  higher results for ISEs. They are most probably related to the volumetric dilution and the volume displacement effect by macromolecules relevant even in diluted samples.

Whenever the comparability between the two methods is impaired by the effects reported, the composition of whole blood is still relatively stable and predictable. The discussions of what should be measured and what should be reported are due to the extremely high quality requirements for most analytes. The evaluation of standard specimens for ion active molality determinations in a symmetric cell arrangement may be considered for future developments.

## Interpretation and Clinical Decision-Making

The comprehensive and critical examination of analytical and clinical decision-making is discussed and reported in numerous papers and volumes. Fundamentally, the evaluation of the efficiency of analytical tests has to be distinguished from the evaluation of procedures for medical decision-making. However, data evaluation in clinical chemistry is based on a stepwise process that starts with the value measured by the analytical instrument. The instrumental output is most frequently represented by a metric signal on a continuous scale. By calibration and standardization this value is transformed to a reliable analytical result. In contrast, clinicians intuitively regard an assay in terms of its ability to discriminate differences between two values. The medical validation reduces the full information to a binary decision that is coded 'healthy' or 'ill' as well as 'continue therapy' or 'change therapy' in most cases. The cutoff value for decisions is denoted as a reference value specifically evaluated for the population involved. Assuming that analytical problems such as drifts, inaccuracies induced by the biological matrix and interfering ions have been minimized, the analytical performance is monitored in terms of the long-term reproducibility denoted by the coefficient of variation  $cv_a$ . However, the discriminatory power of an assay in any biological setting point involves the analytical and the biological variation ( $cv_b$ ) assuming near Gaussian distributions for both. The difference between two single values is statistically significant (2.5% level for the one-tailed normal deviate with  $Z = 1.96$ ) provided that the difference between ion concentrations  $\Delta c_{i, \text{tot}}$  or active ion molalities,  $\Delta \bar{m}_i$  exceeds  $2.8 \times$  the total variation  $((cv_b^2 + cv_a^2)^{1/2})$  within the considered range of values. Since the intraindividual variation ( $cv_{b, \text{intra}}$ ) is smaller than the interindividual variation ( $cv_{b, \text{inter}}$ ), with few exceptions, individual longitudinal studies are considered more reliable for interpretation (see Table 1).

## Applications

### Microelectrodes

The miniaturization of ISEs allows the possibility of local measurements of ion activities in tissues, on surfaces, and through catheters. However, the diameter of the tip of microelectrodes is in many cases too large for *in vivo* measurements. A rule of thumb is that this diameter should be  $< 1/20$  of the larger extension of a cell. In the range  $< 1 \mu\text{m}$  a very low ion-exchange current of the primary ion as well as leaking currents of glass electrodes limit the selectivity and the response of microelectrodes. Potassium, calcium, and magnesium pH-selective microelectrodes were successfully applied for measurements in the microenvironment.

### Nonsymmetrical Ion Sensors and Solid-State Electrodes

The general trend to miniaturize chemical sensors and to make them cheap and easier for handling has led to the development of potentiometric sensors with solid internal contact. This class of sensors includes different types of solid-state electrodes, solid-state contact, and coated wire electrodes as well as ion-selective field-effect transistors (FETs).

Solid-state sensors for chloride, iodide, and fluoride are based on the solubility product of silver chloride or silver iodide particles in silicone rubber and a doped lanthanum fluoride single crystal, respectively. The fluoride-selective electrode was applied for the analysis of urine and bone tissue of people exposed to industrial sources as well as for control of therapeutic fluoride application for osteoporosis, whereas the chloride-selective sensor was applied to the analysis of sweat for the diagnosis of cystic fibrosis. In solid-state contact electrodes the solvent polymeric membrane is directly contacted to the solid field transducing element, although the reference electrode is separated from the ion-selective sensing pad.

In ion-selective field-effect transistors (ISFETs) a solvent polymeric membrane is directly contacted with the FET and the reference electrode is integrated in the sensor. A sandwich structure built on semiconducting layers operates as an electrochemical potentiometric cell. With this approach the high impedance signal from the highly resistive miniaturized electrode can be amplified *in situ* and ideally only one capacitive interface in the measuring circuit has to be respected. Concomitantly the key issue is this single interface between the ion-selective membrane and the contacted gate insulator. A prerequisite is a conductivity of the membrane due to the permselectivity and the transition of the analyte species only, related to a minimum thickness of the membrane in the range of  $> 50 \mu\text{m}$  at least. The gas permeability of the membrane ( $\text{CO}_2$ ) is often responsible for unpredicted deviations. Light sensitivity and drift problems in ISFET structures are common. Several attempts have been made to modify the ISFET response by surface modifications and by immobilizing bioselective compounds such as enzymes and antibodies. Some of the most critical drawbacks have been overcome and applications of semiconducting layers for on-line blood electrolyte monitoring by ChemFET arrays have been demonstrated.

### Automation

Crucial factors affecting the use of analytical tools in the clinic are speed and reproducibility. A clear advantage of automated systems is the feasibility of the compensation of inaccuracies due to the properties

of the biological sample by the continuously updated calibration procedure, the chemometric compensation of the effects of interfering ions within the dynamic range, and the adjustment of the liquid junction potential by calibration and computing procedures. A number of instruments for measurements in whole blood, serum, or plasma directly with ISE arrays are available. The core of the analytical flow-through system is a thermostatted and in most cases completely transparent measuring chamber ( $T = 310\text{ K} \pm 0.1$ ). The required sample volume normally ranges from 50 to 300  $\mu\text{L}$ .

In most large automated biochemistry analyzers electrode concentrations are evaluated in the diluted sample by ISEs, by the so-called indirect method. A special compartment for the determination of sodium, potassium, and chloride is integrated. The ISEs have replaced flame photometry as well as coulometry to a large extent. However, the compatibility as well as interpretability of the results is problematic in many cases (see above). Since the ion-selective assays strictly respond to molal single-ion activities in the aqueous phase, the comparability to direct measurements is weak. A volume displacement effect by lipids and proteins even affects the accuracy and comparability in diluted samples with buffered ionic strength.

See also: **Blood and Plasma. Quality Assurance: Instrument Calibration.**

## Further Reading

- Ammann D (1986) *Ion-Selective Microelectrodes*. Berlin: Springer.
- Byrne TP (1988) Ion-selective electrodes in direct potentiometric clinical analysers. *Selective Electrode Review* 10: 107–124.
- College of American Pathologists (CAP) (1977) In Elevitch FR (ed.) *Proceeding of the 1976 Aspen Conference on Analytical Goals in Clinical Chemistry*, Aug. 25–27, 1976, College of American Pathologists.
- Dinten O, Spichiger UE, Chaniotakis N, *et al.* (1991) Lifetime of neutral-carrier-based liquid membranes in aqueous samples and blood and the lipophilicity of membrane components. *Analytical Chemistry* 63: 596–603.
- Fluka Chemie AG (1991) *Selectophore<sup>®</sup> Ionophores for Ion-Selective Electrodes and Optodes*. Buchs, Switzerland: Fluka Chemie.
- Janata J (1989) *Principles of Chemical Sensors*. New York: Plenum Press.
- Katay R, Nicholson PE, Parker D, and Covington AK (1991) Comparative performance of 14-crown-4 derivatives as lithium-selective electrodes. *Analyst* 116: 135–140.
- Kobiro K, Tobe Y, Watanabe K, Yamada H, and Suzuki K (1993) Highly selective lithium ion electrode based on decalino-14-crown-4. *Analytical Letters* 26: 49–54.
- Koryta J (1980) *Medical and Biomedical Applications of Electrochemical Devices*. Chichester: Wiley.
- Koryta J (1990) Theory and applications of ion-selective electrodes, part 8. *Analytica Chimica Acta* 233: 1–30.
- Oesch U, Ammann D, and Simon W (1986) Ion-selective membrane electrodes for clinical use. *Clinical Chemistry* 32: 1448–1459.
- Rumpf G, Spichiger UE, Bühler H, and Simon W (1992) Calibration-free measurement of sodium and potassium in undiluted human serum with an electrically symmetric measuring system. *Analytical Sciences* 8: 533–559.
- Simon W and Spichiger UE (1993) Ion-selective membrane phases for microelectrodes. In: Kettenmann H and Grantyn R (eds.) *Practical Electrophysiological Methods: A Guide for in vitro Studies in Vertebrate Neurobiology*. New York: Wiley.
- Spichiger UE (1994) *Chemical Sensors and Biosensors for Medical and Biological Applications: A Developing Field in Analytical Chemistry*. Habilitation thesis. Zurich: Swiss Federal Institute of Technology.
- Suzuki K, Hayashi K, Tohda K, *et al.* (1991) Sodium ion-selective electrodes based on lipophilic 16-crown-5-derivatives. *Analytical Letters* 24: 1085–1091.
- Tanaka M, Kobayashi T, Yamashoji Y, *et al.* (1991) Polymer membrane sodium-selective electrodes based on calix[4]aryl decyl ester. *Analytical Sciences* 7: 817–818.
- Wang J (1988) *Electroanalytical Techniques in Clinical Chemistry and Laboratory Medicine*. Weinheim: VCH Publishers.

## Food Applications

**M Maj-Zurawska and A Hulanicki**, University of Warsaw, Warsaw, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Ion-selective electrodes (ISEs) offer several possibilities in food analysis, mainly for the determination of

microcomponents such as nutrients, contaminants of various origin, and chemicals added during food processing. To obtain meaningful analytical results special protocols for sample preparation and measurement must be carefully followed. The determination of sodium, potassium, calcium, ammonium, fluoride, chloride, and nitrate, as well as the measurement of pH, are the most common applications of ISEs in foodstuffs analysis.



of the biological sample by the continuously updated calibration procedure, the chemometric compensation of the effects of interfering ions within the dynamic range, and the adjustment of the liquid junction potential by calibration and computing procedures. A number of instruments for measurements in whole blood, serum, or plasma directly with ISE arrays are available. The core of the analytical flow-through system is a thermostatted and in most cases completely transparent measuring chamber ( $T = 310\text{ K} \pm 0.1$ ). The required sample volume normally ranges from 50 to 300  $\mu\text{L}$ .

In most large automated biochemistry analyzers electrode concentrations are evaluated in the diluted sample by ISEs, by the so-called indirect method. A special compartment for the determination of sodium, potassium, and chloride is integrated. The ISEs have replaced flame photometry as well as coulometry to a large extent. However, the compatibility as well as interpretability of the results is problematic in many cases (see above). Since the ion-selective assays strictly respond to molal single-ion activities in the aqueous phase, the comparability to direct measurements is weak. A volume displacement effect by lipids and proteins even affects the accuracy and comparability in diluted samples with buffered ionic strength.

See also: **Blood and Plasma. Quality Assurance: Instrument Calibration.**

## Further Reading

- Ammann D (1986) *Ion-Selective Microelectrodes*. Berlin: Springer.
- Byrne TP (1988) Ion-selective electrodes in direct potentiometric clinical analysers. *Selective Electrode Review* 10: 107–124.
- College of American Pathologists (CAP) (1977) In Elevitch FR (ed.) *Proceeding of the 1976 Aspen Conference on Analytical Goals in Clinical Chemistry*, Aug. 25–27, 1976, College of American Pathologists.

- Dinten O, Spichiger UE, Chaniotakis N, *et al.* (1991) Lifetime of neutral-carrier-based liquid membranes in aqueous samples and blood and the lipophilicity of membrane components. *Analytical Chemistry* 63: 596–603.
- Fluka Chemie AG (1991) *Selectophore<sup>®</sup> Ionophores for Ion-Selective Electrodes and Optodes*. Buchs, Switzerland: Fluka Chemie.
- Janata J (1989) *Principles of Chemical Sensors*. New York: Plenum Press.
- Katay R, Nicholson PE, Parker D, and Covington AK (1991) Comparative performance of 14-crown-4 derivatives as lithium-selective electrodes. *Analyst* 116: 135–140.
- Kobiro K, Tobe Y, Watanabe K, Yamada H, and Suzuki K (1993) Highly selective lithium ion electrode based on decalino-14-crown-4. *Analytical Letters* 26: 49–54.
- Koryta J (1980) *Medical and Biomedical Applications of Electrochemical Devices*. Chichester: Wiley.
- Koryta J (1990) Theory and applications of ion-selective electrodes, part 8. *Analytica Chimica Acta* 233: 1–30.
- Oesch U, Ammann D, and Simon W (1986) Ion-selective membrane electrodes for clinical use. *Clinical Chemistry* 32: 1448–1459.
- Rumpf G, Spichiger UE, Bühler H, and Simon W (1992) Calibration-free measurement of sodium and potassium in undiluted human serum with an electrically symmetric measuring system. *Analytical Sciences* 8: 533–559.
- Simon W and Spichiger UE (1993) Ion-selective membrane phases for microelectrodes. In: Kettenmann H and Grantyn R (eds.) *Practical Electrophysiological Methods: A Guide for in vitro Studies in Vertebrate Neurobiology*. New York: Wiley.
- Spichiger UE (1994) *Chemical Sensors and Biosensors for Medical and Biological Applications: A Developing Field in Analytical Chemistry*. Habilitation thesis. Zurich: Swiss Federal Institute of Technology.
- Suzuki K, Hayashi K, Tohda K, *et al.* (1991) Sodium ion-selective electrodes based on lipophilic 16-crown-5-derivatives. *Analytical Letters* 24: 1085–1091.
- Tanaka M, Kobayashi T, Yamashoji Y, *et al.* (1991) Polymer membrane sodium-selective electrodes based on calix[4]aryl decyl ester. *Analytical Sciences* 7: 817–818.
- Wang J (1988) *Electroanalytical Techniques in Clinical Chemistry and Laboratory Medicine*. Weinheim: VCH Publishers.

## Food Applications

**M Maj-Zurawska and A Hulanicki**, University of Warsaw, Warsaw, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Ion-selective electrodes (ISEs) offer several possibilities in food analysis, mainly for the determination of

microcomponents such as nutrients, contaminants of various origin, and chemicals added during food processing. To obtain meaningful analytical results special protocols for sample preparation and measurement must be carefully followed. The determination of sodium, potassium, calcium, ammonium, fluoride, chloride, and nitrate, as well as the measurement of pH, are the most common applications of ISEs in foodstuffs analysis.

## Characteristics of Potentiometric Sensors in Food Analysis

The analysis of foodstuffs using ISEs must take into account the special nature of the sample, the characteristics of ISEs, and the conditions of measurements. In contrast to most analytical techniques, in potentiometric measurements the analytical signal is directly correlated to the activity of the sensed species. This could be of advantage in food analysis as the bioavailability of a species is closely related to the activity of the ion. However, measuring the activity of analytes in real samples requires the fulfillment of special conditions that are rarely attained in practical analysis. The activity of the analyte ion depends on the characteristics of the medium, e.g., its electric permittivity, temperature, and ionic composition, the latter being expressed in dilute solutions by the value of the ionic strength. The presence of nonaqueous solvents also influences the activity/concentration relationship. Therefore for samples that are dilute aqueous solutions alone a potentiometric measurement may give a relatively reasonable evaluation of the activity. In the case of real food samples, we rarely have such ideal situations. In most cases samples, both liquid and solid, need pretreatment, such as dissolution, dilution, extraction of the analyte, or separation of interferents. After such a treatment the only information that may be gained from the potentiometric measurement is the concentration of the analyte.

The correct interpretation of the relation between the activity and the concentration requires a proper choice of experimental conditions and a careful interpretation of the results. The most common procedure involves the use of an ionic strength adjustment solution (buffer). It consists of an inert salt at constant concentration that does not interact with the analyte (e.g.,  $\text{KNO}_3$ ), providing a solution of constant ionic strength, not influenced by small changes in the composition of the sample solution. Usually the other components have additional functions. For example, in total fluoride determination, metal ions, such as  $\text{Al}^{3+}$  or  $\text{Fe}^{3+}$ , which bind fluoride ions, should be complexed with cyclohexane-1,2-diamine- $N,N,N',N'$ -tetraacetic acid to liberate fluoride ions. In determination of total calcium content, controlled complexation of the analyte by addition of iminodiacetic acid eliminates the effects caused by the presence of other weaker complexants of various strengths and concentrations. Usually such an ionic strength adjustment solution also contains a pH buffer, which provides the optimal pH value for the ISE used. Under conditions of constant ionic strength there exists an exact correlation

between the potential reading and the logarithm of concentration of the analyte ion according to the Nernst equation.

A very special situation exists in the case of pH measurements based on measurements made using glass electrodes or other membrane sensors selective to hydrogen ions. By definition, the pH value is directly related to the logarithm of the activity of the hydrogen ion in solution. This is determined by calibrating the meter with primary (or secondary) pH standard buffers. However, aqueous standards may be usable only under specific conditions. When measurements are carried out under conditions differing from these, as in the presence of ethanol, the result of the measurement has only a relative value, e.g., indicating whether the samples differ from one another. Good results, less dependent on the sample material, are obtained from differential pH measurements. The possibility of measuring pH potentiometrically with high precision, coupled with the high specificity of enzymatic reactions, permits the determination of a variety of analytes (urea, citric acid, reducing sugars, alkaline phosphatase, pesticides, etc.) in milk, wine, and other fluid samples, with no or little sample pretreatment.

In principle ISEs are sensitive only to free (hydrated) ions of the analyte and do not respond to various complexed species of the analyte. This property could be exploited in speciation analysis, distinguishing the free and bound analyte species. However, it must be remembered that most procedures of sample pretreatment, dilution, extraction, and addition of reagents, change the initial speciation, corresponding to the nontreated sample. Therefore, only when the species are kinetically inert and the sample pretreatment is carried out with special caution is there a possibility of analyte speciation. For example, calcium in milk is present as a free (hydrated) ion or in combination with proteins, lipids, or better molecular mass species. The evaluation of each of the species is the aim of speciation analysis. The task is relatively simple when the species investigated are inert, do not dissociate, or otherwise undergo decomposition during sample pretreatment. A similar problem is common in the analysis of other biological samples, e.g., in blood analysis, where even a 20-fold dilution results in nearly complete dissociation of calcium complexes with various ligands.

Potentiometric (as well as other electrochemical) measurements have an important advantage over optical measurements in that the color or opacity of the sample solution do not obscure the correct determination of the analyte. On the other hand, the functioning of the electrodes in the presence of colloids and suspensions may also produce erroneous

results. This is described as a suspension effect that is due to blocking the electrode surface or the porous plug connecting the reference electrode. This requires frequent checking of the proper functioning of the electrode system, or even using special electrode systems to avoid trouble. Measurements using flow systems often diminish such effects.

The main limitation of the use of ISEs in practical analysis is in the range of analyte determination and the selectivity of the electrodes. Most commercially available electrodes allow precise measurements of the analyte down to  $10^{-5}$ – $10^{-6}$  mol l<sup>-1</sup> concentrations. For some crystalline electrodes this may be shifted down by some orders of magnitude, but only in the case when the species measured remains in a labile equilibrium in an ion (e.g., metal ion) buffer. Obviously measurements in the lowest concentration range suffer all the difficulties typical for trace analysis (e.g., contamination), which may affect the determination. The concentration range of analyte that may be measured using ISEs is usually between  $10^{-2}$  and  $10^{-4}$  mol l<sup>-1</sup>, and the sample size and its subsequent dilutions should be adjusted to these conditions.

Several electrodes of differing selectivity are commercially available at present for measurements of most common ions. Preliminary information may be obtained on the basis of tabulated values of the selectivity coefficients; nevertheless, it must be remembered that their values depend on the method of determination, on the manner in which the electrodes are maintained, and on the life of the electrode. Therefore it is a good habit to check the selectivity under standard conditions. After long periods of use the selectivity of the electrode may also change, giving rise to erratic results. For reliable measurements the potentiometric selectivity coefficient should not be greater than  $10^{-2}$  for ions of the same charge and for comparable concentration levels of the analyte and interferent.

Quantification of the measurements made using ISEs can be carried out in different ways. The most common one is based on the calibration graph, prepared using standard calibrating solutions adjusted to the expected concentration range of the analyte, with a similar addition of the ionic strength adjustment solution. Another possibility is the known addition or analyte addition mode. These modes may often help eliminate possible interferences. For determination of larger concentrations, ISEs are often used as endpoint indicators in titrations. The acidity of foodstuffs is determined routinely using acid–base titration with a pH-sensitive electrode, glass electrodes being used most commonly as the indicator. It should be mentioned that titration gives the concentration of the analyte directly.

## Electrode Types and Measurement Systems

The analytical characteristics of ISEs are essentially independent of the electrode construction. The classical electrode design with an internal reference solution and electrode is still most commonly used for electrodes with crystalline membranes as well as for electrodes with a polymeric membrane with a neutral or charged carrier. Polymeric electrodes of the ‘coated wire’ type are simple, but over a long period their standard potential may shift, and they need frequent calibration. The use of an intermediate layer, e.g., based on conducting polymers, significantly improves their response. Miniaturized electrodes with micrometer-size sensor tips are becoming common, especially for extremely small sample sizes. However, this is not common in the case of food sample analysis. Electrodes constructed using ion-selective field effect transistors (ISFETs) and those produced using screen-printed or thin-film technology are gaining interest.

Multimembrane systems are another type of potentiometric sensor based on ISEs, wherein the ISE response is modified by an additional membrane of a different function. To this group of sensors belong gas-sensitive electrodes, with a hydrophobic membrane isolating the analyte from the complex sample medium. The chemical reaction of the gaseous analyte with a component of the solution surrounding the ISE membrane proper directly produces the species to be sensed by the electrode. In food analysis, electrodes used for sulfur dioxide or ammonia belong to this group of sensors.

The next group of multimembrane systems comprises membranes sensitized biologically using immobilized enzymes or microorganisms. Species that are directly sensed by an ISE are produced in the enzymatic reaction of the analyte. Examples of such sensors are those used for determination of urea in milk, based on immobilized urease and measurement of a pH change. An example of the application of bacteria strains is the use of immobilized recombinant *Escherichia coli* coupled with a pH electrode. Such electrodes have been used for determination of cephalosporins. When this bacterial strain is coupled with a CO<sub>2</sub> gas sensor, glutamic acid determination can be carried out.

Techniques based on micromachining technology are being increasingly exploited in the construction of commercially available sensor arrays, which find interesting applications in food analysis. The individual potentiometric sensors composing an array may be based on principles similar to those of classical potentiometric sensors. For practical

**Table 1** Examples of applications of sensor arrays – electronic tongue and electronic nose – in food analysis

<i>Samples</i>	<i>Type of sensor array</i>
Fortified feed	Enzymatic (lysine oxidase) sensor + seven solid state sensors for elimination of ionic interferences on determination of lysine
Apple varieties	ISFET sensor array for determination of $H^+$ , $K^+$ , $Na^+$ , and $Ca^{2+}$
Soft drinks	Eight cross-sensitive polymeric membranes on a screen-printed carbon paste electrode
Beers	Twelve cross-sensitive polymeric membranes on a chip
Coffee varieties	A sensor array of 30 sensors

reasons they are used in a miniaturized form, e.g., ISFETs or metal oxide semiconductor field effect transistors. In simpler systems they may permit the determination of individual components, with the elimination of interference due to cross-sensitivity of different sensors. Their unique and main advantages are the possibility of performing a qualitative analysis and recognition and discrimination of complex liquid and gaseous samples. These are based on the use of a sensor array composed of a number of nonselective sensors, together with mathematical signal processing methods based on artificial intelligence and pattern recognition, such as artificial neural networks. Such systems when used in analyses of liquid samples are termed electronic tongues, and in the case of gaseous samples they are called electronic noses. They have been applied e.g., for distinguishing beers and varieties of fruits or for detection of adulteration of foodstuffs, as shown in Table 1.

## Applications

In this section applications related to drinking water and mineral water are not discussed as they generally pose no special problems and such samples may be similarly treated as model solutions.

Sodium, potassium, calcium, ammonium, fluoride, chloride, and nitrate determinations and measurement of pH are among the applications of ISEs in foodstuffs analysis. Sodium and potassium determinations are two of the assays most easily performed using ISEs.

Sodium is usually present in the ionized form and in consequence in solutions to be measured as the free hydrated ion, often in high concentrations. Under these conditions, the method of analyte addition is the most preferred one. For example, an amount of analyzed sample containing from 2 to 200 mg of sodium should be dissolved or leached with 100 ml

of distilled water. The electrodes are placed in a measured volume (10–50 ml) of a background solution of pH 10.2, containing a known concentration of  $0.1\text{--}10\text{ mmol l}^{-1}$  sodium chloride and  $0.5\text{ mol l}^{-1}$  triethanolamine. After the potential stabilizes, a small volume, 0.1–1.0 ml, of the sample is added. The conditions are suitable for measurements when the potential change after the sample addition is in the range 6–20 mV. The concentration of sodium in the sample solution is calculated from the potential increase and the slope of the sodium electrode calibration curve. At low levels of sodium the results may be erroneous due to inevitable contamination with sodium.

When the sodium and potassium levels are lower, the multiple known addition method is preferred. An example is in the determination of sodium and potassium in wine. The presence of ethanol at a level of 10% can affect the performance of the solvent polymeric membrane electrode and change to some extent the selectivity of every type of electrode. Therefore a tenfold dilution of wine samples or an addition of a comparable amount of alcohol to the standards is recommended. Samples can be diluted with a triethanolamine solution that provides proper pH buffering. A wine sample can also be mixed with an ionic strength adjustment solution. Several additions of standard solution are made and the results are calculated using Gran's method. In the case of potassium determination the results may be seriously distorted when diffusion of potassium ions from the reference electrode is not prevented using a double junction bridge containing lithium acetate.

Determination of calcium is usually based on the total calcium content as differentiation between free and bound calcium is possible only when the initial liquid sample is measured directly and has not been modified by pH adjustment. The optimal pH range is usually in the range from 5 to 9. Calcium is often determined in milk and milk products. Usually a  $4\text{ mol l}^{-1}$  solution of KCl is added to the sample as an ionic strength adjustment solution. The potential readings are compared against a calibration curve. For determination of total calcium content the sample is ashed and the residue is dissolved in a small volume of dilute hydrochloric acid and passed through an ion exchange column to remove pyrophosphates and hydrated silicates. Then the resulting solution is diluted to have the calcium concentration in the optimal concentration range, adjusted for pH value and ionic strength. The potential is measured and compared against a calibration curve. Alternatively, the standard addition method can be used for both determinations.



Fluoride is important for dental health but in excess amounts is known to be toxic. The only interfering ion for a fluoride electrode is the hydroxide ion. The determination of the fluoride ion is not disturbed by the presence of most accompanying ions except those that complex fluoride ions, such as aluminum or iron. To eliminate such interferences the ionic strength adjustment solution usually contains acetic acid buffer of pH  $\sim 4.5$  and a ligand, such as polyaminopolyacetate or citrate, which should complex the interfering metal ions. As the fluoride level in foodstuffs is usually low, the only problem is preparation of the sample to match the concentration in the final solution with the optimal determination range. The total fluoride content is measured after digestion of the sample. This can include ashing, fusion, oxygen flask combustion, and hot acid digestion. The best procedures involve the use of closed systems where the sample is decomposed with concentrated nitric acid at 100–120°C. Under these conditions loss of fluoride from the sample is avoided. The best procedures use the multiple standard addition method for fluoride determination. In analysis of flour or milk the treatment of the samples with perchloric acid makes it possible to determine fluoride at a level below 0.4  $\mu\text{g per g}$ . The content of free fluoride may be determined with minimum sample handling and treatment in a liquid sample e.g., wine.

The chloride content in foodstuffs can be determined by titration with silver nitrate using either a chloride ISE (crystalline or positively charged site membrane) or a silver electrode as the indicator. Direct determination with ISEs is convenient for low chloride contents; however, both procedures are affected by the presence of bromide or iodide when their concentrations are significantly larger than that of chloride. In protein-rich samples some irregularities of electrode functioning are often observed. Interference from bromide and iodide as well as from protein adsorption can be avoided by boiling the blended and mixed sample with 0.1  $\text{mol l}^{-1}$  nitric acid. For complicated food matrices a microdiffusion cell technique has been used to simplify the sample. The food sample is digested with cold concentrated sulfuric acid, and the diffusion process is allowed to proceed for  $\sim 24$  h. Thus chloride is converted into hydrochloric acid, which is transported in the diffusion cell into the receiving reagent. Then the chloride content is directly determined in the receiving reagent by comparison of a chloride ISE potential with the calibration curve.

Determination of nitrate is in principle simple; however, there are several interferences, the main ones being chlorides and hydrogencarbonates. They

may be eliminated by the addition of an ionic strength adjustment solution that is composed of 0.01  $\text{mol l}^{-1}$  silver sulfate, 0.06  $\text{mol l}^{-1}$  potassium sulfate, and sulfuric acid to acidify the solution to a pH value less than 4. Determination in potatoes needs only blending of the sample and extraction with distilled water and addition of the ionic strength adjustment solution. Some disruption may occur when outflow of chloride from the reference electrode is not prevented.

A rather rare example indicating the possibility of speciation analysis is the determination of ionic copper, in the range from 20 to 90  $\mu\text{g l}^{-1}$  using a copper-selective electrode in a sample of wine when the total copper content is in the range from 0.10 to 1  $\text{mg l}^{-1}$ . The sample is modified very little by the addition of a 10% volume of 1  $\text{mol l}^{-1}$   $\text{KNO}_3$ , as should be done with standard solutions.

ISEs have been used for determination of food additives such as saccharin and cyclamate. The sensor used is based on a poly(vinyl chloride) membrane containing the analyte salt in a suitable ionic form as a positively charged site e.g., with an ammonium or basic dye cation. Such sensors are not commercially available but can be made easily in the laboratory.

Enzyme electrodes, having an immobilized enzyme or a microbial culture containing an enzyme, operate on the basis of catalytic activity with subsequent detection of the product of the enzymatic reaction. The outcome, depending on the reaction, is e.g., a product like ammonia or a pH change in the medium, which is sensed by a corresponding sensing element.

Sensors used for determination of pesticide (proprhex, paraoxon) residues in vegetables are enzymatic multimembrane devices whose functioning is based on the principle of inhibition of the activity of an enzyme such as acetylcholine esterase. This reaction is monitored by a pH sensor. The response of such biosensors to herbicides and pesticides opens a new area of testing possibilities in food analysis.

Among the more complicated procedures involving ISEs, one can mention the determination of nitrite using a procedure that combines the derivatization of nitrite by diazotization of sulfanilic acid and coupling with 1-naphthylamine. The product of this reaction is detected potentiometrically using an ion-pair electrode with a membrane containing the anion paired with a nickel-phenanthroline complex. Such a procedure has excellent selectivity and permits detection below the microgram per gram of nitrite level in meat. Another nontrivial procedure is based on stoichiometric oxidation of glycerine using an excess of periodate and determination of the excess oxidant using an  $\text{IO}_4^-$ -selective electrode.



**Table 2** Examples of applications of ISEs in food analysis

Analyte	Sample	Electrode type
Acetate (Ac)	Vinegar	Ac <sup>-</sup> positively charged site polymer electrode
Ammonium, ammonia	Tea, juices, wine, shrimps	NH <sub>4</sub> <sup>+</sup> neutral carrier polymer electrode NH <sub>3</sub> gas electrode
Aspartame	Processed food, dietary food	Enzyme aspartase + NH <sub>3</sub> gas electrode
Benzoate (Bz)	Beverages, juices	Bz <sup>-</sup> neutral carrier polymer electrode
Bromide	Alfalfa	Br <sup>-</sup> crystalline electrode
Calcium	Meat, sugar, milk, fruits, wine, seaweed	Ca <sup>2+</sup> neutral carrier polymer electrode
Carbon dioxide	Beverages, wine	CO <sub>2</sub> gas electrode
Chloride	Various foods, cheese, meat, fish, cakes, canned vegetables	Cl <sup>-</sup> crystalline electrode or Cl <sup>-</sup> positively charged site electrode
Copper	Wine	Cu <sup>2+</sup> crystalline electrode
Cyanide	Alcoholic beverages	HCN gas electrode
Cyclamate (Cy)	Processed food	Cy <sup>-</sup> positively charged site electrode
Fluoride	Grain, milk, beer, cheese, fish, fruits, vegetable, wine, tea	F <sup>-</sup> crystalline electrode
Glycerol, glycol	Spirits, wine	IO <sub>4</sub> <sup>-</sup> positively charged site electrode
Iodide	Milk	I <sup>-</sup> crystalline electrode
$\beta$ -Lactams	Fermentation broth, milk	Microbial <sup>a</sup> + pH sensor
Nitrate	Various foods, wine, meat, sugar, spinach, potato	NO <sub>3</sub> <sup>-</sup> positively charged site electrode.
Nitrite	Meat	NO <sub>2</sub> <sup>-</sup> derivative positively charged site electrode
pH	Fruit juices, meat, milk, milk products, vinegar, beverages	Glass electrode or H <sup>+</sup> neutral carrier polymer electrode
Potassium	Wine, fish	K <sup>+</sup> neutral carrier electrode
Propoxur	Lettuce, onions	Enzymatic (AChE <sup>b</sup> ) inhibition + pH electrode
Saccharin (Sac)	Dietary products	Sac <sup>-</sup> positively charged site electrode
Sodium	Soup stock, dried milk, infant formulations, canned food	Na <sup>+</sup> neutral carrier electrode Na <sup>+</sup> glass electrode
Sulfide	Wine, fruit, vegetable	Ag <sub>2</sub> S crystalline electrode
Sulfur dioxide	Wine, processed food	SO <sub>2</sub> gas electrode
Urea	Milk	Enzyme urease + NH <sub>3</sub> gas electrode Bacterial cell <sup>c</sup> + NH <sub>4</sub> <sup>+</sup> neutral carrier electrode

<sup>a</sup> Containing  $\beta$ -lactamase.<sup>b</sup> AChE, acetylcholine esterase.<sup>c</sup> Containing urease.

A selection of some procedures used in food analysis is presented in Table 2.

Another example in food analysis is the use of solid state copper and silver electrodes for evaluation of meat freshness. This is due to the change of putrescine and dimethyl sulfide, the concentration of which changes during meat putrefaction. Such applications indicate the variety of applications of ISEs in food control and analysis.

See also: **Chemometrics and Statistics:** Multivariate Classification Techniques. **Elemental Speciation:** Waters, Sediments, and Soils. **Enzymes:** Overview; Enzyme-Based Electrodes. **Food and Nutritional Analysis:** Overview; Sample Preparation; Additives; Pesticide Residues; Soft Drinks; Coffee, Cocoa, and Tea; Alcoholic Beverages; Wine; Meat and Meat Products; Dairy Products; Vegetables and Legumes; Fruits and Fruit Products. **Ion-Selective Electrodes:** Overview; Glass. **pH. Sensors:** Overview. **Sweeteners. Titrimetry:** Potentiometric.

## Further Reading

- (1981) *A Guide to Ion Selective Electrode Analysis*. Halstead, Essex: Corning Ltd.
- (1982) *Handbook of Electrode Technology*. Cambridge, MA: Orion Research Incorporated.
- Hassan SSM, Marzouk SAM, and Sayour HEM (2003) Selective potentiometric determination of nitrite using a novel (4-sulphophenylazo-)-1-naphthylamine membrane sensor. *Talanta* 59: 1237–1244.
- Kaneki N, Miura T, Shimada K, *et al.* (2004) Measurement of pork freshness using potentiometric sensor. *Talanta* 62: 217–221.
- Kośmider J and Wyszynski B (2000) Artificial olfaction. *Chemia Analityczna – Chemical Analysis* 45: 483–500.
- Luca GC, Reis BF, Zagatto EAG, *et al.* (2002) Development of a potentiometric procedure for determination of glycerol and 2,3-butanediol in wine by sequential injection analysis. *Analytica Chimica Acta* 366: 193–199.
- Patel PD (2002) (Bio)sensors for measurement of analytes implicated in food safety: A review. *Trends in Analytical Chemistry* 21: 96–115.

- Schaller E, Bosset JO, and Escher F (1998) Electronic noses and their application to food. *Lebensmittel-Wissenschaft und Technologie* 31: 305–316.
- Strike DJ, Meijerink MGH, and Koudelka-Hep M (1999) Electronic noses: A mini-review. *Fresenius Journal on Analytical Chemistry* 364: 499–505.

Vlasov Yu and Legin A (1998) Nonselective chemical sensors in analytical chemistry: From 'electronic nose' to 'electronic tongue'. *Fresenius Journal on Analytical Chemistry* 361: 255–260.

## Water Applications

R O Ansell, Glasgow Caledonian University, Glasgow, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

It would be difficult to underestimate the importance of the analysis of waters. It has long been recognized that the use of water for human consumption depended on its purity. More recently environmental considerations have led to recognition of the importance of water purity in preservation of the aquatic environment and food chains within it. Over the past few decades the importance of water quality has come to be appreciated in all of the applications to which water is put. The analyses that are required to ensure that the water is of the correct quality are numerous and legislative standards are in place for many of these analyses. Ion-selective electrodes play two kinds of role in water analysis, the first in those cases where the ion-selective electrode method is the most analytically suitable and the second where the ion-selective electrode is the most appropriate because of the dynamic real-time analytical capability of the ion-selective electrode which allows continuous monitoring of the water.

The requirements for monitoring water quality occur in several environmental situations, in the monitoring of river quality, in the analysis and control of potable water supplies, and in sewage treatment. In addition, the quality of water is important in many industrial situations, for example, for monitoring the quality of boiler feed water, in food production, and in pharmaceutical production. Ion-selective electrodes play an important role in many of these areas.

### Measurement Considerations in Using Ion-Selective Electrodes for Concentration Measurement in Waters

Ion-selective electrodes measure activities rather than concentrations. They require a buffer to maintain the

ionic strength constant so that concentrations may be used instead of activities. The ion-selective electrode measures the free ion, chemical interferences can occur as a result of, for example, complexation or protonation. The ion-selective electrode has a response to the ion of interest but also responds to other ions in the solution. The ion-selective electrode's response to the ion that it is designed for is greater than that to the other ions present. Whilst the term ion-selective electrode is used in this article in common with general terminology the term ion-sensitive electrode would be a preferable description of the electrode's response. The response of the electrode to concentration is logarithmic, rapid (less than 1 min to stabilize unless concentrations are very low or for some liquid ion-exchange electrodes), and temperature dependent. The electrodes are not ultra-sensitive and frequent calibration is required for all electrodes. The electrodes and their associated measurement equipment are small and readily portable.

The ion-selective electrodes will function satisfactorily, if the precautions suggested by the manufacturer are followed; however, faults can be readily identified in ion-selective electrode measurement systems. The response of the electrode should not be slow and the theoretical slope of the electrode potential against logarithm of concentration plot at 25°C should be  $(58 \pm 1)/n$  mV where  $n$  is the number of electrons in the electrode reaction. Slow response times and deviations of slope from the theoretical value can indicate electrode faults and in water analysis particularly in *in situ* or continuous monitoring this can be due to fouling of the electrode by contaminants in the solution. The contaminants can sometimes be removed and the electrode will then function normally. The response time in flowing systems has to be carefully investigated because a slow response time may simply be associated with the passage of solution through the system flow cell rather than an electrode fault. Drift in the potential value obtained from the electrode can be due to temperature variation or in some cases inadequate stirring or flow. Failure to obtain suitable readings on

- Schaller E, Bosset JO, and Escher F (1998) Electronic noses and their application to food. *Lebensmittel-Wissenschaft und Technology* 31: 305–316.
- Strike DJ, Meijerink MGH, and Koudelka-Hep M (1999) Electronic noses: A mini-review. *Fresenius Journal on Analytical Chemistry* 364: 499–505.

Vlasov Yu and Legin A (1998) Nonselective chemical sensors in analytical chemistry: From 'electronic nose' to 'electronic tongue'. *Fresenius Journal on Analytical Chemistry* 361: 255–260.

## Water Applications

R O Ansell, Glasgow Caledonian University, Glasgow, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

It would be difficult to underestimate the importance of the analysis of waters. It has long been recognized that the use of water for human consumption depended on its purity. More recently environmental considerations have led to recognition of the importance of water purity in preservation of the aquatic environment and food chains within it. Over the past few decades the importance of water quality has come to be appreciated in all of the applications to which water is put. The analyses that are required to ensure that the water is of the correct quality are numerous and legislative standards are in place for many of these analyses. Ion-selective electrodes play two kinds of role in water analysis, the first in those cases where the ion-selective electrode method is the most analytically suitable and the second where the ion-selective electrode is the most appropriate because of the dynamic real-time analytical capability of the ion-selective electrode which allows continuous monitoring of the water.

The requirements for monitoring water quality occur in several environmental situations, in the monitoring of river quality, in the analysis and control of potable water supplies, and in sewage treatment. In addition, the quality of water is important in many industrial situations, for example, for monitoring the quality of boiler feed water, in food production, and in pharmaceutical production. Ion-selective electrodes play an important role in many of these areas.

### Measurement Considerations in Using Ion-Selective Electrodes for Concentration Measurement in Waters

Ion-selective electrodes measure activities rather than concentrations. They require a buffer to maintain the

ionic strength constant so that concentrations may be used instead of activities. The ion-selective electrode measures the free ion, chemical interferences can occur as a result of, for example, complexation or protonation. The ion-selective electrode has a response to the ion of interest but also responds to other ions in the solution. The ion-selective electrode's response to the ion that it is designed for is greater than that to the other ions present. Whilst the term ion-selective electrode is used in this article in common with general terminology the term ion-sensitive electrode would be a preferable description of the electrode's response. The response of the electrode to concentration is logarithmic, rapid (less than 1 min to stabilize unless concentrations are very low or for some liquid ion-exchange electrodes), and temperature dependent. The electrodes are not ultra-sensitive and frequent calibration is required for all electrodes. The electrodes and their associated measurement equipment are small and readily portable.

The ion-selective electrodes will function satisfactorily, if the precautions suggested by the manufacturer are followed; however, faults can be readily identified in ion-selective electrode measurement systems. The response of the electrode should not be slow and the theoretical slope of the electrode potential against logarithm of concentration plot at 25°C should be  $(58 \pm 1)/n$  mV where  $n$  is the number of electrons in the electrode reaction. Slow response times and deviations of slope from the theoretical value can indicate electrode faults and in water analysis particularly in *in situ* or continuous monitoring this can be due to fouling of the electrode by contaminants in the solution. The contaminants can sometimes be removed and the electrode will then function normally. The response time in flowing systems has to be carefully investigated because a slow response time may simply be associated with the passage of solution through the system flow cell rather than an electrode fault. Drift in the potential value obtained from the electrode can be due to temperature variation or in some cases inadequate stirring or flow. Failure to obtain suitable readings on

the pH meter often arises from there being a loss of contact in the system. This can be due to incompatible leads, bubbles in the system, and draining of solution from either external or internal reference electrodes. Electrical noise is often significant in ion-selective electrode measurements and care should be taken over the shielding of the electrodes. Particular problems arise in continuous flow systems often associated with noise from the pumping system.

The sequence of analytical measurements made in using the ion-selective electrode is carried out in the following stages: sampling, sample treatment, analysis, data acquisition, and data presentation. The sampling situation may involve discrete sampling, *in situ* sampling, or continuous monitoring.

## Sampling

The need for correct sampling procedures is particularly acute in the analysis of water samples. It is not possible to analyze the whole of a water body. The analysis may vary with position and time within the water body. In general, it is necessary to select positions so that the required information on the whole of the water body can be inferred from the sample. It is also necessary to ensure that the concentrations within the sample do not change from sampling to determination. The time variation in sampling can be overcome by the use of *in situ* measurement ('in line' or 'in stream' measurement) and ion-selective electrodes are particularly useful from this viewpoint; however, the use of this kind of measurement is not yet widespread. An alternative approach is to divert a stream of the water and carry out continuous analysis and this has been done in several cases. This type of measurement is known as online or on-stream measurement.

## Ion-Selective Electrodes for Use in Water Analysis

The ion-selective electrodes used in analysis are based on a system that comprises the ion-selective electrode membrane and a reference electrode. A variety of electrode types can be used for the analysis of species in water and these are described below, but the most important is undoubtedly the glass electrode. The glass electrode is used primarily for the determination of the pH of water samples and is widely used in water analysis.

### The Glass Electrode for the Determination of pH

The glass electrode is used for the measurement of the pH of natural waters. The electrode type can be

quite varied in terms of the size of the electrode probe, but the most common has a 1.2 cm thick stem that is ~12 cm long with a spherical bulb electrode at the end. The majority of electrodes have a silver/silver chloride internal reference electrode. The internal filling solution is a pH buffer solution and a chloride salt such as potassium chloride. The internal fill solution is chosen so that the electrode electromotive force (emf) is zero when the electrode is used with the appropriate external reference electrode, which is often a calomel electrode (either with a saturated or  $3.5 \text{ mol dm}^{-3}$  potassium chloride filling solution). Combined electrodes are now very common for pH measurement and these have an integral external reference electrode to which the solution to be measured has access through a ceramic frit liquid junction at the side of the lower part of the electrode. There are a wide variety of commercial pH meters available. The pH meter must be a specially designed meter that has an input impedance in excess of  $10^{12} \Omega$ .

The electrodes are normally stored in distilled or deionized water and should be rinsed with deionized water before use. The electrodes can be dried by draining the excess water off onto a tissue. The electrode must be standardized by immersion in buffer solutions whose pH are well separated. The procedure is to immerse the electrodes in the buffer solution for a period of 1 min with continuous stirring and observe the value which is then adjusted to the buffer value by the pH or calibration control. The electrode is then rinsed and dried and immersed in the second buffer solution and allowed to equilibrate for 1 min with stirring and the second buffer value is read. If this differs from that which is expected then the slope control of the pH meter should be altered. This whole procedure is then repeated until the values obtained in the buffer solutions at each end of the pH range agree with those of the known buffer solution.

**Measurements in low-conductivity waters** There has been a recent increase in interest in the measurement of pH in low ionic strength waters from upland areas because of the need to assess acidification of these waters. In low ionic strength waters, i.e., those that have a conductivity of less than  $100 \mu\text{S cm}^{-1}$ , problems in pH measurement can arise that are associated either with the low-conductivity solution or variation in the liquid junction potential. The variation in liquid junction potential is caused by blockage of the junction with precipitates and this is more of a problem for the silver/silver chloride reference electrode than the saturated calomel electrode. The problem of measurement in



low-conductivity solutions can be overcome if the electrodes used for measurement are performing correctly and this can be shown by a series of tests on them. At 25°C, the electrode slope factor should be within the range of  $58 \pm 1$  mV/pH unit. The residual liquid junction error that is estimated by measurement in a  $10^{-4}$  mol dm $^{-3}$  hydrochloric acid or sulfuric acid solution (a low-conductivity analytical quality control standard) and subtraction of the value 4.005 should be less than 0.05 pH. The pH shift on stirring obtained by comparison of the value obtained in the  $10^{-4}$  mol dm $^{-3}$  hydrochloric acid or sulfuric acid solution should be less than 0.02 pH and the noise in the pH measurement should be less than 0.02 pH.

The problem of the low conductivity of the water samples has also been addressed by modifications to the glass electrode and special low-resistance glasses have been developed to alleviate the problems associated with the low conductivity of the sample. These electrodes have impedance in the region of 5 M $\Omega$ .

Automated systems for low-conductivity waters are commercially available and have been produced specifically for use with the low-conductivity waters that are found in power stations.

**The glass electrode and sodium ion determination** The glass electrode can also be used for sodium ion determination and several commercial glass electrodes suitable for sodium ion determination are available. There is a very significant danger of sodium contamination of samples in a laboratory since there are many possible sources of sodium contamination. It is essential to avoid glassware in sodium ion determinations and plastic apparatus should be used. Equipment is commercially available for the online monitoring of sodium. Care is taken in such equipment to avoid the possibility of contamination of the sample and in one commercial instrument the water only comes into contact with Perspex, silicone rubber, stainless steel, and nonhalogenated plastics.

### The Fluoride Ion-Selective Electrode

Fluoride is naturally occurring in some waters and is added in small amounts to drinking waters. A level of 1 mg dm $^{-3}$  is recommended as being beneficial to the health of teeth and the limits in the United Kingdom for fluoride lie between 0.8 and 1.2 mg dm $^{-3}$ .

The fluoride ion-selective electrode is a particularly useful electrode in water analysis and is the basis of legislative procedures for the analysis of waters and effluents. The determination of fluoride ion in waters using the fluoride ion-selective electrode is not free from interferences. Consequently, although many potable waters can be determined directly some

cannot. The limit of detection of the ion-selective electrode method is 0.1 mg dm $^{-3}$ .

The fluoride ion-selective electrode responds directly to hydroxide ion but to few other common ions. The effect of the hydroxide ion is of concern if the fluoride ion concentration is not greater than 10 times that of the hydroxide ion. There are also problems, however, if the solution is too acidic as the formation of hydrogen fluoride reduces the concentration of fluoride. As a consequence of this, a total ionic strength adjustment buffer is used to maintain the pH between 5 and 6.

The fluoride ion-selective electrode is subject to interference by polyvalent cations, e.g., Ca $^{2+}$ , Fe $^{2+}$ , and Al $^{3+}$ , either acting on their own or in combination with other species and a complexant 1,2-diamino-cyclohexane-*N,N,N',N'*-tetraacetic acid is used to preferentially complex the interfering metal ions.

The fluoride ion-selective electrode requires equilibration time that is  $\sim 7$  min at concentrations of 0.1 mg dm $^{-3}$ .

**Electrode requirements** The calomel reference electrode should be of the sleeve tip rather than the fiber or ceramic plug type and care should be taken with the replacement of the reference electrode solution. The fluoride ion-selective electrode consists of a europium-doped lanthanum fluoride crystal. The emf response of the electrode to standard solutions should not be less than 55 mV over the range of concentrations of 0.2–200 mg dm $^{-3}$ . Values of the emf response lower than this can arise from poor matching of the sensor and reference electrodes, deterioration of the sensor electrode membrane, poor maintenance of the reference electrode, or poor maintenance of the sensor electrode. The analysis should be carried out in plastic vessels.

**Precautions in handling the fluoride ion-selective electrode** The lanthanum fluoride electrode is susceptible to mechanical damage and care should be taken to avoid situations where the electrode is subject to mechanical shock while in storage or use. If the electrode becomes coated with impurities it is possible to clean it by polishing with alumina on a suitable cloth. The electrode should not be subjected to solutions of pH less than 4 or fluoride concentrations greater than 20 g dm $^{-3}$ . The electrode will deteriorate depending on the amount of use it is subjected to within  $\sim 6$  months to 1 year.

### Ion-Selective Electrodes for the Determination of Ammonia

The electrodes that can be used for determining ammonia are the gas-sensing membrane electrode, the



ammonium-sensitive liquid membrane electrode, and the ammonium-sensitive glass electrode. The gas-sensing membrane electrode is the most sensitive and selective and this consists of a glass inner electrode that senses the change in pH of a film of ammonium salt solution trapped between the end of the glass electrode and a polymer membrane. The ammonium-sensitive liquid membrane electrode has been successfully used in pure waters such as boiler feed waters where it is more subject to alkali metal effects and has a higher limit of detection than the gas-sensing electrode. The ammonium-sensitive glass electrode is only suitable for use in high-purity waters as it is subject to serious interferences from alkali metal ions.

#### Liquid Ion-Exchange Electrodes for the Determination of Nitrate

These electrodes have nitrate-sensitive ion-exchange material incorporated into poly(vinyl chloride)-based membrane electrodes. Care is necessary to avoid contamination by the chloride from the saturated calomel reference electrode and a mercury/mercurous sulfate electrode is preferable as a reference electrode. Industrial monitors using nitrate ion-selective electrodes are commercially available.

#### Water Hardness Measurement

Water hardness is the total calcium and magnesium ion concentration in a water sample and is expressed as the concentration of calcium carbonate. Temporary hardness is that part of the total hardness that disappears on boiling. Whilst not being accepted as a standard method, the use of ion-selective electrodes allows a rapid measurement of water hardness and can be used to determine changes in hardness. The direct potentiometric method is not recommended for the ion-selective electrode but an indirect potentiometric method involving ethylenediaminetetraacetic acid titration is recommended. The ion-selective electrode that is used is a liquid ion-exchange electrode that responds to the divalent ions magnesium and calcium.

#### Carbon Dioxide Determination

Electrodes for the measurement of carbon dioxide are of the gas permeable membrane type. The membrane of the electrode is permeable to carbon dioxide which diffuses into a layer of sodium hydrogen carbonate solution that is trapped between a glass electrode and the membrane. The solution reacts with the carbon dioxide producing a pH change that is sensed by the glass electrode. The probe measures the free carbon dioxide in the sample. If the sample is acidified to pH 5 on the introduction of the probe, then the total carbon dioxide in the sample can be determined.

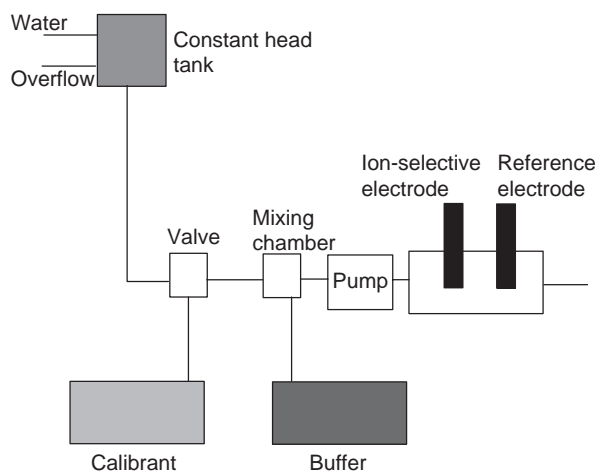
### Continuous Monitoring and Analysis in Flowing Solutions

The experimental arrangement required for the use of the ion-selective electrode in continuous monitoring of natural waters is shown in Figure 1. This type of measurement allows 'online' or 'on stream' measurement. The water stream to be analyzed is fed to a constant head tank fitted with an overflow. The water is fed from the tank into the system where it is pumped through the flow cell and buffer is added to allow measurement to take place. The standard solution (calibrant) is fed in at timed intervals by opening the valve. One of the most important parts of the continuous monitoring apparatus is the flow cell and flow cells have been developed with low sample requirements to minimize the need for refilling the buffer and calibrant reservoirs.

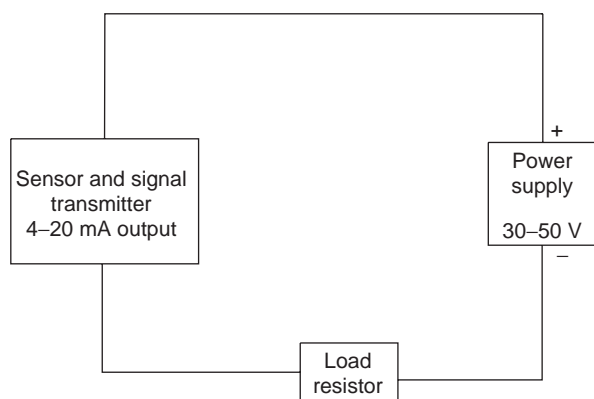
#### Instrumentation for Process Measurement

There are certain requirements for the instrumentation used in process measurement and standard measurement systems have been designed to allow these to be fulfilled. One of the most important aspects is ensuring that the signal can be read at a distance from the pH or ion-selective electrode without distortion due to electrical noise from the surrounding plant. The industrial standard method of doing this is to use a 4–20 mA loop (Figure 2).

The power supply supplies the power to the electronics associated with the sensor. The sensor returns a current that lies between 4 and 20 mA corresponding to its full output range. This current is converted to a voltage by the load resistor in the 4–20 mA loop which is typically a 500  $\Omega$  resistor giving a signal range from 2 to 10 V. The 4 mA



**Figure 1** Continuous monitoring using an ion-selective electrode.



**Figure 2** The 4–20 mA loop for signal transmission in process plant.

output that corresponds to the minimum output of the sensor allows a fault condition giving rise to 0 V to be distinguished from the minimum sensor output.

One of the primary advantages of the 4–20 mA analog communication is that the same two wires that are used to power the sensor also carry the signal output. The 4–20 mA system can transmit the signal over some kilometers. In addition, it is difficult to create unwanted current flows in the signal wire due to induced voltages as such voltages are small and are swamped by the loop driving voltage. With current signal-processing technology the sensor and associated electronics can be powered by a single shielded twisted wire pair that can also carry the output signal and in addition can also carry set-up calibration and sensor integrity information between the sensor and the control system. The electronics that convert the ion-selective or pH electrode signal to the 4–20 mA output are normally separate from the electrode and form part of the pH or ion-selective electrode meter.

### Noise Sources

In any process plant there will be several sources of electrical noise. The most common source is from 240 V, 50 Hz alternating current power circuits that can often be near to the electrodes and these result in sinusoidal interference signals at 50 Hz. Fluorescent lighting is also a common interference where the frequency of the interference is twice that of the mains supply voltage, i.e., 100 Hz. There may be additional noise sources in plant such as pumps or machinery, which require higher power levels than the normal main supply. It is important to have a clear understanding of the noise sources that can affect the measurement system and more details of these are given in the work by Bentley cited in the Further Reading section.

### Continuous pH Measurement

The pH meters used for continuous industrial measurement commonly have a standard 4–20 mA output. The amplifier unit will usually drive a separate display element remote from the measurement point. The pH electrode needs to be close to the amplifier since the electrode has very high impedance and hence requires short leads and careful screening. The amplifier will need to be protected against its environment and also the environment will need to be protected from the electrical hazard posed by the amplifier. Suitable intrinsically safe pH amplifier systems can be obtained commercially. Only the electrodes and the amplifier need to be located in the hazardous area where the measurement is taking place whereas the other equipment and display element can be located in a safe area.

The glass electrode for pH measurement is sufficiently stable to allow calibration at reasonably large intervals of time, most probably of the order of once a week and hence it is generally the case that pH electrodes are calibrated manually.

### In Situ pH Measurement

The pH electrode may be used to make measurements of the hydrogen ion activity by directly immersing the electrodes in the water flow. In this case, it is not possible to buffer the solution and since the pH measurement is normally made in this way there is no problem for pH measurement. Other ion-selective electrodes tend not to be used in this way as activity measurements are not normally used for other ions. Special cells are required for pH measurements, which are carried out *in situ*. These are of two types: cells for use in pipework where the cell is part of the pipe that carries the water, and dip cells that are used to dip the electrodes into ponds, lakes, reservoirs, and other water supplies.

### Continuous Ion-Selective Electrode Measurement

For ion-selective electrodes, in general, more frequent calibration is needed than for the glass electrode for the measurement of pH. As a consequence, automatic calibration systems have been developed. These systems allow automatic chemical standardization at set intervals, e.g., 3, 6, 12 h or 1, 2, 4, 8 days. The systems also allow manual or remote initiation of the standardization sequence at any time. The standardization systems are generally operated by solenoid valves, which stop the sample flow and allow the calibration solution to be fed into the cell. When the sensor has stabilized in the new solution the amplifier output is compared with a preset calibration value and any difference between the

calibration value and the preset value is backed off the amplifier output to adjust it to the correct value. The solenoid valve is then returned to its original position so that the sample can be monitored.

## Other Electrode Types

### Ion-Selective Field-Effect Transistors

Ion-selective field-effect transistors (ISFETs) are potentiometric sensors in which there is direct contact between the electroactive coating and the electrolyte. The ISFET works on the principle that the chemically sensitive layer responds to the analyte, and its potential with respect to the reference electrode is altered. The chemically sensitive layer is mounted on an insulator above the space between two n-type contacts to a silicon chip. The reference electrode is also connected to the silicon chip so that when the potential of the coating changes the potential in the region of the two n-type contacts also changes. The current flowing between the two n-type contacts is very dependent on the potential in this region and as a consequence of this the current is also sensitive to the presence of the analyte. Significant amplification of the signal due to the analyte is achieved in the ISFET.

The ISFETs are electrodes that are currently commercially available and include electrodes for the measurement of pH. The small size of the electrode and its lack of fragility are advantages that are important in other areas than water analysis, but the local amplification of the signal is likely to prove useful in water monitoring situations.

### Enzyme Electrodes

The use of enzyme-based electrodes has been particularly important in electrode development in the biosensing field and enzymes greatly enhance the specificity of the electrodes. The enzyme electrodes can be either potentiometric or voltammetric. Since electrodes based on enzymes are often stable over relatively short periods, electrodes with replaceable enzyme tips may prove useful in water monitoring.

*See also:* **Ion-Selective Electrodes:** Overview; Glass; Solid-State; Liquid Membrane; Gas Sensing Probes; Enzyme Electrodes; Clinical Applications; Food Applications.

### Further Reading

Bentley JP (1995) *Principles of Measurement Systems*, 3rd edn. London: Harlow Longman.

HMSO Series Methods for the Examination of Waters and Associated Materials: *Determination of the pH Value of Sludge, Soil, Mud and Sediment; and the Lime Requirement of Soil*, 2nd edn., 1992; *The Determination of the pH in Low Ionic Strength Waters*, 1988; *Fluoride in Waters, Effluents, Sludges, Plants and Soils*, 1982; *The Determination of Carbon Dioxide in Natural, Treated and Beverage Waters with a Supplement of Sampling Bottled and Canned Waters*, 1986; *5 Day Biochemical Oxygen Demand (BOD<sub>5</sub>)*, 2nd edn., 1988.

Midgely D and Torrance K (1991) *Potentiometric Water Analysis*, 2nd edn. Chichester: Wiley.

Yu TR and Ji GL (1993) *Electrochemical Methods in Soil and Water Research*. New York: Pergamon Press.

## ISE

*See* **ION-SELECTIVE ELECTRODES:** Overview; Glass; Solid-State; Liquid Membrane; Gas Sensing Probes; Enzyme Electrodes; Clinical Applications; Food Applications; Water Applications

## ISOTOPE DILUTION ANALYSIS

**E H Evans and R Clough**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Isotope dilution analysis (IDA) involves the modification of the natural isotopic composition of a target

analyte, contained within a sample, by the addition of an accurately known amount of an isotopically enriched analog (or spike) of the analyte. The concentration of the target analyte can then be calculated by measuring the resultant modified isotope amount ratio. The process is shown schematically in **Figure 1**. The modified isotope amount ratio can be measured using either mass spectrometry or radiochemical methods. It is isotope dilution mass

calibration value and the preset value is backed off the amplifier output to adjust it to the correct value. The solenoid valve is then returned to its original position so that the sample can be monitored.

## Other Electrode Types

### Ion-Selective Field-Effect Transistors

Ion-selective field-effect transistors (ISFETs) are potentiometric sensors in which there is direct contact between the electroactive coating and the electrolyte. The ISFET works on the principle that the chemically sensitive layer responds to the analyte, and its potential with respect to the reference electrode is altered. The chemically sensitive layer is mounted on an insulator above the space between two n-type contacts to a silicon chip. The reference electrode is also connected to the silicon chip so that when the potential of the coating changes the potential in the region of the two n-type contacts also changes. The current flowing between the two n-type contacts is very dependent on the potential in this region and as a consequence of this the current is also sensitive to the presence of the analyte. Significant amplification of the signal due to the analyte is achieved in the ISFET.

The ISFETs are electrodes that are currently commercially available and include electrodes for the measurement of pH. The small size of the electrode and its lack of fragility are advantages that are important in other areas than water analysis, but the local amplification of the signal is likely to prove useful in water monitoring situations.

### Enzyme Electrodes

The use of enzyme-based electrodes has been particularly important in electrode development in the biosensing field and enzymes greatly enhance the specificity of the electrodes. The enzyme electrodes can be either potentiometric or voltammetric. Since electrodes based on enzymes are often stable over relatively short periods, electrodes with replaceable enzyme tips may prove useful in water monitoring.

*See also:* **Ion-Selective Electrodes:** Overview; Glass; Solid-State; Liquid Membrane; Gas Sensing Probes; Enzyme Electrodes; Clinical Applications; Food Applications.

### Further Reading

Bentley JP (1995) *Principles of Measurement Systems*, 3rd edn. London: Harlow Longman.

HMSO Series Methods for the Examination of Waters and Associated Materials: *Determination of the pH Value of Sludge, Soil, Mud and Sediment; and the Lime Requirement of Soil*, 2nd edn., 1992; *The Determination of the pH in Low Ionic Strength Waters*, 1988; *Fluoride in Waters, Effluents, Sludges, Plants and Soils*, 1982; *The Determination of Carbon Dioxide in Natural, Treated and Beverage Waters with a Supplement of Sampling Bottled and Canned Waters*, 1986; *5 Day Biochemical Oxygen Demand (BOD<sub>5</sub>)*, 2nd edn., 1988.

Midgely D and Torrance K (1991) *Potentiometric Water Analysis*, 2nd edn. Chichester: Wiley.

Yu TR and Ji GL (1993) *Electrochemical Methods in Soil and Water Research*. New York: Pergamon Press.

## ISE

*See* **ION-SELECTIVE ELECTRODES:** Overview; Glass; Solid-State; Liquid Membrane; Gas Sensing Probes; Enzyme Electrodes; Clinical Applications; Food Applications; Water Applications

## ISOTOPE DILUTION ANALYSIS

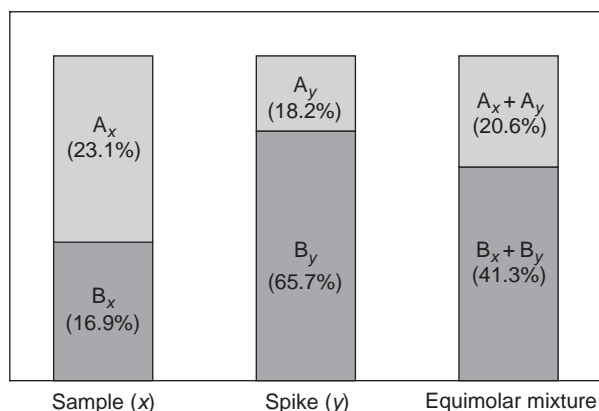
**E H Evans and R Clough**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Isotope dilution analysis (IDA) involves the modification of the natural isotopic composition of a target

analyte, contained within a sample, by the addition of an accurately known amount of an isotopically enriched analog (or spike) of the analyte. The concentration of the target analyte can then be calculated by measuring the resultant modified isotope amount ratio. The process is shown schematically in **Figure 1**. The modified isotope amount ratio can be measured using either mass spectrometry or radiochemical methods. It is isotope dilution mass



**Figure 1** Schematic representation of isotope dilution analysis showing the relative amounts of two isotopes, A and B, in the sample, the spike, and an equimolar mixture of the two. The figures in parentheses are the percent abundances of the isotopes before and after mixing. The example shown reflects the isotopic abundance of Hg and the figures do not add to 100% because several other isotopes contribute to the total (Table 3).

spectrometry (ID-MS) that will be described in detail here, though the general principles are much the same.

IDA is regarded as a definitive analytical technique because:

- it is directly traceable to SI units;
- the precision and accuracy obtainable are unsurpassed by alternative analytical methods; and
- it can account for analyte losses or incomplete extraction.

In ID-MS a spike solution, containing the isotopically enriched analog of the analyte, is added to the sample containing the natural abundance analyte. In order to measure the isotope ratio two isotopes (for inorganic ID-MS) or isotopomers (for organic ID-MS) are chosen, the sample (A) and the spike (B). The sample isotope, or isotopomer, is usually most abundant in the sample and the spike isotope, or isotopomer, is most abundant in the spike solution.

It should be noted here that there are some fundamental differences between organic and inorganic ID-MS. In organic ID-MS, the isotopically enriched analog is an organic molecule that has been labeled with (usually) either  $^{13}\text{C}$ ,  $^2\text{H}$ , or  $^{15}\text{N}$  on a nonlabile site. In order to prevent interference from low abundance isotopomers multiple labeling is usually performed so that the mass of the labeled spike differs from the analyte by at least 3 mass units. The spike isotopes have extremely low abundances in nature (Table 1), so the labeled spike isotopomer is of extremely low abundance in the sample. Conversely,

**Table 1** Natural atom fraction abundances of H, C, and N

Isotope	Natural abundance (%)
$^1\text{H}$	99.9
$^2\text{H}$	0.01
$^{12}\text{C}$	98.9
$^{13}\text{C}$	1.1
$^{14}\text{N}$	99.6
$^{15}\text{N}$	0.36

the sample isotopomer will be of extremely low abundance in the spike (Table 2).

In inorganic IDMS, the isotopically enriched analog is not an array of individual labeled molecules, but rather a mixture of individual isotopes that contribute directly to the enrichment of the array, but in different proportion to the sample (Table 3 and Figure 1). Ideally, the sample isotope should be of low abundance in the spike, and the spike isotope should be of low abundance in the sample, but this is very often not the case (Table 3). The isotope ratio is measured by ratioing the signal strengths of the spike and sample isotopes at their corresponding masses.

## Theory

### Inorganic ID-MS

The sample containing the analyte is spiked with the isotopically enriched analog. The sample isotope (A) and the spike isotope (B) are ratioed to give an isotope amount ratio ( $R_B$ ) in the sample/spike blend. This is a ratio of the number of moles of isotope A to the number of moles of isotope B, originating from both the sample and spike, and is given by eqn [1]:

$$R_B = \frac{A_y \cdot n_y + A_x \cdot n_x}{B_y \cdot n_y + B_x \cdot n_x} \quad [1]$$

where  $A_y$  is the isotope amount fraction of isotope A in the spike,  $A_x$  the isotope amount fraction of isotope A in the sample,  $B_y$  the isotope amount fraction of isotope B in the spike,  $B_x$  the isotope amount fraction of isotope B in the sample,  $n_y$  the number of moles of analyte in the spike solution, and  $n_x$  the number of moles of analyte in the sample.

Note that the subscript  $x$  is used to denote material of natural isotopic abundance and the subscript  $y$  is used to denote isotopically modified material.

Equation [1] can be solved for  $n_x$ , the amount in moles of analyte originally present in the unspiked sample (eqn [2]):

$$n_x = n_y \cdot \frac{A_y - R_B \cdot B_y}{R_B \cdot B_x - A_x} \quad [2]$$



**Table 2** Reference and spike masses of unlabeled and phenyl-<sup>13</sup>C6 (atom 90% enriched) labeled sulfamethazine

Isotopomer mass (Da)	Structure	Relative abundance	
		Sample	Spike
278 (sample)		1.0	<0.001
284 (spike)		<0.001	1.0

The relative abundances of the sample and spike isotopomers are given at the masses shown, not the abundances of all possible isotopomer masses.

**Table 3** Overall abundances in an array of <sup>199</sup>Hg-enriched Hg atoms in the sample and spike

Isotope	Overall Isotopic abundance in the array (%)	
	Sample	Spike
<sup>196</sup> Hg	0.15	0.15
<sup>198</sup> Hg	9.97	2.93
<sup>199</sup> Hg (spike)	16.87	65.98
<sup>200</sup> Hg (reference)	23.1	18.15
<sup>201</sup> Hg	13.18	3.96
<sup>202</sup> Hg	29.86	7.43
<sup>204</sup> Hg	6.87	1.44

It is more usual to quote the analytical result as a mass fraction, thus eqn [3] is used for the IDA calculation:

$$c_x = \frac{c_y w_y M_x}{w_x M_y} \frac{A_y - R_B \cdot B_y}{R_B \cdot B_x - A_x} \quad [3]$$

where, in addition,  $c_x$  is the mass fraction of analyte in the sample,  $c_y$  the mass fraction of analyte in the spike,  $w_y$  the mass of spike,  $w_x$  the mass of sample,  $M_x$  the molar mass of analyte in the sample, and  $M_y$  the molar mass of analyte in the spike.

Alternatively, the atom fraction ( $f$ ) can be calculated according to eqn [4], where one isotope is chosen as a reference isotope ( $r$ ) and the other isotopes are referenced to it (including the reference isotope). Hence, for an element having  $i$  isotopes the atom fraction of isotope  $q$  ( $f_q$ ) is given by

$$f_q = \frac{n_q/r}{\sum_i n_i/r} \quad [4]$$

where  $n_q$  is the isotope amount fraction of isotope  $q$ ,  $n_i$  the isotope amount fraction of isotope  $i$ , and  $r$  the isotope amount fraction of the reference isotope.

By using the terminology in eqn [2], where the subscript  $y$  denotes the enriched spike material and the subscript  $x$  denotes the natural isotopic abundance analyte in the sample, the isotope amount fractions can be expressed in the form of eqn [4]. If the spike isotope ( $B$ ) is used as the reference isotope then the isotope amount fraction of isotope  $A$  to isotope  $B$  in the spike ( $A_{\text{spike}}$ ) can be expressed as (eqn [5])

$$A_{\text{spike}} = \frac{A_y/B_y}{\sum_i n_y/B_y} \quad [5]$$

which can be simplified to (eqn [6])

$$A_{\text{spike}} = \frac{R_y}{\sum_i R_{iy}} \quad [6]$$

where  $R_y$  is the isotope amount ratio of isotope  $A$  to  $B$  in the spike,  $\sum_i R_{iy}$  the sum of the atom fraction ratios of all isotopes to the reference isotope.

Similarly, the isotope amount fractions of the spike and sample isotopes in the spike and sample ( $B_{\text{spike}}$ ,  $A_{\text{sample}}$ ,  $B_{\text{sample}}$ ) can be expressed as

$$B_{\text{spike}} = \frac{B_y/B_y}{\sum_i n_y/B_y} = \frac{1}{\sum_i R_{iy}} \quad [7]$$

$$A_{\text{sample}} = \frac{A_x/B_x}{\sum_i n_x/B_x} = \frac{R_x}{\sum_i R_{ix}} \quad [8]$$

$$B_{\text{sample}} = \frac{B_x/B_x}{\sum_i n_x/B_x} = \frac{1}{\sum_i R_{ix}} \quad [9]$$

where  $R_x$  is the isotope amount ratio of  $A$  to  $B$  in the natural sample and  $\sum_i R_{ix}$  the sum of the atom fraction ratios of all isotopes to the reference isotope.

Thus, the isotope amount ratio,  $R_B$ , of the sample to spike isotopes in the sample/spike blend can also be written as

$$R_B = \frac{(R_y / \sum_i R_{iy})n_y + (R_x / \sum_i R_{ix})n_x}{(1 / \sum_i R_{iy})n_y + (1 / \sum_i R_{ix})n_x} \quad [10]$$

which can be solved for  $n_x$ , the number of moles of analyte present in the unspiked sample giving eqn [11]:

$$n_x = n_y \frac{R_y - R_B \sum_i R_{ix}}{R_B - R_x \sum_i R_{iy}} \quad [11]$$

In this form of the ID-MS equation, the amount of substance can be directly replaced by an amount content (e.g.,  $\text{mol kg}^{-1}$ ) or mass fraction (e.g., ng per g), or any other appropriate unit provided the use of units remains constant throughout. Thus, eqn [11] can be expressed in terms of a mass fraction:

$$c_x = c_y \frac{m_y R_y - R_B \sum_i R_{ix}}{m_x R_B - R_x \sum_i R_{iy}} \quad [12]$$

where, in addition,  $c_x$  is the mass fraction of the analyte in the unspiked sample,  $m_x$  the mass of the sample,  $c_y$  the mass fraction of the isotopically modified spike, and  $m_y$  the mass of the isotopically modified spike added to the sample.

The advantage of using eqn [12] rather than eqn [3] is that the isotope amount fractions in the sample and the spike are not correlated, which simplifies the calculation of the measurement uncertainty of the analytical result.

### Organic ID-MS

The theory is identical for organic ID-MS but can be simplified somewhat. Organic compounds are usually labeled with  $^{13}\text{C}$ ,  $^2\text{H}$ , or  $^{15}\text{N}$ , each of which have only two isotopes (Table 1), so eqn [12] can be further simplified to

$$c_x = c_y \frac{m_y R_y - R_B R_x + 1}{m_x R_B - R_x R_y + 1} \quad [13]$$

A further simplification can be made if the spike isotope is of extremely low natural abundance and there is very little of the spike isotopomer in the sample. In this case,  $R_x$  tends toward infinity, so eqn [13] can be simplified to

$$c_x = c_y \frac{m_y R_y - R_B}{m_x R_y + 1} \quad [14]$$

If the isotopically enriched analog is of high purity and there is very little of the sample isotopomer present in the spike, then  $R_y$  tends to zero and a

further simplification can be made:

$$c_x = c_y \frac{m_y}{m_x} R_B \quad [15]$$

### Practice

The successful practice of IDA by mass spectrometry depends on a number of factors:

- more than one stable isotope (or isotopomer) in the spike and sample must exist;
- the isotopically differentiated peaks in the mass spectrum must be free from interference;
- an isotopically enriched analog, or spike, of the analyte must be available;
- complete equilibration between the spike and sample must be achieved;
- the mass fractions and isotopic abundances of sample and spike isotopes must be well characterized; and
- the spike and sample must be chemically stable.

### Isotopic Standards

The isotopic abundance of the naturally occurring isotopes in the sample and the isotopically enriched analog in the spike must be well characterized. This can be a particular problem in inorganic ID-MS if isotopes of the analyte in the sample vary in nature, or are man-made (e.g., as a result of nuclear reactions). If this is the case the isotopic abundances must first be determined before IDA can proceed. Isotopically enriched spike compounds can normally be purchased with a certificate stating concentration and isotopic abundance; however, it is good practice to at least verify the concentration. If the purity of the isotopically enriched analog has been characterized with sufficient accuracy and minimal uncertainty, the concentration can be calculated simply from knowledge of the masses of the compound and solvent employed. If the purity of the spike material is not certain, then reverse IDA is employed. In this case, the enriched spike material is treated as the sample and the isotopic abundance is modified by the addition of a certified standard of natural isotopic abundance that acts as the spike material.

### Spiking Procedure

In order to reduce systematic errors, such as ion counting errors due to detector dead time, it is better to fix the isotope amount ratio as close to unity as possible. In order to do this it is necessary to know the approximate concentration of the analyte in the sample prior to spiking. An exact 1:1 isotope amount ratio can be achieved by using an iterative matching

procedure; however, this can add considerable time to the method and is not always necessary. An approximate matching procedure confers many of the benefits of the exact matching procedure but is less time consuming. To avoid matrix effects and dilution of the sample, the spiking procedure follows the convention that the spike should be added in a small volume of a relatively high concentration.

For inorganic ID-MS, error propagation plots can be used to calculate the optimum analyte-spike isotope amount ratio for the minimization of errors during the measurement of the isotope amount ratio. This ratio can be calculated for a particular isotope pair using eqn [16], examples of which are shown in Figure 2. In practice, the isotope amount ratio should lie between 1:4 and 4:1 to give a reasonable signal for each isotope, and it is probably simplest to use an isotope amount ratio as close to unity as possible:

$$\sqrt{(B_y/A_y) \times (B_x/A_x)} \quad [16]$$

where  $A_y$  is the abundance of the sample isotope in the spike,  $B_y$  the abundance of spike isotope in the spike,  $A_x$  the abundance of sample isotope in the sample, and  $B_x$  the abundance of spike isotope in the sample.

### Extraction and Equilibration

A major advantage of IDA is that complete extraction of the analyte from the sample matrix is not

necessary to achieve accurate results; however, complete equilibration of the spike and sample isotopes is vital.

For the determination of organic compounds by IDA it is necessary to extract the analyte from the sample without causing chemical breakdown of the analyte. Complete extraction of the analyte from the sample is not necessary for successful IDA but it is important to ensure complete equilibration of the spike and sample, such that partitioning between solid, liquid, and gaseous phases is the same for both the spike and sample isotopomers. For example, if solvent extraction of a compound from a soil sample is undertaken, the spike and analyte should partition to the same extent between the solvent and soil particles.

In comparison, for elemental analysis, the integrity of the analyte is not an issue, so much harsher sample pretreatment can be performed using inorganic acids. This breaks down the sample and extracts the analyte completely into the liquid phase. However, it is still necessary to ensure that the spike and sample isotopes are in the same chemical form, so repeated oxidation/reduction cycles may be necessary. This is particularly true if thermal ionization mass spectrometry (TIMS) is used, because different inorganic oxidation states can have quite different thermal properties.

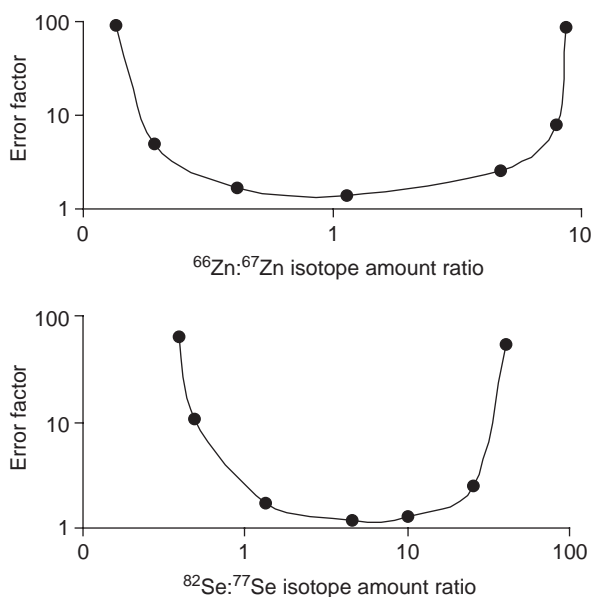
### Measurement of Isotope Amount Ratios

Once the sample has been spiked, equilibrated, and prepared in a way suitable for introduction into the instrument, the sample-spike isotope amount ratio ( $A:B$ ) must be measured. The observed isotope amount ratio must be corrected for the effects of spectroscopic interferences, mass fractionation in the sample introduction system and/or ion source, and mass bias in the mass spectrometer and associated ion optics.

**Spectroscopic interferences** Nonspectroscopic instrumental interferences are accounted for because they have the same effect on both the spike and sample signals, so are cancelled in the measured isotope amount ratio. However, spectroscopic interferences often apply to either the spike or sample isotopes alone and so are not accounted for. In inorganic ID-MS, simple isobaric interferences such as  $^{204}\text{Hg}$  on  $^{204}\text{Pb}$  can be corrected for by measuring the signal at  $^{202}\text{Hg}$  and using an elemental equation to obtain the true signal for Pb at  $m/z$  204:

$$^{204}\text{Pb}_{\text{true}} = ^{204}\text{Pb}_{\text{obs}} - (^{202}\text{Hg}_{\text{obs}} \times (a_{204\text{Hg}}/a_{202\text{Hg}})) \quad [17]$$

where  $^{204}\text{Pb}_{\text{true}}$  is the true signal for  $^{204}\text{Pb}$  at  $m/z$  204,  $^{204}\text{Pb}_{\text{obs}}$  the observed signal for  $^{204}\text{Pb}$  at  $m/z$  204,



**Figure 2** Error factors calculated for different isotope amount ratios of Zn and Se. (Adapted from Clough RC, Truscott J, Belt ST, *et al.* (2003) Isotope dilution ICP-MS for speciation studies. *Applied Spectroscopy Reviews* 38: 101–132.)

$^{202}\text{Hg}_{\text{obs}}$  the observed signal for  $^{202}\text{Hg}$  at  $m/z$  202,  $a_{204\text{Hg}}$  the atom fraction of the  $^{204}\text{Hg}$  isotope, and  $a_{202\text{Hg}}$  the atom fraction of the  $^{202}\text{Hg}$  isotope. If the interference is caused by a polyatomic ion (e.g.,  $^{40}\text{Ar}^{16}\text{O}^+$  on  $^{56}\text{Fe}$ ) the interference can sometimes be removed instrumentally or resolved if the mass spectrometer has sufficient resolving power.

Likewise, in organic ID-MS, multiple labeling of the isotopically enriched spike is usually performed so that its mass differs from the analyte by at least 3 mass units. This ensures that interference from the  $^{13}\text{C}$  sample isotopomer at  $M + 1$  does not occur.

**Mass fractionation** It occurs when the spike and sample isotopes, or compounds that contain them, undergo differential separation during the sample introduction or ionization process. For example, if the analyte is introduced as a gas, molecules containing the lighter of the sample or spike isotopes will diffuse slightly quicker than molecules containing the heavier isotope, hence mass fractionation will occur. This is a particular problem when gas chromatography is used as the sample introduction method and, in extreme cases, two distinct chromatographic peaks will result. For the same reason, mass fractionation can occur during the combustion stage in isotope ratio mass spectrometry, or in ionization sources that rely on thermal ionization processes such as TIMS.

If fractionation occurs in a chromatographic step then each complete peak must be integrated and the integral used to calculate the isotope amount ratio.

**Instrumental mass bias** It occurs because mass spectrometers and their associated ion optics do not transmit ions of different mass equally. In other words, if a sample composed of two isotopes with an exactly 1:1 molar ratio is analyzed, a 1:1 isotope amount ratio will not necessarily be observed. This so-called mass bias depends on mass and the type of mass spectrometer used, but generally tends to be greatest at low mass, and decreases with increasing mass. Even very small mass-biases can have deleterious effects on the accuracy of ID-MS, so a correction must often be made, usually in one of two ways as follows.

**Bracketing** If an isotopic standard of known composition for the isotope pair under study is available, then a correction can be applied as shown in eqn [18]:

$$C = \frac{R_{\text{true}}}{R_{\text{obs}}} \quad [18]$$

where  $C$  is the mass bias correction factor,  $R_{\text{true}}$  the true or certified isotope amount ratio for the isotope pair, and  $R_{\text{obs}}$  the observed or measured isotope amount ratio for the isotope pair. In practice, the isotopic standard is analyzed before and after the sample (i.e., the sample is bracketed) and the mean correction factor calculated for the bracketing pair is applied to the sample.

**Interpolation** In inorganic ID-MS, it is also possible to use an alternative element, with an isotope pair of similar mass to the isotope under study, and which has a certified isotope amount ratio. The mass bias correction can be performed by interpolating (with increasing accuracy), using either linear, power, or exponential equations:

$$\text{Linear} \quad \frac{R_{\text{obs}}}{R_{\text{true}}} = 1 + B\Delta m \quad [19]$$

$$\text{Power} \quad \frac{R_{\text{obs}}}{R_{\text{true}}} = (1 + B)^{\Delta m} \quad [20]$$

$$\text{Exponential} \quad \frac{R_{\text{obs}}}{R_{\text{true}}} = \exp(B\Delta m) \quad [21]$$

where  $B$  is the mass bias and  $\Delta m$  the mass difference between the isotopes used to calculate the ratio,  $R$ .

The Russell correction is a particular case of the generalized power law:

$$\text{Russell} \quad \frac{R_{\text{obs}}}{R_{\text{true}}} = \left(\frac{m_2}{m_1}\right)^f \quad [22]$$

where  $m_1$  and  $m_2$  are the masses of the isotopes used to calculate the ratio,  $R$ ; and  $f$  is the mass bias correction factor. For example, the  $^{205}\text{Tl}/^{203}\text{Tl}$  ratio can be used to correct the mass bias of  $^{200}\text{Hg}/^{199}\text{Hg}$  using the Russell correction as shown in eqn [23]:

$$\left(^{200}\text{Hg}/^{199}\text{Hg}\right)_{\text{true}} = \frac{\left(^{200}\text{Hg}/^{199}\text{Hg}\right)_{\text{obs}}}{\left[\frac{\left(^{205}\text{Tl}/^{203}\text{Tl}\right)_{\text{obs}}}{\left(^{205}\text{Tl}/^{203}\text{Tl}\right)_{\text{true}}}\right]^{\left[\frac{\ln(M^{200}\text{Hg}/M^{199}\text{Hg})}{\ln(M^{205}\text{Tl}/M^{203}\text{Tl})}\right]}} \quad [23]$$

where  $M$  is molar mass of the isotope.

The advantage of this approach is that the mass bias correction can be performed by spiking the sample with a mass bias correction standard (e.g., Tl) and measuring this isotope amount ratio at the same time as the sample.

**Detector dead time** Also important is the effect of detector dead time when using certain types of detector. For example, when ions are detected using an

electron multiplier in pulse counting mode the resultant electronic pulses are  $\sim 10$  ns long. During and after each pulse there is a period of time during which the detector is effectively 'dead' (i.e., it cannot detect any ions). The dead time is made up of the time for each pulse and recovery time for the detector and associated electronics. Typical dead times vary between 20 and 100 ns. If dead time is not taken into account there will be an apparent reduction in the number of pulses at high count rates, which would cause an inaccuracy in the measurement of isotope ratios when abundances differ markedly. However, a correction can be applied as shown in eqn [24]:

$$C_{\text{true}} = \frac{C_{\text{obs}}t}{t - C_{\text{obs}}D} \quad [24]$$

where  $C_{\text{true}}$  is the true number of counts,  $C_{\text{obs}}$  the observed number of counts,  $t$  the time spent monitoring each mass, and  $D$  the dead time.

The corresponding equation for count rate is

$$R_{\text{true}} = \frac{R_{\text{obs}}}{1 - R_{\text{obs}}D} \quad [25]$$

where  $R_{\text{true}}$  is the true count rate and  $R_{\text{obs}}$  the observed count rate.

If no dead time correction is applied then a linear calibration would not be possible, since the higher count rates between  $10^4$  and  $10^6$  Hz would be underestimated. This provides a way of determining the dead-time empirically, i.e., by re-integrating the isotopic signals with different dead times until a linear calibration is obtained for a series of accurately known standards.

A better method, which accounts for any instrumental drift, is to measure the isotope amount ratio of two isotopes with different abundances, and use the following expression:

$$(R_m - CR_M) = D[R_m R_M(1 - C)] \quad [26]$$

where  $R_m$  is the count rate of the minor isotope,  $R_M$  the count rate of the major isotope, and  $C = R_m/R_M$ .

The isotope amount ratio  $C$  must be calculated in absence of dead-time effects (i.e., at low count rates, but not so low as to give bad counting statistics) by repeated measurements of the blank subtracted isotope amount ratio. This is the instrumental isotope amount ratio, and no correction is made for mass bias. The count rates are then measured for a series of concentrations and the data, which has not been corrected for dead time, can be used to plot  $(R_m - CR_M)$  against  $[R_m R_M(1 - C)]$ , the slope of the line being equivalent to the dead time,  $D$ , in seconds. The data that are plotted must fall within the

range of count rates at which dead-time effects become significant (i.e., between  $\sim 10^3$  and  $\sim 10^6$  Hz); otherwise, a curve rather than a straight line will result. In practice, the effect of dead time will not be significant as long as the count rate is below  $10^5$  counts per second. More important is the effect of low count rate on precision, which means that the longest possible time should be allowed for ion counting when the count rate is below  $\sim 500$  counts per second, though this will depend on the amount of sample that is available.

**Blank correction** In order to achieve the best accuracy and precision, blank determinations should be made using the same procedure as for the sample; however, complications can arise.

If the blank signal is small, then it may not be possible to spike the blank to achieve the optimal ratio and still observe a measurable signal for both isotopes. In this case, the spike-sample ratio should be  $\sim 10:1$  to achieve best precision.

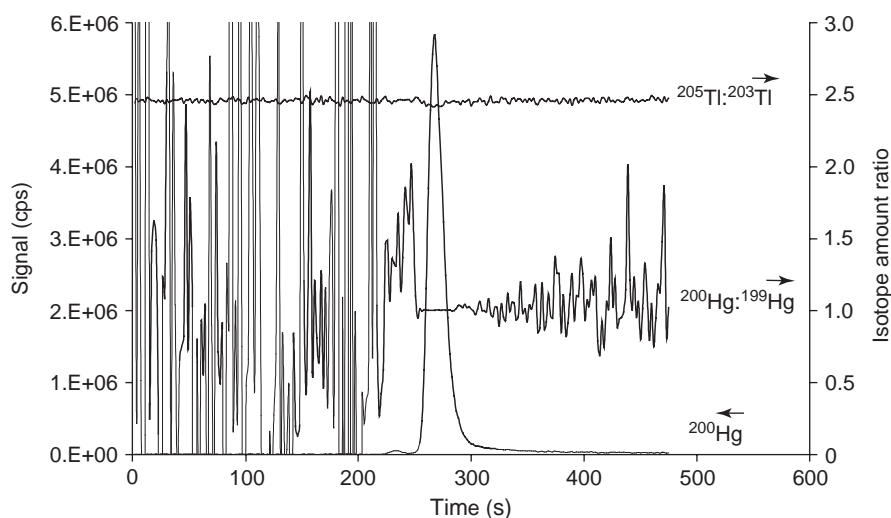
If the isotope abundances of the analyte isotopes vary in nature then the situation may arise where the isotopic abundances in the blank and sample differ (e.g., Pb contamination in reagents used for sample preparation). If the blank is large relative to the sample then it will make a significant contribution to the overall isotopic abundance of the sample when presented for analysis, which will be intermediate between the original sample and the blank. In this case, the isotopic abundances in the blank and the prepared sample (i.e., including the blank contribution) must be first determined, before the concentration of each can then be determined by IDA for subsequent blank subtraction. Alternatively, the blank signals for the spike and reference isotopes can be subtracted from the sample signals before the isotope amount ratio is calculated, though this will possibly cause a degradation in precision and accuracy.

## Applications of IDMS

### Elemental IDMS

Isotope dilution as a quantitative technique for elemental analysis was first applied by Reynolds, in 1950, to determine the amounts of decay products after neutron irradiation of  $\text{Cu}^{64}$ ,  $\text{Br}^{80}$ ,  $\text{Br}^{82}$ , and  $\text{I}^{128}$ . TIMS was the mainstay of this type of analysis for many years until the advent of inductively coupled plasma mass spectrometry (ICP-MS), which allows more rapid determination of a wider range of elements. The technique of IDA is traceable to the mole and so is the method of choice whenever a high





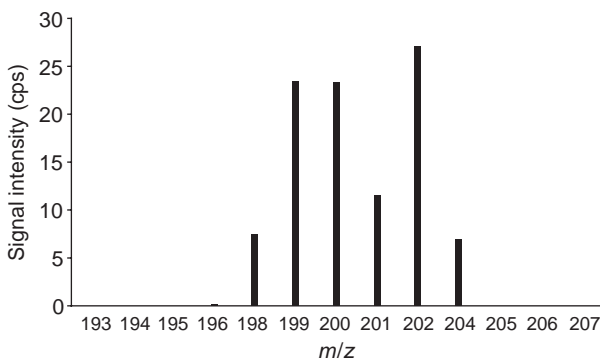
**Figure 3**  $^{200}\text{Hg}$ : $^{199}\text{Hg}$  and  $^{205}\text{Tl}$ : $^{203}\text{Tl}$  isotope amount ratios and  $^{200}\text{Hg}$  signal of 900 ng per g methyl mercury monitored using multicollector, magnetic sector ICP-MS. (Adapted from Clough *et al.* (2003) *Journal of Analytical Atomic Spectrometry* 18: 1039–1046.)

degree of accuracy, precision, and traceability is required, such as for the preparation of certified reference materials. It is also useful when analyte losses during sample preparation are likely to occur. The latest generation of ICP-MS instruments, equipped with double focusing magnetic sector analyzers and multiple, simultaneous ion detectors are capable of isotope amount ratio precision of 0.01% or better.

A recent development is species-specific IDA, whereby organometallic compounds or elemental species can be chromatographically separated, introduced into the ICP-MS instrument, and the isotope amount ratio of the elemental moiety can be determined for the purposes of IDA. An example of this is the speciation of methylmercury in environmental samples, where the sample is spiked with an analog of methylmercury enriched in  $^{199}\text{Hg}$ . The resulting multiple ion chromatogram obtained after online chromatographic separation is shown in **Figure 3**. The mass spectrum taken at the chromatographic peak is shown in **Figure 4**, with isotopic abundances of the Hg isotopes reflecting the mixing of the sample and spike (i.e., intermediate between the two).

### Organic IDMS

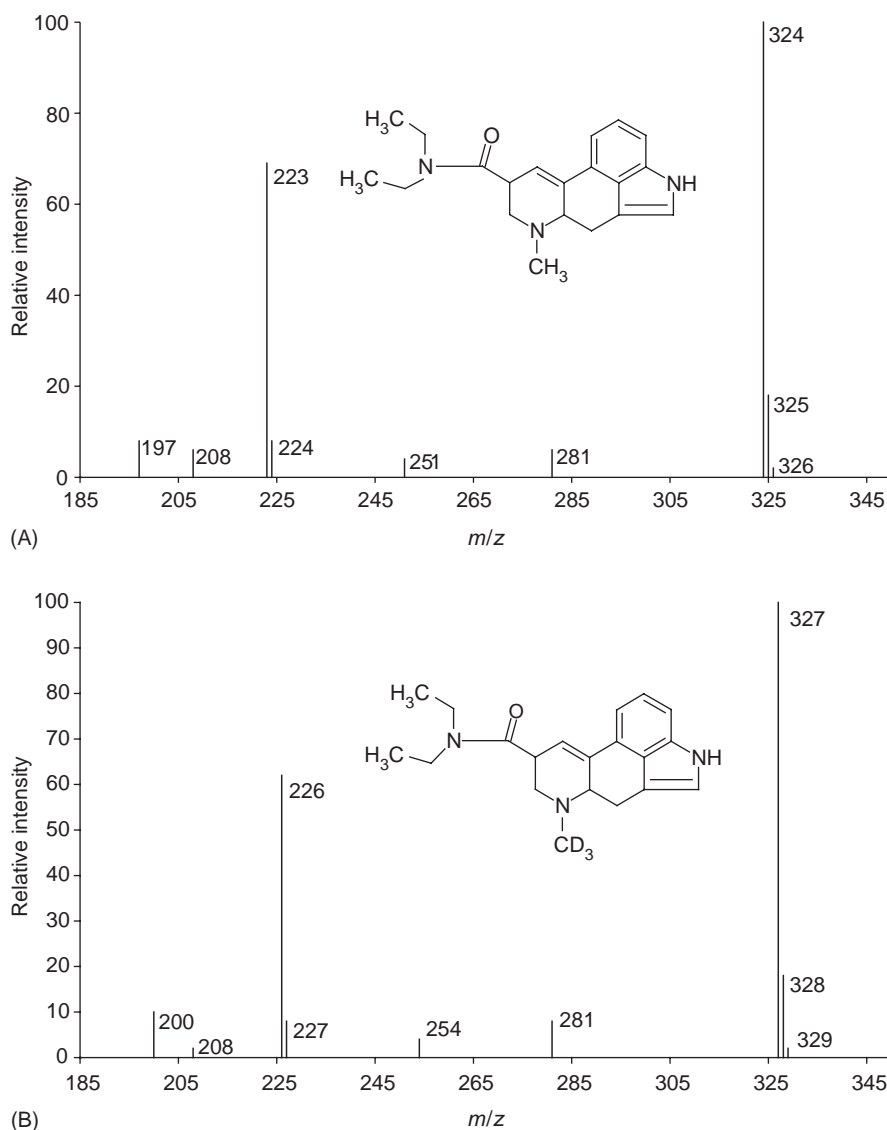
During the 1970s, isotope dilution was introduced for the quantitation of organic analytes by gas chromatography–mass spectrometry (GC–MS), with spike compounds labeled with  $^{13}\text{C}$ ,  $^2\text{H}$ , or  $^{15}\text{N}$ . Since then it has been used for the determination of compounds ranging from ethanol in water to cholesterol in human serum. ID-MS is ideally suited to this type of measurement because the complex sample matrix



**Figure 4** Mass spectrum of Hg taken at the chromatographic peak maximum shown in **Figure 3**.

presents difficulties when external calibration is applied, due to factors such as different analyte ionization efficiencies, unless the standards are closely matrix matched to the samples. For example, electrospray ionization mass spectrometry (ESI-MS), which is one of the most widely used methods for the determination of nonvolatile compounds, has greatly varying ionization efficiencies depending on the polarity and charge in solution of the analyte compound.

An example of using ESI-MS for IDA of lysergide is shown in **Figure 5**. In this case, the isotopically enriched analog has been triply deuterated so that the protonated molecular ion ( $M + H$ )<sup>+</sup>, which normally occurs at  $m/z$  324, is shifted to  $m/z$  327. Because the natural abundance of deuterium is so low (0.01%) the degree of enrichment of the deuterated spike isotopomer is extremely high, and so two distinct peaks appear in the mass spectrum, separated by 3 mass units.



**Figure 5** Structure and fragmentation induced ESI mass spectrum of lysergide, showing: (A) the natural abundance and (B) the deuterated analog. (Adapted from White SA, Kidd AS, and Webb KS (1999) The determination of lysergide (LSD) in urine by high-performance liquid chromatography-isotope dilution mass spectrometry (IDMS). *Journal of Forensic Science* 44: 375–379.)

See also: **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Gas Chromatography:** Mass Spectrometry. **Liquid Chromatography:** Liquid Chromatography–Mass Spectrometry. **Mass Spectrometry:** Ionization Methods Overview; Electrospray; Stable Isotope Ratio.

## Further Reading

Ingle CP, Sharp BL, Horstwood MSA, Parrish RR, and Lewis DJ (2003) Instrument response functions, mass bias and matrix effects in isotope ratio measurements and semi-quantitative analysis by single and multi-collector ICP-MS. *Journal of Analytical Atomic Spectrometry* 18: 219–229.

Platzner IT (1997) *Modern Isotope Ratio Mass Spectrometry*. New York: Wiley.

Sargent M, Harte R, and Harrington C (2002) *Guidelines for Achieving High Accuracy in Isotope Dilution Mass Spectrometry (IDMS)*. Cambridge: Royal Society of Chemistry.

White SA, Kidd AS, and Webb KS (1999) The determination of lysergide (LSD) in urine by high-performance liquid chromatography-isotope dilution mass spectrometry (IDMS). *Journal of Forensic Science* 44: 375–379.

Wolff-Briche CSJ, Hernandez H, O'Connor G, Webb KS, and Catterick T (2000) High accuracy isotope dilution analysis for the determination of ethanol using gas chromatography-combustion-isotope ratio mass spectrometry. *Analyst* 125: 2189–2195.

# ISOTOPE RATIO MEASUREMENTS

**J F Carter and R Sleeman**, Mass Spec Analytical Ltd., Bristol, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Atomic nuclei (nuclides) consist of positively charged protons and uncharged neutrons; particles collectively known as nucleons, which interact through a short-range attractive force that holds the nucleus together. As the number of protons in the nucleus (the atomic number  $Z$ ) increases more neutrons are required to prevent the nucleus from breaking apart under the strain of proton–proton repulsion. Nuclides are most commonly referred to by the elemental symbol preceded by the mass number ( $A$ ), e.g.,  $^{12}\text{C}$ ,  $^{238}\text{U}$ . Isotopes are atomic species of the same element that differ in atomic mass; that is, nuclei with a given number of protons but differing numbers of neutrons. The term ‘isotope’ was first used by Soddy in 1913, the same year in which the existence of naturally occurring isotopes was first reported by J.J. Thompson, who discovered two stable isotopes of neon (mass 20 and 22) using one of the earliest mass spectrometers. Today, a range of isotopes has been discovered or made for every known element, the vast majority being made by nuclear reactions using man-made sources of nucleons.

The principle technique used for the measurement of isotope ratios is mass spectrometry (MS) using various combinations of ion source and mass analyzer for the analysis of many organic and inorganic materials. Less frequently employed are spectroscopic techniques such as optical absorption and emission, and nuclear magnetic resonance (NMR).

## Isotope Ratios

The precise and accurate determination of isotope ratios is applied to many different fields, such as: measurements of stable isotopes in nature, especially for the investigation of isotope variation in geological and cosmic samples or age dating; determining isotope ratios of radiogenic elements in the nuclear industry; quality assurance of fuel material for reprocessing plants; nuclear material accounting and radioactive waste control; and tracer experiments using highly enriched stable isotopes or long-lived radionuclides in biological or medical studies.

The elements that form biological molecules (hydrogen, carbon, nitrogen, oxygen, and sulfur) have

two or more stable isotopes, the lighter isotope in each case being far more abundant. Natural and artificial chemical reactions fractionate isotopes, thus leading to the occurrence of organic and inorganic materials with different isotopic compositions. Studies over 50 years have shown that isotopic fractionation due to physiological processes, specifically  $\text{CO}_2$  transport within plants and photosynthesis, leads to variations in  $^{13}\text{C}/^{12}\text{C}$  isotope ratios in natural compounds.

These variations result from a kinetic isotope effect (KIE), which is defined as the ratio of reaction rate constants for reactant molecules containing the light and heavy isotope:

$$\text{KIE} = k_{\text{light}}/k_{\text{heavy}}$$

KIEs depend upon differences in vibrational energy levels and are proportional to the square root of the inverse of the atomic mass. Hence, for heavy atoms the typical KIE will be very small ( $^{35}\text{Cl}/^{37}\text{Cl} \sim 1.01$ ) but increases rapidly as the atomic mass decreases ( $^{13}\text{C}/^{12}\text{C} \sim 1.04$  and  $^2\text{H}/^1\text{H} \sim 8$ ).

Isotopic composition can be expressed as either absolute or relative measurement, depending on the particular application. When isotopically labeled materials are used as tracers absolute values are determined. The isotopic (or radiochemical) purity is the percentage of label present in the specified chemical form that may include the position of the label or the enantiomorphic form of the compound. In contrast, radionuclidic purity is the percentage of the total radioactivity present as the specified radionuclide and implies nothing about the chemical form of the radionuclides present.

Relative measurements are most commonly used to express small variations in the natural abundance of ‘light’ stable isotopes and find applications in areas such as the geosciences and ecological research. No consensus exists on the appropriate units in which isotope ratios are reported, especially in the biomedical field. Most isotope ratio mass spectrometers (IRMS) report isotopic abundance in terms of delta notation (‘parts per thousand’ or ‘per mil’), which is a convention determined by geochemistry, because most of the original IRMS instruments were developed in isotope geochemistry laboratories. Delta units are not SI units. The SI base unit for quantity is the mole, from which atom fraction and mole fraction are derived. The units of stable isotope abundance, at.% and mol.%, are the atom and mole

fractions expressed as percentages. At.% excess and mol.% excess are the SI units of enrichment and are recommended for use in tracer studies.

## Absolute Abundance

Atom per cent (at.%) is an absolute measure of the number of atoms of a particular isotope present in 100 atoms of that element:

$$\begin{aligned}\text{For example, at.\% } ^{15}\text{N for air} \\ &= \left( \frac{^{15}\text{N}/^{14}\text{N}}{^{15}\text{N}/^{14}\text{N} + 1} \right) \times 100\% \\ &= \left( \frac{0.0036765}{0.0036765 + 1} \right) \times 100\% \\ &= 0.3663\%\end{aligned}$$

Fractional abundance is a simple expression of isotope abundance, very similar to at.%, and is used to express isotopic abundance in ppm:

$$\begin{aligned}\text{For example, ppm } ^2\text{H of standard water} \\ &= \left( \frac{^2\text{H}/^1\text{H}}{^2\text{H}/^1\text{H} + 1} \right) \times 10^6 \text{ ppm} \\ &= \left( \frac{0.00015576}{0.00015576 + 1} \right) \times 10^6 \text{ ppm} \\ &= 155.74 \text{ ppm}\end{aligned}$$

## Relative Isotope Ratio

Natural abundance measurements often require measurements of absolute variations in the third or fourth decimal place of an at.% measurement. It is, therefore, better to use a relative measurement where the sample isotopic composition is compared to an international standard in terms of per mil (‰) versus an international standard:

$$\begin{aligned}\delta (\text{‰}) &= \left( \frac{R_{\text{sample}} - R_{\text{reference}}}{R_{\text{reference}}} \right) \times 1000\text{‰} \\ &= \left( \frac{R_{\text{sample}}}{R_{\text{reference}}} - 1 \right) \times 1000\text{‰}\end{aligned}$$

where  $R$  is the absolute ratio of the heavy and light isotopes and it follows that all international standards have a delta value of 0‰ as their ratios cancel out.

In practice, an instrument records the ratio of two signals corresponding to the two isotopes,  $Rm$ .

For example

$$Rm \text{ of CO}_2 = \frac{\text{Minor Beam}}{\text{Major Beam}} = \frac{m/z \text{ 45}}{m/z \text{ 44}}$$

The 'raw' delta value of the sample is given by

$$\delta_{\text{sample(raw)}} = \left( \frac{Rm_{\text{sample}}}{Rm_{\text{reference}}} - 1 \right) \times 1000\text{‰}$$

Reference materials are used to calibrate appropriate 'working standards' that are used in individual laboratories for routine analyses. On most occasions the reference material is not an international standard but will have a known delta value ( $\delta_{\text{reference}}$ ). However, all values measured relative to the working standard must be reported relative to the primary standard. The true delta value of the sample is then given by

$$\begin{aligned}\delta_{\text{sample(true)}} &= \delta_{\text{sample (raw)}} + \delta_{\text{reference}} \\ &\quad + \left( \frac{\delta_{\text{sample (raw)}} \times \delta_{\text{reference}}}{1000} \right) \times 1000\end{aligned}$$

It is possible to rearrange this equation to deduce absolute ratios in terms of at.% and fractional isotope abundance.

The primary standards for the five major elements in the biosphere are standard mean ocean water (SMOW) for hydrogen and oxygen, Pee Dee belemnite (PDB) for carbon (and sometimes oxygen), atmospheric air for nitrogen and Canyon Diablo troilite (CDT) for sulfur (Table 1). Supplies of both SMOW and PDB have been exhausted and replaced by reference materials that have been calibrated against the original standards, available from the International Atomic Energy Agency (IAEA), Vienna, Austria. To note this change relative isotopic measurements should now be reported as versus VPDB or VSMOW ('V' for Vienna). PDB is a carbonate mineral and since chemical processes discriminate against  $^{13}\text{C}$  the vast majority of organic compounds are depleted in  $^{13}\text{C}$  versus VPDB. Hence, a number of suitable reference materials are now available, e.g., polyethylene foil (PEF1  $\delta^{13}\text{C} = -31.77\text{‰}$  versus VPDB) and sucrose (Sucrose ANU  $\delta^{13}\text{C} = -10.47\text{‰}$  versus VPDB). Similarly, because of the huge variation in  $^2\text{H}/^1\text{H}$  in global precipitation further reference materials have been made available including; Greenland ice sheet precipitation (GISP,  $\delta\text{D} = -189.5\text{‰}$  versus VSMOW) and standard light Antarctic precipitation (SLAP,  $\delta\text{D} = -428\text{‰}$  versus VSMOW).

## Mass Spectrometry

The ability to determine isotopic abundances is a main feature of MS. Even with low-resolution instruments the presence of isotopes of known natural

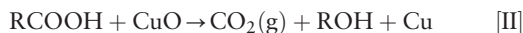
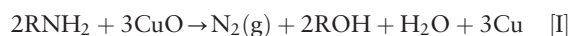
**Table 1** The primary standards for the five major elements in the biosphere

Element	Natural abundance (at.%)	International standard	Ratio of standard	<i>m/z</i> Ratio measured
Hydrogen	<sup>1</sup> H 99.985 <sup>2</sup> H 0.015	SMOW	<sup>2</sup> H/ <sup>1</sup> H 0.00015576	H <sub>2</sub> 3/2
Carbon	<sup>12</sup> C 98.892 <sup>13</sup> C 1.108	PDB	<sup>13</sup> C/ <sup>12</sup> C 0.0112372	CO <sub>2</sub> 45/44
Nitrogen	<sup>14</sup> N 99.6337 <sup>15</sup> N 0.3663	Air	<sup>15</sup> N/ <sup>14</sup> N 0.0036765	N <sub>2</sub> 29/28
Oxygen	<sup>16</sup> O 99.759 <sup>17</sup> O 0.0374 <sup>18</sup> O 0.2039	PDBSMOW	PDB <sup>18</sup> O/ <sup>16</sup> O 0.00206710 SMOW <sup>18</sup> O/ <sup>16</sup> O 0.00200520	CO <sub>2</sub> 46/44 CO 30/28 O <sub>2</sub> 34/32
Sulfur	<sup>32</sup> S 95.018 <sup>34</sup> S 4.215	CDT	<sup>34</sup> S/ <sup>32</sup> S 0.0450045	SO <sub>2</sub> 66/64 SO 50/48

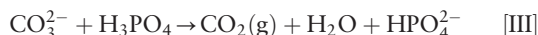
abundance makes possible a method for deducing the elemental composition of many ions derived from organic compounds. The isotopic abundance of the elements can be classified into three general categories: 'A' those elements with only one natural isotope in appreciable abundance, 'A + 1' those elements with two natural isotopes the second of which is 1 Da heavier than the most abundant, and 'A + 2', those elements that have an isotope 2 Da heavier than the major isotope. The presence of elements in an ion is often easily recognized from the 'isotopic cluster' produced in the spectrum. For example, the presence of 1.1% <sup>13</sup>C contributes to the ion 1 Da heavier than the molecular ion. Increasing the number of carbon atoms in the ion increases the probability that one of these atoms will be a <sup>13</sup>C isotope. A C<sub>10</sub> ion will, therefore, contribute  $10 \times 1.1\% = 11\%$ . This fact allows the number of carbon atoms to be deduced. When interpreting organic mass spectra the natural variation in <sup>13</sup>C content of organic compounds is generally ignored, as is the contribution from the abundance of <sup>2</sup>H, <sup>15</sup>N, and <sup>17</sup>O.

High-precision IRMS for the determination of variations in isotope abundance was described by Nier in 1940. The essential components of the instrument remain largely unchanged to the present day. IRMS instruments are highly specialized for the analysis of <sup>13</sup>C/<sup>12</sup>C, <sup>2</sup>H/<sup>1</sup>H, <sup>15</sup>N/<sup>14</sup>N, <sup>18</sup>O/<sup>16</sup>O, and <sup>34</sup>S/<sup>32</sup>S via analysis of gases CO<sub>2</sub>, H<sub>2</sub>, N<sub>2</sub>, CO, and SO<sub>2</sub> using a tight electron ionization ion source, high transmission magnetic sector, and multiple collectors, delivering relative standard deviations of less than 0.01%. The use of multiple Faraday cup collectors, as opposed to a single electron multiplier, leads to reduced sensitivity but greatly enhanced precision. An essential feature of instruments designed for precise measurement of relative isotope ratios is a system that admits gas to the ion source, through a capillary leak, from a dual inlet or continuous flow. This permits the rapid and repeated comparisons of sample and reference gases.

Organic forms of an element can be converted directly to a gaseous form by modified Dumas combustion in evacuated sealed tubes using copper(II) oxide as the oxygen donor (reactions [I] and [II]), requiring ~10 µg to 1 mg of sample for combustion followed by N<sub>2</sub> or CO<sub>2</sub> transfer to the MS:



Alternatively, an inorganic form of the element can be conveniently converted to the desired gaseous form. A standard method for preparing CO<sub>2</sub> is the controlled reaction of carbonate with phosphoric acid (reaction [III]):



Oxygen isotope ratios may be directly measured as CO<sub>2</sub> when produced by acid reaction with carbonate or as CO when a compound is pyrolyzed. Water samples are commonly analyzed for <sup>18</sup>O/<sup>16</sup>O by equilibration with a headspace of CO<sub>2</sub>.

When measuring <sup>13</sup>C/<sup>12</sup>C ratios of CO<sub>2</sub> it is necessary to correct the interference from the isobaric ion at *m/z* 45 corresponding to <sup>12</sup>C<sup>17</sup>O<sup>16</sup>O. This is achieved by measuring *m/z* 46 (<sup>12</sup>C<sup>18</sup>O<sup>16</sup>O), since the abundance of <sup>18</sup>O may be used to correct the 45/44 ratio. The formation of HCO<sub>2</sub><sup>+</sup> from CO<sub>2</sub> and background H<sub>2</sub>O in IRMS is also troublesome because it is not resolved from <sup>13</sup>CO<sub>2</sub><sup>+</sup> and the resulting artificial enhancement of *m/z* 45 can cause systematic errors.

## Continuous Flow Mass Spectrometry Techniques

The introduction of continuous flow techniques to IRMS instrumentation permits a wide range of samples to be analyzed by highly automated systems



and has heralded a huge expansion in the use of stable isotopes in biomedical and environmental sciences. Continuous flow isotope ratio mass spectrometry (CF-IRMS) minimizes sample preparation and reduces the sample size requirements and time for the determination of carbon and nitrogen. In the most common configuration of CF-IRMS an elemental analyzer is coupled to the MS via an open-split. Using modified Dumas combustion samples are converted to  $N_2$  and  $CO_2$ , which are separated by gas chromatography (GC) and sequentially introduced to the IRMS allowing both  $\delta^{15}N$  and  $\delta^{13}C$  to be measured from a single aliquot of sample.

Online oxygen and deuterium isotope ratio determinations for organic and inorganic substances are now possible through high-temperature conversion to  $H_2$  and  $CO$  gases. Samples are pyrolytically decomposed at temperatures up to  $1450^\circ C$  in the presence of chromium or nickelized graphite. Gas/liquid equilibrators have automated H and O isotopic analysis of water in untreated aqueous fluids as complex as urine. Automated cryogenic concentrators permit analysis at ppm concentrations in environmental samples.

The development of GC coupled via a combustion furnace to an IRMS (GC-C-IRMS) has allowed the analysis of individual compounds occurring at trace levels in very complex mixtures. Sometimes referred to as compound specific isotope analysis or isotope ratio monitoring MS (GC-irmMS), this technique has opened new fields of research in areas such as organic geochemistry, food science, medicine, nutrition, sport, forensic science, archaeology, soil science, and extraterrestrial science. The chromatographic separation in connection with the combustion of the analyte, however, exerts the strongest influence on the uncertainty of the measurement. Multidimensional GC (GC/GC) has also been coupled to IRMS for the authentication of flavor components.

Since GC is limited to the analysis of volatile substances, liquid chromatography has been coupled to an IRMS allowing the analysis of high molecular weight, polar, and thermally labile materials. Because of its technical complexity this technique has found only limited use to date.

## 'Inorganic' Isotopes

Inorganic MS methods are widely used for the determination of isotope ratio measurements in an extensive range of materials, e.g., conducting, semiconducting, and nonconducting solid samples, water, organic solvents, and solutions.

Traditionally, isotope ratios have been determined using thermal ionization MS (TIMS) and to some extent glow discharge MS (GDMS). TIMS and GDMS use high-resolution mass analyzers but differ in analyte ionization methods. TIMS uses electrons from a hot filament, whereas GDMS uses direct atomization and ionization via cathodic sputtering of the sample surface in a low-pressure argon atmosphere. The GD ion source is highly efficient and is capable of measuring 99% of the elements in the periodic table. Its chief limitation is that samples must conduct electricity. TIMS can be used in negative or positive ion modes with high sensitivity and precision for a broad range of isotopes. However, time-consuming sample preparation and sample determination processes limit the applications for routine analysis.

The development of plasma source MS has enabled the determination of isotope ratio composition of metallic solid samples without requiring sample purification and has effectively replaced TIMS and GDMS in many areas. Plasma source MS uses an Ar inductively coupled plasma (ICP) as an ionization device. The quadrupole ICP-MS is less precise than TIMS but is a popular method for isotope ratio determination because of its speed and convenience. Samples may be introduced as solutions, slurries, or suspensions of very fine particles generated by laser ablation and nearly complete analytical coverage of the periodic table is available. Sample matrices, memory effects, and some instrument parameters, however, may affect the accuracy and precision of isotope ratio determinations if adequate precautions are not taken. A new generation of ICP-MS instruments using high-resolution mass analyzers now provides better sensitivity and precision than quadrupole ICP-MS. The most significant instrument improvement for isotope analysis by sector field ICP-MS has been the application of multiple ion collection devices (MC-ICP-MS) to achieve better precision of isotope ratio measurements. MC-ICP-MS is used for the precise measurement of variations in the isotopic composition of a wide range of elements and diverse applications include measurement of radiogenic and stable isotope ratios and the determination of trace element concentrations by isotope dilution.

Secondary ion MS (SIMS) employs an energetic primary ion beam of  $Ga^+$ ,  $Cs^+$ , or  $O^-$  for analyte ionization. SIMS can determine isotope ratios in intact solid samples without destroying them but has poorer resolution and sensitivity than TIMS and is difficult to standardize for biological samples. The instrument is often termed an 'ion microprobe', because it can display images showing the distribution

of specific masses on a specimen with sub-micrometer spatial resolution. SIMS is a widely used analytical method with demonstrated applications in forensic science but is being superseded by laser ablation ICP-MS (LA-ICP-MS), which provides excellent sensitivity, precision, and good accuracy for isotope ratio measurements.

Both ICP-MS and LA-ICP-MS are increasingly replacing other MS techniques that have been dominant analytical methods for precise isotope ratio measurements for many decades. For all types of inorganic IRMS, however, precision and accuracy are primarily limited by sample preparation, introduction, and analytical methodology.

## Radioisotopes

The majority of isotopes of the elements are unstable with respect to other nuclides, and spontaneously transform to more stable nuclides. These unstable species are referred to as radionuclides or radioisotopes. Two key applications of isotope ratio measurements to radioisotopes are the characterization of fissionable materials and radiocarbon dating.

The reaction between atmospheric  $^{14}\text{N}$  nuclei and cosmic rays (neutrons) produces the radionuclide  $^{14}\text{C}$ , accompanied by the emission of a proton.  $^{14}\text{C}$  produced in this manner enters the Earth's carbon cycle from which some carbon may be trapped in a form that ceases to exchange with atmospheric carbon (e.g., as rocks or plant material). As  $^{14}\text{C}$  has a half-life of 5730 years, the determination of the  $^{14}\text{C}$  to ( $^{12}\text{C} + ^{13}\text{C}$ ) ratio has become a useful technique for estimating the age of objects from 1000 to 10 000 years timescale. The  $^{14}\text{C}/^{12}\text{C}$  ratio of the atmosphere is only  $1.2 \times 10^{-12}$  and determination of this ratio requires the use of accelerator mass spectrometry (AMS). AMS utilizes a type of mass spectrometer with a van der Graaff electrostatic accelerator to accelerate negative ions produced in a SIMS or FAB type ion source at MV potentials. The ions are stripped of electrons at the terminal of the accelerator to destroy molecules isobaric with the isotope being analyzed and then reaccelerated as positive ions. The high-energy positive atomic ions are then separated using magnets and velocity filters followed by identification of the total energy and energy loss of each ion as it is individually counted in the detector. AMS can measure isotope ratios as low as  $1:10^{-14}$  with precision better than 1% and is applied to the minor isotopes of other elements such as  $^{10}\text{Be}/^9\text{Be}$ ,  $^3\text{H}/^1\text{H}$ , or  $^{36}\text{Cl}/^{35}\text{Cl}$ . AMS can also reduce the degree of radioisotope labeling required for

medical studies to measure the amount of chemical substance absorbed, distributed and eliminated from various parts of the body.

## Other Analytical Techniques

Although spectroscopic methods are far less precise than IRMS for isotope ratio measurements, they can have advantages in terms of costs and ease of operation. These techniques also remove the laborious chemical preparation required for conventional IRMS. They are most commonly applied to the analysis of samples that are artificially enriched in the heavier isotopes of nitrogen, carbon, and oxygen.

Atomic emission spectroscopy may be used to distinguish different isotopic species, e.g.,  $^{14}\text{N}^{14}\text{N}$ ,  $^{14}\text{N}^{15}\text{N}$ , and  $^{15}\text{N}^{15}\text{N}$ , present in a low-pressure gas. The sample molecules are excited by an external microwave energy source and emit ultraviolet electromagnetic radiation at specific wavelengths: 297.7, 298.3, and 298.9 nm, corresponding to the nitrogen isotopomers above. Ratios of the intensities at these wavelengths are related to the isotopic composition of the nitrogen sample.

Direct infrared laser absorption spectroscopy has been applied to low-pressure gas-phase water samples for the accurate, simultaneous determination of the relative  $^2\text{H}/^1\text{H}$ ,  $^{17}\text{O}/^{16}\text{O}$ , and  $^{18}\text{O}/^{16}\text{O}$  isotope abundance ratios of water in highly enriched samples, as used in the biomedical doubly labeled water method to quantify energy metabolism. In contrast, photoacoustic spectroscopy has also been applied to isotope ratio measurements of  $^{18}\text{O}/^{16}\text{O}$  in water vapor samples using a pulsed tuneable dye laser to provide a transportable system for *in situ* and rapid measurements in the environment.

Both nondispersive infrared spectroscopy and cavity ring down spectroscopy (employing a near-infrared external cavity diode laser) can be used to determine the  $^{13}\text{C}/^{12}\text{C}$  ratio in the  $\text{CO}_2$  of expired breath. This technique can be used to diagnose liver cirrhosis or *Helicobacter pylori* infection after feeding  $^{13}\text{C}$  labeled substrates to patients.

Laser absorption spectroscopy may also be applied to the analysis of heavy elements. For example, the isotopes of  $^{235}\text{U}$  and  $^{238}\text{U}$  may be simultaneously measured by absorption in laser-induced plasma using two separate diode lasers tuned to the absorption lines of  $^{235}\text{U}$  and  $^{238}\text{U}$ . Using this technique uranium-containing minerals may be probed directly without any preceding preparation steps.

Whilst IRMS and spectroscopic techniques precisely measure the overall isotopic content of a

molecular species, the determination of  $^2\text{H}/^1\text{H}$  and  $^{13}\text{C}/^{12}\text{C}$  ratios at each atomic site of a pure substance can be performed by site-specific natural isotope fractionation NMR spectroscopy (SNIF-NMR). The isotopic pattern of a given molecule is a rich source of information on reaction pathways and on genealogy of the atoms, on the relative rates of certain reaction steps, on the percentage of intra- and intermolecular hydrogen transfer, and possibly on the stereospecificity of reactions. The disadvantage of the technique arises from the fact that the sample must be very pure (95–100%) and since deuterium is a relatively insensitive nucleus and is present in low abundance, relatively large sample sizes and long experiment times are required. In addition, deuterium NMR spectra are not well resolved and quantification of overlapping peaks can be problematic.

Natural abundance isotope ratios provide powerful analytical criteria for authenticating the origins of products. Used alone or in combination with IRMS the SNIF-NMR technique has been used for many years for determining geographical origin, adulteration, and authenticity of many natural products (wine, tobacco, fruit juice, etc.). For example, the site-specific  $^2\text{H}$  abundance of the methyl and methylene sites of ethanol present in alcoholic drinks, determined by SNIF-NMR, may be used to infer meteorological parameters associated with the growing period and evaporation processes in casks during the aging period.

## Applications

One of the major uses of isotopically labeled molecules is as tracers for chemically similar nonlabeled compounds to follow the behavior of organic molecules in complex media such as living organisms and ecosystems. Isotope labeling provides many methods to study mineral metabolism in humans, ranging from relatively simple measurements of absorption, and fecal excretion, to complex studies using multicompartmental modeling. Most studies use highly enriched stable isotope tracers administered intravenously and/or orally, coupled with precise measurements of the resulting isotopic enrichment in blood, urine, or feces. The use of illicit drugs such as testosterone can now be detected by isotope ratio analysis when referenced against endogenous steroids.

Researchers have used isotope ratio measurements to study animal migrations, for example, by analyzing wing membranes from migratory butterflies. The isotopic composition of the wing membrane does not change after they are formed at the breeding site and provides a good marker of geographical origin.

Stable isotopes are now increasingly used for the control of the origin or authenticity of food products. For example, the  $^{13}\text{C}$  content of sugars indicates the photosynthetic metabolism of the plant that synthesized them; the deuterium content is more characteristic of secondary metabolism and of environmental factors. SNIF-NMR has been used in conjunction with stable carbon IRMS to detect sugars that have been added to monofloral honey from C4 plants (cane or corn). A comparison of the  $^{13}\text{C}$  stable isotope composition of sparkling wines and the  $\text{CO}_2$  bubbles produced by secondary fermentation from a number of countries has also been used to estimate the addition of sugar. The determination of Sr isotope ratio  $^{87}\text{Sr}/^{86}\text{Sr}$  in table and fortified wines has provided a fingerprint of wine origin for regions of France and Portugal. Despite the high complexity of seized heroin samples, determinations of the  $^{13}\text{C}/^{12}\text{C}$  and  $^{15}\text{N}/^{14}\text{N}$  ratios by IRMS provide information for the determination of their origin.

Different chemical processes lead to isotopic fractionation in the generation of petroleum. IRMS analysis can, in many cases, distinguish petroleum derived from different geographical locations that are otherwise identical. The carbon and hydrogen isotopic composition of individual compounds in crude oil or refined products provide an isotopic fingerprint for use in correlation studies and do not appear to undergo significant isotopic fractionation as a result of weathering.

The advent of laser ablation MC-ICP-MS technology allows the rapid *in situ* determination of the stable isotope ratios of heavy metals commonly found in sulfide ore deposits (e.g., Cu, Zn, Fe, Sb, Ag) providing important information on the source, transport, and depositional mechanisms of these metals.  $^{206}\text{Pb}$ ,  $^{207}\text{Pb}$ , and  $^{208}\text{Pb}$  are formed as the end product of radioactive decay and the isotopic variability of lead results because the elements from which the isotopes form were not evenly distributed in ore bodies. Hence, the analysis of stable lead isotopes in annually laminated lake-sediments is a useful method to study lead pollution history as the relative contribution of pollution and natural lead in sediment samples can be calculated. The analysis of lead isotopes by SIMS has also been used to identify the geographical origin of bullets.

The thermodynamic properties of isotopic substances are well described and oxygen isotopes are known to be sensitive tracers of climatic change, as  $\text{H}_2^{16}\text{O}$  is more readily evaporated from the ocean than its  $\text{H}_2^{18}\text{O}$  counterpart. During periods of cool climate or mini 'ice-ages' the easily evaporated  $\text{H}_2^{16}\text{O}$  is trapped on land in the form of ice or other

reservoirs changing the  $^{18}\text{O}/^{16}\text{O}$  ratio of the ocean. This record is preserved in ice cores and in the calcium carbonate shells of fossil plankton and has been used since the 1940s for the reconstruction of past climates by oxygen isotope paleothermometry.

*See also:* **Mass Spectrometry:** Stable Isotope Ratio. **Quality Assurance:** Reference Materials; Production of Reference Materials.

### Further Reading

- Coleman DC and Fry B (eds.) (1991) *Carbon Isotope Techniques*. San Diego: Academic Press.
- Knowles R and Blackburn TH (eds.) (1993) *Nitrogen Isotope Techniques*. San Diego: Academic Press.
- Platzner IT (1997) *Modern Isotope Ratio Mass Spectrometry*. Wiley.
- Rundel PW, Ehleringer JR, and Nagy KA (eds.) (1989) *Stable Isotopes in Ecology Research*. New York: Springer.

# KINETIC METHODS

## Contents

### Principles and Instrumentation

#### Noncatalytic Techniques

#### Catalytic Techniques

## Principles and Instrumentation

**D Pérez-Bendito**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 2416–2424,  
© 1995, Elsevier Ltd., with revisions made by G Werner.

## Introduction

There are many areas in analytical chemistry in which kinetics play a significant role. This is true not only for reaction chemistry in solution, in separation processes, and diffusion controlled methods, but also for physical excitation. All analytical methods where measurements are influenced by the occurrence of a transient (kinetic) process could therefore strictly be considered a kinetic method.

In a narrower sense, however, the term ‘kinetic method’ is confined to methods based on direct or indirect measurements of the rate of a chemical reaction, which ought to be called ‘reaction-rate methods’. This view is followed in this article.

In the next section, the general principles of the analytical use of reaction-rate methods are described; in the subsequent sections the application of non-catalytic and catalytic techniques are treated. Several methods, instruments, and techniques for which separate entries can be found in this encyclopedia are only briefly mentioned, e.g., enzymatic catalysis, chemiluminescence, sensors, data processing.

## General Considerations

Signal measurements made under dynamic conditions on systems approaching equilibrium compete in

efficiency with static and equilibrium measurements. Inasmuch as not all reactions meet the requirements imposed by static measurements, the choice is usually dictated by the specific problem addressed since there are no instrumental limitations for avoiding kinetic methodologies. Thus, kinetic determinations based on catalytic (whether enzymatic or nonenzymatic) and differential reaction-rate methods are by now well established.

The development of kinetic methods can essentially be ascribed to (1) the increasingly frequent need to quantify minute amounts (occasionally of a few nanograms or even less) of substances contained in high-purity materials or samples of environmental or biological interest, (2) the increasing knowledge available on reaction mechanisms, and (3) recent breakthroughs in instrumentation, particularly with regards to automation and computerization capabilities.

Kinetic methods have been classified according to a number of criteria. One classification distinguishes between catalytic and noncatalytic methods (see **Table 1**). The former are further divided according to the type of reaction involved, while the latter are categorized according to whether they are used to determine a single species or several components in mixtures (differential reaction-rate methods)

**Table 1** General classification of kinetic methods of analysis

### *Catalytic*

Nonenzymatic (homogeneous)  
Enzymatic (homogeneous and heterogeneous involving immobilized enzymes)  
Electrochemical (heterogeneous)

### *Noncatalytic*

For determination of a single species  
For determination of multicomponents (differential reaction-rate methods)



although this occasionally may deal with catalyst mixtures. Other possible classifications are dealt with in the pertinent sections.

The principles behind kinetic methods are described below on the basis of uncatalyzed reactions in homogeneous solutions. The rate at which a given chemical reaction develops depends on several factors including temperature, reactant concentrations, the presence or absence of catalysts, activators, and inhibitors, and dielectric constant or ionic strength. Most of the reactions employed in kinetic analysis are influenced by temperature, which usually accelerates reaction development. Hence, a thermostating device is typically needed for kinetic applications.

Kinetic methods are usually associated with two types of reactions, namely slow reactions (those with half-lives of 10 s or longer) and fast reactions (those reaching half-completion in less than 10 s), which dictate both the sample-reagent mixing procedure and the instrumentation to be used.

## Theoretical Background

### Reaction Rate and Kinetic Equations

The determination of a given species using a kinetic method based on direct or indirect measurements of its reaction rate entails monitoring changes in the reactant or product concentration. The reaction rate is defined as the number of moles of substance that is consumed or formed per unit volume per unit time. Thus, for a straightforward reaction of the type



the rate at time  $t$  will be given by

$$\text{Rate} = d[P]/dt = -d[A]/dt = -d[B]/dt \quad [1]$$

where the derivatives of  $[A]$  and  $[B]$  are negative because both species disappear as the reaction develops. The reaction rate is proportional to the concentrations of all the species that take part in the reaction and the kinetic equation for the above reaction can be written as

$$\text{Rate} = d[P]/dt = k[A][B] \quad [2]$$

where  $k$  is the so-called rate constant and denotes the reaction rate per unit concentration of reactants. The sum of exponents of  $[A]$  and  $[B]$  in eqn [2] is known as reaction order; thus reaction [1] is second order. Also, the individual exponents of  $[A]$  and  $[B]$  in eqn [2] denote their respective partial reaction orders.

If one of the reactants (e.g.,  $B$ ) is present in a large excess (over 50-fold) with respect to the other, then

its concentration changes will be negligible relative to the other. In such a case,  $[B]$  can be included in constant  $k$  in eqn [2] and the reaction assumed to be pseudo-first order in  $A$  or pseudo-zero order in  $B$ , i.e.,

$$\text{Rate} = d[P]/dt = k_A[A] \quad [3]$$

where  $k_A$  is the pseudo-first-order rate constant.

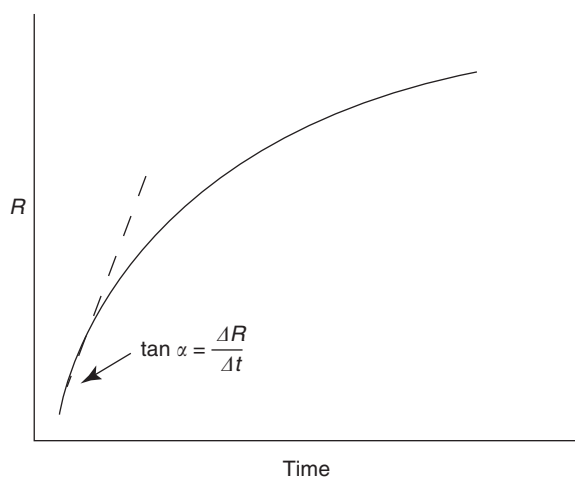
In applying most kinetic methods, experimental parameters are usually set so as to ensure pseudo-first-order reaction conditions, which is why only this type of reaction is dealt with here for simplicity.

The concentrations in eqn [1] can be replaced with any measurable quantity  $R$  provided it is directly proportional to concentration. Temporal changes in the reactant or product concentrations can be monitored physically or chemically. Physicochemical techniques (e.g., those based on absorbance, potential, temperature, luminescence, and conductivity measurements) are more commonly used for this purpose. Figure 1 shows the variation of a measured property,  $R$  (a signal), as a function of time. The reaction rate is given by the slope of the rising exponential curve at each point.

Most kinetic methods are applied to the initial portion of the curve, i.e., when the reaction has only developed by 1–3%. Such a portion is usually linear and its slope proportional to the concentration of the measured species (initial-rate methods).

### Determination of Partial Reaction Orders and Rate Constants

Reaction orders and the rate constant are typically determined experimentally under preset, controlled conditions. Establishing the kinetic or rate equation



**Figure 1** Typical kinetic curve showing the initial straight portion.

of a chemical system entails determining the partial reaction orders in the variables that influence the system. This can be done in two ways, depending on whether the rate equation is to be expressed in integral or differential form. As a rule, differential methods, which involve measuring the initial reaction rate, are usually employed for determining partial orders, whereas integral methods are preferentially used for calculating rate constants.

Differential methods rely on the assumption that, for a partial reaction order  $n$ , the initial rate is related to the concentration by

$$\text{Rate} = \tan \alpha = -d[A]/dt = k_A[A]^n \quad [4]$$

where A is the species whose partial reaction order is to be determined and  $k_A$  its pseudo- $n$ th-order rate constant.

Equation [4] can be expressed in logarithmic form as

$$\log(\tan \alpha) = \log k_A + n \log[A] \quad [5]$$

which is the equation of a straight line of slope  $n$  and intercept  $\log k_A$ . Therefore, one can obtain the partial order in A and its reaction rate from the slope and intercept, respectively, of a  $\log(\tan \alpha)$  versus  $\log[A]$  plot.

The determination of the rate constant by use of integration methods involves plotting the integral form of the kinetic equation for an assumed reaction order. Thus, the integral rate equation for a hypothetical first-order reaction is

$$\ln[A]_t = \ln[A]_0 - k_A t \quad [6]$$

where  $[A]_t$  is the concentration of species A at reaction time  $t$ ,  $[A]_0$  the initial concentration of this species, and  $k_A$  its rate constant which coincides with the slope of this plot. If, for instance, the reaction concerned is monitored photometrically via the product P, then

$$\ln(D_\infty - D_t) = \ln D_\infty - k_A t \quad [7]$$

since  $[A]_t = [P]_\infty - [P]_t$  and  $[A]_0 = [P]_\infty$ , i.e.,  $[P]$  and  $D$  (the absorbance) are proportional.

Once the influence of each variable and the concentration on the reaction rate has been determined, the kinetic equation can be formulated according to the general form

$$\nu = \text{rate} = d[P]/dt = k[A]^a[B]^b[C]^c \quad [8]$$

where P is the reaction product through which the reaction is monitored and  $k$  the rate constant, and superscripts denote the partial reaction orders in the species to which they refer.

The partial order in a given reactant will be unity if a plot of initial rate versus reactant concentration is a straight line; in fact, at  $n = 1$ , eqn [4] is representative of a straight line of zero intercept.

## Kinetic Determination of a Single Species

Inasmuch as most chemical kinetic systems involve bimolecular reactions and orders above two are a rare and analytically unusable, kinetic methods for the determination of a single species can be classified according to the kinetic order of the bimolecular reaction  $A + B \rightarrow P$  into (1) differential or pseudo-zero-order methods, and (2) integral, first-order, or pseudo-first-order methods.

### Differential Methods

Differential kinetic methods for the determination of a single species rely on measurements made at the start of reaction, i.e. when changes in the reactant or product concentrations are still negligible and hence the reaction rate is not influenced by the concentration of either reactant, so eqn [2] can be simplified to a pseudo zero-order expression

$$\text{Initial rate} = \nu_0 = d[P]/dt \approx k[A]_0[B]_0 \approx \text{constant} \quad [9]$$

which is the starting point for application of the initial-rate method. Since  $[B]_0$  is constant, a plot of the initial rate against  $[A]_0$  must be a straight line that can be used as a calibration plot for determining  $[A]_0$  from eqn [9] in incremental form, i.e.,

$$\nu_0 = \Delta[P]/\Delta t = k_A[A]_0 \quad [10]$$

Both the fixed-time and the variable-time differential methods are based on eqn [10] by making  $\Delta t$  or  $\Delta[P]$  constant, respectively. Therefore,  $[A]_0$  will be linearly related to  $\Delta t$  (or  $\Delta[P]$ ), which allows one to construct the pertinent calibration graph.

### Integral Methods

Determinations of single species by integral methods are based on the use of the integrated pseudo-first-order eqn [6] or [7], both of which allow  $[A]_0$  to be calculated from the intercept of the curve obtained by plotting  $\ln[A]_t$  as a function of time.

Application of the fixed-time and variable-time integral methods relies on eqn [6] in the form

$$\ln[A]_1/[A]_2 = k_A \Delta t \quad [11]$$

and requires that  $t_1 = 0$  (and hence  $[A]_t = [A]_0$ ) and keeping  $\Delta t$  and  $\Delta[A]$  constant, respectively.

## Kinetic Determination of Multicomponents

Differential reaction-rate methods are based on the different rate at which two or more species react with a common reagent and allow the determination of several components without the need for a prior separation.

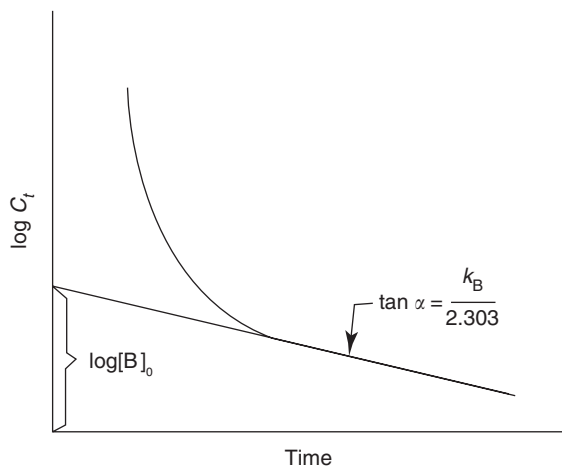
Let A and B be two substances in a mixture that react with a reagent R to yield two products P and P', which, though different in nature, are essentially similar as regards the analytical property to be measured. If the reaction rates of the two substances are different and independent of each other, and correspond to first-order kinetics in each mixture component (i.e.,  $[R]_0 > [A]_0$  and  $[R]_0 > [B]_0$ ), then the sum of the concentrations of A and B at time  $t$  will be given by

$$\begin{aligned} C_t &= [A]_t + [B]_t = [P]_\infty - [P]_t \\ &= [A]_0 \exp(-k_A t) + [B]_0 \exp(-k_B t) \end{aligned} \quad [12]$$

where  $k_A$  and  $k_B$  are the pseudo-first-order reaction rate constants for A and B, respectively.

There are several types of differential kinetic methods and the most frequently used of which are the logarithmic-extrapolation and the proportional-equation methods, both of which are based on the application of eqn [12].

The logarithmic extrapolation method involves taking logarithms in eqn [12] and plotting  $\log C$  or  $\log([P]_\infty - [P]_t)$  as a function of time (Figure 2). The result is a curve unless  $k_A = k_B$ , in which case a straight line is obtained. If species A disappears at a higher rate than does B (i.e.,  $k_A > k_B$ ), then  $[A]_t \rightarrow 0$  and the curve eventually becomes a straight line



**Figure 2** Application of the logarithmic-extrapolation method to first-order reactions.

(Figure 2); so

$$\log C_t = \log([P]_\infty - [P]_t) = \log[B]_0 - k_B t / 2.303 \quad [13]$$

The intercept of the  $\log C_t$  or  $\log([P]_\infty - [P]_t)$  versus  $t$  plot allows the initial concentration of B,  $[B]_0$ , to be calculated, whereas that of the other component,  $[A]_0$ , can be obtained by difference once  $[A]_0 + [B]_0$  has been obtained from  $[P]$ . If a photometric technique is used for measurements, eqn [13] can be alternatively written as

$$\log(D_\infty - D_t) = \log \varepsilon_B t [B]_0 - k_B t / 2.303 \quad [14]$$

where  $D_\infty$  and  $D_t$  are the absorbances measured after the reaction has developed to completion and after a time  $t$  close to the start of the linear segment of the curve, respectively, and  $\varepsilon_B$  denotes the molar absorptivity of B, which should be known beforehand. The path length of the photometric cell is assumed to be 1 cm.

This method requires no prior knowledge of the rate constants of A or B; however, it does require at least 99% of the more reactive species to be consumed in the process if reliable results are to be obtained, and the total initial concentrations of A and B to be accurately known.

Application of the proportional-equation method to a binary mixture entails measuring  $C_t$  at two reaction times and formulating two equations similar to [12], the resolution of which provides the initial concentration of the two components as a function of their respective rate constants, which should be known in advance. Not only time, but any other experimental variable such as the temperature or one of the physicochemical properties of the reactants are usable for this purpose. The two equations in question will be of the form

$$f_1 = k_{A,1} [A]_0 k_{B,1} [B]_0 \quad [15]$$

$$f_2 = k_{A,2} [A]_0 k_{B,2} [B]_0 \quad [16]$$

where  $f$  denotes a measurable parameter determined under two different sets of experimental conditions, and  $k_{A,1}$ ,  $k_{A,2}$ ,  $k_{B,1}$ , and  $k_{B,2}$  are empirical or proportionality constants obtained separately for components A and B under different conditions.

For these methods to be applicable in practice, the rate-constant ratio,  $k_A/k_B$ , should generally be at least 3–4. However, novel mathematical approaches such as the multipoint curve-fitting method, which takes full advantage of the information provided by the entire kinetic curve (e.g., the Kalman filter), are applicable to ratios of only 1.5–2.

## Instrumentation

The present status of kinetic methods, which are currently competitive with static (equilibrium) methods in many respects, has been chiefly reached thanks to recent breakthroughs in instrumentation that have enabled their automated implementation. Electronics and computers have played decisive roles in this context as they have solved the typical problems encountered in measuring reaction time accurately – particularly in relation to fast processes – by allowing automated acquisition of data throughout.

Kinetic methods call for no measurements of absolute values of the parameter typically used to monitor reactions (absorbance, fluorescence intensity, potential), but rather for their temporal variations; as a result, kinetic measurements are free from the effects of factors that introduce errors in absolute values (e.g., turbidity, the liquid–liquid junction potential, and the presence of other absorbing or fluorescent substances provided they do not take part in the reaction of interest or alter the parameter response). However, strict timing and temperature control (to within 0.01–0.1°C) are essential to kinetic methods, which thus require modern, powerful instrumentation.

The choice of a given method for measuring the rate of a reaction is dictated by its half-life. Thus, the instrumentation required to monitor slow reactions is typically simpler than that needed for fast reactions.

### Primary Components of Kinetic Instrumentation

Kinetic methods use variously complex instruments from conventional straightforward systems such as spectrophotometers, spectrofluorimeters, or potentiometers to highly sophisticated setups, e.g., fully automatic instruments capable of collecting samples, transferring them to the detector, and transducing and processing the results obtained.

The process involved in measuring the rate of a reaction comprises the following stages:

(1) Preparation, measurement, transport, and mixing of reactants. These preliminary operations are common to other analytical methods and are a key step in kinetic methods as regards measurement of the parameter by which the process is monitored.

(2) Signal monitoring and transducing at a constant temperature, both of which are effected by the measuring instrument (e.g., photometer, potentiometer), which follows the course of the reaction concerned by measuring changes in a property of one reactant or product.

(3) Timed acquisition of data (viz., collection of signals at different times), which are processed

simultaneously or sequentially, either by manual computation or with the aid of a computer.

Fully automated systems execute all three stages above, whereas partly automated instruments usually perform stages (2) and (3) only, and process data in a sequential rather than simultaneous manner.

Each of the above stages is related to one of the basic elements of the instrumentation typically used in kinetic methods. In fact, the first stage, i.e., mixing of sample and reagents, is done differently depending on whether the external experimental conditions will remain unchanged (closed systems) or not (open systems).

### Temperature Control

Strict temperature control is compulsory at stages (1) and (2), whichever type of system is used, because fluctuations in this parameter detract from the accuracy and precision of the results. A change in the temperature by as little as 1°C may result in a variation in the reaction rate by up to 10% in some cases.

There are basically two ways of controlling the temperature: by circulating a liquid kept at a constant temperature (e.g., from a thermostatic bath) and by electrical heating (or cooling) via the Peltier effect.

### Mixing of Reactants

The manner in which reactants are mixed to start the analytical reaction depends on the type of system concerned.

**Closed and open systems** While closed systems are unsuitable for fast reactions, open systems are especially suited, so much so that they have significantly eased the use of kinetic procedures.

The sample must be thoroughly mixed with the reagent(s) at a constant temperature before it is transferred to the detection system for measurement. This can be done manually, which is usually the case with closed systems, or automatically. Automated mixing of sample and reagents calls for special mechanical devices when fast reactions, which require the use of open systems as noted earlier, are to be addressed. On the other hand, manual mixing can be accomplished by simply homogenizing the reactants in the reaction vessel, which can be a spectrophotometric cuvette or an electrode cell. The instant at which the last reactant is added is taken as the reaction start time ( $t = 0$ ) and must therefore be determined highly accurately. Alternatively, the reactants can be incorporated into the measuring system automatically in order to increase the reproducibility and throughput of closed or batch systems.

The open systems more commonly used can be classified into continuous or flow systems and discrete or batch systems (Table 2). The most popular and widely employed methods in connection with such systems for kinetic-based determinations in routine analyses are stopped-flow, stat, and continuous-addition-of-reagent methods.

**Stopped-flow methods** This type of method essentially mixes a sample solution and the reagent(s) by means of two driving syringes that are actuated manually or automatically (by a pneumatic device in the latter case) into a flow cell or mixing chamber that can also act as the observation cell. The flow is stopped abruptly by using a third syringe and the analytical signal is then recorded as a function of time. This type of method is mandatory in dealing with fast reactions with half-lives of a few milliseconds to several seconds. Its performance relies heavily on the dead time, the length of which is determined by the mixing efficiency (viz., the mixing time) and the transport and stop times. For practical

purposes, the dead time of the instrument used should be smaller than the half-life of the reaction to be addressed by approximately two orders of magnitude. Some European and American manufacturers market specific instruments and accessories including stopped-flow mixing modules that can be fitted to commercially available spectrophotometers and spectrofluorimeters for implementation of the stopped-flow technique.

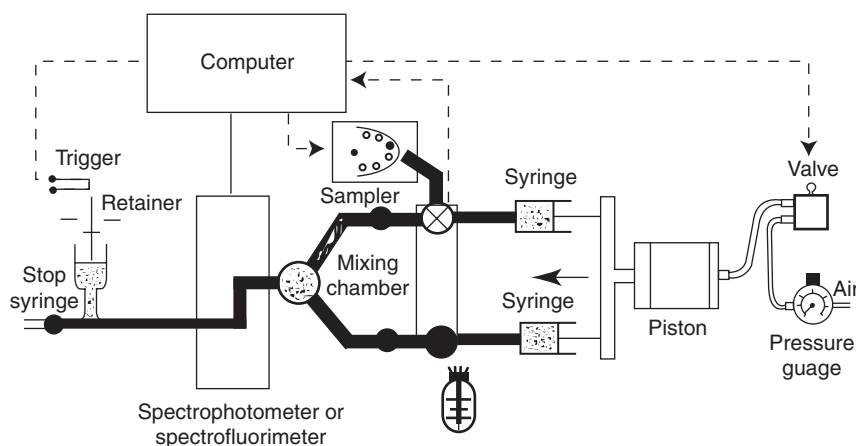
Figure 3 shows a schematic diagram of a stopped-flow spectrophotometer (or spectrofluorimeter) coupled to an automatic analyzer controlled by a computer that also handles data acquisition and processing.?

**Stat methods** Stat methods, which are suitable for slow reactions, involve addition of a reactant to the reaction vessel at a given rate such that a characteristic property of the monitored reaction is maintained constant. In practice, a small amount of one of the reaction ingredients is added until a given value of the monitored parameter (pH, absorbance, luminescence, etc.) is reached. Any deviation from this state as a result of the reaction developing further is immediately compensated for by automatically adding an extra amount of the ingredient in question. The rate of addition is used to measure the reaction rate. This type of method has so far been chiefly used quantitatively with catalyzed (including enzymatic) reactions.

**Continuous-addition-of-reagent methods** In continuous-addition-of-reagent (CAR) methods, a reagent is added at a constant rate to the species of interest and the kinetic graph (absorbance or fluorescence as a function of time) is recorded simultaneously with the aid of an automatic, computer-controlled system. A CAR system typically consists of (1) an addition unit composed of an autoburette, a fan stirrer, and a thermostatted vessel as reaction cell, (2) a photometric

**Table 2** Classification of open systems

<i>Continuous</i>
Continuous flow
Accelerated flow
Pulsed flow
Stopped flow
<i>Discrete</i>
Based on slow reactions
Stat methods
Steady-state methods
Based on fast reactions
Continuous-addition-of-reagent methods
Relaxation methods
Pulse methods



**Figure 3** Diagram of a stopped-flow spectrophotometer or spectrofluorimeter. (Reproduced with permission from Pérez-Bendito D and Silva M (1988) *Kinetic Methods in Analytical Chemistry*. Chichester: Ellis Horwood.)



(or fluorimetric) detection unit including an immersion probe for optical signal measurements, and (3) a computer. **Figure 4** shows a block diagram of the instrumental components.

The CAR technique is a major alternative to the stopped-flow technique in tackling direct rate measurements on reactions with half-lives of a few milliseconds as it requires no sophisticated instrumentation and is specially suited to routine analyses. In addition, it allows mixtures of two or more analytes to be resolved. For a binary mixture, the response versus time graph typically obtained shows two consecutive linear segments of different slope from which each reaction rate can be determined and related to the concentration of each component.

### Detection Systems

Although the course of a reaction can be monitored by chemical or even visual means, most kinetic methods rely on instruments for this purpose, optical (photometric and fluorimetric) and electroanalytical devices being by far the most common choices. In this context, it is worth emphasizing the ability of stopped-flow mixing methodology to boost the performance of the chemiluminescence reaction to which it is specially suited on account of the fast transient nature of chemiluminescence reactions.

In selecting a given instrument for a specific purpose, the user should take into account the nature of the analyte, the most suitable species for measuring its concentration, and the type of reaction involved and its half-life. Also, the response time should be short enough for the signal to be as close as possible in time to the event that gives rise to it, especially with fast reactions or initial-rate methods.

**Ultraviolet–visible molecular absorption spectrometers** This type of instrument owes much of its popularity to the availability of a large number of

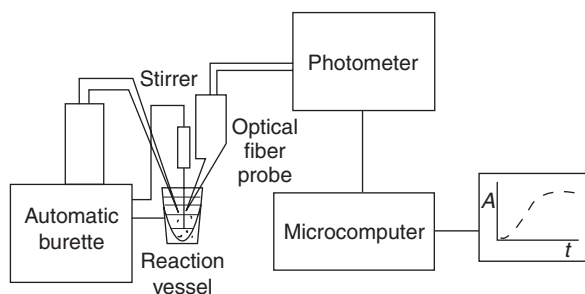
reactions involving substances that absorb in the near-ultraviolet or visible region. Spectrophotometric sensing, unlike fluorimetric detection, allows the simultaneous recording of absorbance readings from the unknown sample and a blank by using a dual-beam spectrophotometer which is of special significance to monitoring development of the catalyzed reaction against the uncatalyzed reaction. However, this procedure is inadvisable for fast reactions as it may provide delayed responses that would detract from the precision of measurements.

Rapid scanning spectrometry offers an appealing potential for kinetic methods of analysis. The associated instrumentation is based on various operational principles; thus, the vidicon image detector bears some relationship to television, while photodiode array detectors consist of a very large number (a few hundreds or thousands) of individual detectors, each of which deals with a narrow band of the transmitted spectrum. Diode array systems allow the simultaneous kinetic determination of two or more dissolved species giving rise to as many reactions with the same or different reagents. Provided the species in question absorb at different wavelengths, direct resolution of the mixture only requires simultaneously measuring the initial rates at the corresponding maximum absorption wavelengths using spectrophotometers equipped with diode array detectors. **Figure 5** shows the three-dimensional plot obtained for the products afforded in the reactions of hydrazine and phenylhydrazine with p-dimethylaminobenzaldehyde (DAB) and their respective kinetic curves, which were processed simultaneously by using a computerized initial-rate method. This methodological principle has also been used in molecular fluorescence spectroscopy and kinetic fluorimetric determinations.

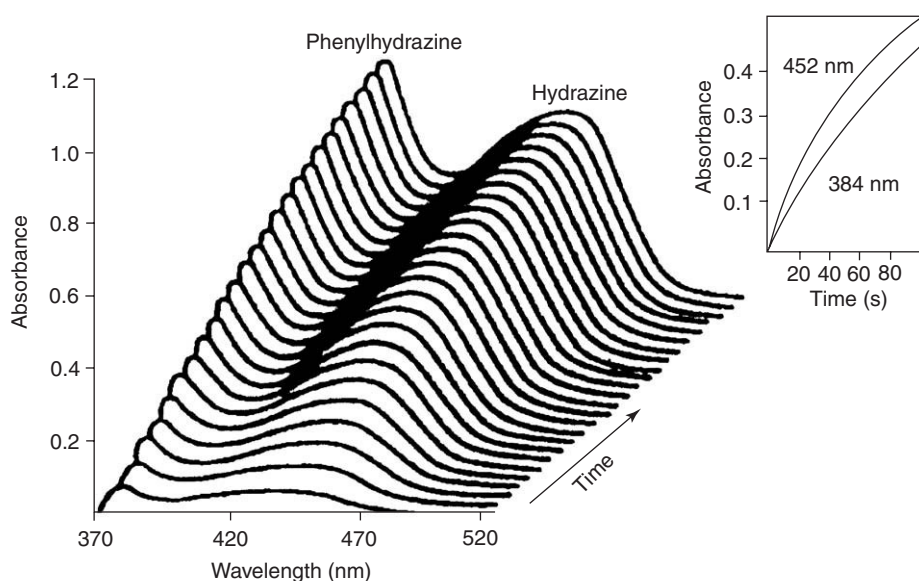
**Potentiometric and amperometric detectors** Potentiometric sensing usually entails the use of the variable-time approach, i.e., measuring the time elapsed until the required change between two pre-set potentials has occurred.

Selective electrodes are of special use in potentiometric techniques as applied to kinetic analysis on account of their usually high sensitivity to concentration changes. The glass (pH) electrode is very frequently used to determine enzymes and their substrates, which take part in reactions that typically involve release or uptake of hydrogen ions. The ammonia-selective membrane electrode has also been used for enzymatic analyses.

Such amperometric electrodes as the oxygen electrode are also frequently employed in enzymatic analyses involving consumption or release of oxygen



**Figure 4** Experimental setup for implementation of the continuous-addition-of-reagent technique. (Reproduced with permission from Márquez M, Silva M, and Pérez-Bendito D (1990) Kinetic determination of sulfonamides at the millimolar level by continuous addition of reagent technique. *Analytica Chimica Acta* 237: 353–359; © Elsevier.)



**Figure 5** Temporal evolution of the absorption spectrum of the phenylhydrazine(A)–hydrazine(B)–DAB system. (A)  $A_{\max} = 452 \text{ nm}$ ; (B)  $A_{\max} = 384 \text{ nm}$ ;  $t = 30 \text{ s}$ . (Reproduced with permission from Pérez-Bendito D (1990) Approaches to differential reaction-rate methods. *Plenary Lecture: Analyst* 115: 689–697.)

or hydrogen peroxide. The widespread usage of these electrodes is partly a result of advances in enzyme immobilization techniques. In their simplest form, these electrodes are composed of a membrane that accommodates an artificial enzyme fixed onto the transducer. The substrate diffuses across the thin catalyst layer and yields an electroactive substance that is sensed potentiometrically or amperometrically. Potentiometric sensing is more economical and easier to implement as regards the electrode preparation, albeit rather slow compared to amperometric detection, which also surpasses the former in linear range. Differential potentiometry and biamperometry are also commonplace in kinetic analyses.

### Data Acquisition and Processing

This final stage of the analytical process is a key to obtaining acceptable results, particularly in dealing with fast reactions and in differential kinetic analysis. Generated signal/time data pairs are converted into reaction rates or, directly, into analyte concentrations. How this final operation is performed depends on the nature of the applied method.

Kinetic methods for the determination of a single species can be classified into various categories according to the type of data processing system used (see Table 3).

Direct-computation methods, on which most kinetic analyzers rely, use one or more signal readouts to calculate generally the sought reaction rate.

**Table 3** Classification of kinetic methods according to data acquisition and processing

<i>Direct computation methods</i>
One-point methods
Two-point methods
Multipoint (regression) methods
<i>Curve-fitting methods</i>
Linear responses
Nonlinear responses

Depending on the number of measurements used, these methods can be further divided into one-, two-, and multipoint methods. Multipoint regression methods employ a mathematical algorithm to ensure the best possible fit between three or more data points by using a given mathematical function that coincides with the equation of a straight line for calculation of the reaction rate and hence should be fitted to the linear portion of the kinetic curve.

Until recently, a vast majority of kinetic procedures relied on direct computation methods. However, there is a current trend to increasing usage of curve-fitting methods, which compute parameter values that represent the best fit of some mathematical models to signal/time data in order to obtain the analyte concentration, the rate constant, and even the partial reaction orders. Although least-squares methods have been used most frequently for this purpose, other curve-fitting approaches such as the Kalman filter, a recursive algorithm that allows data processing in real time, are gradually gaining ground in this context.

Data acquisition by a computer-controlled instrument requires the two components to be appropriately interfaced. Analog interfaces convert a voltage or any other analog output into a binary (digital) form that can be managed by the computer. The wide availability and present affordability of computers makes this formerly complicated operation quite easy.

Recent advances in instrumentation and data acquisition and processing have put kinetic methods of analysis in prominence by virtue of their simplicity, rapidity, and precision, in addition to their ready automation for use in routine analyses.

See also: **Fluorescence:** Overview. **Spectrometry:** Overview.

## Further Reading

- Crouch SR, Cullen TF, Scheeline A, and Kirkor ES (1998) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 70: 53R–106R.
- Crouch SR, Scheeline A, and Kirkor ES (2000) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 72: 53R–70R.

- Mottola HA (1988) *Kinetic Aspects of Analytical Chemistry*. New York: Wiley.
- Mottola HA and Pérez-Bendito D (1992) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 64: 407R–428R.
- Mottola HA, Pérez-Bendito D, and Mark HB, Jr. (1990) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 62: 441R–461R.
- Mottola HA and Pérez-Bendito D (1992) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 64: 407R–428R.
- Mottola HA and Pérez-Bendito D (1994) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 66: 131R–162R.
- Mottola HA and Pérez-Bendito D (1996) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 68: 257R–289R.
- Pardue HL (1989) Kinetic aspects of analytical chemistry. *Analytica Chimica Acta* 219: 69–107.
- Pérez-Bendito D, Silva M, and Gómez-Hens A (1989) Automated kinetic-based determinations for routine analyses: recent developments. *Trends in Analytical Chemistry* 8: 302–308.

## Noncatalytic Techniques

**M Silva**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 2424–2437, © 1995, Elsevier Ltd., with revisions made by G Werner.

## Introduction

Noncatalytic reactions are less frequently used in kinetic-based determinations than are those involving a catalytic effect. However, recent advances in instrumentation mean that noncatalytic kinetic methods are powerful alternatives to equilibrium (nonkinetic) methods. This type of reaction is of especial relevance to the analysis of mixtures of closely related compounds, for which a number of differential reaction rate methods have been developed. Whether for individual or joint determinations of species, the main field of application of noncatalytic reactions is organic analysis, unlike catalytic reactions, where a metal ion usually acts as the catalyst; this has also contributed to their current wide acceptance.

## Principles of Methods

Analytical methods based on noncatalytic reactions are useful for both the kinetic determination of a

single species and the simultaneous kinetic determination of several species in a mixture with no prior separation. The kinetic determination of a single species in a mixture is a special instance of the latter. These concepts are classified in Table 1 and discussed in detail below.

### Determination of a Single Species: Classical Methods

Consider a chemical species A that reacts with a reagent B to yield product P:



The determination of species A can involve two limiting situations: (1) if B is in large excess (over 50-fold) relative to the analyte (A), then the reaction will be pseudo-first order in A; (2) if the concentration of B is similar to or less than 50 times that of A, the reaction will be second order.

**Pseudo-first-order reactions** The rate of a reaction of pseudo-first order in the analyte, A, is given by

$$\text{rate} = -d[A]/dt = d[P]/dt = k[A][B]_0 = k'[A] \quad [1]$$

where  $k' = k[B]_0$ . This expression is normally the basis for the kinetic determination of A. For initial-rate

Data acquisition by a computer-controlled instrument requires the two components to be appropriately interfaced. Analog interfaces convert a voltage or any other analog output into a binary (digital) form that can be managed by the computer. The wide availability and present affordability of computers makes this formerly complicated operation quite easy.

Recent advances in instrumentation and data acquisition and processing have put kinetic methods of analysis in prominence by virtue of their simplicity, rapidity, and precision, in addition to their ready automation for use in routine analyses.

See also: **Fluorescence:** Overview. **Spectrometry:** Overview.

## Further Reading

- Crouch SR, Cullen TF, Scheeline A, and Kirkor ES (1998) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 70: 53R–106R.
- Crouch SR, Scheeline A, and Kirkor ES (2000) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 72: 53R–70R.

Mottola HA (1988) *Kinetic Aspects of Analytical Chemistry*. New York: Wiley.

Mottola HA and Pérez-Bendito D (1992) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 64: 407R–428R.

Mottola HA, Pérez-Bendito D, and Mark HB, Jr. (1990) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 62: 441R–461R.

Mottola HA and Pérez-Bendito D (1992) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 64: 407R–428R.

Mottola HA and Pérez-Bendito D (1994) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 66: 131R–162R.

Mottola HA and Pérez-Bendito D (1996) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 68: 257R–289R.

Pardue HL (1989) Kinetic aspects of analytical chemistry. *Analytica Chimica Acta* 219: 69–107.

Pérez-Bendito D, Silva M, and Gómez-Hens A (1989) Automated kinetic-based determinations for routine analyses: recent developments. *Trends in Analytical Chemistry* 8: 302–308.

## Noncatalytic Techniques

**M Silva**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 2424–2437, © 1995, Elsevier Ltd., with revisions made by G Werner.

## Introduction

Noncatalytic reactions are less frequently used in kinetic-based determinations than are those involving a catalytic effect. However, recent advances in instrumentation mean that noncatalytic kinetic methods are powerful alternatives to equilibrium (nonkinetic) methods. This type of reaction is of especial relevance to the analysis of mixtures of closely related compounds, for which a number of differential reaction rate methods have been developed. Whether for individual or joint determinations of species, the main field of application of noncatalytic reactions is organic analysis, unlike catalytic reactions, where a metal ion usually acts as the catalyst; this has also contributed to their current wide acceptance.

## Principles of Methods

Analytical methods based on noncatalytic reactions are useful for both the kinetic determination of a

single species and the simultaneous kinetic determination of several species in a mixture with no prior separation. The kinetic determination of a single species in a mixture is a special instance of the latter. These concepts are classified in Table 1 and discussed in detail below.

### Determination of a Single Species: Classical Methods

Consider a chemical species A that reacts with a reagent B to yield product P:



The determination of species A can involve two limiting situations: (1) if B is in large excess (over 50-fold) relative to the analyte (A), then the reaction will be pseudo-first order in A; (2) if the concentration of B is similar to or less than 50 times that of A, the reaction will be second order.

**Pseudo-first-order reactions** The rate of a reaction of pseudo-first order in the analyte, A, is given by

$$\text{rate} = -d[A]/dt = d[P]/dt = k[A][B]_0 = k'[A] \quad [1]$$

where  $k' = k[B]_0$ . This expression is normally the basis for the kinetic determination of A. For initial-rate



**Table 1** Analytical methods used in kinetic determinations based on noncatalytic reactions

---

<i>Determination of a single species</i>
Pseudo-first-order reactions
Initial rate method
Fixed-time method
Variable-time method
Second-order reactions
Identical reactant concentrations
Unequal reactant concentrations
Multipoint methods
Curve-fitting methods
Predictive methods
Error-compensated methods
<i>Determination of a single species in a mixture</i>
<i>Simultaneous kinetic-based determinations</i>
Classical differential kinetic methods
Logarithmic-extrapolation method
Proportional-equation method
Multipoint methods
Curve-fitting methods
Kalman filter algorithm
Artificial neural networks
Multivariate calibration methods

---

methods, eqn [1] can be expressed in incremental form as

$$\text{rate} = -\Delta[A]/\Delta t = k[A]_0 \quad [2]$$

from which the concentration of A can be calculated readily (Figure 1A).

The mathematical treatments used by the fixed-time and variable-time methods rely on the exponential version of eqn [1], namely

$$\text{rate} = -d[A]/dt = k[A]_0 \exp(-kt) \quad [3]$$

Taking into account the fact that  $t_2 = \Delta t + t_1$ , integration of this equation over an interval  $t_2 - t_1$  yields

$$[A]_0 = -\Delta[A]/\exp(-kt_1)[1 - \exp(-k\Delta t)] \quad [4]$$

Application of the fixed-time method involves keeping  $t_1$  and  $\Delta t$  constant, and so, according to eqn [4],  $\Delta[A]$  and  $[A]_0$  will be linearly related (Figure 1B). Using the variable-time method (Figure 1C) involves keeping  $\Delta[A]$  constant. Since  $\Delta t$  and  $t_1$  depend on  $[A]_0$  according to eqn [4],  $[A]_0$  will never be linearly related to  $1/\Delta t$ . In this case the minimum error will be made when  $[A]_1$  is virtually identical with  $[A]_0$  and hence  $\exp(-kt_1) \rightarrow 1$ , and at short reaction times (small  $\Delta[A]$  and  $\Delta t \rightarrow 0$ ) eqn [4] can be linearized since  $(1 - \exp(-k\Delta t))$  can be simplified to  $k\Delta t$  by expanding the exponential function as a Maclaurin series. In summary, the fixed-time method

provides wider dynamic ranges than does the variable-time method for first-order or pseudo-first-order reactions (see Figures 1B and 1C).

**Second-order reactions** The rate of a second-order reaction is dependent on the concentrations of two reactants. The rate of reaction [I] is

$$\text{rate} = -d[A]/dt = -d[B]/dt = d[P]/dt = k[A][B] \quad [5]$$

If the concentrations of the two reactants are equal, then  $[A]_0 = [B]_0$ , and so eqn [5] simplifies to

$$\text{rate} = -d[A]/dt = k[A]^2 \quad [6]$$

Integration of this equation between times zero and  $t$  yields the following linear equation:

$$1/[A]_t = (1/[A]_0) + kt \quad [7]$$

from which the concentration of A can be calculated readily (Figure 2A).

For a second-order reaction in which the concentrations of the two reactants are different, integration of eqn [5] over an interval from  $t_1 = 0$  to  $t_2 = t$  leads to the general second-order integrated rate equation:

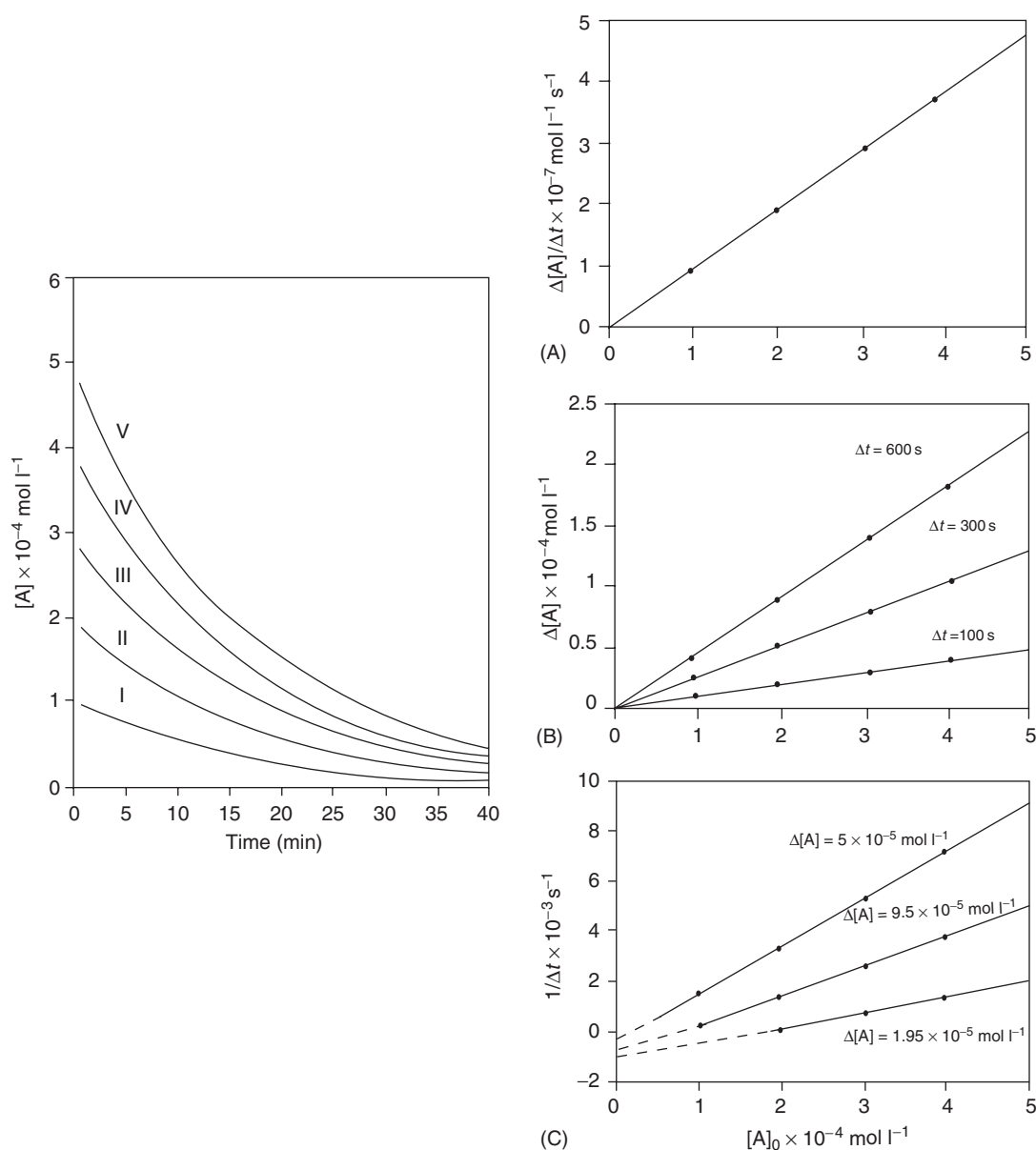
$$\ln([B]_t/[A]_t) = \ln([B]_0/[A]_0) + ([B]_0 - [A]_0)kt \quad [8]$$

A plot of  $\ln([B]_t/[A]_t)$  against time (Figure 2B) will thus provide an intercept from which the concentration of A can be calculated since  $[B]_0$  is known.

### Determination of a Single Species: Multipoint Methods

The high processing power of affordable computers has fostered the development of a variety of improved methods relying on efficient use of analytical information. These include curve-fitting methods, predictive methods, and error-compensated methods. Curve-fitting methods compute signals in order to obtain the best fit for a mathematical model describing the transient signal. A least-squares procedure is commonly used for this purpose. Predictive methods use a large number of data to obtain the best estimate for the signal change to be expected if the system is allowed to reach equilibrium. Error-compensated methods reduce the marked dependence of experimental variables in methods based on time-dependent signals. Such dependence can be minimized by using mathematical algorithms or simultaneously measuring the analytical signals and the variable(s) by which they can be affected in order to establish pertinent corrections.





**Figure 1** Kinetic profiles (left) and determination of a single species based on pseudo-first-order reactions. (A) Initial rate, (B) fixed-time, and (C) variable-time methods. Curves I–V in the kinetic profiles correspond to  $1\text{--}5 \times 10^{-4} \text{ mol l}^{-1}$ , respectively. Simulated data for  $k = 10^{-3} \text{ s}^{-1}$ .

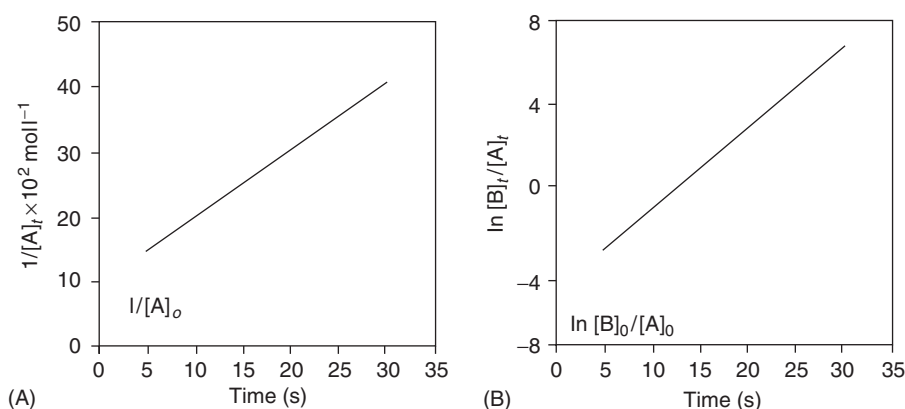
### Determination of a Single Species in a Mixture

For the case where different chemical species in a mixture have different rates of reaction with a reagent, it is possible to treat each species separately. That means that during a certain time interval only one of the species has to be considered because the other species have already reacted or react so slowly that they do not interfere with the rate measurements. The criteria for neglecting reactions of slower or faster reacting components are the ratio of their rate constants, the ratio of their concentrations, and the time interval of measurement.

Consider the determination of A in the presence of B. Their respective integrated first-order rate equations will be

$$\ln([A]_0/[A]_t) = k_A(t - t_0) \quad \ln([B]_0/[B]_t) = k_B(t - t_0) \quad [9]$$

If the extent of reaction of B over the interval from  $t_0$  to  $t$  is short enough, then  $[A]_0$  can be determined readily by simply measuring  $[A]_t$ . However, if B reacts to an appreciable extent, the degree of interference with the analyte reaction can be calculated from the ratio of the two above expressions.



**Figure 2** Determination of a single species based on second-order reactions. (A) Identical reactant concentrations; and (B) unequal reactant concentrations. Simulated data for  $[A]_0 = 10^{-3} \text{ mol l}^{-1}$ ,  $[B]_0 = 5.0 \times 10^{-3} \text{ mol l}^{-1}$ , and  $k = 10^2 \text{ mol l}^{-1} \text{ s}^{-1}$ .

Thus

$$\ln([A]_0/[A]_t)/\ln([B]_0/[B]_t) = k_A/k_B \quad [10]$$

This equation is time dependent and relates the extent of both reactions to their rate constants. If a maximum error of 1% is considered to be acceptable (i.e., if there is still 1% of A present when less than 1% of B has reacted), the  $k_A/k_B$  ratio in eqn [10] should be  $\approx 500$ . (This follows from  $\log 100/\log 1.01 = 463$ .)

The error made in calculating  $[A]_0$  from a given rate constant ratio at a given extent of reaction will be

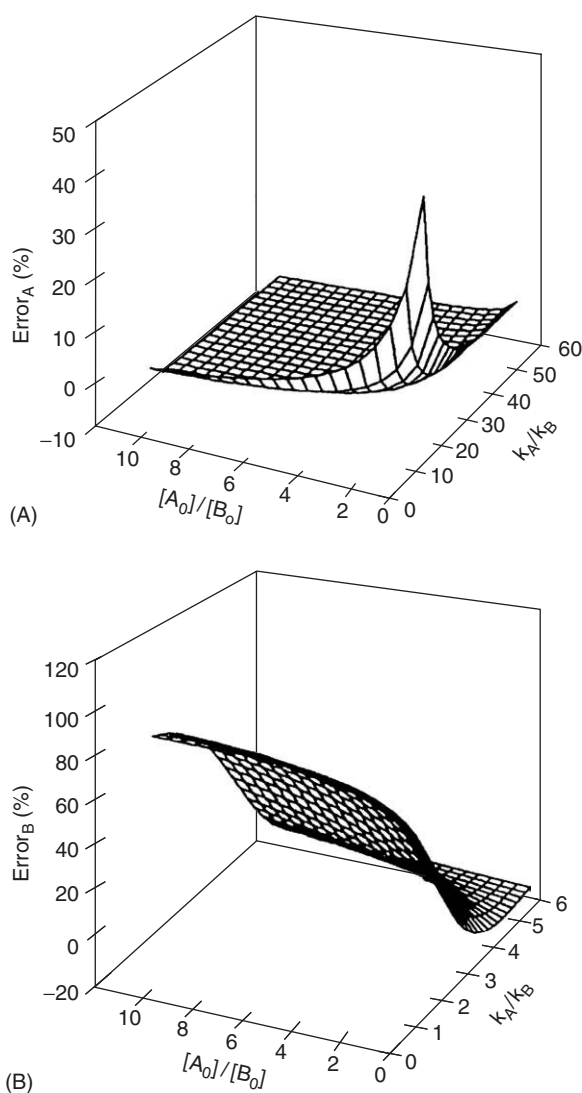
$$\begin{aligned} \%e_A &= \frac{[P_B]_t}{([P_A]_t + [P_B]_t)} \times 100 \\ &= \frac{[B]_0[1 - \exp(-k_B t)]}{[A]_0[1 - \exp(-k_A t)] + [B]_0[1 - \exp(-k_B t)]} \times 100 \end{aligned} \quad [11]$$

where  $[P_A]_t$  and  $[P_B]_t$  are the concentrations of the products formed at time  $t$ . **Figure 3A** is a 3D plot of  $\%error_A$  as a function of  $k_A/k_B$  and  $[A]_0/[B]_0$ . As can be seen, the error made in the determination decreases with increasing rate constant ratio and increasing  $[A]_0/[B]_0$ .

The error made in the determination of B can be calculated from

$$\begin{aligned} \%e_B &= \frac{[A]_t}{([A]_t + [B]_t)} \times 100 \\ &= \frac{[A]_0[1 - \exp(-k_A t)]}{[A]_0[1 - \exp(-k_A t)] + [B]_0[1 - \exp(-k_B t)]} \times 100 \end{aligned} \quad [12]$$

**Figure 3B** is a similar 3D plot for  $\%error_B$ . As can be seen, the relative error decreases with increasing rate constant ratio and increasing concentration of B in the mixture.



**Figure 3** Variation of the relative error in the determination of (A) reactant A in a mixture with B and (B) reactant B in a mixture with A, as a function of  $[A]_0/[B]_0$  and  $k_A/k_B$ . Simulated data at 60 s reaction time.

### Simultaneous Kinetic-Based Determinations

Multideterminations of species are of great interest to kinetic methods of analysis. Various mathematical algorithms allow exploiting of the whole kinetic curve to determine the concentrations of two or more analytes in a mixture.

Curve-fitting methods compute the parameters that give the best fit of some mathematical model to signal-time data. Traditionally, least-squares methods have been most frequently used. The Kalman filter, however, which is one of the most powerful algorithms available for simultaneous kinetic-based determinations, allows reliable calculation of the rate constant,  $k$ , the net signal change,  $\Delta A$ , and the background contribution,  $C$ , from the following equation for a binary mixture:

$$\Delta A_t = \Delta A_A[1 - \exp(-k_A t)] + \Delta A_B[1 - \exp(k_B t)] + C \quad [13]$$

where  $A_t$  is the total absorbance at time  $t$  and  $\Delta A_A$  and  $\Delta A_B$ , and  $k_A$  and  $k_B$  are the net absorbance changes and the rate constants of A and B, respectively. It is superior to classical multicomponent methods (see first section of this topic), especially as regards the rate constant ratio of reactions for a binary mixture, which can be closer to unity than in the classical methods. Thus the Kalman filter provided satisfactory results (error less than 10%) in the resolution of mixtures of 2- and 3-chlorophenol (concentration ratio 1:4 to 4:1), where the rate constant ratio was only 1.4. Linear or nonlinear models are used to resolve mixtures, depending on whether or not the rate constants of the two reactions are accurately known.

The use of multidetection devices such as diode array systems allows the application of multivariate calibration methods. To gain selectivity both spectral and temporal domains are incorporated. The simultaneous determination of analytes is facilitated as no prior knowledge of rate constants and explicit adoption of a kinetic model are necessary.

Recently artificial neural networks have also been successfully applied. So the concentration ratio for the above-mentioned resolution of chlorophenols could be 1:9 to 9:1 with an error of 10%.

### Reaction Types

The different reaction types involved in individual and simultaneous kinetic-based determinations of species can be classified into several categories, which are described in detail below. As a rule, kinetic-based determinations of a single species have chiefly been used for organic species, whereas simultaneous

determinations have been applied to inorganic and organic species alike.

### Complexation Reactions

This group encompasses both formation and substitution reactions, both of which have mainly been applied to inorganic analysis.

**Substitution reactions** Methods based on substitution reactions can in turn be classified according to whether a ligand or a metal is displaced from the starting complex:



The ligands most frequently used in this context are aminopolycarboxylic acids, particularly 1,2-diaminocyclohexane- $N,N,N',N'$ -tetraacetic acid (DCTA) on account of the advantages deriving from the singular kinetic behavior of M-DCTA complexes, which are employed either as chelating agents or scavengers. Classical chromogenic reagents such as (4-(2'-pyridylazo)resorcinol) (PAR) and (1-(2'-pyridylazo)-2-naphthol) are often included in the starting complexes.

Binary, ternary, and even quaternary mixtures of transition metals can be resolved by ligand substitution; on the other hand, metal-substitution reactions permit the resolution of mixtures of alkaline earth metals. These determinations usually involve photometric detection thanks to the chromogenic reagents used; also, signal readings are usually processed by using the logarithmic extrapolation method and stopped-flow instrumentation is required on account of the typically short half-lives of the reactions involved.

**Complex-formation reactions** This type of reaction is mainly used for the determination of a single metal ion and is less commonly employed (together with dissociation reactions) in simultaneous kinetic-based determinations than are substitution reactions.

Regarding determinations of a single species, iron, copper, and aluminum are the metal ions most frequently determined. The reactivity of iron(II) toward ferriox-type ligands (1,10-phenanthroline, 2,2'-bipyridyl, ferrozine, etc.) is the basis for a number of methods. The special interest of aluminum can be ascribed to its fluorescent chelates, which allow the highly sensitive kinetic determination of this metal ion.

Few simultaneous kinetic methods are based on this type of reaction. One typical example is the

simultaneous analysis for nitric oxide and nitrogen dioxide in air by formation of the iron(II)–nitrosyl complexes from nitrogen oxide and nitrogen dioxide. Also, a representative example of a dissociation reaction is the decomposition of the nitrilotriacetic acid (NTA) complexes of molybdenum(VI) and tungsten(VI) by the hydroxide ion.

### Redox Reactions

This type of reaction has been used for the determination of anionic rather than cationic species in inorganic analysis. Thus, halides have been determined using their inhibitory effect on the oxidation of organic compounds by metal ions, as well as in photochemical reactions; on the other hand, the oxidizing power of some metal ions such as cerium(IV) and vanadium(V) has been exploited for their own determination. Photochemical reduction reactions have also been used for the determination of anions such as nitrate in effluents. Most of the few reported simultaneous kinetic determinations relying on redox reactions are based on the reduction of heteropoly acids formed between molybdate and silicate, phosphate, or germanate ions.

Redox reactions are of special interest in organic analysis and rely on the oxidation of organic compounds by a strong oxidant such as permanganate, periodate, or hydrogen peroxide. For example, periodate is a common oxidant for organic hydroxyl compounds including phenols, chlorophenols, and vicinal glycols (Malaprade reaction). Fixed-time (absorbance measurements) and variable-time (perchlorate ion-selective electrode) methods can be used for the determination of these species and related substances such as carbohydrates.

One interesting redox reaction is the reversible interconversion of ferriin/ferroin complexes. The 1,10-phenanthroline/iron(III) complex (ferriin) has been used to oxidize species of clinical and pharmaceutical interest such as analgesics and catecholamines, the reaction being monitored by measuring the initial rate of change of the absorbance of the ferroin formed at 510 nm. This type of reaction has also been used in simultaneous kinetic-based determinations, such as the resolution of mixtures of uric acid and ascorbic acid, paracetamol and oxyphenbutazone, and adrenaline and noradrenaline.

As a rule, these determinations use the initial reaction rate method. Their prominence in recent years can be ascribed to the increasing availability of inexpensive mixing systems such as modular stopped-flow and continuous-addition-of-reagent devices, which enable kinetic monitoring.

### Coupled Reactions

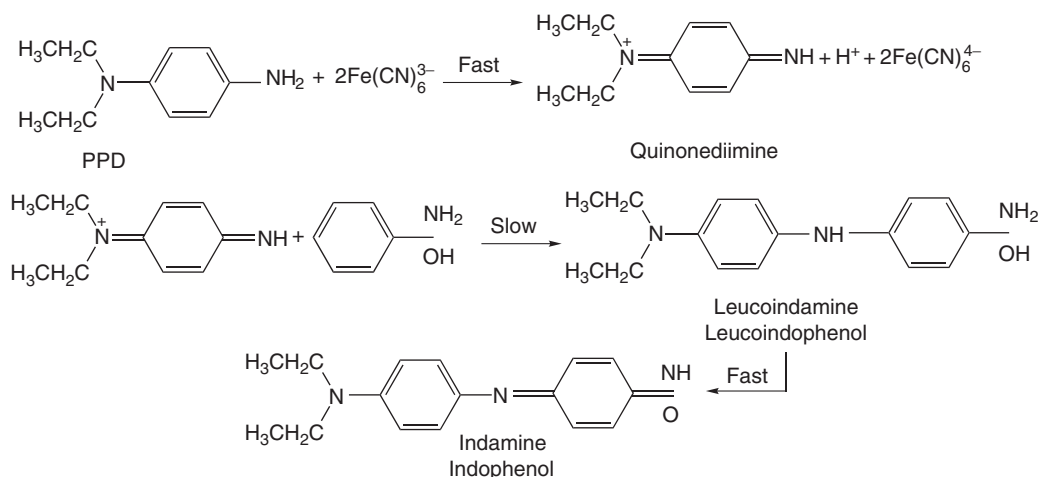
The most representative example of this type of reaction is the so-called Griess reaction, originally developed for the kinetic determination of nitrite. Nitrite in a weakly acidic solution reacts with an amine (usually sulfanilic acid) to form the corresponding diazonium salt, which then couples with a naphthylamine or naphthol. This reaction can also be used for the kinetic determination of nitrate after reduction in a column, as well as for measuring other species such as sulfonamides (Bratton–Marshall reaction), and benzodiazepines and *N*-methylcarbamate pesticides, the hydrolysis of which yields the corresponding benzophenones (which can be diazotized) and naphthols (useful for the coupling reaction), respectively.

Oxidative coupling reactions of *p*-phenylenediamines with amines and phenols are widely used. As can be seen in **Scheme 1** *p*-phenylene diamine is oxidized to its quinone diimine (QDI) by potassium hexacyanoferrate(III) or a similar oxidant in a weakly basic medium. In the rate-limiting step, QDI reacts with the amine or phenol to give leuco-indamines (indophenols), which are rapidly oxidized to colored indamines (indophenols) with the aid of a QDI molecule. This reaction is chiefly used for resolving a large variety of mixtures of aromatic amines, phenols, and chlorophenols.

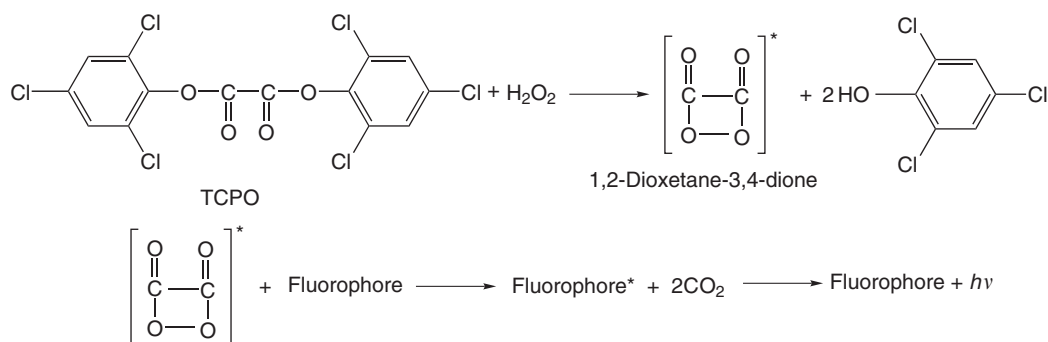
Bromination reactions use bromine generated *in situ* from the  $\text{BrO}_3^-/\text{Br}^-$  couple in an acidic medium. At low analyte concentrations, the rate of bromination and bromine generation are virtually the same, and so methyl orange in the reaction medium will only be decolorized after the bromination process has been completed. The decolorization time will be proportional to the analyte concentration. Bromide can be determined in addition to phenol compounds, which are the typical substrates. Salicylic acid and paracetamol in mixtures with caffeine can also be determined this way.

### Chemiluminescence Reactions

While the main field of application of chemiluminescence (CL) reactions is catalytic processes (e.g., oxidation of luminol in the presence of metal ions), some noncatalytic reactions are also of great interest on account of the high sensitivity they provide. One of the most prominent of such reactions is that involving the hydrogen peroxide-induced oxidation of aryl oxalate esters in the presence of a suitable fluorescent species (**Scheme 2**). The most widely used aryl oxalates for this purpose are bis(2,4,6-trichlorophenyl)oxalate (TCPO) and bis(2,4-dinitrophenyl)oxalate. These CL reactions have been used



Scheme 1



Scheme 2

for liquid chromatographic detection in the determination of a great variety of fluorescent compounds by their native fluorescence (or that of a compound obtained on derivatization).

### Miscellaneous Reactions

Formation of addition compounds from species containing carbonyl groups has been exploited for the determination of organic compounds reacting with hydrogen sulfite ion in the presence of acetone, methanol, or acetic acid. Primary amines can be determined alone and in mixtures with secondary and tertiary amines by condensation with aldehydes such as salicylaldehyde and 4-dimethylaminocinnamaldehyde.

Hydrolysis reactions have also been used for individual (e.g., nitroglycerine in ethanol) and simultaneous kinetic-based determinations (e.g., penicillins and their derivatives).

The determination of the rates of methyl ketal formation in the reaction of dehydroascorbic acid

has been used to determine its content in tablets of vitamin C.

In the Fujiwara reaction chlorohydrocarbons react with pyridine (*N*-alkylation) under complete destruction of the ring. But the reaction path will go through a number of Schiff base intermediates, which, depending of the kind of analyte, have different life times and absorbances in the visible range. Thus mixtures of 1,1,1-trichloroethane, trichloroethylene, and chloroform could be resolved, even in the presence of unknown interfering substances.

### Reactions in Micellar Systems

The use of organic microheterogeneous systems for developing new or improving known kinetic-based determinations is progressing well. Many micellar kinetic determinations based on uncatalyzed reactions have involved aromatic nucleophilic substitutions, e.g., reactions between 1-fluoro-2,2-dinitrobenzene and primary and secondary amines, phenolic compounds, and thiols. These reactions

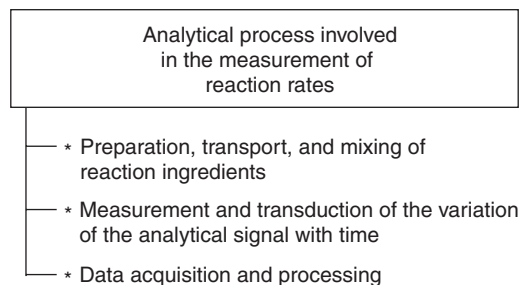


have been considerably accelerated by cationic micelles. A determination of different analytes is possible, even in turbid and colored samples, by using a fluoride ion-selective electrode. A kinetic method for the determination of hydrazine and phenylhydrazine based on their condensation with *p*-dimethylaminobenzaldehyde in anionic sodium dodecyl sulfate solutions has been described.

Very promising is the application of micelles for differential reaction rate methods. Micelles can alter the rate constant ratio of two or more species that interact with a common reagent. Simultaneous kinetic determination of nickel and cobalt based on the complex formation with 5-octyloxymethyl-8-quinolinol in the nonionic micellar medium of Triton X-100 is effective as this surfactant decreases the rates of formation of both complexes compared with an aqueous medium, so permitting their spectrophotometric monitoring. In the micellar medium the formation of the Co complex is 44 times faster than that of Ni, and determinations of both ions in the  $10^{-5}$  mol l<sup>-1</sup> range are possible.

## Methods of Monitoring Reactions

Kinetic-based determinations involving noncatalytic reactions use instruments of variable complexity for monitoring reactions. The great developments in the so-called intelligent instrumentation have fostered automation to a greater or lesser extent of the three essential steps involved in measuring reaction rates (Figure 4), the first of which is the most relevant to automation on account of the high complexity of the operations involved (complete automation of this step, however, is still impossible). Thus, the choice of the type of instrument to be used should be dictated not only by its technical features (mainly related to the second and third steps) but also by the half-life of the reaction concerned (closely related to the first step).



**Figure 4** Basic steps of reaction rate-based determinations. (Reproduced with permission from Silva M (1993) Recent strategies in automated reaction rate-based determinations. Tutorial Review. *Analyst* 118: 681–688.)

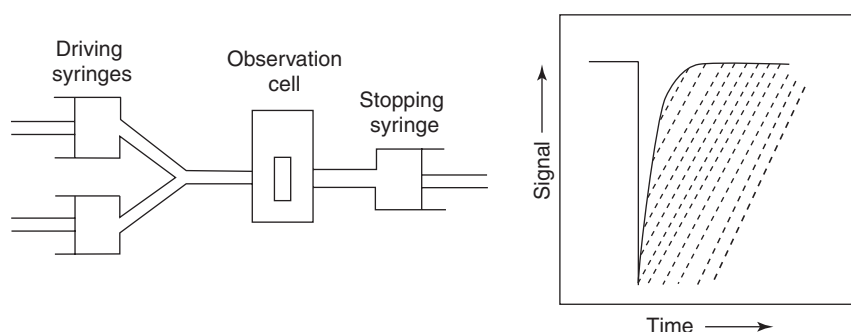
Manual mixing of sample and reagents in the reaction vessel, which may be a spectrophotometer cuvette or an electrode cell, is usually adequate for monitoring slow reactions, whether in batch or closed systems. Flow systems avoid some manipulation, and so they result in increased precision in the determinations.

Flow systems, both continuous and discrete, are used in kinetic-based determinations for monitoring fast reactions mainly. To this end (1) the dead time in the mixing system should be several orders of magnitude lower than the half-life of the reaction concerned; and (2) nearly the whole kinetic curve must be recorded in order to implement reaction rate-based determinations and perform fundamental kinetic studies (e.g., the determination of reaction orders and rate constants). The advent of stopped-flow mixing and the continuous-addition-of-reagent technique has made noncatalytic reactions competitive with equilibrium methods in practical terms.

## The Stopped-Flow Technique

The stopped-flow technique is the most commonly used for studying and implementing application of fast reactions with half-lives between a few milliseconds and a few seconds. The special features of this technique, in which reactants are driven at a high rate into a mixing and/or observation cell, the flow is abruptly stopped, and the extent of reaction monitored (Figure 5), have facilitated studies on the kinetics and mechanism of fast reactions and enabled the development of reaction rate-based determinations.

The stopped-flow technique was introduced by Chance in 1940. Earlier applications to kinetic analysis were concerned with studies on kinetics and reaction mechanisms (e.g., the formation of the iron(III)–thiocyanate complex, that of 12-molybdophosphoric acid, the redox reaction between 2,6-dichlorophenolindophenol and ascorbic acid, etc.) as well as the resolution of mixtures of metal ions using substitution reactions. On the other hand, the inception of commercially available stopped-flow instruments and inexpensive modular mixing systems for adaptation to existing detectors have led to a broad use of this technique in routine kinetic determination of individual species and mixtures in a variety of samples of clinical, pharmaceutical, nutritional, and environmental interest. The analytical features of the methods developed for this purpose usually surpass those of the equilibrium counterparts, as shown by the selected examples given in Table 2. In addition, stopped-flow systems accelerate some slow reactions relative to the conventional kinetic technique as a



**Figure 5** Typical instrumental setup used and signal–time profile provided by the stopped-flow technique. The shaded area denotes reaction development during the stop time. (Reproduced with permission from Silva M (1993) Recent strategies in automated reaction rate-based determinations. *Tutorial Review: Analyst* 118: 681–688.)

**Table 2** Selected applications illustrating the advantages of the stopped-flow technique over equilibrium methods

Analyte/chemical system	Remarks
Urea/biacetyl monoxime	Conventional equilibrium method: reaction time ~20–30 min; uses a high-temperature and a concentrated acid medium. Stopped flow: 2–3 min for acquisition of analytical data; avoids the use of drastic experimental conditions
Carbaryl/diazotized sulfanilic acid	Wider determination range; higher selectivity factors for other <i>N</i> -methylcarbamate pesticides; no blank is required for the stopped-flow technique
Psychotropic drugs/hydrogen peroxide	Addition of an oxidant is unnecessary in the stopped-flow technique since oxidation is effected by dissolved oxygen itself

result of mixing taking place at a higher pressure. Hence the stopped-flow technique can also be advantageously applied to this type of reaction, thereby saving time and reagents and simplifying handling of the reaction ingredients.

### The Continuous-Addition-of-Reagent Technique

This technique is a major alternative to the stopped-flow technique for performing rate measurements on fast reactions as it uses simpler instrumentation (Figure 4, previous article) to add the reagent (*R*) continuously at a constant rate, *u*, over a volume *V*<sub>0</sub> of sample containing the analyte (*A*). The overall reaction rate of the process, which depends on the rate of the reaction and the dilution of the species present in the reaction vessel, can be expressed in integral form as

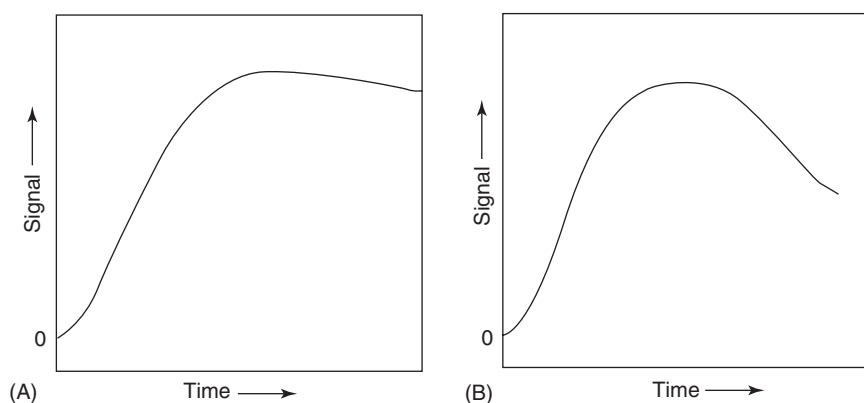
$$\ln([A]/[A]_0) = -k[R]_0 t + (k[R]_0 V_0/u - 1)\ln(V_0 + ut/V_0) \quad [14]$$

where *[A]*<sub>0</sub> and *[A]* are the analyte concentrations at time zero and time *t*, respectively; *[R]*<sub>0</sub> is the reagent concentration in the addition unit; and *k* is the pseudo-second-order reaction rate constant. The special way in which sample and reagents are mixed in this technique allows reaction half-lives to be altered

through changes in the reagent concentration and its rate of addition.

This technique provides a full kinetic profile (Figure 6A) that can be used to implement two reaction rate methodologies, namely (1) the initial-reaction method, based on the initial concave portion of the curve, along which the analytical signal is directly proportional to *t*<sup>2</sup> and (2) the maximum-reaction method, which relies on the linear intermediate portion of the curve. Based on reported results, the maximum-reaction method is preferable. In addition, it offers several advantages over traditional pseudo-first-order initial-rate methods, particularly a much wider linear portion for measurements to be made, where instrumental errors are much smaller.

This technique is currently being used for individual and multiple determinations of species with interesting results. However, it is particularly suitable for reactions where excess reagent might have undesirable effects on the analytical signal (e.g., the determination of sulfonamides using the Bratton–Marshall reaction). Even though the kinetic profile obtained in such a case is somewhat special (Figure 6B), the maximum-reaction method can be accurate over the linear portion of the kinetic curve.



**Figure 6** Kinetic profiles provided by continuous-addition-of-reagent technique. (A) Typical kinetic curve. (B) Recording obtained for a chemical system involving an undesirable interaction between the reagent and some reaction ingredient. (Reproduced with permission from Silva M (1993) Recent strategies in automated reaction rate-based determinations. *Tutorial Review: Analyst* 118: 681–688.)

## Analytes Determined

Originally, noncatalytic reactions were seldom used for the determination of a single inorganic or organic species because equilibrium methods were faster and just as sensitive. Their use was limited to slow reactions, for which kinetic methods offer some advantages over equilibrium methods (e.g., the effect of undesirable side-reactions is avoided and analytical data are acquired in a shorter time). The above-described automated mixing devices have extended the scope of application of these kinetic-based determinations to fast reactions, especially in organic analysis. These methods make interesting alternatives to equilibrium methods for practical purposes. On the other hand, simultaneous kinetic determinations clearly outdo equilibrium methods. Classical differential methods were formerly aimed at the resolution of mixtures of inorganic species but are currently more often applied to the simultaneous determination of organic species, especially using multipoint methodologies.

### Determination of Inorganic Species

Complex-formation and redox reactions have so far been the most frequently used for the determination of both metals and nonmetals (Table 3). A variety of transition metal ions have been thus determined including such common elements as iron, copper, and calcium and rare earths as well as technetium and europium. Common nonmetals such as nitrogen anions and phosphate, bromide, and sulfur anions have also been determined this way. However, the most important applications of noncatalytic reactions in inorganic analysis are the simultaneous determinations of metal ions. As can be seen in Table 4, a wide variety of binary mixtures and some ternary and even

quaternary mixtures have been resolved using complexation reactions mainly. It is worth emphasizing the determinations of alkaline earth metals, both individually and in mixtures, which are unaffordable by methods based on catalytic reactions as the metal ions exert no catalytic effect.

The analytical features of these determinations warrant some comments. Thus, (1) photometric and potentiometric (ion-selective electrodes) techniques are widely used; (2) in many cases, determinations are nearly fully automated by using the stopped-flow technique for mixing sample and reagents; (3) determination ranges typically lie in the microgram per milliliter range except for fluorimetric detection (the limit of detection (LOD) for europium in Table 4 is  $3 \times 10^{-13} \text{ mol l}^{-1}$ ); and (4) the maximum analyte concentration ratio afforded by simultaneous methods is usually  $\sim 5:1$  but also occasionally  $10:1$  or even higher, depending on the rate constant ratio and the particular method used.

### Determination of Organic Compounds

Kinetic-based determinations involving noncatalytic reactions are quite commonly used for the determination of organic substances since these generally exhibit no catalytic properties and so they are elusive to conventional equilibrium methods. A great variety of organic compounds have been determined this way including substances possessing reducible groups or aromatic nuclei, amines, carbonyl compounds, etc. – the former, however, are the most frequently determined on account of their reactivity toward strong oxidants (Table 5). Table 6 includes some other types of reaction. Species of clinical and pharmaceutical interest such as vitamins, catecholamines, analgesics, and psychotropic drugs have been

**Table 3** Selected determinations of inorganic species using noncatalytic reactions

Analyte	Chemical system	Determination range or LOD <sup>a</sup>	Observations
Iron	Complex formation with thiocyanate	1.0–30	Iron(II; III) can be determined using SF
Copper	Complex formation with 5,5-di-methyl-1,3-cyclohexanedione bis(4-phenyl-3-thiosemicarbazone)	0.8–8	Reaction medium: 2.5–3.0 mol l <sup>-1</sup> HCl. Both fixed time and initial rate were used
Calcium	Complex formation with <i>o</i> -cresolphthalein	1.0–50	Endpoint determination using SF
Technetium	Color formation with 1,3,5-triphenyl- $\Delta^2$ -pyrazoline	0.01–12	Initial-rate and fixed-time methods used
Aluminum	Complex formation with F <sup>-</sup>		Stopped-flow technique, fluoride-selective electrode
Europium	Complex formation with 4,7-bis(chlorosulfophenyl)-1,10-phenanthroline-2,9-dicarboxylic acid	$3 \times 10^{-13}$ mol l <sup>-1</sup>	Determination using microsecond time-resolved fluorimetry
Hydrazine	Reaction with Mo(VI)	10 <sup>-3</sup> mol l <sup>-1</sup> level	Use of variable and fixed-time methods. Spectrophotometric detection
Nitrite	Griess reaction	$1.0 \times 10^{-6}$ mol l <sup>-1</sup>	SF mixing. Nitrate can be determined using cadmium
Sulfite	Reaction with Methyl Green	0.05–25	Fixed-time procedure, absorbance measurement at 625 nm
Sulfur anions	Reaction with iodate		The iodide produced is monitored using an iodide-selective electrode
Phosphate	Complex formation with Mo(VI)	20–100	The fast formation of 12-molybdo-phosphoric acid was studied using SF
Bromide	Bromination reaction	0.1–20	Chromotrope 2B can be used instead of methyl orange

<sup>a</sup>In micrograms per milliliter or as stated otherwise.  
LOD, limit of detection; SF, stopped flow.

determined regularly. As regards simultaneous kinetic determinations, early application involving simple organic compounds such as acids, alcohols, amines, and ketones have been extended to more complex substances in the environmental and clinical fields (Tables 7 and 8).

Organic kinetic-based determinations usually involve photometric or fluorimetric detection (many organic compounds and their reaction products exhibit native fluorescence). The determination levels typically achieved range from a few nanograms per milliliter to a few micrograms per milliliter but can be further lowered by using chemical and instrumental means. Examples of chemical and instrumental enhancement of these determinations are the use of micellar media for quantifying organic peroxides (Table 5) and diode array detection for measuring butylated hydroxyanisole (Table 6). Because fast reactions are quite commonplace in this context, the stopped-flow and continuous-addition-of-reagent techniques are frequently employed for mixing the sample and reagents. Finally, mixtures are usually resolved by using the classical differential reaction rate, logarithmic-extrapolation, and proportional-equation methods, even though currently multipoint

methods are interesting alternatives (e.g., for the simultaneous determination of corticosteroids using the Kalman filter algorithm, Table 7).

## Selected Applications

The use of kinetic methods of analysis based on noncatalytic reactions has leveled off in the last few years, especially as regards practical applications; in any case some of them are the best choices available to analytical chemists for individual and multicomponent determinations of species in a wide variety of real samples. This growing use of noncatalytic kinetic methods can largely be ascribed to the increased automation of reaction rate-based determinations, especially in organic analysis. Applications of these methods to real samples lie in various areas; those of environmental, clinical, pharmaceutical, and nutritional interest are discussed in some detail below on account of their great significance.

### Environmental Analysis

Noncatalytic kinetic methods have been widely applied in this area for several reasons, namely

**Table 4** Selected simultaneous determinations of inorganic species

<i>Chemical system</i>	<i>Mixtures analyzed</i>	<i>Observations</i>
M-EGTA + PAR → M-PAR + EGTA	Ni-Co, Zn-Pb, Fe-Ni, Zn-Cd, Cu-Pb, Co-Ni-Zn-Pb	Absorbance data at 500 nm are processed using the logarithmic extrapolation method. Determinations at the micromolar level
M-EGTA + BPDP → M-BPDP + EGTA	Fe(II)-Fe(III)	Progress of reaction was followed by monitoring absorbance at 556 nm
M-EDTA + PAR → M-PAR + EDTA	Ga-In	Determination in river waters
M-DCTA + Pb → Pb-DCTA + M	Mg-Ca, Sr-Ca, Sr-Ba, Mg-Ca-Sr	Determination range: $1 \times 10^{-6}$ – $5 \times 10^{-4}$ mg l <sup>-1</sup> Precision: 5–10%
M-XO + EDTA → M-EDTA + XO	La-Eu-Tb	Following the ligand replacement in a micellar solution. A partial least-squares regression algorithm was used for data processing
M + MTB	Mg-Ca	Photodiode array detection, usage of partial least-squares data analysis
X + Fe <sup>2+</sup> → Fe(X) <sup>2+</sup>	NO-NO <sub>2</sub>	Stopped-flow mixing. The samples assayed contained NO <sub>2</sub> at levels 2.3–5.5 times higher than that of NO
X + MoO <sub>4</sub> <sup>2-</sup> → heteropolymolybdates	PO <sub>4</sub> <sup>3-</sup> -SiO <sub>3</sub> <sup>2-</sup>	Use of the CAR technique. Determination at phosphate/silicate ratios from 12:1 to 1:25. Determination in sunflowers
M + PT → M-PT	Fe-Co, Ni-Co, Cu-Fe	Logarithmic-extrapolation and single-point methods Interpolation method used to overcome synergistic effects Use of the continuous-addition-of-reagent technique
M + TCSP → M-TCSP	Zn-Hg	Using the method of proportional equations in an unsegmented continuous-flow system at two different pH values
Fe(II)DGT	IO <sub>3</sub> <sup>-</sup> -IO <sub>4</sub> <sup>-</sup>	Spectrophotometric determination of the Fe(III)DGT complex at 410 nm
M-NTA + OH <sup>-</sup> → MO <sub>4</sub> <sup>2-</sup> (MO <sub>3</sub> <sup>-</sup> ) + NTA	Mo-W, Mo-V, Mo-W-V	Stopped-flow mixing with ultraviolet detection at 260 nm Determination limit: $\sim 10^{-5}$ mol l <sup>-1</sup> . RSD: 2–4%
X + F <sup>-</sup> → Fluoro complexes	Fe(III)-Zr(IV)	Use of a fluoride-selective electrode. Kalman filter algorithm
XO <sub>3</sub> <sup>-</sup> (XO <sub>4</sub> <sup>-</sup> ) + I <sup>-</sup> → I <sub>2</sub>	IO <sub>3</sub> <sup>-</sup> -BrO <sub>3</sub> <sup>-</sup> , IO <sub>4</sub> <sup>-</sup> -BrO <sub>3</sub> <sup>-</sup>	Use of a continuous-flow system and the logarithmic-extrapolation method
Luminol + K <sub>2</sub> S <sub>2</sub> O <sub>8</sub> Hydrolysis at 70°C	Au-Ag Phosphates	Time-resolved CL Resolution of pyrophosphate/tripolyphosphate mixtures in a 1.0 mol l <sup>-1</sup> HCl medium
X + DTNB	CN <sup>-</sup> -S <sup>2-</sup> -SO <sub>3</sub> <sup>2-</sup>	Application of the cationic surfactant cetyltrimethylammonium bromide. Binary mixtures may be determined in the 10 <sup>-4</sup> mol l <sup>-1</sup> range using the multiple linear regression method
XO <sub>4</sub> <sup>3-</sup> -MoO <sub>4</sub> <sup>2-</sup>	PO <sub>4</sub> <sup>3-</sup> -AsO <sub>4</sub> <sup>3-</sup>	The heteropoly acids formed will be subsequently reduced by ascorbic acid to form heteropoly blue species. Triton X-100 micelles were present
M + morin	Nb-Ta, Ga-In	Use of linear extrapolation methods and fluorescence monitoring. Detection limits of 10 <sup>-8</sup> mol l <sup>-1</sup>
M + PAR	Ln-Pr-Nd	Data obtained at multiple wavelengths over time by use of stopped-flow mixing and photodiode array detector

EGTA, ethyleneglycol bis(2-aminoethyl ether)-*N,N,N',N'*-tetraacetic acid; DTNB, 5,5'-dithiobis-(2-nitrobenzoic acid); MTB, Methylthymol Blue; PT, pyridoxal thiosemicarbazone; DGT, dipyrldylglyoxal dithiocarbazone; HMF, 3-hydroxy-7-methoxyflavone; TCSP, 5,10,15,20-tetrakis-(3-chloro-4-sulfophenyl)porphine; RSD, relative standard deviation; PAR: 4-(2'-pyridylazo)resorcinol; DCTA, 1,2-diaminocyclohexane-*N,N,N',N'* tetraacetic acid.

(1) matrixes are usually quite simple, and so they allow direct determinations; (2) environmental samples are abundant and readily available; and (3) the properties of these samples make them ideal candidates for automated analysis. Table 9 shows most of the species determined kinetically using noncatalytic reactions. As can be seen, most of the entries are

kinetic-based determinations of inorganic species in water samples.

### Clinical Analysis

Serum and urine are the two types of sample most frequently analyzed in this context. The species



**Table 5** Selected determinations of a single organic species based on redox reactions

Analyte	Reagent	Determination range or LOD <sup>a</sup>	Observations
Ascorbic acid	DPIP	$2.0 \times 10^{-7}$ – $2.0 \times 10^{-5}$	Stopped-flow mixing. Measurement time 10 s. RSD 2%
Ascorbic acid	MoO <sub>4</sub> <sup>2-</sup>	$5.0 \times 10^{-5}$ – $6.0 \times 10^{-3}$	Fixed-time spectrophotometric method. Determination in fruits and pharmaceuticals.
Tartaric acid	IO <sub>4</sub> <sup>-</sup>	$4.0 \times 10^{-7}$ – $1.2 \times 10^{-4}$	Use of a liquid-membrane periodate-selective electrode
Vicinal glycols	IO <sub>4</sub> <sup>-</sup>	$1.4 \times 10^{-3}$ – $7.0 \times 10^{-3}$	Malaprade reaction. Use of variable-time method
Catecholamines	IO <sub>4</sub> <sup>-</sup>	$2.0 \times 10^{-4}$ – $2.0 \times 10^{-2}$	Allows mixture resolution
Sucrose	IO <sub>4</sub> <sup>-</sup>	0.01–1.0	Reaction requires 30 min prehydrolysis time
Thiamine	Fe(CN) <sub>6</sub> <sup>3-</sup>	$8.3 \times 10^{-9}$ – $7.5 \times 10^{-6}$	Use of the CAR technique and fluorimetric detection
Morphine	Fe(CN) <sub>6</sub> <sup>3-</sup>	$4.5 \times 10^{-8}$ – $2.5 \times 10^{-6}$	Useful for solving a variety of medical problems
Carbamazepine	Ce(IV)	0.04–140 $\mu\text{g ml}^{-1}$	Stopped-flow mixing and fluorimetric detection
Choline	Ce(IV)	$1.0 \times 10^{-5}$ – $1.0 \times 10^{-3}$	Fluorimetric monitoring of the Ce <sup>3+</sup> formed
Ethanol	MnO <sub>4</sub> <sup>-</sup>	0.2–2.0 $\mu\text{g ml}^{-1}$	Up to a fivefold concentration of methanol is tolerated
Uric acid	H <sub>2</sub> O <sub>2</sub> /TRIAP	0.03–3 $\mu\text{g ml}^{-1}$	Stopped-flow mixing. Reactions carried out at 40°C
Acetaminophen	Ferriin	0.25–25 $\mu\text{g ml}^{-1}$	Stopped-flow mixing. A reaction mechanism is proposed
Organic peroxides	I <sub>3</sub> <sup>-</sup> /CPC	$5.0 \times 10^{-7}$ – $2.5 \times 10^{-6}$	Reaction in a micellar medium
Theophylline	Ce(IV)	$5.0 \times 10^{-5}$ – $3.5 \times 10^{-3}$	Stopped-flow procedure

<sup>a</sup> In moles per liter or as stated otherwise.DPIP, 2,6-dichlorophenol indophenol; TRIAP, 1,1,3-tricyano-2-aminoprop-1-ene; ferriin, 1,10-phenanthroline/Fe<sup>3+</sup> complex; CPC, hexadecylpyridinium chloride; CAR, continuous-addition-of-reagent; RSD, relative standard deviation.**Table 6** Selected determinations of a single organic species based on miscellaneous reactions

Analyte	Chemical system	Determination range or LOD <sup>a</sup>	Observations
Phenols	Bromination reaction	$10^{-3}$ mol l <sup>-1</sup>	Variable-time method
Primary and secondary amines	Reaction with 1-fluoro-2,4-dinitrobenzene	$10^{-4}$ – $10^{-3}$ mol l <sup>-1</sup>	A fluoride-selective electrode is used.
	Reaction with 1,2-naphthoquinone-4-sulfonate		Measurement of color formation in the presence of Triton X-100
N-Methylcarbamate pesticides	Griess reaction with 1-naphthol as coupling reagent	0.35–40	Carbaryl, carbofuran, and propoxur are determined using stopped-flow mixing
Sulfonamides	Griess reaction	0.5–8.5	Practical problems associated with these types of reaction were avoided by using the CAR technique
Benzodiazepines	Griess reaction	0.15–3.65	
Butylated hydroxyanisole	Gibbs reaction	0.5–400	Amplification method using a DA detector
Formaldehyde	Reaction with CN <sup>-</sup>	1.0–6.0	Use of an ion-selective electrode
Methemoglobin	Color formation with CN <sup>-</sup>		Curve fitting
Urea	Biacetyl + Fe <sup>3+</sup> + TSC	0.5–10	Direct determination in serum
Creatinine	Alkaline picrate	10–100	Fluorimetric detection. Stopped-flow mixing
Zineb	Zincon	0.25–30	Monitoring of the fast complex reaction using CAR
Procaine	Condensation with 1,2-naphthoquinone-4-sulfonate	1–200	Spectrophotometric detection and stopped-flow mixing
Cefadroxil	Alkaline hydrolysis	10–100 mg l <sup>-1</sup>	Absorbance measurement at 470 nm. Determination in capsules and tablets
Furazolidone, furaltadone	Alkaline hydrolysis	1–20	Use of method of tangents
Aromatic aldehydes	Condensation with barbituric acid	$5 \times 10^{-4}$ – $10^{-6}$ mol l <sup>-1</sup>	
Morphine	Reaction with 1-fluoro-2,4-dinitrobenzene	15	Potentiometric initial rate method with F <sup>-</sup> -selective electrode detection

<sup>a</sup> In micrograms per milliliter or as otherwise stated.

TSC: thiosemicarbazone; Zineb: dithiocarbamate fungicide; Zincon: 2-carboxy-2'-hydroxy-5'-sulfoformazylbenzene, DA: diode-array; CAR: continuous-addition-of-reagent.

**Table 7** Selected simultaneous determinations of organic species based on redox reactions

<i>Mixtures analyzed</i>	<i>Reagent</i>	<i>Observations</i>
Carbohydrates	$\text{IO}_4^-$	Enthalpimetric determination of glucose and fructose using a reaction cell fitted with a thermistor
Carboxylic acids	$\text{Ag}^{2+}$	Anodic oxidation of $\text{Ag}^+$ at a Pt electrode
Tartaric/formic acid	$\text{MnO}_4^-$	Second-order reactions. Mixtures of formic/tartaric acid in ratios from 1:10 to 4.5:1 can be resolved
Adrenaline/noradrenalin	$\text{Fe}(1,10\text{phen})^{3+}$	Spectrophotometric determination of the $\text{Fe}^{2+}$ complex
Paracetamol/oxyphenbutazone	$\text{Fe}(\text{bipy})^{3+}$	Spectrophotometric determination of the $\text{Fe}^{2+}$ complex
Amoxycillin/clavulanic acid	$\text{Ce}(\text{IV})$	Formation of the corresponding fluorescent derivatives. Stopped-flow mixing
Thiamine and its pyrophosphoric ester	$\text{Hg}^{2+}$	Use of the logarithmic-extrapolation method to resolve mixtures of thiamine and its pyrophosphoric ester in ratios of 4:1 to 1:15
Uric/ascorbic acid	$\text{Fe}(\text{bipy})^{3+}$	The tris(2,2'-bipyridine)iron(III) complex acts as the oxidant. A nonlinear method for estimating the mixture composition is used.
Catecholamines	$\text{IO}_4^-$	Mixtures of epinephrine and L-dopa at the micromole per liter level are resolved by formation of the corresponding aminochromes. Stopped-flow mixing
Phenothiazines	Dissolved $\text{O}_2$	Perphenazine/chlorpromazine mixtures at the micrograms per milliliter level are resolved with fluorimetric detection and stopped-flow mixing
	$\text{Fe}^{3+}$	Methotrimepazine/thioridazine mixtures of 12–88% of one component in the presence of the other can be determined with stopped-flow mixing and diode-array detection. Use of proportional equations
Corticosteroids	Blue Tetrazolium	Use of the Kalman filter algorithm to resolve mixtures of cortisone and hydrocortisone with a pseudo-first order rate constant ratio as low as 1.8 Comparison with logarithmic-extrapolation and proportional-equations methods
Ephedrine/phenylephedrine	$\text{IO}_4^-$	Use of stopped-flow mixing. Absorbance resulting from formazan formation was monitored at 620 nm. Method of proportional equations
Imipramine/desipramine	$\text{Fe}^{3+}$	Oxidative coupling with 3-methyl benzothiazoline-2-hydrazone. Initial rate measurements and change in absorbance measured after 5 min

typically determined in them, both inorganic and organic, are summarized in **Table 9**. Most of the determinations require wet or dry treatment for inorganic species and wet processing for organic compounds. Although direct determinations are occasionally possible (e.g., urea, uric acid, and phosphate in serum and urine), a separation technique is usually needed. Thus, protein precipitation is required in the determination of calcium and magnesium in serum; also Sep-Pack  $\text{C}_{18}$  cartridges are used to remove matrix interference in many determinations. The sensitivity achieved in many cases results in the rather low sample consumption (25–500  $\mu\text{l}$  for serum samples), which is of great significance to analyses in biological fluids.

### Pharmaceutical Analysis

Interest in the application of kinetic-based determinations in this area has grown steadily because

pharmaceutical samples are readily available and their matrixes are usually fairly simple. Virtually all kinds of preparations (e.g., tablets, capsules, syrups, injectables) have been assayed for a great variety of organic compounds (**Table 9**). Direct determinations are feasible in many cases.

### Food Analysis

Application of kinetic methods in this field has grown dramatically over the last few years. As can be seen from **Table 9**, the applications involve inorganic and organic species alike. The types of sample assayed are widely variable, and so no general sample treatment can be discussed. Ascorbic acid has been the organic species most frequently determined, whereas iron, nitrite, and phosphate have been the inorganic ions most commonly assayed.

**Table 8** Selected simultaneous determinations of organic species based on miscellaneous reactions

Type of reaction	Mixtures analyzed	Observations
Griess reaction	<i>N</i> -Methylcarbamate pesticides	Mixtures of carbaryl and its hydrolysis product (1-naphthol) and of carbofuran and propoxur can be resolved using the stopped-flow technique
	1- and 2-Naphthol	Mixtures in ratios from 20:1 to 1:5 were resolved as were ternary mixtures with carbaryl
Condensation	Spermine/spermidine	Mixtures in ratios of 20:1 to 1:20 were analyzed at the microgram per milliliter level. Photometric detection and stopped-flow mixing.
	Hydrazine/phenylhydrazine	Analytes reacted with <i>p</i> -dimethylaminobenzaldehyde and a diode array detector used to monitor the reaction
	Dehydroascorbic acid/ascorbic acid	Methyl ketal formation. Using fast atom bombardment mass spectrometry.
	Amino acids	Use of ninhydrin reaction. Binary or ternary mixtures from glycine, leucine, isoleucine, lysine, histidine can be resolved by using multipoint curve-fitting methods
Oxidative coupling	Cresols	Reaction with DPD/hexacyanoferrate(III). <i>o</i> -Cresol/ <i>m</i> -cresol mixtures determined at the micromoles per liter level
	Chlorophenols	Reaction with DPD/hexacyanoferrate(III). Resolution of mixtures of 2- and 3-chlorophenol, where the pseudo-first-order rate constant was only 1.4
Complex formation	Zineb/Maneb	Decomposition of dithiocarbamate-type pesticides and complexation of the released metal ion by Zincon. CAR technique with photometric detection
Hydrolysis	Penicillins	Hydrolysis reactions in an acidic medium. Mixtures of penicillins and 6-aminopenicillanic acid can be resolved.
CN group exchange	Amines	2-Hydroxybenzaldehyde azine used as reagent. Binary mixtures of hydrazine, hydroxylamine, and ammonia can be resolved. In all cases 2-hydroxybenzaldehyde hydrazone is formed, which is fluorimetrically detected
Fujiwara reaction	Chlorinated hydrocarbons	The red colored products of the reaction with the reagent (pyridine, water, base) were detected at different wavelengths using optical sensors

DPD, *N,N*-diethyl-*p*-phenylenediamine; Zineb and Maneb, dithiocarbamate fungicides; Zincon, 2-carboxy-2'-hydroxy-5'-sulfoformazyl-benzene; CAR, continuous-addition-of-reagent.

**Table 9** Some applications of kinetic noncatalytic methods

<i>Environmental analysis</i>	
Natural waters	Manganese, arsenic, magnesium, hypochlorite, silicate + phosphate, carbaryl, carbofuran, hydrazine, NTA
Wastewaters	Mercury, nitrite, benzaldehyde, formaldehyde
Sea and river waters	Calcium, aluminum, calcium + magnesium, gallium + indium, diaquat
Air	Manganese, nitrogen oxides(NO + NO <sub>2</sub> ), carbon dioxide
Others	Carbofuran (soils), carbaryl (vegetables), Zineb (vine and olive leaves)
<i>Clinical analysis</i>	
Serum	Calcium, magnesium, aluminum, copper + zinc, copper + iron, nitrate, phosphate. Creatinine, paracetamol, albumin, ascorbic acid, total serum protein, urea, uric acid, hemoglobin, methemoglobin, reduced glutathione, imipramine, desipramine
Urine	Nitrate, phosphate, ascorbic acid, sulfonamides, catecholamines, uric acid, morphine
<i>Pharmaceutical analysis</i>	
	Bismuth, ascorbic acid, tartaric acid, hydrazine and related compounds, nitroglycerine, paracetamol, sulfonamides, tetracyclines, thiamine, vitamin B <sub>12</sub> , theophiline, procaine, benzocaine, perphenazine, chlorpromazine, methotrimeprazine, thioridazine, catecholamines, oxazepam, nitrazepam, carbimazole, tobramycin, bromhexine, bromazepam, pyridoxine, amoxyllin, clavulanic acid, furazolidone, furaltadone, ephedrine, phenylephedrine
<i>Food analysis</i>	
	Calcium, magnesium, iron, tryptophan, and crude protein (animal feed), ascorbic acid (orange, lemon and tomato juice, green tea, potatoes, red pepper, spinach, and strawberry), sulfite (fruit juice), iron(II, III) (wines), nitrite (cod liver oil, cured meat), aluminum (spinach leaves, tea leaves), bromate (bread), tryptophan (vegetables and fish), butylated hydroxyanisole (sunflower, corn, and olive oil), kojic acid (fermentation broth), and sucrose (milk and soft drinks)

See also: **Blood and Plasma. Chemiluminescence:** Liquid-Phase. **Clinical Analysis:** Overview. **Food and Nutritional Analysis:** Overview. **Kinetic Methods:** Principles and Instrumentation. **Liquid Chromatography:** Column Technology.

## Further Reading

- Biennial reviews: Kinetic determinations and some kinetic aspects of analytical chemistry.
- Crouch SR (1993) Trends in kinetic methods of analysis. *Analytica Chimica Acta* 283: 453–470.
- Crouch SR, Cullen TF, Scheeline A, and Kirkor ES (1998) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 70: 53R–106R.
- Crouch SR, Scheeline A, and Kirkor ES (2000) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 72: 53R–70R.
- Mottola HA (1988) *Kinetic Aspects of Analytical Chemistry*. New York: Wiley.
- Mottola HA and Pérez-Bendito D (1992) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 64: 407R–428R.
- Mottola HA and Pérez-Bendito D (1994) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 66: 131R–162R.
- Mottola HA and Pérez-Bendito D (1996) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 68: 257R–289R.
- Mottola HA, Pérez-Bendito D, and Mark HB, Jr. (1990) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 62: 441R–461R.
- Pardue HL (1989) Kinetic aspects of analytical chemistry. *Analytica Chimica Acta* 216: 69–107.
- Pérez-Bendito D (1990) Approaches to differential reaction-rate methods. Plenary Lecture. *Analyst* 115: 689–697.
- Pérez-Bendito D and Silva M (1988) *Kinetic Methods in Analytical Chemistry*. Chichester: Ellis Horwood.
- Pérez-Bendito D, Silva M, and Gómez-Hens A (1989) Automated kinetic-based determinations for routine analyses: recent developments. *Trends in Analytical Chemistry* 8: 302–308.

## Catalytic Techniques

**D Pérez-Bendito**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 2437–2449, © 1995, Elsevier Ltd., with revisions made by G Werner.

## Introduction

Catalytic techniques are more significant than non-catalytic ones in the broad scope of applications of kinetic methods due to their superior sensitivity and selectivity. The general principles behind the techniques outlined in the following sections are described in more detail elsewhere in this encyclopedia. Thus, this article is exclusively concerned with reaction rate methods of catalyzed reactions in homogeneous media. No mention is made of enzymatic analysis, which is treated elsewhere.

## Principles of Catalytic Methods

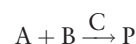
### General Considerations

Some substances including metals, nonmetals, and to some extent even organic compounds are known to accelerate slow reactions by a catalytic effect. Inasmuch as the rate of a catalyzed reaction is directly proportional to the catalyst concentration, such a reaction can be used for the determination of the

catalyst. Catalytic methods of this type are extremely sensitive since the catalyst is not consumed in the reaction but takes part in it in a cyclic manner. The fact that ‘small causes have great effects’ (e.g., on color development, luminescence) is the basis for trace and subtrace analysis using instrumental techniques. A catalyst can be defined as a substance that lowers the free activation energy of a reaction without altering its equilibrium position. The reaction that is catalyzed by the substance to be determined is known as the indicator reaction. One can find some 400 different indicator reactions in the literature for ~50 elements and numerous compounds. The rate of the blank indicator reaction must be very low or negligible relative to that of the catalyzed reaction if it is to be used for analytical purposes.

### Kinetic Equations and Reaction Mechanisms

The rate equation for a reaction



where B is in excess with respect to A and C being the catalyst is given by

$$v = -d[A]/dt = k[A][C]_0 + k_1[A] \quad [1]$$

or

$$v = d[P]/dt = (k[A]_0 - [P])[C]_0 + k_1([A]_0 - [P]) \quad [2]$$

See also: **Blood and Plasma. Chemiluminescence:** Liquid-Phase. **Clinical Analysis:** Overview. **Food and Nutritional Analysis:** Overview. **Kinetic Methods:** Principles and Instrumentation. **Liquid Chromatography:** Column Technology.

## Further Reading

- Biennial reviews: Kinetic determinations and some kinetic aspects of analytical chemistry.
- Crouch SR (1993) Trends in kinetic methods of analysis. *Analytica Chimica Acta* 283: 453–470.
- Crouch SR, Cullen TF, Scheeline A, and Kirkor ES (1998) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 70: 53R–106R.
- Crouch SR, Scheeline A, and Kirkor ES (2000) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 72: 53R–70R.
- Mottola HA (1988) *Kinetic Aspects of Analytical Chemistry*. New York: Wiley.
- Mottola HA and Pérez-Bendito D (1992) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 64: 407R–428R.
- Mottola HA and Pérez-Bendito D (1994) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 66: 131R–162R.
- Mottola HA and Pérez-Bendito D (1996) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 68: 257R–289R.
- Mottola HA, Pérez-Bendito D, and Mark HB, Jr. (1990) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 62: 441R–461R.
- Pardue HL (1989) Kinetic aspects of analytical chemistry. *Analytica Chimica Acta* 216: 69–107.
- Pérez-Bendito D (1990) Approaches to differential reaction-rate methods. Plenary Lecture. *Analyst* 115: 689–697.
- Pérez-Bendito D and Silva M (1988) *Kinetic Methods in Analytical Chemistry*. Chichester: Ellis Horwood.
- Pérez-Bendito D, Silva M, and Gómez-Hens A (1989) Automated kinetic-based determinations for routine analyses: recent developments. *Trends in Analytical Chemistry* 8: 302–308.

## Catalytic Techniques

**D Pérez-Bendito**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 2437–2449, © 1995, Elsevier Ltd., with revisions made by G Werner.

## Introduction

Catalytic techniques are more significant than non-catalytic ones in the broad scope of applications of kinetic methods due to their superior sensitivity and selectivity. The general principles behind the techniques outlined in the following sections are described in more detail elsewhere in this encyclopedia. Thus, this article is exclusively concerned with reaction rate methods of catalyzed reactions in homogeneous media. No mention is made of enzymatic analysis, which is treated elsewhere.

## Principles of Catalytic Methods

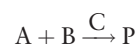
### General Considerations

Some substances including metals, nonmetals, and to some extent even organic compounds are known to accelerate slow reactions by a catalytic effect. Inasmuch as the rate of a catalyzed reaction is directly proportional to the catalyst concentration, such a reaction can be used for the determination of the

catalyst. Catalytic methods of this type are extremely sensitive since the catalyst is not consumed in the reaction but takes part in it in a cyclic manner. The fact that ‘small causes have great effects’ (e.g., on color development, luminescence) is the basis for trace and subtrace analysis using instrumental techniques. A catalyst can be defined as a substance that lowers the free activation energy of a reaction without altering its equilibrium position. The reaction that is catalyzed by the substance to be determined is known as the indicator reaction. One can find some 400 different indicator reactions in the literature for ~50 elements and numerous compounds. The rate of the blank indicator reaction must be very low or negligible relative to that of the catalyzed reaction if it is to be used for analytical purposes.

### Kinetic Equations and Reaction Mechanisms

The rate equation for a reaction



where B is in excess with respect to A and C being the catalyst is given by

$$v = -d[A]/dt = k[A][C]_0 + k_1[A] \quad [1]$$

or

$$v = d[P]/dt = (k[A]_0 - [P])[C]_0 + k_1([A]_0 - [P]) \quad [2]$$



according to whether reactant A or product P is monitored,  $[A]_0$  being the initial concentration of A,  $[P]$  the concentration of product formed, and  $k$  and  $k_1$  the pseudo-first order rate constants of the catalyzed and uncatalyzed reactions, respectively. The reaction must be made pseudo-first order in the monitored reactant since the catalyst concentration, by definition, does not change during the reaction.

The rate of the indicator reaction  $A + B \rightarrow P + Y$  can be increased by the action of a catalyst C in two different ways, i.e., by forming a transient intermediate complex with one of the reactants (e.g., B) or by reaction with one of the reactants (e.g., A) to yield one of the reaction products and the activated form of the catalyst.

In the first way, the catalyst–reactant complex (CB) interacts with the other reactant (A) to yield one of the reaction products while regenerating the catalyst.



Here two different cases can be considered depending on whether the rate-determining step (RDS) is reaction [I] or [II]. The overall rate of the process depends on the relative values of the constants  $k_1$ ,  $k_{-1}$ , and  $k_2$ . As a rule, reaction [II] is the RDS (i.e.,  $k_2 < k_{-1}$ ); the complex dissociation reaction is much faster than the reaction between such a complex and reactant A, and the process is said to be in pre-equilibrium. However, in some processes (chiefly enzymatic reactions), the opposite holds, and so reaction [I] is the rds (i.e.,  $k_2 > k_{-1}$ , and  $k_2 > k_1$ ) and the concentration of CB in the reaction medium is always constant and very small, hence the term steady state. Chelate and complex formation reactions between the catalyst and the substrate are highly significant to the above catalytic cycle and the basis for a host of catalytic reactions.

The other possible way in which the catalyst may interact with the ingredients of the indicator reaction involves reacting with A to yield P and its own activated form,  $C^*$ , from which it is rapidly regenerated on subsequent reaction with B to yield Y according to the following sequence



where [III] is the RDS. Many catalytic methods are based on this type of mechanism, where the formation of species  $C^*$  involves a change in the oxidation

state of the catalyst. For this mechanism to occur, the system concerned must meet two requirements, namely (1) the oxidation potential of the catalytic system,  $E_C$ , must be more positive than that of the P/A couple and more negative than that of the B/Y couple (i.e.,  $E_{B/Y} > E_C > E_{P/A}$ ); and (2) a direct interaction between A and B, though thermodynamically permitted, must be kinetically forbidden. In addition, the reaction between the catalyst and B should be very fast.

The sum of the constituent reactions of the two mechanisms makes up the overall indicator reaction. These mechanisms are deceptively simple; in fact, each step consists of a series of stages, one of which can be the actual RDS.

### Methods of Determination

Usage of a kinetic method to determine a catalyst entails recording (normally in an automatic or semi-automatic fashion) the kinetic curve, i.e., the variation of the measured property as a function of time. Kinetic curves consist of rising or falling segments, depending on whether the monitored species is a product or reactant, respectively.

Kinetic catalytic methods for determination of species can be classified in a manner similar to that of the kinetic noncatalytic methods described elsewhere in this encyclopedia (Table 1). Methods commonly used to measure induction periods are commented on in dealing with Landolt reactions below.

**Initial-rate method** The initial-rate method assumes that  $[P]$  is negligible compared with  $[A]_0$ ; the latter can be considered to remain constant in eqn [2], which can thus be simplified to

$$v_o = d[P]/dt = \tan \alpha = \Delta[P]/\Delta t = k'[C]_0 + k'_1 \quad [3]$$

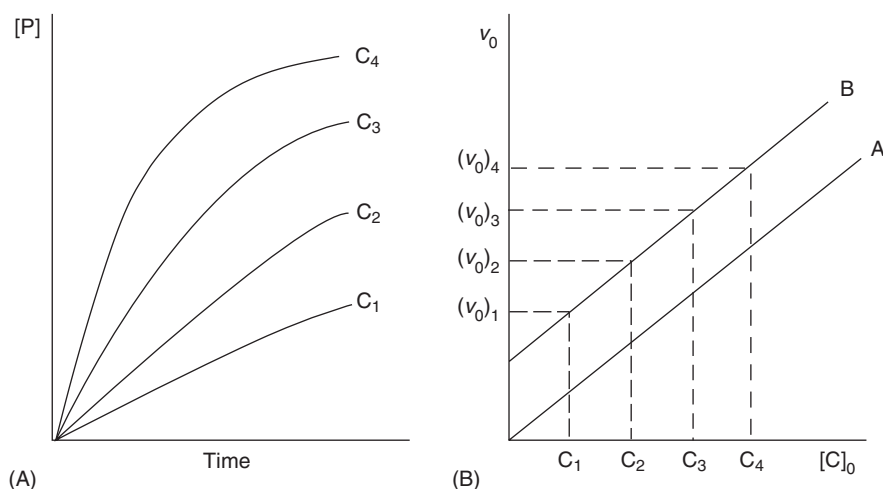
Therefore, the variation of  $[P]$  through the measured property as a function of time will be a straight segment (Figure 1A), which is the ultimate basis

**Table 1** Kinetic methods of analysis based on catalyzed reactions

*Differential*  
Initial rate  
Fixed time  
Variable time

*Integration*  
Tangent  
Fixed time  
Variable time

*Induction period measurements*



**Figure 1** Implementation of the initial-rate method. (Reproduced with permission from Pérez-Bendito D and Silva M (1988) *Kinetic Methods in Analytical Chemistry*. Chichester: Ellis Horwood.)

for this method. By plotting  $v_0$  against the catalyst concentration,  $[C]_0$ , for a series of samples containing known concentrations of the catalyst (Figure 1B) one obtains a straight line of zero (A) or nonzero intercept (B), depending on whether the uncatalyzed reaction develops to a negligible or appreciable extent, respectively.

**Differential fixed-time method** Implementation of this kinetic method involves measuring the concentration of a reactant or product at a preset time from the start of the reaction according to the following equation

$$\Delta[P] = k'[C]_0\Delta t + k_1'\Delta t \quad [4]$$

Since  $\Delta t$  is constant, plotting  $\Delta[P]$  as a function of  $[C]_0$  yields a straight line of slope  $k'\Delta t$  and intercept  $k_1'\Delta t$ , as shown in Figure 2 for  $k_1 = 0$ .

**Differential variable-time method** This method, also known as the 'fixed- or constant-concentration method', entails measuring the time required for a preset change in the reaction medium to take place. Solving eqn [4] for  $1/\Delta t$  yields

$$1/\Delta t = (k'[C]_0 + k_1')/\Delta[P] \quad [5]$$

Consequently, since  $\Delta[P]$  is constant, a plot of  $1/\Delta t$  versus  $[C]_0$  will be a straight line of slope  $k'/\Delta[P]$  and intercept  $k_1'/\Delta[P]$ , as shown in Figure 3 for  $k_1 = 0$ .

If  $[P]$  is not negligible relative to  $[A]_0$  in eqn [2], the equation must be integrated over a finite, though not necessarily short, time interval. Since integral methods are less commonly used than differential methods, they are not dealt with here.

### Catalytic Modified Effects

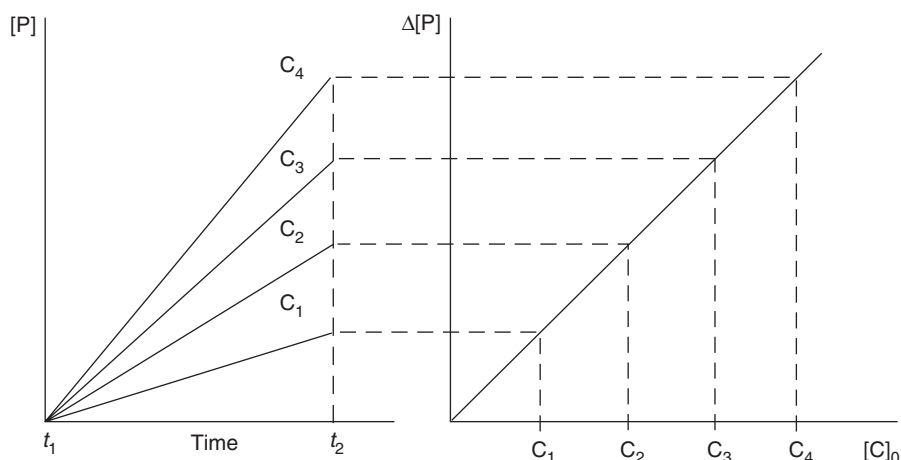
Catalytic determinations are based on either of the following approaches: (1) direct use of primary catalytic effects (determination of catalysts) and (2) use of modified catalytic reactions (improved determination of catalysts or determination of a modifier such as an activator or an inhibitor).

These effects, which change the rate of a catalytic reaction, are of great theoretical and practical significance. Catalytic methods are widely used in trace analysis on the grounds of their low detection limits and in some cases good selectivity. But there is always a wish to improve these figures of merit. The addition of auxiliary reagents may have these desired effects. Activators are substances that do not catalyze the indicator reaction, but their presence significantly increases the rate of the catalyzed reaction. There are several explanations for this effect, the most important being the facilitation of the catalyst-reactant interaction and the participation of the activator in the regeneration process of the catalyst.

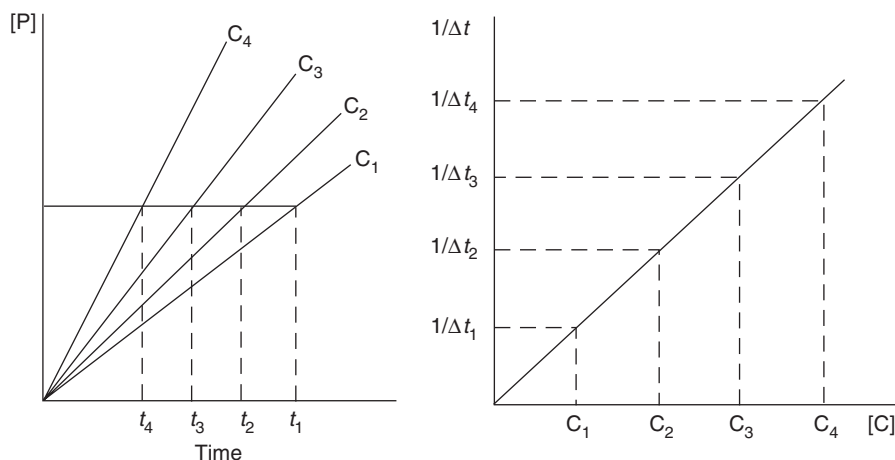
On the other hand some substances combine with the catalyst to form some sort of a complex, which either has less catalytic activity than the free metal ion or renders the catalyst completely inactive.

As described for uncatalyzed reactions, a way to enhance the sensitivity and/or selectivity of catalytic determinations is to have the reaction develop in a micellar medium.

The rate of a metal ion-catalyzed reaction is usually modified by the presence of hydroxylated, polyaminocarboxylic, and chromogenic ligands; however, complexones are preferentially used as inhibitors, while other ligands are employed as inhibitors or activators, depending on the particular



**Figure 2** Implementation of the fixed-time method. (Reproduced with permission from Pérez-Bendito D and Silva M (1988) *Kinetic Methods in Analytical Chemistry*. Chichester: Ellis Horwood.)



**Figure 3** Implementation of the variable-time method (Reproduced with permission from Pérez-Bendito D and Silva M (1988) *Kinetic Methods in Analytical Chemistry*. Chichester: Ellis Horwood.)

needs. Inorganic anions usually act as inhibitors for these reactions, while metal ions are used both to inhibit anion-catalyzed reactions and activate metal-catalyzed reactions.

Activators and inhibitors can be determined kinetically with good sensitivity and acceptable detection limits (between  $10^{-6}$  and  $10^{-5} \text{ mol l}^{-1}$ ), though not so good as those achieved in the determination of the catalyst. The precision, however, is rather poor in many instances. Similarly, inhibitory effects allow the kinetic determination of some substances that act as inhibitors for catalyzed reactions, provided the rate of the modified reaction is proportional to the inhibitor concentration.

### Catalytic Titrations

One of the major applications of inhibitory effects lies in the determination of the inhibitor using catalytic titration. This technique combines kinetic and

equilibrium concepts into an endpoint methodology. Catalytic titrations rely on the inhibitory effect of a given substance (an anion, ligand, or metal) on a metal- or anion-catalyzed reaction and are used for the determination of not only the inhibitor but also the catalyst or other species (usually metal ions) with no catalytic or inhibitory properties such as alkaline earth metals.

This methodology involves two consecutive reactions that can be schematized as follows:

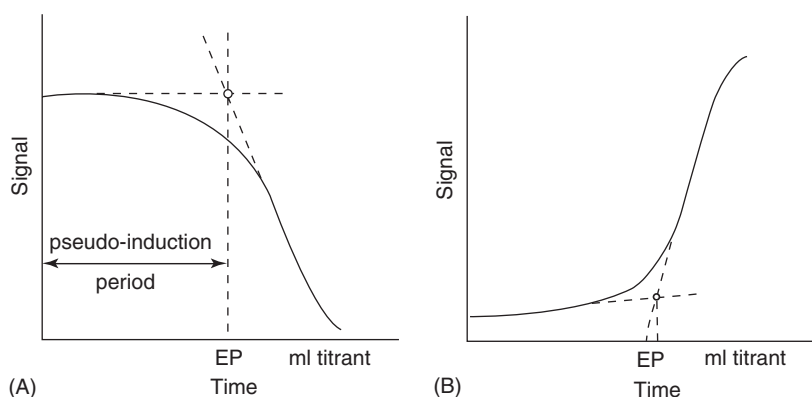
1. the titration reaction,



2. the indicator reaction [V],



The ingredients of the indicator reaction are added to the unknown solution (analyte inhibitor) in the



**Figure 4** Generic types of titration curve obtained by monitoring (A) a reactant and (B) a product of the indicator reaction (EP = end point). (Reproduced with permission from Pérez-Bendito D and Silva M (1988) *Kinetic Methods in Analytical Chemistry*. Chichester: Ellis Horwood.)

reaction vessel, and the catalyst is poured from a burette to start the titration reaction, which must meet the typical requisites for a conventional titration reaction (i.e., it should be fast and stoichiometric and develop to completion). The indicator reaction is only started by an excess of titrant (catalyst) after the endpoint and lends itself to instrumental monitoring via one of its reactants or products. Also, it is essential that the optimum conditions for development of the titration reaction (pH, solvent, ionic strength) should be the same as those for the indicator reaction.

Kinetic titration curves are constructed by plotting the relative analytical signal (absorbance, fluorescence, potential, or temperature) as a function of the volume of titrant added. If the two parameters are linearly related, then the titration curve will consist of two linear segments that can be extrapolated to intersect at the endpoint, as shown in **Figure 4**. Current instrumentation permits the use of automatic or semiautomatic devices capable of delivering the titrant and continuously monitoring the signal obtained. The time elapsed between the start of the titration and the endpoint is known as the pseudoinduction period and is proportional to the inhibitor concentration. These automated procedures are quite rapid and reproducible.

## Reaction Types

Homogeneous catalytic reactions can be classified into two broad categories: (1) ordinary catalytic reactions and (2) Landolt reactions.

### Ordinary Catalytic Reactions

This group of reactions includes redox, chemiluminescence, dissociation, and complex-formation reactions.

**Redox reactions** Redox reactions are by far the most common indicator reactions and typically

involve such oxidants as hydrogen peroxide, oxygen (usually atmospheric),  $\text{NO}_3^-$ ,  $\text{BrO}_3^-$ ,  $\text{ClO}_3^-$ ,  $\text{IO}_3^-$ ,  $\text{IO}_4^-$ , or  $\text{S}_2\text{O}_8^{2-}$  or inorganic reductants such as tin(II), arsenic(III),  $\text{I}^-$ , or  $\text{S}_2\text{O}_3^{2-}$ , or organic reductants such as ascorbic acid, hydrazine, hydroxylamine, amines, phenols, and azo dyes. Reactions between members of both groups with an electron transition in s- or p-orbitals are usually slow and therefore especially appropriate for indicator reactions. The catalysts determined are usually metal ions that possess vacant d-orbitals, have different oxidation states, and/or are capable of forming coordination compounds with one of the ingredients of the indicator reaction. Specifically, they are transition metals of one of the following types: (1) metal(IV) (zirconium, thorium); (2) metal(V) (niobium, tantalum, vanadium); (3) metal(VI) (molybdenum, tungsten); (4) metal(II) and (III) ( $\text{Fe}^{3+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Cu}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Ni}^{2+}$ , etc.); and (5) platinum metals and related metals (platinum, osmium, lead, ruthenium, rhodium, iridium, silver).

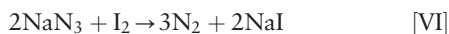
The most illustrative examples of redox indicator reactions are those involving the cerium(IV)/arsenic(III) system (Sandell–Kolthoff reaction), catalyzed by iodide ion; the iodine/sodium azide system, catalyzed by sulfur-containing compounds; the decomposition of hydrogen peroxide by copper, cobalt, manganese, and iron; the  $\text{H}_2\text{O}_2/\text{I}^-$ -system, catalyzed by iron, molybdenum, tungsten, zirconium, and hafnium; the hydrogen peroxide/hydroquinone system, catalyzed by copper; and the periodate/Malachite Green system, catalyzed by manganese. A number of automatic methods based on these systems have become the standards for determination of the corresponding catalysts in a wide variety of real samples.

The catalytic effect of iodide ion on the redox reaction between cerium(IV) and arsenic(III) was first shown and exploited for the determination of

this halide and its parent halogen at the nanogram per milliliter level by Sandell and Kolthoff. The reaction is carried out in a sulfuric acid medium, even though the catalytic activity of iodide is reportedly 20 times higher in nitric acid. The chloride ion inhibits the catalyzed reaction at concentrations of  $\sim 0.34 \text{ mol l}^{-1}$  or higher, whereas low concentrations of this halide increase the reaction rate and even expand the linear range of the calibration curve. The reaction can be monitored photometrically via cerium(IV), fluorimetrically through the cerium(III) produced, or electrochemically by using an amperometric or potentiometric sensor (including an iodide-selective electrode).

Not only iodide ion and iodine but also iodate and even periodate can be determined using this reaction by simply converting the analyte into the catalytically active species (i.e., iodide). The reduction is effected by arsenite ion itself on heating for 30 min. This treatment allows the resolution of iodate/iodide mixtures. The reaction can also be used for the determination of iodine-containing biological substances such as iodoproteins (thyroid hormones) by simply pretreating the samples in order to release iodine.

The reaction [VI] between sodium azide and iodine,



is catalyzed by substances containing sulfur in its  $-2$  oxidation state and has been used for the determination of anions such as sulfide, thiocyanate, and thiosulfate, as well as a variety of sulfur-containing organic compounds, with photometric or electrochemical monitoring.

Most often, these determinations rely on the oxidation of an organic compound, usually a dye, by hydrogen peroxide or an oxoanion such as bromate, iodate, or peroxodisulfate or even a chloramine derivative.

Both the cerium(IV)/arsenic(III) and the azide/iodine systems have also been used for the indirect determination of metal ions that interact with the catalyst, thus inhibiting the catalytic cycle, by use of a kinetic method or catalytic titration.

**Chemiluminescence reactions** Chemiluminescence reactions are usually redox reactions. The catalytic effect is reflected in the release of radiant energy. The best known of this type of reaction is the oxidation of luminol (5-amino-2,3-dihydrophthalazine-1,4-dione) by hydrogen peroxide, catalyzed by metal ions such as cobalt(II), copper(II), iron(II), nickel(II), chromium(III), and manganese(II) at pH 10–11. Detection limits of  $10^{-11}$  to  $10^{-8} \text{ mol l}^{-1}$  are readily achieved this way. Luminol is converted into a doubly charged anion that is subsequently oxidized to an excited singlet state that emits radiation on decomposing to

the aminophthalate ion. This reaction has been used for determination of the above-mentioned catalysts and hydrogen peroxide itself using copper(II) or cobalt(II) as the catalyst. Alternative oxidants include atmospheric oxygen and the peroxodisulfate ion – this latter is used for the determination of silver in the presence of amines. Though much less frequently used than luminol, lucigenin and lophine are also used as substrates in this type of reaction.

### Further Ordinary Catalytic Reactions

Kinetic catalytic methods based on catalytic decomposition, hydrolysis, ligand-exchange, or complex-formation reactions have a promising future as they allow the determination of nontransition metals such as alkaline earths (whether individually or in mixtures), in addition to ammonia and some other species.

### Landolt Reactions

Some catalyzed reactions involve an induction period, defined as the interval between addition of the last reactant (i.e., the start of the catalyzed reaction) and the appearance of a reaction product, during which the reaction appears not to develop. This phenomenon was first observed by Landolt in the reaction between iodate and sulfite ions in an acid medium, which releases iodine after an induction period. This is known as the Landolt effect and is chiefly exhibited by redox reactions involving halogens in various oxidation states. Thus, a slow reaction [VII] is coupled to a fast one [VIII] via the reaction product of the former:



Consequently,  $k_2 > k_1$  and P will only be detected after L (the Landolt reagent) has disappeared altogether in the second reaction. If reaction [VII] is accelerated by a catalyst, then the time elapsed until the product (P) appears will be a measure of the catalyst concentration, which is directly related to the induction period,  $t_1$ , by an empirical equation of the form  $[\text{C}]_0 = k_i/t_i$  or  $[\text{C}]_0 = k_i/t_i^2$ , depending on the particular system. Some slow reactions can be converted into Landolt reactions by using a retardant as the Landolt reagent.

This type of reaction is frequently used for analytical purposes on account of its simplicity and high sensitivity.

### Methods of Monitoring Reactions

Whenever a kinetic method is applied, the sample must be thoroughly mixed with the reagent(s) at a



constant temperature prior to transfer to the detection system. Mixing can be done manually (the usual situation with closed systems, which are applied to slow reactions such as the catalyzed reaction) or automatically. Automated mixing of the sample and reagents calls for special mechanical devices when fast reactions are involved as these require the use of open systems. However, open systems can be used in both catalytic titrations and kinetic catalytic determinations implemented in an automatic fashion for routine analyses.

For manual mixing it usually suffices to mix the reactants under continuous agitation in the reaction vessel, which may be a spectrophotometric or spectrofluorimetric cuvette or an electrode cell, or occasionally a thermometric cell. The time at which the last reactant is added (normally from a syringe) is taken as the start of the reaction ( $t=0$ ) and must therefore be accurately known. As a rule, the reactants and the cell compartment are thermostated prior to mixing. Alternatively, the sample and reagents can be mixed in a volumetric flask and made to the mark, and a portion of the diluted mixture be transferred to the detector cell.

The reactants can be incorporated into the measuring system automatically in order to increase the reproducibility of the above-mentioned closed or batch systems, equipped with photometric detectors, which are widely used in clinical laboratories.

Discrete stat and steady-state methods, whether automatic or semiautomatic, are frequently used for monitoring catalyzed reactions. Stat methods involve addition of a small amount of one of the reaction ingredients until a given value of the monitored parameter (pH, absorbance, luminescence, etc.) is reached. Any deviation from this state as a result of reaction development is immediately compensated for by the automatic addition of such an ingredient, the speed of signal restoration being proportional to the catalyst concentration. The procedure is referred to as pH-stat or absorptiostat method, according to whether the pH or absorbance is the continuously monitored parameter. Alternative techniques used to monitor the catalyzed reaction include biamprometry.

Steady-state methods entail keeping the speed of addition of a reagent constant throughout the process, thereby giving rise to a steady state in which the amount of reagent added is equal to that consumed in the reaction over a given interval. A plot of the steady-state concentration of the reagent as a function of the catalyst concentration can be used as a calibration graph for determining the catalyst.

Both stat and steady-state methods have been used for the kinetic determination of the iodide ion using its catalytic effect on the reaction between cerium(IV)

and arsenic(III) with photometric monitoring of the ceric ion added as the titrant at a variable (stat method) or constant (steady-state method) speed.

Although other commonly used open systems such as the stopped-flow methodology or the continuous-addition-of-reagent (CAR) technique are especially well suited to fast uncatalyzed reactions, they can also be applied to slow reactions.

## Detection Limits and Selectivity

Low limits of detection and high selectivity are important parameters in any analytical procedure. For a rough estimation of the minimum amount of catalyst that can be determined, eqn [4] may be transformed under an assumption of absence of the uncatalyzed reaction ( $k_1' = 0$ ) and  $[A] = [B] = 1 \text{ mol l}^{-1}$  to

$$\Delta[P] = k[C]_0 \Delta t \quad [6]$$

$$[C]_{0,\min} = (\Delta[P]/k\Delta t) \quad [7]$$

For photometric detection with

$$\Delta \text{absorbance} = \varepsilon_\lambda d \Delta[P] \quad [8]$$

it follows that

$$[C]_{0,\min} = (\Delta \text{absorbance} / \varepsilon_\lambda d k \Delta t) \quad [9]$$

Assuming a maximal molar absorbance,  $\varepsilon_\lambda$ , of  $10^5 \text{ (l mol}^{-1} \text{ cm}^{-1}\text{)}$ , a just measurable absorbance difference of 0.05, a reasonable optical length  $d$  of 5 cm, a practical time interval of 1 min, and a value of  $k' = 10^6 \text{ (cm}^{-1}\text{)}$ , the smallest determined catalyst concentration is calculated as  $10^{-13} \text{ mol l}^{-1}$ . The real limits of detection are some orders of magnitude higher than this theoretical value. The main reason is the influence of the background, especially by the extent of development of the uncatalyzed reaction. A comparison of the detection limits afforded by various catalytic methods in the determination of the most common ions reveals that the lowest achieved so far are in the  $10^{-9}$ – $10^{-7} \text{ mol l}^{-1}$  range and occasionally as low as  $10^{-10} \text{ mol l}^{-1}$  (kinetic determination of cobalt and vanadium, this latter in the presence of activators).

In contrast to the excellent detection power of the catalytic reaction, their selectivity is rather poor. Chemically similar species exhibit similar catalytic effects. This may be exemplified by the oxidation of different substrates by  $\text{H}_2\text{O}_2$  in acidic media (Table 2). In many cases a variation of the conditions in the indicator reaction will help to increase selectivity. Change of pH, reagent concentrations or temperature, addition of suitable complexing agents for activation,

inhibition or masking, and selection of appropriate detection techniques may greatly increase the selectivity of such reactions. If these procedures will not help, a preliminary separation step is necessary. Besides chromatographic methods, solvent extraction has proved to be effective, especially when the indicator reaction could be performed in the organic extract without the need for a stripping process (extraction-catalytic determination (Table 3).

## Analytes Determined

Determinations of major analytes using the catalytic reaction typically involve kinetic assays for metal ions as catalysts since for anions the number of possible indicator reactions is limited.

Primary catalytic effects are seldom exploited for the catalytic determination of organic compounds unless they can act as carriers for inorganic catalysts, as is the case with the determination of organic iodine-containing compounds using the Sandell-Kolthoff reaction. The most salient methods of this type for determination of metal ions, inorganic anions, and organic compounds are commented on in the following sections.

### Determination of Metal Ions

Table 4 shows some selected indicator reactions for the kinetic determination of transition metal ions (e.g., copper, cobalt, manganese, iron, vanadium) as

catalysts for various indicator reactions alongside their working conditions and analytical figures of merit. Note that photometric detection and the use of activators to enhance the sensitivity, in addition to the oxidation of organic compounds, are the most common choices in this context. Some ions (e.g., zinc, indium, gallium, selenium, tellurium, mercury) can also be determined kinetically using their activating effects on metal-catalyzed reactions or their inhibitory effects (mercury, gold, silver, copper, nickel, tin, magnesium) on nonmetal-catalyzed reactions (chiefly those involving the cerium(IV)/arsenic(III) system) or metal-catalyzed reactions. Depending on the particular metal and reaction, concentrations between a few nanograms per milliliter and a few micrograms per milliliter can be determined this way.

### Determination of Anions

Most kinetic determinations of anions involve the iodide ion, which exhibits a strong catalytic effect on the reaction between cerium(IV) and arsenic(III) and a few others as a result of the redox properties of the  $I_2/I^-$  couple. Other anions that can be determined using their intrinsic catalytic effect include sulfur-containing species such as sulfite, sulfide, and thiosulfate, which are quantified by means of the iodine/sodium azide system, and phosphates, which are measured through their effect on the formation of molybdenum blue. Table 5 gives illustrative examples of determinations for these anions and a few others.

Several anions such as cyanide, fluoride, and sulfide can also be determined kinetically at concentrations between 0.1 and  $4 \mu\text{g ml}^{-1}$  through their inhibitory effects on various metal-catalyzed reactions or, at somewhat higher concentrations, by direct catalytic titration.

### Determination of Organic Compounds

Most organic compounds of analytical interest in such areas as environmental or clinical chemistry

**Table 2** Indicator reactions with  $H_2O_2$  as oxidant

Indicator reaction	Catalyst
$H_2O_2 + I^-$	Ti, Zr, Hf, Th, V, Nb, Ta, Cr, Mo, W, Fe(III), $Cr_2O_7^{2-}$ , $PO_4^{3-}$
$H_2O_2 + S_2O_3^{2-}$	Ti, Zr, Hf, V, Nb, Ta, Mo, W
$H_2O_2 + PhOH$ , $PhNH_2$	Cr, Mn, Fe, Co, Ru, Os, Cu
$H_2O_2 + H_2(NHCS)_2$	Mo, W
$H_2O_2 + \text{organic dyes}$	Fe, Cu, Cr, Mn, Co, Ni

**Table 3** Extraction catalytic determinations

Analyte	Separation system	Indicator reaction	Observations
Ag	Phen/Bromopyrogallol Red (BPR)/nitrobenzene	$BPR + S_2O_8^{2-} + \text{phen}$ in a mixture of dioxane, nitrobenzene, and water	Extraction from EDTA-complexed solution
Mo	8-Hydroxyquinoline/ $CHCl_3$	1-Naphthylamine + $BrO_3^-$ in water/alcohol/ $CHCl_3$	Used for determination in seawater
Fe	Methyl isobutyl ketone(MIBK) + LiCl	<i>p</i> -Phenetidine + $H_2O_2$ (phen as activator)	Determination of iron traces in reagent grade salts of Co, Cu, Ni, Mg
$Cr_2O_7^{2-}$	MIBK + HCl	<i>o</i> -Dianisidine + $H_2O_2$	Determination in Cr(III) salts
Cu	$Cu(py)_2(Hsal)_2$ in $CHCl_3$	Sulfanilic acid + $H_2O_2$ (py as activator)	Determination in liver biopsy samples

phen, 1,10-phenanthroline; sal, salicylate; py, pyridine

**Table 4** Some kinetic catalytic methods for the determination of copper, cobalt, manganese, iron, and vanadium

Indicator reaction	Dynamic range or detection limit (ng ml <sup>-1</sup> or as otherwise stated)	Observations
<i>Copper</i>		
Decomposition of H <sub>2</sub> O <sub>2</sub> <sup>d,b</sup>	0.6–6 µg	In the presence of cyanide
Fe(CN) <sub>6</sub> <sup>3-</sup> + CN <sup>-a</sup>	< 4 × 10 <sup>-6</sup> mol l <sup>-1</sup>	Constant-rate method
IO <sub>4</sub> <sup>-</sup> + S <sub>2</sub> O <sub>3</sub> <sup>2-a</sup>	10–100	Automated procedure
Hydroquinone + H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	0.2–2	Pyridine as activator
Pyrocatechol Violet + H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	3–13	In 30% acetonitrile
<i>N</i> -phenyl- <i>p</i> -phenylenediamine + <i>N,N</i> -dimethylaniline + H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	0.4	Fixed-time method ( <i>t</i> = 15 min)
Bindschedler Green + H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	3.5–8	In an aqueous medium
	1–5	In a micellar medium (sodium dodecyl sulfate)
	0.8–6	In a micellar medium (dodecyltrimethylammonium bromide)
Ascorbic acid + S <sub>2</sub> O <sub>8</sub> <sup>2-</sup>	0.5–5	pH-stat method
<i>Cobalt</i>		
Decomposition of H <sub>2</sub> O <sub>2</sub> <sup>d</sup>	10	In the presence of 1,10-phenanthroline
1,4-Dihydroxyanthraquinone + H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	0.3	Borate buffer
Catechol + H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	3	Fixed-time method
Quinalizarin + H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	10 <sup>-8</sup> –10 <sup>-7</sup> mol l <sup>-1</sup>	
Gallic acid + O <sub>2</sub> <sup>c</sup>	< 2 µmol l <sup>-1</sup>	Oxygen electrode
Tiron + H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	3–120	Only Ni and Mn interfere
Hydrazine + H <sub>2</sub> O <sub>2</sub> <sup>d</sup>	2–1500	Interference by Mn and Pb
<i>Manganese</i>		
Decomposition of H <sub>2</sub> O <sub>2</sub> <sup>d</sup>	0.1–13	Only Fe(III) interferes at low concentrations
Malachite Green + IO <sub>4</sub> <sup>-a</sup>	0.1–1.0 × 10 <sup>-7</sup> mol l <sup>-1</sup>	In the presence of NTA
	0.02–2.0 µg ml <sup>-1</sup>	Absorptiostat technique
Pyrogallol Red + H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	0–32	2,2'-Bipyridyl as activator
<i>N,N</i> -diethylaniline + KIO <sub>4</sub> <sup>a</sup>	0.01–20	Flow-injection technique (15 samples per hour)
Hydroxynaphthol Blue + H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	0.01–10	
<i>o</i> -Dianisidine + H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	2 × 10 <sup>-8</sup> –2 × 10 <sup>-7</sup> mol l <sup>-1</sup>	Active species Mn(HCO <sub>3</sub> ) <sub>2</sub>
Salicylaldehyde + H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	5–100	
Rhodamine B + KIO <sub>4</sub> <sup>b</sup>	0.05–6.0	1,10-Phenanthroline as activator; pH 2.4
Tiron + H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	5–40	Flow-injection technique (40 samples per hour)
Oxidative coupling of 3-methyl-2-benzothiazolinone hydrazone with <i>N,N</i> -dimethylaniline by H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	2–30	1,10-Phenanthroline and citrate used as activators flow-injection technique (10 samples per hour)
<i>Iron</i>		
Decomposition of H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	50	1,10-Phenanthroline as activator
	1–13	Catalyzed by Fe(III)-triethylenetetramine
I <sup>-</sup> + H <sub>2</sub> O <sub>2</sub> <sup>c</sup>	5–160 µmol l <sup>-1</sup>	Iodide-selective electrode
Pyridoxal 2-pyridylhydrazone + H <sub>2</sub> O <sub>2</sub> <sup>b</sup>	5–60	
<i>N</i> -phenyl- <i>p</i> -phenylenediamine + <i>N,N</i> -dimethylaniline + H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	0–6	Acetate as activator
Sulfanilic acid + KIO <sub>4</sub>	11–56 <sup>a</sup>	1,10-Phenanthroline as activator
	50–220 <sup>d</sup>	Fixed-time method ( <i>t</i> = 10 min)
	5–160 µmol l <sup>-1</sup>	In a micellar medium (cetylpyridinium chloride) with 1,10-phenanthroline as activator
Methylene Blue + H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	0.04–12	1,10-Phenanthroline as activator
Rhodamine B + H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	0.04–150	Accelerated by potassium thiocyanate; fixed-time method ( <i>t</i> = 12 min)
Thymol Blue + H <sub>2</sub> O <sub>2</sub> <sup>a</sup>	0–3.2	2,2'-Bipyridyl as activator

Monitoring: <sup>a</sup>photometric, <sup>b</sup>fluorimetric, <sup>c</sup>potentiometric, <sup>d</sup>thermometric, <sup>e</sup>amperometric, <sup>f</sup>biamperostat, <sup>g</sup>absorptiostat, <sup>h</sup>potentiostat.

are determined on the basis of uncatalyzed reactions as they scarcely exhibit primary catalytic effects. Some organic substances such as thyroid hormones, cysteine, vitamin B<sub>1</sub>, and thiurea derivatives

have been determined because they contain atoms that can act as catalysts (e.g., iodine, sulfur) in the above-mentioned indicator reactions. Thus, vitamin B<sub>12</sub> is determined through its cobalt atom,

**Table 5** Some kinetic catalytic methods for the determination of iodide and other inorganic anions

Indicator reaction	Dynamic range or detection limit ( $\text{ng ml}^{-1}$ or as otherwise stated)	Observations
<i>Iodide</i>		
As(III) + Ce(IV)	4–20 <sup>a</sup> 2–100 <sup>a</sup> 2–80 <sup>b</sup> 4–500 <sup>b</sup> 0.2–100 <sup>c</sup> 3–25 $\mu\text{g}$ 0.01–10 $\mu\text{mol l}^{-1}$ 11–110 <sup>g</sup> 6–60 <sup>h</sup>	Wider range in the presence of $\text{Cl}^-$ Modular stopped-flow system Modular stopped-flow system Flow-injection technique Vitreous carbon electrode Electrogeneration of Ce(IV) (dead-stop method) Unsegmented-flow method Absorbance kept constant by addition of Ce(IV) Fixed potential kept constant by addition of Ce(IV)
As(III) + $\text{Fe}(\text{CN})_6^{3-}$ <sup>a</sup>	12–60	Data on mechanism
$\text{FeSCN}^{2+} + \text{NO}_2^-$ <sup>a</sup>	20–100	
$\text{BrO}_3^- + \text{Br}^-$ <sup>a</sup>	10	
3,3'-Dimethylnaphthylidene + $\text{H}_2\text{O}_2$ <sup>a</sup>	2–20	In the presence of formic acid
Catechol Violet + Chloramine B <sup>a</sup>	1.3	
Malachite Green + $\text{IO}_4^-$ <sup>a</sup>	10–200	EDTA as masking agent
<i>Other anions (in brackets)</i>		
Decomposition of $\text{BrO}_3^-$ ( $\text{Br}^-$ ) <sup>a</sup>	3	Autocatalysis
Decomposition of $\text{BrO}_3^-$ (thio anions) <sup>a</sup>	5–10	
<i>o</i> -Phenylenediamine + $\text{H}_2\text{O}_2$ ( $\text{Br}^-$ ) <sup>a</sup>	100–1000	
Pyrocatechol Violet + $\text{H}_2\text{O}_2$ ( $\text{Br}^-$ ) <sup>a</sup>	4–300 10–600	Fixed-time method ( $t = 2$ min) Flow-injection technique (45 samples per hour)
Fe(III) + $\text{ClO}_3^-$ ( $\text{Cl}^-$ ) <sup>a</sup>	100	Fixed-time method
Formation of sulfochlorophenol-Zr complex ( $\text{F}^-$ ) <sup>a</sup>	20–2000	Flow-injection technique
Pyridoxal 5-phosphate oxalyl hydrazone + $\text{O}_2$ ( $\text{CN}^-$ ) <sup>b</sup>	3–180	Only Hg interferes
Thionine + $\text{KBrO}_3$ ( $\text{NO}_2^-$ ) <sup>a</sup>	$7 \times 10^{-9}$ – $1.2 \times 10^{-6}$ $\text{mol l}^{-1}$	Fixed-time method ( $t = 5$ min)
Methyl Red + $\text{KBrO}_3$ ( $\text{NO}_2^-$ ) <sup>a</sup>	0.09	
Bromination of pyridine-2-carboxaldehyde 2-pyridyl hydrazone ( $\text{NO}_2^-$ ) <sup>b</sup>	7–120	Initial-rate method
	0–70	Induction period measurements
$\text{I}^- + \text{MoO}_4^{2-}$ (silicate) <sup>a</sup>	0.2–1.0 $\mu\text{mol l}^{-1}$	
$\text{I}^- + \text{MoO}_4^{2-}$ (phosphate)	30	
$\text{I}^- + \text{MoO}_4^{2-}$ (germanate)	100	
$\text{I}^- + \text{Cr}_2\text{O}_7^{2-}$ (oxalate) <sup>a</sup>	$10^{-6}$ – $10^{-5}$ $\text{mol l}^{-1}$	In the presence of starch
Naphthol Green B + $\text{BrO}_3^-$ ( $\text{NO}_2^-$ )		Simultaneous determination of nitrate and nitrite using a reductor column and flow-injection
Cr(III) + EDTA ( $\text{CO}_3^{2-}$ )	10–300	Flow-injection, photometric monitoring
Hydroquinone + $\text{H}_2\text{O}_2$ (phosphate)	1.9–4.75 $\mu\text{g ml}^{-1}$	
Azo dye + chromic acid (oxalate) <sup>a</sup>	500–1500	Data on mechanism

Monitoring: <sup>a</sup>photometric, <sup>b</sup>fluorimetric, <sup>c</sup>potentiometric, <sup>d</sup>thermometric, <sup>e</sup>amperometric, <sup>f</sup>biamperostat, <sup>g</sup>absorptiostat, <sup>h</sup>potentiostat.

which catalyzes various indicator reactions involving oxidation of organic dyes by hydrogen peroxide.

Aminopolycarboxylic acids (ethylenediaminetetraacetic acid (EDTA), 1,2-diaminocyclohexanetetraacetic acid, bis-(aminoethyl)glycoether-*N,N,N',N'*-tetraacetic acid, diethylenetriaminepentaacetic acid and nitroacetic acid) can be quantified through their inhibitory effects on metal-catalyzed reactions, whether by kinetic methods (at concentrations between  $10^{-6}$  and  $10^{-7}$   $\text{mol l}^{-1}$ ) or by direct catalytic titration

(at the microgram per milliliter level). Also, NTA has been determined as an activator for some manganese(II)-catalyzed reactions.

Other organic species are determined kinetically at the microgram per milliliter level through their activating (ascorbic acid, citric acid, bipyridyl, 1,10-phenanthroline, pyridine) or inhibitory (8-hydroxyquinoline, dimethylglyoxime, chromotropic acid, tetracyclines, amino acids, nitrogen-containing drugs) effects on various metal-catalyzed reactions at the microgram per milliliter level.

**Table 6** Some applications of kinetic catalytic methods in environmental chemistry

<i>Sample</i>	<i>Species</i>	<i>Observations</i>
Drinking water	Manganese	Compared with NAA Automatic photometric procedure Standard-addition method Fluorimetric detection <sup>a</sup> Segmented-flow system (30 per hour) Prior ion-exchange separation
	Vanadium	Mixture resolution
	Zinc, magnesium, copper, and nickel <sup>b</sup>	Determination range 0.01–2 mg l <sup>-1</sup>
	Iodide	Segmented-flow analyzer (20 per hour)
	Iodate, iodide	FIA (40 per hour)
	Fluoride	
	Sulfide <sup>c</sup>	
Tap water	Phosphate, silicate	Miniature centrifugal analyzer
	Copper	Compared with electrothermal AAS
	Cobalt	Removal of interfering Fe and Mg
	Nitrite	Stopped-flow (360 per hour)
	Fluoride	Content < 0.1 µg ml <sup>-1</sup>
	Chlorinated compounds <sup>c</sup>	Compared with colorimetry
	Hypochlorite <sup>c</sup>	Stopped-flow method
River water	Copper	Compared with electrothermal AAS Compared with liquid–liquid extraction
	Vanadium	Also applied to tap water
	Iron	Also applied to sea and tap water
	Fluoride	Interfered by sulfate and phosphate
	Iodate, iodide	Uses Technicon I AutoAnalyzer
	Carbaryl + 1-naphthol <sup>a</sup>	Stopped-flow method; also determined in vegetables
Seawater	Vanadium	Filtering + concentrated HCl
	Copper	No pretreatment
	Calcium + magnesium <sup>b</sup>	Stopped-flow mixture analysis
	Molybdenum	Preconcentration with chelating resin
	Chlorinated compounds <sup>c</sup>	Also applied to other types of water
Rainwater	Vanadium	Ion-exchange preconcentration (FIA)
	Iodide	Simulated samples, matrix
	Ammonia	1–55 ng ml <sup>-1</sup>
Wastewater	Mercury	Stagnant and electric power plant waters <sup>b</sup> Industrial effluents <sup>a</sup> Stopped-flow method
	Cadmium + manganese	Mining water
	Tungsten	Washing sewage
	Gold	Electrolytic plating sewage
	Cyanide	With and without cyanide distillation
	Bromide	Mine water
	Chlorinated compounds <sup>c</sup>	Analysis for ClO <sup>-</sup> and ClO <sub>2</sub> in disinfected water
	Nitrate <sup>d</sup>	Organic-free effluents
	Benzaldehyde, formaldehyde <sup>a</sup>	Used in the manufacture of synthetic fibers
	Hydrazine + phenylhydrazine <sup>a</sup>	Stopped-flow method
Environmental air	Manganese <sup>a</sup>	Sample collected on cellulose filter (0.8 µm)
	Sulfide <sup>c</sup>	Stopped-flow/microdistillation system
	Nitrogen oxides <sup>a</sup>	Determination of NO–NO <sub>2</sub> mixtures
	Mercury <sup>b</sup>	Sample collected in aqueous solution
	Carbon monoxide <sup>a</sup>	
Environmental dust	Vanadium	Prior ion-exchange separation
	Fluoride <sup>e</sup>	Isolation from matrix by microdiffusion
Soils	Carbofuran <sup>a</sup>	Stopped-flow method
	Iodide	Photometric determination

Type of reaction: <sup>a</sup>uncatalyzed, <sup>b</sup>ligand exchange, <sup>c</sup>chemiluminescence, <sup>d</sup>photochemical inhibition. AAS, atomic absorption spectrometry; FIA, flow injection analysis; NAA, neutron activation analysis.

## Selected Applications

The literature on kinetic methods of analysis places special emphasis on the major assets of these methods for

the analysis of real samples, particularly when an automatic procedure is used to facilitate routine analyses. This is a result of the high sensitivity of catalytic and chemiluminescence methods and the



**Table 7** Some applications of kinetic catalytic methods in clinical and pharmaceutical chemistry

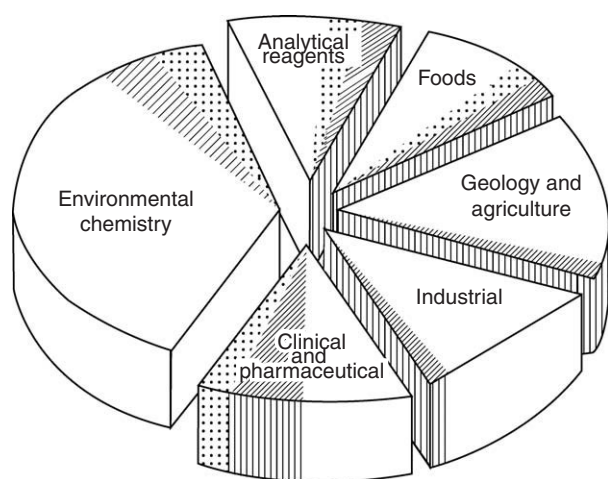
<i>Sample</i>	<i>Species</i>	<i>Observations</i>
Serum	Copper	Prior separation of albumin Automatic analyzer (20 per hour) Ashing + HNO <sub>3</sub> <sup>b</sup> Stopped-flow method CAR <sup>a</sup> CAR
	Copper + iron <sup>a</sup>	
	Cobalt	
	Molybdenum	Ashing and extraction of Mo
	Chloride <sup>b</sup>	Prior electrolysis
	Iodide	Prior mineralization Serum volumes of 12.3 $\mu$ l
	Carbon monoxide <sup>a</sup>	Range 0.05–20 mg per 100 ml
	Iodoproteins	Acidification + bromine
	Uric acid <sup>a</sup>	Stopped-flow method
	Uric and ascorbic acid <sup>b</sup>	Sensitivity increased through use of oxidoreductase
	Creatinine <sup>a</sup>	Ion-exchange separation (Dowex 50W-X8) Stopped-flow method
	Urea <sup>a</sup>	Stopped-flow method
	Copper <sup>a</sup>	Statistical study on the population
	Iodide	Dialysis at 37°C; AutoAnalyzer With and without pretreatment
Urine	Uric acid	Simple, selective, sensitive stopped-flow method
	Sulfonamides <sup>a</sup>	Stopped-flow method
	Morphine	Compared using liquid chromatography
	Epinephrine + norepinephrine <sup>a</sup>	Digestion with HNO <sub>3</sub> + Cu extraction Stopped-flow method
	Mercury <sup>a</sup>	CAR
Liver tissue	Copper	
Human hair	Selenium	Ashing at 600°C + HNO <sub>3</sub> + boiling to dryness + H <sub>2</sub> O
Pig viscera	Copper	Range 0.76–6.83 $\mu$ g ml <sup>-1</sup>
Rat brain	Dopamine <sup>a</sup>	Indirect determination
Pharmaceuticals	Bismuth <sup>b</sup>	Catalytic titration with potentiometric detection
	Mercury	Treatment with NaOH in methanol
	Nitroglycerine	Mean recovery, 100.5–102.6%
	Paracetamol	Fluorimetric detection Photometric stopped-flow method
	Thiamine <sup>a</sup>	By oxidation to thiochrome
	L-Thyroxine	Treatment with NaOH + filtration
	T <sub>3</sub> and T <sub>4</sub> <sup>c</sup>	Stopped-flow method
	Theophylline <sup>a</sup>	Stopped-flow method
	Procaine <sup>a</sup>	Stopped-flow method
	Chlorpromazine + perphenazine <sup>a</sup>	Stopped-flow method
	Epinephrine + norepinephrine <sup>a</sup>	Stopped-flow method
	Sulfonamides <sup>a</sup>	CAR
	Vitamin B <sub>12</sub> <sup>a</sup>	CAR
	Tetracyclines	Formation of anhydrides (0.8–4.0%)
	Catecholamines, sulfonamides, and barbiturates	Catalytic titration Excipient influences titration

Type of reaction: <sup>a</sup>uncatalyzed; <sup>b</sup>chemiluminescence; <sup>c</sup>T<sub>3</sub>, 3,3',5-triiodo-L-thyronine; T<sub>4</sub>, 3,3',5,5'-tetraiodo-L-thyronine (L-Thyroxine); CAR, continuous-addition-of-reagent technique; LC, liquid chromatography.

ability to resolve mixtures of closely related species using differential reaction rate methods. These advantageous features allow the determination of traces of species in environmental wastes and foods with no preconcentration, as well as in minute amounts of biological, forensic, or geological samples. In addition, the joint use of chromatographic techniques and kinetic catalytic methods is the most

effective approach to solving the problems arising from the low selectivity of some catalytic and chemiluminescence methods.

Although the applications of kinetic methods span many areas (industrial products, geochemistry and agricultural chemistry, food analysis, analytical-grade reagents), many of them are concerned with environmental, clinical, and pharmaceutical chemistry.



**Figure 5** Significance of kinetic methods in their different areas of application: catalytic reactions (clear areas), chemiluminescent reactions (dotted areas), and uncatalyzed reactions (shaded areas). (Reproduced with permission from Pérez-Bendito D and Silva M (1988) *Kinetic Methods in Analytical Chemistry*. Chichester: Ellis Horwood.)

### Environmental Chemistry

The features of environmental samples make them especially suitable for automated kinetic analyses. The determination of metal ions in environmental samples is of great ecological significance as these species cannot be degraded biologically or chemically. Kinetic methods are useful tools for determining heavy metals in environmental water and air, where toxic species occur at low concentration levels. **Table 6** collates some relevant applications, among which those concerned with water analyses are prominent.

### Clinical and Pharmaceutical Chemistry

Although kinetic methods of analysis are of special significance to clinical and pharmaceutical analysis in relation to enzymatic methods, they are equally applicable to nonenzymatic processes. **Table 7** shows the most salient contributions reported in this respect over the past few years. Most of the clinical determinations shown have been performed on blood and/or urine samples and have involved inorganic or organic ions in similar proportions. On the other hand, pharmaceutical determinations have chiefly been concerned with organic substances. The Sandell–Kolthoff reaction has by far been the most commonly employed for the analysis of biological samples, so much

so that some clinical handbooks have recommended it for the determination of thyroid hormones such as tri- or tetraiodothyronine ( $T_3$  or  $T_4$ ). **Figure 5** summarizes developments in kinetic methods for the analysis of real samples reported over the past 10 years in various areas of interest. Uncatalyzed reactions are included for comparison. As can be seen, catalyzed reactions are the most commonly used, followed by chemiluminescent and uncatalyzed reactions.

**See also:** **Air Analysis:** Sampling. **Blood and Plasma.** **Chemiluminescence:** Overview. **Enzymes:** Overview.

### Further Reading

- Crouch SR (1993) Trends in kinetic methods of analysis. *Analytica Chimica Acta* 283: 453–470.
- Crouch SR, Cullen TF, Scheeline A, and Kirkor ES (1998) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 70: 53R–106R.
- Crouch SR, Scheeline A, and Kirkor ES (2000) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 72: 53R–70R.
- Mottola HA (1988) *Kinetic Aspects of Analytical Chemistry*. New York: Wiley.
- Mottola HA and Pérez-Bendito D (1992) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 64: 407R–428R.
- Mottola HA and Pérez-Bendito D (1994) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 66: 131R–162R.
- Mottola HA and Pérez-Bendito D (1996) *Analytical Chemistry* 68: 257R–289R.
- Mottola HA, Pérez-Bendito D, and Mark HB Jr. (1990) Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry* 62: 441R–461R.
- Müller H (1995) Catalytic methods of analysis: characterization, classification and methodology. *Pure and Applied Chemistry* 67: 601–613.
- Pardue HL (1989) Kinetic aspects of analytical chemistry. *Analytica Chimica Acta* 216: 69–107.
- Pérez-Bendito D (1990) Approaches to differential reaction-rate methods. *Analyst* 115: 689–697.
- Pérez-Bendito D, Silva M, and Gómez-Hens A (1989) Automated kinetic-based determinations for routine analyses: recent developments. *Trends in Analytical Chemistry* 8: 302–308.
- Silva M (1993) Recent strategies in automated reaction rate based determinations. *Analyst* 118: 681–688.



## LAB-ON-A-CHIP TECHNOLOGIES

**U Bilitewski**, German Research Centre for Biotechnology (GBF), Braunschweig, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Concepts of ‘micro-total analysis systems’ or analytical ‘labs-on-chips’ stemmed from the experience that chemical analysis in liquids can be performed in capillary networks fabricated on planar substrates. The first systems utilized the technological developments of the microelectronic industries, i.e., silicon wafers were used as substrate materials, and capillaries were formed by etching. Flow of liquid was controlled by miniaturized pumps and valves. Further integration of functionalities was achieved by the replacement of hydrodynamic liquid transport by the electroosmotic control of liquid flow, as mechanical parts such as pumps and valves could be replaced by electrodes, which were integrated in the chips. Moreover, the utilization of electric fields for liquid transport led to the separation of mixtures of compounds due to different electrophoretic mobilities, and thus opened the possibilities for the performance of different analytical assays within those chips. However, as the separation of compounds required high electric field strengths, alternative substrate materials had to be investigated with a higher electric resistance than silicon. Thus, nowadays glass is frequently used as substrate material due to its favorable electric and optical properties and ease of structuring by wet etching. Nowadays, polymers are also investigated for their applicability as chip materials, for example, poly(methyl methacrylate) (PMMA), polycarbonate (PC), and poly(dimethylsiloxane) (PDMS).

As miniaturization proved to be advantageous with respect to analysis time and volumes of sample and reagents required, research with these technologies spread and at the end of the 1990s the first instrument was introduced to the market for capillary electrophoretic separations of double-stranded DNA fragments by Caliper Technologies Corp./Agilent Technologies.

The establishment of an analytical assay in a microsystem is the transfer of the assay in a capillary flow-through system, leading to the necessity to consider some particular aspects:

1. Due to the reduction of geometrical dimensions the surface-to-volume ratio is larger than in conventional systems and surface properties are more important. Thus, the chemical nature of the surface, surface pretreatment procedures, and the adsorption of components on the surface are of utmost importance.
2. Due to the small dimensions of capillaries the flow resistance for pressurized flow of liquid increases. Thus, alternative methods for liquid transport are used and at present the most important principle is electroosmosis requiring the integration of electrodes, usually platinum electrodes, and the application of electric fields. Additionally, centrifugal forces are used for propelling liquid through capillary systems. In some systems these principles are combined with pressurized liquid transport.
3. The reduction of dimensions also reduces volumes that are accessible for the detector. Thus, detection principles that are related to the volume of the detector cell, as absorbance measurements, are not ideally suited for coupling to microsystems, whereas surface sensitive principles, such as electrochemical methods or optical methods utilizing the evanescent field of a waveguide, or methods that can be focused on a small amount of liquid, such as fluorescence, are better suited. That is why fluorescence detectors are most often combined with microsystems, with an increasing number of reports appearing on the integration of electrochemical detectors.

### Principles

Labs-on-chips are planar capillary systems in which the capillaries are typically 10–50  $\mu\text{m}$  deep and 10–400  $\mu\text{m}$  wide. The channels are several centimeters long to allow mixing of components, the performance of (bio)chemical reactions, or the separation of compounds. A typical layout is shown

schematically in **Figure 1**, with a cross of capillaries being the simplest version, such as the part of the network formed by reservoirs 2, 3, 6, and 7. Reservoirs serve as interfaces to all required solutions. The design of the intersection of capillaries influences the volume of liquid introduced into reaction or separation channels. Often a double-T layout (insert in **Figure 1**) is used to increase volumes.

### Substrate Materials

Early reports were based on devices produced by etching and bonding of silicon, as these technologies were well established from the fabrication of micro-electronic devices. However, electroosmotic transport of liquid requires the application of electric fields (see below) and, thus, suitable substrates have to have a high electric resistance. Moreover, the combination with optical detectors is facilitated by glass as substrate material. Thus, etching processes and bonding procedures were also established for glass substrates.

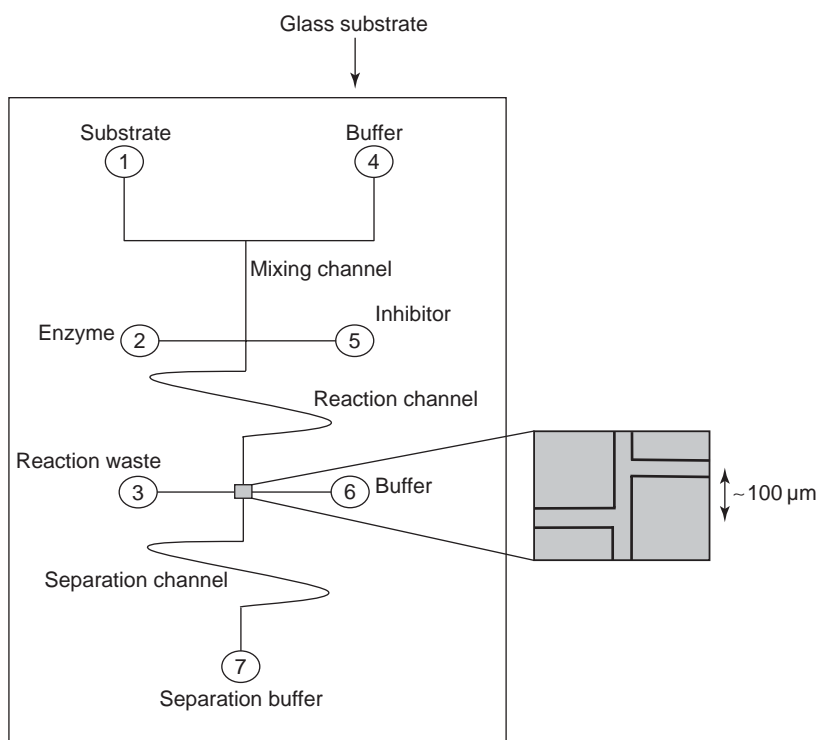
Polymers are discussed as less-expensive alternatives to glass, in particular with respect to mass production of structures. Injection molding, hot embossing, wire imprinting, and laser ablation are among the technologies that were established to

produce required capillary structures in PC, polystyrene, PMMA, poly(ethylene terephthalate), poly(vinyl chloride), or cellulose acetate. Major problems are related to bonding a cover plate to the structured substrate to close the capillaries tightly, without deforming capillary dimensions. Solutions range from adhesive tape lamination to gluing, and the application of heat or pressure.

Besides these rigid polymers, PDMS is widely used in recent studies, because structures are easily fabricated without the need for specialized equipment and PDMS shows good adhesion to a number of materials, e.g., glass, facilitating the tight closure of channels. Compared to glass, polymers are often more hydrophobic, which hinders an easy filling of structures with aqueous solutions and leads to an increased adsorption of hydrophobic compounds on capillary surfaces. Thus, fabrication conditions have to be chosen resulting in hydrophilic surfaces, or surfaces are treated after processing, such as by modification with charged polymers or by treatment with plasma or with  $\text{SO}_3$  vapor from fuming sulfuric acid (**Table 1**).

### Electrokinetic Control of Liquid Flow

**Principle** If an electric field is applied parallel to the long axis of a capillary, two migration mechanisms



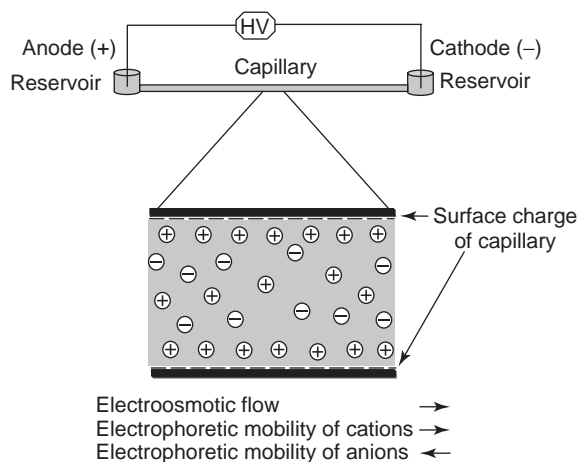
**Figure 1** Layouts of a microfluidic chip combining mixing and reaction zones (capillary system connecting reservoirs 1–5) with an integrated electrophoretic separation system (reservoirs 6 and 7).

**Table 1** Influence of the material and fabrication technology used for the production of microfluidic chips (labs-on-chips), and of surface pretreatments or additives to the running buffer on the electroosmotic mobilities

Material	Fabrication process	Pretreatment	$\mu_{osm}$ ( $10^{-4} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ )
Fused silica	Wet chemical etching		6.2–10.2
Borosilicate glass			2.2–4.2
		Hydrophob. silan.	2.0
PDMS	Molding	Protein adsorption	1.5
			0–6.4
		oxidation	4–8.2 (pH 9.0)
		Polybrene addition	– 4.29 to – 1.94
		Dextran sulfate addition	2.47–3.69
		MES addition	2
PET	Imprinting	SDS addition	4.5–6.5
			4.3
		Polyallylamine	– 1.8
PC	Laser ablation	Polystyrene sulfonate	4.2
	Molding		4.4–6.3
			0.7
	Hot embossing	UV irradiation	2.7
			7–8 (pH > 8)
	Laser ablation	SO <sub>3</sub> treatment	7–8 (pH > 6)
PMMA	Imprinting		2.8–5.3
		Tween addition	0.6
	Imprinting	SDS addition	2.8
			3.0
	LIGA		1.3–2.5
	Hot embossing		1.9
Co-polyester	Imprinting		1.8–2.2
		SDS addition	2.8
	Imprinting		4.3
			1.8–2.5
	Imprinting	Polyallylamine	– 1.3
		Polystyrene sulfonate	4.1
Cellulose acetate	Laser ablation		4.47
PVC	Laser ablation		4.74
Zeonor plastic	Hot embossing		3.53–5.24
Ceramic (LTCC)	Mechanical milling	Oxidation	1.1
			3–4.2

PDMS, poly(dimethylsiloxane); PET, poly(ethylene terephthalate); PC, polycarbonate; PMMA, poly(methyl methacrylate); PVC, poly(vinyl chloride); LTCC, low-temperature cofirable ceramics; MES, 2-morpholinoethanesulfonic acid; SDS, sodium dodecyl sulfate.

for compounds are induced: electrically charged molecules move within the electric field according to their charge, i.e., positively charged molecules migrate to the cathode, negatively charged molecules to the anode. This mobility is called the electrophoretic mobility of compounds and is dependent on the electric field strength, the viscosity of the solution, and the charge and size of the molecule. It is the reason for the separation of compounds on the basis of the charge-to-size (mass) ratio in electrophoresis. This mobility is superimposed on the electroosmotic mobility of the buffer. Within a capillary with a charged surface an electric double layer is formed by the ions present in solution, with oppositely charged ions from the solution being most close to the capillary surface (Figure 2).

**Figure 2** Schematic representation of the principles of capillary electrophoresis and electroosmotic flow.



This results in the so-called  $\zeta$  potential, which is dependent on the charge density on the capillary surface and the type and concentration of ions present in the solution. Application of an electric field perpendicular to this double layer, i.e., parallel to the surface or the long axis of the capillary, forces the ions next to the capillary surface to migrate according to their charge. Thus, the whole layer of ions neighboring the capillary surface moves in one direction, together with the corresponding shell of water molecules. By viscous forces this movement of the surface layer is transferred to adjacent solution layers so that the whole liquid within the capillary is moved toward the electrode charged with the same polarity as the capillary surface (Figure 2). This electroosmotic flow (EOF) rate  $\nu_{\text{osm}}$  is dependent not only on the electric field strength  $E$ , but also on the  $\zeta$  potential of the capillary and thus on the capillary material and treatment of the capillary surface. The Helmholtz–Smoluchowski equation relates these parameters according to

$$\nu_{\text{osm}} = -(\epsilon\epsilon_0\zeta/\eta)E = \mu_{\text{osm}} \times E = \mu_{\text{osm}} \times V/L$$

where  $\mu_{\text{osm}}$  is the electroosmotic mobility,  $L$  the length of the capillary,  $V$  the applied voltage,  $\epsilon$  and  $\eta$  the dielectric constant (78 for water) and viscosity of the medium, respectively, and  $\epsilon_0$  the permittivity of the vacuum ( $8.85 \times 10^{-12} \text{ C}^2 \text{ N}^{-1} \text{ m}^{-2}$ ).

$\nu_{\text{osm}}$  is the rate of liquid movement along the length of the capillary and is determined via the time required to transport solutions from the entrance of the capillary to the waste reservoir. In most materials the capillary surface is negatively charged so that the EOF is directed from the anode to the cathode.

**Influence of surface properties** As mentioned above, polymeric surfaces are usually more hydrophobic than glass surfaces due to the reduced density of polar or even ionized groups. Thus, the EOF is lower in most polymeric capillaries. The addition of ionic detergents to the running buffer not only allows a smooth filling of these capillaries with aqueous solutions, but also leads to the adsorption of detergents on the surface influencing the charge density on the surface, and thus the electroosmotic mobility (Table 1). Depending on the charge of the detergent, the EOF can even be reversed, which is marked by a negative sign in Table 1. Moreover, these modifications lead to a stabilization of the EOF in polymeric chips and reduce the unwanted adsorption of hydrophobic analytes on the capillary surface, thus improving the analytical properties (Table 2).

As the EOF is based on the presence of ionic groups on the surface of the capillary it is dependent on the pH of the solution, usually increasing with pH. Silanol groups on the surfaces of glass capillaries are minimally ionized at low pH, resulting in a low EOF; increasing the pH increases the density of negative charge on the capillary surface and thus the EOF. Depending on the  $\text{pK}_a$  of functional groups present on the capillary surface, the pH dependence is more or less pronounced in the pH range investigated and more or less shifted compared to glass capillaries.

### Injection of Samples

The resolution obtained in microfabricated separation systems and signals resulting from (bio)chemical transformations are limited not only by experimental conditions, such as composition of

**Table 2** Examples for the influence of additives to the running buffer on the analytical performance of microfluidic chips; abbreviations see Table 1

Additive	Chip material	Example of an effect of the additive
Hydroxypropylmethylcellulose	PMMA	Avoids adsorption of fluorescently labeled oligosaccharides to capillary surfaces
Methylcellulose		
Hydroxyethylcellulose		
Hydroxypropylmethylcellulose + mannitol, glucose, or glycerol		Improves separation of PCR products by low-viscosity sieving matrix
SDS	Glass	Allows micellar electrokinetic chromatography of peptides or proteins
Hydroxyethylcellulose		Allows seizing/separation of PCR products
Poly(vinyl pyrrolidone)	Pyrex glass	Improves separation efficiency of labeled amino acids and small peptides
Polybrene, dextran sulfate	PDMS	Avoids adsorption of neutral compounds to the capillary surface
MES	PDMS	Reduces the electroosmotic mobility and the adsorption of analytes, such as glucose or amino acids
Poly(vinyl alcohol)	Borofloat glass	Improves separation of labeled amines, allows separation of chiral amines

the running background electrolyte or electric fields applied during separation or reaction, but also by the sample volume, in particular by the chip layout. The volume of the injected sample is mainly determined by the channel dimensions and the offset between channels 3 and 6 (Figure 1), which typically ranges from 0 (no offset, simple cross) to  $\sim 250\ \mu\text{m}$ .

Additionally, the amount of injected sample can be influenced by the electric field strengths applied to sample and buffer reservoirs during injection. In the simplest injection mode first an electric field ( $> 100\ \text{V cm}^{-1}$ ) is applied between one of the upper reservoirs and reservoir 3 leading to filling of the capillary with sample, in particular of the intersection between the buffer capillary (6 and 7) and the capillary introducing the sample (grounded in reservoir 3). In the analysis step the electric field is switched to reservoirs 6 and 7, so that the sample is moved to the detector placed near buffer waste reservoir 7. For most substrate materials (negative surface charge) and analytes, the waste reservoirs (3 and 7) are electrically grounded and a positive voltage is applied to the sample or buffer reservoir, i.e., during analysis, injection respectively, the waste reservoirs are connected to the cathode, the sample and buffer reservoirs to the anode of the high-voltage supply.

With fluorescently labeled compounds and corresponding camera systems it was shown that in the intersection of the two capillaries, broadening of the sample plug during the injection step occurred due to diffusion of sample constituents into the long channel. During the analysis step sample still migrated into the long channel from the sample channels again by diffusion, but also by viscous effects. Both effects lead to less well defined, broadened sample plugs with reduced separation efficiencies of sample constituents. Thus, alternative injection regimes to shapen the sample plug by more complex control of electric fields were introduced. In the 'floating injection mode' the long channel is still floating during injection, whereas during analysis an electric field is generated in the sample channel so that buffer moves from the buffer reservoir 6 not only to the buffer waste reservoir 7, but also to sample reservoirs. This avoids leakage of sample into the long analysis channel during analysis (pushback effect). Diffusion of sample into the long channel during injection is suppressed by the 'pinched injection mode'. This requires application of an electric field to the long channel during injection so that not only a sample but also a buffer flow to the sample waste reservoir 3 is generated. This not only avoids the leakage of sample into the long channel during injection but

also reduces the amount of sample present in the intersection. Voltages to be applied to the different electrodes and the chronological sequence of the different steps have to be adapted to the respective chip layout and sample matrix.

## Detectors

Conventional chromatographic or capillary electrophoretic devices are based on analyte detection by absorbance, for example, of ultraviolet (UV) light. However, in chip-based systems absorbance measurements are the exception as detection cells have to be developed leading to an extension of the path-length of the transmitted light.

Thus, the detectors most often used in microchips are fluorescence detectors, in which laser light is focused in the channel to excite corresponding fluorescent dyes and the emitted light is collected by microscope objectives and detected by photomultiplier tubes. Though these detectors allow the detection of low numbers of molecules, the general application is hindered by the need for fluorescent compounds. To circumvent this problem, the principle of indirect fluorescence is used, in which a fluorophore is added to the background electrolyte and replaced by nonfluorescent analytes during analysis. Thus, these compounds are detected via the reduction of the high background signal.

Due to the ease of fabrication of miniaturized electrodes and the compatibility of corresponding technologies with fabrication technologies for microfluidic chips the combination of electrochemical detectors with microchips gains in interest. The major problem is the decoupling of the electrochemical detection from the electric field used for liquid control. This is achieved by positioning the detector electrodes close to the exit of the capillary so that the electric field generated by the high-voltage supply is dropped almost completely. Alternatively, electrodes are placed within the capillary requiring additional decouplers, which may be a small fracture in the capillary, or a palladium or platinum film placed in front of the working electrodes. However, though electrochemical detectors can be integrated in the chip already during fabrication, they are also connected to the capillary system via suitable chip holders allowing the applicability of electrode materials optimal for the selective and sensitive detection of the analytes and not limited by the fabrication technology.

There are reports on additional detectors connected to microfluidic devices the importance of which will increase with the robustness of interfaces. In particular with respect to applications in the life

sciences microchips were connected to mass spectrometry, and recently a microthermocouple was described allowing the determination of heat generated by a chemical reaction.

### World-to-Chip Interface

Though a number of steps of the analytical procedure can be performed automatically within the chip, samples and reagents have to be introduced into the chip; thus, the interface between the 'macroworld' and the microsystem has to be designed. The first chips comprised reservoirs at the ends of the capillaries, i.e., in the cover plate used for closing the capillaries, holes were drilled allowing access to the ends of the capillaries so that electrodes needed for electroosmotic fluid flow control could be placed from the top into these holes. The volumes were given by the diameter of the holes (typically 1–3 mm) and by the thickness of the respective cover plate (<1 mm) and were in the order of magnitude of microliters. Thus, liquid movement for several minutes may reduce the level in these reservoirs significantly, which causes pressure differences and thus additional hydrostatic effects. To increase the volumes of reservoirs and simplify placement of electrodes in first reports plastic pipette tips or tubings were glued to the substrates. However, this procedure is feasible only for single reservoirs and single chips and is replaced nowadays by chip holders with integrated reservoirs. Sample throughput is increased by parallelization of analysis channels and the application of multichannel pipettes. Those chips are used for

quality control, for example, of protein – or RNA – preparations, or in food or water analysis (Table 3).

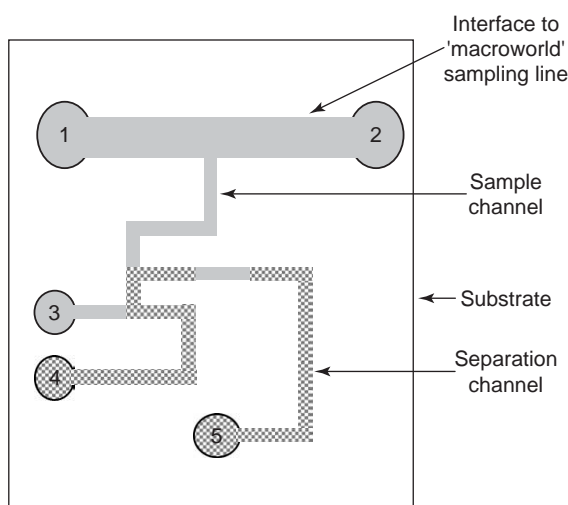
Labs-on-chips that are developed for drug screening need access to microtiter plates, in which the compounds of interest are stored. These chips contain sipper capillaries through which samples are introduced into the microchannels by vacuum. Continuous analysis of samples is possible if the chip contains a sample channel of larger dimensions (approximately in the millimeter scale), through which sample solutions are pumped hydrodynamically and which is connected to the separation or reaction channel. The principle is illustrated in Figure 3. To increase the impedance against pressure-induced flows the separation channel may be filled with a gel.

As illustrated by the examples given in Table 3, the application of labs-on-chips to real samples is still limited. This is partly due to the fact that the analytical assay is only the final step of the whole procedure, which includes sample pretreatment protocols such as filtration, analyte cleanup, or analyte preconcentration. However, also the integration of corresponding microfabricated elements is described. Filtration was achieved by porous membranes or arrays of thin channels preventing particulates to enter the analytical device. Analyte preconcentration in combination with removal of other sample constituents is achieved by solid-phase extraction modules, which are either capillaries or beads coated with a suitable adsorbent, such as a C18 phase originating from coating with octadecyltrimethoxysilane, from which the analyte is

**Table 3** Examples of the application of microcapillary electrophoresis to the determination of various groups of analytes in different samples

<i>Compounds</i>	<i>Chip material</i>	<i>Electrophoretic method</i>	<i>Detection principle</i>	<i>Sample matrix</i>
Inorganic anions	PMMA	ITP, ITP–CZE	Conductivity	Tap water
Organic acids, inorganic anions	PMMA	ITP	Conductivity	Wine, soups, drinks, sauces
Amino acids/peptides	PMMA	CZE	Conductivity	
Neurotransmitters	Glass	CZE	Amperometry (Pt electrode)	Standards
Carbohydrates	Glass	CE	Amperometry (Cu electrode)	
Amino acids	Glass	CE	Indirect fluorescence	Urine
Phenylenediamines	Glass	CE	Indirect fluorescence	Photographic developer
Amino acids/peptides	Glass	RP-CEC	Fluorescence	Standards
Proteins/peptides	PDMS	IEF	Fluorescence	
Polycyclic aromatic compounds	Glass	OCEC	Fluorescence	
Oligosaccharides	PMMA	CE	Fluorescence	
PCR products	PMMA	CE	Fluorescence	Serum
Proteins	Glass	CE	Fluorescence	Cell culture supernatants
Double-stranded DNA	Glass	CE	Fluorescence	DNA sequencing

ITP, isotachopheresis; CZE, capillary zone electrophoresis; RP-CEC, reversed-phase capillary electrochromatography; IEF, isoelectric focusing; OCEC, open channel electrochromatography.



**Figure 3** Chip layout comprising a larger channel for automated sample introduction via pressure-driven flow (connecting reservoirs 1 and 2), the width and depth of the sample channel are less than 10% of the sample introduction channel, reservoir 3 is the sample waste, reservoirs 4 and 5 are connected by the separation channel (reservoir 5: buffer waste).

eluted in a subsequent step. Another widespread method for analyte enrichment involves control of the electrophoretic velocities of analytes and includes field-amplified sample stacking (FASS) as the simplest method. In FASS, the sample is prepared in a low-conductivity buffer, whereas the running buffer is of higher conductivity. Thus, when voltage is applied the electric field strength in the sample zone is higher than in the running buffer zone, leading to a higher mobility of charged analytes in the sample zone. Reaching the interface to the running buffer analyte ions decelerate abruptly and stack in a narrow discrete sample band resulting in significant signal gains of up to three orders of magnitude.

Integration of the whole analytical procedure ranging from sample pretreatment procedures to the analytical signals leads to true analytical labs-on-chips.

## Assays

Microchips offer various possibilities for the performance of analytical assays: they comprise a capillary system, to which electric fields can be applied, which is the principle of capillary electrophoresis (CE). Thus, capillary electrophoretic separation assays can be transferred to microchips leading to the accelerated separation of compounds, in particular of carbohydrates, amino acids, peptides, proteins, and nucleic acids. In Table 3, some representative assays are summarized with special emphasis on the practical applications.

The broad range of possible analytes and sample matrices stems from the possibilities to connect the chip to different detectors, though fluorescence detectors are the predominant, and to use different separation modes (CE, isotachopheresis, reversed-phase electrochromatography) as different buffer additives, sieving matrices, capillary wall coatings, or capillary packings can be used. As mentioned above and summarized in Table 2, detergents and polymers added to the running background electrolyte interact with the capillary wall and prevent the adsorption of proteins to glass surfaces or of neutral analytes to polymeric surfaces. However, detergents form micelles, if they are used in concentrations above the critical micelle concentration, leading to the entrapment of analytes and their separation based on micellar electrokinetic chromatography and polymers, such as hydroxyethylcellulose, serve as sieving matrix allowing the separation of nucleic acids. In addition to these dynamic coating procedures, capillaries are permanently coated by silanization or filled with gels or beads. The interaction of compounds with these permanent coatings or fillings may lead to a strong adsorption of analytes within the capillary requiring the subsequent elution by appropriate solutions, such as acetonitrile gradients. This phenomenon leads, on the one hand, to the implementation of chromatographic methods in chips and, on the other hand, to the integration of solid-phase extraction procedures for analyte cleanup and preconcentration.

However, labs-on-chips are used not only in separation-based assays but also as flow-through devices allowing the performance of (bio)chemical reactions, such as enzymatic transformations or antigen–antibody interactions. Usually products of the reaction are quantitatively determined allowing the estimation of enzyme substrates, enzyme inhibitors, the respective kinetic and thermodynamic constants, or of antigens/antibodies. The assays can be performed in a homogeneous format or with immobilized biomolecules. The immobilization substrate can be the capillary surface leading to open-tubular enzyme or antibody reactors. The surface area, to which the immobilized proteins are restricted, is defined by channels through which solutions are pumped for a permanent coating of the surface with compounds introducing functional groups for covalent or adsorptive protein coupling. As an alternative approach, proteins can be immobilized on microbeads, which are entrapped within reaction chambers defined by weirs, which leave an open space too small to allow the beads to leave the system but allow solutions to pass. Due to the versatility of the layout of microchip capillary systems,

**Table 4** Examples for biochemical analytical assays performed in microfluidic chips

<i>Application</i>	<i>Biological element</i>	<i>Format</i>	<i>Detection</i>
Enzyme inhibitor screening	$\beta$ -Galactosidase $\beta$ -Glucuronidase	Mixing substrate, protein, and inhibitor	Fluorescence
Enzyme inhibitor screening	Acetylcholinesterase	Mixing substrate, protein, inhibitor; detection of product after further derivatization	Fluorescence
Enzyme inhibitor screening	Protein kinase A	Mixing fluorescently labeled peptide, protein, inhibitor, electrophoretic separation of phosphorylated product and nonphosphorylated substrate	Fluorescence
Enzyme inhibitor screening	$\beta$ -Galactosidase	Mixing inhibitor and substrate, enzyme immobilized on membrane	Fluorescence
Enzyme activity determination	Extracellular signal-regulated protein kinase	Mixing fluorescently labeled peptide, protein, inhibitor, electrophoretic separation of phosphorylated product and nonphosphorylated substrate	Fluorescence
Glucose determination	Glucose oxidase	Injection of sample in enzyme containing running electrolyte, electrophoretic separation from interferents	Amperometry (Au electrode)
Glucose determination	Glucose oxidase + horseradish peroxidase	Both enzymes co-immobilized via biotin–streptavidin on capillary wall	Fluorescence (Resorufin)
Sucrose	Invertase, glucose oxidase, soybean peroxidase	Enzymes covalently co-immobilized on capillary wall	Fluorescence (poly ( <i>p</i> -cresol))
Proteins	Trypsin	A reactor bed for immobilized trypsin in combination with a separation channel	Electrospray ionization mass spectroscopy
Trinitrotoluene	Specific antibodies	Affinity capillary electrophoresis by mixing antibody, analyte, and a labeled analyte analog, electrophoretic separation of antibody complex from noncomplexed tracer	Fluorescence
Mouse antibodies	Antimouse antibodies	Mixing of analyte with antibodies labeled with horseradish peroxidase, electrophoretic separation of the noncomplexed antibody, detection of label via mixing with enzyme substrates	Chemiluminescence
Interferon- $\gamma$	Specific antibodies	Sandwich immunoassay with capture antibody immobilized on beads, colloidal gold as tracer	Thermal lens microscope
Rabbit antibodies	Protein A	Competitive immunoassay with protein A immobilized on the walls of the capillary and Cy5-labeled antibodies as tracer	Fluorescence
Gene analysis	DNA polymerase	Amplification of gene sequences by PCR and electrophoretic separation of PCR products	Fluorescence
Drug screening	T lymphocytes (whole cells)	Mixing samples with cells, and additionally with membrane potential fluorescent dyes; detects ion channel activity	Fluorescence

electrophoretic separations can be combined with biochemical reactions (a possible layout is shown in **Figure 1**). This feature is utilized if signals from the educt and product of the reaction, or from sample constituents and the reaction product cannot be distinguished by the detector. Examples are the phosphorylation of fluorescently labeled peptides by kinase reactions, which does not change the fluorescence of the peptide, but its electrophoretic mobility

so that the phosphorylated form can be separated from the nonphosphorylated one, or the separation of polymerase chain reaction (PCR) products when the PCR is performed also on the chip. Some examples of biochemical assays performed in microchips are collected in **Table 4**.

*See also:* **Extraction:** Solid-Phase Extraction. **Immunoassays:** Overview. **Immunoassays, Techniques:** Enzyme



Immunoassays. **Polymerase Chain Reaction. Proteins:** Overview. **Quality Assurance:** Quality Control.

## Further Reading

- Andersson H and van den Berg A (2003) Microfluidic devices for cellomics: A review. *Sensors and Actuators B* 92: 315–325.
- Auroux PA, Iossifidis D, Reyes DR, and Manz A (2002) Micro total analysis systems. 2. Analytical standard operations and applications. *Analytical Chemistry* 74: 2637–2652.
- Becker H and Locascio LE (2002) Polymer microfluidic devices. *Talanta* 56: 267–287.
- Bilitewski U, Genrich M, Kadow S, and Mersal G (2003) Biochemical analysis with microfluidic systems. *Analytical and Bioanalytical Chemistry* 377: 556–569.
- Dolnik V, Liu S, and Jovanovich S (2000) Capillary electrophoresis on microchip. *Electrophoresis* 21: 41–54.
- Lacher NA, Garrison KE, Martin RS, and Lunte SM (2001) Review: Microchip capillary electrophoresis/electrochemistry. *Electrophoresis* 22: 2526–2536.
- Lichtenberg J, de Rooij NF, and Verpoorte E (2002) Sample pre-treatment on microfabricated devices. *Talanta* 56: 233–266.
- Manz A and Becker H (eds.) (1998) *Microsystem Technology in Chemistry and Life Sciences*. Topics in Current Chemistry, vol. 194. Berlin: Springer.
- McDonald JC, Duffy DC, Anderson JR, *et al.* (2000) Fabrication of microfluidic systems in poly(dimethylsiloxane). *Electrophoresis* 21: 27–40.
- Ng JMK, Gitlin I, Stroock AD, and Whitesides GM (2002) Components for integrated poly(dimethylsiloxane) microfluidic systems. *Electrophoresis* 23: 3461–3473.
- Reyes DR, Iossifidis D, Auroux PA, and Manz A (2002) Micro total analysis systems. 1. Introduction, theory, and technology. *Analytical Chemistry* 74: 2623–2636.
- Vandaveer WR, Pasas SA, Martin RS, and Lunte SM (2002) Recent developments in amperometric detection for microchip capillary electrophoresis. *Electrophoresis* 23: 3667–3677.
- Wang J (2002) Electrochemical detection for microscale analytical systems: a review. *Talanta* 56: 223–231.
- Weigl BH, Bardell RL, and Cabrera CR (2003) Lab-on-a-chip for drug development. *Advanced Drug Delivery Reviews* 55: 349–377.

# LASER-BASED TECHNIQUES

**A Quentmeier**, ISAS – Institute for Analytical Sciences, Dortmund, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

In analytical atomic spectrometry, lasers are employed either as sources of tunable radiation of precise wavelength or as sources of collimated, intense radiation. Laser sources that emit tunable radiation of low divergence in a well-defined spectral wavelength interval are the most powerful new techniques introduced in optical spectroscopy. They enable experiments to be carried out in basic atomic and molecular spectroscopies that were not possible using classical methods of optical spectroscopy. Such experiments include the evaluation of fundamental spectroscopic parameters, more precise determinations of spectral wavelengths of optical transitions, direct measurements on atomic and molecular beams, and Doppler-free high-resolution spectroscopy. Moreover, the new laser techniques improve the analytical capability of the classical spectrochemical methods, yielding much higher selectivity, larger dynamic range, and lower limits of detection.

In atomic laser spectroscopy, the laser radiation, which is tuned to a strong dipole transition of the atoms under investigation, penetrates the volume of species evaporated from the sample. The presence of analyte atoms can be measured by means of the specific interaction between atoms and laser photons, such as by absorption techniques (laser atomic absorption spectrometry, LAAS), by fluorescence detection (laser-induced fluorescence spectroscopy, LIFS), or by means of ionization products (electrons or ions) of the selectively excited analyte atoms after an appropriate ionization process (**Figures 1A and 1B**). Ionization can be achieved in different ways: (1) by interaction with an additional photon of the exciting laser or of a second laser (resonance ionization spectroscopy, RIS, or resonance ionization mass spectrometry, RIMS, respectively, if combined with a mass detection system); (2) by an electric field applied to the atomization volume (field-ionization laser spectroscopy, FILS); or (3) by collisional ionization by surrounding atoms (laser-enhanced ionization spectroscopy, LEIS).

Since its commercial availability, the laser has been used extensively as a source of powerful radiation. If the laser radiation is focused on to a solid sample, a small volume of the sample material can easily be

Immunoassays. **Polymerase Chain Reaction. Proteins:** Overview. **Quality Assurance:** Quality Control.

## Further Reading

- Andersson H and van den Berg A (2003) Microfluidic devices for cellomics: A review. *Sensors and Actuators B* 92: 315–325.
- Auroux PA, Iossifidis D, Reyes DR, and Manz A (2002) Micro total analysis systems. 2. Analytical standard operations and applications. *Analytical Chemistry* 74: 2637–2652.
- Becker H and Locascio LE (2002) Polymer microfluidic devices. *Talanta* 56: 267–287.
- Bilitewski U, Genrich M, Kadow S, and Mersal G (2003) Biochemical analysis with microfluidic systems. *Analytical and Bioanalytical Chemistry* 377: 556–569.
- Dolnik V, Liu S, and Jovanovich S (2000) Capillary electrophoresis on microchip. *Electrophoresis* 21: 41–54.
- Lacher NA, Garrison KE, Martin RS, and Lunte SM (2001) Review: Microchip capillary electrophoresis/electrochemistry. *Electrophoresis* 22: 2526–2536.
- Lichtenberg J, de Rooij NF, and Verpoorte E (2002) Sample pre-treatment on microfabricated devices. *Talanta* 56: 233–266.
- Manz A and Becker H (eds.) (1998) *Microsystem Technology in Chemistry and Life Sciences*. Topics in Current Chemistry, vol. 194. Berlin: Springer.
- McDonald JC, Duffy DC, Anderson JR, et al. (2000) Fabrication of microfluidic systems in poly(dimethylsiloxane). *Electrophoresis* 21: 27–40.
- Ng JMK, Gitlin I, Stroock AD, and Whitesides GM (2002) Components for integrated poly(dimethylsiloxane) microfluidic systems. *Electrophoresis* 23: 3461–3473.
- Reyes DR, Iossifidis D, Auroux PA, and Manz A (2002) Micro total analysis systems. 1. Introduction, theory, and technology. *Analytical Chemistry* 74: 2623–2636.
- Vandaveer WR, Pasas SA, Martin RS, and Lunte SM (2002) Recent developments in amperometric detection for microchip capillary electrophoresis. *Electrophoresis* 23: 3667–3677.
- Wang J (2002) Electrochemical detection for microscale analytical systems: a review. *Talanta* 56: 223–231.
- Weigl BH, Bardell RL, and Cabrera CR (2003) Lab-on-a-chip for drug development. *Advanced Drug Delivery Reviews* 55: 349–377.

# LASER-BASED TECHNIQUES

**A Quentmeier**, ISAS – Institute for Analytical Sciences, Dortmund, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

In analytical atomic spectrometry, lasers are employed either as sources of tunable radiation of precise wavelength or as sources of collimated, intense radiation. Laser sources that emit tunable radiation of low divergence in a well-defined spectral wavelength interval are the most powerful new techniques introduced in optical spectroscopy. They enable experiments to be carried out in basic atomic and molecular spectroscopies that were not possible using classical methods of optical spectroscopy. Such experiments include the evaluation of fundamental spectroscopic parameters, more precise determinations of spectral wavelengths of optical transitions, direct measurements on atomic and molecular beams, and Doppler-free high-resolution spectroscopy. Moreover, the new laser techniques improve the analytical capability of the classical spectrochemical methods, yielding much higher selectivity, larger dynamic range, and lower limits of detection.

In atomic laser spectroscopy, the laser radiation, which is tuned to a strong dipole transition of the atoms under investigation, penetrates the volume of species evaporated from the sample. The presence of analyte atoms can be measured by means of the specific interaction between atoms and laser photons, such as by absorption techniques (laser atomic absorption spectrometry, LAAS), by fluorescence detection (laser-induced fluorescence spectroscopy, LIFS), or by means of ionization products (electrons or ions) of the selectively excited analyte atoms after an appropriate ionization process (**Figures 1A and 1B**). Ionization can be achieved in different ways: (1) by interaction with an additional photon of the exciting laser or of a second laser (resonance ionization spectroscopy, RIS, or resonance ionization mass spectrometry, RIMS, respectively, if combined with a mass detection system); (2) by an electric field applied to the atomization volume (field-ionization laser spectroscopy, FILS); or (3) by collisional ionization by surrounding atoms (laser-enhanced ionization spectroscopy, LEIS).

Since its commercial availability, the laser has been used extensively as a source of powerful radiation. If the laser radiation is focused on to a solid sample, a small volume of the sample material can easily be

ablated into vapor-phase constituents such as atoms, ions, and particles. Because of that laser ablation (LA) represents one of the most promising micro-probe techniques of direct vapor generation and solid sample introduction in conjunction with a number of different analytical techniques. In combination with inductively coupled plasma (ICP), LA enables the

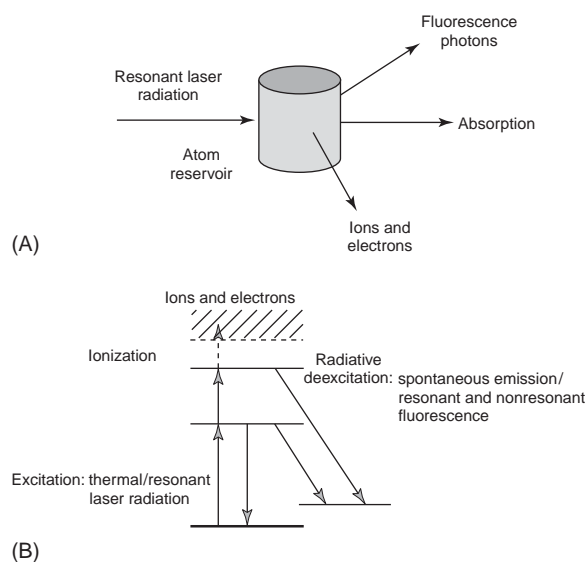
multielement analysis of solid samples by optical emission spectroscopic measurement (ICP-OES) or the ultimate trace analysis if coupled with mass spectrometric detection (ICP-MS). Ablated sample constituents that are excited in the laser plasma can be detected directly by emission spectroscopic techniques denoted as laser-induced breakdown spectroscopy (LIBS) or laser-induced plasma spectroscopy (LIPS).

The laser-based techniques commonly used in analytical atomic spectroscopy are summarized in Table 1.

The laser-based methods make use of different types of laser sources. Many of them, such as the solid-state laser or the semiconductor laser, are under continuous development because of the discovery of new lasing materials or the optimization of frequency conversion techniques. The type of the analytical application and the nature of the required information determine the choice of a special laser source, particularly pulsed or continuous-wave (CW) operating mode. Pulsed laser sources with fixed wavelength are usually used for LA, whereas both tunable pulsed and CW laser sources are preferred for element-specific detection or plasma diagnostics.

Some of the most frequently used laser sources in atomic spectroscopy and their typical parameters are listed in Table 2.

There is a whole range of other laser-based analytical techniques that are not within the scope of this



**Figure 1** Principles of laser-based atomic spectroscopic methods using resonant laser excitation: (A) scheme of different detection methods; (B) energy level diagram and transitions for excitation, ionization, and radiative deexcitation processes.

**Table 1** Laser-based techniques commonly used in analytical atomic spectroscopy

Ablation	Minute amounts of sample material ablated with the focused radiation of a pulsed laser are transported into an independent excitation source, e.g., inductively coupled plasma (ICP) for further atomization, excitation, or ionization. The detection of target atoms after laser ablation (LA) is performed by hyphenated techniques using optical emission or mass spectrometry LA-ICP-OES: laser ablation-ICP-optical emission spectroscopy LA-ICP-MS: laser ablation-ICP-mass spectrometry Ablated target constituents can be detected directly in the laser-generated plasma by their atomic and ionic emission LA-OES: laser ablation-optical emission spectroscopy LIBS: laser-induced breakdown spectroscopy LIPS: laser-induced plasma spectroscopy
Absorption	Tunable lasers (preferentially dye lasers and diode lasers) are used as primary sources for atomic absorption spectroscopy with various atomizers such as flames, furnaces, or plasmas LAAS: laser atomic absorption spectrometry CRS: cavity ring-down spectroscopy
Fluorescence	Tunable lasers are used to populate excited states of analyte atoms, which are detected by their radiative deexcitation LIFS: laser-induced fluorescence spectroscopy LEAFS: laser-excited atomic fluorescence spectroscopy
Ionization	One or several lasers are used to populate highly excited levels of analyte atoms, which are ionized by collisions, an electric field, or resonant photo ionization. The ions are detected by the charge current formed or by mass spectrometry LEIS: laser-enhanced ionization spectroscopy FILS: field ionization laser spectroscopy RIS: resonance ionization spectroscopy RIMS: resonance ionization mass spectrometry

**Table 2** Laser sources commonly used in atomic spectroscopy

Type of laser	Operating mode	Lasing medium	Wavelength range (nm)	Pulse width (typ.) (ns)
<i>Tunable laser</i>				
Diode laser	cw or pulsed	GaN	360–425	
		InGaN	420–570	
		InGaP	630–730	
		AlGaAs	720–860	
Dye laser	cw or pulsed <sup>a</sup>	Organic dyes	320–1000 218–380 <sup>b</sup>	< 10
<i>Solid-state laser</i>				
Ti:sapphire	cw	Al <sub>2</sub> O <sub>3</sub> :Ti <sup>+</sup>	720–950	
OPO-system	Pulsed	Y <sub>3</sub> Al <sub>5</sub> O <sub>12</sub> :Nd <sup>3+</sup>	410–710	< 10
<i>Ablating laser</i>				
Nd:YAG	Pulsed	Y <sub>3</sub> Al <sub>5</sub> O <sub>12</sub> :Nd <sup>3+</sup>	1064	< 10
			532; 355; 266 <sup>c</sup>	< 10
Ti:Sapphire	Pulsed	Al <sub>2</sub> O <sub>3</sub> :Ti <sup>+</sup>	790–810	< 100
Excimer	Pulsed	ArF	193	15–25
		KrF	248	15–25
		XeCl	308	15–25

<sup>a</sup> Depending on pump source (cw Ar<sup>+</sup> ion or pulsed Excimer/N<sub>2</sub>/Nd:YAG laser).

<sup>b</sup> Using second harmonic generation (SHG).

<sup>c</sup> Using frequency doubling, tripling, or quadrupling.

article, e.g., optogalvanic spectroscopy in flames or low-pressure discharges, and especially the numerous applications of molecular laser spectroscopy. The unsurpassed properties of commercially available lasers such as spectral brightness, small spectral linewidth, and variable pulse duration enable the improvement of existing methods and the development of new techniques of molecular spectroscopy not possible with conventional light sources.

The following examples of laser-based techniques and their relevant cross-references are summarized briefly for the sake of completeness. Some of the typical applications are:

- The absorption and fluorescence measurements of molecular gases with high spectral resolution by use of CW lasers or time-resolved spectroscopy with pulsed laser sources that enable the determination of the life time of short-lived excited molecular levels and radicals.
- The various techniques of Raman scattering that enable laser-based diagnostics of technical combustion processes as well as species identification on the micrometer scale or remote sensing of molecular species and pollutant in the atmosphere.
- Near-field optical microscopy in combination with Raman spectroscopy for the analysis of surfaces and interfaces and – because of the outstanding lateral resolution – for subcellular monitoring of molecular dynamics and single molecule detection.

- Photoacoustic spectroscopy and thermal lensing spectroscopy, which have proved to be very sensitive techniques for the detection of gaseous molecules.
- Matrix-assisted laser desorption and ionization process with mass spectrometric detection, which has been applied, for instance, for the identification of proteins and biological macromolecules.

An overview of important laser-based techniques in molecular spectroscopy is listed in Table 3.

## Radiation Sources for Laser Ablation

So far the most commonly used radiation sources for LA has been the Nd:YAG laser operating at the fundamental wavelength of 1064 nm. The advantages of this laser are good beam quality and a high degree of reliability. In principle, different types of lasers with a variety of wavelengths can be used for sampling, provided they deliver sufficient pulse energy. There has recently been a trend toward laser sources, which emit in the shorter wavelength region because of greater absorption of ultraviolet (UV) laser radiation by most solid materials and less absorption by the plasma formed above the sample surface. Both enable more efficient coupling of laser energy to the solid sample. Because of this the lasers most widely used nowadays for LA in spectrochemistry are the Nd:YAG lasers operated in the fourth harmonic (266 nm) and different types of excimer lasers such as ArF (197 nm), KrF (248 nm), or XeCl (308 nm).

**Table 3** Selected examples of laser-based techniques in molecular spectroscopy and typical applications

Absorption spectroscopy/ fluorescence spectroscopy	Detection techniques with high spectral resolution (tunable CW lasers). Most sensitive with multipass absorption cells and intracavity absorption. Measurements of chemical reaction diagnostics, lifetimes, and relaxation processes with high temporal resolution (pulsed lasers). Used also in combination with capillary electrophoresis and capillary chromatography
Raman spectroscopy	Makes use of various types of inelastic light-scattering phenomena such as spontaneous and stimulated Raman scattering induced by CW or pulsed laser sources. Diagnostics of combustion processes with high spatial and temporal resolution by coherent anti-Stokes Raman scattering (CARS). Structural analysis of polymers and biological compounds with high lateral resolution by laser Raman microanalysis (LRMA) or with lateral resolution below the diffraction limit in combination with scanning near-field optical microscopy (SNOM)
Remote sensing	Evaluation of various scattering processes (such as Mie, Rayleigh, Raman scattering, and resonance fluorescence) induced by the propagating of one or two pulsed laser beams. Mapping of trace constituents and pollutants in the atmosphere or measurement of meteorological parameters by LIDAR techniques (light detection and ranging)
Photoacoustic spectroscopy	Measurement of the photo- (or opto-) acoustic signal following the partial nonradiative dissipation of laser radiation absorbed by one or more compounds in a gaseous or liquid medium. Detection of ultralow concentrations of impurities. Used also in combination with gas and liquid chromatographies
Thermal lens spectroscopy	Evaluation of the transverse temperature profile ('the thermal lens') induced by laser radiation passing an absorbing (liquid) medium. Optical detection of organic compounds by the change of refractive index
Ionization spectroscopy	Dissociation and ionization by means of one or several lasers. Sensitive detection of fragments and ions with resonance photoionization and mass spectrometric techniques (tunable lasers). Matrix-assisted laser desorption/ionization mass spectrometry (MALDI) with pulsed ablating laser sources. High lateral resolution enables molecular microprobing of biological cell compounds

The main advantage of LA is the direct ablation capability from preselected areas of a sample, thereby avoiding laborious sample preparation and contamination. The focal spot size achieved with UV lasers of adequate (Gaussian) beam profile can be reduced to a few micrometers or less, which enables the microanalysis of solids. Quasicontinuous sampling is possible if the laser is operated with an adequate repetition rate.

## Tunable Laser Radiation Sources

If the laser is used as the primary excitation source the most important prerequisite is the tunability, which is the facility of preselecting the laser wavelength. To yield an appropriate tuning range, the lasing medium should have a wide gain profile. Soluble organic dyes such as rhodamine or coumarin are well suited for this purpose since they show strong absorption bands in the visible part of the spectrum, thus enabling optical pumping by means of flash lamps or other pump lasers. Solid-state lasers such as the optically pumped titanium-sapphire laser are also of increasing importance because of their broad tunability. The semiconductor diode lasers (DLs), on the other hand, which are commercially available nowadays, represent a fascinating excitation source combining compact size, reliability, low cost, and ease of handling. These qualities ensure

that such lasers are increasingly employed in the fields of plasma diagnostic and high-resolution spectroscopy.

As well as tunability, the radiation power delivered by the laser source is a characteristic parameter. In the case of high peak laser power, nonlinear optical effects can be utilized to expand the tuning range, predominantly into the short wavelength region. For this purpose one has to consider nonlinear optical processes, which are based on the interaction of high-intensity laser radiation with an appropriate optical medium. Well-known techniques in this respect are frequency doubling of the fundamental laser radiation, performed by means of second harmonic generation (SHG) in anisotropic crystals, Raman shifts in gases (like  $H_2$ ), or mixing techniques using two laser beams, which generate the sum or the difference frequencies of the exciting lasers. The efficiency of the nonlinear optical effects mentioned above varies quadratically with the laser intensity or with the product of the radiant power of the contributing lasers.

## Pulsed Dye Lasers

The properties of pulsed dye lasers are determined by the dye as well as by the nature of the pumping source. Common to all dyes is their broad amplification profile. By proper selection of dyes (which easily can be exchanged by a closed-loop design of



the dye flow circuit), the laser wavelength can be varied continuously within the range from  $\sim 350$  to  $1000$  nm. The spectral bandwidth of the laser radiation is usually confined (in commercial laser systems to  $\sim 1$  pm) by dispersive optical elements in the laser resonator, such as gratings and wavelength-selective transmission filters (etalons). This ensures the high selectivity that is essential for the excitation of atomic dipole transitions. Wavelengths below  $350$  nm are accessible by SHG, which is very effective because of the high peak power of  $\sim 1$ – $10$  MW, available in a single laser pulse. Frequency doubling techniques are restricted by the transmission of the nonlinear optical material to a shortest wavelength of  $\sim 210$  nm. In this way, most elements can be analyzed by making use of their most sensitive transitions.

Depending on the pumping source (either flash-lamps or other laser systems, such as excimer, Nd:YAG, nitrogen, or copper vapor lasers), the repetition rate of the dye laser can be varied between  $\sim 10$  Hz and as high as a megahertz. It is obvious that at low repetition rates only a small fraction of analyte atoms, which pass the interaction volume, is irradiated by the laser. This restricts the detection limit of the measurement, in particular if only small amounts of sample material are available, as in microanalysis.

### Tunable Solid-State Lasers

The pulsed dye lasers tend to be complex and have several shortcomings such as the use of toxic solvents and the relatively short lifetime of the dyes. For this reason these systems are gradually replaced by tunable solid-state laser systems. Their tuning range varies widely depending on the choice of the implanted ions and the host material. The tuning range of the titanium-sapphire laser ( $\text{Ti}^{3+}:\text{Al}_2\text{O}_3$ ), for instance, is typically  $720$ – $950$  nm. The output power, frequency stability, and linewidth of the titanium-sapphire laser are comparable with or superior to the dye laser. Repetition rates in the kilohertz range are available if pumped with a pulsed source such as a diode-pumped solid-state laser.

The wavelength region below  $720$  nm cannot be accessed directly by the titanium-sapphire laser but is covered by another laser system, e.g., the optical parametric oscillator (OPO). The OPO system consists of a birefringent crystal, e.g., BBO ( $\text{BaB}_2\text{O}_4$ ), placed within an optical cavity irradiated by an intense pump beam such as a Nd:YAG laser that is operated at its third harmonic ( $355$  nm). Through nonlinear interaction the pump beam is split into an idler beam and a signal beam, which generate a

signal beam tunable within the wavelength range of typically  $410$ – $710$  nm. This wavelength range can be extended significantly by additional frequency mixing techniques. To cover the same region with a dye laser, several dyes would be required.

### Continuous-Wave Dye Lasers

CW dye lasers are generally pumped by an argon ion laser. The tuning range of the CW dye lasers covers the wavelength interval  $260$ – $950$  nm, including frequency doubling, and is somewhat smaller than that of pulsed dye lasers. The most important feature of CW dye lasers in spectrochemistry is their very narrow spectral line width, which in the case of mono-mode operation and frequency-stabilized lasers is  $\sim 1$  fm (i.e.,  $10^{-15}$  m), and is thus smaller by three orders of magnitude than the linewidth of pulsed laser systems. Since the laser linewidth is even narrower than the Doppler profiles of spectral lines, this allows experiments of Doppler-free spectroscopy, which overcome the limitation of thermal broadening of spectral lines. The radiant power delivered by the CW dye lasers in mono-mode operation is typically less than  $1$  W. If one considers the laser linewidth, however, the spectral radiance is sufficient to excite optical transitions effectively. Another advantage of CW dye lasers is the long interaction time of analyte atoms in the radiation field, which yields improved detection limits.

### Semiconductor Diode Lasers

The disadvantage of all dye lasers is their complexity, their physical size, and their costs, which preclude the adoption of these systems for routine analysis. Semiconductor DLs, however, meet the requirements of practical spectrochemical analysis. They are small, reliable devices, which do not need to be aligned and have very stable power output. Produced on a large scale (as the scanning device in compact disc players), the costs of semiconductor DLs have been greatly reduced.

In contrast to the dye lasers mentioned above, lasing of the DL is accomplished easily by an electric current, which is sent in a forward direction through the p–n junction of the diode. The range of the laser wavelengths is determined by the semiconducting material, see Table 2, and comprises  $420$ – $570$  nm for InGaN,  $630$ – $730$  nm for InGaAlP/InGaP, and  $720$ – $860$  nm for AlGaAs/GaAs. Recently, the wavelength range has been further expanded by the development of GaN lasers, which emit wavelengths in the region  $360$ – $425$  nm. In general, wavelength tuning can be controlled by a computer and is performed by variation of the diode current as well as of the

temperature of the semiconductor. The DL can be tuned continuously within typically 0.05–0.1 nm. Tuning beyond this limit causes the laser to operate in another longitudinal mode ('mode hopping') with a further region of continuous tuning. The wavelength gaps, occurring between mode hopping, can be covered by additional DLs, which have been selected with regard to their spectral properties.

Because of the restricted spectral range covered by the fundamental wavelength of DLs commercially available nowadays, only a limited number of elements can be analyzed using their most sensitive resonance lines. Other elements of the periodic table, which have their most sensitive lines in the UV region, have to be excited by use of weaker transitions or by frequency doubling. This can easily be achieved yielding up to 1–10  $\mu\text{W}$  radiant power, whereas the radiant power of the fundamental wavelength emitted in the visible is in the range 50–100 mW.

The semiconductor DLs can be operated not only in the CW mode but also in a pulsed mode, e.g., as a pump source for solid-state lasers, or can be modulated at frequencies up to the gigahertz range by modulation of the diode current (see section on Laser-induced plasma spectrometry).

## Methods of Atomic Laser Spectrochemistry

Before spectroscopic techniques can be applied to elemental analysis, the sample has to be atomized effectively. Thus, the commonly used atomization sources such as flames, electrothermal atomizers (e.g., filaments or graphite furnaces), and ICPs are also prerequisites for the laser-based techniques of atom spectrochemistry. In addition, atomization by LA or electrothermal evaporation in low-pressure gas atmospheres should also be considered. The atomization source must be optimized with regard to efficiency, stability, uniformity, and chemical interference in order to obtain the best analytical results.

### Laser-Induced Plasma Spectrometry

The commercial laser sources used for LA ensure an irradiance typically  $>1 \text{ GW cm}^{-2}$ , far above the breakdown threshold, which is  $\sim 10^7 \text{ W cm}^{-2}$  for many materials. The nanogram to microgram amounts of sample material ablated with an individual laser pulse are effectively atomized, excited, and partially ionized in the hot plasma generated by the breakdown. Analyte atoms and ions can then be detected either by monitoring the spectral emission of the laser-induced plasma (LIPS, LIBS) or by the measurement of absorption, fluorescence, or

ionization signals that come into being if the laser radiation of properly selected wavelength intersects the plasma.

The plasma formation and excitation of sample atoms are complex processes, which are affected by the working parameters of the laser, the type of the sample material, and the ambient atmosphere. The mass removal and the plasma temperature, for example, have to be controlled carefully because they are affected by the sample composition. The calibration step requires, therefore, the use of matrix-matched standards, which are not always available. To correct for small variations of ablated mass and plasma temperature the signals from the analyte and a common matrix element are measured simultaneously applying the technique of the internal standard.

During the laser-surface interaction with lasers of nanosecond pulse width the material is removed mainly by an explosion-like evaporation from the overheated spot of the sample. The disadvantage of nanosecond lasers for accurate chemical analysis is the predominantly thermal character of the ablation process. As a consequence, fractionation may occur during the laser pulse depending on the sample and laser properties. Ideally, the entire laser-irradiated volume would be ablated without melting of adjacent sample regions and the ablated mass would represent the stoichiometric composition of the sample. With femtosecond pulse duration the deposition of the laser energy to the sample is so rapid that thermal effects can be clearly reduced or totally avoided. Femtosecond solid-state lasers with pulse duration of 100–200 fs are commercially available nowadays and are the basis for material processing on nanometer scale as well as for spectrochemical analysis with reduced matrix effects and improved lateral and even depth resolution.

**Applications** Since the laser plasma can excite the emission of all the elements ablated from the sample, LIBS has the potential for simultaneous multielement analysis. Typical applications are the rapid characterization of scrap and plastics prior to recycling, the minor or trace element detection for the bulk analysis of alloys, minerals, ceramics, or glass, the micro-sampling for pigment identification in artwork or geological inclusion chemistry, the compositional mapping of steel or of metal impurities in silicon wafers, and the *in situ* surface or depth analysis of technical coatings. If the laser radiation as well as the radiation from the plasma emission are transmitted by an optical fiber link, LIBS can be utilized for rapid noncontact sampling as in the online analysis of steel melt, for characterization of soils and ores with field instruments, or for remote sensing in hazardous

environments. Recently, the determination of  $^{239}\text{Pu}/^{240}\text{Pu}$  isotope ratios by LIBS was demonstrated using high-resolution emission spectrometry.

### Laser Atomic Absorption Spectrometry

Atomic absorption spectrometry (AAS) is a simple and reliable analytical method, but limited by a rather low dynamic range of about two orders of magnitude and by the fact that only one element can be analyzed at a time. These drawbacks can be overcome by LAAS if a tunable laser is used as the primary radiation source and if the hollow cathode lamp (HCL) is replaced by a DL. With regard to stability and spectral linewidth, the DL exceeds the specification of the HCL. Because of the high spectral radiance of the laser source, as compared to the background radiation from the atomizer (flame or graphite tube), no dispersive element such as a monochromator or a filter is necessary, and the collimated laser radiation (beam diameter typically 1 mm) is monitored by a low-cost semiconductor photodiode. The typical linewidth of the DL is  $\sim 30$  times less than the atomic linewidths that are observed in room temperature atom reservoirs (low-pressure plasmas, HCL) and two orders of magnitude less than the linewidths found in atmospheric pressure flames or graphite tubes. This allows the expansion of the linear dynamic range into higher concentration of an analyte if the laser wavelength is tuned precisely to the wing of the absorption profile, thus reducing the optical thickness, without changing to another less-sensitive absorption line. The wavelength of the DL can be easily modulated at frequencies up to gigahertz by modulation of the diode current. The wavelength modulation provides an effective correction of nonspecific absorption if the laser wavelength is switched from resonance to the background next to the absorption line and enables greatly improved limits of detection.

Simultaneous multielement LAAS can be performed if the radiation from several DLs is passed through the absorbing volume after alignment to a collinear beam by means of optical fibers or mirrors. The radiation of each laser is tuned separately to an absorption line and modulated in amplitude by an individual frequency. The absorption signals of different laser wavelengths are integrated at the common photodiode and evaluated by Fourier analysis.

**Applications** The potentials of DL-AAS have been demonstrated by measurement of trace metals in various atom reservoirs such as analytical flames and graphite tube furnaces. The limits of detection are typically two orders of magnitude lower than with

conventional AAS using HCLs. Absorbances of smaller than  $10^{-6}$  have been achieved with the double modulation approach when both the source and the absorption process are modulated, thus eliminating completely background signal and baseline drift. DL-AAS offers, moreover, a promising alternative as an element-selective detector for gas chromatography or liquid chromatography techniques. Elements that cannot be measured by conventional AAS with HCL such as hydrogen, oxygen, sulfur, noble gases, and halogens have been detected by coupling a gas chromatographic unit to a low-pressure plasma. DL-AAS can be utilized for elemental speciation as has been demonstrated by the detection of Cr(III) and Cr(VI) species in water using a commercial high-performance liquid chromatography unit. The limits of detection of  $0.5 \text{ ng ml}^{-1}$  observed for Cr(VI) are quite comparable with detection limits obtained with ICP-MS. Isotope selective analysis by DL-AAS of elements with isotope shifts larger than the Doppler broadening of the spectral lines has been demonstrated on the detection of isotopes of light and heavy elements such as Li, Pb, and U in low-pressure atomizers, where the pressure broadening is small.

### Laser-Induced Fluorescence Spectrometry

LIFS serves as a sensitive monitor for the selective absorption of laser photons as in LAAS. Since the LIF measurement reflects the population of the excited atomic state of analyte atoms it is well suited to gain information about competing transitions and collisional excitation or deexcitation processes. Compared with LAAS, LIFS techniques offer the advantage of a large concentration range of four to five orders of magnitude and low detection limits, particularly with CW tunable lasers. For weak radiation fields, such as in conventional sources, the fluorescence intensity is proportional to the source irradiance. In contrast to this, for a strong laser radiation field, the absorption of the laser radiation and consequently the fluorescence intensity is no longer linear. In this case the transition can be saturated by the exciting radiation field and the fluorescence intensity is entirely independent of the irradiance of the laser source. Consequently, fluctuations of the laser output do not affect the stability of the LIF signal. Moreover, radiative transition rates are usually some orders of magnitude greater than collisional depopulation rates due to quenching in a surrounding gas. This allows measurements to be made on atmospheric pressure sources, like flames, with good detection limits. Since saturation of the resonance transition can be attained most easily by pulsed laser sources, in this case the total number of fluorescence photons is

limited mainly by the laser pulse length and by the duty-cycle of the laser. If the measuring time is not restricted by the amount of available sample material, pulsed laser systems can also yield good analytical results if the fluorescence intensity is integrated over an appropriate number of single laser shots. In this case, however, it is essential to have a stable and reproducible atomization source.

Detection of analyte atoms with a more complex transition scheme requires the careful selection of the most favorable fluorescence transition, particularly if competing branching transitions are to be considered. Because of the high irradiance of the laser source scattering of the laser radiation on optical surfaces or on incompletely evaporated sample material deteriorates the detection limit of LIFS. The contribution of scattering can be neglected entirely if nonresonant fluorescence transitions are measured, which originate in branching transitions or in levels that have been populated by collisional transfer.

**Applications** The great potential of LIFS for diagnostics on spectrochemical excitation sources and atomizers, preferentially when using pulsed dye lasers, enable the measurement of various important spectroscopic quantities, such as transition lifetimes, collisional coefficients, and quenching rates. Tunable narrowband CW lasers, however, allow the determination of the profile and shift of spectral lines, thus yielding the excitation temperature, electron number density, and information about the different processes responsible for the observed line broadening. The analytical applications make use of the unsurpassed sensitivity of LIFS, especially in combination with electrothermal atomizers, such as the graphite tube furnace or the tungsten filament. The detection limits obtained with the analysis of metal traces in environmental samples or in biological liquids (e.g., Pb, Pt, Se in blood or serum) are in the nanogram per gram to picogram per gram range or even better. LIFS has also been applied to the determination of lithium and uranium isotopes in solid samples using a Nd:YAG laser for LA, and a pulsed tunable dye laser or a CW DL for excitation.

#### **Intracavity Absorption and Cavity Ring-Down Spectroscopy**

The sensitivity of AAS measurements can be increased significantly by increasing the absorption pathlength. A very effective multipass system can be achieved if the collimated radiation of a tunable laser is used as the primary light source. If the absorbing medium is placed inside the resonator cavity of the laser the sensitivity of the absorption measurement

can be enhanced considerably due to the multiple pass of the laser radiation within the resonator. For analytical purpose the intracavity absorption is measured by means of LIFS or by monitoring the output power of the laser, which is particularly advantageous, if the laser is operated near to the threshold level.

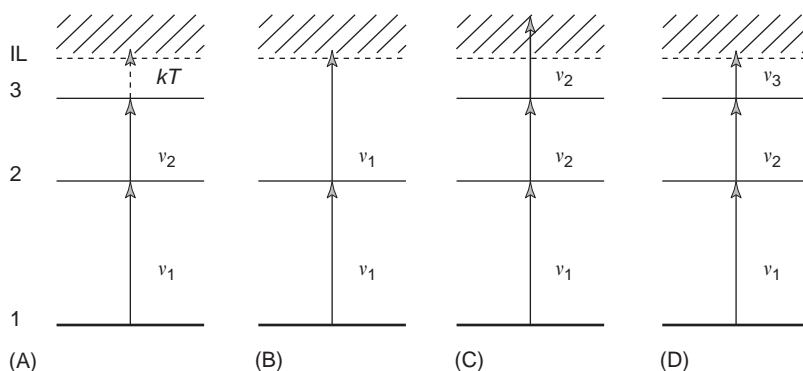
Cavity ring-down spectroscopy (CRS) is based on the measurement of the time constant for a laser pulse to decay inside an optical cavity consisting of two highly reflective dielectric mirrors. The laser pulse is injected through an end mirror of the cavity where it remains trapped between the mirror surfaces. The intensity of the laser radiation in the cavity as measured by a photomultiplier tube decays exponentially with time at a rate determined by the round trip losses experienced by the laser pulse. If absorbing sample atoms fill the optical cavity the decay time constant is reduced further because of the cavity losses originating from selective absorption of the sample following the Beer-Lambert law. Other sources of cavity losses like scattering can be determined easily by measuring the ring-down time with atomizer on and without analyte present.

**Applications** So far, intracavity laser spectroscopy has been applied primarily to the detection of absorption spectra of gaseous impurities such as  $\text{NH}_3$  and  $\text{CH}_4$  in the near-infrared region using a tunable broadband laser. Special DLs designed with an external cavity have also been investigated recently for this purpose. CRS has been applied successfully to trace element detection using the ICP as the atomization system. The detection limits observed are at sub-parts per billion level (e.g.,  $0.3 \text{ ng ml}^{-1}$  for lead) and comparable to the detection limits achieved with ICP-MS.

#### **Ionization Spectroscopy**

The strong radiation of laser sources enables the excitation of analyte atoms into high energetic levels just below the ionization limit. Ionization can then be accomplished by additional photons, by collisions, or by an external electric field (**Figure 2A**). In most cases the energy of one laser photon is not sufficient to ionize the ground state atom in a single step; ionization, therefore, has to be performed stepwise either by nonresonant multiphoton or by resonant ionization via a strong transition by use of a second or third tunable laser. The technique of stepwise resonance ionization by photons of different laser energy noticeably enhances the selectivity of the measurement by some orders of magnitude. This is essential if the ion contribution from matrix elements





**Figure 2** Excitation schemes of several laser ionization methods applying two- or three-step ionization with one, two, or three tunable lasers with different wavelength. In LEIS the final step of ionization is achieved by the thermal energy  $kT$  provided by collisional excitation (A), whereas in RIS the ionization is achieved with laser photons (B–D).

needs to be reduced, particularly in LEIS and RIS measurements without preceding matrix separation. Generally, ionization spectroscopy is one of the most sensitive detection techniques, provided that the ionization probability is high and the collection efficiency of the electrons or ions is nearly complete.

**Laser-enhanced ionization spectroscopy** LEIS makes use of the collisional ionization of analyte atoms that have been excited by resonant one- or two-step laser radiation close to the ionization limit. The ionization energy of the order  $kT$  can be raised by thermal collisions in chemical flames or in the surrounding buffer gas atmosphere. If the laser is tuned to a resonance transition of the atom the electrons and ions following the ionization step can be detected by two collecting electrodes, between which a voltage of  $\sim 1\text{--}2\text{ kV}$  is applied. In LEI spectrometry on flames the electron current is detected. This is because of the higher mobility of the electrons, which yield sharper analytical signals in the case of pulsed excitation. Different electrode configurations (plates, rods, water-cooled electrodes) have been tested.

**Field-ionization laser spectroscopy** The techniques of FILS are performed on collimated atom beams of an electrothermally evaporated sample. One or two pulsed laser beams traverse the beam of atoms, which are excited into high-lying Rydberg states by a one- or two-step excitation process. A high-voltage pulse, coinciding with the laser pulse, is applied to the electrodes, which are passed by the atom beam. If the field strength is properly chosen (depending on the effective quantum number  $n'$ ), a field gradient of  $\sim 400\text{ V cm}^{-1}$  is sufficient for complete ionization of analyte atoms with  $n' \gg 30$ . The ions are extracted from the atom beam by the same electric field pulse and measured by an ion detector.

**Resonance ionization spectroscopy** The ionization energy of most of the elements can be supplied only by two or more laser photons. To make use of the techniques of resonant ionization spectroscopy (RIS) ground-state atoms are excited via a dipole transition, which can be saturated easily. In the case of the single-laser excitation, an additional photon of the same laser can ionize the excited atoms (Figure 2B). High-intensity pulsed laser sources are needed for the efficient ionization, since the photoionization cross-section is less than the cross-section for the resonant absorption from the ground state by about five orders of magnitude. The single-laser resonance ionization process, however, may reduce the selectivity of analysis if off-resonant multiphoton excitation and ionization have to be considered. The transition rates for off-resonant two-photon ionization are rather small, but matrix atoms may contribute significantly because of their large number density, particularly in the case of trace analysis. To improve the selectivity, excitation processes with two or even three lasers with different wavelengths need to be considered. For this purpose the first excitation step is performed resonantly by a high-energy photon with moderate laser intensity. The succeeding excitation and ionization steps can be accomplished by low-energy photons of a second tunable laser with high radiant flux (Figure 2C). The selectivity can further be increased if the ionization is performed by a third laser (Figure 2D). This can be either a fixed wavelength laser with low photon energy, but high radiant flux (e.g.,  $10\text{ }\mu\text{m}$  of a  $\text{CO}_2$  laser) or a tunable laser of low intensity, if the wavelength is tuned to an autoionizing state. The third laser may have only moderate radiant flux, since the transfer rates for photoionization exceed the transfer rates for photoionization by two or three orders of magnitude.

For monitoring electrons or ions, RIS techniques make use of ionization chambers, ion multipliers, or



proportional counters as ionization detectors, which yield a nonselective analytical signal. The selectivity can be improved significantly by RIMS techniques, where a mass spectrometer is placed between the ion source and the ion detector. For this purpose a quadrupole mass filter can be used or a time-of-flight mass spectrometer, respectively, if the photoionization is performed by a pulsed laser. Spectrometers with high mass resolution, on the other hand, such as sector field instruments, offer the advantage of isotopic analysis.

**Applications** Although LEIS offers low detection limits, its application is somewhat restricted. This is due to quenching by molecular species in the flame or by matrix effects, which are introduced by accompanying elements with low ionization energy (such as alkali metals). Quenching can be avoided by use of an electrothermal atomizer such as a graphite tube furnace instead of a flame. Matrix interferences have been reduced by temperature programming of the electrothermal atomizer yielding a detection limit for lead in blood, for example, in the picogram per gram range. The coupling of LEI detection to flow-injection techniques has also been reported.

Because of their low absolute detection limits and their high selectivity, RIS and RIMS techniques have found important applications in the field of ultra-trace detection of metals in biological, environmental, and geological samples including single atom detection as has been proved with mercury atoms. The capability of the resonance ionization techniques was demonstrated on the isotope-selective detection of  $^{90}\text{Sr}$  in solid strontium samples combining the high sensitivity and high resolution of RIS with low-cost DLs and a low-resolution quadrupole mass filter. This technique was shown to be both spectroscopically and mass selective and by that advantageous over traditional ionization methods for atomic mass spectrometry since isobaric interferences, such as  $^{90}\text{Zr}$  or doubly ionized  $^{180}\text{Ta}$  and  $^{180}\text{W}$ , are reduced.

### Isotope-Selective Elemental Analysis by High-Resolution Laser Spectroscopy

The isotope-selective analysis by optical detection methods is almost impossible unless transitions with sufficiently large isotope shifts as observed with light and heavy elements are available. In contrast to traditional emission or absorption techniques the high-resolution laser spectroscopy enables Doppler-free measurements since the spectral linewidth of tunable CW lasers is commonly less than the Doppler profile

of the absorption lines introduced by the thermal velocity distribution of the atoms in the atomizer. A basic requirement for Doppler-free spectroscopy is the suppression of additional broadening effects, which implies that the analyte is atomized into vacuum or under low gas pressure. Typical applications are measurements on collimated atom beams or two-photon techniques, which are confined to an ensemble of atoms of uniform velocity.

#### Spectroscopy on Collimated Atom Beams

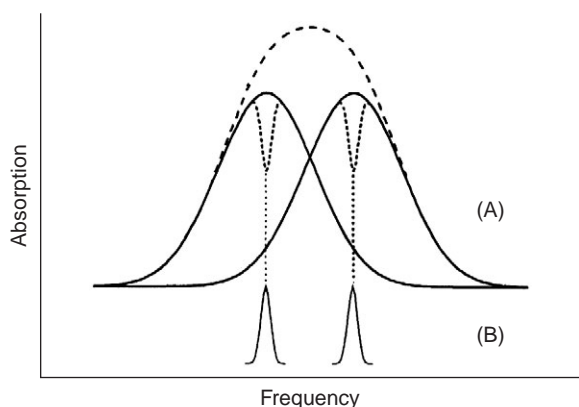
If a collimated atom beam expands into vacuum, the velocity component of the atoms in the beam direction is much greater than the thermal velocity component perpendicular to the beams axis, in accord with the low divergence of the beam. Consequently, the Doppler width of atom transitions, as seen by an intersecting laser field perpendicular to the axis of the atom beam, is reduced significantly. Isotopes, which have been excited selectively by the tunable laser, can then be detected by LIFS, by ionization techniques, such as RIS and FILS, or by RIMS after photoionization. LEI spectroscopy cannot be applied since collisional ionization does not occur in the atom beam.

#### Saturation Spectroscopy

Doppler-free profiles of spectral lines can be achieved by the method of saturation spectroscopy. The radiation of a tunable laser is split into two counter-propagating beams, which intersect a volume of absorbing atoms. The first strong pump beam saturates an optical resonance transition yielding the maximum population of the excited level, while the second weaker laser beam probes the remaining ground-state atoms. If the laser is tuned to the center of the resonance line, the absorbing medium is bleached and the probing beam cannot be absorbed. Tuning the probe laser across the absorption profile of the observed transition results in a narrow dip ('Lamb dip'), which has a width much smaller than the Doppler width. This enables the spectral resolution of transitions even when their Doppler profiles overlap (Figure 3). The method of saturation spectroscopy has been demonstrated on low-pressure gas discharges by LEIS techniques, where the change of plasma conductivity by the ionization products is detected.

#### Nonresonant Two-Photon Spectroscopy

Laser excitation of atoms can be achieved resonantly not only by one laser photon with appropriate energy, but also by nonresonant two-photon excitation, if the resulting photon energy equals the energy

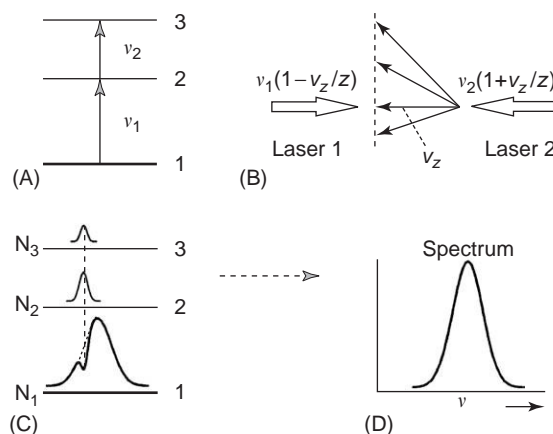


**Figure 3** Spectral separation of two Doppler-broadened profiles by saturation spectroscopy: (A) profiles of Doppler-broadened absorption lines with Lamb dips; (B) observed Doppler-free profiles (cross-over signals omitted).

of the dipole transition. The cross-section of nonresonant two-photon excitation is very low, as compared to resonance excitation. Since the transition rate is determined by the product of the intensities of the participating radiation fields, high-intensity lasers have to be used. If the laser radiation is focused, the interaction volume and thus the number of excited analyte atoms remains small, resulting in a poor detection limit. If only one laser source is available, the laser beam intersects the absorbing volume twice by means of a reflecting mirror. In this case atoms with a relative velocity component  $v_z$  in the direction of the laser beam, which see the laser frequency  $\nu_0$  Doppler-shifted by  $+\nu_0 \times (v_z/c)$ , can be further excited by a photon of the retrograding beam and the Doppler-shift is compensated by the corresponding component  $-\nu_0 \times (v_z/c)$ , and vice versa. This allows Doppler-free spectroscopy, since all atoms, which absorb photons from both directions, contribute to the analytical signal, irrespective of their velocity.

### Resonant Two-Photon Spectroscopy

In contrast to nonresonant two-photon excitation, the cross-section for resonant two-photon excitations is relative large if the atoms are excited via a strong resonance transition. To achieve resonant two-photon excitation, however, two tunable lasers are necessary with sufficiently narrow spectral bandwidths. The laser beams intersect the absorbing volume in co- or counter-propagating direction. If the first laser is tuned to the Doppler profile of the lower transition, atoms are excited with a well-defined velocity component in beam direction, whereas the second laser probes the population density of the excited atoms within this velocity group (**Figure 4**). The basic arrangement for isotope-selective analysis makes use of two absorption volumes, which are intersected by



**Figure 4** Principle of Doppler-free two-photon excitation: (A) resonant two-photon excitation steps; (B) frequencies of counter-propagating laser beams, as seen by atoms of the velocity component  $v_z$ ; (C) population densities of competing levels according to a well-defined class of velocity of analyte atoms; (D) resulting Doppler-free spectrum.

the laser beams simultaneously. In one absorption volume atoms with known isotopic composition are evaporated continuously. While the frequency of the first laser is locked to a line of an isotope, the second laser is tuned properly, yielding a Doppler-free spectrum of the isotopes in the second absorption volume. In this way isotope ratios can be determined optically if the first laser is switched in wavelength from isotope to isotope. The first atom reservoir is essential for fixing the laser wavelength, since the intensities of isotope components are strongly dependent on the frequency position of the first laser within the Doppler profile.

The spectroscopic methods described above have the disadvantage that not all excited atoms contribute to the analytical signal at the same time, but only those of a confined velocity group. This degrades the limit of detection; however, this can then be improved if efficient methods for ion detection are applied, such as LEIS combined with the thermionic diode. In the thermionic diode the ions are detected by means of the diode current, which is induced by charge effects with internal amplification by a factor of  $10^5$ – $10^7$ .

**Applications** Doppler-free high-resolution laser spectroscopy on low-pressure atom reservoirs and gas discharges allows the measurement of isotopic trace content and isotope ratios such as the determination of  $^7\text{Li}/^6\text{Li}$  isotope ratio, and thus offers the advantage of isotope dilution techniques for calibration.

**See also:** **Fluorescence:** Overview. **Mass Spectrometry:** Matrix-Assisted Laser Desorption/Ionization.

**Photoacoustic Spectroscopy. Raman Spectroscopy: Instrumentation. Remote Gas Sensing: Overview. Thermal Lensing Spectrometry.**

## Further Reading

- Bushaw BA and Cannon BD (1997) Diode laser based resonance ionization mass spectrometric measurement of strontium-90. *Spectrochimica Acta B* 52: 1839–1854.
- Butcher DJ, Zybin A, Bolshov MA, and Niemax K (2001) Diode laser atomic absorption spectrometry as a detector for metal speciation. *Review of Analytical Chemistry* 2: 79–100.
- Koch J, Miclea M, and Niemax K (1999) Analysis of chlorine in polymers by laser sampling and diode laser atomic absorption spectrometry. *Spectrochimica Acta B* 54: 1723–1735.
- Margetic V, Bolshov M, Stockhaus A, Niemax K, and Hergenröder R (2001) Depth profiling of multi-layer samples using laser ablation with femtosecond laser. *Journal of Analytical Atomic Spectrometry* 16: 616–621.
- Panne U, Neuhauser RE, Haisch C, Fink H, and Niessner R (2002) Remote analysis of a mineral melt by laser-induced plasma spectroscopy. *Applied Spectroscopy* 56: 375–380.
- Smith BW, Quentmeier A, Bolshov M, and Niemax K (1999) Measurement of uranium isotope ratios in solid samples using laser ablation and diode laser-excited atomic fluorescence spectrometry. *Spectrochimica Acta B* 54: 943–958.
- Smith CA, Martinez MA, Veirs DK, and Cremers DA (2002) Pu-239/Pu-240 isotope ratios determined using high-resolution emission spectroscopy in a laser-induced plasma. *Spectrochimica Acta B* 57: 929–937.
- Stchur P, Yang KX, Hou XD, Sun T, and Michel RG (2001) Laser excited atomic fluorescence spectrometry – A review. *Spectrochimica Acta B* 56: 1565–1592.
- Tognoni E, Palleschi V, Corsi M, and Cristoforetti G (2002) Quantitative micro-analysis by laser-induced breakdown spectroscopy: A review of the experimental approaches. *Spectrochimica Acta B* 57: 1115–1130.
- Wang CJ, Mazzotti FJ, Miller GP, and Winstead CB (2002) Cavity ringdown spectroscopy for diagnostic and analytical measurements in an inductively coupled plasma. *Applied Spectroscopy* 56: 386–397.
- Winefordner JD, Gornushkin IB, Pappas D, Matveev OI, and Smith BW (2000) Novel uses of lasers in atomic spectroscopy. *Journal of Analytical Atomic Spectrometry* 15: 1161–1189.
- Wizemann HD and Niemax K (2000) Measurement of Li-7/Li-6 isotope ratios by resonant Doppler-free two-photon diode laser atomic absorption spectroscopy in a low-pressure graphite furnace. *Spectrochimica Acta B* 55: 637–650.
- Zybin A, Schnürer-Patschan C, Bolshov MA, and Niemax K (1998) Elemental analysis by diode laser spectroscopy. *Trends in Analytical Chemistry* 17: 513–520.

# LASERS

See **ATOMIC MASS SPECTROMETRY: Laser Microprobe. LASER-BASED TECHNIQUES. SURFACE ANALYSIS: Laser Ionization**

# LEAD

**J Ruiz Encinar and M Moldovan**, CNRS UMR 5034, Pau, France

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Organolead compounds have been extensively used for many years as antiknock agents in gasoline. Although this use has been banned in developed countries, they are still in use in many developing countries. As a result, enormous amounts of these compounds have been, and still are, discharged into the atmosphere. Therefore, there is a significant

interest in understanding the origin, pathways, toxicity, and biological effects of organolead compounds in the environment. In order to properly identify organolead species, precise analytical methods capable of determining them at trace levels are necessary. The determination of individual organolead species requires the extraction and separation of the compounds, and their measurement with highly sensitive detection techniques that are able to perform analyses at the nanogram per liter level. During the last decades important advances were achieved in the development of hyphenated techniques for the analysis of organolead compounds, as well as in the production of certified reference materials for

**Photoacoustic Spectroscopy. Raman Spectroscopy: Instrumentation. Remote Gas Sensing: Overview. Thermal Lensing Spectrometry.**

## Further Reading

- Bushaw BA and Cannon BD (1997) Diode laser based resonance ionization mass spectrometric measurement of strontium-90. *Spectrochimica Acta B* 52: 1839–1854.
- Butcher DJ, Zybin A, Bolshov MA, and Niemax K (2001) Diode laser atomic absorption spectrometry as a detector for metal speciation. *Review of Analytical Chemistry* 2: 79–100.
- Koch J, Miclea M, and Niemax K (1999) Analysis of chlorine in polymers by laser sampling and diode laser atomic absorption spectrometry. *Spectrochimica Acta B* 54: 1723–1735.
- Margetic V, Bolshov M, Stockhaus A, Niemax K, and Hergenröder R (2001) Depth profiling of multi-layer samples using laser ablation with femtosecond laser. *Journal of Analytical Atomic Spectrometry* 16: 616–621.
- Panne U, Neuhauser RE, Haisch C, Fink H, and Niessner R (2002) Remote analysis of a mineral melt by laser-induced plasma spectroscopy. *Applied Spectroscopy* 56: 375–380.
- Smith BW, Quentmeier A, Bolshov M, and Niemax K (1999) Measurement of uranium isotope ratios in solid samples using laser ablation and diode laser-excited atomic fluorescence spectrometry. *Spectrochimica Acta B* 54: 943–958.
- Smith CA, Martinez MA, Veirs DK, and Cremers DA (2002) Pu-239/Pu-240 isotope ratios determined using high-resolution emission spectroscopy in a laser-induced plasma. *Spectrochimica Acta B* 57: 929–937.
- Stchur P, Yang KX, Hou XD, Sun T, and Michel RG (2001) Laser excited atomic fluorescence spectrometry – A review. *Spectrochimica Acta B* 56: 1565–1592.
- Tognoni E, Palleschi V, Corsi M, and Cristoforetti G (2002) Quantitative micro-analysis by laser-induced breakdown spectroscopy: A review of the experimental approaches. *Spectrochimica Acta B* 57: 1115–1130.
- Wang CJ, Mazzotti FJ, Miller GP, and Winstead CB (2002) Cavity ringdown spectroscopy for diagnostic and analytical measurements in an inductively coupled plasma. *Applied Spectroscopy* 56: 386–397.
- Winefordner JD, Gornushkin IB, Pappas D, Matveev OI, and Smith BW (2000) Novel uses of lasers in atomic spectroscopy. *Journal of Analytical Atomic Spectrometry* 15: 1161–1189.
- Wizemann HD and Niemax K (2000) Measurement of Li-7/Li-6 isotope ratios by resonant Doppler-free two-photon diode laser atomic absorption spectroscopy in a low-pressure graphite furnace. *Spectrochimica Acta B* 55: 637–650.
- Zybin A, Schnürer-Patschan C, Bolshov MA, and Niemax K (1998) Elemental analysis by diode laser spectroscopy. *Trends in Analytical Chemistry* 17: 513–520.

# LASERS

See **ATOMIC MASS SPECTROMETRY: Laser Microprobe. LASER-BASED TECHNIQUES. SURFACE ANALYSIS: Laser Ionization**

# LEAD

**J Ruiz Encinar and M Moldovan**, CNRS UMR 5034, Pau, France

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Organolead compounds have been extensively used for many years as antiknock agents in gasoline. Although this use has been banned in developed countries, they are still in use in many developing countries. As a result, enormous amounts of these compounds have been, and still are, discharged into the atmosphere. Therefore, there is a significant

interest in understanding the origin, pathways, toxicity, and biological effects of organolead compounds in the environment. In order to properly identify organolead species, precise analytical methods capable of determining them at trace levels are necessary. The determination of individual organolead species requires the extraction and separation of the compounds, and their measurement with highly sensitive detection techniques that are able to perform analyses at the nanogram per liter level. During the last decades important advances were achieved in the development of hyphenated techniques for the analysis of organolead compounds, as well as in the production of certified reference materials for

the validation of the results. Lately, lead isotope ratios have been identified as an important tool for the study of lead contamination studies.

## Organolead Pollution

### Sources and Uses of Organolead Compounds

The massive use of lead in kitchen utensils and water pipes goes back to the period of the Roman Empire. Later on, the industrial revolution promoted its use on a global scale. However, it was the discovery of the antiknocking properties of the organolead compounds and the fast development of the auto industry in the last century that turned lead contamination

into a global problem. Tetraethyllead (TEL), tetramethyllead (TML), and, sometimes mixtures of methylated and ethylated organolead species, are added to gasoline to prevent the premature and spontaneous explosion of the air/gasoline mixture in the engine. Table 1 shows the different organolead species that can be found in the environment. The release of TML and TEL was drastically reduced with the introduction of unleaded gasoline in the late 1970s in the USA and followed by other parts of the world. However, leaded gasoline is still available in many countries and contributes to the emission of TEL and to a less extent TML into the environment. Besides the anthropogenic sources cited below, there is limited evidence that under some circumstances, natural methylation of lead salts may also occur.

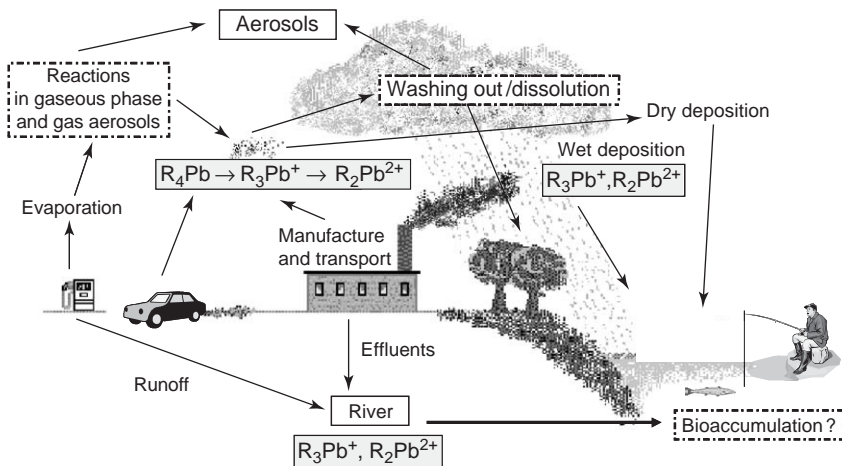
**Table 1** Organolead compounds found in the environment

Compound	Name (abbreviation)
<b>Tetraalkyllead (TAL)</b>	
Me <sub>4</sub> Pb	Tetramethyllead (TML)
Me <sub>3</sub> EtPb	Trimethylethyllead (TMEL)
Me <sub>2</sub> Et <sub>2</sub> Pb	Dimethyldiethyllead (DMDEL)
MeEt <sub>3</sub> Pb	Methyltriethyllead (METL)
Et <sub>4</sub> Pb	Tetraethyllead (TEL)
<b>Ionic alkyllead (IAL)</b>	
<i>Trialkyllead</i>	
Me <sub>3</sub> Pb <sup>+</sup>	Trimethyllead salt (TriML)
Me <sub>2</sub> EtPb <sup>+</sup>	Dimethylethyllead salt (DMEL)
MeEt <sub>2</sub> Pb <sup>+</sup>	Methyldiethyllead salt (MDEL)
Et <sub>3</sub> Pb <sup>+</sup>	Triethyllead salt (TriEL)
<i>Dialkyllead</i>	
Me <sub>2</sub> Pb <sup>2+</sup>	Dimethyllead salt (DML)
MeEtPb <sup>2+</sup>	Methylethyllead salt (MEL)
Et <sub>2</sub> Pb <sup>2+</sup>	Diethyllead salt (DEL)
<i>Monoalkyllead</i>	
MePb <sup>3+</sup>	Methyllead salt (ML)
EtPb <sup>3+</sup>	Ethyllead salt (EL)

### Emission, Transport, and Partitioning of Organolead Compounds

TML and TEL are introduced into the environment through car exhaust fumes (~1% of the organolead content of the gasoline is expelled unchanged), evaporation losses from gas tanks and stations, and accidental spilling. Whereas vehicle emissions led to a ubiquitous distribution of low concentrations of organolead compounds and in greater amounts of inorganic lead in the upper layers of soil, spillages of highly toxic tetraalkyllead (TAL) compounds during production, transportation, or blending at oil refineries and petrol stations caused more severe soil and groundwater contaminations. Figure 1 shows the biogeochemical cycle of the organolead compounds.

In the atmosphere, TAL is decomposed photocatalytically by ultraviolet irradiation, by ozone, or by hydroxyl radicals to the more stable, water-soluble ionic tri- and dialkylated species. Monoalkyllead cations are assumed to be very unstable. The



**Figure 1** Simplified biogeochemical cycle for organolead compounds in the environment.



dialkyllead species are thermodynamically more stable, and are therefore the main lead species found in air samples collected far away from the anthropogenic sources. The enrichment of organometallics over the inorganic lead as the air mass is moving away from the source can be explained on the basis of the species residence times. The organolead species ( $R_4Pb$  and  $R_nPb^{(4-n)+}$ ) present in the vapor phase are scavenged less efficiently from the atmosphere than the inorganic lead present in the aerosol.

### Toxicity of Alkyllead Species

Due to their more covalent nature, and thus higher lipophilicity, organometallic compounds usually accumulate more efficiently and they are more toxic than the inorganic form of the metal. The toxicity of alkyllead compounds decreases in the sequence:  $R_4Pb > R_3Pb^+ > R_2Pb^{2+}$  (with  $R = -CH_3$  or  $-C_2H_5$ ), the methylated lead species being less toxic than the ethylated compounds. Thus, even though their stability is not high, the final toxicological impact of the organolead species is very important as they are much more toxic than inorganic lead. Minimal bioaccumulation was observed for TEL and TML in shrimps, mussels, and plaice.

Organolead compounds can be absorbed through the skin and the respiratory tract. Their lipophilic nature allows them to easily go through biological membranes of the organism by passive diffusion. Oxidative dealkylation catalyzed by cytochrome P-450 in the liver results in the formation of trialkyl metabolites and finally inorganic lead. These trialkyl species are transported by the blood to the kidney, muscles, and brain, leading to a great variety of physiologic disorders. After some weeks, these metabolites finally settle in bones and teeth. The acute toxicity of TEL and TML is moderate in mammals and high in aquatic biota.

### Techniques for the Determination of Organolead Compounds

The concentrations of organolead compounds found in the different environmental compartments are often in the  $pg\ m^{-3}$ ,  $ng\ l^{-1}$ , and  $ng$  per  $g$  levels for air, water, and sediment and biological tissues, respectively. However, inorganic lead forms can simultaneously be present in the same samples at a 1000-fold higher level. For this reason, the instrumentation necessary to carry out lead speciation analysis in real samples requires the isolation, separation, and sensitive detection of the individual organolead species in the presence of thousands of

matrix components. This goal is usually achieved by coupling powerful chromatographic techniques with sensitive and specific detection systems based on atomic detectors.

### Detection Techniques

Atomic absorption spectrometry (AAS) was established as the most popular gas chromatography (GC) detection technique for lead speciation analysis in the first years of speciation studies. The increase of the residence time of the species in the flame using a ceramic tube inside the flame and, later, the use of electrically heated tubes, made out of graphite or quartz where electrothermal atomization was achieved, provided lower detection limits but still not sufficiently low. Later, the boom of plasma detectors, mainly microwave induced plasma atomic emission (MIP-AES) and, above all, inductively coupled plasma atomic emission and mass spectrometry (ICP-AES and ICP-MS, respectively) allowed the sensitivity requirements for reliable organolead speciation analysis in environmental and biological samples (typically subfemtogram levels) to be achieved. These sensitivity requirements makes speciation analysis of organolead compounds by molecular detection techniques such as electrospray mass spectrometry (ES-MS) a very difficult task and, therefore, the number of applications in the literature is very limited.

### Separation Techniques

Generally, GC is superior to liquid chromatography (LC) with regard to resolution, and very often in terms of separation time and sensitivity as well. This is especially true when an atomic detector is coupled to the chromatographic technique since LC requires a nebulization step whose efficiency is most often much less than 100%.

Prior to their separation by GC, organolead compounds need to be extracted from the sample matrix and, if necessary, derivatized into a volatile and thermally stable form. Volatile nonpolar tetraalkyl compounds can be extracted from environmental samples with nonpolar solvents such as benzene or hexane and injected directly onto the gas chromatograph. The case of nonvolatile, ionic alkyl compounds ( $R_2Pb^{2+}$  and  $R_3Pb^+$ ) is more complex. They should first be extracted and preconcentrated in the presence of NaCl and sodium diethyldithiocarbamate and then derivatized into a volatile form. The most widespread derivatization methods are of two distinct types: those based on the use of tetraalkylborates that can be directly performed in the aqueous solution (ethylation and propylation) and those applying

alkylation through Grignard reagents that requires nonaqueous media.

LC has not been so widely used for the determination of organolead species, although it does not require the derivatization of the analytes. The need for a nebulizer/spray chamber system usually restricts the detection limits achievable using standard LC. Almost all the methods use a reverse-phase column, and in some of the cases an ion-pairing reagent is added to the mobile phase. Several types of detection systems have been employed, such as AAS and ICP-AES, but the sensitivity of both approaches is disappointing. The coupling of LC with ICP-MS is the only means capable of detecting ultratrace levels of organolead species in typical environmental samples. The development of isotope dilution LC-ICP-MS has been successfully applied in the determination of TML and TEL in the presence of large amounts of inorganic lead in synthetic rainwater samples. An interesting alternative consists in the extraction of the organolead species in capillary tubing using solid-phase microextraction (in-tube SPME) and reversed-phased LC directly coupled to ES-MS allowing the detection of the molecular ions of TML and TEL (Table 2).

The use of supercritical-fluid chromatography (SFC) coupled to ICP-MS was applied for the separation and detection of TEL and tributyl lead acetate, although the apparent instability of tributyllead acetate under SF conditions and the presence of an additional peak when the mixture is injected does not favor its determination by SFC-ICP-MS. The use of SFC-ICP-MS for the quantification of TEL in SRM NIST 2715 'Lead in Reference Fuel' gave results in excellent agreement with the certified values.

In order to shorten the sample preparation steps, the use of a flow injection (FI) system for the continuous preconcentration of organolead compounds has been studied. It is based on a C<sub>60</sub> fullerene sorbent column and derivatization of the eluent analytes using a Grignard reagent prior to measurement by GC-MS.

Organolead speciation analysis has also been achieved by the use of modified silica capillaries following nebulization with a direct injection nebulizer (DIN), which is very suitable for the introduction of liquid samples into ICP-MS.

## Organolead Compounds in the Environment

It is estimated that ~1% of organolead in gasoline is emitted from vehicles in the form of TAL and ionic alkyllead. The presence of alkyllead compounds was confirmed in different types of environmental samples, such as air and rainwater, which contained organolead compounds in appreciable concentrations. The presence of organolead compounds in aquatic ecosystems, and especially in the marine environment, has not been studied extensively.

### Air

In the period 1970–90, considerable efforts were made to establish the fate of these compounds in air samples. At most urban sites, the typical contribution of organolead to total lead lies within the range 1–10%, with absolute concentrations in the range 10–200 ng Pb m<sup>-3</sup>, and a strong correlation with local gasoline could be observed in the composition of the atmospheric TAL. At roadside sites, the percentage of TAL to total lead is determined by the predominant driving mode; alongside motorways where vehicles drive at constant high speeds the TAL concentration can be from 5 to 10 times lower than at urban sites with interrupted traffic flows. The concentration of TAL decreases with an increase of distance from the road. In the vicinity of petrol (gasoline) stations a high percentage of TAL may be observed, mainly due to evaporative losses from spillage, and the displacement of TAL from fuel tanks. A high percentage of TAL has been found in enclosed car parks where vehicles are started from cold, with consequent inefficient fuel combustion and low engine temperatures. Measurements performed at rural sites are very scarce. For example, TAL concentrations were found to be below 0.1 ng (Pb) m<sup>-3</sup> in air samples collected from two rural sites near Beijing (China). In rural Denmark, TML concentrations were between 0.5 and 2.5 ng (Pb) m<sup>-3</sup>.

Organolead compounds can be associated with airborne particulate matter. A recent study determined the concentration of organolead in airborne particulate matter collected in Oviedo (Spain). The results show concentrations of 1.6–3.6 and 5.5–23.5 pg m<sup>-3</sup> DML and DEL, respectively.

**Table 2** Detection limits, expressed as ng (Pb) ml<sup>-1</sup>, for the determination of organolead species using different analytical techniques

Technique	TriML	DML	TriEL	DEL
GC-MS	0.6	1.2	0.4	0.8
SPME-GC-MS	0.13		0.08	
GC-MIP-AES	0.04	0.05	0.06	0.08
GC-ICP-MS	0.003	0.002	0.002	0.009
GC-ICP-TOF-MS	0.04	0.05	0.05	0.06
LC-ICP-MS	3		14	
SPME-LC-ESI-MS	11.3		12.6	

### Water, Snow, and Ice

Few measurements have been made of organolead compounds in natural waters. Apart from direct emission into waters, from fuel spillage or accident during transport, the tri- and dialkyllead decomposition products of TAL are quite soluble in water and so may be washed out by rain into surface waters. A survey of waters in Birmingham (UK) revealed concentrations of  $2 \mu\text{g(Pb)}\text{l}^{-1}$  of TAL in several road drainage grids. The analysis of runoff waters from the M6 motorway (northwest England) showed levels of TAL in the range  $0\text{--}36 \mu\text{g(Pb)}\text{l}^{-1}$ . Rainwater samples collected simultaneously at six sites in or near Antwerp (Belgium) were found to contain  $28\text{--}330 \text{ ng(Pb)}\text{l}^{-1}$  trialkyllead with an apparent correlation with local traffic density. Samples were also taken from surface waters and the trialkyllead concentration was found to be less than  $20 \text{ ng(Pb)}\text{l}^{-1}$ , except for road drainage waters from a highway ( $70\text{--}140 \text{ ng(Pb)}\text{l}^{-1}$ ). A lead speciation study in rainwater collected in central England shows that the total lead in rainwater has exhibited a strong decline since the 1980s because of the phasing out of gasoline lead, but the decline in organolead in rainwater is much less. A study performed during 1992–95 in Šibenik (Croatia) showed concentrations of  $37 \pm 28$  and  $2.7 \pm 2.3 \text{ ng l}^{-1}$  of organic lead in rainwater and seawater, respectively. These concentration levels correspond to 0.1% and 1.8% of total lead in rainwater and seawater, respectively.

Speciation analysis of organolead compounds was carried out in ice core samples collected in Greenland and on Mont Blanc. Organolead concentrations in the Greenland record ranged from 0.05 to  $0.5 \text{ pg per g}$ . The concentrations found show an increasing trend since the early 1970s until the beginning of the 1980s. In the Mont Blanc samples the measured concentrations ranged from 0.1 to  $3 \text{ pg per g}$  for DML, from 0.08 to  $3.4 \text{ pg per g}$  for TriML, from 0.01 to  $0.57 \text{ pg per g}$  DEL, and from 0.01 to  $0.13 \text{ pg per g}$  TriEL. No organolead concentrations were detected in ice deposited before 1962.

### Dust, Soil, and Sediments

Levels of tetra-, tri-, and dialkyllead were measured in street dust samples collected in 1976 in Lancaster (UK), and the concentrations ranged between 0.4 and  $7.4 \mu\text{g Pb g}^{-1}$ . With regards to the study carried out in Šibenik (Croatia), in contrast to the total lead ( $19.8 \pm 17 \text{ ng per g}$ ), organolead compounds were not found to be accumulated in sediments ( $0.2 \pm 0.2 \text{ ng per g}$ ).

Four organolead species (DEL, TriEL, DML, and TriML) were determined in a peat bog collected in

the Jura Mountains (Switzerland). The first quantifiable occurrence of organolead in this peat is from year 1943, and this agrees with the introduction of leaded gasoline in Switzerland.

### Biological Samples

In an extensive survey of fish from various lakes and rivers in Ontario (Canada) in the late 1970s, TAL compounds at levels of  $1\text{--}10 \text{ ng(Pb)}\text{g}^{-1}$  were found in  $\sim 16\%$  of the samples. All five TAL compounds were found. Samples of water, vegetation, algae, weeds, and sediments collected during the same survey did not contain TAL above the detection limits ( $0.1\text{--}0.5 \text{ ng(Pb)}\text{g}^{-1}$ ). A high concentration of trialkyllead was found in birds found dead or dying on the Mersey estuary (UK) in 1979. The analysis of the birds' invertebrate prey, bivalve baltic tellin and polychaete worms, also showed increased concentrations of trialkyllead species. Mussels and fish collected in Šibenik (Croatia) revealed a concentration of organolead compounds of  $6.1 \pm 12.1$  and  $0.9 \pm 0.9 \text{ ng per g}$ , respectively.

In mammals, inhalation or adsorption of TAL compounds results in the formation of trialkyllead in tissues and body fluids. With the exception of methylleads found in the blood of petrol workers, they are usually below the detection limits in blood and urine samples.

## Assessment of the Origin of the Organolead Pollution

### Record of Organolead Compounds in Environmental Archives

Lead released to the atmosphere consists of a complex mixture containing inorganic lead (natural and anthropogenic sources) and organic lead used in leaded gasoline. Thus, the determination of the total lead concentration is likely to be too simple for a reliable and definitive identification of the polluting source, since it is clear that the use of TAL in gasoline brought about an increase in lead industrial activities as well. Therefore, the determination of the organolead species seems to be the only valid way for assessing the contribution of gasoline to global lead pollution. In this way, the analysis of organolead species recorded by Greenland ice, high-altitude alpine snow, Swiss peat bogs, and old wine vintages allowed tracing the introduction, growth, and decline of the use of antiknocking agents in gasoline and their influence on the overall global and local environmental geochemistry of lead. In fact, no detectable concentrations for these compounds in samples predating the introduction of leaded gasoline

strongly suggest that natural sources of these species are negligible.

Moreover, the use of the configuration of the different organolead species in the sample (referred to as speciation pattern) may be used to obtain valuable information of the sample. For example, the DEL/TriEL and  $Pb_{org}/Pb_{inorg}$  ratios could help in assessing the age of the air masses transporting the species, the distance away from emission sources, and time spent over the ocean.

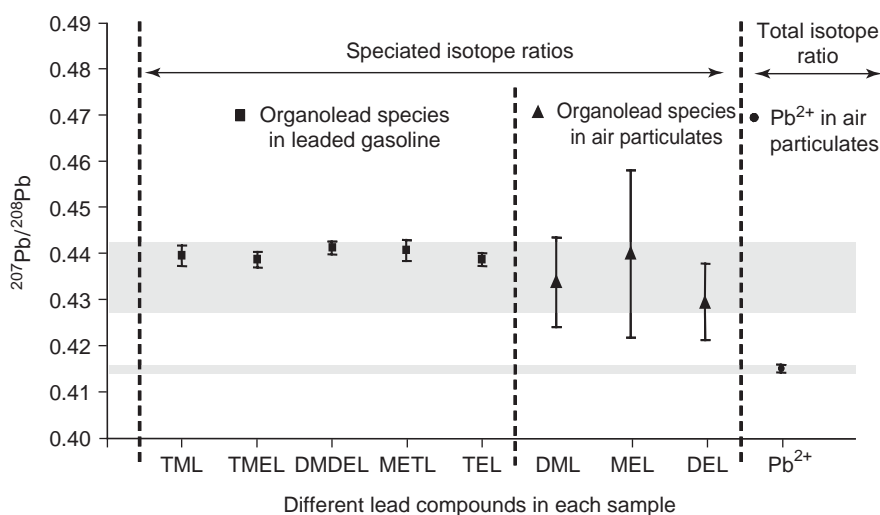
### Isotope Ratio Measurements

Lead isotope ratios can provide analytical information about sources of lead contamination whereas lead concentration measurements cannot. Due to this fact, studies of the isotopic composition of lead are commonly used in environmental science as well as in geological and anthropological studies. Among all of the naturally occurring lead isotopes only  $^{204}Pb$  is nonradiogenic, whereas  $^{206}Pb$ ,  $^{207}Pb$ , and  $^{208}Pb$  are daughter products from the radioactive decay of  $^{238}U$ ,  $^{235}U$ , and  $^{232}Th$ , respectively. This fact produces small Pb isotope abundance variations in nature, and it was observed that the isotopic composition of lead in environmental materials is clearly dependent on the ore deposits from which it came.

The differentiation between sources of lead organometallic compounds was recently accomplished using quadrupole ICP-MS. The measurement of lead isotope ratios by GC-ICP-MS of organolead standards allowed the differentiation between lead sources. The excellent precision of results obtained using a multicollector ICP-MS for lead isotope ratio

measurements by GC-ICP-MS is very promising for this kind of study.

In this way, characterization of the source of different organolead compounds in atmospheric particulate samples can be carried out by measuring lead isotope ratios for the various lead species. It is worth indicating that organolead species account for  $\sim 0.1\%$  of the total lead present in these samples. On the one hand, the glass fiber filters used in a standard high volume sampler were digested and the isotopic composition of total lead was measured by nebulization for ICP-MS using thallium as an internal isotope standard. On the other hand, organolead species were extracted and analyzed by GC-ICP-MS. Lead isotope ratios, measured for different organolead species present in these airborne particulates (DML, MEL, and DEL) and in leaded gasoline (TML, TMEL, DMDEL, METL, and TEL) collected in the same city, could be compared. The results obtained are illustrated in Figure 2. As can be observed, the isotopic composition of organolead species in the urban atmosphere was very close to that found in leaded fuel sold in the same city. On the other hand, total lead in these airborne particulates had an isotopic signature clearly different. Thus, the measurement of the isotope ratios in the organolead species allowed the influence of the leaded gasoline in the air pollution of the city under study to be traced whereas the measurement of the isotope ratios in the total lead failed to do this. Moreover, this speciated isotopic information greatly complements isotopic information on total lead, since it allows the direct and unambiguous identification of the origin of the



**Figure 2** Demonstration of the additional information provided by speciated isotope ratios for the analysis of total lead and organolead species in airborne particulate matter, and organolead species in leaded gasoline. Both types of samples were collected in Oviedo (Spain). Error bars correspond to one standard deviation.



pollution from leaded gasoline while the second one also allows the assessment of the contribution of the natural and industrial lead to the global lead contamination.

## Validation of the Analytical Methodology

### Certified Reference Materials

In order to improve the quality control of lead speciation analysis, two certified reference materials have been produced by the EC Measurement and Testing Programme. BCR CRM 604 (TriML in artificial rainwater) with a reference value of  $55.2 \pm 3.5$  ng per kg, and BCR CRM 605 (TriML in urban dust) with a reference value of  $7.9 \pm 1.2$  µg per kg. The latter is useful for testing the extraction and derivatization procedures in addition to the testing of the effect of large presence of inorganic lead. As an example, Figure 3 shows a GC-ICP-MS chromatogram of the CRM 605 material. As can be seen, a wide range of organolead species together with the certified TriML were detected (TEL was added as internal standard for quantification purposes).

### Isotope Dilution

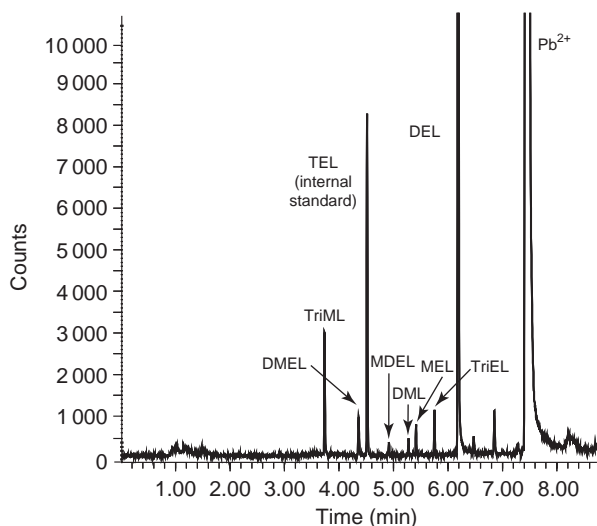
An emerging area of interest in analytical chemistry is the use of isotopically labeled species for validating organometallic compound speciation analysis by means of a primary method of measurement such as isotope dilution analysis. The synthesis of the species using an isotopically labeled metal is a prerequisite. This unique way of quantification, based also

on isotope ratio measurements in each separate species, leads to valuable analytical advantages if the isotope equilibration between the isotopically labeled species and the incurred one is achieved. If so, any loss that occurs in the subsequent steps, nonquantitative separations during sample preparation and preconcentration, matrix effects and detector instabilities are totally compensated. Up to now, this innovative methodology has been applied to organolead speciation only using LC as a separation technique. For this purpose, TriML enriched in  $^{206}\text{Pb}$  was synthesized and applied for the accurate and precise determination of TriML in rainwater.

**See also:** **Air Analysis:** Outdoor Air. **Atomic Emission Spectrometry:** Microwave-Induced Plasma. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Elemental Speciation:** Overview; Practicalities and Instrumentation. **Gas Chromatography:** Environmental Applications. **Isotope Dilution Analysis.** **Isotope Ratio Measurements.**

## Further Reading

- Bouysiere B, Szpunar J, and Lobinski R (2002) Gas chromatography with inductively coupled plasma mass spectrometric detection in speciation analysis. *Spectrochimica Acta Part B* 57: 805–828.
- Cornelis R, Caruso J, Crews H, and Heumann K (2003) *Handbook of Elemental Speciation. Techniques and Methodology*. Chichester: Wiley.
- Craig PJ (1986) *Organometallic Compounds in the Environment. Principles and Reactions*. Essex: Longman Group Limited.
- Ebdon L, Hill SJ, and Rivas C (1998) Lead speciation in rainwater by isotope dilution high performance liquid chromatography inductively coupled plasma mass spectrometry. *Spectrochimica Acta Part B* 53: 289–297.
- Ebdon L, Pitts L, Cornelis R, et al. (2001) *Trace Element Speciation for Environment, Food and Health*. Cambridge: Royal Society of Chemistry.
- Łobiński R, Boutron CF, Candelone JP, et al. (1993) Speciation analysis of organolead compounds in Greenland snow at the femtogram-per-gram level by capillary gas chromatography atomic emission spectrometry. *Analytical Chemistry* 65: 2510–2515.
- Montes-Bayon M, DeNicola K, and Caruso JA (2003) Liquid chromatography inductively coupled mass spectrometry. *Journal of Chromatography A* 1000: 457–476.
- Quevauviller Ph, Harrison R, Adams F, and Ebdon L (2000) An overview of methods for the determination of trimethyllead in rainwater and urban dust reference materials. *Trends in Analytical Chemistry* 19: 195–199.
- Radojevic M and Harrison RM (1987) Concentrations and pathways of organolead compounds in the



**Figure 3** GC-ICP-MS chromatogram obtained for the CRM 605 (urban dust) reference material.



environment: a review. *Science of the Total Environment* 59: 157–180.

Ruiz Encinar J, Leal Granadillo IA, Garcia Alonso JI, and Sanz-Medel A (2001) Isotope ratio measurements using gas chromatography inductively coupled plasma mass spectrometry for the assessment of organolead species. *Journal of Analytical Atomic Spectrometry* 16: 475–480.

Shotyk W, Weiss D, Heisterkamp M, *et al.* (2002) New peat bog record of atmospheric lead pollution in Switzerland: Pb concentrations, enrichment factors, isotopic composition and organolead species. *Environmental Science and Technology* 36: 3893–3900.

Spuznar J and Łobiński R (2003) *Hyphenated Techniques in Speciation Analysis*. Cambridge: Royal Society of Chemistry.

## LIMS

See **QUALITY ASSURANCE: Laboratory Information Management Systems**

## LIPIDS

Contents

**Overview**

**Fatty Acids**

**Polar Lipids**

**Determination in Biological Fluids**

### Overview

**A Kuksis**, University of Toronto, Toronto, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The domain of lipids comprises substances ranging from simple fatty acids to the complex phospholipids and from fat-soluble pigments and vitamins to sterols and steroid hormones. In the past these compounds have been frequently considered related mainly because of their mutual solubility in organic solvents and not because of their physical, chemical, or biochemical characteristics. Many of them, however, occur together in lipoproteins and cell membranes where they apparently form highly organized structures of a liquid crystalline type. It is now known that many lipid molecules are specifically bound by protein carriers, receptors, and enzymes. Faulty binding of lipids by carriers and receptors are known to lead to metabolic abnormalities and disease. It is obvious that the knowledge of complete lipid composition

will be necessary for the elucidation of the molecular relationships in lipoproteins and cell membranes in order to gain full insight and understanding of the normal metabolic and disease state of lipid metabolism.

Due to the chemical diversity of lipids, the major difficulty associated with any study of their chemistry or biochemistry has been that of isolation and separation into individual components. As a result, most advances in the understanding of lipid composition, structure, and function had to await the development of appropriate methodology. By the early 1960s new and improved procedures based upon the principle of chromatography had been forthcoming and had allowed a better fractionation and more complete identification of most lipids. The chromatographic efforts culminated in the development of combined or integrated thin-layer (TLC) and gas chromatography (GC) systems, which permitted an extensive resolution and quantification of many components of natural mixtures. During the next 40 years, both TLC and GC became complemented by high-performance liquid chromatography (HPLC) and by combinations of chromatography and mass spectrometry (MS), which further improved the certainty of identification and quantification of lipids

environment: a review. *Science of the Total Environment* 59: 157–180.

Ruiz Encinar J, Leal Granadillo IA, Garcia Alonso JI, and Sanz-Medel A (2001) Isotope ratio measurements using gas chromatography inductively coupled plasma mass spectrometry for the assessment of organolead species. *Journal of Analytical Atomic Spectrometry* 16: 475–480.

Shotyk W, Weiss D, Heisterkamp M, *et al.* (2002) New peat bog record of atmospheric lead pollution in Switzerland: Pb concentrations, enrichment factors, isotopic composition and organolead species. *Environmental Science and Technology* 36: 3893–3900.

Spuznar J and Łobiński R (2003) *Hyphenated Techniques in Speciation Analysis*. Cambridge: Royal Society of Chemistry.

## LIMS

See **QUALITY ASSURANCE: Laboratory Information Management Systems**

## LIPIDS

Contents

**Overview**

**Fatty Acids**

**Polar Lipids**

**Determination in Biological Fluids**

### Overview

**A Kuksis**, University of Toronto, Toronto, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The domain of lipids comprises substances ranging from simple fatty acids to the complex phospholipids and from fat-soluble pigments and vitamins to sterols and steroid hormones. In the past these compounds have been frequently considered related mainly because of their mutual solubility in organic solvents and not because of their physical, chemical, or biochemical characteristics. Many of them, however, occur together in lipoproteins and cell membranes where they apparently form highly organized structures of a liquid crystalline type. It is now known that many lipid molecules are specifically bound by protein carriers, receptors, and enzymes. Faulty binding of lipids by carriers and receptors are known to lead to metabolic abnormalities and disease. It is obvious that the knowledge of complete lipid composition

will be necessary for the elucidation of the molecular relationships in lipoproteins and cell membranes in order to gain full insight and understanding of the normal metabolic and disease state of lipid metabolism.

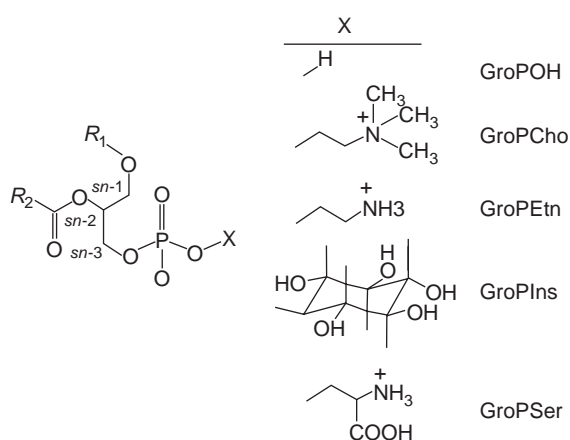
Due to the chemical diversity of lipids, the major difficulty associated with any study of their chemistry or biochemistry has been that of isolation and separation into individual components. As a result, most advances in the understanding of lipid composition, structure, and function had to await the development of appropriate methodology. By the early 1960s new and improved procedures based upon the principle of chromatography had been forthcoming and had allowed a better fractionation and more complete identification of most lipids. The chromatographic efforts culminated in the development of combined or integrated thin-layer (TLC) and gas chromatography (GC) systems, which permitted an extensive resolution and quantification of many components of natural mixtures. During the next 40 years, both TLC and GC became complemented by high-performance liquid chromatography (HPLC) and by combinations of chromatography and mass spectrometry (MS), which further improved the certainty of identification and quantification of lipids

with high sensitivity. Thus, GC in combination with electron ionization MS (GC-MS) and HPLC with online electrospray ionization MS (LC/ESI-MS) became established as the definitive methods of lipid analyses. In recent years, the latter systems have become largely replaced by tandem mass spectrometry, which abandons the chromatographic step for speed and efficiency although at some sacrifice of resolution and quantification. In food and pharmaceutical industry, normal phase HPLC has been extensively replaced by packed column supercritical fluid chromatography (SFC), which has resulted in significant savings in solvents.

The present overview briefly considers the more successful applications of the individual methodologies in lipid analyses, the improvements brought about by their complementary combinations, along with the progress toward further resolution and more sensitive analyses. The overview concludes with a brief consideration of the lipid analyses demanded by lipidomics, and recent developments in proteomics and genomics.

## Types of Lipids Encountered

The common lipids are of two general types: neutral lipids and polar lipids. Both are esters of the common fatty acids, such as palmitate, stearate, oleate, linoleate, arachidonate, and docosahexaenoate, as well as others. The major neutral lipids are made up of the fatty acid esters of glycerol (mono-, di-, and triacylglycerols) and of cholesterol or other sterols. The triacylglycerols may contain one-, two-, or three different fatty acids per molecule, which are non-randomly distributed and give rise to a complex mixture of molecular species. The polar lipids consist of glycerophospholipids, which are mixed organic and inorganic acid esters of glycerol and sphingomyelins, which are derivatives of the base sphingosine. The natural glycerophospholipids represent a single enantiomer and contain either a fatty acid or an alkyl or vinyl ether at the *sn*-1-position and a fatty acid at *sn*-2-position, while the *sn*-3-position is occupied by phosphoric acid (stereospecific numbering). Also, esterified to the phosphate group is one of a series of polar head groups, which include choline, ethanolamine, serine, and one of several forms of the polyhydroxycyclitol inositol. The phosphate can also be esterified to phosphatidylglycerol resulting in four fatty acyl substituted glycerophospholipid (cardiolipin). Molecular species diversity in these phospholipids results from a pairing of different fatty chains in a glycerophospholipid molecule, with a saturated fatty chain usually occupying the *sn*-1 and an unsaturated fatty chain the *sn*-2-position. A



**Figure 1** Major neutral glycerolipids and glycerophospholipids. The general formula on the left depicts an *sn*-1-alkyl-2-acylglycerophospholipid with a head group  $X$ , which may represent  $H$ , as in GroPOH, choline, as in GroPCho, ethanolamine, as in GroPEtn, inositol, as in GroPIns, glycerol, as in GropGro or serine, as in GroPSer, all as *sn*-1,2-enantiomers. Substitution of the *sn*-1-alkyl and the *sn*-3 phosphoryl group with fatty acids gives a triacylglycerol, which upon removal of one or two fatty acids yields a diacyl- or monoacylglycerol, the isomer being determined by the acyl group removed. The three positions of the glycerol molecule may be substituted by one, two, or three different types of fatty acids yielding mixed acid triacylglycerols. Redrawn from Murphy and Harrison (*Mass Spectrometric Reviews* 13: 57–75. 1994).

reversal of this substitution results in reverse isomer formation. The sphingomyelins contain a long chain base of the sphingosine class. These phospholipids contain a fatty acyl group in an amide bond with the basic nitrogen atom of the long chain base and phosphocholine residue esterified to the primary alcohol group of the long chain base. Molecular species diversity in this phospholipid class results from a combination of various long chain bases with various common fatty acyl groups. Representative structures of the common glycerophospholipids are found in **Figure 1**.

## Historical Developments

The earliest attempts of lipid analyses are found in the isolation and identification of individual fatty acids, phospholipids, and triacylglycerols. Although a number of fatty acids were isolated and identified by purely chemical methods, such as metal salt formation, bromination, and crystallization, effective separation of fatty acids was first achieved by distillation, which permitted resolution of fatty acid homologs. Molecular distillation of fatty acids became a forerunner of their eventual resolution by GC. Likewise, chemical methods were used for the initial isolation of the naturally occurring glycerophospholipids as individual lipid classes or mixtures

of related phospholipid classes. Although it was recognized that all these preparations represented mixtures of molecular species, there were no means available that would permit their resolution. The early isolations of natural triacylglycerols were obtained by crystallization, which also led to mixtures of molecular species precipitating a debate, whether or not these preparations represented mixtures of single acid or mixed acid triacylglycerols. Sterols and steroids were isolated in relatively pure form by crystallization, although close isomers were not resolved until chromatographic methods became available.

The field of fatty acid separation was revolutionized by the development of GC, which allowed nearly complete separation of the common fatty acids. Later workers developed methods for GC separation of steroids, and of the steryl and glyceryl esters by high-temperature GC. The introduction of the flame ionization detector (FID) transformed GC from a method of qualitative separation into a method of quantification, which has remained unsurpassed to the present day. In parallel, TLC was developed as both a preparative and analytical technique for the resolution of the nonvolatile lipids. Because of the nearly universal applicability to all lipid analyses and the use of inexpensive apparatus, TLC made lipid separations accessible to all laboratories and it has remained unsurpassed as a method of sample preparation for modern analytical methods of lipid analyses.

## Current Methodology

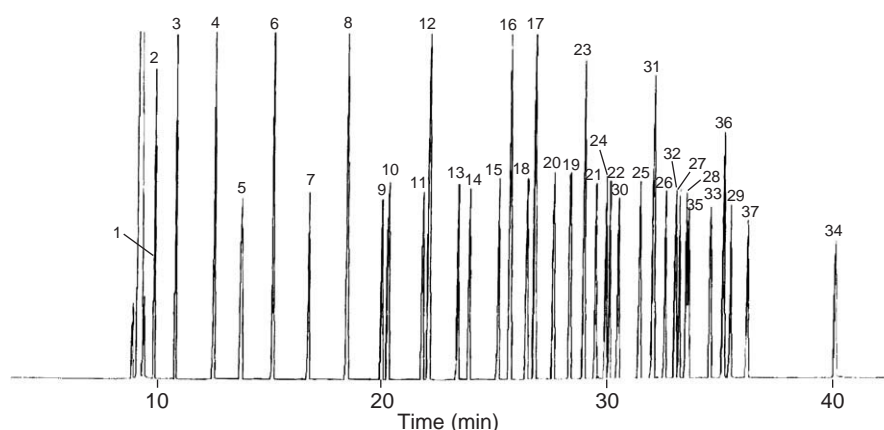
High-performance TLC (HPTLC) has now largely replaced conventional TLC as a method of lipid resolution and quantification. Scanning of charred and radioactive spots provides improved quantitative estimates due to increased resolution of all components compared to conventional TLC. It serves as an efficient method of sample isolation for sensitive MS identification of various lipids. Modern TLC is a fully instrumental technique distinguished from conventional TLC. Because the technique is optimized to achieve maximum separation with high levels of automation, the results are largely independent of operator skill. Both one-dimensional and two-dimensional HPTLC are commonly employed. Normal phase TLC on Chromarods (latroscan) with FID constitutes a special application to separation and quantification of both neutral and polar lipid classes. The order of resolution of the lipid classes is directly related to the polarity of the solutes, which may be modified to some extent by the preparation of derivatives or altering the polarity of the adsorbent.

Various complexing agents may be incorporated into the adsorbent to permit improved resolution of saturated and unsaturated lipid classes and molecular species (silver nitrate) and the resolution of *vic*-hydroxylated and nonhydroxylated lipid classes (borate). A practical combination of HPTLC with MS, however, has not yet materialized. In contrast, normal phase HPLC has been effectively combined with online ESI-MS and atmospheric pressure chemical ionization (APCI)-MS for analyses of both lipid classes and molecular species (see below).

Conventional GC currently serves mainly for the separation and quantification of fatty acids as the methyl esters and sterols. Capillary columns, up to 100 m in length, coated with liquid phases of moderate polarity allow extensive but not complete resolution of all homologs and isomers. **Figure 2** shows a GC separation of a complex mixture of standard fatty acid methyl esters on a 100 m fused silica capillary column (SP-2560, Supelco). A temperature program from 140°C to 240°C permitted a separation of C4 to C24 saturated and polyunsaturated fatty acids. A total of 37 peaks were resolved and quantified by FID. Helium was used as a carrier gas. Presumably, further resolution and earlier elution would have been obtained with hydrogen as a carrier. However, many peaks remain unresolved in the 18:0 and 18:2 $n$ 6 region, including the methylene interrupted and noninterrupted 18:2 isomers. The picolinyl esters and dimethyloxazoline derivatives of unsaturated fatty acids have been used for a GC-MS determination of the location of the double bonds.

GC-MS and GLC are commonly used methods for the identification and quantification of sterols from samples of biological origin. **Figure 3** gives the retention times relative to 5 $\alpha$ -cholestane for the trimethylsilyl (TMS) ether derivatives of unsaturated C<sub>27</sub> sterols and related sterols on two 60 m capillary columns of moderate polarity. An inspection of the retention data indicates that both liquid phases fail to separate numerous sterols, even on highly efficient capillary columns. A preliminary resolution of the sterols into desmethyl, monomethyl, and dimethyl sterols is, therefore, required to provide supporting evidence for the identification of sterol structures. Careful attention to GC and MS conditions is necessary as is calibration of GC and MS data with authentic sterol standards. Furthermore, possible sterol decomposition under GC conditions, as well as the contribution of (Parent  $\pm$  2)<sup>+</sup> to the single ion mass chromatograms must also be kept in mind.

High-temperature GC on nonpolar capillary columns permits the resolution of such natural lipid esters as the mono- and diacylglycerols as the TMS ethers, and intact triacylglycerols and cholesteryl



**Figure 2** A gas-liquid chromatogram (with temperature programming from 140 to 240°C) of a complex mixture of standard fatty acid methyl esters separated on a 100-m fused silica capillary column (SP-2560, Supelco). Peak identification by carbon and double bond number: 1, 4:0; 2, 6:0; 3, 8:0; 4, 10:0; 5, 11:0; 6, 12:0; 7, 13:0; 8, 14:0; 9, 14:1; 10, 15:0; 11, 15:1; 12, 16:0; 13, 16:1; 14, 17:0; 15, 17:1; 16, 18:0; 17, 18:1n9c; 18, 18:1n9t; 19, 18:2n6c; 20, 18:2n6t; 21, 18:3n6; 22, 18:3n3; 23, 20:0; 24, 20:1n9; 25, 20:2; 26, 20:3n6; 27, 20:3n3; 28, 20:4n6; 29, 20:5n3; 30, 21:0; 31, 22:0; 32, 22:1n9; 33, 22:2; 34, 22:6n3; 35, 23:0; 36, 24:0; 37, 24:1n9. (Reproduced from *Chromatography* (2003/2004) Sigma-Aldrich Canada.)

esters based on carbon number, while high-temperature GC on capillary columns containing polarizable liquid phases allows the resolution of the above lipid classes on the basis of carbon number and degree of unsaturation. Both types of GC separations can be readily interfaced with online MS and MS/MS to provide unequivocal peak identification and quantification in relation to appropriate internal standards. High-temperature GC/FID may also be utilized for the analysis of molecular species of glycerophospholipids and sphingomyelins, provided the polar head groups are first removed by phospholipase C and the hydroxyl groups of the released diacylglycerols and ceramides are protected by silylation. Polar capillary GC/FID provides the highest resolution of the molecular species of glycerophospholipids via their diradylglycerol moieties along with superior quantification. **Figure 4** shows the polar capillary GC profiles of the diacylglycerol (A), alkylacylglycerol (B), and alkenylacylglycerol (C) moieties of plasma diradyl glycerophosphoethanolamines. Prior to GC, the diradylglycerols, as the TMS ethers, were resolved into the diacyl, alkylacyl, and alkenylacyl subclasses by normal phase HPLC. The major chromatographic peaks are identified in the figures by the most likely pairing of the fatty chains based on the retention times of the peaks. Recent work has demonstrated that the latter separations and quantification agree fully with the results of LC/ESI-MS analyses, which, however, are more rapid and more sensitive (see below).

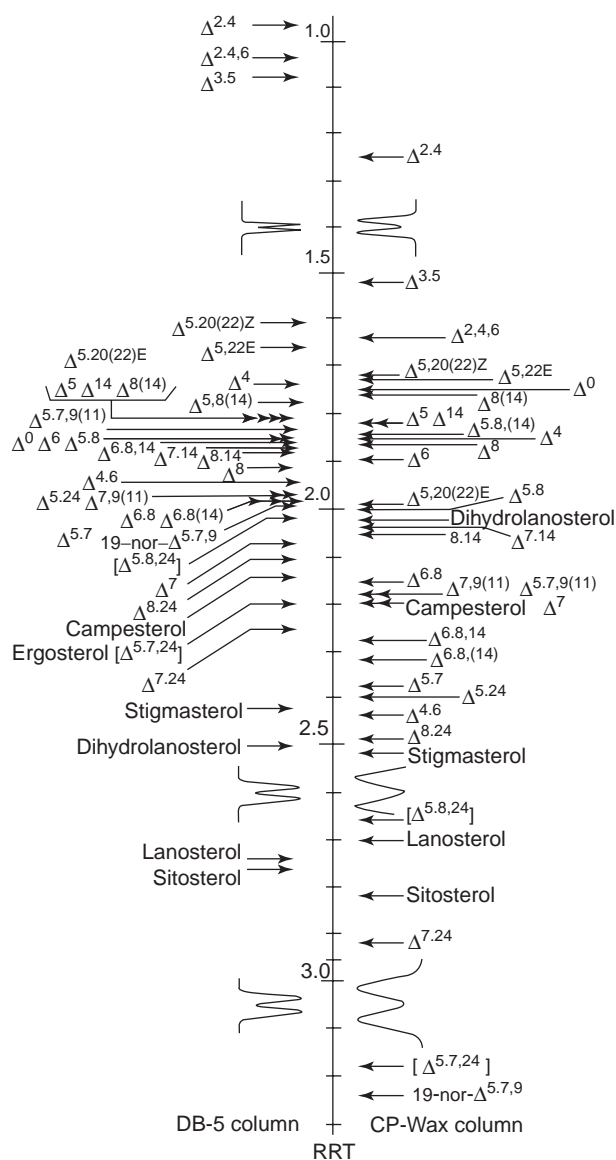
At present, the chromatography of high molecular weight neutral lipids and polar lipids is usually performed by reversed phase HPLC, although a lack of a universal detector prevents the full utilization of this

analytical system. Light scattering detectors are not sufficiently sensitive, while manufacture of FIDs for HPLC applications has been discontinued. Online MS can provide the necessary detection capacity, but its high cost and variable response severely limits its general utilization. Reversed phase HPLC with ultraviolet (UV) detection is well suited for the analyses of molecular species of fatty acids, monoacylglycerols, and diacylglycerols, for which the derivatization step can be combined with the introduction of a UV absorbing or fluorescent ester or ether group, respectively. Thus, the pentafluorobenzoyl esters of fatty acids and pentafluorobenzyl ethers of acylglycerols have been well resolved and quantitatively measured by reversed phase HPLC/UV. Other UV absorbing or fluorescent fatty acid derivatives may be prepared for structural studies using HPLC-MS, as it is done using GC-MS.

Reversed phase HPLC with light scattering detection has been extensively utilized for the resolution of molecular species of triacylglycerols, where sufficiently large samples have been available. **Figure 5** shows a reversed phase HPLC separation of the complex molecular species of bovine milk fat triacylglycerols. A randomized sample of butterfat, the composition of which can be calculated on basis of random distribution of fatty acids, is used for determination of the component molecular species to serve as standard for identification of the naturally occurring triacylglycerol species.

Reversed phase HPLC has recently provided excellent resolution of reverse isomers within an enantiomeric diacylglycerol class. **Figure 6** demonstrates the reversed phase HPLC resolution as 3,5-dinitrophenylurethanes (3,5-DNPU) of standard





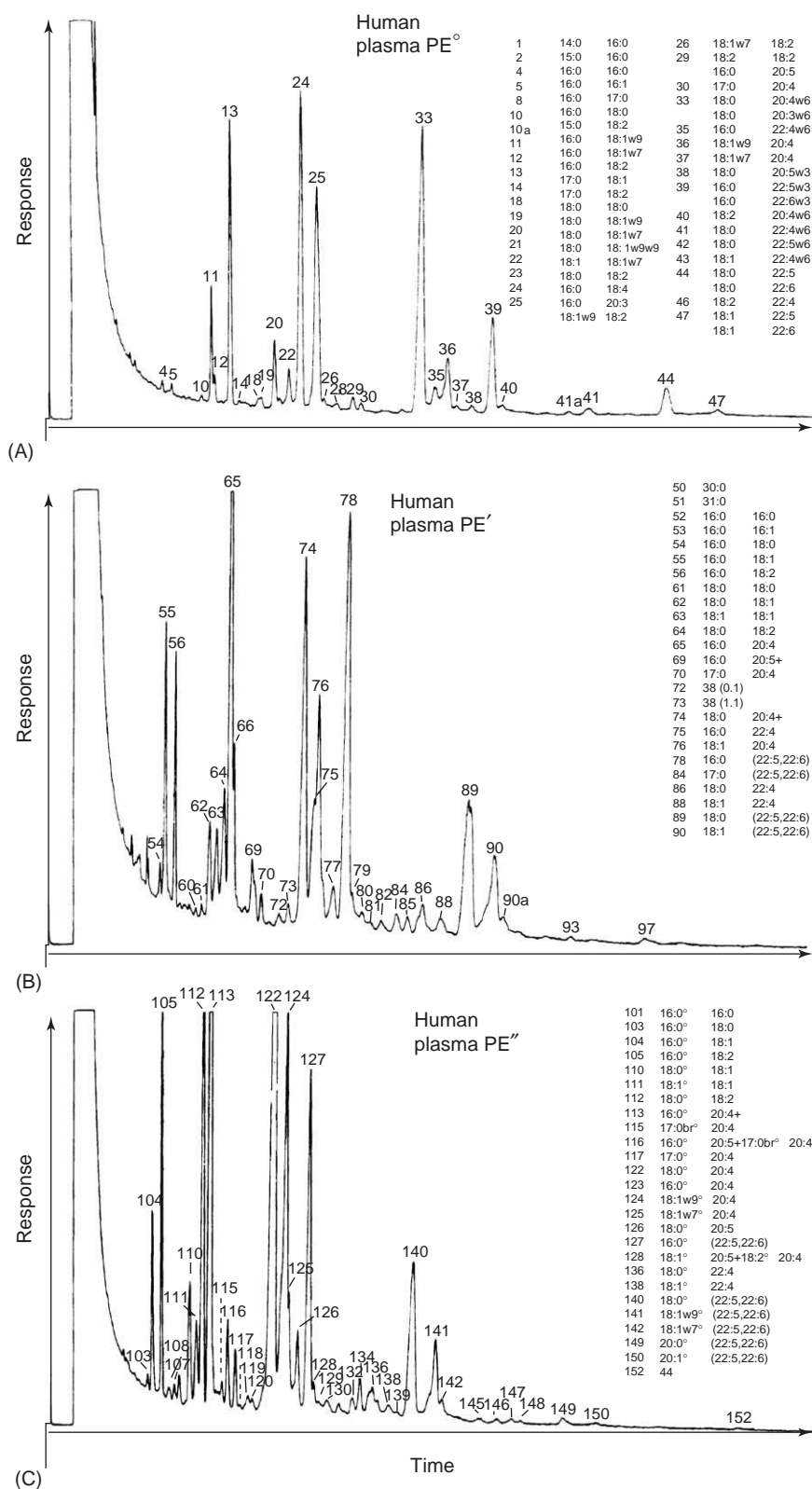
**Figure 3** Retention times relative to 5 $\alpha$ -cholestane for TMS ether derivatives of unsaturated C<sub>27</sub>-sterols and related sterols on DB-5 (60 m) and CP-Wax columns. Unsaturated sterols are all of 5 $\alpha$  configuration (or  $\Delta^4$  or  $\Delta^5$ ) and are designated by their unsaturation. (Reproduced with permission from Gerst *et al.* (1997) *Journal of Lipid Research* 38: 1685–1701.)

*rac*-1,2-diacylglycerols made up of a pair of reverse isomers, 1-palmitoyl-2-linolenoyl- and 1-linolenoyl-2-palmitoyl-*sn*-glycerols, as well as 1-palmitoyl-2-docosahexaenoyl- and 1-docosahexaenoyl-2-palmitoyl-*sn*-glycerols, and two other pairs of reverse isomers. These separations have been recently extended to the diacylglycerol moieties of naturally occurring glycerophospholipids in a search for the natural occurrence of the reverse isomers, which are conventionally determined by hydrolysis with phospholipase A<sub>2</sub>, which requires large amounts of sample, or MS, which is less accurate (see below).

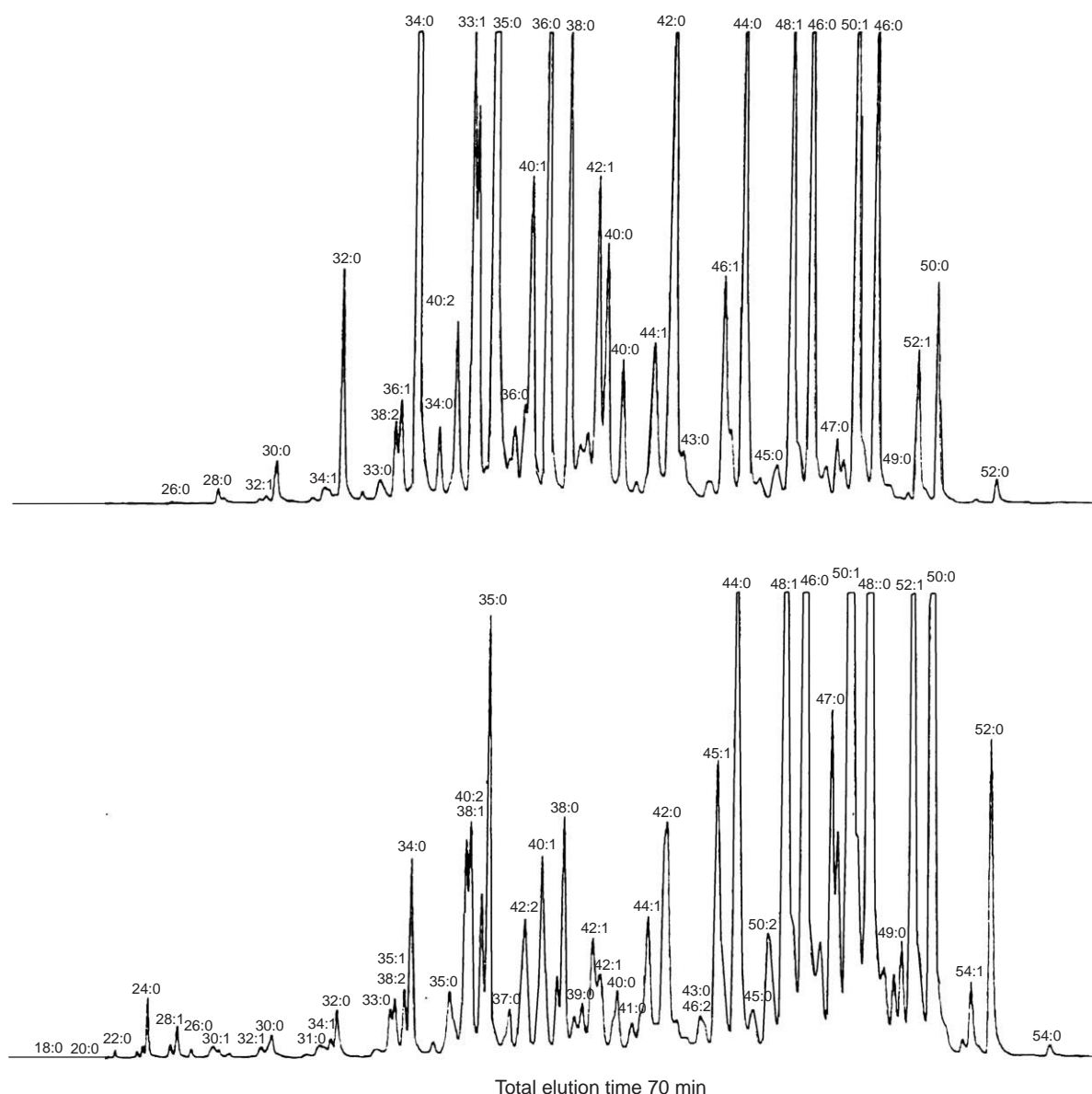
Reversed phase HPLC has also been utilized for the resolution of molecular species of intact glycerophospholipids and sphingomyelins. Excellent resolutions have been obtained, but the monitoring of molecular species by UV of short wavelength has been unsatisfactory. Furthermore, the use of solvents including nonvolatile counterions has prohibited the interfacing of these chromatographic separations with online MS.

HPLC using chiral columns has proved well suited for the resolution of enantiomeric monoacylglycerols, monoalkylglycerols, monoalkenylglycerols, diacylglycerols, and alkylacylglycerols. The monoacylglycerols are resolved following the preparation of the di-dinitrophenylurethane derivatives, while the diacylglycerol separation is achieved with the mono-dinitrophenylurethanes. The racemates of even complex mixtures of diacylglycerols, such as those derived from fish oils, may be resolved into enantiomers without overlapping. The resolved enantiomers are identified and quantified by ESI-MS and the method has been widely applied in the analyses of the structure of natural fat and oils following partial Grignard deacylation. **Figure 7** shows the chiral phase HPLC separation of the enantiomeric *sn*-1,2- and *sn*-2,3-diacylglycerols derived from manhaden oil. The most unsaturated *sn*-1,2-diacylglycerols are eluted well ahead of the most saturated of *sn*-2,3-diacylglycerols. The X-1,3-diacylglycerols overlap with the *sn*-1,2-diacylglycerols, unless they have been first resolved by TLC or HPLC using borate-treated adsorbents. Using a chiral column of opposite configuration HPLC results in a reverse order of elution of the enantiomers, which may be taken advantage of in those instances where the tail end of the *sn*-1,2-enantiomers overlaps in part with the leading edge of the *sn*-2,3-enantiomers. The composition of the molecular species is best deduced from the fatty acid pairing seen by online LC/ESI-MS.

Chiral phase HPLC can also provide a resolution of the reverse isomers within an enantiomeric diacylglycerol. Thus, under appropriate elution conditions, full baseline resolution may be obtained between pairs of enantiomeric diacylglycerols, such as 1-docosahexaenoyl-2-palmitoyl- and 1-palmitoyl-2-docosahexaenoyl-*sn*-glycerols, which are present among the diacylglycerols derived from fish oils, leading to confusion during mass spectrometric identification of the molecular species of enantiomeric diacylglycerols resolved by chiral phase HPLC. **Figure 8** shows the chiral phase HPLC resolution of the reverse isomers and enantiomers of 1,2-diacyl-*rac*-glycerols containing polyunsaturated acyl chains as the 3,5-DNPs on (*R*)-1-(1-naphthyl)ethylamine



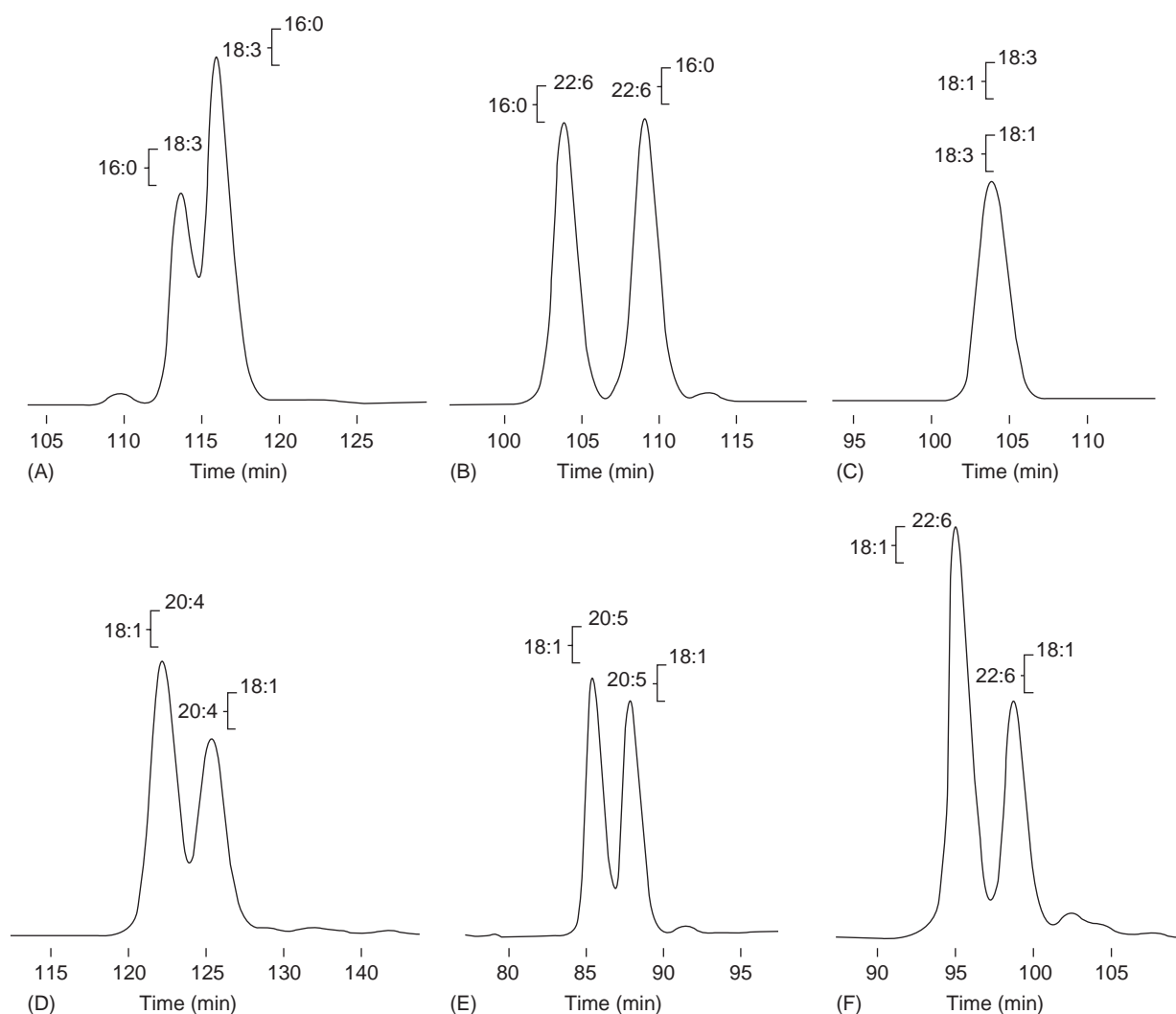
**Figure 4** Polar capillary GC profiles of the diradylglycerol moieties of human plasma glycerophosphoethanolamines: (A) diacylglycerols; (B) alkylacylglycerols; (C) alkenylacylglycerols. Peak identifications are given in the figures. GC conditions: columns, 15 m  $\times$  0.32 mm fused silica capillary coated with cross-bonded RTx 2330; carrier gas, H<sub>2</sub>, 3 psi; temperature, 250°C, isothermal. Sample: 1  $\mu$ l of 0.1% diradylglycerol TMS ethers in hexane. (Reproduced with permission from Myher JJ, Kuksis A, and Pind S (1989) Molecular species of glycerophospholipids and sphingomyelins of human plasma: comparison to red blood cells. *Lipids* 24: 408–418.)



**Figure 5** Reversed phase HPLC (linear gradient of 10–90% isopropanol in acetonitrile at 25°C) of natural (top) and rearranged (bottom) butterfat triacylglycerols as obtained with light-scattering detector. Peak identification by carbon and double bond number. (Reproduced with permission from Marai L, Kuksis A, and Myher JJ (1994) Reversed-phase liquid chromatography–mass spectrometry of the uncommon triacylglycerol structures generated by randomization of butteroil. *Journal of Chromatography A* 672: 87–99.)

liquid phase. There is reasonably good resolution of the reverse isomers of 1-docosahexaenoyl-2-palmitoyl-*rac*-glycerol (22:6–16:0), 1-docosahexaenoyl-2-oleoyl-*rac*-glycerol (22:6/18:1), and 1-eicosapentaenoyl-2-oleoyl-*rac*-glycerol (20:5/18:1) only within their *sn*-2,3-enantiomers, for which the isomers with the acyl group of the higher degree of unsaturation are in the secondary position. There was a similar resolution of the reverse isomers within *sn*-1,2-enantiomers on the chiral stationary phase having the opposite configuration. Again the reverse isomers

with the acyl groups of higher degree of unsaturation in the secondary position were retained more strongly. The presence of long chains of saturated or monounsaturated fatty acids in the secondary position, which are close to 3,5-DNPU group, apparently interfered with diastereomeric hydrogen-bonding and pi-pi donor-acceptor interaction between the 3,5-DNPU and (*R*)- or (*S*)-1-(1-naphthyl)ethylamine stationary phases and resulted in impaired separation of diacylglycerols having long acyl chains in the secondary position.

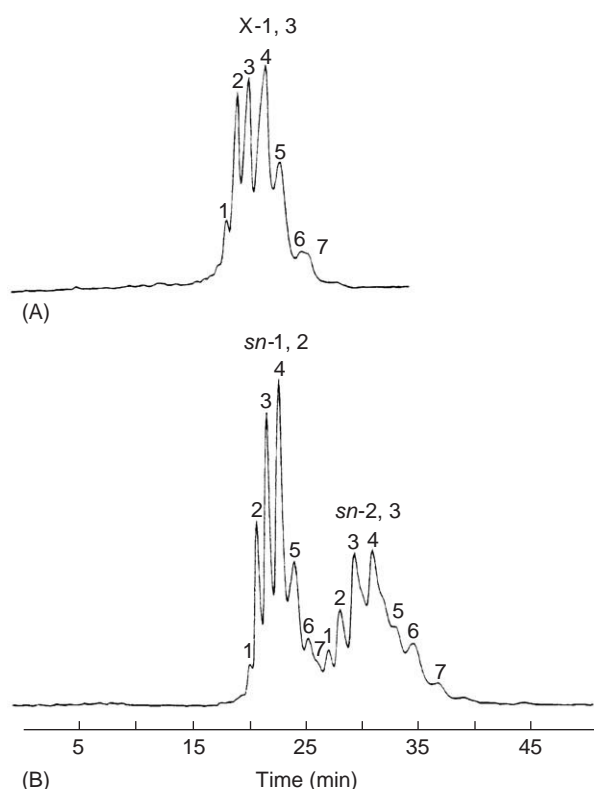


**Figure 6** Reversed phase HPLC resolution (isocratic acetonitrile at 18°C.) of the reverse isomers of 1,2-diacyl-*rac*-glycerols containing polyunsaturated acyl chains as 3,5-dinitrophenylurethanes. Peak identification: A, 1-linolenoyl-2-palmitoyl-*rac*-glycerol and 1-palmitoyl-2-linolenoyl-*rac*-glycerol; B, 1-docosahexaenoyl-2-palmitoyl-*rac*-glycerol and 1-palmitoyl-2-docosahexaenoyl-*rac*-glycerol; C, 1-linolenoyl-2-oleoyl-*rac*-glycerol and 1-oleoyl-2-linolenoyl-*rac*-glycerol; D, 1-arachidonoyl-2-oleoyl-*rac*-glycerol and 1-oleoyl-2-arachidonoyl-*rac*-glycerol; E, 1-eicosapentaenoyl-2-oleoyl-*rac*-glycerol and 1-oleoyl-2-eicosapentaenoyl-*rac*-glycerol; F, 1-docosahexaenoyl-2-oleoyl-*rac*-glycerol and 1-oleoyl-2-docosahexaenoyl-*rac*-glycerol. (Reproduced with permission from Itabashi Y, Myher JJ, and Kuksis A (2000) High performance liquid chromatographic resolution of reverse isomers of 1,2-diacyl-*rac*-glycerols as 3,5-dinitrophenylurethanes. *Journal of Chromatography A* 893: 261–279.)

An effective resolution of *rac*-diacylglycerols may be obtained by normal phase HPLC following preparation of the diastereomeric dinitrophenylethylurethanes. In this instance the *R* and *S* derivatives may be prepared, which results in opposite elution order of the diastereomers. Again, online ESI-MS is necessary for identification of the molecular species as the elution order of the molecules is complicated by both adsorption and partition effects. Normal phase HPLC on a silver-loaded cation-exchange column has permitted resolution of the reverse isomers of *rac*-diacylglycerols as the 3,5-DNPU, such as

1-palmitoyl-2-oleoyl-*rac*-glycerol and 1-oleoyl-2-palmitoyl-*rac*-glycerol, as well as other pairs of reverse isomers containing fatty chains of greater differences in their length or unsaturation. The separation of chiral compounds by packed column SFC has been a great success for the pharmaceutical industry. Almost all of the chiral selectors used in gas and liquid chromatography have been successfully applied in SFC.

Normal phase and to a lesser extent reversed and chiral phase HPLC have been combined online with thermospray ionization (TSI)-MS, ESI-MS, or APCI-MS for the purpose of identification and



**Figure 7** Chiral phase ((*R*)-(+)-1-(1-naphthyl)ethylamine polymer) HPLC resolution of the diacylglycerol moieties of menhaden oil triacylglycerols, as dinitrophenylurethanes. (A) X-1,3-diacylglycerols; (B) *sn*-1,2(2,3)-diacylglycerols. Peak identification confirmed by LC/NCI-MS as equivalent carbon numbers: 1, 32; 2, 30; 3, 28; 4, 26; 5, 24; 6, 22; 7, 20. Peaks monitored at 254 nm during isocratic elution with hexane–1,2-dichloroethane–ethanol (40:10:1, by vol). (Reproduced with permission from Itabashi Y, Kuksis A, Marai L, and Takagi T (1990) HPLC resolution of diacylglycerol moieties of natural triacylglycerols on a chiral phase consisting of bonded (*R*)-(+)-1-(1-naphthyl)ethylamine. *Journal of Lipid Research* 31: 1711–1717.)

quantification of lipid classes and molecular species. The success has depended on the lipid class involved and the exact methodology applied. Because of the ease of interfacing of the above methodologies, online TSI-MS, ESI-MS, and APCI-MS have been utilized most often. Online ESI-MS has given excellent results with both neutral and polar lipids, while APCI-MS has given best results when applied to triacylglycerols. In the case of packed column SFC, a combination with time-of-flight MS has been especially successful, although other MS methods have also been coupled to it.

The choline-containing phospholipids (PtdCho, lysoPtdCho, and sphingomyelin) are commonly detected by ESI-MS in the positive ion mode although in the negative ion mode they may show higher response as the chloride adducts. In the negative ion mode, PtdCho may be determined also as the

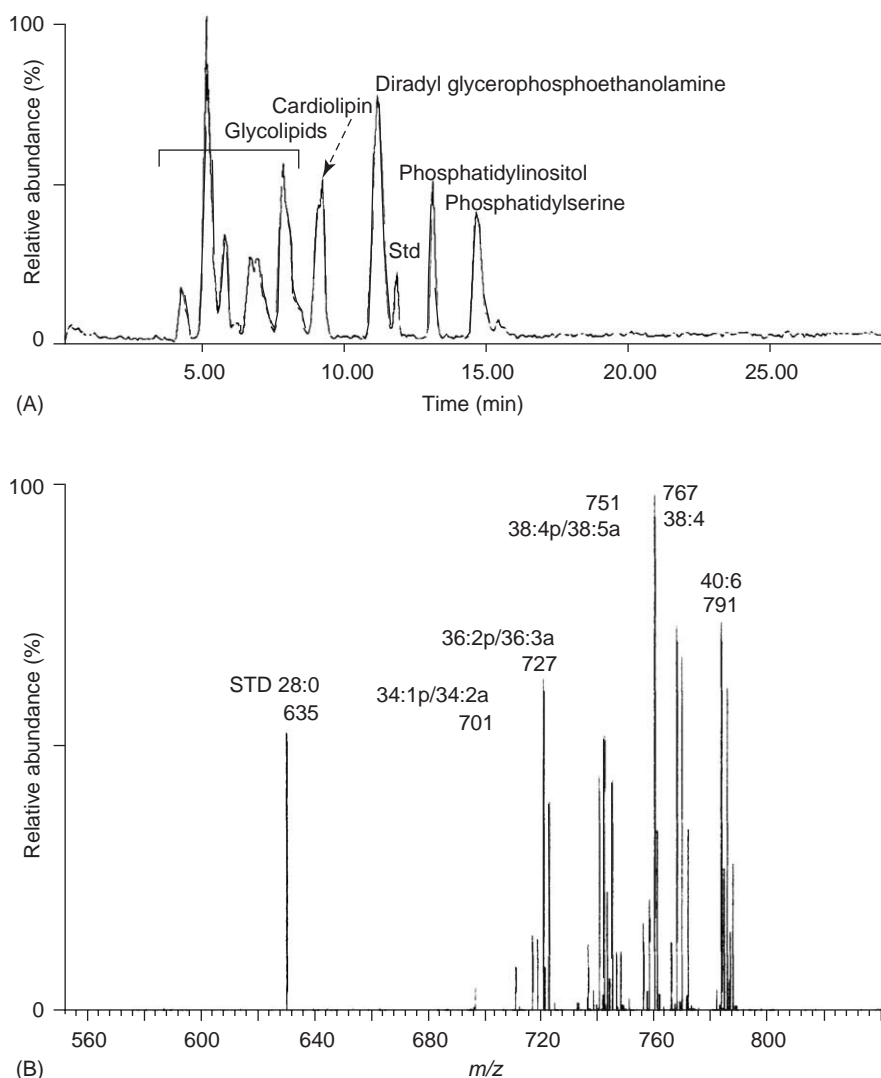
$[M - 15]^-$  ion resulting from the loss of a methyl ( $-\text{CH}_3$ ) group from the choline moiety at cone ionization voltages somewhat higher than those necessary for the ionization of the acidic glycerophospholipids. The individual molecular species making up the various choline-containing peaks may be readily extracted from the total ion current profiles and quantified in relation to an appropriate internal standard for each phospholipid class. The serine, inositol, and ethanolamine glycerophospholipids are commonly determined in the negative ion mode, although the ethanolamine phospholipids also give strong signal in the positive ion mode. In both instances the prior HPLC resolution of the lipid classes, including partial resolution of isobaric molecular species, serves to facilitate the subsequent identification of the molecular species based on the molecular ion and the fragment ion spectra. The interpretation of the mass spectra of the glycerophospholipids is greatly facilitated by the development of a software algorithm for automatic interpretation of the mass spectra. The algorithm interprets spectra according to user-specified identification of ions, corrects the ion intensities for abundances of naturally occurring isotopes, applies response correction factors to each identified ion, quantifies the identified compounds relative to an internal standard, and applies sample dilution factors to obtain quantitative results. **Figure 9** shows the LC/ESI-MS analysis of glycerophospholipids in a total lipid extract of guinea pig brain. The total ion chromatogram recorded in negative ionization mode shows the separation of the phospholipid classes (A). The mass spectrum averaged over the diacylglycerophosphoethanolamine peak displaying  $[M - H]^-$  ions of molecular species along with the identification of the major species is also shown (B). In the interpretation process the algorithm searches the input spectrum against the user-defined auxiliary file and accepts all identifications in which the measured  $m/z$  value fulfills the criteria ( $m/z \pm \text{range}$ ). The allowed error in measurement of  $m/z$  values of  $[M - H]^-$  or  $[M + H]^+$  ions must be less than 0.5 Th. The compound identification process results in a list of identified compounds, their measured  $m/z$  values, and abundances (**Table 1**). Automatic processing of the spectra was shown to be accurate and reliable by testing with spectra of glycerophospholipids obtained by LC/ESI-MS with manual interpretation of the spectra.

At present many laboratories have abandoned LC-MS in favor of MS/MS or MS/MS/MS as a method of determining molecular species of phospholipids. Total lipid extracts are subjected to MS analysis for molecular ions, from which fragment ions are obtained by collision-induced dissociation (CID) for



**Figure 8** Chiral phase HPLC resolution of the reverse isomers and enantiomers of 1,2-diacyl-*rac*-glycerols containing polyunsaturated acyl chains as 3,5-dinitrophenylurethanes on (*R*)-1-(1-naphthyl)ethylamine. Peak identification (from left to right) A, 1-linolenoyl-2-palmitoyl + 1-palmitoyl-2-linolenoyl-*sn*-glycerols and 2-palmitoyl-3-linolenoyl + 2-linolenoyl-3-palmitoyl-*sn*-glycerols; B, 1-docosahexaenoyl-2-palmitoyl + 1-palmitoyl-2-docosahexaenoyl-*sn*-glycerols and 2-palmitoyl-3-docosahexaenoyl + 2-docosahexaenoyl-3-palmitoyl-*sn*-glycerols; C, 1-eicosapentaenoyl-2-oleoyl + 1-oleoyl-2 eicosapentaenoyl-*sn*-glycerols and 2-oleoyl-3-eicosapentaenoyl + 2-eicosapentaenoyl-3-oleoyl-*sn*-glycerols; D, 1-docosahexaenoyl-2-oleoyl + 1-oleoyl-2-docosahexaenoyl-*sn*-glycerols and 2-oleoyl-3-docosahexaenoyl + 2-docosahexaenoyl-3-oleoyl-*sn*-glycerols. A mixture of the reverse isomers of 1,2-diacyl-*rac*-glycerols was analyzed separately. HPLC conditions as in **Figure 6**. (Reproduced with permission from Itabashi Y, Myher JJ, and Kuksis A (2000) High performance liquid chromatographic resolution of reverse isomers of 1,2-diacyl-*rac*-glycerols as 3,5-dinitrophenylurethanes. *Journal of Chromatography A* 893: 261–279. © Elsevier.)

chromatographic adsorbent. It is frequently claimed that the MS/MS data are superior to that obtained by chromatography in terms of accuracy of identification and quantification of molecular species. It is unfortunate that some of these comparisons have been based on less than adequate chromatographic



**Figure 9** HPLC/ESI-MS analysis of glycerophospholipids from a total lipid extract of guinea pig brain. (A) Total ion chromatogram recorded in negative ionization mode showing separation of phospholipid classes; (B) mass spectrum averaged over the diradylglycerophosphoethanolamine peak displaying  $[M-H]^-$  ions of molecular species and the identity of the major species. Std, internal standard (1,2-dimyristoyl-*sn*-glycero-3-phosphoethanolamine). (Reproduced from Kurvinen J-P, Aaltonen J, Kuksis A, and Kallio H (2002) Software algorithm for automatic interpretation of mass spectra of glycerolipids. *Rapid Communications in Mass Spectrometry* 16: 1812–1820.)

analyses. When high-quality chromatography results have been compared to modern MS/MS analyses, closely similar results have been obtained. Table 2 demonstrates that the composition of molecular species of plasma PtdIns obtained by polar capillary GC matches closely that obtained by normal phase LC/ESI-MS. Similar close agreement has been realized between the GC analyses of the molecular species of the diradylglycerols derived from plasma choline and ethanolamine phospholipids and the molecular species of these phospholipids derived from normal phase LC/ESI-MS analyses. There was no evidence that the preliminary isolation of the phospholipid classes by TLC and subsequent subfractionation of

the diradylglycerol subclasses by normal phase HPLC would have resulted in significant losses of the polyunsaturated species. The MS/MS analyses suffer from inability to cope adequately with the identification of isobaric species and the sodium adducts (positive ion mode), the masses of which can be mistaken for molecular species having two additional methylene units and two double bonds. The MS/MS assays are also unable to cope effectively with the quantification of reverse isomers and are not at all able to distinguish between enantiomers. Nevertheless, the high speed of MS/MS analyses and minimal sample requirements provide a definite advantage for studies in lipidomics (see below).

**Table 1** An example of a single input file generated by the interpretation as a result of processing the molecular mass distribution spectrum of a diradylglycerophosphoethanolamine

Identified compounds					Molecular species			
MW	m/z	1	2	3	4	nmol ml <sup>-1</sup>	μg ml <sup>-1</sup>	mol%
673.92	673.1	#	32:1p	32:2a	#	0.33	0.22	0.08
699.95	698.8	#	34:2p	34:3a	#	0.87	0.61	0.21
701.97	701.1	#	34:1p	34:2a	#	22.59	15.85	5.38
703.99	703	#	34:0p	34:1a	#	1.27	0.9	0.3
718.01	717	34:1	#	#	#	7.6	5.46	1.81
723.98	722.97	#	36:4p	36:5a	#	8.82	6.38	2.1
728.01	727	#	36:2p	36:3a	#	23.59	17.18	5.62
730.02	729.0	#	36:1p	36:2a	#	26.24	19.15	6.25
744.05	742.87	36:2:	#	#	#	2.19	1.63	0.52
746.07	745	36:1	#	#	#	8.53	6.36	2.03
748.08	747.06	36::0	38:6p	#	#	9.09	6.8	2.16
750.02	749.1	#	38:5p	38:6a	#	33.2	24.9	7.91
752.03	751.2	#	38:4p	38:5a	#	57.97	43.6	13.81
754.05	753.1	#	38:3p	38:4a	#	3.41	2.57	0.81
756.06	754.89	#	38:2p	38:3a	#	9.4	7.1	2.24
758.08	757.3	#	38:1p	38:2a	#	1.29	0.98	0.31
764.0	763.3	38:6	#	#	#	3.99	3.05	0.95
766.0	765.1	38:5	#	#	#	7.83	6	1.87
768.07	767.1	38:4	#	#	#	39.51	30.35	9.41
772.11	770.7	38:2	#	#	#	0.54	0.42	0.13
774.12	773.27	38:1	#	#	#	2.18	1.69	0.52
776.14	775.06	38:0	40:6p	#	#	22.08	17.14	5.26
778.07	776.99	#	40:5p	40:6a	#	47.67	37.09	11.36
780.09	778.91	#	40:4p	40:5a	#	28.67	22.36	6.83
792.1	791	40:6	#	#	#	11.13	8.82	2.65
794.11	793.3	40:5	#	#	#	30.45	24.18	7.26
796.13	795.1	40:4	#	#	#	9.29	7.4	2.21

Identifications of  $[M - H]^-$  ions of molecular species are given according to the selected  $m/z$  range. Data in range  $m/z$  745–780 are shown. Molecular species are identified by number of acyl carbons:number of double bonds. p, alkenylacyl; a, alkylacyl; #, disabled. (Reproduced with permission from Kurvinen J-P, Aaltonen J, Kuksis A, and Kallio H (2002) Software algorithm for automatic interpretation of mass spectra of glycerolipids. *Rapid Communications in Mass Spectrometry* 16: 1812–1820.)

Both LC–MS and MS/MS have permitted greatly improved analyses of various lipid oxidation products in the form of the intact neutral and polar lipid molecules or their partial degradation products, which was not possible by chromatographic methods alone. Thus, the hydroperoxides, epoxides, hydroxides, isoprostanes, and the core aldehydes and acids generated during nonenzymatic peroxidation have been identified in plasma lipoproteins and atheroma samples and have provided a new basis for hypotheses about the origin and progression of vascular disease.

Current LC–MS or MS/MS, however, has not been found adequate for the analysis of the polyphosphates of PtdIns, although PtdIns itself has been readily analyzed (see above). This is due to the extremely low concentrations in tissue and cell extracts of the PtdIns polyphosphates, which must be detected in their radioactive form. This can be best done by radio-chromatography on anion-exchange HPLC columns with standards serving as chromatographic carriers.

Nuclear magnetic resonance (NMR) spectroscopy is a powerful, nonselective, and nondestructive technique that provides an alternative method for analysis of components in complex mixtures of lipids. Specifically, lipophilic compounds may be determined either by  $^1\text{H}$  NMR, which offers high sensitivity but limited resolution, or by  $^{13}\text{C}$  NMR, which enables a much more detailed analysis of various lipids owing to its higher resolution, but suffers from the drawback of the low natural abundance of the isotope. Recently, the sensitivity of localized  $^{13}\text{C}$  NMR has been improved up to fivefold by applying proton decoupling and nuclear Overhauser enhancement in experiments on human tissues. Thus, lipid extracts of liver tissue have been analyzed by  $^{13}\text{C}$  and  $^1\text{H}$  NMR at 9.4 T. The spectra allowed the measurement of the free fatty acid to total fatty acid chain ratio, the polyunsaturated to monounsaturated lipid ratio, the glycerophospholipid to triacylglycerol ratio, the total cholesterol to total fatty acid molar ratio, the acylcholesterol to cholesterol ratio, the PtdCho to PtdEtn molar ratio, and the unsaturation

**Table 2** Quantitative recoveries of molecular species of plasma PtdIns by normal phase LC/ESI-MS of intact phospholipids and by high temperature polar capillary GLC/FID of derived diacylglycerol TMS ethers (mol%)

<i>m/z</i>	<i>Molecular species</i>	<i>LC/ESI-MS</i> <sup>a</sup>	<i>LC/ESI-MS</i> <sup>b</sup>	<i>GC/FID</i> <sup>c</sup>
781	16:0/14:0	1.8±0.1	0.5	0.0
833	16:0/18:2	3.4±0.2	3.7	4.1±0.2
859	16:0/20:3	5.0±0.5	3.2	3.5±0.1
857	16:0/20:4	2.8±0.2	5.0	4.2±0.2
865	18:0/18:0	1.7±0.1	0.4	0.2±0.1
863	18:0/18:1	5.4±0.5	4.6	5.1±0.5
861	18:1/18:1			2.5±0.1
861	18:0/18:2	11.9±1.0	11.6	11.6±1.5
859	18:0/18:3	2.7±0.1		3.5±0.2
889	18:0/20:2	2.5±0.2		0.2±0.1
887	18:0/20:3	13.6±1.5	13.9	10.1±1.5
885	18:0/20:4	44.3±2.5	46.8	46.5±2.8
883	18:1/20:4		3.1	3.9±0.2
913	18:0/22:4	2.7±0.1	1.4	0.8±0.2
911	18:0/22:5	1.0±0.2	2.7	2.8±0.2
909	18:0/22:6		3.1	1.0±0.5
905	20:0/20:4	0.7±0.2		
		100	100	100

<sup>a</sup>Kuksis and Pruzanski (2003), unpublished results.

<sup>b</sup>Recalculated from Uran *et al.* (2001) *Journal of Chromatography B* 758: 265–275.

<sup>c</sup>Recalculated from Myher JJ, Kuksis A, and Pind S (1989) Molecular species of glycerophospholipids and sphingomyelins of human plasma: comparison to red blood cells. *Lipids* 24: 408–418.

ratio of the fatty acid chains. Other methods, such as infrared, Fourier transform infrared, and optical rotation have also provided critical data for the identification and quantification of specific lipid classes and molecular species but have been employed to a limited extent only because of lack of general applicability.

## Emerging Techniques

The lipid analyses for the immediate future require increased speed of analysis and automated data processing. The demands of lipidomics and studies on lipid metabolome are such that they can be met only by hundreds or even thousands of parallel analyses of minute samples. This requirement is clearly beyond the capability of the current instrumentation and methodology, although some progress has already been made. Multiparallel analyses of mRNA and proteins are central to present day functional genomics initiatives. Metabolite profiling is a new tool for a comparative display of gene function, as it has the potential to provide deeper insight into complex regulatory processes as well as to determine the phenotype directly. Recently, GC–MS has been used to quantify over 300 distinct compounds from *Arabidopsis thaliana* leaf extracts. Comparisons of four *Arabidopsis* genotypes (two homozygous ecotypes and a mutant of each ecotype) showed that each

genotype possesses a distinct metabolic profile. Others have demonstrated that quantification of the change of several metabolite concentrations relative to the concentration change of one selected metabolite can reveal the site of action, in a metabolic network, of a silent gene. A similar approach to analysis of phenotype of silent mutations in yeast has been attempted by LC/ESI-MS of the molecular species of the major glycerophospholipids. Another approach to an assessment of the lipid metabolome (the concentration of each lipid class and each of its constituent fatty acids) has been taken by a combination of TLC resolution of the lipid classes followed by capillary GC of the fatty acids. Application of the method to evaluate the effect of chronic feeding of a low dose of rosiglitazone on lipid metabolism of diabetic F1 males revealed key targets of the action of the drug on lipid metabolism. It may be recalled that only a slightly less sophisticated approach to lipid profiling was described in 1966, where a packed GC column was used for the fatty acid analyses of the lipid classes resolved by TLC. Since repeated attempts to integrate TLC with MS/MS have been unsuccessful, a substitution of normal phase HPLC for the TLC step would appear to be clearly in order. This would allow integration with MS and analyses of the component fatty acids in the form of the molecular species of the various lipid esters. Except for the LC/ESI-MS analyses of the yeast metabolome, the above analyses do not include determination of

molecular species, which provide the basis of the lipid–protein interaction and metabolic function of glycerolipids. Clearly, future analyses of the lipid metabolome of any biological system will require analyses of molecular species, which may very well be provided by instant MS/MS or MS/MS/MS analyses with computerized data assessment that would yield quantitative estimates of the various lipid classes, the component molecular species, and detailed fatty acid composition. Furthermore, the determination of the molecular species of the neutral glycerolipids, in the future, will require the quantification of the reverse isomers as well as the enantiomers. Thus, triacylglycerols in lipid-rich lipoproteins could have come from the meal itself, could have been re-esterified in the intestine, or could have arisen from *de novo* synthesis in the liver. Only chiral chromatography of enantiomeric diacylglycerols has been able to distinguish among these triacylglycerol species. It is hoped that automated radio-chromatographic systems will be developed for the profiling of the PtdIns polyphosphates, which serve as lipid messengers, protein receptors, and membrane anchors.

See also: **Lipids:** Fatty Acids; Polar Lipids; Determination in Biological Fluids. **Liquid Chromatography:** Overview. **Mass Spectrometry:** Overview. **Nuclear Magnetic Resonance Spectroscopy:** Overview. **Thin-Layer Chromatography:** Overview.

## Further Reading

Byrdwell WC and Neff WE (2002) Dual parallel electrospray ionization and atmospheric pressure chemical

ionization mass spectrometry (MS), MS/MS and MS/MS/MS for the analysis of triacylglycerols and triacylglycerol oxidation products. *Rapid Communications in Mass Spectrometry* 16: 300–319.

Christie WW (2003) *Lipid Analysis. Isolation, Separation, Identification and Structural Analysis of Lipids*, 3rd edn. Bridgewater, UK: The Oily Press, PJ Barnes & Associates.

Kuksis A (1966) Quantitative lipid analyses by combined thin-layer and gas–liquid chromatographic systems. *Chromatography Reviews* 8: 172–207.

Kuksis A (2001) Plasma non-cholesterol sterols. *Journal of Chromatography A* 935: 203–236.

Kuksis A (2003) *Inositol Phospholipid Metabolism and Phosphatidylinositol Kinases*, vol. 30. Laboratory Techniques in Biochemistry and Molecular Biology. Amsterdam: Elsevier Science Publishers.

Kuksis A, Kamido H, and Ravandi A (2003) Glycerophospholipid core aldehydes: Mechanism of formation, methods of detection, natural occurrence, and biological significance. In: Kamal-Eldin A (ed.) *Lipid Oxidation Pathways*, pp. 138–189. Champaign: AOCS Press.

Murphy RC, Fiedler J, and Hevko J (2001) Analysis of non-volatile lipids by mass spectrometry. *Chemical Reviews* 101: 479–526.

Pool CF (2000) Progress in packed column supercritical fluid chromatography: Materials and methods. *Journal of Biochemical and Biophysical Methods* 43: 3–23.

Schiller J and Arnold K (2002) Application of high resolution  $^{31}\text{P}$  NMR spectroscopy to the characterization of the phospholipid composition of tissues and body fluids – a methodological review. *Medical Science Monitor* 8(11): MT205–222.

Watkins SM, Reifsnnyder PR, Pan H-j, Bruce-German J, and Leiter EH (2002) Lipid metabolome-wide effects of the PPAR $\gamma$  agonist rosiglitazone. *Journal of Lipid Research* 43: 1809–1817.

## Fatty Acids

I Brondz, University of Oslo, Oslo, Norway

© 2005, Elsevier Ltd. All Rights Reserved.

## Trivial Names and Systematic Nomenclature

Fatty acids are known by several names: volatile fatty acids from  $\text{C}_1$  to  $\text{C}_5$ , fatty acids from  $\text{C}_6$  to  $\text{C}_{24}$ , long-chain fatty acids from  $\text{C}_{25}$  to  $\text{C}_{40}$ , and very long chain fatty acids above  $\text{C}_{40}$ . One can also find other names in the chemical nomenclature. The correct name for this group of carboxylic aliphatic acids (with or without any additional functions) is simply,

fatty acids. It is possible to find trivial, Liege, and Geneva nomenclatures remaining in modern texts, despite the acceptance of the IUPAC nomenclature in 1957. Fatty acids are carboxylic aliphatic acids with the general formula,  $\text{H}(\text{CH}_2)_n\text{COOH}$ . Aliphatic fatty acids are named by dropping the final ‘e’ from the parent alkane, and then adding the term ‘-oic acid’ to the root. The nomenclature for any additional functions on the principal chain follows the rules of the IUPAC convention of 1957. The first members in the homologous series are usually named using their trivial names (Figure 1).

Fatty acids can have different chemical functional groups substituting for H- in the aliphatic principal chain, leading to different classes of fatty acids.



molecular species, which provide the basis of the lipid–protein interaction and metabolic function of glycerolipids. Clearly, future analyses of the lipid metabolome of any biological system will require analyses of molecular species, which may very well be provided by instant MS/MS or MS/MS/MS analyses with computerized data assessment that would yield quantitative estimates of the various lipid classes, the component molecular species, and detailed fatty acid composition. Furthermore, the determination of the molecular species of the neutral glycerolipids, in the future, will require the quantification of the reverse isomers as well as the enantiomers. Thus, triacylglycerols in lipid-rich lipoproteins could have come from the meal itself, could have been re-esterified in the intestine, or could have arisen from *de novo* synthesis in the liver. Only chiral chromatography of enantiomeric diacylglycerols has been able to distinguish among these triacylglycerol species. It is hoped that automated radio-chromatographic systems will be developed for the profiling of the PtdIns polyphosphates, which serve as lipid messengers, protein receptors, and membrane anchors.

*See also:* **Lipids:** Fatty Acids; Polar Lipids; Determination in Biological Fluids. **Liquid Chromatography:** Overview. **Mass Spectrometry:** Overview. **Nuclear Magnetic Resonance Spectroscopy:** Overview. **Thin-Layer Chromatography:** Overview.

## Further Reading

- Byrdwell WC and Neff WE (2002) Dual parallel electrospray ionization and atmospheric pressure chemical ionization mass spectrometry (MS), MS/MS and MS/MS/MS for the analysis of triacylglycerols and triacylglycerol oxidation products. *Rapid Communications in Mass Spectrometry* 16: 300–319.
- Christie WW (2003) *Lipid Analysis. Isolation, Separation, Identification and Structural Analysis of Lipids*, 3rd edn. Bridgewater, UK: The Oily Press, PJ Barnes & Associates.
- Kuksis A (1966) Quantitative lipid analyses by combined thin-layer and gas–liquid chromatographic systems. *Chromatography Reviews* 8: 172–207.
- Kuksis A (2001) Plasma non-cholesterol sterols. *Journal of Chromatography A* 935: 203–236.
- Kuksis A (2003) *Inositol Phospholipid Metabolism and Phosphatidylinositol Kinases*, vol. 30. Laboratory Techniques in Biochemistry and Molecular Biology. Amsterdam: Elsevier Science Publishers.
- Kuksis A, Kamido H, and Ravandi A (2003) Glycerophospholipid core aldehydes: Mechanism of formation, methods of detection, natural occurrence, and biological significance. In: Kamal-Eldin A (ed.) *Lipid Oxidation Pathways*, pp. 138–189. Champaign: AOCS Press.
- Murphy RC, Fiedler J, and Hevko J (2001) Analysis of non-volatile lipids by mass spectrometry. *Chemical Reviews* 101: 479–526.
- Pool CF (2000) Progress in packed column supercritical fluid chromatography: Materials and methods. *Journal of Biochemical and Biophysical Methods* 43: 3–23.
- Schiller J and Arnold K (2002) Application of high resolution  $^{31}\text{P}$  NMR spectroscopy to the characterization of the phospholipid composition of tissues and body fluids – a methodological review. *Medical Science Monitor* 8(11): MT205–222.
- Watkins SM, Reifsnnyder PR, Pan H-j, Bruce-German J, and Leiter EH (2002) Lipid metabolome-wide effects of the PPAR $\gamma$  agonist rosiglitazone. *Journal of Lipid Research* 43: 1809–1817.

## Fatty Acids

I Brondz, University of Oslo, Oslo, Norway

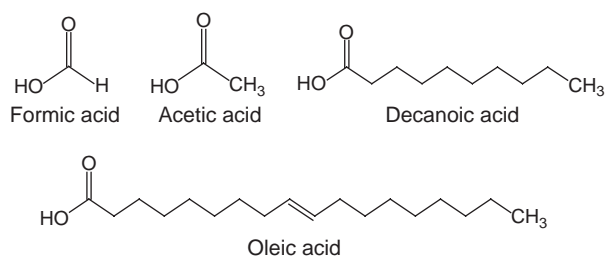
© 2005, Elsevier Ltd. All Rights Reserved.

## Trivial Names and Systematic Nomenclature

Fatty acids are known by several names: volatile fatty acids from  $\text{C}_1$  to  $\text{C}_5$ , fatty acids from  $\text{C}_6$  to  $\text{C}_{24}$ , long-chain fatty acids from  $\text{C}_{25}$  to  $\text{C}_{40}$ , and very long chain fatty acids above  $\text{C}_{40}$ . One can also find other names in the chemical nomenclature. The correct name for this group of carboxylic aliphatic acids (with or without any additional functions) is simply,

fatty acids. It is possible to find trivial, Liege, and Geneva nomenclatures remaining in modern texts, despite the acceptance of the IUPAC nomenclature in 1957. Fatty acids are carboxylic aliphatic acids with the general formula,  $\text{H}(\text{CH}_2)_n\text{COOH}$ . Aliphatic fatty acids are named by dropping the final ‘e’ from the parent alkane, and then adding the term ‘-oic acid’ to the root. The nomenclature for any additional functions on the principal chain follows the rules of the IUPAC convention of 1957. The first members in the homologous series are usually named using their trivial names (Figure 1).

Fatty acids can have different chemical functional groups substituting for H- in the aliphatic principal chain, leading to different classes of fatty acids.



**Figure 1** Aliphatic saturated and monounsaturated fatty acids.

## The Discovery and History of Use of Fatty Acids and Their Distribution in Nature

A narration on the use of fatty acids is contained in ancient historical texts, such as those on the history of the war of Rome with Hannibal in about 200 BC. The Phoenicians made soap by heating animal fat or vegetable oil with wood ash more than 2000 years ago. The alkali salts of fatty acids obtained from the hydrolysis of fats and oils are known as soaps.

Formic acid was first described by S. Fisher in 1670, who obtained it from the distillate of ants. The first nine members of the aliphatic carboxylic acids homologous series are liquids at room temperature. The first solid aliphatic carboxylic acid is decanoic acid, which has a melting point of 31.5°C. An unsaturated principal chain considerably lowers the melting point temperature. For example, *cis*-oleic acid is a liquid at 5°C, and the *cis* and *trans* isomers have notably different melting temperatures. The *trans* isomer of oleic acid is eladic acid, which has a melting point of 44–45°C. The first three members of the aliphatic carboxylic acid homologous series are easily soluble in water.

Fatty acids exist in nature as pure free fatty acids (FFAs), or as part of more complex molecules, known as lipids. Those fatty acids that are covalently bonded to proteins, alcohols, sugars, or other organic molecules are termed ‘bound fatty acids’ (BFAs). There are also a number of synthetic molecules that can be classified as substituted fatty acids or synthetic fatty acids (SFAs), such as 2-(*p*-chlorophenoxy)-2-methylpropionic acid.

Formic acid, a C<sub>1</sub> fatty acid, is found in ants, and acetic acid, a C<sub>2</sub> fatty acid, is found in some scorpions and plants, such as in stinging nettle. The C<sub>4</sub>–C<sub>16</sub> fatty acids are found in different fats and oils. Pelargonic acid, C<sub>9</sub>H<sub>18</sub>O<sub>2</sub>, is part of an ester found in the plant, pelargonium. Capric acid, C<sub>10</sub>H<sub>20</sub>O<sub>2</sub>, was hydrolyzed from *Cuphea llavea* Llave et Lex, and *Lythaceae* seed oil. *Myristicaceae* are known to contain myristic acid. Behenic, lignoceric, and

hexacosanoic acids are usually components of animal and human serum. Most of the natural FAs have an odd number of carbon atoms in their chain, while some natural fatty acids and SFAs may have an even number of carbon atoms in their chain. Branched fatty acids are common in microorganisms, and they can be found with between one and upwards of 20 carbon atoms in their branches.

Branched fatty acids exist as either *iso*-acids or *anteiso*-acids, where the number of branches can vary from one (monobranched) to many (multibranched). There are a number of functional groups that may be present in the principal aliphatic chain of fatty acids, such as hydroxyl-, epoxy-, keto-, fluoride (e.g.,  $\omega$ -fluoro- $\Delta^9$ -octadecanoic acid), chloride, bromide, and iodine, and as sulfur (e.g., in tetradecylthioacetic acid), cyclopropane and cyclopropene rings, and monoenoic, polyenoic, or acetylenic (triple) bonds, and cyclopentane rings (e.g., in prostaglandin E<sub>1</sub>). Other functional groups that may be present include cyclopentene rings (e.g., in chaulmoogric acid), furanoid rings (e.g., in furanoid acid F<sub>6</sub>), and cyclohexene rings (e.g., in  $\alpha$ -retinoic acid). Two more groups are phenyl rings in some synthetic pesticides (e.g., in  $\alpha$ -(2,4,5-trichlorophenoxy)propionic acid, or in the drug ibuprofen ( $\alpha$ -methyl-4-(2-methylpropyl)benzeneacetic acid), and naphthyl rings as in the drug naproxen ([S]-6-methoxymethyl-2-naphthaleneacetic acid). Fatty acids may have unsaturation in chains or in rings. The character of the unsaturated chains can be single, non-conjugated polyunsaturated bonds, conjugated polyunsaturated bonds, or polyunsaturated with cumulative bonds. The unsaturated fatty acids can be in either the *trans* or *cis* configuration.

## Normal Saturated Fatty Acids

Aliphatic fatty acids have the general formula, H(CH<sub>2</sub>)<sub>*n*</sub>COOH, where *n* ranges from 0 to over 40 (Table 1).

## Monobranched Saturated Fatty Acids

Monobranched *iso*-fatty acids have the general formula, (CH<sub>3</sub>)<sub>2</sub>CH(CH<sub>2</sub>)<sub>*n*</sub>COOH (Table 2); the *anteiso*-fatty acids have the general formula, CH<sub>3</sub>CH<sub>2</sub>CHCH<sub>3</sub>(CH<sub>2</sub>)<sub>*n*</sub>COOH; and in addition, there are branched *anteiso*-fatty acids with branches occurring in the middle of the principal chain, which have the general formula, CH<sub>3</sub>(CH<sub>2</sub>)<sub>*n*</sub>CHCH<sub>3</sub>(CH<sub>2</sub>)<sub>*m*</sub>COOH. In nature, the D-form is the dominant form.

**Table 1** Saturated aliphatic fatty acids

<i>Common name</i>	<i>Systematic nomenclature name</i>	<i>Gross formula</i>	<i>Molecular weight</i>	<i>Boiling point (°C)</i>
Methanoic acid	Formic acid	CH <sub>2</sub> O <sub>2</sub>	46.03	100.7
Ethanoic acid	Acetic acid	C <sub>2</sub> H <sub>4</sub> O <sub>2</sub>	60.05	117.9 <sup>a</sup>
Propanoic acid	Propionic acid	C <sub>3</sub> H <sub>6</sub> O <sub>2</sub>	74.08	140.99 <sup>a</sup>
Butanoic acid	Butyric acid	C <sub>4</sub> H <sub>8</sub> O <sub>2</sub>	88.12	163.53 <sup>a</sup>
Pentanoic acid	Valerianic acid	C <sub>5</sub> H <sub>10</sub> O <sub>2</sub>	102.13	186.05 <sup>a</sup>
Hexanoic acid	Caproic acid	C <sub>6</sub> H <sub>12</sub> O <sub>2</sub>	116.16	205 <sup>a</sup>
Heptanoic acid	Enanthic acid	C <sub>7</sub> H <sub>14</sub> O <sub>2</sub>	130.19	223 <sup>a</sup>
Octanoic acid	Caprylic acid	C <sub>8</sub> H <sub>16</sub> O <sub>2</sub>	144.22	239.3 <sup>a</sup>
Nonanoic acid	Pelargonic acid	C <sub>9</sub> H <sub>18</sub> O <sub>2</sub>	158.24	255 <sup>a</sup>
Decanoic acid	Capric acid	C <sub>10</sub> H <sub>20</sub> O <sub>2</sub>	172.27	270 <sup>a</sup>
Undecanoic acid	Undecylic acid	C <sub>11</sub> H <sub>22</sub> O <sub>2</sub>	186.30	280 <sup>a</sup>
Dodecanoic acid	Lauric acid	C <sub>12</sub> H <sub>24</sub> O <sub>2</sub>	200.33	225 <sup>b</sup>
Tridecanoic acid		C <sub>13</sub> H <sub>26</sub> O <sub>2</sub>	214.35	236 <sup>b</sup>
Tetradecanoic acid	Myristic acid	C <sub>14</sub> H <sub>28</sub> O <sub>2</sub>	228.38	250.5 <sup>b</sup>
Pentadecanoic acid		C <sub>15</sub> H <sub>30</sub> O <sub>2</sub>	242.41	257 <sup>b</sup>
Hexadecanoic acid	Palmitic acid	C <sub>16</sub> H <sub>32</sub> O <sub>2</sub>	256.43	390 <sup>a</sup>
Heptadecanoic acid	Margaric acid	C <sub>17</sub> H <sub>34</sub> O <sub>2</sub>	270.46	227 <sup>b</sup>
Octadecanoic acid	Stearic acid	C <sub>18</sub> H <sub>36</sub> O <sub>2</sub>	284.50	360 d <sup>c</sup>
Nonadecanoic acid		C <sub>19</sub> H <sub>38</sub> O <sub>2</sub>	298.52	297 <sup>b</sup>
Eicosanoic acid	Arachidic acid	C <sub>20</sub> H <sub>40</sub> O <sub>2</sub>	312.54	328 d <sup>c</sup>
Heneicosanoic acid		C <sub>21</sub> H <sub>42</sub> O <sub>2</sub>	326.57	
Docosanoic acid	Behenic acid	C <sub>22</sub> H <sub>44</sub> O <sub>2</sub>	340.60	306 <sup>d</sup>
Tricosanoic acid		C <sub>23</sub> H <sub>46</sub> O <sub>2</sub>		
Tetracosanoic acid	Lignoceric acid	C <sub>24</sub> H <sub>48</sub> O <sub>2</sub>		
Pentacosanoic acid	Hyenic acid	C <sub>25</sub> H <sub>50</sub> O <sub>2</sub>	382.68	
Hexacosanoic acid	Cerotic acid	C <sub>26</sub> H <sub>52</sub> O <sub>2</sub>	396.71	
Heptacosanoic acid		C <sub>27</sub> H <sub>54</sub> O <sub>2</sub>		
Octacosanoic acid	Montanic acid	C <sub>28</sub> H <sub>56</sub> O <sub>2</sub>	424.76	
Nonacosanoic acid		C <sub>29</sub> H <sub>58</sub> O <sub>2</sub>		
Triacantanoic acid	Melissic acid	C <sub>30</sub> H <sub>60</sub> O <sub>2</sub>		
Hentriacantanoic acid		C <sub>31</sub> H <sub>62</sub> O <sub>2</sub>	466.84	
Dotriacantanoic acid	Lacceroic acid	C <sub>32</sub> H <sub>64</sub> O <sub>2</sub>		

<sup>a</sup>Boiling point at 760 mmHg.<sup>b</sup>Boiling point at 100 mmHg.<sup>c</sup>Decomposes.<sup>d</sup>Boiling point at 60 mmHg.

The molecular weights and boiling points shown are taken from: Lide DR (ed.) (2001–2002) *CRC Handbook of Chemistry and Physics*, 82nd edn., Boca Raton, FL: CRC Press.

**Table 2** Some of the monobranched fatty acids

<i>Iso-acids</i>	<i>Trivial name</i>	<i>Anteiso-acids</i>	<i>Trivial name</i>
3-Methylbutanoic	Isovaleric	2-Methylbutanoic	
4-Methylpentanoic	Isocapric	3-Methylpentanoic	
5-Methylhexanoic		4-Methylhexanoic	
6-Methylheptanoic	Isocaprylic	5-Methylheptanoic	
7-Methyloctanoic		6-Methyloctanoic	
8-Methylnonanoic	Isocapric	7-Methylnonanoic	
9-Methyldecanoic		8-Methyldecanoic	
10-Methylundecanoic	Isolauric	9-Methylundecanoic	
11-Methyldodecanoic		10-Methyldodecanoic	
12-Methyltridecanoic	Isomyristic	11-Methyltridecanoic	
13-Methyltetradecanoic	Isopentadecanoic	12-Methyltetradecanoic	Sarcinic
14-Methylpentadecanoic	Isopalmitic	13-Methylpentadecanoic	
15-Methylhexadecanoic		14-Methylhexadecanoic	
16-Methylheptadecanoic		15-Methylheptadecanoic	
17-Methyloctadecanoic	Tuberculostearic	16-Methyloctadecanoic	
18-Methylnonadecanoic		17-Methylnonadecanoic	
20-Methylheneicosanoic		19-Methylheneicosanoic	
22-Methyltricosanoic		21-Methyltricosanoic	
24-Methylpentacosanoic			
26-Methylheptacosanoic			

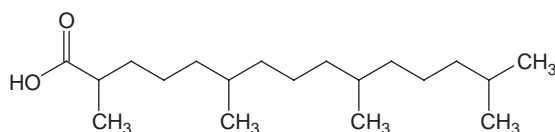
## Multibranched Saturated Fatty Acids

Multibranched saturated fatty acids can have the following configurations:  $R-(CH_2CHCH_3CH_2CH_2)_n-CH_2CHCH_3CH_2COOH$  (where  $R=H$  and  $n=2$  for farnesic acid) or  $R-(CH_2CHCH_3CH_2CH_2)_n-COOH$  (where  $R=H$  and  $n=3$  for 2,8,12-TMTD), or  $R-(CH_2CHCH_3CH_2CH_2)_n-CH_2CHCH_3COOH$  (where  $R=H$  and  $n=3$  for pristanic acid) (Figure 2), and  $R-(CH_2CHCH_3CH_2CH_2)_n-CH_2CHCH_3CH_2COOH$  (where  $R=H$  and  $n=3$  for phytanic acid).

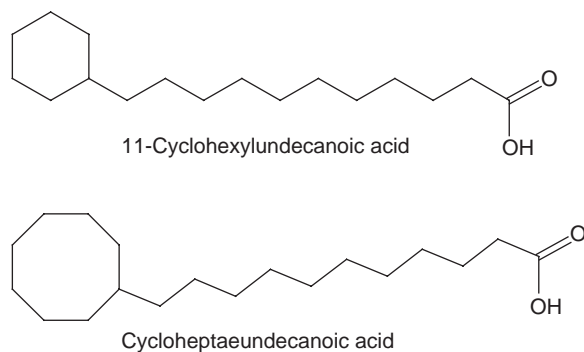
The common carbon sequence in all these molecules has the  $(CH_2CHCH_3CH_2CH_2)$  architecture. Mycrocercosic acid has the formula  $C_{32}H_{64}O_2$  (2,4,6,8-tetramethyloctacosanoic acid) and has four branches in the principal chain. Branched fatty acids (the mycolic acids) are found in bacteria belonging to the *Mycobacterium-Nocardia-Corynebacterium* group. They have often other functional groups in their chain, and should formally belong to the multifunctional fatty acids.

## Saturated Monocarboxylic Fatty Acids with a Cyclohexane or Cycloheptane Ring

Some saturated FAs have a ring system connected to the chain, such as 11-cyclohexylundecanoic acid, 13-cyclohexyltridecanoic acid, and cycloheptaneundecanoic acid (Figure 3).



**Figure 2** Pristanic acid, or 2,6,10,14-tetramethylpentadecanoic acid.



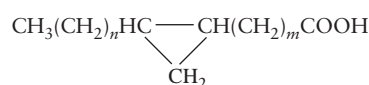
**Figure 3** Monocarboxylic acids with a cyclohexane or cycloheptane ring.

## Partially Unsaturated Multibranched Fatty Acids

This class of fatty acids has the general formula of  $R-(CH_2CH-CH_3CH_2CH_2)_n-CH-C=CH-COOH$ , where  $R=H$ . These saturated and partially unsaturated multibranched fatty acids belong to the isoprenoid acid family, and in nature, isoprenoid acids are often found in microorganisms and plants. Phytanic acid, whose systematic name is 3,7,11,15-tetramethyl-2-hexadecenoic acid (Figure 4), is one member of this class.

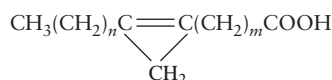
## Cyclopropane and Cyclopropene Fatty Acids

The general formula for fatty acids with a cyclopropane ring is

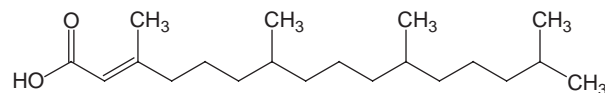


The homologous series with  $n=5$  and  $m=7$  includes *cis*-9,10-methylenehexadecanoic acid and the homologous series with  $n=5$  and  $m=9$  includes lactobacillic acid (*cis*-11,12-methyleneoctadecanoic acid). The homologous series with  $n=7$  and  $m=6$  includes dihydromalvalic acid (systematic name is 2-octyl-cyclopropaneheptanoic acid) (Figure 5) and the homologous series with  $n=7$  and  $m=7$  includes dihydrosterculic acid (systematic name is *cis*-9,10-methyleneoctadecanoic acid).

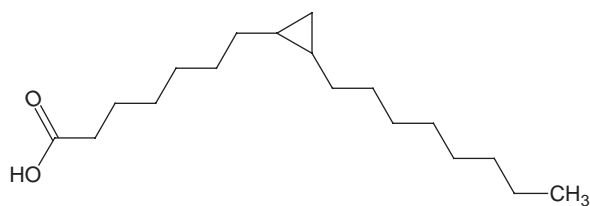
Fatty acids with a cyclopropene ring have the general formula,



Malvalic acid (systematic name is 2-octyl-1-cyclopropene-1-heptanoic acid) has  $n=7$  and  $m=6$



**Figure 4** Phytanic acid, or 3,7,11,15-tetramethyl-2-hexadecenoic acid.



**Figure 5** Dihydromalvalic acid or 2-octyl-cyclopropaneheptanoic acid.

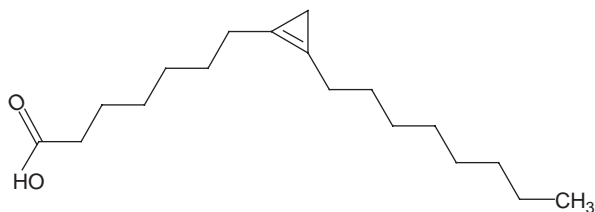
(Figure 6), while sterculic acid (systematic name is *cis*-9,10-methyleneoctadec-9-enoic acid) has  $n=7$  and  $m=7$ .

## Cyclopentene Fatty Acids

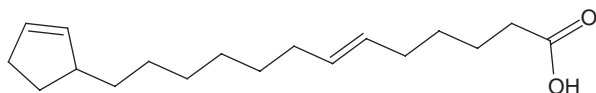
Some of the fatty acids have the general formula,  $R-CH_2(CH_2)_nCOOH$ , where R is a cyclopentene ring. Examples are, alepric acid ( $C_{14}H_{24}O_2$ ), hydrocarpic acid ( $C_{16}H_{28}O_2$ ), and chaulmoogric acid ( $C_{18}H_{32}O_2$ ). Gorlic acid (systematic name is 13-(2-cyclopenten-1-yl)-6-tridecenoic acid) (Figure 7) has an unsaturation in principal chain, with the general formula  $R-(CH_2)_6CH=CH(CH_2)_4COOH$ , where R is a cyclopentene ring.

## Normal Monounsaturated Fatty Acids

The family of monounsaturated fatty acids (Table 3) with one double bond (general formula  $CH_3(CH_2)_nCH=CH(CH_2)_mCOOH$ ), may exist as



**Figure 6** Malvalic acid or 2-octyl-1-cyclopropene-1-heptanoic acid.



**Figure 7** Gorlic acid or 13-(2-cyclopenten-1-yl)-6-tridecenoic acid.

**Table 3** Some of the monounsaturated fatty acids

Systematic name	Trivial name	Abbreviation
<i>Trans</i> -2-butenoic	$\alpha$ -Crotonic	<i>Trans</i> -4:1 (2)
<i>Cis</i> -5-dodecenoic	Denticetic	<i>Cis</i> -12:1 (5)
<i>Cis</i> -9-dodecenoic	Lauroleic	<i>Cis</i> -12:1 (9)
<i>Cis</i> -5-tetradecenoic	Physeteric	<i>Cis</i> -14:1 (5)
<i>Cis</i> -9-tetradecenoic	Myristioleic	<i>Cis</i> -14:1 (9)
<i>Cis</i> -9-hexadecenoic	Palmitoleic	<i>Cis</i> -16:1 (9)
<i>Cis</i> -9-octadecenoic	Oleic	<i>Cis</i> -18:1 (9)
<i>Trans</i> -9-octadecenoic	Eladic	<i>Trans</i> -18:1 (9)
<i>Trans</i> -11-octadecenoic	Vaccenic	<i>Trans</i> -18:1 (11)
<i>Cis</i> -9-icosenoic	Gladoleic	<i>Cis</i> -20:1 (9)
<i>Cis</i> -13-docosenoic	Erulic	<i>Cis</i> -22:1 (13)
<i>Trans</i> -13-docosenoic	Brassicidic	<i>Trans</i> -22:1 (13)
<i>Cis</i> -15-tetracosenoic	Nervonic	<i>Cis</i> -24:1 (15)

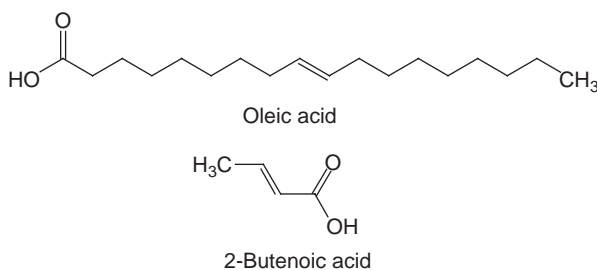
*trans*-isomers, but in nature, the *cis*-isomers predominate.

Some of the monounsaturated fatty acids may also have branched chains, as does, for example, 3-methylcrotonic acid (systematic name is 2-butenic acid) (Figure 8).

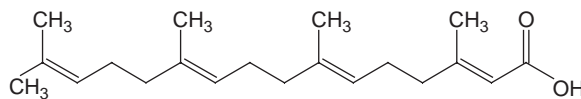
## Polyunsaturated Fatty Acids with and without a Cyclohexene Ring

The polyunsaturated fatty acids family is widely distributed throughout nature. They have several double bonds in the principal chain. Common dienic acids are *cis*-,*cis*-9,12-octadecadienoic acid (linoleic acid), *trans*-,*trans*-9,12-octadecadienoic acid (linoleic acid), 5,9-heneicosadienoic acid, 5,9-docosadienoic acid, and 5,9-tricosadienoic acid.

The common trienoic fatty acids are *cis*-,*cis*-,*cis*-9,12,15-octadecatrienoic acid ( $\alpha$ -linolenic acid), *cis*-,*cis*-,*cis*-6,9,12-octadecatrienoic acid ( $\gamma$ -linolenic acid), *cis*-,*cis*-,*cis*-8,11,14-icosatrienoic acid (dihomo- $\gamma$ -linolenic acid), 5,9,13-eicosatrienoic acid, and 5,9,13-docosatrienoic acid. An example of a tetraenoic fatty acid is *cis*-,*cis*-,*cis*-,*cis*-5,8,11,14-icosatetraenoic acid (arachidonic acid). The discovery of new polyunsaturated acids is an ongoing process, with the recent discovery of 5,9-heneicosadienoic acid, 5,9-docosadienoic acid, 5,9-tricosadienoic acid, trienoic 5,9,13-eicosatrienoic acid, and 5,9,13-docosatrienoic acid from *Stoichactis helianthus*. Geranoic acid (systematic name is 3,7,11,15-tetramethyl-2,6,10,14-hexadecatetraenoic acid) (Figure 9), nerolic and farnesoic acids are polyunsaturated fatty acids with branches in their structures.



**Figure 8** Monounsaturated fatty acids.



**Figure 9** Geranoic acid or 3,7,11,15-tetramethyl-2,6,10,14-hexadecatetraenoic acid.



The principal chain and the ring can both be unsaturated, as is found in  $\alpha$ -retinoic and  $\beta$ -retinoic acids. The systematic name of  $\beta$ -retinoic acid is (all-*E*)-3,7-dimethyl-9-(2,6,6-trimethyl-1-cyclohexen-1-yl)-2,4,6,8-nonatetraenoic acid) (Figure 10). It has four double bonds in the principal chain and one in the cyclohexene ring.

## Acetylenic Fatty Acids

Triple bonds are often found in fatty acids of plant origin. Fatty acids, such as stearolic 9-octadecynoic (Figure 11), behenolic 13-docosynoic, and clupanodonic 4,7,11-docosatrien-18-ynoic acids have a triple bond in their chains. Clupanodonic acid has three double bonds and an acetylenic bond.

## Hydroxy Fatty Acids

Hydroxy fatty acids (HFAs) (Table 4) have hydroxyl functional groups attached to the principal chain, and they can be both saturated and unsaturated. The principal chain can be branched, and these branches are sometimes quite long.

The main positions of attachment of the hydroxyl to the chain are the  $\alpha$ -attachment to the  $C_2$  site, and the  $\beta$ -attachment to the  $C_3$  site. The general formula

for  $\alpha$ -hydroxy fatty acids is  $\text{CH}_3(\text{CH}_2)_n\text{CH}(\text{OH})\text{COOH}$ , while the  $\beta$ -hydroxy fatty acids have the general formula of  $\text{CH}_3(\text{CH}_2)_n\text{CH}(\text{OH})\text{CH}_2\text{COOH}$ .

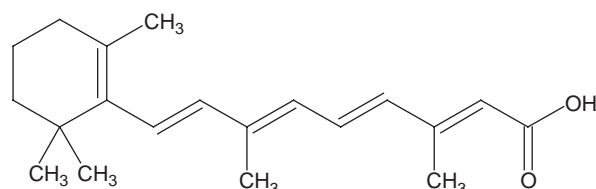
There are two other HFA homologous series: the  $\omega$ - and  $(\omega - 1)$ . The general formula for  $\omega$ -hydroxy fatty acids is  $\text{HOCH}_2(\text{CH}_2)_n\text{COOH}$ ; and the general formula for  $(\omega - 1)$  HFAs is  $\text{CH}_3\text{CH}(\text{OH})(\text{CH}_2)_n\text{COOH}$ . Other functions, such as branches or double bonds, are often found in HFAs, for example, in 3-hydroxy-13-methyltetradecanoic and 3-hydroxy-15-methylhexadecanoic acids. An unsaturated hydroxy fatty acid is ricinoleic acid (12-hydroxy-*cis*-9-octadecenoic acid).

Two or more hydroxy functions can be present in the principal chain, for example, in 9,10-dihydroxyoctadecanoic acid and in ustilic acid (2,15,16-trihydroxyhexadecanoic acid) (Figure 12).

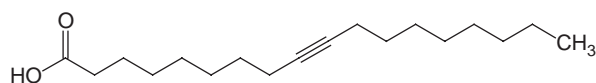
Some of bacterial cell walls and bacterial lipopolysaccharides both contain  $\beta$ -hydroxy branched and unbranched fatty acids.

## Fatty Acids with a Keto Function

The homologous series of keto fatty acids (KFAs) are made up from the  $\alpha$ -,  $\beta$ -, and  $(\omega - 1)$  KFAs. The  $\alpha$ -KFAs have the general formula  $\text{CH}_3(\text{CH}_2)_n\text{COCOOH}$ ; the  $\beta$ -KFAs  $\text{CH}_3(\text{CH}_2)_n\text{COCH}_2\text{COOH}$ ; and the  $(\omega - 1)$  KFAs  $\text{CH}_3\text{CO}(\text{CH}_2)_n\text{COOH}$ .



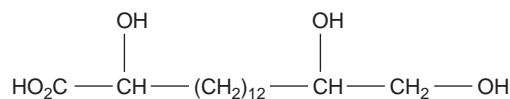
**Figure 10**  $\beta$ -Retinoic acid or (all-*E*)-3,7-dimethyl-9-(2,6,6-trimethyl-1-cyclohexen-1-yl)-2,4,6,8-nonatetraenoic acid.



**Figure 11** Stearolic acid or 9-octadecynoic acid.

## Prostaglandins, Prostacyclin, and Thromboxanes

The prostaglandins, prostacyclin, and the thromboxanes are prostanoids, and from a chemical perspective, the prostanoids are fatty acids. The



**Figure 12** Ustilic acid or 2,15,16-trihydroxyhexadecanoic acid.

**Table 4** Some of the hydroxy fatty acids

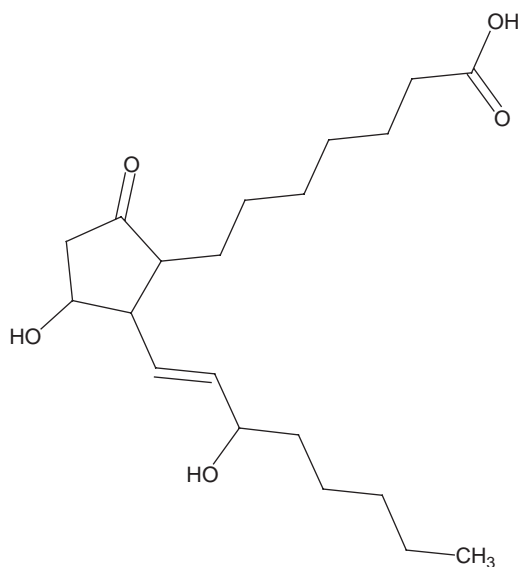
Systematic name	Trivial name	Abbreviation
3-Hydroxybutanoic acid	$\beta$ -Hydroxybutyric acid	3-OH-4:0
2-Hydroxytetradecanoic acid	$\alpha$ -Hydroxymyristic acid	2-OH-14:0
3-Hydroxytetradecanoic acid	$\beta$ -Hydroxymyristic acid	3-OH-14:0
3-Hydroxy-13-methyltetradecanoic acid		3-OH- <i>iso</i> 15
2-Hydroxyhexadecanoic acid	$\alpha$ -Hydroxypalmitic acid	2-OH-16:0
3-Hydroxyhexadecanoic acid	$\beta$ -Hydroxypalmitic acid	3-OH-16:0
3-Hydroxy-15-methylhexadecanoic acid		3-OH- <i>iso</i> 17
3-Hydroxyoctadecanoic acid	$\beta$ -Hydroxystearic acid	3-OH-18:0
17-Hydroxyoctadecanoic acid	17-Hydroxystearic acid	17-OH-18:0

systematic name of prostaglandin E<sub>1</sub> is 3-hydroxy-2-(3-hydroxy-1-octenyl)-5-oxo-cyclopentaneheptanoic acid. Prostanoids have a cyclopentane ring in their chain, as is found, for example, in prostaglandin E<sub>1</sub> (Figure 13) prostaglandin F<sub>1 $\alpha$</sub> , prostaglandin E<sub>2</sub>, prostaglandin E<sub>3</sub>, and in prostacyclin (PGI<sub>2</sub>).

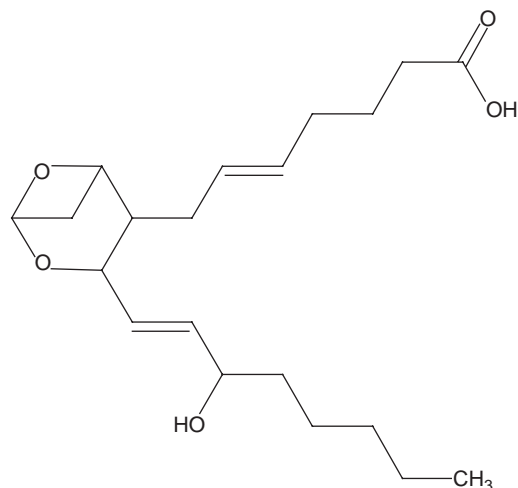
Thromboxane A<sub>2</sub> (Figure 14) and thromboxane B<sub>2</sub> both have a tetrahydropyran system in their chains.

## Furanoid Fatty Acids

Furanoid fatty acids (FuFAs) have a furanoid ring in their principal chain. One example of a furanoid fatty acid has the formula, C<sub>22</sub>H<sub>38</sub>O<sub>3</sub>, and has the systematic name of 3,4-dimethyl-5-pentyl-2-furanundecanoic acid (Figure 15).



**Figure 13** Prostaglandin E<sub>1</sub> or 3-hydroxy-2-(3-hydroxy-1-octenyl)-5,15,16 acid.



**Figure 14** Thromboxane A<sub>2</sub>.

A large variety of FuFAs are commonly found in the lipids derived from fish and invertebrates.

## Some Synthetic Pesticides and Drugs with a Phenyl or a Naphthyl Ring Attached to the Principal Chain of Fatty Acids

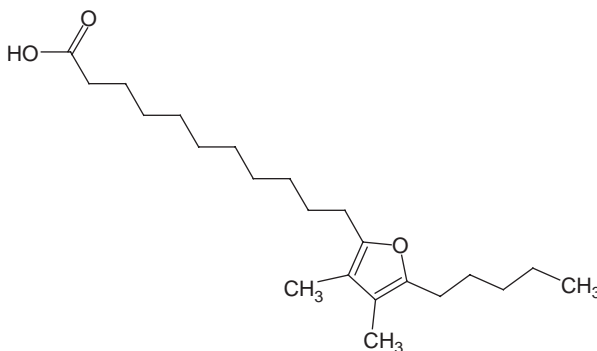
Some synthetic pesticides and drugs have a phenyl or a naphthyl ring attached to the principal chain of fatty acids. The phenoxy acid pesticides, which are derivatives of formic, acetic, propionic, or butyric acids, belong to the synthetic fatty acids (SFAs) family. These acids are usually referred to as phenoxy acids and not as fatty acids, although they should be included in the fatty acids family, because they are derivatives of fatty acids. One such example is (2-methyl-4-chlorophenoxy)acetic acid (Figure 16).

Synthetic drugs with a phenyl or a naphthyl ring system, such as ibuprofen ( $\alpha$ -methyl-4-(2-methyl-propylbenzeneacetic acid) (Figure 17) and naproxen, are also derivatives of fatty acids.

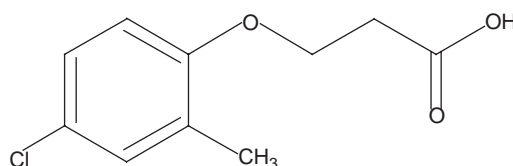
Acitretin (Figure 18) is a synthetic retinoid that is also a fatty acid derivative.

## Fatty Acids with Amino Function in the Principal Chain

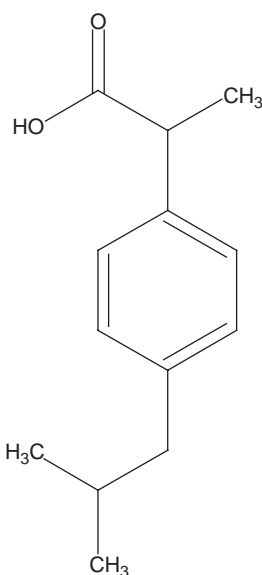
Small molecular weight fatty acids with an amino function in the principal chain as found in proteins and peptides are classified as amino acids. It should



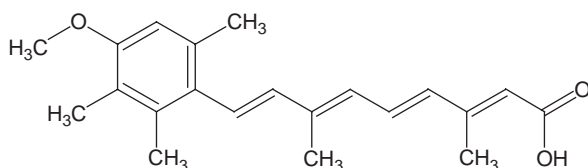
**Figure 15** 3,4-Dimethyl-5-pentyl-2-furanundecanoic acid.



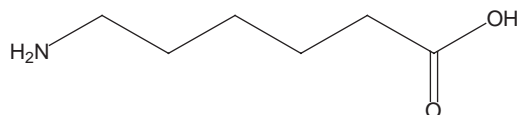
**Figure 16** (2-Methyl-4-chlorophenoxy)acetic acid (MCPA).



**Figure 17** Ibuprofen or  $\alpha$ -methyl-4-(2-methylpropyl)-benzeneacetic acid.



**Figure 18** Acitretin.



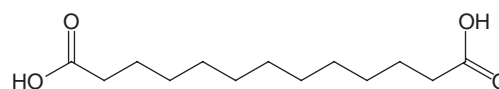
**Figure 19** 6-Amino-*n*-hexanoic acid.

be recognized that non-protein and peptide AAS still belong to the fatty acids family.

The chemical formula for 6-aminohexanoic acid is  $\text{NH}_2(\text{CH}_2)_5\text{COOH}$ , and its common name is  $\alpha$ -aminocaproic acid (systematic name is 6-amino-*n*-hexanoic acid) (Figure 19). It may be referred to as a fatty acid with an amino group in the principal chain.

### Fatty Acids with an Additional Carboxylic Function in the Principal Chain

Some of the small molecular weight dicarboxylic acids belong to the Krebs cycle acid family. There are many dicarboxylic acids in addition to the Krebs cycle acids. Examples of this family are azelaic acid ( $\text{HOOC}(\text{CH}_2)_7\text{COOH}$ ), brassylic acid (systematic



**Figure 20** Brassylic acid or 1,13-tridecanedioic acid.

name is 1,13-tridecanedioic acid), and thapsic acid (systematic name is 1,16-hexadecanedioic acid) (Figure 20).

### Halogenated Fatty Acids

Marine organisms are the source of brominated fatty acids, such as 6-bromo-5,9-heneicosadienoic and 6-bromo-5,9-docosadienoic acids. Other fatty acids with different halogens, such as F, Cl, and I are also present in marine animals.

### Thio Fatty Acids

An example of the thio fatty acid family is tetradecylthioacetic acid ( $\text{CH}_3(\text{CH}_2)_{13}\text{S-CH}_2\text{-COOH}$ ).

### Stability, Sample Storage and Preparation, and Occurrence of Artifacts

Saturated fatty acids are stable in dry, cold atmospheres. However, the presence of additional functions or unsaturated bonds in the fatty acid chains can make them chemically reactive to the oxygen and moisture in ambient air. The general rule is that fatty acid samples should be kept at temperatures between  $-20^\circ\text{C}$  and  $-70^\circ\text{C}$ , and the samples should be protected from exposure to direct light. To protect the samples against oxygen and moisture, they should be stored under  $\text{N}_2$ , He, or Ar in a sealed container. Unprotected polyunsaturated fatty acids will auto-oxidize rapidly. Any peroxides present in the extraction liquids can produce artifacts.

Fatty acids can be extracted either as FFAs or as BFAs. FFAs are obtained by the hydrolysis of BFAs. Liquid-liquid extraction has been successfully applied in the isolation of several classes of lipids, or of single lipids from complex mixtures. Popular extraction methods for lipids are the 'Folch' extraction technique, or the Bligh and Dyer method. Solid-phase extraction (SPE), and solid-phase microextraction (SPME), are also available as simple and economical time- and solvent-efficient sample preparation methods. Prefractionation can be performed using SPE silica or aminopropyl-silica columns.

The argentation SPE method can be used for the isolation of unsaturated FAs. One very useful

extraction method is the supercritical fluid carbon dioxide extraction technique.

### Ultraviolet, Infrared, and Proton Nuclear Magnetic Resonance Spectroscopic Characteristics of Free Fatty Acids

The carboxyl group in aliphatic FA is a poor chromophore with absorption maximum  $\sim 210$  nm in ethanol, with  $\epsilon = 50$ . The UV absorption properties improve with the presence of unsaturated bonds; especially in the case of conjugated double or triple bonds.

The IR absorption of the functional groups in the principal chains of fatty acids follows general rules, and these rules for given groups are summarized in Table 5.

In the nuclear magnetic resonance (NMR) spectra, the proton signal for the terminal methyl group appears at 9.1 ppm, and the proton signal for the methylene groups in the principal chain appears at  $\sim 8.7$  ppm. The proton signals from other functional groups, if present in the principal chain of the molecule, depend on the solvent.

### Analytical Approach to Separation and Structure Determination of Fatty Acids

Chromatography is the most powerful tool for the separation of complex mixtures of either natural or synthetic origin. Retention time is a tool for the identification of compounds of interest. Five different chromatography techniques will be discussed in this section: liquid–liquid separation, thin-layer chromatography (TLC), high-performance liquid chromatography (HPLC), gas chromatography (GC), and supercritical-fluid chromatography (SFC). All these techniques will be discussed in connection with their ability to detect, quantify, and enable structural determination of fatty acids.

Liquid–liquid separation of fatty acids from complex mixtures is the oldest known technique. Hydrolyzed samples are acidified and extracted using a suitable organic phase, which is often diethyl ether.

TLC is used to separate unsaturated fatty acids on plates impregnated with Ag, and is a very popular and useful tool. Less well known is the separation of fatty acids on plates modified with Cu(I), Cu(II), Co(III) ions, and with  $\text{ZnCl}_2$ . TLC may be used as an analytical or preparative tool. Preparative TLC can be used to recover a significant amount of substance for later analysis using other techniques, such as HPLC, GC, or mass spectrometry (MS).

The separation of FFAs using HPLC does not present a problem. A broad spectrum of stationary phases can be used:  $\text{SiO}_2$  phases, stationary phases impregnated with cations, reversed phases, ion-exchange resins, or cellulose. However, detection of underivatized aliphatic fatty acids may be difficult to perform because of a lack of suitable chromophores, luminophores, or fluorophores in the molecule.

Preparation of derivatives to enhance the sensitivity of detectors or to provide useful mass fragmentation for LC–MS analysis may be necessary.

### HPLC Detectors Used in Fatty Acids Analysis

Using an ion-exchange separation in conjunction with a suppressed electrical conductivity detector, based on an injected volume of  $100\ \mu\text{l}$  and a 3:1 signal-to-noise ratio, it is possible to detect haloacetic acids down to the following limits: monochloroacetic acid = 8.0 ppb, dichloroacetic acid = 16.0 ppb, trichloroacetic acid = 80.0 ppb, trifluoroacetic acid = 12.0 ppb, bromoacetic acid = 21.0 ppb, and dibromoacetic acid = 30.0 ppb.

The conductivity detector can be used with gradient elution but is sensitive to pump pulsation, changes in flow rate, and changes in pressure. The analysis should be carried out at constant ionic strength. Refractive index (RI) detectors require isocratic elution. An RI detector is a universal detector.

**Table 5** The IR absorption of the functional groups in the principal chains of fatty acids

Functional groups	Type of absorbance	Frequency ( $\text{cm}^{-1}$ )
Carboxyl $\text{C}=\text{O}$	Normal dimeric carboxylic stretching vibration	1720–1690
$\text{C}-\text{H}$	Stretching vibration superimposed on the $\text{O}-\text{H}$ stretching vibration	2960–2850
$\text{C}-\text{H}$	Bending vibration of the methyl or methylene groups	1375 or 1465
$\text{C}-\text{O}-\text{H}$	In-plane bending vibration	1408
$\text{O}-\text{H}$	Out-of-plane bending vibration of dimeric carboxylic acids	920
$\text{C}=\text{C}$	Unconjugated double bond stretching vibration	1660–1640
$\text{C}\equiv\text{C}$	Triple bond stretching vibration	2260–2100

A new technique developed to detect fatty acids without derivatization, is the use of evaporative light-scattering detector (ELSD). The detector offers several advantages over traditional detectors, as the detection response does not depend on the optical characteristics of the sample. The ELSD can be used with multisolvent gradients. However, the sensitivity is not very high.

Today, ultraviolet detectors (UVDs) are commonly used for the detection of natural and derivatized fatty acids.

The fluorescence detector (FD) is one of the most sensitive detectors in use, and can record both excitation and emission spectra. The excitation spectra are identical to the UV-visible absorption spectra; however, emission spectra can provide additional information. Detectors that are used to detect isotope-labeled molecules measure the radioactivity present. Other analytical instruments, such as IR spectrometers, mass spectrometers, NMR spectrometers, electron spin resonance spectrometers, plasma emission and plasma absorption spectrometers can be connected to HPLCs for use as detectors, to provide further information on molecular structure.

## Supercritical-Fluid Chromatography

From a technical perspective, SFC is similar to HPLC; however, the mobile phase is a supercritical fluid. The SFC method possesses all the advantages of the HPLC method, and some advantages of the GC technique. High-sensitivity detectors used in HPLC and GC can also be used in SFC. However, some restrictions on the detectors arise from the use of condensed gas as the supercritical fluid. The most widely used ELSD.

Modification of the mobile phase with water or an organic solvent can improve the separations. However, it is not possible to use the FID when organic solvent is used as a modifier.

## Derivatization to Enhance Spectrophotometric Detection for HPLC Analysis

The aim of derivatization is to introduce into the molecule a suitable chromophore or fluorophore. Precolumn and postcolumn derivatization are used in HPLC. Some of the reagents, their properties, separation conditions, and reactions are summarized below.

Phenacyl bromide or (2-bromoacetophenone) (PB). Reactions with phenacyl bromide afford quantitative yields, and are easily carried out. Detection

limits are in the nanogram range. Maximum detection wavelength is 242 nm.

A wide choice of stationary phases is available for separation, such as reversed-phase or silver-ion impregnated phases. The separation of phenacyl derivatives of saturated, monoenoic, polyenoic, and monohydroxy fatty acids, and of the geometric isomers of fatty acids is possible.

Other phenacyl reagents are: *p*-bromophenacyl bromide, or  $\alpha$ -bromo-*p*-bromoacetophenone, *p*-chlorophenacyl bromide, or  $\alpha$ -bromo-*p*-chloroacetophenone, *p*-nitrophenacyl bromide, or  $\alpha$ -bromo-*p*-nitroacetophenone, *m*-methoxyphenacyl bromide, or  $\omega$ -bromo-3-methoxyacetophenone.

[*p*-(9-Anthroyloxy)]phenacyl bromide is the reagent of choice, as it has a high detection sensitivity, a detection wavelength maximum of 253 nm, and  $\epsilon$  of  $174,280 \text{ mol}^{-1} \text{ l}^{-1} \text{ cm}^{-1}$ .

## Derivatization of Monohydroxy Fatty Acids for HPLC Analysis

Derivatives of HFAs can be hydroxamic acids, phenacyl derivatives, nitrobenzoyl derivatives, and naphthacyl derivatives. Derivatization of the hydroxy group can be achieved using 1-anthroynitrile. The naphthacyl derivatives can be separated by reversed-phase chromatography and detected at 246 nm, with a detection limit of  $\sim 0.1 \text{ ng}$ .

PBr derivatives, exhibit fluorescence when excited with at 375 nm, emission wavelength at 470 nm. It is possible to achieve femtomolar detection limits by using 4-bromomethyl-7-methoxycoumarin (BrMMC) derivatives and laser excitation.

Other reagents used are: 4-bromomethyl-7-acetoxycoumarin (BrMAC), 3-Bromoacetyl-7-methoxycoumarin (BrAMC). The use of BrAMC allows the derivatization of fatty acids at room temperature. BrAMC has a higher reactivity than other bromomethylcoumarins: 4-bromomethyl-6,7-dimethoxycoumarin, 4-bromomethyl-6,7-methylene-dioxycoumarin, 4-bromomethyl-7,8-benzocoumarin, 4-diazomethyl-7-methoxycoumarin, and 7-(diethyl-amino)-coumarin-3-carbohydrazide.

Separation of  $C_1$  to  $C_{20}$  fatty acids as 9-(2-hydroxyethyl)carbazole derivatives can be achieved by HPLC using gradient elution on a C18 column in a relatively short time. The  $\lambda_{\text{ex}}$  was at 293 nm, and the maximum fluorescence emission for the derivatized fatty acids was at 365 nm. Sensitive detection is possible by fluorescence excitation at  $\lambda_{\text{ex}} = 293$  or 335 nm and emission at 365 nm.

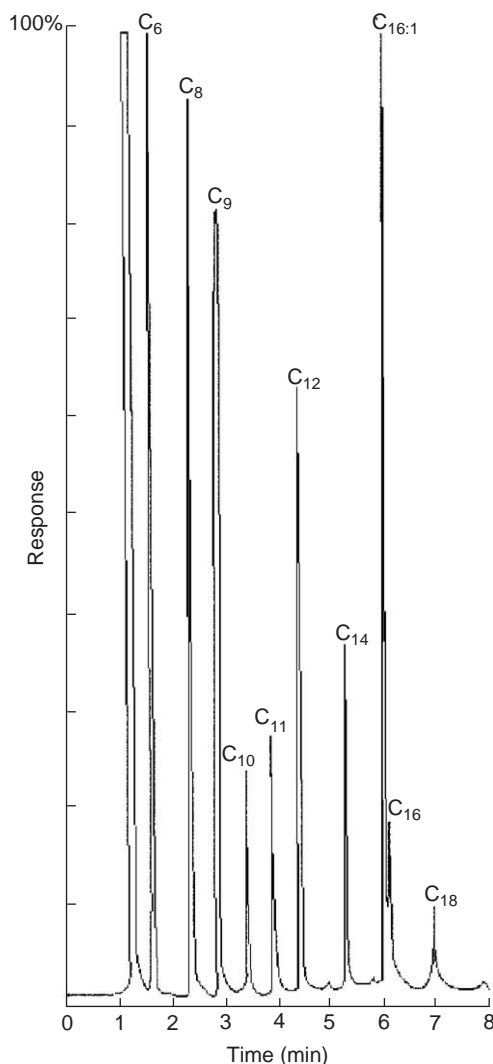
Derivatization with 5-bromomethylfluorescein (5-BMF). Allows use of UV, conductivity, and also laser-induced fluorescence (LIF) detectors.



## GC Analysis of Fatty Acids

The evaporation temperature of high-molecular weight FFAs is close to their decomposition temperatures (see Table 1). Thus, the adsorption of FFAs on the injector or syringe can be a problem as a memory effect. The separation of a broad range of FFAs by GC is shown in Figure 21.

Derivatization of fatty acids is performed to increase the volatility of the products, to reduce dimerization in the vapor phase, to reduce the adhesion to the apparatus' construction materials and columns, to improve separation, and to reduce tailing. The



**Figure 21** GC separation of underivatized saturated and unsaturated FFAs. Peaks: C<sub>6:0</sub> = caproic acid; C<sub>8:0</sub> = caprylic acid; C<sub>9:0</sub> = nonanoic acid; C<sub>10:0</sub> = capric acid; C<sub>11:0</sub> = undecanoic acid; C<sub>12:0</sub> = lauric acid; C<sub>14:0</sub> = myristic acid; C<sub>16:1</sub> = palmitoleic acid; C<sub>16:0</sub> = palmitic acid; C<sub>18:0</sub> = stearic acid. (Reproduced with permission from Brondz I, Olsen I, and Greibroekk T (1983) Direct analysis of free fatty acids in bacteria by gas chromatography. *Journal of Chromatography* 274: 299–304.)

importance of derivatization is that it can improve the detection limit. It can also provide additional, or in some cases clearer, information under fragmentation in MS experiments.

## Derivatizations to Enhance Volatility, to Reduce Adhesion, and to Reduce Tailing

Esterification can increase the volatility of fatty acids, reduce dimerization in the vapor phase, and reduce adhesion. Esterification improves the peak configuration, the separation, and sample detectability.

The methyl, ethyl, propyl, *iso*-propyl, *n*-butyl, and *iso*-butyl esters of fatty acids are recommended.

Methylation of fatty acids is a well-known technique. A wide range of methylation and transesterification procedures are available in the literature. Defined IUPAC methods for the preparation and GC analysis of fatty acid methyl esters (FAMES) have been published. The methylation with diazomethane, as modified by Schlenk and Gelleman is in common use. Under the proper conditions, methylation with diazomethane gives good results.

It is known that methylation with diazomethane selectively methylates FFAs in the presence of FA esters. Caution is required, as diazomethane is a dangerous substance. Under inappropriate conditions, derivatization with diazomethane leads to the formation of artifacts. Functional groups such as phenol, enol, and carbonyl, or olefinic bonds can be affected. The less-dangerous methylation agents such as trimethylsilyldiazomethane (TMS-CHN<sub>2</sub>) and lithiated TMS-CHN<sub>2</sub> have been suggested to replace diazomethane. Benzyl esters can be produced by the use of phenyldiazomethane.

Anhydrous MeOH in the presence of an electrophilic catalyst, such as HCl, concentrated sulfuric acid, dichloroacetic acid, trifluoroacetic acid, benzene sulfonic acid, polyphosphoric acid, *p*-toluene sulfonic acid, thionyl chloride, and perchloric acid is often used for the methylation of fatty acids.

Adding liquid acetyl chloride to anhydrous MeOH to prepare anhydrous MeOH-HCl is a simple procedure that is recommended. Hydrochloric acid, H<sub>2</sub>SO<sub>4</sub>, and BF<sub>3</sub> in anhydrous MeOH can be used for acid-catalyzed transmethylation. In addition, transmethylation can be catalyzed by KOH, NaOCH<sub>3</sub>, and NaOH in MeOH. Cyclopropane fatty acids are sensitive to acid-catalyzed methanolysis, and artifacts can be produced. Alkaline esterification is preferred when the sample contains cyclopropane fatty acids. The preparation of ethyl, propyl, *iso*-propyl, *n*-butyl, *iso*-butyl-, and benzyl esters for GC

analysis of mixtures of low and high molecular weight fatty acids is preferred to methyl esters. It is possible to obtain satisfactory results using boron trifluoride ( $\text{BF}_3\text{-MeOH}$ ) and boron trichloride ( $\text{BCl}_3\text{-MeOH}$ ) for the methylation of cyclopropane fatty acids. Sometimes, it is necessary to perform the methylation of carboxylic acids in an aqueous phase, and then trimethyloxonium tetrafluoroborate (TMO) is the reagent of choice.

## Derivatization of Hydroxy Fatty Acids for GC Analysis

Alcoholic functional groups in HFAs can be derivatized using a trifluoroacetic acid anhydride (TFAA)–pyridine mixture after esterification of the carboxylic function. Methyl iodide ( $\text{CH}_3\text{I}$ ) can be used for preparation of the methyl esters of HFAs.

Trialkylsilylation of HFAs was popular; however, it has drawbacks. *N*-Trimethylsilylimidazole (TSIM) preferentially derivatizes hydroxyl groups. TSIM can be used for the trimethylsilylation of HFAs in the presence of water. *N*-Methyl-*N*-trimethylsilyl-heptafluorobutyramide (MSHFBA) does not produce any deposit in a flame ionization detector, and can be used for the silylation of acid and alcohol functions.

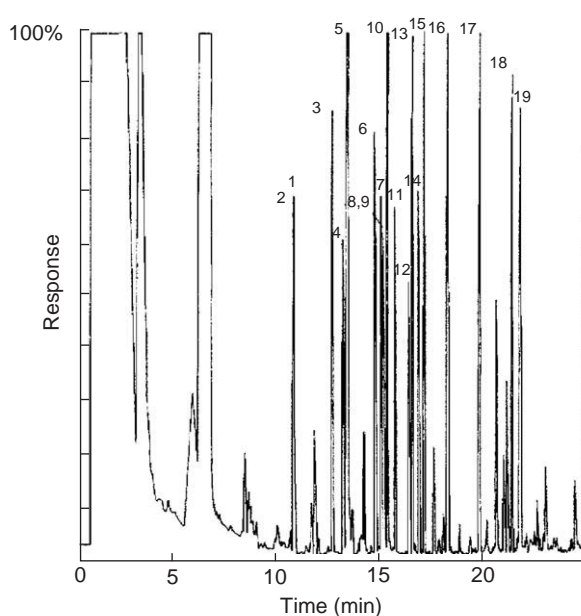
Hexamethyldisilazane (HMDS) is not very effective by itself, but in combination with TMCS, it is a fast and quantitative silylating reagent. *N*-(*tert*-butyldimethylsilyl)-*N*-methyltrifluoroacetamide (MTBSTFA) can react with the hydroxyl on HFAs, and with carboxyl to form relatively stable *tert*-butyldimethylsilyl derivatives which are useful in GC–MS analysis.

Tetramethylammonium hydroxide, tetramethylammonium acetate, tetramethylanilinium hydroxide, trimethylsulfonium hydroxide, and tetramethylguanidine have been used in intrainjector derivatization as shown in Figure 22. This technique has become popular for thermally assisted hydrolysis, and in the methylation of complex mixtures of lipids or for a complex matrix, such as bacterial cell walls. Tetramethylguanidine (TMG) also reacts with FFA to produce methyl esters in the hot GC injector.

*N,N*-Dimethylformamide dialkyl acetals in mixture with pyridine can react with fatty acids in a hot GC injector to methylate fatty acids. However, the hydroxyl groups of HFAs do not react.

## Derivatives to Enhance Detectability of Fatty Acids in GC Analysis

It is well known that formic acid has a low FID response. Derivatization to form the  $\alpha$ -hydroxyphosphonate ester permits the use of NPD for detection

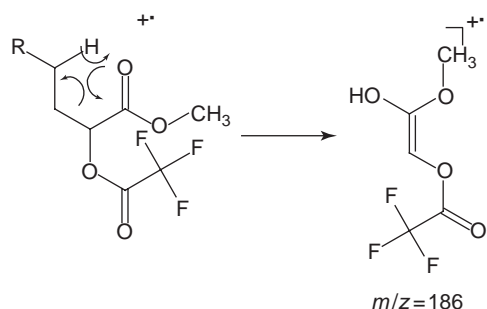


**Figure 22** GC chromatogram of TMAH intrainjector derivatized phenoxy acids. Peaks: 1 = 2-phenoxypropionic acid; 2 = *p*-fluorophenoxyacetic acid; 3 = 2-methylphenoxyacetic acid; 4 = 3-methylphenoxyacetic acid; 5 = 4-methylphenoxyacetic acid; 6 = 2-formylphenoxyacetic acid; 7 = 2,5-dimethylphenoxyacetic acid; 8 = 2-(4-chlorophenoxy)propionic acid; 9 = 2,4-dimethylphenoxyacetic acid; 10 = 2-methoxyphenoxyacetic acid; 11 = 4-phenoxybutyric acid; 12 = 3-methoxyphenoxyacetic acid; 13 = 4-methoxyphenoxyacetic acid; 14 = 2-(4-chloro-2-methylphenoxy)propionic acid; 15 = 4-chloro-2-methylphenoxyacetic acid; 16 = 2,4-dichlorophenoxypropionic acid; 17 = iodophenoxyacetic acid; 18 =  $\alpha$ -(2,4,5-trichlorophenoxy)propionic acid; 19 = 2,4,5-trichlorophenoxyacetic acid. (Reproduced with permission from Brondz I and Olsen I (1992) Intra-injector formation of methyl esters from phenoxy acid pesticides. *Journal of Chromatography* 598: 309–312.)

in the low picogram range. The 2,2,2-trichloroethyl esters of FAs can be detected using an ECD at high sensitivity in the nanogram range. The pentafluorobenzyl bromide (PFB-Br) derivatives of fatty acids have been detected using an ECD in the same sensitivity range. The use of iodomethyltetramethyldisiloxane esters allows the more sensitive detection using an ECD than do chlorine or bromine derivatives.

## Derivatives to Facilitate MS Fragmentation Patterns

Cyanomethyl esters are produced in the reaction between fatty acids and iodoacetoneitrile or bromoacetoneitrile. The cyanomethyl esters produce visible molecular ions  $\text{M}^+$ , and characteristic fragments, e.g.,  $[\text{M} - 15]^+$ ,  $[\text{M} - 31]^+$ ,  $[\text{M} - 43]^+$ , and  $[\text{M} - 99]^+$ .



**Figure 23** McLafferty rearrangement ion  $m/z$  186 of the trifluoroacetic acid anhydride derivatives of 2-hydroxy fatty acid methyl esters.

### Localization of the Hydroxyl in the Principal Chain

HPLC–MS or GC–MS are techniques that are normally used for the localization of hydroxyl groups in the principal chains of HFAs. A characteristic fragment ion at  $m/z = 186$  for trifluoroacetyl derivatives indicates the presence of an  $\alpha$ -hydroxyl group (Figure 23).

Derivatization with dimethyloxazoline (DMOX) assists in the correct localization of the hydroxyl group in the principal chain of the HFAs.

### Localization of Unsaturation or Epoxidation

The techniques used for the localization of unsaturation or rings in the principal chain of fatty acids are the von Rudloff oxidation and the oxidation reactions described by Ackman. The products of these reactions can be analyzed using GC or GC–MS. However, the low molecular weight fatty acids produced in the reactions can be lost. Preparation of pyrrolidids from unsaturated fatty acids avoids this problem. DMOX derivatives can also help in the localization of unsaturation, epoxidation, and hydroxyl groups.

### Localization of Branching

The localization of branching is achieved by careful comparison of the MS fragmentation of normal and branched fatty acids. The intensity of some fragments is a useful diagnostic tool. However, a reduction of esters to alcohols can improve the interpretation.

### Localization of the Keto Function in Fatty Acids

The localization of keto functions in KFA's can be achieved by conversion to methyl hexahydro-3-alkyl-6-thioxo-1,2,4,5-tetrazine-3-alkanoates.

*See also:* **Clinical Analysis:** Sample Handling. **Derivatization of Analytes.** **Extraction:** Solvent Extraction Principles; Supercritical Fluid Extraction; Solid-Phase Extraction. **Lipids:** Overview.

### Further Reading

- Barry CE III, Lee RE, Mdluli K, *et al.* (1998) Mycolic acids: structure, biosynthesis and physiological functions. *Progress in Lipid Research* 37(2/3): 143–179.
- Bremer J (2001) The biochemistry of hypo- and hyperlipidemic fatty acid derivatives: metabolism and metabolic effects. *Progress in Lipid Research* 44: 231–268.
- Brondz I (2002) Development of fatty acid analysis by high-performance liquid chromatography, gas chromatography, and related techniques. *Analytica Chimica Acta* 465: 1–37.
- Dembitsky VM (2002) Natural halogenated fatty acids: their analogues and derivatives. *Progress in Lipid Research* 41: 315–367.
- International Union of Pure and Applied Chemistry (IUPAC) (1992) *Standard Methods for the Analysis of Oils, Fats and Derivatives*, 7th edn., Suppl. 1, Oxford: Pergamon.
- Hiroshi M (2002) High-performance liquid chromatographic determination of free fatty acids and esterified fatty acids in biological materials as their 2-nitrophenylhydrazides. *Analytica Chimica Acta* 465(1/2): 237–255.
- Pinho O, Pérès C, and Ferreira IMPLVO (2003) Solid-phase microextraction of volatile compounds in ‘Terreiro’ ewe cheese. Comparison of different fibers. *Journal of Chromatography A* 1011: 1–9.
- Siddiqui RA, Jensi LJ, Harvey KA, *et al.* (2003) Cell-cycle arrest in Jukat leukaemic cells: a possible role for docosahexaenoic acid. *Biochemical Journal* 371: 621–629.
- Toshimasa T (1995) Use of derivatization to improve the chromatographic and detection selectivity of physiologically important carboxylic acids. *Journal of Chromatography B Biomedical Applications* 671: 91–112.
- Weiss J and Reinhard S (1995) Aliphatic carboxylic acids. A difficult chromatographic problem? *GIT Spezial Chromatographie* 15(2): 105–111.

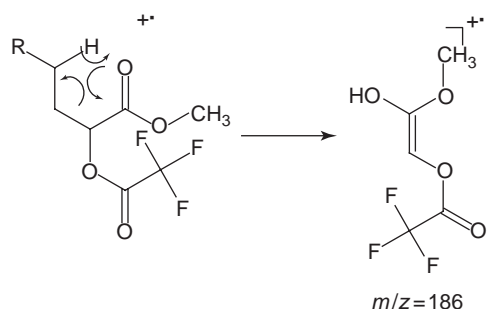
## Polar Lipids

**M H Gordon**, The University of Reading, Reading, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Polar lipids occur as complex mixtures in biological membranes. This category includes phospholipids,



**Figure 23** McLafferty rearrangement ion  $m/z$  186 of the trifluoroacetic acid anhydride derivatives of 2-hydroxy fatty acid methyl esters.

### Localization of the Hydroxyl in the Principal Chain

HPLC–MS or GC–MS are techniques that are normally used for the localization of hydroxyl groups in the principal chains of HFAs. A characteristic fragment ion at  $m/z = 186$  for trifluoroacetyl derivatives indicates the presence of an  $\alpha$ -hydroxyl group (Figure 23).

Derivatization with dimethyloxazoline (DMOX) assists in the correct localization of the hydroxyl group in the principal chain of the HFAs.

### Localization of Unsaturation or Epoxidation

The techniques used for the localization of unsaturation or rings in the principal chain of fatty acids are the von Rudloff oxidation and the oxidation reactions described by Ackman. The products of these reactions can be analyzed using GC or GC–MS. However, the low molecular weight fatty acids produced in the reactions can be lost. Preparation of pyrrolidids from unsaturated fatty acids avoids this problem. DMOX derivatives can also help in the localization of unsaturation, epoxidation, and hydroxyl groups.

### Localization of Branching

The localization of branching is achieved by careful comparison of the MS fragmentation of normal and branched fatty acids. The intensity of some fragments is a useful diagnostic tool. However, a reduction of esters to alcohols can improve the interpretation.

### Localization of the Keto Function in Fatty Acids

The localization of keto functions in KFA's can be achieved by conversion to methyl hexahydro-3-alkyl-6-thioxo-1,2,4,5-tetrazine-3-alkanoates.

**See also:** **Clinical Analysis:** Sample Handling. **Derivatization of Analytes.** **Extraction:** Solvent Extraction Principles; Supercritical Fluid Extraction; Solid-Phase Extraction. **Lipids:** Overview.

### Further Reading

- Barry CE III, Lee RE, Mdluli K, *et al.* (1998) Mycolic acids: structure, biosynthesis and physiological functions. *Progress in Lipid Research* 37(2/3): 143–179.
- Bremer J (2001) The biochemistry of hypo- and hyperlipidemic fatty acid derivatives: metabolism and metabolic effects. *Progress in Lipid Research* 44: 231–268.
- Brondz I (2002) Development of fatty acid analysis by high-performance liquid chromatography, gas chromatography, and related techniques. *Analytica Chimica Acta* 465: 1–37.
- Dembitsky VM (2002) Natural halogenated fatty acids: their analogues and derivatives. *Progress in Lipid Research* 41: 315–367.
- International Union of Pure and Applied Chemistry (IUPAC) (1992) *Standard Methods for the Analysis of Oils, Fats and Derivatives*, 7th edn., Suppl. 1, Oxford: Pergamon.
- Hiroshi M (2002) High-performance liquid chromatographic determination of free fatty acids and esterified fatty acids in biological materials as their 2-nitrophenylhydrazides. *Analytica Chimica Acta* 465(1/2): 237–255.
- Pinho O, Pérès C, and Ferreira IMPLVO (2003) Solid-phase microextraction of volatile compounds in 'Terreiro' ewe cheese. Comparison of different fibers. *Journal of Chromatography A* 1011: 1–9.
- Siddiqui RA, Jensi LJ, Harvey KA, *et al.* (2003) Cell-cycle arrest in Jukat leukaemic cells: a possible role for docosahexaenoic acid. *Biochemical Journal* 371: 621–629.
- Toshimasa T (1995) Use of derivatization to improve the chromatographic and detection selectivity of physiologically important carboxylic acids. *Journal of Chromatography B Biomedical Applications* 671: 91–112.
- Weiss J and Reinhard S (1995) Aliphatic carboxylic acids. A difficult chromatographic problem? *GIT Spezial Chromatographie* 15(2): 105–111.

## Polar Lipids

**M H Gordon**, The University of Reading, Reading, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Polar lipids occur as complex mixtures in biological membranes. This category includes phospholipids,



sphingolipids, glycolipids, and sulfolipids. Each of these classes comprises molecules that vary in the fatty acids and the polar groupings that are present. Thus, the phospholipids include phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine, phosphatidylinositol, phosphatidylglycerol, and phosphatidic acid as well as plasmalogens containing phosphate; sphingolipids include sphingosine, sphingomyelin, ceramides, gangliosides, cerebroside, and cerebroside sulfates; sulphoquinovosyldiacylglycerol and mono- and di-galactosyldiacylglycerols are also members of the polar lipids class (**Figure 1**).

The great complexity of natural lipid mixtures normally necessitates the use of some form of chromatography for their separation. Since the nonpolar lipid molecules are often better resolved and recovered from chromatographic systems, the polar head group may be removed or masked with nonpolar groups prior to the chromatographic separation.

## Solubility and Extraction of Polar Lipids

Polar lipids are only sparingly soluble in hydrocarbon solvents but most polar lipids dissolve in more polar solvents such as methanol, ethanol, or chloroform. Chloroform–methanol mixtures dissolve most polar lipids but polyphosphoinositides or lysophospholipids are poorly soluble in this solvent mixture. Water–methanol mixtures may dissolve significant amounts of the most polar lipids such as gangliosides. Acetone is a poor solvent for phospholipids but glycolipids are more soluble in this solvent.

Extraction of lipids with chloroform–methanol (2: 1) as in the Folch or Bligh and Dyer procedures is commonly sufficient to extract most lipids from biological tissues. The Folch procedure removes contaminating compounds by shaking the solvent mixture with 25% of their volume of potassium chloride solution. The lower chloroform-rich phase contains the purified lipids whilst the upper phase contains nonlipid contaminants together with gangliosides and some glycolipids. Gangliosides can be removed from the upper phase by dialysis and lyophilization.

## Detection and Quantification of Polar Lipids

Lipids are commonly characterized by their solubility properties, i.e., they can be extracted by the Folch or Bligh and Dyer procedures. The next stage in the analytical procedure for mixtures involves

chromatography, commonly thin-layer chromatography (TLC), and a range of selective spray reagents are useful for detecting polar lipids (see **Table 1**).

The concentration of phospholipids in a lipid mixture can be estimated as phosphate after ashing the sample. The ash can be treated with nitric acid, dissolved in hydrochloric acid, and the phosphorus determined by inductively coupled plasma-atomic emission spectrometry at 214.9 nm. Alternatively, the phosphorus can be determined by atomic absorption spectrometry at 213.547 nm with a graphite furnace. Traditional methods for phosphorus determination include the precipitation of phosphate as quinolinium molybdophosphate and gravimetric determination, or colorimetric methods such as the reaction with ammonium molybdate and acidic ammonium vanadate solution followed by determination of the absorbance at 460 nm.

The glycosidic bond of glycolipids is readily hydrolyzed by acid, and the carbohydrates liberated can be analyzed by instrumental methods, e.g., gas chromatography (GC), liquid chromatography (LC), or colorimetric procedures can be used. In one colorimetric procedure, the total amount of hexose units in a glycolipid can be quantified by reaction with anthrone in concentrated sulfuric acid and determination by the absorbance at 635 nm.

## Spectroscopic Methods

Spectroscopic methods are mainly useful for identifying pure lipids. Mass spectrometry (MS), nuclear magnetic resonance (NMR) spectroscopy, and infrared (IR) spectroscopy provide useful information but the ultraviolet (UV) region is mainly useful for LC detection of unsaturated lipids by the weak absorbance of *cis*-unsaturated bonds at 205 nm. Although MS may give a molecular ion and fragments that allow identification of the fatty acids in glycerolipids, identification of the positional distribution of the fatty acids may require enzymic hydrolysis with phospholipase A<sub>2</sub>, which hydrolyzes fatty acids at position 2 of glycerophospholipids.

### Nuclear Magnetic Resonance Spectroscopy

The <sup>1</sup>H, <sup>13</sup>C, and <sup>31</sup>P NMR spectra of complex lipids have been studied. <sup>1</sup>H and <sup>13</sup>C provide useful information for single phospholipids but due to the complexity of the spectra, it is not possible to use these techniques to analyze mixtures.

The amide proton of sphingosine and its derivatives such as cerebroside shows a doublet at low field (~7.6 ppm, *J* = 9 Hz), which is exchangeable



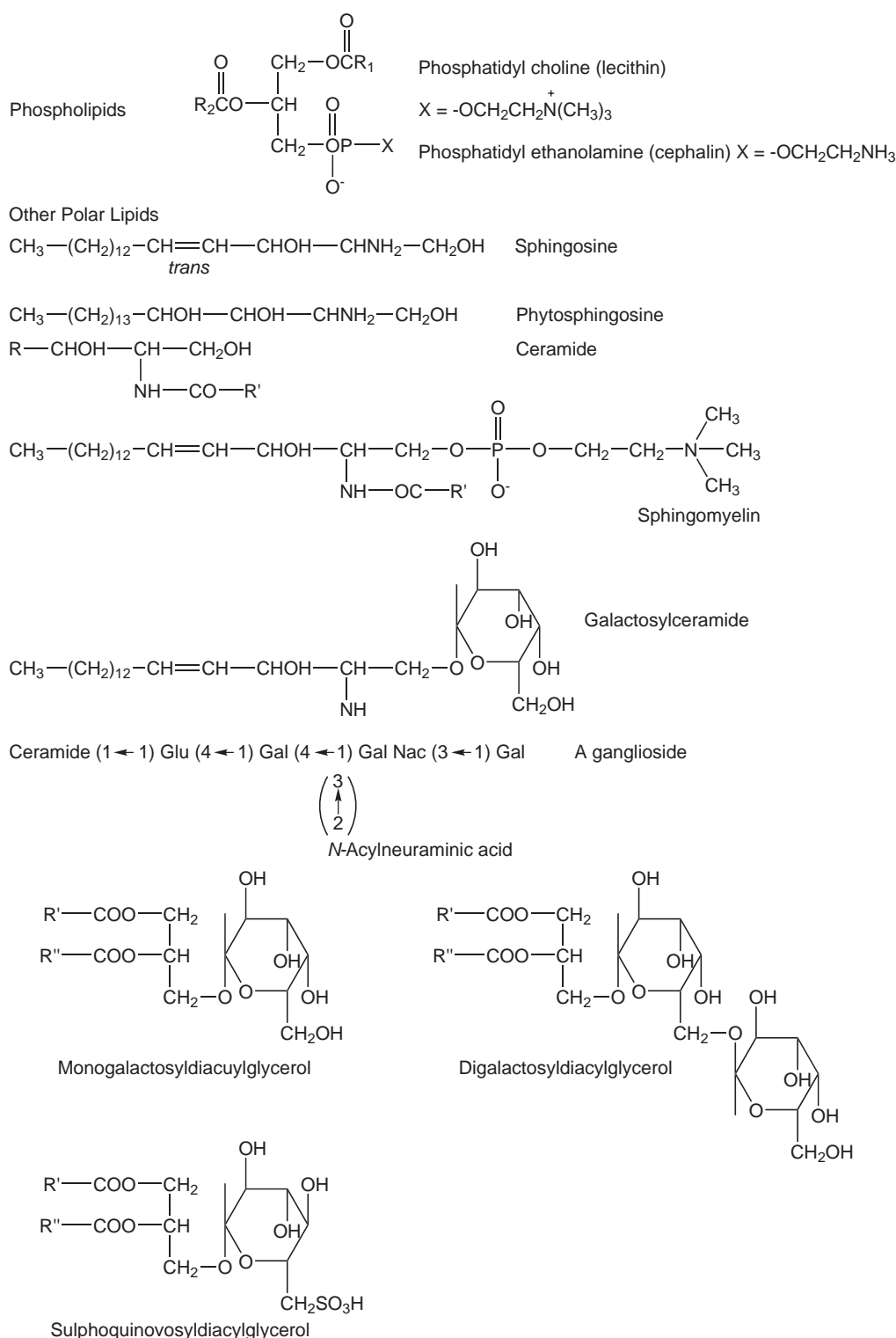


Figure 1 Polar lipids.

with  $\text{D}_2\text{O}$ . The fatty acid substituent shows a long methylene chain at 1.3 ppm, with the terminal methyl group at 0.85 ppm. Acetylation of glycolipids allows the number of sugar residues to be determined

from the ratio of the integral of the acetyl signal at  $\delta$  1.97–2.5 ppm and other characteristic signals, e.g., the amide proton. High-resolution  $^1\text{H}$  NMR of glycosphingolipids using spin decoupling difference

**Table 1** Reagents for detection of lipids in TLC

Reagent	Color	Specificity
50% H <sub>2</sub> SO <sub>4</sub> in methanol Heat at 110°C	Brown-black	Nonspecific
0.25% Potassium dichromate in 15% H <sub>2</sub> SO <sub>4</sub> Heat at 125°C/30 min	Brown-black	Nonspecific
Iodine vapor	Brown	Nonspecific
0.01% 2',7'- Dichlorofluorescein in ethanol/UV irradiation	Pale	Nonspecific
Molybdic anhydride, H <sub>2</sub> SO <sub>4</sub> molybdenum	Blue	Phospholipids
Ammonium pentachlorooxomolybdate/ H <sub>2</sub> SO <sub>4</sub>	Blue	Phospholipids
Ninhydrin/ <i>n</i> -butanol/acetic acid. Heat at 110°C	Red	Amino group
Basic bismuth nitrate/acetic acid/potassium iodide	Orange	Choline
Anthrone/thiourea/H <sub>2</sub> SO <sub>4</sub> Heat at 110°C, 10 min	Red-purple	Glycolipid
Diphenylamine/HCl/acetic acid Heat at 105°C, 5 min	Blue-gray	Glycolipid
Resorcinol/copper sulfate/HCl Heat at 95°C, 30 min	Blue-purple	Ganglioside

spectroscopy or the nuclear Overhauser effect allows identification of most of the sugar ring protons for glycosphingolipids.

<sup>13</sup>C NMR with broadband <sup>1</sup>H decoupling is useful in identifying the sugar residue in glycolipids. <sup>13</sup>C NMR spectroscopy of phospholipids in solution has also been reported. Changes in the polar head group do not affect the chemical shifts of carbon atoms in the rest of the molecule. The carbon  $\alpha$  to the phosphate group in the polar head group is well separated from the  $\beta$  carbon atom. Chemical shifts of C $\alpha$  and C $\beta$  are 59.58 and 66.80 ppm for phosphatidylcholine and 62.16 and 40.80 ppm for phosphatidylethanolamine. The methyl groups in the polar head group of phosphatidylcholine have a chemical shift of 54.47 ppm. Coupling between <sup>13</sup>C and <sup>31</sup>P gives information about the conformation of phospholipids in the region of polar head groups.

High-resolution <sup>31</sup>P NMR spectra of phospholipids are typically determined with spectrometers operating at 202.4 MHz and a magnetic field of 11.75 T. The use of detergents has been shown to narrow the <sup>31</sup>P linewidth considerably. Proton broadband decoupling removes splitting due to <sup>31</sup>P-<sup>1</sup>H coupling, and signals of <3.5 Hz width at half-height can be obtained. Phosphodiester resonance signals normally occur in the chemical shift range of +1.5 to -1.5 ppm relative to 85% inorganic orthophosphoric acid. Lyso derivatives

show an increase in  $\delta$  of ~0.5 ppm. The signals due to each family of phospholipids are well separated and integration can be used to quantify each family of phospholipids. The chemical shifts of the base and conjugate acid forms of phospholipids in aqueous media differ by ~0.5 ppm.

Also, <sup>31</sup>P NMR is useful for studying the conformation and dynamics of phospholipid head groups as well as lipid polymorphism.

### Mass Spectrometry

Degradation and derivatization of complex lipids commonly interferes with the complete analysis of intact molecules when electron impact is used for ionization due to the low volatility and thermal lability of these molecules.

Various classes of phospholipids have been examined by field-desorption MS. Most phospholipids except phosphatidylserine provide intense MH<sup>+</sup> ions under these conditions but many ions of higher mass may also occur due to association of the polar groups.

Desorption chemical ionization with ammonia provides useful information about fatty acyl groups, while fast atom bombardment (FAB) gives diagnostic peaks about the various head groups. Ammonia is better than methane as a reagent gas for chemical ionization, since it provides more useful fragmentations. The desorption chemical ionization spectra of phospholipids other than phosphatidylserines show protonated molecular ions. Phosphatidylcholines show an MH<sup>+</sup>-42 peak arising from MH<sup>+</sup> - N-(CH<sub>3</sub>)<sub>3</sub> + NH<sub>3</sub>. Phospholipids cleave at the phosphorus-glycerol oxygen linkage and ions are evident from the 1,2-diacylglycerol ammonia adduct, which occurs together with protonated ions and ions formed by the elimination of water from the diacylglycerol. The latter species also suffers additional loss of a fatty acyl chain.

In addition, FAB-MS gives protonated ions for phospholipids other than phosphatidylserine. Dimeric ions plus association ions are formed. Common fragments include [H<sub>3</sub>PO<sub>4</sub> + H]<sup>+</sup>, [M - H<sub>2</sub>PO<sub>4</sub>]<sup>+</sup>, [M + H - fatty acid]<sup>+</sup>, [M + H - fatty acyl chain]<sup>+</sup>. Phosphatidylcholines give the choline phosphate ion of mass 184 as the base peak. A base peak of 142 is evident in the mass spectra of phosphatidylethanolamines.

It is clear that the mass spectral behavior of phospholipids using desorption chemical ionization and FAB techniques provides valuable information about their molecular structures.

These techniques are also useful in determining the structure of glycolipids where loss of a sugar residue is a characteristic ion.

## Infrared Spectra

Most complex lipids are acylglycerol derivatives and therefore their IR spectra show bands characteristic of esters and hydrocarbon chains (Table 2).

The phosphate group of phospholipids shows a strong  $\text{P}=\text{O}$  stretching band at  $1350\text{--}1250\text{ cm}^{-1}$ , which is sometimes a doublet and may be shifted to lower frequencies with an increase in intensity. Vibration of the  $\text{P}\text{--}\text{OH}$  group gives rise to absorption at  $2700\text{--}2560\text{ cm}^{-1}$ . Differences occur in the  $\text{P}=\text{O}$  stretching frequency between phosphatidylethanolamine ( $1227\text{ cm}^{-1}$ ) and phosphatidylserine ( $1220\text{--}1180\text{ cm}^{-1}$ ).

The *trans* double bond of sphingosine derivatives gives an absorbance band at  $970\text{ cm}^{-1}$ , although another band occurs at this frequency in some sphingosine derivatives such as sphingomyelin and this interferes with the use of this band for quantitative determination of the *trans* unsaturated content.

The IR spectra of complex lipids are not sensitive to variations in acyl chain length, unsaturation, or the presence of  $\alpha,\beta$ -unsaturated ether links.

However, the nitrogenous bases of phospholipids can be identified by their IR absorption spectra in the region of  $900\text{--}1100\text{ cm}^{-1}$ . Free amine groups as present in phosphatidylserine or phosphatidylethanolamine show a single sharp absorption band at  $1075\text{ cm}^{-1}$ . Quaternary amines such as phosphatidylcholine or sphingomyelin have a doublet at  $1087$  and  $1053\text{ cm}^{-1}$  and a sharp band at  $961\text{ cm}^{-1}$ . In addition, the range  $1430\text{--}1820\text{ cm}^{-1}$  shows characteristic features.

The spectrum of a solid phospholipid may show fine structure at low temperatures, which disappears on heating. For example, the  $\text{CH}_2$  rocking band at  $720\text{ cm}^{-1}$  of 2,3-dipalmitoyl-DL-1-phosphatidylethanolamine is split into a doublet at low temperatures but this band changes into a singlet at room temperature. This is similar to the behavior observed with soaps and is due to the flexing and twisting of the hydrocarbon group on heating. At some temperature below the melting point of the lipid, the fine structure disappears completely due to movement of the acyl chain.

**Table 2** Assignment of acylglycerol bands in infrared spectra

Vibration	Wavenumber range ( $\text{cm}^{-1}$ )
C—H stretch ( $\text{CH}_2$ and $\text{CH}_3$ )	3030–2967
C=O stretch	1751–1720
C—H bend ( $\text{CH}_2$ and $\text{CH}_3$ )	1464–1453
C—H bend (symmetrical deformation of $\text{CH}_3$ )	1383–1361
C—H in plane wagging or rocking of $\text{CH}_2$	1261–1250
C—O stretch	1179–1166
$\text{CH}_2$ rocking	730–717

The infrared spectra of solid phospholipids vary due to the occurrence of polymorphism. The ester carbonyl frequency and bands between  $670$  and  $1050\text{ cm}^{-1}$  may shift due to the variations in the crystal structure arising from different polymorphs.

## Chromatographic Separation of Intact Lipids

Removal of residual water and other nonlipid material may be achieved with solid-phase extraction cartridges. The major lipid classes namely neutral lipids, glycolipids, and phospholipids are commonly separated using column chromatography, TLC, or LC as a preliminary separation procedure. Column chromatography with silicic acid has been widely used to separate phospholipids, cerebroside, and sulfatides. The degree of hydration of the silicic acid has an important influence on the quality of the separation. Sufficient water is required to avoid tailing of the lipid bands, but quantities of water above  $\sim 5\%$  lead to poorer retention and separation. Detection of lipids from the chromatographic column may be gravimetric, or it may involve the removal of aliquots from each fraction for testing with acidified potassium dichromate, which suffers a reduction in absorbance at  $350\text{ nm}$  in the presence of lipids. Diethylaminoethylcellulose chromatography is a useful alternative to adsorption chromatography for the separation of polar lipids. This form of chromatography depends partly on ion exchange and partly on adsorption.

## Thin-Layer Chromatography

TLC is probably the most common method for the separation of polar lipids. One-dimensional TLC is commonly used for rapid group separation or for small-scale preparative purposes, whilst two-dimensional TLC gives improved resolution. Silica gel G plates developed in chloroform–methanol–water (25:10:1 by volume) give useful separations of phospholipids. Solvent mixtures containing acetic acid or ammonia, e.g., chloroform–methanol–28% aqueous ammonia (65:35:5) or chloroform–acetone–methanol–acetic acid–water (10:4:2:2:1) are among the wide range of solvent mixtures that have been used for TLC separations of polar lipids.

For the separation of glycolipids, either chloroform–methanol–water or *n*-propanol–water are commonly used. The mobilities of acidic glycolipids such as gangliosides and the width of their bands are affected by the presence of salts or ammonia, whereas neutral glycolipids are well separated both in the presence and absence of these additives. The

improved resolution of high-performance TLC is valuable in the separation of complex polar lipid mixtures.

Lipids separated on TLC plates are usually detected with chromogenic or fluorescent reagents. Spraying with acidified potassium dichromate followed by heating at 125°C, which yields brown–black spots, or treating with iodine vapor are nonspecific methods of detecting lipids and other organic compounds. Spraying with a fluorescent reagent such as 2',7'-dichlorofluorescein solution followed by inspection under a UV lamp is another nonspecific method. Specific reagents for the detection of phospholipids, amino group-containing lipids, choline-containing lipids, glycolipids, and gangliosides are shown in Table 1.

Polar lipids separated by TLC are commonly quantified by densitometry or the lipids may be scraped off and quantified by colorimetry or gas chromatography after derivatization. The Iatroscan TLC-flame ionization detector system is well suited to the quantification of polar lipids. In this procedure, TLC is performed on quartz rods coated with silica. The rods are then passed through a flame ionization detector to quantify each lipid band in turn.

### Liquid Chromatography

Normal-phase LC analysis is an effective method of separating phospholipid classes. Columns containing silica (3–10 µm diameter) as the stationary phase are used in combination with a mobile phase of chloroform–methanol–water; acetonitrile–methanol–water; or hexane–isopropanol–water. UV absorbance at 205 nm due to the carbon–carbon double bonds of

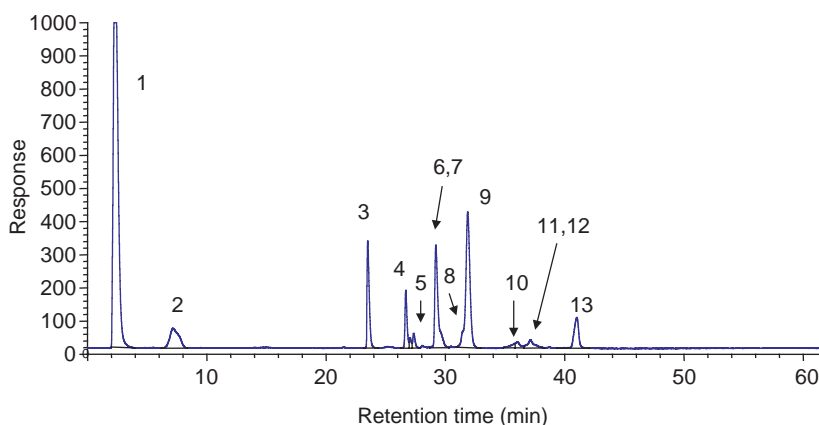
the fatty acids in the phospholipid can be used as the basis of detection for the latter two solvent mixtures. A light scattering detector or a flame ionization detector is required if chloroform is present. After separation of phospholipids into classes by normal-phase chromatography, reversed-phase LC with a C18 column can be used to separate the individual molecular species of a particular class. Effective separation of several classes of polar lipids can be achieved by HPLC (Figure 2).

Phospholipids can be separated from other polar lipids by normal-phase LC. Thus, on isocratic elution with acetonitrile–methanol–sulfuric acid (135:5:0.2), phosphatidylinositol elutes first followed by monogalactosyldiacylglycerol, *N*-acylphosphatidylethanolamine, *N*-acyl-lysophosphatidylethanolamine, phosphatidylethanolamine, and digalactosyldiacylglycerols.

There are very few reports of the application of supercritical fluid chromatography to the analysis of polar lipids in the literature.

### Electrophoresis

The importance of phospholipids in biological membranes and in cell signaling has provided a strong impetus for the development of methods capable of analyzing the complex mixtures that are present in biological systems. Electrophoresis using methods based on micellar electrokinetic capillary chromatography (MEKC) has strong potential for the separation of complex lipids. Baseline resolution of mixtures of phosphatidylethanolamine, phosphatidylserine, and lysophosphatidylethanolamine was achieved in 7 min using MEKC following labeling with a fluorogenic dye.



**Figure 2** HPLC chromatogram of chloroform-extracted wheat flour lipid analyzed on a silica column (5 µm; 100 mm × 4 mm) with gradient elution with a mixture of (A) hexane–tetrahydrofuran 99:1; (B) isopropanol, and (C) water. 1 = triacylglycerols; 2 = free fatty acids; 3 = diacylglycerols; 4 = monoacylglycerols; 5 = acylated steryl glycosides; 6 = monogalactosyldiacylglycerols; 7 = monogalactosylmonoacylglycerols; 8 = digalactosylmonoacylglycerols; 9 = digalactosyldiacylglycerols; 10 = unidentified; 11 = phosphatidylglycerol; 12 = phosphatidylethanolamine; 13 = phosphatidylcholine.

## Enzymatic Hydrolysis of Phospholipids

Glycerophospholipids may be hydrolyzed to phosphatidic acid with phospholipase D, or to diacylglycerols with phospholipase C. The latter hydrolysis allows the analysis of phospholipids by TLC, GC, or LC, whereas the hydrolysis with phospholipase D produces molecules that are best analyzed by LC.

If silver nitrate is incorporated into the silica gel used to coat a TLC plate, good separations of diacylglycerols according to degree of unsaturation can be achieved on elution with a mobile phase such as chloroform–methanol (99:1).

Silver ion chromatography has been extended to LC and is useful for separating polar lipids differing in unsaturation of the fatty acids. Thus, ceramides containing saturated fatty acids have been separated from ceramides containing unsaturated fatty acids on a column based on silver ions bound to an ion-exchange support.

## Stability of Polar Lipids

Polar lipids commonly include polyunsaturated fatty acids and consequently are sensitive to oxidation. Addition of an antioxidant such as tertiary butylhydroquinone is recommended during the isolation of polar lipid mixtures. If the polar lipids are to be extracted from biological tissues, extraction should preferably be performed as soon as possible after removal from the living organism in order to avoid enzyme-catalyzed changes in the lipids. If this is not possible, the samples should be stored in sealed glass containers under nitrogen at  $-20^{\circ}\text{C}$ . It may also help to inactivate lipases by plunging the sample into boiling water or boiling with acetic acid or isopropanol. Careful handling of the polar lipids is required at all stages during the isolation and analysis in order to avoid oxidation. The analysis of polar lipids is a very challenging exercise due to the complexity of natural lipid mixtures, but recent developments in analytical techniques, especially the widespread use of LC and the development of  $^{31}\text{P}$  NMR spectroscopy, have had a big impact on this area of analysis.

*See also:* **Extraction:** Solvent Extraction Principles. **Infrared Spectroscopy:** Overview. **Ion Exchange:** Overview; Principles. **Liquid Chromatography:** Overview; Column Technology; Normal Phase; Reversed Phase. **Nuclear Magnetic Resonance Spectroscopy-Applicable Elements:** Phosphorus-31. **Thin-Layer Chromatography:** Overview.

## Further Reading

- Ayanoglu E, Wegman A, Pilet O, Hass JR, and Djerassi C (1984) Mass spectrometry of phospholipids. *Journal of the American Chemical Society* 106: 5246–5251.
- Christie WW (1992–1997) *Advances in Lipid Methodology*, vols. 1–4. Bridgewater: Oily Press.
- Christie WW (2003) *Lipid Analysis: Isolation, Separation, Identification, and Structural Analysis of Lipids*, 3rd edn. Bridgewater: Oily Press.
- Cremesti AE and Fischl AS (2000) Current methods for the identification and quantitation of ceramides: An overview. *Lipids* 35: 937–945.
- Gelpi E (1995) Biomedical and biochemical applications of liquid-chromatography mass spectrometry. *Journal of Chromatography A* 703: 59–80.
- Geyer H and Geyer R (1998) Strategies for glycoconjugate analysis. *Acta Anatomica* 161: 18–35.
- Gunstone FD, Harwood JL, and Padley FB (1994) *The Lipid Handbook*, 2nd edn. London: Chapman & Hall.
- Hammond EW (1993) *Chromatography for the Analysis of Lipids*. Boca Raton: CRC Press.
- Harvey DJ (1999) Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrometry Reviews* 18: 349–450.
- Meneses P and Glonek T (1988) High resolution P-31 NMR of extracted phospholipids. *Journal of Lipid Research* 29: 679–689.
- Murari R, Abd Ed-Rahman MMA, Wedmid Y, Parthasany S, and Baumann WJ (1982) Carbon-13 NMR of phospholipids in solution. *Journal of Organic Chemistry* 47: 2158–2163.
- Ruiz-Gutierrez V and Perez-Camino MC (2000) Update on solid-phase extraction for the analysis of lipid classes and related compounds. *Journal of Chromatography A* 885: 321–341.
- Zhang L, Hu S, Cook L, and Dovichi NJ (2002) Analysis of aminophospholipid molecular species by methyl-beta-cyclodextrin modified micellar electrokinetic capillary chromatography with laser-induced fluorescence detection. *Electrophoresis* 23: 3071–3077.

## Determination in Biological Fluids

**A Kuksis**, University of Toronto, Toronto, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Lipid measurements in body fluids are of great interest because they can indicate abnormalities in



## Enzymatic Hydrolysis of Phospholipids

Glycerophospholipids may be hydrolyzed to phosphatidic acid with phospholipase D, or to diacylglycerols with phospholipase C. The latter hydrolysis allows the analysis of phospholipids by TLC, GC, or LC, whereas the hydrolysis with phospholipase D produces molecules that are best analyzed by LC.

If silver nitrate is incorporated into the silica gel used to coat a TLC plate, good separations of diacylglycerols according to degree of unsaturation can be achieved on elution with a mobile phase such as chloroform–methanol (99:1).

Silver ion chromatography has been extended to LC and is useful for separating polar lipids differing in unsaturation of the fatty acids. Thus, ceramides containing saturated fatty acids have been separated from ceramides containing unsaturated fatty acids on a column based on silver ions bound to an ion-exchange support.

## Stability of Polar Lipids

Polar lipids commonly include polyunsaturated fatty acids and consequently are sensitive to oxidation. Addition of an antioxidant such as tertiary butylhydroquinone is recommended during the isolation of polar lipid mixtures. If the polar lipids are to be extracted from biological tissues, extraction should preferably be performed as soon as possible after removal from the living organism in order to avoid enzyme-catalyzed changes in the lipids. If this is not possible, the samples should be stored in sealed glass containers under nitrogen at  $-20^{\circ}\text{C}$ . It may also help to inactivate lipases by plunging the sample into boiling water or boiling with acetic acid or isopropanol. Careful handling of the polar lipids is required at all stages during the isolation and analysis in order to avoid oxidation. The analysis of polar lipids is a very challenging exercise due to the complexity of natural lipid mixtures, but recent developments in analytical techniques, especially the widespread use of LC and the development of  $^{31}\text{P}$  NMR spectroscopy, have had a big impact on this area of analysis.

*See also:* **Extraction:** Solvent Extraction Principles. **Infrared Spectroscopy:** Overview. **Ion Exchange:** Overview; Principles. **Liquid Chromatography:** Overview; Column Technology; Normal Phase; Reversed Phase. **Nuclear Magnetic Resonance Spectroscopy-Applicable Elements:** Phosphorus-31. **Thin-Layer Chromatography:** Overview.

## Further Reading

- Ayanoglu E, Wegman A, Pilet O, Hass JR, and Djerassi C (1984) Mass spectrometry of phospholipids. *Journal of the American Chemical Society* 106: 5246–5251.
- Christie WW (1992–1997) *Advances in Lipid Methodology*, vols. 1–4. Bridgewater: Oily Press.
- Christie WW (2003) *Lipid Analysis: Isolation, Separation, Identification, and Structural Analysis of Lipids*, 3rd edn. Bridgewater: Oily Press.
- Cremesti AE and Fischl AS (2000) Current methods for the identification and quantitation of ceramides: An overview. *Lipids* 35: 937–945.
- Gelpi E (1995) Biomedical and biochemical applications of liquid-chromatography mass spectrometry. *Journal of Chromatography A* 703: 59–80.
- Geyer H and Geyer R (1998) Strategies for glycoconjugate analysis. *Acta Anatomica* 161: 18–35.
- Gunstone FD, Harwood JL, and Padley FB (1994) *The Lipid Handbook*, 2nd edn. London: Chapman & Hall.
- Hammond EW (1993) *Chromatography for the Analysis of Lipids*. Boca Raton: CRC Press.
- Harvey DJ (1999) Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrometry Reviews* 18: 349–450.
- Meneses P and Glonek T (1988) High resolution P-31 NMR of extracted phospholipids. *Journal of Lipid Research* 29: 679–689.
- Murari R, Abd Ed-Rahman MMA, Wedmid Y, Parthasanth S, and Baumann WJ (1982) Carbon-13 NMR of phospholipids in solution. *Journal of Organic Chemistry* 47: 2158–2163.
- Ruiz-Gutierrez V and Perez-Camino MC (2000) Update on solid-phase extraction for the analysis of lipid classes and related compounds. *Journal of Chromatography A* 885: 321–341.
- Zhang L, Hu S, Cook L, and Dovichi NJ (2002) Analysis of aminophospholipid molecular species by methyl-beta-cyclodextrin modified micellar electrokinetic capillary chromatography with laser-induced fluorescence detection. *Electrophoresis* 23: 3071–3077.

## Determination in Biological Fluids

**A Kuksis**, University of Toronto, Toronto, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Lipid measurements in body fluids are of great interest because they can indicate abnormalities in

metabolite levels and transport, and provide valuable clues for clinical diagnosis. A detailed determination of the lipid composition of biological fluids differs little from tissue lipid analysis, which requires combinations of two or more analytical systems and is laborious and time consuming. However, much information about the lipids in biological fluids can be frequently obtained by rapid total lipid profiling that is possible following a limited preliminary work-up. Thus, thin-layer chromatography (TLC) and gas-liquid chromatography (GC) continue to be extensively utilized for this purpose, while more recently, high-performance liquid chromatography (HPLC) and tandem mass spectrometry (MS/MS) have also been employed. Both GC and HPLC are readily combined with MS as online GC-MS and LC-MS for multi-component analyses of body fluids.

Most determinations of lipids in biological fluids are presently being performed using state-of-the-art methodology. The analytical routines have usually been adopted from tissue analyses, but in a few instances specific methods have been developed for biofluid analyses (e.g., plasma total lipid profiling). While the lipid analyses of a few biofluids have been reviewed in depth, in most instances background reviews are available only in leading applied publications. As a result, the present encyclopedic account is based largely on reports of the more successful analyses of lipids in the major biofluids. Combination of HPLC with electrospray MS has proved to be the most generally applicable methodology, but other chromatographic and mass spectrometric techniques have also served well for specific applications. Recently, nuclear magnetic resonance (NMR) and LC-NMR have been adopted for this purpose and offer special advantages for phospholipid class analyses. Unfortunately, the modern methods of lipid analyses have not provided the most reliable quantitative results, occasionally contradicting each other. In such instances, resolution of the problem of quantification has been sought in conventional chemical analyses and radio- or stable isotope dilution assays.

## Improvements in Strategy

Improvements in the strategy of analyses have come from the realization that body fluids have structure and are made up of different physicochemical phases, which can be separated into subfractions for a more meaningful assay than that provided by a total fluid analysis. Thus, blood can be subdivided by centrifugation into cellular (erythrocytes, lymphocytes, and platelets) and noncellular (plasma or serum) fractions, while plasma or serum can be subdivided

into various lipoprotein subfractions (high-density lipoprotein (HDL), low-density lipoprotein (LDL), very low density lipoprotein (VLDL), etc.) by ultracentrifugation. In other instances, the protein molecules of a biofluid may be resolved by two-dimensional gel chromatography and the lipid ligands associated with it determined. Furthermore, the total lipid extracts themselves can be resolved into polar and nonpolar fractions by silica gel cartridge separation prior to a quantitative lipid profiling. Finally, depending on the analytical needs, the lipid analysis of a biofluid can be performed on a highly concentrated sample, which would favor the detection of minor components, or on specific components (cholesterol, unesterified fatty acids) of the total lipid extract or the original biofluid (differential extraction and derivatization). A comparable prefractionation can be performed on milk, lymph, bile, and other body fluids and cell cultures. In addition, major developments in instrumental analyses now permit precise analyses on progressively smaller biofluid samples, which have permitted a revision and simplification of the original strategies for lipid analyses in biofluids. As a result, the isolation of specific protein components of a biofluid has been occasionally combined with a mass spectrometric identification of its lipid ligand. In other instances, the biofluid may be analyzed following incubation with endogenous or exogenously added enzymes.

Of equal or greater importance has been the replacement of the laborious and time-consuming multidimensional-chromatographic systems by flow injection and matrix-assisted laser-desorption ionization mass spectrometry (MALDI-MS) and MS/MS. These developments have been rivaled by the recent explosion in the application of NMR methodology for identification and quantification of lipid classes in body fluids. With these advances in the analytical technology has come the possibility of expansion of the lipid analyses to include many more samples than previously practical or possible. Large-scale study of lipids and surveying of the entire lipidome of an organism is now possible. Quantification of the change of several metabolite concentrations relative to the concentration change of one selected metabolite can reveal the site of action, in the metabolic network, of a 'silent gene'. Because of the easy access and a near universal contact with other tissues, blood plasma has long been known as a source of information about the metabolic state of the body. It is therefore no wonder that recent efforts of surveying the lipid metabolome have become centered on blood plasma. In a more limited but more specific way, the lipid metabolites present in other body fluids may also become of interest.

## Selected Applications

In the following, reference is being made to modern analyses of both major and minor lipids in selected biofluids, which may serve as examples for analyses of other internal and external secretions of animal bodies. Since many of the body fluids are exudates of plasma, the improvements in plasma lipid analyses are discussed in the greatest detail.

### Blood Cells, Plasma, and Plasma Lipoproteins

Detailed analysis of phospholipids, as well as neutral lipids and nonesterified fatty acids, has become progressively more important as these molecules have come to represent intercellular messenger molecules that are involved in a number of disease states. The analysis of lipid patterns of blood and lipoproteins may therefore constitute a useful diagnostic tool. As noted in the previous edition, suitable methods for phospholipid analysis are often time consuming and tedious, since most of them include several separation and derivatization steps.

**Phospholipids** The laborious combinations of thin-layer or liquid chromatography (LC) with gas chromatography (GC) for the analysis of phospholipids in biological fluids have been replaced by combinations of HPLC with MS (LC–MS). In many instances, the convenience of MS/MS mode of mass spectrometric identification of the erythrocyte and platelet phospholipids has been preferred to the LC–MS combination. A collision-induced dissociation of specific parent ions allows the deconvolution of the composition of the common mixtures of molecular species. For more complex phospholipid mixtures, a combined LC–MS approach is recommended.

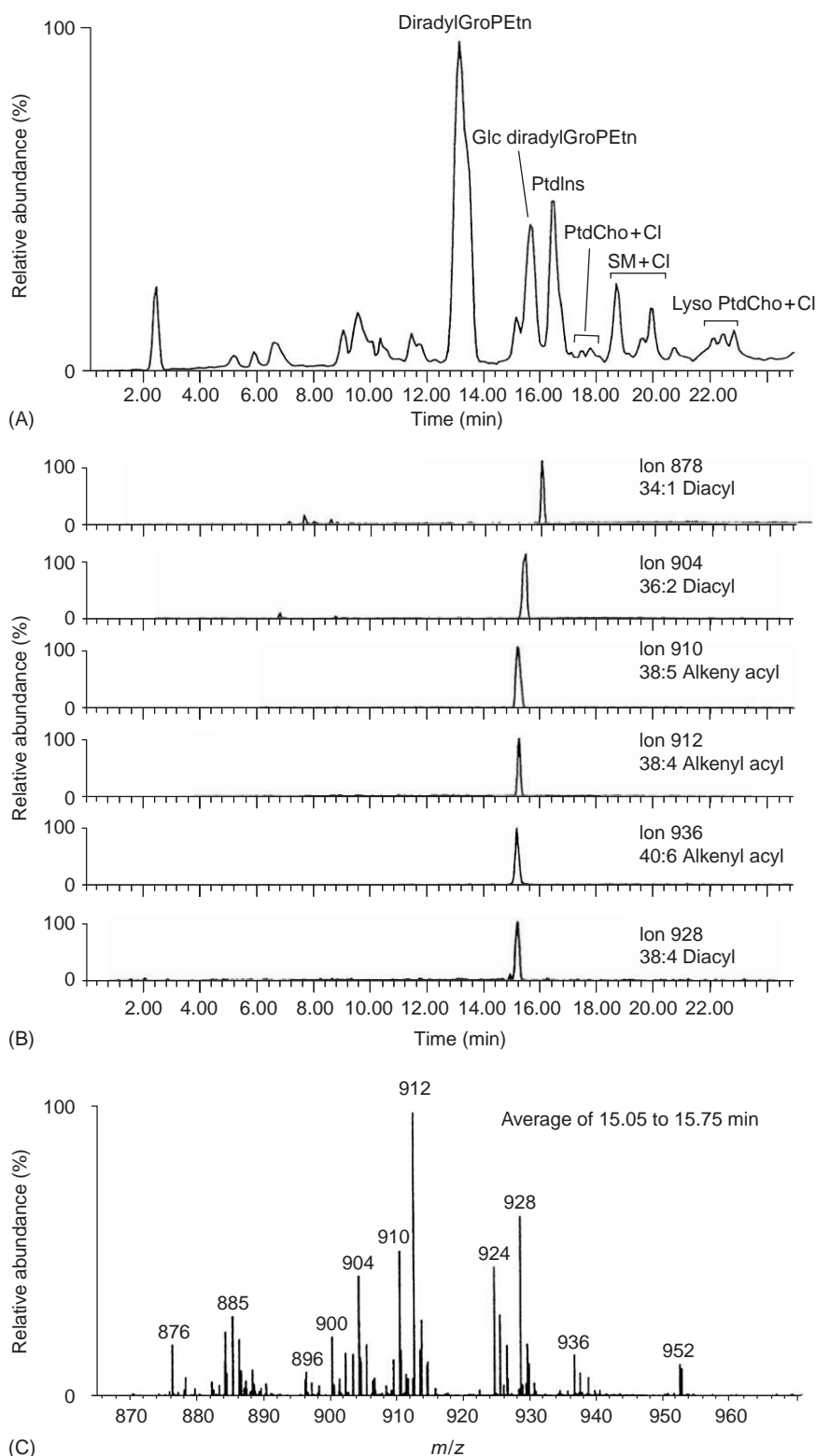
Both lipid class and molecular species composition of glucosylated and nonglucosylated LDL phospholipids can be determined by a combination of normal-phase HPLC with online electrospray MS. **Figure 1** illustrates the normal-phase LC–MS resolution and identification of the major molecular species of the glucosylated ethanolamine phospholipids of LDL in the negative ion mode, while **Table 1** gives the relative proportions of the glucosylated and nonglucosylated phospholipid species. The glucosylation of the ethanolamine phospholipids of LDL was largely indiscriminate; both phosphatidylethanolamine (PtdEtn) and alkenylacyl GroPEtn components were glucosylated.

Schiller and Arnold have reviewed the analysis of phospholipids in HDL and LDL by MALDI time-of-flight-MS (MALDI-TOF-MS) and have compared the results with the data obtained by high-resolution  $^{31}\text{P}$  NMR spectroscopy. Differences between LDL and

HDL in sphingomyelin (SM) and phosphatidylcholine (PtdCho) content could be monitored by NMR and MS, and differences with respect to the extraction efficiency were found by MALDI-TOF-MS. Although MALDI-TOF-MS provided a reliable separation and reproducible lipid profile, only limited information was obtained on fatty acid composition.

LysoPtdCho can be determined by MALDI-TOF-MS and this technique is suitable for monitoring the digestion of PtdCho by pancreatic phospholipase  $\text{A}_2$ . The lysoPtdCho was quantified in relation to an internal standard of known concentration and by signal-to-noise ratios. All classes of lysophospholipids could be easily and sensitively analyzed by MALDI-TOF-MS. The smallest detectable amount of lysophospholipids was 0.09 pmol. This is about two orders of magnitude lower than the amount detectable by standard chromatographic methods. Earlier workers had used a simple and sensitive radioenzymatic assay for lysoPtdOH quantification. For this purpose, a recombinant rat lysoPtdOH acyl transferase (LPAAT) produced by *Escherichia coli* was used. In the presence of [ $^{14}\text{C}$ ]oleoyl-CoA, LPAAT selectively catalyzes the transformation of lysoPtdOH and alkyl-lysoPtdOH into [ $^{14}\text{C}$ ]PtdOH. Acylation of lysoPtdOH was complete and linear from 0 to 200 pmol with a minimal detection of 0.2 pmol. The method was used to quantify lysoPtdOH in lipids from bovine sera, as well as from human and mouse plasma. Cesium chloride (CsCl) had earlier been proposed as an auxiliary reagent for the analysis of PtdCho mixtures by MALDI-TOF-MS. In biological samples various cations are present (mainly  $\text{H}^+$ ,  $\text{Na}^+$ , and  $\text{K}^+$ ) and, therefore, a mixture of different adducts is formed. Since phospholipids exhibit a wide distribution of different fatty acid residues a considerable peak overlap may occur. This problem can be overcome by mixing the analyte with CsCl, which yields naturally nonoccurring  $\text{Cs}^+$  adducts that are apparent due to the large shift of the molecular mass.

Plasma lysophospholipids of diagnostic significance can be determined by electrospray ionization (ESI) MS. LysoPtdOH, lysoPtdIns, lysoPtdSer, and lysoPtdCho were detected at high sensitivity (in low picomole range) by this approach. The major ions and their identities were 409 (16:0-lysoPtdOH), 433–437 (18:2-, 18:1-, and 18:0-lysoPtdOH), 457 41–482 571 (16:0-lysoPtdIns), 599 (18:0 lysoPtdIns), and 619 (20:4-lysoPtdIns). Patients with either ovarian or peritoneal cancer have higher levels of both lysoPtdOH and lysoPtdIns compared with healthy controls: 457 (20:4-lysoPtdOH?), 481–482 (22:6-lysoPtdOH?), 571 (16:0 lysoPtdIns), etc.



**Figure 1** LC/ESI-MS analysis of oxidized PtdCho in oxidized LDL. Total positive ion current profile of oxidized LDL; (A) single ion plots of representative PtdCho oxidation products. Peak identification is as given in figure. LDL was oxidized by incubation with  $5 \mu\text{mol l}^{-1}$   $\text{CiSO}_4$  in  $0.1 \text{ mol l}^{-1}$  phosphate buffered saline (PBS) for 12 h at  $37^\circ\text{C}$ . The total lipid extract of the oxidized LDL was dissolved in chloroform/methanol (2:1, v/v) and  $20 \mu\text{l}$  of the sample containing  $10 \mu\text{g}$  lipid was analyzed. Structural assignment for aldehydes and hydroperoxides are according to reference standards. Ions 832 and 830 were identified on the basis of retention time and molecular weight. Reproduced with permission of publisher from Ravandi A, Kuksis A, and Shaikh NA (2000) *Arteriosclerosis, Thrombosis, and Vascular Biology* 20: 467–477.



**Table 1** Molecular species of glucosylated and non-glucosylated ethanolamine containing phospholipids of LDL (mol.%)

Species	Glucosylated (mol.%)	Nonglucosylated (mol.%)
<i>Alkenylacyl</i>		
16:0/20:4	1.23 ± 0.64	4.25 ± 1.39
18:0/18:3		2.09 ± .97
16:0/20:3		
16:0/22:6	2.36 ± 1.38	5.51 ± 1.86
18:2/20:4		
18:0/20:5	2.56 ± 0.89	10.75 ± 2.14
18:0/20:4	3.71 ± 0.54	19.30 ± 4.53
18:0/20:3	0.58 ± 0.19	3.63 ± 0.23
18:0/22:6	0.24 ± 0.10	3.45 ± 1.87
Total	10.68 ± 3.69	48.98 ± 3.37
16:0/18:2	12.71 ± 2.83	3.84 ± 1.24
16:0/18:1	3.78 ± 0.88	1.20 ± 0.25
16:0/20:4	7.35 ± 2.01	3.45 ± 1.32
18:0/18:3	3.54 ± 1.61	2.29 ± 1.68
18:0/18:2	15.34 ± 4.22	8.11 ± 3.17
18:0/18:1	3.28 ± 1.25	1.99 ± 0.79
16:0/22:6	6.98 ± 1.98	5.51 ± 2.43
18:2/20:4		
18:0/20:5	11.78 ± 3.49	6.10 ± 1.81
16:0/22:5		
18:0/20:4	18.49 ± 4.65	12.36 ± 3.92
18:0/20:3	2.18 ± 0.75	1.89 ± 0.65
18:0/22:6	1.65 ± 0.58	1.41 ± 1.64
18:0/22:5	2.24 ± 1.19	2.88 ± 0.85
Total	89.32 ± 5.33	51.00 ± 4.34

All values in mean ± SD ( $n = 4$ ).

<sup>a</sup>CN:DB, acyl carbon number: double bond number.

Lipids were extracted and different phospholipid classes were separated by normal-phase silica column HPLC and resolved peaks analyzed by online ESI-MS.

Reproduced with permission in modified form from Ravandi A, Kuksis A, and Shaikh NA (2000) *Arteriosclerosis, Thrombosis, and Vascular Biology* 20: 467–477.

Analyses of lysophospholipids, including lysoPtdOH, lysoPtdIns, lysoPtdCho, lysoPtdSer, sphingosine-1-phosphate, and sphingosine phosphorylcholine in human body fluids from subjects with different pathophysiological conditions, revealed not only the relevance of lysophospholipids in human disease, but also their potential application as biomarkers and/or therapeutic targets. These compounds are detected by a variety of methods, no single method is available.

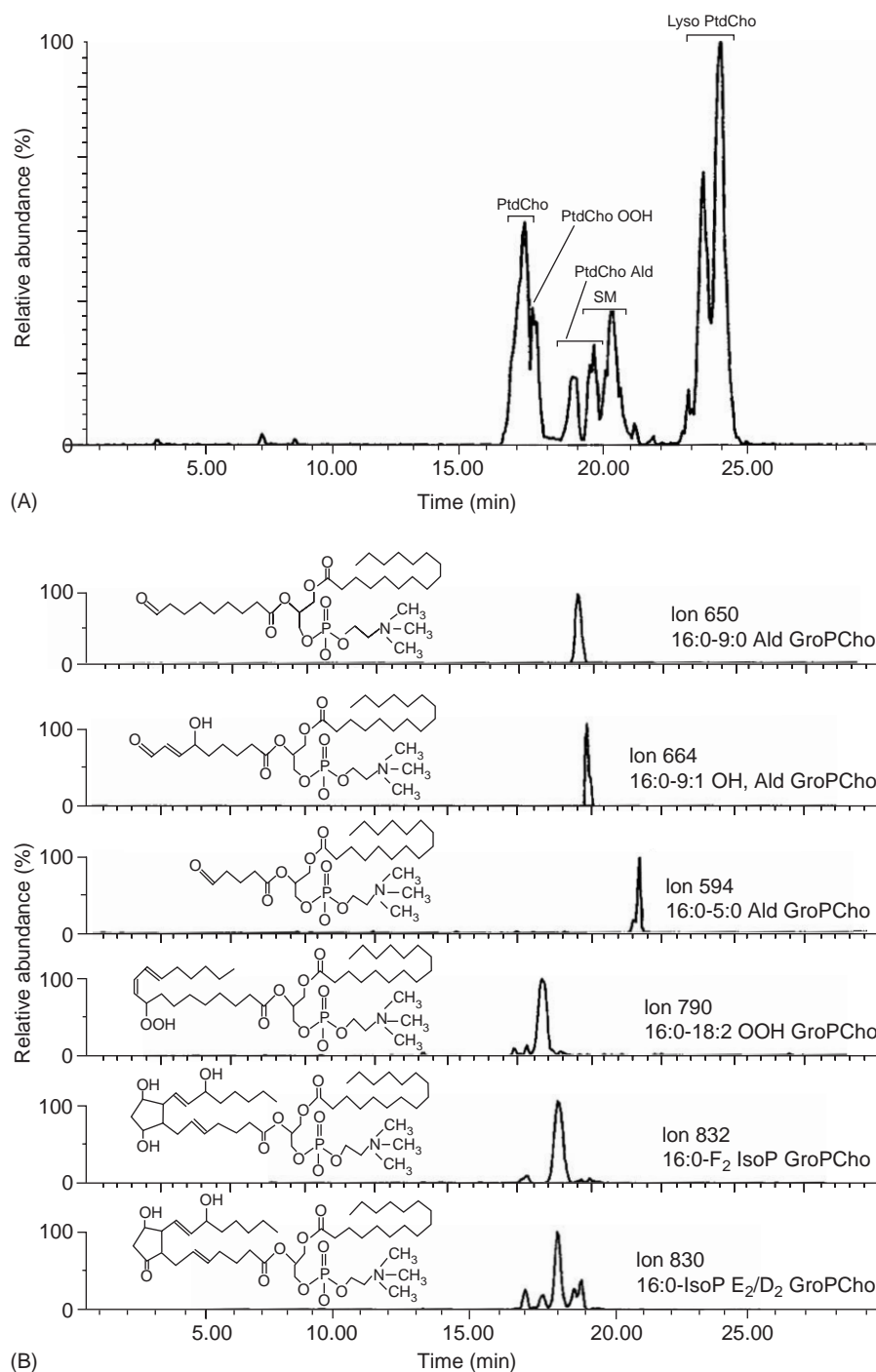
The oxidation products of plasma and lipoprotein phospholipids have been the subject of extensive study by chromatographic and mass spectrometric methods. Earlier, workers had reported the simultaneous determination of phospholipid hydroperoxides and cholesteryl ester hydroperoxides in human plasma by HPLC with chemiluminescence monitoring. Lipid hydroperoxides were quantitatively extracted from human plasma with a mixture of *n*-hexane and

ethyl acetate, and separated by column-switching HPLC using one aminopropyl column and two octyl columns followed by chemiluminescence detection. More specific identification of the hydroperoxides and isoprostanes in plasma phospholipids are obtained by GC coupled with MS (GC–MS), as described for urine and other body fluids (see below). **Figure 2** shows the total ion current profile and the characteristic single ion chromatograms representing the major core aldehyde, hydroperoxide, and isoprostane-containing PtdChos.

<sup>31</sup>P NMR studies of lipoprotein subfractions were first performed in the early 1970s and were dedicated to the investigation of the structure and motion of phospholipids in human plasma lipoproteins. In the 1990s, it was shown that <sup>31</sup>P NMR investigations of blood and of isolated lipoprotein subfractions gave results that agreed well in accuracy and reproducibility with data obtained by enzymatic and TLC methods. <sup>31</sup>P NMR spectroscopic characterization of erythrocyte phospholipids is a well-established technique. Both one- and two-dimensional NMR spectroscopy have been applied to determine quantitatively the lipids extracted from human erythrocyte membranes. The molecular species of the phospholipids were determined after a HPLC fractionation of the lipid extracts and NMR analysis of each lipid class. The results of proton NMR analysis show good agreement with classical lipid analytical techniques. **Figure 3** shows <sup>31</sup>P NMR spectrum of sodium cholate-treated human hemolysate. It is evident that several resonances can be assigned to the various phospholipid classes in human erythrocytes, whereas these resonances are totally absent in the relevant section of the <sup>31</sup>P NMR spectrum of untreated human erythrocytes. <sup>31</sup>P NMR spectroscopy is a convenient and precise analytical tool for the phospholipid analysis of extracts from biological samples of body fluids. Because of its limited sensitivity, <sup>31</sup>P NMR spectroscopy is best suited for work with lipid-rich body fluids. Thus, <sup>31</sup>P NMR spectroscopy has been applied to human blood to monitor chemotherapy-associated changes of serum phospholipids in patients with malignant lymphomas. During therapy leading to remission of cancer, resonances of phospholipids increased progressively, resulting in spectra similar to those seen in the normal sera. The spectral resolution is not high enough to identify molecular species of glycerophospholipids.

**Cholesterol, plant sterols, and bile acids** Measurements of cholesterol and plant sterols in plasma and lymph are obtained by GC, but HPLC has also gained in importance as a method of sensitive

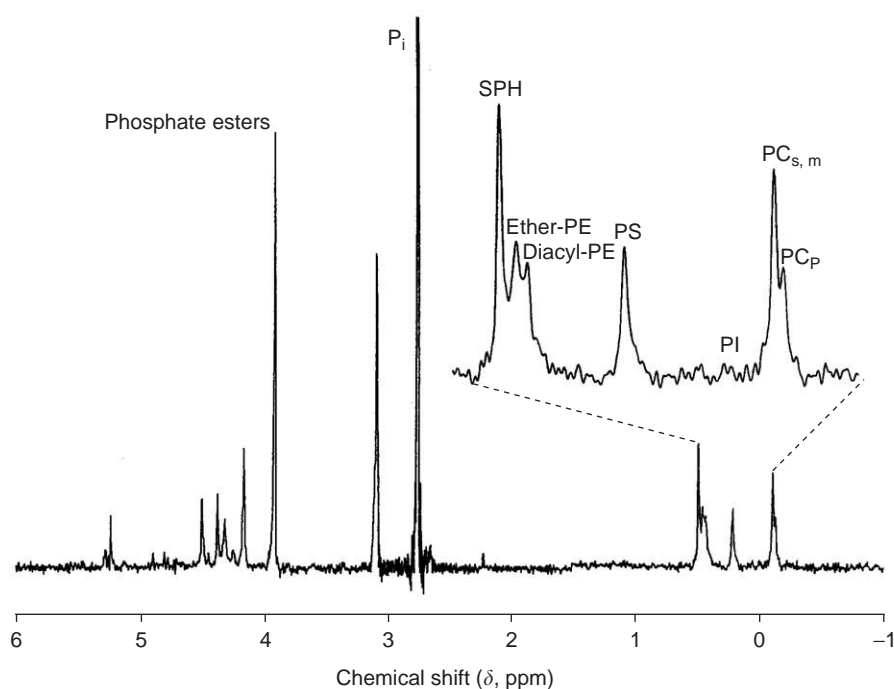




**Figure 2** LC/ESI-MS analysis of oxidized PtdCho in oxidized Glc-PtdEtn LDL. Positive total ion current profile of oxidized glucosylated (Glc)-PtdEtn LDL (A) Single ion plots of representative PtdCho oxidation products, (B) Structural assignments for aldehydes and hydroperoxides are according to reference standards. Ions 832 and 830 were identified on the basis of HPLC retention time and molecular weight. Reproduced with permission from Ravandi A, Kuksis A, and Shaikh NA (1999) *Journal of Biological Chemistry* 274: 16494–16500.

determination following conversion to UV absorbing derivatives. The content of oxysterols in human circulation can be estimated by GC-MS after conversion into trimethylsilyl ether derivatives. **Table 2** illustrates the ratio of oxysterol to cholesterol

(nanogram oxysterol per milligram cholesterol) in isolated lipoprotein fractions and in unfractionated plasma. Unconjugated C<sub>27</sub> bile acids in plasma and blood samples can be determined by MS/MS. The unconjugated di- and tri-hydroxy C<sub>27</sub> bile acids in



**Figure 3**  $^{31}\text{P}$  NMR spectrum of sodium cholate-treated human hemolysate. Resonance assignments for the phospholipids indicated on the spectra were confirmed by adding authentic phospholipids to the NMR samples; and selective hydrolysis of alkenylacyl (plasmalogen) and diacyl phospholipids with HCl and phospholipase  $\text{A}_1$ , respectively.  $\text{PC}_{\text{s,m}}$  and  $\text{PC}_{\text{p}}$  resonances were tentatively assigned. NMR parameters are given in original reference; pH of samples was 8.  $\text{PC}_{\text{p}}$ , polyunsaturated-diacyl glycerophosphocholine;  $\text{PC}_{\text{s,m}}$  saturated and monounsaturated diacyl glycerophosphocholine; PS, phosphatidylserine; P, phosphatidylethanolamine; SPH, sphingomyelin; and  $\text{P}_i$ , inorganic phosphate. Reproduced with permission from Nouri-Sorkhabi MH, Wright LC, Sullivan DR, and Kuchel PW (1996) *Lipids* 31: 765–770.

**Table 2** Content of oxysterols in isolated lipoprotein fractions (VLDL, LDL, HDL), lipoprotein-free plasma (LFP), and unfractionated plasma (plasma)

Oxysterol	$\text{ng ml}^{-1}$ equivalent of plasma				
	VLDL	LDL	HDL	LFP	Plasma
7 $\alpha$ -Hydroxycholesterol	2	32	13	5	44
7 $\beta$ -Hydroxycholesterol	1	15	7	4	8
Cholesterol-5,6 $\alpha$ -epoxide	2	17	8	6	9
Cholesterol-5,6 $\beta$ -epoxide	4	60	26	16	30
Cholestane-3 $\beta$ ,5 $\alpha$ ,6 $\beta$ -triol	1	8	3	9	21
7-Oxcholesterol	1	16	7	8	12
24-Hydroxycholesterol	3	33	31	12	83
25-Hydroxycholesterol	1	5	2	4	7
27-Hydroxycholesterol	3	62	75	18	159
Cholestenic acid	0	2	4	118	118

Mean of seven samples (from seven individual subjects, sample volume 1.0 ml).

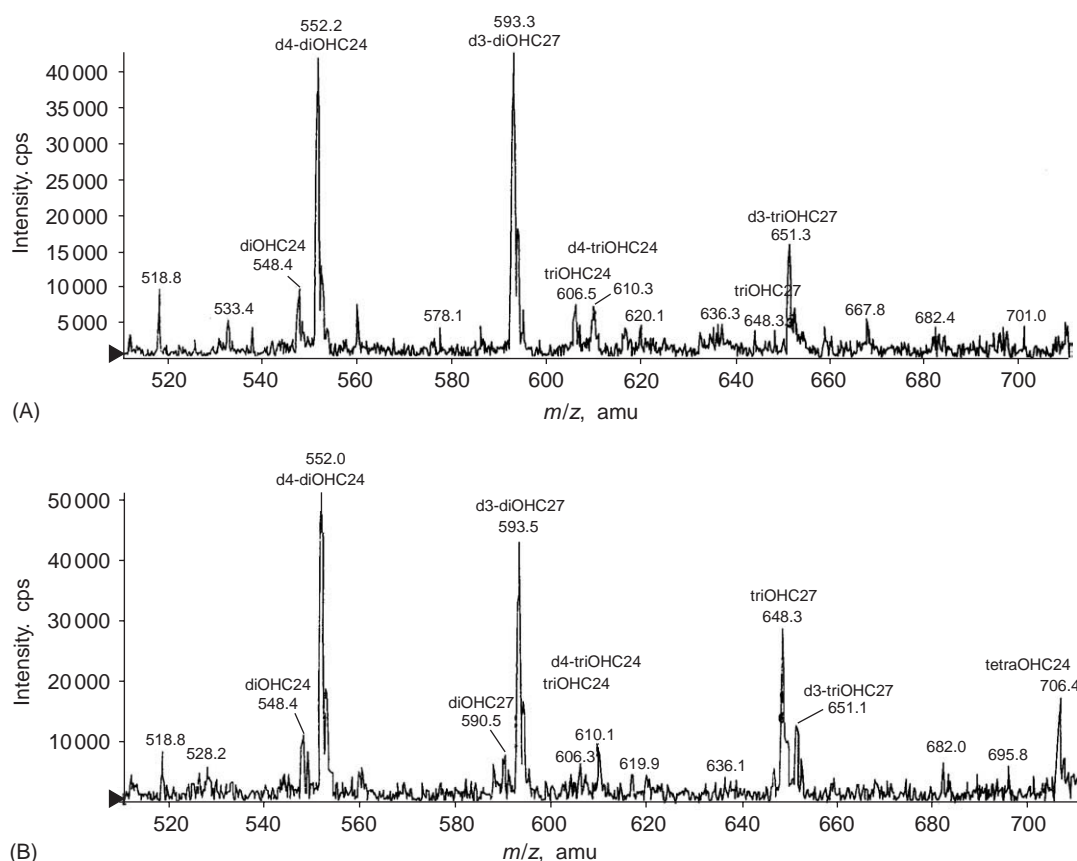
VLDL, very low-density lipoproteins; LDL, low-density lipoproteins; HDL, high-density lipoproteins.

Reproduced with permission in modified form from Babiker A and Diczfalussy U (1998) *Biochimica Biophysica Acta* 1392: 333–339.

5  $\mu\text{l}$  plasma samples and 3 mm blood spots are quantified using deuterium-labeled internal standards. Patients with peroxisomal disorders, who lack the ability to chain-shorten the  $\text{C}_{27}$  bile acid

intermediates into  $\text{C}_{24}$  bile acids, have elevated levels of  $\text{C}_{27}$  bile acids. **Figure 4** shows the ESI-MS/MS analysis of the acetyl dimethylaminoethyl esters extracted from 3 mm dried blood spots. The neutral loss of 60 Da scan for a control blood spot and for a blood spot from a patient is characteristic of a peroxisome biogenesis defect. The reversed-phase LC-ESI-MS method is also used for the analysis of conjugated bile acids.

**Antioxidant levels** HPLC is a well-suited technique for the determination of tocopherols, tocotrienols, ubiquinols, and ubiquinones in biological samples. An automated method for the determination of the total antioxidant capacity (TAC) of human plasma, based on the crocin bleaching assay, is now widely used. The TAC method shows good agreement with the total antioxidant status (TAS) test. Reference value derived from samples of normal blood donors was  $1.175 \pm 0.007 \text{ mmol l}^{-1}$ , while a diet rich in antioxidants more than doubled this value. **Table 3** illustrates the plasma carotenoid and fat-soluble vitamin concentrations of men and women consuming plant steryl esters. There were significant decreases in  $\beta$ -carotene,  $\alpha$ -carotene, and  $\beta$ -cryptoxanthin (females only) not associated with changes



**Figure 4** ESI-MS/MS analysis of the acetyl dimethylaminoethyl esters extracted from 3 mm dried blood spots. A neutral loss of 60 Da scan is shown for (A) a control blood spot, and (B) a blood spot from a patient with peroxisome biogenesis defect (Zellweger syndrome variant). The x-axis shows the molecular weights of the protonated molecular ions that lose acetic acid. Reproduced with permission from Johnson DW, ten Brink HJ, Schuit RC, and Jakobs C (2001) *Journal of Lipid Research* 42: 9–16.

in plasma lipids. Total plasma carotenoids decreased 9.6% with plant steryl ester consumption.

### Lymph

High-temperature GLC profiles of polyunsaturated triacylglycerols are greatly improved by hydrogenation, resulting in a higher recovery of the long-chain species. Lymph lipids can be analyzed by reversed-phase HPLC using evaporative light scattering detection to quantify the lymph triacylglycerols. The triacylglycerols are identified by atmospheric pressure chemical ionization MS. **Figure 5** shows the total ion current chromatogram of lymph triacylglycerols 2 h after administration of various structured triacylglycerols. The triacylglycerols are identified by their ammonium adduct molecular ions and diacylglycerol fragment ions, and by the relative intensity of the fragment ions.

### Milk

No single method or combination of methods is adequate to prepare fatty acid methyl esters (FAME)

from all classes of milk or rumen lipids, which does not simultaneously affect the conjugated dienes. The availability of long polar capillary columns (50–100 m) for GC has improved the resolution of many positional and geometric fatty acid isomers. Prior separations with argentation chromatography, however, still show several regions with overlapping peaks, especially in the mono- and diunsaturated fatty acid region. **Figure 6** shows a typical GC separation of total bovine milk FAMES prepared by NaOCH<sub>3</sub>/methanol followed by HCl/methanol, and separated on a 100 m fused silica capillary column. **Figure 7** shows the expanded methyl eicosanoate to methyl arachidonate region of the GC of total bovine FAMES.

### Urine

Oxidant stress has been widely implicated as a mechanism of disease, but clinical trials of antioxidants have failed to include a biochemical basis for dose selection or patient inclusion. Commonly employed *in vivo* indices of lipid peroxidation are constrained

**Table 3** Plasma carotenoid and fat-soluble vitamin concentrations of men and women consuming plant steryl esters ( $3.6 \text{ g day}^{-1}$ )

Carotenoids and fat-soluble vitamins	<i>LSMean</i> $\pm$ <i>SEE</i> <sup>a</sup>	
	Control	Plant steryl esters
Total carotenoids <sup>b</sup>	116 $\pm$ 2.7	105 $\pm$ 2.7
$\beta$ -Carotene	21 $\pm$ 0.7	18 $\pm$ 0.7
$\alpha$ -Carotene	11 $\pm$ 0.4	9 $\pm$ 0.4
Lutein + Zeaxanthin	15 $\pm$ 0.4	15 $\pm$ 0.4
Anhydrolutein	11 $\pm$ 0.3	11 $\pm$ 0.3
$\alpha$ -Cryptoxanthin	7 $\pm$ 0.2	6 $\pm$ 0.3
$\beta$ -Cryptoxanthin, males	10 $\pm$ 0.5	10 $\pm$ 0.5
$\beta$ -Cryptoxanthin, females	11 $\pm$ 0.6	10 $\pm$ 0.6
Phytofluene	6 $\pm$ 0.2	5 $\pm$ 0.2
$\alpha$ -Tocopherol	1360 $\pm$ 55	1276 $\pm$ 55
$\gamma$ -Tocopherol	205 $\pm$ 11	204 $\pm$ 11
Retinol	41 $\pm$ 2	39 $\pm$ 2

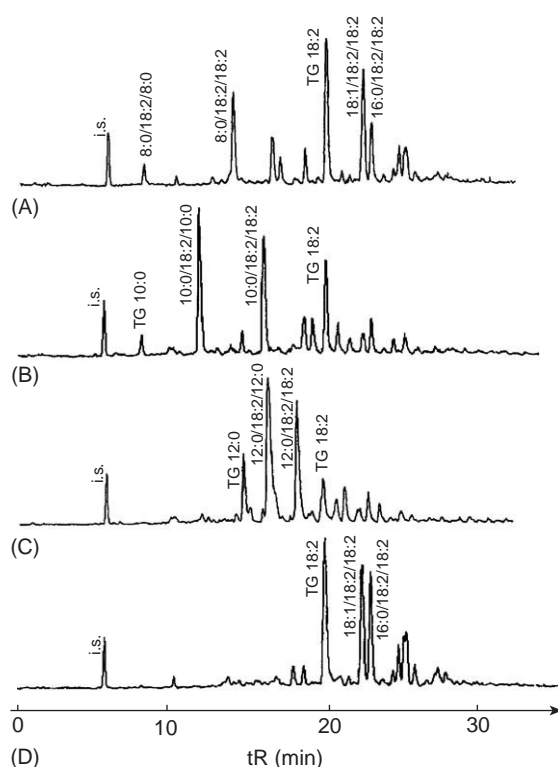
<sup>a</sup> Estimated mean and standard error of the estimate from a mixed model analysis of variance that adjusted for baseline subject characteristics, type of dressing (ranch or Italian), period, and carryover and covariance with blood lipids and lipoproteins;  $n = 53$  (26 men, 27 women).

<sup>b</sup> There was a significant interaction of treatment (plant steryl esters) with body mass index for leucopene.

Reproduced with permission in modified form from Judd JT, Baer DJ, Chen SC, *et al.* (2002) *Lipids* 37: 33–42.

by such issues as the nonspecificity or instability of the target analyte, contamination of the analyte by events *ex vivo*, and nonspecificity of analytical methodology. Specific methodology based on MS applied to both 4-hydroxynonenal and a variety of isoprostanes in human biological fluids, and urinary isoprostanes, resulted in their selection as potential markers of oxidant stress in atherothrombotic disease.

Thus, enhanced urinary excretion of 8-isoprostaglandin  $F_{2x}$  (8-iso-PGF $_{2x}$ ) has been associated with cardiac reperfusion injury and with cardiovascular risk factors, including cigarette smoking, diabetes mellitus, and hypercholesterolemia. The determination of human urinary iso-PGF $_{2x}$  type III by GC–electron impact MS (GC–EI–MS) has been compared to enzymatic assays. The results of GC–EI–MS did not agree with those determined by the enzyme immunoassay and it was apparent that the two methods did not measure the same compounds. An improved assay for the quantification of the major urinary metabolite of the isoprostane 15-F(2t)-8-iso-PGF $_{2x}$  has been proposed using stable isotope dilution GC–MS with negative ion chemical ionization. This assay provides a sensitive and accurate method to assess endogenous isoprostane generation and can be used to further explore the role of oxidant injury in human disease. Evaluation of the isoprostane levels in plasma and urine in type 2 diabetic patients,

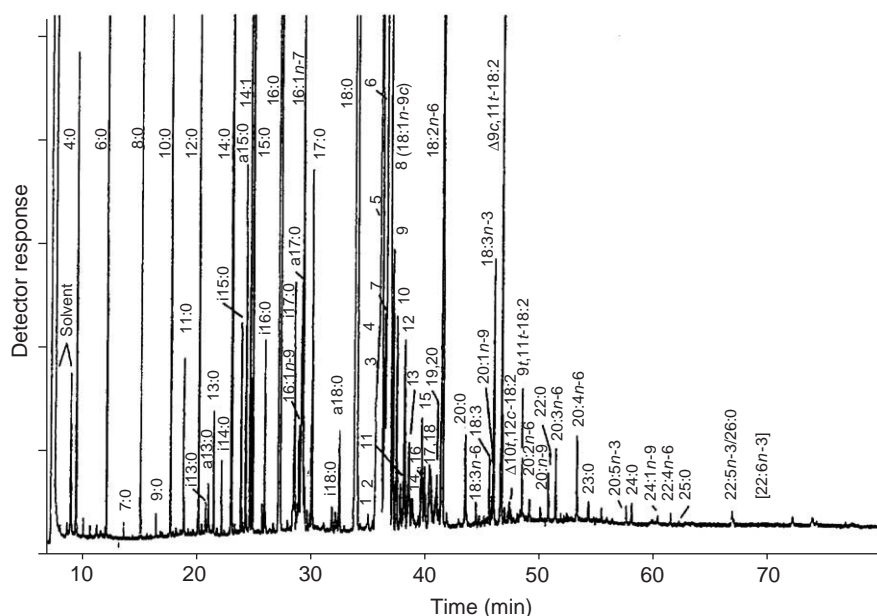


**Figure 5** Total ion current chromatogram of acylglycerols 2 h after administration of structured triacylglycerols (A) 1,3-dioctanoyl-2-linoleoyl-*sn*-glycerol (8:0/18:2/8:0), (B) 1,3-didecanoyl-2-linoleoyl-*sn*-glycerol (10:0/18:2/10:0), and (C) 1,3-didodecanoyl-2-linoleoyl-*sn*-glycerol (12:0/18:2/12:0), and of safflower oil (D). All the triacylglycerols were identified by atmospheric pressure chemical ionization (APCI)–LC–MS. Reproduced with permission from Mu H and Hoy C-E (2001) *Journal of Lipid Research* 42: 792–798.

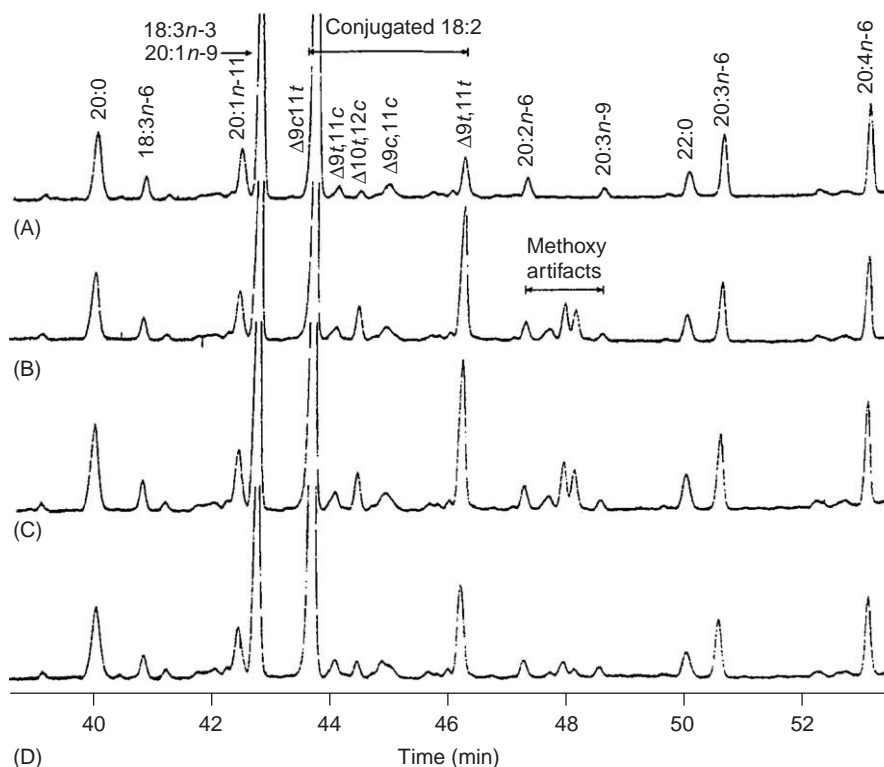
and comparison to other currently used biomarkers of oxidative stress, has indicated that the measurement of same oxidative stress biomarker, isoprostane, in two different biological fluids, plasma and urine, leads to divergent results. Therefore, it is important to measure a biomarker both in circulation fluid (plasma) and in the elimination fluid (urine) to have a general idea of what is occurring in the organism.

### Bronchoalveolar Lavage

Bronchoalveolar Lavage (BAL) fluid covers the alveolae of the lungs as a thin film. The bulk of the BAL lipids are phospholipids with a very high proportion of palmitic acid.  $^{31}\text{P}$  NMR has shown that besides the dipalmitoyl glycerophosphocholine (GroPCho) and glycerophosphoglycerol (GroPGro), minor lipid classes such as plasmalogens are also present and can be quantified. Organic solvent extracts of BAL from different animal species in combination with MALDI-TOF-MS is a very good

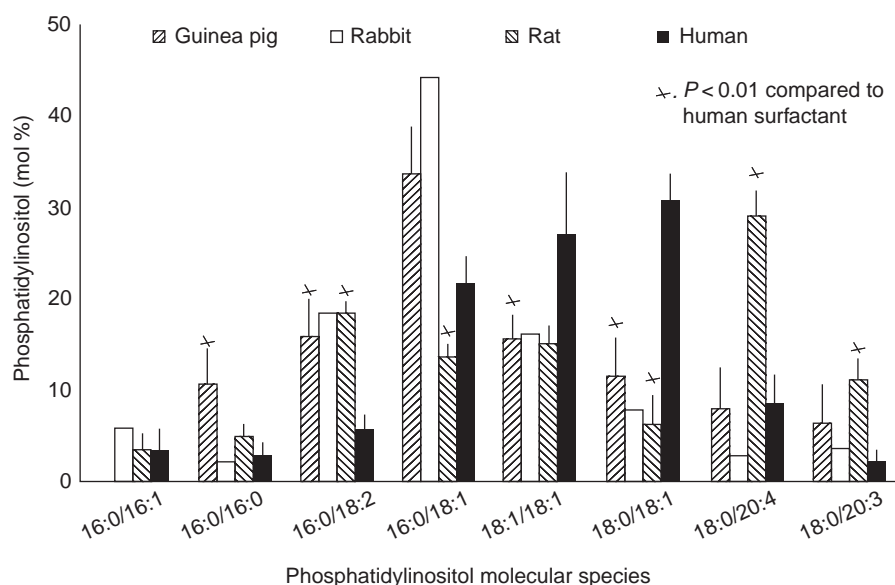


**Figure 6** A typical gas-liquid chromatogram of total bovine milk FAME prepared by  $\text{NaOCH}_3$ /methanol (10 min at  $50^\circ\text{C}$ ) followed by  $\text{HCl}$ /methanol (10 min at  $80^\circ\text{C}$ ), and separated on a 100 m fused silica capillary column (SP-2560; Supelco Inc., Bellefonte, PA). I, iso; a, anteiso; numbers 1–20 are arbitrary consecutive numbering of all peaks in the region between 18:0 and 18:2n-6 using this column. (22:6n-3) is not present in these milk samples but the position at which it should emerge in the chromatogram is indicated. Insert the methyl eicosanoate (20:0) to methyl arachidonate (20:4n-6) region of the GLC of total bovine fatty acid methyl esters prepared by using conventional catalysts. Reproduced with permission from Kramer JKG, Fellner V, Dugan MER, Sauer FD, Mossoba MM, and Yurawecz MP (1997) *Lipids* 32: 1219–1228.



**Figure 7** The methyl eicosanoate (20:0) to methyl arachidonate (20:4n-6) region of the GLC of total bovine fatty acid methyl esters prepared by using the following catalysts in anhydrous methanol: (A)  $\text{NaOCH}_3$ ; (B)  $\text{HCl}$ ; (C) acetyl chloride; or (D)  $\text{BF}_3$ . Other analytical conditions as in **Figure 6**. The fatty acids are identified by reference to both methyl terminal ( $n$  system) and carboxyl terminal ( $\Delta$  system). Reproduced with permission from Kramer JKG, Fellner V, Dugan MER, Sauer FD, Mossoba MM, and Yurawecz MP (1997) *Lipids* 32: 1219–1228. See also: Kramer JKG, Blackadar CB, and Zhou J (2002) *Lipids* 37: 823–835.





**Figure 8** The composition of individual molecular species of phosphatidylinositol in mammalian surfactants as determined by mass spectrometry. Reproduced with permission from Postle AD, Heeley EL, and Wilton DC (2001) *Comparative Biochemical Physiology, Part A* 129: 65–73.

method for the identification of the phospholipid classes along with the molecular species composition. Both  $^{31}\text{P}$  NMR and MALDI-TOF-MS provided fast and reliable information on the lipid composition of the BAL. However, despite of its comparable insensitivity  $^{31}\text{P}$  NMR spectroscopy detected all phospholipid classes in a single experiment and with the same sensitivity, whereas MALDI-TOF-MS failed in the detection of PtdEtn in the presence of higher quantities of PtdCho. While the BAL of all humans contains dipalmitoyl GroPCho, the composition of BAL differs from animal to animal. **Figure 8** shows the composition of the molecular species of phosphatidylinositol (PtdIns), which is a minor component of mammalian surfactants. Many others have measured platelet activation factor (PAF)-like compounds in BAL. Hyperoxia increased the concentration of PAF-like compounds, lipid hydroperoxides, and malondialdehyde in plasma but not in BAL. **Table 4** lists the phospholipids identified by ESI and ESI-MS/MS in Cursurf®, a natural pulmonary surfactant. Cursurf® contains low molecular weight hydrophobic apoproteins and a series of phospholipids. It is used in the treatment or prophylaxis of neonatal respiratory distress syndrome, associated with limited capacity of surfactant production in premature newborns.

### Amniotic Fluids

Using a solvent-reagent system gives narrow line-widths in  $^{32}\text{P}$  NMR spectra of phospholipid extracts when applied to human amniotic fluid. Resolution of the major components is achieved by manipulating

the solvent composition, and assignments can be made by spiking samples with standard compounds. Specifically, estimates of PtdCho/SM ratio and phosphatidylglycerol (PtdGro) and PtdIns were obtained by high-resolution of  $^{31}\text{P}$  NMR.

### Cerebrospinal Fluid

Koch *et al.* determined total cholesterol, phospholipids, and fatty acids in CSF samples from 216 individuals in order to establish the lipid and apolipoprotein levels in Cerebrospinal Fluid (CSF) in a large group of individuals, on the basis of which a classification of CSF lipoproteins was made. The cholesterol and phospholipids are measured enzymatically by fluorometric detection of the reaction products. Earlier work had shown reduced levels of cholesterol, phospholipids, and free fatty acids in Cerebrospinal Fluid (CSF) of Alzheimer disease patients. Urine levels of  $\text{F}_2$ -isoprostanes or their major metabolite were not significantly different between Alzheimer's disease patients and controls. In addition, urine and CSF  $\text{F}_2$ -isoprostanate levels in Alzheimer's disease patients did not correlate. These results indicate that plasma and urine  $\text{F}_2$ -isoprostanes and  $\text{F}_2$ -neuroprostanes do not accurately reflect central nervous system levels of these biomarkers and are not reproducibly elevated in body fluids outside of the central nervous system in Alzheimer's disease patients.

### Saliva

Cholesteryl esters, cholesterol, triacylglycerols, diacylglycerols, monoacylglycerols, and free fatty acids

**Table 4** List of phospholipids identified by ESI and ESI–MS/MS in Curosurf<sup>®</sup>, a natural pulmonary surfactant

Observed ion	Phospholipid class	Molecular species	Molecular weight
490.1	Lyso-PtdCho	14:0	467.3
516.2		16:1	493.3
518.2		16:0	495.3
530.4		17:1	507.3
542.0		18:2	519.3
544.1		18:1	521.4
546.2		18:0	523.4
566.1		20:4	543.3
728.4		14:0/16:0	705.5
742.3		15:0/16:0	719.6
754.3	PtdCho	14:0/18:1 or 16:0/16:1	731.6
756.4		16:0/16:0 (DPPC)	733.6
768.4		15:0/18:1	745.6
780.4		16:0/18:2 or 16:1/18:1	757.6
782.4		16:0/18:1 or 16:1/18:0	759.6
784.4		16:0/18:0	761.6
790.4		15:0/20:4	767.6
804.5		16:0/20:4 or 18:2/18:2	781.6
808.4		18:1/18:1 or 18:0/1:2	785.6
810.4		18:0/18:1	787.6
725.4	SM	16:0	702.6
809.5		22:0	786.6
833.5		24:2	810.6
835.4		24:1	812.7
837.5	LysoPtdEtn	24:0	884.7
500.1		18:2	477.3
502.2		18:1	479.3
524.2	PtdEtn	20:4	501.3
551.9		22:4	529.3
738.4		16:0/18:2 or 16:1/18:1	715.5
740.4		16:0/18:1 or 16:1/18:0	717.5
762.3		18:2/18:2	739.5
764.4		18:1/1:2	741.5
766.5		18:1/18:1 or 18:0/18:2	743.6
768.4		18:0/18:1	745.6
774.7		17:1/20:4 or 18:1/19:4	751.5
788.6		18:1/20:4	765.5
790.6	PtdEtn(vinyl)	18:0/20:4	767.6
724.2		18:1v/18:0	701.3
746.3	PtdSer	18:0v/20:4	723.3
788.3		18:0/18:1	789.6
599.7	LysoPtdIns	18:0	600.3
833.5	PtdIns	16:0/18:2	834.5
835.7		6:0/18:1	836.5
859.7		18:1/18:2	860.6
861.6		18:0/18: or 18:1/18:1	862.6
863.7	PtdGro	18:0/18:1	864.6
883.5		18:1/20:4	884.6
885.5		18:0/20:4	886.6
719.5		16:0/16:1	720.5
721.7		16:0/16:0	722.5
747.7		16:0/18:1	748.5
773.6		18:0/18:2 and 18:1/18:1	774.5

The choline and ethanolamine containing phospholipids yielded Na<sup>+</sup> adducts (plus 23 u); in the negative ion mode, the choline containing phospholipids gave Cl<sup>−</sup> adducts (plus 35 and 37 u). Reproduced with permission in a modified form from Pelizzi N, Catinelli S, Barbosa S, and Zanol M (2002) *Rapid Communications in Mass Spectrometry* 16: 2215–2220.

account for 96–99% of total salivary lipids. The salivary lipids do not float on centrifugation like blood plasma lipids and appear to exist in a different state of aggregation from lipids in blood and lymph. The analyses were performed by TLC using both neutral and polar solvents. The lipids were located by charring and the lipid spots quantified by scanning densitometry at 350 nm.

**See also:** **Gas Chromatography:** Overview. **Lipids:** Overview; Fatty Acids; Polar Lipids. **Mass Spectrometry:** Overview. **Nuclear Magnetic Resonance Spectroscopy-Applicable Elements:** Phosphorus-31.

## Further Reading

- Bessard J, Cracowski JL, Stanke-Labesque F, and Bessard G (2001) Determination of isoprostaglandin F<sub>2α</sub> type III in human urine by gas chromatography-electronic impact mass spectrometry. Comparison with enzyme immunoassay. *Journal of Chromatography B* 754: 333–343.
- Feillet-Coudray C, Chone F, Michel F, *et al.* (2002) Divergence in plasmatic and urinary isoprostane levels in type 2 diabetes. *Clinica Chimica Acta* 324: 25–30.
- Kampa M, Nistikaki A, Tsaousis V, *et al.* (2002) A new automated method for the determination of the total antioxidant capacity (TAC) of human plasma, based on the crocin bleaching assay. *BMC Clinical Pathology* 2: 3.
- Kissmeyer AM and Sonne K (2001) Sensitive analysis of 1α,25-dihydroxyvitamin D<sub>3</sub> in biological fluids by liquid chromatography–tandem mass spectrometry. *Journal of Chromatography A* 935: 93–103.
- Koch S, Donarski N, Goetze K, *et al.* (2001) Characterization of four lipoprotein classes in human cerebrospinal fluid. *Journal of Lipid Research* 42: 1143–1151.
- Kuksis A (2001) Plasma non-cholesterol sterols. *Journal of Chromatography A* 935: 203–236.
- Larsson B, Olivecrona G, and Ericson T (1996) Lipids in human saliva. *Archives of Oral Biology* 41: 105–110.
- Montine J, Quinn JF, Mlatovic D, *et al.* (2002) Peripheral F<sub>2</sub>-isoprostanes and F<sub>4</sub>-neuroprostanes in Alzheimer's disease. *Annals of Neurology* 52: 175–179.
- Morales CR, Terry ES, Zackert WE, Montine TJ, and Morrow JD (2001) Improved assay for the quantification of the major urinary metabolite of the isoprostane 15-F(2t)-isoprostane (8-iso-PGF<sub>2α</sub>) by a stable isotope dilution mass spectrometric assay. *Clinica Chimica Acta* 314: 93–99.
- Pearce JM, Krone JT, and Komoroski RA (1993) Analysis of saturated phosphatidylcholine in amniotic fluid by <sup>31</sup>P NMR. *Magnetic Resonance in Medicine* 30: 476–484.
- Podda M, Weber C, Traber MG, Milbrandt R, and Packer L (1999) Sensitive high-performance liquid chromatography techniques for simultaneous determination of tocopherols, tocotrienols, ubiquinol, and ubiquinones in biological samples. *Methods in Enzymology* 299: 330–341.
- Schiller J and Arnold K (2002) Application of high resolution <sup>31</sup>P NMR spectroscopy to the characterization of the phospholipid composition of tissues and body fluids – a methodological review. *Medical Science Monitor* 8(11): MT205–222.

Wang Y, Liu Y, Ito Y, *et al.* (2001) Simultaneous measurement of anandamine and 2-arachidonoylglycerol by polymyxin B-selective adsorption and subsequent high-performance liquid chromatographic analysis: increase in endogenous cannabinoids in the sera of patients with endotoxic shock. *Analytical Biochemistry* 294: 73–82.

Watkins SM, Reifsnnyder PR, Pan H-j, German BJ, and Leiter EH (2002) Lipid metabolome-wide effects of the

PPAR $\gamma$  agonist rosiglitazone. *Journal of Lipid Research* 43: 1809–1817.

Xiao Y, Chen Y, Kennedy AW, Belinson J, and Xu Y (2000) Evaluation of plasma lysophospholipids for diagnostic significance using electrospray ionization mass spectrometry (ESI-MS) analyses. *Annals of the New York Academy of Science* 905: 242–259.

# LIQUID CHROMATOGRAPHY

Contents

**Overview**

**Principles**

**Column Technology**

**Packed Capillary**

**Mobile Phase Selection**

**Normal Phase**

**Reversed Phase**

**Ion Pair**

**Micellar**

**Size Exclusion**

**Chiral**

**Affinity Chromatography**

**Multidimensional**

**Instrumentation**

**Liquid Chromatography–Mass Spectrometry**

**Liquid Chromatography–Fourier Transform Infrared**

**Liquid Chromatography–Nuclear Magnetic Resonance Spectrometry**

**Amino Acids**

**Chiral Analysis of Amino Acids**

**Biotechnology Applications**

**Clinical Applications**

**Food Applications**

**Pharmaceutical Applications**

**Isotope Separations**

## Overview

**M D Palamareva**, University of Sofia, Sofia, Bulgaria

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The classical methods of recrystallization and distillation fail to provide isolation and proper

purification of more than simple mixtures. Chromatography is a more efficient method in this respect. Any chromatographic system consists of three components: a stationary phase, a mobile phase, and the sample. Liquid chromatography (LC) operates with a liquid mobile phase as opposed to gas chromatography (GC). LC is widely used to follow the course of reactions, to separate complex mixtures, to establish the presence and quantity of specific compounds in physiological materials, etc. Thus, this chromatographic method is daily applied in

Wang Y, Liu Y, Ito Y, *et al.* (2001) Simultaneous measurement of anandamine and 2-arachidonoylglycerol by polymyxin B-selective adsorption and subsequent high-performance liquid chromatographic analysis: increase in endogenous cannabinoids in the sera of patients with endotoxic shock. *Analytical Biochemistry* 294: 73–82.

Watkins SM, Reifsnnyder PR, Pan H-j, German BJ, and Leiter EH (2002) Lipid metabolome-wide effects of the

PPAR $\gamma$  agonist rosiglitazone. *Journal of Lipid Research* 43: 1809–1817.

Xiao Y, Chen Y, Kennedy AW, Belinson J, and Xu Y (2000) Evaluation of plasma lysophospholipids for diagnostic significance using electrospray ionization mass spectrometry (ESI-MS) analyses. *Annals of the New York Academy of Science* 905: 242–259.

# LIQUID CHROMATOGRAPHY

Contents

**Overview**

**Principles**

**Column Technology**

**Packed Capillary**

**Mobile Phase Selection**

**Normal Phase**

**Reversed Phase**

**Ion Pair**

**Micellar**

**Size Exclusion**

**Chiral**

**Affinity Chromatography**

**Multidimensional**

**Instrumentation**

**Liquid Chromatography–Mass Spectrometry**

**Liquid Chromatography–Fourier Transform Infrared**

**Liquid Chromatography–Nuclear Magnetic Resonance Spectrometry**

**Amino Acids**

**Chiral Analysis of Amino Acids**

**Biotechnology Applications**

**Clinical Applications**

**Food Applications**

**Pharmaceutical Applications**

**Isotope Separations**

## Overview

**M D Palamareva**, University of Sofia, Sofia, Bulgaria

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The classical methods of recrystallization and distillation fail to provide isolation and proper

purification of more than simple mixtures. Chromatography is a more efficient method in this respect. Any chromatographic system consists of three components: a stationary phase, a mobile phase, and the sample. Liquid chromatography (LC) operates with a liquid mobile phase as opposed to gas chromatography (GC). LC is widely used to follow the course of reactions, to separate complex mixtures, to establish the presence and quantity of specific compounds in physiological materials, etc. Thus, this chromatographic method is daily applied in

laboratories such as in analytical chemistry, organic synthesis, biochemistry, medicine, and ecology and in production processes in industry. This article deals briefly with the history, importance, and different classifications of LC. Moreover, the parameters connected with the retention and separation of analytes are discussed along with the correlations among them. These parameters are related to the structure of analytes and the chromatographic conditions used. Their adjustment is crucial for obtaining good or optimum separation.

## LC versus GC

In 1903, a botanist Mikhail Tswett achieved unexpectedly the separation of chlorophyll and xanthophylls into their components. In the method he used, a petroleum ether solution of any of these mixtures is passed through calcium carbonate placed in a glass tube (column). This new separation method was called 'Chromatography'; i.e. 'colored writing', owing to the development of separated colored zones within the length of the column. Any of these zones corresponds to a specific component. This experiment is the beginning of both chromatography and LC. In this case, calcium carbonate was the solid stationary phase and petroleum ether (a mixture of hydrocarbons) is the liquid mobile phase. Since then, column liquid-solid chromatography (LSC) has been widely used. In the case of column chromatography (CC), the stationary phase is a three-dimensional bed. The column contains the stationary phase and the analyte. The mobile phase is added at the top of the column under atmospheric pressure. Its movement results in different retention of the analyte components according to their structure.

The development of LC was rapid with the appearance of several different methods and techniques. Some trends are outlined. Liquid-liquid chromatography (LLC) was introduced in 1941. In this case, both the stationary phase and mobile phase are immiscible liquids. The stationary phase is a porous material (support) covered with a thin film of liquid. A variation of LLC is paper chromatography (PC), where the stationary phase is paper with the water included in its pores. PC was the first planar technique. The term 'planar' comes from the fact that the stationary phase (paper) is a two-dimensional bed. A subsequent planar technique is thin-layer chromatography (TLC). In this case, the stationary phase is a thin layer of solid material, composed of small particles, spread on a glass plate or an aluminum sheet.

High-performance liquid chromatography (HPLC) is the modern version of CC.

GC requires elevated temperatures. It is an indispensable method for the separation of volatile organic compounds that do not decompose at higher temperatures.

LC is usually performed at room temperature. Thus, it is suitable for analysis of all kinds of compounds: organic and inorganic compounds, compounds of low and high molecular mass, and labile compounds such as explosives and stable compounds. The conventional CC and planar techniques do not require the use of expensive instruments. HPLC competes with GC in precision and effectiveness. However, the reproducibility is usually lower. If necessary, a high reproducibility is obtained with more precautions.

## Interaction of Analyte with Both Phases

Let us consider that the analyte is composed of one compound. The analyte interacts in a specific way with both the stationary phase (s) and the mobile phase (m). The interactions are usually weak (solvation, adsorption, etc.) without formation of chemical bonds. An electrostatic interaction occurs in specific cases only. According to its structure, an analyte X interacts better with the stationary phase by sorption or mobile phase by desorption. Equilibrium processes between (1) the analyte and stationary phase and (2) the analyte and mobile phase take place. These processes are represented in a simpler way by a single equilibrium process:



Within this equilibrium, the molecules of X are sorbed and desorbed and thus move to some extent with the flow of the mobile phase. The equilibrium is characterized by a distribution coefficient  $K_D$ :

$$K_D = \frac{[X]_s}{[X]_m} \quad [2]$$

where  $[X]$  is the molar concentration of compound X in the corresponding phase. Thus,  $K_D$  is a measure of the retention of X in the chromatographic system. The greater  $K_D$  is, the greater the retention of X and vice versa. The plot of  $[X]_s$  versus  $[X]_m$  is called a sorption isotherm. Its shape is different, but usually there is some part that is linear. Performance of LC in this part of the isotherm is most effective. The capacity, and thus efficiency, of a chromatographic system depends on the ratio of the masses of the analyte and the stationary phase. If the quantity of the analyte is greater, the chromatographic system is



overloaded and operates, with a decreased efficiency, in the nonlinear part of the isotherm.

## Mechanisms of Interaction

Depending on the chromatographic system, the analyte interacts in a different way or through a different mechanism with the stationary phase and mobile phase. Four main mechanisms of interactions are known for LC.

**Partition mechanism:** It concerns partition between two immiscible liquids. The analyte has different solubility in each phase. If it is better dissolved in the stationary phase than in the mobile phase, its  $K_D$  is greater (Figure 1A).

**Adsorption mechanism:** The stationary phase contains, on its surface, active sites, and the analyte adsorbs on them (Figure 1B). The mobile phase tries to desorb them.  $K_D$  is greater if adsorption dominates over desorption.

**Ion-exchange mechanism:** The stationary phase contains, on its surface, ions (cations or anions), and the analyte exchanges its own ions with the counterions of the stationary phase. Figure 1C shows an example of separation using an anion exchanger. Such an analyte has a greater  $K_D$  than does a compound where such an exchange is not possible.

**Size-exclusion mechanism:** The stationary phase is a solid material composed of porous particles with specific inner pore diameters. An analyte with smaller size goes into the pores and has a greater  $K_D$ . An analyte larger than the pores moves with the mobile phase outside the porous particles and has a smaller  $K_D$  (Figure 1D).

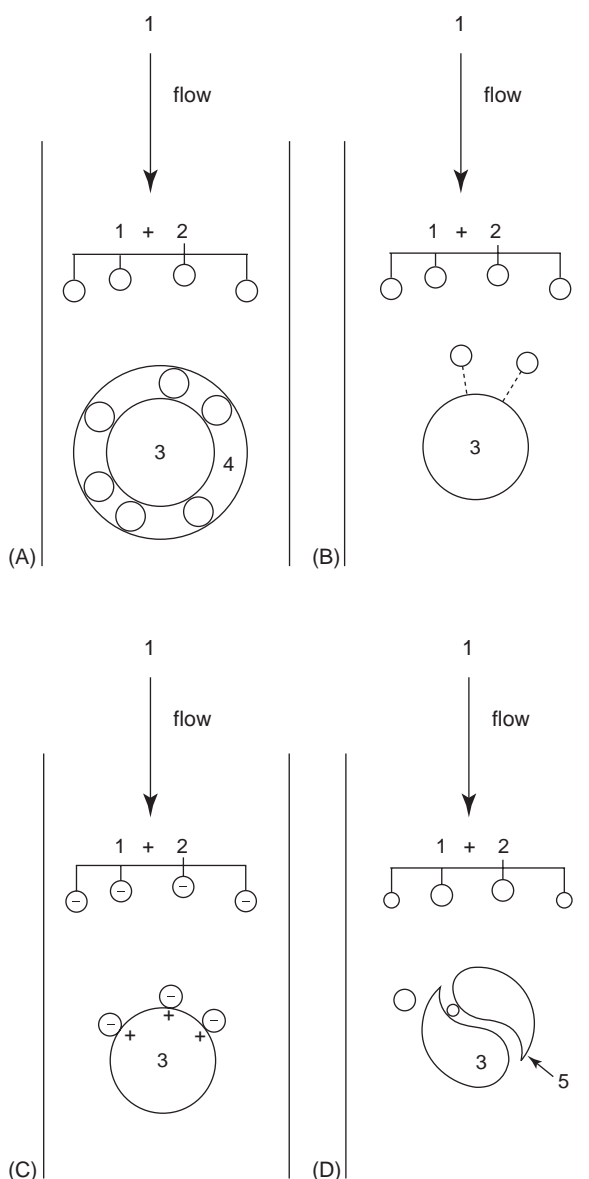
**Other mechanisms:** Affinity and ion-pair LC are based on modifications of the adsorption mechanism and ion-exchange mechanism, respectively.

## Classifications of Liquid Chromatography

LC involves various methods and techniques. This requires its classification from different points of view.

### Classification by the Physical State of Phases

The mobile phase is always liquid. The stationary phase is liquid or solid. Table 1 shows in detail this classification. There are two modifications of LSC: normal phase (NP) and reversed-phase (RP), depending on the relative polarity of the two phases. The same is valid for LLC, but this subdivision is rarely



**Figure 1** Retention mechanisms in LC: (A) partition; (B) adsorption; (C) ion-exchange; and (D) size-exclusion. 1, liquid mobile phase; 2, sample molecules or ions, shown as circles (the latter and '2' are connected by lines); 3, stationary phase (one particle; in case (A), 3 is a support particle covered with thin film of liquid, 4 being the stationary phase itself); 5, an inner pore in a stationary phase particle 3.

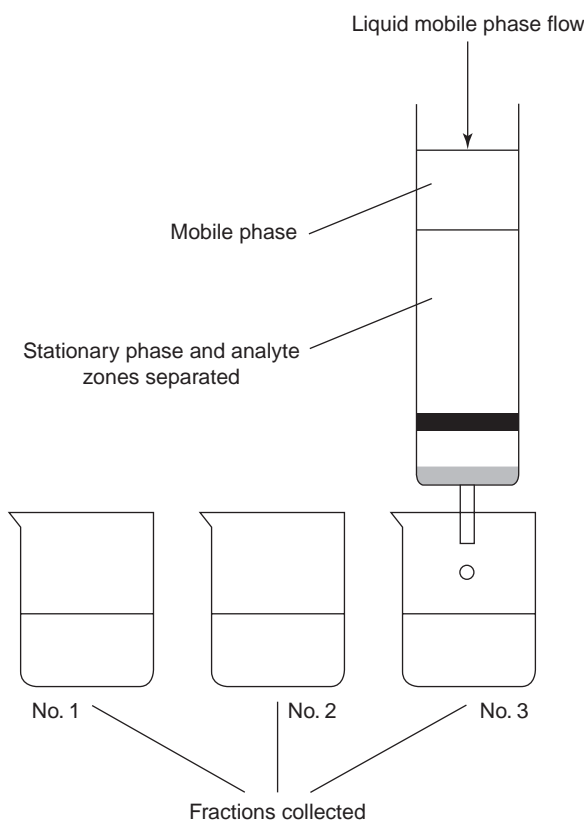
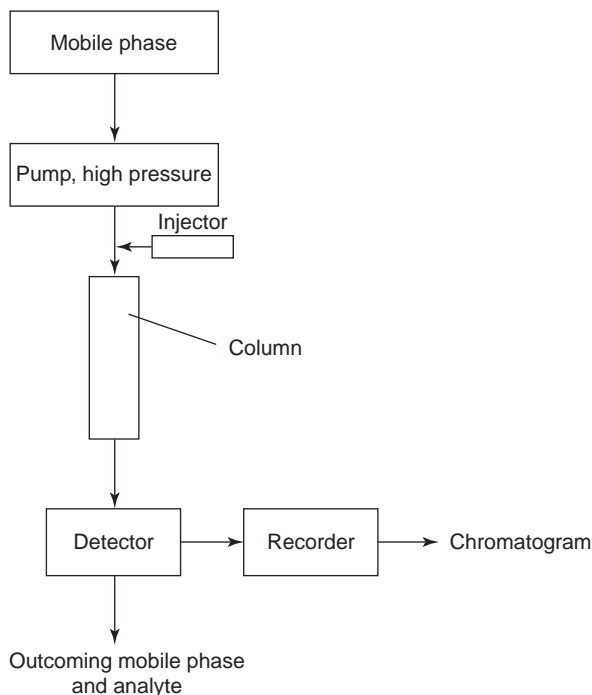
pointed. Methods 1, 2a, and 2b are of general applicability. Methods 3 and 4 apply in specific cases. Method 1 is suitable for separation of organic and inorganic ionic compounds. Methods 2a and 2b are indispensable and widely applied for analysis of nonionic organic compounds. Method 2a is better for separation of isomers including stereoisomers. A modification of LSC is supercritical fluid chromatography, where the mobile phase is a supercritical fluid.

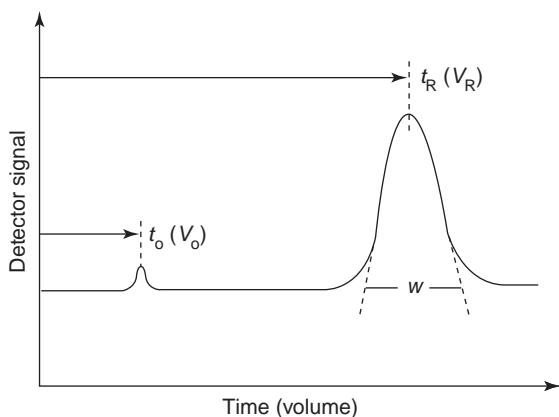
**Table 1** Classification and mechanisms of LC methods

No.	Method	Stationary phase	Main sorption mechanism
1	LLC	Thin film of liquid (on solid)	Partition between two immiscible liquids
2	LSC	Solid	
2a	NP LC	Polar solid	Adsorption
2b	RP LC	Nonpolar solid	Adsorption
3	Ion-exchange chromatography	Ion-exchange resin	Ion-exchange
4	Size-exclusion chromatography	Solid	Sieving

### Classification by the Bed of the Stationary Phase

CC, three-dimensional bed. Conventional CC is performed at atmospheric pressure. The stationary phase particles are large compared with HPLC: the particle diameter ( $d_p$ ) is usually in the range 60–200  $\mu\text{m}$ . Thus, atmospheric pressure is sufficient to overcome the flow resistance of the packed column. This ensures a normal flow rate ( $1\text{--}3\text{ ml min}^{-1}$ ) of the mobile phase. HPLC runs at the higher pressure necessary to overcome the resistance of the smaller particles (usually with  $d_p\ 5\ \mu\text{m}$ ) of the stationary phase. A normal flow rate of the mobile phase is obtained at a pressure of 10–20 MPa. The high efficiency in HPLC is due to the small and uniform size of particles. **Figures 2** and **3** illustrate the equipment necessary for performing conventional CC and HPLC, respectively. In the first case, a unique instrument (chromatograph) is not used. The analyte is applied to the top of a glass column containing the stationary phase. The mobile phase, called the eluent, passes through the column, and this leads to separation of the solute into its components. The outcoming solvent (eluate) from the column is collected in separate fractions, and the compositions are followed using another method. In the case of HPLC, a chromatograph is used. It is composed of six parts: (1) a mobile-phase delivery vessel(s), (2) a pump for producing a high pressure, (3) an injector for application of the analyte, (4) a column containing the stationary phase, (5) a detector (usually UV) giving signals for the composition of the mobile phase exiting from the column, and (6) a data station and data processing unit: the record of the separation is called a chromatogram. The latter is composed of peak(s): any peak corresponds to a specific compound if a complete separation is achieved (**Figure 4**).

**Figure 2** Schematic representation of conventional CC.**Figure 3** Schematic representation of the components of HPLC.



**Figure 4** HPLC chromatogram of a compound.

### Classification by the Composition of the Mobile Phase

To tune the retention of the vast number of known compounds, the mobile phase in LC is usually a mixture of two or more solvents.

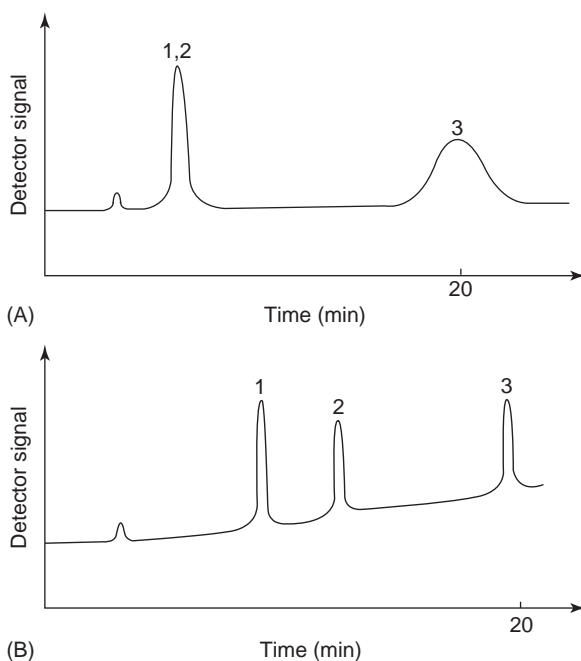
**Isocratic technique.** In this technique, the mobile phase has a constant composition, for example acetonitrile–water 72:28.

**Gradient technique.** In this case, the composition of the mobile phase varies in a specific way, resulting in better separation. It is especially useful if there is a greater difference in the retention of analyte components. It reduces the peak broadening especially for compounds with greater retention times (**Figure 5**). For instance, the gradient change of the mobile-phase composition can be expressed in the following way: mobile phase A = hexane–ethyl acetate 70:30, mobile phase B = hexane–ethyl acetate 30:70, linear gradient from A to B in 16 min.

### Classification by the Analyte Quantity

**Analytical technique.** The quantity of the analyte is small (usually a few micrograms). This chromatographic technique gives an analysis of the analyte composition.

**Preparative technique.** The quantity of the analyte is greater (usually  $\sim 0.5$ –1 g). Thus, the individual components of the analyte are isolated in some quantity. Conventional column LC operates in this mode only. The ratio between the mass of the analyte and stationary phase, i.e., the column capacity, varies usually from 1:50 to 1:100. Chromatographic filtration is a technique where this ratio is smaller, for instance 1:20. It is applied when the analyte is composed of compounds with a greater difference in retention. The efficiency of conventional CC is improved by an increase in pressure (1.2–1.5 atm.) over



**Figure 5** Comparison between (A) isocratic technique and (B) gradient technique.

the column when the stationary phase has smaller particles ( $d_p$  40–60  $\mu\text{m}$ ). To this end, the column is connected with a bottle of liquid nitrogen or helium. This technique is called low-pressure CC or flash chromatography.

HPLC and TLC operate in both modes. In the case of preparative TLC, thick layers of the stationary phase are used.

### Classification by the Possibility of Structural Determination

LC itself does not give the possibility of determining the structure of the individual analyte components in the course of their separation. However, the combination of LC with a specific spectral method enables such a determination. In such a case, after the detector, the mobile phase passes online through the relevant spectrometer. The spectrometer records the spectrum of any peak. LC–mass spectrometry and LC–infrared spectrometry are the most popular techniques. LC–nuclear magnetic resonance spectrometry is becoming increasingly important. If such instruments are not available, structural determination is performed in the classic way. This requires quantitative separation of the analyte and isolation of the individual components in milligram quantities. Spectroscopic instruments that are not connected with the chromatographic system are used to elucidate the structure of the separated compounds.

### Miscellaneous Methods or Techniques

Modification of the mobile phase or the stationary phase leads to new methods or techniques. For instance, inclusion of silver ions from silver nitrate mainly in the stationary phase is used in argentation LC. This method enables a better separation of analytes containing one or more double bonds. The double bond forms a complex with the silver ion and this complex has greater retention, giving the possibility of differentiating the retention of compounds with double bond(s). Modification of the stationary phase with chiral compounds enables separation of the chiral compounds, which are of primary importance nowadays. This technique is called chiral separation.

### Chromatographic Parameters Characterizing Analyte Retention

The distribution coefficient,  $K_D$ , is difficult to determine from the chromatogram. Thus, other chromatographic parameters are defined to characterize the analyte retention directly from the chromatogram.

In the case of HPLC, the retention volume ( $V_R$ ) and retention time ( $t_R$ ) of a compound are equal to the volume of the mobile phase and the time passed until its peak appears in the chromatogram, respectively (see Figure 4). These two parameters are not constant. Their values depend on the column type being characterized by the parameters  $V_o$  (breakthrough or dead volume) and the corresponding time,  $t_o$  (breakthrough or dead time), for an unretained compound. The retention factor,  $k$ , as defined in eqn [3], is a constant measuring the retention.

$$k = \frac{t_R - t_o}{t_o} \quad \text{or} \quad k = \frac{V_R - V_o}{V_o} \quad [3]$$

The greater  $k$  is, the stronger the retention of the compound.

In the case of planar chromatography, the first defined parameter is  $R_F$ :

$$R_F = \frac{S}{S_o} \quad [4]$$

where  $S$  and  $S_o$  are the distances from the start line to the center of the zone and to the front line, respectively. This parameter measures, in fact, the mobility of the analyte since the greater  $R_F$  is, the smaller the analyte retention. The retention factor in TLC,  $R_M$ , is related to  $k$  and  $R_F$ :

$$R_M = \log k = \log \left( \frac{1}{R_F} - 1 \right) \quad [5]$$

In the case of conventional CC, the retention is approximately measured by  $V_R$ . This technique is preparative; the chromatographic system is overloaded by the analyte and thus  $V_R$  varies from separation to separation.

### Chromatographic Parameters Characterizing Analyte Separation

Let us assume that the analyte is composed of two or more compounds. The main goal in LC is to separate a pair of compounds.

Referring to compounds 1 and 2 analyzed under same conditions, the separation factor,  $\alpha$ , is defined in the following way:

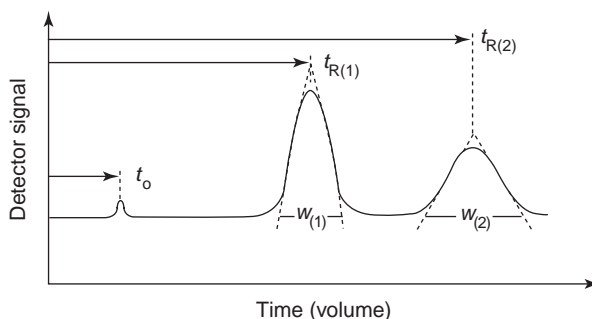
$$\alpha = \frac{k_2}{k_1} \quad \text{or} \quad \log \alpha = \log k_2 - \log k_1 = R_{M(2)} - R_{M(1)} \quad [6]$$

where the subscript specifies the compound. The greater  $\alpha$  is, the better is the separation. No separation is achieved if  $\alpha = 0$  or  $\log \alpha = 1$ . Parameter  $\alpha$  does not take into account the peak width or zone width. However, this factor is important since the wider the peaks are, the poorer the separation. The resolution,  $R$ , depends on the retention and peak width (Figure 6):

$$R = \frac{2[t_{R(2)} - t_{R(1)}]}{W_{(1)} + W_{(2)}} \quad [7]$$

where  $W$  is the width of the corresponding peak. In analogy with a fractionation column, the chromatographic system is considered to consist of a vast number of imaginary plates, called theoretical plates,  $N$ , being easily found from the HPLC chromatogram. The equilibrium between sorption and desorption is assumed complete within any theoretical plate.  $R$  is a function of  $N$ , the mean value of  $k$  for both compounds, and  $\alpha$ :

$$R = \frac{\sqrt{N}}{4} \left( \frac{\alpha - 1}{\alpha} \right) \left( \frac{k}{1 + k} \right) \quad [8]$$



**Figure 6** Parameters necessary for calculation of resolution,  $R$ , using eqn [7].

In LC, the value of  $N$  varies usually from 1000 to 10 000. This factor depends mainly on the stationary phase particle size and is the main reason for the better separation achieved by LC compared with distillation. The values of  $k$  and  $\alpha$  should adjust in such a way as to give a greater value of  $R$  for a reasonable analysis time. To this end, variation of the mobile phase composition and its flow rate,  $F$ , is of significant importance. According to eqn [8],  $R$  is greater at better separation,  $\alpha$ , and stronger retention (greater values of  $k$ ). Reasonable values of  $k$  fall in the range 1–20. Greater values of  $k$  require a longer analysis time, and this is an unfavorable factor. Moreover, the peaks at greater  $k$  become diffuse and asymmetric. This phenomenon is known as peak (band) broadening, being undesired in LC. It leads to partial separation or overlap of two or more peaks. Thus, the suitable conditions for performance of LC are a compromise of various factors.

As mentioned, the flow rate of the mobile phase in the chromatographic system is an important factor for its separation selectivity. It is connected with diffusion and mass transfer in the chromatographic system since LC is a dynamic process. A small flow rate results in greater retention times and strong peak broadening due to diffusion effects. A high flow rate leads to short retention times and peak broadening because the mass transfer between the stationary and mobile phases needs some time. Thus, an optimum flow rate is applied. It is established by the Van Deemter equation. The theoretical plate height,  $H$  (equal to the ratio of bed length,  $L$ , to  $N$ ), is expressed as a function of the linear velocity,  $u$ , of the mobile phase. The flow rate is a product of  $u$  (in  $\text{cm s}^{-1}$ ) and the column diameter. A recent modification of the Van Deemter

equation is

$$H = A + B/u + C_s \cdot u + C_m \cdot u \quad [9]$$

where  $A$  represents the flow anisotropy created by the lack of homogeneity of the column packing, the second term on the right side results from longitudinal diffusion, and the third term from mass transfer into and out of the stationary phase. A plot of  $H$  versus  $u$  gives a curve showing a minimum. The use of the linear velocity,  $u$  (and the relevant flow rate), corresponding to that minimum results in the best separation efficiency of the chromatographic system regarding diffusion and mass transfer.

See also: **Chromatography: Principles. Gas Chromatography: Overview. Liquid Chromatography: Principles; Biotechnology Applications.**

## Further Reading

- Bidlingmeyer BA (1992) *Practical HPLC Methodology and Applications*. New York: Wiley.
- Giddings JC (1991) *Unified Separation Science*. New York: Wiley.
- McMaster MC (1994) *HPLC: A Practical User's Guide*. New York: VCH Publishers.
- Meyer VR (1999) *Practical High-performance Liquid Chromatography*, 3rd edn. Chichester: John Wiley.
- Neue UD (1997) *HPLC Columns: Theory, Technology, and Practice*. New York: Wiley-VCH.
- Poole CF (2003) *The Essence of Chromatography*. Amsterdam: Elsevier.
- Snyder LR, Glajch JL, and Kirkland JJ (1994) *Practical HPLC Method Development*, 2nd edn. New York: Wiley.
- Snyder LR and Kirkland JJ (1979) *Introduction to Modern Liquid Chromatography*, 2nd edn. New York: Wiley-Interscience.
- Touchstone JC (1992) *Practice of Thin Layer Chromatography*, 3rd edn. New York: Wiley.

## Principles

**M D Palamareva**, University of Sofia, Sofia, Bulgaria

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The process in liquid chromatography (LC) is the result of the interaction among an analyte, a stationary phase, and a mobile phase. Moreover, LC is subdivided into liquid–liquid chromatography (LLC) and liquid–solid chromatography (LSC). Both methods are performed in normal-phase (NP) mode and reversed-phase (RP) mode depending on the relative

polarity of both phases. Two other methods are known: ion exchange and size exclusion. The chromatographic parameters  $k$ ,  $R_F$ ,  $R_M$ ,  $\alpha$ , and  $R$  measure retention of analytes and their separation. The relationships among these parameters are important to obtain good separation for reasonable analysis time. All methods are performed by the conventional column chromatography (CC), modern column chromatography, denoted as high-performance liquid chromatography (HPLC), or planar thin-layer chromatography (TLC).

This article deals with the essential theory of LC. Attention is paid to the stationary-phase types,



In LC, the value of  $N$  varies usually from 1000 to 10 000. This factor depends mainly on the stationary phase particle size and is the main reason for the better separation achieved by LC compared with distillation. The values of  $k$  and  $\alpha$  should adjust in such a way as to give a greater value of  $R$  for a reasonable analysis time. To this end, variation of the mobile phase composition and its flow rate,  $F$ , is of significant importance. According to eqn [8],  $R$  is greater at better separation,  $\alpha$ , and stronger retention (greater values of  $k$ ). Reasonable values of  $k$  fall in the range 1–20. Greater values of  $k$  require a longer analysis time, and this is an unfavorable factor. Moreover, the peaks at greater  $k$  become diffuse and asymmetric. This phenomenon is known as peak (band) broadening, being undesired in LC. It leads to partial separation or overlap of two or more peaks. Thus, the suitable conditions for performance of LC are a compromise of various factors.

As mentioned, the flow rate of the mobile phase in the chromatographic system is an important factor for its separation selectivity. It is connected with diffusion and mass transfer in the chromatographic system since LC is a dynamic process. A small flow rate results in greater retention times and strong peak broadening due to diffusion effects. A high flow rate leads to short retention times and peak broadening because the mass transfer between the stationary and mobile phases needs some time. Thus, an optimum flow rate is applied. It is established by the Van Deemter equation. The theoretical plate height,  $H$  (equal to the ratio of bed length,  $L$ , to  $N$ ), is expressed as a function of the linear velocity,  $u$ , of the mobile phase. The flow rate is a product of  $u$  (in  $\text{cm s}^{-1}$ ) and the column diameter. A recent modification of the Van Deemter

equation is

$$H = A + B/u + C_s \cdot u + C_m \cdot u \quad [9]$$

where  $A$  represents the flow anisotropy created by the lack of homogeneity of the column packing, the second term on the right side results from longitudinal diffusion, and the third term from mass transfer into and out of the stationary phase. A plot of  $H$  versus  $u$  gives a curve showing a minimum. The use of the linear velocity,  $u$  (and the relevant flow rate), corresponding to that minimum results in the best separation efficiency of the chromatographic system regarding diffusion and mass transfer.

**See also:** **Chromatography:** Principles. **Gas Chromatography:** Overview. **Liquid Chromatography:** Principles; Biotechnology Applications.

## Further Reading

- Bidlingmeyer BA (1992) *Practical HPLC Methodology and Applications*. New York: Wiley.
- Giddings JC (1991) *Unified Separation Science*. New York: Wiley.
- McMaster MC (1994) *HPLC: A Practical User's Guide*. New York: VCH Publishers.
- Meyer VR (1999) *Practical High-performance Liquid Chromatography*, 3rd edn. Chichester: John Wiley.
- Neue UD (1997) *HPLC Columns: Theory, Technology, and Practice*. New York: Wiley-VCH.
- Poole CF (2003) *The Essence of Chromatography*. Amsterdam: Elsevier.
- Snyder LR, Glajch JL, and Kirkland JJ (1994) *Practical HPLC Method Development*, 2nd edn. New York: Wiley.
- Snyder LR and Kirkland JJ (1979) *Introduction to Modern Liquid Chromatography*, 2nd edn. New York: Wiley-Interscience.
- Touchstone JC (1992) *Practice of Thin Layer Chromatography*, 3rd edn. New York: Wiley.

## Principles

**M D Palamareva**, University of Sofia, Sofia, Bulgaria

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The process in liquid chromatography (LC) is the result of the interaction among an analyte, a stationary phase, and a mobile phase. Moreover, LC is subdivided into liquid–liquid chromatography (LLC) and liquid–solid chromatography (LSC). Both methods are performed in normal-phase (NP) mode and reversed-phase (RP) mode depending on the relative

polarity of both phases. Two other methods are known: ion exchange and size exclusion. The chromatographic parameters  $k$ ,  $R_F$ ,  $R_M$ ,  $\alpha$ , and  $R$  measure retention of analytes and their separation. The relationships among these parameters are important to obtain good separation for reasonable analysis time. All methods are performed by the conventional column chromatography (CC), modern column chromatography, denoted as high-performance liquid chromatography (HPLC), or planar thin-layer chromatography (TLC).

This article deals with the essential theory of LC. Attention is paid to the stationary-phase types,

details of retention mechanisms, and theory of chromatographic process including some quantitative relationships. This knowledge is of importance for selection of a proper method and suitable conditions for a specific separation known as method development. Recommendations in this respect are given. Application of theory requires an understanding of operating principles of the techniques in LC.

## Main Stationary Phases

A stationary phase suitable for LC should meet various requirements such as to be: (1) incompressible, (2) insoluble in the mobile phase, and (3) does not interact with the analyte. Relatively few stationary phases meet these requirements and thus are widely used. They are briefly described. All of them are commercially available in different particle size, being almost uniform or varying in smaller or greater range. In general, the particle diameter in LC is from 1 to 200  $\mu\text{m}$ .

### Silica ( $\text{SiO}_2$ )

The active sites (centers) of silica are mainly the free and bonded silanol groups as shown in **Figure 1**. Two vicinal silanol groups connected by hydrogen bonds form the bonded hydroxyl groups. In the case of normal-phase liquid chromatography (NPLC), these sites interact with the polar analyte functional group/s. Thus, silica is considered as a polar stationary phase. The properties of a specific silica depend on its surface area and the number of active sites.

### Alumina ( $\text{Al}_2\text{O}_3$ )

The active sites of alumina are the aluminum cations ( $\text{Al}^{3+}$ ) and oxygen anions ( $\text{O}^{2-}$ ). Thus, this stationary phase is also polar.

Chemically bonded silica are available. Silica is especially useful in LC and modifications of its active sites are well known.

### Polar-Bonded Silica

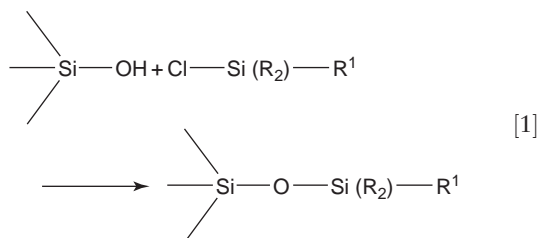
These are spacer-bonded propanediol silica, amino-propylsiloxane-bonded silica, and cyanopropylsiloxane-bonded silica having on their surface, as active sites, hydroxyl (OH) groups, amino ( $\text{NH}_2$ ) groups, and cyano (CN) groups, respectively.



**Figure 1** Main active sites of silica: (A) free silanol group and (B) bounded silanol groups.

### Nonpolar-Bonded Silica

These silica have, as active sites, a siloxane-bonded alkyl group R instead of the silanol groups. There are various reactions leading to such a modification of silica. For example, the reaction of silica with a tri-alkylchlorosilane:



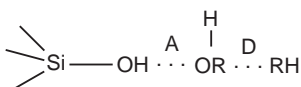
$\text{R}^1$  is usually octyl ( $\text{C}_8\text{H}_{17}$ ) or octadecyl ( $\text{C}_{18}\text{H}_{37}$ ) and is the active site in the case of reversed-phase liquid chromatography (RPLC). The notation RP-18 means that the stationary phase is a siloxane-bonded silica with octadecyl groups on its surface. This stationary phase is nonpolar and thus is suitable for RPLC.

The stationary phase for LLC is usually silica coated with a thin film of liquid. The latter is the stationary phase itself. The coating is performed by treatment of the silica by a solution of this liquid in a suitable solvent and subsequent displacement of the solvent by mobile phase. NP-LLC occurs when the stationary phase is polar contrary to RP-LLC where the stationary phase is nonpolar.

## Detailed Mechanisms

The mechanism in LLC is partition of the analyte between two immiscible liquids with different polarity. Analyte retention is governed by the different dissolution of the analyte components in both liquid phases.

In the case of NPLC, the stationary phase is polar (silica, alumina, or polar-bonded phase) and the mobile phase is nonpolar (hexane, heptane, etc.). Adsorption mechanism operates and involves adsorption of an analyte by its polar groups on the polar active sites of the stationary phase. Adsorption occurs by polar interactions and formation of hydrogen bonds. The nonpolar mobile phase cannot interact with the adsorbent-active sites and thus its adsorption practically does not occur. However, the mobile phase favors the analyte desorption interacting by dispersion forces with its nonpolar groups that are not involved in the adsorption. This is illustrated in **Figure 2** for analysis of an alcohol on silica. This case will be discussed again below. Thus, analyte molecules adsorb and desorb constantly, being



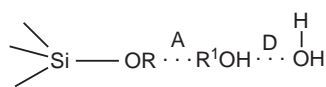
**Figure 2** Detailed adsorption mechanism of retention of an alcohol ROH in NPLC. A, adsorption by hydrogen bonds; D, desorption by dispersion forces.

connected with the stationary phase and mobile phase, respectively. For the time of analysis, the analyte, according to its structure, moves with the mobile phase and this determines its retention in the chromatographic system used. For instance, let us consider NPLC of two alcohols ROH with different length of the R group: hexyl ( $C_6H_{13}$ ) and octyl ( $C_8H_{17}$ ). Hexanol is more polar than octanol. Consequently the first alcohol will be adsorbed stronger (greater retention) than octanol, i.e., the retention order is  $C_6H_{13}OH > C_8H_{17}OH$ . This is an established fact. The adsorption mechanism accounts for the following retention order within organic compounds with the same R group:

- alkane (RH) < haloalkane (RHal)
- < alkene (RH as R contains one double bond)
- < ether ( $ROCH_3$ ) < ester ( $RCO_2CH_3$ )
- < ketone ( $RCOCH_3$ ) < aldehyde ( $RCHO$ )
- < carboxylic acid ( $RCO_2H$ )

Moreover, a higher homolog is less retained than a preceding homologue.

In the case of RPLC, the stationary phase is non-polar (nonpolar-bonded silica) and the mobile phase is polar. Polar mobile phase is, e.g., methanol/water. The mechanism in this case is considered as partition or adsorption. Partition in the sense of different solubility in both phases does not operate. In some cases water is the stationary phase enabling different partition of an analyte in water and the mobile phase. In RPLC, water is a mobile-phase component and thus it cannot be both stationary and mobile phase. Consequently, adsorption seems to govern retention in this case. **Figure 3** shows an illustration of the expected adsorption mechanism of an alcohol  $R^1OH$ . The alkyl groups R on the surface of the stationary phase, being its active sites, adsorb alcohol molecules. Adsorption between R and  $R^1$  takes place by dispersion forces of interaction. The polar mobile phase has negligible adsorption on the nonpolar stationary phase. However, it can lead to analyte desorption interacting with its polar (OH) group. Such a mechanism is supported by the fact that the retention of an analyte is greater on RP-18 than on RP-8. The longer  $C_{18}$  carbon chain is expected to lead to a stronger adsorption by dispersion forces. Moreover, separation of hexanol and octanol in this case will



**Figure 3** Detailed adsorption mechanism of retention of an alcohol  $R^1OH$  in RPLC. A, adsorption by dispersion forces; D, desorption by hydrogen bonds.

lead to greater retention of octanol owing to its longer carbon chain:  $C_8H_{17}OH > C_6H_{13}OH$ . This is indeed established in RPLC. Thus, the opposite retention order in this case relative to NPLC is reasonable on the basis of the adsorption mechanism described.

In the case of ion-exchange chromatography, the mechanism involves ion exchange with a competition between the analyte X ions and mobile-phase Y ions for interaction with R ions at the surface of the stationary phase. Electrostatic attractions require interaction of ions of opposite charges. Two cases are possible: anion-exchange chromatography where X and Y are anions and R is positively charged. The opposite is true for cation-exchange chromatography:



Let us consider a concrete example: separation of acetic acid and propionic acid on a cation exchanger with  $SO_3H$  groups on its surface and water containing NaOH ( $pH > 7$ ) as a mobile phase. Mobile phase passing through the column converts the strong acidic  $SO_3H$  group in its salt form and thus  $Y^+ = Na^+$ . Acetic acid is a stronger acid with a greater ionization than that of propionic acid and this determines the better retention of acetic acid. Thus, the retention expected in this case is: acetic acid  $>$  propionic acid.

In the case of size-exclusion chromatography, retention mechanism depends on geometrical factors only and does not include any interaction between the three components of the chromatographic system. This chromatographic method is useful for analysis of high molecular mass compounds. Let us consider a simpler case. Separation of hexanol and octanol will lead to a greater retention of hexanol:  $C_6H_{13}OH > C_8H_{17}OH$ . This is due to the smaller size of hexanol and its possibility to enter the inner pores of the stationary phase particles. On the other hand, octanol moves faster with the mobile phase outside the particles.

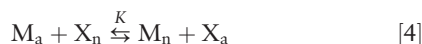
Retention in LC depends on retention mechanism and conditions used such as: stationary phase type, mobile-phase composition including its pH, flow rate, gradient applied if any, etc.

## Theory

There are many theories describing the LC system from different points of view. Analyte is characterized by its retention in such a system. Theoretical treatment of retention is closely related with the important problem for method development. Brief description of retention as a function of analyte structure or other factors is given.

### Retention versus Analyte Structure

The displacement model elaborated by Snyder in the 1960s to 1980s is well suited to explain retention in NPLC. On the basis of this model, chromatographic system is described quantitatively with various equations. According to the displacement model there is a competition between an analyte X and a mobile phase M for adsorption on the stationary-phase (adsorbent) active sites:



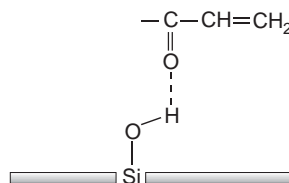
where subscriptions a and n denote adsorbed and nonadsorbed state, respectively.  $K$  is the equilibrium constant. The use of a mobile phase with a strong adsorption will leave X nonadsorbed (small retention) since equilibrium [4] is shifted to the left. An increase or decrease in retention is performed by the use of mobile phase with smaller or greater adsorption. This criterion is used in selection of suitable mobile phases for a specific separation. It allows retention to be adjusted to a favorable range ( $1 < k < 20$ ). To tune retention of the vast number of organic compounds, mobile phases, composed of two or more solvents, are usually used. For instance the retention of an alcohol is significant and its analysis cannot be done with hexane alone as a mobile phase. Mixtures of hexane and other polar solvents such as acetone are necessary. This increases the adsorption of the mobile phase and thus reduces retention of the alcohol. In such a case, acetone from the mobile phase competes with the alcohol for adsorption on the active sites.

Retention is a function of the parameters related with the adsorption properties of stationary phase, mobile phase, and analyte being characterized in the Snyder theory by various dimensionless parameters.

**Parameters characterizing the adsorption properties of stationary phase (adsorbent)** These parameters are selectivity  $\alpha'$  and a complex parameter  $\log (V_a W / V^o)$ .  $V_a$ ,  $V^o$ , and  $W$  are the volume of an adsorbed monolayer of mobile phase per gram adsorbent, volume of the pores surrounding the adsorbent particles (the so-called void volume), and mass of

the adsorbent in the chromatographic system, respectively. The complex parameter is denoted as  $R_{M(\text{shift})}$  since the retention of an analyte is changed or shifted when different adsorbents and the same mobile phase are used. Parameter  $\alpha'$  is related to the number of active sites per unit adsorbent area and parameter  $R_{M(\text{shift})}$  is related to the surface area of the adsorbent.

**Parameters characterizing the adsorption properties of mobile phase** Mobile-phase strength  $\varepsilon$  measures its dimensionless energy of adsorption ( $\Delta G^o / RT \ln 10$ ) per unit area. The greater the value of  $\varepsilon$ , the greater the adsorption of the mobile phase and this results, on the basis of eqn [4], in smaller retention. Parameters that tune the adsorption properties of mobile phase are localization  $m$  and polarity  $P'$  measuring the relative geometry under adsorption and solvating ability of the mobile phase. Localization is connected with the most favorable orientation for adsorption. It is characteristic for both mobile phase and analyte that adsorb in turn according to the displacement model. The strongest adsorbing group in the structure of a solvent or sample is localized and the remaining groups are delocalized. Localization is illustrated in Figure 4 (the carbonyl group only is localized). Mobile phases having a given value of  $\varepsilon$  and different values of  $m$  have different selectivity, i.e., different possibility to separate compounds. In general, a mobile phase having a higher value of  $P'$  dissolves the analyte better and thus decreases its retention  $k$ . The values of  $\varepsilon$ ,  $m$ , and  $P'$  of pure solvents used in LC are known. Some examples are shown in Table 1 where the elutropic series of solvents is given, i.e., the solvents are arranged according to an increase of their value of  $\varepsilon$ . The elutropic series begins with pentane and hexane having  $\varepsilon = 0$  and ends with methanol ( $\varepsilon = 0.70$ ). Table 1 shows also the relevant values of  $m$  and  $P'$ . Calculation of  $\varepsilon$  and  $m$  of mobile phases composed of two or more solvent requires a lot of calculations. Details are not included in this article. The values of  $P'$  are easily found on the basis of percent quantity,  $P_i$ , and individual polarity  $P'_i$  of any solvent  $i$  participating in the mobile-phase composition.



**Figure 4** Illustration of localization,  $m$ . The carbonyl group has a greater energy of adsorption and is localized while the double bond is delocalized.



**Table 1** Elutropic series of some pure solvents used as mobile phases in liquid chromatography and the relevant values of  $\epsilon$ ,  $m$ , and  $P'$ 

No.	Solvent	$\epsilon$	$m$	$P'$
1	Pentane	0.00	—	0.00
2	Hexane	0.00	—	0.10
3	Heptane	0.00	—	0.20
4	Tetrachloromethane	0.11	−0.35	1.60
5	Toluene	0.22	−0.43	2.40
6	Benzene	0.25	−0.42	2.70
7	Chloroform	0.26	0.10	4.10
8	Dichloromethane	0.30	0.10	3.10
9	Diethyl ether	0.43	0.66	2.80
10	Tetrahydrofuran	0.48	1.00	4.00
11	Ethyl acetate	0.48	0.60	4.40
12	Methyl- <i>tert.</i> butyl ether	0.48	—	2.70
13	Acetonitrile	0.52	1.50	5.80
14	Acetone	0.53	0.95	5.10
15	Isopropanol	0.60	—	3.90
16	Ethanol	0.65	1.05	4.30
17	Methanol	0.70	—	5.10

**Parameters characterizing the adsorption properties of the analyte** These parameters are energy of adsorption  $S_X$  and area under adsorption  $A_X$  of an analyte X. They are additive parameters and are calculated by summation of the relevant contributions of the individual structural fragments or functional groups  $i$  available in the analyte structure. A group  $i$  is characterized by its energy of adsorption  $Q_i^0$  and area under adsorption  $a_i$ :

$$S_X = \sum_i Q_i^0 \quad [5]$$

$$A_X = \sum_i a_i \quad [6]$$

There are literature data about  $Q_i^0$  and  $a_i$  of the usual groups  $i$  participating in the structure of organic compounds. The values differ if the group is connected with an alkyl or aryl group. Table 2 shows some examples. On the basis of these data,  $S_X$  and  $A_X$  of a specific compound become known. The greater the value of  $S_X$ , the stronger the analyte adsorption.

Parameters  $k$ ,  $\alpha'$ ,  $R_{M(\text{shift})}$ ,  $S_X$ ,  $A_X$ , and  $\epsilon$  are related by the Snyder theory in an equation having two forms:

$$R_M = \log k = R_{M(\text{shift})} + \alpha'(S_X - \epsilon A_X) \quad [7]$$

and

$$R_M = \log k = R_{M(\text{shift})} + \alpha' \left( \sum_i Q_i^0 - \epsilon \sum_i a_i \right) \quad [8]$$

Equation [7] is derived on the basis of eqn [4]. Its equilibrium constant  $K$  is expressed as a function of

**Table 2** Adsorption properties of some analyte structural elements or functional groups  $i$  as measured by the relevant values of  $Q_i^0$  and  $a_i$ 

No.	Group $i$	$Q_i^0$	$a_i$
1	Al = Alkyl		
2	Al-CH <sub>3</sub>	0.07	1.6
3	Al-CH <sub>2</sub> -Al	−0.05	0.9
4	Al <sub>2</sub> -C = (olefinic carbon)	0.25	1.0
5	Al-Cl	1.74	1.2
6	Al-Br	1.94	1.8
7	Al-I	1.94	2.1
8	Al-OCH <sub>3</sub>	3.61	9.0
9	Al-CHO	4.97	9.2
10	Al-COCH <sub>3</sub>	5.27	9.8
11	Al-CN	5.27	8.7
12	Al-CO <sub>2</sub> CH <sub>3</sub>	5.27	10.5
13	Al-OH	5.60	8.5
14	Al <sub>3</sub> -N	5.80	10.5
15	Al-NH <sub>2</sub>	8.00	8.7
16	Al-CONH <sub>2</sub>	9.60	10.3
Ar = Aryl			
16	Ar-Br	−0.17	1.0
17	Ar-CN	3.33	8.4
18	Ar-CHO	3.48	8.3
19	Ar-OH	4.20	7.6
20	Ar-NH <sub>2</sub>	5.10	8.7

(1) molar fractions of the four components of eqn [4], (2) free energy of adsorption by Gibbs equation, and (3) any one of the chromatographic parameters  $K_D$  and  $k$ . Equations [7] and [8] are used to determine one or two of their parameters when the other parameters are known.

In the Snyder theory, eqn [8] is the simplest equation that gives the relationship between retention and analyte structure. It is used in an approach for selection of suitable mobile phases for a given separation. The adsorbent, that will be used, is to have known values of  $\alpha'$  and  $R_{M(\text{shift})}$ . Otherwise, these values can be easily found. The relevant structural elements or functional groups (see Table 2) are used to express the structure of an analyte and find its values of  $S_X$  and  $A_X$  by eqns [5] and [6], respectively. The retention of an analyte on the selected adsorbent is calculated by eqn [8] when successive values, in the range of 0–0.70, are ascribed to  $\epsilon$ . An analysis of the retention as a function of  $\epsilon$  is performed for all compounds that will be separated. It allows predicting the optimum value of  $\epsilon$  for the concrete separation. Specific mobile phases having that value of  $\epsilon$  are found.

This approach is applicable if the analyte is a non-ionic organic compound and its structure is known. It is already successfully applied by means of a software (LSChrom) to normal-phase liquid chromatography (NRLC) of compounds with relatively simple or complex structure.



## Retention versus Other Factors

Retention indirectly reflects analyte structure. Consequently, it can be related to other parameters not connected directly to structure. There are various relationships in this respect. Attention will be focused on some relationships of this type where retention is a function of mobile-phase composition, keeping stationary phase and analyte the same. For instance, eqns [9] and [10] express the relationships between retention and mobile-phase polarity  $P'$  in the case of NPLC and RPLC, respectively:

$$\frac{k_2}{k_1} = 10^{(P'_1 - P'_2)/2} \quad [9]$$

$$\frac{k_2}{k_1} = 10^{(P'_2 - P'_1)/2} \quad [10]$$

when an analyte is analyzed with mobile phases 1 and 2 with different  $P'$  values. The subscripts to  $k$  and  $P'$  correspond to the mobile phases used. It is worth noting that mobile-phase strength in NPLC is better expressed by  $\varepsilon$ .

These and similar relationships are useful in selection of an optimum mobile phase necessary to separate complex mixtures of compounds. In general, the strategy used includes: (1) preliminary analysis with some concrete mobile phases on a given stationary phase, (2) application of a specific numeric method for analysis of the data collected using appropriate software, (3) selection of the optimum mobile-phase composition ensuring retention in the favorable range ( $1 < k < 20$ ) and maximum resolution  $R$  of all compound pairs, and (4) performance of an analysis with the optimum mobile phase selected. It is clear that this strategy requires preliminary experiments contrary to the strategy described when the analyte structure is directly taken into account. The strategy discussed is applicable both to any kind of LC since the precise mechanism of retention is indirectly taken into account by the mathematical modeling of the chromatographic system (point 2). It is an application of chemometrics in LC. The outlined strategy is used for separation of a complex mixture of compounds with the optimum mobile phase. The latter is selected on the basis of seven preliminary experiments with mobile phases with different composition. The mobile phases are composed of three to four solvents from different groups and properties. Mobile-phase properties are taken into account by localization,  $m$ . Numeric method used is overlap resolution map: the optimum mobile phase avoids overlap of the peaks of all pairs of compounds.

The strategy outlined has some modifications. It is developed by including other factors such as flow

rate, bed length, mobile-phase pH, design of the gradient change of mobile-phase composition, etc. The DRYLAB software is of importance in this respect. Such use of the multifactorial analysis is applicable to select suitable conditions in all liquid chromatographic methods: NPLC and RPLC, ion-exchange chromatography, and size-exclusion chromatography.

## Troubleshooting

The HPLC chromatograph is a complex instrument requiring maintenance of all its parts. Air bubbles in the mobile phase seen in the inlet tubings and leaking fittings prevent the proper operation of the chromatograph and give unreliable data. Mechanical particles in the analyte solution can damage both the syringe for its injection and column. Thus precautions in this respect are necessary. The quantity of the sample injected and the operating wavelength of the UV detector, if used, are to adjust properly. A high quality column is important. Taking these factors into account makes the analyses easier. Any specific problem that arises requires a specific response for its elimination. This is known as troubleshooting.

## Recommendations

A new separation by LC includes several steps. An important step is selection of suitable LC method and technique. The following recommendations are useful.

1. Try NPLC for separation of nonionic organic compounds which are dissolved well in mixtures of organic solvents. Apply it for separation of isomers and preparative separations. Adjust mobile phase pH if analytes are bases ( $\text{pH} > 7$ ) or acids ( $\text{pH} < 7$ ), thus avoiding chemisorption and tailing.
2. Try RPLC for separation of all kinds of nonionic organic compounds.
3. Try ion-pair LC for ionic or ionizable organic and inorganic compounds especially bases.
4. Try ion-exchange LC for separation of mixtures of inorganic ions, proteins, and related compounds.
5. Try size-exclusion LC for separation of high molecular mass samples (proteins, synthetic polymers, etc.).

Adjustment of pH is recommended in items 2, 3, and 4, if necessary.

If the first attempted LC method does not give good results, try another. Keep in mind the efficiency of the chromatographic system used. Adjust its selectivity with the aid of mobile-phase composition, number of theoretical plates, and flow rate (by means of Van Deemter equation).

Full method development is necessary in case of difficult separation with peak overlap and routine analyses.

The available software products in LC will assist one in dealing with chromatograms and selection of suitable experimental conditions.

Patience is required since any separation needs more or less time to settle the proper conditions. Going into depth of LC theory will assist one. However, this can be omitted if one has no time for it. Instead one should rely on the software products where relevant theoretical aspects are incorporated.

*See also:* **Chromatography:** Principles. **Liquid Chromatography:** Overview; Column Technology; Instrumentation; Biotechnology Applications; Clinical Applications; Food Applications; Pharmaceutical Applications.

## Further Reading

Dolan JW and Snyder LR (1989) *Troubleshooting HPLC Systems*. Clifton, NJ: Humana Press.

LSChrom (1999) Ver. 2.1 <http://www.members.tripod.com/LSChrom>.

Meyer VR (1999) *Practical High-performance Liquid Chromatography*, 3rd edn. Chichester: Wiley.

Neue UD (1997) *HPLC Columns: Theory, Technology, and Practice*. New York: Wiley-VCH.

Poole CF (2002) *The Essence of Chromatography*. Amsterdam: Elsevier Health Sciences.

Sadek PC (2000) *Troubleshooting HPLC Systems*. New York: Wiley & Sons.

Snyder LR (1983) Mobile-phase effects in liquid-solid chromatography. In: Horvath C (ed.) *High-Performance Liquid Chromatography*, vol. 3 (p, p. 157. New York: Academic Press.

Snyder LR and Kirkland JJ (1979) *Introduction to Modern Liquid Chromatography*, 2nd edn. New York: Wiley-Interscience.

Snyder LR, Glajch JL, and Kirkland JJ (1994) *Practical HPLC Method Development*, 2nd edn. New York: Wiley. (DRAYLAB applications included).

Snyder LR, Barker J, Glajch JL, Ando DJ, and Kirkland JJ (2002) *Practical HPLC Method Development and Mass Spectrometry*, 2nd edn. New York: Wiley.

## Column Technology

**U D Neue**, Waters Corporation, Milford, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

High-performance liquid chromatography (HPLC) has become one of the more important tools in the chemical analysis of mixtures, especially in the pharmaceutical industry. This would not have become possible without the workhorse of HPLC, the column. In the following, we will describe the characteristic properties of HPLC columns, starting with column physics, followed by column chemistry. In the column physics section, we discuss the choices in column length and column diameter, together with particle size, pore size, and particle porosity. In the discussion of column chemistry, we begin with a description of the base particles, followed by the different options available for surface chemistry. In the last section, we identify the chromatographic selectivity arising from this surface chemistry for packings used in reversed-phase chromatography. For a deeper discussion, the reader is referred to textbooks on column technology or chromatography.

## Column Physics

Most high-performance packings are based on silica or related hard particles, such as inorganic-organic hybrid packings. Therefore, the focus of the following discussion is the performance of incompressible packings.

The correct selection of column length and particle size is important for the separation power as well as the speed of analysis that is achievable with a given column. The compromise between speed and resolving power is shown in **Figures 1A** and **1B** for columns packed with 5  $\mu\text{m}$  and 3  $\mu\text{m}$  particles respectively. The achievable resolving power (= the square root of the column plate count) is plotted as a function of the analysis time. We implicitly assume an isocratic analysis to a retention factor of 10. Longer columns provide a higher resolving power, but cannot provide short analysis times due to the pressure limitations of the HPLC instrument. This is especially important for the smaller particles. A 25 cm column packed with 3  $\mu\text{m}$  particles can deliver much resolving power, but at the expense of a high pressure drop. The pressure limit constrains the flow-rate range over which the column can be used. For faster analyses, on the other hand, it is best to use shorter columns. For example, the performance of a

Full method development is necessary in case of difficult separation with peak overlap and routine analyses.

The available software products in LC will assist one in dealing with chromatograms and selection of suitable experimental conditions.

Patience is required since any separation needs more or less time to settle the proper conditions. Going into depth of LC theory will assist one. However, this can be omitted if one has no time for it. Instead one should rely on the software products where relevant theoretical aspects are incorporated.

*See also:* **Chromatography:** Principles. **Liquid Chromatography:** Overview; Column Technology; Instrumentation; Biotechnology Applications; Clinical Applications; Food Applications; Pharmaceutical Applications.

## Further Reading

Dolan JW and Snyder LR (1989) *Troubleshooting HPLC Systems*. Clifton, NJ: Humana Press.

LSChrom (1999) Ver. 2.1 <http://www.members.tripod.com/LSChrom>.

Meyer VR (1999) *Practical High-performance Liquid Chromatography*, 3rd edn. Chichester: Wiley.

Neue UD (1997) *HPLC Columns: Theory, Technology, and Practice*. New York: Wiley-VCH.

Poole CF (2002) *The Essence of Chromatography*. Amsterdam: Elsevier Health Sciences.

Sadek PC (2000) *Troubleshooting HPLC Systems*. New York: Wiley & Sons.

Snyder LR (1983) Mobile-phase effects in liquid-solid chromatography. In: Horvath C (ed.) *High-Performance Liquid Chromatography*, vol. 3 (p, p. 157. New York: Academic Press.

Snyder LR and Kirkland JJ (1979) *Introduction to Modern Liquid Chromatography*, 2nd edn. New York: Wiley-Interscience.

Snyder LR, Glajch JL, and Kirkland JJ (1994) *Practical HPLC Method Development*, 2nd edn. New York: Wiley. (DRAYLAB applications included).

Snyder LR, Barker J, Glajch JL, Ando DJ, and Kirkland JJ (2002) *Practical HPLC Method Development and Mass Spectrometry*, 2nd edn. New York: Wiley.

## Column Technology

**U D Neue**, Waters Corporation, Milford, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

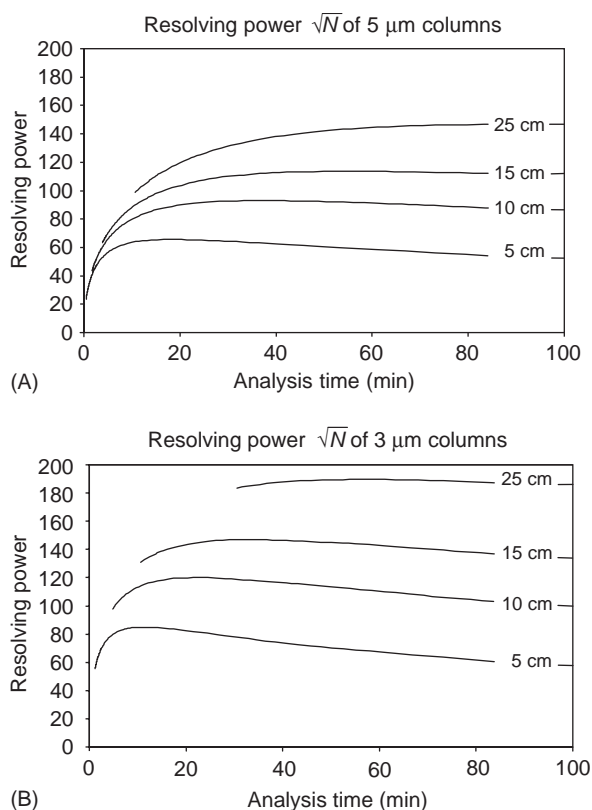
## Introduction

High-performance liquid chromatography (HPLC) has become one of the more important tools in the chemical analysis of mixtures, especially in the pharmaceutical industry. This would not have become possible without the workhorse of HPLC, the column. In the following, we will describe the characteristic properties of HPLC columns, starting with column physics, followed by column chemistry. In the column physics section, we discuss the choices in column length and column diameter, together with particle size, pore size, and particle porosity. In the discussion of column chemistry, we begin with a description of the base particles, followed by the different options available for surface chemistry. In the last section, we identify the chromatographic selectivity arising from this surface chemistry for packings used in reversed-phase chromatography. For a deeper discussion, the reader is referred to textbooks on column technology or chromatography.

## Column Physics

Most high-performance packings are based on silica or related hard particles, such as inorganic-organic hybrid packings. Therefore, the focus of the following discussion is the performance of incompressible packings.

The correct selection of column length and particle size is important for the separation power as well as the speed of analysis that is achievable with a given column. The compromise between speed and resolving power is shown in **Figures 1A** and **1B** for columns packed with 5  $\mu\text{m}$  and 3  $\mu\text{m}$  particles respectively. The achievable resolving power (= the square root of the column plate count) is plotted as a function of the analysis time. We implicitly assume an isocratic analysis to a retention factor of 10. Longer columns provide a higher resolving power, but cannot provide short analysis times due to the pressure limitations of the HPLC instrument. This is especially important for the smaller particles. A 25 cm column packed with 3  $\mu\text{m}$  particles can deliver much resolving power, but at the expense of a high pressure drop. The pressure limit constrains the flow-rate range over which the column can be used. For faster analyses, on the other hand, it is best to use shorter columns. For example, the performance of a



**Figure 1** Resolving power ( $= \sqrt{N}$ ) of various columns as a function of analysis time. It is assumed that the analysis is completed at  $k=10$ . Performance of (A) 5  $\mu\text{m}$  columns and (B) 3  $\mu\text{m}$  columns. The column lengths are 25, 15, 10, and 5 cm.

5 cm 5  $\mu\text{m}$  column matches that of a 10 cm 5  $\mu\text{m}$  column for a 2 min analysis, but at a lower pressure drop. The analysis time can be reduced to less than half a minute if only a performance of 600 plates is needed.

An interesting aspect of column design is the fact that the same maximum plate count is achievable when the column length and the particle size are changed in the same proportion. A 10 cm 10  $\mu\text{m}$  column, a 5 cm 5  $\mu\text{m}$  column, and a 3 cm 3  $\mu\text{m}$  column all achieve the same maximum resolving power,  $\sim 4500$  plates. However, the analysis time at which this performance can be reached decreases with the square of the particle size. The same resolution that can be achieved in  $\sim 75$  min with the 10  $\mu\text{m}$  column can be reached in under 7 min with the 3  $\mu\text{m}$  column. In addition, for a particular analysis time, the backpressures of a 10 cm 10  $\mu\text{m}$  column, a 5 cm 5  $\mu\text{m}$  column, and a 3 cm 3  $\mu\text{m}$  column are identical. This is the primary advantage of small particles.

In principle, the column diameter does not have an influence on the relationship between plate count, pressure, and analysis time. Only in very narrow columns, where the column diameter is less than 10

times the particle diameter, a loose bed with a higher permeability (= lower backpressure) is formed. If the available sample mass is limited, higher sensitivities are achieved if the available sample mass is injected onto a smaller-diameter column. However, if the sample volume is not constrained, better sensitivity can be reached with a standard column with standard detectors. This is because the signal-to-noise ratio decreases with the detector volume for concentration-sensitive detectors.

To maximize the retention, fully porous particles are used in most HPLC applications. The pore size determines the surface area that is available. Packings with a smaller pore size have a larger specific surface area, but this restricts the access for larger molecules. Therefore, packings designed for the analysis of small molecules have typically a pore size of 10 nm, while packings used in the separation of proteins have a pore size of 30 nm or even 100 nm. The increase in pore size leads to a reduction in surface area. For the separation of very large molecules, nonporous particles with a particle diameter around 1–2  $\mu\text{m}$  are advantageous.

The specific pore volume determines the strength of the particle. Packings with a pore volume  $\sim 1.5 \text{ ml g}^{-1}$  are very fragile, and those with values  $\sim 0.5 \text{ ml g}^{-1}$  are indestructible under normal HPLC conditions. Many HPLC packings have a specific pore volume around or just under  $1 \text{ ml g}^{-1}$ . The retentivity of a packing is best described by the ratio of surface area to pore volume. A packing with a specific surface area of  $350 \text{ m}^2 \text{ g}^{-1}$  and a specific pore volume of  $0.9 \text{ ml g}^{-1}$  is about as retentive as a packing with a surface area of  $190 \text{ m}^2 \text{ g}^{-1}$  and a pore volume of  $0.5 \text{ ml g}^{-1}$ . There is a close link between the specific surface area  $A_{\text{sp}}$ , the specific pore volume  $V_{\text{sp}}$ , and the pore size  $d_{\text{po}}$ :

$$\frac{A_{\text{sp}}}{V_{\text{sp}}} \approx \frac{c}{d_{\text{po}}}$$

The value of  $c$  is  $\sim 3.75$ . Thus, the pore size is the determining factor for the retentivity of a packing. The catalogs of column manufacturers supply at least two of these values.

In recent years, monolithic columns have been developed. While classical columns are packed with particles, monolithic columns are continuous structures that look like corrals. The skeleton can be porous, like particles, and contains the surface area responsible for retention. The space between the skeleton provides the flow passages. The key advantage of monolithic columns lies in the fact that the interstitial space can be manipulated independently of the domain size that is responsible for retention. Therefore, the compromise between column

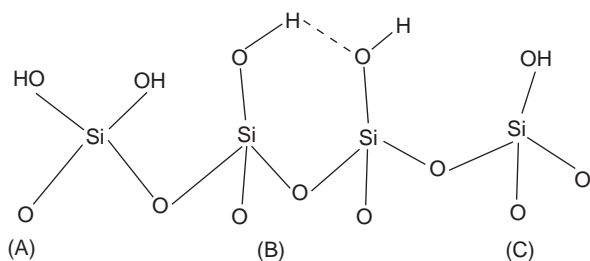


performance and backpressure is more favorable for well-designed monolithic columns. However, the mere fact that a column device is monolithic does not yet mean that a favorable performance has been achieved. The best measure for the comparison of a monolithic device to a packed bed is created if a particle size equivalent is generated from the backpressure of the device. With this measure, a well-designed monolith outperforms a packed bed by a factor of  $\sim 2$ , meaning that current monoliths achieve the same performance as 5  $\mu\text{m}$  columns at a backpressure typical of 10  $\mu\text{m}$  columns.

## Column Chemistry

### The Base Packing

HPLC packings may be based on several different raw materials. Silica, zirconia, titania, and alumina are the inorganic carriers that are used. Silica hybrids are also predominantly inorganic at the time of this writing. Organic polymeric packings, e.g., packings based on styrene–divinylbenzene or methacrylate, are used mainly for size-exclusion chromatography or for the analysis of biomolecules such as proteins. Porous graphitic carbon is used for some specific applications in reversed-phase chromatography. In contrast to inorganic packings, polymeric packings deform under pressure. Therefore, only lower pressures can be applied. The pressure limit depends on the details of the design of the packing.



**Figure 2** Silanol groups on the surface of silica: (A) geminal; (B) vicinal; and (C) lone silanols. Silicon atoms are connected to each other via siloxane bridges.

Silica can be used in the acidic pH range up to pH 8. Zirconia and styrene/divinylbenzene-based packings can be exposed to any pH without damage. Methacrylate packings are stable from pH 2 to pH 12. The pH stability of hybrid organic–inorganic packings depends on the design of the packing, and extends from the acidic pH to pH 12. The pH stability of alumina extends to pH 13.

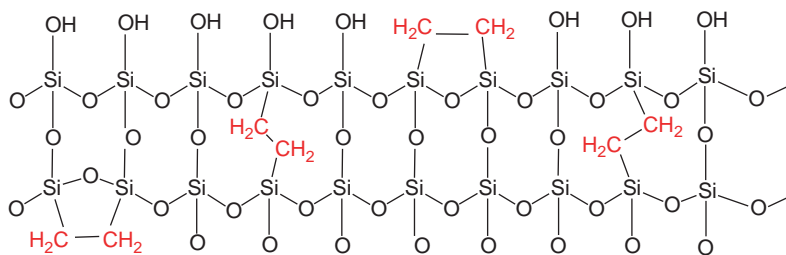
Silica and silica hybrids are the bases for the most popular HPLC packings. The surface of silica is covered with silanol groups. These are subdivided into geminal, vicinal or bridged, and lone silanols, as shown in **Figures 2A–2C** respectively.

The concentration of surface silanols depends on the treatment of the silica. It is usually assumed that  $\sim 8 \mu\text{mol m}^{-2}$  cover the surface of a fully hydroxylated silica.

In silica hybrid packings, some of the siloxane bridges are substituted by methyl groups or by ethylene bridges, as shown in **Figure 3**.

### The Stationary Phase

While silica or silica hybrid packings are active sorbents on their own, their surfaces are commonly derivatized with a silanization reagent. The chromatographic properties of these bonded phases are determined by the reagent as well as the derivatization technique. A reaction with monofunctional, difunctional, and trifunctional silanes is possible (**Figure 4**). The reagent, which can be either an alkoxy silane or a chloro silane, is shown on the left, and a cartoon of the resulting surface derivatization is found on the right. The surface of the silica itself is indicated with a horizontal bar. With trifunctional reagents (**Figure 4A**), additional silanols are produced during the bonding step, because the reagent reacts with the surface only via two surface silanols. Difunctional reagents (**Figure 4B**) react with the surface silanols predominantly in a two-bite reaction, but additional silanols are created for the fraction of the reagent that reacts only with one bite. With a monofunctional silane (**Figure 4C**), only  $\sim 4 \mu\text{mol m}^{-2}$  or less

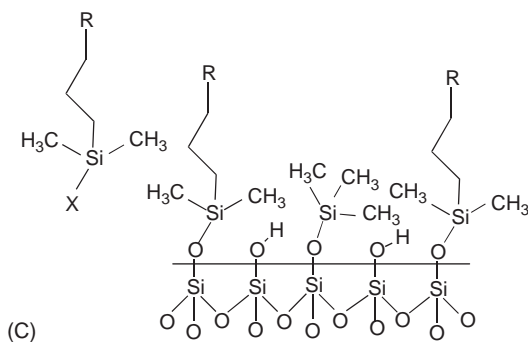
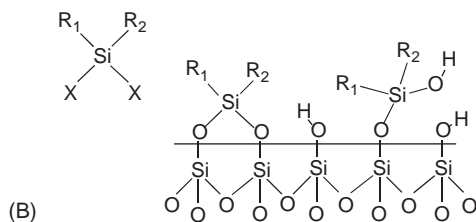


**Figure 3** Structure of an ethyl-bridged silica hybrid packing. The ethyl groups reduce the number of surface silanols and provide chemically stable bridges between the different layers of the backbone of the packing.



With difunctional reagents (**Figure 4B**), both ligands  $R_1$  and  $R_2$  can be the same. However, for most bonded phases based on a difunctional silane, the second ligand is commonly a methyl group. In **Figure 4C**, the methyl group is shown as the common side chain for monofunctional bonded phases. Alternatively, isopropyl or isobutyl groups provide an increase in the hydrolytic stability of the packing under acidic conditions. At the same time, the bulky side groups limit the surface coverage that can be achieved, which results in a larger silanol activity

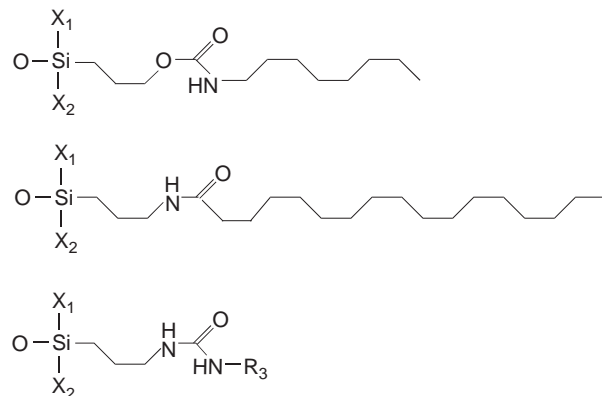
(A) Chemical structure of a siloxane network. The network consists of a horizontal chain of silicon (Si) atoms connected by oxygen (O) atoms. Each silicon atom in the chain is also bonded to two oxygen atoms, which are further bonded to a second layer of silicon atoms. The second layer of silicon atoms is also connected by oxygen atoms. The first layer of silicon atoms has pendant groups: the first silicon atom has two 'X' substituents, and the other three silicon atoms have one 'R' group and one 'H' atom. The second layer of silicon atoms has pendant groups: the first silicon atom has one 'R' group and one 'H' atom, and the other three silicon atoms have one 'R' group and one 'H' atom.



A range of bonded phases are commercially available. They can be classified into bonded phases for reversed-phase chromatography and bonded phases with specific ligands that can be used for other types of chromatography. A selection of ligands used in reversed-phase chromatography is shown in **Figures 5 and 6**. As above, the nature of  $X_1$  and  $X_2$  varies, depending on whether monofunctional, difunctional, or trifunctional ligands are used. The classical

Chemical structures of various siloxane compounds:

- Linear siloxane with a long alkyl chain (10 carbons).
- Linear siloxane with a medium alkyl chain (6 carbons).
- Linear siloxane with a short alkyl chain (4 carbons).
- Trimethylsiloxane.
- Phenylsiloxane with a 3-phenylpropyl chain.



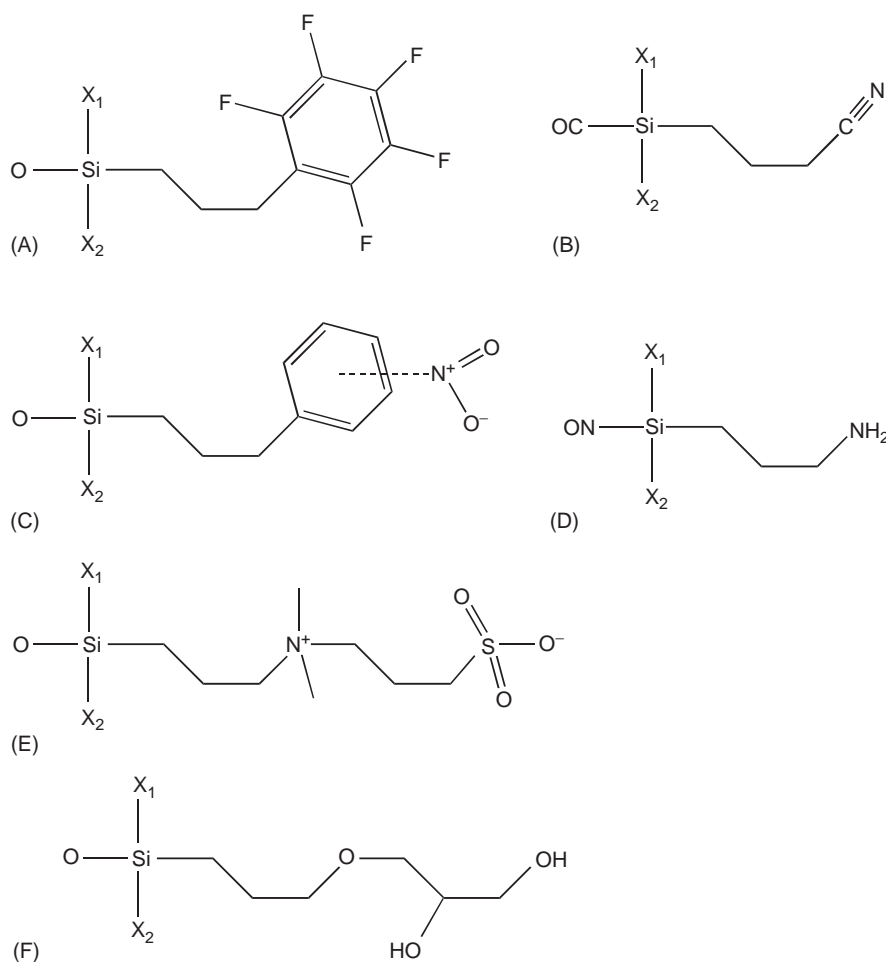
ligands shown in Figure 5 range from the popular  $C_{18}$  phases to shorter chains such as  $C_8$ ,  $C_4$ , or  $C_1$  phases. The latter two are most frequently applied for the separation of proteins. Commercial phenyl phases have different alkyl chains between the phenyl groups and the silicon atom. Only one possible structure is shown. Due to a unique steric selectivity, phases with a  $C_{30}$  ligand (not shown) are used for the separation of carotenoids.

The reversed-phase ligands in Figure 6 contain an embedded polar group, which reduces substantially the interaction of analytes with residual silanols on the surface and thus improves the peak shapes. Some of the packings with an embedded polar group are synthesized in two steps. The first step is a surface derivatization with a propylaminosilane. In the second step, the reversed-phase ligand is attached to the amino group via an amidation reaction. Due to steric hindrance, surface reactions are never complete. Consequently, residual  $-NH_2$  groups remain on the surface of those packings that are synthesized in two

steps. This can be avoided if the ligand is assembled first and reacted onto the silica surface in a single step.

Occasionally, one would like to separate rather polar compounds with a reversed-phase technique. This requires packings that can function well in a 100% aqueous mobile phase. This is not the case for highly covered and fully endcapped  $C_{18}$  phases. The unfavorable wetting angle of these packings toward water drives the mobile phase out of the pores and causes loss of retention. This can be manipulated by a judicious choice of the coating level (Atlantis d $C_{18}$ ), the incorporation of polar groups into the ligand (e.g., SymmetryShield RP $_{18}$ ), polar endcapping, or no endcapping (e.g., Resolve  $C_{18}$ ).

Ligands with polar functional groups and fluorinated phases are shown in Figure 7. Pentafluorophenyl packings and perfluorinated alkyl packings are utilized in reversed-phase chromatography. Cyano and nitrophenyl packings are useful both in normal-phase and in reversed-phase HPLC. The primary use



**Figure 7** Packings with other functional groups: (A) pentafluorophenyl, (B) cyanopropyl, (C) nitrophenyl, (D) aminopropyl, (E) zwitterionic phase, and (F) diol packing.

of aminopropyl packings is in the analysis of carbohydrates in hydrophilic interaction chromatography (HILIC). HILIC is an extension of normal-phase chromatography to mobile phases containing 10–50% water. This technique requires very polar stationary phases. The zwitterionic stationary phase shown in **Figure 7E** is also predominantly employed in HILIC separations. Diol packings can be used for both aqueous size-exclusion chromatography as well as for normal-phase separations.

The hydrolytic stability of silica-based packings depends on the nature of the ligand and the pH of the mobile phase. At acidic pH, the ligand is cleaved off the surface. Under alkaline conditions, the stability depends primarily on the nature of the matrix. Above pH 8, silica dissolves. Methyl hybrids or ethyl-bridged hybrids can be used up to pH 12. At acidic pH, packings based on trifunctional silanes have a better stability than packings based on monofunctional silanes. The best lifetime under acidic conditions is obtained with packings with a bulky side group. A longer chain length, i.e., a C<sub>18</sub> packing, is more stable than a shorter chain length, e.g., a CN packing. Most manufacturers recommend a pH range from 2 to 8 for standard silica-based packings.

## Selectivity of Reversed-Phase Packing

Several hundred reversed-phase materials are commercially available. The majority of them are C<sub>18</sub>-type packings. However, they are by no means interchangeable with each other. The reasons for this are differences in the coating levels, the purity of the base silica, the efficiency of the endcapping procedures, the ligands used, and the nature of the

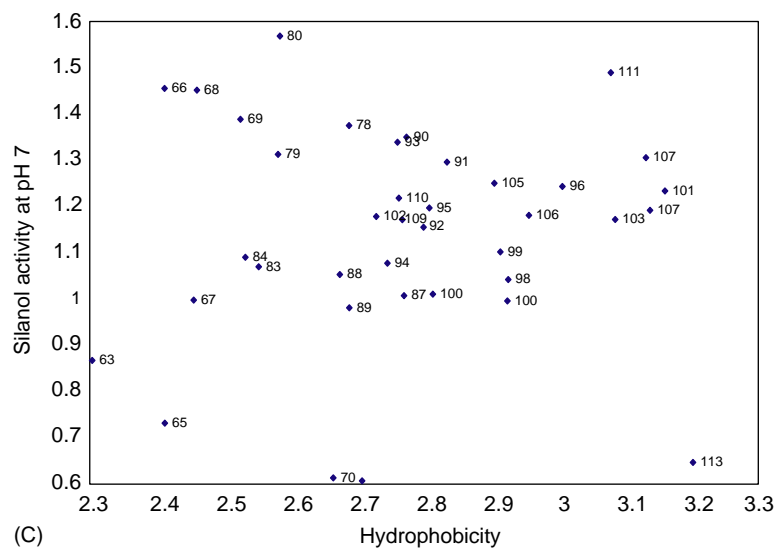
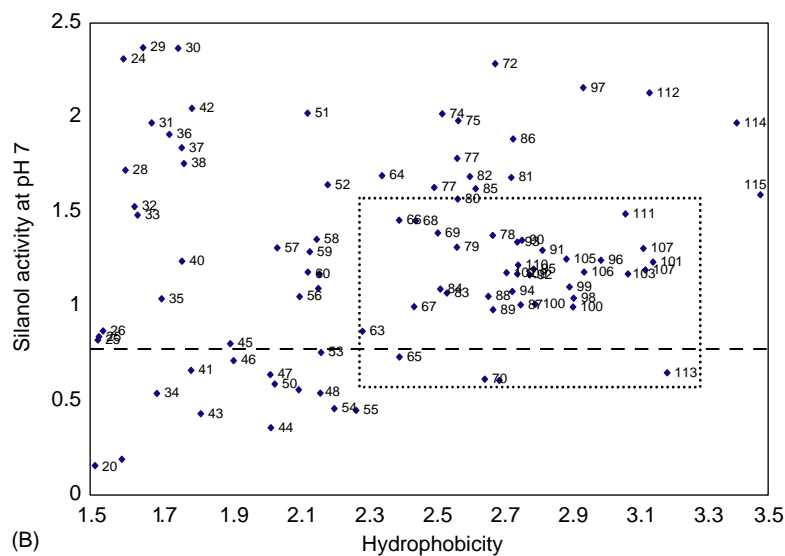
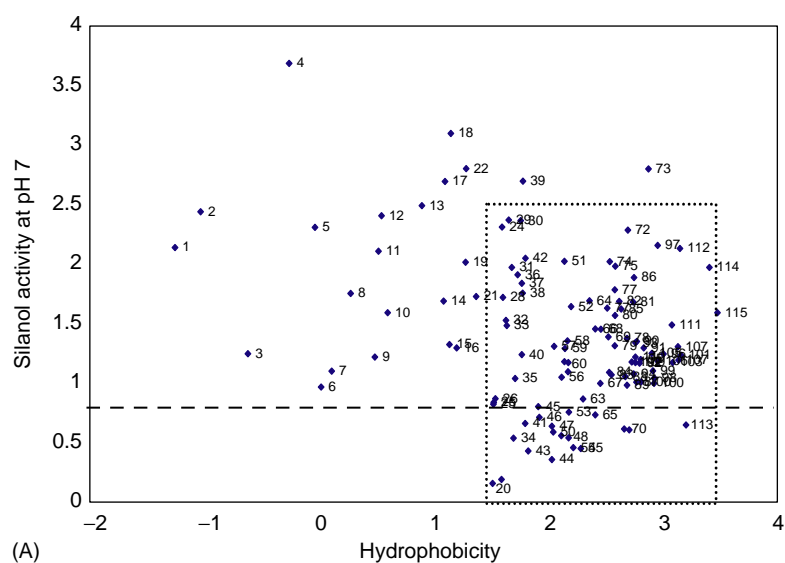
endcapping reagents. This results in a broad range of retention and selectivity properties, as well as differences in the hydrolytic stability and reproducibility of these packings.

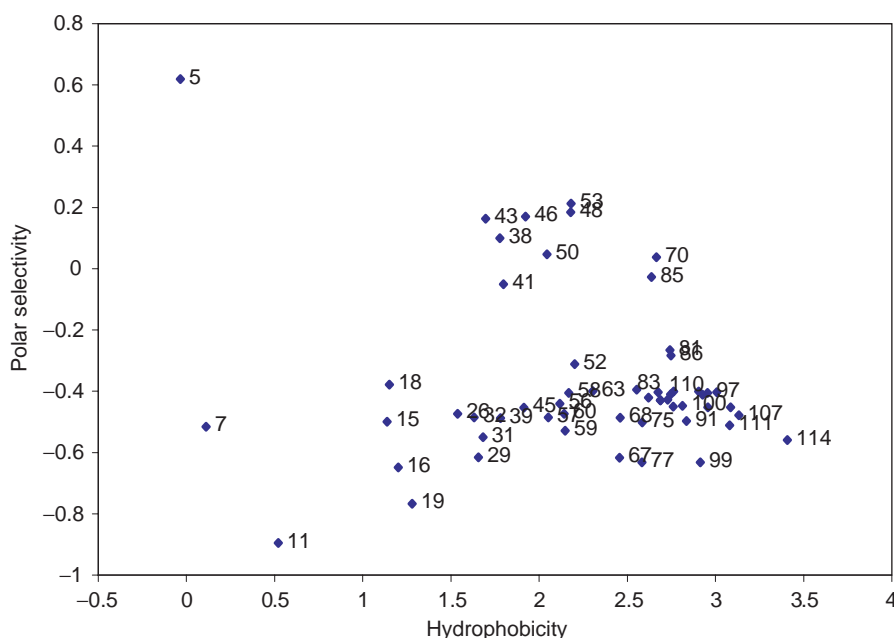
For the user of HPLC columns, the most important features are the chromatographic selectivities of the different packings. In the following, we will categorize reversed-phase packings using measures of their hydrophobicity, silanophilic interaction at pH 7, hydrogen-bond formation, and their steric selectivity.

The hydrophobic retentivity of a packing can simply be measured using the retention factor of a neutral hydrocarbon such as acenaphthene. The silanophilic interaction can be derived from the relative retention between amitriptyline, a hydrophobic base, and acenaphthene. From both parameters, a map of two of the fundamental properties of an HPLC packing can be constructed. A chart of these two primary properties of reversed-phase packings is shown in **Figure 8** for over 100 different commercially available columns. The x-axis represents the hydrophobicity of the packings, and the y-axis is a measure of the silanophilic interaction at pH 7. This parameter may also be viewed as a measure of the ion-exchange properties due to residual silanols. The right-hand side of the graphs contains packings with a higher hydrophobicity, and the left-hand side contains packings with a decreased hydrophobicity. In **Figure 8A**, the left-hand side is occupied by cyano and phenyl packings. **Figure 8B** is an expansion of **Figure 8A** and contains predominantly C<sub>8</sub> packings and C<sub>18</sub> packings with a low surface coverage on the left side, and fully coated C<sub>18</sub> packings on the right. The final expansion, **Figure 8C**, comprises, for the most part, high-purity C<sub>18</sub> packings.

The vertical axis is a measure of the deactivation of surface silanols. Packings with a high silanol

**Figure 8** (see overleaf) Hydrophobicity (x-axis) and silanophilic interaction at pH 7 (y-axis) for 106 commercially available packings. The insets indicate the scales of the subsequent figure. 1 Nova-Pak CN HP, 2 Waters Spherisorb CN RP, 3 Hypersil CPS CN, 4 Waters Spherisorb Phenyl, 5 Keystone Fluofix 120N, 6 YMC-Pack CN, 7 Ultra PFP, 8 Zorbax SB-CN, 9 Hypersil BDS Phenyl, 10 Inertsil 3 CN, 11 Fluophase RP, 12 Hypersil Phenyl, 13 Zorbax SB-Aq, 14 YMC-Pack Ph, 15 YMC Basic, 16 Ultra Phenyl, 17 Inertsil Ph3, 18 Platinum EPS C<sub>18</sub>, 19 Synergi Polar-RP, 20 XTerra RP<sub>8</sub>, 21 Nova-Pak Phenyl, 22 Zorbax SB-Phenyl, 24 Zorbax Rx C<sub>8</sub>, 25 XTerra MS C<sub>8</sub>, 26 Prodigy C<sub>8</sub>, 28 Zorbax Eclipse XBD Phenyl, 29 Zorbax SB C<sub>8</sub>, 30  $\mu$ Bondapak C<sub>18</sub>, 31 YMC J'Sphere L80, 32 Supelcosil LC DB-C<sub>8</sub>, 33 ZirChrom PBD, 34 Discovery RP Amide C<sub>16</sub>, 35 Hypersil BDS C<sub>8</sub>, 36 HydroBond AQ, 37 Lichrospher Select B, 38 Allure Ultra IBD, 39 Platinum C<sub>18</sub>, 40 Nova-Pak C<sub>8</sub>, 41 Capcell Pak C<sub>18</sub>, 42 Alltima C<sub>8</sub>, 43 Discovery RP Amide C<sub>16</sub>, 44 XTerra RP<sub>18</sub>, 45 Symmetry300 C<sub>18</sub>, 46 Spectrum, 47 Zorbax Bonus RP, 48 Supelcosil LC-ABZ Plus, 50 SymmetryShield RP<sub>8</sub>, 51 Lichrosorb Select B, 52 Polyencap A, 53 Prism, 54 Supelcosil LC-ABZ+, 55 Supelcosil LC-ABZ, 56 Luna C<sub>8</sub>(2), 57 Inertsil C<sub>8</sub>, 58 Kromasil C<sub>8</sub>, 59 Zorbax Eclipse XDB C<sub>8</sub>, 60 Symmetry C<sub>8</sub>, 63 Hypersil HyPurity Elite C<sub>18</sub>, 64 Hypersil ODS, 65 Polaris C<sub>18</sub>-A, 66 Luna Phenyl-Hexyl, 67 Hypersil BDS C<sub>18</sub>, 68 Supelcosil LC DB-C<sub>18</sub>, 69 Aqua C<sub>18</sub>, 70 SymmetryShield RP<sub>18</sub>, 72 Nucleosil C<sub>18</sub>, 73 Waters Spherisorb ODS-2, 74 Waters Spherisorb ODSB, 75 YMC J'Sphere M80, 77 Zorbax SB-C<sub>18</sub>, 78 Synergi Max RP, 79 YMC Hydrosphere C<sub>18</sub>, 80 Nova-Pak C<sub>18</sub>, 81 Poly-Encap C<sub>18</sub>, 82 TSK-Gel 80Ts, 83 Ace C<sub>18</sub>, 84 XTerra MS C<sub>18</sub>, 85 Fluophase PFP, 86 Purospher RP<sub>18</sub>, 87 Develosil C30 UG 5, 88 Develosil ODS UG 5, 89 Hypersil Elite C<sub>18</sub>, 90 Zorbax Rx C<sub>18</sub>, 91 Zorbax Eclipse XDB C<sub>18</sub>, 92 L-Column ODS, 93 YMC ODS AQ, 94 Prodigy C<sub>18</sub>, 95 Luna C<sub>18</sub>(2), 96 Kromasil C<sub>18</sub>, 97 Allure PFP Propyl, 98 Discovery HS C<sub>18</sub>, 99 Inertsil ODS-2, 100 Symmetry C<sub>18</sub>, 101 L-column ODS, 102 Puresil C<sub>18</sub>, 103 Cadenza CD-C<sub>18</sub>, 105 Luna C<sub>18</sub>, 106 Zorbax Extend C<sub>18</sub>, 107 Inertsil ODS-3, 109 Zorbax Eclipse XDB C<sub>18</sub>, 110 YMC Pack Pro C<sub>18</sub>, 111 Purospher RP<sub>18</sub>, 112 Alltima C<sub>18</sub>, 113 ODPerfect, 114 YMC J'Sphere H80, 115 Develosil ODS SR 5.





**Figure 9** Plot of polar selectivity versus hydrophobicity. Designations as in **Figure 8**.

activity are at the top, and those with a low activity are shown at the bottom. The inserted horizontal dotted line in **Figures 8A** and **8B** separates packings with embedded polar groups below the line from packings without this feature above the line. This demonstrates clearly the reduction of silanol activity with this type of bonded phase.

Both properties shown in **Figure 8** represent the most important attributes of a packing, but not the only ones. Packings exhibit other selectivity features as well. For example, the hydrogen-bond donor and acceptor properties can be measured using the relative retention between butylparaben and dipropylphthalate. These properties affect the selectivity of a packing for polar functional groups in a manner that is independent of the activity of residual silanols. A similar map as in **Figure 8** has been constructed for this polar selectivity in **Figure 9**. Generally, packings with embedded amide, carbamate, and urea groups excel at this feature, while classical  $C_{18}$  packings, shown at the bottom, lack this property. Packings below a polar selectivity value of  $\sim -0.4$  are classical bonded phases. In addition to these hydrogen-bond selectivity features, these packings are excellent for the separation of basic compounds, because they exhibit much less tailing than classical bonded phases.

Another important feature of a packing is its steric selectivity. It reflects the ability of a packing to discriminate between compounds that differ only in shape. The best-documented examples are polycyclic aromatic hydrocarbons, but the same feature is also

relevant for the separation of carotenoids and steroids, for example. For polynuclear aromatic hydrocarbons, it has been demonstrated that the shape selectivity correlates clearly with the bonding density of a packing. Thus, one can use the surface coverage, measured in  $\mu\text{mol m}^{-2}$ , as a measure of the steric selectivity of a packing. This property is commonly reported by the manufacturers of packings.

In summary, four parameters can be used to differentiate commercial silica-based HPLC packings: hydrophobicity as measured by the retention factor of a neutral hydrophobic compound; silanol activity as measured by the retention characteristic of a basic compound at neutral pH; hydrogen-bond acceptor/donor properties as measured by the relative retention of a hydrogen-bond donor and a suitable polar reference compound; and the steric selectivity measurable by the surface coverage of a packing. Packings that are close to each other in every one of these parameters are likely to give similar separations. On the other hand,  $C_{18}$  packings that are far away from each other have different retention properties. It is therefore important to specify the specific brand of a packing in publications or other documents.

**See also:** **Liquid Chromatography:** Principles; Normal Phase; Reversed Phase; Instrumentation. **Nucleic Acids:** Chromatographic and Electrophoretic Methods; Spectroscopic Methods. **Optical Spectroscopy:** Stray Light. **Ozone.** **Paints:** Water-Based; Organic Solvent-Based. **Peptides.** **Personal Monitoring:** Passive.



## Further Reading

- Cabrera K, Lubda D, Eggenweiler H-M, Minakuchi H, and Nakanishi K (2000) *Journal of High Resolution Chromatography* 23: 93.
- Cheng Y-F, Walter TH, Lu Z, *et al.* (2000) *LC/GC North America* 18(11): 1162–1172.
- Dunlap CJ, Carr PW, McNeff CV, and Stoll D (2001) *Analytical Chemistry* 73: 598A–607A.
- Meyer VR (1999) *Practical High-Performance Liquid Chromatography*. New York: Wiley.
- Neue UD (1997) *HPLC Columns – Theory, Technology, and Practice*. New York: Wiley-VCH.
- Neue UD (2000) Silica gel and its derivatization for liquid chromatography. In: Meyers RA (ed.) *Encyclopedia of Analytical Chemistry*, pp. 11450–11472. Chichester: Wiley.
- Neue UD, Tran K, Iraneta PS, and Alden BA (2003) *Journal of Separation Science* 26: 174–186.
- O’Gara JE, Walsh DP, Alden BA, Casellini P, and Walter TH (1999) *Analytical Chemistry* 71: 2992–2997.
- Poole CF (2003) *The Essence of Chromatography*. Amsterdam: Elsevier.
- Poole CF and Poole SK (1991) *Chromatography Today*. Amsterdam: Elsevier.
- Sander LC, Pursch M, and Wise SA (1999) *Analytical Chemistry* 71: 4821–4830.
- Snyder LR, Kirkland JJ, and Glajch JL (1997) *Practical HPLC Method Development*. New York: Wiley.
- Tanaka N, Kobayashi H, Nakanishi K, Minakuchi H, and Ishizuka N (2001) *Analytical Chemistry* 73: 420A–429A.
- Unger KK (1979) *Porous Silica*. Amsterdam: Elsevier Scientific.
- Zou H, Huang X, Ye M, and Luo Q (2002) *Journal of Chromatography A* 954: 5–32.

## Packed Capillary

**N W Smith**, King’s College, London, UK

**C Legido-Quigley, N D Marlin, and V Melin**, Imperial College London, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Liquid chromatography (LC) is the term used to describe the separation of a solution following differential migration in a liquid flowing through a column packed with solid particles. Miniaturized liquid separation techniques began in the late 1960s. In 1967, Horváth *et al.* developed stainless steel columns ranging from 0.5 to 1.0 mm internal diameter (i.d.) packed with pellicular particles for the separation of nucleotides. In the mid-1970s, Ishii *et al.* prepared 0.5 mm i.d.  $\times$  1.5 m long columns using pellicular particles and introduced slurry-packed Teflon microcolumns. In the same decade, Scott *et al.* prepared 10 mm i.d.  $\times$  10 m long columns for the separation of alkylbenzenes. In the late 1970s and the early 1980s, improvements were made to the technique and in this way the criteria were set for the next three decades of research.

In conventional high-performance liquid chromatography (HPLC), columns are usually of 10–25 cm length and 2–4 mm i.d. Recently, capillary HPLC columns have become available in the market. The terminology, abbreviations, and definitions have not yet been standardized. In the literature, we can find microbore, micro-LC, semimicro-LC,

capillary-LC, and nano-LC. In general, microbore or narrow-bore columns are considered as those with an internal diameter of less than 2 mm. From 1 mm down, micro-LC is more appropriate. Capillary columns are described as those of less than 0.5 mm i.d. and nano-LC columns usually have an internal diameter of less than 0.1 mm.

The advantages of using microflows instead of normal flows are significant. An important advantage is the use of less mobile phase as illustrated in **Table 1**. This addresses both the cost of purchase and the cost of disposal of these solvents. It also diminishes the environmental impact of the toxic solvents. Arguably, the most important advantage of using microflow through capillary columns is the higher sensitivity that can be achieved. When scaling down the internal diameter of a column, the analysis becomes more sensitive. The increased mass sensitivity is attributed to the reduction in column internal diameter, which results in reduced dilution of the chromatographic band during analysis. This is very important when determining compounds present at low concentrations in limited sample volumes. Further, another important direct consequence of the

**Table 1** Dimensions in chromatography

Application	Column ID	Flow rate
Micro-LC	1 mm to 500 $\mu$ m	40–20 $\mu$ l min <sup>−1</sup>
Capillary LC	500–200 $\mu$ m	20 $\mu$ l min <sup>−1</sup> to 300 nl min <sup>−1</sup>
Nano-LC	100–75 $\mu$ m	300–180–80 $\mu$ l min <sup>−1</sup>

## Further Reading

- Cabrera K, Lubda D, Eggenweiler H-M, Minakuchi H, and Nakanishi K (2000) *Journal of High Resolution Chromatography* 23: 93.
- Cheng Y-F, Walter TH, Lu Z, *et al.* (2000) *LC/GC North America* 18(11): 1162–1172.
- Dunlap CJ, Carr PW, McNeff CV, and Stoll D (2001) *Analytical Chemistry* 73: 598A–607A.
- Meyer VR (1999) *Practical High-Performance Liquid Chromatography*. New York: Wiley.
- Neue UD (1997) *HPLC Columns – Theory, Technology, and Practice*. New York: Wiley-VCH.
- Neue UD (2000) Silica gel and its derivatization for liquid chromatography. In: Meyers RA (ed.) *Encyclopedia of Analytical Chemistry*, pp. 11450–11472. Chichester: Wiley.
- Neue UD, Tran K, Iraneta PS, and Alden BA (2003) *Journal of Separation Science* 26: 174–186.
- O’Gara JE, Walsh DP, Alden BA, Casellini P, and Walter TH (1999) *Analytical Chemistry* 71: 2992–2997.
- Poole CF (2003) *The Essence of Chromatography*. Amsterdam: Elsevier.
- Poole CF and Poole SK (1991) *Chromatography Today*. Amsterdam: Elsevier.
- Sander LC, Pursch M, and Wise SA (1999) *Analytical Chemistry* 71: 4821–4830.
- Snyder LR, Kirkland JJ, and Glajch JL (1997) *Practical HPLC Method Development*. New York: Wiley.
- Tanaka N, Kobayashi H, Nakanishi K, Minakuchi H, and Ishizuka N (2001) *Analytical Chemistry* 73: 420A–429A.
- Unger KK (1979) *Porous Silica*. Amsterdam: Elsevier Scientific.
- Zou H, Huang X, Ye M, and Luo Q (2002) *Journal of Chromatography A* 954: 5–32.

## Packed Capillary

**N W Smith**, King’s College, London, UK

**C Legido-Quigley, N D Marlin, and V Melin**, Imperial College London, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Liquid chromatography (LC) is the term used to describe the separation of a solution following differential migration in a liquid flowing through a column packed with solid particles. Miniaturized liquid separation techniques began in the late 1960s. In 1967, Horváth *et al.* developed stainless steel columns ranging from 0.5 to 1.0 mm internal diameter (i.d.) packed with pellicular particles for the separation of nucleotides. In the mid-1970s, Ishii *et al.* prepared 0.5 mm i.d.  $\times$  1.5 m long columns using pellicular particles and introduced slurry-packed Teflon microcolumns. In the same decade, Scott *et al.* prepared 10 mm i.d.  $\times$  10 m long columns for the separation of alkylbenzenes. In the late 1970s and the early 1980s, improvements were made to the technique and in this way the criteria were set for the next three decades of research.

In conventional high-performance liquid chromatography (HPLC), columns are usually of 10–25 cm length and 2–4 mm i.d. Recently, capillary HPLC columns have become available in the market. The terminology, abbreviations, and definitions have not yet been standardized. In the literature, we can find microbore, micro-LC, semimicro-LC,

capillary-LC, and nano-LC. In general, microbore or narrow-bore columns are considered as those with an internal diameter of less than 2 mm. From 1 mm down, micro-LC is more appropriate. Capillary columns are described as those of less than 0.5 mm i.d. and nano-LC columns usually have an internal diameter of less than 0.1 mm.

The advantages of using microflows instead of normal flows are significant. An important advantage is the use of less mobile phase as illustrated in **Table 1**. This addresses both the cost of purchase and the cost of disposal of these solvents. It also diminishes the environmental impact of the toxic solvents. Arguably, the most important advantage of using microflow through capillary columns is the higher sensitivity that can be achieved. When scaling down the internal diameter of a column, the analysis becomes more sensitive. The increased mass sensitivity is attributed to the reduction in column internal diameter, which results in reduced dilution of the chromatographic band during analysis. This is very important when determining compounds present at low concentrations in limited sample volumes. Further, another important direct consequence of the

**Table 1** Dimensions in chromatography

Application	Column ID	Flow rate
Micro-LC	1 mm to 500 $\mu$ m	40–20 $\mu$ l min <sup>−1</sup>
Capillary LC	500–200 $\mu$ m	20 $\mu$ l min <sup>−1</sup> to 300 nl min <sup>−1</sup>
Nano-LC	100–75 $\mu$ m	300–180–80 $\mu$ l min <sup>−1</sup>

small flow rates delivered is the excellent conditions for coupling micro-HPLC to a mass spectrometer.

## The Rate Theory of Chromatography and Key Parameters

The rate theory of chromatography, also known as the van Deemter model, examines the factors affecting band broadening, which is the amount of dispersion of a sample as it migrates through a column.

The efficiency of the column can be approximated by the following expression known as the classical Knox equation:

$$H = Au^{1/3} + \frac{B}{u} + Cu \quad [1]$$

The equation tries to account for all the kinetic processes that occur when a separation is undertaken.  $A$ ,  $B$ , and  $C$  are coefficients that relate to the three different processes:  $A$  being the Eddy diffusion,  $B/u$  the longitudinal axial diffusion,  $Cu$  the mass transfer term, and  $u$  the mean linear velocity.

Van Deemter *et al.*, Giddings, Huber, and Horváth *et al.* have used alternative models to account for the relationship between the solute concentration in the stationary phase, its mobile phase concentration, and the various parameters characterizing the chromatographic system used. In general, when all the contributions of the parameters are accounted for, the separation will be performed at the flow depicted by the smallest  $H$  value, i.e., the greatest efficiency.

Efficiency is a measure of how well the chromatographic column is performing. Another term that accounts for efficiency is the value  $N$ , which is called the number of theoretical plates. The term 'plates' originates from the theory of distillation columns, where equilibrium between two phases occurs in each plate. For LC, a column can be thought of as a series of 'plates' where equilibration of the analyte occurs between two phases. The greater the number of theoretical plates the more efficient the separation.

A useful way of describing efficiency is by taking into account the length of the column  $L$ , in this way calculating the height equivalent to a theoretical plate  $H$  or HETP:

$$H = \frac{L}{N} \quad [2]$$

Hence, an efficient separation will have a small  $H$  value. When columns of various lengths are compared, it is more useful to use plates per meter  $N'$  as a measure of column efficiency:

$$N' = \frac{N}{l} \quad [3]$$

where  $l$  is the length of the column in meters.

$N$  is widely used to compare columns, but this can only be done when the particle size and test analytes are the same. The reduced plate height  $h$  can also be calculated, which takes into account the diameter of the particles  $d_p$ :

$$h = \frac{H}{d_p} \quad [4]$$

Next, we will consider the important factors when using these smaller diameters.

The down-scale factor was studied by Ryan in 1996 and it applies to all components and parameters of the system, such as flow rates, injection, detection volumes, and connecting capillaries. This factor is very important for the successful transfer of a conventional method to a microscale equivalent. The equation below shows that this factor is equal to the partition of the squares of both diameters:

$$f = \frac{d_{\text{conv}}^2}{d_{\text{micro}}^2} \quad [5]$$

where  $d_{\text{conv}}$  and  $d_{\text{micro}}$  are the diameters of the conventional and microscale columns.

### Flow Rates

When the column internal diameter is reduced, the mobile phase flow rate should also be reduced to maintain the same efficiency.

The column vacant volume can be expressed as

$$V_0 = \frac{1}{4}\pi d_c^2 \epsilon L \quad [6]$$

and the volumetric flow rate can be defined as

$$u = \frac{4F}{\pi d_c^2 \epsilon} \quad [7]$$

In order to maintain the same linear velocity for columns of different internal diameters, the volumetric flow rate should be decreased in proportion to the square of the ratio of the column internal diameters.

The linear velocity can also be expressed, based upon other column conditions, as

$$u = \frac{\Delta P d_p^2}{\Phi \eta L} \quad [8]$$

where  $\Delta P$  is the pressure drop across the column,  $d_p$  is the particle diameter,  $\Phi$  is the flow resistance parameter, and  $\eta$  is the viscosity of the mobile phase. Rearranging this equation gives:

$$\Delta P = \frac{\Phi \eta L u}{d_p^2} \quad [9]$$

This demonstrates that the column backpressure is independent of the column internal diameter. Therefore, microcolumns and analytical columns of the same length, packed with the same particles, and operated at the same linear velocity using the same mobile phase will generate similar backpressures. However, microcolumns are more susceptible to clogging; therefore, care must be taken by filtering solutions and samples. The use of low dead-volume inline-filters is also recommended.

### Injection and Detection Volumes

As the column diameter decreases, the loading capacity of the column also decreases, thus the maximum sample injection volume has to be reduced accordingly. The following equation describes the maximum injection volume that can be injected into a column:

$$V_{\max} = \frac{\theta K \pi \epsilon d_c^2 L (k + 1)}{\sqrt{N}} \quad [10]$$

where  $\theta$  is the fractional loss of the column plate number caused by the injection,  $K$  is a constant defining the injection profile,  $L$  is the length of the column,  $k$  is the retention factor, and  $N$  is the theoretical plate number.

The following equation relates the maximum sample mass to the maximum sample concentration that can be eluted from the column:

$$M_{\max} = \frac{C_m \pi \epsilon d_c^2 L k}{2\sqrt{N}} \quad [11]$$

In practice, as stated by Chervet, for a 75  $\mu\text{m}$  i.d.  $\times$  15 cm packed column, the maximum injected sample mass (loading capacity) for an unretained compound ( $k = 0$ ) should not exceed 50 ng.

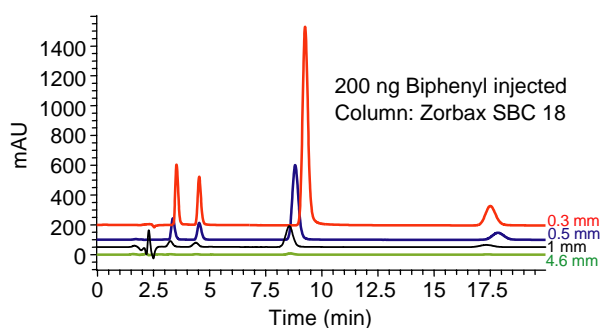
With regards to the optimization of the detection volume, the scale-down factor can also be applied to eqn [2], and a good compromise between sensitivity, noise, and peak dispersion would consist in using detection volumes similar to injection volumes.

### Sensitivity Gain

What is the maximum sensitivity gain possible in liquid chromatography by decreasing the internal diameter of columns?

In 1995, Ryan proposed a model to adapt normal-flow chromatography to microflow chromatography. The theory states that when going from a column of diameter  $X$  to a smaller diameter  $Y$ , the gain in sensitivity would be equivalent to the partition of the squares of the internal diameters [ $X^2/Y^2$ ].

Thus, in practical terms, when adapting a method from a 2.1 mm i.d. column to a 0.5 mm i.d. column,



**Figure 1** The response on four columns of different diameters after injecting the same amount of sample; the sensitivity increases greatly with the smaller diameters.

there would be a sensitivity gain of  $2.1^2/0.5^2 = 17$ , i.e., 17 times more response would be observed. From a 2.1 mm column to a 0.3 mm column the sensitivity increase could be as big as 49 times. **Figure 1** illustrates the sensitivity gain going from a 1.0 mm i.d. column to a 0.3 mm i.d. column, after injecting the same amount of sample.

### Instrumentation

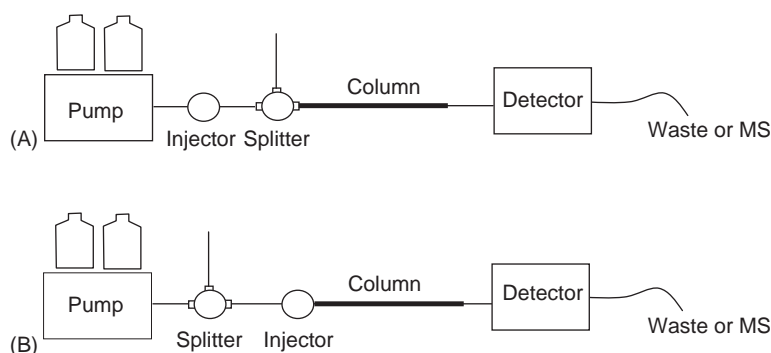
Today, micro-LC is a valuable analytical tool for sample-constrained applications such as proteomics and bioanalysis. The microliter flow rates are ideally suited for direct, splitless coupling with electrospray ionization mass spectrometry (ESI-MS). Many benefits can be demonstrated from well-established HPLC theory, which allows for direct method transfer to micro-LC.

Micro-HPLC has some key differences to conventional LC. Microbore columns require a very low flow rate, which needs to be delivered in a continuous flow. In generating sufficient flow through the columns, much higher backpressures are experienced, compared to standard HPLC. Also, the amount of analyte injected onto the column must be much lower, and all internal dead volumes must be minimized and reduced, otherwise band broadening effects will cause a significant problem.

Conventional HPLC instruments cannot meet these specific requirements, and therefore are simply not suitable for micro-LC. It is possible to downscale standard HPLC equipment by employing flow splitters, capillary flow cells, microconnectors, and fittings; however, these instruments can be unreliable.

The problems can be grouped into three categories:

- detector signal-to-noise,
- gradient formation, and
- flow splitting.



**Figure 2** Flow splitting device: (A) shows the injector before the splitter, which means that both the flow and the sample are split before going on to the column; (B) shows the injector after the splitter, so just the flow is being split, and a full sample injection is going on to the column.

The immediate result of attempts to scale down separations is a lack of detector sensitivity. Analytical detector flow cells introduce excessive band spreading and low peak response, drastically reducing peak capacity and making peak integration all but impossible. By simply using online detection methods, for example, passing the light across the capillary at the end of the column minimizes band spreading, but the short pathlengths can yield very low absorbance.

The other problem arises from the formation of accurate gradients under such low flow conditions. For gradients to be formed, one solvent must be accurately metered at only fractions of the total flow. One solution to this problem, ensuring microliter per minute flow rates, is to employ a flow splitter. Here, the flow is split into two separate paths depending on the backpressure of the restrictor used, allowing flow to the column and to waste. While the ratio of these two pressures remains constant, a consistent flow rate can be achieved; however, this ratio may change over time, due to the backpressure altering, resulting in the flow to the column changing. Another problem arising from splitting the flow as opposed to a system designed specifically to deliver microflows is the greater costs due to the high solvent consumption. **Figure 2** presents the two different configurations for flow splitting in micro-LC.

Micro-HPLC instruments are now commercially available and are designed to meet the exact requirements needed for microflows and offer the highest performance with ease of use. There are several micro/capillary LC instruments available in the market, and these include the Agilent 1100 Capillary LC™ system, LC packings Ultimate™ System, and Waters CapLC® System.

The Agilent 1100 Capillary LC works by splitting the relatively high flow rates to the column. The instrument utilizes a novel flow-monitoring device, called an electronic flow control, which monitors the

flow rate and adjusts it accordingly. The device works by measuring the actual flow rate, rather than using calculations.

This instrument is capable of providing flow rates as low as  $1 \mu\text{L min}^{-1}$ , running accurate gradients, and has a diode-array detector with a 500 nL flow cell.

#### Waters CapLC® System

The solvent delivery system in this instrument uses a syringe-style, positive-displacement, continuous delivery design. Each syringe is motor driven and software controlled so that crossover-related flow phenomena are eliminated. Gradients can also be performed, and are mixed under high pressure to minimize the gradient delay.

There is a dual wavelength detector, with patented low-volume flow cells (250 and 40 nL), which combines a long pathlength and high light throughput (CapLC cell). This focuses and guides light from a deuterium lamp source along the length of the flow cell.

Isocratic flow rates can be as low as  $250 \text{ nL min}^{-1}$ , while gradient flow rates can be generated from  $1 \mu\text{L min}^{-1}$ , while injection volumes can be as low as 20 nL.

#### Ultimate™ Fully Integrated Capillary HPLC System from LC Packings/Dionex

This instrument uses a helium sparging device with four separately controlled solvent lines, ensuring optimal solvent degassing and improving check valve reliability at low flow rates, diminishing baseline disturbances at low ultraviolet (UV) wavelengths.

The pumping system uses a high-precision reciprocating pump with proprietary microflow processing based on flow splitting. There are various calibrators that can be inserted to alter the extent of the flow splitting. This can generate flow rates from  $200 \mu\text{L min}^{-1}$  to  $50 \text{ nL min}^{-1}$ . The system uses a



microautosampler system to perform automated sample injection, but can also use a six-port low-dispersion injection valve for manual sample injection (1  $\mu\text{L}$  minimum).

The instrument has a specially designed scanning UV-visible absorbance detector, which can simultaneously monitor up to four different wavelengths. There are various interchangeable flow cells with volumes from 180 to 3 nL. They are used in conjunction with capillaries of 150 and 20  $\mu\text{m}$  i.d., respectively.

### Micropumps

Pumps are now available that can deliver micro- and nanoflows down to 100 nL min<sup>-1</sup>, without the need for flow splitters. These are generally reciprocating or syringe pumps.

One such pump is the MicroTech Ultra-Plus, which can deliver ultralow flows using a micro-reciprocating piston, which is precise and digitally controlled. This can deliver flow rates to a minimum increment of 0.1  $\mu\text{L}$  min<sup>-1</sup>, and gradient flow rates as low as 0.01  $\mu\text{L}$  min<sup>-1</sup> with a split, and 5  $\mu\text{L}$  min<sup>-1</sup> without a split.

Another pump, the Evolution 200, can also deliver low flow rates with high precision, with isocratic flows down to 0.1  $\mu\text{L}$  min<sup>-1</sup>, and gradient flows from 1  $\mu\text{L}$  min<sup>-1</sup>. This system uses binary high-pressure gradient formation with two linear drive piston pumps.

With the use of very low diameter columns, dead volume becomes a key problem, so manufacturers have gone to great lengths to try and minimize these effects. There are various column formats that have been employed for micro-LC, including fused silica capillary columns, glass-lined stainless steel, and glass-lined peek such as peeksil.

Columns for capillary LC are often prepared in fused-silica capillaries. These capillaries have a polyimide outer coating to increase the robustness of the column. Generally, the columns are packed with conventional HPLC chromatographic phases, with particle sizes of 3 and 5  $\mu\text{m}$  diameter. Frits have to be manufactured in order to retain the packing material, and can be made from sintering silica or directly sintering the packing material.

The column can be packed by various methods, but the most common method is to use a slurry of the stationary phase. Once the initial end frit has been manufactured, the capillary can then be attached to a packing reservoir, which contains a slurry of the stationary phase. The packing reservoir in turn is attached to a high-pressure packing pump that delivers the packing solvent. The packing material is

generally ultrasonicated during packing, to ensure a well-packed chromatographic bed, and then the second retaining frit is sintered directly from the packing material.

More recently, columns have been developed where the stationary phase is formed of a porous polymer network inside the capillary. These are called monolithic phases, and have emerged as an alternative to traditional packed bed columns for use in micro-HPLC. They hold many advantages over traditional packed bed columns, being easy to manufacture since the monolith is formed *in situ*, often via a one-step reaction process, and its properties such as porosity, surface area, and functionality can be tailored. Another major advantage is that they eliminate the need for retaining frits. These columns can be manufactured from a variety of materials, but the most common include sol-gel, methacrylate-based, acrylamide-based, and styrene-based polymeric structures.

There are a few commercially available monolithic columns on the market today, one of the first being the Chromolith<sup>TM</sup> from Merck, which uses a silica-based polymeric network for the high-speed separation of compounds of low to medium molecular weight at high flow rates with only low backpressures. More recently, LC Packings (a Dionex company) have released the Monolith<sup>TM</sup>, which is a polystyrene-divinylbenzene based monolith for the fast separation of proteins and peptides by micro LC.

## Applications of Micro-HPLC

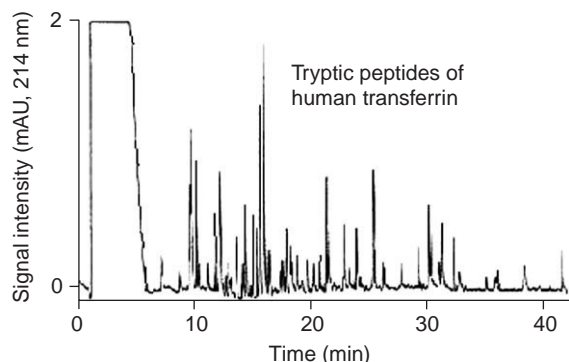
Micro-HPLC has been used for the analysis of a wide variety of analytes, ranging from polyaromatic hydrocarbons (PAHs) and agrochemicals, to biomolecules. In these applications, the microcolumn (i.d. less than 1 mm) consisted of either a traditional packed bed or a monolith, i.e., a polymer network.

### Proteomics

Chen *et al.* reported the enantioseparation of dansyl amino acids on 100  $\mu\text{m}$  i.d. ligand exchange-chiral monolithic microcolumns. Using continuous beds modified with chiral selectors such as L-phenylalaninamide, L-alaninamide, L-prolinamide, he achieved efficiencies of  $\sim 1000$  plates per meter.

Czerwenka *et al.* successfully performed the enantiomeric separations of a series of alanine peptides derivatized with different N-protection groups on packed beds containing a chiral stationary phase modified with the chiral selector *tert*-butyl-carbamoylquinine.

Huber's research group reported the high-resolution separation of peptides and proteins on poly(styrene-divinylbenzene) (PS/DVB) monoliths coupled with ESI-MS detection. More particularly, they were capable of separating and detecting very small



**Figure 3** High-resolution capillary RP-HPLC separation of tryptic peptides of human transferrin in a monolithic column. Column, monolithic PS/DVB,  $60 \times 0.20$  mm i.d.; mobile phase: (A) 2.0% acetonitrile, 0.050% TFA in water; (B) 80% acetonitrile, 0.050% TFA in water; linear gradient, 0–40% B in 30 min, 40–100% B in 10 min; flow rate,  $1.7 \mu\text{L min}^{-1}$ ; temperature,  $50^\circ\text{C}$ ; detection, UV, 214 nm; sample, tryptic digest of 1.0 pmol of human transferrin. (Reprinted with permission from Premstaller *et al.* (2001) *Analytical Chemistry* 73: 2392–2394; © 2003 American Chemical Society.)

amounts (lower femtomole range) of tryptic peptides of human transferrin (Figure 3), bovine catalase (Figure 4), standard proteins (Figure 5), as well as membrane proteins from the photosystem II of higher plants.

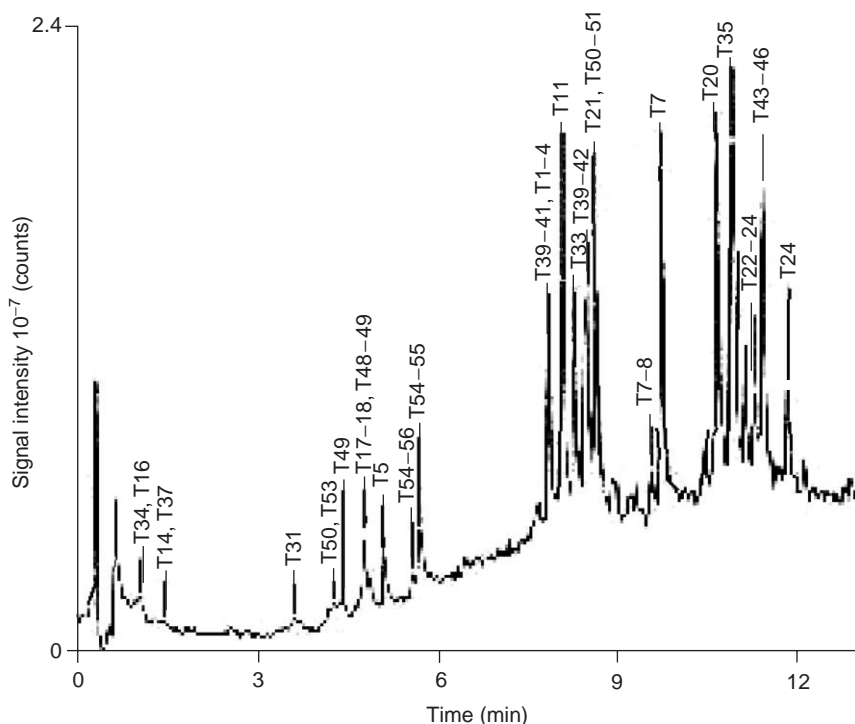
Similarly, Huang *et al.* reported the separation of standard proteins, tryptic digests of cytochrome *c*, and myoglobin on both macroporous underivatized and octadecylated PS/DVB monoliths.

Hjerten's research group described the very fast separations (less than 100 s) of standard proteins and peptides on acrylamide-based monolithic capillary columns derivatized with  $\text{C}_{18}$  ligands.

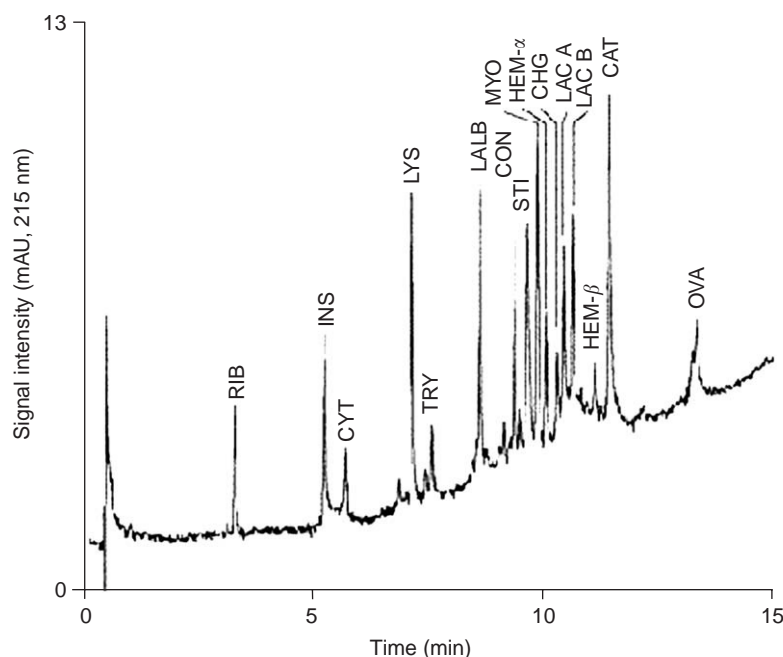
Schenk *et al.* reported the quantification of cytokines in cell extracts using micro-HPLC packed microcolumns coupled with immunochemical detection. Various types of chromatographies were associated with the immunochemical detection system in order to perform the separation of cytokines. These included cation-exchange, size exclusion, and reversed-phase chromatography, with the latter being the most successful.

### Nucleic Acids

Huber's research group reported the rapid and highly efficient (efficiencies of up to 190 000 plates per



**Figure 4** Separation and mass analysis of tryptic peptides of bovine catalase. Column, monolithic PS/DVB,  $60 \times 0.20$  mm i.d.; mobile phase: (A) 0.050% TFA in water; (B) 80% acetonitrile, 0.050% TFA in water; linear gradient, 2–60% B in 15 min; flow rate,  $1.8 \mu\text{L min}^{-1}$ ; temperature,  $50^\circ\text{C}$ ; scan,  $m/z$  400–2000; electrospray voltage, 4.0 kV; sample, 5 pmol of tryptic digest of bovine catalase. (Reprinted with permission from Premstaller *et al.* (2001) *Analytical Chemistry* 73: 2392–2394; © 2003 American Chemical Society.)



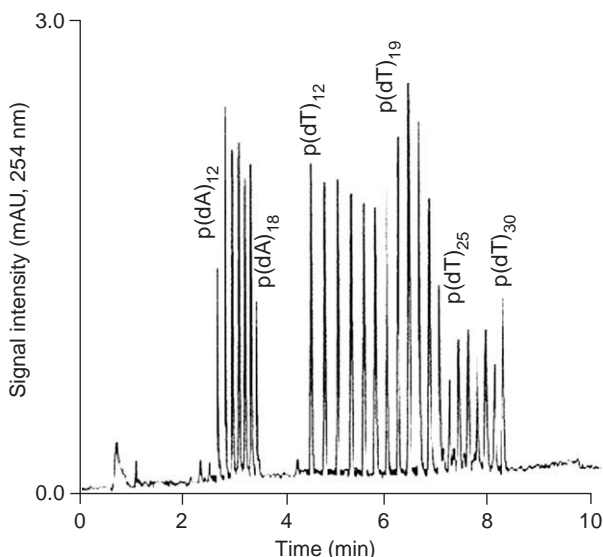
**Figure 5** High-resolution capillary RP-HPLC separation of 16 proteins in a monolithic capillary column. Column, monolithic PS/DVB,  $60 \times 0.20$  mm i.d.; mobile phase: (A) 15% acetonitrile, 0.20% TFA in water; (B) 60% acetonitrile, 0.20% TFA in water; linear gradient, 28–93% B in 15 min; flow rate,  $3.2 \mu\text{L min}^{-1}$ ; temperature,  $80^\circ\text{C}$ ; detection, UV, 214 nm; sample, mixture of 16 proteins, 200–350 fmol of each protein. (Reprinted with permission from Premstaller *et al.* (2001) *Analytical Chemistry* 73: 2392–2394; © 2003 American Chemical Society.)

meter) separations of single-stranded oligodeoxynucleotides and double-stranded DNA fragments by ion-pair reversed-phase micro-HPLC on PS/DVB monoliths (Figure 6) and conventional capillary columns packed with micropellicular octadecylated PS/DVB particles. The coupling with ESI-MS allowed the high resolution and identification of very small quantities of samples (femtomole amounts).

Simek *et al.* described the analysis and quantification of the purine bases hypoxanthine, xanthine, and guanine in excreta of ticks using a traditional reversed-phase  $\text{C}_{18}$  packed microcolumn.

### Agrochemicals

Cappiello's research group applied micro-HPLC to the screening of water samples for pesticide contamination. With electron ionization mass spectrometric detection and the use of two microcolumns packed with a  $\text{C}_{18}$  silica-based stationary phase, they performed the preconcentration of water samples and successfully detected trace levels of pollutants in similar samples. They also showed that ion-interaction micro-HPLC on a  $\text{C}_{18}$  stationary phase with hexylamine as the ion-pair reagent and coupling to particle beam mass spectrometry could be successfully used for the analysis of herbicides ranging from acidic species such as



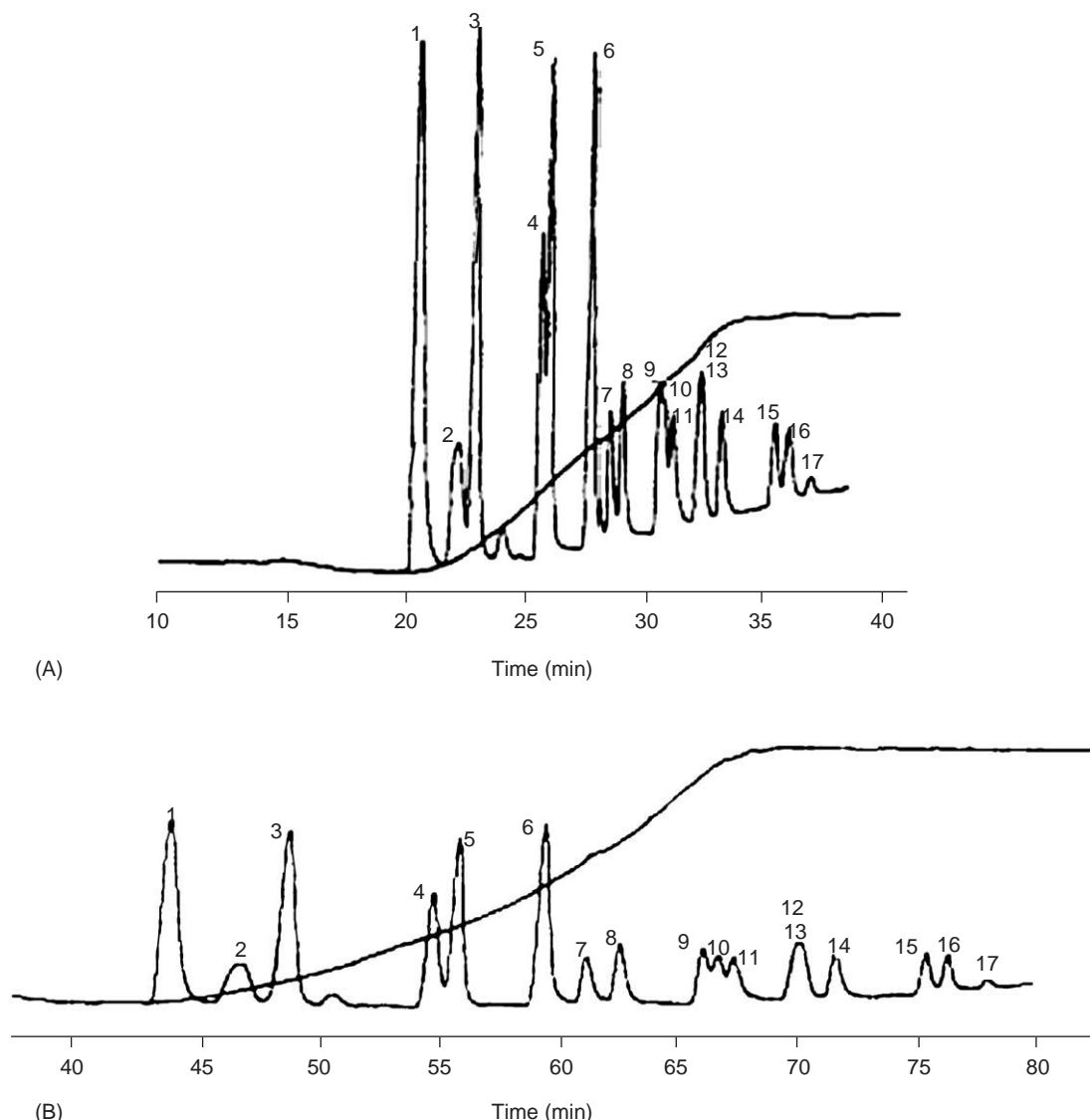
**Figure 6** High-resolution capillary IP-RP-HPLC separation of phosphorylated oligodeoxynucleotide ladders in a monolithic capillary column. Column, continuous PS/DVB,  $60 \times 0.20$  mm i.d.; mobile phase: (A)  $100 \text{ mmol L}^{-1}$  TEAA, pH 6.97; (B)  $100 \text{ mmol L}^{-1}$  TEAA, pH 6.97, 20% acetonitrile; linear gradient, (a) 15–45% B in 3.5 min, 45–55% B in 2.5 min, 55–65% B in 4.0 min; flow-rate,  $2.5 \mu\text{L min}^{-1}$ ; temperature,  $50^\circ\text{C}$ ; detection, UV, 254 nm; sample,  $\text{p(dA)}_{12-18}$ ,  $\text{p(dT)}_{12-30}$ , 40–98 fmol of each oligodeoxynucleotide. (Reprinted with permission from Premstaller *et al.* (2000) *Analytical Chemistry* 72: 4388; © 2003 American Chemical Society.)

phenoxy acids, to weak basic species such as phenol ureic herbicides, in a single analysis. An example of the separation of 17 acidic pesticides is featured in Figure 7.

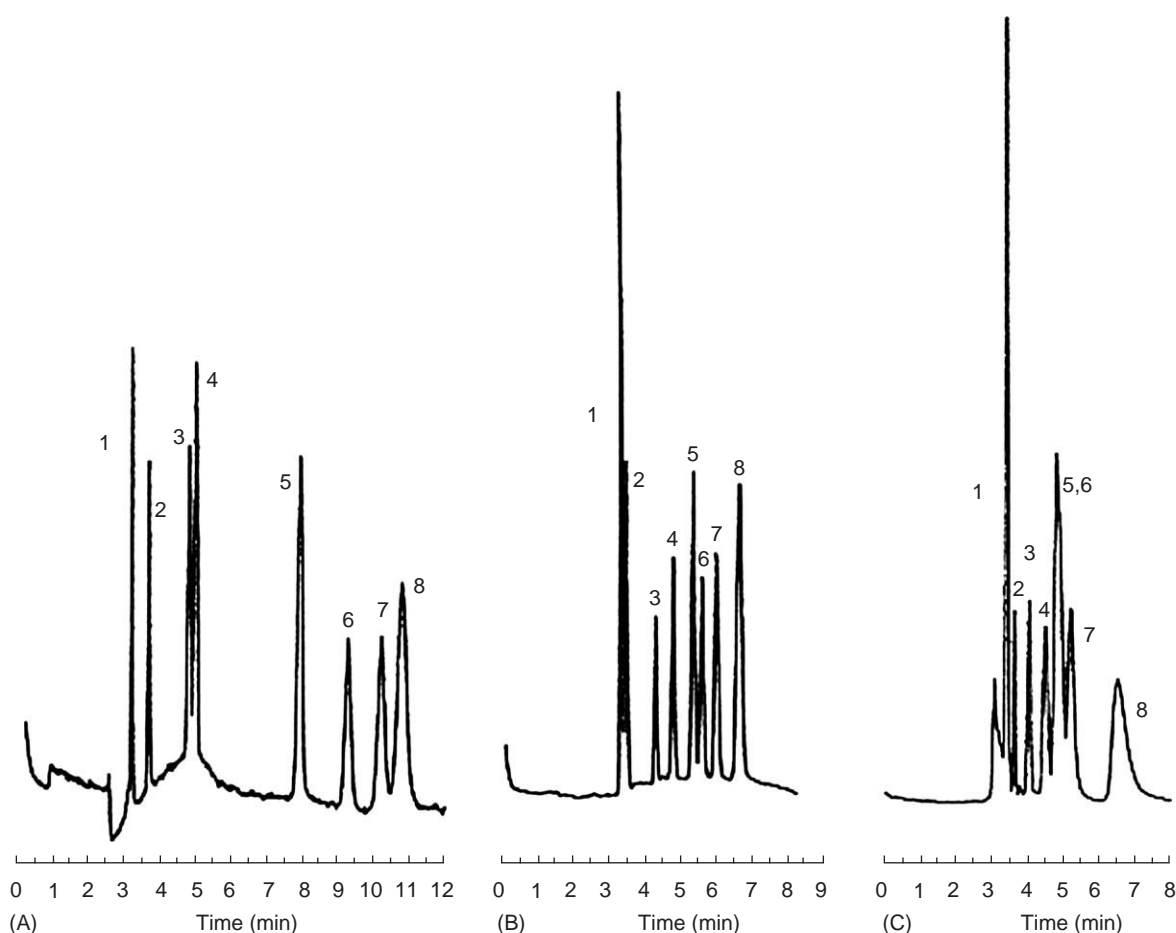
Collins *et al.* described the ultrafast (less than 1 min) separations of triazine herbicides and benzodiazepines on sol-gel bonded continuous beds filled with large pore octadecylsilica particles coupled with time-of-flight mass spectrometric detection. Efficiencies of up to 50 000 plates per meter were achieved with the benzodiazepines.

#### Other Aromatic Compounds: PAHs and Alkylbenzenes

Maruska *et al.* described the separations of polar aromatic solutes such as pyridine, 4-pyridylmethanol, 4-methoxyphenol, 2-naphthol, catechol, hydroquinone, resorcinol, and 2,7-dihydroxynaphthalene by normal-phase micro-HPLC on acrylic polymer-based continuous beds as illustrated in Figure 8. Efficiencies of up to 150 000 plates per meter were achieved on columns II and III, which only differ by their monomer composition.



**Figure 7** Chromatographic separation of 17 acidic pesticides obtained injecting 50  $\mu\text{L}$  of aqueous sample in different injection mode: (A) excluding and (B) including the loop in the mobile phase path after completion of the injection process. Chromatographic conditions: (water + 0.05% TFA)–(methanol + 0.025% TFA), 100:0 to 20:80 in 40 min; flow rate, 2  $\mu\text{L min}^{-1}$ ; sample concentration, 1  $\text{ng } \mu\text{L}^{-1}$ . UV detection of pesticides was performed at 225 nm. UV detection of mobile phase concentration (solid line) was performed at 190 nm. 1 = Picloram, 2 = 4-nitrophenol, 3 = Chloramben, 4 = 2,4-dinitrophenol, 5 = Dicamba, 6 = Bentazone, 7 = 2,4-D, 8 = MCPA, 9 = Mecoprop, 10 = Dichlorprop, 11 = 2,4,5-T, 12 = 2,4-DB, 13 = MCPB, 14 = 2,4,5-TP, 15 = Dinoseb, 16 = Dinoterb, 17 = pentachlorophenol. (Reprinted with permission from Cappiello *et al.* (1997) *Journal of Chromatography A* 768: 215–222; © 2003 Elsevier.)



**Figure 8** Normal-phase capillary chromatography of polar aromatic compounds: pyridine (1), 4-pyridylmethanol (2), 4-methoxyphenol (3), 2-naphthol (4), catechol (5), hydroquinone (6), resorcinol (7), 2,7-dihydroxynaphthalene (8). UV detection at 220 nm. (A) Column II, mobile phase: hexane–ethanol–methanol, pressure: 62 bar; (B) Column II, mobile phase: pure methanol, pressure: 56 bar; (C) Column III, mobile phase: pure methanol, pressure: 5 bar. Column II: 100  $\mu\text{m}$  i.d., 125 mm effective length, 175 mm total length, manufactured from a polymerization mixture containing 300 mg *N*-isopropylacrylamide, 50 mg methacrylamide, 125 mg piperazine diacrylamide, and 10 mg ammonium persulfate dissolved in 1 ml of 50  $\text{mol l}^{-1}$  sodium phosphate, pH 7. Column III: 100  $\mu\text{m}$  i.d., 126 mm effective length, 174 mm total length, manufactured from a polymerization mixture containing 180 mg 2-hydroxyethyl methacrylate, 5 mg vinylsulfonic acid, 150 mg piperazine diacrylamide, and 50 mg ammonium persulfate dissolved in 1 ml of 50  $\text{mol l}^{-1}$  sodium phosphate, pH 7. (Reprinted with permission from Maruska *et al.* (1999) *Journal of Chromatography A* 1837: 25–33; © 2003 Elsevier.)

Ericson *et al.* described the separation of a series of PAHs (naphthalene, 2-methylnaphthalene, fluorine, phenanthrene, and anthracene) on a 25  $\mu\text{m}$  i.d. continuous bed with  $\text{C}_{18}$  ligands and immobilized dextran sulfate. Efficiencies of up to 105 000 plates per meter were obtained.

Tanaka's research group reported the analysis of a series of alkylbenzenes ( $\text{C}_6\text{H}_5-(\text{CH}_2)_n\text{H}$ ,  $n=0-6$ ) and PAHs (naphthalene, fluorine, phenanthrene, anthracene, pyrene, triphenylene, and benzo[*a*]pyrene) on tetramethoxysilane-based macroporous silica gel monoliths and achieved efficiencies up to 80 000 plates per meter.

Tang *et al.* successfully performed the separation of aromatic standards and aromatic amines on

monolithic columns containing sol-gel bonded octadecylsilica particles.

At present, micro-HPLC is not widely used despite its obvious advantages. For years, the lack of suitable instrumentation/hardware inhibited the use of micro-HPLC. Modification of conventional equipment helped kick-start the technique but the lack of suitable hardware (injectors/capillary columns) still prevented its widespread use. These issues have been addressed in the laboratories of micro-LC practitioners in academia and industry and by a few commercial suppliers. However, there still remains a perception that micro-LC is difficult to perform and therefore the exciting advantages of the technique go unrealized by many in the separation sciences. With



the advance in instrumentation and the successful manufacture of robust capillary columns, it is believed that the technique is going to expand and within a few years may become the dominant technique of choice for analysis.

See also: **Liquid Chromatography**: Overview; Principles; Reversed Phase; Ion Pair; Liquid Chromatography–Mass Spectrometry.

## Further Reading

- Chervet JP, Ursem M, Salzmann JP, and Vannoort RW (1989) Ultra sensitive uv detection in micro separation. *Journal of High Resolution Chromatography* 12(5): 278–281.
- Davis MT and Lee TD (1992) Analysis of peptide mixtures by capillary high performance liquid chromatography: A practical guide to small-scale separations. *Separations in Protein Science* 1: 935.

- Henzel WJ, Bourell JH, and Stults JT (1990) Analysis of protein digests by capillary high-performance liquid chromatography and on-line fast atom bombardment mass spectrometry. *Analytical Biochemistry* 187: 228.
- Ishii D (1988) *Introduction to Micro-Scale High-Performance Liquid Chromatography*. Weinheim: Wiley-VCH Publishing.
- Krejci M (ed.) (1992) *Trace Analysis with Microcolumn Liquid Chromatography*. New York: Dekker.
- Neue UD (1997) *HPLC Columns: Theory, Technology and Practice*. New York: Wiley-VCH.
- Novotny M (1981) Microcolumns in liquid chromatography. *Analytical Chemistry* 53: 1294A.
- Snyder L and Kirkland J (1979) *Introduction to Modern Liquid Chromatography*, 2nd edn. New York: Wiley-Interscience.
- Weston A and Brown P (1997) *HPLC, CE: Principles and Practice*, 1st edn. San Diego: Academic Press.
- Yang FJ (ed.) (1988) *Microbore Column Chromatography A Unified Approach to Chromatography*. New York: Dekker.

## Mobile Phase Selection

**M Swartz**, Waters Corporation, Milford, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by J-L Rocca, pp. 2579–2584, © 1995, Elsevier Ltd.

## Introduction

In liquid chromatography (LC), solute retention is governed by the solute distribution factor, which reflects the different interactions of solute–stationary phase, solute–mobile phase, and mobile phase–stationary phase. For a given stationary phase, the retention of given solutes depends directly on the mobile phase, the nature and composition of which have to be judiciously selected in order to get the appropriate solute retentions and the required solute separation. The mobile phase has to be adapted in terms of elution strength (solute retention) and solvent selectivity (solute separation). This adaptation of the solvent requires some preliminary experiments, and, according to the complexity of the mobile phase that is to be optimized, the use of manual or computer-based optimization procedures needs to be considered. Hundreds of solvents are available to the chromatographer for use as mobile phase components. Solvent properties that relate to detection, miscibility (mobile phases prepared for high-performance liquid chromatography analyses

must be miscible, preferably at all proportions), viscosity, and purity are all important considerations.

## Solvent Polarity

Polarity is a key word in many chromatographic separations since a polar mobile phase will give rise to a low solute retention in normal phase LC (liquid–solid chromatography, LSC, of adsorption chromatography), or to a high solute retention in reversed–phase LC (RPLC). Nevertheless, it is often unclear exactly what this term means. One way to define the concept of polarity is to consider the Hildebrand solubility parameter; another is to consider the Snyder solvent parameter.

### Hildebrand's Solubility Parameter

Hildebrand has defined a regular solution in which deviations from ideality are attributed only to the enthalpy of mixing, the intermolecular forces being dispersion forces. The equation describing this model is

$$\Delta H_{\text{mix}} = \varphi_S^2 V_A (\delta_S - \delta_A)^2 \quad [1]$$

where  $\varphi_S$  is the volume fraction of the solvent,  $V_A$  is the molar volume of solute A, and  $\delta_S$  and  $\delta_A$  are the so-called solubility parameters for the solvent and the solute, respectively, measuring the internal

the advance in instrumentation and the successful manufacture of robust capillary columns, it is believed that the technique is going to expand and within a few years may become the dominant technique of choice for analysis.

See also: **Liquid Chromatography**: Overview; Principles; Reversed Phase; Ion Pair; Liquid Chromatography–Mass Spectrometry.

## Further Reading

- Chervet JP, Ursem M, Salzmann JP, and Vannoort RW (1989) Ultra sensitive uv detection in micro separation. *Journal of High Resolution Chromatography* 12(5): 278–281.
- Davis MT and Lee TD (1992) Analysis of peptide mixtures by capillary high performance liquid chromatography: A practical guide to small-scale separations. *Separations in Protein Science* 1: 935.

- Henzel WJ, Bourell JH, and Stults JT (1990) Analysis of protein digests by capillary high-performance liquid chromatography and on-line fast atom bombardment mass spectrometry. *Analytical Biochemistry* 187: 228.
- Ishii D (1988) *Introduction to Micro-Scale High-Performance Liquid Chromatography*. Weinheim: Wiley-VCH Publishing.
- Krejci M (ed.) (1992) *Trace Analysis with Microcolumn Liquid Chromatography*. New York: Dekker.
- Neue UD (1997) *HPLC Columns: Theory, Technology and Practice*. New York: Wiley-VCH.
- Novotny M (1981) Microcolumns in liquid chromatography. *Analytical Chemistry* 53: 1294A.
- Snyder L and Kirkland J (1979) *Introduction to Modern Liquid Chromatography*, 2nd edn. New York: Wiley-Interscience.
- Weston A and Brown P (1997) *HPLC, CE: Principles and Practice*, 1st edn. San Diego: Academic Press.
- Yang FJ (ed.) (1988) *Microbore Column Chromatography A Unified Approach to Chromatography*. New York: Dekker.

## Mobile Phase Selection

**M Swartz**, Waters Corporation, Milford, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by J-L Rocca, pp. 2579–2584, © 1995, Elsevier Ltd.

## Introduction

In liquid chromatography (LC), solute retention is governed by the solute distribution factor, which reflects the different interactions of solute–stationary phase, solute–mobile phase, and mobile phase–stationary phase. For a given stationary phase, the retention of given solutes depends directly on the mobile phase, the nature and composition of which have to be judiciously selected in order to get the appropriate solute retentions and the required solute separation. The mobile phase has to be adapted in terms of elution strength (solute retention) and solvent selectivity (solute separation). This adaptation of the solvent requires some preliminary experiments, and, according to the complexity of the mobile phase that is to be optimized, the use of manual or computer-based optimization procedures needs to be considered. Hundreds of solvents are available to the chromatographer for use as mobile phase components. Solvent properties that relate to detection, miscibility (mobile phases prepared for high-performance liquid chromatography analyses

must be miscible, preferably at all proportions), viscosity, and purity are all important considerations.

## Solvent Polarity

Polarity is a key word in many chromatographic separations since a polar mobile phase will give rise to a low solute retention in normal phase LC (liquid–solid chromatography, LSC, of adsorption chromatography), or to a high solute retention in reversed–phase LC (RPLC). Nevertheless, it is often unclear exactly what this term means. One way to define the concept of polarity is to consider the Hildebrand solubility parameter; another is to consider the Snyder solvent parameter.

### Hildebrand's Solubility Parameter

Hildebrand has defined a regular solution in which deviations from ideality are attributed only to the enthalpy of mixing, the intermolecular forces being dispersion forces. The equation describing this model is

$$\Delta H_{\text{mix}} = \varphi_S^2 V_A (\delta_S - \delta_A)^2 \quad [1]$$

where  $\varphi_S$  is the volume fraction of the solvent,  $V_A$  is the molar volume of solute A, and  $\delta_S$  and  $\delta_A$  are the so-called solubility parameters for the solvent and the solute, respectively, measuring the internal

pressure of the corresponding substances. This solubility parameter is defined as the square root of the cohesive energy density according to

$$\delta = (-E/V)^{1/2} \quad [2]$$

where  $E$  is the cohesive energy and  $V$  is the molar volume,  $\delta$ , commonly expressed in  $\text{J}^{1/2}\text{ml}^{-1/2}$  or  $\text{cal}^{1/2}\text{ml}^{-1/2}$  is a measure of the interactive forces between molecules. The stronger the intermolecular interaction, the more effectively the compound acts as a solvent for polar solutes. Table 1 gives the solubility parameter of some compounds of chromatographic interest.

Chromatographers have used this solubility parameter to predict chromatographic retention in gas chromatography and in LC. Assuming that both phases, stationary and mobile, are bulk fluids (i.e., LC), the retention factor  $k_i$  of a solute is related to solubility parameters according to

$$\ln k_i = (V_i/RT)(\delta_m + \delta_x - 2\delta_i)(\delta_m - \delta_x) + \ln(n_x - n_m) \quad [3]$$

where  $\delta_m$ ,  $\delta_x$ , and  $\delta_i$  are the solubility parameters of the mobile phase, the stationary phase, and the solute, respectively. Consequently, in RPLC,  $\delta_m$  being larger than  $\delta_x$ , a mobile phase possessing a large solubility parameter will give rise to high solute retention. The elution strength of a mobile phase varies as the inverse of its solubility parameter. In LSC,  $\delta_m$  being lower than  $\delta_x$ , the elution strength of mobile phases varies according to their solubility parameters.

To a first approximation, the solubility parameters for mixed mobile phases can be written as

$$\delta_m = \Sigma \varphi_i \delta_i \quad [4]$$

where  $\varphi_i$  is the volume fraction of component  $i$  in the mixture, possessing the solubility parameter  $\delta_i$ . For a binary mixture of water ( $w$ ) and an organic modifier

(a), the solubility parameter is expressed as

$$\delta_m = \varphi_w \delta_w + \varphi_a \delta_a \quad [5]$$

and because  $\varphi_w + \varphi_a = 1$ , eqn [5] becomes

$$\delta_m = \delta_w + \varphi_a(\delta_a - \delta_w) \quad [6]$$

In RPLC, by considering eqn [5] and the  $\delta_a$  values from Table 1, to a first approximation, it may be expected that mixtures of the same solubility parameter will yield the same solute retention factor. In other words, if acetonitrile (ACN) or tetrahydrofuran (THF) is used instead of methanol (MeOH) in a binary mixture with water, a composition of  $\varphi_{\text{ACN}} = 0.78\varphi_{\text{MeOH}}$  or  $\varphi_{\text{THF}} = 0.62\varphi_{\text{MeOH}}$  is expected to yield roughly the same retention factor as the composition  $\varphi_{\text{MeOH}}$ .

Combining eqns [3] and [6] we find

$$\ln k_i = A\varphi_x^2 + B\varphi_a + C \quad [7]$$

where  $A$ ,  $B$ , and  $C$  are constants in the chromatographic system.

### Snyder's Solvent Parameter

According to the model of adsorption in LSC provided by Snyder, the solute distribution constant  $K$  is given by

$$\log K = \log V_a + E_a(S^0 - A_s \varepsilon^0) \quad [8]$$

where  $V_a$  is the adsorbent surface volume,  $E_a$  is the average surface activity of the adsorbent,  $S^0$  is the adsorption energy of the solute,  $A_s$  is the area of solid adsorbent required by the adsorbed solute, and  $\varepsilon^0$  is the Snyder eluent strength parameter. Relative partition coefficients obtained with two different mobile phases for a given solute and a given stationary phase are given by

$$\log(K_2 - K_1) = E_a A_s (\varepsilon_1^0 - \varepsilon_2^0) \quad [9]$$

**Table 1** Solubility parameters and partial solubility parameters (expressed in  $\text{cal}^{1/2}\text{ml}^{-1/2}$ ) for some compounds

Compound	$\delta_t$	$\delta_d$	$\delta_a$	$\delta_{ird}$	$\delta_a$	$\delta_b$
Water	25.52	7.2	—	—	21.7	14.2
Alumina	— 16	10.8	9.8	—	11.4	2.5
Methanol	15.85	7.2	3.9	0.1	17.1	5.4
Acetonitrile	13.14	7.3	5.8	0.2	10	4
Methylene chloride	10.68	8.0	4.38	1.0	4	1
1,4-Dioxane	10.65	8.1	1.2	3.4	Small	2.0
Tetrahydrofuran	9.88	8.0	3.3	0.3	6.2	1.5
Toluene	9.57	8.5	0.8	0.7	8	0.4
Ethyl acetate	9.53	7.6	3.6	1.4	3.5	1
Alkanes	~7	~7	0	0	0	0
Perfluorinated alkanes	— 5	— 5	0	0	0	0

The ratio of the distribution factors obtained with the two mobile phases depends on their difference in eluent strength ( $\epsilon_1^0 - \epsilon_2^0$ ) since  $E_a$  and  $A_s$  are constant. By definition,  $\epsilon^0$  has a value of zero for *n*-pentane, and consequently the eluent strength parameters  $\epsilon^0$  were determined for different solvents: the larger the value of  $\epsilon^0$ , the less retained is a solute by that solvent. Table 2 illustrates some eluent strength parameter for alumina ( $\text{Al}_2\text{O}_3$ ) as the stationary phase, given that

$$\epsilon^0(\text{silica}) = 0.77\epsilon^0(\text{alumina}) \quad [10]$$

For the elution strength of a binary mobile phase, a mixture of solvents 1 and 2, having  $\epsilon_1^0$  and  $\epsilon_2^0$  as the respective eluent strengths and knowing that component 2 is the more strongly sorbed component, Snyder obtained the following relationship:

$$\epsilon_{12}^0 = \epsilon_1^0 + [\log(X_2 10^W (\epsilon_2^0 - \epsilon_1^0) + 1 - X_2)] / W \quad [11]$$

with  $W = E_A S_2$ , where  $E_A$  is the activity of the adsorbent,  $S_2$  is the area occupied by solvent 2 of the mobile phase, and  $X_2$  is the mole fraction of solvent 2. Figure 1 illustrates the elution strengths of some binary mobile phases, with silica as stationary phase.

## Solvent Selectivity

Although the solubility parameters or the solvent strength parameters are useful for quantifying the polarity of a mobile phase, they characterize only global interactions of mobile phases with solutes and stationary phases. The concepts have been refined by considering the different types of interaction that together contribute to interactions between species.

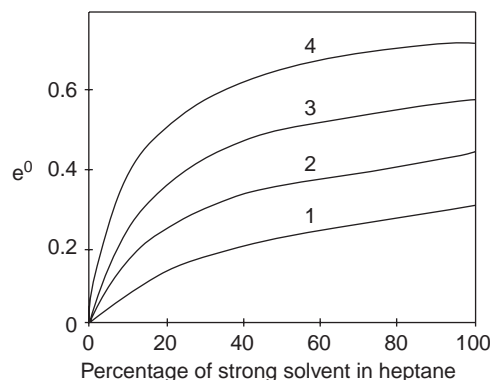
Interactions can be classified as follows:

- Dispersion interactions, which arise from the mutual interaction of the electron clouds of interacting species: all compounds, polar or nonpolar, exhibit such a dispersion interaction.
- Dipole orientation between molecules possessing as permanent dipole moment.
- Induced-dipole interactions, which occur when a permanent dipole induces a temporary dipole in a neighboring molecule.
- Acid–base interactions including all the processes in which electrons or protons are exchanged between molecules.

## Partial Solubility Parameter

Considering the different kinds of interaction, the solubility parameter can be written as a function of partial solubility parameters as

$$\delta_r^2 = \delta_d^2 + \delta_a^2 + 2\delta_{ms}\delta_d + 2\delta_a\delta_b \quad [12]$$



**Figure 1** Eluent strength ( $\epsilon^0$ ) of mixed solvents vs. the composition of a strong solvent in heptane. 1, Diisopropyl ether; 2, methylene chloride or chloroform; 3, ethyl acetate or acetone; 4, tetrahydrofuran or acetonitrile.

**Table 2** Eluent strength parameters,  $\epsilon^2$ , polarity index, and selectivity parameters for some classical solvents

Solvent	$\epsilon^0$ (alumina)	$P'$	$X_c$	$X_d$	$X_n$	Group
<i>n</i> -Hexane	0.00	0.1	—	—	—	
Toluene	0.22	2.4	0.25	0.28	0.47	VIII
Benzene	0.25	2.7	0.23	0.32	0.45	VIII
Diethyl ether	0.29	2.8	0.53	0.13	0.34	I
Methylene chloride	0.32	3.1	0.29	0.18	0.53	V
Isopropanol	0.63	3.9	0.55	0.19	0.27	II
Tetrahydrofuran	0.35	4.0	0.38	0.20	0.42	III
Chloroform	0.31	4.1	0.25	0.41	0.33	VIII
Ethanol	0.68	4.3	0.52	0.19	0.29	II
Ethyl acetate	0.45	4.4	0.34	0.23	0.43	VI
Dioxane	0.43	4.8	0.36	0.24	0.40	VI
Acetone	0.43	5.1	0.35	0.23	0.42	VI
Methanol	0.73	5.1	0.48	0.22	0.31	II
Acetonitrile	0.50	5.8	0.31	0.27	0.42	VI
Nitromethane	0.49	6.0	0.28	0.31	0.40	VII
Water	Large	10.2	0.37	0.37	0.25	VIII

where the subscripts refer to total (T), dispersion (d), orientation (o), induction (ind), acid (a), and base (b). Two different types of terms appear in eqn [12], viz. quadratic terms for dispersion and orientation interactions, which are symmetrical, and double product terms for induction and acid–base interactions, which are nonsymmetrical because the two molecules involved in the mutual interaction are not equivalent. Looking at Table 1, a distinction can now be made between solvents of similar overall polarity (similar total solubility parameter).

### Snyder Solvent Classification Scheme

Another way of classifying solvents in terms of their polarity and specific chemical interactions is the empirical scheme proposed by Snyder and based on experimental measurements of test solutes (probes) in gas–liquid chromatography, using LC mobile phases as stationary phases. The retention of probes reflects their interaction with the liquid stationary phase, so by selecting probes with selective interactions, Snyder could determine a set of numbers that characterized the liquids studied. Distribution constants were corrected for dispersion interactions and molecular weight effects ( $K^w$ ), and, assuming that there are three additive interaction (proton donor, proton acceptor, and strong dipole interactions), the distribution constants corrected for the three effects should sum to a total value called the polarity index  $P'$  defined according to

$$P' = \log(K^w)_d + \log(K^w)_e + \log(K^w)_n \quad [13]$$

where subscripts d, e, and n refer to these three interactions, respectively. Selecting dioxane, ethanol, and nitromethane as probes for measuring the proton donor effect, proton acceptor effect, and strong dipole effect, respectively,  $P'$  values were determined for the main solvents classically used in LC. Defining for each interaction a parameter such as

$$x_i = \log(K^w)_i / P' \quad [14]$$

$x_i$  will represent a selectivity parameter for measuring the proton donor character ( $i = d$ ), the proton acceptor character ( $i = e$ ), and the strong dipole character ( $i = n$ ). The sum of the three parameters is thus normalized to 1. Polarity index and selectivity parameters for some solvents are listed in Table 2. Using this approach, solvents that are chemically similar yield similar selectivity parameters, but solvents that are chemically different may appear in a single selectivity group. Some discrepancies can occur between the solvent classification according to the solubility parameters and to the polarity

index: acetonitrile, for example, appears as more polar than methanol in Table 2, whereas the reverse order is observed in Table 1. Nevertheless, this classification of solvents is very useful as the selectivity of solvent can be depicted in a triangular diagram because the sum of the selectivity parameters is equal to 1. Figure 2 illustrates such a representation: solvents can be classified into eight main classes; a summary of the resulting classification is given in Table 3.

To use the selectivity parameters in selecting solvent mixtures for LC, Snyder assumes that the mixture polarity is given by

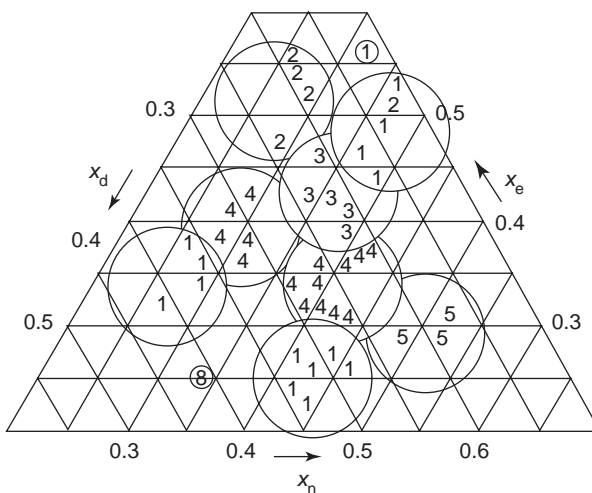
$$P'_{AB} = (V_A/V_B)P'_A + (V_B/V_A)P'_B \quad [15]$$

The solvents A and B should be from two of the eight solvent groups in Figure 2 and should be as different as possible. The best solvents to be used are

- In LSC: diethyl ether (or methyl *t*-butyl ether), methylene chloride (dichloromethane), and chloroform from groups I, V, and VIII, respectively, the nonpolar solvent being hexane.
- In RPLC: methanol, acetonitrile and tetrahydrofuran from groups 2, 6, and 3, respectively, the polar solvent being water.

### Solvent Viscosity

Solvent viscosity affects the pressure drop across a column and dictates the flow rate. For nonaqueous mobile phase component mixtures, the viscosity ( $\eta$ ) can be estimated from the viscosity of the individual A and B solvents ( $\eta_A$  and  $\eta_B$ ) and their mole fractions



**Figure 2** Solvent classification according to Snyder. (Reproduced with permission from Snyder LR (1978) Classification of the solvent properties of common liquid. *Journal of Chromatogr. Sci.* 16: 223–234; © Preston Publications Division.)



**Table 3** Solvent classification according to Snyder (see Figure 2)

Group	Solvents
I	Aliphatic ethers Trialkylamines (somewhat more basic than group I solvents)
II	Aliphatic alcohols
III	Pyridine derivatives, THF, sulfoxides
IV	Glycols, acetic acid
V	Methylene chloride (dichloromethane), 1,2-dichloroethane
VI	a: Aliphatic ketones and esters, dioxane b: Sulfones, nitriles
VII	Aromatic hydrocarbons, halosubstituted aromatic hydrocarbons, nitro compounds, aromatic ethers
VIII	Fluoroalknols, water Chloroform (somewhat less basic than group VIII solvents)

( $x_A$  and  $x_B$ ):

$$\log(\eta) = x_A \log(\eta_A) + x_B \log(\eta_B) \quad [16]$$

Viscosity also decreases with temperature. Equation [16] does not hold for aqueous mobile phases as a result of the strong interactions between water molecules. For aqueous mobile phases, and temperature effects see Snyder (in the Further reading section).

### Detection

In LC, a means of detection is employed for identification and quantification. While detection is covered elsewhere in this volume, mobile phase selection can play an important role in the detectability of the compounds of interest, and vice versa. A solvent's lowest usable (cutoff) wavelength is important for UV detectors, solvent refractive index (RI) effects the sensitivity of RI detection, and solvent volatility is an important consideration for evaporative light scattering and mass spectrophotometric based detectors. Table 4 lists some common LC mobile phase spectral data.

## Solvent Optimization

The use of optimization methods makes it possible to separate complex mixtures in a short analysis time with binary, ternary, or quaternary mobile phases. Only procedures concerning the optimization of selectivity are described in this section.

### Simultaneous Method of Optimization

In such methods, all the experiments that have to be performed are defined in a first step. After all the corresponding chromatograms have been recorded,

the optimum can be located. This grid search requires that the entire parameter space is covered, usually at regular intervals. If two isoeluotropic solvents are selected (ternary mixtures) there is only one degree of freedom, i.e., one parameter. In this case, intervals of 10% lead to 11 experiments. For three isoeluotropic solvents (quaternary composition), the parameter space is located in an equilateral triangle because the sum of the solvent volume fractions is equal to 1: intervals of 10% require 66 experiments. Consequently, such a procedure is very time consuming; no effort of computation is required but complete automation of the method is recommended and is easily possible.

### Sequential Method of Optimization

This optimization procedure, based on the simplex technique, requires a number of initial experimental conditions that is one more than the number of parameters considered. The resulting chromatograms are qualified and the worst conditions corresponding to a given location inside the parameter space are replaced by new conditions (new location), and the process is repeated until the quality of the latest chromatogram does not differ from the quality of the preceding one. This procedure does not require any chromatographic model but does require a large number of experiments and may lead to only a local optimum.

### Interpretive Methods

Interpretive methods involve a given set of pre-planned experiments from which chromatographic results are used for predicting retention surfaces of individual solutes according to given mathematical or statistical models. The model for the retention surfaces of individual solutes is then used to calculate the response surface over the parameter space explored, and to locate the optimum.

**Window diagram** Window diagrams, developed by Laub and Purnell for optimizing the composition of mixed stationary phases in gas chromatography, can be used for optimizing mobile phase composition in LC. From two initial experiments (if a linear relationship is assumed between  $\log k$  and mobile phase composition) or more (in the case of a quadratic relationship), the retention models are calculated for all solutes, and a response function (selectivity between every possible pair of solutes) is calculated and plotted versus the mobile phase composition. Areas or windows in which all solutes are separated can be located graphically. No particular effort of computation is required in such a procedure.

**Table 4** Mobile phase spectral data

	Absorbance (AU) at specified wavelengths (nm)									
	200	205	210	215	220	230	240	250	260	280
Acetonitrile	0.05	0.03	0.02	0.01	0.01	<0.01				
Methanol	2.06	1.00	0.53	0.37	0.24	0.11	0.06	0.02	<0.01	
Degassed	1.91	0.76	0.35	0.21	0.15	0.06	0.02	<0.01		
Isopropanol	1.80	0.68	0.34	0.24	0.19	0.08	0.04	0.03	0.02	0.02
Tetrahydrofuran										
Fresh	2.44	2.57	2.31	1.80	1.54	0.94	0.42	0.21	0.09	0.05
Old	>2.5	>2.5	>2.5	>2.5	>2.5	>2.5	>2.5	>2.5	>2.5	1.45
Acetic acid 1%	2.61	2.63	2.61	2.43	2.17	0.87	0.14	0.01	<0.01	
Hydrochloric 6 mmol l <sup>-1</sup> (0.02%)	0.11	0.02	<0.01							
Phosphoric acid, 0.1%	<0.01									
Trifluoroacetic acid										
0.1% in water	1.20	0.78	0.54	0.34	0.20	0.06	0.02	<0.01		
0.1% in acetonitrile	0.29	0.33	0.37	0.38	0.37	0.25	0.12	0.04	0.01	<0.01
Ammonium phosphate, dibasic 50 mmol l <sup>-1</sup>	1.85	0.67	0.15	0.02	<0.01					
Triethylemine, 1%	2.33	2.42	2.50	2.45	2.37	1.96	0.50	0.12	0.04	<0.01
Ammonium acetate, 10 mmol l <sup>-1</sup>	1.88	0.94	0.53	0.29	0.15	0.02	<0.01			
Ammonium tricarbonat 10 mmol l <sup>-1</sup>	0.41	0.10	0.01	<0.01						
EDTA (ethylenediaminetetraacetic acid), disodium, 1 mmol l <sup>-1</sup>	0.11	0.07	0.06	0.04	0.03	0.03	0.02	0.02	0.02	0.02
HEPES ( <i>N</i> -[2-hydroxyethyl]piperazine- <i>N</i> -2-ethanesulfonic acid), 10 mmol l <sup>-1</sup> , pH 7.6	2.45	2.50	2.37	2.08	1.50	0.29	0.03	<0.01		
MES (2-[ <i>N</i> -morpholino]ethanesulfonic acid), 10 mmol l <sup>-1</sup> (pH 6.0)	2.42	2.38	1.89	0.90	1.45	0.06	0.06	<0.01		
Potassium phosphate										
Monobasic, 10 mmol l <sup>-1</sup>	0.03	<0.01								
Dibasic, 10 mmol l <sup>-1</sup>	0.53	0.16	0.05	0.01	<0.01					
Sodium acetate, 10 mmol l <sup>-1</sup>	1.85	0.96	0.52	0.30	0.15	0.03	<0.01			
Sodium chloride 1 mmol l <sup>-1</sup>	2.00	1.67	0.40	0.10	<0.01					
Sodium citrate, 10 mmol l <sup>-1</sup>	2.48	2.84	2.31	2.02	1.49	0.54	0.12	0.03	0.02	0.01
Sodium formate, 10 mmol l <sup>-1</sup>	1.00	0.73	0.53	0.33	0.20	0.03	<0.01			
Sodium phosphate, 100 mmol l <sup>-1</sup> (pH 6.8)	1.99	0.75	0.19	0.06	0.02	0.01	0.01	0.01	0.01	<0.01
Tris-hydrochloric acid, 20 mmol l <sup>-1</sup>										
pH 7.0	1.40	0.77	0.28	0.10	0.04	<0.01				
pH 8.0	1.80	1.90	1.11	0.43	0.13	<0.01				
Brij (23 lauryl ether), 1%	0.06	0.03	0.02	0.02	0.02	0.01	<0.01			
CHAPS (3-[3-cholamidopropyl]-dimethylammonio]-1-propanesulfonate), 0.1%	2.40	2.32	1.48	0.80	0.40	0.08	0.04	0.02	0.02	0.01
SDS (sodium dodecyl sulfate), 0.1%	0.02	0.02	<0.01							
Triton X-100 (octoxynol), 0.1%	2.48	2.50	2.43	2.42	2.37	2.37	0.50	0.25	0.67	1.42
Tween 20 (polyoxyethylenesorbitan monolaurate), 0.1%	0.21	0.14	0.11	0.10	0.09	0.06	0.05	0.04	0.04	0.03

**Critical band method** The same development is adopted to determine the retention models for all solutes. Below every retention surface it is possible to determine a forbidden zone (critical band): any solute whose retention would fall into a critical band would interfere with the compound considered with a resolution value ( $R_S$ ) lower than a threshold value. The separation with  $R_S$  value higher than the threshold value is achieved at a composition for which none of the critical bands overlap. This is a graphical method for the optimization of a single parameter, such as the composition of a binary mobile phase or of a ternary mobile phase constituted of mixtures of two isoelutropic binary mobile phases.

**Overlapping resolution mapping** This method was developed by Glajch *et al.* for the optimization of quaternary mobile phase composition in LSC or in RPLC. The first step is to restrict the parameter space to isoelutropic solvents corresponding to the appropriate  $k$  value for the last-eluted solute. In such a way, the parameter space is defined by an equilateral triangle, the three apices of which correspond to binary mobile phases, and the three sides to ternary mobile phases. Any composition inside the triangle is a quaternary mobile phase composition. Seven experiments, corresponding to the composition of each apex, the composition of each half-side, and the composition of the center of the triangle, are carried

out. The surface retention of each solute is characterized by a quadratic model, the coefficients of which are determined by regression analysis. The response surface (threshold minimum resolution) is calculated by means of the retention models: the overlapping resolution mapping procedure consists of eliminating zones inside the triangle that correspond to  $R_s$  values lower than the threshold value and repeating the procedure for all pairs of solutes. The remaining zones inside the triangle will define the optimized compositions of the mobile phase. This method is relatively fast and simple, and may be automated. Nevertheless, more experiments may be required if a more accurate description of the response surface around the optimum is required.

### Iterative Methods

Iterative methods are interpretive methods in which new experiments can be added to the small number of starting experiments in order to refine the surface retention for all the solutes, and to predict more accurately the optimum location in the parameter space. For single-parameter optimization both graphical and mathematical methods can be used, but for multiparameter optimization computerized methods are necessary.

### Gradient Elution

In most cases, solute separations can be achieved in the isocratic mode by optimizing the mobile phase composition. Nevertheless, for some complex cases of solutes possessing very different polarities, separations can require a gradient elution mode: in such cases, mobile phase compositions and gradient slopes have to be determined.

For a chromatographic experiment carried out in gradient elution mode, the composition of the mobile phase is varied during the analysis. This is achieved by a gradual increase of the proportion of methanol, acetonitrile, or tetrahydrofuran in water for RPLC, or of diisopropyl ether, methylene chloride, or chloroform in *n*-hexane for LSC. During such a programmed analysis, solutes of a wide range of polarities can be eluted from the column under optimum conditions in a single experiment. According to the concept of linear solvent strength (LSS) gradients developed by Snyder, the shape of an LSS gradient is given by

$$\log k_m = \log k_a - b(t/t_0) \quad [17]$$

where  $k_m$  is the capacity factor of the solute under the isocratic conditions at the column inlet at time  $t$ ,  $k_a$  is the capacity factor of the solute at the initial composition of the gradient program,  $b$  is the

gradient steepness of the gradient,  $t$  is the time elapsed since the arrival of the starting gradient at the column inlet, and  $t_0$  is the column dead time. Optimized  $b$  values range from 0.2 to 0.4.

In LSC, according to eqns [8] and [11], an LSS gradient is concave. In RPLC, eqn [7] can be simplified into a linear relationship between  $\ln k$  and  $\varphi$ :

$$\ln k = \ln k_0 - S\varphi \quad [18]$$

with  $5 < S < 10$  for low molecular mass solutes in water-methanol mixtures. Consequently, in RPLC, an LSS gradient is given by a linear gradient such as

$$\varphi = A + Bt \quad [19]$$

Considering eqns [16]–[18], it appears that the optimum gradient slope  $B$  expressed in  $\text{min}^{-1}$  is given by

$$B = 2.303b/(St_0) \quad [20]$$

Initial conditions of gradient can be optimized by a trial-and-error approach in such a way that solutes are eluted neither too late nor too early during the analysis. Not only are many complex samples better separated in a gradient elution mode, but gradient runs can also be used to predict the relationship between retention of solutes and mobile phase composition in the isocratic mode.

Additional considerations in the choice of whether or not to use gradient mobile phases include:

- It requires an instrument capable of generating gradients.
- Because of differences in instrumentation, method transfer and validation can be more problematic than for isocratic methods.
- Some detection methods are not compatible.
- Mobile phases need to be sparged or degassed due to bubble formation when mixing aqueous and organic phases.
- Baseline shifts may be encountered due to UV absorbing modifiers.

### Buffered Mobile Phases

For the reversed phase separation of acidic or basic compounds, it is strongly recommended to include a buffer to control the pH of the mobile phase. Capacity, UV absorbance, volatility, solubility, stability, and sample and stationary phase interaction are important to consider when selecting a buffer. For reversed phase separation, a buffer concentration in the range of  $10\text{--}50 \text{ mmol l}^{-1}$  is usually sufficient, when used within  $\pm 1.0$  pH units of the  $\text{p}K_a$ . Stationary phases recently introduced to the market easily withstand exceedingly higher pH, allowing

the chromatographer to use pH as a true selectivity tool.

**See also: Chemometrics and Statistics:** Experimental Design; Optimization Strategies.

## Further Reading

- Deyl Z, Macek K, and Janak J (1975) *Liquid Column Chromatography*. Amsterdam: Elsevier Scientific.
- Heinisch S, Riviere P, and Rocca J-L (1991) Computer routine for the selection and the optimization of multisolvent mobile phase systems in reversed phase liquid chromatography. *Chromatographia* 32: 559–565.

- Hildebrand JH, Prausnitz JM, and Scott RL (1970) *Regular and Related Solutions*. New York: Van Nostrand Reinhold.
- Miller JM (1988) *Chromatography: Concepts and Contrasts*. New York: Wiley Interscience.
- Neue UD (1997) *HPLC Columns, Theory, Technology, and Practice*. New York: Wiley-VCH.
- Schoenmakers PJ (1986) *Optimization Chromatographic Selectivity*. Amsterdam: Elsevier.
- Snyder LR and Kirkland JJ (1979) *An Introduction to Modern Liquid Chromatography*. New York: Wiley Interscience.
- Snyder LR, Kirkland JJ, and Glajch JL (1997) *Practical HPLC Method Development*, 2nd edn. New York: Wiley Interscience.

## Normal Phase

**P Jandera**, Univerzita Pardubice, Pardubice, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

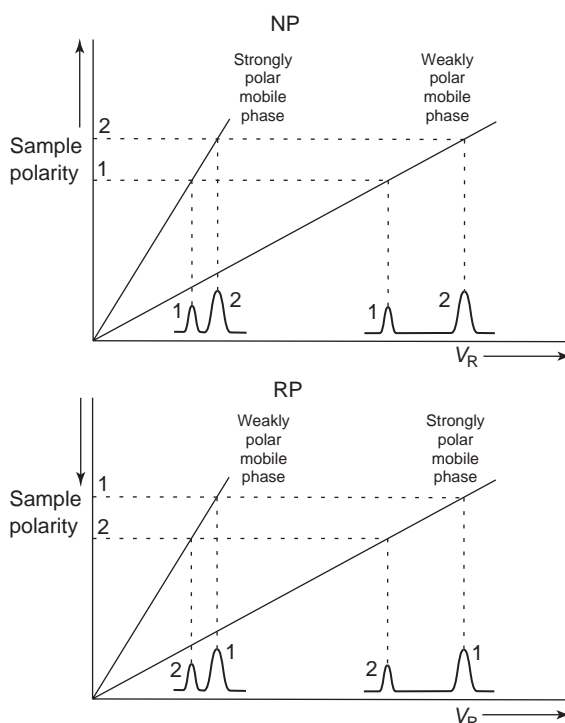
## Introduction

Normal-phase liquid chromatography (NPLC) is the oldest chromatographic mode, discovered by M.S. Tswett more than 100 years ago. It has been the predominant mode in thin-layer chromatography (TLC) and low-pressure dry-column liquid chromatography before the introduction of reversed-phase technique, which is a preferred mode in high-performance liquid chromatography (HPLC).

In normal-phase chromatography, the stationary phase is more polar than the mobile phase. As illustrated in **Figure 1**, the retention increases as the polarity of the mobile phase decreases and thus polar analytes are more strongly retained than nonpolar analytes, the opposite of that in reversed-phase liquid chromatography (RPLC). The column packing is either an inorganic adsorbent (silica gel or, less often, aluminum, titanium, or zirconium oxides) or a moderately polar bonded phase (cyanopropyl,  $-(CH_2)_3-CN$ ; diol,  $-(CH_2)_3-O-CH_2-CHOH-CH_2-OH$ ; or aminopropyl,  $-(CH_2)_3-NH_2$ ) chemically bonded on a suitable support material, which is usually silica gel. The mobile phase is usually a mixture of two or more organic solvents: nonpolar solvent and strongly or weakly polar solvents.

Moderately polar bonded phases may be used either in the normal-phase mode with organic mobile phases, such as *n*-hexane + 2-propanol, or in the reversed-phase mode with aqueous–organic mobile phases (**Figure 2**). However, normal-phase behavior can also be sometimes observed in nonaqueous

RPLC with chemically bonded C18, C8, or other alkyl-bonded phases. The reason is the activity of polar residual silanol groups that remain on the support surface after incomplete reaction of silica gel with organosilanes. Thus, such a stationary phase may behave as a deactivated polar adsorbent in nonpolar or weakly polar organic solvents.



**Figure 1** Schematic diagram of the effects of sample and mobile phase polarities on the retention in normal-phase (NP) and reversed-phase (RP) liquid chromatography.  $V_R$  – retention volume.

the chromatographer to use pH as a true selectivity tool.

**See also: Chemometrics and Statistics:** Experimental Design; Optimization Strategies.

## Further Reading

- Deyl Z, Macek K, and Janak J (1975) *Liquid Column Chromatography*. Amsterdam: Elsevier Scientific.
- Heinisch S, Riviere P, and Rocca J-L (1991) Computer routine for the selection and the optimization of multisolvent mobile phase systems in reversed phase liquid chromatography. *Chromatographia* 32: 559–565.

Hildebrand JH, Prausnitz JM, and Scott RL (1970) *Regular and Related Solutions*. New York: Van Nostrand Reinhold.

Miller JM (1988) *Chromatography: Concepts and Contrasts*. New York: Wiley Interscience.

Neue UD (1997) *HPLC Columns, Theory, Technology, and Practice*. New York: Wiley-VCH.

Schoenmakers PJ (1986) *Optimization Chromatographic Selectivity*. Amsterdam: Elsevier.

Snyder LR and Kirkland JJ (1979) *An Introduction to Modern Liquid Chromatography*. New York: Wiley Interscience.

Snyder LR, Kirkland JJ, and Glajch JL (1997) *Practical HPLC Method Development*, 2nd edn. New York: Wiley Interscience.

## Normal Phase

**P Jandera**, Univerzita Pardubice, Pardubice, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

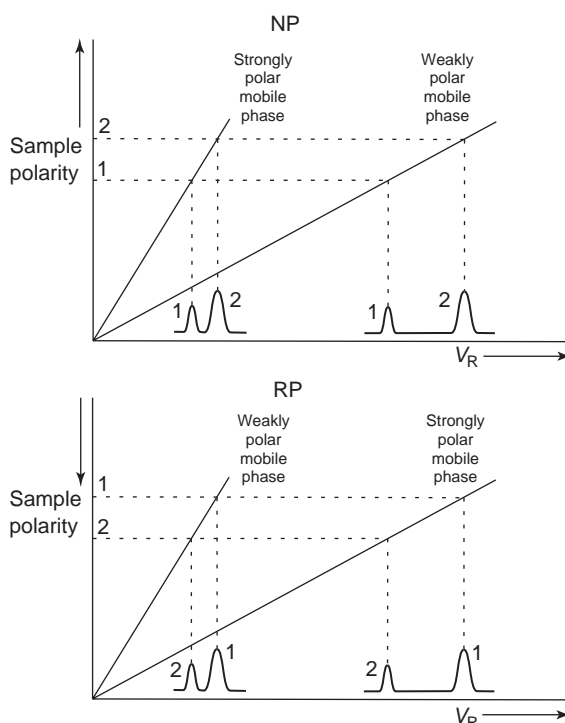
## Introduction

Normal-phase liquid chromatography (NPLC) is the oldest chromatographic mode, discovered by M.S. Tswett more than 100 years ago. It has been the predominant mode in thin-layer chromatography (TLC) and low-pressure dry-column liquid chromatography before the introduction of reversed-phase technique, which is a preferred mode in high-performance liquid chromatography (HPLC).

In normal-phase chromatography, the stationary phase is more polar than the mobile phase. As illustrated in **Figure 1**, the retention increases as the polarity of the mobile phase decreases and thus polar analytes are more strongly retained than nonpolar analytes, the opposite of that in reversed-phase liquid chromatography (RPLC). The column packing is either an inorganic adsorbent (silica gel or, less often, aluminum, titanium, or zirconium oxides) or a moderately polar bonded phase (cyanopropyl,  $-(\text{CH}_2)_3\text{-CN}$ ; diol,  $-(\text{CH}_2)_3\text{-O-CH}_2\text{-CHOH-CH}_2\text{-OH}$ ; or aminopropyl,  $-(\text{CH}_2)_3\text{-NH}_2$ ) chemically bonded on a suitable support material, which is usually silica gel. The mobile phase is usually a mixture of two or more organic solvents: nonpolar solvent and strongly or weakly polar solvents.

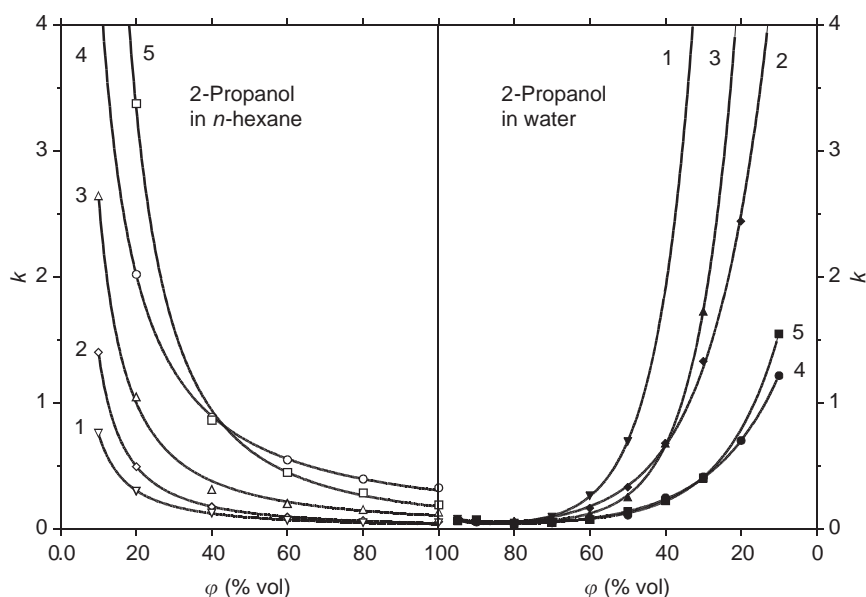
Moderately polar bonded phases may be used either in the normal-phase mode with organic mobile phases, such as *n*-hexane + 2-propanol, or in the reversed-phase mode with aqueous–organic mobile phases (**Figure 2**). However, normal-phase behavior can also be sometimes observed in nonaqueous

RPLC with chemically bonded C18, C8, or other alkyl-bonded phases. The reason is the activity of polar residual silanol groups that remain on the support surface after incomplete reaction of silica gel with organosilanes. Thus, such a stationary phase may behave as a deactivated polar adsorbent in nonpolar or weakly polar organic solvents.



**Figure 1** Schematic diagram of the effects of sample and mobile phase polarities on the retention in normal-phase (NP) and reversed-phase (RP) liquid chromatography.  $V_R$  – retention volume.





**Figure 2** Effects of the concentration of 2-propanol,  $\phi$ , in the mobile phase on the retention factors,  $k$ , of phenylurea herbicides on a bonded nitrile column in normal-phase (2-propanol/hexane) and reversed-phase (2-propanol-water) modes. Column: Silasorb Nitrile, 7.5  $\mu\text{m}$ , 300  $\times$  4.2 mm ID. Solutes: bis-*N,N*-(3-chloro-4-methylphenyl)urea (1); 2-*N*-butyl-*N'*-phenylurea (2); isoproturon (3); *N*-phenylurea (4); hydroxymethoxuron (5). (Adapted with permission from Fischer J and Jandera P (1994) *Journal of Chromatography A* 684: 77–92; © Elsevier.)

As the retention on inorganic adsorbents originates due to the interactions of the polar adsorption centers on the surface with the polar functional groups of the analytes, this mode was also called adsorption or liquid–solid chromatography.

## Practical Considerations

In the reversed-phase mode, higher separation selectivity than in normal-phase chromatography can be usually achieved for compounds with even minor differences in the size of the molecules. This is the main reason for the predominant role of reversed-phase separations in contemporary HPLC. On the other hand, NPLC usually offers significantly better separation of isomers than RPLC.

Normal-phase chromatography has several practical advantages: (1) Because of a lower organic mobile phase viscosity, the pressure drop across the column is lower than with aqueous–organic mobile phases commonly used in RPLC at a comparable flow rate. (2) HPLC columns are usually more stable and have longer lifetimes in organic solvents than in aqueous–organic mobile phases. (3) Columns packed with nonmodified inorganic adsorbents are not subject to ‘bleeding’, i.e., gradual loss of the chemically bonded stationary phase originating from hydrolysis, which slowly decreases the retention during the lifetime of a chemically bonded column, especially in aqueous mobile phases and at elevated temperatures.

(4) Many samples are more soluble or less prone to decompose in organic than in aqueous mobile phases. Such samples do not cause direct-injection problems in NPLC, unlike to RPLC. (5) The recovery of organic-soluble sample components is easy in preparative NPLC. (6) Very large changes in separation selectivity are possible by changing either the mobile phase or the stationary phases in NPLC.

On the other hand, reversed-phase chromatography generally offers better selectivity than NPLC for the separation of molecules differing in carbon number. Further, ionic samples are usually more easily separated in RPLC. The purchase and disposal of organic solvents are more expensive in NPLC than in RPLC, and lower-boiling solvents are more prone to evaporation and bubble formation, especially at elevated temperatures. Finally, controlling the retention by adjusting the solvent strength can be less predictable and reproducible in NPLC than in RPLC, because of significant preferential adsorption of polar solvents, especially water, on polar adsorbents.

## Principles of Retention in NPLC

In RPLC, the retained solute molecules are not localized on the surface of the column packing (this behavior resembles a partition process), whereas the adsorbed polar sample and solvent molecules in NPLC are localized on specific polar adsorption sites, to which they are strongly attracted. Neutral

compounds can be separated by both RPLC and NPLC, generally with the opposite order of elution, but some differences in the separation selectivity and elution order are usually observed between the two modes. This is illustrated in **Figure 2** for the retention factors,  $k$ , of some phenylurea herbicides on a bonded nitrile column in propanol–hexane (normal-phase mode) and propanol–water (reversed-phase mode) mobile phases.

The retention model in adsorption chromatography developed by Snyder and Soczewinski is based on the assumption that there is flat adsorption in a monomolecular layer on a homogeneous adsorption surface. The adsorption is understood as a competition phenomenon between the molecules of the solute and the solvent on the adsorbent surface, so that the retention of a sample molecule requires the displacement of one or more previously adsorbed polar solvent molecules. Later, the model was corrected for adsorption on a heterogeneous surface of adsorbent. To first approximation, the solute–solvent interactions in the mobile and stationary phases are assumed to compensate each other and possible liquid–liquid partition effects are neglected. In this case, the retention in a mixed binary mobile phase comprising a nonpolar solvent, A (usually an aliphatic hydrocarbon), and a polar solvent, B, can be described by eqn [1]:

$$\log k = \log k_a + \alpha' A_s (\epsilon_a - \epsilon_{ab}) \quad [1]$$

Here,  $k$  is the retention factor in a binary mobile phase containing solvents A and B, which can be calculated from the retention volume of the analyte,  $V_R$ , and the column hold-up volume,  $V_M$  (determined as the elution volume of a nonretained compound, such as trichloroethylene),  $k = V_R/V_M - 1$ ,  $k_a$  is  $k$  in pure weak solvent, A,  $\alpha'$  is the activity of the adsorbent in the column,  $A_s$  is the specific surface of the adsorbent, and  $\epsilon_a$  and  $\epsilon_{ab}$  are the solvent strength (polarity) parameters of the weak (nonpolar) solvent and of the mixed binary mobile phase, respectively.

The retention in NPLC increases with increasing polarity (activity) and is proportional to the specific surface area,  $A_s$ , which controls the number of the adsorption sites available for contact with the sample solutes in the column. Water in mobile phases deactivates the adsorbent (decreases  $\alpha'$ ), as the sites with strongly held adsorbed water are no longer available for adsorption of moderately polar organic compounds. The elution times of analytes generally increase in the following sequence: alkanes < alkenes < aromatic hydrocarbons  $\approx$  chloroalkanes < alkylsulphides < ethers < ketones  $\approx$  aldehydes  $\approx$  esters < alcohols < amides < phenols, amines, carboxylic acids. The

retention also depends to some extent on the hydrocarbon part of solutes and generally slightly decreases as the size of alkyl groups increases. However, the adsorption energy of alkyl groups is low and the adsorbed molecules adhere to the adsorbent surface by the polar groups, whereas the hydrocarbon parts of the molecules point from the adsorbent surface toward the mobile phase. The separation in a homologous series is far less effective than in reversed-phase chromatography and usually only a few lowest homologs can be successfully resolved by NPLC. If the localization of the adsorption sites on the surface of a polar adsorbent fits a spatial distribution of polar functional groups in a solute molecule with multiple functional groups, simultaneous interactions of two or more functional groups are possible, which are weaker or absent for molecules with other positions of functional groups. Hence, NPLC (especially on silica gel columns) is usually suitable for separation of positional isomers. Further, differences in the retention of molecules of similar polarities, but different shapes (rigid planar, rod-like, or of flexible chain structure), are often observed and utilized in normal-phase chromatography. Both steric hindrance to the access of a polar functional group to the adsorption site and interactions between vicinal polar groups decrease the adsorption and the retention. Hence, *ortho*-substituted benzenes are less strongly retained than *meta*- and *para*-isomers. For example, *o*-nitrophenol elutes earlier than *m*-nitrophenol, whereas *p*-nitrophenol is retained most strongly in NPLC because of increased acidity caused by mesomeric effect.

The retention decreases at increased temperature, but high temperatures are rarely used in NPLC because of relatively low boiling points of polar organic solvents. Usually, a change in temperature has only minor effects on the selectivity and band spacing in NPLC, except for mobile phases containing localizing solvents (e.g., acetonitrile), which are desorbed at elevated temperatures.

### Columns and Stationary Phases in NPLC

Nowadays, most HPLC separations are performed on conventional analytical columns, 10–25 cm long, 3–4.6 mm in diameter, and packed with 3–10  $\mu$ m uniform particles (recently even with 1–2  $\mu$ m size). With short ‘high-speed’ columns, 1.5–5 cm long, simple separations can be accomplished in 1–3 min, with significantly reduced analysis time and solvent consumption. Separations on ‘microbore’ columns, 15–25 cm long and 1–2 mm ID, need even less mobile phase and allow detection of high mass sensitivity. These columns are useful for analyses of small

sample amounts and with mass-spectrometric detection. Recently, packed capillary HPLC columns of internal diameter 0.1–0.5 mm have become commercially available. However, the microbore and especially the capillary columns need miniaturized instrumentation to suppress extracolumn contributions of the injector, the detector, and the connecting capillaries to band broadening.

NPLC stationary phases include metal oxides and moderately or strongly polar chemically bonded phases. Unmodified silica gel and silica-based bonded phases are most frequently used nowadays. Considerable effort in the development of new HPLC column packing materials in the past years has resulted in significant improvement of the column efficiency, reproducibility, and increased stability at elevated temperatures and at higher pH, enabling better compatibility with HPLC/mass spectrometry techniques and rapid analyses. Even though the new column technologies were primarily focused on RPLC separations, normal-phase HPLC also benefits from the improved properties of the support materials with uniform small particles and well-defined pore size.

Relatively inexpensive irregular shape materials are still used extensively for preparative separations. However, they are being gradually replaced by spherical particles in analytical HPLC, since they form more homogeneous, stable, and permeable column beds, requiring lower working pressures. Because of superior performance of spherical microparticles with diameters less than 5  $\mu\text{m}$ , irregular materials of this size are only rarely used, if at all. The porosity of particles suitable for packing HPLC columns depends on the size of the separated molecules. Totally porous particles with pore size 7–12 nm and specific surface area 150–400  $\text{m}^2 \text{g}^{-1}$  are suitable for separations of small molecules, but wide-pore particles with pore size 15–100 nm and relatively low specific surface area (10–150  $\text{m}^2 \text{g}^{-1}$ ) are required for HPLC of macromolecules to allow easy access to the interactive surface within the pores.

**Silica gel** Macroporous silica gel is the material most often used in NPLC and is also the most important support material for preparing chemically bonded phase packings with a large variety of functionalities. The main advantages of silica particles are their high mechanical strength, compatibility with water and all organic solvents, and dimensional stability of column beds during use with various solvent types and gradient elution, as the wet silica particles do not swell.

Noncrystalline silica has the general chemical formula  $\text{SiO}_2 \cdot x\text{H}_2\text{O}$ , in which  $x$  depends on the degree

of hydration. The unmodified hydrated silica contains a surface layer of  $-\text{SiOH}$  (silanol) groups with a maximum concentration of  $\sim 8 \mu\text{mol m}^{-2}$ . The silanol groups are free, vicinal (on two vicinal silicon atoms), or geminal (two  $-\text{SiOH}$  groups borne by the same silicon atom). Vicinal silanols may be associated by hydrogen bonds. In addition, siloxane bonds ( $-\text{Si}-\text{O}-\text{Si}-$ ) may be present on the silica surface. Silanol groups may be associated with water molecules by hydrogen bonding. By heating to 115–200°C, the silica surface is first dehydrated by removing the adsorbed water. Further heating up to 400°C causes associated silanols to reversibly condense to siloxanes and at higher temperatures, irreversible siloxane formation occurs connected with lattice rearrangement and sintering.

An unfavorable property of silica is its solubility at high pH, which, however, is rarely applied with the mobile phases used in NPLC. Silica gel has weak cation-exchange properties (the silanols on the silica gel surface may be ionized to  $\text{SiO}^-$  groups), whereas the bonded amine stationary phases behave as weak anion exchangers. This mixed retention mechanism may cause excessive retention and tailing or even split peaks of ionic or ionizable compounds.

Several technologies are used for preparation of silica gel for HPLC. ‘Sol-gel’ silica with irregular particles can be prepared by acidifying solutions of silicates (type A silicas) or by hydrolysis of tetraalkylsilanes and precipitation of a hydrosol, followed by aging, which results in gelling of the colloidal particles to form larger aggregates connected by siloxane bonds. After washing free of the residual sodium salts and crushing to the desired particle size, a hard, porous xerogel is produced. Spherical microparticles can be prepared by emulsification of a silica sol in a water-immiscible organic liquid. Type A silica gels are more or less acidic because of contamination with certain metals, which activate surface silanol groups and can complex with some chelating solutes, causing strong retention or asymmetrical peaks.

Newer, highly purified, less-acidic ‘sol-gel’ spherical silica particles (type B silicas) are formed by the aggregation of silica sols exposed to air, contain very low amounts of metals and are more stable at intermediate and higher pH than the xerogel-type materials, to at least pH 9. They generally provide better separations, especially of basic samples. Recently, silica gel with a surface populated with nonpolar silicon hydride  $\text{Si}-\text{H}$  groups instead of silanol groups (type C) has been introduced for preparation of chemically bonded phases bonded via  $\text{Si}-\text{C}$  instead of  $\text{Si}-\text{O}-\text{C}$  siloxane bonds, providing improved stability at elevated temperatures and low bleed of the

organic-bonded phase. These materials are less polar than the materials with silanol groups and can be used for separation of acids or bases in the normal-phase mode in buffered mobile phases containing ~50–70% organic solvent (acetonitrile).

**Polar-bonded phases** Cyano, diol, and amino bonded phases used in NPLC are generally prepared by chemical modification of the silica gel surface, like the C18 or C8 phases for RPLC. NPLC separations on these columns are usually more convenient and reproducible than on unmodified silica, as they are less sensitive to the control of mobile-phase water content and require shorter equilibration times after the change of the mobile phase. However, silica columns usually have a better stability and a longer lifetime, show better isomer selectivity, higher loadability for preparative separations, and are less expensive. Generally, the strength of the interactions with analytes increases in the order: cyanopropyl < diol < aminopropyl < silica  $\approx$  alumina stationary phases, but selective interactions may change this order. Basic analytes are generally very strongly retained by the silanol groups in silica gel and acidic compounds show increased affinities to aminopropyl columns. Aminopropyl and diol columns prefer compounds with proton-acceptor or proton-donor functional groups (alcohols, esters, ethers, ketones, etc.), while dipolar compounds that cannot have proton-donor or proton-acceptor interactions are usually more strongly retained on cyanopropyl than on aminopropyl silica. Chemically bonded phases with other functionalities such as polyethylene glycol, pentafluorobenzene, or alkyls with embedded amide or carbamate groups are generally intended for reversed-phase applications in water-rich mobile phases. However, at high percentage of organic solvent(s) in the mobile phases, the retention of many compounds increases with increasing concentration of acetonitrile or methanol, showing a typical normal-phase behavior.

**Alumina** Several types of porous alumina are available with pore diameters from 6 to 15 nm and surface areas of 70–250 m<sup>2</sup> g<sup>-1</sup>. Its surface is less homogeneous than silica and contains both hydroxyl groups and aluminum ions with a function of Lewis acid sites. By conditioning with acids or bases, the pH of the surface can be adjusted between pH 3 and 9 and can exhibit both anion-exchange (at low pH) and cation-exchange (at high pH) properties. Heating at 200°C activates the surface of alumina by dehydration; heating at temperatures higher than 900°C causes irreversible loss of activity. Even though the adsorption properties of alumina are

similar to those of silica, there are some differences in selectivity. Alumina favors interactions with  $\pi$  electrons and often yields excellent separation selectivity for compounds with different numbers or spacing of unsaturated (double) bonds, such as polycyclic aromatic hydrocarbons or some steroids. However, it is much less widely used than silica nowadays, because of catalytic properties of its surface, which may cause decomposition or irreversible adsorption of some samples. It is more difficult to bond stationary phases to the surface of alumina than to silica gel, but alumina-based stationary phases are more stable in basic mobile phases.

**Zirconium dioxide** Zirconium dioxide (zirconia) has a polar surface with zirconium atoms as strong Lewis acid sites. Because of empty orbitals, Zr attracts electron-rich Lewis bases such as phosphate anions. Hence, the surface of zirconia has significant cation-exchange properties, giving rise to a mixed-mode retention mechanism. Recently, 3 and 5  $\mu$ m porous and nonporous spherical zirconia particles have become available for HPLC, with 30 nm pore size and 30 m<sup>2</sup> g<sup>-1</sup> specific surface. The primary advantage of zirconia relative to silica is its stability over the entire pH range from 1 to 14 and at temperatures up to 200°C. Like silica, it is mechanically stable and provides high chromatographic efficiency. Stationary phases based on zirconia surface coated with polybutadiene, polystyrene, carbon, or carbon covalently modified with octadecyl groups are intended mainly for RPLC applications. Even though unmodified zirconia is a much weaker adsorbent than silica in nonaqueous solvents, its NPLC applications in the analysis of steroids and plant genins were reported.

### Mobile Phase in Normal-Phase Chromatography

The polarity and the elution strength, i.e., the ability to enhance the elution, generally increases in the following order of most common normal-phase chromatography solvents: hexane  $\approx$  heptane  $\approx$  octane < methylene chloride < methyl-*t*-butyl ether < ethyl acetate < dioxane < acetonitrile  $\approx$  tetrahydrofuran < 1- or 2-propanol < methanol. Large changes in selectivity of normal-phase chromatographic separations can be achieved by selecting the solvent with appropriate type of polar interactions. For chromatography on silica gel, normal-phase chromatographic solvents can be classified as nonlocalizing (e.g., alkanes, aromatic hydrocarbons, chloroalkanes), basic-localizing (e.g., amines and ethers), and nonbasic-localizing (e.g., esters, nitriles, or nitro compounds). Localizing solvents are strongly attracted to adsorption sites,

while nonlocalizing solvents are more or less regularly distributed on the whole adsorbent surface. The basicity of a solvent is understood as its ability of hydrogen-acceptor interactions with the silanol groups.

The selection of the polar solvent **B** strongly affects the separation selectivity in NPLC. A change from a nonlocalizing to a localizing solvent causes large changes in mobile-phase selectivity. Significant differences in selectivity occur also between basic and nonbasic localizing solvents. Changing the type of the solvent **B** may even cause a reversal in the order of elution of sample compounds. **Table 1** lists characteristic properties of some useful solvents for NPLC mobile phases.

**Binary mobile phases** A single solvent only rarely provides suitable separation selectivity and retention in normal-phase systems, which should be adjusted by selecting appropriate composition of a two- or multicomponent mobile phase. Most often, a binary mobile phase is used, containing a weak solvent **A**, usually an alkane, and a more polar organic solvent, **B**, whose concentration controls the elution strength of a mixed mobile phase. Change in retention is a regular function of the concentration of **B** in the mobile phase. As a rule, a twofold increase in percentage of **B** will cause a two- to threefold decrease in the retention factors, *k*, depending on the polarity of the strong solvent **B**. The dependence of the retention in NPLC on the composition of the mobile

**Table 1** Properties of some solvents used in chromatography

<i>Solvent</i>	<i>Relative molar mass</i>	<i>Boiling point (°C)</i>	<i>Viscosity at 20° C (cP)</i>	<i>Refractive index at 25° C</i>	<i>UV cut-off (nm)</i>	<i>Polarity<sup>a</sup></i>	<i>ε (Al<sub>2</sub>O<sub>3</sub>)<sup>b</sup></i>
Isooctane	114	99	0.50	1.391	200	− 0.4	0.01
Decane	142	174	0.92	1.412	200	− 0.3	0.04
Heptane	100	98	0.42	1.388	200	0.0	0.01
Cyclohexane	84	81	0.98	1.426	210	0.0	0.04
Hexane	86	69	0.33	1.375	200	0.0	0.01
Pentane	72	36	0.23	1.358	200	0.0	0.00
Tetrachloromethane	154	76	0.97	1.466	265	1.7	0.18
Dibutyl ether	130	143		1.400		1.7	
Di-isopropyl ether	102	67.8	0.37	1.368	220	2.2	0.28
Toluene	92	111	0.59	1.496	285	2.3	0.29
<i>p</i> -Xylene	106	139	0.62	1.500	290	2.4	0.26
Chlorobenzene	112	132	0.80	1.525	280	2.7	0.30
Diethyl ether	74	35	0.23	1.353	202	2.9	0.38
Benzene	78	80	0.65	1.501	280	3.0	0.32
Fluorobenzene	96	85		1.468		3.3	
Dichloromethane	85	40	0.44	1.424	233	3.4	0.42
2-Pentanol	88		4.10	1.410	210	3.6	0.61
1,2-Dichloroethane	97	84	0.79	1.445	230	3.7	0.49
1-Butanol	74	117		1.400		3.9	
1-Propanol	60	97	2.30	1.380	207	4.1	0.82
Tetrahydrofuran	72	66	0.51	1.407	230	4.2	0.45
Ethyl acetate	88	77	0.45	1.370	260	4.3	0.58
2-Propanol	60	82	2.30	1.380	207	4.3	0.82
Chloroform	119	61	0.57	1.443	245	4.4	0.40
Methyl ethyl ketone	72	80	0.40	1.379	330	4.5	0.51
Dioxane	88	101	1.54	1.422	215	4.8	0.56
Diethylene glycol	106	245		1.449		5.0	
Ethanol	46	78	1.20	1.361	205	5.2	0.88
Acetone	58	56	0.32	1.359	330	5.4	0.56
Ethylene glycol	62	198	2.10	1.432	210	5.4	1.11
Methoxyethanol	76	125		1.402		5.7	
Methyl formamide	59	180				6.2	
Acetonitrile	41	82	0.37	1.344	212	6.2	0.65
<i>N,N</i> -Dimethylacetamide	87	165		1.435		6.3	
Dimethylformamide	73	155	0.90	1.428	310	6.4	
Methanol	32	65	0.60	1.329	208	6.6	0.95
Formamide	45	210	3.76	1.448	210	7.3	
Water	18	100	1.00	1.333	180	9.0	

<sup>a</sup> Solvent polarity parameters derived from Rohrschneider data.

<sup>b</sup> Solvent strength parameter for NPLC on alumina.



phase can be characterized more exactly on the basis of the simple adsorption model described by eqn [1]. With some simplification, the dependence of the retention factor,  $k$ , on the concentration (volume fraction) of the stronger (more polar) solvent,  $\varphi$ , in binary mobile phases comprising two solvents of different polarities can be described by eqn [2]:

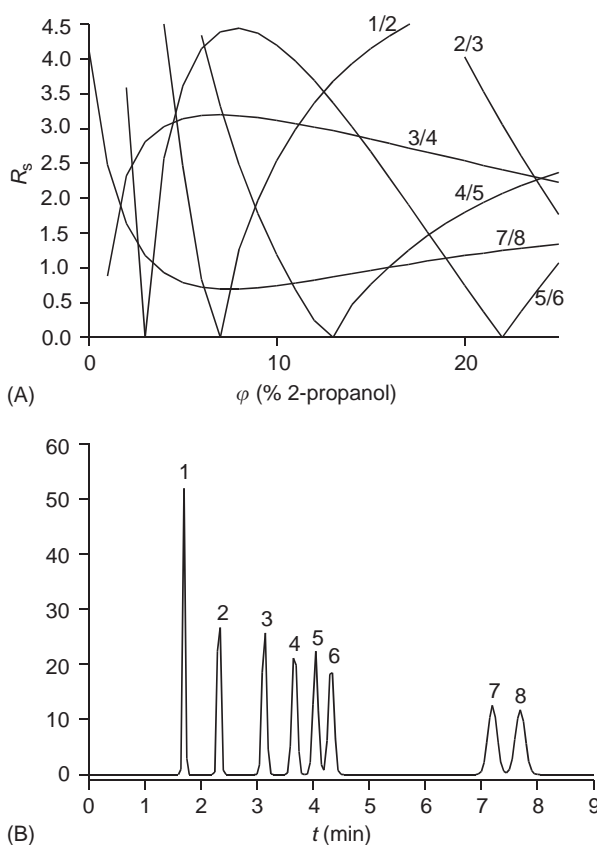
$$k = k_0 \varphi^{-m} \quad [2]$$

Here,  $k_0$  and  $m$  are experimental constants, which depend on the nature of the solute and on the chromatographic system, but are independent of the concentration of the strong solvent,  $\varphi$ , in the mobile phase.  $k_0$  is the retention factor in pure strong (polar) solvent, B. The parameter  $m$  is a measure of stoichiometric exchange in the displacement model of adsorption, i.e., the number of molecules of the strong solvent B necessary to displace one adsorbed sample molecule. Equation [2] applies only for samples strongly retained in the pure nonpolar solvent, A. If the analytes are less strongly retained in pure solvent A, a three-parameter equation [3] can often describe the effect of changing the mobile phase composition on the retention:

$$k = (b + a\varphi)^{-m} \quad [3]$$

The constants  $a$ ,  $b$ , and  $m$  in eqn [3] depend on the solute and on the chromatographic system.  $b = (k_a)^{-m}$ , where  $k_a$  is the retention factor in a pure nonpolar solvent. Equation [2] or [3] can be used as the basis of optimization of the composition of two-component (binary) mobile phases in NPLC, using a common window diagram or overlapping resolution mapping approach, as illustrated in an example in Figure 3.

**Complex mobile phases** Complex mobile phases containing two or more different polar solvents in a nonpolar one can be used to enhance the selectivity in NPLC. Fine tuning of NPLC separations can be based on the 'selectivity triangle' approach, selecting the optimum composition of four-component mobile phases with hexane or heptane as a nonpolar solvent, dichloromethane as a nonlocalizing solvent, methyl *tert*-butyl ether as a basic localizing solvent, and acetonitrile or ethyl acetate as a nonbasic localizing solvent. The concentration of the nonpolar solvent (a diluter) controls the solvent strength and the concentration ratios of the three polar solvents adjust the selectivity for various sample components. To describe the retention in ternary and more complex mixtures, it is principally possible to use the theoretical model of adsorption chromatography, with elution strength contributed to by all solvents in the



**Figure 3** Resolution diagram for the isocratic separation of eight herbicides on a Silasorb SPH silica gel column, 7.5  $\mu\text{m}$ , 300  $\times$  4.2 mm ID as a function of the concentration,  $\varphi$ , of 2-propanol in heptane as the mobile phase (A) and the separation under optimized conditions yielding best resolution of the pairs of compounds 5/6 and 7/8 determined from the window diagram A as 19% 2-propanol (B). Solutes: neburon (1); chlorobromuron (2); 3-chloro-4-methylphenylurea (3); desphenuron (4); isoproturon (5); diuron (6); methoxuron (7); deschloromethoxuron (8). (Reprinted with permission from Jandera P (1998) *Journal of Chromatography A* 797: 11–22; © Elsevier.)

mobile phase, but it would be necessary to consider the competition between various solvents in the mobile phase for localized adsorption centers on the adsorbent surface and to correct correspondingly the solvent strength, which is not straightforward in practice.

**Uptake of water and polar solvents by the column** From mixed mobile phases, polar solvent(s) are preferentially adsorbed on the surface of polar adsorbents, sometimes giving rise to multilayer solvent adsorption on the adsorbent support. In such a case, the retention is contributed to by a liquid–liquid partition mechanism between the adsorbed liquid layer and the bulk mobile phase, in addition to the adsorption. Such a mixed-mode mechanism can be intentionally utilized for separation of strongly polar or even ionic compounds.

The uptake of the polar solvents by the column increases with the polarity of both the solvent and the adsorbent and this effect is especially strong with silica or alumina adsorbents and mobile phases containing minor concentrations of water. Because of the ambient humidity, any organic solvent contains a certain quantity of water if it is in contact with the surrounding atmosphere. This dissolved water is extracted from the mobile phase by the column. As the water content of the column increases, the activity of the adsorbent decreases (eqn [1]), and sample retention times can be greatly reduced compared to a dry column. Ambient humidity seldom remains constant over a longer period of time and as the equilibration of the column can be a very long process, column water content can vary from run to run. Consequently, the reproducibility of the retention data is impaired and long equilibration times are required when the separation conditions are changed. For reproducible results in NPLC, it is necessary to keep a constant adsorbent activity, which can be accomplished by intentional preequilibration of the mobile phase components with water (in the concentration range of a few parts per million). Unfortunately, using 'isohydric' organic solvents with equilibrium water concentrations is time consuming and not very practical. Using mobile phases containing at least 0.1–0.5% polar organic solvent, such as propanol, which mimics the effects of added water, is much simpler to use and often solves the problem.

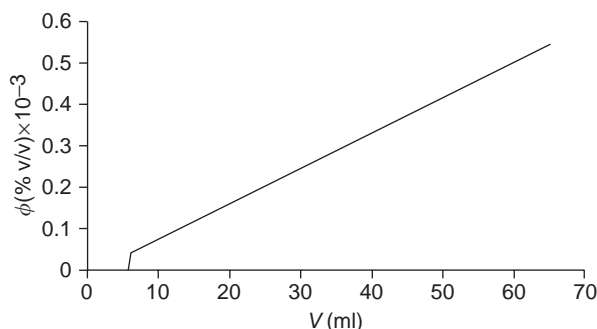
The reproducibility in normal-phase chromatography can be significantly improved and the column reequilibration times can be short if the temperature is controlled to  $\pm 0.1^\circ\text{C}$  during the separation and dry mobile phases with nonlocalizing polar solvents being used, such as dichloromethane, dioxane, or *tert*-butyl methyl ether. Dehydrated solvents can be kept dry over activated molecular sieves and filtered just before the use to improve the reproducibility. These measures can result in reproducible retention data over a long period of column use, even when using gradient elution technique.

**Gradient elution** During gradient-elution chromatography in normal-phase systems the concentration of one or more polar solvents in a nonpolar solvent is increased. Mixtures of analytes whose polarities span over a large polarity range (such as samples containing fatty acid esters, sterols, and sugars) can be separated in a single run using the so-called incremental gradient elution (or relay gradient elution) with a series of consecutive linear binary gradient steps employing the final solvent **B** from the previous step as the starting solvent **A** in the next step, e.g., a four-solvent three-step gradient elution

starting from hexane to ethyl acetate, continuing from ethyl acetate to acetonitrile, and finishing from acetonitrile to water. Reequilibration of the column between the consecutive runs in 'incremental gradient elution' is tedious and time consuming, hence a single-step gradient elution technique with two miscible solvents such as hexane and 2-propanol is more practical. However, if the two solvents largely differ in polarities, the strong solvent may be completely removed from the mobile phase by adsorption at the first stage of the gradient run, until the breakthrough of the polar solvent occurs, as illustrated in **Figure 4**. This behavior is similar to the situation in TLC or in dry-column liquid chromatography techniques, where it has been known for a long time as the 'solvent-demixing' effect. The extent of the polar solvent uptake by the column during gradient elution depends on the type of the strong solvent B, on the stationary phase, and on the gradient program. This effect may give rise to irreproducible retention, which was the main reason for a strong bias against using gradient elution in normal-phase chromatography.

The preferential uptake of polar solvents by the column can be largely eliminated when starting the gradients at an initial concentration of the polar solvent higher than 3%. If this is not possible, the effects of the preferential adsorption of the polar solvent on the retention can be predicted by numerical calculation from the parameters of the adsorption isotherm of the polar solvent on the column packing material (usually the Langmuir type).

**Normal-phase separation of hydrophilic and ionic compounds** Normal-phase chromatography can sometimes provide satisfactory resolution of hydrophilic and even ionic samples. Very hydrophilic

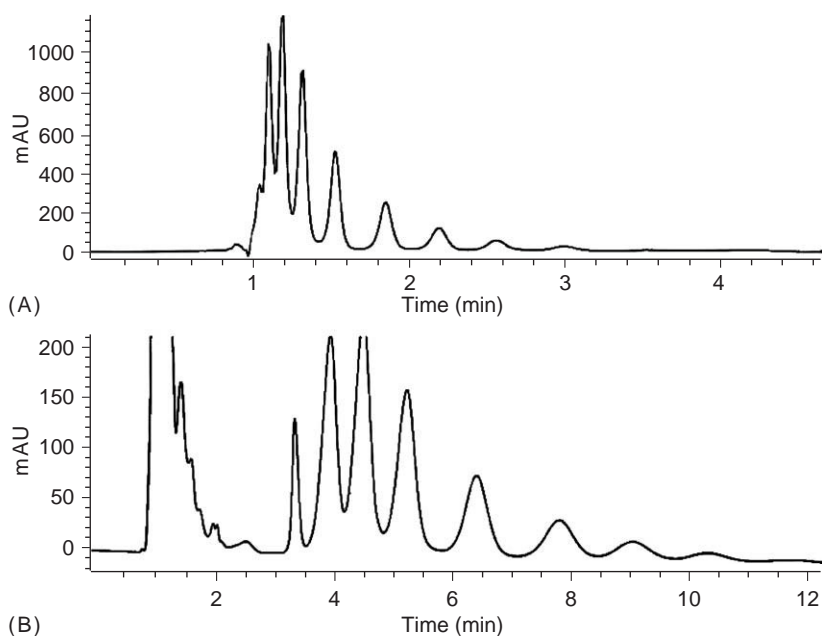


**Figure 4** Breakthrough curve of 2-propanol in heptane on a Separon SGX silica gel column in normal-phase gradient-elution HPLC calculated using the experimental isotherm data. Gradient 0–50% 2-propanol in 30 min,  $1 \text{ ml min}^{-1}$ .  $V$  – milliliters of eluate from the start of gradient elution,  $\phi$  – concentration of 2-propanol at the outlet from the column. (Adapted with permission from Jandera P (2002) *Journal of Chromatography A* 965: 239–261; © Elsevier.)

samples such as carbohydrates may be too weakly retained in RPLC systems to allow their separation. They are well retained in normal-phase systems, but usually are not sufficiently soluble in the nonaqueous mobile phases. This problem can be solved by using bonded amine or special stationary phases, such as polyhydroxyethyl aspartamide with aqueous–organic mobile phases in the so-called hydrophilic interaction (HILIC) chromatography. The separation of polar compounds on polar stationary phases in mobile phases containing water is based rather on liquid–liquid partition than on adsorption, but shows characteristic normal-phase behavior – the retention increases in mobile phases with higher concentrations of the organic solvent (acetonitrile, less polar than water) and is higher for more polar samples. Gradients with decreasing concentrations of organic solvents or increasing concentrations of salts (often volatile ones) were successfully applied for the separation of various biological samples. The separation of ionic samples in the HILIC mode can be improved by using weak acids or bases as the mobile phase additives, such as triethylamine for basic compounds and acetic or formic acid for acidic compounds. Good separations of sugars, oligosaccharides, amino acids, or peptides can be achieved by HILIC. **Figure 5A** shows an example of separation of ethoxylated surfactants on an aminopropylsilica

column using an acetonitrile–dichloromethane–water mobile phase. An example of a HILIC separation of sulfonated anionic surfactants in a mobile phase with an ion-pairing reagent is shown in **Figure 5B**.

The ion-exchange capacity of silica and alumina can be used for its dynamic modification by interactions with ions in the mobile phases consisting of an aqueous buffer mixed with an organic solvent. These systems can be used for separations of basic drugs and other weakly basic compounds. Here, the retention mechanism is complex and involves ion-exchange interactions. Organic acids can be separated using normal-phase ion-pairing chromatography with nonaqueous mobile phases, such as organic solutions of tetraalkylammonium hydroxide, which becomes adsorbed on the silica surface. In aqueous–organic mobile phases containing cetyl triethylammonium bromide, the ion pairing reagent is adsorbed in such a way on the silica gel surface that the hexadecyl groups point away from the surface, forming thus an apolar layer similar to chemically bonded alkylsiloxane phases, except that the adsorbed layer is in dynamical equilibrium with the alkylammonium ions in the mobile phase. Anionic solutes are retained as ion pairs and their separation is more reproducible than in ion-pair chromatography on nonpolar chemically bonded phases. However, a relatively long time is required to reach equilibrium compared to



**Figure 5** HILIC separation of the oligomers in a Serdiox NNP4 sample of ethoxylated nonionic surfactants on a Separon SGX Amine column, 7  $\mu\text{m}$ , 150  $\times$  3.3 mm ID, with acetonitrile–water–dichloromethane 49:1:50, mobile phase (A) and of a partially sulfated Serdiox NNP4 sample with acetonitrile–water–dichloromethane 68.6:1.4:30 mobile phase containing 0.04 mol l<sup>-1</sup> cetyl trimethylammonium bromide (B). Flow rate 0.5 ml min<sup>-1</sup>. Detection: UV, 230 nm. (Adapted from Jandera P (2001) Comparison of various modes and phase systems for analytical HPLC. In: Valko K (ed.) *Separation Methods in Drug Synthesis and Purification*, pp. 1–71. Amsterdam: Elsevier, with permission; © Elsevier.)

reversed-phase separations on chemically bonded phases, so that the method development and optimization may be lengthy.

## Applications of Normal-Phase HPLC

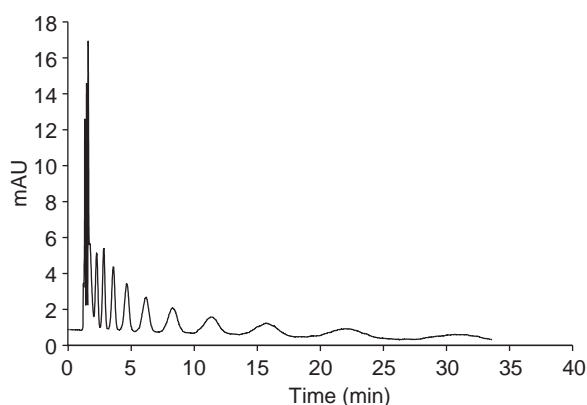
NPLC is most suitable for separation of nonionic and moderately polar compounds, especially for lipophilic samples that are too strongly retained by RPLC. Lipids differing in the number and position of double bonds, tocopherol, carotenoids, fat-soluble vitamins, and steroids in pharmaceuticals can be successfully separated by NPLC on silica gel or alumina columns. Mixed lipid classes in the extracts of animal or plant tissues can be analyzed on silica columns or on columns with bonded polyvinyl alcohol using complex solvent gradients. Gradient-elution

NPLC methods were reported for the analysis of natural phospholipids in nervous tissues, meat, etc. Mono-, di-, and triacylglycerol classes in natural oils can be distinguished on silica gel columns using normal-phase gradient elution. In these applications, either low-wavelength UV detection at 205 nm, or evaporative light-scattering detection is used for quantitative analysis.

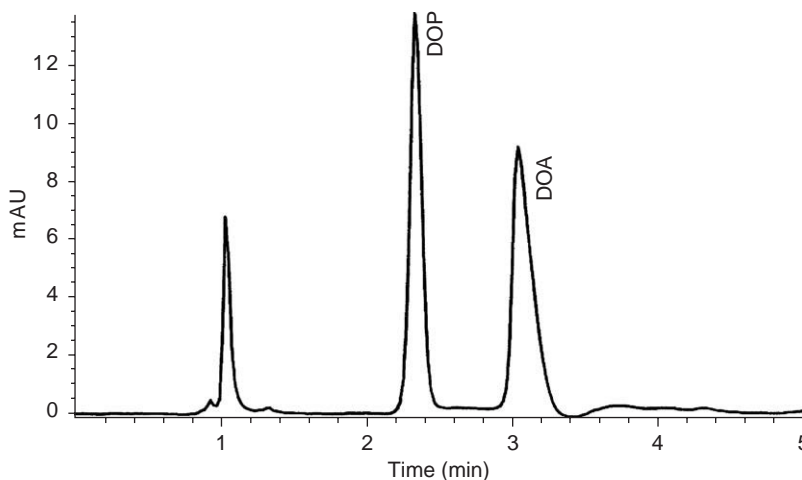
Normal-phase HPLC usually offers much improved separation of positional isomers or stereoisomers with respect to RPLC. This is also the reason why normal-phase liquid chromatographic mode with nonaqueous mobile phases is often used for the separation of enantiomers on chiral bonded stationary phases.

NPLC is often the method of choice for separation of water-insoluble synthetic polymers and oligomers with polar repeat monomer units or with polar end groups by the so-called interactive liquid chromatography, which usually offers much better resolution than size-exclusion chromatography for lower polymers with molar masses up to 10 000–20 000. Because the effect of increasing concentration of the polar solvent **B** is more significant for polymers than for small molecules (larger values of the parameter *m* in eqn [2] or [3]), successful separation over a broad range of molar masses requires gradient elution, as demonstrated by imperfect isocratic separation of a polystyrene sample with mean molar mass 2350 (Figure 6).

Normal-phase chromatography may provide rapid analyses of samples whose pretreatment involves extraction into a nonpolar solvent, as such samples can be injected directly onto the column, whereas they require evaporation and reextraction into a polar solvent before reversed-phase separation. Figure 7



**Figure 6** Separation of a polystyrene sample with mean molar mass 2350 Da, on two Nova Pak silica gel columns in series, 4  $\mu\text{m}$ , 150  $\times$  3.9 mm ID, in dioxane–hexane 5:95 at 1 ml min<sup>-1</sup>. Detection UV, 254 nm. Halama M and Jandera P, unpublished results.



**Figure 7** Separation of dioctyl phthalate (DOP) and dioctyl adipate (DOA) plasticizers in exhaust gases from an industrial plant. Separon SGX silica gel column, 5  $\mu\text{m}$ , 150  $\times$  3 mm ID; mobile phase: 0.3% 2-propanol in *n*-hexane (v/v), 1 ml min<sup>-1</sup>; temperature: 40°C. Sample volume: 20  $\mu\text{l}$ , UV detection at 230 nm.

shows an example of separation of plasticizers as pollutants in the work place atmosphere, after trapping on a cartridge containing a 60  $\mu\text{m}$  particle C18 sorbent. Hexane should be used for efficient extraction and the extract can be injected directly onto a silica gel column, providing complete separation of dioctyl adipate from dioctyl phthalate, whereas their reversed-phase separation is difficult and the sample requires reextraction into methanol before the HPLC analysis.

**See also:** **Liquid Chromatography:** Principles; Column Technology; Mobile Phase Selection. **Thin-Layer Chromatography:** Method Development.

### Further Reading

Jandera P (2002) Gradient elution in normal-phase high-performance liquid chromatographic systems. *Journal of Chromatography A* 965: 239–261.

Jandera P and Churáček J (1985) *Gradient Elution in Liquid Column Chromatography*. Amsterdam: Elsevier.

Jandera P, Petránek L, and Kučerová M (1997) Characterization and prediction of retention in isocratic and gradient-elution normal-phase high-performance liquid chromatography on polar bonded stationary phases with binary and ternary solvent systems. *Journal of Chromatography A* 791: 1–19.

Nawrocki J (1991) Silica surface controversies, strong adsorption sites. *Chromatographia* 31: 177–192. (see also pp. 193–203)

Snyder LR (1968) *Principles of Adsorption Chromatography*. New York: Dekker.

Snyder LR (1974) Role of the solvent in liquid–solid chromatography – a review. *Analytical Chemistry* 46: 1384–1393.

Snyder LR, Kirkland JJ, and Glajch JL (1997) *Practical HPLC Method Development*. New York: Wiley-Interscience.

Unger KK (1979) *Porous Silica*. Amsterdam: Elsevier.

## Reversed Phase

**R A Shalliker and S Kayillo**, University of Western Sydney, South Penrith, NSW, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Liquid chromatography (LC) was first developed around the turn of the twentieth century and is defined by the International Union of Pure and Applied Chemistry (IUPAC) as a means of physically separating components in mixtures by distributing the components between two phases. One of these phases is moving and percolates through the second, which is stationary. In LC the moving phase or mobile phase is a liquid. The technique has since grown enormously to include the areas of capillary LC, microbore LC, high performance LC (HPLC), and preparative LC. In the following discussion we shall detail the process of solute retention in the technique referred to as reversed-phase liquid chromatography. This is the most widely employed form of liquid chromatography, and in 1997, more than 80% of all HPLC separations were reportedly reversed phase. In this mode of chromatography the stationary phase is a non-polar adsorbent, while the mobile phase is more polar than the stationary phase.

### Reversed-Phase Liquid Chromatography

The first reported use of reversed-phase style LC came in 1948 where Boldingh used a rubber powder stationary phase and an acetone and methanol mobile phase to separate fatty acids. The term reversed phase was subsequently coined by Howard and Martin in 1950 with the separation of fatty acids on a paraffin/*n*-octane stationary phase using aqueous eluents. Because the technique used methods that were opposite to those normally employed, the technique bore the name reversed phase, distinguishing it from the normal-phase separation approach.

The popularity of reversed-phase LC has since escalated, largely because the technique is suitable for the separation of a large number of compounds with varying degrees of polarity and structure. There is a wide range of stationary phase surfaces that support reversed-phase applications and the technique is generally easy to use with fast solvent/stationary phase equilibration times. Typically mobile phases consist of a binary mixture of water plus an organic modifier of methanol, acetonitrile, or tetrahydrofuran. The most common method for preparing nonpolar stationary phases has been through the chemical modification of silica support materials, even though the use of silica-based supports is limited to pHs between 2 and 7 (although some silica-based supports have been prepared that



shows an example of separation of plasticizers as pollutants in the work place atmosphere, after trapping on a cartridge containing a 60  $\mu\text{m}$  particle C18 sorbent. Hexane should be used for efficient extraction and the extract can be injected directly onto a silica gel column, providing complete separation of dioctyl adipate from dioctyl phthalate, whereas their reversed-phase separation is difficult and the sample requires reextraction into methanol before the HPLC analysis.

**See also:** **Liquid Chromatography:** Principles; Column Technology; Mobile Phase Selection. **Thin-Layer Chromatography:** Method Development.

### Further Reading

Jandera P (2002) Gradient elution in normal-phase high-performance liquid chromatographic systems. *Journal of Chromatography A* 965: 239–261.

Jandera P and Churáček J (1985) *Gradient Elution in Liquid Column Chromatography*. Amsterdam: Elsevier.

Jandera P, Petránek L, and Kučerová M (1997) Characterization and prediction of retention in isocratic and gradient-elution normal-phase high-performance liquid chromatography on polar bonded stationary phases with binary and ternary solvent systems. *Journal of Chromatography A* 791: 1–19.

Nawrocki J (1991) Silica surface controversies, strong adsorption sites. *Chromatographia* 31: 177–192. (see also pp. 193–203)

Snyder LR (1968) *Principles of Adsorption Chromatography*. New York: Dekker.

Snyder LR (1974) Role of the solvent in liquid–solid chromatography – a review. *Analytical Chemistry* 46: 1384–1393.

Snyder LR, Kirkland JJ, and Glajch JL (1997) *Practical HPLC Method Development*. New York: Wiley-Interscience.

Unger KK (1979) *Porous Silica*. Amsterdam: Elsevier.

## Reversed Phase

**R A Shalliker and S Kayillo**, University of Western Sydney, South Penrith, NSW, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

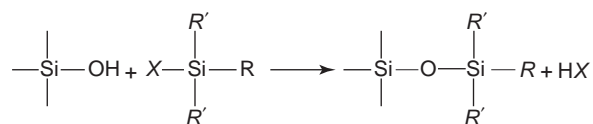
Liquid chromatography (LC) was first developed around the turn of the twentieth century and is defined by the International Union of Pure and Applied Chemistry (IUPAC) as a means of physically separating components in mixtures by distributing the components between two phases. One of these phases is moving and percolates through the second, which is stationary. In LC the moving phase or mobile phase is a liquid. The technique has since grown enormously to include the areas of capillary LC, microbore LC, high performance LC (HPLC), and preparative LC. In the following discussion we shall detail the process of solute retention in the technique referred to as reversed-phase liquid chromatography. This is the most widely employed form of liquid chromatography, and in 1997, more than 80% of all HPLC separations were reportedly reversed phase. In this mode of chromatography the stationary phase is a non-polar adsorbent, while the mobile phase is more polar than the stationary phase.

### Reversed-Phase Liquid Chromatography

The first reported use of reversed-phase style LC came in 1948 where Boldingh used a rubber powder stationary phase and an acetone and methanol mobile phase to separate fatty acids. The term reversed phase was subsequently coined by Howard and Martin in 1950 with the separation of fatty acids on a paraffin/*n*-octane stationary phase using aqueous eluents. Because the technique used methods that were opposite to those normally employed, the technique bore the name reversed phase, distinguishing it from the normal-phase separation approach.

The popularity of reversed-phase LC has since escalated, largely because the technique is suitable for the separation of a large number of compounds with varying degrees of polarity and structure. There is a wide range of stationary phase surfaces that support reversed-phase applications and the technique is generally easy to use with fast solvent/stationary phase equilibration times. Typically mobile phases consist of a binary mixture of water plus an organic modifier of methanol, acetonitrile, or tetrahydrofuran. The most common method for preparing nonpolar stationary phases has been through the chemical modification of silica support materials, even though the use of silica-based supports is limited to pHs between 2 and 7 (although some silica-based supports have been prepared that

are stable to pH 9). The surface modification of silica has largely been based on organosilanization, according to a generalized reaction scheme shown below:



where X is a reactive group, such as a halide (usually chloride),  $R'$  is typically a methyl or ethyl group, and R is the main group that imparts a hydrophobic (nonpolar) nature to the stationary phase surface (for instance an octyldecyl (C18) chain). When the surface modification takes place to yield a monomeric surface coverage, the reactive group on the organosilane reacts with hydroxyl groups located on the silica surface, at a concentration that usually does not exceed  $4 \mu\text{mol m}^{-2}$ . Such a surface coverage usually results in approximately only half the silica hydroxyl groups undergoing reaction, even less for lower surface coverages. These residual hydroxyl groups can lead to detrimental chromatographic performance, particularly in the analysis of basic solutes, hence it is important to protect the solute interactions from these hydroxyl groups. This can be achieved in a number of ways; through a process referred to as endcapping, the employment of polymeric surface coverages, or by using silanizing reagents that have  $R'$  groups that are longer and more hydrophobic than methyl groups in combination with endcapping. When endcapping is employed, a reversed-phase moiety such as trimethyl silane is reacted with the residual hydroxyl groups in the same process as for the coating of the particles with the reversed-phase ligand. The endcapping reagent has a smaller hydrodynamic volume than the main reversed-phase ligand and is able to access a higher proportion of the unreacted hydroxyl groups that underlie the main surface coverage. Protection of solutes from residual hydroxyl groups can be achieved by coating the particles with a polymeric surface coverage. A polymeric surface coverage is produced when organosilanes with multiple reactive X groups (i.e., trichloro-organosilanes) are used instead of organosilanes that have a single reactive X group (i.e., chloro-organosilanes), in which case a monomeric surface coverage is produced. This effectively results in a cross-linking process that protects the surface. However, polymeric coatings have been attributed to a decrease in the performance of the chromatographic system due to increases in mass transfer into the stationary phase layer. The third

approach to protect solutes from interactions with residual hydroxyl groups in which organosilanes that have longer and more hydrophobic  $R'$  groups (i.e., hexyl) than the methyl  $R'$  counterparts, in conjunction with endcapping, results in a surface whereby the extended  $R'$  groups are purportedly capable of providing a hydrophobic surface environment in close proximity to the silica surface.

Examples of reversed-phase HPLC bonded phases are given in Table 1. These include: alkyl bonded silica, where the R group is an alkyl chain, the length of which determines the degree of hydrophobicity (C18, C8, C4, etc.); the phenyl stationary phase,

**Table 1** List of some commonly used reversed-phase bonded stationary phases

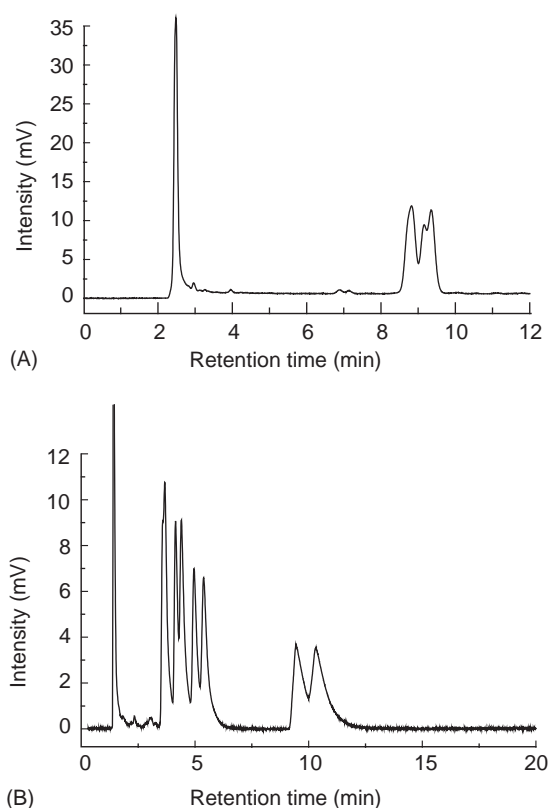
Type of reversed-phase stationary phase	Structure
Octadecyl (ODS) C18	$\begin{array}{c} R' \\   \\ \text{Si—O—Si—} \text{CCCCCCCCCCCCCCCCCC} \\   \\ R' \end{array}$
Octyl C8	$\begin{array}{c} R' \\   \\ \text{Si—O—Si—} \text{CCCCCCCC} \\   \\ R' \end{array}$
Phenyl	$\begin{array}{c} R' \\   \\ \text{Si—O—Si—} \text{C}_6\text{H}_5 \\   \\ R' \end{array}$
Propyl phenyl	$\begin{array}{c} R' \\   \\ \text{Si—O—Si—} \text{CH}_2\text{CH}_2\text{C}_6\text{H}_5 \\   \\ R' \end{array}$
Hexyl phenyl	$\begin{array}{c} R' \\   \\ \text{Si—O—Si—} \text{CCCCC—C}_6\text{H}_5 \\   \\ R' \end{array}$
Amino ( $\text{NH}_2$ )	$\begin{array}{c} R' \\   \\ \text{Si—O—Si—} \text{CH}_2\text{CH}_2\text{NH}_2 \\   \\ R' \end{array}$
Nitrile (cyano – CN)	$\begin{array}{c} R' \\   \\ \text{Si—O—Si—} \text{CH}_2\text{CH}_2\text{C}\equiv\text{N} \\   \\ R' \end{array}$

which may offer a change in the selectivity due to the possibility of  $\pi$ - $\pi$  interactions; cyano (nitrile) and amino surfaces, which can be employed in both reversed- and normal-phase mode since they possess a medium polarity. More recently, new 'polar end-capped' reversed-phase stationary phases have been developed, although applications based on the use of these materials are presently limited. The exact nature of the 'polar' endcapping agents is currently proprietary and little has been reported on the mechanisms of retention. However, the general consensus with regard to the mode of operation is that the polar endcapping agents repel basic species from within the region near the support surface. By far the most popular reversed-phase material is the C18 stationary phase and a multitude of applications using C18 columns are reported in the literature.

Other stationary phase supports that are employed for reversed-phase separations include a variety of polymeric resins (such as polystyrene-divinylbenzene). These types of stationary phases were originally developed in an effort to overcome the problems associated with residual hydroxyl groups on the surface of silica and also allow for separations to be performed in alkaline solvents. Polymeric resins are generally less well received than silica-based supports because initial manufacture produced columns that suffered from problems associated with swelling in changing solvent environments. This affected the pore structure and the pressure within the column. The modern polymeric column is, however, far more robust than its predecessor and in many respects chromatographers who may have initially employed polymeric resins in their early days may find new joy in the modern polymeric stationary phase. The use of polymeric columns finds a niche in HPLC/MS/MS, mainly as a consequence of the pH stability of the polymeric resins. This allows these columns to be employed in either very basic or acidic solvents, which are often required in chemical ionization procedures. Another advantage of the polymeric resins is that the surface is homogeneous, resulting in a decreased likelihood of secondary elution mechanisms through interactions with residual hydroxyl groups. However, the homogeneity of the surface may also be a disadvantage as it somewhat limits the selectivity of the surface.

Porous graphitic carbon has also been used as a nonbonded reversed-phase surface. This has a rigid structure, is chemically stable, hydrophobic, and displays different selectivity compared to silica-based stationary phases. Porous graphitic carbon is composed of sheets of hexagonally arranged carbon atoms showing  $sp^2$  hybridization and has a regular, homogeneous, and flat surface. These types of

surfaces are particularly useful for the separation of isomeric mixtures, such as geometric and diastereomers and in the analysis of highly polar analytes normally not retained on C18-silica. A number of review articles have been published on porous graphitic carbon. One particular advantage of porous graphitic carbon is that it is stable at high temperatures, and across the pH range 1–14 and is resistant to organic solvents. In addition to graphitic carbon surfaces, carbon clad zirconia is a second type of carbonaceous reversed-phase surface. The carbon clad zirconia stationary phase is a relatively new surface in which zirconia microspheres are coated with a light hydrocarbon (rather than the conventional silica support). The hydrocarbon is then pyrolyzed onto the zirconia surface, resulting in a superficial carbonaceous surface in many ways resembling a pure carbon support. However, the carbon clad zirconia support is more rigid and has a more easily controlled pore structure than its porous graphitic carbon counterpart. Carbon-clad zirconia supports are hydrophobic in nature and retain solutes based on electronic ( $\pi$ - $\pi$ ) interactions. They are often selective in the separation of isomers, particularly stereoisomers. Both the carbon-clad zirconia and porous graphitic carbon supports offer unique retention mechanisms selective for isomer separations. While these surfaces do behave as a reversed-phase adsorbent, they possess many unique chromatographic characteristics and therefore translations between chromatographic methods from conventional supports may be difficult. An interesting illustration as to the differences in retention processes that are apparent for various carbon surfaces, such as carbon-clad zirconia in comparison to bonded phase surfaces, such as C18-silica can be found in the elution behavior of oligostyrenes. The chromatograms illustrated in **Figure 1** show this effect for an oligostyrene with four configurational repeating units and a *sec*-butyl end group. For such an oligomer there are eight diastereomers, four of which result from the variation in tacticity associated with the stereochemistry of the main chain, while the remaining four diastereomers result from the stereochemical site in the *sec*-butyl end group. On a C18 column using an acetonitrile mobile phase, the oligostyrenes elute with some diastereomer selectivity (**Figure 1A**). However, when a carbon-clad zirconia column is employed, the diastereomer selectivity is highly expressed and seven diastereomers are apparent (**Figure 1B**). Of significance is the fact that a C18 column was 250 mm in length, packed with 5  $\mu$ m particles compared to the carbon-clad zirconia column, which was 100 mm in length, packed with 3  $\mu$ m particles.



**Figure 1** Illustration of the isomeric separation of oligostyrene containing four configurational repeating with a *sec*-butyl end group on (A) a C18 column (250 mm × 4.6 mm, Pd 5 μm) and (B) a carbon clad zirconia column (100 mm × 4.6 mm, Pd 3 μm). Both separations were in a mobile phase of 100% acetonitrile at 30°C.

## Retention in Reversed-Phase Liquid Chromatography

While virtually any compound, including, for example, amino acids, peptides, and proteins, may potentially be separated in a reversed-phase mode, the technique is ideally suited to the separation of hydrocarbons, substituted aromatics, homologs, and polymers. Generally, reversed- and normal-phase methods of separation are considered to be restricted to molecules with molecular weights less than 5000 Da; however, there are many reports in the literature on the reversed-phase separation of polymers with molecular weights up to one million Daltons.

Despite reversed phase chromatography being more than 50 years mature, there still exists considerable debate regarding the retention processes involved. Much of this debate has discussed processes of adsorption versus those of partitioning. Here the basic difference between these models lies in the solute association with the stationary phase. In partitioning, the solute is embedded within the stationary phase, as distinct from adsorption where the solute is

in contact with the stationary phase, but not fully embedded as is the case with partitioning. In partitioning, there are three distinct processes that occur: (1) the creation of a solute sized cavity within the stationary phase; (2) the transfer of the solute from the mobile phase into the cavity created within the stationary phase; and (3) the closure of the solute-sized cavity in the mobile phase that was vacated by the solute upon transfer to the stationary phase. A lattice model theory was used to describe this process taking into account the relevant contact interactions between solute molecules and those of the stationary and mobile phases. This process is driven by the relative chemical affinities between the solute and stationary and mobile phase molecules. These interactions may be considered as being coulombic, if for instance, the interacting molecules possess a net charge, or they may be dipole-dipole interactions if the interacting molecules contain temporary dipole moments. These types of interactions may also be governed by specific molecular orientation associated with the alignment of molecules undergoing, say, hydrogen bonding. The associated enthalpic and entropic contributions need then be considered in the retention process. The lattice model theory also describes the process of adsorption. However, the driving force for adsorption is much weaker than that of partitioning because the contact surface area of the solute contained on the surface of the stationary phase is lower than that of the solute when embedded within the stationary phase. The process for adsorption differs from partitioning in regards to the creation of a cavity adjacent to the stationary phase, rather than within the stationary phase.

In reality, retention need not be governed solely by either adsorption or partitioning. A combination of these retention processes may result, depending upon the nature of the solute molecule and the stationary phase environment. Because stationary phases are fixed to a support material, they do not behave as a bulk liquid. Instead the bonded moiety contains a degree of order that increases as the surface density increases. Solute partitioning then becomes entropically unfavourable within the increasing ordered system, especially since the inclusion of a solute molecule into the stationary phase further increases the state of order within the system. Hence adsorption processes dominate on higher density stationary phases, or when the size of the solute increases to the point whereby it is in effect excluded from within the bulk phase. Plots of  $\log k$  versus the carbon number of a homologous series undergoing chromatographic retention on alkyl chain (C18, C8, and C4 for example) modified silica illustrate this affect clearly.  $\log k$  increases linearly with respect to the number of

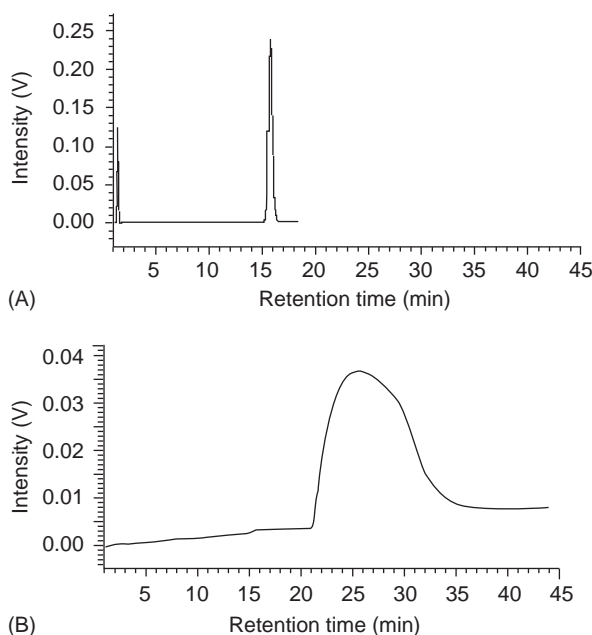
carbon atoms of the homolog, until the carbon number of the solute approximately equals the length of the organic ligand of the stationary phase. This discontinuity indicates a change in retention mechanism from that of pure partitioning to a mixed process of partitioning and adsorption, after which steric factors prevent solute molecules undergoing partitioning, and adsorption consequently dominates. Adsorption also dominates on stationary phases prepared from short chain ligands, such as the C4, as these ligands are highly restricted in their degree of freedom.

Partitioning and adsorptive type retention mechanisms only explain in part the nature of solute interactions with the stationary phase. The separation of isomers, for example, must be explained by the shape selectivity of the reversed-phase surface. Stationary phase ligands bonded to the support surface are able to exist in a considerable number of molecular orientations. The molecular shape of solute molecules and their orientation with respect to the stationary phase ligands determines their degree of retention. The relative change in free energy is more substantial for a solute molecule that aligns parallel to the stationary phase surface, rather than normal, hence the shape selectivity of the surface. Studies using isomeric polyaromatic hydrocarbons have illustrated this effect. Retention has been shown to increase according to rods > disks > flexible chains.

Any discussion regarding reversed-phase LC would be incomplete without comment regarding the solvophobic theory. This theory was essentially the first attempt to fully describe the retention process in reversed-phase liquid chromatography. The solvophobic theory predicts that the stationary phase plays a passive role and retention is largely governed by the exclusion of the solute from the mobile-phase environment. In this theory, the mobile-phase molecules are considered to exist in an extensive network of hydrogen bonding due to the hydro-organic nature of the mobile phase. The water molecules more favourably prefer to bond with each other rather than interact with solute molecules. Consequently, nonpolar solute molecules distributed within the aqueous mobile phase are believed to exist in a highly ordered network. Thus, it is entropically unfavourable to put a nonpolar molecule within the aqueous mobile phase. Consequently, exclusion of this nonpolar molecule from the aqueous phase and its subsequent association with the nonpolar reversed-phase ligand is said to be entropically favourable as the ordered mobile phase attains a disordered configuration following expulsion of the solute. The expulsion of the solute from the mobile phase is thought to be the driving force for retention. The solvophobic

theory, however, offers an incomplete description of the retention process. The independence of solute retention on the type of stationary phase is not supported by the ample experimental evidence that shows retention being determined by the nature of the stationary phase, as well as the nature of the mobile phase.

Whether the mechanism of retention is dominated one way or the other by partitioning or adsorption may in many respects be considered immaterial to the separation process. What is important, however, is attaining resolution. For this part correct selection of the stationary phase and mobile phase combinations is paramount. Factors that lead to loss in resolution include (the obvious) lack of selectivity and (the not so obvious) secondary elution processes that lead to band distortions often revealing themselves in band tailing. An example of poor column selection may be found in the analysis of polyaromatic hydrocarbons. On an alkyl-bonded column (C18) band shape is uniform (Figure 2A), but on a carbon-clad zirconia column significant peak tailing is observed (Figure 2B). This effect is also apparent, but to a lesser extent on phenyl columns. For these types of solutes, as the  $\pi$ - $\pi$  interactions become more significant, band tailing increases and the separation performance decreases. Incidentally, the difference in retention behavior of these solutes on the different stationary phases is also an indicator that the role played by the stationary phase is more



**Figure 2** Chromatograms of pyrene illustrating the band tailing that is observed on a carbon clad zirconia reversed-phase column in comparison to a C18 reversed-phase column.



substantial than considered in the solvophobic theory.

The development of reversed-phase chromatographic media has intensified during the last decade. Now, many manufacturers pride themselves on their high degree of the batch to batch reproducibility, both with respect to the retention properties of the stationary phase and also in regards to the quality of the packed bed. Recent years have seen the development of a variety of polar embedded and polar endcapped reversed-phase materials. There is in reality no limitation placed on the type of reversed-phase material that could be used for reversed-phase applications, other than commercial availability. Therefore expansion of the available reversed-phase media that enter the market place continues, as chromatographers continually seek new ways to gain selectivity.

**See also:** **Paints:** Water-Based. **Pharmaceutical Analysis:** Drug Purity Determination.

## Further Reading

- Braithwaite A and Smith FJ (1996) *Chromatographic Methods*. Glasgow: Chapman and Hall.
- Dorsey JG and Dill KA (1989) *Chemical Reviews* 89: 331–346.
- Horvath C, Melander W, and Molnar I (1976) *Journal of Chromatography* 125: 129–156.
- Neue UD (1997) *HPLC Columns: Theory, Technology, and Practice*. New York: Wiley-VCH.
- Ranatunga RPJ and Carr PW (2000) *Analytical Chemistry* 72: 5679–5692.
- Rimmer CA, Simmons CR, and Dorsey JG (2002) *J. Chromatogr. A* 965: 219.
- Ying PT, Dorsey JG, and Dill KA (1989) *Analytical Chemistry* 61: 2540–2546.

## Ion Pair

**B-A Persson**, Astra Zeneca R&D Mölndal, Mölndal, Sweden

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Ion pair extraction is a technique used for the distribution of ionic compounds to an organic phase with the aid of counterions of opposite charge. Ion pairs are formed in the organic solvent, e.g., the extraction of a cation  $\text{HA}^+$  with an anion  $\text{X}^-$  making up an ion pair HAX, and the distribution of the ion pairs is dependent on their solvation properties. In organic analysis, this ion pair extraction methodology was originally used to extract basic and acidic compounds with counterions of different hydrophobicity as well as a means to provide selectivity and enable detection for analytes with low inherent detectability. Counterion-aided extraction has been widely applied in liquid–liquid extraction but during the last two decades the main use of the ion pair technique has been in reversed-phase liquid chromatography (LC), an important separation technique for ionic compounds in organic and biomedical analysis.

In chromatography, the ion pair technique was first used in liquid–liquid chromatography where the counterion was supplied through the aqueous phase and the ion pair distributed to the organic phase. In the normal or straight-phase mode, the ionic solutes are transported as ion pairs in the mobile organic

phase; in the reversed-phase mode the solutes are retained as ion pairs in the stationary organic liquid phase. However, due to strong demands on temperature control and liquid–liquid equilibration, liquid–liquid chromatography was early on substituted by liquid–solid chromatography owing to the relative ease of this technique.

In the normal-phase mode ion pair adsorption or liquid–solid chromatography found a couple of applications but since the entry of bonded phases, reversed-phase LC has been a quite dominating separation technique and the main one employing the ion pair concept. Ion pair chromatography has been the most widely used name for this separation method, but terms such as dynamic ion exchange, ion interaction, and paired ion chromatography have also been used, where retention is due to electrostatic, hydrophobic, and polar interactions.

## Principles of Ion Pair Formation

The ion pair formation process in the extraction of two oppositely charged ionic compounds to an organic phase is most often expressed by the equilibrium



where  $\text{HA}^+$  is a protonated amine, a quaternary ammonium ion, or any other cation and  $\text{X}^-$  is an ionized acid or other anions. HAX is the ion pair

substantial than considered in the solvophobic theory.

The development of reversed-phase chromatographic media has intensified during the last decade. Now, many manufacturers pride themselves on their high degree of the batch to batch reproducibility, both with respect to the retention properties of the stationary phase and also in regards to the quality of the packed bed. Recent years have seen the development of a variety of polar embedded and polar endcapped reversed-phase materials. There is in reality no limitation placed on the type of reversed-phase material that could be used for reversed-phase applications, other than commercial availability. Therefore expansion of the available reversed-phase media that enter the market place continues, as chromatographers continually seek new ways to gain selectivity.

**See also:** **Paints:** Water-Based. **Pharmaceutical Analysis:** Drug Purity Determination.

## Further Reading

- Braithwaite A and Smith FJ (1996) *Chromatographic Methods*. Glasgow: Chapman and Hall.
- Dorsey JG and Dill KA (1989) *Chemical Reviews* 89: 331–346.
- Horvath C, Melander W, and Molnar I (1976) *Journal of Chromatography* 125: 129–156.
- Neue UD (1997) *HPLC Columns: Theory, Technology, and Practice*. New York: Wiley-VCH.
- Ranatunga RPJ and Carr PW (2000) *Analytical Chemistry* 72: 5679–5692.
- Rimmer CA, Simmons CR, and Dorsey JG (2002) *J. Chromatogr. A* 965: 219.
- Ying PT, Dorsey JG, and Dill KA (1989) *Analytical Chemistry* 61: 2540–2546.

## Ion Pair

**B-A Persson**, Astra Zeneca R&D Mölndal, Mölndal, Sweden

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Ion pair extraction is a technique used for the distribution of ionic compounds to an organic phase with the aid of counterions of opposite charge. Ion pairs are formed in the organic solvent, e.g., the extraction of a cation  $\text{HA}^+$  with an anion  $\text{X}^-$  making up an ion pair HAX, and the distribution of the ion pairs is dependent on their solvation properties. In organic analysis, this ion pair extraction methodology was originally used to extract basic and acidic compounds with counterions of different hydrophobicity as well as a means to provide selectivity and enable detection for analytes with low inherent detectability. Counterion-aided extraction has been widely applied in liquid–liquid extraction but during the last two decades the main use of the ion pair technique has been in reversed-phase liquid chromatography (LC), an important separation technique for ionic compounds in organic and biomedical analysis.

In chromatography, the ion pair technique was first used in liquid–liquid chromatography where the counterion was supplied through the aqueous phase and the ion pair distributed to the organic phase. In the normal or straight-phase mode, the ionic solutes are transported as ion pairs in the mobile organic

phase; in the reversed-phase mode the solutes are retained as ion pairs in the stationary organic liquid phase. However, due to strong demands on temperature control and liquid–liquid equilibration, liquid–liquid chromatography was early on substituted by liquid–solid chromatography owing to the relative ease of this technique.

In the normal-phase mode ion pair adsorption or liquid–solid chromatography found a couple of applications but since the entry of bonded phases, reversed-phase LC has been a quite dominating separation technique and the main one employing the ion pair concept. Ion pair chromatography has been the most widely used name for this separation method, but terms such as dynamic ion exchange, ion interaction, and paired ion chromatography have also been used, where retention is due to electrostatic, hydrophobic, and polar interactions.

## Principles of Ion Pair Formation

The ion pair formation process in the extraction of two oppositely charged ionic compounds to an organic phase is most often expressed by the equilibrium



where  $\text{HA}^+$  is a protonated amine, a quaternary ammonium ion, or any other cation and  $\text{X}^-$  is an ionized acid or other anions. HAX is the ion pair

formed in the organic phase. The equilibrium constant summarizing the process is defined by

$$K_{\text{ex(HAX)}} = \frac{[\text{HAX}]_{\text{org}}}{[\text{HA}^+]_{\text{aq}}[\text{X}^-]_{\text{aq}}} \quad [2]$$

and expresses the combined effect of phase transfer of the ions and ion pair formation in the organic phase. The magnitude of the extraction constant,  $K_{\text{ex(HAX)}}$ , depends on the hydrophobic properties of the two ions forming the ion pair, the extent of electrostatic and hydrogen-bonding interaction and on the nature of the organic phase.

If processes other than the one given in eqn [2] can be disregarded, the distribution ratio for  $\text{HA}^+$ ,  $D_{\text{HA}}$ , is expressed by the product of the extraction constant and the concentration of the counterion,  $\text{X}^-$ , in the aqueous phase:

$$D_{\text{HA}} = \frac{[\text{HAX}]_{\text{org}}}{[\text{HA}^+]_{\text{aq}}} = K_{\text{ex(HAX)}}[\text{X}^-]_{\text{aq}} \quad [3]$$

For a given organic solvent the extraction constant and the concentration of the counterion will accordingly determine the distribution ratio.

In a normal-phase liquid-liquid chromatographic system with a nonpolar mobile phase, in which the ion pair is migrating, the retention factor (capacity factor) is given by

$$k_{\text{HA}} = q/D_{\text{HA}} = q/K_{\text{ex(HAX)}}[\text{X}^-]_{\text{aq}} \quad [4]$$

where  $q$  is the volume ratio  $V_{\text{stat}}/V_{\text{mob}}$ .

In a corresponding reversed-phase system, where the ion pairs are retained in the stationary organic phase, the expression will be

$$k_{\text{HA}} = qD_{\text{HA}} = qK_{\text{ex(HAX)}}[\text{X}^-]_{\text{aq}} \quad [5]$$

Equations [4] and [5] can be used to calculate the retention factors for a given ion pair system or to choose conditions for suitable retention of the ionic solute. A more hydrophobic counterion  $\text{X}^-$  (higher  $K_{\text{ex(HAX)}}$ ) in higher concentration will increase the distribution of  $\text{HA}^+$  to the organic phase. This will decrease (eqn [4]) or increase (eqn [5]) the retention factor. The accuracy of the predictions made depends on the inertness of the chromatographic packing material used as support for the liquid stationary phase. Deviations between calculated and obtained retention data are often quite large owing to interactions of the solid phase with the ionic compounds.

The ion pair concept as described in eqns [1] and [2] implies transfer of ionic compounds between two liquid phases with maintenance of electroneutrality, ion pairs being formed in the nonpolar organic

phase. The process could be named counterion-aided extraction rather than ion pair extraction, since the latter may indicate that ion pair formation in the aqueous phase is preceding the phase transfer. However, it should be emphasized that this is by no means an inevitable first step, although the mechanism is not fully elucidated. It is reasonable to assume that ion pairs are formed when ionic species are distributed to a nonpolar organic solvent as in liquid-liquid extraction or liquid-liquid chromatography. As regards reversed-phase adsorption chromatography, it is more difficult to imagine ion pair formation in the hydrophobic bonded phase. The role of an adsorbed counterion is to induce a charged surface or, by the adsorption of an ionic solute, to bring about electroneutrality. The following will mainly be devoted to ion pair LC in the reversed-phase mode, which is the technique of most interest.

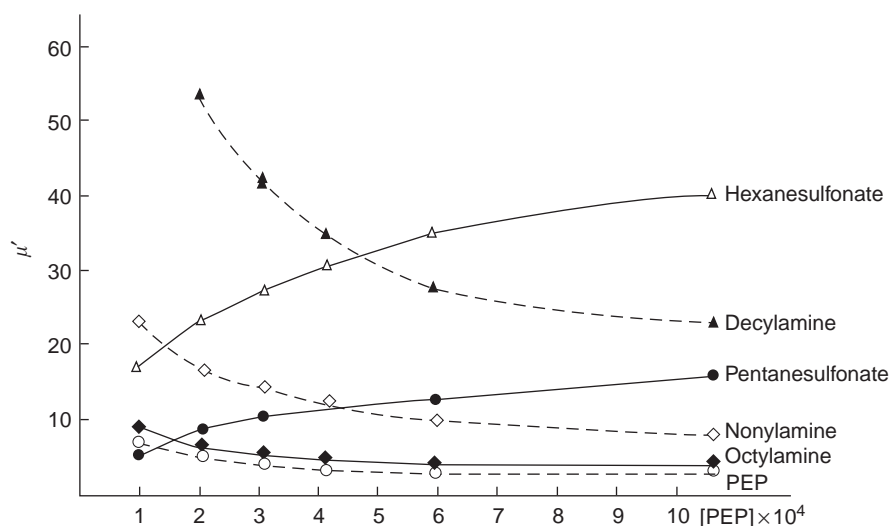
## Comparison of Theoretical Models

In ion pair separations with a hydrophobic solid stationary phase, most often alkyl-bonded silica, the binding capacity is limited. The principles of retention in such a system can be expressed as follows: (1) the retention of an ionic solute,  $\text{HA}^+$ , is accompanied by the displacement of a system ion of the same charge or by the binding of a counterion,  $\text{X}^-$  and (2) the solute and the components in the mobile phase may compete for the binding capacity of the adsorbing solid phase.

The overall retention of  $\text{HA}^+$  can be expressed in a simplified way assuming a Langmuir adsorption on to a homogeneous solid phase

$$k_{\text{HA}} = \frac{qK^\circ K_{\text{HAX}}[\text{X}^-]_{\text{aq}}}{1 + K_{\text{QX}}[\text{Q}^+]_{\text{aq}}[\text{X}^-]_{\text{aq}}} \quad [6]$$

where  $K^\circ$  is the capacity of the adsorbent,  $K_{\text{HAX}}$  and  $K_{\text{QX}}$  are distribution constants, and  $\text{Q}^+$  and  $\text{X}^-$  are ionic components in the mobile phase. Equation [6] shows that the retention factor of  $\text{HA}^+$  can be influenced by the nature of the counterion,  $\text{X}^-$ , and its concentration. However, an increase of the hydrophobicity and concentration of the counterion will also increase the denominator, giving a curved relationship as demonstrated in **Figure 1**. When unity in the denominator can be disregarded, e.g., with a hydrophobic counterion  $\text{X}^-$  in increasing concentration, eqn [6] will express an ion-exchange process between  $\text{HA}^+$  and  $\text{Q}^+$ .  $\text{Q}^+$  is a cation in the mobile phase such as potassium in a phosphate buffer or sodium with octyl sulfate counterion, and may also be a hydrophobic cation such as dimethyloctylammonium.



**Figure 1** Retention of anionic and cationic compounds with 1-phenethyl-2-picolinium (PEP) in the mobile phase. Stationary phase:  $\mu$ Bondapak Phenyl. Mobile phase: PEP in  $0.1 \text{ mol l}^{-1}$  acetic acid. UV detection at 254 nm. (Reprinted with permission from Denkert M, Hackzell L, Schill G, and Sjögren E (1981) Reversed-phase ion-pair chromatography with UV-absorbing ions in the mobile phase. *Journal of Chromatography* 218: 31–43; © Elsevier.)

The presence in the mobile phase of a competing alkylammonium ion,  $Q^+$ , will decrease the retention of  $HA^+$  as illustrated in **Figure 1**, and may also, in particular for hydrophobic cationic solutes, decrease tailing. The unsatisfactory performance of certain aliphatic ammonium compounds was more pronounced in the past but modern and improved reversed-phase packings have largely eliminated the need for precautions by including an ionic modifier in the mobile phase. The observed effects indicating how the retention is influenced by the concentration and hydrophobicity of a counterion  $X^-$  and competing cation  $Q^+$  are largely in accordance with eqn [6]. However, it must be emphasized that this model does by no means explain the mechanism of 'ion pair' chromatography, neither do alternative stoichiometric models such as 'dynamic ion exchange' and 'ion interaction' chromatography. Ion pair formation in the mobile phase is still in some instances proposed as preceding ion pair adsorption. However, equilibrium constants for ion pair formation in aqueous phases reported in literature are, with a few exceptions, small and this process is not a prerequisite for ion pair extraction to an organic solvent or to adsorption of an ionic solute to an organic solid phase. Neither is it probable that micelle formation in the mobile phase is responsible for unanticipated retention effects unless conditions for micelle formation are at hand.

The retention of a solute in chromatography is determined by thermodynamic equilibria, and non-stoichiometric theories have been developed for qualitative evaluation of ion pair chromatography.

These theories are based on the interaction of the solute ion with the charged surface layer established by the adsorbed counterion and by adsorbed competing ions. The nonstoichiometric models apply the Poisson–Boltzmann equation to estimate retention from an electrostatic point of view. The electrical double-layer model applied uses different approaches such as 'liquid partition', 'surface adsorption, diffuse layer ion-exchange', and 'surface adsorption' double-layer models. It is not possible to draw conclusions about the ion pair process from chromatographic retention data, but each model and theory may find use in describing experimental results under the particular conditions studied.

The variability in hydrophobicity and other properties of the solid phase and in concentration of the ionic species involved will make little sense for elucidations of more general validity. The heterogeneous nature of the adsorbing phase, most often  $C_8$  or  $C_{18}$  bonded silica, contributes to the complexity of the systems. Whatever approach is being used it must be emphasized that ion pairs are not fixed complexes in solution and even less on a surface. They represent a dynamic equilibrium in which the solute ion is transported or retained by the aid of the counterion, and electrostatic, hydrophobic, and other interaction provide overall electroneutrality. In this context it should be made clear that although the treatment so far has dealt with  $HA^+$  as solute ion and  $X^-$  as counterion, the considerations made are as valid for the opposite system with  $X^-$  as ionic solute and  $HA^+$  as counterion in the mobile aqueous phase.

## Range of Anion and Cation Pairing Reagents

There is a wide range of counterions that can be used in reversed-phase ion pair chromatography. Besides the hydrophobicity, the concentration of the counterion will influence retention. Concentrations used in the mobile aqueous phase are often in the region of  $0.001\text{--}0.1\text{ mol l}^{-1}$  depending on the hydrophobicity and solubility of the counterion. If either of the ionic species originates from a weak protolyte, pH has to be controlled by a pH-buffering electrolyte or the protolyte being present along with the counterion. Strong or weak acids may act as both ion pairing and pH buffering agents in the pH region of 2–8, available for most bonded silica packings. As a rule both the solute and the counterion should be in ionic form at the pH chosen. This means that strong protolytes like quaternary ammonium ions and sulfonates are well suited since they are aprotic and can be used irrespective of pH. Some anions and cations are listed in Table 1 to exemplify the counterions that are used. For normal separation the counterion should preferably be invisible to the detector. As can be seen in Table 1, there is a wide difference in physicochemical properties among the counterions listed. The choice of counterion should, of course, be adapted to the hydrophobic properties of the ionic solute. Lately, the evolving dominant position of mass spectrometry (MS) as detection mode in LC has put strong demands also on volatility of the ion pairing agent and the pH buffering electrolyte.

### Retention

Dihydrogenphosphate as a pH-buffering ion may give appropriate retention and separation of hydrophobic ammonium compound such as tricyclic

alkylamines. Hydrophilic catecholamines require a counterion like octyl sulfate or dodecyl sulfate (lauryl sulfate) to achieve sufficient retention. Perfluorinated carboxylic acids have found increased use in LC–MS applications due to their volatility; not only trifluoroacetate and heptafluorobutyrate but also higher perfluorinated carboxylic acids have been used.

Strongly lipophilic anionic solutes such as arylalkanesulfonates may not need a separate counterion beside sodium or potassium to be added for sufficient retention. Tetrabutylammonium has been used for moderately hydrophilic anionic solutes and cetyltrimethylammonium (cetrimide) for strongly hydrophilic ones. In recent reports, a couple of amines, such as triethylamine and dimethylhexylamine, have been introduced as cationic agents in LC–MS applications.

For a given solute there are in most instances alternative ionic systems, e.g., a more hydrophobic counterion in low concentration or a less hydrophobic ion in higher concentration, which may give similar retention. For major adjustment of the retention in an ion pair system, a different counterion should be selected that is either more hydrophobic or more hydrophilic depending on whether the solute is eluting too fast or too slowly. If only minor changes in retention are required, the counterion concentration in the mobile phase can be increased or decreased.

The discussion so far has not concerned the presence of an organic solvent in the aqueous mobile phase as a modifier. This is as obvious here as in other forms of reversed-phase LC. The same kind of organic modifiers are used, such as acetonitrile, methanol, 2-propanol, and tetrahydrofuran. Increased contents of these solvents will give decreased retention of the solutes, and their concentration can be used to regulate the eluent strength. In ion pair reversed-phase chromatography, minor changes in the retention of the solute are preferably made in this way. The organic modifier is, along with the nature and concentration of the counterion, decisive for the retention of the solute. Different combinations of the variables may give retention factors of the same magnitude. A hydrophobic counterion and a high content of organic solvent may give the same retention as a more hydrophilic counterion combined with a low degree of organic modifier.

Different combinations of phase systems may provide the gain in selectivity needed for difficult separations between ionic solutes of the same charge. The selectivity effects are even more pronounced for the resolution of two interfering sample components that are retained by different principles, e.g., one uncharged solute and one subjected to ionic interaction. For the selectivity of ion pair liquid

**Table 1** Counterions in ion pair LC

#### Anions

Dihydrogenphosphate

Acetate

Perchlorate

Trifluoroacetate, heptafluorobutyrate

Alkyl sulfates: butyl, octyl, dodecyl

Alkanesulfonates: methane-, hexane-, decane-

#### Cations

Sodium, potassium

Tetralkylammonium, tetraethyl-, tetrabutyl-

Triethylammonium

Dimethylhexylammonium

Trimethyloctylammonium

Cetyltrimethylammonium



chromatographic systems, the same rules or lack of rules apply as for other kinds of reversed-phase chromatography. Specific interactions, besides electrostatic, between the counterion and the ionic solute cannot be excluded but is by no means as obvious as in liquid-liquid distribution systems. Small selectivity changes that are often sufficient to separate two co-eluting components may be induced by using another organic modifier or by testing different combinations of counterion and content of organic solvent in the aqueous mobile phase. However, by the introduction of MS detection the issue of selectivity in LC separations has become less demanding than with other detection principles.

Ion pairing agents in liquid-liquid systems in reversed-phase mode have included dihydrogenphosphate for separation of tricyclic amines, octyl sulfate for catecholamines, and tetrabutylammonium for aromatic carboxylates and anions of sulfonamides, to exemplify some of the comparatively few applications. Liquid stationary phases coated on the alkyl-bonded phase include 1-pentanol, butyronitrile, and tributylphosphate. In normal-phase liquid-liquid ion pair chromatography aqueous perchlorate solution has been coated on to silica particles for ion pair separation of catecholamines and related compounds and tetrabutylammonium ion at neutral pH for carboxylates and anions of sulfonamides. The organic mobile phase often contained dichloromethane and butanol. In the normal-phase mode on silica alternative separation systems have been described with aqueous perchloric acid in methanol added to dichloromethane as mobile phase for separation of amines such as drug substances. This is not an extensively utilized, but quite useful, kind of separation, which has been named ion pair adsorption chromatography.

### Detection

In most instances it is preferable that the ion pairing reagent is invisible for the detection principle used and gives no background. The ionic agents discussed here have this property, being either alkyl sulfates or alkanesulfonates, or ammonium ions or inorganic electrolytes. In the past, strongly ultraviolet (UV)-absorbing anions such as naphthalenesulfonate, picrate, and also some arylalkylammonium cations have been used as counterions in liquid-liquid chromatography with an organic mobile phase. Nonvisible ionic solutes like acetylcholine, aliphatic amines, dipeptides, and alkyl carboxylates and alkanesulfonates could then be separated and detected by UV absorption through the ion pairing agent. The complexity of these systems with strong requirements on temperature control and careful equilibration

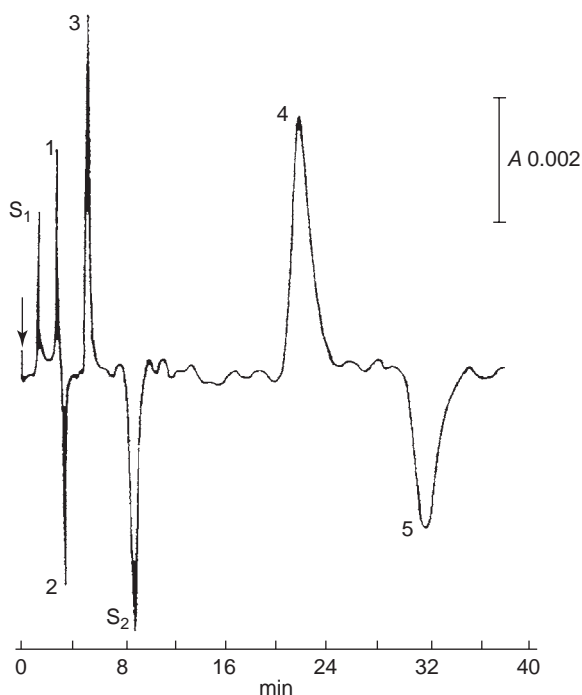
between the two liquid phases has limited the use of this technique. Also, here the introduction of the both universal and selective MS detector has given a new and superior alternative methodology.

In reversed-phase liquid-solid chromatography an indirect detection technique has been applied to charged and noncharged analytes. An ionic component with affinity for the adsorbing phase and with detection properties is included in the aqueous mobile phase. The equilibrium of the detectable ion, the probe, between the mobile and stationary phase will become disturbed by the injection of the analytes into the system. There will be one peak for each of the solutes and at least one additional system peak. When ionic solutes and the probe have opposite charges, peaks eluting before the system peak in the chromatogram have negative direction and those after have positive direction, while this is reversed with ionic solutes of the same charge as the probe. An example is given in **Figure 2**. Apart from the detector response of the probe ion, its fractional coverage of the adsorbent and the relative retention of the solute will influence the detector response. UV-absorbing ions such as naphthalene-2-sulfonate and 1-phenethyl-2-picolinium have been used as probes.

A narrow application field for ion pair chromatography has been the separation of enantiomers in normal-phase liquid-solid systems. Chiral counterions are then used as selectors and their ability to interact with the ionic solute is decisive for the enantioselectivity. (+)-10-Camphorsulfonic acid and a dipeptide derivate, *N*-benzoxycarbonylglycyl L-proline, have been used as ion pairing agents for the resolution of enantiomers of amino alcohols, and quinine and cinchonidine for the separation of enantiomers of chiral carboxylic acids. In recent years, efficient chiral separations employing the ion pair technique on a hydrophobic stationary phase have been demonstrated using different N-blocked dipeptides as counter ions for the resolution of enantiomers of amino alcohols. These systems consist of porous graphitized carbon, Hypercarb, as the solid phase and methanol containing the chiral selector as the mobile phase.

### Selected Applications

Ion pair LC in the reversed-phase mode is the overall dominating separation technique, not least within the biomedical field. Many compounds of interest occur in ionic form or are ionizable, and chromatographic separation can be performed by counterion-aided distribution. The range of applications is extremely wide and includes small hydrophilic aminophenols like catecholamines, carboxylic acids,

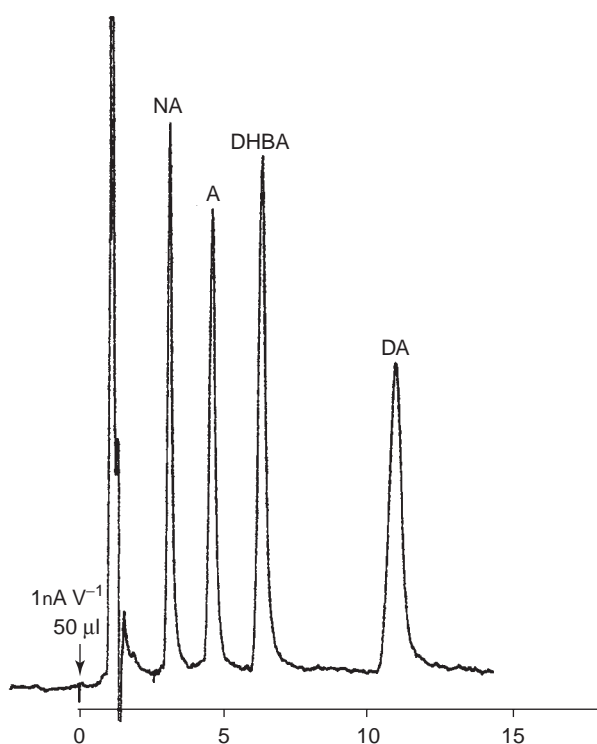


**Figure 2** Separations of anionic and cationic compounds with naphthalene-2-sulfonate in the mobile phase. A, absorbance. Stationary phase:  $\mu$ Bondapak Phenyl. Mobile phase: NS  $4 \times 10^{-4} \text{ mol l}^{-1}$  in  $0.05 \text{ mol l}^{-1}$  phosphoric acid. UV detection at 254 nm. Peaks: 1, pentanesulfonate; 2, di-2-propylamine; 3, hexanesulfonate; 4, heptylamine; 5, octanesulfonate; S<sub>1</sub> and S<sub>2</sub> are system peaks. (Reproduced with permission from Hackzell L and Schill G (1982) Detection by ion-pairing probes in reversed-phase liquid chromatography. *Chromatographia* 15: 437–444; © Vieweg-Publishing.)

hydrophobic tricyclic amines, amino acids, peptides, nucleic acid derivatives, and proteins to mention a few groups of compounds normally present in ionic form and subjected to ion pair chromatography.

Catecholamines can be retained and separated on an octadecyl-bonded phase in the presence of a hydrophobic counterion like C<sub>8</sub>-sulfate in the mobile aqueous phase (Figure 3). For this kind of system it has been shown that the degree of retention is largely correlated to the amount of counterion adsorbed onto the solid phase, forming a charged surface there. The porous graphitic carbon stationary phase is very hydrophobic and particularly suited and used for polar compounds. Reports indicate that on this stationary phase, relative to the silica-based bonded phases, much less of hydrophobic counterions is accumulated on the solid phase, which influences the retention model and also the ranking between different counterions and the effect of the organic modifier. An example of separation of polar guanidino compounds on a graphite column is shown in Figure 4.

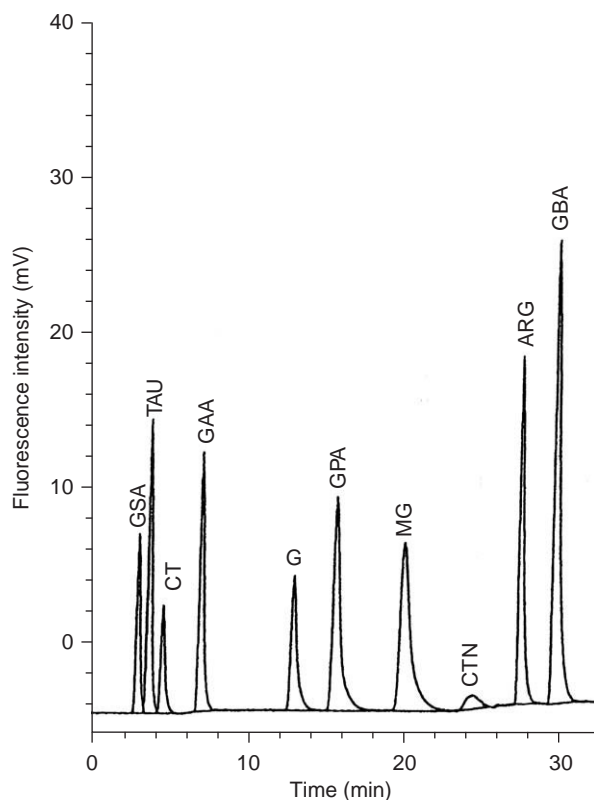
The separation of moderate to strongly hydrophobic amines and ammonium compounds on



**Figure 3** Separation of noradrenaline (NA), adrenaline (A), dihydroxybenzylamine (DBH), and dopamine (DA), 0.65 pmol each. Stationary phase: LiChrosorb RP18, 5  $\mu\text{m}$ . Mobile phase: 3,5-dimethylcyclohexyl sulfate  $2 \times 10^{-3} \text{ mol l}^{-1}$  and 8% of methanol in citrate-acetate buffer solution, pH 5.2. Electrochemical detection at +0.7 V.

silica-bonded phases can often be made with a hydrophilic counterion or just dihydrogenphosphate as both counterion and pH-buffering electrolyte. As mentioned above, many hydrophobic amine solutes showed a chromatographic behavior that was not very satisfactory but it could be corrected by including an aliphatic amine modifier in the mobile aqueous phase. However, the improved properties of modern reversed-phase materials have largely made such precautions unnecessary. For ion pair chromatography of amines, a low pH of 2–3 in the aqueous mobile phase is often used but higher pH, up to neutral, is also possible, e.g., ammonium acetate buffers often used in LC–MS methods.

For the separation of anions of weak protolytes such as carboxylates, a pH close to neutral is chosen to ensure that the solutes are in ionic form. Quaternary ammonium ions such as tetrabutylammonium are most often used as counterions, as in the example given in Figure 5, which shows the separation of benzoic acid and three structurally related carboxylic acids. A mixture of acetonitrile and phosphate buffer of pH 6.0 is the mobile phase. These few applications illustrate the kind of phase systems being used with anionic or cationic counterions. The hydrophobicity

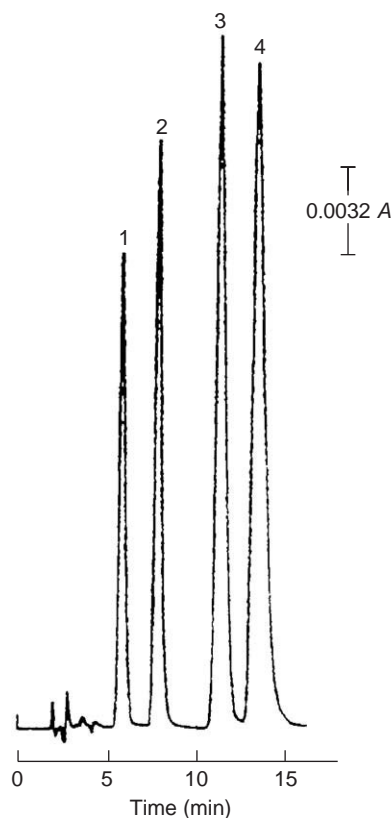


**Figure 4** Separation of 10 guanidino compounds on a porous graphitic carbon column in ion pair LC. Column:  $50 \times 4.6$  mm ID packed  $3.5 \mu\text{m}$  graphitic carbon. Eluent: four step-wise gradient from  $10 \text{ mmol l}^{-1}$  sodium citrate buffer (pH 4.50) containing  $5 \text{ mmol l}^{-1}$  sodium octanesulfonate to  $10 \text{ mmol l}^{-1}$  sodium citrate buffer (pH 4.50) containing 20% (v/v) acetonitrile; flow rate:  $0.8 \text{ ml min}^{-1}$  at  $40^\circ\text{C}$ . Flow rate of 0.6% ninhydrin and  $1 \text{ mol l}^{-1}$  NaOH solution:  $0.2 \text{ ml min}^{-1}$ . Fluorescence detection: ex.  $392 \text{ nm}$  and em.  $500 \text{ nm}$ . Peaks: GSA, guanidino succinate; TAU, taurocyamine; CT, creatine; GAA, glycosylamine; G, guanidine hydrochloride; GPA,  $\beta$ -guanidino propionic acid; MG, methylguanidine; CTN, creatinine; ARG, arginine; GBA,  $\gamma$ -guanidino butyric acid. (Reprinted with permission from Inamoto Y, Inamoto S, Hanai T, Tokuda M, Hatase O, Yoshii K, Sugiyama N, and Kinoshita T (1998) *Journal of Chromatography B* 707: 111–120; © Elsevier.)

and concentration of the counterion and the nature and content of organic modifier in the aqueous mobile phase are the parameters that can be selected according to the character of the ionic solute and can be combined to affect the selectivity.

## Problems of Analysis

A major application field for reversed-phase LC is ionic or ionizable organic compounds. The addition of a counterion to the aqueous mobile phase to aid the distribution and retention of the ionic solute does not demand ion pair LC of any special precautions or create any separate problems. The use of a guard column, exchanged frequently enough, to protect the



**Figure 5** Separation of carboxylates as ion pairs with tetrabutylammonium (TBA); A, absorbance. Stationary phase: LiChrosorb RP8,  $10 \mu\text{m}$ . Mobile phase: TBA  $3.2 \times 10^{-2} \text{ mol l}^{-1}$  in acetonitrile–phosphate buffer, pH 6.0 (20+80). UV detection at  $254 \text{ nm}$ . Peaks: 1, 4-hydroxybenzoic acid; 2, 3-hydroxybenzoic acid; 3, acetylsalicylic acid; 4, benzoic acid. (Reprinted with permission from Tilly-Melin A, Askemark Y, Wahlund K-G, and Schill G (1979) *Analytical Chemistry* 51: 976–983; © American Chemical Society.)

analytical column is as obvious as in other forms of LC. Temperature control is advantageous in all types of chromatography. It is necessary in liquid–liquid ion pair chromatography and is crucial in systems with an ionic probe and indirect detection. For regular separations ambient temperature is most often sufficient.

The solvent or solution used for the sample to be introduced into the separation column has to be compatible with the aqueous mobile phase. If it is too strong an eluent and injected in relatively large volumes, peak deformation may occur as in other kinds of chromatography. The addition to the sample solution of a hydrophilic counterion or a strongly competing ionic component of the same charge as the ionic solute may give severe peak distortion, but peak compression under optimized conditions. Column deterioration is probably no more pronounced in ion pair LC than in other systems with bonded stationary phases and mixtures of aqueous buffer solutions and

organic modifiers as mobile phases. Hydrochloric acid is known to be aggressive to all kinds of solid phases and is avoided. Aqueous phase at pH above neutral are believed to affect the bonded phase packing and there are indications that the presence of a quarternary ammonium ion such as tetrabutylammonium potentiates the degradation. Intermittent reports of decreased column stability in ion pair LC can be linked to the nature of injected samples, such as the degree of soluble impurities and microparticulate material in the solution. A guard column should take care of most of this.

From a general point of view, ion pair LC systems can be expected to give a performance and durability comparable to other bonded-phase systems. However, a small proportion of the ion pair reagent may become irreversibly bound to the stationary phase, possibly to free silanol groups, and if so the column cannot be regenerated to its original condition. In practice, this means that once a column has been used for an ion pair separation it should be retained for the specific analysis in question. Attempts to reuse the column for conventional reversed-phase chromatography may lead to unexpected and spurious results.

**See also:** **Ion Exchange:** Principles. **Liquid Chromatography:** Normal Phase; Reversed Phase; Chiral.

## Further Reading

Chaimbault P, Petritis K, Elfakir C, and Dreux M (2000) Ion-pair chromatography on a porous graphitic carbon

stationary phase for the analysis of twenty underivatized protein amino acids. *Journal of Chromatography A* 870: 245–254.

Chen JG, Weber SG, Glavina LL, and Cantwell FF (1993) Electrical double-layer models of ion-modified (ion-pair) reversed-phase liquid chromatography. *Journal of Chromatography A* 656: 549–576.

Hanai T (2003) Separation of polar compounds using carbon columns. *Journal of Chromatography A* 989: 183–196.

Hearn MTW (ed.) (1985) *Ion Pair Chromatography*, Chromatographic Science Series, vol. 31. New York: Dekker.

Karger BLK, Le Page JN, and Tanaka N (1980) Secondary equilibria in high-performance liquid chromatography. In: Horváth C (ed.) *High Performance Liquid Chromatography: Advances and Perspectives*, vol. 1. New York: Academic Press.

Persson BA and Lagerström PO (1990) Packings and stationary phases for ion pair chromatography. In: Unger KK (ed.) *Packings and Stationary Phases in Chromatographic Techniques*. New York: Dekker.

Pettersson C, Zief M, and Schill G (1988) Enantiomer separation in ion pairing systems. In: Zeif S and Crane LJ (eds.) *Chromatographic Chiral Separations*, Chromatographic Science Series, vol. 40. New York: Dekker.

Schill G (1974) Isolation of drugs and related organic compounds by ion-pair extraction. In: Marinsky JA Marcus Y (eds.) *Ion Exchange and Solvent Extraction*, vol., pp. 1–57. New York: Dekker.

Schill G and Arvidsson E (1989) Application of indirect detection methods in biomedical analysis. *Journal of Chromatography* 429: 299–318.

Ståhlberg J (1999) Retention models for ions in chromatography. *Journal of Chromatography A* 855: 3–55.

## Micellar

**M C García-Alvarez-Coque and J R Torres-Lapasió**, University of Valencia, Valencia, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

In the late 1970s, ionic surfactants were added for the first time to polar aqueous–organic mobile phases in reversed-phase liquid chromatography (RPLC) to form ion pairs (IPs). In IP chromatography, surfactant monomers adsorbed on a bonded silica-based stationary phase associate with neutral and ionic solutes, modifying their retention. The concentration of surfactant in the mobile phases is kept below the critical micellar concentration (CMC). In 1980, Daniel W. Armstrong reported the possibility of using

solutions containing surfactants above the CMC for separation purposes. The technique was called pseudo-phase liquid chromatography, adopting later the term micellar liquid chromatography (MLC).

Some qualities of the new technique were outlined in the first descriptions: the relatively low cost, the ease of operation, and the less time required for a complete run. Other qualities were revealed as the technique was developed: the possibility of analyzing solutes presenting a wide range of polarities within a single isocratic run, the high reproducibility of retention times, and their accurate prediction through modeling. The latter feature facilitates the optimization of the separation conditions. The existence of micelles in the mobile phase and the adsorption of surfactant monomers on the stationary phase produce notable changes in the chromatographic

organic modifiers as mobile phases. Hydrochloric acid is known to be aggressive to all kinds of solid phases and is avoided. Aqueous phase at pH above neutral are believed to affect the bonded phase packing and there are indications that the presence of a quarternary ammonium ion such as tetrabutylammonium potentiates the degradation. Intermittent reports of decreased column stability in ion pair LC can be linked to the nature of injected samples, such as the degree of soluble impurities and microparticulate material in the solution. A guard column should take care of most of this.

From a general point of view, ion pair LC systems can be expected to give a performance and durability comparable to other bonded-phase systems. However, a small proportion of the ion pair reagent may become irreversibly bound to the stationary phase, possibly to free silanol groups, and if so the column cannot be regenerated to its original condition. In practice, this means that once a column has been used for an ion pair separation it should be retained for the specific analysis in question. Attempts to reuse the column for conventional reversed-phase chromatography may lead to unexpected and spurious results.

**See also:** **Ion Exchange:** Principles. **Liquid Chromatography:** Normal Phase; Reversed Phase; Chiral.

## Further Reading

Chaimbault P, Petritis K, Elfakir C, and Dreux M (2000) Ion-pair chromatography on a porous graphitic carbon

stationary phase for the analysis of twenty underivatized protein amino acids. *Journal of Chromatography A* 870: 245–254.

Chen JG, Weber SG, Glavina LL, and Cantwell FF (1993) Electrical double-layer models of ion-modified (ion-pair) reversed-phase liquid chromatography. *Journal of Chromatography A* 656: 549–576.

Hanai T (2003) Separation of polar compounds using carbon columns. *Journal of Chromatography A* 989: 183–196.

Hearn MTW (ed.) (1985) *Ion Pair Chromatography*, Chromatographic Science Series, vol. 31. New York: Dekker.

Karger BLK, Le Page JN, and Tanaka N (1980) Secondary equilibria in high-performance liquid chromatography. In: Horváth C (ed.) *High Performance Liquid Chromatography: Advances and Perspectives*, vol. 1. New York: Academic Press.

Persson BA and Lagerström PO (1990) Packings and stationary phases for ion pair chromatography. In: Unger KK (ed.) *Packings and Stationary Phases in Chromatographic Techniques*. New York: Dekker.

Pettersson C, Zief M, and Schill G (1988) Enantiomer separation in ion pairing systems. In: Zeif S and Crane LJ (eds.) *Chromatographic Chiral Separations*, Chromatographic Science Series, vol. 40. New York: Dekker.

Schill G (1974) Isolation of drugs and related organic compounds by ion-pair extraction. In: Marinsky JA Marcus Y (eds.) *Ion Exchange and Solvent Extraction*, vol., pp. 1–57. New York: Dekker.

Schill G and Arvidsson E (1989) Application of indirect detection methods in biomedical analysis. *Journal of Chromatography* 429: 299–318.

Ståhlberg J (1999) Retention models for ions in chromatography. *Journal of Chromatography A* 855: 3–55.

## Micellar

**M C García-Alvarez-Coque and J R Torres-Lapasió**, University of Valencia, Valencia, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

In the late 1970s, ionic surfactants were added for the first time to polar aqueous–organic mobile phases in reversed-phase liquid chromatography (RPLC) to form ion pairs (IPs). In IP chromatography, surfactant monomers adsorbed on a bonded silica-based stationary phase associate with neutral and ionic solutes, modifying their retention. The concentration of surfactant in the mobile phases is kept below the critical micellar concentration (CMC). In 1980, Daniel W. Armstrong reported the possibility of using

solutions containing surfactants above the CMC for separation purposes. The technique was called pseudo-phase liquid chromatography, adopting later the term micellar liquid chromatography (MLC).

Some qualities of the new technique were outlined in the first descriptions: the relatively low cost, the ease of operation, and the less time required for a complete run. Other qualities were revealed as the technique was developed: the possibility of analyzing solutes presenting a wide range of polarities within a single isocratic run, the high reproducibility of retention times, and their accurate prediction through modeling. The latter feature facilitates the optimization of the separation conditions. The existence of micelles in the mobile phase and the adsorption of surfactant monomers on the stationary phase produce notable changes in the chromatographic



behavior (retention, elution strength, peakshape, selectivity, and resolution).

Pure micellar solutions are generally useless as mobile phases, except for highly polar compounds, due to the weak elution strength and poor efficiencies. An organic solvent should be added to decrease the analysis times to acceptable values and enhance efficiencies. However, the concentration of the organic solvent needed is appreciably smaller than in RPLC with conventional organic solvent–water mixtures. In addition, the organic solvent is stabilized by the micellar medium. This decreases the risk of evaporation. Micellar mobile phases are thus stable for a longer time and their toxicity is reduced. Finally, direct injection into the chromatographic system is possible using conventional alkyl-bonded stationary phases. As a result, MLC procedures are faster and more economical.

The term MLC is usually given to the use of micellar mobile phases with RPLC columns. A few examples have been described in other chromatographic modes that use similar mobile phases, with normal or reverse microemulsions, bile salts, and surfactants in supercritical fluids. Also, studies using non-RPLC stationary phases and micellar mobile phases have been reported in size exclusion and gel permeation chromatography. These topics are beyond the scope of this article.

## The Chromatographic System

### The Mobile Phase

Instead of the organic solvent–water mixtures used conventionally in RPLC, the mobile phase in MLC consists of an aqueous solution of surfactant-forming micelles, or a ternary mixture of water, surfactant, and organic solvent (called hybrid micellar mobile phase). In the latter case, the concentration of organic solvent is maintained low enough to permit the formation of micelles; the micellization process is, however, altered.

Several surfactants of different types (nonionic, anionic, cationic, and zwitterionic) have been assayed, but most reported procedures use the anionic surfactant sodium dodecyl sulfate (SDS,  $\text{CMC} = 8.2 \times 10^{-3} \text{ mol l}^{-1}$  at  $25^\circ\text{C}$ ), which is readily available. Other common surfactants are the cationic cetyltrimethylammonium bromide (CTAB,  $\text{CMC} = 9 \times 10^{-4} \text{ mol l}^{-1}$ ), and the nonionic Brij-35 (polyoxyethylene (23) dodecyl ether,  $\text{CMC} = 1 \times 10^{-4} \text{ mol l}^{-1}$ ). Common solvents in RPLC (methanol, ethanol, propanol, acetonitrile, and tetrahydrofuran) are suitable modifiers. Other less polar solvents, such as butanol and pentanol, can

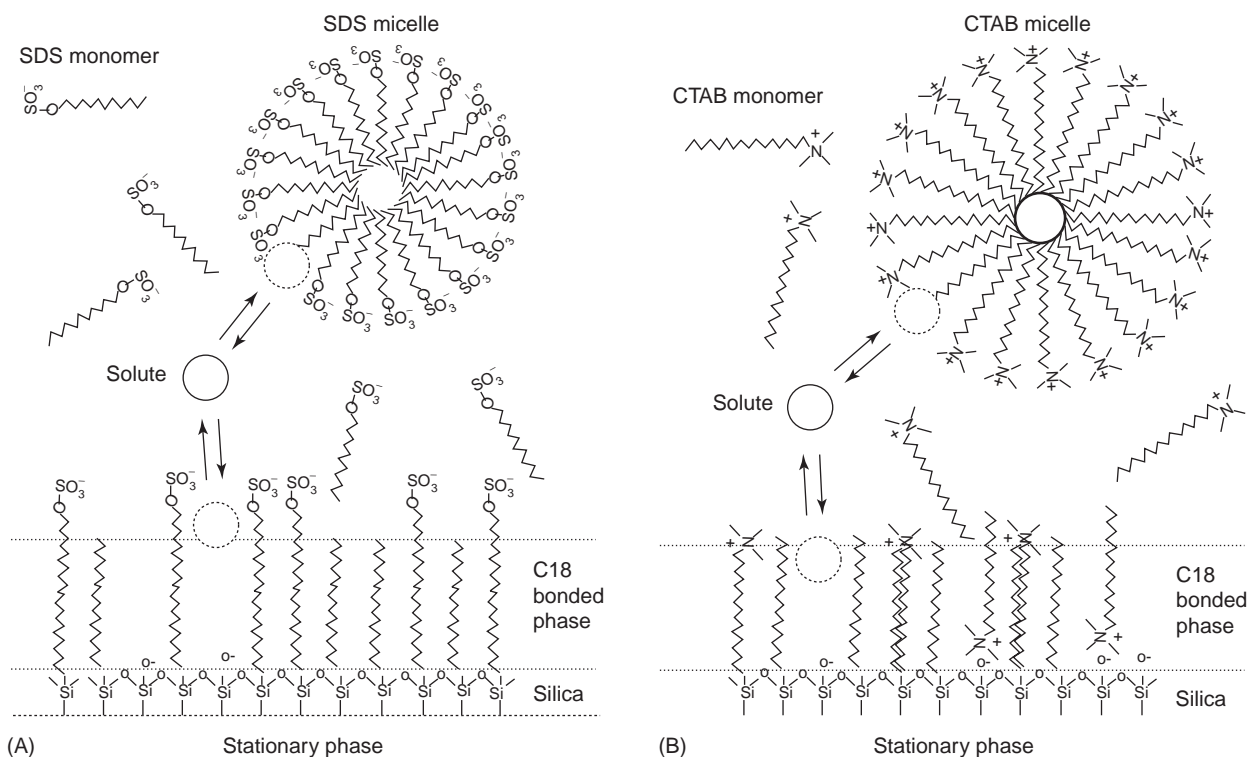
also be added up to useful concentration levels, owing to the solubilizing effect of micelles. Propanol (especially 1-propanol) is the most frequent, although acetonitrile has similar elution strength and usually improves peakshape. Butanol and pentanol, which are stronger solvents, permit the elution of apolar solutes at sufficiently short retention times.

The extreme concentrations of surfactant and organic solvent are imposed by practical limitations in the chromatographic system. The lower concentration of surfactant must be well above the CMC. Its solubility and the viscosity of the resulting mobile phase determine the upper concentration. On the other hand, the concentration of organic solvent is conditioned by the maximal value that ensures the integrity of micelles. For SDS, the maximal concentration of acetonitrile, propanol, butanol, and pentanol should be approximately below 20, 15, 10, and 7% (v/v), respectively.

### The Stationary Phase

Separations are usually carried out in conventional octyl- and octadecylsiloxane (C8 and C18) columns. A significant number of surfactant molecules may be adsorbed on these stationary phases, giving rise to a structure similar to an open micelle. Consequently, column properties change radically, although the subjacent stationary phase (the bonded moiety) still plays a role in the interaction with solutes. Cyano-propylsiloxane-bonded columns are useful for some specific applications. Efficiencies are improved with fluorinated-bonded phases. More recently, ultrawide pore and monolithic octadecylsiloxane columns have been shown to enhance the eluting power of micellar mobile phases.

The amount of adsorbed anionic and cationic surfactants on the column increases rapidly at increasing surfactant concentration in the mobile phase, reaching a plateau nearby the CMC. This is not the case for nonionic surfactants, whose adsorption increases beyond the CMC. Nuclear magnetic resonance studies have indicated that on the densely grafted stationary phase, the hydrophobic chain of anionic surfactants is inserted in the bonded organic layer with the functional group protruding toward the mobile phase (Figure 1A). Cationic surfactants give rise to two kinds of interactions with the stationary phase: hydrophobic association with the alkyl-bonded layer like anionic surfactants, and electrostatic attraction of the positively charged surfactant heads to the residual free silanols, which are buried inside the thickness of the bonded layer (Figure 1B). This explains why the SDS-covered alkyl phases are more polar, with a negative charge, than the CTAB-covered phases, which bear positive charge.



**Figure 1** Solute interactions with stationary phase and micelle in pure MLC systems containing the surfactants: (A) SDS and (B) CTAB.

The adsorption of surfactant monomers changes the stationary phase polarity, structure, surface, and pore volume. Surfactant molecules coat the interior walls of the pores without completely filling them.

Ionic compounds added to the micellar system for pH or ionic strength adjustment increase usually the surfactant adsorption. Organic solvents, in contrast, decrease the amount of adsorbed surfactant. The alkyl chains of propanol and longer *n*-alcohols form on the stationary phase a monolayer similar to that of adsorbed surfactant molecules, with the hydroxyl group oriented toward the aqueous phase. Alcohol and surfactant molecules compete for adsorption sites. For ionic surfactants, the desorbing ability depends on the organic solvent polarity (methanol < ethanol < propanol < butanol < pentanol). Also, the amount of adsorbed surfactant molecules decreases linearly with the organic solvent concentration.

Some care is needed to preserve the column performance for long time periods of intensive MLC use, which can be comparable or even longer than in conventional RPLC. First, since most micellar solutions are able to dissolve minute amounts of silica, the mobile phase should be saturated in silica by placing a short precolumn before the injection valve. Second, the micellar solution should never stay motionless in a chromatographic system to avoid the formation of surfactant crystals that can clog the

system or ruin the column. When the system is not being used, the flow can be reduced to a minimal value (e.g.,  $0.1 \text{ ml min}^{-1}$ ). Finally, before stopping the flow, the column should be cleaned with water to later allow the passage of pure methanol or a 75:25 methanol–propanol mixture for a few minutes. Operating in this way, the layer of SDS monomers can be completely stripped off the stationary phase. CTAB and Brij-35 cannot be removed so easily. With convenient experimental caution, hundreds of injections can be made without modification of the chromatographic system or pressure buildup.

The existence of a plateau of adsorbed ionic surfactant above the CMC permits rapid analyses using a micelle gradient. Beyond the CMC, changes in total surfactant concentration only change the concentration of micelles in the mobile phase; the stationary phase is not affected. This means that the initial conditions can be recovered without re-equilibration time. However, normally, the particular elution strength behavior of MLC makes gradient elution unnecessary.

## Partitioning Behavior

MLC is a good example of the use of secondary equilibria in liquid chromatography. Solutes are partitioned between the mobile and stationary phases,

which are modified by the presence of surfactant monomers and micelles. Depending on polarity and steric factors, solutes can remain outside the micelle, associate to the polar head of the surfactant, form a part of the outer palisade layer, or penetrate into the micelle core. Organic solvents added to the micellar solution experience the same interactions: the association of acetonitrile and methanol with micelles is weak, while butanol and pentanol are inserted in the intermonomer spaces of the micelle palisade.

Neutral solutes interact hydrophobically with micelles. An increase in the concentration of surfactant in the mobile phase results in decreased retention. Water-insoluble apolar solutes exhibit a large affinity for the micelles and the surfactant-coated stationary phase, and, consequently, can be only transported between them by direct transfer. By addition of organic solvent to the mobile phase, the solubility of the apolar solutes increases and this transfer mechanism loses importance with regard to the usual two-step mechanism (micelle–bulk solvent and bulk solvent–stationary phase) (Figure 1).

Charged solutes experience hydrophobic and electrostatic interactions with ionic surfactants. Two distinct situations are possible:

- Charges on solute and surfactant are of the same sign. Electrostatic repulsion from the surfactant-modified stationary phase should cause a low retention (the solute may be even eluted with the void volume), unless sufficiently strong hydrophobic interaction is established. A particular behavior has been observed with methylsiloxane (C1) and cyanopropylsiloxane-bonded phases, where the amount of adsorbed surfactant is low: repulsion of solutes from micelles is dominant and the retention increases at increasing micelle concentration (antibinding behavior). In this case, the electrical double layer surrounding the micelles can be narrowed in a solution containing higher concentration of ions, which facilitates the approximation of the solute to the micelle assembly to establish hydrophobic interactions. By increasing the ionic strength, antibinding solutes can thus adopt nonbinding or even binding behavior.
- Charges on solute and surfactant are opposite. Electrostatic and hydrophobic interactions with the stationary phase may be sufficiently large to offset the increase in micelle attraction, and retention will be strong.

## Correlations between Retention and Polarity

MLC has demonstrated to be useful for measuring the polarity of compounds. The regular linear

increase in  $\log k$  with the number of carbon atoms,  $n_C$ , in a homologous series, observed in aqueous–organic RPLC is not usually valid with micellar mobile phases;  $k$  versus  $n_C$  is linear instead. This behavior has been explained by the location in the micelle of different members in the homologous series. The more hydrophobic homologues tend to locate in more apolar microenvironments, experiencing smaller change in polarity upon transfer from the micelle to the stationary phase.

Good correlation has been found between retention in MLC and the logarithm of octanol–water partition coefficient,  $\log P_{o/w}$ , for series of neutral compounds of diverse polarities. The type of general relationship ( $\log k$  versus  $\log P_{o/w}$  or  $k$  versus  $\log P_{o/w}$ ) seems to depend on the selected set of compounds and the characteristics of both the mobile and stationary phases. The linearity is improved by addition of alcohols. The organic solvent apparently provides an environment closer to the octanol–water mixture than pure micellar systems. Correlation of retention with  $\log P_{o/w}$  is poorer for ionizable compounds, due to the extra ionic interactions. However, acceptable correlation has been found for groups of basic compounds in acidic mobile phases of anionic surfactant, where the positively charged protonated species dominate. The retention of cationic solutes is superior than expected from the correlation established for nonionizable compounds, which are retained due only (or mainly) to hydrophobic interactions.

Alternatively, retention can be predicted from solute  $\log P_{o/w}$  values. The most suitable organic solvent to be used as modifier of the mobile phase should be chosen according to the polarity of the eluted compound. For SDS, a low propanol content ( $\sim 1\%$ , v/v) is useful to separate compounds with  $\log P_{o/w} < -1$ , such as amino acids. A larger amount of propanol ( $\sim 5\text{--}7\%$ ) is needed for compounds in the range  $-1 < \log P_{o/w} < 2$ , such as diuretics and sulfonamides. Other alcohols ( $< 10\%$  butanol or  $< 6\%$  pentanol) are required for apolar compounds with  $\log P_{o/w} > 3$ , such as steroids. This rule of thumb is, however, not always valid: propanol is too weak for cationic solutes, such as phenethylamines ( $0 < \log P_{o/w} < 1.7$ ) or  $\beta$ -blockers ( $1 < \log P_{o/w} < 3$ ). For these compounds, the additional electrostatic attraction to the anionic surfactant molecules adsorbed on the stationary phase makes a stronger solvent necessary.

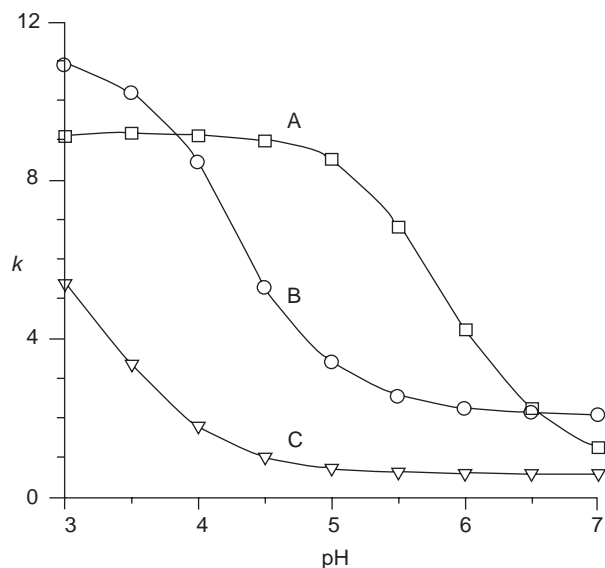
Micelles are recognized as simple biomembrane models. Biomembrane and micelles have amphiphilic properties and are anisotropic, providing hydrophobic and electrostatic interaction sites. MLC combines the unique characteristics of micelles and the capabilities of RPLC in quantitative retention–activity relationships studies.

## Effect of pH on Retention

Chromatographic retention of ionizable compounds depends on the pH of the mobile phase. Micelles and organic solvent shift the protonation equilibria, due to the modification of the polarity of bulk solvent and the association of solutes with micelles. Since the intrinsic retention of both acid–base species is different, a sudden change in retention time will happen at pH values close to the logarithm of the apparent protonation constant in the mobile phase medium,  $K'_H$ . The full protonation process covers several pH units and is shifted more or less depending on the concentrations of surfactant and organic solvent. The change in retention is not properly observed when the working pH of the column is narrow (e.g., pH 3–7 for conventional alkyl-bonded stationary phases). In this case, retention versus pH plots may show different patterns (Figure 2).

Usually,  $K'_H$  diminishes as the organic solvent concentration increases. For ionic surfactants, solute–surfactant electrostatic interactions are responsible for the shift in sign in  $K'_H$  at increasing surfactant concentration. These interactions also explain the sign in the trend of the  $k$ –pH dependence, which is sigmoidal and resembles conventional acid–base titration curves. The observed behaviors indicate that the interaction of solutes with the surface of the surfactant-modified stationary phase is stronger than that with micelles.

For weak bases having neutral acidic and anionic basic species (or weak bases having cationic acidic and neutral basic species),  $K'_H$  increases with



**Figure 2** Retention versus pH plots for: (A) xipamide, (B) ethacrynic acid, and (C) tyrosine. Chromatographic system: ODS-2/0.10 mol l<sup>-1</sup> SDS/4% (v/v) 1-propanol.

surfactant concentration for anionic surfactants, since the acidic species is stabilized by interaction with micelles and adsorbed monomers. Also, retention decreases with pH, due to the weaker association of the basic species with the surfactant-modified stationary phase (Figure 2). The shift in  $K'_H$  to higher pH benefits the observation of the maximal retention (the retention of the acidic species) within the operable limits of silica-based columns. When a cationic surfactant is used instead, both the shift in  $K'_H$  and the  $k$  versus pH dependence are opposite to those observed for anionic surfactants, since cationic surfactant molecules stabilize the basic species.

The pH of the micellar mobile phase is usually buffered using the phosphoric or citric acid–base system. Potassium ion cannot be used with SDS as potassium dodecyl sulfate precipitates from aqueous solutions. Only one peak is observed for each solute in the chromatograms because acid–base equilibria are much faster than solute–micelle or solute–stationary phase dynamics.

## Modeling of Retention

### Pure Micellar Mobile Phases

The retention behavior is the final result of the competition of the equilibria existing in the separation system. In MLC, solutes partition between three environments: water, micelle, and stationary phase (Figure 1). If the solute interacts with the micelle, retention is explained by

$$\frac{V_e - V_0}{V_s} = \frac{k}{\phi} = \frac{P_{AS}}{1 + v(P_{AM} - 1)[M]} \quad [1]$$

where  $V_e$  represents the total volume of mobile phase needed to elute the solute from the column,  $V_s$  the volume of active surface of the stationary phase,  $V_0$  the column void volume,  $\phi = V_s/V_0$  the phase ratio,  $v$  the partial specific volume of surfactant monomers in the micelle,  $P_{AS}$  and  $P_{AM}$  are the partition coefficients between water and stationary phase, and between water and micelle, respectively, and  $[M]$  is the concentration of micellized surfactant (total concentration of surfactant minus CMC).

For  $[M] = 0$ , eqn [1] reduces to the partition equation of aqueous–organic RPLC:

$$V_e = V_0 - V_s P_{AS} \quad [2]$$

A more practical model can be written by considering the interactions of solute inside the column in terms of association equilibria of solute in bulk solvent, with stationary phase binding sites and



surfactant monomers in micelles:

$$k = \frac{\phi P_{AS}}{1 + K_{AM}[M]} \quad [3]$$

which results in the following linear model:

$$\frac{1}{k} = \frac{1}{K_{AS}} + \frac{K_{AM}}{K_{AS}}[M] = c_0 + c_1[M] \quad [4]$$

where  $K_{AS} = \phi P_{AS}$ ;  $K_{AM}$  is the association constant of the solute with a surfactant monomer in the micelle and should be multiplied by the aggregation number to be referred to the whole micelle. MLC provides a convenient mode of estimating  $K_{AM}$ , since solute concentration need not be known, the impurities in the sample are separated, and the chromatographic process makes the simultaneous determination of the constants for several solutes possible.

Equations [1], [3], and [4] have been checked to be valid for apolar, uncharged polar, and ionic solutes chromatographed with anionic, cationic, and nonionic surfactants, in octylsiloxane- (C8), octadecylsiloxane- (C18), and cyanopropylsiloxane-bonded columns.

### Hybrid Micellar Mobile Phases

The addition of organic solvent to the micellar mobile phase shifts the partition equilibria of the solute in bulk solvent with both the stationary phase and the micelle. The retention is described by

$$k = \frac{K_{AS}[(1 + K_{SD}\phi)/(1 + K_{AD}\phi)]}{1 + K_{AM}[(1 + K_{MD}\phi)/(1 + K_{AD}\phi)][M]} \quad [5]$$

$\phi$  being the volumetric fraction of organic solvent. The constants  $K_{AD}$ ,  $K_{SD}$ , and  $K_{MD}$  measure the relative variation in the concentration of solute in bulk water, stationary phase, and micelle, respectively, due to the diminution in bulk solvent polarity and the modification of the stationary phase and micelle produced by the organic solvent. The  $K_{SD}$  term is only needed for highly apolar solutes, which are strongly associated to the stationary phase. In other cases,  $K_{SD} = 0$  and eqn [5] can be simplified to

$$k = \frac{1}{c_0 + c_1[M] + c_2\phi + c_{12}[M]\phi} \quad [6]$$

The analogous effects of organic solvent on both microenvironments (stationary phase and micelle) are evident in the parallel variations of solute-stationary phase and solute-micelle partition coefficients, as the concentration of organic solvent changes.

### Simultaneous Effect of Surfactant, Organic Solvent, and pH

Very often, chromatographic separations are carried out in a buffered medium. In this case, eqns [4]–[6] are appropriate to describe the retention of ionizable solutes. For compounds exhibiting acid–base behavior, retention in pure micellar medium is a weighted mean of the retention of the basic and acidic species. At varying pH,

$$k = \frac{k_A + k_{HA}K'_Hb}{1 + K'_Hb} \quad [7]$$

where  $k_A$  and  $k_{HA}$  are the retention factors of the basic and acidic species, and  $b$  is the concentration of hydrogen ion. When the simultaneous effect of the three factors (surfactant, organic solvent, and pH) is considered, the retention is given by

$$k = \frac{K_{AS} \frac{1 + K_{SD}\phi}{1 + K_{AD}\phi} + K_{HAS} \frac{1 + K_{HSD}\phi}{1 + K_{HAD}\phi} K_Hb}{\left(1 + K_{AM} \frac{1 + K_{MD}\phi}{1 + K_{AD}\phi} [M]\right) + \left(1 + K_{HAM} \frac{1 + K_{HMD}\phi}{1 + K_{HAD}\phi} [M]\right) K_Hb} \quad [8]$$

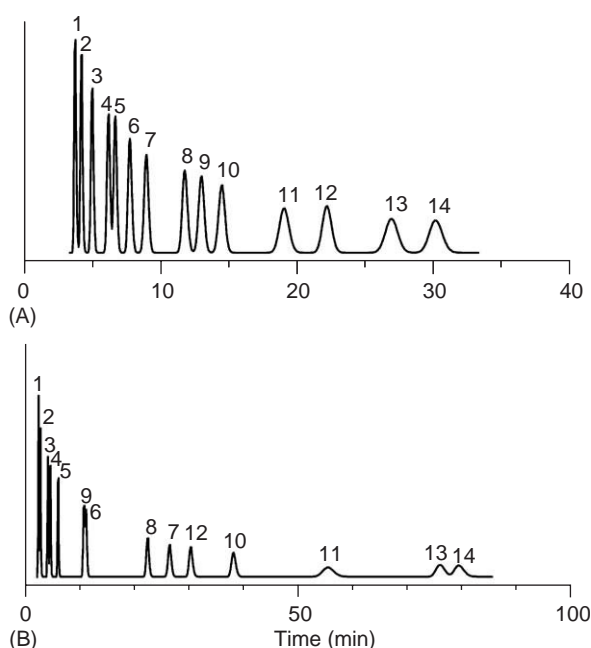
$K_{AS}$ ,  $K_{AM}$ ,  $K_{SD}$ ,  $K_{AD}$ , and  $K_{MD}$  are equilibrium constants associated to the basic species, and  $K_{HAS}$ ,  $K_{HAM}$ ,  $K_{HSD}$ ,  $K_{HAD}$ , and  $K_{HMD}$  are constants associated to the acidic species.

### Peak Shape

The thick surfactant film formed on the surface of the organic bonded layer is partially responsible for the loss of efficiency (i.e., broad and tailing peaks) observed when pure micellar mobile phases are used, with regard to aqueous–organic mixtures, which is especially true for apolar compounds. Peak Shape deteriorates at larger surfactant concentration. This is a major drawback that hindered the development of MLC. The addition of a small amount of propanol was first proposed to remediate this problem. Other organic solvents, such as butanol, pentanol, and acetonitrile, were later used with different degrees of success. Efficiency increases in some cases dramatically, and for polar compounds it becomes often similar (or eventually larger) to that achieved in classical RPLC. Peak tailing is also partially or totally suppressed.

The reason for the favorable behavior produced by the addition of organic solvent is the reduction of the viscosity and thickness of adsorbed surfactant on the stationary phase, and the increase in solute–micelle exit rate constants. This enhances the rate of solute mass-transfer between bulk solvent and both the stationary phase and micelles. Higher temperature decreases also the viscosity, and increases the chemical reaction rates.



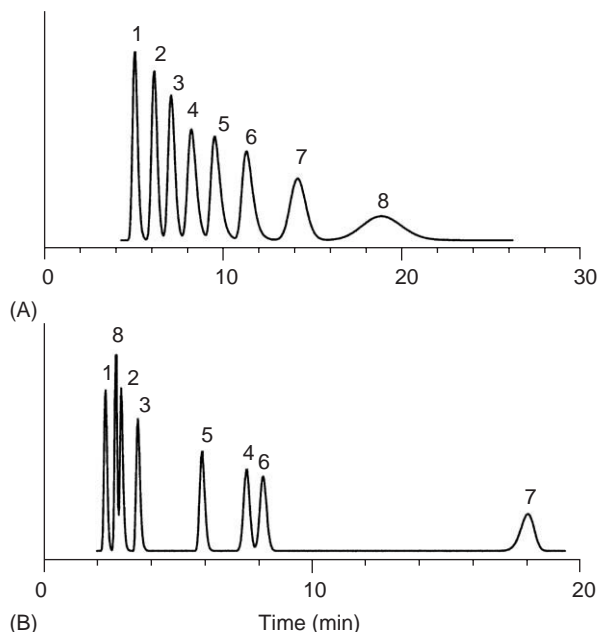


**Figure 3** Chromatogram of a mixture of 14  $\beta$ -blockers. Chromatographic systems: (A) ODS-2/0.10 mol l<sup>-1</sup> SDS/15% (v/v) 1-propanol, and (B) X-Terra MS/15% (v/v) acetonitrile. Compounds: (1) atenolol, (2) sotalol, (3) carteolol, (4) nadolol, (5) pindolol, (6) acebutolol, (7) celiprolol, (8) esmolol, (9) timolol, (10) bisoprolol, (11) labetalol, (12) oxprenolol, (13) propranolol, and (14) alprenolol.

Basic compounds are difficult to quantify by classical RPLC due to the ionic interaction of the positively charged species with free silanol groups of the packings. Ion exchange can be avoided, at least partially, by lowering the pH of the mobile phase to suppress the ionization of silanols. This process can be also blocked by addition of amine modifiers (mainly tertiary amines) that associate to silanol sites. Efficiency and peak tailing for several basic compounds, such as phenethylamines, tetracyclines, and  $\beta$ -blockers, eluted from conventional octadecylsiloxane columns with hybrid SDS mobile phases, are thus similar to those achieved with polymeric columns and aqueous-organic mixtures (Figure 3). The thinner surfactant layer on the modified column, in the presence of organic solvent, permits good diffusion of solutes. Also, the interaction of the protonated basic compound with the anionic heads of adsorbed SDS monomers reduces its penetration depth into the bonded phase. The kinetics of this electrostatic association seems to be easier than ion-exchange processes involving the silanols on the silica surface.

## Separation Performance

In MLC, the retention behavior (elution strength and selectivity) can be quite different from the observed



**Figure 4** Chromatogram of a mixture of eight steroids. Chromatographic systems: (A) ODS-2/0.12 mol l<sup>-1</sup> SDS/6% (v/v) 1-pentanol, and (B) ODS-2/52% (v/v) acetonitrile. Compounds: (1) dehydrotestosterone, (2) testosterone, (3) methyltestosterone, (4) medroxyprogesterone acetate, (5) dydrogesterone, (6) progesterone, (7) testosterone propionate, and (8) nandrolone.

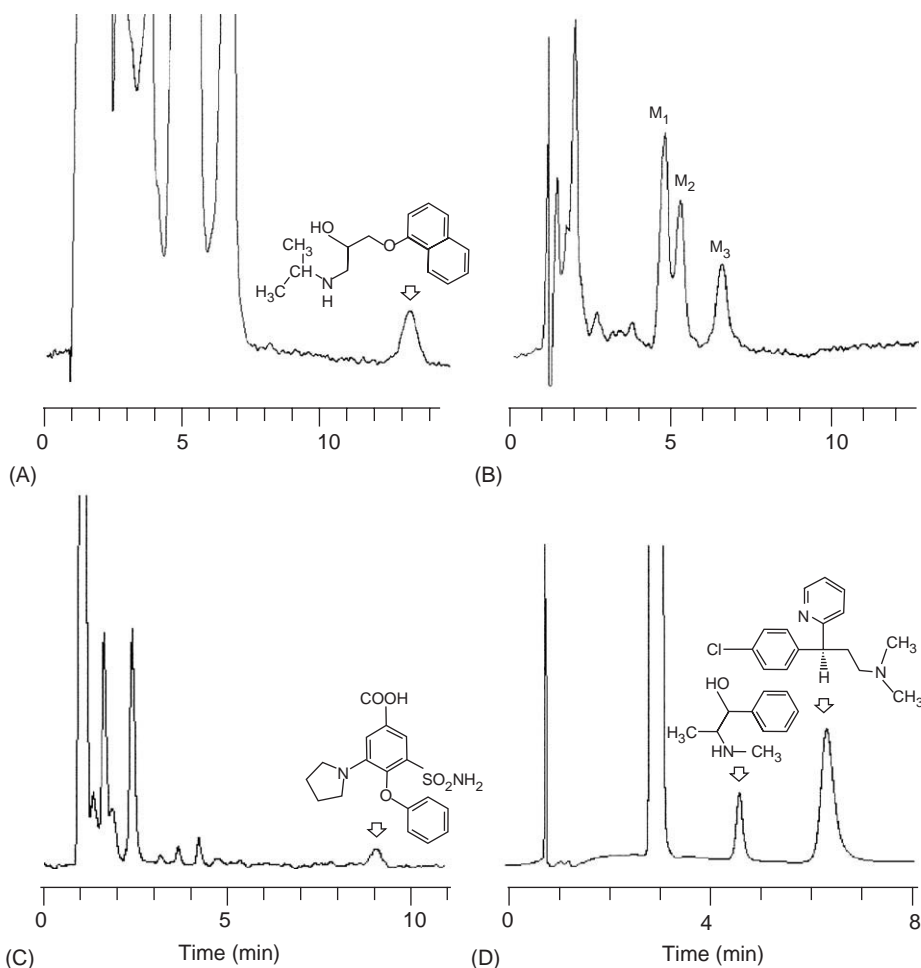
in classical RPLC, even after the addition of an organic modifier. The range of interactions provided by MLC is by far superior to that in aqueous-organic systems. This feature confers micellar mobile phases the ability of making cationic, anionic, uncharged polar, and apolar solutes compatible within the same chromatographic run. Furthermore, perhaps the main strength of MLC lies in the capability of performing the isocratic separation of mixtures of compounds exhibiting a relatively wide range of polarities (Figure 3).

The best resolution is usually found at low surfactant concentration due to the concurrence of longer retention times and better efficiencies. In contrast, moderate or high organic solvent concentrations (with the limitation of maintaining the micelles) are required, due mainly to the improvement in peak shape. For polar compounds, the elution strength of the surfactant is larger or similar to the organic solvent; the opposite situation is found for apolar compounds, which can be eluted at sufficiently low retention times using a small amount of butanol or pentanol in the micellar mobile phase (Figure 4).

The selectivity can be maximized by choosing the nature and adequate concentration of both the surfactant and organic modifier and, occasionally, the pH. The larger possibilities of tuning the

**Table 1** Representative examples of analytical use of micellar chromatographic systems

Analytes	Samples	Chromatographic system
Proteins	Myoglobin	C <sub>8</sub> /Neodol 91–6
Recombinant human growth hormone	<i>Escherichia coli</i>	C <sub>4</sub> /SDS
Folypolyglutamate hydrolase activity	Mouse kidney	C <sub>18</sub> /SDS
Fungicides	Pond water	Cyano/CTAB
Vanillins	Tobacco	C <sub>18</sub> /Brij-35
Maleic hydrazide	Tobacco	C <sub>18</sub> /CTAB
Berberine alkaloids	Coptis rhizome and pharmaceuticals	Phenyl/SDS
Antihistamines, benzodiazepines, $\beta$ -blockers, catecholamines, diuretics, phenethylamines, steroids, and sulfonamides	Pharmaceuticals	C <sub>18</sub> /SDS
Cephalosporins, theophylline, methotrexate, 6-mercaptopurine	Serum	C <sub>18</sub> /SDS
Antipyrine metabolites, diuretics	Plasma	C <sub>18</sub> /SDS
$\beta$ -Blockers, caffeine, diuretics, and sulfonamides	Urine	C <sub>18</sub> /SDS
Sulfonamides	Cow milk	C <sub>18</sub> /SDS



**Figure 5** Analysis of several drugs in urine (A–C) and in a cough-cold preparation (D): (A) propranolol (2 h after a 10 mg dose intake), determined by direct injection of a urine sample without dilution on 0.1 mol l<sup>−1</sup> SDS/15% propanol/1% triethylamine mobile phase and fluorimetric detection; (B) propranolol metabolites (M<sub>1</sub>–M<sub>3</sub>) determined after a 1:25 dilution of the previous sample; (C) piritanide (2 h after a 6 mg dose intake), determined after a 1:25 dilution of a urine sample on 0.055 mol l<sup>−1</sup> SDS/8% propanol and fluorimetric detection; (D) pseudoephedrine (preceding peak) and dexchlorpheniramine on 0.15 mol l<sup>−1</sup> SDS/6% pentanol and spectrophotometric detection.

selectivity, together with the more even distribution of peaks in the chromatograms, compensate often the poorer efficiencies achieved in MLC, making this technique competitive against classical RPLC. Actually, both chromatographic modes show complementary resolving capabilities. Equations [3]–[8] are useful for predicting the retention. Optimizations based on these models can be carried out using economical experimental designs.

## Analytical Applications

Most reported applications correspond to the assay of drugs in pharmaceutical preparations and physiological fluids. Some examples are given in **Table 1** and **Figure 5**. Micellar solutions can be used to dissolve samples or extract analytes. Apolar and polar compounds, derivatization reagents, and products can be co-solubilized. These media can also induce favorable shifts in the equilibrium constants and expedite reactions through micellar catalysis. Micelles improve the analytical performance of various spectroscopic (ultraviolet–visible spectrophotometry, fluorimetry, phosphorimetry, and atomic absorption) and electrochemical (amperometry) detection methods, by increasing the sensitivity, modifying the selectivity, and overcoming some problems associated with the use of aqueous–organic solvents. An interesting recent example is the use of MLC with nonionic surfactant to separate amino acids or cholesterol derivatives with postcolumn immobilized enzyme reactors. The micellar phases are gentler for the enzyme activity than aqueous–organic mobile phases or aqueous buffers.

One of the main appealing factors of MLC is the possibility of determining drugs in physiological fluids in a few minutes, without the need of previous separation of the proteins present in the samples. Micelles bind to proteins releasing bound drugs, whereas the proteins rather than precipitating on the column are solubilized and swept away, eluting with or shortly after the solvent front. However, although hundreds of repetitive serial injections are possible, the sample should be diluted when possible before performing the direct injection to keep the column operative for a longer time. The anionic SDS is the most common surfactant in these analyses, but the nonionic Brij-35 can also be employed. Cationic surfactants are not compatible because they cause the precipitation of proteins.

See also: **Chromatography: Overview. Liquid Chromatography: Overview; Principles; Reversed Phase; Ion Pair. Micellar Electrokinetic Chromatography. Surfactants and Detergents.**

## Further Reading

- Armstrong DW and Nome F (1981) Partitioning behavior of solutes eluted with micellar mobile phases in liquid chromatography. *Analytical Chemistry* 53: 1662–1666.
- Arunyanart M and Cline Love LJ (1984) Model for micellar effects on liquid chromatography capacity factors and for determination of micelle–solute equilibrium constants. *Analytical Chemistry* 56: 1557–1561.
- Berthod A (1997) Causes and remediation of reduced efficiency in micellar liquid chromatography. *Journal of Chromatography A* 780: 191–206.
- Berthod A and García-Alvarez-Coque MC (2000) *Micellar Liquid Chromatography*. New York: Marcel Dekker.
- Borgerding MF, Hinze WL, Stafford LD, Fulp GW Jr, and Hamlin WC Jr (1989) Investigation of stationary phase modification by the mobile phase surfactant in micellar liquid chromatography. *Analytical Chemistry* 61: 1353–1358.
- Caballero RD, Ruiz-Angel MJ, Simó-Alfonso E, and García-Alvarez-Coque MC (2002) Micellar liquid chromatography: Suitable technique for screening analysis. *Journal of Chromatography A* 947: 31–45.
- Esteve-Romero J, Gil-Agustí M, Carda-Broch S, Torres-Lapasió JR, and García-Alvarez-Coque MC (2001) Analysis of pharmaceuticals by micellar liquid chromatography. *Recent Research Developments in Pure & Applied Analytical Chemistry* 3: 157–168.
- García-Alvarez-Coque MC and Carda-Broch S (1999) Direct injection of physiological fluids in micellar liquid chromatography: A review. *Journal of Chromatography B* 736: 1–18.
- García-Alvarez-Coque MC and Torres-Lapasió JR (1999) Quantitation of hydrophobicity in micellar liquid chromatography. *Trends in Analytical Chemistry* 18: 533–543.
- García-Alvarez-Coque MC and Torres-Lapasió JR (2002) Hybrid micellar mobile phases. In: Cazes J (Ed.) *Encyclopedia of Chromatography*. New York: Marcel Dekker.
- García-Alvarez-Coque MC, Torres-Lapasió JR, and Baeza-Baeza JJ (1997) Modeling of retention behaviour of solutes in micellar liquid chromatography. *Journal of Chromatography A* 780: 129–148.
- Kord AS and Khaledi MG (1992) Controlling solvent strength and selectivity in micellar liquid chromatography: Role of organic modifiers and micelles. *Analytical Chemistry* 64: 1894–1900.
- López-Grío S, García-Alvarez-Coque MC, Hinze WL, Quina FH, and Berthod A (2000) Effect of a variety of organic additives on retention and efficiency in micellar liquid chromatography. *Analytical Chemistry* 72: 4826–4835.
- Marina ML and García MA (1997) Evaluation of distribution coefficients in micellar liquid chromatography. *Journal of Chromatography A* 780: 103–116.
- Tomer S, Dorsey JG, and Berthod A (2001) Nonionic micellar liquid chromatography coupled to immobilized enzyme reactors. *Journal of Chromatography A* 923: 7–16.

## Size Exclusion

J Janča, Université de La Rochelle, Cedex, France

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Size-exclusion chromatography (SEC) is the conventional name for a separation method used most frequently for the fractionation and analytical characterization of macromolecules of biological or synthetic origin and less frequently for the separation of colloidal particles. The history of SEC began in the late 1950s with the observation that proteins can be separated on porous cross-linked dextran gels, swollen in aqueous media. The first separation of a synthetic polymer by SEC was described by Vaughan. Moore was the first to separate polymers in a column using cross-linked polystyrene gel. The invention of small porous particles with a typical diameter between 1 and 10  $\mu\text{m}$  brought about an important technological improvement in SEC. The consequent miniaturization of the columns allowed the reduction of the analysis time to minutes or even to tens of seconds.

The development of this separation method is reflected by the numerous names that have been given to this process; for example, gel filtration, gel chromatography, gel filtration chromatography, gel exclusion chromatography, gel permeation chromatography, restricted diffusion chromatography, size separation chromatography, and molecular sieve chromatography. The term SEC, which has been most frequently used in the recent literature, describes very concisely the basic mechanism governing the separation. Nevertheless, this mechanism, separating the species according to their effective dimensions in the solution, is often complicated by other interactions that can, in some instances, make the separation more efficient but can also deteriorate it. This fact must be taken into consideration when selecting the experimental conditions in SEC.

Theoretical bases of the separation and dispersion in SEC are briefly reviewed in order to allow an understanding of the performances of this method when applied to the separation and characterization of macromolecular or particulate species. Among various retention mechanisms theoretically describing the separation, the most general thermodynamic approach represents an advantage due to the fact that it covers the cases where the entropy-driven size exclusion is not dominating the separation of the retained species and the enthalpy-driven interactions

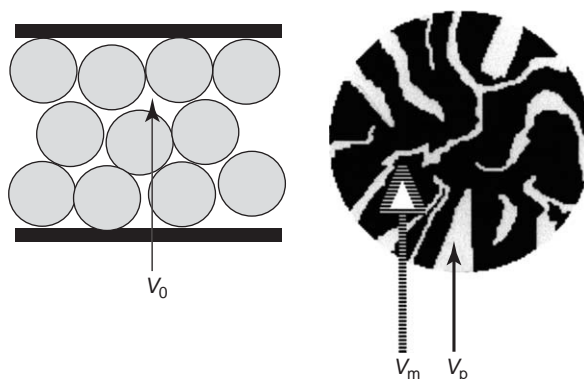
also take place in the chromatographic system. The SEC method can be used not only for purely separation purposes but also for the quantitative characterization of the molar mass or size of the separated species, provided an accurate calibration procedure of the separation system and a correction for the zone broadening are applied.

### Fundamentals

The basic principle of the SEC can be best explained with the use of **Figure 1**. The separation is considered as a specific type of distribution of the separated species (solute) between the solvent filling the pores of the column packing particles (stationary phase) and the solvent outside the particles (mobile phase). The total volume of a packed column,  $V_T$ , is the sum of the total volume of all pores,  $V_p$ , of the volume of the particles matrix,  $V_m$ , and of the interstitial volume,  $V_0$ , outside the particles:

$$V_T = V_m + V_0 + V_p \quad [1]$$

The biggest solute molecules are excluded from the penetration into the pores and move with the mobile phase only in the interstitial volume. Their elution (retention) volume is thus equal to  $V_0$ . The small molecules (including the mobile phase) can penetrate all available pores and their retention volume is the sum ( $V_0 + V_p$ ). The solute molecules of intermediate size can, according to the ratio of their size to the pore size, penetrate a part of the pores while they are excluded from another part. Their retention volume lies within the above limit volumes. Consequently,



**Figure 1** Schematic representation of a chromatographic column for SEC with the void volume between the spherical particles of the column packing and the structure of one porous particle with the pore and matrix volumes.

the retention volume  $V_R$  of a uniform size solute species is

$$V_R = V_0 + K_{SEC} V \quad [2]$$

where  $K_{SEC}$  is a formal distribution coefficient that can also be written in a standard thermodynamic form as

$$K_{SEC} = \frac{c_p}{c_0} \quad [3]$$

where  $c_p$  and  $c_0$  refer to the solute concentrations in the pores and in the interstitial volume, respectively. For the solute molecules that are excluded from the pores,  $K_{SEC} = 0$ . For the molecules that can penetrate the pores and where no other interactions occur in addition to size exclusion,  $0 < K_{SEC} < 1$ . Totally penetrating molecules exhibit  $K_{SEC} = 1$ . Whenever  $K_{SEC} > 1$ , other interactions (e.g., adsorption) participate in the separation mechanism. The thermodynamic interpretation of the distribution coefficient  $K_{SEC}$ , given in the following paragraph, allows one to explain accurately the dependence of  $K_{SEC}$  on the molar mass or size of the solute molecules.

## Theory of Retention

A number of theoretical approaches explained the retention by a simple size-exclusion mechanism presented in the Introduction. Such explanations were associated with more or less complicated pore shapes that were considered as a system of cylindrical, conical, etc., holes that spherical solute molecules can penetrate. The size distribution of the pores was not given much importance. Nevertheless, experiments whose objective was to correlate the molar mass dependence of the retention volume of different solute molecules with the pore size distribution curves did not provide satisfactory results. Although more sophisticated models of this type were more successful when compared with the experiments, they are not considered now as they are not sufficiently general and do not accurately describe the physical reality.

The above-mentioned mechanistic models of separation were based on the assumption that the time necessary to transport the solute molecules between the phases is much shorter than the time spent by them at a given position inside the column. To overcome the difficulty to explain the differences in  $K_{SEC}$  measured under static and dynamic conditions, a mechanism of restricted diffusion has been suggested. Unfortunately, the restricted diffusion model is incoherent with the independence of the  $K_{SEC}$  on the flow rate and temperature observed within a wide range of experimental conditions.

In the so-called hydrodynamic models, the porous structure is approximated by a system of open cylindrical capillaries of different diameters. A flow velocity profile formed inside, combined with the stochastic movement of the solute molecules, results in different average velocities of the molecules transported down the capillary. Larger molecules are eluted at higher average velocities. The model assumes that the percolating solvent can pass through the pores. Such a situation, nevertheless, cannot be true, especially with polymeric gels. Even if a flow through the open pores is possible, the flow rate that is proportional to the fourth power of the radius of the capillary should be too low with respect to the ratio of the diameter of the capillaries in the interstitial volume to that inside the pores, which is  $\sim 10$ – $1000$ .

The stochastic models were developed by calculating the probability with which the solute molecules can penetrate the pores. A comparison of these theoretical models with some experimental data has rendered good agreement.

The above-mentioned mechanisms of SEC separation represent the evolution of theoretical opinion and can still, under specific conditions, provide an explanation of the particular phenomena emerging in experimental SEC. More details on these theories together with the corresponding references can be found in references cited in the Further Reading section. The most general theoretical concept is based on the thermodynamic approach. All transport phenomena in SEC are supposed to take place near the thermodynamic equilibrium. Thus, the equilibrium concentration distribution of the solute represents a good approximation of the actual state of the system. It holds for a change in the Gibbs free energy for a displacement of the mole of a species between two locations at a constant temperature  $T$  and pressure  $P$ :

$$dG = dH - T dS \quad [4]$$

where  $H$  and  $S$  are enthalpy and entropy, respectively. It holds for an exchange of mass due to the separation processes:

$$dG = \left( \frac{\partial G}{\partial n_i} \right)_{T,P,n_{j \neq i}} dn_i \quad [5]$$

where  $dn_i$  is the number of the species  $i$  entering the system. The chemical potential is

$$\mu_i = \mu_i^\circ + RT \ln c_i \quad [6]$$

The use of the concentration  $c_i$  instead of the activity is justified by the very low concentrations concerned. It holds for the equilibrium between the stationary



and mobile phases that  $dG = 0$  and

$$\mu_i^m = \mu_i^s \quad [7]$$

$$\frac{c_{si}}{c_{mi}} = \exp\left(\frac{\Delta\mu_i^\circ}{RT}\right) = \exp\left(\frac{-\Delta H^\circ}{RT}\right) \exp\left(\frac{\Delta S^\circ}{R}\right) \quad [8]$$

The enthalpic term in eqn [8] can be considered as a partition coefficient  $K_p$ , whose value is unity ( $\Delta H^\circ = 0$ ) if the size exclusion is the only interaction of the solute molecules with the column packing. If the adsorption and/or partition of the solute takes place in the chromatographic system,  $K_p > 1$  ( $\Delta H^\circ$  is negative), or if an incompatibility between the solute and the column packing exists,  $K_p$  lies between zero and unity ( $\Delta H^\circ$  is positive). Consequently, the entropic term in eqn [8] represents the distribution coefficient  $K_d$  of pure size exclusion.  $K_d$  acquires values between zero and unity and, in agreement with the experimental findings, is independent of the temperature. Hence, the distribution coefficients can be re-defined as

$$K_p = \exp\left(\frac{-\Delta H^\circ}{RT}\right) \quad [9]$$

$$K_d = \exp\left(\frac{\Delta S^\circ}{R}\right) \quad [10]$$

If  $K_p \neq 1$ , the nonexclusion interactions take place in the separation mechanism. The absence of such interactions requires the SEC to be performed under conditions of only entropic, size exclusion interactions. In such a case,  $K_p = 1$  and  $K_d = K_{SEC}$ . It follows from this thermodynamic approach that the distribution coefficient depends on the chemical character of the solute molecules and that of the solvent, as well as on the matrix constituting the porous particles.

Several other theories were proposed, each one being more or less conforming with a particular case and sometimes with a singular case, such as the flexible, noninteracting macromolecular chains or rigid, freely rotating macromolecules, etc., retained in the pores of a column packing. Each particular approach explains (often in a very sophisticated manner) the retention of the considered species on a well-defined porous structure. The problem is that none of these theories can be applied to resolve the practical analytical question: how to associate the experimental retention data obtained on a given chromatographic column to the molecular characteristics of the separated species.

In practice, the only viable method is the calibration of the separation system with a series of molecules, macromolecules, etc., of the well-known molar masses and chemical nature and the use of the

calibration function to calculate the molar mass of an unknown sample from its chromatogram. From this point of view, the above described general thermodynamic approach is very useful because it provides a real possibility to understand the practical problems of the separations by SEC, the role of the dominating size-exclusion mechanism, and also the importance of the casual enthalpy-driven (non-size exclusion) interactions that are almost always present at different degrees. This approach is justified by the increasing applications of the SEC to the separation and characterization of more and more complex macromolecules of natural biological or synthetic origin, associating systems such as the micelles, polyelectrolytes, etc. However, the references to the existing theoretical models can be found in the literature dealing with SEC and its practical use for analytical purposes.

## Theory of Zone Broadening

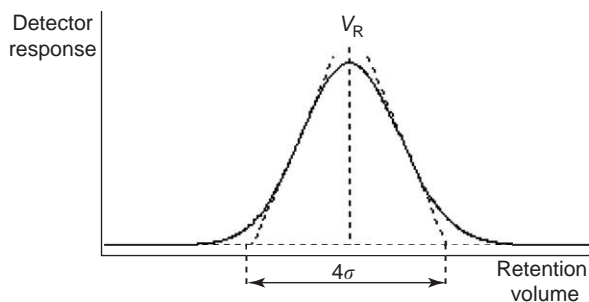
The zones of the separated species migrating down the column undergo dynamic processes and mechanisms that are caused by the natural entropic tendency to disperse any concentration gradient formed. A short pulse of a solution of uniform solute species introduced into the column is subject to a broadening during the elution. Consequently, the concentration distribution of the solute inside the column changes permanently from the initial distribution at the moment of the injection until the final one at the end of the elution. The exhaustive theory of the zone dispersion in chromatography is frequently described in the specialized literature. This section covers the explanation of the basic principles of zone-broadening mechanisms necessary to understand and interpret correctly the experimental results of SEC.

It follows from the general theory of chromatography that the elution curve  $F(V)$  (chromatogram) of a uniform solute can be approximated by a Gaussian distribution function:

$$F(V) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(\frac{(-1/2)(V - V_R)^2}{\sigma^2}\right) \quad [11]$$

where  $\sigma$  is the standard deviation of the Gaussian distribution function. The graphical representation of a model chromatogram is shown in Figure 2. The Gaussian shape of the elution curve is at the origin of the definition of the height equivalent to a theoretical plate,  $H$ :

$$H = L \left( \frac{\sigma}{V_R} \right)^2 \quad [12]$$



**Figure 2** Schematic representation of a model SEC chromatogram characterized by the retention volume of the peak maximum and the standard deviation of the peakwidth.

which is used to describe the efficiency of the chromatographic column. Lower its value, the more efficient is the column, because the broadening of the zone is less important. The physical meaning of  $H$  follows from the random-walk theory of chromatography. In terms of this theory, several processes contribute to the final width of the elution curve. This can quantitatively be described by

$$H = \frac{B}{v} + \frac{C}{v} + \sum \frac{1}{1/A_i + 1/(C_{mi}v)} \quad [13]$$

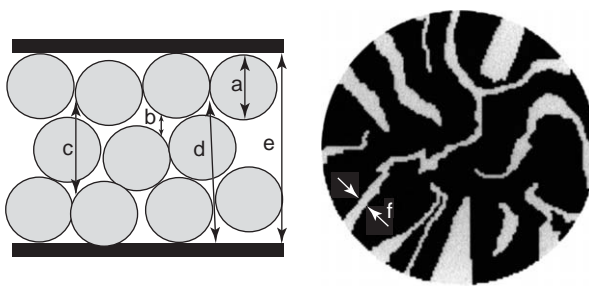
where  $v$  is the velocity of the mobile phase in the column. The first, second, and third terms on the right-hand side of eqn [13] describe, respectively, the effect of the longitudinal diffusion, nonequilibrium effects in the stationary phase, and eddy diffusion and nonequilibrium effects in the mobile phase. The diffusion coefficient  $D$  of the solute molecules in the pores is reduced by an obstructive factor  $\gamma_s$  and by a factor  $\gamma_0$  in the interstitial volume. Hence,  $B$  is given by

$$B = 2\gamma_0 D + \frac{2\gamma_s D(1 - v_R/v)}{v_R/v} \quad [14]$$

where  $v_R$  is the velocity of the zone of the retained solute molecules. It holds for the spherical particles of the diameter  $d_p$  of the column packing:

$$C = \frac{v_R/v}{30} (1 - v_R/v) \frac{d_p^2}{\gamma_s D} \quad [15]$$

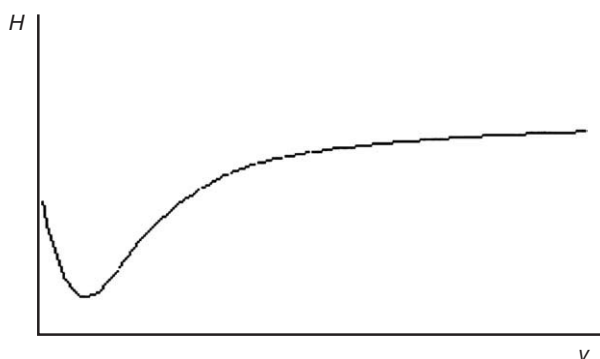
The terms  $1/A_i$  and  $1/C_{mi}v$  describe the effect of the eddy diffusion and nonequilibrium in the mobile phase, respectively. The contribution to zone broadening due to the eddy diffusion is dominating at high flow rates. The importance of both these terms is related with the particle size distribution (PSD) and with the homogeneity of the column packing or, in other words, with the relative differences in the flow



**Figure 3** Schematic representation of a chromatographic column with various scales of the relative flow velocity differences: a – trans-particle scale; b – trans-channel scale; c – short-range interchannel scale; d – long-range interchannel scale; e – trans-column scale; f – trans-pore scale.

velocities caused by the flow velocity profiles and irregularities in the column packing. Five different  $A_i$  and  $C_{mi}$  terms can be distinguished according to the scale, over which the flow velocity differences are considered, namely, trans-channel, trans-particle, short-range interchannel, long-range interchannel, and trans-column. The schematic representation of the domains corresponding to these five different contributions is shown in Figure 3. The trans-particle scale (a) takes into account the average ratio of the maximum to the average flow velocities across the particles under the condition that there is no negligible flow through the open pores (f) in the particles. The trans-channel scale (b) takes into account the average ratio of the maximum to the average flow velocities in the channels forming the interstitial volume and the three later contributions concern the heterogeneity of the average flow velocities on various distances across the column, short-range interchannel (c), long-range interchannel (d), and trans-column (e) scales. All these contributions are additive, as demonstrated by the third term in eqn [13].

In the absence of the specific interactions of the solute with the column packing, the chemical nature of the packing has no effect on the zone broadening and the effect of the pore size distribution is negligible. The effect of the temperature is related exclusively with the changes of the mobile phase viscosity and, consequently, with the changes of the solute diffusion coefficients. The effect of the solute molar mass on the efficiency is related to the mobility of the macromolecules in the solution that decreases with decreasing solute diffusion coefficient. The chromatograms of high molar mass solute exhibit a tendency to an important asymmetry. The presented theoretical model agrees well with the experimental results. A typical shape of the  $H$  versus  $v$  dependence is shown in Figure 4.

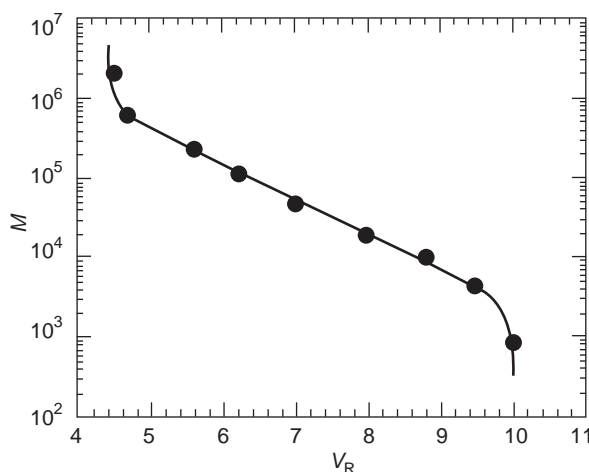


**Figure 4** Schematic representation of the dependence of the height equivalent to a theoretical plate,  $H$ , on the velocity of the mobile phase,  $v$ .

## Information Contents of a Chromatogram

The position of the maximum, the width, and the whole shape of the chromatogram contains information on various processes related with the separation and dispersion, as demonstrated in the above discussion. As long as a retained species is uniform (monomolecular), which means that the sample does not exhibit a molar mass distribution (MMD) or that the MMD is negligibly narrow, the position of the chromatogram maximum can be simply related with the molar mass or size of the sample and the calibration graph, such as that shown in **Figure 5**, can easily be constructed by relating, e.g., the molar mass to the retention volume of a series of standard samples. It has to be stressed, however, that the molar mass is not a universal molecular characteristic on the basis of which the calibration curve obtained with the use of polymer standards of a given chemical composition and structure is applicable to other polymers. It is rather a product of the molar mass and intrinsic viscosity,  $M[\eta]$ , which is proportional to the hydrodynamic volume of a macromolecular chain in solution; that is, the effective molecular size governing the separation under the conditions that the entropy-driven size exclusion is the dominating mechanism.

The width of each particular peak of a chromatogram, in the case that the individual peaks are not overlapping, characterizes only the dispersion phenomena and thus the efficiency of the SEC column. The molar mass or size of the separated species can easily be determined from the position of the maximum of each particular peak. Such a situation is typical for either low molar mass molecules or for some macromolecules of biological origin. Most of the polymers, including the natural ones, are not uniform with respect to their molar mass or size but exhibit the MMD or the PSD. In these cases, the



**Figure 5** Schematic representation of a calibration graph of the column for SEC.

resulting chromatogram represents a superposition of all overlapping peaks, each one corresponding to a particular molar mass, and of the zone broadening. Such a situation is schematically represented in **Figure 6** showing a simplified model chromatogram of a polydisperse macromolecular sample.

As a result, in order to extract accurate information concerning the average molar mass and MMD of a polydisperse sample from the SEC data, the contribution of the dispersive processes should be subtracted from the raw chromatogram. Such a procedure is mathematically rather complex but, fortunately, most of the producers of the SEC chromatographs include the corresponding software into the data treatment computer pack. A simplified explanation of the correction procedure is given here.

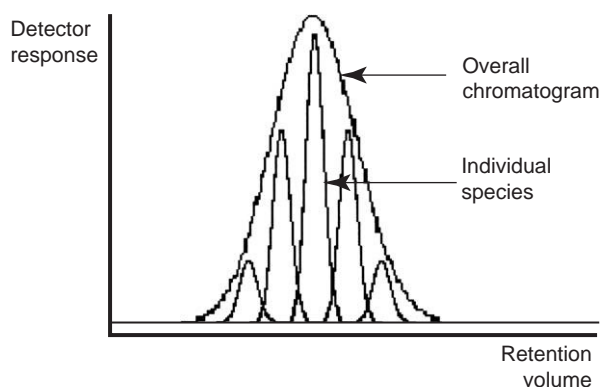
As mentioned in preceding paragraphs, the raw chromatogram of a polydisperse sample represents a convolution of the separation and of the zone broadening. This fact can be described mathematically by

$$h(V) = \int_0^\infty g(Y)G(V - Y) dY \quad [16]$$

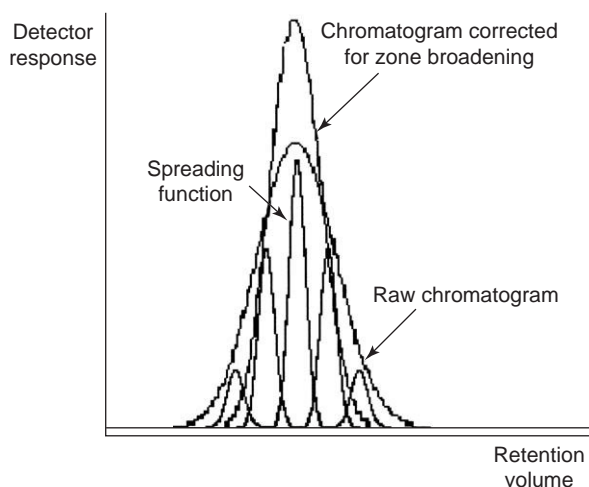
where  $h(V)$  represents the raw chromatogram,  $g(Y)$  is the chromatogram corrected for the zone broadening, and  $G(V, Y)$  is the spreading function or, in other words, the detector response to a uniform species having the elution volume  $Y$ . The spreading function can be, in agreement with the above presented theory of zone broadening, approximated by a Gaussian function:

$$G(V, Y) = (1/(2\pi\sigma_s^2))^{1/2} \exp[-(V - Y)^2/(2\sigma_s^2)] \quad [17]$$

where the standard deviation  $\sigma_s$  should be independent of the retention volume. Whenever  $\sigma_s$  depends on



**Figure 6** Schematic representation of the superposition of five different molar mass (or size) separated species resulting in an overall chromatogram.



**Figure 7** Schematic representation of the difference between a raw chromatogram and the chromatogram corrected for the zone broadening.

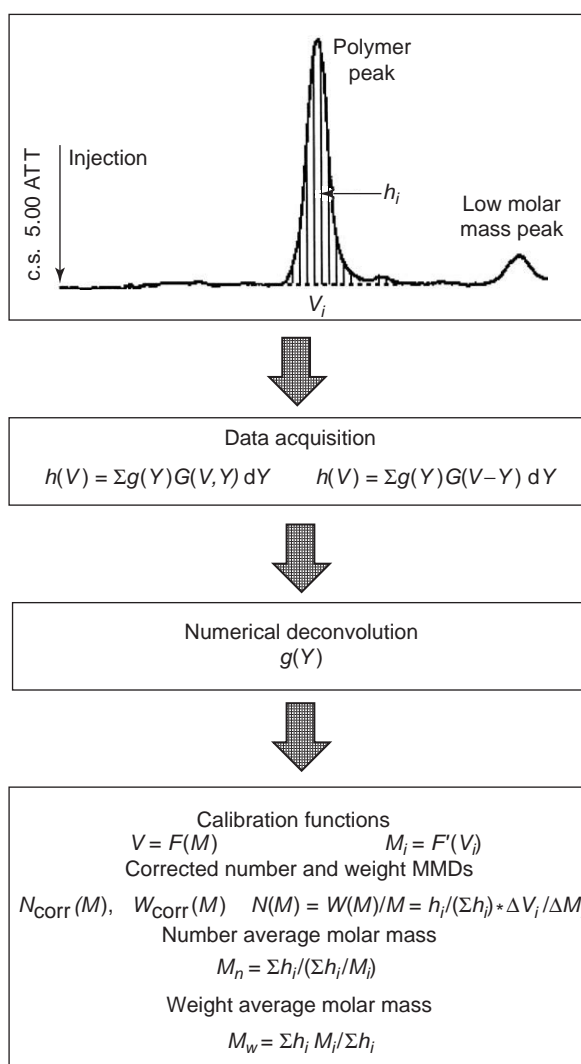
the retention volume, the relationship to be applied is

$$h(V) = \int_0^{\infty} g(Y) [1 / (2\pi(\sigma_s(Y))^2)]^{1/2} \times \exp[-(V - Y)^2 / (2(\sigma_s(Y))^2)] dY \quad [18]$$

Equation [17] or [18] can be solved numerically in order to obtain the chromatogram corrected for zone broadening (see Figure 7), which can be used further to calculate the accurate MMD and average molar mass or the PSD and average particle size of the analyzed sample.

## Applications of SEC

The modern SEC is a powerful separation method that, nevertheless, cannot render absolute information on the molar mass or particle size of the



**Figure 8** Schematic representation of a procedure of SEC data treatment.

separated species. The main chromatographic data, the retention volume of the peak, the width and shape of the chromatogram are, however, very precisely measured quantities that, either by means of a calibration procedure and/or through a combination of the SEC separation system with a suitable absolute detector, provide the required information on the molecular characteristics or size of the samples under investigation.

As mentioned in the previous paragraph, the calibration curve constructed by relating an average molar mass to the retention volume of a series of preferentially narrow MMD polymer samples can be used only for the determination of the MMD and average molar masses of an unknown sample of the same chemical composition and structure. The  $[\eta]$  is the intrinsic viscosity determined experimentally. On the other hand, the use of the product  $M[\eta]$ ,

proportional to the hydrodynamic volume of a macromolecular chain in solution, instead of  $M$ , makes the calibration curve independent of the polymer composition and structure in the majority of practical applications of SEC. Consequently, the universal calibration curve constructed with the use of convenient polymer standards can be used for other polymers under condition that the empirical relationships between  $M$  and  $[\eta]$ , called Mark–Houwink equations, for all polymers in question are known. The application of the principle of universal calibration allows also the estimation of the branching of an analyzed unknown polymer due to the fact that the branched polymer chains in solution exhibit lower intrinsic viscosity in comparison with linear chains of the same molar mass. Thus, the use of only an appropriate Mark–Houwink equation for the interpretation of the experimental SEC data of an unknown branched polymer can result in an agreement of the calculated average molar masses and intrinsic viscosities with those measured by independent methods.

This type of measurements can very elegantly be realized online by coupling several detectors at the end of the SEC column such as a concentration detector (refractive index detector, spectrophotometric detector, etc.) and an absolute detector measuring the molar mass or related property of the separated species such as laser light scattering detector or capillary viscometer detector. These modern sophisticated separation systems allow not only the separation of the analyzed species but also their very detailed analysis and characterization as concerns the MMD or PSD, as well as other structural and compositional characteristics of simple polymers, co-polymers, etc. A schematic representation of a procedure of SEC data treatment from an experimental chromatogram to the final MMD or PSD data is shown in **Figure 8**.

Although other separation techniques have become competitive and preferred for particular separations

and detailed analyses of polymers and colloidal particles, such as field-flow fractionation, hydrodynamic chromatography, and various other liquid chromatographies, SEC is still the most extensively applied technique for the molecular characterization of synthetic and natural macromolecules.

**See also:** **Liquid Chromatography:** Overview; Principles; Column Technology; Mobile Phase Selection; Instrumentation.

## Further Reading

- Altgelt KH and Segal L (1971) *Gel Permeation Chromatography*. New York: Dekker.
- Belenkii BG and Vilenchik LZ (1983) *Modern Liquid Chromatography of Macromolecules*. Amsterdam: Elsevier.
- Cazes J (ed.) (2001) *Encyclopedia of Chromatography*. New York: Dekker.
- Dawkins JV (1978) In: Epton RE (ed.) *Chromatography of Synthetic and Biological Polymers*. Chichester: Ellis Horwood.
- Determann H (1967) *Gel Chromatography*. Berlin: Springer.
- Fisher L (1980) *Gel Filtration Chromatography*. Amsterdam: Elsevier.
- Giddings JC (1965) *Dynamics of Chromatography*. New York: Dekker.
- Heftmann E (ed.) (1992) *Chromatography*. Amsterdam: Elsevier.
- Heftmann E (ed.) (2003) *Journal of Biochemical and Biophysical Methods* 56(1–3).
- Janča J (ed.) (1984) *Steric Exclusion Liquid Chromatography of Polymers*. New York: Dekker.
- Wilson ID, Adlard TR, Poole CF, and Cook M (eds.) (2000) *Encyclopedia of Separation Science*. London: Academic Press.
- Wu CS (ed.) (1995) *Handbook of Size Exclusion Chromatography*. New York: Dekker.
- Yau WW, Kirkland JJ, and Bly DD (1979) *Modern Size-Exclusion Liquid Chromatography*. New York: Wiley.

## Chiral

**V R Meyer**, EMPA St. Gallen, St. Gallen, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

For the separation of enantiomers or optical isomers by chromatography, a chiral setup is necessary. This is obtained by the temporary or permanent

formation of diastereoisomers. For the latter approach the sample is derivatized with a chiral reagent, and the separation can then be achieved with a nonchiral chromatographic system; this is the ‘indirect’ method. For the ‘direct’ separation of enantiomers a number of chiral solid stationary phases have been developed and are commercially available, but it is also possible to use a chiral additive to the mobile phase in combination with a



proportional to the hydrodynamic volume of a macromolecular chain in solution, instead of  $M$ , makes the calibration curve independent of the polymer composition and structure in the majority of practical applications of SEC. Consequently, the universal calibration curve constructed with the use of convenient polymer standards can be used for other polymers under condition that the empirical relationships between  $M$  and  $[\eta]$ , called Mark–Houwink equations, for all polymers in question are known. The application of the principle of universal calibration allows also the estimation of the branching of an analyzed unknown polymer due to the fact that the branched polymer chains in solution exhibit lower intrinsic viscosity in comparison with linear chains of the same molar mass. Thus, the use of only an appropriate Mark–Houwink equation for the interpretation of the experimental SEC data of an unknown branched polymer can result in an agreement of the calculated average molar masses and intrinsic viscosities with those measured by independent methods.

This type of measurements can very elegantly be realized online by coupling several detectors at the end of the SEC column such as a concentration detector (refractive index detector, spectrophotometric detector, etc.) and an absolute detector measuring the molar mass or related property of the separated species such as laser light scattering detector or capillary viscometer detector. These modern sophisticated separation systems allow not only the separation of the analyzed species but also their very detailed analysis and characterization as concerns the MMD or PSD, as well as other structural and compositional characteristics of simple polymers, co-polymers, etc. A schematic representation of a procedure of SEC data treatment from an experimental chromatogram to the final MMD or PSD data is shown in **Figure 8**.

Although other separation techniques have become competitive and preferred for particular separations

and detailed analyses of polymers and colloidal particles, such as field-flow fractionation, hydrodynamic chromatography, and various other liquid chromatographies, SEC is still the most extensively applied technique for the molecular characterization of synthetic and natural macromolecules.

**See also:** **Liquid Chromatography:** Overview; Principles; Column Technology; Mobile Phase Selection; Instrumentation.

## Further Reading

- Altgelt KH and Segal L (1971) *Gel Permeation Chromatography*. New York: Dekker.
- Belenkii BG and Vilenchik LZ (1983) *Modern Liquid Chromatography of Macromolecules*. Amsterdam: Elsevier.
- Cazes J (ed.) (2001) *Encyclopedia of Chromatography*. New York: Dekker.
- Dawkins JV (1978) In: Epton RE (ed.) *Chromatography of Synthetic and Biological Polymers*. Chichester: Ellis Horwood.
- Determann H (1967) *Gel Chromatography*. Berlin: Springer.
- Fisher L (1980) *Gel Filtration Chromatography*. Amsterdam: Elsevier.
- Giddings JC (1965) *Dynamics of Chromatography*. New York: Dekker.
- Heftmann E (ed.) (1992) *Chromatography*. Amsterdam: Elsevier.
- Heftmann E (ed.) (2003) *Journal of Biochemical and Biophysical Methods* 56(1–3).
- Janča J (ed.) (1984) *Steric Exclusion Liquid Chromatography of Polymers*. New York: Dekker.
- Wilson ID, Adlard TR, Poole CF, and Cook M (eds.) (2000) *Encyclopedia of Separation Science*. London: Academic Press.
- Wu CS (ed.) (1995) *Handbook of Size Exclusion Chromatography*. New York: Dekker.
- Yau WW, Kirkland JJ, and Bly DD (1979) *Modern Size-Exclusion Liquid Chromatography*. New York: Wiley.

## Chiral

**V R Meyer**, EMPA St. Gallen, St. Gallen, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

For the separation of enantiomers or optical isomers by chromatography, a chiral setup is necessary. This is obtained by the temporary or permanent

formation of diastereoisomers. For the latter approach the sample is derivatized with a chiral reagent, and the separation can then be achieved with a nonchiral chromatographic system; this is the ‘indirect’ method. For the ‘direct’ separation of enantiomers a number of chiral solid stationary phases have been developed and are commercially available, but it is also possible to use a chiral additive to the mobile phase in combination with a

nonchiral stationary phase; a further technique is the use of a chiral liquid stationary phase that is coated on a solid support. All the common liquid chromatography (LC) detectors can be used, but chiroptical devices such as the polarimeter and the circular dichroism spectrometer can also be especially helpful for qualitative analysis. It is possible to obtain circular dichroism spectra online during the elution of the peaks.

## The Analytical Problem

Enantiomers are asymmetric molecules such that one isomer is the mirror image of the other. As long as their environment is symmetrical they have identical physical and chemical properties and they cannot be distinguished. This changes as soon as they interact with asymmetric objects or fields. Such a chiral (from Greek *cheir* for hand) environment is typically found in biological processes because enzymes, receptors, etc., are three-dimensional asymmetric objects. (An important chiral field is polarized light.) Therefore, it is more a rule than an exception that the two forms of enantiomers that interact with a living organism show different actions: leucine has a sweet or bitter taste, amphetamine is a potent or an inactive stimulant of the central nervous system, and permethrin is a toxic or an almost ineffective insecticide. The demand for efficient analytical and preparative methods for enantiodifferentiation is well established. Chromatographic techniques have brought important improvements in both areas.

For all chromatographic enantioselective techniques it is necessary to create a chiral environment. If the method of separation is LC, especially high-performance liquid chromatography, four different techniques are available; ordered in decreasing importance they are given as follows:

1. The use of a chiral solid stationary phase (usually termed chiral stationary phase, CSP).
2. The formation of diastereoisomers by derivatization with a chiral reagent prior to chromatography on a nonchiral separation system.
3. The use of a chiral mobile phase, which is obtained by the addition of a small amount of chiral reagent to the eluent.
4. The use of a chiral liquid stationary phase. The packing of an LC column, e.g., silica, is coated with a film of a liquid, optically pure compound that must not be soluble in the mobile phase used. A possible phase pair is a tartrate derivative in combination with an aqueous buffer eluent. The handling of such liquid-liquid chromatographic systems is not simple, which is why they are not often used, and they are

not discussed further here. They have been replaced by CSPs.

The fact that the separation of enantiomers is possible with chiral chromatographic systems can be explained by the temporary formation of diastereoisomeric adducts or complexes between the analytes and the CSP or chiral liquid stationary phase. To obtain enantiodifferentiation it is necessary that the two forms of the isomer differ in the number of binding points or in their bonding energy. These non-permanent bondings or interactions can be of the dipole-dipole (induced or permanent), hydrogen bonding,  $\pi$ - $\pi$  interaction, or d-orbital (metal complexation) type. Obviously, a separation is possible if one enantiomer is retained by three binding interactions and the other one by only two of them. This principle is shown in **Figure 1** and is known as three-point interaction rule. Although there is no doubt that it is a valid concept, it needs expansion in various respects: the points of interaction must be directly related to the chiral center but in differing directions; both enantiomers can interact at three points but with differing energy at least at one point; or one of the interactions can be repulsive or even neutral and effective only by guiding the sample molecules sterically.

For the characterization of the separation of a pair of enantiomers the separation factor  $\alpha$  is usually used;  $\alpha$  is the ratio of retention factors  $k$ , which are the measure of retention (eqn [1]):

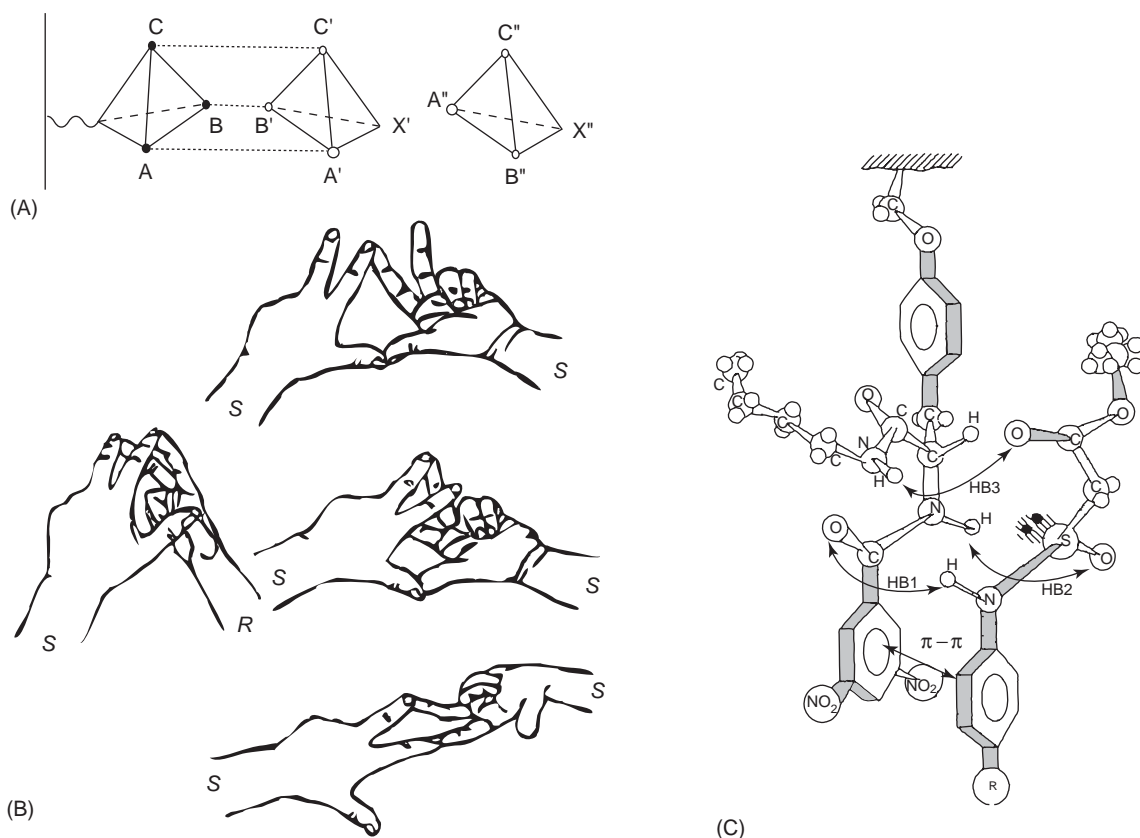
$$\alpha = \frac{k_2}{k_1} = \frac{t_{R2} - t_0}{t_{R1} - t_0} \quad \text{with } k_2 \geq k_1 \quad [1]$$

where  $t_R$  is the retention time of a compound and  $t_0$  the breakthrough time.

The separation factor  $\alpha$  is directly related to thermodynamic parameters, i.e., to the difference in free energy  $\Delta(\Delta G)$ , enthalpy  $\Delta(\Delta H)$ , and entropy  $\Delta(\Delta S)$  of adsorption of the two enantiomers, gas constant  $R$ , and absolute temperature  $T$  (eqn [2]):

$$-RT \ln \alpha = \Delta(\Delta G) = \Delta(\Delta H) - T\Delta(\Delta S) \quad [2]$$

In most cases, the separation of enantiomers on CSPs is controlled by the difference of adsorption enthalpies and  $\alpha$  increases with decreasing temperature. The effects and thus the separation factors are small (which means unfortunately that nonchiral retention is dominating). An  $\alpha$  of 1.05 corresponds to a free energy difference of adsorption of  $120 \text{ J mol}^{-1}$  and a minimum of 8000 theoretical plates is necessary to obtain good resolution (gas constant,  $R = 8.314 \text{ J mol}^{-1} \text{ K}^{-1}$  with  $k_1 = 10$ ; the calculation is performed by means of the chromatographic resolution equation).



**Figure 1** Three representations of the three-point interaction rule for chiral recognition: (A) All molecules involved have tetrahedral structure; one enantiomer can interact with three points because of the fit of AA', BB', CC', whereas with the other no more than a two-point interaction is possible if the molecule is turned suitably; (B) The same principle visualized using hands. Here a three-point interaction is only possible with the *RS* adduct, the *SS* (or *RR*) adduct allows two interactions only; and (C) A model that explains the elution order of *N*-arylsulphinamoyl esters (right) on a dinitrobenzoylamide type CSP (left). It is assumed that four interactions occur (but two of them belong to the same direction, as seen from the chiral center of the sample molecule which is the sulfur atom), three hydrogen bonds and a  $\pi$ - $\pi$  interaction, between the CSP and the more strongly retained enantiomer. In the other enantiomer two groups at the chiral center have changed position and fewer interactions can occur. ((A) and (B) after Meyer VR and Rais M (1989) *Chirality* 1: 167; (C) Reproduced with permission from Caude M, Tambuté A, and Siret L (1991) Chiral stationary phases derived from tyrosine. *Journal of Chromatography* 550: 357.)

## Chiral Stationary Phases

A large number of CSPs have been developed and many of them are commercially available. The most important of them are listed in Table 1. They can be grouped as follows:

1. 'Brush'-type CSPs where small molecules, usually with  $\pi$ -active groups, are bonded to silica.
2. Helical polymers, mainly cellulose and its derivatives.
3. Cavity phases such as cyclodextrins, crown ethers, and macrocyclic compounds.
4. Protein phases.
5. Ligand exchange phases.

A great number of separation problems can be solved by the use of CSPs. They are expensive but can be very effective and simple to use, although the user

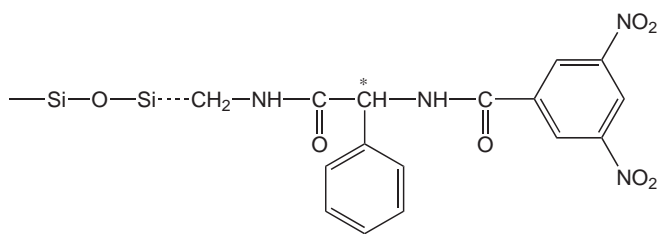
needs to discover which type is best suited for a given problem. They were the key that opened the door to contemporary enantioselective chromatography and only subsequently were chiral additives to the mobile phases and the indirect approach (the formation of diastereoisomers) studied more profoundly.

### 'Brush-Type' CSPs

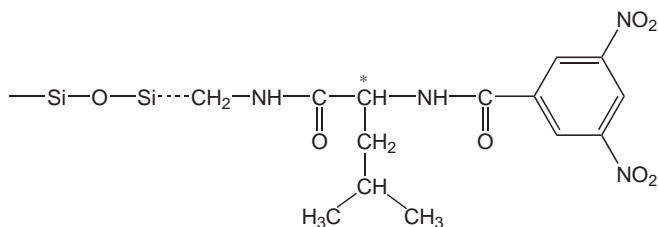
The OH groups of silica can be derivatized by converting them to  $-\text{O}-\text{Si}-\text{O}-\text{R}$  groups where R can be of any type. If R is small and not cross-linked, the new moieties can be considered as bristles fixed to the silica surface. The CSP shown in Figure 1C is of this type; the hatched part at the top of the drawing represents silica. The first broadly used and still very important CSP is the brush-type dinitrobenzoylphenylglycine (DNBPG), shown first in Table 1. After its

**Table 1** Some chiral stationary phases for LC<sup>a</sup>*'Brush-type' phases*

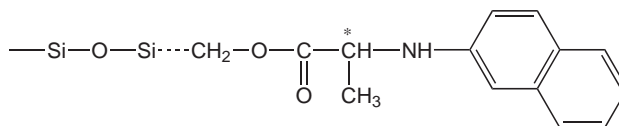
Dinitrobenzoylphenylglycine



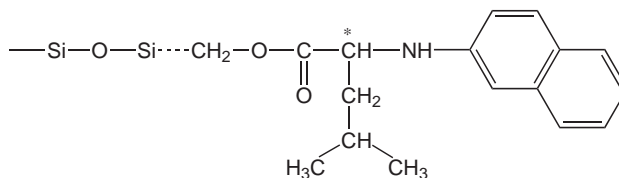
Dinitrobenzoylleucine



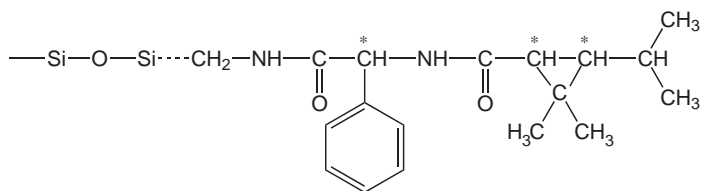
Naphthylalanine



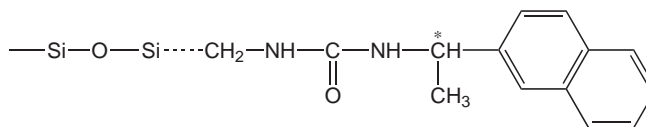
Naphthylleucine



Chrysanthemoylphenylglycine



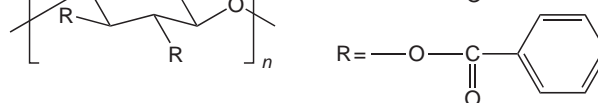
Naphthylethylurea

*Helical polymer phases*

Cellulose triacetate

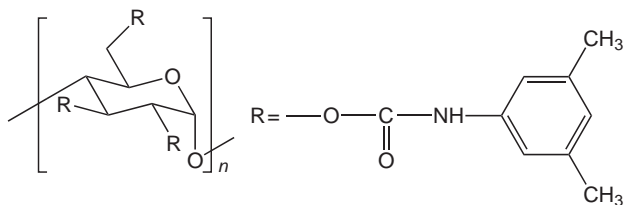


Cellulose tribenzoate

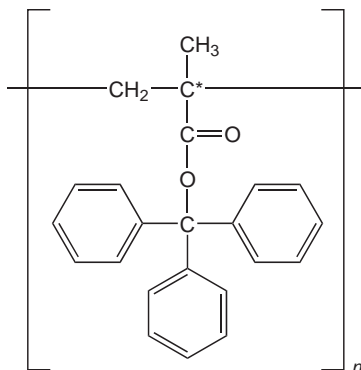
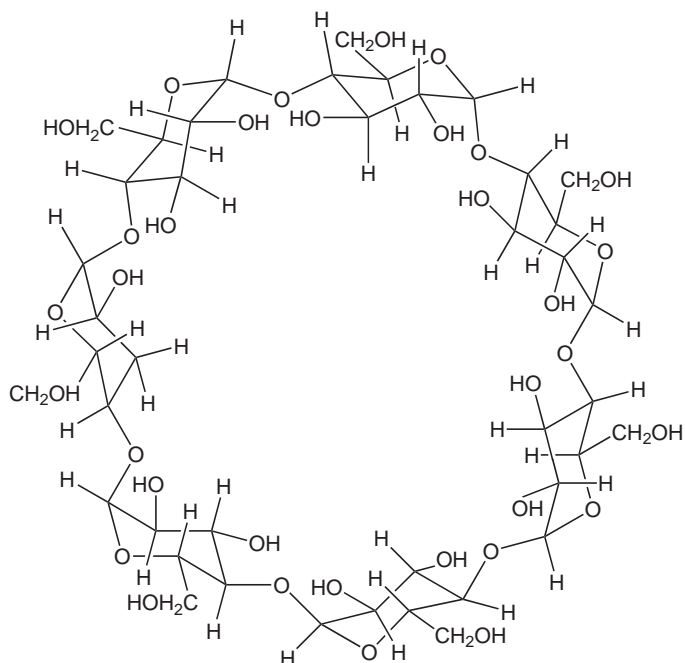


**Table 1** Continued

Amylose tris(dimethylphenylcarbamate)



Poly(triphenylmethylmethacrylate)

*Cavity phases* $\beta$ -Cyclodextrin

Crown ether

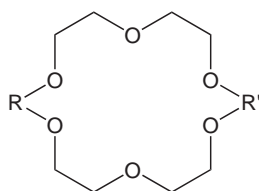
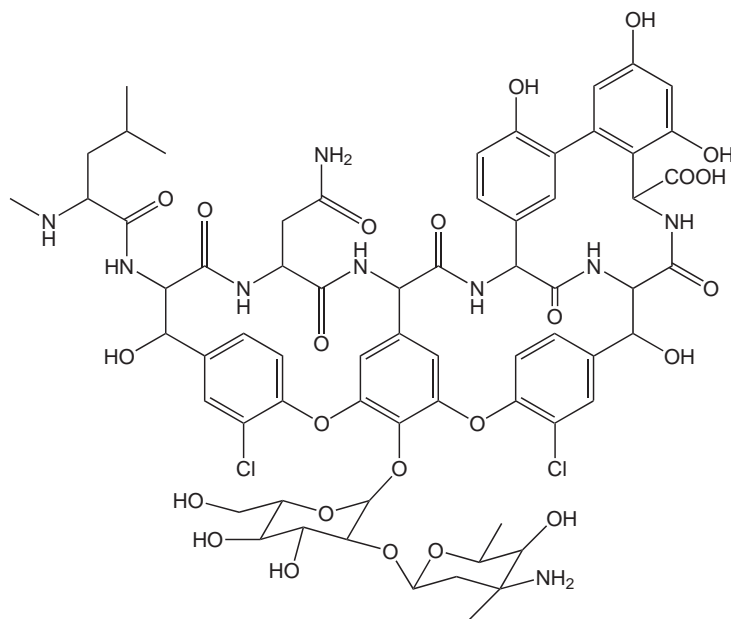
*Continued*



Table 1 Continued

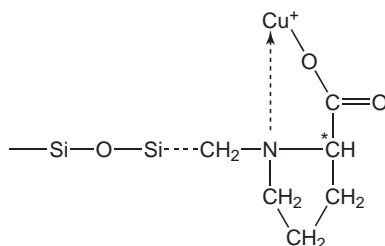
Vancomycin

*Protein phases*

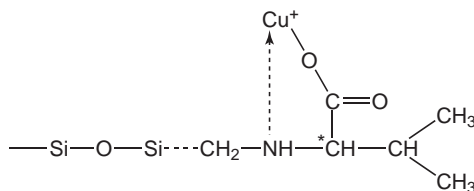
$\alpha_1$ -Acid glycoprotein  
 Bovine serum albumin  
 Human serum albumin  
 Cellulase  
 Ovomuroid

*Ligand exchange phases*

Proline-copper



Valine-copper



<sup>a</sup>Most of these phases are chemically bonded to silica. - - - indicates a spacer group, e.g., a CH<sub>2</sub> chain.

\*Indicates chiral carbons. The chirality of the phases is not specified here; some of the brush-type phases are available in both enantiomeric forms. The chirality is given for cellulose, amylose, cyclodextrin, vancomycin, and the proteins.

inventor, William H. Pirkle, it is often called 'Pirkle-phase', although a more correct name is 'Pirkle I' because it is not the only one of his phases on the market.

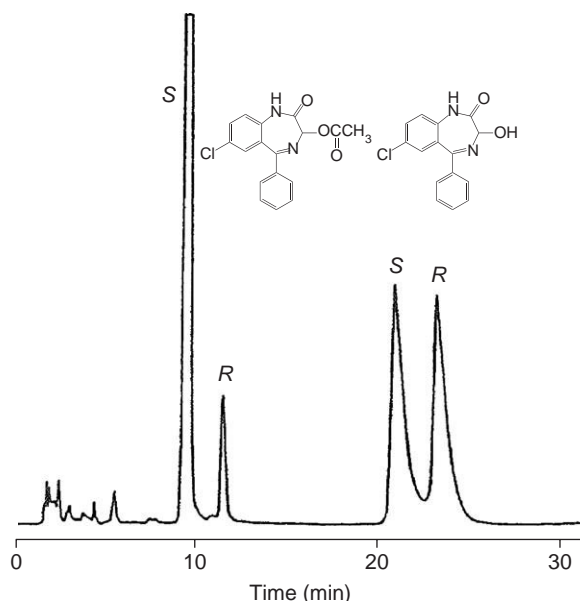
The DNBPG phase has a number of features that are typical of almost all of the brush-type CSPs. It has two amide groups that are rigid (planar), and

therefore the whole chiral moiety will prefer a limited (i.e., not an unlimited) number of conformations, which is important for chiral recognition. The amide groups can undergo dipole-dipole interactions as well as hydrogen bonding with suitable sample molecules. The dinitrobenzoyl group is a  $\pi$ -acceptor (the ring has a slightly positive charge) and will

preferentially interact with  $\pi$ -donors such as anilines, phenols, chlorobenzenes, and naphthalenes. This interaction is assumed to be the most important one and there is no practical chance of obtaining a separation of enantiomers that do not bear this type of group. Therefore, nonchiral derivatization is popular: amines and amino acids are converted to naphthamides, alcohols to naphthyl carbamates, and carboxylic acids to anilides. Usually, the derivatization is simple and improves detectability because the groups brought into the molecule have excellent ultraviolet (UV) absorption. The DNBPG phase is relatively cheap and is available in both enantiomeric forms. This can be important for trace analysis, where the small peak should be eluted ahead of the neighboring large peak, or in preparative applications, where the first eluted peak can usually be obtained in pure form whereas the second peak is contaminated by its forerunner.

Other brush-type CSPs are used in a similar manner: dinitrobenzoylleucine, naphthylalanine, naphthylleucine, chrysanthemoylphenylglycine, naphthylethylurea to name but a few (see Table 1). They all have their specificities for certain classes of compounds, although this is sometimes difficult to predict. For an idea of their individual abilities the literature or the brochures provided by the manufacturers may be consulted.

Brush-type CSPs are robust and allow high sample loads. As an example, Figure 2 shows the separation of oxazepam acetate and its hydrolysis product on dinitrobenzoylleucine.



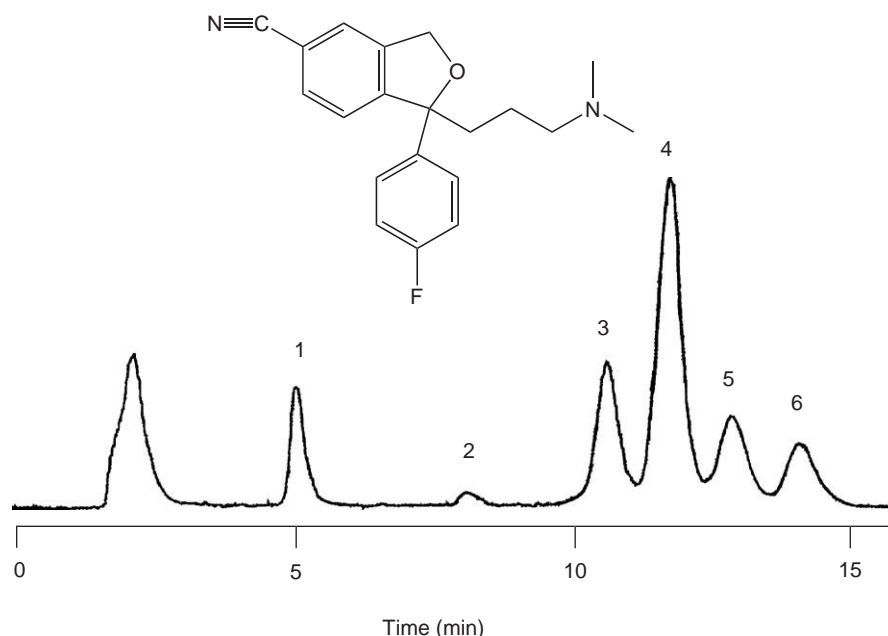
**Figure 2** Separation of the enantiomers of oxazepam acetate and oxazepam on a chiral stationary phase. Column: 4.6 mm  $\times$  25 cm; stationary phase: (*S*)-dinitrobenzoylleucine covalently bonded to 5  $\mu$ m silica; mobile phase, hexane-ethanol-acetonitrile (91.5:5.7:2.8, v/v/v), 2 mL min<sup>-1</sup>; detector, UV 231 nm. Racemic oxazepam 3-acetate was hydrolyzed by human liver microsomes, giving racemic oxazepam. The microsomes strongly preferred the (*R*) enantiomer; after incomplete hydrolysis the remaining oxazepam acetate had a composition of 91.8% (*S*) and 8.2% (*R*) enantiomers. (Reproduced from Yang SK, Liu K, and Güengerich FP (1990) Enantioselective hydrolysis of oxazepam 3-acetate by esterases in human and rat liver microsomes and rat brain S9 fraction. *Chirality* 2: 150.)

of spirobichromene is performed on cellulose tribenzoate; this is not by chance, but helical polymers are suited in an excellent manner for the resolution of enantiomers of 'twisted' molecules (although a vast number of 'flat' compounds can be resolved as well).

### Cavity Phases

Underivatized microcrystalline cellulose is not an effective agent for the LC separation of enantiomers but its derivatives have been shown to be a most interesting and versatile family of CSPs. Cellulose triacetate is available as a bulk material (in this form it is rather cheap) as well as coated on silica. Other derivatives are cellulose tribenzoate (also shown in Table 1), tris(phenylcarbamate), tris(dimethylphenylcarbamate), tris(chlorophenylcarbamate), tris(toluylate), and tricinnamate. They are expensive and perhaps less robust than the brush-type CSPs, but the range of enantiomers that can be separated by them (as a family) is the broadest and most impressive one of all the CSP groups. Their range of application is expanded by poly(triphenylmethacrylate), its derivative where one of the phenyl groups is replaced by a pyridyl moiety, and by amylose derivatives. The retention and separation mechanisms are complex. Figure 5 is presented to demonstrate the possibilities of a circular dichroism detector, but the separation

There are three classes of cyclic chiral selectors in use as CSPs, cyclodextrins, crown ethers, and macrocyclic glycopeptide antibiotics. They can undergo host-guest complexes with small molecules if these can fit into the ring structure; for enantiodifferentiation this fit needs to be stereochemically controlled. Cyclodextrins are oligoglucoses with six, seven, or eight units. Table 1 shows  $\beta$ -cyclodextrin, which is the heptamer and which is used most frequently.  $\alpha$ -Cyclodextrin has six glucose units and  $\gamma$ -cyclodextrin has eight. Glucose itself is chiral and in the cyclodextrin molecule, which in fact is a truncated cone, the superstructure of directed primary (at the smaller rim of the cone) and secondary (at the wider rim) hydroxy groups yields chiral binding points that seem to be essential for enantioselectivity. These



**Figure 3** Separation of citalopram and its demethylated metabolites in a human plasma extract on a vancomycin phase. Column: 4.6 mm  $\times$  15 cm; stationary phase: Chirobiotic V, 5  $\mu$ m; mobile phase: methanol–acetic acid–triethylamine (99.9:0.055:0.06, v/v/v), 1 ml min<sup>-1</sup>; detector: fluorescence 240/296 nm. The structure of the drug is shown; it is used as the racemate. Peaks: 1, alprenolol (internal standard); 2, *R*-(–)-didesmethylcitalopram; 3, *R*-(–)-desmethylcitalopram; 4, *R*-(–)-citalopram; 5, *S*-(+)-citalopram; 6, *S*-(+)-desmethylcitalopram. *S*-(+)-didesmethylcitalopram is not present. (Reproduced with permission from Kosel M, Eap CB, Amey M, and Baumann P (1998) Analysis of the enantiomers of citalopram. *Journal of Chromatography B* 719: 234.)

groups can also be derivatized, e.g., acetylated. Cyclodextrins are used like reversed phases, i.e., with very polar mobile phases. They can separate a wide range of chiral samples, but it is difficult to predict their suitability for a given compound.

Chiral crown ethers of the 18-crown-6 type can resolve amino acids as well as primary amines with close neighborhood of amino function and the center of asymmetry. The interaction occurs between the amino protons and the crown ether oxygens. The groups R and R' of the crown ether need to be large and rigid, e.g., binaphthyls, in order to force the small guest molecules into a well-defined interaction with the host.

The antibiotics vancomycin (shown in Table 1), teicoplanin, and ristocetin A can be bonded to silica, giving a unique class of macrocyclic glycopeptide CSPs. They can be used in the normal-phase mode with a nonpolar eluent as well as in the reversed-phase mode with an aqueous eluent. They show unique selectivity for a large number of analytes. Figure 3 gives an example with the separation of the antidepressant citalopram and its metabolites on a vancomycin CSP.

### Proteins

It is well known in biochemistry that many proteins, especially enzymes and also transport proteins such as albumin, show high enantioselectivities in their

interactions with small chiral molecules. It is possible to bind proteins to silica and to obtain a valuable class of CSPs that are mainly suited for the separation of chiral drugs. Several protein phases are commercially available: albumins,  $\alpha_1$ -acid glycoprotein, ovomucoid, cellobiohydrolase, pepsin, and avidin. They differ in their chromatographic and enantioselective properties, which is to be expected as their biological functions, and their sizes, shapes, or isoelectric points are quite different.

Protein phases are expensive and delicate; their performances (as plate numbers) and capacities are low. For many applications these drawbacks are outweighed by their excellent enantioselectivity.

### Ligand Exchange Phases

Amino acids bonded to silica and loaded with Cu<sup>2+</sup> ions can interact in stereoselective manner with amino acids in aqueous solution. The copper ion forms a complex with both the bound and the sample amino acids. Ligand exchange phases are suited for the separation of amino acids as well as of some  $\beta$ -amino alcohols and similar molecules because these compounds bear two polar functional groups with appropriate spacing. This approach has found limited interest because the column efficiencies are rather low, the detectability of the nonderivatized sample

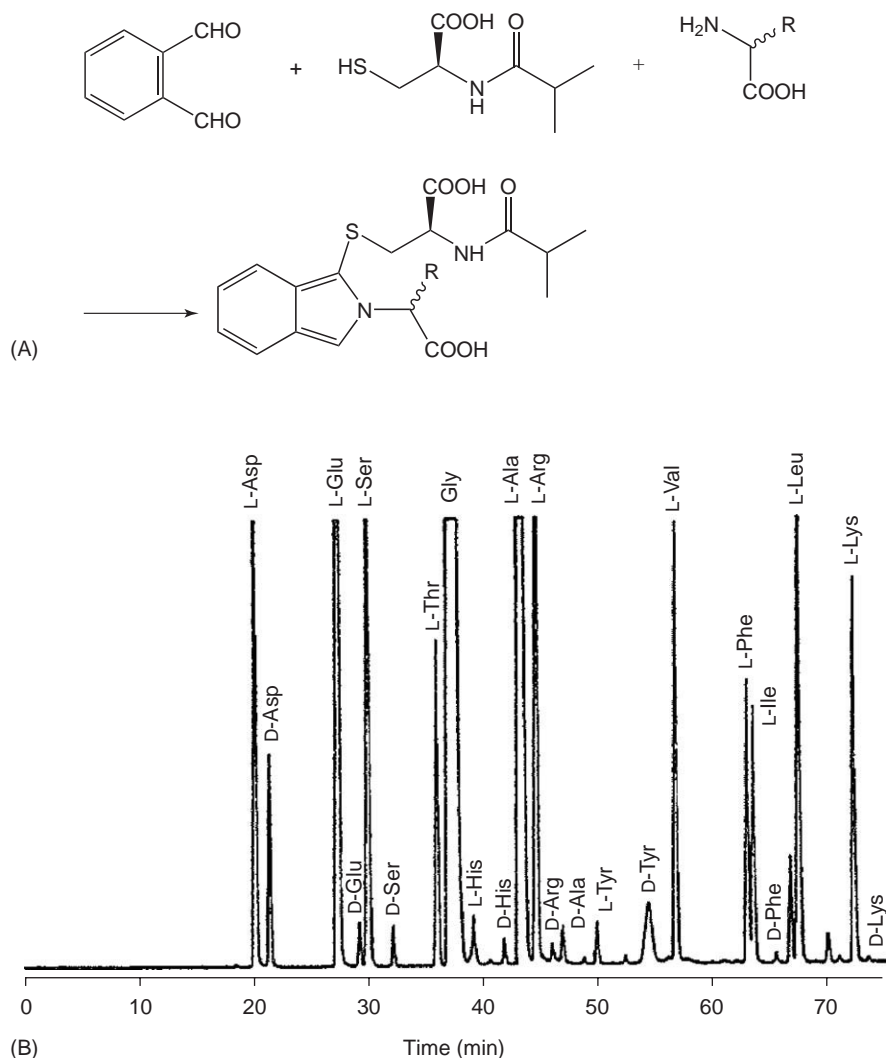
compounds can be a problem, and the mobile phase needs to contain copper.

## Derivatization with Chiral Reagents to Obtain Diastereoisomers

This is the 'indirect' approach for the chromatographic separation of enantiomers. The reaction of the two forms of an enantiomer with an optically pure chiral reagent gives a mixture of diastereoisomers that are not mirror images of each other and therefore can be separated on a nonchiral, common LC phase. This is shown in Figure 4 where amino acids have been derivatized with *o*-phthalaldehyde and *N*-isobutyl-

L-cysteine and can be resolved on an octadecyl-silica phase. The analyst can choose from a wide range of reagents, among which only the price may sometimes be a problem; in contrast, the stationary phase is cheap. Besides the fact that the sample molecule needs to have a derivatizable functional group, there are several other points needing consideration.

The chiral reagent must be optically pure. Otherwise four reaction products are formed, namely two pairs of enantiomers, and the enantiomers will not be resolved but will influence the analytical result. If the reagent has a purity of 99.5%, the maximum optical purity of the sample that can be determined is 99.0%. The reaction should run quantitatively,



**Figure 4** Separation of amino acid enantiomers on a nonchiral stationary phase after derivatization to diastereoisomers with a chiral reagent. (A) Derivatization scheme. The chiral moiety of the reagent is *N*-isobutyl-L-cysteine, which together with *o*-phthalaldehyde and amino acids forms a fluorescing indole. (B) Amino acid enantiomers found in gelatin hydrolysate. Column: 4.0 mm × 25 cm; stationary phase, Hypersil ODS 5 μm; mobile phase, eluent A, 23 mmol l<sup>-1</sup> sodium acetate pH 6.0; eluent B, methanol–acetonitrile (92.3:7.7 v/v), gradient from 0 to 53.5% B in 75 min, 1 ml min<sup>-1</sup>; detector, fluorescence 230/445 nm. (Reproduced with permission from Brückner H, Wittner R, and Godel H (1991) Fully automated HPLC separation of D,L-amino acids derivatized with OPA together with *N*-isobutyl-cysteine. *Chromatographia* 32: 383.)

without racemization and without giving by-products. The necessary excess of reagent usually needs to be removed prior to injection because otherwise a reagent peak will be found in the chromatogram. Diastereoisomers can have different detector properties (e.g., UV absorptivities) and a calibration graph must be obtained for quantitative analysis.

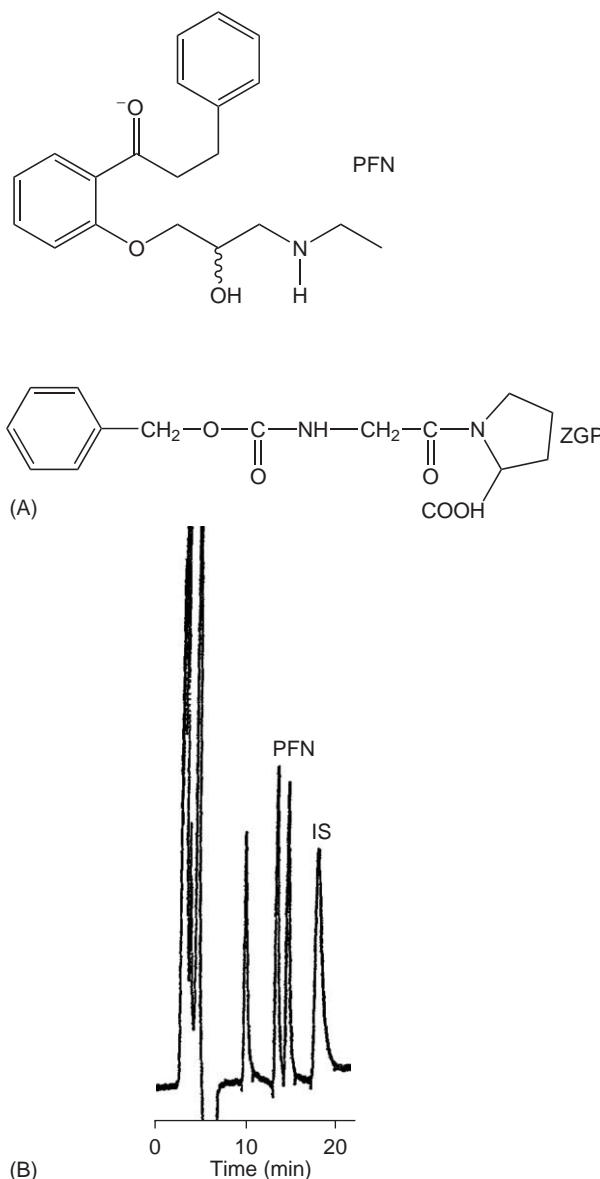
The functional group to be derivatized should be situated close to the chiral center. It is advantageous to obtain diastereoisomers with a certain rigidity such as amides, carbamates, or ureas. If one can choose and needs to determine small amounts of one enantiomer in the presence of an excess of the other, it is often necessary to use the enantiomeric form of the reagent that elutes the small peak first. (This was not done and was not necessary for the analysis presented in Figure 4.)

For biological samples in particular, the derivatization step can often be coupled to the necessary sample cleaning procedures in an elegant and highly effective manner. By the incorporation of fluorescing or otherwise highly detector-active groups, the detectability of the compound of interest can be improved by several orders of magnitude.

## Chiral Additives to the Mobile Phase

Another approach to creating an asymmetrical environment is the use of a chiral mobile phase where the stationary LC phase is of the usual, nonchiral, type. It is not necessary to use a very expensive chiral solvent and it is sufficient to dissolve rather small amounts of a chiral compound in the eluent. In Figure 5, propafenone, a secondary amine and anti-arrhythmic drug, could be separated into the enantiomers by adding  $3 \times 10^{-3} \text{ mol l}^{-1}$  *N*-benzyloxycarbonylglycyl-L-proline to the mobile phase. This reagent allows multipoint interaction because it has a carboxylic group, which forms an ion pair with propafenone, and a carbamate and amide group that are suitable for hydrogen bonding or dipole-dipole interactions.

The advantages are similar to those of the indirect method: the additive can be chosen from a wide range and in some cases its chirality can be adapted to the separation problem; the stationary phase is cheap. In fact, the reagent does not necessarily need to be optically pure (although it should not be a racemate), but decreasing enantiomeric purity reduces the separation factor. The price of the reagent can be high. The interaction between the chiral selector and the sample can be based on inclusion (e.g., with cyclodextrins), on complexation (e.g., with amino acid-copper additives), on ion pair formation (e.g.,

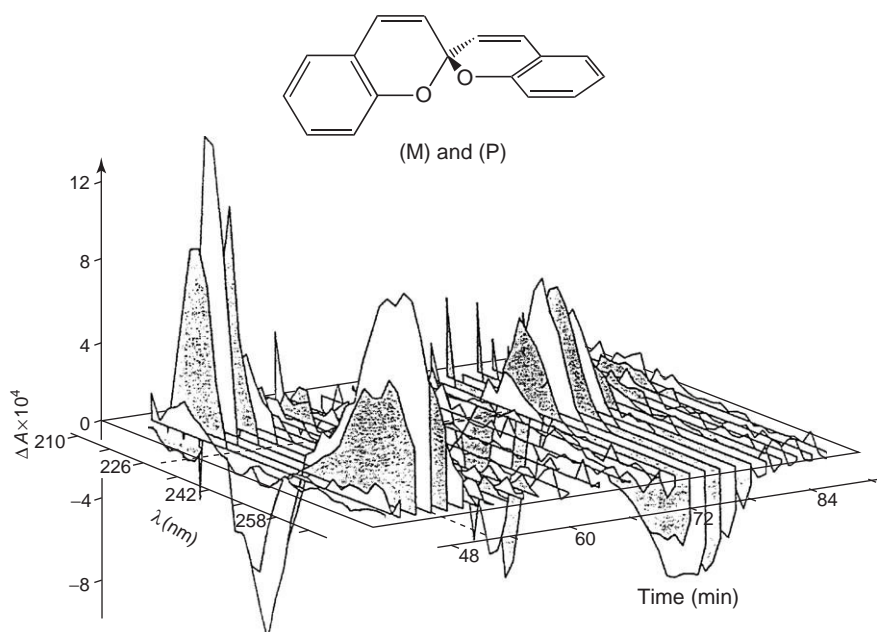


**Figure 5** Separation of propafenone in plasma by means of a chiral additive to the mobile phase. (A) Structures of propafenone (PFN) and the reagent *N*-benzyloxycarbonylglycyl-L-proline (ZGP). (B) Chromatogram of spiked plasma. Stationary phase, Nucleosil 5 cyano; mobile phase, dichloromethane with  $3 \times 10^{-3} \text{ mol l}^{-1}$  ZGP,  $1.5 \times 10^{-3} \text{ mol l}^{-1}$  triethylamine, and  $250 \mu\text{g ml}^{-1}$  water,  $1 \text{ ml min}^{-1}$ ; detector, UV 300 nm; internal standard (IS), (–)-propranolol. (Reproduced with permission from Prevot M, Tod M, Chalom J, Nicolas P, and Petitjean O (1992) Separation of propafenone enantiomers by LC with a chiral counter ion. *Journal of Chromatography* 605: 33.)

with quinine or the reagent used in Figure 5), or on hydrogen bondings and dipoles (e.g., with amino acid derivatives).

It must be considered that this approach is based on chemical equilibria in solution that may be strongly dependent on temperature, pH, or concentration. It is good practice to investigate the influence





**Figure 6** Separation of a racemic spirobichromene and nonstop circular dichroism (CD) spectral acquisition. The two enantiomeric forms are eluted at 52 and 79 min, respectively, and their CD spectra have two bands of same sign and one band of opposite sign each. In two-dimensional representation the dashed lines correspond to a spectrum (line at 52 min) or to a chromatogram (line at 228 nm). Stationary phase, microcrystalline cellulose tribenzoate; mobile phase, methanol,  $0.7 \text{ mL min}^{-1}$ . (Reproduced from Mannschreck A (1992) Chiroptical detection during LC. *Chirality* 4: 163.)

of these parameters on the ruggedness of the method and to control them very strictly if necessary.

## Chiroptical Detection

Enantioselective chromatographic separation can be supported effectively by chiroptical detection. Unfortunately, progress in this field is slower than in separation techniques, although the need is proven.

Optical rotation and circular dichroism can both be utilized for detection. It may be advantageous to couple the chiroptical device to an ordinary LC detector such as a UV or fluorescence detector. A polarimeter measures the rotation, giving positive and negative signals if a resolved pair of enantiomers is passing through its flow cell. One thereby obtains confirmation that a given chromatographic peak is indeed from a chiral compound, and additionally information about the absolute configuration (in  $+/-$  terms) is obtained. It is not even necessary to have complete chromatographic resolution and it is possible to find chiral compounds that are eluted together with a nonchiral solute.

Circular dichroism, i.e., the difference in absorption of right and left circularly polarized light of the same wavelength, can also be used for detection. Again, the two forms of an enantiomer give a positive and a negative signal as in the polarimeter but the sensitivity can be higher and the range of detectability is extended to lower wavelengths. It is also

possible to obtain circular dichroism spectra with stopped-flow or even true online techniques. This allows positive identification of a compound or gives information about the structure of an unknown. **Figure 6** presents an example of this technique.

The ratio of circular dichroism  $\Delta A$  and absorbance  $A$  is termed the  $g$  factor ( $g = \Delta A/A$ ) and is directly related to the enantiomeric excess of a mixture of enantiomers. It is possible to build a  $g$  detector that gives information on the enantiomeric composition of the eluate at any moment during the chromatogram. This is helpful for fraction collection during the preparative chromatographic separation of enantiomers.

**See also:** Chiroptical Analysis. Derivatization of Analytes.

## Further Reading

- Ahuja S (2000) *Chiral Separations by Chromatography*. Washington, DC: American Chemical Society.
- Allenmark S (1991) *Chromatographic Enantioseparation. Methods and Applications*, 2nd edn. Chichester: Ellis Horwood.
- Beesley TE, Lee JT, and Wang AX (2001) Method development and optimization of enantiomeric separations using macrocyclic glycopeptide chiral stationary phases. In: Subramanian G (ed.) *Chiral Separation Techniques: A Practical Approach*, pp. 25–54. Weinheim: Wiley-VCH.

- Bobbitt DR and Linder SW (2001) Recent advances in chiral detection for high performance liquid chromatography. *Trends in Analytical Chemistry* 20: 111–123.
- Gasparini F, Misiti D, and Villani C (2001) High-performance liquid chromatography chiral stationary phases based on low-molecular-mass selectors. *Journal of Chromatography A* 906: 35–50.
- Haginaka J (2001) Protein-based chiral stationary phases for high-performance liquid chromatography. *Journal of Chromatography A* 906: 253–273.
- Kurganov A (2001) Chiral chromatographic separations based on ligand exchange. *Journal of Chromatography A* 906: 51–71.
- Nakano T (2001) Optically active synthetic polymers as chiral stationary phases in HPLC. *Journal of Chromatography A* 906: 205–225.
- Szymura-Oleksiak J, Bojarski J, and Aboul-Enein HY (2002) Recent applications of stereoselective chromatography. *Chirality* 14: 417–435.
- Ward TJ and Farris AB III (2001) Chiral separations using the macrocyclic antibiotics: A review. *Journal of Chromatography A* 906: 73–89.
- Yashima E (2001) Polysaccharide-based chiral stationary phases for high-performance liquid chromatographic enantioseparation. *Journal of Chromatography A* 906: 105–125.

## Affinity Chromatography

**D S Hage**, University of Nebraska, Lincoln, NE, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Affinity chromatography can be defined as a liquid chromatographic method in which a biological agent or biomimetic ligand is used for the selective retention of complementary compounds. This form of liquid chromatography was originally used by Starckenstein in 1910 for the purification of amylase through the use of starch as a solid support. This method continued to slowly develop over the next 50 years. However, it was not until the 1960s that suitable supports like beaded agarose, as developed by Hjerten, became available as well as relatively simple immobilization techniques for these supports. An important advancement in this latter area was a report in 1967 by Axen, Porath, and Ernback in which the cyanogen bromide method for protein and peptide immobilization was first reported. This approach was then used in 1968 by Cuatrecasas, Anfinsen, and Wilchek to purify enzymes through the use of immobilized enzyme inhibitors. It was also at this time that the term ‘affinity chromatography’ was proposed to describe this technique.

Affinity chromatography is relatively simple to perform and is a powerful tool for the separation of biological macromolecules. The high selectivity of this approach often allows single-step purification strategies to be developed, even when working with dilute and highly complex mixtures. This simplicity and the variety of ligands that can be used with this approach have made it an important tool in process-scale separations. However, modern affinity chromatography also plays an important role in the

analysis and study of biological systems. For instance, most forms of chiral liquid chromatography, such as those using immobilized cyclodextrins or serum proteins, can be considered a subcategory of affinity chromatography.

### Basic Principles of Affinity Chromatography

#### The Process

Affinity chromatography relies on the specific recognition that occurs between many biological species, such as the binding of an antibody with an antibody or hormone with its receptor. These interactions are used in affinity chromatography by permanently bonding (or ‘immobilizing’) onto a solid support an appropriate binding agent. This immobilized agent is known as the affinity ligand and represents the stationary phase for this method. Once the affinity ligand and its support have been placed within a column, this column can be used to retain any substance that will form a strong but reversible complex with the ligand. The optimum association constant ( $K_a$ ) for such a system in purification work is generally  $10^4$ – $10^8$  l mol<sup>-1</sup>. However, stronger and weaker interactions can be used in analytical applications of affinity columns.

The most common format employed with affinity columns is given in **Figure 1**. In this scheme, a solution containing the target of interest is passed through a column containing an immobilized ligand capable of binding the target. This is performed in the presence of an application buffer that allows such binding to occur. Since this interaction is usually selective in nature, the ligand will recognize and retain the target while allowing other compounds in

- Bobbitt DR and Linder SW (2001) Recent advances in chiral detection for high performance liquid chromatography. *Trends in Analytical Chemistry* 20: 111–123.
- Gasparini F, Misiti D, and Villani C (2001) High-performance liquid chromatography chiral stationary phases based on low-molecular-mass selectors. *Journal of Chromatography A* 906: 35–50.
- Haginaka J (2001) Protein-based chiral stationary phases for high-performance liquid chromatography. *Journal of Chromatography A* 906: 253–273.
- Kurganov A (2001) Chiral chromatographic separations based on ligand exchange. *Journal of Chromatography A* 906: 51–71.
- Nakano T (2001) Optically active synthetic polymers as chiral stationary phases in HPLC. *Journal of Chromatography A* 906: 205–225.
- Szymura-Oleksiak J, Bojarski J, and Aboul-Enein HY (2002) Recent applications of stereoselective chromatography. *Chirality* 14: 417–435.
- Ward TJ and Farris AB III (2001) Chiral separations using the macrocyclic antibiotics: A review. *Journal of Chromatography A* 906: 73–89.
- Yashima E (2001) Polysaccharide-based chiral stationary phases for high-performance liquid chromatographic enantioseparation. *Journal of Chromatography A* 906: 105–125.

## Affinity Chromatography

D S Hage, University of Nebraska, Lincoln, NE, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Affinity chromatography can be defined as a liquid chromatographic method in which a biological agent or biomimetic ligand is used for the selective retention of complementary compounds. This form of liquid chromatography was originally used by Starckenstein in 1910 for the purification of amylase through the use of starch as a solid support. This method continued to slowly develop over the next 50 years. However, it was not until the 1960s that suitable supports like beaded agarose, as developed by Hjerten, became available as well as relatively simple immobilization techniques for these supports. An important advancement in this latter area was a report in 1967 by Axen, Porath, and Ernback in which the cyanogen bromide method for protein and peptide immobilization was first reported. This approach was then used in 1968 by Cuatrecasas, Anfinsen, and Wilchek to purify enzymes through the use of immobilized enzyme inhibitors. It was also at this time that the term ‘affinity chromatography’ was proposed to describe this technique.

Affinity chromatography is relatively simple to perform and is a powerful tool for the separation of biological macromolecules. The high selectivity of this approach often allows single-step purification strategies to be developed, even when working with dilute and highly complex mixtures. This simplicity and the variety of ligands that can be used with this approach have made it an important tool in process-scale separations. However, modern affinity chromatography also plays an important role in the

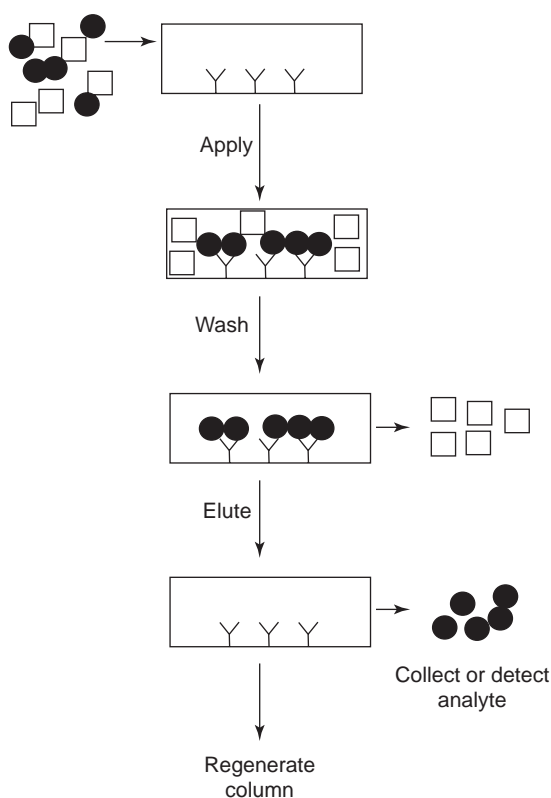
analysis and study of biological systems. For instance, most forms of chiral liquid chromatography, such as those using immobilized cyclodextrins or serum proteins, can be considered a subcategory of affinity chromatography.

### Basic Principles of Affinity Chromatography

#### The Process

Affinity chromatography relies on the specific recognition that occurs between many biological species, such as the binding of an antibody with an antibody or hormone with its receptor. These interactions are used in affinity chromatography by permanently bonding (or ‘immobilizing’) onto a solid support an appropriate binding agent. This immobilized agent is known as the affinity ligand and represents the stationary phase for this method. Once the affinity ligand and its support have been placed within a column, this column can be used to retain any substance that will form a strong but reversible complex with the ligand. The optimum association constant ( $K_a$ ) for such a system in purification work is generally  $10^4$ – $10^8$  l mol<sup>-1</sup>. However, stronger and weaker interactions can be used in analytical applications of affinity columns.

The most common format employed with affinity columns is given in **Figure 1**. In this scheme, a solution containing the target of interest is passed through a column containing an immobilized ligand capable of binding the target. This is performed in the presence of an application buffer that allows such binding to occur. Since this interaction is usually selective in nature, the ligand will recognize and retain the target while allowing other compounds in



**Figure 1** Typical separation scheme for affinity chromatography, showing the steps for sample application, column washing, analyte/target elution, and column regeneration.

the sample to pass through the column in the non-retained peak. However, due to the strong binding between the target and ligand, the target is held within the column until the mobile phase or chromatographic conditions are varied. This is often accomplished by later passing an elution buffer through the column. As the target elutes from the column, it is captured for further use or monitored by an online detector. The column is then cleaned, allowed to regenerate in the application buffer, and used for the next sample. As this scheme suggests, there are several factors to consider in the design of an affinity separation. These include the ligand, support material, immobilization method, and application or elution conditions.

### The Affinity Ligand

Examples of ligands used in affinity chromatography are given in **Table 1**. The ligand can be a naturally occurring biomolecule, an engineered macromolecule, or synthetic substance. Examples of naturally occurring ligands employed in this technique include enzyme inhibitors, hormones, lectins, antibodies, and nucleic acids, which are used to bind enzymes,

**Table 1** Affinity ligands and their target compounds

Type of affinity ligand	Typical targets
<i>Biological agents</i>	
Lectins	Glycoproteins, cells
Carbohydrates	Lectins and other carbohydrate-binding proteins
Nucleic acids	Exonucleases, endonucleases, and polymerases
Cofactors, substrates, and inhibitors	Enzymes
Protein A and protein G	Antibodies
Hormones and drugs	Receptors
Antibodies	Antigens
<i>Nonbiological ligands</i>	
Metal-ion chelates	Proteins and peptides that bind metal ions
Synthetic dyes	Enzymes and various nucleotide-binding proteins

receptors, polysaccharides, antigens, and nucleic acid-binding proteins, respectively. Nonbiological ligands that have been used in affinity columns include immobilized chelates of metal ions (e.g.,  $\text{Ni}^{2+}$ ), reactive dyes, and molecularly imprinted polymers.

As shown in **Table 1**, many ligands are biological macromolecules. These can take part in several types of interactions as they bind to their targets, including electrostatic forces, hydrogen bonding, dipole–dipole interactions, and van der Waals forces. The specific recognition between the ligand and target is due to a combination of these interactions as well as the fit between these species. This is the source of the high affinity seen for many ligands in affinity columns.

All affinity ligands can be grouped into one of two categories: group-specific ligands and high-specificity ligands. High-specificity ligands are generally based on biological agents. A common example is the binding of an antibody with the foreign substance (or antigen) to which the antibody was initially raised. These ligands tend to have relatively large association constants and generally require a step gradient for elution. Group-specific (or general) ligands are agents that bind to a class of related molecules. These recognize a common structural feature on their targets and can be either biological or nonbiological in origin. Examples include protein A, protein G, lectins, boronates, dyes, and immobilized metal-ion chelates. Depending on their affinity, group-specific ligands may require either step elution (e.g., lectins, protein A, and protein G) or may allow the use of isocratic elution (e.g., as occurs in some applications of boronates and immobilized metal-ion chelates).

There are several terms used to classify affinity methods based on the ligands they employ. For instance, bioaffinity chromatography (or biospecific

adsorption) is a term used to describe any affinity method that has a biological molecule as the ligand. This category can be further subdivided into other techniques. As an example, the use of antibodies (or immunoglobulins) as affinity ligands is often referred to as immunoaffinity chromatography, and the use of lectins as ligands is commonly known as lectin affinity chromatography. Other classifications based on the type of ligand include (1) dye-ligand or biomimetic affinity chromatography, which generally uses an immobilized synthetic dye; (2) immobilized metal-ion affinity chromatography, in which the ligand is a metal ion complexed with an immobilized chelating agent; and (3) boronate affinity chromatography, in which boronic acid or one of its derivatives is utilized as the ligand.

### The Support

The support is the material or matrix that holds the ligand within the affinity column. Table 2 shows various materials that have been employed for this purpose. Ideally, this support should have low non-specific binding for sample components but should be easy to modify for ligand attachment. This material should also be stable under the flow rate, pressure, and solvent conditions to be employed in the analysis or purification of samples. In addition, the support should be readily available and simple to use in method development.

Depending on what type of support material is being used, affinity chromatography can be characterized as either a low- or high-performance technique. In low-performance (or column) affinity chromatography, the support is usually a large

diameter, nonrigid gel. Many of the carbohydrate-based supports and synthetic organic materials listed in Table 2 fall within this category. The low back-pressure and reasonable cost of these supports makes them useful for large-scale processing and small-scale preparative work. However, they also tend to have slow mass transfer and limited stability at high flow rates and pressures. These factors limit the usefulness of these supports in analytical applications, where both rapid and efficient separations are often desired. In high-performance affinity chromatography, the supports generally consist of small, rigid particles capable of withstanding the flow rates and/or pressures characteristic of high-performance liquid chromatography (HPLC). Examples of supports suitable for such work include modified silica or glass and hydroxylated polystyrene media. The mechanical stability and efficiency of these supports allows them to be used with standard HPLC equipment. This results in a separation with good speed and precision, making it useful in analytical applications.

One desirable characteristic for the support is that it should allow easy access of the target to the ligand. For a porous material, this requires that the support have pores at least several-fold larger than the target. However, the use of a support with large pores will also have a low surface area per unit volume, which will limit the number of ligands that can be attached to the surface. In process-scale work another requirement for the support is that it should be possible to routinely sanitize this without causing damage. This generally requires that the support be resistant to reagents such as concentrated sodium hydroxide or 8 mol l<sup>-1</sup> urea.

### Immobilization Methods

The third item to consider in the development of an affinity technique is the way in which the ligand is attached to the solid support, or the immobilization method. There are many ways immobilization can be accomplished. These approaches include simple adsorption to a solid support, bioselective adsorption to a secondary ligand (e.g., the noncovalent binding of antibodies to immobilized protein A), entrapment, imprinting, and covalent attachment.

There are several criteria that should ideally be met for the immobilization method. First, this approach must allow the affinity ligand to be coupled to the support without significantly affecting its binding properties. Second, this method must allow the immobilized ligand to still be accessible to its target. In addition, the immobilization method should not introduce any groups to the support that can give rise to nonspecific binding. Fourth, the amount of ligand

**Table 2** Materials used as supports in affinity chromatography

<i>Support material</i>	<i>Approximate usable pH range</i>
<i>Carbohydrate-based supports</i>	
Agarose	2–14
Cellulose	1–14
Dextran	2–14
<i>Synthetic organic polymers</i>	
N-Acryloyl-2-amino-2-hydroxymethyl-1,3-propane diol	1–11
Hydroxyethylmethacrylate polymer	2–12
Oxirane-acrylic polymer	0–12
Polyacrylamide	3–10
Polytetrafluoroethylene	0–14
Poly(vinyl alcohol)	1–14
Styrene-divinylbenzene polymer	1–13
<i>Inorganic supports</i>	
Glass	2–8
Silica	2–8



**Table 3** Examples of immobilization methods based on the covalent coupling of a ligand to the support

---

<i>Techniques for ligands with amine groups</i>
Cyanogen bromide (CNBr) method
<i>N,N'</i> -Carbonyl diimidazole (CDI) method
Divinylsulfone method
Epoxy (bisoxirane) method
Ethyl dimethylaminopropyl carbodiimide (EDC) method
<i>N</i> -Hydroxysuccinimide ester (NHS) method
Schiff base (reductive amination) method
Tresyl chloride/tosyl chloride method
<i>Techniques for ligands with free sulfhydryl groups</i>
Divinylsulfone method
Epoxy (bisoxirane) method
Iodoacetyl/bromoacetyl methods
Maleimide method
Pyridyl disulfide method
TNB-thiol method
Tresyl chloride/tosyl chloride method
<i>Techniques for ligands with carboxylate groups</i>
Ethyl dimethylaminopropyl carbodiimide (EDC) method
<i>Techniques for ligands with hydroxyl groups</i>
Cyanuric chloride method
Divinylsulfone method
Epoxy (bisoxirane) method
<i>Techniques for ligands with aldehyde groups</i>
Hydrazide method
Schiff base (reductive amination) method

---

coupled to the support should be optimal for binding to the target and the ligand should be immobilized in a manner that prevents the loss of ligand over time.

Covalent attachment is the most common approach used for ligand immobilization. Some techniques available for this are given in Table 3. For a protein or peptide this generally involves coupling the ligand through amine, carboxylic acid, or sulfhydryl residues present in its structure. Immobilization of a ligand through other functional sites (e.g., aldehyde groups produced by carbohydrate oxidation) is also possible.

### Application and Elution Conditions

The final set of items to consider in the use of affinity chromatography is the selection of application and elution conditions. Most application buffers in affinity chromatography are solvents that mimic the pH, ionic strength, and polarity experienced by the target and ligand in their natural environment. These conditions generally allow for optimum binding between the target and ligand. Any cofactors or metal ions required for solute–ligand binding should also be present in this solvent. In addition, surfactants and blocking agents may be added to the buffer to prevent nonspecific retention on the support or affinity ligand.

The elution buffer can be either a solvent that produces weak binding or one that decreases the extent of this binding by using a competing agent that displaces the target from the column. These two approaches are known as nonspecific elution and biospecific elution, respectively. Biospecific elution is the gentler of these two methods since it is carried out under essentially the same solvent conditions as used for sample application. This makes this approach attractive for purification work, where a high recovery of active target is desired.

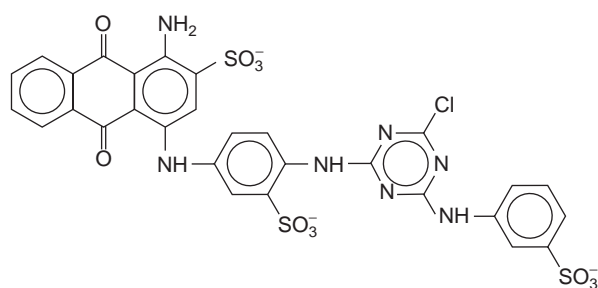
In nonspecific elution, a change in column conditions is used to weaken the interactions between retained compounds and the ligand. This can be accomplished by changing the pH, ionic strength, or polarity of the mobile phase. The addition of denaturing or chaotropic agents to the mobile phase can also be employed. This alters the interactions of the target with the ligand, leading to lower retention. Nonspecific elution tends to be much faster than biospecific elution in removing compounds from affinity columns. This results in sharper peaks and lower limits of detection, making nonspecific elution common in analytical applications of affinity chromatography. This method can also be used in purifying solutes, but there is a greater risk of target denaturation with this approach than there is with biospecific elution. Furthermore, care must be taken to avoid using conditions that may harm the support or result in an irreversible loss of ligand activity.

## Applications of Affinity Chromatography

### Preparative Applications

The earliest use of affinity chromatography and its most popular application is in the purification of proteins and other biological agents. The use of this method in enzyme purification is particularly important, with hundreds to thousands of applications having been reported in this field alone. Ligands used for this purpose include enzyme inhibitors, coenzymes, substrates, and cofactors. For instance, nucleotide mono-, di-, and triphosphates can be used for the purification of various kinases, NAD has been used to isolate dehydrogenases, and RNA or DNA has been used for the preparation of polymerases and nucleases.

Antibodies have also been popular ligands for the purification of biological compounds. There are now thousands of examples of immunoaffinity methods that have been developed for the isolation of hormones, peptides, enzymes, recombinant proteins, receptors, viruses, and subcellular components. In



**Figure 2** Structure of Cibacron Blue 3G-A, a stationary phase often used in dye-ligand affinity chromatography.

addition, immobilized antigens are frequently used to isolate specific types of antibodies. A more general purification scheme for antibodies can be obtained by using antibody-binding proteins like protein A and protein G. These latter ligands have the ability to bind to the constant region of many types of immunoglobulins. Both protein A and protein G have their strongest binding to immunoglobulins at or near neutral pH but readily dissociate from these when placed into a lower pH buffer.

Dye-ligand affinity chromatography is often used in large-scale protein and enzyme purification, with over 500 such compounds having been isolated by this technique. In this method, an immobilized synthetic dye is used that binds to the active site of a target by mimicking the structure of its substrate or co-factor. The most common dye used for this purpose is Cibacron Blue 3G-A (see **Figure 2**). Other dyes used include Procion Blue MX-3G or MX-R, Procion Red HE-3B, Thymol Blue, and Phenol Red. Although these ligands were originally discovered on a trial and error basis, recent work in the area of biomimetic affinity chromatography has used computer modeling and three-dimensional protein structures to develop dyes that compliment the binding pockets of specific target proteins.

### Analytical Applications

Although affinity chromatography was originally created as a preparative method, the past few decades have seen this method also become an important tool in analytical applications. **Table 4** summarizes some strategies that can be employed in these analyses and gives examples of representative applications. The simplest format for using affinity chromatography in analysis involves the traditional step gradient mode, as shown in **Figure 1**. The advantages of using this approach in analytical applications, particularly when performed by HPLC, include its speed, relative simplicity, and good precision.

**Table 4** Analytical applications of affinity chromatography<sup>a</sup>

General application	Examples of analytes
Direct detection	Anti-idiotypic antibodies, antithrombin III, bovine growth hormone, fibrinogen, fungal carbohydrate antigens, glucose tetrasaccharide, glutamine synthetase, granulocyte colony stimulating factor, group A-active oligosaccharides, human serum albumin, immunoglobulin G, immunoglobulin E, interferon, interleukin-2, lymphocyte receptors, $\beta_2$ -microglobulin, tissue-type plasminogen activator, transferrin
Offline affinity extraction	Aflatoxin, albuterol, benzodiazepines, cytokinins, fumonisin, human chorionic gonadotropin, ivermectin and avermectin, nortestosterone, ochratoxin A, oxytocin, phenylurea herbicides, sendai virus protein, trenbolone, triazine herbicides
Online affinity extraction	Aflatoxin M1, $\beta$ -agonists, $\alpha_1$ -antitrypsin, atrazine, atrazine metabolites, benzylpenicilloyl-peptides, bovine serum albumin, carbendazim, carbofuran, chloramphenicol, clenbuterol, cortisol, dexamethasone, diethylstilbestrol, digoxin, estrogens, hemoglobin, human epidermal growth factor, human growth hormone variants interferon $\alpha$ -2, LSD, lysozyme variants
Chromatographic immunoassays	Adrenocorticotrophic hormone, $\alpha$ -amylase, atrazine/triazines, 2,4-dinitrophenyl lysine, human chorionic gonadotropin, human serum albumin, immunoglobulin G, isoproturon, parathyroid hormone, testosterone, theophylline, thyroid stimulating hormone, thyroxine, transferrin, transferrin, trinitrotoluene

<sup>a</sup>The information in this table is based on data provided in Hage (1998).

Affinity extraction is another approach that can be used for solute detection. In this method, an affinity column is used for the removal of a specific solute or group of solutes from a sample prior to their determination by a second method. This employs the same operating scheme as shown in **Figure 1**, but now involves combining the affinity column either offline or online with some other method for the actual quantitation of analytes. This often involves the use of antibodies as ligands, but other binding agents can also be employed.

Offline extraction is the easiest and most common way for combining affinity columns with other analytical techniques. This typically involves the use of antibodies that are immobilized and packed into a disposable syringe or solid-phase extraction cartridge. After conditioning the affinity column with the necessary application buffer or conditioning solvents,

the sample is applied and undesired sample components are washed away. An elution buffer is then applied and the retained target is collected. If desired, the collected fraction can be analyzed directly or first dried down and reconstituted in a solvent that is more compatible with the method to be used for quantitation.

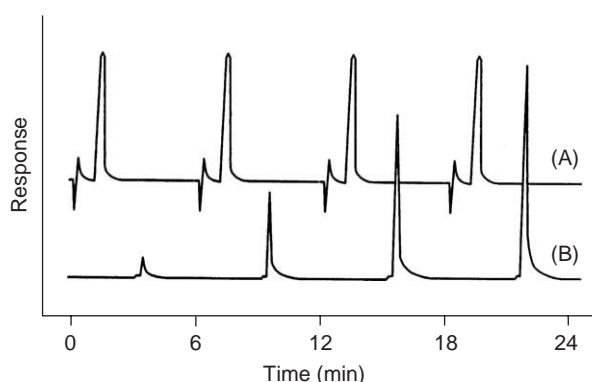
Online affinity extraction can also be used. A typical scheme for performing online immunoextraction with reversed-phase liquid chromatography (RPLC) is shown in **Figure 3**. This involves injecting the sample onto the immunoextraction column, with this column later being switched online with a RPLC column. An elution buffer is then applied to dissociate any retained analyte, which will be captured and reconcentrated at the head of the RPLC column. After all solutes have left the immunoaffinity column, this column is switched back offline and regenerated by passing through the initial application buffer. Meanwhile, the RPLC column is developed with either an isocratic or gradient elution scheme that uses a mobile phase with increased organic modifier content. As the solutes elute through the RPLC column, they are monitored and quantitated through the use of an online detector.

A third way in which affinity chromatography has been used in analytical applications has been in the area of chromatographic, or flow-injection, immunoassays. This method is used in determining trace analytes that do not directly produce a readily detectable signal. The competitive binding assay is the most common format used in performing such an assay. This is generally accomplished by mixing the sample with a fixed amount of a labeled analyte analog (i.e., the 'label') and simultaneously or sequentially injecting these onto an immunoaffinity column that contains a limiting amount of antibody.

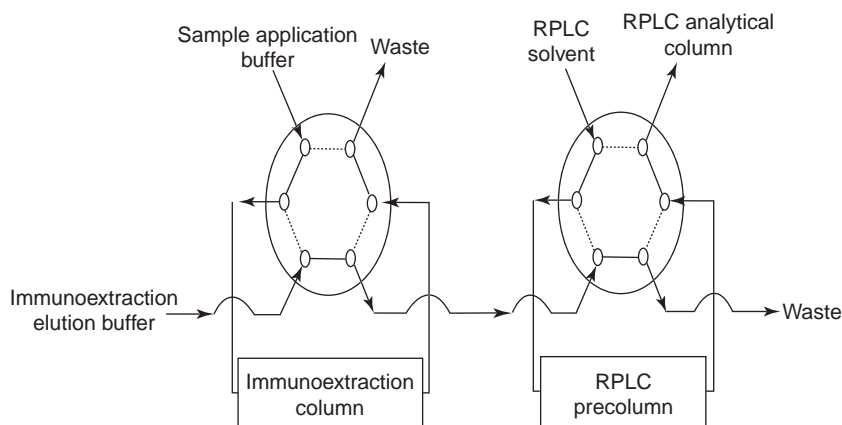
Immunometric assays have also been performed on chromatographic systems. For instance, in a sandwich immunoassay, two different types of antibodies are used (see **Figure 4**). The first of these two antibodies is attached to a solid-phase support and used to extract the analyte from samples. The second antibody contains an easily measured tag and serves to place a label on the analyte, thus allowing it to be quantitated.

### Biophysical Applications

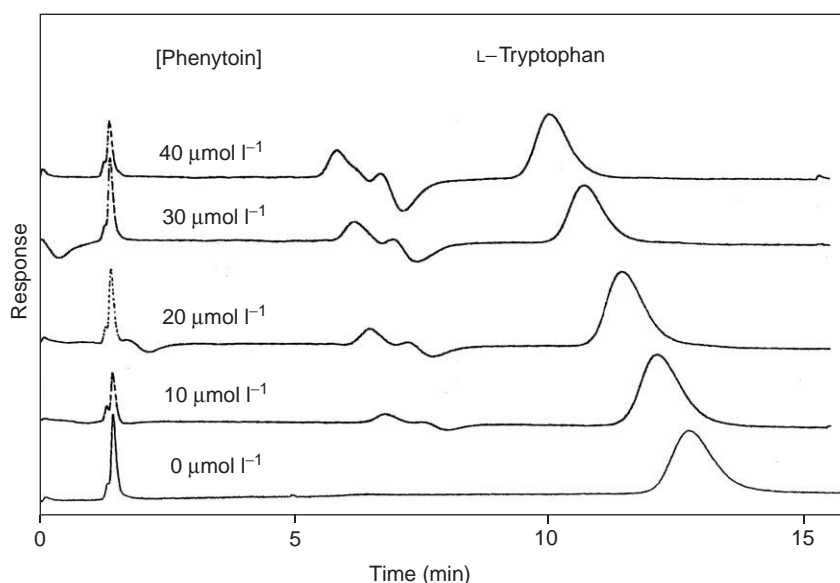
Besides its use in separating and quantitating sample components, affinity chromatography can also be employed as a tool for studying solute–ligand interactions. This approach is called analytical, or quantitative, affinity chromatography. Using this technique, information can be obtained regarding



**Figure 4** Detection of a trace hormone (parathyrin) in human plasma by a chromatographic sandwich immunoassay. The results in (A) show the injection spikes and nonretained fractions for four sequential injections of human plasma samples with increasing parathyrin levels. The results in (B) show the response due to the retained parathyrin. This plot was obtained from Hage and Kao (1991).



**Figure 3** A general scheme for interfacing an immunoextraction column with reversed-phase liquid chromatography (RPLC). The two circles represent six-port switching valves, with the solid and dashed lines showing the flow of sample and solvents in each of the two positions. The operation of this system is described in the text.



**Figure 5** An example of a zonal elution experiment, in which small injections of L-tryptophan are made on to an immobilized human serum albumin column in the presence of increasing amounts of phenytoin in the mobile phase.

the equilibrium and rate constants for biological interactions, as well as the number and types of sites involved in these interactions.

Information on the equilibrium constants for a solute–ligand system can be acquired by using the methods of zonal elution or frontal analysis. Zonal elution involves the injection of a small amount of solute onto an affinity column in the presence of a mobile phase that contains a known concentration of competing agent. The equilibrium constants for the ligand with the solute (and competing agent) can then be obtained by examining how the solute's retention changes as the competing agent's concentration is varied (see **Figure 5**). This method has been used to examine a number of biological systems, such as enzyme–inhibitor binding, protein–protein interactions, and drug–protein binding. Frontal analysis is used in a similar manner but is performed by continuously applying a known concentration of solute to the affinity column. The moles of analyte required to reach the mean point of the resulting breakthrough curve is then measured and used to determine the equilibrium constant for solute–ligand binding.

Information on the kinetics of solute–ligand interactions can also be obtained using affinity chromatography. A number of methods have been developed for this, including techniques based on band-broadening measurements, the split-peak effect, and peak decay analysis. These methods are generally more difficult to perform than equilibrium constant measurements but represent a powerful means for examining the rates of biological interactions. Systems studied by these techniques have

included the binding of lectins with sugars, protein A, or protein G with immunoglobulins, antibodies with antigens, and drugs with serum proteins. The recent creation of commercial sensors for biointeraction studies is one result of such work.

**See also:** **Chromatography:** Overview; Principles. **Electrophoresis:** Affinity Techniques. **Enzymes:** Immobilized Enzymes. **Immunoassays:** Overview. **Liquid Chromatography:** Overview; Principles; Reversed Phase; Chiral; Multidimensional.

## Further Reading

- Axen R, Porath J, and Ernback S (1967) Chemical coupling of peptides and proteins to polysaccharides by means of cyanogen halides. *Nature* 214: 1302–1304.
- Chaga GC (2001) Twenty-five years of immobilized metal ion affinity chromatography: Past, present and future. *Journal of Biochemical and Biophysical Methods* 49: 313–334.
- Chaiken IM (ed.) (1987) *Analytical Affinity Chromatography*. Boca Raton: CRC Press.
- Cuatrecasas P, Wilchek M, and Anfinsen CB (1968) Selective enzyme purification by affinity chromatography. *Proceedings of the National Academy of Sciences of the USA* 61: 636–643.
- Dean PDG, Johnson WS, and Middle FA (1985) *Affinity Chromatography: A Practical Approach*. Oxford: IRL Press.
- De Frutos M and Regnier FE (1993) Tandem chromatographic immunological analyses. *Analytical Chemistry* 65: 17A–25A.
- Hage DS and Kao PC (1991) High-performance immunoaffinity chromatography and chemiluminescent detection

- in the automation of a parathyroid hormone sandwich immunoassay. *Analytical Chemistry* 63: 586–595.
- Hage DS (1998) Recent advances in analytical applications of immunoaffinity chromatography. *Journal of Chromatography B* 715: 3–28.
- Hage DS (ed.) (2004) *Handbook of Affinity Chromatography*. New York: Dekker.
- Hage DS and Nelson MA (2001) Chromatographic immunoassays. *Analytical Chemistry* 73: 198A–205A.
- Hermanson GT, Mallia AK, and Smith PK (1992) *Immobilized Affinity Ligand Techniques*. San Diego: Academic Press.
- Hjerten S (1964) The preparation of agarose spheres for chromatography of molecules and particles. *Biochimica Biophysica Acta* 79: 393–398.
- Porath J, Carlsson J, Olsson I, and Belfrage G (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* 258: 598–599.
- Scouten WH (1981) *Affinity Chromatography: Bioselective Adsorption on Inert Matrices*. New York: Wiley.
- Starkenstein E (1910) Ferment action and the influence upon it of neutral salts. *Biochemische Zeitschrift* 24: 210–218.

## Multidimensional

**F Hernández and J V Sancho**, University Jaume I, Castellon, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Multidimensional chromatography is a very attractive technique for the analysis of complex mixtures where a mono-dimensional separation cannot be sufficient to resolve all the components of interest. Obvious advantages are the higher peak capacity and resolution offered by these systems. Typically, one part of the chromatogram from the first column is transferred to another column via a suitable interface.

Multidimensional liquid chromatography (MDLC), also known as coupled-column liquid chromatography (LC–LC) or column-switching chromatography, is a particular case of multidimensional chromatography. It can be performed either in offline or online mode. In offline operation, the fractions eluted from the primary column are collected manually or by a fraction collector and then reinjected (either with or without preconcentration) into a second column. This approach has the advantage of being simple, does not need any switching valve, and the mobile phases used in each column need not to be mutually compatible. However, offline procedures usually are laborious and time consuming and the recovery is often low.

Online techniques have the advantage of automation by using pneumatic or electronically controlled valves, which switch the column effluent directly from the primary column into the secondary column. Automation improves reliability and sample throughput and shortens analysis time, as well as minimizes sample loss. Online techniques are commonly preferred although they are not always

feasible from an operational point of view, mainly by limitations in the compatibility of mobile phases.

The term MDLC usually refers to the systems involving different separation modes (i.e., varying the nature of the stationary phase), and because of this reason some authors use the concept of heteromodal LC–LC. Obviously, when orthogonal separation mechanisms are used, the resolving power is significantly increased in relation to mono-dimensional LC. However, if the mechanisms involved lead to an inversion of elution order from the first to the second column, the separation achieved with the whole system can be substantially decreased. Thus, although column selectivity is critical in MDLC, coupling dissimilar stationary phases do not always ensure an orthogonal separation. Tuning operation parameters, such as mobile phase strength, temperature, or buffer strength, is also critical in MDLC. Among the experimental parameters, mobile-phase strength is the most significant one influencing the separation and is the easiest to tune.

In homomodal LC–LC, the chromatographic improvement occurs by switching columns of similar selectivity. This approach is not a genuine multidimensional chromatography but some authors consider it as MDLC. The objective is to improve the analytical methodology pursuing, generally, the concentration of the sample components (sample enrichment) or to shorten the analysis time.

Sample enrichment is one of the most frequent applications of MDLC. These methods are based on the possibility that analytes are retained as a narrow band on the top of the first column when a large volume of sample is pumped. Therefore, the separation power of the first column is not critical, and typically short precolumns or small cartridges are employed. This allows the use of relatively high flow-rates during the sample enrichment step keeping



- in the automation of a parathyroid hormone sandwich immunoassay. *Analytical Chemistry* 63: 586–595.
- Hage DS (1998) Recent advances in analytical applications of immunoaffinity chromatography. *Journal of Chromatography B* 715: 3–28.
- Hage DS (ed.) (2004) *Handbook of Affinity Chromatography*. New York: Dekker.
- Hage DS and Nelson MA (2001) Chromatographic immunoassays. *Analytical Chemistry* 73: 198A–205A.
- Hermanson GT, Mallia AK, and Smith PK (1992) *Immobilized Affinity Ligand Techniques*. San Diego: Academic Press.
- Hjerten S (1964) The preparation of agarose spheres for chromatography of molecules and particles. *Biochimica Biophysica Acta* 79: 393–398.
- Porath J, Carlsson J, Olsson I, and Belfrage G (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* 258: 598–599.
- Scouten WH (1981) *Affinity Chromatography: Bioselective Adsorption on Inert Matrices*. New York: Wiley.
- Starkenstein E (1910) Ferment action and the influence upon it of neutral salts. *Biochemische Zeitschrift* 24: 210–218.

## Multidimensional

**F Hernández and J V Sancho**, University Jaume I, Castellon, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Multidimensional chromatography is a very attractive technique for the analysis of complex mixtures where a mono-dimensional separation cannot be sufficient to resolve all the components of interest. Obvious advantages are the higher peak capacity and resolution offered by these systems. Typically, one part of the chromatogram from the first column is transferred to another column via a suitable interface.

Multidimensional liquid chromatography (MDLC), also known as coupled-column liquid chromatography (LC–LC) or column-switching chromatography, is a particular case of multidimensional chromatography. It can be performed either in offline or online mode. In offline operation, the fractions eluted from the primary column are collected manually or by a fraction collector and then reinjected (either with or without preconcentration) into a second column. This approach has the advantage of being simple, does not need any switching valve, and the mobile phases used in each column need not to be mutually compatible. However, offline procedures usually are laborious and time consuming and the recovery is often low.

Online techniques have the advantage of automation by using pneumatic or electronically controlled valves, which switch the column effluent directly from the primary column into the secondary column. Automation improves reliability and sample throughput and shortens analysis time, as well as minimizes sample loss. Online techniques are commonly preferred although they are not always

feasible from an operational point of view, mainly by limitations in the compatibility of mobile phases.

The term MDLC usually refers to the systems involving different separation modes (i.e., varying the nature of the stationary phase), and because of this reason some authors use the concept of heteromodal LC–LC. Obviously, when orthogonal separation mechanisms are used, the resolving power is significantly increased in relation to mono-dimensional LC. However, if the mechanisms involved lead to an inversion of elution order from the first to the second column, the separation achieved with the whole system can be substantially decreased. Thus, although column selectivity is critical in MDLC, coupling dissimilar stationary phases do not always ensure an orthogonal separation. Tuning operation parameters, such as mobile phase strength, temperature, or buffer strength, is also critical in MDLC. Among the experimental parameters, mobile-phase strength is the most significant one influencing the separation and is the easiest to tune.

In homomodal LC–LC, the chromatographic improvement occurs by switching columns of similar selectivity. This approach is not a genuine multidimensional chromatography but some authors consider it as MDLC. The objective is to improve the analytical methodology pursuing, generally, the concentration of the sample components (sample enrichment) or to shorten the analysis time.

Sample enrichment is one of the most frequent applications of MDLC. These methods are based on the possibility that analytes are retained as a narrow band on the top of the first column when a large volume of sample is pumped. Therefore, the separation power of the first column is not critical, and typically short precolumns or small cartridges are employed. This allows the use of relatively high flow-rates during the sample enrichment step keeping

reasonable retention capacity. As the main objective is the extraction/preconcentration of analytes, the term online solid-phase extraction (SPE) is often used leading to a SPE–LC coupling. Obviously, some cleanup can also be obtained during this step depending on the type of stationary phase used in the preconcentration cartridge.

LC–LC coupling using two full-size separation columns provides a greatly enhanced peak capacity in relation to one-dimensional techniques. The LC–LC separation can be used in either profiling or target mode. In the first case, the aim is to separate all single components from a complex mixture, in such a way that every fraction from the first column is transferred to the second column. However, the purpose of LC–LC in the target mode is to isolate a specific narrow fraction containing the analyte(s) and to transfer it to the second column after performing an efficient cleanup. The fraction of interest may be formed by early-eluting analytes (front-cut), analytes eluting in the middle of the chromatographic effluent (heart-cut) or late-eluting analytes (end-cut).

This article focuses on these two coupled systems, SPE–LC and LC–LC, considering either homomodal or heteromodal approaches. The main characteristics of these techniques are summarized in **Table 1** and will be discussed in more detail in the subsequent sections.

## SPE–LC Coupling

Online coupling of SPE with LC followed by various modes of detection represents a fast, modern, and

suitable approach for trace analysis. Important advantages are the increase in sensitivity and automation, as well as the lower risks of sample contamination or analyte losses by evaporation. In contrast to LC–LC, no continuous elution occurs in SPE–LC during the first dimension. Ideally, the analytes are fully retained on the SPE phase during the sorption and washing steps, and subsequently, quickly and completely desorbed in a small volume during the elution step. The main difference in relation to the offline approach is the desorption step, since the trapped compounds are eluted directly from the first to the second column by a suitable mobile phase that also acts on the chromatographic separation. The analysis of the complete sample allows the dramatic reduction of the sample volume, in contrast to offline SPE where only an aliquot of the extract is injected into the chromatographic system. Online SPE–LC is a robust technique to be applied in different fields, due to the good compatibility of the LC aqueous mobile phases with the SPE of biological or environmental samples, for example, which are mainly aqueous in nature.

## SPE Sorbents

Typically, trace enrichment has been applied to relatively nonpolar analytes present in aqueous solutions by using small reversed-phase precolumns as C-1 (normally, C<sub>18</sub> silicas). However, other approaches are nowadays applied, mainly due to the development of new sorbents to be used as stationary phases in C-1.

**Table 1** Main characteristics of MDLC techniques: SPE–LC versus LC–LC

<i>Item/issue</i>	<i>SPE–LC</i>	<i>LC–LC</i>
General approach	The whole sample is preconcentrated in a short column/cartridge (C-1) and backflush/forwardflush transferred to the analytical column (C-2)	Target mode where only the fraction containing the analytes is forwardflush transferred from the first column (C-1) to the second analytical column (C-2)
Main objectives	Preconcentration	Selectivity improvement
Advantages	Automation Sensitivity improvement Sample treatment reduction	Automated cleanup Sensitivity improvement by large volume injection (LVI) Sample treatment simplification
Critical/weak points	Retention/preconcentration of interferents Breakthrough volume	Stability/robustness of C-1 Periodical checking of coupling conditions (cleanup and transfer times)
Main application fields	Environment (organic pollutants in water)  Biomedical (drugs and related compounds in biological fluids)	Environment (target analysis of single or a few compounds in complex analyte/matrix combinations) Biomedical (drugs and related compounds, and endogenous compounds in biological fluids)
Trends and developments	Use of more selective retention mechanisms in the preconcentration column (immunosorbents, restricted access media, molecularly imprinted polymers, etc.) Hyphenation to mass spectrometry (MS and MS <sup>n</sup> )	Comprehensive LC–LC applied to the entire profiling of proteins and peptide mixtures (proteomics)  Hyphenation with mass spectrometry (MS and MS <sup>n</sup> )

The extraction of polar analytes in environmental water has been partly solved by the introduction of carbon-based sorbents (e.g., graphitized carbon blacks) and highly cross-linked styrene-divinyl benzene copolymers, and a new generation of polymers (Oasis, Absolut Nexus, etc.) has been designed to extract a wide spectrum of analytes (including hydrophilic, lipophilic, acidic, basic, and neutral). Ion-pair and ion-exchange sorbents allow the extraction of ionic or ionizable analytes. Cation exchangers include weak carboxylic acid and strong aromatic (or nonaromatic) sulfonic acid groups, and anion exchangers are normally made of primary and secondary amino groups (weak) or quaternary amine forms (strong).

Today, there is strong interest in the development of online sample treatment techniques that allow the handling of untreated biological samples. Thus, in online SPE-LC, deproteinization of plasma and serum is required before extraction, especially if the same cartridge is used for repeated analysis. For this purpose, restricted-access materials (RAMs) have been developed, which combine size-exclusion and reversed phase mechanisms, allowing extraction and cleanup of samples in the same step. RAMs have become quite popular for the direct injection of biological fluids, since they prevent the access of matrix components (e.g., proteins) while retaining the analytes in the interior of the sorbent.

There is a considerable interest in using highly selective SPE sorbents that allow extraction, cleanup, and preconcentration in a single step. As a consequence, new selective sorbents have emerged in recent years that are based upon molecular recognition. Immunoaffinity sorbents use immobilized antibodies as packing material and allow a high degree of molecular selectivity. Immunoaffinity SPE-LC in combination with mass spectrometry (MS) or tandem MS is a powerful approach for the determination of organic analytes in biological fluids and offers the very high selectivity required when analytes have to be determined/confirmed at the trace levels in complex matrices.

Recently, molecular imprinted polymers (MIPs) have gained attention as new selective sorbents for chromatography. The cavities in the polymer selectively recognize and bind a specific compound in a similar way as immunoaffinity sorbents, but can be tailor-made. They have the advantages of being prepared more rapidly and easily, and to be stable at high temperatures in large pH ranges. However, most current MIPs require the analyte to be dissolved in nonprotic organic solvents, limiting their use in online SPE of aqueous samples.

## Experimental Setup and Method Development

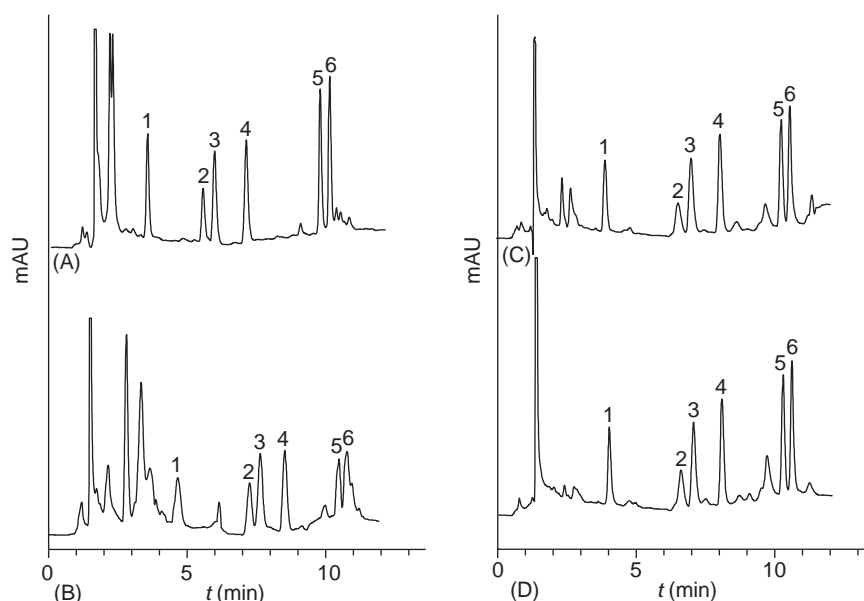
Online SPE-LC can be easily automated by using commercially available devices as Prospekt (from Spark Holland) or OSP-2 (Merck), which have the capability to switch to a fresh disposable precolumn for every sample. These systems improve the productivity since the next sample is automatically prepared while the previous sample is being analyzed.

The size of the precolumn is of great importance since the profile of the extracted species transferred from the precolumn to the analytical column should be as narrow as possible at the beginning of the separation in order to avoid band broadening. Dimensions of the precolumn should be adapted to those of the analytical column, and are typically 2–15 mm long  $\times$  1–4.6 mm ID for 50–250 mm long analytical columns. Dimensions can be slightly increased without band broadening if an appropriate mobile phase gradient is used.

Desorption of analytes from the precolumn (C-1) and subsequent loading on the analytical column (C-2) can be carried out either in the back-flush or forward-flush modes. The back-flush mode provides better peak shape, but it may lead to precolumn packing disturbances, as well as clogging of the analytical column when real samples are analyzed. Forward desorption allows the precolumns to also act as guard columns; however, depending on the dimension and particle size, this mode can provide additional band broadening of analyte. Both desorption modes are used in practice.

In SPE-LC method development, several parameters have to be optimized. Most effort is directed toward the optimization of the SPE step. Obviously, coupling with LC and experimental conditions for the correct chromatographic separation in the second (analytical) column have to be also optimized. The main parameters involved in the SPE process are: type and amount of sorbent, the sample volume that can be applied without losses of analytes, the composition and volume of the washing solution, and the composition and volume of the elution (desorption) solution. A helpful parameter to characterize the usefulness of a given precolumn is the breakthrough volume, which represents the maximum sample volume that can be loaded into the precolumn with a theoretical 100% recovery (the ratio between the amount of analyte extracted and the amount applied).

As an example, **Figure 1** shows SPE-LC chromatograms for a water sample containing several herbicides, obtained under different experimental conditions (volume of sample passed throughout the



**Figure 1** SPE–LC chromatograms for HPLC water spiked with different herbicides: (1) bromacil, (2) metabenzthiazuron, (3) chlortoluron, (4) diuron, (5) linuron, and (6) chloroxuron. (A) 50 ml sample spiked at  $0.2 \mu\text{g l}^{-1}$  passed throughout a  $\text{C}_{18}$  precolumn ( $30 \times 4 \text{ mm}$ ); (B) 100 ml sample spiked at  $0.1 \mu\text{g l}^{-1}$  passed throughout a  $\text{C}_{18}$  precolumn ( $30 \times 4 \text{ mm}$ ); (C) 50 ml sample spiked at  $0.2 \mu\text{g l}^{-1}$  passed throughout a PRP-1 precolumn ( $10 \times 2 \text{ mm}$ ); and (D) 100 ml sample spiked at  $0.1 \mu\text{g l}^{-1}$  passed throughout a PRP-1 precolumn ( $10 \times 2 \text{ mm}$ ). In all cases, C-2 is a  $\text{C}_{18}$  column ( $125 \times 4 \text{ mm}$ ). Mobile phase: acetonitrile–water gradient.

cartridge and type of sorbent). Increasing the sample volume when using  $\text{C}_{18}$  cartridges led to band broadening, while PRP-1 polymeric sorbent provided satisfactory peak shape making feasible the determination of analytes below  $0.1 \mu\text{g l}^{-1}$ .

In comparison with offline methods there are two main limitations when developing online SPE–LC. First, the small columns used obviously contain small amount of sorbent (typically between 20 and 100 mg). Under these conditions, when analytes are poorly retained in the precolumn, the amount of sorbent cannot be sufficiently increased, as in offline SPE procedures, as it would lead to an increase in the precolumn size and unacceptable band broadening. Normally, the best solution is to select a more retentive sorbent. Second, the mobile phase selected should allow the elution of analytes from the precolumn being compatible with the analytical column, and adequate to obtain the chromatographic separation required in each particular case. In practice, the selection of sorbents in the analytical columns has more restrictions than in precolumns. As an example, if many compounds have to be separated over a wide range of polarity in a SPE–LC multiresidue method, highly efficient analytical columns with both water-rich and organic-rich mobile phases are required when conventional detection techniques (fluorescence, ultraviolet–visible) are employed. At present, only  $\text{C}_{18}$  silica columns meet this requirement. In fact, this packing material is the

most widely used as stationary phase in the analytical columns employed in SPE–LC methods.

One advantage of automated online SPE–LC methods is the good reproducibility in comparison with offline methodologies, even when breakthrough volumes are exceeded. Quantification in online SPE–LC should be performed using standards (in solvent or, preferably, in matrix) after application of the whole procedure, under the same experimental conditions than in samples. Therefore, it is not necessary to calculate the recovery of the extraction process for every analyte, as quantitative analysis is not carried out by comparison with direct injections. Of course, it is more convenient to handle a sample volume lower than the breakthrough volume for more reproducible results, but this is not always feasible. For example, in a multiresidue analysis where analytes have to be determined at very low concentration levels, relatively high sample volumes have to be extracted to get satisfactory limits of quantification. Under this situation, polar analytes might not be completely recovered after extraction in the precolumn, but they could be satisfactorily determined, with acceptable reproducible results, in automated SPE–LC procedures.

### Applications and Trends

Nowadays online SPE–LC is a powerful approach for sample preparation, as it satisfies most of analytical demands usually required in modern

laboratories: selectivity, sensitivity, reliability, and rapidity. Simplification of sample preparation is one of the trends in analytical chemistry that is clearly fulfilled by automated SPE-LC. Miniaturization of the overall SPE-LC system reduces analysis time and decreases solvent consumption, two important requirements in the biological field. The development of more selective sorbents is also one of the main research areas, and clear examples are the MIPs or immunoextraction sorbents.

Apart from interesting applications to the biological field (biomedical and pharmaceutical applications), SPE-LC is a very attractive approach in the environmental field, where it has been widely applied mainly because of the improvement in sensitivity. Additionally, it is suited to multiresidue analysis, one of the main trends in the environmental field, as there is an increasing interest in monitoring the environment and preserving the water quality, and increasing the number of analytes to be controlled. At present, the sensitive and accurate determination of a wide range of neutral, acidic, and basic pollutants, and even ionic substances, is a demand in environmental field. The extraction of polar analytes in water is a research area that is being investigated by introduction of new polymers and carbons with high specific surface areas. The identification of transformation products and metabolites of different pollutants in the environment will be a key issue in the next years, and SPE sorbents able to extract these polar analytes will be required. Obviously, all analytical demands in biological, environmental, and any other application field will be better satisfied by using hyphenated techniques, i.e., by coupling MDLC with MS (single or tandem). In fact, SPE-LC (and LC-LC) in combination with MS/MS is one of the most powerful tools for the majority of problems encountered in real-world samples.

## LC-LC Coupling

LC-LC refers to the conventional two-dimensional mode of chromatography using two full-size separation columns where the fractions from one column are selectively and precisely transferred (online or offline) to one secondary column for a further separation. An LC-LC separation system may be used in either the profiling or targeted mode. The purpose of the profiling mode, also known as comprehensive LC-LC, is to fractionate all components of the analyzed mixture, while the purpose of the targeted LC-LC mode is to isolate either a single or a few specific components (that present similar retention on the first dimension) from a complex mixture. Thus, the transferred volume of the first

mobile phase can contain all the components of the sample, a group of peaks, a single peak, or even a fraction of a peak.

### Comprehensive LC-LC

In this approach, the first column is generally operated at either a low-mobile-phase velocity or the flow is interrupted intermittently, while the second column is operated at high velocity to analyze column fractions as they elute from the first column. The rate-limiting step in profiling is analysis time in the second dimension, as it should be able to generate at least one chromatogram during the time required for a peak to elute from the primary column. Thus, apart from other obvious advantages, the use of MS as a detection system is very appropriate in this technique.

Nowadays, the main application area of profiling LC-LC is the proteomic field, where the intention is to characterize the entire range of proteins expressed in a cell or tissue under defined metabolic state, ~10 000–20 000 proteins. Additionally, the proteins are enzymatically cleaved yielding ~20 peptides per protein; so a high-peak-capacity system is required to cope with these extremely complex samples. Various configurations have been proposed for the comprehensive multidimensional chromatography separation of proteins and peptides, mainly size-exclusion chromatography (SEC)/reversed-phase liquid chromatography (RPLC), affinity chromatography (AC)/RPLC, and ion-exchange chromatography (IEC)/RPLC. Other configurations include capillary electrophoresis, used as the second dimension, in combination with SEC or RPLC as first dimension, or isoelectric focusing and chromatofocusing, used as first dimension, coupled with RPLC as second dimension. Coupling IEC to RPLC as first and second dimensions, respectively, is possibly the most frequent approach as it shows several advantages as complementary selectivities, solvent compatibility, and easy coupling of the second column to electrospray tandem mass spectrometry systems, which relaxes the constraints on LC resolution because additional separation also occurs in the mass domain. The majority of chromatographic systems developed in this field employ capillary LC because of the improvement in sensitivity and the easy hyphenation with MS detection systems.

Several strategies have been employed in MDLC separations for proteomic studies. These strategies include discontinuous offline or online MDLC using fraction collection, direct online coupling of two dimensions, and MDLC using column switching. A major advantage of MDLC for proteomic studies



is its ability to fully automate multidimensional LC-MS<sup>n</sup> systems. However, the complete resolution of all components still remains a challenge.

### Targeted LC-LC

In contrast to comprehensive MDLC, the goal of the targeted LC-LC mode is to isolate a single, or a few specific components, from a complex mixture. Target analysis is normally performed in the 'heart-cutting' mode, where an eluting zone from the first column (C-1) is switched online to a second column (C-2) for a subsequent separation of the transferred compounds (Figure 2). This combination leads to an enhanced selectivity and sensitivity, although the use of compatible mobile phases restricts LC-LC to certain chromatographic combinations.

In comparison to automated online trace enrichment on precolumns (SPE-LC), one of the most favorable aspects of automated sample processing using LC-LC is the utilization of the separation power of the first column (C-1). In fact, the application range of the LC-LC technique will be determined by the separation power of the first column: low resolution favors multiresidue methods, while high resolution leads to methods for single analytes or a group of compounds with fairly similar properties.

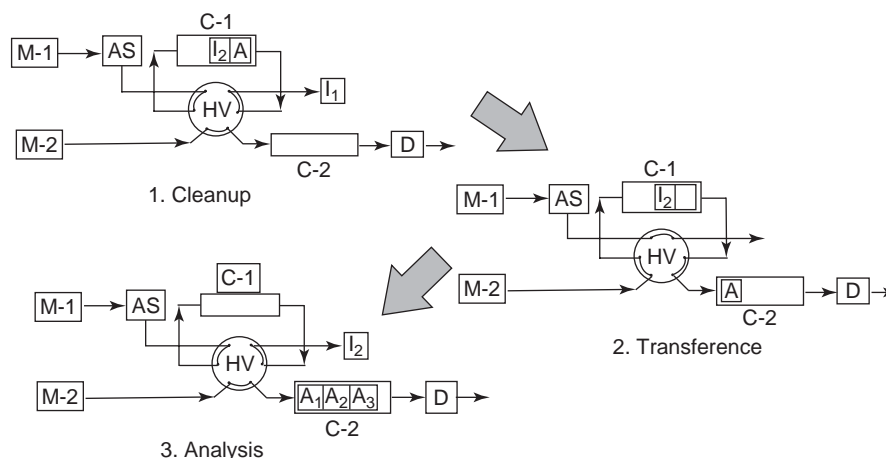
LC-LC offers the opportunity to enlarge the sample injection volume up to several milliliters for improved sensitivity. Additionally, it gives the possibility to remove a large excess of interferences by appropriate selection of the LC-LC conditions (for example, early-eluting polar interferences encountered in environmental samples). This pre-separation on C-1 contributes largely to the enhancement of the selectivity.

**Method development** A critical operation in target component analysis by LC-LC is the selection and transfer of the eluent fraction from C-1 to C-2. Connecting the first column directly to the detector, after injecting a standard solution, the volume of the first mobile phase (M-1) required for the elution of the first analyte gives the cleanup volume/time. On the other hand, the volume of M-2 required to complete the elution of the last eluting analyte is used to estimate the transfer volume.

Usually, the forward-flush mode is preferred for the transfer from C-1 to C-2, although the backflush mode has also been used. The latter has two main drawbacks: (1) it will lead to an additional band broadening when analyzing polar to moderately polar compounds, and (2) for more retained analytes reversing the flow will diminish the separation obtained earlier in the process.

Assuming that analytes retention times and peak widths are constant, the selection of coupling conditions is usually achieved by time-based valve switching. Dirty samples (or extracts) may affect the retention times of analytes on C-1 during the processing of large series of samples. In these cases, due to C-1 column aging and possible increasing peak width, it is recommended to periodically check the heart-cut timing parameters (cleanup and transfer times).

The attainable selectivity will be determined by the effectiveness of cleanup on C-1 for the removal of early-eluting interferences (cleanup time) and by the volume of the analytes fraction (transfer time), which limits the transfer of late-eluting interferences. In other words, optimal selectivity will be obtained at a minimal transfer volume, and part of the required separation should have been taken place in C-1



**Figure 2** Different stages of the LC-LC coupling in the targeted mode using the 'heart-cutting' approach. M-1 and M-2: mobile phases of the first and the second columns; AS: autosampler; C-1 and C-2: first and second columns; HV: six-port high-pressure valve; D: detector; I<sub>1</sub>: early-eluting interferences; A: analytes; I<sub>2</sub>: late-eluting interferences.

before transferring the analyte(s) fraction to C-2. However, in complex samples it is unavoidable that part of the interferences is transferred together with the analytes. Thus, it is important to combine complementary separation mechanisms in the first and second columns to achieve satisfactory separation between analytes and cotransferred interferences.

**Applications** Major applications of targeted LC–LC deal with the biomedical, pharmaceutical, and environmental fields. In biomedical analysis, LC–LC has been used extensively and successfully in the heart-cut mode for the determination of drugs and related compounds in matrices as plasma, serum, or urine. LC–LC of endogenous compounds in biological fluids has also been reported. Chromatographic combinations applied include RPLC, ion exchange, polar bonded-phase, chiral, or affinity columns.

Although the term ‘multidimensional’ refers to LC systems with different retention mechanisms in the columns used, column switching without change in retention mode can be a useful approach, easy to implement, for separating complex mixtures. For example, coupling two RPLC columns is a practical substitute for linear gradient elution and is particularly useful for analyte enrichment and sample cleanup, as has been demonstrated for environmental analysis or bioanalysis.

Most of applications of targeted LC–LC in the environmental field deal with analysis of a single analyte. Thus, various polar ‘difficult’ analytes, as the herbicides bentazone, glyphosate and glufosinate, and pesticides degradation products, as ethylenetiourea, have been successfully determined in water. The effectiveness of this approach is illustrated in **Figure 3**, relative to the determination of MCPA residues using diode array detection (DAD). Using standards, the cleanup time was established to be 3.6 min in the C<sub>18</sub> column used as C-1, and the transfer time 1.3 min for the multiresidue determination of the herbicide MCPA and two of its main metabolites (**Figure 3A**). However, the single analysis of parent MCPA allowed increasing the cleanup up to 4 min, decreasing the transference time to 0.5 min. **Figure 3B** shows the presence of matrix interferences in olive leaves extracts when connecting directly the C<sub>18</sub> column to the detector. The small transfer window used for the single determination of the MCPA still contained a considerable amount of coextracted interferent compounds. However, coupling of this C<sub>18</sub> column (C-1) to a RAM column (C-2) provided satisfactory results allowing the trace-level determination of this herbicide in complex matrices (**Figure 3C**).

In multiresidue analysis, dealing with several analytes with a wide polarity range, larger transfer

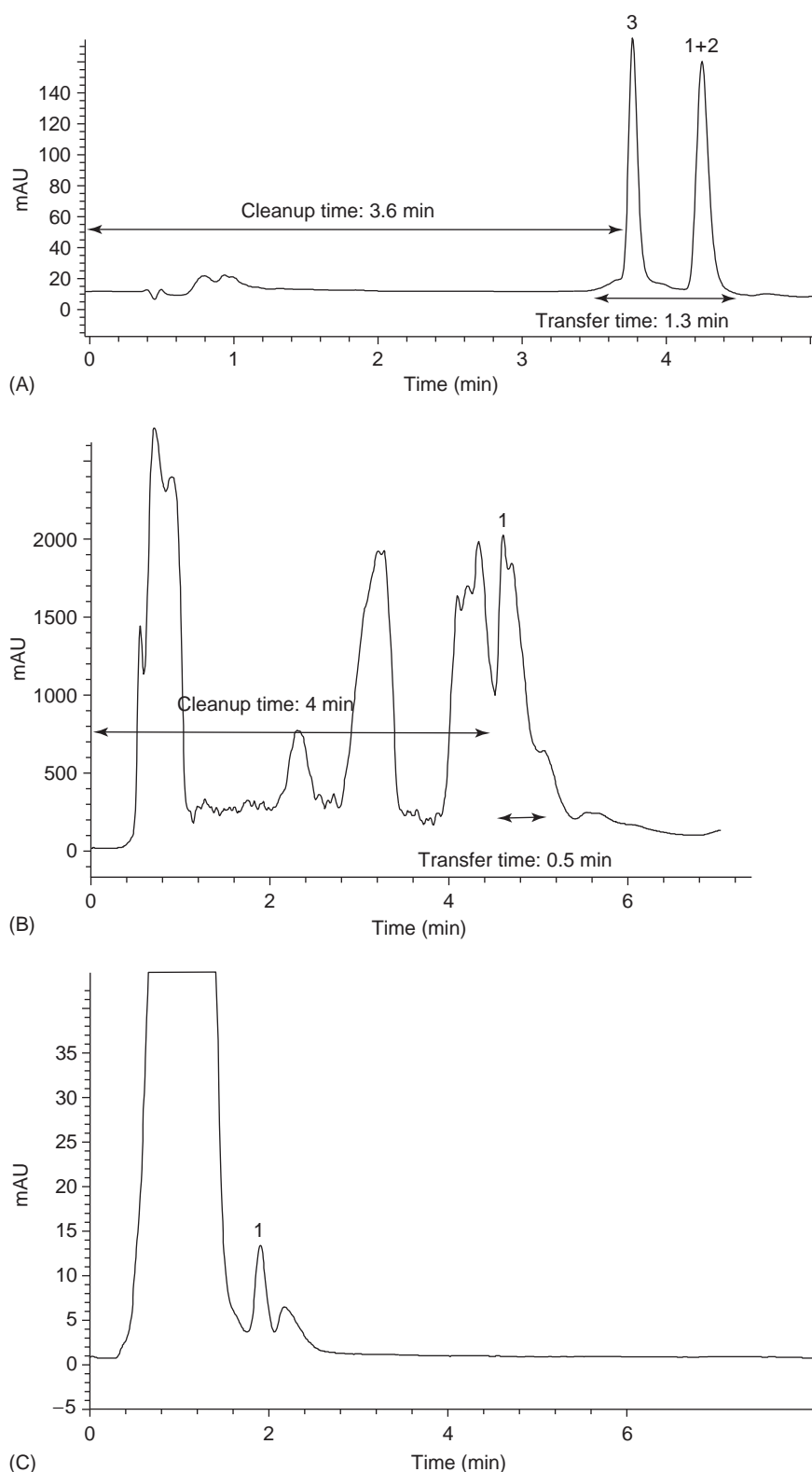
volumes are required, and consequently the selectivity is lower. However, the major source of interferences in environmental samples are usually the more polar compounds, as humic and fulvic acids (at least using conventional ultraviolet detection). In these cases, the use of coupled-column RPLC allows the effective removal of these early-eluting interferences. Therefore, multiresidue methods involving large transfer volumes, which, in principle, would lead to less selectivity, can still be quite attractive. Thus, several triazines and their transformation products, or a group of chlorofenoxiacids, have been determined at the sub-ppb levels in environmental samples with minimum sample treatment by using this approach.

## MDLC in Combination with MS

Multidimensional chromatography is one of the best options to prevent unwanted coelutions of analytes with other matrix components present in a variety of complex samples, and also to incorporate automated online sample preparation in the analytical process. Nowadays, well-developed methodologies already exist using multidimensional modes as LC–LC or SPE–LC in combination with spectroscopic or electrochemical detection for the selective and sensitive determination of many analytes, mainly in the biomedical, pharmaceutical, environmental, and food fields.

At present, the selectivity improvement is easily achieved by coupling simple mono-dimensional chromatographic systems to MS. Today, easy-to-operate benchtop MS (/MS) systems (single/triple quadrupole or ion-trap) with robust interfaces are commercially available, and LC–MS(/MS) has become the method of choice in many quantitative bioanalytical and food applications, its use also being increased in the environmental field. In spite of the high selectivity (and sensitivity) of tandem MS techniques, some kind of sample preparation and preseparation still remains necessary in order to avoid the analyte response decrease by matrix effects (ion suppression) observed in the majority of analyte/sample combinations as well as to avoid source contamination by matrix loading.

The use of LC–MS(/MS) could certainly decrease the interest of multidimensionality in many applications, as no highly efficient separations will be normally required. However, extremely complex samples still need the improved peak capacity offered by MDLC. Thus, comprehensive LC–LC in combination with MS(/MS) is one of the most powerful approaches for the entire profiling of proteins and



**Figure 3** (A) LC chromatogram ( $\lambda = 228$  nm) for a mixture of MCPA (1) and its metabolites 4-chloro-2-methylphenol (2) and 4-chlorophenol (3) obtained by injection of a  $400 \text{ ng ml}^{-1}$  standard solution, connecting C-1 ( $30 \times 4.6 \text{ mm ID C}_{18} 5 \mu\text{m}$ ) directly to a DAD detector. Optimization of the LC-LC coupling conditions. (B) LC chromatogram ( $\lambda = 228$  nm) of an olive leaves extract spiked with MCPA at  $4 \text{ mg kg}^{-1}$ , obtained connecting C-1 ( $30 \times 4.6 \text{ mm C}_{18} 5 \mu\text{m}$ ) directly to the detector. (C) LC-LC chromatogram ( $\lambda = 228$  nm) of an olive leaves extract containing  $0.4 \text{ mg kg}^{-1}$  MCPA. A  $30 \times 4.6 \text{ mm}, 5 \mu\text{m}, \text{C}_{18}$  column used as C-1, and a  $50 \times 4.6 \text{ mm}, 5 \mu\text{m}$  ( $100 \text{ \AA}$  pore size) Hisep column used as C-2. Coupling conditions shown in (B).

peptide mixtures of unknown composition. Moreover, the automated and efficient cleanup offered by targeted LC–LC justifies its use in combination with MS/MS systems in the development of advanced and accurate methods with very little sample treatment. As an example, the use of LC–LC–MS/MS has been recently proposed for the rapid determination of low concentration levels of xenobiotics and metabolites in biological human fluids. Additionally, highly diluted samples take advantage of the high sensitivity obtainable by online trace enrichment (SPE–LC). The combination of SPE–LC with MS/MS leads to sensitive and selective analytical methods, which are suitable for the monitoring of a variety of organic micropollutants in environmental waters, as a typical application.

*See also:* **Chromatography:** Multidimensional Techniques. **Environmental Analysis.** **Extraction:** Solid-Phase Extraction. **Food and Nutritional Analysis:** Sample Preparation; Contaminants; Pesticide Residues. **Forensic Sciences:** Drug Screening in Sport; Illicit Drugs. **Herbicides.** **Liquid Chromatography:** Instrumentation; Clinical Applications; Food Applications. **Mass Spectrometry:** Peptides and Proteins. **Pesticides.** **Pharmaceutical Analysis:** Sample Preparation. **Proteomics.** **Sample Handling:** Automated Sample Preparation. **Water Analysis:** Organic Compounds.

## Further Reading

- Barcelo D and Hennion MC (1997) On-line sample handling strategies. In: *Trace Determination of Pesticides and Their Degradation Products in Water*, pp. 357–427. Amsterdam: Elsevier.
- Cortes HJ (ed.) (1990) *Multidimensional Chromatography: Techniques and Applications*. New York: Dekker.

- Hennion MC (1999) Solid phase extraction: Method development, sorbents and coupling with liquid chromatography. *Journal of Chromatography A* 856: 3–54.
- Hernández F, Hidalgo C, Sancho JV, and López FJ (1998) Coupled-column liquid chromatography applied to the trace-level determination of triazine herbicides and some of their metabolites in water samples. *Analytical Chemistry* 70: 3322–3328.
- Hernandez F, Hidalgo C, Grimalt S, and Sancho JV (2001) Application of column-switching liquid chromatography to the determination of polar herbicides in different environmental matrices. MCPA as a case study. *Química Analítica* 20: 81–91.
- Hogendoorn EA and van Zoonen P (1995) Coupled column liquid chromatography in environmental analysis. *Journal of Chromatography A* 703: 149–166.
- Mondello L, Lewis AC, and Bartle KD (eds.) (2002) *Multidimensional Chromatography*. New York: John Wiley.
- Ramsteiner KA (1988) Systematic approach to column switching. *Journal of Chromatography A* 456: 3–20.
- Regnier F and Huang G (1996) Future potential of targeted component analysis by multidimensional liquid chromatography–mass spectrometry. *Journal of Chromatography A* 750: 3–10.
- Venkatramani CJ and Zelechunok Y (2003) An automated orthogonal two-dimensional liquid chromatograph. *Analytical Chemistry* 75: 3152–3162.
- Wang H and Hanash S (2003) Multidimensional liquid phase based separations in proteomics. *Journal of Chromatography B* 787: 11–18.
- Wehr T (2003) Multidimensional liquid chromatography in proteomic studies. *LC–GC Europe* 16: 154–162.
- Wu D, Berna M, Maier G, and Johnson J (1997) An automated multidimensional screening approach for rapid method development in high performance liquid chromatography. *Journal of Pharmaceutical and Biomedical Analysis* 16: 57–68.

## Instrumentation

R P W Scott, Sedlescombe, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The modern liquid chromatograph appears as an impressive, very elaborate, and technically advanced combination of electronic, mechanical, and hydraulic devices. This impression, however, is a little misleading as, although the overall instrument can be made very complex by the addition of a large number of ancillary devices, the ‘basic’ liquid chromatograph

is relatively simple incorporating well-established standard components. Even in its simplest form, the basic liquid chromatograph can probably still deal with over 80% of all liquid chromatography (LC) applications. In the discussion of LC instrumentation, therefore, the complex versatile instrument will be considered as an embellished form of the basic liquid chromatograph. In this way, as the sample becomes more challenging, it will be clear which specific ancillary equipment will be necessary to achieve the required separation and it will be possible to choose an optimum system based on rational discrimination. A diagram of a modern versatile

peptide mixtures of unknown composition. Moreover, the automated and efficient cleanup offered by targeted LC–LC justifies its use in combination with MS/MS systems in the development of advanced and accurate methods with very little sample treatment. As an example, the use of LC–LC–MS/MS has been recently proposed for the rapid determination of low concentration levels of xenobiotics and metabolites in biological human fluids. Additionally, highly diluted samples take advantage of the high sensitivity obtainable by online trace enrichment (SPE–LC). The combination of SPE–LC with MS/MS leads to sensitive and selective analytical methods, which are suitable for the monitoring of a variety of organic micropollutants in environmental waters, as a typical application.

*See also:* **Chromatography:** Multidimensional Techniques. **Environmental Analysis.** Extraction: Solid-Phase Extraction. **Food and Nutritional Analysis:** Sample Preparation; Contaminants; Pesticide Residues. **Forensic Sciences:** Drug Screening in Sport; Illicit Drugs. **Herbicides.** **Liquid Chromatography:** Instrumentation; Clinical Applications; Food Applications. **Mass Spectrometry:** Peptides and Proteins. **Pesticides.** **Pharmaceutical Analysis:** Sample Preparation. **Proteomics.** **Sample Handling:** Automated Sample Preparation. **Water Analysis:** Organic Compounds.

## Further Reading

- Barcelo D and Hennion MC (1997) On-line sample handling strategies. In: *Trace Determination of Pesticides and Their Degradation Products in Water*, pp. 357–427. Amsterdam: Elsevier.
- Cortes HJ (ed.) (1990) *Multidimensional Chromatography: Techniques and Applications*. New York: Dekker.

- Hennion MC (1999) Solid phase extraction: Method development, sorbents and coupling with liquid chromatography. *Journal of Chromatography A* 856: 3–54.
- Hernández F, Hidalgo C, Sancho JV, and López FJ (1998) Coupled-column liquid chromatography applied to the trace-level determination of triazine herbicides and some of their metabolites in water samples. *Analytical Chemistry* 70: 3322–3328.
- Hernandez F, Hidalgo C, Grimalt S, and Sancho JV (2001) Application of column-switching liquid chromatography to the determination of polar herbicides in different environmental matrices. MCPA as a case study. *Química Analítica* 20: 81–91.
- Hogendoorn EA and van Zoonen P (1995) Coupled column liquid chromatography in environmental analysis. *Journal of Chromatography A* 703: 149–166.
- Mondello L, Lewis AC, and Bartle KD (eds.) (2002) *Multidimensional Chromatography*. New York: John Wiley.
- Ramsteiner KA (1988) Systematic approach to column switching. *Journal of Chromatography A* 456: 3–20.
- Regnier F and Huang G (1996) Future potential of targeted component analysis by multidimensional liquid chromatography–mass spectrometry. *Journal of Chromatography A* 750: 3–10.
- Venkatramani CJ and Zelechunok Y (2003) An automated orthogonal two-dimensional liquid chromatograph. *Analytical Chemistry* 75: 3152–3162.
- Wang H and Hanash S (2003) Multidimensional liquid phase based separations in proteomics. *Journal of Chromatography B* 787: 11–18.
- Wehr T (2003) Multidimensional liquid chromatography in proteomic studies. *LC–GC Europe* 16: 154–162.
- Wu D, Berna M, Maier G, and Johnson J (1997) An automated multidimensional screening approach for rapid method development in high performance liquid chromatography. *Journal of Pharmaceutical and Biomedical Analysis* 16: 57–68.

## Instrumentation

R P W Scott, Sedlescombe, UK

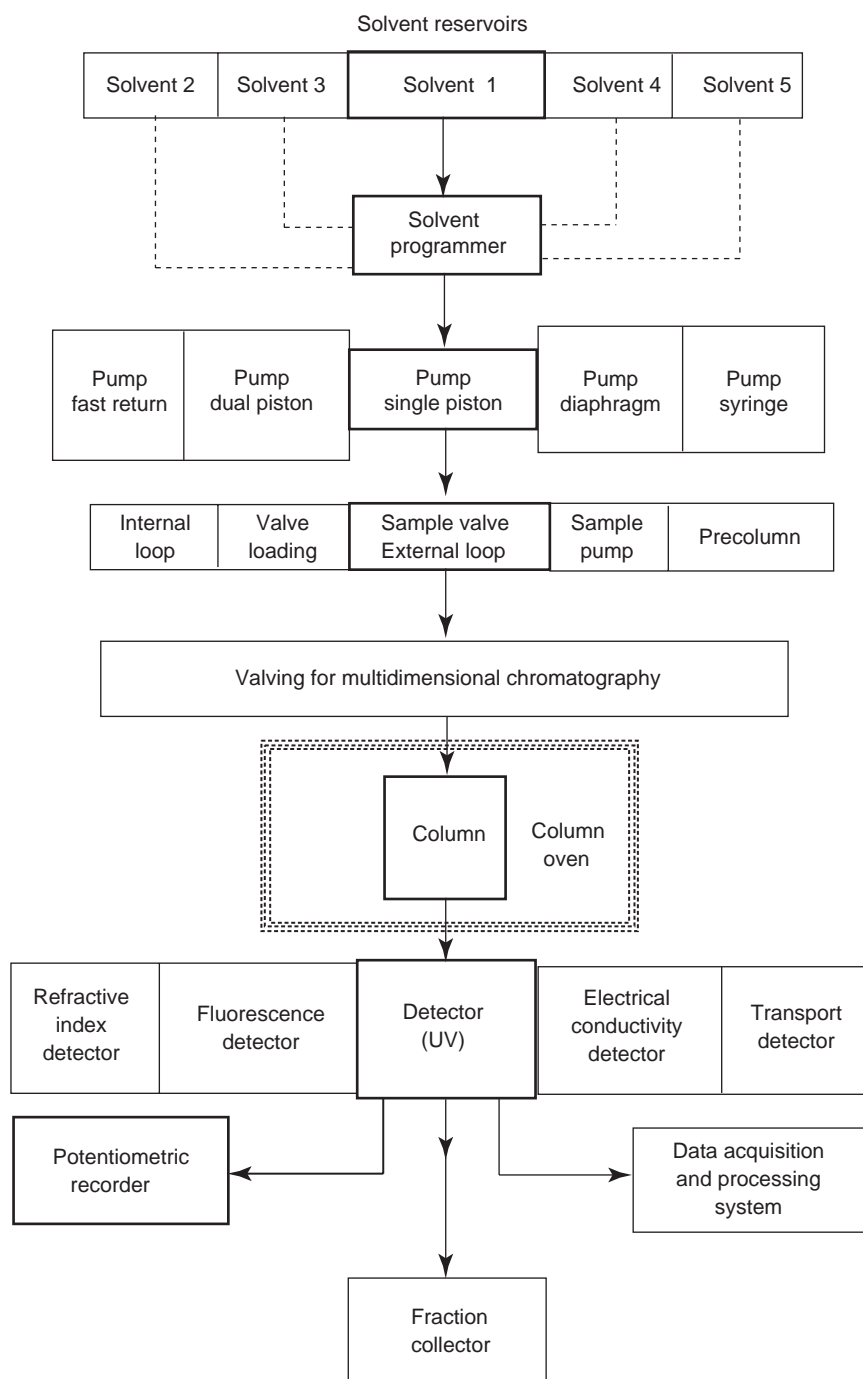
© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The modern liquid chromatograph appears as an impressive, very elaborate, and technically advanced combination of electronic, mechanical, and hydraulic devices. This impression, however, is a little misleading as, although the overall instrument can be made very complex by the addition of a large number of ancillary devices, the ‘basic’ liquid chromatograph

is relatively simple incorporating well-established standard components. Even in its simplest form, the basic liquid chromatograph can probably still deal with over 80% of all liquid chromatography (LC) applications. In the discussion of LC instrumentation, therefore, the complex versatile instrument will be considered as an embellished form of the basic liquid chromatograph. In this way, as the sample becomes more challenging, it will be clear which specific ancillary equipment will be necessary to achieve the required separation and it will be possible to choose an optimum system based on rational discrimination. A diagram of a modern versatile





**Figure 1** A modern liquid chromatograph.

liquid chromatograph is shown, in block form, in **Figure 1**.

It is seen that the basic liquid chromatograph consists of a solvent reservoir, a pump, a sample valve, column, detector, and a potentiometric recorder. The main differences between the simple device and the versatile instrument is, first, the chromatographer who designs the separation methodology must have a good understanding of molecular interactive

processes between the sample components and the two phases so that the separation can be achieved using a single solvent (i.e., isocratically). Second, the components of the mixture of interest must contain an ultraviolet (UV) chromophore so that it can be detected by a UV detector (or a property appropriate for any other type of detector that might be chosen). Third, the results must be calculated from direct measurements on the recorder chart as opposed to a

computer data acquisition and processing system. The relative advantages of these latter two procedures will be discussed later.

## The Solvent Reservoir(s)

The solvent reservoirs are normally made of glass or stainless steel. Glass reservoirs are usually necessary for ionic mobile phases (such as those used in ion chromatography and 'paired ion' chromatography) to avoid corrosion. If only nonionic solvents are to be used, then stainless steel is quite satisfactory. The reservoir volume can vary from 250 to 1000 ml and each reservoir should be fitted with an appropriate 'sparging' apparatus. Most solvents contain a significant amount of dissolved air (oxygen and nitrogen) and on leaving the column, bubbles are often formed in the mobile phase that, when passing through the detector, can seriously impair its performance. The sparging procedure consists of bubbling a continuous stream of an 'insoluble' gas (e.g., helium) through the solvent that elutes the dissolved gases before entering the column.

Some samples contain solutes that have a wide range of interactivity with the stationary phase. Under such conditions solvent programming may be necessary to elute the more strongly held solutes in reasonable time. Thus, a number of solvent reservoirs may be necessary. Generally, by careful choice of stationary phase and solvents, even challenging separations can often be achieved with only two solvents but, sometimes, for certain difficult mixtures (e.g., chiral separations) very subtle interactions between solutes and solvents may be necessary and three different solvents may be required. It follows that three or more solvent reservoirs will be included in the more versatile chromatograph.

## Solvent Programmers

In the basic chromatograph the solvent flows directly from the reservoir to the pump, in the more versatile solvent systems different solvents are passed to a solvent programmer. The solvent programmer adjusts the flow of each solvent so that, at any given time, the resulting mixture has a known and predefined composition. The program is usually controlled either by the main chromatograph computer or a microprocessor associated with the programmer itself. There are two basic types of programmer, arbitrarily referred to as the 'high-pressure programmer' and the 'low-pressure programmer'. The high-pressure programmer is not commonly used as it requires a pump for each solvent and is, thus, rather expensive. The pumps are actuated by stepping

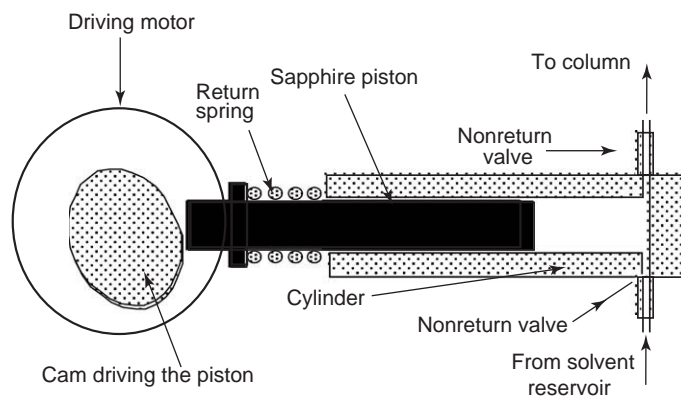
motors that are supplied with square wave pulses of different frequency that provide different flow rates. The different solvent flows are then passed through a mixer or mixing chamber to the sample valve. The frequency of the square wave pulses is computer or microprocessor controlled. The low-pressure programmer involves the use of a single pump that draws solvent from a series of valves (one valve for each solvent) that are also actuated by square wave pulses. The relative amount of solvent drawn through a valve depends on the period it is open (active) and the period it is closed (inactive) (i.e., the 'mark-to-space' ratio of the square wave). The mark-to-space ratio of the square wave is also computer or microprocessor controlled. The solvent from each valve passes to a mixing chamber and then to the pump. The gradient, gradient form, initial flow rate, and final flow rate are all user defined by the gradient program. Linear, concave, and convex gradient profiles are usually available and reverse gradients facilities are also possible.

## Pumps

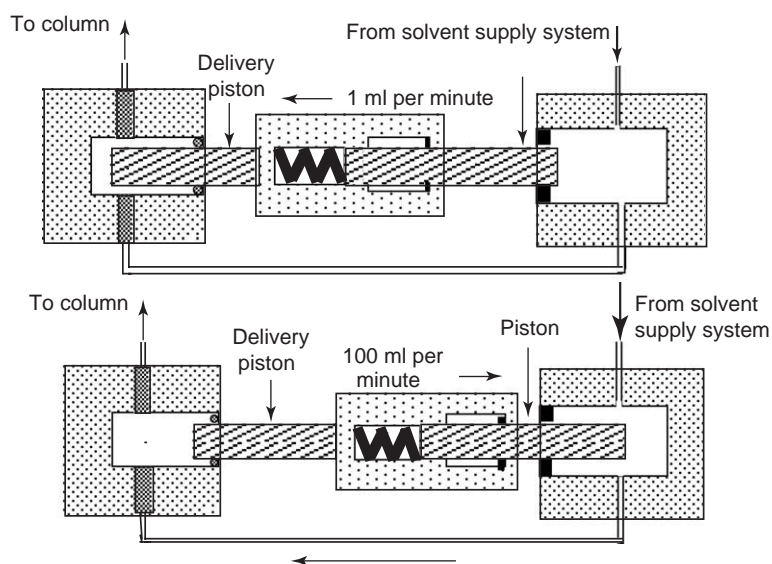
Pump deliveries vary with the application for which they are to be used. For general use, flow rates of  $0.1\text{--}10\text{ ml min}^{-1}$  are common but if microbore columns are used then flow rates as low as  $5\text{--}500\text{ }\mu\text{l min}^{-1}$  might be more appropriate. Pumps should operate at pressures up to at least 6000 psi or even 10 000 psi but pressures above these are not practical as sample valves cannot operate for very long at pressures in excess of 6000 psi without leaking. In fact, for a reasonably long life, most sample valves should not be operated continuously much above 3000 psi, which also limits the maximum operating pressure of the pump. There are a number of different types of pump that can provide the necessary pressures and flow rates required by the modern liquid chromatograph. Most chromatographs are fitted with reciprocating pumps actuated by either diaphragms or, more commonly, pistons. Examples of three of the more common pumps employed in contemporary liquid chromatographs are shown in Figure 2.

The single piston reciprocating pump was the first of its type to be used in liquid chromatographs, it is the simplest, the least expensive, and is still very popular today. The single piston pump is shown at the top of Figure 2. The piston is made of synthetic sapphire; in fact, most pistons of modern LC pumps are made of sapphire to reduce wear and extend the working life of the pump. The cylinder is usually made of stainless steel, which incorporates two non-return valves in line with the inlet and outlet connections to the pump. The piston is driven by a

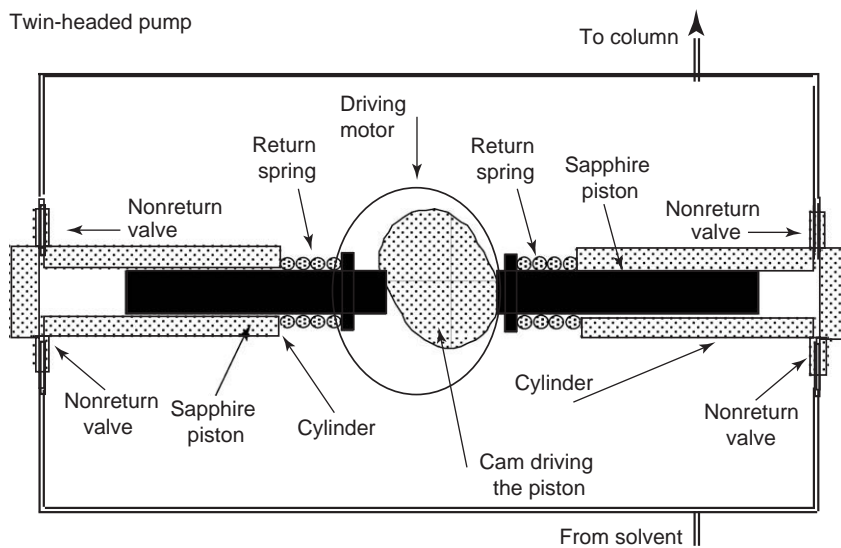
Single piston pump



Rapid refill pump



Twin-headed pump

**Figure 2** Different types of LC pumps.

stainless steel cam that forces the piston into the cylinder expressing the solvent through the exit non-return valve. After reaching the maximum movement, the piston follows the cam where it meets a step and, as a result of the pressure exerted by the return spring, rapidly returns the piston and refills the cylinder. The shape of the cam is cut to provide a linear movement of the piston during expression of the solvent, but a sudden return movement on the refill stroke. In this way the pulse effect is reduced. However, although the pulses are not completely eliminated in this manner, the damping effect of the modern high-efficiency LC column, packed with particles a few micrometers in diameter, ensures that any remaining pulses are strongly attenuated before they reach the detector. In addition, most detector sensors are designed to be both flow and pressure insensitive and so single piston pumps can usually be employed with confidence.

The rapid refill pump, shown at the center of **Figure 2**, consists of two cylinders and a single common piston. The expression of the solvent to the column is depicted in the upper part of the figure. As the piston progresses to the right, solvent is pumped to the column system, but at the same time fresh solvent is being drawn into the right-hand chamber from the solvent supply system. At the point where the piston arrives at the extent of its travel, a step in the driving cam is reached and the piston is rapidly reversed. As a result the contents of the chamber on the right-hand side are conveyed to the left-hand chamber. This situation is depicted in the lower part of the central figure. The transfer rate of the solvent to the left-hand chamber is 100 times as fast as the delivery rate to the column and consequently reduces the pulse on refill very significantly. Furthermore, if a solvent gradient is being used while the right-hand chamber is being filled, excellent mixing is achieved. However, the refill cylinder volume will cause a series of stepped changes in solvent composition during gradient elution instead of a smooth continuous change that may or may not be deleterious to the quality of the separation.

An alternative approach to the removal of pump pulses, which is probably the more successful but, as one might expect, the more expensive, is the use of twin pump heads. In a two-headed pump, one cylinder can be filled, while the other is delivering solvent to the column. The cylinders and pistons of a two-headed pump are constructed in a very similar manner to the single-piston pump with a sapphire piston and a stainless steel cylinder. Each cylinder is fitted with nonreturn valves both at the inlet and the outlet. The cams that drive the two pistons are carefully cut to provide an increase in flow from one pump while the

other pump is being filled to compensate for the loss of delivery during the refill process and, thus, a fall in pressure. A diagram of a twin-headed pump is shown at the bottom of **Figure 2**. It is seen that there is a common supply of mobile phase from the solvent reservoir or solvent programmer to both pumps and the output of each pump joins and the solvent then passes to the sample valve and then to the column. In the diagram, a single cam drives both pistons, but in practice, to minimize pressure pulses, each pump usually has its own cam drive from the motor. The displacement volume of each pump can vary from 20 or 30  $\mu\text{l}$  to over 1 ml but the usual displacement volume is  $\sim 250 \mu\text{l}$ . The pump is driven by a stepping motor and, thus, the delivery depends on the frequency of the supply fed to the motor, which is controlled by a microprocessor or the main computer. Some pumps are fitted with pulse suppressers that can take a variety of forms. Pulse suppressers, however, do introduce extra column volumes in the solvent conduit to the column, which can significantly distort the expected concentration profile during gradient elution.

There are other pumps used in LC. The 'diaphragm' pump is similar to a piston pump except that the piston is replaced by a moving diaphragm. This type of pump is often used in preparative and semipreparative chromatography. One of the early pumps to be used was the 'syringe' pump, which was, in effect, a large, mechanically operated hypodermic syringe. The capacity of this type of pump is limited and, thus, frequently requires refilling. It is now largely used for injecting reagents in postcolumn derivatization. Another pump is the 'pneumatic' pump, which can probably provide the highest pressures. This pump is driven by compressed air and, by using widely different piston cross-sections for the air and solvent cylinders, can provide very large pressure amplification and, thus, very high pressures. This type of pump provides very large pulses when refilling and is now mostly used for column packing.

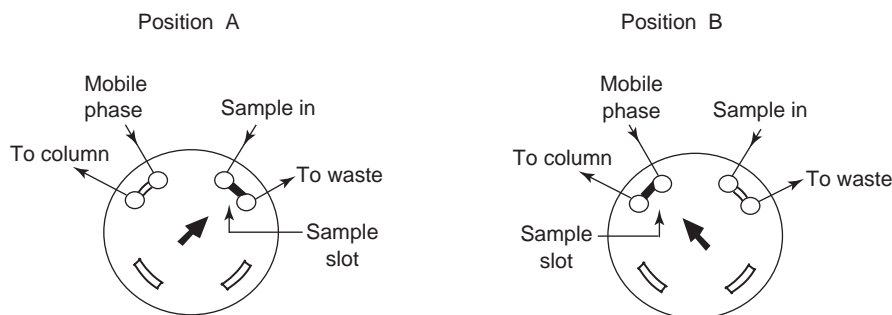
## Sample Valves

There are two types of sample valve in general use in analytical LC: the internal loop valve and the external loop valve. The two types of sample valve are shown diagrammatically in **Figure 3**.

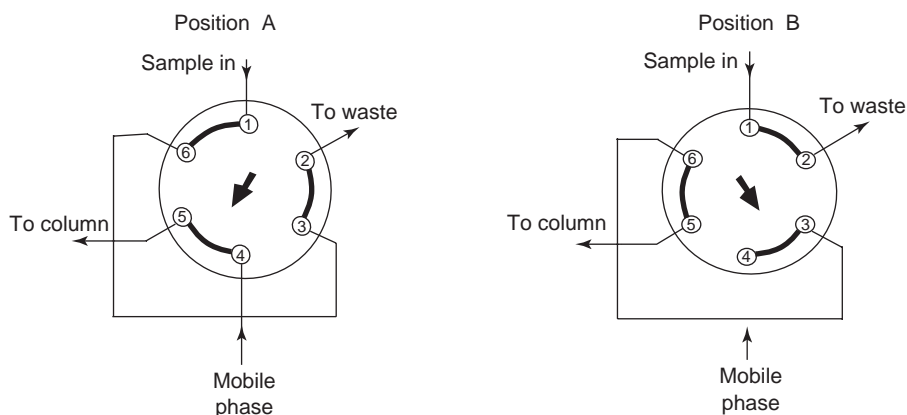
### The Internal Loop Valve

The internal loop valve has only four ports and the sample volume is contained in the connecting slot of the valve rotor. As a consequence, this type of valve is used only for relatively small sample volumes. Internal sample loop valves deliver samples ranging from 0.1 to  $\sim 0.5 \mu\text{l}$  in volume. The operation of the valve

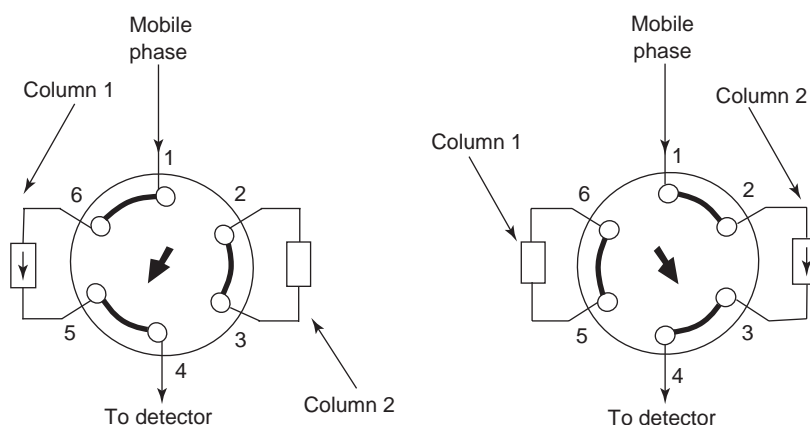
The internal loop valve



The external loop valve



Column-switching valve arrangements

**Figure 3** Different sample valve types.

is depicted in **Figure 3**, the diagram on the left shows the sample being loaded into the valve. The sample occupies the rotor slot on the upper right and has been filled by passing the sample from an appropriate syringe through the rotor slot to waste. While loading the sample, the mobile phase passes through the valve and direct to the column. The valve is then

rotated and the sample position is depicted on the right of **Figure 3**. The valve slot is now imposed between the mobile phase supply and the column and, consequently, the sample is swept onto the column by the solvent. This type of valve is used when the sample volume can have a critical effect on the net peak dispersion. A typical example would be in the



use of short columns, perhaps 4.6 mm in diameter and 2–5 cm long packed with support particles only 3  $\mu$ m in diameter. Alternatively, they would be used with relatively long columns (10–30 cm in length) but only  $\sim$ 1 mm or less in diameter.

It must be emphasized that if the sample contains labile material of biological origin, then all parts of the chromatographic system that can come in contact with the sample must be constructed of a biocompatible material. In practice, this means that every part between the sample valve and the detector should be made from titanium or some other appropriate material such as PEEK, or where there is little or no pressure, Teflon.

### The External Loop Valve

The external sample loop injector is the more popular type of sample valve and has six ports as depicted in **Figure 3**. Three slots are cut in the rotor so that any adjacent pair of ports can be connected. In the loading position shown on the left, the mobile phase supply is connected by a rotor slot to port (4) and the column to port (5), allowing the mobile phase to flow directly through the column. In this position, the sample loop is connected to ports (3) and (6). Sample flows from a syringe into port (1) through the rotor slot to the sample loop at port (6). At the same time the third slot in the rotor connects the exit of the sample loop to waste at port (2).

The sampling position is shown on the right of the diagram. On rotating the valve, the sample loop is interposed between the column and the mobile phase supply by connecting ports (3) and (4) and ports (5) and (6). As a consequence, the sample is swept onto the column by the mobile phase. In the sampling position, the third rotor slot connects the syringe port to the waste port. The sample loop can take a wide range of volumes, those used for analytical purposes usually falling within the range of 1–20  $\mu$ l, but those for preparative separations may be as large as 100 ml or more. Some external loop valves are fitted with a seventh port with a silicone septum on the face of the valve. This allows the sample to be injected directly into the sample loop by a hypodermic syringe.

Samples can be placed on the column in other ways. In preparative chromatography, the sample is often very large and may be pumped directly onto the column from a sample reservoir employing a high-pressure pump similar to that used in the mobile phase supply system. Alternatively, with particularly difficult separations, the sample may be placed on the column from the eluent of a precolumn, using column switching (often described as two-dimensional chromatography). For example, the sample may be first

separated on the basis of molecular size using size exclusion chromatography and, by a suitable valve arrangement, a sample of selected molecular weight may be diverted to a reversed-phase column and the fraction separated on the basis of dispersive interactions. This significantly increases the discriminating capability of the overall chromatography system.

After passing through the sample valve and prior to entering the column, a separate valve can be used to select a particular type of column that can increase the versatility of the liquid chromatograph significantly. An example of a six-port valve arranged for column switching is shown at the base of **Figure 3**. The same valve is used as that for external loop sampling. Column (1) is connected between ports (5) and (6) and column (2) between ports (2) and (3). The mobile phase from a sample valve (or sometimes from another column) enters port (1) and the detector is connected to port (4). In the initial position of the rotor shown in the diagram on the left-hand side, the rotor slots connect ports (1) and (6), (2) and (3), and (4) and (5). This results in the mobile phase passing from port (1) to port (6), through column (1) to port (5), from port (5) to port (4), and out to the detector. Thus, the separation takes place in column (1). The ports connected to column (2) are themselves connected by the third slot and thus isolated. When the valve is rotated, port (1) is connected to port (2), port (3) is connected to port (4), and port (5) is connected to port (6). Thus, the mobile phase from either a sample valve or another column enters port (1) passes to port (2) through column (2) to port (3), then to port (4), and then to the detector. Ports (5) and (6) are connected, this time isolating column (1). This arrangement allows either one of two columns to be selected for an analysis or part of the eluent from another column pass to column (1) for separation and the rest passed to column (2).

For multiple analyses of a routine nature, automatic samplers are often employed. One type consists of a carousel of septum-capped vials containing different samples interspersed by vials containing pure solvent for washing purposes. The septum is automatically pierced by two needles one dipping into the liquid and one dipping only into the air space above the liquid. A small pressure of gas is applied to the air space which forces some liquid through the sampling tube that is connected to a normal external loop sample valve and then passes through the sample loop. After an appropriate time the pressure is removed and the next vial is placed in the sampling position. The valve loop is first washed with solvent then filled with sample that is then injected onto the column by a pneumatic actuator. This procedure is repeated for each subsequent sample.

## Columns

Columns vary widely in size depending on the nature of the sample and the scale of the separation. The radius can range from a fraction of a millimeter to several centimeters and in large-scale separations column diameters may be 10 cm or more. Columns can also vary in length from 2 or 3 cm to a meter. For most analytical separations columns of 2–4 mm ID and 3–25 cm length are used, packed with particles 3–10  $\mu\text{m}$  in diameter. Small diameter columns (1 mm or less) provide the highest mass sensitivity and are used when there is limited sample. Wide columns are used for preparative and semipreparative work. Most stationary phases are bonded to silica although polymeric materials are also very popular. The bonded phase can range in character from strongly dispersive to strongly polar and include anion and cation exchangers. However, the most popular ion-exchange media are based on cross-linked polystyrene beads.

In the early days of LC, the effect of temperature on separations was not considered to be very profound and columns were rarely thermostatted, but operated at room temperature. Over the last decade, however, with the advent of fast, high-efficiency columns and the challenges of very difficult separations (such as those met in chiral chromatography) it was found that temperature can have a very strong effect on retention; particularly 'relative' retention. In fact, although it could be theoretically predicted that in many separations there will be a temperature of co-elution for a given pair of solutes at a particular solvent composition, this was not experimentally verified until the 1980s. It follows that a column oven became an important feature of all liquid chromatographs and that it was advisable to thermostat an LC column to  $\pm 0.5^\circ\text{C}$  for accurate and precise results. All modern chromatographs are fitted with temperature-controlled ovens, many with refrigeration units allowing temperatures ranging from 0 to  $100^\circ\text{C}$  to be employed if so desired.

## Detectors

The detector is a device that senses the presence of a solute in the eluent from a chromatographic column and, thus, the output, suitably presented (for example, on a chart recorder) can display the nature of the separation and permit a quantitative assessment of the sample components. Excluding the column, the detector is the most important component of the liquid chromatograph as its design determines the accuracy and precision of the LC analysis. Many detectors have been developed based on a variety of different sensing principles, but only  $\sim 12$  are

presently used in modern LC and, of those 12, perhaps only four dominate in general use. The more popular four are the 'UV detector' (fixed and variable wavelength), the 'electrical conductivity detector', the 'fluorescence detector', and the 'refractive index detector'. The refractive index, although the least sensitive of the three detectors, is included as it has an almost universal response. It is particularly useful in the detection of polymeric material (particularly biopolymers) as its response is directly proportional to the concentration of the polymer and virtually independent of the molecular weight.

### Detector Specifications

There are seven major detector specifications which are as follows:

- detector linearity,
- linear dynamic range,
- detector noise level,
- detector sensitivity, or minimum detectable concentration,
- pressure sensitivity,
- flow sensitivity, and
- temperature sensitivity.

The two most important specifications are sensitivity (minimum detectable concentration) and linearity.

Detector sensitivity has been defined as the minimum concentration of an eluted solute that can be unambiguously differentiated from the noise. The ratio of the signal to the noise for a peak that is considered decisively identifiable has been arbitrarily chosen to be two (a ratio employed in electronic theory). Thus, the sensitivity or minimum detectable concentration is that concentration that provides a signal equivalent to twice the noise level. Unfortunately, the sensitivity defined in this way will usually depend on the physical properties of the solute used for measurement and, thus, must be defined.

The detector linearity can be defined from the following relationship:

$$v = Rc^\alpha$$

where  $v$  is the output from the detector,  $R$  is a constant, and  $\alpha$  is the 'response index'.

For a truly linear detector the value of  $\alpha$  will be unity, and for accurate quantitative analysis, the value of  $\alpha$  should be between 0.97 and 1.03. The general specifications for the four, more common detectors are given in Table 1.

The detector may be incorporated in the total chromatograph or connected to the column as a separate component. The latter is quite popular as it allows the detector to be changed, if so required, and permits greater flexibility. For preparative work the

**Table 1** Typical specifications for the four commonly used LC detectors

<i>The UV detectors</i>	
Fixed wavelength UV detector	
Sensitivity (toluene)	$5 \times 10^{-8} \text{ g ml}^{-1}$
Linear dynamic range	$5 \times 10^{-8}$ to $5 \times 10^{-4} \text{ g ml}^{-1}$
Response index	0.98–1.02
The multi-wavelength UV detector	
Sensitivity	$1 \times 10^{-7} \text{ g ml}^{-1}$
Linear dynamic range	$5 \times 10^{-7}$ to $5 \times 10^{-4} \text{ g ml}^{-1}$
Response index	0.98–1.02
Typical specifications for a refractive index detector	
Sensitivity (benzene)	$1 \times 10^{-6} \text{ g ml}^{-1}$
Linear dynamic range	$1 \times 10^{-6}$ to $1 \times 10^{-4} \text{ g ml}^{-1}$
Response index	0.97–1.03
Typical specifications for a fluorescence detector	
Sensitivity (anthracene)	$1 \times 10^{-9} \text{ g ml}^{-1}$
Linear dynamic range	$1 \times 10^{-9}$ to $5 \times 10^{-6} \text{ g ml}^{-1}$
Response index	0.96–1.04
Typical specifications for an electrical conductivity detector	
Sensitivity	$5 \times 10^{-9} \text{ g ml}^{-1}$
Linear dynamic range	$5 \times 10^{-9}$ to $1 \times 10^{-6} \text{ g ml}^{-1}$
Response index	0.97–1.03

detector outlet is often connected to a fraction collector. For semipreparative work, this usually consists of a carousel of collecting tubes that can be alternately and automatically filled by the eluent from the detector outlet. The carousel rotation can either be actuated on a time basis or by the detector signal. For large-scale separations, usually only one or two products are of interest, in which case a three-way valve is connected to the detector outlet that can divert the column eluent to waste or to one of two collection vessels. This valve can also be actuated on a time basis or from the detector signal.

## Data Processing and Recording

The best method of recording chromatographic data and its processing depends heavily on the complexity of the sample and the quality of the separation. If the sample is relatively simple, or only a small number of solutes are of interest, and those of interest are well

resolved, then peakheight measurement can be used for quantitative assessment. Under such circumstances, manual measurements (made on the recorder chart) will provide as fast and as accurate an analysis as a complex computer data processing system. In fact, it was shown many years ago that the most accurate results were obtained manually by cutting the peak out of a recorder chart and weighing it. This procedure is, however, very tedious and time consuming and not really practical in modern analytical laboratories. If the sample is complex, however, and the solutes are not completely resolved then, without doubt, computer data acquisition and processing will be, by far, the fastest procedure. The accuracy of the results will, however, depend heavily on the 'quality of the separation' and the 'integrity' of the processing software that may differ significantly from one supplier to another.

It should be emphasized that clever software is no substitute for good chromatography. Good chromatography is attained from a sound, fundamental knowledge of the technique coupled with experience and experimental skill. It is not achieved solely by the astute choice of complex and expensive apparatus.

See also: **Liquid Chromatography:** Packed Capillary; Multidimensional; Liquid Chromatography–Mass Spectrometry.

## Further Reading

- Beesley TE and Scott RPW (1998) *Chiral Chromatography*. Chichester: Wiley.
- Beesley TE, Buglio B, and Scott RPW (2001) *Quantitative Chromatographic Analysis*. New York: Dekker.
- Engelhardt H (1987) *Practice of High Performance Liquid Chromatography*. New York: Springer.
- Katz ED (1994) *High Performance Liquid Chromatography: principles and methods in biotechnology*. Chichester: Wiley.
- Scott RPW (1996) *Chromatographic Detectors*. New York: Dekker.
- Scott RPW (1992) *Liquid Chromatography Column Theory*. Chichester: Wiley.
- Scott RPW (1994) *Liquid Chromatography for the Analyst*. New York: Dekker.
- Scott RPW (1995) *Techniques and Practice of Chromatography*. New York: Dekker.
- Snyder L, Glajch JL, and Kirkland J (1997) *Practical HPLC Method Development*. New York: Wiley.

# Liquid Chromatography–Mass Spectrometry

W M A Niessen, Hyphen MassSpec Consultancy, Leiden, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Combined high-performance liquid chromatography–mass spectrometry (LC–MS) was originally

**Table 1** Typical specifications for the four commonly used LC detectors

<i>The UV detectors</i>	
Fixed wavelength UV detector	
Sensitivity (toluene)	$5 \times 10^{-8} \text{ g ml}^{-1}$
Linear dynamic range	$5 \times 10^{-8}$ to $5 \times 10^{-4} \text{ g ml}^{-1}$
Response index	0.98–1.02
The multi-wavelength UV detector	
Sensitivity	$1 \times 10^{-7} \text{ g ml}^{-1}$
Linear dynamic range	$5 \times 10^{-7}$ to $5 \times 10^{-4} \text{ g ml}^{-1}$
Response index	0.98–1.02
Typical specifications for a refractive index detector	
Sensitivity (benzene)	$1 \times 10^{-6} \text{ g ml}^{-1}$
Linear dynamic range	$1 \times 10^{-6}$ to $1 \times 10^{-4} \text{ g ml}^{-1}$
Response index	0.97–1.03
Typical specifications for a fluorescence detector	
Sensitivity (anthracene)	$1 \times 10^{-9} \text{ g ml}^{-1}$
Linear dynamic range	$1 \times 10^{-9}$ to $5 \times 10^{-6} \text{ g ml}^{-1}$
Response index	0.96–1.04
Typical specifications for an electrical conductivity detector	
Sensitivity	$5 \times 10^{-9} \text{ g ml}^{-1}$
Linear dynamic range	$5 \times 10^{-9}$ to $1 \times 10^{-6} \text{ g ml}^{-1}$
Response index	0.97–1.03

detector outlet is often connected to a fraction collector. For semipreparative work, this usually consists of a carousel of collecting tubes that can be alternately and automatically filled by the eluent from the detector outlet. The carousel rotation can either be actuated on a time basis or by the detector signal. For large-scale separations, usually only one or two products are of interest, in which case a three-way valve is connected to the detector outlet that can divert the column eluent to waste or to one of two collection vessels. This valve can also be actuated on a time basis or from the detector signal.

## Data Processing and Recording

The best method of recording chromatographic data and its processing depends heavily on the complexity of the sample and the quality of the separation. If the sample is relatively simple, or only a small number of solutes are of interest, and those of interest are well

resolved, then peakheight measurement can be used for quantitative assessment. Under such circumstances, manual measurements (made on the recorder chart) will provide as fast and as accurate an analysis as a complex computer data processing system. In fact, it was shown many years ago that the most accurate results were obtained manually by cutting the peak out of a recorder chart and weighing it. This procedure is, however, very tedious and time consuming and not really practical in modern analytical laboratories. If the sample is complex, however, and the solutes are not completely resolved then, without doubt, computer data acquisition and processing will be, by far, the fastest procedure. The accuracy of the results will, however, depend heavily on the 'quality of the separation' and the 'integrity' of the processing software that may differ significantly from one supplier to another.

It should be emphasized that clever software is no substitute for good chromatography. Good chromatography is attained from a sound, fundamental knowledge of the technique coupled with experience and experimental skill. It is not achieved solely by the astute choice of complex and expensive apparatus.

See also: **Liquid Chromatography:** Packed Capillary; Multidimensional; Liquid Chromatography–Mass Spectrometry.

## Further Reading

- Beesley TE and Scott RPW (1998) *Chiral Chromatography*. Chichester: Wiley.
- Beesley TE, Buglio B, and Scott RPW (2001) *Quantitative Chromatographic Analysis*. New York: Dekker.
- Engelhardt H (1987) *Practice of High Performance Liquid Chromatography*. New York: Springer.
- Katz ED (1994) *High Performance Liquid Chromatography: principles and methods in biotechnology*. Chichester: Wiley.
- Scott RPW (1996) *Chromatographic Detectors*. New York: Dekker.
- Scott RPW (1992) *Liquid Chromatography Column Theory*. Chichester: Wiley.
- Scott RPW (1994) *Liquid Chromatography for the Analyst*. New York: Dekker.
- Scott RPW (1995) *Techniques and Practice of Chromatography*. New York: Dekker.
- Snyder L, Glajch JL, and Kirkland J (1997) *Practical HPLC Method Development*. New York: Wiley.

# Liquid Chromatography–Mass Spectrometry

W M A Niessen, Hyphen MassSpec Consultancy, Leiden, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Combined high-performance liquid chromatography–mass spectrometry (LC–MS) was originally



**Table 1** Milestones in the development of combined liquid chromatography–mass spectrometry

1972	First paper on LC–MS by the group of Tal'roze
1974	A variety of LC–MS interfaces presented at the <i>9th International Symposium on Advances in Chromatography</i> , Houston, Texas
1976	Commercial introduction of the moving-belt interface
1980	Commercial introduction of the direct liquid introduction interface
1984	Commercial introduction of the thermospray interface
1987	Commercial introduction of continuous-flow fast-atom bombardment
1988	Commercial introduction of ionspray, the heated nebulizer, and the particle-beam interfaces
1990	Commercial introduction of the electrospray interface

developed as an alternative to gas chromatography–mass spectrometry (GC–MS) for those compounds that are not amenable to GC–MS. As such, the field of applications of LC–MS concerns the analysis of highly polar, thermolabile, and/or ionic compounds. The development of technologies to couple LC and MS started in the early 1970s along various lines simultaneously. The coupling of LC and MS takes place by means of a so-called interface, which is a dedicated piece of instrumentation installed onto the mass spectrometer to which the LC system is connected. A variety of LC–MS interfaces have been developed. Important milestones in the development of the LC–MS interface are summarized in **Table 1**.

## Principles

In developing online LC–MS, three major problems require attention: the amount of mobile phase that has to be introduced into the high vacuum of the MS, the composition of the mobile phase, and the nature of the components to be analyzed.

In order to avoid the loss of analyte, either as much of the LC effluent as possible has to be introduced into the mass spectrometer, or efficient analyte enrichment must be achieved in the LC–MS interface. Without modification, only  $\sim 50 \text{ nL s}^{-1}$  of liquid can be introduced into a differentially pumped MS vacuum system. Possible solutions to this flow-rate incompatibility problem are: enlargement of the pumping capacity of the MS vacuum system, removal of solvent prior to the introduction of the analytes into the high-vacuum region, splitting away part of the effluent at the cost of analyte loss, and miniaturization of the internal diameter of the LC column. The currently applied interface, based on atmospheric-pressure ionization (API), is discussed in more detail below.

There are rather severe restrictions with respect to the composition of the mobile phase. Although most

current API interfaces can deal with phosphate buffers in the mobile phase without significant instrumental problems, the use of nonvolatile buffers can lead to problems related to severe adduct formation and signal suppression. For routine operation in LC–MS, mobile phases containing nonvolatile mobile-phase additives, e.g., phosphate buffers, ethylenediaminetetraacetic acid, and ion-pairing agents like tetraalkylammonium ions or alkylsulfonates, are not recommended.

The problems related to the nature of the analytes appeared most difficult in the early years of LC–MS, but meanwhile a variety of so-called soft-ionization techniques have been developed, sometimes as a spin-off or in conjunction with LC–MS developments. Electrospray ionization (ESI) is by far the most successful of these techniques. They allow the mass spectrometric analysis of highly polar, ionic, and (thermo)labile compounds. As a result, the problem can be considered to be solved. In the majority of the liquid-based soft-ionization methods, the analytes, either as neutrals or as preformed ions in solutions, are transferred to the gas phase by nebulization of the liquid and subsequent desolvation of the droplets. In this way, the amount of heat required in the transfer has been minimized, which is favorable for the type of compounds to be analyzed.

LC–MS should be considered as a hybrid technique, in which the features of both LC and MS are optimally used to get the best possible result. Optimum separation conditions in terms of mobile-phase composition and stationary phase for a ultraviolet (UV) detector may widely differ from the separation conditions ideally suiting the mass spectrometer. Unlike in the early days of LC–MS, performing LC–MS in most laboratories is no longer the task of specialized mass spectrometrists. The technique is entering the chromatography laboratory as well. And in many instances separate LC–MS laboratories are set up, where specialists from the two disciplines work together to get the best results.

## LC–MS Interfaces

In the past 10 years, the manner in which LC–MS analysis is performed has significantly changed. While in the past it was necessary to choose the most appropriate LC–MS interface for a particular application from a list of five possibilities, e.g., the moving-belt interface, the direct-liquid introduction interface, the thermospray interface, the particle-beam interface, and the continuous-flow fast-atom bombardment interface, today all LC–MS technologies are based on API. The two most important



techniques are ESI and atmospheric-pressure chemical ionization (APCI); in addition, atmospheric-pressure photoionization is also used.

An API interface consists of five parts (Figure 1):

- the liquid introduction device, where nebulization of the LC column effluent is performed;
- the actual ion source region, where the ions are generated at atmospheric pressure;
- the ion-sampling aperture;
- the atmospheric-pressure to high-vacuum interface: the transition region; and
- the ion-optical system, where the ions generated in the source are analyte-enriched and transported toward the high-vacuum mass analyzer.

The operational principle of an electrospray API interface is as follows.

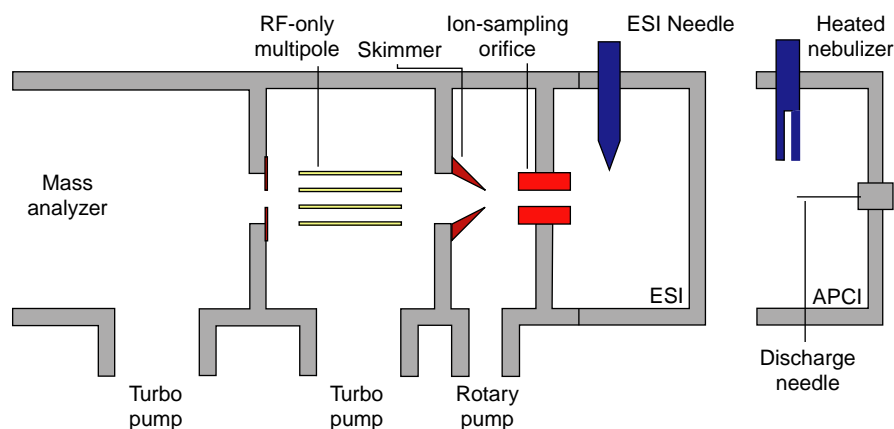
The column effluent from the LC is nebulized into the atmospheric-pressure ion source region. Typical solvents used are mixtures of water and methanol or acetonitrile, containing up to  $10 \text{ mmol l}^{-1}$  electrolyte, such as formic acid or ammonium acetate. Nebulization can be performed either by means of a strong electrical field, or by a combination of the strong electric field and pneumatic nebulization. The latter is sometimes named pneumatically assisted electrospray or ionspray<sup>®</sup>. The pure electrospray process is limited to flow rates up to  $10 \mu\text{l min}^{-1}$ . Pneumatically assisted electrospray enables the introduction of higher flow rates, up to  $1 \text{ ml min}^{-1}$ . Since ion production mechanisms between the two modes are identical, the term electrospray is used here throughout.

The electrospray nebulization is the result of a 2–5 kV electrical potential applied to the solvent emerging from a capillary. The field pulls out the liquid from the inlet capillary into a conical tip,

frequently called the Taylor cone. At the tip of the cone, one or more fine liquid threads appear, which subsequently disintegrate into small droplets. Uniform droplets in the  $1 \mu\text{m}$  diameter range are produced in a breakup process that results from autorepulsion of the electrostatically charged surface, which overcomes the cohesive forces of surface tension.

In the next step, ions are produced from the evaporating droplets. The processes involved are not completely understood. The general picture is as follows. As a result of the electrospray nebulization process, small charged droplets are generated, positively charged droplets when the source is operated in positive-ion mode and negatively charged droplets in the negative-ion mode. These droplets travel toward the counter-electrode. In the positive-ion mode, this results in transport of positive charges from the inlet capillary, also called electrospray needle, to the counter-electrode, which is the source block in most cases. At the same time, the negative ions are attracted toward the needle. At the needle, oxidation of these negative ions takes place, while reduction of the positive charges takes place at the counter-electrode. In negative-ion mode, this picture is reversed. The counter-electrode contains an  $\sim 200 \mu\text{m}$  ID hole, the so-called ion-sampling orifice, which samples part of the ions generated into the vacuum interface toward the mass analyzer.

In their flight between needle and counter-electrode, the droplets will start to evaporate, which means that neutral solvent molecules are evaporated from the surface of the charged droplets. The charge in the droplets will preferentially sit at the surface of the droplet, in order to achieve minimum interaction between the charges, in accordance with the laws of Coulomb. As a result of the solvent evaporation, the droplet size will decrease, which in turn leads to



**Figure 1** Schematic diagram of an atmospheric-pressure ionization interface.

shorter distances between the charges at the surface of the droplets. This process of solvent evaporation will continue as long as the surface tension of the liquid can accommodate the Coulomb repulsion between the charges. When this is no longer possible, a so-called Coulomb explosion will take place. The process, also called field-induced electrohydrodynamic droplet disintegration, results in the emission of a number of microdroplets with a radius of less than 10% of that of the initial droplets.

With these microdroplets, the solvent evaporation process will continue. Another Coulomb explosion may follow, leading to further reduction of the droplet size. The microdroplets play an important role in the analyte ionization. However, there are two theoretical models describing the actual ionization event, leading to gas-phase ions amenable to MS. In both models, the concept of preformed ions in solution plays an important role.

In the mid-1960s, Dole and co-workers were the first to combine electrospray nebulization with MS. Their hypothesis was that the nebulization of a dilute protein solution would lead to gas-phase protein ions as a result of continued solvent evaporation and Coulomb explosion. This would in the end lead to a very small droplet containing only one charged protein molecule. This small droplet can be considered as a solvated protein ion. Desolvation of this ion by solvent evaporation results in a gas-phase protein ion. This model is nowadays called the charge-residue model.

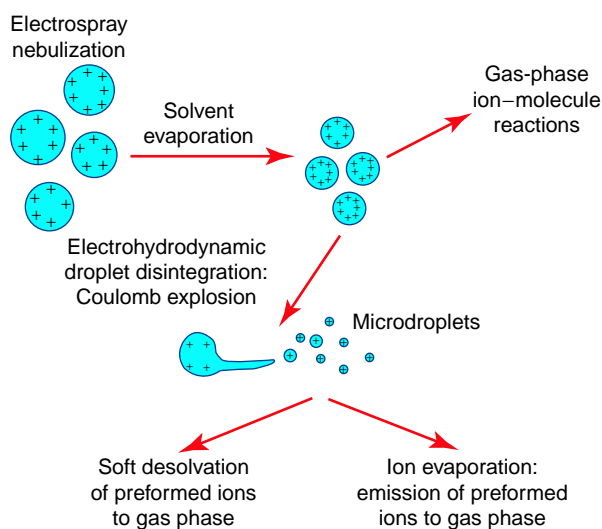
When, 15 years later, the Nobel-prize laureate Fenn and his co-workers continued the work of Dole with other analytes than proteins, they found results not in agreement with the charge-residue model. Therefore, they adopted another model, described in the mid-1970s by two Canadian scientists Iribarne and Thomson. In this so-called ion-evaporation model, it has been thermodynamically demonstrated that ions may escape from a sufficiently small droplet with a sufficient number of charges. This model also requires the analyte to be present as a preformed ion in solution.

Significant research has been done to prove the correctness of either of these models and/or to falsify one of them. From these studies, it appears that both models contribute to ion formation in ESI. The importance of either effect depends on the analyte. Therefore, ESI is best considered as a mixed-mode ionization. Some effects are more readily explained from the charge-residue model, some other from the ion-evaporation model. An important prerequisite for both models is that analytes should be present as preformed ions in solution. This indicates an analytical strategy to enhance the analyte ionization

in electrospray – adjust the solvent pH to stimulate the formation of either protonated or deprotonated molecules for organic bases and acids, respectively.

Unfortunately, this picture got corrupted when data were shown of good ESI performance for organic bases in positive-ion mode from basic solutions. This so-called wrong-way-around electrospray indicates that yet another mechanism will be operative. Nebulization of analyte solutions was initially adopted in LC–MS to achieve a gentle transfer of neutral molecules from the liquid phase to the gas phase by soft desolvation, which is a process similar to the processes described by the charge-residue model, but now for neutral species. Gas-phase ion–molecule reactions between these neutral analyte molecules and ion-evaporated buffer ions, for instance,  $\text{NH}_4^+$ , will also lead to protonated molecules. It appears that this gas-phase chemical ionization rather than the liquid-phase process is just another process involved in ESI. A summary of the ionization processes is given in **Figure 2**.

Whatever the exact mechanism is, the ionization leads to the formation of either positive or negative gas-phase ions, which move toward the ion-sampling orifice. In practice, we can only sample a very small portion of the ions produced. However, the ions are not expected to reach the ion-sampling orifice as a free gas-phase ion. In the humid ion source, the gas-phase ion is expected to be solvated. This solvated ion, which can be considered as a very tiny droplet, will pass through the ion-sampling orifice toward the skimmer and through the radio frequency (RF)-only multipole element. In this vacuum transition region, the solvated ion must be desolvated again, which is due to many ion–molecule collisions, with solvent and nitrogen molecules in the transition region. Ionic



**Figure 2** Summary of the electrospray ionization process.

species like  $[M + \text{MeCN} + \text{H}]^+$  at  $m/z$   $M + 42$ , which are sometimes observed, can be considered to be due to insufficient desolvation of the analyte ion.

Different instrument manufacturers have developed a variety of source designs. Important design characteristics in this respect are:

- the position of the spray probe relative to the ion sampling orifice, which is nowadays orthogonal in most cases;
- the manner in which heat is applied to assist in the evaporation of the solvent, e.g., by concurrent, orthogonal, or countercurrent-heated nitrogen gas;
- the type of ion-sampling orifice, e.g., an orifice in a flat plate, a cone, or a capillary, and the means to protect the orifice against clogging;
- the presence of ion-optical devices between ion-sampling orifice and skimmer;
- the position of the skimmer relative to the ion-sampling orifice; and
- the type of RF-only multipole device present in the second (and eventually third) vacuum chamber, e.g., quadrupole, hexapole, octapole, or ion tunnel.

Although such design characteristics can lead to significant different performance for a particular analyte, which in solving particular analytical problems may direct the selection of the instrument of that particular manufacturer, in general all systems currently commercially available show adequate performance in LC–MS. Discussing these differences in detail is beyond the scope of this article.

Next to ESI, APCI is an alternative, especially for less polar compounds. The APCI interface consists of a different solvent inlet probe, the so-called heated nebulizer, and a discharge electrode, but uses exactly the same atmospheric-pressure ion source and vacuum transition region as in ESI (see **Figure 1**). In a heated nebulizer, the LC column effluent is first pneumatically nebulized into an aerosol, which subsequently is evaporated in a heated vaporizer zone. Solvent and analyte vapor meet the discharge needle, where a potential of a few kilovolts initiates a continuous discharge that results in the ionization of solvent molecules, which in turn ionize the analyte molecules by gas-phase ion–molecule reactions (common chemical ionization processes).

## Qualitative and Quantitative Analysis

LC–MS can be used for both qualitative and quantitative analysis. A large number of examples

have been described in the literature. In this respect, it must be emphasized that the availability of LC–MS interfaces also enables the rapid introduction of samples into the mass spectrometer, e.g., by means of an autosampler, without online chromatographic separation. This approach has, for instance, been applied for the rapid screening of combinatorial libraries and rapid molecular-weight determination of synthetic products during multistep organic synthesis.

LC–MS has been used in qualitative analysis in a wide variety of fields. The most important application areas comprise:

- pesticides, herbicides, and insecticides;
- environmental contaminants such as sulfonated azo dyes, surfactants, antifouling agents, and drugs;
- support of drug development within pharmaceutical industry, i.e., in drug discovery, identification of metabolites, reaction by-products, and degradation products;
- steroids and steroid conjugates;
- natural products such as alkaloids, flavonoids, mycotoxins, and neurotoxins; and
- biochemically important compounds such as prostaglandins, peptides, proteins, (phospho)lipids, (oligo)saccharides, and (oligo)nucleotides.

Furthermore, LC–MS is used in a wide variety of other identification studies, e.g., in the identification of impurities and by-products in bulk chemicals used in industry, in the characterization of frequently applied fluorescent derivatives for LC in combination with fluorescence detection, for studying simulated biochemical processes via online electrochemical conversion units, and mass spectrometry via an LC–MS interface.

Identification is primarily based on molecular-mass determination, while for an actual structure elucidation LC–MS must be used in combination with tandem mass spectrometry (MS–MS). ESI and APCI are soft-ionization techniques, generating only intact molecule-derived ions, but no fragment ions for most molecules. Therefore, it is frequently applied in combination with MS–MS to achieve more structural information. With respect to qualitative analysis, the use of electrospray LC–MS–MS for peptide sequencing as part of proteomics research is currently an important area.

Although quantitative analysis can be based on the response for an intact molecule-derived ion, e.g., the protonated molecule, in most cases MS–MS is applied as well. In that case, the MS–MS is operated in the so-called selected-reaction monitoring mode

(SRM). In this mode, the first MS stage is used to select the ion of the intact molecule of interest from all other ions generated in the source, e.g., ions due to coeluting interferences from a complex matrix. The selected  $m/z$  is collided with neutral gas to induce fragmentation in a process called collision-induced dissociation (CID). While in product-ion scan mode the second MS stage is scanned to obtain the product-ion mass spectrum, in the SRM mode only one structure-specific  $m/z$  is transmitted to the detector. The SRM mode is the most selective generally applicable detection mode available – one will only get a signal when after sample pretreatment and LC within a particular retention time window, a molecule is present which is ionized to the  $m/z$  selected in the first MS stage, subsequently fragmented by CID to the  $m/z$  selected in the second MS stage. SRM is routinely applied in many studies for quantitative analysis. It is, for instance, the method of choice in analytical support of preclinical and clinical studies during drug development by pharmaceutical industries.

Similar to GC–MS, isotopically labeled internal standards can be used in quantitative analysis by LC–MS. Examples are given in the section below.

## Selected Applications

Environmental analysis is another important application area of LC–MS. The analysis is often a multiresidue strategy directed at many different compounds from various compound classes. A good example is the screening of surface and ground waters for environmental contamination prior to their use in the production of drinking water. According to European regulations, the determination of any individual pesticide at the level of  $0.1 \mu\text{g l}^{-1}$  is required. Both quantitation and identification is required at this level. Detection of pesticides at these levels can only be achieved by the use of preconcentrating sample pretreatment, e.g., offline or online liquid–liquid extraction or solid-phase extraction (SPE), in combination with selected-ion monitoring or SRM. Such a procedure is obviously not a general screening procedure. In addition, the various compound classes show quite different responses in ESI or APCI, e.g., some compounds should be analyzed in positive-ion mode, while others only provide sufficient response in negative-ion mode.

A typical approach in environmental analysis is an integrated system called SAMOS, which enables the automated, unattended analysis of filtered 100 ml surface water samples by means of online SPE and subsequent gradient-LC–UV detection. Any

compound-detected by a SAMOS system must be confirmed and/or identified by LC–MS. In this step, a more target-compound oriented approach can be applied. SAMOS-like systems can be coupled online to an LC–MS interface, as, for instance, demonstrated for a priority group of pesticides by the group of Barceló. An example of a chromatogram obtained with the application of this method to 200 ml of tap water spiked with 15 pesticides from various compound classes at  $0.04 \text{ ng ml}^{-1}$  level each is shown in Figure 3.

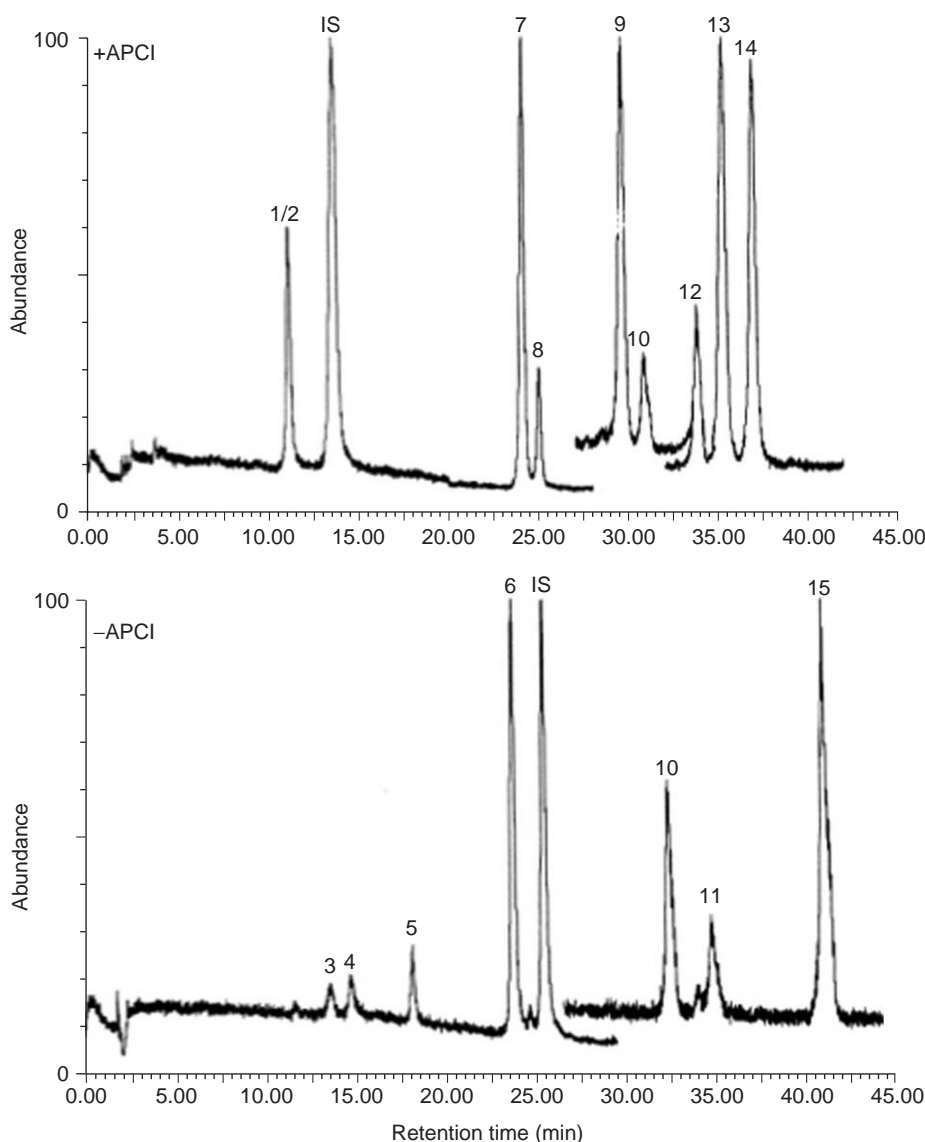
Due to its sensitivity, selectivity, and potential analysis speed, LC–MS is a tool of utmost importance in virtually every stage of drug development. The role of LC–MS already starts in the drug discovery stage. Three powerful tools have been developed for LC–MS to support in drug discovery.

The first and most generally applicable tool is Open Access, which transforms an LC–MS system into a walk-up ‘black-box’ for synthetic chemist in need of a rapid confirmation of the good progress of their synthesis by molecular-mass determination of their product.

To a growing extent, drug discovery is based on synthesis by combinatorial chemistry procedures. The combinatorial libraries need to be screened for biological activity, while a rapid characterization of identity is also required. LC–MS technology is frequently applied for the latter step. The analytical system consists of an  $x$ – $y$  autosampler, enabling sample introduction from a 96-well plate, connected to a fast LC–MS system. The software allows the rapid analysis of large series of samples at a high sample throughput, i.e., up to 60 samples per hour. The data-processing software provides a data browser, which allows rapid answers on whether or not the expected products are present. In some applications, the screening for biological activity and the mass spectral characterization is combined into one system by the application of integrated biodetection systems based on antigen–antibody or ligand–receptor interaction.

The third tool combines the rapid screening of combinatorial libraries or of a series of extracts from natural products to preparative-scale purification of biologically active compounds. The fractionation is controlled by the response of the compound of interest in the LC–MS.

While the structure confirmation in these approaches is primarily based on the molecular mass of the intact compound, more elaborate structure elucidation is required in subsequent stages of drug development, i.e., the stages related to impurity screening, identification of drug metabolites, and the search for degradation products in drug substances



**Figure 3** SPE–LC–APCI–MS chromatogram of 200 ml tap water spiked with  $0.04 \text{ ng ml}^{-1}$  of pesticides and  $0.05 \text{ ng ml}^{-1}$  IS (*tert*-butylazine in positive-ion mode and dinoterb in negative-ion mode). Peak identification: 1, bentazone; 2, Vamidothion; 3, 4-nitrophenol; 4, MCPA; 5, mecoprop; 6, dinoseb; 7, atrazine; 8, isoproturon; 9, ametryn; 10, malathion; 11, fenotrothion; 12, molinate; 13, prometryn; 14, terbutryn; and 15, parathion-ethyl. (Reprinted with permission from Aguilar C, Ferrer I, Borrull F, Marcé RM, and Barceló D (1998) Comparison of automated on-line solid-phase extraction followed by liquid chromatography–mass spectrometry with atmospheric-pressure chemical ionization and particle-beam mass spectrometry for the determination of a priority group of pesticides in environmental waters. *Journal of Chromatography A* 794: 147–163; © Elsevier.)

and drug formulations. A comprehensive and multi-instrument strategy for the screening and identification of drug metabolites has been outlined by Clark and co-workers. This approach is readily applicable to impurities and degradation products as well. In the first step, the MS and product-ion MS–MS spectrum of the parent drug is acquired and interpreted using a triple-quadrupole instrument. Discussion with toxicologists and molecular modeling results in a list of expected metabolites. This list

and the interpretation of the MS–MS spectrum enables the design of a number of precursor-ion scan and neutral-loss scan experiments to screen biological samples for metabolites using the triple-quadrupole instrument. In the third step, product-ion MS–MS spectra are acquired for the metabolites in the previous step. At this stage, interpretation of these data starts. Most likely, the interpretation will not lead to unambiguous identification of all metabolites found. Further studies can be performed



in the steps four and five – structure elucidation by multiple stages of MS–MS on an ion-trap instrument, and accurate mass determination of metabolites and product ions in the MS–MS spectra using a hybrid quadrupole–time-of-flight instrument. The final step involves a complete interpretation of the data, synthesis of the metabolites identified, checking of these by further MS–MS and NMR studies.

In addition to these qualitative studies, quantitative bioanalysis, e.g., in preclinical and clinical studies to provide pharmacokinetic and pharmacodynamic data, is an essential part of drug development. Quantitative bioanalysis is the most important application area of LC–MS, in terms of number of instruments applied and the number of analyses performed. Fast, high-throughput, and routine quantitative analysis by LC–MS also demands fast and automated sample pretreatment strategies and advanced data-processing software.

The keys to the success of LC–MS in quantitative bioanalysis are: (1) typical detection limits in the picogram and in favorable cases even subpicogram range, (2) excellent selectivity against possibly interfering compounds in the biological matrix by the use of the SRM mode, (3) enhanced confidence of identity of the compound(s) analyzed, and (4) the ability to use the ideal internal standards: isotopically labeled compounds. LC–MS–MS is often as easy to operate as LC–UV–PDA, but provides better selectivity. As a result, LC–MS–MS has become the method of choice in quantitative bioanalysis within pharmaceutical industries.

The higher selectivity achievable due to the use of SRM procedures is often immediately given away by decreasing the quality of the sample pretreatment and/or the chromatography. The rationale of this is an increase in the sample throughput. Advanced and selective sample pretreatment methods are often time consuming and thereby limit the sample throughput.

Serious matrix problems may be experienced in quantitative bioanalysis, especially in ESI. Signal suppression due to unknown matrix interferences is often observed. Changes in the sample pretreatment procedures may be successful in solving the problem, but in some cases changing over to APCI, when applicable, appears to be the only feasible solution.

ESI is frequently applied in the various stages of the characterization of peptides and proteins: molecular mass determination, amino acid sequencing, determination of nature and position of chemical and posttranslational modifications of proteins, investigation in protein tertiary and quaternary conformations, and the study of noncovalent associates. In most cases, no online separation is applied, but the sample solution is introduced directly via the ESI or

nano-ESI interface. Impressive results have been achieved in this area. Multidimensional LC in combination with MS–MS is rapidly becoming an important alternative to two-dimensional gel electrophoresis in proteomics research.

LC–MS is applied in many other fields as well, e.g., in the study of natural products, such as in the dereplication of flavonoids and several other compound classes in plant material and the identification and quantification of natural toxins, of endogenous compounds like acylcarnitines and arachidonic acid metabolites, and DNA adducts. LC–MS has become a routinely applicable technique. It is rapidly entering the chromatography laboratories to act as an LC detector in a variety of analyses. As such, LC–MS is appreciated for its sensitivity and selectivity, its specificity, and the information obtained, e.g., on the molecular mass of the analyte. The operation of an LC–MS system is no longer reserved to the MS specialist.

**See also:** **Mass Spectrometry:** Atmospheric Pressure Ionization Techniques; Electrospray.

## Further Reading

- Ardrey B (2003) *Liquid Chromatography–Mass Spectrometry: An Introduction*. Chichester: Wiley.
- Careri M, Bianchi F, and Corradini C (2002) Recent advances in the application of mass spectrometry in food-related analysis. *Journal of Chromatography A* 970: 3–64.
- Clarke NJ, Rindgen D, Korfmacher WA, and Cox KA (2001) Systematic LC–MS metabolite identification in drug discovery. *Analytical Chemistry* 73: 430A.
- Cole RB (ed.) (1997) *Electrospray Ionization Mass Spectrometry*. New York: Wiley.
- Matuszewski BK, Constanzer KL, and Chavez-Eng CM (1998) Matrix effect in quantitative LC–MS–MS analyses of biological fluids: A method for determination of finasteride in human plasma at picogram per milliliter concentrations. *Analytical Chemistry* 70: 882–889.
- Niessen WMA (1999) *Liquid Chromatography–Mass Spectrometry*, 2nd edn. New York: Dekker.
- Niessen WMA (2003) Progress in liquid chromatography–mass spectrometry and its impact on high-throughput screening. *Journal of Chromatography A* 1000: 413–436.
- Reemstma T (2003) Liquid chromatography–mass spectrometry and strategies for trace-level analysis of polar organic pollutants. *Journal of Chromatography A* 1000: 477–501.
- Romijn EP, Krijgsveld J, and Heck AJR (2003) Recent liquid chromatographic–(tandem) mass spectrometric applications in proteomics. *Journal of Chromatography A* 1000: 589–608.

Willoughby R, Sheehan E, and Mitrovich S (1998) *A Global View of LC–MS*. Pittsburgh, PA: Global View Publishing.

Wolfender J-L, Ndjoko K, and Hostettmann K (2003) Liquid chromatography with ultraviolet absorbance-

mass spectrometric detection and with nuclear magnetic resonance spectrometry: A powerful combination for the on-line structural investigation of plant metabolites. *Journal of Chromatography A* 1000: 437–455.

## Liquid Chromatography–Fourier Transform Infrared

**S J Kok**, TNO Nutrition and Food Research, Zeist, The Netherlands

**G W Somsen**, Utrecht University, Utrecht, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Fourier-transform infrared (FTIR) spectroscopy deals with the quantitative measurement of the absorption of IR radiation by a compound. FTIR exhibits molecular–vibrational transitions and provides characteristic information on molecular structure. Therefore, the combination of liquid chromatography (LC) and FTIR, in principle, can be highly useful when specific detection or identification of separated compounds is required. The high speed and multiplex nature of FTIR allows spectra to be recorded at any point in the chromatogram, and during or after the LC separation software can be used to calculate a total IR-absorbance based chromatogram (via Gram–Schmidt vector orthogonalization) or to reconstruct functional-group chromatograms at one or more specific wavelengths.

The combination of LC and FTIR has been a subject of research for more than 25 years now. In the development of LC–FTIR techniques, two basically different coupling methodologies can be discerned, namely one that involves flow cells and one that involves solvent-elimination. In the flow-cell approach, the eluent is led directly through a flow cell where IR spectra are recorded continuously offering fast and relatively easy detection of eluting analytes. The significant IR absorption of the eluent, however, may obscure large parts of the IR spectrum and dictates the use of small optical pathlengths. The solvent-elimination approach involves an evaporation interface for the removal of the interfering eluent and subsequent analyte deposition onto a suitable substrate, prior to FTIR detection of the analyte. In this case, detection is no longer affected by the IR characteristics of the mobile phase and

full spectra of relatively low amounts of compound can be obtained. The challenge of an effective solvent-elimination technique lies in the simultaneous eluent evaporation and analyte deposition, while maintaining the integrity of the obtained LC separation. This article provides an overview of the principles, practical aspects, and current status of LC–FTIR covering both flow-cell and solvent-elimination interfaces.

### Flow-Cell Interfaces

Flow cells offer a simple and straightforward means for the online coupling of LC and FTIR. The effluent of the LC is passed directly through a flow cell and IR spectra are acquired on-the-fly. The merits of the approach include low cost, instrumental simplicity, ease of operation, low maintenance, and the possible use of nonvolatile buffers. The analyte can be studied without any orientation or crystallization effects, oxidative degradation, or evaporation, which might occur upon collection and storage after solvent elimination. Furthermore, flow-cell detection takes place in real time making it potentially useful for online reaction monitoring and fast IR detection. On the other hand, the dynamic nature of on-the-run IR measurements leaves less time to collect spectra, limiting the signal-to-noise ratio (SNR) and sensitivity.

Despite simplicity of flow-cell LC–FTIR, its major drawback is the rather limited choice of eluents. For example, water strongly obscures most of the mid-IR region, prohibiting a practical combination of reversed-phase (RP)-LC and FTIR via a flow cell. Only some organic solvents (e.g., chloroform), which show sufficient transparency in (parts of) the IR spectrum, can actually be used. This essentially limits the application area of flow-cell LC–FTIR to normal phase (NP)-LC, size-exclusion chromatography (SEC), and critical chromatography (CC). Gradient elution cannot be applied, as accurate background subtraction with changing eluent composition is virtually impossible.

Willoughby R, Sheehan E, and Mitrovich S (1998) *A Global View of LC–MS*. Pittsburgh, PA: Global View Publishing.

Wolfender J-L, Ndjoko K, and Hostettmann K (2003) Liquid chromatography with ultraviolet absorbance-

mass spectrometric detection and with nuclear magnetic resonance spectrometry: A powerful combination for the on-line structural investigation of plant metabolites. *Journal of Chromatography A* 1000: 437–455.

## Liquid Chromatography–Fourier Transform Infrared

**S J Kok**, TNO Nutrition and Food Research, Zeist, The Netherlands

**G W Somsen**, Utrecht University, Utrecht, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Fourier-transform infrared (FTIR) spectroscopy deals with the quantitative measurement of the absorption of IR radiation by a compound. FTIR exhibits molecular–vibrational transitions and provides characteristic information on molecular structure. Therefore, the combination of liquid chromatography (LC) and FTIR, in principle, can be highly useful when specific detection or identification of separated compounds is required. The high speed and multiplex nature of FTIR allows spectra to be recorded at any point in the chromatogram, and during or after the LC separation software can be used to calculate a total IR-absorbance based chromatogram (via Gram–Schmidt vector orthogonalization) or to reconstruct functional-group chromatograms at one or more specific wavelengths.

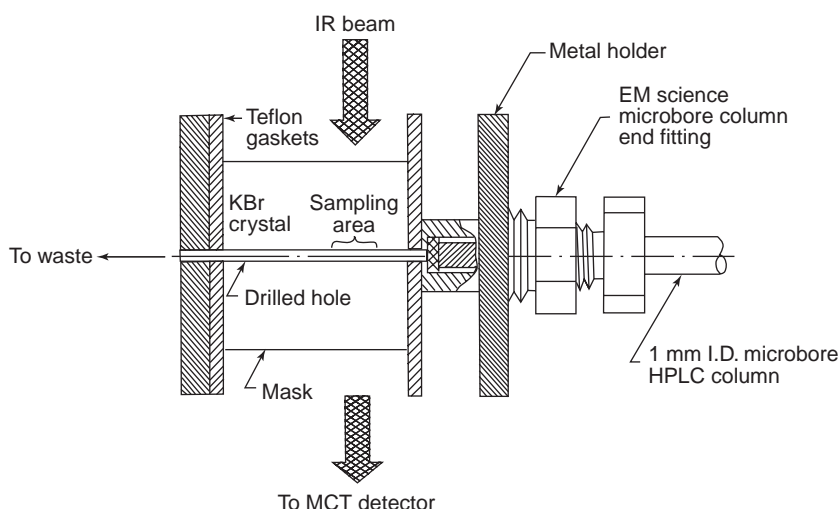
The combination of LC and FTIR has been a subject of research for more than 25 years now. In the development of LC–FTIR techniques, two basically different coupling methodologies can be discerned, namely one that involves flow cells and one that involves solvent-elimination. In the flow-cell approach, the eluent is led directly through a flow cell where IR spectra are recorded continuously offering fast and relatively easy detection of eluting analytes. The significant IR absorption of the eluent, however, may obscure large parts of the IR spectrum and dictates the use of small optical pathlengths. The solvent-elimination approach involves an evaporation interface for the removal of the interfering eluent and subsequent analyte deposition onto a suitable substrate, prior to FTIR detection of the analyte. In this case, detection is no longer affected by the IR characteristics of the mobile phase and

full spectra of relatively low amounts of compound can be obtained. The challenge of an effective solvent-elimination technique lies in the simultaneous eluent evaporation and analyte deposition, while maintaining the integrity of the obtained LC separation. This article provides an overview of the principles, practical aspects, and current status of LC–FTIR covering both flow-cell and solvent-elimination interfaces.

### Flow-Cell Interfaces

Flow cells offer a simple and straightforward means for the online coupling of LC and FTIR. The effluent of the LC is passed directly through a flow cell and IR spectra are acquired on-the-fly. The merits of the approach include low cost, instrumental simplicity, ease of operation, low maintenance, and the possible use of nonvolatile buffers. The analyte can be studied without any orientation or crystallization effects, oxidative degradation, or evaporation, which might occur upon collection and storage after solvent elimination. Furthermore, flow-cell detection takes place in real time making it potentially useful for online reaction monitoring and fast IR detection. On the other hand, the dynamic nature of on-the-run IR measurements leaves less time to collect spectra, limiting the signal-to-noise ratio (SNR) and sensitivity.

Despite simplicity of flow-cell LC–FTIR, its major drawback is the rather limited choice of eluents. For example, water strongly obscures most of the mid-IR region, prohibiting a practical combination of reversed-phase (RP)-LC and FTIR via a flow cell. Only some organic solvents (e.g., chloroform), which show sufficient transparency in (parts of) the IR spectrum, can actually be used. This essentially limits the application area of flow-cell LC–FTIR to normal phase (NP)-LC, size-exclusion chromatography (SEC), and critical chromatography (CC). Gradient elution cannot be applied, as accurate background subtraction with changing eluent composition is virtually impossible.



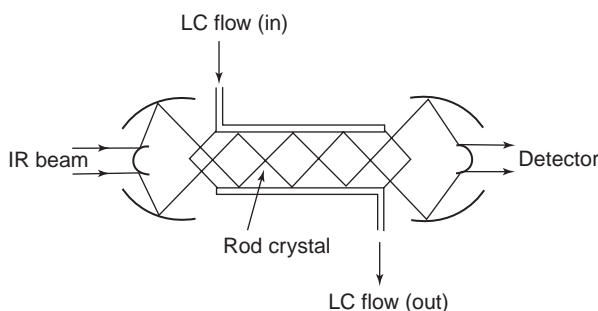
**Figure 1** Principle of a transmission zero-dead volume microbore LC-FTIR flow cell (cross-sectional view). (Reprinted with permission from Johnson CC and Taylor LT (1984), *Analytical Chemistry* 56(14): 2642–2647; © American Chemical Society.)

### Flow-Cell Types

Three types of flow cells can be discerned for online LC-FTIR coupling, and these are respectively based on transmission, attenuated-total-reflection (ATR), or specular reflection measurements. The spectral range (i.e., detection wave number range) of these interfaces is determined by the IR characteristics of the applied cell-window material and by the mobile phase used for the chromatographic separation.

The most frequently used type of flow cell is the transmission cell, which can either consist of an IR-transparent cavity or two IR-transparent windows separated by a metal or Teflon spacer. The LC eluent enters and exits the cell via capillary tubing and is sampled by the IR beam passing perpendicularly. Depending on the application, the optical pathlength can be adjusted, ranging from 0.001 to 2 mm. Transmission flow cells are available from several manufacturers and can include high-temperature options. For use in microbore LC, specially designed zero-dead volume flow-cells with an internal volume of 0.33  $\mu\text{L}$  have been developed (Figure 1). The eluent is led through a sample cavity consisting of a 0.75 mm hole drilled in a block of potassium bromide or calcium fluoride. The IR beam interrogates the eluent stream under  $90^\circ$  yielding detection limits in the range of 40–50  $\mu\text{g}$  when chloroform is used as mobile phase.

The second category of flow cells is based on the ATR principle. The cell consists of a cylindrically shaped ATR crystal with cone-shaped ends (Figure 2). The crystal is incorporated in a flow cell with the cone ends outside the cell body and the eluent can enter the flow-cell cavity surrounding the crystal. Cassegrain optics are used to focus the IR beam on



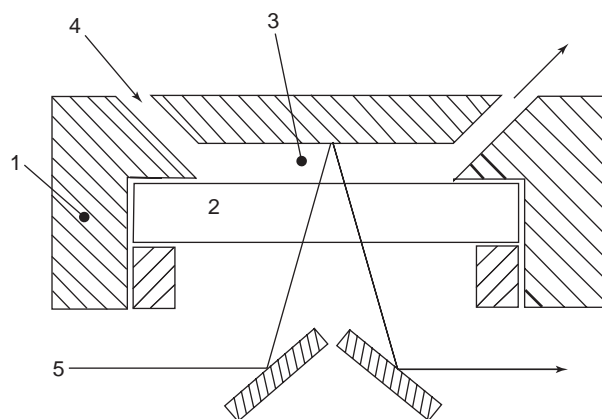
**Figure 2** Principle of an ATR flow cell.

the crystal at one end, and to collect the IR radiance emerging from the other end and direct it toward the detector. To maintain an effective optical pathlength, the number of reflections in the optical element is typically 10 or 11 and the internal volume of the flow cell is 1–25  $\mu\text{L}$ . Spectra collected from ATR flow cells may exhibit typical bandshape distortion due to the refractive-index changes around absorption bands, complicating spectral interpretation. In addition, the wavelength-dependent penetration depth of IR radiation complicates quantitation. However, ATR techniques can be very useful when spectral information from aqueous eluents has to be obtained, as the optical pathlength (i.e., penetration depth) is in the low micrometer range, thereby limiting the absorption by the eluent.

The third type of flow cell is based on specular reflection measurements and consists of a trough-shaped stainless-steel cell body, covered with an IR-transparent window (Figure 3). An external mirror is used to direct the IR beam toward the flow cell under near-normal incidence angles, reducing the reflection losses at the air-window interface. The IR beam is

reflected via a mirror surface inside the cell cavity and directed toward the detector via a second external mirror. The actual optical pathlength actually is twice the thickness of the sample cavity and can be adjusted from 50  $\mu\text{m}$  to 2 mm (cell volumes, 1–40  $\mu\text{l}$ ). AABSpec (Waterford, Ireland) supplies this type of cells.

The commonly used detector-cell volumes, though much smaller than required to minimize band broadening in conventional LC, render flow-cell FTIR detection compatible with miniaturized LC systems. In microbore-LC chromatographic peak volumes are several orders of magnitude smaller and peak concentrations are higher than in conventional LC. However, the sample capacity (both in mass and volume) of LC columns decreases in proportion to their cross-sectional area and the advantage of microbore-LC therefore is only partly fulfilled.



**Figure 3** Principle of a reflection flow cell. 1, cell body; 2, IR-transparent window; 3, flow-cell cavity; 4, LC-flow path; 5, IR-beam path.

### Cell Window Materials

Cell windows or crystals are available from many materials and the choice depends on the application (Table 1). The materials must be chemically resistant to the eluent used in the chromatographic method, withstand high pressures, and offer sufficient transmittance to maintain a reasonable IR-energy throughput. Calcium fluoride ( $\text{CaF}_2$ ), zinc selenide ( $\text{ZnSe}$ ), and, to a lesser extent, germanium ( $\text{Ge}$ ), are frequently applied, but rather expensive flow-cell materials. Potassium bromide ( $\text{KBr}$ ) and sodium chloride ( $\text{NaCl}$ ) are cheap alternatives and offer complete transparency in the mid-IR range. Besides, their low refractive indices minimize the risk of spectral fringes at certain optical pathlengths. However, these materials cannot resist excessive pressures and their strongly hygroscopic properties limit their use to nonaqueous eluents. High refractive index materials (like  $\text{ZnSe}$ ) are required in ATR flow cells in order to maintain total reflection at the crystal boundaries and allow sufficient reflections.

### Eluent Absorption

Ideally, the mobile phase used in flow-cell LC–FTIR should not exhibit serious background absorption that may obscure analyte absorption bands. Unfortunately, most organic solvents used in LC – highly chlorinated and fluorinated solvents being the exception – show intense IR spectra. Furthermore, in frequent cases the choice of eluent is largely determined by the required chromatographic properties of the involved solvent(s). As a consequence, the obtainable qualitative and molecular information often is limited and determined by the spectral window(s) provided by the eluent. The magnitude of solvent

**Table 1** Optical and physical properties of window materials for use in IR flow cells and as deposition substrate in solvent-elimination interfaces

Material	Transmission range ( $\text{cm}^{-1}$ )	% Transmittance (thickness)	Refractive index (at $1000\text{ cm}^{-1}$ )	Hardness ( $\text{kg mm}^{-2}$ )	Sensitive to	Solubility in water
Calcium fluoride ( $\text{CaF}_2$ )	50 000–1111	90.0 (4 mm)	1.39	158	Ammonium salts, acids	Slightly soluble ( $0.013\text{ g l}^{-1}$ )
Germanium ( $\text{Ge}$ )	5500–475	50 (2 mm)	4.0	550	Sulfuric acid, aqueous reagents	Insoluble
Potassium bromide ( $\text{KBr}$ )	40 000–400	90.5 (4 mm)	1.52	7	Lower alcohols, water	Highly soluble
Sodium chloride ( $\text{NaCl}$ )	40 000–625	91.5 (4 mm)	1.49	15	Lower alcohols, water	Highly soluble
Zinc selenide ( $\text{ZnSe}$ )	20 000–454	65 (1 mm)	2.4	137	Acids, strong alkalis	Insoluble



absorption can be decreased by adjusting the optical pathlength of the cell, although this obviously will affect the analyte absorbance too. The optimum pathlength also depends on the analytical query at hand. For example, when specific, accurate, and sensitive detection of an analyte is required at a particular wave number where the solvent shows absorption, an optical pathlength resulting in an eluent absorption of  $\sim 0.4$  AU (i.e., a transmission of  $e^{-1}$ ) would be selected, because then an optimum SNR is obtained. On the other hand, when the primary goal of the experiment is the characterization or identification of the analyte(s), an optical pathlength is chosen in such a way that the eluent absorptions are minimized throughout the spectrum in order to ensure all characteristic absorption bands can be detected for reliable structure elucidation. Clearly, there is always a trade-off between structural information and high sensitivity, and there is no single pathlength suitable for all eluents used in LC. For organic solvents typical optical pathlengths are 100–2000  $\mu\text{m}$  while much shorter optical pathlengths (10–50  $\mu\text{m}$ ) have to be used for water.

In order to correct for background absorption by the eluent, background subtraction often can be carried out quite reliably. Provided that isocratic LC is used, the analytes are present in a constant matrix (eluent) and IR bandshifts due to changes in environment are absent. In addition, FTIR allows the acquisition of spectral data on a precise wave number scale. However, one must be aware of 'ghost bands' or spikes in the region where the eluent is completely opaque leading to the faulty interpretation that there is analyte absorption.

To circumvent problems associated with excessive eluent absorptions that prohibit FTIR transmission detection, some remedies are possible. ATR flow cells can be used to inherently reduce the optical pathlength. Another option is postcolumn extraction of the LC effluent in order to transfer the analytes to a more IR-transparent solvent. Also, deuterated solvents could be used in which eluent absorption bands are shifted to lower wave numbers and potential analyte absorption bands are revealed.

A more recent option to cope with eluent absorption is the increase of the IR source intensity using quantum cascade lasers operating in the mid-IR region. Though molecular structure information cannot be obtained when using a monochromatic source, quantitative measurements of specific functional groups can be achieved. The powerful emission of the IR laser allows larger optical pathlengths to be used in combination with aqueous eluents and improves the SNR with a factor of 50.

**Applications** Notwithstanding the limitations, there are a limited number of specific applications in which flow-cell LC-FTIR can be quite useful to obtain specific quantitative and structural information in a convenient manner. The application area of flow-cell FTIR is limited to samples with relatively high analyte concentrations, as is the case in, for instance, the analysis of carbohydrates, alcohols, and organic acids in wines and sugars in soft drinks. SEC, as used for the separation of synthetic polymers, is also well suited to be coupled with FTIR by flow cells. Polymer samples are often available in large quantities and low detection limits are usually not required. In addition, the separation process in SEC is essentially independent of the choice of eluent, provided the sample is fully soluble and no analyte-stationary phase interactions take place. Consequently, IR-favorable eluents can be selected. Therefore, SEC-flow-cell-FTIR is a valuable tool for the rapid, selective, and quantitative analysis of the chemical composition of polymers as a function of their hydrodynamic volume.

## Solvent-Elimination Interfaces

The strong IR absorption of most eluents reduces the attainable detection limits in flow-cell FTIR and has directed LC-FTIR research toward a solvent-elimination approach in which the eluent is removed prior to detection. To accomplish this, generally the eluent is directed to a nebulizer, often aided with (heated) nebulizer gas, to achieve solvent evaporation. Almost simultaneously, the separated analytes are deposited (immobilized) on a moving substrate to collect the analytes individually and retain the chromatographic integrity. After deposition, IR spectra from the immobilized chromatogram are acquired. Dependent on the type of substrate used (see below) and/or size of the deposited spots, for detection often special optics, such as a (diffuse) reflection unit, a beam condenser, or an IR microscope, are used.

Solvent-elimination LC-FTIR offers a number of distinct advantages when compared with flow-cell LC-FTIR approaches. First of all, the absence of interfering eluent absorption bands permits spectral interpretation over the full wave number region, thereby fully exploiting the identification possibilities of IR spectroscopy. Secondly, the chromatogram is still available after the chromatographic run has completed, and accordingly, the sensitivity (SNR) and/or spectral resolution can be greatly enhanced by employing increased scanning times. Besides, the sensitivity can be further increased by producing concentrated analyte deposits in combination with appropriate IR optics. These aspects make

solvent-elimination the LC–FTIR methodology of choice when structural information of relatively small amounts of analyte has to be obtained. Finally, solvent-elimination interfaces may offer the possibility of gradient LC by varying the nebulizer temperature during the chromatographic run to obtain a constant deposit quality.

### Deposition Substrates and Spectral Quality

Deposition of analytes in solvent-elimination LC–FTIR is performed on powdered substrates, mirrors, or IR-transparent windows. Correspondingly, diffuse reflection Fourier-transform infrared (DRIFT) detection, reflection–absorption (R–A) spectroscopy, or transmission measurements are applied to investigate the analyte deposits.

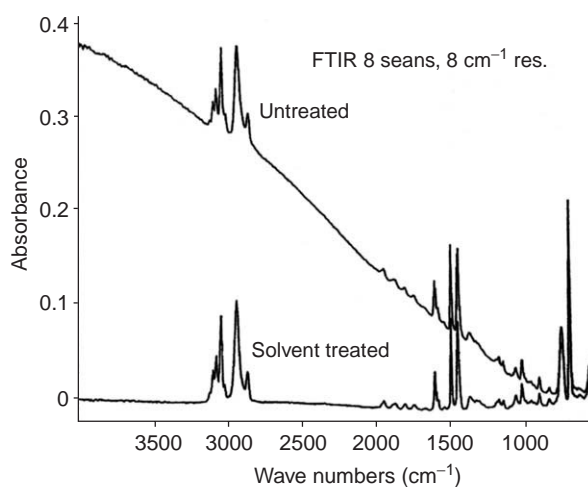
In early solvent-elimination interfaces, powdered potassium chloride (KCl) was used as substrate. DRIFT, being one of the most sensitive IR modes, was subsequently used for detection and sub-microgram detection limits could be achieved. However, when the eluent is not completely evaporated during analyte deposition, the remaining eluent can draw analyte into the underlying powder layers, which cannot be penetrated by the interrogating IR beam thereby frustrating the detection. Moreover, the homogeneity of the powder, the nature, and load of sample, and the reorientation of the powder surface during deposition may strongly affect the quality and reproducibility of the IR spectra acquired. Furthermore, KCl cannot be used with aqueous eluents, and thus not with RP-LC. As an alternative diamond powder can be used, but this is very expensive and difficult to recycle.

Water resistant, front-surface aluminum mirrors can be used as deposition substrates, followed by spectral acquisition in the R–A mode. The smooth and hard surface properties of mirrors complicate efficient analyte deposition (spreading of the analytes) when the eluent is not completely evaporated. The spectral data recorded from these substrates should be in close resemblance to the spectra obtained from transmission measurements, because the band intensities are controlled by a double-pass transmittance mechanism. However, still spectral differences between R–A and KBr disk spectra can be observed, including absorption-band shifts and asymmetries. Furthermore, the effect of light scattering (Christiansen effect) may become apparent when the spot thickness exceeds a certain level and anomalous relative band intensities may be observed in R–A spectra of certain analytes deposited on flat substrates when compared to transmission spectra acquired from KBr disks. In order to minimize these

effects, a rear-surface aluminum-coated IR-transparent germanium disc can be used as deposition substrate, although the adverse spectral effects are never completely eliminated. A postdeposition annealing procedure with dichloromethane has been proposed in order to minimize the effect of light scattering and to produce homogeneous deposits (Figure 4).

Most favorable spectral results in solvent-elimination LC–IR are obtained when analytes are deposited on flat IR-transparent substrates (ZnSe,  $\text{CaF}_2$ , KBr) and measured in the transmission mode. Preferably, ZnSe is used as deposition substrate as this material is inert and insoluble in water (favoring application in RP-LC) and offers a wide transmission range (Table 1). The transmission spectra acquired from analytes deposited on ZnSe are of good quality, free from spectral distortions, show a better SNR than R–A spectra of the same amount of material, and closely resemble KBr-disc transmission spectra allowing reliable spectral interpretation and automated library searches.  $\text{CaF}_2$  can be used as a cheap alternative when no spectral information has to be obtained in the low wave number region.

The quality and appearance of spectra are influenced by the morphology and thickness of the deposited analytes. The morphology will depend primarily on parameters such as eluent composition, evaporation rate, temperature, and nature of the substrate and the analytes. During solvent evaporation some compounds will form nice crystals while



**Figure 4** Overlay of IR spectra obtained from polystyrene deposit on Ge–Al substrate before (top) and after (bottom) dichloromethane vapor annealing demonstrating the effect on IR scattering. (Reprinted with permission from Johnson Schunk TC, Balke ST, and Cheung P (1994) *Journal of Chromatography A* 661: 227–238; © Elsevier.)

others will deposit as an amorphous layer. Over time, the morphology can change to its energetically most favorable state. Analyte morphology must be taken into consideration, as library entries may contain analyte spectra only with a particular morphology. Analyte deposits may be well defined to form a smooth film, whereas other analytes may form discontinuous spots showing many small irregular domains. Due to this spatial inhomogeneity, it can be difficult to obtain a continuous chromatogram with an IR microscope, as empty substrate areas may be sampled by the narrow IR beam. Alternatively, scanning over a larger substrate area using less beam-condensing optics can average out the spatial inhomogeneity, although the sensitivity may be compromised. Effective and distinct deposition of low-viscosity and liquid-like compounds may be a problem when hard and smooth substrates are used. Spreading and remixing of these analytes can be accomplished by deposition/trapping in the pores of either low-density polyethylene or polytetrafluoroethylene membranes as supplied by Lab Connections (Northborough, MA, USA).

### Types of Solvent-Elimination Interface

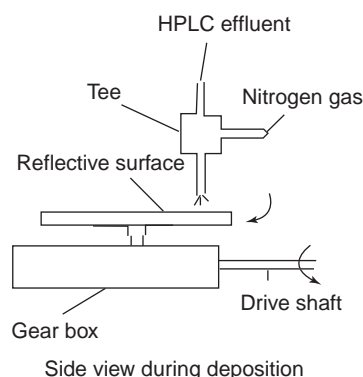
Considering the variety of solvent-elimination interfaces developed over the past decades, substantial efforts have been made to accomplish successful evaporation of the LC eluent and subsequent deposition of analytes. Early interfaces employed KCl powder as substrate for DRIFT detection. The LC eluent was dropped via a heated tube into discrete KCl-filled cups and residual solvent was removed by aspiration before acquisition of spectra. These early LC-DRIFT systems have demonstrated for the first time that solvent-elimination LC-FTIR is more sensitive and produces spectra of better quality than flow-cell-based LC-FTIR. However, the drawbacks of DRIFT discussed previously redirected the focus on the use of flat deposition substrates. In the first designs of this kind, small flows ( $<5 \mu\text{l min}^{-1}$ ) of organic solvents from micro-LC columns were led directly to a moving KBr plate that was placed under a stream of warm nitrogen to enhance solvent evaporation. After deposition, analyte spectra were acquired by IR transmission detection. In this setup, the complete chromatogram was stored as a continuous trace, but the analyte spots obtained on the KBr plate were quite large leading to a rather poor IR sensitivity.

In order to permit higher (aqueous) flow rates ( $>5 \mu\text{l min}^{-1}$ ) in LC-FTIR, use of interfaces with an enhanced evaporation capacity are essential. As effective solvent elimination is an important issue as

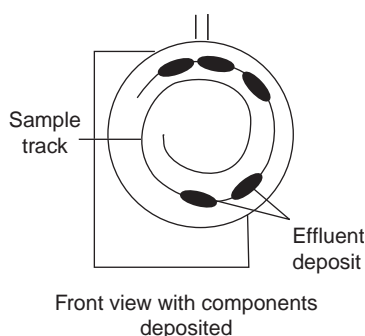
well when LC is combined with mass spectrometry (MS), several LC-MS interface types have also been utilized for LC-FTIR. For instance, the thermospray interface has been used to evaporate aqueous eluents at  $0.5 \text{ ml min}^{-1}$  and simultaneously deposit separated analytes on a metal IR-reflective ribbon that is continuously moved through an FTIR spectrometer equipped with a reflection accessory for spectra acquisition. Detection limits were in the low-microgram range. The particle beam interface originally developed for LC-MS was successfully used for the deposition of LC-separated compounds on KBr substrates. Aqueous eluents could be effectively evaporated at flow rates of up to  $0.3 \text{ ml min}^{-1}$ , and microgram-amounts of analytes (including thermolabile compounds) could be analyzed using FTIR transmission detection. Micro-LC-FTIR via an electrospray interface is also possible: flow rates of up to  $20 \mu\text{l min}^{-1}$  can be eliminated while depositing analytes on a ZnSe plate. Although the potential usefulness of LC-MS interfaces for solvent-elimination LC-FTIR has been demonstrated, the developed systems have never really matured and essentially were used by their designers only.

Most successful solvent-elimination LC-FTIR is achieved by employing pneumatic nebulization (Figure 5). These nebulizers use a high-speed gas flow to break up the eluent into small, fast-moving droplets, thereby greatly enhancing the solvent evaporation. In this way, organic eluents can be readily eliminated, while the nebulizer gas is heated when (almost) complete removal of aqueous eluents is required. Following eluent evaporation, the analytes are deposited on a step-wise or continuously moving IR-transparent substrate. Depending on the focusing capacity of the nebulizer, deposition trace widths of 200–500  $\mu\text{m}$  are achieved, resulting in IR detection limits in the sub-microgram range. Several LC-FTIR designs based on pneumatic nebulization are commercially available from Lab Connections. A rotating germanium disc with a rear-surface reflective coating is utilized as deposition substrate in combination with a reflection accessory for IR measurements.

With pneumatic nebulization for LC-FTIR, optimum mass sensitivity is achieved when microbore-LC (typical flow rate,  $20\text{--}50 \mu\text{l min}^{-1}$ ) is used in combination with a ZnSe deposition substrate and IR microscopy for detection. With such systems it is possible to acquire full spectra from 1 to 10 ng of analyte. In order to achieve complete evaporation of 100% aqueous eluents, enhanced solvent elimination power is required. One solution to this problem is the placement of the nebulizer inside a vacuum chamber to facilitate the evaporation of water. Another option is the online liquid-liquid extraction of the LC eluent

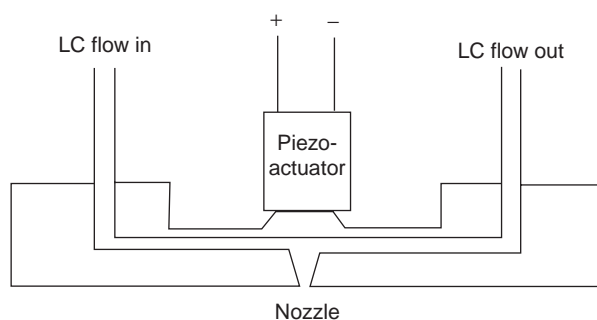


(A)



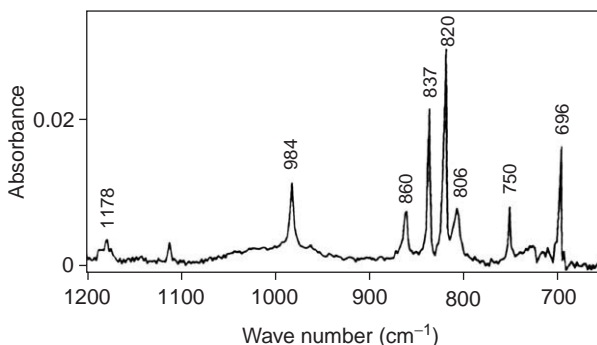
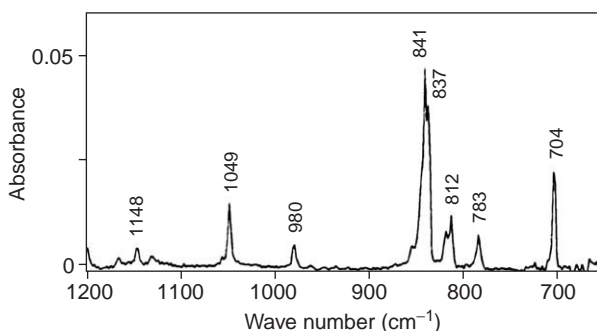
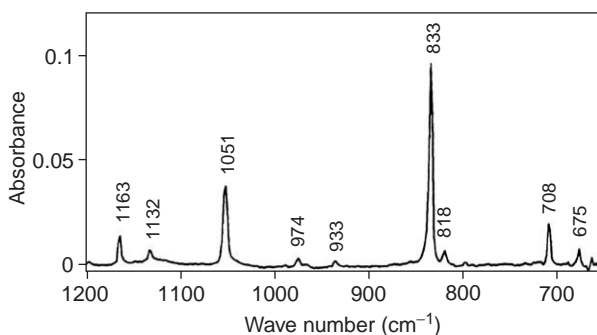
(B)

**Figure 5** Schematic representation of a solvent-elimination interface. (A) Side-view during solvent-elimination and analyte deposition; (B) Top-view from analyte deposits on substrate. (Reprinted with permission from Gagel JJ and Biemann K (1986) *Analytical Chemistry* 58(11): 2184–2189; 86 © American Chemical Society.



**Figure 6** Schematic representation of a piezoactuated flow-through microdispenser interface.

with a volatile organic solvent that, after phase separation, is being directed to the pneumatic interface. An additional advantage of this approach is that nonvolatile buffers can be used in the LC eluent since they are not extracted. Further reduction of the LC flow rate to  $1\text{--}2\ \mu\text{L min}^{-1}$  while adding a make up flow of  $20\ \mu\text{L min}^{-1}$  of methanol is also a way to handle highly aqueous eluents. Then, evaporation



**Figure 7** FTIR spectra of isomeric chloropyrenes recorded during solvent-elimination LC–FTIR of a chlorinated pyrene sample. Based on the spectral data (from the top downward) the isomers could be identified as 1,6-dichloropyrene, 1,8-dichloropyrene, and 1,3-dichloropyrene, respectively.

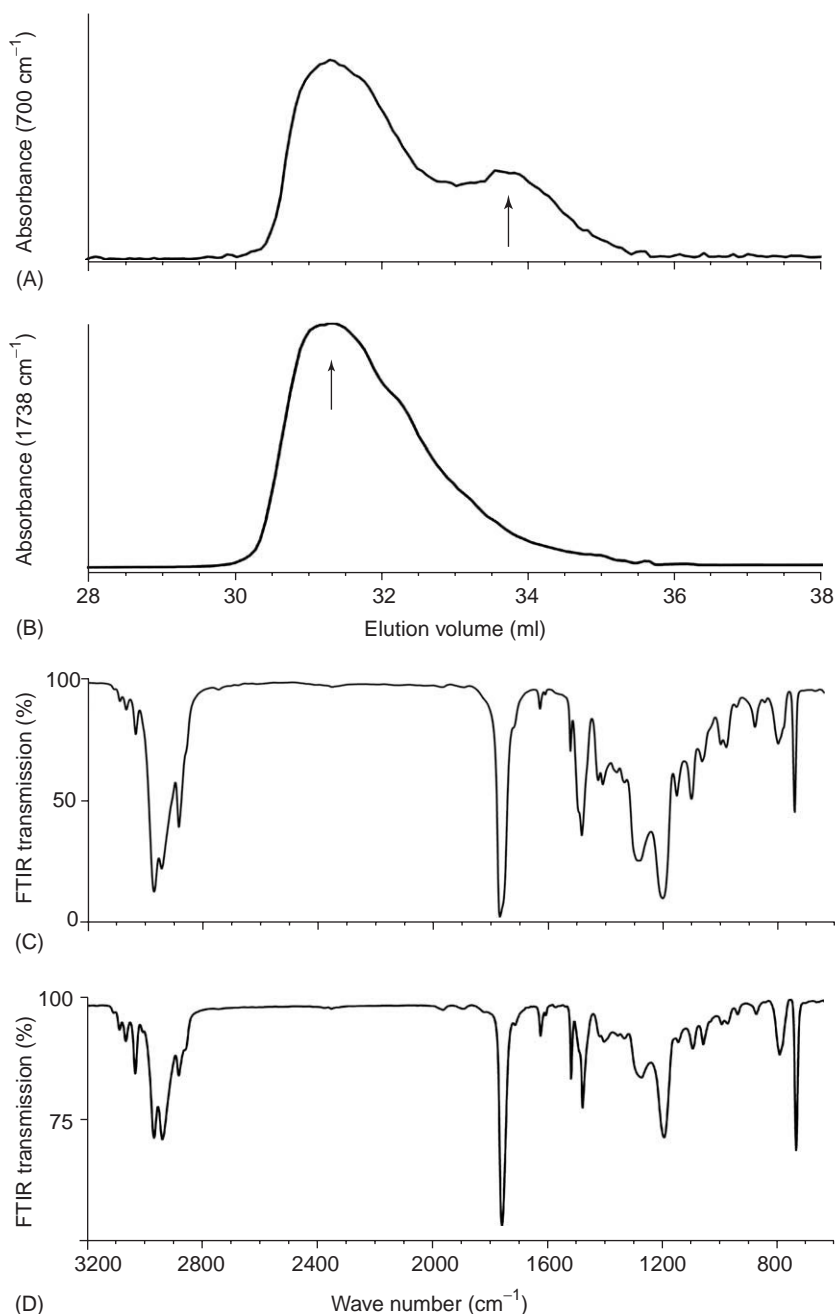
conditions are essentially independent of the water content and even gradient elution can be used.

Next to pneumatic nebulization, ultrasonic nebulization can be applied for solvent-elimination LC–FTIR. The eluent spray is now formed by disrupting the liquid surface at ultrasonic frequencies. Carrier gas can be used to enhance eluent evaporation and to focus the spray toward the deposition substrate. A further increase in the evaporation capacity is accomplished by placing the ultrasonic nebulizer and substrate in a vacuum chamber. Such a system is suitable for the evaporation of high-boiling eluents at relatively high flow rates ( $100\text{--}200\ \mu\text{L min}^{-1}$ ). Various manufacturers have commercialized ultrasonic nebulizers for LC–FTIR: Lab Connections offers the LC Transform 300 Series

and The InfraRed Chromatograph is manufactured by Bourne Scientific (Acton, MA, USA). The latter utilizes an additional desolvation chamber for elimination of the LC eluent.

As mentioned earlier, the most favorable sensitivity in LC-FTIR is achieved when analytes are deposited on the IR substrate in compact spots, because then the advantages of IR microscopy can be fully exploited. Effectively evaporating a stream of eluent

and, at the same time, depositing analytes in a narrow trace is not an easy task, but developments in this direction are ongoing. This is illustrated, for example, by the use of a state-of-the-art piezoactuated flow-through microdispenser in the analysis of glucose and fructose by LC-FTIR. The interface is based on the principle used for inkjet printing and its design has been adapted to operate in the flow-through mode for the use in LC (Figure 6). The



**Figure 8** Solvent-elimination SEC-IR from poly(styrene-butylacrylate) revealing changes in chemical composition as function of hydrodynamic volume. (A) Functional-group chromatogram for styrene; (B) functional-group chromatogram for butylacrylate. FTIR spectra taken from (C) the peak top and (D) the peak tail at positions indicated by arrows in A and B.



droplets produced by the interface are 53 pl in volume and are readily evaporated at room temperature and atmospheric pressure without additional heating or nebulizer gas, offering mild deposition conditions. As a result the deposits are concentrated in 40–80  $\mu\text{m}$  narrow spots on a deposition substrate of calcium fluoride, which perfectly meets the dimensions for optimum detection by IR microscopy.

### Application

The LC–FTIR detection limits obtained with pneumatic and ultrasonic nebulizers are sufficient for a number of practical applications. For real trace analysis, a sample enrichment procedure such as solid-phase extraction prior to LC–FTIR will be necessary to allow analyte detection. Nevertheless, the usefulness of solvent-elimination LC–FTIR has been successfully demonstrated in a variety of analytical queries where structural information and/or identification of (unknown) compounds are required. The wide range of compounds analyzed by LC–FTIR employing eluent removal comprises environmental pollutants (polycyclic aromatic hydrocarbons, pesticides, and herbicides), pharmaceuticals (e.g., steroids and analgesics) and their impurities, drug metabolites, polymer additives, dyes, and fullerenes. FTIR detection can be especially useful when isomeric compounds have to be distinguished (Figure 7). Even the secondary structure of proteins, such as  $\beta$ -lactoglobulin and lysozyme, has been studied by solvent-elimination LC–FTIR. This technique can be particularly beneficial in synthetic-polymer analysis in order to reveal the chemical composition of (co-)polymers (Figure 8). A special application area is high-temperature SEC–FTIR for the analysis of polyolefins (e.g., composition determination of ethylene–propylene rubbers) with the high-boiling trichlorobenzene as eluent.

The choice for the type of LC–FTIR coupling (flow cell or solvent-elimination) depends on the particular application of the user, where aspects such as type of spectral information needed, required sensitivity, and ease of use are main criteria. Flow-cell LC–FTIR is relatively simple and straightforward, and has developed into a special-purpose technique that can be used in a routine fashion for the monitoring of major mixture constituents with specific functional groups. Solvent-elimination LC–FTIR is somewhat more complicated requiring (sometimes complex) evaporation interfaces, but allows characterization of minor sample components with a high level of confidence. At present, the practical application of FTIR detection in LC is still quite limited.

Nevertheless, developments over the last years have led to the situation that almost every type of LC has been and can be effectively coupled to FTIR. Furthermore, there always will be particular applications (e.g., discrimination of isomers or polymer analysis) where IR data on separated compounds can be highly valuable. Furthermore, for the solution of complex analytical problems, the possible integration of the information on molecular structure provided by FTIR, MS, and/or nuclear magnetic resonance (NMR) would be highly advantageous. Illustrative for this statement is the recent development of hyphenated systems employing multiple interfacing of the same LC system to several spectrometric detectors (UV absorbance, MS, NMR, and FTIR). It is the complementary nature of the data provided by each spectrometric technique that leads to an enormous information content of the total system.

**See also:** **Fourier Transform Techniques. Infrared Spectroscopy:** Overview. **Liquid Chromatography:** Overview.

### Further Reading

- Fraser DJJ, Norton KL, and Griffiths PR (1988) HPLC/FT-IR measurements by transmission, reflection-absorption, and diffuse reflection microscopy. In: Messerschmidt RG and Harthcock MA (eds.) *Infrared Microspectroscopy: Theory and Applications*, pp. 197–210. New York: Dekker.
- Griffiths PR and Conroy CM (1986) Solvent elimination techniques for HPLC/FT-IR. *Advances in Chromatography* 25: 105–138.
- Griffiths PR and De Haseth JA (1986) The HPLC/FTIR interface. In: *Fourier Transform Infrared Spectrometry*, pp. 611–647. New York: Wiley.
- Kalasinsky KS and Kalasinsky VF (2002) High-performance liquid chromatography/Fourier transform infrared spectroscopy. In: *Handbook of Vibrational Spectroscopy*, vol. 2, pp. 1641–1660. Chichester: Wiley.
- Kalasinsky VF and Kalasinsky KS (1992) HPLC detection using Fourier transform infrared spectrometry. In: Patonay G (ed.) *HPLC Detection: Newer Methods*, pp. 127–161. New York: VCH.
- Somsen GW, Gooijer C, Velthorst NH, and Brinkman UAT (1998) Coupling of column liquid chromatography and Fourier transform infrared spectrometry. *Journal of Chromatography A* 811: 1–34.
- Vidrine DW (1979) Liquid chromatography detection using FT-IR. In: Ferraro JR and Basile LJ (eds.) *Fourier Transform Infrared Spectroscopy: Applications to Chemical Systems*, pp. 129–164. New York: Academic Press.
- White R (1990) Liquid chromatography/Fourier transform infrared spectroscopy. In: *Chromatography/Fourier Transform Infrared Spectroscopy and its Applications*, pp. 95–136. New York: Dekker.

## Liquid Chromatography–Nuclear Magnetic Resonance Spectrometry

J-L Wolfender, K Ndjoko, and K Hostettmann,  
Université de Lausanne, Lausanne, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The coupling of high-performance liquid chromatography with nuclear magnetic resonance spectroscopy (LC/NMR) is one of the most powerful methods for the separation and structural elucidation of unknown compounds in mixtures. The recent progress in pulse field gradients and solvent suppression, the improvement in probe technology, and the construction of high-field magnets have given a new impulse to this technique which has emerged since the mid-1990s as a very efficient method for the on-line identification of organic molecules. It took nearly 20 years to establish LC/NMR and this long period can be mainly attributed to the intrinsic low sensitivity of NMR. LC/NMR represents an interesting complementary technique to LC/UV/MS for on-line identification of LC peaks. Recent applications have demonstrated its usefulness in various fields of analytical chemistry.

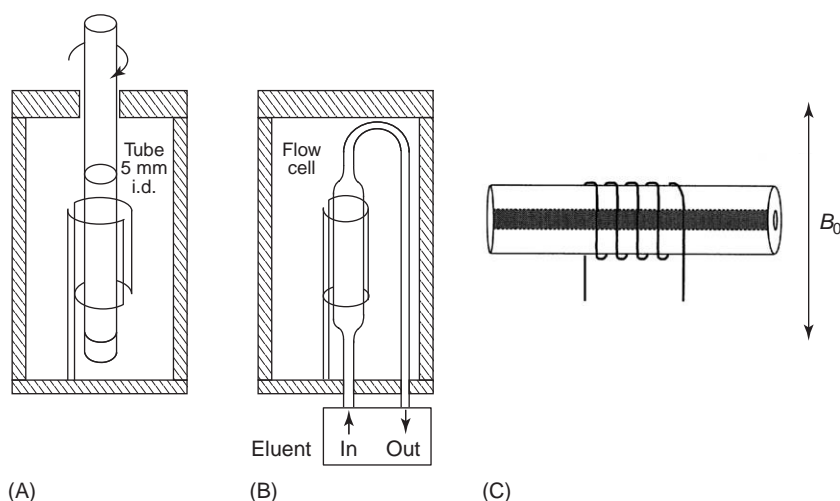
### NMR Flow Cell Design

The heart of an LC/NMR system includes a dedicated continuous-flow probe. In a classical NMR

experiment, compounds of interest are dissolved in a deuterated solvent and the sample is rotated in a 5 mm i.d. tube to eliminate field inhomogeneities (Figure 1A). With modern NMR instrument the tube is not rotated anymore since for highly sophisticated two-dimensional (2D) experiments rotation causes severe distortions. The design of continuous-flow NMR probes for cryomagnets, introduced in the early 1980s, was in contradiction with this classical conventional tubes setup. In this case, a U-type glass tube was fixed in the dewar of the NMR probe body, breaking the central symmetry of the magnetic field. The first application of this design ('saddle'-shaped geometry) (Figure 1B) showed, however, an excellent resolution approaching those registered with the rotation of the NMR tubes. Saddle-shaped continuous-flow NMR probes for routine applications employs detection volumes between 40 and 120  $\mu\text{L}$ , which are much larger than conventional ultraviolet (UV) detection volumes (8  $\mu\text{L}$ ).

The use of such important detection volumes resides on a compromise that has to be made between the needs for chromatography and those required by NMR. Indeed, while conventional high-performance liquid chromatography (HPLC) can accommodate detection volumes of 5–10  $\mu\text{L}$ , NMR needs extremely large detection volumes for two reasons: (1) line-shape quality and (2) sensitivity.

1. Line-shape quality: When operated in the flowing mode (on-flow), there is a distinct residence time



**Figure 1** Schematic of (A) conventional NMR probes, (B) saddle-type continuous-flow LC/NMR probes, and (C) solenoidal continuous-flow LC/NMR probes. (Reproduced with permission from Albert K (2002) *On-line LC–NMR and Related Techniques*. Chichester: Wiley; © John Wiley & Sons Ltd.)

(dwell time:  $\tau$ ) of the nuclei in the NMR detection cell which has to be considered. This factor was not affecting the standard NMR experiment since nuclei were staying for infinite time period in the tube. In LC/NMR,  $\tau$  is defined by the ratio of the detection volume to the flow rate. If  $\tau$  is below 5 s a line broadening of the NMR signal will be observed which will reduce the NMR spectral resolution. For example, the measured signal half-width of chloroform at 400 MHz will be of 0.55 Hz if the flow is stopped, of 0.75 Hz with a flow of  $0.5 \text{ ml min}^{-1}$  and 1.05 Hz at  $1.0 \text{ ml min}^{-1}$ . In order to maintain a tolerable line broadening the detection volume of the LC/NMR probe was  $44 \text{ }\mu\text{l}$  in this case, which gives a residence time  $\tau$  of 5.28 s at  $0.5 \text{ ml min}^{-1}$ .

2. Sensitivity: The second reason is that LC/NMR is a volume-sensitive detection technique and requires maximization of NMR-active nuclei by extension of the detection volume. The concentration in later LC eluting peaks due to the diffusion of the LC chromatographic process is reduced by a factor of at least 3 and NMR detection of these dilute solutions would require very high detection volumes in the milliliter range.

The design of the actual flow cells thus results in a compromise between needs for resolution and sensitivity on the NMR side (large volume) and reasonable LC peak separation on the chromatographic side (small volume).

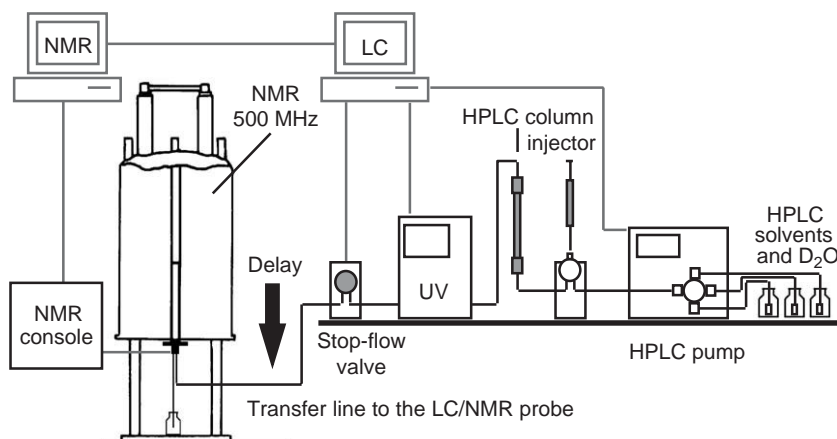
### Small-Volume LC/NMR

With the constant need for analysis in the smaller domains, challenges for inherently insensitive methods such as LC/NMR become increasingly severe. Indeed, in the downscaling from  $200 \text{ }\mu\text{l}$  to  $20 \text{ nL}$ ,

10 000-fold fewer spins are present for a given concentration which causes severe sensitivity problems. In this respect, the development of small-volume radiofrequency (RF) probes has experienced a renaissance in the last decade. The design of these microliter-scale probes consists mainly in a microcoil wound directly on to a capillary tube which acts as both a sample container and a coil form (Figure 1C). The fabrication material used has near-zero magnetic susceptibility for decreasing linewidth, improving lineshape and S/N. The LC flow axis of these cells is perpendicular to the magnetic field  $B_0$ , contrary to the conventional LC/NMR probes of larger volume which adopt the saddle-type geometry (Figure 1B). These new microliter-scale probes operating with nanoliter volumes are becoming commercially available and offers interesting possibilities for sample-limited applications.

### Experimental Arrangement

From a practical point of view, a standard LC/NMR experiment can be performed on an NMR system with magnet field strength higher than 9.4 T ( $^1\text{H}$  resonance frequency of 400 MHz) with a dedicated flow-probe and a standard HPLC system. In conventional setups, the HPLC system will be installed at 2–4 m from the magnet; however, with shielded magnets it can be installed directly close to it. The main constraint will be the installation of a valve before the probe for the recording of continuous- or stop-flow spectra (Figure 2). Indeed, while LC/NMR was originally designed for continuous-flow NMR acquisition, the need for full structural assignment of unknown constituents led to major applications in the stop-flow mode, which enable the recording of 2D correlation experiments referred to in this article later.



**Figure 2** Typical setup used for on-flow and direct stop-flow LC/NMR experiments. The control of the stop-flow valve is achieved by the computer of the LC/UV system which triggers both the valve of the HPLC pump and the NMR acquisition computer. A calibrated delay is used for parking the LC-peak of interest with precision at the center of the LC/NMR flow probe.

## Practical Considerations

### Solvent Signal Suppression

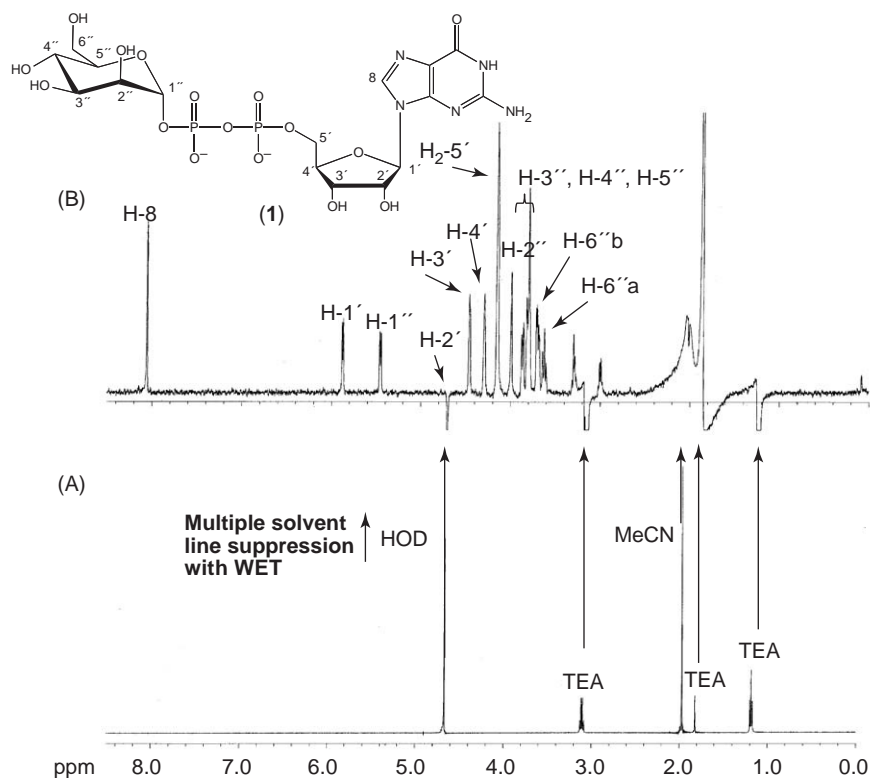
The main problem of LC/NMR is the difficulty in observing analyte resonances in the presence of the much larger resonances of the mobile phase. This problem is even worsened in the case of typical LC reversed-phase operating conditions, where more than one protonated solvent is used and where the resonances change frequencies during analysis in the gradient mode. Furthermore, the continuous flow of sample in the detector coil complicated solvent suppression. These problems have now been overcome thanks to the development of fast, reliable, and powerful solvent suppression techniques. These techniques are mainly:

- presaturation (NOESY presaturation);
- soft pulse multiple irradiation; and
- WET (water suppression enhanced through  $T_1$  effects) presaturation.

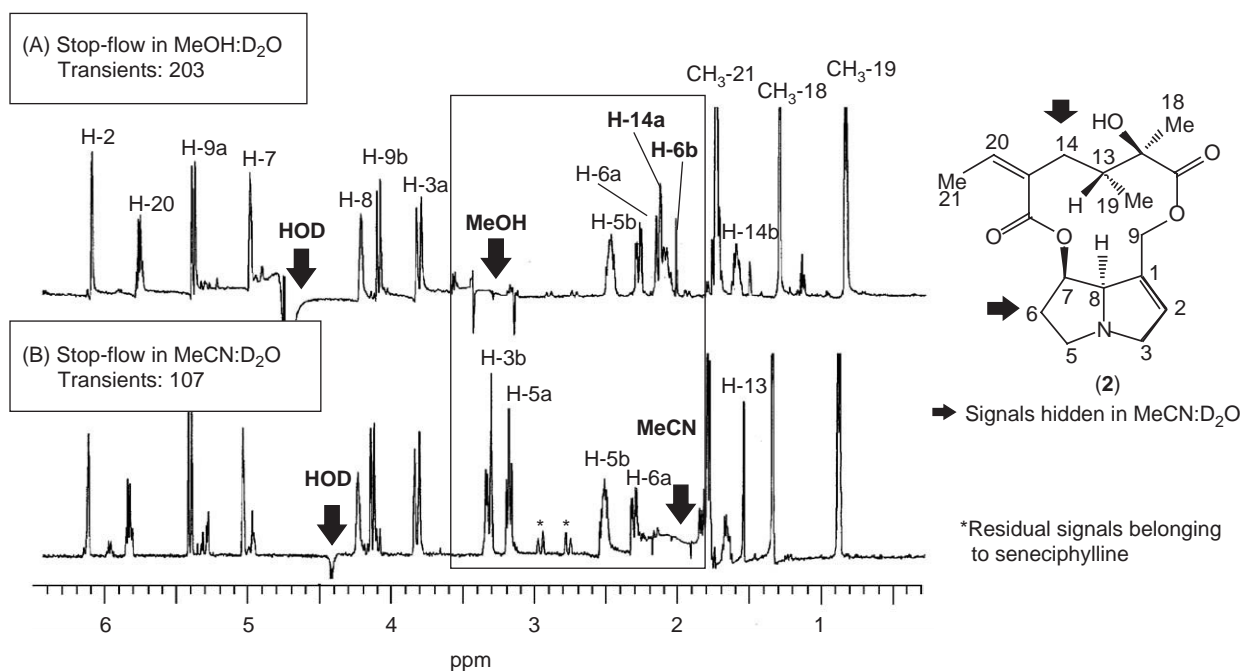
All these techniques can be used either in the continuous- or the stop-flow mode. WET, in particular,

consists of a combination of pulsed field gradients, shaped RF pulses, and shifted laminar pulses. By varying the tip angle of the selective RF pulses, the WET sequence can be optimized. This sequence provides a fast and efficient saturation of multiple solvent resonances. It can be used in combination with  $^{13}\text{C}$  decoupling for efficient removal of the  $^{13}\text{C}$  satellites of the solvent, which will otherwise strongly affect the spectral quality. An example of the application of the WET sequence for suppression of multiple signals from solvent and ion-pair reagents is shown in Figure 3 for the analysis of GDP-mannose (1).

The main drawback of these solvent suppression procedures is that the analyte signals localized under the solvent resonances will also be suppressed and thus the whole range of  $^1\text{H}$ -NMR signals are not directly observable. This can be a major drawback when dealing with unknown constituents. In order to detect all analyte signals, one alternative is to carry out the solvent suppression in two independent solvent systems such as MeCN– $\text{D}_2\text{O}$  and MeOH– $\text{D}_2\text{O}$ . Such an approach is shown for the analysis of hepatotoxic pyrrolizidine alkaloids in the crude



**Figure 3** Multiple suppression of solvent and ion-pair reagents in the LC/NMR spectrum of GDP-mannose with the WET sequence. (A) Stop-flow spectrum recorded without solvent suppression, solvent signals and ion-pair reagents hinder the detection of the signals of the analyte GDP-mannose. (B) Stop-flow spectrum of GDP-mannose (1) (200  $\mu\text{g}$ , NT = 1024) with suppression of all signals due to the LC eluent. (Reprinted with permission from Ramm M, Wolfender J-L, Queiroz EF, Hostettmann K, and Hamburger M (2004) Rapid analysis of nucleotide activated sugars by high performance liquid chromatography coupled with diode-array detection, electrospray ionization mass spectrometry and nuclear magnetic resonance. *Journal of Chromatographic A* 1034:139–148; © Elsevier.)



**Figure 4** Stop-flow spectra measured for senecionine (**2**) in the lipophilic extract of *Senecio vulgaris* (Asteraceae) in both MeCN–D<sub>2</sub>O and MeOH–D<sub>2</sub>O solvent systems. HPLC conditions: Inj. 3 mg; Col. Symmetry C-18 (150 × 3.9 mm i.d., 5 μm); (A) MeOH–D<sub>2</sub>O gradient (10:90 to 70:30 in 30 min); (B) MeCN–D<sub>2</sub>O gradient (5:95 to 50:50 in 20 min); 0.015 mol l<sup>−1</sup> NH<sub>3</sub>; 0.8 ml min<sup>−1</sup>. LC/NMR conditions: flow cell (60 μl, 3 mm i.d.), 500 MHz. (Reproduced with permission from Wolfender J-L, Ndjoko K, and Hostettmann K (2001) The potential of LC–NMR in phytochemical analysis. *Phytochemical Analysis* 12: 2–22; © John Wiley & Sons Ltd.)

extracts of *Senecio* species (Figure 4). In this case, the LC/<sup>1</sup>H-NMR analyses performed in both solvent systems have enabled the recording of the whole range of <sup>1</sup>H-NMR signals after combination of the information found in both spectra. For senecionine (**2**), the signals in MeCN–D<sub>2</sub>O corresponding to H-14a and H-6b were hidden by the suppression of the acetonitrile signal (Figure 4B) while these resonances were, however, clearly visible when LC/NMR was performed in MeOH–D<sub>2</sub>O.

One way to overcome the problem of solvent suppression is to perform the separation in fully deuterated solvents. This, however, cannot be envisaged routinely on standard HPLC columns, operated at typical flow rates of 1 ml min<sup>−1</sup>, due to the cost of deuterated solvents. Nevertheless, separations may be envisaged on micro or capillary columns, which have a very low solvent consumption. Another alternative consists in column-trapping on solid phase extraction (SPE) and back-flushing into the LC/NMR flow cell using an appropriate deuterated solvent (SPE-NMR), which will yield very 'clean' spectra.

#### HPLC Conditions

In comparison with LC/UV or LC/MS analyses, the on-column loading for LC/NMR is of several orders

of magnitude higher. In order to achieve the best sensitivity, the goal is to bring the highest amount of separated analyte in the lowest elution volume into the LC/NMR cell. Ideally, the chromatographic peak elution volume should be equivalent to the NMR flow cell volume (see above). For crude plant extract analysis, for example, the amount injected on-column can be of several milligrams while only a few micrograms are necessary for LC/UV. In order to cope with these important sample loadings, long HPLC columns (250 mm) or columns with an important diameter (8 mm) are used for standard LC/NMR measurements. In consequence, specific chromatographic conditions have to be developed for LC/NMR in order to separate large amounts of sample with satisfactory LC resolution and maintain flow conditions compatible with the detection cell used.

Most of the applications are performed under typical reversed-phase HPLC conditions. Nondeuterated solvents such as MeOH or MeCN are used. Great care in selecting the appropriate solvent grades is important since most solvents contain small impurities such as stabilizing chemicals, which will not be detected in UV but will cause interferences in the LC/NMR spectra. Water is usually replaced by D<sub>2</sub>O, which is a relatively 'cheap' deuterated solvent, and gives better quality spectra.



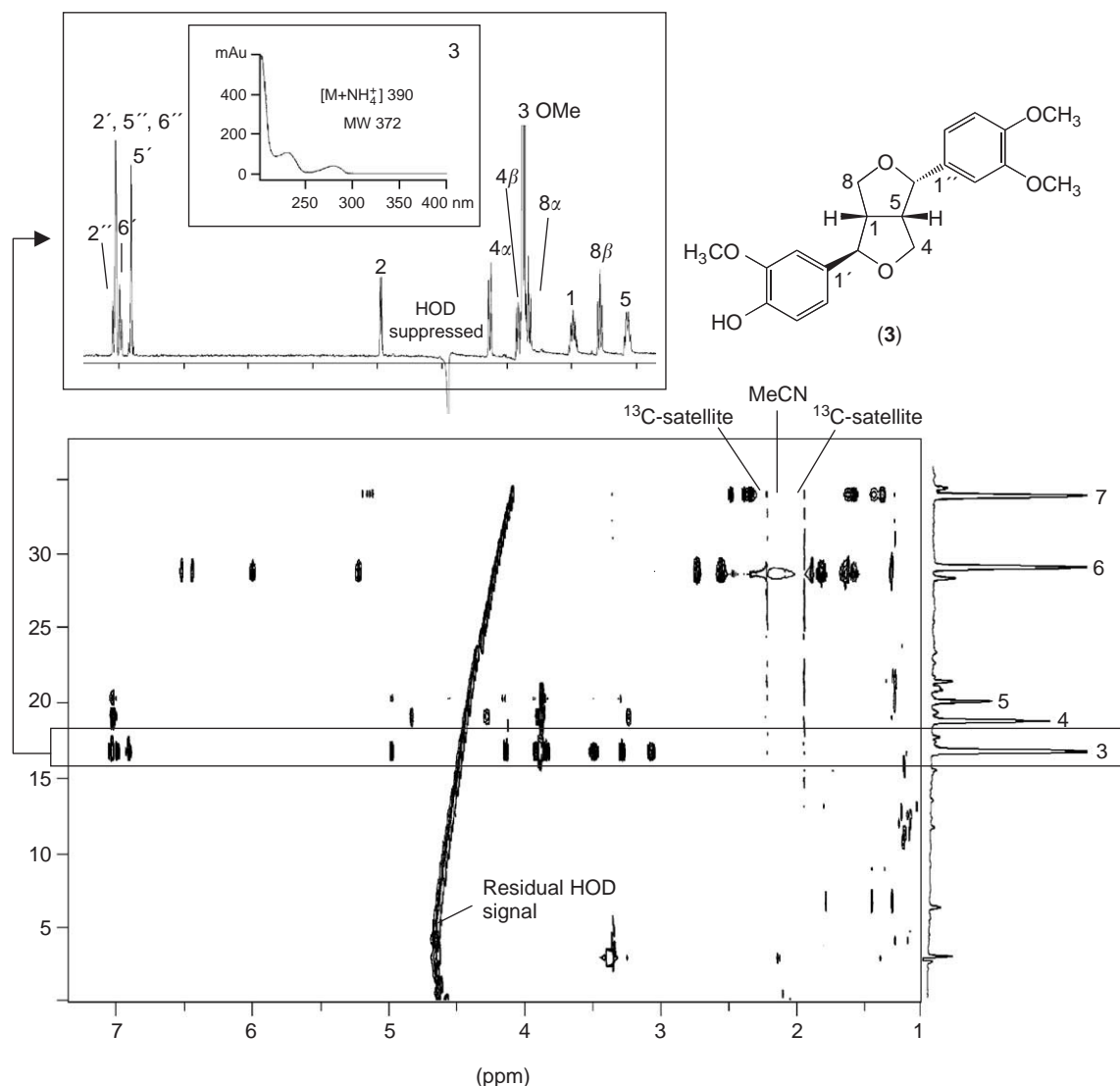
## Gradient Elution

During gradient elution, the frequencies of solvent signals are shifted due to changes of dielectric constants related to the modification of solvent composition. For an effective solvent suppression on-the-fly, a single exploratory scan ('scout scan') is acquired to determine the exact position solvent signals and their suppression is consequently performed by calculation of an adapted selective pulse. The shift of the residual HOD line during gradient elution in the on-flow contour plot of Figure 5 illustrates this phenomenon.

In this case the residual line of MeCN was locked at  $\delta$  2.00.

## Sensitivity

Compared to other analytical techniques (LC/UV or LC/MS), LC/NMR can be regarded as relatively insensitive due to intrinsic properties of NMR detection, which is a consequence of the small energy gap between ground and excited nuclear spin states. On the other hand, limits of detection will also depend on the type of magnet used, LC/NMR applications have been reported on magnets of 400 up to



**Figure 5** On-flow 2D LC/ $^1\text{H}$ -NMR chromatogram of the antioxidant fraction from the  $\text{CH}_2\text{Cl}_2$  extract of *Orophea enneandra* on NovaPak RP-18 ( $150 \times 3.9$  mm i.d.), MeCN- $\text{D}_2\text{O}$  gradient (20:80 to 95:5 in 50 min,  $1 \text{ ml min}^{-1}$ ), 0.05% TFA, 2 mg injected. UV trace was recorded in the same conditions with  $20 \mu\text{g}$  injected. The on-flow spectrum of (–)-phylligenin (3) (NT=32) is displayed. The shift of residual signal of HOD is indicated. LC/NMR conditions: flow cell ( $60 \mu\text{L}$ , 3 mm i.d.), 500 MHz. (Reprinted with permission from Cavin A, Potterat O, Wolfender J-L, Hostettman K, and Dyetmike W (1998) Use of on-flow LC/ $^1\text{H}$  NMR for the study of an antioxidant fraction from *Orophea enneandra* and isolation of a polyacetylene, lignans and a tocopherol derivative. *Journal of Natural Products* 61: 1497–1501; © American Chemical Society.)

900 MHz, and the type of flow probes employed (standard LC/NMR probes: 40–120  $\mu\text{l}$  or microcoils; 20 nl to 1  $\mu\text{l}$ ). The use of new cryo-flow-probes enhances also considerably the sensitivity and a gain of a factor 4 in sensitivity can be roughly obtained compared to noncryogenated systems. Furthermore, the use of new techniques of NMR data acquisition, such as digital filtering and oversampling, also enhance the signal-to-noise ratio (S/N).

The mode of operation also strongly affects sensitivity since in continuous-flow measurements the time of acquisition is limited, while in stop-flow important number of transients can be recorded.

As an example, absolute  $^1\text{H}$ – $^{13}\text{C}$  detection limits of a 100  $\mu\text{l}$  commercially available LC/NMR probe was estimated for 3'-deoxy-3'-azidothymidine (AZT) (MW 252) on a 500 MHz magnet in the stop-flow mode. The S/N ratio of the H'1 proton of the ribose moiety of this molecule was used to define the detection limits. An S/N ratio of 3.0 for this signal was obtained for 1  $\mu\text{g}$  of AZT after 64 scans (5 min) and for 85 ng after an overnight experiment. With 21  $\mu\text{g}$  of AZT, a  $^1\text{H}$ – $^1\text{H}$  2D TOCSY was acquired in 2.6 h and all the cross-peaks of  $^1\text{H}$ – $^{13}\text{C}$  HMQC could be measured overnight. Recently, it was demonstrated that it was possible with a new capillary microcoil probe to detect amounts of  $\alpha$ -pinene as low as 37 ng in the on-flow mode and that detection of an impurity in the subnanomole range was made possible in the stop-flow mode.

## Modes of Operations

Due to the lack of sensitivity, most of the LC/NMR applications requires the accumulation of many transients for achieving a satisfactory detection of the compounds of interest. For the recording of detailed structural information, measurement of 2D correlation experiments is also often mandatory. Consequently, LC/NMR can be operated not only in dynamic but also in static conditions. These main operation modes are described below and require different level of automation:

- on-flow,
- direct stop-flow,
- time-slice,
- loop storage/loop transfer.

### On-Flow Mode

In the on-flow mode, the LC/NMR spectra are acquired continuously during the separation and are stored as a set of scans in discrete increments. The

on-flow data are processed as a 2D NMR experiment. One dimension of the contour plot represents the NMR scale ( $\delta$  ppm) and the other the time scale (min) (Figure 5). According to the analysis, a compromise between the number of scans per increment and the LC resolution has to be made. The main drawback of on-flow LC/NMR is its low sensitivity. In our experience, a rough estimation of the detection limit in the on-flow mode with a 60  $\mu\text{l}$  cell on a 500 MHz instrument was realized for the secoiridoid glycoside swertiamarin (MW 374). An injection of 20  $\mu\text{g}$  of this compound on-column was necessary in order to obtain an S/N ratio of 3 for the olefinic proton H-3 (16 scans/increment 1 ml min<sup>-1</sup>). Thus, practically on-flow LC/NMR measurements will be mainly restricted to the direct measurement of the main constituents of complex mixtures and this often under 'overloaded' LC conditions (typically 1–5 mg of crude mixture injected on-column in the case of crude plant extract analysis).

A simple example of the type of data that can be obtained in the on-flow mode is shown for the online identification of the antioxidant constituents of the  $\text{CH}_2\text{Cl}_2$  extract of an Indonesian medicinal plant *Orophea eneandra* (Annonaceae). This extract had a relatively simple composition consisting of five main constituents. The separation was achieved at 1 ml min<sup>-1</sup> with a MeCN–D<sub>2</sub>O gradient (20:80–95:5 in 50 min). A loading of 2 mg of the extract on-column was necessary and spectra were recorded with 32 transients per increment on a 500 MHz instrument equipped with a 60  $\mu\text{l}$  flow cell. Solvent suppression was performed on-the-fly by applying the WET sequence.

As shown in Figure 5, the LC/ $^1\text{H}$ -NMR spectra were successfully recorded for the five major constituents (3–7) on the LC/NMR contour plot. The important trace starting from 4.8 ppm (at 0 min) and ending at 4 ppm (at 35 min) was due to the change of the chemical shift of the residual negative water (HOD) signal during the LC gradient. The traces around 2 ppm are due to residual  $^{13}\text{C}$  satellite signals of MeCN.

A detailed interpretation of LC/ $^1\text{H}$ -NMR in combination with LC/UV-DAD, LC/MS, and chemotaxonomical data (see the case of 3 in Figure 5) enables the identification of three known lignanes (–)-phylligenin (3), (–)-eudesmin (4), (–)-epieudesmin (5) as well as polycerasoidol (6). Compound (7) could be completely identified based on online data and was finally characterized as a new polyacetylenic natural product named oropheic acid (7). This example also demonstrates the use of LC/UV LC/MS and LC/NMR in phytochemical analysis; known natural products can be dereplicated by comparison

of the spectroscopic information obtained online with literature data and isolation of putative new compounds having particular spectroscopic features can be efficiently targeted.

### Low-Flow Mode, Time-Slice Mode

The detection limits of the on-flow mode can be improved by performing the analysis at low-flow ( $<0.1 \text{ ml min}^{-1}$ ) or by running time-slice experiments over a whole chromatogram. In this latter case, the flow is stopped at defined time intervals. Both modes of operation enable higher number of transients per increments to be recorded and thus a significant improvement in S/N ratio is obtained.

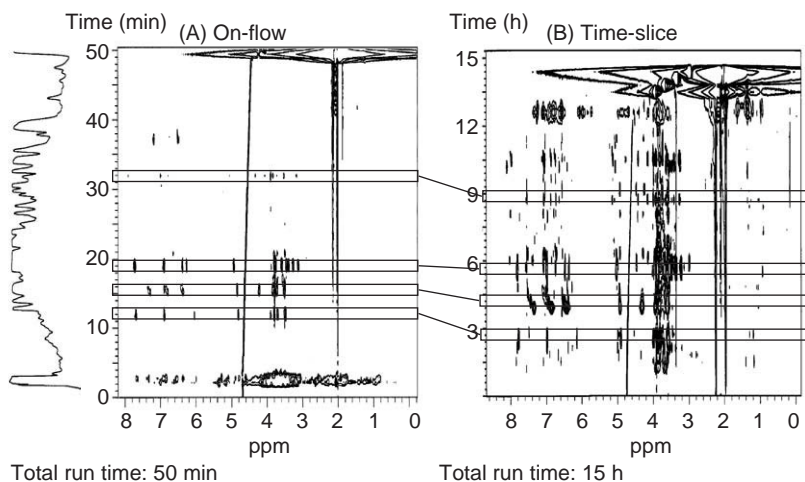
An example of a time-slice experiment on a whole crude plant extract LC chromatogram is shown for the analysis of the constituents of *Gnidia involucrata* (Thymelaceae). As shown in Figure 6, a comparison of the contour plot of the on-flow LC/NMR analysis and the time-slice stop-flow analysis clearly demonstrates the gain in sensitivity that can be obtained in this latter mode. In the on-flow mode (Figure 6A) ( $nt = 24$ , per increment), only the four main constituents of *G. involucrata* were detected but in the time-slice LC/NMR analysis ( $nt = 1024$ ,  $\sim 15 \text{ min}$  acquisition per slice), an important number of minor constituents ( $>20$ ) were revealed (see compounds eluting between 7–12 h in Figure 6B). These different constituents correspond to various xanthone and flavone O- and C-glycosides.

### Stop-Flow Mode

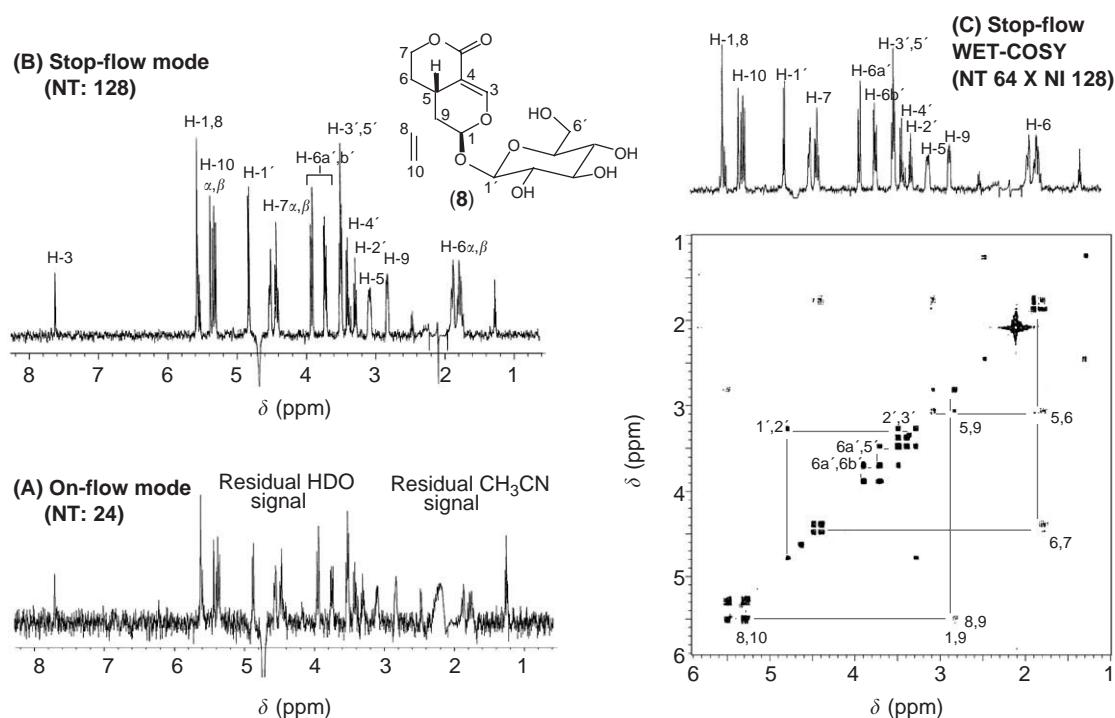
Another way to improve the S/N ratio is the recording of the spectra in stop-flow mode. Operation in the direct stop-flow mode requires that the retention times of the analytes of interest are known or that a sensitive method of detection such as LC/UV or LC/MS is used prior to LC/NMR to trigger the detection. In practice, one of these detectors is connected online before the NMR instrument and the signal of the analytes of interest passing through this detector is used to trigger a valve that will stop the LC flow exactly when the peak reaches the NMR cell after a calibrated delay (for setup, see Figure 2).

The stop-flow mode gives the possibility to acquire an important number of transients on a given LC peak and LC/ $^1\text{H}$ -NMR spectra of compounds present even in the nanogram range can be obtained. An example of the improvement of spectral quality obtained in the stop-flow mode is shown in Figure 7 where the spectra of the bitter principle sweroside (8) recorded in the crude extract of the Gentianaceae *Swertia calycina* in the on-flow and stop-flow modes are compared. The measurement of various 2D correlation experiments such as COSY, NOESY, HSQC, and HMBC is also possible in the stop-flow mode provided that the amount of analyte is high enough. An example of the WET-COSY spectrum recorded for sweroside (8), recorded directly from the crude extract of *S. calycina*, is also shown in Figure 7.

As mentioned earlier, stop-flow can also be operated in the time-slice mode where the flow is



**Figure 6** Comparison of the on-flow and time-slice LC/NMR contour plots of the MeOH extract of the aerial parts of *Gnidia involucrata* (Thymelaceae). As shown, the on-flow experiment allows the detection of four major compounds while the time-slice procedure reveals  $\sim 20$  components. Amount injected: 20 mg. HPLC conditions: C-18 column with radial compression, Waters RCM  $8 \times 10$  ( $100 \times 8 \text{ mm i.d.}$ ); MeCN- $\text{D}_2\text{O}$  gradient (5:95 to 20:80 in 50 min);  $0.9 \text{ ml min}^{-1}$ . LC/NMR conditions: 24 scans/increment (on-flow); 1024 scans/increment (time-slice);  $60 \mu\text{l}$  flow cell ( $3 \text{ mm i.d.}$ ); 500 MHz. (Reproduced with permission from Wolfender J-L, Ndjoko K, and Hostettmann K, (2001) The potential of LC-NMR in phytochemical analysis. *Phytochemical Analysis* 12: 2–22; © John Wiley & Sons Ltd.)



**Figure 7** Comparison of on-flow (A) and stop-flow (B) LC/ $^1\text{H}$ -NMR spectra of sweroside (**8**) in the  $\text{CH}_2\text{Cl}_2$  extract of *Swertia calycina*. (C) WET-COSY spectrum (NT 64  $\times$  NI 128) recorded in the stop-flow mode. LC/NMR conditions: 60  $\mu\text{L}$  flow cell (3 mm i.d.); 500 MHz. (Reproduced with permission from Wolfender J-L, Ndjoko K, and Hostettmann K (2001) The potential of LC–NMR in phytochemical analysis. *Phytochemical Analysis* 12: 2–22; © John Wiley & Sons Ltd.)

automatically stopped at given time intervals through a given peak or on a whole chromatogram. This mode is particularly convenient for analyzing LC peaks that are partially overlapped.

The stop-flow mode can also be operated in a loop storage loop–loop transfer mode. In this case, the LC peaks are stored in a storage device consisting of multiple loops. The analytes are parked in these loops without interrupting the separation. At a later stage the content of the loop is transferred in the LC/NMR flow cell like in the direct stop-flow mode. This process in two steps has the advantage that no disturbance due to the multiple stops in the separation is recorded and that no limitation due to diffusion effects occurs; furthermore, the transfer in the loops is completely independent from the NMR measurements and can be made separately. The mode of operation can be completely automated and different series of NMR experiments can be programmed in advance for each analyte of interest.

### SPE–NMR

The idea of working with LC/NMR at-line, instead of online, gives the possibility to trap the LC/peak of interest on a SPE column installed postcolumn by adding an additional amount of water in the case of typical reversed-phase separations. The SPE cartridge

is dried and the content is transferred into the LC/NMR flow probe with the aid of an appropriate deuterated solvent. This approach enables a preconcentration of the sample of interest on the SPE, multiple trapping are possible, and the S/N ratio is thus enhanced. The separation can be carried out in normal HPLC grade solvents while deuterated solvents are used only to transfer analytes into the NMR probe. This provides ‘cleaner’ spectra without need for solvent suppression.

### Other Coupling

Besides coupling with HPLC, NMR can also be hyphenated with other separation techniques such as gel permeation chromatography (GPC/NMR), supercritical fluid chromatography (SFC/NMR), capillary HPLC (CapLC/NMR), and electrodriven separation techniques such as capillary electrophoresis (CE/NMR).

GPC/NMR coupling opens new opportunities mainly for the characterization of polymers. SFC is very often considered as a niche separation technique mainly for nonpolar analytes. Coupled to NMR, however, SFC/NMR demonstrates the inherent advantage that no solvent suppression is needed and the whole proton chemical shift range can be used



without distortions of solvent signals or impurities. With the development of new NMR flow probes for applications in the microdomains (see above), CapLC/NMR and CE/NMR represents new exciting research fields that enable the characterization of analytes in the nanogram range while carrying separation in fully deuterated solvents.

## Applications of LC/NMR

LC/NMR is now well established and has been integrated in many laboratories often as a complement to LC/UV-DAD and LC/MS for complete characterization of unknown metabolites. Applications of this technique are thus manifold. LC/NMR/MS has demonstrated its potential for many biomedical and pharmaceutical applications and strategies for drug impurities identification, reactivity study, or characterization of mixtures issued from combinatorial chemistry have been described. In drug metabolism studies, applications of LC/NMR are numerous. In this field, LC/NMR was found to be very efficient since structures of the parent drugs are known and precise assignment of the modification occurring after metabolization can be ascertained based on LC/MS and LC/NMR with very good accuracy. In drug metabolism, the use of LC/ $^{19}\text{F}$ -NMR as a very specific detection method for the localization of the metabolites of a fluorinated parent drug was found very efficient since detection of  $^{19}\text{F}$  is much more sensitive than  $^1\text{H}$  in NMR.

LC/NMR in various combinations with LC/UV-DAD, LC/MS, LC/MSMS, LC/IR, and/or LC/CD has been used in many applications related to the online identification of natural products. In this field, the challenge for hyphenated techniques is important since often the characterization of completely unknown molecules is required in very complex biological matrices. In this case, LC hyphenated techniques are used for the chemical evaluation of biologically active fractions or extracts and for de-replication purposes. As full structure assignment is often needed, all online spectroscopic data are taken into consideration. Most applications are performed in the stop-flow mode and extensive 2D NMR correlation experiments are measured. For unknown online determination the need for  $^{13}\text{C}$  data is often mandatory. This type of information can be deduced from HSQC and HMBC indirect measurements and very recently it has been demonstrated that even direct  $^{13}\text{C}$  measurements were possible in a crude plant extract. For this application the LC peak of interest was preconcentrated by trapping on SPE and the measurement was performed on a cryogenic flow

probe. In this field, LC/NMR has also revealed to be very useful for the structural assignment of unstable natural products or compounds not isolable at the preparative scale.

Different applications of LC/NMR in environmental problems have also demonstrated the interest of the approach for nontarget analysis of organic compounds in the samples. On the basis of the LC/NMR and LC-MS, a screening of various pollutants can be performed and target analysis of specific pollutants can be efficiently developed in a second step with more sensitive methods for a definitive identification.

As shown in this section, LC/NMR has been used successfully in various field of analytical chemistry. Further optimizations of the technique are, however, still awaited, especially on sensitivity issues. Very recently impressive progress, particularly with the introduction of liquid helium cooled cryoprobes and flow probes solenoid coils designs, has already improved considerably the detection limits.

LC/NMR is often used within multi-hyphenated systems in different fields of application, especially LC/NMR/MS. In this case, MS is used as an efficient and selective detection method to trigger NMR experiments only on the peaks of interest, which render the targeted measurement extremely valuable.

LC/NMR is probably still not mature for real demanding high throughput studies, but it is a safe bet to say that, considering the very recent development recorded, this powerful tool will play a major role in many aspects of analytical chemistry where the identification of LC peaks is mandatory.

**See also:** **Liquid Chromatography:** Overview; **Liquid Chromatography–Mass Spectrometry.** **Nuclear Magnetic Resonance Spectroscopy:** Overview; **Instrumentation.** **Nuclear Magnetic Resonance Spectroscopy Techniques:** Multidimensional Proton.

## Further Reading

- Albert K (1999) Liquid chromatography–nuclear magnetic resonance spectroscopy. *Journal of Chromatography A* 856: 199–211.
- Dalvit C, Shapiro G, Bohlen J-M, and Parella T (1999) Technical aspects of an efficient multiple solvent suppression pulse sequence. *Magnetic Resonance in Chemistry* 37: 7–14.
- Duarte IF, Godejohann M, Braumann U, Spraul M, and Gil AM (2003) Application of NMR spectroscopy and LC/NMR/MS to the identification of carbohydrates in beer. *Journal of Agricultural Food Chemistry* 51(17): 4847–4852.
- Lacey ME, Tan ZJ, Webb AG, and Sweedler JV (2001) Union of capillary high-performance liquid chromatography and microcoil nuclear magnetic resonance spectroscopy applied to the separation and identification



- of terpenoids. *Journal of Chromatography A* 922: 139–149.
- Lindon JC, Nicholson JK, and Wilson ID (2000) Directly coupled HPLC–NMR and HPLC–NMR–MS in pharmaceutical research and development. *Journal of Chromatography B* 748: 233–258.
- Pusecker K, Shewitz J, Gfrörer P, *et al.* (1998) On-line coupling of capillary electrochromatography, capillary electrophoresis and capillary HPLC with nuclear magnetic resonance. *Analytical Chemistry* 70: 3280–3285.
- Schneider B (2000) Natural products: Liquid chromatography–nuclear magnetic resonance. In: Wilson ID, Adlard TR, Cooke M, and Poole CF (eds.) *Encyclopedia of Separation Science*, pp. 3434–3445. London: Academic Press.
- Smallcombe SH, Patt SL, and Keiffer PA (1995) WET solvent suppression and its application to LC–NMR and high-resolution NMR spectroscopy. *Journal of Magnetic Resonance Series A* 117: 295–303.
- Wilson I (2000) Multiple hyphenation of liquid chromatography with nuclear magnetic resonance spectroscopy, mass spectrometry, and beyond. *Journal of Chromatography A* 892: 315–327.
- Wolfender F, Ndjoko K, and Hostettmann K (2003) Liquid chromatography with ultraviolet absorbance–mass spectrometric detection and with nuclear magnetic resonance spectroscopy: A powerful combination for the on-line structural investigation of plant metabolites. *Journal of Chromatography A* 1000: 437–455.

## Amino Acids

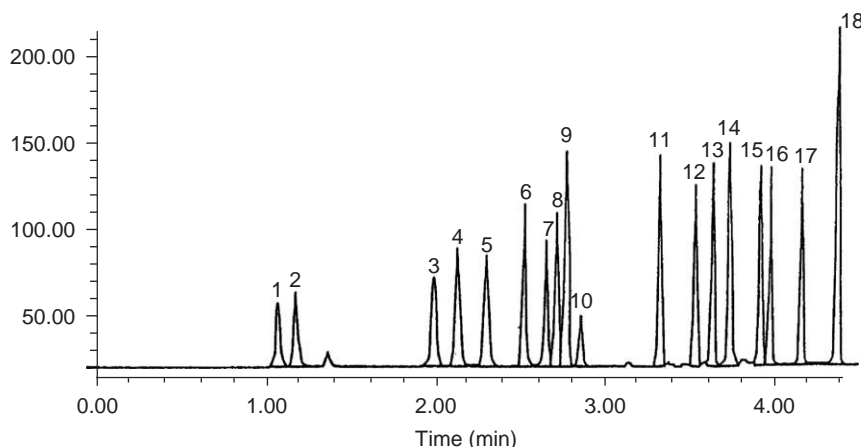
I Molnár-Perl, L Eötvös University, Budapest, Hungary

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The first approach to automatic liquid chromatography (LC) of amino acids (AAs) – known today as ion-exchange chromatography – was published by Spackman, Stein, and Moore in 1958, which led to the award of the Nobel Prize in 1972 to Stanford Moore and William H. Stein. Forty-five years later, it now takes less than 5 min (**Figure 1**) to separate and quantitate the essential protein AAs instead of 2 days. Early separations were carried out by postcolumn derivatizations.

Over the last 20 years, LC has offered unlimited possibilities in both preparative and analytical scales. The wide range of sophisticated columns, detectors, derivatization procedures, modern instrumentation, and data handling systems reduce operating time and costs, and give versatility and automation in Good Laboratory Practice conditions for selectivity, sensitivity, and reproducibility. It is the responsibility of the researcher to choose the most appropriate method for a given task. The most popular LC method for analysis of both free AAs (present in many natural matrices, biological fluids, and tissues, feedstuffs, and foodstuffs) and of constituents of protein hydrolysates is now reverse-phase (RP) chromatography after precolumn derivatization of the AAs. Numerous methods are available for derivatization.



**Figure 1** Separation of the phenylthiocarbamyl (PTC) AAs separated on TSKgel Super-ODS (details in **Table 3**), peaks: 1 = Asp, 2 = Glu, 3 = Ser, 4 = Gly, 5 = His, 6 = Arg, 7 = Thr, 8 = Ala, 9 = Pro, 10 =  $\text{NH}_4^+$ , 11 = Tyr, 12 = Val, 13 = Met, 14 = Cys, 15 = Ile, 16 = Leu, 17 = Phe, 18 = Lys. (Reproduced with permission from Catalogue (2003) *TosoHaas Biosciences*, p. 16.)

- of terpenoids. *Journal of Chromatography A* 922: 139–149.
- Lindon JC, Nicholson JK, and Wilson ID (2000) Directly coupled HPLC–NMR and HPLC–NMR–MS in pharmaceutical research and development. *Journal of Chromatography B* 748: 233–258.
- Pusecker K, Shewitz J, Gfrörer P, *et al.* (1998) On-line coupling of capillary electrochromatography, capillary electrophoresis and capillary HPLC with nuclear magnetic resonance. *Analytical Chemistry* 70: 3280–3285.
- Schneider B (2000) Natural products: Liquid chromatography–nuclear magnetic resonance. In: Wilson ID, Adlard TR, Cooke M, and Poole CF (eds.) *Encyclopedia of Separation Science*, pp. 3434–3445. London: Academic Press.
- Smallcombe SH, Patt SL, and Keiffer PA (1995) WET solvent suppression and its application to LC–NMR and high-resolution NMR spectroscopy. *Journal of Magnetic Resonance Series A* 117: 295–303.
- Wilson I (2000) Multiple hyphenation of liquid chromatography with nuclear magnetic resonance spectroscopy, mass spectrometry, and beyond. *Journal of Chromatography A* 892: 315–327.
- Wolfender F, Ndjoko K, and Hostettmann K (2003) Liquid chromatography with ultraviolet absorbance–mass spectrometric detection and with nuclear magnetic resonance spectroscopy: A powerful combination for the on-line structural investigation of plant metabolites. *Journal of Chromatography A* 1000: 437–455.

## Amino Acids

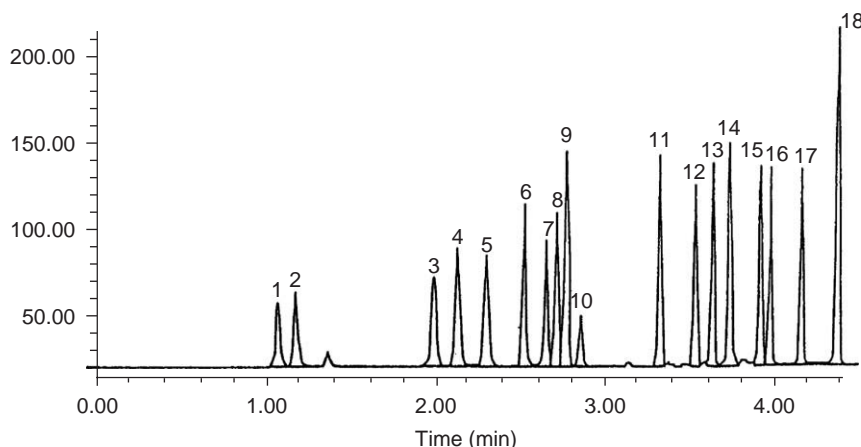
I Molnár-Perl, L Eötvös University, Budapest, Hungary

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The first approach to automatic liquid chromatography (LC) of amino acids (AAs) – known today as ion-exchange chromatography – was published by Spackman, Stein, and Moore in 1958, which led to the award of the Nobel Prize in 1972 to Stanford Moore and William H. Stein. Forty-five years later, it now takes less than 5 min (**Figure 1**) to separate and quantitate the essential protein AAs instead of 2 days. Early separations were carried out by postcolumn derivatizations.

Over the last 20 years, LC has offered unlimited possibilities in both preparative and analytical scales. The wide range of sophisticated columns, detectors, derivatization procedures, modern instrumentation, and data handling systems reduce operating time and costs, and give versatility and automation in Good Laboratory Practice conditions for selectivity, sensitivity, and reproducibility. It is the responsibility of the researcher to choose the most appropriate method for a given task. The most popular LC method for analysis of both free AAs (present in many natural matrices, biological fluids, and tissues, feedstuffs, and foodstuffs) and of constituents of protein hydrolysates is now reverse-phase (RP) chromatography after precolumn derivatization of the AAs. Numerous methods are available for derivatization.



**Figure 1** Separation of the phenylthiocarbamyl (PTC) AAs separated on TSKgel Super-ODS (details in **Table 3**), peaks: 1 = Asp, 2 = Glu, 3 = Ser, 4 = Gly, 5 = His, 6 = Arg, 7 = Thr, 8 = Ala, 9 = Pro, 10 =  $\text{NH}_4^+$ , 11 = Tyr, 12 = Val, 13 = Met, 14 = Cys, 15 = Ile, 16 = Leu, 17 = Phe, 18 = Lys. (Reproduced with permission from Catalogue (2003) *TosoHaas Biosciences*, p. 16.)

In this article, special emphasis is placed (1) on the recently developed detector systems that have opened a new perspective/area for the direct quantitation of AAs in their underivatized form, and (2) on the elucidation of the background and contradictions of the most commonly used and popular *o*-phthaldialdehyde (OPA) derivatization method.

Current trends in AA analysis identify the best conditions for the separation of enantiomers and the development of LC-mass spectrometry (LC-MS, LC-MS/MS), detailed elsewhere.

## LC of Underivatized AAs

One of the main advantages of LC, which has appealed to chromatographers, is the separation of the 'classical 20' as underivatized AAs. In spite of a number of efforts using both earlier methods (indirect ultraviolet (UV) or fluorescence (FL) detection, *in situ* complexation with Cu(II), and electrochemical detections) and those using recent methods, which apply special detections (chemiluminescence nitrogen detector (CLNSD), ELSD, MS, MS-MS), the simultaneous LC of underivatized AAs has remained of secondary importance so far. The determination of a few selected AAs, such as tryptophan or sulfur-containing AAs, has proved to be significant for special tasks.

The aim of various investigations has been to render the time-consuming derivatization techniques unnecessary. However, the characteristics of the free AAs are considerably different from each other and their various structural properties have not permitted their easy resolution. Thus, in attempting to achieve better separation of free AAs, further means of discriminations are needed. For this purpose special techniques have been introduced, such as the use of various phase systems, ion pairs, and ligand-exchange chromatography, column-switching techniques, or anion-exchange chromatography with electrochemical detection.

Recently, a comprehensive work was published that applied uniform experimental conditions, providing a basis for comparison and critical evaluation of the state of the art of detection (not quantitation) of AAs in their underivatized form. Based on the comparison of seven different detection systems under the same chromatographic conditions, the following conclusions could be drawn: due to the intrinsic specificity of these detection systems, in order of listing, they ensured increasing selectivity associated with increasing sensitivity (characterized with the limit of detection (LOD) as follows: refractive index detection:  $\text{LOD} = 50 \text{ mg l}^{-1}$ ; indirect

conductivity:  $\text{LOD} = 1\text{--}25 \text{ mg l}^{-1}$  or  $11^{-1}$ ; chemiluminescence nitrogen:  $\text{LOD} = 0.33\text{--}0.7 \text{ mg l}^{-1}$ ; MS:  $\text{LOD} = 0.2\text{--}5 \text{ mg l}^{-1}$ ; tandem mass spectrometric, MS-MS:  $\text{LOD} = 50.08\text{--}0.8 \text{ mg l}^{-1}$ ) during detections.

For tryptophan and the sulfur-containing AAs (cysteine/cystine, methionine, glutathione, etc.), the fast isocratic elution of the underivatized samples has gained wide acceptance and is a powerful tool in their quantitation.

Tryptophan can be measured directly, within 8 min, in neutralized alkaline hydrolysates of feed- and foodstuffs, using an RP column, 5% methanol containing acetate buffer ( $\text{pH} \sim 4.0$ ) and UV detection (280 nm). The pulsed amperometric detection of sulfur-containing AAs, at the low picomole level, was carried out with an Au working electrode and an Ag/AgCl reference electrode, subsequent to their separation on both cation-exchange and RP columns, applying as mobile phase  $0.1 \text{ mol l}^{-1} \text{ HClO}_4/0.15 \text{ mol l}^{-1} \text{ NaClO}_4/5\% \text{ acetonitrile}$ .

## LC of Derivatized AAs

Derivatization studies have concerned the optimization of parameters, such as the yield and stability of derivatives, to separate and quantitate all AAs with a simple and fast elution procedure.

### Postcolumn Derivatization

Drawbacks of the postcolumn techniques (Table 1) are long elution times and the need of expensive devices, such as a delivery system for the derivatizing reagent (one or more extra pumps): (1) mixing chamber for the column effluent and the reagent(s); (2) special thermostable reactors (packed bed, air segmented, and/or coir reactors) ensuring the necessary delay for quantitative reactions accompanied with as small band broadening as possible; (3) finally, the mobile phase was probably incompatible with the derivatizing reagent. The preferred mobile phase might be inappropriate for the optimum conditions of the derivatization reaction. The early and current stage of postcolumn methods can be illustrated by the elution followed by postcolumn reaction with ninhydrin (NHYD; Figure 2A), with *o*-phthaldialdehyde/2-mercaptoethanol (OPA/MCE; Figure 2B), or with 1,2-naphthoquinone-4-sulfonyl chloride (NQS; Figure 2C). All three types of derivatives have been separated in most cases on ion-exchange resin columns. Recent postcolumn methods are, without exception, slow separations (Table 1). However, the efficiency of the recently published methods of NHYD derivatives using short columns proved to

**Table 1** Advances in the LC of postcolumn derivatized AAs, obtained with *o*-phthaldialdehyde/ $\beta$ -mercaptoethanol (OPA/MCE), with ninhydrine (NHYD) and 1,2-naphthoquinone-4-sulfonate (NQS)

Date	Column, cm $\times$ mm	Size ( $\mu$ m)	Type	Eluents (elution temperature in $^{\circ}$ C)	Detector UV (nm), <i>F</i> Ex/ <i>Em</i>	Reagent ( $^{\circ}$ C)	Analyte (nmol l $^{-1}$ )	RSD (%)	Matrix	No. of AAs/ el. time (min)
1958, Moore	150 $\times$ 0.9	40	Amberlite IR-120, IE	Citrate buffers, 0.2 mol l $^{-1}$ : pH 4.25 for first day (30 $^{\circ}$ C), and pH 4.25 for the second day (55 $^{\circ}$ C)	UV 440, 570	NHYD (–)	100–3000	–	Hydrolysates AAs	20/24–48 h
1992, Grunau	15 $\times$ 3	5	Pickering 'fast run'	Pickering Eluents A, (Li280), B, (Li750), C, (RG003) (42 $^{\circ}$ C)	UV 570	NHYD (130 $^{\circ}$ C)	20	–	Plasma AAs	59/~ 150
1995, Iwase	6 $\times$ 4.6	3	2622, Hitachi, IE	Five eluents: PF-1-PF-4, PF-RG, cont. Li salts, ethanol, benzyl alcohol, thiodiglycol, Brij-35 buffer with pH 2.8, 3.7, 3.6, 4.1, – (gradient program: 28–40 $^{\circ}$ C)	UV 440, 570	NHYD (130 $^{\circ}$ C)	50	<3	Plasma AAs	40/120
1973, Roth	25 $\times$ 6		Aminex 6, IE	Citrate buffers pH 3.20, 4.25, and 6.40 for 40, 60, and 70 min (34 $^{\circ}$ C for 100 min, then raised to 55 $^{\circ}$ C)	<i>F</i> , –	OPA/MCE (55 $^{\circ}$ C)	10	–	Model study	14/170
1986, Elrifi	60 $\times$ 9	–	IE	Pierce Pico-Buffer system, Li citr. buffer; pH of A, B, C, D, and E = 2.9, 3.1, 3.5, 3.4, and 2.3; temperature gradient: 0–44 min (34 $^{\circ}$ C), 44–128 min (63 $^{\circ}$ C)	<i>F</i> *, no data	OPA/MCE (40 $^{\circ}$ C)	0.63–45.0	–	AAs in foods	23/128
1988, Haginaka	30 $\times$ 4.6 + guard 4.6 $\times$ 5	5 3	ODS-5	A: 15 mmol l $^{-1}$ naoctanesulfonate/ 21 mmol l $^{-1}$ H <sub>3</sub> PO <sub>4</sub> / 9 mmol l $^{-1}$ NaH <sub>2</sub> PO <sub>4</sub> / CH <sub>3</sub> OH (20/20/20/1, v/v), pH 2.8; B: as A, except (1/1/ 1/6, v/v), pH 4.2 (60 $^{\circ}$ C)	<i>F</i> , 340/450	OPA/MCE (60 $^{\circ}$ C)	0.25–2.5	<4.5	AAs in hydrolysates	18/~ 120
1993, Møller	15 $\times$ 3 + guard 2 $\times$ 3	5 5	Pickering, IE	A: 0.24 mol l $^{-1}$ Licitr., pH 2.27, B: 0.64 mol l $^{-1}$ Licitr., pH 7.50 (50 $^{\circ}$ C)	<i>F</i> , 340/448	OPA/MCE (4 $^{\circ}$ C)	0.6	<11	Physiological AAs	39/180

Continued

Table 1 Continued

Date	Column, cm × mm	Size ( $\mu\text{m}$ )	Type	Eluents (elution temperature in $^{\circ}\text{C}$ )	Detector UV (nm), F Ex/ Em	Reagent ( $^{\circ}\text{C}$ )	Analyte ( $\text{nmol l}^{-1}$ )	RSD (%)	Matrix	No. of AAs/ el. time (min)
1994, Saurina	15 × 4.6	5	Spherisorb ODS 2	A: 20 $\text{mmol l}^{-1}$ $\text{H}_3\text{PO}_4$ + 20 $\text{mmol l}^{-1}$ $\text{NaH}_2\text{PO}_4$ + 15 $\text{mmol l}^{-1}$ SDS; B: 25 $\text{mmol l}^{-1}$ $\text{H}_3\text{PO}_4$ + 25 $\text{mmol l}^{-1}$ $\text{NaH}_2\text{PO}_4$ + 18.5 $\text{mmol l}^{-1}$ SDS/PrOH (4:1, v/v) (50 $^{\circ}\text{C}$ )	UV 305	NQS (65 $^{\circ}\text{C}$ )	32	<5	AAs in food + feed <sup>a</sup>	18/105

<sup>a</sup> hydrolysates; Indications: – = no data available; IE = ion exchange resin; el. = elution. (Reproduced with permission from Vasanits A and Molnár-Peti I (2000) *Journal of Chromatography A* 870: 271–287; © Elsevier)

be superior, determining 99 compounds in 150 min (Figure 2A).

### Precolumn Derivatization

Precolumn derivatization offers numerous advantages (Tables 2 and 3, Figures 3–6). It requires less equipment and allows the evaluation of the derivatives in an easier way from the point of view of their selectivity, sensitivity, various means of detection, derivatization yield, stability, and storability. All of these phenomena can be controlled and improved by means of modern instrumental techniques and computerization, both individually and simultaneously. Potential disadvantages in precolumn derivatization procedures can be completely avoided, such as contamination from the reagents (due to their insufficient purity) and loss of analyte from incomplete interactions, undesirable side reactions, and sample handling losses.

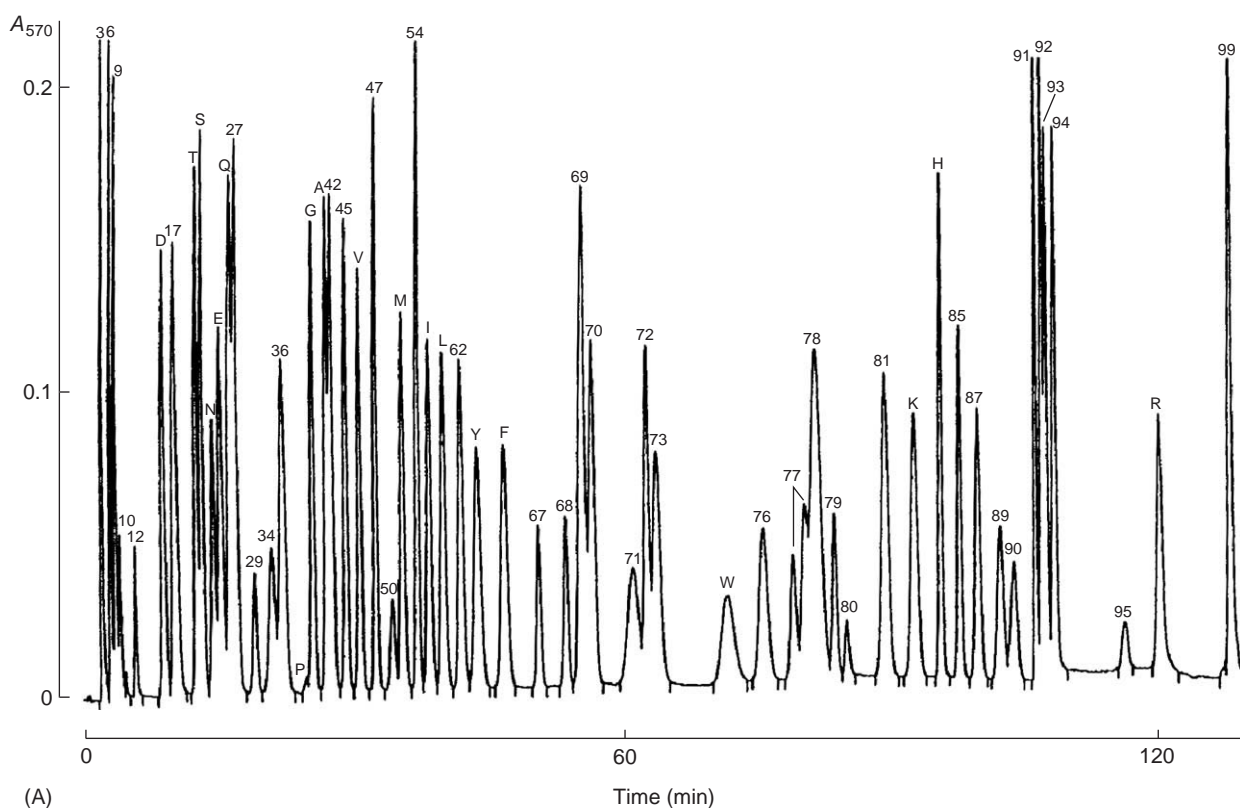
Although numerous precolumn derivatization techniques have been introduced in the last 35 years, none complies with the criteria of an ideal procedure: providing rapid and quantitative interaction in aqueous media, permitting mild conditions, ensuring interaction with both primary and secondary AAs, and resulting in single and stable derivatives in the case of all AAs.

### OPA Derivatives

The pioneering work of Roth (1971) on the very fast reaction of AAs in aqueous solutions with OPA/MCE detectable by both UV and FL, without the need to remove excess reagent, represented a great advance (Table 2, Figure 3).

Because of the different and low stability of the isoindoles obtained from the reaction of AAs with OPA/MCE reagent, alternative precolumn derivatizations reagents, such as 3-mercaptopropionic acid (MPA) and several *N*-alkyl-L/D-cysteines, were proposed. The OPA/MPA and OPA/*N*-acetyl-L-cysteine (NAC) reagents provide more stable isoindoles compared to those formed with the OPA/MCE, and the optical resolution of enantiomeric AAs with the OPA/NAC, as well as with further *N*-alkyl-L/D-cysteine reagents, have opened a new area in separation of AA enantiomers. Due to robotic auto-samplers, which provide excellent reproducibility even for moderately quantitative interactions, most AA analyses are performed with the OPA derivative. The unexplainable contradictions of this most popular process – relating to the particularly low stability of the OPA derivatives of six very important AAs (glycine,  $\gamma$ -aminobutyric acid (GABA),  $\beta$ -alanine,



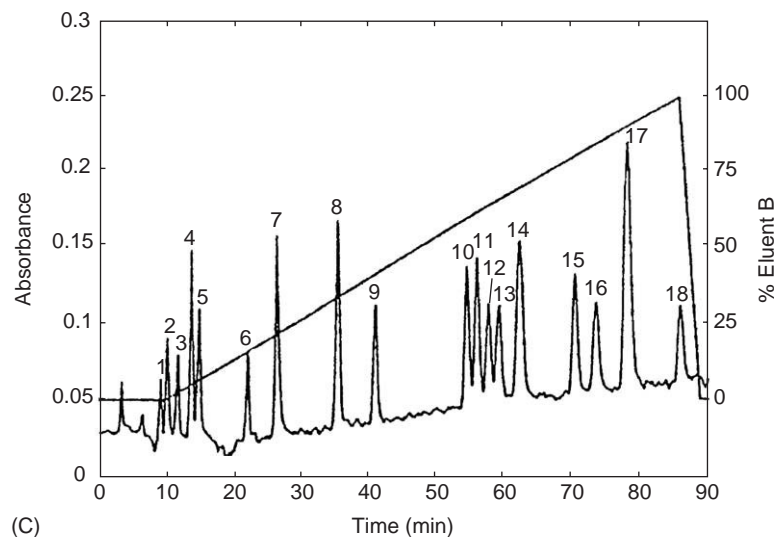
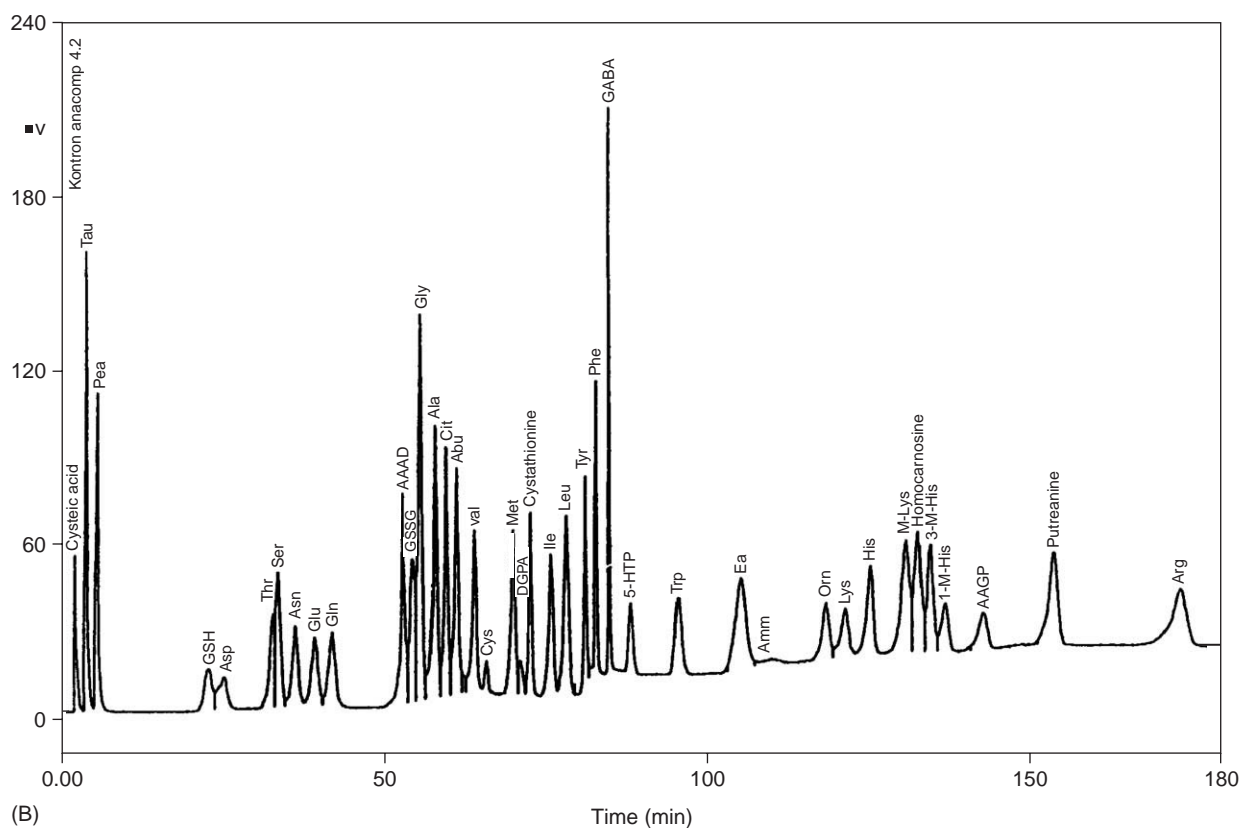


**Figure 2** (continued overleaf) LC of postcolumn derivatized AAs (details in **Table 1**). (A) Chromatographic profile of 99 AAs and related compounds as their ninhydrin derivatives; peaks: 3 = *O*-phospho-DL-serine, 6 = taurine, 9 = *O*-phosphoethanolamine, 10 = *N*<sup>2</sup>-(1-D-mannityl)-L-glutamine (mannopine), 11 = urea, 12 =  $\beta$ -ciano-L-alanine, D = L-aspartic acid, 17 = *O*-acetyl-L-serine, T = L-threonine, S = L-serine, N = L-asparagine, E = L-glutamic acid, Q = L-glutamine, 27 = L-homoserine, 29 = sarcosine, 34 = DL- $\alpha$ -aminoadipic acid, 36 = S-methyl-L-cysteine, P = L-proline, G = glycine, A = L-alanine, 42 = L-citrulline, 45 = L- $\alpha$ -aminobutyric acid, V = L-valine, 47 = L-cystine, 50 =  $\alpha$ -methyl-DL-methionine, M = L-methionine, 54 = L-cystathionine, I = L-isoleucine, L = L-leucine, 62 = L-norleucine, Y = L-tyrosine, F = L-phenylalanine, 67 =  $\beta$ -alanine, 68 = DL-aminoisobutyric acid, 69 = DL-homocystine, 70 =  $\delta$ -aminolevulinic acid, 71 = 5-hydroxy-L-tryptophan, 72 =  $\gamma$ -aminobutyric acid, 73 = DL-kynurenine, W = L-tryptophan, 76 = ethanolamine, 77 =  $\delta$ -hydroxylysines (DL- and DL-allo), 78 = ammonia, 79 =  $\epsilon$ -amino-*n*-caproic acid, 80 = creatinine, 81 = L-ornithine, K = L-lysine, H = L-histidine, 85 = 3-methyl-L-histidine, 87 = 1-methyl-L-histidine, 89 = L-carnosine, 90 = L-anserine, 91 = L-canavanine, 92 = S-methyl-DL-methionine, 93 = L- $\alpha$ -amino- $\beta$ -guanidinopropionic acid, 94 = L-leucinamide, 95 = *N*<sup>G1</sup>-dimethyl-L-arginine, R = L-arginine, 99 = L-homoarginine. (Reproduced with permission from Grunau JA and Swiader JM (1992) Chromatography of 99 amino acids and other ninhydrinreactive compounds in the Pickering lithium gradient system. *Journal of Chromatography* 594: 165–171; © Elsevier.) (B) Separation of OPA/MCE derivatives by gradient IEC chromatography. (Reproduced with permission from Møller SE (1993) *Journal of Chromatography* 613: 223; © Elsevier.) (C) Determination of AAs by ion-pair liquid chromatography with postcolumn derivatization using 1,2-naphthoquinone-4-sulfonate (NQS); peaks: 1 = Asp; 2 = Ser, 3 = Glu, 4 = Gly, 5 = Thr, 6 = Ala, 7 = Pro, 8 = Tyr, 9 = Met, 10 = Ile, 11 = Phe, 12 = Leu, 13 = Nle, 14 = Trp, 15 = His, 16 = Orn, 17 = Lys, 18 = Arg, Line = elution gradient profile. (Reproduced with permission from Saurina J and Hernandez-Cassou S (2002) *Journal of Chromatography A* 961: 9–21; © Elsevier.)

histidine, ornithine, and lysine) – have recently been studied and clarified by following two approaches.

First, stoichiometric investigations – based on simultaneous photodiode array (PDA) and FL detections – revealed that the above-selected AAs provided more than one OPA derivative. Thus, taking into account that their initially formed derivatives are transformed, these six AAs provide acceptable stability: to achieve their reproducible quantitation with the same molar responses as other AAs, only a proper elution protocol is needed and they can be quantitated on the basis of the total on their derivatives (Figures 3A and 3B).

Second, structure of the transformed OPA derivatives (containing two OPA molecules) and the reaction mechanism leading to the transformation are determined by online high-performance liquid chromatography (HPLC)–UV–MS evidences. It has been confirmed that the common characteristics of the presumably less stable OPA derivatives are associated with their original molecular structure: they contain the  $\text{NH}_2\text{--CH}_2\text{--R}$  moiety, and all of them furnish more than one OPA derivative. In the light of the common structure of the multiple OPA derivatives providing compounds, a plausible reaction pathway has been introduced, fully supported by



**Figure 2** Continued

MS fragmentation patterns of all; in their initial structure the  $\text{NH}_2\text{-CH}_2\text{-R}$  moiety containing primary amino compounds.

The essential shortage of an OPA/S<sub>H</sub> group reagent (reactive toward primary AAs only) can be eliminated by Shuster's principle – the automatic two-step precolumn derivatization method – by applying the OPA/MPA/fluorenylmethyl-chloroformate

(FMOC) reagent, which also ensures derivatization of the secondary AAs. A high-speed elution of OPA/MPA/FMOC derivatives has been observed recently (Table 2, Figure 3B: 19 compounds/8 min). In the new improved methods, shorter, thermostated columns of smaller particle size with autosamplers are used, which provide greater sensitivity and reproducibility.

**Table 2** Advances in the LC of precolumn derivatized AAs, obtained with OPA/MCE, OPA/ethanethiol, OPA/mercaptopropionic acid, OPA/*N*-acetyl-L-cysteine, with OPA/*N*-isobutyryl-L-cysteine, and with OPA/MPA/fluorenylmethylchloroformate (OPA/ET; OPA/MPA; OPA/NAC; OPA/NIBC; OPA/MPA/FMOC)

Date	Column (cm × mm)	Size (μm)	Type	Eluents (elution temperature °C)	Detector UV (nm), <i>F</i> , Ex/Em	Reagent (°C)	Analyte (pmol l <sup>-1</sup> )	RSD (%)	Matrix	No. of AAs/ el. time (min)
1983, Jones	75 × 4.6 + guard 4.5 × 2.1	3 40	Ultrasphere ODS	A: THF/CH <sub>3</sub> OH/NaAc (pH 7.2) (5:95:900, v/v); B: CH <sub>3</sub> OH (–)	<i>F</i> , 305–395, 420–650	OPA/MCE (–)	0.1–80	< 1.5	AAs in hydrolysates	48/50
1995, Fekkes	12.5 × 4.6	5	Spherisorb ODS-2	A (pH 6.72–6.77) and B (pH 5.95–6.00): 250 mmol l <sup>-1</sup> Na <sub>2</sub> HPO <sub>4</sub> /250 mmol l <sup>-1</sup> propionic acid/ACN/THF/ H <sub>2</sub> O = (20:20:7:2:51, v/v); C: ACN/CH <sub>3</sub> OH/DMSO/ H <sub>2</sub> O = (28:24:5:43, v/v) (25–35 °C)	<i>F</i> , 337/452	OPA/MCE (3 °C)	50	< 2	Plasma AAs	40/49
1979, Hill	30 × 3.9	–	μ-Bondapak, C-18	A: 12.5 mmol l <sup>-1</sup> Na <sub>2</sub> HPO <sub>4</sub> (pH 7.2), B: A eluent/ACN in gradient (–)	<i>F</i> , 229/470	OPA/ET (–)	5	–	AAs in human serum	20/40
1987, Eslami	50 × 4.5	3	ODS IBM	Buffer: ~ 2 mol l <sup>-1</sup> Na <sub>2</sub> HPO <sub>4</sub> (pH 7), A: ACN/H <sub>2</sub> O/ buffer = (50:425:25, v/v); B: ACN/H <sub>2</sub> O/ buffer = (275:200:25, v/v) (22 °C)	<i>F</i> , 330/480	OPA/ET (–)	40–100	–	Model study	22/14
1984, Godel	25 × 4	4	Supersphere CH-8	A: 12.5 mmol l <sup>-1</sup> Na <sub>2</sub> HPO <sub>4</sub> (pH 7.2), B: 12.5 mmol l <sup>-1</sup> Na <sub>2</sub> HPO <sub>4</sub> (pH 7.2)/ACN (1:1, v/v) (–)	<i>F</i> , 330/445	OPA/MPA (–)	1–10	< 4.2	AAs in biological fluid	28/40
1993, van Eijk	15 × 4.6 + guard 1 × 4	2–3	Spherisorb ODS-2	A: 12.5 mmol l <sup>-1</sup> Na <sub>2</sub> HPO <sub>4</sub> (pH 7.0) to 7 ml THF/11, eluent B: 12.5 mmol l <sup>-1</sup> Na <sub>2</sub> HPO <sub>4</sub> (pH 7.0)/ACN/ THF = (57:43:7, v/v) (35 °C)	<i>F</i> , 335/440	OPA/MPA (–)	35	< 3	Plasma AAs	30/28
1994, Teerlink	10 × 4.6 + guard 1 × 2	3	Microsphere ODS	A: 4.5 mmol l <sup>-1</sup> K <sub>2</sub> HPO <sub>4</sub> (pH 6.9) + 2 ml THF/1 l, B: 4.5 mmol l <sup>-1</sup> K <sub>2</sub> HPO <sub>4</sub> (pH 6.9)/CH <sub>3</sub> OH/ ACN = (50:35:15, v/v) (–)	<i>F</i> , 230/389	OPA/MPA (–)	100	< 3.2	Plasma AAs	25/17

Continued

**Table 2** Continued

Date	Column (cm × mm)	Size (μm)	Type	Eluents (elution temperature °C)	Detector UV (nm), F, Ex/Em	Reagent (°C)	Analyte (pmol l <sup>-1</sup> )	RSD (%)	Matrix	No. of AAs/ el. time (min)
2000, Vasanits	15 × 4.0 + guard 2 × 4.0	5	Hypersil ODS	A: 50 mmol l <sup>-1</sup> NaAc cont. 1% THF (pH 7.2); B: ACN/CH <sub>3</sub> OH/ 0.1 mmol l <sup>-1</sup> NaAc = (46:44:10, v/v, pH 7.2); (OPA/MPA: 50°C); (OPA/NAC: 30°C)	UV 334; F, 337/434	OPA/MPA (50°C), OPA/NAC (30°C)	6–12\~000	< 3.4	Model study; apple AAs	31/40
1989, Schuster	20 × 2.1 20 × 4.6	5 5	Hypersil ODS	Protein hydrolysates, A: 30 mmol l <sup>-1</sup> NaAc cont. 0.5% THF (pH 7.2); B: ACN/0.1 mol l <sup>-1</sup> NaAc = (4:1, v/v); (42°C); Plasma AAs, A: 60 mmol l <sup>-1</sup> NaAc cont. 0.6% THF (pH 8.0); B: ACN/0.1 mol l <sup>-1</sup> NaAc/ CH <sub>3</sub> OH = (14:4:1, v/v) (43°C)	UV 338/266; F, 230/ 455, 266/ 310	OPA/MPA/ FMOC (4°C)	UV 2–5; F, 0.02–0.05	< 2–5	AAs in protein hydrolysates plasma AAs	19/20, 38/60
1994, Bartók	10 × 4	3	Hypersil ODS	A: 18 mmol l <sup>-1</sup> NaAc (pH 7.2) + 0.02% (v/v) TEA + 0.3% THF (v/v); B: ACN/CH <sub>3</sub> OH/NaAc 0.1 mol l <sup>-1</sup> (pH 7.2) = (2:2:1, v/v) (40°C)	F, 340/450, 264/313	OPA/MPA/ FMOC (4°C)	50	< 1.1	Plant AAs	21/8

Indications: — = no data available; el. = elution; cont. = containing. (Reproduced with permission from Vasanits A and Molnár-Perl I (2000) *Journal of Chromatography A* 870: 271–287; © Elsevier.)

**Table 3** Advances in the LC of precolumn derivatized AAs, obtained with phenylisothiocyanate (PITC), 5-dimethylaminonaphtalene-1-sulfonyl-Cl (DANS), 4-dimethylaminoazobenzene-4-sulfonyl-Cl (DABS), and with 9-fluorenylmethyl chloroformate (FMOC)

Date	Column, cm × mm	Size (μm)	Type	Eluents (elution temperature °C)	Detector UV (nm), F Ex/Em	Reagent (°C)	Analyte (pmol l <sup>-1</sup> )	RSD (%)	Matrix	No. of AAs/el. time (min)
1982, Koop	25 × 4.6	5	Ultrasphere ODS	A: 70 mmol l <sup>-1</sup> NaH <sub>2</sub> PO <sub>4</sub> (adjusted to pH 6.45 with TEA), B: ACN (27°C)	UV 254	PITC (–)	6000	–	AAs in protein hydrolysates	18/130
1995, Toso Haas	10 × 4.6	2	TSKgel Super- ODS	A: 50 mmol l <sup>-1</sup> NaAc (pH 6.0)/ ACN = (97:3, v/v), B: 50 mmol l <sup>-1</sup> NaAc (pH 6.0)/ ACN = (40:60, v/v) (40°C)	UV 254	PITC (–)	250	–	Model study	17/4.5
1996, Shang	15 × 3.9	5	PicoTag ODS	A: NaAc(pH 6.4), B: ACN; A and B performed in gradient (38°C)	UV 254	PITC (–)	5	<1.9	AAs in kelp	17/12
1976, Bayer	50 × 3	10	LiChrosorb, RP 8	Eluent: 10 mmol l <sup>-1</sup> Na <sub>2</sub> HPO <sub>4</sub> /CH <sub>3</sub> OH = (50:20, v/v) to which 1.5 ml CH <sub>3</sub> OH/min is added (45°C)	F, 340/510	DANS (amb)	0.1	–	Model study	17/40
1996, Martins	15 × 3.9	4	Nova Pak C 18	A: 30 mmol l <sup>-1</sup> phosphate buffer(pH 7.4) + 5 ml CH <sub>3</sub> OH + 6.5 ml THF adjusted to 100 ml with distilled water; B: CH <sub>3</sub> OH/ H <sub>2</sub> O = (70/30, v/v) (25°C)	F, 338/445	DANS (40°C)	60	–	AAs in polypeptides	17/35
1983, Chang	–	5	–	A: 25 mmol l <sup>-1</sup> NaAc (pH 6.5) containing 4% dimethylformamide; B: ACN (40°C)	UV 436	DABS (70°C)	5	–	AAs in protein hydrol-sates	17/40
1993, Yang	15 × 4.6	5	Hypersil ODS	A: 25 mmol l <sup>-1</sup> NaAc (pH 6.35) containing 4% dimethylformamide; B: CAN (40°C)	UV 436, 580	DABS (70°C)	50	–	AAs in polypeptides	17/40
1983, Einarsson	50 × 4.6 500 × 2.26	3 5	Sphereisorb ODS-2	Eluent: 20 mmol l <sup>-1</sup> NaAc buffer (pH 4.08–4.31)/ACN gradient (–)	F, 265/315	FMOC (–)	(–)	<6.6	AAs in protein hydrolysates, and in urine	17/10 and 33/ 100
1996, Qu	15 × 4.6		Hypersil ODS	A: 30 mmol l <sup>-1</sup> phosphate buffer (pH 6.5) in 15% CH <sub>3</sub> OH (v/v); B: 15% CH <sub>3</sub> OH (v/v); C: 90% ACN (v/v) (38°C)	F, 270/316	FMOC (–)	125	<1.0	AAs in protein hydrolysates, biol. samples	15/35

Continued



Table 3 Continued

Date	Column, cm × mm	Size ( $\mu\text{m}$ )	Type	Eluents (elution temperature °C)	Detector UV (nm), F Ex/Em	Reagent (°C)	Analyte ( $\mu\text{mol l}^{-1}$ )	RSD (%)	Matrix	No. of AAs/el. time (min)
1996, Bank	15 × 4.6	5	Micropak ODS- 80TM	A: 20 mmol l <sup>-1</sup> citric acid NaAc buffer (pH 2.85); B: 20 mmol l <sup>-1</sup> NaAc (pH 4.5)/ CH <sub>3</sub> OH = (80:20, v/v); A and B both, cont. 0.01% (w/v) NaNa <sub>3</sub> + 5 mmol l <sup>-1</sup> (CH <sub>3</sub> ) <sub>4</sub> NCl; C: ACN (40°C)	F, 254/630	FMOC (-)	50	<3.6	AAs in protein hydrolysates	21/35

Indications: - = no data available; el. = elution; biol. = biological. (Reproduced with permission from Vasanitis A and Molnár-Pertl I (2000) *Journal of Chromatography A* 870: 271–287; © Elsevier.)

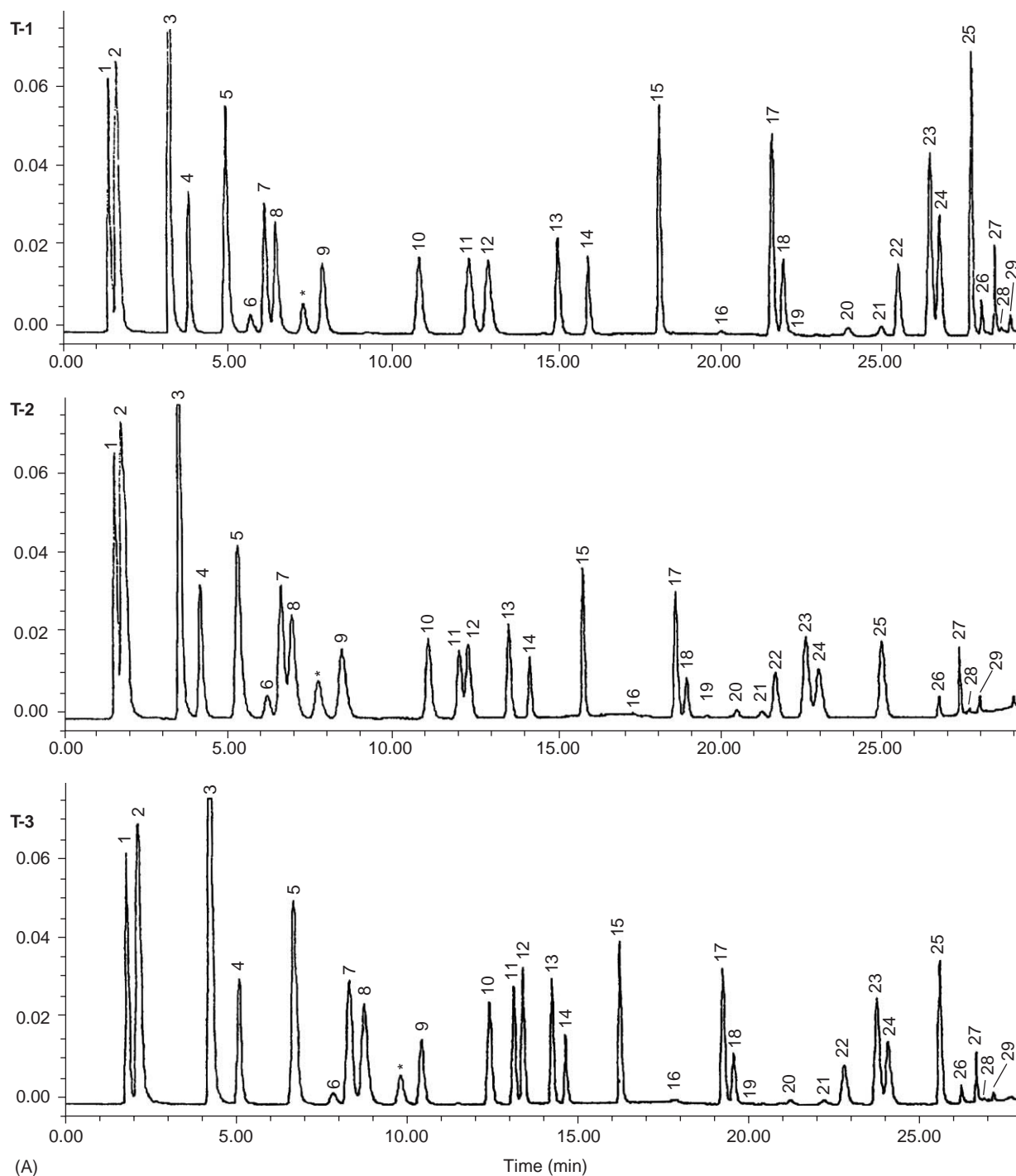
### Phenylthiocarbamyl, FMOC, 1-*N,N'*-Dimethylaminonaphthalene-5-sulfonyl, and Dimethylaminoazobenzenesulfonyl Derivatives

Judging by the number of publications in the last decade, the interest in Phenylthiocarbamyl (PTC) and FMOC derivatives has proved to be lasting, that in OPA derivatives is increasing, while the application of Dimethylaminonaphthalene-5-sulfonyl (DANS) and Dimethylaminoazobenzenesulfonyl (DABS) derivatives is decreasing. However, in the direct enantiomer separation of AAs, the use of the DANS derivatives is preferred.

The reaction of AAs with phenylisothiocyanate (Table 3, Figures 1, 4, 5), in water-free media and at ambient temperature, is quantitative and fast (10 min), resulting in the highly stable single PTC derivatives (except for cyst(e)ines in hydrolysates that elute in one to four peaks). The excess of the reagent is removed by vacuum, and the PTC derivatives can be stored in the freezer for long periods of time, and for a day after dissolution in buffer at 4°C. UV detection at 254 nm allows their quantitation. The short TSK gel (Figure 1) and the short PicoTag (Figure 4) column can separate 17 AAs within 4.5 and 12 min, respectively, while applying a multistep gradient provides the separation of 27 AAs, including the hardly separable GABA, histidine, threonine, alanine, and ACPA (Figure 5).

The first LC separation of the strongly fluorescent DANS AAs has been used earlier in protein chemistry and in thin-layer chromatography. The decreased popularity of this technique in LC can be explained by its main disadvantages: long reaction times, elevated temperatures for derivatization, generation of FL side products (DANS hydroxide, DANS amide), and interference from excess reagent. The disturbing effect of these compounds cannot be completely eliminated, and they elute between the AA derivatives. No significant improvement has been obtained and is not expected. DABS AAs were first separated by applying precolumn labeling. Derivatization was performed using Na<sub>2</sub>CO<sub>3</sub>/NaHCO<sub>3</sub> buffer, at pH ~8.9, with DABS chloride dissolved in acetone under continuous stirring at 80°C for 10 min. DABS AAs can be stored for 4 weeks in solution at 25°C, without any changes. In spite of the unique stability of DABS AAs in aqueous media and the improvement in their chromatographic conditions, the use of the DABS AAs is dwindling.

FMOC was introduced in 1983 as a fluorescent labeling agent, reacting rapidly with both primary and secondary AAs under mild conditions (borate buffer, pH 7.7–8.0) to give stable derivatives. The excess reagent is extracted by pentane. Recent derivatization studies have shown that, depending



**Figure 3** (continued overleaf) (A) Retention quantitation properties of the *o*-phthalaldehyde-3-mercaptopropionic acid amino acid derivatives in reversed-phase high-performance liquid chromatography; separation of 25 + 4 OPA/MPA-AAs (columns: T-1–T-3; T-1, 150 + (20 guard)  $\times$  4 mm, C18 Hypersil 5  $\mu$ m; T-3, 200 + (20 guard)  $\times$  4 mm;  $T=50^{\circ}\text{C}$ , eluent A: 0.05 mol l $^{-1}$  NaAc, pH = 7.2, B: A eluent/acetonitrile/methanol = 46/44/10 (pH 7.2, containing 1% (v/v) THF), flow rate: between 0 and 9 min, 1.3 ml min $^{-1}$ , from 9 to 40 min, 2.3 ml min $^{-1}$ ; Peaks: 1 = aspartic, 2 = glutamic acids, 3 = asparagine, 4 = serine, 5 = glutamine, 6 = histidine, 7 = glycine, 8 = homoserine, 9 = threonine, 10 =  $\beta$ -alanine, 11 = arginine, 12 = alanine, 13 = GABA, 14 = homoarginine, 15 = tyrosine, 16 = glycine2, 17 = valine, 18 = methionine, 19 = cyst(e)ine, 20 =  $\beta$ -alanine2, 21 = GABA2, 22 = tryptophan, 23 = henylalanine, 24 = *i*-leucine, 25 = *n*-leucine, 26 = ornithine2, 27 = lysine2, 28 = ornithine3, 29 = lysine3; \* = system peaks. (Reproduced with permission from Vasani A *et al.* (2000) *Journal of Chromatography A* 870: 271; © Elsevier.) (B) High-speed RP-HPLC analysis of the OPA/MPA/FMOC derivatives. (Reproduced with permission from Bartók T *et al.* (1994) *Journal of Liquid Chromatography* 17: 4391; © Marcel Dekker, Inc.)

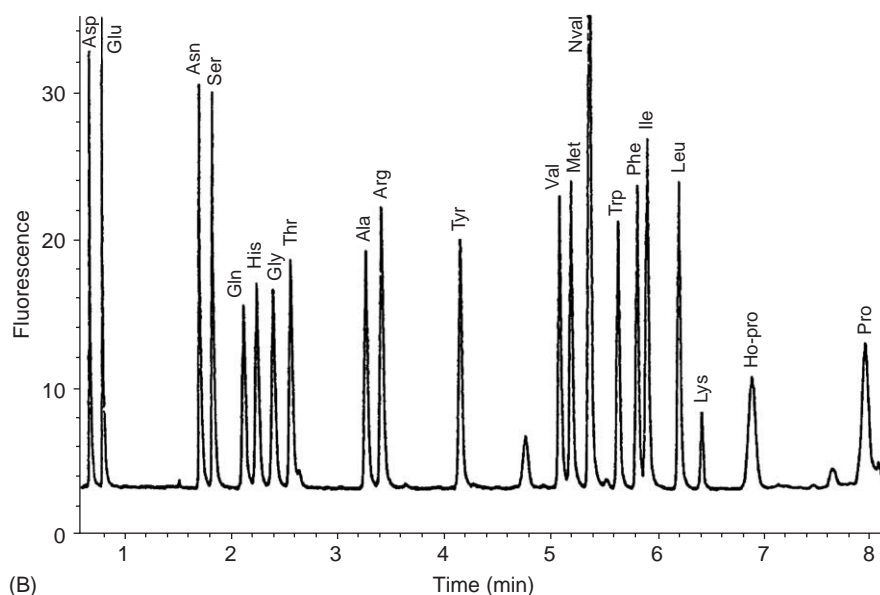
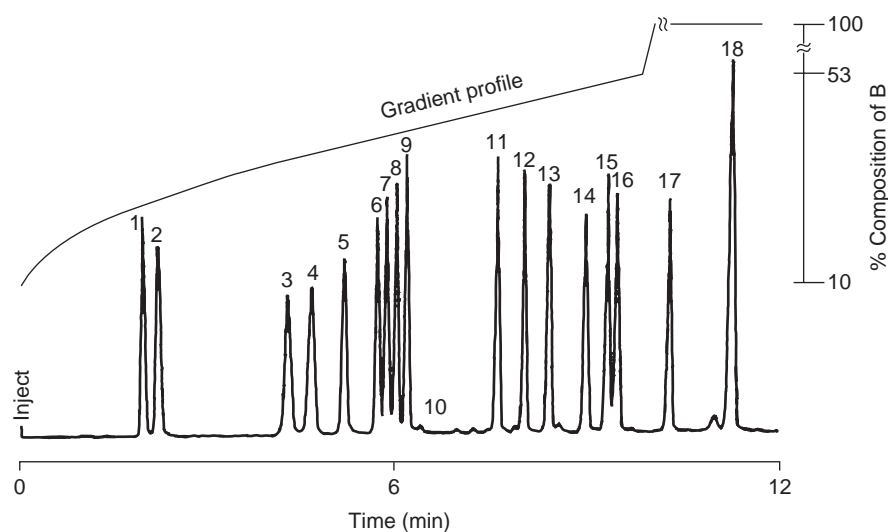


Figure 3 Continued

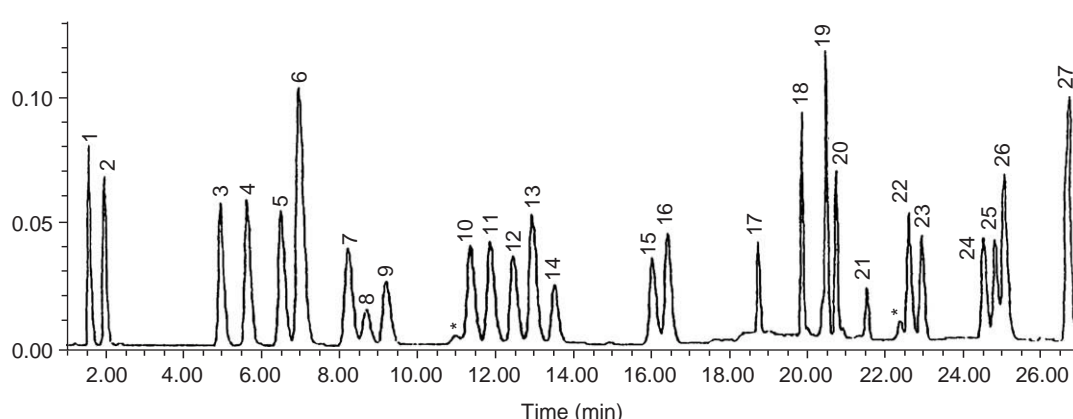


**Figure 4** HPLC separation of 17 PTC AAs (Pico-Tag column  $150 \times 3.9$  mm, eluent A:  $0.14 \text{ mol l}^{-1}$  NaAc, pH 6.4, containing  $0.5 \text{ ml TEAL}^{-1}$ , B: 60% acetonitrile in water; flow rate:  $1 \text{ ml min}^{-1}$ ; peaks: 1 = aspartic, 2 = glutamic acids, 3 = serine, 4 = glycine, 5 = histidine, 6 = arginine, 7 = threonine, 8 = alanine, 9 = proline, 10 = ammonia, 11 = tyrosine, 12 = valine, 13 = methionine, 14 = cystine, 15 = isoleucine, 16 = *n*-leucine, 17 = phenylalanine, 18 = tryptophan. (Reproduced with permission from Bidlingmayer BA *et al.* (1984) *Journal of Chromatography* 336; © Elsevier.)

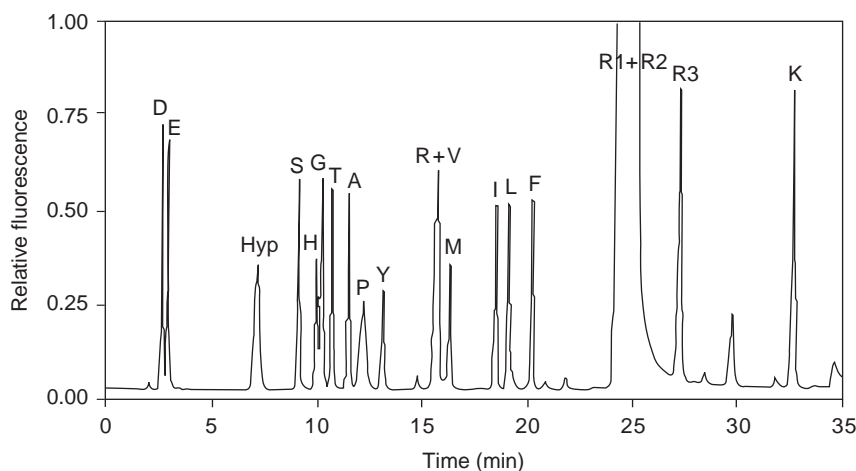
on the time (2 and 40 min) and pH (8.0 and 11.4), considerable differences can be found. At pH  $\sim 8$  acidic AAs manifest low responses, and a slow reaction was experienced; histidine and tyrosine yield their mono- and disubstituted derivatives in varying ratios. With longer reaction times, the amount of disubstituted histidine decreases and that of tyrosine increases, together with interfering hydrolysis products of the reagent. At pH  $\sim 11.4$ , faster reaction and less interfering hydrolysis products were found. After 40 min reaction time, the monosubstituted histidine

and the disubstituted tyrosine are formed in quantitative yield. Also, 30% less hydrolysis product was obtained, favoring the resolution of the neighboring alanine. The separation of the Fmoc derivatives and the presence of interfering substances are shown in Figure 6.

**See also:** Derivatization of Analytes. **Fluorescence:** Derivatization. **Ion Exchange:** Overview. **Liquid Chromatography:** Clinical Applications. **Proteins:** Traditional Methods of Sequence Determination.



**Figure 5** Separation of 27 PTC AAs (column 150 + (20 guard)  $\times$  4 mm, C18 Hypersil 5  $\mu$ m,  $T = 50^\circ\text{C}$ , eluent A: 0.05 mol l $^{-1}$  NaAc, pH 7.2, B: A eluent/acetonitrile/methanol = 46/44/10 (pH 7.2); flow rate: 2.1 ml min $^{-1}$ ; peaks; 1 = aspartic, 2 = glutamic acids, 3 = hydroxyproline, 4 = serine, 5 = glycine, 6 = asparagine, 7 =  $\beta$ -alanine, 8 = glutamine, 9 = homoserine, 10 =  $\gamma$ -aminobutyric acid (GABA), 11 = histidine, 12 = threonine, 13 = alanine, 14 = 1-amino-1-cyclopropane carboxylic acid (ACPCA), 15 = arginine, 16 = proline, 17 = homoarginine, 18 = tyrosine, 19 = valine, 20 = methionine, 21 = cyst(e)ine, 22 = isoleucine, 23 = *n*-leucine, 24 = phenylalanine, 25 = tryptophan, 26 = ornithine, 27 = lysine; \* = system peaks. (Reproduced with permission from Vasánits A and Molnár-Perl I (2000) *Journal of Chromatography A* 870: 271–287; © Elsevier.)



**Figure 6** (details in Table 3) HPLC of AAs derivatized with 9-fluorenylmethyl chloroformate (FMOC). Peaks labeled with one letter abbreviations for protein AAs, as well as: Hyp = hydroxyproline, R1 = FMOC-hydroxylamine, R2 = FMOC-hydroxide, R3 = reagent peak present in blank derivatization. (Reproduced with permission from Qu K *et al.* (1996) *Journal of Chromatography A* 723: 219; © Elsevier.)

## Further Reading

- Blau K and Halket (eds.) (1993) *Handbook of Derivatives for Chromatography*. Chichester: Wiley.
- Lunn G and Hellwig LC (1998) Amino acids. In: Lunn G and Hellwig LC (eds.) *Handbook of Derivatization Reactions for HPLC*, pp. 625–821. New York: Wiley.
- Mengerink Y, Kutlán D, Tóth F, Csámpai A, and Molnár-Perl I (2002) Advances in the evaluation of the stability and characteristics of the amino acid/amine derivatives obtained with the *o*-phthaldialdehyde (OPA)/3-mercaptopropionic acid and OPA/*N*-acetyl-L-cysteine reagents: HPLC/MS study. *Journal of Chromatography A* 949: 99–124.
- Molnár-Perl I (2001) Derivatization and chromatographic behavior of the *o*-phthaldialdehyde amino acid derivatives obtained with various SH-group-containing additives. *Journal of Chromatography A* 913: 283–302.
- Spackman DH, Stein WH, and Moore S (1958) Automatic recording apparatus for use in the chromatography of amino acids. *Analytical Chemistry* 30: 1190–1206.
- Vasánits A and Molnár-Perl I (1999) Temperature, eluent flow-rate and column effects on the retention and quantitation properties of phenylthiocarbamyl derivatives of amino acids in reverse-phase high-performance liquid chromatography. *Journal of Chromatography A* 832: 109–122.
- Zhou J, Hefta S, and Lee TD (1997) High sensitivity analysis of phenylthiohydantoin amino acid derivatives by electrospray mass spectrometry. *Journal of the American Chemical Society of Mass Spectrometry* 8: 1165–1174.

## Chiral Analysis of Amino Acids

**J Haginaka**, Mukogawa Women's University,  
Nishinomiya, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

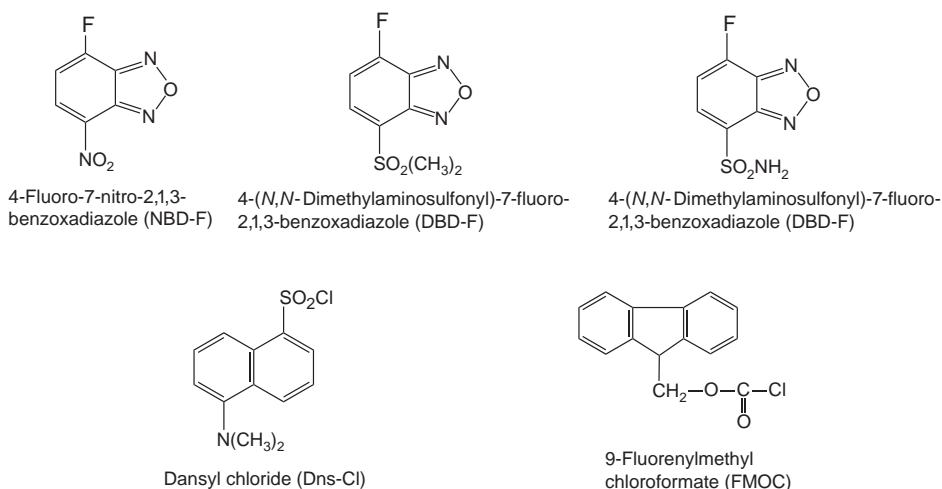
### Introduction

Methods for the chiral separation of amino acids can be divided into two categories: one is an indirect method based on the formation of diastereomers by the reaction of amino acids with a chiral derivatization reagent and separation of the diastereomeric derivatives on an achiral stationary phase. The other is a direct method based on the formation of diastereomers on a stationary phase or in a mobile phase, where the former uses a chiral stationary phase (CSP) and the latter adds a chiral selector in a mobile phase and uses an achiral stationary phase.

The indirect method is an efficient technique for the enantioseparation of amino acids. However, it is essential that the chiral derivatization reaction proceeds completely in both enantiomers, and that the racemization reaction does not occur. Furthermore, if the optical purity of the derivatization reagent is not known, and/or is not taken into consideration, the optical purity of the amino acid will not be determined precisely. Thus, the indirect method is unsuitable for the analysis of amino acid enantiomers in a standard sample and pharmaceutical preparations, where a low amount of antipode, at a level of 0.1 or 0.05%, should be determined.

Furthermore, it is also unsuitable for preparative purposes. However, it is suitable for the trace analysis of amino acid enantiomers in complex matrices such as biological samples because of the introduction of a highly sensitive ultraviolet-visible or fluorescence tag.

The direct method based on CSPs is divided into two categories: one is the direct enantioseparation of underivatized amino acids on a CSP and the other is the enantioseparation on a CSP following derivatization with an achiral reagent (**Figure 1**). At present, both methods are complementarily used for the enantioseparation of amino acids by liquid chromatography. The direct method, using a CSP, is suitable for the analysis of underivatized amino acid enantiomers in a standard sample and pharmaceutical preparations, where low amounts of the antipode should be determined. Furthermore, it is suitable for preparative purposes. For trace analysis of the amino acid enantiomers, especially D-forms, in complex matrices such as biological samples, derivatization with an achiral reagent combined with liquid-liquid extraction, solid-phase extraction, or column switching technique is followed by the enantioseparation on a CSP. However, in this case it is important that the derivatization reaction proceeds completely in both enantiomers and that the racemization reaction does not occur. On the other hand, the direct method using chiral mobile phase additives has the same drawback with regard to the optical purity of chiral selectors. This article deals with two chiral separation methods of amino



**Figure 1** Achiral derivatization reagents used for the derivatization of amino acids followed by the enantioseparation on chiral stationary phases.



acid enantiomers by liquid chromatography: indirect and direct methods.

## Indirect Method

Many chiral derivatization reagents have been developed for the enantioseparation of amino acids wherein ultraviolet-visible or fluorescence tags are introduced. The fluorescence derivatization is more effective for the determination of amino acid enantiomers in complex matrices in terms of sensitivity and/or selectivity. **Table 1** shows the chiral derivatization reagents, whose structures are shown in **Figure 2**, used for the enantioseparation of amino acids.

Amino acids have two functional groups, amino and carboxylic acid groups, which react with chiral derivatization reagents. Many chiral derivatization reagents that react with primary and/or secondary amino groups have been developed. It is well known that primary amino acids react with *o*-phthalaldehyde (OPA) in the presence of an achiral sulfhydryl reagent such as 2-mercaptoethanol or ethanethiol and yield the fluorescent products. The reaction of amino acids with a chiral sulfhydryl reagent such as *N*-acetyl-L-cysteine, *N*-butyloxycarbonyl-L-cysteine, *N*-acetyl-D-penicillamine, or 2,3,4,6-tetra-*O*-acetyl-1-thio- $\beta$ -glucopyranoside, instead of the achiral sulfhydryl reagent gives the corresponding diastereomeric amino acids, which are separated on an achiral, C8, or C18 column. The disadvantage of this reagent is that the secondary amino acids do not react and that the reaction products have limited stability. On the other hand, the reaction of 1-(9-fluorenyl)ethyl chloroformate (FLEC) with primary and secondary amino acids gives stable fluorescent products at room temperature within 4 min. **Figure 3**

shows the enantioseparation of 17 primary amino acids derivatized with FLEC on a C18 column, where a linear gradient with tetrahydrofuran is employed to facilitate the separation of derivatized asparagine and glutamine.

Many chiral derivatization reagents based on an isothiocyanate group have been developed for the enantioseparation of primary and secondary amino acids. These include 2,3,4,6-tetra-*O*-acetyl- $\beta$ -D-glucopyranosyl-isothiocyanate (GITC), 4-(3-isothiocyanatopyrrolidin-1-yl)-7-nitro-2,1,3-benzoxadiazole (NBD-PyNCS), 4-(3-isothiocyanatopyrrolidin-1-yl)-7-(*N,N*-dimethylaminosulfonyl)-2,1,3-benzoxadiazole (DBD-PyNCS), 1,3-diacetoxy-1-(4-nitrophenyl)-2-propylisothiocyanate (DANI), and *N*-[(2-isothiocyanato)cyclohexyl]-6-methoxy-4-quinolinylamide (CDITC). The derivatization of amino acids with the corresponding isothiocyanates yielded highly fluorescent thioureas. These derivatives were separated on an achiral C18 column.

The chiral derivatization reagents bearing an amino group include 1-(4-dimethylamino-1-naphthyl)ethylamine (DANE), 4-(3-aminopyrrolidin-1-yl)-7-nitro-2,1,3-benzoxadiazole (NBD-APy), and 4-(3-aminopyrrolidin-1-yl)-7-(*N,N*-dimethylaminosulfonyl)-2,1,3-benzoxadiazole (DBD-APy). The derivatization of amino acids with these reagents yielded the corresponding amide derivatives, and the diastereomeric amide derivatives produced were separated by normal or reversed-phase chromatography.

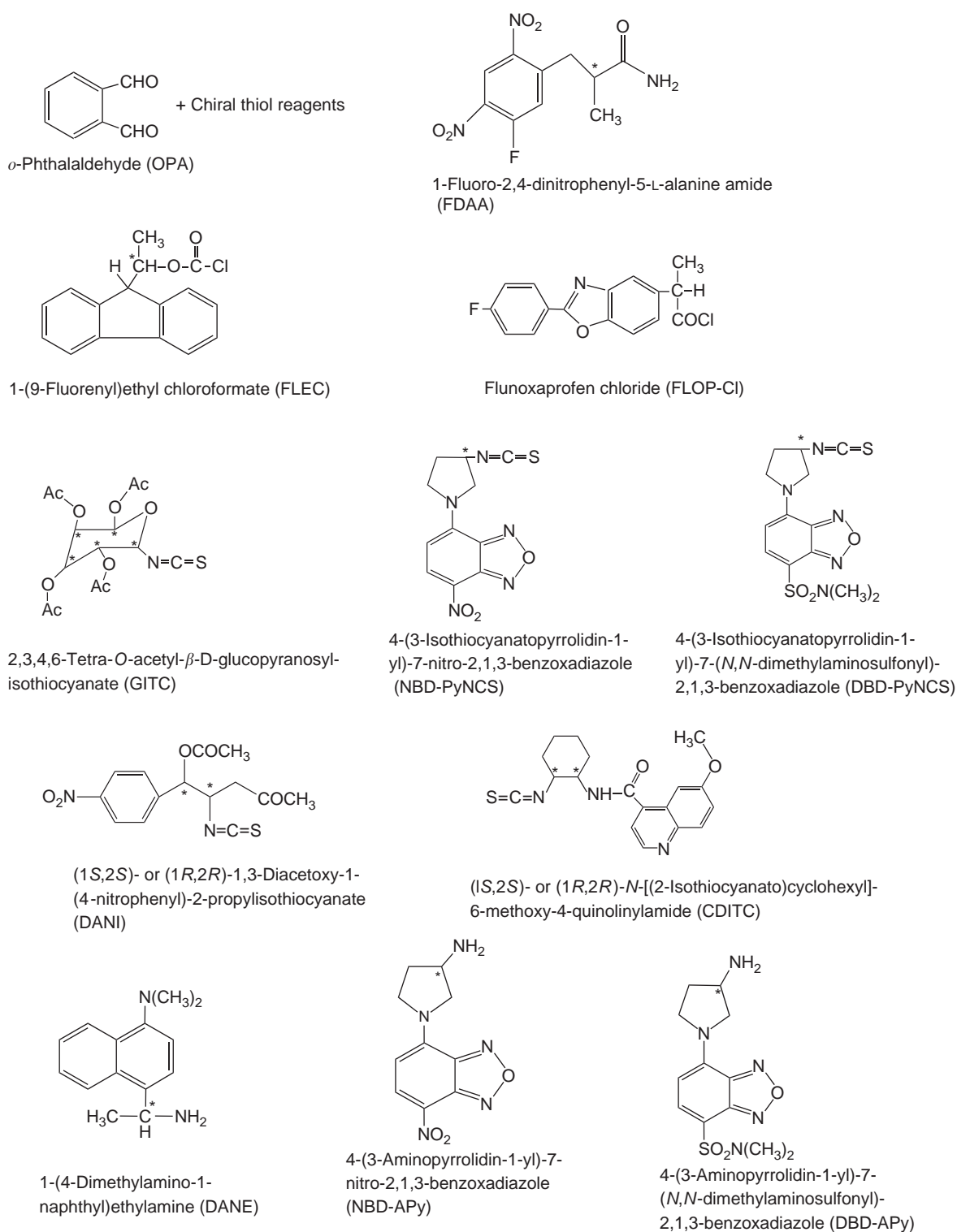
## Direct Method

### Chiral Stationary Phases

The CSPs based on aromatic  $\pi$ -acid (3,5-dinitrobenzene) and  $\pi$ -base ( $\alpha$ -naphthalene moiety) derivatives are called Pirkle-type stationary phases. These CSPs cannot separate underivatized amino acid enantiomers. However, after derivatization of amino acids with a fluorogenic benzofurazan reagent such as 4-fluoro-7-nitro-2,1,3-benzoxadiazole (NBD-F), 4-(*N,N*-dimethylaminosulfonyl)-7-fluoro-2,1,3-benzoxadiazole (DBD-F), or 4-aminosulfonyl-7-fluoro-2,1,3-benzoxadiazole (ABD-F), the resulting products were separated on Pirkle-type stationary phases in normal or reversed-phase mode. D-Amino acids in biological samples were determined by derivatization with NBD-F, enantioseparation and fluorometric detection following deproteinization of serum or brain homogenates with methanol. Also, column switching techniques using C18 and Pirkle-type columns were developed for the determination of D-amino acids in biological fluids. The former was used for the separation of the individual amino acids

**Table 1** Chiral derivatization reagents used for the enantioseparation of amino acids

Reagent	Reactive functional group	Separation mode
OPA + <i>N</i> -protected thiol	CHO + SH	Reversed phase
FDAA	F	Reversed phase
FLEC	COCl	Reversed phase
FLOP-CI	COCl	Reversed phase and normal phase
GITC	NCS	Reversed phase
NBD-PyNCS	NCS	Reversed phase
NBD-PyNCS	NCS	Reversed phase
DANI	NCS	Reversed phase
CDITC	NCS	Reversed phase
DANE	NH <sub>2</sub>	Normal phase
NBD-Apy	NH <sub>2</sub>	Reversed phase
DBD-Apy	NH <sub>2</sub>	Reversed phase

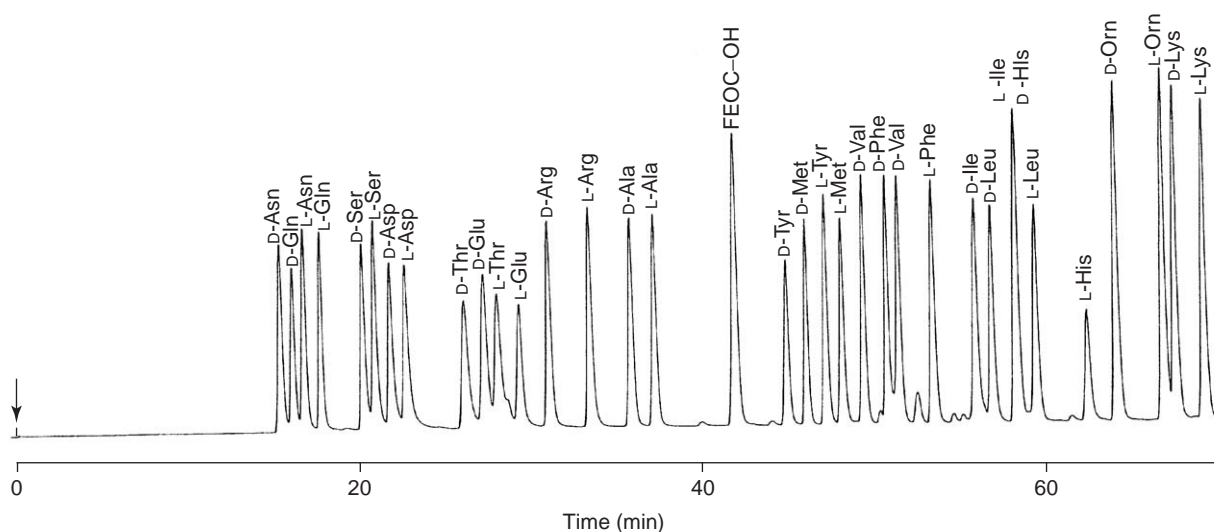


**Figure 2** Chiral derivatization reagents used for the enantioseparation of amino acids.

derivatized with NBD-F, while the latter for the enantioseparation of the respective amino acid derivatives. Furthermore, 3,5-dinitrobenzoyl- or 3,5-dinitroanilido-amino acids were separated on Pirkle-type CSPs. In Pirkle-type CSPs, in addition to

$\pi$ - $\pi$  interactions, hydrogen bonding and/or dipole-dipole interactions could help in chiral recognition of amino acid derivatives.

CSPs based on carbamoylated derivatives of quinine and quinidine as chiral selectors were found



**Figure 3** Separation of a standard solution of D- and L-amino acids by chiral derivatization with (+)-1-(9-fluorenyl)ethyl chloroformate followed by high-performance liquid chromatography on an achiral column. Column: Spherisorb octyl material (150 × 4.6 mm). Mobile phase: acetonitrile (ACN), tetrahydrofuran (THF), and an acetic acid buffer (1.8 ml of glacial acetic acid in 1 l of water; pH adjusted to 4.35 with NaOH): Gradient: 0–8 min, 8% ACN, 17% THF, 75% buffer (8/17/75); 8–22 min, 8/17/75 to 0/30/70; 22–70 min, 0/30/70 to 0/50/50. Flow rate: 0.8 ml min<sup>-1</sup>. (Reprinted with permission from Einarsson S, Josefsson B, Möller P, and Sanchez D (1987) *Analytical Chemistry* 59: 1194; © American Chemical Society.)

to be highly enantioselective for *N*-derivatized amino acids. Since carbamoylated quinine or quinidine has a positive charge, ionic interactions between the positively charged selectors and negatively charged amino acid enantiomers could work for chiral recognition of *N*-derivatized amino acids.

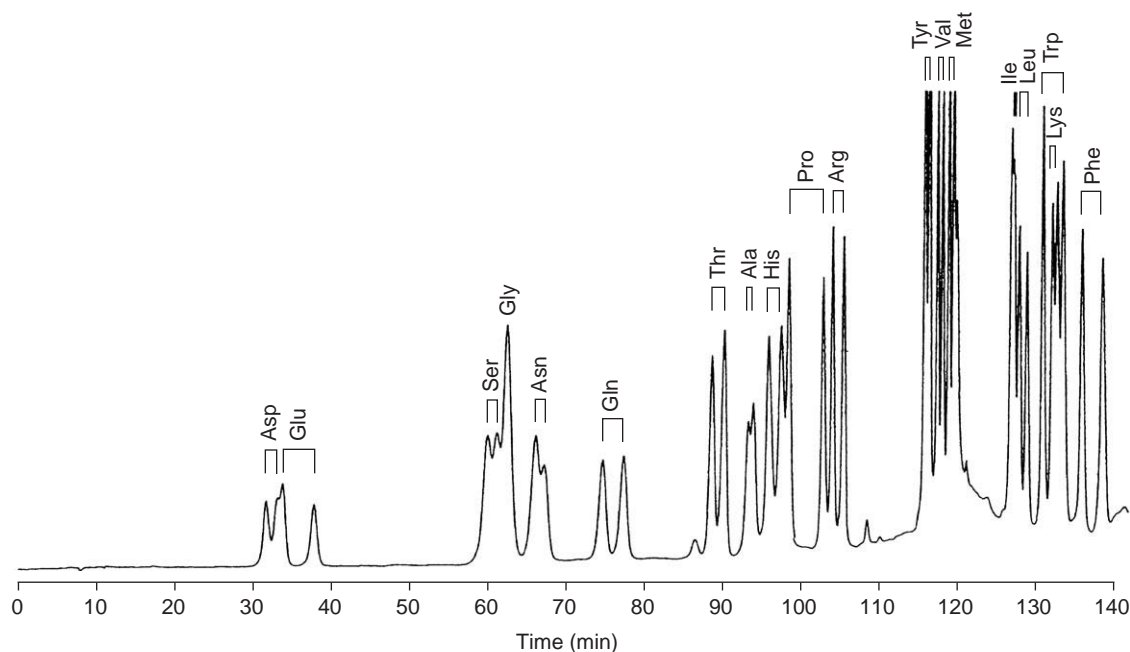
Recently, CSPs based on macrocyclic antibiotics such as vancomycin, teicoplanin, or ristocetin A have been introduced. These CSPs can separate many enantiomers of underivatized and derivatized amino acids in the normal and reversed-phase modes. In addition to  $\pi$ - $\pi$  interactions, hydrogen bonding, electrostatic and hydrophobic (in the case of reversed-phase mode) interactions could help in the chiral recognition of amino acid derivatives.

Cellulose and amylose derivatives are the best chiral selectors for the enantioseparation of various compounds. However, tris(3,5-dimethylcarbamate)s of cellulose and amylose only showed high resolving power for *N*-benzyloxycarbonyl  $\alpha$ -amino acid esters.

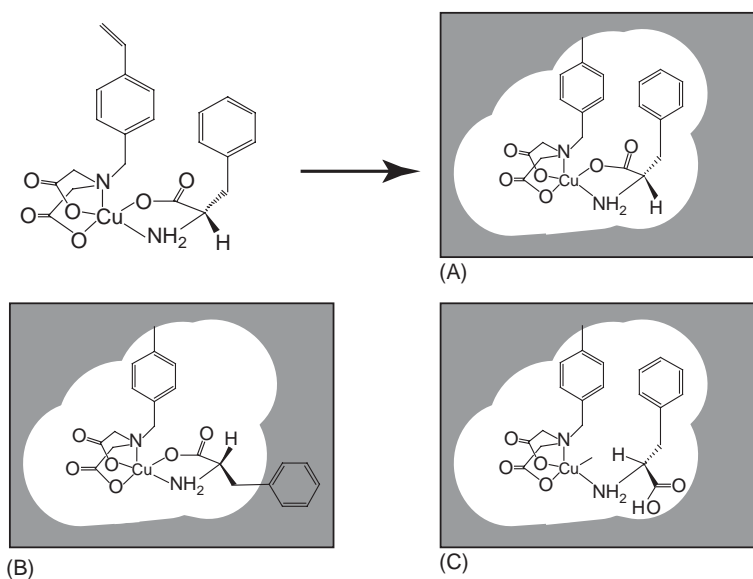
Crown ethers are synthetic macrocyclic polyethers that can form selective complexes with various cations. Chiral crown ethers such as (diphenyl-substituted 1,1'-binaphthyl) crown ether or (+)-(18-crown-6)-2,3,11,12-tetracarboxylic acid, which are bound to silica gels or coated on reversed-phase materials, were utilized for the enantioseparation of underivatized primary amino acids and their esters. On the other hand, the enantiomers of underivatized and derivatized amino acids enter into

hydrophobic chiral cavities of cyclodextrins (CDs) to form inclusion complexes, and relative stability constants of the resulting diastereomeric complexes are different. Underivatized amino acids such as tryptophan, phenylalanine, tyrosine, and their analogs can be separated on  $\alpha$ -CD bonded stationary phases using aqueous or hydro-organic mobile phases. However, other CDs ( $\beta$ - and  $\gamma$ -CDs) cannot separate underivatized amino acids. Amino acids derivatized with phenylisothiocyanate (PITC), OPA, dansyl chloride (DNS-Cl), or 9-fluorenylmethyl chloroformate (FMOCl) were separated on underivatized CDs ( $\alpha$ -,  $\beta$ -, and  $\gamma$ -CDs), and derivatized CDs (naphthylethyl carbamoylated and amino CDs). The enantioseparation of phenylthiocarbonyl derivatives of amino acids (amino acids derivatized with PITC) was attained on a series of reversed-phase HPLC columns and derivatized  $\beta$ -CD columns. **Figure 4** shows a chromatogram of the separation of 37 phenylthiocarbonyl amino acids on octyl and phenylcarbamoylated  $\beta$ -CD columns coupled in series. As described above, the former was used for the separation of the individual amino acids derivatized with PITC and the latter for the enantioseparation of the respective amino acid derivatives.

Chiral ligand-exchange chromatography is based on the formation of diastereomeric ternary complexes that involve a transition metal ion, chiral ligand, and the amino acid enantiomers. Among transition metals, Cu(II) formed the most stable complexes



**Figure 4** Separation of a mixture of 37 phenylthiocarbamoyl amino acids by achiral derivatization followed by high-performance liquid chromatography on a chiral stationary phase (second column). Column and temperature: Octyl-80Ts ( $150 \times 4.6$  mm) and phenylcarbamoylated  $\beta$ -CD ( $150 \times 6.0$  mm) columns were coupled in series at column temperatures of 30 and 20°C, respectively. Mobile phase: (A)  $100 \text{ mmol l}^{-1}$  ammonium acetate (pH 6.5) containing  $1 \text{ mmol l}^{-1}$  sodium butanesulfonate; (B)  $100 \text{ mmol l}^{-1}$  ammonium acetate (pH 6.5)/methanol = 50/50 (v/v) containing  $1 \text{ mmol l}^{-1}$  sodium butanesulfonate. Gradient: isocratic elution until 50 min (B, 0%), linear gradient elution from 50 to 80 min (B, 0–20%), then from 80 to 110 min (B, 20–80%), isocratic elution from 110 to 150 min (B, 80%). Flow rate:  $0.7 \text{ ml min}^{-1}$ . (Reprinted with permission from Iida T, Matsunaga H, Fukushima T, *et al.* (1997) *Analytical Chemistry* 69: 4467; © American Chemical Society.)



**Figure 5** Source of enantioselectivity in ligand-exchange MIPs. (Reprinted with permission from Vidyasankar S, Ru M, and Arnold FH (1997) *Journal of Chromatography A* 775: 61; © Elsevier.)

when cyclic amino acids such as L-proline and L-4-hydroxyproline were used as the ligand. For chiral ligand-exchange chromatography of amino acids, chemically bound L-proline, L-4-hydroxyproline,

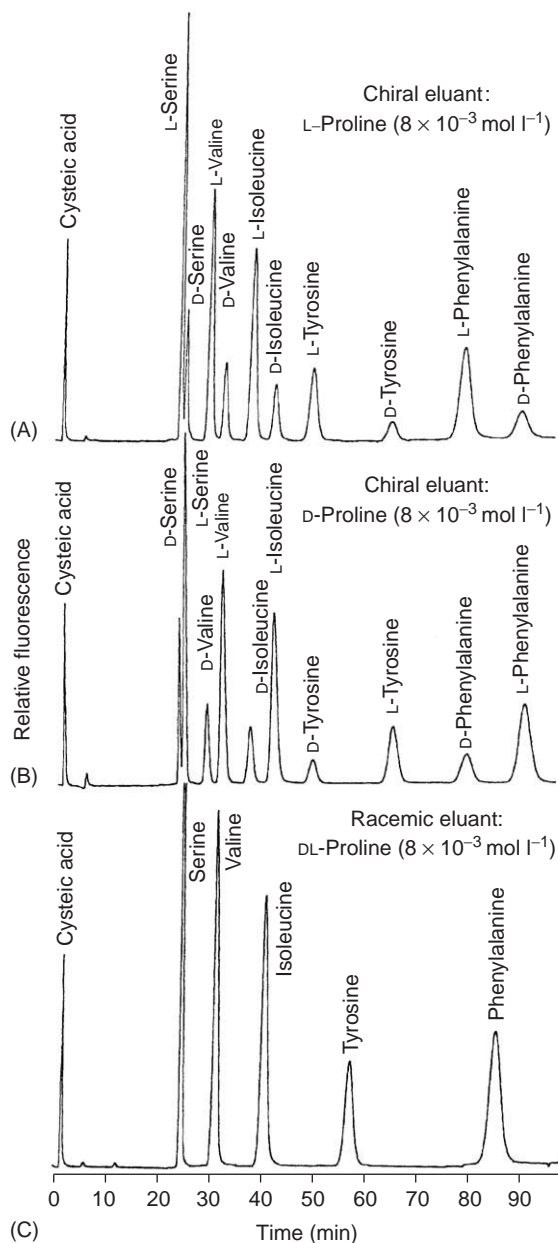
D-penicillamine, L-phenylalanine amide, L-alanine amide, or L-proline amide to silica gels or polymers was used as a CSP, and Cu(II) ion was added in a mobile phase, or adsorbed on the CSP. The

enantioseparation of underivatized amino acids and derivatized amino acids (dansyl amino acids or *N*-methyl amino acids) was attained.

Many CSPs based on a protein have been developed. CSPs based on bovine serum albumin or human serum albumin could be only utilized for the enantioseparation of underivatized tryptophan, while other underivatized amino acids were not enantioseparated. However, they can separate derivatized amino acids such as dansylated ones. A CSP based on  $\alpha$ -chymotrypsin and trypsin could be used for the enantioseparation of underivatized amino acid (tryptophan, arginine, or histidine), and *N*- or *O*-derivatized amino acids. Ovomucoid-based CSP could separate only derivatized amino acids. Recently, antibodies against protein conjugates of *D*- and *L*-*p*-aminophenylalanine were prepared, and applied to the enantioseparation of racemic amino acids such as phenylalanine, tryptophan, tyrosine, and histidine. The antibodies only recognized  $\alpha$ -amino acids that have the correct configuration; i.e., anti-*D*-amino acid antibodies bound to *D*-amino acids and anti-*L*-amino acid antibodies bound to *L*-amino acids. Chiral recognition mechanisms of protein CSPs are based on combinations of hydrophobic, electrostatic, and hydrogen bonding interactions. The disadvantages of these CSPs are low column efficiency because of too strong interactions of CSPs and enantiomers, especially late eluted ones, and lack of repetitive usage of the column.

In the molecular imprinting technique, a template molecule is polymerized in the presence of a functional monomer, which interacts with the template molecule. Thus, the obtained molecularly imprinted polymers (MIPs) could be used for the specific recognition of the template molecule. The MIPs for *N*- or *O*-derivatized amino acid enantiomers were prepared using methacrylic acid (MAA) as a functional monomer and ethylene glycol dimethacrylate as a crosslinker. The enantiomer used as the template molecule was strongly recognized on the obtained MIPs, and eluted latter than the antipode. With regard to enantiomeric resolution of underivatized amino acids, a ligand-exchange MIP based on the MAA-derivatized silica for *D*-phenylalanine was prepared using Cu(II)-*N*-(4-vinylbenzyl)iminodiacetic acid as an achiral monomer. The MIP obtained could separate racemic phenylalanine and tyrosine, but could not resolve racemic tryptophan, alanine, leucine, and isoleucine. The chiral recognition mechanism could be explained as follows: re-binding of the *D*-enantiomer proceeds through chelation of the metal ion, in addition to which the aromatic side chain fits into a cavity that selects for both the size and shape of this group, as shown in

Figure 5A. In contrast, metal chelation by the *L*-enantiomer would be sterically hindered (Figure 5B). If the side group of the *L*-enantiomer fits into the cavity, binding to the metal would be obstructed (Figure 5C). Racemic tryptophan and aliphatic



**Figure 6** Effect of chirality of the mobile phase additives on the separation of *D*- and *L*-amino acid enantiomers by ligand-exchange chromatography. Column: DC 4a cation-exchange resin ( $12 \times 0.2$  cm). Mobile phase: sodium acetate buffer ( $0.05 \text{ mol l}^{-1}$ , pH 5.5) containing  $4 \times 10^{-3} \text{ mol l}^{-1} \text{ CuSO}_4$  and  $8 \times 10^{-3} \text{ mol l}^{-1}$  proline. Column temperature:  $75^\circ\text{C}$ . Flow rate:  $10 \text{ ml h}^{-1}$ . (Reprinted (abstracted/experted) with permission from Hare PE and Gil-Av E (1979) *Science* 204: 1227; © American Association for the Advancement of Science.)



amino acids could not be resolved on the MIP because of the loose fit into the cavity.

### Chiral Mobile Phase Additives

There are no fundamental differences between the techniques using CSPs and chiral mobile phase additives. This means that all chiral selectors covalently bound to or coated on supports can be used for the enantioseparation of amino acids by addition to the mobile phase.

The most frequently used additives in chiral ligand-exchange chromatography are Cu(II) complexes of chiral ligands such as D- or L-proline, esters of L-proline or L-arginine,  $N^2$ -octyl-(S)-phenylalaninamide, or *N*-(*p*-toluenesulfonyl)-L-phenylalanine. Underivatized amino acids, amino acid amides, amino acid esters, and dansyl amino acids were separated using achiral C18 or ion-exchange columns. **Figure 6** shows the effect of the chiral ligand on the enantioseparation of underivatized amino acids. When a Cu(II) complex of L-proline ligand was used as the mobile phase additive, the L-enantiomers of amino acids eluted before the corresponding D-enantiomers (**Figure 6A**). When the antipode was used as the ligand, the elution order was reversed (**Figure 6B**). Furthermore, no enantiomeric resolution of amino acids was observed with a racemic Cu(II)-proline complex (**Figure 6C**).

CDs or derivatized CDs were used for the enantioseparation of dansyl- or *t*-butyloxycarbonyl-amino acids as the chiral mobile phase additive. One macrocyclic antibiotic, LY33328, was used for the enantioseparation of nine dansyl amino acids.

**See also:** Derivatization of Analytes. **Fluorescence:** Derivatization. **Liquid Chromatography:** Amino Acids. Pharmaceutical Applications.

### Further Reading

- Allenmark S (1991) *Chromatographic Enantioseparation: Methods and Applications*, 2nd edn. Chichester: Ellis Horwood.
- Bhushan R and Joshi S (1993) Resolution of enantiomers of amino acids by HPLC. *Biomedical Chromatography* 7: 235–250.
- Hamase K, Morikawa A, and Zaitse K (2002) D-Amino acids in mammals and their diagnostic value. *Journal of Chromatography B* 781: 73–91.
- Krstulovic AM (ed.) (1989) *Chiral Separations by HPLC: Applications to Pharmaceutical Compounds*. Chichester: Ellis Horwood.
- Lough WJ (ed.) (1989) *Chiral Liquid Chromatography*. London: Blackie.
- Stevenson D and Wilson ID (eds.) (1988) *Chiral Separations*. New York: Plenum.
- Taylor DR and Maher K (1992) Chiral separations by high-performance liquid chromatography. *Journal of Chromatographic Science* 30: 67–85.
- Toyooka T (1996) Recent progress in liquid chromatographic enantioseparation based upon diastereomer formation with fluorescent chiral derivatization reagents. *Biomedical Chromatography* 10: 265–277.
- Wainer IW (1993) *Drug Stereochemistry: Analytical Methods and Pharmacology*, 2nd edn. New York: Dekker.
- Zief M and Crane LJ (eds.) (1988) *Chromatographic Chiral Separations*. New York: Dekker.

## Biotechnology Applications

**R Freitag**, Ecole Polytechnique Federal Lausanne, Lausanne, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Biotechnology, i.e., the development and application of expertise from the disciplines of biology, chemistry, and engineering, represents a rapidly growing modern industry, which is increasingly being considered an important driving factor in many of today's economies. The spectrum of biotechnological products covers a wide range from simple substances such as ethanol and citric acid through antibiotics and vaccines to the most advanced 'biopharmaceuticals'

such as recombinant proteins, antibodies, or DNA-based therapeutics. Concomitantly a wide range of production organisms and conditions are used by the industry, while the application spectrum of 'biotechnology' covers areas as diverse as agriculture, food science, general industry, and, perhaps most importantly, medicine.

However, there are certain common motifs in analytical biotechnology, which is mainly focused on process monitoring and control on the one side and quality control (identity, purity, stability) of the final product on the other. Monitoring the downstream process, i.e., product isolation and purification, then combines aspects of both areas. Analytical biotechnology has to operate within rather strict boundaries. In particular, the most precious

amino acids could not be resolved on the MIP because of the loose fit into the cavity.

### Chiral Mobile Phase Additives

There are no fundamental differences between the techniques using CSPs and chiral mobile phase additives. This means that all chiral selectors covalently bound to or coated on supports can be used for the enantioseparation of amino acids by addition to the mobile phase.

The most frequently used additives in chiral ligand-exchange chromatography are Cu(II) complexes of chiral ligands such as D- or L-proline, esters of L-proline or L-arginine,  $N^2$ -octyl-(S)-phenylalaninamide, or *N*-(*p*-toluenesulfonyl)-L-phenylalanine. Underivatized amino acids, amino acid amides, amino acid esters, and dansyl amino acids were separated using achiral C18 or ion-exchange columns. **Figure 6** shows the effect of the chiral ligand on the enantioseparation of underivatized amino acids. When a Cu(II) complex of L-proline ligand was used as the mobile phase additive, the L-enantiomers of amino acids eluted before the corresponding D-enantiomers (**Figure 6A**). When the antipode was used as the ligand, the elution order was reversed (**Figure 6B**). Furthermore, no enantiomeric resolution of amino acids was observed with a racemic Cu(II)-proline complex (**Figure 6C**).

CDs or derivatized CDs were used for the enantioseparation of dansyl- or *t*-butyloxycarbonyl-amino acids as the chiral mobile phase additive. One macrocyclic antibiotic, LY33328, was used for the enantioseparation of nine dansyl amino acids.

**See also:** Derivatization of Analytes. Fluorescence: Derivatization. Liquid Chromatography: Amino Acids. Pharmaceutical Applications.

### Further Reading

- Allenmark S (1991) *Chromatographic Enantioseparation: Methods and Applications*, 2nd edn. Chichester: Ellis Horwood.
- Bhushan R and Joshi S (1993) Resolution of enantiomers of amino acids by HPLC. *Biomedical Chromatography* 7: 235–250.
- Hamase K, Morikawa A, and Zaitse K (2002) D-Amino acids in mammals and their diagnostic value. *Journal of Chromatography B* 781: 73–91.
- Krstulovic AM (ed.) (1989) *Chiral Separations by HPLC: Applications to Pharmaceutical Compounds*. Chichester: Ellis Horwood.
- Lough WJ (ed.) (1989) *Chiral Liquid Chromatography*. London: Blackie.
- Stevenson D and Wilson ID (eds.) (1988) *Chiral Separations*. New York: Plenum.
- Taylor DR and Maher K (1992) Chiral separations by high-performance liquid chromatography. *Journal of Chromatographic Science* 30: 67–85.
- Toyooka T (1996) Recent progress in liquid chromatographic enantioseparation based upon diastereomer formation with fluorescent chiral derivatization reagents. *Biomedical Chromatography* 10: 265–277.
- Wainer IW (1993) *Drug Stereochemistry: Analytical Methods and Pharmacology*, 2nd edn. New York: Dekker.
- Zief M and Crane LJ (eds.) (1988) *Chromatographic Chiral Separations*. New York: Dekker.

## Biotechnology Applications

**R Freitag**, Ecole Polytechnique Federal Lausanne, Lausanne, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Biotechnology, i.e., the development and application of expertise from the disciplines of biology, chemistry, and engineering, represents a rapidly growing modern industry, which is increasingly being considered an important driving factor in many of today's economies. The spectrum of biotechnological products covers a wide range from simple substances such as ethanol and citric acid through antibiotics and vaccines to the most advanced 'biopharmaceuticals'

such as recombinant proteins, antibodies, or DNA-based therapeutics. Concomitantly a wide range of production organisms and conditions are used by the industry, while the application spectrum of 'biotechnology' covers areas as diverse as agriculture, food science, general industry, and, perhaps most importantly, medicine.

However, there are certain common motifs in analytical biotechnology, which is mainly focused on process monitoring and control on the one side and quality control (identity, purity, stability) of the final product on the other. Monitoring the downstream process, i.e., product isolation and purification, then combines aspects of both areas. Analytical biotechnology has to operate within rather strict boundaries. In particular, the most precious

bioproducts (recombinant), proteins and DNA molecules, are not susceptible to analysis by methods such as gas chromatography and other standard high-resolution methods of analytical chemistry. In this context, liquid chromatography (LC) – together with the various electrophoretic methods – has been and will continue to be one of the major analytical techniques utilized for such work.

## LC

The most significant advantages of LC as an analytical technique in general are the diversity of parameters that can be used to achieve separation. For applications in analytical biotechnology, the most important of these parameters are size (gel permeation chromatography (GPC), size exclusion chromatography), charge (cation/anion exchange chromatography (IEC)), hydrophobicity/polarity (reversed-phase chromatography (RPC), hydrophobic interaction chromatography (HIC), normal-phase chromatography), and finally certain 'biospecific' interactions (affinity chromatography). By employing several of these techniques in series (multidimensional chromatography, e.g. IEC–HIC), highly orthogonal separation schemes can be drawn up that allow resolving even the most complex sample mixtures, e.g. for applications such as in 'proteomics'. An alternative to such 2D LC techniques is coupling a high-resolution LC technique such as RPC with a second analytical technique such as matrix-assisted laser desorption/ionization or electrospray ionization mass spectrometry, i.e., creating a so-called hyphenated technique. Especially when micro- or nano-LC techniques are used, hyphenated techniques have advantages in terms of sample size, speed, detection limit, and solvent consumption. The most recent developments, true microanalytical LC-systems, that work with capillary columns or perform the entire analysis on a microfluidic chip (total analytical system). In such systems the content of a single cell becomes in principle accessible to a detailed analysis.

### Separation Mechanisms

The various LC techniques available today to the analytical chemist are discussed in detail in other chapters and sections of this encyclopedia. Below therefore only the methods and features relevant to analytical biotechnology are briefly discussed. As the typical separation problems of analytical biotechnology are similar to those encountered in other areas of (bio)analytical chemistry, commercially available standard stationary phases and prepacked columns are used normally. The stationary phases are

normally porous or pellicular (solid), spherical particles with a diameter of 2 or 10  $\mu\text{m}$  and a very narrow particle size distribution (monosized beads) to ensure a high separation efficiency and minimum column back pressure simultaneously. Even complex separations can be carried out within minutes using such columns. In addition, so-called monolithic materials (UNO<sup>TM</sup>, BioRad USA, CIM<sup>TM</sup>-disks, BIA d.d.o, Slovenia) are being used increasingly, where the entire column consists of a single porous polymer block. Due to the improved mass transfer properties, such columns can be operated at extremely high flow rates without losing their efficiency.

**RPC** Reversed-phase high-performance liquid chromatography (RP-HPLC) is perhaps the most important LC method in analytical biotechnology. The stationary phase is usually a silica composite bearing hydrophobic hydrocarbon groups ( $\text{C}_2$ ,  $\text{C}_4$ ,  $\text{C}_8$ , or most often  $\text{C}_{18}$ ). Normally hydro-organic mobile phases (acetonitrile/water, methanol/water) are used, although variants using nonaqueous mobile phases have been reported. Elution is achieved by decreasing the polarity, i.e., increasing the organic content of the mobile phase. In the case of charged analytes, ion-pairing reagents such as trifluoroacetic acid (TFA) are added in small amounts to the mobile phase in order to reduce the nonspecific interaction between the analytes' charges with residual accessible silanol groups on the silica surface.

Due to their hydrophilicity, many biomolecules elute early in the gradient. While  $\text{C}_{18}$ -phases are still used in the majority of analytical applications, less hydrophobic phases bearing shorter hydrocarbon groups are gradually becoming available. Another consequence of the low hydrophobicity of typical biomolecules is the need for compatibility of the stationary phase with mobile phases of low organic solvent content. While most of the earlier  $\text{C}_{18}$ -phases, for example, could not be used with a mobile phase consisting of 100% water (irreversible collapse of the  $\text{C}_{18}$ -brush structure), today several column types are available that can be used with gradients ranging from 0 to 100% of an organic modifier in the mobile phase.

In analytical biotechnology, RP-HPLC is used for process monitoring, especially of media compounds such as certain amino acids but also for analysis of products such as antibiotics, enzymes, and other proteins. In addition, RP-HPLC constitutes the most important method for establishing the identity of many biotech products. In the case of recombinant proteins this may include the retention time of the product itself, the analysis of the amino acid composition, the analysis of the peptide map, as well

those of N-terminal variants, glycovariants, disulfide isomers, and proteolytic clips. Due to the volatility of its typical mobile phases, RP-HPLC has also become the method of choice for LC–mass spectrometry (MS) coupling.

**Ion-exchange chromatography** Ion exchangers can be divided into cation exchangers and anion exchangers, with a further division into strong and weak ones according to the functional groups. The choice between a cation and anion exchanger will depend upon the net charge of the analyte under the conditions used – anion exchange for negatively charged analytes and cation exchange for positively charged ones. Whether to use a strong or weak exchanger is a more difficult decision. The charge density of a strong ion exchanger is essentially independent of pH, but for a weak ion exchanger it is dependent upon the  $pK_a$  of the stationary phase. Ion exchange chromatography is carried out normally with an aqueous mobile phase (buffer) and tends to preserve biological activity to a high extent. Elution in IEC is accomplished by changing the pH or – most commonly – by increasing the ionic strength of the mobile phase.

In analytical biotechnology IEC is used for process monitoring (amino acids, sugars, certain media compounds and products) and for quality control purposes, for example, for determination of glycosylation variants or detection of deamidated product molecules (proteins). In addition, IEC type stationary phases are also used in many sample preparation schemes for rapid enrichment of the target molecule-containing fraction prior to further analysis.

While being a most powerful and widely used preparative separation technique, analytical IEC has the disadvantage of being difficult to couple to MS due to the need for a high salt elution buffer. For the same reason, miniaturization is difficult since high salt buffers are intrinsically unsuited for electrochromatographic applications, i.e., chromatographic separation, where the electro-osmotic flow is used to drive the mobile phase. Pressure-driven  $\mu$ -LC is possible in the IEC mode but is more difficult to perform.

**Gel filtration (size exclusion)** Gel filtration chromatography (also called size exclusion chromatography) employs porous beads with a defined pore size distribution as the stationary phase. Small molecules can enter the entire intraparticle pore space and hence elute last, whereas large molecules are excluded from all pores and hence elute first. Molecules of intermediate size are found that can enter a certain

portion of the intraparticle pores. For these molecules the column residence time is a direct function of their size and they can hence be separated according to this parameter. Retentive (or repulsive) interactions between the analytes and the stationary phases have to be avoided, and the chromatographic conditions (mobile phase composition, stationary phase material) have to be chosen accordingly. In analytical biotechnology, GPC is used mainly for quality control, e.g., to differentiate between dimers/product aggregates and the actual product during purity and stability testing.

**Affinity chromatography** In affinity chromatography the high specificity associated with the formation of a complex by two biological molecules is exploited. Ligands with specific binding sites, e.g., the antibody–antigen system, or group-specific molecules, e.g., lectins for the isolation of glycoproteins or Protein A for certain antibodies, can be used. Increasingly, affinity ligands are coming out of specific screening programs, based on, for example, combinatorial chemistry or biological systems such as phage display. The affinity ligand is linked covalently to an inert matrix. A spacer arm may be required, particularly for small molecules for decreasing matrix effects or steric hindrance to the interaction with the target molecules. Due to its high selectivity, affinity chromatography is a very important preparative technique in the biotechnological downstream process. In analytical biotechnology it is used in some cases to verify biological activity but mostly as a first capturing step for enriching a highly diluted product from the production stream prior to analysis using other techniques or for specific production rate monitoring during fermentation. The fact that often product variants (dimers, fragments, deamidated molecules, glycovariants) bind equally well to the affinity column may present a problem. Recently the coupling of affinity chromatography to MS has become more feasible, with in-line dialysis modules becoming available.

## Amino Acids

Determination of the amount and distribution of amino acids is required for a variety of bioanalytical questions. In bioprocess monitoring, the amount of certain free amino acids has to be monitored in order to avoid media depletion. In the case of a protein or a protein hydrolysate, analysis or verification of the amino acid composition may be a first step toward identification or characterization. The early work on amino acid analysis has been dominated by Stein and Moore, who in 1972 were awarded the Nobel prize



in chemistry for their work. The first dedicated amino acid analyzer was introduced by D.H. Spackman in 1950. Even today, the analysis of free amino acids is still done very much as proposed by these researchers, although considerable progress has been made in terms of the detection limit and the required sample volume. Analysis of free amino acids has to face two major difficulties. First, amino acids are small molecules that cover a wide range of polarity; they are therefore inherently difficult to handle for most LC techniques except IEC. Second, save for the aromatic amino acids, they possess almost no ultraviolet (UV) or fluorescence activity and can therefore only be detected using indirect methods, i.e., after labeling with a suitable dye. The introduction of a label prior to chromatography (precolumn labeling) will change the chromatographic behavior of the molecules and make them, for example, suitable for analysis using RP-HPLC. Compared with IEC, the separation will become faster and the resolution should improve in such cases.

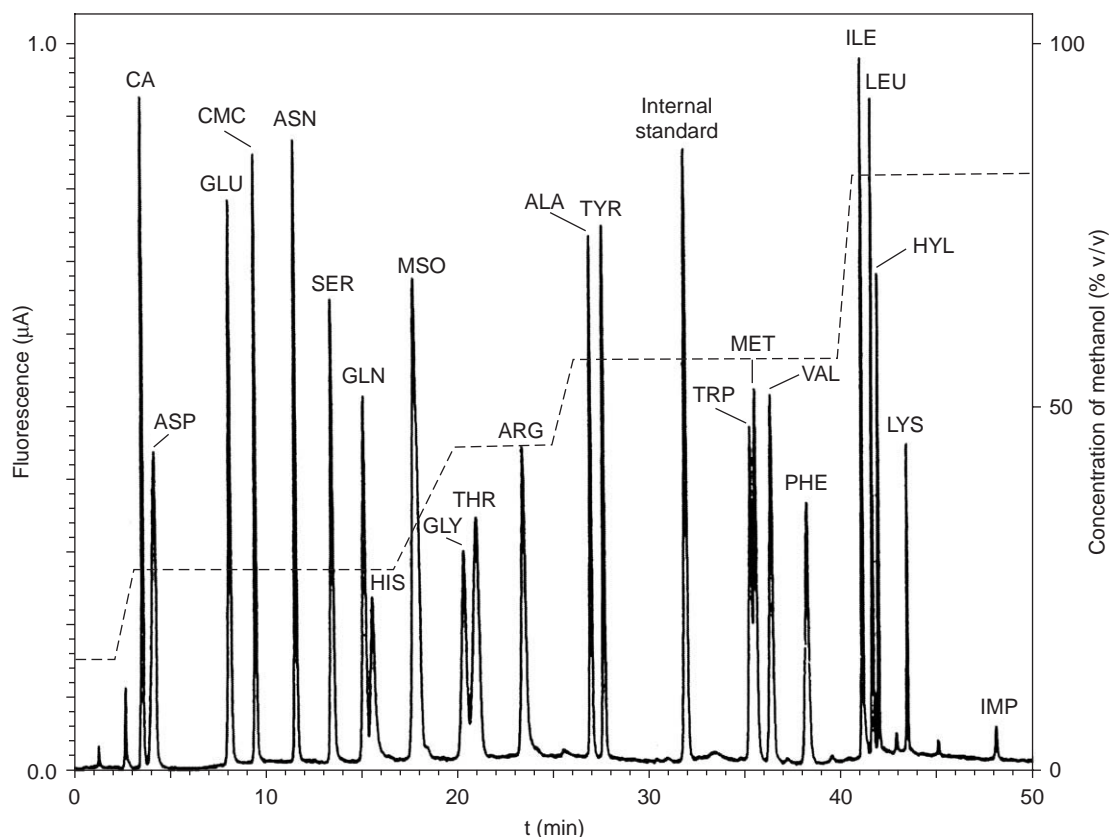
Today, analysis of free amino acids is normally carried out using IEC together with postcolumn derivatization for detection or by RP-HPLC together

with precolumn derivatization (Figure 1). The detection limit of precolumn techniques tends to be one order of magnitude lower ( $<1$  pmol). At such levels, interference by impurities (microbial growth in column and/or buffer) may become a severe problem. The majority of the commercially available amino acid analyzers use IEC resins for the separation and ninhydrin as the postcolumn chromogenic reagent, although *ortho*-phthalaldehyde is gaining in popularity.

### Derivatization Reactions

Amino acids are derivatized in order to enable their detection by standard chromatographic detectors. Even today the perfect label has not been found. Such an ideal label should be easy to introduce in a defined manner (only one reaction product per amino acid) and should result in a stable and highly active (UV-absorbance, fluorescent activity) amino acid derivative, and it should be universal and react with all possible amino acids of interest.

**Dyes for postcolumn derivatization** Ninhydrin has been the standard reagent for detecting amino acids



**Figure 1** Separation of a physiological mixture of amino acids. (Reproduced with permission from Agnes H (ed.) (1985) *High-Performance Liquid Chromatography in Biochemistry*. Wiley-VCH, p. 156.)



for over 50 years. However, not all amino acids react equally well with ninhydrin, and the detection limit is  $\sim 50$  pmol ( $\sim 6$  ng) of amino acid derivative. The product of the reaction of *ortho*-phthaldialdehyde (OPA) with amino acids in the presence of thiol will fluoresce with excitation at 340 nm and emission at 450 nm. The reaction is fast (1–3 minutes), however, the OPA-derivatives are not stable. Secondary amino acids do not react and must first be converted into primary amino acids (oxidation with hypochlorite or chloramines-T (N-chloro-*p*-toluenesulfonamide)). The detection limit in the case of postcolumn derivatization with OPA is  $\sim 10$  pmol (1.2 ng). Fluorescein has also been proposed for postcolumn derivatization of free amino acids but is not widely used due to the many problems associated with its application.

**Dyes for precolumn derivatization** The Edman reagent (phenylisothiocyanate) can be used to form thiohydantoin derivatives of primary and secondary amino acids, which can be separated using RPC (detection UV 254 nm). The detection limit is  $\sim 1$  pmol ( $\sim 0.1$  ng). OPA is also used for the precolumn derivatization. Detection limits (fluorescence) down to the femtomole level are possible. Some other dyes are of limited importance in precolumn derivatization. 4-(Dimethylamino)azobenzene-4'-sulfonyl chloride reacts with amino acids to produce a derivative that is stable at room temperature and can be detected at the 3–5 pmol level using detection at 436 nm. 4-(Dimethylamino)azobenzene-4'-isothiocyanate in combination with RPC has a detection limit of  $\sim 5$  pmol with detection at 436 nm. Fluorenylmethoxycarbonyl chloride forms derivatives of all amino acids, including proline. UV detection is possible at 260 nm, and fluorescence detection (excitation 266 nm, emission 305 nm) is also possible, the detection limit being  $\sim 50$  fmol in that case. 5-(Dimethylamino)naphthalene-1-sulfonyl chloride also reacts with both primary and secondary amines. The derivatives can be detected using UV at 254 nm or using fluorescence (excitation 385 nm, emission 460 nm). The sensitivity is in the low picomole range. Due to some problems with the specificity and the completeness of the reaction, the dye is rarely used.

### Enantiomeric Separations

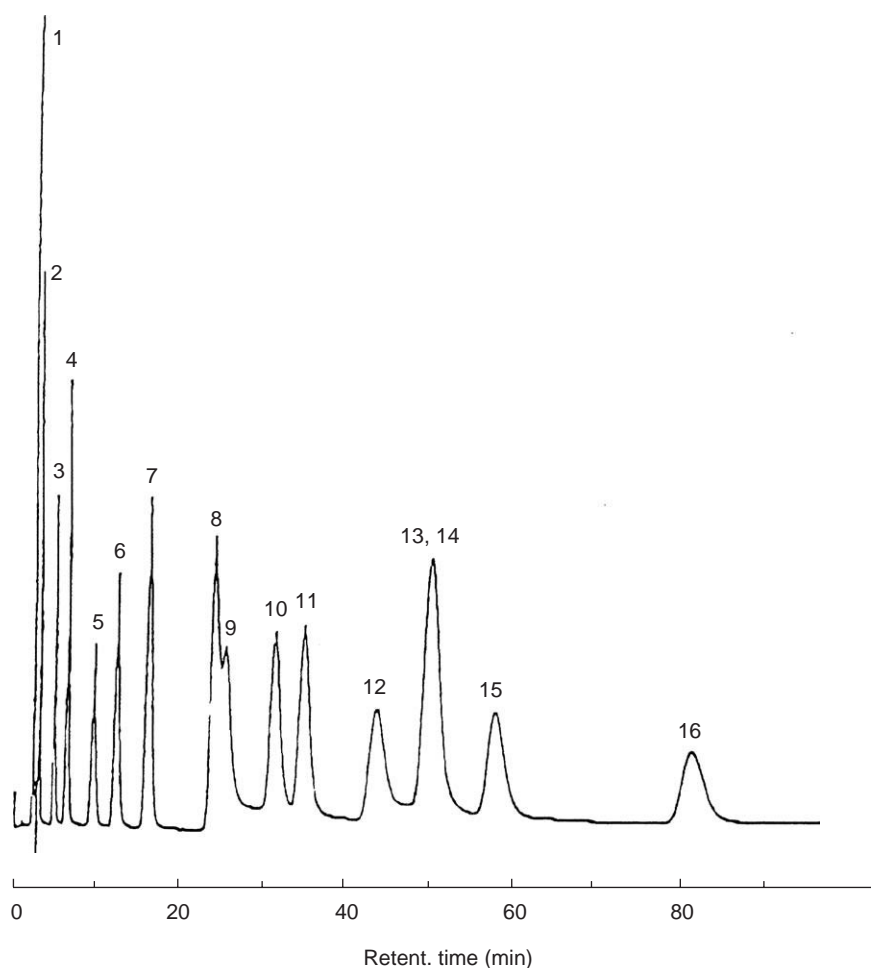
The need to resolve the D- and L-enantiomers of amino acids has grown in recent years as it has been recognized that they differ in biological and physicochemical properties. Amino acid enantiomers cannot be resolved in achiral systems; it is therefore necessary to have a second chiral center either in the chromatographic system or to create one by derivatization in the molecule to be separated (Figure 2).

Conversion of amino acids to produce a diastereoisomeric peptide by reaction with an *N*-carboxy-anhydride, e.g., L-Phe-*N*-carboxy-anhydride, has been used for determination of optical purity using RPC. Alternatively, for analysis of amino acid enantiomers without derivatization, two options are available: (1) chiral mobile phases such as the *N,N*-di-*n*-propyl-L-alanine-copper(II) complex can be used with reversed-phase columns, and (2) stationary phases with a covalently bound ligand capable of stereo recognition can be used. Such ligands include cyclodextrins, albumins, glycoproteins, and copper(II) complexes.

### Peptides, Proteins, 'Proteomics'

The analysis of peptides and proteins naturally has some analogies with that of their main building blocks, amino acids (Figure 3). The analysis of proteomes, i.e., the entire gene product spectrum expressed by a given organism (cell) at a given time, is a very recent development in protein analysis. Proteome analysis is largely dominated by 2D gel electrophoresis (orthogonal separation of the protein mixture according to isoelectric point and mass) as a literature survey will show that still well over 90% of the pertinent separations are carried out using this method. However, the application of orthogonal LC–LC and LC–MS techniques for this purpose is gaining ground.

For analysis of proteins and (synthetic) peptides, on the other hand, LC techniques are the method of choice (Figure 4). The problems encountered in analysis of free amino acids, namely the lack of an intrinsic detection method and the wide polarity range, are much less significant in the case of proteins/peptides. Since most of these molecules contain at least one aromatic amino acid and several peptide bonds, their UV detection at 280 nm and 214 nm, respectively, is straightforward. Since most peptides/proteins are made for a mixture of polar and non-polar amino acid residues, their separation using RP-HPLC in a gradient of increasing organic solvent is the most commonly used approach. The resolution of RP-HPLC is high enough to show clear differences between a protein and its variants (reduced, deamidated form, etc.) and also between the correct product of a peptide synthesis and peptides where one or several addition steps have failed. A second important method for protein characterization and identification is the peptide map. For this purpose the protein is digested by a protease such as trypsin (tryptic map) or chymotrypsin, and the resulting peptide mix is analyzed using RP-HPLC. Even single



**Figure 2** Analysis of eight amino acid racemates. Chromatography conditions: column MCI GEL CRS10W, eluent  $0.5 \text{ mmol l}^{-1}$  copper (II) sulfate, flow rate  $1.0 \text{ ml min}^{-1}$ , detection  $254 \text{ nm}$ , peaks 1 D-Ala, 2 L-Ala, 3 D-Pro, 4 D-Val, 5 L-Pro, 6 L-Val, 7 D-Leu, 8 D-Nle, 9 D-Tyr, 10 L-Leu, 11 D-Eth, 12 L-Tyr, 13 L-Nle, 14 D-Phe, 15 L-Eth, 16 L-Phe. (Reproduced with permission from Weston A and Brown Ph (1997) *HPLC and CE – Principles and Practice*. Academic Press, p. 61; © Elsevier.)

amino acid deviations can normally be detected as differences in the peak pattern given by the protein product and the variants. Peptide mapping is in addition very important for an indirect evaluation of the glycosylation pattern as the glycoform of a given peptide will show significant differences in the retention behavior compared with the nonglycosylated form. Finally, in combination with MS, the identity of the separated peptides can be determined and compared with the theoretical one in the case of proteins with known primary structure.

## Carbohydrates

Carbohydrate analysis has been slow to develop in comparison with that of other biologically important molecules. However, the more recent developments in biotechnology and renewable resources of raw materials have promoted interest in this diverse class

of compounds. Indeed, the term ‘carbohydrate’ encompasses a wide range of compounds from the simple building blocks (monosaccharides) to carbohydrate polymers, oligosaccharides, and polysaccharides and to various carbohydrate-containing species including glycoproteins, proteoglycans, nucleic acids, glycolipids, and antibiotics.

### Monosaccharides

There is a great variety in structure as well as chemical and physical properties of monosaccharides. There is no single method that is applicable to the qualitative and quantitative analysis of all monosaccharides; instead, the method must be chosen according to the chemical and physical characteristics of the solutes of interest. Borate complexes of monosaccharides can be separated using strong (quaternary amine, Q-type) anion-exchange columns. Alternatively, if the monosaccharide is acidic

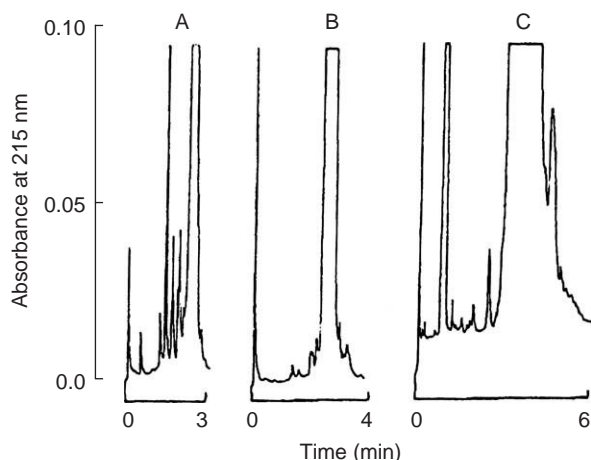
or negatively charged at high pH, anion exchangers can be used directly. Cation-exchange chromatography is normally used for analysis of amino sugars and – using calcium as counterion – alditols and monosaccharides.

### Oligosaccharides

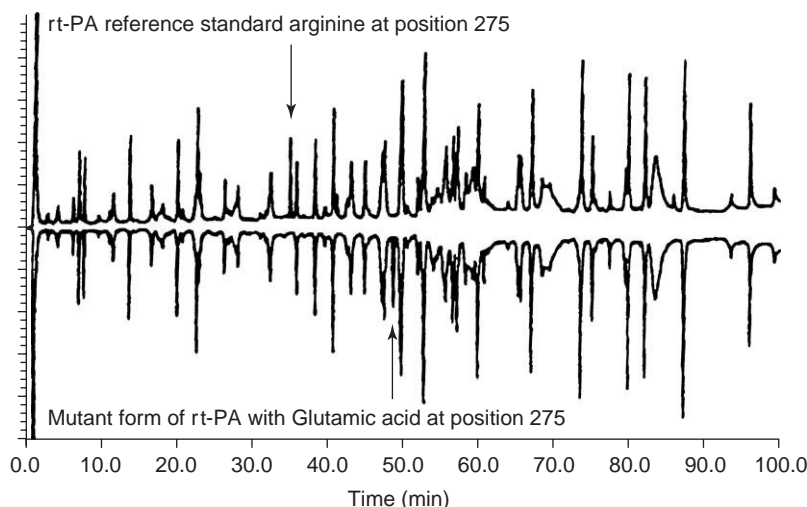
Oligosaccharides are traditionally defined as polymers of monosaccharides containing from two to ten

residues, although this range is often extended to 25 as naturally occurring polysaccharides rarely contain less than 25 repeat units. Anion-exchange resins with acetate, hydroxyl, or chloride counterions are usually used for oligosaccharide analysis. For a homologous series, the elution increases with increasing chain length. In an attempt to overcome the problems of interconversion of the terminal reducing residues of oligosaccharides, cation-exchange resins, in particular sulfonated poly(styrene–divinyl–benzene) materials, have been used for oligosaccharide separations, with water as the mobile phase. The effect of the counterion, e.g., lithium, barium, potassium, silver, or lead, is considerable as the separation mechanism is not purely cation exchange but also encompasses ligand exchange.

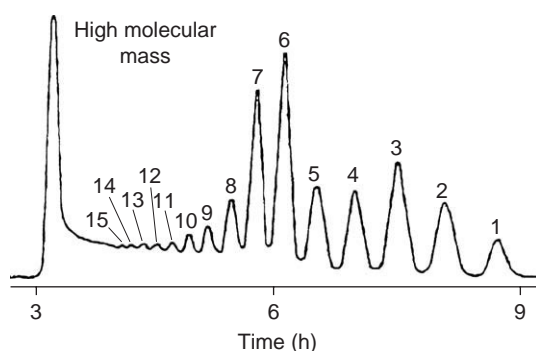
Gel filtration chromatography is another very useful technique for analysis of oligosaccharides. By a judicious choice of the stationary phase, a resolution of up to DP 15 can be achieved. The use of RPC for oligosaccharide analysis, on the other hand, has in general been limited to smaller oligomers owing to problems of solubility in the typical hydro-organic mobile phases. Normal-phase chromatography using stationary phases with amino or cyano functionalities has been used extensively for oligosaccharide analysis, again with aqueous acetonitrile mobile phases. However, unlike in RPC, elution takes place in a gradient of increasing water content of the mobile phase. Higher oligomers or polysaccharide material may be bound irreversibly to the column or have excessively long elution times.



**Figure 3** Determination of protein purity. Chromatographic conditions: column Hy-Tach micropellicular C-18 silica, eluent A 0.1% TFA in water, eluent B 95% acetonitrile in water containing 0.1% TFA, flow rate  $2.0 \text{ ml min}^{-1}$ , sample A  $40 \mu\text{g}$  carbonic anhydrase (gradient 15 to 55% B in 3 min), B  $40 \mu\text{g}$  L-asparaginase (gradient 30 to 40% B in 4 min), C  $40 \mu\text{g}$  myoglobin (gradient 23 to 45% B in 6 min). (Reproduced with permission from Horvath C and Nikelly JG (eds.) (1990) *Analytical Biotechnology*, ACS Symposium Series 434; © American Chemical Society.)



**Figure 4** Comparison of the tryptic map of a recombinant tissue plasminogen activator reference standard and a mutant form with a glutamic acid residue instead of the normal arginine residue at position 275. (Reproduced with permission from Weston A and Brown Ph (1997) *HPLC and CE – Principles and Practice*, Academic Press, p. 37; © Elsevier.)



**Figure 5** Gel filtration chromatography of starch-derived oligosaccharides on a Bio-Gel P2 column. Peak numbers refer to the degree of polymerization.

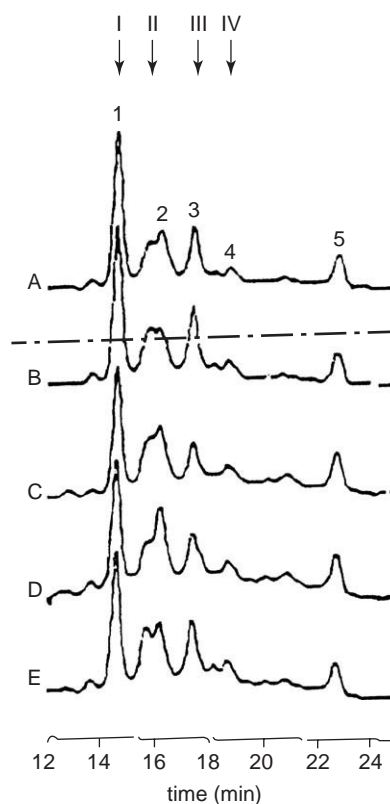
### Polysaccharides

The number of liquid chromatographic methods available for analysis of polysaccharides is limited compared with those for monomers or oligomers. Polysaccharides comprise a diverse group of complex macromolecules that may range in molecular mass from several thousand to several million Daltons and may be composed of neutral, acidic, or basic monomers or mixtures thereof arranged in a variety of structures. As polysaccharides are polydisperse macromolecules, gel filtration is often the first liquid chromatographic method to be used for their characterization. As the separation mechanism is one based on solute size, columns can be calibrated with polysaccharides of known molecular masses, and estimates for unknowns can be made. IEC is used to assess the homogeneity of a polysaccharide, often in combination with gel filtration chromatography, to separate molecules of similar size on the basis of differences in their charges (Figure 5).

### Carbohydrate-Containing Macromolecules, Glycoproteins

The methods previously discussed for mono-, oligo-, and polysaccharides are also applicable to analysis of carbohydrate-containing macromolecules. For example, proteoglycans, which consist of a central core protein to which a number of glycosaminoglycan chains are attached, which carry a significant negative charge due to a high concentration of carboxyl and sulfate ester groups, can be separated using anion-exchange chromatography (Figure 6).

Of increasing importance in medical biotechnology is analysis of the glycosylation pattern, i.e., the composition and structure of the oligosaccharide chains of glycoproteins. While the amino acid sequence is encoded by the genetic code and can hence be transferred by genetic engineering from one

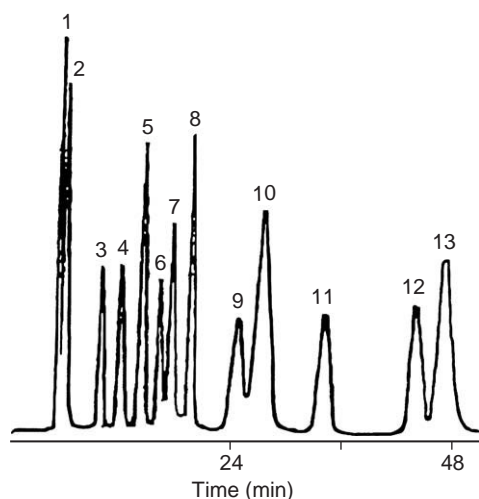


**Figure 6** HPAEC-PAD mapping of N- and O-glycans of glycoproteins after release by hydrazinolysis. (Reproduced with permission from Animal cell technology: products from cells, cells as products in *Proceedings of the 16th ESACT Meeting, April 25–29, Lugano, Switzerland*; © Kluwer Academic Publishers.)

species to the next, glycosylation is a post-translational event and hence shows pronounced differences from species to species and even between the different cell types of a given organism. At the same time glycosylation does have significant consequences for the biological activity as well as for the pharmacokinetics and -dynamics. A typical procedure in analysis of glycosylation patterns is comparison of the peptide map of the native glycoprotein and of that of the residual polypeptide structure after enzymatic release of the sugar residues together with a direct analysis of the released glycans using high-performance anion-exchange chromatography (HPAEC) with pulsed amperometric detection (PAD).

### Detection Methods

In general carbohydrates do not have a UV chromophore and do not fluoresce. Monosaccharides do absorb in the far UV, with the absorbance maxima at  $\sim 188$  nm. Detection is normally performed at wavelengths between 192 and 200 nm, where there is less residual noise in the detector signal, if the absorbance of the solvent permits this. The UV



**Figure 7** Ion exclusion isocratic separation of organic acids. Peaks: 1 chloride, 2 oxalate, 3 pyruvate, 4 tartrate, 5 malonate, 6 lactate, 7 malate, 8 acetate, 9 isocitrate, 10 citrate, 11  $\beta$ -hydroxy-*n*-butyrate, 12 succinate, 13 propionate. (Reproduced with permission from Dionex Product Selection Guide 1991.)

response is dependent upon the carbonyl group and will differ for the individual monosaccharides, fructose and glucose being the two extremes, with the response of fructose being six times that of glucose. Differential refractive index detection has often been used in the past; however, the method is not particularly sensitive and is only suitable for isocratic separations. Increasingly, PAD is being used as the universal method in the LC of mono- and oligosaccharides (Figure 7). The carbohydrate is oxidized at a gold electrode at a pH greater than 11, with the electrode surface being reduced back to gold using a negative potential sequence.

In addition, carbohydrates can be derivatized to produce UV or fluorescent species to improve detection. As with the amino acids, both precolumn and postcolumn derivatization are used. Precolumn derivatization influences the chromatographic behavior of the compounds. For UV detection, phenylisocyanate derivatization can be used together with RP-HPLC or benzoyl chloride and 4-bromobenzoyl chloride derivatization together with normal-phase chromatography. When fluorescence detection is preferred, dansyl hydrazine and aminopyridine derivatives can be analyzed using RPC.

Postcolumn derivatization is normally preferred for carbohydrate analysis as it is applicable to mono-, oligo-, and polysaccharides. It also improves the detection limit, and of course analysis selectivity, but without altering the intrinsic separation characteristics of the carbohydrates by introducing a group or groups that may dominate the separation process. The most common derivatization reactions are

automated versions of the spectrophotometric assays traditionally used for identification and quantification of carbohydrates. Methods of detection for total carbohydrate include the following: L-cysteine-sulfuric acid- and phenol-sulfuric acid-based assays for neutral carbohydrates, the carbazole assay or Warren assay for acid carbohydrates, the Elson Morgan assay for basic carbohydrates, and the 3,5-dinitrosalicylic acid assay for reducing groups.

## Organic Acids

Short-chain mono-, di-, and polycarboxylic acids are of significance to scientists studying amino acid degradation, metabolites of carbohydrate oxidation, and lipid breakdown. They can be indicators of human metabolic disorders and of importance in the biotechnology of food. LC is now the preferred method of analysis for these analytes irrespective of the sample origin – biological fluid, fermentation broth, or food.

IEC, or perhaps more correctly ion-exclusion chromatography using sulfonated polymeric resins, has become the method of choice for separating organic acids. The stationary phase can withstand extremes of pH, and the separation of the organic acids is normally according to  $pK_a$  values, but partition effects can also occur. Dilute acids are used as the mobile phase. Normally only isocratic mobile phases are required, with the addition of a small quantity of an organic modifier (usually acetonitrile) for analysis of nonpolar acids. In the case of weak organic acids, ion pairing agents can be used and the analysis can then also be achieved using RP-HPLC. However, this separation system is limited to certain organic acids, and the reproducibility can be poor. Detection can be achieved at nanogram or picogram levels, using electrochemical and conductivity detectors.

## Monitoring Fermentations and Fermentation Products

LC continues to be an important tool in fermentation monitoring. Using modern technology, the methods used are sufficiently fast and robust even for fermentation control, for example, to supervise nutrient addition and hence to maintain an optimum environment for the production organisms. Carbohydrates, amino acids, and the content of organic acids, all critical to either the production or the inhibition of production, can all be monitored as described above.

In the production of wine, monitoring the pyruvic acid content during malo-lactic fermentation gives an



indication of the health of the bacteria, a pyruvate being an indirect measure of NADH, the reducing co-factor that stimulates bacterial growth. Malates and tartrates are also simultaneously monitored, being indicators of the maturity of the fermentation process. Lactose fermentation can yield lactic, acetic, pyruvic, formic, propionic, and butyric acids, which are indicators of the flavor and stability of dairy products. All these acids can be determined in one LC-analysis run.

In the food industry LC is frequently used to determine the quality of the finished product. For example, it has been shown that the flavor quality of grape juice is directly linked to the amount and spectrum of organic acids present. Likewise, in the production of sour cream by fermentation, the quality of the finished product quality is related directly to the organic acid profile. Last, but not least and as pointed out above, LC techniques are also crucial in determining the quality of recombinant protein products.

See also: **Carbohydrates:** Overview; Sugars – Chromatographic Methods. **Chromatography:** Overview; Principles. **Derivatization of Analytes.** **Electrophoresis:** Two-Dimensional Gels; Proteins. **Fluorescence:** Derivatization; Fluorescence Labeling. **Ion Exchange:** Overview; Principles; Isolation of Biopolymers. **Liquid Chromatography:** Overview; Principles; Column Technology; Mobile Phase Selection; Normal Phase; Reversed Phase; Size Exclusion; Chiral; Affinity Chromatography; Multidimensional; Instrumentation; Liquid Chromatography–Mass Spectrometry; Amino Acids; Clinical Applications. **Micro Total Analytical Systems.** **Peptides.** **Pharmaceutical Analysis:** Overview. **Process Analysis:** Chromatography; Bioprocess Analysis.

**Proteins:** Traditional Methods of Sequence Determination. **Proteomics.**

## Further Reading

- Chaplin MF (1986) Monosaccharides. In: Chaplin MF and Kennedy JF (eds.) *Carbohydrate Analysis – A Practical Approach*, pp. 37–69. Oxford IRL Press.
- El Rassi Z (ed.) (1995) *Carbohydrate Analysis*. Journal of Chromatography Library, vol. 58. Elsevier.
- Heftmann E (ed.) (1991) *Chromatography*, 5th edn., Journal of Chromatography Library, vol. 51. Elsevier.
- Horvath C and Nikelly JG (eds.) (1990) *Analytical Biotechnology*. ACS Symposium Series 434. Washington, DC: American Chemical Society.
- Kastner M (ed.) (2000) *Protein Liquid Chromatography*. Journal of Chromatography Library, vol. 67. Elsevier.
- Keller R, Lottspeich F, Meaer HE (eds.) (1999) *Microcharacterization of Proteins*. Wiley–VCH.
- Lloyd LL, Warner FP, and Kennedy JF (1990) The application of HPLC in biotechnology. In: *Biotechnology International*, pp. 221–226. London: Century Press.
- Schwarzenbach R (1982) High-performance liquid chromatography of carboxylic acids. *Journal of Chromatography* 251: 339–358.
- Sofer G and Hagel L (1997) *Handbook of Process Chromatography – A Guide to Optimization, Scale-up and Validation*. Academic Press.
- Tukova J (ed.) (1993) *Bioaffinity Chromatography*, 2nd completely revised edn., Journal of Chromatography Library, vol. 55. Elsevier.
- Weston A and Brown Ph (1997) *HPLC and CE – Principles and Practice*. Academic Press.
- White CA and Kennedy JF (1986) Oligosaccharides. In: Chaplin MF and Kennedy JF (eds.) *Carbohydrate Analysis – A Practical Approach*, pp. 37–69. Oxford IRL Press.

## Clinical Applications

**R A Sherwood**, King's College Hospital, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by R A Sherwood and B F Rocks, pp. 2677–2685, © 1995, Elsevier Ltd.

## Introduction

The diagnosis and monitoring of many disease states relies on the detection and quantitation of a particular compound(s) or metabolite(s) in body fluids, usually blood or urine. A wide variety of techniques are used in the clinical chemistry laboratory with spectrophotometric and immunoassay methods on automated analyzers measuring single compounds

predominating. In some instances a family of compounds may need to be measured and it is in these situations where the chromatographic techniques are important to avoid the necessity for a separate assay for each component of interest in the mixture.

Initially chromatography was carried out using paper or slab gels. The use of columns to contain the stationary phase came more recently with the advent of high-performance liquid chromatography (HPLC). Typical applications of HPLC in clinical chemistry include:

1. Biogenic amines
  - (a) Catecholamines and metabolites
  - (b) Serotonin and metabolites

indication of the health of the bacteria, a pyruvate being an indirect measure of NADH, the reducing co-factor that stimulates bacterial growth. Malates and tartrates are also simultaneously monitored, being indicators of the maturity of the fermentation process. Lactose fermentation can yield lactic, acetic, pyruvic, formic, propionic, and butyric acids, which are indicators of the flavor and stability of dairy products. All these acids can be determined in one LC-analysis run.

In the food industry LC is frequently used to determine the quality of the finished product. For example, it has been shown that the flavor quality of grape juice is directly linked to the amount and spectrum of organic acids present. Likewise, in the production of sour cream by fermentation, the quality of the finished product quality is related directly to the organic acid profile. Last, but not least and as pointed out above, LC techniques are also crucial in determining the quality of recombinant protein products.

See also: **Carbohydrates:** Overview; Sugars – Chromatographic Methods. **Chromatography:** Overview; Principles. **Derivatization of Analytes.** **Electrophoresis:** Two-Dimensional Gels; Proteins. **Fluorescence:** Derivatization; Fluorescence Labeling. **Ion Exchange:** Overview; Principles; Isolation of Biopolymers. **Liquid Chromatography:** Overview; Principles; Column Technology; Mobile Phase Selection; Normal Phase; Reversed Phase; Size Exclusion; Chiral; Affinity Chromatography; Multidimensional; Instrumentation; Liquid Chromatography–Mass Spectrometry; Amino Acids; Clinical Applications. **Micro Total Analytical Systems.** **Peptides.** **Pharmaceutical Analysis:** Overview. **Process Analysis:** Chromatography; Bioprocess Analysis.

**Proteins:** Traditional Methods of Sequence Determination. **Proteomics.**

## Further Reading

- Chaplin MF (1986) Monosaccharides. In: Chaplin MF and Kennedy JF (eds.) *Carbohydrate Analysis – A Practical Approach*, pp. 37–69. Oxford IRL Press.
- El Rassi Z (ed.) (1995) *Carbohydrate Analysis*. Journal of Chromatography Library, vol. 58. Elsevier.
- Heftmann E (ed.) (1991) *Chromatography*, 5th edn., Journal of Chromatography Library, vol. 51. Elsevier.
- Horvath C and Nikelly JG (eds.) (1990) *Analytical Biotechnology*. ACS Symposium Series 434. Washington, DC: American Chemical Society.
- Kastner M (ed.) (2000) *Protein Liquid Chromatography*. Journal of Chromatography Library, vol. 67. Elsevier.
- Keller R, Lottspeich F, Meaer HE (eds.) (1999) *Microcharacterization of Proteins*. Wiley–VCH.
- Lloyd LL, Warner FP, and Kennedy JF (1990) The application of HPLC in biotechnology. In: *Biotechnology International*, pp. 221–226. London: Century Press.
- Schwarzenbach R (1982) High-performance liquid chromatography of carboxylic acids. *Journal of Chromatography* 251: 339–358.
- Sofer G and Hagel L (1997) *Handbook of Process Chromatography – A Guide to Optimization, Scale-up and Validation*. Academic Press.
- Tukova J (ed.) (1993) *Bioaffinity Chromatography*, 2nd completely revised edn., Journal of Chromatography Library, vol. 55. Elsevier.
- Weston A and Brown Ph (1997) *HPLC and CE – Principles and Practice*. Academic Press.
- White CA and Kennedy JF (1986) Oligosaccharides. In: Chaplin MF and Kennedy JF (eds.) *Carbohydrate Analysis – A Practical Approach*, pp. 37–69. Oxford IRL Press.

## Clinical Applications

**R A Sherwood**, King's College Hospital, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by R A Sherwood and B F Rocks, pp. 2677–2685, © 1995, Elsevier Ltd.

## Introduction

The diagnosis and monitoring of many disease states relies on the detection and quantitation of a particular compound(s) or metabolite(s) in body fluids, usually blood or urine. A wide variety of techniques are used in the clinical chemistry laboratory with spectrophotometric and immunoassay methods on automated analyzers measuring single compounds

predominating. In some instances a family of compounds may need to be measured and it is in these situations where the chromatographic techniques are important to avoid the necessity for a separate assay for each component of interest in the mixture.

Initially chromatography was carried out using paper or slab gels. The use of columns to contain the stationary phase came more recently with the advent of high-performance liquid chromatography (HPLC). Typical applications of HPLC in clinical chemistry include:

1. Biogenic amines
  - (a) Catecholamines and metabolites
  - (b) Serotonin and metabolites

2. Drugs
  - (a) Anticonvulsants
  - (b) Miscellaneous
3. Amino acids
4. Porphyrins
5. Vitamins
6. Carbohydrates
7. Glycohemoglobin

These applications will be described in greater detail in this article. Many other specific applications have been developed for particular projects, including separations of nucleosides and collagen breakdown products, but these have not commonly transferred from the research environment to the routine repertoire of tests.

Whilst 10–15 years ago the use of HPLC in the clinical laboratory was only found in the larger laboratory or where it was the particular interest of an individual, improvements in the equipment have permitted the extension of the technique to many more laboratories. In particular, HPLC with electrochemical detection is now the method of choice for the measurement of catecholamines and their metabolites and this has resulted in a much wider user base. HPLC does not readily lend itself to the analysis of large numbers of samples nor does it typically produce rapid results, although recent developments in the analysis of glycohemoglobins are changing this. The type of application most suited to a HPLC solution, therefore, has a small number of samples requiring a complex analysis without the need for a very rapid throughput. Reversed-phase chromatography is particularly suited to the separation of uncharged species. One of the problems encountered with the use of HPLC with clinical samples is that many of the compounds of interest are charged at physiological pH. This may be overcome by the use of ion-pairing agents in the extraction buffers and/or mobile phase.

## Sample Preparation for Biological Samples

The majority of applications of HPLC in clinical laboratories are based on the use of blood, plasma/serum, or urine samples, although feces and tissue samples are also used. The high protein content of blood samples ( $\sim 70 \text{ g l}^{-1}$  in plasma or serum) necessitates some form of sample pretreatment prior to injection onto the column to reduce background interference and to avoid column clogging. The specific preparation steps found most appropriate for the applications listed above will be described in the

relevant sections, but typically deproteinization and/or extraction of the compounds of interest is carried out. Both liquid–liquid and solid-phase extractions are used in the clinical laboratory. When urine is used, the presence of colored/fluorescent pigments in the sample may cause difficulties when ultraviolet (UV) or fluorescence detection is used and some form of extraction is usually required to eliminate this interference.

## Biogenic Amines

**Catecholamines and their metabolites** The catecholamines, adrenaline, noradrenaline, and dopamine are essential components of the central nervous system acting as neurotransmitters both within the brain and at peripheral nerves. All are synthesized in the adrenal medulla from phenylalanine or tyrosine and are metabolized by a mixture of enzymatic side chain oxidation and methylation of the hydroxy groups on the ring. If the metabolism is complete, adrenaline and noradrenaline are degraded to 4-hydroxy-3-methoxy mandelic acid (HMMA, commonly called vanillylmandelic acid – VMA), while dopamine is broken down to homovanillic acid (HVA). Urinary excretion of these metabolites and their conjugates is the major route of elimination of catecholamines from the body, although small amounts are excreted unchanged as the free catecholamines.

Measurement of the catecholamines and a variety of the intermediate metabolites in blood or tissue is common in psychiatric and neurochemical research. In clinical chemistry the major interest in catecholamine metabolism surrounds the detection and location of the tumors of neural crest origin – pheochromocytoma and neuroblastoma. These tumors are fortunately rare but their identification is important because if treated promptly they may be curable. Pheochromocytoma is associated with secretion of adrenaline and/or noradrenaline into the bloodstream and can either be detected by the increase in the parent compounds or by increased VMA excretion. The most common presentation by the patient is hypertension unresponsive to conventional therapy. Neuroblastoma is the commonest malignant soft tissue tumor of childhood, arising from ectodermal neuroblasts. These tumors secrete abnormally high concentrations of dopamine which is largely metabolized to HVA.

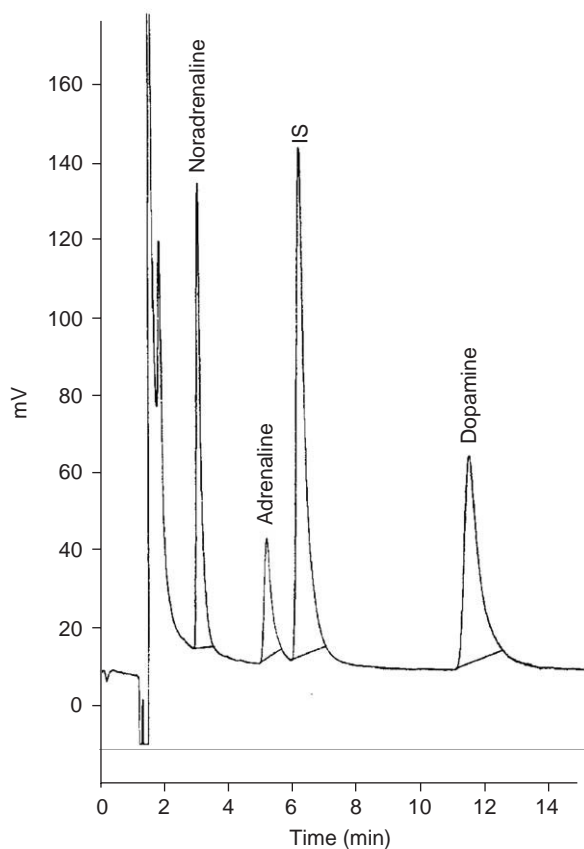
Various strategies for the identification of these tumors have been employed. Prior to the advent of reliable HPLC methods, thin-layer chromatography (TLC), or colorimetric assays for VMA and HVA in urine were the most frequently used

measurements, although fluorimetric assays for catecholamines were also utilized by some laboratories. Plasma catecholamines, urine catecholamines, and the metabolites can be measured by HPLC. Two detection techniques, fluorimetric (with and without derivatization) and electrochemical (both amperometric and coulometric), provide the required sensitivity and specificity for quantitation of the catecholamines.

**Plasma catecholamines** The catecholamines are present in human blood plasma in the low nanomolar range, hence some preconcentration, extraction, and sample cleanup of endogenous electrophores is required prior to analysis. The catecholamines from 1 to 3 ml of plasma are adsorbed onto alumina and then eluted into 100  $\mu$ l of acid to give a 10–30-fold increase in concentration. Preparation of trihydroxyindole derivatives provides sufficient sensitivity for fluorescent detection. Reversed-phase (C18) columns with addition of an anionic ion-pair reagent or specific ion-exchange columns are suitable, producing separation of all three catecholamines within 15–20 min with isocratic elution. Electrochemical detection provides acceptable specificity with an oxidation potential of +0.75 V when using amperometry.

**Urine catecholamines** The catecholamines are present in urine in low micromolar concentrations, 1000-fold higher than in plasma, their measurement therefore being considerably easier than in plasma. A wide variety of extraction methods have been published including both liquid–liquid and solid-phase extraction. Many of the liquid–liquid extraction methods are based on the formation of complexes between the catecholamines and diphenylborate in an organic solvent with subsequent back-extraction into acid. Ion pairing with tetrabutylammonium bromide or a similar ion-pairing reagent produces good resolution of the catecholamines when a C18 reversed-phase column is used (Figure 1). Dihydroxybenzylamine is suitable as an internal standard as it behaves in a similar manner to the catecholamines in most extraction procedures and can be resolved from the majority of interfering substances found in urine.

**Urinary metabolites VMA and HVA** The metabolites are present in urine at much higher concentrations than the catecholamines – 10–50  $\mu$ mol l<sup>-1</sup>. They are readily extracted from acidified urine into ethyl acetate which can then be evaporated off and the residue dissolved in the mobile phase of choice. VMA is only weakly retained on C18 columns and

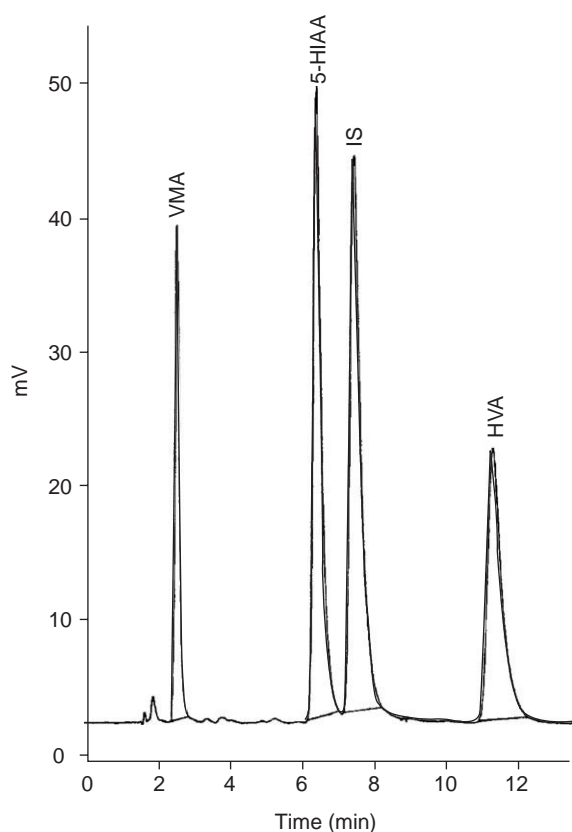


**Figure 1** Chromatographic profile of a urinary extract of catecholamines using a C18 reversed-phase column. The catecholamines are extracted into a hexane:octanol mixture by complex formation with diphenylborate in alkaline solution and ion-pair formation with tetrabutylammonium bromide. The catecholamines are back-extracted into hydrochloric acid. Mobile phase: phosphate buffer pH 5.5/methanol (87.5:12.5). Flow rate 1.2 ml min<sup>-1</sup>. Detection: electrochemical oxidation at +0.65 V vs Ag/AgCl with glassy carbon electrode. Column: ODS Hypersil 5  $\mu$ m, 15 cm  $\times$  4.6 mm.

careful choice of conditions is required to resolve it from the solvent front and other early eluting compounds. Phosphate buffer (pH 3.0) modified by the addition of methanol (15–20%) resolves VMA adequately without producing unacceptably long retention times for HVA (Figure 2). Virtually all the methods described utilize electrochemical detection with an oxidation potential between +0.65 and +0.75 V.

**The indoleamines** The indole of greatest significance in human metabolism is serotonin (5-hydroxytryptamine (5-HT)) which is synthesized from tryptophan and metabolized by the monoamine oxidase system to 5-hydroxyindole acetic acid (5-HIAA). Serotonin is believed to play a significant role in maintenance of mood and many antidepressant drugs act via the serotonergic receptor system.





**Figure 2** Chromatographic profile of a urinary extract of catecholamine and indoleamine metabolites using a C18 reversed-phase column. VMA, HVA, and 5-HIAA are extracted with ethyl acetate, evaporated to dryness, and redissolved in mobile phase. Mobile phase: phosphate buffer pH 3.0/methanol (82:18). Flow rate  $1.0 \text{ mL min}^{-1}$ . Detection: electrochemical oxidation at  $+0.75 \text{ V}$  vs Ag/AgCl with glassy carbon electrode. Column: ODS Hypersil  $5 \mu\text{m}$ ,  $15 \text{ cm} \times 4.6 \text{ mm}$ . The internal standard (IS) is 5-hydroxy-2-indolecarboxylic acid.

Carcinoid tumors arising from the argentaffin cells of the gastrointestinal (GI) tract secrete excessive amounts of serotonin, the local action of which in the gut provokes diarrhoea. Measurement of serotonin or 5-HIAA is useful in the detection of carcinoid tumors and in monitoring therapy.

**Serotonin** In blood serotonin is predominantly located in the platelets, with levels in plasma being  $\sim 100$ -fold lower. Most analysts interested in serotonin measurements have concentrated on determining serotonin in whole blood or platelet rich plasma. The presence of monoamine oxidase enzymes in blood necessitates the use of an enzyme inhibitor, e.g., pargyline to prevent the conversion of serotonin to 5-HIAA prior to analysis. Similar chromatographic conditions to those used for catecholamines can be used, e.g., C18 reversed-phase columns and electrochemical detection at an oxidation potential of  $+0.75 \text{ V}$  or fluorescence detection. Sample

preparation is complicated by the use of whole blood as the sample matrix, but both protein precipitation methods and solid-phase extraction have been used successfully. Measurement of serotonin in platelet poor plasma is much more difficult due to the much lower concentrations and protein precipitation or direct injection methods have been described.

**5-HIAA** The Urinary concentration of 5-HIAA is of the same order as VMA and HVA, e.g.,  $10\text{--}50 \mu\text{mol l}^{-1}$ . Many of the methods developed for urine VMA and HVA can also be used for the assay of 5-HIAA (Figure 2). Relatively few methods have been reported for the determination of 5-HIAA in plasma as both the concentration and the difficulties are similar to those encountered for HVA and 3-methoxy-4-hydroxyphenylglycol (MHPG). Both fluorimetric and amperometric methods have been described.

## Drugs

Drug analysis is of interest to clinical chemists for two main purposes; the identification and quantitation of drugs used for recreational purposes and the monitoring of therapeutic drugs. The former is usually confined to the investigation of patients suspected of drug misuse or where there is a possibility of accidental or deliberate overdose. Toxicological analysis is a specialized area and will not be discussed further in this article. Therapeutic drug monitoring (TDM) has become a routine part of most clinical chemistry laboratory repertoires particularly for anticonvulsant, antiasthmatic, and antiarrhythmic drugs. A mixture of immunoassay and chromatographic methods can be found for these drugs with individual laboratory's preferences being defined by a number of factors including: cost, availability of equipment, staff resources, and required turnaround times. Assay of the antibiotics usually falls within the remit of microbiology rather than clinical chemistry departments with immunoassay techniques being used in most instances.

The general principles for selecting whether TDM is appropriate for a particular drug are based on the occurrence of interindividual variation in the rate and extent of drug absorption, distribution, metabolism, and excretion of the drug. The dosage required to maintain an effective serum concentration may vary considerably between individuals particularly for drugs with a narrow therapeutic index. Plasma or serum drug concentrations can therefore help in establishing initial dosage regimens and in making adjustments to optimize therapy.



**Anticonvulsants** Anticonvulsants were one of the first groups of drugs for which TDM was indicated. The common anticonvulsants encountered in clinical practice are: phenobarbitone, primidone (metabolized to phenobarbitone), phenytoin, carbamazepine, ethosuximide, and valproate. The majority of assays for these drugs are immunoassays these days as 24 h availability of measurements of these drugs has become the general expectation.

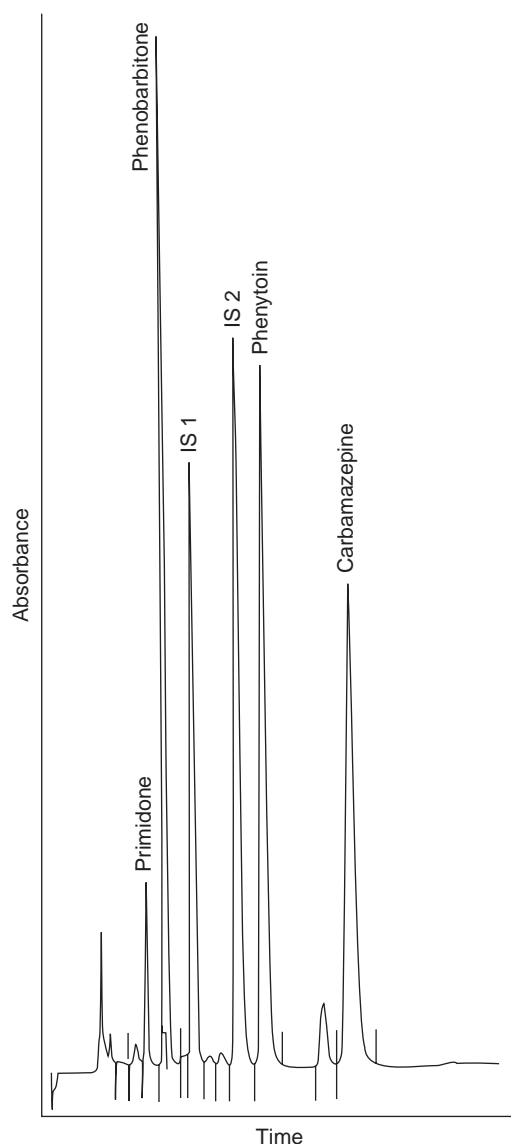
The major disadvantage of immunoassays is the need to carry out one assay for each drug or drug metabolite of interest. For example, carbamazepine has a metabolite carbamazepine-10-11-epoxide which may be biologically active and which may need to be measured on occasion. HPLC can be used to profile the first four to five of the listed drugs in a single analytical run (Figure 3). Numerous liquid-liquid or solid-phase extractions have been described for the anticonvulsants, the difficulty being to obtain acceptable recoveries from the differing compounds. Reversed-phase chromatography on C18 columns has been almost universally used with an isocratic methanol/buffer system. Detection is relatively easy with strong UV maxima between 200 and 250 nm.

There is a new generation of anticonvulsants for which the need for TDM is currently being determined including: lamotrigine, gabapentin, vigabatrin, topiramate, and leviteracetam. For gabapentin and vigabatrin the plasma concentration has little relationship with clinical efficacy, and TDM is only useful for assessing compliance. For the others there may be other indications for TDM. All can be measured by reversed-phase HPLC, usually with UV detection.

**Miscellaneous drugs** The number of drugs that can be measured in human samples runs into thousands. Most of these are only measured for specific projects or initial pharmacokinetic studies. The other types of drug routinely measured for TDM purposes include:

- antidepressants, e.g., amitriptyline, clozapine, etc.
- immunosuppressive drugs, e.g., cyclosporin A
- cardioactive drugs, e.g., amiodarone

The above-mentioned examples all have active or inactive metabolites which can be measured by HPLC, but not by typical immunoassay methods. Where the metabolites are active, their measurement can be beneficial in interpreting the results. For cyclosporin A up to 21 metabolites are known, but the biological activity of most of these is uncertain. HPLC is the reference method for the parent



**Figure 3** Chromatogram of a plasma extract showing separation of the anticonvulsants primidone, phenobarbitone, phenytoin, and carbamazepine by isocratic reversed-phase chromatography. Solid-phase extraction with a C18 column was used for sample preparation. Methanol (45%) in dilute acid is used as the mobile phase. Flow rate  $1.0 \text{ ml min}^{-1}$ . Detection: UV absorption at 225 nm. Column: ODS Hypersil  $5 \mu\text{m}$ ,  $15 \text{ cm} \times 4.6 \text{ mm}$ . Two internal standards are used; IS 1 is 4-methyl primidone and IS 2 is 5-ethyl-5-*p*-tolylbarbituric acid.

compound as most antibodies currently available have some cross-reactivity with the metabolites. Mass spectrometry (MS) has been used as a mode of detection in combination with gas chromatography for many years. The large fluid volumes in HPLC in comparison prevented the combination of HPLC and MS initially. The introduction of microbore HPLC and developments in the production of very fine liquid aerosols (electrospray) have resulted in

LC-MS becoming widely available. LC-MS or tandem MS (MS-MS) may replace HPLC alone as the standard methodology for the immunosuppressive drugs.

### Amino Acids

Amino acid analysis in clinical chemistry is of importance in the diagnosis and monitoring of the inherited disorders of amino acid metabolism and is also of value in many nutritional studies. Qualitative or semiquantitative identification of the most important amino acids may be achieved using TLC but quantitation is often necessary particularly for the monitoring of therapy in disorders such as phenylketonuria or maple syrup urine disease. The traditional amino acid analyzer (the first specialist HPLC system) is based on separation of the amino acids by cation-exchange chromatography with spectrophotometric detection following postcolumn derivatization with ninhydrin. Although this approach is still the commonest technique found in clinical chemistry laboratories, it has relatively poor sensitivity and long analysis time (up to 4 h) and HPLC methods are now beginning to appear that could replace it.

The presence of large amounts of protein can have a major effect on the determination of free amino acids in plasma and samples must be deproteinized prior to analysis. Protein precipitation with either acid or organic solvent may be used, although care should be taken when acids are used to avoid interference in the chromatography from the acid itself.

To achieve increased sensitivity and reduced analysis time it is necessary to produce derivatives of the amino acids which can be separated by reversed-phase HPLC and then detected using fluorescent, spectrophotometric, or electrochemical detection. Both precolumn and postcolumn derivatization methods have been described with a range of derivatizing agents including:

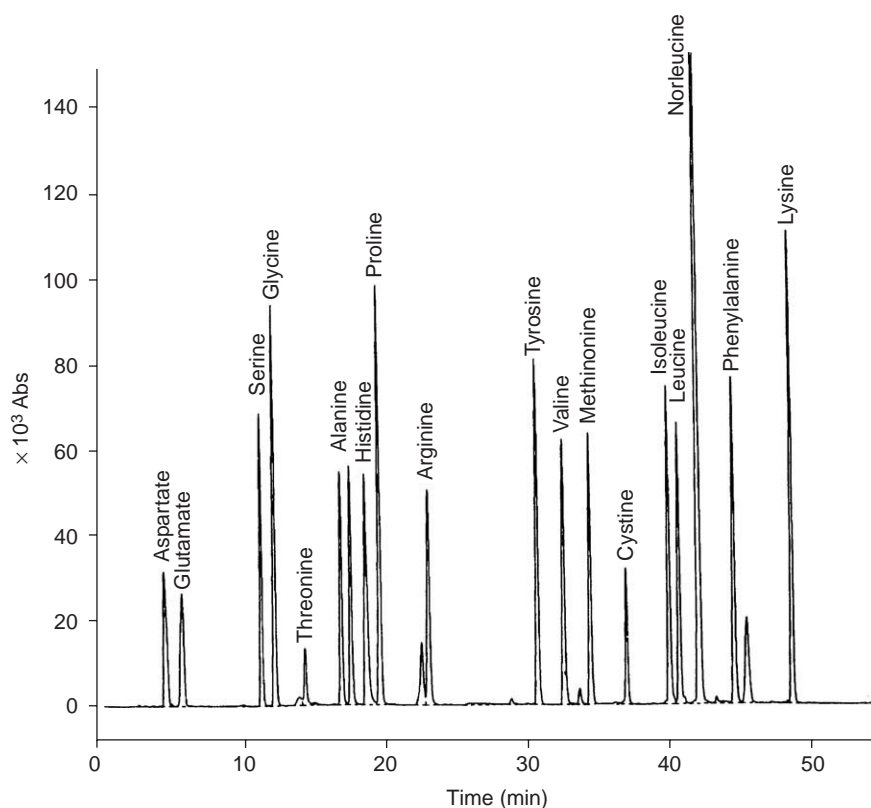
- dabsyl chloride,
- dansyl chloride,
- 9-fluorenylmethyl chloroformate (FMOC),
- *o*-phthalaldehyde (OPA), and
- phenylisothiocyanate (PITC).

The first four derivatizing agents listed above produce products suitable for fluorescent detection while PITC derivatives can be measured spectrophotometrically or electrochemically. Both dansyl and dabsyl chloride must be added in considerable excess to give reasonable sensitivity necessitating the removal of

the excess sulfonic acid by-products which cause interference if allowed to remain. A similar problem occurs with the use of FMOC as it forms fluorenylmethyl alcohol which must be removed prior to chromatography. Until recently OPA was the most commonly used derivatizing agent for HPLC amino acid analysis. In the presence of a thiol reagent such as mercaptoethanol, OPA, which is itself not fluorescent, reacts with primary amines to form highly fluorescent isoindoles well suited to reversed-phase separation. The OPA derivatives with aliphatic amines are unstable but the use of automated sample preparation systems has made this less of a disadvantage and stable derivatives with OPA and *N*-acetyl-cysteine have now been described. A much more significant disadvantage for clinical chemists is the nonreactivity of OPA with the imino acids proline and hydroxyproline and the low sensitivity to cysteine. Several approaches have been reported to overcome this problem including the postcolumn addition of hypochlorite and a dual derivatization method using both OPA and FMOC.

The Edman reagent, PITC, has for many years been used in protein sequencing and has more recently been adapted for amino acid analysis in body fluids. The phenylthiocarbamyl derivatives formed can be detected either by their UV absorption at 254 nm or by electrochemical detection at an oxidation potential of +1.1 V. Derivatives are formed with both primary and secondary amino acids although sensitivity to cysteine and homocysteine is lower than for the other amino acids. Use of gradient elution with a mobile phase containing acetate buffer and acetonitrile permits resolution of the majority of the physiological amino acids within 50 min (**Figure 4**), although the time required for reequilibration of the system results in a total cycle time of 70–80 min. Modification of the gradient profile can significantly reduce the overall run time if only a single amino acid or related group of amino acids are required. For example a total run time of 25 min can be achieved for the measurement of glycine and hydroxyproline. The derivatives are stable at 4°C for several days and for longer periods at –20°C. The derivatization process is relatively simple but cannot be readily automated. Very few compounds found in body fluids interfere, although some of the cephalosporin antibiotics produce peaks as does the anticonvulsant vigabatrin (a gamma-amino-butyric acid analog). Quantitation is reproducible with between batch CVs of 8–12% for most amino acids.

As for the immunosuppressive drugs, the advent of affordable and reliable tandem MS has meant that many laboratories providing a neonatal screening



**Figure 4** Chromatogram of PITC derivatized amino acid standards. Initially the mobile phase is a phosphate buffer, pH 6.2, containing 0.7% acetonitrile. A linear gradient is used to increase the acetonitrile concentration to 70%. The internal standard is norleucine. Flow rate  $1.0 \text{ ml min}^{-1}$ . Detection is at 254 nm. Column: ODS Hypersil  $5 \mu\text{m}$ ,  $25 \text{ cm} \times 4.6 \text{ mm}$ .

service for phenylketonuria are evaluating, changing from TLC/HPLC to LC-MS methods.

### Porphyryns

The porphyrin family of compounds are cyclic tetrapyrroles many of which are intermediates of the heme biosynthetic pathway. The majority of the compounds of interest to the clinical chemist are not in fact porphyrins (with the exception of protoporphyrin) but porphyrinogens in which all four methylene bridges are in the reduced form. The significance of the porphyrins in medicine relates to the class of inherited disorders – the porphyrias – where there is accumulation of the precursors due to an enzyme deficiency in the heme pathway. The clinical presentation of the porphyrias varies from a chronic photosensitivity to severe acute abdominal pain as in acute intermittent porphyria, which may be precipitated by exposure to certain drugs. Characterization of the type of porphyria depends upon the identification of particular patterns of porphyrins in blood, urine, and feces.

Early methods for porphyrin HPLC were based on the formation of methyl esters followed by normal

phase chromatographic separation. These had the disadvantage of not distinguishing between the types I and III porphyrin isomers and with the advent of reversed-phase systems for porphyrin separation are only used for unusual cases today. Porphyrins can be extracted and concentrated from urine by adsorption with acidified acetone. An acid/ether extraction will separate the porphyrins in feces from interfering chlorophyll or carotenoid pigments, whilst ethyl acetate-acetic acid is the preferred extraction method for red blood cell or plasma porphyrins. Once extracted the porphyrin free acids can be separated using a C18 reversed-phase column with an ammonium acetate-acetonitrile mobile phase. Varying the proportion of acetonitrile in the mobile phase from 13% to 30% permits the separation of particular groups of the porphyrins with the uroporphyrins eluting at the lower acetonitrile concentration and the coproporphyrins at the higher concentration. Gradient elution using a linear gradient of acetonitrile-ammonium acetate will resolve the 17 most important porphyrins including the types I and III isomers within 40 min. Detection of porphyrins can be achieved by measuring their UV absorption at

404 nm or by fluorescence emission at 580 nm after excitation at 395 nm.

## Vitamins

The technique of HPLC has had a marked impact on the assay of vitamins in body fluids. Whilst linked by function the vitamins are a diverse group of compounds, but virtually all have been measured in blood or urine by HPLC. All the fat soluble vitamins, A, D, E, and K, are measurable by HPLC as are most of the water-soluble B vitamins and vitamin C. In some cases, notably folate, B<sub>12</sub> and vitamin D immunoassay methods are routinely used to achieve sufficient sensitivity for clinical purposes and these vitamins have not been discussed further in this article.

**Vitamin A and the carotenoids** The majority of the dietary provitamin carotenoids, e.g., B-carotene, are oxidatively cleaved to vitamin A (retinol), although an appreciable fraction circulates intact in plasma. Retinol in plasma forms a complex with retinol-binding protein and it is this that is usually measured as an indication of vitamin A status in individuals. The complex can be disrupted by ethanol and retinol or B-carotene can then be extracted into hexane. Separation can be achieved with C18 reversed-phase columns with elution by methanol. Detection of retinol may be achieved either from its UV absorption at 325 nm or its native fluorescence, while B-carotene absorbs in the visible region.

## The B vitamins

**Thiamin (vitamin B<sub>1</sub>)** Thiamin in the body is chiefly found in the phosphorylated form thiamin pyrophosphate (TPP) which is a coenzyme. The majority (80%) of thiamin in the blood is found in the erythrocytes and assay of blood thiamin is a more reliable indicator of deficiency than assay of erythrocyte transketolase. The phosphorylated vitamers are enzymically converted to thiamin in samples using diastase following deproteinization. To reach the low picomolar concentrations the thiamin compounds are oxidized by ferricyanide to form thiochromes, which are highly fluorescent. The thiochromes are then separated by reversed-phase HPLC and detected by their emission at 425–450 nm.

**Riboflavin (vitamin B<sub>2</sub>)** Riboflavin exists primarily as the coenzyme forms flavin mononucleotide and flavin adenine dinucleotide. Not commonly measured, both can be measured by their natural fluorescence at 530 nm following reversed-phase

separation. Urine may be injected directly, while in other fluids deproteinization is necessary.

**Nicotinic acid and nicotinamide** These are sometimes known by the generic term niacin. Their importance is in combination with tryptophan, as the coenzyme forms nicotinamide adenine dinucleotide (NAD<sup>+</sup>) and nicotinamide adenine dinucleotide phosphate (NADP). HPLC is too insensitive to measure endogenous plasma levels, but the urinary metabolites *N*-methyl-2-pyridone-5-carboxylamide and *N'*-methylnicotinamide can be measured to assess niacin status. Preliminary cleanup of urine by anion-exchange resins is followed by reversed-phase HPLC with UV detection.

**Vitamin B<sub>6</sub>** Vitamin B<sub>6</sub> is the generic name for six naturally occurring vitamers which are derivatives of 2-methyl-3-hydroxy pyridone. All forms of vitamin B<sub>6</sub> have been detected in human plasma with the exception of pyridoxine phosphate. Both free and protein bound forms can be extracted with trichloroacetic acid or perchloric acid, the excess of which should be removed prior to chromatography. Cation-exchange chromatography with linear gradients of hydrochloric acid and phosphate buffers produces a profile of the vitamin B<sub>6</sub> compounds in 50 min.

**Vitamin C** Vitamin C activity resides in two naturally occurring compounds; ascorbic acid and its oxidation product, dehydroascorbic acid. In human tissues ascorbic acid predominates. Ascorbic acid is labile in most samples, oxidizing to dehydroascorbic acid and then degrading to 2,3-diketogluconic acid. Various reagents can be used to prevent this oxidation in plasma or whole blood samples. Extraction with 5% metaphosphoric or trichloroacetic acid is the usual initial preparation. Only ascorbic acid may be detected by UV spectrophotometry at 245–265 nm, the absorption maxima of dehydroascorbic acid being 210 nm. A similar problem exists with electrochemical detection where ascorbic acid oxidizes at +0.7 V with carbon electrodes. Fluorescent derivatives may be formed with 2-4-dinitrophenylhydrazine or *o*-phenyldiamine. These derivatives can be assayed by reversed-phase HPLC.

**Vitamin E** In man  $\alpha$ -tocopherol is the major form of vitamin E activity found in plasma and cells with  $\beta$ - and  $\gamma$ -tocopherol being minor components all of which are bound to lipoproteins. With a plasma concentration of 5–15 mg l<sup>-1</sup>,  $\alpha$ -tocopherol has the highest plasma concentration of any of the fat soluble vitamins. For the separation of the tocopherols, reversed-phase HPLC can be used. Though the molar



extinction coefficient of vitamin E in the UV region is relatively low, the plasma concentrations are sufficiently high to allow UV detection. Tocopherols exhibit native fluorescence and can also be detected electrochemically at an oxidation potential of  $+0.7\text{ V}$ .

**Vitamin K** HPLC has provided the first assay of the phyloquinones and menaquinones that constitute vitamin K in plasma. Phyloquinone circulates bound to lipoproteins from which it can be extracted with hexane after ethanol protein precipitation. Removal of co-eluted lipids can be achieved with normal-phase cartridge columns. Reversed-phase HPLC is almost universally used for vitamin K measurement. Either UV (270 nm) or electrochemical detection is suitable. Electrochemical detection uses the reductive mode ( $-1.3\text{ V}$ ) to convert the quinone moiety to hydroquinone; the main disadvantage being the need to remove oxygen from the mobile phase.

### Carbohydrates

Apart from the measurement of glucose, perhaps one of the most frequently undertaken tests in medicine, the need to quantitatively assay the many carbohydrates in the body has fortunately been uncommon. TLC methods have sufficed for the screening of samples for the inherited disorders of carbohydrate metabolism such as galactosemia or fructosuria. Recent developments of multiple sugar absorption tests to assess small bowel permeability have necessitated a reassessment of the available methods for separation and quantitation of specific carbohydrates in urine. Lactulose, rhamnose, xylose, and 3-*o*-methylglucose cross the wall of the gut via different mechanisms, and consumption of a solution containing these sugars combined with measurement of the percentage of each sugar excreted in urine over a 5 h period can assist in the differential diagnosis of GI tract disease.

The carbohydrates are a problematical group of compounds from the point of view of detection, having neither useable spectrophotometric or fluorescent properties. The method of choice for detection of carbohydrates in pure solution has been refractometry but interference from other compounds in body fluids and its relative insensitivity has limited its application in clinical chemistry. The development of pulsed amperometric detection at noble metal electrodes and improvements in anion-exchange columns has now made possible the separation and quantitation of most carbohydrates of interest in urine samples. As ion-exchange columns require a constant ionic concentration to achieve consistent separations, urines are first desalted with

Amberlite resin prior to injection into a dilute sodium hydroxide mobile phase. This provides the alkaline environment suitable for pulsed amperometry, although if reversed-phase chromatography provides better resolution of the carbohydrates under study, postcolumn addition of hydroxide can be carried out.

### Glycohemoglobin

There are relatively few applications of HPLC to protein separations in clinical chemistry, electrophoresis being the preferred technique. An exception to this is the separation of glycohemoglobins using weak cation-exchange chromatography. Separation of the glycosylated fraction HbA1c from the nonglycosylated HbA1a and HbA1b can be achieved using HPLC, and dedicated automated systems are available with analytical cycle times of 4–8 min; suitable for use in the clinic setting. A disadvantage of the ion-exchange chromatographic methods is their inability to separate the abnormal hemoglobins, HbS and HbC and their glycosylated equivalents, from the normal HbA and HbA1c peaks without significantly lengthening the run time. A similar automated column chromatography method is available based on the principle of affinity chromatography. The glucose residues on the hemoglobin bind to a boronate agarose affinity gel and are eluted subsequently with sorbitol. This method is not affected by the charge on the hemoglobin and can measure the glycated fractions of the variant hemoglobins. A reference method has been established which involves cleavage of hemoglobin into peptides by the endoproteinase Glu-C followed by separation and quantitation of the hexapeptides by HPLC–electrospray ionization mass spectrometry (ESI-MS) or HPLC–capillary electrophoresis (HPLC–CE).

*See also:* **Carbohydrates:** Sugars – Spectrophotometric Methods; Sugars – Chromatographic Methods. **Clinical Analysis:** Overview. **Forensic Sciences:** Systematic Drug Identification. **Liquid Chromatography:** Amino Acids. **Vitamins:** Overview; Fat-Soluble; Water-Soluble.

### Further Reading

- Baldwin RP (1999) Electrochemical determination of carbohydrates: Enzyme electrodes and amperometric detection in liquid chromatography and capillary electrophoresis. *Journal of Pharmacological and Biomedical Analysis* 19: 69–81.
- Binder SR (1996) Analysis of drugs of abuse in biological fluids by liquid chromatography. *Advances in Chromatography* 36: 201–271.
- Chace DH (2001) Mass spectrometry in the clinical laboratory. *Chemical Review* 101: 445–477.



- Garcia Alvarez-Coque MC and Carda Broch S (1999) Direct injection of physiological fluids in micellar liquid chromatography. *Journal of Chromatography B Biomedical Science Applications* 736: 1–18.
- Ghosh MK (1992) *HPLC Methods on Drug Analysis*. Berlin: Springer-Verlag.
- Honour JW (2003) Benchtop mass spectrometry in clinical biochemistry. *Annals of Clinical Biochemistry* 40: 628–638.
- Jeppsson JO, Kobold U, Barr J, *et al.* (2002) Approved IFCC reference method for the measurement of HbA1C in human blood. *Clinical Chemistry and Laboratory Medicine* 40: 78–89.
- Maurer HH (1998) Liquid chromatography–mass spectrometry in forensic and clinical toxicology. *Journal of Chromatography B Biomedical Science Applications* 713: 3–25.
- Sherwood RA, Richards DA, and Titheradge AC (1990) Measurement of plasma and urine amino acids by high performance liquid chromatography with electrochemical detection using PITC derivatisation. *Journal of Chromatography Biomedical Applications* 528: 293–303.
- Walker V and Mills GA (2002) Solid phase extraction in clinical biochemistry. *Annals of Clinical Biochemistry* 39: 464–477.
- Willemsen JJ, Ross HA, Wolthers BG, Sweep CGJ, and Kema JP (2001) Evaluation of specific high-performance liquid-chromatographic determinations of urinary metanephrene and normetanephrene by comparison with isotope dilution mass spectrometry. *Annals of Clinical Biochemistry* 38: 722–730.

## Food Applications

**E Ibañez**, Instituto de Fermentaciones Industriales, CSIC, Madrid, Spain

**F J Señoráns**, University Autonoma Madrid, Madrid, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Liquid chromatography (LC) has become a very useful tool in the area of food analysis. Its use and development have increased in parallel with consumers' concern about the composition and the safety of food. Moreover, the Nutrition Labeling and Education Act in the United States and the new regulations in the European Union countries (E.G. Regulation EC 258/97 or EN 29000 and subsequent issues) have had a major impact on food laboratories, consequently increasing their requirements in terms of faster, more powerful, and cleaner procedures to analyze foods. Analytical techniques such as LC must address a number of important problems and must provide fast and accurate information about processing and food quality while ensuring compliance with laws relating to food, trade, adulteration, contamination, food tampering, and chemical composition of foods. All these requirements have increased interest in LC. The reasons for its popularity are easy to find: wide range of applicability, allowing the separation of compounds ranging from ions to biopolymers, mild separation conditions (low temperature), which preclude the decomposition of thermally labile compounds such as vitamins, good sensitivity and resolution, and a wide range of separation mechanisms, which favor the separation of different types of compounds.

High-performance LC (HPLC) is the technique used most frequently in food analysis for measuring carbohydrates, vitamins, additives (sweeteners, antioxidants, colorants, preservatives, etc.), mycotoxins, amino acids, proteins, triglycerides in fats and oils, lipids, chiral compounds and pigments, among others (Table 1). Some of these applications will be discussed in this article.

One of the main problems that LC has to encounter is the complexity of the samples that have to be analyzed, with the selected analyte being usually present in very low levels. There are several ways of overcoming this problem in LC, e.g., increasing the specificity of the separation method and improving

**Table 1** Compound classes typically analyzed using HPLC

Food additives	Acidulants
	Antioxidants
	Preservatives
	Artificial sweeteners
	Colorants
Food residues and contaminants	Flavors
	Residues of chemotherapeutics
	Antiparasitic drugs
	Mycotoxins
	Pesticides
Lipids	Triglycerides
	Hydroperoxides
	Fatty acids
Carbohydrates	
Vitamins	Fat-soluble vitamins
	Water-soluble vitamins
Amino acids	
Peptides	
Proteins	

- Garcia Alvarez-Coque MC and Carda Broch S (1999) Direct injection of physiological fluids in micellar liquid chromatography. *Journal of Chromatography B Biomedical Science Applications* 736: 1–18.
- Ghosh MK (1992) *HPLC Methods on Drug Analysis*. Berlin: Springer-Verlag.
- Honour JW (2003) Benchtop mass spectrometry in clinical biochemistry. *Annals of Clinical Biochemistry* 40: 628–638.
- Jeppsson JO, Kobold U, Barr J, *et al.* (2002) Approved IFCC reference method for the measurement of HbA1C in human blood. *Clinical Chemistry and Laboratory Medicine* 40: 78–89.
- Maurer HH (1998) Liquid chromatography–mass spectrometry in forensic and clinical toxicology. *Journal of Chromatography B Biomedical Science Applications* 713: 3–25.
- Sherwood RA, Richards DA, and Titheradge AC (1990) Measurement of plasma and urine amino acids by high performance liquid chromatography with electrochemical detection using PITC derivatisation. *Journal of Chromatography Biomedical Applications* 528: 293–303.
- Walker V and Mills GA (2002) Solid phase extraction in clinical biochemistry. *Annals of Clinical Biochemistry* 39: 464–477.
- Willemsen JJ, Ross HA, Wolthers BG, Sweep CGJ, and Kema JP (2001) Evaluation of specific high-performance liquid-chromatographic determinations of urinary metanephrene and normetanephrene by comparison with isotope dilution mass spectrometry. *Annals of Clinical Biochemistry* 38: 722–730.

## Food Applications

**E Ibañez**, Instituto de Fermentaciones Industriales, CSIC, Madrid, Spain

**F J Señoráns**, University Autonoma Madrid, Madrid, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Liquid chromatography (LC) has become a very useful tool in the area of food analysis. Its use and development have increased in parallel with consumers' concern about the composition and the safety of food. Moreover, the Nutrition Labeling and Education Act in the United States and the new regulations in the European Union countries (E.G. Regulation EC 258/97 or EN 29000 and subsequent issues) have had a major impact on food laboratories, consequently increasing their requirements in terms of faster, more powerful, and cleaner procedures to analyze foods. Analytical techniques such as LC must address a number of important problems and must provide fast and accurate information about processing and food quality while ensuring compliance with laws relating to food, trade, adulteration, contamination, food tampering, and chemical composition of foods. All these requirements have increased interest in LC. The reasons for its popularity are easy to find: wide range of applicability, allowing the separation of compounds ranging from ions to biopolymers, mild separation conditions (low temperature), which preclude the decomposition of thermally labile compounds such as vitamins, good sensitivity and resolution, and a wide range of separation mechanisms, which favor the separation of different types of compounds.

High-performance LC (HPLC) is the technique used most frequently in food analysis for measuring carbohydrates, vitamins, additives (sweeteners, antioxidants, colorants, preservatives, etc.), mycotoxins, amino acids, proteins, tryglicerides in fats and oils, lipids, chiral compounds and pigments, among others (Table 1). Some of these applications will be discussed in this article.

One of the main problems that LC has to encounter is the complexity of the samples that have to be analyzed, with the selected analyte being usually present in very low levels. There are several ways of overcoming this problem in LC, e.g., increasing the specificity of the separation method and improving

**Table 1** Compound classes typically analyzed using HPLC

Food additives	Acidulants
	Antioxidants
	Preservatives
	Artificial sweeteners
	Colorants
Food residues and contaminants	Flavors
	Residues of chemotherapeutics
	Antiparasitic drugs
	Mycotoxins
	Pesticides
Lipids	Triglycerides
	Hydroperoxides
	Fatty acids
Carbohydrates	
Vitamins	Fat-soluble vitamins
	Water-soluble vitamins
Amino acids	
Peptides	
Proteins	

the selectivity of both the sample preparation prior to LC analysis and the LC analysis itself.

In terms of the specificity of the separation method, it is possible to select the most appropriate separation mechanism by choosing from a wide range of stationary phases, detectors, and mobile phases. Thus, not only can elution power be altered by changes in mobile-phase composition but selectivity can also be changed using mobile-phase additives such as chiral compounds or ion pair reagents. A new and very specific type of LC is affinity chromatography, which provides high selectivity and specificity for separating compounds able to form a complex with a ligand (immobilized on the chromatographic support), thus allowing specific determination of the major compounds, such as proteins, lipids, and polysaccharides, as well as minor biologically active components, such as drugs, hormones, and toxins. The highest specificity is provided by the antibody-antigen interaction (which is called immunochromatography), which has been applied in food analysis to detect small concentrations of aflatoxins in peanuts. For even more demanding separations, new molecularly imprinted polymers are being synthesized and used to perform chiral and affinity separations.

As for sample preparation, it is widely accepted that this is one of the most important steps in warranting the quality of the analysis. This step becomes critical when analyzing organic substances in complex matrices like foods. In most cases, clean-up and concentration of the analytes is required prior to separation using LC. Among the useful sample preparation techniques, solid-phase extraction (SPE) is probably the most widely used. SPE is based on the selective retention of the compounds of interest on a sorbent placed in a disposable extraction minicolumn (or cartridge). This technique has been developed to replace classic liquid-liquid extraction and adsorption-based cleanup methods. By using this sample preparation method, interfering compounds can be removed and the analytes of interest enriched. To cover a wide range of analytes, several sorbents of polar, nonpolar, ionic, and polymeric materials can be used. SPE has been widely used in food analysis as a sample preparation or cleanup procedure prior to analysis using LC; among the compounds that have been extracted from foods are aromatic amines formed during heating, insecticides in fruits, fat contained in cereals, anthocyanins in plants, and pesticides in milk.

An extraction technique that has also been used mainly for fatty compounds (such as lipids and fat-soluble vitamins) is supercritical fluid extraction (SFE). SFE uses compounds that are held at

pressures and temperatures above their critical parameters, providing some advantages such as a reduced use of organic solvents, the possibility of tuning the selectivity of the extractant simply by changing the pressure or the temperature of the supercritical fluid, a reduced extraction time (because of the higher diffusivity associated with the supercritical fluids), higher quantitative yields (recoveries), and lower cost per extraction. Also important in terms of food sample preparation is the ability of SFE to be performed at low temperatures using a nonoxidant medium ( $\text{CO}_2$ ), which allows the extraction of thermally labile or easily oxidized compounds.

## Sugars (Carbohydrates)

Carbohydrates can be considered a major source of energy for the human body and on account of their diverse functional properties (considering the totally soluble carbohydrates such as glucose, fructose, galactose, lactose, maltose, and sucrose along with starch, gums, and pectins, among others) are used in the food industry as sweeteners, to enhance viscosity, acceptability, and palatability. Also, through well-known reactions induced by heating, reducing sugars can produce aromas and pigments. There is a need to monitor carbohydrates levels in food to predict their activity (also associated with human health) and to control the properties of foods. To determine quantitatively their content is also crucial for specially formulated foods (low-calorie, dietetic, suitable for diabetics, and with no lactose (in specially formulated milks)).

Prior to analysis by LC, carbohydrates can be extracted from food with water at a low temperature (to avoid risk of hydrolysis of oligosaccharides). After extraction, the crude aqueous extract should be cleaned up, usually by adding a volume of an organic solvent such as acetonitrile or by using a solid-phase cartridge.

Considering the wide range of molecular weights and types of carbohydrates, it is not possible to find a single LC method to cover all the applications. On the other hand, nearly all chromatographic modes can be used in separating carbohydrates. For example, LC analysis on chemically bonded amino, diol, or cyano groups have been used for separation of groups of monosaccharides and oligosaccharides. To achieve the desired separations, aqueous acetonitrile is used as a mobile phase, and the amount of water depends on the resolution needed. Reversed-phase (RP)-LC using  $\text{C}_{18}$  columns and using water as the mobile phase can also be employed to separate sugars with a different number of saccharide units.

Refractive index detection is normally used, but no gradient elution is allowed; therefore, new detection strategies are needed when challenging separations must be performed. In this sense, short-wavelength ultraviolet (UV) absorption (between 180 and 200 nm) can be used, but high-purity solvents are required. In order to overcome this problem, a derivatization reaction forming UV derivatives absorbing at higher UV wavelengths or fluorescent derivatives is carried out. Recently, a new detection device has been studied that is based on mass detection (evaporative light scattering detection (ELSD)). ELSD offers some advantages in terms of gradient compatibility, flat baselines, and improved peak shapes and resolution for mono- and oligosaccharides with a certain degree of polymerization (up to 22).

Nevertheless, the most common method of analyzing carbohydrates from foods is through ion exchange chromatography (IEC) using strongly or weakly basic anionic resins or cationic resins. IEC is often combined with amperometric detection, which allows electrochemical detection of sugars. The detection mechanism involves the oxidation of the sugar in an electrode (i.e., platinum or nickel-chromium (80:20) alloy). This type of detection has been proved to be very sensitive, allowing, for example, the analysis of disaccharides and oligosaccharides present in honey as minor components. **Figure 1** shows an example of a separation of oligosaccharides (maltodextrines) found in enteral formulas using IEC with amperometric detection.

Size-exclusion chromatography (SEC) has been applied to separation of sugars, particularly when the mixture contains sugars with a different number of saccharide units. In this sense, different methods have been developed for routine analysis of glucose, fructose, and sucrose in fruit juices; the analysis time is less than 8 min. In these separations, not only the molecular weight but also the molecular volume (spatial configuration) is taken into account. Two packing materials have been used, one based on a polystyrene polymer with an exclusion limit of  $\sim 1000$  and another one on a sulfonated monodisperse resin-based column in protonated form. Both systems allow the complete separation of small carbohydrates and also of some sugar alcohols such as mannitol, sorbitol, and xylitol.

## Vitamins

The nutritional role of vitamins is crucial to human health, and analytical methods have been developed in order to attempt complete separation of the different types of vitamins. The interest in vitamin

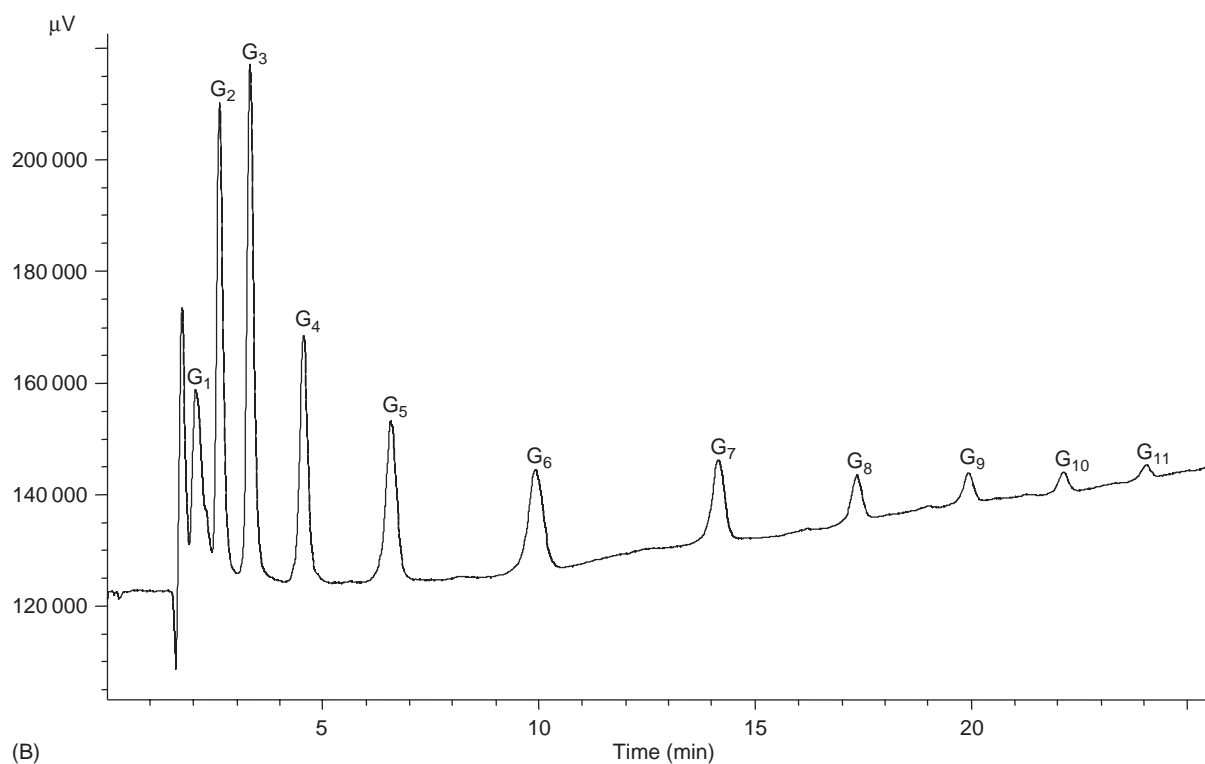
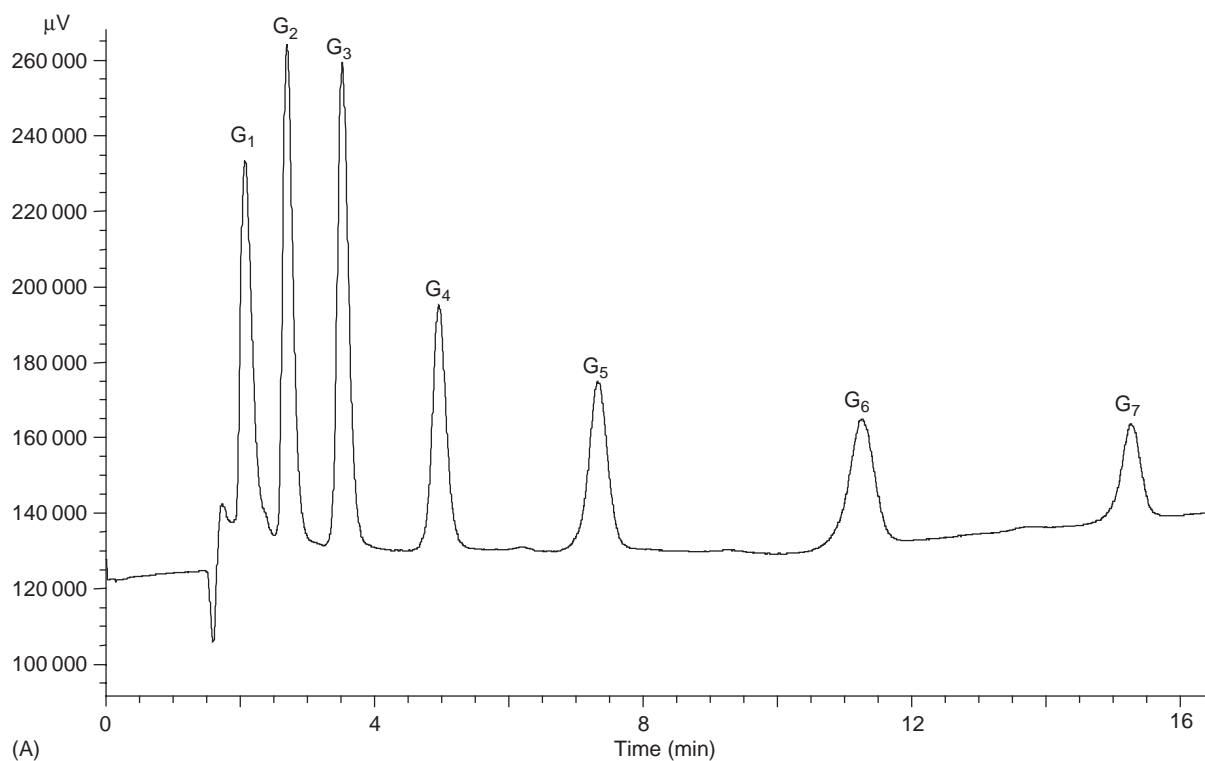
analysis is related to different purposes: implementing regulatory enforcement, checking compliance with nutrient labeling regulations, providing quality assurance for supplemented products, studying changes in vitamin content attributable to food processing, packaging, storage, etc. At present, LC is the method of choice, but the results are dependent on the extraction method used prior to chromatographic separation. Usually, vitamins are difficult to analyze because they are thermally labile compounds, readily oxidized by heat, oxygen, or UV light and, at the same time, are present in foods as minor components with physicochemical properties similar to those of the major compounds. LC offers some advantages for vitamin separation such as operation at low temperatures in the absence of light and oxygen, short exposure to UV light during detection, short analysis times, and the possibility of using selective detectors that allow simplification and shortening of the sample preparation step. In this article, fat-soluble and water-soluble vitamins will be discussed separately.

### Fat-Soluble Vitamins

The fat-soluble vitamins can be divided into four different groups, based on biological activity: vitamin A (retinol and carotenoids), vitamin D (cholecalciferol and ergocalciferol), vitamin E (tocopherols and tocotrienols), and vitamin K (phyloquinone and menaquinone), all of which may be present in a number of closely related forms.

**Vitamin A** Naturally occurring sources of vitamin A provide different forms of vitamin A (retinol, retinyl esters, and retinaldehyde) and provitamin A carotenoids. An adequate intake of dietary vitamin A is essential for normal vision, growth, functioning of the immune system, etc. All natural sources of vitamin A are derived from provitamin A carotenoids, which are synthesized exclusively by higher plants and photosynthetic organisms. The synthetic retinyl acetate or retinyl palmitate is used to supplement foods with vitamin A; unless knowledge of the ester form is important, the food is saponified to liberate free retinol prior to analysis; after saponification, an extraction step is needed. Sometimes (such as for the case of fortified milk) total vitamin A can be extracted directly using hexane and injected into the LC system.

Retinol can be chromatographed on normal-phase or RP columns, the latter being more appropriate when the sample extracts contain small amounts of triglycerides (after saponification). Different LC methods allow separation of various isomers such as the *cis-trans* retinol forms. Although retinol and



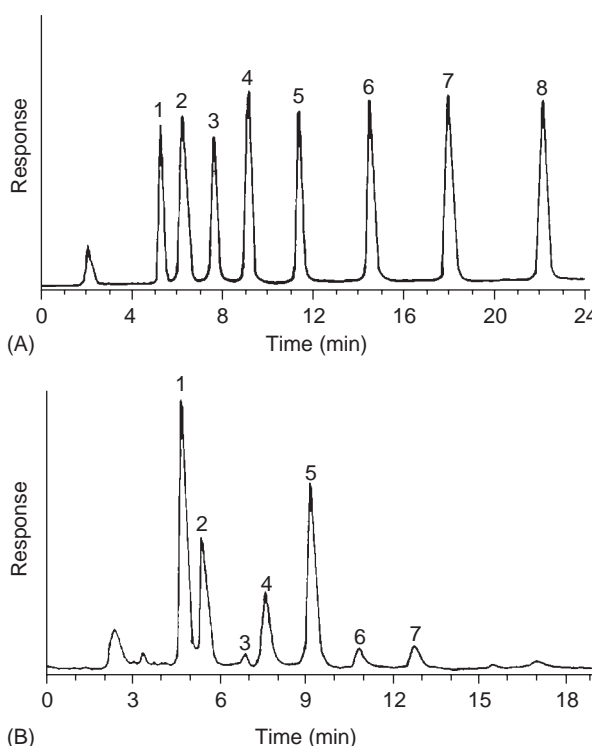
**Figure 1** High-performance anion exchange chromatography separation of (A) standard solution of maltodextrins and (B) enteral formulation. Column, Carbowac PA-100 (250 × 4 mm); mobile phase, gradient from water at pH 5.8 (mobile phase 1), 1 mol l<sup>-1</sup> sodium acetate (mobile phase 2), and 500 mmol l<sup>-1</sup> sodium hydroxide (mobile phase 3). Gradient program: mobile phase 1, constant at 57% from 0 to 6.5 min, then reduced to 42% at 31.5 min; mobile phase 2, constant at 10% from 0 to 6.5 min, then increased to 25% at 31.5 min; mobile phase 3, constant at 33% from 0 to 31.5 min; 1 ml min<sup>-1</sup>. Amperometric detection (gold electrode): pulse potentials for  $E_1$  of 0.05 V,  $E_2$  of 0.75 V, and  $E_3$  of -0.15 V were applied for duration times  $T_1$  of 0.4 s,  $T_2$  of 0.2 s, and  $T_3$  of 0.4 s. Peaks from G1 to G11 maltooligosaccharides. (Reproduced with permission from Moreno FJ, Olano A, Santa-Maria G, and Corzo N (1999) Determination of maltodextrins enteral formulations by three different chromatographic methods. *Chromatographia* 50: 705–710.)



retinyl esters exhibit native fluorescence, the quantum efficiency is low, and UV detection is commonly used in most LC systems (the wavelength of maximum absorbance is 328 nm).

**Vitamin D** Vitamin D ( $D_2$  or  $D_3$ ) is derived from UV irradiation of sterols. An adequate supply of vitamin D is essential in the prevention of rickets in children and of osteomalacia in adults. Dietary sources of vitamin D are essential for those people with inadequate skin exposure to sunlight. Vitamin D is present in foods only at trace levels. Because its physicochemical properties are similar to those of other lipid components present at higher levels, extensive sample preparation is needed prior to LC analysis. Saponification is mandatory for foods with high triglyceride (TG) content, followed by extraction using organic solvents. Further purification is usually needed to remove compounds such as sterols, carotenoids, and other constituents of the unsaponifiable fraction that could interfere in the LC determination of vitamin D. A wide variety of cleanup techniques have been used, including sterol precipitation, thin-layer chromatography (TLC), open-column chromatography, and SPE. Vitamins  $D_2$  and  $D_3$  can only be separated on RP columns using methanol:water or methanol:acetonitrile as the mobile phase. Detection is performed at 264 nm, providing enough sensitivity for detection in fortified foods and in those foods rich in vitamin D.

**Vitamin E** Vitamin E is found in vegetable oils and other food products such as nuts, seeds, and fruits in eight forms, four with saturated side-chains (tocopherols) and four with unsaturated side-chains (tocotrienols). Vitamin E plays an important role as a lipid antioxidant in stabilizing subcellular membranes. The different forms ( $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$ ) are commonly found in the above natural products, but their biological activities differ from one another. Thus it is necessary to separate completely the various forms to obtain an assessment of the total activity. Direct solvent extraction or saponification can be used, depending on the food matrix. Oils and fats can be simply dissolved in hexane and analyzed directly using normal-phase LC with either silica or polar-bonded stationary phases. Under these conditions, the eight forms of vitamin E can be isocratically separated (Figure 2). Since RP-LC is not capable of separating  $\beta$ - and  $\gamma$ - isomers, it can only be used to determine the total  $\alpha$ -tocopherol after sample saponification. Fluorescence detection is commonly used on account of the strong natural fluorescence provided by tocopherols and tocotrienols, this type of detection having a high sensitivity and selectivity



**Figure 2** HPLC of vitamin E. (A) Standards of vitamin E vitamers. Column, 5  $\mu$ m Supelcosil LC-Si (250  $\times$  4.6 mm inner diameter); mobile phase, isooctane/ethyl acetate (97.5:2.5), 1.6 mL min<sup>-1</sup>; fluorescence detection, excitation 290 nm, emission 330 nm. Peaks: 1,  $\alpha$ -tocopherol; 2,  $\alpha$ -tocotrienol; 3,  $\beta$ -tocopherol; 4,  $\gamma$ -tocopherol; 5,  $\beta$ -tocotrienol; 6,  $\gamma$ -tocotrienol; 7,  $\delta$ -tocopherol; 8,  $\delta$ -tocotrienol. (B) Saponified rice bran sample. Chromatographic conditions as in (A) except for the mobile phase, isooctane/ethyl acetate/2,2-dimethoxypropane (98.15:0.90:0.85:0.1). (Reproduced from AOCS Press.)

against other lipidic material present in the sample. Fluorescence detection is performed using 290 nm (excitation) and 330 nm (emission) for tocopherols and tocotrienols. UV absorbance (at 280 nm) has also been used with enough sensitivity, and electrochemical detection has provided the highest sensitivity, allowing the detection of low picomoles of vitamin E in human plasma.

**Vitamin K** Vitamin K exhibits an important anti-hemorrhagic activity, which has increased the interest in developing analytical methods to determine its content in foods. The sample preparation includes various steps such as enzymatic hydrolysis and cleanup using SPE cartridges. Normal-phase LC and RP-LC have been used to separate vitamin K in conjunction with UV detection at 270 nm; fluorescence detection can be used only after phyloquinone has been converted to the corresponding hydroquinone after electrochemical or chemical reduction.

### Multiple Fat-Soluble Vitamins

Recent developments have been directed toward the simultaneous determination of multiple vitamins in foods. Separation of vitamins A, E, D<sub>2</sub>, and D<sub>3</sub> in lacteal matrices has been performed using RP-LC with methanol:water as the mobile phase and UV detection at 265 nm (or at the wavelength of maximum absorption of each individual form), after exhaustive saponification. Normal-phase LC has also been used to analyze multiple fat-soluble vitamins in seed oils after extraction or even direct injection of the sample prior to analysis. UV detection has also been used. Sometimes a combination of two detection systems such as UV-visible detection and fluorescence have been used to, for example, determine, respectively, the carotenoids and fat-soluble vitamins in foods.

### Water-Soluble Vitamins

In general, the methods available for water-soluble vitamins are less successful for real samples than those described for the fat-soluble vitamins. The problems are, in general, related to sample preparation prior to LC analysis since naturally occurring vitamins are often bound to other food constituents such as carbohydrates or proteins.

**Vitamin C** In food, vitamin C exists as two vitamers with equivalent biological activities: l-ascorbic acid (AA) and its oxidation product, dehydro-l-ascorbic acid (DHAA), while the stereoisomer, iso-ascorbic acid, has very low activity. In addition to their known activities as vitamins, their role in other health-promoting functions such as prevention of cancer, heart disease, hypertension, etc. is at present under investigation. The most important problem encountered in vitamin C analysis is the lack of stability and ease of oxidation that makes it necessary to use extraction procedures that avoid degradation; usually acids (metaphosphoric, trichloroacetic, etc.) are employed along with metal chelators such as ethylenediaminetetraacetic acid. Vitamin C can be analyzed using RP-LC (with C<sub>18</sub> columns) using water at low pH values (acidified with sulfuric acid or different buffers). Detection is performed through UV absorption at 254 nm when AA has to be detected. DHAA has its maximum absorbance between 210 and 230 nm, and therefore it is susceptible to interferences with different natural food constituents. Fluorescence detection (after derivatization with o-phenylenediamine) is often used to attain maximum sensitivity.

**Vitamin B<sub>1</sub>** In food, vitamin B<sub>1</sub> (thiamine) exists as four vitamers, the nonphosphorylated form,

thiamine (predominating in plants), and its phosphorylated esters (monophosphate, pyrophosphate, and triphosphate), which are the principal forms in animal products. Thiamine deficiency has been associated with beriberi disease. Extraction of vitamin B<sub>1</sub> includes acid hydrolysis to release vitamers from food matrices, although sometimes enzymatic hydrolysis is used to remove excess starch or protein in the matrix. Further cleanup is performed through ion exchange column chromatography or SPE on C<sub>18</sub> cartridges. Separation can be performed using RP ion pair chromatography. Detection in real food samples is done using a fluorescence detector (excitation 365–375 nm and emission 425–435 nm) after online or offline derivatization with alkali potassium hexacyanoferrate(III) to convert thiamine to thiochrome. UV detection has lately been rejected because of the interference of other peaks corresponding, for example, to nucleic acids.

**Vitamin B<sub>2</sub>** Food contains three B<sub>2</sub> vitamers, riboflavin and its two coenzyme forms, flavin mononucleotide and flavin adenine dinucleotide, which are the predominant vitamers in foods and are usually bound to proteins. Their analysis usually takes place after extraction with dilute mineral acids with or without enzymatic hydrolysis of the coenzymes (which is necessary to convert all forms to riboflavin and to quantify them as total riboflavin). The extracts may be purified using SPE with C<sub>18</sub> cartridges. All the operations performed prior to analysis need to be done under subdued lighting to avoid decomposition of riboflavin upon exposure to light. RP chromatography with C<sub>18</sub> columns is used along with fluorescence detection (excitation, 440 nm; emission, 520 nm).

**Vitamin B<sub>3</sub>** Vitamin B<sub>3</sub> (niacin) has different physiologically active forms, nicotinic acid, nicotinamide, and their coenzymes, which are very stable at ambient temperature. Usually acid or alkaline hydrolysis are used to convert nicotinamide to nicotinic acid for quantitation of both vitamers as nicotinic acid; the first treatment (acid) is used to quantitate biologically available niacin, while alkaline hydrolysis provides an estimate of the total niacin content. Nicotinic acid has been analyzed using ion exchange or RP chromatography with amino columns. Detection is performed with UV absorbance at 254 nm.

**Vitamin B<sub>6</sub>** Pyridoxine is the principal vitamer of vitamin B<sub>6</sub> in plants products, while pyridoxal phosphate is the predominant form in animal-based foods. The analysis of vitamin B<sub>6</sub> is quite complex due to the presence of six possible vitamers of

different structures. Sometimes enzymatic hydrolysis is used to convert phosphorylated vitamers to simple forms. They are light sensitive, and therefore analysis must be performed under subdued lighting. LC has the extraordinary capability of separating vitamers (including glycosylated ones) using  $C_{18}$  columns and low-pH (2.2) mobile phases with alkane sulfonates to favor ion pairing. Fluorescence is used as detection method because of its increased sensitivity and specificity relative to UV absorbance; nevertheless, even if vitamers fluoresce naturally in slightly acidic to neutral solutions, postcolumn derivatization with, for example, monobasic potassium orthophosphate is often used to adjust the pH for enhanced fluorescence sensitivity of  $B_6$  vitamers.

**Folacin** Folacin gives rise to a large number of related compounds having the same basic structure as folic acid but differing in the state of reduction and the number of glutamate residues. Folacin vitamers are sensitive to heat, oxygen, light, and extremes of pH. Their analysis is complicated by the low levels at which vitamers occur in food, their lack of stability, and the multiplicity of forms of naturally occurring vitamers. Extraction takes place after homogenization in a neutral or mildly acidic solution and usually contains one antioxidant such as ascorbic acid. Sometimes folacin in polyglutamate forms is enzymatically deconjugated to the corresponding monoglutamates prior to quantitation. Sample cleanup is accomplished by means of SPE on anion or cation resins, although affinity chromatography using a folate-binding protein isolated from milk has also been used. LC using ion exchange or RP ion pair chromatography is the method of choice, but due to the complexity of folacin vitamer analysis, at present there is no single LC method to achieve such separation. UV absorption is often sufficient to allow detection at the levels at which folic acid is added in fortified foods; if more sensitivity is needed, fluorescence detection can be used simply by using an acidic pH (pH value of 2.3) or after postcolumn oxidative cleavage of a pterin fragment.

### Multiple Water-Soluble Vitamins

At present, much effort is being devoted to simultaneous separation and detection of water-soluble vitamins. Undoubtedly, multiple water-vitamin analysis using LC as a separation method is effective with RP ion pair chromatography with acidified methanol or acetonitrile:water as the mobile phase. Detection is performed using combined systems such as UV absorbance and fluorescence systems, depending on the vitamins to be determined. Analysis in real food

samples requires complex sample preparation with multiple steps (extraction, hydrolysis, oxidation, cleanup, etc.) that are reduced when synthetic preparations have to be analyzed.

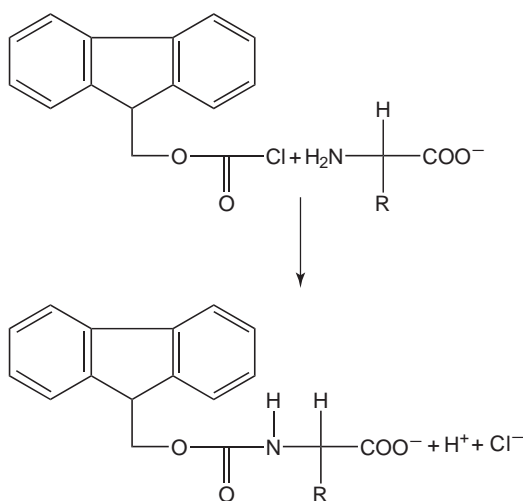
### Amino Acids, Peptides, and Proteins

The importance of amino acids, peptides, and proteins is widely known and it arises from different aspects ranging from an important role in human nutrition and health to different technological functions that can widely influence the quality of food products (mainly for peptides and proteins). Therefore, their analysis is crucial not only from the nutritional point of view (for example, to warranty and verify the content of an specific amino acids) but also for product quality assessment and product development (where information about the composition of the protein fraction is mandatory). Also important is assessment of the authenticity of foods through enantiomer separation and identification.

Amino acid analysis has a different approach, in terms of sample preparation, if free or total amino acids have to be determined. For amino acids bound to proteins, it is necessary to liberate them, using acid hydrolysis or acid digestion with microwave heating, prior to chromatographic analysis. A clean-up step is always needed to remove interferences and/or to concentrate the sample. Usually sample clean-up is performed using SPE with different resins (cation exchange) or  $C_{18}$  cartridges that can retain lipids and large proteins and allow elution of free amino acids.

One of the most commonly used methods in the separation of free amino acids using LC is ion exchange chromatography (cation exchange resin), which in its new format of small particles and pellicular packings has improved both the resolving power and the analysis time. Detection is performed through UV absorption after postcolumn reaction with ninhydrin ( $\lambda_{\max} = 570 \text{ nm}$ ) or *o*-phthalaldehyde (OPA), which allows the detection limit to be decreased to 5 pmol. RP-LC can also be used after precolumn derivatization with OPA, phenylisothiocyanate (PITC), or 9-fluorenylmethyl chloroformate (FMOC), among others. A typical derivatization scheme of an amino acid with FMOC is shown in **Figure 3**. Derivatization allows spectroscopic detection of amino acids while increasing their degree of hydrophobicity (and subsequently their retention), allowing their separation using RP-LC.

Separation of chiral amino acids has become very important for assessing authenticity: only L-amino acids seem to have a natural origin and D-isomers



**Figure 3** An example of a typical derivatization scheme of amino acids with Fmoc-Cl.

cannot be utilized by humans, so that the associated proteins have a decreased digestibility. Enantiomeric separation of the d- and l-amino acids can be carried using three different approaches: precolumn derivatization with a chiral agent followed by separation of the resulting diastereomers on regular RP columns, direct separation of underivatized enantiomers on a column with a chiral bonded phase (such as the Pirkle column), and using a chiral crown ether added to the mobile phase.

Peptide analysis requires an exhaustive sample preparation consisting of two steps: sample extraction and removal of proteins, usually by ultrafiltration, which allows fractionation of peptides at the same time. Further clean-up through SPE with C<sub>18</sub> cartridges is often necessary. Peptide LC analysis can be done using RP columns with ion pairing agents such as trifluoroacetic acid (TFA), IEC, and SEC. Detection is performed through UV absorption (for aromatic amino acids and tyrosine and tryptophan) and, more commonly, after derivatization with dansyl chloride or PITC (for primary and secondary amino groups). One of the most important advances in peptides detection has been the online coupling of LC with mass spectrometry (MS) through an electrospray ionization (ESI) interface, which allows the identification of high-molecular weight biopolymers generating multiple charged molecular ions, being, at the same time, compatible with liquid mobile phases. Matrix-assisted laser desorption/ionization coupled to a time-of-flight mass analyzer (MALDI-TOF-MS) has also been used for peptide analysis, allowing accurate determination of the mass of peptides and proteins with a high sensitivity (subpicomole range). The data obtained from the above-mentioned soft

ionization methods along with database searching have allowed the characterization of peptides of great biological significance and also their sequencing.

Separation of peptide stereoisomers can be accomplished using the same approach as for chiral amino acids, as mentioned before.

Proteins are often analyzed after isolation using different systems such as dialysis, ultrafiltration, or protein precipitation with organic solvents or addition of high concentrations of salts. Once isolated, proteins can be analyzed using RP-LC, IEC, and SEC. Recent studies have shown the possibility of fast separation of proteins using LC with pellicular packings, perfusion packings, and monolithic columns. These are different approaches to the same problem, which is the low diffusivity of some molecules into the pores of conventional packings, which ends up with a dramatic band broadening and, therefore, with a loss of efficiency. Pellicular packings consist of nonporous sorbents with a fluid impervious core that allow rapid analysis of proteins due to the suppression of diffusion inside the pore structure. Perfusion packings are based on packing materials with different pores, the throughpores (>500 nm) and a network of smaller pores (30 nm < diameter of pores < 70 nm) branching from the throughpores; this structure allows a huge increase in flow rate without loss of resolution. The most recent approach consists of using monolithic columns, which are agglomerates of polyacrylamide particles, and silica-rod columns: these columns provide a high efficiency and better permeability than the conventional columns used in LC, allowing also a higher flow rate and therefore reducing the total analysis time. Because of the flat plate height versus flow velocity curves at high velocity values, it is possible to work at higher flow rates without losing efficiency. These columns can be used for rapid purification of proteins.

UV absorbance and fluorescence and, lately, LC-MS are the most commonly means of detecting proteins. LC-ESI-MS and MALDI-TOF-MS have largely improved our knowledge of the primary structure of proteins and have become the methods of reference in both food control and food research laboratories.

## Lipids

Lipids are involved in many important biological processes in humans, and since the absorption and metabolism of numerous lipids are closely related to several diseases, analysis of lipids has grown greatly in recent years. Diverse definitions for lipids can be considered, but all of them include fatty acids and



their derivatives and substances related biosynthetically to these compounds; other definitions include also carotenoids, terpenes, and other classes of compounds. As components of oils and fats, lipids are almost ubiquitous food ingredients, with an important role in nutrition and the taste and functional properties of the final product.

The main area of application of HPLC in lipid analysis is the determination of larger molecules like triglycerides and phospholipids, while the analysis of smaller lipid classes like sterols and fatty acids is usually carried out using gas chromatography (GC), though it is possible to analyze these molecules using LC too. The subject can be divided in two topics: separation of simple lipid classes from a lipid fraction and separation of the individual components, like the different sterols of a specific fraction.

Sample preparation for analysis of lipids poses few problems. If the sample is liquid, often it is not necessary to extract or fractionate the food, and even direct injection of liquid samples such as oils can be done. For solid foods, a simple extraction with a nonpolar solvent or mixture of solvents is enough to isolate the fat containing triglycerides, followed by a simple cleanup if needed, before injection into the LC. For detection, the refractive index has been traditionally used for triglyceride analysis with isocratic elution, though the separation of lipids using RP-LC with gradient elution excludes the use of refractive index detectors, and mass detectors or light-scattering detectors are being regularly used to obtain better sensitivity than with UV detection.

The separation of simple lipid classes from a lipid fraction aims to obtain distinct fractions of sterol esters, triacylglycerols, diacylglycerols, free sterols, free fatty acids, monoacylglycerols, and wax esters. There are numerous methods for lipid class separation, traditionally employing adsorption chromatography with silica gel columns, with increasing use currently of bonded phases such as the nitrile, diol, and polyvinylalcohol phases and of ELSD. These bonded phases give much better reproducibility of retention times than do the usual silica gel columns. Nevertheless, refractive index detection with silica gel columns and isocratic elution is still frequently employed for routine applications.

The separation of triglycerides using partition chromatography can be done both in RP and normal-phase columns. RP columns have given better results with nonaqueous mobile phases (for instance, acetonitrile:acetone), which allow separation of triglycerides according to the equivalent carbon number. Normal-phase silica columns have been used for analysis of triglycerides, which is a classical application of this type of column. The main

drawbacks are the problems of reproducibility and time of stabilization. Oxidized triacylglycerols (monomers, dimers, and polymers) are separated using SEC with two columns in series and refractive index detection after isolation of the polar fraction from fats using SPE.

The preferred method for analysis of free fatty acids is GC, while TLC and HPLC can be used to isolate the fatty acid fraction. The alternative determination using HPLC needs fatty acid derivatives to be prepared using a chromophore to allow detection using the most sensitive fluorescent detector. Many examples of different derivatives can be found in the literature.

Analysis of the components of other fractions of simple lipids like free sterols, sterol esters, and wax esters by means of HPLC has some drawbacks, and it is usually performed using high-temperature GC. However, the most remarkable field of application of HPLC is the analysis of oxidized sterol and their esters, of biological interest, that used to be performed earlier using RP-LC with postcolumn fluorometric detection and is nowadays be carried out using LC coupled to MS, providing structural information on these complex derivatives.

## Minor Components of Food

A developing area of analytical development in HPLC is the determination of several groups of minor food components, like flavonoids and polyphenols in general, xanthophylls, and other carotenoids. Most of them have important bioactivities and have stimulated much interest in food and nutrition studies.

These food components have the common property of being labile and so are degraded by light, heat, or oxygen, which limits the sample extraction and cleanup, and especially its great complexity, with dozens (or even hundreds) of very similar components or isomers, which sometimes makes it extremely difficult to separate the components of each family in the sample. An additional problem is the lack of standard compounds for most of them, which makes the identification very complicated. The difficulty is usually overcome by coupling LC with MS, or with MS/MS, to determine the structure of the separated compounds.

### Phenolic Compounds

Phenolic compounds have been traditionally of interest due to their contribution to the color and taste of food, but nowadays the main interest in their determination is the proven bioactivity of many of



them, which includes beneficial functions such as antioxidative, antimicrobial, and anticarcinogenic activities. The phenolic family includes thousands of compounds, from simple (monocyclic) phenolic acids to flavonoids, which can be divided into many classes (flavone, flavanone, flavonol, anthocyanin, isoflavone, etc.), present as aglycone or as glycoside. An important group of polyphenols is that of the anthocyanins, the largest group of water-soluble pigments in nature, responsible for the intense color of many fruits and flowers. Because of their chemical structure and polarity, RP-LC is the most regularly used technique in the analysis of polyphenols.

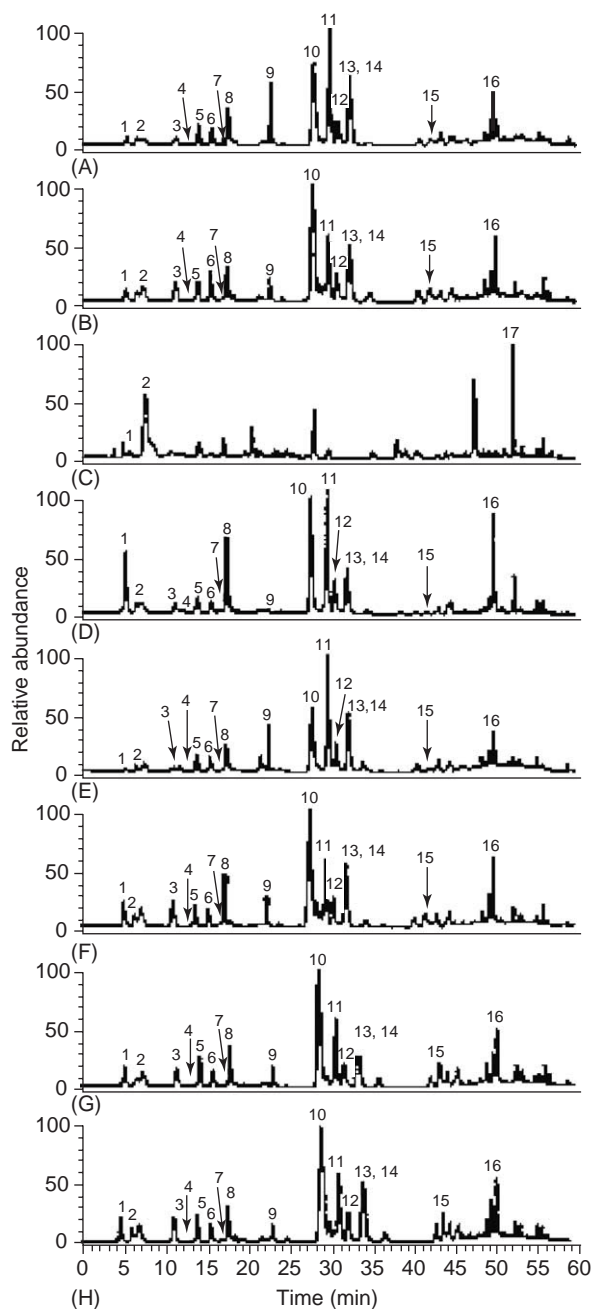
There are countless methods of analysis of phenolic compounds in the literature, the majority of them employing  $C_{18}$  columns with binary gradients (with acetonitrile or methanol) in aqueous RPs. Frequently, the ionic strength and, especially, the acid pH of the mobile phase must also be optimized. Many types of detector have been used for this application, with UV detection (or, more conveniently, detection using a photodiode array detector) being the most commonly used form of detection and with an increasing number of methods that use LC-MS with ESI. An example of the analysis of phenolic compounds in ethanolic extracts of propolis (a natural substance collected by honeybees from buds and exudates of certain trees and plants) of different geographical origin using LC-MS is shown in Figure 4.

### Carotenoids

The analysis of carotenoids, including carotenes (without oxygen) and the oxygenated derivatives xanthophylls, poses a difficult challenge to

the analyst, mostly due to the great variety of carotene isomers including *cis-trans* isomers, cyclic and acyclic isomers, and the number of oxygen functions, which can produce hundreds of combinations (where lycopene,  $\beta$ -carotene, zeaxanthin, lutein,  $\beta$ -cryptoxanthin, and violaxanthin are just some known examples), and the very low oxidative stability of carotenoids once they have been isolated.

Carotenoids can be analyzed using normal-phase LC or RP-LC, but RP-LC is preferred because carotenes are strongly retained and the separation of  $\alpha$ - and  $\beta$ -carotenes is easily achieved. Since carotenes



**Figure 4** HPLC chromatograms of ethanol extracts of propolis of various geographic origin. (A) Argentina; (B) Australia; (C) Brazil; (D) Bulgaria; (E) Chile; (F) China (Hebei); (G) China (Hubei); (H) China (Zhejiang). Column:  $5\mu\text{m}$  Capcell Pak ACR 120  $C_{18}$  column ( $250 \times 2\text{ mm}$  inner diameter); mobile phase 0.1% formic acid in water (A) and 0.08% formic acid in acetonitrile (B); gradient 20–30% (B) (15 min), 30% (B) (15–35 min), 30–80% (B) (35–60 min) at a flow rate of  $0.2\text{ mL min}^{-1}$ . Mass spectrometric detection (LC quadrupole ion trap) with ESI source: voltage 5 kV; electrospray capillary voltage,  $-10\text{ V}$ ; capillary temperature,  $260^\circ\text{C}$ ; all MS data were acquired in the negative ionization. In each line the numbers of peaks represent the same compounds. 1, caffeic acid; 2, *p*-coumaric acid; 3, 3,4-dimethoxycinnamic acid; 4, quercetin; 5, pinobanksin 5-methyl ether; 6, apigenin; 7, kaempferol; 8, pinobanksin; 9, cinnamylideneacetic acid; 10, chrysin; 11, pinocembrin; 12, galangin; 13, pinobanksin 3-acetate; 14, phenethyl caffeate; 15, cinnamyl caffeate; 16, tectochrysin; 17, artemillin C. (Reprinted from Shigenori Kumaza, Tomoko Hamasaka, and Tsutomu Nakayama (2004) Antioxidant activity of propolis of various geographic origins. *Food Chemistry* 84: 329–339.)

are only sparingly soluble in typical RP solvents (methanol or acetonitrile), different mixtures containing acetonitrile/dichloromethane/methanol (70:20:10) or methanol/chloroform have been successfully used along with octadecyl silica (ODS) columns to separate different carotenoids and/or isomers. Polymeric stationary phases (especially C30- and C34-bonded phases) have distinct advantages over the common C<sub>18</sub> material for separation of similar carotenoids, though in these cases control of the temperature of analysis is critical.

Detection has been usually accomplished by light absorption in the visible region ( $\lambda_{\text{max}}$  453 nm for  $\beta$ -carotene), while LC-MS examples for identification of complex samples of carotenoids are becoming common.

### Other Pigments

Chlorophylls and their derivatives in food are separated using HPLC with great advantages compared with other analytical techniques, and consequently RP-LC with spectrophotometric detectors is the most used technique for their analysis. On the other hand, several researchers have developed methods to analyze simultaneously chlorophylls and carotenoids of fruits and vegetables in a single run using RP HPLC with mobile phases in gradient and ODS columns. The determination of other pigments in food is usually less complex than in the case of anthocyanins or carotenoids, mainly because of the limited number of components of other groups like betalains (natural pigments) or azo dyes (synthetic), and routine HPLC applications for them can be found in the literature.

### Food Contaminants

The growing concerns regarding food safety have made the analysis of food contaminants an area of intense development in recent years, with a focus on chromatographic techniques, where LC has an essential role. The number of possible food contaminants grows without limit, including mycotoxins, antibiotic residues, and other potentially dangerous chemicals.

Analysis of pesticide residues is usually performed using GC, and the main field of application of LC is the simultaneous detection of very different pesticides in a single analysis, due to the lack of limitations of volatility or stability compared with GC. Compared with other analyses of minor components of food, determination of residues in food needs lower detection limits and, usually, laborious sample preparation and fractionation before the LC separation can take place.

Separation of organochlorine and organophosphorus pesticides residues is mainly performed in RP-LC columns with water and acetonitrile or methanol or both. The usual detector is a UV detector or a diode array detector, with a growing number of applications that use MS with APCI interface.

The analysis of mycotoxins, toxic substances produced by some fungi, is essential to guaranteeing the safety of foods such as cereals, nuts, oilseeds, and spices since the discovery of their significance to human health. The main mycotoxins, like aflatoxins and ochratoxin A, are determined using HPLC with octyl- or octadecyl-bonded phases and polar solvents such as methanol, acetonitrile, and water. For detection with the required sensitivity and selectivity, spectrofluorimetric detection is the most common method, with increasing use of mass spectrometric interfaces.

Another source of concern in food safety control is the increasing amounts of residues of veterinary drugs that can be found in foods. These drugs are mostly antibiotics of different structures, such as tetracyclines, macrolides, quinolones, sulphonamides, and  $\beta$ -lactams. The preferred analytical technique for antibiotic residue determination is HPLC, and specific methods for each family of antibiotics can be found in the scientific literature. Most of the published methods use RP columns with gradient elution. Every available detection method can be used, depending on the application, though MS is a powerful technique for the identification and confirmation of veterinary drug residues in food samples, and at the moment, LC-MS is the method of reference.

**See also:** **Carbohydrates:** Overview. **Extraction:** Solid-Phase Extraction; Supercritical Fluid Extraction. **Food and Nutritional Analysis:** Sample Preparation; Antioxidants and Preservatives; Mycotoxins; Oils and Fats. **Lipids:** Overview. **Peptides. Proteins:** Overview. **Toxins:** Mycotoxins; Neurotoxins. **Vitamins:** Overview; Fat-Soluble; Water-Soluble.

### Further Reading

- Christie WW (2003) Lipid analysis. *Isolation, Separation, Identification and Structural Analysis of Lipids*, 3rd edn. Bridgwater, England: The Oily Press.
- Hancock WS (ed.) (1990) *High Performance Liquid Chromatography in Biotechnology*. San Francisco: Wiley-Interscience.
- Ibañez E and Cifuentes A (2001) New analytical techniques in food science. *Critical Reviews in Food Science and Nutrition* 41(6): 413–450.
- Macrae R (1988) *HPLC in Food Analysis*, 2nd edn. London: Academic Press.
- Matissek R and Wittkowski R (eds.) (1993) *High Performance Liquid Chromatography in Food Control*

and Research. Pennsylvania: Technomic Publishing Co.

Moreno-Arribas MV, Pueyo E, and Polo MC (2002) Analytical methods for the characterization of proteins

and peptides in wines. *Analytica Chimica Acta* 458: 63–75.

Nollet LML (ed.) (2000) *Food Analysis by HPLC*, 2nd edn. New York: Dekker.

## Pharmaceutical Applications

**D M Diehl**, Waters Corporation, Milford, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by T D Wilson, pp. 2702–2711, © 1995, Elsevier Ltd.

### Introduction

Liquid chromatography (LC) has been applied in the pharmaceutical industry almost from the inception of the technique and certainly through its adolescence up to its current mature status. Liquid chromatography in particular has grown along with the demands of the pharmaceutical industry because of its unique capabilities and strengths, including the obvious precision, accuracy, low levels of detection, selectivity (stability-indicating properties), robustness, versatility in terms of column and mobile phase chemistries and mechanisms of separation, and range of detection methods.

### Overview of Requirements

In order to understand the place of liquid chromatography in the pharmaceutical industry, the evolution of a drug in the drug development cycle should be briefly discussed first. Several of the major stages a drug passes through from discovery to commercialization of a dosage form and further, are outlined in **Table 1**.

From the synthesis of a drug onwards, its identity strength, purity, and quality are continuously assessed and monitored, commonly using chromatography. The organic chemist discovering the compound is likely to use analytical LC to study reaction progress and condition optimization in synthetic and isolation schemes. Preparative LC is also used at various stages of discovery to provide adequate quantities of purified intermediates to take to subsequent steps and quantities of purified products for pharmacological testing. When larger than laboratory scale quantities of bulk drug are required in chemical process development, pilot production scale and later, full production scale batches are prepared. The resulting drugs and intermediates are assayed by

analytical LC and similarly release-tested at later stages.

Analytical research department involvement with bulk drug LC methods development commonly begins during early toxicology and safety evaluation, when it is necessary to document dose administered in the vehicle used and stability of the drug in the vehicle over the dosing period. This soon leads to preclinical method validation studies for the bulk drug assay and for the development of an impurities assay, both of which receive full validation at the later time.

Metabolism, pharmacokinetics, and bioavailability studies, initially on a preclinical level but later after human testing is approved, are also supported by LC analytical methods of drug and metabolite levels determination in various biological fluids and tissues. These studies and those to support further toxicology testing are the subject of other articles.

Chromatographic studies in pharmaceuticals development departments generally begin with preformulation testing and evaluation of bulk-drug properties such as solid state and solution stability and degradation mechanisms. Dosage form development is monitored by LC stability studies on short-term and later formal long-term bases. LC analytical methods to support these latter studies and validation thereof are normally carried out in analytical chemistry departments.

Drug substance impurity profiling is conducted on several batches by high-sensitivity LC methods prior to Investigational New Drug application (IND) submission, while the number of batches increases substantially by New Drug Application (NDA) time. In addition, formal drug substance stability studies are conducted with potency and impurity level testing by LC as integral parts.

Once the product formula and process of manufacture are characterized, they are documented and sets of tests and specifications are compiled. Specific testing is used to release incoming raw materials and bulk drug as well as finished product. Here again in release testing, LC plays a major role. No matter what stage of development the drug has reached,

and Research. Pennsylvania: Technomic Publishing Co.

Moreno-Arribas MV, Pueyo E, and Polo MC (2002) Analytical methods for the characterization of proteins

and peptides in wines. *Analytica Chimica Acta* 458: 63–75.

Nollet LML (ed.) (2000) *Food Analysis by HPLC*, 2nd edn. New York: Dekker.

## Pharmaceutical Applications

**D M Diehl**, Waters Corporation, Milford, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by T D Wilson, pp. 2702–2711, © 1995, Elsevier Ltd.

### Introduction

Liquid chromatography (LC) has been applied in the pharmaceutical industry almost from the inception of the technique and certainly through its adolescence up to its current mature status. Liquid chromatography in particular has grown along with the demands of the pharmaceutical industry because of its unique capabilities and strengths, including the obvious precision, accuracy, low levels of detection, selectivity (stability-indicating properties), robustness, versatility in terms of column and mobile phase chemistries and mechanisms of separation, and range of detection methods.

### Overview of Requirements

In order to understand the place of liquid chromatography in the pharmaceutical industry, the evolution of a drug in the drug development cycle should be briefly discussed first. Several of the major stages a drug passes through from discovery to commercialization of a dosage form and further, are outlined in **Table 1**.

From the synthesis of a drug onwards, its identity strength, purity, and quality are continuously assessed and monitored, commonly using chromatography. The organic chemist discovering the compound is likely to use analytical LC to study reaction progress and condition optimization in synthetic and isolation schemes. Preparative LC is also used at various stages of discovery to provide adequate quantities of purified intermediates to take to subsequent steps and quantities of purified products for pharmacological testing. When larger than laboratory scale quantities of bulk drug are required in chemical process development, pilot production scale and later, full production scale batches are prepared. The resulting drugs and intermediates are assayed by

analytical LC and similarly release-tested at later stages.

Analytical research department involvement with bulk drug LC methods development commonly begins during early toxicology and safety evaluation, when it is necessary to document dose administered in the vehicle used and stability of the drug in the vehicle over the dosing period. This soon leads to preclinical method validation studies for the bulk drug assay and for the development of an impurities assay, both of which receive full validation at the later time.

Metabolism, pharmacokinetics, and bioavailability studies, initially on a preclinical level but later after human testing is approved, are also supported by LC analytical methods of drug and metabolite levels determination in various biological fluids and tissues. These studies and those to support further toxicology testing are the subject of other articles.

Chromatographic studies in pharmaceuticals development departments generally begin with preformulation testing and evaluation of bulk-drug properties such as solid state and solution stability and degradation mechanisms. Dosage form development is monitored by LC stability studies on short-term and later formal long-term bases. LC analytical methods to support these latter studies and validation thereof are normally carried out in analytical chemistry departments.

Drug substance impurity profiling is conducted on several batches by high-sensitivity LC methods prior to Investigational New Drug application (IND) submission, while the number of batches increases substantially by New Drug Application (NDA) time. In addition, formal drug substance stability studies are conducted with potency and impurity level testing by LC as integral parts.

Once the product formula and process of manufacture are characterized, they are documented and sets of tests and specifications are compiled. Specific testing is used to release incoming raw materials and bulk drug as well as finished product. Here again in release testing, LC plays a major role. No matter what stage of development the drug has reached,

**Table 1** The drug development cycle

Discovery: organic, pharmacology	Drug substance characterization, salt selection, thin layer chromatography, preparative and analytical LC	
Chemical process development	Process improvement, intermediates characterization, impurities, solvent selection, scale-up	
Toxicology and safety testing	Preclinical Clinical	
Metabolism, bioavailability and pharmacokinetics	Preclinical Clinical	IND (CTA) Phase I Phase II Phase III
Environmental impact studies		
Preformulation studies	Solid state Solution state Compatibility	
Formulation studies	Laboratory scale, pilot scale	
Process improvement	Manufacturing	
Stability studies	Short-term, drug substance Long-term, drug product Clinical product Marketed product	
Tests and specifications development		
Drug substance	Release testing	NDA (MAA)
Excipients	Shelf-life testing	Submission
Drug product	In-process testing	Approval
Technology transfer	Manufacturing Quality control Process validation Cleaning validation	
Commercial production		
Marketed product	Line extensions, new package, new manufacturing site, new vendors	Phase IV

CTA, Clinical Trial Application; IND, Investigational New Drug; MAA, Marketing Authorization Application; NDA, New Drug Application. (Reproduced with permission from Willi AV (1983) *Isotope Effects in Chemical Reactions* [in German]. Stuttgart G; Thieme. Worth GK and Retallack EW (1988) Tritium isotope effect in high-pressure liquid chromatography. *Analytical Biochemistry* 174: 137–141.)

preclinical, clinical (Phases I–III) or commercial (Phase IV), release testing is required.

Technology transfer generally begins during Phase III clinical studies. This includes both transfer of the manufacturing process of drug substance and product and of testing procedures to specific company facilities or contract laboratories for implementation as chemical and commercial product sources.

Following approval to market the product, any change or upgrade in the process or reformulation of the product can require stability work and LC methods development to support studies. Any new manufacturing sites for bulk drug or drug product also need stability support. New vendors for packaging components or excipients must be qualified by Good Manufacturing Practice (GMP) requirements, which usually involves stability studies on product

manufactured with the changes. Product that is approved for marketing also requires formal stability studies, which are conducted to the extent of the proposed shelf-life aside from release testing. These also include LC analysis of product potency.

Other tests that require LC methods in the drug development cycle include assays for key process intermediates, in-process testing, cleaning validation studies, and compatibility testing. Synthetic process intermediates must be characterized structurally, by means of spectroscopy, but purity information and assays are applied as well with results supplied in regulatory filings. In-process testing may include an assay of activity in some pivotal step of processing the dosage form, although this is not always the case. These steps may be at a granulation stage prior to tableting or encapsulation or in a bulk solution stage



prior to filling. Cleaning validation studies are required as part of GMP procedures and these usually include testing at maximum sensitivity for the presence, or absence, of active components. Compatibility testing is commonly carried out on parenteral products that require mixing with available intravenous diluents (water for injection, normal saline, 5% dextrose, etc.). These are short-term stability tests that need drug assay and often impurities measurements conducted chromatographically. In addition, a significant portion of the activities of quality control departments on bulk drug, raw materials, and finished products involves LC measurements.

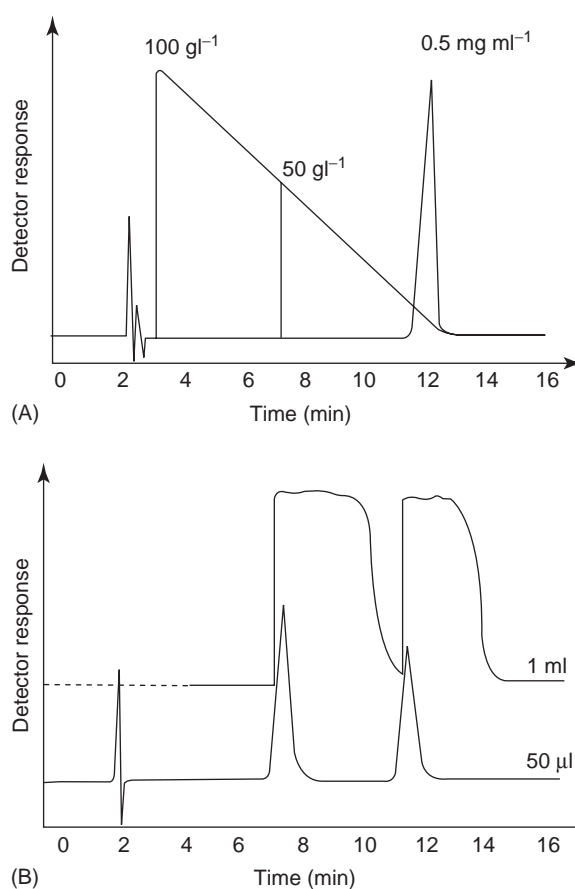
## Drug Substance Methods

### Preparative LC

In the synthesis of biologically active compounds, purification, isolation and identification procedures are necessary. The chromatographic methods that have traditionally been used by organic chemists (thin-layer chromatography (TLC), low-pressure open-column and flash chromatography) are supplemented with preparative LC. Major differences between preparative and analytical LC are column dimensions, mobile phase flow rates, and sample injection masses and volumes. Preparative column diameters often go from 8 mm to over 100 mm with lengths of 250–1000 mm. These large columns can require mobile phase flow rates of up to  $100 \text{ ml min}^{-1}$  and sample injection volumes up to several milliliters. Sample concentration can routinely run as high as hundreds of  $\text{mg ml}^{-1}$ . In the isolation of intermediates or final products in reaction schemes or of unknown impurities or related compounds, use is made of automatic fraction collectors and replicate injections.

Two different modes of operating preparative LC systems, touching bands and overlapping bands, were described along with optimization schemes for each in a summary paper (see bibliography). While production rates and yields can be balanced by an appropriate choice of sample size and flow velocity, preparative separations often use overload conditions. The results of concentration or sample volume overload on chromatographic peak shape are shown in **Figures 1A** and **1B**. Concentration overload will change a normal analytical-shaped peak into a fronted peak at earlier retention times, while sample volume overload will result in characteristic flat-topped peaks at later retention times than in analytical runs.

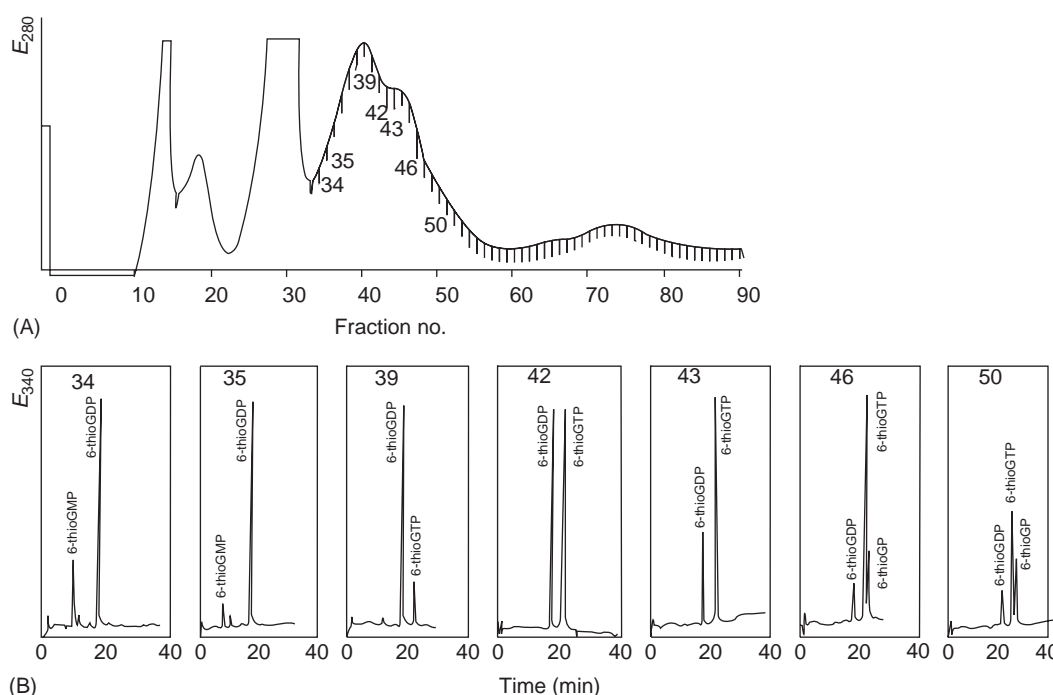
The use of low-pressure liquid column chromatography is an integral part of organic synthesis, with several recent examples described for pharmaceutically related compounds. For example, columns



**Figure 1** Preparative LC peak shapes resulting from (A) sample concentration overload and (B) sample volume overload.

packed with DEAE Sephadex A25 have been used to purify 6-thio-guanosine 5'-monophosphate, -guanosine-5'-diphosphate and -guanosine-5'-triphosphate from reaction mixtures with fractions collected and assayed by analytical LC as shown in **Figure 2**. The UV absorbance detector was set at 280 nm for the low-pressure separation, while LC of the collected fractions was monitored at 340 nm. It can be observed that 6-thioGTP reached a maximum content in fraction 43.

Prior to the routine use of mass spectrometry in the pharmaceutical laboratory, fraction collection in preparative LC was driven by UV detection. Every UV-active compound above a certain threshold setting will trigger fraction collection. While this method ensures nearly perfect capture of all analytes of interest, in many cases excess fractions are collected. Many commercial instrument companies have developed preparative LC systems that use mass spectrometry to detect and trigger the fraction collection. In this case, only the target molecule is selected, enabling higher throughput sample preparation. If target molecules do not respond in the MS, then



**Figure 2** (A) Elution profile of 6-thiopurines obtained on a low-pressure DEAE-Sephadex column showing times of 90 fractions collected. (B) Liquid chromatograms of seven collected fractions showing content of specific 6-thiopurines. GMP, guanosine-5'-monophosphate; GDP, guanosine-5'-diphosphate; GTP, guanosine-5'-triphosphate. (Reproduced with permission from Breter H and Mertes H (1990) *Biochimica et Biophysica Acta* 1033: 124–132; © American Society of Health-System Pharmacists.)

other detection techniques such as UV or ELS can be added to the system.

### Impurities Testing

The second important aspect of drug substance liquid chromatographic analysis in the pharmaceutical industry, and one of great importance to the control of raw materials and stability of the drug, is that of impurities analysis. While continuous monitoring of drug substance purity is conducted in the development of a chemical synthetic process, it is also used in stability testing and release testing of bulk-drug lots. This latter information is required by the respective regulatory authorities in drug product submissions to show the lack of variation in purity profile between lots of drug substance produced at different sites by different manufacturing processes, and to show stability of the drug itself. Process impurities and degradation products along with structural analogs have been termed 'related substances'. In general, process impurities include trace substances resulting from the drug substance synthesis, trace side-reaction products, or contaminants not removed by the isolation and purification scheme. These should remain at a constant level or possibly decrease as the bulk drug is incorporated into a dosage form and tested on a stability program. Degradation

products, on the other hand, which are assayed using the same or similar methods to the drug substance purity method, can accompany the bulk drug at release as a process impurity. More importantly, they can arise by the usual instability reaction mechanisms over time: hydrolysis, oxidation, photolysis, isomerization, rearrangement, reduction, substitution, elimination, etc. These products must be investigated, isolated, identified, and monitored in stability protocols to provide sufficient data for filing to show that their levels are under control.

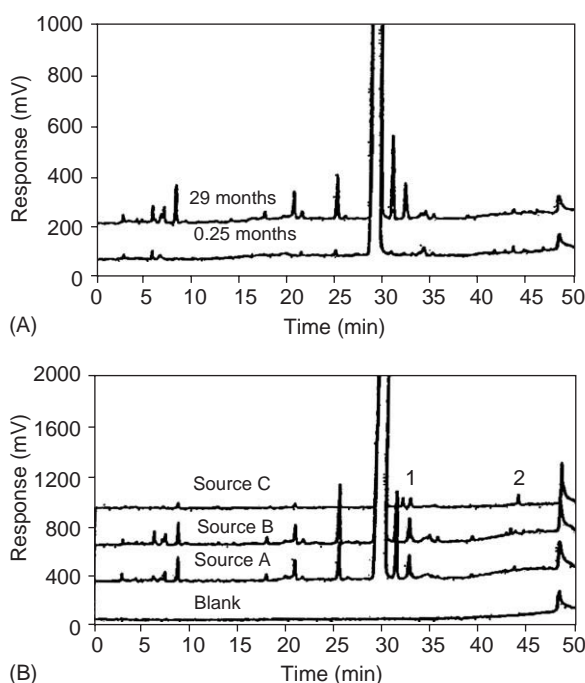
Regulatory authorities set various levels for individual and total impurities in both drug substance and drug product that must be reported and investigated. Usually >0.1% of the active component is set for measurement, with limits imposed for total and individual impurities in batch release of drug substance and possibly drug product as well. This testing is required in addition to the normal purity assay method, usually analytical LC, where specifications of  $\pm 1$ –2% are found for drug substance and drug product. Every pharmaceutical company must therefore be concerned with levels of impurities and degradation product in their compounds of interest and must have the required information on file both in house and with the regulatory authorities; however little of it has been published outside.

When the impurities or degradation products have been identified by spectroscopic means, their structure and often proof of structure are also included in the filing. The application of impurities information is in control of the synthetic process for the drug substance. In this way the homogeneity of the drug can be assessed as vendors of raw materials, reaction steps, recrystallization solvents, synthetic scale, and other process parameters are varied. The application of degradation product determination in drug substance and drug products will be further discussed in the section on stability below.

Several pharmaceutically relevant studies have been published that discuss impurity profiling by LC. The drugs ethynodiol diacetate, enalapril maleate, and piceurionium bromide were subjected to analysis of impurities utilizing diode-array detection with structure determination by spectroscopy. *E* and *Z* isomers of the 4-oestrene 17 $\alpha$ -ethynyl impurities, the *p*-tolyl analog and 2-dehydro-piceurionium were identified in these drugs, respectively. Related substances in the antibiotic cefaclor were measured using a reversed-phase gradient method with UV absorbance detection at 220 nm. Altogether, seven degradation products and 11 other process impurities were separated. Applications from this study in drug substance stability testing and in drug sourcing are illustrated in Figures 3A and 3B.

Among the related substances in many antibiotics are various structurally related components in the drug substance, the composite mixture of which is obtained in the synthetic or semisynthetic scheme and which gives rise to the drug efficacy. Control of the relative content of components is therefore necessary for these drugs. Test specification limits for the components are normally stated in terms of area percent of each, as maximum, minimum or a range of values for each or for the sums of several components. The types of components seen in these antibiotics were studied as impurities in the semisynthetic antibiotic clarithromycin, where detection limits of 0.1% w/w were found. Normalization factors were determined for each of 15 known related substances using ratios of the slope of linear calibrations for each substance to that for the reference impurity.

Routine impurities testing employs the use of HPLC with UV detection. However, this methodology requires additional testing to identify these analytes. By coupling HPLC to mass spectrometry, these impurities can be identified during the LC run. MS also offers a second confirmation (in addition to the retention time) as to the identity of the impurity.



**Figure 3** (A) Chromatograms showing related substances in cefaclor bulk drug at 0.25 and 29 months' storage at room temperature in amber glass. (B) Chromatogram showing related substances in cefaclor bulk drug from three different sources. (Reproduced with permission from Lorenz LJ, Bashore FN, and Olsen BA (1992) *Journal of Chromatographic Science* 30: 211–216; © Preston Publications.)

## Preformulation Studies

Preformulation testing involves the physical chemical characterization of drug substances. This consists of testing on the drug in the solid state, on the drug in solution and on the drug in prototype dosage forms. Solid-state testing includes thermal (differential scanning calorimetry and thermogravimetric analysis) and microscopic (including hot-stage) particle size, surface area, solid volume, cohesivity-flow, and batch variation studies. Stability of the drug substance is studied in relation to heat, light and humidity stress. Spectral properties of the drug substance as measured in spectroscopy or physical chemistry sections using mass spectrometry, nuclear magnetic resonance or infrared spectrometry, etc., are not usually considered preformulation studies. Solution studies can be classified into solubility studies (pH/solubility, non-aqueous and intrinsic dissolution), stability studies (pH, photochemical, thermal, nonaqueous, mechanistic), and ionization constant, partition coefficient and solubility product determinations. Preformulation testing can also encompass compatibility testing between the drug and individual excipients or groups of excipients, or between mixed products as occurs in hospital settings where more than one intravenous

product can be mixed for administration to patients. Stability properties of the drug substance, prototype dosage forms, and mixtures of dosage forms will be addressed below in the section on stability, although mention might be made here of one representative preformulation study that considered the stability of an antiviral compound. Results found using a reversed-phase LC method were compared to spectrophotometric analysis in order to determine reaction kinetics. Drug stability was found to be vehicle-dependent with suspensions in carboxymethylcellulose and methylcellulose being less prone to degradation than others.

An extremely fruitful area of research and publication in the field of preformulation testing by LC is that of partition coefficient or lipophilicity determination. Several standard procedures are available, including the relationship between partition coefficient and reversed-phase retention behavior ( $\log P$  versus  $\log k$ ). Many variations in method and calculations have been proposed, such as comparison with an internal standard partitioning, variation in organic content of the mobile phase for the retention model, and rudimentary assay of organic and aqueous phase in the shake-flask experiment by LC.

A representative publication using these methods is a study in which lipophilicities and solubilities were measured for four 3-hydroxy-3-methylglutaryl-coenzyme A reductase inhibitors by LC. Partition coefficients were determined in this study by LC analysis of drug content in the aqueous phase before and after equilibration with *n*-octanol and were then compared to retention factors in a reversed-phase LC system. The hydroxyacid form of pravastatin was found to have the lowest oil in water partition coefficient ( $P$  o/w) with simvastatin the highest.

## Formulation Development – Stability

In the formulation of prototype dosage forms, clinical supplies and proposed products, many considerations must be met such as sourcing, processing, manufacturing scale, dose requirements, marketing directives, bioavailability, packaging and documentation, besides the formula. These must be met no matter which dosage form is developed – oral solid (tablet, capsule), oral liquid (solution, suspension), ophthalmic, parenteral (solution, suspension), suppository, topical, aerosol, spray, transdermal, or implant. In the development of any dosage form, an overriding question, aside from the above requirements, is product stability along with the necessary drug substance stability. These must be demonstrated at each stage of development with short-term stress testing being useful in formulation prototype

selection and long-term formal stability studies required for submission to regulatory authorities in support of both clinical (IND) testing and proposed product (NDA) marketing. The stability studies on clinical products are conducted with the product contained in packaging not necessarily identical to the desired market package. Final long-term studies for drug product, however, must be conducted in the proposed package for marketing. Accumulated data from these studies in the form of reports are compiled along with formulation, manufacturing instructions, process information, specifications, tests and methods for the drug substance, drug product and excipients, results of batch analysis, certificates of analysis, components description, manufacturer description diagrams, organizational chart, drug master file references, manufacturing batch records, reprocess options, process validation, packaging components description, specifications, tests and sampling procedures, drug substance synthesis, drug reference standard characterization, drug substance preformulation report and other sections as part of the chemistry, manufacturing and control (CMC) submission.

Following IND approval and during clinical tests, drug product used in this testing must also be stability tested over the use period. This involves subjecting product stored in bulk or in its clinical packaging to testing on a schedule with storage at specified temperatures. Results are filed in support of clinical batch manufacturer.

Once a drug product is approved for marketing it is further stability tested with a commitment to test the first three batches for each dosage strength and the first production batch each quarter or year in the concurrent stability testing program. Testing can proceed each three months for the first year, each six months for the second year and at the end of the third year. Results obtained from these production batches stored under specified conditions are entered into annual reports that are submitted to the authorities to support the marketed product.

When reformulation of approved products, new packaging or new sources of drug substance are investigated, stability studies must be conducted as well. Needless to say, all of the above-mentioned stability testing requirements involve assays for potency, whether of drug product or drug substance, and many involve impurities and degradation product determinations as well. LC has been shown to be the technique of choice for most of this testing.

## Drug Substance Stability

Tests demonstrating drug substance stability take several forms, some of which were mentioned above,



but almost all of which utilize LC. Drug substance stability is addressed in preformulation studies on the solid-state bulk drug as well as the drug in various solutions. When the solid is stressed with heat, light, or humidity, assays and degradation product testing will show the extent of the drug resistance to these insults. When the resulting degradation products match retention times and spectra with known authentic related compounds, this serves as evidence for particular reaction pathways. Alternately, fractions containing unknown degradation products can be collected and structurally identified by spectral means. The combined techniques of LC-mass spectrometry (MS) and LC-MS-MS are routinely being applied successfully to this type of study in pharmaceutical firms.

Preformulation studies of a drug in solution give information on the ability of a drug to survive in a parenteral or oral solution product or the need for lyophilization or dry powder fill to avoid instability. The effects of solution pH, ionic strength, metal ions, common ions, oxygen levels, light intensity and wavelength, cosolvents and excipients can also be determined from this type of study. In fact, results from preformulation solution-state stability testing are extremely useful in elucidating possible bioavailability and drug absorption problems in relation to preclinical and clinical testing.

Drug substance stability is also addressed in analytical methods development validation studies that require a demonstration of selectivity as part of specificity testing. These results are included in CMC submissions in support of the analytical methods. In this testing, drug substance is commonly stressed with acid, base, oxidation, air, heat, light and other conditions to give rise to products that can be separated from the parent compound.

A fourth type of drug substance stability testing that is conducted on pharmaceuticals is formal drug substance long-term stability testing. In these procedures the bulk drug is stored in packaging representative of what would be found on a large scale in the production setting, that is, double polyethylene (PE) bags contained in a fiber box or tube. In certain cases, for shipping purposes bulk drug is packaged in double PE bags in plastic or metal containers, each of which should be simulated in the formal stability study. For comparison, drug substance is also stored in glass ampoules or vials, usually under refrigerator conditions, to provide an analytical reference. A formal protocol of required testing is followed, results are accumulated, and reports are written delineating any stability problems encountered.

The conditions of storage for these formal stability studies are enunciated by regulatory authorities, with

international harmonization tending towards the most encompassing set of rules. The 2003 Revision 2 draft of the Guidance for Industry, Q1A Stability Testing of New Drug Substances and Products by the International Conference on Harmonization (ICH) contains guidelines on the conditions of storage for these formal stability studies. Depending on whether the substance is stored at room temperature, in the refrigerator or freezer, a set of temperatures and relative humidities are given. The current revision of the draft document should always be consulted for specific details. Three batches of drug substance of at least pilot-scale size should be tested in this program, also to include stress testing on a single batch, which will show the effect of high temperature, susceptibility to hydrolysis (in solution), pH stability, oxidation, and light stress in formal studies.

Although much research has been conducted on drug substance stability in relation to formulation stability as described below, the area of drug substance stability alone that has probably received most study is that of solution stability. This typical preformulation testing along with the required LC methods development often includes kinetic analysis of the degradation pathway and products. Some examples of this type of work are outlined here. A full validation study on apidrine indicated that the stability of the raw drug and its dosage form are the same. The samples were stress tested by exposure to heat, oxidation, and acidic and basic environments. LC-MS was used in a stability study on cloxacillin in order to identify degradation products.

The stability of drug substances either as dry powder or in solution and in formulations under light stress is also an area of current interest for pharmaceuticals, often using LC to study drug degradation reactions. Benzydamine, for example, was shown to be unstable under UV radiation in gel, cream, or mouthwash preparations.

A related field of research that has found certain applications in pharmaceutical analysis is flash photolysis. This technique has been adapted with UV absorbance, fluorescence, and electrochemical detection of photolysis products separated by LC to enhance limits of detection and quantification. Products of these reactions have also been further investigated following LC separation.

### Drug Product Stability

During bulk drug stability and preformulation testing, preliminary formulations and processes are developed that are first stability tested on a short-term, accelerated basis. Drug product intended for clinical use after IND approval must be stability



tested before IND submission, with preferably 3–6 months' data minimum at submission time. This means that specific analytical methods must be developed during the latter stages of formulation development of the clinical product.

Further formulation changes, selection, and approvals occur during clinical testing with the intent of developing the proposed product for marketing. Each formulation must have new stability-indicating methods developed for assay and related compound measurement. Stability properties of all products developed for commercialization must be studied and reported in NDA (MAA) or supplemental submissions.

The stability of excipients in dosage forms has not been studied extensively in the past by LC methodology. Regulatory authorities, however, are presently tending towards recommendations of more stability specific assays for excipients, which will supplement the required identification testing. Of the few studies reported, one used ion chromatographic methods to measure sodium metabisulfite in parenteral products. The antioxidant was measured following conversion to the sulfite ion in aqueous solution along with its oxidation product (sulfate ion) by conductivity detection.

### Compatibility Testing

The compatibility testing that will be dealt with here is that of liquid products, both oral and parenteral, with single drug substances in solution or suspension and with mixtures of drug substances in solution or suspension. This type of preparation is common in hospital or clinical settings and is also common in formal stability studies on drug products where demonstrations of stability following dilution or storage in plastic containers, infusion sets, etc., are necessary. Compatibility studies between individual bulk excipients and mixtures of bulk excipients and one or more bulk drug substances will not be considered, although several pharmaceutical firms conduct this type of testing with analysis of drug stability in these mixtures by LC.

Among single drug component studies, usual vehicles for dilution are 5% dextrose (D5W), 0.9% sodium chloride (normal saline, NS), aqueous buffers, peritoneal dialysis fluid, nonaqueous solvents, water for injection, phosphate-buffered saline, bacteriostatic water for injection, bacteriostatic sodium chloride, Ringer's injection, and lactated Ringer's. Stability studies can also be carried out on the drug product solution as such or in specific containers or injection devices. Solutions and suspensions can also be prepared extemporaneously and stability tested to show worthiness for oral, ophthalmic, or rectal administration. The following is an example of the first

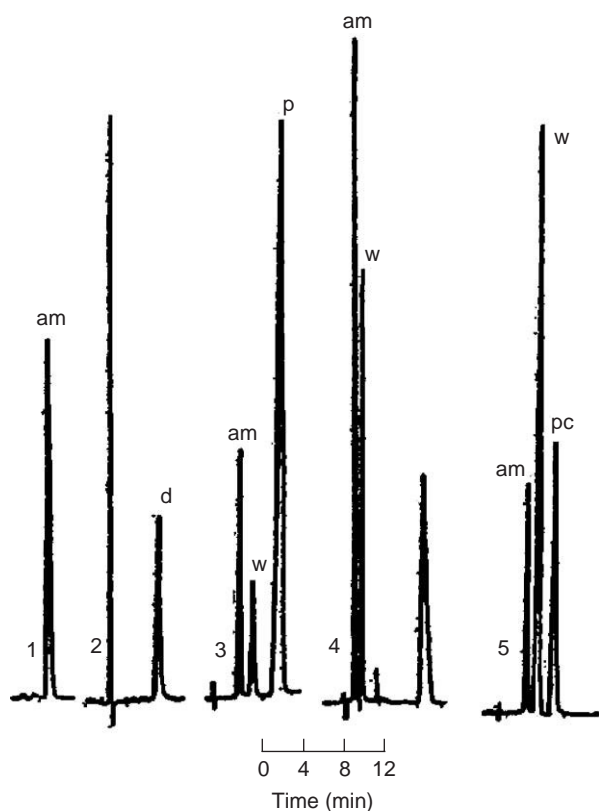
listed type of research with analysis conducted by stability-indicating LC methods. Pibenzimol-HCl was diluted to  $0.15 \text{ mg ml}^{-1}$  in 5% dextrose, 0.9% NaCl, and lactated Ringer's solution, where stability was found for 24 h each at 25°C. Stability in glass and poly(vinyl chloride) (PVC) containers was compared with the finding of increased loss after 24 h in lactated Ringer's in glass and plastic and in normal saline in PVC bags. Stability testing of injectable solutions prepared without dilution with common large-volume parenterals (LVPs) or other solutions has also been reported.

Compatibility testing has also been conducted on mixtures of solution dosage forms and parenteral drug products. These mixtures have been stability tested in one or more intravenous solvents as listed above. Methods utilized were reported to be stability-indicating by reference or by actual inclusion of data on separation of components from stressed drug substance. One such study concerned the stability of amrinone and digoxin, procainamide-HCl, propranolol-HCl, sodium hydrogencarbonate, potassium chloride, and verapamil-HCl in intravenous admixtures prepared in 0.45% NaCl or 5% dextrose. **Figure 4** shows five chromatograms from this study run by different methods as necessitated for quantification of the separate components.

### Chiral Separations of Pharmaceuticals

Current national and proposed ICH directives on stability testing of drug substances and drug products suggest that enantiomeric purity be controlled in drug manufacture, storage, and shipment for both bulk drug and product for molecular entities with chiral centers. While development of racemic drugs is still permitted, examples are known in which enantiomeric pairs both have the same biological activity, or have different biological activity, or in which one member is active and the opposite isomer is completely inactive. Because of the possibility of interconversion and racemization, stereochemically specific identity tests and assay methods are to be in place for both drug substance and drug product in order to develop data for submission. If a lack of interconversion was once demonstrated under proposed storage conditions, it might not be necessary to collect further data on each batch of drug or product.

The specific chiral LC methods that have been in use over the past decade have involved three general techniques. The first is precolumn derivatization of the enantiomer or mixture with single-enantiomer reagents to yield diastereomers that are



**Figure 4** Chromatograms of (1) amrinone; (2) digoxin; (3) amrinone, Win 55870 (degradation product of amrinone) and propranolol; (4) amrinone, Win 55870 and verapamil; and (5) amrinone, Win 55870 and procainamide by five different methods. am = amrinone, d = digoxin, w = Win 55870, p = propranolol, v = verapamil and pc = procainamide. (Originally published in Riley and Junkin (1991). © 1991 American Society of Pharmacists, Inc. All rights reserved. Reprinted with permission (reference 92144).)

then separated by nonchiral LC. Secondly, constituents have been added to mobile phases to afford separation of enantiomer pairs on nonchiral stationary phases. Finally, chiral stationary phases have been developed commercially and utilized for the separation of diverse types of drug optical isomers. Abundant references can be found on either of these techniques and several high-quality reviews have been written.

Chiral stationary phases that are currently available can be classified into those containing cavities (cellulose derivatives, cyclodextrins, synthetic polymers, crown ethers, and chiral imprinted gels), affinity phases (bovine serum albumin, human serum albumin,  $\alpha$ -glycoprotein, enzymes), multiple hydrogen-bond phases,  $\pi$ -donor and  $\pi$ -acceptor phases, and chiral ligand exchange phases. This classification scheme was used in a review that gave numerous pharmaceutical examples of separation by

each mechanism. Another review included much information on the use of nonchiral stationary phases with mobile phases developed for chiral separations.

## Quality Control

The control of pharmaceutical quality is one of the primary objectives of drug manufacturers. Quality control (QC) has been defined as actions such as testing, monitoring, and inspecting taken to detect and control defects. In contrast, quality assurance has been defined as all actions necessary, including tests for quality of design and conformance, to provide adequate confidence that an item is fit for use. The control of drug substance and drug product quality, as monitored routinely in operational division QC departments, involves developing sets of tests and specifications for raw materials, intermediates of interest and final products, performing the tests on the subject material, compiling the results and submitting them to regulatory authorities on a regular basis. Reasons for rejection of batches and deviations from specification must be well documented and explained.

The use of LC in the quality control of raw materials and finished products (release and stability testing) is primarily in the area of potency assay and impurity testing. It has also been used in dissolution testing, for which several commercially available automatic and robotic systems incorporate LC with dissolution sampling. Very few useful publications exist on QC applications of LC other than several examples that list quality control or quality assurance in their titles as part of a methods development study. One report, however, did summarize all LC methods of use in quality control found in the US Pharmacopeia (USP) XX and its supplements and addenda. This listed the drug or drug product, stationary phase, mobile phase, detector, and uses of each test. These included identification, purity assay, content assay, content uniformity, and dissolution. A summary paper has also been written on dissolution testing with analysis by LC that would find application in QC work. This study discussed advantages of LC over conventional spectrophotometric analysis as well as multicomponent and low-dose product applications.

## Method Validation

Validation of LC methods in the pharmaceutical industry follows prescribed guidelines as published by regulatory authorities and compendia. The FDA

guideline (2001) has mandated that demonstrations of precision of 20% and accuracy of 80–120% of theoretical method specificity, detection limits for degradation products, and structures of those products be submitted in validations of drug substance LC methods. Methods for drug products are to contain this same information along with recovery data from spiked placebos, proof of no placebo interference (stressed and unstressed), precision results (interlaboratory, interday and intercolumn), and demonstrations of degradation product separation. The USP includes recommendations on method precision, accuracy, limits of detection and quantification as well as selectivity. Proof of method robustness (ruggedness) and linearity of detector response with concentration are also suggested. Short validation packages are made available to the authorities in the IND submission, while all requirements must be fulfilled by NDA submission time.

Although several reports have appeared on the validation of LC methods for specific pharmaceuticals, one particularly comprehensive study discussed system suitability, peak purity, system resolution, system selectivity, and stability-indicating properties. The example used in this work was the method developed for pipecuronium bromide. A final comprehensive method validation review has also shown the importance of each facet: specificity, accuracy, precision, sensitivity, and robustness.

## Acknowledgment

The author wishes to acknowledge the expert assistance of Dr J T Stewart, University of Georgia, for his critical evaluation of this manuscript.

**See also:** **Liquid Chromatography:** Chiral; Liquid Chromatography–Mass Spectrometry. **Pharmaceutical Analysis:** Drug Purity Determination. **Quality Assurance:** Quality Control.

## Further Reading

- Bauser M (2002) Preparative high-performance liquid chromatography-mass spectrometry for the high-throughput purification of combinatorial libraries. *Journal of Chromatographic Science* 40: 292–296.
- Bouma M, Nuijen B, Jansen MT, *et al.* (2003) Development of a LC method for pharmaceutical quality control of the antimetastatic ruthenium complex NAMI-A. *Journal of Pharmaceutical and Biomedical Analysis* 31: 215–228.
- Brown R, Caphart M, Faustino P, *et al.* (2001) Analytical procedures and method validation: Highlights of the FDA's draft guidance. *LC GC North America* 19: 74–77.
- Gazdag M, Babjik M, Brlik J, *et al.* (1998) Estimation of impurity profiles of drugs and related materials. Part 18. Impurities and degradation products of mazipredone. *Journal of Pharmaceutical and Biomedical Analysis* 17: 1029–1036.
- Gubitz G and Schmid MG (2001) Chiral separation by chromatographic and electromigration techniques. A Review. *Biopharmaceutics and Drug Disposition* 22: 291–336.
- Houri JJ, Gouaillard G, Acar V, *et al.* (2002) LC determination of diphemanil methylsulfate: application to stability study of stored pharmaceutical formulations. *Journal of Pharmaceutical and Biomedical Analysis* 28: 811–818.
- Igo DH, Brennan TD, and Pullen EE (2001) Development of an automated in-line microfiltration system coupled to an HPLC for the determination of solubility. *Journal of Pharmaceutical and Biomedical Analysis* 26: 495–500.
- Nicolas EC and Scholz TH (1998) Active drug substance impurity profiling: Part I. LC/UV diode array spectral matching. *Journal of Pharmaceutical and Biomedical Analysis* 16: 813–824.
- Nicolas EC and Scholz TH (1998) Active drug substance impurity profiling: Part II. LC/MS/MS fingerprinting. *Journal of Pharmaceutical and Biomedical Analysis* 16: 825–836.
- Nuijen B, Rodrigues-Campos IM, Noain CP, *et al.* (2001) HPLC-UV method development and impurity profiling of the marine anticancer agent aplidine in raw drug substance and pharmaceutical dosage form. *Journal of Liquid Chromatography and Related Technologies* 24: 3119–3139.
- Okamoto M (2002) Reversal of elution order during the chiral separation in high-performance liquid chromatography: Review. *Journal of Pharmaceutical and Biomedical Analysis* 27: 401–407.
- Ozkan Y, Savascediler A, Tascetil C, *et al.* (2003) Drug dissolution studies and determination of deflazacort in pharmaceutical formulations and human serum samples by RP-HPLC. *Journal of Liquid Chromatography and Related Technologies* 26: 2141–2156.
- Savascediler A, Ozkan CK, Ozkan Y, *et al.* (2003) Development and validation of an RP-HPLC method for the determination of valacyclovir in tablets and human serum and its application to drug dissolution studies. *Journal of Liquid Chromatography and Related Technologies* 26: 1755–1767.
- Serajuddin ATM, Ranadive SA, and Mahoney EM (1991) Relative lipophilicities, solubilities, and structure-pharmacological considerations of 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase inhibitors pravastatin, lovastatin, mevastatin, and simvastatin. *Journal of Pharmaceutical Sciences* 80: 830–834.
- Sun XX, Sun LZ, and Aboul-Enein HY (2001) Chiral derivatization reagents for drug enantioseparation by high performance liquid chromatography based upon precolumn derivatization and formation of diastereomers: Enantioselectivity and related structure. Review. *Biomedical Chromatography* 15: 116–132.

## Isotope Separations

L Lešetický, Charles University, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The application of isotopes in science, in general, and in analytical chemistry, in particular, is based on two rather contradictory suppositions. Tracer experiments demand the same or very similar behavior of the isotopically modified and unmodified compounds. The study of the effects of isotopes requires the recognition, measurement, and interpretation of minute differences in the chemical and physical properties of compounds that differ only in their isotopic composition (isotopomers). These differences produce a different behavior of the isotopomers in all types of chromatographic processes, and these isotope effects have been demonstrated experimentally.

As early as in 1938, Urey and co-workers published a paper about enrichment of  $^6\text{Li}$  by means of ion-exchange chromatography and since then work in this field has been continuing till date. Gas-solid chromatography has been used for the separation of hydrogen isotopes, noble gases, some other gaseous elements, and simple compounds. Low-molecular-mass volatile compounds labeled mainly with hydrogen isotopes have been separated by gas-liquid chromatography. During the late 1950s and the 1960s, papers were published concerning small but observable isotopic fractionation with liquid chromatography (LC) of isotopically labeled organic compounds. Over the last 20 years, during which high-performance liquid chromatography (HPLC) has become a common analytical technique, many papers have been published dealing with separation of labeled compounds from their unlabeled counterparts. A review of these works, with a possible explanation of the mechanism of the separation process, is presented.

### Isotope Effects

Chromatography can be considered as a process in which the compounds to be separated interact with a stationary phase and a mobile phase. These mutual interactions differ in magnitude even for different isotopes of the same element. The magnitude of the interaction energy for an isotopic pair depends on many parameters that are discussed below.

It is possible to draw the following general conclusions from the experimental data on isotope effects on the physical properties of molecules:

1. The  $\text{C}-^2\text{H}$  bond is shorter than the  $\text{C}-^1\text{H}$  bond (e.g., in ethane the  $\text{C}-^1\text{H}$  bond length is 111.2 pm; in hexadeuterioethane the  $\text{C}-^2\text{H}$  bond length is 110.7 pm) and exhibits higher electron density than the  $^1\text{H}$  bond. The deuterium atom appears to be smaller than the hydrogen atom. There are small, but measurable differences in the dipole moments, e.g.,  $\mu(\text{CH}_3-^2\text{H}) = 3.7 \times 10^{-32} \text{ C m}$ ;  $\mu(^1\text{HCl}) - \mu(^2\text{HCl}) = 1.7 \times 10^{-32} \text{ C m}$ .

2. The  $\text{C}-^2\text{H}$  bond has a polarizability lower than that of the  $\text{C}-^1\text{H}$  bond.

3. The  $\text{C}-^2\text{H}$  stretching vibrational frequency ( $\sim 2200 \text{ cm}^{-1}$ ) is lower than the corresponding  $\text{C}-^1\text{H}$  frequency ( $\sim 3000 \text{ cm}^{-1}$ ).

4. Heavier isotopes have smaller atomic volume; deuterated compounds have smaller molar volumes than the corresponding unlabeled compounds. The molar volume differences between, e.g., deuterated and ordinary benzene are due to a molecular size effect caused by differences in zero point intermolecular motion of the molecules.

5. Deuterium ( $^2\text{H}$ ) is more electropositive than protium ( $^1\text{H}$ ) and thus some isotope effects can be discussed in terms of inductive effects. This has been clearly demonstrated in the measurement of the secondary isotope effect on ionization equilibria of some deuterated carboxylic acids and protonated amines as shown in Table 1.

6. Deuterated compounds are less lipophilic than the corresponding unlabeled compounds.

7. The above-mentioned experimentally proved statements can also be generalized for other isotope pairs.

**Table 1** Isotope effect on dissociation constants<sup>a</sup>

Compound <sup>b</sup>	$\log(K_{\text{H}}/K_{\text{D}})$
$^2\text{HCOOH}$	$0.035 \pm 0.002$
$\text{C}^2\text{H}_3\text{COOH}$	$0.014 \pm 0.002$
$(\text{C}^2\text{H}_3)_3\text{CCOOH}$	$0.018 \pm 0.001$
$\text{C}_2^2\text{H}_5\text{COOH}$	$0.035 \pm 0.002$
$(2,6-^2\text{H}_2)\text{benzoic acid}$	$0.003 \pm 0.001$
$\text{C}^2\text{H}_3\text{N}^+\text{H}_3$	0.051
$(\text{C}^2\text{H}_3)_2\text{N}^+\text{H}_2$	0.117
$(\text{C}^2\text{H}_3)_3\text{N}^+\text{H}$	0.207

<sup>a</sup>Willi AV (1983) *Isotope Effects in Chemical Reactions* [in German]. Stuttgart: G. Thieme.

<sup>b</sup>Only the isotope-modified compound is specified in the table.



All these experimentally demonstrated phenomena may influence the chromatographic separation processes. Isotope effects on the chromatographic behavior of compounds labeled with isotopes of heavier elements (carbon, nitrogen, oxygen, etc.) are so small that they can only be detected for simple compounds of low relative molecular mass. The following discussion is limited to deuterated and tritiated compounds and only a few examples are given of separation of heavier isotopes.

It should be pointed out that discussion of isotope effects in terms of inductive or other electronic effects is sometimes helpful but represents a gross simplification. A more exact description of the origin of isotope effects uses zero-point energy and vibrational frequencies of molecules in question.

## Separation Processes

Two basic types of the LC separation processes can be distinguished. By adsorption chromatography on 'classic column' or 'straight phase' HPLC, usually 'ordinary' (unlabeled) compounds elute first, i.e., the separation factor is less than unity. ( $\alpha = k_1/k_2 < 1$ ; subscripts 1, 2, and 3 refer to protium, deuterium, and tritium ( $^3\text{H}$ ), respectively;  $k_i$  is a capacity factor). As mentioned above, deuterated or tritiated compounds are more polar and thus are more strongly bound to polar stationary phases.

Partition chromatography and reversed-phase LC usually lead to the opposite result, i.e., a heavier isotopomer is eluted first ( $\alpha > 1$ ). It has been suggested that a major contribution to the isotope effect is a hydrophobic interaction. The fact that the separation factors are higher (see Table 2) when the mobile phase contains more water demonstrates that they are affected by more restricted motion of C-H bonds (in solute), caused by tighter solvation of C-H bonds within the aqueous mobile phase relative to the hydrophobic stationary phase. On the other hand, a less restricted motion of the C-H bonds in the stationary phase would tend to favor protium over deuterium, and this could contribute to the observed isotope effect.

Another important factor is the position of the label in a molecule. Because these effects are essentially primary isotope effects, the best case is when the isotopic atom is pushed out from the rest of the molecule and is well accessible for an interaction. When an atom in question is in the shadow of a bulky group (e.g., *ortho* hydrogen in benzoic acid) or in an inconvenient configuration, neither hydrophobic interactions nor adsorption can take place, or are greatly suppressed and no separation is attained. Other contributions such as differences in solubility,

**Table 2** LC separation factors of some hydrocarbons versus their perdeuterated analogs

Compound	$\alpha = k_1/k_2$	Conditions <sup>a,b</sup>
[ $^2\text{H}_{14}$ ]Hexane <sup>a</sup>	1.036	51
	1.049	40
[ $^2\text{H}_{18}$ ]Octane <sup>a</sup>	1.054	51
[ $^2\text{H}_{12}$ ]Cyclohexane <sup>a</sup>	1.044	33
[ $^2\text{H}_6$ ]Benzene <sup>a</sup>	1.043	23
	1.048	18
[ $^2\text{H}_6$ ]Benzene <sup>b</sup>	1.049	30
[ $^2\text{H}_8$ ]Toluene <sup>a</sup>	1.046	33
	1.057	23
[ $^2\text{H}_8$ ]Toluene <sup>b</sup>	1.052	40
[ $^2\text{H}_{10}$ ]Phenathrene <sup>b</sup>	1.067	70
[ $^2\text{H}_{14}$ ]Durene <sup>b</sup>	1.071	70
[ $^2\text{H}_{10}$ ]Biphenyl <sup>b</sup>	1.062	55

<sup>a</sup>  $\mu$ -Bondapak C<sub>18</sub> column; methanol in water (mol%); Tanaka (1977).

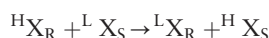
<sup>b</sup> Ultrasphere C<sub>18</sub> column; acetonitrile in water (v/v); Baweja R, *et al.* (1987) Application of reversed-phase HPLC for the separation of deuterium and hydrogen analogues of aromatic hydrocarbons. *Analytica Chimica Acta* 192: 345-348.

molar volume, and adsorption on the residual non-derivatized sites on the silica particles seem to be of far less importance for the magnitude of the isotope effect. The above discussion dealt with a direct interaction of an isotopic atom or bond with the stationary and mobile phases (primary isotope effect). However, there also exist chromatographic separations based on secondary isotope effects. In molecules deuterated or tritiated in the position adjacent to an ionizable functional group, the magnitude of the isotope effect is often greatly enhanced. As follows from Table 1  $\alpha$ -deuterated amines are stronger bases than unlabeled amines, and  $\alpha$ -deuterated acids are weaker acids than their protiated counterparts. In an elution system with pH value close to the  $\text{pK}_a$  values of the deuterated/unlabeled amine pair, this isotope-induced base strengthening would alter the ratio of protonated and unprotonated forms, leading to differences in chromatographic mobility. However, at pH much higher or much lower than the  $\text{pK}_a$  value, where both labeled and unlabeled amines are completely protonated or unprotonated, the  $\text{pK}_a$  effect of isotopic substitution on chromatographic separation is greatly suppressed. A similar effect was observed for labeled carboxylic acids.

Ion-exchange chromatography could be regarded as one form of LC because of the similarity of the experimental procedures. The mechanism of the separation process for labeled polar (ionizable) organic compounds can be simply explained in terms of the primary or secondary isotope effect on the ionization constant. For the separation of the isotopic metal ions other mechanisms can be involved.



The isotopic exchange reaction taking place during the separation process can be expressed as



where L and H refer to lighter and heavier isotopes, respectively, and R and S refer to the resin and solution phases, respectively. The equilibrium constant of this isotopic exchange reaction is larger than unity in the case where the heavier isotope is enriched in the solution phase and the lighter one in the resin phase.

The single-stage (elemental) separation coefficient is then defined as

$$\varepsilon = \alpha - 1 = (C_{L,R} \cdot C_{H,S} / C_{L,S} \cdot C_{H,R}) - 1$$

where C is the concentration or activity.

In almost all cases, the experiments yielded values  $\alpha > 1$ . The heavier isotope is enriched in the solution phase and is eluted first. The theory of isotope separation by elution chromatography and the procedure for the calculation of the separation factor from experimental data was elaborated by Glueckauf. For the case of displacement chromatography the theory was developed by Kakihana and Oi.

The main parameters that can influence the separation process are:

1. The mass difference of the isotope pair  $\Delta M$  (or  $\Delta M/M$ ). For the enrichment of magnesium isotopes, the following values were observed:  $\alpha_{24,25} = 1.015$ ;  $\alpha_{24,26} = 1.029$ ;  $\alpha_{25,26} = 1.014$ .
2. Pauling ionic radii of the isotopes in question. The difference of the Pauling radii for  ${}^6\text{Li}$  and  ${}^7\text{Li}$  is 0.04 pm,  ${}^7\text{Li}$  being smaller.
3. The nature and the type of ion exchanger. The classical strongly acidic cation exchanger and strongly basic anion exchanger usually yielded relatively low separation factors. It was found that the separation factor increases as the degree of cross-linking increases. Crown-ethers and azacrown-ethers (cryptands) bound on the Merrifield-type resins were successfully used in the last two decades, the separation factors being at least one order of magnitude higher. Even though the interaction between the cryptand and the cation is not of the pure ionic nature, the exchange ions takes place. The third type of exchanger is the special (hydrated) inorganic oxides, e.g., hydrous manganese(IV) oxide, tin and titanium antimonate, monoclinic antimonite acid, cubic niobic, and tantalate acid.
4. Type of the counterion or complexing agent.
5. Nature, concentration, pH of the eluent. These parameters can influence dramatically not only the magnitude of the separation factor but even the order

of elution. For example, when the zirconium isotopes were eluted by triammonium citrate the heavier  ${}^{96}\text{Zr}$  was eluted first, while in the diammonium citrate system the lighter one ( ${}^{90}\text{Zr}$ ) was enriched in the solution phase and eluted first.

6. Temperature. At lower temperature the separation factors are larger.

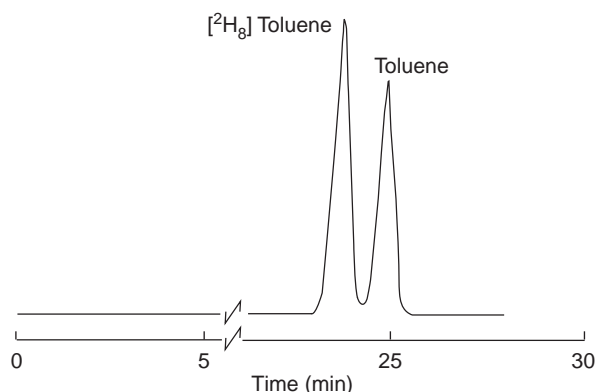
## Survey of Separations

### Hydrocarbons and Their Simple Derivatives

Separation of perdeuterated aliphatic and aromatic hydrocarbons has been performed on reversed-phase columns with similar results (Figure 1). A higher content of water in the mobile phase leads to higher separation factors (Table 2).

Chromatography of various isotopomers of benzoic acid shows a moderate effect of number of deuterium atoms on the separation factor; however, an important effect is exerted by the position of the label (Table 3). Isotopomers containing deuterium in the *ortho* positions display lower isotopic separation than isotopomers with deuterium in *meta* or *para* position. The isotope effects associated with the *ortho* positions are minimal, probably because interactions between *ortho* C–H or C– ${}^2\text{H}$  bonds and the stationary phase are minimized by the steric shielding of the carboxyl group.

[9,10,12,13- ${}^3\text{H}$ ]Linoleic acid methyl ester and [9,10,12,13- ${}^3\text{H}$ ]oleic acid methyl ester were chromatographed on silica impregnated with silver nitrate attaining partial separation from the appropriate [1- ${}^{14}\text{C}$ ]methyl esters. The tritium atom bound on the



**Figure 1** Reversed-phase LC separation of [ ${}^2\text{H}_8$ ]toluene and unlabeled toluene. Mobile phase: water–acetonitrile 60:40 (v/v). (Reprinted with permission from Baweja R, *et al.* (1987) Application of reversed-phase HPLC for the separation of deuterium and hydrogen analogues of aromatic hydrocarbons. *Analytica Chimica Acta* 192: 345–348; © Elsevier.)

olefinic carbon atoms affects the equilibrium constant of the formation of the silver–olefin complex.

Enantiomers of phenyl-[<sup>2</sup>H<sub>5</sub>]phenylmethanol were separated by reversed-phase HPLC based on the isotopic chirality using cellulose tribenzoate-coated silica as a stationary phase and 2-propanol/hexane (5:95, v/v) mixture as a mobile phase. The stationary phase showed preferential retention of (*R*)-(–) enantiomer with a separation factor of 1.008. Complete separation was achieved by a recycle column system after 65 cycles.

### Natural Substances

Tritium-labeled steroids have often been used for the elucidation of biochemical and physiological trans-

**Table 3** LC separation factors of unlabeled versus deuterated carboxylic acids

Labeled compound	$\alpha = k_1/k_2$
[ <sup>2</sup> H <sub>31</sub> ]Palmitic acid	1.076 <sup>a,b</sup>
[ <sup>2</sup> H <sub>21</sub> ]Lauric acid	1.066 <sup>a,c</sup>
[ <sup>2</sup> H <sub>5</sub> ]Benzoic acid	1.040 <sup>a,d</sup>
	1.038 <sup>e</sup>
[3,4,5- <sup>2</sup> H <sub>3</sub> ]Benzoic acid	1.029 <sup>e</sup>
[2,3,5- <sup>2</sup> H <sub>3</sub> ]Benzoic acid	1.023 <sup>e</sup>
[3,4- <sup>2</sup> H <sub>2</sub> ]Benzoic acid	1.019 <sup>e</sup>
[3,5- <sup>2</sup> H <sub>2</sub> ]Benzoic acid	1.019 <sup>e</sup>
[2,4- <sup>2</sup> H <sub>2</sub> ]Benzoic acid	1.013 <sup>e</sup>
[2,5- <sup>2</sup> H <sub>2</sub> ]Benzoic acid	1.010 <sup>e</sup>
[2,6- <sup>2</sup> H <sub>2</sub> ]Benzoic acid	> 1.010 <sup>e</sup>

<sup>a</sup>  $\mu$ -Bondapak C<sub>18</sub> column; Tanaka *et al.* (1977).

<sup>b</sup> 64 mol%; CH<sub>3</sub>OH–H<sub>2</sub>O.

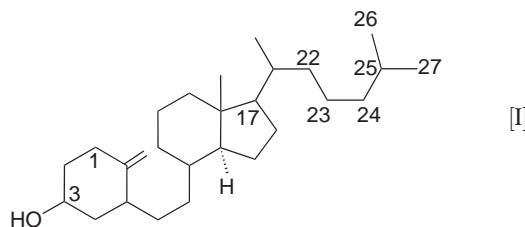
<sup>c</sup> 51 mol%; CH<sub>3</sub>OH–H<sub>2</sub>O.

<sup>d</sup> H<sub>2</sub>O, pH 2.51.

<sup>e</sup> Hypersil C<sub>18</sub> column; CH<sub>3</sub>OH–H<sub>2</sub>O (3:7, v/v) + 1% HCOOH; Lockley *et al.* (1989).

formations. During the chromatographic purification of such compounds, isotopic fractionation has often been observed. Good, sometimes baseline, separations of tritiated aldosterone, cortisone, oestrone, testosterone, prednisolone, and estradiol from their unlabeled or <sup>14</sup>C-labeled analogs were obtained in reversed-phase LC, with tritiated compounds eluting first. Other techniques such as column partition chromatography and paper chromatography have also been successful.

Vitamin D (cholecalciferol, structure [I]) is hydroxylated in living organisms to 1,25-dihydroxy- or 24,25-dihydroxy-derivatives. A significant chromatographic isotope effect was observed when both derivatives were labeled with tritium in positions 26 and 27, whereas labeling in positions 23 and 24 caused a substantially smaller effect. A similar behavior was found in the chromatography of the corresponding trimethylsilyl ethers (Table 4). As indicated in the table, the heavier isotopomers were eluted first in the reversed phase, and later than lighter isotopomers in the normal phase.



Catalytic hydrogenation with protium, deuterium, and tritium of echinocandin B, a macrocyclic peptide

**Table 4** LC separation factors of vitamin D<sub>3</sub> metabolites versus their tritiated counterparts

Vitamin D <sub>3</sub> metabolite	$\alpha = k_1/k_3$	Eluent
1,25-Dihydroxy-[23,24- <sup>3</sup> H]-TMS <sup>a</sup>	0.977 <sup>b,c</sup>	Hexane–CH <sub>2</sub> Cl <sub>2</sub> (85:15)
	0.989 <sup>b,d</sup>	Hexane–CH <sub>2</sub> Cl <sub>2</sub> –CH <sub>3</sub> CN (90:10:0,035)
1,25-Dihydroxy-[23,24- <sup>3</sup> H]	0.991 <sup>b,d</sup>	Hexane–2-propanol
25( <i>R</i> ),26-Dihydroxy-[23,24- <sup>3</sup> H]-TMS <sup>a</sup>	0.965 <sup>b,d</sup>	Hexane–CH <sub>2</sub> Cl <sub>2</sub> (85:15)
1,25-Dihydroxy-[26,27- <sup>3</sup> H <sub>6</sub> ]	0.983 <sup>e,f</sup>	Hexane–ethanol (94:6)
24,25-Dihydroxy-[26,27- <sup>3</sup> H <sub>6</sub> ]	0.965 <sup>e,f</sup>	
25,26-Dihydroxy-[26,27- <sup>3</sup> H]	1.008 <sup>e,g</sup>	Methanol–water (3:1)
1,25-Dihydroxy-[26,27- <sup>3</sup> H <sub>6</sub> ]	1.023 <sup>e,g</sup>	Methanol–water (1:1)

<sup>a</sup> TMS = trimethylsilyl derivative.

<sup>b</sup> Halloran BP, Bikle DD, and Whitney JD (1984) Separation of isotopically labeled vitamin D metabolites by HPLC. *Journal of Chromatography* 303: 229–233.

<sup>c</sup> Zorbax-sil column.

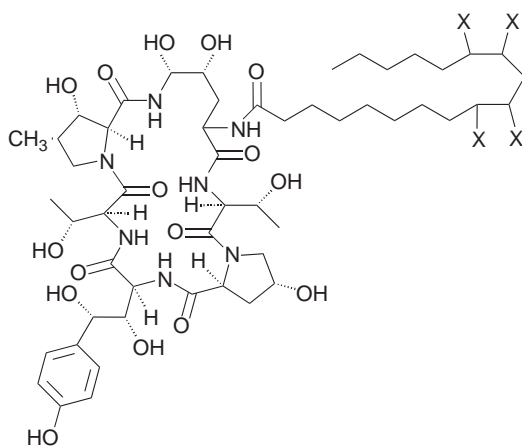
<sup>d</sup>  $\mu$ -Porasil column.

<sup>e</sup> Worth GK and Retallack EW (1988) Tritium isotope effect in high-pressure liquid chromatography. *Analytical Biochemistry* 174: 137–141.

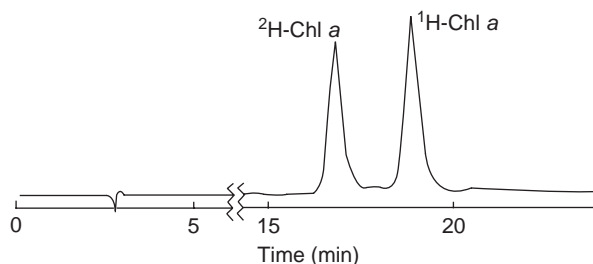
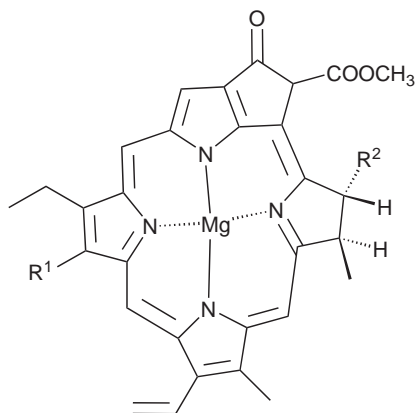
<sup>f</sup> Econosphere silica column.

<sup>g</sup> Radial-pak column.

possessing antibiotic and antifungal properties, leads to the corresponding tetrahydroderivatives [II]. The isotope effect in reversed-phase LC mobility was surprisingly large:  $\alpha = k_1/k_2 = 1.0190$  for the deuterated compound and  $\alpha = k_1/k_3 = 1.0233$  for the tritiated compound. A Partisil 5 C<sub>18</sub> column and a mixture of 0.1% phosphoric acid (45%) with CH<sub>3</sub>CN/THF/H<sub>3</sub>PO<sub>4</sub> (90:20:0.1 v/v/v) (55%) as eluent were used. Even though the chromatographic conditions described do not allow resolution of labeled and unlabeled species, such a possibility exists, e.g., when column recycling system is used.



Reversed-phase LC of deuterated chlorophylls [III] obtained from green algae *Rhodospirillum rubrum* and *Rhodospirillum spheroides* grown in media containing 50%, 80%, 90%, and 99.7% <sup>2</sup>H<sub>2</sub>O exhibited baseline separation from isotopically unmodified chlorophyll. As usual, the labeled species were eluted first, and the greater the percentage of deuterium in the compound the faster it moved along the column (Figure 2, Table 5).



**Figure 2** Reversed-phase LC separation of deuterated chlorophyll *a* and undeuterated chlorophyll *a* on a 25 cm × 4.6 mm i.d. C<sub>18</sub>-Ultrasphere ODS column. Mobile phase: water-methanol-acetonitrile-tetrahydrofuran 5:28:38:23 (v/v/v/v). UV detector 663 nm. (Reprinted with permission from Baweja R, *et al.* (1986) HPLC separation of deuterated photosynthetic pigments from their protio analogues. *Journal of Chromatography* 369: 125–131; © Elsevier.)

**Table 5** LC separation factors of chlorophylls versus deuterated chlorophylls<sup>a</sup>

Labeled compound	$\alpha = k_1/k_2$	Eluent (A/B) <sup>b</sup> (v/v)
[ <sup>2</sup> H]Chlorophyll <i>a</i>	1.132	5:95
[ <sup>2</sup> H]Chlorophyll <i>a'</i>	1.158	5:95
[ <sup>2</sup> H]Chlorophyll <i>b</i>	1.156	5:95
[ <sup>2</sup> H]Bacteriochlorophyll <i>a</i> <sup>c</sup>	1.149	10:90
[ <sup>2</sup> H]Bacteriochlorophyll <i>a'</i> <sup>d</sup>	1.118	7:93
[ <sup>2</sup> H]Pyrochlorophyll <i>a</i>	1.167	0:100

<sup>a</sup>Baweja R, *et al.* (1986) HPLC separation of deuterated photosynthetic pigments from their protio analogues. *Journal of Chromatography* 369: 125–131.

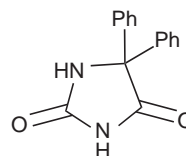
<sup>b</sup>A = H<sub>2</sub>O; B = methanol-acetonitrile-THF (30:40.5:24.5, by vol.).

<sup>c</sup>Contains geranylgeranyl side chain.

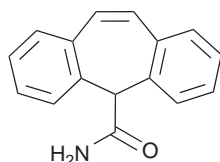
<sup>d</sup>Contains phytyl side chain.

## Drugs

(<sup>2</sup>H<sub>10</sub>)Diphenylhydantoin (phenytoin [IV]) and (<sup>2</sup>H<sub>10</sub>)-5H-dibenz[*b,f*]azepine-5-carboxamide (carbamazepine [V]) were separated from their unlabeled parent compounds by reversed-phase chromatography on a C<sub>18</sub> column with H<sub>2</sub>O/CH<sub>3</sub>CN/THF (80:16:4, v/v/v) as the eluent. Baseline separation was attained even when using extracted serum samples. The calculated resolution for both pairs of isotopomers was 1.3 (resolution  $R_s = (t_2 - t_1)/[0.5(w_1 + w_2)]$ ).

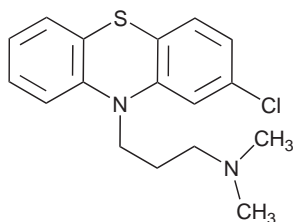


[IV]



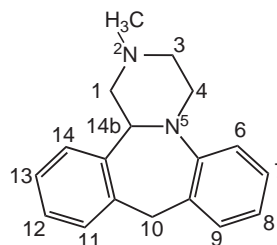
[V]

[CH<sub>3</sub>-<sup>3</sup>H]Chlorpromazine [VI] and its metabolite [CH<sub>3</sub>-<sup>3</sup>H]-7-hydroxychlorpromazine were separated on a chemically bonded stationary phase (Spherisorb CN) from their unlabeled counterparts. Although the specific activity of tritiated compound was  $\sim 40 \text{ Ci mmol}^{-1}$  indicating  $\sim 1.5$  tritium atom per methyl group, the measured separation factors were rather high (Table 6). The separation factors depend strongly on the pH of mobile phase, so that the isotope effect on the basicity of the dimethylamino group is probably significant. The labeled species was eluted later.

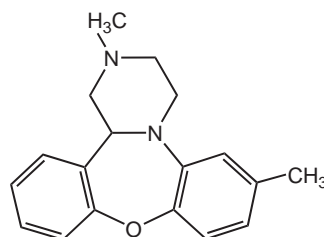


[VI]

An almost complete resolution was attained between unlabeled and di-, tri-, or tetradeuterated drugs active on the central nervous system – 1,2,3,4,10,14b-hexahydro-2-methyl-dibenzo[*c,f*]pyrazino[1,2-*a*]azepine (mianserin [VII]) and 1,3,4,10,14b-hexahydro-2,7-dimethyl-2H-dibenzo[*c,f*]pyrazino[1,2-*a*]oxazepine (Org GC 94 [VIII]). [3,3,4,4-<sup>2</sup>H<sub>4</sub>]Org GC 94 was chromatographed on  $\mu$ -Porasil column and *n*-hexane-2-propanol (90:10, v/v) to which a solution of 4% ethanol and 0.1%

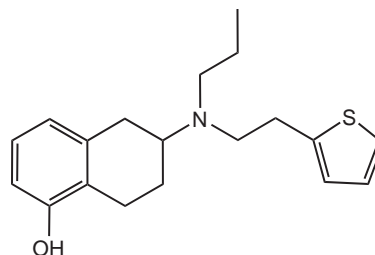


[VII]



[VIII]

2-(*N*-Propyl-*N*-2-thienylethyl)-5-hydroxytetralinamine (N-0437 [IX]), an effective drug against Parkinson's disease and glaucoma, can be separated from its counterparts deuterated and tritiated at the propyl group. The corresponding diastereomeric glucuronides were prepared by enzymatic synthesis and separated in a reversed-phase LC system (Nova-Pak C<sub>18</sub>). The separation factors were pH-dependent and relatively large, considering the remote positions of the labels from the nitrogen atom, e.g.,  $\alpha_2 = 1.033$ ;  $\alpha_3 = 1.037$ .



[IX]

**Table 6** LC separation factors of chlorpromazine versus tritiated chlorpromazine<sup>a</sup>

Labeled compound	$\alpha = k_1/k_2^b$	pH of eluent <sup>c</sup>
[CH <sub>3</sub> - <sup>3</sup> H]Chlorpromazine	0.855	7
	0.901	6
	1.00	5
[CH <sub>3</sub> - <sup>3</sup> H]-7-Hydroxychlorpromazine	0.847	7
	0.910	6
	0.952	5

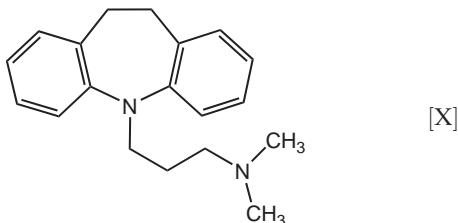
<sup>a</sup>Yeung PKF, Hubbard JW, Baker BW, Looker MR, and Midha KK (1984) Isotopic fractionation of *N*-([<sup>3</sup>H]methyl)chlorpromazine and *N*-([<sup>3</sup>H]methyl)-7-hydroxychlorpromazine by reversed-phase high performance liquid chromatography. *Journal of Chromatography* 303: 412–416.

<sup>b</sup>Spherisorb CN column.

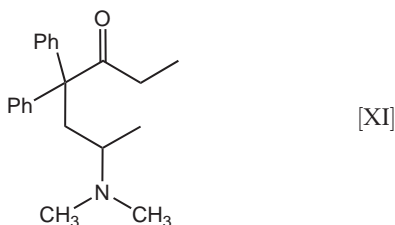
<sup>c</sup>10% 0.05 mol l<sup>-1</sup> sodium acetate buffer in methanol.

The above-mentioned effect of isotope substitution on the basicity of amines was demonstrated by thin-layer chromatography (TLC) of the popular antidepressant imipramine [X]. [<sup>2</sup>H<sub>10</sub>]Imipramine labeled in both methyl groups and in aromatic rings,

and its parent compound were resolved as individual spots on silica plates using different eluents, in which the presence of ammonia was necessary; e.g., benzene/acetone/ $\text{NH}_4\text{OH}$ , 300:60:1;  $R_F$  [ $^1\text{H}$ ] = 0.24;  $R_F$  [ $^2\text{H}$ ] = 0.18. Labeling on the benzene ring exerted no effect on the chromatographic mobility.



Similar TLC behavior has been observed with (R,S)-6-[ $\text{CH}_3\text{-}^2\text{H}$ ]dimethylamino-4,4-diphenylheptane-3-one  $\cdot \text{HCl}$  ( $[\text{H}]$ methadone [XI]) (silica gel Merck; benzene/methanol/ $\text{NH}_4\text{OH}$ ; 85:15:0.65, v/v/v);  $R_F$  [ $^2\text{H}$ ] = 0.58;  $R_F$  [ $^1\text{H}$ ] = 0.79. Reversed-phase LC showed baseline resolution, whereas gas-liquid chromatography afforded no separation.



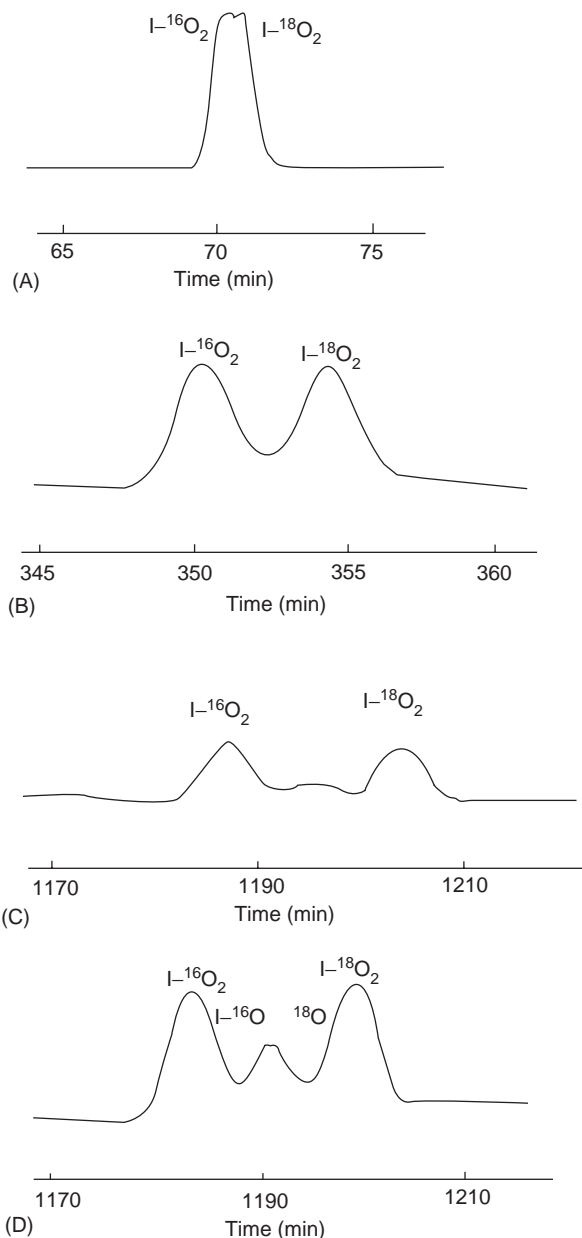
### Isotopes of Heavier Elements

Liquid chromatographic separations of the isotopes of elements other than hydrogen have been rather rare. A high-efficiency liquid-liquid chromatographic system consisting of porous silica microspheres covered with 25% (w/w) bis(2-ethylhexyl)phosphoric acid in dodecane as the stationary phase and nitric acid as the mobile phase provided a certain enrichment of heavier isotopes of calcium in front of the elution curve. Separation factors calculated by Glueckauf for  $^{42}\text{Ca}$ ,  $^{45}\text{Ca}$ ,  $^{48}\text{Ca}$  versus  $^{40}\text{Ca}$  were 1.0012–1.0029.

Better resolution was obtained when isotopic pairs of  $^{144}\text{Sm}$ – $^{154}\text{Sm}$ ,  $^{140}\text{Ce}$ – $^{142}\text{Ce}$ , and  $^{151}\text{Eu}$ – $^{153}\text{Eu}$  were chromatographed on LiChrosorb 60 with di-2-propyl-ether/THF/65%  $\text{HNO}_3$  (100:20:5, v/v/v). Heavier isotopes were always eluted first, the separation factors being two orders of magnitude higher than in ion-exchange chromatography (1.030–1.085).

There are only a few references in literature about the liquid chromatographic separations of organic compounds labeled with isotopes heavier than hydrogen. For example, Tanaka *et al.* demonstrated the

oxygen isotope effect in the reversed-phase LC separation of  $^{18}\text{O}$  labeled benzoic acid (Figure 3). They demonstrated that successful separation was due to differences in the dissociation constants between the labeled and parent compounds. The isotope effect was pH-dependent and attained a maximal value around pH 5 (Table 7). An equation was derived to



**Figure 3** Separation of benzoic acid (I) isotopomers by recycle chromatography. Column: Cosmosil 5- $\text{C}_{18}$ -P  $\times$  4.15 cm  $\times$  4.6 mm i.d. Mobile phase: methanol–0.05 mol  $\text{l}^{-1}$  acetate buffer (pH 4.83) 20:80 (v/v). Cycles: (A) 1, (B) 5, (C) and (D) 17. (Reprinted with permission from Tanaka N, Araki M, and Kimata K (1986) Separation of oxygen isotopic compounds by reversed-phase liquid chromatography. *Journal of Chromatography* 352: 307–314; © Elsevier.)



**Table 7** Dependence of eluent pH on LC<sup>a</sup> separation factors of benzoic acid versus [<sup>18</sup>O<sub>2</sub>] benzoic acid<sup>b</sup>

Eluent pH	2.44	3.13	4.07	4.40	4.90	5.29	5.86	6.27
$\alpha$	1.000	1.000	0.995	0.991	0.988	0.988	0.992	0.995

<sup>a</sup>Cosmosil C<sub>18</sub>-P; methanol–0.05 mol l<sup>-1</sup> acetate buffer (20:80); 30°C.

<sup>b</sup>Tanaka N, Araki M, and Kimata K (1986) Separation of oxygen isotopic compounds by reversed-phase liquid chromatography. *Journal of Chromatography* 352: 307–314.

determine the <sup>18</sup>O isotope effect on the dissociation constants from LC data. This isotope effect was calculated to be  $^{16}K_a/^{18}K_a = 1.020 \pm 0.002$ . A very good, nearly baseline, resolution was attained in a recycle system for the pair benzoic acid–[<sup>18</sup>O<sub>2</sub>]benzoic acid after five cycles, and for the mixture benzoic acid–[<sup>18</sup>O]benzoic acid–[<sup>18</sup>O<sub>2</sub>]benzoic acid after 17 cycles. Similar results were obtained in the chromatography of [<sup>18</sup>O<sub>2</sub>]-4-chlorobenzoic acid and [OH-<sup>18</sup>O]-4-nitrophenol with their unlabeled counterparts.

The same procedure as above was successfully applied to the separation of aniline and [<sup>15</sup>N]aniline on the same column with 0.05 mol l<sup>-1</sup> acetate buffer containing 5% methanol and 0.01% triethylamine as eluent. At pH 4.24, the separation factor was found to be  $\alpha = k_{14}/k_{15} = 1.010$ . The complete separation was accomplished after 20 cycles.

### Ion-Exchange Chromatography

D-Glucose and (1-<sup>2</sup>H)-D-glucose were separated by anion-exchange chromatography on the CarboPac PA1 column. Elution was accomplished with water and required 18–22 min. The baseline separation was achieved using two columns.

There exist large amounts of published experimental works dealing with enrichment of the metal isotopes (from lithium to uranium) by chromatography on cation-exchange resins. Here, we will present some optimistic and interesting examples.

Extensive work has been devoted to the lithium isotopes. The estimated separation factors varied between 0.992 and 1.065. The highest values of the separation factors were achieved by chromatography on the crown ethers bound on the Merrifield resins.

The isotopes of magnesium were chromatographed on the 1-aza-12-crown-4 and bound on the Merrifield peptide resin. The separation factors were  $\alpha(24/25) = 1.015$ ;  $\alpha(24/26) = 1.029$ ;  $\alpha(25/26) = 1.014$ . Similar values were achieved by chromatography on hydrous manganese(IV) oxide:  $\alpha(24/25) = 1.011$ ;  $\alpha(24/26) = 1.021$ ;  $\alpha(25/26) = 1.011$ .

LC of a column packed with the cryptand polymer was used for the separation of zinc isotopes. The enrichment factors were  $\alpha(67/66) = 0.99967$ ;

$\alpha(68/66) = 1.00018$ ;  $\alpha(70/68) = 1.00072$ . Different contributions, i.e., vibrational energy shift, the nuclear mass shift, and field shift caused by the nuclear size and shape of the isotope were also estimated.

An unusual isotope effect of molybdenum in chemical exchange reaction using dicyclohexano-18-crown-6 (batch method) was observed:  $\alpha(94/96) = 1.0086 \pm 0.0007$ . This extremely high value was not ascribed to field shift effect but due to an anomaly of the vibrational levels.

Ion-exchange chromatography of europium has been carried out in a ligand exchange system (LXS) and an electron exchange system (EXS) on highly acidic cation-exchanger AG-MP 50. The observed separation factor for LXS was  $\alpha(153/151) = 1.000015$  at 80°C and  $\alpha(153/151) = 1.000213$  at 60°C for EXS.

**See also:** Gas Chromatography: Overview. Ion Exchange: Ion Chromatography Applications. Isotope Ratio Measurements. Liquid Chromatography: Reversed Phase. Mass Spectrometry: Mass Separation; Stable Isotope Ratio.

### Further Reading

- Filler CN (1999) Isotopic fractionation of organic compounds in chromatography. *Journal of Labelled Compounds and Radiopharmaceuticals* 42: 169–197.
- Glueckauf E (1961) Isotope separation by chromatographic methods. In: London H (ed.) *Separation of Isotopes*, pp. 209–248. London: G. Newnes Ltd.
- Kakihana H and Oi T (1989) Theory of isotope separation by displacement chromatography. *Journal of Chromatography* 483: 179–188.
- Kaspersen FM, van Acquoy J, van de Laar GLM, Wagenaar GN, and Funke CW (1984) Deuterium isotope effects in mianserin. *Recueil Travaux Chimique de Pay-Bas* 103: 32–36.
- Kim DW (2002) Chromatographic separation of magnesium isotopes by monoazacrown bonded Merrifield peptide resins. *Journal of Radioanalytical and Nuclear Chemistry* 252: 129–132.
- Kimata K, Hosoya K, Araki T, and Tanaka N (1997) Direct chromatographic separation of racemates on the basis of isotopic chirality. *Analytical Chemistry* 69: 2610–2612.
- Klein PD (1966) Occurrence and significance of isotope fractionation during analytical separation of large molecules. *Advances in Chromatography* 3: 3–65.

- Lešetický L (1985) Isotope separation by chromatographic methods [in Czech]. *Radioisotopy* 26: 113–126.
- Lockley WJS (1989) Regiochemical differences in the isotopic fractionation of deuterated benzoic acid isotopomers by reversed phase HPLC. *Journal of Chromatography* 483: 413–418.
- Simon H and Palm (1966) Isotope effects in organic chemistry and biochemistry [in German]. *Angewandte Chemie* 78: 993–1007.

- Tanaka N and Thornton (1977) Structural and isotopic effects in hydrophobic binding measured by HPLC. A stable and highly precise model for hydrophobic interaction in biomembranes. *Journal of the American Chemical Society* 99: 7300–7306.
- Wade D (1999) Deuterium isotope effects on noncovalent interaction between molecules. *Chemico-Biological Interactions* 117: 191–217.

## LOW ENERGY ELECTRON DIFFRACTION

See **SURFACE ANALYSIS: Low Energy Electron Diffraction**

## LUMINESCENCE

Contents

**Overview**

**Solid Phase**

### Overview

**N W Barnett and P S Francis**, Deakin University, Geelong, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The observation and investigation of luminescent phenomena has a long and delightful history and some of the more important milestones are briefly described here. Naturally occurring luminescence, such as fireflies, St Elmo's fire, and shining flesh has fascinated humans since the dawn of time, with an early record appearing in Chinese literature between 1500 and 1000 BC. A millennium or so later, Aristotle (384–322 BC) astutely noted that these emanations were produced without heat. Subsequently, Caius Plinius Secundus (Pliny the Elder, AD 23–79) described, in detail, a number of luminous organisms. Notwithstanding the revolutionary nature of these observations of bioluminescence, made by two such great philosophers, a rigorous scientific approach to the subject was not taken until the mid-sixteenth century. This study culminated in the publication, in 1555, of a book by Conrad Gesner (1516–1565) concerned solely with luminescence. However, Sir

Robert Boyle (1627–1691), son of the Earl of Cork, who came to be known as the father of analytical chemistry, first categorized the essential differences between incandescence and luminescence in 1668. The following year saw the first artificial luminescence, which accompanied Hennig Brandt's discovery and isolation of elemental phosphorus. More than three centuries passed before the emanations from white phosphorus were correctly characterized as chemiluminescence. Cambridge professor and president of the Royal Society, Sir George Gabriel Stokes (1819–1903), was the first to characterize the bichromatic nature of crystal fluospar as a true emission (actually phosphorescence) in 1845. He also coined the term fluorescence as an analogy to opalescence and noted that in the production of fluorescence the absorption of shorter wavelengths resulted in emission at longer wavelengths (Stokes' law). In 1877, Raziszewski was probably the first to observe chemiluminescence from synthetically produced organic compounds during the preparation of lophine (2,4,5-triphenylimidazole) from hydrobenzamide.

The earliest usage of the word luminescence is credited to Eilhardt Wiedemann (1852–1928), who in 1888 used the term to describe the emission of light that was not the result of a rise in temperature ('cold light'). He defined six classes of luminescence by the source of energy that stimulated the emission.

- Lešetický L (1985) Isotope separation by chromatographic methods [in Czech]. *Radioisotopy* 26: 113–126.
- Lockley WJS (1989) Regiochemical differences in the isotopic fractionation of deuterated benzoic acid isotopomers by reversed phase HPLC. *Journal of Chromatography* 483: 413–418.
- Simon H and Palm (1966) Isotope effects in organic chemistry and biochemistry [in German]. *Angewandte Chemie* 78: 993–1007.

- Tanaka N and Thornton (1977) Structural and isotopic effects in hydrophobic binding measured by HPLC. A stable and highly precise model for hydrophobic interaction in biomembranes. *Journal of the American Chemical Society* 99: 7300–7306.
- Wade D (1999) Deuterium isotope effects on noncovalent interaction between molecules. *Chemico-Biological Interactions* 117: 191–217.

## LOW ENERGY ELECTRON DIFFRACTION

See **SURFACE ANALYSIS: Low Energy Electron Diffraction**

## LUMINESCENCE

Contents

**Overview**

**Solid Phase**

### Overview

**N W Barnett and P S Francis**, Deakin University, Geelong, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The observation and investigation of luminescent phenomena has a long and delightful history and some of the more important milestones are briefly described here. Naturally occurring luminescence, such as fireflies, St Elmo's fire, and shining flesh has fascinated humans since the dawn of time, with an early record appearing in Chinese literature between 1500 and 1000 BC. A millennium or so later, Aristotle (384–322 BC) astutely noted that these emanations were produced without heat. Subsequently, Caius Plinius Secundus (Pliny the Elder, AD 23–79) described, in detail, a number of luminous organisms. Notwithstanding the revolutionary nature of these observations of bioluminescence, made by two such great philosophers, a rigorous scientific approach to the subject was not taken until the mid-sixteenth century. This study culminated in the publication, in 1555, of a book by Conrad Gesner (1516–1565) concerned solely with luminescence. However, Sir

Robert Boyle (1627–1691), son of the Earl of Cork, who came to be known as the father of analytical chemistry, first categorized the essential differences between incandescence and luminescence in 1668. The following year saw the first artificial luminescence, which accompanied Hennig Brandt's discovery and isolation of elemental phosphorus. More than three centuries passed before the emanations from white phosphorus were correctly characterized as chemiluminescence. Cambridge professor and president of the Royal Society, Sir George Gabriel Stokes (1819–1903), was the first to characterize the bichromatic nature of crystal fluospar as a true emission (actually phosphorescence) in 1845. He also coined the term fluorescence as an analogy to opalescence and noted that in the production of fluorescence the absorption of shorter wavelengths resulted in emission at longer wavelengths (Stokes' law). In 1877, Raziszewski was probably the first to observe chemiluminescence from synthetically produced organic compounds during the preparation of lophine (2,4,5-triphenylimidazole) from hydrobenzamide.

The earliest usage of the word luminescence is credited to Eilhardt Wiedemann (1852–1928), who in 1888 used the term to describe the emission of light that was not the result of a rise in temperature ('cold light'). He defined six classes of luminescence by the source of energy that stimulated the emission.

The present number of classes of luminescent phenomena is somewhere around 20; however, the boundaries between some of these classifications is more than a little blurred. In stark contrast to incandescence, luminescent phenomena are not only cold but are also of relatively low intensity, for example:

- the mysterious St Elmo's Fire and aurora australis (electro- or radioluminescence);
- emission from rubbing or shattering crystals (triboluminescence);
- the ephemeral blue haze from a glass of gin and tonic in the sunlight (fluorescence);
- the enchanting emanations from 'lightening bugs' (fireflies) on a summer evening (chemiluminescence).

Luminescence is generally less intense than incandescence, but it often emanates from extremely small amounts of matter, which has beneficial implications for analytical science. Nevertheless, the utilization of luminescence for analysis is quite a recent innovation. The following commentary describes the fundamental spectroscopic and chemical principles underlying luminescence in relation to its application in analytical science. As other articles will deal with atomic spectroscopy, this discussion will be restricted to analytical molecular luminescence spectroscopy including fluorescence, phosphorescence, and chemiluminescence (bioluminescence being a special case of chemiluminescence).

By monitoring the relative intensity of luminescence as a function of concentration it is possible to quantitatively determine (often at trace levels) a range of inorganic and organic analytes. Contemporary analytical texts reveal that fluorimetric methodology is far more commonplace than either phosphorescence or chemiluminescence. The research literature, however, indicates an increased acceptance of the latter two spectroscopies particularly when combined with either organized media or flow analysis, respectively. Luminescence methodology generally offers superior selectivity, detectability, and linear calibration range compared to that attainable with absorption spectrometry. Unfortunately, there are a relatively limited number of molecules that will exhibit luminescence thus restricting its applicability compared to absorption techniques. Although all forms of luminescence have a common quantum mechanical basis for emission, the route to the excited state defines their respective class. Given that the theory of photoluminescence is better understood than that of the various related phenomena, a basic synopsis of fluorescence and phosphorescence

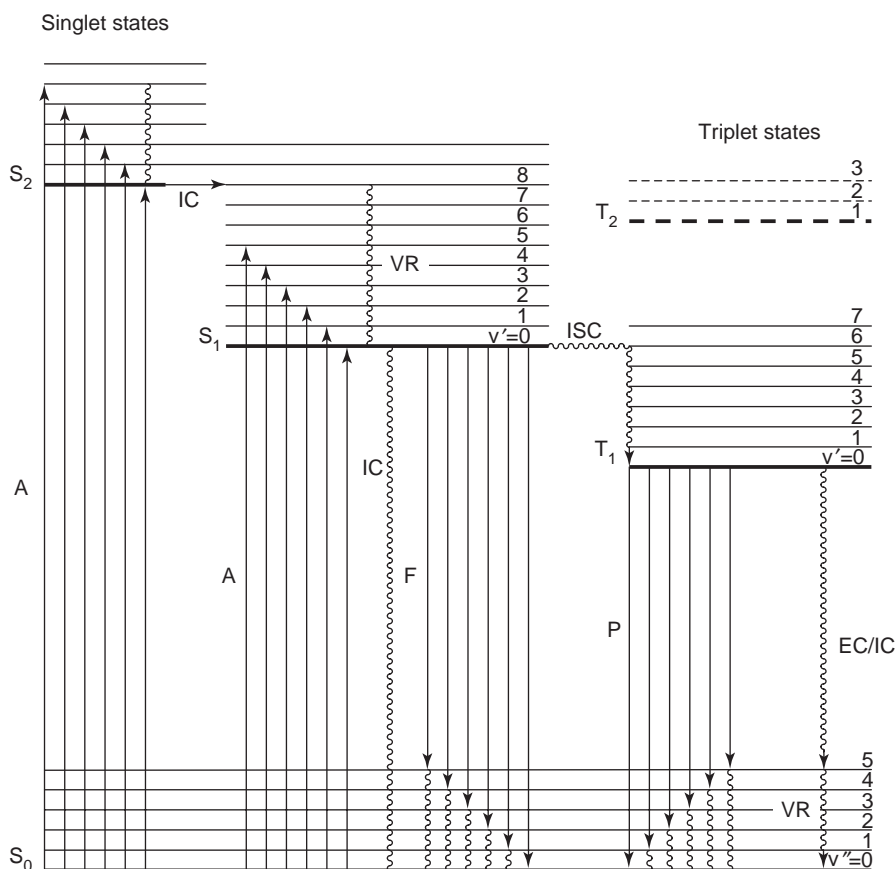
is presented prior to an introduction to chemically induced luminescence.

## Photoluminescence

### General Principles

Photoluminescence occurs as the result of electronic excitation within a molecule brought about by the absorption of a photon. The question therefore arises, as the vast majority of molecules do not luminescence, what are the quantum mechanisms that give rise to fluorescence and phosphorescence and how does molecular structure aid or inhibit these processes? **Figure 1** is a Jabłoński diagram, named in honor of the Ukrainian born physicist, Professor Alexander Jabłoński (1898–1980), who is considered by many to be the father of fluorescence spectroscopy. These schematic energy level diagrams are useful tools for the understanding of molecular electronic excitation and deexcitation leading to photon emission. The first observation to be made from **Figure 1** is that fluorescence emission (F) results from an electron falling from the lowest vibrational level of an electronically excited state ( $S_1$ ) to any of the vibrational excited levels of the ground state ( $S_0$ ). Phosphorescence emission (P) is a similar process except that the upper electronic state is a triplet ( $T_1$ ). The terms singlet and triplet refer to the relative spins of the electrons in the ground and excited states. When the spins are paired (antiparallel) the upper level is termed a singlet, and when the spins are unpaired (parallel) a triplet state exists. The nomenclature arises from the observed multiplicity in spectra measured under the influence of a magnetic field. It is noteworthy that  $T_1$  states are less energetic than  $S_1$ ; this is a direct consequence of electrons with parallel spins being further apart thus exhibiting less mutual repulsion (spin correlation, Hund's rule). Hence, for a given compound the phosphorescence emission will occur at longer wavelengths than the fluorescence. The photo-induced promotion of an electron from  $S_0$  directly to  $T_1$  has not been shown in **Figure 1** as the simultaneous change of molecular orbital and electronic spin has a very low probability of occurrence. In fact, such transitions are often referred to as 'forbidden' by the spin selection rule.

As most compounds are not luminescent, this implies that there must be more efficient alternative deexcitation mechanisms available to molecules in  $S_1$  or  $T_1$  states to return to  $S_0$  other than ejection of a photon. To understand the nature of the spectroscopic processes shown in **Figure 1**, their relative rates of occurrence and the influences on the various pathways must be considered.



**Figure 1** A Jablonski diagram for a hypothetical luminescent molecule, where A is absorption, F is fluorescence, P is phosphorescence, VR is vibrational relaxation, IC is internal conversion, EC is external conversion, ISC is intersystem crossing,  $v''$  and  $v'$  are vibrational levels associated with each electronic state,  $S_0$  is the electronic ground state,  $S_1$  and  $S_2$  are excited singlet states and  $T_1$  and  $T_2$  are excited triplet states.

At room temperature in solution, we may assume that all molecules will be in the lowest vibrational level ( $v'' = 0$ ) of the ground state ( $S_0$ ), thus absorption will originate solely from  $v'' = 0$ . According to the Franck–Condon principle, the initial step (absorption of a photon) is extremely rapid, requiring only  $10^{-15}$ – $10^{-14}$  s. Immediately after absorption; either of the excited singlets ( $S_1$  or  $S_2$ ) will undergo vibrational relaxation to their lowest vibrational level ( $v' = 0$ ). This process is extremely efficient with all the excess vibrational energy being transferred to the solvent molecules as heat in around  $10^{-13}$ – $10^{-11}$  s. In the case of excitation to  $S_2$ , the process of internal conversion to the upper vibrational levels of  $S_1$  (IC in **Figure 1**) is rapid and effective. Thus luminescence in solution will usually result from the lowest vibrational level of the  $S_1$  excited state. This is not always the case in gas-phase luminescence where collisional relaxation is far less likely. The fluorescence emission lifetime ( $\tau$ ) is of similar magnitude to the mean lifetime of the excited singlet state ( $10^{-10}$ – $10^{-7}$  s), the latter being inversely proportional to the molar absorptivity.

As the probability of finding an excited state molecule at time  $t$  after removal of the excitation source is  $e^{-t/\tau}$ , then the relationship between luminescence intensity and the mean excited state lifetime is:

$$I = I_0 e^{-t/\tau} \quad [1]$$

where  $I$  is the luminescence intensity at time  $t$  and  $I_0$  is the maximum luminescence intensity during excitation. In practical terms the difference in  $\tau$  values for various analytes can provide extra selectivity when using time-resolved luminescence spectroscopy. **Figure 1** also illustrates that the fluorescence emission spectrum is shifted to longer wavelengths compared to that of the absorption (excitation) spectrum. This is the result of vibrational relaxation both after excitation and after fluorescence and is known as Stokes' shift. From the zeroth vibrational level of the lowest excited singlet ( $v' = 0$ ,  $S_1$ ) it is also possible for radiationless deexcitation to convert all the energy to heat in preference to light. This is termed internal conversion, the quantum mechanical basis for which is somewhat vague. The internal conversion from  $S_1$



to  $S_0$  for aliphatic compounds can be rationalized by the overlapping of the upper vibrational levels of  $S_0$  with those of  $S_1$ . In this case relaxation is rapid and efficient and explains why aliphatic molecules rarely luminesce. Internal conversion also occurs between excited electronic states (as shown in **Figure 1**) due to the overlapping vibrational levels of  $S_1$  and  $S_2$ . At the crossover point the potential energies of the two excited states are equal and as the efficiency of vibrational relaxation is far greater than emission of a photon, internal conversion to the zeroth vibrational level of  $S_1$  occurs rather than fluorescence from  $S_2$ . Therefore, fluorescence from anything but the  $S_1$  state is rare; the blue hydrocarbon azulene (isomeric with naphthalene) is the most well-known exception. In large molecules internal conversion may also result in predissociation (bond cleavage), when the electron moves from an excited state to a high-vibrational level of a lower electronic state. In this situation the vibrational energy is sufficient to cause cleavage of bonds in the unstable excited state. Together with internal conversions, electronically excited states can be deactivated by virtue of their interactions with solvent molecules of other concomitant species present. Such pathways are termed external conversions. As a general rule, those environmental parameters that lower the probability of collisional deexcitation (such as lower temperature, increased viscosity, and organized media) tend also to enhance luminescence.

Whilst the probability of the direct population of  $T_1$  from  $S_0$  by absorption is virtually zero, a kinetically efficient pathway exists from the  $S_1$  state in a number of molecules. The mechanism is known as intersystem crossing and it can be considered as a spin-dependent internal conversion. It should be borne in mind that singlet-triplet transitions are approximately a million times less likely to occur than singlet-singlet or triplet-triplet processes and also that radiationless vibrational deexcitation occurs in around  $10^{-13}$  s. As the mechanism of intersystem crossing relies upon vibrational coupling between  $S_1$  and  $T_1$ , the time required for this spin forbidden process can be estimated to be around  $10^{-8}$ – $10^{-7}$  s. This is of the same order as the lifetime of the radiative transition and therefore intersystem crossing competes with fluorescence. The probability of intersystem crossing is enhanced when the energy difference between  $S_1$  and  $T_1$  is small and when the lifetime of  $S_1$  is relatively long. Intersystem crossing is also aided by the presence of heavy atoms (e.g., iodine and bromine) as substituents on either the solute or the solvent molecules. This so-called heavy-atom effect arises from increased spin/orbital interactions and as such spin reversal becomes

more probable. After intersystem crossing, the molecule rapidly undergoes internal conversion ( $10^{-13}$ – $10^{-11}$  s) to the lowest vibrational level of  $T_1$ .

In a similar manner to fluorescence, phosphorescence can only occur with a radiative deexcitation from  $T_1$  to  $S_0$  and two factors limit the likelihood of this event. Firstly, because the energy difference between  $S_0$  and  $T_1$  is smaller than that between  $S_0$  and  $S_1$ , the vibrational coupling between  $S_0$  and  $T_1$  may be enhanced resulting in more efficient internal conversion. Of more consequence, however, is the relatively long intrinsic lifetime of the triplet excited-state ( $10^{-4}$ – $10^2$  s), which provides more than ample opportunity for collisional deactivation. In fact, this mechanism is predominant and explains why phosphorescence is rarely observed at room temperature in simple solutions (2,3-butanedione and tris(2,2'-bipyridyl)ruthenium(II) are examples). The relatively long lifetime of  $T_1$  arises from the spin forbidden nature of the triplet-singlet state transition; as a consequence phosphorescence is often characterized by an afterglow. This is not seen in fluorescence. Clearly, radiationless processes are much more likely to deexcite a triplet state compared to the ejection of a photon, therefore, phosphorescence is most commonly observed from molecules that are either at very low temperatures (often 77 K), in organized media (micelles), or adsorbed onto solid surfaces.

### Structural and Environmental Influences on Photoluminescence

From the preceding section, it is evident that the molecule's structure and its chemical environment determine whether or not it will luminesce, and to what extent. In order to discuss these variables it is useful to introduce quantum yield ( $\phi$ ) or quantum efficiency. For a luminescent process,  $\phi$  represents the ratio of the number of molecules that emit to the total number excited. For highly luminescent substances  $\phi$  will approach unity and for species that do not luminesce appreciably,  $\phi$  will tend toward zero. Thus the luminescence quantum yield for a particular molecule in a specified environment will be related to the relative rate constants of those pathways which can deexcite the lowest excited state ( $S_1$  or  $T_1$ ). The fluorescence quantum yield can therefore be expressed as follows:

$$\phi_f = \frac{k_f}{(k_f + k_i + k_r)} \quad [2]$$

in which  $k$  represents the various rate constants as designated by the subscripts, where f is fluorescence, i is intersystem crossing, and r is radiationless energy loss, which includes internal and external

conversions plus predissociation or dissociation. Thus, if  $k_f \gg (k_i + k_r)$  then  $\phi_f \rightarrow 1$  and if  $k_f \ll (k_i + k_r)$  then  $\phi_f \rightarrow 0$ . The phosphorescence quantum yield ( $\phi_p$ ) is dependent upon competition between the radiationless routes from  $T_1$  to  $S_0$  and phosphorescence emission. But  $\phi_p$  also relies upon the rate of intersystem crossing which competes with both fluorescence and radiationless deactivation from  $S_1$  and  $S_0$ . Thus  $\phi_p$  is given by:

$$\phi_p = \frac{k_p}{k_p + k'_r} \times \frac{k_i}{k_f + k_i + k_r} \quad [3]$$

where  $k'_r$  represents the combined rate constant for all radiationless pathways from  $T_1$  to  $S_0$ . In the majority of situations  $k_f$  and  $k_p$  are related to molecular structure with only minor dependence upon environmental variables. The magnitude of  $k_i$  can be affected by both these parameters. Both  $k_f$  and  $k'_r$  have only slight dependence upon molecular structure but are markedly affected by the molecular environment.

Photoluminescence resulting from absorption of wavelengths below 200 nm is not common since the subsequent transitions from interaction with such energetic photons are likely to result in deactivation of the excited states via predissociation or dissociation. It is therefore not surprising that luminescence due to  $\sigma^* \rightarrow \sigma$  transitions are virtually nonexistent, in fact most fluorescent or phosphorescent emissions from organic molecules arise from  $\pi^* \rightarrow \pi$  and sometimes  $\pi^* \rightarrow n$  transitions depending upon which is less energetic. Fluorescence is more commonly observed from compounds where the  $S_1 \rightarrow S_0$  transition corresponds to a  $\pi^* \rightarrow \pi$  emission rather than  $\pi^* \rightarrow n$ , which implies that the quantum efficiency resulting from  $\pi, \pi^*$  states is far greater than that from  $n, \pi^*$  states. This situation can be explained by consideration of some spectroscopic parameters for each state (see Table 1).

Because the lower molar absorptivity of the  $n \rightarrow \pi^*$  transitions translates to lower values of  $k_f$ , the relative importance (efficiency) of intersystem crossing is in turn enhanced. The smaller energy differences between  $S_1$  and  $T_1$  for  $n, \pi^*$  states also increases the rate of intersystem crossing. Therefore, if  $S_1$  is an

$n, \pi^*$  state,  $k_i$  in eqn [2] becomes large with respect to  $k_f$  and as such  $\phi_f$  tends toward zero. On the other hand, when  $S_1$  is a  $\pi, \pi^*$  state, lifetimes are much shorter ( $k_i < k_f$ ). There is also less vibrational level overlap between  $S_1$  and  $T_1$  and consequently fluorescence becomes more probable. In general, molecules with low lying  $n, \pi^*$  states do not fluoresce but may phosphoresce given a suitable environment. In such a case emission would result from a  $\pi^*$  to  $n$  transition, but phosphorescence is also observed from  $\pi^*$  to  $\pi$  transitions.

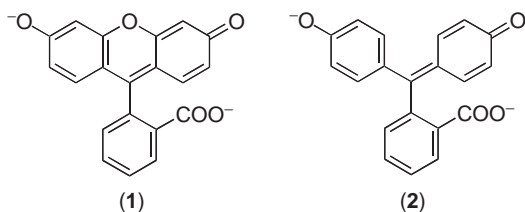
Analytically usable fluorescence is most often observed from compounds with aromatic functionality. Relatively few nonaromatic molecules exhibit fluorescence and those that do often contain carbonyl groups or are highly conjugated. Consistent with the previous discussion, all these species possess low energy  $\pi$  to  $\pi^*$  ( $S_0$  to  $S_1$ ) transitions. The majority of unsubstituted aromatic hydrocarbons exhibit fluorescence in liquid solution and sometimes phosphorescence under specific conditions. Increasing the number of fused rings generally results in emission at longer wavelengths; thus benzene and naphthalene fluoresce in the UV whereas anthracene and naphthacene exhibit blue and green fluorescence, respectively. Substitution on to the aromatic rings causes shifts in the absorption wavelengths, which in turn changes the emission spectra. More importantly, substitution of particular species can drastically affect the quantum efficiency. For example, large (heavy) atoms increase the probability of intersystem crossing to the triplet state and carboxylic acids or carbonyl groups reduce the fluorescence quantum yield since these moieties often possess low level  $n, \pi^*$  transitions. Along with substitution, molecular geometry also affects luminescence. This can be illustrated by considering two structurally similar compounds: fluorescein(1), which is intensely fluorescent; and phenolphthalein(2), which is not fluorescent. The only difference between the two structures is the oxygen bridge present in fluorescein, which imparts rigidity to the molecule. In phenolphthalein electronic excitation can be lost internally via vibration and rotation rather than photon emission. In structurally rigid molecules, energy dissipation via vibration and internal rotation is far less efficient. Enhanced phosphorescence can result from adsorption of the emitting species on to a solid surface to provide added rigidity. Such molecular inflexibility is often used to rationalize the observed increase in fluorescence intensity of organic complexing agents when chelated to a metal cation. Thus, the complexes of  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Zn}^{2+}$ , and  $\text{Al}^{3+}$  with 8-quinolinol-5-sulfonic acid are much more fluorescent than the free ligand molecule itself.

**Table 1** A comparison of  $n, \pi^*$  and  $\pi, \pi^*$  singlet states

	$n, \pi^*$	$\pi, \pi^*$
Molar absorptivity ( $1 \text{ mol}^{-1} \text{ cm}^{-1}$ )	$10\text{--}10^3$	$10^3\text{--}10^5$
Lifetime(s)	$10^{-7}\text{--}10^{-5}$	$10^{-10}\text{--}10^{-7}$
Energy difference between $S_1$ and $T_1$	Small	Often large
Rate constant for intersystem crossing ( $k_i$ )	$> k_f$	$< k_f$

It is worth mentioning that while many metal ions form rigid complexes with 8-quinolinol-5-sulfonic acid, relatively few exhibit analytically useful fluorescence. This is due to the internal heavy atom effect and/or paramagnetism causing intersystem crossing, which is manifested (in some instances) by increased intensity of phosphorescence.

Together with molecular structure, environmental parameters such as temperature, solvent type, viscosity, pH, and dissolved oxygen content can also affect luminescence. As the solution temperature rises, it follows that the number of collisions between the excited state molecule and the solvent molecules will increase, thus greatly improving the likelihood of radiationless deexcitation to the ground state. Therefore,  $\phi_f$  for most compounds decreases with increasing temperature. As mentioned earlier, the effect of temperature upon  $\phi_p$  is even more dramatic due to the vastly greater lifetime of triplet states.



By similar reasoning an increase in viscosity (via organized media) will serve to limit collisional energy transfer and thus enhance luminescence. The heavy atom effect discussed earlier in relation to molecular structure can be used to explain the depression of fluorescence with solvents or solutes containing such species. Carbon tetrabromide and ethyl iodide can be used to limit fluorescence by promoting intersystem crossing with a resultant enhancement of phosphorescence. The absorption of a photon causes a change in electronic distribution and, therefore, molecular geometry. This can often lead to a considerable difference in polarity between the ground and excited states. If the excited state has increased polarity then in polar solvents luminescence will be shifted to longer wavelengths compared to that observed in a nonpolar medium. This results from the increased stabilization of the excited state relative to the ground state. Solvent polarity can also affect the type of luminescent phenomena observed. For example, in cyclohexane, isoquinoline exhibits intense phosphorescence (at 77 K), yet the same compound in either water or ethanol gives only fluorescence. In protic solvents the lone pair on the nitrogen ( $n$ -orbital) is solvated which lowers its energy relative to the  $\pi^*$  orbital; as a consequence the  $\pi, \pi^*$  becomes the lowest lying state and, as discussed earlier, fluorescence is preferred. The reverse is the case in

cyclohexane, with the low lying  $n, \pi^*$  being prone to intersystem crossing and then phosphorescence.

The control of pH in a solution containing a luminescent analyte is also of great importance for sensitive and reliable analysis. For aromatic compounds with acidic or basic substituents, excitation and emission wavelengths of the ionized and free forms are likely to differ. In the case of fluorescence from metal chelates, the pH must be controlled to ensure that the conditional stability constant for the complex is optimal for the particular analytical situation.

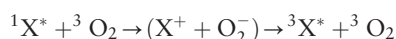
Dissolved molecular oxygen can quench both fluorescence and phosphorescence albeit with different efficiencies and mechanisms. The extent of oxygen quenching on fluorescence is strongly dependent on the fluorophore, whereas for phosphorescence the intensity is always adversely affected. This can be rationalized using the Stern–Volmer equation [4] assuming no other quenching species are present.

$$\frac{\phi_Q}{\phi_0} = \frac{1}{(1 + k_Q \tau [Q])} \quad [4]$$

where  $\phi_Q$  and  $\phi_0$  are the luminescence quantum yields in the presence and absence of the quencher, respectively,  $[Q]$  is the molar concentration of the quenching species and  $k_Q$  is the rate constant for the quenching interaction. Given that quenching is diffusion limited ( $k_Q \approx 10^{10} \text{ mol}^{-1} \text{ s}^{-1}$ ) and that the concentration of oxygen in water at atmospheric pressure is  $\sim 10^{-3} \text{ mol l}^{-1}$  then eqn [4] becomes:

$$\frac{\phi_Q}{\phi_0} = \frac{1}{(1 + 10^7 \tau)} \quad [5]$$

Since phosphorescence lifetimes in the absence of oxygen are in the range  $10^{-4}$ – $10^2 \text{ s}$ , the ratio of the quantum yields (from eqn [5]) will be extremely small for most phosphorescent species. However, as fluorescent lifetimes in the absence of oxygen vary from  $10^{-10}$  to  $10^{-7} \text{ s}$ , the diffusion controlled oxygen quenching of singlet states will be most efficient for the longer-lived species. Numerous mechanisms have been proposed for the quenching of fluorescence by molecular oxygen. The most likely pathway involves enhancement of intersystem crossing by the triplet oxygen, which can be summarized as follows:



The mechanism of intersystem crossing ( $^1X^*$  to  $^3X^*$ ) is thought to occur via a transient charge-transfer species between the excited fluorophore ( $X^*$ ) and molecular oxygen. Although molecular oxygen promotes intersystem crossing, it does not enhance

phosphorescence. As discussed earlier, the long lifetime of the radiative transition precludes phosphorescence being observed in simple solution due to quenching by collisional deactivation. The mechanism for the quenching of triplet states in solution by molecular oxygen (or other triplets) is at best speculative, possibly involving triplet–triplet annihilation. The ability of molecular oxygen to efficiently quench photoluminescence can, however, be exploited for the sensitive determination of oxygen in liquids and solids.

Paramagnetic transition metal ions may also quench solution luminescence by increasing the rate of intersystem crossing. However, there are numerous exceptions to this simple mechanism including manganese(II) ions, which are generally poor quenchers of luminescence but are highly paramagnetic. Diamagnetic main group metal ions are usually inefficient quenching agents. It is therefore clear that other processes may be involved in the interaction of metal ions with luminescent species. Excited state complex formation and energy transfer are both likely in some instances. From a practical analytical point of view, the quenching of luminescence by any concomitant species in real samples must be investigated during method development and either removed or compensated for.

### The Relationship between Photoluminescence Intensity and Analyte Concentration

The relationship between emission intensity and analyte concentration can be derived using the Lambert–Beer law in conjunction with the schematic of a spectroluminometer in Figure 2.

If the incident radiation ( $\lambda_{\text{ex}}$ ) has the power  $P_0$  and a portion of this is absorbed over  $b_1$  then the radiant power striking the central region of the sample ( $P'_0$ ) is given by:

$$P'_0 = P_0 10^{-\epsilon_{\text{ex}} b_1 C} \quad [6]$$

where  $\epsilon_{\text{ex}}$  is the molar absorptivity at  $\lambda_{\text{ex}}$ . It follows that the radiant power of the exciting beam after traversing  $b_2$  ( $P'$ ) is:

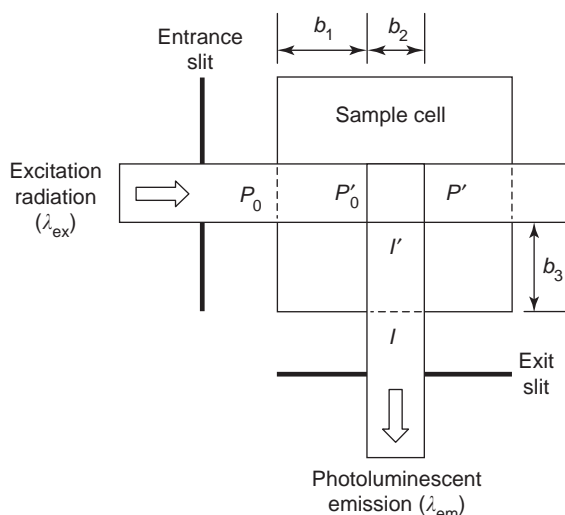
$$P' = P'_0 10^{-\epsilon_{\text{ex}} b_2 C} \quad [7]$$

The luminescence intensity ( $I'$ ) is proportional to both the amount of light absorbed and the quantum efficiency ( $\phi$ ) such that:

$$I' = \phi(P'_0 - P') \quad [8]$$

A fraction of the emitted radiation will be absorbed over the pathlength  $b_3$  and this can be expressed in terms of the observed luminescence intensity ( $I$ ):

$$I = I' 10^{-\epsilon_{\text{em}} b_3 C} \quad [9]$$



**Figure 2** A schematic representation of the sample compartment of a photoluminescent spectrometer.

where  $\epsilon_{\text{em}}$  is the molar absorptivity at  $\lambda_{\text{em}}$ . By rearranging eqns [6]–[9] the following relationship between  $P_0$  and  $I$  can be obtained:

$$I = \phi P_0 10^{-\epsilon_{\text{ex}} b_1 C} (1 - 10^{-\epsilon_{\text{ex}} b_2 C}) 10^{-\epsilon_{\text{em}} b_3 C} \quad [10]$$

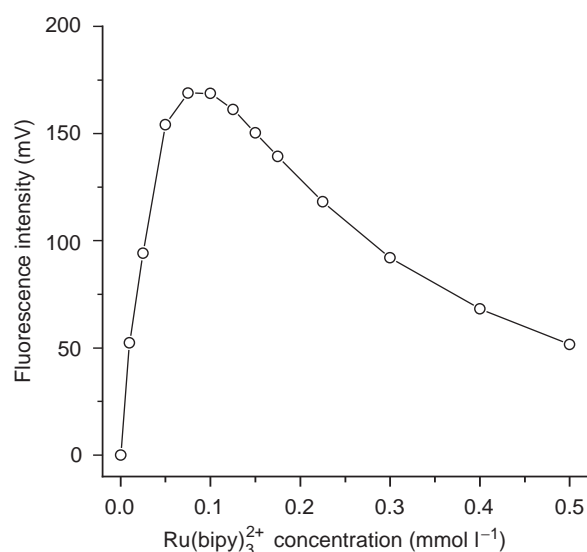
At low analyte concentrations  $I$  is proportional to  $C$  as absorption is small. Since the absorbance increases faster than emission at high analyte concentrations the resultant  $I$  value reaches a maximum and then decreases as seen in Figure 3. This is termed self-absorption and it occurs when  $\lambda_{\text{em}}$  overlaps the absorption band. Self-quenching also decreases the luminescence intensity at high concentrations as it results from a radiationless energy transfer between two excited molecules and the solvent in a similar fashion to external conversion. Under certain conditions self-absorption and self-quenching may result in a maximum in the calibration function. Equation [10] can be simplified at low analyte concentrations; absorbance will be negligible and the three exponents will become very small. Therefore, we can replace the terms  $10^{-\epsilon_{\text{ex}} b_1 C}$  and  $10^{-\epsilon_{\text{em}} b_3 C}$  by unity. Clearly, the term  $10^{-\epsilon_{\text{ex}} b_2 C}$  cannot be treated in this way or eqn [10] would become zero:

$$I = \phi P_0 (1 - 10^{-\epsilon_{\text{ex}} b_2 C}) \quad [11]$$

The term in eqn [11] can subsequently be expanded as a MacLaurin series to give:

$$I = \phi P_0 \left[ \epsilon_{\text{ex}} b_2 C \ln 10 - \frac{(\epsilon_{\text{ex}} b_2 C \ln 10)^2}{2!} + \frac{(\epsilon_{\text{ex}} b_2 C \ln 10)^3}{3!} \dots + \frac{(\epsilon_{\text{ex}} b_2 C \ln 10)^n}{n!} \right] \quad [12]$$

When the solution absorbance is small ( $\epsilon_{\text{ex}} b_2 C < 0.05$ ) it is a good approximation to neglect all



**Figure 3** Photoluminescence calibration for an aqueous solution of tris(2,2'-bipyridyl)ruthenium(II) chloride hexahydrate,  $\lambda_{\text{ex}} = 457 \text{ nm}$ ,  $\lambda_{\text{em}} = 609 \text{ nm}$ .

but the first term in the series:

$$I = \phi P_0 \epsilon_{\text{ex}} b_2 C \ln 10 \quad [13]$$

Therefore, when the absorbance is small (at constant  $P_0$  and under a defined chemical environment) the observed emission intensity is directly proportional to analyte concentration:

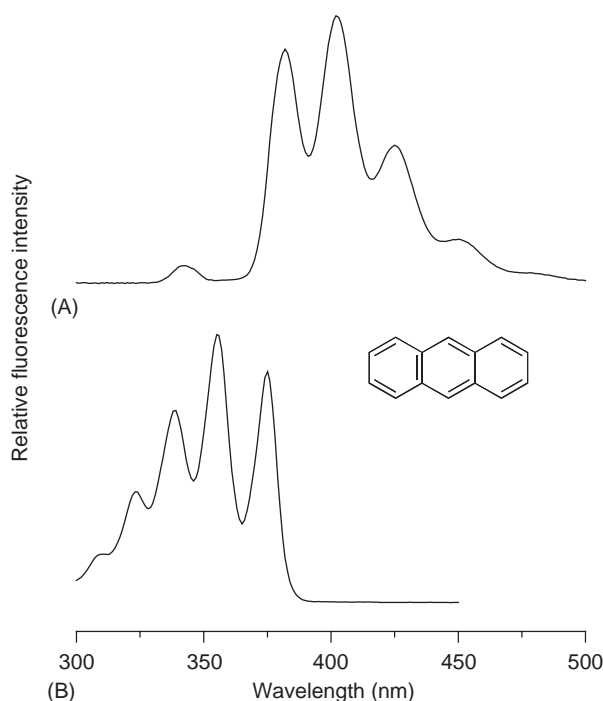
$$I = KC \quad [14]$$

In most analytical applications the above conditions are met; and although the linear dependence of  $I$  upon  $P_0$  is not infinite, it provides a significant advantage for photoluminescence over absorption spectrophotometry.

### Excitation and Emission Spectra

**Figure 4** shows both the emission (A) and excitation (B) spectra of anthracene in ethanol ( $c. 1 \mu\text{g ml}^{-1}$ ). The former was recorded with a fixed excitation wavelength of 340 nm and the latter was monitored at an emission wavelength set to 379 nm. The four peaks in **Figure 4A** are transitions from the zeroth vibrational level of  $S_1$  to the excited vibrational levels of  $S_0$ . Conversely, the four peaks in **Figure 4B** correspond to transitions between the zeroth vibrational level of  $S_0$  and the various vibrational levels of  $S_1$ . The approximate 'mirror image' appearance of the two spectra occurs due to the similarities in the energies of the vibrational levels in  $S_0$  and  $S_1$ .

The difference in wavelength between the excitation and emission spectra is a characteristic of photoluminescent molecules and is known as the Stokes' shift, which can be quantified into wavenumbers by



**Figure 4** Fluorescence spectra of anthracene ( $1 \mu\text{g ml}^{-1}$ ) in ethanol. The emission spectrum (A) was obtained with  $\lambda_{\text{ex}} = 340 \text{ nm}$  and the excitation spectrum (B) was obtained with  $\lambda_{\text{em}} = 379 \text{ nm}$ .

the following relationship:

$$\text{Stokes' shift} = 10^7 \left( \frac{1}{\lambda_{\text{ex}}} - \frac{1}{\lambda_{\text{em}}} \right) \quad [15]$$

where  $\lambda_{\text{ex}}$  and  $\lambda_{\text{em}}$  are the corrected wavelengths of the maximum excitation and emission, respectively (in nanometers). In order to record corrected excitation and emission spectra the wavelength dependence of the source, monochromators, and photomultiplier response must be identified and accounted for. The presence of the extraneous peak in the emission spectrum can be attributed to Rayleigh scattering of the excitation radiation. This is most commonly observed at either the excitation wavelength or twice this value due to second order diffraction from the grating in the emission spectrometer. Anthracene also exhibits phosphorescence in ethanol at 77 K with the wavelength of maximum emission being 462 nm. This is consistent with the lower energy of the lowest excited triplet state.

### Chemiluminescence

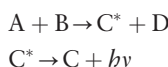
All chemical reactions are accompanied by energy changes. Any excess energy is usually dissipated by collision in around  $10^{-12} \text{ s}$ . Chemiluminescence is commonly observed at wavelengths from the near ultraviolet to the near infrared (which correspond to



excess chemical energy in the range from 340 to 130 kJ mol<sup>-1</sup>, respectively). Whilst these amounts of excess energy are achievable with certain reactions (often redox), the generation of chemiluminescence depends upon the suitable molecular structure of intermediates or products that will facilitate the conversion of chemical potential to electronic excitation.

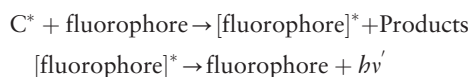
After chemical generation of the excited state has occurred, all of the environmental factors that influence photoluminescence are equally valid for chemiluminescence. It is generally accepted that most chemiluminescence reactions can be thought of as chemically induced fluorescence based upon the rarity of phosphorescence occurring in simple solution. Two noteworthy exceptions are the reduction of (1) tris(2,2'-bipyridyl)ruthenium(III) and (2) acidic potassium permanganate in the presence of certain polyphosphates. The emission in these cases appears to originate from very short-lived triplets. Many chemiluminescent reactions can produce light for several minutes or even longer. It should be emphasized that this is indicative only of the reaction kinetics, which produce the emitting species, and not the lifetime of the excited state.

Chemiluminescence can be rather simply classified as either direct or indirect. The former can be thought of as follows:



where A and B are reactants employed to produce either a product or intermediate in an electronically excited state (C\*) that returns to its ground state by ejection of a photon ( $h\nu$ ).

Indirect chemiluminescence has been used extensively for analysis and in the so-called chemical light sources; the most well known of these reactions involve oxidation of certain diaryl oxalates and oxamides. Instead of C\* returning to the ground state by photon ejection (as in the above reaction scheme), it can undergo energy transfer with a suitable fluorophore, which in turn may then exhibit its characteristic fluorescence emission:

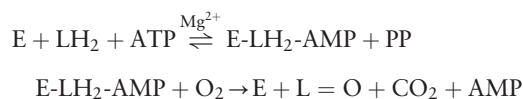


The analytical utility of indirect chemiluminescence depends upon little or no emission resulting from C\* in the region of maximum emission from the electronically excited fluorophore together with efficient transfer of excitation energy.

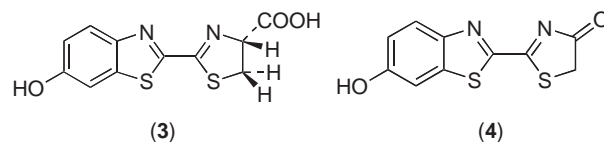
## Bioluminescence

The emission of light from fireflies has enchanted observers for millennia. However, it was not until comparatively recently (1947) that adenosine triphosphate (ATP) was identified as being the key component in the enzymatically controlled luciferin luminescence. The analytical utility of firefly bioluminescence was first demonstrated in 1952. Since that time the luciferin–luciferase reaction has been extensively employed for analysis in a wide variety of scientific disciplines.

While there are numerous species of firefly having similar biochemical pathways for the generation of light, the proposed mechanism for the luciferin–luciferase reaction is based upon the system of *Photinus pyralis*. The oxidation of D-luciferin (3) by oxygen, which is catalyzed by the enzyme luciferase, is thought to proceed by the sequence outlined below:



where E is the enzyme luciferase, L is luciferin (3), AMP is adenosine monophosphate, PP is pyrophosphate and L = O is oxyluciferin (4).



In the initial step, the D-luciferin is activated via interaction with Mg-ATP to give the enzyme bound D-luciferyl adenylate and pyrophosphate. As shown, this reaction is reversible and therefore pyrophosphate can impede the forward reaction. Generally, only low levels of pyrophosphate are formed from the concentrations of ATP and luciferase present under normal reaction conditions. The second step is an oxidative decarboxylation which produces oxyluciferin (4) in an electronically excited state. The excited oxyluciferin returns to the ground state by the ejection of a photon ( $\lambda_{\text{em}} = 562 \text{ nm}$ ) with an impressive overall quantum yield of  $\sim 0.9$  (at 25°C and pH 7.8). The light intensity is directly proportional to the ATP concentration provided that the level of luciferin remains constant.

There are numerous other bioluminescent systems of analytical importance, which have been isolated from various natural sources. A more detailed

**Table 2** Selected analytica applications of other luminescent phenomena

Class	Analyte	Comments	Reference
Sonoluminescence	Dissolved oxygen	No interference from anions present in natural waters	Yan L, Ruo F, and Zhaohua C (1995). <i>Water Research</i> 29: 2014
Thermoluminescence	Tetragonal zirconia in alumina–zirconia powders	Characterized by a peak at $-35^{\circ}\text{C}$ .	Salle C <i>et al.</i> (2003). <i>Journal of European Ceramic Society</i> 23: 667
Candoluminescence	Lead and tellurium	Matrix coated on a rod before placing in a hydrogen flame	Kassir ZM and Taher MB (1985). <i>Analyst</i> 110: 1223
Radioluminescence	Phosphorus	Photoluminescent technique with a light source based on radioluminescence	Leach AM, Burden DL, and Hieftje GM (1999). <i>Analytical Chimica Acta</i> 402: 267
Ionoluminescence	Chemical states of iron in a plagioclase sample	External proton beam induced peaks at 553 nm ( $\text{Fe}^{2+}$ ) and 682 nm ( $\text{Fe}^{3+}$ )	Sha Y <i>et al.</i> (2002). <i>Nuclear Instruments and Methods in Physical Research Section B</i> 189: 408

discussion of these systems is given elsewhere in this encyclopedia.

*See also:* **Bioluminescence.** **Chemiluminescence:** Overview. **Fluorescence:** Overview. **Phosphorescence:** Principles and Instrumentation.

## Other Types of Luminescence

Other classes of luminescence have been used in analytical applications; some examples are provided in Table 2. Thermally stimulated luminescence (or thermoluminescence) is the emission arising during mild heating of a solid material, often after it has been subjected to ionizing radiation. Thermoluminescence has been extensively employed for dating and dosimetry, including (for example) the detection of irradiated foods. Some incandescent solids emit at much shorter wavelengths than expected due to ‘candoluminescence’, which has been exploited for the detection of metals such as manganese, antimony, cerium, europium, terbium, lead, and bismuth.

Radioluminescence – induced by gamma- or X-rays – has also been exploited for dating. Radioluminescent light sources have been designed for spectrometric studies; the radioisotope and scintillation medium allows independent selection of the spectral and temporal characteristics. Sonoluminescence, the emission observed when particular solutions are exposed to ultrasonic waves, has the potential for analytical applications involving the detection of medium-to-high concentrations (over  $10\text{ g dm}^{-3}$ ) of elements that have an ionization energy below 7.65 eV and a boiling point below  $2700^{\circ}\text{C}$ .

Triboluminescence (arising from solids during structural rearrangements such as crushing) has limited applicability in analytical chemistry, but has been examined as a tool for clinical diagnosis with blood samples. The use of triboluminescent materials to detect damage in composite structures during impact has been suggested.

## Further Reading

- Baeyens WRG, de Keukeliere D, and Korkidis K (eds.) (1991) *Luminescence Techniques in Chemical and Biochemical Analysis*. New York: Dekker.
- Burr JG (ed.) (1985) *Chemi- and Bioluminescence*. New York: Dekker.
- Campbell AK (1988) *Chemiluminescence Principles and Applications in Biology and Medicine*. Chichester: Ellis Horwood.
- García-Campaña AN and Baeyens WRG (eds.) (2001) *Chemiluminescence in Analytical Chemistry*. New York: Dekker.
- Guilbault GG (ed.) (1990) *Practical Fluorescence*, 2nd edn. New York: Dekker.
- Harris DC (1991) *Quantitative Chemical Analysis*, 3rd edn., pp. 530–532. New York: W.H. Freeman.
- Harvey EN (1957) *A History of Luminescence from the Earliest Times until 1900*. Philadelphia: The American Philosophical Society.
- Hercules DM (ed.) (1966) *Fluorescence and Phosphorescence Analysis Principles and Application*. New York: Interscience.
- Kricka LJ and Carter TJN (eds.) (1982) *Clinical and Biochemical Luminescence*. New York: Dekker.
- Lakowicz JR (1999) *Principles of Fluorescence Spectroscopy*, 2nd edn. New York: Kluwer Academic/Plenum.
- Schulman SG (ed.) (1985) *Molecular Luminescence Spectroscopy. Methods and Applications, part 1*. New York: Wiley-Interscience.
- Schulman SG (ed.) (1988) *Molecular Luminescence Spectroscopy. Methods and Applications, part 2*. New York: Wiley-Interscience.
- Schulman SG (ed.) (1993) *Molecular Luminescence Spectroscopy. Methods and Applications, part 3*. New York: Wiley-Interscience.

Skoog DA, Holler FJ, and Nieman TA (1998) *Principles of Instrumental Analysis*, 5th edn., pp. 355–379. Fort Worth: Saunders College Publishing.

Valeur B (2002) *Molecular Fluorescence Principles and Applications*. Weinheim: Wiley-VCH.

van Wandruszka K (1992) Luminescence of micellar solutions. *Critical Reviews in Analytical Chemistry* 23: 187–215.

Wolfbeis OS (ed.) (1993) *Fluorescence Spectroscopy. New Methods and Applications*. Heidelberg: Springer.

## Solid Phase

**L F Capitán-Valvey**, University of Granada, Granada, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Luminescence – fluorescence, phosphorescence, and even chemiluminescence – may be emitted from the surface of powdered solid supports with small-sized particles or from membranes coming from compounds immobilized by physical or chemical procedures when exposed to by external radiation. The selectivity, sensitivity, speed, flexibility, and simplicity of solid-phase luminescence spectrometry (SPLS) make it a good analytical tool, especially in trace analysis. Because of its ease in handling, the use of SPLS has increased in different fields of interest and many different formats have emerged.

The reasons for the increase in the use of SPLS lie in its multidimensional character (spectral, lifetime, polarization, and other measurements) as well as in the need for a system that shows a large number of reactions and processes very efficiently. The advantages of SPLS include its sensitivity, low cost, ease in performance, versatility, and that it offers subnanometer spatial resolution with submicrometer visualization and submillisecond temporal resolution.

The reasons for the widespread use of SPLS luminescence are varied and generally make use of: improvements in the photophysical emission process, convenience of use, possibility of preconcentration, and use of a solid phase as a carrier for reactions or as a matrix for depositing or protecting reagents.

Referring here only to analytical purposes, SPLS is used for detection or determination of intrinsically fluorescent or phosphorescent compounds, for non-luminescent compounds that are capable of showing luminescence when they are derivatized, for non-luminescent, nonderivatized compounds that are capable of modifying the luminescent properties of a probe, i.e., via quenching or solvatochromic effect, and for compounds that interact via a binding partner and indicate this reaction with a luminescent label.

The different ways in which luminescence in solid phases is measured are: as a simple support, i.e., in intrinsically fluorescent compounds; as a support that makes a luminescent process possible in certain chemical conditions, i.e., phosphorescence; as a phase for preconcentrating the analyte or a derivative from a diluted solution due to the favorable distribution constant, i.e., polycyclic aromatic hydrocarbons (PAHs) on paper or C<sub>18</sub>-silica; as a phase that contains one or more immobilized reagents in a monolayer or multilayer format, enabling the reaction and retention of the analyte as well as different processes such as separation or others, i.e., test strips; as a phase that contains a compound that modifies its luminescent characteristics upon contact with the analyte, i.e., oxygen acting on a metallophorphyrin embedded in a polymer.

### Principles

Fluorescence arises from transitions that occur between singlet states and is not significantly modified upon being emitted from a solid substrate with respect to the solution, although in general the quantum yield is higher than in solution, while in the case of phosphorescence there are striking differences in quantum yield between solid and solution. Phosphorescence generally involves an intersystem crossing to a triplet state, and a subsequent radiative transition that is rather slow. In this case, nonradiative relaxation is the usual deactivation process of the long lifetime excited triplet state, due to intermolecular collisional quenching processes and intramolecular vibrational–rotational relaxation. Fixing a potentially phosphorescent analyte on a solid phase hinders its motion in this rigid environment and restricts radiationless deactivation, protecting the triplet state and increasing the quantum yield. There are two main solid-phase immobilization techniques: physical trapping in the rigid glass formed at low temperature and retention on a solid phase at room temperature. We are concerned with this second technique.

While in solution luminescence components are dissolved, in SPLS the electromagnetic radiation that

Skoog DA, Holler FJ, and Nieman TA (1998) *Principles of Instrumental Analysis*, 5th edn., pp. 355–379. Fort Worth: Saunders College Publishing.

Valeur B (2002) *Molecular Fluorescence Principles and Applications*. Weinheim: Wiley-VCH.

van Wandruszka K (1992) Luminescence of micellar solutions. *Critical Reviews in Analytical Chemistry* 23: 187–215.

Wolfbeis OS (ed.) (1993) *Fluorescence Spectroscopy. New Methods and Applications*. Heidelberg: Springer.

## Solid Phase

**L F Capitán-Valvey**, University of Granada, Granada, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Luminescence – fluorescence, phosphorescence, and even chemiluminescence – may be emitted from the surface of powdered solid supports with small-sized particles or from membranes coming from compounds immobilized by physical or chemical procedures when exposed to by external radiation. The selectivity, sensitivity, speed, flexibility, and simplicity of solid-phase luminescence spectrometry (SPLS) make it a good analytical tool, especially in trace analysis. Because of its ease in handling, the use of SPLS has increased in different fields of interest and many different formats have emerged.

The reasons for the increase in the use of SPLS lie in its multidimensional character (spectral, lifetime, polarization, and other measurements) as well as in the need for a system that shows a large number of reactions and processes very efficiently. The advantages of SPLS include its sensitivity, low cost, ease in performance, versatility, and that it offers subnanometer spatial resolution with submicrometer visualization and submillisecond temporal resolution.

The reasons for the widespread use of SPLS luminescence are varied and generally make use of: improvements in the photophysical emission process, convenience of use, possibility of preconcentration, and use of a solid phase as a carrier for reactions or as a matrix for depositing or protecting reagents.

Referring here only to analytical purposes, SPLS is used for detection or determination of intrinsically fluorescent or phosphorescent compounds, for non-luminescent compounds that are capable of showing luminescence when they are derivatized, for non-luminescent, nonderivatized compounds that are capable of modifying the luminescent properties of a probe, i.e., via quenching or solvatochromic effect, and for compounds that interact via a binding partner and indicate this reaction with a luminescent label.

The different ways in which luminescence in solid phases is measured are: as a simple support, i.e., in intrinsically fluorescent compounds; as a support that makes a luminescent process possible in certain chemical conditions, i.e., phosphorescence; as a phase for preconcentrating the analyte or a derivative from a diluted solution due to the favorable distribution constant, i.e., polycyclic aromatic hydrocarbons (PAHs) on paper or C<sub>18</sub>-silica; as a phase that contains one or more immobilized reagents in a monolayer or multilayer format, enabling the reaction and retention of the analyte as well as different processes such as separation or others, i.e., test strips; as a phase that contains a compound that modifies its luminescent characteristics upon contact with the analyte, i.e., oxygen acting on a metallophorphyrin embedded in a polymer.

### Principles

Fluorescence arises from transitions that occur between singlet states and is not significantly modified upon being emitted from a solid substrate with respect to the solution, although in general the quantum yield is higher than in solution, while in the case of phosphorescence there are striking differences in quantum yield between solid and solution. Phosphorescence generally involves an intersystem crossing to a triplet state, and a subsequent radiative transition that is rather slow. In this case, nonradiative relaxation is the usual deactivation process of the long lifetime excited triplet state, due to intermolecular collisional quenching processes and intramolecular vibrational–rotational relaxation. Fixing a potentially phosphorescent analyte on a solid phase hinders its motion in this rigid environment and restricts radiationless deactivation, protecting the triplet state and increasing the quantum yield. There are two main solid-phase immobilization techniques: physical trapping in the rigid glass formed at low temperature and retention on a solid phase at room temperature. We are concerned with this second technique.

While in solution luminescence components are dissolved, in SPLS the electromagnetic radiation that



impinges on the solid matrix and the luminescence radiation that is given off is scattered. Usually, a fraction of the analyte penetrates the solid phase to depths that differ depending on the solid phase, and luminescence is excited on the surface and within the matrix, giving off diffuse luminescence due to multiple scattering by particles or fibers of the solid phase. Diffusively transmitted luminescence may appear on the unexcited side of the support; however, it is the diffusively reflected luminescence that is usually measured.

Two different strategies have been employed to put the analyte in contact with the solid phase. The first consists of the use of membranes on which the analyte or a derivative is retained and which contain some or none of the necessary reagents. When placed in contact with the sample via immersion, moistening, or depositing a volume of the sample or an extract of it, either the analyte or a derivative of it is retained on the solid phase or interacts with some reagent present in it, generating the luminescent response that is then measured on the membrane itself. The second approach uses powdered solid supports with small-sized particles that are placed in contact with the sample, usually in a solution. The collection of this solid and the direct measurement of a luminescent property make it possible to then quantify the analyte.

The two methods – membranes and powdered supports – share the same basic concept and both are capable of being automated, though in different forms. In the first case, the solid supports are packed into flow-through cells and use flow-injection analysis (FIA) type systems. Membranes, on the other hand, have been implemented in probe-type systems both in batch or continuous mode.

## Types of Solid-Phase Procedures and Applications

### Batch Procedures

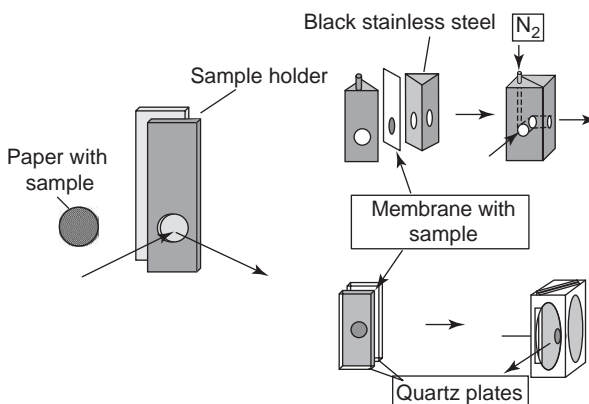
Batch mode can be performed in membranes and in powdered solids. The simplest procedures use the solid phase as a medium to support the analytes and carry out the measurement. Among membranes, a variety of cellulosic supports have been used, and the nature of the analyte plays an important role in producing luminescence. Polar or ionic molecules give off a strong adsorption interaction due to hydrogen bonding and induce intense phosphorescent signals, while nonpolar compounds produce weak dispersion interactions and small shifts in fluorescence spectra and weak phosphorescence. External heavy-atom perturbation can be used to induce

intense phosphorescence. Other membranes include filter paper treated with different chemicals such as NaF,  $\text{NH}_4\text{Cl}$ ,  $\text{NaCH}_3\text{COO}$ , salts containing heavy atoms ( $\text{AgNO}_3$ ,  $\text{Pb}(\text{CH}_3\text{COO})_2$ , NaI), sensitizers used as radiation absorbers as well as ion-exchange papers, and silica gel chromatoplates that interact with adsorbates via silanol groups or unspecific interactions. Powdered supports include sodium acetate, polymers/salt mixture, and cyclodextrins/salt mixture.

In this procedure, a microliter-scale volume of a sample is delivered to a disk of membrane support, then dried and placed in a suitable sample holder to measure the luminescent signal. In the case of powders, the analyte is added as a solution and then the solvent is evaporated. The dried residue is ground into a fine powder and placed in a phosphorescence cell or, alternatively, pressed into pellets (Figure 1).

SPLS is very sensitive and selective for organic trace analysis. Detection limits of a nanogram or even a picogram can be obtained. The methodology is simple, inexpensive, relatively precise (2–20%), relatively rapid, can handle small samples, and can be very selective in mixture analysis when solid-phase fluorescence (SPF) and solid-phase phosphorescence (SPP) are combined or when using derivative, synchronous, or time-resolved SPLS. Additionally, SPLS is well suited to being combined directly with both thin-layer and paper planar chromatography.

Drawbacks include (1) the need for some practice in sample preparation in order to obtain good results; and (2) the presence of background signals that affect detection limits and reproducibility and require the use of some means to lower these signals, such as cleaning the solid phase with polar organic solvents. Moisture and oxygen quenching affect SPP more than SPF but passing a dry inert gas into the cell compartment minimizes these problems. Additionally,



**Figure 1** Sample holders used in SPLS.



in SPP applications, the solvent used, chemical conditions of the analyte, type of solid matrix, and drying system must all be considered.

Different types of membranes such as paper, fiberglass containing C18 silica, or plasticized poly(vinyl chloride) (PVC) have also been used for pre-concentration purposes and subsequent SPF or SPP measurement. The equilibration of the membrane with the solution containing the analyte or a derivative can be carried out by adding the membrane to the solution or by filtration of the treated sample through it (solid-phase extraction). The improvement in the detection limit with respect to conventional procedures is between 100 and 1000 with a precision between 5% and 10%. Polychlorinated biphenyls from water have been preconcentrated in octadecyl-silica fiberglass, detected at subnanogram per milliliter levels and distinguished from PAHs by a combination of time discrimination and synchronous excitation techniques.

SPF and SPP can be applied in many areas such as clinical analysis, biomedical research, agricultural and pharmaceutical chemistry, energy technology, and environmental and industrial control.

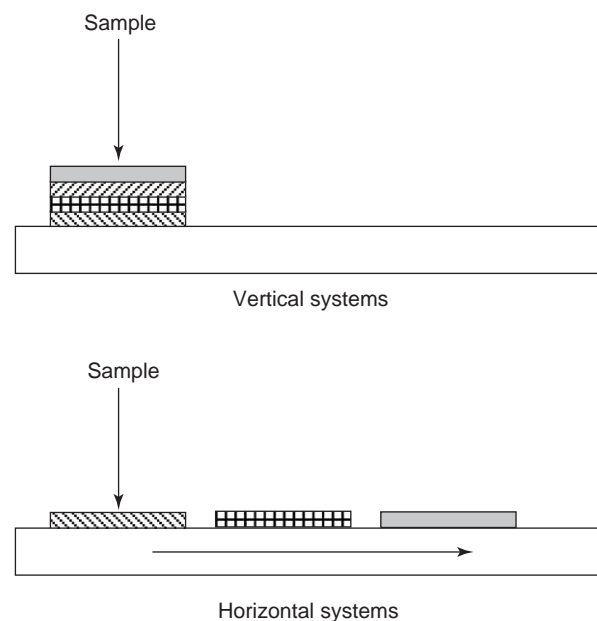
A second group of procedures relies on crystallophors or phosphors that are inorganic compounds (oxide, oxyhalide, or vanadate) acting as a receptor for some luminescence activating ion (analyte), present in a very low concentration. These ions supply energy levels that are responsible for the luminescent process and from which two types can be distinguished: those that have electronic emission levels that interact weakly with the matrix (trivalent rare earth ions) and those that interact in an intense way (transition element ions, electronic configuration ions  $s^2$  (Sb(III) or Bi(III)) and most of the ions of IVB, VB, and VIB groups). The preparation of phosphor matrices that include the analyte can be accomplished in different ways such as evaporation, crystallization, precipitation, or melting. This has been used for the determination of rare earths and transition elements in minerals, industrial products, or wastewaters with very good detections limits ( $3 \times 10^{-6}\%$  of U in  $\text{SiO}_2$  or  $10^{-8}\%$  of Mn in HCl), although these are inconvenient methods considering the preparation of the matrix.

A third type of procedure integrates all the necessary reagents in one or more layers of a reactive zone (test strips). To use these systems, a drop or precise volume of a fluid problem is placed on one of the surfaces of the solid-phase element by dipping or addition, where it diffuses into the reaction zone and dissolves the reagents. After the reaction, the generated luminescent signal is measured on the device itself by means of a compact analyzer or by a fully

automatic analyzer. These test strips have been used mainly for clinical analysis.

It is possible to classify these membranes by how they relate the sample containing the analyte to the test strip; that is, into vertical and horizontal systems. With vertical systems, the fluid goes through the layer or layers that make up the reagent zone in a perpendicular way, undergoing a series of reactions and processes that make it possible to develop an analytical property. In horizontal systems (lateral flow membrane), the sample is deposited on a reception zone and is moved by fluid flow through the test strip, producing chromatographic separations, reactions, and/or analyte retentions or interferences that generate the analytical property (Figure 2).

Among the sensing schemes implemented in luminescent vertical systems are: (1) Ionophores for alkaline and alkaline earth ion determination in biological fluids. The host-guest recognition process of the analyte by the ionophore is transduced by means of a fluorescent dye incorporated in a heterogeneous monolayer membrane or in a multilayer format. (2) Enzymatic reactions for analytes of clinical interest; for example, those based on the use of a dehydrogenase, a diaphorase, and a fluorophor. Enzymatic chemiluminescent test strips have been proposed that use photographic film positioned next to the strip or include a photoresponsive imaging layer in a multilayer strip. (3) Biological recognition elements, mainly antibody/antigen reactions. Here, different heterogeneous assays have been proposed that use the prior separation of bound and free labeled components and the measurement of test strip



**Figure 2** Test strip types.

luminescence. Homogeneous assays use different label systems such as: enzyme prosthetic group labels, enzyme substrate labels, coenzyme labels, enzyme modulator labels, or quenchable fluorescent labels.

Horizontal systems have been used mainly for heterogeneous bioassay where the separation of the free and bound labeled molecules is accomplished by capillary migrations through the carrier element.

Fluorescent labels are typically small organic dye molecules, such as fluorescein, Texas Red, or rhodamine, which can be readily conjugated to probe molecules. They show problems associated with background sample, autofluorescence, and photobleaching. Some alternatives include the use of fluorescent dyes, such as phycobiliproteins, that emit in the far-red or near-infrared where nonspecific fluorescent noise is reduced. Another alternative uses inorganic luminescent materials, like down-converting phosphors (i.e., ZnS activated by Ag or  $Y_2O_3$  activated by Eu) with narrow-band emission, large downward Stokes shift ( $\geq 100$  nm), long lifetime ( $> 1$   $\mu$ s), and stability, but low quantum efficiency in aqueous solutions. Another alternative is the up-converting phosphors (i.e., rare earth doped vitrocereamics) that absorb two or more photons in a frequency that is lower than the frequency of photon emission. Since up-conversion occurs within the host crystal, the optical properties of the phosphors are unaffected by their environment.

A fourth group of SPLS procedures uses a solid phase to preconcentrate the analyte as such or as a derivative from solution and consists of the use of powdered solid supports with small-sized particles ( $\leq 100$   $\mu$ m), which are placed in contact with the sample and to which are added the reagents needed for derivatization, elimination of interferences, adjustment of conditions (pH, ionic strength, solvent, etc.), and retention on the solid phase. The collection of this solid, packed in a 1 mm quartz cell, and the direct measurement of the luminescence signal makes it possible to then quantify the analyte.

Solid supports used include ion exchangers (polystyrene and polydextran type), sorbents (hydrophilic such as polydextran type and hydrophobic such as  $C_{18}$  silica), and modified sorbents, usually chelating sorbents, either with grafted groups or loaded with chelating ligands, i.e., silica modified with lumogallion used for aluminum analysis. Interaction between the analyte and the particulated solid phase can be accomplished in several ways. Retention of the intrinsically luminescent analyte can be accomplished by ion exchange or adsorption using hydrophilic or hydrophobic sorbents, as is the case with PAHs, pesticides, and drugs such as ciprofloxacin. When derivatives are used, the derivatization of the

analyte through complex formation or organic reactions can be accomplished in solution or directly in solid phase.

In the procedure, a variable volume of the sample is equilibrated in the presence of the solid phase and reagents, if necessary, by magnetic stirring in a glass or tube or by shaking in a rotating agitator. Measurement preparation involves attaining an evenly covered layer with the solid phase. The most common procedure uses a 1 mm optical quartz cell with the solid phase packed following solid-phase filtration or decanting. The intensity of the transmitted luminescence is usually measured, for which a cell holder fixed in a position at a  $45^\circ$  angle with the excitation and emissions beam is used.

These assays can be used for organic compounds such as pesticide and metal ion trace analysis, typically in water, mainly by SSF, and show good preconcentration factors and, consequently, good sensitivity and detection limits, with an acceptable reproducibility (1–3%) and equilibration times that are generally not very long. They use less-expensive and conventional instrumentation for analysis from nanogram per liter to microgram per liter levels and do not use solvents. The selection of the solid phase, chemical conditions, and/or use of derivative, and synchronous or variable-angle scanning luminescence makes it possible to resolve mixtures, along with, on occasions, multivariate calibration. Additionally, the sensitivity of the procedures can be adjusted by modifying the volume of the sample being used. The assays are inconvenient in that they can be slower than conventional methods and require, at times, specifically prepared solid phases.

A fifth group of procedures includes the use of an array format. The usual 96-well plate for immunoassay measures the fluorescence in liquid phase either by determining the unbound fraction or by dissolving the bound reporter molecules to overcome problems related to autofluorescence, scattering, and back-reflection, although the selection of an appropriate solid support allows for the measurement of the fluorescence directly on the solid surface. For example, a heterogeneous fluoroimmunoassay using fluorescein-labeled antibodies for total thyroxine in serum uses this procedure.

The miniaturization of microplates produces high-throughput screening arrays. Microarrays (biochip) usually consist of many – from a hundred to several hundred thousand – microscopic spots containing a binding partner, such as receptors, ligands, probes, or targets, immobilized within a small and defined area on a flat solid support of glass or polymer. Each spot contains identical molecules but the array as a whole contains different molecules, according to the

requirement. Fluorescence labels are by far those most often used to detect a binding event. The sample itself is labeled at some stage before incubation. Next the array is read, usually in a dry state, giving a snapshot picture of fluorescence intensity. The most widely used luminescent detection methods in high-throughput screening arrays include laser-induced fluorescence, fluorescence resonance energy transfer (FRET), fluorescence polarization, time-resolved luminescence, and fluorescence correlation spectroscopy.

As the ambient analyte theory predicts, the miniaturization of binding assays gives rise to increased sensitivity (nanomole per liter range) with low sample consumption (submicroliter volume range) and independence from the actual volume of sample. Currently, typical screening programs operate at throughput rates on the order of 10 000 compounds per assay per day.

Different types of molecules can be used as capture molecules in microarray assays, resulting in different interaction assays. Antigen–antibody, protein–protein, aptamers, enzyme–substrate, and receptor–ligand interactions, among others, have been used in miniaturized and parallelized immunoassay for diagnosis, with great possibilities for the diagnostic market. For DNA microarrays, which are based on nucleic acid–nucleic acid interactions, presynthesized oligonucleotides, or polymerase chain reaction fragments, are immobilized on the solid support. The targets of the analysis are extracted from the cell, luminescently labeled and hybridized to their immobilized complementary capture probes and then the capture targets are measured. DNA microarrays enable the production of expression profiles and their correlation with cell function, which is useful in different fields, such as cancer. Protein microarrays or proteome chips are potentially powerful tools for comprehensive analysis of protein–protein interactions. Cell microarray also makes it possible to perform a high-throughput characterization of gene function using cell arrays involving either living cells (microwells) or printed microarrays (slides).

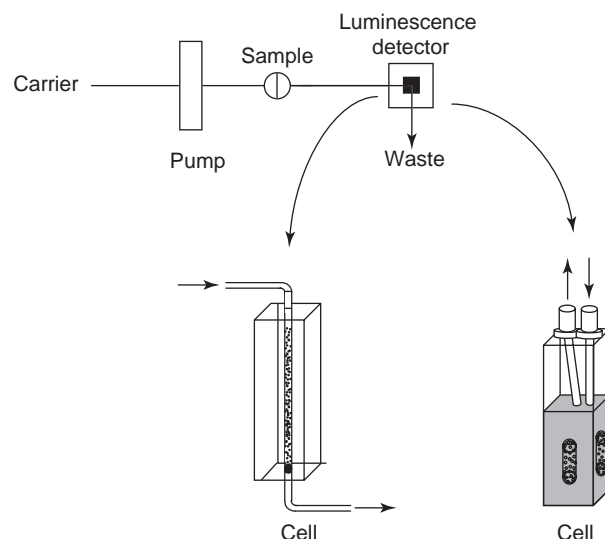
### Continuous Flow Procedures

The location of the membrane at the end of an optical fiber or the use of the flow-through cell of a FIA device with particles (usually larger than 30  $\mu\text{m}$ ) or films means an increase in the degree of integration of the procedure because the measurement stage is included with the reaction and detection and, thus, luminescent sensors can be designed. The solid phases used are widely varied and include ion exchangers, adsorbents, and organic or inorganic polymers.

Flow-through luminescent procedures have several advantages: continuous flow operation, easier automation, improved precision, and selectivity by using online separation chemistries. The drawback that water has in SPP batch procedures, especially with paper, disappears with continuous procedures, making it possible to obtain signals from aqueous flows (Figure 3).

Optical fiber (OF) luminescent sensors use an OF technology to transmit electromagnetic radiation from a sensing region that is in direct contact with the sample and contains a chemical recognition phase that generates an analyte-dependent signal. The advantages of OF luminescent systems come from their long-distance transmission, multiplex capability, immunity to electric or magnetic interference, easy miniaturization, and inherent sensitivity. The principal drawbacks are: leaching and photobleaching of dyes (photodrift), background signals from the fiber, and a long response time if a mass transfer to the membrane occurs. Various approaches have been used to solve the problem of photodrift, such as the ratiometric method, which refers the luminescent signal to the isosbestic point signal or to another analyte-dependent signal, or the use of an internal reference dye, such as the dual luminophore referencing method, wherein the analyte-sensitive fluorescence intensity signal is converted into the phase domain by coimmobilizing an inert long-lifetime reference luminophore with similar spectral characteristics. Sensing schemes based on lifetime measurement both in the frequency and time domains are more insensitive to signal drift.

The measurement schemes for intrinsic or extrinsic luminescence in FIA or OF may be similar and



**Figure 3** Luminescent optosensing manifold.

intensity, decay time, phase, or polarization can be selected as the analytical signal, although OF also offers alternatives such as the use of an evanescent wave to excite the luminescence of luminophors placed on the surface of the side of the fiber core.

Different configurations and chemistries have been used in SPLS continuous procedures with increased advantages over the batch procedures, since it is possible with the former to use systems that are not possible with the latter. The luminescence signal can be obtained from temporary or permanent immobilization of the analyte or reaction product in the solid phase in the case of flow-through systems. The sensing schemes for intrinsic luminescent analytes are not very common because of the lack of intrinsic luminophors, such as with PAH, polychlorinated biphenyls, or trinitrotoluene. It is much more common to use different sensing schemes to obtain a luminescent event from a nonluminescent analyte. Among others are

1. Conventional complex formation, mostly in ion-exchangers, for metal ions or organic complexing molecules. An example is the determination of lead in seawater with 8-hydroxy-7-quinolinesulphonic acid retained in an anion-exchange resin. An interesting option is the use of lanthanide-sensitized luminescence in SPP, as with tetracyclines. The slightly reversible character of most complexes due their high free enthalpy of formation does not make their use in OF formats possible, rather only in irreversible-reusable FIA sensors.
2. Ionophores to reversibly bind cations, anions, or organic compounds, usually incorporated in plasticized hydrophobic membranes. Different approaches for carrier-based fluorosensors are known, such as polarity-sensitive dyes based on the change induced by an analyte in the microenvironment. This is the case with the PVC membrane containing rhodamine B octadecyl ester sensitive to nitrate; other types are ion-exchange or coextraction mechanisms for cation or anions, respectively, based on the coupling, through an electroneutrality condition, of a proton exchange or extraction process highlighted by a lipophilized luminescent pH indicator.
3. Fluoroionophores that consists of a fluorophore, signaling moiety, linked to an ionophore, recognition moieties, through a spacer. The analyte recognition changes the photophysical characteristics of the fluorophore, due to electron transfer, charge transfer, energy transfer, excimer or exciplex formation, disappearance, etc. The most used types are photoinduced electron transfer and photoinduced charge transfer. They are used for cation sensing through chelator, podand, coronand, cryptand, or calixarene units, for anion sensing through hydrogen bonding or metal-ligand interactions, and for neutral analytes using fluorogenic reactants that perform reversible chemical reactions.
4. Chemiluminescent reactions through the immobilization of chemiluminescent reagents on different supports in OF or FIA format, as with the use of NADH oxidoreductase and bacterial luciferase for NADH determination at subpicomolar concentration.
5. Bioreceptor components including enzymes in combination with a luminescent indicator or reagent through a transducer such as oxygen, pH, or NADH; whole cells such as genetically engineered bacteria placed at the distal end of an OF, containing a gene-encoding bioluminescent protein that produces a bioluminescent signal in the presence of the analyte; animal or plant tissues, i.e., algal biosensors based on chlorophyll fluorescence for detection of herbicides; cell receptors, like the combination of bacterial cytoplasmic membrane and an oxygen-sensitive phosphorescent ruthenium complex for lactate detection; antibodies or antigens using auxiliary luminescent labeled reagents in a wide variety of competitive and noncompetitive assay formats.
6. Molecularly imprinted polymers (MIP) containing specific binding clefts with a predetermined size, shape, and three-dimensional arrangement of functional groups obtained by molecular imprinting with covalent or noncovalent interactions and used in different FIA or OF approaches, depending on the analyte. For intrinsically or labeled fluorogenic analytes, the measurement of bound analyte is used, as with *N*-dansyl phenylalanine using a polymer of methacrylic acid and 2-vinylpyridine. For nonluminescent analytes, different strategies are used, such as competition with a reporter molecule, alteration of wavelength, and/or intensity of a luminophore within the MIP or formation of a luminophore within the MIP.
7. FRET, which converts color changes into luminescence information, either intensity or lifetime, based on the overlap of an inert luminescent donor and the absorption spectrum of an analyte-dependent acceptor. High levels of CO<sub>2</sub> in a gas phase can be measured based on FRET between a ruthenium polypyridyl complex and the pH indicator Sudan III.
8. Twisted intramolecular charge transfer based fluorophores that involve states with high dipole moment and strong dependence of its fluorescence emission on the polarity of the microenvironment. These are useful for the determination of lower alcohols, for example, methanol or ethanol, with 4-(diethylamino)benzoic acid.
9. Quenchable luminophore based on dynamic quenching, quite used for oxygen, as well as for



heavy metal ions, humidity determination, or static quenching, as with curcumin for *o*-nitrophenol determination.

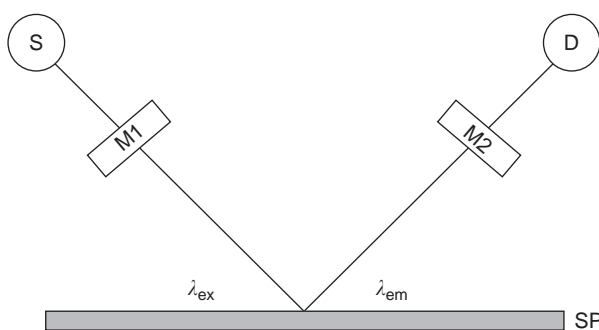
SPLS has been employed as a detection technique for high-performance liquid chromatography (HPLC), either by collecting fractions based on retention times and measuring SPF and/or SPP, as with tetrol metabolites of benzo[*a*]pyrene–DNA adducts or, alternatively, SPP can be carried out in a continuous flow mode with the aid of a two-nebulizer automatic system coupled to an HPLC system and used for the determination of caffeine, theophylline, and theobromine in the nanogram range. The use of a small flow cell loaded with a solid phase also makes HPLC luminescent detection possible. For example, terbium ions immobilized on silica have been used for the detection of nitrogen heterocycles.

## Instrumentation

Commercial instruments can be used for solid-phase luminescence by simply using a solid-phase sample holder that is available from the manufacturer or by constructing a homemade sample holder. The instrumental components are described elsewhere in the encyclopedia.

The instrumental components used for SPF are the same as those used for the measurement of solution fluorescence, although frontal cell geometry is preferred. In this way, the excited and viewed volume element moves next to the cell wall, reducing the attenuation of the signal. As light sources, arc and incandescent lamps, especially low-pressure mercury arc lamps and xenon lamps, are used, and lasers are used for steady-state or time-resolved measurements. The source radiation passes to a wavelength selection device (monochromator or filter) that selects the appropriate excitation radiation. The exciting radiation is passed to the luminophors containing the solid matrix, and from the luminescence that is emitted in all directions, a portion is collected, passed to the emission wavelength selector, and is imaged on the detector. Because the observed luminescence radiant power is usually low, the use of detectors such as photomultiplier tubes for fixed-wavelength or slow-scanning measurements is required. Complete emission spectra can be acquired rapidly with a multichannel detector interfaced to an emission spectrograph. Both intensified diode arrays and silicon-intensified target vidicons have been used (Figure 4).

To measure SPP in the absence of fluorescence and scattered radiation, it is necessary to use a phosphoroscope or a pulse-source gated detector. A phosphoroscope is a mechanical device, at present not used, that



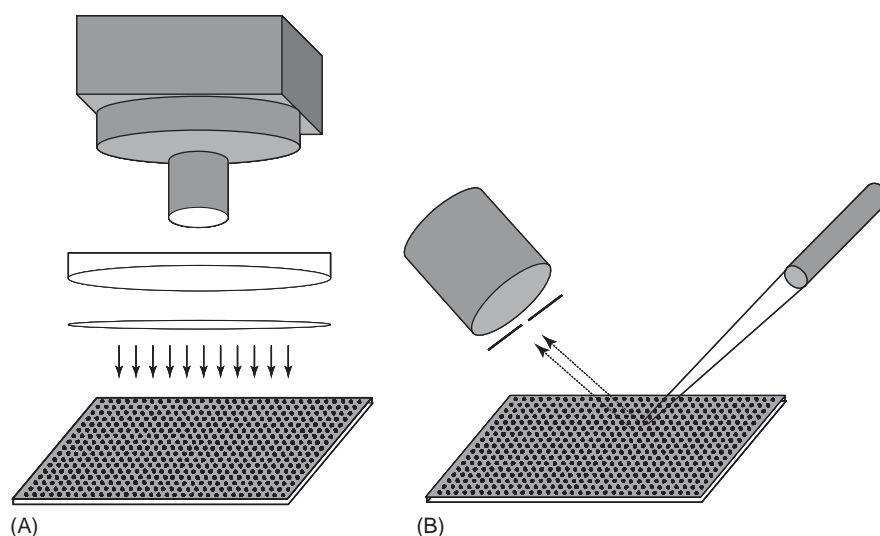
**Figure 4** Schematic diagram of instrumental components for the measurement of SPF. S, source; M1, excitation monochromator; SP, solid phase with sample; M2, emission monochromator; D, detector. (Adapted with permission from Robert JH (1995) *Luminescence: Solid Phase*, pp. 2749–2756; © Elsevier.)

makes out-of-phase excitation and measurement of phosphorescence possible. The pulse-source gated detector makes it possible to measure phosphorescence by time resolution achieved with pulse excitation and electronically gated detection. The solid phase is repeatedly excited with radiation from a pulsed or chopped light source and the resulting phosphorescence from a given pulse is measured following a suitable delay time after the end of the pulse.

In the case of OF luminescent sensors, both traditional optical systems and all solid-state systems have been used. Instruments from this second group operate using the same principle as the conventional system, except that LEDs and laser diodes form the main light sources and the signal processing units are fully electronic systems.

The miniaturization of luminescence assay screening has led to the development of new detection instruments. So far, detection is performed principally by luminescence using charge coupled device (CCD) cameras or laser scanners with confocal detection optics. CCD cameras offer a great amount of flexibility and sensitivity for the detection of luminescent labels in parallel. Quantitative fluorescent imaging depends on the linearity of photon capture and electron release by the CCD. For medium-density arrays (spot size > 200 μm), area imaging systems are rapid and appropriate. Confocal techniques are not effectively sensitive to miniaturization because of the small size of the confocal detection volume (typically femtoliters); consequently, confocal arrangements are preferable for high spatial resolution (high-throughput screening arrays), although they are more time consuming because they focus on each dot in the whole array. A sensitive alternative to confocal detection optics regarding the signal intensity, linearity, signal-to-noise ratio, and background





**Figure 5** Schematic of data acquisition in microarray technology: (A) charge coupled device (CCD) cameras; (B) laser scanners with confocal detection optics.

is the application of planar waveguide excitation devices combined with CCD cameras or photomultipliers as detectors and based on the excitation of confined surface fluorophores by the created evanescent field (Figure 5).

*See also:* **Bioassays:** Overview. **Chemiluminescence:** Overview. **Derivatization of Analytes.** **DNA Sequencing.** **Enzymes:** Immobilized Enzymes. **Flow Injection Analysis:** Detection Techniques. **Fluorescence:** Overview. **Immunoassays:** Overview. **Luminescence:** Overview. **Molecularly Imprinted Polymers.** **Optical Spectroscopy:** Detection Devices. **Phosphorescence:** Principles and Instrumentation. **Sensors:** Photometric.

## Further Reading

Capitán-Vallvey LF and Fernández Ramos MD (2003) Solid-phase spectrometric assays. In: Alegret S (ed.) *Integrated Analytical Systems*, pp. 81–159. Amsterdam: Elsevier.

Hurtubise RJ (1981) *Solid Surface Luminescence Analysis. Theory, Instrumentation, Applications*. New York: Dekker.

Hurtubise RJ (1990) *Phosphorimetry: Theory, Instrumentation and Applications*. New York: VCH.

Marazuela MD and Moreno-Bondi MC (2002) Fiber-optic biosensors – an overview. *Analytical and Bioanalytical Chemistry* 372: 664–682.

Niedbala RS, Feindt H, Kardos K, *et al.* (2001) Detection analytes by immunoassay using up-converting phosphor technology. *Analytical Biochemistry* 293: 22–30.

Sanz-Medel A (1993) Solid surface photoluminescence and flow analysis: a happy marriage. *Analytica Chimica Acta* 283: 367–378.

Templin MF, Stoll D, Schrenk M, *et al.* (2002) Protein microarray technology. *Trends in Biotechnology* 20: 160–166.

Vo-Dinh T (1984) *Room Temperature Phosphorimetry for Chemical Analysis*. New York: Wiley.

Werner T (1997) Fluorescence optical sensors for ion recognition. *EPA Newsletter* 60: 19–29.

# M

## MAGNETIC PROPERTIES

E Sinn, University of Hull, Hull, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 2765–2780, © 1995, Elsevier Ltd.

### Introduction

Mankind has been fascinated by magnetism since before recorded history. Magnetic iron-containing rocks like lodestone show the attraction and repulsion of a natural magnet. Navigating via the magnetic compass was the first widespread application of magnetism, but magnetic navigation predates us by eons: ferromagnetic crystals based on an iron-oxide hydrate complex in the brains of migratory birds allow them to fly halfway around the world without getting lost even when landmarks and stars are obscured.

Everything has magnetic properties, so it is not surprising that magnetism can be used for both quantitative and qualitative analysis. Magnetic studies and data have been widely reviewed and tabulated. Because of the wide range of magnetic properties that quite similar materials can possess, it is important to know about magnetism in general before it can be used analytically. Magnetism can be used for bulk analysis, distinguishing particular elements especially different metals in compounds, and for indicating structure types. Magnetic properties are best used in conjunction with other properties, and the more is known about a sample, the more use can be made of the magnetic properties.

### Types of Magnetism

#### Diamagnetism

This is caused by orbiting electrons and is therefore a property of all matter. Applied magnetic fields induce currents in electric conductors, and this includes electron orbits and the induced currents in turn decrease magnetic field strength within the conductors. These are the shielding currents that are the basis of nuclear magnetic resonance (NMR) spectroscopy.

The effect is small compared to paramagnetism (see below) and therefore treated as a correction to be added to the observed paramagnetism, i.e., how much the observed magnetism has been decreased by the diamagnetic shielding.

The effect will depend on the number and distribution of electrons, and is therefore roughly constant for a given element. The approximate molar diamagnetic susceptibility  $\chi^{\text{dia}}$  can be calculated by taking the tabulated contributions from the known values for each element, called Pascal's constants (Table 1), and adding them up, e.g.,

$$\begin{aligned}\chi_{\text{C}_6\text{H}_6} &= 6 \times \chi_{\text{C}} + 6 \times \chi_{\text{H}} + 6 \times (\text{constitutive correction}) \\ &= -6 \times 10^{-6}(6.00 + 2.97 + 0.24) \\ &= -55.02 \times 10^{-6} \text{ c.g.s.e.m.u}\end{aligned}$$

**Table 1** Molar correction ('Pascal') constants for atoms, ions, and multiple bonds

Transition metal	– 13	N, NRR'	– 2.1
Ag	– 31.0	N, N=CR	– 2.11
Ag(I)	– 24	Na	– 9.2
Al	– 13.0	NCO <sup>–</sup>	– 21
As(III)	– 20.9	NCS <sup>–</sup>	– 35
As(V)	– 43.0	NO <sub>3</sub> <sup>–</sup>	– 20
B	– 7.0	O, R–O–	– 4.61
BF <sub>4</sub> <sup>–</sup>	– 39	O, C=O	+ 1.73
Bi	– 192.0	O, carboxy	– 3.36
Br	– 30.6	O, O <sup>2–</sup>	– 12
Br <sup>–</sup>	– 36	P	– 26.3
C	– 6.0	Pb(II)	– 46.0
Ca	– 15.9	S	– 15.0
Cl	– 20.1	Se	– 23.0
ClO <sub>4</sub> <sup>2–</sup>	– 34	Si	– 20.0
CN <sup>–</sup>	– 18	Sn(IV)	– 30.0
CO <sub>3</sub> <sup>2–</sup>	– 34	SO <sub>4</sub> <sup>2–</sup>	– 40
F	– 6.3	Te	– 37.3
F <sup>–</sup>	– 11	Tl(I)	– 40.0
H	– 2.93	Y <sup>3+</sup>	– 12
Hg(II)	– 33.0	Yb <sup>3+</sup>	– 18
I	– 44.6	Zn	– 13.5
I <sup>–</sup>	– 52		
K	– 18.5	Multiple bonds	
Li	– 4.2	–C=C–	5.5
Mg	– 10.0	–C≡C–	0.8
N, aza chain	– 5.57	–C=C– in phenyl	– 0.24
N, ring	– 4.61	–C=N–	8.2
N, NHR	– 1.54	C=N–N=C	10.2

This is in good agreement with the observed diamagnetism for benzene ( $54.78 \times 10^{-6}$  c.g.s.e.m.u.) based on data collected in the *Handbook of Chemistry and Physics*.

The tables of diamagnetic susceptibilities in the above handbook can be used to obtain even better values for appropriate molecular fragments that are contained in the paramagnetic sample. When extremely accurate corrections are needed (e.g., when the material being analyzed is a protein), it is best to make measurements on the diamagnetic chemical equivalent of the paramagnetic complex in question, e.g., the diamagnetism of a zinc porphyrin could be used to estimate the correction in the iron complex with the corresponding porphyrin ligand.

### Paramagnetism

Electrons and nucleons have angular momentum (and hence the magnetism of a moving charge) as if they had spin as well as orbital motion. Paramagnetism is due to unpaired spins. Nucleon spins, the basis of NMR spectroscopy, have a small angular momentum compared with that of the larger electrons, and is normally ignored when considering the paramagnetism arising from electrons.

Electron spins produce a large effect. For ease of consideration, the electron motion is resolved into two components.

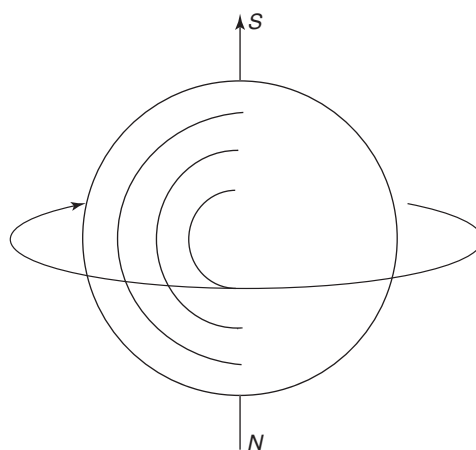
The charged particles are too small to observe directly if they are spinning, and it is not necessary to believe that electrons spin. Dirac showed that the apparent spin is purely a relativistic effect, getting larger for the faster-moving electrons of outer orbitals. However, the real angular momentum is conveniently calculated from the (fictitious in the Dirac theory) spin component and the conventional (i.e., classical mechanical) orbital component (Figures 1 and 2). Since it is the easiest way to obtain the correct results, the convenient fiction of electron spin, used by all and believed by many, will be retained.

In Figures 1 and 2, the effect of  $S$  can be removed only by pairing an electron with another of opposite spin, so the effect is always present when there are unpaired electrons;  $L$  is sometimes quenched, depending on molecular symmetry, in which case we speak of a 'spin-only' system.

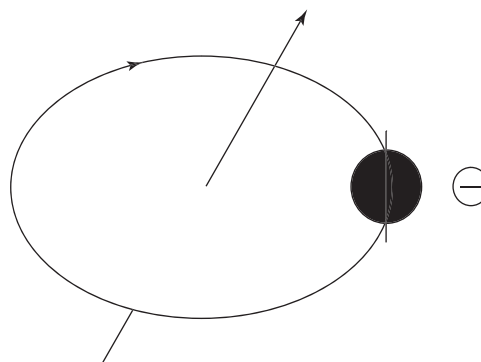
The electron spins tend to align with an applied magnetic field, but in the absence of the field, they lie in random orientation.

### Extended Magnetic Coupling

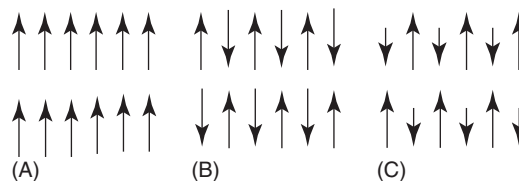
The main most important types of extended, or lattice-wide, magnetic coupling are ferromagnetism, which has spins lining up spontaneously even in the



**Figure 1** Spin angular momentum,  $S$ . Simplified view of spinning charged particle. The current loop generates a magnet.



**Figure 2** Orbital angular momentum,  $L$ . Simplified view of orbiting electron. Again the current loop generates a magnet.



**Figure 3** Line-up of electron spins in (A) ferromagnetism, (B) antiferromagnetism, (C) ferrimagnetism.

absence of an external magnetic field (Figure 3A); antiferromagnetism, which has electron spins on different atoms tending to line up antiparallel resulting in weaker magnetism (Figure 3B); and ferrimagnetism, which has two unequal ferromagnetic lattices lined up against each other, canceling out part of the effect. The last is a special kind of ferromagnetism (Figure 3C).

### Simplified Theoretical Basis

A simple theoretical consideration will indicate what it is that is to be measured.

Magnetization of a substance is defined as

$$\mathbf{B} = \mathbf{H} + 4\pi\mathbf{I}$$

where  $\mathbf{B}$  is the magnetic flux (field within the sample),  $\mathbf{H}$  is the applied magnetic field (*in vacuo*), and  $\mathbf{I}$  is the magnetization intensity or magnetic moment per unit volume.

Chemists often use the symbol  $\mathbf{H}$  in place of  $\mathbf{B}$ , but there is a difference between the two. Rearranging the terms in the equation:

$$\mathbf{B}/\mathbf{H} = 1 + 4\pi\mathbf{I}/\mathbf{H}$$

This gives  $\mathbf{I}/\mathbf{H}$ , the magnetization per unit magnetic field per unit volume, i.e., the magnetic susceptibility,  $\kappa$ , per unit volume.

The magnetic susceptibility per unit mass ('gram susceptibility'),  $\chi_g$ , is defined in terms of  $\kappa$  as

$$\chi_g = \kappa/d$$

where  $d$  is the density.

The magnetic susceptibility per mole ('molar susceptibility') is  $\chi_M = \chi_g$  (mol. wt.).

But as everything has a diamagnetic susceptibility, this must be corrected by subtracting the total diamagnetic susceptibility,  $\chi_M^{\text{dia}}$  (a negative quantity):

$$\chi_M = \chi_M^{\text{obsd}} - \chi_M^{\text{dia}}$$

The quantity most often quoted to express magnetic properties in metal complexes, the effective Bohr magneton number  $\mu_{\text{eff}}$  is defined in terms of  $\chi_M$ :

$$\chi_M = N \left( \alpha_0 + \frac{\mu_{\text{eff}}^2 \beta^2}{3kT} \right) \quad [1]$$

where  $N\alpha_0$ , the induced moment, a temperature-independent paramagnetism, is usually small and often ignored ( $\beta$  is the Bohr magneton,  $k$  the Boltzmann constant, and  $T$  the absolute temperature). (Although  $\mu_{\text{eff}}$  is often assigned the dimensions of Bohr magnetons, this is incorrect.) The magnetic moment in the field direction,  $\mu$ , is defined as  $\mu = -\partial E/\partial H$ , where  $E$  is the energy of the atom or molecule in field  $H$ .

Thus, the moment observed per mole,

$$M = N \frac{\sum_i \mu_i e^{-E_i/kT}}{\sum_i e^{-E_i/kT}}$$

based on  $i$  energy states, each with moment  $\mu_i$ . This gives

$$\chi_M = \frac{M}{H} = -\frac{N}{H} \frac{\sum_i (\partial E_i / \partial H) e^{-E_i/kT}}{\sum_i e^{-E_i/kT}} \quad [2]$$

In general,  $E_i$  can be expanded as a power series in  $H$ :

$$E_i = E_i^0 + E_i^{\text{I}}H + E_i^{\text{II}}H^2 + \quad [3]$$

which when substituted into [2] gives

$$\chi = -\frac{N}{H} \frac{\sum_i (E_i^{\text{I}} + 2E_i^{\text{II}}H + \dots) e^{E_i^0/kT} \cdot e^{-E_i^{\text{I}}H/kT} \dots}{\sum_i e^{-E_i^0/kT} \cdot e^{-E_i^{\text{I}}H/kT} \dots}$$

This is useful if it converges as it does for the 'small'  $H$  values that apply to common laboratory magnetic fields (for  $H = 1$  T (one 'tesla', the SI unit of magnetic field),  $E = g\beta H \approx 0.5 \text{ cm}^{-1}$ , while the thermal energy available at room temperature,  $kT \approx 200 \text{ cm}^{-1}$ ,  $g$ , the Lande factor, has a small numerical value, frequently close to 2).

Using

$$e^{\pm x} = 1 \pm \frac{x}{1!} + \frac{x^2}{2!} \pm \frac{x^3}{3!} + \frac{x^4}{4!} \pm \dots$$

gives

$$e^{-E_i/kT} \approx e^{E_i^0/-kT} \cdot e^{E_i^{\text{I}}/-kT} \approx e^{E_i^0/-kT} \left( 1 - \frac{E_i^{\text{I}}H}{kT} \right)$$

If, as is normally the case, there is no mean residual moment at  $H = 0$ , this gives

$$\Sigma(\mu)_{H=0} e^{-E_i^0/kT} = -\Sigma E_i^{\text{I}} e^{-E_i^0/kT} = 0$$

that is,  $\Sigma_i E_i^{\text{I}}$  always sums to zero (e.g.,  $g\beta H$  and  $-g\beta H$ ); the terms in  $H^2/kT$  are very small, so that

$$\chi_M = N \frac{\Sigma_i \left( (E_i^{\text{I}})^2 / kT - 2E_i^{\text{II}} \right) \cdot e^{-E_i^0/kT}}{\Sigma_i e^{-E_i^0/kT}} \quad [4]$$

This is the van Vleck equation, though strictly van Vleck derived the equation for the closely related electric susceptibility. Equation [4] holds for most cases encountered in the laboratory ( $g\beta H \ll kT$ ). At very large  $H$ , the approximations used in the derivation break down,  $\chi_M$  becomes dependent on the magnetic field, and paramagnetic saturation is said to be reached.

This is best illustrated by a simple example: For the spin-only  $S = \frac{1}{2}$  case, e.g., most simple copper(II) and vanadium(IV)O compounds,  $M_s = \pm \frac{1}{2}$  and  $E_i$  has only two values  $E_1 = -\frac{1}{2}\beta H$  and  $E_2 = +\frac{1}{2}g\beta H$  (Figure 4).

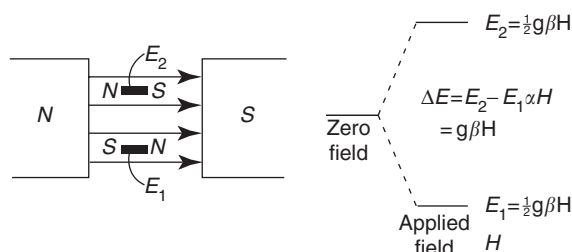


Figure 4  $E_i$  values for  $S = \frac{1}{2}$ .

In the format of eqn [3], these correspond to

$$E_1 = 0 - \frac{1}{2}g\beta H + 0H^2 + 0\cdots$$

$$E_2 = 0 + \frac{1}{2}g\beta H + 0H^2 + 0\cdots$$

Substituting in eqn [4],

$$\chi_M = N \frac{(-\frac{1}{2}g\beta H)^2 + (+\frac{1}{2}g\beta H)^2 \cdot e^0}{2e^0} = \frac{Ng^2\beta^2}{4kT} \quad [5]$$

It is also possible, and instructive, to derive this from eqn [2]:

$$\chi_M = -\frac{N(\partial E_1/\partial H)e^{-E_1/kT} + (\partial E_2/\partial H)e^{-E_2/kT}}{H(e^{-E_1/kT} + e^{-E_2/kT})}$$

$$= -\frac{\frac{1}{2}g\beta(e^{-\frac{1}{2}g\beta H/kT} - e^{\frac{1}{2}g\beta H/kT})}{(e^{-\frac{1}{2}g\beta H/kT} - e^{\frac{1}{2}g\beta H/kT})}$$

Using  $e^{-x} - e^x = -2x - x^3/3 - \cdots \approx -2x$  and  $e^{-x} + e^x = 2 + x^2 + \cdots \approx 2$ , this again gives eqn [5].

In the general case of spin  $S$ :

$$\chi_M = \frac{Ng^2\beta^2}{3kT}S(S+1) \quad \text{or} \quad \mu_{\text{eff}} = g\sqrt{S(S+1)} \quad [6]$$

Equations [5] and [6] show that  $\chi_M$  is inversely proportional to temperature, or  $\chi_M = C/T$  for spin-only systems. This is known as the Curie law ( $C$  is the Curie constant), and requires that  $\mu_{\text{eff}}$  be independent of temperature (see eqn [1]). Slight deviations in real systems from this behavior can be approximated with an extra temperature parameter,  $\theta$  (called the Weiss constant), in the Curie–Weiss law:

$$\chi_M = \frac{Ng^2\beta^2S(S+1)}{3k(T-\theta)} \quad [7]$$

If weak intermolecular coupling of the electrons is responsible for the deviation, a negative  $\theta$  represents a negative, or antiferromagnetic, coupling, while a positive  $\theta$  would mean ferromagnetic coupling.

$\mu_{\text{eff}}$  is readily obtained from  $\chi$  (eqn [1]):

$$\mu_{\text{eff}} = \sqrt{\frac{3k}{N\beta^2}\chi_M T} = \sqrt{7.997\chi_M T} \quad [8]$$

It is worth noting that  $\chi$  can deviate from the values predicted from the Curie law (and  $\mu_{\text{eff}}$  becomes temperature-dependent) if there is

1. Spin–orbit coupling in magnetically isolated metal centers (‘magnetically dilute material’).
2. Antiferromagnetic or ferromagnetic coupling or both involving two or more paramagnetic atoms in a cluster, without further coupling between the clusters.
3. Lattice-wide ferro- or antiferromagnetic coupling between paramagnetic neighbors.

4. Lattice-wide ferro- or antiferromagnetic coupling between clusters that already have internal coupling.

Only in case (3) is the Curie–Weiss law mathematically based as a valid approximation. However, (1), (2), and (4) also cause deviation from the Curie law and all can be approximated by the Curie–Weiss law in the high-temperature limit and it is common to obtain Weiss  $\theta$  values. This is acceptable so long as there is no pretence that the  $\theta$  value has a physical meaning (e.g., purely from mechanism (3)). More often than not, the reason for the deviation is not known.

If the number of unpaired electrons is  $n$ ,  $S = n/2$  in the spin-only case and  $\mu_{\text{eff}} = \frac{1}{2}g\sqrt{n(n+2)}$ .

For free electrons,  $g = 2.00$  (actually 2.0023 with relativistic corrections), so  $\mu_{\text{eff}} = \sqrt{n(n+1)}$ , but in real compounds,  $g$  can differ from 2, e.g., octahedral  $\text{Ni}^{2+}$ ,  $g \approx 2.2$ ;  $\text{Cu}^{2+}$ ,  $g \sim 2.1$ ,  $\text{Mn}^{2+}$  and  $\text{Fe}^{3+}$ ,  $g \approx 2.00$ . In general for transition metals, when the d-subshell is more than half-full,  $g > 2.0$ , and when less than half-full,  $g$  is close to but a little less than 2.0. This makes for a convenient determination of unpaired electron numbers. In octahedral geometry, spin-only cases are the ‘A’ and ‘E’ ground states, i.e.,  $d^3$  (e.g., chromium(III)), high-spin  $d^4$  (e.g., chromium(II)), high-spin  $d^6$  (e.g., iron(III)),  $d^8$  (e.g., nickel(II)), and  $d^9$  (e.g., copper(II)). In tetrahedral geometry,  $d^1$ ,  $d^2$ ,  $d^4$ ,  $d^6$ , and  $d^7$  are spin-only. Square planar, square pyramidal, and trigonal bipyramidal complexes are also normally spin-only.

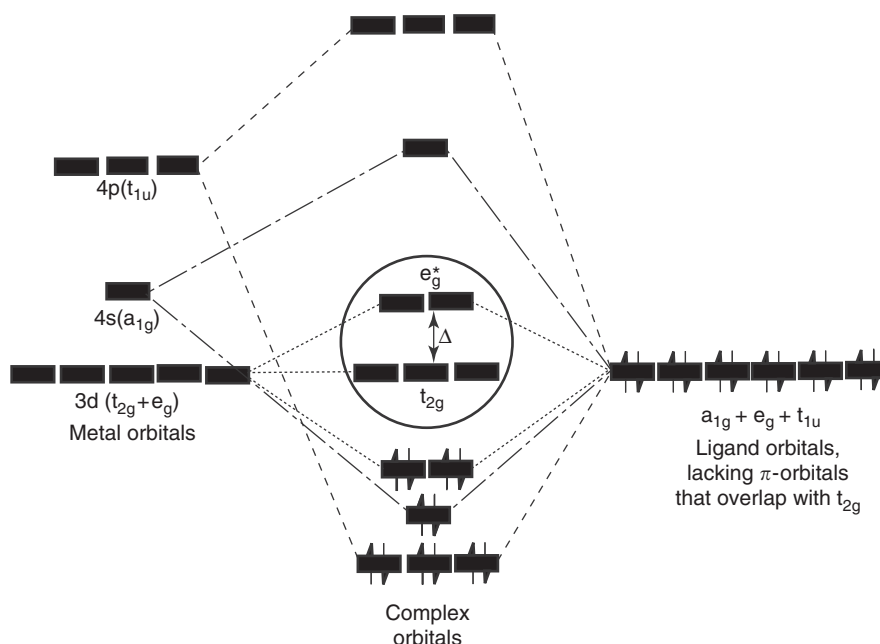
For nonspin-only cases, due to the contribution of orbital angular momentum ( $L$ ), larger (and smaller) values of  $\mu_{\text{eff}}$  may be observed, up to  $\mu_{\text{eff}} = \sqrt{4S(S+1) + L(L+1)}$ . Some specific cases are detailed below.

## Transition Metal Spin States and Magnetism

The molecular orbital diagram for even a simple octahedral complex looks complicated (Figure 5), but most of the time the information needed may be obtained by just considering the circled region, containing five orbitals:  $t_{2g}$  with threefold degeneracy and  $e_g$  with twofold degeneracy.

This is qualitatively equivalent to the results of the simpler crystal field theory, which ignores orbital overlap with the ligand orbitals and uses just the electrostatic repulsion of the metal d-orbitals and the negative charges of the ligands. Electrons in the  $t_{2g}$  orbitals are less repelled by the ligand charges, so these orbitals are lowest. The two approaches give different values for the splitting energy  $\Delta$ , which is normally determined empirically.





**Figure 5** Molecular orbital diagram for a simple octahedral complex.

For the species  $d^4$ ,  $d^5$ ,  $d^6$ , and  $d^7$  there are two choices, high spin or low spin, depending on how  $\Delta$  compares with the electron-pairing energy  $\Pi$ . If it takes more energy ( $\Pi$ ) to put two electrons into the same (lower) orbital than it takes to raise one electron into an upper orbital ( $\Delta$ ), the complex has the highest possible number of unpaired electrons or the highest value of spin ('high spin'). The configurations are:

high spin:

$$(d^7) t_{2g}^5 e_g^2, {}^4T_{1g}; \quad \frac{\text{CFSE}}{\Delta} = -\frac{4}{5}$$

$$(d^6) t_{2g}^4 e_g^2, {}^5T_{2g}; \quad \frac{\text{CFSE}}{\Delta} = -\frac{2}{5}$$

$$(d^5) t_{2g}^3 e_g^2, {}^6A_{1g}; \quad \frac{\text{CFSE}}{\Delta} = -\frac{0}{5}$$

$$(d^4) t_{2g}^3 e_g^1, {}^5E_g; \quad \frac{\text{CFSE}}{\Delta} = -\frac{3}{5}$$

where CFSE is the crystal-field stabilization energy.

The other oxidation states have only a single spin state to choose from:

$$(d^8) t_{2g}^6 e_g^2, {}^3A_{2g}; \quad \frac{\text{CFSE}}{\Delta} = -\frac{6}{5}$$

$$(d^9) t_{2g}^6 e_g^3, {}^2E_g; \quad \frac{\text{CFSE}}{\Delta} = -\frac{3}{5}$$

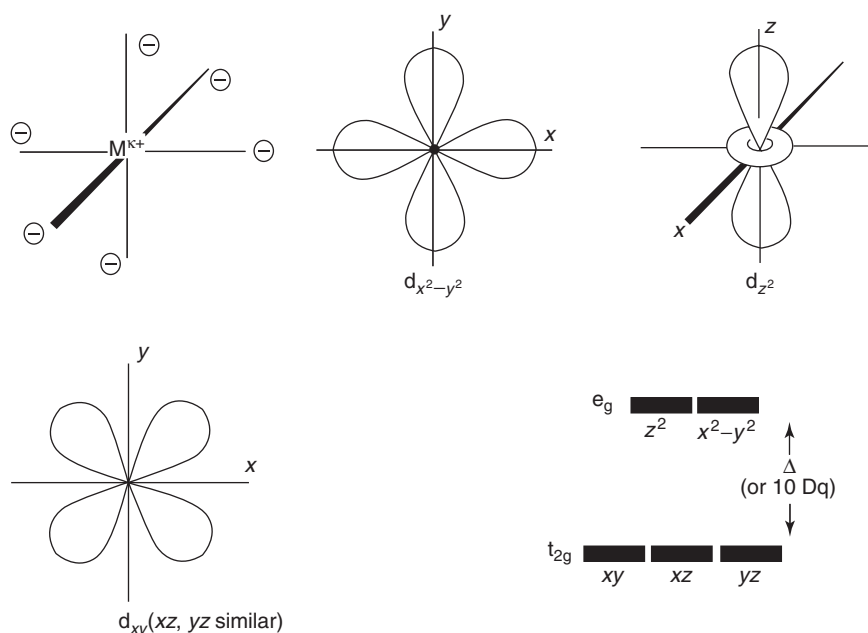
$$(d^3) t_{2g}^3, {}^4A_{1g}; \quad \frac{\text{CFSE}}{\Delta} = -\frac{6}{5}$$

$$(d^2) t_{2g}^2, {}^3T_{1g}; \quad \frac{\text{CFSE}}{\Delta} = -\frac{4}{5}$$

$$(d^1) t_{2g}^1, {}^2T_{2g}; \quad \frac{\text{CFSE}}{\Delta} = -\frac{2}{5}$$

$d^0$  (e.g., vanadium(V), titanium(IV), scandium(III)) and  $d^{10}$  (e.g., palladium(0), platinum(0), gold(I), copper(I), zinc(II)) are the simplest to consider. With empty and filled d-shells, respectively, they have no magnetic properties, no CFSE, and no ligand field spectra, as in nontransition metal complexes. There are some other special cases: low spin iron(II), cobalt(III), and their analogs in the second and third transition series are  $d^6$ , and are also diamagnetic. This makes cyano complexes like  $\text{Fe}[(\text{CN})_6]^{4-}$  and pyrite ( $\text{FeS}_2$ ) magnetically uninteresting, unlike high-spin iron(II), cobalt(II), and iron(III), and low-spin iron(III), which are often characterized by their magnetic properties.

High-spin manganese(III) and iron(III), having no CFSE, resemble the nontransition metals in having no preferred geometry. This enhances their ability to replace nontransition metal ions of similar size in minerals and complexes. An example is in ferrites,  $\text{M(II)Fe(III)}_2\text{O}_4$ , where the spinel structure is generally preferred, except that when M(II) is  $d^{6-9}$ , the inverse spinel structure is stabilized, with all the M(II) in octahedral sites (CFSE) and the iron(III),



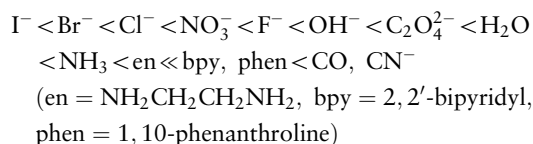
**Figure 6** Orbitals in octahedral complexes.

having no geometry preference, distributed equally over octahedral and tetrahedral sites. This enables strong magnetic coupling between the iron(III) in the octahedral and tetrahedral sites: hence,  $\text{MnFe}_2\text{O}_4$ ,  $\text{CoFe}_2\text{O}_4$ , and  $\text{NiFe}_2\text{O}_4$  have moments near the 6, 5, and 3 BM values expected for high-spin manganese, cobalt, and nickel, respectively, while  $\text{MgFe}_2\text{O}_4$  has a moment approaching the zero value for manganese.

Tetrahedral complexes can be treated in a similar manner to octahedral. Both octahedra and tetrahedra are cubic systems: in octahedral complexes (**Figure 6**) the ligands occupy the centers of the six faces of a cube. In tetrahedral complexes the ligands are diagonally across the cube faces from each other (**Figure 7**), so that the d-electron repulsion by ligands works in exactly the opposite sense, giving rise to exactly the inverse picture for the orbital energies to that for the octahedral case. The  $t_2$  orbitals point closer to the ligands, and are therefore more unfavorable places for electrons to occupy, than the  $e_g$  orbitals. The subscript g, for even parity, applies to undistorted octahedral, but not to tetrahedral geometry (no center of inversion) and is therefore dropped for tetrahedral. By visualizing energy in the inverse direction, the ground state for octahedral  $d^n$  is seen to be the same as that for tetrahedral  $d^{10-n}$ . The state symbols and CFSE formulas also apply as given, except that low-spin tetrahedral complexes tend not to occur: this would require a rigid ligand with donor atoms locked in tetrahedral geometry and a large value of  $\Delta_t$ . Otherwise, any four-coordinated complex with  $\Delta$  large enough to cause spin pairing would

also distort to planar, which offers a much larger CFSE.

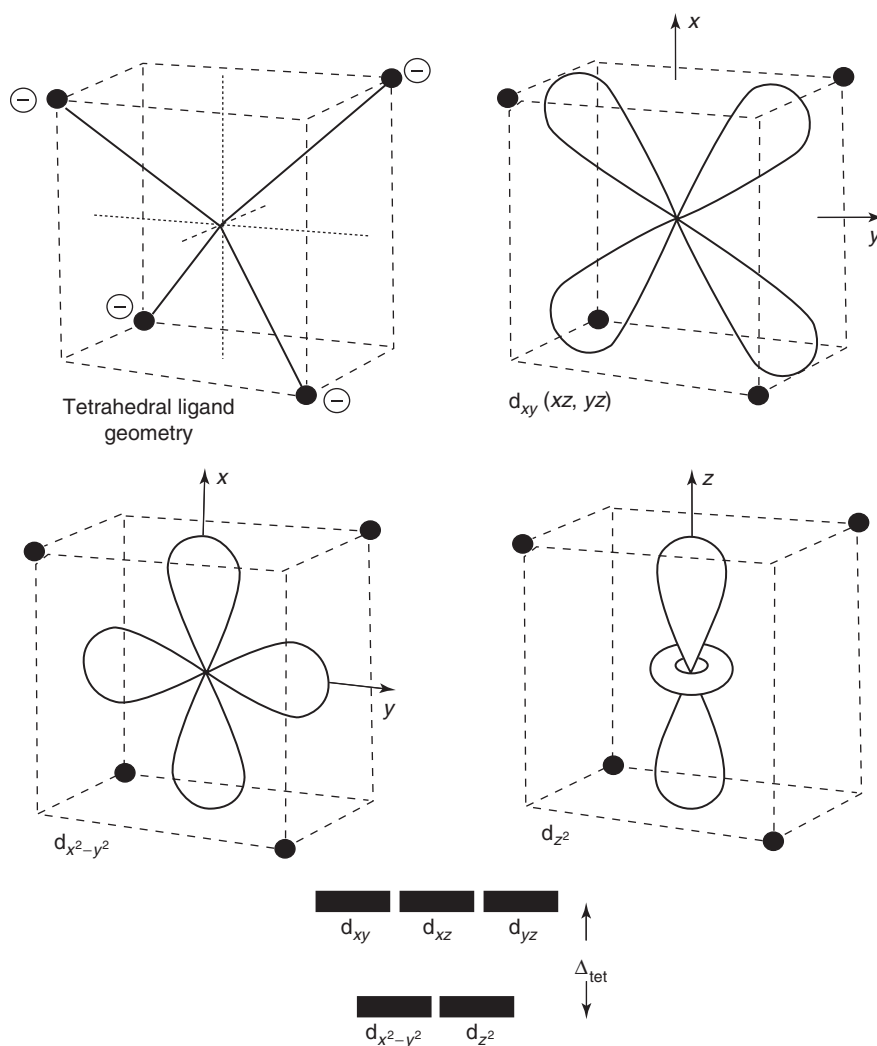
The value of  $\Pi$  does not vary much within complexes within the same transition series, but the size of  $\Delta$ , the ‘strength of the ligand’, varies widely. One way to estimate it is from d–d transitions, or ligand field spectra (hence the name, spectrochemical series):



This list gives an approximate indication of relative ligand field strength, but other factors, such as different geometries, also affect the actual crystal field experienced. In general, ligands before ‘en’ in the list do not produce low-spin compounds with iron, while those listed to the right of it do. Manganese(III) is harder to spin-pair, though  $\text{CN}^-$  does so.

## Methods of Measurement

According to what magnetic properties are to be measured and budgetary constraints, a variety of experimental techniques are available. If amount of sample is no limitation, the Gouy method is the cheapest to set up. In many cases, measurements at a single temperature cannot give unambiguous information, so variable temperature capability must be considered. The simplest quantity to measure is the force on a sample in the presence of a magnetic field.



**Figure 7** Orbitals in tetrahedral complexes.

### Gouy Balance

The Gouy method is the classical one and the easiest to set up. It is based on the apparent increase in mass of a sample hung above the axis of a magnetic field. The mass (sample plus holder) is determined by hanging the sample just over the magnetic field axis of an electromagnet that can be turned off or a permanent magnet that can be moved away (**Figure 8**). The upper end of the suspending chain attaches to a balance. The force is given as the apparent mass difference for field on and field off. The measurements must be made with the holder empty and then full of sample. The advantage is that sample mass is determined at the same time.

The force is given by the gradient of the magnetic field energy,

$$\mathbf{F} = \nabla \left( \frac{1}{2} \int_0^V K H^2 dV \right) \quad [9]$$

For a uniform sample in a tube of constant cross-sectional area  $A$ :

$$F = \frac{1}{2} \kappa A \nabla \int_0^L H^2 dz = \kappa A \int_0^L \frac{dH^2}{dz} dz = \kappa / B$$

where  $B$  is a constant of the experiment, provided the magnetic field is kept constant and the same sample tube, filled to the same point, is always used.

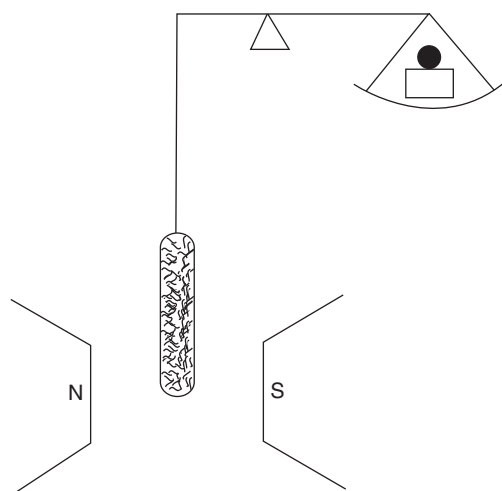
As  $F$  is always in the same direction, it need not be treated as a vector, and its value is the mass difference  $F_{\text{obs}} - F_{\text{empty tube}}$ .

Then  $\kappa = BF = \chi_g d$ , in a vacuum, or to good approximation in nitrogen gas. In air a correction for the paramagnetism of oxygen should be made:

$$\kappa_{\text{true}} = \kappa_{\text{obs}} + \kappa_{\text{air}}$$

i.e.,

$$\chi_g = \frac{BF + \kappa_{\text{air}}}{d}$$



**Figure 8** Schematic view of Gouy method.

This is one of the reasons why magnetic measurements are often made in gases containing no oxygen. For very accurate measurements, even the diamagnetic susceptibility of any surrounding gas must be taken into account.

If the sample tube is always filled to the same point,  $V$  is constant, and

$$\chi_g = \frac{BF}{(m/V)} = \frac{B'F}{m}$$

Calibration may be carried out with a known standard, e.g.,  $\text{HgCo}(\text{NCS})_4$ , for which  $\chi_g = 10.44 \times 10^{-6}$  c.g.s.u. at  $25^\circ\text{C}$ :

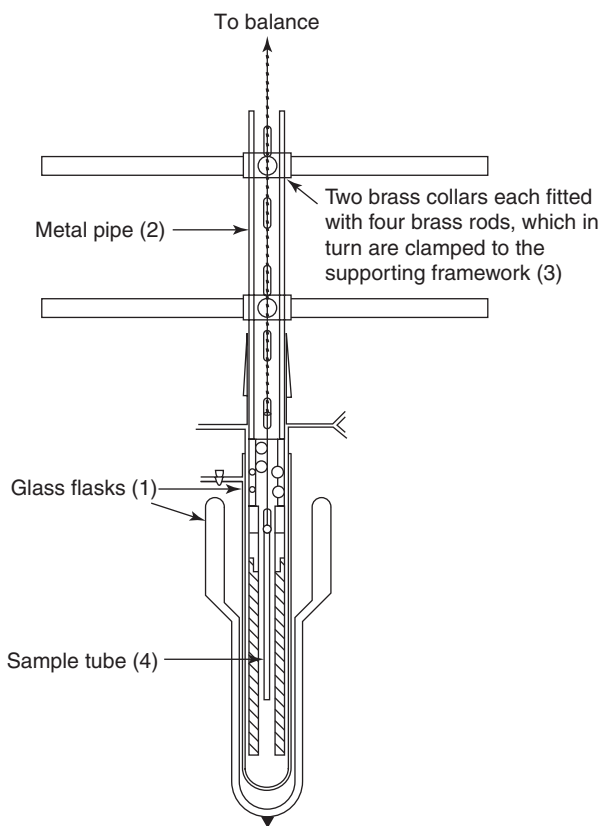
$$B' = \frac{\chi_{g(\text{std})} m_{\text{std}}}{F_{\text{std}}}$$

Then

$$\chi_{g, \text{sample}} = \frac{\chi_{g(\text{std})} m_{\text{std}} F_{\text{sample}}}{F_{\text{std}} m_{\text{sample}}}$$

The apparatus may be designed for variable temperature measurements by surrounding the sample area with a copper tube surrounded by a heater and by coolant liquid. Liquid nitrogen is the most convenient coolant. Lower temperatures may be achieved by pumping but because of the relatively large sample volume, other techniques are more convenient for temperatures down to liquid helium.

The Gouy apparatus is the easiest to make inexpensively. The main components are the magnet and a balance set up to weigh below the standard balance pan. Refinements to the setup include a transparent enclosing tube to shield the sample container and the chain supporting it against air currents. The apparatus can also be constructed for high precision. **Figure 9** shows a simple but moderately high precision setup that includes a metal (preferably copper) central tube. Holes cut on opposite sides in this make it



**Figure 9** Gouy apparatus between magnet poles.

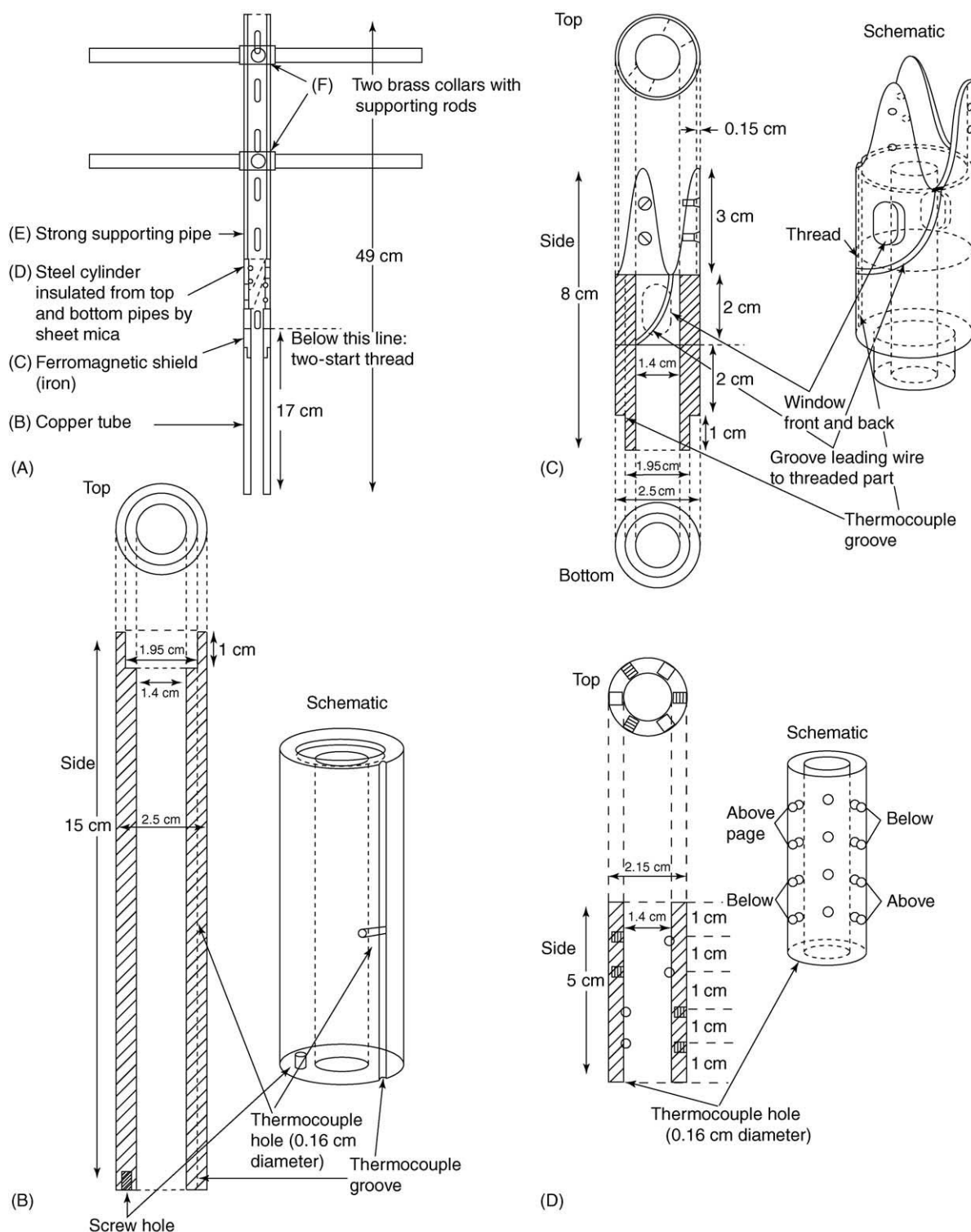
effectively transparent; grooves on this support windings of heating wire, and the high conductivity and heat capacity of the metal enhance thermal stability by smoothing out thermal spikes. The setup shown includes a ferromagnetic shield to allow measurements on shorter samples.

The ferromagnetic shield makes the outer assembly unstable to movement toward either pole face, requiring a stronger clamping arrangement. A pair of clamps, separated by a few centimeters, is best. Here brass collars are tightened around the metal tube with strong brass rods embedded in the collar. **Figure 10** shows the individual parts of the setup.

In particular, **Figure 10F** shows the collars and rods used to clamp the metal assembly in a rigid position. **Figure 10G** shows the glassware cryostat, and **Figure 10H** the balance mounting.

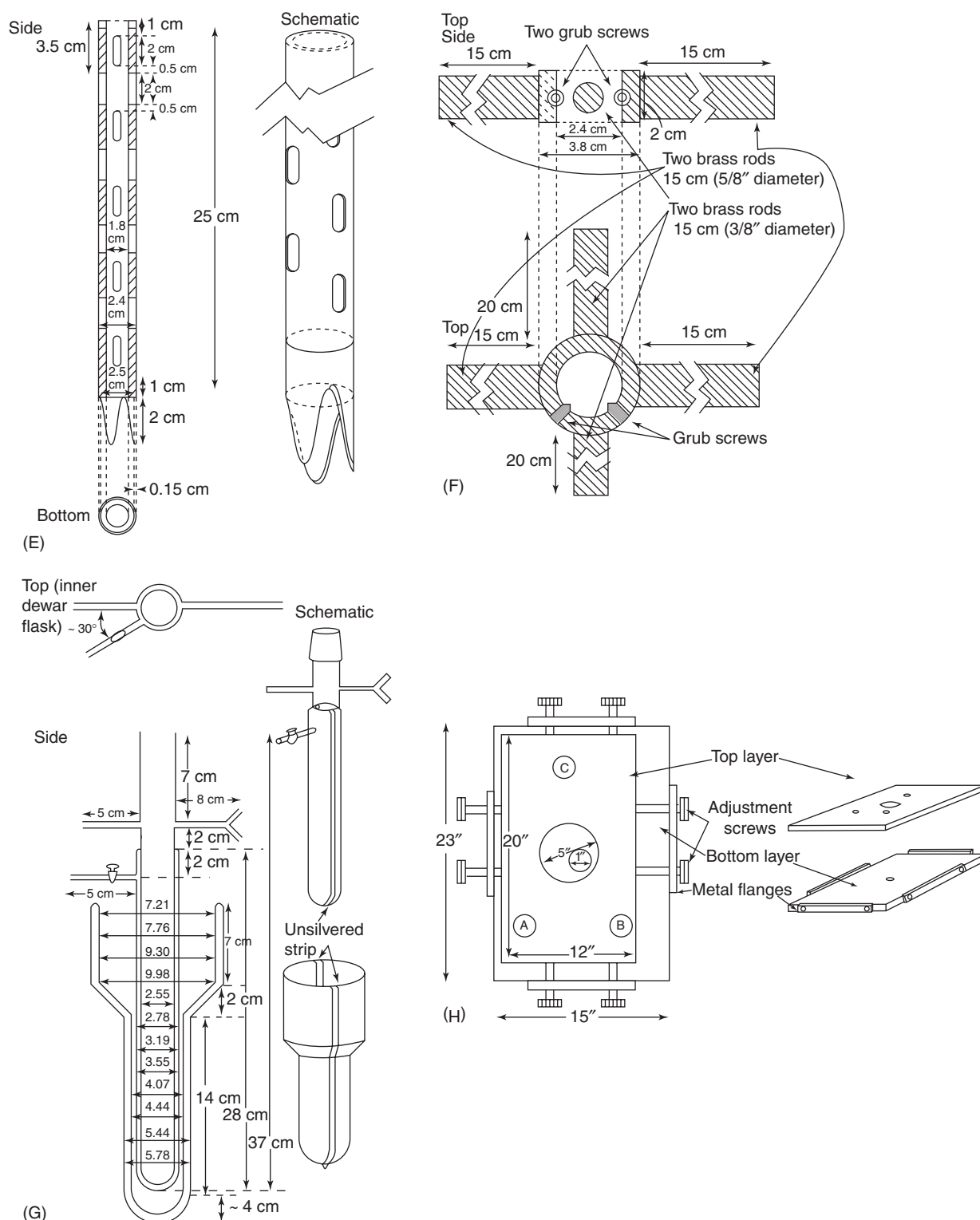
The effect of changing the shapes of the magnet poles is shown in **Figure 11**. A slight conical shape gives a higher field at the center but the rate of decrease with distance away from the center is less. The optimal choice is the conical shape but with a ferromagnetic shield to decrease the field strength away from the poles more rapidly.

With large samples, the Gouy method can be very accurate, but this can also be a disadvantage when



**Figure 10** Individual parts of a Gouy balance set-up: (A) metalwork section; (B) copper tube; (C) ferromagnetic shield; (D) steel cylinder (nonmagnetic); (E) supporting pipe; (F) two brass collars with supporting rods; (G) glass flasks; (H) balance mounting. *Bottom layer* is bolted to the table. The 2.5 cm hole continues through the table and into the sample compartment. Four metal flanges house two adjustment screws each. The adjustment screws can be used to move or to clamp the *top layer* when it is in the desired position. The *top layer* slides freely on the bottom layer, except when clamped by the adjustment screws. The balance rests on the top layer, with three circular recesses to accommodate the balance legs (A, B, C, 0.5 cm deep). The 13 cm hole is required to accommodate the attachments to the balance pan. When the balance has been centrally positioned, minor adjustments may be made to the position with the adjustment screws, without dismantling.

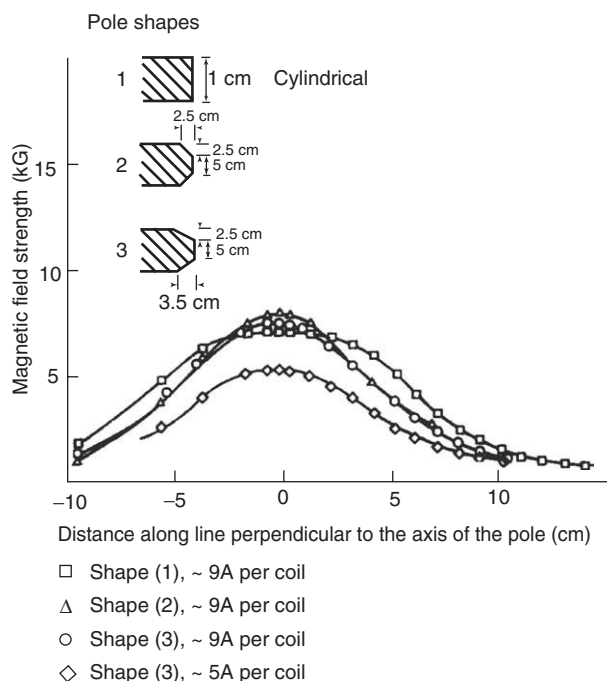




**Figure 10** (Continued).

large samples may not be available. A range of sample tubes is normally calibrated in preparation for a variety of sample sizes. Packing of the sample must be very uniform.

The sample must be large enough to fill at least a thin sample tube, though thin tubes make uniform packing more difficult. The tube must be long enough to allow a large drop-off in magnetic field over the



**Figure 11** Effect of changing the shapes of the magnet poles.

length of the tube, since the method relies on the maximum magnetic field at the lower end of the sample tube and zero at the top. The use of a ferromagnetic shield at the top allows for a faster drop-off in field, so that a shorter sample tube can be used.

The method is also good for solution samples where the packing problem is eliminated, though the dilution of a complex in solution correspondingly reduces accuracy. For this the high-precision version of the Gouy method is preferable. It can even be used for protein solutions where the paramagnetism is very small relative to the molecular weight.

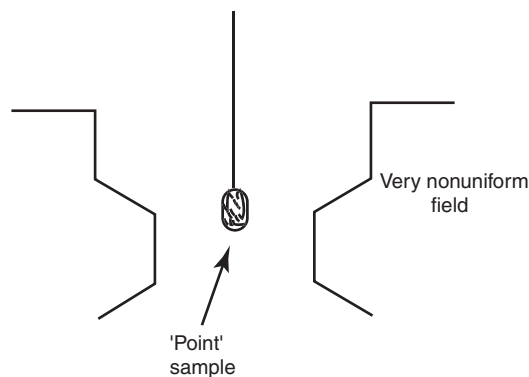
The Gouy balance can readily be adapted for high-pressure measurements. In this case, the high-pressure sample tube ('bomb') must be made of mechanically strong, but weakly magnetic material, e.g., beryllco-25 alloy. The sample, normally in solution, is pressurized in the tube, locked, and put in the Gouy balance.

### The Faraday Method

The Faraday method relies on the use of a point sample in a highly nonuniform magnetic field so that the paramagnetic sample is pulled to the strongest point. The high magnetic field gradient is achieved by unsymmetrically shaped pole faces (Figure 12).

Again using eqn [9] and the fact that  $V$  is very small:

$$F = \nabla\left(\frac{1}{2}KVH^2\right) = \frac{1}{2}KV\frac{\delta H^2}{\delta z} = \chi_g m \frac{\delta H^2}{\delta z} = \chi_g m/B'$$



**Figure 12** Faraday method with magnet poles designed to produce a highly nonuniform field.

where  $B'$  is a constant of the experiment. This leads to the same equations as for the Gouy balance above. Calibration is carried out with a known sample as above. In this method, the small sample size makes it imperative that the sample chosen to put in the Faraday bucket is representative. Because both the sample and the force on it are small, an electrobalance is normally used. Very high accuracy can be achieved, but because the method relies on approximating a point sample, it cannot be scaled up much to improve accuracy on a weakly magnetic sample. Ideally, the sample should correspond to a point but in practice a sample size of a few milligrams is commonly used. Variable temperature measurements can be made by a variation of the cryostat described for the Gouy balance.

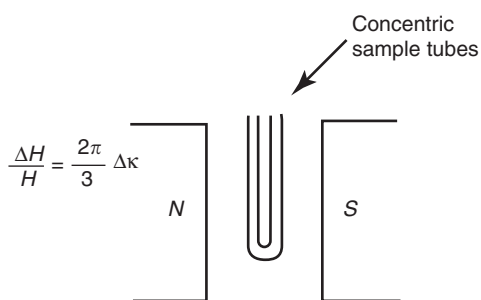
### Alternating Force Magnetometer

An advance over the Faraday method is to use induction coils, instead of specially shaped pole faces, to produce the large magnetic field gradient. The field gradient can then be varied independently of the static magnetic field, and can be reversed by reversing the current. The direction of the field gradient may be alternated periodically with respect to the static field, and the force detection passed through a phase-sensitive detector to decrease noise and improve accuracy. A variable temperature capability can be added as with the Faraday method.

### Evans NMR Method

This uses the shift  $\Delta H/H (= A\nu/\nu)$  in the NMR signal of a standard sensor, e.g., *t*-butanol, caused by a paramagnetic sample over the shift normally experienced by the sensor in the magnetic field alone (Figure 13).

Two concentric sample tubes are used, one of which contains the paramagnetic sample. Both



**Figure 13** Evans method.

contain the tensor. The volume susceptibility  $\Delta\kappa$  represents the increase in paramagnetism in the presence of the sample. This method is very convenient for determination of solution magnetism. Disadvantages are that only temperatures available to the NMR apparatus can be used and the accuracy is reduced to the proportion that the paramagnetic compound constitutes of the total solution.

### Induction Magnetometer

Insertion of a magnetic material into a conducting coil ( $n$  turns, area  $A$ , and length  $S$ ) induces a voltage  $V$  given by

$$V = \frac{pn^2 A}{S} \frac{di}{dt}$$

where  $di/dt$  is the alternating current frequency. The sample permeability,  $p$ , is related to  $\chi$  by  $p = 1 + 4\pi\chi$ . When the alternating frequency is high enough, the sample magnetization lags behind. The lag may be expressed as a phase angle  $\phi$ . If there is a fixed field  $H$  on which is superimposed an oscillation  $H_1$ , frequency  $\omega$ , such that  $H(t) = H + H_1 \cos \omega t$ , the resulting magnetization is  $M(t) = M + M_1 \cos(\omega t - \phi)$ .

### Vibrating Sample Magnetometer

This method uses the dipole field setup by an oscillating magnetic sample in a uniform magnetic field. The sample oscillation is driven by a loudspeaker that vibrates a sample holder rod, the other end of which moves the sample between the magnet poles. A permanent magnet reference sample is vibrated on the same rod. Separate induction coils, oriented with their axes parallel to the vibration monitor to allow phase-sensitive detection, monitor the sample and reference. Only a relatively small sample can be placed fully within a uniform field, but measurements on small single crystals are readily made with this apparatus. Measurements on anisotropic crystals can be made by remounting the crystal to vary the orientation with respect to the sample rod.

Limits of accuracy of commercial instruments are  $\sim 10^{-5}$  c.g.s.e.m.u.

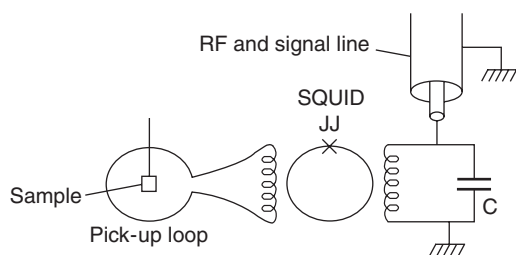
### SQUID Magnetometer

A SQUID (superconducting quantum interference device) consists of a superconducting loop with a weak link through which current continues to flow by quantum mechanical tunneling. The SQUID is very sensitive to small changes in magnetic field, e.g., due to the approach of a paramagnetic sample to produce a sample-dependent change in magnetic flux. This signal is magnified to give the most accurate of all techniques for measuring magnetic susceptibility. The accuracy is frequently described in terms of fractional change of a fluxon (individual line of magnetic force) and the only ultimate limitation is the Heisenberg uncertainty principle. Because of the high sensitivity ( $10^{-12}$  c.g.s.e.m.u.), it is common practice to shield the magnetometer, e.g., with mu-metal around the SQUID, sample, and cryostat assembly. The optimal design has a SQUID coupled inductively to a superconducting detecting loop into which the sample is inserted (Figure 14). The overall design of a SQUID magnetometer used in the author's laboratory is shown in Figure 15.

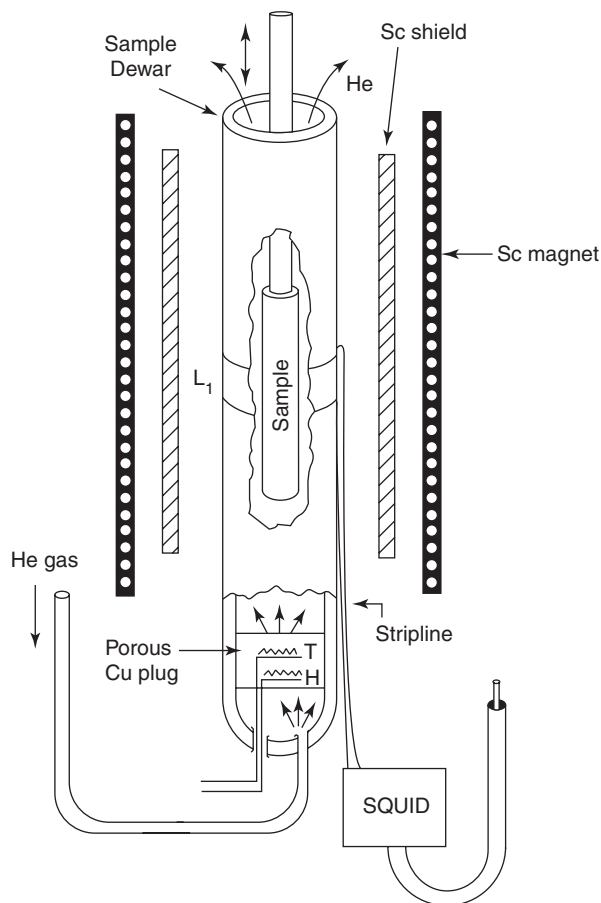
As the SQUID measures changes in magnetic flux, it is possible to measure either a sample with fixed magnetic moment, which must be moved into or out of the detector region, or a sample with varying moment. The latter would be held in place while the perturbing influence, e.g., pressure, temperature, light excitation, chemical reaction, is allowed to change the magnetic moment. The SQUID response time is limited by the frequency (normally RF) used in the detection circuit, which readily allows the study of magnetic transients of less than 1 ms, but faster response times are being produced for shorter-lived transients.

### Applications

Sometimes magnetic properties constitute supplementary information to support other data such as microanalyses. However, by establishing how magnetic properties relate to structure they can be used in cases where structural information cannot be obtained in other ways. An example is the determination of the nature of the active sites in metalloenzymes: nearly half the important enzymes contain one or more transition metals at the active site. Because of the large molecular weight, it is frequently impossible to determine how many metal atoms there are from simple chemical analysis. This was the case with the purple acid phosphatase (p.a.p.) series.



**Figure 14** A SQUID coupled inductively to a superconducting detecting loop into which the sample is inserted.



**Figure 15** Schematic outline of a SQUID magnetometer.

The p.a.p. from animal sources was thought to contain one Fe per molecule by Aisen *et al.* (rounding down from an analysis of 1.3) and by Zerner *et al.* to contain two iron atoms (rounding up from 1.6) based on the same bulk iron analysis result within experimental error. Knowledge of the number of iron atoms present is important in understanding the function and mechanism of the enzyme, author's measurements on this system were predicated on the assumption that if there were two iron atoms the magnetism would be twice as great than if only one iron were present. Surprisingly, the magnetism was

much less than expected for either case, from which it became clear that iron is present but in a very specific way, viz., two magnetically coupled iron atoms linked as Fe(III)–O–Fe(III) (eqn [10], with  $S = 5/2$ ):

$$\chi_M = \frac{Ng^2\beta^2}{3kT} \cdot \frac{\sum_s S(S+1)(2S+1)e^{-E(S)/kT}}{\sum_s (2S+1)e^{-E(S)/kT}} \quad [10]$$

Thus, the magnetic technique not only gave the correct amount of iron but also showed the essential structural feature. The strong magnetic coupling requires that there be no electron spin resonance (ESR) spectrum, yet the enzymes had previously been characterized via ESR. The so-called characterizing spectrum was due to the presence

$$\chi_M = \frac{N\beta^2 \cdot \left[ (S+1)(2S+3)(2Sg_1+g_2)^2 + \frac{8}{x}S(S+1)(g_1+g_2)^2 \right]}{12(2S+1)kT[(S+1)e^x + S]} \quad [11]$$

of a ubiquitous impurity and in this case the magnetic technique is far superior to another (highly sensitive) analytical method (ESR).

When one Fe is replaced by copper(II), the magnetic behavior (eqn [11]) is rather similar to what is observed for another enzyme, cytochrome oxidase (Fe–Cu). Where  $g_1$  and  $g_2$  apply to the copper and iron, respectively, the spin  $S$  for iron would be  $5/2$  in the normal iron(III) state, or  $2$  for iron(II). When one of the irons is replaced by a zinc(II), the magnetism increases strongly, corresponding to a single iron(III), (zinc(II), having no d electrons, contributes nothing) and its behavior conforms to eqn [6], the Curie law. The Zn–O–Fe form still works as a normal enzyme on its substrate. This is very significant because the vegetable acid phosphatase (from kidney beans) analyses for iron and zinc and has Curie law magnetism, implying a Zn–O–Fe center. When the zinc is replaced with iron it behaves like the animal form, and also still works on its substrate; nature had long ago done the metal substitution experiment.

Again, the enzyme urease was found to contain a dimeric nickel(II) center based only on its magnetic properties (eqn [10] with  $S = 1$ ). In large proteins such as ferritin, strong coupling is seen in a Fe–O–phosphate center, rather like a small piece of rust enclosed in a protein sheath. Likewise, nitrogenases contain a number of linked irons.

The use of magnetism in studies of minerals and transition metal complexes is alluded to above. Individual metal atoms reveal their presence and bonding environment (planar, tetrahedral, etc.,

including distortion) in their  $T$ -dependent magnetic properties – deviation from Curie–Weiss (eqn [6]).

$$\chi_M = \frac{Ng^2\beta^2}{3kT} \cdot \frac{(S+1)(S+2)(2S+3)x + S(S+1)(2S+1)y + S(S-1)(2S-1)}{(2S+3)x + (2S+1)y + (2S-1)} \quad [12]$$

$$\chi_M = \frac{Ng^2\beta^2}{12kT} \cdot \frac{(2S+3)(2S+5)(S+2)a + (2S+1)(2S+3)(S+1)b + S(2S-1)(2S+1)c + (2S-3)(2S-1)(S-1)}{(S+2)a + (S+1)b + Sc + S-2} \quad [13]$$

Magnetic coupling, lattice-wide, is likewise revealed by  $\chi$  versus  $T$  dependence, and this shows how closely the metals are linked. This is a large subject, beyond the scope of this article.

Many systems have small-scale linkages where just two or three metals link in a molecule. Typical examples are copper dimers. Thus, the magnetic susceptibility of copper(II) acetate deviates dramatically from the Curie law. It was analyzed as a binuclear rather than monomeric structure long before X-ray crystallography confirmed this. Up to a third of copper(II) compounds are now known to have some such magnetic linkage that can be analyzed as ligand bridging.

Current interest in advanced, or twenty-first century, materials makes coupling between transition metals and lanthanoids important, so that highly magnetic centers (large  $S$  values) must be considered.

For example, for an ( $S=1$ ) metal coupling with another metal having a spin  $S$  of at least 1 (i.e.,  $S_1=1$  and  $S_2=S>\frac{1}{2}$ ), the magnetism is described by eqn [12] where  $x = e^{(2S+4)J/kT}$ ,  $y = e^{2SJ/kT}$ .

In a  $\text{Ni}^{\text{II}}\text{--Co}^{\text{II}}$  ( $S=3/2$ ) system, eqn [12] becomes

$$\chi_M = \frac{Ng^2\beta^2}{4kT} \cdot \frac{35e^{8J/kT} + 10e^{3J/kT} + 1}{3e^{8J/kT} + 2e^{3J/kT} + 1}$$

For an ( $S=3/2$ ) metal coupling with another metal having a spin  $S$  greater than 1 (i.e.,  $S_1=3/2$  and  $S_2=S>1$ ), the magnetism is described by eqn [13] where  $a = e^{(6S+3)J/kT}$ ,  $b = e^{4SJ/kT}$ ,  $c = e^{(2S-1)J/kT}$ .

For example, in coupling an  $S=3/2$  transition metal with an  $S=3$  lanthanoid, eqn [13] becomes

$$\chi_M = \frac{Ng^2\beta^2}{4kT} \cdot \frac{165e^{21J/kT} + 84e^{12J/kT} + 35e^{5J/kT} + 10}{5e^{21J/kT} + 4e^{12J/kT} + 3e^{5J/kT} + 2}$$

From the strength of coupling and the shape of the curve we can discriminate between different molecular forms. Each metal pair has its own characteristic signature: the  $\chi$  versus  $T$  curves represented by these equations can discriminate reliably between the different centers represented. Accurate measurement gives the strengths of coupling ( $J$ -values), which

indicate how strong the link is between the metals, e.g., a single atom bridge or a larger bridge like an

imidazole or carboxyl group. This even discriminates between different types of one-atom bridges, e.g., the oxidized form of p.a.p. has strong coupling  $J$  ( $-80\text{ cm}^{-1}$ ) due to  $\text{Fe(III)–O–Fe(III)}$  whereas the reduced form has much weaker coupling ( $-20\text{ cm}^{-1}$ ) indicating an OH bridge (having weaker Fe–O links and therefore weaker orbital overlap).

**See also:** **Electron Spin Resonance Spectroscopy: Principles and Instrumentation.**

## Further Reading

- Brill AS (1977) Transition metals in biochemistry. *Molecular Biology, Biochemistry and Biophysics*, vol. 26. Berlin: Springer.
- Carlin RL (1986) *Magnetochemistry*. New York: Springer.
- Carlin RL and van Duijneveldt (1977) *Magnetism of Transition Metal Complexes*. New York: Springer.
- Figgis BN and Lewis J (1960) The magnetochemistry of complex compounds. In: Lewis J and Wilkinson RG (eds.) *Modern Coordination Chemistry*, pp. 400–454. New York: Interscience.
- Gerloch M (1983) *Magnetism and Ligand-field Analysis*. Cambridge: Cambridge University Press.
- Mockler GM, O'Connor CJ, deJersey J, Zerner B, and Sinn E (1983) On the active site of pig allantoic fluid acid phosphatase and a model complex for strongly coupled diiron sites. *Journal of the American Chemistry Society* 105: 1891–1893.
- Mulay LN and Boudreaux EA (1976) *Theory and Application of Molecular Diamagnetism*. New York: Wiley-Interscience.
- O'Connor CJ (1982) Magnetochemistry – advances in theory and experimentation. *Progress in Inorganic Chemistry* 29: 203–283.
- O'Connor CJ, Cukauskas EJ, Deaver BS Jr, and Sinn E (1979) Low temperature magnetic properties and antiferromagnetic interactions of the magnetic susceptibility calibrant  $\text{HgCo(NCS)}_4$ . *Inorganic Chimica Acta* 32: 29–32.
- Selwood PW (1956) *Magnetochemistry*. New York: Intersciences.
- Sinn E (1993) Inorganic chemistry of iron. In: Silver J (ed.) *Chemistry of Iron*. Glasgow: Blackie Academic and Professional (Chapman and Hall).



# MALDI

See **MASS SPECTROMETRY: Matrix-Assisted Laser Desorption/Ionization**

# MARINE

See **WATER ANALYSIS: Seawater – Organic Compounds; Seawater – Dissolved Organic Carbon; Seawater – Inorganic Compounds**

# MASS SPECTROMETRY

Contents

**Overview**

**Principles**

**Ionization Methods Overview**

**Electron Impact and Chemical Ionization**

**Atmospheric Pressure Ionization Techniques**

**Electrospray**

**Liquid Secondary Ion Mass Spectrometry**

**Matrix-Assisted Laser Desorption Ionization**

**Mass Separation**

**Ion Traps**

**Time-of-Flight**

**Selected Ion Monitoring**

**Multidimensional**

**Stable Isotope Ratio**

**Pyrolysis**

**Archaeological Applications**

**Clinical Applications**

**Environmental Applications**

**Food Applications**

**Forensic Applications**

**Gas Analysis**

**Peptides and Proteins**

**Polymerase Chain Reaction Products**

## Overview

**R Sleeman and J F Carter**, Mass Spec Analytical Ltd.,  
Bristol, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Mass spectrometry (MS) is an analytical technique that is used to identify unknown compounds,

quantify known materials, and elucidate the structural and physical properties of ions. The technique is associated with very high levels of specificity and sensitivity, somewhat offset by a high degree of technical complexity. Analyses can be accomplished with minute quantities – sometimes less than picogram ( $10^{-12}$  g) amounts of material. MS is highly suited to the identification of individual components in extremely complex mixtures.

The history of MS began when the existence of electrons and ‘positive rays’ was demonstrated by J.J. Thomson in the early part of the twentieth century. Thomson suggested that the technique could be used to perform chemical analysis, but this was not realized for several decades. MS was initially used to determine the relative abundances of gaseous isotopes and to measure their ‘exact masses’, i.e., atomic masses with a precision of 1 part in  $10^6$  or better. These fundamental measurements led to developments in a wide range of physical sciences.

## Principles

A mass spectrometer is an instrument that separates charged atoms or molecules according to their mass-to-charge ( $m/z$ ) ratio. The unit of mass used is the unified atomic mass unit (symbol u), defined as 1/12 of the mass of a single atom of the isotope of carbon-12 ( $^{12}\text{C}$ ). The term Dalton (Da) has become widely accepted in MS as the unit to represent atomic mass. The charge on an ion is denoted by the integer number ( $z$ ) of the fundamental unit of charge, the magnitude of the charge of an electron. In many cases, the ions encountered have just one charge ( $z = 1$ ) so the  $m/z$  value is numerically equal to the molecular (ionic) mass in Daltons. Mass spectrometrists sometimes refer to the ‘mass of an ion’ when they really mean the  $m/z$  ratio. The name Thomson (Th) has recently been proposed as a unit of mass-to-charge ratio in an attempt to alleviate the confusion caused by the increasing importance of multiply charged ions in MS, but does not yet enjoy widespread acceptance.

Formation of gas-phase sample ions is an essential prerequisite to mass selection and detection. Ions are produced and are accelerated toward an analyzer region that is maintained under vacuum. Either positive or negative ions are selected for analysis. In many cases a high proportion of the molecules do not ionize and are simply pumped away and not detected.

The sample, which may be a solid, liquid, or vapor, enters the vacuum chamber. Early mass spectrometers required a sample to be gaseous, but the applicability of MS has been extended to include samples in liquid solutions or embedded in a solid matrix.

The analyte may already exist as ions in solution, or it may be ionized by a variety of methods within the ion source.

Gas-phase ions are separated in the mass analyzer according to their mass-to-charge ( $m/z$ ) ratios and impinge on a detector, where the ion flux is converted to a proportional electrical current. A data system records the magnitude of these electrical signals as a function of  $m/z$  and converts this information into a mass spectrum, a graph of ion intensity as a function of mass-to-charge ratio, often depicted as simple histograms.

The mass scale is calibrated by introducing a reference compound that yields a well-characterized mass spectrum comprising known masses at suitable intervals. A range of calibrants is available for various techniques and applications.

## Compound Identification

The mass spectrum will typically establish the molecular weight and structure of the compound being analyzed. Mass spectra recorded under controlled conditions are highly reproducible such that the spectrum derived from an unknown may reasonably be compared to that of an authenticated standard. When combined with a chromatographic technique mass spectral identification is often regarded as providing unequivocal characterization of a chemical compound.

Large databases or libraries of spectra are commercially available to assist in compound identification in a range of applications, although currently these are largely restricted to electron ionization (EI) spectra.

Since the rules governing the fragmentation of ions in the gas phase are well established, it is also possible to elucidate the structure of an unknown compound solely from its mass spectrum.

## Quantification

Quantification may be achieved by comparing the response of the mass spectrometer from an analyte of interest to the response obtained from the introduction of a known amount of a standard. Standards are typically a closely related substance or may be chemically identical but synthesized by substituting an isotope of one of the elements.

Since the identity of the analyte is already known and the requirement is to measure how much is present, it is not necessary to record the full mass spectrum. Selected ion monitoring (SIM) is often used in such circumstances, where the mass spectrometer (generally coupled with a chromatographic technique) monitors only the sample ion and the

equivalent ion for a suitable internal standard. In this way, very selective conditions for quantifying a known sample can be devised.

### Resolution

The ability of a mass spectrometer to distinguish between ions of different mass-to-charge ratio values is termed as resolution. There are many definitions, depending on specific applications and instrument types being used. Traditionally, multisector mass spectrometers are considered as high-resolution instruments and quadrupole instruments as medium resolution.

A typical multisector instrument can be set to resolve ions at the 20 parts per million level, that is to say an ion of mass-to-charge ratio 500.00 can be separated from an ion of  $m/z$  500.01. Time-of-flight (TOF) instruments can also attain or exceed this resolution.

Exact mass measurement can aid in determining chemical composition. Every isotope (except carbon-12 which is assigned exactly 12.000 00 Da) has a unique, noninteger mass. Exact mass measurement thus allows determination of chemical composition. With sufficient resolution it is possible to distinguish between carbon monoxide (CO, 27.995 Da) and nitrogen ( $N_2$ , 28.006 Da) by exact mass measurement.

## Mass Analyzers

### Sector Field Instruments

Magnetic sectors deflect the trajectories of ions into circular paths of radii that depend on the

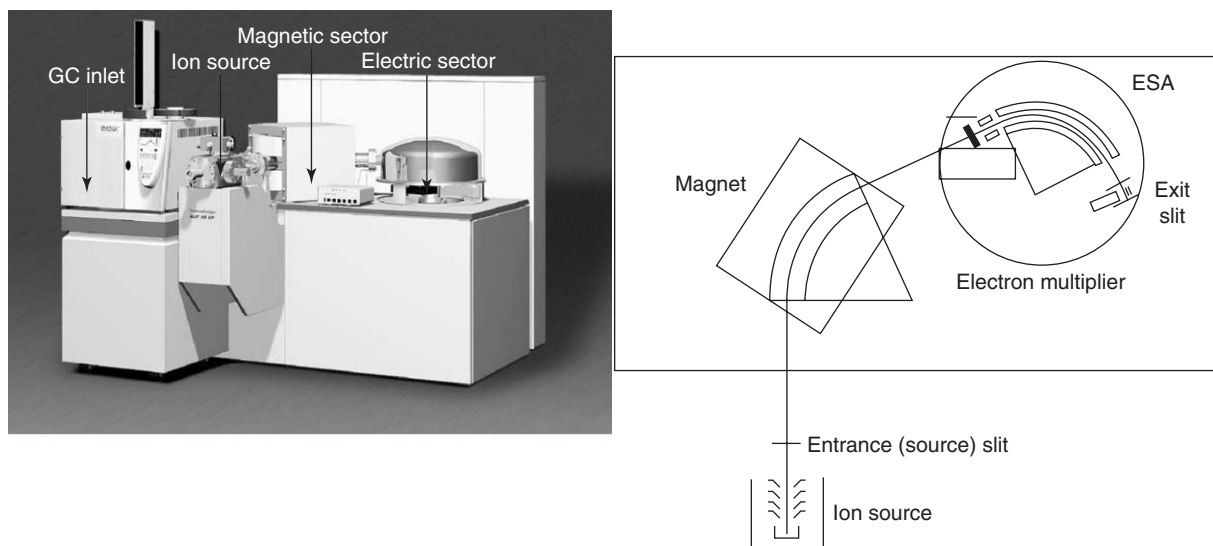
momentum-to-charge ratios of the ions. Ions of larger  $m/z$  values follow larger radius paths than ions of smaller  $m/z$  values, so ions of differing  $m/z$  values are dispersed in space. By changing the ion trajectories through variations of the magnetic field strength, ions of different mass-to-charge ratios can be focused onto a detector.

Double focusing sector field instruments incorporate a combination of electromagnetic fields ( $B$ ) and electric fields ( $E$ ).

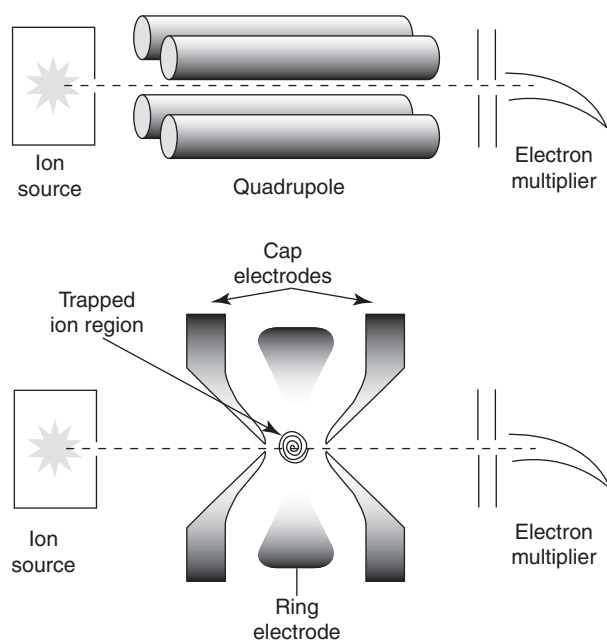
A common configuration ('forward geometry') for a sector instrument is the geometry in which a magnetic sector follows an electric sector analyzer (ESA). This 'double focusing' combination of energy focusing and 'angular' or 'directional' focusing and energy focusing provides mass resolution high enough to separate ions of the same nominal mass but different chemical formulas. In 'reverse geometry,' the magnetic sector precedes the electric sector. **Figure 1** shows a reverse-geometry mass spectrometer coupled to a gas chromatograph (GC). Such instruments are commonly used for the highly specific detection of environmental contaminants such as dioxins or performance-enhancing drugs in athletes.

### Quadrupole Filters

Quadrupole mass filters consist of four parallel rods, illustrated in **Figure 2**. In such instruments, mass selection depends on ion motion resulting from simultaneously applied DC and RF electric fields. Scanning is accomplished by systematically changing the field strengths, thereby changing the  $m/z$  value that is transmitted through the analyzer. Quadrupole mass



**Figure 1** Photograph and schematic diagram of a modern 'reverse geometry' mass spectrometer coupled to a GC. (Reproduced with permission from Thermo Electron (Bremen).)



**Figure 2** Schematic diagram of quadrupole and ion trap mass analyzers.

spectrometers provide lower resolution than double focusing instruments but are less costly. The high scan speeds of quadrupole mass filters render them highly suited for use in combination with chromatographic inlet systems.

### Ion Traps

Ion trap mass spectrometers operate on a principle similar to a quadrupole mass filter. However, it does not operate as a filter; the ions are stored for subsequent experiments and analysis. Electric fields are applied to electrodes arranged as a ring electrode in the middle with cap electrodes on each end. Conceptually, an ion trap can be considered as a conventional quadrupole folded on itself to form a closed loop. A comparison with a quadrupole mass analyzer is shown in **Figure 2**. Within a selected range of  $m/z$  ratios determined by the applied voltages, the device traps ions in the space bounded by the electrodes. A mass spectrum is produced by scanning the applied RF voltages to eject ions sequentially of increasing  $m/z$  ratio through an end cap opening for detection.

### Fourier Transform Mass Spectrometry

In a Fourier transform ion cyclotron resonance (FT-ICR) spectrometer, ions are trapped electrostatically within a cell in a constant magnetic field. An orbital ('cyclotron') motion is induced by the application of a pulse between the exciter (or emitter) plates. The orbiting ions generate a faint signal in the detector (or receiver) plates of the cell. The frequency

of the signal from each ion is equal to its orbital frequency, which in turn is inversely related to its  $m/z$  value. The signal intensity of each frequency is proportional to the number of ions having that  $m/z$  value. The signal is amplified and all the frequency components are determined, yielding the mass spectrum. Since the pressure in the cell is very low, the ion orbital motion can be maintained over many cycles and the frequency can be measured with very high precision. FT-ICR instruments can therefore be used to generate exceptionally high resolution spectra with great mass accuracy.

### Time-of-Flight Mass Spectrometry

TOF mass analyzers separate ions by virtue of their different flight times over a known distance. Pulses of ions are ejected from a source and accelerated so that ions of like charge have equal kinetic energy. They are then directed into a flight tube where lower mass ions have greater velocities and shorter flight times. The travel time from source to detector can be transformed to the  $m/z$  value. All ion masses are measured for each pulse, so TOF mass spectrometers offer high sensitivity as well as very rapid scanning. They can provide mass data for very high-mass biomolecules.

### Ionization Methods

Compounds are converted into gas-phase molecules either before or during the charging or ionization process, which takes place in the ion source.

Many types of ionization mode are available; the type of compound to be analyzed and the specific information required determine which ionization mode is the most suitable. The ionized molecule may subsequently fragment, producing ions of lower mass than the original precursor molecule. These fragment ions are determined by the structure of the original molecule.

### Electron Ionization

In the commonly used EI source (earlier referred to as 'electron impact'), ions are generated by bombarding the gaseous sample molecules with a beam of energetic electrons. EI produces a mixture of positive and negative ions, as well as neutral species. Positive-ion EI mass spectra are more commonly recorded because these ions form more readily.

The energy of the electrons (typically 70 eV) is generally much greater than that of the bonds that hold the molecule together. Ionization by electrons is a highly energetic or 'hard' process that may lead to extensive fragmentation that leaves very little or no trace of a molecular ion. Because molecular mass and

structure are not easily determined in the absence of a molecular ion, lower energy or 'soft' ionization techniques have been developed based on chemical and desorption ionization processes.

### Chemical Ionization

Chemical ionization (CI) source is very similar to the EI source but the beam of electrons is used to create a plasma of ionized reagent gas (e.g., isobutane, methane, ammonia).

Transfer of a proton to a sample molecule  $M$ , from an ionized reagent gas such as methane in the form of  $\text{CH}_5^+$ , yields the  $[M + H]^+$  positive ion. This process is less energetic than EI and generally produces less fragmentation. The fragmentation patterns are not necessarily the same as those of molecular ions,  $M^{+\bullet}$ .

Negative ions can also be produced under CI conditions. Transfer of a proton from  $M$  to reagent gas or ions can leave  $[M - H]^-$ , a negatively charged sample ion. Addition of an electron to  $M$ , a process facilitated by collisionally moderating the energy of electrons generated in the source, can yield an intense  $M^-$  ion. Such ions are often the only ion generated and can be used to detect species with great sensitivity.

### Desorption Ionization

Samples may be desorbed and ionized by an impact process that involves bombardment of the sample with high-velocity atoms, ions, fission fragments, or photons of relatively high energy. The impact deposits energy into the sample, either directly or via a matrix, and leads to both sample molecule transfer into the gas phase and ionization.

Field desorption (FD) is perhaps the simplest technique, the sample is coated as a thin film onto a special filament placed within a very high intensity electric field. Ions created by field-induced removal of an electron from the molecule are extracted into the mass spectrometer. Field ionization (FI) is the equivalent process whereby gas-phase molecules are ionized by a high electric potential. These techniques are sometimes applied to relatively large, polymeric molecules but are not commonplace.

### Fast Atom Bombardment/Liquid SIMS

The techniques of fast-atom bombardment (FAB)/liquid secondary ionization (LSIMS), developed in the early 1980s, revolutionized the range of compounds amenable to analysis by MS and opened up the field to many areas of biomedical research. Although now considered insensitive by comparison with more recently introduced ionization modes, FAB still has a role as a rapid, reliable, and robust

technique for samples where quantity and purity are not a problem. The sample is first dissolved in a liquid matrix. This is typically a viscous, low vapor pressure liquid. A few microliters of this liquid are placed on a small metal target at the end of a probe that is inserted into the mass spectrometer. The liquid surface is then bombarded with a beam of high kinetic energy atoms (xenon) or ions (cesium). Molecules sputtered from the surface enter the gas phase and ionize, either by protonation, deprotonation, or adduction. The resulting ions tend to be stable and exhibit little fragmentation.

### Matrix-Assisted Laser Desorption Ionization

Unlike FAB/LSIMS, matrix-assisted laser desorption ionization (MALDI) uses a crystalline, rather than liquid, matrix, and a beam of photons, rather than atoms or ions. The net result is a dramatic increase in both sensitivity and mass range of compounds that may be analyzed. The sample is dissolved in a matrix and is allowed to crystallize on a stainless-steel target. The target is then inserted into the mass spectrometer and the surface bombarded with a pulsed laser beam. Molecules are desorbed from the surface and ionize, usually by protonation or deprotonation. Any fragment or multiply charged ions are generally of low abundance in this ionization mode. The pulsed nature of the laser excitation renders this technique compatible with TOF, and the combined technique enjoys an almost limitless mass range.

### Atmospheric Pressure Ionization Techniques

Atmospheric pressure ionization (API) techniques encompass a range of techniques in which ionization occurs external to the mass spectrometer vacuum. Ionization can be achieved by a variety of methods, including photoionization, corona discharge at the tip of a needle, or by the use of radioisotopes such as  $^{63}\text{Ni}$ .

Atmospheric pressure chemical ionization (APCI) is a simple and robust technique routinely used to interface the eluent from a high-performance liquid chromatography (HPLC) to a mass spectrometer. The liquid stream passes through a heated nebulizer into a corona discharge region. Analyte molecules are ionized and extracted into the mass analyzer.

### Electrospray

Electrospray (ESI) is another example of an API technique.

The sample is dissolved in a mobile phase and pumped through a fine stainless-steel capillary maintained at high potential. This creates an electrostatic spray of multiply charged droplets containing the



sample. At higher solvent flow rates, heat and drying gas may be needed to increase the rate of droplet evaporation. This technique is sometimes referred to as pneumatically assisted ESI or Ionspray.

After desolvation and subsequent charge concentration, gas-phase ions are produced and propelled toward the high vacuum mass analyzer. ESI is considered to be one of the 'softest' ionization techniques available, i.e., little energy is transferred to the molecule other than that required for ionization. Thus, protonated, deprotonated, or cationized molecules that undergo very little fragmentation are generated, even from highly polar, thermally labile molecules.

ESI can impart many charges ( $z$ ), usually in the form of protons, to amenable large molecules such as proteins, and thus compounds with molecular masses in the tens of thousands can be analyzed with mass spectrometers with  $m/z$  ranges of a few thousand. Molecular mass can often be determined to a precision in the order of one part in 10 000 or better. ESI is particularly compatible with liquid separation methods and has become a widely used method in biological and pharmaceutical analysis, where identification is achieved through deconvolution of the envelope of peaks formed with multiple charge states.

### Choice of Ionization Technique

EI and CI are generally the techniques of choice for small ( $<800$  Da), volatile, thermally stable compounds.

CI tends to give molecular weight information and EI, with the greater fragmentation, provides structural information.

FAB/LSIMS is useful for larger ( $\leq 5000$  Da) involatile, polar, thermally unstable molecules, such as peptides, small proteins, and other biopolymers. However, this technique has now largely been superseded by ESI and MALDI.

MALDI is suitable for similar compounds to those amenable to FAB, but affords much greater sensitivity. Biopolymers with molecular weights above 300 000 Da have been successfully analyzed.

ESI is suitable for similar compounds to MALDI, with possibly a slightly reduced sensitivity and mass range. The tendency to produce multiply charged ions brings the mass-to-charge ratios of high molecular weight proteins well within the range of inexpensive mass spectrometers. This has fuelled an explosion in biochemical applications of MS and has spawned the developing fields of proteomics, genomics, and metabonomics.

ESI is frequently interfaced with chromatographic techniques such as HPLC, capillary zone electrophoresis (CZE), and capillary electrochromatography (CEC).

## Sample Introduction Techniques

### Probe Inlets

A direct insertion probe may be used for reasonably pure volatile solids. The sample is loaded into a quartz tube on the tip of a rod that is inserted into the evacuated source region. The sample is then evaporated or sublimed into the gas phase, usually by heating. The gaseous molecules are then ionized (often with accompanying fragmentation) and the ions are mass analyzed. In some techniques, volatilization and ionization occur at the same time.

### Septum Inlets

Heated reservoir septum inlets may be used for pure gases or volatile liquids, comprising a heated reservoir with a small restriction 'bleed' into the ion source. The sample is injected into the reservoir through a septum. This method is commonly used to introduce reference materials for calibration.

### Chromatographic Techniques

To obtain the mass spectrum of a single constituent of a mixture, the individual components often need to be separated prior to analysis. Separation is necessary for unambiguous identification because two compounds present in the source region simultaneously create a mixed spectrum and even simple compounds can generate many fragment ions. The historical combination of GC and MS (GC-MS) allows compounds in the vapor phase to enter the mass spectrometer so that the components of mixtures can be detected and analyzed sequentially.

The challenge in interfacing a mass spectrometer to a separation system like a gas or liquid chromatograph is maintaining the required vacuum in the mass spectrometer while introducing flow from the chromatograph. Interfaces that restrict or reduce the gas flow into the mass spectrometer (e.g., flow splitters or devices that differentially remove carrier gas from the GC effluent) initially made the combination of GC and MS an extremely widely used technique. The low gas flows typical of capillary GC now permit direct connection to mass spectrometers.

More recently, liquid chromatography (LC), supercritical fluid chromatography (SFC), and CZE devices connected to mass spectrometers have been used to separate components of complex mixtures prior to mass analysis. When vaporized, the solvent from an LC represents a volume of 100–1000 times greater than that of a carrier gas used in GC. Interfaces developed commercially have solved the problem of eliminating this gas load by using combinations of heating and pumping, sometimes with

the assistance of a drying gas stream. The inlets for higher flow rates (as in analytical LC) employed in LC/MS systems in routine use are primarily APCI and ESI.

Particle beam interfaces, thermospray, and 'dynamic FAB' have also been used as LC continuous-flow injection techniques, but these have largely been superseded.

For GC-MS, LC-MS, or other combinations, the data consist of a series of mass spectra acquired sequentially in time. To generate this information, the mass spectrometer scans the mass range repetitively during the chromatographic run. The intensities of all the ions in each spectrum can be summed, and this sum plotted as a function of chromatographic retention time to give a total ion chromatogram (TIC) whose appearance is similar to the output of a conventional chromatographic detector. Each peak in the TIC represents an eluting compound that can be identified by interpretation of the mass spectra recorded for the peak. The intensity at a single mass-to-charge ratio over the course of a chromatographic run can be displayed to yield a selected ion current profile or mass chromatogram. This technique can be used to find components of interest in a complex mixture without having to examine each individual mass spectrum.

## Mass Spectrometric Techniques

### Selected Ion Monitoring

SIM is frequently used for the quantitative determination of specific analytes by MS, usually in combination with a chromatographic separation. The mass spectrometer is used to monitor a limited number of ions characteristic of target compounds, rather than to acquire a complete spectrum. The effect is that the instrument spends a greater time recording ions from the analytes of interest with a resulting increase in both sensitivity and selectivity. This is a very sensitive technique and for some compounds it is possible to detect at the femtogram ( $10^{-15}$  g) level.

### Tandem or Multistage Mass Spectrometry (MS/MS, MS<sup>n</sup>)

Tandem MS is used to provide more information than can be afforded by a single mass spectrometer and is widely used for screening complex matrices such as blood and urine. Analysis is achieved, in effect, by performing two stages of SIM. The first mass spectrometer is set to transmit the 'precursor' ion of interest into a region where fragmentation occurs. One of the 'product' ions is monitored by a second mass spectrometer. Selection of an appropriate internal

standard and switching between the gas phase transitions can lead to very high specificity. This technique is known as selected reaction monitoring (SRM), and is frequently used for quantification. Fragmentation is usually achieved in a collision cell pressurized with an inert gas such as argon. Collision of ions with atoms in the cell produces fragments by a process known as collision induced dissociation (CID). Other approaches have been used to cause fragmentation, such as lasers, electron beams, and surface collisions.

In cases where 'soft' ionization techniques are used, the molecular weight of the sample may be observed but the lack of in-source fragmentation means that little structural information is available. A product ion mass spectrum acquired with a tandem mass spectrometer can yield this structurally significant information.

In the technique of precursor ion scanning the second mass spectrometer is set statically to transmit product ions of only one selected mass-to-charge ratio. This mass is monitored continuously whilst the first mass spectrometer is scanned. A signal will be detected only when a precursor ion fragments to form the product ion that is monitored. This technique is often used to screen for compounds of related structure, such as the metabolites of a known drug.

Another tandem screening method is known as constant neutral loss scanning. Here, both mass spectrometers are scanned simultaneously but are offset corresponding to the difference between precursor and product ion masses. A signal only appears when a precursor ion yields a product ion with the mass difference selected. This technique can be used to screen for compounds that contain a specific structural feature that yields a common fragmentation process.

Tandem MS can be performed using sector, quadrupole, and TOF instruments. However, each stage of mass analysis requires a separate mass analyzer. Different mass analyzers are often combined to form tandem instruments for specific applications, e.g., Q-TOF.

Ion trap or ICR mass spectrometers permit MS/MS product ion experiments to be conducted sequentially in time within a single mass analyzer. A number of sequential experiments, termed MS<sup>n</sup>, may be performed.

Modern TOF mass spectrometers incorporate a reflectron unit and the facility to analyze what are known as postsource decay ions. The spectra produced are similar to product ion spectra and can be enhanced by the inclusion of a collision cell.

### Stable Isotope Ratio Mass Spectrometry

Although often presumed to be constant, natural isotope abundance ratios show significant and

characteristic variations when measured very precisely. In stable isotope ratio mass spectrometry (IRMS), element isotope ratios are determined very accurately and precisely. Typically, single focusing magnetic sector mass spectrometers with fixed multiple detectors (one per isotopomer) are used. Complex compounds are reduced to simple molecules prior to measurement, for example, organic compounds are combusted to CO<sub>2</sub>, H<sub>2</sub>O, and N<sub>2</sub>. Isotope ratio measurements are useful in a wide range of applications, for example, metabolic studies using isotopically enriched elements as tracers; climate studies using measurements of temperature-dependent oxygen and carbon isotope ratios in foraminifera; rock age dating using radiogenic isotopes of elements such as lead, neodymium, or strontium; and source determinations using carbon isotope ratios (for example, to discriminate between naturally occurring substances and petroleum-based synthetic materials).

### Pyrolysis Mass Spectrometry

Pyrolysis is the thermal degradation of complex material in an inert atmosphere or a vacuum. Molecules cleave into smaller, volatile fragments called pyrolysate. In pyrolysis MS (PyMS), the pyrolysate is directly analyzed by MS to produce a chemical profile or fingerprint of the complex material analyzed. The development of PyMS was largely driven by its applicability to the characterization of microorganisms and has now largely been supplanted by the application of MALDI in this field. In contrast, pyrolysis GC/MS (Py-GC/MS) still finds numerous applications in the analysis of complex synthetic and biological polymers.

### Elemental Mass Spectrometry

Elemental MS is applied mostly to inorganic materials, to determine the elemental composition of a sample rather than the structural identities of its

chemical constituents. Elemental MS provides quantitative information about the concentrations of those elements. The decomposition of the sample into its constituent atoms and ionization of those atoms occurs in a specially designed source. The ion source used in elemental MS is ordinarily an atmospheric-pressure discharge such as inductively coupled plasma (ICP) or a moderate-power device such as glow discharge (GD). The resulting atomic-ion beam is then separated by a mass analyzer and the signal used to determine the sample composition. With an ICP employed as an ion source, solution detection limits down to the parts per trillion level are possible in favorable cases, while with the glow-discharge source, solid metal samples can be analyzed directly and their elemental composition determined over a million-fold range of concentrations. Isotopic information is readily available.

*See also:* **Gas Chromatography:** Mass Spectrometry. **Liquid Chromatography:** Mass Spectrometry. **Mass Spectrometry:** Ionization Methods Overview; Mass Separation; Stable Isotope Ratio.

### Further Reading

- American Society for Mass Spectrometry website, [www.asms.org](http://www.asms.org)
- Armentrout PB (2003) In: Gross ML and Caprioli RM (eds.) *Encyclopedia of Mass Spectrometry, vol.1: Theory and Ion Chemistry*. Amsterdam: Elsevier.
- Ashcroft AE (1997) *Ionization Methods in Organic Mass Spectrometry*. Cambridge: Royal Society of Chemistry.
- British Mass Spectrometry Society website, [www.bmss.org.uk](http://www.bmss.org.uk)
- Chapman JR (1993) *Practical Organic Mass Spectrometry*, 2nd edn. New York: Wiley.
- de Hoffmann E and Stroobant V (2001) *Mass Spectrometry Principles and Applications*. Chichester: Wiley.
- Sparkman OD (2000) *Mass spectrometry desk reference*. Pittsburgh, PA: Global View Publishing.

## Principles

**K J Welham**, University of Hull, Hull, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by A Roberts, pp. 2787–2795, © 1995, Elsevier Ltd.

### Introduction

The mass spectrometer is an instrument capable of producing a beam of ions from a sample under

investigation, separating these ions according to their mass-to-charge ratio and recording the relative abundances of the separated ion species as a mass spectrum. Many types of mass spectrometers have been developed to fulfill the above definition. The widely differing nature and physical state of sample materials has resulted in a variety of techniques for sampling, ionization, ion storage, mass separation, and ion recording. This article on the basic principles of mass spectrometry (MS) will cover the historical

characteristic variations when measured very precisely. In stable isotope ratio mass spectrometry (IRMS), element isotope ratios are determined very accurately and precisely. Typically, single focusing magnetic sector mass spectrometers with fixed multiple detectors (one per isotopomer) are used. Complex compounds are reduced to simple molecules prior to measurement, for example, organic compounds are combusted to CO<sub>2</sub>, H<sub>2</sub>O, and N<sub>2</sub>. Isotope ratio measurements are useful in a wide range of applications, for example, metabolic studies using isotopically enriched elements as tracers; climate studies using measurements of temperature-dependent oxygen and carbon isotope ratios in foraminifera; rock age dating using radiogenic isotopes of elements such as lead, neodymium, or strontium; and source determinations using carbon isotope ratios (for example, to discriminate between naturally occurring substances and petroleum-based synthetic materials).

### Pyrolysis Mass Spectrometry

Pyrolysis is the thermal degradation of complex material in an inert atmosphere or a vacuum. Molecules cleave into smaller, volatile fragments called pyrolysate. In pyrolysis MS (PyMS), the pyrolysate is directly analyzed by MS to produce a chemical profile or fingerprint of the complex material analyzed. The development of PyMS was largely driven by its applicability to the characterization of microorganisms and has now largely been supplanted by the application of MALDI in this field. In contrast, pyrolysis GC/MS (Py-GC/MS) still finds numerous applications in the analysis of complex synthetic and biological polymers.

### Elemental Mass Spectrometry

Elemental MS is applied mostly to inorganic materials, to determine the elemental composition of a sample rather than the structural identities of its

chemical constituents. Elemental MS provides quantitative information about the concentrations of those elements. The decomposition of the sample into its constituent atoms and ionization of those atoms occurs in a specially designed source. The ion source used in elemental MS is ordinarily an atmospheric-pressure discharge such as inductively coupled plasma (ICP) or a moderate-power device such as glow discharge (GD). The resulting atomic-ion beam is then separated by a mass analyzer and the signal used to determine the sample composition. With an ICP employed as an ion source, solution detection limits down to the parts per trillion level are possible in favorable cases, while with the glow-discharge source, solid metal samples can be analyzed directly and their elemental composition determined over a million-fold range of concentrations. Isotopic information is readily available.

*See also:* **Gas Chromatography:** Mass Spectrometry. **Liquid Chromatography:** Mass Spectrometry. **Mass Spectrometry:** Ionization Methods Overview; Mass Separation; Stable Isotope Ratio.

### Further Reading

- American Society for Mass Spectrometry website, [www.asms.org](http://www.asms.org)
- Armentrout PB (2003) In: Gross ML and Caprioli RM (eds.) *Encyclopedia of Mass Spectrometry, vol.1: Theory and Ion Chemistry*. Amsterdam: Elsevier.
- Ashcroft AE (1997) *Ionization Methods in Organic Mass Spectrometry*. Cambridge: Royal Society of Chemistry.
- British Mass Spectrometry Society website, [www.bmss.org.uk](http://www.bmss.org.uk)
- Chapman JR (1993) *Practical Organic Mass Spectrometry*, 2nd edn. New York: Wiley.
- de Hoffmann E and Stroobant V (2001) *Mass Spectrometry Principles and Applications*. Chichester: Wiley.
- Sparkman OD (2000) *Mass spectrometry desk reference*. Pittsburgh, PA: Global View Publishing.

## Principles

**K J Welham**, University of Hull, Hull, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by A Roberts, pp. 2787–2795, © 1995, Elsevier Ltd.

### Introduction

The mass spectrometer is an instrument capable of producing a beam of ions from a sample under

investigation, separating these ions according to their mass-to-charge ratio and recording the relative abundances of the separated ion species as a mass spectrum. Many types of mass spectrometers have been developed to fulfill the above definition. The widely differing nature and physical state of sample materials has resulted in a variety of techniques for sampling, ionization, ion storage, mass separation, and ion recording. This article on the basic principles of mass spectrometry (MS) will cover the historical

development of the instrumentation and introduce the topics of ionization methods and fragmentation processes before concluding with a brief summary of the approach to interpreting data.

## Instrumental Development

The origins of MS are to be found in studies on the deflection of charged particles in magnetic fields by some eminent nineteenth century experimentalists. The theory of this effect was developed by several researchers with the Nobel Laureates JJ Thompson, KF Braun, and W Wren being the most prominent. The first spectrometer exhibiting velocity focusing was built by FW Aston in 1919. This machine (the mass spectrograph) recorded the separated ion beams on a photographic plate and became associated with the accurate determination of mass. At the same time, AJ Dempster designed a 180° magnetic deflection machine (the mass spectrometer) employing directional focusing and electrical detection of the ion beam. The first spectrometer employing directional and velocity focusing (double focusing) in the same instrument was demonstrated in 1937 (J Mattauch and R Herzog) and the concept was repeated in the Nier-Johnson machine of 1953, which employed an alternative ion path geometry. It is this latter design that has been developed into the modern chemical analysis spectrometer incorporating high-resolution accurate mass determination. The availability of commercial high-performance chemical analysis mass spectrometers generated a demand for ionization techniques capable of producing positive and negative molecular ions from compounds that provide fragment ion spectra (i.e., no molecular ions) under electron bombardment. The techniques developed include photoionization (PI), fast atom or ion bombardment (FAB), field desorption (FD) or field ionization (FI), plasma desorption, matrix-assisted laser desorption ionization (MALDI), chemical ionization (CI), atmospheric pressure chemical ionization (APCI), and electrospray (ESP). The introduction of samples to the spectrometer, conventionally achieved by gaseous leak for gas and volatile liquid samples and by direct insertion probe for solids and involatile liquids, was expanded by the successful interfacing of a range of chromatographic techniques to the spectrometer. The availability of direct mixture separation via gas, liquid, supercritical, and thin-layer chromatography has removed the necessity for difficult and time-consuming chemical or physical separation outside the spectrometer. At present no universal interface is available for all chromatograph-spectrometer combinations, although the eluent from capillary gas chromatography

can be directly introduced into the mass spectrometer ion source as can the eluent from high-performance liquid chromatography when using APCI or ESP ionization. The two-dimensional separation supplied by chromatography-mass spectrometry is also intrinsic to the linked MS-MS technique. In this approach, normally preceded by soft ionization, the pseudomolecular ion derived from an unknown compound is selected from a mixture of ions by a magnetic field ( $B$ ) combined with an electric sector field ( $E$ ) in a double focusing instrument, or quadrupole field ( $Q$ ), and is passed through a cell containing a high pressure of collision gas, normally helium or argon. In this cell collision-activated decomposition to characteristic daughter (fragment) ions occurs. These ions have the same velocity but different mass and therefore different energies, and thus may be separated by a variety of analyzer types mounted after the collision cell. This area has seen much recent growth and most possible combinations of mass analyzers have been investigated. The conventional mass separation of ion beams by magnetic and electrostatic fields has been extended by the introduction of radio frequency (RF) fields for mass separation in quadrupole mass spectrometers. This technique has itself been further developed in the QUISTOR (quadrupole ion storage) or ion trap. Another technique employing the trapping principle is ion cyclotron resonance Fourier transform mass spectrometry, one of the key features of which is the simultaneous detection of all ions. Coupled with pulsed laser ionization this has proved useful in the analysis of sensitive high-mass molecules of biological origin and is capable of unrivaled resolution at lower mass. Time-of-flight (TOF) spectrometers in which ion separation is achieved in time rather than in space are also particularly amenable to pulsed ionization techniques. Current TOF machines do not need a magnetic focusing field for mass separation and are therefore linear in design and light in weight and have consequently found application in space research. When used with an RF biasing field they also find employment in isotopic ratio measurements or the collection of single isotopes in the nuclear industry. Due to the potentially unlimited mass range parent ions of heavy biological molecules (up to 1 000 000 Da) can also be detected and MALDI-TOF-MS of biological species has become a very important technique. The potentially unlimited mass range of the TOF spectrometer is to some extent offset by the relatively high (up to 20 000) but still limited resolution available from modern instruments. The detection of the separated ion beam on photographic plates is still practiced in spark source spectrometers for the analysis of, among others, high-melting point inorganic samples. This is a



time-consuming process requiring the production of photoplates of standard materials for quantitative work and a significant time period is necessary for plate development. Electrical detection, originally achieved with the Faraday cage or hollow cylinder and still used for precision isotope studies, has been replaced by the high-gain electron multiplier or photomultiplier. Both the discrete dynode and continuous dynode multiplier have successfully lowered the minimum detectable ion current to the  $10^{-18}$  A level, resulting in typical minimum detection levels in the picomole range for many organic compounds. Channel electron multiplier/photodiode arrays enable the counting of single ions and the simultaneous detection of ions in a broad mass range leaving the separation field.

Almost all the major advances evident in the present generation of mass spectrometers can be traced to concurrent advances in semiconductor and data processing technology. For the first 50 years of spectrometer development, all control and handling was achieved by manual means. Offline data handling of the digitized ion current was the first step toward computer control, followed by online real-time processing. Control of the mass separation region then became available producing the capability for single ion monitoring, MS-MS, and linked scan functions. Today, the complete spectrometer comes under precise digital control and commercial 'user-friendly' machines are available for a variety of applications requiring a minimum knowledge of MS by the user.

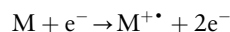
Computer library searching of spectral databases is routinely available. The database is usually a component part of the spectrometer although the search may be undertaken remotely. Several attempts have been made to develop artificial intelligence systems for direct spectral interpretation, but to date these have met with limited success. Advances in computer control have allowed multiexperiment analysis in which the spectrometer will follow a set of experiments sequentially while automatically adjusting operating parameters as directed by the results of the preceding experiment. Further advances in this area are anticipated.

## Ionization Processes

### Electron Bombardment

The first step in the production of a spectrum of ions from a gaseous molecular sample is the removal or addition of one or more electrons from the molecule, resulting in the acquisition of charge with negligible change in mass.

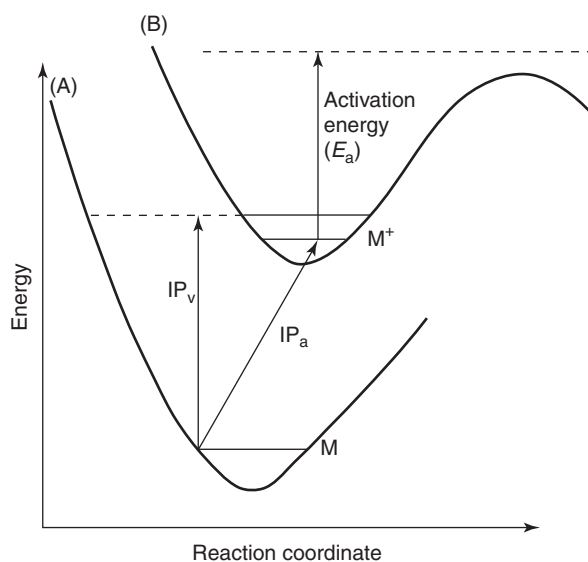
**Positive ion formation** The formation of a positive ion is normally expressed as



where  $M^{+\bullet}$  is known as the molecular ion. The molecular ion is the heaviest ion in the spectrum that can be produced by unimolecular reaction and is of considerable importance as it provides the relative molecular mass of the sample.

The probability of a bombarding electron producing a positive ion is a function of the electron's kinetic energy and a plot of this probability as a function of energy constitutes an ionization efficiency curve. Such curves show that a threshold energy exists (the appearance potential) below which ionization cannot occur. The energy at which the singly charged molecular ion appears is referred to as the first ionization potential. The time taken for interaction between a bombarding and orbital electron is  $\sim 2.4 \times 10^{-16}$  s, in which time the ionization must have occurred. The fastest molecular vibration is C-H stretching, with a period of  $\sim 10^{-14}$  s. Hence, the atoms may be considered to be at rest during the ionization process and thus satisfy the Frank-Condon principle (Figure 1).

The minimum energy required to form molecular ions by electron bombardment is known as the 'vertical' ionization potential ( $IP_v$ ) and may exceed the minimum energy necessary to produce molecular ions in the ground state (the 'adiabatic' ionization potential  $IP_a$ ). The stability of the molecular ion determines its abundance relative to ions that are formed by its fragmentation. The time elapsing between ionization



**Figure 1** Frank-Condon diagram showing (A) vertical, and (B) adiabatic transitions. Appearance potential =  $IP_a + E_a$ .

and detection of any ion is  $\sim 10^{-5}$  s and if a molecular ion fragments in significantly less than this time it will not appear in the recorded spectrum. The rate of fragmentation is a function of the excess energy of the ion on formation; the number of molecular ions reaching the collector will therefore increase as the ionizing energy is reduced toward the ionization potential. The stability of the molecular ion will also decrease with increasing molecular size across a homologous series as a greater number of fragmentation pathways are available to larger molecules.

Addition of a stabilizing entity into a chain will dramatically change molecular ion intensity. The aromatic nucleus and other conjugate systems are particularly notable in this respect as the radical site may be dispersed by conjugation over an extended part of the ion. Stabilization is also found in compounds containing unsaturated functional groups such as carbonyl, nitro, nitrite, and sulfonyl.

**Negative ion formation** During the ionization process described above, in addition to the production of positive ions, a small number of molecules undergo addition of one or more electrons to form negative ions. At the operating ionizing energy of the analytical mass spectrometer (60–100 eV), the sensitivity of negative ion formation is several orders of magnitude less than that for positive ion production. If a neutral molecule is bombarded with electrons of increasing energy a sharp peak in ion production is observed between 0 and 10 eV. The energy at which this peak appears is determined by the electron affinity of the molecule and its narrow nature shows electron attachment to be a resonance process. An interest in analytical negative ion spectra developed when production of a sufficiently high flux of low-energy electrons became readily available. This was achieved by employing a high pressure of reagent gas as an energy moderator for energetic electrons.

Negative ions are produced either directly by electron capture or by ion–molecule reactions with anionic species.

Three methods of direct electron capture may occur:

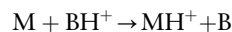
1. associative resonance capture ( $AB + e^- \rightarrow AB^{\cdot-}$ );
2. dissociative resonance capture ( $AB + e^- \rightarrow A^- + B^{\cdot}$ ); and
3. ion pair production ( $AB + e^- \rightarrow A^+ + B^- + e^-$ ).

As the electron affinities of most organic materials do not exceed 2.5 eV, the reagent gas not only acts as a means of producing thermal electrons but also provides collisional stabilization of the anions formed.

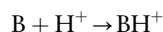
## Chemical Ionization

The process of chemical ionization in the mass spectrometer bears several similarities to the production of negative ions described above. The characteristic ionization of the sample is achieved by ionic reaction and the technique also requires a high pressure (up to 1 Torr) of reagent gas in the source. The partial pressure of sample to reagent gas does not normally exceed 0.01%. The energy transfer in the CI process is low compared to that for electron impact (EI) ionization and in general does not exceed 5 eV. The technique is therefore referred to as ‘soft ionization’ and is capable of producing extremely stable pseudomolecular ions (often protonated) and little fragmentation. Compounds that are thermally labile or unstable to EI ionization readily produce molecular ion data under CI. By varying the nature of the reagent gas to provide more energetic ionization, fragmentation data may also be obtained. The reactions between sample molecules and positively charged reagent ions fall broadly into four groups: proton transfer, charge exchange, electrophilic addition, and anion abstraction.

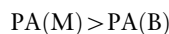
**Proton transfer** The general reaction is



It should be noted that the proton affinity (PA) for the addition



is expressed as  $-\Delta H$  for that reaction and that for the general proton exchange reaction shown above



that is, the reaction is exothermic.

Hydrogen is the most exothermic of all protonating agents ( $PA = 423 \text{ kJ mol}^{-1}$ ), followed by methane ( $PA = 536 \text{ kJ mol}^{-1}$ ). Both of these reagents, however, tend to produce excessive fragmentation. If molecular ion information is required, isobutane or ammonia ( $PA = 810$  and  $847 \text{ kJ mol}^{-1}$ , respectively) provide the necessary ‘soft’ ionization.

## Other Ionization Processes

A variety of other techniques based on the reaction of ionizing agents with sample molecules are readily available to the mass spectrometrists. FAB and secondary ion mass spectrometry (SIMS) are used to analyze nonvolatile solids and use respectively energetic atoms and ions to achieve sputtering of the sample including ionization. Laser and PI and FD/FI are also well-established techniques. Inorganic materials may be ionized by arc discharge,

inductively coupled plasma (ICP) or SIMS, or glow discharge. Perhaps two of the most important techniques today are ESP ionization and MALDI. Both these techniques are heavily utilized in the identification of proteins and other biological polymers at high molecular mass (up to 1 000 000 Da). These techniques have increased in importance as MS has increased its role in biomedical sciences and both are now the basis for the use of MS in studies of the total cellular protein complement (proteomics). A detailed theoretical discussion of the ionization processes involved in these techniques is beyond the scope of this introductory article.

## Fragmentation

The electron bombardment ionization of a molecule as described above produces an excited molecular ion, which, in attempting to gain stability, may decompose in a number of different ways by unimolecular reaction to fragment ions or neutrals. The relative abundance of the molecular ion is determined by its ability to resist decomposition, but the stability of fragment ions is dependent on the relative rates of reaction that form and destroy the ion.

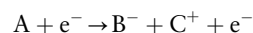
The factors that play a role in determining the fragment ions derived from a particular molecular ion are:

1. the relative stabilities of the bonds in the decomposing ion,
2. the relative stabilities of the ions and neutrals formed by various pathways, and
3. the relative probabilities of decomposition pathways involving cleavage of bonds in transient cyclic intermediates.

The reactivity of a molecular bond is normally mirrored by that bond's tendency to break down under electron impact. For example, in  $\text{CH}_2\text{Cl}-\text{CH}(\text{CH}_3)_2$ , the 1–2 bond is the most chemically labile due to the polarizability of the isopropyl group and the inductive effect of the halogen. In an analogous way, the majority of the fragment ions in the mass spectrum of this compound are produced by 1–2 cleavage. The chemical lability of allylic and benzylic bonds is also closely mirrored by their electron impact-induced cleavage. The simplest indicator of a fragment's stability is its odd- or even-electron nature with even-electron species, whether ion or neutral, showing greatly enhanced stability and therefore exerting most influence on subsequent decompositions. Simple homolytic cleavage of an odd-electron molecular ion normally provides an

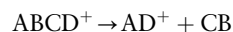
even-electron ion and odd-electron radical. To produce an odd-electron ion, it is necessary for the molecular ion to undergo energetically less-favored multiple-bond cleavage or rearrangement. The positive charge on the product ion is normally located in the fragment with the lowest ionization potential (Simpson's rule). Because even-electron species containing only paired electrons tend to be more stable than odd electron species, fragmentations producing such products tend to predominate.

Fragment ions may also arise directly by the pair-production process without passing through a molecular ion stage:



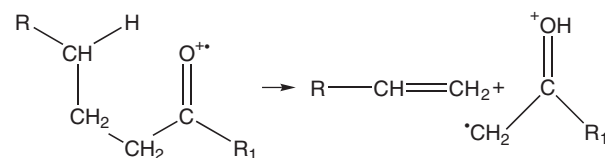
Halogen compounds that can provide stable atomic halogen negative ions are frequently found to fragment in this way.

The fragmentation of an ion often involves a transfer or rearrangement of atoms between products accompanied by the breaking of one bond and the formation of another, and may be represented by



This rearrangement process often involves hydrogen but species as diverse as fluorine, methyl, ethyl, methoxy, and phenyl or hydroxy may be involved, leading to products of enhanced stability. For a rearrangement reaction to produce observable quantities of product, the unfavorable entropy of the reaction must be more than balanced by particularly favorable activation energy.

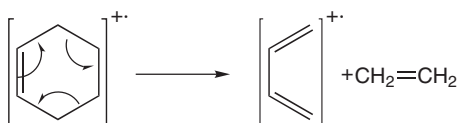
When the fragmenting ion contains a heteroatom, abundant rearrangement ions are often formed in the spectrum via a transition state in which the charge is localized on the heteroatom. The McLafferty rearrangement of ketones may be used to illustrate this process:



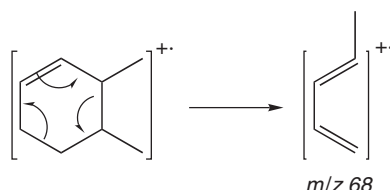
The product ion has lost one of the oxygen lone pair electrons in the ionization process and  $\text{O}^+$  is therefore trivalent. A similar valency increase can occur in rearrangements involving nitrogen as shown in the formation of the ammonium  $\text{R} + \text{NH}_3$  from secondary amines. The McLafferty rearrangement is the most generally applicable rearrangement process observed in the production of mass spectral fragmentations and requires the presence of a multiple bond and a  $\gamma$ -hydrogen if it is to occur. Aldehydes, olefins,

amides, nitrites, esters, substituted aromatics, phosphates and sulfites have all been shown to rearrange by a McLafferty mechanism.

Another well-known process producing abundant rearrangement ions in the mass spectrum is the retro-Diels–Alder reaction of cyclic olefins. The simplest example of this is shown by cyclohexene:



Identification of a molecule from ions produced by this type of rearrangement can be obscured by the migration of hydrogen, and thus the site of the double bond in the molecule before fragmentation may be uncertain. The cyclic olefin 1,2-dimethylcyclohexene is found to produce a base peak at  $m/z$  68, suggesting that the compound is the 1,3-isomer. Production of this peak must be preceded by isomerization of the double bond:



## Interpretation

As with most other spectroscopic techniques, interpretation of the data relies on a combination of an understanding of the fundamental processes going on, the application of logical reasoning, and the use of intuition. There are no hard and fast rules over the best approach; however, certain things should be looked for and the procedure outlined below should yield substantial information:

1. The first and probably most important step in the identification of a compound from its mass spectrum is identification of the molecular ion.

- Note odd or even mass – remember the nitrogen rule.
- Note the isotope pattern – halogens, etc.
- Note the intensity – strong  $M^{+\bullet}$  may mean aromatic.
- If an accurate mass is available, calculate the molecular formula.
- Consider degree of unsaturation, i.e., rings plus double bonds.

2. Identify prominent odd electron ( $OE^{+\bullet}$ ) and even electron ( $EE^+$ ) ions.

- Calculate possible neutral losses and relate to structural units.
- Consider rearrangements.

3. Identify metastable peaks if present.

- Establish connectivity relationships.
- Consider rearrangements.

4. Identify major low mass peaks that relate to neutral losses and structural units.

5. Look for characteristic ion series that identify the compound class, some examples are given below:

27, 41, 55, 69, etc.	Alkenyl	$C_nH_{2n-1}$
15, 29, 43, 57, 71, etc.	Alkyl	$C_nH_{2n+1}$
29, 43, 57, 71, etc.	Aldehydes/ ketones	$C_nH_{2n-1}O$
30, 44, 58, 72, etc.	Amines	$C_nH_{2n+2}N$
39, 50–52, 63–65, 75–78	Aromatics	

All the above procedure may assume less significance if computer-aided library searching is available. The library may not contain an entry for the sample under examination, but will always provide a best-match spectrum. Close scrutiny of the 'goodness of fit' of the library and acquired data is therefore essential if the possibility of an erroneous match is to be avoided.

Many combinations of the basic techniques are known, the exact specification being determined by the function to be addressed. Thus, organic materials, often thermally sensitive and labile, require less energetic (softer) ionization (e.g., CI or laser desorption) than high-melting-point inorganic materials (e.g., ICP or arc discharge ionization). Molecular ion and fragmentation data (from EI ionization) may be required for the identification of unknown molecular materials, while monitoring of a single ion, selected from the complete spectrum, may be all that is required for the identification of a trace impurity. Measurement of isotope ratios in molecular or atomic spectra may be used to identify or quantify polyisotopic species (e.g.,  $^{12}C/^{13}C$  ratio, halogen identification,  $^{32}S/^{34}S$  ratio, and medical tracer identification). The physical state of sample material will determine the method adopted for introduction into the spectrometer, differing techniques being available for gaseous, liquid, and solid samples. Several methods are also available for the separation of sample mixtures before ionization, interfaces being available for gas chromatography, liquid chromatography, supercritical fluid chromatography, capillary electrophoresis, and thin-layer chromatography.

The applications of mass MS are extensive and may be found in many spheres of scientific investigation. The range includes purely physical measurements such as appearance potentials, kinetic studies on ultra-short-lived ionic species, isotopic ratio measurements for dating or health studies, ion structure investigations, and a vast array of qualitative and quantitative analytical procedures employed in chemistry, biochemistry, medicine, nuclear science, and semiconductor engineering.

See also: **Mass Spectrometry:** Overview; Ionization Methods Overview; Electron Impact and Chemical Ionization; Atmospheric Pressure Ionization Techniques; Electrospray; Liquid Secondary Ion Mass Spectrometry; Matrix-Assisted Laser Desorption/Ionization; Mass Separation; Ion Traps; Time-of-Flight.

## Further Reading

Ashcroft AE (1997) *Ionization Methods in Organic Mass Spectrometry*. Cambridge: Royal Society of Chemistry.

Barker J (1999) *Mass Spectrometry*, 2nd edn., Analytical chemistry by open learning series. Chichester: Wiley.

Burlingame AL, Mattley D, Russel DH, and Holland PT (1988) Mass spectrometry. *Analytical Chemistry* 60: 2948–3428.

Chapman JR (1993) *Practical Organic Mass Spectrometry*, 2nd edn. New York: Wiley.

De Hoffman E and Stroobant V (2002) *Mass Spectrometry: Principles and Applications*, 2nd edn. Chichester: Wiley.

Johnstone RAW and Rose ME (1996) *Mass Spectrometry for Chemists and Biochemists*, 2nd edn. Cambridge: Cambridge University Press.

Lee TA (1998) *A Beginner's Guide to Mass Spectral Interpretation*. Chichester: Wiley.

McLafferty FW and Turecek F (1993) *Interpretation of Mass Spectra*, 4th edn. Mill Valley, CA: University Science Books.

Smith RM (2003) *Understanding Mass Spectra: A Basic Approach*, 2nd edn. New York: Wiley-Interscience.

Watson JT (1997) *Introduction to Mass Spectrometry*, 3rd edn. London: Lippincott Williams and Wilkins.

## Ionization Methods Overview

D J Harvey, University of Oxford, Oxford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

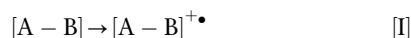
## Introduction

In order to obtain the mass spectrum of a compound, it must first be ionized. Early methods involved thermal evaporation of samples into the ion source of the mass spectrometer where they were ionized by electron beam bombardment or chemical reactions. Thermal instability, where present, was frequently overcome by derivatization. For large and involatile molecules, desorption techniques using high electric fields or bombardment with fission fragments, fast atoms, fast ions, or laser beams were later developed. The newest techniques, developed in the late 1980s, involve laser ionization of samples imbedded in a solid matrix and evaporation of solutions by electrospray. Both of these techniques are able to ionize large molecules such as proteins with masses of up to several hundred kilodaltons. This article describes these techniques in more detail and indicates the type of mass spectrometer that is suitable for analysis of the types of ion that they produce.

## Ionization Methods

### Electron Impact

In the electron impact (EI) method, the ion source consists of a small (~10 mm across) heated chamber located within the vacuum system (Figure 1). A beam of electrons, generated from a heated tungsten or rhenium filament, is sent through the chamber to a second electrode, known as the trap. The voltage difference between the filament and trap governs the energy of the electrons and is usually set at some value above the ionization potential of the sample molecules, typically 70 eV. Small permanent magnets are positioned in line with the electron beam to cause the electrons to execute a spiral path in order to increase the path length. Sample molecules are evaporated either outside the ion source or within it from the sample deposited on a heated probe and, when bombarded with the electron beam, absorb energy and ionize by release of an electron:



The resulting ion is termed an odd-electron molecular ion and is written in the form  $M^{+\bullet}$ . The ionization efficiency is low with only about 1 in  $10^3$  sample molecules producing ions. However, the high



The applications of mass MS are extensive and may be found in many spheres of scientific investigation. The range includes purely physical measurements such as appearance potentials, kinetic studies on ultra-short-lived ionic species, isotopic ratio measurements for dating or health studies, ion structure investigations, and a vast array of qualitative and quantitative analytical procedures employed in chemistry, biochemistry, medicine, nuclear science, and semiconductor engineering.

See also: **Mass Spectrometry:** Overview; Ionization Methods Overview; Electron Impact and Chemical Ionization; Atmospheric Pressure Ionization Techniques; Electrospray; Liquid Secondary Ion Mass Spectrometry; Matrix-Assisted Laser Desorption/Ionization; Mass Separation; Ion Traps; Time-of-Flight.

## Further Reading

Ashcroft AE (1997) *Ionization Methods in Organic Mass Spectrometry*. Cambridge: Royal Society of Chemistry.

Barker J (1999) *Mass Spectrometry*, 2nd edn., Analytical chemistry by open learning series. Chichester: Wiley.

Burlingame AL, Mattley D, Russel DH, and Holland PT (1988) Mass spectrometry. *Analytical Chemistry* 60: 2948–3428.

Chapman JR (1993) *Practical Organic Mass Spectrometry*, 2nd edn. New York: Wiley.

De Hoffman E and Stroobant V (2002) *Mass Spectrometry: Principles and Applications*, 2nd edn. Chichester: Wiley.

Johnstone RAW and Rose ME (1996) *Mass Spectrometry for Chemists and Biochemists*, 2nd edn. Cambridge: Cambridge University Press.

Lee TA (1998) *A Beginner's Guide to Mass Spectral Interpretation*. Chichester: Wiley.

McLafferty FW and Turecek F (1993) *Interpretation of Mass Spectra*, 4th edn. Mill Valley, CA: University Science Books.

Smith RM (2003) *Understanding Mass Spectra: A Basic Approach*, 2nd edn. New York: Wiley-Interscience.

Watson JT (1997) *Introduction to Mass Spectrometry*, 3rd edn. London: Lippincott Williams and Wilkins.

# Ionization Methods Overview

D J Harvey, University of Oxford, Oxford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

In order to obtain the mass spectrum of a compound, it must first be ionized. Early methods involved thermal evaporation of samples into the ion source of the mass spectrometer where they were ionized by electron beam bombardment or chemical reactions. Thermal instability, where present, was frequently overcome by derivatization. For large and involatile molecules, desorption techniques using high electric fields or bombardment with fission fragments, fast atoms, fast ions, or laser beams were later developed. The newest techniques, developed in the late 1980s, involve laser ionization of samples imbedded in a solid matrix and evaporation of solutions by electrospray. Both of these techniques are able to ionize large molecules such as proteins with masses of up to several hundred kilodaltons. This article describes these techniques in more detail and indicates the type of mass spectrometer that is suitable for analysis of the types of ion that they produce.

## Ionization Methods

### Electron Impact

In the electron impact (EI) method, the ion source consists of a small (~10 mm across) heated chamber located within the vacuum system (Figure 1). A beam of electrons, generated from a heated tungsten or rhenium filament, is sent through the chamber to a second electrode, known as the trap. The voltage difference between the filament and trap governs the energy of the electrons and is usually set at some value above the ionization potential of the sample molecules, typically 70 eV. Small permanent magnets are positioned in line with the electron beam to cause the electrons to execute a spiral path in order to increase the path length. Sample molecules are evaporated either outside the ion source or within it from the sample deposited on a heated probe and, when bombarded with the electron beam, absorb energy and ionize by release of an electron:

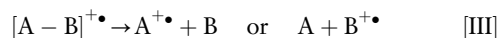
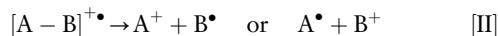


The resulting ion is termed an odd-electron molecular ion and is written in the form  $M^{+\bullet}$ . The ionization efficiency is low with only about 1 in  $10^3$  sample molecules producing ions. However, the high

sensitivity of the mass spectrometer is more than adequate to produce a strong spectrum. The extent to which a molecule ionizes is related to the structure; molecules containing heteroatoms with lone pairs of electrons ionize well by loss of one electron from a lone pair, whereas molecules such as saturated hydrocarbons are difficult to ionize as all electrons are involved in sigma bond formation.

Odd-electron ions are generally unstable and decompose, usually by several consecutive and competing reactions within the ion source in an attempt to stabilize the odd electron. Typical fragmentation

pathways include radical elimination to give an even-electron ion [II], neutral molecule elimination to give another odd-electron ion [III], or fragmentation by rearrangement, often involving migrations of hydrogen atoms. Only the ions, not the neutral species, appear in the spectrum:



The resulting array of parent and fragment ions constitutes the mass spectrum of the sample (Figure 2). Such spectra are very reproducible for a given electron beam energy; 70 eV being chosen as a value that produces negligible changes in the spectrum for small changes in electron energy. Many collections of spectra have been published and a vast amount of published material exists on the chemistry of these fragmentation mechanisms; most reactions are unimolecular although some ion-molecule reactions have been reported for selected compounds.

Ions are ejected from the ion source and focused into a beam by a positive potential on a repeller plate positioned at the back of the ion source and a high potential, in the order of 5–10 kV at the front of the source. The beam then enters the analyzer section of the instrument for mass separation to form the spectrum. As the molecules are essentially ionized in the vapor phase, they must be stable at the temperatures needed to vaporize them. Many molecules do not fulfill these requirements but can be stabilized and made more volatile by derivatization. However, even with derivatization and heated ion sources, the technique is only applicable to molecules with masses less than ~1000.

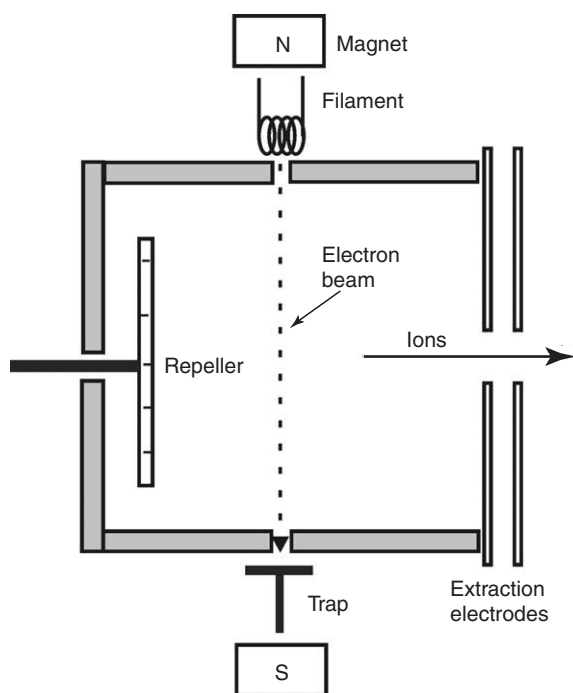


Figure 1 EI ion source.

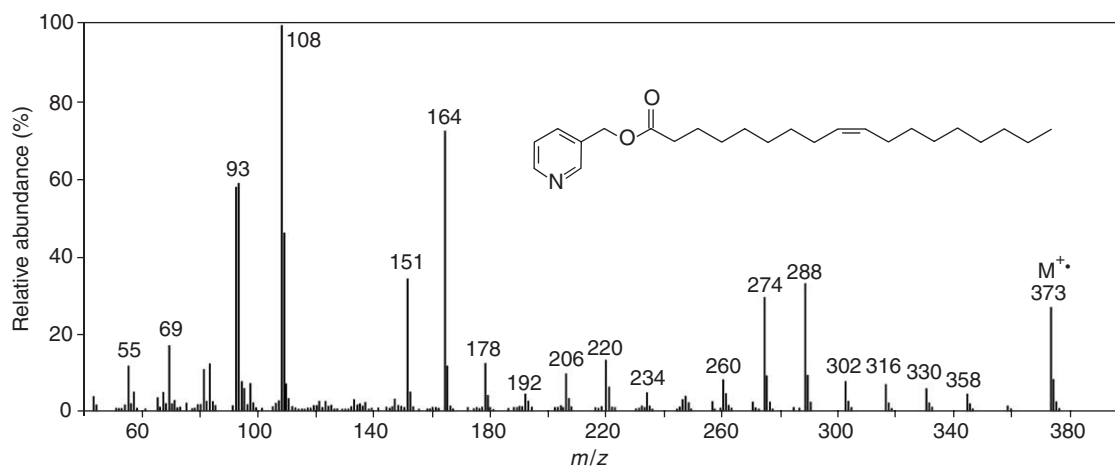
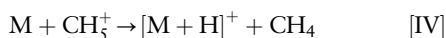


Figure 2 EI mass spectrum of a fatty acid derivatized as its picolinyl ester.

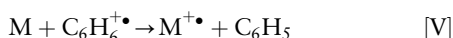
## Chemical Ionization

**Positive ion chemical ionization (CI)** The instability of odd-electron ions formed by EI frequently results in spectra in which the molecular ion peak, the most important in the spectrum, is absent as none of the ions survive long enough to reach the detector. A solution exists in the form of the CI source. This is constructed in a similar manner to the EI source but has a much smaller exit slit, thus making it much more gas-tight, allowing pressures of 0.1–1 Torr to be achieved. A so-called reagent gas is introduced into the source chamber and ionized by the electron beam. Although a large number of suitable gases have been reported, methane, *iso*-butane, and ammonia are the most common. Once ionized, the gases undergo a series of ion–molecule reactions to form species such as  $\text{CH}_5^+$  (from methane) and  $\text{NH}_4^+$  (from ammonia). These species are allowed to react with the sample molecules, present at partial pressures of  $\sim 0.01\%$  of the reagent gas pressure, with a variety of outcomes. Five common reactions are outlined below:

- (i) Proton transfer, typically from hydrocarbon reagent gases and ammonia



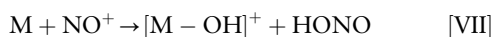
- (ii) Charge exchange from, for example, benzene



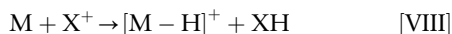
- (iii) Electrophilic addition



- (iv) Anion abstraction as with nitric oxide



- (v) Hydride ion extraction



Of these reactions, the formation of  $[\text{M} + \text{H}]^+$  is the most common. Species formed in these reactions by addition or subtraction of atoms are sometimes referred to as pseudomolecular ions as they do not have the same elemental composition as the charged molecule but the term appears to be falling out of favor.

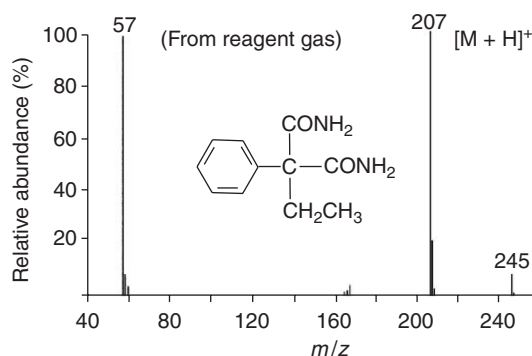
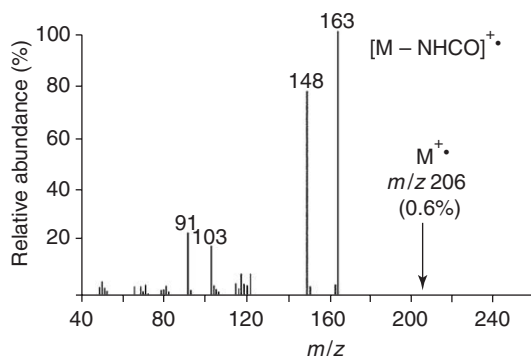
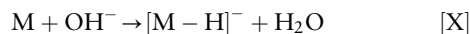
These ionization reactions involve transfer of much less energy to the sample molecules than EI and, consequently, there is much less fragmentation. CI is generally termed a soft ionization technique. In addition, the  $[\text{M} + \text{X}]^+$  or  $[\text{M} - \text{X}]^+$  species do not contain unpaired electrons, i.e., they are even-electron ions. Consequently, they are much more stable than the odd-electron ions and show very little fragmentation (**Figure 3**). EI and CI are, thus, complementary ionization techniques for the same types of volatile molecule and have been combined into a single ion source in a technique termed ACE (alternate CI/EI). Essentially, a CI spectrum of the sample is acquired to produce a strong molecular ion, after which the CI reagent gas is removed and an EI spectrum is acquired to record the fragments. The reagent gas is then pulsed back and the process is repeated. Another variant of the technique is desorption CI in which the sample is introduced as a solid on a probe and subsequently ionized by the reagent gas.

**Negative ion CI** For mainly historical reasons, mass spectrometry is commonly regarded as a positive ion technique. However, many ionization processes produce both positive and negative ions, sometimes favoring the latter. CI is one such technique and, again, several reactions are recognized:

- (i) Electron capture from thermal electrons

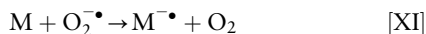


- (ii) Proton abstraction



**Figure 3** (Left) EI spectrum of a drug molecule exhibiting no molecular ion. (Right) CI mass spectrum of the same compound with *iso*-butane as the reagent gas. The ion at  $m/z$  57 is from the reagent.

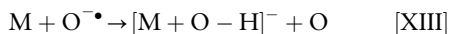
## (iii) Charge exchange



## (iv) Nucleophilic addition from, for example, chlorine produced from methylene dichloride



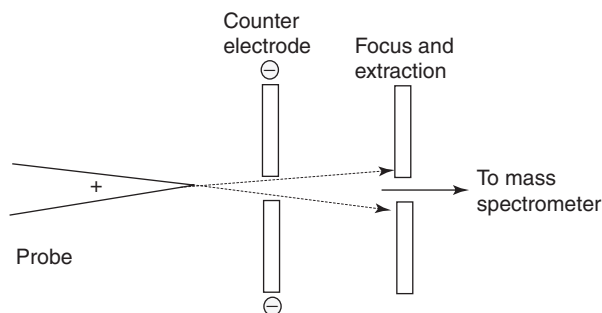
## (v) Nucleophilic displacement with reagents such as nitrous oxide



Capture of electrons that have been thermalized to 0–2 eV can be a very efficient process with suitable compounds and can, thus, often produce greater sensitivity than EI. Many molecules that are not efficiently ionized by this technique can be derivatized with electron-capturing reagents such as fluorocarbon esters in order to increase their susceptibility. This technique is often used for trace analysis, particularly as the derivatization reaction can be compound specific.

**Field Ionization**

A major problem with the above ionization techniques is that they require volatilization of the sample, usually by heating, and are, thus, only suitable for relatively small molecules. This fact delayed the introduction of mass spectrometry into biochemical areas where many of the potential target molecules such as proteins and nucleic acids could not be ionized. There are now many techniques capable of ionizing these molecules, one of the earliest being field ionization (FI). Ionization is produced by exposing the molecules, in the vapor phase, to a strong electric field produced by a sharp point or edge. This electrode is placed a short distance from a cathode such that a field of  $\sim 10^7$ – $10^8$  V cm $^{-1}$  is generated (Figure 4). This field is sufficient to abstract an electron from many types of compound with little excess energy transfer such that  $M^{+\bullet}$  or, occasionally,  $[M + H]^+$  ions are generated. The odd-electron nature of the  $M^{+\bullet}$  ions usually leads to some fragmentation.



**Figure 4** FI ion source.

**Field Desorption**

In a related technique, field desorption (FD), a strong electric field (typically  $10^7$ – $10^8$  V cm $^{-1}$ ) is again used to extract ions from the compound of interest but, in this case, the sample is placed onto a probe, termed the emitter, usually by evaporation. The emitter is covered with a large number of short whiskers with tip radii in the region of 0.1 nm in order to increase the surface area and enhance the electric field. Although spectra can often be obtained at ambient temperature, heating the probe often increases the ion yield. As the sample is not evaporated prior to ionization, but relies on desorption of the sample molecules, the technique can be used for many large, thermally unstable compounds such as peptides and carbohydrates. Unfortunately, unlike the ionization methods described above, the beam of ions only lasts for a short time giving little scope for experimentation.

The technique is very soft usually with the production of only  $M^{+\bullet}$  ions. Energy transfer to the molecules is in the order of fractions of an electronvolt and, consequently, fragmentation is minimal. Some thermal fragmentation may, however, occur as the result of probe heating. Unfortunately, preparation of the emitters is difficult and labor intensive and ions are produced with a large energy spread such that only instruments such as magnetic sector spectrometers are suitable for analyzing them. The technique now enjoys little but historical interest as it has been replaced by the more convenient methods of fast atom bombardment, laser ionization methods, and electrospray, as discussed next.

**Secondary Ion Mass Spectrometry, Liquid  
Secondary Ion Mass Spectrometry, and  
Fast Atom Bombardment**

The introduction of fast atom bombardment (FAB) in 1981 revolutionized organic mass spectrometry in that it provided an easy and convenient method for ionizing thermally unstable molecules with masses up to  $\sim 10$  kDa. It developed from surface ionization methods in which beams of fast atoms or ions are used to sputter ions from solid surfaces in a technique known as secondary ion mass spectrometry (SIMS). However, although organic molecules can often be ionized directly in a similar manner, the resulting ion beam is too transient to be of any use in the types of mass spectrometer available in the early 1980s (mainly magnetic sector and quadrupole instruments). The difference with FAB is that the sample is dissolved in a high-boiling solvent such as glycerol, known as the matrix, before being placed

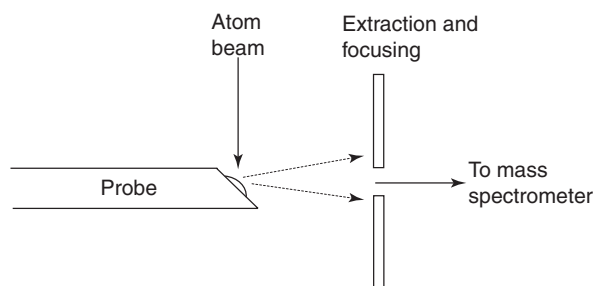
on the probe (Figure 5). If the sample is hydrophobic or can be made so by derivatization, e.g., by permethylation, it is repelled by the solvent molecules and forms a layer on the surface of the matrix. This layer is evaporated and ionized by the fast atom beam (5–8 keV), after which, more sample molecules migrate to the matrix surface to maintain the beam. Ion beams can be maintained for 10–15 min. Because it can take time for the surface layer to reform, a modification of the technique, known as pulsed FAB, in which the atom beam is pulsed, has been used to ensure that there is always sample available for ionization and that the atom beam does not ionize predominantly matrix molecules, which would degrade the signal-to-noise ratio. For large molecules, xenon is the preferred gas because of its mass. However, xenon is expensive and a beam of cesium ions is frequently used as a substitute. With ions as the ionizing

beam instead of atoms, the technique is known as liquid secondary ion mass spectrometry (LSIMS).

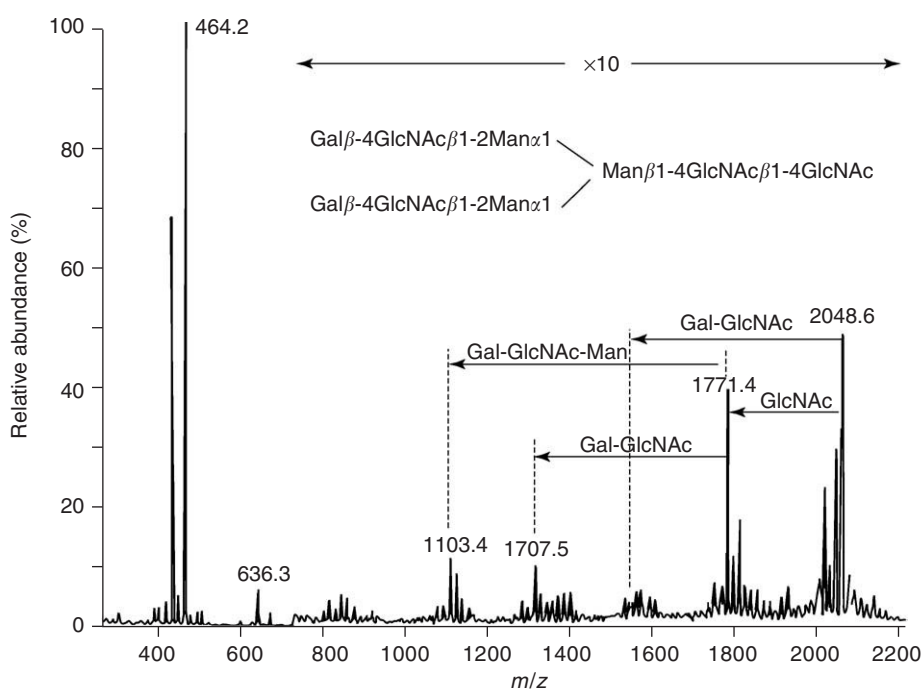
Ionization of matrix molecules is one of the drawbacks of FAB as it results in spectra with a comparatively high background throughout the mass range with adverse effects on sensitivity. Both positive and negative ions are produced and are mainly of the type  $[M + H]^+$  or  $[M + X]^+$  (depending on the matrix) where X can be moieties such as Na, K, or  $NH_4$  or, in negative ion mode,  $[M - H]^-$ . There is often a considerable amount of fragmentation (Figure 6). Matrices vary, depending on the sample. Glycerol was the first to be used and is often the first choice but others, such as thioglycerol and *m*-nitrobenzyl alcohol, are also popular.

Construction of an FAB ion source is comparatively simple and consists mainly of the sample probe and extraction electrode. The sign of the charge on these electrodes governs whether positive or negative ions are extracted. The fast atom beam is generated by ionizing a noble gas such as argon or, less commonly, xenon, forming a beam from the ions and then removing the charge by passing the beam through a relatively high pressure region. The resulting atoms have sufficient kinetic energy to proceed to the target.

**Continuous flow or dynamic-FAB** As FAB essentially works in the liquid phase, it has been used to ionize molecules infused into the mass spectrometer



**Figure 5** Fast atom bombardment ion source.



**Figure 6** FAB spectrum of the permethyl derivative of a complex carbohydrate.



in solution or as effluents from high-performance liquid chromatography (HPLC) columns. The matrix, in these instances, has reduced viscosity and consists of, for example, water containing 5% glycerol. This solution can be effused directly onto the surface of the probe through a narrow capillary. Related to this method is the technique of frit-FAB in which the solution, usually from an HPLC column, passes through a fine stainless steel frit in order to disperse the liquid.

### Plasma Desorption Mass Spectrometry

Another early desorption technique is that of plasma desorption mass spectrometry (PDMS). The sample is deposited onto a thin aluminum or aluminized polyester foil (0.5–1  $\mu\text{m}$  in thickness) and placed just in front of a californium-252 emitter which is located in a time-of-flight (TOF) mass spectrometer (Figure 7). Californium is an  $\alpha$ -emitter that decays into two highly energetic  $\alpha$ -particles that are expelled in diametrically opposite directions. One of these particles collides with and ionizes the sample while the other hits a collection plate and triggers the timing circuit for the TOF mass spectrometer.

Although largely obsolete, PDMS was initially the only ionization technique capable of analyzing high-mass compounds such as small proteins and polar anionic carbohydrates. Proteins were best examined on foils covered with a layer of nitrocellulose to which the protein adhered and from which contaminants could be removed by washing with water. Following analysis, the residual sample could be recovered from the foil for further experiments. A major disadvantage of the technique was the slow rate of ion production resulting in acquisition times of several hours.

### Thermospray and Plasmaspray

The thermospray ion source was originally developed as a method for introduction of the relatively large volumes of liquid (1–1.5  $\text{ml min}^{-1}$ ) from a

liquid chromatographic system into a mass spectrometer and it was only by accident that the device was found to produce ions. The source consists of a heated capillary of inside diameter  $\sim 100 \mu\text{m}$ , through which the solvent flows into a heated source region, an orthogonally positioned sample cone for transferring ions to the mass spectrometer and a large pump, often with a solvent trap to remove the vapor. A repeller electrode opposite the cone or just downstream (known in this case as the retarder) aids ion transit to the spectrometer (Figure 8). The source may also be provided with a discharge electrode, or filament, to facilitate ionization by stripping electrons from the sample molecules. Under these conditions, the technique is known as plasmaspray.

The temperature of the capillary is adjusted so that  $\sim 95\%$  of the solvent evaporates to provide a nebulizing gas that causes the remaining solvent, which normally contains a volatile buffer such as ammonium acetate, to form small droplets that subsequently evaporate. Ionization is by CI processes, is mild, and produces ions such as  $[\text{M} + \text{H}]^+$  and  $[\text{M} + \text{NH}_4]^+$ . Both positive and negative ions are produced, unlike electrospray in which the capillary carries a high potential (see below). Spontaneous fragmentation by loss of neutral molecules is sometimes seen but the discharge electrode can be used to produce additional ionization when no buffer is present. Ionization is very dependent on the sample and ion source conditions and the technique does not work as well for large compounds such as proteins as it does for smaller molecules such as drugs and drug metabolites, an area in which it has had extensive use. Electrospray has now largely replaced thermospray

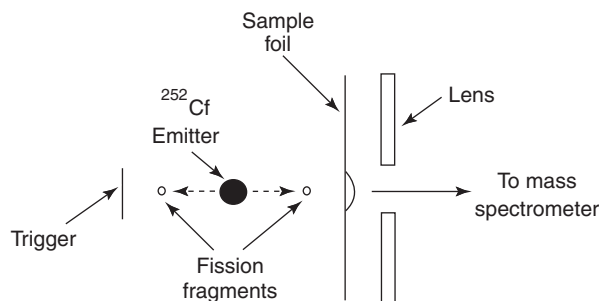


Figure 7 Plasma desorption ion source.

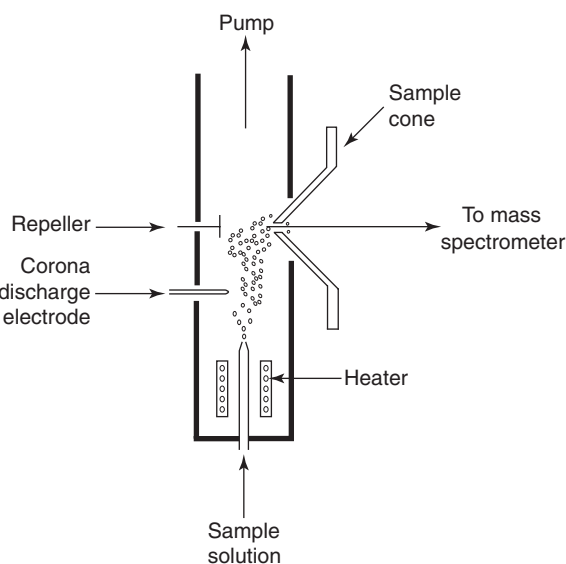


Figure 8 Thermospray ion source.

as a means of introducing LC effluents into a mass spectrometer.

### Electrospray Ionization

In the late 1980s, two techniques, electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI), emerged and caused a further revolution in mass spectrometry as each was capable of ionizing extremely large molecules in excess of several mDa. This led to the joint award of the 2002 Nobel prize for chemistry. The electrospray ion source works at atmospheric pressure and consists of a stainless steel or metalized silica capillary held at a high potential, typically  $\sim 3$  kV, and placed a few millimeters from a counter electrode (Figure 9). A solution of the sample flows through the capillary and charged droplets emerge into the ion source where they meet a counter flow of warm nitrogen gas, sometimes known as a curtain gas, to vaporize the solvent. As the droplet diameter decreases, a point is reached where electrostatic repulsion causes the droplet to disintegrate into smaller droplets. The process continues until only ionized sample molecules remain. These are extracted by a sampling cone placed off-axis or at right angles to the solvent flow and introduced into the mass spectrometer. Typical solvents are mixtures of water and either methanol or acetonitrile; pure water or nonpolar organic solvents are less effective. Efficiency can be improved by use of a coaxial nebulizing flow of nitrogen around the spray; a technique known as ion-spray.

The main feature of the electrospray process is that it produces multiple-charged ions. For a typical protein, an average of one charge is added for each kDa in mass. Both positive and negative ions of the type  $[M + nH]^{n+}$ ,  $[M + nX]^{n+}$ , or  $[M - nH]^{n-}$  are produced with a typical spectrum consisting of an array of ions, each differing by one charge state (Figure 10). Although instrumental resolution is not usually sufficient to enable the charge state to be determined directly by observation of the isotopic peak spacing, measurement of the  $m/z$  values of each peak and knowledge that the charge difference between any pair of peaks is one charge enables the mass to be

calculated. As many pairs of peaks are available for measurement, an extremely accurate measurement of the mass can be obtained, usually better than 0.01%; this measurement is made in the background by the instrument's data system and a reconstructed mass spectrum consisting of one peak for each component is produced.

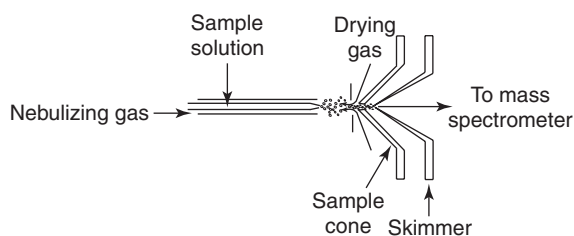
The essential feature of the electrospray spectrum that enables compounds of high mass to be examined is the multiple charging. Typical  $m/z$  values are in the region below  $\sim 5$  kDa and, consequently, typical quadrupole or related instruments can easily analyze the ions with no modification being needed. As the method accepts a flowing solvent, it is ideally coupled to liquid chromatographic columns and has major applications in the field of protein and peptide analysis.

**Nanospray** Nanospray is a variant of electrospray in that the solvent is sprayed from a very fine capillary at flows that can be as low as  $20 \text{ nl min}^{-1}$ . Sensitivity is good and desolvation is enhanced because of the small initial droplet size. Furthermore, for very polar compounds such as carbohydrates, the larger molecules appear to be more efficiently ionized than with conventional electrospray.

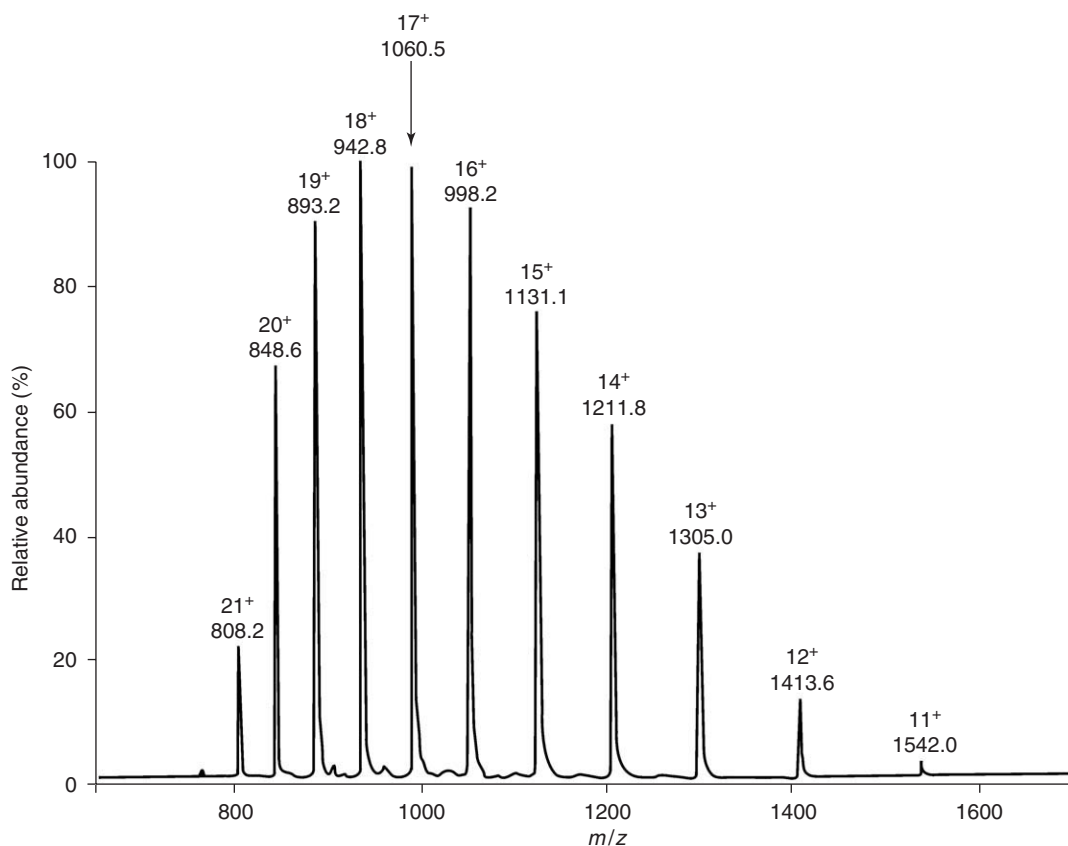
**Atmospheric pressure chemical ionization (APCI)** APCI is another variation of electrospray where, in a similar manner to thermospray, the incorporation of a corona discharge into the spray increases the ion yield by electron stripping from both sample and solvent molecules. Much of the ion formation thus occurs by CI processes.

### Matrix-Assisted Laser Desorption/Ionization

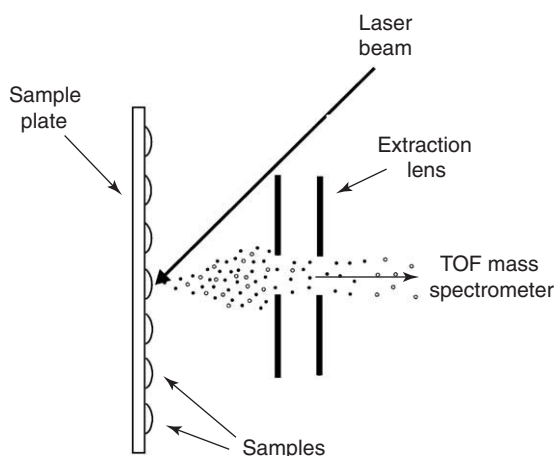
In this technique, the sample is mixed with an ultraviolet (UV) absorbing matrix and ionized with a UV laser (Figure 11). Ions are extracted with a high potential (in the order of 20 kV) and accelerated into the mass spectrometer. The purpose of the matrix is to dilute the sample and absorb the laser energy for transmission to the sample but the exact mechanism by which this occurs is, as yet, unclear. The main products are singly charged ions such as  $[M + H]^+$ ,  $[M - H]^-$ , or  $[M + Na]^+$  (Figure 12). As the technique produces a pulsed beam, it is ideally interfaced to TOF mass spectrometers. However, magnetic sector instruments equipped with array detectors have been used and, recently, collisional cooling of the ion beam has enabled MALDI to be linked to quadrupole mass spectrometers. Commercial instruments invariably use nitrogen lasers emitting at 337 nm although several other wavelengths including those in the infrared region of the spectrum have been used. The



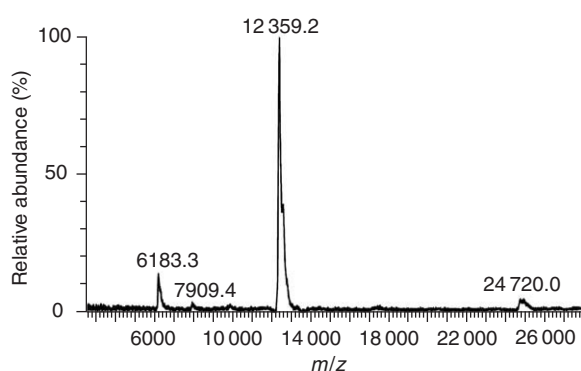
**Figure 9** Electrospray ion source.



**Figure 10** Electrospray spectrum of myoglobin (16951.5 Da). Measured masses of the peaks and the charge states are shown.



**Figure 11** MALDI ion source.



**Figure 12** Typical MALDI mass spectrum of a protein. The peak at  $m/z \sim 12359$  is the  $[M+H]^+$  ion. Peaks at  $m/z \sim 6183$  and  $24720$  are from the  $[M+2H]^{2+}$  ion and the dimer, respectively.

mass range of the technique is limited, not by its ability to ionize molecules but by the drop in signal with increasing mass imposed by most mass spectrometric detectors. In reality, the method is useful up to  $\sim 300$  kDa on commercial instruments although ions can sometimes be seen at much higher mass.

Most types of molecule can be ionized given a suitable matrix and a considerable amount of work

has been performed on the determination of the best matrices for various compounds. 2,5-Dihydroxybenzoic acid appears to be one of the best all-round matrices and is the standard for carbohydrates. Proteins and glycoproteins are best ionized from sinapinic acid whereas peptides give strong signals from  $\alpha$ -cyano-4-hydroxycinnamic acid. In order to obtain a spectrum, the sample is mixed with the matrix in a

ratio of roughly 1:5000 sample–matrix and allowed to crystallize on the mass spectrometry target. This target, which is usually in the form of a plate with up to several hundred sample spots, is inserted into the mass spectrometer and illuminated by the laser.

MALDI offers high sensitivity; some samples can be analyzed from as little as a few hundred attomoles applied to the target and often, after spectral acquisition, the remaining sample can be recovered. The technique is reasonably tolerant to the presence of contaminants such as salts and buffers but for strong signals these contaminants should be removed. MALDI, along with electrospray, has now become one of the major methods for examination of proteins and peptides and is extensively used in proteomic research. Measurement of synthetic polymer distributions is another growth area and large carbohydrates can be ionized without the need for derivatization as required by FAB mass spectrometry.

**Atmospheric pressure matrix-assisted laser desorption/ionization (ap-MALDI)** A variant of the conventional MALDI technique is ap-MALDI in which the sample is ionized outside the vacuum system and ions are captured through a small orifice, usually into an ion trap mass spectrometer. Its advantage appears to be a considerable amount of rapid collisional cooling of the ions by the high atmospheric pressure, leading to stabilization of sensitive compounds. On the other hand, some compounds, such as carbohydrates, appear to suffer increased fragmentation under these conditions.

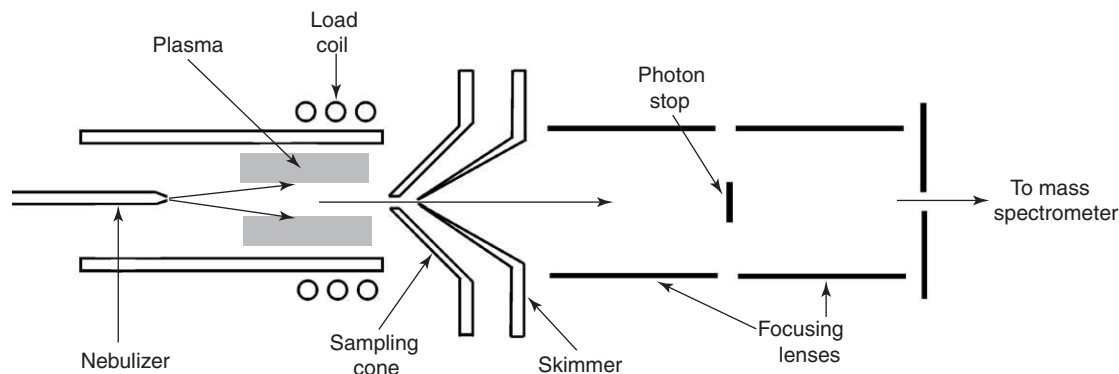
**Surface enhanced laser desorption/ionization (SELDI)** SELDI is a variant of MALDI in which the MALDI probe is derivatized with various substances that have affinity for the analyte. The probes are then used to extract the analyte directly from mixtures thus avoiding sample loss through more complicated cleanup procedures. Contaminants can

be washed from the probe with appropriate buffers or solvents leaving the purified analyte ready for analysis. Many adsorbents have been used; typical examples are hydrophobic or ionic compounds, enzymes, antibodies, DNA, and various receptors. Although most applications have been reported with proteins, the technique is applicable to any type of compound for which a specific adsorbent can be attached to the probe.

Three types of SELDI are recognized, surface-enhanced affinity capture (SEAC), surface-enhanced neat desorption (SEND), and surface-enhanced photolabile attachment and release (SEPAR). With SEND, the energy-absorbing molecules typical of the MALDI process are bound to the probe surface, often covalently although, recently, porous silica has been found to be an effective SEND surface. Additional binding is involved with SEPAR and analytes are released photochemically during the ionization process. SEAC is the most widely used SELDI technique; the probe essentially acts as the purification device via affinity capture and then spectra are recorded as in MALDI following the addition of a typical MALDI matrix. Various affinity probes are now commercially available and the technique offers great potential for examination of trace material in complex biological matrices.

### Spark Source Mass Spectrometry

This is a technique for analysis of inorganic material. The sample, as a powder, is mixed with graphite and pressed into rods. Two rods carrying high potentials are placed close together such that a spark is generated between them. The result is mainly the production of atomic species such as  $\text{Mg}^{2+}$ ,  $\text{K}^+$ , and  $\text{Cl}^+$  and large numbers of elements can be analyzed simultaneously. However, the method has drawbacks such as the presence of incompletely dissociated molecular fragments and the production of ions



**Figure 13** ICP ion source.

in multiple charge states leading to very complex spectra.

### Inductively Coupled Plasma Mass Spectrometry

Most elemental analyses performed by mass spectrometry now employ inductively coupled plasma mass spectrometry (ICPMS), a technique that produced mainly singly charged ions against a very low background. Acidic solutions of the analyte, at concentrations of  $1 \mu\text{g ml}^{-1}$  or less, are sprayed axially from a nebulizer into an inductively coupled argon plasma operating in the region of 8–10 000 K. Ions are sampled through a metal sampling cone, usually nickel, with an aperture of 1–1.2 mm (Figure 13). A sampling skimmer with a  $5 \mu\text{m}$  aperture placed  $\sim 6 \text{ mm}$  behind this cone captures ions into a lens system that focuses them into the quadrupole mass spectrometer. A photon stop that casts a shadow over the final aperture into the spectrometer blocks photons that would otherwise increase the background noise.

Spectra generated by these instruments are comparatively simple because of the predominantly singly charged ions. Most of the naturally occurring elements have three or fewer isotopes and isotope overlap is minimal. There are no coincidences in mass below 40 ( $^{40}\text{Ar}$ ,  $^{40}\text{Ca}$ , and  $^{40}\text{K}$ ) and above this mass, only indium ( $^{113}\text{In}$  and  $^{115}\text{In}$ ) has no isotope free of isobaric interference. Detection limits are in the region of  $50 \text{ pg ml}^{-1}$ .

See also: **Derivatization of Analytes. Gas Chromatography:** Mass Spectrometry. **Liquid Chromatography:** Liquid Chromatography–Mass Spectrometry. **Mass Spectrometry:** Overview; Matrix-Assisted Laser Desorption/Ionization; Time-of-Flight; Peptides and Proteins.

### Further Reading

- Ardrey RE (2003) *Liquid Chromatography–Mass Spectrometry: An Introduction*. Chichester: Wiley.
- Ashcroft AE (1997) *Ionization Methods in Organic Mass Spectrometry*. Cambridge: Royal Society of Chemistry.
- Beckey HD (1977) *Principles of Field Ionization and Field Desorption Mass Spectrometry*. New York: Pergamon.
- Caprioli RM (1990) *Continuous-Flow Fast Atom Bombardment Mass Spectrometry*. Chichester: Wiley.
- Cole RB (1997) *Electrospray Ionization Mass Spectrometry, Fundamentals, Instrumentation and Applications*. New York: Wiley.
- Date AR and Gray AL (1989) *Applications of Inductively Coupled Plasma Mass Spectrometry*. Glasgow: Blackie (New York: Chapman and Hall).
- Harrison AG (2000) *Chemical Ionization Mass Spectrometry*. Boca Raton, FL: CRC Press.
- Harvey DJ (1999) Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrometry Reviews* 18: 349–451.
- Hill SD (1999) *ICP Spectrometry and its Applications*. Sheffield: Sheffield Academic.
- Holland G and Tanner SD (1999) *Plasma Source Mass Spectrometry: New Developments and Applications*. Cambridge: Royal Society of Chemistry.
- Mark TD and Dunn GH (1985) *Electron Impact Ionization*. Berlin: Springer.
- Merchant M and Weinberger SR (2000) Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. *Electrophoresis* 21: 1164–1167.
- Montaser A (1998) *Inductively Coupled Plasma Mass Spectrometry*. Berlin: Wiley-VCH.
- Prokai L (1989) *Field Desorption Mass Spectrometry*. New York: Dekker.
- Vickerman JC, Brown A, and Reed NM (1989) *Secondary Ion Mass Spectrometry: Principles and Applications*. Oxford: Clarendon.

## Electron Impact and Chemical Ionization

**E Eljarrat and D Barceló**, Institut d' Investigacions Quíiques i Ambientals de Barcelona, Consejo Superior de Investigaciones Científicas, Barcelona, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Electron impact (EI) is not only the oldest but also is still the most frequently used ionization technique in environmental organic analysis. EI is, theoretically, capable of ionizing all organic compounds. The practical limitations arise from vaporizing the sample in the

source. Highly involatile compounds, with large or very polar molecules, cannot be evaporated from a probe, while thermally labile substances decompose on heating. Among the alternative ionization modes, chemical ionization (CI) is by far the most frequently used method. CI is capable of ionizing a wide range of organic molecules, although ionization efficiency varies greatly, depending upon the type and degree of functionalization. Molecules that support protonation work best, whereas hydrocarbons and haloalkanes ionize very poorly. Combined EI/CI sources have been commercially available for many years. It is of particular importance that both EI and CI sources can easily be coupled



in multiple charge states leading to very complex spectra.

### Inductively Coupled Plasma Mass Spectrometry

Most elemental analyses performed by mass spectrometry now employ inductively coupled plasma mass spectrometry (ICPMS), a technique that produced mainly singly charged ions against a very low background. Acidic solutions of the analyte, at concentrations of  $1 \mu\text{g ml}^{-1}$  or less, are sprayed axially from a nebulizer into an inductively coupled argon plasma operating in the region of 8–10 000 K. Ions are sampled through a metal sampling cone, usually nickel, with an aperture of 1–1.2 mm (Figure 13). A sampling skimmer with a  $5 \mu\text{m}$  aperture placed  $\sim 6 \text{ mm}$  behind this cone captures ions into a lens system that focuses them into the quadrupole mass spectrometer. A photon stop that casts a shadow over the final aperture into the spectrometer blocks photons that would otherwise increase the background noise.

Spectra generated by these instruments are comparatively simple because of the predominantly singly charged ions. Most of the naturally occurring elements have three or fewer isotopes and isotope overlap is minimal. There are no coincidences in mass below 40 ( $^{40}\text{Ar}$ ,  $^{40}\text{Ca}$ , and  $^{40}\text{K}$ ) and above this mass, only indium ( $^{113}\text{In}$  and  $^{115}\text{In}$ ) has no isotope free of isobaric interference. Detection limits are in the region of  $50 \text{ pg ml}^{-1}$ .

See also: **Derivatization of Analytes. Gas Chromatography: Mass Spectrometry. Liquid Chromatography: Liquid Chromatography–Mass Spectrometry. Mass Spectrometry: Overview; Matrix-Assisted Laser Desorption/Ionization; Time-of-Flight; Peptides and Proteins.**

### Further Reading

- Ardrey RE (2003) *Liquid Chromatography–Mass Spectrometry: An Introduction*. Chichester: Wiley.
- Ashcroft AE (1997) *Ionization Methods in Organic Mass Spectrometry*. Cambridge: Royal Society of Chemistry.
- Beckey HD (1977) *Principles of Field Ionization and Field Desorption Mass Spectrometry*. New York: Pergamon.
- Caprioli RM (1990) *Continuous-Flow Fast Atom Bombardment Mass Spectrometry*. Chichester: Wiley.
- Cole RB (1997) *Electrospray Ionization Mass Spectrometry, Fundamentals, Instrumentation and Applications*. New York: Wiley.
- Date AR and Gray AL (1989) *Applications of Inductively Coupled Plasma Mass Spectrometry*. Glasgow: Blackie (New York: Chapman and Hall).
- Harrison AG (2000) *Chemical Ionization Mass Spectrometry*. Boca Raton, FL: CRC Press.
- Harvey DJ (1999) Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrometry Reviews* 18: 349–451.
- Hill SD (1999) *ICP Spectrometry and its Applications*. Sheffield: Sheffield Academic.
- Holland G and Tanner SD (1999) *Plasma Source Mass Spectrometry: New Developments and Applications*. Cambridge: Royal Society of Chemistry.
- Mark TD and Dunn GH (1985) *Electron Impact Ionization*. Berlin: Springer.
- Merchant M and Weinberger SR (2000) Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. *Electrophoresis* 21: 1164–1167.
- Montaser A (1998) *Inductively Coupled Plasma Mass Spectrometry*. Berlin: Wiley-VCH.
- Prokai L (1989) *Field Desorption Mass Spectrometry*. New York: Dekker.
- Vickerman JC, Brown A, and Reed NM (1989) *Secondary Ion Mass Spectrometry: Principles and Applications*. Oxford: Clarendon.

## Electron Impact and Chemical Ionization

**E Eljarrat and D Barceló**, Institut d' Investigacions Quíiques i Ambientals de Barcelona, Consejo Superior de Investigaciones Científicas, Barcelona, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Electron impact (EI) is not only the oldest but also is still the most frequently used ionization technique in environmental organic analysis. EI is, theoretically, capable of ionizing all organic compounds. The practical limitations arise from vaporizing the sample in the

source. Highly involatile compounds, with large or very polar molecules, cannot be evaporated from a probe, while thermally labile substances decompose on heating. Among the alternative ionization modes, chemical ionization (CI) is by far the most frequently used method. CI is capable of ionizing a wide range of organic molecules, although ionization efficiency varies greatly, depending upon the type and degree of functionalization. Molecules that support protonation work best, whereas hydrocarbons and haloalkanes ionize very poorly. Combined EI/CI sources have been commercially available for many years. It is of particular importance that both EI and CI sources can easily be coupled

**Table 1** Guideline to select the best ionization technique for each class of substance

Nature of sample	Ionization technique	Example
Gaseous samples and volatile thermally stable substances	EI, CI (EI may well not show a molecular ion)	CO <sub>2</sub> , NO, solvents, PAHs, dioxins
Small molecules (<1000), which are pure and sufficiently stable and volatile to desorb from a probe	EI, CI (EI may well not show a molecular ion)	Many organic molecules
Small molecules (<1500), which are not volatile or thermally stable. Molecules need to have a degree of proton affinity or potential to form a cation or an anion	FAB, if pure, from a probe. ESI or APCI online in a solvent stream, or with an LC separation if required	Amino acids, carbohydrates, lipids, and many organic molecules
Small molecules as above, which have been derivatized to give stable volatile products	EI or CI using GC online with MS	Acids as esters, alcohols as silyl ethers
Peptides, proteins, oligonucleotides	ESI from solution or online with micro-LC	Intact proteins
Proteins and peptides and mixtures of same. Polymers	MALDI	Tryptic digests. PPG, PEG
'Non-covalent' interactions	ESI (nanoflow) with MS/MS	Drug–drug or drug–protein interactions

APCI, atmospheric pressure chemical ionization; CI, chemical ionization; EI, electron impact; ESI, electrospray ionization; FAB, fast atom bombardment; GC, gas chromatography; LC, liquid chromatography; MALDI, matrix-assisted laser desorption ionization; MS, mass spectrometry; MS/MS, tandem mass spectrometry; PAHs, polycyclic aromatic hydrocarbons; PEG, polyethylene glycol; PPG, polypropylene glycol.

with capillary gas chromatography (GC), thus combining the high separation efficiency of GC with the high sensitivity and specificity of mass spectrometry (MS).

Whereas EI is an energetic ionization technique, CI is a softer ionization applied to volatile samples where no or a very small molecular ion is observed due to excessive fragmentation. For both EI and CI, the molecular weight range is 10–1000 Da. In rare cases, it is possible to analyze samples of higher molecular weight. Accuracy of the mass measurement at low resolving power is  $\pm 0.1$  Da and in the high-resolution mode,  $\pm 5$  ppm.

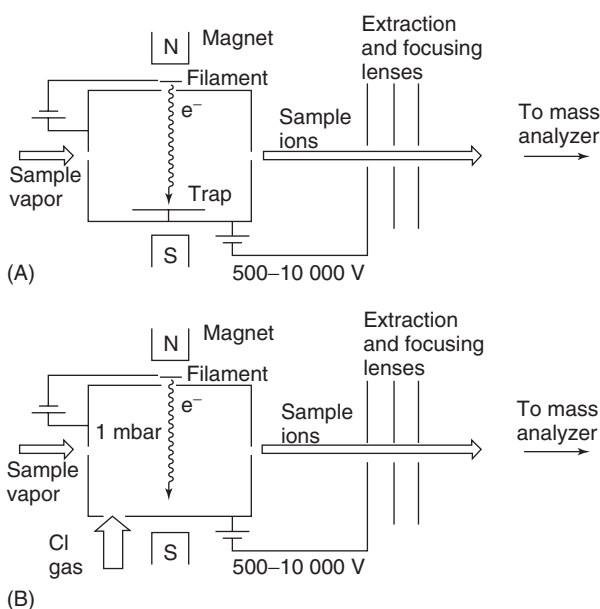
EI and CI methods can be used if the compound to be studied is sufficiently volatile and stable to be vaporized intact. However, only 20% of the organics found in surface water are volatile enough to be amenable to GC–EI–MS or GC–CI–MS. Today, there are a variety of other ionization techniques available: electrospray ionization (ESI), atmospheric pressure chemical ionization, matrix-assisted laser desorption ionization, and fast atom bombardment. Each of these has its advantages and disadvantages. A simple guideline to the most likely optimum ionization technique for a given class of substance is given in Table 1.

In this article, the basic theory of EI and CI is introduced. Then, some applications using both ionization techniques will be presented.

## Theory of Operation

### Electron Impact

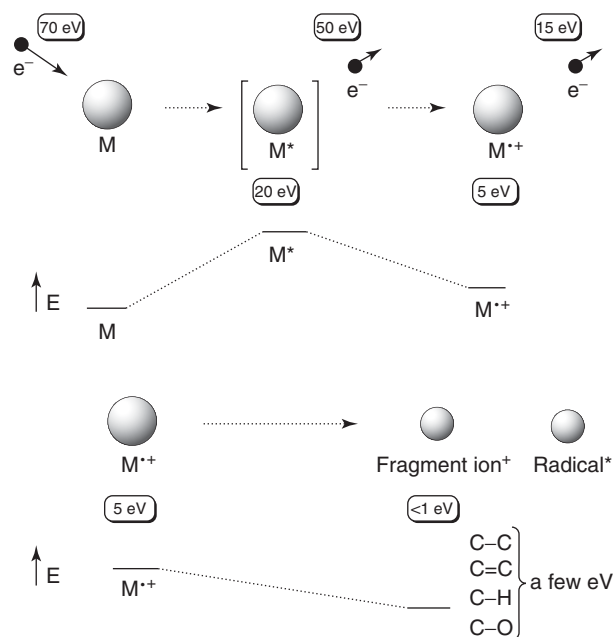
EI is the classical ionization method in MS. The sample for analysis is introduced into the ion

**Figure 1** (A) EI ionization source and (B) CI ionization source.

source (Figure 1) (held under high vacuum,  $10^{-7}$ – $10^{-5}$  mbar) from a reservoir (in the case of gases and volatile liquids), from a heated probe (for involatile liquids and solids), or as the eluent from a GC. It is essential that the sample enters the ion source in the gaseous state and the ability to heat the source and solids probe are important to successful sample analysis. Sample molecules collide with high-energy electrons (typically 70 eV), produced by a glowing filament of either tungsten or rhenium, and if the energy transferred exceeds the molecules' ionization

energy, ions are formed. Because the pressure is kept low ( $10^{-7}$ – $10^{-5}$  mbar), ion–molecule reactions do not occur, e.g., the  $[M+H]^+$  signal due to proton transfer is not observed. Application of an extraction voltage (500–10 000 V, depending on the type of instrument) to a lens stack accelerates ions out of the source to the mass analyzer. Ionization efficiency is increased by placing the inner source between the poles of a small magnet, causing the electrons to travel with a helical trajectory.

EI ion formation is outlined in Figure 2. In this example, 20 eV is transferred to the sample molecule following its collision with a 70 eV electron. Since the ionization energy of most organic molecules is of the order of 15 eV, an electron is expelled to produce a radical cation  $M^{+\bullet}$  with 5 eV excess energy. Once formed, the radical cation can be accelerated out of the source, mass analyzed, and detected as the molecular ion. Given that the mass of an electron is trivial in comparison to that of an organic molecule, the mass of the molecular ion is effectively the molecular mass  $M_r$  of the sample. If we consider that covalent bond strength in an organic molecule is of the order of a few electron volts, any individual molecular ions possessing an internal energy greater than this level will be prone to fragmentation. In the example given in Figure 2, the molecular ion is 5 eV above the ground state and there is easily sufficient energy to cause fragmentation. Fragments may be lost as radicals or as neutral molecules. MS is only capable of detecting ions, hence neutral molecules and radicals do not appear in the spectrum.



**Figure 2** Overall reaction:  $M + e^- \rightarrow M^+ + 2e^-$ .

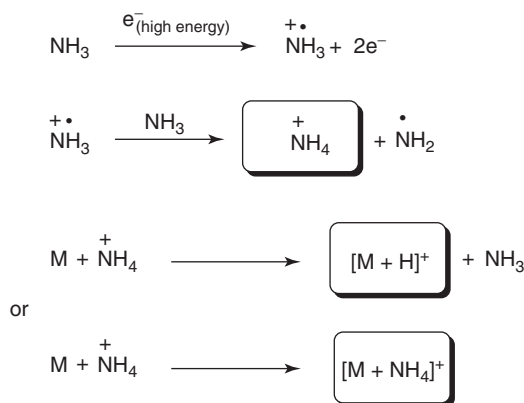
Since the ion source temperature (200–250°C) and the bombarding electron's energy are kept constant, the number and amount of fragments is constant for almost every mass spectrometer. Therefore, the number and amount of ionic fragments and the amount of the  $M^+$  is characteristic for each substance. For this reason, EI-MS is particularly valuable for analysis as both the molecular ion (if present) and abundant fragments can be used to characterize compounds. Library searches are extensively used to identify 'unknown' compounds. Huge mass spectral libraries are commercially available, such as the NIST Library, which contains 230 000 spectra, and the Wiley Library, with 275 000 spectra. Such libraries usually contain only EI mass spectra, which further support the importance of EI-MS. The identification process is based on search algorithms that compare the obtained spectra with those of a library. A spectral match and fit factor defines the certainty of the identification. However, not all compounds have a unique mass spectrum. For example, the EI mass spectra of such distinct compounds as 9-fluorenone and benzo(c)chinoline are almost indistinguishable. Moreover, it is often impossible to distinguish between isomers by EI-MS.

### Chemical Ionization

CI is an ionization technique similar to the classical EI but the knowledge and results of ion–molecule reactions are exploited. CI is carried out in an ion source similar to that used for EI (Figure 1). The principal difference between the two techniques is the presence, during operation in the CI mode, of a CI reagent gas (typically ammonia, methane, or isobutene). Dedicated CI sources also tend to have a narrower exit slit to maintain a higher CI gas pressure in the inner source ( $10^{-3}$ –1 mbar). Electrons from the filament ionize the CI gas in an EI source and the free radicals produced undergo various possible ion–molecule reactions with the sample molecules present in order to enhance the abundance of the molecular ion. The major reaction in positive chemical ionization (PCI) is proton transfer, which takes place in sample molecules with a higher proton affinity than the reagent ions. Other reactions that can also occur include charge exchange, if the reagent gas does not contain available hydrogen, electrophilic addition, and anion abstraction. Table 2 shows the different reactive species generated for each reagent gas normally used in CI. The major difference between isobutene CI and methane CI is the fact that protonated methane can transfer 284 kJ mol $^{-1}$  more energy during ionization than *t*-butyl cation, the main reagent ion of isobutene.

**Table 2** Different reactive species generated for each reagent gas

Reagent gas	Reactive species	Characteristics
H <sub>2</sub> CH <sub>4</sub>	H <sub>3</sub> <sup>+</sup> CH <sub>5</sub> <sup>+</sup> , C <sub>2</sub> H <sub>5</sub> <sup>+</sup>	Good for most organic compounds Usually produces [M + H] <sup>+</sup> , [M + CH <sub>3</sub> ] <sup>+</sup> adducts Adducts are not always abundant Extensive fragmentation
C <sub>2</sub> H <sub>4</sub> C <sub>3</sub> H <sub>8</sub> H <sub>2</sub> O <i>i</i> -C <sub>4</sub> H <sub>10</sub>	C <sub>2</sub> H <sub>5</sub> <sup>+</sup> C <sub>3</sub> H <sub>7</sub> <sup>+</sup> H <sub>3</sub> O <sup>+</sup> <i>t</i> -C <sub>4</sub> H <sub>9</sub> <sup>+</sup>	Usually produces [M + H] <sup>+</sup> , [M + C <sub>4</sub> H <sub>9</sub> ] <sup>+</sup> adducts and some fragmentation Adducts are relatively more abundant than for methane Not as universal as methane Fragmentation virtually absent Polar compounds produce [M + NH <sub>4</sub> ] <sup>+</sup> adducts Basic compounds produce [M + H] <sup>+</sup> adducts Non-polar and non-basic compounds are not ionized
NH <sub>3</sub>	NH <sub>4</sub> <sup>+</sup>	

**Figure 3** Positive ion CI with ammonia.

Thus, there is usually less fragmentation with isobutene than with methane. The ammonium ion can transfer a proton only to compounds with a higher proton affinity than ammonia, mainly nitrogen-containing compounds. As with EI, the sample is introduced from a reservoir, from a heated probe, or as the eluent from a GC.

The CI of ammonia is outlined in **Figure 3**. First, the ammonia radical cation is generated by EI, and this reacts with neutral ammonia to form the ammonium cation, the reactive species in ammonia CI. Once produced, NH<sub>4</sub><sup>+</sup> reacts with sample molecules by proton transfer or adduct formation. The [M + H]<sup>+</sup> and [M + NH<sub>4</sub>]<sup>+</sup> ions are often referred

to as molecular species. They have lower internal energy than radical cations produced by EI and, hence are less likely to fragment. CI is not as soft as ESI or field ionization, and some fragmentation is usually observed.

**Negative chemical ionization** Many important compounds of environmental and biological interest can produce negative ions under the right conditions. Negative ions can be produced by a number of processes. Negative ions are formed by ion–molecule reactions between sample and reagent gas ions. Such reactions include proton transfer, charge exchange, nucleophilic addition, or nucleophilic displacement. Moreover, the capture of the thermal electrons generated under CI conditions allows for the formation of molecular anions from compounds with a positive electron affinity. The electron energy is very low, and the specific energy required for electron capture depends on the molecular structure of the analyte. Electron attachment is an important mode of formation of negative ions, which frequently is used in CI. Negative ions are produced as a result of electron–molecule interactions by three general processes:

- ion-pair formation:  $e + \text{MX} \rightarrow \text{M}^+ + \text{X}^- + e$
- electron attachment:  $e + \text{MX} \rightarrow \text{MX}^-$
- dissociative electron attachment:  $e + \text{MX} \rightarrow \text{M} + \text{X}^-$

Benefits of negative CI (NCI) are efficient ionization, higher sensitivity, and less fragmentation than positive-ion EI or CI. There is also a greater selectivity for certain environmentally or biologically important compounds. The limitations are that not all volatile compounds produce negative ions and a poor reproducibility of the measurements.

### Coupling GC–MS

If a mixture is introduced into an ion source a mixed spectrum will result. If a technique that only gives a single ion from the molecule is used and if all the components of the mixture are similar in chemical structure then it is possible to do a mixture analysis; if an energetic ionization is used such as EI the presence of all the fragment ions will be confusing in the resulting spectrum. Also, if two or more very disparate components are present, the ionization of one may suppress the ionization of all the others. Thus, if possible it is best to introduce the analytes discretely from a chromatographic separation. GC is the original method linked to MS. We also have the capability to perform tandem mass spectrometry (MS/MS) with both EI and CI techniques.

## Applications for EI-MS and CI-MS

GC-MS continues to play an important role in the identification and quantification of analytes. Several ionization techniques are also used in GC-MS. Among them, EI is the most popular because it often produces both molecular and fragment ions. One important feature of EI spectra is that they are highly reproducible, which means that mass spectral libraries can be used for the identification of unknowns. However, in some cases, EI does not provide the sensitivity required for the analysis of very small amounts of compounds in environmental samples. To solve this problem, softer ionization techniques such as CI are applied. Moreover, CI is the technique of choice for the analysis of isomers, because different isomers have different reactivities toward the reagent gas, resulting in different spectra. With EI, very similar spectra are obtained for different isomer compounds.

In this section, some recent environmental examples are presented in order to view the actual applicability of each one of the ionization techniques.

### Differentiation between Isomers

CI-MS is a powerful tool for steric structure identification that often allows differentiation between isomers. In addition, compound selectivity or sensitivity can be increased by choice of reagent gases with different proton affinities and ion-molecule reaction properties. However, in practice, only few reagent gases are used such as methane or isobutene. Occasionally, ammonia with a higher proton affinity than methane is employed to enhance selectivity. Methane and argon are most frequently used in electron capture negative ionization (ECNI).

The use of less frequently applied reagent gases can result in gas-phase ion-molecule reactions that lead to increased sensitivity, selectivity, or both for certain compounds. Such reactions can be observed in ion traps (IT) at reagent gas levels of a few hundreds of ppm due to the longer residence time of the ions. CI in conventional ion sources requires a higher reagent gas pressure in the range of 0.7–2.0 mbar (0.5–1.5 Torr). This does not allow the use of corrosive reagent gases or those leading to contamination by excessive polymerization reactions.

**Polychlorinated alkane determinations** NCI has frequently been used to identify chlorinated hydrocarbons and other chlorinated compounds such as toxaphene, chlorinated paraffins, chlorinated styrenes, chlorinated diphenylethers, pentachlorophenol, chlorinated pesticides, and pesticide metabolites. Even if the molecular ion cannot be detected in all

instances, the NCI spectra are usually much simpler than the EI spectra.

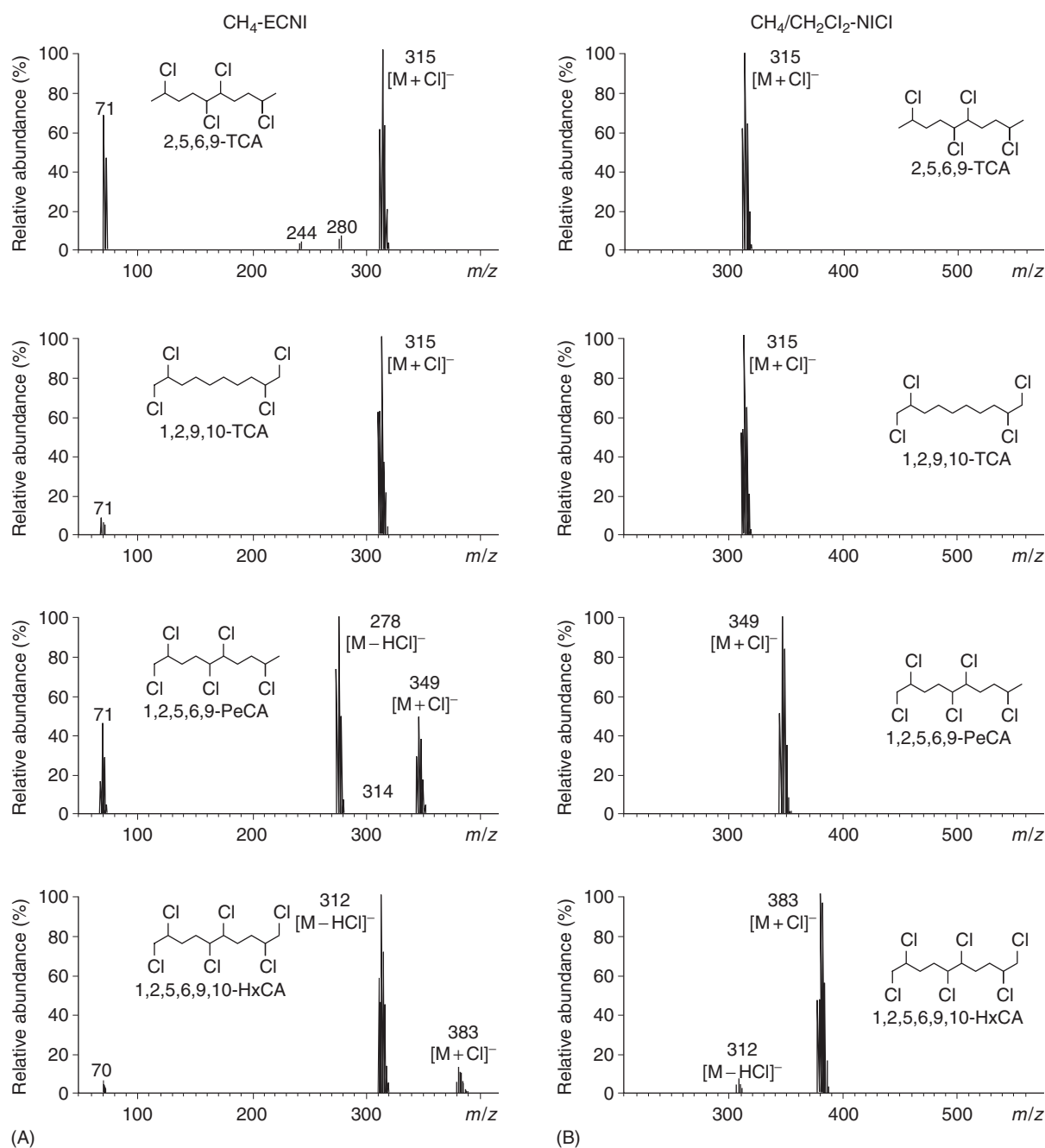
Chlorinated paraffins are complex technical mixtures consisting of polychlorinated alkanes (PCAs). The number of theoretically possible congeners, homologs, diastereomers, and enantiomers is huge (far beyond 10 000 compounds), which explains the complexity of the analysis. Most PCA analyses have been performed by GC-ECNI-MS. Methane or argon were mostly used as reagent gases. Under these conditions, the mass spectra of PCAs contain the three main ions  $[M - HCl]^-$ ,  $[M - Cl]^-$ , and  $[M + Cl]^-$ . The molecular ion  $[M]^-$  has generally an intensity of <10% (Figure 4). The use of dichloromethane/methane reagent gas mixtures is an alternative to conventional ECNI. A nearly exclusive formation of  $[M + Cl]^-$  adduct ions was observed suppressing the generation of other fragment ions and enhancing selectivity and sensitivity. Moreover, interferences from other chlorinated compounds present in environmental samples such as toxaphenes or chlordanes were suppressed.

**Polycyclic aromatic hydrocarbon determinations** Polycyclic aromatic hydrocarbons (PAHs) refer to a large and diverse group of a few hundreds of compounds with two or more fused benzene rings that are derived from endogenous and anthropogenic sources. A good number of them are known to cause cancer and genetic damage. PAHs are generally lipid-soluble, usually persistent, and in some cases bioaccumulating. Typically, these environmental pollutants arise from sources that subject organic material to high temperatures involving the incomplete combustion of such organic matter.

IT-MS combined with CI and full scan operation mode has not been frequently applied in the analysis of organic pollutants in the environment. Nevertheless, several works that compare CI and EI modes have been carried out. Methane is the most widely used chemical ionization reagent, although other compounds have also been used. For example, 1,1-difluoroethane was used as a reagent gas to differentiate PAH isomers through the formation of specific adducts by ion-molecule reaction between the cations  $CH_3CHF^+$  and  $CH_3CF_2^+$  and PAHs. This method allowed for the separation and determination of coeluting isomers.

**Lanolin steryl ester determinations** Lanolin is the wool grease secreted by the sheep sebaceous glands, and represents a complex mixture of high molecular mass lipids, including fatty acids and alcohols, sterols, hydroxyacids, diols, and aliphatic and steryl esters. Lanolin is widely used in cosmetic and pharmaceutical formulations. The study of minor lipids





**Figure 4** Mass spectra of single PCA congeners recorded with  $\text{CH}_4$ -ECNI (A) and  $\text{CH}_4/\text{CH}_2\text{Cl}_2$ -NICI (B). TCA, tetrachloroalkane; PeCA, pentachloroalkane; HxCA, hexachloroalkane. (Reprinted from Zencak Z, Reth M, and Oehme M (2003) Dichloromethane-enhanced negative ion chemical ionization for the determination of polychlorinated *n*-alkanes. *Analytical Chemistry* 75: 2487–2492; © American Chemical Society.)

has shown interest in order to assess the quality and authenticity of cosmetic and pharmaceutical products and also the steryl esters have already been used to prove the authenticity of eatable oil.

Those compounds have been reported as difficult to analyze due to their high molecular mass and poor information given by EI-MS. Also, the isomers *normal*, *iso*, and *anteiso* have been identified for each

family by means of chromatographic retention data. A comparison between different ionization techniques has been carried out in order to optimize target compound identification. EI-MS has shown to be the most sensitive but providing poor structural information for cholesteryl esters not permitting to identify the acid moiety or molecular ion. On the other hand, CI offered the best results in terms of

structural information obtained in the PCI with ammonia as reagent gas.

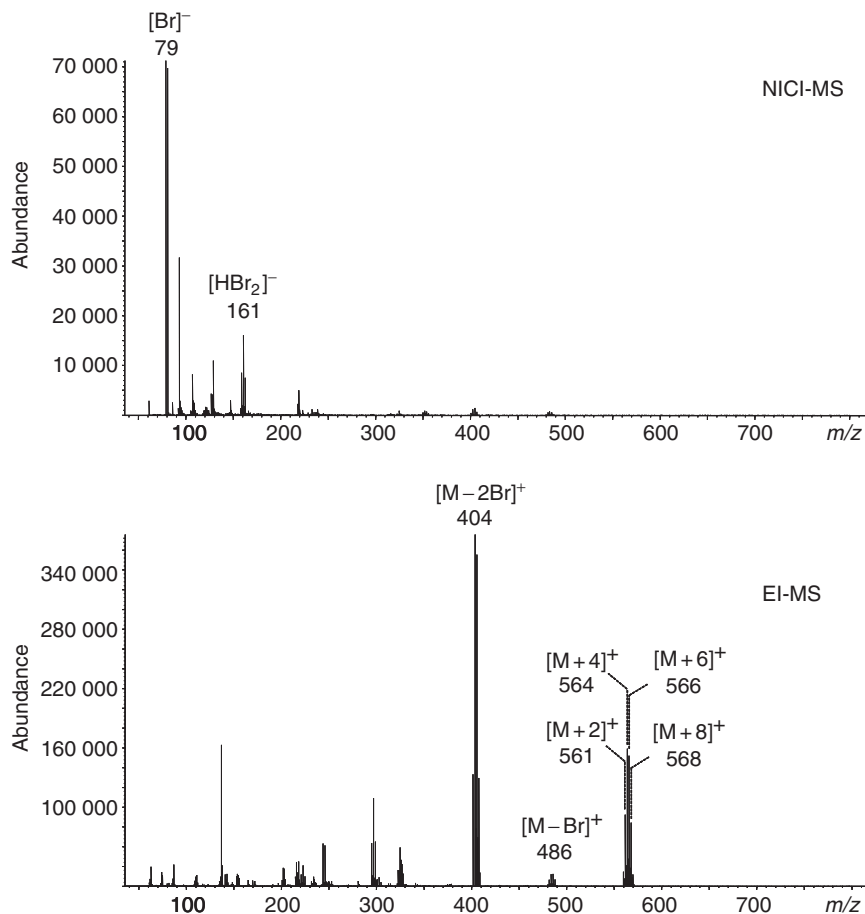
Several examples showed the suitability of CI versus EI in order to achieve a satisfactory separation between isomers. However, there are also some examples demonstrating that EI afforded better structural information. Fully acetylated methyl  $\alpha$ -deoxy- $\alpha$ -fluoro- $\alpha$ -D-glucopyranosides have been studied using EI and ammonia CI. Characteristic differences in the fragmentation of positional isomers were noted on analysis of the spectra, and these make it possible to determine the location of fluorine in the molecules. Fragmentation of  $[M - OCH_3]^+$  ions produced by EI provides an alternative method for localization of the fluorine atoms. On the contrary, spectra of  $[M + NH_4]^+$  cluster ions produced by CI did not afford such structural information.

### Environmental Organic Contaminants

**Phthalates** Phthalate acid esters (PAEs) are an estrogenic class of chemical compounds widely used in different industrial applications mainly as plasticizers

for polyvinyl chloride resins, adhesives, and cellulose film coatings and with minor applications in cosmetics, medical products, and insecticide carriers. They comprise a large group of compounds, several of them considered as priority pollutants: dimethyl, diethyl, dibutyl, butylbenzyl, di(2-ethylhexyl), and di-*n*-octyl phthalate. The worldwide production of PAEs  $\sim 2.7$  million metric tons a year and considerable direct (production of plastic materials) and indirect emission, via leaching and volatilization from plastic products after their usage, disposal, and incineration, explains their ubiquity in the environment.

GC-MS with EI and CI, using methane or isobutene as reagent gas has been used for the analysis of these compounds in water, soil, and sewage sludge samples. PCI with methane was found to be very useful in obtaining molecular mass information, but EI-MS is the most sensitive method. Using GC-MS in EI mode, limits of detection of  $2\text{--}30\text{ ng l}^{-1}$  in river and sea water,  $125\text{--}250\text{ ng per kg}$  in sediments, and  $10\text{--}600\text{ ng per g}$  in sewage sludge have been reported.



**Figure 5** Mass spectra of BDE-99 using NCI-MS and EI-MS approaches. (After E. Eljarrat, A. de la Cal, and D. Barceló (2003) Potential chlorinated and brominated interferences on the polybrominated diphenyl ether determinations by gas chromatography-mass spectrometry. *Journal of Chromatography A* 1008: 181–192).

**Polybrominated diphenyl ethers** Polybrominated diphenyl ethers (PBDEs) are used in large quantities for several applications due to their fire retarding properties, including electrical appliances such as television and computers, building materials, and textiles. The similarity in molecular structure of PBDEs with that of environmental toxic pollutants such as polychlorinated biphenyls and dioxins, and their resistance to degradative processes, gives rise to concern that they may lead to similar environmental problems. PBDEs are suspected to cause endocrine dysfunction by interfering with the thyroid hormone metabolism. PBDE comprises a total of 209 theoretically possible congeners. Qualitative and quantitative methods for PBDEs have been developed involving GC–NCI–MS or GC–EI–MS. Generally, NCI presents higher sensitivity than EI: NCI gave detection limits between 30 fg and 1.72 pg, whereas EI gave detection limits between 0.53 and 32.09 pg. The main advantage of EI–MS is that it provides better structural information. No structural information on the degree of bromination was obtained by NCI. The mass spectra of all PBDEs were dominated by the molecular ion  $[\text{Br}]^-$  and did not show any molecular ion. However, the EI provided better structural information, giving the molecular ions and the sequential losses of bromine atoms (Figure 5). Furthermore, it allows the quantification by isotopic dilution technique whereas NCI does not, making the analysis more reliable at trace levels.

**Selenium stable isotopes** Selenium is recognized as an essential trace element for humans. Se status is determined by dietary Se intake and its bioavailability. Se bioavailability can be estimated by assessing absorption and retention of Se stable isotopes. These studies require analytical techniques that allow precise and accurate determination of stable isotope ratios at low levels of total selenium.

Stable isotopes of Se were determined by different analytical techniques; however, GC–MS is probably the most widely used technique. GC–MS employs

selective derivatization of selenite, with an aromatic *o*-diamine to form in acidic medium a volatile piazselenole. The piazselenole is extracted into an organic solvent and analyzed by GC–MS. The analysis using EI as ionization mode was first described; however, NCI was also tested for this kind of determination. Analytical characteristics of the NCI method demonstrated the ability of this technique to accurately determine total Se and Se isotope ratios in biological samples. The detection limit for any selenium isotope was found to be 1 pg for NCI and 90 pg for EI. This technique can be applied to determine absorption and retention of Se in humans, and due to its lower detection limit compared to EI, NCI is preferential when the sample size is small (e.g., biological samples from infants).

*See also:* **Chromatography:** Overview; Principles. **Derivatization of Analytes.** **Endocrine Disrupting Chemicals.** **Environmental Analysis.** **Gas Chromatography:** Overview; Mass Spectrometry. **Mass Spectrometry:** Overview; Ionization Methods Overview; Environmental Applications. **Water Analysis:** Overview.

## Further Reading

- Aschroft AE (1997) *Ionization Methods in Organic Mass Spectrometry*. Cambridge: Royal Society of Chemistry.
- Gelpi E (2001) *Advances in Mass Spectrometry*, vol. 15. New York: Wiley.
- Harrison AG (1986) *Chemical Ionization Mass Spectrometry*. Boca Raton, FL: CRC Press.
- Petrovic M, Eljarrat E, López de Alda MJ, and Barceló D (2002) Review: recent advances in the mass spectrometric analysis related to endocrine disrupting compounds in aquatic environmental samples. *Journal of Chromatography A* 974: 23–51.
- Santos FJ and Galceran MT (2003) Modern developments in gas chromatography–mass spectrometry-based environmental analysis. *Journal of Chromatography A* 1000: 125–151.
- Vestel ML (2001) Methods of ion generation. *Chemical Reviews* 101: 361–376.

## Atmospheric Pressure Ionization Techniques

**A P Bruins**, University of Groningen, Groningen, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by D J Bell, pp. 2811–2822, © 1995, Elsevier Ltd.

## Introduction

Mass spectrometer ion sources are normally located inside a high-vacuum envelope. Such low-pressure ion sources, which can make use of a range of different ionization methods, are in routine use in

**Polybrominated diphenyl ethers** Polybrominated diphenyl ethers (PBDEs) are used in large quantities for several applications due to their fire retarding properties, including electrical appliances such as television and computers, building materials, and textiles. The similarity in molecular structure of PBDEs with that of environmental toxic pollutants such as polychlorinated biphenyls and dioxins, and their resistance to degradative processes, gives rise to concern that they may lead to similar environmental problems. PBDEs are suspected to cause endocrine dysfunction by interfering with the thyroid hormone metabolism. PBDE comprises a total of 209 theoretically possible congeners. Qualitative and quantitative methods for PBDEs have been developed involving GC–NCI–MS or GC–EI–MS. Generally, NCI presents higher sensitivity than EI: NCI gave detection limits between 30 fg and 1.72 pg, whereas EI gave detection limits between 0.53 and 32.09 pg. The main advantage of EI–MS is that it provides better structural information. No structural information on the degree of bromination was obtained by NCI. The mass spectra of all PBDEs were dominated by the molecular ion  $[\text{Br}]^-$  and did not show any molecular ion. However, the EI provided better structural information, giving the molecular ions and the sequential losses of bromine atoms (Figure 5). Furthermore, it allows the quantification by isotopic dilution technique whereas NCI does not, making the analysis more reliable at trace levels.

**Selenium stable isotopes** Selenium is recognized as an essential trace element for humans. Se status is determined by dietary Se intake and its bioavailability. Se bioavailability can be estimated by assessing absorption and retention of Se stable isotopes. These studies require analytical techniques that allow precise and accurate determination of stable isotope ratios at low levels of total selenium.

Stable isotopes of Se were determined by different analytical techniques; however, GC–MS is probably the most widely used technique. GC–MS employs

selective derivatization of selenite, with an aromatic *o*-diamine to form in acidic medium a volatile piazselenole. The piazselenole is extracted into an organic solvent and analyzed by GC–MS. The analysis using EI as ionization mode was first described; however, NCI was also tested for this kind of determination. Analytical characteristics of the NCI method demonstrated the ability of this technique to accurately determine total Se and Se isotope ratios in biological samples. The detection limit for any selenium isotope was found to be 1 pg for NCI and 90 pg for EI. This technique can be applied to determine absorption and retention of Se in humans, and due to its lower detection limit compared to EI, NCI is preferential when the sample size is small (e.g., biological samples from infants).

*See also:* **Chromatography:** Overview; Principles. **Derivatization of Analytes.** **Endocrine Disrupting Chemicals.** **Environmental Analysis.** **Gas Chromatography:** Overview; Mass Spectrometry. **Mass Spectrometry:** Overview; Ionization Methods Overview; Environmental Applications. **Water Analysis:** Overview.

## Further Reading

- Aschroft AE (1997) *Ionization Methods in Organic Mass Spectrometry*. Cambridge: Royal Society of Chemistry.
- Gelpi E (2001) *Advances in Mass Spectrometry*, vol. 15. New York: Wiley.
- Harrison AG (1986) *Chemical Ionization Mass Spectrometry*. Boca Raton, FL: CRC Press.
- Petrovic M, Eljarrat E, López de Alda MJ, and Barceló D (2002) Review: recent advances in the mass spectrometric analysis related to endocrine disrupting compounds in aquatic environmental samples. *Journal of Chromatography A* 974: 23–51.
- Santos FJ and Galceran MT (2003) Modern developments in gas chromatography–mass spectrometry-based environmental analysis. *Journal of Chromatography A* 1000: 125–151.
- Vestel ML (2001) Methods of ion generation. *Chemical Reviews* 101: 361–376.

## Atmospheric Pressure Ionization Techniques

**A P Bruins**, University of Groningen, Groningen, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by D J Bell, pp. 2811–2822, © 1995, Elsevier Ltd.

## Introduction

Mass spectrometer ion sources are normally located inside a high-vacuum envelope. Such low-pressure ion sources, which can make use of a range of different ionization methods, are in routine use in

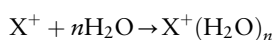
analytical mass spectrometers. An ion source operating at atmospheric pressure is better suited, if not essential for a growing number of applications. Whilst atmospheric pressure ionization (API) mass spectrometry (MS) was originally developed during the 1960s, it was the refinements made some 20 years later that initiated an explosion of interest in the technique. Early applications of API included direct monitoring of air for the mass-selective analysis of pollutants, explosives, and illegal drugs. The subsequent development of liquid introduction systems and modifications to the ions source, ion optics, and vacuum systems resulted in ionization techniques that were orders of magnitude more sensitive than established liquid-phase ionization techniques. In addition, these developments expanded the range of polarity and thermal lability of samples that could be analyzed. The major advantage to emerge was the increase in molecular mass range of samples amenable to MS by orders of magnitude from 15 kDa to beyond 1 MDa. Established chromatographic separation methods are now readily coupled to mass spectrometers with atmospheric pressure ion sources. As API opened new areas of mass spectrometric analysis, the technique rapidly became established, generating both qualitative and quantitative mass spectrometric data with unmatched sensitivity.

## Instrumentation

### The Combination of Ionization at Atmospheric Pressure with Separation of Ions in Vacuum

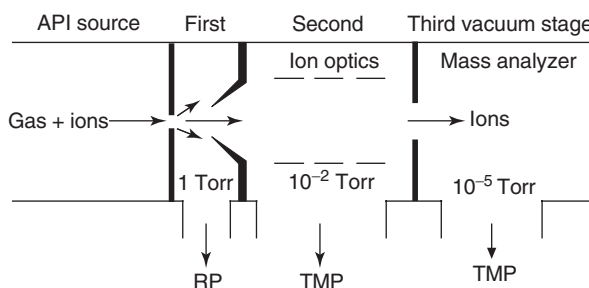
In a mass spectrometer that is used for organic analysis, the ion source is located inside a high-vacuum envelope. Such an ion source, which makes use of electron impact or chemical ionization, can conveniently and efficiently emit sample ions into a mass analyzer, which also operates under a high vacuum. The combination of an API source with a mass analyzer requires a  $10^7$ - to  $10^8$ -fold pressure reduction along the ion passageway between the source and the mass analyzer, a demanding task for the vacuum pumping system. The most common layout of an API mass spectrometer is shown in Figure 1.

When a gas expands into a vacuum, strong cooling takes place. This effect is widely used for the spectroscopic study of 'cold' molecules. When a mixture of ions and gas and water or solvent vapor expands into vacuum the cooling in the expansion region will promote the formation of ion-solvent clusters that may be so large that they fall outside the mass range of the mass analyzer:

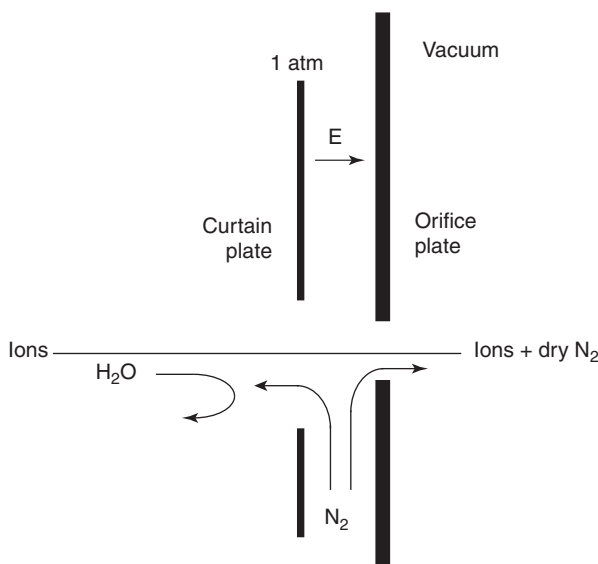


Different instrument designs address this clustering problem in different ways: by keeping water and solvent vapor away from the vacuum system by a gas curtain or countercurrent flow of nitrogen, by keeping the source at an elevated temperature, or by breaking the initially formed clusters.

Figure 2 depicts a typical arrangement of an ion source equipped with a gas curtain. The space between the curtain plate and the orifice plate is at atmospheric pressure and is flushed with dry nitrogen. Part of the nitrogen flows to the left into the ionization region of the source, sweeping neutrals such as water molecules or dust particles to the left. Under the influence of the electric field between the curtain plate and the orifice plate ions are moved to the right toward the orifice plate. There the ions are entrained in the flow of dry nitrogen into the vacuum system. Since water and other polar molecules



**Figure 1** API source and vacuum system of a mass spectrometer. RP, rotary vacuum pump; TMP, turbomolecular vacuum pump. (Reprinted with permission from *Trends in Analytical Chemistry* 13 (1994), 37; © Elsevier.)



**Figure 2** Curtain gas arrangement for the prevention of solvent vapor from entering the vacuum envelope of the mass spectrometer.



cannot penetrate the opening in the curtain plate, the formation of ion–solvent clusters is prevented.

The undesired formation of ion–solvent clusters is due to the temperature drop in the expansion into the vacuum system. By heating the source, or using of a heated tube for the passage of sample ions from the source into the vacuum, the temperature in the expansion region will not reach such a low value, and ions may cluster with only a few molecules, which can be removed in the expansion region by moderately energetic collisions.

### Mass Analyzers

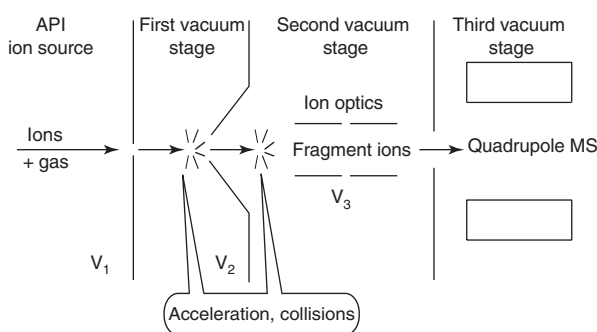
Initial API studies were performed on quadrupole instruments with either single- or triple-stage analyzers and the first commercial instruments were also of this form. The next most popular combination is API with an ion trap mass analyzer. Time-of-flight (TOF) mass analyzers offer medium resolution that is sufficient for accurate mass determination in many cases. TOF also allows the observation of high  $m/z$  ions generated from noncovalently bound complexes of proteins. For high-resolution API, the Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer is an excellent instrument. Magnetic sector mass spectrometers have found limited application in API, because of the technical complexity of the combination of an API source with the necessary high accelerating voltage. The interest in magnetic sector instruments with an extended mass range that were developed for fast atom bombardment and field desorption has been reduced since the advent of electrospray. High-molecular-mass materials can readily be detected either by electrospray ionization on quadrupoles and TOF analyzers, or by matrix-assisted laser desorption ionization (MALDI).

### Tandem Mass Spectrometry (MS/MS)

Full-scan API spectra can be accumulated to generate molecular mass information, but in order to generate fragment ions of significant intensity it is usually necessary to use one of two commonly applied methods.

First, if a multiple mass analyzer instrument is used, the pseudomolecular ion can be collided with a stream of gas to generate collisionally induced dissociation (CID) and the formation of structurally significant fragment ions. MS/MS with selected reaction monitoring in a triple quadrupole is the method of choice for quantitation. Structure determination or confirmation is done by means of any of the available MS/MS fragmentation techniques: triple quadrupole, ion trap, quadrupole-TOF, or FT-ICR.

Second, an alternative method for the generation of fragmentation is to increase the electric field



**Figure 3** Collision-induced dissociation of ions inside the first and second vacuum stages of an API mass spectrometer.

strength between elements of the ion path inside the vacuum regions between the atmospheric pressure source and the mass analyzer (Figure 3). This ‘up front’ CID method is oftentimes incorrectly called ‘in source CID’. It can have the dramatic effect of changing the spectrum from being dominated by ions related to the intact analyte to one containing many intense fragment ions. If spectra are acquired at each of two potentials the spectra will be complementary. This ‘poor man’s MS/MS’ technique is of limited application for direct mixture analysis, but a single mass analyzer can be operated very effectively in this way during liquid chromatography–mass spectrometry (LC–MS) analysis to generate fragmentation that can be related to a particular component.

If ‘up front’ fragmentation is used in combination with real MS/MS instrumentation, the fragments of ions generated just outside the source can be fragmented again in the collision cell, resulting in spectra of MS/MS/MS type.

The most widely used method of achieving multiple, consecutive fragmentations is the repeated isolation and fragmentation inside a quadrupole ion trap.

### Techniques for Ionization at Atmospheric Pressure

The considerations given above about ion transport, declustering, fragmentation, and mass separation are independent from the way ions are generated in the API source.

Sample ions can be formed in the gas phase by the ionization of sample vapor, or can be taken from the condensed phase, which is either the liquid phase (sample in solution) or the solid phase (sample in a solid matrix).

#### Atmospheric Pressure Chemical Ionization

The development of API has yielded atmospheric pressure chemical ionization (APCI) with unique

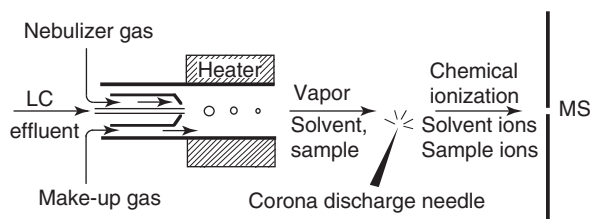
application in high-sensitivity analysis of organic compounds. Air monitoring was one of the original fields of application of APCI-MS. Suppliers of research-grade gases use APCI for the determination of impurities at trace levels. APCI is a gas-phase ionization technique that makes use of ion-molecule reactions between reactant ions and sample molecules. Early APCI-MS instrumentation used a  $^{63}\text{Ni}$   $\beta$ -emitter or a corona discharge to ionize the sample. Modifications to the corona discharge system have produced the modern APCI instrumentation. Problems of the addition of solvent ions to the analyte to form ion clusters were alleviated by the use of a stream of gas before the mass analyzer vacuum and the application of electric field to induce collisional decomposition of such clusters (Figures 2 and 3). A further refinement of a heated nebulizer resulted in a practical, robust ionization technique capable of accepting flow rates up to  $2\text{ ml min}^{-1}$  with detection limits often in the femtogram region (Figure 4).

Ionization in positive-ion mode occurs by reaction of the solute with protonated solvent molecules, generally giving rise to an intense protonated analyte molecule or pseudomolecular ion. If ammonium salts are added to the solvent, ammonium adducts may be observed. This can be exploited to yield molecular mass information in cases where the protonated molecule readily fragments or there is an absence of basic sites at which protonation can occur.

Ions are generated in the negative-ion mode by proton abstraction by oxygen ions or by the formation of adducts with anions such as acetate or chloride present in the sample.

APCI is a gas-phase ionization technique, which requires that a sample is completely evaporated before ionization can take place. Thus, APCI can be applied to samples that do not decompose upon heating during evaporation.

The impact of APCI has been enhanced by its compatibility with separation techniques routinely used in the analysis of organic compounds.



**Figure 4** Atmospheric pressure chemical ionization source with corona discharge needle and heated nebulizer interface for combined liquid chromatography-mass spectrometry. (Reprinted with permission from *Trends in Analytical Chemistry* 13 (1994), 81; © Elsevier.)

## Atmospheric Pressure Photoionization

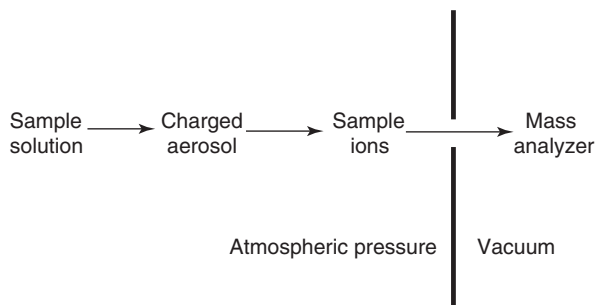
APCI is mostly done by means of protonation or deprotonation. If gas-phase basicity or acidity is too low for these acid-base reactions, the association with ammonium ions, formate ions, or acetate ions can be used to generate sample ions. Many samples do not have polar functional groups required for such ion-molecule reactions.

High-energy photons can remove an electron from a sample molecule and generate a molecular ion. Inexpensive gas discharge lamps, originally developed for photoionization detection in gas chromatography, are suitable sources of photons, e.g., 10.0 eV photons from a krypton discharge lamp.

Since most solvents used in reversed-phase LC have ionization energies above 10.0 eV, while many organic molecules of interest have an ionization energy below 10.0 eV, it is possible, in principle, to selectively ionize a sample. Atmospheric pressure photoionization (APPI) is a new technique first published in 2000, but has already been developed into a commercial product. APPI is a gas-phase ionization technique, which requires that a sample is completely evaporated before ionization can take place. Thus, APPI can be applied to samples that do not decompose upon heating during evaporation.

## Electrospray

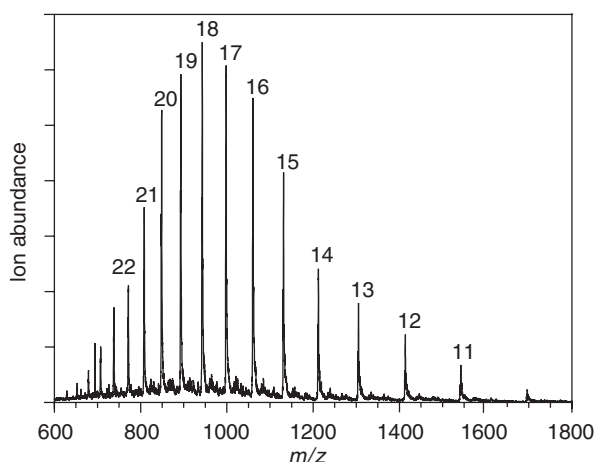
In parallel with the development of APCI, investigations were being made into a process by which a solution is sprayed into a high electric field at atmospheric pressure (Figure 5). Charged droplets result from the nebulization of the solution in an electric field, with both solvent and solute ions being detected resulting in the technique referred to as electrospray ionization (ESI). ESI is a liquid-phase ionization technique. It does not involve the evaporation of a neutral sample, so the input of heat is not required. Therefore, ESI is the method of choice for the ionization of samples that would decompose under the input of heat for evaporation. Because ESI



**Figure 5** Principle of electrospray ionization inside an atmospheric pressure ion source.

makes use of sample ions in solution, the composition of the sample solution has to be adjusted in order to convert a neutral sample into a sample ion in solution. Basic samples, for example, are ionized by the addition of an acid to the solution.

Refinement by the use of a nebulizing gas and a cross-flow of hot gas allows high liquid flow rates to be used. The high sensitivity of the technique and compatibility with liquid separation methods gave rise to much interest in the pharmaceutical applications of the technique, but an explosion of interest was initiated by the observation that high-molecular-mass compounds of biological interest could be ionized with great efficiency and readily detected on instrumentation previously considered to be of low mass range. This was possible as larger molecules were subject to multiple protonation or deprotonation. Thus, whilst a peptide of molecular mass 1000 Da may be protonated once, a 3000 Da sample may be protonated three times, and a protein of molecular mass 17 000 Da 17 times (Figure 6). The resulting mass-to-charge ratios, as recorded by mass analyzers, are below  $m/z$  2000 and within the mass range of a medium-specification instrument. In addition to increasing sensitivity, electrospray effectively extended the mass range of materials amenable to MS by more than an order of magnitude to beyond 150 kDa. Subsequent work was concentrated in two fields of applications: qualitative and quantitative analysis of small molecules in the pharmaceutical industry and the newly accessible area of protein analysis. It is now apparent that electrospray, APCI, and APPI are inherently more sensitive and reliable than thermospray and fast atom bombardment. The latter two methods have essentially become obsolete for LC-MS.



**Figure 6** Electrospray mass spectrum of the protein myoglobin, FW 16950. The numbers at the peaks denote the number of charges on the ions.

### Nanoelectrospray (Nanospray)

In its original form, ESI was suitable for liquid flow rates between 1 and  $10 \mu\text{l min}^{-1}$ . Pneumatic assistance allowed flow rates up to  $50 \mu\text{l min}^{-1}$  with operation at room temperature. A further increase of flow rate requires the input of heat for the desolvation of charged droplets in order to promote the release of free sample ions.

An unexpected feature of ESI is its behavior as a concentration-sensitive detector in LC. The ion signal in ESI depends on sample concentration, not on the amount of sample per unit time passing through the source. This implies that a reduction of liquid flow rate of a given sample solution does not decrease the signal. However, by working at  $50 \text{ nl min}^{-1}$  instead of  $5 \mu\text{l min}^{-1}$  one can spend 100 times as much time on recording mass spectra from the same volume of sample solution. Thus, one can increase sensitivity by multichannel averaging of many spectra, or perform a number of different fragmentation experiments.

ESI hardware has to be adapted for low flow rate operation. The so-called nanospray systems are commercially available and can be used both for infusion of a small volume of a sample solution, and for the combination with microbore and nanobore HPLC columns.

Nanospray is used extensively in the field of proteomics.

### Atmospheric Pressure MALDI

Because of the large installed base of API instruments there is an opportunity to try and use the source and mass analyzer for new techniques and applications. One such technique is MALDI, which is normally combined with TOF mass separation under vacuum. AP-MALDI can be combined with ion traps and quadrupoles to give a low-cost expansion of the capabilities of those instruments, in particular in laboratories that cannot justify the purchase of several instruments dedicated to a specific task. AP-MALDI on an ion trap is a cost-effective way of obtaining fragmentation in combination with MALDI.

## Interfacing Separation Sciences to API Techniques

### APCI

Of the API techniques, APCI is readily amenable for online mass spectrometric detection with LC separation methods. The heated nebulizer interface can accommodate aqueous mobile phases of up to 100% water at flow rates of up to  $2 \text{ ml min}^{-1}$ . As volatile buffers can be used with the interface, many LC

separation methods can be employed without modification. Low levels of nonvolatile additives in the eluent can be tolerated for a limited time, but not in long series of samples in LC-MS/MS quantitation. Judicious choice of solvents and additives will enhance sensitivity of APCI, or at least avoid a serious loss of efficiency of sample ionization. Relative gas phase basicity and acidity of sample and solvents should be taken into account.

If thermal degradation occurs one has to use ESI instead of APCI. Supercritical fluid chromatography can also be readily linked to APCI without major modification to the chromatograph or the separation method as the ion source is designed to operate at high pressure.

### APPI

The next gas-phase ionization technique for LC-MS is APPI. It is a relatively new method, and its strengths and weaknesses will become clear as the number of applications grows over time. Thus far APPI appears to work best at flow rates of 100–200  $\mu\text{L min}^{-1}$ . At higher flow rates ionization efficiency is reduced. To increase sensitivity a dopant such as acetone, toluene, or anisole is introduced into the ionization region. First, the dopant is ionized by photons and next the dopant ions undergo ion-molecule reactions with ultimate ionization of the sample. Solvent components can have a major influence on the cascade of ion-molecule reactions, and should be selected with care.

### Electrospray

As electrospray can be used without the input of heat, this is the method of choice for the detection of thermally labile materials resolved by separation techniques. Flow rates of up to 50  $\mu\text{L min}^{-1}$  can be taken into the pneumatically assisted ESI interface at room temperature. Ion sources capable of operating with up to 1  $\text{mL min}^{-1}$  include a cross-flow of hot gas. A simple splitting device can be used at high flow rate conditions. The excess flow can be diverted for online detection using an alternative detection method or a fraction collector can be used to allow offline studies. Although 100% aqueous solutions can be sprayed there is loss in sensitivity; this can often be avoided by the use of a postcolumn make-up solvent containing a high proportion of organic liquid.

The fact that the interface can operate at low flow rates is an advantage in that micro-LC can be directly coupled for situations where the amount of sample is limited. Suitability for low flow rate studies enables capillary electrophoresis (CE) to be interfaced to the electrospray ion source. Despite the low sample

loadings associated with CE, selected ion monitoring detection of compounds can be performed with an increasing number of analyses involving the acquisition of full scan spectra. This is clearly an immense aid to CE studies as the nature of the separation precludes the use of most other qualitative techniques from both on- or offline analysis.

## Applications of APCI, APPI, and ESI

### Analytical Biotechnology

The greatest impact of electrospray has been in analytical biotechnology, as this is a new area of mass spectrometric study rather than an existing area in which sensitivity has been increased. The electrospray spectrum of a protein consists of an envelope of ions originating from protein molecules possessing different numbers of protons and hence mass-to-charge ratios. The number of charges on a molecule is linked to the number of basic sites and hence to the molecular mass. An injection of 5 pmol of protein dissolved in 10  $\mu\text{L}$  of solvent will produce an intense spectrum. In order to calculate the molecular mass of the protein, one of several algorithms can be used. The simplest involves the calculation of the molecular mass from the mass-to-charge ratios of two adjacent charge states. If the lower charge state is  $n$ , then the mass-to-charge ratio ( $m_1$ ) of this lower charge state is

$$m_1 = \frac{M + n}{n}$$

where  $M$  is the molecular mass of the compound.

For the higher charge state ( $n + 1$ ), the mass-to-charge ratio  $m_2$  is given by

$$m_2 = \frac{M + (n + 1)}{(n + 1)}$$

Then the number of charges and the molecular mass can be calculated from these two equations with two unknowns. This takes account of the charging species, usually protons. The system is self-checking as several charge states are observed and each adjacent pair will produce a mass measurement. The complete mass calculation process is normally performed by the mass spectrometer data system. Contamination of a protein with buffer salts or with detergents is a serious problem in electrospray. MALDI is more flexible in this respect.

Although of great value, accurate molecular mass measurement is only part of the protein characterization process. In order to confirm a protein sequence the material can be digested to produce peptides that can in turn be characterized by MALDI

or LC–MS. Electrospray is the preferred option for molecular mass determination or, if used in conjunction with LC–MS/MS, for peptide sequencing. In addition to amino acid sequence information, electrospray can be used to improve the detection and characterization of post-translational modifications including N- and C-terminal derivatization, phosphorylation, and, by analysis of peptide with and without a reducing agent, disulfide bonding. Using a combination of traditional carbohydrate analysis techniques and electrospray, the site and nature of protein glycosylation may also be determined.

Being a very soft ionization technique, electrospray does not disrupt noncovalent bonds in complexes in solution. Complexes of an enzyme with a substrate, enzyme with inhibitor, and conglomerates of proteins can be detected by ESI. This development has led to the examination of protein–protein interactions by means of ESI MS.

Conformational changes of proteins may be observed by either studying the number of protons that can be placed on a protein or by determining the number of exchangeable protons by also acquiring data in deuterated solvents and subtracting the calculated molecular masses. This can also be performed in real time, allowing the observation of conformational changes as they occur in biochemical systems.

### **Drug Discovery**

Many applications in drug discovery have evolved from methodology originally developed in analytical biotechnology. In addition to the mass spectrometric analysis of potential therapeutic agents, drug targets may be observed directly by electrospray, which can be exploited to investigate the mode of action of drugs with the aim of designing more effective products. This application includes the measurement of enzyme inhibition or the interaction of a drug with a receptor by simply observing the increase in molecular mass using ESI upon incubation of the enzyme or receptor protein with an appropriate inhibitor or drug. As data can be acquired in a matter of minutes, enzyme–inhibitor or receptor–drug reactions may be studied in real time and thus kinetic data can be derived by plotting the concentration of components against time.

### **Drug Development**

The increased sensitivity available from API techniques, either electrospray or APCI, has enabled lower levels of impurities found in drug substances, or their intermediates, to be characterized by LC–MS without time-consuming isolation of components of

interest. This information in turn allows the modification and optimization of a process to yield the pure drug substance.

### **Drug Quantitation and Metabolism Studies**

Again, the sensitivity of API techniques and the range of samples amenable has been exploited, allowing the detection, characterization, and quantification of drugs and metabolites present in amounts orders of magnitude less than earlier mass spectrometric methods. APCI, ESI, and recently APPI are now used for such ADME (administration, distribution, metabolism, excretion) studies. Although either ionization technique can be used in many cases, the preferred method depends on the type of material to be analyzed; thermally labile materials show molecular mass information by electrospray whereas many small polar molecules show greatest response by APCI or APPI. Probably the majority of triple quadrupole LC–MS/MS instruments in the field is in use for quantitation of drugs and metabolites. Quantitative LC–MS/MS information exceeding the limits of detection of LC with UV absorption is acquired by selected reaction monitoring. Contract labs perform a major part of such quantitations.

### **Natural Product Structure Elucidation**

The sensitivity of API techniques and the compatibility with LC make API ideal for studies when the amount of sample available is limited. In addition to molecular mass determination, MS/MS spectra containing structurally significant data can be generated from sub-picomole amounts of sample. The wide range of materials amenable to electrospray in terms of polarity, molecular mass, and thermal lability endorse electrospray as a first choice for such investigations. Given appropriate chromatographic separations, LC–MS with one of the API techniques can be used as a routine, rapid screening device with the searching of databases by molecular mass and ultraviolet absorption (also acquired on-line) to determine readily whether the components of interest have been previously observed. Chemical techniques such as derivatization or hydrolysis can be used with LC–MS and LC–MS/MS to yield further information relating to the structure of the material of interest.

*See also:* **Drug Metabolism:** Metabolite Isolation and Identification; Isotope Studies. **Liquid Chromatography:** Liquid Chromatography–Mass Spectrometry. **Mass Spectrometry:** Ionization Methods Overview; Matrix-Assisted Laser Desorption Ionization; Mass Separation;



Ion Traps; Multidimensional; Archaeological Applications; Clinical Applications; Environmental Applications; Food Applications; Forensic Applications; Peptides and Proteins. **Proteins:** Overview. **Proteomics.**

## Further Reading

- Bruins AP (1991) Mass spectrometry with ion sources operating at atmospheric pressure. *Mass Spectrometry Reviews* 10: 53–77.
- Cole RB (ed.) (1997) *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation & Applications*. New York: Wiley.
- James P (ed.) (2001) *Proteome Research: Mass Spectrometry*. Berlin: Springer.
- Lee MS (2002) *LC/MS Applications in Drug Development*. New York: Wiley.
- Robb DB, Covey TR, and Bruins AP (2000) Atmospheric pressure photoionization: an ionization method for liquid chromatography-mass spectrometry. *Analytical Chemistry* 72: 3653–3659.
- Willoughby R, Sheehan E, and Mitrowich S (2002) *A Global View of LC/MS*, 2nd edn. Pittsburgh, PA: Global View Publishing.
- Siuzdak G (2003) *The Expanding role of Mass Spectrometry in Biotechnology (preliminary information found on www.lcms.com)*. San Diego: MCC Press.
- <http://webbook.nist.gov>—A compilation of ion energetics (e.g., ionization energy, proton affinity, gas-phase basicity) of commonly used gases, solvents, and many organic molecules.

## Electrospray

**W M A Niessen**, Hyphen MassSpec Consultancy, Leiden, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

In the past 15 years, electrospray ionization has become one of the most important ionization methods in mass spectrometry. It is the method of choice in the coupling of liquid chromatography (LC) and mass spectrometry (MS). By estimation, it is used in over 90% of all LC–MS applications. It is especially useful in the analysis of highly polar, ionic, and macromolecular analytes. In addition, electrospray ionization plays an important role in the characterization of biomacromolecules, especially peptides and proteins. History, principle, instrumentation, practical aspects, and application of electrospray ionization are discussed in this article.

## Principle of Operation

The use of electrospray nebulization and ionization in MS is based on atmospheric pressure ionization. The column effluent from a reversed-phase LC, i.e., a solvent mixture of methanol or acetonitrile and up to  $10 \text{ mmol l}^{-1}$  aqueous buffer, is nebulized into an atmospheric-pressure ion source region. Because pure electrospray nebulization can only be achieved at low flow rates, i.e., typically up to  $10 \mu\text{l min}^{-1}$ , pneumatically assisted electrospray is performed in most cases: the liquid flow is nebulized into small droplets by a combined action of a strong electric

potential between needle and counter-electrode, e.g., 3 kV, and a high-speed concurrent nitrogen flow ( $50\text{--}100 \text{ l h}^{-1}$ ). The electrospray nebulization process results in the formation of small charged droplets, positively charged droplets when the source is operated in positive-ion mode and negatively charged droplets in the negative-ion mode. In their flight between needle and counter-electrode, neutral solvent molecules will evaporate from the surface of the charged droplets. The evaporation of the solvent results in a decreasing droplet size, which in turn leads to shorter distances between the charges at the surface of the droplets. When the surface tension of the liquid can no longer accommodate the Coulomb repulsion between the charges, the droplets will disintegrate by means of a so-called Coulomb explosion or field-induced electrohydrodynamic disintegration process. This leads to the emission of a number of microdroplets with a radius of less than 10% of that of the initial droplets. Gas-phase analyte ions are generated from these microdroplets by a variety of processes, including ion evaporation, soft desolvation, and chemical ionization at the droplet surface (see below). Subsequently, the gas-phase ions can be mass analyzed.

## History of Electrospray Ionization

Although electrospray nebulization was already described in 1917 by Zeleny, the first applications of electrospray in MS date from the late 1960s, when the group of Dole investigated the transfer of macromolecules from the liquid phase to the gas phase by electrospraying dilute solutions in a nitrogen bath

Ion Traps; Multidimensional; Archaeological Applications; Clinical Applications; Environmental Applications; Food Applications; Forensic Applications; Peptides and Proteins. **Proteins:** Overview. **Proteomics.**

## Further Reading

- Bruins AP (1991) Mass spectrometry with ion sources operating at atmospheric pressure. *Mass Spectrometry Reviews* 10: 53–77.
- Cole RB (ed.) (1997) *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation & Applications*. New York: Wiley.
- James P (ed.) (2001) *Proteome Research: Mass Spectrometry*. Berlin: Springer.
- Lee MS (2002) *LC/MS Applications in Drug Development*. New York: Wiley.
- Robb DB, Covey TR, and Bruins AP (2000) Atmospheric pressure photoionization: an ionization method for liquid chromatography-mass spectrometry. *Analytical Chemistry* 72: 3653–3659.
- Willoughby R, Sheehan E, and Mitrowich S (2002) *A Global View of LC/MS*, 2nd edn. Pittsburgh, PA: Global View Publishing.
- Siuzdak G (2003) *The Expanding role of Mass Spectrometry in Biotechnology (preliminary information found on www.lcms.com)*. San Diego: MCC Press.
- <http://webbook.nist.gov>—A compilation of ion energetics (e.g., ionization energy, proton affinity, gas-phase basicity) of commonly used gases, solvents, and many organic molecules.

## Electrospray

**W M A Niessen**, Hyphen MassSpec Consultancy, Leiden, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

In the past 15 years, electrospray ionization has become one of the most important ionization methods in mass spectrometry. It is the method of choice in the coupling of liquid chromatography (LC) and mass spectrometry (MS). By estimation, it is used in over 90% of all LC–MS applications. It is especially useful in the analysis of highly polar, ionic, and macromolecular analytes. In addition, electrospray ionization plays an important role in the characterization of biomacromolecules, especially peptides and proteins. History, principle, instrumentation, practical aspects, and application of electrospray ionization are discussed in this article.

## Principle of Operation

The use of electrospray nebulization and ionization in MS is based on atmospheric pressure ionization. The column effluent from a reversed-phase LC, i.e., a solvent mixture of methanol or acetonitrile and up to  $10 \text{ mmol l}^{-1}$  aqueous buffer, is nebulized into an atmospheric-pressure ion source region. Because pure electrospray nebulization can only be achieved at low flow rates, i.e., typically up to  $10 \mu\text{l min}^{-1}$ , pneumatically assisted electrospray is performed in most cases: the liquid flow is nebulized into small droplets by a combined action of a strong electric

potential between needle and counter-electrode, e.g., 3 kV, and a high-speed concurrent nitrogen flow ( $50\text{--}100 \text{ l h}^{-1}$ ). The electrospray nebulization process results in the formation of small charged droplets, positively charged droplets when the source is operated in positive-ion mode and negatively charged droplets in the negative-ion mode. In their flight between needle and counter-electrode, neutral solvent molecules will evaporate from the surface of the charged droplets. The evaporation of the solvent results in a decreasing droplet size, which in turn leads to shorter distances between the charges at the surface of the droplets. When the surface tension of the liquid can no longer accommodate the Coulomb repulsion between the charges, the droplets will disintegrate by means of a so-called Coulomb explosion or field-induced electrohydrodynamic disintegration process. This leads to the emission of a number of microdroplets with a radius of less than 10% of that of the initial droplets. Gas-phase analyte ions are generated from these microdroplets by a variety of processes, including ion evaporation, soft desolvation, and chemical ionization at the droplet surface (see below). Subsequently, the gas-phase ions can be mass analyzed.

## History of Electrospray Ionization

Although electrospray nebulization was already described in 1917 by Zeleny, the first applications of electrospray in MS date from the late 1960s, when the group of Dole investigated the transfer of macromolecules from the liquid phase to the gas phase by electrospraying dilute solutions in a nitrogen bath

gas. Their hypothesis was that gas-phase ions from macromolecules can be produced by desolvation of the charged droplets produced in electrospray, assuming these macromolecules exist as preformed ions in solution, e.g., positively charged proteins at a pH below their isoelectric point. This ionization mechanism is nowadays called the charge residue model.

After the work of Dole, where the sample solutions were sprayed in an atmospheric-pressure region, a variety of other related approaches have been described. In the mid-1970s, the group of Evans investigated the applicability of electrospraying solutions in vacuum. The method is called electrohydrodynamic ionization. As the result of the interaction of a strong electrostatic field with a liquid meniscus at the end of a capillary tube, solvated ions are emitted from the apex of a so-called Taylor cone, formed at the capillary tip. Due to the high-vacuum conditions, the use of nonvolatile organic solvents, e.g., glycerol doped with sodium iodide to increase the conductivity, is required for analyte introduction. Compounds investigated are saccharides, nucleosides, and small peptides. Ions are formed by attachment of a cation or an anion to the analyte molecules. In 1980, the group of Zolotai reported similar experiments using glycerol or water as solvent and introduced the term 'field evaporation of ions from solution'.

An important contribution to this field was made by Iribarne and Thomson in the late 1970s, who investigated the processes of the direct emission of ions from liquid droplets. In their experimental setup, a liquid solution is pneumatically nebulized in an atmospheric-pressure chamber and the droplets produced are charged by random statistical charging using an induction electrode positioned close to the nebulizer. Solvated singly charged ions are formed in the evaporating spray, which are sampled into a differentially pumped quadrupole mass spectrometer. Their theoretical description of the ionization process, i.e., the so-called ion-evaporation model, was adapted later by researchers in the field of thermospray and electrospray ionization.

The ion formation in thermospray ionization, introduced in the early 1980s by Vestal and co-workers, was explained in terms of ion evaporation. Preformed analyte ions are evaporated from small, fast-evaporating, charged droplets generated by the thermospray nebulization process. Although the importance of ion evaporation in thermospray ionization is questioned, the emphasis put at it at the time certainly stimulated further investigation into liquid-based ionization approaches for MS.

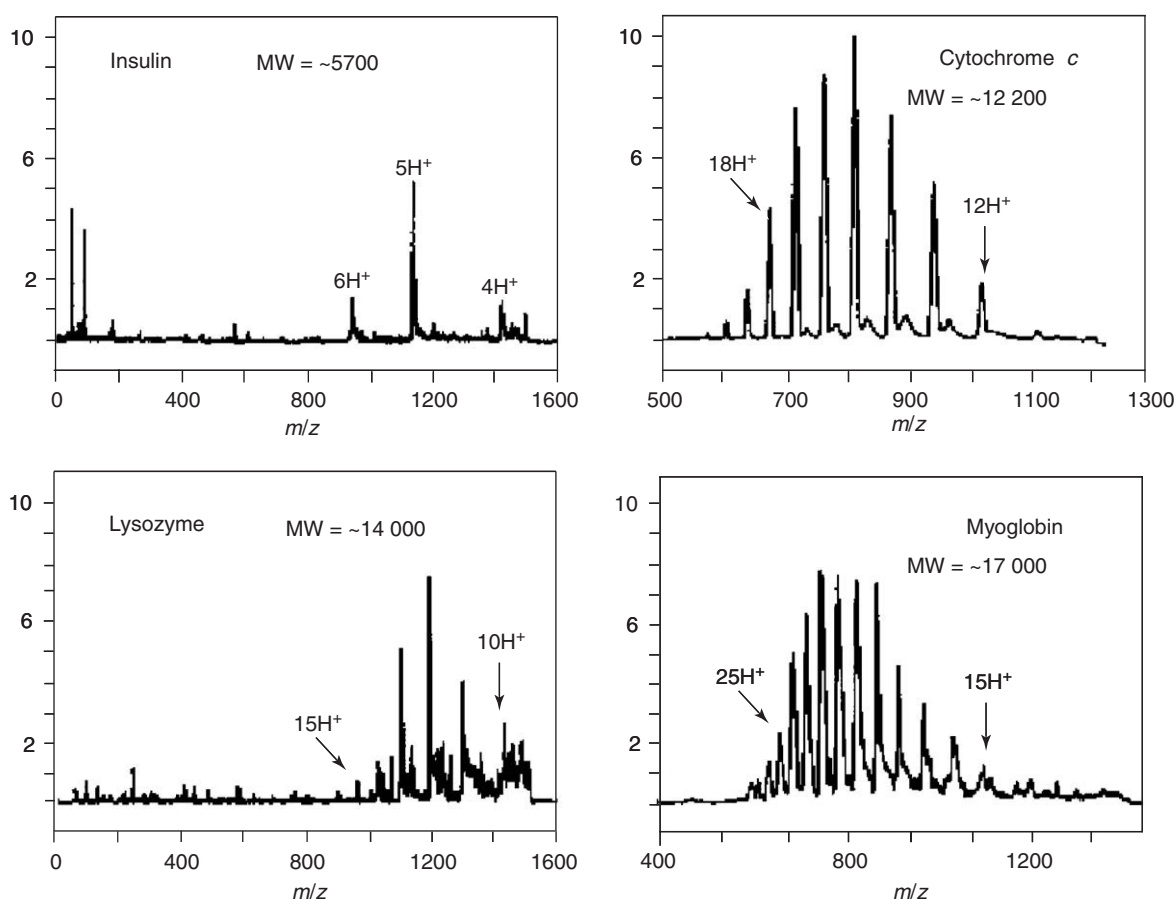
Within this context, Yamashita and Fenn, in the mid-1980s, continued the original electrospray

experiments of Dole and implemented molecular beam technology to solve some of the experimental difficulties experienced by Dole. The liquid is electrosprayed into a bath gas. The dispersion of ions, solvent vapor, and bath gas is expanded into a vacuum chamber, forming a supersonic jet, the core of which is sampled into the mass spectrometer by means of a skimmer. Initially, they tried in vain to reproduce the experiments of Dole with biomacromolecules, but nevertheless continued to investigate the sequence of electrospraying and droplet evaporation with low molecular mass compounds. They concluded that the hypothesis of Dole was not valid in their experiment and they recognized the importance of the work of Iribarne and Thomson in this field. The major breakthrough in this research was achieved in 1988, when the formation of multiply charged protein ions was demonstrated (Figure 1). This revolutionized the application of MS, because biomacromolecules could now be analyzed using simple and relatively low-cost quadrupole mass analyzers. In 2002, Fenn received the Nobel price for chemistry for his contribution to mass spectrometric analysis and characterization of biomacromolecules.

Although Fenn initially emphasized the importance of ion evaporation in ion formation by electrospray ionization, extensive research has demonstrated that, depending on the analyte and the experimental conditions, the charge-residue model and the ion-evaporation model are both important. In addition, especially in the analysis of small molecules, gas-phase ion-molecule reactions appear to play an important role in ion formation by electrospray ionization.

## Instrumentation

After the first demonstration of multiply charged gas-phase proteins ions, all major instrument manufacturers developed atmospheric-pressure ion sources, equipped with electrospray interfaces for both protein characterization and LC-MS applications. Within 5 years, electrospray interfacing became the method of choice in LC-MS coupling. It led to a large increase in the use of MS for the characterization and identification of labile and polar analytes as well as to routine quantitative analysis. The advent of electrospray ionization for peptide and protein analysis stimulated further development and analytical application of existing and new mass analysis approaches, such as quadrupole ion traps, Fourier-transform ion-cyclotron resonance MS, and quadrupole-time-of-flight hybrid instruments. It opened new application areas, such as proteomics. LC-MS

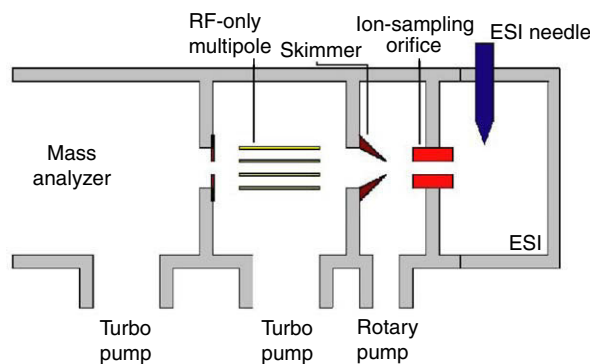


**Figure 1** First spectra demonstrating multiply charged proteins generated by electrospray ionization mass spectrometry. (Reprinted with permission from Mann M, Meng CK, and Fenn JB (1989) Interpreting mass spectra of multiply charged ions. *Analytical Chemistry* 61: 1702; © American Chemical Society.)

with electrospray ionization and in selected-reaction monitoring (SRM) mode in triple-quadrupole MS-MS instruments has become the method of choice in routine quantitative bioanalysis supporting preclinical and clinical studies within the pharmaceutical industry as well as in many other areas.

Since the first introduction of commercial electrospray interfaces for LC-MS, based on the Fenn design, continuous developments have taken place, resulting in more reliable and robust as well as more efficient interfaces. Some aspects of these developments are briefly discussed here.

An atmospheric-pressure ion source for electrospray ionization consists of five parts (**Figure 2**): (1) the pneumatically assisted electrospray needle, used for the introduction of sample solution or LC mobile phase; (2) the actual ion source region, where ions are generated from the microdroplets at atmospheric pressure; (3) the ion-sampling aperture; (4) the atmospheric-pressure to high-vacuum interface; and (5) the ion-optical system, where the ions generated in the source are analyte-enriched and transported toward the high-vacuum mass analyzer.



**Figure 2** Schematic diagram of an atmospheric-pressure ionization interface.

In principle, two types of electrospray interfaces must be distinguished. Most attention is paid to the high-flow devices applied for routine LC-MS. However, an extreme low flow electrospray needle device, i.e., the so-called nanoelectrospray, has been developed as well, especially by Mann in the mid-1990s. Nanoelectrospray is especially important in

the low nanoliter per minute introduction of protein and peptide solutions, but is increasingly important in the coupling to MS of micro- and nano-LC and capillary electromigration techniques, such as capillary electrophoresis and isoelectric focusing. Many precautions described below for the high-flow devices are not necessary in the low-flow devices.

Different instrument manufacturers have developed a wide variety of ion source and interface designs. Initially, the electrospray needle was positioned axial with the ion-sampling orifice. Due to the heavy use of electrospray LC-MS in high-throughput quantitative bioanalysis, alternative positions had to be investigated, in order to prevent a rapid contamination of the ion-sampling orifice by nonvolatiles in the liquid stream. The liquid introduction needle is nowadays positioned orthogonal to the ion-sampling orifice. In addition, a countercurrent nitrogen flow is applied around the ion-sampling orifice to further protect it against precipitation of nonvolatile contaminants. The increase in the flow rate of the liquid introduced by means of pneumatically assisted electrospray required the supply of heat to the spray in order to stimulate the solvent evaporation from the droplets. In most systems, this is achieved by the application of heated nitrogen streams, either concurrent, orthogonal to the spray, like in turbo-ion-spray, or countercurrent. In addition, the ion source block and ion-sampling orifice are heated.

The glass capillary with two-sided conductive layers, as introduced by Fenn, is still applied as the ion-sampling device in some of the commercial interfaces. However, a variety of other ion-sampling devices have been introduced, including orifices in flat plates and sample cones as well as heated stainless-steel capillaries.

While initially only one differentially pumped stage between atmospheric-pressure ion source and mass analyzer was applied, two or even three pumping stages are used in current systems. Significant effort has been put in increasing the pumping speed at the transition zone, partly in order to increase the size of the ion-sampling orifice. Significant progress has been made in turbomolecular pumps, enabling the use of small and relatively efficient dual-inlet pumps in small bench-top LC-MS systems. Skimmers are used between the various pumping regions. In the relatively large pumping regions, evacuated by turbomolecular pumps, a variety of RF-only multipole devices are used, e.g., quadrupole, hexapoles, octapoles, and ion tunnels. These devices are used to transport the ions through these pumping regions and to focus them into the mass analyzer.

## Practical Aspects

Successful operation of the electrospray interface requires the optimization of a variety of interrelated ion-source parameters. In addition, the solvent composition and flow rate may influence the system performance.

### Ion-Source Parameters

The ion-source parameters and their optimum values depend on the specific electrospray ionization source applied, e.g., in some source designs the optimum voltage difference between needle and counter-electrode is around 4.5–5 kV, while in other designs the optimum is around 3 kV. Most systems nowadays allow for an automatic optimization of the voltages in the source and in the mass spectrometer. These procedures search for optimum ion production and transmission toward the detector. In addition, the ion transitions to be used in SRM in quantitative analysis can be optimized in this way. Although very useful, the procedures do not account for the important influence of mobile-phase composition and sample effects on the performance of the electrospray interface.

### Solvent Composition

Both the solvent composition and interfering contaminants in the sample solution can have distinct influence on the electrospray ionization efficiency for a particular analyte. In general, reversed-phase-type mobile phases are applied, i.e., mixtures of water and either methanol or acetonitrile containing an acid, base, or buffer. It is often found that most analytes show better response from solutions containing >50% organic solvent (methanol or acetonitrile in most cases). However, electrospray ionization is not possible from pure organic solvents. Obviously, the organic content of the mobile phase in LC-MS is more determined by retention characteristics in LC than by ionization and MS detection. It has been demonstrated for a large variety of compounds that a methanol-containing mobile phase leads to a better response than an acetonitrile-containing mobile phase. Postcolumn and/or sheath-flow addition of high-organic solvents, such as aqueous mixtures with 2-propanol or methoxyethanol, are sometimes applied, especially in the analysis of proteins by electrospray ionization MS.

### Electrolyte Additives

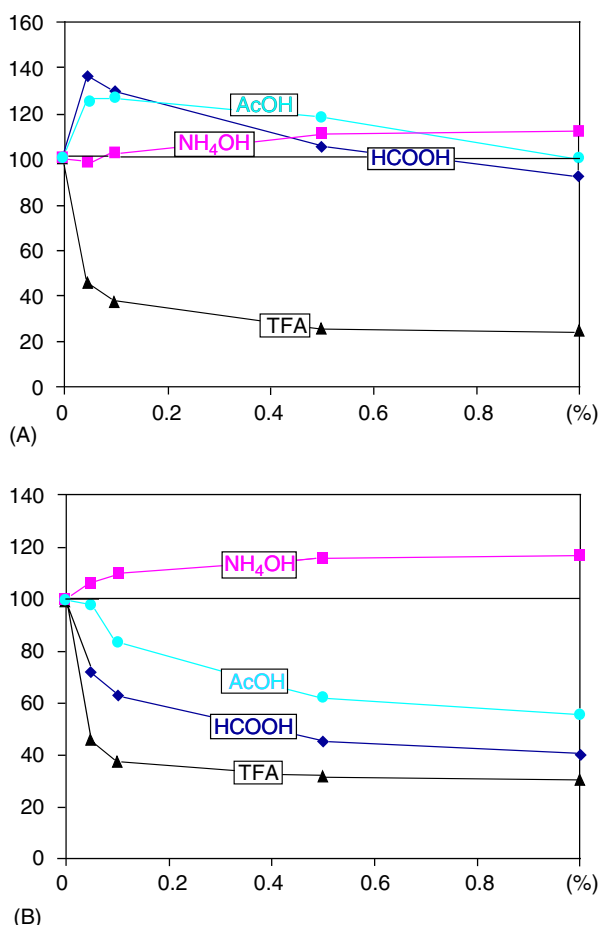
The composition, pH, and concentration of the buffer used also influences the performance of the electrospray interface. Based on the ionization



mechanisms, briefly outlined above, one would conclude that an acidic mobile phase is needed in the analysis of an organic base and a basic solution for an organic acid. In this way, the formation of liquid-phase preformed ions is promoted. However, in the electrospray MS analysis of especially small molecules (<1 kDa), the picture appears to be more complex. Wrong-way-around electrospray has been described, where organic bases are ionized from basic solutions in positive-ion mode without significant loss in response.

Although most of the current LC-MS interfaces can readily cope with nonvolatile mobile-phase additives such as sodium phosphate buffers, they should not be used in routine LC-MS applications. They may lead to more pronounced adduct formation and may lead to significant suppression of the response. Therefore, the use of volatile additives is recommended.

Recently, Mallet and co-workers investigated the influence of a number of mobile-phase additives on the response of eight basic drugs in positive-ion mode and eight acidic drugs in negative-ion mode. The influence of formic, acetic, and trifluoroacetic acids, and ammonium hydroxide, formate, biphosphonate, and bicarbonate were investigated as a function of additive concentration. Some of their results are shown in Figure 3. The ionization of all compounds, both basic and acidic ones, was found to be suppressed by the addition of strong ion-pairing agents like trifluoroacetic acid or nonafluoropentadecanoic acid. Most compounds respond similarly to the addition of acetic acid and formic acid, although the effect is compound dependent. In positive-ion mode, some compounds (propranolol, trimethoprim, terfenadine, methoxyverapamil, and reserpine) show ion enhancement at acid concentrations below 0.5%, while others show ion suppression (risperidone and benextramine). Somewhat surprisingly, the response of some bases (risperidone, terfenadine, methoxyverapamil, benextramine) is enhanced by addition of ammonium hydroxide. The quaternary ammonium compound piperizolate is not significantly influenced by additions of ammonium hydroxide or formic or acetic acids. Most basic compounds are suppressed by the addition of ammonium formate, bicarbonate, and biphosphonate. In negative-ion mode, most compounds are ion suppressed by formic and acetic acid additives. Ammonium hydroxide results in (significant) ion enhancement for cholic acid, raffinose, canrenoic acid, while fumaric acid, malic acid, and etidronic acid are strongly suppressed. The response of cholic acid, raffinose, and canrenoic acid is also enhanced by ammonium formate and bicarbonate. Ammonium biphosphate suppresses the response of all acidic compounds.



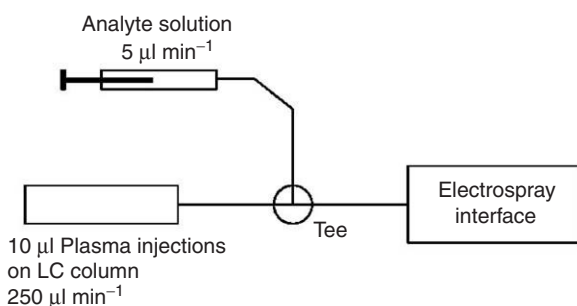
**Figure 3** Effect of additives, formic acid (FA), trifluoroacetic acid (TFA), ammonium hydroxide (NH<sub>4</sub>OH), and acetic acid (AcOH) on the relative response of (A) propranolol and (B) risperidone in electrospray ionization (compared to 50% aqueous methanol, which is 100%). (Data from Mallet CR, Lu Z, and Mazzeo JR (2004) *Rapid Communications in Mass Spectrometry* 18: 49.)

### Detergents

The influence of various detergents on the electrospray ionization of proteins was also investigated in some detail. The detergents are frequently required for the solubilization of the proteins. At concentrations of 0.01% or below, only sodium dodecyl sulfate and taurocholate show significant suppression. At 0.1%, most of the detergents result in significant suppression, e.g., to below 10% for sodium dodecyl sulfate and taurocholate, CTAB, LDAO, and Triton X-100, to the 10–30% range for sodium cholate, and various alkyl glucosides, and to the 30–60% range for CHAPS, hexyl and dodecyl glucosides.

### Sample Interferences

In quantitative analysis, further ion suppression or enhancement may be due to coeluting sample constituents not removed by the sample pretreatment. In



**Figure 4** Schematic diagram of the experimental setup for the study of matrix ion-suppression effects in electrospray ionization mass spectrometry. (Based on Bonfiglio R *et al.* (1999) *Rapid Communications in Mass Spectrometry* 13: 1175.)

most cases, the identity of such components is not clear. It has been demonstrated that nonvolatile residues in the pretreated plasma or urine samples cause ion suppression or enhancement. If the suppression is not too severe, i.e., not compromising the method detection limit too much, it would not be of significant importance. However, it is found that the suppression effects are not reproducible from sample batch to sample batch. Recently, such effects have been studied in detail by analyzing five different plasma batches. Poor relative standard deviations (%RSD) were found for the analyte in electrospray ionization, while the analog internal standard was showing good %RSD. It was found that this effect results from highly erroneous peak area values for the analyte in one of the five plasma batches. Because the internal standard peak area was not affected, and all samples could be analyzed with good %RSD by means of atmospheric-pressure chemical ionization, it was concluded that an unknown interference must be present in one of the plasma batches. Similar effects have been reported by many others. Based on these results, Matuszewski *et al.* proposed that evaluation of ion suppression by matrix effects should not be performed using just one plasma batch, but rather by evaluating the effect in five independent plasma batches.

When in method development and validation a matrix effect is observed which compromises the reliability of the method (poor %RSD), further attention must be paid to improving the sample pretreatment method. The T-in method shown in Figure 4 may be of great help in such investigations: the effect of eluting sample constituents on the response of a continuously infused analyte solution is evaluated.

### Concluding Remarks

Being a liquid-phase ionization method, the performance of electrospray ionization can be significantly

influenced by the solvent composition, especially mobile-phase additives and sample interferences. Nevertheless, electrospray ionization can be successfully applied in the (quantitative) analysis of samples from various application areas. In most cases, the method development for electrospray LC-MS is easier to perform than for LC-UV. However, one should keep in mind that these two detection techniques respond differently to interfering compounds, and therefore require a different optimization strategy. With proper validation, LC-MS is a powerful approach to quantitative analysis. Care must be taken in the selection of mobile-phase additives and in sample pretreatment.

### Applications

The advent of electrospray ionization certainly opened many new application areas for mass spectrometric analysis. This is based on the ability to provide extreme soft liquid-based ionization. Perhaps the most important application area is the analysis of peptides and proteins. The possibility to perform rapid molecular-weight determination of proteins up to 200 kDa stimulated the commercial availability of MS instrumentation featuring atmospheric-pressure ion sources, equipped with electrospray ionization. Other application areas benefited from these developments. LC-MS has become an important analytical tool in many areas of drug development within the pharmaceutical industry, in the study of natural products in plants, in food and environmental analysis. It is about to enter the clinical application area for therapeutic drug monitoring, systematic toxicological analysis, and monitoring of inherited metabolic diseases.

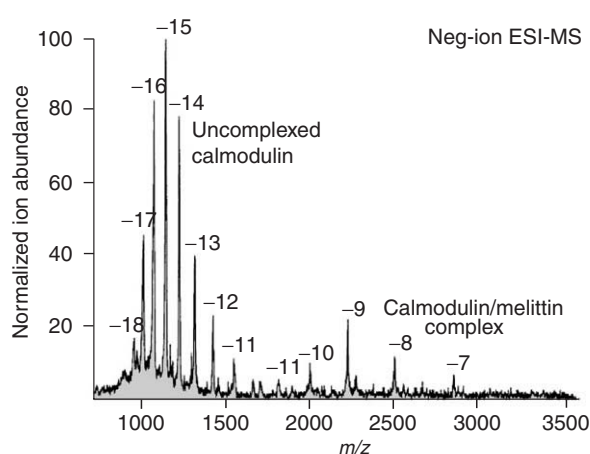
As indicated, electrospray ionization plays an important role in the characterization of proteins. In this area, electrospray ionization is competed by matrix-assisted laser desorption ionization (MALDI), which in some cases appears to be less prone to suppression effects to sample constituents and/or additives. First of all, the formation of multiply charged ions of proteins enables the accurate determination of the protein molecular weight, generally within 0.1% accuracy. The ion envelope of multiply charged proteins ions can actually be a disadvantage, especially in the analysis of proteins with significant heterogeneity due to, for instance, post-translational glycosylation. In current protein research, the determination of the molecular weight of the intact protein appears to play a less important role. In most cases, the protein is enzymatically digested to a peptide mixture. The mass spectrometric characterization of this mixture leads to a peptide map. Due to its ability to produce

singly charged ions, MALDI is preferred in most cases in initial research. Data from the peptide map are applied to a protein database search in an attempt to identify the protein. If further studies are required to identify the protein, e.g., peptide sequencing by means of tandem mass spectrometry (MS/MS), the use of electrospray is often more favorable, because it is coupled more readily to triple-quadrupole or quadrupole–time-of-flight hybrid instruments.

The identification and quantification of proteins in complex biological samples is especially important in the current proteomics research. While for many years MALDI was the method of choice, because of the ease with which it is combined to two-dimensional gel electrophoresis, the role of electrospray ionization in combination with multidimensional LC is increasing.

Already in the early applications of electrospray in the analysis of proteins it was found that electrospray ionization can significantly contribute to the study of protein conformation in solution. While initially solvent-induced conformational changes were investigated, such studies were extended by the use of hydrogen/deuterium exchange and other gas-phase reactions. More recently, ion-mobility spectrometry was implemented between the electrospray ion source and the mass spectrometer as a powerful tool in the assessment of conformational changes of proteins. Such tools are not only applied in fundamental studies, but also applied to detection and monitoring of changes in protein conformation related to, for instance, neurodegenerative diseases like Alzheimer, Parkinson, and Creutzfeldt-Jacobs diseases.

Another important application area of electrospray ionization is the study of noncovalent complexes between proteins and inhibitors, cofactors, metal ions, carbohydrates, other peptides and proteins, substrates, and of nucleic acid complexes. In most cases such complexes exhibit relatively low charge states, resulting in high  $m/z$ . In most cases, the study of such complexes requires the use of either modified quadrupole mass analyzers, to enlarge their  $m/z$  range, or time-of-flight instrument, which show an inherent larger  $m/z$  range. For example, the group of Naylor examined the binding of the calcium-binding protein calmodulin to the calmodulin-dependent protein kinase II (CamK-II) and to melittin. Both peptides form equimolar complexes with calmodulin only in the presence of calcium (stoichiometry was 1:1:4 for calmodulin:peptide:Ca). A typical spectrum of the calmodulin/melittin complex is shown in **Figure 5**. These type of studies are relevant both in the field of structural biology and in the field of drug discovery, enabling the study of drug–protein or inhibitor–protein interactions.



**Figure 5** Negative ion electrospray ionization mass spectrum of calmodulin ( $50 \mu\text{mol l}^{-1}$ ) in the presence of melittin ( $50 \mu\text{mol l}^{-1}$ ) and  $100 \mu\text{mol l}^{-1}$  calcium acetate, pH 5.7. Ions in the range between  $m/z$  900 and 1600 represent uncomplexed calmodulin and ions in the range between  $m/z$  1700 and 3000 represent the noncovalent calmodulin/melittin complex. (Reprinted by permission of Elsevier from Low temperature aqueous electrospray ionization mass spectrometry of noncovalent complexes by Veenstra TD, Tomlinson AJ, Benson L, Kumar R, and Naylor S, *Journal of the American Society of Mass Spectrometry* 9: 580–584, © 1998, by the American Society for Mass Spectrometry.)

See also: **Liquid Chromatography:** Liquid Chromatography–Mass Spectrometry.

## Further Reading

- Fenn JB, Mann M, Men CK, Wong SF, and Whitehouse CM (1990) Electrospray ionization – principles and practice. *Mass Spectrometry Reviews* 9: 37–70.
- Loo JA (2000) Electrospray ionization mass spectrometry: a technology for studying noncovalent macromolecular complexes. *International Journal of Mass Spectrometry* 200: 175–186.
- Mallet CR, Lu Z, and Mazzeo JR (2004) A study of ion suppression effects in electrospray ionization from mobile phase additives and solid-phase extracts. *Rapid Communications in Mass Spectrometry* 18: 49–58.
- Matuszewski BK, Constanzer ML, and Chavez-Eng CM (2003) Strategies for the assessment of matrix effects in quantitative bioanalytical methods based on HPLC–MS–MS. *Analytical Chemistry* 75: 3019–3030.
- Niessen WMA (1999) *Liquid Chromatography–Mass Spectrometry*. New York: Dekker.
- Niessen WMA (2003) Progress in liquid chromatography–mass spectrometry instrumentation and its impact on high-throughput screening. *Journal of Chromatography A* 1000: 413–436.
- Orgorzalek Loo RR, Dales N, and Andrews PC (1994) Surfactant effects on protein structure examined by electrospray ionization mass spectrometry. *Protein Science* 3: 1975–1983.

## Liquid Secondary Ion Mass Spectrometry

N J Haskins, Hants, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Liquid secondary-ion mass spectrometry (liquid SIMS) is an obsolescent ionization technique that revolutionized the analyses of large biological molecules in the 1980s. The term includes fast atom bombardment mass spectrometry (FAB-MS). The technique has been largely superseded by electrospray ionization. This account will cover the history of the technique, the principles of the operation, and the development of continuous flow systems allowing a degree of liquid chromatography–mass spectrometry operation and cover the application of the technique, especially to polar and large molecules such as peptides and small proteins.

### The Development of Liquid SIMS

Liquid SIMS grew out of the development of secondary-ion mass spectrometry. In this technique, ionized atomic or molecular particles are analyzed following their emission from a surface after bombardment with a beam of energetic primary particles. This primary beam may be composed of electrons, ions, fast neutrals, or photons. Secondary ions were observed as early as 1910 by J.J. Thomson after bombarding a metal surface with a beam of ions, but the use of SIMS as an analytical tool came into common use in the 1970s. In this mode of operation, a high-energy beam of ions was directed at a surface to obtain a high yield of secondary ions. The surface itself was eroded quite rapidly. The technique was used especially to study the surface layers of metals and silicon wafers in the growing electronics industry. This technique is known as dynamic SIMS.

To analyze material absorbed on a surface required a less energetic primary beam that would not disrupt the crystalline structure of the substrate. This technique was described as static SIMS. Problems arose inasmuch poorly conducting substrates would become charged, disrupting the primary beam and the emission of secondary ions. Vickerman and co-workers showed in 1979 that the use of a fast atom beam, a stream of energetic neutrals, reduced the charging problem and improved the stability of the secondary ion beam. They called their technique fast atom bombardment SIMS or FAB SIMS. The technique was taken up by Barber and co-workers for the analysis of

intractable organic molecules, and rapidly evolved thereafter into liquid SIMS.

Barber was working at a time when it was required to have a sample in the gas phase before ionization, which is not possible for biologically important materials that are involatile, thermally unstable, and often charged. The most obvious classes are peptides and proteins that are massive, fragile, and often very polar. Some techniques existed for studying these types of samples by introduction in the solid phase and the most used techniques in the 1970s were probably field desorption, static SIMS, and LASER-induced desorption. All these techniques had difficulties with their use and often gave irreproducible results. Barber used static SIMS by coating targets with a thin layer of deposited sample and subjecting this surface to a beam of argon ions with 2–5 keV energy. Both positive and negative ion spectra could be obtained but the sputtered ions leaving the surface had considerable energy spread reducing resolution, and surface charging was a continuing problem. Barber developed a small source producing a beam of fast argon atoms with 2–8 keV energy to strike the sample. Spectra were obtained by depositing a solution of sample on a copper target and introducing this into the fast atom beam. A chance observation showed that spectra could be much improved if the target was covered with an involatile oil (typically glycerol) and the sample dissolved in this. Not only were more stable spectra obtained but also the sample lifetime was considerably extended. Spectra were obtained in both positive and negative ionization modes. The technique was called fast atom bombardment mass spectrometry.

It was quickly realized that it was not essential to use a ‘pure’ atom beam. Acidified glycerol was a sufficiently good conductor where charging of the surface did not occur and several groups, such as M-Scan Ltd. in the UK, developed simpler sources relying on generating beams of fast ions for the primary beam. Initially, gas sources were used for generating beams of xenon ions with energies of 8–10 keV. It was shown that the intensity of the secondary beam depended on the atomic weight of the primary ions, i.e., xenon > argon > neon. Later sources used cesium to generate a primary beam of cesium ions with very high energy (20–30 keV).

The high voltages used to generate the primary beam and hence the secondary beam plus the high concentration of sputtered neutral molecules made magnetic sector instruments the most ideal mass analyzer to use with liquid SIMS sources. Linear

instruments such as quadrupoles are affected by noise from neutrals whilst other analyzers requiring higher vacuum (such as FT-ICR-MS) are badly affected by the relatively high gas pressures.

The use of ion beams precluded the use of the term FAB and the family of techniques became known as liquid SIMS.

## Structure and Operation of the Liquid SIMS System

The liquid SIMS system comprises a target stage mounted on the tip of an introduction probe, a source for the primary ionization beam (commonly known as a 'gun'), and a source comprising beam focusing, extraction, and exit lenses to shape the secondary ion beam and inject it into the mass analyzer (Figure 1).

The target stage originally comprised a thin steel plate mounted on two contact legs in early sources,

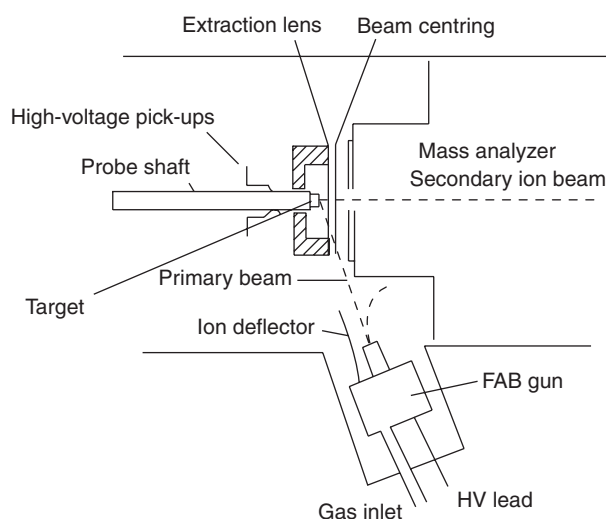


Figure 1 Schematic of a FAB source.

although later more robust solid blocks were used. In operation, the target would be covered with a film of involatile oil known as the 'matrix' in which the sample was dissolved. Matrix materials became an art with special recipes for some selected applications but in the main most samples could be analyzed using a matrix of acidified glycerol. For some biomolecules, thioglycerol or a mixture of thioglycerol and glycerol gave better results whilst *m*-nitrobenzyl alcohol was useful for some apolar compounds that did not dissolve in the hydrophilic matrices. Acidification was usually accomplished using acetic or hydrochloric acid, although occasionally thioacetic was used. In order to obtain good spectra it was essential to ensure that the sample remained in solution within the matrix. Samples were usually added dissolved in a suitable solvent; experience showed that alcohols (methanol, ethanol) and dimethylsulfoxide were useful. Halocarbon solvents were unsuitable as they created flashovers and generally gave weak, noisy spectra.

The primary beams could be generated using three types of guns. Initial work by Barber used a fast atom generator, but later authors used ion guns either using xenon or cesium.

The fast atom generator is a device (Figure 2) in which argon gas is first ionized and accelerated toward the target, then passes through a collision cell in which charge exchange occurs between the fast-moving argon ions and slow argon atoms. This produces a mixed beam of fast and slow argon atoms and ions. After ejection from the gun the remaining ions are deflected from the beam whilst the slow argon atoms diffuse and are lost. The accelerating and exit electrodes are aluminum and erode during use and this type of gun requires cleaning and replacement of these parts every so often. The beam of mainly fast argon atoms hits the surface of the matrix causing sputtering and the expulsion of sample ions (see below).

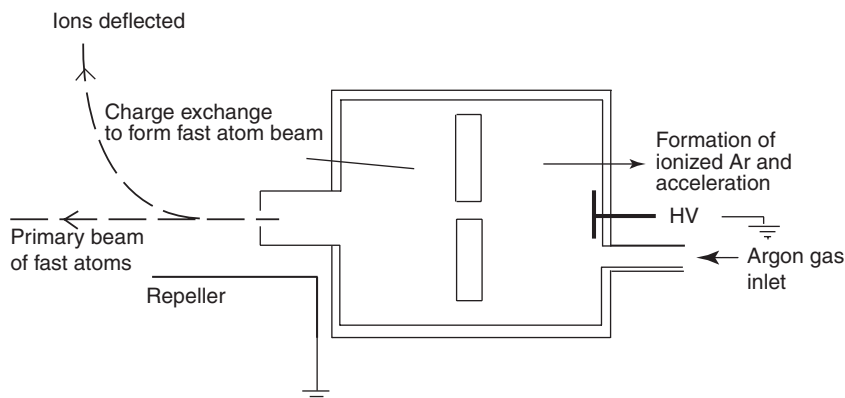


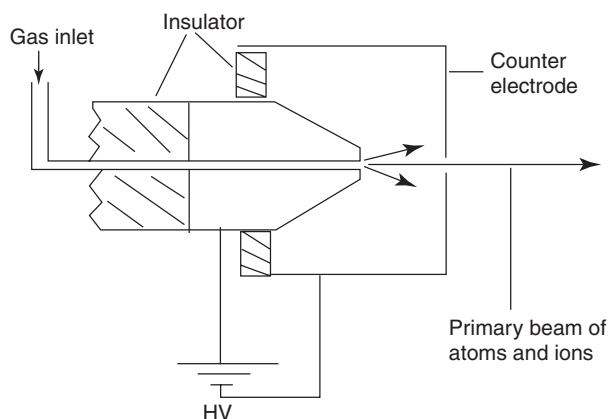
Figure 2 Schematic for a fast atom gun.



The need for a fast atom beam was not necessary as the target was connected into an electrical circuit and the matrix was sufficiently conducting so that surface charging effects caused by using a fast ion beam no longer occurred. Consequently, an alternative primary ion device was developed that used a simple charged nozzle to ionize a jet of gas ions. Gas at a flow rate of  $0.5 \text{ ml min}^{-1}$  passes through a nozzle with a narrow orifice that is held at high voltage to a cap (counter electrode) with an exit hole. Experiments with various gases showed that the intensity of the secondary ion beam depended on the mass of the bombarding gas and therefore xenon was used as the gas of choice. However, the flow of gas into the source chamber was higher than that from the fast atom gun, and care was required to avoid tripping the high vacuum (Figure 3).

The final development of a primary ion source was the cesium ion gun. A pellet of cesium iodide molded around a pin is heated to allow the salt to flow. A high voltage is applied (25–30 kV) and a beam of fast moving cesium ions is formed, which is extracted and focused onto the target through a series of lenses. The great advantage of this type of device is the very low load of gas in the source chamber. The disadvantage is the very high voltages employed, which require special generators and insulators to prevent electrical breakdowns.

All three types of guns generate a beam of high-velocity particles (either atoms or ions). These hit the surface of the matrix with considerable energy and disrupt the surface layers. Any ions contained within the surface layers are sputtered away and, as the target is held at the accelerating voltage for the mass spectrometer, are accelerated through conventional beam shaping optics into the mass analyzer. Holding the target with a positive charge allows the analysis of positive ions, while a negative charge allows the negative ions. With ion guns the accelerating energy



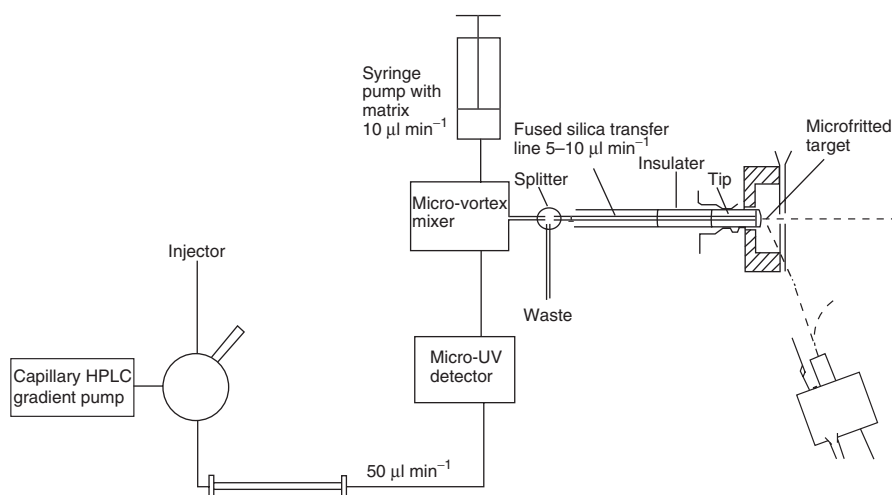
**Figure 3** Schematic for an ion gun.

of the incident beam needs to be greater than the accelerating voltage applied to the target. Thus, in a typical sector instrument with an accelerating voltage of 6 or 8 kV, the incident beam needs to be at least 8 keV. Cesium guns usually operate at much higher voltages so no problems arise with these sources.

The amount of energy released by the incident beam is sufficiently high in the vicinity of the point of impact that sample molecules are ionized by proton transfer from a protonated matrix. Typically, the spectrum obtained shows cluster ions formed from the matrix. In particular, the positive ion spectrum of glycerol (mw 92 u) shows a series of ions  $92n + 1$  where  $n$  can be 1 up to 12 or more. This spectrum can be used to calibrate the mass spectrometer and also provide internal lock masses for accurate mass measurement, described in applications below.

A modified target is used with capillary high-performance liquid chromatography systems, developed by Ito in 1985. In this target, a fused silica line brings the effluent from the liquid chromatogram to a diffuser in the target (Figure 4). This was a difficult technique to operate because flows to the tip had to be kept very low ( $< 1 \mu\text{l min}^{-1}$ ) and the matrix (usually glycerol) had to be added postcolumn. The effluent from a capillary LC system was mixed with a solution of matrix and the combined solution, containing perhaps 10% glycerol by volume, passed through a splitter to keep the flow to the mass spectrometer within limits. The complete system was described as dynamic FAB. The simple system was developed by various manufacturers, especially JEOL and Micromass. The ultimate probe tips had a diffusing mesh across the end of the capillary with an annular wick of tissue outside this to trap excess liquid. The tip was warmed slightly to aid evaporation of liquid from the wick. Results were obtained and published using these systems but it was difficult to retain chromatographic resolution and the quality of the spectra was not great. Data were obtained on tryptic digests of proteins using these techniques, which allowed accurate mass measurements using some polyethylene glycol as internal calibrant. This study showed that dynamic FAB could produce a stable ion beam but better results were obtained by collecting chromatographic fractions offline, concentrating these, and running as normal samples. This aspect of liquid SIMS has been completely superseded by capillary electrospray sources, introduced in the 1990s. Electrospray gave better sensitivity and the generation of multiply charged ions allowed the analysis of much larger molecules than FAB, which usually tends to give protonated molecules.

All liquid SIMS sources work in the same fashion. The primary beam impacts the surface of a liquid



**Figure 4** Schematic for a dynamic FAB probe for LC-MS.

matrix. At the point of impact considerable energy is released, disrupting the lattice of molecules of the matrix making up the meniscus. Indeed, the energy at the point of impact is sufficient to cause molecular fragmentation, so for glycerol successive losses of water are observed (the protonated molecule for glycerol has  $m/z$  93 u; fragments are observed at  $m/z$  75 and 57 u). The impact sputters protonated species away from the surface, both from the matrix and from any analyte molecules contained within the surface layers. If the analyte exists as a protonated species within the matrix then these cations are analyzed directly. Otherwise the energy of the impacting beam allows proton transfer between the protonated matrix and neutral species just above the surface of the matrix. Further away from the impact point clusters are sputtered with insufficient energy to break apart. This is clearly observed for glycerol where a sequence of cluster ions is observed having mass  $92n + 1$ ,  $n$  ranging from 2 up to 15–20. The sequence is sufficiently stable to act as an internal calibrant for accurate mass measurements (see below).

The matrix may contain traces of metal cations, especially alkali metals such as  $\text{Li}^+$ ,  $\text{Na}^+$ , and  $\text{K}^+$ . In these cases, cluster ions are observed for analytes of the type  $\text{M} + \text{Na}^+$ . These are often accompanied by  $\text{M} + \text{K}^+$ , giving a sequence of ions  $\text{M} + 1^+$ ,  $\text{M} + 23^+$ , and  $\text{M} + 39^+$ .

Negative ions can also be observed by reversing the polarities within the ion source and the target. The primary beam remains positive. Deprotonated species are sputtered in a manner analogous to the cationized species. Clusters are sometimes observed such as  $\text{M} + \text{Cl}^-$ .

The elegance of liquid SIMS is that after such disruptive events the surface is restored by the flow of liquid matrix. Further, the conducting properties

mean that any charge buildup in the meniscus is conducted away to the electrically connected target stage. Thus, a sample can be analyzed for as long as liquid remains on the target, often for minutes on end. Spectra can be accumulated by a data system improving the signal-to-noise ratio.

Sample intensity depends on the atomic mass of the bombarding gas. In experiments with both fast atom and ion guns the intensity of the signal was in the approximate ratio of 1:0.28:0.15 for xenon, argon, and neon, respectively (the ratio of atomic masses is 1:0.30:0.15). The approximately linear relationship between atomic mass and intensity is as expected.

Accurate mass measurement could be carried out using suitable calibrants. A popular calibration compound was polyethylene glycol, a mixture of polymers of the form  $\text{H}(\text{OCH}_2\text{CH}_2)_n\text{OH}$ . This commonly used oil has a series of protonated molecules with 44 u spacing, e.g., 151, 195, 239, 283, 372, etc. It also gave a similar negative ion series (149, 193, 237, etc.). Other calibrants were glycerol itself, relying on a series of cluster ions of mass  $92n + 1$  in positive ion mode, and  $92n - 1$  in negative ion mode. High mass calibration was carried out using various alkaline iodide salts, especially cesium iodide with clusters of  $(\text{CsI})_n\text{Cs}^+$  in positive mode and  $(\text{CsI})_n\text{I}^-$  in negative ion mode.

## Some Applications of Liquid SIMS

From the start liquid SIMS was used for the analysis of important biological molecules that, due to their polarity or their mass, were not amenable to existing mass spectral methods. This was exemplified in Barber's original presentation at the Chemical Society Symposium on Soft Ionization Biological Mass

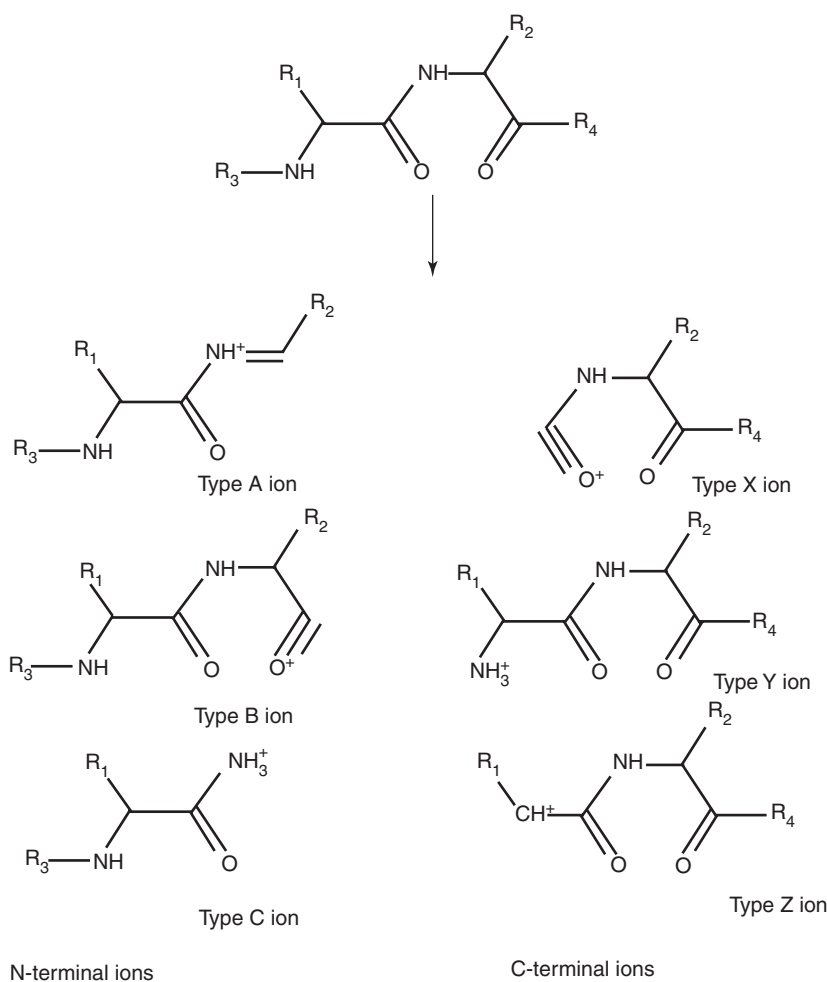
Spectrometry held at Imperial College, London, in July 1980, which was the first announcement of the FAB technique. Barber's examples included a glycoside antibiotic (bluensomycin), a dinucleotide, some penicillins, and some peptides, including gastrin with a molecular weight of almost 2000 Da. These examples set the main application areas for liquid SIMS. The development of high-field magnetic sector instruments with mass ranges up to 10 000 u, and fitted with liquid SIMS sources, revolutionized the analysis of peptides and proteins in the 1980s at a time when the young biotechnology industry required these analyses to be done.

Confirmation of the primary structure of peptides and proteins generated using recombinant DNA was rapidly identified as an important application. Indeed, using a magnetic sector instrument, accurate mass measurements could be carried out. In an early case using peak matching and a resolution of 10 000 the mass of a recombinant octapeptide

(Ac-thr-glu-gly-phe-ser-glu-leu-lys-OH) was determined as 952.4691. The calculated mass for  $C_{42}H_{66}N_9O_{16}$ , the protonated molecule, was 952.4628.

For larger engineered proteins a combination of enzymatic digestion and analysis of the mixture of peptides formed was used to confirm the primary structure of engineered proteins. Simple programs were written to predict the expected masses of the peptides formed and any overlaps due to partial cleavages. The technique was termed 'FAB mapping' in 1981. It was especially useful in determining whether any point mutations had taken place in the sequence of a recombinant protein or if partial cleavage of N- or C-terminal amino acids had taken place, a phenomenon known as 'ragged ends'.

In a typical experiment a protein would be subjected to enzymatic digestion using an enzyme such as trypsin, which cleaves the amino acid chain on the C-terminal side of lysine or arginine. This produces a



**Figure 5** The fragment ions formed by cleavages involving the peptide bond from protonated peptides following ionization using a liquid SIMS source.

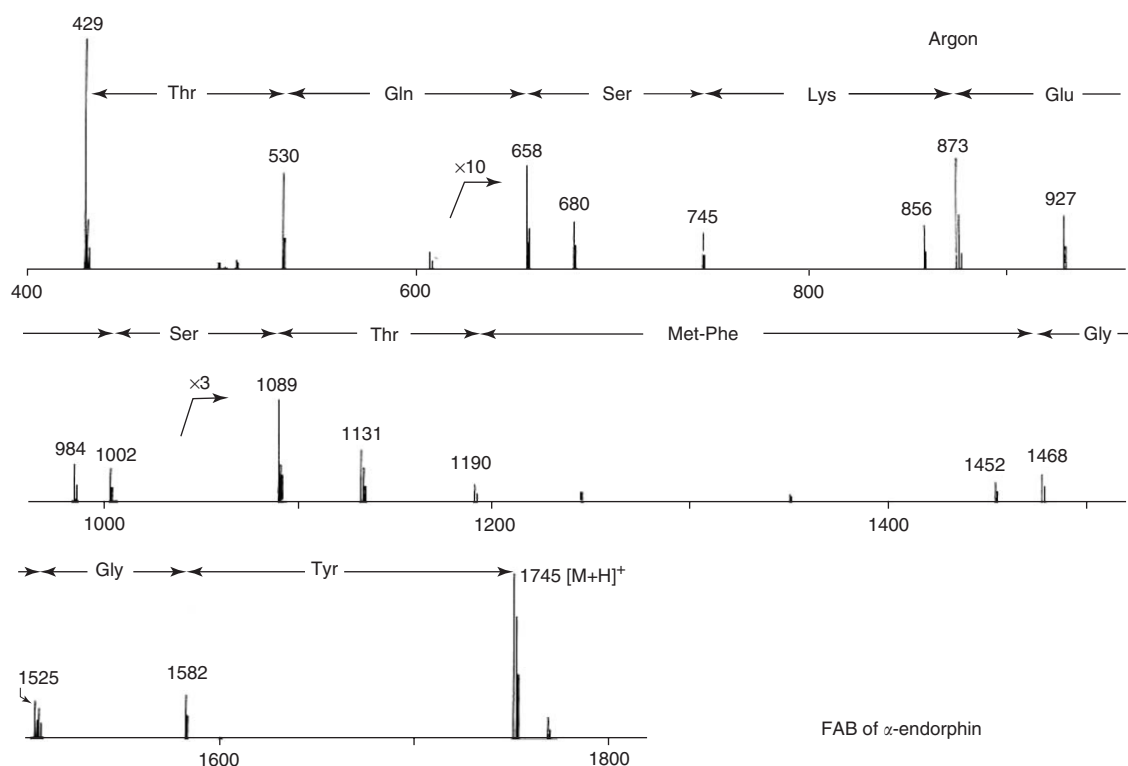
mixture of peptides comprising not only the peptide chains between arginine and lysine residues but also lesser amounts of chains comprising two or more residues due to partial cleavage. The probability of two peptides having the same mass is low, so the 'map' produced is an accurate quality control tool for monitoring the presence of the correct components in the protein sequence. If any discrepancies occur, further maps, using different enzymes giving different mixtures of peptides, can be used to identify a point mutation. Additionally, microchemical manipulation such as methylation or acetylation can be carried out on the probe tip to give additional information. In these cases, it is often useful to use a mixed stable isotopically labeled reagent to give a clear marker signal, for example, acetylation with a 1:1 mixture of  $\text{CH}_3\text{CO}$  and  $\text{C}^2\text{H}_3\text{CO}$  markers. The mass shift and isotopic pattern immediately highlights the number of added acetate groups for each tryptic peptide. This technique was used in the 1980s for monitoring the quality of recombinant insulins and interferons.

Similar techniques were used for determining the structures of unknown peptides and proteins but in these cases the individual peptides were isolated and the primary sequence determined. The mass spectral fragmentation of peptides occurs principally at the peptide linkages and can give rise to six fragment ion types. A nomenclature for this fragmentation (see

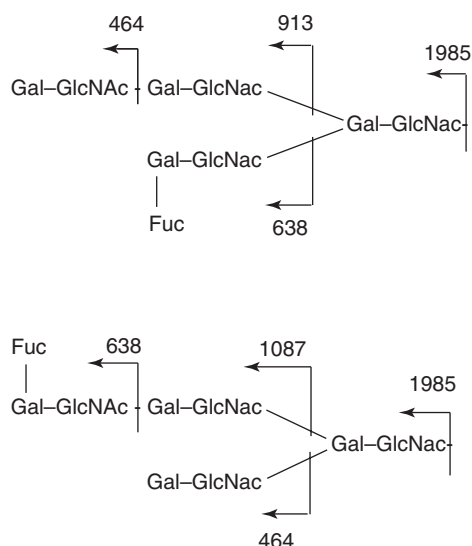
Figure 5) was proposed by Roepstorff and Fohlman in 1985, and has been accepted as a standard since. In the early 1980s, many reports appeared describing the primary structures of peptides that had been derived using a combination of liquid SIMS sources and magnetic sector mass spectrometers. Spectra were produced using MS/MS techniques as well as the spectra from single sector analysis. These spectra were marked by a series of ions spaced with the residual mass of the amino acid residue lost and were easy to read. Only two pairs of amino acids had the same residual mass, namely leucine and isoleucine (isomers with residual mass 113), and glutamine and lysine (different formulae but same nominal mass 128). An example is the spectrum of the C-terminal y-ions from  $\alpha$ -endorphin generated using an ion -gun and a mixed glycerol/thioglycerol matrix (Figure 6).

The revolution in the use of mass spectrometry was extended to other biopolymers, especially sugars. Most enzymes consist of a protein molecule that is modified by the attachment of polysaccharides. The structure of these sugar residues can be vital for the activity of the enzyme and liquid SIMS was used, often in the negative ion mode, to study these. An example showing the fragmentation of two isomeric sugars is shown in Figure 7.

Although the combination of liquid SIMS and magnetic sector mass spectrometry gave rise to the



**Figure 6** The fragmentation of  $\alpha$ -endorphin protonated molecule after FAB ionization (1984; example provided by M-Scan Ltd.).



**Figure 7** The fragmentation patterns induced after FAB ionization that distinguishes two isomeric sugars (1984; example provided by M-Scan Ltd.).

use of mass spectrometry as, arguably, the most important analytical tool for biochemistry, by itself this technique is almost obsolete. The development of atmospheric pressure electrospray (ESI), then micro-electrospray sources, and of analyzers such as orthogonal axis time-of-flight (oa-ToF) has allowed analysis of biopolymers to become a routine in most biology research laboratories. Amenable to the application of sophisticated computing programs to

control the whole system, from separation and introduction devices to complex MS/MS experiments, the ESI-oa-ToF combination is capable of automated analyses on very small quantities of material, far exceeding the sensitivity of liquid SIMS.

*See also: Mass Spectrometry: Overview; Electrospray; Peptides and Proteins. Surface Analysis: Secondary Ion Mass Spectrometry of Polymers.*

## Further Reading

- Barber M, Bordoli RS, and Sedgwick RD (1980) Fast atom bombardment mass spectrometry. In: Morris HR (ed.) *Soft Ionisation Biological Mass Spectrometry*. London: Heyden.
- Clench MR (1992) *A Comparison of Thermospray, Plasmaspray, Electrospray and Dynamic Fab*. Manchester: VG Micromass.
- Gaskell SJ (ed.) (1986) *Mass Spectrometry in Biomedical Research*. Chichester: Wiley.
- Ito Y, Takeuchi T, Ishii D, and Goto M (1985) Direct coupling of micro high performance liquid chromatography with fast atom bombardment mass spectrometry. *Journal of Chromatography* 346: 161–166.
- McNeal CJ (ed.) (1986) *Mass Spectrometry in the Analysis of Large Molecules*. New York: Wiley.
- Morris HR, Panico M, and Haskins NJ (1983) Comparison of ionisation gases in FAB mass spectra. *International Journal of Mass Spectrometry and Ion Physics* 46: 363–366.

## Matrix-Assisted Laser Desorption/Ionization

**D J Harvey**, University of Oxford, Oxford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

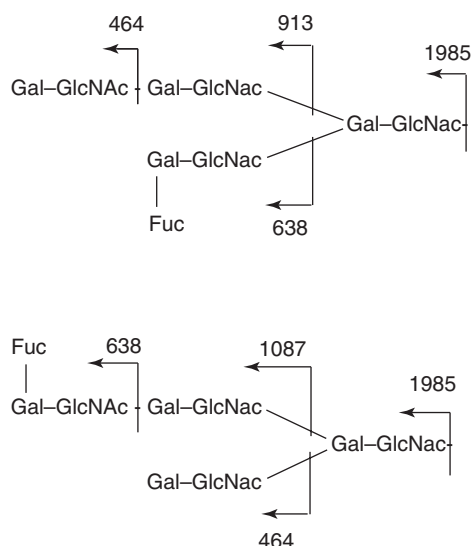
### Introduction

Matrix-assisted laser desorption/ionization (MALDI) is a relatively new ionization technique that is capable of ionizing a large variety of compounds, particularly large proteins, for analysis by mass spectrometry. Its introduction, together with electrospray ionization in the late 1980s, extended the mass range of molecules that could be examined by mass spectrometry into the low megadalton range and provided a new tool that could be used for analysis of both biopolymers and synthetic polymers. The technique is comparatively simple to use; the sample is mixed with an excess of a suitable matrix, usually

a solid, and ionized with pulses from a laser (**Figure 1**). Ion detection is usually, but not exclusively, achieved by use of a time-of-flight (TOF) mass spectrometer. The purpose of the matrix is mainly to dilute the sample and dissipate the laser energy; however, the details of the ionization process are still poorly understood and subject to intensive research. The technique is capable of high-throughput spectral recording and typical target plates, such as those used for proteomics, can accommodate several hundred samples. Since its introduction, MALDI has become one of the major techniques for ionizing organic molecules and has largely replaced older, less sensitive, methods such as fast-atom bombardment. This article briefly reviews the technique and describes applications to several compound types.

Two variations of MALDI mass spectrometry were invented almost simultaneously in the late 1980s. In





**Figure 7** The fragmentation patterns induced after FAB ionization that distinguishes two isomeric sugars (1984; example provided by M-Scan Ltd.).

use of mass spectrometry as, arguably, the most important analytical tool for biochemistry, by itself this technique is almost obsolete. The development of atmospheric pressure electrospray (ESI), then micro-electrospray sources, and of analyzers such as orthogonal axis time-of-flight (oa-ToF) has allowed analysis of biopolymers to become a routine in most biology research laboratories. Amenable to the application of sophisticated computing programs to

control the whole system, from separation and introduction devices to complex MS/MS experiments, the ESI-oa-ToF combination is capable of automated analyses on very small quantities of material, far exceeding the sensitivity of liquid SIMS.

*See also:* **Mass Spectrometry:** Overview; Electrospray; Peptides and Proteins. **Surface Analysis:** Secondary Ion Mass Spectrometry of Polymers.

## Further Reading

- Barber M, Bordoli RS, and Sedgwick RD (1980) Fast atom bombardment mass spectrometry. In: Morris HR (ed.) *Soft Ionisation Biological Mass Spectrometry*. London: Heyden.
- Clench MR (1992) *A Comparison of Thermospray, Plasmaspray, Electrospray and Dynamic Fab*. Manchester: VG Micromass.
- Gaskell SJ (ed.) (1986) *Mass Spectrometry in Biomedical Research*. Chichester: Wiley.
- Ito Y, Takeuchi T, Ishii D, and Goto M (1985) Direct coupling of micro high performance liquid chromatography with fast atom bombardment mass spectrometry. *Journal of Chromatography* 346: 161–166.
- McNeal CJ (ed.) (1986) *Mass Spectrometry in the Analysis of Large Molecules*. New York: Wiley.
- Morris HR, Panico M, and Haskins NJ (1983) Comparison of ionisation gases in FAB mass spectra. *International Journal of Mass Spectrometry and Ion Physics* 46: 363–366.

## Matrix-Assisted Laser Desorption/Ionization

**D J Harvey**, University of Oxford, Oxford, UK

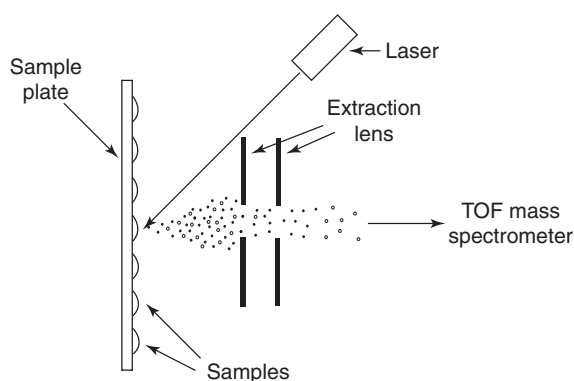
© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Matrix-assisted laser desorption/ionization (MALDI) is a relatively new ionization technique that is capable of ionizing a large variety of compounds, particularly large proteins, for analysis by mass spectrometry. Its introduction, together with electrospray ionization in the late 1980s, extended the mass range of molecules that could be examined by mass spectrometry into the low megadalton range and provided a new tool that could be used for analysis of both biopolymers and synthetic polymers. The technique is comparatively simple to use; the sample is mixed with an excess of a suitable matrix, usually

a solid, and ionized with pulses from a laser (**Figure 1**). Ion detection is usually, but not exclusively, achieved by use of a time-of-flight (TOF) mass spectrometer. The purpose of the matrix is mainly to dilute the sample and dissipate the laser energy; however, the details of the ionization process are still poorly understood and subject to intensive research. The technique is capable of high-throughput spectral recording and typical target plates, such as those used for proteomics, can accommodate several hundred samples. Since its introduction, MALDI has become one of the major techniques for ionizing organic molecules and has largely replaced older, less sensitive, methods such as fast-atom bombardment. This article briefly reviews the technique and describes applications to several compound types.

Two variations of MALDI mass spectrometry were invented almost simultaneously in the late 1980s. In



**Figure 1** MALDI ion source.

the version invented by Koichi Tanaka in Japan, a matrix of glycerol and cobalt powder was mixed with the sample and ionized with a laser. In the other version, invented by Michael Karas and Franz Hillenkamp in Germany, and which has enjoyed much more popularity, the sample was co-crystallized with a ultraviolet absorbing matrix, nicotinic acid, and ionized with a frequency-quadrupled Nd-YAG laser. The main products were singly charged ions such as  $[M + H]^+$  or  $[M + Na]^+$  and the technique was shown to be capable of ionizing proteins with masses in excess of 100 kDa.

## Instrumentation

Commercial instruments now almost invariably use nitrogen lasers emitting at 337 nm although several other wavelengths including those in the infrared (IR) region of the spectrum have been used experimentally. As MALDI naturally produces a pulsed ion beam, it is ideally interfaced to TOF mass spectrometers that are ideally suited to record the massive, largely singly charged ions produced in most analyses. The mass range of the technique is usually limited, not by its ability to ionize molecules but by the drop in signal strength with increasing mass imposed by most common mass spectrometric detectors. In reality, the method is useful up to  $\sim 300$  kDa on commercial instruments although ions can sometimes be seen at much higher mass.

Other types of mass spectrometer have been used to record MALDI spectra; magnetic sector instruments equipped with an array detector initially offered the advantage of higher resolution than was available with the early linear TOF instruments but these spectrometers have now largely been replaced with reflectron-TOF instruments using delayed extraction ion sources. Mass spectrometers operated with higher than normal ion source pressures to

provide collisional cooling of the ion beam have recently allowed hybrid quadrupole-TOF (Q-ToF) instruments to record spectra from MALDI ion sources and to produce high-quality fragmentation spectra by collisionally induced decomposition. The recently introduced TOF/TOF instruments have enabled high-energy fragmentation spectra to be obtained; such spectra provide additional information such as the structure of the amino acid side-chain of peptides and linkage position between the sugar rings of complex carbohydrates.

## Matrices

Most types of molecule can be ionized given a suitable matrix and a considerable amount of work has been performed on the determination of the best matrices for various compounds. Many of the most popular matrices are listed in Table 1. Some of these matrices, such as 2,5-dihydroxybenzoic acid (DHB), are suitable for a broad range of compounds whereas others, such as sinapinic acid, tend to work with only a few compound types, such as proteins in the case of sinapinic acid. As the technique has matured, most analyses for a particular compound type will be conducted with only a limited number of matrices.

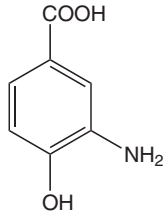
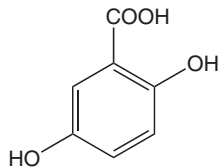
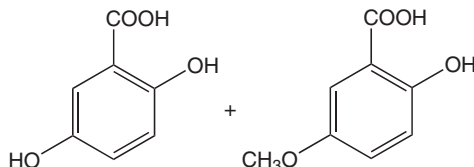
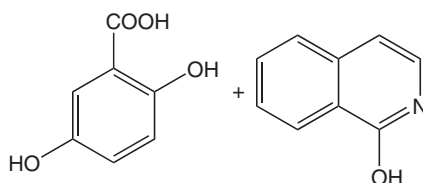
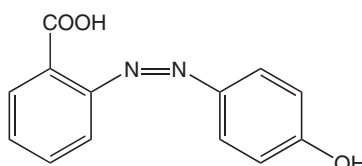
In order to function as a satisfactory matrix, a compound should possess the following properties:

1. The ability to form crystals incorporating the analyte into the crystal lattice or the ability to dissolve the analyte if the matrix is a liquid.
2. The ability to absorb the laser energy.
3. Possess sufficiently poor volatility that sample spots remain on the target in the vacuum system of the mass spectrometer for a reasonable time but, at the same time, be sufficiently volatile that they can be vaporized with the laser.
4. Possess suitable chemistry such that they are able to ionize the sample, usually by proton transfer, but, at the same time, they should not modify the analyte chemically.

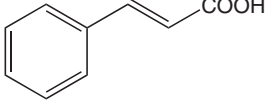
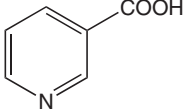
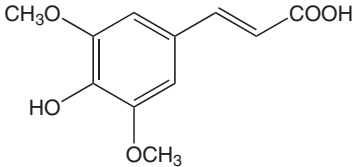
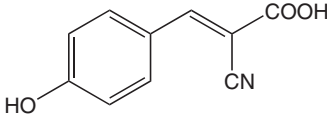
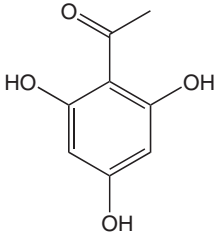
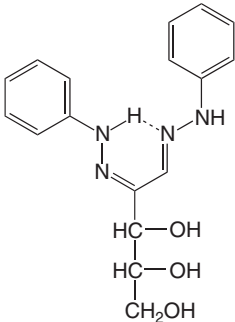
Most of the early matrices were small organic acids capable of protonating the proteins and amino acids that were the main early targets of the technique, but many neutral and even basic compounds are now employed for different compound types.

The most common method for preparing a MALDI target, known as the dried droplet technique, is to mix the sample and matrix, usually in a ratio of about 1:5000 in  $\sim 1$   $\mu$ l of solvent, and allow the mixture to dry on the target. Mixing can occur either before addition to the target or, more commonly, solutions of the sample and matrix are added to the target independently and allowed to mix. Most

**Table 1** Common matrices for MALDI mass spectrometry

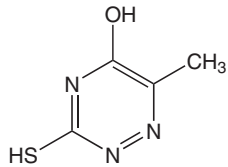
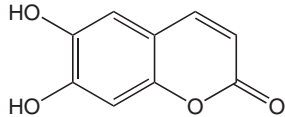
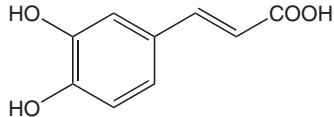
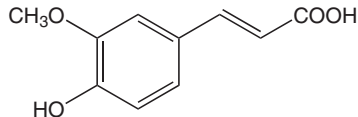
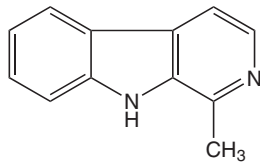
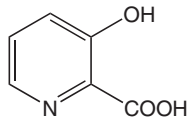
Matrix	Abbreviation	Molecular weight	Solvent	Analyte	Structure
3-Amino-4-hydroxybenzoic acid	–	153.1	MeCN/H <sub>2</sub> O	Sugars	
2,5-Dihydroxybenzoic acid	DHB	154.1	MeCN, acetone	General, particularly sugars	
DHB/2-hydroxy-5-methoxybenzoic acid	Super-DHB	154.1/168.1	MeCN, acetone	Peptides, sugars	
DHB/1-hydroxy- <i>iso</i> -quinoline	DHB/HIQ	154.1/145.2	MeCN/H <sub>2</sub> O	Sugars	
2-(4'-Hydroxy-phenylazo)benzoic acid	HABA	242.2	Acetone	Polymers, sugars	

*Continued*

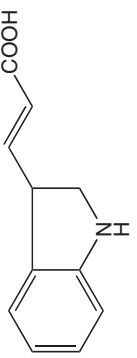
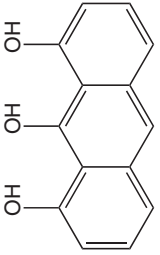
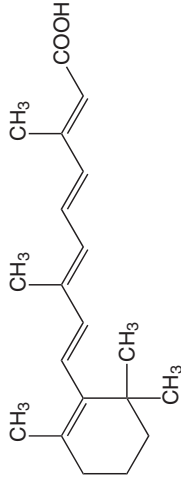
Cinnamic acid	–	148.1	MeCN/H <sub>2</sub> O	General	
Nicotinic acid	–	123.1	H <sub>2</sub> O	Proteins	
Sinapinic acid (3,5-dimethoxy-4-hydroxycinnamic acid)	–	224.2	MeCN/H <sub>2</sub> O	Proteins, polymers	
$\alpha$ -Cyano-4-hydroxycinnamic acid	4-HCCA	189.2	MeCN/H <sub>2</sub> O, acetone	Peptides, lipids	
2,4,6-Trihydroxyacetophenone	THAP	167.2	Ethanol	Sugars. Particularly those containing sialic acid, oligonucleotides	
Arabinosazone	–	328.4	EtOH, MeOH	Sugars	

*Continued*

**Table 1** Continued

<i>Matrix</i>	<i>Abbreviation</i>	<i>Molecular weight</i>	<i>Solvent</i>	<i>Analyte</i>	<i>Structure</i>
6-Aza-2-thiothymine	ATT	143.2	MeCN/H <sub>2</sub> O	Gangliosides	
6,7-Dihydroxycoumarin (Esculetin)	–	178.2	H <sub>2</sub> O	Sugars	
Caffeic acid	–	180.2	MeCN/H <sub>2</sub> O	Glycosaminoglycans	
Ferulic acid	–	194.2	MeCN/H <sub>2</sub> O	Glycoproteins	
Harmane derivatives	–	182.2	MeCN/H <sub>2</sub> O	Cyclodextrins. sugars	
3-Hydroxypicolinic acid	HPA	139.1	Ethanol	Oligonucleotides	



3 $\beta$ -Indoleacrylic acid	—	187.2	Acetone	Synthetic polymers	
1,8,9-Trihydroxyanthracene (dithranol)	—	226.2	Tetrahydrofuran (THF)	Synthetic polymers	
All-trans-retinoic acid	—	300.4	Tetrahydrofuran (THF)	Synthetic polymers	

samples are allowed to dry under ambient conditions but drying is sometimes accelerated by heat, currents of air or nitrogen, or by placing the target in a desiccator.

Although this method works well for the majority of compounds, many variations have been developed to accommodate different analyte properties or simply to improve signal strength or quality. For example, in the layering technique, a film of matrix is deposited on the target in an organic solvent and allowed to dry quickly. The sample solution, with or without additional matrix, is then placed on top and allowed to dry as normal. Sometimes an additional matrix is added at this stage and again allowed to dry. The small matrix crystals from the original layer are thought to catalyze the formation of small crystals in the final target spot, thus producing a more homogeneous target.

Some matrices, such as DHB, produce only a few large crystals, usually originating at the periphery of the target and pointing inwards. Consequently, with instruments that are not fitted with a camera that is able to image the target, it can be difficult to find a suitable spot from which to record a signal. Thus, in addition to the layering technique, several other procedures have been introduced to overcome the problem. One such method, used with carbohydrates, is to recrystallize the sample spot, once dried, with ethanol. This procedure not only reduces the crystal size but is also thought to produce better uptake of the sample by the matrix. Another method is to use a mixed matrix where the added substance, usually in a ratio of about 10:1 sample to additive, forms small seed crystals for the main matrix. Examples are the addition of 2-hydroxy-5-methoxybenzoic acid or 1-hydroxy-*iso*-quinoline (HIQ) to DHB, the former mixture being known as super-DHB. Additives may also be used to form ion pairs with the sample or impurities in the target solution. Ion pairing of sulfated carbohydrates with amino acids (see below) is an example of the first additive whereas the use of citric acid to bind interfering cations in hydroxybenzophenone matrices is an example of the second. L-Fucose has also been used as a matrix adduct to improve spectral quality; it has been proposed that this decomposes to carbon dioxide and water within the laser plume to enhance the number of ion-molecule reactions.

It appears to be important for the sample to enter the crystal lattice of the matrix for satisfactory ion production to occur and proteins tagged with a dye have been used to demonstrate that inclusion does indeed take place. Under ideal conditions high sensitivity can be achieved; some samples can be analyzed from as little as a few hundred attomoles applied

to the target and often, after spectral acquisition, the remaining sample can be recovered. The technique is reasonably sensitive to the presence of contaminants such as salts and buffers but is more tolerant to the presence of these compounds than techniques such as electrospray. A slight drawback to the technique results from the inhomogeneity of the sample surface that causes the signal strength to vary across the target. Consequently, it is sometimes necessary to hunt for so-called sweet spots when acquiring a spectrum. MALDI is a relatively quantitative technique providing that sufficient laser shots are acquired per spectrum to even out the variations in intensity caused by the inhomogeneous target. However, although accuracy can be relatively high, precision is usually only in the order of  $\pm 10\%$ .

Except for some of the larger proteins, MALDI produces almost exclusively singly charged ions. The reason is thought not to be associated with the initial ionization process, but with subsequent reactions in the plume of material desorbed by the laser. It is thought that, as in electrospray ionization, multiply charged ions are initially formed but that these undergo collisions in the plume, predominantly with matrix molecules, allowing neutralization to occur. It has been proposed that such processes are mainly under thermodynamic control, allowing predictions to be made as to the nature of the resulting spectra. Such ion-molecule reactions could explain, for example, why ionization from matrices such as DHB doped with divalent metals such as copper, produce only singly charged ions. Because the second ionization potential of copper (20.29 eV) is above that for DHB (8.05), collisions between ionized copper and DHB result in charge transfer to the matrix. Similar processes are thought to account for phenomena such as matrix suppression and specific analyte ion suppression that are seen under some conditions.

Sample purity is an important factor in obtaining a good MALDI spectrum as contaminants such as salts and buffers tend to inhibit efficient crystallization of sample targets. Among micropurification methods in common use is drop-dialysis in which droplets (about 1  $\mu$ l) of the sample solution are spotted onto a dialysis membrane floating on the surface of water. A variant, used in carbohydrate work, is to use a Nafion-117 membrane that additionally adsorbs hydrophobic compounds. Various ion-exchange resins are in common use; for MALDI work, these are frequently used as microcolumns packed into small pipette tips. Thus, peptides can be efficiently desalted by adsorption onto a C18 solid phase often packed into a pipette tip and sold under the name of ZipTip. Various other such tips including ion-exchange resins have recently become available. Another desalting

technique is to add a few beads of an ion-exchange resin, such as AG-50, to the MALDI target. The beads can either be removed mechanically before introduction of the target into the mass spectrometer, or they can be left on the target with little apparent effect on spectral quality.

## **Variants of the Basic MALDI Technique**

### **Atmospheric Pressure MALDI**

In atmospheric pressure MALDI (ap-MALDI), the sample is ionized outside the vacuum system and ions are captured through a small orifice, usually into an ion-trap mass spectrometer. Its advantage appears to be a considerable amount of rapid collisional cooling of the ions by the high atmospheric pressure, leading to stabilization of sensitive compounds. On the other hand, some compounds, such as carbohydrates, appear to suffer increased fragmentation under these conditions.

### **Surface Enhanced Laser Desorption/Ionization**

Surface enhanced laser desorption/ionization (SELDI) is a variant of MALDI in which the MALDI probe is derivatized with various substances that have affinity for the analyte. The probes are then used to extract the analyte directly from mixtures thus avoiding sample loss through more complicated procedures such as column chromatography. Contaminants can be washed from the probe with appropriate buffers or solvents leaving the purified analyte ready for analysis. Many adsorbents have been used; typical examples are hydrophobic or ionic compounds, enzymes, various receptors, antibodies, and nucleic acids. Although most applications have been reported with proteins, the technique is potentially applicable to any type of compound for which a specific adsorbent can be attached to the probe.

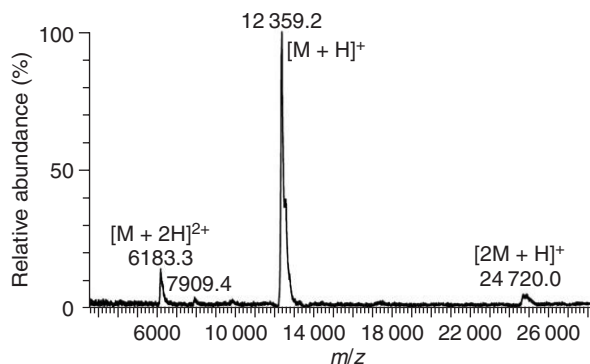
Three types of SELDI are recognized, surface-enhanced affinity capture (SEAC), surface-enhanced neat desorption (SEND), and surface-enhanced photolabile attachment and release (SEPAR). With SEND, the energy-absorbing molecules are bound to the probe surface, often covalently and no additional matrix is necessary. In recent years, porous silica has been found to be an effective SEND surface. Although large molecules can be desorbed, the technique has largely been restricted to small-molecule analysis. Additional binding between the analyte and the energy-absorbing compound is involved with SEPAR and analytes are released photochemically during the ionization process. SEAC is the most widely used SELDI technique; the probe essentially

acts as the purification device via affinity capture and then spectra are recorded, as in classical MALDI, following the addition of an appropriate MALDI matrix. Various affinity probes are now commercially available and the technique offers great potential for examination of trace material in complex biological matrices.

## Applications of MALDI to Specific Compound Types

### Proteins

Proteins produce mainly singly charged ions (Figure 2), usually by protonation and the larger ones are usually ionized from sinapinic acid. However, DHB, ferulic acid, and nicotinic acid can also be effective. A small amount (usually 0.1%) of organic acid, such as trifluoroacetic acid, is often added to aid protonation. In addition to the main singly charged monomeric ions, the resulting spectra usually contain weak doubly charged ions and oligomers, usually dimers and occasionally trimers. For many proteins, these additional oligomeric ion peaks are artifacts of the MALDI process and are more abundant from samples containing high concentrations of protein. However, their observation has prompted research into whether MALDI can be used to study protein-protein interactions. The answer appears to be yes in certain circumstances although, in general, milder techniques such as electrospray, which maintain more realistic physiological conditions, are preferred. Nevertheless, MALDI has been used for measurements of physiologically relevant protein-protein interactions under appropriate conditions. These conditions include proper choice of matrix and pH, the use of chemical cross-linking, or by recording spectra from only one laser shot aimed at any particular region of the target in order to minimize laser-induced dissociation of the complex.



**Figure 2** Typical positive ion MALDI mass spectrum of a protein.

The accuracy of protein mass measurement depends on a number of factors because many proteins are sufficiently massive that resolution is not achieved from additional species such as salts or matrix adducts. Under these conditions, such species cause peak broadening and result in a small shift in measured mass to higher values. Post-translational modifications such as glycosylation can often be resolved for the smaller proteins such as ribonuclease B (Figure 3) with a single glycosylation site, but larger compounds with multiple modifications often produce only broad peaks.

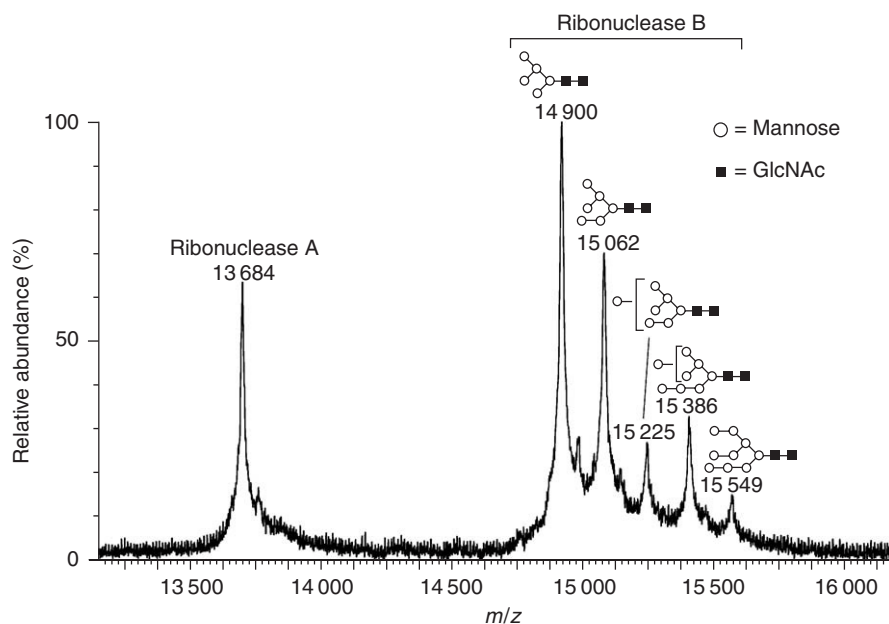
### Peptides

Peptides are usually ionized as  $[M + H]^+$  ions from  $\alpha$ -cyano-4-hydroxycinnamic acid (4-HCCA) or DHB and produce very different molar responses depending on the proton affinity of the constituent amino acids. Arginine-containing peptides, for example, usually produce very strong signals and can be detected at the attomole level. Glycopeptides, on the other hand, frequently give very weak signals in the presence of peptides due to ion suppression phenomena and can ionize as  $[M + H]^+$  and/or  $[M + Na]^+$  species depending on the size of the carbohydrate and the proton affinity of the peptide. A major application of MALDI to peptide analysis is in the area of proteomics where MALDI can provide a very rapid throughput of the resulting tryptic peptides. Unlike electrospray, however, under MALDI conditions, tryptic peptides will produce singly rather than doubly or triply charged ions.

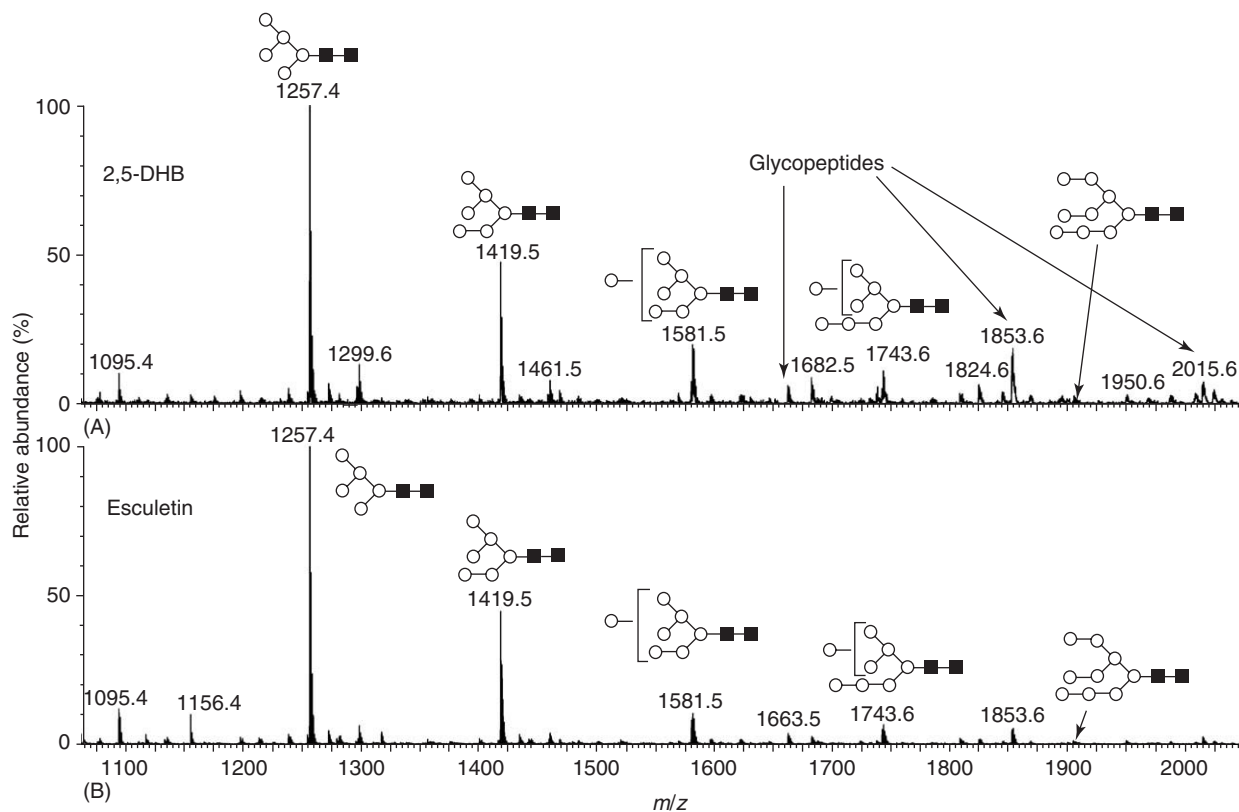
### Carbohydrates

Many matrices have been used for carbohydrate analysis but 2,5-DHB has proved to be the most versatile and widely used. Esculetin is more specific, however, and ionizes sugars in preference to peptides and glycopeptides in mixtures of these compounds (Figure 4). Carbohydrates form  $[M + Na]^+$  ions with most matrices but other adducts can be produced if the matrix is doped with the appropriate salt. Anionic carbohydrates additionally form  $[M - H]^-$  ions. Sensitivity in both modes is lower than with peptides as the latter can be efficiently protonated because of their high proton affinity. However, sensitivity can be increased by permethylation or derivatization at the reducing terminus with a compound that can be protonated, such as a tertiary amine or one that possesses a constitutive charge. Alkyltrimethylammonium compounds have been used in the latter context.

Mixtures of polymeric carbohydrates, such as dextrans, tend to show a drop in sensitivity for the



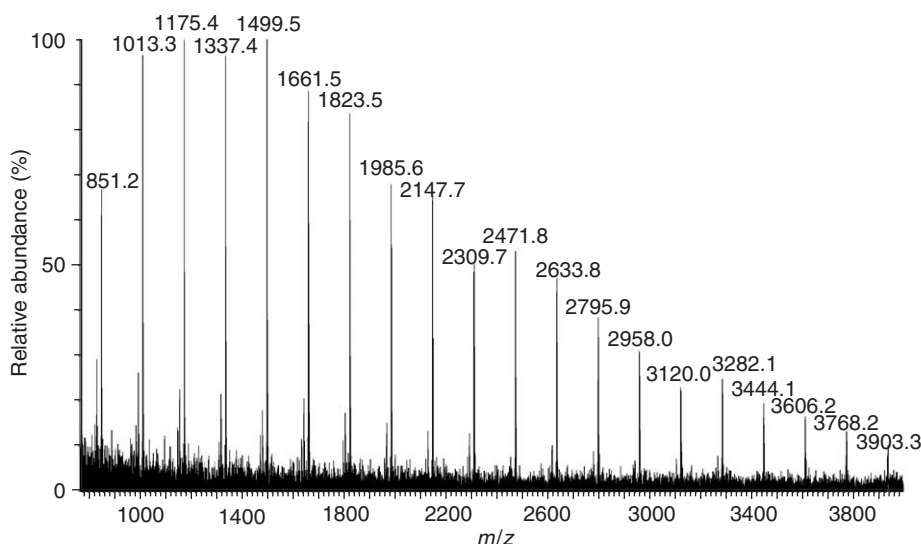
**Figure 3** Positive ion MALDI mass spectrum of ribonuclease B, a glycoprotein with a single *N*-linked glycosylation site.



**Figure 4** Positive ion MALDI mass spectrum of *N*-linked glycans released from ribonuclease B and recorded from (A) 2,5-DHB and (B) esculetin.

larger molecules such as those with masses  $\sim 10$  kDa (Figure 5) but signals can be increased by removal of the shorter polymers by size-exclusion chromatography. For smaller compounds such as

*N*-linked glycans, ionization appears reasonably consistent over the mass range of 2–3 kDa that most of these compounds fall into but smaller O-linked glycans suffer a drop in sensitivity and interference



**Figure 5** Positive ion MALDI mass spectrum of glucose oligomers recorded from 2,5-DHB.

by ions from the matrix. The consistent ionization shown by glycans of similar mass is related to the ionization being produced by sodium adduction rather than protonation, a mechanism that is similar for all compounds. Sialylated glycans are unstable under MALDI conditions and eliminate substantial amounts of sialic acid with the result that broad metastable ions appear in spectra recorded with reflectron-TOF instruments. Methylation or salt formation can stabilize the sialic acids by removing the labile acidic proton that is involved in the sialic acid loss. Sulfated glycans, such as glycosaminoglycans, eliminate sulfate but can be examined by ion pairing with a suitable amino acid such as (RG)<sub>15</sub>.

### Nucleic Acids

Small oligonucleotides with four to six base pairs were first analyzed by MALDI mass spectrometry in 1990 but it was soon clear that oligonucleotide analysis by this technique was beset by two problems: salt formation and fragmentation. Salts were the result of the affinity of the phosphate groups for alkali metals whereas fragmentation appeared to be related to protonation of the base. Salt formation can be reduced by use of ammonium salts, either added to the matrix or exchanged into the sample before target preparation. The ammonium ion transfers a proton to the phosphate to create the free acid by dissociation of the ammonium phosphate ion pairs following desorption. Fragmentation can be reduced by chemical modification. Thus, 7-diaza-analogs of purine groups confer stability whereas adenine and guanine bases can be stabilized by replacement of the 7-nitrogen group with carbon.

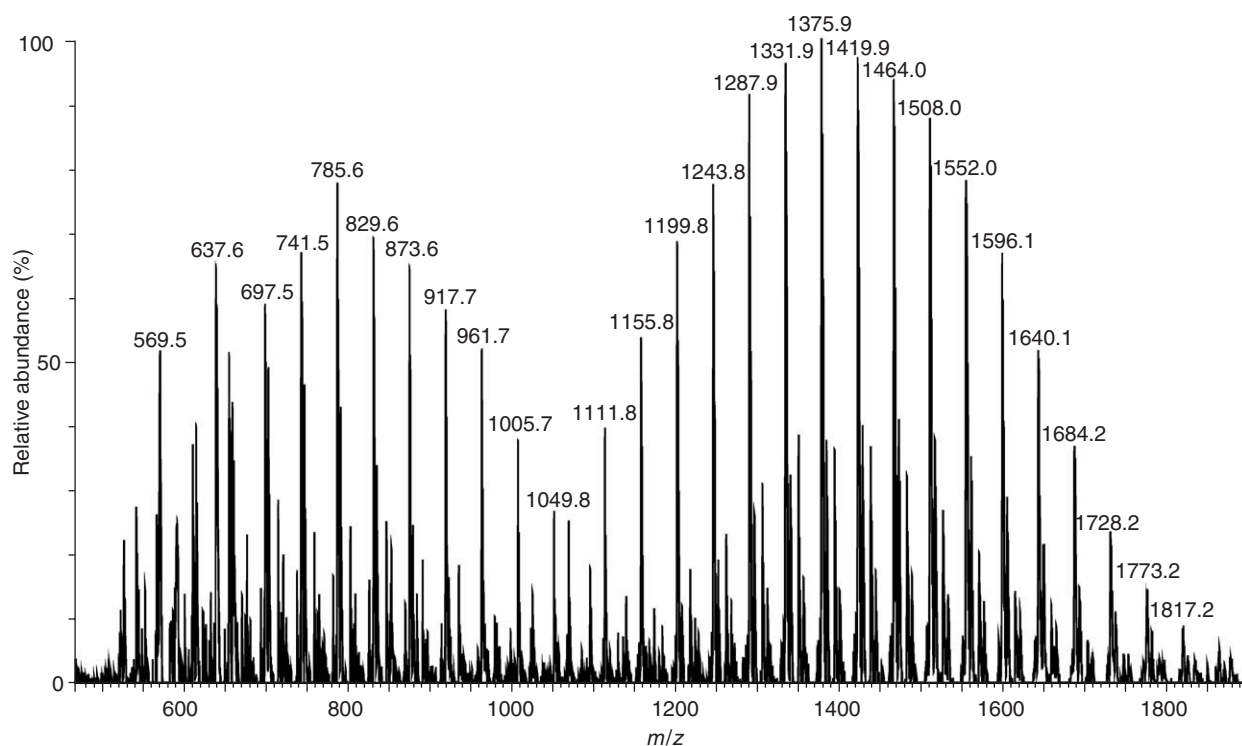
Like other compounds, successful analysis of oligonucleotides depends on selection of a suitable matrix. An early system used frozen aqueous solutions on a copper target with irradiation with visible light at 581 nm. DNA with more than 600 bases could be desorbed intact. However, molecules with over 2000 nucleotides have been detected using IR-MALDI. Possibly the most useful matrix for oligonucleotides is 3-hydroxypicolinic acid (HPA). It produces both positive and negative ions although better results have been reported for the negative ion spectra. The addition of a polyamine comatrix such as spermine has been reported to improve spectral quality.

Double-stranded DNA becomes denatured under MALDI conditions with only ions corresponding to the monomer chains appearing in the spectra. Sensitivity appears to be a function of base composition. Thus, homopolymers containing thymine yield stronger signals than homopolymers containing the other three bases. The reason is thought to be related to fragmentation rather than absorption into the MALDI crystal lattice as thymine is more difficult to protonate than the other bases. Reduced fragmentation has been proposed as the reason why RNA produces stronger signals than DNA; the presence of the 2'-hydroxy group prevents the occurrence of the proposed 1',2'-*trans*-elimination involved in glycosidic cleavage of the ribose moiety from the base.

### Lipids

Analysis of lipids by MALDI mass spectrometry has received much less attention than analysis of proteins and peptides but lipids can, nevertheless, be induced to produce strong signals with suitable matrices.





**Figure 6** Positive ion MALDI mass spectrum of polyethylene glycols recorded from 2,5-DHB.

Phospholipids can be ionized from the popular matrices such as DHB or 4-HCCA but positive ion signal strengths vary considerably with the different types of lipid; phosphatidylcholine, with its quaternary nitrogen atom, produces very strong positive ion signals whereas phospholipids such as phosphatidylinositol, which do not possess a positive charge or amine group, are more difficult to ionize and produce weaker signals. Glycosphingolipids generally behave as derivatized carbohydrates and ionize well from DHB with the production of  $[M + Na]^+$  ions. However, deacylation as with lyso-sphingolipids, converts the amide group into an amine with the result that both  $[M + H]^+$  and  $[M + Na]^+$  ions are formed. MALDI is used extensively in the analysis of bacterial glycolipids such as lipid A, usually in negative ion mode, because of the presence of phosphate or other acidic groups.

### Synthetic Polymers

A considerable amount of work has been reported on the use of MALDI mass spectrometry to examine synthetic polymers. The most popular matrices are DHB, HABA, 1,8,9-trihydroxyanthracene (dithranol), and all-*trans*-retinoic acid, doped with a cationization reagent such as copper or silver. The oligomer distribution, the composition of the end groups, the repeat unit, and the nature of any chemical modifications can all be determined if oligomer

resolution can be attained but one of the major applications is in measurement of the mean molecular weight of polydisperse fractions (Figure 6).

The main problem with analyses of this type is mass discrimination against the larger compounds, caused both by instrumental and matrix factors. Instrumental factors are usually the result of detector saturation as most detectors do not have the dynamic range necessary to accommodate the concentration differences between the various constituents of the polymer mixture. Solvents also have a marked effect on the accuracy of the results. As most synthetic polymers are insoluble in aqueous solvents, the presence of water can have an adverse effect on crystallization. In mixtures containing water, if the organic solvent is more volatile it will evaporate first, causing precipitation of the polymer and inefficient incorporation into the matrix. Nevertheless, for mixtures covering a relatively narrow mass range, MALDI can give an accurate value for the average molecular weight.

**See also:** Laser-Based Techniques. Liquid Chromatography: Size-Exclusion. Mass Spectrometry: Time-of-Flight. Proteomics.

### Further Reading

Ashcroft AE (1997) *Ionization Methods in Organic Mass Spectrometry*. Cambridge: Royal Society of Chemistry.

- Dong X, Gusev A, and Hercules DM (1998) Characterization of polysiloxanes with different functional groups by time-of-flight secondary ion mass spectrometry. *Journal of the American Society for Mass Spectrometry* 9: 292–298.
- Farmer TB and Caprioli RM (1998) Determination of protein–protein interactions by matrix-assisted laser desorption/ionization mass spectrometry. *Journal of Mass Spectrometry* 33: 697–704.
- Festag R, Alexandratos SD, Joy DC, Wunderlich B, Annis B, and Cook KD (1998) Effects of molecular entanglements during electrospray of high molecular weight polymers. *Journal of the American Society for Mass Spectrometry* 9: 299–304.
- Harvey DJ (1999) Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrometry Reviews* 18: 349–451.
- Jackson AT, Yates HT, Scrivens JH, Green MR, and Bate-man RH (1998) Matrix-assisted laser desorption/ionization-collision induced dissociation of poly(styrene). *Journal of the American Society for Mass Spectrometry* 9: 269–274.
- Merchant M and Weinberger SR (2000) Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. *Electrophoresis* 21: 1164–1167.
- Montaudo G and Lattimer RP (eds.) (2002) *Mass Spectrometry of Polymers*. Boca Raton, FL: CRC Press.
- Parees DM, Hanton SD, Cornelio Clark PA, and Willcox DA (1998) Comparison of mass spectrometric techniques for generating molecular weight information on a class of ethoxylated oligomers. *Journal of the American Society for Mass Spectrometry* 9: 282–291.
- Tang N, Tornatore P, and Weinberger SR (2004) Current developments in SELDI affinity technology. *Mass Spectrometry Reviews* 23: 24–44.
- Zhu H, Yalcin T, and Liang L (1998) Analysis of the accuracy of determining average molecular weights of narrow polydispersity polymers by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Journal of the American Society for Mass Spectrometry* 9: 275–281.

## Mass Separation

K J Welham, University of Hull, Hull, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

A mass spectrometer can be broken down into four basic components: sample introduction, ionization, mass separation, and detection/recording. This article will concentrate on mass separation, or ‘analysis’, although the components should not be treated in complete isolation as there are interactions between them.

The modes of mass analysis that will be considered here can be classified as follows:

- Continuous mode:
  - magnetic sectors, possibly with electric sectors
  - quadrupole mass filters
- Pulsed mode:
  - time-of-flight (TOF) analyzers
- Ion trapping devices:
  - quadrupole traps
  - Fourier-transform ion cyclotron resonance traps.

These various methods differ widely in cost, size, and complexity, and also in terms of performance (i.e., mass/charge range and resolving power). The

preferred method of mass analysis depends on the type of problem being investigated.

### Types of Mass Analyzer

#### Magnetic Sectors

Magnetic sector instruments are relatively expensive but they generally give high performance when combined with electric sectors, especially in terms of resolution, mass measurement accuracy, sensitivity, and  $m/z$  range. They employ high accelerating potentials (several kilovolts) in comparison with quadrupole instruments. An ion of mass  $m$  and charge  $ze$  accelerated out of an ion source through potential  $V$  will acquire a velocity  $v$ , and under the influence of a magnetic field of intensity  $B$  will follow a circular path of radius  $r$ . The following equations apply:

$$\text{Ion kinetic energy: } mv^2/2 = zeV \quad [1]$$

$$\begin{aligned} \text{Deflecting force: } BzeV &= \text{centrifugal force} \\ &= mv^2/r \end{aligned} \quad [2]$$

$$\text{Combining these: } m/z = B^2 r^2 e/2V \quad [3]$$

The number of charges,  $z$ , is usually equal to one, so eqn [3] shows that a spectrum of masses can be obtained by changing one of the three variables,  $B$ ,  $r$ ,

- Dong X, Gusev A, and Hercules DM (1998) Characterization of polysiloxanes with different functional groups by time-of-flight secondary ion mass spectrometry. *Journal of the American Society for Mass Spectrometry* 9: 292–298.
- Farmer TB and Caprioli RM (1998) Determination of protein–protein interactions by matrix-assisted laser desorption/ionization mass spectrometry. *Journal of Mass Spectrometry* 33: 697–704.
- Festag R, Alexandratos SD, Joy DC, Wunderlich B, Annis B, and Cook KD (1998) Effects of molecular entanglements during electrospray of high molecular weight polymers. *Journal of the American Society for Mass Spectrometry* 9: 299–304.
- Harvey DJ (1999) Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrometry Reviews* 18: 349–451.
- Jackson AT, Yates HT, Scrivens JH, Green MR, and Bate-man RH (1998) Matrix-assisted laser desorption/ionization-collision induced dissociation of poly(styrene). *Journal of the American Society for Mass Spectrometry* 9: 269–274.
- Merchant M and Weinberger SR (2000) Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. *Electrophoresis* 21: 1164–1167.
- Montaudo G and Lattimer RP (eds.) (2002) *Mass Spectrometry of Polymers*. Boca Raton, FL: CRC Press.
- Parees DM, Hanton SD, Cornelio Clark PA, and Willcox DA (1998) Comparison of mass spectrometric techniques for generating molecular weight information on a class of ethoxylated oligomers. *Journal of the American Society for Mass Spectrometry* 9: 282–291.
- Tang N, Tornatore P, and Weinberger SR (2004) Current developments in SELDI affinity technology. *Mass Spectrometry Reviews* 23: 24–44.
- Zhu H, Yalcin T, and Liang L (1998) Analysis of the accuracy of determining average molecular weights of narrow polydispersity polymers by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Journal of the American Society for Mass Spectrometry* 9: 275–281.

## Mass Separation

**K J Welham**, University of Hull, Hull, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

A mass spectrometer can be broken down into four basic components: sample introduction, ionization, mass separation, and detection/recording. This article will concentrate on mass separation, or ‘analysis’, although the components should not be treated in complete isolation as there are interactions between them.

The modes of mass analysis that will be considered here can be classified as follows:

- Continuous mode:
  - magnetic sectors, possibly with electric sectors
  - quadrupole mass filters
- Pulsed mode:
  - time-of-flight (TOF) analyzers
- Ion trapping devices:
  - quadrupole traps
  - Fourier-transform ion cyclotron resonance traps.

These various methods differ widely in cost, size, and complexity, and also in terms of performance (i.e., mass/charge range and resolving power). The

preferred method of mass analysis depends on the type of problem being investigated.

### Types of Mass Analyzer

#### Magnetic Sectors

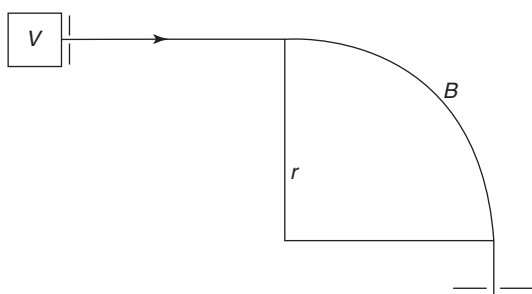
Magnetic sector instruments are relatively expensive but they generally give high performance when combined with electric sectors, especially in terms of resolution, mass measurement accuracy, sensitivity, and  $m/z$  range. They employ high accelerating potentials (several kilovolts) in comparison with quadrupole instruments. An ion of mass  $m$  and charge  $ze$  accelerated out of an ion source through potential  $V$  will acquire a velocity  $v$ , and under the influence of a magnetic field of intensity  $B$  will follow a circular path of radius  $r$ . The following equations apply:

$$\text{Ion kinetic energy: } mv^2/2 = zeV \quad [1]$$

$$\begin{aligned} \text{Deflecting force: } BzeV &= \text{centrifugal force} \\ &= mv^2/r \end{aligned} \quad [2]$$

$$\text{Combining these: } m/z = B^2 r^2 e/2V \quad [3]$$

The number of charges,  $z$ , is usually equal to one, so eqn [3] shows that a spectrum of masses can be obtained by changing one of the three variables,  $B$ ,  $r$ ,



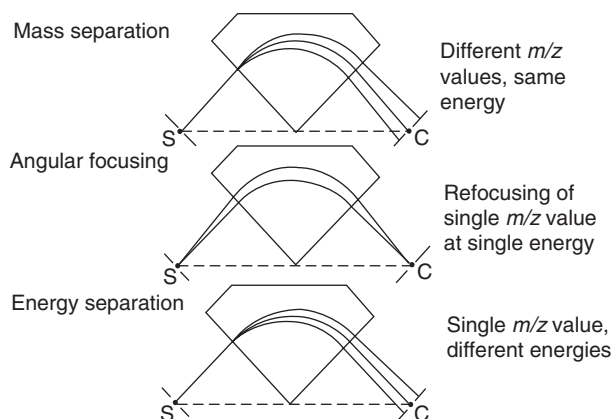
**Figure 1** The variables for separation by a magnetic analyzer.

or  $V$ , as illustrated in **Figure 1**. Commercial instruments have been produced for each of these modes of scanning:

- **Magnetic scanning,  $B$ :** Most commonly used; provides wide mass range without loss of performance at high mass, as in the case of voltage scanning (see below). The disadvantages of hysteresis and slow scanning associated with electromagnets have largely been overcome with modern laminated and/or air core magnet designs.
- **Change of radius,  $r$ :** Not possible with a fixed detector, although special applications such as accurate isotope ratio measurements use two or more fixed detectors. This method was used historically with photoplate detection, and has been revived with microchannel plate photodiode array detectors; this can improve signal-to-noise ratio as all the ions of interest can be monitored simultaneously and hence continuously.
- **Voltage scanning,  $V$ :** No theoretical limit to the mass range as the mass is inversely proportional to the accelerating potential. However, in practice the extraction efficiency of ions from the source depends on  $V$ ; in addition, the ions rely on the kinetic energy imparted by the potential for effective transmission through the instrument, and hence the performance deteriorates at high mass. Advantages are the relatively low cost of permanent magnets (which do not, of course, require electronic circuits), fast scanning, and the reproducibility of the scan function as there is no hysteresis.

Many commercial instruments combine different scan modes, for example, using an electromagnet for normal scanning ( $B$ ), and voltage scanning ( $V$ ) for rapid, reproducible single peak display or multiple peak monitoring. High-performance magnetic scanning instruments are available with array detectors for increased sensitivity of detection over a restricted mass range ( $r$ ).

Dividing eqn [2] by  $v$  shows that a magnetic sector is not a true mass analyzer as the radius of the ion



**Figure 2** The focusing properties of a magnetic sector.

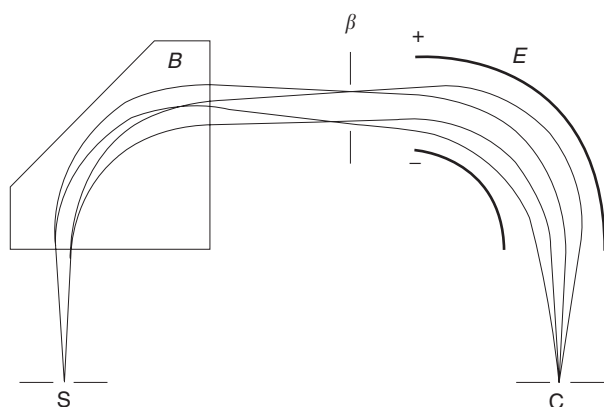
path actually depends on the product of mass and velocity  $mv$  (momentum). Ions of a common mass accelerated through the same potential  $V$  should all have the same momentum, but in practice this will not be entirely true and a single-focusing magnetic sector instrument will give a broadened image at the collector slit, thereby reducing the resolution. Another potential source of loss of resolution is the angular broadening of the beam as it enters the magnetic field. A wedge-shaped magnet has a focusing action, and as long as the source slit and collector slit are on a line passing through the center of the circle where the ion path through the magnet is an arc, a divergent beam will be focused at the collector slit. The field must be homogeneous and the ion beam must enter and leave the magnet at  $90^\circ$ . Angular and energy focusing are illustrated in **Figure 2**.

### Double Focusing Magnetic Sector Instruments

Ions can also be deflected by electric fields, although an electric sector has properties different from a magnetic sector, deflection being dependent upon the kinetic energy of the ions rather than the momentum, i.e., independent of mass. This is described by eqn [4] for an electric field of strength  $E$ :

$$\text{Deflecting force } zeE = \text{centrifugal force} = mv^2/r \quad [4]$$

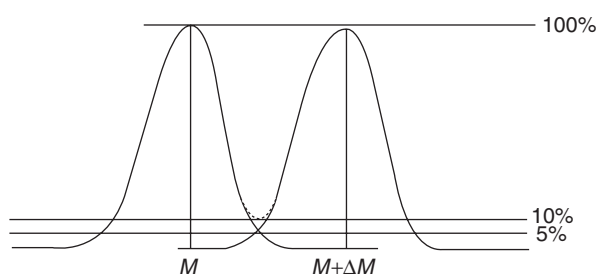
Therefore, the electric field can be used as a focusing element but does not need to be scanned to pass the full range of masses. **Figure 3** shows the double focusing effect of a combination of a magnetic sector and an electric sector for a beam of ions of common mass but being both angularly and energetically divergent. This example has 'reversed' geometry in that the magnet comes before the electric sector, and is the reverse of the original design by Nier and Johnson. This does not affect the focusing properties. Ions with higher kinetic energy are deflected less than



**Figure 3** Double focusing mass spectrometer.

those with lower kinetic energy, so that although the angularly divergent beams are brought to focus at the  $\beta$ -slit (which is equivalent to the collector slit in a single-focusing instrument), the spread of energies has broadened the beam and it cannot pass through a narrow slit without loss of sensitivity. Thus, the  $\beta$ -slit is usually quite wide for maximum transmission. In the electric sector the higher energy ions take a longer path through the electric field and they are deflected more than the lower energy ions, thereby bringing the energetically divergent ions back to a focus at the collector slit.

The geometry of the instrument is accurately calculated so that the angular and energy foci coincide at the collector slit, hence the name double focusing. The design affords a means of compensating for the spread of kinetic energies in the ion beam and allows high mass resolution to be attained by using narrow source and collector slits. The resolving power required to separate a peak due to ions of mass  $M$  from a peak of mass  $M + \Delta M$  is defined as  $M/\Delta M$ . If these peaks differ in mass by one part in 60 000, the resolving power needed will be 60 000, which is the sort of performance that can be attained from a large modern double focusing instrument. Two peaks of equal height are defined as being resolved when the valley between them does not exceed 10% of the peak height (10% valley definition), which for symmetrical peaks corresponds to defining the resolution relative to the peak width at 5% of the height. This is illustrated in **Figure 4**. There are other definitions that include 50% peak height, i.e., full-width at half-maximum. For a sector instrument the resolving power  $M/\Delta M$  is constant throughout the mass range, but the ability to separate peaks that differ by a constant amount deteriorates at high mass, e.g., an instrument of resolving power 1000 can separate peaks differing in mass by less than 1 Da up to mass 1000, but at mass 10 000 it can only separate peaks differing by 10 Da.



**Figure 4** Ten percent valley definition of resolving power.

A high-resolution mass spectrometer allows the accurate determination of ionic masses to within  $\sim 1$  ppm. The mass defects in the atoms that make up organic molecules allow the atomic composition to be determined from the accurate mass, e.g., ions at  $m/z$  28 can be  $\text{CO}^+$  (27.994914),  $\text{N}_2^+$  (28.006148),  $\text{CH}_2\text{N}^+$  (28.018724), or  $\text{C}_2\text{H}_4^+$  (28.031300). This works well up to  $\sim m/z$  500, above which point the decreasing accuracy of mass measurement (as explained above) and the increasing number of possible compositions at each mass make the technique less useful.

The choice of forward or reversed geometry for a double focusing mass spectrometer is not important for the attainment of high resolution. However, there can be other reasons for adopting a particular configuration, especially for studies of metastable ions and tandem mass spectrometry. Ions that are accelerated out of the ion source by field potential  $V$  all have essentially the same kinetic energy and are transmitted by the electric sector at its 'normal' strength. However, ions that fragment in the flight tube lose a fraction of their kinetic energy, which is partitioned between the new ion and neutral fragment, and they can only pass through the electric sector at a reduced field strength. Metastable ions breaking down in the second field-free region between the two sectors will be transmitted by the magnet in a forward geometry instrument, although at reduced magnetic field, to give the familiar broad and rather weak metastable peaks in the normal mass spectra. In a reversed geometry instrument the electric sector will act as a barrier to these ions, which will not have sufficient energy for transmission. Consequently, the normal mass spectra show no metastable peaks.

However, by reducing the electric field the 'normal' ions can be discriminated against and only the ionic products from the metastable fragmentations will be transmitted, giving pure metastable spectra. For both geometries it is possible to scan both sectors simultaneously in a 'linked scan', which provides low-resolution tandem mass spectrometry.



### Time-of-Flight Analyzers

TOF analyzers have recently enjoyed a considerable resurgence with the introduction of desorption ionization techniques such as laser desorption.

In the original design shown in **Figure 5**, the ions are formed in an electron impact source in which the electron beam is pulsed by applying a square wave voltage to a control grid. The acceleration of the ions into the flight tube is similarly pulsed, so that ions are introduced into the mass analyzer in bursts. The accelerating potential of  $\sim 3$  kV gives all the ions the same kinetic energy:

$$zeV = mv^2/2 \quad [5]$$

thus

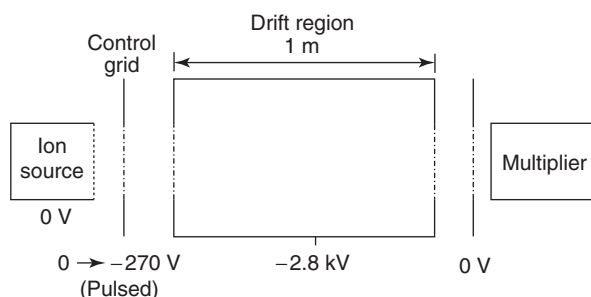
$$m/z \propto 1/V^2 \propto t^2 \quad [6]$$

The ions travel down a 1 m linear drift tube and are separated according to mass, the lighter ions reaching the collector before the heavier ones, and the masses being calculated from the flight times. The output, which is stored in a fast digitizer and sent onto the control computer for display, can respond extremely rapidly to changes in the sample composition as a spectrum can be recorded in milliseconds.

The major problem with early designs was associated with poor resolution, arising from the length of the gating pulses and the spread of kinetic energies due to thermal energies and local differences in the accelerating field. The maximum resolving power was  $\sim 140$ , which in turn restricted the useful mass range although in theory the mass range of a TOF instrument is infinite.

More recently, the unrestricted mass range has been exploited in matrix-assisted laser desorption TOF mass spectrometers for protein analysis at masses up to 1 000 000 Da. Improvements in resolving power have come from several developments:

1. Modern electronics provide much greater accuracy in gating and timing, so that flight times can be recorded with nanosecond precision.



**Figure 5** TOF mass spectrometer.

2. The pulsed nature of ion formation using techniques such as laser desorption, whereby all ions are formed in a very short space of time and can therefore start their flight through the drift tube simultaneously. This has been further enhanced by the use of a gating voltage to delay the ion extraction and has given rise to the current high resolution (up to 20 000) for modern TOF instruments.

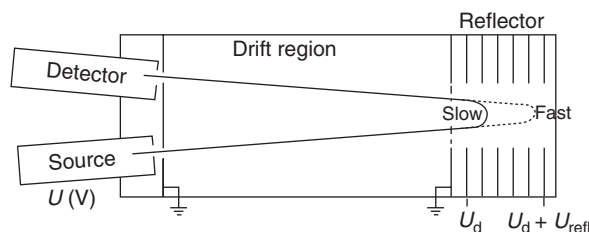
3. The development of reflectron instruments to compensate for differences in kinetic energies. This is illustrated in **Figure 6**, and shows the ions penetrating into an electric field and being repelled or reflected back out again. The higher energy ions penetrate further into the field before being reflected, and therefore follow longer path lengths requiring longer times. This compensates for the shorter time they spend traversing the field-free region, and so eliminates the broadening associated with the spread of kinetic energies, leading to time-focusing.

With these improvements it is possible to obtain resolving powers of 20 000 with TOF instruments, comparable to the performance of sector instruments.

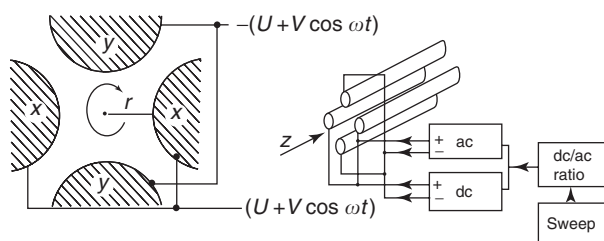
### Quadrupole Analyzers

Quadrupole mass spectrometers are generally less-expensive, lower-performance instruments, not capable of high resolution, but they offer the advantages of flexibility, simplicity, ease of interfacing to liquid chromatography and gas chromatography (as they do not use high potentials in the ion source), ease of computer interfacing and mass calibration (as they have a linear scan law), and speed of scanning and stepping between peaks for multiple ion monitoring. As explained above, the magnetic sector is actually a momentum analyzer and the electric sector is a kinetic energy analyzer, although both can give mass analysis under certain conditions. However, the quadrupole analyzer is a true mass analyzer as it is unaffected by changes in ion velocity, i.e., in momentum or kinetic energy.

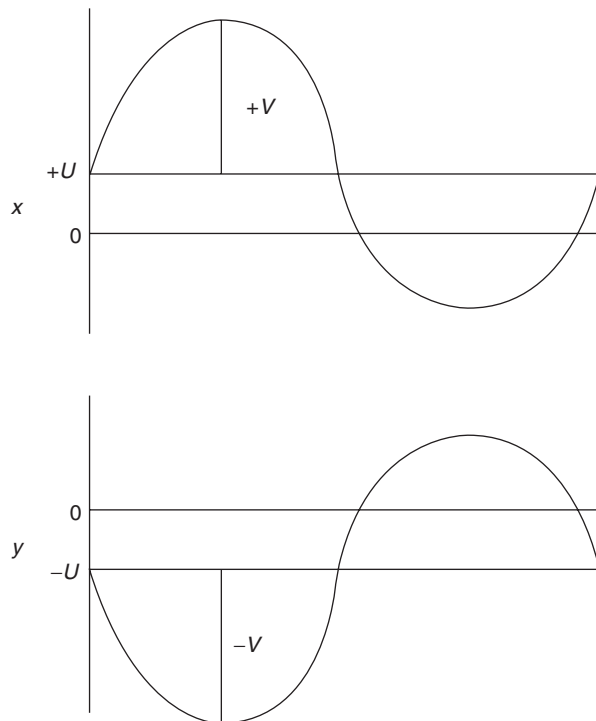
**Figure 7** shows the quadrupole analyzer, which consists of four parallel rods, the surfaces of which should actually be hyperbolic, but circular rods are



**Figure 6** The reflectron TOF mass spectrometer.



**Figure 7** The quadrupole analyzer.

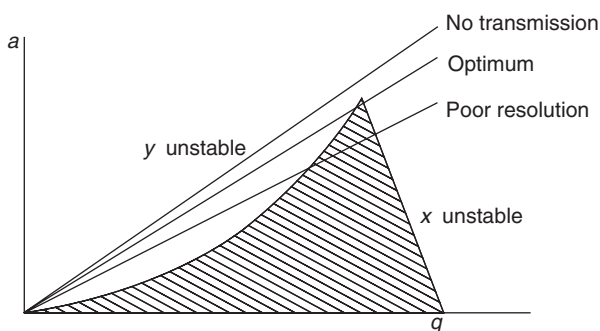


**Figure 8** Ac and dc voltages.

easier to manufacture and give satisfactory performance. The rods are electrically connected in diagonally opposite pairs and mass separation is achieved by applying a dc potential  $+U$  to one pair ( $x$ ) and  $-U$  to the other ( $y$ ), and superimposing a radio-frequency ac potential  $V \cos \omega t$ , which differs in phase by  $180^\circ$  between the pairs of rods. The peak value of the ac voltage is greater than the dc voltage, so the 'positive' pairs are sometimes negative, and vice versa, as shown in **Figure 8**.

Low kinetic energy ions ( $\sim 20$  eV) are introduced and under the influence of the oscillating field follow a spiral path through the analyzer between the rods (along the  $z$ -axis). Their motion on the  $x$  and  $y$  axes is defined by the Mathieu equations, which are of the form shown below for  $x$  (eqn [7]), with an identical equation for  $y$ :

$$\frac{d^2x}{d(\omega t/2)^2} + (a + 2q \cos \omega t)x = 0 \quad [7]$$



**Figure 9** Stability diagram for quadrupole analyzer.

The terms  $a$  and  $q$  are defined as follows:

$$a = 8eU/mr^2\omega^2 \quad [8]$$

and

$$q = 4eV/mr^2\omega^2 \quad [9]$$

where  $r$  is the radius from rod surface to center line; thus,

$$a/q = 2U/V \quad [10]$$

If the ac frequency is kept constant,  $m$  is linearly proportional to  $U$  and  $V$ . In scanning the instrument the ratio  $a/q$  is kept constant by maintaining a fixed ratio between  $U$  and  $V$  (eqn [10]), which are scanned simultaneously to give a spectrum that is linear in mass. The behavior of the ions is defined by the stability diagram shown in **Figure 9**.

Within the shaded triangle the ions follow stable orbits and they are transmitted through the analyzer, whereas outside this triangle the orbits become infinitely large and the ions are lost. The gradient of the scan line determines the proportion of ions transmitted. If  $U = 0$ , i.e., no dc voltage, all ions will pass through and there will be no mass separation. As the gradient of the scan line is increased, i.e., as  $U$  is increased relative to  $V$ , the proportion of ions following stable orbits decreases and the mass resolution increases. If the scan line is too steep it will not intersect with the triangle at all and no ions will be transmitted. The gradient  $a/q$  is therefore selected to cut the shaded triangle near its apex and give the optimum combination of transmission and mass resolution.

The performance of the quadrupole cannot be defined by the  $M/\Delta M$  relationship that applies to sectors, for which the actual peak width increases linearly with mass. With a quadrupole the peak width is constant throughout the mass range, and if the instrument is set up to resolve adjacent mass numbers it is described as having unit resolution. Modern research grade instruments are now available with resolutions up to 5000.

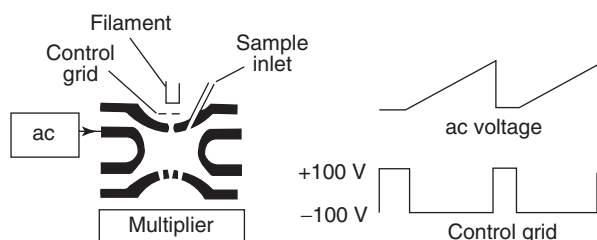
Quadrupoles offer very fast scanning limited only by the impedance involved in changing electric potentials, and are therefore ideally suited to interfacing with high-resolution chromatographic techniques. A typical high-performance quadrupole will have rods of 1 cm diameter and 25 cm in length, possibly preceded by an ac-only quadrupole prefilter. The ac voltage will be in the megahertz frequency range and will have a maximum peak-peak voltage of 10 kV or more.

A quadrupole mass analyzer will not transmit the ionic products of metastable ions breaking down in the analyzer as the change in mass causes the ion orbits to become unstable. However, as explained above, an ac-only quadrupole gives no mass selection and therefore will transmit such ions, and furthermore will do so with high efficiency as the quadrupole is insensitive to ion kinetic energy and is relatively unaffected by angular distribution. Consequently, an instrument known as the triple quadrupole with three quadrupoles in series is routinely used in tandem mass spectrometry. The first and third quadrupoles are mass selective, whereas the second quadrupole is ac-only and is contained within a collision chamber to encourage fragmentation. Quadrupoles can be combined with several other types of analyzers in 'hybrid' instruments.

### Quadrupole Ion Traps

These are essentially a three-dimensional quadrupole in that ions move in the  $x$  and  $y$  directions but there is no  $z$  direction, i.e., the ions oscillate around a fixed point (**Figure 10**). This is a device for storing ions and has been called the QUISTOR (quadrupole ion storage trap). It is constructed from a doughnut-shaped ring electrode with two end-caps, all having cylindrically symmetric hyperbolic surfaces. A radio frequency ac voltage of  $\sim 1$  kV at 1.1 MHz is applied to the ring electrode, which induces the oscillatory motion of the ions. The equations presented for the quadrupole analyzer also apply here, and in the ac-only mode there is no mass discrimination.

Ions can be formed by a variety of techniques either externally in an ion source or directly within the trap.



**Figure 10** The quadrupole ion storage trap (QUISTOR).

Ions flow into the trap controlled by the gating electrode. In the trap, there are stability criteria for the ion trajectories in a manner analogous to the quadrupole, and the application of the ac signal stores ions of all masses down to a low cut-off value of  $m/z$ . Increasing the ac voltage causes the ion trajectories to increase and to approach the end-caps. As this voltage is scanned upwards the ions are ejected from the trap in order of increasing  $m/z$ , and those that emerge through apertures in the lower end-cap are detected by the electron multiplier. Helium is used as a buffer gas in the trap at about  $10^{-3}$  Torr to dampen out the more violent motions and to give improved spectral stability. The storage time can be increased by delaying the start of the voltage ramp.

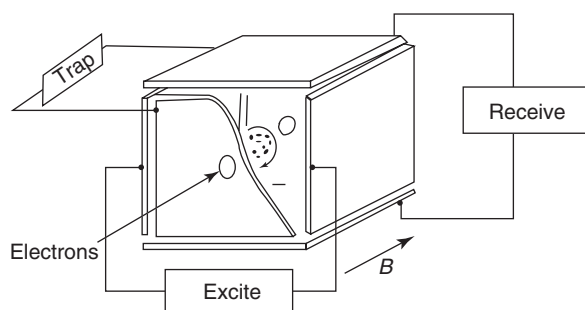
This allows ion/molecule reactions to occur and may be used to effect chemical ionization (CI). The circular electrode has an inner diameter of  $\sim 2$  cm, and the whole device is compact and inexpensive compared with most analyzers. It is ideal as a low-cost mass-selective detector for chromatography. More sophisticated versions are available for which other scan regimes allow different modes of operation for tandem and other studies including MS<sup>n</sup>.

### Fourier-Transform Ion Cyclotron Resonance Mass Spectrometers

Ion cyclotron resonance (ICR) mass spectrometers have become increasingly popular with the use of Fourier transform (FT) techniques, the development of superconducting magnets, and the importance of pulsed desorption ionization methods and are now in the forefront of mass spectrometer development.

The mass analyzer in an Fourier-transform ion cyclotron resonance (FT-ICR) (or Fourier-transform mass spectrometer (FTMS) instrument is illustrated in **Figure 11**. In its simplest form it consists of a cubic cell made up of six isolated plates,  $2.5 \text{ cm}^2$ , contained in a homogenous magnetic field generated by a superconducting magnet. Sample gas molecules may be ionized within the cell by pulsed electron bombardment, or ions may be formed in a second adjacent cell or more remotely in a separate ionization region and then injected into the analyzer. A small potential applied to two opposite plates (trapping plates) parallel to the axis of the magnetic field traps the ions within the field. Any thermal energy they have will cause them to spiral close to the center of the cube with cyclotron frequencies  $\omega_c$  and velocities  $v$  inversely proportional to  $m/z$ :

$$\omega_c = v/r = zeb/m \quad [11]$$



**Figure 11** Fourier-transform mass analyzer.

The cyclotron frequencies range from megahertz down to kilohertz. The application of a burst of low-amplitude radio frequency power to the side plates (excitation plates) will excite the ions into larger orbits. They will now pass close to the top and bottom plates (receiver plates), and so will induce an electrical signal in the circuit connecting these plates. This signal is amplified, digitized, and stored in a computer for subsequent processing. This involves an analysis of the constituent frequencies of the complex waveform that arises from having a complete range of  $m/z$  values, and is carried out by Fourier transformation. From this the mass spectrum is derived; a key feature of this means of detection is that, unlike all other mass spectrometers, it is nondestructive. The instrument can be operated in a number of different modes. The excited ions will gradually lose kinetic energy and will fall back toward the center of the cell. They can be ejected (quenched) by applying a voltage pulse to the trapping plates before new ions are produced, or they can be re-excited to produce a new signal that allows spectrum averaging in order to obtain an enhanced signal-to-noise ratio. Alternatively, by application of only certain frequencies, some ions can be excited to such large orbits that they strike the sides of the cube and are destroyed. This will leave only selected ions for further study, e.g., in tandem mass spectrometry experiments. It is possible to vary the time between ion formation and excitation for CI or to study the kinetics of unimolecular fragmentation or ion/molecule reactions.

FTMS instruments can operate at very high resolving power, and if the residual pressure is low and the density of ions within the cell is small, the measurement of the cyclotron frequency can be very precise. This allows very accurate measurement of  $m/z$  and very precise separation of adjacent peaks. In a 4.7 T magnetic field at  $10^{-8}$  Pa, the time domain signal measured for  $m/z$  18 over 51 s has been recorded to give a resolving power of  $M/\Delta M = 100\,000\,000$ .

In FTMS, the cell can contain both positive and negative ions simultaneously, and  $\text{Cl}^+$  can be resolved from  $\text{Cl}^-$ , a difference of two electrons ( $\sim 0.001$  Da). Resolving power for the FTMS is inversely proportional to mass, and so is lower for high mass ions.

The mass range for FTMS instruments depends on the strength of the magnetic field, and with large superconducting magnets 10 000 Da for singly charged ions is well within reach; in this respect performance is comparable to large sector instruments. The dynamic range is poor compared with sector instruments as there is a space-charge limit to the number of ions that can be stored simultaneously, i.e., ions repel each other and are lost. The dynamic range achievable with a sector instrument can be greater than  $10^6:1$ , whereas FTMS gives only  $10^3:1$ . The pressure in the FTMS cell must remain low as ions will collide with residual gas molecules, leading to a randomization of their motion. This limits the ionization methods available unless the ions are prepared remotely and injected into the cell.

**See also:** Mass Spectrometry: Ion Traps; Time-of-Flight; Multidimensional; Stable Isotope Ratio.

## Further Reading

- Asamoto B (ed.) (1991) *FT-ICR/MS: Analytical Applications of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry*. New York: VCH Publishers.
- Chapman JR (1993) *Practical Organic Mass Spectrometry: A Guide for Chemical and Biochemical Analysis*, 2nd edn. Chichester: Wiley.
- Cole RB (1997) *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation and Applications*. New York: Wiley.
- Cotter RJ (ed.) (1993) *Time-of-Flight Mass Spectrometry*. ACS Symposium Series. American Chemical Society.
- Dawson PH (ed.) (1995) *Quadrupole Mass Spectrometry and its Applications*. Berlin: Springer.
- De Hoffman E, Stroobant V, and Charette JJ (2001) *Mass Spectrometry: Principles and Applications*. New York: Wiley.
- Gelpi E (ed.) (2001) *Advances in Mass Spectrometry*. Chichester: Wiley.
- March RE and Todd JFJ (eds.) (1995) *Practical Aspects of Ion Trap Mass Spectrometry: Ion Trap Instrumentation*. Modern Mass Spectrometry Series. Boca Raton: Interpharm/CRC.
- McNeal CJ (ed.) (1986) *Mass Spectrometry in the Analysis of Large Molecules*. New York: Wiley.
- Vertes A, Gijbels R, and Adams F (eds.) (1993) *Laser Ionization Mass Analysis*. New York: Wiley.

## Ion Traps

**R E March**, Trent University, Peterborough, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Wolfgang Paul and co-workers at the University of Bonn, Germany, obtained a patent in the 1950s for a quadrupole mass filter that is the most common analytical mass spectrometer in use today. A mass filter consists of a square parallel array of four round (rather than the ideal hyperbolic form) rods to which a radiofrequency (rf) potential is applied such that opposite rods are held at the same potential and adjacent rods are out-of-phase by  $180^\circ$  (Figure 1). The mass filter is a two-dimensional quadrupole device because there is a quadrupole ion trapping field in the  $x$ - and  $y$ -directions only. Described in the same patent as the mass filter was its three-dimensional (3D) analog, the 3D quadrupole ion trap, wherein three quadrupole fields act. The quadrupole ion trap is an enormously versatile mass spectrometer that is capable of multiple stages of mass selectivity (tandem mass spectrometry, MS/MS and MS<sup>n</sup>), high sensitivity, high mass resolution, and high mass range. In combination with electrospray ionization, the quadrupole ion trap is applied widely for the study of polar molecules such as peptides and proteins.

In recent years, there has been relatively little further development of the quadrupole ion trap; rather, the range of applications has expanded by making use of the versatility of the instrument. Bruker Daltonik has reported recently on a high charge ion trap wherein the hyperbolic angle of the ion trap has been modified. This change in geometry, combined with resonant ion ejection under the influence of combined dipolar, hexapolar, and octopolar fields, permits rapid mass scanning at high mass resolution

of a greater ion charge in the absence of space charge perturbation.

### Theory

The mathematics of ion trajectory stability within a quadrupole field follows the Mathieu second-order differential equation:

$$\frac{d^2 u}{d\xi^2} + (a_u - 2q_u \cos 2\xi)u = 0, \quad \xi = \frac{\Omega t}{2} \quad [1]$$

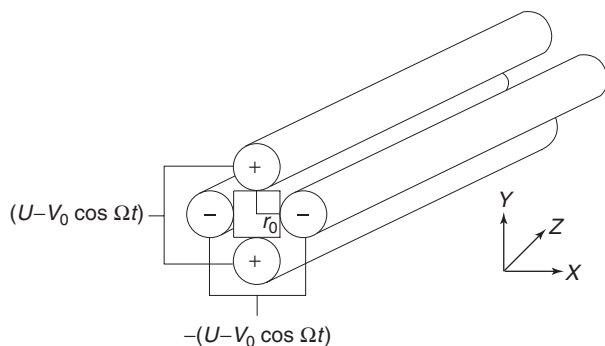
where the subscript  $u$  refers to ion motion in the  $x$ ,  $y$ , and  $z$  coordinates,  $\xi$  is a dimensionless quantity,  $\Omega$  is the radial frequency of the rf potential, and  $t$  is time. A great strength of quadrupolar devices is that the quadrupole fields are independent of one another. Thus, we can apply solutions of the Mathieu equation to 2D or 3D devices provided the weighting factors in each dimension sum to zero. For a 2D device, the trapping parameters are given as

$$a_x = -a_y = \frac{8eU}{mr_0^2\Omega^2}; \quad q_x = -q_y = -\frac{4eV_{0-p}}{mr_0^2\Omega^2} \quad [2]$$

where  $U$  is a dc voltage,  $V_{0-p}$  is the zero-to-peak amplitude of the rf potential,  $m$  is the ion mass,  $e$  is the electronic charge, and  $r_0$  is the radius of the inscribed circle of the rod array.

### New Types of Ion Traps

In the operation of 2D and 3D quadrupole instruments, the modification of ion trajectories by ion/neutral collisions must be considered. In rf quadrupole fields, an ion/neutral collision reduces both ion kinetic energy and ion excursions such that the ions are cooled and focused to the center of each field. Collisional cooling is an important aspect of the behavior of ions in a quadrupole field; for example, in a 2D quadrupole device, collisional cooling is employed to limit the excursions of ions so as to form a tightly focused ion beam of diminished kinetic energy constrained close to the central axis. A focused beam of ions may be transmitted through a relatively small orifice from one section of an instrument to the next such that pumping requirements are reduced, and a focused ion beam can be accelerated with reduced ion loss. When the axial motion of a focused ion beam is arrested within a rod array such that the confined ions can be excited resonantly, a linear ion quadrupole trap is obtained.



**Figure 1** An ensemble of circular rods that forms a quadrupole mass filter.



There has been substantial development of miniature cylindrical ion traps in response to the demands for a small mass spectrometer for use in the field. The remainder of this article is given over to discussion of novel linear ion traps and miniature cylindrical ion traps that employ quadrupole ion trapping fields.

### Linear Quadrupole Ion Trap Mass Spectrometer

In 2002, two linear ion traps were reported; they had the basic structure of a quadrupole mass filter. The first linear ion trap instrument was described by J.W. Hager of MDS SCIEX and a second linear ion trap instrument was described by J.C. Schwartz, M.W. Senko, and J.E.P. Syka of Thermo Finnigan. Both instruments employ mass-selective ion ejection. Axial ion ejection is employed in the SCIEX instrument while in the Thermo Finnigan instrument ion ejection occurs radially. In addition, ion trapping in the SCIEX instrument can occur either in a pressurized collision cell region or in a low-pressure quadrupole rod array downstream of the collision cell.

The principal advantage of a linear ion trap is that more ions can be confined in the larger device than in a 3D ion trap. Thus, the onset of space-charge repulsion is experienced only at greater ion loading of the trapping device. Virtually all of the operating characteristics of the 3D ion trap are retained in the linear device. Both linear ion trap instruments were developed using existing experimental platforms that permitted prior collisional cooling of the ion beam.

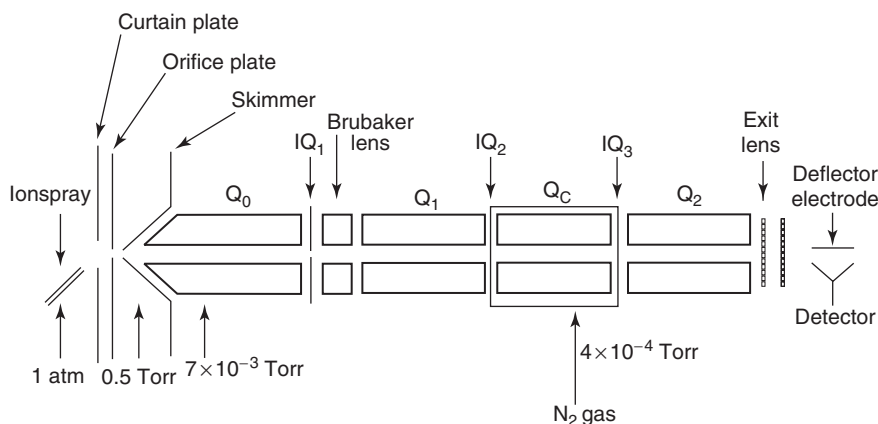
**SCIEX linear ion trap** The SCIEX instrument shown schematically in Figure 2 is based on a triple-stage quadrupole (TSQ) mass spectrometer wherein, normally, the ion beam travels from left to right. Ions

are focused collisionally close to the axis in  $Q_0$ , mass selected in  $Q_1$ , and collisionally dissociated to fragment ions in the collision cell  $Q_C$ .  $Q_C$  is a quadrupolar array of round rods of length 127 mm enclosed in a pressurized container to which a collision gas is introduced.  $Q_2$  is a quadrupole mass filter that has rods of length 127 mm also. The round rod arrays were fabricated with a field radius,  $r_o$ , of 4.17 mm.

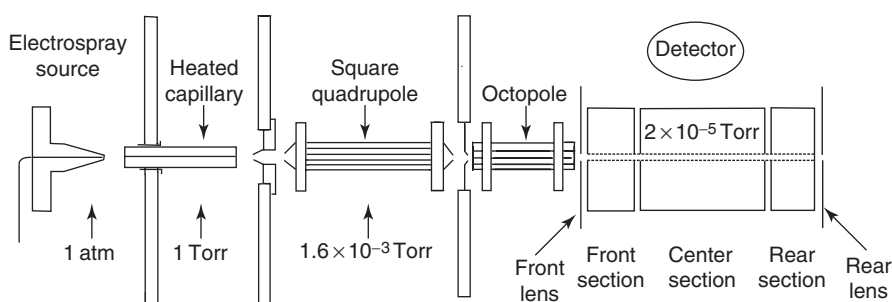
When  $Q_C$  or  $Q_2$  is operated as a linear ion trap, an rf drive voltage of 1 MHz is applied. An auxiliary ac potential is applied in quadrupolar fashion to the rod arrays of either  $Q_C$  or  $Q_2$  so as to excite ions at secular frequencies. Ions are trapped radially in the linear ion trap  $Q_C$  by the rf potential, and are confined axially within the rod array by dc potentials applied to the aperture plates,  $IQ_2$  and  $IQ_3$ . Thus, a linear ion trap resembles a bathtub having a parabolic cross-section and near vertical ends. Ions enter the linear ion trap close to the zero-field centerline of the device and encounter a series of momentum-dissipating collisions with collision gas.

When an auxiliary ac field is applied to induce radial resonant excitation, coupling of radial and axial motions effected axial ion ejection when the ion radial secular frequency matched that of the ac field.

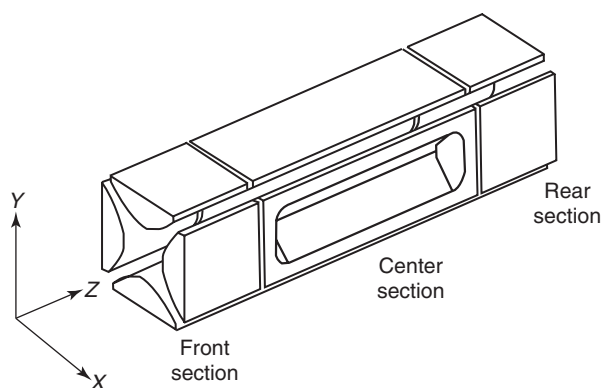
**Thermo Finnigan linear ion trap** In the Thermo Finnigan instrument shown in Figure 3, an ion beam from an electrospray source is directed through a heated capillary, two successive rod arrays, a front lens, and into a linear ion trap composed of three sections of quadrupole hyperbolic rod arrays with  $r_o = 4$  mm. Radial confinement of ions is effected by the trapping potential well in the center section and axial confinement by dc potentials applied to the front and rear lenses. The basic design of the linear



**Figure 2** A schematic diagram of the SCIEX linear ion trap mass spectrometer. Ion trapping can be wrought in either of  $Q_C$  or  $Q_3$ . The apparatus is based on the ion path in a TSQ mass spectrometer and the ion beam travels from left to right. Typical operating pressures are given. (Reproduced with permission from Hager JW (2002) A new linear ion trap mass spectrometer. *Rapid Communications in Mass Spectrometry* 16: 512–526. © 2002, Wiley & Sons Ltd.)



**Figure 3** A schematic diagram of the Thermo Finnigan linear ion trap mass spectrometer. The ion beam travels from left to right. Typical operating pressures are given. (Reproduced with permission from Elsevier from Schwartz JC, Senko MW, and Syka JEP (2002) A two-dimensional quadrupole ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry* 13: 659–669. Copyright 2002, by the American Society for Mass Spectrometry.)



**Figure 4** An angled view of the three sections of the two-dimensional linear ion trap. The detector faces the center section. (Reproduced with permission from Elsevier from Schwartz JC, Senko MW, and Syka JEP (2002) A two-dimensional quadrupole ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry* 13: 659–669. Copyright 2002, by the American Society for Mass Spectrometry.)

ion trap is shown in **Figure 4** where the front and rear sections are of length 12 mm and the center section is of length 37 mm. An rf potential at 1 MHz is applied to the linear ion trap, and an auxiliary ac potential is applied in dipolar mode across the rods in the  $x$ -direction of the center section. Ions are directed radially from the center section; in order to detect the ions externally, a 0.25 mm high slot was cut along the middle 30 mm of the center section of one rod as shown in **Figure 4**; a standard detector system is mounted in front of the slot.

**Modes of operation** In **Figure 2**,  $Q_0$  is an rf-only quadrupole in which the ion beam is focused collisionally to enhance the transmission of ions from  $Q_0$  to  $Q_1$ , an rf/dc quadrupole mass filter. Precursor ions are mass selected in  $Q_1$  and transmitted to  $Q_C$  where they are confined and subjected to resonant dissociation; the product ions are ejected axially and mass-selectively. In an MS/MS experiment, where

collision-induced dissociation (CID) is brought about by resonant excitation in  $Q_C$ ,  $Q_2$  serves merely to transmit the product ions at low pressure to the detector.

In an alternative arrangement, ions fall through a potential difference as they pass from  $Q_1$  to  $Q_C$  so that CID occurs in  $Q_C$ . The nascent product ions are accumulated in the linear ion trap  $Q_C$  and then ejected mass selectively to the detector.

In yet a further arrangement made possible by the serendipitous selection of the original experimental platform, ions fall through a potential difference as they pass from  $Q_1$  to  $Q_C$  so that, again, CID occurs in  $Q_C$ . The nascent product ions pass through  $Q_C$  and enter into and are accumulated within the linear ion trap  $Q_2$  that is operated at rf only. Ions confined in  $Q_2$  are excited radially and, in the fringing field at the exit of  $Q_2$ , are ejected axially and mass-selectively. In this mode, since  $Q_2$  is not pressurized directly, it is identified as a low-pressure linear ion trap.

**Ion trap capacities** A simple model for the comparison of ion trap capacities based on the volumes of the trapped ion clouds subjected to collisional focusing in each instrument yields

$$\begin{aligned} N_{2D,A} : N_{2D,B} : N_{3D} &= (\pi r_A^2 l_A) : (\pi r_B^2 l_B) : \frac{4\pi r_C^3}{3} \\ &= l_A : l_B : \frac{4r_C}{3} \end{aligned} \quad [3]$$

where  $N_{2D,A}$ ,  $N_{2D,B}$ , and  $N_{3D}$  are the ion capacities in the SCIEX, Thermo Finnigan, and quadrupole ion trap instruments, respectively;  $(\pi r_A^2 l_A)$ ,  $(\pi r_B^2 l_B)$ , and  $4\pi r_C^3/3$  are the occupied volumes of the same instruments, respectively. For collisionally focused ion clouds at a common rf potential and radial frequency,  $r_A = r_B = r_C = 1.0$  mm based on ion tomography experiments, the ratio  $N_{2D,A} : N_{2D,B} : N_{3D} = 95:22:1$ . This ratio suggests an ion capacity for the linear ion trap two orders of magnitude greater than that of the 3D ion trap.

### Characteristics of linear ion trap operation

**Trapping efficiency** The efficiency of trapping in  $Q_C$  approaches 100% for ions with 5–15 eV of kinetic energy over the pressure range  $1\text{--}17 \times 10^{-4}$  Torr. The high acceptance of ions into a linear trap is attributed to the absence of a quadrupole field along the  $z$ -axis.

**Mass discrimination** Ions injected along the  $z$ -axis into a linear ion trap as shown in **Figures 2** and **4** encounter minimal rf field in the  $z$ -direction. Thus, mass discrimination during ion injection and ion trajectory stabilization processes is much reduced in a linear ion trap; a wide mass range from 150 to 2000 Th can be trapped simultaneously with high efficiency.

**Ion isolation** Ion isolation has been demonstrated to be >95% efficient.

**Ion activation** Activation of ions isolated in a linear ion trap is carried out at  $q_x = 0.25$ , which is a compromise for efficient fragmentation (74%) and a relatively large fractional (72%) mass range for confinement of product ions.

**Tandem mass spectrometry**  $MS^n$  can be achieved by repeated cycles of isolation and activation.

**Spectral space charge limit** The onset of space charge perturbation corresponds to an ion density of  $\sim 460 \text{ ions mm}^{-3}$  that compares well with the value of  $\sim 400 \text{ ions mm}^{-3}$  for the 3D ion trap.

**Ion ejection** The efficiency of mass-selective radial ion ejection was 44% relative to the number of ions detected under mass-independent ion ejection conditions, yielding an overall efficiency of 12.7%. Mass-selective axial ejection yielded an overall extraction efficiency of 8% at  $5.0 \times 10^{-4}$  Torr to 18% at  $1.2 \times 10^{-3}$  Torr. The ejected ion signal intensity increased threefold as the auxiliary AC frequency was increased from 334 to 762 kHz.

**Enhanced mass resolution** Mass spectra with enhanced mass resolution are obtained with linear ion traps by reduction of the mass scanning rate and axial modulation amplitude. At a scanning rate of  $5 \text{ Th s}^{-1}$  and with an auxiliary frequency of 450 kHz applied to  $Q_C$  of **Figure 2**, a mass resolution ( $m/\Delta m$ ) of  $\sim 6000$  was observed for the protonated reserpine molecule.

**Sensitivity** A fivefold improvement in the detection limits of a LCQ Deca 3D ion trap has been demonstrated with the Thermo Finnigan linear ion trap. The observed extraction efficiency of 18% from  $Q_C$

indicates that the linear ion trap to be some 10 times more sensitive than a 3D device. The linear ion trap has the capability to become a highly sensitive scanning mass spectrometer.

**Low-pressure linear ion trap,  $Q_2$**  The operating pressure of the  $Q_2$  linear ion trap is <1% of that of the  $Q_C$  linear ion trap, yet the trapping efficiency for the low-pressure  $Q_2$  linear ion trap is  $\sim 45\%$  for ions with kinetic energies in the range 5–15 eV.

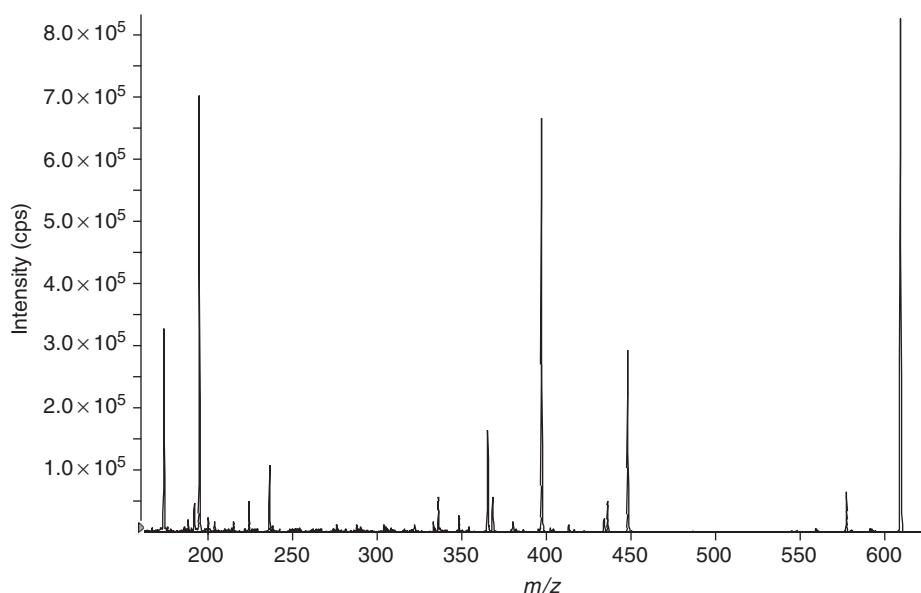
**Comparison of performance** The product ion mass spectrum of protonated reserpine can be used to affect a comparison of the performances of a TSQ, a  $Q_C$  linear ion trap, and a low-pressure  $Q_2$  linear ion trap. The TSQ mass spectrum shows fragment ions to  $m/z$  174; the base peak of  $m/z$  195 was observed at a signal intensity of  $1375 \text{ counts s}^{-1}$ . The  $Q_C$  linear ion trap mass spectrum shows the middle range ( $m/z$  350–450) fragment ions having higher relative signal intensities; the base peak of  $m/z$  195 was observed at a signal intensity of  $15\,700 \text{ counts s}^{-1}$ ; that is, some 11 times more intense. The low-pressure  $Q_2$  linear ion trap mass spectrum is shown in **Figure 5**; the relative fragment ion intensities are similar to those obtained with the  $Q_C$  linear ion trap and the base peak,  $m/z$  195, was observed at a signal intensity of  $7000 \text{ counts s}^{-1}$ ; that is, some five times more intense than the TSQ. The  $Q_2$  linear ion trap showed the highest mass resolution and signal-to-noise ratio.

### Cylindrical Ion Trap Mass Spectrometer

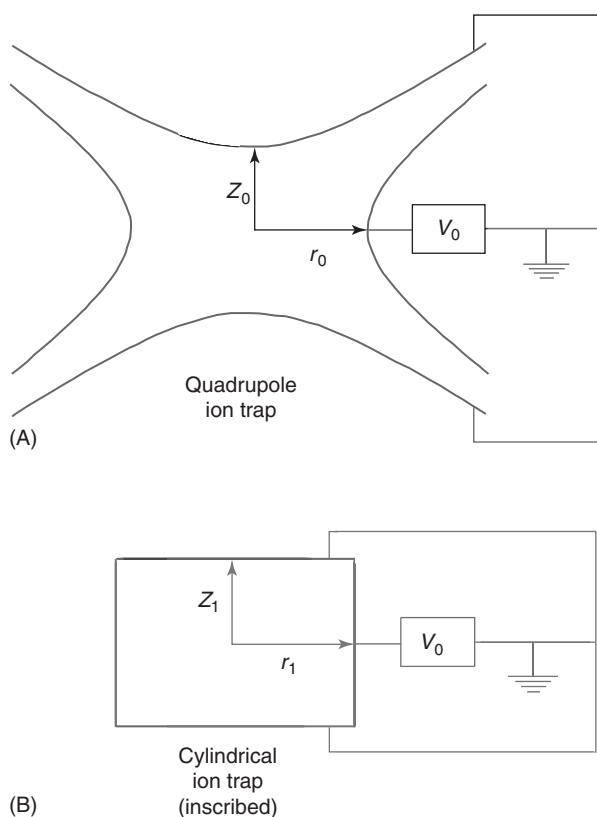
A cylindrical ion trap, CIT, configured as the inscribed cylinder to a 3D quadrupole ion trap is shown in **Figure 6**. A CIT can be fabricated readily from a polished metal cylinder with an internal radius of a few millimeters and two disks perforated for admission of electrons or ions and for ejection of ions.

The cylindrical device is easier to construct than a 3D ion trap having hyperbolic electrodes, and variation of the ratio  $r_c:2z_c$  (radius of cylinder to height of CIT) is facile. Research into CITs has focused upon small CITs, described as mini-CITs, for application as tandem mass spectrometers and for the development of arrays of mini-CITs. In order to achieve in a mini-CIT a given value of  $q_z (= -q_x \text{ from eqn [2]})$ , an elevated drive frequency is required to compensate for the reduced value of  $r_c$ . In this manner, a potential well of sufficient magnitude is facilitated that permits excitation of ions confined within the mini-CIT.

**Miniature cylindrical ion traps (mini-CITs)** A mini-CIT is defined according to the radius,  $r_c$ , of the ring or barrel electrode; when  $r_c$  is <5 mm, the CIT is classed as miniature. Mass analysis using the



**Figure 5** Product ion mass spectrum of protonated reserpine,  $m/z$  609, obtained with the  $Q_2$  linear ion trap at a mass scan rate of  $5200 \text{ Th s}^{-1}$ . (Reproduced with permission from Hager JW (2002) A new linear ion trap mass spectrometer. *Rapid Communications in Mass Spectrometry* 16: 512–526. © 2002, Wiley & Sons Ltd.)



**Figure 6** Comparison of geometries and operating modes for the three-dimensional quadrupole ion trap (QUISTOR) and a cylindrical ion trap. (A) QUISTOR,  $r_0^2 = 2z_0^2$ ; (B) 'inscribed' cylindrical ion trap,  $r_1^2 = 2z_1^2$ . (Reproduced with permission from March RE, Hughes RJ, and Todd JFJ (1989) *Quadrupole Storage Mass Spectrometry*. Chemical Analysis Series, vol. 102. New York: Wiley. © 1989 Wiley & Sons.)

mass-selective axial instability scan in a CIT mass spectrometer has been demonstrated as has  $MS^n$ . The driving force for the development of CITs as field-portable instruments is the need for *in situ* analysis of toxic pollutants and detection of chemical warfare (CW) agents. A field-portable CIT must have high sensitivity and specificity, good precision and accuracy, high speed of analysis, be capable of continuous operation *in situ* for extended periods, and be of low weight, compact, and battery powered.

**Miniaturization** Cooks and co-workers have described the construction and performance of a mini-CIT mass spectrometer that has a mass/charge ratio range of  $\sim 250 \text{ Th}$  and tandem mass spectrometric capability. The cylindrical barrel is of internal radius,  $r_c = 2.5 \text{ mm}$  and the separation ( $2z_c$ ) of the end-cap electrode disks is  $5.77 \text{ mm}$ . A schematic representation of the ion source, ion optics, mini-CIT, and detector is shown in Figure 7. The end-cap electrode spacing was adjusted to  $5.77 \text{ mm}$  to add a small octapolar field to compensate for field defects due to perforations in the end-cap electrodes.

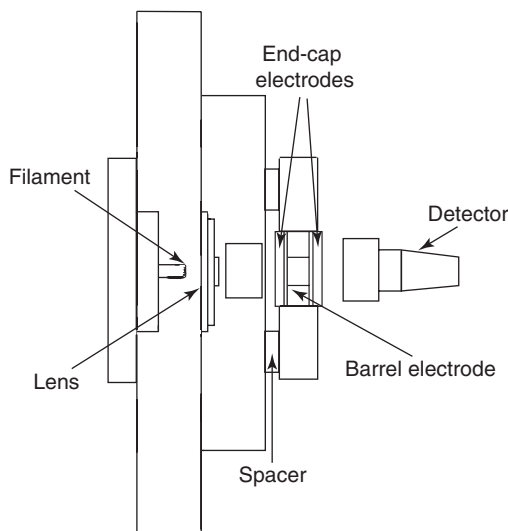
These electrode dimensions provided a mass/charge ratio range of  $\sim 250 \text{ Th}$  at maximum drive RF voltage amplitude of  $800V_{0-p}$  oscillating at  $2 \text{ MHz}$ . For a typical tandem mass spectrometer experiment, these experimental parameters correspond to irradiation of  $m/z$  250 at, say,  $q_z = 0.4$  ( $352V_{0-p}$ ) to yield a product ion mass spectrum from  $m/z$  110 to  $m/z$  250. The trapping well depth,  $D_z$ , for  $m/z$  250

at  $q_z = 0.4$  is given as

$$D_z \approx \frac{V_{0-p} q_z}{8} \quad [4]$$

such that  $D_z \approx 352 V_{0-p} \times 0.4/8 = 17.6 \text{ eV}$ , that is, more than sufficient for CID.

**Portable miniature CIT** The mini-CIT mass analyzer accounts for <1% of the total volume and weight of the instrument; thus, the major miniaturization effort

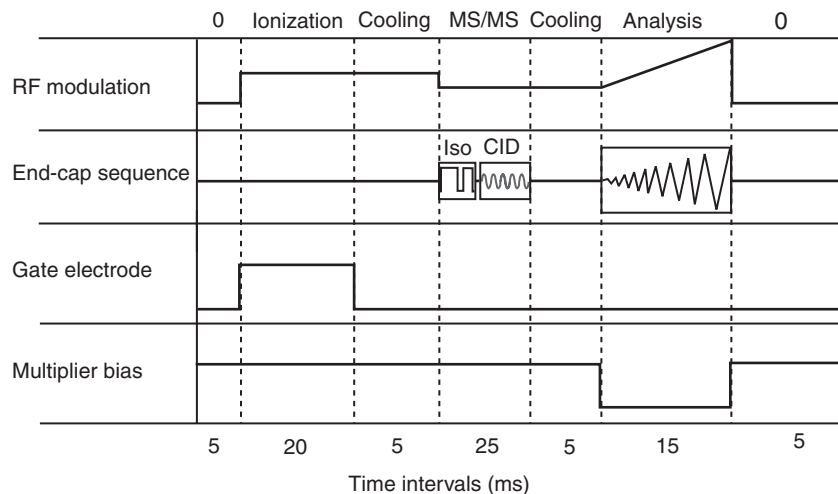


**Figure 7** A schematic representation of the ion source, ion optics, CIT, and detector. The filament, mounting hardware, and gate electrode are based on those used in the Finnigan ITS-40-instrument. (Reproduced with permission from Patterson GE, Guymon AJ, Riter LS, *et al.* (2002) Miniature cylindrical ion trap mass spectrometer. *Analytical Chemistry* 74: 6145–6153. Copyright (2002) American Chemical Society.)

was directed to the ancillary components. The weight of version 5 mini-CIT was 35 kg including the computer (plus 20 kg for the system battery), and measured  $\sim 45 \times 60 \times 71 \text{ cm}$ ; a smaller version 7 instrument of 16 kg and measuring  $28 \times 70 \times 18 \text{ cm}$  has also been constructed. In a mini-CIT, the oscillation amplitude of an ion at a given secular frequency is smaller than in an ion trap of standard size; hence, ions experience fewer ion/neutral collisions per rf cycle and these collisions are of diminished collision energy. Therefore, a mini-CIT can be operated at a higher-than-normal pressure and the pumping requirements are reduced.

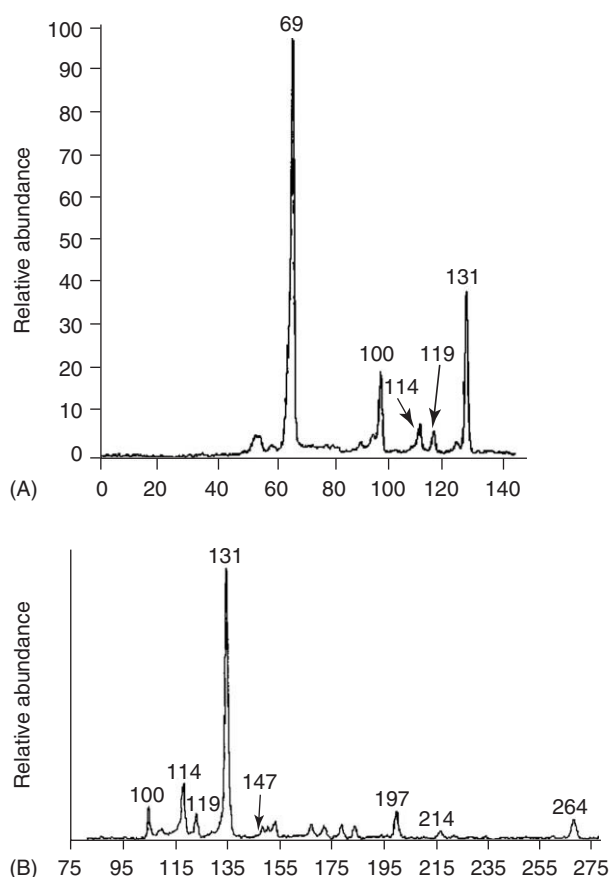
**Ionization source** Electron impact (EI) ionization is employed using a standard EI filament assembly from a Finnigan ITS-40 ion trap instrument. The entire circuit, including power unit and test points, is enclosed in a  $5 \times 5 \times 1 \text{ cm}$  volume. The electron flow into the CIT is gated during ion cooling and mass analysis.

**Waveforms** Operation of a mini-CIT requires five waveforms: drive rf (2.000 MHz) and rf amplitude modulation potentials, ion isolation waveform, and ac potentials for ion excitation and for mass-selective axial ejection. Typical ac amplitude used in the mini-CIT for CID is  $\sim 2V_{0-p}$  and is applied in dipolar mode to the end-cap electrodes. The frequency used for resonant ejection is an octapole resonance that causes rapid ejection of ions. In **Figure 8** is shown a typical timing diagram of scan function by which an EI mass spectrum may be obtained. A scan function is a diagram showing the temporal relationships of the various potentials applied during



**Figure 8** Timing diagram or scan function for a miniature CIT. (Reproduced with permission from Patterson GE, Guymon AJ, Riter LS, *et al.* (2002) Miniature cylindrical ion trap mass spectrometer. *Analytical Chemistry* 74: 6145–6153. Copyright 2002, by American Chemical Society.)





**Figure 9** Mass spectra obtained from PFTBA with two miniature CITs. (A) A lower mass range obtained with version 5 miniature CIT. (B) A higher range mass spectrum, also of PFTBA, obtained with version 7 miniature CIT. (Reproduced with permission from Patterson GE, Guymon AJ, Riter LS, *et al.* (2002) Miniature cylindrical ion trap mass spectrometer. *Analytical Chemistry* 74: 6145–6153. Copyright 2002, by American Chemical Society.)

the course of an ion trap experiment. The duration of application of each of the potentials and of each cooling period is indicated. The isolation waveform is indicated schematically as a square wave with an interposed slot corresponding to the ion secular frequency of the isolated species. The CID waveform is a single frequency sine wave. The analysis waveform is a fixed frequency sine wave increasing in amplitude at a constant rate.

### System performance

**Mass calibration** In Figure 9 are shown two mass spectra obtained with mini-CITs using perfluorotri-*n*-butylamine (PFTBA), a commonly used mass calibrant. The low range mass spectrum in Figure 9A was obtained with version 5 mini-CIT using a scan function similar to that shown in Figure 8. A mass range from 48 to 250 Th was scanned in 15 ms, giving a ramp rate of  $\sim 13\,500\text{ Th s}^{-1}$ . The expected base peak is  $m/z$  69 ( $\text{CF}_3^+$ ). The higher range mass spectrum of PFTBA in Figure 9B was obtained using the smaller version 7 mini-CIT. The mass calibrations were linear with  $R^2 = 0.9999$ .

**Tandem mass spectrometry** In Table 1 are shown mass spectra from some mini-CIT experiments together with mass spectra obtained by MS/MS and MS/MS/MS. The intensity of each ion species is given in parentheses in units of volts to illustrate the absolute performance of the instrument.

The major peaks observed from *p*-nitrotoluene are shown in the MS column; the base peak is  $m/z$  137 of intensity 4.8 V. The MS/MS isolation column gives the peak heights of ions after isolation but prior to CID. For *p*-nitrotoluene, the  $m/z$  137 ion has an

**Table 1** Single-stage MS, MS/MS, and MS/MS/MS spectra of *p*-nitrotoluene and acetophenone<sup>a</sup>

Compound	MS <sup>b</sup>	MS/MS		MS/MS/MS	
		Isolation <sup>b</sup>	Dissociation <sup>c</sup>	Isolation <sup>b</sup>	Dissociation <sup>c</sup>
<i>p</i> -Nitrotoluene	137 (4.8)	137 (4.6)	137 (2.0)	121 (0.5)	121 (0.5)
		121 (0.4)	121 (2.0)		107 (0.08)
			107 (0.2)		91 (0.1)
			91 (0.3)		
	121 (1.1)				
Acetophenone	107 (1.4)				
	91 (1.2)				
	120 (1.3)	120 (1.0)	120 (0.4)	105 (0.6)	105 (0.4)
		105 (0.2)	105 (0.7)		77 (0.06)
			77 (0.06)		
	105 (2.0)				
	77 (1.2)				

<sup>a</sup>The data were obtained from pure samples of the compounds. In all cases, the data are given as mass/charge ratio with the peak height given in volts in parentheses.

<sup>b</sup>No bath or buffer or collision gas.

<sup>c</sup>Helium was used as the bath/buffer/collision gas.

intensity of 4.6 V following isolation; thus the isolation procedure has an efficiency of 96%. The signal intensity (0.4 V) of  $m/z$  121 constitutes a percentage residual ion abundance of 8%.

In the MS/MS dissociation column are presented the components of the product ion mass spectrum of  $m/z$  137; CID of  $m/z$  137 yields a residual ion signal intensity of 2.0 V for  $m/z$  137, and intensities of 2.0, 0.2, and 0.3 V for  $m/z$  121, 107, and 91, respectively. Thus, the attenuation of the molecular ion of  $m/z$  137 is 56% (i.e., a dissociation efficiency of 44%). The total ion intensity of the components of the product ion mass spectrum is 4.5 V that, when compared with the total ion intensity after the isolation stage, 5 V, yields an overall ion retention efficiency during CID of 90%.

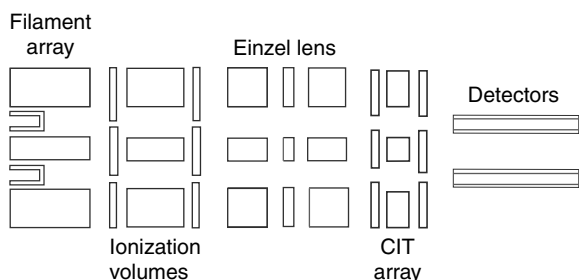
The fifth and sixth columns show data from MS/MS experiments that permit evaluation of the isolation and CID efficiencies and retention of total

ion charge; acetophenone provides an almost ideal example. From the fourth and fifth columns, the isolation efficiency for  $m/z$  105 is shown to be 86% and, from the fifth and sixth columns, the dissociation efficiency is 33% while the ion retention efficiency is 77%. The MS<sup>3</sup> product ion mass spectrum shows unchanged  $m/z$  105 and a fragment ion of  $m/z$  77. The overall efficiency of the formation of  $m/z$  77, through two repeated processes of isolation and CID, from  $m/z$  120 is 5%. With complete dissociation of  $m/z$  120 and  $m/z$  105, the overall efficiency could be increased to 19%; this MS<sup>3</sup> achievement is remarkable for a mini-CIT.

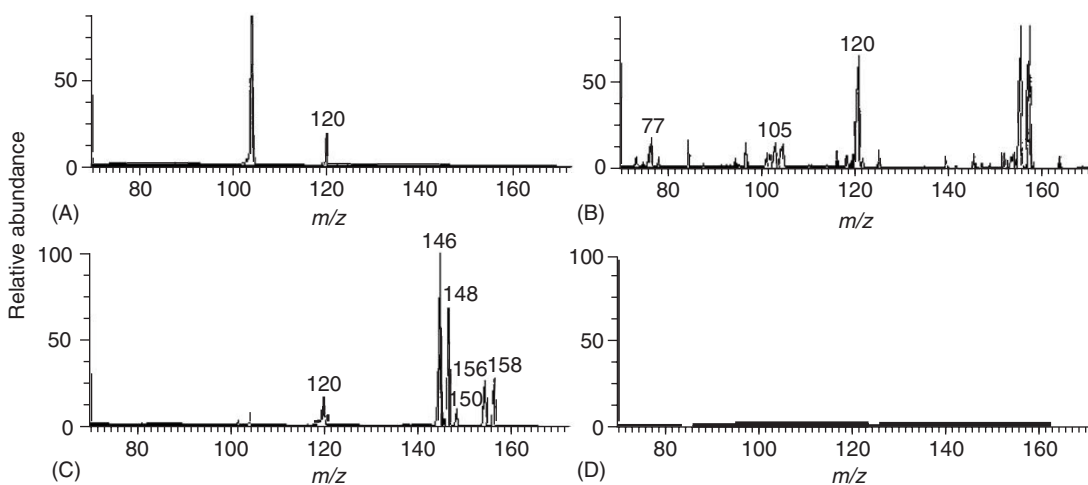
**Limit of detection** The limit of detection for the CW simulant, methyl salicylate, was estimated as 1 pg.

**Miniature cylindrical ion trap array** An approach to the rapid screening of large numbers of samples in the areas of proteomics, industrial process monitoring, and metabolomics is to use a multiplexed inlet system with a multiplexed mass spectrometer having an equal number of parallel sample channels. To this end, a high-throughput mini-CIT array mass spectrometer has been developed and tested.

The array consists of four mini-CITs arranged like the rods of a mass filter shown in **Figure 1**. Each mini-CIT has a barrel electrode with  $r_c = 1.25$  mm, its own inlet system, filament assembly, and detector, all within a common vacuum manifold and with a single set of control electronics. Sample ions are generated externally to the mini-CIT and transported through an Einzel lens into each mini-CIT, as shown



**Figure 10** A schematic representation of two of the four parallel analysis channels in the  $2 \times 2$  array of mini-CITs. (Reproduced with permission from Tabert AM, Griep-Raming J, Guymon AJ, and Cooks RG (2003) High-throughput miniature cylindrical ion trap array mass spectrometer. *Analytical Chemistry* 75: 5656–5664. Copyright 2003, by American Chemical Society.)



**Figure 11** Mass spectra obtained from a high-throughput experiment during which three channels of the four-channel array instrument were used to analyze three different compounds simultaneously: 1, acetophenone; 2, bromobenzene; 3, 1,3-dichlorobenzene; 4, idle. (Reproduced with permission from Tabert AM, Griep-Raming J, Guymon AJ, and Cooks RG (2003). High-throughput miniature cylindrical ion trap array mass spectrometer. *Analytical Chemistry* 75: 5656. Copyright 2003, by American Chemical Society.)

in Figure 10 for two of the four parallel channels. The mass/charge ratio range is  $m/z$  50–500 and a mass resolution ( $m/\Delta m$ ) of 1000 is obtained at  $m/z$  300. Because each channel has its own inlet system, four compounds can be examined in the array. In Figure 11 are displayed the results of a high-throughput experiment in which three of the four channels were operated simultaneously; acetophenone, bromobenzene, and 1,3-dichlorobenzene were introduced into channels 1, 2, and 3, respectively. Channel 4 was idle. The successful operation of this array of mini-CITs is a remarkable achievement.

See also: **Gas Chromatography:** Mass Spectrometry. **Mass Spectrometry:** Mass Separation; Selected Ion Monitoring; Environmental Applications; Peptides and Proteins.

## Further Reading

- Badman ER and Cooks RG (2000) A parallel miniature cylindrical ion trap array. *Analytical Chemistry* 72: 3291–3297.
- Badman ER and Cooks RG (2000) Miniature mass analyzers. *Journal of Mass Spectrometry* 35: 659–671.
- Hao C and March RE (2001) A survey of recent research activity in quadrupole ion trap mass spectrometry. *International Journal of Mass Spectrometry* 212: 337–357.
- Henry CM (2002) Itsy-bitsy mass spectrometers. *Chemical Engineering News*, April 8, 34.
- March RE (2000) Quadrupole ion trap mass spectrometry: a view at the turn of the century. *International Journal of Mass Spectrometry* 200: 285–312.
- Ritter LS, Peng Y, Noll RJ, *et al.* (2002) Analytical performance of a miniature cylindrical ion trap mass spectrometer. *Analytical Chemistry* 74: 6154–6162.

## Time-of-Flight

**M Guilhaus**, University of New South Wales, Sydney, NSW, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Patrick M Thompson, pp. 2923–2932, © 1995, Elsevier Ltd.

## Introduction

The measurement of the time-of-flight (TOF) of ions of near constant energy has been used to determine ion mass-to-charge ratios ( $m/z$ ) since the 1940s. However, it was not until 50 years later that TOF quickly became a mainstream device in mass spectrometry (MS). The reasons for the renaissance of this technique include: the development of supporting technologies of high-speed timing electronics (especially high-speed signal digitizers); high-speed data processing capability of computers; development of focal plane electron multipliers that deliver pulses with nanosecond rise-times; the invention of the ion mirror; development of matrix-assisted laser desorption ionization (MALDI) sources; and the development orthogonal acceleration instrument geometries for gating ions from a range of continuous ion sources that are widely used. Prior to these developments, TOF-MS was not able to provide selectivity (mass resolution) or sensitivity that was competitive with scanning mass spectrometers based on quadrupole and magnetic sector technologies.

During the 1990s, the increasing utility of MALDI and electrospray ionization (ESI) sources to deliver

intact gas-phase ions of polar, thermally labile, and high molecular weight biomolecules provided impetus to develop TOF-MS. This was because of the intrinsic ability of TOF-MS to transmit and detect ions of high  $m/z$  and its emerging high sensitivity and speed. The potential of TOF for bioanalysis was foreshadowed by the  $^{252}\text{Cf}$  Plasma Desorption mass spectrometers in the 1970s; however, the greater utility of MALDI and ESI and the existence of practical approaches to combine these sources with TOF have led to a proliferation of TOF-MS.

The majority of applications of TOF-MS are currently in the area of bioanalytical chemistry and particularly the analysis of whole proteins, their peptide fragments, and nucleic acids. The importance of protein analysis is a consequence of the explosion of data from mapping of the human genome and genomes of several other species. The genes express proteins and the function of specific proteins (hence that of their genes) is now a critical concern in many areas of medicine, drug discovery, and biochemistry. TOF-MS together with advancing separation techniques has become an enabling technology in the life sciences because of its remarkable ability to characterize these molecules even when they are present in very low abundance in complex mixtures.

The growing demands from life science research is a major force driving the development TOF-MS technology. The maturation of the technology has allowed TOF-MS to be adopted as a mainstream mass spectrometer with the special advantages of high speed as well as high resolution and mass

in Figure 10 for two of the four parallel channels. The mass/charge ratio range is  $m/z$  50–500 and a mass resolution ( $m/\Delta m$ ) of 1000 is obtained at  $m/z$  300. Because each channel has its own inlet system, four compounds can be examined in the array. In Figure 11 are displayed the results of a high-throughput experiment in which three of the four channels were operated simultaneously; acetophenone, bromobenzene, and 1,3-dichlorobenzene were introduced into channels 1, 2, and 3, respectively. Channel 4 was idle. The successful operation of this array of mini-CITs is a remarkable achievement.

See also: **Gas Chromatography:** Mass Spectrometry. **Mass Spectrometry:** Mass Separation; Selected Ion Monitoring; Environmental Applications; Peptides and Proteins.

## Further Reading

- Badman ER and Cooks RG (2000) A parallel miniature cylindrical ion trap array. *Analytical Chemistry* 72: 3291–3297.
- Badman ER and Cooks RG (2000) Miniature mass analyzers. *Journal of Mass Spectrometry* 35: 659–671.
- Hao C and March RE (2001) A survey of recent research activity in quadrupole ion trap mass spectrometry. *International Journal of Mass Spectrometry* 212: 337–357.
- Henry CM (2002) Itsy-bitsy mass spectrometers. *Chemical Engineering News*, April 8, 34.
- March RE (2000) Quadrupole ion trap mass spectrometry: a view at the turn of the century. *International Journal of Mass Spectrometry* 200: 285–312.
- Ritter LS, Peng Y, Noll RJ, *et al.* (2002) Analytical performance of a miniature cylindrical ion trap mass spectrometer. *Analytical Chemistry* 74: 6154–6162.

## Time-of-Flight

**M Guilhaus**, University of New South Wales, Sydney, NSW, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Patrick M Thompson, pp. 2923–2932, © 1995, Elsevier Ltd.

## Introduction

The measurement of the time-of-flight (TOF) of ions of near constant energy has been used to determine ion mass-to-charge ratios ( $m/z$ ) since the 1940s. However, it was not until 50 years later that TOF quickly became a mainstream device in mass spectrometry (MS). The reasons for the renaissance of this technique include: the development of supporting technologies of high-speed timing electronics (especially high-speed signal digitizers); high-speed data processing capability of computers; development of focal plane electron multipliers that deliver pulses with nanosecond rise-times; the invention of the ion mirror; development of matrix-assisted laser desorption ionization (MALDI) sources; and the development orthogonal acceleration instrument geometries for gating ions from a range of continuous ion sources that are widely used. Prior to these developments, TOF-MS was not able to provide selectivity (mass resolution) or sensitivity that was competitive with scanning mass spectrometers based on quadrupole and magnetic sector technologies.

During the 1990s, the increasing utility of MALDI and electrospray ionization (ESI) sources to deliver

intact gas-phase ions of polar, thermally labile, and high molecular weight biomolecules provided impetus to develop TOF-MS. This was because of the intrinsic ability of TOF-MS to transmit and detect ions of high  $m/z$  and its emerging high sensitivity and speed. The potential of TOF for bioanalysis was foreshadowed by the  $^{252}\text{Cf}$  Plasma Desorption mass spectrometers in the 1970s; however, the greater utility of MALDI and ESI and the existence of practical approaches to combine these sources with TOF have led to a proliferation of TOF-MS.

The majority of applications of TOF-MS are currently in the area of bioanalytical chemistry and particularly the analysis of whole proteins, their peptide fragments, and nucleic acids. The importance of protein analysis is a consequence of the explosion of data from mapping of the human genome and genomes of several other species. The genes express proteins and the function of specific proteins (hence that of their genes) is now a critical concern in many areas of medicine, drug discovery, and biochemistry. TOF-MS together with advancing separation techniques has become an enabling technology in the life sciences because of its remarkable ability to characterize these molecules even when they are present in very low abundance in complex mixtures.

The growing demands from life science research is a major force driving the development TOF-MS technology. The maturation of the technology has allowed TOF-MS to be adopted as a mainstream mass spectrometer with the special advantages of high speed as well as high resolution and mass

accuracy. Consequently, it is experiencing increasing use in areas such as environmental sciences and is enabling new areas such as high-speed gas chromatography/mass spectrometry (GC-MS) and comprehensive GC-MS (GC  $\times$  GC-MS). The ability of benchtop GC-TOF instruments to work with field ionization sources and provide accurate molecular weight information for many compounds that do not give molecular ions under standard conditions of electron ionization is also noteworthy.

This article will present the basic theory of TOF-MS and will consider the most important instrument configurations and applications. The utility of TOF-MS compared to other mass analyzers will be discussed.

## Theory of Operation

In the simplest TOF-MS system an ideal population of ions of different masses is located in a plane at a fixed distance from a detector and accelerated from rest toward the detector. Acceleration is caused by the application of typically one or two electric fields set up between grids in the accelerator as shown in **Figure 1**. Ions of mass  $m$  and charge  $q$  (where  $q$  is accounted for by an integer number of electron charges  $ze$ ) are either generated in bunches or accelerated as a bunch by pulsing the first field. The two fields –  $E_1$  and  $E_2$  – are aligned in the TOF direction. Initiation of acceleration is a start event and this triggers a timing device. An ion will travel a distance of  $s_1$  then  $s_2$  in the fields gaining kinetic energy  $q(E_1s_1 + E_2s_2)$ . The length of the combined acceleration regions is typically a few centimeters. This is related to velocity,  $u_{\text{TOF}}$ , in the TOF direction by  $0.5u_{\text{TOF}}^2$  (eqn [1]). A drift region of length  $D$  follows and the time for ions to traverse it is given by  $t_D$  in eqn [3]. Drift regions vary in length from about 0.5 m to 3 m. Since the velocity is inversely proportional to the square root of  $m/z$  (eqn [2]), the ions will separate as lighter ions traverse the drift region faster than heavier ions (eqn [3]). A planar detector is normally

used to convert the ion arrival event into an electrical signal whose time relative to the start event is recorded digitally. The ion arrival events distribute along a  $(m/z)^{0.5}$  axis that is reliably calibrated:

$$0.5mu^2 = ze(E_1s_1 + E_2s_2) \quad [1]$$

$$u_{\text{TOF}} = \sqrt{\frac{2ze(E_1s_1 + E_2s_2)}{m}} \quad [2]$$

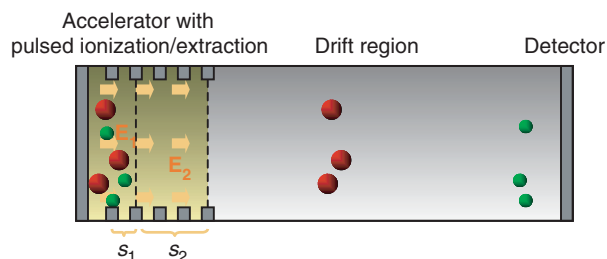
$$t_D = \frac{D}{e^{0.5}(E_1s_1 + E_2s_2)^{0.5}} \times (m/z)^{0.5} \quad [3]$$

The total ion drift energy in TOF-MS is typically 2–30 keV and this gives a flight time for the heaviest ion in the order of 50–100  $\mu\text{s}$ . Thus, the repetition rate of the TOF-MS is  $\sim 10$ –20 kHz and one repetition is often referred to as a ‘shot’. Ions differing in mass by 1 Da are typically separated in arrival time by several nanoseconds and thus a digitizer with a sampling rate of at least  $2 \times 10^8 \text{ s}^{-1}$  is required.

Ions’ arrival times for a group of ions of the same  $m/z$  distribute to form a peak centered at a time  $t$  easily computed from the laws of motion. A distribution of ion arrival time with a full-width at half-maximum height (FWHM) of  $\Delta t$  results from nonideal starting parameters: the ions do not all start at the same distance from the detector; the ions do not all start from rest but are typically subject to a Boltzmann distribution of molecular speed in all directions; and the ions do not all form at the same time and/or the registration of their start events is subject to error. These three factors are, respectively, referred to as the initial spatial, velocity, and temporal dispersions. In addition, the arrival of the ion at the detector is registered as a pulse with a more-or-less constant rise-time and fall-time making a further contribution to temporal spread. The effect of  $\Delta t$  is to place a limit on the difference in  $m/z$  that can be discerned by TOF-MS. Resolution in MS is expressed as  $m/\Delta m$  where for singly charged ions,  $\Delta m$  is the FWHM mass width of a peak at mass  $m$ . Differentiating eqn [3] with respect to  $m$  allows resolution to be expressed conveniently in terms of time:

$$m/\Delta m = t/2\Delta t \quad [4]$$

The temporal dispersion is typically a constant contributor to  $\Delta t$  (i.e., independent of  $m/z$ ); however, the contribution of spatial and velocity dispersions to  $\Delta t$  scales with  $m/z$ . The initial dispersions are typically uncorrelated so their effects add in quadrature. It is usual that the temporal dispersion is small compared to the others except at very low  $m/z$  and therefore it has an insignificant effect on resolution at higher masses. Therefore, the resolution of a TOF mass



**Figure 1** Schematic diagram of a linear TOF-MS showing light (green) and heavy (red) ions in the accelerator and later in the drift region. Two accelerator fields are shown.



spectrometer is nearly constant over most of its mass range and is significantly lower than this value only at very low mass.

It is possible to correct for initial dispersions in isolation. When two or more dispersions are significant and not correlated there is no generally practical way to compensate for their effects. Mass resolution is therefore limited. However, some fundamental focusing methods to address individual dispersions together with novel ion-optical configurations provide mass resolution in the range 5000–20 000 in commercial TOF-MS systems.

### Spatial Focusing

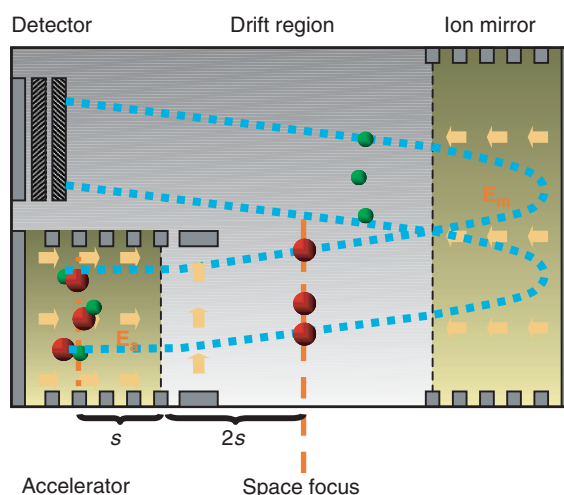
This method to correct for a spread of initial ion position was first described by Wiley and McLaren in the 1950s. The acceleration fields are arranged so that ions (of the same  $m/z$ ) that are further from the detector are accelerated through a longer distance ( $s_1$ ) and reach a higher velocity than their counterparts that are initially closer to the detector. The faster ions catch up with the slower ions at the detector and arrive simultaneously. The residual arrival-time spread from spatial spread can be kept insignificant for typical spatial dispersions of  $\sim 2$  mm.

### Time-Lag Focusing

The initial velocity spread is converted to a spatial spread by allowing a delay between ionization and acceleration. Ions separate to create a correlation between position and velocity. Spatial focusing can be applied to correct this. This method was also developed by Wiley and McLaren and the conditions (length of delay) must be adjusted for each  $m/z$ . Time-lag focusing does not work well when a significant spatial and velocity spread are present simultaneously. As will be discussed later, time-lag focusing has recently been adapted to greatly increase the resolution in MALDI TOF-MS.

### Energy and Spatial Focusing Devices

In a TOF-MS configuration named the ‘mass reflectron’, Mamyryn and Shmikk found a way to increase resolution by employing an electrostatic ion mirror in the drift region. The effect of initial velocity and spatial dispersions is minimized by applying a strong single acceleration field and this creates a spatial focus at  $2s_1$  in the drift region. Ions of the same mass pass through this point in a spatially and temporally tight bunch but with a large energy spread. The more energetic (faster) ions penetrate more deeply into the ion mirror before they turn around. The time to travel through the mirror is inversely proportional to



**Figure 2** Schematic diagram of a reflecting TOF-MS showing light (green) and heavy (red) ions in the accelerator and later in the drift region. Accelerator, deflection, and mirror fields are shown.

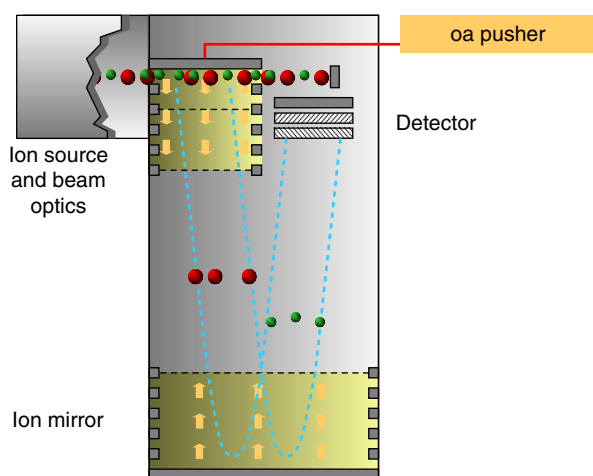
the square root of the ion's drift energy. The slower ions are allowed to catch up with faster ones that have further to travel. The arrangement is configured so the point at which the ions catch up is at the detector. The configuration of the ion mirror is shown in Figure 2. Another advantage of the ion mirror is that the drift region is effectively used twice and this reduces instrument size. A long total drift length is important because it increases  $t$  while the correction devices keep  $\Delta t$  small. Resolution is substantially increased by increasing the drift distance. One recent instrument is configured with a second mirror at the opposite end of the drift region relative to the main mirror. Ions therefore are made to traverse the drift region four times before detection. This device is called a ‘W’ mirror because it creates a trajectory shaped like the letter ‘W’. It significantly increases resolution (a factor of  $\sim 2$ ) but the cost is that a significant loss of ion transmission occurs because ions make more passes through grids (typical grid transmission 50–90%).

Another device that provides isochronous focusing with respect to energy in TOF-MS is the electric sector as introduced by Poschenrieder. As with the mirror, the more energetic ions travel a longer trajectory and therefore take longer to travel through the sector. Slower ions are once again allowed to catch up at the detector. Electric sectors usually require slits that substantially reduce ion transmission. Ion mirrors have generally higher transmission than the electric sector and together with the convenience of doubling the drift distance without increasing the size of the instrument they are the preferred energy correcting device in most TOF-MS systems.

## Orthogonal Acceleration

A continuous or pseudocontinuous ion beam is arranged orthogonally to a TOF mass spectrometer. Orthogonal acceleration (oa) was developed independently by Guilhaud and Dodonov in the 1980s and it is an excellent way to couple continuous ion sources to a TOF mass analyzer. In oa-TOF-MS, a nearly parallel ion beam has ideally no velocity spread in the direction orthogonal to it (Figure 3). This direction is chosen for the TOF mass analysis and the finite spatial spread is corrected with a linear or reflecting instrument geometry. The arrangement has several important advantages:

- The small zero-centered velocity spread in the TOF direction allows excellent resolution and mass accuracy after the spatial spread is corrected.
- A substantial length of the continuous ion beam may be sampled and the energies of the ion beam and the orthogonally accelerated ion packets can be arranged so that the flight time in the mass analyzer is about the same as the time for a new section of beam to enter the accelerator. This results in an increase in the duty cycle, i.e., a substantial increase in sensitivity compared to scanning mass spectrometers because much greater proportion of ions produced can be detected.
- Most continuous ion sources are suited to oa-TOF-MS; it can also be used effectively as the final mass analyzer in hybrid instruments which combine two mass spectrometers in an approach called tandem mass spectrometry or MS/MS.



**Figure 3** Arrangement of an oa-TOF-MS. A continuous ion beam enters from the source (top left) and sections of this beam are periodically accelerated from the orthogonal accelerator (oa) in a direction orthogonal to the beam. The packet of ions so formed is analyzed by the reflecting TOF-MS while (simultaneously) a new beam is forming in the oa.

## Ion Detection

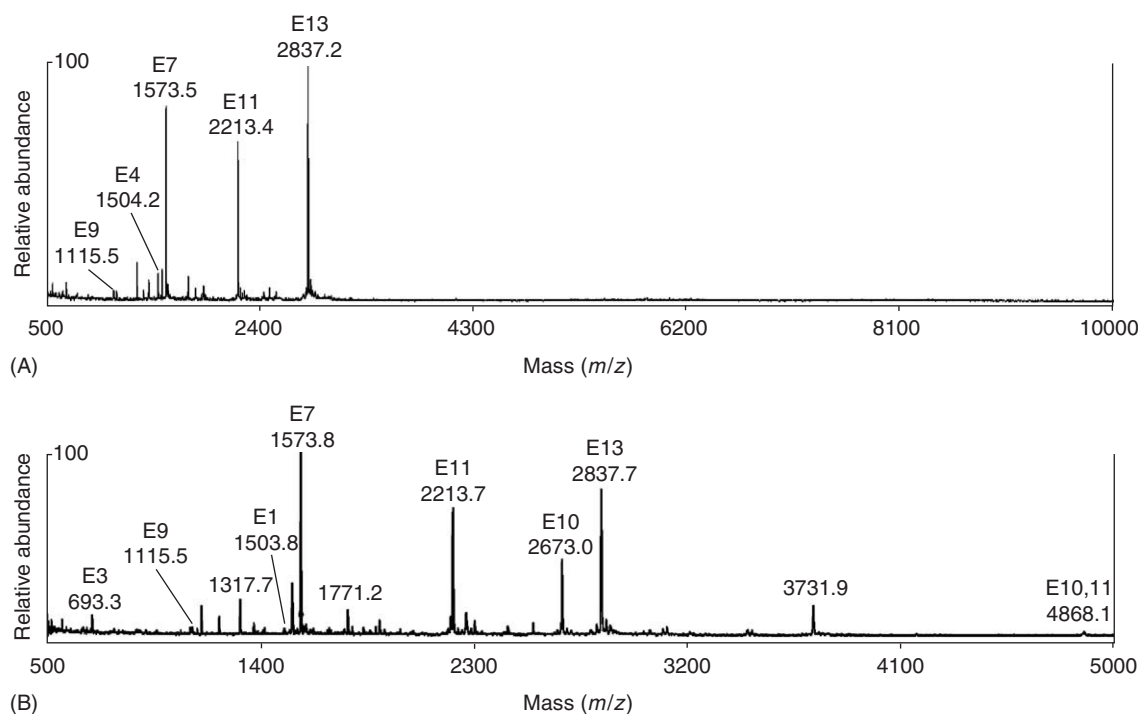
Either a discrete dynode electron multiplier or, more commonly a microchannel plate converts the arrival of an ion into a cascade of secondary electrons. A gain of  $10^5$ – $10^7$  in charge occurs in these detectors. The pulse that results from this charge is detected by a digitizer. Digitizers generally fall into two categories. Time-to-digital converters (TDC) simply record the time of a stop event in a list. The spectrum is generated after several shots have been acquired and a histogram of stop events is constructed. This approach has the advantages that null data are not processed and the rise-time of the stop pulses from the detector is sensed over a threshold giving a very small temporal contribution to the peak width. The TDC has the disadvantage of coincidence losses. If two or more ions arrive within the same time-bin they are seen as one. This limits the dynamic range for high count rates. Transient recorders are an alternative type of digitizer. Here, a fast analog-to-digital converter converts the output of the detector into a time transient. Null data are recorded and must be summed into the final spectrum usually over a few tens of shots. Transient recorders typically operate at sampling rates of  $0.5$ – $4 \times 10^9 \text{ s}^{-1}$  and at these rates the digital amplitude resolution is limited to 8 bits. Summing data from subsequent shots into 16 bit memory increases the dynamic range to  $\sim 10^3$ – $10^4$ . The limitation comes at the low count rates due to digital noise.

## TOF-MS Instrument Configurations and Applications

Applications of MS tend to be based on a few main types of ion sources. The most important of these will be discussed in detail in this article.

### MALDI TOF-MS

In MALDI, a pulsed laser is fired at a solid sample mixed with a large molar excess of a matrix (e.g., 1000:1). Typically, MALDI instruments use a 337 nm pulsed ultraviolet laser and the matrix is a substituted aromatic acid (e.g., 2,5-dihydroxybenzoic acid). A small amount of matrix in an irradiated area of  $\sim 100 \mu\text{m}$  (diameter) efficiently absorbs the laser energy and vaporizes. Sample molecules, typically biopolymers or peptides, are desorbed and ionized in this process (Figure 4). MALDI TOF-MS is an important application accounting for much of TOF's recent revival. The first generation of MALDI instruments suffered from low resolution that further degraded with mass. This was largely due to the energy spreads imparted to desorbing ions in



**Figure 4** Peptide mass fingerprints of recombinant proteins (EPR) used in doping: MALDI TOF-MS spectrum of (A) Glu-C digested EPR and (B) Glu-C digested and deglycosylated EPR. (Reprinted with permission from Stanley and Poljak (2003) *Journal of Chromatography B* 785: 205–218; © Elsevier.)

collisions with the desorbing matrix. Since ions are generated from a surface at a well-defined time (laser pulse) it is possible to correct for the initial spread of energy by delaying the acceleration. The velocity spread prior to acceleration leads to a correlated spatial spread that may be corrected as already discussed. The correction is mass dependent so that optimal resolution is available over a limited mass range. Care must also be used in calibration as the mean desorption velocity depends on experimental conditions (e.g., matrix and laser power) in such a way that the mass calibration relationship is less predictable than TOF-MS with ions starting with zero average velocity. Despite these minor shortcomings, delayed extraction (DE) has greatly improved the selectivity and sensitivity of MALDI and has contributed to the establishment of MALDI as an essential tool in bioanalytical chemistry.

It is often useful to observe the fragmentation patterns of ions in MS. This can give insights into their structure. Ions that fragment in the drift region prior to the ion mirror are called metastable ions. The fragment ion carries a proportion of the original ion energy relative to its fraction of the mass of the precursor. The fragment therefore has a shorter flight-time through the mirror than the precursor and it appears at a lower apparent mass. Conditions in

DE-MALDI TOF-MS systems can be set up to detect these ions and determine their  $m/z$  for structure determination. The technique is called postsource decay and it is particularly important in determining the sequence of amino acids in peptides formed from proteolytic digestion of proteins.

MALDI has recently been shown to be highly suited to orthogonal acceleration TOF-MS. The energy spread of desorbed ions is collapsed by a process of collisional cooling. Ions undergo several nondestructive collisions with a cool collision gas in an ion guide device. The ion guide is typically a radiofrequency-only quadrupole or hexapole. The ions cool to the temperature of the collision gas and achieve a correspondingly narrow energy spread. After collisional cooling the ions are reaccelerated with a dc electric field and formed into a pseudocontinuous ion beam that is the source for an oa-TOF-MS. As well as eliminating the problems associated with energy spreads in the MALDI process, this geometry also provides ions with zero average velocity in the TOF direction and this helps make the calibration of the mass scale more accurate. Resolution of 5000–15 000 can be achieved with these instruments. Another advantage of MALDI oa-TOF-MS is that MS/MS can be done readily with this instrument configuration as discussed under Tandem TOF-MS instruments later in this article.

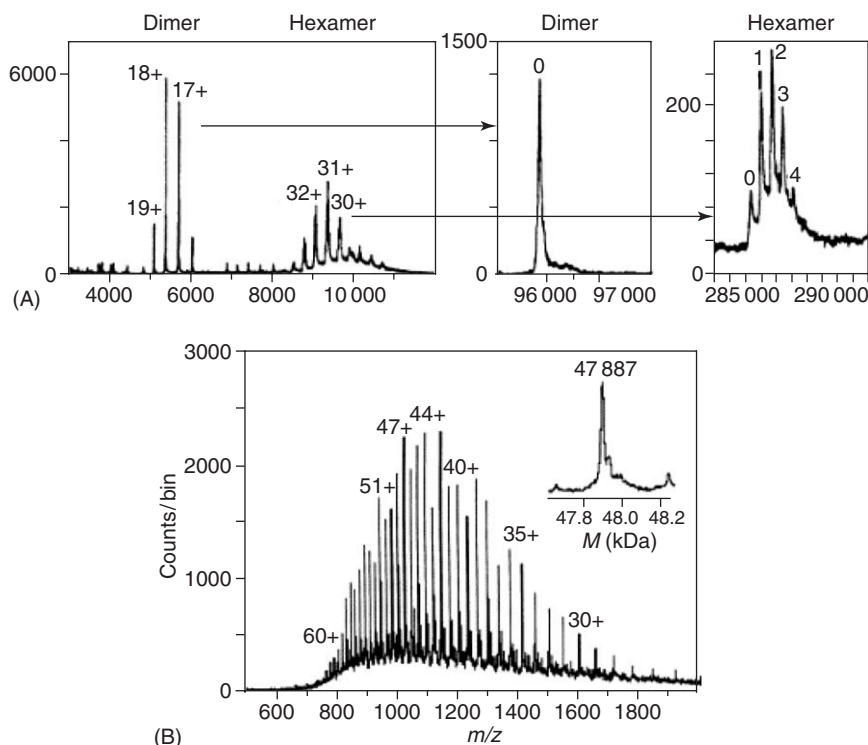
### Electrospray TOF-MS

ESI sources have become widespread in MS because they produce abundant molecular ions for polar and thermally labile molecules over a very wide mass range and the atmospheric pressure ionization technique is ideally suited to coupling with liquid chromatography. Molecules are typically ionized in acidic solution and sprayed from a needle to which high voltage is applied. Tiny droplets of ion containing liquid evaporate and leave intact ions of involatile molecules in the gas phase. These are sampled into the mass spectrometer through an orifice. Multiple charging occurs for larger species. For example, peptides up to a mass of 2 kDa typically form in charge states from +1 to +3. Large intact proteins and nucleic acids with masses  $10^4$ – $10^5$  Da can display charge states in the  $+10^1$  to  $+10^2$  charges. Negative ions may also be selected by reversing ion source potentials and using basic buffers. Initially, ESI was used with quadrupole mass analyzers. Though having a limited mass range compared to the masses of the molecules typically analyzed by ESI, the multiple charging of molecular ions almost always brought  $m/z$  into the range of the quadrupole. In

1991, Dodonov's group demonstrated the orthogonal acceleration instrument geometry with an ESI source. It soon became apparent that the higher sensitivity and the superior resolution of oa-TOF-MS was a compelling reason to develop these instruments commercially. Moreover, within 5 years, extremely useful hybrid tandem quadrupole oa-TOF systems established a new benchmark in sensitivity and selectivity for bioanalytical MS. In addition to these advantages, the ESI TOF-MS configuration has shown that ESI ions of proteins can be generated to retain their structure and noncovalent complexes are observed in relatively low charge states. This gives rise to  $m/z$  well outside the range of quadrupole mass analyzers. For example, the ESI TOF-MS spectrum in Figure 5 reveals the interaction between citrate synthase hexamer and an enzyme inhibitor. The hexamer ions are observed at  $m/z \sim 7000$ .

### MS/MS on Electrospray and MALDI Hybrid Quadrupole TOF-MS Instruments

Tandem mass spectrometry (MS/MS) is now widely used in analytical science as a way to increase specificity and sensitivity in chemical analysis. The

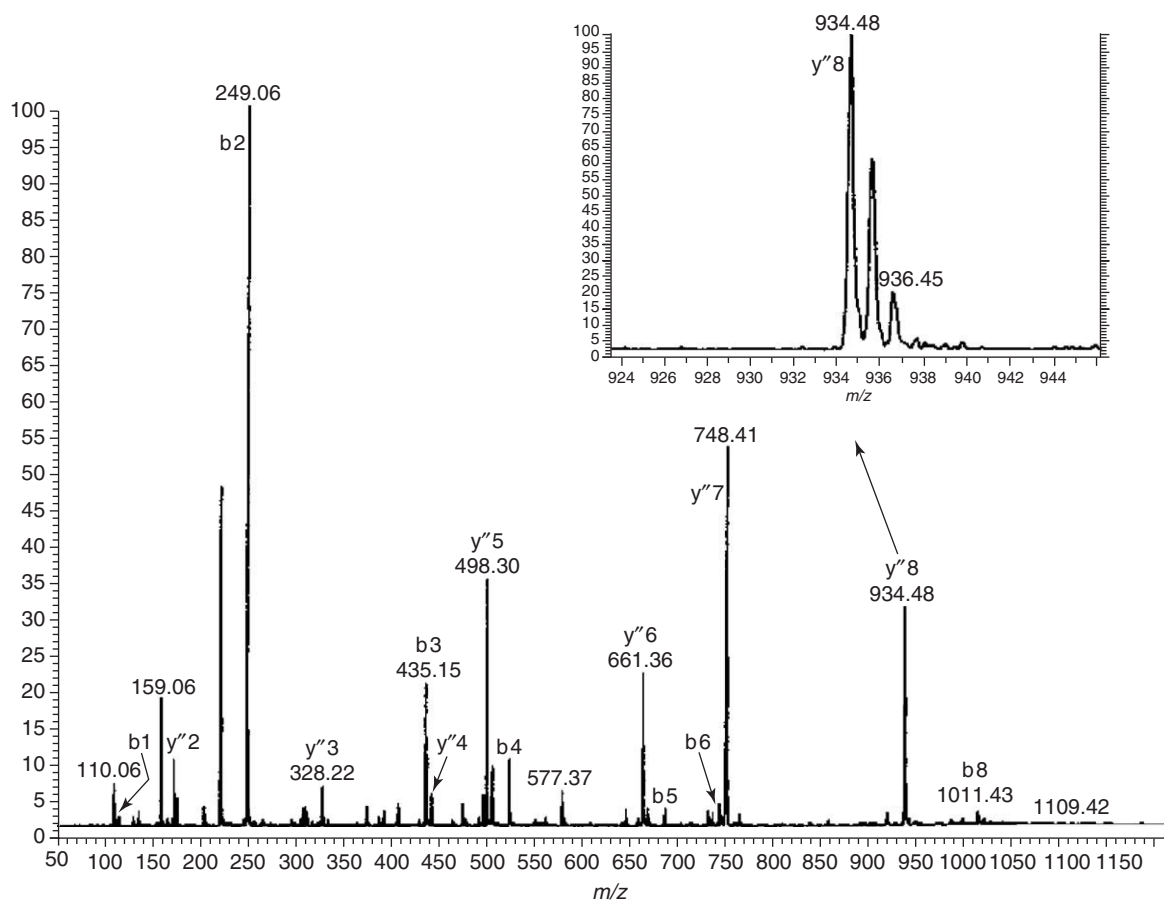


**Figure 5** (A) Electrospray mass spectra of (9 mmol L<sup>-1</sup> subunit concentration) recombinant *Escherichia coli* citrate synthase in a bicarbonate buffer at pH 7.5 with 4.5 mmol L<sup>-1</sup> nicotinamide adenine dinucleotide (NADH) inhibitor. Dimer and hexamer complexes are evident as are the addition of NADH adducts (1–4) to the hexamer. (Reprinted with permission from Ayed *et al.* (1998) *Rapid Communications in Mass Spectrometry* 12: 339–344; © John Wiley & Sons Ltd.) (B) In contrast to A, the spectrum obtained under denaturing conditions (water/methanol 1/1 v/v + 5% acetic acid, pH ~2.5) in which only the monomer in high charge states is observed. (Additional data for B provided by the authors referred to in A.)

use of two or more mass analyzers in series, with a collision cell to promote ion fragmentations between them, allows additional 'purification' by selecting specific precursor molecular ions from a mixture for further analysis and characterization. This is often required when complex mixtures of tryptic peptides are separated by LC and two or more species co-elute. The most widely used MS/MS method in biomolecular analysis is to obtain the product ion spectrum of a precursor molecular ion. For example, **Figure 6** shows the product ions formed when the doubly charged molecular ion of a neuroendocrine peptide is fragmented to reveal the characteristic collisional fragmentation of the amide bonds along the peptide backbone, producing 'b' and 'y' series ions corresponding to charge retention on the N and C terminus, respectively. The resolution and signal-to-noise ratio are impressive in these results from the first reported hybrid quadrupole TOF-MS system shown in **Figure 7**. MS(1) is static and the quadrupole transmits ~100% of the ions to the TOF-MS(2). The efficient duty cycle of the TOF-MS(2)

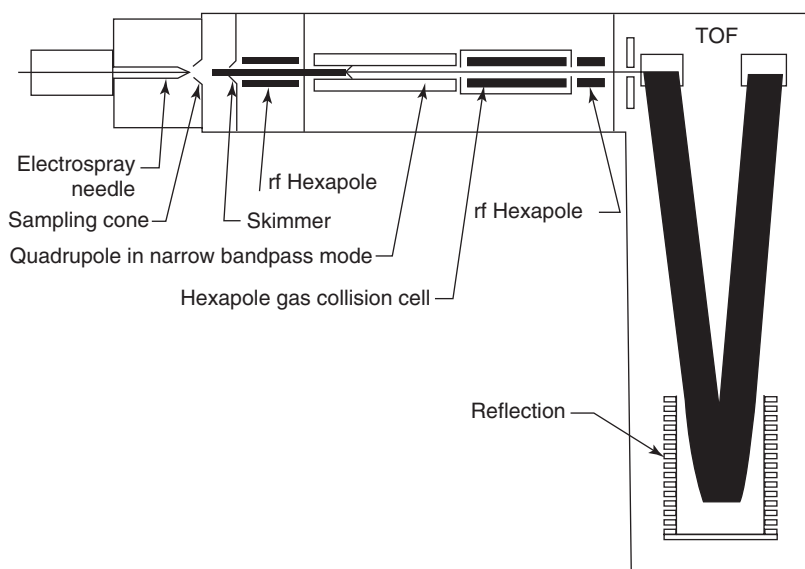
together with its high resolution combine to make this configuration preferable to MS/MS systems based on a scanning MS(2).

Product scans of this type are very important in protein characterization because the difference in mass of the y and b series ion can reveal the amino acid sequence for a domain of the peptide. This process called *de novo* sequencing involves searching for the domain in protein databases and usually a sequence of a few amino acids can lead to successful identification of the original native protein. A particularly powerful feature of hybrid quadrupole TOF-MS systems is their ability to perform information dependent analysis (IDA). During the elution of an LC peak, software automatically identifies the molecular ions and sets up to collect MS/MS product spectra of (typically) doubly charged molecular ions above a preset abundance. TOF-MS is fast and sensitive enough in this application to cope with several (e.g., up to five) peptides co-eluting. More are possible but the quality of the data starts to become marginal. An application where this capability of



**Figure 6** A single acquisition representing 500 amol of a peptide in a neuroendocrine peptide mixture obtained by CID of  $[M + 2H]^{2+}$  at  $m/z$  591.8. (Reprinted with permission from Morris *et al.* (1996) *Rapid Communications in Mass Spectrometry* 10: 889–896; © John Wiley & Sons Ltd.)





**Figure 7** The Q-TOF mass spectrometer. The MS(1) stage and collision region are similar in design to those found on a triple quadrupole instrument. MS(2) is a reflecting oa-TOF design. (Reprinted with permission from Morris *et al.* (1996) *Rapid Communications in Mass Spectrometry* 10: 889–896; © John Wiley & Sons Ltd.)

TOF-MS is particularly important is when multidimensional protein identification technology (MuDPIT) is used. Mixtures of many proteins (e.g., 100s) are digested and the peptides are loaded onto a strong cation exchange column. Sequential elution with mobile phase of increasing ionic strength fractionates the complex mixture of peptides for separate introduction on a reversed-phase column. IDA is performed on the reverse-phase peptide separations and in a typical run over several hours as many as 5000 individual MS/MS spectra can be recorded for automatic interpretation of amino acid sequence and database searching to identify the original proteins. Hundreds of proteins can be identified in such experiments.

In addition to product ion spectra, MS/MS systems are capable of modes to show precursors of a specific product ion or to screen for a constant neutral loss or presence of a product/precursor pair. These modes involve scanning both MS(1) and MS(2) or not scanning MS(2) so the duty cycle advantage is not as significant as in the product scan mode.

As mentioned above, MALDI oa-TOF-MS is feasible and the hybrid quadrupole TOF systems are available to provide the additional feature of MS/MS. One of the key advantages of MALDI TOF-MS is that it can provide a simple spectrum that reveals the molecular ions from a complex mixture. This is a common experiment in protein analysis, where a list of peptide masses that is referred to as a peptide mass fingerprint (PMF) is generated after digestion of an isolated whole protein with a proteolytic reagent (see Figure 4). The PMF is compared to databases of

predicted proteolysis products when a particular enzyme or reagent acts on a known protein sequence. MS/MS offers the ability to *de novo* sequence individual peptide ions and this can greatly increase the success of protein identification. The combination of rapid acquisition of the PMF followed by MS/MS if it is required is powerful.

MALDI produces predominantly singly charged molecular ions  $[M + H]^+$  as precursors for MS/MS. This is a disadvantage compared to ESI MS/MS where abundant doubly charged ions  $[M + 2H]^{2+}$  are typical. These tend to fragment in CID to give a more extensive sequence of ions than singly charged precursors.

### MALDI TOF/TOF Instruments

MS/MS spectra of MALDI ions can be obtained in a tandem TOF/TOF instrument. These instruments are essentially a linear TOF followed by a reflecting TOF MS system. The first TOF focuses MALDI ions into a collision cell. An ion gate in the first stage allows a specific precursor or all ions to pass into MS(1). Fragments of collisions (if desired) are passed to MS(2), which records the product ion spectrum. This mass spectrometer utilizes a high proportion of the ions generated in the source. It also subjects precursor ions to much higher collision energies than they would experience in a hybrid quadrupole TOF-MS system. As such there is a greater likelihood of seeing sufficient sequence ions even though the precursors are only singly charged. On the other hand, there is a tendency for CID spectra to be complicated by

fragmentation of the amino acid residues. This can make interpretation difficult. Despite these limitations the system can be automated to analyze very large numbers of samples in IDA mode and well suited for high throughput protein analysis laboratories.

### GC-MS, Electron Ionization, and Field Ionization

TOF-MS has long been promising to facilitate faster GC-MS. This is a growing area though not as fast as for the bioanalytical applications. The advantage of high-speed GC/MS is countered by the loss of sample capacity when using narrow-bore capillary GC.

Notwithstanding this, there are commercial GC-TOF-MS instruments and most of these are based on the oa-TOF-MS principle. The most successful of these offer the key advantages of better sensitivity in the full scan mode (oa-TOF-MS has a higher duty cycle than a scanning MS) and much greater mass accuracy than for other small mass spectrometers. Mass accuracy of 10 ppm or better is possible with oa-TOF-MS and this allows sufficient significant figures in ion mass to deduce the empirical formula. This could otherwise be achieved only with much larger and more expensive double-focusing magnetic mass spectrometer or Fourier transform mass spectrometers.

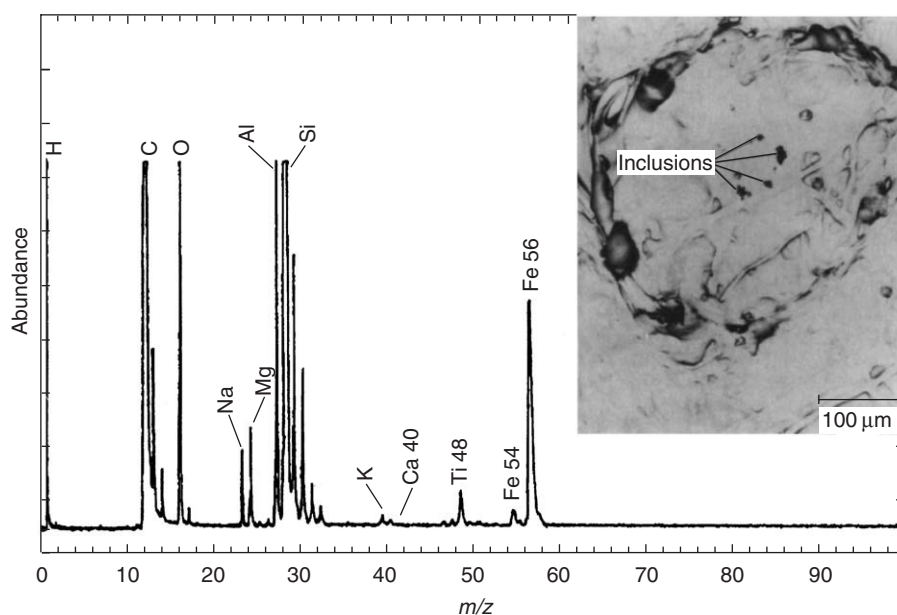
An interesting development in GC-TOF-MS is its combination with field ionization (FI). Many compounds of interest do not provide good molecular weight information by normal ionization methods,

whereas FI produces primarily molecular ions. The mass and full spectrum sensitivity of oa-TOF-MS combines well with FI as does the ability of TOF-MS to provide signal averaged spectra faithfully even when the ionization process is unsteady. RMS mass errors of less than 1.5 amu are attainable for small molecules ( $m/z < 400$ ).

Another GC-MS application of considerable potential is with comprehensive gas chromatography (GC  $\times$  GC). A short fast column follows a longer standard column and fractions eluting from the first column are trapped and periodically released into the second (fast) column. The second column has a stationary phase that is different from that of the first so that there is a tendency for the compounds that co-elute from the first column to separate in the second column. GC  $\times$  GC requires full scan spectra for separations less than 20–30 s and TOF-MS is ideal for this speed of analysis.

### Inductively Coupled Plasma Ionization TOF-MS

The coupling of TOF-MS to the inductively coupled plasma (ICP) ion source is challenging because of the large velocity spreads present in the intense ion beam. Notwithstanding this, the advantage of full spectrum sensitivity and the ability to perform full element analysis with transient inlet systems such as laser ablation and chromatography has provided a considerable driving force for ICP-TOF-MS. An additional advantage is a considerably improved precision of isotope ratios in the TOF systems compared

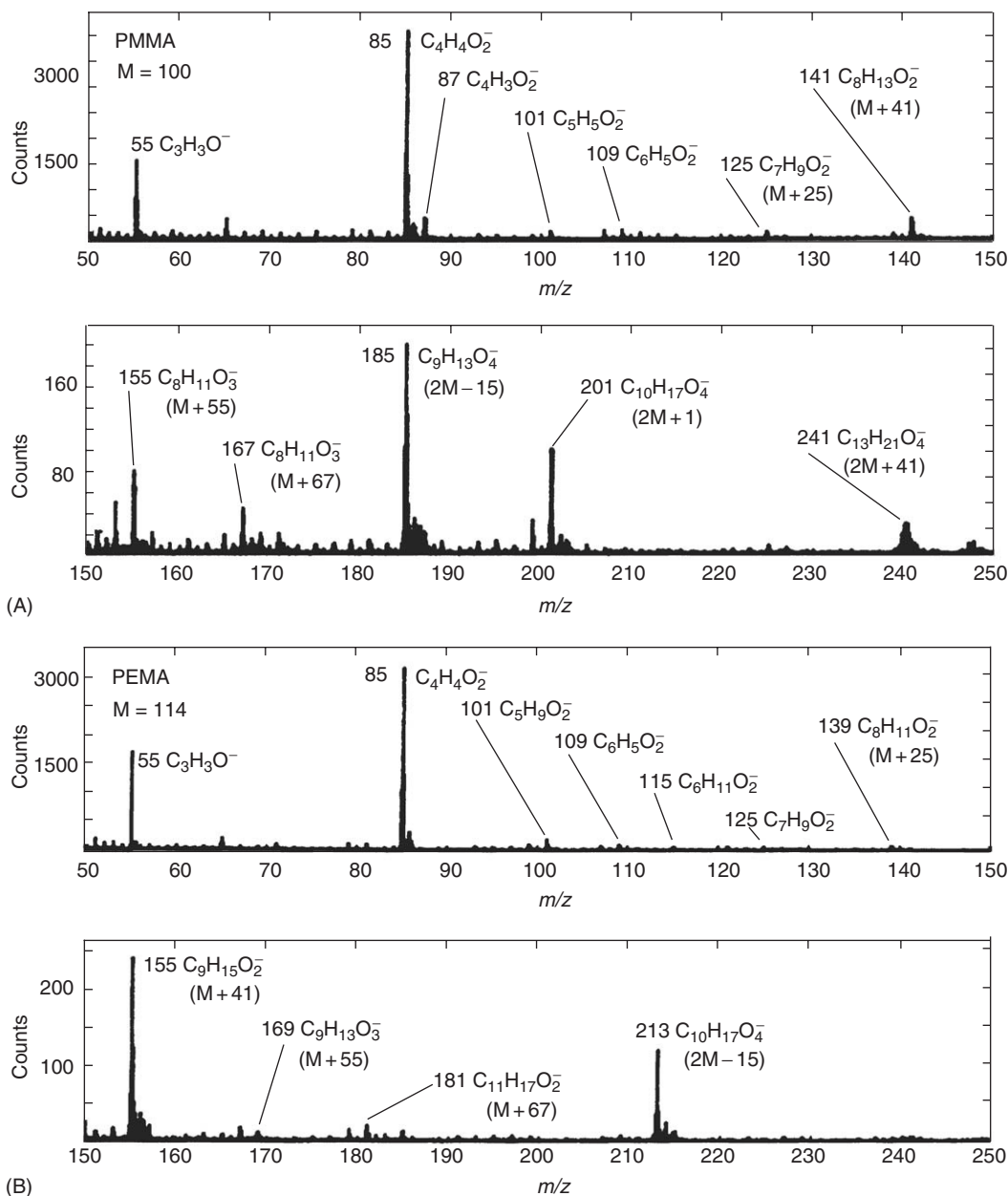


**Figure 8** A 0–100  $m/z$  mass spectrum of a particle inclusion in a polymer matrix obtained from an LIMA (Laser Ionization Mass Analyzer, Cambridge Mass Spectrometry, UK) instrument. From the LIMA spectrum it is suspected that these inclusions are paint-coated rust particles. Inset micrograph shows inclusions typical of those analyzed.

to quadrupole and magnetic mass spectrometers. The reason for this is that the latter instruments are slow scanning and the conditions in the source can change considerably as the scan proceeds from one isotope to another. TOF, on the other hand, samples ions from the plasma almost simultaneously. Two commercial ICP-TOF-MS instruments have recently emerged. One uses an axial (ion-bunching) geometry while the other is based on oa-TOF-MS. Both give resolving powers of  $\sim 2000$  FWHM and they deliver the promised speed of analysis.

### Surface Analysis

TOF-MS has played an important role in analyzing ions generated from surfaces for the purpose of characterizing the chemical composition of the surface or bulk analysis of small inclusions at the surface of a solid. This role extends back well before TOF became famous in bioanalytical chemistry. Ions are most typically generated from surfaces by irradiation with a laser beam or an ion beam.



**Figure 9** Negative-ion static SIMS spectra over a 50–250  $m/z$  range for (A) PMMA (monomer repeat mass ( $M$ ) of 100  $m/z$ ) and (B) PEMA (monomer repeat mass ( $M$ ) of 114). Polymers containing electronegative heteroatoms, such as oxygen in this case, often yield negative-ion spectra rich in information that can uniquely identify the polymer being analyzed.

**Laser ionization** The identification of unknown inclusions in a variety of matrices is very important in an industrial environment. One method used to analyze these unwanted small ( $\sim 1\text{--}50\text{ }\mu\text{m}$  diameter) inclusions is laser desorption/ionization mass analysis, also known as laser microprobe mass spectrometry and laser microprobe analysis. Trade names used are LIMA (laser ionization mass analyzer) and LAMMA (laser microprobe mass analysis). A variable-power laser is focused on to the inclusion. As discussed above, the desorption/ionization process is very sensitive to the power density of the laser at the sample surface. Firing the laser with an appropriate power setting generates ions containing information about the inclusion, which are then mass-analyzed using TOF-MS.

The spectra obtained from small ( $\sim 10\text{ }\mu\text{m}$  diameter) particles of unknown composition in a polymer matrix are shown in **Figure 8**. These spectra were obtained on a LIMA (Cambridge Mass Spectrometry, UK), which incorporates a reflecting TOF. It was found that all of the particles contained iron, titanium, aluminum, silicon, sodium, potassium, calcium, magnesium, carbon, and oxygen. The four most significant elements present were iron, titanium, carbon, and oxygen. Although this technique is semiquantitative at best, the observed distribution of these elements is consistent with that associated with a low-grade oxidized carbon steel and not from a stainless steel. The titanium level was abnormally high and was believed to be from paint, which contains high levels of  $\text{TiO}_2$ . Therefore, it appears likely that the unknown particles are from painted-over rust deposits that are easily dislodged and settle in the polymer product.

**Static secondary ion mass spectrometry (SIMS) of polymers** Static SIMS is a technique that obtains a mass spectrum of the upper 2 nm of a surface. Knowledge about surfaces is very important in studies of adhesion, corrosion, catalytic activity, biological compatibility, etc. By limiting the total primary ion dose from a primary ion beam to less than  $10^{13}$  ions per  $\text{cm}^2$ , each primary ion interacts with a sample that has not been damaged by previous primary ion impacts. In this way less than 1% of the total surface will be damaged by primary ion impacts.

Static SIMS is capable of obtaining very reproducible fingerprint spectra from bulk polymer surfaces. For example, static SIMS can easily distinguish between poly(methyl methacrylate) (PMMA) and the structurally similar poly(ethyl methacrylate) (PEMA) (**Figure 9**) based upon the unique negative ion fragmentation pattern obtained from each polymer. PEMA has a monomer repeat unit mass of 114 amu;

this is 14 amu heavier than that for PMMA, which is 100 amu. Thus, many of the fragments observed in the PMMA spectrum have corresponding fragments at masses 14 amu higher in the PEMA spectrum.

These spectra were obtained from a Poschenrieder TOF instrument equipped with a 30 keV Ga liquid-metal ion gun. The area analyzed was  $\sim 300 \times 300\text{ }\mu\text{m}$  using a total ion dose of  $\sim 10^{12}$  ions per  $\text{cm}^2$ .

It is interesting to note that from each pulse of the ion gun  $\sim 300$  gallium ions strike a spot less than 100 nm in diameter on the sample. Secondary ion yields for organic materials are typically between  $10^{-3}$  and  $10^{-4}$  secondary ions per primary ion impact. Thus, only a few fragment ions from the polymer surface are produced per pulse. The high transmission of the TOF allows for  $\sim 50\%$  of these secondary ions to be mass-analyzed. To obtain the final mass spectrum, hundreds of thousands of individual mass spectra are integrated. With TOF-MS, usable static SIMS mass spectra from areas as small as  $25 \times 25\text{ }\mu\text{m}$  can be obtained from polymer surfaces.

*See also:* **Atomic Mass Spectrometry:** Inductively Coupled Plasma; Laser Microprobe. **Gas Chromatography:** Mass Spectrometry. **Liquid Chromatography:** Liquid Chromatography–Mass Spectrometry. **Mass Spectrometry:** Ionization Methods Overview; Atmospheric Pressure Ionization Techniques; Electrospray; Liquid Secondary Ion Mass Spectrometry; Matrix-Assisted Laser Desorption/Ionization. **Surface Analysis:** Secondary Ion Mass Spectrometry of Polymers; Laser Ionization.

## Further Reading

- Benninghoven A, Hagenhoff B, and Niehuis (1993) Surface MS: probing real-world samples. *Analytical Chemistry* 65(14): 630A–640A.
- Cotter RJ (1997) *Time-of-Flight Mass Spectrometry: Instrumentation and Applications in Biological Research*. Washington: American Chemical Society.
- Dawson JHJ and Guilhaus M (1989) Orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Communications in Mass Spectrometry* 3: 155–159.
- Dodonov AF, Chernushevich IV, and Laiko VV (1994) Electrospray ionization on a reflecting time-of-flight mass spectrometer. In: Cotter RJ (ed.) *Time of Flight Mass Spectrometry*, pp. 108–123. Washington DC: American Chemical Society.
- Guilhaus M (2000) ‘Essential elements’ of time-of-flight mass spectrometry in combination with the inductively coupled plasma ion source. *Spectrochimica Acta* 55: 1511–1525.
- Guilhaus M (1995) Principles and instrumentation in time of flight mass spectrometry (Special Feature Tutorial Article). *Journal of Mass Spectrometry* 30: 1519–1532.

Guilhaus M, Selby D, and Mlynski V (2000) Orthogonal acceleration time-of-flight mass spectrometry. *Mass Spectrometry Reviews* 19: 65–107.

Mamyrin BA, Karataev VI, Shmikk DV, and Zagulin VA (1973) The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution. *Soviet Physics-JETP* 37(1): 45–48.

Poschenrieder WP and Oetjen GH (1972) New directional and energy focusing time of flight mass spectrometers for special tasks in vacuum and surface physics. *The Journal of Vacuum Science and Technology* 9(1): 212–215.

Wiley WC and McLaren IH (1955) Time-of-flight mass spectrometer with improved resolution. *Review of Scientific Instrumentation* 26: 1150–1157.

## Selected Ion Monitoring

**R Sleeman and J F Carter**, Mass Spec Analytical Ltd., Bristol, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The technique of selected ion monitoring (SIM) (not to be confused with secondary ion mass spectrometry, SIMS), also referred to as selected ion recording or multiple ion detection (MID), is the use of a mass spectrometer to record the intensities of a limited number of ions. This allows the instrument to dwell for a greater proportion of the analysis time on those mass-to-charge ratios of greatest significance in the mass spectrum of a compound of interest. The major application of the technique is in organic mass spectrometry (MS), where SIM data are typically acquired using instruments coupled to chromatographic inlet systems, allowing the qualitative or quantitative determination of organic compounds at extremely low concentrations. Inorganic mass spectrometry (inductively coupled plasma-MS and thermal ionization MS) and stable isotope ratio mass spectrometry (IRMS) also, typically, record only the intensity of selected ions. The term SIM is, however, generally only applied to organic MS.

In an early report (1966), the accelerating voltage of a magnetic sector instrument was rapidly switched by an 'accelerating voltage alternator' in order to focus the ions of interest onto the detector slit. Combined gas chromatography-mass spectrometry (GC-MS) was used to monitor the elution, from a packed column GC, of trimethylsilyl glucose and an internal standard comprising the closely eluting deuterated ( $d_7$ ) analog (**Figure 1**). The chromatographic peaks corresponding to the deuterated compound were observed at a slightly earlier retention time than the corresponding nondeuterated compound. Since the analyte and internal standard had very similar chemical and chromatographic properties, quantitation was achieved from a simple ratio of the two isotopomers. The fundamental principles

of the technique remain unchanged to the present day.

The identification of a compound by means of MS is normally achieved by recording the full mass spectrum of the compound of interest (full data collection). To identify a compound as unambiguously as possible, using a single analytical technique, this acquired spectrum may be compared either with the mass spectrum of an authentic standard or with a library spectrum. Authorities concur that 10 complete scans of the mass analyzer are typically required to characterize a compound. In practice, however, it is not always possible to acquire a full mass spectrum, either due to low sample concentration or matrix interference. During full data collection, the intensities of all mass-to-charge ratios, between selected mass limits, are recorded to produce a complete mass spectrum. The instrument scans through these masses spending a certain time recording the intensity of ions reaching the detector at each mass-to-charge ratio. The scanning cycle is then repeated, following a brief reset period. During this process, however, a proportion of the analysis time is used to monitor mass-to-charge regions in which no ions are present or ions that are of little or no diagnostic value or structural significance. In addition, the occurrence of certain ions is too ubiquitous to be of analytical value. The proportion of such ions may be considerable, and will depend on the analyte, the ionization process used, and other compounds present in the sample. The time available for scanning is limited; specifically only a restricted number of scans may be acquired during the elution of a chromatographic peak. This is especially true of modern chromatographic techniques, such as capillary GC (including 'fast GC') and capillary electrophoresis (CE), in which the chromatographic peakwidth may be less than a second. Using SIM, only a restricted number of ions are monitored; generally, those that are diagnostic for a particular compound or class of compounds. In this way, the proportion of time spent monitoring each of these ions may be substantially increased, leading to an improvement in sensitivity toward the compound



Guilhaus M, Selby D, and Mlynski V (2000) Orthogonal acceleration time-of-flight mass spectrometry. *Mass Spectrometry Reviews* 19: 65–107.

Mamyrin BA, Karataev VI, Shmikk DV, and Zagulin VA (1973) The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution. *Soviet Physics-JETP* 37(1): 45–48.

Poschenrieder WP and Oetjen GH (1972) New directional and energy focusing time of flight mass spectrometers for special tasks in vacuum and surface physics. *The Journal of Vacuum Science and Technology* 9(1): 212–215.

Wiley WC and McLaren IH (1955) Time-of-flight mass spectrometer with improved resolution. *Review of Scientific Instrumentation* 26: 1150–1157.

## Selected Ion Monitoring

**R Sleeman and J F Carter**, Mass Spec Analytical Ltd., Bristol, UK

© 2005, Elsevier Ltd. All Rights Reserved.

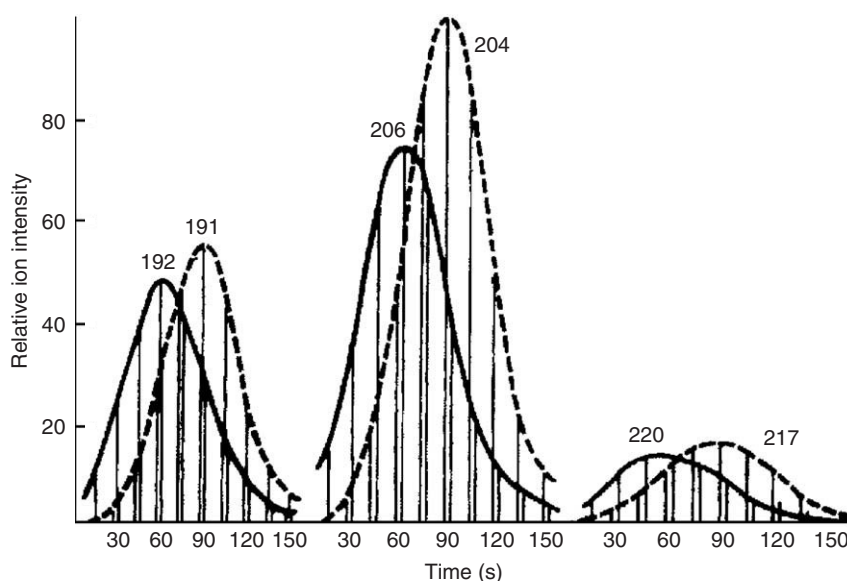
### Introduction

The technique of selected ion monitoring (SIM) (not to be confused with secondary ion mass spectrometry, SIMS), also referred to as selected ion recording or multiple ion detection (MID), is the use of a mass spectrometer to record the intensities of a limited number of ions. This allows the instrument to dwell for a greater proportion of the analysis time on those mass-to-charge ratios of greatest significance in the mass spectrum of a compound of interest. The major application of the technique is in organic mass spectrometry (MS), where SIM data are typically acquired using instruments coupled to chromatographic inlet systems, allowing the qualitative or quantitative determination of organic compounds at extremely low concentrations. Inorganic mass spectrometry (inductively coupled plasma-MS and thermal ionization MS) and stable isotope ratio mass spectrometry (IRMS) also, typically, record only the intensity of selected ions. The term SIM is, however, generally only applied to organic MS.

In an early report (1966), the accelerating voltage of a magnetic sector instrument was rapidly switched by an 'accelerating voltage alternator' in order to focus the ions of interest onto the detector slit. Combined gas chromatography-mass spectrometry (GC-MS) was used to monitor the elution, from a packed column GC, of trimethylsilyl glucose and an internal standard comprising the closely eluting deuterated ( $d_7$ ) analog (**Figure 1**). The chromatographic peaks corresponding to the deuterated compound were observed at a slightly earlier retention time than the corresponding nondeuterated compound. Since the analyte and internal standard had very similar chemical and chromatographic properties, quantitation was achieved from a simple ratio of the two isotopomers. The fundamental principles

of the technique remain unchanged to the present day.

The identification of a compound by means of MS is normally achieved by recording the full mass spectrum of the compound of interest (full data collection). To identify a compound as unambiguously as possible, using a single analytical technique, this acquired spectrum may be compared either with the mass spectrum of an authentic standard or with a library spectrum. Authorities concur that 10 complete scans of the mass analyzer are typically required to characterize a compound. In practice, however, it is not always possible to acquire a full mass spectrum, either due to low sample concentration or matrix interference. During full data collection, the intensities of all mass-to-charge ratios, between selected mass limits, are recorded to produce a complete mass spectrum. The instrument scans through these masses spending a certain time recording the intensity of ions reaching the detector at each mass-to-charge ratio. The scanning cycle is then repeated, following a brief reset period. During this process, however, a proportion of the analysis time is used to monitor mass-to-charge regions in which no ions are present or ions that are of little or no diagnostic value or structural significance. In addition, the occurrence of certain ions is too ubiquitous to be of analytical value. The proportion of such ions may be considerable, and will depend on the analyte, the ionization process used, and other compounds present in the sample. The time available for scanning is limited; specifically only a restricted number of scans may be acquired during the elution of a chromatographic peak. This is especially true of modern chromatographic techniques, such as capillary GC (including 'fast GC') and capillary electrophoresis (CE), in which the chromatographic peakwidth may be less than a second. Using SIM, only a restricted number of ions are monitored; generally, those that are diagnostic for a particular compound or class of compounds. In this way, the proportion of time spent monitoring each of these ions may be substantially increased, leading to an improvement in sensitivity toward the compound



**Figure 1** Changes in fragment ion intensities during the elution of  $d_0$ - and  $d_7$ -trimethylsilyl glucose from a packed column. (Reproduced from Sweeley *et al.*, 1966. *Analytical Chemistry*, copyright (1966) American Chemical Society.)

or compounds of interest. Estimates of this increase vary between 10- and 1000-fold depending on the analyte, matrix, ionization technique, and mass analyzer used. This enhancement in sensitivity, however, is necessarily achieved at the expense of unambiguous identification, since the entire mass spectrum is not recorded. If the ions to be monitored are selected prudently, however, the sensitivity gain may be significant without a concomitant diminution in the amount of analytically significant information derived. Comparisons of acquired data obtained using full data collection with authenticated spectra are typically based on the intensities of a limited number (normally eight) of the most significant ions in the mass spectrum. Library matching is normally based on the degree of agreement between the ratios of the ion intensities measured in the acquired spectrum and those in the library spectrum. If, therefore, the same ions are monitored in SIM mode, the same comparisons can be made with little reduction in analytically useful information. Indeed, the comparisons may be thought to be more reliable due to the improvement in signal-to-noise (S/N) ratio achieved by use of the SIM technique.

In general, the fewer ion peaks monitored, the greater the improvement in sensitivity. Taking the SIM technique to its extreme, sometimes only one ion is monitored; this is known as single ion monitoring (also abbreviated to SIM).

Most quantitative MS of organic compounds is conducted using a chromatographic inlet system and SIM. Compounds are separated from the matrix by the chromatographic system and enter the source of

the mass spectrometer after a characteristic retention time. The intensity of each characteristic ion may thus be monitored by SIM a number of times as the compound elutes. The use of SIM for quantification is more reliable than using full data collection because more time is spent monitoring the significant ions present in a chromatographic peak.

## Quantitative Mass Spectrometry with SIM

Quantification is effected by comparing the integrated peak areas or heights for the ions of interest with the responses obtained from a known aliquot of a standard compound. Either external or internal standards may be used. The response recorded for an external standard would be monitored in a preceding or subsequent analysis. Repeated measurements of the response for a known aliquot of an external standard may additionally be collated in order to monitor and correct for fluctuations in instrument response between analyses. More commonly internal standards are added at the beginning of the sample work up, and prior to any derivatization or wet-chemical process. Suitable compounds for use as internal standards include homologs or structurally related compounds possessing characteristic fragment ions, and which display similar chromatographic properties. These are distinguished from the analyte either by chromatographic separation or by differences in their mass spectra. For quantitative studies, the characteristic ion or ions

from the internal standards are monitored in a rapid sequence with those ions corresponding to the analytes. Modern instrumentation allows a number of SIM experiments to be performed at different time periods during the elution of a chromatogram, increasing the number of compounds that may be detected. Certain instrumentation also allows the simultaneous acquisition of full scan and SIM data.

A typical SIM experiment will monitor at least two ions (or ion transitions), resulting in the noncontinuous sampling of the chromatographic eluent. The quantified peak height obtained by SIM can be regarded as an estimate of the true height of a Gaussian curve based on a limited number of observations across the peak. To ensure that the error associated with this estimate is  $<1\%$ , at least 10 observations are required across the width of the peak, at half the observed height. Using SIM to monitor a pair of ions, at a typical scan speed of 20 ms per ion, 25 observations are generated each second facilitating quantification with a high degree of precision. In theory, error can also be reduced by using peak areas rather than heights for quantification. In practice, especially for small peaks, the errors introduced in attempting to delineate the baseline from a typical multicomponent chromatographic profile are usually much greater than those introduced in attempting to estimate a baseline level from which to measure the peak height.

Similar chemical behavior is a desirable attribute for an internal standard and, hence, isotopically labeled analogs of an analyte are frequently used as internal standards since they exhibit similar chemical and chromatographic behavior to the analyte. Deuterium-,  $^{13}\text{C}$ -, and  $^{15}\text{N}$ -labeled compounds are widely used as internal standards, the fragment ions monitored occurring at different mass-to-charge ratios (assuming that the charge-retaining fragments are isotopically labeled to some degree). The natural abundance of stable heavy isotopes present in the analyte make it desirable that the internal standard is at least three Daltons greater in mass than the analyte. When isotopically labeled analogs are employed, the relative response of the analyte with respect to its isotopically labeled counterpart is generally close to unity, since losses during sample preparation can be assumed to be similar for the two compounds.

The greatest accuracy attainable by MS is achieved by Isotope Dilution Mass Spectrometry using an 'exact signal matching' method. Like other SIM techniques, an isotopically labeled internal standard is added gravimetrically to the analyte, prior to extraction from its matrix. The quantity of internal standard is then adjusted, over a number of iterations, such that the signal obtained exactly matches that of

the analyte. This process is repeated with the internal standard and an authenticated sample of analyte (one of known purity and isotopic composition). Although time consuming, the instrumental variations, associated with all aspects of the method, are identical for both the analyte and standard, and this approach allows the quantitative determination of an analyte with a total uncertainty as low as 0.1–0.2%.

## Instrumentation and Techniques

The SIM technique is most commonly applied using quadrupole or magnetic sector instruments because of their ability to rapidly monitor different mass-to-charge ratios, even over significant mass ranges. Magnetic sector instruments achieve this by switching the accelerating voltage of the ion source and are limited in the range of masses that may be recorded. These instruments are, however, inherently more sensitive than quadrupole devices (typically a factor of 10) since the latter act as mass filters and incur transmission losses. In principle, other mass analyzers, including time-of-flight, ion trap, and Fourier-transform mass spectrometers can be adapted to acquire SIM data. In essence, these instruments take a 'sample' of the ions instantaneously present in the source for subsequent mass analysis and are potentially more sensitive than scanned instruments. Due to the fundamental nature of operation, however, little advantage is accrued by operating these instruments in the SIM mode.

### Sample Introduction Techniques

Most applications of SIM involve the coupling of the mass spectrometer to a means of chromatographic separation. This leads to a reduction in matrix effects and introduces an additional means of discriminating between compounds, namely the chromatographic retention time. Techniques such as GC, liquid chromatography (LC), CE, supercritical fluid chromatography, and thin-layer chromatography are all used in conjunction with mass spectrometers to form 'hyphenated' instruments such as GC-MS and LC-MS. The use of a chromatographic stage helps to resolve the compounds of interest from those that interfere with their determination, leading to a diminution in background chemical noise. Chromatographic conditions such as temperature gradient, column length, and choice of stationary and mobile phases are normally selected for optimum resolution and analysis time. The resolution of compounds in time also allows the possibility of monitoring different selected ions representative of various target compounds in different regions of the chromatogram.

The results from a chromatographic analysis by means of SIM are usually presented in the form of 'mass chromatogram', i.e., a plot of the intensity of a particular ion as a function of analysis time (Figure 2). Often this is plotted in conjunction with the total ion current (TIC), the sum of the intensities of all ions monitored over the same time period. The TIC is analogous to a chromatogram obtained from a 'universal' GC detector system, such as a flame ionization detector. Mass chromatograms may also be derived from full scan data. Whilst this technique is commonly employed for both the identification and quantitation of specific compounds, this process does not yield the improvement in sensitivity afforded by SIM.

The use of a chromatographic inlet technique leads to an additional means of characterizing and identifying a particular compound in a complex mixture, and hence to improved specificity. For a compound to be confidently identified by GC-MS-SIM, all the mass chromatograms representative of a particular

compound should maximize at the retention time corresponding to that compound, with the expected intensity ratios. Whilst it is commonplace to employ a chromatographic inlet system in conjunction with SIM, other methods are sometimes used. The elimination of a chromatographic stage can save considerable effort in sample preparation and derivatization, leading to a reduction in total analysis time, and can form the basis of rapid screening methods often known as 'fast and dirty methods'. Both flow injection MS and membrane inlet MS are frequently employed with SIM to monitor chemical processes and reactions in real time.

### Derivative Formation

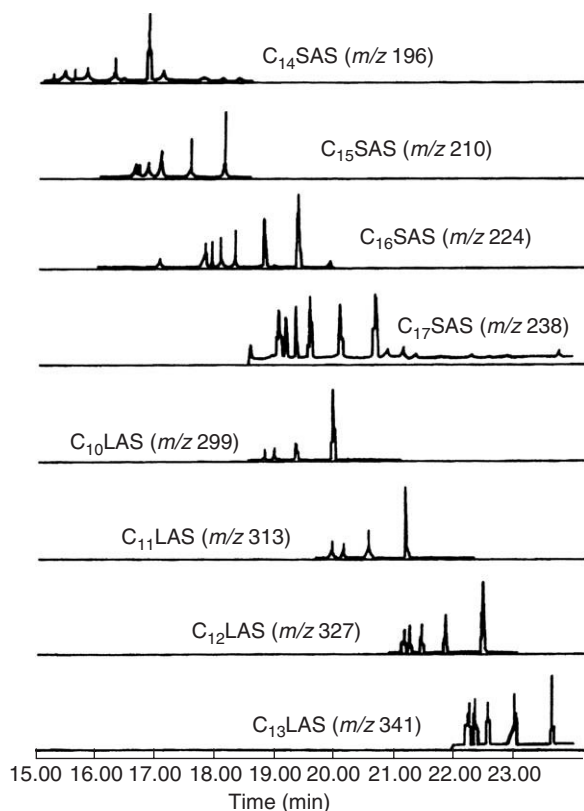
In cases where compounds are too polar or thermally unstable to be amenable to GC analysis, the situation may be improved by the formation of a suitable derivative with characteristics that render it more amenable to chromatographic analysis. In addition, derivative formation can enhance sensitivity and selectivity by altering the fragmentation mechanism of the molecule. A further beneficial effect of derivatization is that it normally results in a compound possessing characteristic ions at mass-to-charge ratios different from those of the original compound. Selection of an appropriate derivative can, therefore, lead to the choice of ions for monitoring by SIM in a region of the mass spectrum (usually at higher mass), which is less prone to chemical noise, and consequently to a further improvement in S/N ratio.

Derivatization may, in certain cases, be effected within the injection port of a GC-MS system. An ion pair/supercritical fluid extraction and injection port derivatization process has been presented as a method for the quantitative determination of aromatic and aliphatic sulfonated surfactants in sewage sludge by GC-MS with SIM (Figure 2).

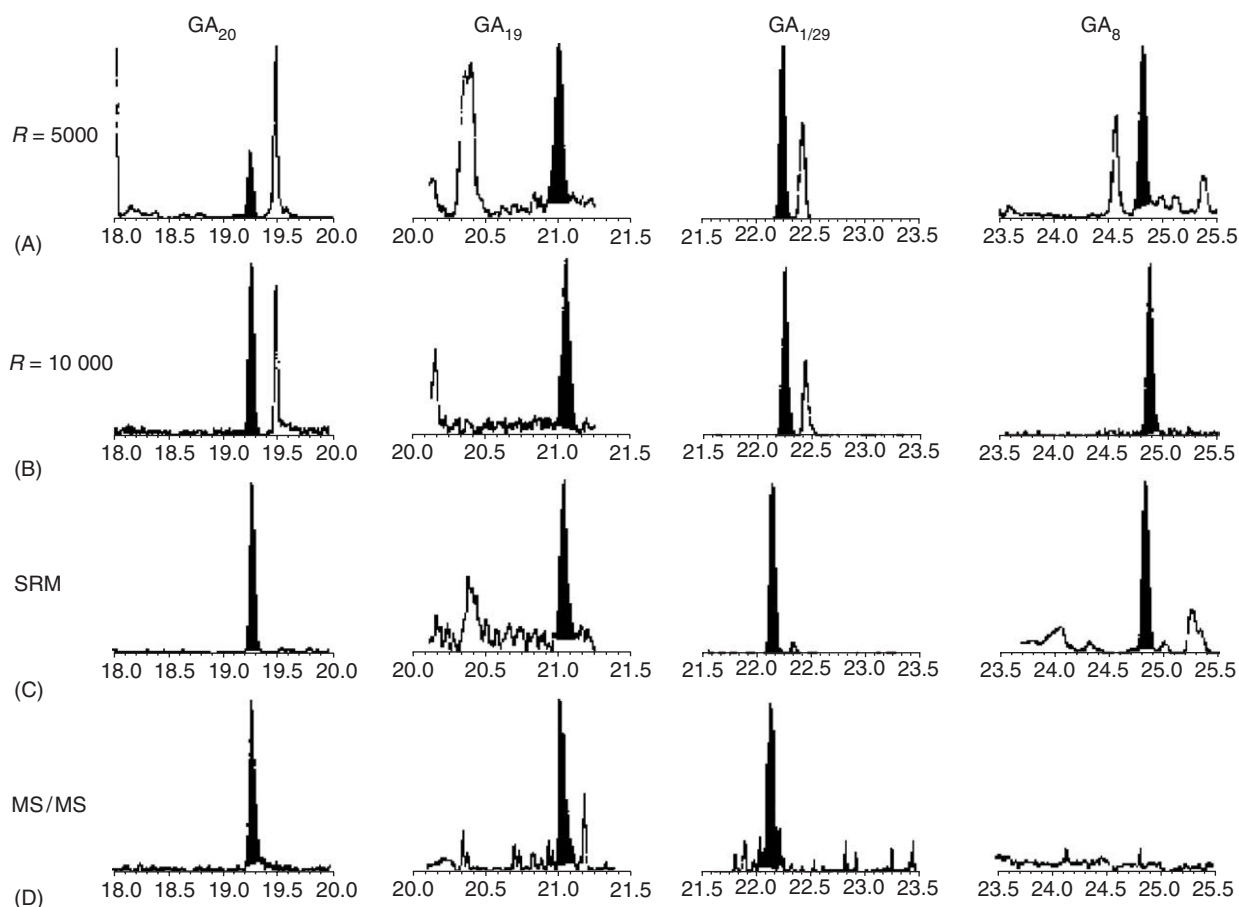
Careful selection of the derivatization technique may also be employed to enhance the detection limit of an analyte. The conversion of amphetamines to tetrafluorophthaloyl derivatives yields compounds possessing high electron capture cross-sections and the sensitivity toward such compounds, using negative ion chemical ionization (NICI) can be better than that achieved using positive ion formation by two orders of magnitude. Employing NICI in conjunction with SIM, the detection of amphetamine derivatives in quantities as small as 10 fg (30 amol) may be effected.

### High-Resolution Selected Ion Monitoring

An improvement in the specificity of the SIM technique may be achieved by the use of high-resolution



**Figure 2** Selected ion chromatograms for homologues of secondary alkanesulfonate surfactants (SAS) using  $[M - 138]^+$  ions at  $m/z$  196, 210, 224, and 238, and of linear alkylbenzenesulfonate surfactants (LAS) using  $[M - 55]^+$  ions at  $m/z$  299, 313, 327, and 341 in an unfractionated ion pair/SFE extract of sewage sludge. (Reproduced from Field *et al.*, 1992. *Analytical Chemistry*, copyright (1992) American Chemical Society.)



**Figure 3** SIM detection of gibberellins using: (A) 5000 resolution; (B) 10 000 resolution; (C) SRM using a sector instrument; and (D) SRM using a four-sector instrument. (Reproduced with permission from *Analytical Chemistry* 76: 1711–1716. Copyright (1995) the American Chemical Society.)

MS. When using an MS with unit resolution, the intensity recorded for a selected ion may be affected by contributions from the isotopomers of neighboring ions or nominally isobaric species. Any response measured may also contain a contribution from an adjacent mass peak rather than the intended peak. Increasing the resolution of the mass spectrometer reduces the number of background ions recorded, hence improving the detection limit. Typically, high resolution (usually 5000–10 000) SIM affords an increase in S/N of 2–10-fold. The certainty with which a compound can be identified by SIM is also greatly improved by high resolution, since the ions measured may correspond to only a single elemental composition. High-resolution sector instruments are now routinely employed in applications requiring the ultimate sensitivity and confidence attainable, often for analytes in extremely complex matrices. Typical applications include the determination of pesticides, polychlorinated biphenyls, and dioxins in environmental materials and the determination of controlled substances such as anabolic steroids in human and

animal sports. **Figure 3** illustrates the effect of increased mass resolution for the determination of endogenous growth regulators (gibberellins) in crude extracts of plant material analyzed by GC–MS SIM. Increasing the resolution from 5000 to 10 000 reduces the background ion current leading to an improvement in S/N ratio and hence detection limit.

### Selected Reaction Monitoring

An alternative means of enhancing both the specificity and detection limits attainable by SIM is afforded by use of tandem mass spectrometry (MS/MS). The MS/MS equivalent of SIM is termed selected reaction monitoring (SRM), sometimes incorrectly referred to as multiple reaction monitoring. In this technique, selected precursor-product gas phase ion reactions, specific to the analyte and an internal standard, are recorded in rapid succession.

Although many designs of MS/MS instruments exist, SRM is typically performed with sector instruments (in linked scan mode), triple quadrupoles, or



hybrid instruments, including four sector instruments. Like conventional SIM operation, each type of instrument offers certain advantages and limitations. Triple quadrupole instruments are the most widely adopted because of their ability to record ion transitions in a very rapid sequence. Hybrid instruments, incorporating high-resolution capability, offer the advantage of high-specificity precursor ion selection. SRM offers significant reduction in instrument noise (and hence increase in S/N ratio) even when compared to high-resolution SIM (Figure 3). In this example, SRM provided greater sensitivity and selectivity than high-resolution SIM. Use of a four-sector mass spectrometer led to a further increase in specificity (since both precursor and product ions are transmitted at high resolution) but a reduction in sensitivity.

### **Ionization Methods with Selected Ion Monitoring**

Any ionization technique may be used in combination with SIM, monitoring either positive or negative ions. Electron ionization (EI) is the most commonly used ionization technique and employs energetic electrons that generally induce a high degree of fragmentation in the analyte. This typically generates a range of ions that may be monitored in SIM. The use of 'softer' ionization techniques, such as chemical ionization (CI), often leads to a higher proportion of the ion current being carried by the higher-mass ions, typically the protonated molecule. The monitoring of these species affords improved sensitivity and less interference from other components. Since molecular mass is one of the most characteristic features of a molecule, selectivity is also often enhanced. In particular, NICI has been extensively used with SIM in quantitative organic MS. The negatively charged molecular ion is often more stable than its cationic radical counterpart, and may fragment by a different pathway, yielding supplementary information to that derived from the positive EI or CI mass spectrum. Negative ions formed by either electron capture or ion-molecule reactions in the ion source may be monitored. Nucleophilic addition reactions in the source with anions of low proton affinity, such as  $\text{Cl}^-$ , may lead to the formation of stable adducts. These  $[\text{M} + \text{Cl}]^-$  complexes have been used, for example, to search for the presence of traces of certain explosives in security surveillance applications. The formation of chloride addition complexes, in an atmospheric pressure chemical ionization (APCI) source, leads both to improved detection limits for the explosives in question and to enhanced selectivity, since two adduct ions may be monitored for each explosive, corresponding to complexes formed with the  $^{35}\text{Cl}$  and  $^{37}\text{Cl}$  isotopes. The observation of ions at

masses corresponding to adducts of both isotopes, in the expected relative abundance, gives a far more reliable indication of the presence or absence of an analyte than monitoring a single ion.

Most LC-MS sample introduction techniques, such as thermospray, electrospray, and APCI, do not employ an electron beam to induce ionization, and lead to the formation of predominantly molecular ion species. This leads to the formation of higher mass fragments that may be monitored by SIM or SRM. Such techniques are frequently used in biomedical applications, and find particular application in the quantitative analysis of labile or polar compounds such as biomolecules from complex biological matrices.

In some circumstances, a depletion in the amounts of a particular CI reagent ion has been monitored by SIM as compounds elute from a chromatographic inlet. Although this is neither a particularly sensitive nor selective technique, semiquantitative data may be obtained without calibration for a number of compounds. This technique has been referred to as 'reagent ion monitoring'. A development of this technique is the selected ion flow tube-mass spectrometer, which employs a quadrupole mass spectrometer to introduce selected reactant ions into a reaction flow tube. Volatile organic compounds (VOCs) introduced into this region react with the selected reagent ions and the products are mass analyzed by a subsequent quadrupole mass analyzer. This technique has already found wide application in the quantitative analysis of parts per billion amounts of VOCs present in breath gases or evolved from environmental samples.

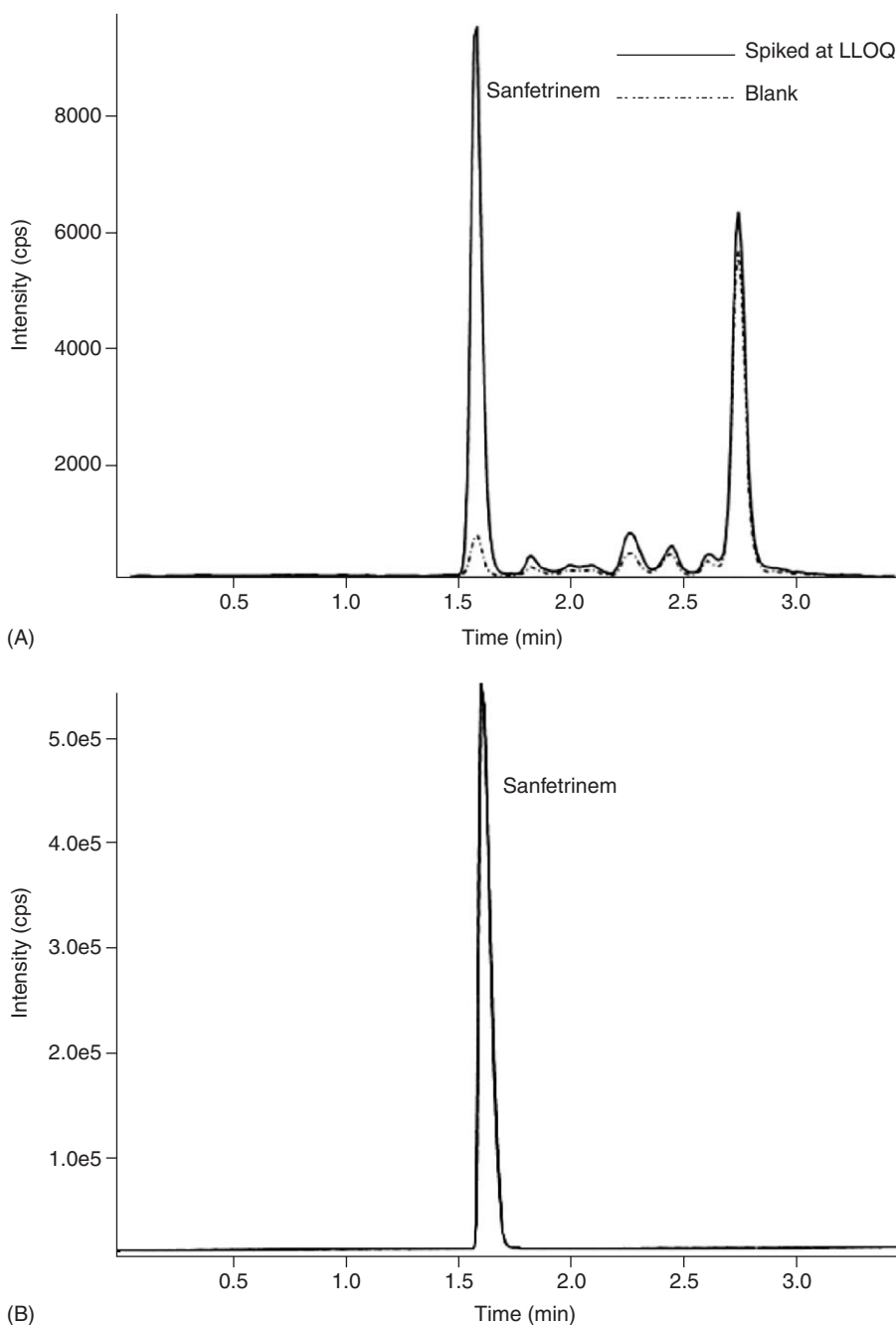
### **Applications**

The SIM technique finds use primarily in quantitative studies of trace analytes in complex mixtures. The technique has been utilized in a diverse range of applications including forensic science, toxicology, environmental monitoring, biochemistry, clinical chemistry, and geochemistry. In applications such as these, the analyte is often present in a complex matrix comprising a great number of compounds that may mask its presence or otherwise inhibit its detection.

The presence or absence of specific trace components in such complex matrices is often difficult to establish due to interference from a high abundance of compounds that co-elute from the chromatographic system used. This makes the acquisition of 'clean' mass spectra difficult. Some compounds may simply be present in insufficient quantities to allow a high-quality spectrum to be obtained. Hence, it is

desirable to increase the sensitivity of the mass spectrometer toward the analytes sought. SIM data may in such cases provide a qualitative solution to the question of the presence or absence of a particular analyte. It is not usually possible simply to increase the amount of analyte entering the source of the mass spectrometer by injecting a more concentrated solution of the extract into the GC-MS. In mass spectrometric techniques employing a chromatographic

inlet, the quantity of a particular component, which may be introduced into the source, is limited by its concentration relative to the most abundant component present. In the search for trace components, therefore, the chromatographic stage can be the limiting factor. Minor components may thus become masked by other substances in the mixture, and in adverse cases may not be discerned from the background noise. The SIM technique is commonly used



**Figure 4** SIM detection of sanfetrinem in plasma: (A) a blank sample compared to spiked plasma; (B) a representative chromatography of a real patient sample. (Reprinted with permission from Nardi *et al.*, 2001. *Journal of Chromatography B* 767: 193–201; © from Elsevier.)

to establish the presence of trace components in such mixtures. SIM and SRM are recognized as the most commonly used biomedical mass spectrometric techniques, specifically in the fields of drug development and screening. **Figure 4** shows a typical application of LC–MS/MS SIM to the determination of sanfetrinem, a broad spectrum antibacterial agent, in human plasma. Because of the high specificity of the technique, minimal sample preparation and chromatography are required. The sample is simply deproteinized with acetonitrile, diluted with water, and injected onto a short chromatography column (5 cm – C<sub>18</sub>), the analyte eluting in less than 2 min. The extreme simplicity of the chromatogram obtained reflects the high selectivity of the technique, even when the analyte is present in a highly complex biological matrix. Using this simple and rapid technique, this antibiotic could be quantified in the range 10 ng–5 µg ml<sup>−1</sup> in human plasma.

**See also:** Derivatization of Analytes. Gas Chromatography: Mass Spectrometry. Liquid Chromatography: Liquid

Chromatography–Mass Spectrometry. Mass Spectrometry: Atmospheric Pressure Ionization Techniques.

## Further Reading

- Chapman JR (1993) *Practical Organic Mass Spectrometry*, 2nd edn. Chichester: Wiley.
- Sargent M, Harrington M, and Hart R (eds.) (2002) *Guidelines for Achieving High Accuracy in Isotope Dilution Mass Spectrometry (IDMS)*. Cambridge: The Royal Society of Chemistry.
- Busch KL (1997) Selected ion monitoring. *Spectroscopy* 12(9): 30–35.
- Petite BR (1986) Expected errors of quantification from selected ion monitoring assays using capillary columns. *Biomedical and Environmental Mass Spectrometry* 13: 473–475.
- Freeman CG and McEwan MJ (2002) Rapid analysis of trace gases in complex mixtures using selected ion flow tube-mass spectrometry. *Australian Journal of Chemistry* 55: 491–494.

## Multidimensional

**R M Twyman**, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Multidimensional mass spectrometry (MS) involves the use of two or more mass analysis procedures one after the other in order to facilitate the selective fragmentation of specific ions in a complex mixture. Many different types and configurations of instruments can be used, some of which separate the consecutive mass analysis events spatially, while others carry out two or more analyses in the same compartment but separate them temporally. Tandem mass spectrometry (MS/MS or MS<sup>2</sup>) often involves two physically distinct mass analyzers separated by a collision cell to induce the fragmentation of precursor (primary or parent) ions. Higher-order analysis (MS<sup>3</sup> and above) involves multiple rounds of collision and ion extraction, and is usually carried out in ion trap or ion cyclotron resonance instruments, which have ion storage and selection capability. Multidimensional MS is particularly useful where the analyte is very complex, where ionization produces relatively few structurally diagnostic ions, or where useful ions are obscured or suppressed by others

introduced either in the analyte itself or in the sample matrix. Depending on the scan mode, MS/MS can be used to characterize individual ions, or to search for specific ion fragments that are diagnostic for certain classes of chemical compounds. This article discusses the instrumentation and scan modes used in multidimensional MS and briefly outlines some of its major applications.

## Instrumentation

There are two major groups of instruments used in tandem and multidimensional MS. The first group encompasses those instruments in which two or more mass analyzers are assembled in series, such that parent or precursor ions emerging from the ion source are selected in the first analyzer, fragmented in a collision cell, and the product (secondary or daughter) ions are then characterized in the second and subsequent analyzers. The second group encompasses instruments capable of ion storage, such that precursor ions are retained, fragmented, and the product ions analyzed in the same chamber. Selected product ions can then be retained and fragmented, and the second-generation product ions can be analyzed. This procedure can be repeated as many times as

to establish the presence of trace components in such mixtures. SIM and SRM are recognized as the most commonly used biomedical mass spectrometric techniques, specifically in the fields of drug development and screening. **Figure 4** shows a typical application of LC–MS/MS SIM to the determination of sanfetrinem, a broad spectrum antibacterial agent, in human plasma. Because of the high specificity of the technique, minimal sample preparation and chromatography are required. The sample is simply deproteinized with acetonitrile, diluted with water, and injected onto a short chromatography column (5 cm – C<sub>18</sub>), the analyte eluting in less than 2 min. The extreme simplicity of the chromatogram obtained reflects the high selectivity of the technique, even when the analyte is present in a highly complex biological matrix. Using this simple and rapid technique, this antibiotic could be quantified in the range 10 ng–5 µg ml<sup>−1</sup> in human plasma.

**See also:** Derivatization of Analytes. Gas Chromatography: Mass Spectrometry. Liquid Chromatography: Liquid

Chromatography–Mass Spectrometry. Mass Spectrometry: Atmospheric Pressure Ionization Techniques.

## Further Reading

- Chapman JR (1993) *Practical Organic Mass Spectrometry*, 2nd edn. Chichester: Wiley.
- Sargent M, Harrington M, and Hart R (eds.) (2002) *Guidelines for Achieving High Accuracy in Isotope Dilution Mass Spectrometry (IDMS)*. Cambridge: The Royal Society of Chemistry.
- Busch KL (1997) Selected ion monitoring. *Spectroscopy* 12(9): 30–35.
- Petite BR (1986) Expected errors of quantification from selected ion monitoring assays using capillary columns. *Biomedical and Environmental Mass Spectrometry* 13: 473–475.
- Freeman CG and McEwan MJ (2002) Rapid analysis of trace gases in complex mixtures using selected ion flow tube-mass spectrometry. *Australian Journal of Chemistry* 55: 491–494.

## Multidimensional

**R M Twyman**, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Multidimensional mass spectrometry (MS) involves the use of two or more mass analysis procedures one after the other in order to facilitate the selective fragmentation of specific ions in a complex mixture. Many different types and configurations of instruments can be used, some of which separate the consecutive mass analysis events spatially, while others carry out two or more analyses in the same compartment but separate them temporally. Tandem mass spectrometry (MS/MS or MS<sup>2</sup>) often involves two physically distinct mass analyzers separated by a collision cell to induce the fragmentation of precursor (primary or parent) ions. Higher-order analysis (MS<sup>3</sup> and above) involves multiple rounds of collision and ion extraction, and is usually carried out in ion trap or ion cyclotron resonance instruments, which have ion storage and selection capability. Multidimensional MS is particularly useful where the analyte is very complex, where ionization produces relatively few structurally diagnostic ions, or where useful ions are obscured or suppressed by others

introduced either in the analyte itself or in the sample matrix. Depending on the scan mode, MS/MS can be used to characterize individual ions, or to search for specific ion fragments that are diagnostic for certain classes of chemical compounds. This article discusses the instrumentation and scan modes used in multidimensional MS and briefly outlines some of its major applications.

## Instrumentation

There are two major groups of instruments used in tandem and multidimensional MS. The first group encompasses those instruments in which two or more mass analyzers are assembled in series, such that parent or precursor ions emerging from the ion source are selected in the first analyzer, fragmented in a collision cell, and the product (secondary or daughter) ions are then characterized in the second and subsequent analyzers. The second group encompasses instruments capable of ion storage, such that precursor ions are retained, fragmented, and the product ions analyzed in the same chamber. Selected product ions can then be retained and fragmented, and the second-generation product ions can be analyzed. This procedure can be repeated as many times as

necessary to gain useful structural information from subsequent generations of product ions.

### Multisector Instruments

The first type of instruments developed for MS/MS were double-sector mass spectrometers consisting of magnetic (B) and electrostatic (E) sectors arranged in series. The two sectors can be arranged in either configuration, i.e., BE or EB, although the BE configuration (reversed geometry) has become the most widespread. The principle of MS/MS analysis in such instruments is that precursor ions fragmenting in the field-free region downstream of the magnetic sector will yield product ions with approximately the same velocity as the precursor. Since the electrostatic analyzer separates ions according to their kinetic energies, and kinetic energy is a product of mass and velocity, product ions with identical velocities will be separated according to mass. As well as double-sector instruments, a number of manufacturers produce multisector instruments with three, four, or five magnetic and electrostatic analyzers in series. Two double-focusing instruments have been combined to generate a four-sector instrument with the configuration BEEB, incorporating multiple collision cells to facilitate both high- and low-energy collisions (Figure 1). The Autospect-T, produced by Micromass Inc., is an example of a commercially produced mass spectrometer which has five sectors in the configuration EBEBE.

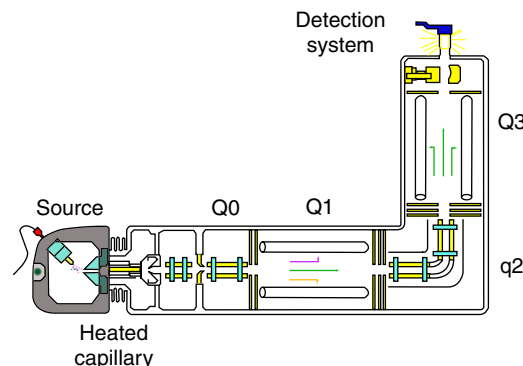
### Triple Quadrupole Instruments

The most straightforward way to carry out MS/MS analysis is to arrange two mass analyzers in tandem, placing an ion source at one end, a detector at the other, and a collision cell in the middle. The most widely used MS/MS instrument in this category is the triple quadrupole, in which three quadrupole analyzers are arranged in series (Figure 2). The first and

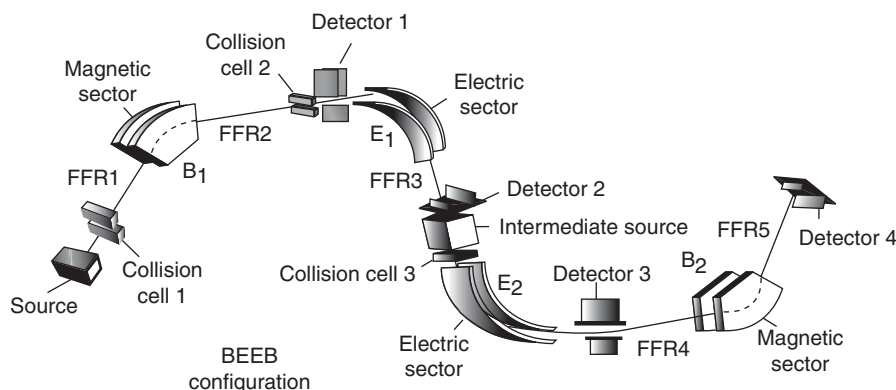
third quadrupoles are used as analyzers, and may either scan the ion stream or select ions of a fixed  $m/z$  ratio, while the second quadrupole is run in radio frequency (rf) mode and operates as the collision cell. The conventional representation for this geometry is QqQ, where Q represents a quadrupole operating as a mass filter and q represents a quadrupole operating in rf mode and letting ions of any  $m/z$  value through. The second quadrupole also focuses the ion stream, preventing the loss of ions scattered during collision. The number of collisions occurring in a triple-quadrupole mass spectrometer is generally much larger than that observed in a sector instrument, resulting in a more efficient conversion of precursor ions into product ions. Other advantages of the triple quadrupole include economy, ease of use, and convenient switching between different operational modes. However, the instruments have a restricted mass range, low sensitivity, and a limited resolution.

### Hybrid Instruments

Hybrid instruments combining magnetic sector and quadrupole analyzers have been developed, which



**Figure 2** Design of a triple-quadrupole mass spectrometer, where q2 operates as the collision cell.



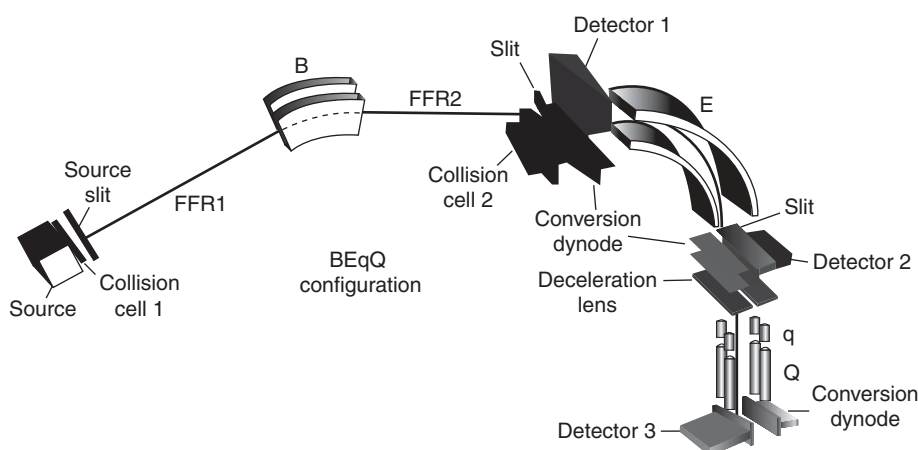
**Figure 1** Design of a four-sector MS/MS instrument with the configuration BEEB. FFR = field-free region.



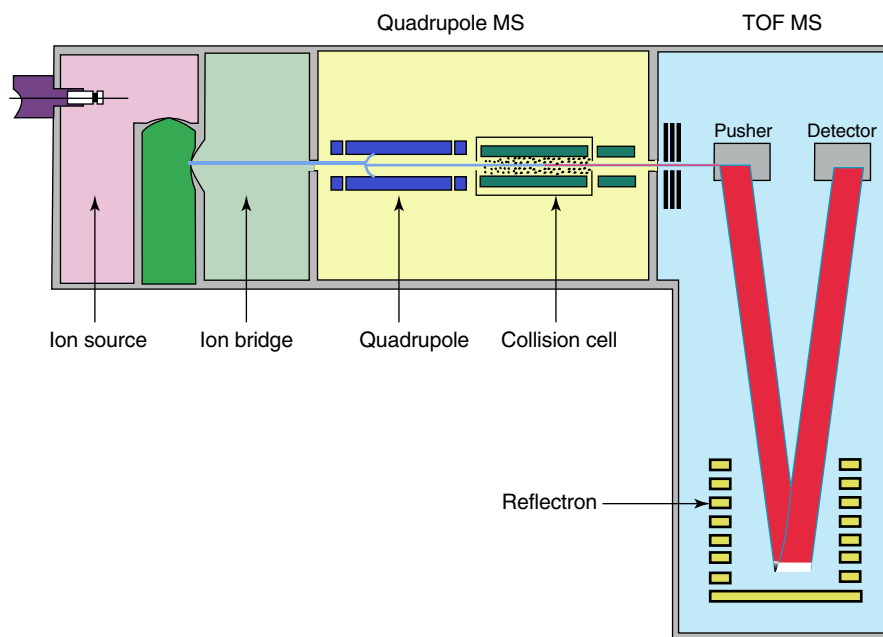
are simpler to operate than the multisector instruments described above. The configuration BEqQ, which uses the magnetic and electrostatic sectors to separate the precursor ions and a quadrupole collision chamber and mass analyzer to generate and separate the product ions, is shown in Figure 3. This instrument allows high-resolution selection of parent ions, facilitates both high- and low-energy collisions, and achieves unit resolution of product ions. Other configurations, such as EBqQ, have also been used.

Both quadrupole and sector analyzers have also been combined with time-of-flight (TOF) analyzers, to give BE-TOF and QqTOF configurations. In the

former case, the BE analyzer is joined to an orthogonal TOF analyzer, such that product ions with the same velocity as the precursor ion are accelerated orthogonally into a field-free tube, thus traveling at a velocity proportional to their  $m/z$  ratio. The hybrid QqTOF instrument operates on similar principles, and is renowned for its high resolution and sensitivity across the full mass range (Figure 4). Such instruments can be more than 100 times more sensitive than triple quadrupoles for MS/MS experiments and have an  $m/z$  upper limit in excess of 20 000 mass units. One disadvantage is that only low-energy collisions are possible, resulting in incomplete fragmentation.



**Figure 3** Design of a hybrid sector-quadrupole instrument with the configuration BEqQ. FFR = field-free region.



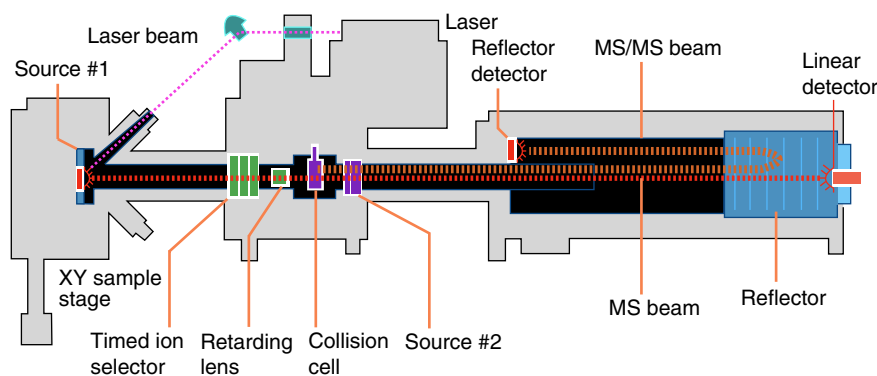
**Figure 4** Design of a hybrid Q-TOF instrument.

### TOF-TOF (Tandem TOF)

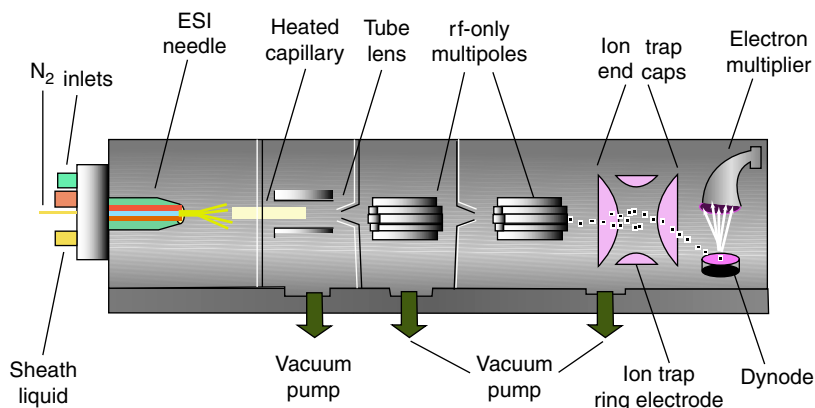
To overcome the limitations of QqTOF and other instruments, mass spectrometers have been designed with two orthogonally arranged TOF analyzers separated by a collision cell. The high-energy collisions that are possible in such instruments are particularly useful in proteomics applications, where full fragmentation can reveal differences between the amino acids leucine and isoleucine, which have identical molecular masses. In order to focus all the product ions without changing the reflectron voltage, conventional TOF/TOF instruments initially used a low-power source to generate the primary ions, and then re-accelerated the product ions after fragmentation, using either a second source or by lifting the voltage of the collision cell (Figure 5). Contemporary instruments have two curved reflection analyzers separated by a collision cell, which accommodates the full range of product kinetic energies. Collision energies can be as high as 20 keV for an instrument with a 20 kV ion acceleration voltage, therefore facilitating complete fragmentation of the precursor ions.

### Ion-Storage Devices

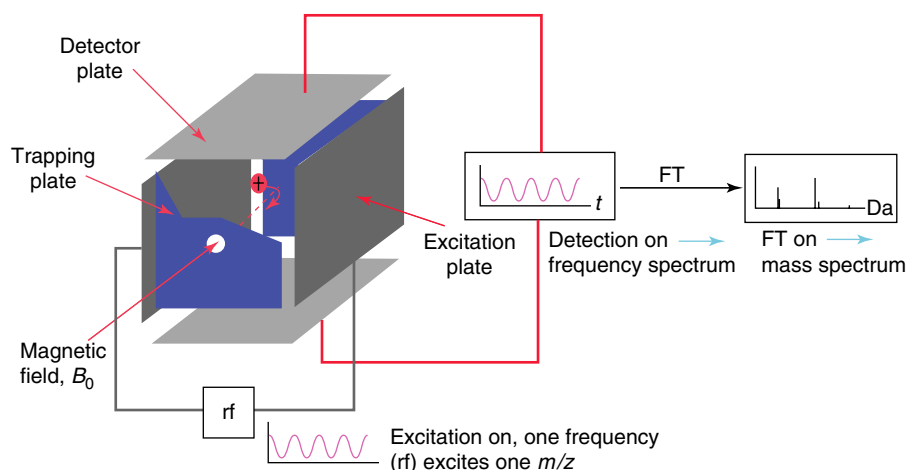
Two types of ion storage instrument have been used for MS<sup>n</sup> applications: the quadrupole ion-trap (QIT) (Figure 6) and the Fourier transform ion cyclotron resonance (FT-ICR) analyzer (Figure 7). The ion trap consists of a chamber surrounded by a ring electrode and two end-cap electrodes. The voltage applied to the ring electrode determines which ions remain in the trap. Ions above the threshold  $m/z$  ratio remain in the trap while others are ejected through small holes in the distal end-cap electrode. A mass spectrum of intact ions can be obtained by gradually increasing the voltage in the ring electrode so that ions of progressively increasing  $m/z$  ratios are ejected over time. For MS<sup>n</sup> analysis, the trapped precursor ions can be fragmented by injecting a stream of helium gas, and the resulting product ions can be ejected by ramping the voltage of the ring electrode to generate a spectrum. Multiple rounds of analysis can be carried out because one of the fragment ions from the first experiment can be retained in the trap and subjected to further collisions.



**Figure 5** Design and layout of the tandem time-of-flight mass spectrometer, coupled to a MALDI source.



**Figure 6** Design and layout of the quadrupole-ion trap mass spectrometer, coupled to an ESI source.



**Figure 7** Design principles of the Fourier transform ion cyclotron resonance mass spectrometer.

The FT-ICR analyzer is the most complex and difficult to operate, but has by far the highest resolution, mass accuracy, and sensitivity. The operating principle is that ions in a magnetic field will orbit at a frequency that is related to the ion's mass ( $m$ ), charge ( $z$ ), and the strength of the magnetic field ( $B$ ). This is called the cyclotron frequency ( $f_c$ ). The relationship can be described by the following equation:

$$m/z = B/2\pi f_c$$

Therefore, all ions with the same  $m/z$  value will orbit with the same cyclotron frequency in a uniform magnetic field, and this collection of ions is known as an ion packet.

Orbiting ions of a particular  $m/z$  value are then excited using an applied rf field, which causes the cyclotron radius to expand. If the frequency of the applied field is the same as the cyclotron frequency of the ions, the ions absorb energy thus increasing their velocity (and the orbital radius) but keeping a constant cyclotron frequency. As the selected ions cycle between the two electrodes, electrons are attracted first to one plate and then the other, with the same frequency as the cycling ions (i.e., in resonance with the cyclotron frequency). This movement of electrons is detected as an image current on a detector. The image current is then converted, by Fourier transformation, into a series of component frequencies and amplitudes of the individual ions. Finally, the cyclotron frequency values are converted into  $m/z$  values to produce the mass spectrum. Various forms of collisional activation can be used to fragment the precursor ions in an FT-ICR instrument (see below) but in simple terms the principle of  $MS^n$  analysis is the same as described above for the ion-trap analyzer. Selected precursor ions are fragmented and gradually increasing frequencies are used to induce resonance in product ions of progressively increasing  $m/z$  ratios,

producing current images that are transformed into mass spectra. At each stage of the analysis, product ions can be retained and subjected to further rounds of fragmentation and excitation. The major difference between ion traps and FT-ICR instruments is that the latter are nondestructive. Ions are not expelled to a detector during analysis, but are cycled within the device.

### The Collision Cell

Fragmentation occurs spontaneously in many ionization sources, but the product ions generated in this way are neither diverse nor abundant enough to allow the complete characterization of the precursor ion. Fragmentation can also be induced, e.g., by in-source collision in electrospray sources or by post-source decay in matrix-assisted laser desorption/ionization (MALDI) sources, but fragmentation prior to the separation of precursor ions can make the analysis of complex analytes very difficult. The fragmentation of selected precursor ions is achieved by placing a collision cell between the first and second mass analyzers in a tandem mass spectrometer. Collision occurs when the ion stream passes through a jet of inert gas, such as argon or xenon. The fragmentation of precursor ions by collision with neutrals in this manner is called collision-induced dissociation (CID). (The term collision-activated dissociation is also used.) In  $MS/MS$  instruments with tandem analyzers, the collision cell is a discrete chamber between the two analyzers. It may be a simple chamber filled with gas, in which case the ions are scattered as well as fragmented, but more usually the collision cell is contained within a quadrupole running in rf mode, which allows the ion stream to be focused. In  $MS^n$  instruments such as the ion trap and the ICR analyzers, there is no separate collision

cell. Instead, gas is introduced into the ion-trap chamber using an operator-controlled valve.

In an FT-ICR instrument, fragmentation may be achieved by colliding ions with neutrals (CID) but various other strategies are available, such as collision with surfaces (surface-induced dissociation) or bombardment with ultraviolet or infrared radiation from a laser (ultraviolet photodissociation and multiphoton infrared photodissociation). Dissociation may also be achieved by the absorption of blackbody radiation produced by a heated vacuum chamber walls (blackbody infrared dissociation). An advantage of these radiation-induced fragmentation methods is that gas pulses are no longer required. Sustained off-resonance irradiation is the preferred, radiation-based method for FT-ICR MS<sup>n</sup> because it is the simplest to implement and tune. Very low energy and multiple excitation collisional activation techniques are also available.

## MS/MS and MS<sup>n</sup> Scan Modes

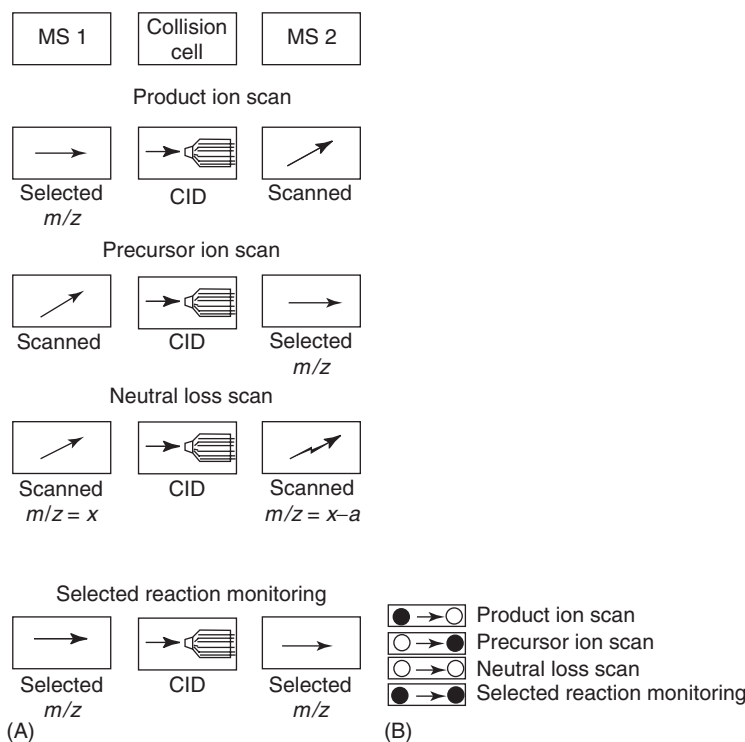
Four scan modes are available in MS/MS analysis. The three that are used most frequently are product ion scanning, precursor ion scanning, and neutral loss scanning, all of which generate mass spectra across a given mass range. The remaining mode is used for selected reaction monitoring and results in

the production of a single type of fragment ion. All of these scan modes are possible in tandem instruments but only product ion scanning is possible in trapping instruments. Kondrat and colleagues proposed a simple symbolic nomenclature to represent these scan modes, in which a filled circle represents mass analysis at a fixed  $m/z$  ratio and an open circle refers to a mass analyzer scanning the ion stream. The nomenclature is shown in Figure 8.

### Product Ion Scan Mode (MS/MS)

In this mode, also known as daughter ion scanning, the mode scheme is select–dissociate–scan ( $\bullet \rightarrow \circ$  in Kondrat symbols). For sequential instruments where mass analysis events are separated in space, the first analyzer is fixed to select intact precursor ions of a particular  $m/z$  ratio from the source, these are directed into the collision cell, which fragments the chosen ion and focuses the products into the second analyzer. The second analyzer scans the incoming stream of product ions over an appropriate mass range and produces a mass spectrum. Product ion scanning is one of the most widely used scan modes in QqQ, Q-TOF, and tandem TOF instruments.

In sector instruments with reversed geometry, product ion scanning as depicted above is known as MILES (mass-analyzed ion kinetic energy spectrometry). The magnet is used to select the precursor



**Figure 8** (A) Schematic representation of the four scan modes used in tandem mass spectrometry. (B) Kondrat symbolism, where a filled circle indicates a fixed mass analyzer and an open circle indicates a scanning analyzer.

ion, which is fragmented in the field-free region between the magnetic and electrostatic sectors, and the product ions are then separated according to their kinetic energies by the electrostatic analyzer. However, because the precursor and product ions have the same velocity, scanning both sectors at the same time while maintaining a constant ratio between  $B$  (the magnetic field strength) and  $E$  (the electrostatic field strength) allows precursor ions and derivative product ions to be detected at the same time, even if fragmentation occurs between the source and the first detector. This is one example of a linked scan, where both analyzers are scanned simultaneously to preserve a predetermined relationship between their fields.

In ion-trap and FT-ICR instruments, product ion scanning is achieved by the ‘tandem-in-time’ route. Ions entering the trap at time 1 are subjected to a static field that selects for ions of one particular  $m/z$  ratio. These ions are collided with a stream of neutral atoms, a surface, or with radiation (see above) at time 2, which activates them and causes fragmentation. The resulting product ions are then scanned with a steadily increasing rf, which (in the case of a QIT instrument) accelerates them toward a sequential  $m/z$  hit detector.

### Product Ion Scan Mode ( $MS^n$ )

$MS^n$  analysis in ion traps and FT-ICR instruments is achieved by extending the ‘tandem-in-time’ principle, and iterating the time sequence discussed above. Product ions generated in the first round of fragmentation are scanned to give the product ion mass spectrum, but individual product ions can then be selected as second-round precursors, and fragmented again. Similar experiments up to  $MS^5$  have been carried out using sequentially aligned quadrupole analyzers (penta-quadrupole) but this is cumbersome and rarely used. In contrast,  $MS^n$  analysis in ion traps and FT-ICR is routinely carried out to  $MS^{10}$  and above, with at least one report describing the  $MS^{12}$  analysis of a complex glycoprotein sample.

### Precursor Ion Scan Mode

The precursor ion scan mode scheme is scan–dissociate–select ( $\circ \rightarrow \bullet$  in Kondrat symbols, and the opposite of the product ion scan mode). This mode is used in instruments containing two or more analyzers in series, but is unavailable in ion storage devices. The first analyzer scans the ion stream over a particular mass range and ions of increasing  $m/z$  value are passed to the collision cell and fragmented. The second analyzer is fixed to select product ions of a specific  $m/z$  ratio. Therefore, ions passing through

the first analyzer will be detected only if they yield the appropriate product ions during fragmentation. This mode is often used where the aim is to identify compounds in the analyte that carry a particular functional group. For example, organic compounds containing a phenyl group can be detected if the second analyzer is set to detect positive ions with an  $m/z$  value of 77, which corresponds to  $[C_6H_5^+]$ . Similarly, when MS/MS is used to analyze proteins or peptides in negative ion mode, the second analyzer can be set to detect ions with an  $m/z$  value of 79, which corresponds to  $[PO_3^-]$  and thus identifies proteins containing a phosphate group. False positives can occur in precursor ion scan mode if ions with the same  $m/z$  value as the selected group-specific fragment are generated by other fragmentation events. Precursor ion scans can also be performed using sector instruments if the magnetic and electrostatic sectors are linked in the appropriate fashion.

### Neutral Loss Scan Mode

The scheme for neutral loss scan mode is scan–dissociate–scan ( $\circ \rightarrow \circ$  in Kondrat symbols) and is an example of a linked scan using contemporary triple-quadrupole and hybrid instruments. Like the precursor ion scan, this mode is not available in trapping instruments but can be achieved in sector instruments by adjusting the relationship between the magnetic and electrostatic field strengths. In this mode, both the first and second analyzers are used to scan the ion stream, but the second analyzer is set out of register with the first so that it scans a parallel ion range at an  $m/z$  ratio that is lower by a fixed amount. This achieves the same aim as the precursor ion scan, i.e., it can help to identify compounds that carry specific functional groups, but does so in a different way – by scanning for the loss of those groups. In proteomics applications, for example, the second analyzer can be set to scan at 98 mass units below the first, which will specifically detect peptides that have lost a neutral  $H_3PO_4$  group. This helps to identify proteins that have phosphate groups attached to them.

### Selected Reaction Monitoring

The Kondrat representation for selected reaction monitoring is  $\bullet \rightarrow \bullet$ , i.e., select–dissociate–select. This mode can be used to monitor either single reactions (single reaction monitoring) or multiple reactions (multiple reaction monitoring) and is essentially a quantitative target analyte scan. In the case of a triple-quadrupole instrument, Q1 is set to select a particular  $m/z$  ratio, ions of this type are fragmented in the collision cell, and Q3 is set to detect specific products of that reaction, again by selecting a particular  $m/z$  ratio. For the detection of



particular compounds, selective reaction monitoring is  $\sim 10$  times more sensitive than neutral loss scanning, which in turn is  $\sim 10$  times more sensitive than product ion scanning.

## Applications of MS/MS and MS<sup>n</sup>

Tandem mass spectrometry has impacted on many areas of science, including organic chemistry, biochemistry, genetics, pharmacology, and medicine. In organic chemistry, the ability of MS/MS to fragment complex organic molecules and determine fragment ion masses helps not only to reveal molecular structures but also to distinguish between very similar organic molecules and perform quantitative analysis.

With a high-resolution mass spectrometer it is possible to achieve an accuracy of approximately one part in one million, which allows ions such as  $C_6H_{12}^+$  and  $C_4H_4O_2^+$  to be distinguished even though their molecular masses differ by only 0.0728 mass units. Although the fragmentation pattern of a given organic molecule cannot be predicted with accuracy, peaks can often be associated with particular chemical groups on an empirical basis. MS/MS can be applied to endpoint samples and reactions in progress, the latter providing important information about the structures of intermediates and by-products. When such studies are combined with isotopic labeling, the distribution of labeled atoms in the fragment ions can show which atoms are involved in particular reactions and whether intramolecular or intermolecular processes are involved. Specific applications in organic chemistry include the analysis of metabolic profiles, the detection of chemicals and explosives in the environment and in forensic samples, and the characterization of chemical libraries, as discussed below.

In recent years, MS/MS has played an important role in combinatorial chemistry, a discipline in which large libraries of molecules are created so that they can be screened as drugs. MS/MS is one of the most useful techniques for the analysis of structure–activity relationships in potentially therapeutic molecules, helping to determine which particular functional groups, revealed as product ions, are responsible for positive therapeutic effects. Although bioassays are needed to confirm the therapeutic properties of drugs, MS/MS can help to identify the most active components of libraries by correlating structural data with the results from functional assays. This can determine which features of each molecule are likely to be most important for interactions with drug targets. Soft ionization techniques such as electrospray ionization (ESI) and MALDI coupled to MS/MS have been at the forefront of new discovery

paradigms, including the identification and analysis of drug–target interactions, receptor–ligand binding, and enzyme–substrate–inhibitor relationships.

MS/MS has also become established in biochemistry and genetics, both as a discovery tool and a diagnostic technique. The high sensitivity of contemporary tandem and multidimensional mass spectrometers means they can be used to detect disease-specific molecules in tissue and blood samples down to the femtomole level. For example, MS/MS neutral loss scanning is used to detect increased levels of phenylalanine in the blood of newborn infants, to increase the detection rate and lower the false-positive rate associated with screening for phenylketonuria. Selected reaction monitoring is used widely to detect drug residues, such as steroids in athletes and domestic animals and in adulterated foods.

Perhaps the most important application of MS/MS in biology at the beginning of the new millennium is in the analysis of proteins, which until 1995 could only be sequenced using laborious methods based on cyclical chemical degradation. The advent of MALDI and ESI ion sources now permits the soft ionization of peptides, whose masses can be determined and correlated to those calculated from protein sequences stored in sequence databases. Unfortunately, the information provided by MS analysis is often insufficient for protein identification, which is where MS/MS comes into its own. Where peptide mass fingerprinting fails to identify any proteins matching those present in a given sample, the CID spectrum of one or more individual peptides may provide important additional information. The data can be used either in its uninterpreted form, to search for peptides whose sequences could yield fragments of identical mass to the experimentally determined peptide fragments, or the mass spectrum can be interpreted to give short peptide tags that can be used to search sequence databases for matching sequences. Taken to its extreme, this method can even be used for the *de novo* sequencing of proteins, since fragment ions can be assembled in order of increasing size, with differences between adjacent members of a nested set separated by masses that can be correlated with those of specific amino acids. As discussed above, precursor ion scanning and neutral loss scan modes can be used to identify proteins with particular forms of modification, such as phosphoproteins and glycoproteins. Multidimensional MS analysis has been especially useful for the sequencing and structural analysis of complex carbohydrates and glycoproteins.

**See also:** Mass Spectrometry; Mass Separation; Ion Traps; Time-of-Flight; Selected Ion Monitoring; Proteomics.

## Further Reading

- Aebersold R and Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422: 198–207.
- Chernushevich IV, Loboda AV, and Thomson BA (2000) An introduction to quadrupole–time-of-flight mass spectrometry. *Journal of Mass Spectrometry* 35: 1069–1090.
- De Hoffman E and van Stroobant (2001) *Mass Spectrometry: Principles and Applications*, 2nd edn. New York: Wiley.
- De Hoffmann E (1996) Tandem mass spectrometry: A primer. *Journal of Mass Spectrometry* 31: 129–137.
- Marshall AG, Hendrickson CL, and Jackson GS (1998) Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Review* 17: 1–35.
- McLafferty FW and Turecek F (1993) *Interpretation of Mass Spectra*, 4th edn. Herndon, VA: University Science Books.
- Shukla AK and Futrell JH (2000) Tandem mass spectrometry: Dissociation of ions by collisional activation. *Journal of Mass Spectrometry* 35: 1069–1090.
- Triolo A, Altamura M, Cardinali F, Sisto A, and Maggi CA (2001) Mass spectrometry and combinatorial chemistry: A short outline. *Journal of Mass Spectrometry* 36: 1249–1259.

## Stable Isotope Ratio

**K J Welham**, University of Hull, Hull, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by I P Wright, pp. 2869–2877, © 1995, Elsevier Ltd.

## Introduction

The term ‘stable isotope ratio’ in its normal context is restricted to a small group of elements (i.e., hydrogen, carbon, nitrogen, oxygen, and sulfur) that are readily converted into simple gaseous compounds for the purpose of isotope ratio determination using a mass spectrometer. This article will focus on the instrumentation used to measure stable isotope ratios, from dual inlet systems operating with dynamic gas flow through later developments for interfacing to gas chromatography and on to newer static vacuum designs. The fields of application of stable isotope ratio mass spectrometry are broad, transcending the life sciences, atmospheric chemistry, oceanography, agriculture, archaeology, cosmochemistry, geological sciences, etc. In these examples, stable isotopes are used, among other things, to elucidate metabolism, the sources of materials, to constrain the extent of chemical reactions, monitor diffusion, assess formation temperatures, and so on.

## Dual Inlet Mass Spectrometry

Dual inlet mass spectrometers designed for the purpose of analyzing gas samples were first built in the 1940s. Despite improvements in hardware components such as power supplies, electronic detection systems, and computer control, the modern commercially available mass spectrometer resembles its

forebears in many respects. A schematic diagram of a typical dual inlet mass spectrometer is shown in **Figure 1**. There are two identical inlets, one containing a sample gas of unknown isotopic composition and one containing a reference gas.

Each inlet consists of a variable volume, a cold finger, and a capillary crimped near to its termination. Both inlets are connected to a changeover valve mounted in close proximity to the ion source of the mass spectrometer, which is pumped to a ‘background pressure’ of  $<10^{-10}$  Torr ( $\sim 10^{-8}$  Pa; 1 Torr  $\sim 133$  Pa).

Sample and reference gases, stored at relatively high pressures in each of the inlets, flow continuously through the capillaries and into the changeover valve. At the end of each capillary a crimp allows fine-scale adjustment of flow rate (a procedure carried out only once, when the instrument is first commissioned). The changeover valve selects one of the gases for entry to the mass spectrometer, resulting in a typical ‘source pressure’ of  $10^{-6}$ – $10^{-7}$  Torr ( $\sim 10^{-4}$ – $10^{-5}$  Pa), while the other gas flows to waste. Some older mass spectrometers suffered from cross-contamination of the gases due to changeover valve design, necessitating a valve-mixing correction to adjust the measured isotopic composition; this problem has been eliminated in modern designs.

The isotope ratio of interest is obtained by comparing the sample measurement with that from the reference. Actually, several comparisons are made – sample and reference gases are repeatedly and alternately admitted to the mass spectrometer, with analyses taking 10 min and involving 20 or so sample/reference comparisons. Since the sample and reference gases are analyzed identically, there is no need for the mass spectrometer to determine the absolute isotope ratio; the comparative process ensures

## Further Reading

- Aebersold R and Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422: 198–207.
- Chernushevich IV, Loboda AV, and Thomson BA (2000) An introduction to quadrupole–time-of-flight mass spectrometry. *Journal of Mass Spectrometry* 35: 1069–1090.
- De Hoffman E and van Stroobant (2001) *Mass Spectrometry: Principles and Applications*, 2nd edn. New York: Wiley.
- De Hoffmann E (1996) Tandem mass spectrometry: A primer. *Journal of Mass Spectrometry* 31: 129–137.
- Marshall AG, Hendrickson CL, and Jackson GS (1998) Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Review* 17: 1–35.
- McLafferty FW and Turecek F (1993) *Interpretation of Mass Spectra*, 4th edn. Herndon, VA: University Science Books.
- Shukla AK and Futrell JH (2000) Tandem mass spectrometry: Dissociation of ions by collisional activation. *Journal of Mass Spectrometry* 35: 1069–1090.
- Triolo A, Altamura M, Cardinali F, Sisto A, and Maggi CA (2001) Mass spectrometry and combinatorial chemistry: A short outline. *Journal of Mass Spectrometry* 36: 1249–1259.

## Stable Isotope Ratio

K J Welham, University of Hull, Hull, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by I P Wright, pp. 2869–2877, © 1995, Elsevier Ltd.

## Introduction

The term ‘stable isotope ratio’ in its normal context is restricted to a small group of elements (i.e., hydrogen, carbon, nitrogen, oxygen, and sulfur) that are readily converted into simple gaseous compounds for the purpose of isotope ratio determination using a mass spectrometer. This article will focus on the instrumentation used to measure stable isotope ratios, from dual inlet systems operating with dynamic gas flow through later developments for interfacing to gas chromatography and on to newer static vacuum designs. The fields of application of stable isotope ratio mass spectrometry are broad, transcending the life sciences, atmospheric chemistry, oceanography, agriculture, archaeology, cosmochemistry, geological sciences, etc. In these examples, stable isotopes are used, among other things, to elucidate metabolism, the sources of materials, to constrain the extent of chemical reactions, monitor diffusion, assess formation temperatures, and so on.

## Dual Inlet Mass Spectrometry

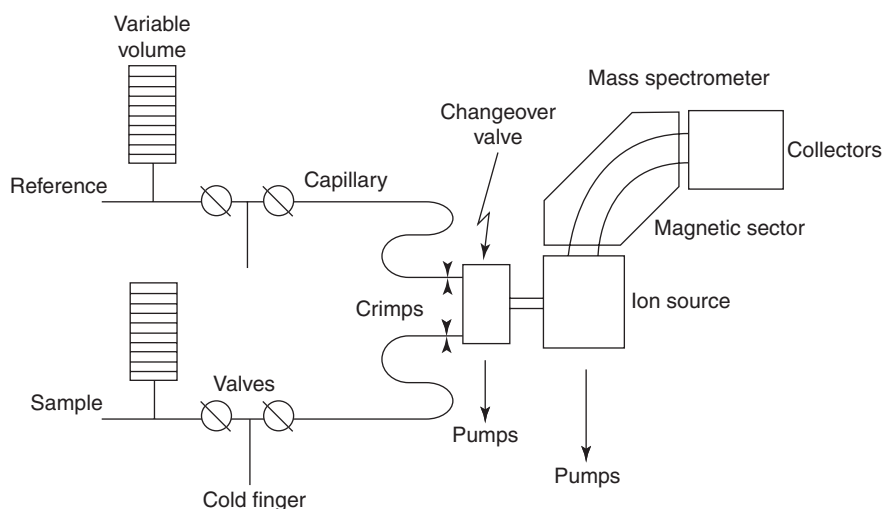
Dual inlet mass spectrometers designed for the purpose of analyzing gas samples were first built in the 1940s. Despite improvements in hardware components such as power supplies, electronic detection systems, and computer control, the modern commercially available mass spectrometer resembles its

forebears in many respects. A schematic diagram of a typical dual inlet mass spectrometer is shown in **Figure 1**. There are two identical inlets, one containing a sample gas of unknown isotopic composition and one containing a reference gas.

Each inlet consists of a variable volume, a cold finger, and a capillary crimped near to its termination. Both inlets are connected to a changeover valve mounted in close proximity to the ion source of the mass spectrometer, which is pumped to a ‘background pressure’ of  $<10^{-10}$  Torr ( $\sim 10^{-8}$  Pa; 1 Torr  $\sim 133$  Pa).

Sample and reference gases, stored at relatively high pressures in each of the inlets, flow continuously through the capillaries and into the changeover valve. At the end of each capillary a crimp allows fine-scale adjustment of flow rate (a procedure carried out only once, when the instrument is first commissioned). The changeover valve selects one of the gases for entry to the mass spectrometer, resulting in a typical ‘source pressure’ of  $10^{-6}$ – $10^{-7}$  Torr ( $\sim 10^{-4}$ – $10^{-5}$  Pa), while the other gas flows to waste. Some older mass spectrometers suffered from cross-contamination of the gases due to changeover valve design, necessitating a valve-mixing correction to adjust the measured isotopic composition; this problem has been eliminated in modern designs.

The isotope ratio of interest is obtained by comparing the sample measurement with that from the reference. Actually, several comparisons are made – sample and reference gases are repeatedly and alternately admitted to the mass spectrometer, with analyses taking 10 min and involving 20 or so sample/reference comparisons. Since the sample and reference gases are analyzed identically, there is no need for the mass spectrometer to determine the absolute isotope ratio; the comparative process ensures



**Figure 1** Dual inlet isotope-ratio mass spectrometer. The instrument is shown as a magnetic sector device with a 90° deflection angle. The mass spectrometer is operated in a dynamic mode, i.e., continuously open to the pumps.

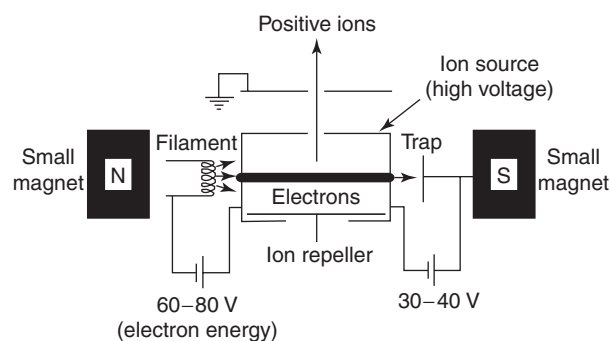
measurements of the appropriate accuracy. High levels of precision result from repeating the comparative procedure many times, this is important since tremendous significance may ultimately be attached to relatively small isotopic variations.

The operating protocol of the dual inlet ensures that when the flow rates of the two gases are equalized, by having equivalent pressures, any accompanying isotopic fractionations are the same in each case, and so the effects cancel out. In order to equalize the pressures of gases in both inlets, either one, or two, variable volumes are used to compress the gas(es) until identical source pressures are obtained. The variable volumes are usually stainless steel bellows, which can be mechanically opened or closed under computer control allowing very accurate 'balancing' of the source pressures. The balancing process is assessed by comparing the magnitude of two parameters for each gas – the source pressure and the intensities of the 'major' ion beams (i.e., those from the isotopically most abundant species, e.g.,  $^{12}\text{C}^{16}\text{O}^{16}\text{O}^+$  in the case of carbon dioxide). An instrument known to be operating satisfactorily will give similar major ion beam intensities, as measured by the mass spectrometer, for equivalent source pressures of sample and reference gases. Any discrepancy implies that one of the gases is contaminated, a situation that could result in a spurious isotope ratio measurement. The instrument itself can be used to assess whether contaminants are present by determining the mass spectrum of the gases in each inlet. The variable volume is potentially wasteful and dual inlet instruments generally incorporate cold fingers for the cryogenic concentration of the gases of interest. This allows measurements of the isotopic composition on 100 nmol quantities of gas.

The selected gas is ionized on entry into the mass spectrometer and the resultant ions separated according to their mass-to-charge ( $m/z$ ) ratio. Ionization is normally effected via an electron impact source of 'closed' design, whereby the flowing gas is admitted to a small box in the mass spectrometer ion source (the ion source is not completely closed; a small conductance to the pump ensures an equilibrium pressure is reached as gases flow into the device). This arrangement results in a greater sensitivity of the instrument since a localized high pressure is obtained in the ion source, which in turn results in more intense ion beams. Too high a source pressure can result in an unwanted ion-molecule reaction, producing the  $\text{H}_3^+$  ion, which causes interference at the masses detected during hydrogen isotope ratio measurement. In this case, it may be necessary to employ an ion source of more traditional, 'open' design, or to decrease the flow rate.

A schematic diagram of an ion source is shown in **Figure 2**. Regardless of type (open or closed), the ionizing electrons are provided by a heated filament, generally tungsten, or thoriated-iridium coated on tungsten. A 3–5 A current boils electrons off the filament by thermionic emission, which are drawn into the ion source by a 60–80 V potential difference between the filament and the rest of the ion source (referred to as the electron energy, typically 70 eV).

A further voltage difference, between the ion source and the trap plate (see **Figure 2**), produces a partially collimated electron beam. Electronic circuitry ensures that the trap current remains at a constant, preselected value (typically 100–1000  $\mu\text{A}$ ). Thus, any changes in filament current due to aging, or the effect of gas being admitted to the ion source, are counteracted by the preservation of a constant



**Figure 2** Typical electron impact ion source configuration for a stable isotope ratio mass spectrometer.

trap current. Small magnets produce a magnetic field parallel to the direction of the electron beam, increasing the effective path length by causing the electrons to spiral. The ion source is operated at a positive voltage of 2–5 kV above the ground potential. Thus, when an electron interacts with a neutral gas molecule and produces a positively charged ion this is ejected from the ion source and subsequently enters the magnetic field region of the mass spectrometer.

An isotope ratio measurement is derived from a determination of the relative intensities of the ion beams as registered at the detector(s). Since the ion beams are due to nuclides that differ in mass by approximately one, the mass resolution does not need to be greater than  $\sim 100$  Da. Thus, stable isotope ratio mass spectrometers are generally small, comprising a magnet that produces a flight path of the ions of the order of 6–12 cm radius, with a deflection angle of  $60^\circ$  or  $90^\circ$ . Typical ion beam intensities are generally greater than  $10^{-12}$  A; therefore, stable isotope ratio mass spectrometers do not require high-gain devices such as electron multipliers. Faraday bucket-type detectors referred to normally as a ‘collector’ are sufficient. Almost all stable isotope ratio mass spectrometers are fitted with multiple collectors capable of measuring the desired ion beams simultaneously. In the case of carbon dioxide, the three ion beams of interest are at  $m/z$  44 ( $^{12}\text{C}^{16}\text{O}^{16}\text{O}^+$ ),  $m/z$  45 ( $^{13}\text{C}^{16}\text{O}^{16}\text{O}^+$ ,  $^{12}\text{C}^{16}\text{O}^{17}\text{O}^+$ ), and  $m/z$  46 (predominantly  $^{12}\text{C}^{16}\text{O}^{18}\text{O}^+$ ). The use of multiple collectors reduces the analysis time and circumvents the problem of instability in the instrument following a ‘peak jumping’ routine, whereby the ion beams of interest are successively focused onto a single collector (i.e., by changing the magnitude of the accelerating voltage, or the magnetic field strength). Some older stable isotope ratio mass spectrometers intended for carbon dioxide analysis may only have two collectors although there are three ion beams that need to be detected. In this case it is necessary to peak-jump in order to measure the appropriate ion

**Table 1** Relative abundances on Earth of the stable isotopes used in stable isotope ratio mass spectrometry

Element	Isotope	Relative abundance (%)
Hydrogen	$^1\text{H}$	99.9844
	$^2\text{H}$ (D)	0.0156
Carbon	$^{12}\text{C}$	98.89
	$^{13}\text{C}$	1.11
Nitrogen	$^{14}\text{N}$	99.64
	$^{15}\text{N}$	0.36
Oxygen	$^{16}\text{O}$	99.7630
	$^{17}\text{O}$	0.0375
	$^{18}\text{O}$	0.1995
Sulfur	$^{32}\text{S}$	95.02
	$^{33}\text{S}$	0.75
	$^{34}\text{S}$	4.21
	$^{36}\text{S}$	0.02

beam ratios, which results in a doubling of the analysis time and, correspondingly, isotope ratio measurements that are not as precise as those obtained from instruments with three collectors. Stable isotope ratio instruments are intended for the analysis of individual, uncomplicated gas species, such as carbon dioxide. Thus, as long as the less intense (minor) beams are adequately focused and collected, the major ions can essentially just be directed toward a wide collector (i.e., there will not be any interfering ions from other species to contribute to the signal).

Each detector is connected to a high-gain amplifier mounted immediately outside the high-vacuum chamber of the mass spectrometer. In sophisticated designs these amplifiers are mounted in their own vacuum chamber (at  $\sim 10^{-3}$  Torr), which helps to stabilize the environmental conditions surrounding the amplifiers and reduces instability in the measured ion beam signal. The amplifiers convert the small ion currents into dc voltages ( $\sim 0.1$ – $10$  V), which can then be further amplified if necessary, processed by an analog-to-digital converter, and read into a desktop computer. More recently, quadrupole instruments have also been used for isotope ratio measurements.

The gases that are normally used for making isotopic measurements are  $\text{H}_2$  (for the D/H, or  $^2\text{H}/^1\text{H}$ , ratio),  $\text{CO}_2$  (for  $^{13}\text{C}/^{12}\text{C}$  and  $^{18}\text{O}/^{16}\text{O}$ ),  $\text{N}_2$  ( $^{15}\text{N}/^{14}\text{N}$ ), and  $\text{SO}_2$  ( $^{34}\text{S}/^{32}\text{S}$ ). In some instances, it is desirable to analyze gases such as  $\text{O}_2$  ( $^{17}\text{O}/^{16}\text{O}$  and  $^{18}\text{O}/^{16}\text{O}$ ) or  $\text{SF}_6$  ( $^{34}\text{S}/^{32}\text{S}$ ,  $^{33}\text{S}/^{32}\text{S}$ ,  $^{36}\text{S}/^{32}\text{S}$ ). A list of the natural abundances of the various isotopes is given in Table 1.

In principle, it should be possible to design a mass spectrometer capable of analyzing all the above gases, i.e., there are no special requirements in terms of the materials used to construct the instrument (although for  $\text{SF}_6$  it is necessary to determine masses of  $\sim 130$ , the parent ion being  $\text{SF}_5^+$ ). In reality, since the various gases may have different requirements in



terms of gas inlets, or instrumental operating conditions, it is commonplace for individual instruments to be dedicated to the analysis of single gas species. Constraints imposed by gas extraction systems lead to sulfur isotope analyses being carried out on dedicated sulfur dioxide instruments. Nitrogen too needs specialized treatment because of the obvious possibilities of atmospheric contamination. A carbon dioxide machine, on the other hand, is able to measure both carbon and oxygen isotopic compositions. Since a major interest in oxygen isotope analyses arises from studies of water samples, there is a demand that any facility capable of analyzing oxygen should also be able to measure hydrogen isotopes as well. In the past this was resolved by the use of a 'Siamese' head, i.e., two mass spectrometer flight tubes (one for carbon dioxide, the other for hydrogen) with only the one dual inlet. More recently, the problem has been circumvented by the evolution of the 'split' flight tube that allows carbon dioxide and hydrogen to be admitted to a single mass spectrometer, the magnetic sector being configured to accept ion beams from either species. Any compromise in the precision to which the D/H ratio can be measured under these circumstances is outweighed by the relatively large range in hydrogen isotopic compositions that exist in nature.

As stated above, the isotope ratio measurements are made by comparing the ion beam intensities obtained from an unknown sample gas with those from a reference gas. In the case of nitrogen, which in the gaseous form exists as the diatomic molecule  $N_2$ , and which has only two stable isotopes ( $^{14}N$  and  $^{15}N$ ), the isotope ratio measurement is made simply by a comparison of sample and reference ion beams at  $m/z$  29 and 28 (i.e.,  $^{14}N^{15}N^+$  and  $^{14}N^{14}N^+$ ). In a more complicated molecule such as carbon dioxide it is necessary to consider both the stable isotopic composition of carbon (which has isotopes  $^{12}C$  and  $^{13}C$ ) and oxygen ( $^{16}O$ ,  $^{17}O$ , and  $^{18}O$ ). Thus, a  $^{13}C/^{12}C$  measurement cannot simply be made by analyzing the ion beams at  $m/z$  45 ( $^{13}C^{16}O^{16}O^+$ ) and  $m/z$  44 ( $^{12}C^{16}O^{16}O^+$ ) since the former beam partly consists of  $^{12}C^{16}O^{17}O^+$ . Thus, it is always necessary to make an oxygen isotopic determination as well (by measuring  $m/z$  46) so that a correction can be applied to the measured  $m/z$  44/45 ratio.

### Isotope Ratio Measurements, Atomic Percentage Enrichments, the $\delta$ Value, and Precision

In their most straightforward usage, isotope ratio mass spectrometers are used to detect artificially

enriched isotopic components as they are introduced into a microenvironment of some description. For instance, isotopically labeled components (normally enriched in heavy isotopes, e.g., D,  $^{13}C$ ,  $^{15}N$ ) may be administered to living organisms in an attempt to trace metabolic pathways. In this case it may not always be necessary to determine the isotope ratio to a high degree of precision and it is customary to quote measured values according to their atomic enrichments, as follows:

$$\text{Atomic \%} = 100/(1 + 1/R)$$

where  $R$  is the ratio of the heavy isotope to the light isotope. Note that for the elements considered herein, the heavy isotopes are always of lesser abundance than the lightest one (this is not true of all elements of course). In the case of carbon a typical value of  $R$  ( $^{13}C/^{12}C$ ) in the terrestrial environment is 0.0112; thus, the atomic percentage of  $^{13}C$  is  $\sim 1.11\%$ . In the same way, a measurement of  $^{13}C$ -enriched material (i.e., elevated  $^{13}C/^{12}C$  ratio) can be quoted in terms of its atomic percentage enrichment and compared with the typical terrestrial value.

In more sophisticated uses of isotope ratio mass spectrometry, the objective will be to detect the relatively small differences in isotopic composition that may exist (1) between different natural occurrences of an identical compound, or (2) when an element is partitioned between two constituents of a natural system. The hydrogen and oxygen isotopic compositions of natural water samples are an example of type (1). On the other hand, type (2) is exemplified by studies that aim to determine the oxygen isotopic composition of calcium carbonate deposited by living marine organisms; knowing the difference in oxygen isotopic composition between carbonate and seawater enables calculation of the temperature of formation. By analyzing fossilized organisms it is possible to document historical variations in temperature.

In studies that require measurement of relatively minor variations in isotopic compositions, it is necessary for the analyses to have errors that are small compared with the anticipated range. Since isotopic variations in natural systems are sometimes quite limited, differences are quoted in per mil (parts per 1000) rather than percent (parts per 100). The symbol for per mil is ‰, where 10‰ is equivalent to 1%. The precision needed for isotopic measurements is generally less than  $\pm 1\%$ , and more probably of the order of  $\pm 0.1\%$ . Note that the ultimate precision possible using a dual inlet mass spectrometer is of the order of  $\pm 0.05\%$ .

Isotope ratios are quoted as per mil variations from a reference point according to the following

formula:

$$\delta \text{ (in ‰)} = \left( \frac{R_{\text{Sam}}}{R_{\text{Ref}}} - 1 \right) \times 1000$$

where  $R_{\text{Sam}}$  is the ratio of heavy to light isotope in the sample and  $R_{\text{Ref}}$  is the equivalent ratio in a reference material. Thus, for example, in the case of carbon, where  $R = {}^{13}\text{C}/{}^{12}\text{C}$ ;

$$\delta^{13}\text{C} = \left( \frac{({}^{13}\text{C}/{}^{12}\text{C})_{\text{Sam}}}{({}^{13}\text{C}/{}^{12}\text{C})_{\text{Ref}}} - 1 \right) \times 1000\text{‰}$$

To facilitate the  $\delta$  values being discussed in an international forum it is necessary to cross-calibrate the data obtained from individual laboratories. This is accomplished by knowing the  $\delta$  value of each laboratory's reference material relative to an internationally accepted standard. The standards used for stable isotopic measurements are given in Table 2. In the case of carbon, where the standard is Pee Dee belemnite (PDB),

$$\delta^{13}\text{C}_{\text{PDB}} = \left( \frac{({}^{13}\text{C}/{}^{12}\text{C})_{\text{Sam}}}{({}^{13}\text{C}/{}^{12}\text{C})_{\text{PDB}}} - 1 \right) \times 1000\text{‰}$$

Since, in this particular case, there is none of the PDB standards remaining, the  $\delta^{13}\text{C}_{\text{PDB}}$  value of an unknown sample is obtained from:

$$\delta^{13}\text{C}_{\text{Sam-PDB}} = \delta^{13}\text{C}_{\text{Sam-Ref}} + \delta^{13}\text{C}_{\text{Ref-PDB}} + \frac{\delta^{13}\text{C}_{\text{Sam-Ref}} \times \delta^{13}\text{C}_{\text{Ref-PDB}}}{1000}$$

where  $\delta^{13}\text{C}_{\text{Sam-Ref}}$  is the measurement of the unknown sample relative to the reference, and  $\delta^{13}\text{C}_{\text{Ref-PDB}}$  is the value for the reference relative to PDB (a datum that may have been derived in the past when PDB was still in evidence, or via other reference materials). Appropriate reference materials are available from the National Institute of Standards and Technology (Gaithersburg, Maryland, USA).

## Gas Preparation Techniques

In all but a few cases it is necessary to convert the material under investigation into a gas for stable isotope mass spectrometric measurement. Even if the element already exists in the appropriate form (e.g.,  $\text{CO}_2$  trapped in ice cores, or as the product of respiration) purification may still be necessary. It is impossible to list full details of all gas extraction procedures here and, as such, only a brief outline is given for each element. Hydrogen isotopic compositions are derived from hydrogen ( $\text{H}_2$ ) gas; this is generally obtained via the reduction of water, using heated uranium or zinc. Sample waters thus require only minimal preparation, whereas hydrogen from more refractory sources is converted to water either by direct heating (pyrolysis), or under an atmosphere of oxygen (combustion). For carbon, the analysis gas is carbon dioxide,  $\text{CO}_2$ . This can be obtained by combustion techniques or, as in the case of carbonate analyses, reaction of the mineral with acid to produce  $\text{CO}_2$ . Nitrogen is prepared either by combustion, pyrolysis, or, in the case of organic materials, a more complex technique known as the Kjeldahl method. A particular problem for nitrogen is to ensure that no gaseous oxides are formed during the preparation procedure. Oxygen isotopic compositions of carbonates are prepared by acid treatment (as in the case of carbon) whereas the oxygen in water samples is measured by an isotopic equilibration procedure involving  $\text{CO}_2$ . In order to determine the oxygen isotopic composition of a silicate rock sample, the minerals are treated with a fluorinating agent (such as  $\text{F}_2$ ,  $\text{ClF}_3$ , or  $\text{BrF}_5$ ), which displaces the oxygen, producing  $\text{O}_2$  gas. This is then either measured directly, or converted to  $\text{CO}_2$  by reaction with graphite. Sulfur bearing constituents are generally separated and combusted to  $\text{SO}_2$ , or fluorinated to produce  $\text{SF}_6$ .

In all the above procedures it is necessary to purify the gases prior to analysis. This is accomplished by chemical, or cryogenic, means. The pressure of the purified gas can then be measured, for quantification

**Table 2** Reference materials used for making stable isotope ratio measurements

Element	Reference	Isotopes	Absolute ratio	Comments
Hydrogen	SMOW	D/H	$1.5576 \times 10^{-4}$	Standard mean ocean water (an artificial standard closely representing sea water)
Carbon	PDB	${}^{13}\text{C}/{}^{12}\text{C}$	$1.12372 \times 10^{-2}$	Pee Dee belemnite ( <i>Belemnitella Americana</i> , a fossil from S. Carolina, USA)
Nitrogen	AIR	${}^{15}\text{N}/{}^{14}\text{N}$	$3.67647 \times 10^{-3}$	Atmospheric nitrogen (known to be isotopically homogeneous)
Oxygen	SMOW	${}^{18}\text{O}/{}^{16}\text{O}$	$2.0052 \times 10^{-3}$	Standard mean ocean water (also use PDB for carbonate studies)
		${}^{17}\text{O}/{}^{16}\text{O}$	$3.8288 \times 10^{-4}$	
Sulfur	CDT	${}^{34}\text{S}/{}^{32}\text{S}$	$4.50045 \times 10^{-2}$	Canyon Diablo Troilite ( $\text{FeS}$ from an iron meteorite)
		${}^{33}\text{S}/{}^{32}\text{S}$	$8.09976 \times 10^{-3}$	
		${}^{36}\text{S}/{}^{32}\text{S}$	$1.79058 \times 10^{-4}$	

purposes, following which it is admitted to the mass spectrometer. The laborious and time-consuming nature of gas preparation has led to the development of certain automated gas-extraction techniques (e.g., batch combustion, multiple sample acid dissolution devices, etc.). On a more research-orientated front, the laser microprobe uses a focused laser beam ( $\sim 50\text{ }\mu\text{m}$ ) to heat a localized area of the sample of interest. The benefits of this technique lie in its inherent rapidity and also the low associated blank levels. Laser pyrolysis has been used successfully to liberate nitrogen from various samples, and to deprecipitate carbonates to  $\text{CO}_2$  gas. When laser heating is carried out under a pressure of oxygen, it is possible to conduct microscale combustion. A further implementation involves using a laser beam in conjunction with a fluorinating agent to allow oxygen isotopic analyses of small quantities of silicate samples.

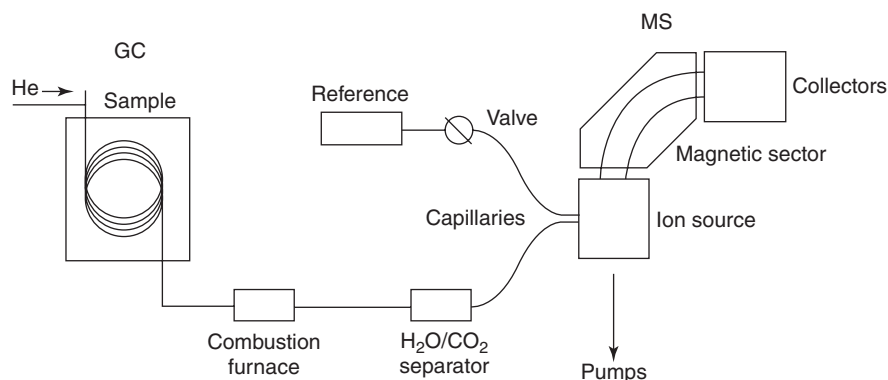
## Gas Chromatography–Isotope Ratio Mass Spectrometry

Stable isotope ratio mass spectrometers that interface directly to a gas chromatograph are commercially available. This type of analysis is known variously as gas chromatography–isotope ratio mass spectrometry (GC–IRMS), gas chromatography–combustion isotope ratio mass spectrometry, or compound-specific isotope analysis. The same principle of operation has also been used for more general isotope ratio analyses, where the components of interest are combusted as bulk samples. This type of arrangement, known as continuous flow–isotope ratio mass spectrometry, allows rapid automated analyses of  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$ .

The mass spectrometers used for GC–IRMS have dual inlets, although in this case there are no

changeover valves (see **Figure 3**). Rather, a flow of helium gas through two capillaries transports small quantities of sample and reference gases into the mass spectrometer ion source. Complex mixtures of organic compounds are injected into a capillary column gas chromatograph and then separated according to normal principles. Each compound is then swept through a combustion furnace (comprising copper oxide and a platinum catalyst, operated at a temperature of  $\sim 800^\circ\text{C}$ ) where it is converted to  $\text{CO}_2$  and  $\text{H}_2\text{O}$  (and  $\text{N}_2$ , if the compounds are nitrogen bearing). In principle, it should be possible, using GC–IRMS, to make stable isotope ratio measurements of hydrogen, carbon, and nitrogen; oxygen is excluded since, under these circumstances, it arises from the combustion furnace. Considerable successes have been achieved with the determination of carbon and nitrogen isotope ratios using GC–IRMS.

An immediate problem that has to be addressed regarding the mass spectrometry in a GC–IRMS system is the separation of  $\text{N}_2$ ,  $\text{CO}_2$ , and  $\text{H}_2\text{O}$  out of a flowing stream of gas. This is accomplished in practice by various means, including cryogenic removal, and/or the use of water-permeable membranes. Purified  $\text{N}_2$  or  $\text{CO}_2$  is then carried into the ion source of the mass spectrometer using the flow of helium. The flow of carrier gas into the mass spectrometer results in higher source pressures and turbo-molecular pumps provide a higher pumping capacity than that used for dual inlet machines. In other respects the operation of the ion source and mass spectrometer can be considered similar to the conventional case. A fundamental difference is that the ion beam intensities for an individual  $\text{N}_2$  or  $\text{CO}_2$  sample can be measured for only a brief period of time (a few seconds). Furthermore, since there is an isotopic fractionation associated with the flowing gas, the measured isotopic composition changes markedly



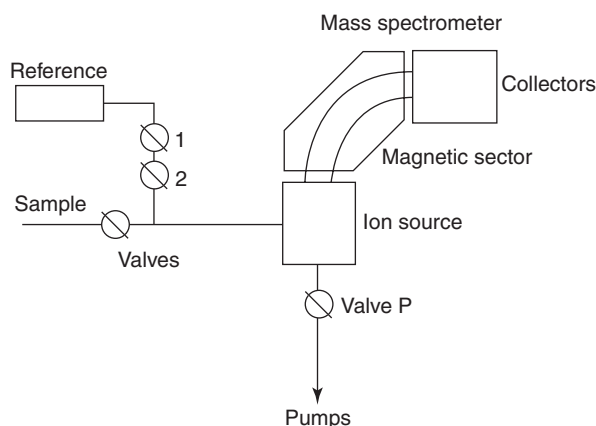
**Figure 3** Gas chromatography–isotope ratio mass spectrometer (GC–IRMS). The mass spectrometer (MS) depicted here is similar in many respects to that shown in **Figure 1**. The instrument operates on a dynamic principle, although in this case the capillaries are not connected to a changeover valve.

during the course of analysis. In order to combat this effect it is necessary to integrate the ion beam yields obtained for each isotopic species during the analysis. It goes without saying that multiple collections are essential since there is no time available for peak jumping.

The sensitivity of the GC-IRMS system is very high compared with a normal dual inlet mass spectrometer. It has been shown that isotope ratio measurements of carbon in 100 pmol samples of  $\text{CO}_2$  can be measured to precisions of  $\pm 0.5\%$ . The level of precision is limited by the sample size and also by the nature of the operation of the GC-IRMS. Because eluted compounds are continuously flowing from the gas chromatograph column and into the combustion furnace, it is not possible to derive the isotopic measurements from rapid and repeated sample/reference comparisons, as in the case of a dual inlet mass spectrometer. Thus, it is generally only possible to contemplate an analysis of a reference gas at the beginning, or end, of a run (or preferably both). Under normal circumstances this is affected by admitting an aliquot of reference gas (of, say, 1–10 nmol) into the mass spectrometer via the second capillary. This gas is swept into the ion source by its own stream of helium carrier gas and analyzed in the same way as gases produced by the combustion furnace. If an aliquot of reference gas is analyzed before and after the sequence of combusted organic compounds elutes from the gas chromatograph, the results serve as a check on the stability of the instrument over this time period. It is then possible to derive  $\delta$  values for individual compounds by comparing ion beam intensities with those obtained from the reference gas at the start and end of the run. In order to obtain results of the highest precision it is desirable to calibrate the instrument with respect to the compounds under investigation. This is carried out by injecting reference compounds of known isotopic compositions into the gas chromatograph and measuring the isotopic compositions of the gases produced by the combustion furnace. Ideally, these reference compounds need to have isotopic compositions that are similar to the unknown samples under investigation.

## Static Mass Spectrometry

The GC-IRMS system achieves its high level of sensitivity by virtue of the fact that the entire gas sample is admitted to the mass spectrometer for analysis. However, because the mass spectrometer operates dynamically, the small quantities of gas that enter the ion source are very quickly pumped away (i.e., the gas may flow through the ion source for only a few



**Figure 4** Static isotope ratio mass spectrometer. The mass spectrometer depicted here, which is similar in many respects to that shown in **Figure 1**, is isolated from the pumping system during the analysis, by means of closing a valve, P (i.e., a static mode of operation).

seconds). An alternative approach to the analysis of small samples of gas is to use a static vacuum mass spectrometer (see **Figure 4**) where the entire sample (or a portion thereof contained between valves 1 and 2) is admitted to the instrument after it has been isolated from the pumps.

This is essentially the way in which isotopic compositions of noble gases are measured, where the unreactive species are able to remain in the static vacuum for long periods of time, the only losses being due to ionization and subsequent implantation. In contrast, 'active' gases have certain special problems, not least of all because they may undergo reactions within the mass spectrometer during analysis. Furthermore, gases may leak from the surrounding air into the high vacuum of the instrument; there may also be outgassing effects. Notwithstanding these problems, static mass spectrometers have been successfully used to analyze gases such as  $\text{CH}_4$ ,  $\text{CO}_2$ , and  $\text{N}_2$ . Thus so far it has proved possible to measure carbon and nitrogen isotope ratios routinely using static mass spectrometry; furthermore, it has been demonstrated that oxygen and hydrogen isotopic measurements are also feasible.

A static mass spectrometer only has a single inlet and no changeover valve. The required  $\delta$  value is obtained by comparing the ion beam intensities of the unknown sample with aliquots of a reference gas admitted to the instrument before and after the sample. Since it is essential to pump the mass spectrometer to high vacuum ( $\sim 10^{-10}$  Torr) after each analysis, the total time required to produce one  $\delta$  value is  $\sim 30$  min. Notwithstanding any instabilities that may operate over this timescale, precisions in  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  of  $\pm 0.5\%$  can be obtained routinely

using static mass spectrometry, from measurements of 100 pmol quantities of CO<sub>2</sub> and N<sub>2</sub>. Furthermore, the instruments' potential capabilities have been demonstrated down to the level of 10 pmol.

The techniques used to prepare the gases for analysis are essentially the same as those used for dual inlet mass spectrometry although a major consideration here is the blank levels that are incurred. Because of current blank limitations static mass spectrometry is normally restricted to investigations at the level of 0.1–1 nmol although the advent of new methods of gas extraction, such as the laser microprobe technique, may circumvent this problem.

See also: **Gas Chromatography: Mass Spectrometry. Mass Spectrometry: Principles; Mass Separation.**

## Further Reading

Bridger NI, Craig RD, and Sercombe JSF (1974) New mass spectrometer for isotopic analysis of small gas samples. *Advances in Mass Spectrometry* 6: 365–375.

Brenna JT, Corso TN, Tobias HJ, and Caimi RJ (1997) High precision continuous-flow isotope ratio mass spectrometry. *Mass Spectrometry Reviews* 16: 227–258.

Griffiths H (ed.) (1997) *Stable Isotopes: The Integration of Biological, Ecological and Geological Processes. Environmental Plant Biology Series*: BIOS Scientific Publishers.

Gross ML (ed.) (2004) *The Encyclopedia of Mass Spectrometry: Elemental and Isotope Ratio Mass Spectrometry*, vol. 5. Amsterdam: Elsevier.

Habfast K (1991) *The Methodology and Instrumentation of GC IRMS*. Bremen: Finnigan-MAT.

Hoefs J (1987) *Stable Isotope Geochemistry*. Berlin: Springer.

Meier-Augenstein W (1999) Applied gas chromatography coupled to isotope ratio mass spectrometry. *Journal of Chromatography A* 842: 351–371.

Platzner IT (1997) *Modern Isotope Ratio Mass Spectrometry*. Chichester: Wiley.

Prosser SJ, Wright IP, and Pillinger CT (1990) A preliminary investigation into the isotope measurement of carbon at the picomole level using static vacuum mass spectrometry. *Chemical Geology* 83: T1–T88.

Wright IP (1984)  $\delta^{13}\text{C}$  measurements of smaller samples. *Trends in Analytical Chemistry* 3: 210–215.

## Pyrolysis

**R M Twyman**, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Pyrolysis mass spectrometry (PyMS) is a sensitive analytical technique in which samples are subjected to rapid thermal degradation (pyrolysis) before the derivative ions are separated in a mass spectrometer. The technique is particularly suitable for the analysis of otherwise nonvolatile compounds in complex samples. Pyrolysis causes the weakest bonds in each molecule to break, forming volatile fragments (the pyrolysate) whose masses are expressed as a pyrolysis mass spectrum. This can be interpreted and used to characterize the structures of the parent molecules in the sample, or used in an uninterpreted manner to provide a diagnostic fingerprint that can easily distinguish between related complex samples, such as different microbial strains. Quantitative analysis is possible using sophisticated statistical methods based on neural networks and supervised learning, making PyMS useful for a very wide variety of applications.

Two critical aspects of the technique are the requirement for rapid and reproducible heat input, so

that degradation occurs in the context of a steep thermal gradient, and the need to remove primary degradation products from the pyrolysis zone immediately after they are formed. Under these circumstances, the primary pyrolysis events can be described as instantaneous, and secondary events (such as further degradation and recombination between primary fragments) can be avoided. Additionally, the time constant of the analysis should ideally be compatible with that of the degradation process itself. The coupling of *in vacuo* Curie-point pyrolysis (see below) with high-speed quadrupole or ion-trap mass spectrometry achieves all of these requirements, and is therefore the most widely practiced form of PyMS today. Curie-point pyrolysis is also the basis of all commercially available PyMS instruments.

## Development and Instrumentation

### Early Development

The origins of PyMS can be traced back to 1952, when Zemany first investigated the use of thermal degradation products for the characterization of biopolymers. In these initial experiments, the pyrolysis step was carried out offline and the pyrolysate was



using static mass spectrometry, from measurements of 100 pmol quantities of CO<sub>2</sub> and N<sub>2</sub>. Furthermore, the instruments' potential capabilities have been demonstrated down to the level of 10 pmol.

The techniques used to prepare the gases for analysis are essentially the same as those used for dual inlet mass spectrometry although a major consideration here is the blank levels that are incurred. Because of current blank limitations static mass spectrometry is normally restricted to investigations at the level of 0.1–1 nmol although the advent of new methods of gas extraction, such as the laser microprobe technique, may circumvent this problem.

See also: **Gas Chromatography: Mass Spectrometry. Mass Spectrometry: Principles; Mass Separation.**

## Further Reading

Bridger NI, Craig RD, and Sercombe JSF (1974) New mass spectrometer for isotopic analysis of small gas samples. *Advances in Mass Spectrometry* 6: 365–375.

Brenna JT, Corso TN, Tobias HJ, and Caimi RJ (1997) High precision continuous-flow isotope ratio mass spectrometry. *Mass Spectrometry Reviews* 16: 227–258.

Griffiths H (ed.) (1997) *Stable Isotopes: The Integration of Biological, Ecological and Geological Processes. Environmental Plant Biology Series*. BIOS Scientific Publishers.

Gross ML (ed.) (2004) *The Encyclopedia of Mass Spectrometry: Elemental and Isotope Ratio Mass Spectrometry*, vol. 5. Amsterdam: Elsevier.

Habfast K (1991) *The Methodology and Instrumentation of GC IRMS*. Bremen: Finnigan-MAT.

Hoefs J (1987) *Stable Isotope Geochemistry*. Berlin: Springer.

Meier-Augenstein W (1999) Applied gas chromatography coupled to isotope ratio mass spectrometry. *Journal of Chromatography A* 842: 351–371.

Platzner IT (1997) *Modern Isotope Ratio Mass Spectrometry*. Chichester: Wiley.

Prosser SJ, Wright IP, and Pillinger CT (1990) A preliminary investigation into the isotope measurement of carbon at the picomole level using static vacuum mass spectrometry. *Chemical Geology* 83: T1–T88.

Wright IP (1984)  $\delta^{13}\text{C}$  measurements of smaller samples. *Trends in Analytical Chemistry* 3: 210–215.

## Pyrolysis

**R M Twyman**, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Pyrolysis mass spectrometry (PyMS) is a sensitive analytical technique in which samples are subjected to rapid thermal degradation (pyrolysis) before the derivative ions are separated in a mass spectrometer. The technique is particularly suitable for the analysis of otherwise nonvolatile compounds in complex samples. Pyrolysis causes the weakest bonds in each molecule to break, forming volatile fragments (the pyrolysate) whose masses are expressed as a pyrolysis mass spectrum. This can be interpreted and used to characterize the structures of the parent molecules in the sample, or used in an uninterpreted manner to provide a diagnostic fingerprint that can easily distinguish between related complex samples, such as different microbial strains. Quantitative analysis is possible using sophisticated statistical methods based on neural networks and supervised learning, making PyMS useful for a very wide variety of applications.

Two critical aspects of the technique are the requirement for rapid and reproducible heat input, so

that degradation occurs in the context of a steep thermal gradient, and the need to remove primary degradation products from the pyrolysis zone immediately after they are formed. Under these circumstances, the primary pyrolysis events can be described as instantaneous, and secondary events (such as further degradation and recombination between primary fragments) can be avoided. Additionally, the time constant of the analysis should ideally be compatible with that of the degradation process itself. The coupling of *in vacuo* Curie-point pyrolysis (see below) with high-speed quadrupole or ion-trap mass spectrometry achieves all of these requirements, and is therefore the most widely practiced form of PyMS today. Curie-point pyrolysis is also the basis of all commercially available PyMS instruments.

## Development and Instrumentation

### Early Development

The origins of PyMS can be traced back to 1952, when Zemaný first investigated the use of thermal degradation products for the characterization of biopolymers. In these initial experiments, the pyrolysis step was carried out offline and the pyrolysate was

introduced manually into the mass spectrometer. Even so, it was shown that the thermal degradation products of proteins such as albumin yielded reproducible molecular fingerprints that, in principle, could be used to identify these proteins in other samples.

Throughout the 1950s and 1960s, there were continuing developments in pyrolysis techniques and in the use of mass spectrometry to analyze thermal degradation products. Although many investigators demonstrated the principle of online pyrolysis followed by mass spectrometry, most of the work carried out during this period focused on the use of relatively slow heating methods, which made the pyrolysis conditions difficult to reproduce and did not prevent the occurrence of secondary events. Even rapid heating methods, such as galvanic filament pyrolysis and laser-induced pyrolysis, were found to be difficult to control in the context of PyMS and were shown to induce many competing processes.

The Curie-point pyrolyzer, described by Giacobbo and Simon in 1964, was to provide the key to solving many of these problems. In this method, pyrolysis is achieved by the rapid inductive heating of a ferromagnetic substrate (e.g., a wire or foil) using an electromagnetic coil. The temperature increases steeply and in a linear fashion until the metal reaches its Curie point. When this temperature is achieved, the substrate loses its ferromagnetic properties and inductive heating ceases. However, as soon as the temperature drops below the Curie point, the ferromagnetic properties return and heating resumes. Curie-point pyrolysis, therefore, has the powerful advantage that skin heating rates are constant, e.g., for wires of identical cross-section, in a given electromagnetic field. This means that the heating step is both rapid and highly reproducible, two of the most important requirements for analytical PyMS. Furthermore, the use of wires or other disposable metal holders means that each pyrolysis event can be carried out on a clean surface, reducing the likelihood of cross-contamination between samples.

The first meaningful study involving Curie-point PyMS was published in 1970, and dealt with the analysis of dyes, fatty acid derivatives, and substituted benzoic acids. The apparatus used in this study consisted of a Curie-point pyrolyzer coupled to a magnetic sector mass spectrometer by means of a short length of capillary tubing. The tubing led from the pyrolyzer to a conventional molecular separator positioned in front of the ionization chamber. This instrument became the benchmark for future developments in PyMS and led directly to the introduction of the first automated Curie-point pyrolysis mass spectrometer in 1973.

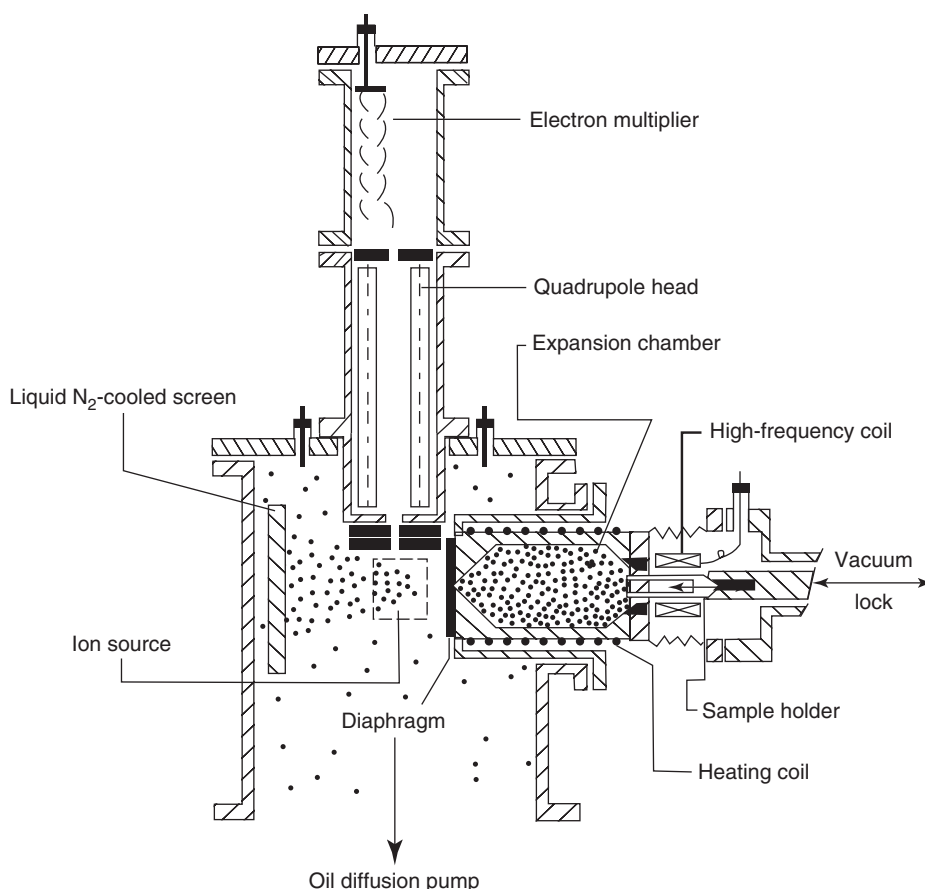
### Later Development: Automated PyMS

From 1970 onwards, most of the impetus for further developments in PyMS instrumentation was confined to a group working at the FOM-Institut in Amsterdam. In 1973, Meuzelaar and Kistemaker described the differentiation of bacterial strains using PyMS. They developed an instrument in which the pyrolyzer was integral with the vacuum system of a high-speed quadrupole mass spectrometer, and was connected directly into an electron-impact ionization source operating in low-voltage mode so that the emerging pyrolysis products would be ionized but not subject to further fragmentation. The ionization source was connected to the quadrupole by means of a heated, gold-plated expansion chamber, which was surrounded by a liquid-nitrogen-cooled screen to condense out neutrals and prevent them fouling the mass analyzer. From there, it was a short step to the design and construction of a fully automated system (**Figure 1**) in which a rotary magazine of 40 sample wires in glass carriers (**Figure 2**) was loaded into the extended vacuum system of the quadrupole instrument. The coated sample wires, in their protective glass carriers, were automatically introduced into the pyrolysis zone and heated sequentially. A series of mass scans over the appropriate mass range was then recorded.

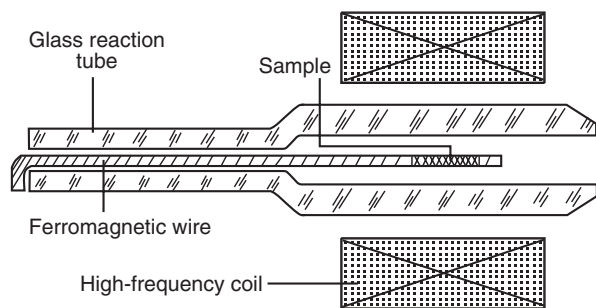
Further notable developments include renewed interest in the use of laser source pyrolysis, particularly in combination with time-of-flight mass spectrometry, which offers the possibility of targeting particular chemical bonds for degradation, and the coupling of analytical pyrolysis and tandem mass spectrometry, to induce the fragmentation of specific pyrolysate ions. The latter has culminated in the construction of a number of novel instruments, including a pyrolysis ion trap mass spectrometer capable of MS<sup>n</sup> analysis.

### Overview of the PyMS Method

Correct sample preparation and presentation are very important factors for the accuracy of PyMS, and sample preparation in quadruplicate is often carried out to ensure reproducibility. Samples readily soluble in a volatile solvent cause few problems other than a tendency to bead. Thin films are required to avoid primary degradation products diffusing through the undegraded or partially degraded sample, which presents an opportunity for secondary events. Beading can be avoided by adding a small amount of tetrahydrofuran or dioxan to the sample, since these substances act as wetting agents. Insoluble samples can be ground to powder and dispersed in a volatile liquid before placing on the Curie-point substrate, although care must be exercised to avoid mechanical degradation.



**Figure 1** Schematic section of a pyrolysis mass spectrometer.



**Figure 2** Detail showing sample carrier. (Reprinted with permission from Meuzelaar HLC, Haverkamp J, and Hileman FD (1982) Techniques and instrumentation in analytical chemistry. In: *Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials*, vol. 3, p. 14. Amsterdam: Elsevier; © Elsevier.)

Biological samples such as microbial cultures must be obtained carefully using disposable plastic wires or picks, and smeared on the Curie-point substrate. If the sample is from a liquid culture, a small amount of liquid is applied directly to the substrate. In each case, the samples are dried by heating in an oven or by vacuum desiccation. The samples are loaded into the instrument sequentially and the pyrolysate is

generated in a vacuum. A radio-frequency current is passed through a pyrolysis coil, which surrounds the Curie-point substrate (**Figure 2**). The substrate reaches its Curie-point temperature within 0.2–0.3 s, and the pyrolysate expands into a gold-plated chamber heated to 150°C, from where it diffuses to the ionization chamber of the mass spectrometer. A common approach with biological samples is to use alloy foils made of equal parts iron and nickel, which have a Curie-point temperature of 530°C. This has been shown to provide a good balance between the fragmentation of carbohydrates and proteins. For other samples, alternative Curie-point substrates can be used. For example, iron/nickel alloys with different ratios of the two metals can be used to achieve Curie-point temperatures of, e.g., 350, 610, and 770°C.

In the mass spectrometer, the pyrolysate is bombarded with low-energy electrons (~25 eV) producing molecular and fragment ions, most of which carry a single charge. The ions are focused and accelerated toward the quadrupole analyzers, where they are separated on the basis of their mass to charge ( $m/z$ ) ratios to generate the mass spectrum. Neutral molecules and fragments are trapped by the

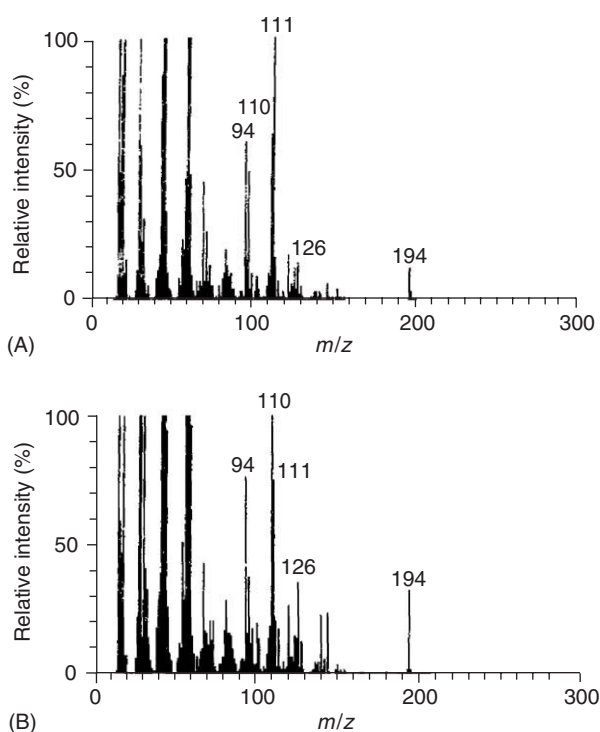
nitrogen-cooled screen and are removed from the instrument.

## Data Analysis in PyMS

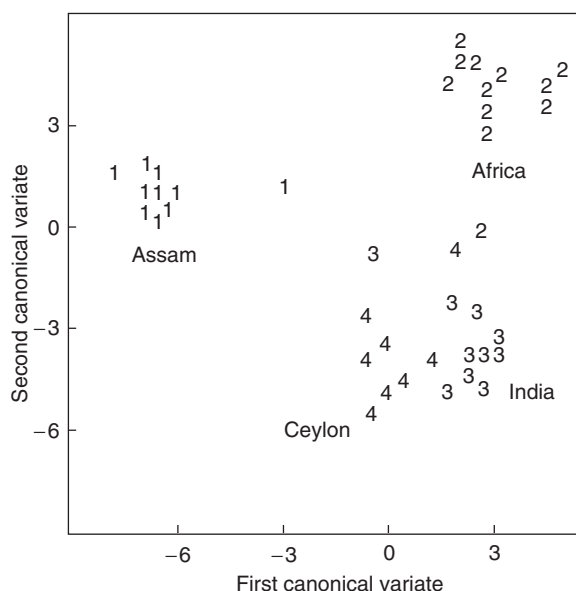
### Multivariate Analysis

The major application of PyMS is sample profiling to identify diagnostic fingerprints of particular analytes. In analytical chemistry and geochemistry, where the fingerprints of specific organic molecules are relatively easy to identify, standards can be used as reference profiles to facilitate sample matching to known compounds. In the case of more complex biological samples, where PyMS has been widely used to identify metabolites, adulterants or different microbial species in environmental, biological, and industrial samples, principal component analysis (PCA) is often used to simplify the multivariate data and identify key differences and similarities between samples. The canonical variates analysis (CVA) method is very similar, separating the samples into different categories based on the principal components and some *a priori* knowledge of the appropriate number of groupings. As an example, **Figure 3** shows pyrolysis mass spectra for Ceylonese and African teas, the inversion of the intensities at masses 110 and 111 being one of the most obvious differences between them. **Figure 4** shows the results of CVA procedure, i.e., a map obtained by plotting the first canonical variate against the second, resulting in clustering that is self-evident. For microbial specimens, clustering methods may be used to map the relationships between different samples. The first step is to produce a similarity matrix, which shows the percentage similarity between all possible pairs of samples, and then this is converted into a dendrogram in which the most closely related samples are linked together through the use of hierarchical clustering algorithms. If reference strains are included in this analysis, then unknown isolates can be compared to a variety of known genera and species to increase the likelihood of identification.

In a clinical setting, the multivariate analysis of pyrolysis mass spectra is widely used to identify clinical isolates in a manner that is much more rapid, reliable, and convenient than conventional typing systems. Conventional typing generally involves a battery of specialized culture-based tests to identify the species and determine antibiotic sensitivity. This process is not always specific because a limited range of properties is tested, such as particular markers, cell-surface proteins or metabolic capabilities. It can take more than 3 days to generate sufficient information to enable a rational choice of antibiotic



**Figure 3** Pyrolysis mass spectra of extracts derived from Ceylonese and African teas: (A) Tea Kep (Africa) and (B) Tea Dya (Ceylon). (Reprinted with permission from Vorhees KJ (ed.) (1994) *Analytical Pyrolysis: Techniques and Applications*. London: Butterworths; © Elsevier.)



**Figure 4** Canonical variates analysis of tea extract data. (Reprinted with permission from Vorhees KJ (ed.) (1994) *Analytical Pyrolysis: Techniques and Applications*. London: Butterworths; © Elsevier.)

therapy for a seriously ill patient. PyMS is advantageous in this respect because the entire cell is used to derive the fingerprint, which potentially provides

much more useful information than tests designed to identify individual biological features. Furthermore, while conventional tests for different microbes differ widely in their methodology and the amount of expertise required to perform them, PyMS is a single technique that can be applied to all microbial isolates. The same advantages apply to the identification of microbes in environmental samples or spoiled food, and the identification of novel bacterial strains producing potentially therapeutic compounds. In many ways, PyMS is complementary to the use of DNA analysis for taxonomic assignment, since the latter is also a simple and universal method although based on the analysis of genotype rather than phenotype. Where the two methods have been compared there is often close agreement in the relationships that are identified. For example, studies of *Bradyrhizobium japonicum* and *Clostridium acetobutylicum* isolates using PyMS and 16S rRNA profiling have shown very similar clustering hierarchies.

### Supervised Learning

PCA and other multivariate methods achieve the clustering of related data points so that samples can be grouped together according to similarity. The overall purpose of this type of analysis is to distinguish individual components of the dataset, which means it is difficult or impossible to generate quantitative data. Quantitative interpretations of pyrolysis mass spectra can be achieved using a different approach in which the multivariate spectra are subjected to various types of regression analysis. Several techniques have been applied, including multiple linear regression (MLR), partial least squares regression (PLS) and principal components regression (PCR). These techniques are often grouped under the term 'supervised learning', because the algorithms must be trained using calibration data (spectral standards and training sets), which allow them to predict results from new spectra. Similar methods have been developed using artificial neural networks, which have also been employed to correct for instrumental drift – a phenomenon that reduces long-term reproducibility and makes it difficult to compare samples that have been collected more than a few days apart.

The first use of neural networks for the interpretation of PyMS spectra was reported by Goodacre (see Further Reading) for the identification of adulterated olive oils. The neural network was trained with standard pyrolysis mass spectra from 12 virgin and 12 adulterated olive oils, and then tested against a bank of further samples. The method successfully discriminated between virgin oils and those which had been adulterated with other seed oils, while PCA

and other multivariate techniques failed to do so. Since this initial report, the combination of PyMS and supervised learning methods has been widely used both in the identification of contaminants and adulterants, and in the analysis of microbial isolates (see later).

### Applications of PyMS

PyMS is an extremely versatile analytical method, which has been applied in a wide range of areas including organic and inorganic chemistry, geochemistry, environmental biology, archaeology, forensic science, food analysis, clinical biology, and biotechnology. In geochemistry, PyMS has been used for the analysis of coals, shales, peats, and sediments, but perhaps the most interesting application in this area is in oil exploration, where its ability to detect and characterize volatile hydrocarbons in small pieces of rock, such as drill chips, has greatly enhanced the analysis of cored samples. In archaeology, PyMS has been used to determine how ancient artifacts were used. For example, the examination of pottery fragments and other such items has revealed the presence of sealed-in volatile compounds which allowed the original use of the vessel from which the fragments came to be determined, e.g., wine jugs, food utensils, and oil containers. In another example, PyMS was used to determine the likely purpose of a 20 000-year-old Neolithic Egyptian grinding stone, which had been buried in the sand since the dawn of history.

Many workers have used PyMS to study the structures of polymers, both natural and artificial. Understanding the performance of polymers in terms of cohesion and substrate adhesion is of immense commercial significance in the paint and adhesive industries. Similarly, the behavior of polymers under stress and when exposed to external factors such as ultraviolet light has been extensively studied by PyMS and is useful in the development of novel materials that have desirable properties, e.g., fire-retardant coatings and biodegradable fibers. There is much interest in polyhydroxyalkanoates as potentially biodegradable plastics, and PyMS has been a principal method used to study thermal degradation profiles of this material. Similarly, in forensic science, PyMS has been used to analyze fibers and to help match samples of automotive finishes to paint chips found at crime scenes.

In environmental science, PyMS is used for the identification of microbial isolates (see below) and to study the origins and impact of environmental pollutants. For example, carbon-rich material condensing on airborne particulates adjacent to roads can be positively linked to emissions from traffic plus a



small contribution from the degradation of tires. Samples taken from lake and river sediments, or from the inlet and outlet systems of sewerage plants, can be used to judge the efficiency of sewerage treatment and pollution limitation in factory effluents. PyMS has also been widely applied in the food industry, particularly in quality control and the detection of adulteration. **Figures 3 and 4** show how PyMS can be used to identify the sources of teas, and in similar fashion the technique has been used to authenticate different types of whiskey, olive oil, orange juice, oysters, cheese, and honey. Indeed, PyMS is an ideal method for quality control in a number of bioprocesses including fermentation, metabolic engineering, and the production of recombinant proteins by cultured cells.

In this context, the applications of PyMS in biotechnology are of particular interest. The technique has been used to detect and measure the level of metabolites produced in cell culture, such as indole produced by cultured *Escherichia coli* cells grown in different types of media, and to predict the yield of recombinant protein in culture systems, such as *E. coli* producing recombinant cytochrome B5 or  $\alpha$ -interferon. An analogous application is the use of PyMS to profile drugs and their metabolic breakdown products in serum (e.g., cyclosporin, vancomycin), or to profile microbial isolates for novel drug-producing strains.

However, the major application of PyMS in many areas of biology remains the rapid identification of particular microbes at the genus, species, or subspecies levels. This approach has been used to analyze environmental specimens, clinical specimens, and food samples for the rapid identification of pathogens, spoilage organisms, and the characterization of ecological diversity. Interesting examples include the successful identification of coffee rust spores responsible for the large-scale destruction of coffee crops, the characterization of algal diversity in estuarine environments, the analysis of cell-wall mutants of the yeast *Saccharomyces cerevisiae*, the predication of bacterial populations in different types of yogurt,

and the identification of novel actinobacters in sewerage sludge.

The myriad of applications in diverse fields, coupled with the continuing advances in instrumentation, sensitivity, quantitative analysis, long-term reproducibility, and data-handling methods, will ensure that PyMS plays a crucial role in many analytical settings in the future.

*See also:* **Mass Spectrometry:** Multidimensional.

## Further Reading

- Barshick SA, Wolf DA, and Vass AA (1999) Differentiation of microorganisms based on pyrolysis ion trap mass spectrometry using chemical ionization. *Analytical Chemistry* 71: 633–641.
- Basile F, Beverly MB, Vorhees KJ, and Hadfield TL (1998) Pathogenic bacteria: Their detection and differentiation by rapid lipid profiling with pyrolysis mass spectrometry. *Trends in Analytical Chemistry* 17: 95–109.
- Goodacre R and Kell DB (1996) Pyrolysis mass spectrometry and its applications in biotechnology. *Current Opinion in Biotechnology* 7: 20–28.
- Goodfellow M, Freeman R, and Sisson PR (1997) Curie-point pyrolysis mass spectrometry as a tool in clinical microbiology. *Zentralblatt fur Bakteriologie* 285: 133–156.
- Guillou C, Lipp M, Radovic B, *et al.* (1999) Use of pyrolysis-mass spectrometry in food analysis: Applications in the food analysis laboratory of the European Commissions' Joint Research Centre. *Journal of Analytical and Applied Pyrolysis* 49: 329–335.
- Irwin WJ (1982) *Analytical Pyrolysis: A Comprehensive Guide*. New York: Dekker.
- Lattimer RP (2003) Pyrolysis mass spectrometry of acrylic acid polymers. *Journal of Analytical and Applied Pyrolysis* 68–69: 3–14.
- Schnitzer M and Schulten HR (1995) Analysis of organic matter in soil extracts and whole soils by pyrolysis mass spectrometry. *Advances in Agronomy* 55: 167–217.
- Tsuge S (1995) Analytical pyrolysis – past, present and future. *Journal of Analytical and Applied Pyrolysis* 32: 1–6.

## Archaeological Applications

**T C O'Connell and A W G Pike**, University of Oxford, Oxford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Z A Stos-Gale and N Gale, pp. 2939–2944, © 1995, Elsevier Ltd.

## Introduction

One of the most influential developments in archaeology in the latter part of the twentieth century has been the application of sophisticated analytical

small contribution from the degradation of tires. Samples taken from lake and river sediments, or from the inlet and outlet systems of sewerage plants, can be used to judge the efficiency of sewerage treatment and pollution limitation in factory effluents. PyMS has also been widely applied in the food industry, particularly in quality control and the detection of adulteration. Figures 3 and 4 show how PyMS can be used to identify the sources of teas, and in similar fashion the technique has been used to authenticate different types of whiskey, olive oil, orange juice, oysters, cheese, and honey. Indeed, PyMS is an ideal method for quality control in a number of bioprocesses including fermentation, metabolic engineering, and the production of recombinant proteins by cultured cells.

In this context, the applications of PyMS in biotechnology are of particular interest. The technique has been used to detect and measure the level of metabolites produced in cell culture, such as indole produced by cultured *Escherichia coli* cells grown in different types of media, and to predict the yield of recombinant protein in culture systems, such as *E. coli* producing recombinant cytochrome B5 or  $\alpha$ -interferon. An analogous application is the use of PyMS to profile drugs and their metabolic breakdown products in serum (e.g., cyclosporin, vancomycin), or to profile microbial isolates for novel drug-producing strains.

However, the major application of PyMS in many areas of biology remains the rapid identification of particular microbes at the genus, species, or subspecies levels. This approach has been used to analyze environmental specimens, clinical specimens, and food samples for the rapid identification of pathogens, spoilage organisms, and the characterization of ecological diversity. Interesting examples include the successful identification of coffee rust spores responsible for the large-scale destruction of coffee crops, the characterization of algal diversity in estuarine environments, the analysis of cell-wall mutants of the yeast *Saccharomyces cerevisiae*, the predication of bacterial populations in different types of yogurt,

and the identification of novel actinobacters in sewerage sludge.

The myriad of applications in diverse fields, coupled with the continuing advances in instrumentation, sensitivity, quantitative analysis, long-term reproducibility, and data-handling methods, will ensure that PyMS plays a crucial role in many analytical settings in the future.

*See also:* **Mass Spectrometry:** Multidimensional.

## Further Reading

- Barshick SA, Wolf DA, and Vass AA (1999) Differentiation of microorganisms based on pyrolysis ion trap mass spectrometry using chemical ionization. *Analytical Chemistry* 71: 633–641.
- Basile F, Beverly MB, Vorhees KJ, and Hadfield TL (1998) Pathogenic bacteria: Their detection and differentiation by rapid lipid profiling with pyrolysis mass spectrometry. *Trends in Analytical Chemistry* 17: 95–109.
- Goodacre R and Kell DB (1996) Pyrolysis mass spectrometry and its applications in biotechnology. *Current Opinion in Biotechnology* 7: 20–28.
- Goodfellow M, Freeman R, and Sisson PR (1997) Curie-point pyrolysis mass spectrometry as a tool in clinical microbiology. *Zentralblatt fur Bakteriologie* 285: 133–156.
- Guillou C, Lipp M, Radovic B, *et al.* (1999) Use of pyrolysis-mass spectrometry in food analysis: Applications in the food analysis laboratory of the European Commissions' Joint Research Centre. *Journal of Analytical and Applied Pyrolysis* 49: 329–335.
- Irwin WJ (1982) *Analytical Pyrolysis: A Comprehensive Guide*. New York: Dekker.
- Lattimer RP (2003) Pyrolysis mass spectrometry of acrylic acid polymers. *Journal of Analytical and Applied Pyrolysis* 68–69: 3–14.
- Schnitzer M and Schulten HR (1995) Analysis of organic matter in soil extracts and whole soils by pyrolysis mass spectrometry. *Advances in Agronomy* 55: 167–217.
- Tsuge S (1995) Analytical pyrolysis – past, present and future. *Journal of Analytical and Applied Pyrolysis* 32: 1–6.

## Archaeological Applications

**T C O'Connell and A W G Pike**, University of Oxford, Oxford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Z A Stos-Gale and N Gale, pp. 2939–2944, © 1995, Elsevier Ltd.

## Introduction

One of the most influential developments in archaeology in the latter part of the twentieth century has been the application of sophisticated analytical

techniques to excavated archaeological material, greatly increasing the range of information that can be derived from archaeological remains at the post-excavational stage. This has enabled archaeologists and archaeological scientists to probe past societies and cultures in greater detail, examining social and economic structures, diet and subsistence, technological development, trade contacts, and sources of wealth. The archaeologically valuable information that can be derived from ancient remains by means of mass spectrometry (MS) falls into three broad categories: provenance, biomolecular studies, and dating.

## Provenance Studies

Archaeological provenance studies – the tracing of artifacts to the source of their raw materials and production – plays an important role in understanding ancient systems of trade and exchange, and the organization of past societies. The quest for provenance is to find a characteristic ‘fingerprint’ unique to a particular raw material source, and to identify that same fingerprint in finished artifacts. Artifacts such as stone axes and coarse pottery can be characterized by optical microscopy, but more homogenous materials such as flint are generally distinguished by trace element composition. Although much of the groundwork for trace element characterization of materials was done prior to its invention, the multielement capability of plasma mass spectrometers has proved useful for the provenance of a range of materials. In some cases, however, the trace element composition of an artifact is changed in its manufacture, or the material is simply too homogenous (or conversely a source too variable) to characterize by trace elements alone. In these cases isotopic analyses can often provide the required ‘fingerprint’.

### Provenance of Marble

The provenance of marble is one such case. Early attempts at provenance using trace elements found that trace element concentrations varied by factors of over 100 within the same quarry. However, the stable isotopic ratios of oxygen and carbon have been used to discriminate between marble sources. These isotopic ratios are fractionated during initial formation of the protolith. For example, the  $\delta^{18}\text{O}$  will vary with the source of the carbonate (a chemical precipitate, or of biogenic origin), and will be further fractionated during metamorphism.  $\delta^{18}\text{O}$  and  $\delta^{13}\text{C}$  are measured using isotope ratio MS, and isotopic databases of values for Early Bronze Age Cycladic marble, and Classical marble exist for the Mediterranean region. In cases where the oxygen and carbon isotope values overlap and sources cannot be discriminated,

$^{87}\text{Sr}/^{86}\text{Sr}$  measured using thermal ionization mass spectrometry (TIMS) can often be successfully used to provide an additional signature. Oxygen and carbon isotopes have also been used to match broken fragments of marble inscriptions, to identify later repairs to statues, and identify probable fake marble sculpture.

### Provenance of Metals Using Lead isotopes

The provenance of metal artifacts provides particular problems, since there are several processing stages from the ore to the finished artifact in which the elemental composition will change. Ores are generally processed before smelting, and two or more ores (along with a flux) can be co-smelted. The resulting metal is often remelted and refined, and potentially alloyed with another metal before the final artifact is produced. Each of these steps can alter the composition of the metal to the point where the trace element composition can be significantly different from the initial ore. However, in general, the lead isotopic composition of an ore is transmitted to the finished metal, and even in the case of the mixing of two metals from different sources, the isotopic ratios will change in a predictable fashion.

The ratios of  $^{208}\text{Pb}/^{206}\text{Pb}$ ,  $^{207}\text{Pb}/^{206}\text{Pb}$ , and  $^{206}\text{Pb}/^{204}\text{Pb}$  in an ore are a function of its geochemical history.  $^{206}\text{Pb}$ ,  $^{207}\text{Pb}$ , and  $^{208}\text{Pb}$  are the stable products of the radioactive decay series with parent isotopes  $^{238}\text{U}$ ,  $^{235}\text{U}$ , and  $^{232}\text{Th}$ , and will thus be different for deposits of different ages and different radioactive parent compositions. For copper and lead ores with a simple geochemical history, the isotopic ratios of lead can provide a distinct fingerprint by which artifacts can be attributed to that source. However, some ore fields have complex geochemical histories and widely varying isotopic signatures that may overlap with ore deposits from other regions. However, with supporting archaeological evidence (e.g., to eliminate mines that were not in production at a particular time), and provided the known sources of ore are well characterized, copper, lead, and silver artifacts have been attributed to ore sources with only a small degree of uncertainty.

Perhaps the best characterized are the Cypriot copper ore deposits exploited from the second millennium BC. By the Late Bronze Age (LBA), Cypriot mines were providing the ores for the production of large ‘ox-hide’ copper ingots weighing 25–28 kg, which were traded all over the Mediterranean. TIMS lead isotope studies have traced the majority of LBA ox-hide ingots found in the Mediterranean, including those on the Ulu Burun and Gelidonya shipwrecks, to the Apliki ores of Cyprus.

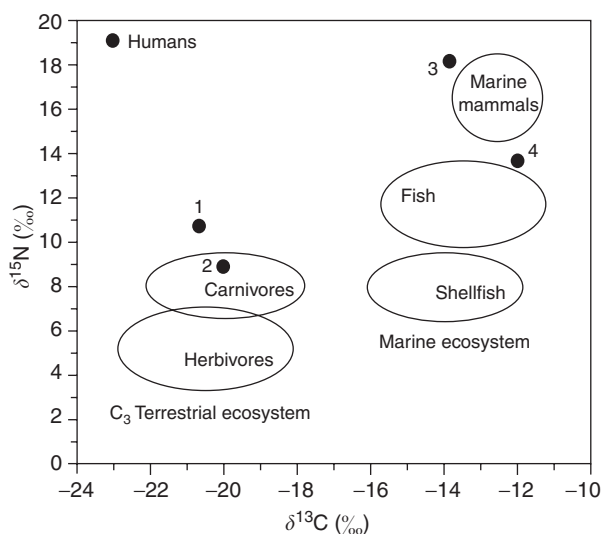
## Biomolecular Studies

The two primary applications of MS to biomolecular archaeology are the study of human diet and life histories using light stable isotopes and trace elements, and the analysis of residues on ceramics and other artifacts.

### Human Diet and Life Histories

Palaeodietary reconstruction using light stable isotopes is based on the principle that 'you are what you eat' – that molecules consumed as food are incorporated into the consumer's body tissues and therefore that a chemical signal passed either unchanged or altered in a quantifiable fashion from food into the body can provide dietary information. The natural variation in the distribution of stable isotopes of carbon, nitrogen, and sulfur throughout different ecosystems makes it possible to use them as natural dietary tracers, whilst oxygen and hydrogen isotopic distributions reflect environmental temperature and water sources, and ratios of radiogenic isotopes such as strontium and lead relate to the local geological environment, and can provide information as to an individual's place of origin.

**Human diet** When interpreted in relation to the isotopic signatures of available food sources, the combination of carbon and nitrogen isotope ratios in archaeological bone provides an objective and direct measure of the diet of an individual (see **Figure 1**). Most of the isotope analyses for palaeodiet are of the bone and tooth protein collagen, since bone is the most abundantly surviving human and animal material on archaeological sites. Collagen also has the significant advantage of being chemically characterizable, independent of the isotope measurement, which allows an assessment of the integrity of the material being analyzed, and affords confidence that the signal being measured is biogenic in origin rather than diagenetic contamination. The carbonate fraction of bone and tooth mineral can also be analyzed, and though prone to contamination in the case of bone, it is strongly resistant to it in the case of tooth enamel. Soft tissues such as hair, skin, and flesh may be analyzed on the rare occasions where they are preserved, usually through mummification (intended or naturally occurring through desiccation) or freezing (e.g., permafrost burials). Each tissue represents a different time-slice of an individual's diet, depending on growth period and turnover rate: bone with its slow turnover integrates long-term diet, while hair provides a linear record of short- to medium-term diet, and teeth record a snapshot of childhood diet consumed during their short formation.



**Figure 1** Diagram to show correlation between carbon and nitrogen isotopic values of collagen from archaeological humans (●), and their expected food sources. Human populations are: 1, Neanderthals; 2, British Neolithic farmers; 3, Native American marine mammal hunters; 4, British Mesolithic hunter-fishers. (Data replotted from: Richards *et al.* (2000) *Proceedings of the National Academy of Sciences of the United States of America* 97: 7663–7666; Richards and Hedges (1999) *Journal of Archaeological Science* 26: 717–722.)

Samples are analyzed by either dual inlet isotope ratio MS, or more commonly nowadays by continuous flow isotope ratio MS. The development of the latter technique has greatly increased the number of palaeodietary studies through the reduced analysis time and the resulting greater number of analyses performed.

An individual's carbon isotope signal indicates the plant types at the base of their ecosystem. When atmospheric carbon dioxide is incorporated through photosynthesis into plants, the heavier isotope is discriminated against, resulting in plants being depleted in  $^{13}\text{C}$  relative to the atmosphere. Most plants photosynthesize using the Calvin–Benson or  $\text{C}_3$  pathway, and such plants typically have  $\delta^{13}\text{C}$  values of  $-27\text{‰}$  compared to atmospheric  $\text{CO}_2$  at  $-8\text{‰}$ . A smaller group of plants use the Hatch–Slack or  $\text{C}_4$  pathway, which results in less discrimination against  $^{13}\text{C}$ , producing plants with  $\delta^{13}\text{C}$  of typically  $-11\text{‰}$ .  $\text{C}_4$  plants are adapted to hot, dry environments, and include the important human food crops of maize, millet, sorghum, and sugar cane, as well as tropical savannah grasses and sedges. The  $\delta^{13}\text{C}$  signature is passed almost unchanged up the food chain, allowing the differentiation of  $\text{C}_3$  consumers from those that eat  $\text{C}_4$ , either directly (eating plants) or indirectly (eating herbivorous animals).

The timing of the introduction of maize agriculture to eastern North America was revealed through a

combination of radiocarbon dating and carbon isotopic analysis: human bone collagen  $\delta^{13}\text{C}$  values indicated a dramatic shift from a  $\text{C}_3$ -based diet, typical of hunter-gatherers, to one rich in  $\text{C}_4$  plants such as maize at around AD 1000, much later than had been predicted.

Hair samples from Nubian burials excavated in modern-day Sudan allow us a glimpse of the agricultural year 2000 years ago: carbon isotopic analyses along each hair strand show a shift from a  $\text{C}_3$  to  $\text{C}_4$  signal within a matter of months, echoing the modern practice of alternating winter  $\text{C}_3$  and summer  $\text{C}_4$  crops in this harsh environment.

Further back in time, carbon isotopes have provided clues as to the diet of early hominids in Africa. Analyses of the tooth enamel of *Australopithecus robustus* from Swartkrans produced a surprising  $\text{C}_4$  signal, as *A. robustus* was assumed to have a  $\text{C}_3$  vegetarian diet of fruits and nuts. Since dental morphology and microwear do not indicate direct grass consumption, *A. robustus* must have incorporated a  $\text{C}_4$  signal either directly by eating  $\text{C}_4$  sedges and their tubers, or indirectly by consuming animals living in the  $\text{C}_4$  savannah grassland environment. But both of these suggest that *A. robustus* utilized a wider range of resources than previously thought.

Marine environments are enriched in  $^{13}\text{C}$  relative to  $\text{C}_3$  terrestrial ecosystems, due to the thermodynamic isotope fractionation in the dissolution of  $\text{CO}_2$  at the ocean/atmosphere interface, and this difference allows us to distinguish whether humans were eating marine or terrestrial foods. In northwest Europe, the transition from hunting, gathering, and fishing in Mesolithic period to horticulture and pastoralism in the Neolithic is indicated by a shift in human bone collagen  $\delta^{13}\text{C}$  values, from those indicating marine resource consumption (ca.  $-14\text{‰}$  to  $-12\text{‰}$ ) to those showing a dependence on terrestrial plants and animals (around  $-19\text{‰}$  to  $-23\text{‰}$ ).

Nitrogen is incorporated from atmospheric  $\text{N}_2$  in the soil into plants. Little isotopic fractionation occurs during this process, but at subsequent steps in the food chain (plant to herbivore, herbivore to carnivore, etc.) there is an observed enrichment in  $^{15}\text{N}$ , resulting in an  $\sim 3\text{--}5\text{‰}$  increase in  $\delta^{15}\text{N}$  with each trophic level. Measurement of  $\delta^{15}\text{N}$  therefore allows an assessment of an individual's position in the food chain, and therefore in humans, an indication of how much animal protein is consumed (either as meat or as secondary products such as milk and cheese). Neanderthal specimens from Belgium and Croatia have been found to have  $\delta^{15}\text{N}$  similar to contemporary large carnivores such as wolf, hyena, and cave lion, providing quantitative support for the belief, based on other archaeological evidence, that

Neanderthals were efficient large-game hunters in Palaeolithic Europe.

Besides answering questions such as what past people ate, and how subsistence changed through time, isotopic analyses can help to investigate differences within populations, related to age, gender, and status. Studies of an Anglo-Saxon population at Berinsfield in Oxfordshire showed that, based on nitrogen isotopic values, poor people seemed to eat more fish than richer individuals (those buried with more grave goods): although initially counterintuitive, as fish is often assumed to be a high-status food, one explanation is the exploitation of free riverine resources by those too poor to own cattle, sheep, or goats.

**Human life histories** By analyzing different skeletal elements, formed at different points in an individual's life, we can gain an insight into diet and residence at various points in life.

The major dietary shift that all humans undergo early in life is the transition from breastfeeding to solid foods, and the timing of this weaning has important demographic connotations. As breastfeeding infants are a trophic level above their mothers, a decrease in infant  $\delta^{15}\text{N}$  values is observed when breast milk is replaced by solid foods, paralleled by a change in  $\delta^{18}\text{O}$  related to the replacement of water derived from breast milk by surface water. A study of teeth from the Guatemalan site of Kaminaljuyú, dating from 700 BC to AD 1500 showed that solid foods were introduced earlier than 2 years of age, but breast milk was a major source of water for at least a further year.

A combination of trace element analyses such as strontium with stable carbon and nitrogen isotope analyses can show whether individuals moved geographically, as well as changing their diet. Isotopes were measured on bones and teeth of the  $\sim 5300$ -year-old Neolithic human mummy preserved in an alpine Glacier in northern Italy in an attempt to reconstruct aspects of his life. Radiogenic Sr, Nd, and Pb isotopes passed into the body from food relate to the local geological environment, and the stable isotopes of O and C provide information about altitude and position relative to the main Alpine watershed. Differences in the  $^{87}\text{Sr}/^{86}\text{Sr}$  of his tooth enamel (formed in early childhood) and bone (probably representing the final decade of life) suggest that as an adult he had moved to a different location from that of his childhood, whilst the  $\delta^{18}\text{O}$  data suggest the Iceman had spent his childhood at lower altitudes than his burial site.

Using a similar methodology but at a larger scale, trace element analyses can be used to look at the movement of people around the landscape. A comparison of strontium isotope analyses of teeth



and bone from a number of individuals from a Bell Beaker population in Germany showed that ~20–25% of the population were incomers to the region in which they were buried, as their teeth and bone  $^{87}\text{Sr}/^{86}\text{Sr}$  were significantly different. Based on the gender distribution, females appeared to be more likely to have moved, suggesting a possible pattern of exogamy.

**Animals** Palaeodietary reconstruction usually focuses on humans, but in certain situations, the study of animal remains can provide information on human activities such as herd management and transhumance. Studies of carbon, oxygen, and strontium isotopes in sheep teeth from Late Stone Age Kasteelberg in Southern Africa indicate that there was little seasonal movement of animals from the coast to inland, suggesting that the transhumance practiced in that region by the historical Khoekhoe peoples developed only later in prehistory, when cattle pastoralism began.

### Residue Analysis

The ability to analyze organic residues in and on archaeological artifacts such as ceramics and stone tools is a result of the development of compound-specific analytical techniques such as gas chromatography (GC), gas chromatography–mass spectrometry (GC–MS), and more recently gas chromatography–combustion–isotope ratio mass spectrometry (GC–C–IRMS). These techniques enable individual compounds, generally lipids, to be chemically characterized from amorphous and often degraded residues left on artifacts by human activities. From such ‘biomarker’ molecules, the original material can then be identified. Lipids have been shown to be well preserved when adsorbed within an unglazed ceramic matrix, and a variety of different compounds have been measured including fatty acids, waxes, and sterols of both animal and plant origin. Studies on stone tools have been less successful, due to problems of diagenesis and lack of reproducibility.

Whereas palaeodietary reconstruction using light stable isotope and trace element analysis is generally focused on the consumer, giving an idea of overall diet, analysis of organic remains from the preparing, cooking, and storing of food provides an insight into specific foodstuffs that were consumed. The types of fats and their distribution within pots of different shapes and sizes from the Saxon settlement of Raunds in Northamptonshire showed that form was correlated with function: large shallow dishes had a high density of animal fats on their bases, suggesting they were placed under roasting meats to collect the dripping fat, whereas taller ‘top hat’ jars had the distinctive leaf waxes derived from cabbage leaves

adsorbed around their rims, suggesting that they were vegetable boiling pots.

Pots dating to 6000 BC from the Early Neolithic site of Çatal Höyük in Turkey were found to have contained a mixture of ruminant and pig fats, but no dairy fats, based on GC–C–IRMS analyses of palmitic ( $\text{C}_{16:0}$ ) and stearic ( $\text{C}_{18:0}$ ) fatty acids.

Residue analysis is not limited to the study of food remains, and analysis of nondietary related residues such as waxes, resins, and pitches can give an idea of daily life in prehistory, as well as trade and technology. GC–MS has been used to identify that Apache, Washo, and Paiute peoples in California waterproofed bottles and baskets using pinyon and lac resins. In the Old World, analyses have shown that pine pitches and tars had a wide variety of uses, such as sealing Hellenistic amphorae, repairing Roman jars, and waterproofing ships from pre-Roman times to the Tudor warship ‘Mary Rose’. In Minoan Crete, GC–MS and GC–C–IRMS demonstrated that beeswax was burnt as lamp fuel, whilst in Saxon Britain, lamps were more likely to contain sheep or cow tallow.

### Chronometric Dating

Assessment of an accurate and precise chronology is crucial for resolving many questions in human prehistory, and MS underlies three important dating techniques.

#### Radiocarbon and Other Cosmogenic Isotopes

Radiocarbon provides the most common archaeological dating method, and can be applied back to ~40 ky. Organic remains suitable for dating (e.g., wood, charcoal, bone) are often abundant on archaeological sites. In the past, the  $^{14}\text{C}$  concentration in a sample was determined by counting beta particles emitted when  $^{14}\text{C}$  decays to  $^{14}\text{N}$ . This required a large sample size (1–10 g), and measurement times could be as long as weeks or months. The development of radiocarbon accelerator mass spectrometry (AMS) reduced sample requirements to a few milligrams and measurement times to less than an hour, and has pushed the practical limit of radiocarbon back to beyond 50 ky.

The particular challenge in measuring  $^{14}\text{C}$  by MS stems from its extremely low abundance (1 part in  $10^{12}$ ), in particular relative to  $^{14}\text{N}$  which is isobaric. The problem can be solved using tandem accelerators, where negative ions are produced at the source, accelerated to a stripper where they are converted to positive ions (a transition that excludes most of the  $^{14}\text{N}$ ) for detection. High acceleration voltages (> 1 MeV) ensure molecular interferences are minimal.

The application of accelerators to radiocarbon dating has been deemed a revolution in archaeology. This is partly because of the much larger sample throughput possible, but also because the integrity of the samples can be more closely controlled. For example, a single carbonized seed can be dated, rather than an amalgamation of many tens (which may not be of the same date) required for conventional beta counting. Organic remains from a single growing season (e.g., a seed) are much more representative of the date of a deposit, whereas charcoal may represent wood already several hundreds of years old. The technique is obviously less destructive, allowing AMS to be used in the dating of small samples, including the Turin shroud which was dated to 1260–1390 AD.

Radiocarbon is produced in the upper atmosphere from various interactions with cosmic nucleons, and there are other isotopes produced in similar reactions that are of interest to archaeology. The interaction and spallation of atmospheric nuclei with cosmogenic nucleons generates a flux of high-energy neutrons reaching the earth's surface which can give rise to the *in situ* production of 'cosmogenic isotopes' in materials of archaeological relevance. It has been proposed that the production of stable cosmogenic isotopes such as  $^3\text{He}$  and  $^{21}\text{Ne}$  can be used to calculate the time elapsed since a material was exposed to the neutron flux. If a material has been exposed long enough to produce measurable cosmogenic isotopes, and then reburied to a depth of  $>1.5\text{ m}$ , the unstable cosmogenic isotopes (e.g.,  $^{10}\text{Be}$ ,  $^{26}\text{Al}$ , and  $^{36}\text{Cl}$  with half-lives in the range  $10^5$ – $10^6$  years) may provide chronological information. Production rates for these isotopes are low ( $10$ – $100\text{ atoms g}^{-1}\text{ year}^{-1}$ ), however, requiring lengthy chemical purification and the use of AMS for measurement. To date, archaeological applications have been few, largely giving maximum exposure dates for stone, and by implication maximum dates for human activity such as rock painting or construction of monuments.

### Uranium-Series Dating

Uranium-238 and Uranium-235 are the parent isotopes of decay chains that can be used to provide a chronology back to  $\sim 500\text{ ky}$ . In an old system ( $>500\text{ ky}$ ) a radioactive secular equilibrium is established between the parent  $^{238}\text{U}$  or  $^{235}\text{U}$  and their daughter radioisotopes. Natural processes can disrupt this equilibrium, and the subsequent in-growth of the long-lived daughter isotopes, as equilibrium is re-established, provides the chronological information.

Of particular interest to archaeology are the long-lived daughters of  $^{238}\text{U}$  ( $^{234}\text{U}$  and  $^{230}\text{Th}$ , with half-lives of 245 and 75.7 ky, respectively) and the long-lived  $^{231}\text{Pa}$  daughter (with a half-life of 32.7 ky)

of  $^{235}\text{U}$ . Uranium is relatively soluble, so in aqueous systems, the parent uranium isotopes are separated from their insoluble  $^{230}\text{Th}$  and  $^{231}\text{Pa}$  daughters. Hence, materials such as calcium carbonate precipitates (e.g., stalagmites or flowstones in caves) will initially contain no daughter isotopes. The daughters in-grow at a rate defined by their half-life, and the measurement of  $^{230}\text{Th}/^{238}\text{U}$  and  $^{231}\text{Pa}/^{235}\text{U}$  can be used to calculate the time since precipitation of the material. An additional complication for the  $^{238}\text{U}$  decay series is that  $^{234}\text{U}$ , intermediate to  $^{238}\text{U}$  and  $^{230}\text{Th}$ , is fractionated by geochemical processes, and is rarely in secular equilibrium with  $^{238}\text{U}$ . Hence, both  $^{234}\text{U}/^{238}\text{U}$  and  $^{230}\text{Th}/^{238}\text{U}$  are required to calculate a date.

Conventionally, alpha spectrometry was used to determine the isotopic ratios, but more recently thermal ionization and plasma mass spectrometric methods have been employed, greatly increasing precision, decreasing sample size requirements, and pushing back the older limit of datable samples.

An additional requirement of the method is that the sample remains a 'closed system', that is, there is no further input or loss of uranium after initial formation of the sample. Generally, calcium carbonate deposits remain a closed system and provide the most accurate uranium-series chronologies. However, dating carbonate deposits only indirectly dates the archaeology. In the best scenario, cultural horizons will be interstratified by flowstones or stalagmites. This is the case for some of the Neanderthal bearing layers at the cave site of La Chaise de Vouthon in France, which are sandwiched between flowstones, allowing a date to be measured to between 98 and 122 ky. Where just a single flowstone layer is present, only a minimum or maximum age can be given for the archaeological horizons above or below it, but these ages are still invaluable. Minimum dates of 400 and 510 ky for *Homo erectus* fossils from Zhoukoudian and Nanjing demonstrated a significant gap between the last *H. erectus* and the earliest *H. sapiens*, forcing a re-evaluation of the view that they may have been contemporary in this region.

The co-occurrence of archaeological deposits with datable calcite is unfortunately rare, and because of this, attempts have been made to date more controversial materials (e.g., bones and teeth). Bones and teeth present particular problems for uranium-series dating because they remain open systems. Uranium is taken up from the burial environment, and can also be lost, and a model of this migration is required to calculate a date from the measured isotope ratios. One approach utilizes a model of the diffusion of uranium into a bone that predicts specific distributions of uranium and uranium-series isotopes across

a bone section under particular geochemical conditions. Measurement of these distributions has recently been facilitated by the development of laser ablation plasma mass spectrometry. The distributions are compared with those predicted by the model, and where there is good agreement a uranium-series date can be calculated. The use of laser ablation also has the advantage of being virtually nondestructive (the laser spot size can be as small as 5  $\mu\text{m}$ ), which allows dating of the most valuable fossil specimens.

### Potassium–Argon Dating

Potassium-40 decays to  $^{40}\text{Ar}$  with a half-life of 1250 My. Extreme heating events such as volcanic eruption will drive off the gaseous  $^{40}\text{Ar}$ , allowing use of the  $^{40}\text{K}/^{40}\text{Ar}$  ratio to determine the time elapsed since the heating. In the traditional approach, the potassium content of a sample was measured using techniques such as atomic absorption spectrometry, and the  $^{40}\text{Ar}$  in a high-resolution gas source mass spectrometer. Greater precision, however, can be achieved by irradiating the sample in a neutron flux to convert some of the K to  $^{39}\text{Ar}$  and measuring  $^{40}\text{Ar}/^{39}\text{Ar}$  in the mass spectrometer. Single grains of potassium-rich minerals can be heated with a laser to release the argon allowing detection of older contaminating grains.

Archaeological applications of K–Ar dating require volcanic material to be associated with cultural layers. K–Ar on tuffs within Bed I at the Olduvai Gorge in Tanzania provided a date of  $\sim 1.8$  My for *Homo habilis*, and was also used to calibrate the Olduvai magnetic event. New archaeological applications are being developed with the increasing sensitivity of the technique, which enables younger samples to be dated. For example,  $^{40}\text{Ar}/^{39}\text{Ar}$  was used to date anatomically modern *Homo sapiens* fossils from Herto in Ethiopia to between 154 and 160 ky. According to current morphological classification, these fossils represent the earliest known examples of modern humans.

**See also:** Archaeometry and Antique Analysis: Dating of Artifacts; Metallic and Ceramic Objects. **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Gas Chromatography:** Mass Spectrometry. **Mass Spectrometry:** Time-of-Flight; Stable Isotope Ratio; Clinical Applications; Environmental Applications; Food Applications; Forensic Applications.

### Further Reading

- Aitken MJ (1990) *Science-Based Dating in Archaeology*. Harlow, UK: Longman.
- Aitken MJ, Stringer CB, and Mellars PA (eds.) (1993) *The Origin of Modern Humans and the Impact of Chronometric Dating*. Princeton: Princeton University Press.
- Bourdon B, Henderson GM, Lundstrom CC, and Turner SP (eds.) (2003) Uranium-series geochemistry. *Reviews in Mineralogy & Geochemistry* 52.
- Brothwell DA and Pollard AM (eds.) (2001) *Handbook of Archaeological Sciences*. New York: Wiley.
- Evershed RP, Dudd SN, Charters S, *et al.* (1999) Lipids as carriers of anthropogenic signals from prehistory. *Philosophical Transactions of the Royal Society of London B* 354: 19–31.
- Hertz N (1992) Provenance determination of Neolithic to Classical Mediterranean marbles by stable isotopes. *Archaeometry* 34: 185–194.
- Price TD, Burton JH, and Bentley RA (2002) The characterization of biologically available strontium isotope ratios for the study of prehistoric migration. *Archaeometry* 44: 117–135.
- Schwarcz HP and Schoeninger MJ (1991) Stable isotopic analyses in human nutritional ecology. *Yearbook of Physical Anthropology* 34: 283–321.
- Stos-Gale ZA, Malotis G, Gale NH, and Annetts N (1997) Lead isotope characteristics of the Cyprus copper ore deposits applied to provenance studies of copper oxhide ingots. *Archaeometry* 39: 83–123.
- van der Merwe NJ (1992) Light stable isotopes and the reconstruction of prehistoric diets. In: Pollard AM (ed.) *New Developments in Archaeological Science*. Oxford: Oxford University Press.

## Clinical Applications

N J Haskins, Denmead, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Mass spectrometry (MS) is a powerful tool for the structure characterization of small amounts of

material. Its high sensitivity and rapid response make it ideal for the analysis of trace components obtained from biofluids and tissues, especially when it is combined with a chromatographic separation. Since the 1990s, advances in technology have resulted in the availability of cheaper systems that have the potential to revolutionize the application of MS in clinical analysis. Also, the advance of powerful software

a bone section under particular geochemical conditions. Measurement of these distributions has recently been facilitated by the development of laser ablation plasma mass spectrometry. The distributions are compared with those predicted by the model, and where there is good agreement a uranium-series date can be calculated. The use of laser ablation also has the advantage of being virtually nondestructive (the laser spot size can be as small as 5  $\mu\text{m}$ ), which allows dating of the most valuable fossil specimens.

### Potassium–Argon Dating

Potassium-40 decays to  $^{40}\text{Ar}$  with a half-life of 1250 My. Extreme heating events such as volcanic eruption will drive off the gaseous  $^{40}\text{Ar}$ , allowing use of the  $^{40}\text{K}/^{40}\text{Ar}$  ratio to determine the time elapsed since the heating. In the traditional approach, the potassium content of a sample was measured using techniques such as atomic absorption spectrometry, and the  $^{40}\text{Ar}$  in a high-resolution gas source mass spectrometer. Greater precision, however, can be achieved by irradiating the sample in a neutron flux to convert some of the K to  $^{39}\text{Ar}$  and measuring  $^{40}\text{Ar}/^{39}\text{Ar}$  in the mass spectrometer. Single grains of potassium-rich minerals can be heated with a laser to release the argon allowing detection of older contaminating grains.

Archaeological applications of K–Ar dating require volcanic material to be associated with cultural layers. K–Ar on tuffs within Bed I at the Olduvai Gorge in Tanzania provided a date of  $\sim 1.8$  My for *Homo habilis*, and was also used to calibrate the Olduvai magnetic event. New archaeological applications are being developed with the increasing sensitivity of the technique, which enables younger samples to be dated. For example,  $^{40}\text{Ar}/^{39}\text{Ar}$  was used to date anatomically modern *Homo sapiens* fossils from Herto in Ethiopia to between 154 and 160 ky. According to current morphological classification, these fossils represent the earliest known examples of modern humans.

**See also:** Archaeometry and Antique Analysis: Dating of Artifacts; Metallic and Ceramic Objects. **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Gas Chromatography:** Mass Spectrometry. **Mass Spectrometry:** Time-of-Flight; Stable Isotope Ratio; Clinical Applications; Environmental Applications; Food Applications; Forensic Applications.

### Further Reading

- Aitken MJ (1990) *Science-Based Dating in Archaeology*. Harlow, UK: Longman.
- Aitken MJ, Stringer CB, and Mellars PA (eds.) (1993) *The Origin of Modern Humans and the Impact of Chronometric Dating*. Princeton: Princeton University Press.
- Bourdon B, Henderson GM, Lundstrom CC, and Turner SP (eds.) (2003) Uranium-series geochemistry. *Reviews in Mineralogy & Geochemistry* 52.
- Brothwell DA and Pollard AM (eds.) (2001) *Handbook of Archaeological Sciences*. New York: Wiley.
- Evershed RP, Dudd SN, Charters S, *et al.* (1999) Lipids as carriers of anthropogenic signals from prehistory. *Philosophical Transactions of the Royal Society of London B* 354: 19–31.
- Hertz N (1992) Provenance determination of Neolithic to Classical Mediterranean marbles by stable isotopes. *Archaeometry* 34: 185–194.
- Price TD, Burton JH, and Bentley RA (2002) The characterization of biologically available strontium isotope ratios for the study of prehistoric migration. *Archaeometry* 44: 117–135.
- Schwarcz HP and Schoeninger MJ (1991) Stable isotopic analyses in human nutritional ecology. *Yearbook of Physical Anthropology* 34: 283–321.
- Stos-Gale ZA, Malotis G, Gale NH, and Annetts N (1997) Lead isotope characteristics of the Cyprus copper ore deposits applied to provenance studies of copper oxhide ingots. *Archaeometry* 39: 83–123.
- van der Merwe NJ (1992) Light stable isotopes and the reconstruction of prehistoric diets. In: Pollard AM (ed.) *New Developments in Archaeological Science*. Oxford: Oxford University Press.

## Clinical Applications

N J Haskins, Denmead, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Mass spectrometry (MS) is a powerful tool for the structure characterization of small amounts of

material. Its high sensitivity and rapid response make it ideal for the analysis of trace components obtained from biofluids and tissues, especially when it is combined with a chromatographic separation. Since the 1990s, advances in technology have resulted in the availability of cheaper systems that have the potential to revolutionize the application of MS in clinical analysis. Also, the advance of powerful software

tools to help in the analysis of the large quantities of data produced during a mass spectrometric analysis, with a high level of automation, has made MS more available as a routine analytical tool. It has given rise to new areas of research, especially the so-called proteomics and metabonomics.

Nonroutine use of MS in clinical analysis covers a large number of applications involving the analysis of all kinds of biofluids and tissues for analytes ranging in mass from expired gases to proteins. The emphasis in this article will be on the use of MS in drug metabolism and pharmacokinetic studies, the identification and quantification of endogenous components in both normal and diseased states, and the increasing use of MS in the analysis of biopolymers. Although these techniques have been available for many years the introduction of new technology will make them less costly and easier to apply, and this will be indicated throughout this article.

## Sample Types

All kinds of clinical samples can be analyzed using MS combined with suitable extraction and cleanup procedures. Noninvasive samples include expired air, urine, feces, milk, saliva, and sweat. Invasive procedures yielding samples of blood, bile, cerebrospinal fluid, amniotic fluid, and tissue samples of many types are also used to provide material for mass spectrometric analysis. Samples can be analyzed for endogenous components, for instance, profiling disease types, or for exogenous components and metabolites, for instance, drug metabolism studies. The recent development of high-mass MS combined with powerful information systems, driven by the needs of biomedical research, has brought the analysis of peptides, proteins, and other biopolymers within reach and also allowing investigation of abnormal proteins. The examination of total protein contents of various cell types is undertaken. This technique is called 'proteomics' and will have application in clinical analysis.

## Sample Preparation

The direct analysis of some samples is now possible. Breath analysis has always been a direct technique but the direct analysis of other biological samples such as urine and bile has required the development of new methods for handling such samples. Hence, direct analysis of urine using liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS) procedures is often used in metabolism studies with

either electrospray ionization (ESI) or atmospheric pressure chemical ionization (APCI) sources providing the ionization mechanism. Direct analysis of bile has been described with the use of fast atom bombardment-MS (FAB-MS) for the analysis of bile acids, but FAB-MS has been largely superseded by either ESI- or APCI-MS. The same analysis of bile acids can be carried out with more ease using negative-ion APCI.

Most other relatively low molecular mass analytes, such as steroids, fatty acids, require separation and often derivatization prior to analysis. Various steps that may be used in sample preparation are shown in **Table 1**. This is not meant to be an exhaustive list and many references in the primary literature use specific extraction and derivatizing agents.

Care always has to be taken with sample cleanup because of the complexity of biological matrices. Almost all biological samples contain sodium chloride and this can severely disrupt FAB, ESI, or APCI analyses, for instance. Even if salt is removed by the use of appropriate absorption columns, impurities may be introduced at this stage from poor-quality solvents and absorbents. Thin layer chromatography (TLC) is often used as a separation technique but most commercial plates yield extracts containing a

**Table 1** Examples of the various stages in the extraction and analysis of body fluids and/or tissues by MS

Stage	Procedure
Pretreatment	Blood: centrifuge to yield plasma Tissue: solubilize Tissue: homogenize
Extraction	Solvent partition Solid phase, e.g., Lipidex, etc. Immunoabsorption
Fractionation	Dialysis to remove inorganic salts Chromatography (LC, TLC, column, gel) Ion-exchange chromatography
Derivatization	Ketones: oximes Hydroxyl groups: trimethylsilyl or <i>t</i> -butyldimethylsilyl ethers Carboxylic acids: methyl esters
Chromatography-mass spectrometry	Amines: trifluoroacetates Capillary GLC with EI, CI-MS LC: with thermospray, FAB-MS, ESI-MS, APCI-MS SFC (supercritical fluid chromatography)-EI, CI-MS Capillary electrophoresis and related techniques – nanoflow ESI
Accurate mass measurement systems for use with GLC or LC	Sector instruments Ion cyclotron resonance analyzers Orthogonal time-of-flight analyzers (preferred), single analyzers, and in hybrid systems such as QqT



range of binders, modifiers, etc.; again this is detrimental to most ionization techniques. The act of derivatization also introduces impurities due to partial reactions, formation of by-products, and introduction of other impurities. Indeed, in many steroid analyses relying on a final gas chromatographic-mass spectrometric (GC-MS) analysis of a trimethylsilyl ether/oxime, the act of derivatization introduces so many impurities that additional clean-up is often used before the final GC-MS step.

## Types of Analyte

The types of analyte can vary from expired gases to high relative molecular mass proteins but considerable differences in the type of mass spectrometer used for such analyses means that no single instrument will cover the whole range. **Table 2** lists some typical analytes and the kind of mass spectrometric analyses that would be used to characterize them.

## Applications

### Isotope Ratio Mass Spectrometry

Stable isotopes and isotope ratio mass spectrometry are used extensively in nutritional and metabolic research, especially in pediatric research. For example, in an experiment to determine the glucose oxidation rate, the patient swallows an aqueous solution of glucose. The earliest workers used naturally enriched  $^{13}\text{C}$ -glucose obtained from maize sugar, but more recently the availability of synthetic  $^{13}\text{C}$ -glucose with up to 98% atom enrichment has gained favor. The administered glucose is assumed to enter a single plasma pool from where it is either oxidized to  $\text{CO}_2$ , stored as glycogen or fat, or the carbon enters various body cycles including production of protein. At steady state the amount of enrichment of  $^{13}\text{C}$  in both plasma glucose and in expired air remains constant. Since the amount of glucose entering the

pool is known, the glucose oxidation rate can be determined.

Similar studies are conducted using  $^{13}\text{C}$ -enriched triglycerides to study fatty acid metabolism, again following the expiration of  $^{13}\text{CO}_2$ . Protein turnover has been studied by administering  $^{15}\text{N}$ -enriched amino acids and measuring enrichment in excreted urea.  $^{13}\text{C}$ -leucine has also been used but on this occasion  $^{13}\text{CO}_2$  enrichment in expired air was followed.

In all experiments the mass spectral analysis remains similar. Very small amounts of  $^{13}\text{C}$  are measured, requiring high precision abundance measurements. Classically, the instruments have used fixed detectors for  $m/z$  44 ( $^{12}\text{C}^{16}\text{O}_2$ ),  $m/z$  45 ( $^{13}\text{C}^{16}\text{O}_2$  and  $^{12}\text{C}^{16}\text{O}^{17}\text{O}$ ), and  $m/z$  46 ( $^{12}\text{C}^{16}\text{O}^{18}\text{O}$ ). A standard  $\text{CO}_2$  sample is used to calibrate the detector responses and the analyte is introduced by a batch inlet process. The correction at  $m/z$  45 for  $^{17}\text{O}$  is made from the amount of  $^{18}\text{O}$  present at  $m/z$  46 by assuming that the  $^{17}\text{O}/^{18}\text{O}$  ratio remains constant.

More recently continuous-flow devices have been introduced using a combustion chamber to oxidize the carbon to  $\text{CO}_2$ , followed by scrubbers to remove water and oxygen. This is equivalent to interfacing an automated CHN analyzer to the mass spectrometer. Such devices can be used to oxidize carbon-containing compounds separated by GC, for instance. Measurements of naturally expired  $\text{CO}_2$  can also be made using such apparatus. They do not require combustion and enter the system after the combustion chamber but pass through the scrubbers. This system is usually automated.

A number of disease states and physiological processes have been studied using these techniques; a representative selection is shown in **Table 3**.

### Endogenous Components – Identification

The introduction of GC-MS in the 1960s revolutionized the analysis for endogenous components in blood and urine. Indeed, one of the earliest uses of GC-MS was the analysis of steroid conjugates in

**Table 2** Typical analytes derived from biological systems and analytical methods used

Analyte	Example	Mass spectrometry
Gases	Carbon dioxide, anesthetics	Gas analyzer Isotope ratio MS
Low relative molecular mass (100–600)	Steroids, fatty acids	GC-MS, LC-MS
Polar materials (200–2000)	Sugars, nucleosides, small peptides	LC-MS, FAB-MS
Drugs and metabolites	Glucuronides, sulfates	LC-MS/MS, FAB-MS/MS
Peptides, proteins	Endorphins, albumin, proteomics	Digest followed by LC-ESI-MS/MS, nanoflow-MS/MS, MALDI-ToF-MS

GC-MS, gas chromatography-mass spectrometry; LC-MS, liquid chromatography-mass spectrometry; FAB-MS, fast atom bombardment-mass spectrometry; MS/MS, tandem mass spectrometry; ESI-MS, electrospray ionization-mass spectrometry; MALDI-ToF-MS, matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry.

**Table 3** Some physiological processes and disease states studied by isotope ratio mass spectrometry

Study	Administered substrate	Analysis
Glucose oxidation	Natural $^{13}\text{C}$ glucose	Expired $^{13}\text{CO}_2$
Fatty acid metabolism	L- $^{13}\text{C}$ trioctanoate	Expired $^{13}\text{CO}_2$
Protein metabolism	$^{15}\text{N}$ -glycine	$^{15}\text{NH}_3$ , $^{15}\text{N}$ -urea
Amino acid metabolism	$^{13}\text{C}$ -leucine	Expired and plasma $^{13}\text{CO}_2$
Amino acid metabolism	$^2\text{H}_5$ -phenylalanine	Urine/plasma phenylalanine and tyrosine
<i>Campylobacter pylori</i>	$^{13}\text{C}$ -urea	Expired $^{13}\text{CO}_2$
Hypoglycemia	$^{13}\text{C}$ -glucose, $^2\text{H}_2$ -glucose	Plasma glucose
Zellweger's syndrome	$^{13}\text{C}$ -phytanic acid (PA)	Plasma PA and HO-PA
Maple syrup urine disease	$^2\text{H}_5$ -Phenylalanine	Urine/plasma phenylalanine and tyrosine

blood by Sjövall and co-workers. Since that time the mass spectrometric determination of steroids and other small molecules has continued as a research tool but, probably due to the cost and complexity of the assays, has not become a universally accepted routine clinical diagnostic aid.

The isolation and analysis of steroids from biofluids requires an extraction step, followed by purification and derivatization. Extraction is commonly carried out by passing the biofluid through a solid-phase absorbent, usually a C18 cartridge. Up to 50 ml of urine can be passed through the cartridge. After washing with water, steroids are recovered by elution with methanol. Plasma is treated with solvent to precipitate protein (Shackleton, for example, uses acetone-ethanol 1:1 (v/v) and adds 1–2 ml plasma to 20 ml solvent in a sonic bath). After 10 min the solution is cooled and centrifuged. The pellet is resuspended in further solvent and extracted again. The combined extract is evaporated to dryness and chromatographed to separate conjugated and free steroids using an ion-exchange resin. The conjugated fraction is hydrolyzed using  $\beta$ -glucuronidase (*H. Pomatia*) and then cleaned up again using a second C18 cartridge separation.

Prior to GC-MS analysis, the steroid fractions require derivatization. Ketosteroids are usually converted to oximes using methoxylamine, benzylamine, or hydroxylamine hydrochloride in pyridine. Two epimers of the oxime are formed (*E* and *Z*). Hydroxyl groups are then derivatized as trimethylsilyl (TMS) ethers. After formation of the TMS derivative, the extract is further purified by chromatography on a small column of Lipidex<sup>TM</sup> 5000. An alternative to TMS is the use of *t*-butyldimethylsilyl (*t*-BDMS) ethers, but although conferring desirable mass spectrometric properties, namely an intense  $M - 57$  ion, the *t*-BDMS group is 42 U heavier than TMS and the formation of poly *t*-BDMS ethers may increase the relative molecular mass of the steroid derivative beyond the mass range of small quadrupoles.

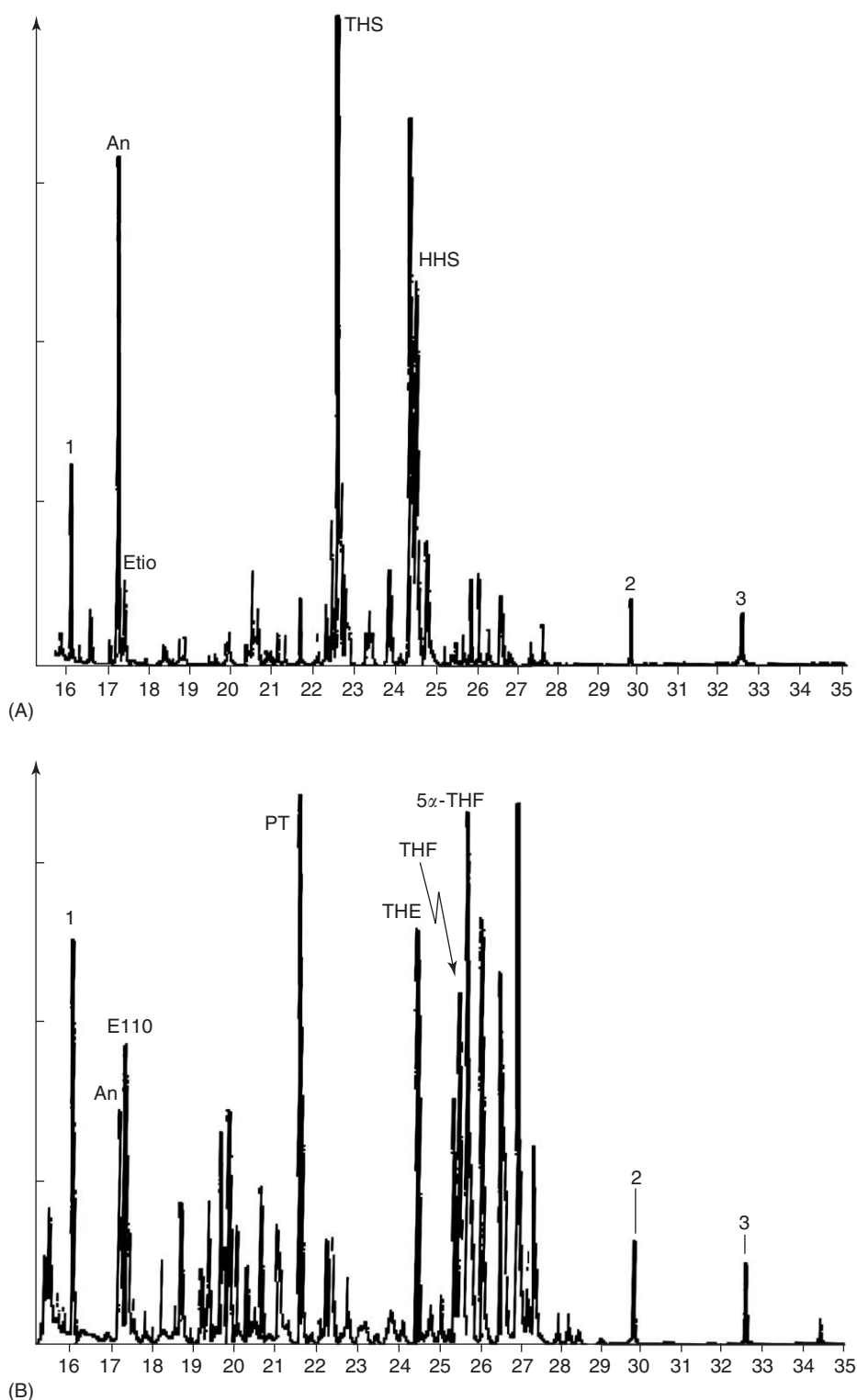
GC-MS profiles obtained from individuals with disorders affecting steroid synthesis may be very different from normal (Figure 1). A range of clinical conditions have been identified, many due to congenital defects affecting cortisol synthesis, which may be analyzed using these methods. Some of these have been tabulated by Shackleton and co-workers (Table 4).

Steroids and similar lipids do not ionize well using ESI or APCI and hence GC-MS remains the favored analytical procedure. The introduction of systems combining a gas chromatograph with an orthogonal axis time-of-flight mass spectrometer (oa-ToF/MS) gives better resolution and the ability to use accurate mass measurement. Thus, an extra parameter is introduced into the analysis, namely the components accurate rather than nominal mass, which improves the identification of the analyte. It also gives important extra data used for the structural characterization of unrecognized steroids or other endogenous components.

A number of other endogenous materials can be analyzed using these techniques, i.e., extraction, derivatization, and GC-MS analysis, to provide a clue to a particular error of metabolism based on the presence or absence of certain components. The scope of such analyses is shown in Table 5.

The increasing application of LC to generate chromatographic profiles enables more-polar compounds such as acids to be analyzed without derivatization. Sensitive LC-MS systems have been developed but reported clinical applications are not yet as numerous as GC-MS. Two LC-MS techniques have been widely accepted for MS applications, namely ESI and APCI interfaces. These have almost completely replaced earlier thermospray and 'flow-FAB' interfaces (Table 6).

APCI ionization is one of the most widely used techniques for LC-MS. It relies on the formation of protonated or cationized molecules in positive-ion mode, or on deprotonated or anionized molecules in the negative-ion mode. However, it is by no means a



**Figure 1** GC-MS Profile analysis. Total ion chromatogram profiles produced by scanning methoxy-trimethylsilyl derivatives of urinary steroids. (A) Profile from a patient with 11 $\beta$ -hydroxylase deficiency in which the principal steroids are metabolites of 11 $\beta$ -deoxycortisol (substance S) such as tetrahydrosubstance S (THS) and hexahydrosubstance S epimers (HHS). (B) Profile from the patient's father; this profile is essentially normal. Some of the major metabolites are C<sub>19</sub> steroids such as androsterone (An), etiocholanone (Etio), pregnanetriol (PT), and cortisol metabolites (THE, tetrahydrocortisone; THF, tetrahydrocortisol and 5 $\alpha$ -THF, 5 $\alpha$ -tetrahydrocortisol). Peaks 1–3 are internal standards. (Reprinted from Shackleton CHL, Merdinck J, and Lawson AM (1990) In: McEwan CN and Larsen BS (eds.) *Mass Spectrometry of Biological Materials*, pp. 297–377. New York: Dekker; courtesy of Marcel Dekker Inc.)

**Table 4** Some disorders affecting steroid synthesis and metabolism

<i>Disease</i>	<i>Cause</i>	<i>Change in urinary steroid profile</i>
Congenital adrenal hyperplasia	21-Hydroxylase deficiency	Metabolites of cortisol grossly reduced; 21-deoxysteroids (17-hydroxypregnanolone, pregnanetriol and pregnanetriolone) and androgen metabolites highly elevated
	11 $\beta$ -Hydroxylase deficiency	Metabolites of cortisol greatly reduced; 11-deoxysteroids (e.g., tetrahydrosubstances S) and 11-deoxyandrogens elevated
	3 $\beta$ -Hydroxysteroid dehydrogenase-isomerase defect	Metabolites of cortisol greatly reduced; 3 $\beta$ -hydroxy-5-ene steroid sulfates markedly increased
	17 $\alpha$ -Hydroxylase deficiency	Metabolites of cortisol and androgens greatly reduced. Dominant urinary steroids are metabolites of corticosterone (e.g., tetrahydrocorticosterone and 5 $\alpha$ -tetrahydrocorticosterone)
Mineralocorticoid (or apparent mineralocorticoid) excess syndromes	11-Hydroxysteroid dehydrogenase deficiency (cortisol oxidase deficiency)	Relatively low cortisol metabolite excretion. Tetrahydrocortisone greatly reduced relative to tetrahydrocortisol and 5 $\alpha$ -tetrahydrocortisol
	Adrenal adenoma	High excretion of tetrahydroaldosterone and/or 18-hydroxycortisol
	Dexamethasone-suppressible hyperaldosteronism	High excretion of 18-hydroxycortisol and tetrahydroaldosterone
	Aldosteronism (elevated aldosterone production)	High excretion of tetrahydroaldosterone and 18-hydroxytetrahydroaldosterone
Hypercortisolism	Excess adrenal cortisol secretion (Cushing's disease)	High excretion of all cortisol metabolites
	Cortisone reductase deficiency	Elevated excretion of tetrahydrocortisone, cortolones, and androgen metabolites. Extremely low excretion of tetrahydrocortisol, 5 $\alpha$ -tetrahydrocortisol, and cortols
Adrenal tumors		Various patterns of elevated steroid excretion, e.g., high 3 $\beta$ -hydroxy-5-ene steroid excretion or high 11 $\beta$ -hydroxyandrosterone excretion
Gonadal tumors		Various types, e.g., ovarian with high excretion of pregnanetriol and 17-hydroxypregnanolone or testicular with high excretion of testosterone and metabolites and 3 $\beta$ -hydroxy-5-ene steroids
Low pregnancy estrogen production	Placental sulfatase deficiency	High excretion of 3 $\beta$ -hydroxy-5-ene steroids in pregnancy urine. Low urinary estriol excretion
Deficient androgen production	5 $\alpha$ -Reductase deficiency	Low excretion of 5 $\alpha$ -reduced steroids. Diagnosed by high 5 $\beta$ /5 $\alpha$ urinary metabolite ratios, e.g., tetrahydrocortisol/5 $\alpha$ -tetrahydrocortisol
	17 $\beta$ -Hydroxysteroid dehydrogenase deficiency	High excretion of androsterone and etiocholanolone
	Testicular feminization	High excretion of androsterone and etiocholanolone
Mineralocorticoid deficiency syndrome	Pseudohypoaldosteronism (renal tubules unresponsive to aldosterone)	High tetrahydroaldosterone and 18-hydroxytetrahydroaldosterone
	Hypoaldosteronism (corticosterone $\rightarrow$ aldosterone biosynthetic defect)	Low tetrahydroaldosterone, 5 $\alpha$ -tetrahydrocortisone and 18-hydroxytetrahydroaldosterone excretion

**Table 5** Various endogenous components that can be analyzed by extraction, derivatization, and GC-MS

Type of analyte	Derivatives	Applications
Steroids	Oximes, TMS, <i>t</i> -BDMS ethers	Steroid profiles
Bile acids	Methyl esters, TMS ethers	Liver disease, peroxisomal deficiencies
Fatty acids	Methyl esters, TMS, <i>t</i> -BDMS esters	Phospholipids, triglycerides
Urinary acids	Methyl esters, TMS, <i>t</i> -BDMS esters	Acidurias, diabetes
Ketones	Oximes	Diabetes
Prostaglandins	Methyl esters, oximes, TMS, ethers	Healing processes, pain

**Table 6** Characteristics of LC-MS methods used in biological mass spectrometric analyses

Interface	Liquid flow	Ionization technique	Applications
Thermospray	<2 ml min <sup>-1</sup>	Thermospray or discharge assisted thermospray	Steroid sulfates, bile acids
Atmospheric pressure chemical ionization (APCI)	10–1000 µl min <sup>-1</sup>	APCI, positive or negative ion	Conjugates, peptides, small polar molecules
Electrospray	10–500 µl min <sup>-1</sup>	Electrospray	Conjugates, peptides, proteins
Nano-electrospray	<10 µl min <sup>-1</sup>	Electrospray	Conjugates, peptides, proteins with limited amounts

universal source and thermal degradation can be a considerable problem for sensitive molecules (such as polyhydroxylated steroids, which may lose water). Additionally, although many compounds will ionize under a variety of probe temperature and solvent conditions, some have little tolerance to temperature change and only give a useful signal within a narrow temperature window.

Electrospray appears to be useful especially when dealing with compounds that give poor APCI spectra due to thermal degradation (such as glucuronide conjugates). Meanwhile, electrospray and its derivative, 'nanoflow electrospray', have developed a niche in the analysis of proteins. They are especially useful as an ionization method combined with capillary electrophoresis, and for offline analysis of spots retrieved from chromatographic plates.

LC-MS analyses using ESI or APCI have tended to be used for those compounds that are not amenable to GC-MS, such as intact conjugates (especially sulfates and phosphates) and quaternary ammonium salts (phosphatidylcholines, acetylcholine).

Other LC-MS ionization techniques are available but their utility for clinical research is low. Until the advent of ESI and APCI the most popular LC-MS interface was thermospray. Other techniques used with varying degrees of success included flow-FAB, transporter belts, and particle beam interfaces. These have been almost wholly superseded by ESI and APCI.

All these chromatographic mass spectrometric techniques are time consuming because a separation has to be made before the sample can be ionized. Direct analysis of biological samples by MS without

**Table 7** Deprotonated molecular anions observed for bile acids in FAB-MS and APCI-MS

Mass	Assignment	Example
432	Glycomonohydroxycholates	Glycolithocholic acid
448	Glycodihydroxycholates	Glycodeoxycholic acid
464	Glycotrihydroxycholates	Glycocholic acid
482	Tauromonohydroxycholates	Taurolithocholic acid
498	Taurodihydroxycholates	Taurochenodeoxycholic acid
514	Taurotrihydroxycholate	Taurocholic acid

prior separation is very restricted because of the complexity of the samples but some useful work has been reported using FAB-MS for the analysis of bile acids in urine or bile. After passing the biological fluid through a C18 cartridge to desalt it, spectra are obtained by dissolving the mixture in methanol, and adding to glycerol. Negative ions are generated and a partial spectrum over the mass range 400–600 U will show the major conjugated bile acids (Table 7). The introduction of electrospray and nanoflow electrospray means these analyses can be carried out on small samples using a syringe pump to supply the extract to the electrospray needle. It has been shown that changes in the pattern of bile acids secreted can give indications of peroxisomal dysfunction and other liver disease. Thus, direct analysis can be used as a rapid screen. Samples showing abnormalities can then be more rigorously analyzed using LC-MS or GC-MS.

### Drug Metabolism Studies – Identification

All new chemical entities (NCEs) proposed as pharmaceuticals require extensive testing before they can



be introduced to the clinic. The study of metabolism is a major requirement in order to identify any metabolites that may have pharmacological, toxicological, or other properties affecting the efficacy, safety, and use of the drug. MS has played a unique role in the development of metabolism studies and is an essential tool in the development of any NCE. This is in contrast to the slow acceptance of MS in clinical diagnosis, although much of the technology and methodology is common to both.

Traditionally, the identification of metabolites has relied on extraction of biofluids, derivatization, and either LC-MS or GC-MS analysis of the final sample. However, refinement of the methodology has led to the adoption of isotopic labeling and MS/MS techniques as the current state of the art.

In a typical study the compound is administered with a trace level of radioisotope ( $^{14}\text{C}$  or  $^3\text{H}$  normally) and preferably labeled with a moderate amount of stable isotopes ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{18}\text{O}$ ,  $^2\text{H}$ ) to generate a characteristic molecular ion pattern in the mass spectrometer. After collecting biofluids such as plasma, urine, and feces, extraction may be carried out using the radiolabel as a tracer to follow drug-derived material. The extracts, or often the biofluid itself, are subjected to LC separation with radiation, UV, and mass spectrometric detection. This allows any components that contain drug-derived material to be readily identified. Significant ions can be subjected to daughter or precursor ion analysis using MS/MS techniques to confirm the relative molecular mass and often to identify the structure of the metabolite even at very low concentrations.

An example of the use of this approach is the study of a sample of horse urine thought to contain furosemide (frusemide). The LC-UV chromatogram at 254 nm showed no clear evidence for the presence of furosemide compared to a control urine sample. LC-MS gave a total ion chromatogram, which again failed to show a clear chromatogram peak for furosemide. However, selected ion monitoring (SIM) of mass 329 u (the deprotonated molecule from furosemide) showed a peak with the expected retention time for furosemide. The analysis was confirmed by a LC-MS/MS experiment. The peak was shown to give a daughter spectrum similar to that obtained from furosemide.

The combination of stable isotopes and MS gives a powerful tool with extra features for the study of drug metabolism. For example, unlike natural products, most synthetic drugs are racemic mixtures. Administering a racemate with one enantiomer labeled with stable isotopes can show up any enantioselective route of metabolism. It is important to administer a racemic mixture because the

enantiomer may affect the metabolism of its optical twin.

However, care must be taken with stable isotopes. The rates of metabolism, especially of deuterated substrates, can be significantly different from those of the unlabeled substrates and the site of deuteration should be chosen with these 'isotope effects' in mind. A classical study with antipyrine (2,3-dimethyl-1-phenyl-3-pyrazalin-5-one) showed that substitution of the trideuteromethyl group in the 3 position switched the dominant metabolite from 3-dimethyl to 4-hydroxyantipyrine. Far fewer isotope effects are observed with  $^{13}\text{C}$  or  $^{15}\text{N}$  labels and these are to be preferred for *in vivo* work.

The technique of LC-MS/MS is a very powerful tool in drug metabolism studies. Mass shifts in daughter ion peaks can often provide sufficient information to enable the structure of a metabolite to be postulated. Software programs have been developed which allow automated screening of LC-MS data sets for the presence of possible metabolites. These often used simple tables of mass shifts obtained after metabolic modification but powerful computing tools such as principal components analysis (PCA) have enabled any changes in the postdose profile to be detected (see below, metabonomics). The introduction of hybrid spectrometers combining a quadrupole and an oa-ToF/MS enables the accurate mass of the parent compound to be determined, confirming its empirical formula. The protonated or deprotonated molecular ion can then be selected using the quadrupole mass spectrometer, fragmented in the collision cell, and analyzed using the oa-ToF, enabling the accurate masses of any fragment ions to be determined. Once a potential metabolite has been proposed it can be synthesized and confirmation of the structure of the metabolite can be made with authentic material.

## Quantification

### Endogenous Components

The quantification of endogenous components in biofluids, in contrast to the quantification of xenobiotic substances, raises problems as no true blanks of a biofluid can be obtained. However, despite the problems this poses, adequately validated assay procedures have been developed, and used as reference methods for the assessment of simpler more routine methods such as colorimetric or radioimmunoassay.

The use of isotopically labeled steroids as internal standards for the quantification of steroids in biofluids has been refined to a high degree. Four crucial steps in the analysis have been identified, namely

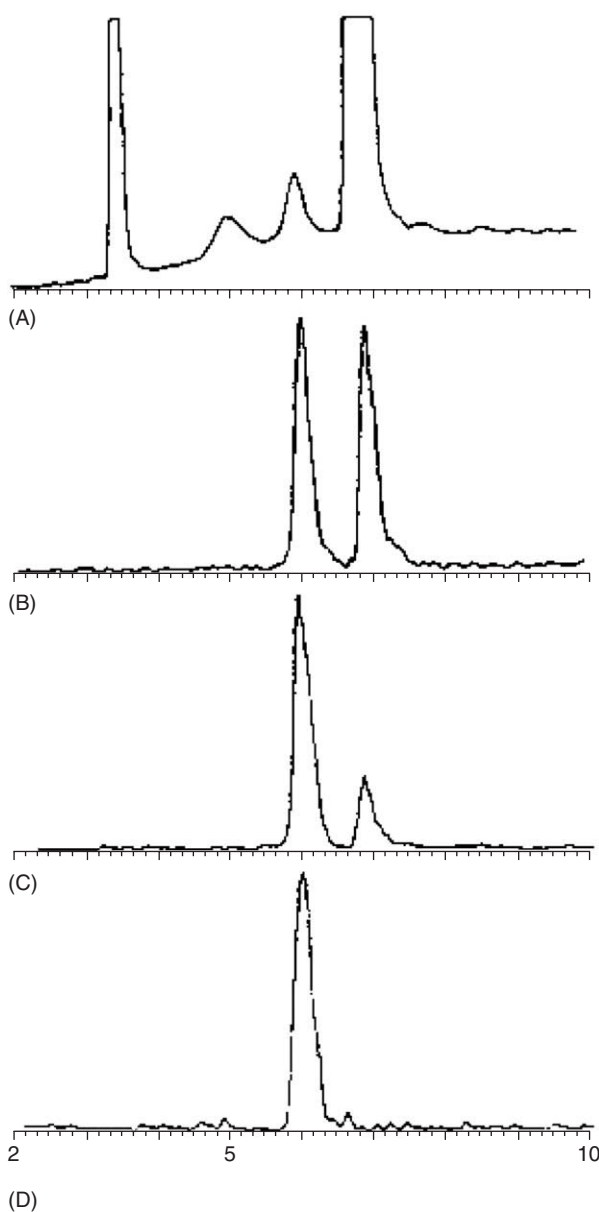
extraction, fractionation, derivatization, and GC-MS. Solvent extraction can be used but the use of immunoabsorbents (i.e., using microcellulose-coupled antisera for specific components) affords a high degree of selectivity as well as extraction. Fractionation of the extract may be required followed by derivatization to produce a derivative having good chromatographic properties as well as a significant ion in the mass spectrum. The use of *t*-butyldimethylsilyl ethers is often favored due to the intensity of the  $M - 57$  ion.

Mass spectrometric detection can be of three types:

1. Early work with quadrupoles used SIM at unit resolution, monitoring perhaps two ions for the analyte and two for the standard. However, the selectivity of the ion is not high and often the chromatograms contained many peaks due to isobaric ions (ions of the same nominal mass but different empirical formula, e.g., CO and C<sub>2</sub>H<sub>4</sub>) derived from other components.

2. Although the use of high resolution causes a reduction in absolute sensitivity of detection, the improved selectivity of the ion monitored often yields a higher signal-to-noise ratio. Thus, the use of high-resolution SIM may improve the overall limit of detection. The recent introduction of orthogonal time-of-flight analyzers has brought the cost of instruments capable of accurate mass measurements down dramatically, and small bench top systems are capable of generating accuracies of 5–10 ppm. These analyzers also have the advantages of faster response than scanning systems, enabling spectra to be obtained on sharp chromatographic peaks of 1–2 s width. Also, as a packet of ions are ejected in an instant from the primary beam for analysis in the ToF, no distortion of spectra are seen due to changing quantities of the analyte as the chromatographic peak passes through the ionization source. Mass chromatograms are selected for quantitation; there is no means for ToF analyzers to monitor single ions. These advantages make the combination of GC with oa-ToF a powerful and affordable new tool for the analysis of biofluid extracts.

3. Finally, the use of linked scan or MS/MS techniques can impart even greater selectivity. In this approach, the use of collisionally activated decomposition of a parent ion to give a suitable daughter ion is monitored. For instance, in a typical MS/MS experiment on *t*-BDMS ethers, the first mass spectrometer MS1 would pass the parent ion whilst MS2 would be set to pass the  $M - 57$  ion. This technique is known as selective reaction monitoring. An example is shown in Figure 2, which shows the analyses of a plasma extract of oestradiol-17 $\beta$  obtained using the



**Figure 2** GC-MS analyses of a plasma extract for oestradiol-17/3, as the bis-tertiarybutyldimethylsilyl ether. (A) Low-resolution (1000) selected ion monitoring of  $m/z$  500. (B) Selected reaction monitoring of  $m/z$  500  $\rightarrow$  443; parent resolution 1000. (C) as (B), parent ion resolution 2500. (D) as (B), parent ion resolution 5000. (Reproduced with permission from Gaskell SJ, Gould VJ, and Leith HM (1986) In: Gaskell SJ (ed.) *Mass Spectrometry in Biomedical Research*, pp. 347–361. Chichester: Wiley; © John Wiley & Sons Ltd.)

three techniques. The introduction of hybrid quadrupole–oa-ToF mass spectrometers means a modern more affordable spectrometer can also be used for this type of analysis.

### Toxicity Studies

The study of the ‘metabolome’; i.e., the total mixture of low molecular weight materials present within a

cell, has followed from the development of proteomics (see below). It has application in natural products research to identify active components and in pharmacognosy. It is also used in a semiquantitative manner to identify any compounds that have increased or decreased in concentration following ingestion of xenobiotic substances or in disease states. A particular approach concerning toxicity studies has been termed 'metabonomics'. The definition of metabonomics given by Nicholson is "the quantitative measurement of time related multiparametric responses of multicellular systems to pathophysiological stimuli."

Studying the downregulation or upregulation of these small molecules can provide markers for clinical diagnosis, which can be developed to provide clinical tests. The identification of these markers can be carried out by a combination of LC-MS(/MS) analyses followed by a detailed statistical review of the data sets using, for example, PCA.

Thus, following a normal toxicity study using increasing doses of a xenobiotic substance to elicit a toxic response, biological samples are collected at various time intervals after dosage. After minimal sample preparation LC/MS analyses are carried out leading to large, complex data files. The individual scans from these files are combined into a single scan file. This file is divided into intensity and mass data. The file is sometimes reduced further by grouping the mass data into bins covering 3 or 5 Da ranges. These data are then subjected to PCA. The scores plots are compared between samples and controls and used to identify those mass bins that have increased or decreased in relative concentration compared with the control. This information can be used to construct mass chromatograms using the original LC/MS data files and highlight the components that have changed in concentration. Sometimes it is possible to identify the required component from its simple MS spectrum, but chemical noise is often too great to obtain a clean interpretable spectrum. In these cases product ion spectra are used. An example is the analysis of urine after a toxic insult. PCA analysis indicated that the bin  $m/z$  113–115 was enhanced. Mass chromatography indicated a component giving an ion  $m/z$  114, which was identified as creatinine. An increase in creatinine is expected from liver toxicity. In addition, an increase in a nicotinic acid analog was also reported.

Apart from its use in toxicity studies, this approach (i.e., MS analysis followed by statistical processing of the data set followed by reanalysis of the sample to confirm the identity of perturbed metabolites) has obvious applications in clinical diagnosis following infection or some genetic abnormality.

To meet this demand most instrument companies are now offering automated systems that carry out the analysis and perform PCA processing with minimal intervention.

### Drug Metabolism Studies

The measurement of the concentration of drug-derived materials in biofluids, especially blood (plasma) and urine, is essential for the pharmacokinetic study of the rates of absorption, distribution, metabolism, and excretion. These are known as ADME studies. All such analyses are based on the addition of a known amount of an internal standard to an aliquot of the biofluid, followed by coextraction of the standard and the analyte after a period of equilibrium. The extract is purified and normally a chromatographic separation is carried out prior to detection.

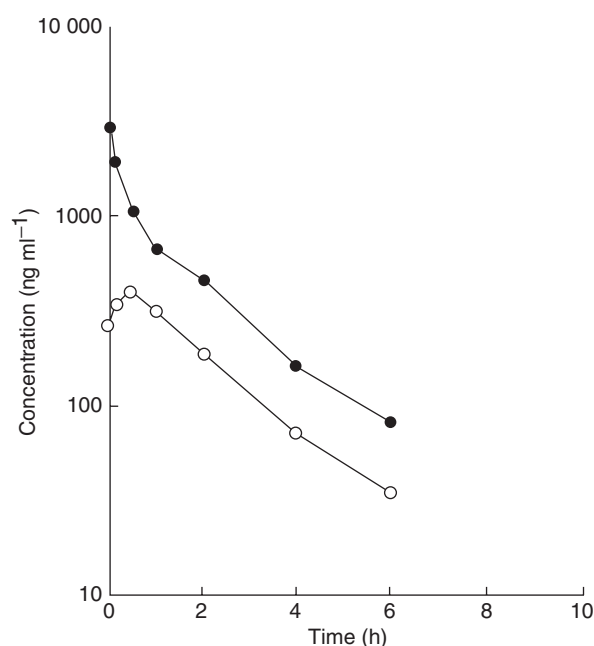
Such assays are closely monitored for both accuracy and precision. Calibration graphs are prepared by adding known amounts of analyte to blank samples of the biofluid under study and extracting these. A plot of the ratio of the response of the analyte to the response of the standard against concentration normally yields a straight line. The equation of this line can be used to calculate the concentration of analyte in unknown samples. These unknown samples should also include reference samples of analyte in biofluid where the concentration of the analyte is known to a quality assurance monitor but not to the analyst. These steps are necessary to ensure the accuracy of the results of the unknown.

These procedures often involve many samples, most of which will be analyzed in duplicate. This in turn has led to the development of automated systems for carrying out these analyses and interpreting the data.

The mass spectrometer's unique ability to differentiate between isotopes has led to the development of the 'stable isotope dilution' method for quantification. Although close analogs and homologs, or even isomers, of analytes are often used as internal standards, problems can arise because the physicochemical properties are not identical for the standard and the analyte. However, careful selection of a stable isotopically labeled analog can give an internal standard that has almost identical chemical properties. Use of such standards can lead to an improvement in both accuracy and precision, but especially precision.

The use of stable isotopes can be extended one stage further. Formulation pharmacists are interested in the amount of drug that is absorbed after oral administration of a tablet, capsule, or liquid

preparation. Experiments to determine this parameter (termed the 'absolute bioavailability') usually involve comparing the areas under the plasma curve after intravenous and oral administration of the compound. However, such studies require a 'wash-out' period between one administration and the other during which time the ADME profile may alter. Examples of this effect include drugs whose rate of excretion is critically dependent on urinary pH. These effects can be nullified by administering the intravenous dose and the oral dose at the same time. Obviously, one would not be able to distinguish between drugs from either route, except by labeling with stable isotopes. An example is shown in Figure 3 when disopyramide phosphate was administered to



**Figure 3** The plasma concentration versus time curves for disopyramide (●) and [<sup>13</sup>C, <sup>15</sup>N]-disopyramide (○) after coadministration of disopyramide phosphate (5 mg kg<sup>-1</sup>) intravenously (i.v.) and [<sup>13</sup>C, <sup>15</sup>N]-disopyramide phosphate (5 mg kg<sup>-1</sup>) by mouth (p.o.) to the rat. Data are the averages from nine animals. The areas under the curves were determined as 3522 ng h<sup>-1</sup> ml<sup>-1</sup> after i.v. and 1381 ng h<sup>-1</sup> ml<sup>-1</sup> after p.o. administration. This gave an absolute bioavailability figure of about 39%. (From Haskins NJ and Waddell KA (1981) Unpublished results.)

rats as an intravenous bolus (5 mg kg<sup>-1</sup>) and its <sup>13</sup>C, <sup>15</sup>N-labeled isotopomer was administered as an oral solution, also 5 mg kg<sup>-1</sup>. This is a versatile technique and can be extended to a range of studies when chemically identical materials are being compared. Thus, studies of various formulations against an orally administered reference solution, of enantioselective metabolism after labeling one enantiomer with label, and establishing kinetics of a single dose during the course of a chronic dosing schedule can all be carried out using a stable isotopically labeled drug.

## Peptides and Proteins

The biggest development in MS in the 1980s to the 1990s has been the introduction of methods for the analysis of larger peptides and proteins. In this context, the development of FAB ionization in the 1980s led to the use of mass spectrometers to determine the primary sequence of peptides, usually following enzymatic digestion. The later developments in the 1990s of matrix-assisted laser desorption/ionization (MALDI)-time-of-flight and ESI-MS allow the analysis of intact peptides and proteins. These developments led to the award of the Nobel Prize for chemistry to Professors Takeda and Fenn in 2002.

Since the 1990s the concentration of technological research on the analysis of large biomolecules has led to the development of a range of methods employing different ionization techniques and mass analysers. The most significant of these have been the combination of MALDI sources with high-performance mass analyzers (oa-ToF and FT-ICR/MS) improving the mass resolution of the spectra; the miniaturization of LC and electrospray systems to enable LC/MS analyses on small amounts of material; and the advent of static or very low flow electrospray using flows of 20 nl min<sup>-1</sup> (the so-called nanoflow electrospray). The attributes and application of these techniques are summarized in Table 8.

The advantage of MALDI-ToF is its mass range, but this is outweighed by its lack of resolution. ESI is technically more difficult because multiply charged ions are observed, but has the advantages of high

**Table 8** Typical parameters for the techniques available for the mass measurement of peptides and proteins

<i>Ionization technique</i>	<i>FAB</i>	<i>MALDI-ToF</i>	<i>Electrospray</i>	<i>Nanoflow electrospray</i>
Mass range	<20 kDa	<500 kDa	<200 kDa	<200 kDa
Mass accuracy	±0.001%	±0.01%	±0.001%	±0.001%
Sensitivity	nmol	<pmol	pmol	<fmol
Sample introduction	Solution in matrix	Dried in matrix	Solution	Solution

FAB, Fast atom bombardment; MALDI-ToF, matrix-assisted laser desorption/ionization time-of-flight.

mass range and good resolution. Examples of the clinical applications of these techniques include the use of ESI-MS to identify sickle-cell hemoglobin from adults and neonates, clearly showing the multi-charged species for the sickle  $\beta$ -chain ( $M_r$  15 837) separated from the normal  $\beta$ -chain ( $M_r$  15 867). Spectra of hemoglobins have also been obtained with MALDI-ToF.

The second technological innovation has been the introduction of powerful computing techniques. PCA of complex data sets has been used to detect up- or downregulation of proteins in complex mixtures in a similar manner to small molecules (see metabolomics above). In addition, application of information technology has revolutionized the rapid identification of peptides and proteins. This has led to MS becoming the method of choice for the study of proteomics.

## Proteomics

Proteomics is the study of the total protein complement of the cell and has obvious applications in clinical diagnosis and research. Once a protein has been recognized as either up- or downregulated it is necessary to try and identify it. Two techniques are being routinely used for this. Firstly separation of the protein mixture using 2D gel plates and a combination of electrophoretic and chromatographic separations. Spots that are changed in intensity can be excised and the protein recovered. After digestion the peptide mixtures can be analyzed by direct analysis using MALDI-ToF or nanoflow ESI. The mixture of peptides can be compared with a library and identification often made. Failing this, peptides can be subjected to MS/MS analysis and a sequence obtained. Again the sequence can be compared with libraries.

The second technique is to digest the total protein mixture and analyze the peptide mixture using LC-MS(/MS) and PCA analysis. After identifying peptides that have been up- or downregulated, comparison with libraries can usually give a positive identification.

A more recent separation technique uses the so-called 'isotope coded affinity tags' (ICAT<sup>TM</sup> Applied Biosystems, Foster City, CA). In this technique, a tag comprising a stable isotopically labeled marker and a biotin molecule is reacted with the protein extract from a perturbed sample whilst an unlabelled tag is used to mark proteins from a control sample. Following purification on an affinity column to isolate the biotin labeled peptides the biotin is cleaved leaving a mixture of tagged peptides. These either have a labeled (clinical sample) or unlabelled marker (control). The mixture is subject to LC-MS(/MS)

analysis. The ratio of label/unlabelled material should be constant if no perturbation has occurred. Differences in relative concentration between the sample and its control will identify target proteins. Sequence data from MS/MS analysis in combination with a library search should identify the protein. Once again commercial systems with a high degree of automation are being developed.

The identification of changes in protein concentration can give crucial information concerning the effect of disease, genetic problems, or exposure to xenobiotic substances (drugs, toxins, etc.). It is becoming a powerful clinical diagnostic tool and again instrument manufacturers are producing automated systems that can carry out the full analysis with minimal operator intervention. This is probably the area of application of MS in clinical research that will see the greatest advance in the next decade.

*See also:* **Clinical Analysis:** Sample Handling. **Gas Chromatography:** Mass Spectrometry. **Liquid Chromatography:** Liquid Chromatography–Mass Spectrometry. **Mass Spectrometry:** Ionization Methods Overview; Atmospheric Pressure Ionization Techniques; Time-of-Flight; Selected Ion Monitoring; Stable Isotope Ratio. **Peptides. Proteins:** Traditional Methods of Sequence Determination. **Quality Assurance:** Internal Standards.

## Further Reading

- Burlingame AL and McCloskey JA (1990) *Biological Mass Spectrometry*. Amsterdam: Elsevier.
- Chapman TE, Berger R, Reijngiud DJ, and Okken A (1990) *Stable Isotopes in Paediatric Nutritional and Metabolic Research*. Andover: Intercept.
- Cottee F, Haskins N, Bryant D, Eckers C, and Monté S (2000) The use of accurate mass measurement by orthogonal time-of-flight mass spectrometry in pharmaceuticals research. *European Mass Spectrometry* 6: 219–224.
- Han D, Eng J, Zhou H, and Aebersold R (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope coded affinity tags and mass spectrometry. *Nature Biotechnology* 19: 946–951.
- Haskins NJ (1982) The application of stable isotopes in biomedical research. *Biomedical Mass Spectrometry* 9: 269–277.
- Horning MG, Nowlin J, Thenot JP, and Bauwsma OJ (1979) Effect of deuterium substitution on the rate of caffeine metabolism. In: Klein RL and Klein PD (eds.) *Stable Isotopes. Proceedings of the 3rd International Conference*, pp. 379–384. New York: Academic Press.
- Major H, Nicholls A, Shockor J, and Perez J-C (2002) An HPLC–MS metabonomics approach to the study of rat urine after dosing with hydrazine. Presented at RSC-DMDG, Edinburgh, December 12.



- McEwen CN and Larsen BS (1990) *Mass Spectrometry of Biological Materials*. New York: Dekker.
- Millard BJ (1978) *Quantitative Mass Spectrometry*. London: Heyden and Sons.
- Nicholson JK, Connelly J, Lindon JC, and Holmes E (2002) Metabonomics: A platform for studying drug

toxicity and gene function. *Nature Reviews Drug Discovery* 1: 153–161.

- Paust H, Park W, Helge H, and Scigalla P (1988) *Use of Stable Isotopes in Clinical Research and Practice*. Munich: W. Zuckerschwerdt-Verlag.
- Sjövall J and Axelsson M (1982) *Vitamins and Hormones* 39: 31.

## Environmental Applications

**D Barceló, E Eljarrat, and M Petrovic**, Institut d'Investigacions Químiques i Ambientals de Barcelona, Consejo Superior de Investigaciones Científicas, Barcelona, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The analysis of pollutants in environmental samples is complicated by the fact that these compounds are often part of a very complex mixture of hundreds of chemicals and that they are present at extremely low concentrations. Thus, any instrumental technique that is to be used successfully in environmental analysis should allow the separation of such complex mixtures and the identification of the constituents at trace level. Mass spectrometry (MS) is particularly suited for this purpose as it can be coupled with the most important instrumental separation techniques – gas chromatography (GC) and liquid chromatography (LC) – while at the same time it is one of the most sensitive instrumental methods available for the analysis of contaminants. Moreover, it is not only the sensitivity, but also the specificity that makes MS the most powerful instrumental method in environmental analysis.

During the last 10 years the environmental analysis field has changed significantly. Impressive improvements in detection limits are such that parts-per-trillion (ppt) detection is routine for many organic analytes, and detection of a few hundred femtograms of some analytes is now possible, although such detection levels cannot yet be considered routine. Moreover, technologies that have significantly advanced in the last few years include tandem (MS–MS) systems, time-of-flight MS (TOF–MS) and quadrupole-time-of-flight MS (Q–TOF–MS). This article attempts to survey current state-of-the-art in the application of modern mass spectrometric techniques in the environmental analysis.

### Target Compounds

A wide range of man-made chemicals designed for use in industry, agriculture, and consumer goods and

chemicals unintentionally formed or produced as by-products of industrial processes or combustion are potentially of environmental concern. Beside recognized pollutants, numerous new chemicals are synthesized each year and released into the environments with unforeseen consequences. Over the past few decades, analytes having acute toxic or carcinogenic effects and that are persistent in the environment as a result of anthropogenic activity have been widely studied. Examples of these compounds include BTEX (benzene, toluene, and xylene), DDT and its metabolites, polycyclic aromatic hydrocarbons (PAHs), and chlorinated biphenyls (PCBs). While toxic, carcinogenic, and persistent compounds will always be of concern, the last several years have brought an increased interest in endocrine disrupting (ED) compounds – those that interrupt normal hormone function and cause adverse biological effects. ED chemicals encompass a large range of organic, metallic, and organometallic substances (Table 1). There are two classes of substances that can cause endocrine disruption: natural substances (hormone found naturally in the body of humans and animals, and phytoestrogens, substances contained in some plants) and man-made substances. The group of man-made substances comprises synthetically produced hormones designed intentionally to interfere with the endocrine system (e.g., oral contraceptives) and man-made chemicals designed for use in industry, agriculture, and consumer goods that may have unforeseen adverse or synergistic effects. Man-made chemicals also include chemicals unintentionally formed or produced as a by-product of industrial processes or combustion.

Other compounds of interest are the so-called 'emerging contaminants' (Table 1). Emerging contaminants are unregulated contaminants, which may be candidates for future regulation depending on research on their potential health effects and monitoring data regarding their occurrence. This group is mainly composed of products used in everyday life, such as surfactants and surfactant residues, pharmaceuticals and personal care products, gasoline

- McEwen CN and Larsen BS (1990) *Mass Spectrometry of Biological Materials*. New York: Dekker.
- Millard BJ (1978) *Quantitative Mass Spectrometry*. London: Heyden and Sons.
- Nicholson JK, Connelly J, Lindon JC, and Holmes E (2002) Metabonomics: A platform for studying drug

toxicity and gene function. *Nature Reviews Drug Discovery* 1: 153–161.

- Paust H, Park W, Helge H, and Scigalla P (1988) *Use of Stable Isotopes in Clinical Research and Practice*. Munich: W. Zuckerschwerdt-Verlag.
- Sjövall J and Axelsson M (1982) *Vitamins and Hormones* 39: 31.

## Environmental Applications

**D Barceló, E Eljarrat, and M Petrovic**, Institut d'Investigacions Químiques i Ambientals de Barcelona, Consejo Superior de Investigaciones Científicas, Barcelona, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The analysis of pollutants in environmental samples is complicated by the fact that these compounds are often part of a very complex mixture of hundreds of chemicals and that they are present at extremely low concentrations. Thus, any instrumental technique that is to be used successfully in environmental analysis should allow the separation of such complex mixtures and the identification of the constituents at trace level. Mass spectrometry (MS) is particularly suited for this purpose as it can be coupled with the most important instrumental separation techniques – gas chromatography (GC) and liquid chromatography (LC) – while at the same time it is one of the most sensitive instrumental methods available for the analysis of contaminants. Moreover, it is not only the sensitivity, but also the specificity that makes MS the most powerful instrumental method in environmental analysis.

During the last 10 years the environmental analysis field has changed significantly. Impressive improvements in detection limits are such that parts-per-trillion (ppt) detection is routine for many organic analytes, and detection of a few hundred femtograms of some analytes is now possible, although such detection levels cannot yet be considered routine. Moreover, technologies that have significantly advanced in the last few years include tandem (MS–MS) systems, time-of-flight MS (TOF–MS) and quadrupole-time-of-flight MS (Q–TOF–MS). This article attempts to survey current state-of-the-art in the application of modern mass spectrometric techniques in the environmental analysis.

### Target Compounds

A wide range of man-made chemicals designed for use in industry, agriculture, and consumer goods and

chemicals unintentionally formed or produced as by-products of industrial processes or combustion are potentially of environmental concern. Beside recognized pollutants, numerous new chemicals are synthesized each year and released into the environments with unforeseen consequences. Over the past few decades, analytes having acute toxic or carcinogenic effects and that are persistent in the environment as a result of anthropogenic activity have been widely studied. Examples of these compounds include BTEX (benzene, toluene, and xylene), DDT and its metabolites, polycyclic aromatic hydrocarbons (PAHs), and chlorinated biphenyls (PCBs). While toxic, carcinogenic, and persistent compounds will always be of concern, the last several years have brought an increased interest in endocrine disrupting (ED) compounds – those that interrupt normal hormone function and cause adverse biological effects. ED chemicals encompass a large range of organic, metallic, and organometallic substances (Table 1). There are two classes of substances that can cause endocrine disruption: natural substances (hormone found naturally in the body of humans and animals, and phytoestrogens, substances contained in some plants) and man-made substances. The group of man-made substances comprises synthetically produced hormones designed intentionally to interfere with the endocrine system (e.g., oral contraceptives) and man-made chemicals designed for use in industry, agriculture, and consumer goods that may have unforeseen adverse or synergistic effects. Man-made chemicals also include chemicals unintentionally formed or produced as a by-product of industrial processes or combustion.

Other compounds of interest are the so-called 'emerging contaminants' (Table 1). Emerging contaminants are unregulated contaminants, which may be candidates for future regulation depending on research on their potential health effects and monitoring data regarding their occurrence. This group is mainly composed of products used in everyday life, such as surfactants and surfactant residues, pharmaceuticals and personal care products, gasoline

**Table 1** List of endocrine disruptors and emerging contaminants, analyzed in environmental samples using mass spectrometric techniques

<i>Chemical group</i>	<i>Example</i>	<i>MS technique</i>
Benzenes	Hexachlorobenzene	GC-MS
Bisphenols	Bisphenol A	LC-MS
Chlorinated cyclodienes and camphenes	Chlordane, mirex, toxaphene	GC-MS, HRMS
DDT, derivatives and metabolites	DDT, DDE, tetrachloro-DDT	GC-MS
Disinfection by-products	THMs, HAAs, haloacetonitriles, haloketones, haloaldehydes	GC-MS (EI)
Flame retardants	PBDEs, HBCD, TBBPA	GC-MS (EI and NCI), HRMS
Gasoline additives	Dialkyl ethers, MTBE	GC-MS
Hexachlorocyclohexane and isomers	Lindane	GC-MS
Industrial additives and agents	Chelating agents (EDTA), aromatic sulfonates	LC-MS (ESI)
Organotin compounds	Tributyltin, tributyltin oxide, triphenyltin	GC-ICP-MS, TOF-MS
PBBs	PBB-153, PBB-180, PBB-209	GC-MS (EI and NCI)
PCBs	PCB-77, PCB-153, PCB-169	GC-MS (EI), HRMS
PCDDs/PCDFs	2378-TCDD, 12378-PeCDD, 23478-PeCDF	GC-HRMS (EI), IT-MS
Personal care products	Nitro musks, polycyclic musk, chlorophene	GC-MS (EI and NCI)
Pharmaceuticals	Trimethoprim, erythromycine, sulfamethaxazole, ibuprofene, acetylsalicylic acid, diazepam, estradiol, estrone, estriol	GC-MS (EI and NCI), IT-MS, HRMS, LC-MS (ESI and APCI), MS-MS (ESI and APCI), GC-MS (EI and PCI)
Phthalates	BBP, DEHP, DBP	GC-MS (EI and PCI), LC-MS (ESI)
Surfactants and metabolites	Alkylphenol ethoxylates and carboxylates, NP, OP	GC-MS, LC-MS (APCI)
Triazines and other pesticides	Atrazine, acetochlor, alachlor	

additives, fire retardants, plasticizers, etc. The characteristic of these group contaminants is that they do not need to be persistent in the environment to cause negative effect since their high transformation/removal rates can be compensated by their continuous introduction into environment.

GC-MS and LC-MS are the most commonly used techniques for the environmental analysis of these target compounds. Generally, GC-MS is used for apolar or moderate polar compounds, while for polar compounds LC-MS is the preferred technique. Other compounds are analyzed using both techniques.

## Gas Chromatography-Mass Spectrometry

GC-MS continues to play an important role in the identification and quantification of contaminants in environmental samples. Different mass analyzers, such as linear quadrupole, quadrupole ion-trap, double-focusing sectors, and TOF analyzers, are employed in different applications. Several ionization techniques are also used in GC-MS. Among them, electron ionization (EI) is the most popular because it often produces both molecular and fragment ions. One important feature of EI spectra is that they are highly reproducible, which means that mass spectral libraries can be used for identification of unknowns. However, in some cases, EI does not provide the

sensitivity required for the analysis of very small amounts of compounds in environmental samples. To solve this problem, softer ionization techniques such as chemical ionization (CI) are applied.

A large number of publications have resulted from research on environmental applications of GC-MS. The compounds most commonly analyzed include alkanes, PAHs, pesticides, volatile organic compounds (VOCs) including off-flavor and water disinfection by-products (DBPs), PCBs, polychlorinated dibenzo-*p*-dioxins, and furans (PCDDs/Fs), as well as other ED chemicals such as phthalates and short ethoxy alkylphenol ethoxylate. GC-MS is also the technique of choice for the analysis of emerging contaminants, such as polybrominated diphenyl ethers (PBDEs) or polychlorinated alkanes, as well as for the analysis of some pharmaceuticals and organo-metallic compounds.

## Inorganic Compounds

The analysis of inorganic compounds including metals, mercury, and volatile and nonvolatile organometallics using MS methodologies has been reported in the literature. Inductively coupled plasma (ICP)-MS is enjoying an increasing interest in inorganic trace and ultratrace analysis because of its unmatched sensitivity, elemental specificity, multielement measurement capability, and the possibility of the accurate determination of elemental

isotope ratios. ICP-MS was coupled to GC for speciation analysis in environmental samples, including speciation of organomercury, organolead, and organotin applications. Moreover, GC-ICP-MS is the only technique to date enabling direct speciation analysis of As and Hg in natural gases and gas condensates at the micrograms per liter level. Some recent papers have shown the potential of ICP-TOF-MS for the separation and speciation of organotin and organolead compounds. Detection limits in the femtogram level with good reproducibility were obtained.

### Pharmaceuticals (Steroid Sex Hormones, Human and Veterinary Drugs)

GC-MS has been the technique most commonly implicated in the determination of estrogens in environmental samples. Both conventional MS and ion trap (IT)-MS detection have been accomplished in the EI mode at 70 eV, usually in the selected ion monitoring (SIM) mode, and, in most cases, after sample derivatization, while the use of negative (N)CI has seldom been reported. The sensitivity of GC-MS-MS was reported to be in the range  $20 \text{ pg } \mu\text{l}^{-1}$  (limit of quantification). A recent study has also discussed the possibilities of high-resolution (HR)MS for analysis of steroid sex hormones in environmental samples, showing that, with 20 000 resolution, 10 pg of ethynyl estradiol can be detected. This amount corresponds to a detection limit of 1 ppt if the ethynyl estradiol is in 1 l of water (concentrated into a 0.1 ml extract, followed by injection of 1  $\mu\text{l}$ ).

The often incomplete derivatization of highly polar drugs makes inappropriate their determination by GC. However, a method based on solid-phase extraction (SPE), derivatization by silylation, and detection by GC-MS has been developed for the determination of 22 different neutral and weakly basic drugs belonging to several different medicinal classes like antiphlogistics,  $\beta$ -blockers,  $\beta_2$ -sympathomimetics, lipid regulators, antiepileptic agents, psychiatric drugs, and vasodilators in wastewater as well as in river and drinking water. This method permits detection down to  $5 \text{ ng l}^{-1}$  with recovery rates mostly exceeding 70%.

### Polybrominated Diphenyl Ethers

Several methods for qualitative and quantitative analysis have been developed involving GC-NCI-MS and GC-EI-MS. Until recently, quantitative work has been performed by using technical PBDE products, i.e., Bromkal 70-5DE (BK70), due to the lack of pure reference standards for most PBDE congeners. The major three components in BK70

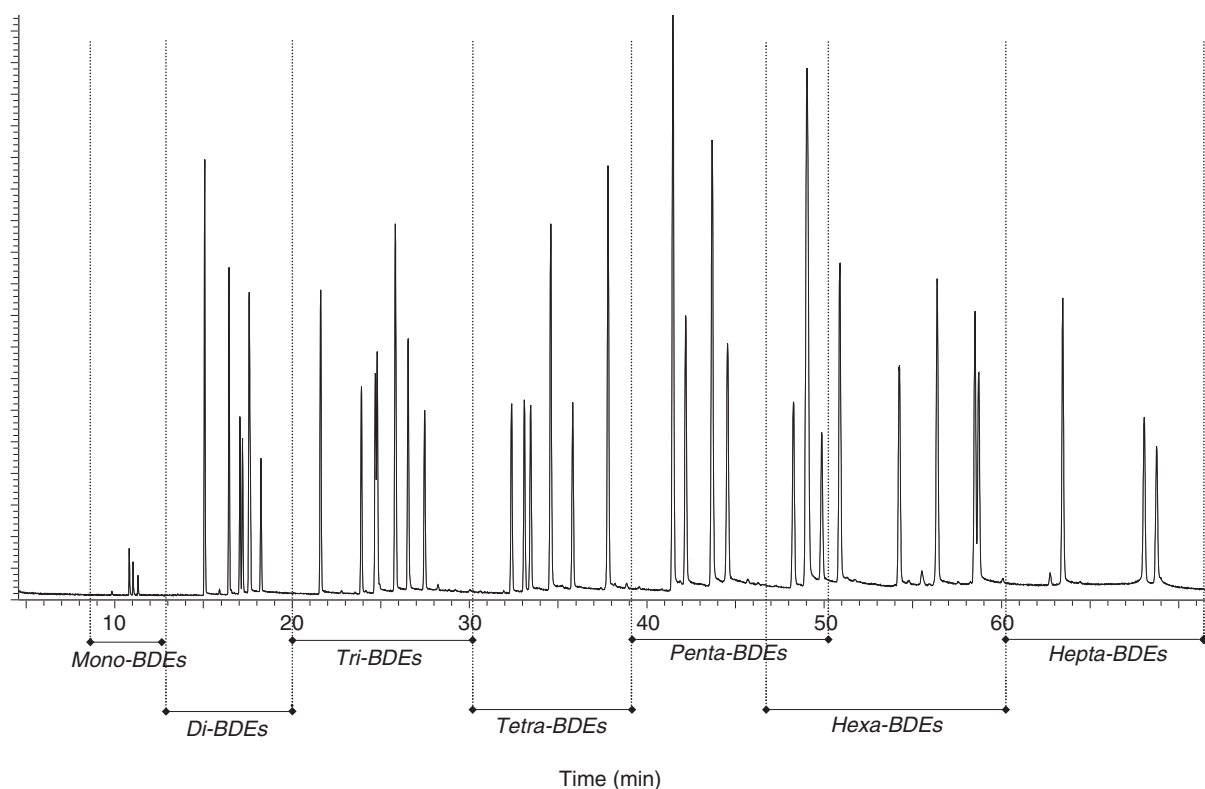
have been identified as BDE-47, BDE-99, and BDE-100, and only these congeners could be quantified. Since more than 30 PBDE congeners now are available it has become possible to analyze for additional PBDEs (Figure 1). Moreover, the availability of some  $^{13}\text{C}$ -labeled standards allows the development of a methodology based on the quantification by the isotopic dilution method. Previous reports had used other  $^{13}\text{C}$  or  $^{12}\text{C}$ -PCB and other organochlorine surrogates that, in general, give poorer precision and accuracy in the determination of analytes.

Some studies have been carried out comparing the capabilities of the two MS approaches (EI and NCI) for the PBDE determinations. Comparison of detection limits for PBDEs clearly indicates that NCI offers better sensitivity than does EI: NCI-MS gave detection limits between 30 fg and 1.72 pg, whereas EI-MS gave detection limits between 0.53 pg and 32.09 pg. The main advantage of EI-MS is that it provides better structural information. No structural information on the degree of bromination was obtained by NCI. The mass spectra of all PBDEs were dominated by the molecular ion  $[\text{Br}]^-$  and did not show any molecular ion. However, EI provided better structural information, giving the molecular ions and the sequential losses of bromine atoms. Moreover, the use of EI-MS allowed the use of an isotopic dilution method for quantification, making the analysis more reliable at trace levels.

### Polychlorinated Biphenyls

Different determinations of PCBs have been developed. A number of countries have chosen to monitor PCBs as a set of seven indicator PCBs (IUPAC Nos. 28, 52, 101, 118, 138, 153, and 180). These PCBs were analyzed by GC-low resolution (LR)MS. However, when dioxin-like PCBs (non-*ortho* and mono-*ortho* PCBs) were studied, comprehensive analytical procedures are necessary because they occur at concentrations lower than the indicator PCBs mentioned above, and are therefore very elaborate and complicated to analyze. In these cases, a GC-HRMS with quantification by isotopic dilution method was required.

The powerful HRMS instrumentation that is used for the analysis of these contaminants suffers some limitations that are related to the limited accelerating voltage working range for a given group of ions in the SIM mode. For example, co-eluting compounds with wide differences in their masses cannot be effectively monitored in the same window and the chromatographic run thus needs to be prolonged for adequate component speciation. Due to their non-scanning character, TOF-MS is a valuable tool for fast GC because it is able to monitor the entire mass



**Figure 1** GC-NCI-MS-SIM ( $m/z$  79) obtained for a standard solution containing from mono- to hepta-BDEs. Different chromatographic windows could be defined in order to enhance sensitivity.

range in very short times. Different recent studies showed the capabilities of TOF for the analysis of PCBs in different type of samples. Thirty different PCB congeners were analyzed in biological samples above the detection limit in less than 7 min. Routinely around 40 congeners are detected above the detection limit by using SIM on a quadrupole GC-MS in 45 min. The detection limit using the GC-TOF seemed to be somewhat higher.

Another advantage of GC-TOF analyses is that full scan spectra are always acquired. The acquisition of full scan spectra during the analysis opens the possibility of screening for 'unknown' pollutants present in the extracts. A GC-MS-SIM run always need input on the masses to be monitored, and thus this kind of quantitative analysis is restricted to compounds known to be present in the extracts. By using GC-TOF new compounds can be identified by their mass spectra while doing quantitative analysis of known compounds at the same time.

#### Polychlorinated Dibenzo-*p*-dioxins and Furans

MS has been the leading technique used in dioxin analysis. Until the 1990s, ultratrace determinations of polychlorinated dibenzo-*p*-dioxins and furans (PCDDs/Fs) were routinely performed using GC-HRMS, operating in the EI mode (electron energy

$\sim 38$  eV) at 10 000 resolving power. Quantification was carried out by an isotopic dilution technique, based on the addition of known amounts of  $^{13}\text{C}$ -labeled standards.

The application of HRMS has proved to provide the required sensitivity and specificity. The required specificity could be provided by MS-MS as well. While MS-MS with sector or quadrupole instruments needs a series of mass analyzers in space, ITs use one mass analyzer to perform MS-MS in time. Recently, IT-MS was developed to analyze PCDDs and PCDFs. The advantage of IT-MS systems is the much lower price, which could reduce analyses costs. However, sensitivity of IT-MS instruments is considerably lower than of HRMS instruments. The limit of determination for IT-MS systems for 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (signal-to-noise ratio 3:1) can be assumed to be in the range of  $\sim 100$ – $300$  fg, whereas modern HRMS instruments have a detection limit of  $\sim 3$  fg.

#### Polycyclic Aromatic Hydrocarbons

Different analytical methods have been published for the PAH determinations. PAHs in environmental samples are usually determined by GC-MS, using both EI and CI modes, or by LC with either fluorescence or diode-array detection. In recent years,



LC-MS using atmospheric-pressure chemical ionization (APCI) has been seldom used for the PAH analyses. However, detection limits obtained using LC-MS were approximately one order of magnitude higher than those obtained by GC-MS.

In the last few years, new procedures were developed for PAH analysis using MS-based detection methods. GC-IT-MS and a laser desorption-IT-MS screening method were applied to the analysis of airborne particulate-associated PAHs. Direct screening of aerosol filter samples were carried out using laser desorption TOF-MS. Isotope dilution techniques using GC-MS were also developed for some PAHs, such as benzo[a]pyrene.

### Surfactants

Numerous GC-MS applications for the analysis of surfactants have come into focus since the early 1960s when the first GC papers on separation of nonionic surfactants were published. However, the major drawback for application of GC to surfactants is their lack of volatility.

Among different nonionic surfactant classes, alkylphenol ethoxylates (APEOs) comprise the class meriting special attention with respect to environmental issues. The analysis of underivatized alkylphenolic compounds by GC-MS is restricted to the most volatile degradation products, such as alkylphenols and APEOs with less than 4-ethoxy groups. To overcome the problem of volatility, different offline and online derivatization protocols have been developed. Two complementary MS techniques, one using EI and another using the less commonly used positive (P)CI, have been evaluated for the analysis of APEOs, their acidic (APECs) and neutral metabolites (APs), and halogenated derivatives.

Linear alkylbenzene sulfonates (LAS) are the most commonly used anionic surfactants; however, as nonvolatile compounds, they cannot be analyzed directly by GC-MS and pretreatment is required to form volatile derivatives. Capillary GC of derivatized (or desulfonated) LAS allows separation according to alkyl chain length, with the highest chain lengths giving longer retention times and with internal isomers of each homolog eluting first. The higher separation efficiency of GC over LC has been particularly useful to study the mechanism of biodegradation of homologs and isomers in laboratory-scale experiments.

### Volatile Organic Compounds

Miniature instrumentation and field portability is an area of increasing interest in MS. These instruments were ideal for a variety of *in situ* analysis in environmental applications. *In situ* analysis of volatile

organic compounds (VOCs) is a good option because it can solve the problems related with sampling, evaporative loss, and transport. For instance, a GC-MS field instrument has been used for the analysis of BTEXs in urban outdoor environments. The mass analyzer most frequently used in GC-MS field instruments is the linear quadrupole due to its simplicity, small size, durability, and compactness. However, other analyzers were also tested. Direct sampling IT-MS systems were evaluated for monitoring trace levels of halogenated VOCs in air samples. It was found that this method offers comparable results with those obtained using MS(SIM) and MS-MS operation modes.

In water, GC-MS is coupled to purge and trap or headspace sample preparation for the analysis of VOCs like BTEX and MTBE. Another important group of volatile analytes in water are DBPs. Attention has been directed to volatile chlorinated compounds such as trihalomethanes (THMs), as well as other semivolatile compounds such as haloacetic acids (HAAs), haloacetonitriles, halo ketones, and haloaldehydes. The methods used to determine these compounds include GC-EI-LRMS, where a after derivatization step is necessary due to the low volatility and high polarity of these analytes. Using this technique, limits of detection were in the microgram per liter range.

The most obvious taste and odor compounds, such as hydrocarbons, solvents, iodoforms, and various chloro- and bromophenols, which have odor thresholds in the microgram per liter range, are currently analyzed by GC-EI-MS using quadrupole instruments.

## Liquid Chromatography-Mass Spectrometry

Currently, the main breakthrough in environmental analysis is observed in the application of LC-MS and LC-MS/MS techniques. One of the obstacles to routine analytical applications of LC-MS had been the unavailability of rugged and reliable LC-MS interfaces. The development of atmospheric pressure ionization (API) overcame such limitations as poor structural information or sensitivity seen with thermospray (TSP) or particle-beam (PB). API is used as a generic term for soft ionization obtained by different interface/ionization types, such as APCI and electrospray (ESI) that operate under atmospheric pressure conditions. Today, LC-MS has become a routine analytical tool, allowing the detection of polar and nonvolatile compounds not amenable to GC analysis.

## Pesticides

Most of the modern pesticides and their degradation products are characterized by medium to high polarity and thermal lability. Neutral and basic compounds (phenylureas, triazines) are more sensitive using APCI (especially in positive ion mode), while cationic and anionic herbicides (e.g., bipyridylium ions, sulfonic acids) are more sensitive using ESI (especially in negative ion mode). Based on comparative analysis of 75 pesticides, the so-called ionization-continuum diagram illustrates the relationship between compound acidity and appropriate ionization modes (Figure 2).

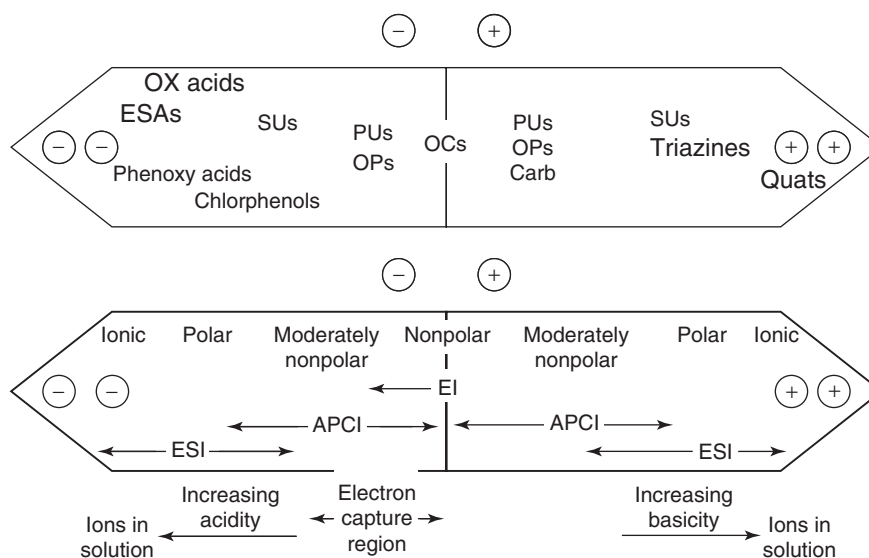
Some studies have reported the possibilities of screening surface water and identification of water pollutants using a modern data acquisition of Q-TOF-MS, such as data-dependent MS to MS-MS switching. Additionally, online SPE followed by LC with the combination of orthogonal-acceleration TOF-MS is also shown to be a powerful tool in the determination of pesticides from various compound classes in surface water in the concentration range of  $0.1\text{--}10\text{ }\mu\text{g l}^{-1}$ . LC-TOF-MS seemed to be a promising method for the identification of the unknown pesticides or their degradation products, which were almost impossible by a single quadrupole LC-MS. The TOF-MS gives the accurate mass of the molecular ion, which in turn may be used to determine the empirical formula of the unknown compound. This is helpful in the determination of pesticides, when no standards exist, especially for the ones where the mass of the molecular ion is typically less than 350 amu.

## Steroid Sex Hormones and Other Pharmaceuticals

The principal analytical methods employed in the analysis of pharmaceuticals in aqueous environmental samples include both GC and LC coupled to MS or MS/MS. In general, LC-MS methods have demonstrated lower relative standard deviations than GC-MS methods that have derivatization prior to analysis. LC-ESI-MS/MS is presented as the technique of choice for polar, unstable, and high molecular mass compounds, such as most pharmaceuticals and their metabolites. MS detection has preferably been performed with the ESI interface. An analytical protocol applying LC-ESI-MS/MS with previous filtration of the water sample, with addition of  $\text{Na}_2\text{EDTA}$  to avoid complexation of analytes (especially tetracyclines) with metals present in the water, was proposed for the analysis of antibiotics corresponding to the classes of macrolides, sulfonamides, penicillins, and tetracyclines in water samples.

For LC-MS determination of estrogens, ESI operating in the negative ion mode of ionization has been the most widely used interface because of its observed better sensitivity compared to the same interface operating in the positive ion mode of ionization and to APCI.

In the LC-MS analysis of estrogens, selected ion monitoring of the  $[\text{M}-\text{H}]^-$  ion in ESI (NI) and the  $[\text{M} + \text{H} - \text{H}_2\text{O}]^+$  ion ( $[\text{M} + \text{H}]^+$  for estrone) in APCI (positive) is usually carried out for maximum sensitivity. Progestogens can be determined in the positive ion mode of operation with both ESI and APCI. However, the sensitivity achieved with the



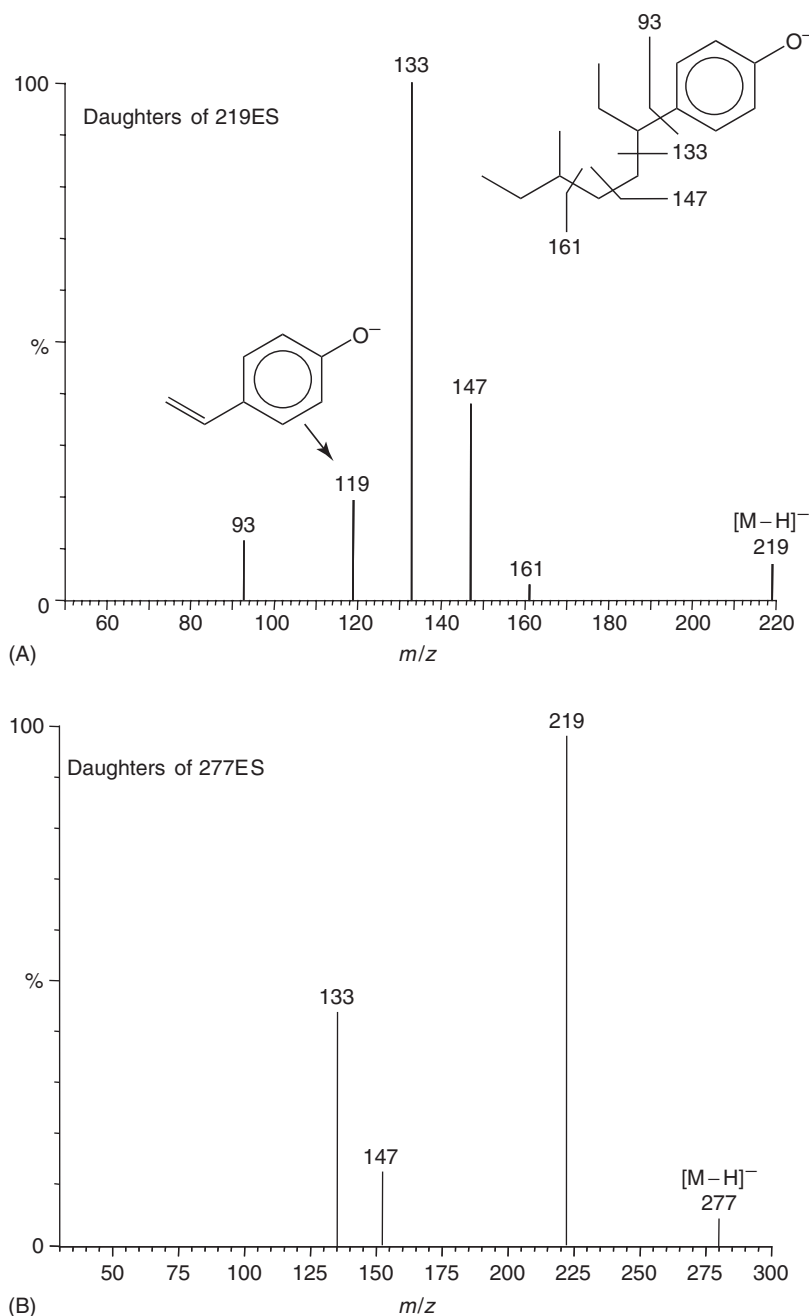
**Figure 2** Ionization continuum diagram showing the interrelationship between pesticide properties and the appropriate API mode. (Reprinted with permission from Thurman EM, Ferrer I, and Barceló D (2001) *Analytical Chemistry* 73: 5441; © American Chemical Society.)

APCI interface is about 10-fold lower than that of the ESI interface. Application of MS-MS improves the sensitivity of the detection (limits of quantification  $5 \text{ pg } \mu\text{l}^{-1}$ ) as compared to LC-MS (limits of detection  $200 \text{ pg } \mu\text{l}^{-1}$ ).

### Surfactants

Owing to the anionic character of LAS, LC coupled with an ESI-MS operated in negative mode is

particularly attractive for the quantitative determination of this surfactant type. LAS analysis by ESI-MS under negative conditions led to mass spectra with the four  $[\text{M}-\text{H}]^-$  ions of  $m/z$  297, 311, 325, and 339, corresponding to the LAS homologs C10–C13, respectively. With increasing cone voltage, additional fragment ions at  $m/z$  183 and 80 were obtained, which were assigned to styrene-4-sulfonate and  $[\text{SO}_3]^-$ , respectively. No side chain-specific



**Figure 3** Product ion scan of (A) NP with  $m/z$  219 and of (B) NPE<sub>1</sub>C with  $m/z$  277 and the proposed fragmentation scheme under CID conditions obtained using argon as collision gas at collision energy of 40 eV. Note: the exact branching of the alkyl chain is unknown and the alkyl isomer shown here is just one of several possibilities.

fragment ions were produced under API conditions; thus a mass spectrometric distinction of positional isomers of LAS could not be achieved. Though the sulfonate group of LAS is quite acidic, the long hydrophobic alkyl chain provides sufficient retention in RPLC; however, the more polar carboxylated degradation products, the sulfophenyl carboxylates (SPCs), require additives such as triethylamine or tetraethylammonium acetate for a sufficient retention. To avoid LC conditions involving the use of ion-pairing agents, because of their adverse effects on the process of ion formation, an alternative method based on the conversion of the carboxylic groups into methyl esters, which resulted in successful separation and a 10-fold enhancement of the sensitivity of the ESI-MS detection, was proposed.

During the 1990s, LC-MS and LC-MS/MS have become a routine method for analysis of alkylphenolic nonionic surfactants and their potentially estrogenic degradation products. In normal-phase systems, the APEOs are separated according to the increasing number of ethylene oxide units, while corresponding oligomers with the same number of ethoxy units but different alkyl substituents (e.g., nonylphenol ethoxylates (NPEOs) and octylphenol ethoxylates) coelute. Normal-phase LC-ESI-MS was used for the quantitative determination of individual NPEOs since complete chromatographic separation of NPEO oligomers is achieved using gradient elution with nonpolar solvents. However, such mobile phases are not compatible with ESI, and post-column addition of a polar solvent and a modifier is required to facilitate ionization of the target analytes and thus to enhance the signal and system stability. On the other hand, RPLC allows separation according to the character of the hydrophobic moiety and it is particularly well suited to separate surfactants containing various hydrophobic moieties (separation of alkyl homologs). In this case, the length of the ethylene oxide chain does not influence the separation and the various oligomers containing the same hydrophobic moiety elute in one peak.

The typical MS spectrum of polyethoxylates yields a characteristic pattern of equally spaced signals with mass differences of 44 Da (one ethylene oxide unit), which is a diagnostic fingerprint for this group of compounds. The nonionic surfactants easily form adducts with alkaline and other impurities. Using an ESI interface and aprotic solvent APEOs predominantly give evenly spaced sodium adducts  $[M + Na]^+$  due to the ubiquity of sodium in the solvents and surfaces. Sodium adducts are relatively

stable and generally no further structurally significant fragmentation is provided in the mass spectrum. However, formation of distinct adducts with ions originating from the buffer, the sample, and/or the introduction system (e.g.,  $H^+$ ,  $Na^+$ ,  $K^+$ ), water clusters (especially when using APCI interface), dimeric complexes, and doubly charged ions, such as disodium adducts, are also reported.

When using MS/MS and more selective multiple reaction monitoring detection, it is recommended that the formation of sodium adducts is suppressed, as these fragment poorly. ESI-MS/MS permits unambiguous identification and structure elucidation under negative ionization conditions of acidic alkylphenolic compounds (i.e., APECs) and fully deethoxylated alkylphenols. An example of MS/MS spectra of NP and NP<sub>1</sub>EC is shown in Figure 3.

**See also:** **Air Analysis:** Sampling; Outdoor Air; Workplace Air. **Dioxins.** **Headspace Analysis:** Static; Purge and Trap. **Isotope Dilution Analysis.** **Liquid Chromatography:** Overview. **Mass Spectrometry:** Overview. **Pesticides.** **Surfactants and Detergents.** **Water Analysis:** Overview; Industrial Effluents.

## Further Reading

- Badman ER and Cooks RG (2000) Miniature mass analyzers. *Journal of Mass Spectrometry* 35: 659–671.
- Bouyssiere B, Szpunar J, and Lobinski R (2002) Gas chromatography with inductively coupled plasma mass spectrometric detection in speciation analysis. *Spectrochimica Acta Part B* 57: 805–828.
- Clement RE, Yang PW, and Koester CJ (2001) Environmental analysis. *Analytical Chemistry* 73: 2761–2790.
- Eljarrat E and Barceló D (2002) Congener specific analysis of dioxins and related compounds by gas chromatography coupled to LRMS, HRMS, MS-MS and TOF-MS. *Journal of Mass Spectrometry* 37: 1105–1117.
- Petrovic M and Barceló D (2001) Analysis of ethoxylated non-ionic surfactants and their metabolites by liquid chromatography–atmospheric pressure ionization mass spectrometry. *Journal of Mass Spectrometry* 36: 1173–1185.
- Petrovic M, Eljarrat E, López de Alda MJ, and Barceló D (2002) Recent advances in the mass spectrometric analysis related to endocrine disrupting compounds in aquatic environmental samples. *Journal of Chromatography A* 974: 23–51.
- Santos FJ and Galceran MT (2003) Modern developments in gas chromatography–mass spectrometry-based environmental analysis. *Journal of Chromatography A* 1000: 125–151.

## Food Applications

**F A Mellon**, Institute of Food Research, Norwich, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Mass spectrometry (MS) is a very versatile technique with an established and rapidly growing record of successful applications in the food and nutrition sciences. Major advances in organic MS over the last decade, include the refinement and widespread adoption of electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), and matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) MS. These techniques have substantially extended the scope of mass spectrometric applications in the food and nutrition sciences. Furthermore, alternative mass spectrometric techniques, including isotope ratio MS (CIRMS), inductively coupled plasma MS (ICP-MS), proton transfer reaction MS (PTR-MS), and accelerator MS (AMS) are having an increased impact in the study of food authenticity, nutrient metabolism, and, in the case of PTR-MS, flavor research. This article will briefly review applications of MS in the analysis of a wide range of food components, including nutrients, structural constituents of foods, flavors, biologically active nonnutrients, and toxicants.

### Sample Preparation

Sample preparation techniques vary from rudimentary to elaborate because of the very wide range of analytical problems presented in food analysis and research. Preparation methods are highly dependent on both the nature of the analyte and on the type of mass spectrometric procedure used. At one extreme, the processing of food samples prior to pyrolysis MS (PyMS) may entail little more than the drying of a few micrograms of material, followed by introduction of the sample into the mass spectrometer on a pyrolysis probe. Flavor components are frequently isolated by solvent extraction, headspace sampling, or solid-phase microextraction (SPME), or in the case of PTR-MS analysis, simply by direct sampling of volatile components of the analyte. In contrast, linkage analysis of cell wall polysaccharides in food plants requires extensive chemical derivatization before samples are analyzed by gas chromatography-mass spectrometry (GC-MS). Combined liquid chromatography-mass spectrometry (LC-MS) and tandem mass spectrometry (MS/MS or MS<sup>n</sup>)

can sometimes be conducted on samples that have undergone minimal preparation. Finally, sample preparation for inorganic MS, zinc or chromium in food or biological fluids, for example, can be highly demanding because it is essential to avoid contamination.

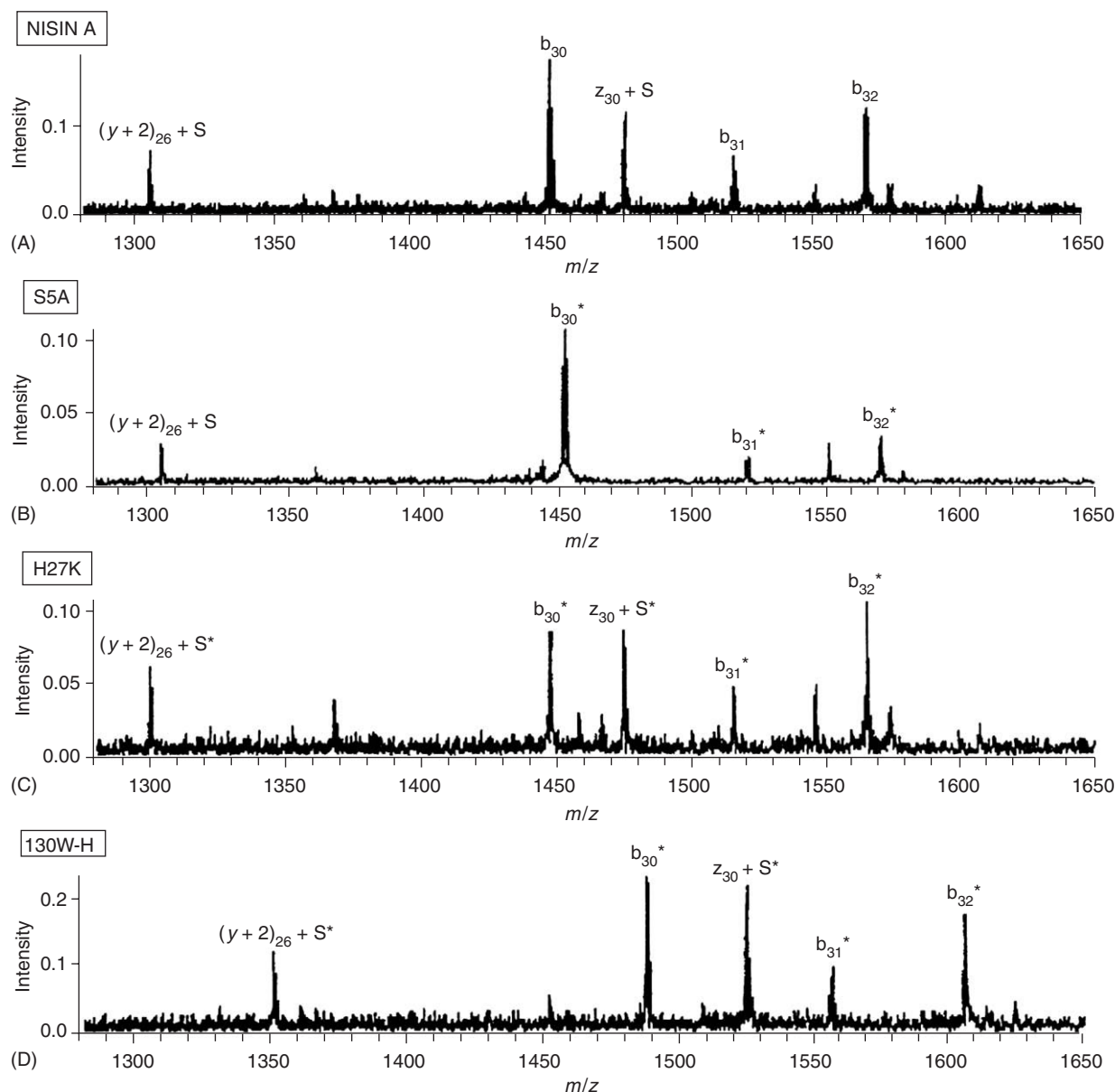
### Major Food Components

#### Amino Acids, Peptides, and Proteins

The analysis of amino acids by GC-MS of suitable volatile derivatives is a well-established technique. Despite the availability of amino acid analyzers, GC-MS is still occasionally used because it is useful for qualitative and quantitative analysis of unusual (i.e., nonprotein) amino acids. GC-MS is also valuable for conducting studies of the racemization of amino acids during cooking or food processing, when used in conjunction with chiral GC columns and deuterium labeling. GC-MS methods are also often used for determining stable isotope labeled amino acids in nutritional, metabolic studies. ESI LC-MS may also be used for determining amino acids.

Peptides and proteins are of interest in several areas of food and nutrition science, including allergenic response, food material properties, flavorings, toxicology, and nutritional value. An increasing number of analytical papers in the food and nutrition field now describe applications of modern MS techniques such as ESI and MALDI-TOF for molecular weight determination or (in MS/MS mode) full or partial sequencing. The variety of applications appearing in the current literature is also impressive: representative examples include the effects of irradiation on egg proteins, pH-related conformational changes in egg lysozyme, measurement of lactosylation of whey proteins, determination of the heterogeneity of casein proteins, authentication of cheeses by determining protein profiles, characterization of flavor peptides in protein hydrolysates, quantification of wheat gliadins, characterization of soya and other legume proteins, and the characterization of the food preservative peptide antibiotic nisin, its genetically modified variants and its inactivation by glutathione by Fourier transform ion cyclotron resonance (FT-ICR) MS and MALDI-TOF, respectively. **Figure 1** shows a comparison of the sustained off-resonance irradiation collisionally activated decomposition (SORI-CAD) spectra of 'wild-type' nisin A and three genetically modified variants. The ions shifted in mass (marked by an asterisk) relative to the wild-type





**Figure 1** Comparison of parts of the SORI-CAD ESI FT-ICR mass spectra of the  $[M + 3H]^{3+}$  precursor ions of (A) nisin A (wild-type nisin), (B) nisin S5A, (C) nisin H27K, and (D) nisin I30W-H, showing the region containing doubly charged fragment ions. The ions in the transgenic variants that are shifted in mass compared to nisin A are marked with asterisks. (Reprinted from Lavanant H, Heck A, Derrick PJ, *et al.* (1998) Characterization of genetically modified nisin molecules by Fourier transform ion cyclotron resonance mass spectrometry. *European Mass Spectrometry* 4: 405–416, with permission; © IM Publications.)

molecule are clearly visible and were used to aid structural confirmation of the variant molecules.

### Carbohydrates and Sugars

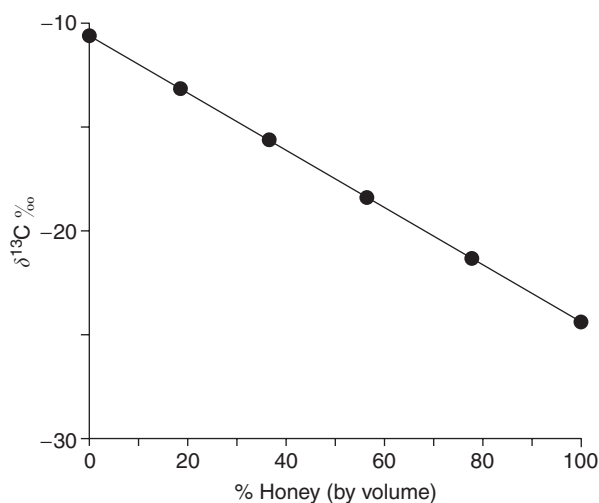
Carbohydrate analysis is of great importance in the food sciences because of the significant role of polysaccharides as macronutrients, as major constituents of dietary fiber, and as food structure components contributing to textural properties, and food additives. Because of structural similarities between

many monosaccharide residues, detailed mass spectrometric analysis is more difficult than for other biopolymers. Furthermore, the anomeric configuration and the nature of the linkages between sugar residues cannot generally be determined directly by MS. Many different schemes are available to convert sugars to derivatives sufficiently volatile for GC–MS analysis. These often involve trimethylsilylation or a combination of this derivatization technique with oxime formation. A classic method for compositional and linkage analysis of large polysaccharides

involves GC-MS of partially methylated alditol acetates (PMAAs). Polysaccharides are methylated, hydrolyzed, and reduced with NaBD<sub>4</sub> to deuterated O-methyl alditols, which are then acetylated and analyzed qualitatively and quantitatively by GC-MS. PMAAs yield characteristic mass spectra that allow determination of the number, nature, and linkage position of the monosaccharides that comprise the polysaccharide.

The molecular weights of large polysaccharides can be measured with the aid of ESI, in LC-MS, capillary electrophoresis/mass spectrometry (CE-MS), or direct injection modes, and by MALDI MS. For example, MALDI-TOF analysis of fructooligosaccharides has enabled the development of a rapid method for analyzing these molecules in food plants. Examples of other important applications of organic MS in food carbohydrate analysis include determination of the structures of plant oligosaccharides that may be involved in eliciting allergic response and direct analysis of plant cell wall xyloglucans.

Concerns over the adulteration of high value foods, for example, dilution of maple syrup, fruit juices, or of protected denomination of origin honeys with corn syrup, have been addressed for some time by combusting food samples and then measuring <sup>13</sup>CO<sub>2</sub>/<sup>12</sup>CO<sub>2</sub> ratios by high-precision IRMS. Subtle differences in isotope ratio are found in sugars derived from different sources and this property can be used to confirm their origin. Figure 2 exemplifies this by demonstrating the linear relationship between δ<sup>13</sup>C‰ and the ratio of an adulterant sugar, high fructose corn syrup, to honey sugar.



**Figure 2** Plot of changes in <sup>13</sup>CO<sub>2</sub>:<sup>12</sup>CO<sub>2</sub> ratio (δ<sup>13</sup>C‰, defined in text) of honey against concentration of high fructose corn syrup adulterant. (Reproduced with permission from PDZ Europa, Norwich, UK.)

δ<sup>13</sup>C‰ is a measure of natural carbon-13 levels against a reference standard and is defined by the equation:

$$\delta^{13}\text{C}\text{‰} = \left[ \frac{{}^{13}\text{C}/{}^{12}\text{C}_{\text{sample}}}{{}^{13}\text{C}/{}^{12}\text{C}_{\text{standard}}} - 1 \right] \times 1000$$

## Lipids

Dietary lipids may be analyzed mass spectrometrically as long-chain fatty acids, esters, triglycerides, and phospholipids. Although the degree of unsaturation of fatty acids may be determined readily from electron ionization (EI) mass spectra, double bond location presents a more challenging problem. Long-chain unsaturated fatty acids rearrange extensively under EI, so fragment ions do not yield reliable information regarding double bond position. A number of strategies have been devised to overcome these problems: one of the most popular is derivatization followed by GC-MS. Several different procedures are available, for example, oxidation followed by silylation, epoxidation, and the analysis of methoxy or dimethyldisulfide derivatives. Alternatively, chemical derivatives that localize charge may be generated, thus preventing charge-induced double bond migration. A good example is the formation of volatile 2-substituted 4,4-dimethyloxazoline derivatives by condensation of the fatty acid with 2-amino-2-methylpropanol. The 'charge-localizing' derivative is sufficiently volatile for GC-MS analysis and yields ion clusters separated by 14 Da, reducing to 12 Da when a double bond interrupts the chain. Alternative techniques for locating double bond position include high collision energy 'charge-remote' MS/MS fragmentation of molecules ionized by negative-ion chemical ionization (CI). Possible alternative techniques employ ESI of lithiated adducts of unsaturated fatty acids and low-energy MS/MS and acetonitrile CI ion trap MS/MS of polyunsaturated fatty acid methyl esters.

Capillary GC-MS is useful for identifying the carbon and unsaturation number of acylglycerols. However, APCI LC-MS can be used to analyze less volatile acylglycerols that may be unsuitable for GC-MS. APCI LC-MS has become the method of choice for qualitative and quantitative analysis of acylglycerols and, when combined with MS/MS, is capable of distinguishing fatty acid chains in the sn-2 position from those in the sn-1/3 positions.

The sensitivity, selectivity, and convenience of positive and negative-ion ESI LC-MS makes this the current benchmark technique for analyzing phospholipid mixtures. Additional structural information can be obtained by MS/MS. Where these advanced

techniques are unavailable, a more laborious approach of chromatographic class separation, chemical degradation, and GC-MS analysis of fatty acid can be used to generate useful structural information.

### Complete Foods and Food-Related Materials

PyMS is capable of generating useful data on intact foodstuffs or associated microorganisms (food poisoning bacteria, for example). PyMS is based on the controlled thermal degradation of samples under inert conditions to produce mixtures of volatile compounds that are swept into the mass spectrometer ion source and ionized by EI or CI. The resulting 'fingerprint' spectrum of the analyte contains characteristic features that can be classified by chemometric methods. Pyrolysis is sometimes combined with GC-MS in order to extract more information from the analyte. Applications of PyMS in the food sciences include quality assurance and authentication of food and drinks, analysis of cell wall material in food plants, and identification of food microorganisms.

Direct analysis of microorganisms, including food poisoning microorganisms, has also been conducted by MALDI-TOF MS. This promising technique can be used to characterize bacteria rapidly by genus, species, and strain.

### Flavors and Taints

Because most flavor components are highly volatile, EI and CI GC-MS have been primary techniques in flavor analysis and research since the early 1960s. For example, ~1000 discrete compounds have been identified in coffee volatiles using capillary GC-MS methods.

The ease with which SPME GC-MS studies may be conducted has made this an important and useful technique for sampling flavors and taints. GC-MS has been supplemented by LC-MS techniques for studying involatile flavor precursors or semivolatile or involatile food components that have important flavor characteristics.

The availability of a large knowledge base of flavor profiles (largely defined by GC-MS analysis) and recent instrumental advances have resulted in a shift in emphasis of mass spectrometric applications in flavor research. Instead of characterizing complex mixtures of volatiles, several researchers are now focusing on flavor release and, more specifically, on sampling volatiles released into the mouth and nose. By using a special APCI probe coupled to a mouth or nose piece, it is possible to conduct dynamic, breath-by-breath analyses of air expired during eating. The recent development of PTR-MS for online trace gas monitoring has considerable potential for application

to flavor research because of its ability to sample air directly. PTR-MS has already been applied to analysis of flavor compounds. Other food-related applications include the control of food production by determining volatile organic compounds produced during fruit ripening and aging, the study of coffee volatiles and monitoring of meat degradation.

Mass spectrometric methods are also useful for authenticating flavor components, using methods for accurately measuring  $^{13}\text{CO}_2/^{12}\text{CO}_2$  ratios, similar to those already described for authenticating sugars (above). If more detailed analysis is required GC/combustion/isotope ratio MS will yield accurate isotope ratios on individual components of a flavor sample.

## Nutrition Studies

### Stable Isotope Methods

Interest in the use of stable isotope MS for studying both the nutritional value of foods and diets and fundamental aspects of nutrient metabolism in humans has increased considerably. The only major drawbacks of stable isotope studies are associated with the presence of endogenous isotopes of the elements under investigation. Sufficient label must be administered to generate a measurable increase in isotope ratio. However, care is needed to ensure physiological dosing as administration of excessive doses yield physiologically contentious data. Conversely, administration of small quantities of label requires careful and accurate measurement of isotope ratios because the enriched material is diluted by the endogenous nutrients. In the case of mineral metabolism studies, additional precautions are necessary because of the possibility of contamination during sample processing.

Stable isotope studies usually involve administration of the enriched stable isotope in or with a meal. The method of labeling depends on the type of study undertaken. For elements such as selenium, where absorption and metabolism are highly dependent on chemical form, it is usually necessary to use an intrinsic label (i.e., one that is biosynthetically incorporated into the food). Conversely, minerals believed to form a 'common pool' in the digestive system may be mixed directly with the food (extrinsic labeling). In some cases a second isotope is injected or infused intravenously to correct for endogenous losses. Samples of breath, blood, urine, saliva, or feces are then collected for an appropriate period and subjected to isotopic analysis. Isotope ratio measurement often requires specialized instrumentation such as

high-precision IRMS or dedicated inorganic instruments such as ICP-MS.

### Mineral Nutrients

Although several different mass spectrometric methods have been deployed to determine enriched stable isotopes in human studies of nutrient mineral metabolism, thermal ionization mass spectrometry (TIMS) and particularly ICP-MS are now used almost exclusively. ICP-MS is rapid, very sensitive, and sample preparation and introduction is often simplified. Furthermore, ICP-MS can be coupled directly to separation techniques such as size-exclusion chromatography (SEC), high-performance liquid chromatography (HPLC), or CE so that speciation, the determination of the chemical form of particular elements, may also be studied. The two major drawbacks of ICP-MS, low precision relative to TIMS and interference from polyatomic ions in the argon plasma, have largely been overcome by new generations of instruments equipped with multiple collectors and collision/reaction cells, respectively.

A wide range of human studies has been conducted using enriched stable isotopes of nutrient minerals. Applications include the determination of the absorption and metabolism of iron, zinc, calcium, copper, selenium, and molybdenum. An example of the type of information that can be obtained is provided by a study of iron absorption from different weaning foods, and the effects of vitamin C on iron absorption. These measurements, conducted using enriched stable isotopes of  $^{57}\text{Fe}$  and  $^{58}\text{Fe}$ , demonstrated a doubling of iron absorption when a drink containing 50 mg of vitamin C was administered with the food.

A recent development, especially useful for conducting long-term studies of the effect of diet on metabolism, is to administer extremely low levels of long-lived radioisotopes, e.g.,  $^{41}\text{Ca}$ , that are then measured by AMS. Because the activity of the radioisotope is extremely low (typically  $<10^{-6}$  of annual background radiation dose), it is considered to be safe for human use. AMS opens up the exciting possibility of conducting long-term metabolic and dietary intervention studies where required; for example, in studies of bone metabolism.

### Vitamins

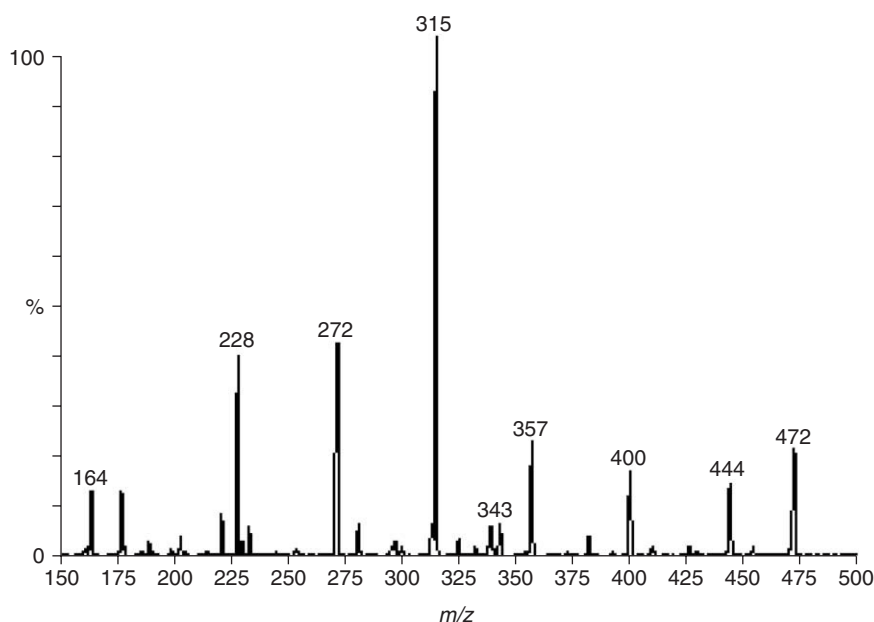
Although vitamins can be determined, both qualitatively and quantitatively, by MS, routine analysis is usually best conducted by other means (e.g., HPLC with ultraviolet (UV) or fluorescence detection, immunoassay methods, or microbiological methods). Analytically, MS does have an important role as a reference technique, especially when used in isotope

dilution-based analyses, for evaluating and calibrating alternative, non-MS techniques. However, the most prominent role for MS in vitamin research is in studies of the metabolism of vitamins in humans, especially by LC-MS or LC-MS/MS. The newer LC-MS techniques of APCI and ESI are particularly useful in this respect because many vitamins are too labile for EI and CI MS.

Vitamins  $\text{D}_2$  and  $\text{D}_3$  and their major metabolites have been studied extensively by GC-MS of volatile derivatives; these studies include quantitative determination by isotope dilution. More recently, ESI LC-MS and LC-MS/MS have been used in the qualitative and quantitative measurement of vitamin D and vitamin D analogs.

The association of folates with reduced chronic disease risk and prevention of neural tube defects has generated increased interest in studies of the metabolism of these B group vitamins. Recent activity has focused on developing improved mass spectrometric methods for determining the absorption, metabolism, and bioavailability of these molecules. The folates are polar, involatile molecules that are unstable in solution and must be derivatized if they are to be analyzed by EI or CI MS. Pioneering studies of the human metabolism of folates were first conducted by selected ion monitoring (SIM) GC-MS of derivatized extracts from biological fluids. Although the GC-MS method has provided useful data on the kinetics of folate metabolism and on urinary excretion of folate and metabolites, the method has several drawbacks. Improvements both in mass spectrometric techniques and in speed of analysis were clearly desirable and a number of groups have focused on developing LC-MS methods. Negative-ion ESI techniques developed to determine the four main food and supplementary folates, folic acid, 5-methyltetrahydrofolic acid, tetrahydrofolic acid, and folinic acid, in selected foodstuffs and a vitamin supplement demonstrated the feasibility of using ESI LC-MS in folate analysis. Other laboratories have subsequently reported successful measurement of stable isotope labeled folates in human plasma and/or urine by positive-ion ESI LC-MS, negative-ion LC-MS/MS (in multiple reaction monitoring mode), and SIM negative ion LC-MS. Limits of quantification were sufficient to conduct human metabolic studies using  $^{13}\text{C}$ ,  $^{15}\text{N}$ , and  $^2\text{H}$  folates, labeled both extrinsically and intrinsically, on spinach and fortified cereal grain. The low-collision-energy negative-ion tandem ESI mass spectrum of 5-formyl tetrahydrofolate is shown in Figure 3.

AMS has also been used to determine the metabolic fate of folates labeled with low levels of the radioactive isotope  $^{14}\text{C}$ .



**Figure 3** The MS/MS product ion spectrum of  $m/z$  472, the  $[M - H]^-$  ion of 5-formyl tetrahydrofolate, at a collision energy of 35 V. (© Institute of Food Research, reproduced with permission.)

The ability to conduct bioavailability studies on foods labeled intrinsically with stable isotopes is particularly important as this type of labeling is the 'gold standard' for metabolic studies, i.e., intrinsic label should mimic the behavior of endogenous nutrients most closely. However, the mass spectrometric measurement of the metabolism of intrinsically stable isotope labeled materials is far more challenging than measuring extrinsically labeled nutrients. Low levels of isotope incorporation must be determined in small samples and it is only quite recently that modern LC-MS and LC-MS/MS methods, as demonstrated by the folate studies described above, have begun to rise to this challenge.

Vitamin A is a generic descriptor for a family of fat-soluble vitamins that has the biological activity of retinol, one of the most active and bioavailable members of the group. Retinol, retinoids, and their derivatives yield characteristic EI spectra, a property that has been exploited for over 25 years by GC-MS methods that determine liver stores of retinol non-invasively, following oral administration of deuterated retinol. A simple mathematical formula derives liver stores of retinol from isotope composition in plasma. Various EI and positive ion CI GC-MS procedures have been used; the most successful techniques involve measurement of silylated derivatives of retinol. Although these methods have provided useful nutritional data, further improvements in sensitivity were still desirable for conducting serial infant studies or when sampling under nonideal field conditions. A technique based on negative ion

electron capture GC-MS of trimethylsilyl derivatives fulfills these increased demands and yields. The method worked on 200  $\mu$ l blood samples, the amount of blood typically collected in a heel-prick sample from an infant, and yielded measurements on enrichments as low as 0.01% of circulating retinol- $D_8$ . More recently, both APCI LC-MS and LC-MS<sup>n</sup> methods have been developed to measure retinol conjugates such as retinyl palmitate (which appears in the blood soon after vitamin A containing meal is consumed). These techniques have advantages over GC-MS methods because there is no need to hydrolyze samples to release free retinol, or for derivatization for GC-MS. Carotenoid (pro-vitamin A) metabolism is also being studied by LC-MS techniques because of epidemiological evidence of their putative role in cancer prevention. Although ESI LC-MS has shown promise in preliminary mass spectrometric studies, APCI LC-MS is now the method of choice because of its robustness and tolerance of a wider range of organic solvents.

Many other vitamins have been determined by mass spectrometric methods; representative examples include APCI LC-MS/MS measurement of tocopherols and LC-MS measurement of K vitamins.

## Trace Components in Foods

### Food Additives

Food additives comprise a wide variety of compounds that are generally monitored by techniques



other than MS; however, occasional quantitative applications do occur, exemplified by GC–MS determination of antioxidants in stored products.

### Biologically Active Nonnutrients (Dietary Phytochemicals)

Biologically active (bioactive) nonnutrients in foods comprise a wide range of phytochemical substances. Lack of space permits discussion of mass spectrometric analysis of only a small range of these compounds here.

Glucosinolates (Figure 4) and their biologically active breakdown products are found in many plant foods. More than 100 different types of glucosinolate have been isolated from plants. Although some glucosinolates may have toxic (e.g., goitrogenic) properties, they are known to be potent inducers of Phase II enzymes that protect against carcinogens and other toxic electrophiles.

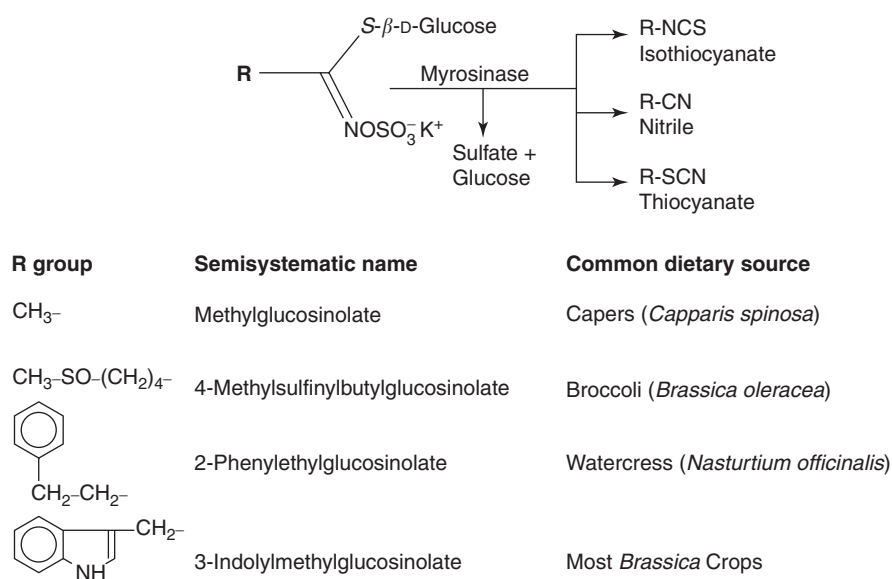
The structural variability is in the aglycone, R that may comprise linear or branched alkyl and alkenyl side chains, alcohols, methylthioalkyl, methylsulfinyl, aralkyl, or heterocyclic groups. GC–MS is a useful mass spectrometric technique for analyzing volatile glucosinolate breakdown products. Negative-ion ESI LC–MS is now the method of choice for determining intact glucosinolates and positive ion APCI LC–MS for analyzing the more thermally labile breakdown products (e.g., sulforaphane) and their human metabolites. ESI LC–MS and MS/MS are also useful for conducting metabolic studies, for example, by detecting and determining glutathione

conjugates of isothiocyanate breakdown products of glucosinolates.

Phenolics are a distinctive feature of all plant tissues and are of interest because they can affect palatability, taste, nutritional value, and particularly the health properties of foods. For example, the flavonoids occur very widely in plants and there is great interest in their role as protective factors in the diet. More recently, the modern methods of LC–MS analysis, APCI and electrospray, have been applied to the detection and quantification of flavonoids and isoflavonoids. Although early mass spectrometric techniques for analyzing these molecules focused on GC–MS methods (after appropriate derivatization), the newer LC–MS methods are now widely used. For example, isoflavones and their conjugates have been determined in soy foods by positive and negative ion APCI LC–MS. ESI and APCI LC–MS methods are also now used for conducting studies of flavonoid metabolism, by following the disappearance and deconjugation of flavonoid glycosides and the appearance of glucuronides, sulfates, and methylated metabolites.

### Natural Toxicants

Mycotoxins secondary metabolites produced by *Aspergillus* (aflatoxins) and *Fusarium* (tricothecenes, etc.) food-spoilage molds comprise two of the most prominent groups of natural toxicants. Many different mass spectrometric techniques, including EI, positive and negative ion CI, GC–MS, LC–MS, supercritical fluid chromatography/mass spectrometry, and



**Figure 4** General structure of the glucosinolates and their common myrosinase hydrolysis products. (Reproduced with permission from Institute of Food Research; © Institute of Food Research.)

MS/MS have been devised to monitor the levels of mycotoxins in foods, body fluids, and tissues. The most recent developments include ESI LC-MS determination of aflatoxins down to low picogram levels. Additional selectivity was provided by MS/MS selected reaction monitoring. Examples of other naturally occurring food toxicants determined with the aid of MS include mutagenic compounds related to quinoxaline that may be formed in cooked meats. The most notorious recent example of a food natural toxicant is the discovery, made with the aid of GC-MS, ESI LC-MS, and LC-MS/MS, that the genotoxic carcinogen acrylamide is formed during some cooking processes.

### Anthropogenic Toxicants

Many examples of mass spectrometric methods for determining toxic or potentially deleterious anthropogenic compounds in foods can be found in the scientific literature. The range of compounds analyzed is wide and includes dioxins, polyaromatic hydrocarbons, pesticide and veterinary drug residues, plasticizers from packaging materials, and environmental contaminants. LC-MS, MS/MS, GC-MS, and GC-MS/MS techniques are used widely for determining these compounds down to parts per billion or even parts per trillion levels. For example, polar organophosphorus pesticides can be quantified by LC-MS/MS down to levels of  $0.01 \text{ mg kg}^{-1}$  in fruits and vegetables and dioxins are regularly determined in food matrices at femtogram levels.

Isotope dilution mass spectrometry is an accurate and sensitive technique for determining toxic trace elements in food matrices. Lead, cadmium, and thallium have been analyzed rapidly down to very low levels by ICP-MS. The latter technique is particularly useful for simultaneous measurement of a wide range of elements. Because the toxicity of an element can be highly dependent on its chemical form, ICP-MS is also useful in the speciation of toxic minerals in foodstuffs by combination with HPLC or SEC.

### Future Trends

Enormous advances in the scope and sensitivity of mass spectrometric techniques have occurred in the last decade. The use of modern techniques of organic MS, particularly APCI and ESI LC-MS, CE/MS, and MS/MS, has grown apace in the food and nutritional sciences. Applications of MALDI-TOF in the food sciences are also increasing, although at a lower rate (the frequent need for chromatographic separation in food-related applications has led to an understandable bias toward LC-MS methods).

The importance of ICP-MS in multielement analysis of foodstuffs is already well established, as is the use of this technique in human metabolic studies. Both types of application are being augmented by increased applications of combined chromatography/ICP-MS for conducting speciation studies. The advent of a new generation of high-precision isotope ratio ICP-MS instruments also opens up the exciting possibility of conducting metabolic studies without using enriched labels, by observing isotopic fractionation between different body compartments. High-precision ICP-MS also has considerable potential in food authentication (by linking isotopic composition to geographical origin, for example).

Fundamental studies of the metabolism of proteins, fats, and starches by GIR-MS are increasing and it is anticipated that practical applications to the metabolism of extrinsically labeled foods will make an impact in the future.

Mass spectrometry has a central role in the postgenomic science of proteomics, the qualitative and quantitative comparison of the entire protein complement of a genome under different conditions. The new science of metabolomics, the global, quantitative analysis of all the low and intermediate molecular weight metabolites expressed by a genome under specific conditions, has obvious applications in food science, for example, in assessing the safety of transgenic plants or in studying environmental effects on food poisoning microorganisms and mass spectrometry is one of the pivotal analytical techniques in metabolomics. This type of application is expected to increase in the future.

**See also:** **Carbohydrates:** Overview. **Elemental Speciation:** Overview. **Food and Nutritional Analysis:** Overview. **Gas Chromatography:** Mass Spectrometry. **Lipids:** Overview. **Liquid Chromatography:** Liquid Chromatography-Mass Spectrometry; Food Applications. **Mass Spectrometry:** Overview; Principles; Ionization Methods Overview; Atmospheric Pressure Ionization Techniques; Electrospray; Matrix-Assisted Laser Desorption/Ionization; Pyrolysis. **Proteins:** Traditional Methods of Sequence Determination. **Vitamins:** Overview.

### Further Reading

- Alomirah HF, Alli I, and Konishi Y (2000) Applications of mass spectrometry to food proteins and peptides. *Journal of Chromatography A* 893: 1-21.
- Byrdwell WC (2001) Atmospheric pressure chemical ionization mass spectrometry for analysis of lipids. *Lipids* 36: 327-346.
- Careri M, Bianchi F, and Corradini C (2002) Recent advances in the application of mass spectrometry in

- food-related analysis. *Journal of Chromatography A* 970: 3–64.
- Dennis MJ (1998) Recent developments in food authentication. *Analyst* 123: 151R–156R.
- Fenselau C and Demirev PA (2001) Characterization of intact microorganisms by MALDI mass spectrometry. *Mass Spectrometry Reviews* 20: 157–171.
- Garrido-Frenich A, Arrebola FJ, Gonzalez-Rodriguez MJ, Vidal JLM, and Diez NM (2003) Rapid pesticide analysis in post-harvest plants used as animal feed by low-pressure gas chromatography-tandem mass spectrometry. *Analytical and Bioanalytical Chemistry* 377: 1038–1046.
- Guillou C, Lipp M, Radovic B, *et al.* (1999) Use of pyrolysis mass spectrometry in food analysis: Applications in the food analysis laboratory of the European Commissions' Joint Research Centre. *Journal of Analytical and Applied Pyrolysis* 49: 329–335.
- Kuiper HA, Kleter GA, Noteborn HPJM, and Kok EJ (2001) Assessment of food safety issues related to genetically modified foods. *Plant Journal* 27: 503–528.
- Lindinger W, Hansel A, and Jordan A (1998) On-line monitoring of volatile organic compounds at pptv levels by means of proton-transfer-reaction mass spectrometry (PTR-MS) medical applications, food control and environmental research. *International Journal of Mass Spectrometry and Ion Processes* 173: 191–241.
- Linforth RST, Ingham KE, and Taylor AJ (1996) *Flavour Science: Recent Developments*. Cambridge: Royal Society of Chemistry.
- Mellon FA, Self R, and Startin JR (2000) *Mass Spectrometry of Natural Substances in Food*. Cambridge: Royal Society of Chemistry.
- Pico Y, Blasco C, and Font G (2004) Environmental and food applications of LC-tandem mass spectrometry in pesticide-residue analysis: An Overview. *Mass Spectrometry Reviews* 23: 45–85.
- Van der Werf MJ, Schuren FHJ, Bijlsma S, Tas AC, and van Ommen B (2001) Nutrigenomics: Application of genomics technologies in nutritional sciences and food technology. *Journal of Food Science* 66: 772–780.
- Wolfender JL, Terreaux C, and Hostettmann K (2000) The importance of LC–MS and LC–NMR in the discovery of new lead compounds from plants. *Pharmaceutical Biology (Supplement)* 38: 41–54.

## Forensic Applications

**R J Lewis**, Civil Aerospace Medical Institute, Oklahoma City, OK, USA

**R H Liu**, Fooyin University, Kaohsiung Hsien, Taiwan

Published by Elsevier Ltd.

This article is a revision of the previous-edition article by R H Liu, pp. 2989–2996, © 1995, Elsevier Ltd.

## Introduction

Forensic science is the application of science to problems encountered in the courts of law. Forensic laboratories specialize in the analysis, identification, and interpretation of physical evidence. Typically, analyses performed in such laboratories are concerned with determining (1) component identification of a sample, (2) quantitation or purity of sample components, or (3) similarities or differences between two or more samples. With automated instrumentation widely available, highly specific mass spectrometry (MS)-based technologies are now the most valuable tools used to achieve these desired analytical goals.

MS-based techniques, in the forensic setting, are primarily used for the identification of specific components in drug-related samples. These samples may contain drugs in dosage forms or in biological matrices. The drugs analyzed may be abused by the general population, a specific age group, or those used as performance enhancers in human or animal

sporting events. MS-based methodologies are also routinely used for the analysis of materials related to arson, explosives, and synthetic polymers. Applications in the characterization of inorganic elements for forensic science purposes have also been reported.

Most MS applications utilize electron ionization (EI) or chemical ionization (CI), a quadrupole analyzer, and a gas chromatograph (GC) as the sample introduction device. However, other MS methodologies such as tandem MS (MS/MS) and isotope-ratio are becoming more common. Recent advances in sample evaporation/ionization, high-mass ion resolution, and liquid chromatography (LC)–MS interfacing technologies have significantly expanded the potential forensic applications of MS-based approaches to the analysis of biomolecules, i.e., proteomics.

## Applications Resulting from Advances in Ionization and Ion Resolution Technologies

With recent advances in matrix assisted laser desorption/ionization (MALDI) and electrospray LC–MS interface technologies (awarded the 2002 Nobel prize in chemistry), MS can now be successfully applied to the analysis of compounds with high molecular weights or polar functional groups. Time-of-flight (TOF) MS is typically used to resolve and

- food-related analysis. *Journal of Chromatography A* 970: 3–64.
- Dennis MJ (1998) Recent developments in food authentication. *Analyst* 123: 151R–156R.
- Fenselau C and Demirev PA (2001) Characterization of intact microorganisms by MALDI mass spectrometry. *Mass Spectrometry Reviews* 20: 157–171.
- Garrido-Frenich A, Arrebola FJ, Gonzalez-Rodriguez MJ, Vidal JLM, and Diez NM (2003) Rapid pesticide analysis in post-harvest plants used as animal feed by low-pressure gas chromatography-tandem mass spectrometry. *Analytical and Bioanalytical Chemistry* 377: 1038–1046.
- Guillou C, Lipp M, Radovic B, *et al.* (1999) Use of pyrolysis mass spectrometry in food analysis: Applications in the food analysis laboratory of the European Commissions' Joint Research Centre. *Journal of Analytical and Applied Pyrolysis* 49: 329–335.
- Kuiper HA, Kleter GA, Noteborn HPJM, and Kok EJ (2001) Assessment of food safety issues related to genetically modified foods. *Plant Journal* 27: 503–528.
- Lindinger W, Hansel A, and Jordan A (1998) On-line monitoring of volatile organic compounds at pptv levels by means of proton-transfer-reaction mass spectrometry (PTR-MS) medical applications, food control and environmental research. *International Journal of Mass Spectrometry and Ion Processes* 173: 191–241.
- Linforth RST, Ingham KE, and Taylor AJ (1996) *Flavour Science: Recent Developments*. Cambridge: Royal Society of Chemistry.
- Mellon FA, Self R, and Startin JR (2000) *Mass Spectrometry of Natural Substances in Food*. Cambridge: Royal Society of Chemistry.
- Pico Y, Blasco C, and Font G (2004) Environmental and food applications of LC-tandem mass spectrometry in pesticide-residue analysis: An Overview. *Mass Spectrometry Reviews* 23: 45–85.
- Van der Werf MJ, Schuren FHJ, Bijlsma S, Tas AC, and van Ommen B (2001) Nutrigenomics: Application of genomics technologies in nutritional sciences and food technology. *Journal of Food Science* 66: 772–780.
- Wolfender JL, Terreaux C, and Hostettmann K (2000) The importance of LC–MS and LC–NMR in the discovery of new lead compounds from plants. *Pharmaceutical Biology (Supplement)* 38: 41–54.

## Forensic Applications

**R J Lewis**, Civil Aerospace Medical Institute, Oklahoma City, OK, USA

**R H Liu**, Fooyin University, Kaohsiung Hsien, Taiwan

Published by Elsevier Ltd.

This article is a revision of the previous-edition article by R H Liu, pp. 2989–2996, © 1995, Elsevier Ltd.

### Introduction

Forensic science is the application of science to problems encountered in the courts of law. Forensic laboratories specialize in the analysis, identification, and interpretation of physical evidence. Typically, analyses performed in such laboratories are concerned with determining (1) component identification of a sample, (2) quantitation or purity of sample components, or (3) similarities or differences between two or more samples. With automated instrumentation widely available, highly specific mass spectrometry (MS)-based technologies are now the most valuable tools used to achieve these desired analytical goals.

MS-based techniques, in the forensic setting, are primarily used for the identification of specific components in drug-related samples. These samples may contain drugs in dosage forms or in biological matrices. The drugs analyzed may be abused by the general population, a specific age group, or those used as performance enhancers in human or animal

sporting events. MS-based methodologies are also routinely used for the analysis of materials related to arson, explosives, and synthetic polymers. Applications in the characterization of inorganic elements for forensic science purposes have also been reported.

Most MS applications utilize electron ionization (EI) or chemical ionization (CI), a quadrupole analyzer, and a gas chromatograph (GC) as the sample introduction device. However, other MS methodologies such as tandem MS (MS/MS) and isotope-ratio are becoming more common. Recent advances in sample evaporation/ionization, high-mass ion resolution, and liquid chromatography (LC)–MS interfacing technologies have significantly expanded the potential forensic applications of MS-based approaches to the analysis of biomolecules, i.e., proteomics.

### Applications Resulting from Advances in Ionization and Ion Resolution Technologies

With recent advances in matrix assisted laser desorption/ionization (MALDI) and electrospray LC–MS interface technologies (awarded the 2002 Nobel prize in chemistry), MS can now be successfully applied to the analysis of compounds with high molecular weights or polar functional groups. Time-of-flight (TOF) MS is typically used to resolve and

identify ions produced by MALDI, while conventional or tandem MS are used to identify LC eluates. While the analyzer in a tandem MS serves as a 'filter', the TOF detector collects all ions produced and is, therefore, theoretically more sensitive.

A recent report described MALDI-TOF MS characterization of ricin, an extremely potent poison from the seeds of the castor bean plant. Ricin has been mentioned in terrorism literature and was reportedly used in a high profile international spy homicide case. In the MS application, samples were first purified by LC fractionation or molecular weight cutoff filtration. Confirmation was carried out by comparing the MALDI-TOF MS spectrum of the tryptic digest against a standard and GC-MS identification of an alkaloid marker, ricinine.

In another forensic science application, electrospray LC-MS was used to determine elevated levels of insulin-like growth factor 1 (IGF-1) in serum as an indication of growth hormone (GH, somatotropin) abuse in racehorses. Structural identity of IGF-1 was confirmed by LC-MS and LC-MS/MS characterization of the endoproteinase Asp-N digestion product.

With improved interface technology, LC-MS is now commonly advocated for the analysis of smaller molecules with polar functional groups, especially drug metabolites and their conjugates. This approach can provide intact conjugate information not obtainable using parallel GC-MS approaches. It also bypasses the hydrolysis and chemical derivatization steps that are often required in GC-MS applications. The application of LC-MS to small, polar metabolites is exemplified in current studies investigating differential serotonin metabolism as a tool to differentiate between ethanol ingestion and postmortem ethanol formation in human specimens.

## Drug Analysis Using MS

Drug-related samples commonly encountered in forensic science laboratories may be present in dosage forms with varying degrees of purity, in plant materials or in biological matrices. Conventional EI-MS techniques are widely used to identify drugs in relatively pure forms, made possible either by prior purification procedures or via the use of a chromatographic device for sample introduction. CI techniques are also commonly employed to provide molecular weight information and occasionally to probe the number and the nature of active hydrogens attached to heteroatoms in the drugs of concern. Special chemical derivatizations in conjunction with GC-MS are effective in characterizing structural features of optical isomers of certain drugs, thus making possible their identification.

The use of an MS as the detector significantly improves the quantitative determination of drugs and their metabolites in biological matrices by making possible the use of isotope-labeled analogs (ILAs) of the analytes as internal standards (ISs). The most commonly employed ILAs are deuterated analogs. However, some studies have indicated that  $^{13}\text{C}$ -analogs may be more appropriate. Examples of these applications are presented below.

### Identification of Drugs and Drug Metabolite Using MS

Drug identifications are commonly carried out using both EI and CI MS. The most common applications involve the generation of a spectrum of the compound in question, similar to those shown in **Figure 1**, followed by a comparison against an authentic drug standard spectrum. An authentic spectrum can be obtained from a drug standard itself or retrieved from a library database. Useful compilations of mass spectra are given in the Further reading section.

Since authentic spectra are not always available, skillful and cautious interpretation of EI spectra can aid in the identification of an unknown compound. Mass spectrum interpretation of ephedrine-related compounds serves as an interesting example. Ephedrine and norephedrine derivatized with *N*-tri-fluoroacetyl-L-prolyl chloride (L-TPC) undergo an interesting migration/cleavage process (path A in **Figure 2**) in addition to the common ' $\beta$ -fission' process (path B in **Figure 2**), where the location of fragmentation is ' $\beta$ ' to the amino function and to the aromatic system. Path A is more favorable for norephedrine than for ephedrine as a result of the steric effect of ephedrine's methyl group. Path A is hindered if the hydroxyl hydrogen on norephedrine is replaced with a bulky addition such as the trimethylsilyl (TMS) group.

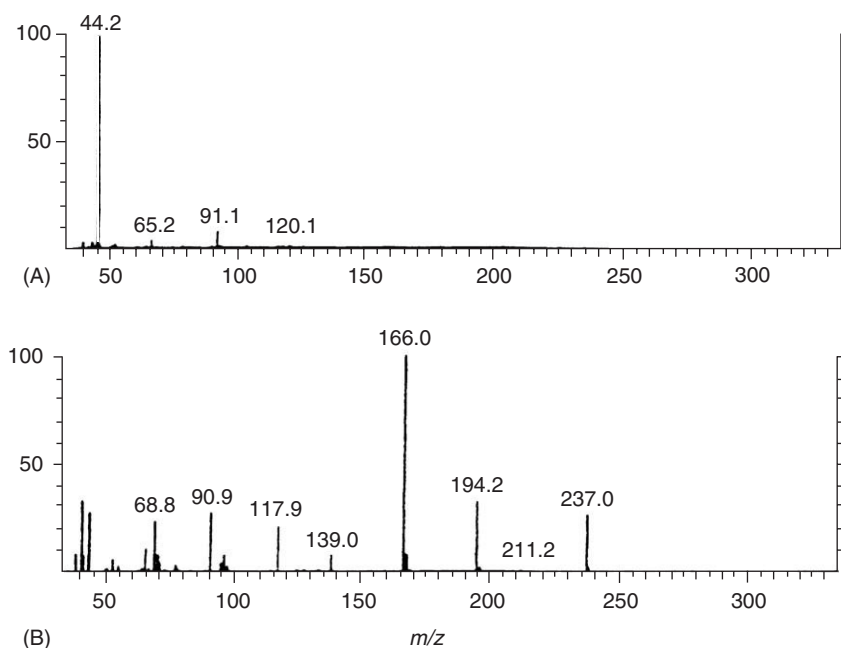
Chemical ionization MS, with various deuterated reagent gases, has been used to probe the number and nature of active hydrogens in a molecule. For example, based on the mass shifts of the 'protonated' ions that are observed when switching from ammonia ( $\text{NH}_3$ ) to deuterated ammonia ( $\text{ND}_3$ ), the number of active hydrogens in morphine and several other drugs have been demonstrated.

The total number of active hydrogens in a compound is derived from the observed mass shifts ( $n$ ) of the 'protonated' ion when  $\text{NH}_3$  is replaced with  $\text{ND}_3$  as the reagent gas:

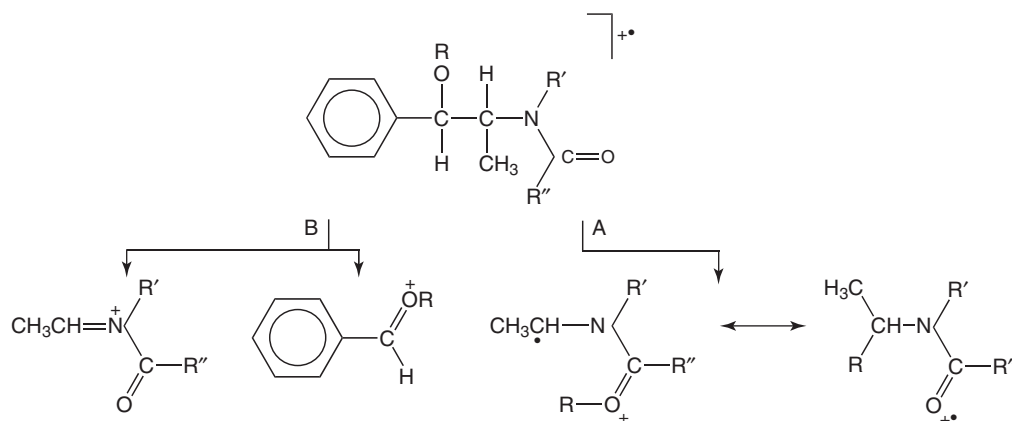
$$n = M_{\text{ND}_3} - M_{\text{NH}_3} - 1$$

where  $M_{\text{ND}_3}$  and  $M_{\text{NH}_3}$  are the mass to charge ratios of the 'protonated' ions observed with  $\text{ND}_3$  and  $\text{NH}_3$  as the reagent gases, respectively.





**Figure 1** Comparison of the EI mass spectra of amphetamine (A) and L-TPC derivatized amphetamine (B). (Reproduced with permission from Liu RH (1989) Applications of mass spectrometry to forensic sciences. In: Lee HC and Gaensslen RE (eds.) *Advances in Forensic Science*, vol. 2 – Instrumental Analysis, p. 75. Year Book Medical Publications: Chicago, IL.)



**Figure 2** Electron impact fragmentation pathways of ephedrine-related compounds. (Reproduced with permission from Gadzala DE, Liu RH, Legendre MG, and Ku WW (1988) Mass spectra of derivatized ephedrine and norephedrine. *Organic Mass Spectrometry* 23, 851–852.)

The combined information provided by (1) mass shifts, (2) the tendencies of the protonated and adduct ions to lose molecules of water, and (3) the intensity ratio of adduct ion to protonated ion, may further reveal the nature of the active hydrogens involved, particularly in differentiating a phenolic from an alcoholic hydrogen.

### Chemical Derivatization in MS

Chemical derivatization of drugs is commonly employed when conducting GC–MS analysis. In

addition to improving chromatographic and stability characteristics of the analytes, derivatized drugs may also produce characteristic mass spectra that are not available from parent compounds; thereby, better achieving the desired analytical objectives. Derivatization approaches also form the basis for optical isomer differentiations.

The detection limit and quantitation of compounds may be improved by derivatization, yielding intense high-mass ions not available from the parent compounds. In general, for any given compound the higher the mass of the ion monitored, the lower the

background noise. For example, the spectrum of amphetamine (**Figure 1A**) exhibits low intensities of ions at higher mass range, while L-TPC derivatized amphetamine shows significant ion intensity at  $m/z$  237. Considering the probability of contributions from interfering analytes, the base peak from un-derivatized amphetamine,  $m/z$  44, is not suitable for quantitative purposes and thus, requires derivatization.

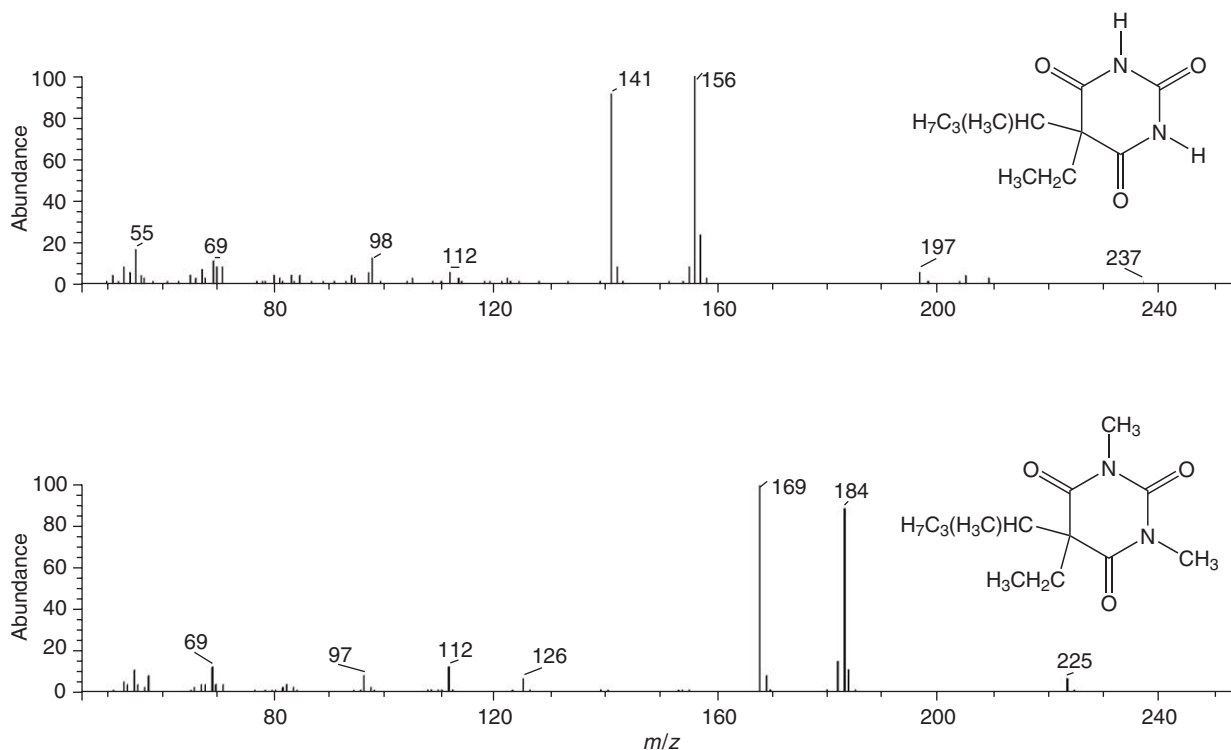
Detection limits can also be improved through the introduction of elements of high electron affinities into the compound and the monitoring of negative ions using negative CI-MS. This approach has been successfully applied in the analysis of cannabinoids in biological samples. Following derivatization with pentafluoropropionic anhydride (PFPA),  $\Delta^1$ -tetrahydrocannabinol has been quantitatively reported in the mid-pg/ml range in postmortem blood.

Mass shifts in the spectra produced by different derivatizing agents may provide extremely useful information for the elucidation of certain structures. For example, the number of TMS groups attached to the parent compound can often be deduced by observing the mass shifts that occur upon replacing *N,O*-bis-(trimethylsilyl)-acetamide (BSA) with  $d_9$ -BSA as the derivatizing agent. This information facilitates the identification of desoxymorphine-A,

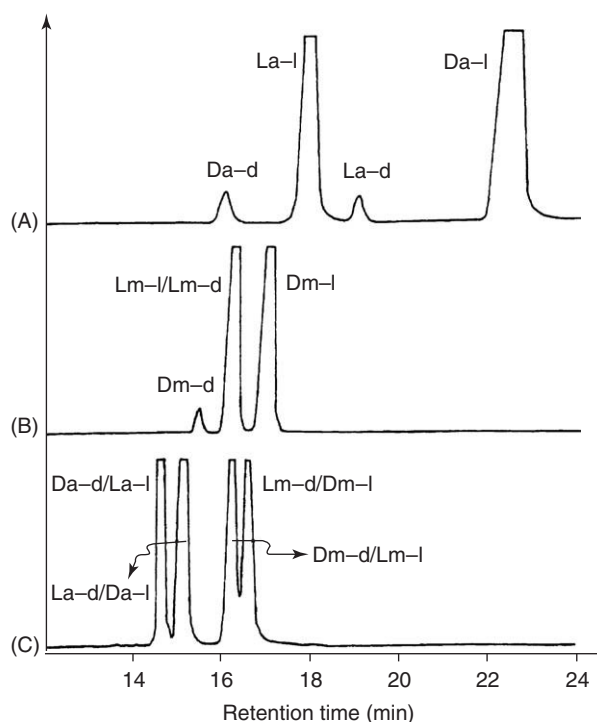
monoacetyldesoxy-morphine-A, and diacetyldesoxy-morphine-A as the impurities in an illicit heroin sample. The same approach is used to characterize  $O^6$ - and  $O^3$ -acetylmorphine. Similarly, the 28-amu mass shift observed in the mass spectrum of derivatized secobarbital indicates the replacement of two hydrogen atoms by two methyl groups (**Figures 3A** and **3B**).

Proper selection of derivatizing agents can also help determine the molecular weight of a compound. Compared to parent compounds, TMS derivatives of *N*-substituted barbiturates are found to generate less olefin radical elimination ( $[M-41]^+$  and  $[M-55]^+$ ). Instead, the formation of the  $[M-15]^+$  ion is favored, making it easier to recognize the molecular weight of the compound under examination.

MS is generally ineffective in differentiating enantiomers by itself; however, MS can be made effective by the proper derivatization of enantiomers followed by chromatographic separation. Although chiral stationary phases are available for direct separation of enantiomers, derivatization with chiral or achiral agents can often improve chromatographic characteristics. As an example, the total ion chromatograms of the L-TPC-derivatized amphetamine and methamphetamine are shown in **Figure 4**. The



**Figure 3** Mass spectra and chemical structures of pentobarbital (A) and methyl derivative of pentobarbital (B). (Reproduced with permission from Liu RH and Gadzala DE (1997) *Handbook of Drug Analysis – Applications in Forensic and Clinical Laboratories*, p. 225. Washington, DC: American Chemical Society.)



**Figure 4** Total ion chromatograms of L-TPC derivatized amphetamine (A), methamphetamine (B), and amphetamine and methamphetamine mixture (C). (A) and (B) are obtained from a 25 m (0.30 mm ID) glass capillary Chirasil-Val column; (C) is obtained from a 25 m (0.20 mm ID) fused silica glass capillary SP-2100 column. La, L-amphetamine; Da, D-amphetamine; Lm, L-methamphetamine; Dm, D-methamphetamine; l, L-TPC; and d, D-TPC. (Reproduced with permission from Liu RH (1981) *Approaches to Drug Sample Differentiation*, p. 32. Taiwan: Central Police College Press.)

four possible isomers resulting from the reaction of D- and L-amphetamine with D- and L-TPC are completely resolved by the Chirasil-Val column (Figure 4A). The methamphetamine counterparts are resolved into three peaks (Figure 4B). Since the achiral column (SP-2100) cannot resolve enantiomers, only two peaks are observed (Figure 4C).

### ILAs as ISs

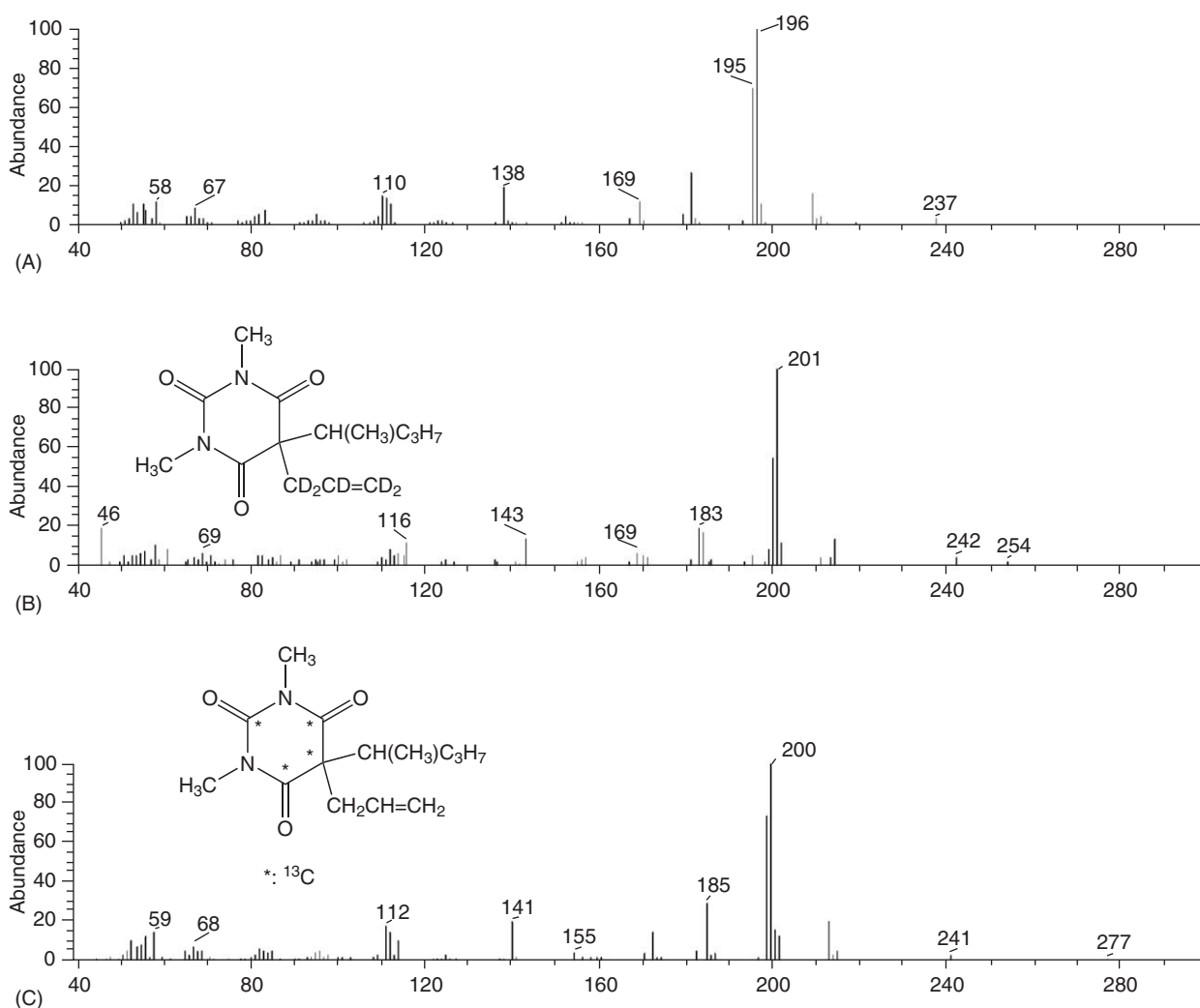
The use of MS-based methodologies for quantitative analysis is now a routine practice in forensic science laboratories. In most applications, ILAs of the analytes are used as the ISs, and the MS is operated in the selected ion monitoring (SIM) mode. Although  $^2\text{H}$ -analogs are the most commonly used ILAs, recent comparative studies have suggested that  $^{13}\text{C}$ -analogs (Figures 5A–5C) could be more effective. With virtually identical chemical properties and MS fragmentation characteristics, an ILA is a preferred IS because it offers the following advantages.

1. Errors derived from incomplete recovery of the analyte in the sample preparation process and/or varying GC and MS conditions are compensated for.
2. The presence of interfering materials (or mechanisms) which affect the detection or quantitation of the analyte will result in an altered response and ion intensity ratios or the absence of the IS in the final chromatogram; thus, alerting the analyst to conduct further investigation.

Before an ILA of the analyte is adopted as the IS, several important parameters should be carefully considered.

1. The ILA should be labeled with a sufficient number of atoms of a selected isotope so that the corresponding ions selected from the IS and analyte possess a significant mass difference. This is especially critical when sulfur atoms are present or a trimethylsilyl derivative is used.
2. The ILA should be manufactured with extremely high isotopic purity. Otherwise, the addition of the IS may result in the appearance of a significant amount of the analyte in a true negative sample and may also introduce errors in quantitation.
3. The labeling isotopes must be positioned at appropriate locations in the molecular framework of the compounds so that, after the fragmentation process, a sufficient number of high-mass ions, retaining the labeling isotopes, are present with significant intensities. Furthermore, these ions should not contribute to the intensities of the corresponding ions derived from the nonlabeled analyte. These ions and their counterparts in the analyte may then be monitored for ion ratio evaluation to facilitate qualitative compound identifications and quantitative determinations.

In order to establish a linear calibration curve, the intensities of the ions designated for the analyte and the IS must not have cross-contribution. For example, in a secobarbital/ $^{13}\text{C}_4$ -secobarbital study, calibration curves were evaluated using two pairs of ions:  $m/z$  196/200 and 181/185 (Figures 5A and 5C). For  $m/z$  196 (designed for the analyte), 0.23% of the measured intensity is contributed by the IS, while 0.017% of the measured intensity of  $m/z$  200 (designed for the IS) is contributed by the analyte. On the other hand, 1.6% of the measured intensity of  $m/z$  181 (designed for analyte) is contributed by the IS, while 0.29% of the measured intensity of  $m/z$  185 (designed for the IS) is contributed by the analyte. Figure 6 clearly demonstrates the difference in linearity of the calibration curves established by these two ion pairs. When cross-contribution occurs, non-linear models can best describe the observed data and should be seriously considered.



**Figure 5** Mass spectra and chemical structures of secobarbital (A),  $^2\text{H}_5$ -secobarbital (B), and  $^{13}\text{C}_4$ -secobarbital (C) (All as methyl-derivatives). (Reproduced with permission from Chang W-T, Smith J, and Liu RH (2002) Isotopic analogues as internal standards for quantitative GC/MS analysis – molecular abundance and retention time difference as interference factors. *Journal of Forensic Sciences* 47: 873–881.)

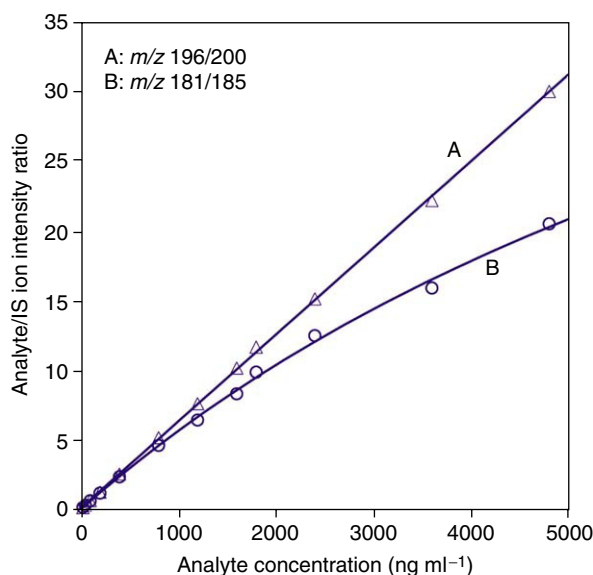
## Analysis of Nondrug Samples of Forensic Science Interest

MS in conjunction with GC is commonly employed during fire investigations. The GC–MS is capable of determining whether an observed hydrocarbon chromatogram pattern is derived from an accelerant or from pyrolysis products of various polymers. If the chromatogram is indicative of an accelerant, GC–MS is able to characterize what category of accelerant was present. Additionally, MS has been used to identify explosives and their postblast residues.

Distinct ion chromatograms are observed from debris derived from fire accelerant and those resulting from the pyrolysis of various polymers. Selected ions representing the following categories of

compounds are used to display the distributions of these compounds in a chromatogram: aliphatics ( $m/z$ : 57, 71, 85, 99), alicyclics ( $m/z$ : 55, 69, 83, 97), alkylbenzenes ( $m/z$ : 91, 105, 119, 183), alkylstyrenes ( $m/z$ : 104, 116, 132, 146), naphthalenes ( $m/z$ : 128, 142, 156, 170), terpenes ( $m/z$ : 93, 121, 136), biphenyls ( $m/z$ : 154, 168, 182, 196), and polynuclear aromatics ( $m/z$ : 178, 192, 206).

Common fire accelerants show distinct ion chromatogram profiles of alkylbenzenes and alkyl-naphthalenes. High percentages of these compounds are present in evaporated gasoline. Aviation fuel contains low-boiling alkylbenzenes similar to gasoline, but lacks significant quantities of high-boiling alkylbenzenes and alkyl-naphthalenes. Fuel oil and diesel fuel have a boiling range similar to that of evaporated gasoline, but are deficient in aromatics. Charcoal



**Figure 6** Linear and hyperbolic calibration results using ion-pairs with different degrees of cross-contribution – Secobarbital (SB)/ $^{13}\text{C}_4$ -secobarbital ( $^{13}\text{C}_4$ -SB):  $m/z$  196/200, 181/185. A total of 0.23% of the measured intensity of  $m/z$  196 (designed for SB) is contributed by  $^{13}\text{C}_4$ -SB; while 0.017% of the measured intensity of  $m/z$  200 (designed for  $^{13}\text{C}_4$ -SB) is contributed by SB. A total of 1.6% of the measured intensity of  $m/z$  181 (designed for SB) is contributed by  $^{13}\text{C}_4$ -SB, while 0.29% of the measured intensity of  $m/z$  185 (designed for  $^{13}\text{C}_4$ -SB) is contributed by SB. (Reproduced with permission from Liu RH, Lin T-L, Chang W-T, *et al.* (2002) Isotopically labeled analogues for drug quantitation. *Analytical Chemistry* 74: 618A–626A.)

lighter fluid and low-boiling naphthas are devoid of aromatic hydrocarbons. Paint thinners have significant amounts of middle-boiling alkylbenzenes, but few alkyl-naphthalenes.

In the forensic science laboratory, thermogravimetry and pyrolysis techniques are often combined with MS for the analysis of fiber and paint. Limited MS studies conducted on nylons, polyesters, acrylics, polyolefins, chlorine-containing polymers, cellulose, and various natural fibers concluded that it may be possible to differentiate some of these materials. However, MS techniques do not generally provide the overall discrimination power that has been demonstrated by IR-based methodologies.

A substantial number of studies on the characterization of vehicle paints, mainly alkyd and acrylic resins, have been reported. Alkyd resins, which are complex polyesters formed by co-polymerizing phthalic anhydride, polyols, and saturated and unsaturated fatty acids, can usually be classified with MS techniques on the basis of hydrocarbon chain length and modification components such as melamine. Characterizations of acrylic resins are based on CI MS analysis of monomers that are reasonably reproducible by pyrolysis techniques.

Mass spectrometric characteristics of common explosive substances, such as nitrate esters, nitramines, and nitroaromatic compounds, have been studied with the intention of developing effective approaches and protocols for preventive detection of hidden explosives and the analysis of postexplosion residues in debris materials. It has been strongly advocated that the analysis of explosives should follow the same criteria applied to drugs, i.e., MS identification regardless of how many other chromatographic techniques have been applied. Using GC–MS in the EI and negative CI mode, ethylene glycol dinitrate (EGDN), nitroglycerine (GTN), 2,4,6-trinitrotoluene (TNT), and hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) can be identified at the picogram to low nanogram level. The identification of GTN, TNT, PETN, RDX, and tetryl (2,4,6-trinitrophenolmethyl-nitramine) in postblast residue by GC–MS have also been reported. The development of portable MS configurations for detecting vapors of hidden explosives is an area that attracts much interest and may provide valuable service in the future.

## Tandem MS in Forensic Science

Earlier applications of tandem MS methods included the determination of the exact structure of ‘China White’ ( $\alpha$ -methylfentanyl) and the screening and confirmation of multiple drugs in biological fluids. The exact structure of  $\alpha$ -methylfentanyl was pieced together through successive pairing of EI ‘complementary ions’, whose sum of masses equals that of the molecular ion, and the determination of the structures of these paired ions by their collisionally induced dissociation spectra, i.e., MS/MS spectra.

Tandem MS procedures have been employed for the rapid screening and confirmation of up to 50 drugs in a single sample, at drug levels of nanograms per milliliter to micrograms per milliliter. In this screening method, the molecular ion is selected by the first MS stage, followed by collisionally induced dissociation, which produces daughter fragments that are then analyzed by the second stage of the MS. Such a procedure involves the use of (1) selected reaction monitoring for preliminary screening of possible drugs, (2) neutral loss screening to indicate the presence of any member of a given class of drugs, and (3) the acquisition of complete daughter spectra for confirmation. Using the first stage MS to isolate a specific ion, not only saves time, but also significantly improves the limits of detection by reducing chemical noise in the MS/MS spectrum.

Applications of GC–MS/MS and LC–MS/MS by forensic scientists are now frequently reported. For



example, a recent report on the analysis of accelerants in fire debris demonstrates that information derived from a GC coupled with ion trap MS/MS can greatly reduce interference caused by pyrolysis of construction and furnishing materials. Another example is the confirmation and quantitation of sildenafil and its metabolite using LC-MS/MS in post-mortem tissues and fluids at levels down to 1 ng per g and 1 ng per ml, respectively. By removing the majority of interfering clutter in the first stage of MS, the MS/MS spectrum generated is extremely clean. Thus, both detectability and specificity are increased with MS/MS.

## Stable Isotope Ratio MS

Under certain circumstances standard MS techniques are unable to elucidate the necessary results. For example, sample comparisons based simply on compound identification and quantitation can be misleading if sample doping has occurred. Additionally, determining the regional origin of plant-derived drugs has been a difficult, if not impossible task. An approach that has proven successful in achieving these investigative goals is the comparison of naturally occurring isotope compositions in the samples of concern. The most convenient isotope to measure is the  $^{13}\text{C}$ -enrichment level.

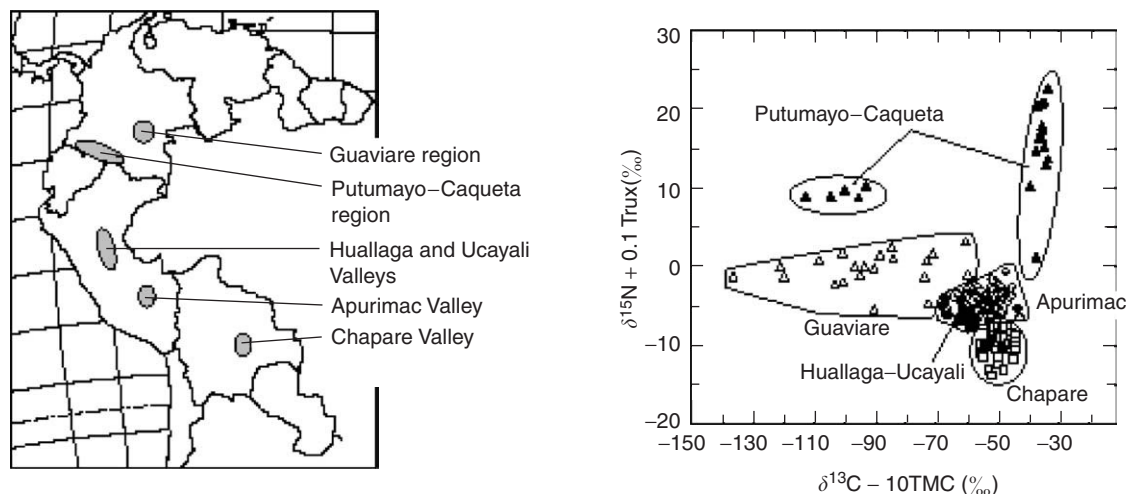
Differences in  $^{13}\text{C}$ -enrichment level are universally expressed using the delta notation shown below:

$$\delta^{13}\text{C}(\text{‰}) = \left[ \frac{R_{\text{sample}}}{R_{\text{standard}}} - 1 \right] \times 1000$$

where  $R$  is the ratio of the number of atoms of the minor isotope to that of the major isotope. In order

to compare isotope enrichment levels, a standard reference material is needed. The  $^{13}\text{C}$ -enrichment level found in Pee Dee belemnite (PDB), i.e., a calcium carbonate deposit of the fossil *Belemnitella americana* from the Peedee formation in South Carolina, is used as the reference.

The determination of  $^{13}\text{C}$ -content has recently become a major area of interest in the antidoping testing community. This interest is derived from the desire to develop a subjective parameter for detecting misuse of androgens, especially testosterone. Since testosterone concentration varies among individuals, the testosterone/epitestosterone, i.e., the  $17\alpha$  epimer and metabolite of testosterone, ratio was used as an indirect marker of testosterone administration. However, this approach has been circumvented through doping oneself with epitestosterone. By determining the androgen  $^{13}\text{C}$ -content during the testing process, differentiation between endogenous and synthetic testosterone and epitestosterone can be achieved. This procedure is considered the most effective approach in identifying androgen doping. Its effectiveness is based on the fact that synthetic androgens are generally made from a single plant species, which have lower  $^{13}\text{C}$ -contents than those derived from endogenous sources. For example, it has been reported that the  $\delta^{13}\text{C}$  values for all four synthetic epitestosterones examined were  $\leq -30.3\text{‰}$ , while the mean  $\delta^{13}\text{C}$  value of epitestosterone obtained from 43 healthy male urine samples was  $-23.8\text{‰}$  (SD,  $0.93\text{‰}$ ). Nine of 10 athletes' urine samples, with epitestosterone concentrations  $>180\text{ }\mu\text{g l}^{-1}$ , had  $\delta^{13}\text{C}$  values within  $\pm 3$  SD of the control group. The other athlete's  $\delta^{13}\text{C}$  value was  $-32.6\text{‰}$  and was suggested to have administered synthetic epitestosterone.

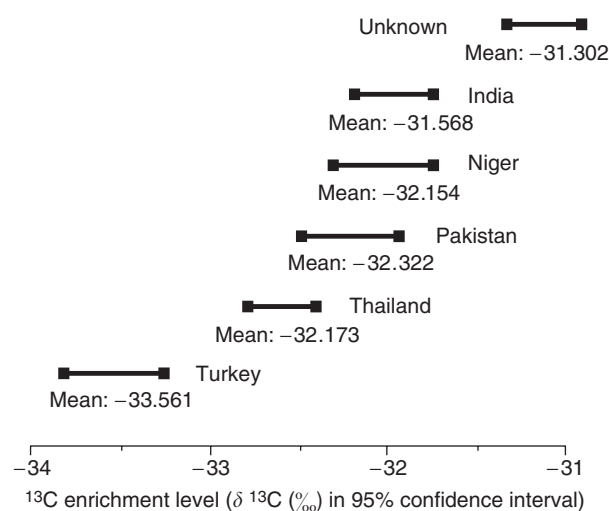


**Figure 7** Regional grouping of cocaine samples based on  $^{15}\text{N}/^{14}\text{N}$  ratio and truxilline (Trux) content and  $^{13}\text{C}/^{12}\text{C}$  ratio and trimethoxycocaine (TMC) content. (Reproduced with permission from Ehleringer JR, Casale JF, Lott MJ, and Ford VL (2000) Tracing the geographical origin of cocaine. *Nature* 408: 311–312.)

**Table 1** Variations of  $^{13}\text{C}$ -enrichment level in methamphetamine synthesized with different percentage of  $^{13}\text{C}$ -methylamine

Sample no.	Methamphetamine yield (%)	$^{13}\text{C}$ -methylamine used (%)	$^{13}\text{C}$ to $^{12}\text{C}$ ratio in $[\text{C}_3\text{H}_8\text{N}]^+$		
			Calculated	Measured	
				Average	SD
7	31	0	0.039 61	0.038 73	0.000 68
9	22	0		0.039 44	0.000 45
14	7.5	0.25	0.041 99	0.041 40	0.000 74
16	6.4	0.25		0.041 56	0.000 71
17	58	0.25		0.041 66	0.000 71
13	26	0.50	0.044 39	0.044 14	0.000 66
12	18	10	0.049 32	0.048 44	0.000 83
15	40	10		0.048 71	0.001 03
10	31	2.0	0.059 32	0.058 03	0.000 96
11	23	40	0.079 56	0.076 23	0.001 43

From Low IA, Liu RH, Piotrowski EG, and Furner RL (1986) Gas chromatographic/mass spectrometric determination of carbon isotope composition in unpurified samples: methamphetamine example. *Biomedical and Environmental Mass Spectrometry* 13: 531–534.



**Figure 8** Constructed based on data reported in Desage M, Guillury R, Brazier JL, *et al.* (1991) Gas chromatography with mass spectrometry or isotope-ratio mass spectrometry in studying the geographical origin of heroin. *Analytica Chimica Acta* 247: 249–254.

This isotope approach has also been very successful in differentiating between beet sugar and cane sugar, due to differences of carbon fixation pathways of these two plants. Additionally, it has been used to detect the addition of sugar to honey and natural fruit products. Typically, the  $\delta^{13}\text{C}$  value for cane sugar is  $-11.5\text{‰}$ , while most fruit sugars are lower than  $-20\text{‰}$ .

$^{13}\text{C}$ -enrichment levels have been successfully applied to samples in the field of forensic science. Coca leaves from South America were found to vary in their  $\delta^{13}\text{C}$  ( $-32.4\text{‰}$  to  $-25.3\text{‰}$ ) and  $\delta^{15}\text{N}$  ( $0.1$ – $13.0\text{‰}$ ) values. Humidity levels and the length of the rainy season, and differences in soils were thought to affect the fixation processes and cause the observed

subtle variations in  $^{13}\text{C}$  and  $^{15}\text{N}$  contents, respectively. In conjunction with the variations of trace alkaloids (truxilline and trimethoxycocaine) found in cocaine, researchers were able to correctly identify 96% of 200 cocaine samples originating from the regions studied (Figure 7). Additionally, the  $^{13}\text{C}$ -enrichment levels in heroin samples derived from various geographic regions have been investigated. Reported data are converted into a graphical presentation by this author (Figure 8).

Sample differentiation approaches based on  $^{13}\text{C}$ -enrichment levels have also been advocated as a basis for a potential sample tagging mechanism. An exploratory study utilized different percentages of  $^{13}\text{C}$ -enriched N-methylamine during the methamphetamine synthesis process. Data shown in Table 1, obtained using a general purpose GC–MS system, indicate that a variation step of 0.25% in  $^{13}\text{C}$ -N-methylamine was sufficient for product differentiation. It was estimated that the variation step could be reduced by a factor of 50-fold if an isotope ratio mass spectrometer was used.

**See also: Forensic Sciences:** Arson Residues; Drug Screening in Sport; Explosives; Illicit Drugs; Paints, Varnishes, and Lacquers; Systematic Drug Identification; Volatile Substances.

## Further Reading

- Aguilera R, Hatton CK, and Catlin DH (2002) Detection of epitestosterone doping by isotope ratio mass spectrometry. *Clinical Chemistry* 48: 629–636.
- Cheng MT, Kruppa GH, McLafferty FW, and Cooper DA (1982) Structural information from tandem mass spectrometry for China White and related fentanyl derivatives. *Analytical Chemistry* 54: 2204–2207.

- Darby SM, Miller ML, and Allen RO (2001) Forensic determination of ricin and the alkaloid marker ricinine from castor bean extracts. *Journal of Forensic Sciences* 46: 1033–1042.
- de Kock SS, Rodgers JP, and Swanepoel BC (2001) Growth hormone abuse in the horse: preliminary assessment of a mass spectrometric procedure for IGF-1 identification and quantitation. *Rapid Communication in Mass Spectrometry* 15: 1191–1197.
- de Vos B-J, Froneman M, Rohwer E, and Sutherland DA (2002) Detection of petrol (gasoline) in fire debris by gas chromatography/mass spectrometry/mass spectrometry (GC/MS/MS). *Journal of Forensic Sciences* 47: 736–756.
- Ehleringer JR, Casale JF, Lott MJ, and Ford VL (2000) Tracing the geographical origin of cocaine. *Nature* 408: 311–312.
- Ho MH (ed.) (1990) *Analytical Methods in Forensic Chemistry*, chs. 1–6. New York: Ellis Horwood.
- Lewis RJ, Johnson RD, and Blank CL (2000) A novel method for the determination of sildenafil (viagra) and its metabolite (UK-103,320) in postmortem specimens using LC/MS/MS and LC/MS/MS/MS. Department of Transportation, Federal Aviation Administration, DOT/FAA/AM-00/20.
- Liu RH and Gadzala DE (1997) *Handbook of Drug Analysis – Applications in Forensic and Clinical Laboratories*, chs. 9 and 11. Washington, DC: American Chemical Society Press.
- Liu RH, Lin T-L, Chang W-T, *et al.* (2002) Isotopically labeled analogues for drug quantitation. *Analytical Chemistry* 74: 618A–626A.
- Moore JM (1978) The application of derivatization techniques in forensic drug analysis. In: Klein M, Kruegel AV, and Sobol SP (eds.) *Instrumental Applications in Forensic Drug Chemistry*, pp. 180–201. Washington, DC: US Government Printing Office.
- Nowicki J (1990) Accelerant classification scheme based on analysis by gas chromatography/mass spectrometry (GC–MS). *Journal of Forensic Sciences* 35: 1064–1086.
- Pfleger K, Maurer HH, and Weber A (eds.) (1992) *Mass Spectral and GC Data of Drugs, Poisons, Pesticides, Pollutants and Their Metabolites*, 2nd edn. Weinheim: VCH.
- Yinon J (2001) Forensic analysis of explosives by LC/MS. *Forensic Science Review* 13: 19–28.
- Yinon J (ed.) (1995) *Forensic Applications of Mass Spectrometry*. Boca Raton, FL: CRC Press.

## Gas Analysis

**R Wright**, Thermo Electron Corporation, Winsford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Mass spectrometry has become established as a very powerful method of gas analysis, being fast, accurate, multicomponent, and flexible. During the past two decades, advances in technology have enabled mass spectrometers to be designed specifically for continuous or fast online compositional measurement of gases to facilitate industrial plant process control or for environmental monitoring. Modern instruments are typically compact, robust, very reliable, and simple to use with flexible user-friendly software. They are often equipped with multistream sampling devices, to measure samples from many sample points with high frequency, to benefit from the speed of analysis. Gas analysis mass spectrometers can be configured with suitable protection for operation in hazardous areas in industrial plants.

### Instrumentation

The sample of a gas analysis mass spectrometer usually passes through an enclosed electron ionization

(EI) ion source, in which the ionization volume is operated at a pressure of typically  $10^{-4}$  mbar, which is higher than that of the mass analyzer and the thermionic filament. This gives improved detection limits, since the effect of background pressure in the analyzer is reduced. Interaction between the hot filament and the sample is minimized; this might otherwise give rise to artifact peaks or, in the case of reactive gases, changes in the sensitivity of the instrument.

The process of ionization of an EI source typically produces fragment ions and multiply-charged ions as well as molecular ions. Additionally the existence of isotopes produces ions having different masses. The relative numbers of ions formed at different  $m/z$  from a given gas component, known as the ‘fragmentation pattern’ or ‘cracking pattern’, is used for identification and quantification. The fragmentation pattern can be influenced by instrumental factors such as the energy of the bombarding electrons, temperature, the geometry of the ion source, mass spectrometer lens voltages, pressure, mass analyzer type and geometry, and detector type. To facilitate comparison, reference data on fragmentation patterns usually originate from measurements using 70 eV electrons and at a source pressure of  $10^{-4}$  mbar or lower. Under these conditions, an analytical mass spectrometer

- Darby SM, Miller ML, and Allen RO (2001) Forensic determination of ricin and the alkaloid marker ricinine from castor bean extracts. *Journal of Forensic Sciences* 46: 1033–1042.
- de Kock SS, Rodgers JP, and Swanepoel BC (2001) Growth hormone abuse in the horse: preliminary assessment of a mass spectrometric procedure for IGF-1 identification and quantitation. *Rapid Communication in Mass Spectrometry* 15: 1191–1197.
- de Vos B-J, Froneman M, Rohwer E, and Sutherland DA (2002) Detection of petrol (gasoline) in fire debris by gas chromatography/mass spectrometry/mass spectrometry (GC/MS/MS). *Journal of Forensic Sciences* 47: 736–756.
- Ehleringer JR, Casale JF, Lott MJ, and Ford VL (2000) Tracing the geographical origin of cocaine. *Nature* 408: 311–312.
- Ho MH (ed.) (1990) *Analytical Methods in Forensic Chemistry*, chs. 1–6. New York: Ellis Horwood.
- Lewis RJ, Johnson RD, and Blank CL (2000) A novel method for the determination of sildenafil (viagra) and its metabolite (UK-103,320) in postmortem specimens using LC/MS/MS and LC/MS/MS/MS. Department of Transportation, Federal Aviation Administration, DOT/FAA/AM-00/20.
- Liu RH and Gadzala DE (1997) *Handbook of Drug Analysis – Applications in Forensic and Clinical Laboratories*, chs. 9 and 11. Washington, DC: American Chemical Society Press.
- Liu RH, Lin T-L, Chang W-T, *et al.* (2002) Isotopically labeled analogues for drug quantitation. *Analytical Chemistry* 74: 618A–626A.
- Moore JM (1978) The application of derivatization techniques in forensic drug analysis. In: Klein M, Kruegel AV, and Sobol SP (eds.) *Instrumental Applications in Forensic Drug Chemistry*, pp. 180–201. Washington, DC: US Government Printing Office.
- Nowicki J (1990) Accelerant classification scheme based on analysis by gas chromatography/mass spectrometry (GC–MS). *Journal of Forensic Sciences* 35: 1064–1086.
- Pfleger K, Maurer HH, and Weber A (eds.) (1992) *Mass Spectral and GC Data of Drugs, Poisons, Pesticides, Pollutants and Their Metabolites*, 2nd edn. Weinheim: VCH.
- Yinon J (2001) Forensic analysis of explosives by LC/MS. *Forensic Science Review* 13: 19–28.
- Yinon J (ed.) (1995) *Forensic Applications of Mass Spectrometry*. Boca Raton, FL: CRC Press.

## Gas Analysis

**R Wright**, Thermo Electron Corporation, Winsford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Mass spectrometry has become established as a very powerful method of gas analysis, being fast, accurate, multicomponent, and flexible. During the past two decades, advances in technology have enabled mass spectrometers to be designed specifically for continuous or fast online compositional measurement of gases to facilitate industrial plant process control or for environmental monitoring. Modern instruments are typically compact, robust, very reliable, and simple to use with flexible user-friendly software. They are often equipped with multistream sampling devices, to measure samples from many sample points with high frequency, to benefit from the speed of analysis. Gas analysis mass spectrometers can be configured with suitable protection for operation in hazardous areas in industrial plants.

### Instrumentation

The sample of a gas analysis mass spectrometer usually passes through an enclosed electron ionization

(EI) ion source, in which the ionization volume is operated at a pressure of typically  $10^{-4}$  mbar, which is higher than that of the mass analyzer and the thermionic filament. This gives improved detection limits, since the effect of background pressure in the analyzer is reduced. Interaction between the hot filament and the sample is minimized; this might otherwise give rise to artifact peaks or, in the case of reactive gases, changes in the sensitivity of the instrument.

The process of ionization of an EI source typically produces fragment ions and multiply-charged ions as well as molecular ions. Additionally the existence of isotopes produces ions having different masses. The relative numbers of ions formed at different  $m/z$  from a given gas component, known as the ‘fragmentation pattern’ or ‘cracking pattern’, is used for identification and quantification. The fragmentation pattern can be influenced by instrumental factors such as the energy of the bombarding electrons, temperature, the geometry of the ion source, mass spectrometer lens voltages, pressure, mass analyzer type and geometry, and detector type. To facilitate comparison, reference data on fragmentation patterns usually originate from measurements using 70 eV electrons and at a source pressure of  $10^{-4}$  mbar or lower. Under these conditions, an analytical mass spectrometer

will typically display fragmentation patterns consistent with library data to within perhaps  $\pm 30\%$  relative.

Other ionization techniques are sometimes used, such as atmospheric pressure ionization (API), photoionization, and chemical ionization. Of these, API has been the most widely applied, and involves the production of ions at high pressure in a corona discharge. The ions are then swept by the bulk gas flow into a region where they are extracted by lens voltages into the vacuum part of the mass spectrometer for mass separation and detection.

There are many types of mass analyzers for gas analysis, but the magnetic sector and the quadrupole are the most widely used, with time-of-flight, Fourier transform ion cyclotron resonance, and ion trap types being advantageous for special applications, for example, those requiring very fast measurement, high resolution, and high sensitivity, respectively. Since mass spectrometers generally operate at a pressure of  $10^{-5}$  mbar or lower, they require a high vacuum system. In contrast, the sample will usually originate at close to atmospheric pressure (1000 mbar). It is therefore necessary to admit the sample into a high-vacuum environment, in a way which reduces the sample pressure by many decades. During this transfer the composition must remain sufficiently consistent for the analysis in hand.

After passing through the mass analyzer, the ions impinge directly onto a Faraday electrode or, for enhanced sensitivity, a secondary electron multiplier. The typical range of ion currents measured are  $10^{-18}$ – $10^{-9}$  A. Almost invariably, a computer will record, process, and display the data. In principle, quite complex calculations may be executed by the computer to derive the compositional results required, but the physical phenomena involved are essentially linear, so well-established algorithms are generally employed.

## Gas Analysis by Mass Spectrometry

In order to perform an analysis, one must establish a correlation between the molecular species present and the peaks observed in the mass spectrum. The gas may consist of known components, whose relative concentrations are to be determined, or alternatively it may be required to identify unknown species.

If the species are known, the first step is to identify suitable mass peaks to monitor. 'Standard' tabulations of mass spectra are readily available. For example, consider the analysis of breathable air. The main species of interest are nitrogen, oxygen, argon,

and carbon dioxide. **Table 1** shows standard spectra for these four gases. It is conventional to normalize each spectrum to the base peak (that is, the most intense). (It should be mentioned again that these values are not fixed and vary from instrument to instrument and depend on actual operating parameters and also may change with time.)

One then identifies peaks that will permit analysis of each individual species, if possible avoiding or minimizing overlaps between the spectra. In the present example, the base peaks for each of the four species are suitable since the amount of carbon dioxide is sufficiently low so that the overlap or 'interference' with nitrogen at mass 28 is small. This overlap is easily corrected for by simple subtraction if the ratio of peaks from carbon dioxide is determined by calibration with a suitable gas containing carbon dioxide. For some gas analysis applications, there is considerable overlap and development of an analytical method then becomes more complex, involving a search for peaks which have less overlap but still have worthwhile intensity. The analysis of several overlapping or 'interfering' gas mixtures inevitably involves measurement by deconvolution of the overlapping peaks. An important assumption is made (and is generally obeyed) that the overlapping peaks when combined obey the principle of linear peak superposition. The principle of linear peak superposition is that the composite peak height at a particular mass is simply equal to the sum of the peak heights which correspond linearly to the individual concentrations of the contributing components in the complex mixture. This may be represented as follows:

$$i_1 = s_1 f_{11} c_1 + s_2 f_{21} c_2 + s_3 f_{31} c_3 + \dots + t s_n f_{n1} c_n$$

**Table 1** Mass spectra of gas mixture components (relative intensities)

	<i>N</i> <sub>2</sub>	<i>O</i> <sub>2</sub>	<i>Ar</i>	<i>CO</i> <sub>2</sub>	Structure
12				99	<sup>12</sup> C <sup>+</sup>
14	137				<sup>14</sup> N <sup>+</sup>
16		218		115	<sup>16</sup> O <sup>+</sup>
20			134		<sup>40</sup> Ar <sup>2+</sup>
22				17	<sup>12</sup> C <sup>16</sup> O <sub>2</sub> <sup>+</sup>
28	999			117	<sup>14</sup> N <sub>2</sub> <sup>+</sup> , <sup>12</sup> C <sup>16</sup> O <sup>+</sup>
29	7				<sup>14</sup> N <sub>2</sub> <sup>+</sup>
32		999			<sup>16</sup> O <sub>2</sub> <sup>+</sup>
36			3		<sup>36</sup> Ar <sup>+</sup>
40			999		<sup>40</sup> Ar <sup>+</sup>
44				999	<sup>12</sup> C <sup>16</sup> O <sub>2</sub> <sup>+</sup>
45				11	<sup>13</sup> C <sup>16</sup> O <sub>2</sub> <sup>+</sup>

Data taken from NIST/EPA/MSDC Mass Spectral Database PC Version 3.0.

Note standard air composition is N<sub>2</sub>, 78.1%; O<sub>2</sub>, 21.0%, Ar, 0.93%; CO<sub>2</sub>, 0.03%.



where  $i$  is a composite peak height, and for 1 to  $n$  components contributing to this peak,  $s$  is the base peak sensitivity,  $f$  its cracking pattern, and  $c$  its concentration. Values of  $s$  and  $f$  are determined for each component during calibration. During analysis, the peak heights ( $i$ ) are measured for 1 to  $n$  (or  $>n$ ) masses so that there are at least  $n$  simultaneous equations to determine the  $n$  unknown values of  $c$ . Matrix inversion is generally used to solve these sets of simultaneous equations. These calculations are normally executed easily and rapidly by a computer within a fraction of a second, and typically the complete analysis going from sample introduction to giving a reading of gas concentrations takes only seconds.

If calibration cannot be performed, then estimates can be made for values of  $f$  and  $c$  above, but subsequent errors may be considerable. The greater the overlap, the more serious is the possibility of error. Generally, for reliable quantitative work, complete calibration should be made. Even with calibration, errors can be significant due to changes in sensitivity and fragmentation patterns. There can be adverse influences from poor environmental control, ionizer instability, mass scale instability, and detector instability. There may also be unexpected errors from a number of phenomena such as tails from much larger peaks (an effect often referred to as 'abundance sensitivity'), nonlinearity in response, detector noise and drift, vacuum system out-gassing, permeation and other background phenomena (such as poor high vacuum pump compression for light gases), inaccurate or unstable calibration gases, memory, adsorption, reaction and conditioning effects, and generally poor sampling. Sometimes there can be interferences from unexpected components due to reactions occurring in the ion source to generate unexpected species (e.g., cracking of hydrocarbons to generate molecular hydrogen) which need to be considered. Where two or more components share very similar spectra, their quantification tends to become very inaccurate or impossible (e.g., *cis*- and *trans*-alkenes). However, despite these various possibilities for measurement error, a carefully installed and maintained mass spectrometer, with the analytical method correctly configured, is able to produce very accurate and reproducible analysis for a wide range of applications. Accuracy is typically better than 1% relative over 1 month with standard deviations over 24 h typically better than 0.1% relative.

When analyzing for unknown components, similar considerations apply, but of course with a different emphasis. An experienced operator will be able to identify individual gases from the measured spectrum, especially if only a few components are

present. If the mixture is complex, the measured signals can be compared with a library of 'standard' spectra, and a 'most likely' composition inferred. A linear least-squares calculation is typically used, but the error can be considerable, due to actual spectra sometimes being significantly different from the 'standard' spectra. It is, however, normally necessary, where the task is to reliably identify unknowns in a multicomponent mixture, that the system includes a gas chromatograph to provide a preseparation, to enable introduction of the components into the mass spectrometer one at a time, so that individual component mass spectra can be identified more easily in a sequential manner.

## Inlets

There are many ways of introducing a sample into a mass spectrometer. Those most commonly used for general gas analysis and vacuum work are discussed below. In order to describe these, it is first necessary to review some basic gas properties.

## Gas Properties

The physical properties of the sample gas, such as pressure, density, temperature, and flow rate are described by classical kinetic theory. Thus the ideal gas equation  $PV = nRT$  applies throughout the system. However, gas flow patterns change as the pressure is reduced, giving rise to effects that must be considered in the design of inlets.

The behavior of gas at high pressure is dominated by intermolecular collisions. Each molecule travels only a short distance, characterized by the mean-free-path (MFP) between collisions. The MFP can be estimated from kinetic theory, and is, for example,  $\sim 6.5 \times 10^{-5}$  m for nitrogen at atmospheric pressure. If a gas is contained in a vessel whose shortest dimension is much longer than the MFP, then only molecules close to the wall of the vessel interact with it directly. These considerations lead to the concept of gas viscosity, and the gas is said to undergo 'viscous' or 'laminar' flow. This is a bulk property of the gas, and the flow rate depends upon the difference in total pressure and upon the bulk viscosity. In the vacuum system, the MFP is usually longer than the dimensions of the containing vessel. Intermolecular collisions seldom occur, and collisions with the wall of the vessel dominate. The flow rate for each species then depends upon the difference in partial pressure for the individual species, and upon the molecular weight. This condition is called 'molecular flow'.

As the sample passes to the ionizer, the nature of the gas flow will alter from predominantly viscous flow at high pressure to molecular flow within the

ionizer. A consequent change in composition, known as 'fractionation', can be avoided by employing special inlet techniques.

Typically the ionizer is pumped at a rate of  $\sim 1 \text{ l s}^{-1}$ . Therefore a gas introduction rate of  $\sim 0.1 \mu\text{l s}^{-1}$ , is typically used to create a pressure in the ionizer of  $10^{-4}$  mbar.

### Capillary with By-Pass

A particularly versatile inlet for continuous online analysis is the 'capillary with by-pass'. This is shown schematically in **Figure 1**. The sample, initially at atmospheric pressure, passes down a length of narrow diameter tubing. The pressure reduces under viscous flow. At a location where the pressure has reduced to a suitable value (typically 10 mbar), a small fraction of the flow is diverted through a leak (a plug of sintered material or small orifice) into the mass spectrometer. The remainder of the flow passes on to a vacuum pump that maintains the flow. The gas flow is essentially viscous from the atmospheric pressure end of the tubing to past the location of the leak. At the leak, the pressure is low enough for the entire flow through the leak to be molecular. By this means, fractionation is avoided, and the relative concentrations of the components are the same in the ionizer as in the origin of the sample at atmospheric pressure. The inlet is thus suitable for applications where the gas composition will alter significantly. It is relatively tolerant of changes in sample pressure, and responds rapidly to changes in composition, since the flow down the capillary is rapid. Memory effects are minimized, especially if trace heating is employed.

### Membrane

A method of enhancing the response preferentially to trace organic compounds is with a membrane inlet. Typically a poly(dimethylsiloxane) membrane is used

as the interface between the sample at atmospheric pressure and the ionization chamber at  $10^{-4}$  mbar. The membrane inlet serves to provide (when the sample gas is air) approximately the same gas introduction rate as with the other types of inlets, but with a typically 10–100 $\times$  preferential transmission of organic compounds to give significantly improved trace detection.

An interesting variation of the membrane inlet is to use it as a sampling device for sampling dissolved organic gases in water by flowing a carrier gas through a capillary of poly(dimethylsiloxane) immersed in the water sample.

### Direct Capillary Inlet

A particularly compact and simple inlet consists of a short length of narrow-bore tubing. For example a 0.5 m length of 0.03 mm capillary provides  $\sim 0.1 \mu\text{l s}^{-1}$  flow rate. The gas flow changes from viscous to molecular flow as it passes through the constriction, resulting in fractionation of the sample. The relative concentrations in the ion source are inversely proportional to the square root of the molecular weights. For example with a gas sample of hydrogen and nitrogen, the relative concentrations will be altered by a factor of  $(28/2)^{0.5}$ , i.e., 3.7. Therefore, the hydrogen/nitrogen ratio will be 3.7 $\times$  less in the ionizer compared with the original sample. This means the response to lighter components is less favorable. The direct capillary inlet is easily blocked by particulates or condensation and may take several seconds to equilibrate when the gas composition alters.

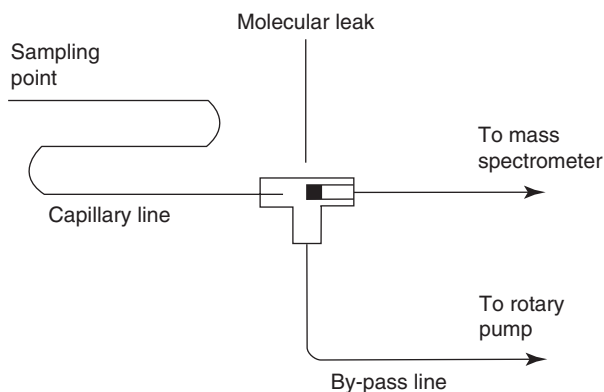
A direct capillary inlet can be configured rapidly, and will often be considered for experimental work or if small dimensions are important; but it must be used with caution. The response is slow and mass dependent.

### Leak Valve

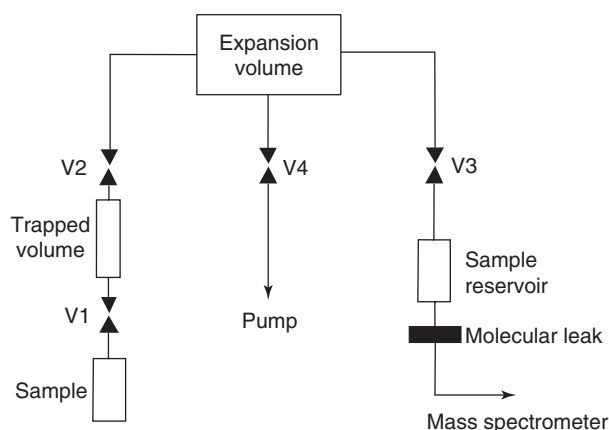
A leak valve gives capability over a wide range of pressures, and, if motorized, gives the possibility of automatic compensation for varying sample pressures. Fractionation effects are similar to those of the direct capillary inlet.

### Batch Inlet

A basic batch inlet is shown in **Figure 2**. A small quantity of gas is expanded into an evacuated volume, so that the pressure is low enough for molecular flow to occur in the leak to the mass spectrometer. No mixing of flow regimes occurs, so fractionation is absent. However, it involves more hardware than other inlets, and cannot give truly



**Figure 1** Schematic diagram of capillary inlet with by-pass.



**Figure 2** Schematic diagram of batch inlet.

continuous readings. Typically a batch inlet is employed for laboratory analysis of small samples, e.g., 1 mbar l of gas.

## Data Acquisition and Processing

A modern commercial mass spectrometer will incorporate a computer or will offer a facility to communicate with a computer via a data link, for example, to a plant control system using serial, ethernet or analog, and/or digital signals. This allows data to be stored and manipulated, often in 'real-time', so that results can be displayed to the user in a convenient form or to be utilized for process control either automatically via a closed loop or for operator 'fine-tuning'. A mass spectrum may be displayed as a continuous scan over a specified mass range similar to a strip-chart, but this is difficult to interpret, and is not directly quantitative.

It is usually preferable to monitor only a few selected masses, using computer control to switch between them. This simplifies the presentation of the data and allows much more rapid measurement. An alternative method is to measure every peak within a specified mass range. In either case, the rate of acquisition of data may be as high as several hundred separate measurements per second. Such high data rates are not always appropriate or convenient, so facilities will be provided to make measurements intermittently, or to accumulate averaged signals. Typically for precise quantitative gas analysis, each mass takes about 0.3–1 s to be measured. Most gas analysis mass spectrometers are designed to operate at low resolution (less than 100), so that peaks can be measured by peak jumping with infrequent mass calibration. Analysis stability and accuracy is dependent on a number of factors. By adopting a design of mass spectrometer which exhibits flat-topped peaks, such as with a magnetic sector instrument

(as seen in Figure 3) makes the system much more tolerant to errors in mass alignment due to drift, due to noise, temperature changes, or spurious surface potentials. Thereby, the peak height measurement reproducibility is enhanced and the analysis precision is improved.

## Calibration

In common with other analytical techniques, the best accuracy will be obtained by reference to calibrated gas mixtures whose composition is similar to that of the sample. This is straight forward if the sample composition remains fairly constant. If the gas composition varies more widely, one may need to calibrate by reference to a small number of selected gas mixtures. In a typical case, for a complete calibration, to provide accurate composition monitoring, it is necessary to use a minimum number  $N$  calibration gases where  $N$  is calculated as:

$$N = M + 2$$

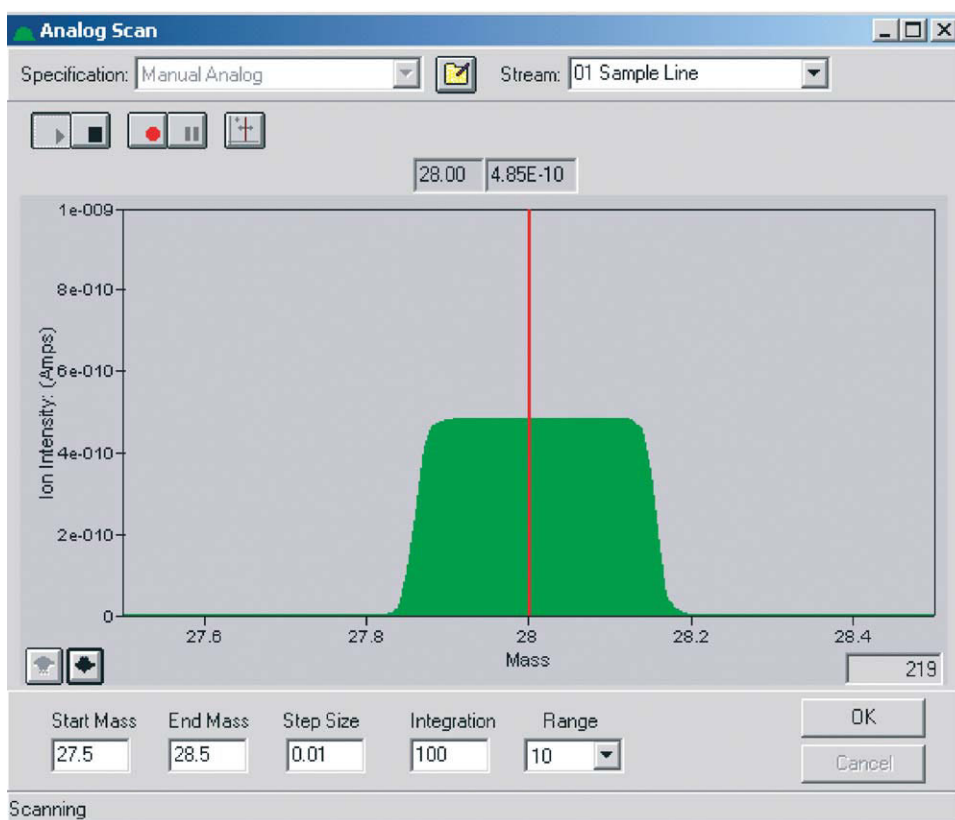
where  $M$  is the total number of components having fragmentation patterns that are used in the analysis.

The two additional calibration gases required are typically a multicomponent mixture to accurately determine the relative sensitivities and a 'background' gas for measuring the background signals at the various masses used in the analysis. As an example, consider the case of the reactor outlet gas from an ethylene oxide (EO) plant. Table 2 shows the components and their typical concentration levels, relative sensitivities, and fragmentation patterns. Note only a minimum set of masses has been considered for the analysis, i.e., the number of masses = the number of components = 8.

It is seen that there are five components having fragmentation patterns that are used, namely  $C_2H_4$ ,  $C_2H_6$ ,  $O_2$ , EO, and  $CO_2$ . Therefore  $5 + 2 = 7$  calibration gases are required, for example, as shown in Table 3.

Calibration gas 1 is for determining the relative sensitivities, calibration gases 2–6 are for calibrating the fragmentation patterns, and calibration gas 7 is for calibration for the background peak heights.

It is normal practice to periodically check the validity of the calibration by analyzing a 'check gas', which is essentially representative of the sample gas. This check gas is often the same gas that is used to calibrate for the relative sensitivities. Analysis of this gas will also indicate when recalibration is required. Normally certain tolerances are applied as criteria for determining when recalibration is required. For components in the 1–100% range, these tolerances



**Figure 3** Flat-topped peaks obtained with magnetic sector instrument are advantageous for longer term reproducibility when analysis is performed by peak jumping because precise mass calibration is unnecessary. Also peak overlaps are simpler to deconvolute because the peak tops coincide over a more extensive mass.

**Table 2** Typical components and their fragmentation patterns used for analysis of ethylene oxide production

	Concentration (mol%)	Relative sensitivity	16	26	28	30	32	40	43	44
CH <sub>4</sub>	45.5	0.7	100							
N <sub>2</sub>	2	1			100					
C <sub>2</sub> H <sub>4</sub>	25	1		52.4	100					
C <sub>2</sub> H <sub>6</sub>	1	1.1		18.9	100	28.6				
O <sub>2</sub>	5	0.8	5.4				100			
Ar	10	1.4						100		
EO	1.5	0.5		4	6.6	2.3			25.2	100
CO <sub>2</sub>	10	1.2	5.8		5.7				0.02	100

are typically 1–2% relative. One normally expects results to remain within the tolerances over periods of a week or longer. However, in practice this very much depends on environmental factors such as ambient temperature and pressure, and on the exposure of the mass spectrometer ion source to various types of gases: effects are often referred to as ‘conditioning’. For example, online mass spectrometers used in the iron and steel industry for measuring mainly reducing gases such as H<sub>2</sub>, CO, and CH<sub>4</sub> are typically observed to remain within 1% (relative) tolerance with the check gases over months; however, if

**Table 3** Calibration gases required for analysis of ethylene oxide production

Calibration gas	1	2	3	4	5	6	7
% CH <sub>4</sub>	45.5						
% N <sub>2</sub>	2	95					
% C <sub>2</sub> H <sub>4</sub>	25		25				
% C <sub>2</sub> H <sub>6</sub>	1			5			
% O <sub>2</sub>	5	5					
% Ar	10						
% EO	1.5				1.5		
% CO <sub>2</sub>	10					100	
% He			75	95	98.5		100

such a system is exposed to oxygen containing samples, e.g., air, then larger deviations may be observed and recalibration may be required more frequently.

It is good practice to maintain a log of check gas analysis data. Ideally the data should be plotted and calculations performed to determine mean accuracies and values of standard deviations.

The most convenient way to assess the accuracy of analysis is to use the following expression:

$$\begin{aligned} \text{accuracy} = & (\text{mean accuracy of check gas}) \\ & + (\text{gas mixture manufacturer's} \\ & \quad \text{concentration tolerance}) \\ & + 3 \times (\text{standard deviation}) \end{aligned}$$

## Applications of Gas Analysis Mass Spectrometry

### Fermentation

Fermentation is defined as a chemical change brought about using microorganisms, e.g., in the biotechnology industry for production of pharmaceuticals, food additives, and animal feed-stuffs. Respiration is the process whereby an organism oxidizes food to produce energy. Control of fermentation processes to give enhanced yields requires the measurement of the respiratory gases oxygen and carbon dioxide. Mass spectrometry has become established as the preferred technique for accurate monitoring of oxygen consumption and carbon dioxide evolution. In recent years, with considerable growth in the biotechnology industry, this has become the most common application of online mass spectrometry. The analytical requirement for this application is for the measurement of oxygen and carbon dioxide to be better than  $\pm 0.01$  mol%, with a linearity over a decade of at least 0.2% relative and a calibration interval of at least 30 days.

Other components, which are sometimes monitored, include methanol, ethanol, water, ammonia, and sulfur compounds.

### Iron and Steel

In a blast furnace, coke is used to reduce the ore (iron oxide) to iron. Large quantities of gas, 'top gas', are formed at the top of the blast furnace as a result of the reduction of iron ore and the combustion of injected coal by injected blast air. Gas analysis is used to optimize the blast furnace process, including reducing the consumption rate of coke and controlling the temperatures inside the furnace. Analysis of hydrogen is very important for early detection of cooling water leaks which can be a considerable hazard.

In a steel plant, the basic processes for conversion of iron (from blast furnace) and scrap into steel are removal of sulfur and reduction in the carbon content to the required level. In the basic oxygen steel making process (BOS), basic oxygen furnace process (BOF), or Linz-Donawitz process (LD), typically the carbon content is reduced from 4–0.05%. In this process high purity oxygen is blown through molten hot metal via a lance at supersonic speed, in a process that lasts only  $\sim 30$  min. The oxygen reacts with carbon to produce CO and CO<sub>2</sub>; analysis of these gases is used to control operation of the lance to improve process efficiency and also to detect the process endpoint. The benefits are very great – process turnaround is improved, reblows are avoided, as is undesirable oxidation of the metal.

Ruhrstahl-Hausen (RH) and vacuum oxygen decarburization (VOD) are secondary steel-making processes to produce specialty steels. The process occurs under vacuum, which enables reduction of the carbon content to very low levels ( $< 0.01\%$ ) without undesirable oxidation of chromium. Special inlet systems have been designed to sample from these variable pressure processes and still maintain a constant flow of sample gas into the ion source.

Frequently in an integrated iron and steel works, there is a requirement to measure, very rapidly on-line, the calorific value and stoichiometric air requirement of fuel gas mixtures, in order to optimize burner performance and this is typically performed using mass spectrometry.

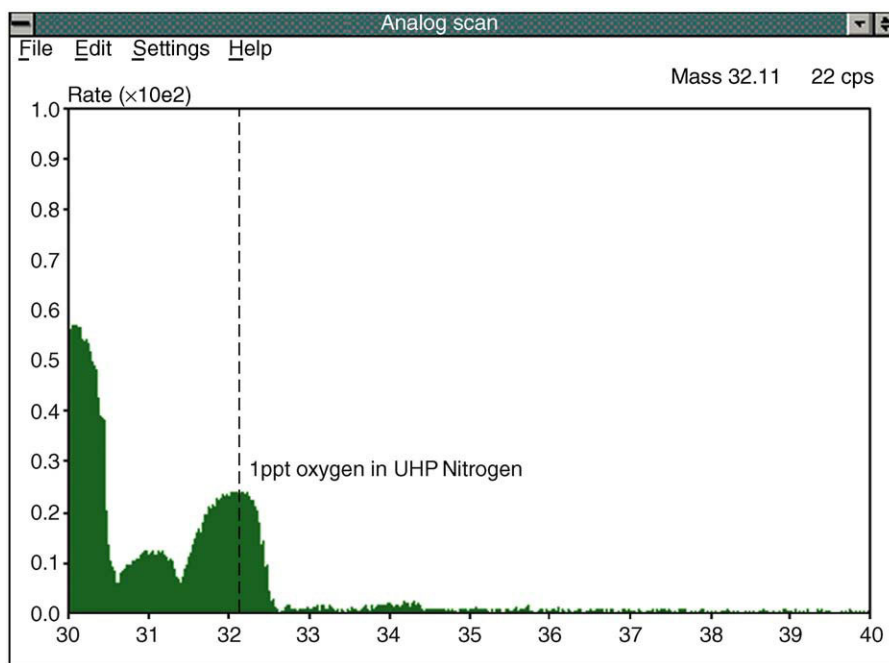
Other gases analyzed by online mass spectrometers in the iron and steel industry include those involved in coke ovens and direct reduction iron.

### Chemical

Examples of chemical processes using online gas analysis include those involving the production of the following materials: ethylene, ethylene oxide, polyolefins, ammonia, methanol, acrylonitrile, and vinyl chloride. Mass spectrometry is frequently used for control of steam reformers and coal gasification plants. Examples of species being monitored are nitrogen, oxygen, argon, carbon monoxide, carbon dioxide, and a variety of saturated and unsaturated organic compounds up to C<sub>6</sub>. Fast online mass spectrometric compositional gas analysis is being applied in many chemical applications to improve process control and, therefore, to give enhanced yield and improved product quality, but also frequently to ensure safety of operation.

Catalyst research where gases need to be analyzed often uses mass spectrometry because of its speed and flexibility.





**Figure 4** Peak observed with 1 ppt (parts per trillion) oxygen in nitrogen using an API source mass spectrometer.

### Semiconductor Manufacturing

Clean surfaces and pure reagent gases are essential requirements for semiconductor processing. Many fabrication techniques take place under vacuum, so partial pressure analysis is used to ensure that conditions are acceptable. For example, in a sputtering process, operating at a pressure of 0.01 mbar of argon, the presence of oxygen or water at a concentration as low as  $10^{-6}$  mole fraction may have an adverse effect on the process. It is therefore necessary to achieve a base pressure in the vacuum system better than  $10^{-7}$  mbar before back-filling with argon.

The gas composition can also be monitored during the fabrication process. The operating pressure is generally too high for direct operation of a mass spectrometer head, so the sample is taken to a separately pumped analyzer system. It is advantageous to employ an enclosed ion source, in which the ionization volume is operated at a pressure higher than that of the mass analyzer. This gives improved detection limits, since the effects of background pressure in the analyzer are reduced. Some designs can be operated with the ionization volume at or close to the process pressure, which makes for relatively non-invasive sampling and high sensitivity.

Detection limits of parts per million (mol) or even lower can be achieved for the common gases, such as hydrogen, low molecular weight hydrocarbons, nitrogen, oxygen, and water vapor.

Special techniques, such as API and collision-induced dissociation (CID) have been applied to detect

contaminants in extremely pure gases. For example, **Figure 4** shows the detection of parts per trillion (mol) of oxygen in nitrogen using a commercial instrument.

### Environmental Trace Gas Analysis

In trace gas analysis, it is required to monitor compounds, frequently organics, in atmosphere at low concentrations (parts per million (mol) or lower). Mass spectrometry offers versatile instrumentation for this, particularly when sensitivity and selectivity are enhanced by special inlet techniques or ionization processes.

A silicone-rubber membrane can be used to enhance the sensitivity to organics, detection limits of less than parts per million (mol) being obtainable for many species, and as low as parts per billion (mol) in some cases. When monitoring reactive species, an enclosed ion source is recommended for improved linearity. By using special ionization methods, the performance can be enhanced for specific compounds. For example, high-pressure ionization, followed by detection of negative ions, can yield a detection limit of parts per trillion (mol) for sulfur hexafluoride, which is strongly electronegative.

### Residual Gas Analysis in Vacuum Systems

Vacuum techniques are widely used in scientific research and in specialized industrial processes, such as manufacture of semiconductors. A pumping system is used to evacuate a housing to a pressure

sufficiently low for work to proceed. The process of evacuation may take several hours or even days if a very low pressure is required.

The pressure is limited by the rate at which material enters the gas phase in the housing, and by the capacity of the pumping system. If the gas load is too high, the system will fail to achieve its design pressure. In such a case, it is often helpful to establish the composition of the gas, since this will usually suggest appropriate remedial actions. A small mass-spectrometer, known as a 'residual gas analyzer' or a 'partial pressure analyzer' is well-suited to this task. Using helium as a tracer gas, the origin of leaks can be readily identified.

### Other Applications

Mass spectrometry has been extensively used to monitor the composition of the upper atmosphere. The requirements of compact size, light weight, and low power consumption stimulated the development of many different designs. Obtaining a representative sample presents unique problems, especially if radicals or ions are to be measured. The quadrupole is particularly suitable for monitoring energetic ions and neutrals.

The gases evolved during thermogravimetric analysis may be analyzed by mass spectrometry. A typical installation consists of a quadrupole mass spectrometer, fitted with an enclosed ion source, and sampling the effluent from a thermal analyzer via a heated capillary inlet with by-pass.

Isotope ratio monitoring, often needing very high precision and therefore requiring a magnetic sector

analyzer, is used for a variety of investigations in biological, medical, and geological research.

**See also:** **Air Analysis:** Outdoor Air. **Mass Spectrometry:** Electron Impact and Chemical Ionization; Mass Separation; Environmental Applications. **Process Analysis:** Overview. **Thermal Analysis:** Temperature-Modulated Techniques.

### Further Reading

- Batey JH (1987) Quadrupole gas analysers. *Vacuum* 37: 658–659.
- Colin T (1997) Process mass spectrometry as a tool for process understanding. *Journal of Process Analytical Chemistry* II: 410–415.
- Dawson PH (ed.) (1976) *Quadrupole Mass Spectrometry and its Applications*. Amsterdam: Elsevier.
- Ketola RA, Kotiaho T, Cisper ME, and Allen TM (2002) Environmental applications of membrane introduction mass spectrometry. *Journal of Mass Spectrometry* 37: 457–476.
- Pollard D and Salmon P (1999) Vent gas analysis. In: Michael CF and Stephen WD (eds.) *Encyclopedia of Bioprocess Technology: Fermentation, Biocatalysis, and Bioseparation*, pp. 2622–2637. New York: Wiley.
- Traynor P (1998) Developments in process mass spectrometry. *Journal of Process Analytical Chemistry* IV: 33–38.
- White FA and Wood GM (1986) *Mass Spectrometry Applications in Science and Engineering*. New York: Wiley-Interscience.
- Wright RG (1998) On-line analysis of complex gas mixtures by mass spectrometry. *Journal of Process Analytical Chemistry* IV: 71–78.

## Peptides and Proteins

**G P McMahon**, Dublin City University, Dublin, Republic of Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Since the late 1950s, mass spectrometry (MS) has been used to study peptides. However, until as recently as 15 years ago, mass spectrometers were used almost exclusively by organic chemists to investigate small molecules using field desorption techniques. The reason for this was that there were many problems with derivatization of the analytes and the low mass range was restrictive. But in the 1980s, two techniques became popular, which allowed direct

MS without derivatization – fission-fragment-induced desorption and fast atom bombardment, also known as liquid secondary ion mass spectrometry. Both were for proteins but still limited to some extent by sensitivity and mass range (up to 30 kDa). As most biomolecules are sensitive to organic solvents, substantial effort then went into bringing the biomolecules into the gas phase gently and in their natural state using 'softer' ionization techniques. So, in the past decade, protein identification and characterization became possible due to the soft ionization methods of matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) that allowed researchers to put molecules into the gas phase much more gently than previously. They also had good sensitivity and better mass ranges

sufficiently low for work to proceed. The process of evacuation may take several hours or even days if a very low pressure is required.

The pressure is limited by the rate at which material enters the gas phase in the housing, and by the capacity of the pumping system. If the gas load is too high, the system will fail to achieve its design pressure. In such a case, it is often helpful to establish the composition of the gas, since this will usually suggest appropriate remedial actions. A small mass-spectrometer, known as a 'residual gas analyzer' or a 'partial pressure analyzer' is well-suited to this task. Using helium as a tracer gas, the origin of leaks can be readily identified.

### Other Applications

Mass spectrometry has been extensively used to monitor the composition of the upper atmosphere. The requirements of compact size, light weight, and low power consumption stimulated the development of many different designs. Obtaining a representative sample presents unique problems, especially if radicals or ions are to be measured. The quadrupole is particularly suitable for monitoring energetic ions and neutrals.

The gases evolved during thermogravimetric analysis may be analyzed by mass spectrometry. A typical installation consists of a quadrupole mass spectrometer, fitted with an enclosed ion source, and sampling the effluent from a thermal analyzer via a heated capillary inlet with by-pass.

Isotope ratio monitoring, often needing very high precision and therefore requiring a magnetic sector

analyzer, is used for a variety of investigations in biological, medical, and geological research.

**See also:** **Air Analysis:** Outdoor Air. **Mass Spectrometry:** Electron Impact and Chemical Ionization; Mass Separation; Environmental Applications. **Process Analysis:** Overview. **Thermal Analysis:** Temperature-Modulated Techniques.

### Further Reading

- Batey JH (1987) Quadrupole gas analysers. *Vacuum* 37: 658–659.
- Colin T (1997) Process mass spectrometry as a tool for process understanding. *Journal of Process Analytical Chemistry* II: 410–415.
- Dawson PH (ed.) (1976) *Quadrupole Mass Spectrometry and its Applications*. Amsterdam: Elsevier.
- Ketola RA, Kotiaho T, Cisner ME, and Allen TM (2002) Environmental applications of membrane introduction mass spectrometry. *Journal of Mass Spectrometry* 37: 457–476.
- Pollard D and Salmon P (1999) Vent gas analysis. In: Michael CF and Stephen WD (eds.) *Encyclopedia of Bioprocess Technology: Fermentation, Biocatalysis, and Bioseparation*, pp. 2622–2637. New York: Wiley.
- Traynor P (1998) Developments in process mass spectrometry. *Journal of Process Analytical Chemistry* IV: 33–38.
- White FA and Wood GM (1986) *Mass Spectrometry Applications in Science and Engineering*. New York: Wiley-Interscience.
- Wright RG (1998) On-line analysis of complex gas mixtures by mass spectrometry. *Journal of Process Analytical Chemistry* IV: 71–78.

## Peptides and Proteins

**G P McMahon**, Dublin City University, Dublin, Republic of Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Since the late 1950s, mass spectrometry (MS) has been used to study peptides. However, until as recently as 15 years ago, mass spectrometers were used almost exclusively by organic chemists to investigate small molecules using field desorption techniques. The reason for this was that there were many problems with derivatization of the analytes and the low mass range was restrictive. But in the 1980s, two techniques became popular, which allowed direct

MS without derivatization – fission-fragment-induced desorption and fast atom bombardment, also known as liquid secondary ion mass spectrometry. Both were for proteins but still limited to some extent by sensitivity and mass range (up to 30 kDa). As most biomolecules are sensitive to organic solvents, substantial effort then went into bringing the biomolecules into the gas phase gently and in their natural state using 'softer' ionization techniques. So, in the past decade, protein identification and characterization became possible due to the soft ionization methods of matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) that allowed researchers to put molecules into the gas phase much more gently than previously. They also had good sensitivity and better mass ranges

(MALDI >300 kDa, ESI >100 kDa). These inventions triggered the widespread use of MS in proteomics (and earned the inventors John Fenn and Koichi Tanaka a shared Chemistry Nobel prize in 2002). There have also been developments in the detector end of the mass spectrometer with time-of-flight (TOF), ion trap, ion cyclotron resonance, and tandem mass spectrometers now available. Also, the exponentially growing amount of sequence information available in databases is helping enormously.

Proteomics is a far more complex field than genomics as it encompasses the characterization and functional analysis of all proteins that are expressed by the genome at a certain moment, under certain conditions. The proteome, unlike the genome, is also highly dynamic. Cellular processes, especially in disease states, are determined by multiple proteins, and thus it is important to be able to investigate not only on one single gene product (protein) but to be able to study the complete set of gene products (the proteome). Mass spectrometers are a valuable analytical tool used for measuring the molecular weight of a sample. This procedure is useful for the structural elucidation of organic compounds, for peptide or oligonucleotide sequencing, and for monitoring and detecting specific drug metabolites in biological matrices, to name but a few applications. Protein identification can also be carried out using certain types of mass spectrometers, usually tandem mass spectrometers, and fragmenting the sample and analyzing the products generated. Quantitative proteomics is now also possible using labeling techniques. MS is also extremely useful for the detection of post-translational modifications such as phosphorylation and glycosylation. Further investigation of protein structure such as folding and binding can also be achieved using mass spectrometric methods.

## **Protein and Peptide Molecular Weight**

Many analytical problems in protein and peptide research can be answered by obtaining the molecular mass of the compound. The mass accuracy is disputed somewhat but for a few picomoles of a >10 kDa molecule, somewhere ~0.1–0.01% would be expected depending on the ionization/desorption technique used and the mass analyzer employed. For large samples such as biomolecules, molecular weights can be measured to within an accuracy of 0.01% of the total molecular weight of the sample, i.e., within 4 Da or atomic mass units (amu) error for a sample of 40 000 Da. This is sufficient to allow minor mass changes to be detected, e.g., the substitution of one amino acid for another, or a post-translational modification. For small organic

molecules the molecular weight can be measured to within an accuracy of 5 ppm, which is often sufficient to confirm the molecular formula of a compound, and is also a standard requirement for publication in a chemical journal. Other important applications for molecular mass are confirmation of an expected structure, verification of an amino acid sequence (only in terms of the sums of the masses of the amino acids present), and monitoring of the formation or disappearance of reaction products.

## **Protein Identification**

In recent years, as a complement to Edman sequencing, mass spectral identification has emerged as a means to identify proteins. Proteomic analyses are usually carried out by using two-dimensional gel electrophoresis (2DGE) followed by MS and database searches. This approach is most suitable for the analysis of pure proteins or very simple mixtures. Since the mass of a protein alone is not enough information to confirm its identity, there are two main procedures used for identifying a protein by MS. The first is called peptide mass fingerprinting and is usually carried out using MALDI-TOF MS. The second, often employed when the first has not been totally successful, is called sequence tagging and is usually performed on an ESI-ion trap or ESI-quadrupole-TOF mass spectrometer. There is a third, more recent multi-dimensional or hyphenated technique called liquid chromatography–mass spectrometry (LC–MS) that can also be used under certain circumstances.

### **Technique 1 – Peptide Mass Fingerprinting**

Protein identification by in-gel digestion and peptide mass fingerprinting is almost a routine method in MS laboratories. The protein sample is digested with a specific enzyme, e.g., trypsin, which cuts in a sequence-specific manner to produce a defined set of peptides. Trypsin is useful for mass spectrometric studies because each proteolytic fragment contains a basic arginine or lysine amino acid residue, and thus is eminently suitable for positive ionization mass spectrometric analysis. The peptides are extracted from the gel, sometimes desalted and then the digest mixture can be directly deposited onto the MALDI target plate and the rest of the sample can be stored for subsequent analysis, e.g., by ESI-MS. The peptide mixture is then analyzed by MS to produce a rather complex spectrum from which the molecular weights of all of the proteolytic fragments can be read. This spectrum, with its molecular weight information, is called a peptide map or peptide mass fingerprint (PMF). This PMF can be used to search protein

databases. If the protein already exists on a database, then the peptide map is often sufficient to confirm the protein (Figure 1).

A search algorithm is used that carries out virtual digests of protein sequences based on the sequence-specificity of trypsin and then calculates the masses of the predicted peptides from first principles (for example, by adding up the masses of the individual atoms). The hits are generally ranked according to the number of peptides that match. Unique identification does not require the whole protein to be covered by the tryptic peptides. Usually 10–20% coverage is sufficient. The great precision of mass spectrometers nowadays allows the discrimination of even extremely similar proteins, differing in structure by only one amino acid. The mass accuracy is an important aspect of this approach and with a mass accuracy of 10–30 ppm for peptides in a MALDI-TOF system, usually four or five peptides are enough to identify the protein unambiguously. MALDI produces singly charged ions and hence the spectra are easy to interpret.

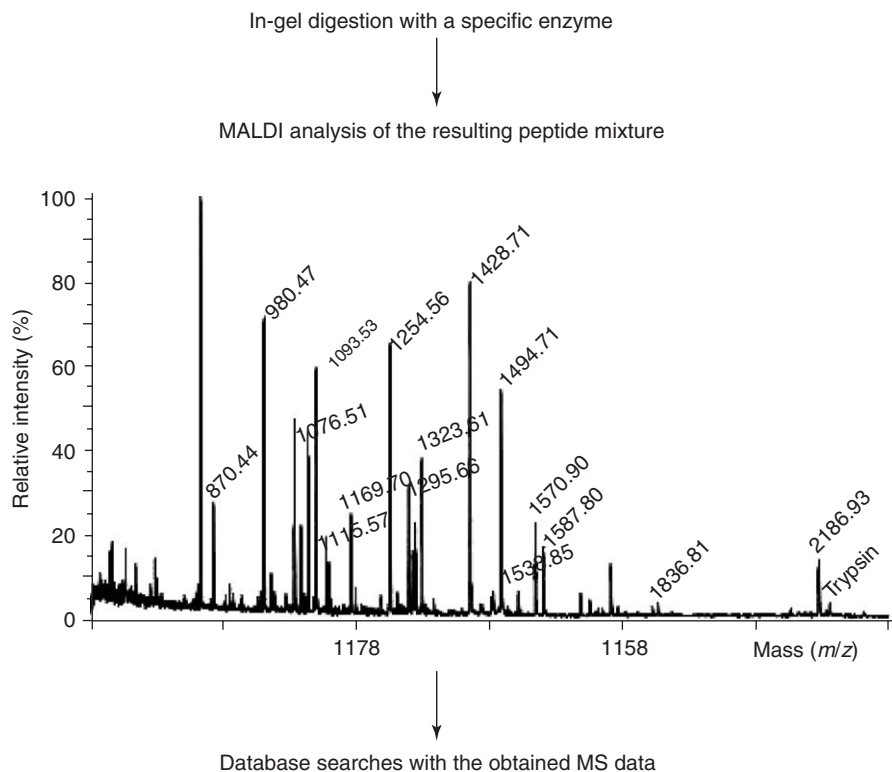
There are now available automated systems capable of cutting the spots out of the gel, digesting the samples, desalting, and then spotting onto the

MALDI plates. Data processing and database searching can also be automated but in practice, when small amounts of protein digests are analyzed, the spectra have to be checked manually before database searching. Instead of excising individual spots from the gel, some researchers proteolytically digest all proteins on the gel at the same time, transfer to a membrane, and directly scan by MALDI-TOF. This molecular scanning approach takes an enormous amount of time and memory, e.g., for a  $16 \times 16 \text{ cm}^2$  membrane, 36 days of continuous scanning and more than 40 Gb of raw data.

PMF is only successful when the digested protein exists in the protein or genomic databases, if the fingerprint is unique, and when four or more peptides are obtained by MALDI. If PMF is not successful, it will be necessary to obtain sequence information from the peptides. The most common way to do this is by ESI-MS/MS that is described below.

## Technique 2 – Sequence Tagging

This second approach takes the peptides after digestion and fragments them in the tandem mass spectrometer yielding partial random amino acid



**Figure 1** The mass accuracy of the peptides measured is critical in PMF; therefore, the spectra have to be internally calibrated. If trypsin is used for in-gel digestion, the enzyme will also cleave itself into fragments, which can be used for calibration. (Reprinted with permission from Nyman TA (2001) The role of mass spectrometry in proteome studies. *Biomolecular Engineering* 18(5): 221–227; © Elsevier.)



sequences from the peptides (sequence tags). Database searches are performed using both the molecular weight information and the sequence tags, and the chance of a unique hit in the database is considerably enhanced. The advantage of using the sequence tag approach is that database searches are much more accurate when both peptide molecular weight and sequence information are used compared with molecular weight only. A sequence tag from one peptide is often sufficient to identify the protein. The fragment ions can also be ordered by size and the masses of sequential fragments used to establish which amino acids have been cleaved off. In this way, the protein sequence can be established and used to search sequence databases to find related rather than exact matches. If a match is not found, additional peptides have to be mass spectrometrically sequenced. Determination of the total sequence of a protein takes considerably more effort. Compared to MALDI-MS, ESI-MS is more time consuming and laborious. The analysis time is also limited since the liquid sample is being consumed throughout the ESI-MS experiment. ESI-MS can be automated by using LC prior to the MS. LC separation before MS is especially useful when complex mixtures are being analyzed. One useful strategy is to get the software to switch from MS mode to MS/MS mode when a peak elutes from the LC into the MS causing a surge in the ion current. This is a product ion scan. This generates very useful data in an automated way and can be set up to automatically fragment and collect MS/MS data on any number of peptides observed in the MS spectrum based on their intensity,  $m/z$  value, or charge state.

### Technique 3 – LC-MS Approach

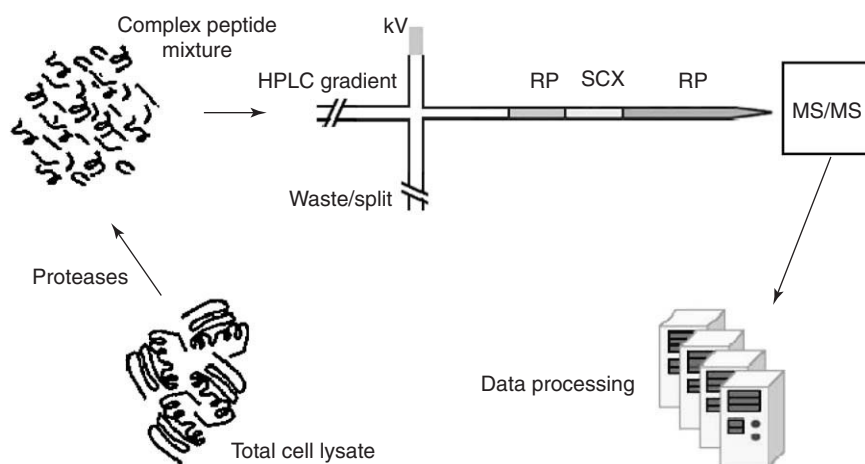
In this approach, the whole cell lysate is digested chemically or with a protease. This generates a very complex set of peptides. For the analysis of such complex mixtures, several multidimensional techniques are possible involving combinations of chromatography and/or capillary electrophoresis (CE). Multidimensional chromatography coupled to MS/MS for proteomics has been called multidimensional protein identification technology (MudPIT). One group of researchers has used MudPIT to analyze the proteome of *Saccharomyces cerevisiae* in which a total of 1484 proteins were detected and identified. Even low abundance proteins such as transcription factors and protein kinases were identified. Multidimensional LC-MS has also proved very important for the analysis of membrane proteins that are usually underrepresented on conventional 2D gels.

MudPIT uses the ability of tandem mass spectrometers to select and analyze peptides from protein

mixtures that have been proteolytically digested. The peptides then serve as surrogate markers for the protein sequence. Proteins are identified by searching the resulting tandem mass spectra through sequence databases. Because digested protein mixtures create complex mixtures of peptides, high-performance separation techniques are required to resolve the peptides prior to entering the tandem MS. The online MudPIT process fully automates the separation and identification of proteins from complex mixtures. The first step in such an experiment is the enzymatic digestion of a protein mixture to produce a more complex peptide mixture. A specially packed bi- or triphasic column containing sections of strong cation exchange (SCX) and reversed-phase (RP) materials is used to separate the peptides. Peptides are eluted from the SCX phase onto the RP part of the column in a series of salt steps that increase in concentration. A gradient in the latter RP part of the column separates and delivers the peptides onto the MS in manageable numbers. MudPIT appears to introduce very little bias in its sampling of proteins from the cell but the sample preparation procedure is very important. MudPIT-type analyses require high throughput database searching techniques (Figure 2).

### Quantitative Proteomics

An increasing emphasis in proteomics is the quantitation of protein content rather than simple determination of presence or absence. The traditional way to do this is by comparing the densities of corresponding spots in different 2D gels. This provides only a rough estimate, as it is very dependent on sample processing, staining efficiency of individual gels, and subsequent spot detection. As MS is not inherently quantitative, other methods have appeared that allow it to be. These methods are based on the direct/linear relationship between the amount of peptide/protein in a particular sample and its peak height in the mass spectrum. However, direct quantitation of proteins and peptides is very difficult because the absolute intensities of the detected ions are determined by their physiochemical properties. Therefore, internal standards are needed with chemical properties very similar to the analyte. Absolute quantities of peptides can also be measured using known quantities of peptide standards, but the method is not suitable for proteins. Protein quantitation can be achieved by labeling with a mass label to differentiate the same proteins for relative quantitation. This technique is called isotope dilution and involves spiking the sample with the analyte modified with a stable isotope, e.g., H-2, C-13, or N-15. The intensity ratio of the internal standard to the analyzed



**Figure 2** Schematic overview of the three-phase multidimensional protein identification technology (MudPIT). A total cell lysate is digested by a protease (i.e., trypsin). The complex peptide mixture is loaded on a capillary column. The LC gradient is delivered to a junction and part of the flow is split off through the column to reduce the flow to a couple of hundreds of nanoliters per minute. The peptides are separated by a tri-phasic column, with in line a reversed phase, strong cation exchange, and another reversed phase, before they enter the ionization source of the mass spectrometer. Each peptide elutes in a time window of several seconds allowing the mass measurements and the determination of peptide sequences by tandem mass spectrometry. These data are used for protein identification using database searches. (Reprinted with permission from Romijn EP, Krijgsveld J, and Heck AJR (2003) Recent liquid-chromatographic-(tandem) mass spectrometric applications in proteomics (review). *Journal of Chromatography A* 1000(2): 589–608; © Elsevier.)

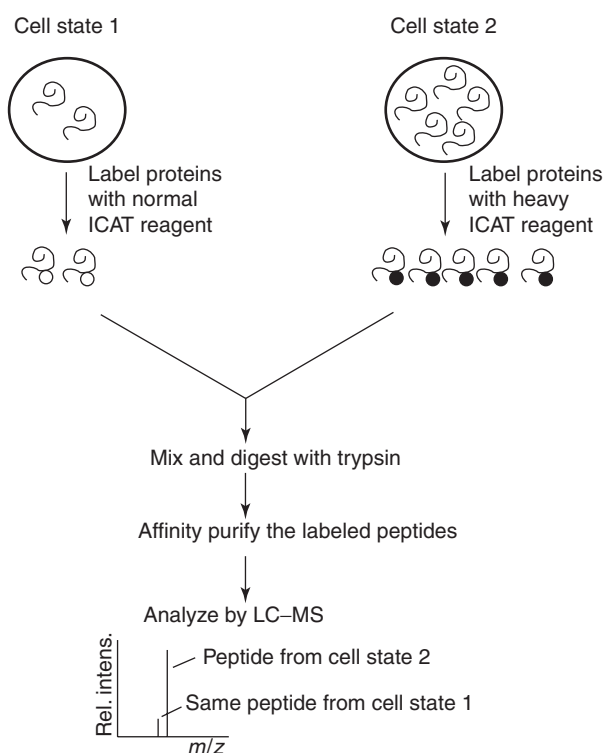
compound is then used for quantitation. If the absolute concentration of the standard is known, then the quantity of the analyte can be determined accurately. There are four main types of labeling strategies. The first method uses metabolic labeling to incorporate N-15 into peptides. The second method uses proteolytic digestion to generate peptides in the presence of O-18-containing water to incorporate a label. The third method uses covalent labeling that is usually targeted at reactive amino acid residues and generally the label contains deuterium atoms in place of hydrogen atoms. The last method is protein derivatization with isotope-coded affinity tags (ICAT) (Figure 3).

In ICAT labeling, tags of different isotope composition (heavy and light) are reacted with two protein populations to quantitatively differentiate between proteins altered in abundance. The ICAT reagent itself consists of a thiol-specific group, a linker region containing either eight hydrogen atoms (d, light) or eight deuterium atoms (d8, heavy), and a biotin group. A protein sample from one source is reacted with the d form of the tag and another with the d8 version. These samples are then combined and digested with a proteolytic enzyme. The biotin moiety is then used to selectively capture reacted peptides using avidin chromatography. Since the method targets cysteine-containing peptides, this results in a substantial reduction in sample complexity so that a reduced number of peptides need to be analyzed. Combined LC-MS/MS can be used at this stage for

identification by peptide fragmentation and the abundance ratio of peptides derivatized by either label can be determined by integration of the peaks separated by 8 Da. One example of a large-scale study applying ICAT showed the differential analysis of microsomal fractions of human cells that resulted in the identification and quantitation of 491 proteins.

### Analysis of Co- or Post-Translational (Chemical) Modifications

Further complicating the analysis of proteins are the chemical modifications that can occur co-translationally or post-translationally to modify or regulate their functions. There are more than 100 modification types already described in the literature and it is most likely that more will be found. MS has been used to study many of them such as phosphorylation, glycosylation, methylation, formylation, acetylation, palmitoylation, lipidation, and sulfonation. MS is particularly important for this application as recent studies have shown that mRNA expression and protein levels do not always correlate. Proteins can be post-translationally modified or present as different isoforms, and these modifications cannot be predicted from the corresponding DNA sequences. Post-translational modifications play an important role in determining protein function, but identifying peptides containing these modifications poses many challenges. Modified peptides such as those bearing



**Figure 3** The principle of ICAT. The ICAT reagents have a thiol-specific reactive group and a biotin group. This way it is possible to label all peptides containing cysteines and then avidin affinity purify them. The cysteines in proteins in cell state 1 are labeled with normal ICAT reagent, and the cysteines in cell state 2 with deuterated, heavy ICAT reagent. Then the samples are pooled, digested with trypsin, and ICAT-labeled peptides are isolated with affinity chromatography followed by analysis with LC-MS. Protein levels can be quantitated by measuring peak ratios between normal ICAT-derivatized peptide and deuterated ICAT-derivatized peptide. For protein identification peptides are sequenced by MS/MS. (Reprinted with permission from Nyman TA (2001) The role of mass spectrometry in proteome studies. *Bio-molecular Engineering* 18(5): 221–227; © Elsevier.)

the important phosphorylation modification may produce fragment ion spectra more complex than normal peptides, complicating their interpretation. This means that augmented forms of existing databases matching algorithms typically take far longer to run on spectra when modification possibilities are taken into account. The use of MS to investigate phosphorylation, glycosylation, and ubiquitination modifications are discussed below.

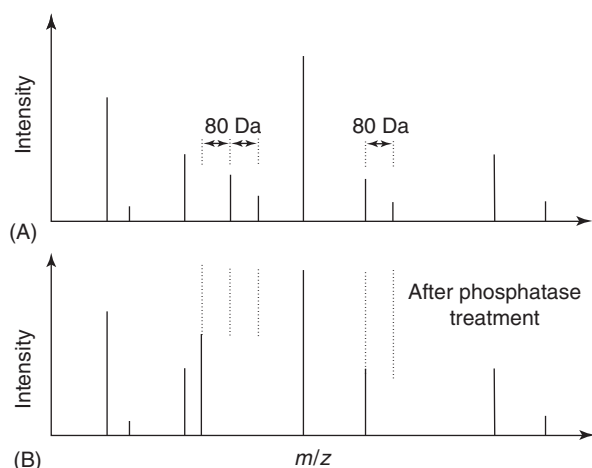
### Phosphorylation

Although phosphorylation is observed on a number of amino acid residues, by far the most common and most important sites are on serine, threonine, and tyrosine residues. Reversible phosphorylation and dephosphorylation is an extremely important modulator of protein function. MS has emerged as

an increasingly viable tool for the task of elucidating the presence and sites of O-phosphorylation. MS can firstly and accurately provide the molecular mass of the intact phosphorylated protein. When the molecular mass is compared with that of the unmodified protein and/or the protein treated with phosphatase, the number of attached phosphate groups can be determined. To fully understand protein phosphorylation, it is necessary to identify the exact amino acid residues that become phosphorylated.

Protein phosphorylation can be studied by MS provided there is a good amount of starting material (more than a spot on a gel) and that the resolution of the mass spectrometer is good. The first step in analysis of phosphoproteins by MS is enrichment. This is because MS of proteolytic digests of proteins rarely provides 100% coverage of the protein sequence. Phosphorylation is often sub-stoichiometric such that the phosphopeptide is present in lower abundance than other peptides from the initial protein. Also, the MS response to a phosphopeptide may be suppressed relative to its unphosphorylated counterpart, so analysis of phosphopeptides is easier if they are first enriched and the best way to achieve this is to use immobilized metal affinity chromatography. In this technique, metal ions are bound to a chelating support. Phosphopeptides are selectively bound because of the affinity of the phosphate groups for metal ions. The phosphopeptides can then be released by high pH or phosphate buffer. Other methods of enrichment for phosphopeptides include the use of antibodies for immunoprecipitation and chemical tagging of phosphorylation sites. The second step in the analysis of phosphoproteins is to perform trypsin digestion and generate a PMF by MALDI-MS. Examination of the peptide map for peptides shifted by multiples of 80 Da ( $\text{HPO}_3$ ) or 98 Da ( $\text{H}_3\text{PO}_4$ ) can be effective in providing a signature for phosphopeptides. Differences in the peptide map before and after phosphatase treatment can further aid analysis since their molecular weight will be reduced by 80 or 98 Da after the phosphatase treatment. The peptides that lose their phosphate group can be sequenced from the MALDI sample. Subsequent MS analysis can then be offline or online, perhaps with a hyphenated technique in between such as LC or CE. In order to quantify phosphorylation, some workers have used the relative intensities of the peptide signals by ESI-MS as a measure of the extent of phosphorylation. Isotope labeling or the introduction of a mass tag are alternative methods of quantitation.

Other mass spectrometric experiments that can be useful for studying phosphorylation include post-source decay, precursor ion scan, neutral loss scanning, and stepped skimmer potential. In order to



**Figure 4** Phosphopeptide identification by MALDI-TOF-MS mapping combined with alkaline phosphatase treatment. (A) The MALDI-TOF-MS spectrum of a proteolytic digest. Phosphopeptides are indicated by peaks shifted by multiples of 80 Da ( $\text{HPO}_3 = 80 \text{ Da}$ ) relative to predicted unphosphorylated peptide masses. (B) The disappearance of such peaks upon treatment with a phosphatase confirms their identity as phosphopeptides. (Reprinted with permission from McLachlin DT and Chait BT (2001) Analysis of phosphorylated proteins and peptides by mass spectrometry (review). *Current Opinion in Chemical Biology* 5(5): 591–602; © Elsevier.)

identify the positions of phosphorylation sites, collision induced dissociation (CID) MS/MS by ESI is commonly used. Fragment ions can be measured and loss of phosphate as  $\text{HPO}_3$  or  $\text{H}_3\text{PO}_4$  is a favored fragmentation event. Separation methods can be coupled to the MS online in order to improve the CID analysis of complex proteolytic mixtures (Figure 4).

### Glycosylation

Protein glycosylation is another post-translational modification that is essential for *in vivo* functions. Glycoproteins play an important role in cell–cell and cell–substratum recognition events in multicellular organisms. Glycosylation is probably the most complex post-translational modification. There are a number of types of glycosylation but one in particular – O-Glc-NAc modification – is of great biological relevance and can be studied by MS. This type of modification is believed to be a regulatory modification. Most known O-Glc-NAc proteins are also phosphoproteins. Historically, the method used to study these modifications involved galactosyltransferase radiolabeling of O-Glc-NAc moieties, generation of glycopeptides by proteolytic cleavage, purification by LC, and Edman protein sequencing. A more elegant approach involves enrichment of O-Glc-NAc-modified peptides from complex mixtures

using *Ricinus Communis* (RCA 1) affinity chromatography followed by galactosyltransferase labeling. The labeling of an O-Glc-NAc group with galactose using galactosyltransferase produces a disaccharide substituent with a diagnostic mass of  $m/z$  366. After enrichment, the peptides are separated by LC and analyzed by MS. This is achieved by specifically searching for peptides carrying the O-Glc-NAc modification by the monitoring of a specific reporter ion and the identification of the original protein from which the peptide came by CID and database searching. The removal of this disaccharide from the glycopeptide requires much less energy than the fragmentation of the peptide backbone. It is possible to take advantage of this difference in fragmentation energy. Both operations are performed automatically and simultaneously with femtomole sensitivity. In glycosylation studies, the glycans are liberated from the protein and the glycan structures are determined from the free glycans. The glycan structures can be characterized by combining MALDI measurements with exoglycosidase digestions. If more detailed structural information is required, ESI-MS can be used with product ion scans.

### Ubiquitination

Protein ubiquitination is among the most common of all post-translational modifications. Ubiquitin plays an important role as a signal molecule for protein degradation in eukaryotes and as a regulator of protein activity and localization. One strategy that has been used to analyze protein ubiquitination is by coupling LC and MS. Ubiquitin conjugates were purified from a yeast strain expressing Histidine-tagged ubiquitin or a control strain by denaturing nickel affinity chromatography. Following trypsin digestion, a ubiquitin-conjugated protein contains a di-glycine residue of ubiquitin covalently attached to a lysine residue that is resistant to proteolysis. The resulting peptides are separated by SCX chromatography. All fractions are sequentially analyzed by nano-LC–MS/MS. Not only can the proteins be identified in this way, but the exact sites of ubiquitination can also be identified in a percentage of the cases.

### Further Study of Protein Structure

Primary protein structure is the sequence of the amino acids in that protein and this has been discussed in the section on Protein identification. Secondary protein structure is the arrangement of the amino acid sequence that makes up the polypeptide backbone of the protein. Tertiary protein structure is the overall

three-dimensional folded shape of the protein when it is in its native state. Quaternary structure means the arrangement of subunits within multi-subunit proteins, e.g., hemoglobin. Therefore, together, secondary, tertiary, and quaternary structures govern how proteins fold and interact with each other and other molecules. Methods employing ESI-MS have been able to provide information on a protein's higher-order structure such as the use of hydrogen/deuterium exchange experiments to study the folding rates of hen egg-white lysozyme. Hydrogen/deuterium exchange experiments have also allowed determination of the relative amounts of  $\alpha$ -helices and  $\beta$ -sheets in peptides.

Protein-protein interactions such as those between antibody and antigen and enzyme and substrate pairs can be studied by MS. Since many cellular functions are performed by protein complexes and not by individual proteins, it is important to be able to identify the interacting protein components. As ESI is a gentle technique, the detection of weakly bound (noncovalent) complexes in the gas phase is possible. There is evidence to suggest that these ESI-MS observations can reflect the expected interactions in solution phase and can even determine the strength of these interactions. Mass spectrometric experiments with multimeric proteins can yield information on both the stoichiometry and molecular nature of subunit interactions.

Peptide-metal and protein-metal interactions can also be studied using MS and the specific binding of metals can be distinguished from nonspecific interactions such as adduct formation.

## Future Prospects

At the moment, proteome analyses are normally carried out by 2DGE followed by MS, and MS has proven to be the method of choice for large-scale, high-throughput protein identification. The main drawback of this procedure is the 2DGE step where low abundance proteins often cannot be detected and it is also quite a laborious and sometimes irreproducible technique. LC linked to MS has gone some way toward overcoming the shortcomings of the 2DGE step. The technology for proteome analysis is becoming more and more a three-step procedure – separation science chromatography for separation of the proteins and peptides, MS for the identification and characterization of these biomolecules, and bioinformatics for data management. It is expected that nano-LC combined with tandem MS and database searching will dominate the field of protein identification going forward.

Another approach that is being tried is called *de novo* peptide sequencing by MS and it differs substantially from the peptide identification processes that rely on the presence of the target peptide sequence in a supplied database. The *de novo* challenge is to infer a peptide sequence directly from the spectrum without any other information as to what the sequence may be. Several attempts have been made but success has been limited by data quality and hindered by the complexity of accurately predicting fragment ion spectra. It is unlikely ever to be as sensitive as current bioinformatics and database searching methods but with constantly emerging improvements being made in the field of MS for peptides and proteins, there may come a time in the future when this becomes the method of choice.

**See also:** **Mass Spectrometry:** Ionization Methods Overview; Atmospheric Pressure Ionization Techniques; Electrospray; Matrix-Assisted Laser Desorption/Ionization; Time-of-Flight. **Proteins:** Traditional Methods of Sequence Determination. **Proteomics.**

## Further Reading

- Burlingame AL, Carr SA, and Baldwin MA (eds.) (1999) *Mass Spectrometry in Biology and Medicine*. Totowa, NJ: Humana Press.
- Chapman JR (ed.) (2000) *Mass Spectrometry of Proteins and Peptides (Methods in Molecular Biology)*. Totowa, NJ: Humana Press.
- Dass C (ed.) (2000) *Principles and Practice of Biological Mass Spectrometry*. New York: Wiley.
- De Hoffmann E (ed.) (2001) *Mass Spectrometry: Principles and Applications*. New York: Wiley.
- James P (ed.) (2000) *Proteome Research: Mass Spectrometry (Principles and Practice)*. New York: Springer.
- Kinter M (ed.) (2000) *Protein Sequencing and Identification Using Tandem Mass Spectrometry*. New York: Wiley.
- Lin D, Tabb DL, and Yates JR (2003) Large-scale protein identification using mass spectrometry (review). *Biochimica et Biophysica Acta – Proteins and Proteomics* 1646(1–2): 1–10.
- Loo JA and Ogorzalek Loo RR (1997) Electrospray ionization mass spectrometry of peptides and proteins. In: Cole RB (ed.) *Electrospray Ionization Mass Spectrometry – Fundamentals, Instrumentation and Applications*, pp. 385–419. New York: Wiley.
- Marvin LF, Roberts MA, and Fay LB (2003) Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in clinical chemistry (review). *Clinica Chimica Acta* 337(1–2): 11–21.
- McLachlin DT and Chait BT (2001) Analysis of phosphorylated proteins and peptides by mass spectrometry (review). *Current Opinion in Chemical Biology* 5(5): 591–602.



Nyman TA (2001) The role of mass spectrometry in proteome studies. Review. *Biomolecular Engineering* 18(5): 221–227.

Romijn EP, Krijgsveld J, and Heck AJR (2003) Recent liquid-chromatographic–(tandem) mass spectrometric

applications in proteomics (review). *Journal of Chromatography A* 1000(2): 589–608.

Siuzdak G (ed.) (2003) *The Expanding Role of Mass Spectrometry in Biotechnology*. MCC (Mass Consortium Corporation) Press, San Diego.

## Polymerase Chain Reaction Products

**A Fox, K F Fox, E Castanha, and W T Muhammad,**  
University of South Carolina, Columbia, SC, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Generally, techniques to detect genetic diversity are based on the known genomic sequence. Specific genes are often amplified by the polymerase chain reaction (PCR). The detection of genetic variants involves properties of DNA that indirectly reflect sequence: e.g., recognition of a region in a strand by a complementary sequence (as observed with microarrays), distinct mobility on electrophoretic separation, or differential digestion of specific DNA regions by enzymes (e.g., restriction enzymes). Alternatively, mass spectrometry (MS) directly determines DNA structure. Electrospray ionization (ESI)-MS discriminates two structurally related PCR products by molecular weight (MW) difference. Whilst tandem mass spectrometry (MS/MS) differentiates PCR products by changes in molecular fragmentation pattern (analogous to an enzymatic restriction digest but providing much greater information) directly reflecting sequence differences at each position on both DNA strands. Examples are given here to illustrate the type of information that can be obtained using ESI-MS and MS/MS analysis involving: (1) discrimination of structurally related PCR products from genetically related bacterial species or strains with implications for detection of a biological warfare attack or forensics; or (2) recognizing molecular diversity among normal or diseased human populations (e.g., cancer patients) with a role in understanding etiology, diagnosis, and treatment.

The state-of-the-art technology to be discussed here employs two techniques that are each powerful in their own right: PCR and tandem mass spectrometry. However, when used together they display synergism that allows information to be obtained, in real time, that is difficult to obtain by either alone. Classical PCR involves detection of a PCR by electrophoretic mobility on a gel, which is time

consuming. Real-time PCR is distinct from classical PCR, in that electrophoresis is avoided and the PCR product is detected simply by an increase in fluorescence. For example, the PCR reaction mixture contains a dye that is not fluorescent, except when bound to double stranded (ds) DNA. Thus, as the amount of ds DNA increases on amplification, fluorescence also increases. Unfortunately, PCR is a technique generally used by molecular biologists, whereas MS is employed traditionally by analytical chemists. Thus, methodologies, concepts, and terms used by one discipline may be unfamiliar to the other. This explains why, in certain places, definitions are given here that might be obvious to one group but not the other. It also helps explain why, to this point, PCR-MS and PCR-MS/MS has largely been used by a handful of researchers rather than as a routine diagnostic tool. It is argued here that PCR-MS and MS/MS technology has matured to a stage that widespread application is indeed not only possible but desirable. However, mutual education and/or collaboration of molecular biologists and analytical chemists is essential, and this is part of what we attempt to accomplish here in this article.

For the non-mass spectrometrists, it should be pointed out that nowadays the analysis of large molecules (e.g., PCR products) is primarily based on matrix-assisted laser desorption ionization (MALDI) mass spectrometry or ESI ionization. In the former case, the sample is present within an energy-absorbing matrix on a surface. When hit with a laser beam, the matrix absorbs the energy, transferring it to the molecule of interest (e.g., DNA). Generally, only a singly ionized species is produced having a single charge. In contrast, ESI-MS is performed in solution and the sample is introduced into the source of the mass spectrometer using a syringe pump. The solution is sprayed, in the presence of an electrical field, and charges are transferred to the resulting droplets. As the droplets evaporate, the charge is transferred to molecules present within the droplet. Ions are produced that can have multiple charge states (e.g., for a 30-mer, anywhere from  $-1$  to  $-30$  charges). Since mass analyzers (the next stage after the source)

Nyman TA (2001) The role of mass spectrometry in proteome studies. Review. *Biomolecular Engineering* 18(5): 221–227.

Romijn EP, Krijgsveld J, and Heck AJR (2003) Recent liquid-chromatographic–(tandem) mass spectrometric

applications in proteomics (review). *Journal of Chromatography A* 1000(2): 589–608.

Siuzdak G (ed.) (2003) *The Expanding Role of Mass Spectrometry in Biotechnology*. MCC (Mass Consortium Corporation) Press, San Diego.

## Polymerase Chain Reaction Products

**A Fox, K F Fox, E Castanha, and W T Muhammad,**  
University of South Carolina, Columbia, SC, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Generally, techniques to detect genetic diversity are based on the known genomic sequence. Specific genes are often amplified by the polymerase chain reaction (PCR). The detection of genetic variants involves properties of DNA that indirectly reflect sequence: e.g., recognition of a region in a strand by a complementary sequence (as observed with microarrays), distinct mobility on electrophoretic separation, or differential digestion of specific DNA regions by enzymes (e.g., restriction enzymes). Alternatively, mass spectrometry (MS) directly determines DNA structure. Electrospray ionization (ESI)-MS discriminates two structurally related PCR products by molecular weight (MW) difference. Whilst tandem mass spectrometry (MS/MS) differentiates PCR products by changes in molecular fragmentation pattern (analogous to an enzymatic restriction digest but providing much greater information) directly reflecting sequence differences at each position on both DNA strands. Examples are given here to illustrate the type of information that can be obtained using ESI-MS and MS/MS analysis involving: (1) discrimination of structurally related PCR products from genetically related bacterial species or strains with implications for detection of a biological warfare attack or forensics; or (2) recognizing molecular diversity among normal or diseased human populations (e.g., cancer patients) with a role in understanding etiology, diagnosis, and treatment.

The state-of-the-art technology to be discussed here employs two techniques that are each powerful in their own right: PCR and tandem mass spectrometry. However, when used together they display synergism that allows information to be obtained, in real time, that is difficult to obtain by either alone. Classical PCR involves detection of a PCR by electrophoretic mobility on a gel, which is time

consuming. Real-time PCR is distinct from classical PCR, in that electrophoresis is avoided and the PCR product is detected simply by an increase in fluorescence. For example, the PCR reaction mixture contains a dye that is not fluorescent, except when bound to double stranded (ds) DNA. Thus, as the amount of ds DNA increases on amplification, fluorescence also increases. Unfortunately, PCR is a technique generally used by molecular biologists, whereas MS is employed traditionally by analytical chemists. Thus, methodologies, concepts, and terms used by one discipline may be unfamiliar to the other. This explains why, in certain places, definitions are given here that might be obvious to one group but not the other. It also helps explain why, to this point, PCR-MS and PCR-MS/MS has largely been used by a handful of researchers rather than as a routine diagnostic tool. It is argued here that PCR-MS and MS/MS technology has matured to a stage that widespread application is indeed not only possible but desirable. However, mutual education and/or collaboration of molecular biologists and analytical chemists is essential, and this is part of what we attempt to accomplish here in this article.

For the non-mass spectrometrists, it should be pointed out that nowadays the analysis of large molecules (e.g., PCR products) is primarily based on matrix-assisted laser desorption ionization (MALDI) mass spectrometry or ESI ionization. In the former case, the sample is present within an energy-absorbing matrix on a surface. When hit with a laser beam, the matrix absorbs the energy, transferring it to the molecule of interest (e.g., DNA). Generally, only a singly ionized species is produced having a single charge. In contrast, ESI-MS is performed in solution and the sample is introduced into the source of the mass spectrometer using a syringe pump. The solution is sprayed, in the presence of an electrical field, and charges are transferred to the resulting droplets. As the droplets evaporate, the charge is transferred to molecules present within the droplet. Ions are produced that can have multiple charge states (e.g., for a 30-mer, anywhere from  $-1$  to  $-30$  charges). Since mass analyzers (the next stage after the source)

generally separate by the mass-to-charge ratio, simple spectra are generated for MALDI-MS but ESI spectra are considered to be complex. However, as noted below, a simple computer-assisted operation (deconvolution) converts the complex plots (mass spectra) of mass/charge ratios versus abundance to molecular weight also versus abundance. When plotted in this fashion, ESI mass spectra resemble MALDI spectra and are readily interpreted. The longer the DNA sequence that can be analyzed the more information that can be obtained, and generally MALDI-MS has been most successfully used for analysis of low mass DNA (e.g., primer extension reactions). Furthermore, it is only in the past couple of years that MALDI-MS/MS has begun to become more widely available, thus analysis of PCR products has been very limited to this point. Thus, this article will focus primarily on ESI-MS and MS/MS analysis of PCR products.

PCR is the most widely used technique, as the first step, in technology to analyze structure diversity in DNA. Minor variations in sequence (e.g., single nucleotide polymorphisms or the presence of tandem repeats (variable numbers of a short sequence present sequentially in a single stretch of DNA)) are detected or the presence of more than one copy of a gene (e.g., heterozygotes (two copies differing in sequence, one from the mother and one from the father) versus homozygotes (both copies are the same)). PCR employs specific sets of primers to amplify only a gene of interest. Femtogram to picogram amounts of DNA are used to generate high nanogram or even microgram amounts of the gene of interest. Classically, in trace analysis, the analytical chemist uses liquid chromatography, or other separation techniques, to concentrate on analyte present in low concentration and to separate it from high abundance background peaks that would otherwise confound the analysis. However, using PCR such separations are often unnecessary even when performing trace analysis on complex clinical or environmental samples (e.g., detecting bacterial infection). As noted above, generally, PCR products are characterized by classical molecular biology techniques (e.g., arrays or electrophoresis). Alternatively, one of the most powerful techniques to obtain structural information is MS. This is evidenced by the proteomics revolution where the engine driving this has been MS. MS could have a similar impact in the molecular diagnostics arena.

Over the past decade, ESI-MS analysis of PCR products has become a mature technology for genetic discrimination. The more mainstream technique, real-time PCR, is also a powerful approach that provides a simple yes or no answer, by generation of a fluorescent signal, as to the presence of a

characteristic sequence. However, there may be too many false positives in certain situations (e.g., detection of a terrorist attack) to be acceptable. It is also difficult to use real-time PCR to detect minor sequence variations. The only difference between the stages in sample preparation, for real-time PCR or PCR-MS relate to an additional clean-up step for PCR-MS and PCR-MS/MS. However, a great deal of structural information is additionally obtained by mass spectrometric analysis. The primary steps are first sample collection. This can involve an air sampling pump and either an impinger (liquid collection) or an impactor (for collection onto a filter) for environmental samples. Clinical samples (e.g., body fluids) are, of course, more simply collected. The next step involves release of DNA from cells using organic solvents to lyse cells. For bacteria, the cell membrane is encased in a tough cell wall that must be broken (e.g., shaking with glass beads). The DNA is purified to remove potential contaminants that otherwise will inhibit the PCR reaction. Next, real-time PCR is performed to generate the gene of interest. In real-time PCR, the fluorescent signal is generated as the DNA is amplified, i.e., amplification and detection occur together in a standard PCR instrument. However, additionally for MS analysis of PCR products, purification of the PCR product must be performed prior to MS or MS/MS analysis. As noted below this is primarily to eliminate low mass constituents of the PCR reaction mixture that otherwise inhibit ionization in the mass spectrometer.

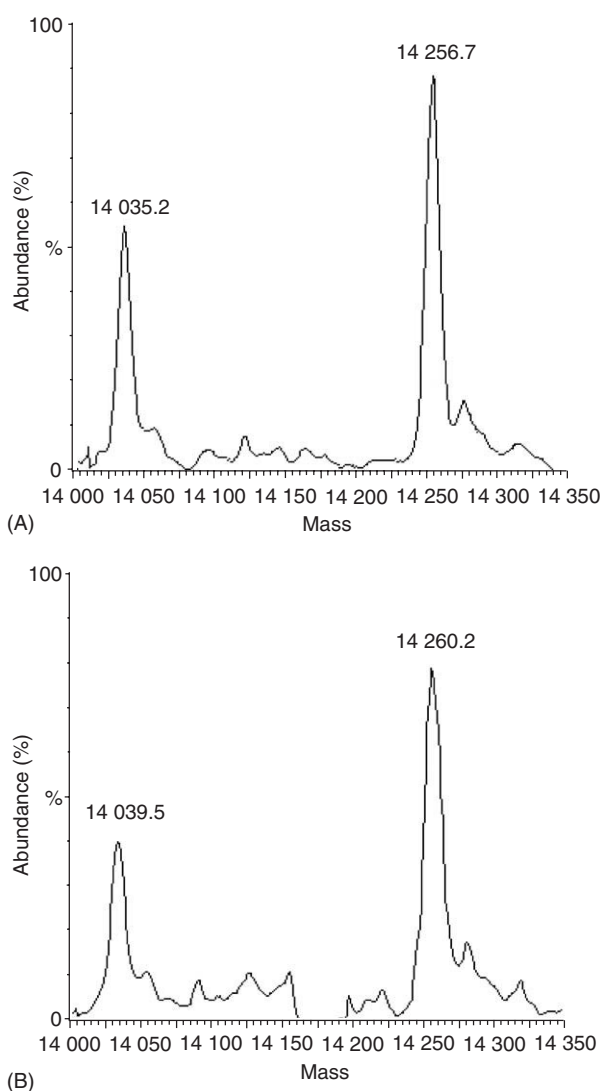
Currently, PCR and MS are performed using two separate instruments. However, there is no reason that PCR followed by simple automated clean up and MS could not be incorporated into a single integrated instrument. Essentially, every configuration of modern ESI mass spectrometer has been used successfully for analysis of PCR products, including from the highest to the lowest resolution, i.e., Fourier transform ion cyclotron resonance, magnetic sector, triple quadrupole, quadrupole-time of flight (Q-TOF), and ion trap. MS discriminates two structurally related PCR products by MW difference. Mass accuracy is needed to differentiate the pairs of coding or noncoding strands (each pair may differ by 9–40 Da, adenine to thymine (A to T) and guanine to cytosine (G to C) switch, respectively). Large PCR products in the range 223–500 nucleotides have been analyzed with the required precision. MS/MS allows for the differentiation of PCR products of closely related sequence, by fingerprint, even if they have the same nucleotide composition; fragmentation of the parent molecule occurs in the tandem mass spectrometer (analogous to a restriction digest pattern). A particular mass is selected that is collided with a collision

gas (usually argon or helium). The DNA strand breaks into fragments, at each nucleotide position, by scission of phospho-diester bridges. Thus, a series of fragments are generated, rather like a Sanger sequencing ladder, each differing from one another by a single nucleotide. Unfortunately, each piece can differ by charge and mass and there is further complication that there are four possible breakage points within a phospho-diester bridge. Thus, MS/MS spectra are much more complicated than MS spectra and currently it is difficult to visually read sequence, although this is possible with computer-assisted software. Alternatively, MS/MS spectra can be compared to detect sequence variation without the need for sequencing.

## MS Analysis of PCR Products

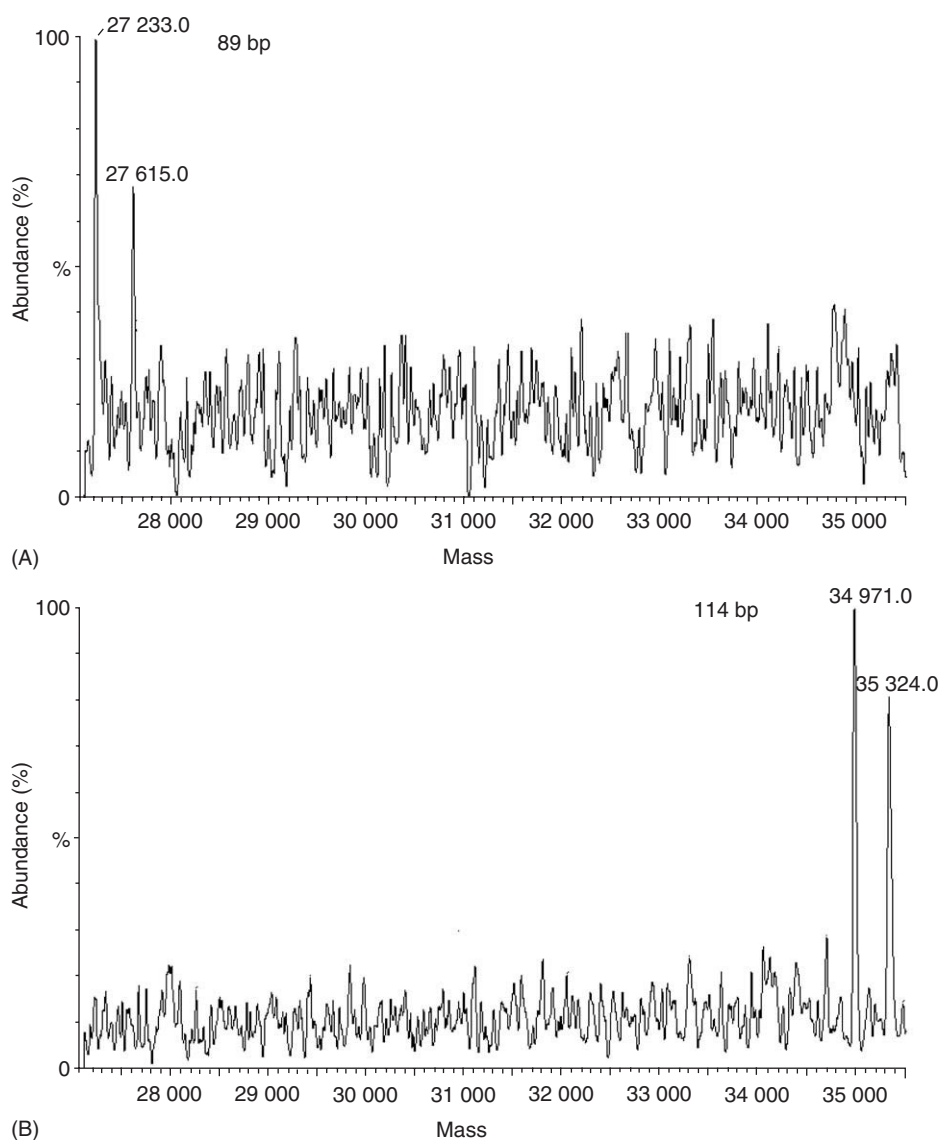
Samples must be cleaned up prior to MS analysis. Other components of the PCR reaction (e.g., metal ions and nucleotides) inhibit ionization in the mass spectrometer source or preferentially ionize. Furthermore, metal ions can bind to the phosphate backbone of DNA, thus changing the mass and complicating mass spectral interpretation. When PCR-MS was first introduced, in the mid-1990s, there was a great deal of interest in developing a simple approach for clean-up of PCR products for MS analysis. As regards ESI-MS analysis, a single approach, involving ethanol precipitation, was used reliably among a variety of laboratories and rapidly became the gold standard. Generally, clean-up is performed with ethanol precipitation in the presence of ammonium hydroxide. Ammonia replaces the metal ions and dissociates from the DNA in the mass spectrometer. Unfortunately, this is time consuming, taking a few hours or more commonly overnight for the precipitation. Although after precipitation the pellet is rapidly washed with ethanol and ethanol-water. Thus, when rapid analysis is required (e.g., for biodection) this approach is inadequate. Recently, a new approach was introduced that shows great promise. The PCR product is bound to a weak anionic exchange column (e.g., a Zip-Tip, Millipore, Bedford, MA) and washed to remove low mass materials (e.g., nucleotides and dideoxynucleotides) that do not bind to the column. The column is then washed with solutions containing high concentrations of ammonia to remove metal ions. After elution with methanol-water, the sample is ready for MS analysis. Comparable results are obtained by the two approaches (see Figure 1).

Two variations for analysis of PCR products by ESI-MS have emerged: (1) direct-injection MS and tandem mass spectrometry and (2) liquid



**Figure 1** MW of 46bp PCR product of mutant human *K-ras* codon 12 variant TGT determined using ESI-quadrupole MS after clean-up of PCR product by (A) ethanol precipitation or (B) Zip-Tip. The spectra have been deconvoluted to simplify visual discrimination. Predicted masses for the coding and noncoding strands are 14 261.3 and 14 039.1 Da, respectively.

chromatography-mass spectrometry (LC-MS) and tandem mass spectrometry (LC-MS/MS). In the former approach, sample clean-up is performed manually and, as noted above, must be simple and preferably rapid. While in the latter approach, clean-up is carried out automatically by the LC. The advantages of LC-MS and LC-MS/MS are that more complex samples can be analyzed, since there is the opportunity to separate a sample into components by LC, before entry into the mass spectrometer. Additionally, it is possible to run large batches of samples in an entirely automated fashion. However, LC analysis is much more complicated instrumentally than direct injection. The molecular biology community is not



**Figure 2** Measured MW of PCR products, 16S–23S interspace region determined using ESI-quadrupole MS. (A) *B. anthracis* (89 bp, predicted mass of coding and noncoding strands; 27 237.8 and 27 618.9, respectively) and (B) *B. subtilis* (114 bp, predicted mass of coding and noncoding strands; 34 972.8 and 35 329.0, respectively).

generally well-versed in LC instrumentation, preferring to use electrophoresis. Thus, they are more likely to employ direct MS analysis of PCR products. It is worthy of note that modern MS instruments (e.g., MALDI-TOF or ESI-ion traps) are run by Windows-based computer software. There is very little to learning how to introduce a DNA sample into either instrumental configuration. Indeed, when MS spectra are plotted as mass versus abundance, there is little data interpretation required. MS/MS spectra are quite complex, and further simplification of computer-assisted data handling is still needed. Although developments in MALDI-MS/MS may help, simpler spectra than for ESI MS/MS should be generated.

### Examples of PCR-MS and PCR-MS-MS Analyses for Genetic Analysis

Regardless of the source of a PCR product, e.g., human, plant, or microbe, sample preparation is identical and the final product is an MS or MS/MS spectrum. Certain complex samples, e.g., environmental samples, contain inhibitors of the PCR reaction (e.g., humic acids and metal ions) and developments in this area are continuing which are beyond the scope of this article. The examples below are only provided to illustrate what can be achieved with PCR-MS or MS/MS analysis.

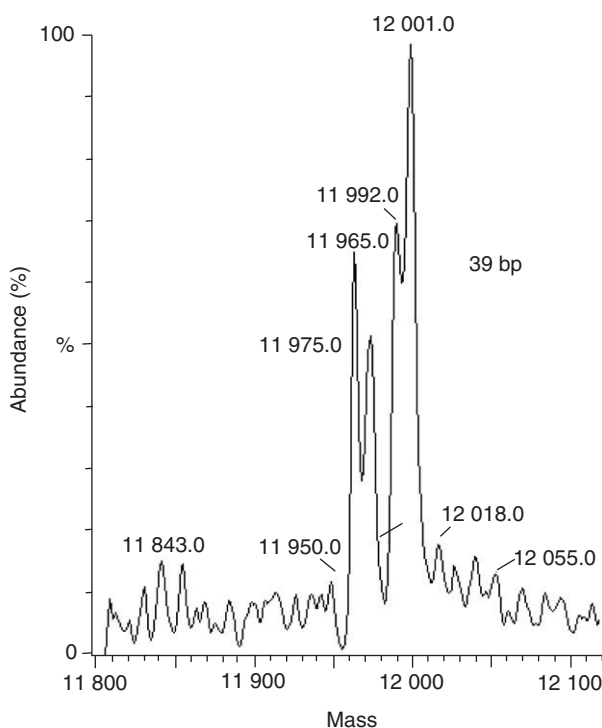


### Chemotaxonomic Differentiation of Bacteria

Due to the nature of PCR, this can be accomplished either directly from clinical or environmental samples or after culture. However, since environmental samples contain numerous bacterial species, among which bacilli are the most common, design of primers is important since otherwise complex mixtures of PCR products would be generated. Such mixtures would be extremely difficult to analyze by direct MS analysis. Furthermore, sensitivity of analysis would be compromised by spreading the signal among numerous components. Indeed, successful analysis directly from environmental samples is still a topic for current research.

As an example of bacterial discrimination, using pure bacterial cultures, the ribosomal RNA (*rrn*) operon consists of 16S and 23S rRNA genes flanking an interspace region (ISR). This operon is present in all bacteria but the sequence varies among bacteria in a species-specific fashion. For example, variation in ISR size readily distinguishes the *B. cereus* group of organisms (which includes *Bacillus anthracis*, *B. cereus*, and *B. thuringiensis*) from the closely related *B. subtilis* group (which includes *B. subtilis* and *B. atrophaeus*) (Figure 2). With appropriate primers, 89 bp products are generated for the former and 111–119 bp products for the latter. Based on sequencing of single operons, until recently, it was felt that *B. anthracis* and *B. cereus/B. thuringiensis* were identical in all regions of the *rrn* operon. However, work from the Centers for Disease Control comparing sequences of the 11 operons of the 16S rRNA gene in the *B. anthracis* genome, at position 1146, revealed five genes contain Ts whereas six contain As. However, for *B. cereus* and *B. thuringiensis* only Ts were observed at this position. It is demonstrated here that MS has the discriminating power to detect such fine genetic differences (an A to T switch) (Figure 3). Synthetic oligonucleotides, having the same sequence as the bacterial genes, were used. Thus, further work will be needed to develop a routine procedure utilizing PCR products.

Unfortunately, the ability to detect small genetic changes becomes more difficult as mass increases. Thus, there is an upper mass range where analysis is impractical. For low-resolution instruments, this limit is around a 100-mer. Thus, the mass should be minimized (or a high-resolution instrument employed). Alternatively, the smaller the piece of DNA analyzed the more it chemically resembles a primer or nucleotide monomer; thus, separation of the two on cleanup becomes more difficult. If the primers and nucleotides are not removed, they can provide a massive background on MS analysis or inhibit ion-



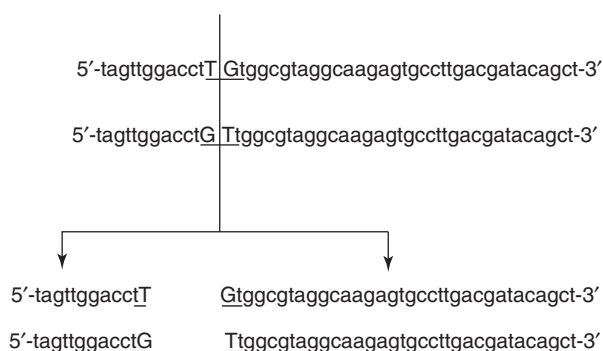
**Figure 3** Measured MW of four synthetic oligonucleotides (39-mers) determined using ESI-quadrupole MS. The sequences correspond to two regions of the *B. anthracis* genome coding for 16S rRNA and differing by a single nucleotide substitution at one position (A to T switch). Predicted mass of noncoding strands: 11,967.8 and 11,976.8; and coding strands: 11,993.8 and 12,002.8, respectively.

ization of the PCR product by preferential ionization. Thus, for practical reasons, it is extremely difficult to employ a PCR product below a 40–50-mer for direct ESI MS or ESI MS/MS analysis.

### Human Genomics and Cancer Diagnostics

Mutations in the oncogene *K-ras* can occur at positions one or two of codon 12. Thus, at each of two adjacent positions, there can be variation from G to C, A or T (six in total). It has been suggested that knowing the exact mutation could have important diagnostic or therapeutic implications for certain forms of cancer. Variations at position one are single nucleotide substitutions, and changes at position one versus complementary changes at position two constitute isomeric changes. In the latter case, the mass is the same and MS/MS is required for discrimination of genetic variants (Figure 4).

Although 50–100-mers have been sequenced, interpretation of fragmentation patterns of large pieces of DNA still remains challenging at the present time. Indeed, for complete sequence information, PCR products must currently be <80-mers in length, even



**Figure 4** Discrimination of isomeric PCR products using tandem mass spectrometry. Two mutant sequences of regions of the *K-ras* gene are shown that have the same base composition (due to sequence change at two adjacent positions within a codon) and thus the same mass on MS analysis. However, scission within this codon, on MS–MS analysis, generates two fragments for each respectively differing in sequence and mass.

with the assistance of advanced computer software. In diagnostic applications or studies of population diversity, the sequence difference between normal and cancer DNA or between genes of two individuals is determined; the rest of the sequence is identical.

Algorithms have also been developed that use automated software to compare product ion spectra (abundances of each fragment ion) of PCR products displaying single nucleotide polymorphisms or isomeric sequence changes. To find the reference sequence most closely related to the experimental spectrum, the algorithm replaces each nucleotide by the other three possible nucleotides, and determines the best fit. More simply, relative abundances of specific fragment ions, for the discrimination of closely related PCR products of different structure, are directly compared for similarities or differences. In this simple comparison of mass spectral patterns, interpretation is not necessary. All that is required is finding a single MS/MS peak that occurs in one genetic variant but not in another. Unfortunately, MS/MS spectra are extremely complex and visual discrimination is extremely time consuming and, in some cases, impossible. In order to automate this process, a new computer algorithm, SpecDiff, was recently created that automatically compares spectra searching for such differences (<http://fields.scripps.edu/SpecDiff>).

### Future Directions

There is great potential for integration of all stages of PCR–MS and PCR–MS/MS analysis including sample collection, DNA isolation, PCR, sample cleanup, MS/MS, and computer-assisted data handling. As noted above, with the exception of sample clean-up, the other stages are identical for real-time PCR and

PCR–MS. However, PCR–MS and MS/MS provide a great deal of additional structure information. This is absolutely required in detection of a specific genetic variant (for genotyping or forensic analysis) of a human or bacterial cell or detection of a particular bacterial species (e.g., *B. anthracis*) in a complex clinical or environmental matrix.

The availability of integrated commercial instrumentation could be of immeasurable benefit to biomedical science and homeland security. For human genotyping, MS and MS/MS represent a powerful approach for defining genetic variation. In forensics, the fine discriminating power might be used to define the origin of particular bacterial or viral strain. A particularly difficult analytical problem remains detection of a bioterrorist attack. The specificity of MS/MS analysis could be of major importance in solving the problems of false positives, which is of much greater concern for environmental than clinical analysis, due to the complexity of the microbial population in the latter.

**See also:** **Mass Spectrometry:** Electrospray; Matrix-Assisted Laser Desorption/Ionization; Multidimensional. **Polymerase Chain Reaction.**

### Further Reading

- Johnson YA, Nagpal M, Krahmer MT, Fox KF, and Fox A (2000) Precise molecular weight determination of PCR products of the intergenic spacer region using electrospray quadrupole mass spectrometry for differentiation of *B. subtilis* and *B. atrophaeus*, closely related species of bacilli. *Journal of Microbiological Methods* 40: 241–254.
- Krahmer MT, Walters JJ, Fox KF, *et al.* (2000) MS for identification of single nucleotide polymorphisms and MS/MS for discrimination of isomeric PCR products. *Analytical Chemistry* 72: 4033–4040.
- Laken SJ, Jackson PE, Kinzler KK, *et al.* (1998) Genotyping by mass spectrometric analysis of short DNA fragments. *Nature* 16: 1352–1356.
- McLucky SA and Habibi-Goudarzi S (1993) Decompositions of multiply charged oligonucleotide anions. *Journal of the American Chemical Society* 115: 12085–12095.
- Muddiman DC, Wunschel DS, Liu C, *et al.* (1996) Characterization of PCR products from bacilli using electrospray ionization FTICR mass spectrometry. *Analytical Chemistry* 68: 3705–3712.
- Muhammad WT, Fox F, Fox A, Cotham W, and Walla M (2002) Electrospray ionization quadrupole time-of-flight mass spectrometry and quadrupole mass spectrometry for genotyping single nucleotide substitutions in intact polymerase chain reaction products in *K-ras* and *p53*. *Rapid Communications in Mass Spectrometry* 16: 2278–2285.

- Muhammad WT, Tabb DL, Fox KE, and Fox A (2003) Automated discrimination of polymerase chain reaction products with closely related sequences by software-based detection of characteristic peaks in product ion spectra. *Rapid Communications in Mass Spectrometry* 16: 1–8.
- Muddiman DC, Null AP, and Hannis JC (1999) Precise mass measurement of a double stranded 500 base-pair (309 kDa) polymerase chain reaction product by negative ion electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Communications in Mass Spectrometry* 13: 1201–1204.
- Naito Y, Ishikawa K, Koga Y, *et al.* (1995) Molecular mass measurement of polymerase chain reaction products amplified from human blood DNA by electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* 9: 1484–1486.
- Oefner PJ and Huber CG (2002) A decade of high-resolution liquid chromatography of nucleic acids on styrene–divinylbenzene co-polymers. *Journal of Chromatography B* 782: 27–55.
- Walters JJ, Fox KE, and Fox A (2002) Mass spectrometry and tandem mass spectrometry, alone or after liquid chromatography, for analysis of polymerase chain reaction products in the detection of genomic variants. *Journal of Chromatography B* 782: 57–66.
- Yiang Y and Hofstadler SA (2003) A highly efficient and automated method of purifying and desalting PCR products by electrospray ionization mass spectrometry. *Analytical Biochemistry* 316: 50–57.

## MEAT AND MEAT PRODUCTS

See **FOOD AND NUTRITIONAL ANALYSIS: Meat and Meat Products**

## MECA

See **MOLECULAR EMISSION CAVITY ANALYSIS**

## MEMBRANE TECHNIQUES

Contents

**Dialysis and Reverse Osmosis**

**Ultrafiltration**

**Liquid Membranes**

**Pervaporation**

### Dialysis and Reverse Osmosis

**U Spohn**, Martin-Luther Universitat Halle-Wittenberg, Halle, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Dialysis is a separation process with many applications in clinical, biochemical, and environmental

analysis. Donor and acceptor solutions are separated by a membrane, which is permeable to one or more solutes. When the solutions have different solute activities, a diffusional transport from the higher to the lower activity  $a$  ( $a = cf$ , where  $c$  is the concentration and  $f$  the activity coefficient) through the membrane takes place (**Figure 1**). The main driving force for the diffusion of small molecules or ions is their concentration gradient. In most cases, dialysis is employed to retain large molecules while exchanging or eliminating small ones.

Different solvent activities generate an osmotic pressure, which causes a solvent flow to the solution

- Muhammad WT, Tabb DL, Fox KE, and Fox A (2003) Automated discrimination of polymerase chain reaction products with closely related sequences by software-based detection of characteristic peaks in product ion spectra. *Rapid Communications in Mass Spectrometry* 16: 1–8.
- Muddiman DC, Null AP, and Hannis JC (1999) Precise mass measurement of a double stranded 500 base-pair (309 kDa) polymerase chain reaction product by negative ion electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Communications in Mass Spectrometry* 13: 1201–1204.
- Naito Y, Ishikawa K, Koga Y, *et al.* (1995) Molecular mass measurement of polymerase chain reaction products amplified from human blood DNA by electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* 9: 1484–1486.
- Oefner PJ and Huber CG (2002) A decade of high-resolution liquid chromatography of nucleic acids on styrene–divinylbenzene co-polymers. *Journal of Chromatography B* 782: 27–55.
- Walters JJ, Fox KE, and Fox A (2002) Mass spectrometry and tandem mass spectrometry, alone or after liquid chromatography, for analysis of polymerase chain reaction products in the detection of genomic variants. *Journal of Chromatography B* 782: 57–66.
- Yiang Y and Hofstadler SA (2003) A highly efficient and automated method of purifying and desalting PCR products by electrospray ionization mass spectrometry. *Analytical Biochemistry* 316: 50–57.

## MEAT AND MEAT PRODUCTS

See **FOOD AND NUTRITIONAL ANALYSIS: Meat and Meat Products**

## MECA

See **MOLECULAR EMISSION CAVITY ANALYSIS**

## MEMBRANE TECHNIQUES

Contents

**Dialysis and Reverse Osmosis**

**Ultrafiltration**

**Liquid Membranes**

**Pervaporation**

### Dialysis and Reverse Osmosis

**U Spohn**, Martin-Luther Universitat Halle-Wittenberg, Halle, Germany

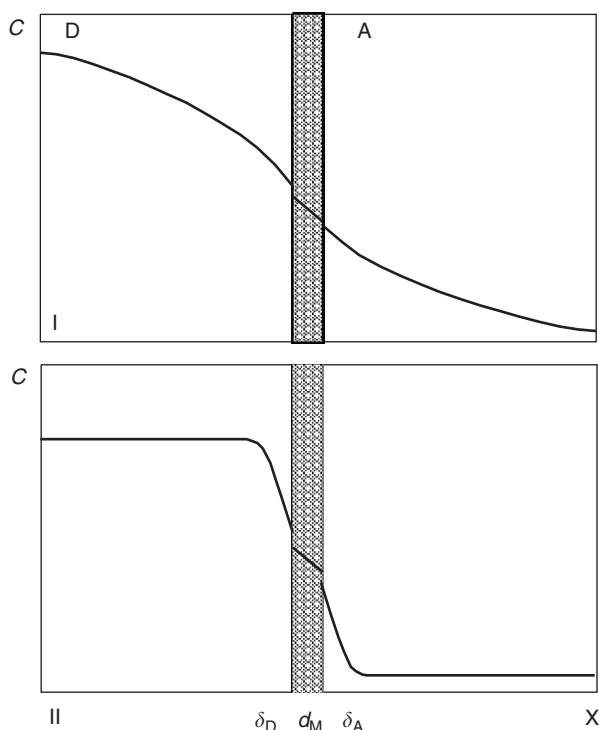
© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Dialysis is a separation process with many applications in clinical, biochemical, and environmental

analysis. Donor and acceptor solutions are separated by a membrane, which is permeable to one or more solutes. When the solutions have different solute activities, a diffusional transport from the higher to the lower activity  $a$  ( $a = cf$ , where  $c$  is the concentration and  $f$  the activity coefficient) through the membrane takes place (**Figure 1**). The main driving force for the diffusion of small molecules or ions is their concentration gradient. In most cases, dialysis is employed to retain large molecules while exchanging or eliminating small ones.

Different solvent activities generate an osmotic pressure, which causes a solvent flow to the solution



**Figure 1** Dialysis between quiescent (I) and stirred or flowing solutions (II) solutions. C, concentration; D, donor solution; A, acceptor solution; M, membrane;  $d_M$ , thickness of the membrane;  $\delta_A$  and  $\delta_D$ , thicknesses of the boundary layers.

of the lower solvent activity. This process is termed osmosis. Reverse osmosis is the process during which an outer pressure is applied to force the solvent through a membrane, which rejects the solute. Reverse osmosis is used almost exclusively to desalt and purify water. In the ideal case, the separation membrane is only permeable to water.

## Thermodynamics

The difference between the chemical activities of the solute  $i$  is the precaution for its diffusional transport through the separation membrane. The free enthalpy  $\Delta G$  of diffusion can be calculated from the difference between the chemical potentials  $\mu_{i,A}$  and  $\mu_{i,D}$  of the solute  $i$  in the acceptor (A) and donor solution (D), respectively (eqn [1]).  $\mu_{i,D}^0$  and  $\mu_{i,A}^0$  are the standard chemical potentials,

$$\Delta G = \mu_{i,D} - \mu_{i,A} = \mu_{i,D}^0 - \mu_{i,A}^0 + RT \ln(a_{i,D}/a_{i,A}) \quad [1]$$

The equilibrium constant  $K_i^*$  depends on the activity coefficients  $f_{i,A}$  and  $f_{i,D}$ , the concentrations  $c_{i,A}$  and  $c_{i,D}$ , and the absolute temperature  $T$  according to eqn [2].  $R$  is the molar gas constant.  $K_i^*$  is given as

$$K_i^* = \frac{a_{i,A}}{a_{i,D}} = \frac{f_{i,A} \times c_{i,A}}{f_{i,D} \times c_{i,D}} = \exp \left[ \frac{\mu_{i,D}^0 - \mu_{i,A}^0}{RT} \right] \quad [2]$$

The effective separation ratio  $K_i$  is calculated according to eqn [3]

$$K_i = \frac{c_{i,A}}{c_{i,D}} \quad [3]$$

At  $K_i^* = 1$  a stationary state, called the thermodynamic equilibrium is reached and no directed diffusion can take place.  $K_i$  differs from unity when the activity coefficients of the solute  $i$  are different, e.g., if the donor and acceptor solutions have different compositions and/or the dialysis equilibrium is coupled with other equilibria. Without coupled reactions, the enrichment  $E$  and the purification  $P$  can be calculated according to eqns [4] and [5],

$$E = \frac{c_{i,A}}{c_{i,D_0}} = \frac{(c_{i,D_0} - c_{i,D})V_D}{c_{i,D_0} \times V_A} \quad [4]$$

$$P = \frac{c_{i,D_0} - c_{i,D}}{c_{i,D_0}} = \frac{(c_{i,A} - c_{i,A_0})V_A}{c_{i,D_0} \times V_D} \quad [5]$$

where  $c_{i,D_0}$  and  $c_{i,A_0}$  are the initial solute concentrations in the donor and in the acceptor solution, respectively.  $V_A$  and  $V_D$  are the volumes of the corresponding solutions. Without trapping reaction in the acceptor solution, during which the received analyte is converted into a substance that cannot permeate the membrane, the enrichment  $E$  approaches 0.5 for  $V_A = V_D$ . If the activity coefficients are equal (i.e.,  $f_{i,A} = f_{i,D}$ ) in the equilibrium state,  $E$  will be 0.5. Also the chemical activities of the solvent  $S$  are equal at dialysis equilibrium. Because

$$\mu_{S,D} = \mu_{S,A} \quad [6]$$

and

$$\mu_{S,D}^0 = \mu_{S,A}^0 \quad [7]$$

The thermodynamic equilibrium constant  $K_S^*$  is given as

$$K_S^* = \exp[(V_{mS,D}\pi_D - V_{mS,A}\pi_A)/RT] = \frac{a_{x,S,A}}{a_{x,S,D}} \quad [8]$$

where  $V_{mS,D}$  and  $V_{mS,A}$  are the partial molar volumes of the solvent, which are equal to the  $V_{mS}$  for dilute solutions.  $\pi_D$  and  $\pi_A$  refer to the corresponding osmotic pressures.  $a_{x,S,A}$  and  $a_{x,S,D}$  are the activities of the solvent with respect to its mole fraction  $x_S$  in the acceptor and donor solutions, respectively. It follows that

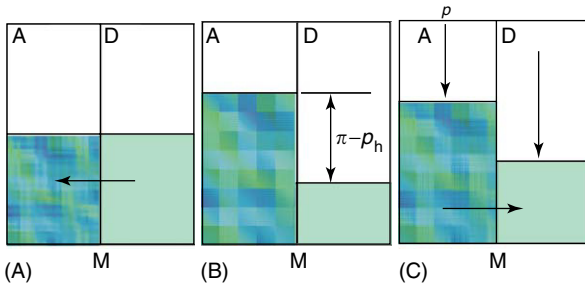
$$K_S^* = \exp[(\pi_D - \pi_A)V_{mS}/RT] \quad [9]$$

The resulting pressure difference  $\pi = \pi_D - \pi_A$  can be calculated according to eqn [10]:

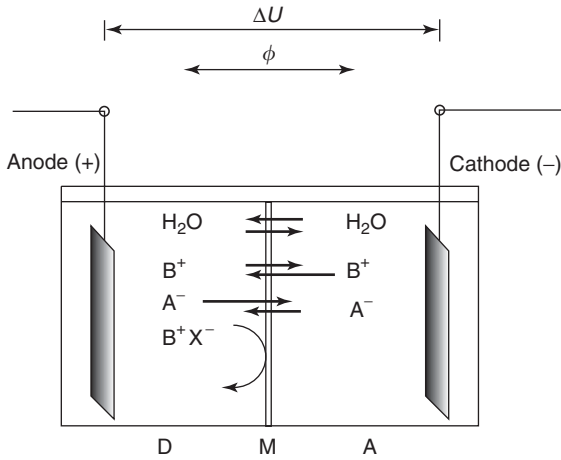
$$\pi = -(RT/V_{mS}) \ln a_{x,S,A} \quad \text{for } a_{x,S,D} = 1 \quad [10]$$

This pressure difference causes a solvent flow through the separation membrane (Figure 2A) up





**Figure 2** (A) Osmosis; (B) osmotic equilibrium; (C) reverse osmosis. M, semipermeable membrane;  $p_h$ , hydrostatic pressure;  $\pi$ , osmotic pressure;  $p$ , external pressure, D, aqueous salt solution, A, water.



**Figure 3** Donnan dialysis. M, semipermeable membrane; D, donor solution; A, acceptor solution,  $\phi$ , Donnan potential;  $U$  applied voltage.

to the state at which the hydrostatic or another back-pressure compensates for the osmotic pressure, e.g., in batch-type arrangement shown in **Figure 2B**. If the outer pressure  $p$  is higher than the osmotic pressure difference  $\pi$ , the osmosis will be reversed (**Figure 2C**). An equilibrium described by eqn [11] will be reached:

$$K_S = \exp \left[ \frac{V_{mS}}{RT} (\pi - p_h) \right] \quad [11]$$

The rejected solute is enriched in the donor solution. To take into account the electric charge  $z_i$  of the solutes and of the separation membrane, the Donnan effect has to be considered. The chemical potential  $\mu_i$  is extended by the electric potential gradient,  $\Delta\phi$ . When a membrane separates two solutions of a dissociating salt  $BA$  ( $BA \rightleftharpoons B^+ + A^-$ ), one of which contains a rejected ion  $X^-$ , an electric potential difference  $\Delta\phi = \phi_D - \phi_A$  is built up between the solutions (**Figure 3**). The thermodynamic equilibrium constant can be calculated in the absence of osmotic pressure differences according to eqn [12] with the

Faraday constant  $F$ :

$$K_i^* = \exp(z_i F \Delta\phi / RT) \quad [12]$$

An additionally applied electric voltage  $U$  generates an electromigration of the anions to the positively charged anode and of the cations in the opposite direction. This process is termed electrodialysis.

## Mass Transport

### Dialysis

For quiescent donor and acceptor solutions, the equilibration time ranges from minutes to several hours and is dependent on the thickness and other geometrical parameters of the corresponding chambers, the membrane permeability, and the temperature. The equilibration is dominated by diffusion. The equilibration can be accelerated by convective mass transport according to **Figure 1** by achieving thin diffusional boundary layers on the membrane.

The overall flux of the substance  $i$  with its molar amount  $n_i$  from the donor to the acceptor solution can be described according to eqn [13]:

$$J_i = \frac{dn_i}{dt} = kA \left( \beta_{i,D} c_{i,D} - \beta_{i,A} c_{i,A} \frac{K_{AM}}{K_{DM}} \right) \quad [13]$$

where  $\beta_{i,D}$  and  $\beta_{i,A}$  are the fractions of the substance  $i$  in the donor and in the acceptor solutions, respectively, that can permeate through the separation membrane.  $K_{DM}$  and  $K_{AM}$  are the distribution constants between the donor solution and the membrane and between the acceptor solution and the membrane, respectively. In the classical case of dialysis of low molecular substances through a membrane with symmetric and open pores eqn [14] holds

$$J_i = \frac{dn_i}{dt} = kA (c_{i,D} - c_{i,A}) \quad [14]$$

The general overall mass transfer coefficient  $k$  can be derived from eqn [15]

$$\frac{1}{k} = \frac{1}{k_D} + \frac{1}{k_M K_{DM}} + \frac{K_{AM}}{k_A K_{DM}} \quad [15]$$

where  $k_D$ ,  $k_M$ , and  $k_A$  are the mass transfer coefficients of the permeating fractions of substance  $i$  for the donor, the membrane, and the acceptor phases, respectively. In many cases, the mass transfer across the phase boundary also needs to be considered, e.g., for gas-filled microporous membranes. Equation [16] has an additional term on the right-hand side:

$$a = \frac{1}{k_{pDM}} + \frac{1}{k_{pAM}} \frac{K_{AM}}{K_{DM}} \quad [16]$$

where  $k_{pDM}$  and  $k_{pAM}$  are the phase transfer coefficients from the donor solution into the membrane phase and from the membrane into the acceptor phase, respectively.

From the general eqns [15] and [16], some cases of practical interest can be deduced.

**Dialysis through hydrophilic and microporous membranes** As  $K_{DM} = K_{AM} = 1$  and no phase transfer occurs, eqn [15] simplifies to eqn [17] as follows:

$$\frac{1}{k} = \frac{1}{k_D} + \frac{1}{k_M} + \frac{1}{k_A} \quad [17]$$

For large concentration gradients and very intensive stirring or in flow-through dialysis cells at high enough flow rates, the membrane diffusional transport becomes rate determining particularly for relatively thick membranes with small pores (eqn [18])

$$k = k_m \quad [18]$$

**Dialysis of volatile and volatilizing substances through hydrophobic and microporous membranes** To separate a nonvolatile base  $AH^+$ , its corresponding volatile base  $A$ , e.g.,  $NH_3$  is produced according to the following equilibrium:



The donor pH value is adjusted according to eqn [20] so that there is 99.9% degree of conversion into the volatile and the permeable base:

$$\beta_{i,D} = \frac{K_a}{10^{-pH} + K_a} > 0.999 \quad [20]$$

where  $K_a$  is the acid-dissociation constant of  $AH^+$ . The acceptor pH value should be adjusted according to eqn [21] to trap the analyte in its nonvolatile form:

$$1 - \beta_{i,A} = \frac{10^{-pH}}{10^{-pH} + K_a} > 0.999 \quad [21]$$

Analogous equations can be derived for volatile acids, e.g.,  $HCN$ . Since only the volatile part of the analyte amount permeates the membrane, it follows that

$$J_i = kA c_{i,D} \quad [22]$$

with

$$\frac{1}{k} = \frac{1}{k_D} + \frac{1}{K_M K_{DM}} + \frac{K_{AM}}{K_A K_{DM}} + a \quad [23]$$

The distribution ratios  $K_{DM}$  and  $K_{AM}$  are inversely proportional to the concentrations  $\beta_{i,D} \times c_{i,D}$  and  $\beta_{i,A} \times c_{i,A}$  of the volatile forms in the donor and in

the acceptor solutions, respectively:

$$K_{DM} = \frac{c_{i,g,D}}{\beta_{i,D} \times c_{i,D}} \cong \frac{p_{i,D}}{\beta_{i,D} RT c_{i,D}} \quad [24]$$

$$K_{AM} = \frac{c_{i,g,A}}{\beta_{i,A} \times c_{i,A}} \cong \frac{p_{i,A}}{\beta_{i,A} RT c_{i,A}} \quad [25]$$

For low partial pressures  $p_{i,D}$  and  $p_{i,A}$ ,  $K_{DM}$  and  $K_{AM}$  approximate the right-hand terms of eqns [24] and [25].  $c_{i,g,D}$  and  $c_{i,g,A}$  are the gas-phase concentrations at the interface between the donor or the acceptor solution and the membrane gas phase, respectively. Because the nonvolatile forms of the analyte cannot permeate and  $K_{AM}$  approximates zero at pH precalculated by eqn [21]. Equation [15] simplifies to eqn [26] as follows:

$$\frac{1}{k} = \frac{1}{k_D} + \frac{1}{k_M K_{DM}} + a \quad [26]$$

For fast-flowing donor solutions with  $1/k_D \rightarrow 0$  the mass transport is determined by the gas diffusion through the membrane, the partial pressure of the analyte in the donor solution, and the phase-transfer resistances.

### Reverse Osmosis

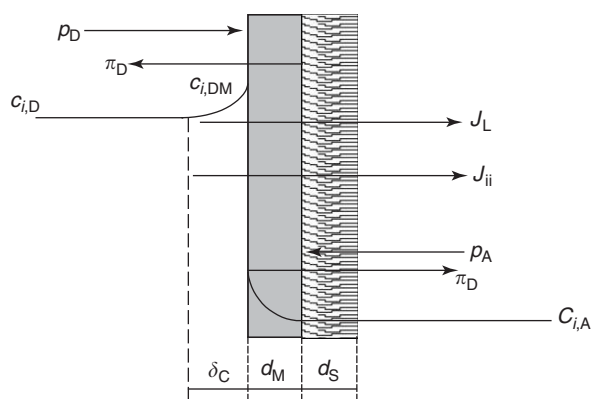
The driving force of reverse osmosis is the difference between the outer pressure  $p_h$  and the osmotic pressure difference  $\pi$ . The mass transfer can be described according to the following equation:

$$J_s = n_s/t = PA(p_h - f_R \pi) \quad [27]$$

where  $J_s$ ,  $P$ , and  $A$  are the mass flux of the solvent through the separation membrane, the water permeability of the membrane, and the membrane area, respectively. The mass flux is influenced by the reflection factor  $f_R$ , which is a measure of the solute rejection by the membrane. The osmotic pressure difference increases continuously during the enrichment of the solutes in the donor solution up to  $p_h = f_R \pi$ . The rejection ratio  $R$  is defined by eqn [28]:

$$R = 1 - \frac{c_{i,A}}{c_{i,D,o}} \quad [28]$$

with  $c_{i,A}$  as the solute concentration in the filtrate, and  $c_{i,D,o}$  as the initial concentration in the donor solution. The rejected solutes accumulate on the membrane surface (Figure 4). This phenomenon is called concentration polarization. On achieving the saturation level, the solute will start to precipitate forming a secondary layer on the membrane, which drastically reduces the mass flux through it. The analytical usefulness of the reverse osmosis is based on the high enrichment factor  $E$ , which can be



**Figure 4** Reverse osmosis with concentration polarization on the asymmetric separation membrane.  $J_i$ , mass flux of the solute  $i$ ;  $p_D$ , outer pressure from the donor side;  $p_A$ , outer pressure from the acceptor side;  $\pi_A$  and  $\pi_D$ , osmotic pressure in the acceptor and donor chamber, respectively;  $d_M$  and  $d_S$ , thickness of the membrane layer and the membrane supporting layer, respectively;  $\delta_C$ , thickness of the polarization layer;  $c_{i,DM}$ ,  $c_{i,A}$ ,  $c_{i,D}$ , concentration of the solute  $i$  in the acceptor solution, in the donor solution, and at the separation membrane, respectively.

calculated according to eqn [29]:

$$E = \frac{c_{i,D}}{c_{i,D0}} = \frac{V_{D,o}}{V_{D,o} - V_A} \quad [29]$$

where  $V_{D,o}$  is the initial volume of the donor solution.

### Geometric Aspects

The geometric shape and volume both of the donor and the acceptor chambers as well as of the separation membrane have an essential influence on the duration of the separation. To minimize the separation time, the thickness of the donor solution layer should be as thin as possible. The ratio of the membrane exchange area to the donor solution volume should be maximized. To maximize the enrichment factor for dialysis with enhanced selectivity, e.g., gas dialysis, the volume ratio between the donor solution and the acceptor solution has to be maximized. The miniaturization of the membrane exchange area up to the micro- or the ultramicroscale enables reproducible sampling from quiescent or slowly flowing solutions to be performed. This is an essential aspect for *in vivo* sampling with microdialytic probes. Figure 5 shows frequently used hollow-fiber and flat-membrane setups.

### The Separation Membrane

The dialytic transport of solute  $i$  through thin membranes can be described by eqn [30]:

$$J_i = \frac{dn_i}{dt} = k_{MA}(c_{i,MD} - c_{i,MA}) \quad [30]$$

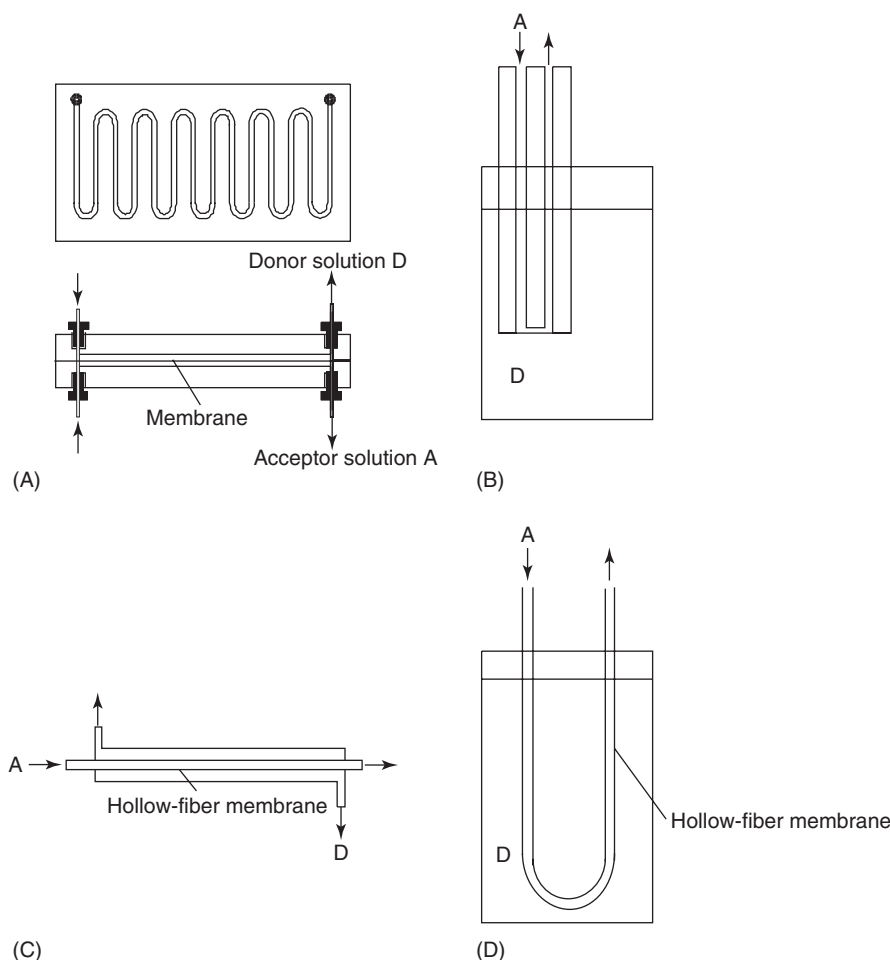
where  $c_{i,MD}$  and  $c_{i,MA}$  are the solute concentrations in the membrane at the interfaces between donor and acceptor solutions, respectively. The separation membrane is chosen with respect to its selectivity and mass transfer kinetics. The membrane material determines the transport mechanism. Table 1 gives an overview about membrane materials, their modifications, and the dominant transport mechanism. In most cases the selectivity is based on the sieve effect of the membrane pores. The dialysis across gas-filled membranes, ion-exchange membranes, or supported liquid membranes (SLMs), the pores of which are filled with organic solvents or solvent mixtures with and without additives, enables the enrichment of the analyte to be detected. The selectivity of SLM techniques is enhanced by the addition of selectively reacting ligands to the liquid membrane phase. When charged ions are complexed and transported through such membranes electroneutrality must be maintained. In some cases ion pairs with selected counterions are transported through the liquid membrane phase. When a binding ligand is dissolved in this phase and the counterion cannot transverse the membrane, the analyte ion transport is coupled with back-diffusion of an ion from the acceptor solution. A similar situation occurs in ion-exchange membranes, which are used to enrich ions by Donnan dialysis.

Using microporous membranes with a hydrophobic inner surface, the selectivity of gas dialysis against other volatiles results predominately from the ratio of the partial pressure of the analyte  $p_i$  to the total pressure. In homogeneous membranes, the different solubilities of the gases in the membrane material are the main selecting factor.

## Applications

### Dialysis through Microporous Membranes

In analytical chemistry this kind of dialysis is applied for purifying, diluting, and conditioning the sample solutions. The dialysis module can be placed prior to or in the sample insertion unit, between the sample insertion unit and the reaction/separation zone and also into the detector zone. Figure 5A shows a typical flow dialysis cell, which can be inserted in many on-line configurations with liquid chromatography and flow injection analysis (FIA), as shown in Figure 6. Two meander channels (depth:  $d = 0.1\text{--}0.5\text{ mm}$ , width:  $b = 1\text{--}2\text{ mm}$ ) are separated by a membrane. The main advantages are: easy exchange of the membrane and the use of very different membrane materials. In a second version of dialysis cell (Figure 5B) a hollow-fiber membrane is concentrically placed



**Figure 5** Frequently applied dialysis setups. (A) Thin layer flow cell, (B) immersion membrane probe, (C) hollow-fiber membrane module, and (D) immersing hollow-fiber membrane probe.

**Table 1** Membrane transport and selectivity

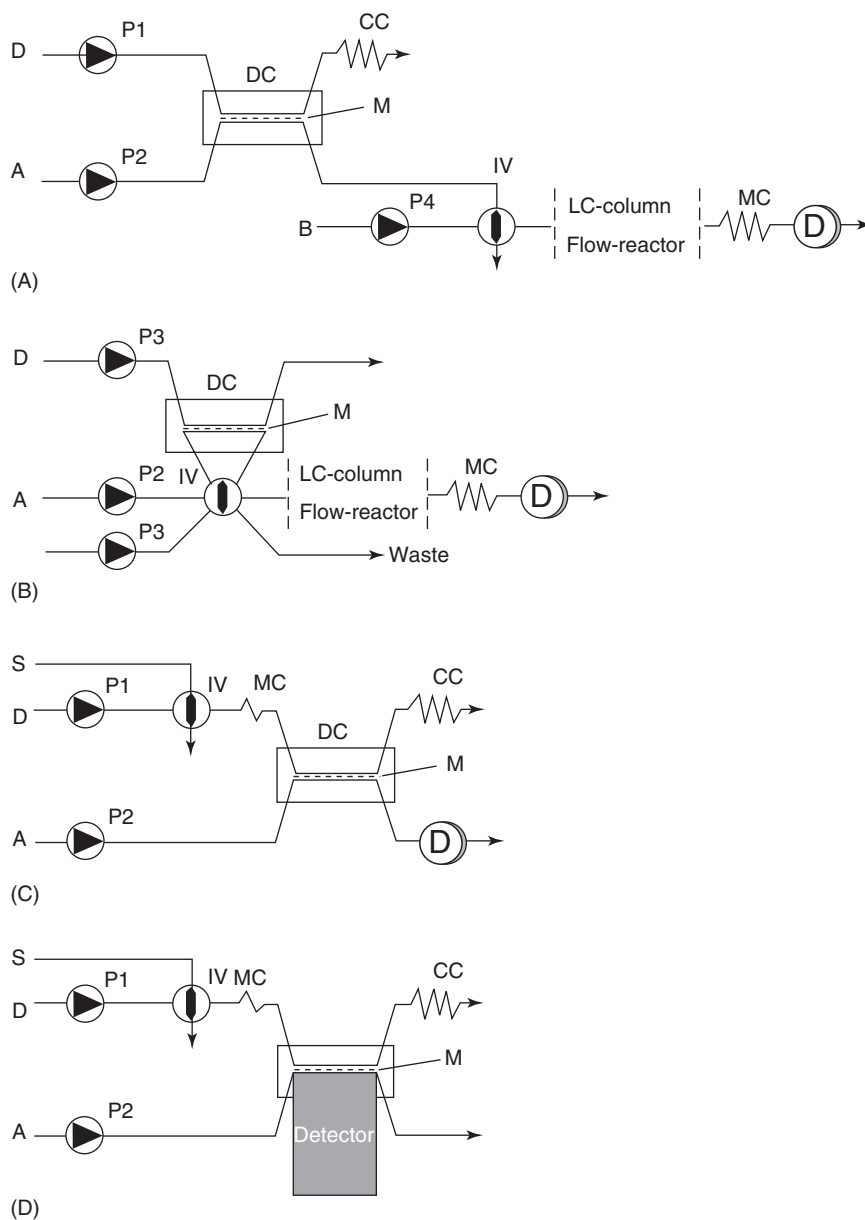
Membrane material	Transport mechanism	Factors determining selectivity
Hydrophilic and electrically uncharged	Diffusion through micropores ( $k_M = D_m \epsilon / \xi d_M$ )	Sieve effect
Ion-exchange membranes	Retarded diffusion through micropores	Ion exchange, Donnan exclusion
Hydrophobic and microporous	Gas diffusion and flow	Volatility
As before but filled with organic solvent phase with and without reactants	Diffusion through Micropores	Solubility and chemical reaction
Homogeneous membranes	Solvation and diffusion	Solubility in the membrane

$D_m$ , molecular diffusion coefficient;  $\epsilon$ , membrane porosity;  $\xi$ , tortuosity of the pores;  $d_M$ , thickness of the membrane.

inside a tube forming a thin-layer acceptor chamber. Higher mass transfer efficiencies are achieved with respect to the membrane area. There is a growing interest in dialysis immersion probes working either with a small membrane window or a hollow-fiber membrane (Figures 5C and 5D). Immersion dialysis probes with integrated sensors, e.g., biosensors open up a way to adapt the sample solution to the sensor with respect to the determination range and the

conditions of detection. The sensor can be calibrated *in situ* by pumping standard solutions through the immersion probe.

Figure 6 shows typical online dialysis configurations. A continuously working dialysis cell can be coupled directly to the valve IV (Figure 6A) injecting the prepurified and conditioned sample solution into a nonsegmented carrier flow stream, e.g., of liquid column chromatography. The analytes are separated



**Figure 6** Continuous and pulsed dialysis in flow analytical setups. P1–P4, pumps; IV, injection valve; DC, thin layer dialysis cell; M, separation membrane; S, sample solution; B, carrier solution; CC, compensation coil; MC, mixing coil.

from high molecular weight substances, e.g., proteins or from precipitates and microparticles. Blocking of the separation column is prevented and its life time is prolonged. To improve the sensitivity, the transferred amount of the analyte can be increased by the loop-dialysis setup shown in **Figure 6B**. The analyte is accumulated in the acceptor loop followed by injection into the detector channel D. To reduce the contact time of sample solution with the separation membrane, the setup shown in **Figure 6C** is used. The probability of membrane fouling and clogging is decreased considerably. A dialysis cell (**Figure 6D**) with

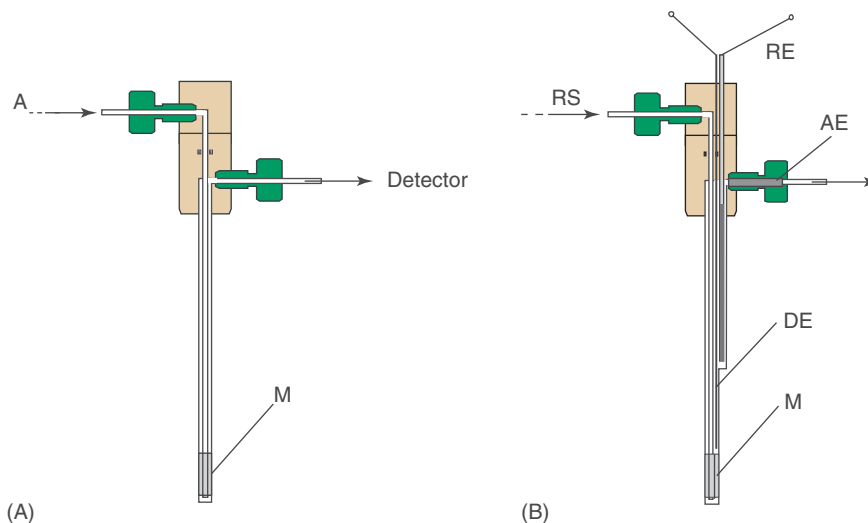
integrated detector can be used to increase the sensitivity and to shorten the analysis time.

A very promising and even more expanding field of application was opened up more than 15 years ago by the microdialysis technique, which is the most widely used dialysis technique in bioanalytical and clinical chemistry. The main advantages are: *in situ* sampling, *in situ* calibration, and avoiding the influence of complex samples matrices on to the indicator reaction. **Table 2** summarizes the application areas. Miniaturized dialysis probes with tip diameters smaller than 1 mm are implanted into different



**Table 2** Application of microdialysis

	Analyte
Neuroscience	Glucose, L-lactate, L-glutamate and other amino acids, ascorbic acid, acetylcholine, choline, NO
Clinical chemistry and Pharmacokinetics	Glucose, L-lactate, drugs and their metabolites
Bioprocess analysis	Glucose, L-lactate, D-lactate, ethanol, enzyme activities, peptides, proteins
Environmental analysis	Aniline, 2-chloroaniline

**Figure 7** Microdialysis probes: (A) without detector; (B) with internal amperometric detector. DE, detector electrode; AE, auxiliary electrode; RE, reference electrode; A, acceptor solution; RS, reagent solution.

tissues of living animals for sampling low molecular weight substances from the interstitial fluid. The analyte is separated across small membrane windows, e.g., of microporous cellulose acetate and polysulfone membranes with cutoffs between 1000 and 60 000 Da into a slowly flowing acceptor stream. The microdialysis probes are coupled online to micro-column LC systems, to capillary FIA setups and to capillary electrophoresis devices. **Figures 7A** and **7B** show a simple microdialysis probe and a probe with integrated amperometric detection, respectively. In the latter version, an enzyme solution can be used to catalyze a reaction generating a substance, e.g.,  $\text{H}_2\text{O}_2$  or an oxidized redox mediator, which can be detected selectively and sensitively. Much shorter response times are achieved in comparison to the on-line coupling to external analytical setups. According to Hamberger *et al.* (1991) the solute concentration  $c_{\text{out}}$  at the outlet of microdialysis probes can be calculated with eqn [31] with the sample concentration  $c$ , the membrane area  $A$ , the volumetric flow rate  $v$  and the overall mass transfer coefficient  $k$ , which depends on the mass transport resistances in the sample tissue, the membrane, and the dialysate.

$$c_{\text{out}} = c(1 - \exp[-kA/v]) \quad [31]$$

### Selective Dialysis across Solid and Liquid Membranes

The analyte transport through such membranes is based on a solvation/diffusion mechanism in a lipophilic phase. The liquid membranes have to be supported by a microporous and hydrophobic layer. The flow separation cells shown in **Figure 5** can be used to apply these membranes for separation procedures that can be coupled online to flow analytical, liquid, and gas chromatographic setups.

In comparison to homogeneous membranes, e.g., of silicon rubber, the SLMs have the advantages of faster membrane diffusion, and easier and wider modification of the liquid phase, which determines the transport mechanism and the selectivity. Further advantages are the low amount of organic solvent and of extractants, low operating costs, easy automation, and high enrichment factors. Dialysis across SLMs has a wide and continuously growing field of application in environmental analysis. **Table 3** summarizes some applications. Three modes of separation are used:

1. An electrical neutral substance diffuses from the donor solution through the liquid membrane phase into the acceptor solution.

**Table 3** Applications of supported liquid membranes

Analytes	Conversion with	Membrane	Trapping by	Detection
<i>Mode A</i>				
Alcohols		n-Heptane, PTFE		GC
<i>Mode B</i>				
Amines	OH <sup>-</sup>	n-Decane, supported by microporous PTFE	H <sup>+</sup>	GC
Phenols	H <sup>+</sup>	n-Dodecane, PTFE	OH <sup>-</sup>	LC, GC
Carboxylates	H <sup>+</sup>	n-Nonane, PVDF	OH <sup>-</sup>	LC
Thiolates	H <sup>+</sup>	n-Dodecane, PVDF	OH <sup>-</sup>	phot
Atrazine, Simazine	pH ≤ 7	Dihexyl ether, PTFE	H <sub>2</sub> SO <sub>4</sub>	MEKC
Caffeine <sup>c</sup>	Neutral	Dihexyl ether/n-undecane in PTFE	H <sub>2</sub> SO <sub>4</sub>	phot
Nicotine <sup>c</sup>	Neutral	n-Undecane in polypropylene	H <sub>2</sub> SO <sub>4</sub>	phot
<i>Mode C</i>				
Al(III), Cd(II), Cu(II)	pH 4.0	Di(2-ethylhexyl)phosphate in kerosene, PTFE	pH = 0.3	AAS
Nd(III)	Weakly acidic	As in former lines	0.1 mol l <sup>-1</sup> HNO <sub>3</sub>	phot
Cu(II) <sup>a</sup>	PAR	n-Pentane containing di-2-ethylhexylphosphoric acid, PVDF	H <sup>+</sup>	
Pb(II) <sup>b</sup>	Anions	Phenylhexane, containing bis(1-hydroxylheptylcyclohexano)-18-crown-6, PTFE	EDTA	AAS
Cr(VII)	pH 3.0	Aliquat 336 in tri-n-butyl-phosphate/kerosene, PTFE	NaNO <sub>3</sub>	AdSV

<sup>a</sup> Jönsson JA and Mathiasson L (1999) *Trends in Analytical Chemistry* 18: 318–334 and references cited therein; Barnes DE and van Staden JF (1992) *Analytica Chimica Acta* 261: 441–451.

<sup>b</sup> Izatt RM, Bruening RL, Bruening ML et al. (1989) *Analytical Chemistry* 61: 1140–1148.

<sup>c</sup> Luque-Perez E, Rios A, Valcarcel M, Danielsson L-G, and Ingman F (1999) *Laboratory Automation* 34: 131–142; Idem (1999) *Analytica Chimica Acta* 387: 155–164.

AAS, atomic absorption spectrometry; AdSV, adsorptive stripping voltammetry; Amp, amperometric; GC, gas chromatography; LC, liquid chromatography; MEKC, micellar electrokinetic chromatography; PAR, 4-(2'-pyridylazo)-resorcinol; phot, photometric; PTFE, polytetrafluoroethylene; PVDF, polyvinylidenedifluoride.

- The analyte is converted into a membrane soluble substance by pH shift or chemical reaction, which diffuses through the membrane and is trapped as a substance that is insoluble in the membrane.
- Co-ion-mediated transport on the basis of a carrier substance, which is dissolved in the liquid-membrane phase. The carrier molecule takes up the analyte molecules or ions, whereby a hydrophobic complex or an ion-pair is formed.

## Gas Dialysis

Gas dialysis is based on the diffusion of a volatile solute from the donor solution through a gas-filled membrane or a membrane in which the volatile substance is soluble, into an acceptor solution.

By trapping a volatile analyte as a nonvolatile form these substances can strongly be enriched. Gas dialysis is used in FIA procedures and other flow-analytical methods to enhance their selectivity for volatile substances or substances, which can be converted into volatiles. The configurations shown in **Figures 5A–5D** are used and adapted. **Table 4** summarizes the applications of the gas dialysis technique. Several nonvolatile species, e.g., acetates can be separated after acidification. Gas dialysis membranes can

separate aqueous solutions with very different pH values and ionic strengths, which enables extreme samples also to be adapted to primarily unsuitable detection procedures. Microporous polytetrafluoroethylene or polypropylene membranes are used in most cases. However, it should be noted that, for example, surfactants and many water-soluble organic compounds adsorb on the membrane surface, which then becomes increasingly hydrophilized. In some of such situations, homogeneous membranes, e.g., silicon-rubber membranes can be advantageous. It should be noted, that silicon-rubber membranes are permeable to hydrogen sulfide, hydrogen cyanide, carbon dioxide, and many organic volatiles and they also have higher selectivity against other hydrophilic gases, e.g., NH<sub>3</sub> and SO<sub>2</sub> in comparison to microporous membranes with a hydrophobic inner surface.

## Reverse Osmosis

Despite its advantages, reverse osmosis is seldom used in analytical procedures. However, diluted sample solutions can be concentrated. Not only high molecular weight substances, but also low molecular weight substances, are rejected by ultrafiltration

**Table 4** Applications of gas dialysis

Analyte	Conversion to	Trapping as or by	Detection
Without conversion			phot, cheml, cond, pot, amp
Cl <sub>2</sub> , Br <sub>2</sub> , I <sub>2</sub> , ClO <sub>2</sub> , O <sub>3</sub> , H <sub>2</sub> O <sub>2</sub> , NO, NO <sub>2</sub>		A color or chemiluminescence reaction or reduction	
N <sub>2</sub> H <sub>4</sub>		Oxidation	cheml, pot
Ethanol, methanol, acetaldehyde, formaldehyde, acetic acid		Oxidation	phot, cheml
Propylene oxide		Oxidation, enzymatic conversion	phot, amp, fluor, cheml
Conversion by acid–base reactions		Hydrolysis, enzymatic conversion, oxidation	fluor
CN <sup>−</sup>	HCN	CN <sup>−</sup> , Ag(CN) <sub>2</sub> <sup>−</sup> , 1-cyanoisindole	amp, pot, fluor
SCN <sup>−</sup>	HSCN	SCN <sup>−</sup> , color-forming reaction	phot, pot
CO <sub>2</sub> , HCO <sub>3</sub> <sup>−</sup> , CO <sub>3</sub> <sup>2−</sup>	CO <sub>2</sub>	HCO <sub>3</sub> <sup>−</sup>	phot, pot
NH <sub>3</sub> , NH <sub>4</sub> <sup>+</sup> , RNH <sub>2</sub>	NH <sub>3</sub> , RNH <sub>2</sub>	NH <sub>4</sub> <sup>+</sup> , RNH <sub>3</sub> <sup>+</sup> , derivative of isoindole	pot, pot, cond
HS <sup>−</sup> , H <sub>2</sub> S	H <sub>2</sub> S	HS <sup>−</sup> , S <sup>2−</sup> , color-forming reaction	phot, pot, fluor
HSO <sub>3</sub> <sup>−</sup> , SO <sub>3</sub> <sup>2−</sup>	SO <sub>2</sub>	HSO <sub>3</sub> <sup>−</sup> , color-forming reaction	pot, coul, phot, fluor, cheml
F <sup>−</sup>	HF	F <sup>−</sup>	phot, pot
Conversion by redox reactions			
Cl <sup>−</sup> , Br <sup>−</sup> , I <sup>−</sup>	Cl <sub>2</sub> , Br <sub>2</sub> , I <sub>2</sub>	Reduction to Cl <sup>−</sup> , Br <sup>−</sup> , I <sup>−</sup>	phot, pot, amp
OCi <sup>−</sup> , ClO <sub>2</sub> <sup>−</sup> , ClO <sub>3</sub> <sup>−</sup> , BrO <sub>3</sub> <sup>−</sup>			

Phot, photometry; cheml, chemiluminescence; cond, conductivity; pot, potentiometry; amp, amperometry; fluor, fluorimetric detection.

membranes. The typical application of reverse osmosis is the separation of low molecular weight substances from aqueous solutions to purify water or to concentrate the substances that are to be determined. The water is propelled by a relatively high pressure gradient through a membrane, which is permeable to water but rejects dissolved molecules and ions. Bundles of hollow-fiber membranes are used in most technical applications, e.g., to desalt seawater.

Highly diluted sample solutions can be concentrated to analyte concentrations, which can be determined by the available determination method. The analyte can theoretically be concentrated up to the precipitation level. Then an additional filter layer is used, which is exchanged and directly analysed, e.g., by X-ray fluorescence spectrometry. For example, transition metals could be analysed in drinking water up to the micrograms per liter level.

See also: **Membrane Techniques:** Ultrafiltration; Liquid Membranes; Pervaporation.

## Further Reading

- Jönsson JA, Lövquist P, Audunsson, and Nilvé G (1993) Mass transfer kinetics for analytical enrichment and sample preparation using supported liquid membranes in a flow system with stagnant acceptor liquid. *Analytica Chimica Acta* 277: 9–24.
- Jönsson JA (2002) Liquid membrane techniques. In: Pawlczyn J (ed.) *Comprehensive Analytical Chemistry*, vol. XXXVII, pp. Amsterdam: Elsevier Science.
- Robinson T and Justice JB (1991) *Microdialysis in the Neurosciences*. New York: Elsevier Science.
- Torto N, Laurell Th, Gorton L, and Marko-Varga G (1999) *Analytica Chimica Acta* 379: 281–305.
- Ungerstedt U (1986) Microdialysis – a new bioanalytical sampling technique. *Current Separations* 7: 43–46.
- Valcarcel M and Luque de Castro MD (1991) *Nonchromatographic Continuous Separation Techniques*. Cambridge: Royal Society of Chemistry.

## Ultrafiltration

**B Spivakov and V Shkinev**, Russian Academy of Sciences, Moscow, Russia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The use of ultrafiltration (UF), microfiltration (MF), and other filtration techniques using semipermeable

**Table 4** Applications of gas dialysis

Analyte	Conversion to	Trapping as or by	Detection
Without conversion			phot, cheml, cond, pot, amp
Cl <sub>2</sub> , Br <sub>2</sub> , I <sub>2</sub> , ClO <sub>2</sub> , O <sub>3</sub> , H <sub>2</sub> O <sub>2</sub> , NO, NO <sub>2</sub>		A color or chemiluminescence reaction or reduction	
N <sub>2</sub> H <sub>4</sub>		Oxidation	cheml, pot
Ethanol, methanol, acetaldehyde, formaldehyde, acetic acid		Oxidation	phot, cheml
Propylene oxide		Oxidation, enzymatic conversion	phot, amp, fluor, cheml
Conversion by acid–base reactions		Hydrolysis, enzymatic conversion, oxidation	fluor
CN <sup>−</sup>	HCN	CN <sup>−</sup> , Ag(CN) <sub>2</sub> <sup>−</sup> , 1-cyanoisindole	amp, pot, fluor
SCN <sup>−</sup>	HSCN	SCN <sup>−</sup> , color-forming reaction	phot, pot
CO <sub>2</sub> , HCO <sub>3</sub> <sup>−</sup> , CO <sub>3</sub> <sup>2−</sup>	CO <sub>2</sub>	HCO <sub>3</sub> <sup>−</sup>	phot, pot
NH <sub>3</sub> , NH <sub>4</sub> <sup>+</sup> , RNH <sub>2</sub>	NH <sub>3</sub> , RNH <sub>2</sub>	NH <sub>4</sub> <sup>+</sup> , RNH <sub>3</sub> <sup>+</sup> , derivative of isoindole	pot, pot, cond
HS <sup>−</sup> , H <sub>2</sub> S	H <sub>2</sub> S	HS <sup>−</sup> , S <sup>2−</sup> , color-forming reaction	phot, pot, fluor
HSO <sub>3</sub> <sup>−</sup> , SO <sub>3</sub> <sup>2−</sup>	SO <sub>2</sub>	HSO <sub>3</sub> <sup>−</sup> , color-forming reaction	pot, coul, phot, fluor, cheml
F <sup>−</sup>	HF	F <sup>−</sup>	phot, pot
Conversion by redox reactions			
Cl <sup>−</sup> , Br <sup>−</sup> , I <sup>−</sup>	Cl <sub>2</sub> , Br <sub>2</sub> , I <sub>2</sub>	Reduction to Cl <sup>−</sup> , Br <sup>−</sup> , I <sup>−</sup>	phot, pot, amp
OCi <sup>−</sup> , ClO <sub>2</sub> <sup>−</sup> , ClO <sub>3</sub> <sup>−</sup> , BrO <sub>3</sub> <sup>−</sup>			

Phot, photometry; cheml, chemiluminescence; cond, conductivity; pot, potentiometry; amp, amperometry; fluor, fluorimetric detection.

membranes. The typical application of reverse osmosis is the separation of low molecular weight substances from aqueous solutions to purify water or to concentrate the substances that are to be determined. The water is propelled by a relatively high pressure gradient through a membrane, which is permeable to water but rejects dissolved molecules and ions. Bundles of hollow-fiber membranes are used in most technical applications, e.g., to desalt seawater.

Highly diluted sample solutions can be concentrated to analyte concentrations, which can be determined by the available determination method. The analyte can theoretically be concentrated up to the precipitation level. Then an additional filter layer is used, which is exchanged and directly analysed, e.g., by X-ray fluorescence spectrometry. For example, transition metals could be analysed in drinking water up to the micrograms per liter level.

See also: **Membrane Techniques:** Ultrafiltration; Liquid Membranes; Pervaporation.

## Further Reading

- Jönsson JA, Lövquist P Audunsson, and Nilvé G (1993) Mass transfer kinetics for analytical enrichment and sample preparation using supported liquid membranes in a flow system with stagnant acceptor liquid. *Analytica Chimica Acta* 277: 9–24.
- Jönsson JA (2002) Liquid membrane techniques. In: Pawlczyn J (ed.) *Comprehensive Analytical Chemistry*, vol. XXXVII, pp. Amsterdam: Elsevier Science.
- Robinson T and Justice JB (1991) *Microdialysis in the Neurosciences*. New York: Elsevier Science.
- Torto N, Laurell Th, Gorton L, and Marko-Varga G (1999) *Analytica Chimica Acta* 379: 281–305.
- Ungerstedt U (1986) Microdialysis – a new bioanalytical sampling technique. *Current Separations* 7: 43–46.
- Valcarcel M and Luque de Castro MD (1991) *Nonchromatographic Continuous Separation Techniques*. Cambridge: Royal Society of Chemistry.

## Ultrafiltration

**B Spivakov and V Shkinev**, Russian Academy of Sciences, Moscow, Russia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The use of ultrafiltration (UF), microfiltration (MF), and other filtration techniques using semipermeable

membranes enable separation that cannot be performed using other means. Membrane-based size fractionation is useful for determination of the distribution pattern of trace elements in natural waters and other fluids. Elements associated with suspended particles, colloids, and polymers as well as the ionic forms of these elements can be determined separately after membrane filtration separation and preconcentration. The possibility of preconcentration and separation of different species without using separating agents is another advantage of some membrane techniques. This is of particular interest in trace and ultratrace analysis to decrease the chance of contamination and in biological and clinical assays, where particles to be analyzed should not suffer changes during prepreparation.

Separation using UF and MF membranes is most selective, however, if soluble reagents are added. Such techniques may supplement two-phase distribution methods (e.g., liquid–liquid extraction, sorption, and precipitation), which are frequently used to extract species from dissolved matrices, industrial fluids, or natural waters. Although many such methods have been developed and successfully used, their application is sometimes troublesome. Some problems are caused by heterogeneous reactions and transfer between phases. Other problems can arise from the composition of the solution finally obtained, which is analyzed using the final determination method. In such cases, additional procedures may be required, e.g., back-extraction or desorption, which make the analytical procedure more complex and can cause additional contamination. Membrane separation can yield a homogeneous aqueous phase suitable for subsequent analysis using a number of methods.

## Principles

UF and MF are pressure-driven membrane transport processes used to separate macromolecules, colloids, microorganisms, and solid particles from fluids. Generally MF refers to processes used to remove particles greater than 0.5  $\mu\text{m}$ . For UF of macromolecules and colloidal particles, membrane filters of 0.002–0.5  $\mu\text{m}$  pore size are available. The membranes are necessary to filter viruses (0.03–1  $\mu\text{m}$ ) and bacteria (0.5–20  $\mu\text{m}$ ). Size cutoffs for MF are usually expressed in terms of micrometers, while UF cutoffs are often expressed in terms of molecular mass.

The first systematic study of UF was undertaken by Bechhold in 1907. Applications of UF are found in the food, drug, chemical, water, and wastewater treatment industries. The major use of MF is probably in sterile filtration. During the last 30–40 years

UF and MF have been successfully employed in analytical chemistry. The principal factors that should be considered for analytical and technological use of membranes with aqueous media are the pore size and pore size distribution, solution flow, and degree of hydrophilicity. The solution flow (flux) through an MF or UF membrane is given by the equation  $J = P/R$ , where  $J$  is the solution flux,  $R$  is a phenomenological resistance coefficient, and  $P$  is the transmembrane pressure drop. A pump or a gas (e.g., nitrogen) bottle can be used as a pressure source (50–500 kPa). The other main features of the filtration system are a membrane filtration unit and reservoirs.

## Membranes and Equipment

Flat-sheet asymmetric-skinned membranes made from synthetic polymers (also copolymers and blends), track-etched polymer membranes, inorganic membranes with inorganic porous supports and inorganic colloids such as  $\text{ZrO}_2$  or alumina with appropriate binders, and melt-spun ‘thermal inversion’ membranes (e.g., hollow-fiber membranes) are in current use. The great majority of analytically important UF membranes belong to the first type. They are usually made of polycarbonate, cellulose (esters), polyamide, polysulfone, poly(ethylene terephthalate), etc.

The pore size distribution in the membranes is another factor that may influence the reliability of membrane separations. Membranes with a relatively narrow distribution are produced. For example, the pore size distribution in some MF Millipore membranes are 0.025–0.05, 0.1–0.22, 0.45–0.65, 1.2–3.0, and 5.0–8.0  $\mu\text{m}$ . High uniformity of the pore size is attained in plastic track-etched (nucleoporous) filters, prepared by bombardment with high-energy heavy ions in cyclotrons with subsequent etching using an alkali solution. Since the open area is often low (0.1–1% of the surface), the housing geometries for membranes are designed to ensure high total areas and low feed pressures.

The so-called hollow-fiber modules meet some of these requirements. A hollow-fiber module consists of a large number of hollow fibers cemented on both ends to the module housing. The fibers may be some 500–1500  $\mu\text{m}$  in diameter, and they are formed of a self-supporting UF membrane.

Membranes are subject to concentration polarization. During UF, the solute accumulates at the membrane surface and the actual concentration may be much higher than that which corresponds to the bulk concentration. This process is called concentration polarization of the membrane. It can be minimized by use of stirred filtration cells or cross-flow units. In



standard stirred cells the liquid flow is parallel to the direction of applied pressure. In this case particles and large molecules may block the pores. Cross-flow or tangential-flow filtration units, where the liquid flow is in a narrow gap between two filters, are currently used. Such units enable faster filtration without blinding the filters as the accumulated solute is swept by the flow parallel to the membrane surface.

Membrane fouling is caused by capture of particles or formation of a gel-like layer on the membrane and the resultant decrease in the area available for fluid flow. Prefiltration, pH adjustment, or other means may be used to prevent fouling.

In multistage separations, minimization of concentration polarization, membrane fouling, and loss of separated components, which can occur in cascades of stirred cells or hollow-fiber modules due to the use of relatively lengthy tubing and a number of connectors and valves, is achieved by using an online multistage membrane filtration device (OMFD). The device consists of several disks coupled together in which different membranes with pore size from  $8\text{ }\mu\text{m}$  to  $1\text{ kDa}$  are inserted (Figure 1). The disks have fine channels on the upper part and a drainage system connected with the under-membrane space of the next disk in the lower part. The sample solution is fed from a reservoir to the first (lowest) filtration step, where it partially moves in parallel to the membrane surface (tangential flow). The filtrate penetrates into the chamber

(reservoir) of the next (upper) disk and is partially filtered and partially recirculated; the process repeats at each next step with a membrane of smaller pore size. Therefore, both the particulate and dissolved components of each fraction remain in the aquatic media after the filtration run and thus are taken for subsequent analyses.

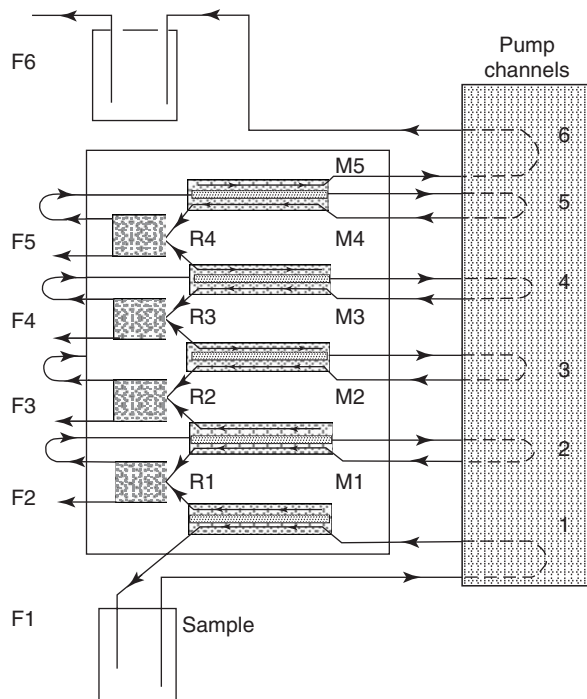
Matteson and Cheryan have provided detailed lists of commercially available membranes and described filtration units appropriate for analytical applications. The methods described below are based on utilization of 'inert' membranes, i.e., membranes that hardly absorb any of the solutes. Of course, any membrane has some adsorption ability, and so a certain conditioning should be provided to minimize the adsorption.

## Membrane Separation without Chemical Reagents

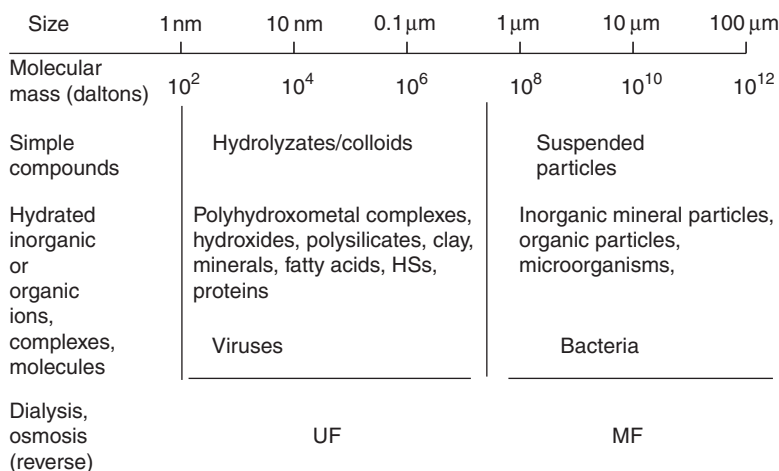
### Studies of Size Distribution Patterns for Trace Elements

Trace elements in natural waters may be present in different physicochemical forms varying in size, charge, and density. They may be associated with ligands ranging from simple ions to large molecules, or they may be absorbed onto or incorporated in organic colloids including humic and fulvic acids or inorganic particles, such as clay minerals and iron and manganese oxides. The binding of trace elements and radionuclides can enable the determination of transport, distribution, and biological uptake. Because of their impact on human and animal organisms, the distribution patterns of trace elements are important for environmental and biological studies. Methods using UF and MF are appropriate for the study of such patterns.

Figure 2 gives some characteristics of the size separation techniques that can be used to study the distribution of trace elements associated with various constituents of natural waters. It is obvious that the dimensions given in the figure are tentative as various factors influence the association/dissociation and aggregation/dispersion processes. However, preservation of real equilibria and labile species of elements, especially at concentrations of less than  $10^{-6}\text{ g l}^{-1}$ , prior to analysis is a much more serious problem encountered with methods that are not based on a direct physical separation. From this point of view, membrane filtration as well as some variants of field-flow fractionation (FFF) have advantages, although some uncertainties connected with equilibria shifts always exist.



**Figure 1** Flow diagram of a tangential-flow OMFD. M, membrane; F, fraction; R, retentate.



**Figure 2** Scheme for the speciation analysis of trace elements in river water using size fractionation techniques.

The pore size distribution is an essential characteristic of filters used to study size distribution patterns. As track-etched filters have a very narrow pore distribution (geometric standard deviation being less than  $\sim 1.1$ ), they are often used if fine solid particles are to be separated.

Track-etched polyethyleneterephthalate filters with pore sizes ranged from 8 to 0.025 μm and other types of membranes made of polysulfone, mixed cellulose ester, regenerated low binding cellulose, polycarbonate, and polyvinylidene fluoride, with pore sizes from 1.2 μm to 1 kDa, have been tested on OMFDs using soluble polymers and particles of known sizes, and the particle size distributions obtained have been checked using laser light scattering measurements. This technique has been subsequently applied to river and other natural waters of Germany, Russia, Switzerland, the Netherlands, and other European countries. Using such filtration systems five or six fractions of different particle sizes have been obtained, subsequently to be analyzed, e.g., using inductively coupled plasma (ICP)–atomic emission spectroscopy (AES) or mass spectrometry (MS), yielding the distribution patterns of the trace and major elements in waters. Suspensions of fine particles derived from polluted areas, e.g., from former uranium and tin mines located in central Germany, have also been analyzed after fractionation on an OMFD.

Informative distribution patterns can be obtained using only two or three filters having different pore sizes if the membrane filtration is combined with acidification of water fractions. Thus, an analytical scheme for determining the speciation of aluminum in natural waters has been described. The scheme is based on sizing by filtration of acidified and unacidified water sample portions through 0.1 and 0.4 μm

polycarbonate membrane filters. The following aluminum species have been detected using ultraviolet (UV)–visible spectrophotometric and atomic adsorption measurements after the separation: particulate, colloidal and soluble, large colloidal, fine colloidal and soluble, fine colloidal complexed, and labile (aluminum forms such as  $\text{Al}^{3+}$ ,  $\text{Al}(\text{OH})^{2+}$ , and  $\text{Al}(\text{OH})_2^+$  as well as total aluminum).

In the experiments with aluminum, the filtration was carried out under stirring. Acidification procedures and utilization of stirred filtration cells should be used with care because they may result in equilibria shifts.

Stirring can influence not only the equilibria between solutions and suspended colloidal forms but also the flux through UF membranes. Besides, the membranes of stirred UF cells clog relatively quickly in the presence of dissolved organic carbon, and the polarized membrane surface changes its retention behavior. In contrast, in the use of the tangential-flow technique, the surface of the membrane is continuously cleared of small particles and allows high flow rates. Hollow-fiber cartridges having a nominal molecular mass cutoff of  $10^6$ ,  $10^5$ ,  $10^4$ , or  $10^3 \text{ g mol}^{-1}$  have been used for successive filtration separations in natural water analysis.

Clogging and sorption effects have also been investigated for aluminum in natural waters. On the basis of the results obtained and studies of distribution patterns for aluminum, iron, and zinc as well as some radionuclides in natural waters, it has been shown that the main advantages with hollow-fibre fractionation are minimal clogging and sorption problems and a high filtering capacity in comparison to those of stirred-cell techniques.

Different methods for determination of the physicochemical forms of metals in waters have been

compared. If the fractionation is rapid, the risk of equilibria shifts during membrane filtration is less than for other fractionation techniques. The absence of gels or other separating materials that can adsorb trace elements and their species more readily than do polymer membranes enables minimization of sorption. The adsorption effects may also be reduced if preconditioning of the filtration unit with a sample aliquot is used. This can be easily done because MF, UF and FFF are faster than gel filtration, to say nothing of such a fractionation technique as sedimentation.

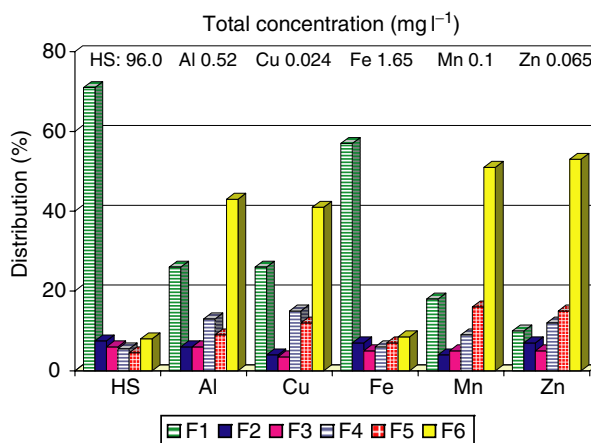
Different techniques are in use for analysis of the filtration fraction and concentrates. Most often multielement techniques such as ICP AES/MS and X-ray fluorescence (XRF) spectrometry are applied. Use of electrochemical methods and different techniques of atomic absorption spectrometry (AAS) is also possible.

It is obvious that the use of one membrane filtration method as a single separation technique cannot give comprehensive information on metal speciation in aqueous media. Combinations of UF with FFF, sedimentation, ion exchange, and liquid-liquid extraction are useful for the identification of the species of an element. However, membrane filtration itself is a rather powerful means if a number of membranes having different pore size are used for studying one sample.

### Membrane Filtration of Organic and Bioorganic Substances and Their Metal Complexes

Not only the molecular mass but also the conformation of organic molecules plays an important role in their retention by membranes; clearly the retention depends on the molecular shape and increases with the rigidity of the molecule. Similar results have been obtained for synthetic water-soluble polymers, proteins, and other biological macromolecules. It has been shown that the nature of the solvent, the nature and concentration of the electrolytes or complexable ions, and an increased concentration of polymer at the membrane surface due to concentration polarization may modify the conformation of the polymer and therefore its retention.

A special group of soluble organic macromolecules comprises humic substances (HSs; humic and fulvic acids), which are important transport agents for inorganic and organic contaminants, including radionuclides, in natural waters. According to current estimates, more than 50% of the dissolved organic carbon in natural waters consists of polyelectrolytes of the HS type. Therefore, characterization of HSs and their complexes is important for environmental



**Figure 3** Molecular-size classification of a sample of an HS (Venner Moor, Muensterland, Germany, 48 mg l<sup>-1</sup> DOC, pH 3.7) and its metal species in a selected bog water using an OMFD device. Fractions: 1, > 100 kDa; 2, 50–100 kDa; 3, 10–50 kDa; 4, 5–10 kDa; 5, 1–5 kDa; 6 < 1 kDa.

analytical chemistry. The possibility of characterizing HS in the colloidal and macromolecular size ranges after size fractionation on sequential-stage UF devices like OMFDs and on cascades of hollow-fiber modules has been shown. As an example, in **Figure 3** the size distribution patterns of an HS (Venner Moor, Muensterland, Germany) and some naturally occurring metal species in a typical bog water obtained on an OMFD, with UV spectrophotometric and AAS studies, are presented. Using UF, sized samples of the HS were chemically characterized using nuclear magnetic resonance spectroscopy, Fourier transform infrared spectroscopy, and other techniques. The studies have demonstrated that sufficient material can be obtained with UF with minimal alteration to the material. Inorganic analysis performed on the size fractions has shown that UF can also allow the binding capacity of the HS in each fraction to be measured.

UF has also been applied to the determination of stability constants for trace metal-HS complexes, equilibrium constants for the distribution of organic compounds between water and synthetic micelles, the binding of metal ions and small compounds to proteins and other macromolecules in biochemistry, etc. This type of application is essential to know precisely to what degree the complexing agent, the complexed species, and the free ions and molecules can pass through the membrane.

MF is used in bacteriological analysis of municipal waters and waste treatments. The identification of coliform bacteria in water indicates fecal contamination and the possible presence of pathogenic enteric organisms, which are responsible for causing

various diseases. The procedure commonly used in bacteriological assays in sanitation inspection consists of filtering the water sample, inhibiting or selectively growing the retained organisms, and counting these organisms. A 0.45  $\mu\text{m}$  pore diameter membrane is usually used in the water analysis. Microfilters are also applied in medical and clinical techniques in the detection of specific cells. Cytological examination with membrane filtration is performed on blood and urine.

The concentration of viruses in water is much lower than that of bacteria; therefore their preconcentration and separation from other suspended matter is important in virological analysis. Viruses may be enriched by UF by use of tangential-flow or hollow-fiber filtration units most suitable for preconcentration from large-volume water samples. Another possibility consists in adsorption of the viruses on the microfilter matrix followed by elution with diluents containing cell extracts, serum, or surfactants.

### Membrane Preseparations in Continuous-Flow Instruments

Physical separation of inorganic ions and other low-molecular analytes from the particulate and dissolved macromolecular fractions of the sample is useful in different flow instruments. Dialysis and UF membrane cells are amongst practical systems for flow injection analysis (FIA), which serve to exclude any unwanted sample material so that only the analyte reaches the reaction or sensing zone. Solid particles and large interfering molecules can be separated using membranes from the smaller species of analytical interest. Such cells can also be used in the continuous-flow, stopped-flow, and sequential injection modes of flow analysis. The membrane preseparation units are very attractive because they are simple, repeatable, require little or no pretreatment of sample, and show no interferences from sample color and turbidity; in most cases no reaction in the membrane interface is involved.

Membrane filtration can be used in sample preparation for liquid chromatography, e.g., in the ion-chromatographic determination of various inorganic anions and cations in water samples. Tubular membrane interfaces have been used for sample introduction in capillary zone electrophoresis to separate the low-molecular mass organic constituents of blood plasma.

Reagent delivery cells with a track-etched polycarbonate membrane filter (pore size 0.1–0.01  $\mu\text{m}$ ) for online dilution of concentrated salt solutions have been described. By using a reagent delivery cell, the

ageing of solutions is avoided, leading to long-term stability, and pipetting errors and waste are minimized. Such cells are useful for calibration purposes, required in many areas of analytical chemistry.

## Membrane Separation Using Chemical Reagents

### Collection of Trace Elements on a Membrane Filter

A chemical reagent added to the aqueous sample, added to the recipient solution, or built in the membrane can improve the separation or increase the concentration efficiency. If the reagent is built in, the membrane is called adsorptive or chemically active. If an 'inert' solid membrane is used, an inorganic reagent or organic monomeric or polymeric agent is added to the sample to obtain insoluble compounds of analytes. Different cellulose-based membranes (pore diameter 0.45–3.0  $\mu\text{m}$ ) have been utilized for the purpose.

A sensitive spectrophotometric method for the determination of phosphorus has been proposed. The ion associate of molybdophosphate with Malachite Green is filtered through a membrane filter (cellulose nitrate; pore size 1.0  $\mu\text{m}$ , diameter 9 mm) and then dissolved together with the membrane filter in methyl Cellosolve and the absorbance measured using a FIA system equipped with an autosampler. For filtration of the ion associate, a special filtration apparatus has been designed. Other procedures have been developed for determination of phosphate, arsenate, arsenite, and silicate as heteropoly species collected on the filter and dissolved, together with the filter, in dimethyl formamide or another organic solvent. The analytes can also be washed out from the membrane surface for subsequent analysis of the liquid concentrate, or they can be directly analyzed on the filter using XRF. Various metal ions have been determined after preconcentration on a membrane in the form of metal chelates.

Removal of the analyte-containing precipitate using a thin membrane filter is similar to conventional filtration using a filter paper. However, there is a difference: the precipitate does not penetrate the membrane filter material, thus facilitating its quantitative removal from the membrane surface for subsequent analysis.

### Liquid-Phase Polymer-Based Retention

Soluble complexing polymers are being used in selective membrane separations of inorganic ions in a homogeneous aqueous phase. Such polymeric reagents form soluble polymer complexes that are

separated from low-molecular mass compounds through membrane filtration processes. Thus, uncomplexed inorganic ions can be removed with the filtrate, whereas the water-soluble polymer complexes are retained. To separate and preconcentrate metal ions, their solution, to which a polymer is added, can be passed from a reservoir into the filtration unit. Another procedure of absolute preconcentration has been recommended in which the polymer reagent is added directly to the cell and then a large volume sample is passed through a solution of a relatively high and constant polymer concentration.

However, interfering components of the test solution partly remain in the cell after the filtration run, even when they do not interact with the reagent. This can cause difficulties, e.g., in trace analysis of highly mineralized waters. Therefore, a combined procedure is of interest: a sample with a large volume is passed through a much smaller volume of polymer solution in the cell and then washed with pure water to remove remaining amounts of constituents not complexed by the polymer. Thus, both absolute and relative preconcentration are achieved.

Conventional stirred filtration cells or a specially designed radial-flow cell equipped with a pump can be used. Membranes made of polysulfone, polyamide, cellulose, etc. are suitable. The essential parameters are the molecular mass exclusion rate in wide pH ranges (1–10) and an appropriate permeate flow rate ( $1\text{--}10\text{ ml min}^{-1}$ ), retentate volume (2–10 ml), and gas pressure (300 kPa is a suitable pressure in most cases). A nominal exclusion rate of  $10\text{ kg mol}^{-1}$  has been shown to be convenient for polymers having a molecular mass between 30 and  $50\text{ kg mol}^{-1}$ . A polymer concentration of 1–4% (w/v) in the cell solution is most appropriate for both retention of elements and their subsequent determination in the retentate.

Poly(ethylene imine) (PEI), poly(acrylic acid), poly(1-vinyl-2-pyrrolidone), and other widely used polymers have been used for preconcentration and separation of various elements. Reagents with typical acidic chelating groups have also been used: 8-quinolinol and iminodiacetic groupings, neutral thiourea groups and trimethylammonium-based anion-exchange groups. Introduction of strong chelating groups into the basic polymer backbone results in more effective retention of metal ions.

The technique described above is called liquid-phase polymer-based retention (LPR) or polyelectrolyte-enhanced UF. Another variant of the separation

and preconcentration of low-molecular mass solutes in an UF cell, called micellar-enhanced ultrafiltration, is based on solute retention in the micellar phase obtained from surfactants. Procedures have been developed for the preconcentration of Al as a complex with lumogallion, of Co and Ni with organophosphorus reagents in the presence of a cationic or nonionic surfactant, and of various heavy metals with PEI in the presence of inorganic colloids formed by iron(III) hydroxide.

Preconcentration of elements using the LPR method can be combined with different determination methods, e.g., with AAS, ICP AES or MS, stripping voltammetry, and other techniques applicable to the analysis of aqueous solutions. The final solution (concentrate) obtained after the separation is an aqueous solution of 1–4% polymer that does not interfere significantly with the determination methods mentioned above.

**See also:** **Membrane Techniques:** Dialysis and Reverse Osmosis; Liquid Membranes; Pervaporation.

## Further Reading

- Buffle J, Perret D, Newman M (1992) The use of filtration and ultrafiltration for size fractionation of aquatic particles, colloids, and macromolecules. In: Buffle J, van Leeuwen HP (eds.) *Environmental Particles*, vol. 1, pp. 171–230. Chelsea, MI: Lewis Publishers.
- Burba P, Aster B, Nifant'eva T, Shkinev V, and Spivakov BYa (1998) Membrane filtration studies of aquatic humic substances and their metal species: a concise overview. Part 1. Analytical fractionation by means of sequential-stage ultrafiltration. *Talanta* 45: 977–988.
- Cheryan M (1998) *Ultrafiltration and Microfiltration Handbook*. Basel: CRC Press.
- Matteson MJ and Orr C (eds.) (1987) *Filtration: Principles and Practices*. New York: Marcel Dekker.
- Shkinev VM, Fedorova OM, Spivakov BYa, Mattusch J, and Wennrich R (1996) Speciation of metals associated with natural water components by on-line membrane fractionation combined with inductively coupled plasma atomic emission and mass spectrometries. *Analytica Chimica Acta* 327: 167–175.
- Spivakov BYa, Shkinev VM, and Geckeler K (1994) Separation and preconcentration of trace elements and their forms in aquatic media using inert porous membranes. *Pure and Applied Chemistry* 66: 631–640.
- Van Staden JF (1995) Membrane separation in flow injection systems. *Fresenius Journal of Analytical Chemistry* 352: 271–302.



## Liquid Membranes

**S D Kolev**, The University of Melbourne, Melbourne, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The analysis of complex samples frequently necessitates pretreatment of the sample to separate the analyte from the matrix. In some cases analyte separation also results in preconcentration. Conventional analytical separation involves physicochemical processes, such as precipitation, adsorption, ion exchange, and extraction.

Extraction, in its various forms, is one of the most frequently used separation processes in the analytical practice. In this process the analyte is transferred from the sample, usually a liquid, to another immiscible phase, which is usually a condensed phase. If the condensed phase is a liquid the extraction technique is called liquid-liquid or solvent extraction. In the case of a solid condensed phase, the corresponding extraction technique can be either solid-phase extraction (SPE) or solid-phase microextraction (SPME).

In liquid-liquid extraction (LLE), the analyte is partitioned between two immiscible liquid phases, which are usually an aqueous sample and an organic solvent. The possibility of dissolving organic compounds in the solvent capable of chemically interacting with the analyte (e.g., forming complexes or ion pairs) can improve substantially the selectivity and efficiency of the extraction process on the one hand and widen the range of analytes (e.g., from ionic inorganic species to nonpolar organic compounds), which can be separated by LLE, on the other. The main disadvantages of LLE are associated with the use of relatively large volumes of organic solvents, which are often expensive and toxic and may form stable emulsions with the aqueous phase. The low sample to solvent volume ratio does not allow high enrichment factors. These drawbacks of LLE can be overcome to a considerable extent by SPE, SPME, and solvent microextraction (SME), known also as liquid-phase microextraction. When the analytical procedure involves further separation and detection in the aqueous phase (e.g., reversed-phase high-performance liquid chromatography) following the LLE step, an additional back-extraction step from the organic phase into a new aqueous phase must be introduced, which complicates the separation process.

In SPE, a solid sorbent material (e.g., resin or alkyl bonded silica) packed into a cartridge or imbedded

into a polytetrafluoroethylene (PTFE) or glass fiber disk, replaces the organic solvent used in LLE. Separation in SPME is based on absorption or adsorption of the analyte into the coating of a fused silica fiber, which may be a thin nonvolatile liquid film (e.g., polydimethylsiloxane) or a porous solid (e.g., C18-bonded silica). The relatively small number of suitable sorbents reduces the applicability of both SPE and SPME to the separation of mainly organic analytes of low polarity and limits the selectivity of these techniques compared to that offered by LLE. Analytes are usually thermally desorbed in SPME while analyte adsorption in SPE is followed by elution with a relatively small amount of an organic solvent.

In SME, a single microdrop of the organic phase (e.g., n-octane, toluene), which may contain an internal standard, is usually suspended in the sample solution on the tip of a microsyringe needle. After a predetermined extraction time the drop is retracted back into the microsyringe to be injected later into a gas chromatograph. Repeated suspending and retracting of the organic microdrop in the microsyringe can induce convective mass transfer within both the microdrop and the sample solution adjacent to it, resulting in high enrichment factors. Alternatively, the concentration of the extracted analyte can be determined spectrophotometrically directly in the microdrop if it forms a colored compound. SME has been successfully applied to headspace analysis. Similarly to LLE, SME offers higher selectivity compared to SPE and SPME, but it is not readily adaptable to analytical separations where back-extraction is required.

Membrane separation involving liquid membranes is an attractive alternative to the extraction techniques outlined above. Like SPE and SPME, separation based on liquid membranes requires small amounts of organic solvents, if any, while at the same time it offers the high selectivity, flexibility, and enrichment factors of LLE and SME. Liquid membranes are suitable for online analytical separation, since the extraction and back-extraction are integrated into a single step, unlike the sequential arrangement of these processes when other extraction techniques are used.

### Membranes and Related Processes Used in Analytical Separation

In most membrane separation processes, the membrane separates two miscible or immiscible fluid

phases, each one of which can be either static or mobile. Most frequently the two fluids are liquids, one of which is the liquid sample (often referred to as the feed or donor solution) while the other liquid is usually referred to as the receiver, strip, or acceptor solution. The membrane prevents mixing and, very often, direct contact between the two solutions. The latter function of the separation membrane is particularly important when the feed and receiver solutions are miscible fluids or when analyte preconcentration is also required.

The membranes used in analytical separation processes are usually single- or dual-phase membranes.

### Single-Phase Membranes

Depending on its aggregate state a single-phase (SP) membrane can be solid or liquid. Typical examples of solid SP separation membranes are nonporous polymer membranes made of silicone rubber. These membranes have been used successfully in membrane introduction mass spectrometry or gas chromatography where a receiver gaseous phase is used. Mechanical stability and relatively long lifetime compared to other types of separation membranes are among the main advantages of nonporous polymer membranes. However, these membranes are hydrophobic in character and thus are limited to the separation of mainly nonpolar analytes. Another drawback of nonporous polymer membranes is their low permeability, resulting in slow mass transfer between the feed and receiver solutions.

Emulsion and bulk liquid membranes, which will be outlined later in this article, are typical liquid SP membranes.

### Dual-Phase Membranes

The dual-phase (DP) membrane used in analytical separation usually consists of a polymer, or in some cases a ceramic solid-phase support impregnated with a fluid (i.e., gaseous or liquid phase). If the fluid is air the DP membrane is known as a gas-diffusion membrane. DP membranes incorporating a liquid phase can be considered in a broader sense as liquid membranes. The liquid phase in a liquid DP membrane can be identical to the feed and/or receiver solution (e.g., dialysis membranes, membrane-assisted LLE (MALLE)) or it can form a third immiscible liquid phase in the membrane separation system (e.g., supported and polymer liquid membranes). Membranes incorporating a liquid phase immiscible with the feed and receiver solutions are usually referred to as liquid membranes. This narrower definition of liquid membranes, currently accepted in the literature, will be used in subsequent discussions. The

membranes covered by this definition, which can be either SP or DP membranes, will be outlined in more detail separately.

**Gas-diffusion membranes** Hydrophobic porous polymer membranes with air filling the membrane pores have been used successfully in the online separation of volatile and semivolatile analytes between two miscible liquid streams in flow injection analysis (FIA) systems. The corresponding technique is frequently referred to as gas-diffusion FIA. The mass transfer of an analyte across a gas-diffusion membrane is controlled by the membrane pore size and the solubility of the analyte in the feed and receiver solutions. The latter can be manipulated by appropriately modifying the chemical composition of the two solutions. In this way it is possible to enhance both the evaporation of the analyte from the feed solution into the membrane pores and its subsequent absorption into the receiver solution.

**Membranes for dialysis and membrane-assisted liquid-liquid extraction** Depending on the hydrophobicity of a porous polymer membrane and the solutions in contact with it, the membrane pores can be filled with the feed solution, the receiver solution, or a mixture of both solutions. If both solutions are miscible (e.g., aqueous) the separation of the analyte is based on dialysis, while if they are immiscible the separation process is often referred to as MALLE.

Dialysis describes a separation process usually involving two aqueous phases (i.e., feed and receiver solutions) divided by a hydrophilic (e.g., cellulose acetate) porous membrane. The membrane pores are filled with water and solutes from both the feed and the receiver solutions. Typically, dialysis is utilized for the determination of low molecular weight analytes in samples containing macromolecular species like proteins (e.g., chloride in milk). The average pore size determines the molecular mass range within which chemical species can cross the membrane. The driving force in dialysis is the analyte concentration gradient across the membrane (i.e., passive dialysis). If an external electric field is applied electrodialysis takes place. Both dialysis and electrodialysis have been frequently used as online separation techniques in FIA systems. When the analytes are charged chemical species, they can be separated from the sample matrix and preconcentrated in an acceptor solution by means of Donnan dialysis. In this process, an appropriately selected ion-exchange membrane (e.g., Nafion) separates the sample from an acceptor solution of a smaller volume and higher ionic strength. Ions of appropriate charge from the receiver solution are transported into the sample

solution as a result of the existing ionic strength gradient while co-ions from the sample solution including the analyte ions diffuse in the opposite direction in order to maintain electroneutrality. Electrodialysis and Donnan dialysis, unlike passive dialysis, allow preconcentration of the analyte as well.

In MALLE a hydrophobic or hydrophilic porous polymer membrane separates two immiscible phases, usually an organic receiver phase and an aqueous feed (sample) phase. If the membrane is hydrophobic, its pores are filled with the organic receiver phase and the actual analyte partitioning takes place at the membrane/feed solution interface. The analyte then diffuses across the membrane into the bulk of the organic receiver solution. Unlike LLE, MALLE does not require mechanical mixing of the two phases followed by phase separation, as a result of which the extraction process can be miniaturized and conducted online. However, similarly to LLE, analyte separation and preconcentration in MALLE is governed by the corresponding aqueous/organic partition coefficient. MALLE has been implemented in FIA as an attractive alternative to the technically more complex segmented LLE flow injection approach, involving phase segmentation, extraction, and phase separation prior to detection in the organic phase. Another recent analytical application of MALLE, based on SME, is the so-called extracting syringe or hollow fiber protected liquid-phase microextraction where a hollow fiber connected to the stainless steel needle of a microsyringe is filled with a suitable organic solvent (e.g., cyclopentane) and exposed to a flowing or mechanically stirred aqueous sample. After extraction, the solvent in the fiber is retracted back into the syringe and subsequently injected into a gas chromatograph.

## Liquid Membranes

By definition the membrane liquid phase of a liquid membrane must be immiscible with the solutions in contact with it. This allows the chemical composition of all three liquid phases (including the membrane liquid phase) to be altered independently of one another. For this reason liquid membranes offer a higher degree of control over the membrane separation process compared to other types of separation membranes.

### Classification

The liquid membranes most often employed in analytical and industrial separation are: supported liquid membranes, polymer liquid membranes, emulsion liquid membranes, and bulk liquid membranes.

**Supported liquid (SL) membranes** The SL membrane is the most frequently used type of liquid membrane in analytical separation. An SL membrane can be constructed by impregnating the pores of a porous hydrophobic membrane (e.g., PTFE; poly(vinylidene fluoride) (PVDF); polypropylene) with a suitable organic liquid (**Figure 1A**). The porous solid membrane phase acts as a support for the liquid phase retained in the membrane pores by capillary forces.

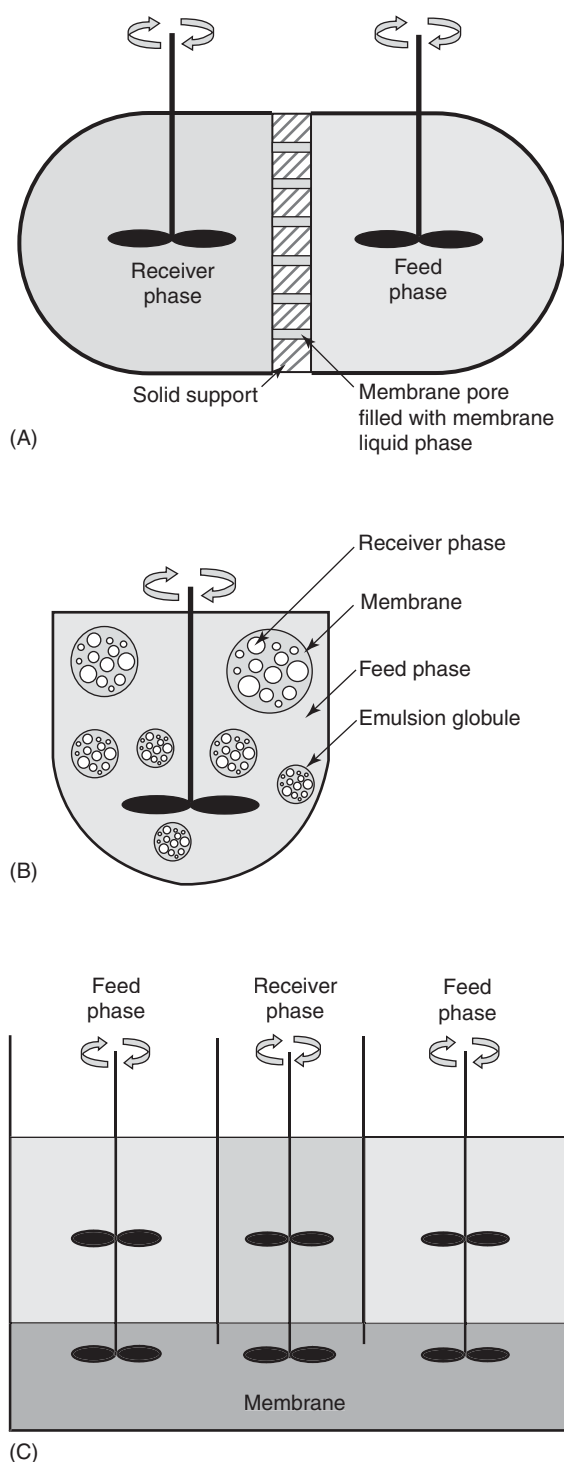
The main disadvantage of the SL membrane is the slow leaching of the membrane liquid phase into the receiver and feed solutions, thus reducing its lifetime. Attempts have been made to slow down the loss of the membrane liquid phase by converting it to a polymer gel. The corresponding SL membrane is known as gelled SL membrane.

The liquid membrane in a recently introduced solvent microextraction/back-extraction (SME/BE) technique, where the organic phase separating the feed and receiver aqueous phases is located within a PTFE ring (i.d. 6.4 mm), can be viewed as an SL membrane with a single macropore. Very high enrichment factors can be achieved if the aqueous receiver solution is suspended as a microdrop from the tip of a microsyringe needle in the organic membrane. SME/BE allows the use of fresh liquid membrane for each extraction, thus eliminating memory effects and long-term stability problems.

**Polymer liquid (PL) membranes** The PL membrane is a relatively new type of self-supporting liquid membrane, which resembles the SL membrane. Similarly to the SL membrane, the PL membrane incorporates a liquid extractant in the membrane polymeric structure. In some cases the incorporation of additional organic compounds (plasticizers) is required to achieve homogeneity and sufficient flexibility of the membrane.

PL membranes are usually cast from a homogeneous solution of the dissolved solid polymer material (e.g., poly(vinyl chloride)), the liquid extractant (e.g., Aliquat 336 chloride; DEHPA) and a plasticizer (e.g., dioctylphthalate; n-decanol) in a suitable solvent (e.g., tetrahydrofuran) by the slow evaporation of the solvent. By increasing the concentration of the extractant the permeability of the PL membrane increases while at the same time its mechanical stability deteriorates. For example, the optimal concentration range for Aliquat 336 chloride in a polyvinyl chloride based PL membrane is between 40 and 50%.

The structures of PL membranes are currently being elucidated. Surface analysis results have suggested the presence of networks of nano-sized channels filled with the liquid extractant and diluent. The small size of these channels explains the facts that PL



**Figure 1** Schematic of: (A) SL membrane sandwiched between two mechanically stirred extraction cells; (B) EL membrane; and (C) BL membrane.

membranes are homogeneous at macrolevel, and that the rate of loss of the liquid membrane phase components is small compared to that in SL membranes. The latter fact is responsible for the longer lifetime of PL membranes.

**Emulsion liquid (EL) membranes** A typical EL membrane consists of emulsion globules (extracting emulsion) of the receiver solution and the organic membrane liquid phase. The emulsion globules are stabilized by a surfactant (e.g., sorbitan monooleate) and are dispersed in the feed aqueous solution (Figure 1B). The resulting double emulsion is characterized by a large interfacial surface area between the emulsion globules and the feed solution on the one hand, and within the emulsion globules, between the receiver phase, and the organic liquid membrane phase on the other. This configuration allows fast mass transfer between the feed and receiver phases. By appropriate selection of the composition of the extracting emulsion (i.e., electrolyte composition of the receiver phase, concentration of emulsifier, adsorption of solid particles on the interface of the emulsion globules, formation of nanodispersion droplets by using microemulsifiers) high stability to coalescence and sedimentation can be achieved. These properties of EL membranes are crucial in avoiding mixing between the feed and receiver solutions during the separation process.

After equilibrium in the system has been attained, the aqueous feed phase is separated from the receiver phase emulsion globules, which are further demulsified (e.g., by adding *n*-butanol) to form an organic liquid membrane layer and an aqueous receiver layer. After the separation of these two liquid layers the aqueous receiver solution is analyzed.

EL membranes allow high enrichment factors due to the large ratio of feed solution to receiver solution volume. The presence of emulsifiers and the relatively larger volume of the membrane liquid phase in EL membranes compared to that in SL membranes allow the organic diluents and carriers in EL membranes to be less hydrophobic than those normally used in SL membranes. This extends the range of membrane liquid phases that can be used in EL membrane extraction.

The currently existing difficulties in handling EL membranes, associated mainly with the formation and breakdown of the emulsion itself, have prevented the widespread use of this type of liquid membrane in the analytical practice despite the advantages outlined above.

**Bulk liquid (BL) membranes** A BL membrane consists of a layer of membrane liquid phase in contact with both the aqueous feed and receiver solutions, located in separate compartments. Depending on its density, the membrane liquid phase can be placed either above or below the two aqueous phases (Figure 1C). Mechanical stirring of the feed and receiver solutions, and in some cases of the membrane



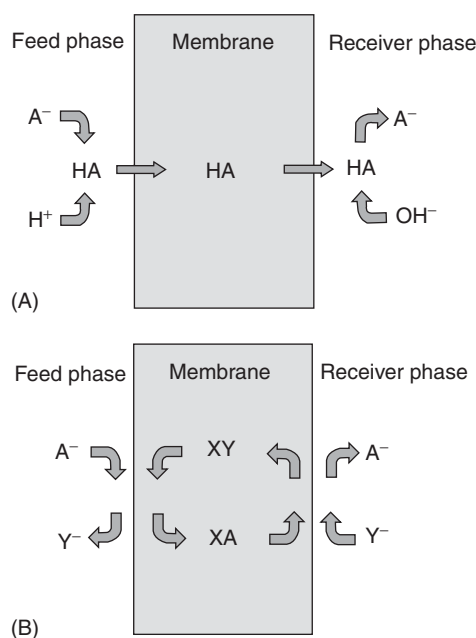
liquid phase, is required to enhance the mass transfer processes in the three-phase system. BL membranes are not suitable for routine analyte separation because of the complex manual operations involved in conducting this analytical procedure. The main analytical interest in this type of liquid membrane is based on the fact that they use small quantities of carrier and diluent compared to traditional LLE, and therefore are suitable for screening carrier properties.

### Liquid Membrane Separation

**Mechanism** Membrane separation involves the following membrane-based processes: (1) transfer of the analyte from the feed solution to the membrane liquid phase as a result of partitioning or chemical reaction, producing a lipophilic complex or ion pair; (2) membrane transport of the analyte or its complex by Fickian diffusion driven by the corresponding concentration gradient across the membrane; and (3) transfer of the analyte from the membrane liquid phase to the aqueous receiver solution. The membrane concentration gradient can be maintained at its maximum by minimizing the analyte concentration in the receiver solution at the corresponding membrane/solution interface. This result can be achieved by (1) converting the analyte entering the receiver solution into another chemical species, usually using a suitable protolytic or complexation reaction; (2) inducing convective mass transfer (e.g., mechanical stirring or flow); or (3) combining approaches (1) and (2). Approach (2) can also be used to enhance the mass transfer of the analyte from the bulk of the feed solution to the corresponding membrane/solution interface.

Depending on the chemical composition of the membrane liquid phase, the mass transfer of the analyte from the feed to the receiver solution can be either passive or facilitated.

**Passive mass transfer** In the case of passive mass transfer, the membrane liquid phase consists of an organic solvent or a mixture of organic solvents. The transfer of the analyte across both membrane/solution interfaces is governed by its partition coefficient. **Figure 2A** illustrates schematically the passive transport of an organic acid across a liquid membrane involving suitable protolytic reactions in both the feed and receiver solutions. If the volume of the receiver solution is smaller than the volume of the feed solution the analyte of interest can be concentrated as well. The mechanism of passive liquid membrane separation is analogous to that involved in separation based on solid SP and gas-diffusion membranes. However, unlike these membranes, liquid membranes



**Figure 2** Schematic representation of: (A) passive membrane mass transfer; and (B) facilitated membrane mass transfer ( $A^-$  is the analyte anion and  $X^+$  is a cationic carrier).

exhibit a higher degree of flexibility due to the ease with which the membrane liquid phase composition can be modified to meet the specific requirements of the analyte of interest (e.g., polarity).

Organic solvents used in liquid membranes for both passive and facilitated transfer can be either nonpolar aliphatic or aromatic hydrocarbons such as toluene and kerosene, or polar organic compounds such as dihexylether. The polarity of the membrane liquid phase can be modified readily by the addition of small amounts of hydrophobic polar organic solvents (e.g., trioctyl phosphine oxide, dodecanol). Liquid membranes are generally more permeable than solid SP membranes thus allowing faster separation. They are also suitable for the separation of both volatile and nonvolatile analytes.

**Facilitated mass transfer** Similarly to LLE, the selectivity and efficiency of the liquid membrane separation process can be considerably improved if a suitable extractant with a high selectivity for the analyte of interest is used. This extractant, often referred to as the carrier, facilitates the mass transfer of the analyte between the feed and receiver solutions. The liquid membrane phase in this case usually consists of a suitable liquid extractant or an extractant dissolved in an organic solvent (diluent). The extractant facilitates the transport of the analyte from the feed phase to the liquid membrane phase by chemically interacting with it. This interaction leads to the selective extraction of the analyte into the



membrane liquid phase as a complex or an ion pair. The analyte is released into the receiver solution at the membrane/receiver solution interface as a result of another ion exchange reaction, and the freed carrier diffuses back to the membrane/feed solution interface. A typical liquid membrane cationic carrier is the positively charged quaternary ammonium ion Aliquat 336 (tricaprylmethylammonium cation). Frequently used anionic carriers are the negatively charged lipophilic conjugated bases of organic acids (e.g., di-2-ethylhexyl phosphoric acid, DEHPA). Both cationic and anionic carriers form neutral ion pairs with the oppositely charged ionic analyte. Metal complexes of macrocyclic ligands with hydrophobic functional groups such as sodium 1,10-didecyl-1,10-diaza-18-crown-6 complex (Na-22DD) have been used successfully as carriers in the membrane separation and preconcentration of metal ions. Ligand substitution reactions take place at both membrane/solution interfaces in these cases. Figure 2B shows schematically a typical mechanism of facilitated liquid membrane separation process involving a cationic carrier ( $X^+$ ).

## Analytical Applications of Liquid Membranes

The analytical interest in BL and EL membranes is based mainly on their use as a convenient research

tool in screening carriers and determining the optimal composition of SL membranes. PL membranes, as a relatively new class of liquid membranes, have shown promising results in metal separation and preconcentration. Though there are numerous environmental and clinical applications of SL membranes, these liquid membranes have been used in a few instances only in food and beverage analyses (e.g., biogenic amines in wine, vanilla in chocolate and sugar, nicotine in Swedish snuff, and caffeine in coffee).

### Environmental Applications of SL Membranes

Most of these applications are in the area of water analysis for the determination of various pesticides, carboxylic and amino acids, aromatic compounds, surfactants, aniline derivatives, and metal ions. There are a few applications of SL membranes in soil and air analysis of carboxylic acids (Table 1).

### Clinical Applications of SL Membranes

Most of these applications are focused on the determination of heavy metals, drugs, and other biologically active compounds in blood plasma, breast milk, and urine (Table 2).

### Liquid Membrane Separation Units

Depending on the geometry of the solid polymer support, SL membranes can be produced in the form

**Table 1** Analytes determined in environmental samples after SLM separation

Amino acids	Tryptophan, phenylalanine, tyrosine
Aniline and its derivatives	Aniline, chloroanilines, methylanilines, chloromethylanilines
Aromatic compounds	Nitrophenols, chlorinated phenols, phenolic acids (benzoic, <i>p</i> -hydroxybenzoic, vanillic, caffeic, <i>p</i> -coumaric, ferulic, salicylic, phenazine-1-carboxylic), aromatic anionic surfactants (alkylbenzenesulfonates)
Carboxylic acids	Monocarboxylic acids (formic, acetic, propionic, butyric, iso-butyric, valeric, iso-valeric), dicarboxylic acids (glutaric, dimethylglutaric, adipic)
Metal ions	Cu(II), Pb(II), Cd(II), Zn(II), Ni(II), Al(III), Cr(III), Cr(VI)
Pesticides	Phenoxy acids (2,4-dichlorophenoxy acetic acid, MCPA, and 2,4,5-trichlorophenoxy acetic acid) and similar acid herbicides (benzaton, dicamba, dichlorprop, and mecoprop); sulfonurea herbicides (chlorsulfuron, metsulfuron methyl, tribenuron methyl, and thifensulfuron methyl); triazine herbicides (methoxy- <i>s</i> -triazines, chloro- <i>s</i> -triazines, alkylthio- <i>s</i> -triazines); phenylurea herbicides (fenuron, monolinuron, and diuron), metabolites of the fungicide thiophanate-methyl (carbenazim and 2-aminobenzimidazole)

**Table 2** Analytes determined in biological fluids after SLM separation

Biologically active compounds	Aliphatic and aromatic amines (methyl-, ethyl-, propyl-, and octyl-amines, piperidine, 1-ethylpiperidine, <i>N</i> -methylmorpholine, cyclohexylamine, <i>N,N</i> -dimethylcyclohexylamine, pyrrole, pyrrolidine, pyridine, aniline); phenols (phenol, <i>p</i> -cresol, 4-chlorophenol)
Drugs	Amperozide, amphetamine, methamphetamine, bambuterol, antidepressants (paroxetine, fluvoxamine, mianserin, and citalopram), anabolic steroids (glucuronides); promethazine, methadone, haloperidol, ibuprofen, naproxen, ketoprofen, propofol, ropivacaine
Metal ions	Pb(II), Mn(II)

of flat sheets or hollow tubes. So far PL membranes have been produced only as flat sheets.

### Flat Sheet Membrane Units

Flat sheet membrane units incorporating SL membranes have been used for both offline and online analytical separation.

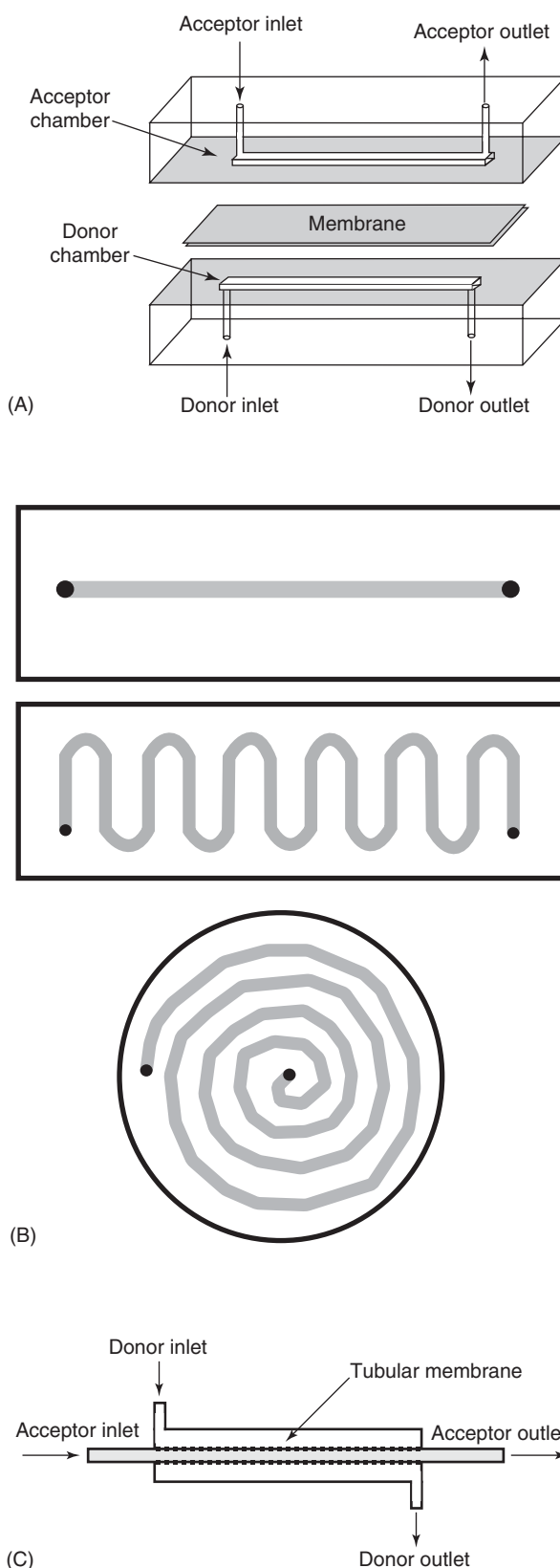
Offline separation can be conducted in a system consisting of two mechanically stirred extraction cells (Figure 1A) attached on both sides of the flat liquid membrane. One of the cells is filled with the sample solution while the other is filled with the receiver solution. Similar extraction systems have also been used in determining the permeability of SL and PL membranes and in studying the extraction mechanism and kinetics involved in membrane separation processes.

A typical flow-through membrane unit allowing online separation usually consists of two halves made of a chemically inert polymer material (e.g., PTFE, PVDF), each furnished with a shallow groove. The two halves are separated by a flat liquid membrane. The groove connected to the feed (donor) solution stream is called the donor chamber, while the other groove is called the acceptor chamber (Figure 3A). The grooves can be linear, meander shaped, or arranged as an Archimedes' spiral (Figure 3B). Typically the groove width and depth can vary in the range from 1 to 3 mm and from 0.1 to 0.5 mm, respectively. Though the two grooves must face each other, they can differ in depth. In most cases, however, they have identical dimensions.

The donor and acceptor solutions usually flow concurrently through the membrane separation unit (continuous mode). Higher efficiency can be achieved by decreasing the acceptor solution flow rate with respect to the donor solution flow rate, or by stopping the acceptor solution for a predetermined period of time to trap the analyte into the stagnant acceptor solution located in the acceptor chamber (stopped-flow mode). The acceptor solution can be flown through a suitable flow-through detector for measuring the analyte concentration or it can be analyzed offline. In many sample clean-up applications the acceptor solution is introduced into a high-performance liquid chromatograph or a gas chromatograph for further separation.

### Tubular Membrane Separation Units

A typical flow-through tubular membrane separation unit frequently consists of two concentric tubes (Figure 3C) and usually operates in a stopped-flow mode. The acceptor solution is located in the inner



**Figure 3** Schematic of: (A) flow-through membrane separation unit for flat sheet membranes; (B) top view of one of the unit's halves with linear, meander, and Archimedes' spiral groove; and (C) tubular flow-through membrane separation unit.

tube, which in most cases is a single porous hollow fiber impregnated with a suitable organic solvent (e.g., 1-octanol, 6-undecanone). The sample (donor) solution flows through the outer tube made of polymer material (e.g., Kel-F) or glass. It is also possible to immerse the hollow fiber SL membrane in a vial containing sample solution, which may be mechanically stirred. A tubular membrane separation unit incorporating a hollow fiber SL membrane connected to the needle of a microsyringe can be operated as an extracting syringe.

Hollow fiber SL membranes allow faster analyte separation and higher enrichment factors compared to flat sheet membranes because of their high surface area to solvent volume ratios. Solvent volumes of the order of several microliters can be employed, which allows the coupling of hollow fiber SL membrane separation to other low sample capacity separation techniques (e.g., packed capillary liquid chromatography, capillary electrophoresis).

*See also:* **Extraction:** Solvent Extraction Principles; Solid-Phase Extraction; Solid-Phase Microextraction. **Flow Injection Analysis:** Principles; Instrumentation. **Ion Exchange:** Principles. **Ion-Selective Electrodes:** Liquid Membrane; Gas Sensing Probes; Enzyme Electrodes. **Membrane Techniques:** Dialysis and Reverse Osmosis; Ultrafiltration; Pervaporation. **Solvents.**

## Further Reading

- Araki T and Tsukube H (1990) *Liquid Membranes: Chemical Applications*. Boca Raton, FL: CRC Press.
- Bartsch RA and Way JD (1996) *Chemical Separation with Liquid Membranes*. ACS Symposium Series 642. Washington: ACS.
- Dean JR (1998) *Extraction Methods for Environmental Analysis*. Chichester: Wiley.
- Fang Z (1993) *Flow Injection Separation and Preconcentration*. Weinheim: VCH.
- Jönsson JÅ and Mathiasson L (2000) Membrane-based techniques for sample enrichment. *Journal of Chromatography A* 902: 205–225.
- Jönsson JÅ and Mathiasson L (2001) Membrane extraction in analytical chemistry. *Journal of Separation Science* 24: 495–507.
- Kolev SD, Sakai S, Catrall RW, Paimin R, and Potter ID (2000) Theoretical and experimental study of palladium(II) extraction from hydrochloric acid solutions into Aliquat 336/PVC membranes. *Analytica Chimica Acta* 413: 241–246.
- Noble DN and Way JD (1987) *Liquid Membranes: Theory and Applications*. ACS Symposium Series 347. Washington: ACS.
- Pedersen-Bjergaard S, Rasmussen KE, and Halvorsen TG (2000) Liquid–liquid extraction procedures for sample enrichment in capillary zone electrophoresis. *Journal of Chromatography A* 902: 91–225.
- Valcarcel M and Luque de Castro MD (1991) *Non-Chromatographic Continuous Separation Techniques*. Cambridge: The Royal Society of Chemistry.

## Pervaporation

**A Jurado-López and M D Luque de Castro**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Pervaporation is a nonchromatographic membrane-based separation technique extensively used in industry. The separation is not based on relative volatilities, as in the case of distillation or evaporation, but rather on the relative rates of permeation through the polymeric membrane. It is based on the diffusion of volatile substances from a donor stream, through a porous hydrophobic membrane, to an acceptor stream caused by a temperature or pressure gradient.

Analytical pervaporation can be defined as a combination of continuous evaporation and gas diffusion through a gas permeable membrane. Both processes

take place in a single step and in the same unit. The volatile analyte (or its volatile reaction product) present in a heated donor phase evaporates through a porous membrane; then, an acceptor stream collects it and, finally, it is conveniently detected. Between the sample in the donor chamber and the membrane there is an air gap, which avoids the clogging of the membrane pores when analyzing dirty samples, and permits the presence in the donor chamber of species such as high molecular weight components, acids, bases, organic solvents, which would have deteriorated the membrane if they got in contact with it. This fact is the most outstanding feature of analytical pervaporation versus its industrial counterpart and other membrane-based nonchromatographic techniques such as dialysis and gas diffusion.

Analytical pervaporation can be carried out in a continuous way if the sample reaches the pervaporator by means of a continuous manifold, or in a discrete mode if the sample is introduced by injection,

tube, which in most cases is a single porous hollow fiber impregnated with a suitable organic solvent (e.g., 1-octanol, 6-undecanone). The sample (donor) solution flows through the outer tube made of polymer material (e.g., Kel-F) or glass. It is also possible to immerse the hollow fiber SL membrane in a vial containing sample solution, which may be mechanically stirred. A tubular membrane separation unit incorporating a hollow fiber SL membrane connected to the needle of a microsyringe can be operated as an extracting syringe.

Hollow fiber SL membranes allow faster analyte separation and higher enrichment factors compared to flat sheet membranes because of their high surface area to solvent volume ratios. Solvent volumes of the order of several microliters can be employed, which allows the coupling of hollow fiber SL membrane separation to other low sample capacity separation techniques (e.g., packed capillary liquid chromatography, capillary electrophoresis).

*See also:* **Extraction:** Solvent Extraction Principles; Solid-Phase Extraction; Solid-Phase Microextraction. **Flow Injection Analysis:** Principles; Instrumentation. **Ion Exchange:** Principles. **Ion-Selective Electrodes:** Liquid Membrane; Gas Sensing Probes; Enzyme Electrodes. **Membrane Techniques:** Dialysis and Reverse Osmosis; Ultrafiltration; Pervaporation. **Solvents.**

## Further Reading

- Araki T and Tsukube H (1990) *Liquid Membranes: Chemical Applications*. Boca Raton, FL: CRC Press.
- Bartsch RA and Way JD (1996) *Chemical Separation with Liquid Membranes*. ACS Symposium Series 642. Washington: ACS.
- Dean JR (1998) *Extraction Methods for Environmental Analysis*. Chichester: Wiley.
- Fang Z (1993) *Flow Injection Separation and Preconcentration*. Weinheim: VCH.
- Jönsson JÅ and Mathiasson L (2000) Membrane-based techniques for sample enrichment. *Journal of Chromatography A* 902: 205–225.
- Jönsson JÅ and Mathiasson L (2001) Membrane extraction in analytical chemistry. *Journal of Separation Science* 24: 495–507.
- Kolev SD, Sakai S, Catrall RW, Paimin R, and Potter ID (2000) Theoretical and experimental study of palladium(II) extraction from hydrochloric acid solutions into Aliquat 336/PVC membranes. *Analytica Chimica Acta* 413: 241–246.
- Noble DN and Way JD (1987) *Liquid Membranes: Theory and Applications*. ACS Symposium Series 347. Washington: ACS.
- Pedersen-Bjergaard S, Rasmussen KE, and Halvorsen TG (2000) Liquid–liquid extraction procedures for sample enrichment in capillary zone electrophoresis. *Journal of Chromatography A* 902: 91–225.
- Valcarcel M and Luque de Castro MD (1991) *Non-Chromatographic Continuous Separation Techniques*. Cambridge: The Royal Society of Chemistry.

## Pervaporation

**A Jurado-López and M D Luque de Castro**, University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Pervaporation is a nonchromatographic membrane-based separation technique extensively used in industry. The separation is not based on relative volatilities, as in the case of distillation or evaporation, but rather on the relative rates of permeation through the polymeric membrane. It is based on the diffusion of volatile substances from a donor stream, through a porous hydrophobic membrane, to an acceptor stream caused by a temperature or pressure gradient.

Analytical pervaporation can be defined as a combination of continuous evaporation and gas diffusion through a gas permeable membrane. Both processes

take place in a single step and in the same unit. The volatile analyte (or its volatile reaction product) present in a heated donor phase evaporates through a porous membrane; then, an acceptor stream collects it and, finally, it is conveniently detected. Between the sample in the donor chamber and the membrane there is an air gap, which avoids the clogging of the membrane pores when analyzing dirty samples, and permits the presence in the donor chamber of species such as high molecular weight components, acids, bases, organic solvents, which would have deteriorated the membrane if they got in contact with it. This fact is the most outstanding feature of analytical pervaporation versus its industrial counterpart and other membrane-based nonchromatographic techniques such as dialysis and gas diffusion.

Analytical pervaporation can be carried out in a continuous way if the sample reaches the pervaporator by means of a continuous manifold, or in a discrete mode if the sample is introduced by injection,

either with the aid of a valve or a syringe. The first option allows simplification and miniaturization of preliminary operations, which clearly improves analytical quality and productivity.

## The Analytical Pervaporator and Auxiliary Units

### The Pervaporator Body

A pervaporation module designed for analytical purposes, **Figure 1**, consists of the following parts:

1. An upper acceptor chamber fitted with inlet and outlet orifices through which the acceptor stream (liquid or gas) is circulated, and in which the pervaporated analyte (or its volatile reaction product) is collected.
2. A membrane support of small thickness.
3. Spacers of varying thickness, if necessary, to increase the volume of the corresponding chambers.
4. A donor chamber (lower part of the unit) where the feed stream is circulated.

The conventional pervaporation unit is usually made of methacrylate, which is a transparent material that allows continuous checking of the performance of the unit. The different parts of the pervaporator are aligned by the insertion of metallic

rods in the drilled orifices and a closer contact is achieved by screwing them together with four screws between two metallic supports. With the aim of integrate pervaporation with detection, some changes must be done on the acceptor chamber in order to place a probe sensor.

### Membranes

The separation of various components from a liquid mixture is not only determined by differences in their vapor pressure, but also by their permeation rate through the membrane. The mechanism of transport through the membrane can be described as comprising three consecutive steps:

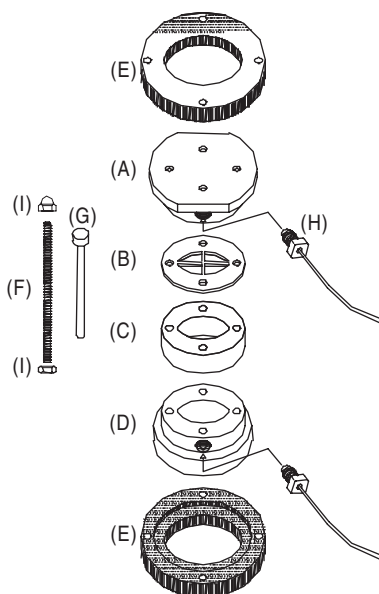
- sorption of the volatile components at the membrane surface;
- diffusion of the sorbed components through the polymer matrix; and
- evaporation from the polymer into the vapor phase on the permeate side of the membrane.

The efficiency of the pervaporation process is controlled mainly by the intrinsic properties of the polymers used for membrane preparation. It has been demonstrated that permeability in glassy polymers (cellulose) is lower than that in rubbery polymers (polydimethylsiloxane). Moreover, as the contribution of the solubility to the permeability dominates in nonglassy polymers, the permeability increases with increasing molecular mass of permeants.

Selectivity depends on the molecular dimensions of the permeating species. Thus, the hydrophobic membranes for analytical pervaporation are usually made of polytetrafluoroethylene and they are similar to those used in processes like ultrafiltration and gas diffusion.

### The Auxiliary Manifold

The simplest way of giving pervaporation the character of a continuous separation technique is its coupling to a dynamic manifold for assistance of both donor and acceptor chambers. In the first instance, when the samples are liquid, the coupling with the manifold (usually a flow-injection (FI) arrangement) is mandatory for driving the sample either by injection or aspiration to the donor chamber. In addition, (bio)chemical and/or physical steps, namely, reactions that convert the analyte into the most appropriate form for being evaporated, physical dispersion, etc., can also be developed in the manifold prior to the arrival of the sample to the donor chamber; meanwhile, a detector can be located in postpervaporator position in order to monitor nonvolatile species. When the sample is a solid, the



**Figure 1** Parts of a pervaporation module: (A) acceptor chamber; (B) membrane support; (C) spacer; (D) donor chamber; (E) aluminum supports; (F) and (G) rods for screwing and aligning the module, respectively; (H) connectors; (I) screws. (Reprinted with permission from De Mattos *et al.* (1995) Pervaporation: an integrated. *Talanta*, 42: 757; © Elsevier.)



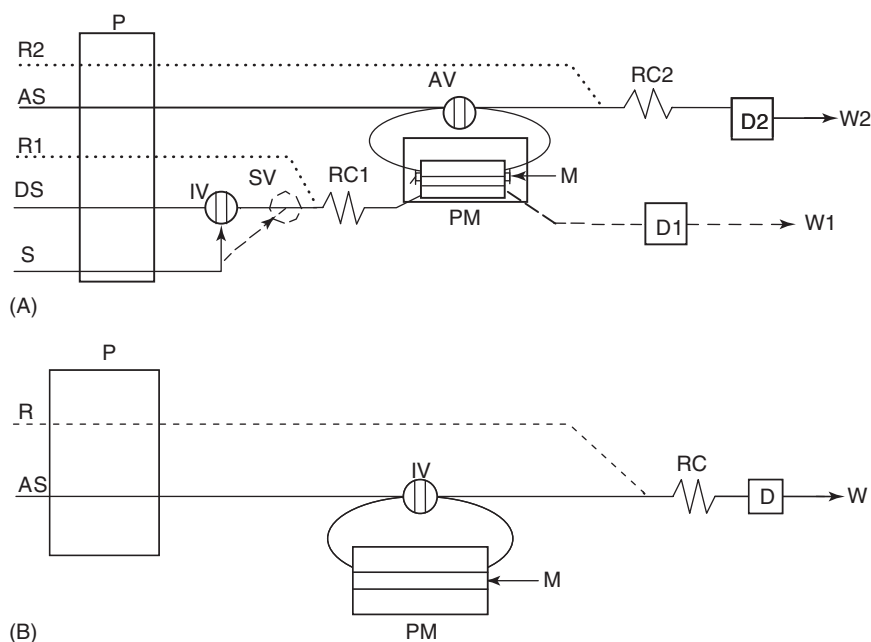
donor chamber works as a multifunctional device in which the sample is weighed and, after adjustment to other parts of the pervaporator, leaching, masking, and (bio)chemical derivatizing reactions take place in this chamber. Concerning the acceptor chamber, it also requires being inline coupled to a dynamic manifold for driving the pervaporated species to the detector. Even when the detector is integrated in this chamber, coupling with the continuous manifold is desirable in order to clean the chamber between input of samples. Derivatization reactions, preconcentration steps, etc., also require the continuous manifold–upper chamber arrangement. Discrimination between species can be based on this coupling.

The characteristic versatility of FI manifolds endows pervaporation with a very valuable aspect in separation techniques as preconcentration capability is. The location of the upper chamber in the loop of a conventional low-pressure injection valve enables the acceptor fluid to remain static during pervaporation by keeping this valve in its filling position. After a preset time, long enough for sufficient enrichment of the acceptor fluid in the pervaporated analyte, the valve is switched to the injection position and the stream leads the content of the loop to the detector. During the enrichment time, the acceptor fluid is continuously reaching the detector, thus establishing the baseline. When the detector is integrated with the

pervaporator through a probe (located at the top of the acceptor chamber, faced to the membrane), the presence of this valve allows monitoring of the mass transfer through the membrane, i.e., monitoring of the kinetics of the pervaporation process. Both types of dynamic manifold pervaporator for liquid and solid samples are shown in Figure 2.

## Pervaporation Efficiency

The efficiency of a pervaporation process can be done in relative terms, i.e., comparing the signals provided by the analyte (or its reaction product) under different experimental conditions, or in absolute terms, comparing the signal obtained under the working conditions with that corresponding to 100% mass transfer of the analyte. The first procedure is usually used in optimization experiments, and it does not require special alterations of the manifold. On the other hand, the absolute procedure requires the use of an auxiliary channel that directly leads the same sample volume as that used in the pervaporation process to the detector, in order to measure the signal corresponding to the whole amount of analyte, and compare it with the signal obtained from the channel coming from the acceptor chamber of the pervaporator, with the collected analyte. The comparison through the peak height is not valid as



**Figure 2** Continuous–discrete approach to implementing analytical pervaporation. The dotted line corresponds to a potential derivatization reaction of the pervaporated species, while the dashed lines represent the continuous manifold used for automatic insertion of liquid samples. P, peristaltic pump; AS, acceptor stream; IV, injection valve; SV, switching valve for changing between continuous and discrete insertion of sample into the pervaporator; R, reagent; DS, donor-sample stream; S, sample; RC, reaction coil; PM, pervaporation module; M, membrane; D, detector; W, waste. (Reproduced with permission from Freund Publishing House Ltd.)

the dispersion of the two plugs (that injected and that from the acceptor chamber) is different. Instead, comparison between peak areas must always be used.

## Variables Influencing Analytical Pervaporation

There are a series of variables influencing on the mass transfer process, which can be divided into two groups, depending on which part of the system is involved.

### Variables Pertaining to the Donor Chamber

**Temperature** An increase in temperature in the donor chamber, carried out either by immersing it into a water bath or with focused microwaves, increases the vapor pressure of the analyte and thus the mass transfer process.

**Sample agitation** Magnetic stirring helps the removal of gases from the donor stream, and so do ultrasounds in a more intense way; however, they are not recommended because of causing leakage and so losses of the gas phase.

The presence of chemically inert beads of a proper size in the donor chamber boosts the transfer of volatile species into the air gap, with both liquid and solid samples because the residence time, in the first instance, is increased due to the tortuous path that the liquid has to go through, and the beads separate the particles of solids, creating a larger surface for the removal of the gas from the sample to the air gap.

**Volume of the air gap** The smaller the gap is the smaller is the amount of analyte required to go through the membrane. The best way of decreasing the volume of the air gap (in liquid samples) is lengthening the waste line because it causes an overpressure in the donor chamber, which raises the level of the solution.

**Flow rate** In liquid samples, the flow rate affects the mass transfer process because the lower it is the longer is the residence time of the analyte in the lower part of the pervaporation unit. The flow can be stopped to allow steady-state mass transfer, but it decreases the sampling frequency, so a compromise between sensitivity and expediency is needed.

**Concentration of reagents** If the measurable volatile substance is a reaction product of the analyte of interest, the concentration of reagents and pH have a dramatic effect on the pervaporation process.

### Variables Pertaining to the Acceptor Chamber or Auxiliary Manifold

**Flow rate** As said before, the flow rate of the acceptor stream strongly affects the mass transfer equilibrium. The continuous use of fresh acceptor fluid maximizes the concentration gradient between the donor and the acceptor chamber. However, if keeping the acceptor fluid static during pervaporation is required, it has to be achieved by accommodating the acceptor chamber in the loop of an injection valve, as commented before. In the filling position, the acceptor stream passes through the bypass to the detector at a preset flow rate. Once the collection time has elapsed, the valve is switched to its load position and the stream passes through the loop to sweep its contents to the detector. The use of a valve is necessary because the direct stopping of the propulsion unit for the acceptor stream disturbs the detector baseline through the start-and-stop sequence, which in turn causes part of the fluid to return to the propulsion unit.

**Preconcentration of pervaporated analytes** The equilibrium between phases can be efficiently displaced by the continuous removal of the pervaporated species from the acceptor chamber, followed by their concentration, either in a minicolumn (packed with suitable material such as an ion exchanger or a sorbent) placed before the detector or in the flow-cell itself.

**Features of the membrane** The membrane thickness and pore size have a strong influence on the pervaporation process. Permeability is inversely proportional to the thickness of the membrane and directly proportional to its pore density. An increased membrane pore size can be expected to require a decreased pore density to ensure acceptable mechanical stability. As a result, the overall area of the pores through which gas diffusion takes place could even decrease with increasing pore size and lead to a lower permeability.

## The Physics of Analytical Pervaporation: Kinetics of the Mass Transfer Process

Monitoring the mass transfer kinetics is a key aspect in order to optimize pervaporation experiments. The main goals of the optimization are the identification of the influencing factors and the establishment of an appropriate pervaporation time to ensure a high throughput without an appreciable decrease in sensitivity. The kinetics of the mass transfer can be

monitored in two ways: (1) integrating pervaporation and detection in a single module, and (2) integrating, after the pervaporation process, a retention step with detection. The first option can be carried out placing either a probe-type sensor, electroanalytical or optical, inside the acceptor chamber, with the active surface faced to the membrane. The second way is achieved with a flow cell packed with a suitable material (ion exchanger or sorbent) and placed inside a nondestructive detector that quantifies the continuous retention of the pervaporated analyte or its reaction product.

## Analytical Features of Pervaporation Methods

The analytical features of pervaporation, concerning the basic analytical properties, namely, sensitivity, selectivity, and precision, strongly rely on experimental variables. Sometimes, a given variable can have an opposite effect on two of the analytical properties, so a compromise between them is needed.

### Sensitivity

The sensitivity of a method involving pervaporation depends mainly on the efficiency of the separation step. The efficiency of pervaporation can be adjusted in order to fit the signal obtained from the analyte or its reaction product to the linear portion of the calibration curve for a given method, thus avoiding the need for a dilution or concentration step.

The analytical signal can be optimized for the sake of quantitation acting on the influencing variables already mentioned: increasing the temperature of the donor chamber, stirring the sample, using chemically inert beads, employing optimal reagent concentrations, increasing the pervaporation time, using an efficient procedure for concentrating the volatile species, and increasing the amount of sample used.

When the concentrations of the target analytes in the sample exceed the upper limit of the linear range of the calibration curve, a dilution or pseudodilution step is mandatory in order to fit the unknown concentration to this portion of the calibration curve, and hence to increase the precision of the measurements. A series of alternatives to the usual previous dilution of the sample have been developed in order to avoid the errors involved in a dilution step or in the weighing of small amounts of a solid. These alternatives are: (1) using of a lower pervaporation temperature; (2) changing the chemical conditions to a less favorable situation if a derivatization reaction is required, in order to reduce the yield of the monitored species; (3) enlarging the air gap between the

sample and the membrane by the addition of the suitable number of spacers; (4) using a thicker membrane; (5) using a smaller loop for the injection valve (if the sample is a liquid); or (6) increasing the flow rate or the donor and/or the acceptor stream.

### Selectivity

The fact that pervaporation can only be used to separate volatile analytes may initially seem a shortcoming rather than an advantage; however, it results in increased selectivity for the target analytes and enables their determination in complex samples such as biological fluids and environmental specimens with no prior treatment, using pervaporation as a cleanup step. If the target analyte is not volatile but can be converted into a volatile substance by means of a selective (bio)chemical reaction, it can also be separated by pervaporation. The possibility of employing the derivatization reaction prior or after the separation step has extended the scope of pervaporation to a variety of analytes and allowed pervaporation units to be coupled to various detector types. Pervaporation can also be used for speciation analysis as it allows the selective determination of different forms of the target analyte.

In multideterminations, and also when the target volatile analytes possess different boiling points, the pervaporation unit can be used at different temperatures to achieve selective separation of the species of interest.

Discrimination between polar and nonpolar volatile analytes can be accomplished by using microwave irradiation of the donor chamber in order to promote the evaporation of polar species. Conventional heating can then evaporate nonpolar ones.

Selectivity can also be improved by changing the membrane used in the pervaporation process, owing to the fact that this property depends on the molecular dimensions of the permeated species.

### Precision

The fact that the placement or the passage of the sample in or through the donor chamber, respectively, which is quite reproducible by means of stirring or the use of chemically inert beads, results in highly reproducible signals following pervaporation.

Repeatability can also be improved by aspirating the detector waste at a flow rate identical to that of the acceptor stream. If a chemical reaction with an additional reagent is required, the flow rate of the detector waste has to be the same as the combined individual flow rates.

The use of long waste tubing in the lower part of the pervaporation module can also improve the

precision of the method since the pressure in the module increases and thus small variations in pressure caused by both internal and external noise have a smaller effect.

The use of gas diffusion membranes instead of ultrafiltration membranes can also improve the precision because the former are more resistant to bending than the latter; this avoids changes in the permeate flux resulting from both irreproducible membrane areas and increased volumes of the upper chamber.

## Scope of Application of Analytical Pervaporation

Analytical pervaporation can be used for the determination of one or more analytes. In multideterminations, analytes can be quantified sequentially or in a simultaneous way.

### Sample Types

The versatility of the design of an analytical pervaporation module, attributable to its changeable donor volume and the air gap present above the sample, has enabled its use with liquid, solid, and semisolid samples. **Table 1** summarizes the applications of analytical pervaporation and the main fields in which it has been used. The manifold used and the way in which the sample is introduced in the system vary depending on whether it is liquid or solid. The main difference between the treatment of liquid and solid samples is the fact that with liquids the whole pervaporation process, from sample introduction to delivery of the results can be fully automated, while some operator's intervention is needed when dealing with solids.

### Derivatization Reactions

The development of chemical or biochemical reactions widens the scope of analytical pervaporation, making possible the determination of nonvolatile species. These reactions can take place in both the donor and the acceptor chamber. In the first case the aim of the reaction is to achieve a volatile species from a nonvolatile analyte either injecting appropriate reagents or merging the donor and the reagents stream; in the second instance the objective is to adapt the monitored product to the characteristics of the detector and the most effective manner of doing so is using the solution containing the reagents as acceptor solution. In this way, the pervaporated analytes react on the acceptor side of the membrane, thus displacing the mass transfer equilibrium. Solid reagents, which are very common in enzyme-catalyzed reactions, are packed in a tube

placed between the injection valve and the donor chamber of the pervaporator.

### Integrated Pervaporation and Detection: Detector Types

The best way of monitoring pervaporated species is by inserting a probe-type sensor in the acceptor chamber with its active side facing the membrane. The most salient advantages of integrating pervaporation and detection are as follows:

1. The response time is shortened with respect to the conventional location of the sensor behind the separation module because there is no need to transport the analytes to the detector.
2. The kinetics of the mass transfer across the membrane can be monitored, obtaining a better understanding of the pervaporation process and an easier optimization.
3. The system is dramatically miniaturized, as the leaching, derivatization, separation, and detection steps take place in a single module.

Pervaporators are amenable to coupling to any type of detector via an appropriate interface such as a transport tube, a microcolumn packed with adsorptive or ion-exchange material, or a gas liquid separator. The acceptor stream can be either liquid or gaseous depending on the characteristics of the detector. The detectors most frequently used are the spectroscopic – atomic or molecular, electroanalytical (potentiometric, voltammetric), electron capture, and flame ionization types. The low selectivity of some of these detection techniques is overcome by that of the pervaporation step, endowing the overall analytical process with the selectivity required for the analysis of complex matrices. The potential use of the pervaporation technique for sample insertion into water-unfriendly detectors such as mass spectrometers or devices such as those based on microwave-induced plasma remains unexplored.

### Coupling a Pervaporator to a Gas Chromatograph: An Alternative to Headspace Sampling

The coupling of a pervaporator to a gas chromatograph is one of the most promising uses of pervaporation and is worth a more detailed discussion, because of the advantages that pervaporation presents as compared with both static and dynamic headspace sampling techniques. In the static approach, the sample is placed in a closed chamber and heated until the volatile compounds in the headspace reach the equilibrium with the sample. Then, part of

**Table 1** Applications of analytical pervaporation in different fields

Analyte	Sample	Derivatization reaction	Detection	Comments
<i>Environmental samples</i>				
F <sup>-</sup>	Trees leaves, waste waters, fertilizers	Chemical	Potentiometry	Continuous/discontinuous manifold; stopped-flow
NH <sub>3</sub>	Industrial fluids	Chemical	Photometry	Determination in the presence of surfactants
NH <sub>3</sub> /urea	Soils	Biochemical	Photometry	Sequential determination; speciation analysis
S <sup>2-</sup>	Kraft liquors	Chemical	Photometry	Improved precision and sensitivity
CN <sup>-</sup>	Mining sample	Chemical	Amperometry	Determination in presence of S <sup>2-</sup>
Phenol	Water	Chemical	Amperometry	Stopped-flow
Diclorvos	Natural water, soils	Biochemical	Fluorimetry	Derivatization by enzymatic inhibition
Metrifonate	Natural water, soils	Biochemical	Fluorimetry	Integration of hydrolysis–pervaporation–derivatization
Hg	MRC		Atomic fluorimetry	Alternative to headspace
VOCs	Soils		ECD	Alternative to headspace
Hg compounds	MRC		Atomic fluorimetry	Microwave-assisted pervaporation
DQO/inorganic carbon	Waste water	Chemical	Photometry/potentiometry	Simultaneous determination
As <sup>III</sup> /As <sup>V</sup>	Dirty waters, soils	Chemical	Atomic fluorimetry	Arsenic speciation
<i>Food samples</i>				
Ethanol	Wines, beers	Biochemical	Fluorimetry, photometry, voltammetry	Immobilized enzyme
Acidity	Wines	Chemical	Photometry	Determination of volatile acidity
SO <sub>2</sub>	Wines	Chemical	Photometry	Free and combined SO <sub>2</sub> speciation
CO <sub>2</sub> /SO <sub>2</sub>	Wines	Chemical	Photometry/potentiometry	Sequential determination
Ethanol/glycerol	Wines	Chemical/biochemical	Photometry/fluorimetry	Simultaneous determination
Diacetyl	Beers	Chemical	Photometry	Preconcentration
Acetaldehyde	Fruit juices, yoghurt	Chemical	Photometry	Continuous monitoring
Acetaldehyde/acetone	Fruit juices, milk, yoghurt		FID	Alternative to headspace
Triethylamine	Fish	Chemical	Photometry	Pretreatment coupled to pervaporation
<i>Pharmaceutical and clinical samples</i>				
NH <sub>3</sub> /urea	Serum, urine	Biochemical	Potentiometry	Sequential determination
Oxalate	Urine, serum	Biochemical	Potentiometry	Continuous/discontinuous manifold
Formaldehyde	Pharmaceuticals, cosmetics	Chemical	Photometry	Stopped-flow
F <sup>-</sup>	Pharmaceuticals, cosmetics	Chemical	Potentiometry	Continuous/discontinuous manifold

the vapor phase is injected in the chromatograph for analysis. The dynamic headspace or purge and trap mode requires the continuous removal of the gas phase from the chamber. Owing to the fact that the separation of the volatile analytes from the sample is a slow process, an intermediate trap is needed in order to concentrate the analytes before their introduction in the chromatograph.

Pervaporation provides a number of advantages versus headspace techniques that can be summarized as follows:

1. The thin air gap above the sample requires very small amounts of the analytes to establish equilibrium with the sample and the mass transfer across the membrane.



2. The continuous removal of the volatilized analytes displaces the equilibrium and increases the separation efficiency.
3. Continuous removal of the pervaporated analytes to the preconcentration column, if used, allows fresh portions of acceptor gas to come into contact with the diffused species, thus displacing the mass transfer equilibrium.
4. The separation step can be totally or partially automated (with liquid or solid samples, respectively) with minimal purchase and maintenance costs.
5. Unlike purge and trap, no water vapor condenser is required, nor is a hydrophobic sorbent, as no water crossing through the hydrophobic membrane occurs at the usual working temperatures.

## Prospects of Analytical Pervaporation

Pervaporation is a useful analytical tool that features simplicity, automation, and miniaturization capabilities. The actual potential of this technique for the analysis of solid, liquid, and slurry samples can be inferred from its intrinsic features and from available methods using a pervaporator.

Consolidation of this technique as an effective, widespread analytical tool needs the gathering of sufficient research experience to compile a cookbook of methods, where potential users can search for solutions to specific problems, and the commercialization of inexpensive pervaporation modules meeting the requirements of a variety of samples and analytes. Once these two aspects are developed, pervaporation can become a useful tool for routine

environmental, pharmaceutical, food, and industrial analyses.

**See also:** **Amperometry.** **Derivatization of Analytes.** **Gas Chromatography:** Overview. **Headspace Analysis:** Static; Purge and Trap. **Membrane Techniques:** Dialysis and Reverse Osmosis; Ultrafiltration; Liquid Membranes. **Potentiometric Stripping Analysis.** **Titrimetry:** Potentiometric.

## Further Reading

- Luque de Castro MD and Papaefstathiou I (1998) Pervaporation: a useful tool for speciation analysis. *Spectrochimica Acta B* 53: 311–319.
- Luque de Castro MD and Luque-García JL (2002) *Acceleration and Automation of Solid Sample Treatment*, pp. 128–154. Amsterdam: Elsevier.
- Mataix E and Luque de Castro MD (2000) Simultaneous determination of ethanol and glycerol in wines by a flow injection-pervaporation approach with in parallel photometric and fluorimetric detection. *Talanta* 51: 489–496.
- Mattos IL and Luque de Castro MD (1994) Study of mass-transfer efficiency in pervaporation processes. *Analytica Chimica Acta* 298: 159–165.
- Priego-López E and Luque de Castro MD (2002) Pervaporation–gas chromatography coupling for slurry samples. Determination of acetaldehyde and acetone in food. *Journal of Chromatography A* 976: 399–407.
- Valcárcel M and Luque de Castro MD (1991) *Non-Chromatographic Continuous Separation Techniques*. Cambridge: Royal Society of Chemistry.
- Vallejo B, Richter P, Toral I, Tapia C, and Luque de Castro MD (2001) Determination of sulphide in liquid and solid samples by integrated pervaporation–potentiometric detection. *Analytica Chimica Acta* 436: 301–307.

# MERCURY

**M Horvat**, Jozef Stefan Institute, Ljubljana, Slovenia

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by M Horvat and W H Schroeder, pp. 3059–3070, © 1995, Elsevier Ltd.

## Introduction

Organomercury compounds may enter the environment from both anthropogenic sources and from production by natural *in situ* biogenic alteration of inorganic or other organomercury compounds.

Commercially used organomercury compounds are listed in **Table 1**. Environmental problems arising from use of commercial products are generally localized. Phenylmercury compounds are still used in exterior paints and as slimicides in the pulp and paper industry. However, their use for other purposes has now been banned by most governments. In some countries, organic mercury compounds are still used in agriculture (mainly phenylmercury, methoxymethylmercury, and small quantities of ethylmercury compounds). Organomercury compounds, often called thiomersal, are still used in medicine, mostly for preservation of vaccines. Although the

2. The continuous removal of the volatilized analytes displaces the equilibrium and increases the separation efficiency.
3. Continuous removal of the pervaporated analytes to the preconcentration column, if used, allows fresh portions of acceptor gas to come into contact with the diffused species, thus displacing the mass transfer equilibrium.
4. The separation step can be totally or partially automated (with liquid or solid samples, respectively) with minimal purchase and maintenance costs.
5. Unlike purge and trap, no water vapor condenser is required, nor is a hydrophobic sorbent, as no water crossing through the hydrophobic membrane occurs at the usual working temperatures.

## Prospects of Analytical Pervaporation

Pervaporation is a useful analytical tool that features simplicity, automation, and miniaturization capabilities. The actual potential of this technique for the analysis of solid, liquid, and slurry samples can be inferred from its intrinsic features and from available methods using a pervaporator.

Consolidation of this technique as an effective, widespread analytical tool needs the gathering of sufficient research experience to compile a cookbook of methods, where potential users can search for solutions to specific problems, and the commercialization of inexpensive pervaporation modules meeting the requirements of a variety of samples and analytes. Once these two aspects are developed, pervaporation can become a useful tool for routine

environmental, pharmaceutical, food, and industrial analyses.

**See also:** **Amperometry.** **Derivatization of Analytes.** **Gas Chromatography:** Overview. **Headspace Analysis:** Static; Purge and Trap. **Membrane Techniques:** Dialysis and Reverse Osmosis; Ultrafiltration; Liquid Membranes. **Potentiometric Stripping Analysis.** **Titrimetry:** Potentiometric.

## Further Reading

- Luque de Castro MD and Papaefstathiou I (1998) Pervaporation: a useful tool for speciation analysis. *Spectrochimica Acta B* 53: 311–319.
- Luque de Castro MD and Luque-García JL (2002) *Acceleration and Automation of Solid Sample Treatment*, pp. 128–154. Amsterdam: Elsevier.
- Mataix E and Luque de Castro MD (2000) Simultaneous determination of ethanol and glycerol in wines by a flow injection-pervaporation approach with in parallel photometric and fluorimetric detection. *Talanta* 51: 489–496.
- Mattos IL and Luque de Castro MD (1994) Study of mass-transfer efficiency in pervaporation processes. *Analytica Chimica Acta* 298: 159–165.
- Priego-López E and Luque de Castro MD (2002) Pervaporation–gas chromatography coupling for slurry samples. Determination of acetaldehyde and acetone in food. *Journal of Chromatography A* 976: 399–407.
- Valcárcel M and Luque de Castro MD (1991) *Non-Chromatographic Continuous Separation Techniques*. Cambridge: Royal Society of Chemistry.
- Vallejo B, Richter P, Toral I, Tapia C, and Luque de Castro MD (2001) Determination of sulphide in liquid and solid samples by integrated pervaporation–potentiometric detection. *Analytica Chimica Acta* 436: 301–307.

# MERCURY

**M Horvat**, Jozef Stefan Institute, Ljubljana, Slovenia

© 2005, Elsevier Ltd. All Rights Reserved.

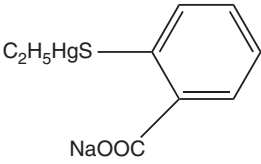
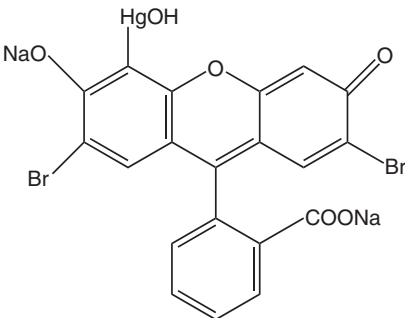
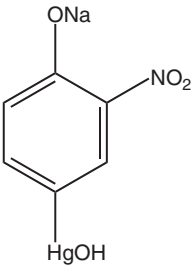
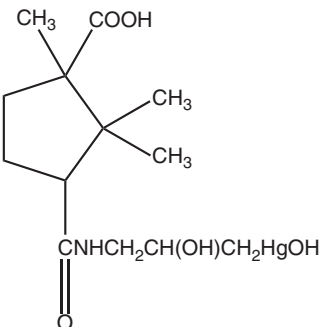
This article is a revision of the previous-edition article by M Horvat and W H Schroeder, pp. 3059–3070, © 1995, Elsevier Ltd.

## Introduction

Organomercury compounds may enter the environment from both anthropogenic sources and from production by natural *in situ* biogenic alteration of inorganic or other organomercury compounds.

Commercially used organomercury compounds are listed in **Table 1**. Environmental problems arising from use of commercial products are generally localized. Phenylmercury compounds are still used in exterior paints and as slimicides in the pulp and paper industry. However, their use for other purposes has now been banned by most governments. In some countries, organic mercury compounds are still used in agriculture (mainly phenylmercury, methoxymethylmercury, and small quantities of ethylmercury compounds). Organomercury compounds, often called thiomersal, are still used in medicine, mostly for preservation of vaccines. Although the

**Table 1** Commercial use of organomercury compounds

Compound	Use	Comments
$\text{CH}_3\text{HgX}$	Agricultural seed dressing, fungicide	Banned in Sweden 1966, USA in 1970, as seed disinfectant. Not used today in Europe or US. Used only in laboratories
$\text{C}_2\text{H}_5\text{HgX}$ $\text{RHgX}$	Cereal seed treatment Catalyst for urethane, vinyl acetate production	Banned in USA and Canada in 1970. Used in UK
$\text{C}_6\text{H}_5\text{HgX}$	Seed dressing, fungicide, slimicide, general bactericide. For pulp, paper, paints	Banned as slimicide in USA in 1970. Banned as seed dressing in Japan in 1970. Used in UK
$p\text{-CH}_3\text{C}_6\text{H}_4\text{HgX}$ $\text{ROCH}_2\text{CH}_2\text{HgX}$ $\text{ClCH}_2\text{CH}(\text{OCH}_3)\text{CH}_2\text{HgX}$	Spermicide Seed dressings, fungicides Fungicide, pesticide, preservative	Banned in Japan in 1968. Used in UK
Thimerosal	Antiseptic, $\text{C}_2\text{H}_5\text{Hg}$ derivative	
Mercurochrome	Antiseptic, organomercury fluorescein derivative	
Mersalyl	Diuretic, methoxyalkyl derivative $\text{RCH}_2\text{CH}(\text{OCH}_3)\text{CH}_2\text{HgX}$ $\text{R} = o\text{-COOHCH}_2\text{OC}_6\text{H}_4\text{CONH-}$	Little used today
Chlormerodrin	Diuretic, methoxyalkyl derivative $\text{NH}_2\text{CONHCH}_2\text{CH}(\text{OCH}_3)\text{CH}_2\text{HgCl}$	Little used today
Mercarbolid Mercurophen	$o\text{-HOC}_6\text{H}_4\text{HgCl}$ $m\text{-NO}_2\text{-}p\text{-ONaC}_6\text{H}_3\text{HgOH}$	$o\text{-Chloromercuriphenol}$ 
Mercurophylline	Diuretic	

X = anionic group. Wide range of X known, e.g.,  $\text{OAc}^-$ ,  $\text{PO}_4^{3-}$ ,  $\text{Cl}^-$ ,  $\text{NHC}(\text{NH})\text{NHCN}^-$ , etc.

Adapted from Craig PJ (1986) *Organometallic Compounds in the Environment – Principles and Reactions*. London: Longman.

safety concerns are mainly theoretical, the perception of risk remains very high and has led to initiatives to eliminate, reduce, or replace thiomersal in all vaccine presentations. In general, however, production and usage of organomercurials is declining.

At present, most studies are focused toward naturally formed organomercury compounds. By far the most common of these are the monomethylmercury compounds. The term 'methylmercury' is used throughout this text to represent monomethylmercury compounds. In many cases the complete identity of these compounds is not known except for the monomethylmercury cation,  $\text{CH}_3\text{Hg}^+$ , which is associated either with a simple anion, like chloride, or a large charged molecule (e.g., a protein).

Many of the environmental aspects of mercury and its compounds have been reviewed. During recent years, new analytical techniques, which have become available, have been used in environmental studies and the understanding of mercury chemistry in natural systems has improved significantly. Mercury can exist in a large number of different physical and chemical forms with a wide range of properties. Conversion between these different forms provides the basis for mercury's complex distribution pattern for local and global cycles, and for its biological enrichment and effects. The most important chemical forms are: elemental mercury ( $\text{Hg}^\circ$ ), divalent inorganic mercury ( $\text{Hg}^{2+}$ ), methylmercury ( $\text{CH}_3\text{Hg}^+$ ), and dimethylmercury ( $(\text{CH}_3)_2\text{Hg}$ ).

Once methylmercury is formed it enters the food chain by rapid diffusion and firm binding to proteins in aquatic biota and attains its highest concentrations in the tissues of fish at the top of the aquatic food chain due to biomagnification through the trophic levels. The main factors that affect the levels of methylmercury in fish are the diet/trophic level of the species, age of the fish, microbial activity and mercury concentration in the upper layer of the sediment, dissolved organic carbon (DOC) content, salinity, pH, and redox potential.

Extensive investigations have recently been conducted in Scandinavia and North America where it was discovered that the long-range transport of  $\text{Hg}^\circ$  and the potential role of acidification (through acid rain) have become major factors concerning future exposure of humans to methylmercury. As a result, methylmercury has exceeded regulatory safety limits in fish (0.5–1.0 mg per kg, fresh weight) in large numbers of remote lakes. Due to long-range transport and deposition high mercury concentrations were observed in communities in remote Northern environments that depend heavily upon mercury-contaminated local diets (fish, birds, marine mammals).

## Types of Compounds and Matrices

Concentrations of mercury in biological and environmental samples are relatively low except in mercury contaminated and industrial areas. Total mercury concentrations in some matrices are elsewhere in this encyclopedia. Only a very brief overview of the presence of organomercury compounds in various compartments of the environment is given below.

### Air

The major form of mercury in air is elemental mercury vapor ( $\text{Hg}^\circ$ ). Typical concentrations of mercury and its compounds in ambient air may range between 1 and 4  $\text{ng m}^{-3}$  for elemental mercury, from 1 to 50  $\text{pg m}^{-3}$  for reactive gaseous mercury (water-soluble gaseous mercury species) and total particulate mercury, whereas MeHg has been found in the range between 1 and 20  $\text{pg m}^{-3}$ . Volatile mercury(II) compounds represent a large part of total gaseous mercury in stack gases from coal burning power plants. Even though  $\text{Hg}^\circ$  and dimethylmercury are the most volatile forms of mercury, dimethylmercury has not unequivocally been detected in the atmosphere. Its presence was detected in gases from landfill waste sites. It is interesting to note that, although significant quantities of dimethylmercury may be emitted to the atmosphere, the lifetime of this species is short, owing to photochemical and/or chemical degradation to  $\text{Hg}^\circ$ . Reliable data on mercury species in air are still very limited and further analytical developments are needed.

### Water

Concentrations of total mercury in water samples are very low (in the nanogram per liter level or below), so that accurate analysis is still a major problem. The theoretical approach via stability calculations can be of great help in making rough estimates of the predominant mercury species under various conditions. Mercury compounds occurring in natural waters are most often defined by their ability to be reduced to elemental mercury. In lake waters methylmercury species account for 1–30% of total mercury. Most of the methylmercury is probably associated with dissolved organic carbon (DOC). Limited data are available on the formation constants between the methylmercury cation and DOC. Thiol groups ( $-\text{RSH}$ ) have been shown, however, to have a higher capability to bind methylmercury in comparison with ligands containing oxygen and nitrogen donor atoms and the inorganic ions ( $\text{CN}^-$ ,  $\text{Cl}^-$ ,  $\text{OH}^-$ ). Methylmercury compounds in surface runoff waters,

soil pore waters, and groundwaters are similar to the species in lake waters and are generally quite strongly associated with DOC. Dimethylmercury has rarely been reported in surface waters except in deep oceans and during some seasons in the slurry of salt marches. Mercury in seawater exists mainly in the form of  $\text{Hg}^{2+}$  complexed with  $\text{Cl}^-$  ions. Methylmercury concentrations in seawater are generally lower than in lake waters. Dissolved gaseous mercury vapor is also present in ocean waters.

### Sediments and Soils

Mercury in these two compartments of the environment is mainly associated with humic matter. Following methylation, methylmercury does not usually build up in sediments to more than  $\sim 1.5\%$  of the total mercury present. This appears to be an approximate equilibrium level between formation and removal. Methylation–demethylation reactions are assumed to be widespread in the environment and each ecosystem attains its own steady state with respect to the individual species of mercury. Dimethylmercury is considered to be unstable in sediments and is assumed to be stabilized by a conjunction of factors, such as high sulfide levels, salinity, anoxic conditions, and constant inputs of methane into the media. It is also important to note that sampling of sediment samples from anoxic/suboxic bottoms must be conducted with extreme care, as, in contrast, changes of environmental conditions (T, redox potential) during sampling may significantly influence the concentrations of mercury species and partitioning (see Sample Handling and Storage).

### Biota

The highest levels found in both fresh and marine organisms are found at the highest trophic levels, where mercury levels can exceed the ‘black list limit’ of 0.5 mg per kg. The percentage of methylmercury to total mercury in fish muscle varies from 80% to 100%, but in other organs its concentration is smaller (in liver and kidney up to 20%). In other aquatic organisms, the percentage of methylmercury is more variable, depending on water depth, locations, and the type of organism. Relatively high mercury and methylmercury concentrations have been reported for marine birds. Birds feeding on wild vegetation generally have much lower mercury in their bodies. There have recently been many studies performed on terrestrial ecosystems (particularly in Canada, Sweden, and the USA). Mercury also passes from vegetation into the food chains of the fauna.

### General Facts about Organomercury Compounds

The vapor pressure of  $\text{CH}_3\text{HgCl}$  is 1.13 Pa and the vapor pressure for dimethylmercury is several times higher. Mercurials also differ greatly in their solubility. The solubility in water increases in the order: mercurous chloride  $< \text{Hg}^\circ < \text{CH}_3\text{HgCl} < \text{HgCl}_2$ . Certain species of mercury are soluble in nonpolar solvents. This includes elemental mercury and halide compounds of arylmercurials.  $\text{CH}_3\text{Hg}^+$  occurs in aqueous solutions as an aqua complex  $\text{CH}_3\text{–Hg–OH}_2^+$  with a covalent bond between mercury and oxygen. The cation behaves as a ‘soft’ acid and has a strong preference for the addition of only one ligand.  $\text{CH}_3\text{Hg}^+$  undergoes rapid coordination reactions with sulfur, phosphorus, oxygen, nitrogen, halogens, and carbon. The rate of formation of  $\text{Cl}^-$ ,  $\text{Br}^-$ , and  $\text{OH}^-$  complexes is extremely fast and is diffusion controlled. Methylmercury, like  $\text{Hg}^{2+}$ , forms strong bonds with sulfur, and it is very likely that all methylmercury in biota is bound to the sulfhydryl groups of proteins. The organomercury–sulfur bond is, however, much less stable than the inorganic Hg–S bond and can be easily cleaved in acid solutions of pH 1. This is used to liberate methylmercury from biological tissues prior to its analytical determination. The  $\text{CH}_3\text{Hg}^+$  unit itself is kinetically remarkably inert toward decomposition. Therefore, methylmercury compounds, once formed, are not readily demethylated. The neutral species derived from  $\text{CH}_3\text{Hg}^+$  can readily pass through both biological and nonbiological boundaries. This, together with the broad tendency to form stable complexes quickly (and the robustness of the  $\text{CH}_3\text{Hg}^+$  unit), enhances the toxicological properties of methylmercury. Like methylmercury, ethylmercury is an organic mercury compound, considered to have similar toxic properties to methylmercury. Ethylmercury is still used in vaccines and its commercial name is Thiomersal, also known as ‘thimerosal’. It has been used in very small amounts in vaccines since the 1930s to prevent bacterial and fungal contamination. It contains  $\sim 49\%$  ethylmercury. Unfortunately, there is a paucity of data on the metabolism, excretion, and toxicity of ethylmercury although there have been reports on toxicity following exposures.

### Sample Handling and Storage

The various sampling procedures must be appropriate to the sample type and methylmercury concentration. Organic mercury is present at much lower concentrations than total mercury, except in fish or seafood. Therefore, apart from the methylmercury-specific separation and detection techniques, careful



handling of samples prior to analysis is necessary. A general difference between total and organomercury analyses is that the contamination of samples by organomercurials prior to analysis is much less probable. They are not usually present at detectable levels in laboratory environments.

### Cleaning Procedures

Rigorous cleaning procedures must be used for all laboratoryware and other equipment, which comes into contact with samples. Reagents that are used for the analyses of total and organomercury species must be of suitable quality (preanalyzed and shown to contribute minimally to blanks). The best materials for sample storage and sample processing are Pyrex and silica (quartz) glass, and Teflon (PTFE or FEP). Plastics such as polypropylene are not recommended since these materials can contribute to either contamination or losses of mercury. There are several cleaning procedures that are recommended: (1) aqua regia treatment followed by soaking in dilute ( $\sim 5$ – $10\%$ ) nitric acid for a week; (2) soaking in a hot oxidizing mixture of  $\text{KMnO}_4$  and  $\text{K}_2\text{S}_2\text{O}_8$ , followed by  $\text{NH}_2\text{O} \cdot \text{HCl}$  rinsing and soaking for a week in  $5 \text{ mol l}^{-1} \text{ HNO}_3$ ; (3) soaking in a 1:1 mixture of concentrated chromic and nitric acids for few days; (4) soaking in  $\text{BrCl}$  (mixture of  $\text{HCl}$  and  $\text{KBrO}_3$ ); and (5) for Teflon, cleaning in hot concentrated  $\text{HNO}_3$  for 48 h, followed by soaking in dilute  $\text{HNO}_3$  ( $50\%$ ) (which is repeated twice). After such treatments, laboratoryware is usually rinsed with mercury-free deionized water or double distilled water, and stored in a mercury-free place, preferably sealed in mercury-free plastic bags. Some authors recommend storage in dilute ( $1\%$ ) hydrochloric acid until used. Laboratoryware that is used for methylmercury analyses should be prepared with extreme caution. It has been shown that final soaking (after using one of the above mentioned cleaning procedure) of laboratoryware, particularly Teflon, in hot ( $70^\circ\text{C}$ )  $1\%$   $\text{HCl}$  removes any oxidizing compounds (e.g., chlorine), which may subsequently destroy methylmercury in solution.

### Air

Most techniques employ sampling of large volumes of air through one or more media (solid or liquid sorbents) that trap the mercury species of interest. Therefore, one of the most critical parameters in obtaining accurate final results is the determination of the sample volume collected under standard conditions (of temperature and pressure). It is important to calibrate flow meters regularly. Usually a high degree of accuracy is achieved when sampling ambient

(indoor and outdoor) air. The presence of excessive amounts of water vapor and  $\text{CO}_2$  (e.g., in fuel or combustion gases) may necessitate a correction to the flow meter readings.

The first operational step is the removal of particulate matter, which is usually separated by glass fiber filters, Teflon or a plug of glass or quartz-wool. If selective adsorption methods (described later) are used, the first step is usually to separate mercury(II) compounds. Various adsorption materials have been used in the past. One of the most common is Chromosorb W, treated by  $\text{HCl}$  alone or coated with a polar stationary phase (e.g., OV-1 or SE-30). In recent years, methods for reactive gaseous mercury(II) compounds were validated and are based on mist chamber techniques, tubular and annular denuders. Following the trapping of mercury(II) species, a second trap is then used to collect methylmercury compounds. Chromosorb W, treated with  $\text{NaOH}$ , has been employed. The use of impingers with aqueous solutions and/or high-flow refluxing mist chambers has recently been optimized for the direct preconcentration of extremely low concentrations of methylmercury in ambient air. Methylmercury (and dimethylmercury if present) can also be trapped using Tenax or Carbotrap and cryogenic trapping followed by one of specific mercury instrumental technique.  $\text{Hg}^0$  is retained or captured on the final trap, usually consisting of silver- or gold-coated adsorbent. In most cases, the total gaseous mercury fraction in air is determined directly by noble metal amalgamation, or by other methods described in the previous section on inorganic mercury determination.

### Water

Collection and handling of aqueous samples for low-level determination of mercury must address factors such as whether the sample is representative of the system sampled, possible interconversion processes, contamination, as well as preservation and storage of the sample before analysis. The measurement (sampling and analysis) protocol must be even more carefully designed if speciation of mercury forms in aqueous samples is intended. There have been remarkable improvements in sampling and analytical techniques that have resulted in a dramatic increase in the reliability of data for mercury levels in water samples over the past 15 years. The stability of mercury in solution is affected by many factors. These include: (1) the concentration of mercury and its compounds, (2) the type of water sample, (3) the type of containers used, (4) the cleaning and pretreatment of the containers, and (5) the preservative added.

Contamination-free sampling devices (e.g., Teflon coated Go-Flo samplers) are commonly used. Alternatively, the water can be pumped through Teflon tubing using a peristaltic pump. Collection of surface waters is usually performed by hand, using arm-long plastic gloves. Samples are taken upwind of a rubber raft or a fiber glass boat. Precipitation samples can be collected by automatic samplers, with inline filtration if desired. Teflon wide-mouth jars have been favored for sampling waters with low mercury concentrations. Containers and other sampling equipment that come into contact with water samples should be made of borosilicate glass, Teflon, or silica glass. These materials have been found to be free from mercury contamination and therefore suitable for work at low, ambient levels. However, Teflon showed the best performance regarding both contamination- and loss-free storage of aqueous samples. The most volatile mercury forms possibly present in the water are  $\text{Hg}^0$  and dimethylmercury. They should be removed from the samples immediately after the sampling step by aeration (purging) with collection on Carbotrap or other suitable adsorption media for subsequent analyses in the laboratory.

Measurement of dissolved mercury compounds requires removal of particulate materials from the sample by either filtration or centrifugation. Various types of filters can be employed: 0.45  $\mu\text{m}$  membrane filters (precipitation, seawater), disposable polystyrene units (Nalgene) with nitrocellulose membrane filters, and pretreated glass fiber filters.

In order to store samples prior to analysis, samples should be preserved, generally by acidification. However, acidification is subject to two limitations: (1) suspended matter has to be removed prior to the acidification; and (2) dimethylmercury and  $\text{Hg}^0$  have to be removed prior acidification, otherwise conversion of these species into methylmercury and subsequently mercury(II), may occur. In general, samples for speciation of gaseous mercury species in water are processed immediately after sampling. However, if water samples do have to be stored for longer periods the mercury present may be stabilized by adding preservatives. For the analyses of organomercurials, preservation with oxidative reagents (as advised for total mercury analyses) should be avoided, since organomercurials are converted into inorganic mercury. Stabilization by  $\text{HNO}_3$  results in decreased methylmercury, while  $\text{Hg(II)}$  remains stable in the presence of this acid.  $\text{HCl}$  was found the most appropriate acid for storing aqueous methylmercury solutions. Sulfuric acid can also be used for preservation of methylmercury solutions in distilled water, although it is not suitable for natural water samples. Inorganic complexing agents, such as  $\text{Cl}^-$ ,

$\text{I}^-$ ,  $\text{CN}^-$ ,  $\text{Br}^-$ , and organic complexes (L-cystein) have also been proved to stabilize organomercury compounds in solution. Some authors claim that for methylmercury determinations, storage of unpreserved samples at low temperatures (or even deep-frozen) is better than adding acid.

### Biological Samples

Relatively little is known about the effects of storage on the stability of methylmercury in biological samples. Significant external contamination of samples with methylmercury is unlikely to occur; however, extreme precautions are necessary to avoid contamination by inorganic mercury.

Blood and hair samples are often analyzed in order to estimate exposure of humans to mercury and its compounds. Blood should be taken by venipuncture. Since some commercial containers may contain mercury compounds added as preservatives it is advisable to check each commercial batch before use. The samples should be refrigerated but not frozen, as it is sometimes useful to measure mercury in plasma and red blood cells separately. The separation of plasma and red blood cells should be performed as soon as possible to avoid hemolysis of the sample. If extensive hemolysis has occurred, the sample should be homogenized before an aliquot is taken for analysis. Blood samples may also be heparinized for total blood, serum, and red blood cell analyses. If unavoidable, samples may be stored deep frozen. However, repeatedly frozen and unfrozen blood samples showed a remarkable decrease in methylmercury concentrations. There is some evidence that methylmercury may be destroyed during lyophilization of blood samples.

Analysis of human hair offers several advantages over analysis of blood samples: e.g., ease of sampling and sample storage, the concentration of methylmercury is  $\sim 250$  times higher than in blood, and analysis of different longitudinal sections of hair can give information on the history of the exposure to methylmercury ingested through food. Adhering dust and grease should be removed by one of the following solvents: hexane, alcohol, acetone, water, diethylether, or detergents. IAEA and WHO recommend the use of only water and acetone. Long-term storage of human hair samples has shown that methylmercury is stable for a period of a few years if stored dry and in darkness at room temperature.

Biological samples are preferably analyzed fresh or after lyophilization. Deep freezing of fresh samples, especially with long storage, should be avoided, since it has been noticed that in some organisms methylmercury may decompose with repeated freezing and

unfreezing (particularly in bivalves). Methylmercury and total mercury in lyophilized biological samples, such as biological certified reference materials (CRMs), are stable for years. The CRMs are, however, sterilized either by autoclaving or by  $\gamma$ -irradiation. This important step prevents bacteriological activity, which may otherwise lead to methylation–demethylation processes. In general, very little is known on the effects of sterilization on the stability of methylmercury compounds. More studies are needed to investigate the stability of organomercury compounds in biological samples, particularly under various sample preparation and long-term storage conditions.

### Sediment and Soil

Sediment and soil samples should also be prepared with caution since the percentage of methylmercury in these samples is very low (e.g., <2% of total mercury) and improper handling and storage may lead to inaccurate results. These samples should be analyzed fresh or, if long-term storage is unavoidable, samples should be kept in the dark at low temperatures or freeze-dried. Drying at elevated temperatures (in the oven) results in high losses of total and methylmercury. Moreover, changes of pH, redox potential, moisture, etc., may significantly influence the stability of methylmercury in sediments. Due to changes during sampling, conversion of mercury species may occur (methylation, demethylation, reduction), particularly in the case of sediments taken from oxygen depleted bottom of water bodies. These samples should better be analyzed fresh or, if long-term storage is unavoidable, samples should be kept in the dark at low temperature, inert atmosphere, and deep frozen.

### Calibration Standards

There have been quite a few studies performed concerning the stability of organomercurials in standard solutions. A decrease of methylmercury in water solutions can be caused by adsorption onto the container walls or decomposition to Hg(II), Hg<sup>0</sup>, or (CH<sub>3</sub>)<sub>2</sub>Hg. Losses of methylmercury chloride due to volatilization is unlikely to occur ( $K_d$  (gas–liquid distribution coefficient:  $C_{\text{gas}}/C_{\text{H}_2\text{O}}$ ) is  $1.07 \times 10^{-5}$  at 20°C). The stability is strongly dependent on the concentration, the container materials, and the storage temperature. An aqueous methylmercury concentration of  $10 \mu\text{g l}^{-1}$ , stored in Pyrex glass (precleaned with HNO<sub>3</sub>) at low temperature (e.g., in a refrigerator) is stable for ~1 month. If Teflon containers are used, the solution is stable for several months if stored in the dark at room temperature. Some authors

recommend storage of methylmercury solutions in 0.2–0.5% HCl. Methylmercury standard solutions in organic solvents stored in Pyrex glass bottles (for gas chromatography (GC) analyses) are even more stable than water solutions.

Calibration standards may also be prepared in the gas phase. Elemental mercury vapor is often used for calibration of detectors for elemental or total mercury determination through preconcentration by amalgamation. Volatile organomercury compounds could also be prepared in the gas phase. This is of importance for the optimization of methodologies for organomercury speciation in air. An aliquot of vapor is removed from a temperature-stabilized vessel using a gas tight syringe. The concentration can be calculated from data on the partial pressure of the individual compounds and the gas-law equation. Calibration standards for inorganic mercury can also be ordered from major producers of CRMs.

## Analytical Methods

During the last 20 years, hundreds of papers dealing with determinations of organomercury compounds in environmental samples have been published. Most of them are based on the method originally developed by Westöö. In recent years, however, significant improvements of analytical methods in terms of specificity and sensitivity have been developed. This has allowed the determination of mercury speciation in all environmental compartments. Only a brief overview is given of methods that have been reported in review articles. Instead, particular emphasis has been placed on more recent analytical developments and future needs.

### Air

The concentration of total mercury in air can be determined with high precision and accuracy. There are still, however, analytical problems in separating and analyzing all the specific mercury compounds that can be present in air. Since concentrations of organomercurials are so low, a preconcentration step has to be employed. In general, two approaches are possible for measurements of mercury compounds in air: (1) Selective adsorption methods (usually coupled with nonselective detectors), in which the separation is operationally defined and a true species identification is not obtained (therefore it is not clear whether the species determined are the actual mercury compounds present in the atmosphere). (2) Gas chromatographic methods, which allow identification of various organic species by their unique retention times and specific mercury detection system. The

most readily determined forms are  $\text{Hg}^\circ$  and dimethylmercury, as these are the most volatile and non-polar species. Water vapor does not disturb their speciation and analysis. More problematic is the separation of methylmercury and mercury(II) compounds, where water vapor affects both the isolation and measurement steps. From the publications cited in the review article of Schroeder it is evident that in studies where selective adsorption methods are used, a maximum of four mercury species ( $\text{Hg}^\circ$ ,  $\text{Hg(II)}$ , monomethyl mercury, and dimethylmercury) are reported. In studies where GC separations are used, only  $\text{Hg}^\circ$ , monomethyl-, and dimethylmercury are reported, although total mercury is usually determined separately. The main difference between these approaches is the identification of  $\text{Hg(II)}$  compounds. The effect of moisture on the performance of the above-mentioned adsorption techniques has been addressed by many authors, although the problem has not, as yet, been solved completely. There is a big risk of  $\text{Hg}^\circ + \text{O}_3$  reaction in sampling bubblers or traps, which may result in too high  $\text{Hg(II)}$  values.

Improvements have been made recently for the preconcentration of gaseous methylmercury in air by preconcentration in mist chambers and subsequent determination by aqueous-phase ethylation, precollection on carbotrap columns, separation by GC, and detection by cold vapor atomic fluorescence spectrometry (CV-AFS). In order to understand the behavior and fate of mercury compounds in the atmosphere and their role in environmental mercury cycling, there is an urgent need for further development of analytical methods to afford identification of mercury species in the atmosphere by compound-specific analytical methods (as opposed to operationally defined protocols).

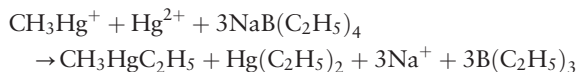
## Water

Only a few analytical techniques have been developed so far for the reliable determination of organomercury species in water samples. In many studies, mercury compounds in aquatic environments are speciated according to their ability to be reduced to the elemental state. Dimethylmercury and  $\text{Hg}^\circ$  (the most volatile mercury compounds) can be isolated by aeration and adsorption on a suitable adsorbent (dimethylmercury on Carbotrap or Tenax and  $\text{Hg}^\circ$  on gold trap) or by cryotrapping, or coupled with noble metal amalgamation (for  $\text{Hg}^\circ$ ), immediately after sampling. Alternatively, they can be directly analyzed by separation on GC columns and detected by one or more suitable mercury detectors. Samples should not be acidified prior to such separations since dimethylmercury and  $\text{Hg}^\circ$  are transformed into

methylmercury and  $\text{Hg(II)}$ , respectively. For specific organomercury compound determinations, a pre-separation and preconcentration method, followed by a very sensitive detection system, is necessary. In a typical extraction method,  $\text{CH}_3\text{HgX}$  (X is a halide ion) is extracted into an organic solvent (benzene or toluene) after acidification. This is followed by derivatization to a water-soluble adduct of methylmercury-cysteine, which is extracted into a water phase. After acidification,  $\text{CH}_3\text{HgX}$  is back-extracted into small amount of organic solvent. An aliquot is then injected onto a gas-liquid chromatography (GLC) column and detected by an electron-capture detector (ECD) or any other suitably sensitive detector (such as a plasma emission detector). Packed or capillary columns can be used. More precise descriptions of the chromatographic conditions are given later. There are quite a few modifications to this extraction procedure. For example, the methylmercury-cysteine compound may be transferred into dithizone chloroform solution and quantified by atomic absorption spectrometry (AAS). Inorganic and organic mercury species can be preconcentrated on dithiocarbamate or sulfhydryl cotton-fiber adsorbent that is then extracted as described above. The common drawbacks of most of these extraction procedures are the large sample requirements, low extraction yields, and nonspecific separation of dimethylmercury, if present.

There are also methods for determination of 'total' organomercury compounds. Inorganic and organic mercury are preconcentrated on a dithiocarbamate resin and are subsequently eluted with thiourea. Separation of organic and inorganic mercury is achieved by differential reduction and detection by CV-AAS. Inorganic and organic mercury can also be separated using anion-exchange resins. Organic mercury is then decomposed (by ultraviolet (UV) irradiation) and measured by CV-AAS. It has, however, been shown that the levels obtained by this method do not necessarily correspond to methylmercury (owing to the lack of specificity of the protocol). The method has recently been improved by the introduction of more specific separations of organic and inorganic mercury species by water vapor distillation.

Another method is based on aqueous-phase ethylation, room-temperature precollection, and separation by GC with CV-AFS detection. Modification of this method is frequently used in laboratories involved in studies of the biogeochemical cycle of mercury. Ionic mercury species in the sample are ethylated according to the following reactions:





Ethylated mercury species are volatile and can therefore be purged from solution at room temperature and then collected on adsorbent materials such as Carbotrap or Tenax. After thermal release, individual mercury compounds are separated by cryogenic or isothermal GC. As the species are eluted they are thermally decomposed (pyrolyzed) at 900°C and measured as  $\text{Hg}^0$  using a CV-AFS detector, which achieves very low detection limits ( $< 10^{-12}$  g). A CV-AAS detector can also be used with poorer limit of detection. In recent years, inductively coupled plasma-mass spectrometry (ICP-MS) is more and more frequently used as it offers numerous advantages over AAS and AFS detectors. Instead of ethylation, propylation was recently shown to be an even more suitable derivatization method, being free from interferences caused by halide ions. Hydration has also proven to be a useful derivatization method, in particular when coupled with preconcentration by cryotrapping.

In any case, the critical part of this procedure is the preparation of samples prior to derivatization. Methylmercury compounds must be removed from bound sites to facilitate the ethylation reaction. Interfering compounds (such as sulfides) must also be removed. Two approaches have so far been used. The first is based on extraction of methylmercury compounds into methylene chloride and then back-extraction into water by solvent evaporation. The second is based on water vapor distillation. Distillation has advantages since it quantitatively releases methylmercury from sulfur and organic-rich water samples. This approach has been proposed for standardization by the USEPA.

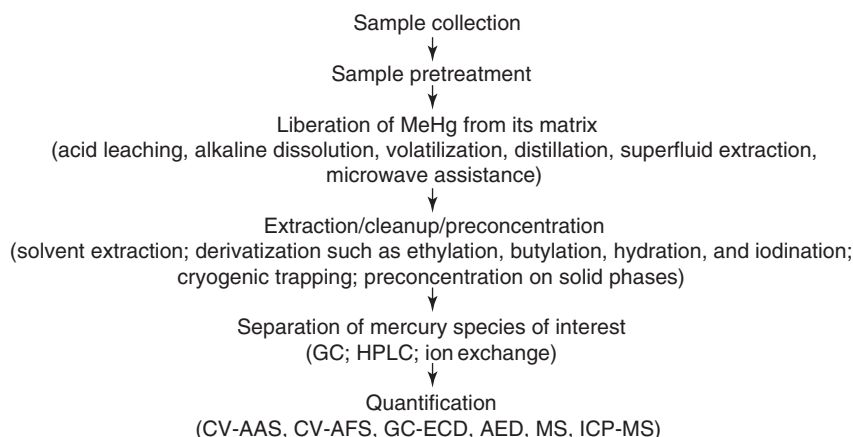
The analysis of different mercury compounds at environmental concentrations should be developed

further. Reliable data can be obtained for monomethyl mercury compounds and dimethylmercury. There is, however, a need to identify biogeochemically important mercury fractions that are currently measured by operationally defined rather than direct ('analytically rigorous') protocols.

### Other Environmental Samples

In general, methods are classified according to the isolation technique and the detection system. Most methods for the isolation/separation of organomercury compounds have been based on solvent extraction, differential reduction, difference calculations between 'total' and 'ionic' mercury, derivatization, or with paper- and thin-layer chromatography (TLC). The most common approaches to organomercury separation and detection are schematically presented in Figure 1.

**Extraction and GC determination** The basis of most present methods was introduced by Japanese and Scandinavian workers. It involves the extraction of organomercury chloride from acidified homogeneous samples into benzene (however, the use of toluene is strongly recommended for health and safety reasons). Organomercury compounds are then back-extracted into an aqueous cysteine solution. The aqueous solution is then acidified and organomercury compounds are reextracted with benzene or toluene. This double partitioning enables removal of many interferents (e.g., benzene-soluble thiols). Finally, methylmercury is analyzed by GC with electron-capture detection. Several modifications have been made to this protocol for the separation and identification of organic mercury in biological and other samples. For example, in the initial step the



**Figure 1** Steps for determination of organomercury compounds. CV-AAS, cold vapor-atomic absorption spectrometry; CV-AFS, cold vapor-atomic fluorescence spectrometry; GC, gas chromatography; GC-ECD, gas chromatography-electron capture detection; AED, atomic emission detector; ICP-MS, inductively coupled plasma-mass spectrometry; HPLC, high-performance liquid chromatography.



addition of copper(II) ions (or mercury(II)) enhances the removal of mercury bound to sulfur. Copper(II) was found superior to mercury(II) since it avoided problems of decomposing dimethylmercury, if present. However, inorganic mercury cannot be determined by this procedure, unless a reagent is added to form, for example, alkyl and aryl derivatives, which can then be extracted and determined by GC. In general, solvent extraction procedures are time consuming, corrections for the recovery of the procedure vary from sample to sample, and with some sample types (e.g., those rich in lipids) phases are difficult to separate due to the presence of persistent emulsions, particularly during the separation of the aqueous cysteine phase. To overcome these problems methylmercury can be adsorbed on cysteine paper (instead of into cysteine solution) during the cleanup stage. Using additional pre-separations prior to extraction, such as volatilization of methylmercury in a microdiffusion cell and distillation may also facilitate separation of phases during extraction.

When speciation is required with insoluble samples (such as sediments and soils), it is difficult to estimate recovery. In such samples, spiked methylmercury is not equivalent to the methylmercury originally present. By comparing various isolation techniques for methylmercury compounds in sediment samples and soils it has been shown that conventional methods based on acid leaching of organomercury compounds prior to extraction of methylmercury compounds into an organic solvent is inadequate in most cases to release methylmercury from sediment samples. Improved recoveries have been achieved by extraction of methylmercury by nitric acid at elevated temperature or assisted by microwave energy. It is important to mention that some protocols may lead to artifact methylmercury production, especially in procedures where methylmercury is isolated at higher temperatures. Quality of the results should therefore be regularly checked by the use of appropriate reference materials, if available or by comparison of the results between the laboratories and/or different analytical approaches.

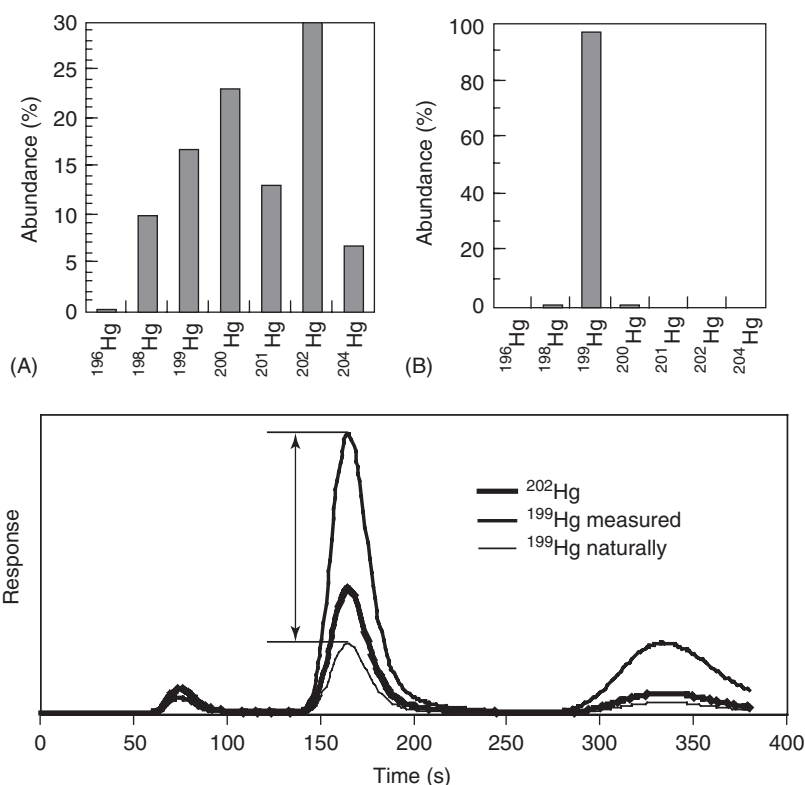
**Chromatographic conditions** Apart from the above-mentioned problems associated with the extraction of organomercurials, problems also exist in the chromatography of organomercurial halides. Many investigations have recommended that columns packed with 5% DEGS-PS on 100–120 mesh Supelcoport be used. Some other polar stationary phases have also been employed, e.g., PEGS, Carbowax 20 M, Durapak, Carbowax 400, PDEAS, HIEFF-2AP, etc. In order to prevent ion-exchange and adsorption processes on the column (which cause

undesirable effects such as tailing, changing of the retention time, and decrease of peak areas/heights), passivation of packing material is needed with Hg(II)chloride in benzene. Although the more inert nature of capillary columns would be expected to minimize such effects, improved chromatographic performance over packed columns cannot be readily achieved. Some workers still prefer to use packed columns since the analytical protocols using capillary columns require additional research to optimize performance.

Several workers have chosen to derivatize the mercury species to their corresponding nonpolar, alkylated analogs such as butyl derivatives, which can then be separated on nonpolar packed or capillary columns.

**Detectors** Various detectors can be used in combination with GC for the determination of mercury species. An ECD is a sensitive detector with an absolute detection limit of a few picograms. It does not, however, measure mercury directly, but responds to the halide ion attached to the  $\text{CH}_3\text{Hg}^+$  ion. The identification of small methylmercury peaks can sometimes be subject to a positive systematic error owing to coeluting contaminants. The use of a plasma atomic emission detector, a mass spectrometric detector, CV-AAS, CV-AFS, or ICP-MS can avoid such problems, since mercury is measured directly.

In recent times, the use of ICP-MS has become more widely used as a powerful tool for mercury speciation. Introducing mercury in the form of gaseous species into a dry plasma greatly reduces memory effects, achieving absolute detection limits of less than 100 pg of Hg. Based on the principle of isotope dilution, ICP-MS is characterized as a very precise and accurate analytical method. Equally important, multiple stable tracer experiments allow studies of the fate of Hg species in the environment and in biological systems. This concept allows the investigation of multiple transformation processes simultaneously. A simplified principle is presented in **Figure 2**. The isotope pattern of ambient Hg is presented in **Figure 2A** and the composition of a typical isotope enriched solution ( $^{199}\text{Hg}^{2+}$ ) in **Figure 2B**. The isotope pattern of the inorganic mercury in the sediment will be changed by adding such solutions into the sample. **Figure 2** also represents a GC-ICP-MS chromatogram of a sediment extract from methylation assay. The concentration of  $\text{CH}_3^{199}\text{Hg}^+$  is enriched in the sample relative to the fraction of  $\text{CH}_3^{199}\text{Hg}^+$  in the original sample and the magnitude of relative change relates to the amount of  $^{199}\text{Hg}^{2+}$  methylated.



**Figure 2** Top: Hg isotope abundance in nature (A) and <sup>199</sup>Hg tracer solution (B). Bottom: GC-ICP-MS chromatogram of a sediment extract after incubation with <sup>199</sup>Hg(II) showing the relative increase in the CH<sub>3</sub><sup>199</sup>Hg<sup>+</sup> signal.

**Derivatization methods** Many methods use the formation of volatile organomercury derivatives (through ethylation, butylation, hydration, and iodination) in order to separate them from the bulk of the sample by simple room-temperature aeration. The same ethylation method as described for water samples has also been applied to biological and sediment samples. An aliquot of alkaline digested sample or a distilled sample is subjected to ethylation by sodium tetraethylborate. Methylmercury is transformed into methylethylmercury and mercury(II) is transformed into diethylmercury. Both species can be determined simultaneously. Volatile ethylated mercury compounds, as well as elemental and dimethylmercury, are removed from solution by aeration and are then trapped on an adsorbent (Carbotrap or Tenax). Mercury compounds are separated on a GC column and pyrolyzed to elemental Hg<sup>0</sup> at 900°C for subsequent mercury determination by CV-AFS, CV-AAS, or ICP-MS. As mentioned previously, very low detection limits may be achieved by CV-AFS and ICP-MS, particularly if methylmercury is pre-separated by distillation (6 pg l<sup>-1</sup> for water and 1 pg g<sup>-1</sup> for biota and sediment samples). Instead of sodium tetraethylborate, sodium borohydride may also be used to form volatile methylmercury hydride, which is then quantified by GC in line with a Fourier

transform infrared spectrophotometer. The CH<sub>3</sub>I formed in a headspace vial may also be introduced onto a GC column and detected by microwave-induced plasma atomic emission spectrometry or AFS detectors. Propylation and hydration has also been applied with great success as described above.

**Differential reduction** There are also a few methods that are based on differential reduction. In the method developed by Magos, the inorganic mercury in an alkaline digested sample is selectively reduced by stannous chloride while organomercury compounds are reduced to elemental mercury by a stannous chloride-cadmium chloride combination. Elemental mercury released can be measured by CV-AAS. The method has been successfully applied for biological samples in toxicological, epidemiological, and clinical studies. CV-AAS has also been used for detection of organomercury compounds after pre-separation of organomercury by (1) anion exchange, (2) volatilization and trapping on cysteine paper, and (3) water vapor distillation. Organomercury compounds have to be destroyed by either UV irradiation or acid digestion prior to detection by CV-AAS. In most biological samples, the organomercury concentrations usually correspond to methylmercury. In some environmental samples such as sediment,

soil, and water samples, the concentrations of organic mercury (particularly if separated by anion exchange) have been found to be much higher than those of methylmercury compounds. This is probably due to the presence of some other organic mercury compounds that have not, as yet, been identified.

**Miscellaneous methods** The first practical method for differentiating between organic and inorganic mercury was a colorimetric method. Organomercury compounds were extracted into an organic solvent and determined spectrophotometrically as dithizone complexes. The method basically suffers from low sensitivity. High-performance liquid chromatography (HPLC) has proven to be of use with reductive amperometric electrochemical detection, UV detection, ICP-AES detection, or AAS detection. Neutron activation analysis has been used for methylmercury determinations in fish, blood, and hair samples after suitable pre-separation procedures. Graphite furnace AAS has also been used for the final determination of methylmercury in toluene extracts to which dithizone was added. An anodic stripping voltammetry technique has been developed for determination of methylmercury. However, the method has never been used for environmental samples. Methylmercury has also been extracted into dichloromethane. This was then evaporated down to 0.1 ml and subjected to GC with an atmospheric pressure active nitrogen detector.

An enzymatic method for specific detection of organomercurials in bacterial cultures has also been developed. It is based on the specific conversion of methylmercury (no other methylmetallo groups are enzymatically converted) to methane by organomercurial lyase. Ethyl- and phenylmercury can also be detected by this procedure.

### Determination of Other Organomercurials

Among organomercury species currently of interest, ethylmercury (EtHg) is a compound that requires further attention as it is still used in Thiomersal for preservation of vaccines. It is important to analyze ethylmercury in vaccines, in wastewater from waste treatment plants in industries using ethylmercury, as well biological samples in order to understand ethylmercury uptake, distribution, excretion, and effects. In principle, methods developed for methylmercury can also be used for ethylmercury, except in the protocols using derivatization by ethylation. In such cases propylation is recommended.

Only a few investigations concerning the determination of other organomercurials used in agricultural and for other purposes (see Table 1) have been

reported. Methoxyethyl- and ethoxyethylmercury have been examined by TLC and GC. It would appear that the only method that can separate and measure many of the compounds simultaneously is HPLC with UV detection. It offers several advantages. The separation of the compounds is performed at ambient temperatures; hence, thermal decomposition of the compounds does not occur. It offers the possibility to separate less volatile or nonvolatile species such as mersalylic acid or the aromatic organomercurials, which usually present a problem for GC. It is, however, very important to isolate these compounds from the environmental samples quantitatively. Methyl- and ethylmercury can easily be isolated from soils by extraction from acidified samples. Several extraction agents have been tested in order to release organomercurials from soils. Methyl- and phenylmercury can be extracted by potassium iodide–ascorbic acid and oxalic acids with satisfactory yields, whereas ethylmercury is only partly extracted. No suitable extraction techniques have been found for methoxyethyl- and ethoxyethylmercury in soils (due to decomposition of these compounds under acidic conditions).

### Toxicological Considerations

Organic mercury compounds are extremely toxic. Direct contact with skin can lead to fatal effects. Precaution during handling of this substance is absolutely necessary. It must, therefore, be handled in glove boxes to avoid inhalation, and gloves as well as other protecting clothes must be worn.

The general population is primarily exposed to methylmercury through the diet, particularly fish and fish products are the dominant source. As a preventive measure, WHO recommends the monitoring of hair levels of methylmercury in women of child-bearing age who consume large amounts of fish (e.g., 100 g per day or more). If the results of such monitoring activities indicate excessive exposure to methylmercury, appropriate and practical measures such as dietary changes should be introduced to reduce exposure, and to maintain it below internationally recommended allowable intakes.

*See also:* **Atomic Absorption Spectrometry:** Flame. **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Quality Assurance:** Reference Materials. **Water Analysis:** Seawater – Dissolved Organic Carbon.

### Further Reading

Bloom NS (1989) Determination of picogram levels of methylmercury by aqueous phase ethylation, followed by

- cryogenic gas chromatography with cold vapour atomic fluorescence detection. *Canadian Journal of Fisheries and Aquatic Sciences* 46: 1131–1135.
- Bulska E, Emteborg H, Baxter DC, and Frech W (1992) Speciation of mercury in human whole blood by capillary gas chromatography with a microwave-induced plasma emission detector system following complexometric extraction and butylation. *Analyst* 117: 657–600.
- Demuth N and Heuman KG (2001) Validation of methylmercury determinations in aquatic systems by alkyl derivatization methods for GC analysis using ICP-MS. *Analytical Chemistry* 73: 4020–4027.
- Fillipelli M, Baldi F, Brinckman FE, and Olson GJ (1992) Methylmercury determination as volatile methylmercury hydride by purge and trap gas chromatography in line with Fourier transform infrared spectroscopy. *Environmental Science & Technology* 26: 1457–1460.
- Hempel M, Hintelman H, and Wilken R-D (1992) Determination of organic mercury species in soils by high-performance liquid chromatography with ultraviolet detection. *Analyst* 117: 117–201.
- Hintelmann H and Ogrinc N (2003) Determination of stable mercury isotopes by ICP-MS and their application in environmental studies. In: Yong C and Olin CB (eds.) *Biogeochemistry of Environmentally Important Trace Elements*, ACS Symposium Series 835, pp. 321–338. New York: American Chemical Society.
- Horvat M and Byrne AR (1992) Preliminary study of the effects of some physical parameters on the stability of methylmercury in biological samples. *Analyst* 117: 665–667.
- Horvat M, Liang L, and Bloom NS (1993) Comparison of distillation with other current isolation methods for the determination of methylmercury compounds in low level environmental samples; Part I. Sediments and Part II. Water. *Analytica Chimica Acta* 281: 135–152.
- Lansens P, Meuleman C, and Baeyens W (1990) Long-term stability of methylmercury standard solutions in distilled, deionized water. *Analytica Chimica Acta* 229: 281–285.
- Lee YH, Wangberg I, and Munthe J (2003) Sampling and analysis of gas-phase methylmercury in ambient air. *The Science of the Total Environment* 304: 107–118.
- Morita M, Yoshinaga J, and Edmonds JS (1998) The determination of mercury species in environmental and biological samples. *Pure and Applied Chemistry* 70: 1585.
- Rapsomanikis S and Craig PJ (1991) Speciation of mercury and methylmercury compounds in aqueous samples by chromatography – AAS after ethylation with sodium tetraethylborate. *Analytica Chimica Acta* 248: 563–567.
- Schroeder WH (1989) Developments in the speciation of mercury in natural waters. *Trends in Analytical Chemistry* 8: 339–342.
- Tseng CM, de Diego A, Pinaly H, Amoroux D, and Donard OFX (1998) Cryofocusing coupled to atomic absorption spectrometry for rapid and simple mercury speciation in environmental matrices. *Journal of Analytical Atomic Spectroscopy* 13: 755.
- Westö G (1966) Determination of methylmercury compounds in foodstuffs. I: Methylmercury compounds in fish, identification and determination. *Acta Chimica Scandinavica* 20: 2131–2135.

## METABOLISM

See **DURG METABOLISM: Overview; Metabolite Isolation and Identification; Isotope Studies. CLINICAL ANALYSIS: Inborn Errors of Metabolism**

# MICELLAR ELECTROKINETIC CHROMATOGRAPHY

**S Terabe and J-B Kim**, University of Hyogo, Hyogo, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Over the last two decades, capillary electrophoresis (CE) has been developed as a powerful separation technique for complex mixtures. Its advantages include a high separation efficiency, short analysis time, small sample requirement, and applicability to a wide range of analytes. The basic modes of CE that are presently being exploited include capillary zone electrophoresis (CZE), micellar electrokinetic chromatography (MEKC), capillary isotachopheresis, capillary isoelectric focusing, and capillary gel electrophoresis.

In CZE, only ionic or charged analytes can be separated in principle since its separation mechanism is based on the difference in electrophoretic mobility of the analytes. MEKC has become popular as a powerful technique for the separation not only of neutral analytes but also of charged ones using a conventional CE instrument without any alteration. In MEKC, an ionic surfactant micelle is used as a pseudostationary phase (PS) that corresponds to the stationary phase in conventional chromatography and the surrounding aqueous phase to the mobile phase. The separation principle of MEKC is based on differential partitioning of analytes between the aqueous phase and the micelle phase.

MEKC is a mode of electrokinetic chromatography (EKC) in which surfactants (micelles) are added to the buffer solution. They have polar head groups that can be cationic, anionic, neutral, or zwitterionic and nonpolar hydrocarbon tails. The formation of micelles is a direct consequence of the hydrophobic effect. The surfactant molecules can self-aggregate if the surfactant concentration exceeds a critical micelle concentration (CMC). The hydrocarbon tails will then be oriented toward the center of the aggregated molecules, whereas the polar head groups point outward. Micellar solutions can solubilize hydrophobic compounds that would otherwise be insoluble in water. Every surfactant has a characteristic CMC and aggregation number (AN), i.e., the number of surfactant molecules making up a micelle (typically in the range of 50–100) (see also Table 1). The size of the

micelles is in the range of 3–6 nm in diameter; therefore, micellar solutions exhibit properties of homogeneous solutions. Micelles have a dynamic structure that is the result of the rapid equilibrium between aggregated and monomeric forms. It is the differences in interaction between the micelle and the neutral solute that cause the separation. Micellar solutions have been employed in a variety of separations and spectroscopic techniques.

## Micelles as Pseudostationary Phase for MEKC

The CMCs and ANs of various kinds of surfactants suitable for PSs in MEKC are listed in Table 1. An anionic surfactant, sodium dodecyl sulfate (SDS), is most widely used as a PS in MEKC. Its popularity can be attributed to its low CMC, high aqueous solubility, low Krafft point, ready availability and the low cost of the pure product.

Cationic surfactants such as tetradecyltrimethylammonium bromide (TTAB), cetyltrimethylammonium bromide (CTAB), and cetyltrimethylammonium chloride (CTAC) have also been useful for MEKC analysis. Most cationic surfactants have an alkyltrimethylammonium group, and their counterions are halides. The addition of cationic surfactants to the background electrolytes (BGE) caused the reversal of electroosmotic flow (EOF) owing to a positively charged capillary wall on account of the adsorption of cationic surfactants. As a result of the reversed EOF, the polarity of the electrodes has to be reversed in order to detect the analytes.

Nonionic surfactants themselves do not possess electrophoretic mobility, but they have the distinct advantage of not contributing appreciably to Joule heating; hence, they may be used at high concentrations. They can have a great influence on the separation of charged analytes and can also be employed as PSs in mixed micelles with ionic surfactants. Nonionic surfactants such as polyoxyethylene (23) dodecyl ether (Brij 35), and polyoxyethylene (20) sorbitan monolaurate (Tween 20) have been effectively used as PSs for the separation of charged compounds through MEKC.

Zwitterionic surfactants are not widely used in MEKC. However, zwitterionic surfactants will be interesting if they are used in mixed micelles or as a modifier of the micelle because they should show



**Table 1** CMCs and ANs of selected surfactants

Surfactant	CMC <sup>a</sup> ( $10^{-3} \text{ mol l}^{-1}$ )	AN
<i>Anionic</i>		
SDS	8.1	62
Sodium tetradecyl sulfate	2.1 (50°C)	138
Sodium decanesulfonate	40	40
Sodium <i>N</i> -lauroyl- <i>N</i> -methyltaurate	8.7	—
Sodium polyoxyethylene dodecyl ether sulfate	2.8	66
Sodium <i>N</i> -dodecanoyl-L-valinate	5.7 (40°C)	—
Sodium cholate	13–15	2–4
Sodium deoxycholate	4–6	4–10
Sodium taurocholate	10–15	5
Sodium taurodeoxycholate	2–6	—
Potassium perfluoroheptanoate	28	—
<i>Cationic</i>		
TTAB	3.5	75
Dodecyltrimethylammonium bromide	15	56
CTAB	0.92	61
CTAC	1.3	—
<i>Nonionic</i>		
Polyoxyethylene(23) dodecyl ether (Brij 35)	0.1	—
Polyoxyethylene(20) sorbitan monolaurate (Tween 20)	0.059	—
<i>Zwitterionic</i>		
3-((3-cholamidopropyl)dimethylammonio)-1-propanesulfonate	4.2–6.3	10
<i>N</i> -dodecyl- <i>N</i> , <i>N</i> -dimethylammonio-3-propanesulfonate	3.3	—

<sup>a</sup> 25°C.

significantly different selectivities from other types of surfactants.

Macromolecular surfactants or high-molecular-mass surfactants (HMMS), butyl acrylate–butyl methacrylate–methacrylic acid copolymers (BMMAs), and sodium 10-undecylenate (SUA) oligomer have been introduced as PSs for MEKC. Since a HMMS forms a molecular micelle, which consists of one molecule, and the CMC value is essentially zero, one can expect a higher reproducibility in a HMMS–MEKC system compared with a low molecular mass surfactant (LMMS)–MEKC system.

Bile salts are biological surfactants synthesized in the liver. The bile salts have both a hydrophilic and a hydrophobic face and tend to combine together at the hydrophobic face in an aqueous phase. To improve selectivity, various types of bile salts have been used as PSs in MEKC. **Figure 1** shows the separation and determination of the ingredients of a cold medicine through MEKC with bile salt.

Although they are not micelles, microemulsions and cyclodextrin (CD) derivatives have also been introduced as PSs for MEKC. Oil-in-water (o/w) microemulsions have been shown to be good PSs for EKC. Microemulsions (o/w) are prepared by mixing oil, water, a surfactant, and a cosurfactant such as a medium alkyl-chain alcohol. They have the characteristic properties of a solvent, such as thermodynamic stability and a high solubilization power. The

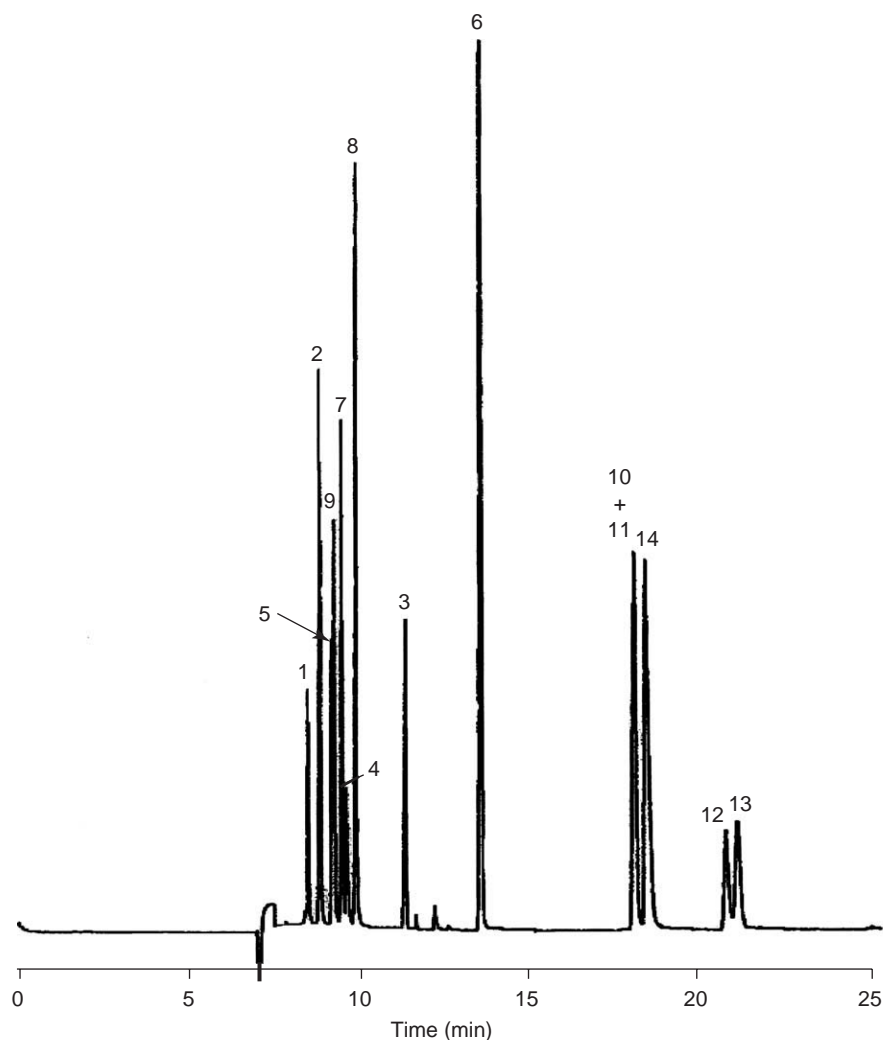
structure of the o/w microemulsion is similar to that of the micelle, except that the microemulsion has a core of a minute droplet of oil. The surfactant and the cosurfactant are located on the surface of the oil droplet to stabilize it. The separation basis of microemulsion EKC (MEEKC) is similar to that involved in MEKC. CD has been widely used not only in liquid chromatography but also in CE as a mobile phase modifier. The use of CD is particularly effective for the separation of aromatic isomers and enantiomers.

## Fundamentals

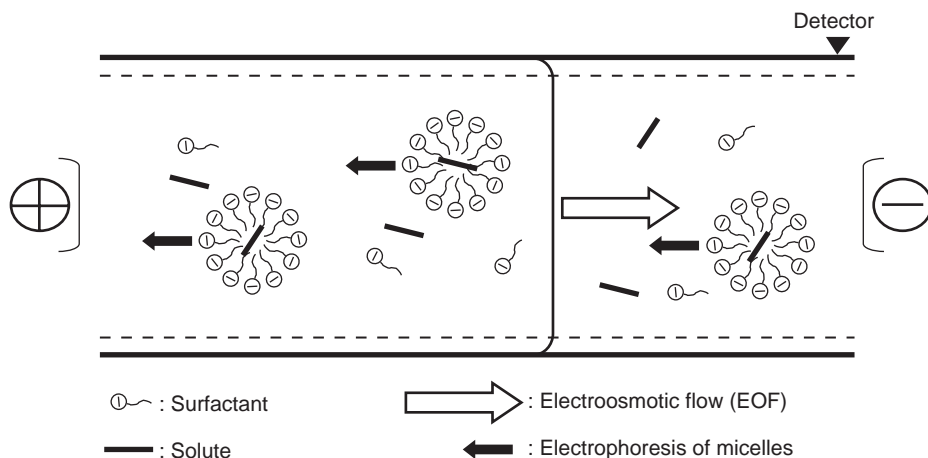
### Separation Principle

**Figure 2** illustrates a schematic representation of the separation principle of MEKC. When an anionic surfactant such as SDS is employed, the micelle migrates toward the anode by electrophoresis. The EOF transports the bulk solution toward the cathode due to the negative charge on the surface of fused silica. The EOF is usually stronger than the electrophoretic migration of the micelle under neutral or alkaline conditions, and therefore, the anionic micelle also travels toward the cathode at a retarded velocity.

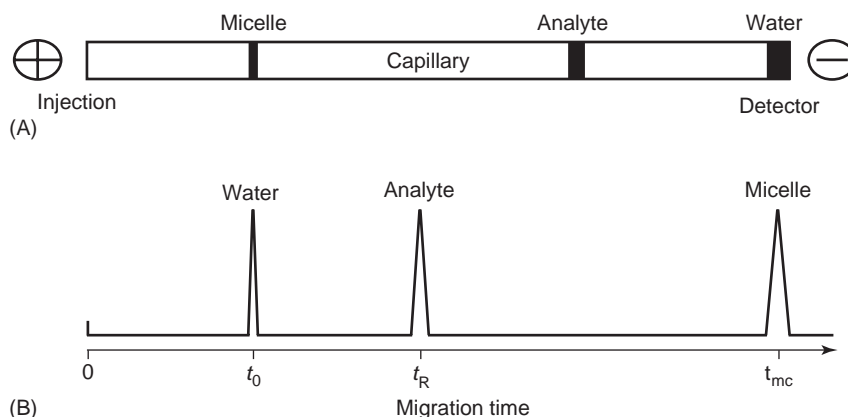
When a neutral analyte is injected into the micelle solution, a fraction of it is incorporated into the micelle and it migrates at the velocity of the micelle.



**Figure 1** Separation of 14 ingredients through MEKC using bile salt. Buffer,  $0.02 \text{ mol l}^{-1}$  phosphate–borate (pH 9.0) containing  $0.05 \text{ mol l}^{-1}$  sodium deoxycholate; applied voltage, +20 V; temperature, ambient; detection wavelength, 210 nm. Solutes: 1, caffeine; 2, acetaminophen; 3, sulpyrin; 4, trimetoquinol hydrochloride; 5, guaifenesin; 6, naproxen; 7, ethenzamide; 8, phenacetin; 9, isopropylantipyrine; 10, noscapine; 11, chlorpheniramine maleate; 12, tipecidine hibenazate; 13, dibucaine hydrochloride; 14, triprolidine hydrochloride. (Reprinted with permission from Nishi H, Fukuyama T, and Terabe S (1990) *Journal of Chromatography A* 498: 313–323. © Elsevier.)



**Figure 2** Schematic illustration of the separation principle of MEKC.



**Figure 3** Schematic diagram of the zone separation in MEKC (A) and chromatogram (B).

The remaining fraction of the analyte migrates at the velocity of the EOF, and thus the migration velocity of the analyte depends upon the distribution coefficient between the micellar and the nonmicellar (aqueous) phases. The greater the percentage of analyte that is distributed into the micelle, the slower it migrates. The analyte must migrate at a velocity between the electroosmotic velocity and the velocity of the micelle (see **Figure 3**), provided the analyte is electrically neutral. In other words, the migration time of the analyte,  $t_R$ , is limited between the migration time of the bulk solution,  $t_0$ , and that of the micelle,  $t_{mc}$ . This is often referred to in the literature as the migration time window in MEKC.

### Retention Factor

The retention factor,  $k$ , is defined as the ratio of the number of analytes incorporated into the micelle,  $n_{mc}$ , and that in the surrounding aqueous solution,  $n_{aq}$ :

$$k = \frac{n_{mc}}{n_{aq}} \quad [1]$$

We can obtain the relation between the retention factor and migration time for an electrically neutral analyte as

$$k = \frac{t_R - t_0}{t_0 \cdot (1 - (t_R/t_{mc}))} \quad [2]$$

It can be rewritten as

$$t_R = \frac{1 + k}{1 + (t_0/t_{mc}) \cdot k} \cdot t_0 \quad [3]$$

When the migration time of the micelle is infinite or the micelle does not migrate in the capillary at all,  $t_0/t_{mc}$  will be zero; then eqns [2] and [3] become identical to those for conventional chromatography,

eqns [4] and [5]:

$$k = \frac{t_R - t_0}{t_0} \quad [4]$$

$$t_R = (1 + k) \cdot t_0 \quad [5]$$

In conventional chromatography,  $k = \infty$  means that the solute is totally retained in the stationary phase and is not eluted at all or the retention time becomes infinite, whereas in MEKC  $k = \infty$  means that the migration time of the solute,  $t_R$ , is equal to  $t_{mc}$ . In that case, such a solute migrates at the same velocity as the micelle or at the lowest velocity and is then detected last.

When  $t_0 = 0$  or the EOF is completely suppressed, eqn [3] becomes

$$t_R = \left(1 + \frac{1}{k}\right) \cdot t_{mc} \quad [6]$$

In this case, the surrounding aqueous phase remains stationary in the capillary and the micelle migrates toward the anode by electrophoresis.

Calculation of the retention factor using eqn [2] requires knowledge of  $t_0$ ,  $t_R$ , and  $t_{mc}$ . For the EOF marker of  $t_0$ , methanol is often used since the distribution coefficient for methanol between the micelle and aqueous phases is negligibly small. Although methanol is UV transparent, it is detectable through the baseline disturbance of the UV detector due to the change in refractive index. Sudan III and Sudan IV are widely used as micelle markers since they are highly lipophilic and totally incorporated into micelle. Timepidium bromide, Yellow OB, and Orange OT are also useful as tracers of micelles.

Note that eqn [3] was derived for the retention characteristics of neutral compounds. Hence, relationships between the migration time or mobility and retention factor will be more complicated when the

solute has an electrophoretic mobility; that is, the migration time of the ionic solute includes a portion ascribed to the migration of the micelle when the solute is incorporated into the micelle and also another portion ascribed to the electrophoresis of the solute itself.

### Resolution

The resolution ( $R_s$ ) in MEKC is given by eqn [7]

$$R_s = \frac{\sqrt{N}}{4} \cdot \left( \frac{\alpha - 1}{\alpha} \right) \cdot \left( \frac{k_2}{1 + k_2} \right) \cdot \left( \frac{1 - (t_0/t_{mc})}{1 + (t_0/t_{mc}) \cdot k_1} \right) \quad [7]$$

where  $N$  is the theoretical plate number,  $\alpha$  is the separation factor equal to  $k_2/k_1$ , and  $k_1$  and  $k_2$  are the retention factors of two closely eluting analytes, 1 and 2, respectively. The resolution increases in proportion to the square root of the plate number. The higher the applied voltage, the higher the plate number, unless the conditions are such that the applied voltage generates excessive Joule heating. The average plate numbers for most analytes are usually in the range of 100 000–200 000. If the plate number is considerably lower, analytes are likely to be adsorbed on the capillary wall. In such cases, experimental conditions must be optimized to produce more efficient separations. Cleaning of the capillary is a possible procedure, as is changing the pH of the run buffer. Hydrophobic analytes, or those having longer migration times, typically yield high theoretical plate numbers because the micelle has a smaller diffusion coefficient. This is explained by assuming the major band broadening is due to the longitudinal diffusion as in CZE. The plate number does not depend significantly on the capillary length. With short capillaries, however, the amount of sample volume injected must be minimized to avoid zone broadening.

The separation factor,  $\alpha$ , is the most important and effective term for maximizing the resolution. The separation factor reveals the relative difference of the distribution coefficient between the two analytes and can be manipulated by chemical means. Since the distribution coefficient is a characteristic of a given separation system, the selectivity can be manipulated by changing either the type of micelle or by modifying the aqueous phase.

The optimum value of the retention factor,  $k_{opt}$ , for maximum  $R_s$  can be calculated by differentiating the following equation, provided  $N$  is independent of  $k$ :

$$f(k) = \left( \frac{k_2}{1 + k_2} \right) \cdot \left( \frac{1 - (t_0/t_{mc})}{1 + (t_0/t_{mc}) \cdot k_1} \right) \quad [8]$$

Then,

$$k_{opt} = \sqrt{\frac{t_{mc}}{t_0}} \quad [9]$$

We can find that  $k_{opt}$  is a function of  $t_{mc}/t_0$ . Under neutral conditions, the optimum value is  $\sim 2$  for SDS micelles as the PS. For practical use, the range of  $k$  recommended is between 1 and 5, or at maximum between 0.5 and 10. Unlike conventional chromatography, in MEKC the micelles as the PS migrate with the bulk liquid due to the EOF, in addition to the electrophoretic migration. This migration of the micelle through the EOF does not affect the separation at all. However, the distance of the migration of the micelle cannot be used for the separation. Since the time available for the solute to interact with the moving micelle depends on its  $k$ , the column availability depends on  $k$  as shown by the last term of the right-hand side of eqn [7] or [8].

### Using Modifiers in MEKC

MEKC is usually performed with a BGE that consists of just a buffer and a surfactant. Other substances may be added to the BGE to alter the selectivity. Modifiers may affect the charge on the micelle or the solute and change the distribution of the solute between the micelles and the surrounding aqueous phase by interacting with the solute. The following categories of additives are applicable in MEKC: CDs, ion-pair reagents, organic solvents, and others.

#### Cyclodextrins

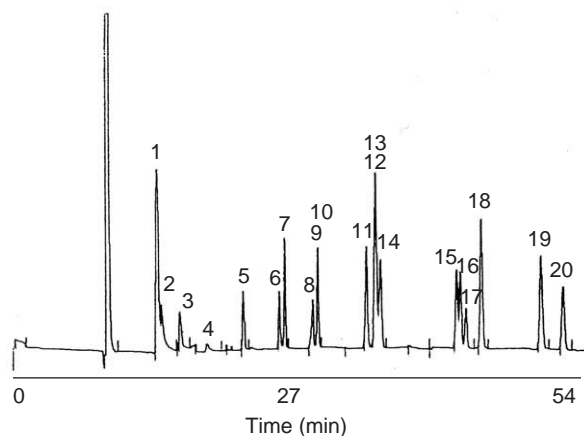
CDs are oligosaccharides with characteristic toroidal molecular shapes. Their outside surfaces are hydrophilic, while their cavities are hydrophobic. CDs tend to include molecules, which fit their cavities by hydrophobic interaction. The most widely used CDs consist of six ( $\alpha$ -CD), seven ( $\beta$ -CD), or eight ( $\gamma$ -CD) glucopyranose units. The size of the cavity differs significantly among the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -CDs. Selectivity results from the inclusion of a portion of a hydrophobic solute into the cavity. The use of CDs is especially effective for the separation of aromatic isomers and aromatic enantiomers that have the chiral center close to the aromatic ring. Many CD derivatives have been developed for increased solubility in water as well as for modifying the cavity shape that provides different selectivities from underivatized CDs and are commercially available.

In CD-modified MEKC (CD-MEKC), separation is achieved because of differences in the distribution of the solute among the micelle, CD, and the aqueous phase. The analyte molecule included by CD migrates at the same velocity as the EOF because

CD electrophoretically behaves as the bulk aqueous phase. The neutral analyte molecule migrates at the same velocity as the EOF, whether included with CD or free from CD. On the other hand, the analyte migrates at a different velocity from that of the EOF when it is incorporated into the ionic micelle. Therefore, the addition of CD reduces the apparent distribution coefficient or  $k$  and enables the separation of highly hydrophobic analytes, which otherwise would be almost totally incorporated into the micelle. The higher the concentration of CD, the smaller the distribution coefficient. In CD-MEKC, therefore, the retention factor can be manipulated by varying either the concentrations of the micelle or CD. Enhancement of the separation selectivity of a group of polycyclic aromatic hydrocarbons (PAHs) using CD-MEKC using mixed CDs is shown in Figure 4.

### Ion-Pair Reagents

Ion-pair reagents are also very useful for selectivity enhancement of both ionic and nonionic analytes in MEKC. For example, when a tetraalkylammonium salt is added to an SDS micellar solution, the migration times of anionic analytes increase with an increase in the concentration of the ammonium salt



**Figure 4** Separation of 20 PAHs using mixed CD-MEKC. Conditions: Capillary, 65 cm  $\times$  50  $\mu$ m inner diameter (50 cm length to the detector); injection by pressure, 0.02 min at 20 mbar, the concentration of each PAH in the mixture was 0.1  $\mu$ g/ $\mu$ l; UV detection at 230 nm; separation solution, 0.042 mol l<sup>-1</sup>  $\beta$ -CD, 0.026  $\gamma$ -CD, 2.5 mol l<sup>-1</sup> urea, and 0.100 mol l<sup>-1</sup> SDS in 0.140 mol l<sup>-1</sup> borate buffer (pH 9); applied voltage, 15 kV. Solute identification: 1, acenaphthene; 2, acenaphthylene; 3, naphthalene; 4, benzo[*g,h,i*]perylene; 5, fluorene; 6, phenanthrene; 7, pyrene; 8, chrysene; 9, perylene; 10, anthracene; 11, benzo[*e*]pyrene; 12, benzo[*a*]pyrene; 13, benzo[*a*]anthracene; 14, fluoranthene; 15, dibenzo[*a,h*]pyrene; 16, benzo[*k*]fluoranthene; 17, triphenylene; 18, benzo[*j*]fluoranthene; 19, benzo[*b*]fluoranthene; 20, indeno[1,2,3-*cd*]pyren. (Reprinted with permission from Jiménez B, Patterson DG, Grainger J, *et al.* (1997) *Journal of Chromatography A* 792: 411–418. © Elsevier.)

because the ammonium ion interacts with the anionic analyte to form a paired ion. Hence, the electrostatic repulsion between the anionic SDS micelle and the anionic analytes is reduced. On the other hand, a cationic analyte competes with the ammonium ion in interacting with the anionic SDS micelle, and so the migration times of cationic analytes decrease with an increase in the ammonium salt concentration. The effect of the ion-pair reagent on selectivity depends strongly on the structure of the reagent, e.g., the length of the alkyl chain.

### Organic Solvents

Organic solvents miscible with water are widely used as mobile phase modifiers in reversed phase high-performance liquid chromatography (RPLC). Organic solvents may also contribute to the enhancement of resolution or alteration of selectivity in MEKC. However, it should be noted that a high concentration of the organic solvent might break down the micellar structure. In general, the maximum content of the organic solvent is 20–30%. Methanol, 2-propanol, and acetonitrile are widely used as organic solvents in MEKC. The addition of organic solvents usually reduces the EOF and, hence, expands the migration time window. The retention factor is also reduced because the solubility of the analyte into the aqueous phase increases.

### Other Additives

A high concentration of urea is known to increase the solubility of the hydrophobic compounds in water. Urea also breaks down hydrogen-bond formation in the aqueous phase. In MEKC, urea slightly reduces the electroosmotic velocity and considerably increases the electrophoretic velocity of the micelle, resulting in an expanded migration time window. The addition of urea is also effective in improving peak shapes, especially in the separation of amino acid derivatives. The selectivity is not remarkably changed by the addition of urea, but minor changes are noticeable, especially for the separation of closely related analytes.

Another approach for changing selectivity in MEKC is the use of metal ions. In particular, the MEKC separation of oligonucleotides is improved by the addition of Mg(II), Cu(II), and Zn(II) ions. Metal ions are electrostatically attracted to the surface of a negatively charged micelle where they can be selectively complexed with analytes. Separation is due to differences in the distribution of solutes between the buffer and the metal-micelle surfaces. Retention is proportional to the complexation constant



of the oligonucleotide and the metal–micelle surface. Using different metal ions can change the selectivity.

## Improving Detection Sensitivity

The most widely used detector in MEKC, which is common to other CE modes, is an ultraviolet (UV) photometric detector because many solutes have absorb UV, and the UV detector is easily set up and is cost efficient. One of the drawbacks of UV detection in MEKC is the low concentration sensitivity resulting from a short optical path length, equal to the capillary diameter, and the small injection volumes needed to maintain high efficiency. This limits the applicability of MEKC to trace analysis. Thus, method development is indispensable for reducing the limit of detection (LOD) or increasing the concentration sensitivity. There are different approaches that have been reported for improving the concentration sensitivity in MEKC. These investigations involve the use of highly sensitive detection methods, the installation of capillaries equipped with extended detection path lengths, and offline, and online sample preconcentration methods.

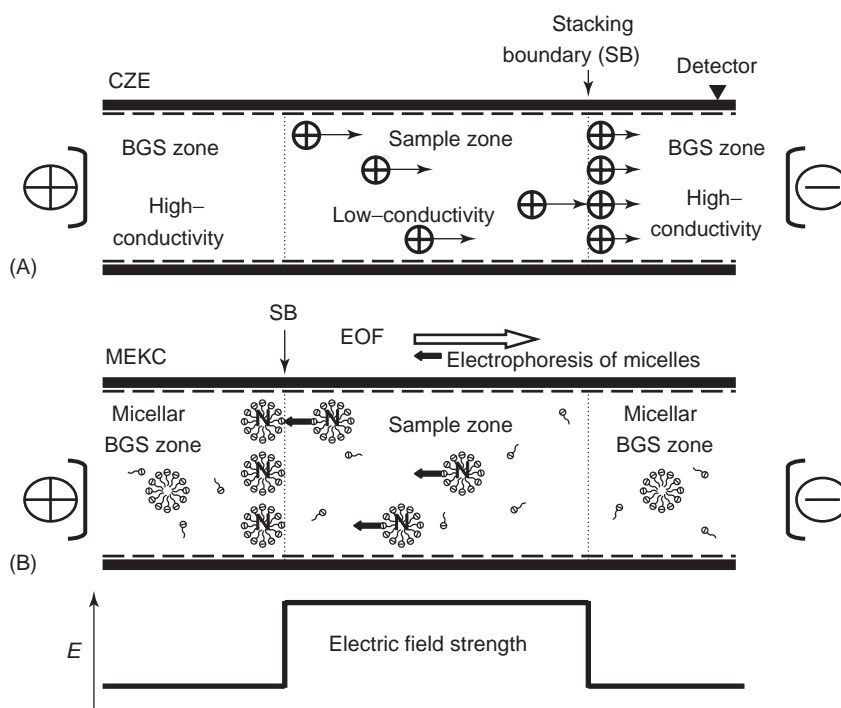
The frequently used highly sensitive detection method in MEKC is laser-induced fluorescence (LIF). LODs for this detection mode are at least an order of magnitude lower than that for UV detection. However, in an ordinary laboratory these detectors may not be affordable and UV detection is widely used as previously mentioned. From Beer's law, the absorbance is proportional to the path length, which is equal to the inner diameter of the capillary. Sensitivity can thus be enhanced through the use of larger inner diameter capillaries. However, this is not a good approach because large-diameter capillaries produce high currents, which cause Joule heating. The separation efficiency deteriorates dramatically when the capillary inner diameter is increased much above 100  $\mu\text{m}$ . Because of problems with large inner diameter capillaries, most approaches to improving the sensitivity have been directed toward extending the path length without increasing the inner diameter of the entire capillary. Bubble cell, Z-shaped cell, and multireflection cell detector configurations have been introduced, which have afforded from twofold to more than 10-fold enhancements in detector response. As with HPLC, a successful MEKC analysis would often rely on dedicated offline sample preparation methods (e.g., liquid–liquid and/or solid-phase extraction). Sample preparation methods usually increase the concentration of the analyte before injection into the capillary. However, these methods require somewhat complex or expensive hardware, with limited applicability, or are time consuming.

On the other hand, online sample preconcentration methods based on electrokinetic focusing of large sample volumes represent one of the most facile ways of sensitivity enhancement in MEKC since the preconcentration step is performed within the same capillary used for separation. Two of the most widely used techniques are sample stacking and sweeping. In general, the application of online preconcentration techniques in MEKC is dictated by the specific properties of the analyte and the sample matrix conditions. Thereafter, a variety of online sample concentration techniques have been reported in CZE. They are field-enhanced sample stacking, large-volume sample stacking, transient-isotachopheresis (ITP), pH-mediated stacking, and dynamic pH junction.

## Sample Stacking

The sample stacking technique was first introduced for ionic analytes in CZE. The basic model of sample stacking is illustrated in **Figure 5**. In CZE (**Figure 5A**), the sample is dissolved in a solution that has a lower conductivity than that of the BGE and injected as a longer plug than in normal injection. When a separation voltage is applied, the electric field strength experienced in the sample zone is higher than in the rest of the capillary. The sample ions in the sample zone will move quickly and then slow down when they reach the BGE zone because they experience a lower electric field strength. Consequently, the ions are focused at the boundary of the two zones. It should be noted that the EOF is assumed to be zero in **Figure 5A**. However, this technique cannot be applied to neutral analytes because neutral analytes have no electrophoretic mobility.

In MEKC (**Figure 5B**), to give effective electrophoretic mobilities to neutral analytes, charged PSs (e.g., anionic SDS micelles) are employed. The sample solution is prepared by dissolving the neutral analytes in a low-conductivity solution and injected into the capillary, which has been previously conditioned with an anionic SDS micellar BGE at neutral pH. The neutral analytes in the sample solution can be quickly carried to the boundary between the BGE and sample solution by the fast migrating anionic micelle entering the sample solution from the cathodic end. Since the electric field strength in the BGE zone is low, the velocity of the micelles is retarded, and the analytes are focused at the boundary between the BGE and the anodic end of the sample solution zone. Note that the neutral analytes are brought to the detector by the EOF since the magnitude of the EOF is greater than the electrophoretic velocity of the micelle.



**Figure 5** Schematic diagram of the principle of sample stacking in CZE (A) and MEKC (B).

However, a dispersive effect brought by the mismatch of local EOF velocities between the high and low electric field zones limits the injection length. To solve this problem, the stacking with reverse migrating micelles (SRMM) technique was developed. SRMM employs an acidic micellar BGE to reduce the EOF. Under this condition, the electrophoretic velocity of SDS micelles is higher than that of the EOF. Samples prepared in purified water or in low conductivity matrix are injected for a much longer time compared to the normal injection after conditioning the capillary with micellar BGE. Sample solutions are introduced at the cathodic end of the capillary, and then the separation voltage is applied with negative polarity at the injection end. Since the negative polarity is applied at the inlet, the sample matrix was slowly pushed out of the capillary by the weak EOF.

### Sweeping

The sweeping technique is defined as a phenomenon where analytes are picked up and concentrated by the micelle, which penetrates the sample zone devoid of micelle. It is independent of the EOF and effective for both charged and uncharged solutes.

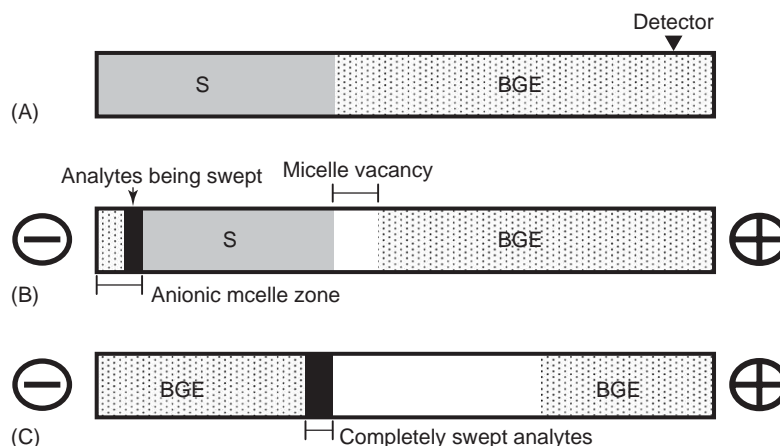
The principle of sweeping under the acidic condition is schematically shown in Figure 6. In step A, test analytes prepared in a matrix, with conductivity similar to that of BGE but devoid of PSs, are pressure-injected into the capillary at the cathodic end. In step

B, once the separation voltage is applied at the negative polarity with the BGE in the inlet vial, the anionic PS will enter the capillary and sweep the analytes. In step C, the analytes are completely swept by the PS and followed by MEKC separation in the reverse migration mode. Under the optimized condition, ~5000-fold enhancements in detection sensitivity were obtained in terms of peak heights by sweeping.

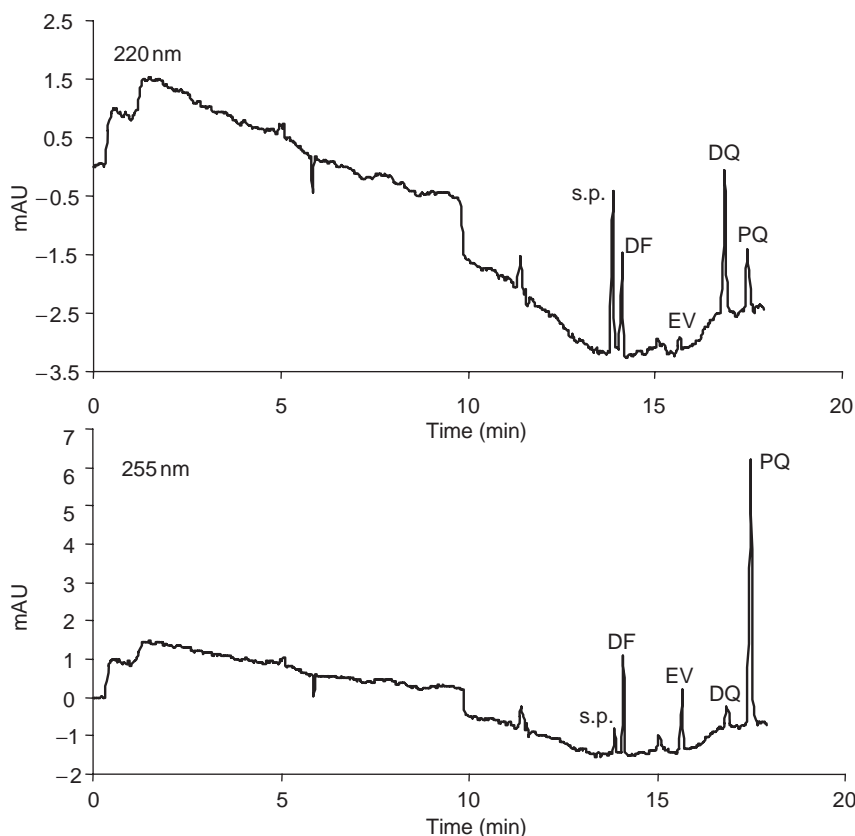
Cation- and anion-selective exhaustive injection sweeping (CSEI-sweep and ASEI-sweep) techniques are also available in MEKC. CSEI-sweep and ASEI-sweep are combinations of two online preconcentration techniques, sample stacking with electrokinetic injection and sweeping, that can provide more than 100,000-fold enhancements in detection sensitivity. An electropherograms of tap water analysis by CSEI-sweep-MEKC is illustrated in Figure 7. The results of this study suggest that CSEI-sweep-MEKC can be used for the analysis of quaternary ammonium herbicides in drinking water.

### MEKC Hyphenation Technique with Mass Spectrometry

The use of mass spectrometry (MS) as a detection method in CE as well as EKC offers several advantages over the UV detection method. Analytes having no strong UV absorption are detected with a high sensitivity by MS detection. Furthermore, MS provides important information not only about the



**Figure 6** Schematic diagram of the principle of sweeping under acidic conditions.



**Figure 7** Electropherogram of tap water spiked with quaternary ammonium herbicides analyzed using CSEI-sweep-MEKC. Non-micellar BGE, 100 mmol l<sup>-1</sup> phosphate buffer (pH 2.5) containing 20% acetonitrile; micellar BGE, 80 mmol l<sup>-1</sup> SDS in 50 mmol l<sup>-1</sup> phosphate buffer (pH 2.5) containing 20% acetonitrile; high conductivity buffer (HCB), 200 mmol l<sup>-1</sup> phosphate buffer (pH 2.5); conditioning solution before injection, nonmicellar BGE; sample concentration, 10 µg l<sup>-1</sup> paraquat (PQ), diquat (DQ), and ethylviologen (EV, internal standard), 50 µg l<sup>-1</sup> difenzoquat (DF). Injection scheme: hydrodynamic injection of HCB for 200 s (5 kPa), hydrodynamic injection of water for 6 s (5 kPa), electrokinetic injection of sample for 400 s (+22 kV); separation voltage, -22 kV with the micellar BGE at both ends of the capillary. s.p., system peak. (Reprinted with permission from Núñez O, Kim JB, Moyano E, Galceran MT, and Terabe S (2002) *Journal of Chromatography A* 961: 65–75. © Elsevier.)

molecular mass but also about the structure of the analytes. The most important issue in CE-MS may be the development of interfaces between CE and MS. MS has been successfully coupled to CE with various

interfaces. An electrospray ionization (ESI) interface is widely employed in CE-MS systems. However, the major drawback in MEKC-MS with an ESI interface is that the introduction of a nonvolatile PS into the

interface reduces the ionization efficiency and contaminates the interface. To solve this problem, the use of the partial filling (PF) technique has been developed. In the PF technique, the micellar solution is filled only in a part of the capillary, and the analyte will interact with the micelle when passing through the micellar zone. In PF-MEKC, the applied voltage for the MEKC run will be cut off before the micellar zone reaches the end of the capillary or the interface to the mass spectrometer, so that the micelle is not introduced into the mass spectrometer.

As an alternate ionization method to ESI, the atmospheric pressure chemical ionization (APCI) has been applied to the MEKC-MS system. In MEKC-APCI-MS, an SDS micellar solution can be introduced directly into the interface without a severe decrease in MS intensity. For highly sensitive analysis of environmental pollutants, an application study of sweeping to MEKC hyphenated with MS using an APCI interface has been reported.

See also: **Capillary Electrochromatography. Liquid Chromatography: Micellar. Solvents.**

## Further Reading

Khaledi MG (1997) *High-Performance Capillary Electrophoresis*. New York: Wiley.

Kim JB and Terabe S (2003) On-line sample preconcentration techniques in micellar electrokinetic chromatography. *Journal of Pharmaceutical and Biomedical Analysis* 30: 1625–1643.

Otsuka K and Terabe S (1989) Micellar electrokinetic chromatography. *Bulletin of the Chemical Society of Japan* 71: 2465–2481.

Quirino JP, Kim JB, and Terabe S (2002) Sweeping: Concentration mechanism and applications toward high sensitivity analysis in capillary electrophoresis. *Journal of Chromatography A* 965: 357–373.

Quirino JP and Terabe S (1999) Electrokinetic chromatography. *Journal of Chromatography A* 856: 465–482.

Terabe S (1989) Electrokinetic chromatography: An interface between electrophoresis and chromatography. *Trends in Analytical Chemistry* 8: 129–134.

Terabe S (1992) *Micellar Electrokinetic Chromatography*. Fullerton, CA: Beckman Instruments.

Terabe S (1992) Selectivity manipulation in micellar electrokinetic chromatography. *Journal of Pharmaceutical and Biomedical Analysis* 10: 705–715.

Terabe S, Chen N, and Otsuka K (1994) Micellar electrokinetic chromatography. *Advances in Electrophoresis* 7: 87–153.

Terabe S, Otsuka S, and Ando T (1985) Electrokinetic chromatography with micellar solution and open-tubular capillary. *Analytical Chemistry* 57: 834–841.

Terabe S, Otsuka K, and Ando T (1989) Band broadening in electrokinetic chromatography with micellar solutions and open-tubular capillaries. *Analytical Chemistry* 61: 251–260.

# MICRO TOTAL ANALYTICAL SYSTEMS

**T McCreedy**, University of Hull, Hull, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Miniaturized total analytical systems ( $\mu$ TAS) for chemical measurement have become important because they are reported to offer significant advantages over analytical systems of more conventional scale. These advantages are reported to include reduced reagent consumption and waste production, increased speed of analysis hence improved rates of sample throughput, and facilitating analysis in the field. There are other reported advantages with such systems, including the need for only very small volumes of sample (important where sample volume is limited) and disposability (particularly important for DNA replication work and clinical applications).

A common alternative name for  $\mu$ TAS is Lab-on-a-Chip. This article will describe typical  $\mu$ TAS, explain how such devices are made, and finally discuss a range of applications.

## Total Analytical Systems and $\mu$ TAS

Chemical analysis does not consist solely of a physical measurement process providing a numerical value, e.g., concentration of pesticide, or identification, e.g., the unknown substance was 3,4-dinitrophenol. The actual measurement is one small part of a far more complex process commencing with sample collection and concluding with data interpretation. The complete analytical process can include filtration, pH adjustment, dilution, reagent addition, calibration, in addition to making the physical measurement and data processing. It has long been recognized that automating all these steps would

interface reduces the ionization efficiency and contaminates the interface. To solve this problem, the use of the partial filling (PF) technique has been developed. In the PF technique, the micellar solution is filled only in a part of the capillary, and the analyte will interact with the micelle when passing through the micellar zone. In PF-MEKC, the applied voltage for the MEKC run will be cut off before the micellar zone reaches the end of the capillary or the interface to the mass spectrometer, so that the micelle is not introduced into the mass spectrometer.

As an alternate ionization method to ESI, the atmospheric pressure chemical ionization (APCI) has been applied to the MEKC-MS system. In MEKC-APCI-MS, an SDS micellar solution can be introduced directly into the interface without a severe decrease in MS intensity. For highly sensitive analysis of environmental pollutants, an application study of sweeping to MEKC hyphenated with MS using an APCI interface has been reported.

See also: **Capillary Electrochromatography. Liquid Chromatography: Micellar. Solvents.**

## Further Reading

Khaledi MG (1997) *High-Performance Capillary Electrophoresis*. New York: Wiley.

Kim JB and Terabe S (2003) On-line sample preconcentration techniques in micellar electrokinetic chromatography. *Journal of Pharmaceutical and Biomedical Analysis* 30: 1625–1643.

Otsuka K and Terabe S (1989) Micellar electrokinetic chromatography. *Bulletin of the Chemical Society of Japan* 71: 2465–2481.

Quirino JP, Kim JB, and Terabe S (2002) Sweeping: Concentration mechanism and applications toward high sensitivity analysis in capillary electrophoresis. *Journal of Chromatography A* 965: 357–373.

Quirino JP and Terabe S (1999) Electrokinetic chromatography. *Journal of Chromatography A* 856: 465–482.

Terabe S (1989) Electrokinetic chromatography: An interface between electrophoresis and chromatography. *Trends in Analytical Chemistry* 8: 129–134.

Terabe S (1992) *Micellar Electrokinetic Chromatography*. Fullerton, CA: Beckman Instruments.

Terabe S (1992) Selectivity manipulation in micellar electrokinetic chromatography. *Journal of Pharmaceutical and Biomedical Analysis* 10: 705–715.

Terabe S, Chen N, and Otsuka K (1994) Micellar electrokinetic chromatography. *Advances in Electrophoresis* 7: 87–153.

Terabe S, Otsuka S, and Ando T (1985) Electrokinetic chromatography with micellar solution and open-tubular capillary. *Analytical Chemistry* 57: 834–841.

Terabe S, Otsuka K, and Ando T (1989) Band broadening in electrokinetic chromatography with micellar solutions and open-tubular capillaries. *Analytical Chemistry* 61: 251–260.

# MICRO TOTAL ANALYTICAL SYSTEMS

**T McCreedy**, University of Hull, Hull, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Miniaturized total analytical systems ( $\mu$ TAS) for chemical measurement have become important because they are reported to offer significant advantages over analytical systems of more conventional scale. These advantages are reported to include reduced reagent consumption and waste production, increased speed of analysis hence improved rates of sample throughput, and facilitating analysis in the field. There are other reported advantages with such systems, including the need for only very small volumes of sample (important where sample volume is limited) and disposability (particularly important for DNA replication work and clinical applications).

A common alternative name for  $\mu$ TAS is Lab-on-a-Chip. This article will describe typical  $\mu$ TAS, explain how such devices are made, and finally discuss a range of applications.

## Total Analytical Systems and $\mu$ TAS

Chemical analysis does not consist solely of a physical measurement process providing a numerical value, e.g., concentration of pesticide, or identification, e.g., the unknown substance was 3,4-dinitrophenol. The actual measurement is one small part of a far more complex process commencing with sample collection and concluding with data interpretation. The complete analytical process can include filtration, pH adjustment, dilution, reagent addition, calibration, in addition to making the physical measurement and data processing. It has long been recognized that automating all these steps would

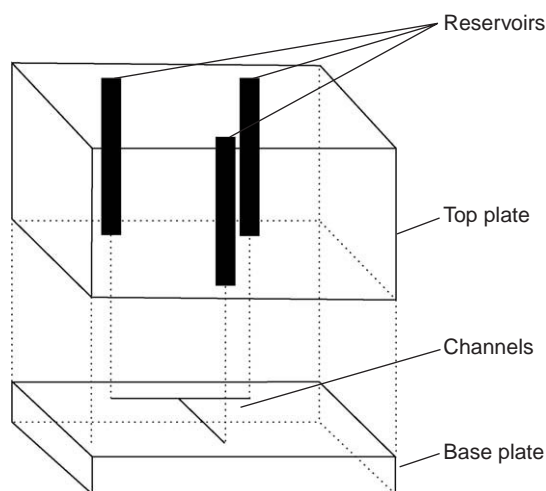


make analysis quicker and less expensive, and total analytical systems (TAS) have been developed to do just that. TAS are instruments that can process a sample, ideally from the point of collection to the production of the analytical report, and include all the sample preparation, instrument calibration, and measurement. Unfortunately, the devices are usually large and expensive to operate, i.e., high consumption of reagents and sample, production of large amounts of waste, and are mechanically complex.

It has long been recognized that reducing the scale of the chemistry (this is not the same as making the instrument smaller) reduces the reagent consumption, increases sample throughput, and shortens reaction times. These advantages, along with developments in fabrication techniques from the electronics industry led to the reality of  $\mu$ TAS; planar systems providing fluidic manipulation and measurement on a small scale (typically in a network of channels of 100  $\mu$ m width or less). The idea of miniaturized instruments was first reported in the 1950s when Terry and co-workers reported the fabrication of a gas chromatograph on a silicon wafer. This had an integral detector and was perhaps the first reported  $\mu$ TAS. After that, while some interest remained, little progress was made until the 1980s when Manz started to develop the idea and interest was rekindled. Such was the potential importance; interest within the scientific community grew very rapidly. During the following decade interest continued to grow and diversify from simple fluidic manipulations and electrophoretic separations to chemical synthesis and reaction chemistry. Devices were often referred to as 'chips' and the term Lab-on-a-Chip was coined. In this article, the focus will be on Lab-on-a-Chip devices for analysis rather than chemical microreactors for synthetic applications.

## Fabrication of Devices

The majority of the fabrication techniques used to manufacture  $\mu$ TAS or Lab-on-a-Chip devices has been developed from the microelectronics industry. A wide range of techniques are available and a comprehensive treatment is beyond the scope of this article. A typical device consists of two layers or plates: the base plate and the cover plate. The channel network is formed in the base plate by a suitable micromachining method. The network is closed by the cover plate, which also contains the fluid transfer connections or the reservoirs and possibly the electrodes. The two plates are either pressed firmly together or irreversibly bonded to form liquid tight seals. A diagram of a typical device can be seen in Figure 1. Devices with multiple channel layers are



**Figure 1** Diagram showing a typical chip. The channel design consists of two intersecting channels with reservoirs at the ends of the channel.

being developed, where the different channel layers are connected via microflow control valves.

## Materials Used

There are a number of materials used for the fabrication of  $\mu$ TAS devices. Perhaps the most common is glass due to its low cost, ease of machining, and suitability for electrophoresis and electroosmotic flow (EOF) applications without requiring surface modifications. It is also chemically inert to most reagents (apart from hydrofluoric acid and concentrated alkali). Silicon is also a valuable material that has similar chemical inertness and can easily be machined by chemical etching. While it is more expensive, it can be easily chemically etched to yield far higher aspect ratios than are possible with glass. Silicon is not suitable for electrophoresis or EOF applications without surface pretreatment. Devices fabricated from polymers such as polymethylmethacrylate (PMMA) and polydimethylsiloxane (PDMS) are also frequently used due to the low cost of the material (especially important for disposable devices) and the ease of fabrication. Perhaps one drawback with polymers is their incompatibility with solvents. They are suited to electrophoretic applications but frequently require surface modification to support EOF. Occasionally, metals are used; however, these are far more frequently encountered in chemical microreactors.

## Micromachining Methods

The choice of machining method is somewhat material dependant, however, wet chemical etching in conjunction with photolithography is a highly

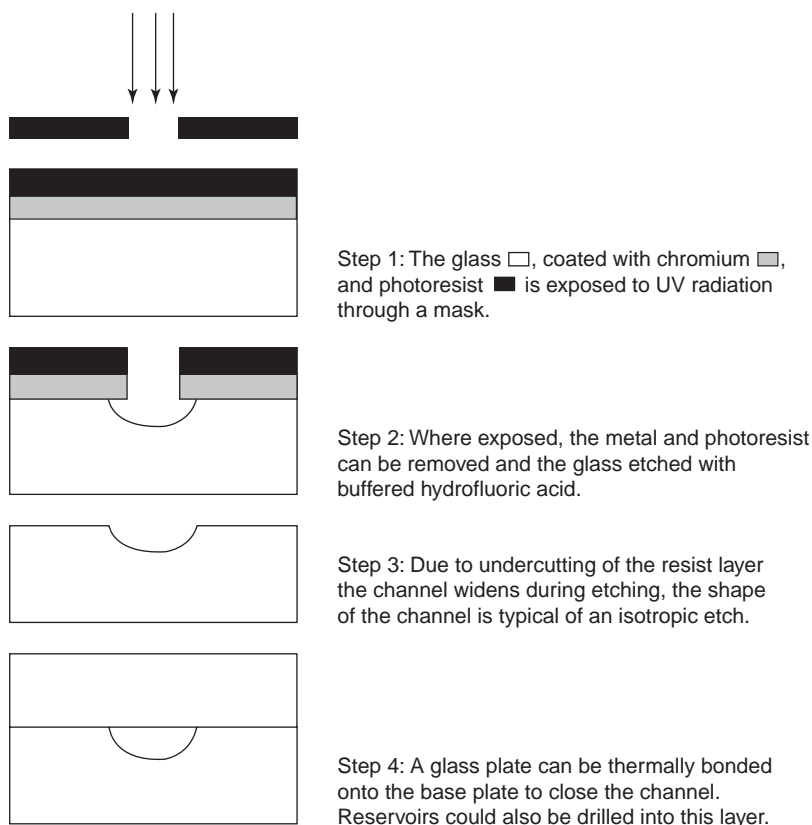
versatile method suitable for both glass and silicon, and can be readily employed in most laboratories. Molding and hot embossing are machining methods reserved for polymers and permit the replication of hundreds of devices from a single master; this can be of significant importance where disposable devices are required. More specialist fabrication techniques include lamination, laser ablation, and dry etching methods using plasmas and reactive ion beams. While all the above methods require some degree of specialist facilities, the latter methods tend to be more specialist and will not be discussed further.

### Wet Chemical Etching and Photolithography

This machining technique (shown schematically in Figure 2) involves etching a channel network into a glass or silicon substrate, the pattern of the channel network is defined by selectively removing areas of a resist layer on the surface. Glass or silicon is first coated with a layer of photoresist, typically  $0.1\text{--}2\text{ }\mu\text{m}$  thick by spin coating. (It should be noted that far thicker layers can be obtained if specialist photoresists are used.) Glass substrates are frequently coated in a thin layer ( $>100\text{ }\text{\AA}$ ) of chromium and

gold prior to the application of the photo resist. The pattern of the required channel network is produced as a photo mask, which is placed between the light sensitive layer and a source of radiation, usually an ultraviolet (UV) source. In a similar way to a photograph, upon exposure, the pattern is transferred to the layer of photoresist. It is also possible to use a direct writing laser to pattern the resist. The exposed resist can be removed using developer and the substrate surface exposed beneath. When the glass is exposed to certain chemicals (etchants), channels are etched into the glass. The etch depth is time dependent. The resultant channel profile will be dependant upon whether the etch process is isotropic or anisotropic.

**Isotropic versus anisotropic etch** Glass is always etched isotropically because the etchant etches all exposed surfaces at the same rate. This is independent of the etchant used and arises because glass does not have different planes. Depending on etchant, silicon is either etched isotropically or anisotropically. Anisotropic etching occurs where material on one plane is removed at a greater rate than material on a different plane. The difference is most apparent in terms of channel shape and aspect ratio. Isotropic

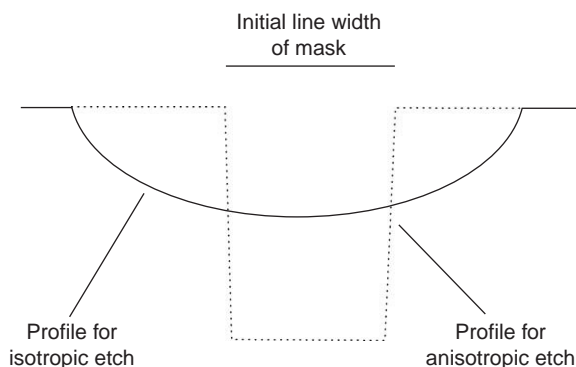


**Figure 2** A schematic representation of device fabrication by photolithography and wet chemical etching of a glass substrate.

**Table 1** The influence of some common etchants

<i>Etchant</i>	<i>Substrate</i>	<i>Effect</i>
HF/HNO <sub>3</sub> <sup>a</sup>	Glass	Isotropic etch
HF/HNO <sub>3</sub> <sup>a</sup>	Silicon	Anisotropic etch
KOH, NaOH, NH <sub>4</sub> OH	Glass	Little effect, isotropic etch if any
KOH, NaOH, NH <sub>4</sub> OH	Silicon	Anisotropic etch

<sup>a</sup>Acetic acid is the preferred diluent since it better maintains the oxidizing properties of HNO<sub>3</sub>.



**Figure 3** The channel profiles for a channel etched isotropically and anisotropically into a silicon substrate. The same mask line width was used for both.

etches form rounded channels with low aspect ratio, e.g., a channel produced in glass by acid etching will have a width equal to the original mask line width plus twice the depth. A channel anisotropically etched into silicon produces deep channels with near vertical sides and high aspect ratios. The influences of etchant on the different materials can be seen in Table 1 and the shapes of channels produced by isotropic and anisotropic etching can be seen in Figure 3.

### Molding and Hot Embossing

These methods of micromachining are ideally suited to polymeric material. Molding is frequently used with polymeric resins such as PDMS that are initially liquid and become solid on heating. A master is first fabricated from silicon, glass, or metal such that the channel network is formed as a series of ridges on the surface. When the polymeric resin is allowed to cure over this, a channel network is formed in the mating face of the polymer. Hot embossing is used to form a channel in a rigid polymer such as PMMA, where pressing a pattern into the surface of the polymer forms the channel network.

## Fluidic Movement within the Device

The method of inducing bulk flow of fluid within the device can take two general forms; pressure driven flow and electroosmotically induced flow. Both methods have advantages and disadvantages and are somewhat dependent on the construction material. Although there are many methods of inducing pressure driven flow, the ultimate flow characteristics are similar.

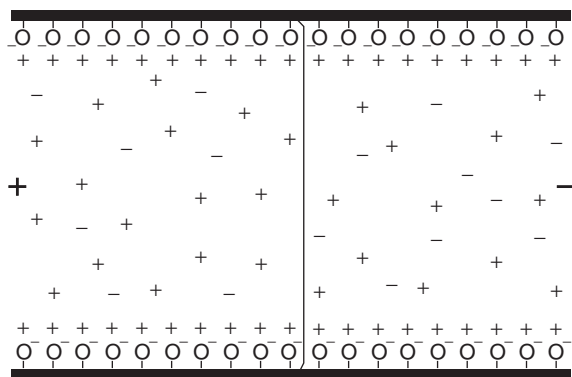
### Pressure-Driven Flow

Over the years, there have been many attempts to develop microfabricated pumps with varying degrees of success. Such devices were based on piezoelectric wafers, etc. Their success was variable and while efforts continue, there are many limitations. Syringe pumps can deliver flow rates typically between 0.1 and 1  $\mu\text{L min}^{-1}$  and are frequently used with  $\mu\text{TAS}$ . The flow is virtually pulse free and, due to the low flow rates used, the problem of refilling the syringe is less significant. An alternative is to use microperistaltic pumps; however, the pulsation can be very high. Pressure-driven flow is not influenced in any way by the material of construction. In all cases, the flow profile is parabolic and can lead to diffusion or turbulent flow, a problem not encountered with EOF. Volume flow rates through the channel are directly proportional to the applied pressure and the cross-sectional area of the channel, but inversely proportional to the length of the channel and the viscosity of the buffer. It is clear that higher pressures will be required to achieve the same volume flow rate through long channels with small cross-sectional area compared to short channels of large cross-sectional area. This can lead to restrictions in channel dimensions (both length and cross-section) when pressure-driven flow is used.

### Electroosmotic Flow

This approach to fluidic propulsion does not rely on mechanical pumping; rather, it is a surface-induced phenomenon. It requires an electrical insulator as the channel material and the ability of the surface to form a double layer. This tends to limit the applicability to glass, surface modified silicon, and surface modifier polymer. EOF is best described by considering the situation occurring in a buffer-filled channel in a glass substrate.

Consider a channel network fabricated in a glass substrate and filled with an aqueous buffer of pH10. The glass surface is covered with a layer of silanol (Si-OH) groups and at this pH; many are dissociated (Si-O<sup>-</sup>). A layer of cations develops near to this surface since they are attracted by the negative



**Figure 4** The formation of the electrical double layer and the flow profile resulting from the application of an applied voltage along the channel.

charge on the surface. In the absence of an applied voltage, nothing else happens. When a voltage is applied along the channel, it causes the cations to migrate toward the negative electrode. This process results in bulk fluid movement (EOF) toward the negative electrode. Unlike pressure-driven flow, the flow profile is laminar, leading to very little turbulence. This process can be seen schematically in **Figure 4**. The applied electric field is usually delivered from electrodes located in the fluid reservoirs. The magnitude of the flow (EOF<sub>m</sub>) is related to the charge on the surface of the channel (often referred to as the Zeta potential ( $\zeta$ )), the viscosity of the buffer ( $\eta$ ), and the dielectric constant of the buffer ( $\epsilon$ ):

$$\text{EOF}_m = \epsilon \zeta / \eta$$

Due to the ionization of the silanol groups, the Zeta potential is strongly influenced by the pH of the buffer; the pH tends to have the greatest influence on flow and therefore needs to be carefully controlled. EOF also increases with increasing field strength, i.e., applied voltage per centimeter. The length and cross-sectional area has less influence on flow rate when EOF is used to pump fluids through channels; hence, this can have significant advantages in certain designs of device.

## Analytical Applications

There have been a vast number of publications in the general area of Lab-on-a-Chip research; however, most devices reported are not true  $\mu$ TAS. Earlier,  $\mu$ TAS were defined as devices that performed the total analytical process, from sample preparation to analysis. Many systems reported perform only some of these functions; frequently the sample preparation aspect is omitted. It remains common practice to use

the term  $\mu$ TAS for systems that do not meet all the steps in the analytical process but are miniaturized devices performing some of the analytical steps.

In the simplest mode of operation,  $\mu$ TAS can be used as a miniature flow injection system either to carry samples from the point of injection to the point of detection with the possible addition of reagents into the stream. Fluid manipulation can be achieved by pressure pumping, EOF, or electrophoresis (though not resulting in bulk fluid movement, it can be very effective for moving ionized species through the bulk). One early example of an application of  $\mu$ TAS as a flow injection system was the determination of orthophosphate by the molybdenum blue reaction. Here a water sample was introduced into a reagent stream by EOF where the colored molybdenum blue complex was formed and the absorbance measured further along the channel. Such systems take advantage of the low reagent consumption and ease of sample manipulation on small planar structures.

The channel sizes typical of  $\mu$ TAS make them ideal for manipulating and analyzing cells. Cells can be transported through the channels by pressure pumping on EOF, then lysed prior to analysis of contents.  $\mu$ TAS has the unique advantage that since the volumes are small, dilution of the cell contents does not occur to any significant extent during the lysing process. This is useful for all cells and microorganisms, but can offer particular advantages when applied to other microorganisms of a more harmful nature.

## Extraction, Preconcentration, and Matrix Removal

The incorporation of extraction methodologies into  $\mu$ TAS can allow unprepared samples to be analyzed. The simplest step is to fabricate filtration units, typically consisting of small sections of shallow channels or porous plugs at the start of analytical channels. These serve to remove particulates from the sample; therefore, unfiltered samples can be introduced with no detrimental effect on the performance of the system. Often, samples can be particulate free but require extraction of the analyte from the matrix, either for preconcentration or for matrix removal. There have been several reports of liquid-liquid extraction being fabricated into  $\mu$ TAS. Usually, two immiscible streams (the sample and the extracting fluid) are pumped down a single channel, where the extraction takes place at the fluid interface. The streams are then separated by splitting the channel. Mixing of the streams does not occur since the flow is laminar; this, however, results in a less efficient extraction system since it is diffusion limited. An

alternative strategy is to incorporate solid phase extraction material into the channel, either as a solid or wall coated to preconcentrate and/or remove interfering species from the sample.

### Separation

This aspect of  $\mu$ TAS is considered by many to be one of the most important aspects permitting separation of complex analytical samples into separate components, eliminating the need for selective detection. While many different strategies have been reported, the two most common are based on electrophoresis and chromatography. In both cases, small volumes of sample are injected into a channel and the various components in the sample are separated on the basis of their electrophoretic mobility or their interaction with a stationary phase. The size of the sample plug will be vital to the reproducibility of the separation, both in terms of resolution and quantification.

### Chromatography

The separation is achieved by pumping a small plug of analyte through a channel either packed, or wall coated with stationary phase. Packed channels yield high back pressures, which is a significant problem with pressure pumped systems. EOF overcomes this problem to a large extent, and the effect observed is comparable to that occurring in capillary electrochromatography. While wall coated stationary phase does not suffer from the back pressure problem, it does lack sample capacity and overloading is likely. This can only be addressed by ensuring that the minimum sample volume is introduced. An alternative to these strategies is to fabricate monoliths in the channel, e.g., by specialist etching of silicon and subsequent chemical modification, to effectively increase the surface area of the stationary phase without filling the channel with small particles. There is increasing use being made of separations based on chromatography particularly where preconcentration and separation can be achieved in a single channel. The first part of the packing can be used to preconcentrate the analyte prior to elution down the remainder of the column with an appropriate solvent.

### Electrophoresis Separations

Electrophoresis separations rely on differing electrophoretic mobilities of the analytes, so clearly this separation approach is only valid for separating charged analytes. The most popular type of separation is based on straightforward differences in migration mobilities of the analytes, i.e., as in capillary

electrophoresis. Where a voltage is applied along the length of a channel, the ions migrate under the influence of that charge. The electrophoretic velocity ( $v$ ) of a given ion is related to the charge on the ion ( $q$ ), the solution viscosity ( $\eta$ ), the radius of the ion ( $r$ ), and the applied electric field ( $E$ ) by the equation below:

$$v = E(q/6\pi\eta)r$$

Since the electrophoretic mobilities (and hence the velocities) are dependent on the radius and charge of the ions, they will become spatially resolved along the channel if the mobilities of each species are sufficiently different, therefore reaching the detector at different times. Since the electrophoretic mobility of all uncharged materials is effectively zero, this approach cannot be used to separate uncharged species. While this approach has been used for many applications, perhaps its most important is its use in separating small biomolecules and fragments of DNA. Analysis of DNA can rapidly be performed on  $\mu$ TAS, offering advantages over other separation techniques of speed, minute sample volume, and disposability. The latter advantage is particularly important where contamination must be avoided.

Another frequently employed electrodriver separation process is isotachopheresis, again based upon differing rates of migration in an applied electric field, but unlike capillary electrophoresis methods employs a leading and tailing electrolyte. All separated species are contained between the two electrolyte bands, thus ensuring all the analytes reach the detector in a controlled period. This is somewhat less frequently used than capillary electrophoresis.

**See also: Capillary Electrochromatography. Capillary Electrophoresis:** Pharmaceutical Applications; Low-Molecular-Weight Ions; Environmental Applications; Food Chemistry Applications; Clinical Applications. **Electrophoresis:** Overview.

### Further Reading

- Fletcher PDI, Haswell SJ, Pombo-Villar E, *et al.* (2002) Micro Reactors: Principles and Applications in Organic Synthesis. *Tetrahedron* 58: 4735–4757.
- Madou M (2000) *Fundamentals of Microfabrication*. Boca Raton, FL: CRC press.
- Microreaction Technology: Proceedings of the Fifth International Conference on Microreaction Technology*, IMRET 5, 2001. Berlin: Springer.
- Oosterbroek E and van den Berg A (2003) *Miniaturised systems for (Bio)Chemical Analysis and Synthesis*. Amsterdam: Elsevier.



# MICROBIOLOGICAL TECHNIQUES

**G L Pettipher**, Reading Scientific Services Limited, Reading, UK

**J M Jay**, University of Nevada Las Vegas, Las Vegas, NV, USA

**H H Wang**, Ohio State University, Columbus, OH, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

There are many different types of microbiological analyses available to the modern microbiologist. These are based on a wide variety of detection principles and range from direct to indirect, manual to automated, fast to slow, and total to specific. The choice of method will be largely affected by the specific application and the user's requirements in terms of determinant and sensitivity. In general, the methods can be used for a wide range of applications, e.g., food, pharmaceutical, water, environmental, clinical, etc. Microbiological methods can be used to enumerate 'total numbers' of bacteria, yeasts, and molds in test materials, 'indicator organisms' of fecal or other contamination, and pathogens that may cause food poisoning or even death. In addition, there are methods for identifying bacteria, yeasts, and molds isolated or enumerated by these techniques. Although a large number of methods are addressed, space limitations do not allow for sufficient descriptions to enable one to discuss most of those covered. For this purpose, the reader should consult one or more of the listed references for the method of choice that pertains to the use or application of the results. Finally, the methods addressed are primarily for chemoheterotrophic bacteria, commonly occurring yeasts and molds, and some viruses and protozoa that cause human gastroenteritis.

## Types of Methods

Methods for estimating microbial numbers can be divided into two groups: direct and indirect. Direct methods count cells directly, either by the ability of viable cells to grow and form colonies, or microscopically. Indirect methods measure a chemical constituent, enzyme, metabolite, or changes produced by organisms during growth. This measurement is converted into numbers by reference to a calibration graph. The 'true' numbers of organisms for the calibration graph are usually assessed by direct methods. Generally, direct methods are more sensitive and accurate than indirect methods and cultural methods

take longer to give a result than microscopic or indirect methods.

While not developed for the enumeration of microorganisms, a number of methods have emerged that can be used to type, characterize, or 'fingerprint' microbial species or strains. These techniques are very valuable in studying the epidemiology of pathogens and in profiling microorganisms that occupy common environmental niches. The bioassay methods are important in assessing certain virulence properties among human and animal pathogens.

## Direct Methods

**Resuscitation and enrichment methods** Some bacteria and fungi that have been frozen, dried, sublethally heated, or exposed to chemical inhibitors may suffer sublethal damage and not replicate when placed on growth media, especially media that contain selective inhibitors. If these cells are not allowed to resuscitate, under-counts may result and, in the case of pathogens, a product may be declared pathogen-free when in fact it is not.

When food or environmental samples are believed to contain numbers of the organism of interest that are below the detection limit of a given viable enumeration method, nonselective enrichment methods may be used. This consists of the addition of, e.g., 25–50 g of sample to a general growth broth to allow the low number of cells to increase. Not only can this increase the original low number to a detectable level, it will allow environmentally stressed cells to resuscitate. Selective enrichments employ agents in the medium that favor a given species or group. For example, a general culture medium that contains 7.5% NaCl becomes selective for *Staphylococci*. Metabolically stressed cells should be allowed to resuscitate in a nonselective broth (e.g., 4–6 h in trypticase soy broth at refrigerator temperatures).

The most widely used colony count methods are further discussed below along with some parameters that affect colony results.

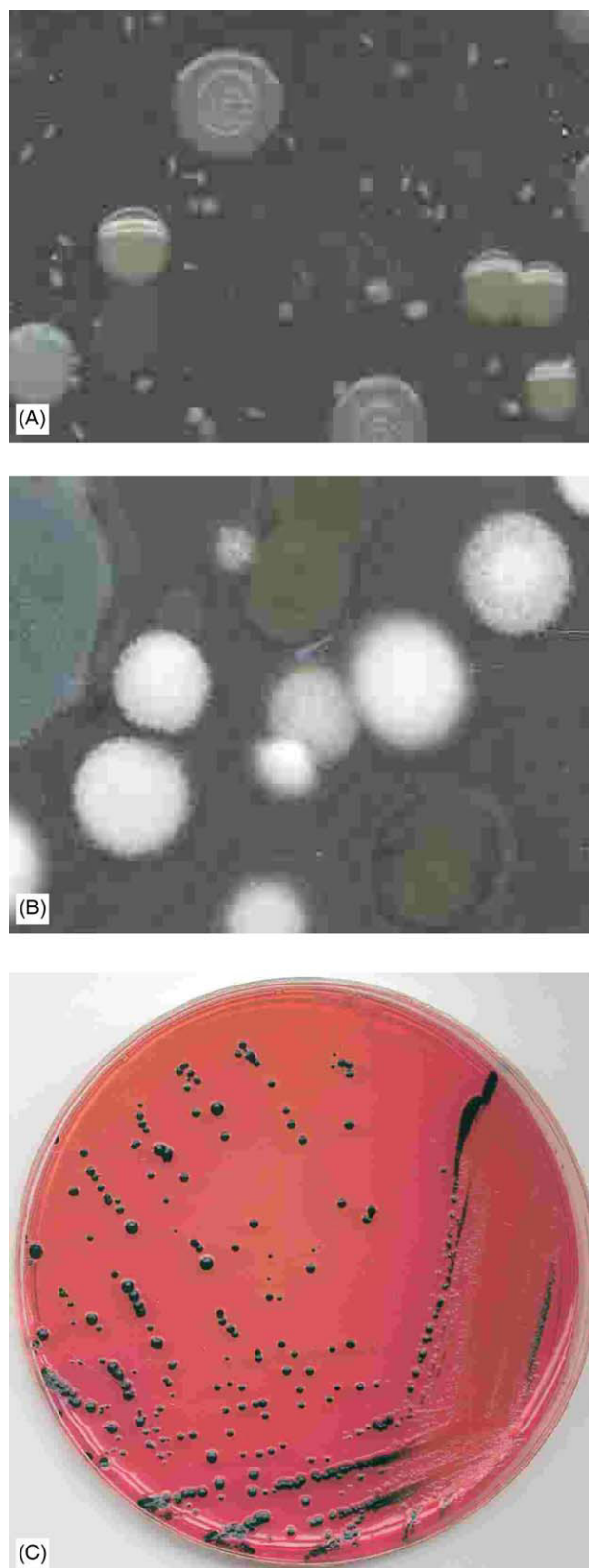
**The pour plate** The plate colony count isolates bacteria, yeasts, or molds in a quantitative manner. A range of dilutions, usually 10-fold of the sample, is prepared in a sterile diluent and 1 ml of each dilution is added to sterile Petri dishes, mixed with melted agar cooled to 45°C, and then allowed to solidify. The organisms present in the sample are fixed within the agar gel. Pour plates are then incubated for 2–5 days at 25–37°C, depending on the application,

during which time viable organisms replicate to form visible colonies. Individual colonies on a plate containing 15–150 colonies are counted and then the count per milliliter or per gram is calculated by multiplying the colony count by the appropriate dilution factor (Figure 1). One of the major disadvantages of the plate colony count is that it takes 2–5 days to give a result. The plate colony count is very sensitive because, in principle, any viable cell when placed in an appropriate medium and incubated at a suitable temperature will give rise to a colony. It is important to note that the plate colony count is subject to errors, and provides only an estimate of the ‘true’ numbers of organisms. These errors, which are mainly due to sampling and decrease with increasing bacterial numbers, are often in the range  $\pm 10\%$  to  $\pm 40\%$ . Automation of the plate colony count enables an operator to perform analyses more rapidly than by the manual method. However, for plate colony count methods, the length of time taken to give the result is mostly due to the incubation period required for colony formation. Automation of the preparation of the plate does little to decrease the overall time required for the technique. Some examples of colony count and related methods for various organisms are summarized in Table 1.

Savings in time and materials can be made by removing the need for the dilution series in the standard plate colony count. Various methods based on the Thompson plate loop method have been described. In the technique, two loops that retain 0.01 and 0.001 ml are dipped into the sample. A Petri dish is positioned under each loop and the loops are flushed with diluent, agar is then added, and the contents of the Petri dish are mixed. These machines can plate  $10^{-2}$  and  $10^{-3}$  dilutions from 300 samples per hour. The technique is suitable for enumerating bacteria in the range  $3000\text{--}300\,000\text{ ml}^{-1}$ . Manufacturers of one commercially available automated plate loop machine (PetriFoss) claim that the relationship with the reference method has a correlation coefficient of 0.99.

**Surface plating** Several methods are now in use that lead to the growth of all colonies on the surface of the culture medium – agar plate, dry film, hydrophobic grid membrane filter (HGMF), etc.

Sample preparations for the agar surface plate are the same as for pour plating except that smaller, e.g., 0.1 ml, volumes are used on plates. Plates are poured with the appropriate medium a day or two in advance of use so that the surface is dry. After 0.1 ml of serial dilutions is planted, they are spread over the entire plate surface with a sterile bent glass or plastic rod (‘hockey stick’). Surface plating is



**Figure 1** Colonies of microorganisms on agar plates after incubation: (A) close-up of bacteria; pour plate, plate count agar; (B) close-up of mold; surface plate, oxytetracycline glucose yeast extract agar; (C) *Salmonella* species; streak plate, xylose lysine decarboxylase agar.

**Table 1** Examples of some conventional colony count methods that may be used to enumerate microorganisms from a variety of samples

Organisms/groups	Culture media <sup>a</sup>	Incubation	Time for results
Anaerobic bacteria	AC agar or PCA	30–35°C	48–72 h
Aerobic sporeformers	PCA after heat shock	35°C	24–48 h
<i>Bacillus cereus</i>	MYP or KG agar	35°C	24 h
<i>Campylobacter jejuni</i>	CCDA-Preston blood-free medium; Campy-Cefex agar	42°C	18–48 h
<i>Clostridium perfringens</i>	SPS, SFP, or TSC agar	35°C	24 h
Coliforms	VRB agar	35°C	24–48 h
<i>Escherichia coli</i>	Petrifilm	35°C	24 h
<i>E. coli</i> O157:H7	Sorb. MacConkey agar	35°C	24 h
Enterobacteriaceae	VRB glucose agar	35°C	18–24 h
Enterococci	KF Strep, FGTC agar	35°C	24–48 h
Fungi (yeasts, molds)	DRBC, Sabouraud, or PCA + antibiotics	22–25°C	3–5 days
Lactic acid bacteria	MRS or KF Strep agar	35°C	48 h
<i>Listeria monocytogenes</i>	LPM or modified McBride agar	35°C	24–48 h
Psychrotrophs	PCA	5–7°C	10 days
Psychrophiles	PCA	< 12°C	3–5 days
Staphylococci	Baird-Parker agar	45°C	24 h
Thermophiles	PCA	55°C	24 h

<sup>a</sup>CCDA, *Campylobacter*-charcoal-differential agar; DRBC, Dichloran Rose Bengal Chloramphenicol; FGTC, fluorogenic gentamicin-thallos-carbonate agar; LPM, lithium-phenylethanol-moxalactam agar; PCA, plate count agar; SPS, sulfite-polymyxin-sulfadiazine; TSC, typtose-sulfite-cycloserine; VRB, violet red bile.

Many or most of these determinations can also be made by use of Petrifilm, HGMF, or SimPlate.

preferred to pour plating for the enumeration of psychrotrophs and psychrophiles. Also, it allows one to observe the size, texture, and pigmentation of all colonies if they are well isolated. Use of the spiral plater is an efficient way to prepare pour plates, and this device is described below.

The spiral plater method for enumerating organisms also avoids the use of a dilution series. By using a varying rate of sample application, it needs only one plate to obtain counts over a range that will require two or three plates in the standard plate colony count. The spiral plater deposits a known volume of sample on a rotating plate in an ever decreasing amount in the form of an Archimedes spiral. After incubation, different colony densities are apparent, closely packed, or confluent in the center to well isolated at the outside. A counting grid that relates the area of the plate examined to the volume of the sample is used to convert the count in a given area of the plate to the numbers of organisms per milliliter or per gram of sample. This can be done automatically in a few seconds by a laser colony counter or image analyzer. Using the spiral plater and laser colony counter, samples containing organisms in the range 500–300 000 ml<sup>-1</sup> can be analyzed in about one-third the preparation time of the standard plate colony count.

**Most probable number (MPN)** This is a widely used culture procedure that provides a statistical estimate of the MPN of viable organisms in a sample.

It is commonly used in the examination of waters and foods for coliforms and *Escherichia coli*. The three-tube method employs nine tubes to which three successive decimal dilutions of specimen are planted into three tubes each. The five-tube method is the same except that five tubes of each dilution are inoculated. After a suitable incubation, positive tubes of each dilution are recorded and used to arrive at the MPN by use of standard three- and five-tube tables.

**Dry films** The most widely used of these methods is Petrifilm<sup>TM</sup>, which consists of a particular culture medium semidried onto a plastic sheet covered with a see-through film. The Petrifilm APC contains triphenyltetrazolium chloride, which aids in viewing colonies. These Petrifilm ‘plates’ are inoculated with 1.0 ml of diluted specimen and then incubated at the appropriate temperature. Selective Petrifilm ‘plates’ are available for coliforms, *E. coli*, and fungi. These methods give results comparable to those obtained by pour or surface plating.

The SimPlate<sup>TM</sup> is a method that resembles a plating method in its setup but individual colonies are not enumerated. It can be viewed as an MPN method that uses special plates rather than tubes. The special plates contain 84 or 198 wells and a diluted sample is added to a single plate. Broth is then added and distributed evenly into the wells with excess medium being decanted. After incubation for 24 h, plates are exposed to ultraviolet light and the fluorescent wells are counted. Special MPN tables are used to

calculate numbers per gram or per milliliter in the original sample. Selective SimPlate media are available for coliforms and *E. coli*.

**Microscope slide methods** A specimen smear on a glass slide followed by simple or differential staining remains one of the fastest ways to determine numbers of bacteria or yeasts. A minimum number of cells of  $\sim 10^4 \text{ cm}^{-2}$  are needed. Both viable and nonviable cells may be enumerated and results can be obtained in  $\sim 5$  min. A number of slide counting devices exist with the Breed slide method being one of the most widely used. The Breed smear was one of the first methods to use microscopy for counting bacteria in milk and involved the preparation of milk films, staining with Methylene Blue, and then microscopic examination.

**Direct epifluorescent filter technique (DEFT)** The DEFT was originally developed as a rapid method for counting bacteria in raw milk and is now used for a range of food, water, pharmaceutical, and medical applications. The DEFT uses membrane filtration, fluorescent staining, and epifluorescence microscopy. Some products require treatment with an enzyme and surfactant to enable a sufficient amount to filter through a  $0.6 \mu\text{m}$  polycarbonate membrane filter. Microbial cells remain intact and are concentrated on the surface of the membrane. After staining with Acridine Orange [3,6-bis(dimethylamino)acridinium chloride], cells fluoresce orange-red under blue light. 4',6-diamidino-2-phenylindole is a DNA-specific stain that fluoresces blue or bluish-white when excited with light of 365 nm wavelength. With either dye, the cells are easily counted using an epifluorescence microscope. The DEFT count is rapid, taking 25 min to complete, is inexpensive, correlates well with the plate colony count ( $r > 0.9$ ), and is suitable for raw milks containing  $5 \times 10^3$  to  $10^8$  bacteria per milliliter.

One of the major disadvantages of the manual DEFT is operator fatigue associated with prolonged use of the microscope. This can be eliminated by the use of a semiautomated counting system based on an image analyzer. An instrument has been marketed (the Biocom Cobra) that automates fully the sample filtration, staining, rinsing, drying, and counting stages of DEFT. One operator can process  $\sim 96$  samples per hour and results are available in  $\sim 30$  min. Compared with the standard DEFT, miniaturization has led to substantial cost savings,  $\sim 80\%$ , on disposables.

One can determine viable cells by use of microcolony-DEFT. This is achieved by placing DEFT membranes (through which homogenates have been filtered) on the surface of an appropriate culture

medium and incubating for 3 h for Gram-negative bacteria and 6 h for Gram-positives to allow microcolonies to develop, which, after fluorescent staining, are then viewed with a microscope. As few as  $\sim 10$  cells can be detected within  $\sim 8$  h. Antibody (Ab)-DEFT for Gram-negative bacteria can be applied to specimens without enrichment and results can be obtained in  $< 1$  h with as few as  $16 \text{ CFU g}^{-1}$ . Ab-DEFT employs a fluorescein-labeled antibody to the organism of interest and results are determined by use of epifluorescence microscopy.

The Bactoscan instrument uses fluorescence microscopy for the automated counting of bacteria in milk. Milk somatic cells and casein micelles are dissolved chemically and then the bacteria are separated by continuous centrifugation in gradients formed with solutions of dextran and sucrose. The bacteria recovered from the gradient are incubated with protease, to remove residual protein, and then stained with Acridine Orange. For counting, the treated sample is applied as a thin film on the surface of a rotary disc and passed under a microscope objective. The fluorescent impulses in the microscope image are converted into electrical impulses and recorded. The instrument can analyze 80 samples per hour with an analysis time of 7 min.

**HGMF** The technique reduces the need for dilutions of the sample before enumeration, and hence reduces processing time, as well as giving better recovery than conventional filters, especially where high numbers of bacteria are involved. The HGMF has a square grid pattern printed in hydrophobic material, such as wax, on a conventional membrane filter. This divides the filter into a number of compartments, usually 2000–4000 depending on the size of the grids. The device functions as an MPN technique, eliminating size variations in colonies and preventing lateral spread. This greatly facilitates the automated counting of the colonies. Growth in a grid cell does not necessarily equal one colony in the plate colony count, since frequently a grid is inoculated with more than one organism. Coincidental inoculation is allowed for in calculating the count per milliliter or per gram of sample.

### Indirect Methods

There is a long list of methods that can be used to detect and/or enumerate microorganisms by measuring parts of cells or by substrate reactions, and the list continues to grow. Some of the most widely used are summarized in Table 2.

**Dye reductions** Dye reduction tests are based on the ability of bacterial enzymes, such as

**Table 2** Summary of some indirect methods for detecting or enumerating microorganisms or their products

Methods	Applications	Time for results	Sensitivity
ATP assay	Total viable cells	5 min	$\sim 10^3$ cells
DNA probes	Cells	1–2 days	1–5 cells
Colony hybridization	Viable bacterial cells	16–24 h	1–10 cells
Dot blot analysis	Viable bacterial cells	2–4 days	1 CFU g <sup>-1</sup>
ELISA	Enterotoxins, mycotoxins, Gram-negative bacteria	1 h for toxins; 20–36 h for cells	0.01–0.1 ng <i>Staph. enterotoxin</i> ; $\sim 5$ ppb for some mycotoxins
Impedance	Total viable bacteria	> 6 h	$10^5$ – $10^6$ cells
	Coliforms and yeasts	in 4 h	10
Immunodiffusion	Soluble toxins	3–5 days	0.01–0.1 $\mu$ g SEs
Limulus assay	Viable and nonviable Gram-negative bacteria	30–60 min	$\sim 300$ <i>E. coli</i> cells; 1–5 pg of <i>E. coli</i> LPS
Polymerase chain reaction	Microbial cells	3–5 h	1–100 cells
Reverse transcription PCR	Enteric viruses in water, food, stools	3–5 h	1–100 pfu
	Viable bacterial cells	3–5 h	1–100 cells
Multiplex PCR	Microbial cells (suitable primers)	3–5 h	1–100 cells
Dye (methylene blue)	Raw milk, meats	$\sim 8$	$10^5$ CFU ml <sup>-1</sup>
Reduction	Seafoods	8–12 h	$10^5$ – $10^4$ CFU
Radioimmunoassay	Botulinal, other toxins	3–4 h	0.1–1.0 ng SEs
Radiometry	Viable cells	2–24 h	1–10 coliforms in 6 h
Thermostable-nuclease	<i>Staph. aureus</i>	$\sim 3$ h	$10^6$ cells or 2–5 ng <i>Staph.</i> <i>enterotoxin</i>
Biosensors			
<i>lux</i> gene	Viable salmonellae	10–30 min	$\sim 10$ CFU
Ice nucleation	Viable salmonellae	24 h	$\sim 25$ cells

dehydrogenases, to transfer hydrogen from a substrate to a redox dye, which then undergoes a change in color. The rate of reduction depends on the enzyme activity, and this has been used as an index of the number of bacteria present in milk. In general, the reduction time is inversely related to the bacterial content of the sample when incubation with the dye commences.

Methylene blue is the most widely used dye and it undergoes a color change from blue to white when reduced. Resazurin, like methylene blue, is a redox indicator but unlike methylene blue, it undergoes a series of color changes from blue and mauve to pink as it is reduced. The use of these dyes is restricted to products that do not contain inherent reducing components.

**Impedance** The growth of microorganisms results in changes in the composition of the culture medium as nutrients are converted into metabolic end products. Complex uncharged molecules, such as carbohydrates or lipids, are catabolized to smaller charged molecules such as lactic and acetic acid. Charged molecules, such as proteins and polypeptides, are converted via amino acids into ammonia and carbon dioxide. As growth proceeds, these processes lead to a decrease in the overall impedance of the medium as conductance and capacitance increase. Electrical charges in microbial cultures provide a means of detecting microorganisms and their metabolic effects.

Electrical charges are a function of bacterial growth and replication.

The threshold for detection by instruments based on electrical methods such as Bactometer, Malthus, and RABIT depends upon the organisms and the media. In general, most media give detection thresholds at  $\sim 10^6$  bacteria or  $10^5$  yeasts ml<sup>-1</sup>. If the sample inoculated into the medium contains fewer organisms than this, there will be no detectable signal until the organisms have replicated sufficiently to reach the threshold number. The detection time depends upon the initial number of viable organisms present in the sample and their specific growth kinetics. Variations in growth rates of different organisms may give rise to errors in the estimation of bacterial numbers. Generally, there is an inverse relationship between log<sub>10</sub> plate colony count per milliliter or per gram and the electrical (impedance or conductance) detection time. The method has an inherent delay whilst growth occurs, generally taking > 6 h to detect  $10^5$  bacteria ml<sup>-1</sup>.

**Adenosine 5'-triphosphate (ATP) assay** The functional significance of ATP in the metabolism of living cells suggests that its measurement should be an excellent monitor of microbial activity in the sample. There are two factors that adversely affect the estimation of bacterial numbers by measurement of ATP: first, nonbacterial ATP and, second, quenching of the emitted light. Failure to destroy the



nonmicrobial ATP will give an elevated estimate of microbial ATP, and hence microbial numbers. The usefulness of ATP methods to enumerate microbial numbers is highly dependent upon the sample type. The BactoFoss (Foss Electric) is a fully automated instrument using ATP-bioluminescence for estimating the numbers of bacteria in raw milk. The instrument can analyze ~20 samples per hour and takes only 3 min to give a result. Methods based on ATP can be used to obtain real-time measurements of the hygienic status of processing equipment. Portable test kits and instruments are marketed by Biotrace, Lumac, and other companies.

**Enzyme-linked immunosorbent assay (ELISA)** ELISAs detect and amplify antigen–antibody reactions by using covalently bound enzyme–antibody molecules. The presence of the enzyme (indicating presence of the antigen) is detected by the addition of the appropriate substrate. Detection systems are usually designed to produce a color change that can be quantified by a microtiter plate reader.

ELISAs can be used in two modes, qualitatively to determine the presence or absence, or quantitatively to determine the amount of antigen present. ELISA kits often depend on the adsorption of either the antibody or antigen to a solid phase, e.g., wells of a microtiter plate, surface of plastic beads, or plastic stick. The choice of antibody (or antibodies) used determines the specificity of the ELISA assay, which can range from genus-specific to strain-specific. The principle on which ELISA methods are based usually prevents them from being used for the determination of total microbial counts. However, they can be used to detect pathogens such as *Salmonella* spp., *Listeria* spp.

**Limulus test** The *Limulus* test can be used to determine rapidly and specifically the cumulative content of Gram-negative bacteria in foods. Gram-negative bacteria produce a lipopolysaccharide (LPS) (endotoxin), which is a high-molecular-weight complex; it is not produced by Gram-positive bacteria. Present in the blue blood of the horseshoe crab, *Limulus polyphemus* is a nucleated cell called an amoebocyte, the cytoplasm of which is densely packed with granules. *Limulus* blood clots in the presence of bacterial LPS. All the necessary clotting factors are contained in the extract of the amoebocyte granules, called *Limulus* lysate. The *Limulus* test is specific for LPSs and is very sensitive. As little as  $10^{-12}$  g LPS per milliliter can be detected, occasionally even  $10^{-15}$  g ml $^{-1}$ . A single Gram-negative bacterium contains  $\sim 10^{-14}$  g of LPS.

For the *Limulus* test, a 10-fold dilution series of the sample is prepared and equal volumes of the *Limulus* lysate and diluted sample are mixed in a test tube. The test tube is then incubated before being inverted and read. If the mixture remains unchanged and runs out of the tube then that dilution of the sample does not contain LPSs. If a firm opaque gel is formed that sticks to the bottom of the tube then that dilution of the sample contains LPSs. Generally, a visual reading of the tenfold or twofold dilutions gives sufficient information about the level of LPSs present in the sample.

The limulus amoebocyte lysate (LAL) test with chromogenic substrate is faster than the gelation method, and it can be automated. The chromogenic substrate is attached to *p*-nitroaniline, that is released when reacted with the endotoxin-activated enzymes. The free *p*-aniline is read at 405 nm.

**Biosensors** The use of certain chemicals or one live organism to detect the presence or activity of another is the basis of what biosensors are. The incorporations of the *lux* and ice nucleation genes are examples of biosensors. A large number of biosensors have been constructed along the lines of the above, and several examples are noted below. *Saccharomyces cerevisiae* has been altered to express the firefly luciferase gene as a means of assessing the presence and toxicity of environmental contaminants in waters. Bioluminescence is decreased by toxins that inhibit this eucaryotic cell and the degree of toxicity is assessed by decreased bioluminescence. In a biosensor that detects the presence of Gram-negative bacteria, the green fluorescent protein (Gfp) was enhanced so that its fluorescence is quenched when it reacts with Gram-negative bacteria or LPS in water. Biosensors have been developed that can detect Gram-negative bacteria that produce acyl-homoserine lactone (AHL) molecules that are often involved in quorum sensing. AHL production is detected on agar plates after ~24 h incubation either by a color change or the production of bioluminescence around the biosensor culture, depending upon the biosensor strains used. A study of the metabolism of *Lactococcus lactis* in the gastrointestinal tract was achieved with an *L. lactis* strain that was infused with the *lux* gene. Some other uses of *lux* and ice nucleation are discussed below.

In some *Vibrio* species, bacterial luciferase catalyzes the flavin-mediated oxidation of tetradecanal with concomitant emission of blue-green light. The genes responsible for reactions of the fatty acid reductase complex (*lux* genes) have been cloned and their functions identified. One novel concept for the enumeration and identification of microorganisms requires the introduction of the *lux* genes into the

genome of a bacteriophage. The recombinant phages lack the intracellular biochemistry necessary for light production and are therefore dark. The result of phage infection is bioluminescent bacteria, which are then readily detectable.

Ice nucleating bacteria are found in a few genera. Certain species of *Pseudomonas*, *Erwinia*, and *Xanthomonas* possess ice nuclei that initiate the formation of ice in water at temperatures as high as  $-2^{\circ}\text{C}$ , compared with most inorganic ice nucleators that do not show activity until  $-8^{\circ}\text{C}$ . The phenotypic characteristic of ice nucleation is probably encoded in a single gene. Ice nucleation genes and proteins have been used as sensitive labels in ELISA immunoassays. The detection of freezing events is simplified by the use of dyes that fluoresce in super-cooled water but are quenched and change color when the water freezes.

Ice nucleation research has led to the development of the Bind assay for the rapid detection of *Salmonella* spp. The Bind assay uses attachment of a phage to its host and the expression of an ice nucleation gene after the injection of phage DNA to achieve highly specific and sensitive detection of the target organism. A range of *Salmonella* spp. has been detected at levels of less than  $10\text{ ml}^{-1}$ . The Bind assay principle could be applied to detecting a range of bacterial pathogens and typically gives results in less than 1 h but will probably require preincubation of the sample before testing. Similar advances are being made in *lux* gene technology.

**Radiometric methods** Most radiometric methods are based on the principle that microbial growth in media can be monitored by measuring  $^{14}\text{CO}_2$  released during the metabolism of radiolabeled nutrients. These methods take from 2 to 24 h to complete depending on the numbers of microorganisms present in the sample. As with impedance measurements, the detection time (the time taken for the measurement of a specific concentration of  $^{14}\text{CO}_2$ ) is generally inversely related to bacterial numbers. Radiometric methods are perhaps best suited to determining the sterility of products as they are particularly sensitive, given a sufficient period of incubation. The Bactec is a radiometric-based system.

**Nucleic-acid based methods** A number of detection and identification methods have been developed that are based on specific nucleotide sequences in RNA or DNA and several are noted in Table 2. Synopses of other related methods are presented in Table 3.

**DNA probes** Using labeled genome sequences, a probe can be made to detect for an organism in a

variety of specimens. Probes are now widely used and this wide usage results in part from the availability of a number of commercial test kits for a variety of microorganisms, mainly bacterial pathogens, and from their sensitivity and effectiveness in identifying species or strains of interest.

To apply a probe, unknown bacterial cells are treated to release single strand DNAs (ssDNA) that are affixed to a cellulose nitrate filter membrane. A laboratory probe (single strand) is made with bases complementary to those of the organism of interest on the membrane, and attached to the probe is a reporter molecule (dye, isotope, or enzyme). After the labeled probe has reacted with the ssDNA on the membrane, unreacted probe is rinsed away followed by detection of the reporter. Probes may bind to rRNA (rRNA:DNA) or to ssDNA (DNA:DNA).

A dipstick probe can be used to identify organisms in clinical or environmental specimens. It consists of adding a capture probe (containing a sequence of poly-dA or -dT bases that identifies it as a capture probe) coupled to dioxygenin as a reporter. Following destruction of unhybridized probe material, a dipstick that can hybridize with the capture probe base sequence is added. After all steps are completed, the presence of the capture probe is positive for the organism of interest. The theoretical sensitivity of a probe is  $<1\text{ }\mu\text{g}$  of nucleic acid or  $\sim 10^6$  cells. However, by use of polymerase chain reaction (PCR), the DNA of a single cell can be amplified several millionfold.

**Fluorescence in situ hybridization (FISH)** FISH permits the microscopic visualization of a nucleic acid target in a cytological context. It is primarily used to map genes to their respective chromosomal locations using labeled probes, and a similar approach might also be adapted to detect the presence of microorganisms. FISH has several advantages including the ability to visualize nucleic acids within a single cell and the sensitivity to detect one copy of the target without destroying the cell structure. The probe is usually a nucleic acid segment that is labeled with a fluorescent marker to allow detection (direct FISH) or labeled with a hapten, which require the development of a secondary reagent coupled to the fluorescent marker (indirect FISH). The target is normally a preparation of DNA or mRNA on a microscope slide. Target nucleic acids and probe are denatured and annealed, allowing the labeled probe to hybridize to the target sequences. Unbound probes are washed away and the specific hybridization signals are revealed.

**Polymerase chain reaction** By use of this method, large quantities of DNA can be made from an initial

**Table 3** Synopses of some of the newer biotyping, fingerprinting, and strain identification methods**Biotyping** (for species and strains)

**Bacteriophage typing:** The very high degree of specificity of a phage for its host cell is used to identify species and strains. Very valuable in epidemiological investigations. About 24 h required for results for most aerobic bacteria.

**Immunofluorescence:** The fluorescent antibody test is the most widely used. Antibody is labeled with a fluorescent dye and results are viewed under a microscope.

**Latex agglutination:** Antibodies are coupled to microspheres that agglutinate in the presence of homologous cells. At least  $10^7$  cells  $\text{ml}^{-1}$  are needed but results may be obtained in  $<1$  h.

**Cellular fatty acid analysis:** Total cellular or membrane-associated fatty acids are determined by GLC and both species and strains can be discerned by reference to a standard library. Several days are required for results.

**Fingerprinting** (strain differentiations)

**Ribotyping:** Using restriction fragment length polymorphism (RFLP) consisting of the restriction fragments around the rRNA genes from a particular genome to characterize bacterial strains. Total DNA is extracted and digested with selected restriction enzyme into multiple fragments ranging from 1 to 20 kb. The fragments are separated by agarose gel electrophoresis and subsequently hybridized with a labeled probe targeting 16S, 23S, or 5S rRNA genes. The length polymorphism of the hybridized fragments can be detected based on the labeling method (dye, radioisotope, or enzyme).

**MicroArray:** This is a hybridization-based technique but offers the advantage of being able to monitor the expression of thousands of genes on a single chip. One version of the MicroArray is to immobilize oligonucleotide or DNA probes to solid surface at defined positions. RNA extract or DNA from sample is eluted over the surface. Complementary binding between the probes and sample is detected by fluorescence following laser excitation. MicroArray is widely used to compare differences in gene expression. Various versions can be used to identify microorganisms as well.

**PCR:** Using a pair of primers targeting a particular gene from a specific organism and employing thermostable DNA polymerases and thermo-cycler to selectively amplify the flanking DNA fragment. The comparison is based on the size of the PCR fragment determined by agarose gel electrophoresis; therefore, a false positive result is possible. The method is very sensitive and rapid. Theoretically, the presence of even a single cell can be detected within a few hours.

**AFPL:** This patented (Keygene NV) amplified fragment length polymorphism assay is a PCR-based method that is very highly discriminating. It employs the ligation of adapters to restriction fragments followed by PCR-amplification with adapter-specific primers. Commercial kits are available, and results can be obtained within 24 h.

**RAPD:** Using a short oligonucleotide fragment the sequence of which is randomly scattered within the genome as PCR primer for random amplification of polymorphic DNA fragments from the genome. The PCR fragments can be separated by agarose gel electrophoresis and the characteristic patterns can be used for strain identification. The method is rather rapid and the procedures can be completed within several hours. However, the repeatability of RAPD sometimes is questionable.

**PFGE:** Whole cells are embedded in agarose gel and genomic DNAs are enzymatically purified and restrict digested by rare cutting enzymes *in situ*. Large restriction fragments (50–1000 kb) are separated by pulse field gel electrophoresis into a characteristic RFLP. The total procedures take several days.

**MEE:** Multilocus enzyme electrophoresis (MEE) can be used to estimate genetic relatedness among strains or species by measuring the relative mobility of cellular enzymes under nondenaturing conditions. The allelic profiles or MEE types may be used to compare isolates of the same species from different sources.

**Identification** (of species and strains)

16S rRNA sequence comparisons are in wide use to identify bacteria to the generic, species, and subspecies levels. The class *Proteobacteria* is based on 16S rRNA data and many of the new bacterial taxa are defined along these genotypic lines in contrast to the phenotypes that results from the use of general biochemical methods noted below.

Substrate test kits include those of the Enterotube, API, BioLog, and other related systems. They can be used to determine biochemical reactions of bacteria, yeasts, and molds within 1–3 days and thus aid in their identification. Some API system kits can be used to characterize organisms at the subspecies level by their capacity to utilize up to 100 substrates. The BioLog systems use carbon source metabolic 'fingerprints' to identify microorganisms.

Chromogenic and fluorogenic substrates are employed in a number of culture media to facilitate strain or species identification. BCIG or X-Gluc (5-bromo-4-chloro-3-indolyl- $\beta$ -D-glucuronic acid) is an example of a chromogenic substrate that turns blue within 24 h with *E. coli* growth while MUG (4-methylumbelliferyl- $\beta$ -D-glucuronide) is a substrate that fluoresces with *E. coli* growth.

small quantity. With the nucleotide sequence of a portion of the desired gene and laboratory-made short oligonucleotide primers that are complementary to the above gene sequence, the various steps in the PCR including 30–50 cycles will lead to a several millionfold increase in the small amount of starting DNA.

Two or more regions of a genome can be amplified during a single PCR procedure and this is referred to as multiplex PCR. If the amplified PCR products have different length, they can be resolved by use of

agarose (0.7–2%, depending upon the length of the fragments) gel electrophoresis in  $\sim 2$ –4 h. This method can be used to distinguish between botulinum toxins A, B, and E in food specimens when the specific genome sequences are employed, and as few as 100 botulinum cells per gram can be detected. If primers were designed to target genomes from different microorganisms, multiplex PCR can also be used to detect the presence of multiple microorganisms during a single PCR procedure.

Reverse transcriptase (RT) PCR (RT-PCR) can be used to detect RNA in specimens, especially ssRNA viruses. Using the ssRNA as a template, cDNA (complementary DNA) can be synthesized with the enzyme reverse transcriptase, which can further be used as a template in PCR amplification. In two-step RT-PCR, extracted RNAs are first mixed with RT and suitable primer to synthesize cDNA and then followed by adding *Taq* polymerase and PCR primer pairs for PCR amplification. In one-step RT-PCR, all the ingredients are mixed together and allowed for one cycle of reverse transcription to synthesize cDNA followed by 30+ cycles of PCR amplification. The product can be analyzed by agarose gel electrophoresis. Commercial RT-PCR test kits are available.

### Other Methods

Microbiological methods are useful for testing naturally occurring water-soluble vitamins such as the vitamin B complex. The growth response of vitamin-dependant strains of microorganisms in vitamin-free media is proportional to the concentration of the vitamin in the added test sample, or standard. Standard calibration curves of turbidity (microbial growth) versus concentration of vitamin are prepared and the concentration in the test sample estimated from its turbidity measurement. The organisms used are often specific strains of *Lactobacillus*, *Streptococcus*, and *Saccharomyces* species. Despite the fact that the microbiological technique is more cumbersome than chromatographic techniques, it is still used for those vitamins that are present in formulations at minute concentrations, as it is more sensitive

(riboflavin can typically be measured at a concentration of less than 1 µg per g) and they are preferred over colorimetric methods for foods containing high levels of Maillard browning products.

A number of microbiological methods are used in the pharmaceutical and healthcare industries. The antimicrobial preservatives – effectiveness assay is used to assess the effectiveness of a given agent within a product formulation to act as a preservative. The test product, complete with the preservative, is assessed in terms of its ability to suppress the growth of a series of test organisms under a defined set of conditions. The microbial limits test is used to estimate the total number of viable organisms within a given finished product, or raw material, and to determine the presence or absence of a specific set of pathogens. Sterility tests are performed in order to determine whether a given Pharmacopeial article is actually sterile. The procedure (direct transfer to test media or membrane filtration) by which the product is tested depends upon the physical nature of the product and the presence of any bacteriostatic or fungistatic agents within the formulation. Microbiological assays are typically used to determine the activity or potency of antimicrobial products such as antibiotics. This activity is demonstrated by a pronounced reduction in growth of the test organism. The microbiological assay remains the definitive technique because it can reveal the presence of subtle changes within the product that may go undetected if assayed by chemical means. There are two standard forms of the microbiological antimicrobial assay: the cylinder-plate assay and turbidometric assay. Both assays rely upon quantifying the degree of growth

**Table 4** Synopses of some of the bioassay methods employed to determine the effect of microorganisms on human and animal tissues

<i>Anton test</i> : A test of virulence, especially for <i>Listeria</i> spp., ~ 10 <sup>6</sup> cells are placed in the eye of a guinea pig or rabbit and conjunctivitis is observed
<i>Mouse lethality</i> : To determine lethality of microbial toxins (e.g., botulinal). Injected mice are observed for up to 72 h for lethality. Toxin serotypes can be determined by use of specific antiserum. The LD <sub>50</sub> of <i>Listeria monocytogenes</i> for normal adult mice is 10 <sup>5</sup> –10 <sup>6</sup> cells, and as few as 50 cells for infant mice
<i>CHO cells</i> . Chinese hamster ovary cell monolayers are used to assess the activity of bacterial enteropathogens after 24–30 h incubation
<i>Ileal loop assays</i> . Rabbits are most commonly used to determine enteropathogenicity of cells or toxins. Injected loop sections are examined for fluid accumulation after 18–24 h
<i>Hep-2 cells</i> . May be used to determine adherence and/or invasive potential of culture filtrates. Results for some bacterial pathogens may be obtained in <2 h while some require 24 h
<i>Vero cells</i> . Used to determine enteropathogenicity of bacteria, especially invasive potential
<i>Sereny test</i> . Used to determine virulence/invasiveness of certain bacteria by administering ~ 10 cells in the eye of a guinea pig. Eyes are examined for up to 5 days for lesions
<i>Skin tests (rabbit, guinea pig)</i> . Live animals are injected i.d. to determine vascular permeability and/or erythelial activity of microbial toxins. Results may be obtained in 18–24 h
<i>Suncus murinus</i> . This small animal (Japanese house shrew) appears to represent a good bioassay of cerulide produced by <i>Clostridium perfringens</i> . The latter has an ed <sub>50</sub> of 10–13 µg per kg
<i>Rhesus monkey</i> . The animal of choice for staphylococcal enterotoxins, which induce emesis with ~ 5 µg per 2–3 kg body weight within 5 h when administered by stomach tube

reduction of the test organism upon exposure to test product; the former detects activity through zones of inhibition on solid agar and the latter through reduced turbidity.

In addition to the direct and indirect methods for enumerating microorganisms, a large number of methods exist that are employed to identify, characterize, or profile individual cultures, and some of these are listed and briefly characterized in **Table 3**.

Although genotypic and phenotypic methods (alone or in combination) can be used to identify species and strains of microorganisms, they do not provide direct evidence for the actual role of an organism in a human or animal infectious process. A number of bioassay methods are employed to this end and they consist of the use of whole animals, animal organs or tissues, or human or animal cells (as monolayers). The more commonly used of these methods are summarized in **Table 4**.

See also: **Water Analysis: Microbiological.**

## Further Reading

- Atlas RM (1996) *Handbook of Microbiological Media*, 2nd edn. Boca Raton, FL: CRC Press.
- Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, and Struhl K (eds.) (1987) *Current Protocols in Molecular Biology*. Indianapolis: Wiley.
- Bacteriological Analytical Manual Online (2001) U.S. Department of Health and Human Services, U.S. Food & Drug Administration, Center for Food Safety & applied Nutrition (<http://www.cfsan.fda.gov/~ebam/bam-toc.html>).
- Balows A, Truper HG, Dworkin D, Harder W, and Schleifer K-H (1992) *The Prokaryotes – A Handbook on the Biology of Bacteria*, vols. 1–4. Berlin: Springer.
- Barnett JA, Barnett LK, Payne RW, and Yarrow D (2000) *Yeasts: Characteristics and Identification*. Cambridge: Cambridge University Press.
- British Standard BS 5763. *Microbiological Examination of Food and Animal Feeding Stuffs*. London: British Standards Institution.
- Clark S, Thompson KC, Keevil CW, and Smith M (2001) *Rapid Detection Assays for Food and Water*. New York: Springer.
- Downes FD and Ito K (eds.) (2001) *Compendium of Methods for the Microbiological Examination of Foods*, 4th edn. Washington, DC: American Public Health Association.
- European Pharmacopoeia Plus Supplements, 4th edn. (2002). France: European Directorate for the Quality of Medicines.
- Jay JM (2000) *Modern Food Microbiology*, 6th edn. New York: Kluwer/Plenum Academic Publishers.
- Kepner RL Jr. and Pratt JR (1994) Use of fluorochromes for direct enumeration of total bacteria in environmental samples: past and present. *Microbiology Reviews* 58: 603–615.
- Murray PR, Barron EJ, Jorgensen JH, Pfaller MA, and Tenover FC (eds.) (2003) *Manual of Clinical Microbiology*, 8th edn. Washington, DC: ASM Press.
- Rose RN, Hamilton RG, and Detrick B (eds.) (2002) *Manual of Clinical Laboratory Immunology*, 6th edn. Washington, DC: ASM Press.
- Samson R, Hoekstra ES, and Van Oorschot CAN (1981) *Introduction to Food-borne Fungi*. Baarn: Centraalbureau voor Schimmelcultures.
- Trask BJ (1991) Fluorescence *in situ* hybridization. *Trends in Genetics* 7: 149–154.

# MICROELECTRODES

**T Zeuthen**, University of Copenhagen, Copenhagen, Denmark

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Microelectrodes are electrodes with tip areas of the order of micrometers, i.e., so small that they do not induce significant damage in the sample to be tested. Such requirements are especially important in living tissue, where microelectrodes have found their widest application; for intracellular recordings tip diameters less than 1  $\mu\text{m}$  are required. Microelectrodes can also be used in combination with recent techniques of molecular biology.

Microelectrodes are basically of four types: glass micropipettes, ion-selective microelectrodes, solid-state microelectrodes, and enzyme microelectrodes (**Figure 1**). Glass micropipettes are used to record steady-state (DC) and alternating (AC) electrical potentials. If the glass micropipette is plugged with an ion-selective membrane, the microelectrode will record a potential that is proportional to the external activity of the ion to which the membrane is sensitive. Metal microelectrodes are well suited to recording AC potentials in situations where mechanical strength is required. Metal microelectrodes can also be used for cyclic voltammetry or amperometry purposes. If a metal microelectrode made from an inert metal or carbon is maintained at a given



reduction of the test organism upon exposure to test product; the former detects activity through zones of inhibition on solid agar and the latter through reduced turbidity.

In addition to the direct and indirect methods for enumerating microorganisms, a large number of methods exist that are employed to identify, characterize, or profile individual cultures, and some of these are listed and briefly characterized in **Table 3**.

Although genotypic and phenotypic methods (alone or in combination) can be used to identify species and strains of microorganisms, they do not provide direct evidence for the actual role of an organism in a human or animal infectious process. A number of bioassay methods are employed to this end and they consist of the use of whole animals, animal organs or tissues, or human or animal cells (as monolayers). The more commonly used of these methods are summarized in **Table 4**.

See also: **Water Analysis: Microbiological.**

## Further Reading

- Atlas RM (1996) *Handbook of Microbiological Media*, 2nd edn. Boca Raton, FL: CRC Press.
- Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, and Struhl K (eds.) (1987) *Current Protocols in Molecular Biology*. Indianapolis: Wiley.
- Bacteriological Analytical Manual Online (2001) U.S. Department of Health and Human Services, U.S. Food & Drug Administration, Center for Food Safety & applied Nutrition (<http://www.cfsan.fda.gov/~ebam/bam-toc.html>).
- Balows A, Truper HG, Dworkin D, Harder W, and Schleifer K-H (1992) *The Prokaryotes – A Handbook on the Biology of Bacteria*, vols. 1–4. Berlin: Springer.
- Barnett JA, Barnett LK, Payne RW, and Yarrow D (2000) *Yeasts: Characteristics and Identification*. Cambridge: Cambridge University Press.
- British Standard BS 5763. *Microbiological Examination of Food and Animal Feeding Stuffs*. London: British Standards Institution.
- Clark S, Thompson KC, Keevil CW, and Smith M (2001) *Rapid Detection Assays for Food and Water*. New York: Springer.
- Downes FD and Ito K (eds.) (2001) *Compendium of Methods for the Microbiological Examination of Foods*, 4th edn. Washington, DC: American Public Health Association.
- European Pharmacopoeia Plus Supplements, 4th edn. (2002). France: European Directorate for the Quality of Medicines.
- Jay JM (2000) *Modern Food Microbiology*, 6th edn. New York: Kluwer/Plenum Academic Publishers.
- Kepner RL Jr. and Pratt JR (1994) Use of fluorochromes for direct enumeration of total bacteria in environmental samples: past and present. *Microbiology Reviews* 58: 603–615.
- Murray PR, Barron EJ, Jorgensen JH, Pfaller MA, and Tenover FC (2003) *Manual of Clinical Microbiology*, 8th edn. Washington, DC: ASM Press.
- Rose RN, Hamilton RG, and Detrick B (eds.) (2002) *Manual of Clinical Laboratory Immunology*, 6th edn. Washington, DC: ASM Press.
- Samson R, Hoekstra ES, and Van Oorschot CAN (1981) *Introduction to Food-borne Fungi*. Baarn: Centraalbureau voor Schimmelcultures.
- Trask BJ (1991) Fluorescence *in situ* hybridization. *Trends in Genetics* 7: 149–154.

# MICROELECTRODES

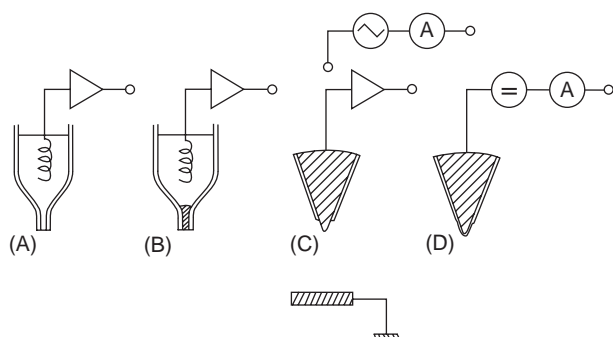
**T Zeuthen**, University of Copenhagen, Copenhagen, Denmark

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

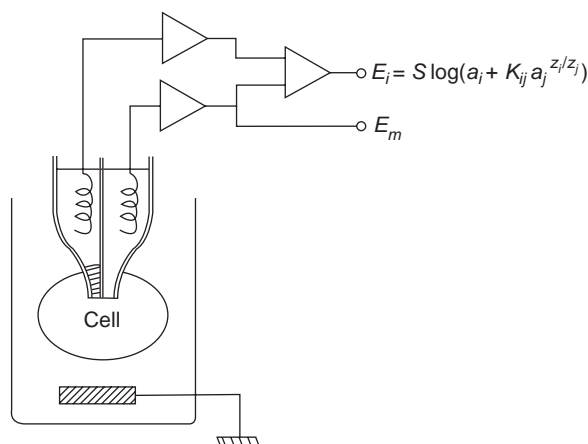
Microelectrodes are electrodes with tip areas of the order of micrometers, i.e., so small that they do not induce significant damage in the sample to be tested. Such requirements are especially important in living tissue, where microelectrodes have found their widest application; for intracellular recordings tip diameters less than 1  $\mu\text{m}$  are required. Microelectrodes can also be used in combination with recent techniques of molecular biology.

Microelectrodes are basically of four types: glass micropipettes, ion-selective microelectrodes, solid-state microelectrodes, and enzyme microelectrodes (**Figure 1**). Glass micropipettes are used to record steady-state (DC) and alternating (AC) electrical potentials. If the glass micropipette is plugged with an ion-selective membrane, the microelectrode will record a potential that is proportional to the external activity of the ion to which the membrane is sensitive. Metal microelectrodes are well suited to recording AC potentials in situations where mechanical strength is required. Metal microelectrodes can also be used for cyclic voltammetry or amperometry purposes. If a metal microelectrode made from an inert metal or carbon is maintained at a given



**Figure 1** Four types of microelectrodes. (A) Glass micropipettes; tip diameters range from 0.01  $\mu\text{m}$  to a few micrometers. The pipette is filled with a conducting salt solution and connected, for example, via a chlorided silver wire (coil) to an amplifier (triangle). The electrical circuit is completed via the reference electrode shown at the bottom of the figure. (B) An ion-selective microelectrode. It consists basically of the glass micropipette shown on the left, but the tip is closed by a liquid ion exchanger (shown hatched) or by an ion-selective glass. Some possible ion-selective microelectrodes are listed in **Table 1**. (C) A metal microelectrode or a carbon fiber electrode. It consists of a sharpened wire (shown hatched) insulated by a layer of, for example, glass or lacquer. The electrode can be connected to an amplifier (triangle) when electrical activity, e.g., neural activity, is recorded. The electrode can also be used for analytical purposes. In this case the electrode is not connected to the amplifier but to a voltage generator and an ammeter (A), shown at the top. If the voltage is varied (symbolized by the saw tooth) relative to the reference electrode, current can flow at certain potentials, e.g., when oxygen can receive an electron from or hydrogen peroxide can deliver an electron to the electrode. The electrode can record oxygen, hydrogen peroxide, or processes that can be oxidized or reduced at the metal or carbon surface. (D) An enzyme-based microelectrode. It is constructed as the previous electrode but its tip area is coated with an enzyme. The enzyme is chosen so as to catalyze a given process the substrate of which we want to quantify in the external medium. The catalysis process should result in a breakdown product that can be measured amperometrically, e.g., hydrogen peroxide as described above. Thus, the electrode is maintained at a given potential symbolized by the DC potential generator and the current is measured by the ammeter (A). In this way a glucose microelectrode can be constructed.

potential, the current can only pass the solution-electrode interphase if electrons can be provided or taken up by the chemical processes at the electrode surface. In cyclic voltammetry the voltage between the metal microelectrode and the reference electrode is varied in a cyclical manner. The voltage at which current can flow and the magnitude of the currents will characterize the chemical composition of the external media. Ambient oxygen concentrations can be measured in this way as the current flow obtained at a voltage of 600 mV relative to an Ag/AgCl reference electrode. If the electrode is prepared with specific enzymes attached to the tip area, it can be used to record substrates that in themselves do not participate in redox processes, e.g., glucose.



**Figure 2** The principle of measurement by means of an ion-selective double-barreled microelectrode inside a cell. The cell is in a bath the solution of which is grounded via a reference electrode. Each barrel is connected via a chlorided silver wire (shown coiled) to amplifiers (triangles). The reference barrel of the double-barreled electrode directly records the intracellular electrical potential, the membrane potential ( $E_m$ ). The ion-selective barrel, indicated by the plug of ion exchanger in the tip, records the sum of the membrane potential and a potential  $E_i$  related to the chemical potential of the ion in question (of activity  $a_i$ ) (see eqn [1]).  $E_i$  is obtained by electronic subtraction. The influence on  $E_i$  from other ions (indicated by index  $j$  and the valencies  $z_j$ ) can be obtained from calibration.

For potentiometric measurements (zero current), the complete measuring circuit consists of the microelectrode, the input amplifier, and a reference electrode. Sometimes a local reference electrode is needed if the electrical potential varies over short distances in the sample, for example, across a cell membrane (Figure 2).

## The Glass Pipette Microelectrode

Glass micropipettes are produced from tubes of aluminum silicate or borosilicate glass, typically with an outer diameter of 2 mm and an inner diameter of 1 mm. If the two ends are pulled away from each other while the center of the tube is heated, the tube will narrow and separate at the middle and two electrodes will form. With proper choice of heating and strength of pull, the narrowed ends of the separated glass tube – the tips – will retain an open lumen and will be continuous with the lumen at the larger end (the shaft) of the glass tube. Tip diameters as small as 0.045  $\mu\text{m}$  have been reported. The thickness of the wall at the tip is typically one-fifth to one-tenth of the tip diameter.

The electrode is filled with a salt solution and connected via a reference electrode (Ag/AgCl or calomel) inserted into the shaft to the input of an amplifier (Figure 1A). The electrode measures the

liquid junction potential between the filling solution and the external solution. In the simple case where the electrode contains the same univalent salt (of concentration  $C_2$ ), as the external solution (of concentration  $C_1$ ), this potential equals approximately  $56 \text{ mV} \times [(U - V)/(U + V)] \log(C_1/C_2)$  at  $20^\circ\text{C}$ , where  $V$  is the mobility of the cation and  $U$  that of the anion.

In the general case where the external solution is unknown, the equation for liquid junction potential takes a more complicated form. Yet, any variation in the potential between various sites of measurements can be suppressed if a filling solution of high concentration and with equal mobilities of the anion and cation (e.g.,  $1\text{--}2 \text{ mol l}^{-1}$  KCl) is employed. Thus, the working principle of the electrode depends on free diffusion of the filling solution from the tip into the sample. This has important implications. KCl will be continuously lost into the sample; if this leakage is prevented, then the electrode does not measure the correct potential. With tip diameters of  $0.1 \mu\text{m}$  (impedance of  $\sim 50 \text{ M}\Omega$ ), this leakage rarely constitutes a problem, even for intracellular recordings.

In addition to the liquid junction potential, which can be suppressed, some electrodes exhibit a so-called tip potential, which is  $\sim 5\text{--}10 \text{ mV}$ . The origin of this potential is unclear. It is associated with narrow tips and is less frequently encountered if the filling solutions are ultrafiltered before use, with a filter limit of  $0.2 \mu\text{m}$ . The tip potential can be defined as the difference in the electrode potential with the tip intact and the tip broken off. The tip potential varies in an unpredictable manner when the tip is placed in different solutions. Such electrodes should be discharged before use.

Glass exposed to water will hydrate and its electrical resistance will decrease. After 2 h of exposure to aqueous media, the resistance of the glass wall may have decreased from say 1000 to  $50 \text{ M}\Omega$ . If the impedance of the electrode is  $10 \text{ M}\Omega$ , this means that a significant shunting of the measured potential may take place. In practice, this means that electrodes ought to be made on the day of the experiment. It is also a common observation that cell debris and other biological material adheres more stubbornly to the sides of micropipettes that have been hydrated. It is often useful to change the electrode after 3–5 h of application in biological fluids.

The input amplifier for the micropipette should have an input resistance of  $10^8\text{--}10^{10} \Omega$ , easily obtained with operational amplifiers with field-effect transistor inputs. If the input is provided with negative capacitance compensation, both the capacitance of the cable that connects the electrode to the amplifier and the capacitance between the inner solution of the electrode and the external bath can be compensated

for. With proper compensation the electrode can follow fast changes in, for example, action potentials from nerves or nerve cells.

## **Ion-Selective Microelectrodes**

The micropipettes described above can be made into ion-selective microelectrodes if the tip is closed by means of an ion-selective membrane (Figure 1B). The membrane may be in the form of an ion-sensitive glass, for example, for the measurement of  $\text{H}^+$  and  $\text{Na}^+$ , or more conveniently a liquid ion exchanger (LIX), which is available for the measurement of  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{H}^+$ ,  $\text{Li}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ , and  $\text{Cl}^-$  (see Table 1). The input amplifier for the ion-selective micropipette should have an input resistance two orders of magnitude larger than the impedance of the electrode itself, i.e.,  $10^{12}\text{--}10^{14} \Omega$ . This can be obtained with operational amplifiers with field-effect transistor or electrometer inputs. The ion-selective membrane may also be placed directly on to the field-effect transistor input of an operational amplifier, a so-called ion-selective field-effect transistor (ISFET). The linear dimensions of the recording area of ISFETs can be as small as  $10 \mu\text{m}$ . To complete the electrical circuit, the amplifier of the ion-selective microelectrode must be connected to a reference electrode in the form of either a locally positioned micropipette or a large external electrode (Figure 2).

## **Construction**

The construction of an ion-selective microelectrode is most conveniently based on a micropipette made from inert glass as described above. This forms the outer housing of the electrode. If the ion-sensitive membrane is to be made from glass, tubes of ion-sensitive glass are pulled into the shape of a pipette with dimensions that allow this pipette to be fitted inside the inert micropipette. The tip of the ion-selective glass micropipette is sealed at the very tip and positioned inside the inert pipette with the tip of the inner pipette as close to the tip of the outer pipette as possible, preferably within  $10\text{--}20 \mu\text{m}$ . The two glasses are sealed together from  $\sim 100 \mu\text{m}$  from the tip for a length of  $\sim 100 \mu\text{m}$ . The electrodes last long but are relatively difficult to make. Unless special problems, such as biocompatibility between LIX and the sample are relevant, the LIX electrodes described below are preferred.

In the LIX microelectrode the ion exchanger is situated at the tip of the inert micropipette by rendering the inside of the micropipette hydrophobic and organophilic. Thus, the ion-sensitive part of the LIX microelectrode is the interphase between the ion

**Table 1** Properties of typical ion-selective microelectrodes

	<i>Principal constituent</i>	<i>Colloquial name</i>	<i>Selectivity log <math>K_{ij}^a</math></i>	<i>Slope (mV)</i>	<i>Range<sup>b</sup></i>	<i>Tip diameter (<math>\mu\text{m}</math>)<sup>c</sup></i>	<i>Impedance (<math>\Omega</math>)<sup>c</sup></i>	<i>Response time<sup>d</sup></i>
H <sup>+</sup>	4-Nonadecylpyrine	Fluka 95279 ETH 1907	K <sup>+</sup> : − 8.8	58	pH 2–9	0.3–1	10 <sup>11</sup>	1 s (1 $\mu\text{m}$ )
Li <sup>+</sup>	<i>N,N'</i> -Diheptyl- <i>N,N'</i> ,5,5-tetramethyl-3,7-dioxanonane diamide	ETH 149 Fluka 62380	Na <sup>+</sup> : − 1.3 Ca <sup>2+</sup> : − 0.6 K <sup>+</sup> : − 2.3 Mg <sup>2+</sup> : − 2.7	59	1–3 mmol	1.5	4 × 10 <sup>10</sup>	
Na <sup>+</sup>	1,1,1-Tris-[1'-(2'-oxo-4'-azo-5'-aza-5'-methyl) dodecanyl] propane <i>N,N'</i> -Dibenzyl- <i>N,N'</i> -1,2-phenylenedioxyacetamide	ETH 227 <sup>f</sup> Fluka 71732 <sup>f</sup> ETH 157 <sup>f</sup> Fluka 71733	Ca <sup>2+</sup> : − 0.2, K <sup>+</sup> : − 2.3 Ca <sup>2+</sup> : − 1.3, K <sup>+</sup> : − 0.5	53	1 mmol upwards	0.5–1	5 × 10 <sup>10</sup>	< 3 s
K <sup>+</sup>	Potassium tetrakis ( <i>p</i> -chlorophenyl) borate	Corning 477317 or Fluka 60398	Na <sup>+</sup> : − 2.0 to − 4.0	58	0.1 mmol upwards	0.2–1	10 <sup>10</sup>	0.2 s (1 $\mu\text{m}$ )
Mg <sup>2+</sup>	<i>N,N'</i> -Dimeptyl- <i>N,N'</i> -dimethylsuccinic-acid diamide	ETH 1117 Fluka 63082	Na <sup>+</sup> : − 1.1, K <sup>+</sup> : − 1.4 Ca <sup>2+</sup> : 1.1	28	0.4 mmol upwards	1	10 <sup>10</sup>	< 5 s (1 $\mu\text{m}$ )
Ca <sup>2+</sup>	(−)-( <i>R,R</i> )- <i>N,N'</i> -bis-[11-(ethoxy carbonyl)-undecyl]- <i>N,N'</i> -4,5-tetramethyl dioxaoctane diamide	ETH 1001 Fluka 21048	Na <sup>+</sup> : − 5.5, K <sup>+</sup> : − 5.4 Mg <sup>2+</sup> : − 5.0 or less	28	10 <sup>−7</sup> mmol upwards	1–2	2 × 10 <sup>10</sup>	< 5 s <sup>g</sup>
Cl <sup>−</sup>	Quaternary ammonium salt	Corning 477315 Corning 477913	HCO <sub>3</sub> <sup>−</sup> : − 0.8 <sup>h</sup>	50	5 mmol upwards	0.5	2 × 10 <sup>10</sup>	< 1 s

<sup>a</sup>These are the major interferences in a physiological experiment, the numbers are log  $K$  (eqn [1]).

<sup>b</sup>These values are relevant for experiments in normal biological fluids.

<sup>c</sup>These are values for typical electrodes, the range of impedances matches the range of tip diameters. Electrodes with smaller tip diameters can be made; this gives higher resistances and slower response time.

<sup>d</sup>These are typical response times determined by the resistance and capacitance of the electrode. There are large variations in reported values. The response time can be reduced by coating the electrode by a thin conducting surface and electronically driving this shield together with the capacitance of the connecting cable.

<sup>e</sup>Interference from carbon dioxide can be avoided if the exchanger is saturated with carbon dioxide prior to use.

<sup>f</sup>BTH 227 is suited for intracellular use; ETH 157 is suited for extracellular use.

<sup>g</sup>The response time depends on stirring around the tip.

<sup>h</sup>There are intracellular substances of unknown composition that equal ~7 mmol of Cl<sup>−</sup>.

exchanger and the external solution (Figure 1B). The siliconization is best achieved by exposing the inside of the pipette to a vapor that incorporates a layer of silicone onto the glass surface. Chloro- or amino-silanes can be used; both utilize the presence of surface-bound  $\text{OH}^-$  groups that bind to Si groups in the glass. By the treatment the  $\text{OH}^-$  is replaced by  $\text{SiC}_3\text{H}_9\text{O}$ , which renders the surface hydrophobic. The LIX is situated by a thin capillary via the shaft or by dipping the tip of the electrode into LIX. The remainder of the electrode is filled with an appropriate filling solution, preferably one that contains the ion in question at a concentration similar to that expected in the sample. This will minimize offset potentials.

### Sensitivity and Selectivity

The electromotive force of the electrode for an ion  $i$  (e.m.f. <sub>$i$</sub> ) can be approximated by the Nernst–Nikolski equation:

$$\text{e.m.f.}_i = E_r + S \log(a_i + K_{ij}a_j^{z_i/z_j}) \quad [1]$$

$E_r$  is the reference potential measured by the local reference electrode. The theoretically possible sensitivity  $S$  equals  $RT/z_iF$  where  $R$  is the gas constant,  $T$  is the absolute temperature,  $F$  is Faraday's constant, and  $z_i$  is the valence of the ion.  $S$  is  $\sim 58$  mV at  $25^\circ\text{C}$ .  $a_i$  is the activity of the ion.  $K_{ij}$  is the selectivity of the electrode toward an interfering ion of activity  $a_j$  and valence  $z_j$ . In practice, the sensitivity (or slope) will be less than the theoretical value. Typical sensitivities are given in Table 1 for electrodes with tip diameters of  $0.5\text{--}1\text{ }\mu\text{m}$ . Smaller the sensitivity the finer the tip, because of the larger relative significance of unselective shunts.

The selectivities for biologically relevant ions are also given in Table 1. For the  $\text{H}^+$  microelectrode it is seen that there is no significant influence from other ions, but the exchanger shown requires saturation with carbon dioxide. There is another  $\text{H}^+$  exchanger that does not require saturation with carbon dioxide (Fluka 95297). Therapeutic doses of  $\text{Li}^+$  are  $\sim 1\text{--}3$  mmol, and it is seen that the influence from  $\text{Na}^+$  and  $\text{K}^+$  at physiological concentrations limits the use of the  $\text{Li}^+$  electrode. With proper calibration, however, data in the physiological range can be obtained.  $\text{Na}^+$  electrodes are of two kinds: the exchanger ETH 227 is well suited for intracellular use since the relatively poor selectivity toward  $\text{Ca}^{2+}$  is irrelevant because of the low intracellular values of  $\text{Ca}^{3+}$ . It will be necessary, though, to calibrate in solutions with and without  $\text{Ca}^{3+}$  in order to obtain the true baseline when the  $\text{Na}^+$  concentration is low outside the cells. Improved selectivity for  $\text{Na}^+$  over

$\text{K}^+$  can be obtained by using an ionophore incorporating a tetramethoxyethyl ester derivate of  $p$ - $t$ -butyl calix[4]arene. The exchanger ETH 157 can be used outside cells where  $\text{K}^+$  is low and the poor selectivity of the exchanger toward  $\text{K}^+$  is of minor importance. Measurements with the  $\text{K}^+$  electrode are influenced by  $\text{Na}^+$  ions only to a minor degree and no special precautions are needed when intra- and extracellular measurements are compared. The useful range of the electrode is limited at low  $\text{K}^+$  activities, where  $0.1\text{ mmol l}^{-1}$  of  $\text{K}^+$  cannot be recognized in a solution that contains  $\text{Na}^+$  at activities of  $100\text{ mmol l}^{-1}$ . Physiological  $\text{Mg}^{2+}$  activities are in the millimolar per liter range and will be affected by  $\text{Na}^+$  and  $\text{K}^+$  concentrations in the physiological range. By careful calibration,  $\text{Mg}^{2+}$  activities as low as  $0.4\text{ mmol l}^{-1}$  can be detected.  $\text{Ca}^{2+}$  is present intracellularly in very low activities ( $10^{-8}\text{--}10^{-7}$  mol), and the electrode requires high selectivities toward  $\text{Na}^+$  and  $\text{K}^+$ . In practice, the electrode tends to be relatively slow, with time constants of several seconds. Methods with fluorescent dyes are more popular at present, yet such methods do not provide the opportunity for simultaneous recording of intracellular electrical potentials.

### Calibration

Calibration should be done in solutions whose compositions lie close to that of the sample to be tested. In this way unnecessary extrapolations are avoided and the greatest accuracy is obtained. Since the electrode measures activity, one compares the activity in the sample with the activity in the calibration solution. Sometimes the measured value is expressed as apparent concentration: it is assumed that there are equal activity coefficients in the sample and in the calibration solutions.

### Tip Diameters, Impedances, and Response Times

Microelectrodes for  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{H}^+$ ,  $\text{Li}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ , and  $\text{Cl}^-$  are usually made with tip diameters of  $\sim 1\text{ }\mu\text{m}$  (Table 1). For significantly smaller tip diameters the sensitivity will be smaller than the optimal value given in Table 1. Smaller tips may be obtained while keeping a high sensitivity by changes in the siliconizing procedure. With large resistances of the electrode, the response time will also be larger. This may be alleviated by using beveled tips, which increase the leading-off area while retaining the sharpness of the tip. Shielding and capacitance compensation built in to the amplifier may improve the response time. In summary, the size of the sample (i.e., the cell) determines the size of the tip that can be used. The physiological phenomenon under study defines the required response time. Response times



smaller than 100 ms can rarely be obtained with electrodes with tip areas of diameter 1  $\mu\text{m}$ .

### Volume Measurements

It is possible to record changes in cell volumes by ion-selective microelectrodes. The method utilizes the fact that the  $\text{K}^+$ -ion exchanger, i.e., Corning 477317, is much more sensitive to choline ions ( $\text{Ch}^+$ ,  $\log K_{\text{K}^+/\text{Ch}^+} \approx 2$ ) and to tetramethylammonium ions ( $\text{TMA}^+$ ,  $\log K_{\text{K}^+/\text{TMA}^+} \approx 3$ ) than to  $\text{K}^+$ . If cells are incubated with  $\text{Ch}^+$  or  $\text{TMA}^+$ , they usually take up these ions in small quantities,  $\text{Ch}^+$  to  $\sim 1\text{--}5 \text{ mmol l}^{-1}$  and  $\text{TMA}^+$  to  $\sim 0.1\text{--}0.5 \text{ mmol l}^{-1}$ . These levels are easily seen by the  $\text{K}^+$  electrode. Since  $\text{Ch}^+$  and  $\text{TMA}^+$  cross the cell membrane slowly, induced changes in their intracellular concentration will in many experiments reflect changes in cell volume. In this way the water permeability of the cell membrane can be obtained from the response to abrupt changes in extra cellular osmolarity.

### Metal Microelectrodes

Metal microelectrodes are used to record electrical activity in, for example, biological tissues. They have advantage over glass micropipettes in mechanical strength and durability. Their disadvantage is that they cannot reproduce slow electrical potential changes (DC changes) since their impedance is mainly capacitive. They are well suited to recording action potentials from neural tissue, but the recorded potential will be distorted owing to the complex impedance of the solution–metal interphase.

### Construction

Platinum microelectrodes can be made from platinum–iridium wire (70% platinum and 30% iridium), 0.25 mm in diameter sharpened electrolytically to a tip diameter of  $\sim 1 \mu\text{m}$  and to a length of the shank of  $\sim 5 \text{ mm}$ . The electrode can be insulated by a layer of glass (Corning 7570, borosilicate glass) that has the same thermal expansion coefficient as the platinum–iridium alloy. The sharpened wire is passed through a drop of molten glass. The tip is bared and platinum-black is deposited on it by applying a negative pulse of 5–15 V for 100 ms to the electrode with the tip immersed in a solution of platinum chloride (0.1%  $\text{H}_2\text{PtCl}_6$ , 0.25% platinum acetate, and 0.025% ammonium chloride).

Tungsten microelectrodes can be made from wire 0.25 mm in diameter that is sharpened electrolytically to a tip diameter of  $\sim 1 \mu\text{m}$  and to a length of the shank of  $\sim 10 \text{ mm}$ . The pointed wire can be insulated by dipping it into lacquer. If the tip is lifted

out first, the tip will be bared for 10–100  $\mu\text{m}$  in length.

### Electrical Properties

The resting potential of metal microelectrodes in biological tissue depends on several factors: the tissue, the material of the particular electrode, the pretreatment of the electrode, and finally the leakage current of the amplifier. The major determinant of the potential will be the magnitude and direction of the leakage current. For the platinum electrode this current cannot pass the electrode–solution interphase unless carried by an electron. The potential will therefore shift to a potential at which this can take place; thus, the potential will depend on the redox potential of the substances in the external medium. For the tungsten electrodes the resting potential will be influenced by the formation of tungsten oxides. With a leakage current of  $\sim 10^{-12} \text{ A}$ , the platinum microelectrode will typically settle somewhere between  $-0.2$  and  $+0.2 \text{ V}$  relative to the  $\text{Ag}/\text{AgCl}$  electrode, and the tungsten electrode somewhere between  $-0.4$  and  $-1.25 \text{ V}$ . For larger positive leakage currents, both electrodes will attain a positive potential of  $\sim 1.3 \text{ V}$  where oxygen and chlorine are formed; larger negative currents will result in a potential of  $1.4 \text{ V}$  where hydrogen is formed.

The small-signal impedances of the platinum and the tungsten electrodes are mainly capacitive. For the platinum electrodes the impedance is given by a capacitance of  $\sim 10 \mu\text{F mm}^{-2}$ ; the average electrode has a surface area of  $10 \mu\text{m}^2$ . The tungsten electrode has a capacitance of  $0.4 \mu\text{F mm}^{-2}$ ; the average electrode surface area is  $30 \mu\text{m}^2$ . It is noticed that platinum-black increases the effective surface area compared to the tungsten electrode.

The electrodes are mainly used to record and stimulate excitable neural tissues. When used for recording of neural activity, the electrodes are connected to an amplifier with a high-input resistance  $> 10^{10} \Omega$  and low leakage current ( $< 10^{-12} \text{ A}$ ); owing to the unstable resting potential, recordings of low frequency are filtered out. When the electrodes are used for stimulation of excitable tissues, current densities of  $100 \mu\text{A mm}^{-2}$  may often be required. In case of positive currents this will shift the electrode potential from  $+1.2$  to  $1.4 \text{ V}$ , the range of redox potentials for  $\text{OH}^-$  and  $\text{Cl}^-$ . This is because transfer of an electron from  $\text{Cl}^-$  and  $\text{OH}^-$  is the only process that can supply sufficient current. For negative currents the potential will move to  $-1.2 \text{ V}$ , because  $\text{H}^+$  will accept sufficient electrons at this potential. For positive potentials the pH around the stimulation electrode will attain values of 3, at negative currents pH 11–14

may be obtained. Such gross changes in pH must be taken into account when the condition of the tissue is considered.

## Oxygen- and Enzyme-Based Microelectrodes

The ambient oxygen concentration can be measured by metal microelectrodes. If the potential of the electrode is maintained at about  $-600$  mV relative to an Ag/AgCl reference electrode, current will flow according to the reactions  $\text{O}_2 + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{H}_2\text{O}_2$  and  $\text{H}_2\text{O}_2 + 2\text{H}^+ + 2\text{e}^- \rightarrow 2\text{H}_2\text{O}$ . The current will therefore be determined by the availability of oxygen and its diffusion toward the electrode surface. The electrode is usually coated with a layer of poly(vinyl chloride) or silicone rubber of a well-defined thickness in order to ensure a constant length of the diffusion for oxygen. This ensures that the current is proportional to the oxygen concentration in the medium irrespective of the position and environment of the electrode.

In principle, microelectrodes can be made specifically for a given substrate provided an enzyme exists that breaks down the substrate with subsequent formation of oxygen, hydrogen peroxide, or hydrogen, substances that can be measured amperometrically by the metal microelectrode. In this way a glucose microelectrode can be constructed from a metal microelectrode. The bared metal surface is coated with a layer of glucose oxidase that is separated from the outer solution by a layer of PVC or silicone rubber of a well-defined thickness. Provided oxygen is freely available, the rate of the process  $\text{glucose} + \text{O}_2 \rightarrow \text{gluconic acid} + \text{H}_2\text{O}_2$  will be catalyzed by the enzyme, and the concentration of glucose will be practically zero at the metal surface. The diffusion of glucose toward the metal will therefore be determined by the glucose concentration in the external medium and by the thickness of the protective membrane. Since this thickness remains constant, the rate of glucose breakdown will depend on the external glucose concentration only. The metal electrode is maintained at a potential of  $-300$  mV relative to an external Ag/AgCl reference electrode and the current will flow according to the reaction  $\text{H}_2\text{O}_2 + 2\text{H}^+ + 2\text{e}^- \rightarrow 2\text{H}_2\text{O}$ . This means that the current will be proportional to the rate of  $\text{H}_2\text{O}_2$  formation and therefore to the glucose concentration in the external solution. It takes considerable skill, patience, and careful calibration to make these electrodes, and so far they have mainly been applied in research laboratories.

In some situations the electrode surface can be coated with ferrocene, which allows the direct

transference of electrons between the enzyme and the surface of the electrode.

## In Vivo Potentiometry

It was seen in the previous section that a metal surface polarized to a given potential can be used to measure oxygen and hydrogen peroxide. The same principle can be used to monitor the concentrations of substrates that are themselves oxidized or reduced, i.e., take up or give up an electron at the metal surface. It should be emphasized that the useful range of potentials in aqueous solutions is limited by the potentials at which  $\text{OH}^-$  is oxidized and  $\text{H}^+$  is reduced, about  $+1$  and  $-1.2$  V relative to an Ag/AgCl reference electrode, respectively. Important possibilities are certain neurotransmitters, the oxidation of which is best performed on carbon electrodes instead of metal electrodes.

Carbon microelectrodes are made from fibers of  $1\text{--}10\text{ }\mu\text{m}$  outer diameter. These can be inserted into and supported by a glass micropipette, and the desired recording area can be obtained by varying the length of the protruding part of the fiber. Ultimately, the fiber can be cut flush with the glass capillary, whereby a small active surface is obtained. The small size decreases the signals obtained, but the spherical diffusion geometry ensures in effect a high degree of stirring around the tip. It is also possible to sharpen the tip by gradual burning with a propane-air mixture at temperatures of  $1300\text{--}1400^\circ\text{C}$ . In this way, sharpened tips with length of  $250\text{--}500\text{ }\mu\text{m}$  can be obtained while the diameter at the very tip is less than  $0.5\text{ }\mu\text{m}$ . This enables insertion into single cells. The tapered part can be coated to expose only the outer  $2\text{--}6\text{ }\mu\text{m}$  of the tip.

It is often necessary to pretreat the carbon surface electrically; positive potentials of up to  $+3$  V and negative ones down to  $-1$  V are applied alternately for various periods of time. The precise treatment should be chosen for each specific application.

For certain purposes it is useful to coat the electrode with a negatively charged membrane. Since the electrode surface is polarized to a positive potential, such a membrane will exclude negatively charged substrates while allowing neutral ones to penetrate. One choice is Nafion<sup>®</sup>, a perfluorosulfonated derivative of Teflon<sup>®</sup>.

Differential pulse voltammetry is the method of choice when compounds with similar oxidation potentials are to be resolved. The method is based on linear sweep voltammetry, where the potential applied to the electrode is increased progressively at a rate of  $5\text{--}20\text{ mVs}^{-1}$  up to about  $+1$  V. On this ramp are superimposed step potentials of constant

amplitude and duration (e.g., 50 mV lasting 50 ms). The resulting current is measured immediately before ( $i_b$ ) and after ( $i_a$ ) the application of the step potential and the difference ( $i_b - i_a$ ) is recorded as a function of the applied potential. The method in fact records the rate of increase in current and specific oxidations will appear as peaks in the current-voltage diagram.

### Specific Applications

Carbon and ceramic electrodes have been used for quantitative and qualitative recordings of the neurotransmitters dopamine (DA), 5-hydroxytryptamine (5-HT), and their metabolites 3,4-dihydroxyphenylacetic acid (DOPAC), 5-hydroxyindoleacetic acid (5-HIAA), and L-glutamate. Nitric oxide can also be recorded. In this case the carbon surface of the electrode is covered with a p-type semiconducting polymeric porphyrin that acts as a catalyst for the oxidation of nitric oxide. Nitric oxide is responsible for the activity of the endothelium-derived relaxing factor associated with hypertension, diabetes, ischemia, and arteriosclerosis.

One of the problems of recording dopamine, 5-HT, and nitric oxide is that other substances present have oxidation potentials similar to the substrate under study. The metabolites of dopamine (DOPAC) and of 5-HT (5-HIAA) are oxidized at the same potential as their parent substances, +0.2 V for DA and +0.4 V for 5-HT. Furthermore, they may be present *in vivo* at concentrations two orders of magnitude higher than that of their parent substances. In addition, ascorbic acid present in large concentrations is oxidized in the same voltage range as DA and 5-HT. Coating the electrode with a negatively charged membrane (Nafion<sup>®</sup>) will to a large extent alleviate these problems since the membrane will keep the metabolites from reaching the electrode surface. In case of the nitric oxide electrode, the Nafion<sup>®</sup> membrane seems to stabilize the oxidation product  $\text{NO}^+$  and to reduce the

sensitivity of the electrode for  $\text{NO}_2^-$  to such an extent that the reactivity of the electrode to  $\text{NO}_2^-$  does not interfere with the recording of nitric oxide.

**See also:** **Ion-Selective Electrodes:** Glass; Liquid Membrane. **Voltammetry:** Linear Sweep and Cyclic.

### Further Reading

- Amman D (1986) *Ion-Selective Microelectrodes, Principles, Design and Application*. Berlin: Springer.
- Boulton AA, Baker GB, and Vanderwolf CH (eds.) (1990) *Neurophysiological Techniques. Basic Methods and Concepts*, pp. 321–383. Clifton, NJ: Humana Press.
- Brown KT and Flaming DG (1986) *Advanced Micropipette Techniques for Cell Physiology*. IBRO Handbook Series: Methods in the Neurosciences, 9. Chichester: Wiley.
- Carden DE, Diamond D, and Miller AJ (2001) An improved  $\text{Na}^+$ -selective microelectrode for intracellular measurements in plant cells. *Journal of Experimental Botany* 52: 1353–1359.
- Katz LC and Grinvald A (2002) New technologies: Molecular probes, microarrays, microelectrodes, microscopes and MRI. *Current Opinion in Neurobiology* 12: 551–553.
- Kessler M, Harrison DK, and Höper J (eds.) (1985) *Ion Measurements in Physiology and Medicine*. Berlin: Springer.
- Malinski T and Taha Z (1992) Nitric oxide release from a single cell measured *in situ* by a porphyrine-based microsensor. *Nature* 358: 676–679.
- Selectophore (1988) *Ionophores for ion-selective electrodes*. Buchs, Switzerland: Fluka Chemie.
- Thomas RC (1978) Ion-sensitive intracellular microelectrodes. How to make and use them. In: Treherne JE and Rubery PH (eds.) *Biological Techniques Series*. Orlando: Academic Press.
- Zeuthen T (1981) The application of ion-selective microelectrodes. In: Dingle JT and Gordon JL (eds.) *Research Monographs in Cell and Tissue Physiology*, vol. 4. Amsterdam: Elsevier/North-Holland Biomedical Press.

# MICROSCOPY

## Overview

**P J Evennett and C Hammond**, University of Leeds, Leeds, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## What is Microscopy?

Microscopy, the use of a microscope, refers to those techniques by which investigations are made on fine details that cannot be discerned with the naked eye. It covers techniques of observation ranging from the use of a single lens or 'magnifying glass' to the

amplitude and duration (e.g., 50 mV lasting 50 ms). The resulting current is measured immediately before ( $i_b$ ) and after ( $i_a$ ) the application of the step potential and the difference ( $i_b - i_a$ ) is recorded as a function of the applied potential. The method in fact records the rate of increase in current and specific oxidations will appear as peaks in the current-voltage diagram.

### Specific Applications

Carbon and ceramic electrodes have been used for quantitative and qualitative recordings of the neurotransmitters dopamine (DA), 5-hydroxytryptamine (5-HT), and their metabolites 3,4-dihydroxyphenylacetic acid (DOPAC), 5-hydroxyindoleacetic acid (5-HIAA), and L-glutamate. Nitric oxide can also be recorded. In this case the carbon surface of the electrode is covered with a p-type semiconducting polymeric porphyrin that acts as a catalyst for the oxidation of nitric oxide. Nitric oxide is responsible for the activity of the endothelium-derived relaxing factor associated with hypertension, diabetes, ischemia, and arteriosclerosis.

One of the problems of recording dopamine, 5-HT, and nitric oxide is that other substances present have oxidation potentials similar to the substrate under study. The metabolites of dopamine (DOPAC) and of 5-HT (5-HIAA) are oxidized at the same potential as their parent substances, +0.2 V for DA and +0.4 V for 5-HT. Furthermore, they may be present *in vivo* at concentrations two orders of magnitude higher than that of their parent substances. In addition, ascorbic acid present in large concentrations is oxidized in the same voltage range as DA and 5-HT. Coating the electrode with a negatively charged membrane (Nafion<sup>®</sup>) will to a large extent alleviate these problems since the membrane will keep the metabolites from reaching the electrode surface. In case of the nitric oxide electrode, the Nafion<sup>®</sup> membrane seems to stabilize the oxidation product  $\text{NO}^+$  and to reduce the

sensitivity of the electrode for  $\text{NO}_2^-$  to such an extent that the reactivity of the electrode to  $\text{NO}_2^-$  does not interfere with the recording of nitric oxide.

**See also:** **Ion-Selective Electrodes:** Glass; Liquid Membrane. **Voltammetry:** Linear Sweep and Cyclic.

### Further Reading

- Amman D (1986) *Ion-Selective Microelectrodes, Principles, Design and Application*. Berlin: Springer.
- Boulton AA, Baker GB, and Vanderwolf CH (eds.) (1990) *Neurophysiological Techniques. Basic Methods and Concepts*, pp. 321–383. Clifton, NJ: Humana Press.
- Brown KT and Flaming DG (1986) *Advanced Micropipette Techniques for Cell Physiology*. IBRO Handbook Series: Methods in the Neurosciences, 9. Chichester: Wiley.
- Carden DE, Diamond D, and Miller AJ (2001) An improved  $\text{Na}^+$ -selective microelectrode for intracellular measurements in plant cells. *Journal of Experimental Botany* 52: 1353–1359.
- Katz LC and Grinvald A (2002) New technologies: Molecular probes, microarrays, microelectrodes, microscopes and MRI. *Current Opinion in Neurobiology* 12: 551–553.
- Kessler M, Harrison DK, and Höper J (eds.) (1985) *Ion Measurements in Physiology and Medicine*. Berlin: Springer.
- Malinski T and Taha Z (1992) Nitric oxide release from a single cell measured *in situ* by a porphyrine-based microsensor. *Nature* 358: 676–679.
- Selectosphere (1988) *Ionophores for ion-selective electrodes*. Buchs, Switzerland: Fluka Chemie.
- Thomas RC (1978) Ion-sensitive intracellular microelectrodes. How to make and use them. In: Treherne JE and Rubery PH (eds.) *Biological Techniques Series*. Orlando: Academic Press.
- Zeuthen T (1981) The application of ion-selective microelectrodes. In: Dingle JT and Gordon JL (eds.) *Research Monographs in Cell and Tissue Physiology*, vol. 4. Amsterdam: Elsevier/North-Holland Biomedical Press.

# MICROSCOPY

## Overview

**P J Evennett and C Hammond**, University of Leeds, Leeds, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## What is Microscopy?

Microscopy, the use of a microscope, refers to those techniques by which investigations are made on fine details that cannot be discerned with the naked eye. It covers techniques of observation ranging from the use of a single lens or 'magnifying glass' to the

imaging of atoms, and provides information about the structure, distribution, organization, and chemical composition of objects. Some newly developed forms of microscopy technically have little in common with traditional types of microscope, but are nevertheless considered to be microscopy since they fulfill the basic function of a microscope, that of providing information about fine details in an object.

Microscopes have been used since the middle of the seventeenth century, and are now applied in all branches of scientific research, and are in routine use in education, hospitals, and industry. Indeed, it is considered that virtually all industrial products rely on the microscope at some stage, in development of the materials and techniques (metals, ceramics, polymers, food), in manufacture and testing, or in failure analysis.

The limitations of the human eye as an instrument for the study of fine detail are overcome by three important attributes of a microscope: resolution, contrast, and magnification. It is clear that magnification is a prime requirement, resulting in an enlarged image from which the sensory elements of the retina of the eye can gather more information than from the smaller, naked-eye image. Though essential, magnification is relatively easily achieved by any form of microscope, and attention should be directed principally toward resolution or resolving power: the ability of a system to make information about fine detail distinguishable in an image. The human eye can detect only variations in brightness and/or color in an image, and these variations, or contrast, must be adequately large. Some optical or chemical properties of objects cannot be discerned by the unaided eye; techniques of contrast enhancement enable information about these properties to be converted by the microscope and presented to the eye in the form of variations in brightness and/or color.

With the exception of lensless devices such as scanning tunneling and near-field microscopes, which have the same fundamental attributes though operating according to different principles, the essential feature of a microscope is the interaction with the specimen of a beam of electromagnetic waves (light) or particles that possess wave-like properties (high-energy electrons). The beam becomes modified as a result of this interaction, carrying information about the specimen, and this information is presented in the form of an image. It is important to note that with a microscope one does not observe the specimen, but an image of the specimen, which will necessarily be a less-than-perfect representation of it. Deficiencies in the quality of the image may be due to inadequate equipment or poor technique, but an important fundamental limitation is set by the laws of

physics and the phenomenon of diffraction. The essential difference between light and the other imaging agents is their wavelength, the factor that sets the ultimate limit of resolution. The conventional light microscope closely approached its theoretical limit of resolution at the end of the nineteenth century; electron microscopes have not yet reached the limit imposed by wavelength, because of the aberrations of the lenses. In addition to the limitations due to the instrument and its operation, the information content of microscopical images is subject to artifacts of specimen preparation, which may include special procedures such as sectioning, staining, and etching. Considerable skill and experience may be necessary in interpreting the image that, though it may be presented as a 'picture of the specimen', should not be accepted without critical consideration of artifacts of preparation and the origins of contrast, and the application of appropriate analytical procedures.

## Theory and Basic Concepts

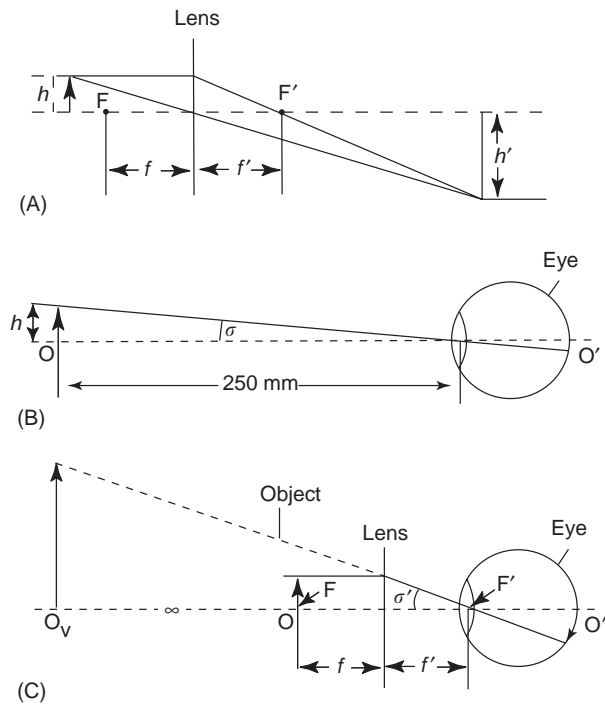
Microscopes that use lenses can be classified in several ways: (1) simple, single-lens instruments (magnifying glasses) or compound microscopes using an objective lens together with an eyepiece; (2) microscopes in which the entire image is presented simultaneously as in normal direct vision, or scanning microscopes in which the image is built up sequentially, point by point, as with television; or (3) classified according to the imaging medium (light, electrons, sound waves, or other forms of radiation or particle). Despite their important differences, all these kinds of microscope have much in common, and are subject to the same fundamental limitations. The description of the fundamentals of microscopy that follows is written in terms of the conventional light microscope; the important features of other forms of microscope are described elsewhere.

### Lenses

The principal lenses of microscopes are positive or converging lenses: thicker in the center than at their edges. In the accompanying diagrams, for simplicity, lenses will be shown as single pieces of glass, or represented by the center plane of the lens (as in **Figure 1**), though in practice they consist of several lens elements of different kinds of glass, carefully designed to minimize the aberrations (defects) of single lenses.

Parallel rays of light, as if from infinity, entering normal to the center plane of a converging lens on one side are refracted so that they intersect on the other side of the lens in the focal plane (**Figure 1A**);





**Figure 1** Magnification by a single lens. (A) Lateral magnification, showing the formation of a magnified image,  $h'$ , of object  $h$ . (B and C) Angular magnification. The viewing angle is increased from  $\sigma$  without the lens, with the object O at the reference viewing distance of 250 mm (diagram b), to  $\sigma'$  when using the lens (diagram C). The size of the image on the retina is thus increased. (Reproduced with permission from Bradbury S, Evennett PJ, Haselmann H, and Piller H (1989) *RMS Dictionary of Light Microscopy*. Oxford: Oxford University Press and Royal Microscopical Society.)

the focal plane crosses the optical axis at the focal point (F). Light originating at a source in the focal plane leaves the lens as parallel rays on the other side. The distance between the focal plane and the center of a (thin) lens is known as the focal length ( $f$  and  $f'$ ). A lens has two focal planes (passing through F and F', respectively), one at each side of the lens (the front and back focal planes); in discussing the operation of the microscope it is frequently necessary to consider both of these focal planes for each lens. Rays from sources situated between infinity and one focal plane (e.g.,  $h$  in Figure 1A) are brought to a focus on the other side of the lens, between the other focal plane and infinity (e.g.,  $h'$ ). The converging action of a lens is due to the curvature of its surfaces: the more highly curved they are (i.e., the smaller their radius of curvature), the 'stronger' the lens (i.e., the shorter its focal length).

### Magnification

The term magnification is used to describe two different phenomena, both of which occur in the

microscope. A lens may form a real image, in the same way as a transparency is projected onto a screen. Distances in the image may be measured and compared with corresponding distances in the object; the ratio of these distances, derived from linear measurements in object and image, is known as linear magnification. Since these measurements are lateral to the optical axis, this should strictly be termed lateral linear magnification. By convention, the scale of a map (a reduced image) is shown as a ratio (e.g., 1:50 000); the scale of a real, magnified image should similarly be indicated by a ratio (e.g., 100:1).

The second use of the term magnification refers to that provided by a single lens acting as a magnifying glass or 'simple microscope'. Here, the lens alone provides no real image (i.e., one that may be received on a screen and measured), though we consider that we see a magnified image when the lens is used in conjunction with the eye, which must, in fact, be considered to be an essential part of the optical system.

The apparent size of an object depends on its distance from the eye, which determines the viewing angle; the closer the object, the greater the angle it subtends at the eye, and the larger the area its image covers on the retina, and the bigger it appears. At distances closer than  $\sim 250$  mm, the unaided normal human eye becomes unable to produce a sharply focused image, even though the object looks larger due to its close proximity. A lens or magnifier enables us to see such a close object in sharp focus, with an enlarged viewing angle (Figure 1B and C). In order to express the degree of angular magnification achieved, it is necessary to have a standard against which the magnified image may be compared. By international agreement, this standard is taken to be the image at the closest distance at which an 'average' person can see clearly, namely 250 mm, known as the reference viewing distance. At this distance the unaided normal eye perceives its largest sharp image of an object, and can resolve  $\sim 0.1$  mm ( $100\ \mu\text{m}$ ). A magnifier with a focal length of 25 mm provides a sharply focused image of an object held 25 mm away from the lens (one-tenth of the reference viewing distance); the viewing angle is enlarged tenfold. A magnifier thus provides angular magnification, which by convention is expressed with a multiplication sign (e.g.,  $10\times$ ).

### The Simple Microscope

Figure 1C shows how a single, positive lens can act as a magnifier. Typically, the object is situated in the first focal plane of the magnifier, and the center of the lens of the eye in the second focal plane.

The shorter the focal length ( $f$ ) of the magnifier, the larger the increase in viewing angle it provides,

and the higher its magnification. Because the focal length of a lens is inversely proportional to the radius of curvature of its surfaces, high-magnification lenses are necessarily small. Moreover, they need to be used uncomfortably close to the eye. The practical upper limit to the magnification of a simple lens is  $\sim 25\times$ , though Leeuwenhoek made his remarkable observations in the seventeenth century using microscopes with tiny single lenses magnifying several hundredfold.

### The Compound Microscope

The compound microscope achieves higher magnifications with greater operating convenience and versatility by dividing the magnifying process into two stages, first at the objective lens, which provides a real image with magnification up to  $\sim 100:1$ , and the second using a 'magnifying glass' (the eyepiece) of relatively low magnification (usually  $\sim 10\times$ ).

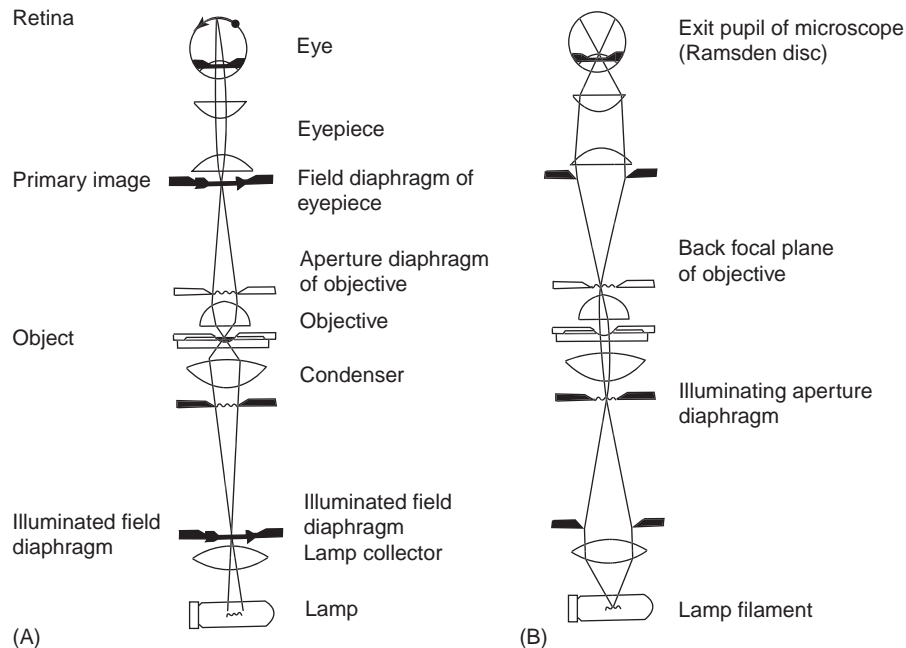
In the traditional design of microscope, the (transparent) *specimen* is situated just before (in front of, or below) the front focal plane of the objective lens (Figure 2A). A magnified image of the specimen (the *primary image*) is formed toward the top of the microscope tube, just as a projector images a slide onto the screen. The primary image is arranged to fall in the front focal plane of the eyepiece, where its extent is sharply defined by a field diaphragm. The eyepiece, operating like a magnifying glass, together with the

lens of the eye, transfers the image onto the *retina*. These three locations within the microscope shown here in *italics* are called conjugate planes; a feature present in one such plane will be imaged into all subsequent planes.

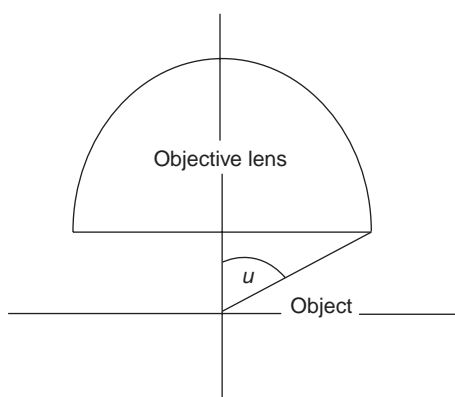
The primary image is a real image, an image that may be received on a screen, and in which distances might be measured. Comparison of measurements of distances in the object plane and image plane gives the lateral magnification of the primary image, which can also be calculated from the ratio of the objective-to-image distance to the specimen-to-objective distance. Since these distances are normally fixed in conventional microscope design, the magnifying power of an objective can also be considered to be fixed; this is the figure engraved on the lens mount. As with the simple microscope, the focal length of the eyepiece lens determines its magnifying power; extremely high or low values are inconvenient and unnecessary, and most eyepieces have a magnifying power between  $\sim 8\times$  and  $15\times$ .

### Numerical Aperture

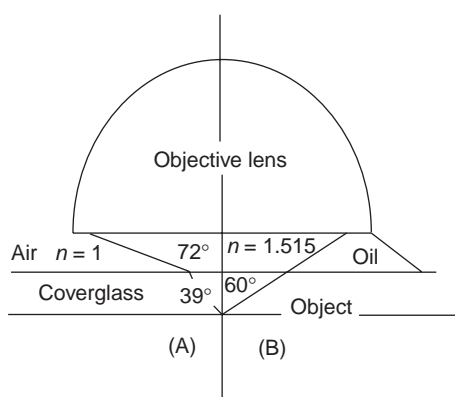
The information-gathering power of an objective lens depends on its ability to collect the oblique rays leaving the object, a factor that depends on the angular aperture ( $u$  in Figure 3) subtended by the lens at the object. Because of the effect of refraction at the



**Figure 2** Ray-paths in the transmitted-light microscope showing (A) the field set of planes that are conjugate with the object and the final image and (B) the aperture set of planes conjugate with the filament and the apertures of condenser and objective lenses. (Reproduced with permission from Bradbury S, Evennett PJ, Haselmann H, and Piller H (1989) *RMS Dictionary of Light Microscopy*. Oxford: Oxford University Press and Royal Microscopical Society.)



**Figure 3** The angular aperture,  $u$ , subtended by an objective lens at the object.



**Figure 4** The effect on numerical aperture of the refractive index,  $n$ , of the medium between the object and the objective lens. In the examples shown, in (A), where the medium is air, the most oblique rays accepted by the objective leave the object at  $\sim 39^\circ$ ; with oil-immersion (B), rays leaving at up to  $60^\circ$  are collected.

surface of the specimen (**Figure 4**), the refractive index ( $n$ ) of the medium between the objective and the object must also be taken into account. Both factors combine to express the aperture of a lens in terms of a number, the numerical aperture (NA), calculated as

$$NA = n \sin u$$

Higher resolving power is provided by objectives of large NA, which in turn derives from a large angular aperture and/or an observing or ‘immersion’ medium of high refractive index (RI). In the common case of observation through air ( $n=1$ ), NA depends solely on  $\sin u$ , the maximum mathematical value of which is 1; a practical maximum value for  $u$  is  $\sim 72^\circ$ , giving  $NA=0.95$ . An oil-immersion lens, which operates through oil of  $n=1.515$ , can provide an NA up to  $\sim 1.4$ . For the mathematical relationship between NA and resolving power, see below.

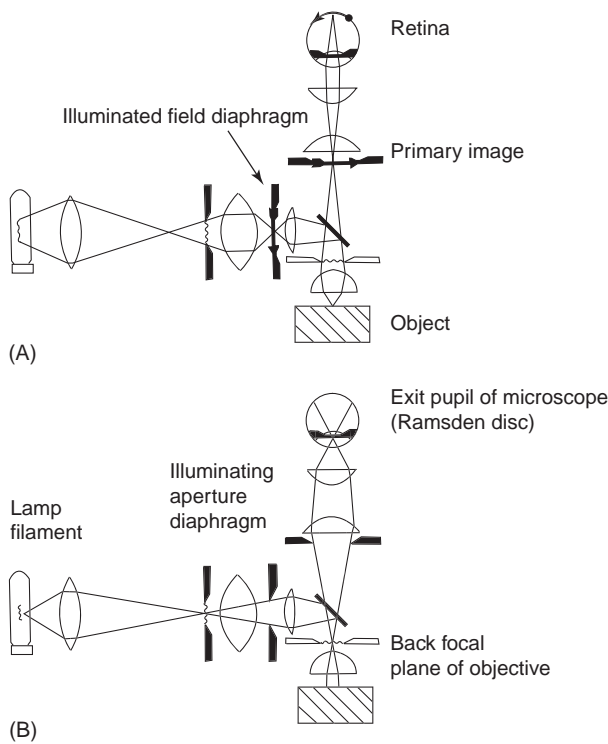
## Illumination

The illuminating system of a microscope must be devised so as to illuminate uniformly a controllable area of the specimen, and also to supply the objective lenses with a cone of light of adjustable angle, to suit their respective NAs. Since the early days of microscopy, uniform illumination has been achieved by throwing onto the object with a condenser lens an image of a uniformly illuminated area, such as a cloud in the sky or the flame of an oil-lamp. Köhler devised a method of illumination suitable for use with nonuniform light sources, such as electric lamps, which would cause disturbing irregularities in illumination across the field of view if imaged directly onto the object. Köhler illumination requires the light source to be placed close to the front focal plane of a lamp collector lens, so that an image of the filament falls on the front focal plane of the condenser; this arrangement causes the lamp collector lens to act as a uniformly illuminated area, which may then be imaged onto the object by the condenser lens. In practice, an iris diaphragm is fitted just after the lamp collector, and this diaphragm, the illuminated field diaphragm, is imaged onto the specimen and used to control the area of the object that is illuminated; the plane of this diaphragm is thus also conjugate with the specimen and the primary and final images (**Figure 2A**).

A second iris diaphragm, located in the front focal plane of the condenser, acts to limit the aperture of the condenser and hence the angular aperture of the cone of rays entering the specimen. This is known as the illuminating aperture diaphragm, and is used to adjust the angle of illumination according to the NA of the objective (**Figure 2B**). Adjustment of this diaphragm has a profound effect on the quality of the image, and its correct setting is necessary to optimize resolution and contrast.

## Illumination in the Reflected-Light Microscope

In addition to the more well-known transparent objects and thin sections, microscopes are also used to study nontransparent specimens. For observations in the simplest imaging mode, bright-field (see below), specimens are typically polished to give a mirror surface, which may then be etched to provide contrast. The principles of illumination for reflected light are the same as for transmitted light, except that the objective lens also acts as the condenser, light being directed into the objective from above, by a partially reflecting surface. Illuminated field and illuminating aperture diaphragms are provided, though in different locations from those for transmitted light (**Figure 5**). A reflected-light microscope



**Figure 5** Ray-paths in the reflected-light microscope showing (A) the field set of planes that are conjugate with the object and the final image, and (B) the aperture set of planes conjugate with the filament and the aperture of the combined condenser and objective lens. (Reproduced with permission from Bradbury S, Evennett PJ, Haselmann H, and Piller H (1989) *RMS Dictionary of Light Microscopy*. Oxford: Oxford University Press and Royal Microscopical Society.)

requires fewer adjustments in use than one for transmitted light, since the condenser is necessarily aligned with and of equal NA to the objective (they are one and the same lens), and area and aperture of illumination become altered automatically with change of objective.

### Diffraction and Resolution

One of the most important fundamental interactions between a specimen and the imaging radiation is that of diffraction, a phenomenon that results from the wave nature of the imaging agent. Diffraction occurs whenever a wave motion encounters an object, but its effects assume particular importance in microscopy, where the dimensions of features being imaged may be close to the wavelength. The role of diffraction as the limiting factor in the performance of the microscope was first elucidated by Ernst Abbe in 1873.

Diffraction is described in textbooks of physics, and no attempt to explain its origin will be made here. The essential features of the phenomenon that

are important to an understanding of microscopy are as follows:

1. When radiation from a distant source encounters an object, a series of new waves are formed, centered about features within the object.
2. Waves from adjacent features reinforce one another (or 'interfere constructively') in certain directions, giving rise to diffracted beams. One of these beams (known as the 'zero-order' beam) proceeds in the same direction as the incident radiation, with 'first-order', 'second-order' (etc.) beams propagated on each side of the zero-order beam.
3. The angles between the zero-order beam and the diffracted beams depend on the wavelength of the radiation and the spacing of the diffracting features. The angles are greater for longer wavelengths and closer-spaced features (finer detail).
4. The diffracted beams arising from the influence of the specimen on the imaging radiation carry the information about the fine details in the specimen. The primary image results from the interference between diffracted beams; it is thus essential for resolution that these beams be collected and allowed to contribute to the image.
5. The ability of an objective lens to collect a cone of diffracted rays emerging at different angles from the specimen is governed by the NA of the objective.

Diffraction thus provides the basis for the dependence of ultimate resolving power on wavelength and NA, though it must be recognized that other limitations due to lens aberrations may be superimposed upon this diffraction limit. Light of shorter wavelengths (e.g., light at the ultraviolet, blue, and green end of the spectrum) is diffracted through smaller angles and is thus more readily collected by an objective than light of longer wavelength (e.g., red); larger NAs admit beams diffracted by finer detail. The relationship between these factors may be calculated from several formulae, a typical one being

$$d_{\min} = 0.61\lambda/\text{NA}$$

where  $d_{\min}$  represents the finest spacing that may be resolved,  $\lambda$  the wavelength of the imaging radiation, and NA the numerical aperture of the objective lens.

Since NA determines the fineness of detail that an objective can present to the observer, it has a relationship with useful magnification (magnification is considered to be useful if it reveals details not discernible at a lower magnification; magnification that reveals no new information, merely enlarging the image, is known as empty magnification).

Low-magnification objectives are generally also of low NA, and higher-magnification objectives require larger NAs. However, these relationships are not fixed; for a given magnification, objectives of a range of NAs are available – for example, objectives giving a magnification of 40 may have NAs ranging from 0.65 up to 1.4.

### Lens Aberrations

A simple lens constructed from a single piece of glass exhibits several defects or aberrations that reduce its performance. These aberrations are minimized in the design of modern lenses, which are constructed using multiple lens elements of different materials and carefully calculated curvatures, the details of which need not concern the user. For the aberration corrections to be effective, however, it is important that the lenses be used as specified in their design.

Chromatic aberration is the inability of a simple lens to bring radiation of different wavelengths to a focus in one plane. It may be corrected to varying degrees of perfection: the simplest objectives are described as achromatic, and the most perfectly corrected as apochromatic, with an intermediate class including the mineral fluorite or its synthetic equivalent in their construction. In the design of the original apochromatic objectives, correction of chromatic aberration was shared between the objective itself and a special kind of eyepiece, the compensating eyepiece. Most modern objectives are virtually completely corrected within themselves, and do not require a compensating eyepiece. Because of this, and the different degrees of compensation used in different designs, it is important to use the correct eyepiece for a given objective.

Spherical aberration is the defect by which radiation passing through the central regions of a lens does not have a common focus with that passing through the periphery of the lens; it is considerably more serious with lenses of larger aperture. Correction of spherical aberration requires two factors to be specified: the distance between the objective lens and the primary image (which is governed by the mechanical tube-length), and the thickness and RI of media between the objective lens and the specimen, especially the coverglass and any immersion medium. Objectives are specified for use at a particular tubelength, previously 160 mm and nowadays more commonly infinity (see below), and thickness of coverglass, typically either 0.17 mm or none at all. Unless these conditions are complied with, a poor image will result.

Recent designs make use of the 'infinity corrected' system, which provides greater versatility in adapting the microscope to new techniques. In this system, the specimen lies precisely in the front focal plane of the

objective. The imaging rays pass parallel through the microscope tube, as if to infinity, and are made to converge and form the primary image by the inclusion of a tubelens. The parallel ray-path between objective and tubelens permits the insertion of beam splitters, filters, and other devices, without requiring special provision to be made for the refractive effects of these items, as required by systems using a finite tubelength. Aberrations are corrected for an image at infinity, and the objectives are thus not interchangeable with those of conventional design.

Simple objective lenses suffer from curvature of field, and are unable simultaneously to focus central and peripheral regions of their field of view of the specimen in a common plane in the image. This defect is especially noticeable in photomicrographs. More complex objectives designed to provide a flat field are commonly available, usually designated by the prefix or suffix 'plan' included in their name.

### Contrast

The degree of contrast observed between features that constitute the fine detail in the field of view depends on the nature of the specimen, its preparation procedures, and the techniques employed in imaging. The generation of contrast depends on the various interactions between the specimen and the imaging radiation, which may be summarized in the following way:

- Absorption, transmission, and reflection
- Scattering and diffraction
- Refraction and polarization
- Phase change
- Fluorescence

When light falls on to an object, some will be absorbed, and the remainder will be transmitted and/or reflected. These effects are familiar in normal vision, and they are important also in microscopy. Small, transparent specimens, and thin sections of larger ones, are frequently studied by transmitted light, and the absorption of light by their features may be differentially enhanced by procedures such as staining. If light of only certain parts of the spectrum (certain wavelengths) is absorbed, the features appear colored. A transmitting specimen may also scatter or diffract light such that some proportion of it is no longer directed into the objective lens.

Thicker, opaque specimens are studied by the light that they reflect, either in a specular manner, as from a mirror, or diffusely. Metal specimens, for example, may be polished to a mirror finish and the features of interest differentiated by means of etching. The specimen surface must be arranged to be perpendicular to



the optical axis. Other, unpolished, specimens will reflect light diffusely.

For both transmitted- and reflected-light microscopy, there are two basic imaging modes: bright field and dark field. A bright-field image is formed when, as already described, light falls on the specimen from any direction within the aperture of the objective – from beneath or from above. With transmitted light, an empty field appears fully illuminated, and absorbing or scattering features reduce intensity locally. Similarly, a featureless, specularly reflecting specimen is a perfect mirror, providing a bright, fully illuminated field: features that scatter or absorb light again appear in dark contrast against the bright field.

In dark-field microscopy, light is arranged to fall onto the specimen from outside the aperture of the objective. An empty field in a transmitting specimen and a featureless polished surface will both appear dark; in both cases, any feature that scatters light into the objective lens will appear bright.

For low-magnification observations in transmitted light, a normal condenser can be converted for dark-field imaging by inserting a central opaque stop of suitable diameter at the level of the illuminating aperture diaphragm; for work with high-magnification, large-aperture objectives, special dark-field condensers are required. Low-magnification, reflected-light, dark-field images may be improvised by directing an intense beam of light obliquely onto the surface of the specimen. For higher-magnification work it is desirable to use a special illuminator, together with objectives that incorporate an annular illuminating path encircling the imaging lenses, thus obtaining uniform illumination from all azimuths.

The RI of specimens that transmit light must be considered in the context of contrast. It is an advantage with certain specimens (notably thin, transparent biological objects or sections in which the contrast has been provided by differential absorption of light by stains) for the specimen to be infiltrated and enclosed in a medium of RI similar to that of the object itself. This provides an undeviated optical path through the specimen, reduces scattering of light, and thus renders the object clear.

Precise determinations of RI can be carried out using samples mounted in media of known RI, or whose changes of RI with temperature have been calibrated. When the RIs of object and medium are mismatched, a characteristic bright line (the Becke line) is seen at the interface. When the RIs are matched, contrast is lost. Since RI is different for light of different wavelengths (the phenomenon of dispersion), it is possible to characterize particles by the colored fringes at their interface with the mounting medium, a procedure known as ‘dispersion staining’.

Materials that have an ordered molecular structure may show birefringence: i.e., they exhibit different refractive indices for light vibrating in different directions. Such specimens may be made to exhibit contrast by polarized light microscopy. Within such birefringent materials, light waves vibrating in different directions with respect to the orientation of the molecules travel at different velocities and have different wavelengths. These waves may interfere, and be made to show contrast in the image by the use of a pair of polarizing filters or polars, inserted into the light path one on each side of the specimen, normally with their vibration directions crossed. Polarized light microscopes may also be used to characterize materials by quantitative determination of birefringence.

Many specimens, notably living cells, are non-absorbing, yet contain features that refract differently, due to variations in RI or thickness, or both. These normally invisible properties of uniformly transparent specimens may be exploited to produce contrast by the phase-contrast technique, devised by Zernike in 1933. Small variations in refraction in the specimen give rise to differences in phase of about a quarter of a wavelength between the diffracted beams and the zero-order beam. The phase-contrast technique increases this difference to a half-wavelength, resulting in destructive interference, and hence contrast, in the image. The extra phase shift for the diffracted beams is achieved selectively, without affecting the zero order, by the simple device of illuminating the specimen with a hollow cone of light that enters only part of the aperture of the objective. A ring of zero-order light thus falls on a defined area of a ‘phase plate’ within the objective. The diffracted light falls on the remainder of the phase plate, which is arranged to have a suitably greater optical thickness than the area through which the zero-order passes, so that the diffracted light acquires the necessary further one-quarter-wavelength phase difference. A characteristic of a phase-contrast image is a bright or dark halo that surrounds each contrasted feature; in direct observation this can be considered to enhance the contrast, but it may become an undesirable artifact if the images are to be used for quantitative work. Phase-contrast microscopy has been of great importance in the development of cell biology. It can also be used for observation of small height differences in reflecting specimens, but for such applications it has been largely superseded by differential interference contrast (DIC) (see below).

A more recent method of generating contrast from small phase-differences between the rays leaving an object is that of DIC, especially as developed by Nomarski. In this system, two images of the object are superimposed, laterally displaced by a distance so

small that it cannot be resolved by the objective in use. This is achieved by dividing each illuminating ray into two components using as a beam-splitter a 'Wollaston prism' situated below the condenser. Differences in RI within a transmitting specimen, or height differences in a reflecting one, cause phase differences between the members of each pair of beams. A second Wollaston prism behind the objective recombines the two beams, which then interfere, their phase-differences giving rise to contrast in the image. In order to function in this way, the Wollaston prisms must be used between crossed polars, as in polarized-light microscope. DIC in reflected light microscopy requires only one Wollaston prism, situated in the objective, which also acts as condenser.

Because its contrast results from a unilateral displacement of beams, DIC provides directional contrast within the image: the specimen appears as if illuminated from one side, and is perceived as if 'three-dimensional', an appearance that may genuinely represent the third dimension, or which may be an artifactual expression of differences of RI. DIC provides a high-resolution image that lacks the haloes of phase contrast, and has excellent 'optical-sectioning' properties for imaging selected planes of focus.

Fluorescence is the selective absorption of energy from radiation of short wavelength, followed very closely by the emission of radiation of longer wavelength. In microscopy, fluorescence is used to demonstrate features that either have the inherent property of fluorescing, or features to which a fluorescent marker (fluorochrome) can selectively be attached (for example, by the techniques of immunology, which are beyond the scope of this review).

Fluorescence microscopy has progressed from bright-field and dark-field transmitted-light configurations, to the now almost universal epi-fluorescence system. In essence, the modern epi-fluorescence microscope is similar to the bright-field reflected-light instrument described above, but with several important differences. It must be provided with a source of intense illumination, particularly rich in the shorter wavelengths that are required to excite most fluorochromes. In place of the partial reflector above the objective, the epi-fluorescence microscope has a special 'dichroic' mirror – a beam-splitter designed to reflect into the objective that part of the spectrum required for excitation of fluorescence, and to transmit the longer-wavelength light emitted by the specimen, from which the image is formed. The dichroic mirror is used in conjunction with carefully chosen filters so that their combined properties match the requirements of the absorption and emission spectra of the fluorochrome. Use of large-aperture objectives is especially desirable in fluorescence microscopy, to

provide not only high resolving power but also the brightest possible image, in order to maximize the sensitivity of the technique. The light transmission of an objective depends on the square of the NA; since the objective lens acts as both condenser and objective, and light passes through it twice, brightness in fluorescence microscopy is proportional to the fourth power of the NA of the objective.

### Confocal Scanning Light Microscopy

Many specimens for microscopy are translucent, with areas of interest lying beneath the surface. When they are observed in bulk, rather than physically cut into thin sections, whether by transmitted or reflected light or by fluorescence, light will reach the image from out-of-focus planes above and below the features of interest within the specimen. Confocal microscopes are designed to exclude from the image light arising from out-of-focus planes.

Illumination from a small source, usually a laser, is imaged as a small, intense spot of light in the plane of focus within the specimen. Since most confocal microscopes are used for epifluorescence or for reflected light, this spot is imaged onto the specimen by the objective lens; a corresponding design for transmitted light is theoretically possible though practically more complex. Light returning from this spot in the plane of focus is imaged onto a small hole that is accurately located in the primary image plane. The small source, the illuminating spot in the specimen and the 'pinhole' in the primary image are all arranged to be in planes that are conjugate, and are thus described as 'confocal' – in focus together. Two important consequences arise:

1. The illumination of features within the specimen that lie above and below the plane of focus will be considerably less intense than that in the plane of focus, since the rays from the source will either not yet have fully converged and concentrated their energy, or they will have diverged after the plane of focus.
2. Features above or below the plane of focus will be imaged as points lying above or below the confocal pinhole in the primary image plane; only a small proportion of the light from each of these out-of-focus features will pass through the pinhole.

Out-of-focus features are thus effectively ignored because they are illuminated less intensely and very little of the light from their images is collected.

Thus described, the confocal microscope collects light from only one spot within the specimen. Larger areas are sampled by scanning the specimen against a stationary beam or, more usually, by using mirrors to scan the beam in a raster pattern against a stationary specimen. Light passing through the primary image

pinhole is collected by a detector, and the output may be used to modulate the brightness of a television tube scanned in unison with the raster on the specimen. Images are normally digitized and stored in computer memory, and may be combined and manipulated to present information about the specimen as required.

In contrast to this electronic system, there is another approach to confocal imaging, the tandem scanning microscope, which uses a spinning disc to scan the specimen with a large number of beams simultaneously, and presents its image directly to the eye or photographic camera. This system may provide faster scanning than the single-beam electronic system, and thus has applications in studying dynamic events.

### Advantages and Limitations of Light Microscopy

The light microscope has many advantages over other forms of microscope. Light microscopes are extremely versatile instruments. They can be used to examine a wide variety of types of specimen, frequently with minimal preparation. Light microscopes can be adapted to examine specimens of any size, whole or sectioned, living or dead, wet or dry, hot or cold, and static or fast-moving. They offer a wide range of contrast techniques, providing information on the physical, chemical, and biological attributes of specimens. The image from a light microscope is presented in color. It can be observed with the eye directly, recorded by photographic, video or computer techniques, and image components can be analyzed. And the instruments themselves are

(relatively) inexpensive, small, require no vacuum, and are undemanding of operating conditions, services and maintenance.

The principal limitation of the light microscope is its resolving power. Using an objective of NA 1.4, and green light of wavelength 500 nm, the resolution limit is  $\sim 0.2 \mu\text{m}$ . This value may be approximately halved, with some inconvenience, using ultraviolet radiation of shorter wavelengths. By good fortune, however, the resolving power of the light microscope is adequate for many areas of work in the biological and physical sciences.

*See also: Microscopy Techniques: Light Microscopy.*

### Further Reading

- Bradbury S and Bracegirdle B (1998) *Introduction to Light Microscopy*. Oxford: Bios Scientific Publishers and Royal Microscopical Society.
- Murphy DB (2001) *Fundamentals of Light Microscopy and Electronic Imaging*. New York: Wiley-Liss, Inc.
- Pluta M (1988) *Advanced Light Microscopy. Vol. 1 – Principles and Basic Properties*. Warsaw: Polish Scientific Publishers.
- Pluta M (1989) *Advanced Light Microscopy. Vol. 2 – Specialized Methods*. Warsaw: Polish Scientific Publishers.
- Pluta M (1993) *Advanced Light Microscopy. Vol. 3 – Measuring Techniques*. Warsaw: Polish Scientific Publishers.
- Richardson JH (1991) *Handbook for the Light Microscope: A User's Guide*. New Jersey, USA: Noyes Publications.
- Robinson PC and Bradbury S (1992) *Qualitative Polarized-Light Microscopy*. Oxford: Oxford University Press and Royal Microscopical Society.

## MICROSCOPY APPLICATIONS

Contents

**Biomedical**

**Food**

**Forensic**

**Environmental**

**Liquid Crystals**

**Semiconductors**

**Proteins and Nucleic Acids**

### Biomedical

**A Warley**, King's College, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Microscopic techniques are primarily used to determine the structural characteristics of a specimen under study. Light microscopy is used to show the overall distribution of different cell types and tissues

pinhole is collected by a detector, and the output may be used to modulate the brightness of a television tube scanned in unison with the raster on the specimen. Images are normally digitized and stored in computer memory, and may be combined and manipulated to present information about the specimen as required.

In contrast to this electronic system, there is another approach to confocal imaging, the tandem scanning microscope, which uses a spinning disc to scan the specimen with a large number of beams simultaneously, and presents its image directly to the eye or photographic camera. This system may provide faster scanning than the single-beam electronic system, and thus has applications in studying dynamic events.

### Advantages and Limitations of Light Microscopy

The light microscope has many advantages over other forms of microscope. Light microscopes are extremely versatile instruments. They can be used to examine a wide variety of types of specimen, frequently with minimal preparation. Light microscopes can be adapted to examine specimens of any size, whole or sectioned, living or dead, wet or dry, hot or cold, and static or fast-moving. They offer a wide range of contrast techniques, providing information on the physical, chemical, and biological attributes of specimens. The image from a light microscope is presented in color. It can be observed with the eye directly, recorded by photographic, video or computer techniques, and image components can be analyzed. And the instruments themselves are

(relatively) inexpensive, small, require no vacuum, and are undemanding of operating conditions, services and maintenance.

The principal limitation of the light microscope is its resolving power. Using an objective of NA 1.4, and green light of wavelength 500 nm, the resolution limit is  $\sim 0.2 \mu\text{m}$ . This value may be approximately halved, with some inconvenience, using ultraviolet radiation of shorter wavelengths. By good fortune, however, the resolving power of the light microscope is adequate for many areas of work in the biological and physical sciences.

*See also: Microscopy Techniques: Light Microscopy.*

### Further Reading

- Bradbury S and Bracegirdle B (1998) *Introduction to Light Microscopy*. Oxford: Bios Scientific Publishers and Royal Microscopical Society.
- Murphy DB (2001) *Fundamentals of Light Microscopy and Electronic Imaging*. New York: Wiley-Liss, Inc.
- Pluta M (1988) *Advanced Light Microscopy. Vol. 1 – Principles and Basic Properties*. Warsaw: Polish Scientific Publishers.
- Pluta M (1989) *Advanced Light Microscopy. Vol. 2 – Specialized Methods*. Warsaw: Polish Scientific Publishers.
- Pluta M (1993) *Advanced Light Microscopy. Vol. 3 – Measuring Techniques*. Warsaw: Polish Scientific Publishers.
- Richardson JH (1991) *Handbook for the Light Microscope: A User's Guide*. New Jersey, USA: Noyes Publications.
- Robinson PC and Bradbury S (1992) *Qualitative Polarized-Light Microscopy*. Oxford: Oxford University Press and Royal Microscopical Society.

## MICROSCOPY APPLICATIONS

Contents

**Biomedical**

**Food**

**Forensic**

**Environmental**

**Liquid Crystals**

**Semiconductors**

**Proteins and Nucleic Acids**

### Biomedical

**A Warley**, King's College, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Microscopic techniques are primarily used to determine the structural characteristics of a specimen under study. Light microscopy is used to show the overall distribution of different cell types and tissues

within an organ, and electron microscopy provides structural details at the cellular and subcellular levels. However, both light and electron microscopy can also provide information about the chemical composition of a specimen under study, although this capability is often not realized. There are a number of occasions in which such information may be required: in the biological field these include identification of elements from external sources, e.g., chemical pollutants, administered drugs, food contaminants; identification of inclusions in pathological diagnoses; the study of physiologically important elements; and as an aid in the study of histochemical methods. In this article, the stains that can be used for detection of elements at the light microscope level will be introduced, the basic theory underlying the ability to detect elements using electron microscopy (EM) will be outlined, precautions that need to be undertaken for analytical studies will be highlighted, and the major areas of application will be reviewed.

## Light Microscopy

Light microscopy can be used to demonstrate the presence of chemical elements at the tissue level. For example Perl's Prussian blue method is used for the detection of iron(III) and Von Kossa's silver reaction for the detection of bound calcium; other reactions can be used to detect gold, copper, lead, and zinc. It should be noted that these methods detect the bound form of an element; any free forms will generally be lost during the fixation period. For the detection of free calcium, light or confocal microscopy using fluorescent dyes have superseded the more traditional staining methods. In general, light microscopy is limited by both the range of elements for which stains are available and by the resolution of the light microscope, which confines observations at best to the level of individual cells. For unambiguous detection of elements and their localization at the cellular and subcellular level, EM techniques are required.

## Electron Microscopy

### Signal Generation

When a specimen is examined using an electron microscope the incident electron beam interacts with atoms in the specimen, leading to the generation of a number of signals that carry information about the chemical nature of the specimen being examined. In elastic scattering events electrons are deflected by atomic nuclei in the specimen and leave the specimen without loss of energy; these electrons contribute to the backscattered electron signal, which gives

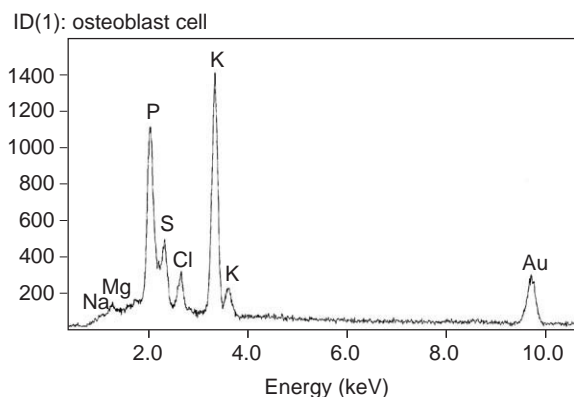
information about the atomic number of the atoms in the specimen. During inelastic scattering events energy is transferred from the incident electron beam to individual atoms in the specimen. These interactions lead to the generation of a number of signals. X-rays are generated as a consequence of the interaction between electrons of the incident beam, causing ejection of orbiting electrons from the atomic nuclei of the specimen, the energy of the emitted X-ray being characteristic of the atomic number of the element in which the original interaction occurs and of the orbital shells involved. Collection of these X-rays and analysis of their energy forms the basis of the technique of electron probe X-ray microanalysis (EPXMA), the EM technique most widely used for the study of biological specimens. Electrons that traverse the specimen suffer either no energy loss or lose the amount of energy that is imparted to the specimen during ionization events. Thus analysis of the transmitted electron signal also yields information about the elemental composition of the specimen and forms the basis of the technique of electron energy loss spectroscopy (EELS), which is increasingly being used for the study of biological materials.

### Detection Systems

Both techniques require that the microscope be fitted with suitable detection systems. For EPXMA two types of detector are available, wavelength-dispersive (WDS) and energy-dispersive (EDS) spectrometers. EDS systems are most widely applied in biology since these systems are easier to use, are capable of detecting a number of elements simultaneously, and require lower beam currents than WDS systems. A typical X-ray spectrum obtained using an EDS system (**Figure 1**) consists of a series of peaks superimposed upon a slowly varying curved background. Peaks in the spectrum denote the presence of a given element in the irradiated area: the area under the peak is proportional to the concentration of the element. Thus EPXMA is capable of providing both qualitative and quantitative information about the elemental content of the area under study. Elements of atomic number ( $Z$ ) greater than 11 can be detected if the usual beryllium-windowed EDS detector system is used; for the detection of elements of lower atomic number a thin-windowed detector is needed.

Two types of detection system are also available for EELS. Parallel detection spectrometers are fitted below the viewing chamber in a transmission electron microscope (TEM) or scanning transmission electron microscope (STEM). These spectrometers allow the detection of EELS spectra, which are plots





**Figure 1** An X-ray energy spectrum obtained from the analysis of a freeze-dried tissue culture cell. The spectrum is essentially a histogram plot of the number of counts against the X-ray energy. The spectrum consists of peaks that correspond to characteristic X-ray emissions from different elements. The curved background under the peaks is derived from the X-ray continuum radiation.

of the scattered electron intensity as a function of the decrease in kinetic energy (i.e., energy loss). The main component of the EELS spectrum is the intense zero loss peak. At higher energy losses steps or edges are superimposed on the decaying background intensity; these correspond to the excitation of inner shell electrons and can be used to detect both the presence of an element in the irradiated area and for quantitative analysis. In addition, variations in the shape and intensity of the edge structure can provide information about the ionization state of the atom.

The second type of EELS system is fitted in the column of a TEM and forms a specialized instrument. This spectrometer is sometimes called an imaging filter, and the technique is called electron spectroscopic imaging (ESI). In this type of instrument the entire inelastic image at a specific energy loss is captured by means of a 2D detector, allowing the distribution of a specific element to be obtained rapidly over a large sample area. In this mode thicker specimens (up to 1  $\mu\text{m}$ ) can be analyzed; however, using this type of instrument, spectral information has to be recorded one energy loss at a time.

### Types of Microscope Used

An EDS system can be fitted to a scanning electron microscope (SEM), TEM, or STEM. A TEM offers the ability to undertake analysis in small areas of thin sections of the specimen and is used when high-resolution analysis is required, whereas a SEM is used for the analysis of larger areas of the surface of bulk specimens and, in general, has much lower spatial resolution. EELS is a technique using a TEM or STEM.

## Capabilities

EPXMA and EELS are complementary techniques for the study of biological specimens. Of the two EPXMA is much simpler, the focusing of the beam onto the area of interest producing a spectrum in which the presence, or absence, of an element of interest is easily determined. All the elements present in the area of analysis are detected in a single analytical run. This has advantages: it eliminates the need for sequential screening for different elements and can also reveal the presence of unsuspected elements. However, it must be realized that the technique can be used to detect nuclides of elements only, and no information about the chemical state of the element can be inferred.

EPXMA is capable of providing information with high spatial resolution. In thin sections (100 nm thick), spatial resolutions of the order of 10–20 nm have been reported. This figure depends on many factors such as the type of instrument used, the internal microscope geometry, the electron emitter, the accelerating voltage, and the thickness and composition of the specimen. The spatial resolution decreases as the specimen thickness and/or the mass under the beam increases; the resolution is also decreased with lower accelerating voltages. A minimum detectable mass of  $10^{-18}$  g of iron has been obtained in a single analytical run of 100 s. However, for biological applications the minimum detectable concentration of an element in the specimen mass is more relevant. In a thin section a minimum detectable concentration in the region of 1 mmol per kilogram of dry mass of the specimen localized under the beam should be possible with a practical analysis time of 100 s. This sensitivity is low compared with other techniques, for example, atomic absorption spectroscopy. In consequence, EPXMA cannot be used for detecting trace amounts of elements evenly dispersed throughout a matrix.

EELS analysis is inherently more difficult to perform than EPXMA, and often EPXMA may be carried out initially for the purpose of localization of a specific element before EELS analysis is carried out. EELS is generally carried out with a field emission source so that probe diameters of the order of 1 nm can be achieved, making this a very high-resolution method of analysis. Due to the higher collection efficiency, EELS is also more sensitive than EPXMA. Recently it has been shown that it is possible to detect single atoms of iron and calcium in isolated biological macromolecules. As stated previously, the two techniques are complementary. The elements Na and K, which are important in many biological systems, are poorly detected in EELS but easily detected

with EPXMA, whereas the superior sensitivity of EELS makes it ideal for the detection of calcium, which is poorly detected in EPXMA because of overlap of the Ca  $K_{\alpha}$  by the K  $K_{\beta}$  peak.

## Factors That Need to Be Taken into Account when Interpreting Microanalytical Data

### Specimen Preparation

Both EPXMA and EELS are used to measure the elemental content of a specimen as it exists under the beam in an electron microscope. Therefore, great care must be taken to ensure that the methods used for specimen preparation neither remove elements from nor add elements to the specimen. The exact method to be used for specimen preparation will depend on the form of the specimen and the objective of the study. Brief outlines of some of the

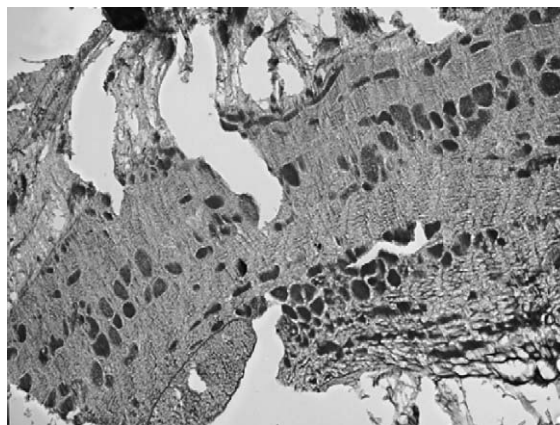
methods that are used for specimen preparation are given below.

Conventional methods used for the preparation of biological material both for light microscopy and EM consist of fixation, dehydration, and embedding. These procedures result in increased permeability of the cell membrane with resultant loss or redistribution of mobile intracellular elements. Sections are then stained, resulting in the deposition of elements onto the specimen so that the resultant spectrum is not representative of the elemental composition *in vivo* (Figure 2). In consequence, such preparation methods are rarely used when the specimen is intended from the outset for microanalysis.

For the study of the diffusible elements the specimen needs to be cryoprepared to arrest the physiological processes and to prevent redistribution of the elements by diffusion during the subsequent steps before analysis. Once the specimen has been

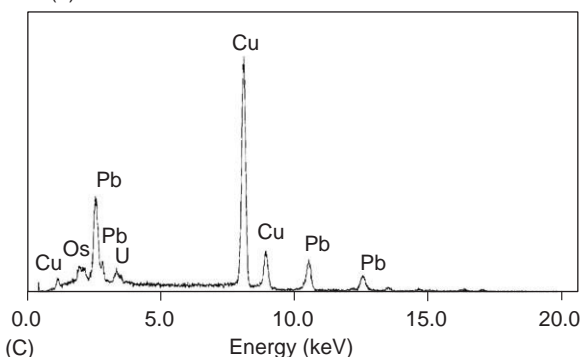


(A)



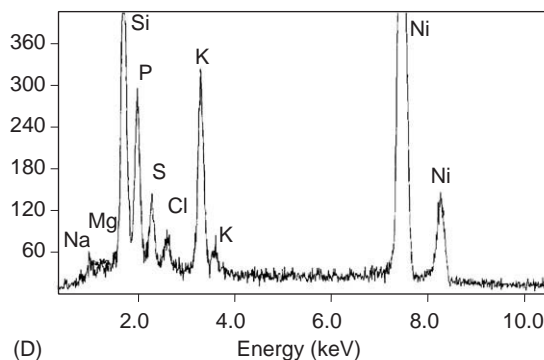
(B)

ID(1):  
ID(2): Fixed heart tissue



(C)

ID(1): Cryofixed heart tissue



(D)

**Figure 2** The effects of specimen preparation on structure and element content of biological tissues. (A) Conventionally prepared section of heart muscle. Individual cells, their nuclei, the mitochondria, the Z-lines, and the myofilaments are clearly seen. (B) A heart muscle cell prepared by cryofixation and cryosectioning. Although the cellular structures are the same as in (A), the details are not as distinct. (C) A spectrum derived from material as in (A). The elements that can be identified are Cu from the grid and Os, Pb, and U from the stain. The only element that can be detected from the specimen itself is S. (D) A spectrum derived from cryoprepared heart tissue. Here the major inorganic elements – Na, Mg, P, S, Cl, and K – are detected.

cryofixed, the low temperatures need to be maintained to prevent ice-crystal growth, which may cause redistribution of ions. The specimen can be maintained in a frozen-hydrated state and analyzed as a bulk specimen in a SEM, provided that a suitable cold stage and an efficient cryotransfer device are available. Alternatively, the frozen-hydrated specimen can be sectioned at a very low temperature (below  $-140^{\circ}\text{C}$ ) and analyzed as a section in the frozen-hydrated state.

### Treatment before Fixation

When investigating diffusible ions it is necessary to consider that treatment of the specimen before cryofixation may affect the results. For example, dicing tissue into small pieces before cryofixation is not suitable for highly metabolically active tissues such as the heart. It has been shown that in this tissue redistribution of ions can occur in times as short as 10 s. Similarly, in other tissues any delay between harvesting and fixation gives the possibility of element redistribution. In addition it has been shown that treatments such as the use of local anesthetics when taking skin biopsy samples can also affect elemental content.

### Specimen Stability

Most elements are stable under the conditions used for analysis; however, some problems have been encountered in the analysis of halides that can be labile if organic matter is absent from the specimen. The loss of sulfur from organic matrices has also been reported.

Under the conditions used for EDS analysis, loss of mass from the specimen matrix inevitably occurs, and this is a special problem with frozen-hydrated material. The loss from thin (100–200 nm) frozen-hydrated sections is so severe that they may be etched away during an analytical run. The loss of specimen mass also has implications for quantitative studies. The Hall method, which measures the mass of a given element per unit mass of specimen, is most frequently used for the purposes of quantification, correction factors being obtained by comparison with suitable standards. Standards that closely resemble the specimen in the composition of the organic matrix need to be used to avoid differences in mass loss between the standard and the specimen, producing errors in quantification. Mass loss can be reduced through the use of a cold stage and by lowering the beam current used for analysis.

### Peak Overlap

When an EDS detector is used, the characteristic X-ray emissions form a Gaussian peak rather than a

**Table 1** Some elements encountered in EPXMA of biological specimens that emit overlapping peaks

<i>Element</i>	<i>Interfering line</i>	<i>Line interfered with</i>	<i>Origin</i>
Pb	M	S $K_{\alpha}$ , Cl $K_{\alpha}$	Stain
U	M	K $K_{\alpha}$	Stain
Os	M	P $K_{\alpha}$ , S $K_{\alpha}$ , Cl $K_{\alpha}$	Fixative
As	L	Na $K_{\alpha}$ , Mg $K_{\alpha}$	Buffer
Ag	L	Cl $K_{\alpha}$ , K $K_{\alpha}$	Precipitating agent
Au	M	P $K_{\alpha}$ , S $K_{\alpha}$	Grid
Cu	L	Na $K_{\alpha}$	Grid
K	$K_{\beta}$	Ca $K_{\alpha}$	Naturally occurring

discrete line in the spectrum (Figure 1). In addition, the spectrum from a sample containing heavy elements may contain a large number of peaks due to L and M emissions. In consequence, some characteristic peaks may be partly or completely overlapped by peaks from a second element (Table 1) making both the identification of the peak and quantification of the number of counts under the peak difficult. The fitting routines in modern EDS analytical software are, however, usually capable of separating such overlaps. It is important that the spectrometer be calibrated on a regular basis to eliminate the effects of spectral drift, which can introduce errors in quantification. Specimen preparation methods that eliminate the addition of contaminating peaks should be adopted.

### Extraneous Radiation

In EDS the production of extraneous radiation is a problem, especially in TEM. During the irradiation of the specimen, electrons are scattered, and both characteristic and continuum X-rays are produced from materials in the surrounding area, such as the grid, the specimen support, and the lens pole pieces. Such extraneous contributions may cause errors by masking elements present in the specimen and by production of excess continuum, which causes problems in quantification.

Production of extraneous radiation from the column can be minimized by ensuring that the instrument is well aligned. Thick apertures should be used to prevent X-rays generated higher in the column from reaching the specimen, and the objective aperture should be removed during analysis to eliminate X-rays from this source. Several different types of grid should be used when pathology specimens are being examined. In quantification X-rays produced by the grid and specimen support can be corrected for, provided they are not excessive and there is a characteristic peak from which the continuum contribution can be determined.

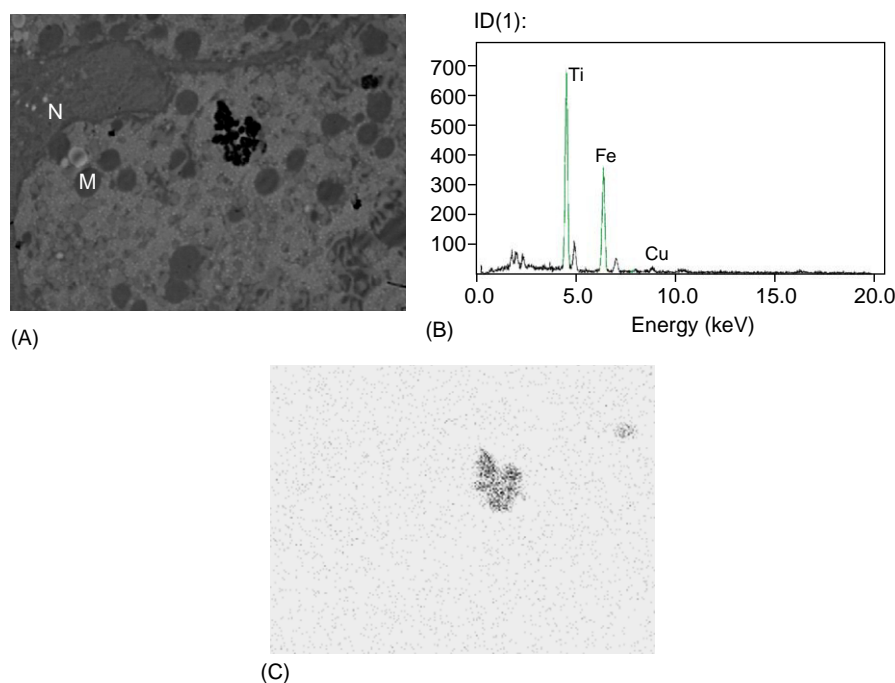
## Applications

### Applications in Pathology and Medicine

One of the major applications of EPXMA is as an aid to pathological studies. Frequently, unidentified objects are found in specimens that have been subjected to routine histopathological examination. Determination of the elemental composition of such objects may be required for several reasons. The elements may be endogenous or may come from exogenous sources. Endogenous sources include errors of metabolism that lead to large accumulations of elements such as copper or iron, which are normally only found in trace quantities. The accumulations are often found in lysosomes of the liver cells (**Figure 3**). Identification of the element involved can aid in diagnosis. However, accumulated particles more usually result from external sources, and the identification of their elemental content may be the first step in an investigation that could have legal, public health, or medical implications. Diseases of the lung (pneumoconiosis) can occur through exposure to metals or other substances in the workplace, for example, asbestosis or silicosis. EM is used both as a means of quantifying the mineral burden in the lung and of identifying the causative agent. Granulomas can occur as a consequence of the application of cosmetics, for example, titanium and zirconium in

deodorants, or even as a result of medication: some ointments contain metals such as zinc. Titanium is also found as a food additive and as a whitening agent in toothpaste; EPXMA studies have linked this element to inflammation in the gut. Drugs taken internally may result in depositions both at and away from the expected site of action, for example, gold-containing compounds used in the treatment of arthritis, platinum-containing drugs used in cancer chemotherapy, and iodine found in contrast media. EPXMA has also been used to investigate reactions of the body to prosthetic implants such as the localization of silicon in the periprosthetic capsule surrounding breast implants. One area of great interest is the interface between joint prostheses and the body. EPXMA has been used to determine the composition of wear particles found in tissue near the prostheses and also to provide evidence that corrosion of the implant may occur within the body. Some of the applications of EPXMA in pathological studies are shown in **Table 2**.

The previous paragraph might suggest that micro-analytical techniques are only used in medical applications when complications occur. However, increasingly these techniques are being used beneficially as research tools. ESI has been used to track the effect of phlebotomy on iron deposits in hemochromatosis patients. Another area in which much



**Figure 3** Identification of the elemental content of an inclusion in a pathological specimen. (A) An unstained transmission electron microscope image of a thin section from a liver biopsy: N, nucleus; M, mitochondria. The lysosomes contain electron-dense deposits. (B) An X-ray spectrum produced by focusing the electron beam onto the specimen shows the deposits contain iron. (C) A map of iron distribution in the specimen shows that the element is confined to areas containing the deposit.



**Table 2** Some applications of EPXMA in pathology using conventionally prepared specimens

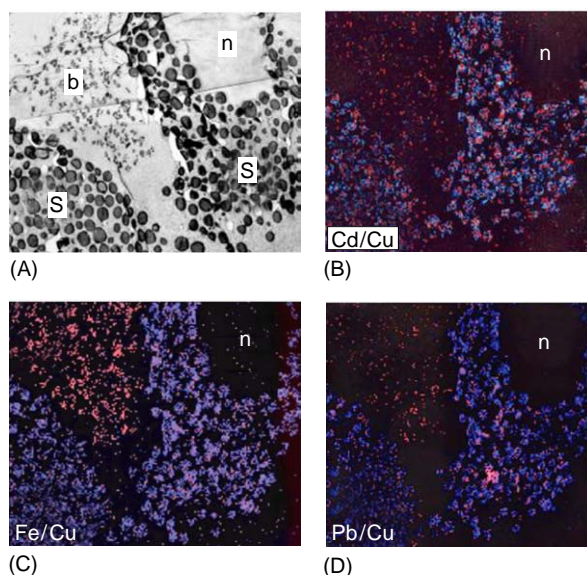
Specimen	Medical problem	Element
Nail clippings	Cystic fibrosis	Na, Cl
Hair, skin	Allergy/dermatology	Cu, Fe, Si, Ni
Liver biopsy	Hemachromatosis	Fe
Liver biopsy	Wilson's disease	Cu
Liver biopsy	Hepatitis/cirrhosis	Fe, Cu
Liver biopsy	$\beta$ -Thalassemia	Fe
Bone, cartilage	Osteoarthritis	Cu, P, Au (drug treatment)
Bone, cartilage	Failed implant	Co, Ti
Intima	Atherosclerosis	Ca, P
Lung biopsy	Pneumoconioses	Asbestos, Si, Pb, etc.
Breast biopsy	Implant	Si
Urine	Renal stones	Ca, P

effort is being devoted is in the development of new biomaterials, with the objective of producing implants with better properties of biocompatibility.

### Pollution Studies

Human activity has begun to affect the environment in which we live, resulting in contamination that affects us, as well as other organisms. Understanding the deleterious effects of pollution requires the ability not only to identify the toxic element itself but also to localize it to its site of action and to correlate the presence of the pollutant with structural and functional changes within cells, tissues, and organelles. From its beginning, EPXMA has been used widely in pollution studies. The more traditional role has been in the study of heavy metal uptake by invertebrates, which, like plants, are the first part of the food chain exposed to the contaminated environment. EPXMA has been and continues to be a valuable tool in elucidating the pathways by which these animals are able to avoid the toxic effects of metals in their diet (Figure 4 and Table 3).

More recently there has been an increased emphasis on air pollution since episodes of pollution have been correlated with increased morbidity and mortality in susceptible groups. Here EM coupled with image analysis and microanalytical techniques gives information at different levels in the pollution pathway (Table 4). The morphology and elemental composition of particles collected from ambient air can be identified and the number and composition of particles present in bronchial lavage fluids and in alveolar macrophages determined. The technique of EELS with its ability to differentiate between oxidative states of metals such as iron, and thus assist in determination of their hazard potential, has great potential in this area.



**Figure 4** (A) Transmission electron micrograph of a region of an ultrathin, freeze-dried cryosection of a propane-plunged hepatopancreas from a terrestrial isopod, *Oniscus asellus*, collected from the vicinity of an abandoned metalliferous mine where the soil was heavily contaminated with Cd, Pb, and Zn. Note the two S-cells (s), containing characteristic, electron-opaque, Cu-rich granules; one cell profile displays a prominent nucleus (n). The field also contains part of a B-cell (b) with floccular, Fe-rich granules. (B)–(D) Pairs of superimposed X-ray distribution speed-maps (i.e., qualitative maps) obtained in a JEOL JEM-1210 TEM equipped with a Link ATW Pentafet detector (138 eV resolution) and Link ISIS analyzer (Oxford Instruments) operated under the following conditions: accelerating voltage, 120 kV; filament current, 10  $\mu$ A; condenser aperture, 200  $\mu$ m; spot size '1' in fine probe mode, producing a convergent beam  $\sim$ 100 nm in diameter scanned by the ISIS system; objective aperture withdrawn; total map acquisition times, 30–60 min. The maps are slightly rotated relative to the imaged area. Note that the three maps (B)–(D) confirm that the S-cell granules are Cu containing and that these granules are capable of sequestering the nonessential metals Cd (B) and Pb (D); the B-cell granules are predominantly Fe containing (C) and, not surprisingly, do also sequester the 'borderline' metal Pb (C) but possibly not the 'Group B' Cd (B), a metal whose affinity for sulfur-donating ligands is much stronger than its affinity for oxygen-donating ligands. (This figure was provided by Dr. AJ Morgan and Ms C Winters of the Department of Applied Biology, University College of Wales, Cardiff.)

### Physiologically Important Ions

The diffusible ions of chemical elements play a central role in cell function. Controlled movement of these ions across cell membranes and sequestration at different sites within the cells are known to underlie many different events, such as the transmission of nerve impulses, the contraction of muscle, and the initiation of cell division. In consequence the study of the cellular and intracellular distributions of these elements is of great interest in physiology. Although microanalytical techniques have traditionally been used to study isolated, physiologically active



**Table 3** Some applications of EPXMA in the study of pollution

<i>Specimen studied</i>	<i>Elements studied</i>	<i>Reference</i>
Woodlouse	Cu, Cd, Fe, Cd, Pb	Hopkin <i>et al.</i> (1989) <i>Microscopy &amp; Analysis</i> 14: 23–27
Earthworm	Ca, Cd, Cu, P, Pb, Zn	Morgan <i>et al.</i> (1994) <i>Scanning Microscopy</i> S8: 231–243
Freshwater snail	Al, Cd, Si, Zn	Desouky <i>et al.</i> (2002) <i>Proceedings of the National Academy of Sciences of the USA</i> 99: 3394–3399
Bronchiolar lavage	Si, S + metals	Falchi <i>et al.</i> (1996) <i>Archives of Environmental Health</i> 51: 157–161
Diesel exhaust	Cr, Fe, Mn, Zn + others	Berube <i>et al.</i> (1999) <i>Atmosphere and Environment</i> 33: 1599–1614
Alveolar macrophages	C	Bunn <i>et al.</i> (2001) <i>Thorax</i> 56: 932–934
Urban aerosol samples	C, O, Zn, Al, Pb	Wawros <i>et al.</i> (2003) <i>Microanalysis</i> 9: 349–358

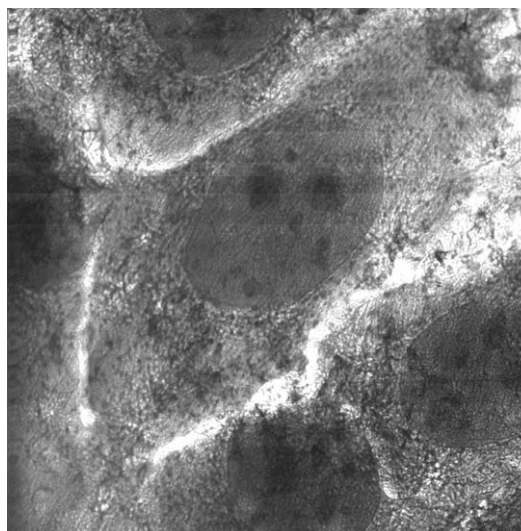
**Table 4** Some applications of microanalytical techniques using cells in culture

<i>Problem investigated</i>	<i>Technique used</i>	<i>Reference</i>
Measurement of [Ca]	EELS	Leapman <i>et al.</i> (1994) <i>Scanning Electron Microscopy Supplement</i> 8: 245–259
Detection of Pt-containing drug	EPXMA	Kirk <i>et al.</i> (1996) <i>Journal of Microscopy</i> 183: 181–186
Apoptosis in monocytes	EPXMA	Fernandez-Segura <i>et al.</i> (1999) <i>Experimental Cell Research</i> 253: 454–462
Ion transport in sweat gland cells	EPXMA	Hongpaisan (1999) <i>Acta Physiol. Scand.</i> 165: 241–250
Effect of cAMP on ion transport	EPXMA	Zhang (1999) <i>Acta Physiol. Scand.</i> 165: 95–101
Apoptosis in macrophages	EPXMA	Skepper <i>et al.</i> (1999) <i>Journal of Pathology</i> 188: 106
Cl secretion by nasal epithelium	EPXMA	Dragomir <i>et al.</i> (2001) <i>Journal of Microscopy</i> 203: 277–284
Boron in neutron capture therapy	EELS	Michel <i>et al.</i> (2003) <i>Journal of Microscopy</i> 210: 25–34

preparations, there has been increasing use of *in vivo* freezing techniques and use of cells in culture.

The point analysis mode is used specifically to determine the concentration of the diffusible elements at the cellular level in complex tissues, in cells in culture, and at the subcellular level in individual cells. Under these circumstances information is provided that is not available from other techniques and that complements the information on ionic activities commonly obtained using ion-selective electrodes or fluorescent dyes. In addition, mapping techniques allow the investigation of 2D spatial distributions of elements at the nanometer level in intracellular compartments, yielding information that again cannot be provided by other techniques.

Cells in culture represent a simple system for study in which experimental conditions are easily controlled. An additional advantage is that some of the problems that are encountered in the preparation of whole tissues for microanalysis can be circumvented. Cells grown in suspension can be prepared quite simply by centrifugation to form a pellet that can then be cryofixed and cryosectioned before analysis using a TEM. Such cells can also be collected on a filter for analysis in a SEM. Cells that normally grow as adherent monolayers can be grown on plastic film-coated gold or titanium EM grids. The grids are then removed from culture and washed rapidly to remove the overlying medium, and the cells freeze-dried and analyzed as whole mounts either in a SEM or a TEM (Figure 5 and Table 4).



**Figure 5** A scanning transmission electron micrograph of osteoblast cells grown on Pioloform-covered gold EM grids. Details of the cell components such as the large ovoid nucleus and mitochondria can be seen. Although the substrate is unusual, it does not affect the growth characteristics of the cells, which readily form a monolayer.

In physiology EPXMA is used to study changes in element distribution in normal tissues in response to different types of stimulation and is also used to investigate the effects of disease, an area sometimes referred to as pathophysiology. There is currently much interest in determining the roles of the mitochondria and the endoplasmic reticulum as sources

**Table 5** Some applications of microanalytical techniques in physiology

Authors	Reference	Comments
Wendt-Gallitelli <i>et al.</i> (1993)	<i>Journal of Physiology</i> 472: 33–44	Isolated cardiac myocytes, subcellular gradients
Wheeler-Clark and Buja (1995)	<i>Journal of Pharmaceutical and Experimental Therapy</i> 274: 1493–1506	Coronary smooth muscle, Ca content gradient
Moravec <i>et al.</i> (1997)	<i>American Journal of Physiology</i> 273: H1432–H1439	Isolated perfused heart, mitochondria
Pozzo-Miller <i>et al.</i> (1997)	<i>Journal of Neuroscience</i> 15: 8792–8838	Brain slices, Ca sequestration in endoplasmic reticulum
Horikawa <i>et al.</i> (1998)	<i>Biophysical Journal</i> 74: 1579–1590	Papillary muscle, Ca content mitochondria
Pipovarova <i>et al.</i> (1999)	<i>Journal of Neuroscience</i> 19: 6372–6384	Isolated ganglia and cells, Ca gradients
Thirion <i>et al.</i> (1999)	<i>Proceedings of the National Academy of Science of the USA</i> 96: 3206–3210	Isolated nerve terminals, Na in vesicles
Hongpaisan <i>et al.</i> (2001)	<i>Journal of General Physiology</i> 118: 101–112	Neurons, Ca uptake and spatial distribution
Reeves <i>et al.</i> (2002)	<i>Nature</i> 416: 291–297	Neutrophil activation by K flux
Silverman <i>et al.</i> (2003)	<i>Cardiovascular Research</i> 57: 1025–1034	Cardiac myocytes, subcellular gradients
Akar <i>et al.</i> (2003)	<i>Circulation</i> 107: 1810–1815	Atrial fibrillation, Cl accumulation

and sinks of the signaling molecule calcium and the interplay between these two organelles under different conditions of stimulation. Undoubtedly a significant contribution of this technique to the understanding of cell physiology has been the demonstration that nonuniform distribution of different elements within the cell can occur. Gradients of sodium concentration have been shown to exist within the first 100 nm of the sarcolemma of heart cells under conditions of stimulation. Similarly it has been shown that mitochondria close to the plasma membrane of neuronal cells accumulate calcium under stimulatory conditions. Some applications of microanalytical techniques in physiological studies are shown in Table 5.

### Botanical Applications

There is now an increase in the number of studies that use EPXMA for the study of plant cell physiology. The main problem with plant material is that large ice-crystals tend to form in the aqueous vacuoles of plant tissue on cryofixation, making cryosectioning difficult, so that the preparation route of cryofixation, cryosectioning, and freeze-drying, which is favored for the preparation of animal cells/tissues, is generally not used. Instead a variety of preparation procedures have been adopted for the study of plant materials. In the botanical field there is considerable interest in determining the composition of sap in relation to transpiration rates in plants, and it has been shown that this can be accomplished successfully by use of planed, bulk frozen-hydrated plant tissue or by using micropuncture techniques followed by analysis of the fluid droplets. Another major area of research is the adaptation of plant cells to their environment. Plants often colonize harsh

environments such as spoil heaps from mines and can also become adapted to increased salinity. EPXMA is an invaluable tool for investigating the mechanisms of tolerance of toxic materials and has been used to show that elements such as lead are sequestered at the cell wall and within vacuoles but not in the cytoplasm, microsomes, nucleus, or nucleoli.

*See also:* **Forensic Sciences:** Gunshot Residues; Hair. **Microscopy Applications:** Environmental. **Microscopy Techniques:** Electron Microscopy; Specimen Preparation for Electron Microscopy. **X-Ray Fluorescence and Emission:** Energy Dispersive X-Ray Fluorescence.

### Further Reading

- Brydson R (2001) *Electron Energy Loss Spectroscopy*, Royal Microscopical Society Handbook 48. Oxford: Bios.
- Hall TA and Gupta BL (1986) EDS quantitation and application to biology. In: Joy DC, Romig AD, and Goldstein JI (eds.) *Principles of Analytical Electron Microscopy*, pp. 219–248. New York: Plenum.
- Ingram P, Shelburne J, Roggli V, and LeFurgey A (1999) *Biomedical Applications of Microprobe Analysis*. San Diego: Academic Press.
- Morgan AJ (1985) *X-ray Microanalysis in Electron Microscopy for Biologists*, Royal Microscopical Society Handbook 05. Oxford: Oxford University Press.
- Roomans GM and von Euler A (1996) X-ray microanalysis in cell biology and cell pathology (Review). *Cell Biology International* 20: 103–109.
- Roomans GM, Gupta BL, Leapman RD, and von Zglinicki (eds.) (1994) The science of biological microanalysis. *Scanning Microscopy* supplement 8. Chicago: Scanning Microscopy International.
- Sigee DS, Sumner AT, Morgan AJ, and Warley A (1993) *X-ray Microanalysis in Biology: Experimental Techniques*

and Applications. Cambridge: Cambridge University Press.

Steinbrecht RA and Zierold K (1987) *Cryotechniques in Biological Electron Microscopy*. Berlin: Springer.

Warley A (1997) X-ray Microanalysis for Biologists, vol. 16, *Practical Methods in Electron Microscopy* (series editor Audrey Glauert). London: Portland Press.

## Food

R M Twyman, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by M Kaláb and S S Miller, pp. 3210–3218, © 1995, Elsevier Ltd.

### Introduction

Most foods are of biological origin, and processing changes their nature to varying degrees. Transitions such as grain → flour → bread, muscle → meat → salami, and milk → curd → cheese may be used as examples of foods that undergo these changes. All foods may be categorized according to specific structural characteristics. These include liquid foods that are in the form of suspensions and emulsions (beverages such as milk, fruit and vegetable juices, dressings, toppings, etc.), foamed foods (whipped cream, marshmallows), viscous foods with a low fat content (yogurts, jams, puddings) or a high fat content (dairy spreads, margarine, peanut butter), solid foods (meat, fish, bread), and powders (flour, milk powders, spices). Frequently, finished foods are combinations of these various groups. The approach used in microscopic analysis depends upon the nature of the food and the objective of the examination. The history of food microscopy, and the principles of microscopic analysis of different food groups were reviewed comprehensively in a book by Aguilera and Stanley in 1990 (see 'Further Reading' section). The microscopic analysis of different food commodities has been reviewed in books published in 1979 edited by Vaughan and in 1994 edited by Flint ('Further reading' section). From 1982 to 1994, the journal *Food Structure* (formerly *Food Microstructure*) specialized in the publication of papers on food microscopy and related techniques. Since 1994 there has been no journal dedicated specifically to the microscopic analysis of food, although the review journal *Trends in Food Science and Technology* carries regular articles on this subject.

## Microscopy Techniques

### Light Microscopy

Light microscopy (LM) is regularly used to obtain rapid, inexpensive qualitative and quantitative information in food analysis. The first routine use of LM in food analysis was for the identification of adulteration (e.g., the presence of chicory root in coffee) or contamination (insect, rodent, microbial, and foreign bodies). Bright-field, polarizing, and fluorescent microscopy are the three traditional LM techniques used most frequently in food analysis. The basic instrument is a conventional compound (bright-field) microscope, to which polarizing and fluorescence accessories are easily attached.

A wide variety of stains can be used in combination with bright-field LM. In food analysis, however, only a few of those stains are used on a regular basis. For example, Toluidine Blue O (TBO) is a metachromatic dye that produces different colors depending on the nature of the component to which it is bound. Pectin-containing plant cell walls stain a rich pink or purple with TBO, while lignified cells of vascular tissues stain dark blue (see **Figure 1A**). TBO is also useful for the examination of meat products, where muscle fibers show up as pale pink, fibroblasts as blue, and elastin fibers as turquoise. The familiar blue coloration produced by staining starch with iodine is widely used, allowing the identification and localization of starch even after processing has destroyed the characteristic granule structure. For fat staining, a lipid-soluble dye such as Oil Red O may be used.

Polarizing microscopy is used to examine food components that exhibit birefringence (an ordered crystalline structure). Many food components are birefringent, e.g., starch, plant cell walls, specialized 'stone cells' in some plant tissues, muscle fibers, fats from both plant and animal sources, and different types of flavor and seasoning components.

The botanical origin of most starches is readily identified on the basis of granule size and shape, the form and position of the hilum, and the

and Applications. Cambridge: Cambridge University Press.

Steinbrecht RA and Zierold K (1987) *Cryotechniques in Biological Electron Microscopy*. Berlin: Springer.

Warley A (1997) X-ray Microanalysis for Biologists, vol. 16, *Practical Methods in Electron Microscopy* (series editor Audrey Glauert). London: Portland Press.

## Food

R M Twyman, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by M Kaláb and S S Miller, pp. 3210–3218, © 1995, Elsevier Ltd.

### Introduction

Most foods are of biological origin, and processing changes their nature to varying degrees. Transitions such as grain → flour → bread, muscle → meat → salami, and milk → curd → cheese may be used as examples of foods that undergo these changes. All foods may be categorized according to specific structural characteristics. These include liquid foods that are in the form of suspensions and emulsions (beverages such as milk, fruit and vegetable juices, dressings, toppings, etc.), foamed foods (whipped cream, marshmallows), viscous foods with a low fat content (yogurts, jams, puddings) or a high fat content (dairy spreads, margarine, peanut butter), solid foods (meat, fish, bread), and powders (flour, milk powders, spices). Frequently, finished foods are combinations of these various groups. The approach used in microscopic analysis depends upon the nature of the food and the objective of the examination. The history of food microscopy, and the principles of microscopic analysis of different food groups were reviewed comprehensively in a book by Aguilera and Stanley in 1990 (see 'Further Reading' section). The microscopic analysis of different food commodities has been reviewed in books published in 1979 edited by Vaughan and in 1994 edited by Flint ('Further reading' section). From 1982 to 1994, the journal *Food Structure* (formerly *Food Microstructure*) specialized in the publication of papers on food microscopy and related techniques. Since 1994 there has been no journal dedicated specifically to the microscopic analysis of food, although the review journal *Trends in Food Science and Technology* carries regular articles on this subject.

### Microscopy Techniques

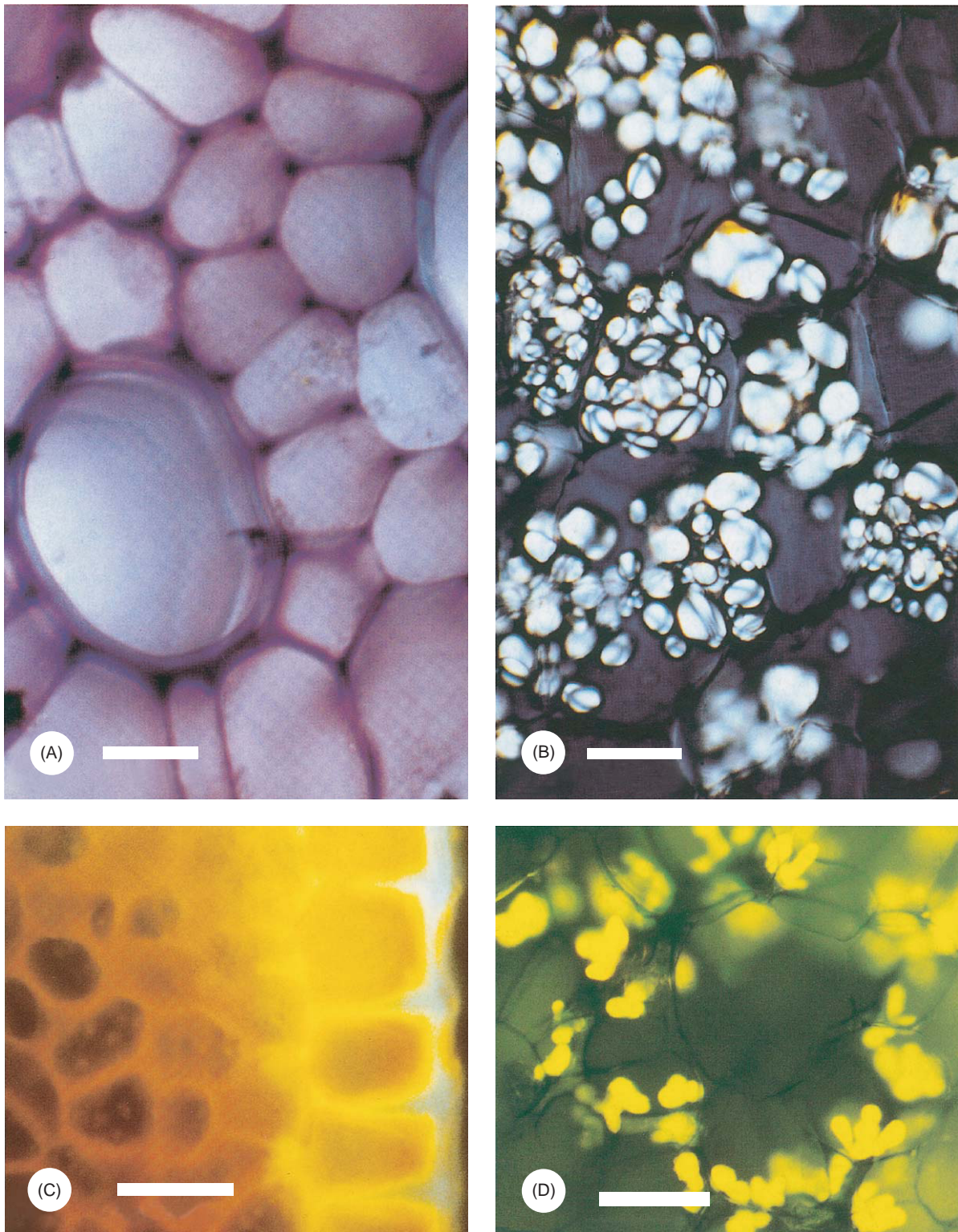
#### Light Microscopy

Light microscopy (LM) is regularly used to obtain rapid, inexpensive qualitative and quantitative information in food analysis. The first routine use of LM in food analysis was for the identification of adulteration (e.g., the presence of chicory root in coffee) or contamination (insect, rodent, microbial, and foreign bodies). Bright-field, polarizing, and fluorescent microscopy are the three traditional LM techniques used most frequently in food analysis. The basic instrument is a conventional compound (bright-field) microscope, to which polarizing and fluorescence accessories are easily attached.

A wide variety of stains can be used in combination with bright-field LM. In food analysis, however, only a few of those stains are used on a regular basis. For example, Toluidine Blue O (TBO) is a metachromatic dye that produces different colors depending on the nature of the component to which it is bound. Pectin-containing plant cell walls stain a rich pink or purple with TBO, while lignified cells of vascular tissues stain dark blue (see **Figure 1A**). TBO is also useful for the examination of meat products, where muscle fibers show up as pale pink, fibroblasts as blue, and elastin fibers as turquoise. The familiar blue coloration produced by staining starch with iodine is widely used, allowing the identification and localization of starch even after processing has destroyed the characteristic granule structure. For fat staining, a lipid-soluble dye such as Oil Red O may be used.

Polarizing microscopy is used to examine food components that exhibit birefringence (an ordered crystalline structure). Many food components are birefringent, e.g., starch, plant cell walls, specialized 'stone cells' in some plant tissues, muscle fibers, fats from both plant and animal sources, and different types of flavor and seasoning components.

The botanical origin of most starches is readily identified on the basis of granule size and shape, the form and position of the hilum, and the



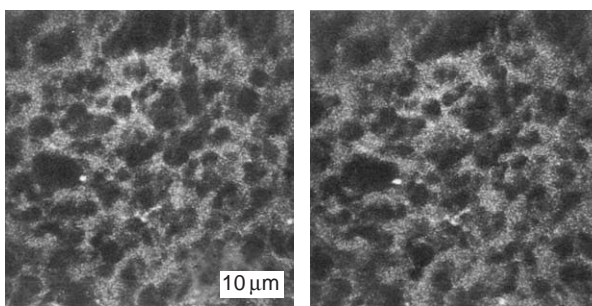
**Figure 1** Light microscopy of some foods of plant origin. (A) Bright-field micrographs of hand-cut section of fresh pumpkin stained with Toluidine Blue. Parenchyma cell walls, which are rich in pectin, are purple, and cell walls of vascular tissue, which are rich in phenolics, are blue. (B) Hand-cut section of fresh potato, viewed using polarizing optics. Birefringent starch granules appear bright against a dark background. (C) Fluorescence micrograph of hand-cut section of pumpkin epidermis, stained with Coriphosphine O. Bars in (A) to (C) represent 25  $\mu\text{m}$ . (D) Fluorescence micrograph showing hand-cut section of infected radish, stained with Acridine Orange (yellow fungal cells) and Methyl Green (green radish cell walls). Bar represents 100  $\mu\text{m}$ . (Miller SS, Agriculture and Agri-food Canada, Ontario, Canada.)



brilliance of the interference cross under polarized light (see **Figure 1B**). Loss of birefringence upon gelatinization, the temperature of which is characteristic for various native and derivatized starches, is determined using a polarizing microscope equipped with a heat stage. The effects of baking and processing on starches are monitored by a combination of polarizing and bright-field microscopy using iodine staining to identify starch after loss of birefringence.

Fluorescence provides a sensitivity that is not available in other forms of LM, allowing the detection of fluorescing compounds present in amounts as little as  $10^{-18}$  mol. A wide range of food components of both plant and animal origin exhibit natural fluorescence (autofluorescence). In plants, these components include pigments such as chlorophylls and carotenoids, as well as various phenolic compounds such as lignins and low-molecular-mass compounds such as ferulic and chlorogenic acids. Many flavor compounds in herbs and spices are also autofluorescent. In animal tissues, the most common sources of autofluorescence are bone, cartilage, collagen, elastin, and some lipids. In addition, fluorescence may be induced by a wide range of compounds such as fluorescent dyes, specific antibodies or lectins that are conjugated to fluorescent markers, and substances that fluoresce only in specific chemical environments. For example, Calcofluor, a fluorescent brightener, can be used as a highly specific probe to localize the mixed linkage (1 $\rightarrow$ 3), (1 $\rightarrow$ 4)- $\beta$ -D-glucan in cereal grains. Nile red is a lipid-soluble dye that becomes intensely fluorescent in the hydrophobic medium of fat droplets. The precise identification and localization of various components permitted by such compounds, coupled with the sensitivity of the fluorescence technique, has made fluorescence microscopy a valuable tool in food analysis.

Confocal microscopy has also been developed as a technique with some advantages for food analysis. The major difference between a confocal and a conventional microscope is the placement of a pin-hole at the focal plane of the image in the case of the confocal instrument. This removes out-of-focus light, generating a clearer image and allowing optical sectioning of the specimen. There are two basic types of confocal microscopy: confocal scanning tandem microscopy (CSTM) and confocal scanning laser microscopy (CSLM). CSTM uses mercury, tungsten, or xenon illuminators and has the advantage of allowing real-time observation of the specimen. Low light intensity can be a problem with this technique. CSLM uses laser illumination. Standard equipment includes an argon laser (488 and 514 nm wavelengths) with or without a



**Figure 2** Confocal laser scanning microscopy of Gouda cheese stained for protein with 1-anilino-8-naphthalene sulfonic acid. The depth resolution of optical sectioning was  $\sim 0.7 \mu\text{m}$ . (Courtesy of I. Heertje and Scanning Microscopy International.)

helium–neon laser (633 nm wavelength). In many of the CSLM systems currently in use, real-time observation is not available because of potential damage to the eye by laser emissions. Images are produced, stored, and manipulated by image-handling software. Light intensity is not a problem with this method.

CSLM can provide focused images to a depth of up to several hundred micrometers, depending on the nature of the sample, so that sequential sections may be obtained for three-dimensional reconstruction of the image (**Figure 2**). In addition, several chemical components (e.g., protein and fat in cheese) can be identified and localized simultaneously using specific fluorescent labels. CSLM has been used for the quantitative analysis of cellular structures in plant material, the structural analysis of emulsions of different complexities, and the location of microorganisms in a wide range of food products.

### Electron Microscopy

Electron microscopy (EM) is used where a higher resolution than that obtainable by LM techniques is required, for example, for the analysis of components such as casein micelles in milk or pectin fibers in fruits and vegetables, which are too small to be visualized by LM. Scanning electron microscopy (SEM) in conjunction with X-ray microanalysis, and transmission electron microscopy (TEM) in the electron energy loss spectrometry (EELS) mode, enable the analysis of foods or their components for elemental composition. SEM of frozen hydrated samples and TEM of replicas obtained from samples fixed by rapid freezing provide images of foods unaffected by chemical fixation. In contrast, chemical fixation of certain components such as proteins (using aldehydes – glutaraldehyde, formaldehyde – and  $\text{OsO}_4$ ), unsaturated fats (using  $\text{OsO}_4$  in the

**Table 1** Overview of microscopical techniques and applications in food analysis

<i>Technique</i>	<i>Mode</i>	<i>Information obtained</i>
Dissecting microscopy	Bright-field, polarizing, and fluorescence	Contamination
Light microscopy		Chemical and structural analysis
Confocal microscopy	Confocal scanning tandem microscopy (CSTM) and confocal scanning laser microscopy (CSLM)	Contamination
		Component identification
Scanning electron microscopy (SEM)	Conventional SEM	Component distribution
		Changes during processing
Energy-dispersive spectrometry (X-ray microanalysis)	Conventional SEM and cryo-SEM	3D reconstruction of low-density foods
		Structural information for food powders, solid foods, and viscous foods
Transmission electron microscopy (TEM)	Cryo-SEM	Structural information for high-fat foods and edible foams
		Chemical information
Electron energy loss spectrometry (EELS)	Conventional SEM and cryo-SEM	Chemical and structural information for liquid foods and emulsions
		Chemical and structural information for all foods
Atomic force microscopy	Thin sectioning	Chemical and structural information for liquid foods, high-fat foods
		Crystallinity
Atomic force microscopy	Freeze-fracture plus replication	Chemical information
		Very high resolution of surface structure

presence of imidazole) and some polysaccharides (using  $\text{RuO}_4$ ) followed by embedding in a resin and staining of thin sections with heavy metals such as uranyl acetate and/or lead citrate, makes it possible to distinguish these components in foods.

### Atomic Force Microscopy

Significant structural changes induced in food ingredients by processing for EM require the development of novel methods, such as environmental stages and metal-free processing, to achieve proper results. Indeed, EM techniques are pushed to their limits when they are used to image complex biopolymers such as the irregular fibrous proteins and polysaccharides that are abundant in foods. Since these types of samples have to be coated with metals, the size of the metal grains restricts the level of detail observable in the final images. Atomic force microscopy (AFM) creates an image by scanning a sharp stylus, which is attached to a flexible cantilever, across the sample surface. When the stylus is held very close to the sample, repulsive forces deflect the cantilever away from the surface. As the cantilever–stylus assembly is scanned over the surface, topological features are translated into movements, which can be recorded. This simple and highly sensitive technique can measure deflections caused even by individual molecules and atoms. It

has been widely used to study complex food components, such as protein fibrils, polysaccharides, and interfaces. Information concerning the different types of microscopy and their applications is presented in Table 1.

## Sample Preparation

### Light Microscopy

Fresh material is examined wherever possible in order to avoid artifacts (i.e., images resulting from faulty preparation) and to prevent the loss of certain components. Fresh material is also valuable where time is an important constraint, as in quality control. The examination of fresh material makes for the most rapid and straightforward analysis.

Powders are examined in their original form by dispersing them in either aqueous or nonaqueous mountants, depending on their solubility and dispersibility. The mounting media may include stains to increase the contrast of individual components for structural observation or discrimination on the basis of chemical and morphological characteristics. The detection of birefringence in powders is also useful for identification.

Smearing, squashing, and comminution are simple preparation techniques that make it possible to obtain extensive information quickly. Depending

on consistency, the product is mixed with appropriate dyes or is comminuted with a small amount of water or dye before it is spread on a slide. The sample may also be prepared for microscopy by squashing between slide and coverslip. Using these methods, appropriate dyes yield valuable chemical information on the components of the food, although the overall microstructure is generally lost.

Where information on chemical composition or structural and spatial relationships is required (e.g., comminuted meat products), the sample is sectioned before examination. The simplest and quickest procedure is to section the sample by hand using a clean razor blade, but this is possible only with rigid samples. Soft samples can be made rigid either by freezing or embedding in paraffin or a plastic resin such as glycol methacrylate. Frozen sections are produced in a relatively short time and, with the exception of ice crystals, have the advantage of few artifacts. Ice-crystal damage can be minimized by freezing the sample in nitrogen slush at  $-205^{\circ}\text{C}$  before mounting and sectioning in a cryostat. Rapid work is essential, since ice crystals grow as the sample warms up to the cryostat temperature ( $-20^{\circ}\text{C}$  to  $-25^{\circ}\text{C}$ ). Enzymatic activity in the samples is arrested at cryostat temperatures but can be observed *in situ* using appropriate chromogenic or fluorogenic substrates. With appropriate instrumentation, it is even possible to measure the rate of the reaction.

Fixation with formaldehyde or glutaraldehyde before sectioning prevents the disintegration of fragile sections on the slide when dyes or mountants are applied. The use of slides coated with albumin or poly-L-lysine to collect sections can be helpful. Some components, particularly certain polysaccharides, cannot be stabilized by fixatives, while others of low molecular mass can migrate in the sections or may be extracted from them when dyes are applied. In such cases, vapor staining can be used (e.g., starch with iodine vapor or fat with  $\text{OsO}_4$  vapor). Embedding in a resin is carried out when better resolution structural details are required. In such cases, sections  $0.5\text{--}5\text{ }\mu\text{m}$  in thickness are examined.

### Electron Microscopy

Sample preparation for EM depends on the nature of the food and the type of scanning or transmission technique used. For SEM, dry foods are prepared in the same way as powders. Solid foods based on proteins (cheese, meat) are fixed in glutaraldehyde solutions (1–3%, buffered near the pH value of the food),

dehydrated in ethanol, defatted by extraction with chloroform or *n*-hexane, returned to ethanol, frozen in liquid isopentane over liquid nitrogen or in nitrogen slush, and cryofractured. Since ethanol does not form crystals on freezing, it is possible to freeze ethanol-impregnated samples directly in liquid nitrogen. Dairy products treated in this way are known to fragment spontaneously, but this does not occur if the products are frozen in isopentane or using Freons (although the latter are discouraged due to their negative environmental impact). The fragments are returned to ethanol and critical-point dried. Dried fragments are mounted, coated with pure or alloyed gold, and examined by SEM. If fat is the food component of interest, samples fixed in glutaraldehyde are postfixed with imidazole-buffered  $\text{OsO}_4$  and the fat extraction step is omitted. Viscous samples, which would disintegrate during sample preparation, are encapsulated in agar gel. The capsules are treated in the same way as solid samples. Cryofracturing provides superior images in comparison with dry fracturing. In hydrated food samples, however, crystals may form during freezing resulting in artifacts. Only a thin surface layer ( $20\text{ }\mu\text{m}$ ) is free from such crystals.

Cryo-SEM (examination at temperatures below  $-80^{\circ}\text{C}$ ) is suitable for high-fat foods and foods of plant origin such as cooked vegetables, which could be altered by other methods involving dehydration. Rapid freezing of the samples is essential for the prevention of artifacts arising from ice-crystal formation.

Low-voltage SEM is valuable when applied to frozen uncoated samples because it allows the observation of food in the near-native (frozen hydrated) state at a high magnification, particularly by field emission SEM, which is best suited for this type of work. Since the samples do not need to be coated with metal, they may be fractured continuously over the course of the observation.

In TEM, negative staining and metal shadowing techniques are suitable for macromolecules (proteins, polysaccharides) and their assemblies (micelles). The TEM technique practiced most frequently is the examination of thin sections. In order to examine foods using this technique, small samples ( $<1\text{ mm}^3$ ) are chemically fixed and embedded in a resin such as Epon, Araldite, or Spurr's. Dense foods such as cheese, dough, comminuted meat products, seeds, and extruded products require longer fixation and impregnation times compared to porous foods. Low-viscosity resins facilitate embedding of the samples. Starch granules usually remain nonimpregnated. During sectioning, when the sections

are floated on the surface of water, they swell and expand. Subsequent drying results in the development of folds in the granules. These result in artifacts (dark, star-like areas in the micrographs). Separate protocols have been developed for most foods.

Freeze-fracturing, freeze-etching, and replication with platinum and carbon, followed by TEM of replicates is the technique most often used for the examination of high-fat dairy products such as cream, oil-in-water emulsions such as ice cream, and water-in-oil emulsions such as margarines and low-fat spreads. Micrographs obtained by any kind of microscopy, which used to be recorded on film, are now recorded as digitized images and may be manipulated using computers and software, which come as part of the EM setup.

## Applications

### Contamination of Food

One of the most important applications of food microscopy is the detection and identification of contaminants in food, usually following a consumer complaint. Macroscopic examination is followed by examination under a binocular dissecting microscope. Foreign bodies, such as metal, rubber/plastic, glass, wood, and crystalline solids may be identified at this stage, as also some biological contaminants (e.g., animal or insect parts). Biological contaminants of microbial origin may not be detected at this stage, unless they have formed colonies. Some foreign bodies may occur in finished food because of crystallization (e.g., tyrosine crystals in anchovy paste) or interactions between components (e.g., calcium oxalate crystals in foods containing milk and vegetable constituents). Sometimes food is unintentionally contaminated while being processed in the kitchen (e.g., by glass or enamel particles from kitchen utensils) or even during consumption (e.g., by chips of tooth enamel or fillings). Many of the more difficult cases of food contamination can now be solved with the aid of elemental X-ray microanalysis and using procedures compiled by Lewis (see 'Further Reading' section).

### Structural Analysis

Food microstructure is intimately related to its sensory properties, e.g., elasticity, firmness, smoothness, and juiciness. Proteins and carbohydrates form gels, the structures of which affect water- and fat-retention properties. Gel structures of various biopolymers (e.g., gluten, myosin, starch, glycogen) have been studied extensively. Fats and oils,

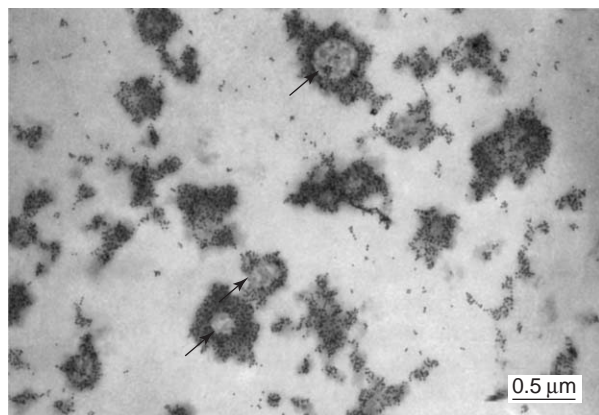
their crystalline forms, emulsions, and foams, represent another important area of food structure studies in which all kinds of microscopy are used. Findings from traditional foods may be applied to newly formulated foods to improve their consumer appeal.

### Dairy Products

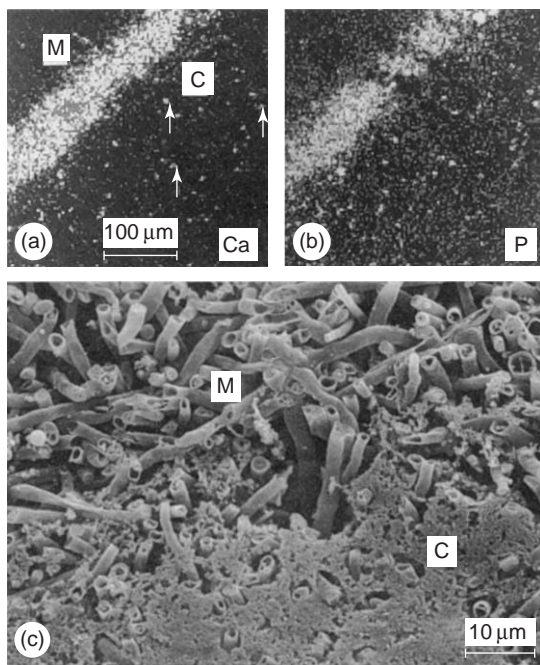
Most dairy products, including yogurts and cheeses, are based on casein micelles (protein globules  $\sim 100$  nm in diameter) and whey proteins. Casein micelles and coagulated whey proteins must be examined by EM because LM techniques do not provide sufficient resolution.

Coagulation of milk results in two different kinds of gel, depending on whether the milk has previously been heated above  $85^{\circ}\text{C}$ . Heating leads to interactions between  $\kappa$ -casein on casein micelle surfaces and a whey protein called  $\beta$ -lactoglobulin. The resulting protein complex occupies sites suitable for interactions with the micelles, and the gel consists of short interlinked chains and minute pores in which the liquid phase is firmly immobilized. This structure retains the liquid phase and is important for soft milk products such as yogurt. Large protein clusters develop in coagula made from unheated milk, and result in syneresis of the gel, i.e., separation of the liquid phase called whey. These structures are important in cheese-making.

Identification of specific food components such as proteins and polysaccharides is possible by immunolocalization using gold-labeled antibodies (Figure 3). In conjunction with X-ray microanalysis, SEM is used to analyze salt crystals in cheeses, including



**Figure 3** Immunogold labeling of egg albumin (minute black dots) which forms a wrapping layer around casein micelles (arrows) in microparticulated protein. The bar represents  $0.5\mu\text{m}$ . (Reproduced from Singer and Dunn (1990) by N.S. Singer and John Wiley & Sons, Inc.; © 1990 Wiley.)



**Figure 4** Results of SEM-EDX analysis showing distribution of (a) calcium and (b) phosphorus at Camembert cheese surface (c) in the mass of mold hyphae M; C image of a freeze-fractured Camembert cheese surface showing the protein matrix C and the hyphae M. (Micrographs (a) and (b) reproduced by courtesy of B.E. Brooker and Scanning Microscopy International.)

deposits of the surface of mold-ripened cheeses such as Camembert. Metabolic activity inside the dense hyphal mat at the cheese surface increases the pH of the cheese and leads to the precipitation of calcium and phosphate ions present in the cheese (**Figure 4**). This process may continue until most of the ions diffuse out of the cheese interior and form a deposit at the cheese surface.

Cryo-SEM or TEM examination of platinum-and-carbon replicas of whipped cream shows stabilization of the air cells by fat globules that adhere to the air-liquid interface. Distribution of intact fat globules and water droplets in butter and other emulsions is also conveniently studied by these methods in which the sample is fixed by rapid freezing.

## Meats

The basic component of meat is muscle tissue, which has a complex microstructure. Individual myofibers are surrounded by connective tissue membranes (endomysium) and bundles of myofibers are separated from each other by another connective tissue membrane, the perimysium. Muscle also contains veins, arteries, and nerve fibers. Skeletal (or striated) muscle differs in microstructure from the smooth muscle that is present in internal organs such as

intestinal walls, blood vessels, etc. Microscopy is used to analyze changes that occur in muscle structures after the animal has been slaughtered, during so-called rigor mortis and then during ageing, freezing, salting, and processing of the meat. These processes cause changes to the muscle microstructure, but the most significant changes are caused by heating, which denatures proteins, leads to shrinkage of the muscle fibers, and the loss of fine structure. Such changes can be observed by LM and EM techniques. Individual components in raw meat, meat heated to over 80°C, and cooked meat can be detected using specific antibodies to native and denatured meat antigens in combination with quantitative fluorescent microscopy.

Comminuted meat products consist of several components and are usually referred to as meat emulsions. The binding of meat and fat particles is improved by the use of fillers, which consist of soy or grain flours, sodium caseinate, starch, gluten, and other polysaccharides and proteins. Their role is to increase yield, improve stability, and modify textural properties. Microscopy is used to detect the presence of fillers, determine the quantity, and establish their distribution.

## Foods of Plant Origin

This group of foods is diverse, encompassing grain flours, dough, bakery products, products derived from starch and/or vegetable proteins, fruits, and vegetables. The consistency of these foods ranges from liquid (fruit and vegetable juices, beer, wine) through moist foods (bread, baked potatoes, tofu) to low-moisture products (potato crisps, bread-crumbs). All these products are derived from plant cells, the walls of which consist of cellulose, hemicelluloses, lignin, pectins, and proteins. The structure and composition of plant cells as foods is discussed in detail by Aguilera and Stanley (see 'Further Reading' section). Foods of plant origin are often prepared by cooking and/or freezing and are amenable to analysis by LM and EM techniques. In contrast, extruded foods are produced by a complex process in which high temperature (100–200°C), high pressure (2–6 MPa), and mechanical shear are applied to the food. The resulting porous as well as compact samples are difficult to embed in resin and to section for LM and TEM. Therefore, SEM is the technique most frequently used to analyze this category of food.

Compared to meat and dairy products, many plant foods contain little or no protein. Starch cannot be fixed by chemical methods, therefore, EM methods involving freeze-fractured samples yield the best results.



## Trends in the Microscopic Analysis of Foods

Over the last few years, there have been significant advances in the analysis of food structures at the microscopic level. These advances have been brought about in part by the development of new techniques and their application to traditional problems, but the inclusion of image analysis as an essential component of quantitative food analysis has also played an important role. Two of the most important developments are the increasing use of AFM and confocal laser scanning microscopy in the analysis of food microstructure. For example, AFM has been used to investigate the structure of fats and oils, and has provided convincing evidence for the presence of microcrystal networks within such substances. This was possible thanks to a recent innovation in AFM technology, known as the tapping mode, in which the cantilever connected to the stylus is oscillated as it scans the sample surface, allowing AFM to be used with soft and hydrated food materials. Another example is the use of CLSM to study the impact of oil frying on potato slices. The ability to generate non-destructive optical sections of the sample facilitated the detection of oil penetration into the crust, followed by three-dimensional reconstructions that showed the oil forming around the cells and penetrating the food through intercellular spaces, while almost no oil was seen in the cells themselves. It is likely that these two key techniques will complement LM and EM methods and play an increasingly important role in food structural analysis in the future.

*See also: Microscopy: Overview. Microscopy Techniques: Light Microscopy; Electron Microscopy; Specimen Preparation for Electron Microscopy; Scanning Electron Microscopy; Atomic Force and Scanning Tunneling Microscopy; X-Ray Microscopy.*

## Further Reading

- Aguilera JM and Stanley DW (1990) *Microstructural Principles of Food Processing and Engineering*. London/New York: Elsevier.
- Aguilera JM, Stanley DW, and Baker KW (2000) New dimensions in microstructure of food products. *Trends in Food Science and Technology* 11: 3–9.
- Ferrando M and Spiess WEL (2000) Confocal scanning laser microscopy: a powerful tool in food science. *Food Science and Technology International* 6: 267–284.
- Flint O (ed.) (1994) *Food Microscopy, Royal Microscopical Society Microscopy Handbooks*. Oxford, UK: BIOS Scientific Publishers.
- Kaláb M, Allan-Wojtas P, and Miller SS (1995) Microscopy and other imaging techniques in food structure analysis. *Trends in Food Science and Technology* 6: 177–186.
- Kirby AR, Gunning AP, and Morris VJ (1995) Atomic-force microscopy in food research – a new technique comes of age. *Trends in Food Science and Technology* 6: 359–365.
- Lewis DF (1993) A tutorial and comprehensive bibliography on the identification of foreign bodies found in food. *Food Structure* 12: 365–378.
- Vaughan JG (1979) *Food Microscopy*. London: Academic Press.
- Wilkinson C, Dijksterhuis GB, and Minekus M (2000) From food structure to texture. *Trends in Food Science and Technology* 11: 442–450.

## Forensic

**C S Palenik and S J Palenik**, Microtrace, Elgin, IL, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The field of forensic microscopy is based upon Edmond Locard's Exchange Principle, which states that when two objects come into contact, a transfer of material will result. Although the amount of material that is transferred in a contact may be minute, microscopical examination can often reveal a great amount of information regarding the materials that came into contact and the way in which they made contact. This article illustrates the way in which forensic microscopy can capitalize on this transfer of

material to aid in establishing the facts within a criminal investigation, describe the types of problems that are addressed, the types of materials encountered, and the analytical approach used to extract information from trace evidence. Finally, a practical example that illustrates the extent to which microscopic evidence can influence the outcome of a case will be described.

## The Role of Forensic Microscopy

The field of forensic microscopy is quite broad in that it deals with any possible type of microscopic residue, which may have biological, geological, or anthropogenic origins. For example, in sorting through the material encountered on a single pair of pants

## Trends in the Microscopic Analysis of Foods

Over the last few years, there have been significant advances in the analysis of food structures at the microscopic level. These advances have been brought about in part by the development of new techniques and their application to traditional problems, but the inclusion of image analysis as an essential component of quantitative food analysis has also played an important role. Two of the most important developments are the increasing use of AFM and confocal laser scanning microscopy in the analysis of food microstructure. For example, AFM has been used to investigate the structure of fats and oils, and has provided convincing evidence for the presence of microcrystal networks within such substances. This was possible thanks to a recent innovation in AFM technology, known as the tapping mode, in which the cantilever connected to the stylus is oscillated as it scans the sample surface, allowing AFM to be used with soft and hydrated food materials. Another example is the use of CLSM to study the impact of oil frying on potato slices. The ability to generate non-destructive optical sections of the sample facilitated the detection of oil penetration into the crust, followed by three-dimensional reconstructions that showed the oil forming around the cells and penetrating the food through intercellular spaces, while almost no oil was seen in the cells themselves. It is likely that these two key techniques will complement LM and EM methods and play an increasingly important role in food structural analysis in the future.

*See also: Microscopy: Overview. Microscopy Techniques: Light Microscopy; Electron Microscopy; Specimen Preparation for Electron Microscopy; Scanning Electron Microscopy; Atomic Force and Scanning Tunneling Microscopy; X-Ray Microscopy.*

## Further Reading

- Aguilera JM and Stanley DW (1990) *Microstructural Principles of Food Processing and Engineering*. London/New York: Elsevier.
- Aguilera JM, Stanley DW, and Baker KW (2000) New dimensions in microstructure of food products. *Trends in Food Science and Technology* 11: 3–9.
- Ferrando M and Spiess WEL (2000) Confocal scanning laser microscopy: a powerful tool in food science. *Food Science and Technology International* 6: 267–284.
- Flint O (ed.) (1994) *Food Microscopy, Royal Microscopical Society Microscopy Handbooks*. Oxford, UK: BIOS Scientific Publishers.
- Kaláb M, Allan-Wojtas P, and Miller SS (1995) Microscopy and other imaging techniques in food structure analysis. *Trends in Food Science and Technology* 6: 177–186.
- Kirby AR, Gunning AP, and Morris VJ (1995) Atomic-force microscopy in food research – a new technique comes of age. *Trends in Food Science and Technology* 6: 359–365.
- Lewis DF (1993) A tutorial and comprehensive bibliography on the identification of foreign bodies found in food. *Food Structure* 12: 365–378.
- Vaughan JG (1979) *Food Microscopy*. London: Academic Press.
- Wilkinson C, Dijksterhuis GB, and Minekus M (2000) From food structure to texture. *Trends in Food Science and Technology* 11: 442–450.

## Forensic

**C S Palenik and S J Palenik**, Microtrace, Elgin, IL, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The field of forensic microscopy is based upon Edmond Locard's Exchange Principle, which states that when two objects come into contact, a transfer of material will result. Although the amount of material that is transferred in a contact may be minute, microscopical examination can often reveal a great amount of information regarding the materials that came into contact and the way in which they made contact. This article illustrates the way in which forensic microscopy can capitalize on this transfer of

material to aid in establishing the facts within a criminal investigation, describe the types of problems that are addressed, the types of materials encountered, and the analytical approach used to extract information from trace evidence. Finally, a practical example that illustrates the extent to which microscopic evidence can influence the outcome of a case will be described.

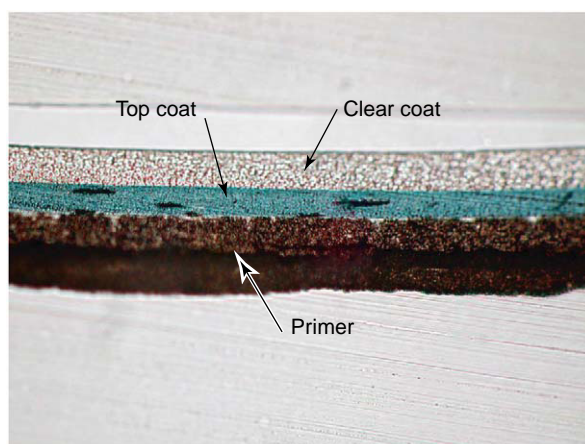
## The Role of Forensic Microscopy

The field of forensic microscopy is quite broad in that it deals with any possible type of microscopic residue, which may have biological, geological, or anthropogenic origins. For example, in sorting through the material encountered on a single pair of pants

recovered from a victim at a crime scene, the microscopist may encounter: soil and pollen from environments to which the victim has been exposed; fibers from clothing, carpeting, or automobiles; biological stains (e.g., blood or semen); hair from the victim and/or suspect; or animal hair. In addition, the victim may have picked up materials more characteristic of a specific anthropogenic environment such as paint, asphalt, glass, wood, metal. These are only a few of the many types of materials that may be encountered, and it is the job of a forensic microscopist to (1) determine the identity of these materials, (2) determine which may have potential value as forensic evidence, and, in some cases (3) establish the significance of the potential evidence in the context of the case. For this reason, a forensic microscopist must possess both a broad general knowledge of materials and a detailed knowledge of the specific microscopic characteristics of these materials so as to be able to classify, if not outright identify the materials encountered as specifically as possible.

Forensic examinations generally progress toward one of two ends: comparison or investigation (i.e., intelligence). Comparative cases are the norm and involve the comparison of suspect material to materials of known origin. In these cases, the goal is to link a suspect to victim or vice versa. An example would be a microscopical comparison between hairs found on a victim to a reference hair sample collected by investigators from a suspect. The less common, and more difficult, investigation involves cases in which police have no suspect and are attempting to establish investigative leads. In these cases, the goal of forensic microscopy is to study microscopic trace evidence to provide police with characteristics of the suspect and/or the type of environment upon which to focus their efforts. In these cases, the forensic microscopist must extract as much information as possible from the trace evidence recovered from a victim. An example of this type of investigation would be a hit-and-run collision in which paint is transferred from the offender's vehicle to the victim. These microscopic chips of transferred paint (Figure 1) can be recovered and studied in terms of the structure of paint layers, color, and chemistry to determine the make, model, year, and color of the involved vehicle. The police can use this information to narrow their search to a particular type of vehicle.

While forensic microscopy is most often considered in regard to the prosecution of a criminal case, microscopic evidence can also be used to advantage by the defense, or even within the realm of civil litigation. Additionally, while this article presents forensic microscopy as an independent topic, the reader must keep in mind that microscopic trace



**Figure 1** Cross-section of a chip of automotive paint. Microscopic examination of the layers and layer thickness, color of each layer, and chemistry of the paint in each layer can be traced back to a specific paint manufacturer and possibly even make, model, and vehicle year. (Courtesy C. Palenik.)

evidence is often used in conjunction with, or is supplemented by, other types of physical evidence (e.g., DNA, fingerprints).

## Materials Encountered

As mentioned above, microscopic evidence spans a vast range of materials (see Table 1). Some material can be easily recognized as hair, a fiber, or soil. Other material, such as biological stains, residues, and white powders, are not immediately recognizable. Take, for example, the recent need to identify anthrax spores and the macroscopic similarity they pose to other common white powders such as wheat flour, starch, or other less dangerous bacteria. In general, materials that cannot be identified by sight are classified and subclassified on the basis of their physical properties. A preliminary microscopic investigation of trace evidence is generally sufficient to classify a material's origin as biological, geological, or anthropogenic. Further classification of a particular sample, or particle, can be made more specific by using morphological characteristics and, if necessary, additional analytical methods to provide more detailed information. For example, even a material such as glass (a transparent, isotropic, conchoidally fractured solid), which can be quickly identified on the basis of morphology and a preliminary examination of optical properties, can be further examined to provide a precise determination of refractive index and trace elemental composition, which will aid in determination of the source of the glass. The procedures by which samples are characterized are dependent on the sample as well as the requirements of a particular investigation (for more information, see

**Table 1** Example of the variety of trace evidence that may encountered in a microscopic examination

<i>Category</i>	<i>Class</i>	<i>Subclass</i>
Biological	Pollen	
	Vegetable fibers	Cordage, fabrics, clothing
	Seeds	
	Hair	Human, animal
	Diatoms	
Geological	Wood	Hardwood, softwood
	Soil	
	Sand	
Anthropogenic	Metals	
	Polymers	Fibers (auto, home, clothing) Paint (auto, home) Plastic
	Industrial	Chemicals, pigments, explosives
	Pharmaceutical	Drugs, herbal supplements
	Paper	Paper fibers
	Glass	Window, bottle, windshield, light bulb
	Food	Flour, starch, prepared foods

the next section ‘Techniques’). The remainder of this section will briefly discuss some of the more commonly encountered microscopic evidence and a survey of the potential information that microscopical study may provide.

### Geological Materials

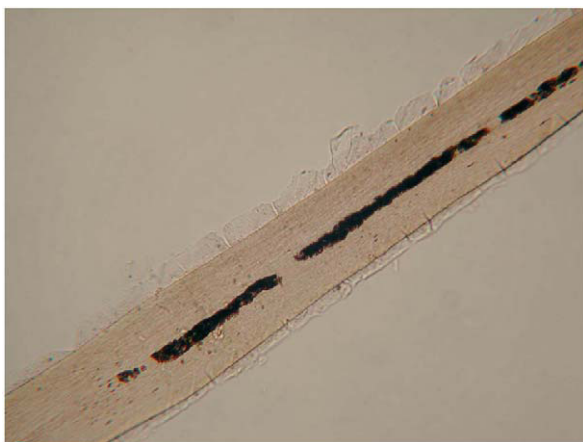
Geological materials include soils, sand, and dust. These substances are commonly recovered from the clothing and shoes of suspects and victims; tires, wheel-wells, and floor mats of vehicles; and from the scene of a crime. The amount of material available for microscopic analysis can vary from several pounds of material accumulated on a vehicle to microscopic amounts of soil vacuumed from the clothing of a victim. These materials can be used for comparative purposes, to demonstrate a suspect or victim has been present at a particular site, or they may aid in establishing investigative leads in cases where no suspect exists. In the first case, soil from a suspected location is compared to a number of known samples from various locations collected by investigators. Although each case is different, a typical examination would begin with a comparison of color and particle size between suspect and known samples. The examination would continue with an examination by polarized light microscopy to identify the mineralogical composition of the suspect soil,

which can then be compared to the minerals identified in known soils. In investigative cases, the police are looking for leads, so soil recovered from a victim may be examined for the purpose of establishing the environment of the location at which the victim picked up this material. Every environment has a characteristic, though not necessarily unique, composition of soil. A forest’s soil will be rich in organic matter; a desert soil will have virtually no organic matter, but may contain characteristic salts. On a regional scale, the type of rock from which the soil originates (igneous, metamorphic, or sedimentary) can be determined from a soil’s mineralogy and may help to define or differentiate specific locations. In addition to purely geological materials, soil often contains other materials such as concrete, asphalt, or pollen, which can be used in conjunction with the geological components to make a comparison or investigative lead stronger or more helpful.

### Biological Materials

The most commonly encountered biological materials in microscopical examinations include hair and fur, pollen, diatoms, wood, vegetable fibers, and residues, such as blood, food, and semen. The solid materials in this list are identified primarily on the basis of morphology, while residues are most easily identified by specific microchemical tests. Similar to geological materials, biological materials can serve to provide an investigative lead, or aid in comparative examinations.

The microscopical identification and examination of hairs can be treated as an independent examination, or can be used in conjunction with DNA analysis. In addition to human hairs, animal hairs can also play a significant role in criminal investigations. The hair of an animal species (human, cat, dog, etc.), as well as the location from which a hair originated (head, limb, pubic, etc.), can normally be identified by morphological characteristics of the cuticle, cortex, and medulla (shown in **Figure 2**). In human hair, various morphological tests such as pigmentation, thickness, and the amount of curl can be suggestive of race. In animal hair, these characteristics can aid to distinguish among species of a particular type of animal, such as the breed of a dog. In addition to these natural comparative features, examination of the root can provide evidence as to the way in which the hair was removed (i.e., forcibly removed, shedded, cut, etc.) and whether the hair was removed postmortem. The tip of a head hair can provide evidence as to relative time of the last haircut and how the hair was cut (i.e., with scissors or a razor). Additionally, hair treatments, such as dying



**Figure 2** Structure of an animal hair. The morphology of the cuticle, cortex, and medulla along with other features such as pigmentation, inclusions, and thickness can aid in the identification of the hair species. (Courtesy S. Palenik.)

or the addition of styling products, can be identified or extracted for further examination. The identification of such characteristics makes a hair more unique, and thus more useful in a microchemical or instrumental examination.

Pollen grains are the male gametophytes of seed-bearing plants. They are produced and released at specific times of the year. This predictable dispersal, coupled with the fact that plants grow in well-defined ecological niches, makes them useful as environmental indicators and, therefore, pollen also serves as an excellent location indicator. In some cases, this information may be as general as to suggest that a victim has been struggling in a field of dandelions, or as specific as to suggest that police seek an indoor greenhouse that contains a rare type of flowering plant not common to the continent.

Diatoms are the siliceous skeletons of certain algae. These may be derived directly from an aqueous source, but may also be found in diatomaceous earth. Commercial uses for diatoms include safe insulation (usually older safes), as mild abrasives, filters, or as a filler in concrete, plastics, paint, and paper. They may also be encountered in certain types of soil samples. The presence (or absence) of diatoms, the number of diatoms, and the species of diatom can all provide useful information.

Wood may be encountered in a variety of forms including: recognizable pieces, splinters, environmental dust, tool dust, or as filler in polymers or paper fibers. Both the species and the end-use of a wood sample can aid in comparative and investigative examinations. Even small slivers of wood can be classified into a proper subdivision as an angiosperm (hardwood) or gymnosperm (softwood) based on fundamental microscopic structural differences.

Identification to the species level may be completed if the sample is large enough to permit observation of other morphological identifiers. Wood that has been treated for industrial uses, such as paper, can be identified as such microscopically. Further microchemical tests can be applied to paper-making fibers to establish the end-use of the paper (i.e., newsprint versus stationery).

Vegetable fibers include materials such as sisal, silk, and cotton, to name a few, which are used in clothing, cordage, carpet backing, or for packing (e.g., jute). Microscopically, these materials can be distinguished from synthetic fibers and further classified on the basis of their microscopic plant anatomy.

### Anthropogenic Materials

Anthropogenic materials encompass the most commonly exploited class of microscopic trace evidence, because the evidentiary significance of a unique anthropogenic material can often be more specific than a biological or geological indicator. As shown in **Table 1**, anthropogenic materials encountered as microscopic evidence include every conceivable material including synthetic and regenerated fibers, paint, glass, concrete, ceramics, metals, plastics, chemicals, drugs, and soot, to name but a few.

The transfer of synthetic fibers from victim to suspect represents one of the most useful types of trace evidence, due to the large amount of information that can be extracted from a single fiber. Useful indicators include the generic class of the fiber (e.g., nylon, polyester, etc.), color (dyed or pigmented), the optical properties (e.g., refractive indices, birefringence), melting point, cross-section, and inclusions, all of which can be quantitatively measured by currently available microscopical, microchemical, and instrumental techniques. Using this information, a single fiber can provide information as to the potential end-use of the fiber (e.g., residential, vehicle, industrial) and, in some cases, even the manufacturer and trade name of fiber. The latter information can increase the evidentiary significance of a fiber by providing a source for information about the availability of a fiber in a particular market. A rare fiber will be potentially more significant than a commonly encountered type of fiber.

Paint evidence is most commonly encountered with regard to automotive paints due to the frequency of transfer in cases involving collisions with automobiles. However, paint recovered from other settings can also provide a wealth of information (see the case description in the final section of this article). **Figure 1** shows a cross-sectional image of



the layers of paint coating an automobile. Each layer in an automotive paint serves a different purpose: the primer layer enables paint to bond to the auto body, the top coat provides the vehicle color, and the clear coat provides ultraviolet and scratch protection to the surface of the car. Each layer, as well as the combination of layers, can aid in the comparison of two paint samples, or in determining the make and model from which an unknown paint chip originated. Within the top coat, or colored layer, the composition of the polymer can be determined by infrared microspectroscopy, which will provide information as to whether the paint is original or repaired. Additionally, the paint color can be quantitatively measured; and the pigments used and their relative proportions can be identified. In the clear coat layer, the polymer used can be identified. Furthermore, the effectiveness of the clear coat layer to protect the colored layer below from ultraviolet (UV) light degrades with time by exposure to the sun. The UV absorption properties of the clear coat layer can provide information about the relative amount of weathering experienced by the paint, which can be characterized by UV-visible microspectrophotometry. Laboratories such as the Bundeskriminalamt in Germany maintain extensive physical collections representing thousands of samples of automotive paint and corresponding chemical analyses. This information is organized in a searchable computer database accessible to laboratories across the nation to assist in quickly identifying the vehicle involved in a hit-and-run case.

Glass is another commonly encountered material in forensic cases. Most glass examinations are undertaken for the purpose of comparison. In some cases, it is more useful to know the source or possible end-use. This can be determined from the refractive index, strain, surface morphology, size, and glass composition (major and trace elements). Common sources of glass include windshields, turn signals and headlight housings, auto-lamp bulbs, household bulbs, windows, drinking glasses, bottles, etc. Each type of glass has characteristic chemical and physical properties. In the case of a comparison between two fragments of glass suspected to be from the same original piece, trace element analysis can be used as a fingerprinting technique to compare the composition of minor elements. In source determinations, the topography of original surfaces as revealed by interference microscopy can be of special value.

As mentioned in the introduction to this section, white powders obtained in small quantities are often among the most difficult materials to identify solely on the basis of morphology. These can include virtually any transparent, colorless solid (organic or

inorganic) and include materials such as salts, chemicals, food products, fertilizers, and explosives. Although some of these materials, such as drugs, are routinely processed by other sections of a crime laboratory (i.e., narcotics), all of these may, from time to time, be extracted or collected in small quantities in a microscopic examination of trace evidence. In these cases, chemical microscopy combined with the measurements of optical properties is typically the best way to approach their identification (see discussion in the techniques section of this article). Additionally, analytical methods such as various spectroscopies can be used to further characterize or identify these substances.

Industrial products such as asphalt, concrete, ceramics, and metals can often provide information about a specific environment or industrial process. In addition to identification of the material, the morphology and surface striations can provide further information, such as the type of tool that produces a metal particle of a particular morphology (e.g., welding spheres produced by an arc-welder, or brass turnings produced by milling of brass).

Finally, microscopic particles that comprise soot and dust can suggest a particular environment. This soot or dust may be collected as residue from a surface, or even from an ear or a nose. Regardless, the soot and dust of many manufacturing processes can be distinguished on the basis of the composition and morphology of the material, or in the case of soot, by the residue of any incompletely combusted material.

## Techniques

The general trends in forensic science are tending toward standardized operating procedures and protocols of analysis. Due to the variety of materials encountered and the wide range of analytical techniques available to characterize samples (as will be discussed in this section), a standardized set of procedure for analysis in forensic microscopy cannot exist as rigorously as in other types of analyses (e.g., narcotics, DNA). The nature of the analysis is highly dependent upon the type of sample, the amount of sample, and the questions that are being investigated. This, in turn, dictates the path of the analysis and the analytical techniques that are applied. For simplicity, this section will discuss the general approach that is followed in a typical microscopical examination. A complimentary discussion of many of these techniques is found elsewhere.

### Sample Collection, Isolation, and Preparation

Samples can be recovered from a variety of sources including clothing, skin, carpeting, and vehicles, to

name but a few. To recover trace evidence, which often occur as only dust or even as a single particle, a variety of techniques have been developed. These include: ‘scraping’, the least elegant technique where material is roughly knocked off of an item of evidence onto a fresh surface where the loose material can be collected; ‘hand picking’, where specific particles removed by means of forceps, needles, or magnets are transferred to a microscope slide or Petri dish; adhesive tape, which, when applied to a surface such as fabric, will lift loose particles from a surface; a vacuum filter, whereby a vacuum pulls particles from the surface of evidence, which are then concentrated on a membrane filter in the path of the vacuum line; and extraction, where soluble residues are dissolved from a piece of evidence and the solvent evaporated.

Once collected, the material concentrated on a microscope slide, adhesive tape, or filter is examined under a low-power stereomicroscope. **Figure 3A** shows a low-magnification photomicrograph of sand taken with a stereomicroscope (**Figure 3B**). Using a stereomicroscope, the forensic microscopist can characterize the heterogeneity of a sample and carry out a preliminary classification to determine the next analytical step. The stereomicroscope also serves as an aid when manipulating and isolating particles for other techniques.

Sample manipulation is typically carried out under the microscope using a sharpened tungsten needle. The manipulation of individual particles as small as a few micrometers in diameter can be carried out with relative ease after training and practice. In addition, microscopic specimens often need to be divided to preserve material for additional analyses, or cut into thin sections to observe a material such as wood or a fiber from a particular orientation. While a microtome can produce extremely thin and uniform sections, a skilled microscopist can cut free-hand thin-sections, suitable for most microscopic analyses, under the stereomicroscope using a razor blade much more quickly than a microtome.

During an examination under the stereomicroscope, particles of particular interest are isolated and mounted for examination by polarized light microscopy, typically the next step in a forensic microscopical examination. By this point, the microscopist typically has a good idea of the material being observed, or at least the class of material, which aids in the choice of method for mounting the particle for additional analyses.

### Advanced Analytical Examination

Typically, the next step is examination using the polarized light microscope (PLM). The PLM (**Figure 4A**) is a compound microscope fitted with two



(A)



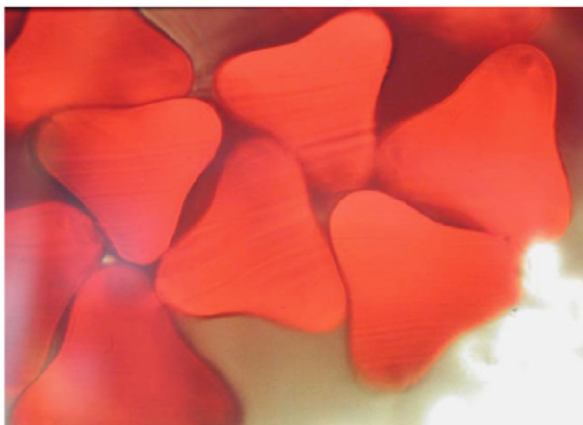
(B)

**Figure 3** (A) Image of sand grains taken through a stereomicroscope. A low-power examination provides an initial characterization and allows for further classification of the material. (Courtesy S. Palenik.) (B) Photomicrograph of a stereomicroscope, a low-power microscope used to conduct a preliminary examination of trace evidence. (Courtesy C. Palenik.)

polarizers. Using polarized light microscopy, the optical properties (e.g., refractive index, birefringence) of a transparent solid compound can be observed and measured. These properties have been tabulated for the most common organic and inorganic compounds and can be used to identify an unknown material.



(A)

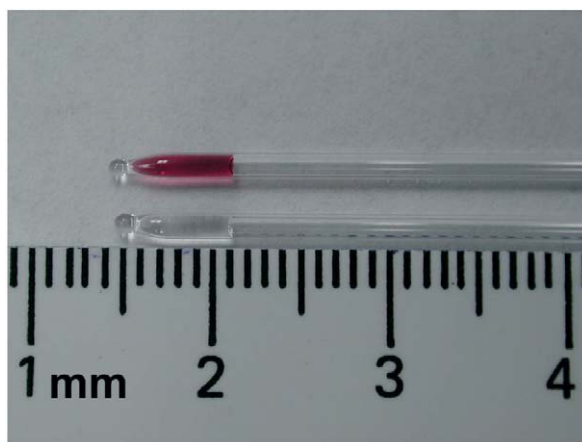


(B)

**Figure 4** (A) Polarized light microscope. (Courtesy C. Palenik.)  
 (B) A photomicrograph showing the trilobal cross-section of a nylon carpet fiber with plane polarized light through a PLM. (Courtesy S. Palenik.)

**Figure 4B** is a photomicrograph of a cross-section of a trilobal nylon carpet fiber photographed in plane polarized light. The geometry of the cross-section is a characteristic that can provide information about the manufacturer and end-use of the carpeting.

Microchemical analysis, a branch of microchemistry, is based on the performance of chemical reactions using single particles and microliters of solution



**Figure 5** Microchemical test for an aldehyde. The two capillary tubes indicate the positive (colored) and negative (colorless) results for this test. Each capillary contains  $\sim 100\ \mu\text{l}$  of solution, indicating the scale at which these reactions can be carried out. (Courtesy S. Palenik.)

to identify functional groups, ions, or compounds. Typically, microchemical reactions are carried out in capillary tubes, on microscope slides, or on filter paper. Developed before the other analytical techniques mentioned below were invented, microchemistry remains a rapid, accurate, and extremely sensitive means by which ions, functional groups, and compounds can be identified. The microchemical test for an aldehyde using Schiff's reagent is illustrated in **Figure 5**. Both capillary tubes in the figure contain  $\sim 100\ \mu\text{l}$  of solution. The lower tube is a blank, illustrating the negative test result. The upper capillary tube has become red upon reaction with Schiff's reagent, and indicates the presence of the aldehyde group in the sample. The scale upon which these reactions can be carried out allows for the detection of picogram quantities of materials.

The scanning electron microscope (SEM) coupled with an energy dispersive X-ray spectrometer (**Figure 6A**) can be used to study materials at magnifications that far exceed the maximum of the light microscope, and, sometimes, most importantly, provide elemental analysis of materials. **Figure 6B** shows a photomicrograph of a quartz grain collected with the SEM. The surface features of the grain shown in **Figure 6B** can be studied in considerably greater detail than is possible with a PLM due to the higher resolution of the SEM. In this grain, sharp, fresh fracture surfaces are visible, suggesting that this grain is freshly broken from its host rock. This grain was deposited by a glacier. Morphological details such as these can be exploited when comparing samples or searching for investigative leads. In addition to imaging, the energy dispersive X-ray spectrometer is capable of



(A)



(B)

**Figure 6** (A) The scanning electron microscope coupled with energy-dispersive spectroscopy pictured here is used to study materials at higher magnifications and determine elemental compositions. (Courtesy C. Palenik.) (B) An SEM photomicrograph of a quartz grain can reveal details about its geologic history and its environment of deposition. (Courtesy S. Palenik.)

qualitatively identifying the elemental composition of a material.

Other useful microscopic analytical techniques include hot stage, fluorescence, and cathodoluminescence microscopies; micro-infrared spectroscopy; micro-Raman spectroscopy; ultraviolet-visible microspectrophotometry; and X-ray diffraction; however, the discussion of these techniques is beyond the scope of this article. Briefly stated, each of these techniques can be used to ascertain additional information about characteristic properties of a material. The microscopist must be aware of all of these techniques, and others, so as to be able to extract the necessary information from a sample when the need arises.

## Application of Scientific Results to Forensic Casework

Many routine analyses of evidence are carried out in forensic science, with DNA, fingerprints, and drug analysis. The cost associated with microscopical examinations relative to some of these more routine analyses, along with the expertise required of a forensic microscopist, and, finally, the lack of understanding as to the amount of information that can be provided by even a preliminary microscopic examination has limited the number of cases to which the techniques discussed in this article are applied. As a result, the power of this field is often reserved until an investigation is well advanced, a case is considered to be 'high profile', or an investigation has reached a dead end. A prime example of a case in which forensic microscopy played a key role was the Green River serial murders in the Seattle, WA area, which stretched over more than 15 years.

Beginning in 1982, police in Washington began to find bodies of murdered women who all fit a similar profile. These murders continued for nearly 20 years as murders attributed to a single murderer began to accumulate until the total exceeded 40 people. Twice during this time, police made inquiries into a suspect, Gary Ridgway; however, they had no specific evidence linking him to the crimes and he could never be charged. In 2001, DNA evidence recovered from the bodies of four victims linked Ridgway to the murders. While DNA evidence proved Ridgway had been with the women, it did not prove Ridgway was the murderer, and Ridgway planned to plead innocent and fight the charges in court.

Looking for more evidence, investigators requested a microscopic review of the victims clothing for trace evidence. When examined, similar types of microscopic particles of spray paint were identified on several of the victims. Using light microscopy, electron microscopy, and micro-infrared spectroscopy, the paint manufacturer was determined. The formulation turned out to be a rare type of paint, used in only in small quantities, mainly by repaint shops and detailers. Ridgway was a painter and detailer in a large truck manufacturing facility in the Seattle area that used this specific type of paint as an original finish.

The paint particles were picked up by Ridgway at work, embedded in his clothing and car, and then transferred to anyone else in contact with Ridgway or his vehicle. The paint provided a direct link between Ridgway and several of the murdered women, as well as a link between the women. As a result of this new evidence, Ridgway was charged with three



new counts of murder. Soon after, he agreed to a plea bargain and plead guilty to 48 counts of murder, the most ever by a single person in the United States, in exchange for avoiding a death sentence.

This case example serves to illustrate several points: first, that forensic microscopy should be applied as an early line of examination rather than a late one. The above case is a prime example of an examination in which an investigative, rather than comparative, approach may have likely led to an earlier resolution of a serial murder case. Second, the scientific approach pursued in each case involving the microscopic examination of trace evidence may be different. Thus, the forensic microscopist must be a true scientist who is able to adapt the techniques and methods available to the requirements of the case at hand and not be bound to a standardized procedure devised for the use of technicians.

See also: **Ceramics. Forensic Sciences:** Blood Analysis; Fibers; Glass; Hair; Paints, Varnishes, and Lacquers; Systematic Drug Identification. **Microscopy:** Overview.

**Microscopy Techniques:** Scanning Electron Microscopy. **Polymers:** Natural Rubber; Synthetic.

## Further Reading

- Catling D and Grayson J (1998) *Identification of Vegetable Fibers*. Los Gatos, CA: Archetype Publishing.
- Deedrick DW and Kock SL (2004) Microscopy of hair. Part 1: A practical guide and manual for human hairs. *Forensic Science Communications* 6 ([www.fbi.gov/hq](http://www.fbi.gov/hq)).
- Houck M (2001) *Mute Witness: Trace Evidence Analysis*. New York: Academic Press.
- Maxen E (2004) Murders under a microscope. *Courier News*, Elgin, IL, p. 1.
- McCrone WC, McCrone LB, and Delly JG (1999) *Polarized Light Microscopy*. Chicago: McCrone Research Institute.
- Palenik SJ (1999) Microscopical examination of fibers. In: Robertson J and Grieve M (eds.) *Forensic Examination of Fibers*, pp. 179–222. London: Taylor & Francis.
- Palenik SJ and Palenik CS (2004) Microscopy and microchemistry of physical evidence. In: Saferstien R (ed.) *Forensic Science Handbook*, vol. II. Englewood Cliffs, NJ: Prentice Hall.

## Environmental

**D Perret**, Man-Earth Interface, Geneva, Switzerland

**G G Leppard**, National Water Research Institute, Burlington, ON, Canada

**D Mavrocordatos**<sup>†</sup>, EAWAG, Urban Water Management, Dübendorf, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

## Brief History

The invention of electron microscopy dates back to 1931 and is credited to Ruska, who succeeded in identifying small biological objects with a transmission electron microscope (TEM) operated at low voltage (50 kV) and exhibiting a resolution far beyond that of photon microscopy. Eventually, Ruska was awarded the 1986 Nobel Prize for his fundamental work in electron optics, shared with Binnig and Rohrer for their invention of the scanning tunneling microscope in the late 1970s.

The developments in lenses and specimen preparation took almost a decade since the invention prior to the first publications on the use of electron microscopy and interpretation of micrographs for the investigation of environmental samples. The

establishment of the earliest scanning-transmission electron microscopes (STEMs) in the late 1930s, the first scanning electron microscope (SEM) in the early 1950s, followed by the development of high-voltage microscopes, the incorporation of different detectors (energy-dispersive X-ray spectroscopy, EDS; electron energy loss spectroscopy, EELS), and the development of brighter electron sources (LaB<sub>6</sub>; field emission guns, FEGs) in the 1960s and 1970s opened increasingly new opportunities for the study and the analysis of a broad range of environmental specimens with various thicknesses and compositions. The emergence of environmental scanning electron microscopes (ESEMs) in the late 1980s dramatically completes the palette of instruments amenable to the study of samples from the environment.

The earliest observation of soil particles by electron microscopy dates back to 1940, when a preparation scheme for the identification of airborne particles at magnifications up to 200 000× was initially published in 1946. It was, however, not until the mid-1970s that the first papers dealing with suspended aquatic particulates finally appeared. Since then more than a 1000 publications have been reported to link the use of electron microscopies to environmental particles and colloids. Historically, electron microscopies have been exploited for the evidencing

<sup>†</sup>Dr. Denis Mavrocordatos is deceased. He tragically disappeared on Lake Geneva on 6 December 2003. "Homme libre, toujours tu chériras la mer!" (Charles Baudelaire).



new counts of murder. Soon after, he agreed to a plea bargain and plead guilty to 48 counts of murder, the most ever by a single person in the United States, in exchange for avoiding a death sentence.

This case example serves to illustrate several points: first, that forensic microscopy should be applied as an early line of examination rather than a late one. The above case is a prime example of an examination in which an investigative, rather than comparative, approach may have likely led to an earlier resolution of a serial murder case. Second, the scientific approach pursued in each case involving the microscopic examination of trace evidence may be different. Thus, the forensic microscopist must be a true scientist who is able to adapt the techniques and methods available to the requirements of the case at hand and not be bound to a standardized procedure devised for the use of technicians.

See also: **Ceramics. Forensic Sciences:** Blood Analysis; Fibers; Glass; Hair; Paints, Varnishes, and Lacquers; Systematic Drug Identification. **Microscopy:** Overview.

**Microscopy Techniques:** Scanning Electron Microscopy. **Polymers:** Natural Rubber; Synthetic.

## Further Reading

- Catling D and Grayson J (1998) *Identification of Vegetable Fibers*. Los Gatos, CA: Archetype Publishing.
- Deedrick DW and Kock SL (2004) Microscopy of hair. Part 1: A practical guide and manual for human hairs. *Forensic Science Communications* 6 ([www.fbi.gov/hq](http://www.fbi.gov/hq)).
- Houck M (2001) *Mute Witness: Trace Evidence Analysis*. New York: Academic Press.
- Maxen E (2004) Murders under a microscope. *Courier News*, Elgin, IL, p. 1.
- McCrone WC, McCrone LB, and Delly JG (1999) *Polarized Light Microscopy*. Chicago: McCrone Research Institute.
- Palenik SJ (1999) Microscopical examination of fibers. In: Robertson J and Grieve M (eds.) *Forensic Examination of Fibers*, pp. 179–222. London: Taylor & Francis.
- Palenik SJ and Palenik CS (2004) Microscopy and microchemistry of physical evidence. In: Saferstien R (ed.) *Forensic Science Handbook*, vol. II. Englewood Cliffs, NJ: Prentice Hall.

## Environmental

**D Perret**, Man-Earth Interface, Geneva, Switzerland

**G G Leppard**, National Water Research Institute, Burlington, ON, Canada

**D Mavrocordatos**<sup>†</sup>, EAWAG, Urban Water Management, Dübendorf, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

## Brief History

The invention of electron microscopy dates back to 1931 and is credited to Ruska, who succeeded in identifying small biological objects with a transmission electron microscope (TEM) operated at low voltage (50 kV) and exhibiting a resolution far beyond that of photon microscopy. Eventually, Ruska was awarded the 1986 Nobel Prize for his fundamental work in electron optics, shared with Binnig and Rohrer for their invention of the scanning tunneling microscope in the late 1970s.

The developments in lenses and specimen preparation took almost a decade since the invention prior to the first publications on the use of electron microscopy and interpretation of micrographs for the investigation of environmental samples. The

establishment of the earliest scanning-transmission electron microscopes (STEMs) in the late 1930s, the first scanning electron microscope (SEM) in the early 1950s, followed by the development of high-voltage microscopes, the incorporation of different detectors (energy-dispersive X-ray spectroscopy, EDS; electron energy loss spectroscopy, EELS), and the development of brighter electron sources (LaB<sub>6</sub>; field emission guns, FEGs) in the 1960s and 1970s opened increasingly new opportunities for the study and the analysis of a broad range of environmental specimens with various thicknesses and compositions. The emergence of environmental scanning electron microscopes (ESEMs) in the late 1980s dramatically completes the palette of instruments amenable to the study of samples from the environment.

The earliest observation of soil particles by electron microscopy dates back to 1940, when a preparation scheme for the identification of airborne particles at magnifications up to 200 000× was initially published in 1946. It was, however, not until the mid-1970s that the first papers dealing with suspended aquatic particulates finally appeared. Since then more than a 1000 publications have been reported to link the use of electron microscopies to environmental particles and colloids. Historically, electron microscopies have been exploited for the evidencing

<sup>†</sup>Dr. Denis Mavrocordatos is deceased. He tragically disappeared on Lake Geneva on 6 December 2003. "Homme libre, toujours tu chériras la mer!" (Charles Baudelaire).

of the morphotypes, textures, and sizes of particulate materials in environmental samples, and the majority of micrographs published had been used until recently for merely illustrative purposes. Nowadays, analytical electron microscopies have emerged from their infancy in the broad field of environmental science, and are used frequently in order to demonstrate and quantify the presence of well-characterized species down to the nanometer scale, or even to assess relationships between the formation or existence of specific types of particles or colloids at the microscopic scale, and the functioning of their natural milieu at the macroscopic scale. Indeed, this article focuses on the specimen preparation methods dedicated to scanning and transmission microscopies, and to the extraordinary potentialities of the latter for solving environmental situations and processes involving complex and heterogeneous species.

## Sampling and Specimen Preparation

Unbiased electron microscopic investigation of particulate material, be it of living (bacteria, protozoa, algae, cellular fragments) or nonliving nature (macromolecular organics, viruses, crystalline or amorphous mineral phases), requires a blend of dedicated sample collection schemes and specimen preparation techniques. **Figure 1** serves as a rough guideline for the selection of the most adequate microscopic approach, from specimen preparation to particle characterization.

Whatever the investigation to be performed on whichever sample, it must be kept in mind that every step of the protocol should be designed to avoid modifications of the native physicochemical characteristics of the sample (e.g., precipitation/coagulation/dissolution due to shifts in particle concentration, ionic strength, pH, redox potential, temperature or uncontrolled dehydration of sample).

### Aerosols and Atmospheric Particles

Atmospheric particles are probably the most straightforward to sample, as a variety of collectors have been devised and optimized for more than half a century, in which particulate material can eventually be size fractionated as a function of the operating conditions of the collector. As opposed to aquatic systems, sampling of particles in a gaseous milieu is less sensitive to physicochemical changes, although biased size fractionation may occur due to particle hydrophilicity/hydrophobicity or surface charges (this should be concurrently checked by scanning mobility particle sizing).

For indoor particles (e.g., asbestos fibers, industrial particles, soot, or combustion smoke of health concern), direct sampling by air pumping through

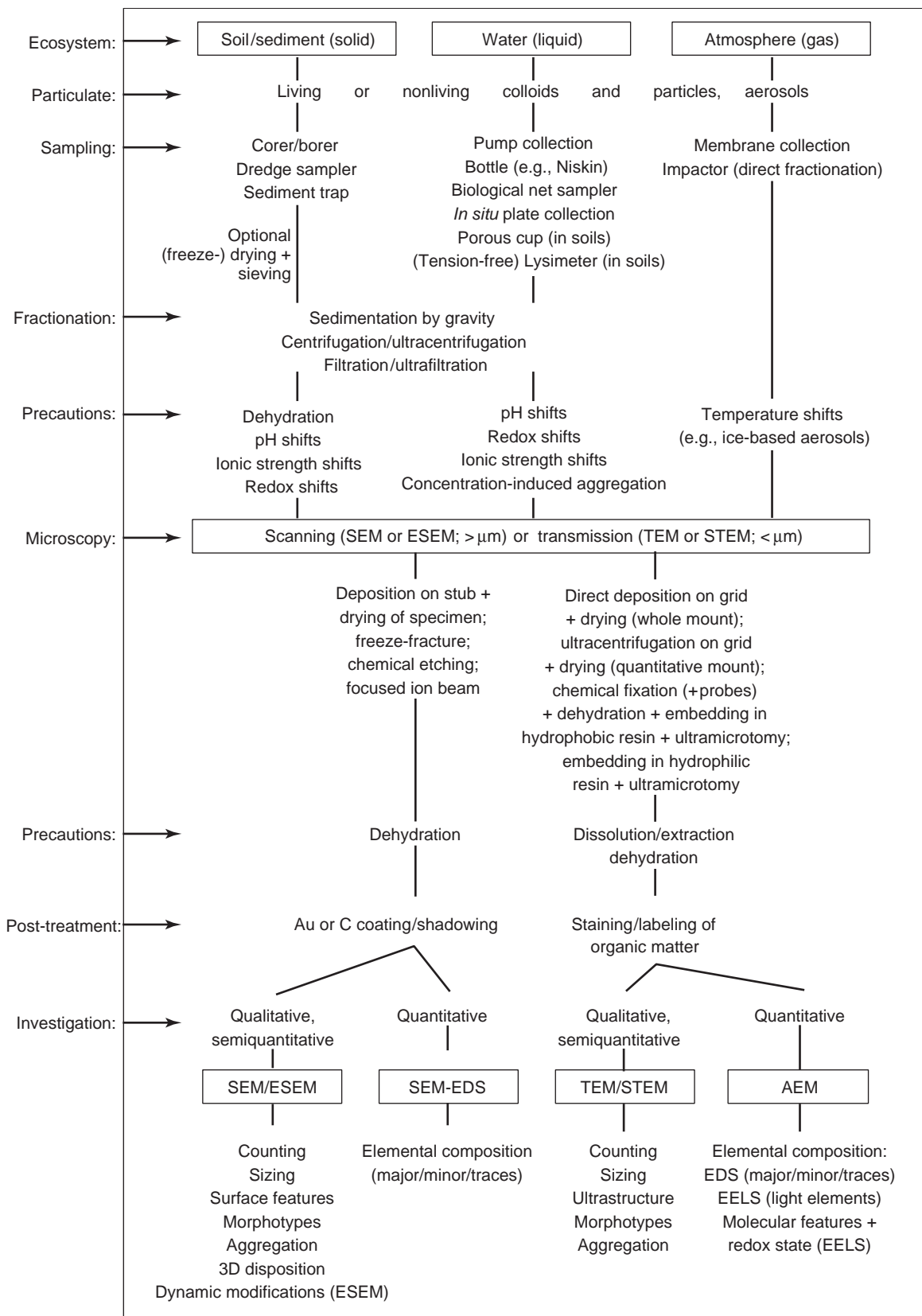
collecting membranes (smooth/flat neutron-impacted Nuclepore<sup>TM</sup>-type filters) without size fractionation is best suited. Outdoor atmospheric or plume particles can be size fractionated and collected by cascade impactors, charged-particle collectors, thermal precipitators, or nephelometers. Particles will have to be transferred by contact from the sampling device to the specimen holder, be it a stub for SEM or a grid for TEM.

The collection of aerosols requires special procedures aimed at maintaining the original temperature and hydration status of the sample, in particular for studies involving wet or iced entities in which the mechanisms of cloud or ice formation are driven by the intimate combination of water, inorganic condensation nuclei, dissolved salts, and organics acting as hygroscopic agents; in these cases, microscopic examination of specimens with a cold stage should be planned within the shortest delay. Transmission (for the smallest colloids) and scanning (for larger particles) electron microscopes, in imaging or elemental analysis modes, are used indiscriminately for the characterization of atmospheric entities, though the ESEM, which allows imaging under various conditions of temperature and pressure, is especially useful for water-containing particles larger than 50 nm.

### Aquatic Particles

The study of aquatic particles (i.e., individual entities and their aggregates) encompasses a broad variety of matrices (marine and freshwaters; surface and groundwater; gravitational and capillary water of soils), particle types (nanometer to millimeter; living and nonliving; organics, organomineral mixtures, and minerals), and processes (structure–composition–function relationships; contaminant transport; mechanisms of formation, and dissolution). Sampling of aquatic particles is thus a complex task that will systematically require adaptation of generic protocols to the specificities of the study. Whatsoever, the most relevant investigations are usually obtained by means of correlative electron microscopic approaches, i.e., schemes involving the use of scanning and transmission microscopes to study in parallel whole mounts and resin-embedded specimens (see below), in order to cover the broadest and most accurate physicochemical fingerprint of particles.

Because of the rapid modifications of the physicochemical and microbiological characteristics of natural water samples, initiated by changes in temperature, pH, depth, dissolved O<sub>2</sub> or CO<sub>2</sub>, light, and convection, aquatic samples should be processed for electron microscopy without any delay to avoid dissolutions, precipitations, coagulations, sedimentations, microbial growth, or shifts in chemical



**Figure 1** Overview of the recommended approaches, crucial steps, and possible artifacts relevant to the physicochemical characterization of particulate material in environmental systems by electron microscopies. This scheme does not take into account precautions required for conventional bulk physicochemical analyses.

equilibria affecting colloids and particles, as it also prevails for conventional bulk chemical analyses. Prefractionation of aquatic particles and colloids into narrower classes can be performed by gravitational sedimentation in thermostatted columns, by single or cascade centrifugations/ultracentrifugations, or by single or cascade filtrations/ultrafiltrations. Both fractionation approaches are subjected to artifacts that can be minimized under carefully controlled conditions: Centrifugations must not be performed with highly concentrated suspensions to avoid problems of differential settling; similarly, tangential-flow filtration at low flow rates usually yields no or less polarization concentration than uncontrolled/unstirred cross-flow filtration. Centrifugation and filtration are usually chosen to eliminate the fractions containing the largest particles, although they may be used instead to collect appreciable amounts of particulate material from poorly charged waters (e.g., groundwater, pristine waters), with possible biases caused by apparent coagulation of particles.

Particulate material can also be sampled from natural waters by direct collection onto vertical or horizontal plates (glass, Teflon<sup>TM</sup>, or other plastics) inserted for a while in the water column. Horizontal plates collect sedimenting particles without the drawbacks of conventional sediment traps (shifts in biological activity and redox conditions due to the absence of mixing at the bottom of the trap); vertical plates selectively collect those entities exhibiting a certain affinity for the plate (e.g., adhering bacteria, polysaccharides, Fe-oxyhydroxides).

Particles in soil water require different sampling approaches, depending on the type of water to be sampled. Gravitational water is better collected by means of tension-free lysimeters; these devices integrate, over time, the sampled water, which should thus be recovered shortly after collection to avoid modifications in the size distribution or chemistry of the particles. It must, however, be underlined that tension-free lysimeters may exhibit fairly low collection efficiencies (as low as 10% of the gravitational water, depending on soil texture and porosity). On the other hand, the capillary water of soils can be recovered by means of suction cups, which are made of a porous material (ceramic or plastic), inserted in the soil and connected to a syringe or pump. Due to their porosity ( $\sim 10\text{--}100\ \mu\text{m}$ ), these devices induce a prefractionation of the sampled particulate, and tend to clog with time.

### Soil and Sediment Particles

These types of particles have been the subject of numerous publications related to the identification of

the phases that constitute the soil or the sediment matrix, and to the formation of soils or sediments and their stratigraphic characteristics, but to a lesser extent to the role of particles as carriers and scavengers of contaminants. For these reasons, sampling and specimen preparation protocols for soil or sediment particles are usually less sophisticated than protocols for aquatic particles, provided that one considers soil and sediment particles as static entities having no degree of freedom in their surrounding water.

While the majority of soil samples are characterized by slow reactivities and thus require fewer precautions, hydromorphic soils subjected to rapid hydration/dehydration processes are highly sensitive to redox changes and should be sampled with the greatest care to avoid precipitation of dissolved species ( $\text{Fe}^{2+}$ ,  $\text{Mn}^{2+}$ ) during accidental aeration. Otherwise, soils are usually dried, sieved, and ground prior to being resuspended in various electrolytes for the selective isolation of certain constituents (e.g., organic matter, clays, silts, sands) by sedimentation or centrifugation. These approaches are indeed not recommended for the identification of trace metals, as drastic morphological and compositional modifications are expected at each step of the protocol. Because of their size, soil particulates are most frequently analyzed by SEM, but the ultrastructural analysis of clay micelles or other finely divided components (e.g., iron oxides) requires TEM examination.

The particulate phases building up aquatic sediments represent an intermediate situation between aquatic particles and soil particles: Except for highly consolidated, deep sediments, their water content is large and their reactivity is comparable to, or even larger than the one of hydromorphic soils. Sediments are sampled by means of vertical corers, from which the different season- or event-dependent strata must be subsampled by slicing in a glove box under controlled atmosphere. Textural analysis of sediment particles can be performed by SEM, but more detailed investigations will require dilution of the samples in an electrolyte of composition similar to the interstitial water, prior to specimen preparation for TEM.

### Specimen Preparation

Specimen preparation for SEM requires collection of particulates onto stubs (either directly, or after precollection of entities onto Nuclepore<sup>TM</sup>-type membranes), followed by postcoating, most frequently with a Au or Pt film (for imaging) or C film (for analysis). The maximum achievable resolution

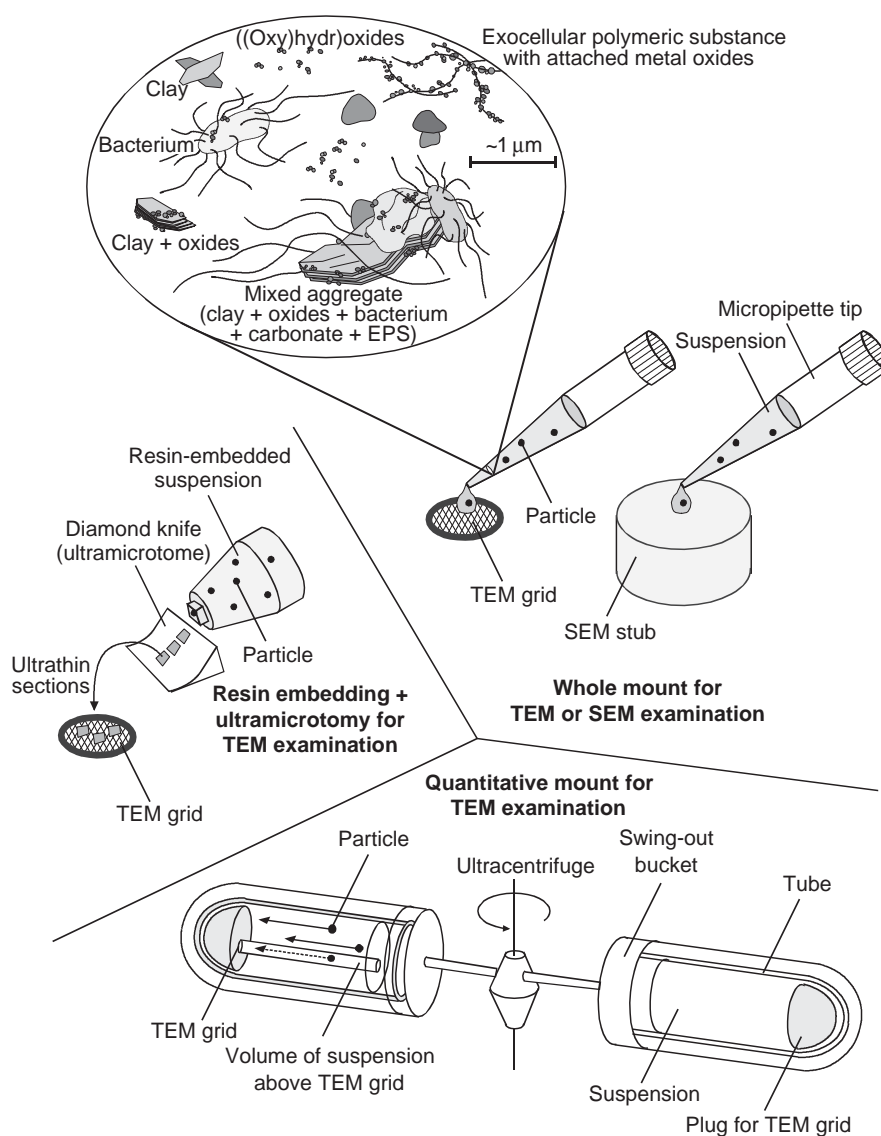
limit that one can expect for SEM of complex heterogeneous environmental particles is  $\sim 10\text{--}50\text{ nm}$ , even for high-intensity FEG-SEM.

Cryotechniques offer an alternative and complementary preparatory technology to chemical fixation and embedding, producing aggregates whose three-dimensional (3D) relationships are spatially 'fixed' by a physical means such as vitrification. The most structurally faithful of the cryotechniques is freeze-etching, which consists of freezing an aquatic sample rapidly enough to vitrify it, mechanically generating a fracture plane through it, and then making a metallic replica of the fracture surface, all the while maintaining the vitrified sample below the recrystallization temperature. The product of the freeze-etch technique is a replica that presents a topographical

image of a colloid or aggregate, untouched by chemical agents and amenable to analysis by TEM.

As compared to SEM, TEM offers a much broader palette of protocols for specimen preparation (see the most prominent procedures in **Figure 2**). Qualitative investigations shall favor the direct deposition of a suspension onto TEM grids (whole mounts), followed by evaporation; this rapid procedure may however induce (1) crystallization of undesirable electrolytes (e.g., salt crystals for marine samples, which can be avoided by rapid rinsing of the grid in ultrapure water), (2) shrinkage of aggregates of flexible organic materials (e.g., exocellular polymeric substances), or (3) coagulation of small colloids.

For the extraction of ultrastructural information (in particular with bacteria, algae, and 3D networks



**Figure 2** Schematic of the most useful qualitative or quantitative specimen preparation schemes for SEM/TEM examination of colloids and particles in aqueous media.



of organic flocs), large size entities (i.e.,  $>1\mu\text{m}$ ) should be visualized by TEM after preparation of thin sections obtained by resin embedding and ultramicrotomy. To avoid artifacts, the choice of an appropriate embedding medium is crucial and should be directed toward a hydrophilic resin (Nanoplast<sup>TM</sup>), rather than more conventional hydrophobic ones (e.g., Spurr<sup>TM</sup>, Epon<sup>TM</sup>, or Araldite<sup>TM</sup>), because the latter require dehydration steps of the sample in organic solvents (acetone, methanol, ethanol, or propylene oxide); these steps are potentially disturbing, as they may cause dissolution of particulate organic moieties and modifications of the morphologies of complex 3D networks containing organic entities. As Nanoplast<sup>TM</sup> produces water molecules during the permeation and polymerization step, it infiltrates readily in porous (e.g., loose and amorphous colloids) and biological (e.g., bacteria) entities, maintaining their fine morphological features.

Ultramicrotomy (see **Figure 2**) of the resin-embedded material should be performed exclusively with a diamond knife to overcome the hardness of mineral particles or biogenic minerals such as silica frustules, which quickly damage conventional glass knives.

Targeted sectioning can also be performed by high-energy focused ion beam (FIB). This emerging technique shall prove highly valuable for environmental science in the near future. In principle, any material (hard inorganic or soft organic entities) is amenable to FIB sectioning.

Ultrastructural investigations require the preparation of ultrathin sections ( $\sim 50\text{--}100\text{ nm}$ ), while larger sections (much easier to obtain, but which should not exceed  $150\text{--}200\text{ nm}$ ) are still amenable to elemental analysis without important biases, even when performing EELS, which theoretically requires the thinnest possible materials.

Probably the most interesting approach to the preparation of specimens for quantitative analysis by TEM is the direct ultracentrifugation of colloidal and particulate entities onto TEM grids. This procedure (see **Figure 2**) yields quantitative whole mounts with evenly distributed particles. For a given suspension, the final coverage of the grid can be finely tuned by varying the volume to be centrifuged, allowing a fairly accurate estimation of the particle concentration in the initial sample, yet avoiding too densely populated grids as is often the case for conventional whole mounts. Entities sensitive to dehydration or redox modifications can be postprotected by horizontally spinning an ultrathin film of hydrophilic Nanoplast<sup>TM</sup> resin. Indeed, the preparation of quantitative mounts can be coupled to sequential fractionation schemes to further narrow the types of particles that are collected onto the grids.

The choice of the grid type and the supporting film is critical for analytical TEM/STEM (in EDS or EELS modes). One must indeed be reminded that the X-rays produced by the analyte under the focused electron beam are emitted in a sphere of volume and will induce secondary X-rays of the materials they hit (e.g., the supporting grid or the pole pieces of the EM column), which may be emitted in the direction of the detector, generating artifact peaks. While Formvar<sup>TM</sup>- or Parlodion<sup>TM</sup>-coated ( $10\text{--}50\text{ nm}$ ; supporting film needed for strength/flexibility), carbon-sputtered ( $3\text{--}10\text{ nm}$ ; needed for thermal/electrical conductivity) copper grids are the best alternative in price and ease of operation for imaging purposes, elemental analysis requires supporting materials that shall not mask the elements of interest: gold grids are preferred for the EDS analysis of trace transition metals; for the identification of carbon-rich entities prepared as whole mounts or quantitative mounts (i.e., without resin embedding), the supporting film must be substituted by a carbon-free  $5\text{--}15\text{ nm}$  SiO film (not adapted for Si-rich entities) or Be film (expensive); large-scale entities (e.g., 3D networks or organic-mineral mixtures) can be collected on holey or Quantifoil<sup>TM</sup> (supporting film with controlled and repetitive holes) or Lacey carbon films for the unconstrained analysis of their unsupported portions, with the risk of a weaker mechanical and electrical stability under the electron beam. Whichever grid is used, the alphanumeric styles shall be preferred in order to keep track of particles of interest over time.

Indeed, staining procedures designed for biological or medical applications can be applied to environmental specimens (in particular in soils, sediments, and natural waters) to either enhance the contrast of poorly electron-opaque organic material (salts of heavy elements, amongst which the most commonly employed are uranyl acetate, lead citrate, phosphotungstate, alcian blue), or selectively stain exocellular polymeric substances (multistep reaction with silver proteinate to yield Ag grains onto polysaccharides). It is regrettable to note that to date it is not possible to clearly identify colloidal humic/fulvic entities, which constitute an important proportion of natural organic materials. On the other hand, highly sophisticated labeling techniques by means of ultraspecific markers (e.g., gold-lectin complexes) should prove, in the near future, to be very valuable to distinguish morphologically similar natural organic macromolecules (e.g., neutral versus acidic polysaccharidic moieties produced by bacteria and algae). In many environmental situations, colloidal organic matter is naturally stained by the major ions present in the electrolyte, and thus requires no staining for simple visualization purposes. It must however be noted

that most existing staining protocols were not designed for environmental specimens (whole mounts or resin embedded) and necessitate careful optimization prior to be used on a routine basis.

## Microscopic Investigation

Electron microscopic investigation of environmental samples has been traditionally performed for years for merely illustrative purposes, although the lateral resolution power and the analytical capabilities of modern electron microscopes are in many situations unique features that can bring unequivocal qualitative and quantitative answers to complex problems in which colloids and particles play a central role. For ease, the qualitative and quantitative approaches of investigation are discussed independently in the following sections, but they should be considered as intimately intermingled within the frame of every cleverly planned correlative electron microscopic investigation, as would be the case for the idealized correlative investigation depicted in Figure 3.

### Qualitative/Semiquantitative Investigation

SEM and ESEM are particularly well suited for the evidencing of surface and textural features of large particulate entities ( $c. >1\mu\text{m}$ ), and for the rough estimate of the particle size distribution. Qualitative 3D morphological information is readily extracted from SEM operated either in secondary electron mode or in backscattered electron mode, and the extent of aggregation between particles (provided that it is not an artifact produced by overloaded specimen stubs) can be documented with a resolution down to  $\sim 50\text{nm}$  without difficulty. Because of its ease of operation, SEM should be selected for the routine survey of samples, in particular for atmospheric and soil particles. Preliminary qualitative surveys should help the operator to focus on either the general trends (e.g., major classes of particle types or sizes or associations) or the significant specificities of the sample (e.g., characteristic aggregation between two types of particles, or prevalence of a narrow size class for a given type of entity).

When combined to conventional bulk experiments and analyses, qualitative use of SEM/ESEM may yield significant progress in the understanding of the studied ecosystem. For example, visualization of the microscopic features of humic and fulvic substances subjected to preliminary changes in pH or ionic strength of the surrounding milieu have greatly contributed to the unequivocal appreciation of the conformational changes of these dominant organics in

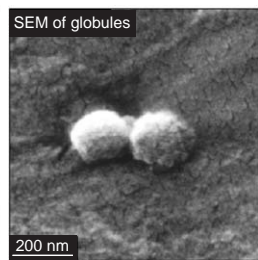
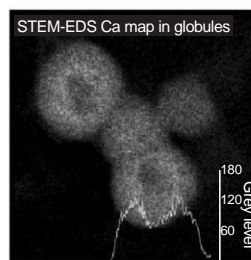
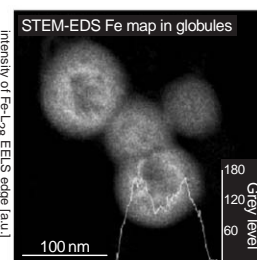
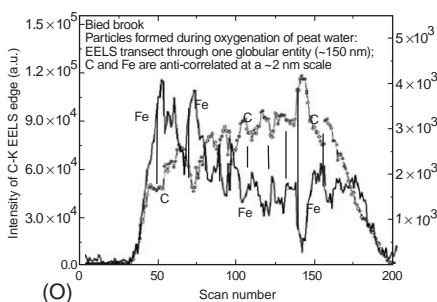
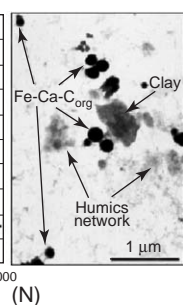
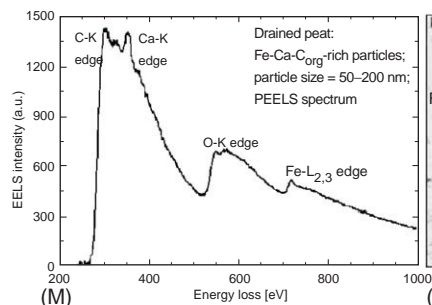
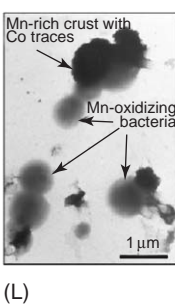
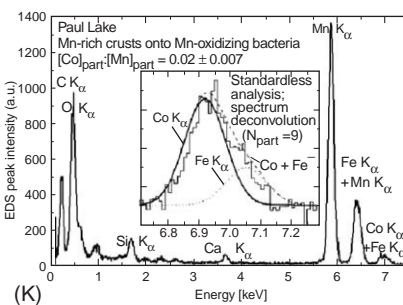
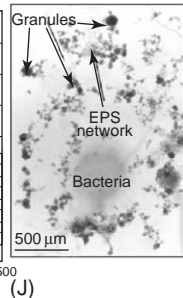
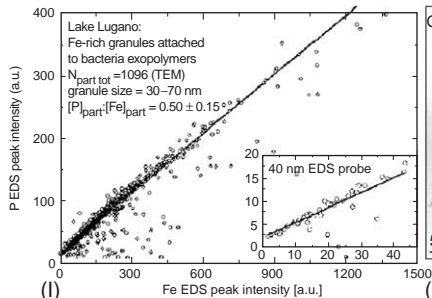
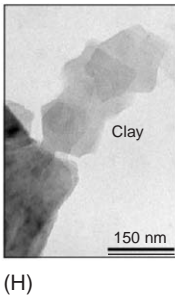
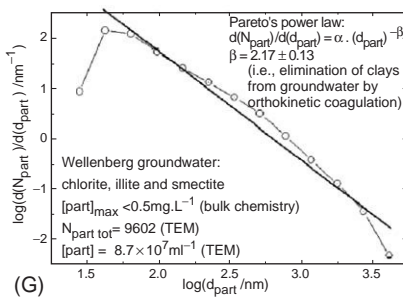
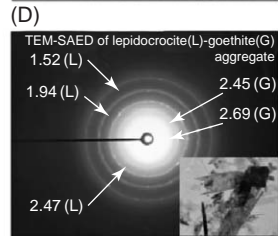
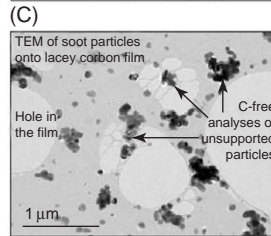
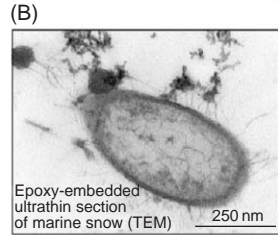
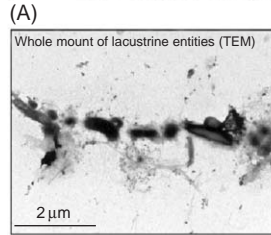
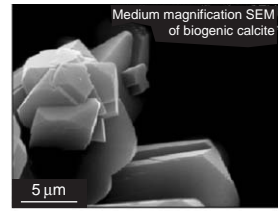
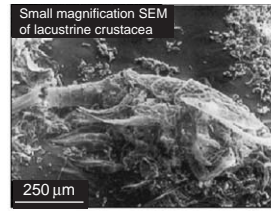
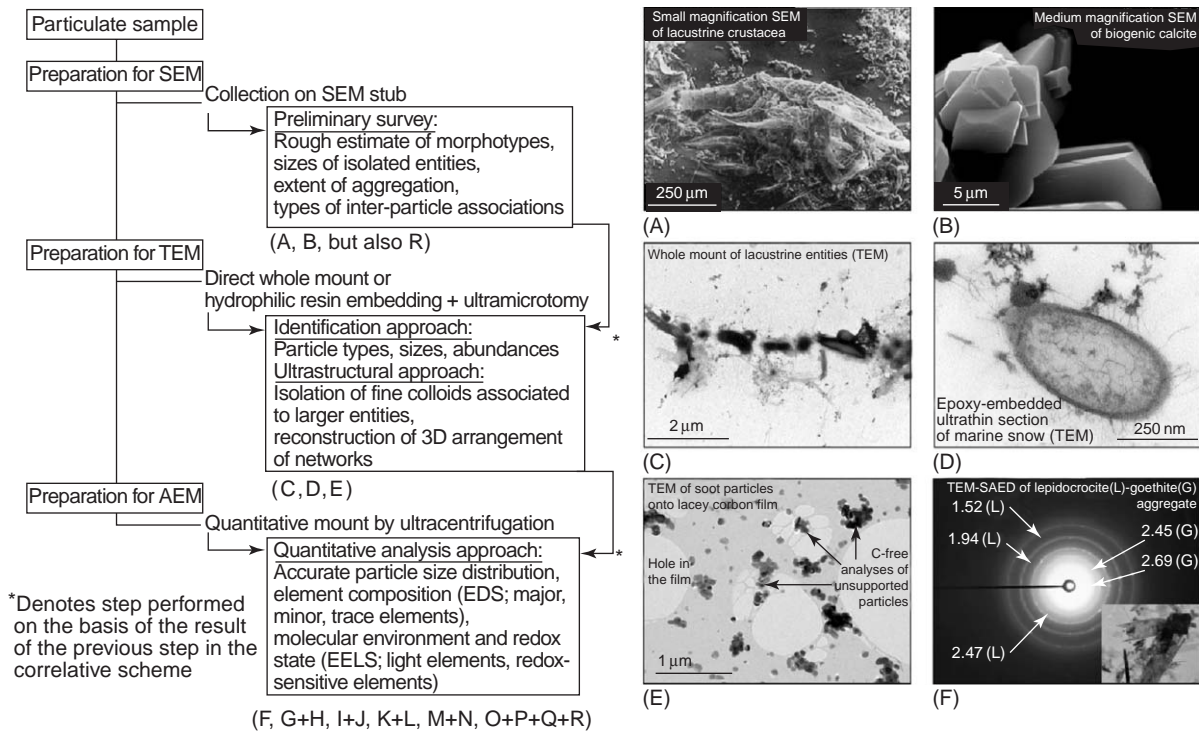
soils, ranging from spherocolloidal entities to flexible and extended fibrils or densely networked 3D systems. Likewise the effect of the intimate associations between clays and fibrillar polysaccharide networks on the stabilization of soil structures has been assessed mainly by means of an SEM. Nonetheless, such investigations require careful specimen preparation techniques (e.g., freeze-drying instead of air drying) aimed at minimizing artifactual conformational changes at every step of the preparation.

Accurate qualitative or semiquantitative investigations on colloidal entities ( $c. <1\mu\text{m}$ ) are better performed by means of a TEM or STEM, with a resolving power down to the nanometer scale, even for complex heterogeneous entities (while a claimed subangstrom resolution is achievable in TEM with aberration corrected electron optics). Whole mounts and quantitative mounts (see Figure 2) are appropriate for semiroutine TEM investigations in the presence of colloidal entities, provided that TEM grids are not too densely populated with large particles and aggregates. In addition, inorganic entities can be checked by qualitative electron diffraction (either in selected area mode or in convergent beam mode) for their potential crystallinity; this is particularly useful for ill-defined particles (e.g., ferrihydrites or partly amorphous oxides) that may reveal locally ordered domains of their atoms, or to distinguish between particles exhibiting the same morphologies and rough composition but different stages of aging.

Ultrastructural TEM characterization of ultrathin resin sections is recommended to reveal, e.g., tiny mineral deposits on bacterial cells, which would be obscured by the thickness-related opacity of bacteria prepared as whole mounts. Likewise, complex large-scale networks of fibrillar exopolymers shall retain their subtle conformation when embedded in the appropriate resin; the reconstruction of their 3D architecture is theoretically achievable by imaging of successive sections, but this requires tedious and extensive image analysis work.

### Quantitative Investigation

Quantitative investigation of natural particulate material necessitates special requirements. In the first place, the specimen must be representative of its original milieu. For example, all particle types initially present in the natural sample must be represented in their right proportions, except if the sample has intentionally undergone controlled fractionation to remove several particle classes. The aggregates present in the specimen must not be artifact expressions of an overloaded TEM grid or of a sample that



would have experienced handling and preparation conditions favoring coagulation of existing entities or precipitation of dissolved species. In that respect, the preparation of quantitative mounts (see **Figure 2**) by direct ultracentrifugation of suspended particles onto TEM grids is certainly the most appropriate approach for quantitative measurements, and most works performed on natural waters, sediments, or soils can be brought to this approach.

Secondly, quantitative investigations need to be performed on a statistically significant number of entities to yield sound measurements. Depending on the expected confidence in the final results, the measurements (size, composition) should be performed on a large set of particles, comprising between 100 and 10 000 entities (see **Figure 3**). When dealing with elemental composition, the measured EDS peak intensities must be calibrated against representative standards; these standards may have to be synthesized in the laboratory under the conditions encountered in the studied ecosystem, because the mechanisms of production of X-rays are influenced by the matrix of the studied material. Standardless analysis is feasible in EDS, but it requires the knowledge of accurate element- and microscope-related parameters. For EELS analysis, relative quantification (i.e., the ratio of one element to another in the particle of interest) is readily obtained by calculation without the need for standards.

Quantitative particle size distributions (either of all types of particles and aggregates present in the specimen, or of a specific type of particle, identified either by morphological criteria or by routine EDS elemental mapping) must be performed on quantitative whole mounts; high-quality micrographs can then be digitized for mathematical morphometry using an image analysis software. The accurate determination of the particle size distribution of atmospheric, aquatic, soil, or sediment entities may yield information on the processes driving their behavior in the milieu (e.g., formation, dissolution, coagulation, sedimentation; see **Figure 3**). It is also possible to estimate the mechanisms of aggregation

of colloids by means of the fractal dimension of aggregates. Provided that micrographs are obtained under carefully controlled conditions of illumination, estimates on the volume of nonspherical particles of known composition can also be obtained by means of image analysis (thickness-dependent opacity of the particles).

Over the past two decades, the clever determination of the composition of nonartifactual environmental particles and colloids by EDS analysis has definitely transformed electron microscopy into a powerful and unique analytical tool. Under optimal conditions, i.e., using an EDS detector equipped with an ultrathin window, even carbon can be quantified, and the detection limits can be as low as  $500\text{--}1000\text{ mg kg}^{-1}$  ( $0.05\text{--}0.1\%$ ), even for submicrometric particles, provided that the trace elements of interest are not overlapped by major or minor elements (see **Figure 3**). The method is particularly well suited for elements emitting  $K\alpha$  peaks in the sensitive  $0\text{--}10\text{ keV}$  energy range, but L-emitting elements can also be measured in this range, provided that they are present as minor or major elements. Quantitative EDS analysis is nowadays the method of choice for the quasiroutine-based detection of trace elements scavenged by particles, or the search for stoichiometric relationships between different elements in similar morphotypes.

Although the ability of EELS techniques (electron spectroscopic imaging, ESI, in imaging mode; EELS in spectrum mode) to detect single atoms with energy resolutions as low as  $0.2\text{ eV}$  in the useful  $0\text{--}1000\text{ eV}$  energy range (in comparison, high-resolution EDS is achieved at  $\sim 150\text{ eV}$ ) has been clearly demonstrated in materials science and even in biological science, EELS is regrettably not commonly used in environmental science. Indeed, the interpretation of spectra is not as clear-cut as in EDS, because (1) the technique is best suited for light elements (although transition metals yield valuable spectral information), (2) the extraction of the EELS K- or L- or even M-edges of elements requires a careful stripping of the substantial background (in the form of a power law

**Figure 3** Idealized correlative electron microscopic procedure for the examination and physicochemical characterization of environmental colloids and particles. Micrographs, spectra, and plots are extracted from published/unpublished work of the authors. (A) Adapted from Jackson TA and Bistricki T (1995) *Journal of Geochemical Exploration* 52: 97–125. (C) and (D) Adapted from Lienemann C-P, Heissenberger A, Leppard GG, and Perret D (1998) *Aquatic Microbial Ecology* 14: 205–213. (E) Adapted from Mavrocordatos D, Kaegi R, and Schmatloch V (2002) *Atmospheric Environment* 36: 5653–5660. (F) Adapted from Mavrocordatos D and Fortin D (2002) *American Mineralogist* 87: 940–946. (G) Adapted from Couture C, Lienemann C-P, Mavrocordatos D, and Perret D (1996) *Chimia* 50: 625–629. (I) and (J) Adapted from Lienemann C-P, Monnerat M, Dominik J, and Perret D (1999) *Aquatic Sciences* 61: 133–149. (K) and (L) Adapted from Lienemann C-P, Taillefert M, Perret D, and Gaillard J-F (1997) *Geochimica Cosmochimica Acta* 61: 1437–1446. (N), (P), and (Q) Adapted from Mavrocordatos D, Mondy-Couture C, Atteia O, Leppard GG, and Perret D (2000) *Journal of Hydrology* 237: 234–247. (O) Adapted from Mondy C, Leifer K, Mavrocordatos D, and Perret D (2002) *Journal of Microscopy* 207: 180–190. All published material reproduced with permission. (B), (H), and (M) Unpublished results.

$I = aE^{-b}$ ), (3) the region below  $\sim 100$  eV (plasmons region) is difficult to model with accuracy, (4) there remain uncertainties in the determination of the cross-sections of the M-edges used for absolute quantification purposes, and (5) EELS is theoretically dedicated to specimens with thicknesses below 20–50 nm. It has, however, been demonstrated that the technique can be used to quantify thicker ( $< 500$  nm) environmental particles, although accuracy drops as specimen thickness increases.

In its simplest expression, EELS can be performed in energy-filtered (EF-TEM) or ESI mode, i.e., to acquire elemental maps, but EELS in spectroscopic mode is best suited for quantitative analysis, with a high energy and lateral resolution.

The main advantage of EELS over EDS is its ability to yield molecular information, in the form of specific features at and beyond the edges (see **Figure 3**). The features referred to as energy loss near edge structure (ELNES; extending up to 50–100 eV beyond the edge; counterpart to X-ray absorption near-edge structure, XANES, in X-ray spectroscopy) are fingerprints that bring qualitative information about the molecular environment of the element giving rise to an EELS edge; for instance, the shape of an EELS-ELNES spectrum is different for aromatic, aliphatic, or amorphous organic carbon centers, as well as for inorganic carbon centers. Likewise, EELS spectra acquired with a high-energy resolution (i.e.,  $< 0.5$ – $1$  eV) may reveal the electronic configuration of redox-sensitive elements (e.g.,  $\text{Fe}^{2+}$  versus  $\text{Fe}^{3+}$ , or their mixtures in the same entity;  $\text{Mn}^{2+}$  versus  $\text{Mn}^{3+}$  versus  $\text{Mn}^{4+}$ ). Although exemplary results have already been obtained on pure, crystalline iron minerals and manganese minerals, the approach is far from routine for the complex heterogeneous particles identified in soils, sediments, and natural waters. Similarly, the use of EELS for the redox discrimination of elements of environmental concern such as  $\text{Cr}^{3+}/\text{Cr}^{6+}$ ,  $\text{Cu}^+/\text{Cu}^{2+}$ ,  $\text{As}^{3-}/\text{As}^{3+}/\text{As}^{5+}$ ,

$\text{Sb}^{3-}/\text{Sb}^{3+}/\text{Sb}^{5+}$ , or  $\text{Se}^{2-}/\text{Se}^{4+}/\text{Se}^{6+}$ , has not been reported yet.

**See also:** **Air Analysis:** Sampling. **Centrifugation:** Analytical Ultracentrifugation; Preparative. **Electron Energy Loss Spectrometry. Microscopy:** Overview. **Microscopy Applications:** Biomedical. **Microscopy Techniques:** Light Microscopy; Scanning Electron Microscopy; Atomic Force and Scanning Tunneling Microscopy.

## Further Reading

- Chen Y (1998) Electron microscopy of soil structure and soil components. In: Huang PM, Senesi N, and Buffle J (eds.) *Structure and Surface Reactions of Soil Particles*. IUPAC Series on Analytical and Physical Chemistry of Environmental Systems, vol. 4, pp. 155–182. New York: Wiley.
- Hawkes PW, Mulvey T, and Kazan B (eds.) (2002) *Advances in Imaging and Electron Physics*. New York: Academic Press.
- Hurd DC and Spencer DW (eds.) (1991) *Marine Particles: Analysis and Characterization*. Geophysical Monograph, vol. 63. Washington, DC: American Geophysical Union.
- Leppard GG (1992) Evaluation of electron microscope techniques for the description of aquatic colloids. In: Buffle J and Van Leeuwen HP (eds.) *Environmental Particles*. IUPAC Series on Analytical and Physical Chemistry of Environmental Systems, vol. 1, pp. 231–289. Ann Arbor: Lewis.
- Smart P and Tovey NK (1981) *Electron Microscopy of Soils and Sediments: Examples*. Oxford: Oxford University Press.
- Smart P and Tovey NK (1982) *Electron Microscopy of Soils and Sediments: Techniques*. Oxford: Oxford University Press.
- Xhoffer C, Wouters L, Artaxo P, Van Put A, and Van Grieken R (1992) Characterization of individual environmental particles by beam techniques. In: Buffle J and Van Leeuwen HP (eds.) *Environmental Particles*. IUPAC Series on Analytical and Physical Chemistry of Environmental Systems, vol. 1, pp. 107–143. Ann Arbor: Lewis.

## Liquid Crystals

**J W Goodby**, University of Hull, Hull, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Liquid Crystal Phases

A liquid crystal mesophase is a state of matter intermediate between a solid and a liquid. Strictly, the liquid crystal or mesomorphic state is made up of a

number of orientationally ordered phases that occur between the breakdown of positional ordering of the molecules in a solid and the formation of the amorphous liquid. The breakdown in ordering on passing from a crystal to a liquid can be achieved either by changing the temperature of the material or by dissolving it in a liquid. Liquid crystals formed on heating and/or cooling processes are called thermotropic liquid crystals, and those formed when a material



$I = aE^{-b}$ ), (3) the region below  $\sim 100$  eV (plasmons region) is difficult to model with accuracy, (4) there remain uncertainties in the determination of the cross-sections of the M-edges used for absolute quantification purposes, and (5) EELS is theoretically dedicated to specimens with thicknesses below 20–50 nm. It has, however, been demonstrated that the technique can be used to quantify thicker ( $< 500$  nm) environmental particles, although accuracy drops as specimen thickness increases.

In its simplest expression, EELS can be performed in energy-filtered (EF-TEM) or ESI mode, i.e., to acquire elemental maps, but EELS in spectroscopic mode is best suited for quantitative analysis, with a high energy and lateral resolution.

The main advantage of EELS over EDS is its ability to yield molecular information, in the form of specific features at and beyond the edges (see **Figure 3**). The features referred to as energy loss near edge structure (ELNES; extending up to 50–100 eV beyond the edge; counterpart to X-ray absorption near-edge structure, XANES, in X-ray spectroscopy) are fingerprints that bring qualitative information about the molecular environment of the element giving rise to an EELS edge; for instance, the shape of an EELS-ELNES spectrum is different for aromatic, aliphatic, or amorphous organic carbon centers, as well as for inorganic carbon centers. Likewise, EELS spectra acquired with a high-energy resolution (i.e.,  $< 0.5$ – $1$  eV) may reveal the electronic configuration of redox-sensitive elements (e.g.,  $\text{Fe}^{2+}$  versus  $\text{Fe}^{3+}$ , or their mixtures in the same entity;  $\text{Mn}^{2+}$  versus  $\text{Mn}^{3+}$  versus  $\text{Mn}^{4+}$ ). Although exemplary results have already been obtained on pure, crystalline iron minerals and manganese minerals, the approach is far from routine for the complex heterogeneous particles identified in soils, sediments, and natural waters. Similarly, the use of EELS for the redox discrimination of elements of environmental concern such as  $\text{Cr}^{3+}/\text{Cr}^{6+}$ ,  $\text{Cu}^+/\text{Cu}^{2+}$ ,  $\text{As}^{3-}/\text{As}^{3+}/\text{As}^{5+}$ ,

$\text{Sb}^{3-}/\text{Sb}^{3+}/\text{Sb}^{5+}$ , or  $\text{Se}^{2-}/\text{Se}^{4+}/\text{Se}^{6+}$ , has not been reported yet.

**See also:** **Air Analysis:** Sampling. **Centrifugation:** Analytical Ultracentrifugation; Preparative. **Electron Energy Loss Spectrometry. Microscopy:** Overview. **Microscopy Applications:** Biomedical. **Microscopy Techniques:** Light Microscopy; Scanning Electron Microscopy; Atomic Force and Scanning Tunneling Microscopy.

## Further Reading

- Chen Y (1998) Electron microscopy of soil structure and soil components. In: Huang PM, Senesi N, and Buffle J (eds.) *Structure and Surface Reactions of Soil Particles*. IUPAC Series on Analytical and Physical Chemistry of Environmental Systems, vol. 4, pp. 155–182. New York: Wiley.
- Hawkes PW, Mulvey T, and Kazan B (eds.) (2002) *Advances in Imaging and Electron Physics*. New York: Academic Press.
- Hurd DC and Spencer DW (eds.) (1991) *Marine Particles: Analysis and Characterization*. Geophysical Monograph, vol. 63. Washington, DC: American Geophysical Union.
- Leppard GG (1992) Evaluation of electron microscope techniques for the description of aquatic colloids. In: Buffle J and Van Leeuwen HP (eds.) *Environmental Particles*. IUPAC Series on Analytical and Physical Chemistry of Environmental Systems, vol. 1, pp. 231–289. Ann Arbor: Lewis.
- Smart P and Tovey NK (1981) *Electron Microscopy of Soils and Sediments: Examples*. Oxford: Oxford University Press.
- Smart P and Tovey NK (1982) *Electron Microscopy of Soils and Sediments: Techniques*. Oxford: Oxford University Press.
- Xhoffer C, Wouters L, Artaxo P, Van Put A, and Van Grieken R (1992) Characterization of individual environmental particles by beam techniques. In: Buffle J and Van Leeuwen HP (eds.) *Environmental Particles*. IUPAC Series on Analytical and Physical Chemistry of Environmental Systems, vol. 1, pp. 107–143. Ann Arbor: Lewis.

## Liquid Crystals

**J W Goodby**, University of Hull, Hull, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Liquid Crystal Phases

A liquid crystal mesophase is a state of matter intermediate between a solid and a liquid. Strictly, the liquid crystal or mesomorphic state is made up of a

number of orientationally ordered phases that occur between the breakdown of positional ordering of the molecules in a solid and the formation of the amorphous liquid. The breakdown in ordering on passing from a crystal to a liquid can be achieved either by changing the temperature of the material or by dissolving it in a liquid. Liquid crystals formed on heating and/or cooling processes are called thermotropic liquid crystals, and those formed when a material

dissolves in a solvent are called lyotropic liquid crystals. In principle, liquid crystal phases have come to be defined as modifications where the molecules are orientationally ordered but yet are still in dynamic motion. Consequently, this definition also includes soft crystals, where the molecules are rotationally disordered.

Thus, in a classical melting process of a solid to a liquid, the lattice of the solid collapses to give the disorganized liquid where the molecules tumble and rotate freely. At the melt point the molecules undergo large and rapid simultaneous changes in rotational, positional, and orientational order. In melting processes mediated by thermotropic liquid-crystalline behavior, however, there is a stepwise breakdown in order. The incremental steps of this breakdown produce a variety of thermodynamically stable, intermediate states known as mesophases. As liquid crystals have fluid-like structures, the defects that are formed in bulk specimens occur on a large enough scale for the mesophases to be investigated by optical microscopic methods. Similarly, when a lyotropic liquid crystal dissolves in a solvent the crystal state will form the solution via a number of intermediate lyotropic phases. The breakdown in this order can also be charted by polarized light microscopy.

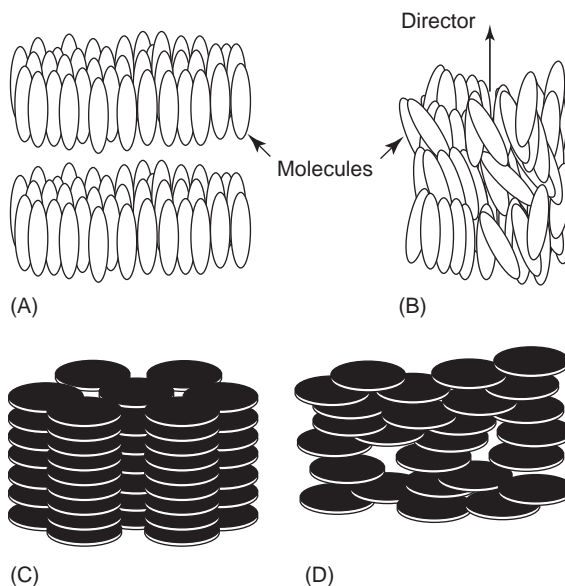
### Thermotropic Liquid Crystals

Like other states of matter, thermotropic mesophases are indefinitely stable at defined temperatures and pressures. Moreover, a thermotropic liquid-crystalline material exhibits reversible phase transitions at well-defined temperatures. For example, the liquid crystal 4-n-pentyl-4'-cyanobiphenyl (5CB) melts from the solid to a nematic liquid crystal at 22.5°C and then from the nematic phase to the liquid at 35.0°C. As a consequence, the characterization and classification of thermotropic phases by microscopy also requires the use of an accurately controlled oven.

The way, or sequence, in which thermotropic transitions occur is defined in the following ways. The liquid crystal to isotropic liquid transition is called the clearing or isotropization point, and this transition, like those between liquid crystal phases, is essentially reversible and occurs with little hysteresis in temperature. The melting point of a material is usually a constant, but the recrystallization process can be subject to supercooling. Mesophases formed on the first heating cycle of a material are thermodynamically stable, and are called enantiotropic phases, whereas phases that are formed below the melt point on cooling cycles, and are revealed

because of the supercooling of the crystal state, are metastable and termed monotropic phases. Both types of phase can be investigated by microscopy.

Thermotropic mesophases are split into two groups as follows. One in which the molecules have rod-like shapes – called calamitic liquid crystals and the other where they have disk-like shapes – called discotic or columnar liquid crystals. Each of these classes can be further subdivided, for example, calamitic mesophases can be split into three groups: disordered or anisotropic plastic crystals, where the molecules have long-range translational order but exhibit rapid dynamic motion; smectic phases, in which the molecules have short-range positional order yet retain layer ordering; and nematic phases, where the molecules are orientationally ordered so that their long axes are on average parallel to one another, this direction is called the director. Again, the molecules exhibit rapid and diffuse molecular motion, principally about their long axes. This motion is related to the Brownian motion observed by microscopy. Examples of the structures of some thermotropic liquid crystals are shown schematically in Figure 1.



**Figure 1** General structures of calamitic and discotic thermotropic liquid crystals. (A) Layered calamitic smectic liquid crystal. The structures of the various types depend on the local packing of the molecules, the extent of the packing, and the orientation of the long axes with respect to the layers. (B) Calamitic nematic liquid crystal. The molecules have no long-range order, and are only orientationally ordered. (C) Ordered columnar discotic liquid crystal. Disk-like molecules form ordered or disordered columns; different column packings give rise to various mesophase structures. (D) Nematic-discotic liquid crystal phase. The disk-like molecules are only orientationally ordered.

### Lytotropic Liquid Crystals

Lytotropic liquid crystals are formed in mixtures of amphiphiles (e.g., surfactants) and solvents, for example, detergents and water. Consequently, these phases are thermodynamically stable at defined temperatures, pressures, and concentrations. Like thermotropics, a variety of structurally distinct modifications exist, which are collectively known as lyotropic liquid crystals.

The quintessential structure of a typical amphiphilic molecule is one where a hydrophilic polar head group is attached to a hydrophobic nonpolar aliphatic moiety. When a lyotropic mesogen dissolves in a solvent the mesophases formed depend on the size of the head group, the number of aliphatic chains present in the amphiphile, and the polarity of the solvent. The phase structures are dependent on the concentration and the degree of curvature produced by the packing arrangements of the amphiphilic molecules, i.e., the ratio of the size of the head group to the aliphatic chain is important. For example, at low curvatures lamellar phases are formed. Increased curvature produces hexagonal phases where the surfactant molecules form columnar structures that pack in hexagonal arrays. Further increase in curvature results in the formation of micellar structures. Typically, when an amphiphile is solvated in a polar solvent, the polar portion of the molecule will interact strongly with the solvent to give a normal micellar structure. Alternatively, in a nonpolar solvent the nonpolar portions of the amphiphile will interact more strongly with the solvent, causing the polar groups to aggregate, thereby producing reversed micellar structures.

### Defects and Textures in Liquid Crystals

Unlike the crystal state, where there are 230 possible space groups, so far only  $\sim 40$  liquid crystal phases have been discovered. Similarly to the way the structure of a crystal can be classified through its birefringent properties seen in the polarized-light microscope, liquid crystal phases (and any accompanying phase sequences) can be classified by thermal polarized-light microscopy. In the case of liquid crystals, however, mesophases are identified from their defect textures, and not from their shapes and/or indexing of crystal faces as is the case for crystals. Moreover, the many defect textures formed by mesophases are dependent on the local ordering of the molecules, and so mesophase structure can be inferred from the analyses of their textures. Thus, it is important to associate microscopic observations with the physical processes and structural changes taking place when a solid melts or dissolves.

As liquid crystals have essentially only weak ordering of the molecules, their textures occur over the whole area of the specimen and the defects that are present appear on a fairly large scale – from about a micrometer or less to hundreds of micrometers in size. This size regime makes the polarizing light microscope an ideal tool for their investigation and classification. The use of a microfurnace, which has viewing ports perpendicular to the direction of the light beam, has made thermal optical microscopy one of the most powerful tools used in the investigation of mesophase morphology for both thermotropic and lyotropic liquid crystals.

The study of defects in liquid crystal systems is rooted in the understanding of defects in the solid state. For instance, crystals are rarely perfect and usually contain a variety of defects, e.g., point defects, line defects, or dislocations, and planar defects such as grain boundaries. In addition to these typical imperfections of the solid state, liquid crystals can also exhibit defects known as disclinations. These defects are not usually found in solids and result from the fact that mesophases have liquid-like structures that can give rise to continuous but sharp changes in the orientations of the molecules, i.e., sharp changes in orientation occur in the director field.

When defects are formed in liquid-crystalline systems the internal energy,  $E$ , rises; however, the entropy,  $S$ , of the system also increases, resulting in a reduction of the free energy. The Gibbs free energy,  $G$ , is related to the internal energy and entropy:

$$G = E - TS \quad [1]$$

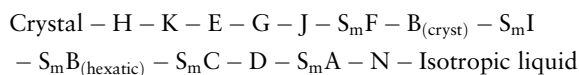
and as the number of defects rises the entropy becomes larger, resulting in a decrease in the free energy. Thus, the presence of defects tends to stabilize the system. In practical applications of liquid crystals, e.g., display devices, the presence of defects is very undesirable; consequently, many investigations have been made into their structures and growth, and the information obtained has been used to minimize their presence in applications. Consequently, through alignment techniques it is now possible to create a large area of nematic liquid crystal that is essentially a defect-free monodomain.

### Polymorphism and Paramorphic Textures

A thermotropic liquid crystal is capable of polymorphism, the situation in which a material exhibits more than one liquid-crystalline modification between the solid and liquid phases. For instance, when a calamitic liquid crystal melts it can do so via transitions between a number of smectic phases and the nematic phase. For the purposes of phase

identification by microscopy it is important to consider the nature of the polymorphism and its effects on defect textures. On heating or cooling cycles the imprint of the preceding phase is usually imparted to the phase under observation. Thus, when a crystal melts to a smectic phase, the defect texture produced will bear the imprint of the preceding crystal phase; this is called a paramorphotic texture. In order to obtain the best textures for the purposes of 'reading' them for phase identification, it is important to start by cooling the liquid into the liquid crystal state. The amorphous liquid has no long-range ordering of the molecules, and so the liquid crystal phase formed first will be obtained in its natural texture. Subsequent transitions to other liquid crystal phases will result in them exhibiting paramorphotic textures based on the natural texture of the mesophase formed first. If the reverse process is used, i.e., heating the crystal, then the textures will be based on the defects present in the solid, and as noted above these will be relatively small-scale defects and hard to see and categorize by microscopy. Similar approaches must also be taken with lyotropic, polymeric, discotic, and chiral liquid crystals.

Another useful tool used in conjunction with identification of liquid crystals by microscopy is a listing of the thermodynamic ordering of mesophases. For calamitic smectic phases this ordering is:



where single capital letters  $\text{S}_m$  denote a smectic-like crystal phase where the molecules have long-range order, phases denoted by  $\text{S}_m$  followed by a capital letter are smectic phases where the molecules are arranged in layers and have short-range order, N denotes the nematic phase, which has no layer ordering; and D is a cubic phase. It is possible to use this thermodynamic ordering to eliminate phases in the process of identification. For instance, if a material cools from the liquid to a smectic I phase, then it is safe to say that any phases formed upon subsequent cooling cannot be classed as nematic, smectics A or C, or cubic D because they all occur at higher temperatures than the smectic I phase. Similar thermodynamic sequences are found for discotic and lyotropic liquid crystals, and for variations in the structures of lyotropic phases with respect to concentration.

## Sample Orientation

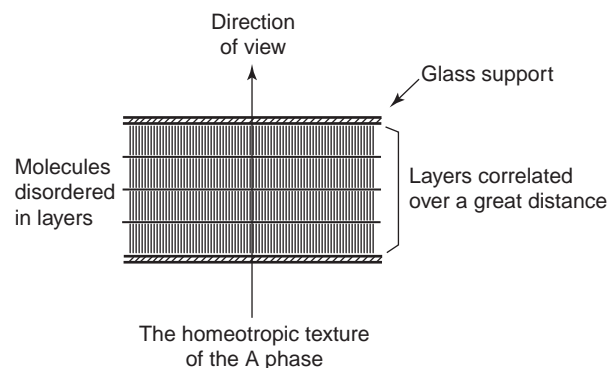
There are various forms of sample preparation of liquid crystals for optical microscopy, and it is

important that as many are used as possible in phase identification. Moreover, it is essential to recognize that some textures complement one another, thereby giving a complete form of information that is not possible to obtain from isolated textures or specimens. There are essentially two main forms of sample preparation: one where the bulk phase is oriented in one of two ways with respect to the glass microscope slides, and the other where the molecules are aligned over a large area. In the first case textures are obtained that possess a large number of defects, which is useful in phase identification. In the second the number of defects is far fewer as the director of the phase does not vary much in its direction throughout the sample, which is important for physical studies, such as electric field investigations. Both of these options are discussed in the following sections.

## Homeotropic Textures

A homeotropic texture is generally defined as one where the optic axis of the mesophase is perpendicular to the glass substrates, and therefore parallel to the light beam of the microscope. More precisely this refers to an optically uniaxial phase where the optic axis is parallel to the light beam, and therefore under crossed polars the texture appears black or optically extinct. Closely related to this form of sample preparation and bulk orientation are structures for smectic phases where the layers of molecules are parallel to the surfaces of the glass slides but the long axes of molecules may be tilted. In the case of the smectic A phase the molecules are arranged in layers with their long axes perpendicular to the glass; thus, the phase is optically uniaxial and exhibits an optically extinct homeotropic texture (see Figure 2).

A similar arrangement of the molecules can be obtained for the E phase, the only difference being



**Figure 2** Side elevation of a specimen of a smectic A phase that exhibits a homeotropic texture. The molecules, shown as rods, are arranged in more disorganized layers than those shown. (Data from Gray and Goodby, 1984.)

that the molecules have long-range order and are packed in an orthorhombic array. Consequently, the E phase is optically biaxial and as such does not exhibit an optically extinct homeotropic texture, but instead gives a mosaic/platelet texture. Sometimes without concession to rigor, therefore, a texture where the layers of a smectic phase are parallel to the glass substrates is referred to as being 'homeotropically oriented'. Typically, however, the nematic, smectics A and B(hexatic), and crystal B phases for rod-like molecules, and the columnar hexagonal phases for disk-like molecules exhibit homeotropic textures.

### Homogeneous Textures

In the homogeneous texture the molecules are arranged so that their optic axes are at an angle to the path of the light beam. In most cases this is when the optic axes are at right angles to the light beam and parallel to the glass surfaces. This arrangement is also called a planar orientation. Although the director in the planar texture lies parallel to the glass, it can wander in the two-dimensional preparation so that the sample is nonuniform. Thus, in a homogeneous texture of the nematic phase, the molecules are organized in such a way that their long axes are on average parallel to the glass; however, throughout the specimen the long axes need not necessarily point in roughly the same direction.

Both of these textures, planar and homogeneous, can also be obtained in aligned specimens. Alignment essentially refers to the use of a surface coating (see later) that is used to order the molecules preferentially in a certain direction relative to the light beam and the glass surfaces. One major addition to the situation described above is that by manipulating the surface of the aligning agent, e.g., by buffing, the liquid crystal molecules can be oriented over the whole area of the specimen. This is unimportant for homeotropically oriented specimens; however, in relation to homogeneous or planar preparations the degeneracy in the plane of the glass is broken and the molecules tend to point in approximately the same direction.

### Sample Preparation

In its simplest form, the microscopic examination of a liquid crystal involves the use of a polarizing light microscope fitted with long-working-distance objectives and condenser. The long working distance, typically  $\sim 0.5$  cm, is required so that a microfurnace can be placed on the microscope stage. It is also useful to have a selection of waveplates available, e.g., a quarter-waveplate, that can be inserted into the

microscope above the objective but below the upper polarizer, for use in conoscopic studies. Typically, a magnification of  $\sim 100\times$  is required for studies to be made; for example, eyepieces of magnification  $10\times$  and an objective with a magnification between  $10\times$  and  $20\times$  is the most suitable combination. It is important that higher magnifications, e.g.,  $400\times$ , are not used for straightforward observations because the defects can become confusing at this level of magnification, leading to misidentification. However, for conoscopic work magnifications of this order are suitable.

A sample of a thermotropic liquid crystal can be prepared by placing a small amount of the material between a clean glass slide and a coverslip, and inserting it into the oven of the microscope. For a lyotropic liquid crystal either the compound can be dissolved in a solvent directly on the microscope stage at room temperature whilst under microscopic observation, or else a solution of known concentration can be made up and sealed in a glass cell, which can then be inserted into the oven. In each of these cases defect textures will be observed. Transition temperatures can be determined by heating or cooling the sample. Typically heating/cooling rates of  $0.1^\circ\text{C min}^{-1}$  are used to determine these values. The reversibility and sharpness of the liquid crystal phase transitions are extremely sensitive to the quality or purity of the material. A pure liquid crystal should show reversible transitions, within a degree, for the heating and cooling cycles.

### Homeotropic Alignment

Homeotropic alignment can be achieved in a variety of ways, for example, by using very clean glass, by using a surface aligning agent, or by using free-standing film techniques. Surface treatments include using materials that have a strongly polar surface interaction but that can be removed from the glass by washing, e.g., lecithin, or aligning materials that can be chemically bound to the glass, e.g., octyldimethylchlorosilane. The surface of the slide is usually treated with a very dilute solution ( $\sim 1\%$ ) of the aligning material in a solvent that will completely wet the glass. The slide is then dipped into the solution and the excess solution and solvent are removed to leave a monolayer of the aligning agent on the glass. Cooling the liquid crystal from the liquid on such a surface will cause the molecules to become aligned with the long aliphatic chains of the surfactant, thereby causing the liquid crystal to stand up on the glass to yield a homeotropic texture.

A homeotropic texture can also be obtained for smectic phases by creating a freely suspended film of



the liquid crystal where the planes of the layers in the phase are parallel to a supporting glass or metal plate. This preparation can be obtained by drawing a liquid crystal, in either its smectic A or smectic C phase, across a small hole (1–2 mm) drilled in a glass or metal plate while the sample is in the oven. This technique creates specimens of extremely good quality with well-defined numbers of layers that can also be used as samples for X-ray analysis.

### Homogeneous Alignment

Homogeneous alignment can be achieved in a variety of ways, for example, from simply using glass slides that have been touched by human hand to using surfaces on which SiO may have been obliquely evaporated. For microscopic observations slightly unclean glass usually works well, as do surface preparations using aligning agents such as poly(vinyl alcohol) (PVA) or nylon. Surface aligning agents have the advantage that they can be unidirectionally buffed to give unidirectionally aligned specimens that can be used in a variety of physical studies. However, for the purposes of phase identification, samples that show a variety of defects are of more practical use.

Nylon or PVA can be coated on to the surface of slides by dissolving a small amount of the polymer (~1%) in a suitable solvent and then dipping the glass slide into the solution or by spin coating the solution directly on to the glass. The excess solution is drained off and the remaining solvent evaporated. Upon drying, the slide can either be used 'as is' to give an unaligned sample, or it can be unidirectionally buffed to produce an aligned specimen. When a liquid crystal is cooled from its isotropic liquid phase on such a surface it will form a preparation in which the long axes of the molecules lie in the plane parallel to the support.

### Shearing Techniques

One good way in which to probe a liquid crystal specimen is to subject it to mechanical shearing while in the microscope oven. This can be done by inserting a metal instrument into the microscope oven and dislodging the upper glass plate of the sample. Furthermore, pressurizing the specimen from above, thereby decreasing its thickness, results in the sample becoming more homeotropic in nature; shearing from the side can result in the conversion of a homeotropic texture to a homogeneous form. In the special case of the cholesteric phase, shearing the homogeneous texture exclusively results in the formation of a Grandjean plane texture, i.e., the helical axes of the structure of the mesophase are converted from parallel or at various angles to the glass to being perpendicular.

## Applications in the Identification of Liquid Crystal Phases

In the following sections some of the more common liquid crystal textures utilized in the classification of thermotropic calamitic mesophases are described. It should be noted that similar examinations of defects are used to classify discotic, polymeric, and lyotropic phases.

### Schlieren Textures

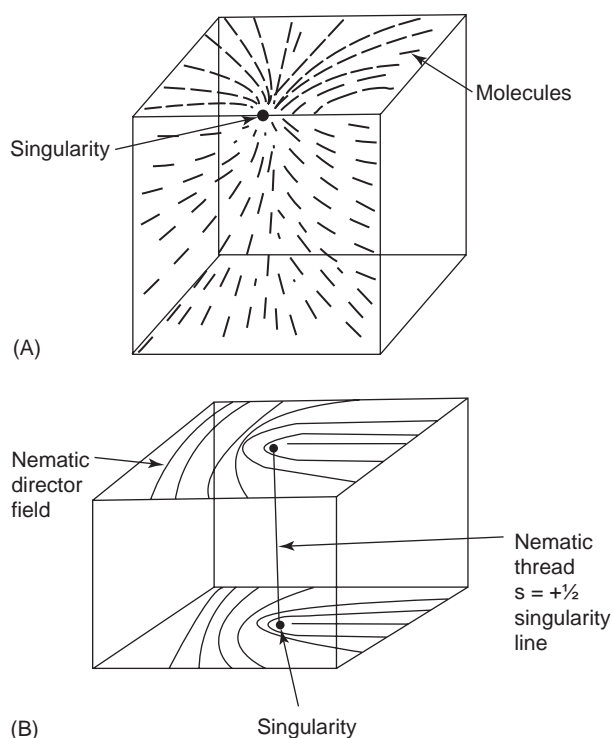
'Schlieren' textures are very common in liquid crystals and are usually easily identified and classified. These patterns are observed in homogeneously oriented nematic phases and for smectic phases where the long axes of the molecules are tilted with respect to the layer planes and the layers are arranged parallel to the surface of the glass. Thus, 'schlieren' textures occur for the nematic phases of both calamitic and discotic materials, and for the tilted smectic phases C, I, and F.

Between crossed polars these defects appear as dark lines or brushes with curved or irregular shapes that correspond to extinction positions of the director and molecular long axes. Thus, the director can be either parallel or perpendicular to the polarizer and analyzer. The brushes tend to cover the specimen in rather a continuous way, indicating the liquid-like nature of the mesophase. The points where the brushes meet are called singularities in the texture (see Figure 3A). For nematic phases two forms of schlieren defect are found, one where two brushes meet at a point and one where four brushes meet. All tilted smectic phases (C, I, F, and ferroelectric C), except for the antiferroelectric phase, exhibit four brush singularities. Therefore, this provides a simple way of distinguishing between smectic and nematic phases. It should be noted that phases such as smectics A and B(hexatic) and crystal phases B(crystal), E, G, H, J, and K do not exhibit schlieren textures and so this narrows down the possibilities for phase identification.

The number of brushes that emanate from point singularities (where  $s$  denotes the strength of the singularity) in schlieren textures are classified as:

$$s = \pm \frac{\text{number of brushes}}{4} \quad [2]$$

where the number of brushes corresponds to the number of dark lines emanating from a point defect in the texture. The sign of the defect ( $\pm$ ) is given by the direction of rotation of the brushes when the polarizer and analyzer are rotated together, i.e., rotation of the brush in the same direction as the rotation of the polarizer and analyzer is defined as

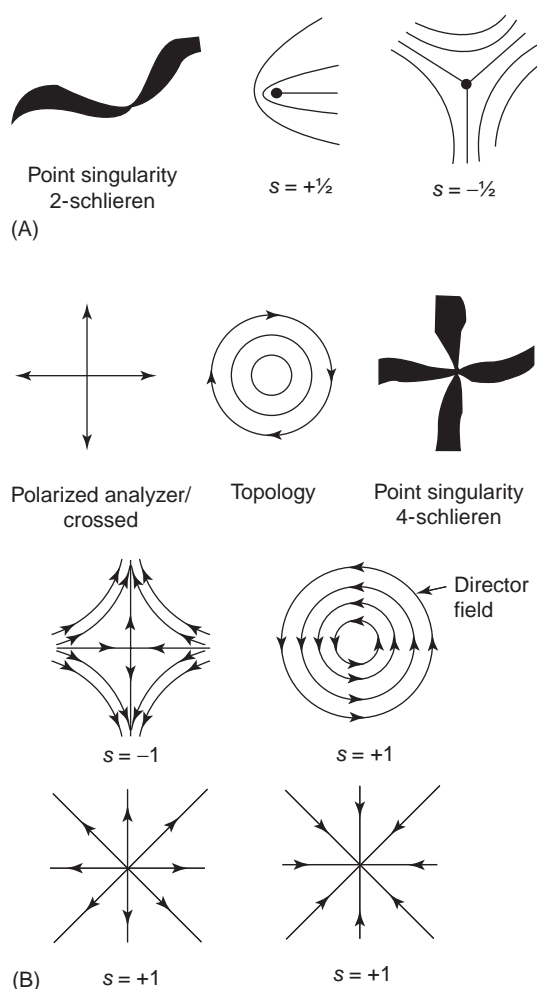


**Figure 3** Singularities in the schlieren textures of nematic phases: (A) a point singularity (the rod-like molecules are shown as short lines), and (B) a nematic thread or  $\pi$  disclination joining two singularities on the surfaces of the glass (the orientation of the director field is shown by the black curved lines).

being positive. Thus, defects with  $s = +1/2$ ,  $-1/2$ ,  $+1$ , and  $-1$  are possible, but if the  $s$  values are added up over the whole sample the resultant value will be approximately zero. Furthermore, neighboring singularities that are connected by brushes tend to have opposite signs and therefore will attract one another. They will annihilate if their respective  $s$  values add up to zero.

Thus, a singularity is a point where the director field abruptly changes, i.e., a disclination occurs in the texture. **Figure 4** shows a variety of different ways in which the director can change about points to give schlieren patterns, examples of which are shown in **Figure 5A**. For nematic phases the lines shown represent the average direction of the long axes of the molecules; however, for smectic phases the lines are associated with the direction of the tilts of the molecules.

It is also possible to have points with  $s = \pm 1/2$  joined by line singularities in the nematic phase: these  $\pi$  disclinations, which are commonly known as threads, pass through the preparation almost perpendicularly with the ends attached to the glass surfaces. **Figure 3B** shows the topology about  $s = +1/2$  singularity line; the end appears as a point on the

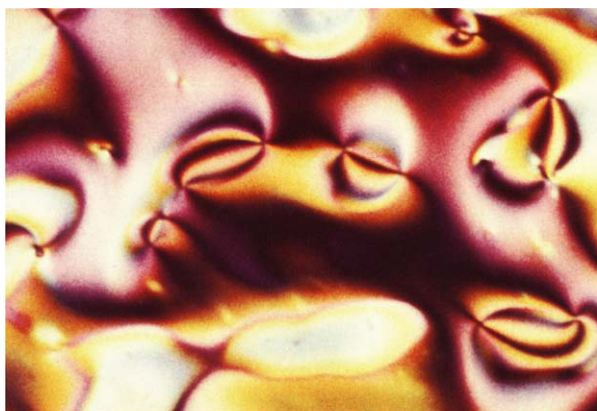


**Figure 4** Schlieren textures in calamitic thermotropic liquid crystals: (A) two brushed singularities seen exclusively in the nematic phase, and (B) four brushed singularities seen for both smectic and nematic phases. The dark lines show the orientation of the director. (Data from Gray and Goodby, 1984.)

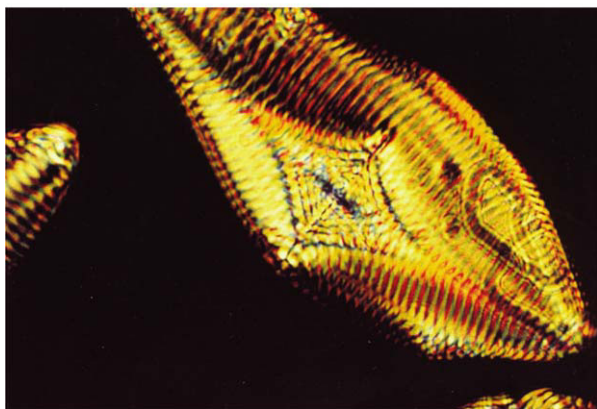
glass as  $s = +1/2$ . In the figure the director field is drawn on the upper and lower glass surfaces. Similarly, for two points joined together that have  $s = \pm 1$  topologies, the resulting line singularity will be a  $2\pi$  disclination. The parameter  $s$  can also be considered as related to the strength of the disclination, and  $s \times 2\pi$  gives the angle by which the director rotates about a closed circuit around the center of the defect.

### Focal-Conic Textures

Focal-conic textures are patterns that are usually observed for smectic phases where the layers are allowed to form curved structures. Such mesophases must therefore have only short-range periodic order, as long-range ordering serves to destabilize curved structures. Typically, focal-conic domains



(A)

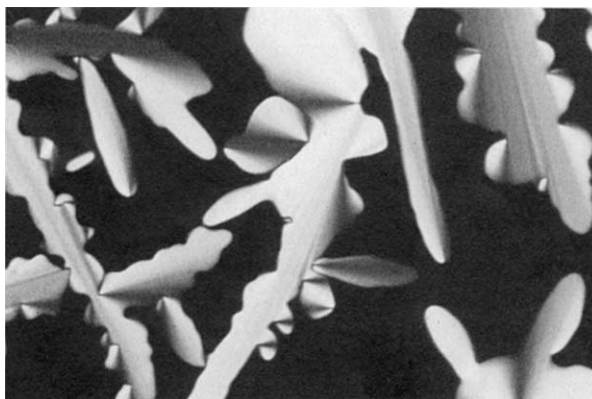


(B)

**Figure 5** Photomicrographs of liquid crystal textures seen in the polarizing microscope. (A) The Schlieren texture of a smectic I phase. (B) The focal conic texture of a chiral smectic C\* phase, which has a helical structure, forming at a transition from the liquid. The pitch of the helix shows up as parallel lines that are parallel to the molecule layers (each line corresponds to about a thousand molecular layers). The pitch lines reflect accurately the layer structure in the focal-conic domain. ((A) Courtesy of JW Goodby, University of Hull, UK, with permission. (B) Reproduced with permission from Gordon and Breach, Switzerland.)

are observed in homogeneously oriented samples where the director is nonuniform in the plane of the preparation. This texture is also the natural texture of the smectic A and smectic C phases; consequently, smectic phases formed on cooling either of these modifications will exhibit paramorphic textures based on, but not necessarily true, focal-conic defects.

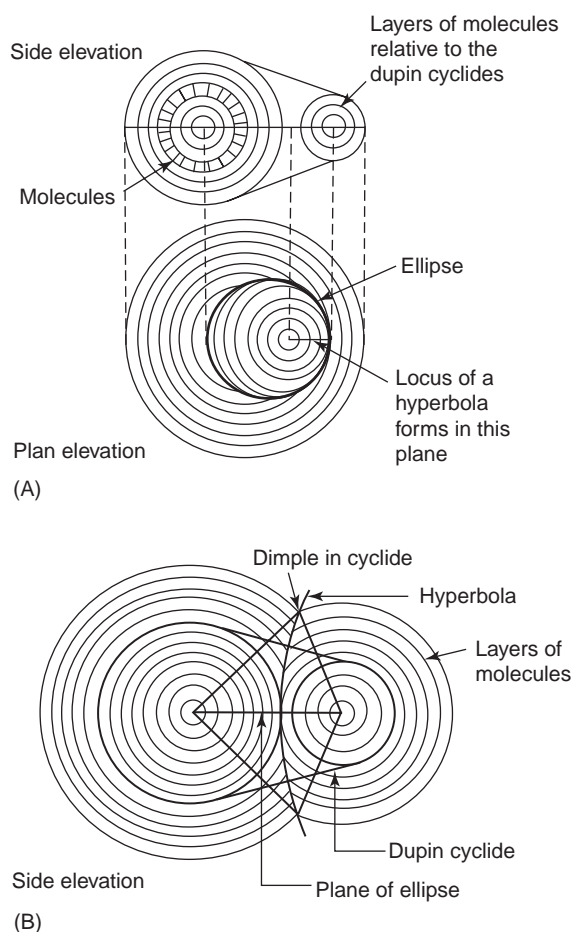
Curvature of the layers in the smectic state is compatible with the formation of cylindrical, tore and Dupin cyclide defects. The most common defect is the Dupin cyclide and therefore is discussed in more detail. When the smectic A phase nucleates from either the liquid or the nematic phase, it can do so with the formation of curved structures called *bâtonnets* (see **Figure 6**), i.e., layers of molecules add to a



**Figure 6** Photomicrograph of *bâtonnets* forming on cooling from the isotropic liquid into a smectic A phase, seen in the polarizing microscope (100 $\times$ ).

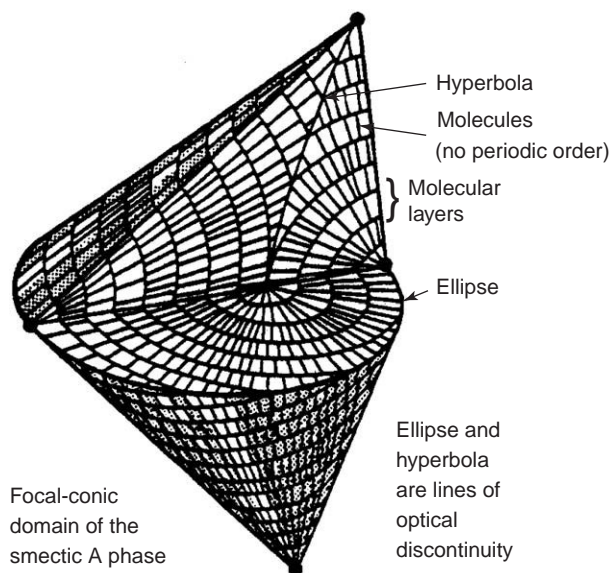
nucleation site almost in the fashion of an onion skin. For a cylindrical structure the growing phase will resemble a jellyroll. However, when the nucleating jellyroll loops back on itself to form a ring, a situation will be produced where the cylindrical layers will form a set of concentric tubes like a number of innertubes one growing on the outside of another, i.e., tores. When the circular tubes grow at a disproportionate rate on one side relative to the other, a lop-sided circular tube, or a Dupin cyclide, is formed. **Figure 7** shows the structure of a focal-conic domain; the side elevation shows the concentric tubes with one side growing at a faster rate than the other. As layers are added they will eventually meet and fill in the hole at the center of the ring. As more layers are added they will cover the outer surface but a dimple will still remain at the point where the layers of the two sides meet. The locus of the dimple is described by a hyperbola. An ellipse is similarly found to describe the growth of the layers in the plan or horizontal section of the defect as shown. Thus, the hyperbola forms the locus of the apices of cones of revolution of which the ellipse is a common section.

**Figure 8** shows a section through a focal-conic domain. On the surface of the domain the long axes of the molecules are radially oriented with respect to the apex of the conical domain, but inside the layers curve so that there is a mismatch of the long axes along the hyperbola. Along the edge of the ellipse the molecules are also radially distributed; therefore, the hyperbola and ellipse show up as lines of optical discontinuity in the polarizing microscope, i.e., they appear as black crosses. **Figure 9** shows a typical texture of focal-conic defects in the smectic A phase. **Figure 5B** shows how the layers are highlighted by in a helical smectic C mesophase.

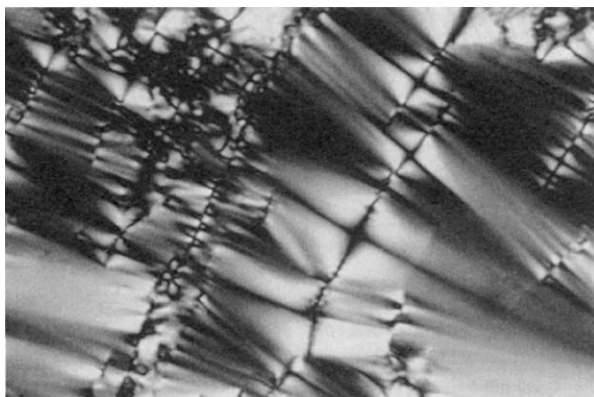


**Figure 7** Structure of a focal-conic domain of a smectic A phase. The growth of a Dupin cyclide (both plan and side elevation) where the layers are depicted as circles and the molecules are only shown in one layer: (A) a plan view, and (B) a section through the fully developed domain. (Data from Gray and Goodby, 1984.)

When a smectic A phase is cooled from the nematic phase or the isotropic liquid and further smectic phases are formed, then the textures produced are paramorphotic of the preceding A phase. For instance, on cooling at a smectic A to smectic C transition, the molecules in the focal-conic domain will tilt and the structure of the domain will become distorted and patchy. This effect is called breaking of the focal-conic domains and is characteristic of changes that occur in the texture when the molecules tilt, i.e., broken focal-conics are formed for the smectic C phase. Alternatively transitory, rather than permanent, changes can occur in the focal-conic texture for certain phase changes. For example, at the smectic A to smectic B transition transitory bands, called transition bars, can appear across the focal-conic domains parallel to the planes of the ellipses. These are due to changes in the positional correlation length of the molecules at the transition. Bands can



**Figure 8** A three-dimensional representation of a focal-conic domain. Part of the domain is cut away to reveal the inner structure. (Data from Gray and Goodby, 1984.)



**Figure 9** Photomicrograph of the focal-conic texture of the smectic A phase seen in the polarizing microscope ( $100\times$ ). The black crosses are optical discontinuities formed by ellipses and hyperbolae.

also be formed at the transition to the E phase, but in this case they are permanent and remain throughout the temperature range of the phase.

By studying paramorphotic patterns and the way that they appear in focal-conics, phase identification can be accomplished and information on mesophase structure can be obtained. However, the problems of phase identification are greatly eased when the focal-conic texture is accompanied by a homeotropically oriented texture. For example, the smectic A phase can exhibit the unbroken focal-conic and optically extinct homeotropic textures together, whereas the smectic C phase exhibits broken focal-conic and schlieren textures, and the E phase exhibits banded focal-conic and mosaic textures. Thus, the





**Figure 10** Photomicrograph of the mosaic texture of a crystal J phase formed in a free-standing film seen in the polarizing microscope (100 $\times$ ).

generation and examination of a combination of various textures together in one preparation is a powerful technique for classifying mesophases.

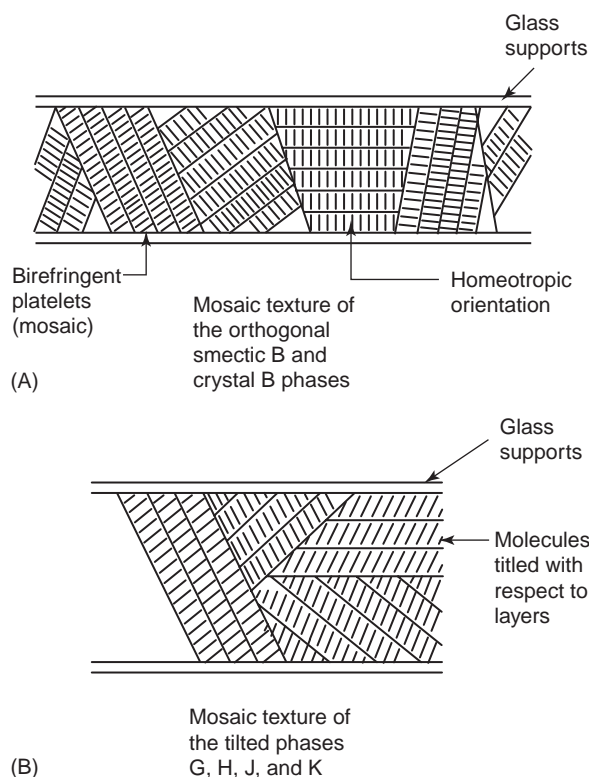
### Mosaic Textures

As noted in the introduction, liquid crystals have differing degrees of order, and for the more ordered crystal-like phases bending of their structures is energetically unfavorable. As a consequence these phases prefer not to form schlieren or focal-conic textures, and instead form mosaic textures, where each domain is separated from neighboring domains by grain boundaries. However, the molecules in each domain have the same relative orientation. The domains in liquid crystal phases can have fairly large sizes (hundreds of micrometers) and irregular shapes; in some cases, where the molecules have extensive long-range order, 'twinning' can be observed.

Mosaics can be observed for a number of phases, for example, smectics F and hexatic B, and crystals B, E, G, J, K, and H, all exhibit mosaics of one form or another (see Figure 10). It is very difficult to identify these phases from their mosaic textures as they all look very similar. Two exceptions are worthy of note: the E phase, because it is biaxial, has a very characteristic mosaic pattern where the domains tend to overlap to give a ghost-like appearance; and the B phases, where the mosaic texture is often accompanied by a homeotropic texture that remains optically extinct on rotation of the microscope stage. Figure 11 shows a variety of mosaic textures where the molecules are either tilted or perpendicular with respect to the layer planes.

### Specific Problems

There are very few problems in using optical microscopy with respect to the preparation of the samples;



**Figure 11** The mosaic textures of smectic phases showing the side elevations of the domains: (A) the orthogonal smectic phases and (B) the mosaic texture of a smectic phase where the long axes of the molecules are tilted with respect to the layer planes. (Data from Gray and Goodby, 1984.)

the only major concerns might be with either decomposition of a material under examination or the quality of its alignment. The first can be minimized by heating the specimen in an inert atmosphere, e.g., nitrogen, and the second can usually be resolved by altering the aligning agent.

### Thermal Optical Microscopy of Liquid Crystals

Thermal polarized light microscopy of liquid crystal systems still primarily involves the identification of phase types. Recently, however, a number of novel phases with complex structures have been discovered and detailed examinations of the configurations of their defects are required in order to provide a basis for future phase classification. Thermal microscopy is also used extensively in examination of the alignment processes of liquid crystals, and, in a related context, electric-field studies on meso-phases are carried out in aligned cells. Electric-field studies are now used as adjuncts to phase classification, e.g., antiferroelectric phases are sometimes identified in the microscope with the aid of electric-field studies.



Thermal optical microscopy also finds extensive use in the analysis of optically active mesophases where the molecules can form macroscopic helical structures. For instance, the pitch length for helical phases can be determined by measuring the distances between the defect lines formed through the interaction of the helical structure and the surface of the preparation. The microscope can also be used as polarimeter to determine helical twist direction for homeotropically aligned helical mesophases.

Thus, since 1921 the polarizing microscope has proved to be a cheap yet powerful tool in the investigation of liquid crystals, and no investigation of a mesogenic material is considered complete without a preliminary examination in the microscope.

See also: **Microscopy**: Overview.

## Further Reading

- Bouligand Y (1980) Defects and textures of hexagonal discotics. *Journal de Physique* 41: 1307–1315.  
 Bouligand Y (1980) Geometry of (non-smectic) hexagonal mesophases. *Journal de Physique* 41: 1297–1306.  
 Brown GH and Wolken JJ (1979) *Liquid Crystals and Biological Structures*. New York: Academic Press.

- Chandrasekhar S and Ranganath GS (1990) Discotic liquid crystals. *Research Progress in Physics* 53: 57–84.  
 Cognard J (1982) Alignment of nematic liquid crystals and their mixtures. *Molecular Crystals and Liquid Crystals (Special Supplement)* 51: 1–77.  
 Collings PJ (1990) *Liquid Crystals. Nature's Delicate State of Matter*. Princeton: Princeton University Press.  
 De Gennes PG (1974) *The Physics of Liquid Crystals*. London: Oxford University Press.  
 Demus D and Richter L (1978) *Textures of Liquid Crystals*. New York: Verlag Chemie.  
 Demus D and Zascke H (1984) *Flüssige Kryistalle in Tabellen, I and II*. Leipzig: VEB Deutscher Verlag für Grundstoffindustrie.  
 Dierking 2003 *Textures of Liquid Crystals*: New York: Wiley-VCH.  
 Gray GW and Goodby JW (1984) *Smectic Liquid Crystals: Textures and Structures*. Glasgow: Leonard Hill.  
 Gray GW and Windsor PA (1974) *Liquid Crystals and Plastic Crystals*, vols 1 and 2. London: Ellis Horwood.  
 Hartshome NH and Stuart A (1970) *Crystals and the Polarizing Microscope*. London: Edward Arnold.  
 Kelker H and Hatz R (1980) *Handbook of Liquid Crystals*. Basel: Verlag Chemie.  
 Priestley EB, Wojtowicz P, and Sheng P (1976) *Introduction to Liquid Crystals*. New York: Plenum.

## Semiconductors

**D J Smith and M R McCartney**, Arizona State University, Tempe, AZ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The transmission electron microscope (TEM) has evolved into a sophisticated instrument capable of providing structural and chemical information from solid materials over a wide range of magnification, to a level of spatial resolution that is unapproachable by most other techniques. The TEM offers a wide variety of imaging, diffraction, and microanalytical modes that can be used individually or in combination to extract essential information about semiconductor materials and devices. Techniques that are widely available and commonly used include transmission and high-resolution electron microscopy, electron-energy-loss and energy-dispersive X-ray spectroscopy, and convergent-beam electron diffraction. Advanced methods include electron

holography, cathodoluminescence (CL) and Z-contrast annular-dark-field (ADF) imaging. The purpose of this article is to review briefly, with selected representative examples, applications of these various imaging, diffraction, and microanalytical modes of the TEM to semiconductor materials. These specific applications have been chosen to demonstrate the versatility and power of the TEM for solving relevant technological problems.

## Background

The TEM utilizes a finely focused electron beam and several intermediate stages of magnification to obtain a highly enlarged image of the specimen region of interest. Electron wavelengths are typically much less than 0.1 nm, but unavoidable lens aberrations restrict optimum image resolutions to the range of 0.25–0.16 nm for operating voltages of 200–400 kV. At this resolution level, details of atomic configurations for many types of inorganic materials are still obtainable in low-index crystallographic projections.

Thermal optical microscopy also finds extensive use in the analysis of optically active mesophases where the molecules can form macroscopic helical structures. For instance, the pitch length for helical phases can be determined by measuring the distances between the defect lines formed through the interaction of the helical structure and the surface of the preparation. The microscope can also be used as polarimeter to determine helical twist direction for homeotropically aligned helical mesophases.

Thus, since 1921 the polarizing microscope has proved to be a cheap yet powerful tool in the investigation of liquid crystals, and no investigation of a mesogenic material is considered complete without a preliminary examination in the microscope.

See also: **Microscopy**: Overview.

## Further Reading

- Bouligand Y (1980) Defects and textures of hexagonal discotics. *Journal de Physique* 41: 1307–1315.  
 Bouligand Y (1980) Geometry of (non-smectic) hexagonal mesophases. *Journal de Physique* 41: 1297–1306.  
 Brown GH and Wolken JJ (1979) *Liquid Crystals and Biological Structures*. New York: Academic Press.

- Chandrasekhar S and Ranganath GS (1990) Discotic liquid crystals. *Research Progress in Physics* 53: 57–84.  
 Cognard J (1982) Alignment of nematic liquid crystals and their mixtures. *Molecular Crystals and Liquid Crystals (Special Supplement)* 51: 1–77.  
 Collings PJ (1990) *Liquid Crystals. Nature's Delicate State of Matter*. Princeton: Princeton University Press.  
 De Gennes PG (1974) *The Physics of Liquid Crystals*. London: Oxford University Press.  
 Demus D and Richter L (1978) *Textures of Liquid Crystals*. New York: Verlag Chemie.  
 Demus D and Zascke H (1984) *Flüssige Kryistalle in Tabellen, I and II*. Leipzig: VEB Deutscher Verlag für Grundstoffindustrie.  
 Dierking 2003 *Textures of Liquid Crystals*: New York: Wiley-VCH.  
 Gray GW and Goodby JW (1984) *Smectic Liquid Crystals: Textures and Structures*. Glasgow: Leonard Hill.  
 Gray GW and Windsor PA (1974) *Liquid Crystals and Plastic Crystals*, vols 1 and 2. London: Ellis Horwood.  
 Hartshome NH and Stuart A (1970) *Crystals and the Polarizing Microscope*. London: Edward Arnold.  
 Kelker H and Hatz R (1980) *Handbook of Liquid Crystals*. Basel: Verlag Chemie.  
 Priestley EB, Wojtowicz P, and Sheng P (1976) *Introduction to Liquid Crystals*. New York: Plenum.

## Semiconductors

**D J Smith and M R McCartney**, Arizona State University, Tempe, AZ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The transmission electron microscope (TEM) has evolved into a sophisticated instrument capable of providing structural and chemical information from solid materials over a wide range of magnification, to a level of spatial resolution that is unapproachable by most other techniques. The TEM offers a wide variety of imaging, diffraction, and microanalytical modes that can be used individually or in combination to extract essential information about semiconductor materials and devices. Techniques that are widely available and commonly used include transmission and high-resolution electron microscopy, electron-energy-loss and energy-dispersive X-ray spectroscopy, and convergent-beam electron diffraction. Advanced methods include electron

holography, cathodoluminescence (CL) and Z-contrast annular-dark-field (ADF) imaging. The purpose of this article is to review briefly, with selected representative examples, applications of these various imaging, diffraction, and microanalytical modes of the TEM to semiconductor materials. These specific applications have been chosen to demonstrate the versatility and power of the TEM for solving relevant technological problems.

## Background

The TEM utilizes a finely focused electron beam and several intermediate stages of magnification to obtain a highly enlarged image of the specimen region of interest. Electron wavelengths are typically much less than 0.1 nm, but unavoidable lens aberrations restrict optimum image resolutions to the range of 0.25–0.16 nm for operating voltages of 200–400 kV. At this resolution level, details of atomic configurations for many types of inorganic materials are still obtainable in low-index crystallographic projections.

The conventional transmission electron microscope (CTEM) uses a broad, usually circular, beam of electrons, with a diameter of  $\sim 0.5\text{--}1.0\text{ }\mu\text{m}$  across, which is transmitted through a suitably thinned specimen. The final magnified image is visible on a fluorescent screen at the base of the microscope or it can be displayed on a television monitor via a suitable image pickup system. The image can then be recorded on photographic film or a digital camera, or else stored on videotape for later review of dynamic events. The scanning transmission electron microscope (STEM) involves the rastering of a focused nanoprobe across the sample in synchronization with the detection of electrons transmitted through the sample, which are then used to form an image by intensity modulation of a display monitor. When the nanoprobe is stopped at specific sample locations, elemental information about the irradiated area can be obtained through detection of the characteristic X-rays emitted by the sample or by examination of the energy spectrum of the transmitted electrons. The complementary information provided by these latter microanalytical signals can be highly useful in elucidating the origins of local defects or inhomogeneities in semiconductor materials.

## Sample Types and Preparation Methods

In the field of semiconductors, there are many different types of specimen that are of interest. Depending on the particular application or problem area, the electron microscopist might focus on such features as dopant clusters, point, line or planar defects, or buried interfaces, etc. A serious restriction of TEM methods is that only relatively thin sections of material can be examined. Since the regions of interest often involve microstructural features that must be evaluated in at least two, and sometimes three dimensions, it often becomes necessary to prepare specimens for observation in two orthogonal directions, i.e., plan-view and cross-section. Plan-view specimens are usually straightforward to prepare, sometimes only involving chemical etching or mechanical polishing. Conversely, the preparation of cross-sections may involve the use of costly equipment such as focused-ion-beam systems, and it can often be quite time consuming.

### Plan-View Specimens

The simplest cases involve passive, and usually selective, wet-chemical etching of material such as the substrate, so that overlayers which may be 100 nm or less in thickness are left for examination.

Typical examples include the selective removal of III–V buffer layers leading to float-off of thin, epitaxially grown III–V overlayers, removal of silicon substrates from diamond films grown by chemical vapor deposition (CVD), and the removal of metal contacts from the underlying semiconductor. In similar fashion, buried layers can often be extracted using a sequence of suitable etchants, and when this layer is not contiguous, a thin amorphous carbon film can be deposited to give structural integrity to the layer. When bulk material needs to be examined, thinned regions must be produced by other methods. Thinning can be accomplished, for example, by mechanical polishing and dimple-grinding followed either by chemical jet-etching or ion-beam milling. The latter approach is somewhat slower but will generally lead to larger areas of thinned material.

### Cross-Sectional Specimens

The cross-sectional geometry is required when the features of interest are buried within the material or else distributed throughout its depth. For example, it may be necessary to assess the planarity of buried interfaces, or to establish the distribution of any interfacial defects, their width and composition, and the presence of secondary phases. Cross-sectional specimens are usually prepared by gluing the material of interest face-to-face with additional supporting material. Mechanical polishing and then dimple-polishing of a thin slice cut from this stack reduce the total overall thickness to  $\sim 5\text{--}10\text{ }\mu\text{m}$ . Finally, argon-ion-beam milling is used to achieve perforation through the cross-sectional film. Thin areas of the interface are then created where the perforated hole intersects the glued faces. Damage to many semiconducting materials during cross-sectioning can be reduced by ion-milling with the specimen held at liquid nitrogen temperature. Some materials, such as indium- or zinc-containing compounds, benefit from additional milling with reactive ions such as  $\text{I}^+$ . Specimens containing materials with very different ion-milling rates, such as those grown on sapphire substrates, have a tendency to form ‘bridges’ of material spanning the perforation. These bridges can often be overcome by ion-milling at very low angles of incidence along the surface ( $\sim 1^\circ\text{--}2^\circ$ ) or by using the technique of chemically assisted ion-beam milling.

In many cross-sectional specimens, especially epitaxial systems, the choice of crystallographic direction that will be aligned parallel to the electron beam direction during TEM observation can be very important. For example, in binary, ternary, and quaternary alloys of cubic (zinc blende) III–V materials, the boundaries between layers are much more clearly

revealed if the sample is observed with the [100] projection parallel to the electron beam direction, rather than [110], because diffracted beams have greater chemical sensitivity in this projection. Specimens having this preferred [100] orientation can often be prepared relatively quickly without ion-milling by using a cleaved-wedge method. This preparation involves applying controlled pressure to small squares of material  $\sim 200\mu\text{m}$  thick, which cleave along the [110] direction to produce  $90^\circ$  wedges about the [100] direction. The thinnest edges of these samples can then be viewed by TEM when the sample is suitably mounted. Wedges with shallower angles, and therefore larger regions of thin area, can sometimes be produced by cleavage along higher-index planes.

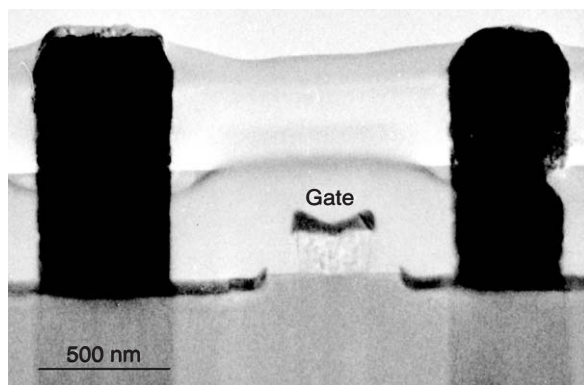
Cross-sectional specimen observation is usually the preferred geometry for studying TEM specimens of actual semiconductor devices. When the devices are relatively large, or when similar structural units are repeated many times, the preparation methods described above can be employed with high success rates. However, when a specific part of a device must be targeted, for example in integrated-circuit failure analysis, then newer techniques based either on carefully controlled mechanical polishing at very shallow wedge angles, or with a focused-ion-beam system, are becoming the preferred method of preparation.

## Imaging and Diffraction

### Transmission Electron Microscopy

The standard operating mode for the TEM is termed amplitude or diffraction contrast imaging. Electrons that have passed through the sample are focused to form the highly magnified image of the region of interest on the final viewing screen. The image contrast results from the use of a suitably small objective aperture that blocks most of the scattered (or diffracted) electrons, which thus do not reach the final image plane. Crystalline samples can be aligned by reference to the electron diffraction pattern (EDP), to ensure that the specimen orientation relative to the incident beam direction will satisfy a strongly diffracting condition. Under such conditions, common types of defects are easily recognized by their characteristic appearance and second-phase materials are often conspicuous. Moreover, characteristic crystal lattice spacings and angles can also often be measured and identified from the EDP.

The availability of a crystalline substrate or supporting medium often provides a useful method for orientation of the sample during observation within the electron microscope. By using the substrate EDP



**Figure 1** Diffraction contrast image showing a cross-sectional view of a Si-based integrated memory device prepared for electron microscopy observation by focused-ion-beam milling.

as a reference, it becomes possible to align the substrate surface normal, or the internal interface of interest, in such a way that they are aligned perpendicular to the electron beam direction. The microstructure of thin films or multilayers can then be characterized as a function of lateral position as well as distance from the substrate (or previous layer). Nevertheless, it must be appreciated that imaging of complex devices represents a serious challenge to the electron microscopist because of the inherent difficulty of preparing electron-transparent regions across the entire structure simultaneously. **Figure 1** is an example of a diffraction contrast image where this thinning has been successfully accomplished: it shows a cross-sectional view of a silicon-based integrated memory device that has been prepared for microscopy using a focused-ion-beam milling system.

### High-Resolution Electron Microscopy

When a large objective aperture (or perhaps none at all) is used, one or more diffracted beams can interfere with the transmitted beam in forming the high-resolution image. The image contrast then depends on the relative phases of these various beams, so this operating mode is usually termed phase-contrast imaging. Under optimum imaging conditions, these phase-contrast images can be interpreted directly in terms of the projected crystal potential, with a spatial resolution that is often better than  $0.2\text{ nm}$ . Individual atomic columns can thus be resolved in many crystalline inorganic materials, at least when they are viewed in low-index zone axes projections where the atomic separations are comparatively large in projection. For highest resolution, the specimen thickness should ideally be less than  $\sim 10\text{ nm}$ . High electron doses (typically  $\sim 500\text{--}2000\text{ electrons }\text{\AA}^{-2}$ ) are also required. Thus, specimens must be relatively

resistant to electron irradiation effects, and organic materials cannot be imaged directly under such conditions. Furthermore, dynamic events can be followed in real time without significant loss of resolution by using a specimen heating holder, and adding a TV image pickup system to the base of the microscope lens column.

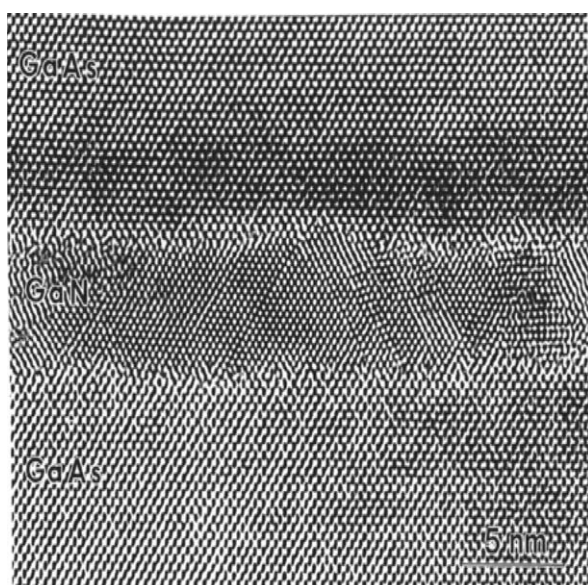
High-resolution electron microscopy has been applied to the characterization of a diverse range of inorganic materials. Important applications include determining the detailed microstructure of defects, interfaces and grain boundaries, investigating nanocrystalline features in amorphous films, and studying small particles in heterogeneous catalysts. From a technological point of view, the characterization of semiconductor interfaces continues to be a very important application. Interfacial structure is critical to the viability and operation of many devices. Examples of such interfaces include superlattices and multiple quantum well (MQW) structures consisting of alternate layers of different III-V semiconductors and having optoelectronic applications; heteroepitaxial layers of II-VI semiconductors grown on III-V substrates; heteroepitaxial systems such as GaAs grown on silicon with potential applications to integrated optoelectronics; and metal/semiconductor contacts for discrete devices. High-resolution images are able to provide details about the interfacial structure that is usually unavailable from any other technique. As an example, **Figure 2** shows a GaAs/GaN/GaAs heterostructure as grown by molecular-beam epitaxy (MBE). There is a large lattice mismatch between the two materials and a high density of lattice

defects in the intermediary GaN layer. Nevertheless, it is clear that high-quality epitaxial growth of the uppermost GaAs layer has been achieved.

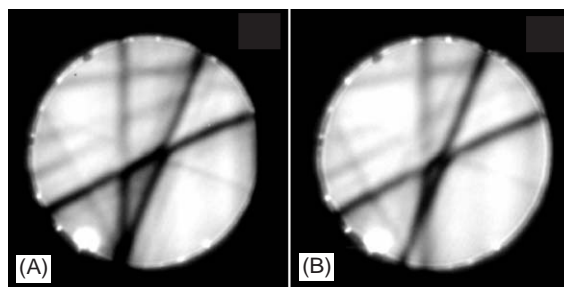
### Convergent-Beam Electron Diffraction

The technique of convergent-beam electron diffraction (CBED) utilizes a convergent focused beam to obtain diffraction patterns from small specimen regions. CBED patterns are comprised of disks of intensity (rather than spots) and these disks are rich in detail that can often be exploited to reveal important aspects about the specimen. Applications of CBED range from the measurement of specimen thickness, to the determination of crystal symmetry (i.e., point and space groups) for analysis of new crystallographic phases and analysis of phase transitions, and determination of amplitudes and phases of low-order structure factors. The technique can also provide very accurate measurements of small changes in lattice parameters resulting from local changes in composition or lattice strain. When the specimen is sufficiently thin ( $< 50$  nm), then the spatial resolution is determined primarily by the size of the focused incident probe. When very small probes are used ( $< 1$  nm or less in diameter), the technique is termed nanodiffraction. A field-emission source of electrons is then essential since the beam intensity at the specimen level would otherwise be too low to be usable. Nanodiffraction is particularly useful for studying very small particles, and for determining local order and crystal symmetry in thin films of near-amorphous materials or disordered crystals.

The high sensitivity of CBED patterns to small changes in lattice parameter is illustrated in **Figure 3**. These two CBED patterns compare the central bright-field CBED disk for a thin GaN film oriented close to  $[10\bar{1}0]$ , with the corresponding disk from a thin epitaxial  $\text{Ga}_{0.96}\text{In}_{0.04}\text{N}$  quantum well layer. The fractional difference in lattice parameter due to the difference in composition is considerably less than 0.5%, but the relative positions of the dark



**Figure 2** High-resolution electron micrograph showing GaAs/GaN/GaAs heterostructure grown by molecular beam epitaxy.



**Figure 3** Comparison of central disks from CBED patterns: (A) GaN and (B)  $\text{Ga}_{0.96}\text{In}_{0.04}\text{N}$ . Note relative shifted positions of dark higher-order Laue zone lines due to differences in lattice parameter.



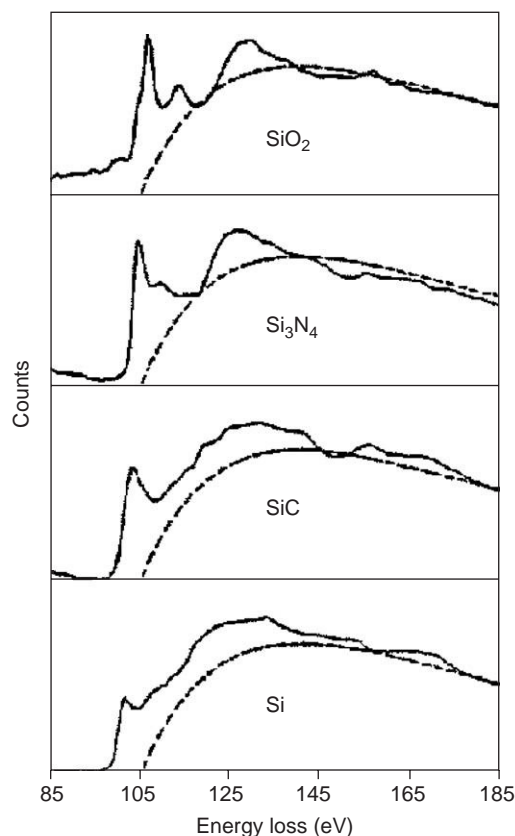
higher-order Laue zone (HOLZ) lines, which are the dark diagonal lines crossing the disk, are readily apparent. The shift in these HOLZ lines as a function of the probe position along the quantum well was used to determine variations in the local In concentration with high precision and spatial resolution.

## Spectroscopy

### Electron-Energy-Loss Spectroscopy

Electron-energy-loss spectroscopy (EELS) uses the characteristic energy-loss spectrum of electrons that have been transmitted through the sample. Since the amount of electron-energy loss is related directly to the ionization energy, which has a well-known characteristic behavior for any given element, information can be obtained about elemental composition, chemical bonding, and electronic structure. The typical EELS spectrum contains a monotonically decreasing background and several broad superimposed peaks, each of which has an energy that is directly related to a particular inelastic scattering process. The spatial resolution is determined primarily by the diameter of the incident focused probe provided that the sample thickness is not excessive. The EELS technique can be used to observe band structure effects, to provide chemical information such as distinguishing between allotropes, and to determine oxidation states, as well as to measure short-range order (i.e., determination of radial distribution function). Quantitative elemental analysis, both fixed-point and time- and position-resolved, is possible with sensitivity levels sometimes approaching or exceeding better than one atomic percent. Specimens that are intended for microanalysis must be thin (<50–100 nm) to avoid multiple scattering effects that can cause loss of spectral resolution and sensitivity. Their surfaces must also be carefully cleaned if quantitative analysis is required since substantial carbon contamination is otherwise liable to occur during small-probe observation. It is becoming common-place to clean samples, and the sample holder, in a plasma-cleaning device immediately prior to insertion into the electron microscope. Finally, it needs to be appreciated that irradiation effects may occur during examination with the intense stationary probe, which can significantly alter the local microstructure and composition.

For applications in the semiconductor field, EELS is commonly used when elemental information is needed about depletion regions or segregation profiles, for example following ion implantation, amorphization, and recrystallization, or as a result of defect cluster precipitation during annealing cycles.



**Figure 4** Electron-energy-loss spectra, with background stripped, showing Si-L fine-edge structure originating from different Si-containing compounds.

Diffusion profiles across interfaces can also be extracted from EELS data provided that the effects of the probe size and shape are properly taken into account. An interesting comparison of the near-edge fine structure in EELS spectra from four Si-containing materials is shown in **Figure 4**. The differences in the shapes of these spectra, which can be replicated theoretically, are due to differences in composition as well as changes in the local chemical bonding (i.e., the local atomic environment).

### Energy-Dispersive X-Ray Spectroscopy

Energy-dispersive X-ray (EDX) spectroscopy utilizes the characteristic spectrum of X-rays that is emitted by a sample, following initial excitation by the high-energy electron beam. Information about the elemental composition of the sample can be obtained, where the spatial resolution is determined by the probe size, any beam broadening occurring within the specimen, and the effects of any backscattered electrons on the specimen around the point of analysis. EDX is relatively simple and can be used to provide rapid qualitative microanalysis. Quantitative elemental analysis can also be achieved, with an

accuracy approaching a few atomic percent. The technique can be used for atomic site and species determination using electron channeling. Note, however, that quantitative EDX analysis implies thin specimens, and (pre)calibration of the EDX analysis system using standards of known composition is required. EDX is generally better suited than EELS for detecting elements of high atomic number ( $Z$ ) but the technique does not usually provide local bonding information, unlike EELS. Low- $Z$  elements ( $Z < 11$ ) are not detectable by some EDX systems depending on absorption by the detector window, and they are only detectable with limited sensitivity by others.

Analysis by EDX represents a useful complementary approach to conventional imaging. Figure 5A shows an ADF image of a Cr-doped AlN film which shows room temperature ferromagnetic behavior. Since the contrast of an ADF image should be sensitive to the atomic number,  $Z$ , of the sample region

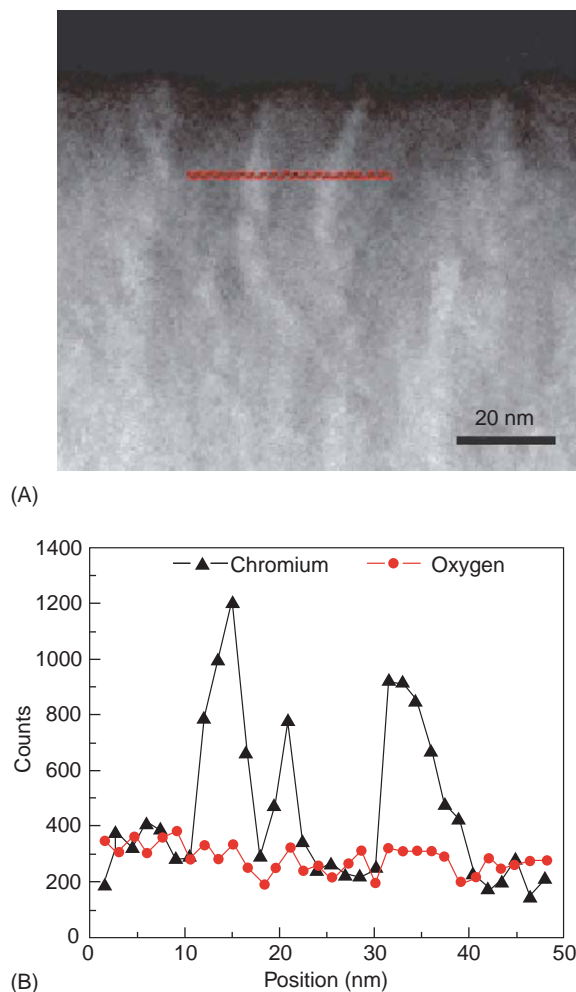
being imaged, the white, roughly columnar regions are interpreted as being Cr-rich. EDX spectra, as shown by the example in Figure 5B, were acquired with a small ( $\sim 0.5$  nm) focused probe. The Cr line profile shows that Cr has large upward and downward swings by factors of 2 or 3, indicative of Cr nanoclusters within the film. In contrast, the matching O profile is relatively flat, implying that no chromium-oxide phase had been formed either during deposition or postgrowth annealing.

## Advanced and Novel Approaches

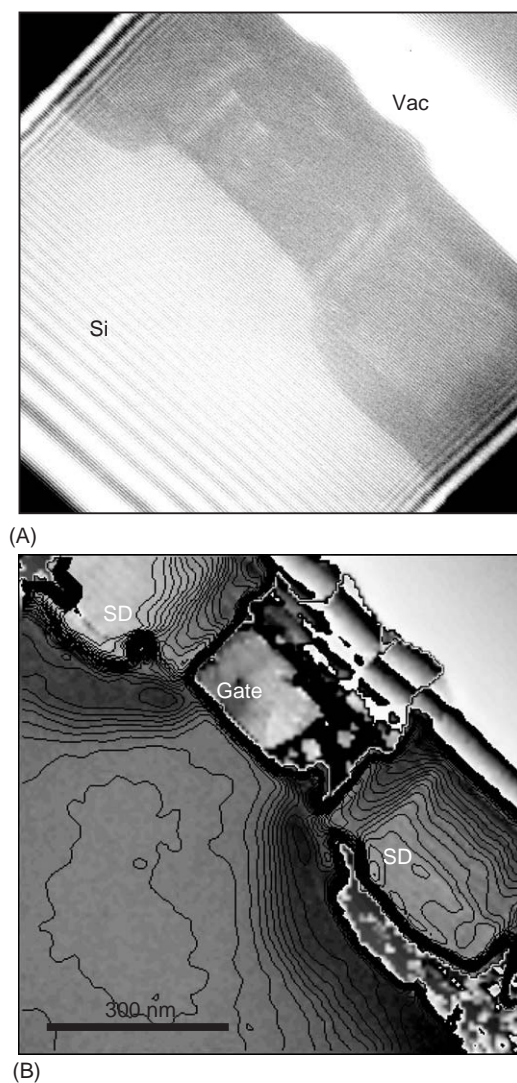
### Electron Holography

Electron holography is an electron interference method that permits the retrieval of both the amplitude and the phase of the electron beam, unlike the case for normal image recording where all of the phase information is lost. A coherent source of electrons, usually provided by a field-emission electron gun, is a necessary requirement. Interference between the reference wave in vacuum and the wave that has passed through the specimen is then achieved using a thin, positively charged wire ('biprism') conveniently located in the imaging system, usually in the selected-area aperture plane. The electron wave present at the exit surface of the sample can be reconstructed, and high-spatial-resolution information about any internal electric and magnetic fields can then be extracted. The technique known as off-axis holography is most commonly used, and allows the mean inner potential of the sample to be accurately measured (to within  $\sim 1\%$ ). The technique is thus useful for studying interface sharpness and any diffusion across interfaces. Specimens should preferably be thin ( $< 200$  nm) in order to minimize inelastic scattering. Moreover, because of the need for a reference beam, the region of interest should typically be located no more than  $\sim 200$  nm from the edge of the specimen.

Knowledge of the electrostatic potential distributions across device junctions is of critical importance to design engineers. Typically, indirect methods such as dopant concentration profiling by secondary ion mass spectrometry, are used to extract relevant information, but the available spatial resolution is inadequate for current and future generation devices. A recent development of off-axis electron holography has been the demonstration that potential distributions within real devices can be imaged with high spatial resolution ( $\sim 6$  nm) and sensitivity ( $\sim \pm 0.1$  V). Figure 6 illustrates an application of the holographic technique to characterize the two-dimensional (2D) potential distribution within a deep submicron p-type field-effect transistor. Figure 6A



**Figure 5** (A) Cross-sectional ADF image of ferromagnetic Cr-doped AlN film; (B) corresponding EDX spectrum indicating presence of Cr-rich regions along line indicated in (A).



**Figure 6** (A) Off-axis electron hologram from a pMOS transistor device; (B) corresponding electrostatic potential. Contour steps of 0.1 V. Gate and source/drain (SD) contacts are indicated.

shows the original hologram and **Figure 6B** shows the reconstructed 2D map of the electrostatic potential within the device resulting from the B doping, where the phase contour steps correspond to 0.1 V. Further studies are still needed to determine whether changes in potential as a result of device biasing can be measured reliably, and there are concerns that need to be resolved about surface ‘dead layers’ sometimes caused during sample preparation. Nevertheless, it is already clear that the technique is likely to have a major impact on future device engineering.

### TEM Cathodoluminescence

CL in the TEM involves the collection and analysis of light that is emitted from a specimen when it has been excited by irradiation with the electron beam.

Whereas all specimens in a TEM will emit X-rays that are detectable by EDX, only those with an appropriate electronic structure, such as a suitable band-gap or energy states in ionic impurities, will emit light. In thin specimens, the CL signal may be too weak to be detected because there is limited material available to luminesce and the presence of surface states may reduce the CL intensity. Details of CL spectra can indicate the presence of electronic states associated with impurities and defects at the parts-per-million level and the distribution of such inhomogeneities can be mapped using the STEM mode. A major attraction of (S)TEM CL is its ability to correlate CL emission with microstructure and thus provide a direct link between electronic structure and crystalline defects. The CL spatial resolution is determined by the probe size and several factors that are specimen-dependent, including thickness, carrier diffusion length, CL generation volume and temperature: 200 nm is attainable in some materials. The technique can be used for detecting and mapping optically/electrically active dopants, impurities, and defects in semiconductors, especially direct band-gap (i.e., not silicon generally) and wide band-gap materials such as diamond, group III nitrides, and some minerals.

### Transmission Electron Microscopy

It has been shown here that the TEM, which has evolved over a period of more than 70 years into a highly sophisticated instrument for microstructural characterization, has many different operating modes that can be important in the field of semiconductor technology. As new types of device structures continue to be developed and as device dimensions are pushed ever smaller, these TEM-based techniques will remain at the forefront of developmental research. And it has been noted that electron holography is already starting to have an important role in submicron device engineering. As well as being applied to the characterization of fabricated materials and structures, the TEM is increasingly being used for studying growth of semiconducting materials in real time within the microscope, both under idealized ultrahigh-vacuum (UHV) conditions and also under less clean but perhaps more technologically relevant situations. Specialized UHV instruments remain the highly restricted domain of a comparatively few university and corporate research laboratories, but they will continue to be important because workers in the field are concerned to understand the properties and behavior of clean semiconductor surfaces. Under vacuum conditions more typical of recent fabrication procedures, the *in situ* imaging of nucleation

and growth of epitaxial layers on semiconductor surfaces, and the process of metallization within a controlled atmosphere environmental cell are exciting, and very promising, developments. Finally, electron microscopes will undoubtedly continue to be heavily utilized as an analytical and diagnostic tool for failure analysis in semiconductor laboratories.

See also: **Microscopy Techniques:** Electron Microscopy.

## Further Reading

- Anderson RM (ed.) (1990) *Specimen Preparation for Transmission Electron Microscopy of Materials*, Materials Research Society Symposium Proceedings, vol. 199. Pittsburgh: Materials Research Society.
- Anderson RM (ed.) (1991) *Specimen Preparation for Transmission Electron Microscopy of Materials*, Materials Research Society Symposium Proceedings, vol. 254. Pittsburgh: Materials Research Society.
- Buseck PR, Cowley JM, and Eyring L (eds.) (1988) *High Resolution Electron Microscopy and Associated Techniques*. New York: Oxford University Press.
- Goodhew PJ (1985) Thin foil preparation for electron microscopy. In: Glauert AM (ed.) *Practical Methods in Electron Microscopy*, vol. 11. Amsterdam: Elsevier.
- Hren JJ, Gldstein JL, and Joy DC (eds.) (1986) *Introduction to Analytical Electron Microscopy*. New York: Plenum.
- Rossiter BW and Hamilton JF (eds.) (1991) *Physical Methods of Chemistry*, vol. 4, Microscopy. New York: Wiley.
- The most comprehensive source of contemporary material in the field of electron microscopy for semiconductors is the series of books entitled *Microscopy of Semiconducting Materials*, which are edited by A.G. Cullis and various colleagues and published on a biennial basis by the Institute of Physics (UK). These books contain short-refereed articles arising from the successive conferences of the same name held in Oxford in April of every odd calendar year.
- Williams DB and Carter CB (1996) *Transmission Electron Microscopy*. New York: Plenum.

## Proteins and Nucleic Acids

J R Harris, Mainz University, Mainz, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 3261–3270, © 1995, Elsevier Ltd.

## Introduction

Since the 1950s, transmission electron microscopy (TEM) has made a significant contribution to the study of biological materials at the cellular, subcellular, and macromolecular levels. From an essentially qualitative beginning, increasingly quantitative data have been obtained, limited in resolution primarily by the specimen preparation techniques and lability of biological molecules within the electron beam, rather than the instrumental resolution. It is the aim of this article to concentrate only on the macromolecular applications, with emphasis being placed on protein molecules, nucleic acids, and their interaction with proteins, membranes, lipids, and lipoprotein complexes. Apart from direct structural analysis, the use of immunolabeling at the molecular level has provided a major extension of the lower-resolution cellular and tissue-based antigenic localization, with immunoferritin or immunogold complexes. Individual protein subunits, subunit domains, and antigenic epitopes have been labeled with specific

immunoglobulins or Fab fragments, from either polyclonal or monoclonal antibodies. Indeed, chemical labels that incorporate nanogold clusters have enabled the position of thiol esters to be defined, for example, within the  $\alpha_2$ -macroglobulin molecule. These very small electron-dense labels are likely to provide further localization of specific sites within protein molecules in the future.

The study of macromolecular structure by scanning transmission electron microscopy (STEM) has, in general, not provided information beyond that already available from TEM. Nevertheless, the determination of the molecular mass by dark-field STEM from the electron scattering of unstained macromolecules, exploited throughout the 1980s by Andreas Engel and others, continues to provide useful comparative if not absolute data. Determination of the mass of the nuclear pore complex and its components has been a valuable contribution of this approach. The problem of varying mass loss with electron dose appears to necessitate extrapolation to zero dose for all biological samples. The introduction of protein stabilization by glutaraldehyde treatment, which is necessary in some cases, also introduces a variable that is difficult to standardize quantitatively. High-resolution scanning electron microscopy (SEM) has begun to make a contribution at the macromolecular level, but is unlikely to ever reach or exceed the resolution attainable by TEM.

and growth of epitaxial layers on semiconductor surfaces, and the process of metallization within a controlled atmosphere environmental cell are exciting, and very promising, developments. Finally, electron microscopes will undoubtedly continue to be heavily utilized as an analytical and diagnostic tool for failure analysis in semiconductor laboratories.

See also: **Microscopy Techniques:** Electron Microscopy.

## Further Reading

- Anderson RM (ed.) (1990) *Specimen Preparation for Transmission Electron Microscopy of Materials*, Materials Research Society Symposium Proceedings, vol. 199. Pittsburgh: Materials Research Society.
- Anderson RM (ed.) (1991) *Specimen Preparation for Transmission Electron Microscopy of Materials*, Materials Research Society Symposium Proceedings, vol. 254. Pittsburgh: Materials Research Society.
- Buseck PR, Cowley JM, and Eyring L (eds.) (1988) *High Resolution Electron Microscopy and Associated Techniques*. New York: Oxford University Press.
- Goodhew PJ (1985) Thin foil preparation for electron microscopy. In: Glauert AM (ed.) *Practical Methods in Electron Microscopy*, vol. 11. Amsterdam: Elsevier.
- Hren JJ, Gldstein JL, and Joy DC (eds.) (1986) *Introduction to Analytical Electron Microscopy*. New York: Plenum.
- Rossiter BW and Hamilton JF (eds.) (1991) *Physical Methods of Chemistry*, vol. 4, Microscopy. New York: Wiley.
- The most comprehensive source of contemporary material in the field of electron microscopy for semiconductors is the series of books entitled *Microscopy of Semiconducting Materials*, which are edited by A.G. Cullis and various colleagues and published on a biennial basis by the Institute of Physics (UK). These books contain short-refereed articles arising from the successive conferences of the same name held in Oxford in April of every odd calendar year.
- Williams DB and Carter CB (1996) *Transmission Electron Microscopy*. New York: Plenum.

## Proteins and Nucleic Acids

**J R Harris**, Mainz University, Mainz, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 3261–3270, © 1995, Elsevier Ltd.

## Introduction

Since the 1950s, transmission electron microscopy (TEM) has made a significant contribution to the study of biological materials at the cellular, subcellular, and macromolecular levels. From an essentially qualitative beginning, increasingly quantitative data have been obtained, limited in resolution primarily by the specimen preparation techniques and lability of biological molecules within the electron beam, rather than the instrumental resolution. It is the aim of this article to concentrate only on the macromolecular applications, with emphasis being placed on protein molecules, nucleic acids, and their interaction with proteins, membranes, lipids, and lipoprotein complexes. Apart from direct structural analysis, the use of immunolabeling at the molecular level has provided a major extension of the lower-resolution cellular and tissue-based antigenic localization, with immunoferritin or immunogold complexes. Individual protein subunits, subunit domains, and antigenic epitopes have been labeled with specific

immunoglobulins or Fab fragments, from either polyclonal or monoclonal antibodies. Indeed, chemical labels that incorporate nanogold clusters have enabled the position of thiol esters to be defined, for example, within the  $\alpha_2$ -macroglobulin molecule. These very small electron-dense labels are likely to provide further localization of specific sites within protein molecules in the future.

The study of macromolecular structure by scanning transmission electron microscopy (STEM) has, in general, not provided information beyond that already available from TEM. Nevertheless, the determination of the molecular mass by dark-field STEM from the electron scattering of unstained macromolecules, exploited throughout the 1980s by Andreas Engel and others, continues to provide useful comparative if not absolute data. Determination of the mass of the nuclear pore complex and its components has been a valuable contribution of this approach. The problem of varying mass loss with electron dose appears to necessitate extrapolation to zero dose for all biological samples. The introduction of protein stabilization by glutaraldehyde treatment, which is necessary in some cases, also introduces a variable that is difficult to standardize quantitatively. High-resolution scanning electron microscopy (SEM) has begun to make a contribution at the macromolecular level, but is unlikely to ever reach or exceed the resolution attainable by TEM.



The correlation of molecular localization/labeling data with biochemical data of a structural nature is very important. Furthermore, it is pertinent that all TEM structural investigations on macromolecular material require a constant two-way interaction with the established biochemical, biophysical, molecular biological, physiological, microbiological, and immunological approaches. To obtain a meaningful understanding of nature, macromolecular structure must always be related to biological function.

## Sample Types

In general, it is essential that the macromolecules or membraneous organelles utilized for TEM studies of the type to be described should be isolated from tissues or single cells by subcellular fractionation procedures and subjected to centrifugal, chromatographic, and other techniques that are able to yield the highest possible homogeneity, purity, and stability in solution.

Having obtained homogeneity/purity, a range of electron microscopical specimen preparation techniques can be applied to organelle or membrane suspensions and solutions of macromolecules, which usually require only a few microliters of fluid

containing a protein/lipid/nucleic acid concentration in the range  $0.01\text{--}1.0\text{ mg ml}^{-1}$ . The presence of buffer salts, surfactants, and organic centrifugation materials such as sucrose, glycerol, and metrizamide can often create problems during specimen preparation and these should be removed, e.g., by centrifugal pelleting or filtration with washing, or dialysis against a low-molarity buffer, if possible. It is important that no dissociation or aggregation of the macromolecule(s) under study occurs in solution, as this can greatly interfere with specimen preparation, as well as subsequent image interpretation.

## Techniques

Several well-established and extensively documented techniques are available for the production of specimens of freely dispersed viruses, macromolecules, and cellular membranes. In all cases, it is important that the material is thinly spread in such a manner that individual particles can be defined with respect to their dimensions and shape, even though several different orientation-dependent image projections may be produced from a single particle by the TEM. Metal shadowing (single-angle and rotary) has been utilized for the production of specimens of nucleic acids,

**Table 1** Some applications of TEM to cell wall and membrane systems

---

Acetylcholine receptor
Bacterial surface layers and outer membrane
Bacterial lipopolysaccharides
Bacterial, eukaryotic plasma membrane, and endomembrane
ATPases (Na/K, H/K, and Ca/Mg)
Bacterial and mitochondrial porins and ion channels
Bacterial spore coat
Bacterial photosystems (PS-I)
Cholesterol, cholesterol/phospholipid/peptide complexes
Chloroplast light-harvesting complexes (LHC-II) and cytochrome <i>b</i> <sub>559</sub>
Chloroplast and bacterial ATP synthase
Complement-membrane lytic lesion (perforins)
Cytochrome oxidase
Cytochrome reductase
Desmosmal and intermediate junction proteins
Erythrocyte membrane and membrane-skeleton
Fungal (yeast) plasma membrane
Gap junction channel proteins (connexon)
<i>Halobacterium halobium</i> purple membrane/bacteriorhodopsin
Mitochondrial F <sub>1</sub> -ATPase
Mitochondrial NADH-dehydrogenase (complex I; NADH-ubiquinol oxidoreductase)
Nuclear envelope, nuclear pore complex and RNP translocation
Phospholipids and lysophospholipids (liposomes, micelles, and lipid polymorphism)
Reconstituted protein-lipid membranes
Surfactants (including saponins; ISCOMs)
Urinary bladder luminal plasma membrane

---

**Table 2** Some applications of TEM to fibril-forming and elongated proteins

---

Amyloid proteins
Bacterial flagella, flagellar basal body, pili, and fimbriae
Carcinoembryonic antigen
Cell adhesion glycoproteins
Cellulose and cellulolysis; cellulose-glucuronoxylans
Chitin
Clathrin (coated vesicles)
Collagen
Dental enamel and bone
Eukaryotic cilia/flagella, sperm tails
Elastin
Epiglycanin
Fibrinogen and fibrin
Fibronectin
Fodrin
Glycosaminoglycans
Hyaluronic acid
Intermediate filaments (vimentin, desmin and cytokeratins, neurofilaments, glial fibrillary acidic protein, and nuclear envelope lamins)
Keratin (extracellular, e.g., feathers, hair, and horn)
Laminin
Microtubules (tubulin and microtubule-associated proteins, Tau, kinesin, and dynein; mitotic spindle and pole body)
Mucins
Muscle proteins (actin, $\alpha$ -actinin, actomyosin, caldesmon, dystrophin, myosin, titin, tropomyosin)
Nuclear matrix proteins
Proteoglycans
Spectrins (erythrocyte and nonerythrocyte)

---

**Table 3** Some applications of TEM to enzymes and soluble proteins

Acetyl-CoA carboxylase	Lectins and hemagglutinins
Acetylcholine esterase	Leucine aminopeptidase
Alcohol oxidase	Lipoproteins, apolipoproteins, lipoporphins, vitellogenins, yolk proteins, chylomicrons
Alkaline phosphatase	Lumazine synthase
Aminoacyl-tRNA synthetase complex	Lysozyme
5-Aminolaevulinate synthase	$\alpha_2$ -Macroglobulin
Annexins	$\alpha$ -Mannosidase
Apo ferritin and ferritin	Methanol dehydrogenase
Arginine decarboxylase	Methylreductase
L-Asparaginase	Multicatalytic proteinase/20S and 26S proteasome
L-Aspartate- $\beta$ -decarboxylase	Neuraminidase
Casein	Nitrogenase
Catalase	Papain
Chaperones and heat-shock proteins (GroEL/ES, cpn 60/10)	
Chaperone-protein complexes	Pepstatin A polymers
Citrate lyase	Phage head-tail connector
Citrate synthase	Phosphofructokinase
Clotting factors (factor V, factor VIII)	5-Phosphoribose-1-pyrophosphate synthetase
Complement proteins	Phosphorylase
Contortin	Phosphorylase kinase
C-Reactive protein	Phycobiliproteins
Creatine kinase	Phytohemagglutinin
Crotoxin	Propionyl-CoA carboxylase
$\alpha$ -Crystalin	Protein kinase
DNA (cytosine-5-) methyltransferase	Pyruvate dehydrogenase complex
DNA B oligomer	Pyruvate carboxylase
Echinonectin	rev Protein
Excinuclease	Reverse transcriptase
Fatty acid synthetase	Riboflavin synthase
Fibronectin receptor	D-Ribulose-1,5-bisphosphate carboxylase/oxygenase (RUBISCO)
$\beta$ -Galactosidase	Ribonucleases
$\beta$ -Glucuronidase	Ribonucleotide reductase
Glutamate dehydrogenase	Seed storage proteins
Glutamine synthetase	sn-Glycerol-3-phosphate-acyltransferase
Glycogen	Soluble Mg-ATPase complexes
Glycogen synthase/kinase	Starch granules
Gramicidin S synthetase	Sucrase/isomaltase
Hemoglobins, hemocyanins, erythrocytins	Thyroglobulin
Haptoglobins	Torin
Immunoglobulins	Toxins (cholera, <i>Staphylococcus aureus</i> , tetanus)
Inclusion-body proteins	Transcarboxylase
Insulin receptor	Tripeptidylpeptidase II
Invertase	Urease
$\alpha$ -Ketoacid dehydrogenase complex	

protein molecules, and membranes. The most satisfactory results have often been obtained directly following freeze-drying of the biological material from water or volatile buffer solutions (e.g., glycerol, ammonium acetate, ammonium hydrogencarbonate). Negative staining with heavy-metal-containing salts, established throughout the 1960s, continues to be one of the principal approaches used for inducing image contrast around thinly spread particulate material. This technique does have limitations with respect to image resolution ( $\sim 1.5$  nm), but because of its wide applicability and simplicity it continues to be a valuable approach. Superior resolution has been achieved,

under low electron dose conditions, using aurothio-glucose or glucose as the 'negative' stain. Recently the author has achieved considerable success using mixtures of negative stain and trehalose.

A technical variation of conventional negative staining, termed the negative staining-carbon film technique, was introduced by Ivonné Pasquali-Ronchetti and Robert Horne in the mid-1970s. Highly purified and concentrated ( $\sim 1\text{--}5$  mg ml $^{-1}$ ) virus or protein is mixed with ammonium molybdate and poly(ethylene glycol) (PEG) and spread on freshly cleaved mica. After air-drying, the material is carbon-coated *in vacuo*, and the layer of carbon plus

**Table 4** Some applications of TEM to nucleic acids and nucleic acid–protein complexes

---

Chromatin, supranucleosomal, and nucleosome structure
DNA and RNA physical mapping (eukaryotic, prokaryotic, mitochondrial, chloroplast, viral, and bacteriophage)
DNA helix destabilizing protein gp32* I
DNA replication
<i>E. coli</i> p32–DNA complex
Histone–DNA complexes
Initiation factor
Initiation complex
Nucleolar proteins (immunolocalization)
Peptide/oligopeptide–DNA complexes
RecA–DNA complex
Replicase–DNA complex
Ribosome (prokaryotic and eukaryotic; protein immunolocalization; RNA localization)
RNA polymerase–DNA interaction
Selenocysteine synthetase–tRNA complex
SnRNP complexes
Transcription (transcription unit and gene analysis)
Transcription factor IIIA–5S rRNA complex
Transcription initiation complex
Transcription terminating factor Rho and Rho–RNA interaction
Vault RNP particles

---

adsorbed protein is floated onto the surface of aqueous negative stain (e.g., uranyl acetate, ammonium molybdate, or even water, for rapid freezing/vitrification). This procedure has enabled the production of two-dimensional (2D) arrays/crystals of the virus particles and protein molecules. From the TEM images of such 2D crystals, image processing has produced averaged reconstructions of the projection (2D) images. The extension of this approach to 3D reconstruction, from a series of tilted images, has severe problems unless the protein molecules are totally surrounded by negative stain; apparently this is not often the case with 2D crystals prepared on mica. An alternative successful approach to protein crystallization has been the formation of crystals on lipid monolayers, followed by negative staining or vitrification. This has been particularly successful for the calcium-binding annexins. The high-resolution electron-crystallographic approach to the study of protein molecules, developed initially by Aaron Klug, Nigel Unwin, Richard Henderson, and others in Cambridge, UK, using tobacco mosaic virus, thin 3D crystals of bovine liver catalase, and the naturally occurring 2D crystal of bacteriorhodopsin in the purple membrane of *Halobacterium halobium*, has continued strongly in several parts of the world. High-resolution studies have been performed with experimentally produced 2D crystals of membrane proteins and thin 3D crystals of soluble protein, such as crotoxin. Integral membrane proteins have been crystallized after solubilization and purification in

the presence of neutral surfactants, usually followed by controlled removal of surfactant by dialysis or conventional salt or PEG crystallization, and replacement of the biological lipids. Considerable success has been achieved using this approach during the study of *Escherichia coli* PhoE porin by Bing Jap in Berkeley and for the chloroplast light-harvesting complex by Werner Kühlbrandt in Heidelberg. The use of spot-scan imaging of crystalline specimens has been introduced to reduce the electron beam-induced radiation damage and specimen mobility while performing photographic image recording, thereby leading to superior resolution.

In parallel with the crystal studies, procedures for single-particle classification and structural analysis have been developed by Joachim Frank, Marin van Heel, and others, with which considerable progress has been made. A good example of this approach is the understanding of ribosome structure, which has led to the localization of individual ribosomal proteins and RNA within this important biological particle. The production of unstained vitrified/frozen-hydrated specimens of a wide range of biological materials, pioneered by Jacques Dubochet and his colleagues in Lausanne, has now become a widely used technique, which maintains the hydration state of the biological material and avoids the interaction with heavy-metal stains. The resolution obtainable from vitrified specimens is usually not greatly superior to that from negative staining. The two techniques tend to provide images that are very much in agreement, but vitrification has greater possibilities for the study of dynamic biological systems, such as microtubule formation and disassembly, actin–myosin interactions, and calcium-induced structural changes in gap junction channels.

## Applications

In view of the extremely large number of current applications of TEM to the study of isolated biological materials, quite apart from the extensive tissue-based ultrastructural studies, only a very limited number of examples can be presented, and, for brevity, several other significant applications are tabulated (Tables 1–4). Individual animal, plant, and bacterial viruses are not listed, because of the accepted applicability of TEM to the analysis of viral structure; some references to the viral literature are, however, included in the Further Reading section.

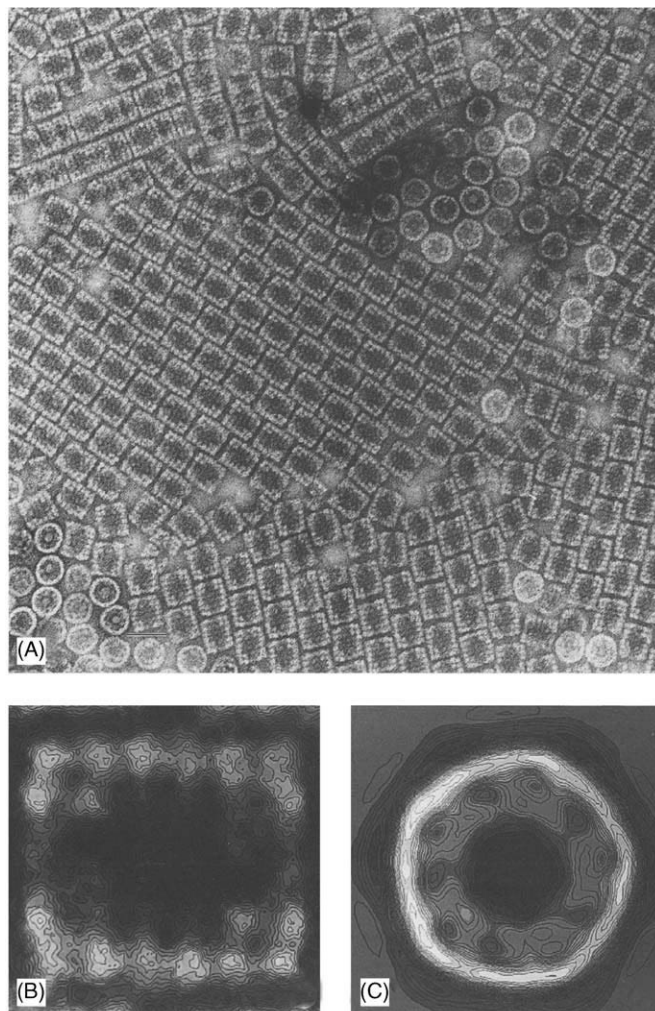
Examples of the application of TEM to the study of respiratory proteins are shown in Figures 1 and 2. The first example (Figure 1A) of negatively stained key-hold limpet (*Megathura crenulata*) hemocyanin (KLH) shows a 2D crystal of the side-on image

projection of this giant cylindrical protein (the didecamer; relative molecular mass  $\sim 7$  MDa) produced by the negative staining–carbon film procedure. The crystallographically averaged 2D reconstruction of this image projection has been obtained by computer processing (Figure 1B). The KLH didecamer has a fivefold rotational symmetry, which tends to interfere with formation of 2D crystals when the cylindrical molecule is orientated end-on. Nevertheless, an averaged reconstructed image can be obtained by single-particle analysis (Figure 1C).

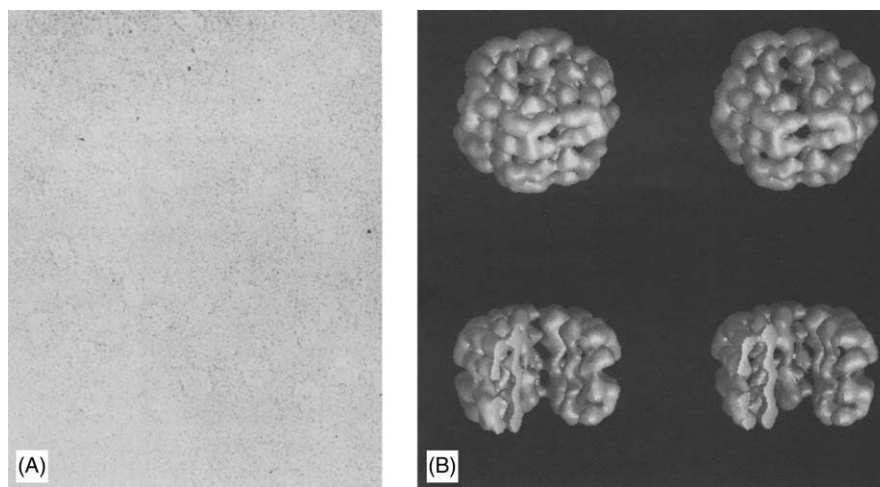
Annelid hemoglobin is also a giant macromolecular assembly, with a relative molecular mass of  $\sim 3$  MDa. Extensive negative staining studies have been performed on this macromolecule, which tend to reveal the same type of stable orientations of the

molecule (side-on/end-on) and image projections as with KLH. However, when preserved unstained in vitreous ice (Figure 2A) the hemoglobin presents a greater range of intermediate images, because of the freely varying rotational angles of the molecule relative to the electron beam, within the surrounding ice. From computer classification and image processing of these different 2D images, a single-particle 3D reconstruction has been obtained (Figure 2B).

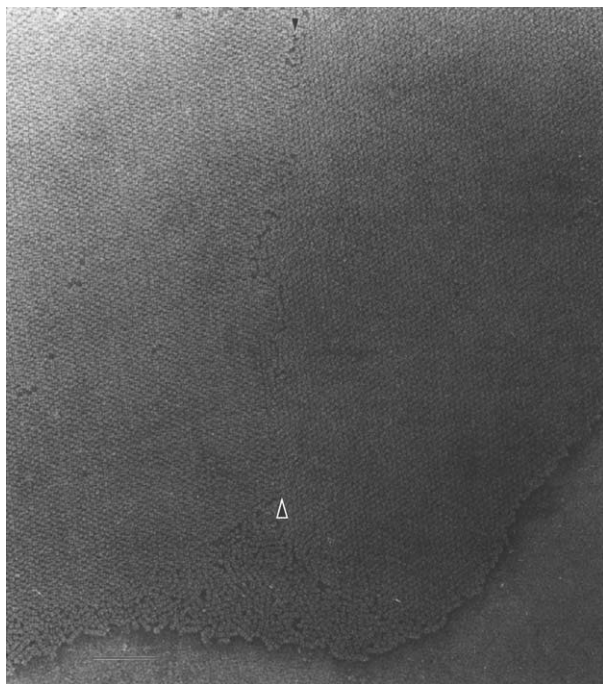
Application of the negative staining–carbon film procedure to the much smaller human erythrocyte catalase (relative molecular mass 256 kDa) has resulted in the formation of several different paracrystalline and truly crystalline 2D forms, an occurrence also encountered with the *E. coli* chaperone GroEL. One of the catalase 2D crystal forms is



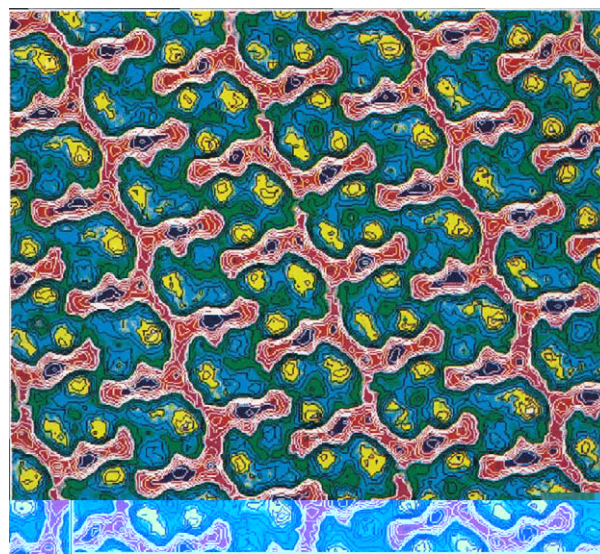
**Figure 1** (A) A 2D crystal of keyhole limpet haemocyanin (KLH) (didecamer) with molecules orientated side-on, produced by the negative staining–carbon film procedure. Bar = 50 nm. (B) The reconstructed 2D image of KLH produced by crystallographic image processing, without symmetrization. (C) The reconstructed 2D image of the end-on KLH molecule, produced by single molecule averaging (without symmetrization). The fivefold rotational symmetry of the molecule is apparent. (Reproduced with permission from Harris JR *et al.* (1992) *Micron and Microscopica Acta* 23: 287–301.)



**Figure 2** (A) A noisy transmission electron micrograph of vitrified earthworm hemoglobin and (B) the 3D single particle reconstruction produced with the IMAGIC-5 image processing software. Note in (A) the faint images of the hemoglobin molecules at different orientations within the vitreous ice, due to varying molecular rotation with respect to the axis of the electron beam. (Micrographs kindly provided by Dr. Marin van Heel, Fritz-Haber Institute, Berlin.)



**Figure 3** Two-dimensional crystals of human erythrocyte catalase, produced by the negative staining–carbon film procedure. These crystals possess a  $p22_12_1$  symmetry. Note the group of free molecules, the line of dislocation (between arrow heads), and the occasional absence of molecules within the crystal lattice. Bar 100 nm. Compare with the crystallographic 2D color image reconstruction shown in Figure 4, which emphasizes the four-molecular row repeat across the crystal lattice. Image processing was performed with the assistance of S. Volker and Dr. H. Engelhardt, Max-Planck-Institute for Biochemistry, Martinsried, Germany, using the Semper software system. (Modified from Harris JR *et al.* (1993) *Journal of Structural Biology* 111: 22–33.)



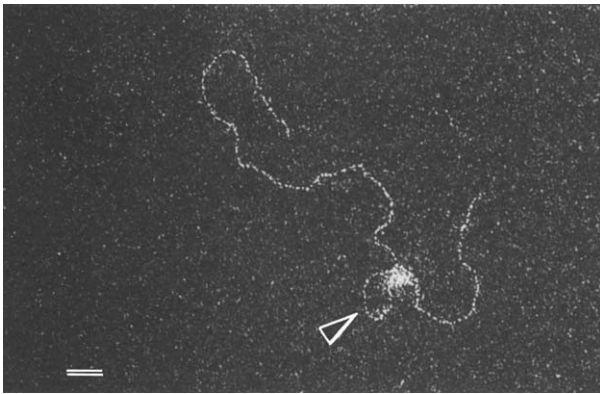
**Figure 4** The crystallographic reconstruction produced by image processing from the  $p20$  2D crystal form of human erythrocyte catalase shown in Figure 3. Protein is denoted by green, light blue, and yellow. Negative stain (uranyl acetate) is represented by red and dark blue, with dark blue for the regions of deepest stain. Note the four-row repeat across the crystal lattice, due to highly specific molecular orientation within the 2D crystal. (Reproduced with permission of Harris JR, Johannes Gutenberg-Universität Mainz, Germany.)

shown in Figure 3; the corresponding crystallographic image reconstruction is illustrated in Figure 4. The phenomenon of 2D crystal and paracrystal diversity should be considered perfectly normal, since it is known that many proteins produce multiple 3D crystal forms. The future study of 2D and



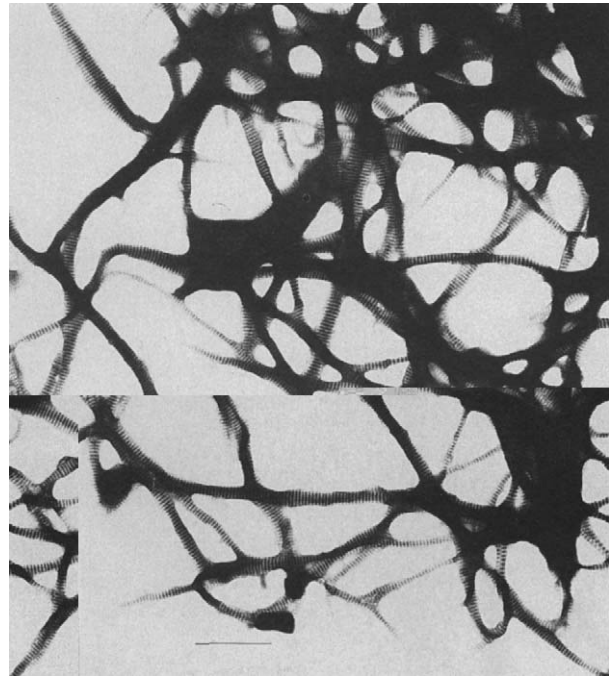
3D protein associations, crystal nucleation, and subsequent crystal formation during the negative staining–carbon film procedure would appear to hold considerable potential.

Although many studies on DNA and DNA–protein interaction have utilized metal shadowing to produce bright-field image contrast, dark-field TEM (annular or spectroscopic) of unstained or uranyl acetate-stained and metal-shadowed nucleic acid–protein complexes also provides useful resolution. In **Figure 5**, a short length of DNA (2356 base pairs), with attached progesterone receptor, is revealed by dark-field TEM. The RecA protein has the ability to

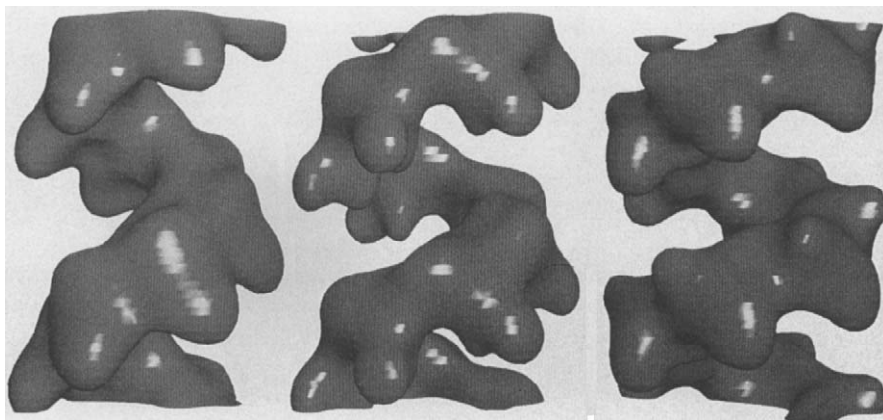


**Figure 5** DNA (2356 base pairs long), representing the regulating region of the uteroglobin gene, and containing two hormone-responsive elements (HREs). Two closely apposed progesterone receptors are bound to these HREs, separated by 260 base pairs, giving rise to a DNA loop (arrow head) due to the interaction between the two receptors. Bar = 50 nm. The DNA–protein image was produced by annular dark-field TEM. (The electron micrograph was taken by A. Barbin and kindly provided by Dr. E. Delain, CNRS and Institute Gustave-Roussy, Villejuif, France.)

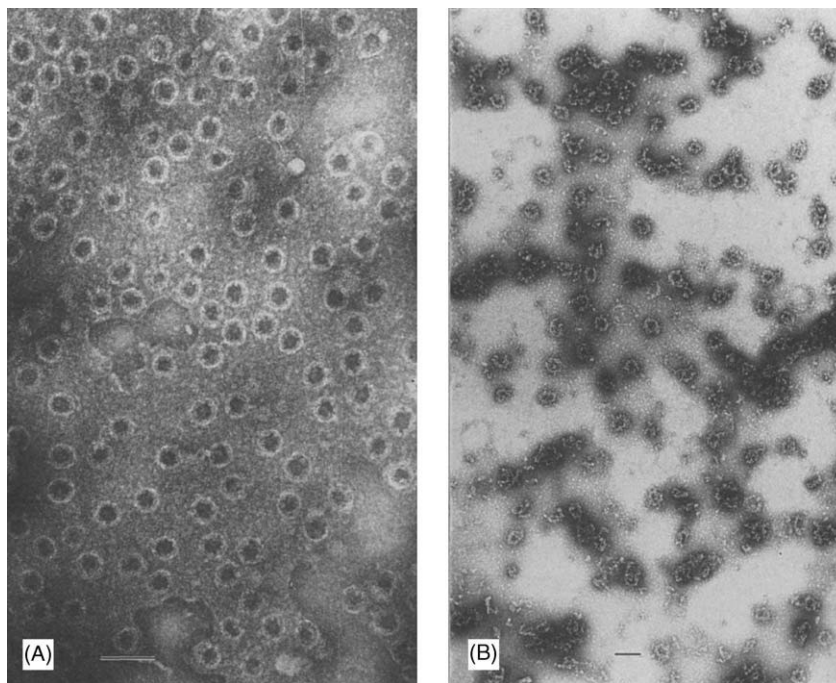
complex with single- and double-stranded DNA, to produce an organized linear array of the RecA on the DNA. Such complexes have been revealed by negative staining, metal shadowing, and in the unstained frozen–hydrated state by vitrification. In **Figure 6**,



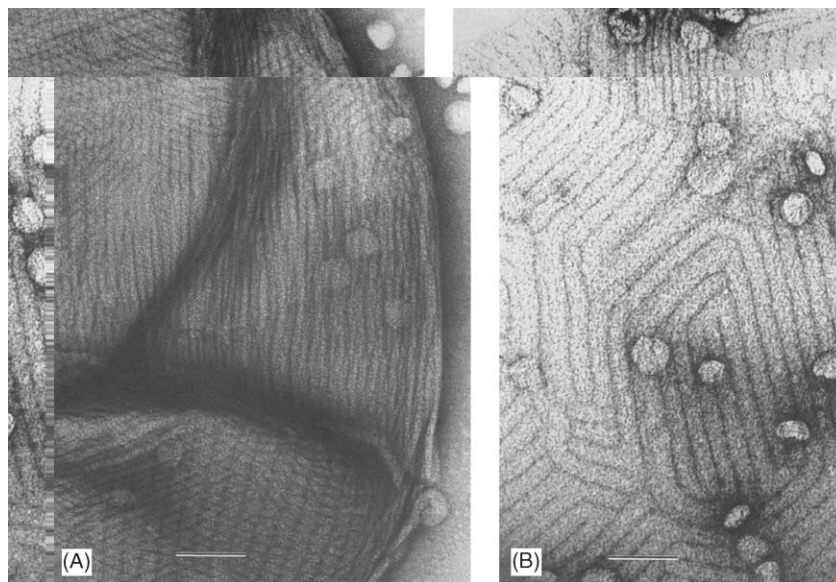
**Figure 7** Part of a large paracrystalline aggregate of *Xenopus* lamin-A, expressed in *E. coli*. Note the characteristic 25 nm periodicity along the filamentous aggregate. Bar = 0.5  $\mu$ m. (The sample material was kindly provided by Gieffers C and Krohne G and the specimen was produced from a solution of protein in 2 mol l<sup>-1</sup> ammonium acetate by the negative staining–carbon film procedure.)



**Figure 6** A comparison of the rendered surfaces of TEM reconstructions of the active RecA–DNA filament (left) and the inactive RecA–DNA filament (right) with the RecA crystal surface (center). The RecA crystal surface, obtained by X-ray crystallography, has been rendered at low resolution, to be comparable with the TEM reconstructions. Both the active and inactive filament reconstructions are from averages of frozen–hydrated specimens imaged with cryo-EM. (Reproduced with permission from Egelman EH and Stasiak A (1993) Electron microscopy of RECA–DNA complexes: Two different states, their functional significance and relation to the solved crystal structure. *Micron* 24: 309–324.)



**Figure 8** (A) Isolated rat liver nuclear envelope, negatively stained with ammonium molybdate. Note the pronounced electron-transparent octagonal annuli of the nuclear pore complexes. (B) A partly purified nuclear pore complex fraction obtained by Triton X-100 extraction and ultrasonication of nuclear envelope, followed by sucrose density gradient centrifugation. Bars = 200 nm.



**Figure 9** The  $P\beta'$  phase of bovine brain sphingomyelin revealed by negative staining with uranyl acetate on (A) a multilamellar body and (B) a single bilayer. Liposome particles are present on the surface of the larger lipid bilayer structures. Bars = 100 nm. (Reproduced with permission from Harris JR (1986) *Micron and Microscopica Acta* 17: 175–200.)

reconstructed rendered surfaces obtained by image processing from vitrified specimens of the active RecA–DNA filament and the inactive RecA–DNA filament are compared with the RecA crystal data.

The filamentous lamin proteins, in a manner similar to collagen,  $\alpha$ -tropomyosin, and intermediate

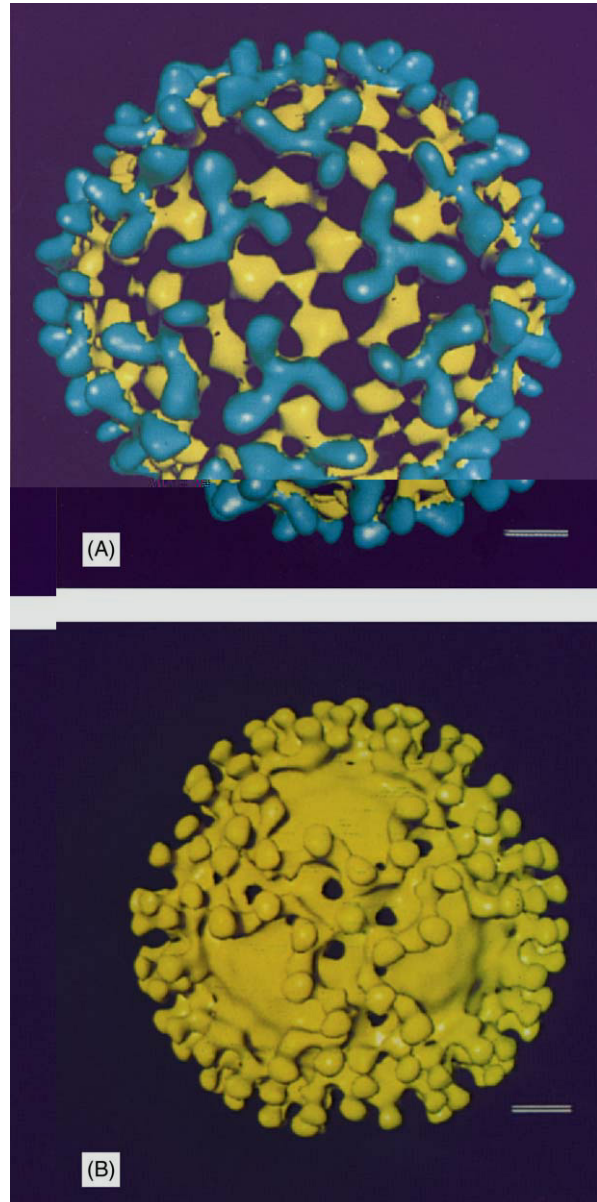
filament fragments, have the ability to form linear multimolecular paracrystalline aggregates by treatment under appropriate ionic conditions *in vitro*. These aggregates exhibit characteristic periodicities determined by the specific intermolecular interactions and repeating linear overlap of the extended peptide chains. One such example of a paracrystal of

bacterially expressed lamin A that possesses ~25 nm periodicity is shown in **Figure 7**.

TEM has contributed to the understanding of the structure of many bacterial cell wall and eukaryotic cellular membrane systems, *in situ* and after isolation as purified subcellular fractions. The nuclear envelope, with its integral and ubiquitous pore complexes, has received much attention over a period of many years. Useful data have been obtained from all the available specimen preparation techniques, including vitrification, from which detailed models of the nuclear pore complex have been presented in recent years. In addition, high-resolution SEM has provided supplementary data that has extended the interpretation available from TEM alone. The functional role of the nuclear pore complex in translocation of proteins and ribonucleoproteins to and from the nucleus is currently being actively addressed using ultrastructural as well as cell-biological approaches. An example of nuclear envelope, isolated from rat liver by biochemical treatments and centrifugation procedures, and subjected to negative staining, is shown in **Figure 8A**. The ~100 nm outer annuli of the nuclear pore complexes are revealed with great clarity, but it is likely that this type of isolated material has been subjected to some morphological disruption during the isolation steps and has been flattened to some extent by the specimen preparation procedure. Nuclear pore complexes extracted by neutral surfactant and ultrasonication and partly purified by density gradient centrifugation are shown in **Figure 8B**. The ultimate goal of producing the nuclear pore complex as an 'intact' and highly purified macromolecular assembly, by biochemical subfractionation of the nuclear envelope, has recently been achieved and valuable information on its subcomponents has already emerged, particularly from the studies of Ueli Aebi, in Basel. In addition, an *in vitro* system for nuclear envelope and nuclear pore complex formation has recently become available.

Aqueous liposomal and micellar suspensions of phospholipid have been studied by TEM for a number of years. The characterization of the single and multiple lipid bilayer nature of liposome particles was greatly assisted by early negative staining TEM studies and was supported by freeze-fracture analysis. In recent years, vitrified specimens have also been produced, which avoid the problems of liposome-stain interaction and distortion (flattening) that may be present in the negatively stained specimens. The study of lipid phase changes and the dynamic membrane fusion event in liposome and reconstituted lipid-protein membrane systems can be studied with some degree of accuracy by TEM. The former (a relatively slow event) can be studied

by freeze-fracture and negative staining, but the latter cannot; it is in the area of lipid bilayer fusion that vitrification of unstained liposomes has made a significant contribution. An example of the undulatory P $\beta$ ' phase of sphingomyelin is shown in **Figure 9**. The production and removal of this lipid phase with



**Figure 10** Reconstructions of Blue tongue virus (BTV) and core-like particle obtained from frozen-hydrated specimens. (A) The surface representation of the outer shell of the reconstructed virus particle viewed along a threefold axis. The globular regions (VP5) are colored yellow and the sail-shaped spikes (VP2) are blue. Bars = 10 nm. (Reproduced from Hewat EA, Booth TF, and Roy P (1992) *Journal of Structural Biology* 109: 61–69.) (B) The surface representation of the BTV core-like particle viewed down a threefold axis from outside the particle. Bar = nm. (Modified from Hewat EA, Booth TF, Loudon PT, and Roy P (1992) *Virology* 189: 10–20.)

increasing temperature has been studied by incubating the lipid suspension and preparing the negatively stained specimens under carefully controlled conditions of incremental temperature.

As the discussion throughout this article has essentially neglected reference to the vast body of important data relating to viral structure, and because there is no separate article within the present Encyclopedia devoted to the analysis of viral structure, an example of the application of TEM to the study of frozen-hydrated Blue tongue virus is given in the color plate section. As with the other areas of biomolecular TEM application discussed above, large contributions to the knowledge of viral particles have come from negative staining and metal shadowing, data that are currently being essentially confirmed and extended by the use of frozen-hydrated/vitrified specimens. It should also be emphasized that conventional thin sectioning of cells and tissues for TEM has made a major contribution to the study of viral morphogenesis, structure, cellular infection, and release (Figure 10).

**See also:** **Microbiological Techniques. Microscopy Techniques:** Specimen Preparation for Electron Microscopy. **Nucleic Acids:** Spectroscopic Methods.

## Further Reading

Amos LA, Henderson R, and Unwin PNT (1982) Three-dimensional structure determination by electron

microscopy of two dimensional crystals. *Progress in Biophysics and Molecular Biology* 39: 183–231.

Cam E le, Theveny B, Mignotte B, Revet B, and Delain E (1991) Quantitative electron microscopic analysis of DNA–protein interactions. *Journal of Electron Microscopy Technique* 18: 375–386.

Dubochet J, Adrian M, Chang J-J, *et al.* (1988) Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics* 21: 129–228.

Frank J (1989) Image analysis of single molecules. *Electron Microscopy Reviews* 2: 53–74.

Harris JR (1981–1987) *Electron Microscopy of Proteins*, vols. 1–6. London: Academic Press.

Harris JR (1991) *Electron Microscopy in Biology: A Practical Approach*. Oxford: IRL Press.

Kiselev NA, Sherman MB, and Tsuprun VL (1990) Negative staining of proteins. *Electron Microscopy Reviews* 3: 43–72.

Kühlbrandt W (1992) Two-dimensional crystallization of membrane proteins. *Quarterly Reviews of Biophysics* 25: 1–49.

Müller SA, Goldie KN, Bürki R, Häring R, and Engel A (1992) Factors influencing the precision of quantitative scanning transmission electron microscopy. *Ultramicroscopy* 46: 317–334.

Nermut MV and Steven AC (1987) *Animal Virus Structure: Perspectives in Medical Virology*, vol. 3. Amsterdam: Elsevier.

Sommerville J and Scheer U (1987) *Electron Microscopy in Molecular Biology: A Practical Approach*. Oxford: IRL Press.

Würz M (1992) Bacteriophage structure. *Electron Microscopy Reviews* 5: 283–310.

# MICROSCOPY TECHNIQUES

## Contents

### Light Microscopy

#### Sample Preparation for Light Microscopy

#### Electron Microscopy

#### Specimen Preparation for Electron Microscopy

#### Scanning Electron Microscopy

#### Atomic Force and Scanning Tunneling Microscopy

#### X-Ray Microscopy

## Light Microscopy

**C Hammond and P J Evennett**, University of Leeds, Leeds, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Michael C Edwards, pp. 3115–3120, © 1995, Elsevier Ltd.

## Introduction

There are several basic types of light microscopes, each appropriate to a particular application. The basic function of a microscope to study the structure of a specimen can be extended to a range of analytical functions. These include measurement and image analysis, and by using techniques such

increasing temperature has been studied by incubating the lipid suspension and preparing the negatively stained specimens under carefully controlled conditions of incremental temperature.

As the discussion throughout this article has essentially neglected reference to the vast body of important data relating to viral structure, and because there is no separate article within the present Encyclopedia devoted to the analysis of viral structure, an example of the application of TEM to the study of frozen-hydrated Blue tongue virus is given in the color plate section. As with the other areas of biomolecular TEM application discussed above, large contributions to the knowledge of viral particles have come from negative staining and metal shadowing, data that are currently being essentially confirmed and extended by the use of frozen-hydrated/vitrified specimens. It should also be emphasized that conventional thin sectioning of cells and tissues for TEM has made a major contribution to the study of viral morphogenesis, structure, cellular infection, and release (Figure 10).

**See also:** **Microbiological Techniques. Microscopy Techniques:** Specimen Preparation for Electron Microscopy. **Nucleic Acids:** Spectroscopic Methods.

## Further Reading

Amos LA, Henderson R, and Unwin PNT (1982) Three-dimensional structure determination by electron

microscopy of two dimensional crystals. *Progress in Biophysics and Molecular Biology* 39: 183–231.

Cam E le, Theveny B, Mignotte B, Revet B, and Delain E (1991) Quantitative electron microscopic analysis of DNA–protein interactions. *Journal of Electron Microscopy Technique* 18: 375–386.

Dubochet J, Adrian M, Chang J-J, *et al.* (1988) Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics* 21: 129–228.

Frank J (1989) Image analysis of single molecules. *Electron Microscopy Reviews* 2: 53–74.

Harris JR (1981–1987) *Electron Microscopy of Proteins*, vols. 1–6. London: Academic Press.

Harris JR (1991) *Electron Microscopy in Biology: A Practical Approach*. Oxford: IRL Press.

Kiselev NA, Sherman MB, and Tsuprun VL (1990) Negative staining of proteins. *Electron Microscopy Reviews* 3: 43–72.

Kühlbrandt W (1992) Two-dimensional crystallization of membrane proteins. *Quarterly Reviews of Biophysics* 25: 1–49.

Müller SA, Goldie KN, Bürki R, Häring R, and Engel A (1992) Factors influencing the precision of quantitative scanning transmission electron microscopy. *Ultramicroscopy* 46: 317–334.

Nermut MV and Steven AC (1987) *Animal Virus Structure: Perspectives in Medical Virology*, vol. 3. Amsterdam: Elsevier.

Sommerville J and Scheer U (1987) *Electron Microscopy in Molecular Biology: A Practical Approach*. Oxford: IRL Press.

Würz M (1992) Bacteriophage structure. *Electron Microscopy Reviews* 5: 283–310.

# MICROSCOPY TECHNIQUES

## Contents

### Light Microscopy

#### Sample Preparation for Light Microscopy

#### Electron Microscopy

#### Specimen Preparation for Electron Microscopy

#### Scanning Electron Microscopy

#### Atomic Force and Scanning Tunneling Microscopy

#### X-Ray Microscopy

## Light Microscopy

**C Hammond and P J Evennett**, University of Leeds, Leeds, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Michael C Edwards, pp. 3115–3120, © 1995, Elsevier Ltd.

## Introduction

There are several basic types of light microscopes, each appropriate to a particular application. The basic function of a microscope to study the structure of a specimen can be extended to a range of analytical functions. These include measurement and image analysis, and by using techniques such



as histochemical staining, the identification and localization of specific chemical components. With appropriate preparations, living processes can be studied and recorded. A temperature-controlled stage may be used to study temperature-dependent changes such as melting or freezing, and, with the addition of differential scanning calorimetry (DSC), energy changes associated with the chemical or physical changes can be directly related to the observed structures.

## Simple Microscope

At its simplest, the microscope is a single lens, consisting of a single glass component, or several components cemented together to reduce spherical and chromatic aberrations. It is generally used as a hand lens for low-magnification examination. Alternatively, the lens may be mounted on a simple stand to free the user's hands for various purposes (e.g., for dissection). Such instruments are cheap and simple, but suffer from the twin disadvantages of low magnification (up to  $25\times$ ) and the fact that they must be brought very close to the eye to achieve a satisfactorily large field of view. All other microscope types achieve higher magnifications and greater ease of use by applying lenses in combination (objective and eyepiece lenses) to give a compound microscope.

## Stereomicroscope

A stereomicroscope, sometimes called a dissecting microscope or a binocular inspection microscope, is a low-power compound instrument used for a closer examination of three-dimensional specimens than is possible with a hand lens (**Figure 1**). It gives a high-quality image with a full three-dimensional effect, produced by viewing the specimen with both eyes from slightly different viewpoints. It usually has a considerable depth of field and a long working distance, giving easy mechanical interaction with the specimen. There is no lateral reversal of the image as with most other compound microscopes, i.e., the image is upright, enabling spatial relationships within the specimen to be easily appreciated.

Many early stereomicroscopes were of the twin-objective or Greenough type, having quite separate optical systems for each eye. Because of the restricted field of view in focus at any one time in that system, it has been largely superseded by instruments having a common large main objective lens, with separate lens systems mounted above for each eye. Older instruments of this type may have a rotating turret of objectives of fixed magnification, whilst most new



**Figure 1** A modern zoom stereomicroscope, for observation of samples in incident light. (Courtesy of Olympus UK Ltd., used by permission.)

examples have a continual adjustment of magnification or zoom lens. Current top-of-the-range instruments may also offer other illumination facilities such as dark field. Available overall magnification is typically up to about  $50\times$ , although common main objective lenses can be replaced with more powerful ones to give overall magnifications of  $100\times$  or more, at the cost of a much reduced working distance and depth of field. The main disadvantage of the common main objective design is that the numerical aperture of the system is limited by the presence of the double beam path above. This has led to the development of the 'macroscope', an instrument with a large diameter objective giving a greatly improved numerical aperture. An alternative solution, which loses the three-dimensional effect of the stereomicroscope, is the 'deep-field microscope' described later.

## Compound Microscope

The transmitted-light compound microscope was first developed for use by biologists, and this remains main use of the microscope, although there are also many applications in fields such as mineralogy, crystallography, and materials science. Because light

is transmitted through the specimen, this is necessarily relatively thin, and thin sections mounted on a glass slide with a coverslip are used for most applications.

### General Arrangement

There are many variations in microscope design but the basic components are usually the same (**Figure 2**). The microscope stand carries a stage at right angles to the optical axis, a tube carrying the objective lenses and eyepieces, and an illumination system consisting of a condenser and a lamp. The parts may be arranged in different ways on different models.

Most modern microscopes are binocular, meaning that the 'same' image is seen by both eyes, reducing user fatigue. Whilst early microscopes had the optical axis tilted toward the user, the optical axis is now normally arranged vertically, with prisms beneath the eyepieces to enable them to be tilted to provide a comfortable viewing angle. Many microscopes also provide an extra tube for the attachment of a camera.

The objective lenses are usually carried on a rotating nosepiece that holds several lenses. Most

objectives for modern microscopes are multielement assemblies, and are designed to correct chromatic aberrations by bringing different colors (wavelengths) to a common focus. Plano-objectives or 'flat field' objectives also compensate for field curvature, bringing the whole field of view into focus; this again reduces user fatigue. Objectives from the same manufacturer should normally be parfocal, i.e., focus at the same position. They are normally designed to work with 0.17 mm thick glass coverslips; any other coverslip thickness may result in some aberration of the image. Lower-magnification objectives are designed to work 'dry', with air between the objective and coverslip. Higher-magnification objectives – marked 'OIL', 'OEL', or 'HI' (homogeneous immersion) – are designed to work with immersion oil of a specific refractive index between lens and coverslip.

Whilst earlier microscopes used an adjustable mirror to reflect light from an external source into the optical axis of the microscope, modern instruments have a built-in lamphouse. The light is focused on the object plane by means of a condenser, an assembly of lenses that matches the light beam to the numerical aperture of the objective.

The microscope stage carries an object slide with provision for holding the slide in place. On earlier microscopes the stage was fixed and the microscope tube was moved up and down to focus the image. More recently, the microscope tube is generally fixed to the stand and the stage is moved up and down. However, both arrangements (fixed and moving stage) are in use.

A specialized form of the transmitted-light compound microscope is the 'inverted microscope' (**Figure 3**), used for objects such as cell cultures, in which the microscope tube and optics are mounted below the specimen, which in turn is illuminated from above. A prism is used to incline the eyepieces at a comfortable viewing angle. Another specialized form is the 'portable microscope', of which the McArthur instrument is the best known. These allow a compound microscope to be taken out into the field.

### Illumination Methods

The conventional transmitted-light compound microscope uses a bright-field system, in which the condenser gives a bright, even illumination across the whole of the field of view. This form of illumination, when correctly set up, is ideal for the examination of a wide range of specimens, particularly stained thin sections of biological material. Where there are no natural variations in color or optical density of the



**Figure 2** A upright transmitted-light compound microscope with binocular eyepieces and an illuminator for epi-fluorescence. (Courtesy of Olympus UK Ltd., used by permission.)



**Figure 3** An transmitted-light inverted microscope. (Courtesy of Olympus UK Ltd., used by permission.)

specimen, and where staining methods cannot be used, for instance with live preparations, other means must be used to impart some form of contrast to the image.

Dark-field or dark-ground illumination is used for objects with low contrast in bright-field illumination and renders images showing a bright specimen against a dark background. The specimen is illuminated with a hollow cone of light that cannot enter the objective lens. The image is formed by light that is diffracted or otherwise scattered by the object, which therefore appears in reverse contrast. This method has largely been replaced by phase-contrast microscopy for weakly scattering objects.

Phase-contrast microscopy was developed to observe living cells without the need for the normal fixing, mounting, and staining procedures that would kill them. Under bright-field illumination, living cells cannot easily be differentiated from the surrounding liquid, because the refractive indices of cells and water are very similar. Phase-contrast microscopes render these unstained cells visible with enhanced contrast of internal details using special objectives with annular phase plates. The condenser also has a series of annular diaphragms, each matched to the phase plate of a particular objective. The phase plate

produces a quarter-wavelength ( $90^\circ$ ) shift of phase. For best contrast, this wavelength must be defined, and so a color filter is often used. The color chosen is normally green, since this lies in the middle of the visible spectrum and hence the optimum wavelength for both the human eye and the microscope optics. Direct light is separated from diffracted light, producing interference and resulting in a dark image on a bright background or vice versa. The dark image on a bright background, known as dark or positive contrast, is used for the observation of the internal cell structure. Phase contrast often gives rise to a 'halo' around features of varying contrast in the image.

Another system for examination of low-contrast objects such as living cells is the Nomarski or differential interference contrast system. It is also particularly useful for materials that cannot be stained satisfactorily for other reasons, such as very thin sections that take up too little stain. This system employs polarizing filters and quartz prisms instead of the annular diaphragm and phase plates used in phase contrast. This eliminates the halo effect seen in phase contrast, rendering sharply defined images with good contrast, having a characteristically (pseudo) three-dimensional appearance. It is rather less suited to routine work than phase contrast however, and is considerably more expensive.

An extremely powerful analytical tool in light microscopy is observation of the specimen in polarized light, in which a piece of polaroid material, called a polarizer, is placed in the illumination path. Light waves vibrate at all planes at right angles to the direction of the ray. Polaroid material will pass only those rays vibrating in one plane at right angles to their direction. If a polarizer made of this material is placed in the light path before the specimen, most specimens will simply appear less bright. If a second piece of polaroid material, called the analyzer, is placed in the light path after the specimen and at right angles to the polarizer, the field of view will appear black except where the specimen has the ability to change the plane of vibration of the light. The specimen is then said to be viewed 'between crossed polars'. This is a property of most, but not all, crystalline materials. Some materials that are not normally thought of as crystalline, including many biological materials, plastics, and liquids (liquid crystals) also display this property. Some samples may produce a range of colors, known as birefringence colors, which can be important clues to the identity of many plastics, fibers, and minerals. It is usual for the analyzer to be mounted in a slide that can be pushed into or out of the light path but for the polarizer to be mounted so that it can be turned through an angle of more than  $90^\circ$ , so that the

changing effects can be observed. A polarizing microscope is also usually fitted with a rotating stage calibrated in degrees to allow the specimen itself to be rotated in relation to the angle of the polarizer and analyzer.

A development of the use of polarized light is the differential staining objective, which is used to impart optical colors to specimens mounted in a liquid of slightly different refractive index. Some specimens, including some artificial fibers and mineral fibers such as asbestos, have different refractive indices at different orientations to the plane of polarized light.

All the illumination methods discussed above are based on transmitted illumination, in which the light passes through the specimen from one side to the other. The alternative to this is incident-light (or epi-) illumination, in which the light passes down the objective before being reflected back through the objective to the eyepieces. Hence, no separate condenser system is required, and the illuminating light beam is always perfectly matched to the objective. These instruments are necessary for the examination of necessary opaque specimens and therefore are chiefly used in materials science, but there are increasing numbers of biological applications, such as the examination of leaf surfaces. The light beam enters the microscope at right angles to the optical axis and is turned through 90° using a dichroic mirror mounted at 45° to the optical axis. The objectives of these microscopes are designed for use on uncovered surfaces, i.e., without coverslip correction. Incident illumination is sometimes combined with the use of polarized light. A particular application of this is the observation of silver grains in autoradiographs, in which only the light reflected from the silver grains is seen against a dark background: the light reflected can be measured using a photometer and related to the number of silver grains present. A similar approach can be used with immunogold-labeled preparations.

The form of incident illumination most widely used in biological applications is fluorescence microscopy, in which light of a relatively short wavelength, often ultraviolet (UV) light, is passed down the objective and excites fluorescence at a longer wavelength in the specimen.

A recent and very sophisticated development of incident illumination is the confocal laser scanning microscope, in which a laser beam is scanned through the specimen, focused at varying depths within it. The signal obtained from the specimen is used to build up a picture on a computer monitor of an 'optical section' at a given depth. A three-dimensional reconstruction of a relatively thin specimen can be built up on the monitor to study variations in structure through its thickness.

## **Analysis Using Light Microscopes**

### **Measurement of Size, Area, and Distribution**

The magnification achieved by the microscope can be calculated accurately and linear measurements can be made by calibrating a graduated scale in the microscope eyepiece against a measurement scale marked on a microscope slide (a stage micrometer). With a camera mounted on the microscope, more complex measurements can be made on the photographs obtained. However, the fitting of a camera to the microscope and subsequent digitization of the image has led to the development of a wide range of image analysis equipment and software for the assessment of the image obtained.

A wide range of parameters can be assessed, including the lengths, widths, areas, and distributions of specific features. Once digitized, the image can be computer enhanced to aid its interpretation and analysis. The measurement of the brightness of light transmitted through the specimen (photometry) or the degree to which density of the specimen or a stain used to color it obscures the light (densitometry) can be used to assess the numbers of particular features in the field of view. Particle size analysis can also be carried out to give a size distribution curve.

### **Chemical Reactions and Structural Changes**

A temperature-controlled stage can be used to study the behavior of specimens at different temperatures or to measure and observe melting or freezing points. A modified version of such a stage allows the combination of microscopy with DSC to relate energy changes in the sample to visual changes such as starch gelatinization or the change of crystal states in fats.

### **Histochemistry**

An important application of the light microscope in biological work is the identification and localization of particular chemical species in microscope sections by the use of histochemical stains. More recently, fluorescent stains have been developed that are highly specific.

### **Confocal Microscopy**

Confocal scanning laser microscopy has been combined with the use of fluorescent stains to measure the distribution and concentration of specific analytes within biological tissue. Stains are used to bind selectively with the analyte under study, and the correct laser beam is selected to excite the stain used to give a fluorescent signal. The technique can also be

used to measure the thickness of translucent layers in a sample such as the enamel layer on teeth.

### Refractive Index Measurement

The refractive indices of glasses, minerals, plastics, and many other samples may be measured by a range of microscopical methods. The simplest of these is the examination of samples mounted in a range of liquids of which the refractive index is known until the sample-liquid interface becomes invisible. A development of this method involves the use of a microscope hot stage to raise the temperature of the liquid and sample gradually. Because in general the refractive index of the sample changes more slowly than that of the liquid as the temperature rises, a specific 'match temperature' can be determined. Either of these methods can be aided by observation of the Becke line, a bright fringe that appears just inside or outside the boundary of the fragment when it is almost in focus. The Becke line appears to move toward the medium of higher refractive index when the distance between specimen and objective is increased, thus indicating whether the mounting liquid has a higher or lower refractive index than the fragment.

## Other Microscopes

### Deep-Field Microscopes

One of the major limitations of conventional light microscopy is the characteristically shallow depth of the field, so that only a limited plane can be observed at one time. The depth of field can be extended by decreasing the lens aperture, but this results in loss of resolution and too little light passing through. This problem is overcome in the deep field microscope, often known as a video microscope, by applying an adjustable beam of highly intense light to the area under illumination. The object is observed using a subminiature color camera held between the fingers like a pencil torch. The achievable depth of field is extended further using two objective lenses, one of which is used to magnify while the other is free to focus the image on the second lens from a long working distance. The image is viewed on a TV monitor. The result is comparable with the depth of field provided by the scanning electron microscope.

### Laser Doppler Microscopy

The laser Doppler microscope is a development of the fluorescence microscope designed to enable the rate of movement of particles to be measured in

living specimens, using a laser beam to provide epi-illumination through the objective lens. The laser light is scattered by moving particles, the angle of scattering depending on particle size, the viscosity of the matrix surrounding the particles, and temperature. When the original laser frequency is mixed with the scattered frequency, a beat frequency is obtained for uniform motion of particles, and a characteristic spectrum of frequency for random Brownian motion. The technique has been applied successfully to the measurement of plastid movement in plant phloem sieve tubes.

### Ultraviolet Microscopy

Examination of specimens in UV light requires the use of optical components capable of transmitting such short wavelengths, which means the use of quartz rather than glass lens components. However, when combined with a UV spectrophotometer, this enables the analysis of small areas of a sample by UV reflectance, absorption, or transmission. An example of this is the identification of materials used by the artist in a painting. The use of UV microscopy for enhanced resolution has been entirely superseded by transmission electron microscopy.

### Infrared Microscopy

The infrared spectra of many organic materials reveal much important information about their chemical nature. The complex waveform produced by passing an infrared beam through a sample or collecting it from the sample may be analyzed by a Fourier transform to reconstruct all the waves that make up the waveform (FTIR microscopy). This enables experimental data to be compared with spectra from known standards. Thus, small amounts of a vast range of organic materials, from plastics through natural polymers to drugs, can be analyzed. The wavelength range of infrared radiation means that the theoretical resolution of an FTIR microscope is of the order of a micrometer, somewhat poorer than that of a visible-light microscope. However, modern instruments are capable of analyzing 10–20  $\mu\text{m}$  wide areas. Most work has been in transmission, but measurements with epi-illumination are now possible, simplifying specimen preparation. Infrared microscopes have also been used to measure the temperature of samples, and scanning versions have been built to enable maps of temperature distribution.

*See also:* **Chiroptical Analysis. Microscopy:** Overview. **Microscopy Techniques:** Specimen Preparation for Light. **Optical Spectroscopy:** Refractometry and Reflectometry. **Particle Size Analysis.**



## Further Reading

Bradbury S and Bracegirdle B (1998) *Introduction to Light Microscopy*. Royal Microscopical Society Microscopy Handbook no. 42. Oxford: Bios Scientific Publishers.

Hemsley DA (ed.) (1989) *Applied Polymer Light Microscopy*. London: Elsevier.

Lacey AJ (ed.) (1989) *Light Microscopy in Biology – a Practical Approach*. Oxford: IRL Press.

Pluta M (1988–1993) *Advanced Light Microscopy*, 3 vols. Amsterdam: Elsevier.

Wallis TE (1965) *Analytical Microscopy*, 3rd edn. London: J. and A. Churchill.

## Sample Preparation for Light Microscopy

**C Hammond and P J Evennett**, University of Leeds, Leeds, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by P J Evennett and C Hammond, pp. 3127–3134, © 1995, Elsevier Ltd.

## Introduction

In this article, the range of specimen preparation techniques is surveyed and classified to provide the reader with an appreciation of the 'context' within which specimen preparation techniques are selected and carried out. However, it cannot provide a solution, or specific information, on the preparation technique for any 'particular' material or sample. Successful specimen preparation, in the end, depends on the judgment, experience, and experimental skill of the preparer, and each individual specimen may present its own particular problems; hence, there may be as many preparation techniques as there are materials or samples to be prepared.

First, the overall criteria for specimen preparation techniques are set down and discussed; these provide the basis on which all the subsequent techniques should be assessed. These fall into two main categories: those for transmitted light and those for reflected light.

Techniques for transmitted light are of major concern in biological light microscopy. This includes the techniques applying to living specimens, the preparation of permanent mounts of dead specimens, and fixation, dehydration and embedding, sectioning, and staining. Staining is a vast topic in its own right and the great variety of stains and staining techniques can only be touched on. The transmitted-light techniques for nonbiological materials (fiber particulates, soils, etc., and petrological thin sections) are also considered; to a certain extent these overlap biological preparation techniques (e.g., cutting polymer thin sections) and reflected-light surface preparation techniques (e.g., preparation of petrological thin sections). Finally, mounting techniques are

discussed in relation to their importance with respect to the microscope imaging system.

Reflected-light techniques for the preparation of 'bulk' specimens are simpler to consider and are confined largely to unmounted biological specimens and nonbiological materials (ceramics, ores, rocks, metals). For the nonbiological specimens there are three main stages. First, there is the process of mounting the specimen in a block to facilitate handling and to ensure that the prepared surface is flat without 'rounded' edges; then follows the initial preparation, grinding, lapping, and polishing, the object of which is to obtain a perfectly flat surface without distortion or obliteration (smearing) of the underlying structure. Finally, surface treatments that are almost invariably required in order to reveal the underlying structure are considered. Under this section come all the etching, electrolytic, and thin-film deposition treatments. This topic is analogous to that of staining of biological materials, and the range of etchants, electrolytes, and deposition techniques can only be touched on.

## Criteria for Specimen Preparation Techniques

The objective of specimen preparation for light microscopy is to allow the structures of materials (biological and nonbiological) to be revealed with sufficient contrast, and features of interest to be described, recorded, and characterized, at a fineness of scale beyond the visual acuity of the naked eye. The fineness of scale is ultimately limited by the limit of resolution (resolving power) of the microscope imaging system.

Hence, specimen preparation techniques and procedures should fulfill, as necessary and appropriate, the following five criteria:

1. They should ensure that the specimen to be examined is a representative sample of the (bulk) material, object, or organism that is to be investigated.

## Further Reading

Bradbury S and Bracegirdle B (1998) *Introduction to Light Microscopy*. Royal Microscopical Society Microscopy Handbook no. 42. Oxford: Bios Scientific Publishers.

Hemsley DA (ed.) (1989) *Applied Polymer Light Microscopy*. London: Elsevier.

Lacey AJ (ed.) (1989) *Light Microscopy in Biology – a Practical Approach*. Oxford: IRL Press.

Pluta M (1988–1993) *Advanced Light Microscopy*, 3 vols. Amsterdam: Elsevier.

Wallis TE (1965) *Analytical Microscopy*, 3rd edn. London: J. and A. Churchill.

## Sample Preparation for Light Microscopy

**C Hammond and P J Evennett**, University of Leeds, Leeds, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by P J Evennett and C Hammond, pp. 3127–3134, © 1995, Elsevier Ltd.

## Introduction

In this article, the range of specimen preparation techniques is surveyed and classified to provide the reader with an appreciation of the 'context' within which specimen preparation techniques are selected and carried out. However, it cannot provide a solution, or specific information, on the preparation technique for any 'particular' material or sample. Successful specimen preparation, in the end, depends on the judgment, experience, and experimental skill of the preparer, and each individual specimen may present its own particular problems; hence, there may be as many preparation techniques as there are materials or samples to be prepared.

First, the overall criteria for specimen preparation techniques are set down and discussed; these provide the basis on which all the subsequent techniques should be assessed. These fall into two main categories: those for transmitted light and those for reflected light.

Techniques for transmitted light are of major concern in biological light microscopy. This includes the techniques applying to living specimens, the preparation of permanent mounts of dead specimens, and fixation, dehydration and embedding, sectioning, and staining. Staining is a vast topic in its own right and the great variety of stains and staining techniques can only be touched on. The transmitted-light techniques for nonbiological materials (fiber particulates, soils, etc., and petrological thin sections) are also considered; to a certain extent these overlap biological preparation techniques (e.g., cutting polymer thin sections) and reflected-light surface preparation techniques (e.g., preparation of petrological thin sections). Finally, mounting techniques are

discussed in relation to their importance with respect to the microscope imaging system.

Reflected-light techniques for the preparation of 'bulk' specimens are simpler to consider and are confined largely to unmounted biological specimens and nonbiological materials (ceramics, ores, rocks, metals). For the nonbiological specimens there are three main stages. First, there is the process of mounting the specimen in a block to facilitate handling and to ensure that the prepared surface is flat without 'rounded' edges; then follows the initial preparation, grinding, lapping, and polishing, the object of which is to obtain a perfectly flat surface without distortion or obliteration (smearing) of the underlying structure. Finally, surface treatments that are almost invariably required in order to reveal the underlying structure are considered. Under this section come all the etching, electrolytic, and thin-film deposition treatments. This topic is analogous to that of staining of biological materials, and the range of etchants, electrolytes, and deposition techniques can only be touched on.

## Criteria for Specimen Preparation Techniques

The objective of specimen preparation for light microscopy is to allow the structures of materials (biological and nonbiological) to be revealed with sufficient contrast, and features of interest to be described, recorded, and characterized, at a fineness of scale beyond the visual acuity of the naked eye. The fineness of scale is ultimately limited by the limit of resolution (resolving power) of the microscope imaging system.

Hence, specimen preparation techniques and procedures should fulfill, as necessary and appropriate, the following five criteria:

1. They should ensure that the specimen to be examined is a representative sample of the (bulk) material, object, or organism that is to be investigated.

2. They should be such as to make the features of interest accessible to the microscope imaging system.
3. They should, where necessary, enhance the contrast between the structural elements of materials and between the structural elements and the undifferentiated background, such that the features of interest may be identified and recorded.
4. They should provide sufficient conditions for specimen stability, and sufficient means to protect the specimen from decay, degradation, corrosion, or contamination within the observational time scale.
5. They should match the specimen to the optical requirements of the microscope imaging system so as to provide the optimum conditions for image contrast and resolution and to avoid optical artifacts that may otherwise arise and that may obscure, or incorporate false detail into, the image.

The particular preparation techniques and procedures relating to these five criteria are summarized below and further details, and their importance with respect to different materials and scientific and technological disciplines, are set out in the subsequent sections.

Specimen preparation may be broadly classified into two main areas: specimen preparation techniques for transmitted-light microscopy (thin sections, smears, fibers, particulate strews) and specimen preparation techniques for reflected-light microscopy (surface preparation of opaque and nearly opaque materials). However, they are by no means exclusive and, although metals, ores, and opaque minerals may be studied by reflected light only, thin sections, smears, etc., may be studied both by transmitted and by reflected light, e.g., transmitted-light and epifluorescence microscopy. The advantages of such dual observational techniques relate in particular to contrast enhancement (criterion (3)).

It is important that the contrast observed should be that arising from features in the specimen and not due to incorrect adjustment of the microscope. The removal of a specimen (transmitted light) or the insertion of a perfectly flat reflecting mirror or speculum (reflected light) should result in a uniformly illuminated field under bright-field imaging conditions and a uniformly dark field under dark-ground imaging conditions.

### Sampling

The need to ensure that the specimen is a representative sample strictly relates rather to overall scientific methodology than to preparation procedures as such. However, the scientific or technological significance of the subsequent microscopical investigation

is ultimately determined by such considerations and indeed may be completely vitiated by unconsidered sampling or dissection procedures.

### Accessibility of Features

The conditions under which the features of interest in a sample are immediately accessible to the imaging system, i.e., such that no further specimen preparation is required (except for an appropriate mounting procedure), are generally limited to low-power techniques, which are normally appropriate to the initial stages of a scientific investigation. Under this category may come 'whole mounts' of biological organisms or parts of organisms, fracture and corrosion surfaces of metals, ceramics, and polymers, and strews of particulate materials (sands, dusts, etc.).

Otherwise the sample must be cut in one or more orientations and sectioned, ground, and polished as appropriate so that the features of interest can be revealed in the thin sections (transmitted light) or from their intersections with the prepared surfaces (reflected light).

### Contrast

Contrast in the image is provided in transmitted-light techniques by variations in the absorption, diffraction, and interference of transmitted light, or in reflected-light techniques by variations in the reflectance, diffraction, and interference of light reflected from the features of interest. Contrast in thin sections (transmitted and reflected light) is enhanced by staining techniques, i.e., the selective absorption or attachment of dyes or fluorochromes to specific structural elements in the material, and from surfaces in reflected light by etching techniques, i.e., the selective attack and staining of the features of interest and the boundaries between them.

In some materials such staining procedures are either unnecessary, difficult, or impossible to carry out, and contrast in transmitted light may be achieved by exploiting (small) refractive index differences between the features of interest, or in optically anisotropic materials by exploiting refractive index differences for light vibrating in different directions. The former is the basis of phase contrast and differential interference contrast techniques and the latter is the basis for polarized-light techniques (for birefringent materials). Such techniques are also applicable to reflected light: phase contrast and differential interference contrast techniques enhance surface topographical detail and polarized-light techniques reveal optical anisotropy between features of interest (for bireflectant materials).

### Stability

The requirement that the specimen should not deteriorate, and should remain representative of the original sample or (living) organism, must be assessed in relation to the appropriate time scale – no specimen is permanent! The problem is far more acute in the case of biological materials; in inorganic materials the surface degradation and corrosion that occurs (particularly in reactive metals) may be removed and the surface simply repaired.

Living organisms must be humanely killed in such a way that their tissues are as nearly as possible unchanged from those of the living state. Unless the material is to be examined immediately, the sample must then be fixed to maintain this state, and it is fixation that largely determines the distinction between ‘permanent’ and ‘temporary’ mounts.

### Optical Requirements

The optical performance of the imaging system of the light microscope is determined primarily by the objective lens, viz., its numerical aperture (NA) and the extent to which the lens is corrected for chromatic and spherical aberrations. The NA determines (1) the minimum resolved distance or resolving power of the system and (2) the extent or depth of field in the object that is in focus. Conventional light microscope objectives have NA values in the range 0.15–1.4, corresponding to a minimum resolved distance approximately in the range  $2\text{--}0.2\text{ }\mu\text{m}$  and a depth of field in the range  $20\text{--}0.4\text{ }\mu\text{m}$ . In addition to depth of field, the other constraint on specimen thickness is the confusion that may arise as a result of overlapping detail in the features of interest. For transmitted light, biological specimens are normally up to about  $10\text{ }\mu\text{m}$  in thickness; nonbiological materials (e.g., mineral thin sections) are generally examined rather thicker, and  $30\text{ }\mu\text{m}$  is a conventionally accepted standard. These thickness requirements are considerably relaxed in the case of ‘confocal’ light microscopy techniques, where the optical-sectioning characteristics enable out-of-focus blur to be suppressed and information about the third dimension to be incorporated in the final image. Specimens for reflected light should have surface topographic detail with height variations of the order of  $3\text{ }\mu\text{m}$  or less. The depth-of-field constraint is substantially less stringent in the case of stereomicroscopes, which normally operate at substantially lower effective NAs, in the range 0.01–0.1, and also at lower magnifications, providing a depth of field of several millimeters. The concomitant advantage of these instruments is that no specimen preparation at all may be required, the limitations to microscopic observation being set only

by the transmission of sufficient light in transmitted-light specimens and the depth of surface topography of reflected-light specimens. In the latter case, in particular, greater variations in surface topography than those given by the depth of field are readily accommodated by coarse-focusing adjustments.

## Preparation Techniques for Transmitted Light

### Techniques for Biological Specimens

Methods of collecting, culturing, handling, killing, and dissecting living organisms are outside the scope of this article and are extremely varied; they should follow normal practice for the organism under investigation. Biological specimens may be examined in many states, as combinations of the following: alive or dead, fresh or preserved/fixed, entire or in section, unstained or stained, hydrated or dehydrated, opaque or cleared.

Specimens sufficiently small to be examined entire may also be examined alive, contrast being provided by optical devices such as phase contrast or interference contrast microscopy, or occasionally by means of ‘vital dyes’ – nontoxic stains that have an affinity for certain components of cells. However, most organisms, their organs and tissues, and even their cells, are too large to provide high-resolution information with the conventional light microscope, and they must be killed and then (usually) dehydrated, embedded, sectioned, stained, and mounted before microscopical examination.

### Temporary mounts of entire, living specimens

Specimens originating as whole organisms or parts of living organisms require special treatment, either to maintain life during observation or to ensure that their death does not destroy the information of interest. Specimens examined by microscopy while still alive are generally small: whole organisms such as bacteria, Protista, and lower animals and plants, or discrete parts of organisms such as cells in culture, blood cells, pollen grains, etc. Depending upon the duration of observation and the nature of the specimen, attention must be paid to providing sufficient oxygen, maintaining concentrations of other substances (e.g., salts and ‘food’), removing carbon dioxide and other wastes, controlling temperature, providing adequate light, and excluding foreign organisms of predation or decay. Specimens of the types considered here will normally be mounted in the medium in which they occur naturally or are cultured, and techniques of preparation may be limited to mounting them on a slide, beneath a

coverslip, in a small drop of medium. If observation is to be prolonged, it may be necessary to retard evaporation by sealing the edges of the coverslip with petroleum jelly, nail varnish, rubber solution, or other suitable sealant. Conversely, the clarity of the image of organisms such as *Paramecium* increases considerably as the preparation becomes thinner and the organism more compressed on evaporation of fluid from beneath the coverslip. Some larger organisms can more easily be accommodated using a 'cavity slide', a slide with a shallow part-spherical depression. Cavity slides, or normal slides with a thin ring attached with an adhesive, can also be used for 'hanging drop' preparations. A small drop of medium containing the organism is placed on a coverslip; a ring of petroleum jelly is applied around the cavity or on the ring of the slide, the slide inverted over the fluid drop on the coverslip, and the whole then raised and turned over so that the drop hangs from the center of the coverslip. Such a preparation allows more prolonged observation before becoming anoxic than does a simple sealed preparation.

#### **Permanent mounts of dead specimens: fixation**

Organisms from which permanent mounts are to be made will necessarily be dead once mounted, so care must be taken that death occurs humanely and with as little alteration to the structure as possible. As soon as possible after death, and in some cases simultaneously with it, living tissues should be 'fixed', a process that 'kills' the cells, inactivates their enzymes (thus halting metabolism and preventing self-digestion), and precipitates and cross-links the molecules that are in simple or colloidal solution or suspension within the cells. All these actions stabilize the structure against damage during subsequent processing. Fixation must also act as a preservative against microbial decay, and may also improve a specimen's affinity for certain stains. Fixation is generally carried out by chemical treatment, though 'physical' fixation by heating or by rapid freezing may be used in certain circumstances.

Chemical fixative agents are many and varied, acting optimally on different constituents of living cells; combinations of several agents are thus frequently used together, a procedure that can also compensate for individual fixatives' shortcomings, such as a tendency to shrink or swell. Common fixative agents include formaldehyde (methanal) (typically a 4% w/v solution of the gas), ethanol (used undiluted), acetic (ethanoic) acid (5% aqueous), mercury(II) chloride (saturated aqueous), osmium tetroxide (1% aqueous), and glutaraldehyde (4% aqueous). Other substances such as picric acid, potassium dichromate, chromium trioxide, and many others have been used

for specialized purposes. Fixatives are generally used in aqueous solution, in saline, or in a buffer when maintenance of a specified pH is considered important. Traditional fixative mixtures still in current use include formaldehyde-saline (4% formaldehyde in 0.7% sodium chloride), and Bouin's solution (picric acid, formaldehyde, acetic acid). For high-quality preparations there is increasing use of fixative systems developed especially for electron microscopy (glutaraldehyde followed by osmium tetroxide). No single fixative mixture can serve all purposes; readers should consult specialist texts for information on choice and preparation of fixatives for specific specimens. Following fixation, it is generally necessary to rinse specimens free of fixative using water, buffer, or dilute ethanol.

**Dehydration and embedding** Whereas it is possible to make satisfactory sections of some plant material simply by skilled use of a razor, other specimens, particularly animal tissue, require more complex preparation. Animal cells lack the support provided by the cell wall of the plant, and thus must be provided artificially with adequate support before sectioning, conventionally by embedding. An embedding medium must be capable of infiltrating the tissue in the form of a liquid, and then solidifying, on cooling or by polymerization, to provide support from within the fine details of the specimen. The most commonly used embedding medium is paraffin wax, of a grade that melts at ~56°C. However, paraffin wax is not miscible with the aqueous fluid into which the fixed specimen has generally been transferred; the technique thus requires the specimen to be dehydrated, generally by passage through increasing concentrations of ethanol until absolute (100%) ethanol is reached. The time required for this will be minutes, hours, or even days, depending on the size of the tissue sample. Although ethanol is generally considered to be a suitable dehydrating agent, it nevertheless is not miscible with molten paraffin wax; replacement of the ethanol with a further, transitional, solvent that is miscible also with wax is thus required – common examples are xylene, toluene, or less-toxic, proprietary substitutes. After such a solvent has replaced the ethanol fully, the tissue is placed in molten wax for a period of ~1 h, during which time the wax infiltrates into the tissue. Finally, the tissue is transferred into fresh molten wax in a suitable container, and the wax is allowed to solidify. Alternatives to paraffin wax for specialized applications exist in the form of proprietary formulations, ester and polyester waxes, and also gelatin and celloidin (nitrocellulose), both of which require specialized sectioning techniques.



**Sectioning** As noted above, it may be possible to cut satisfactory sections freehand; most sections, however, are cut with the aid of a 'microtome' – a heavily built precision instrument designed to move a specimen block across the cutting edge of a sharp knife for sectioning, advancing the specimen toward the knife by a preset distance for each section. With good technique, sections may routinely be cut at 5  $\mu\text{m}$  thickness, and even thinner in some cases. Important factors in the sectioning process include the hardness of the wax block (dependent also on temperature), the cutting speed, the angle at which the knife is set with respect to the path of the block, and, most particularly, the characteristics of the knife (including angle of cutting edge, and its sharpness). Microtome knives are usually made of steel, sometimes in the form of small replaceable blades, and occasionally of tungsten carbide. Glass or diamond knives similar to those devised for electron microscopy may also be used for some specialized applications.

During cutting, successive sections attach together to form a 'ribbon' of sections, which is convenient to handle and mount. Sections are attached to clean microscope slides typically by floating them on a dilute solution of albumen, warmed on a hotplate to soften (but not melt) the wax, thereby to remove wrinkles in the sections, and allowed to dry in a warm place.

**Staining** Sections cut from wax-impregnated blocks must first be 'de-waxed' by immersion in a suitable solvent, traditionally xylene, now often replaced with a less-toxic proprietary substitute. This is generally followed by a succession of increasingly aqueous solutions of ethanol until the specimen is fully hydrated, since most stains are applied in aqueous solution. Once the support of the wax has been removed, the specimen should not be allowed to dry because of damage caused by surface tension.

Staining exploits chemical affinities, in general not well understood, between stains and substances present within the specimen. Some early histological stains were natural substances first used for dyeing textiles, selected for their affinities for cellulose (for dyeing cotton fabrics) or for proteins (the keratin of wool). Since the development of the first aniline dye in 1856, a vast array of stains has been available for experimentation by histologists. Nonetheless, one of the earliest stains, hematoxylin, a natural product extracted from 'logwood', remains one of the most important stains today, generally used to stain cell nuclei blue or black according to the method of application. The successive or combined use of several stains, including hematoxylin, each with its specific

affinity, provides the familiar 'trichrome' or more complex colorations of diverse tissue components.

The chemistry of some 'staining' techniques is better understood. For example, Schiff reagent gives a magenta color on reacting with aldehyde groups made available, for example, from polysaccharides by oxidation with periodic acid, or from DNA by hydrolysis with hydrochloric acid. Enzyme histochemistry relies on the capture of reactive products of enzyme action by appropriate components incorporated in an incubation medium, marking the location of enzyme activity by an inorganic precipitate or a dye. The most specific form of 'staining' in the widest sense of the term is seen in immunocytochemistry, where markers are attached specifically to target molecules within the specimen by means of the immune reaction. The technique requires antibodies specific to the substance under investigation, together with a means of localizing the antibody using a fluorescing substance or other marker.

Following staining, specimens must be dehydrated, cleared, and mounted, as outlined later.

#### **Preparation Technique for Transmitted-Light Microscopy of Nonbiological Materials: Fibers, Particulates, Petrological Thin Sections**

Nonbiological particulate materials (e.g., mineral grains) need no premounting preparation (see next section); biological particulate materials and fibers (e.g., pollen grains, wool fibers) normally require rinsing in alcohols, xylene, etc., to remove waxy surface films and pigmentation that may obscure the surface detail. In situations where it is necessary to maintain the distribution of the grains (e.g., soils, porous ceramics) the bulk specimen must be impregnated with a resin as described for reflected-light microscopy (see below) and a thin section is then prepared. Thin sections of soft materials (e.g., polymers) may be prepared by sectioning with a microtome in a manner analogous to that described above for biological specimens. Contrast between the features of interest (e.g., amorphous and crystalline regions) may be enhanced by staining and/or etching the section, but such techniques are generally rather difficult because of the chemically inert character of the material, and they are less structure-specific and much less developed than for biological material. Thin sections of hard materials (e.g., minerals, rock, ceramics, resin-impregnated ceramics) are prepared by a sequence of steps as follows: (1) cutting a thin plate  $\approx 1\text{ mm}$  thick, or 'chipping' a fragment from the bulk material; (2) mounting the plate on a glass slide using 'Lakeside' or an equivalent proprietary cement; (3) grinding, polishing, or lapping one side of the plate as described below for reflected-light

techniques; (4) mounting the plate polished-face-down on another slide and detaching the first; (5) grinding and polishing the (upper) as-cut surface to the required thickness; and (6) cementing a coverslip to the upper surface. The choice of abrasives and polishing materials is determined partly by the hardness of the material, but diamond abrasives are increasingly used because of their rapid-cutting qualities. It is normal practice in petrology to aim for a section thickness of 30  $\mu\text{m}$  (i.e., considerably greater than the norm for biological thin sections) because using polarized light (crossed polars) the minerals in such sections give polarization or interference colors that enable them to be readily recognized. Staining techniques for minerals and ceramics are not very widespread and are largely confined to those aggregates (e.g., sedimentary rocks) in which interference contrast is weak.

### Mounting

The final step in specimen preparation is to mount the prepared thin section, replica, fibers, or particulate strew, usually on a standard 25 mm  $\times$  76 mm (1 in.  $\times$  3 in.) glass slide, with a mounting material (mountant) (Canada balsam or its proprietary substitutes, glycerin, oils, aqueous liquids), and then to cover the preparation with a coverslip, a thin glass disk, or plate.

The thicknesses of the glass slide and coverslip and the refractive index of the mounting material and its suitability for the specimen are all factors that need to be considered. 'Dry' objectives are designed such that spherical aberrations are minimized when the specimen surface is covered by a glass slip of a given thickness and refractive index. This thickness is conventionally 0.17 mm and this assumes that the film of mounting medium between the specimen surface and coverslip is of negligible thickness. In order to fulfill this condition as closely as possible, it may be good practice to mount the specimen on the coverslip rather than the slide. Deviations from this condition are serious only for dry objectives of high NA (0.65–1.00), and such objectives may be provided with a 'correction collar'. This is adjusted so as to minimize spherical aberration for a range of coverslip-plus-mountant thicknesses, and may be graduated in arbitrary units or in millimeters; in the latter case it should be noted that the correct setting will be less than the coverslip thickness because of the necessity to take into account the mountant thickness. Spherical aberration corrections may also be made by altering the tube length of the microscope imaging system: shorter tube lengths are required when the coverslip-plus-mountant thickness is

greater than the conventional 0.17 mm and vice versa. However, this facility is not normally available in modern binocular instruments.

Glass slides are normally manufactured in a thickness range of  $\sim$ 0.75–1.5 mm. The thickness is an important factor for consideration principally in the case of dark-ground imaging using oil-immersion condensers of high NA that are designed to focus the hollow cone of light on the specimen at a given distance from their (top) working surface. Again, such condensers may be provided with a 'correction collar' so that this distance can be adjusted to suit the particular thickness of slide being used.

Immersion objectives are designed such that spherical aberrations are minimized when the appropriate medium (oil, water, or glycerin) lies between the front element of the objective lens and the coverslip. Immersion objectives may therefore be used equally for covered and uncovered (reflected-light) specimens.

For thin sections, and biological thin sections in particular, the mounting medium is chosen so as to be most appropriate to the specimen – i.e., whether it is a 'permanent' mount (in Canada balsam or its proprietary equivalents) or a 'temporary' mount (in glycerin, glycerin jelly, water-based materials). The distinction between 'permanent' and 'temporary' mounts is not rigid, and 'temporary' mounts may be sealed or ringed with varnishes and sealants of various kinds to make them 'semipermanent'. However, the important consideration is that the mountant should be inert, or act as a preservative against specimen degradation, within the observational time scale.

Transferring a hydrated biological specimen into an aqueous mountant may require little more than the appropriate manipulations, combined with warming the mountant in the case of glycerin jelly. However, nonaqueous mountants require in addition that specimens be dehydrated and transferred to a solvent miscible with the mountant. This is usually achieved using a series of aqueous solutions of ethanol of increasing concentrations leading to 'absolute (nominally 100%) ethanol', followed by xylene or a proprietary substitute for mounting in Canada balsam.

The refractive index of the mountant is a particularly important consideration for all preparations, because of its effects on the quality of the microscopical image, and on contrast in particular. Hydrated biological material frequently has a milky appearance, owing to scattering of light at the many interfaces between the aqueous components and the structures of higher refractive index (proteins, etc.).

Where the image is required to demonstrate optimally the distribution of stained material, the specimen should have approximately uniform refractive index throughout. This is readily achieved when the water is replaced by a suitable organic fluid (e.g., methyl salicylate, xylene, etc.) known as a 'clearing agent', and/or by infiltration and inclusion of the specimen in a resinous medium such as Canada balsam, whose refractive index is close to those of the tissue components.

In examining unstained and otherwise transparent biological objects, where contrast in the image must be achieved by phenomena other than the differential absorption of light, contrast will be increased according to the degree of mismatch between the refractive indices of specimen components and mountant, as discussed below for nonbiological objects.

In the case of fibers or particulate strews, the contrast of the fibers or particles is strongly determined by the refractive index differences between the mountant: nonabsorbing specimens of identical refractive index to the mountant having, of course, zero contrast. This provides an important means of identification of fibers and mineral grains: by immersing specimens in liquids of different refractive indices and by using the 'Becke line test' (which indicates whether the fibers or particles have a greater or lesser refractive index than the liquid), the refractive index (or refractive indices of birefringent materials) can be determined very accurately.

## Reflected-Light Preparation Techniques

### Initial Preparation and Mounting

In order to facilitate handling on the one hand, and to provide a sufficiently large representative specimen surface area on the other, the specimen dimensions should be in the range 5–15 mm in the prepared surface and 1–10 mm thick. Cutting such a specimen from a large piece of rock or a metal component, for example, must be carried out with considerable care; this apparently crude first step in specimen preparation must not be carried out by crude cutting methods that might overheat the specimen and give rise to cracking and changes in the structure. Considerable caution must be exercised, and samples in which it is suspected that inappropriate cutting methods have been used should be rejected. In many cases the orientation of the specimen surface (e.g., in relation to the host rock, or the metal component) is important and should be recorded.

Abrasive wet cutting, using metal or bakelite abrasive 'cut-off' wheels impregnated with alumina, boron nitride, silicon carbide, or diamond, is most appropriate and gives rise to minimum deformation of the cut surface. Wheels may be distinguished as 'hard' or 'soft' depending on their ability to retain or discard the abrasive particles. In general, the harder (more difficult to cut) materials require 'softer' wheels because these provide a continuous supply of new abrasive particles to the surface (but at the same time wear away more quickly).

There are two main mounting or embedding methods: 'hot compressive' mounting in bakelite or a similar thermosetting resin, which is carried out at elevated temperatures and pressures using proprietary equipment; and 'cold' mounting using mixes of resins and hardeners that set by polymerization and are poured around the specimen in a mould. Proprietary embedding materials (epoxies, acrylics, polyesters) are distinguished by their hardening rates, amounts of contraction (shrinkage), and final cured hardness. Epoxy embedding materials are also used, because of their low vapor pressure and viscosity, for the vacuum infiltration of porous materials, a procedure that is required (1) to stabilize the sample and (2) to prevent the ingress of the subsequent grinding and polishing compounds. In difficult cases, the impregnated epoxy may be mixed with a fluorescent dye to allow easy identification of the filled pores.

Clearly, 'hot' mounting methods cannot be used for temperature-sensitive materials, but otherwise they do in general provide a harder mount; this is particularly important in order to maintain a flat polished surface and good edge retention or definition. However, the continuing development, particularly of the polyester/acrylic resins, makes this distinction less evident.

### Grinding, Lapping, and Polishing

These procedures must be carried out with due regard to the precautions noted in the preceding section. Grinding is carried out by hand using paper-backed abrasive strips, or by using abrasive disks mounted on rotating wheels with successively finer grades of grit, normally silicon carbide for the coarser grades and  $\alpha$ - and  $\gamma$ -alumina and diamond for the finer grades. The silicon carbide grit particle sizes are distinguished by their sieve sizes or grades, and normally four such grades are sufficient, e.g., P240 ( $\sim 58 \mu\text{m}$  grit), P400 ( $\sim 35 \mu\text{m}$  grit), P600 ( $\sim 25 \mu\text{m}$  grit), and P1200 ( $\sim 15 \mu\text{m}$  grit). The abrasive papers must be thoroughly lubricated (usually with water) and the specimen (and fingers) thoroughly cleansed in order to avoid carryover of coarser grit to the finer

abrasive papers. Each stage may be regarded as complete when the underlying surface deformation has been completely obliterated. In practice, this means continuing the grinding for about twice the time required to remove the previous set of scratches.

The final lapping or polishing stages may be carried out with  $\alpha$ -alumina (particle sizes in the range 10–1  $\mu\text{m}$ ),  $\gamma$ -alumina (particle sizes  $\sim 0.3 \mu\text{m}$ ), or diamond (particle sizes in the range 6–1  $\mu\text{m}$ ). Normally only two stages are required. A choice has to be made with respect not only to the polishing compound but also to the cloth in which it is embedded, and to the lubricant. In general, soft cloths (those with a nap) and alumina compounds are appropriate to soft, friable, and easily damaged materials, but they do give rise to surface relief between hard and soft phases; this may be either an advantage or a disadvantage depending upon the end requirements. Hard cloths and diamond compounds are in general appropriate to hard materials. However, the distinction is by no means rigid and each specimen may have its own distinctive requirements.

In general, soft materials are more difficult to polish than hard materials. Insufficient lubricant, or cloths in which the abrasive is embedded too deeply to cut the material, may give rise to smearing in metals – the plastic deformation of the surface layers, which effectively obscures the underlying structure. In such cases the final step should be electropolishing, which involves no plastic deformation of the specimen surface.

Lapping is distinguished from grinding and polishing in that the abrasive particles are not stuck to a backing paper, or embedded within a cloth, but are free to move and roll freely across a relatively hard metal surface (e.g., cast iron) or become partially embedded in a soft metal (e.g., lead). Material removal rates during lapping in the former case are in general very low and the technique is generally confined to hard materials (ceramics, hard metals, etc.) in which surface finish is of prime importance. The use of soft laps is of particular importance in the grinding and polishing of minerals where the use of cloths results in unacceptable surface relief. Lapping is in general unsuitable for soft materials because of the problem of grit becoming embedded in the surface.

### Surface Treatments

For many materials (e.g., rocks, ceramics) no further treatment is required and the constituent phases may be sufficiently well revealed by their differences in color, bireflectance, and surface relief. For metals and alloys, etching treatments are required, but in any

event all materials should be examined in the as-polished condition to check for artifacts (e.g., localized smears or ‘comet-tails’) and to detect coarse inclusions and porosity that may be affected by etching processes.

Etching may be defined as the selective attack of grain and interphase boundaries and/or the selective attack of the phases themselves. In complex microstructures no single etchant may be effective and different etchants may be required in order to reveal different elements of the microstructure. Etchants usually consist of aqueous or alcoholic solutions and mixtures of salts, acids, and alkalis; the more corrosion-resistant metals and alloys (e.g., stainless steels) naturally require more aggressive solutions. The etching process consists simply of dipping or swabbing the specimen in the etchant, or of electrolysis, with the specimen acting as the anode of the electrolytic cell. The range of etchants and etching conditions practically equals the range of materials and microstructures to be examined. Like staining, it is very much a matter of experience and experimental skill – incorrect choice of etchants, etching time, etc., simply leads to an overall attack on the specimen surface, revealing no detail at all. Ion bombardment is another method for etching.

The etching process may be partially combined with the final stage of polishing; the etchant is added to the polishing compound and contributes to the overall material removal rate as well as reducing any tendency to smearing. However, a postpolishing etching stage is normally required.

Finally, contrast between phases may be revealed or enhanced by developing a thin (interference) film on the as-polished or polished-and-etched specimen surface. The technique is based on multiple-beam interference between the light reflected from the air-film and film-substrate surfaces, and the colors that occur are determined by such factors as film thickness, refractive index, and absorption coefficient. The film may be developed by anodizing (in which case it is restricted to oxides of the metallic specimen) or it may be deposited by vacuum evaporation or reactive-ion sputtering (in which case there is no restriction either to the deposited film or to the specimen). Anodizing is carried out using a range of oxidizing reagents and is a process analogous to electroetching except that the cell conditions are adjusted to develop an oxide film on the specimen (anode) surface. In practice, the two processes may be carried out sequentially.

Contrast enhancement by deposition of a metal oxide, sulfide, selenide, etc., film is a relatively new technique (interference film microscopy) and also lends itself to quantitative interpretation.

See also: **Fluorescence:** Overview. **Microscopy:** Overview. **Microscopy Techniques:** Specimen Preparation for Electron Microscopy. **Optical Spectroscopy:** Refractometry and Reflectometry. **Sampling:** Theory.

## Further Reading

- Bancroft JD and Stevens A (eds.) (1990) *Theory and Practice of Histological Techniques*, 3rd edn. Edinburgh: Churchill Livingstone.
- Berlyn GP and Miksche JP (1976) *Botanical Microtechnique and Cytochemistry*. Ames, IA: Iowa State University Press.
- Bjerregaard L, Geels K, Ottesen B, and Ruckert M (1992) *Struers Metallog Guide*. Rødovre, Denmark: Struers A/S.
- Bousfield B (1992) *Surface Preparation and Microscopy of Materials*. Chichester: Wiley.
- Bradbury S (1973) *Peacock's Elementary Microtechnique*, 4th revised edn. London: Edward Arnold.
- Gray P (1954) *The Microtome's Formulary and Guide*. New York: Blakiston.
- Humphries DW (1992) *The Preparation of Thin Sections of Rocks, Minerals, and Ceramics*. Oxford: Oxford University Press and Royal Microscopical Society.
- Leco Corporation (1991) *Metallography Principles and Procedures*. St. Joseph, MI: Cornelius A. Johnston.
- Samuels LE (1967) *Metallographic Polishing by Mechanical Methods*. Melbourne: Sir Isaac Pitman & Sons Ltd.
- Sanderson JP (1994) *Biological Microtechnique*. Oxford: Bios Scientific Publications in association with the Royal Microscopical Society.

## Electron Microscopy

**J Pérez-Arantegui**, Universidad de Zaragoza, Zaragoza, Spain

**T Mulvey**, Aston University, Birmingham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Because of the current tendency to develop integrated multipurpose electron microscope systems based on transmission electron microscope (TEM)/scanning transmission electron microscope (STEM) or scanning electron microscope (SEM) instruments equipped with multiple detector systems, and using different diffraction methods and observation modes, conventional TEM, STEM, and SEMs have evolved into a true analytical electron microscopy (AEM) tool. The aim of AEM is to determine the crystallinity, morphology, phase, and elemental compositions of a material using the signals generated in the course of the focused electron-beam interaction with the specimen. In this article, history and development stages of EM are considered; the basic theory of the EM is introduced followed by a discussion on TEM imaging and aberration correction. Finally, some brief considerations in standard TEM and SEM instruments and in the limitations of EM are presented.

## Microscopy: The History of a Challenge

The immediate purpose of microscopy is to enhance our unaided eyesight in examining objects whose details are not otherwise clearly resolvable. It is an old

problem that people have tried to resolve since ancient times. In 1897, Ernst Abbé presented his famous wave theory of the microscope, which predicted that the ultimate limit of the light microscope was set by the finite wavelength of light rather than by the accuracy of current optical manufacturing techniques. As the microscopes, made to Abbé's design, at Carl Zeiss Jena had aberration-free objectives, it meant that further improvement in resolving power was possible. The resolution, i.e., the distance  $d$  between two small adjacent particles that are just distinguishable from each other, is given from optical theory by

$$d = \frac{0.61\lambda}{n \sin \alpha} \quad [1]$$

where  $\lambda$  is the wavelength of the light,  $\alpha$  is the semi-angle of the cone of rays entering the objective lens, and  $n$  is the refractive index of the region between the specimen and the near surface of the objective. In an optical microscope, the semiangle can, in principle, be close to  $\pi/2$  radians and hence  $\sin \alpha$  is close to unity. Since the wavelength of visible light is  $\sim 550$  nm, the theoretical resolution of the microscope is  $\sim 340$  nm in air ( $n = 1$ ). This can be pushed down to 227 nm by oil immersion ( $n = 1.5$ ). In nonvisible ultraviolet illumination, which precludes the use of glass lenses, it might be possible, at a wavelength of 220 nm, to achieve a resolution of  $\sim 100$  nm with quartz lenses. This is still inadequate for the needs of researchers in the fields of materials science and life sciences. The atomic spacing of gold, for example, is 0.2 nm. There was therefore no hope, using the light microscope, of seeing fine detail in



See also: **Fluorescence:** Overview. **Microscopy:** Overview. **Microscopy Techniques:** Specimen Preparation for Electron Microscopy. **Optical Spectroscopy:** Refractometry and Reflectometry. **Sampling:** Theory.

## Further Reading

- Bancroft JD and Stevens A (eds.) (1990) *Theory and Practice of Histological Techniques*, 3rd edn. Edinburgh: Churchill Livingstone.
- Berlyn GP and Miksche JP (1976) *Botanical Microtechnique and Cytochemistry*. Ames, IA: Iowa State University Press.
- Bjerregaard L, Geels K, Ottesen B, and Ruckert M (1992) *Struers Metallog Guide*. Rødovre, Denmark: Struers A/S.
- Bousfield B (1992) *Surface Preparation and Microscopy of Materials*. Chichester: Wiley.
- Bradbury S (1973) *Peacock's Elementary Microtechnique*, 4th revised edn. London: Edward Arnold.
- Gray P (1954) *The Microtometist's Formulary and Guide*. New York: Blakiston.
- Humphries DW (1992) *The Preparation of Thin Sections of Rocks, Minerals, and Ceramics*. Oxford: Oxford University Press and Royal Microscopical Society.
- Leco Corporation (1991) *Metallography Principles and Procedures*. St. Joseph, MI: Cornelius A. Johnston.
- Samuels LE (1967) *Metallographic Polishing by Mechanical Methods*. Melbourne: Sir Isaac Pitman & Sons Ltd.
- Sanderson JP (1994) *Biological Microtechnique*. Oxford: Bios Scientific Publications in association with the Royal Microscopical Society.

## Electron Microscopy

**J Pérez-Arantegui**, Universidad de Zaragoza, Zaragoza, Spain

**T Mulvey**, Aston University, Birmingham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Because of the current tendency to develop integrated multipurpose electron microscope systems based on transmission electron microscope (TEM)/scanning transmission electron microscope (STEM) or scanning electron microscope (SEM) instruments equipped with multiple detector systems, and using different diffraction methods and observation modes, conventional TEM, STEM, and SEMs have evolved into a true analytical electron microscopy (AEM) tool. The aim of AEM is to determine the crystallinity, morphology, phase, and elemental compositions of a material using the signals generated in the course of the focused electron-beam interaction with the specimen. In this article, history and development stages of EM are considered; the basic theory of the EM is introduced followed by a discussion on TEM imaging and aberration correction. Finally, some brief considerations in standard TEM and SEM instruments and in the limitations of EM are presented.

## Microscopy: The History of a Challenge

The immediate purpose of microscopy is to enhance our unaided eyesight in examining objects whose details are not otherwise clearly resolvable. It is an old

problem that people have tried to resolve since ancient times. In 1897, Ernst Abbé presented his famous wave theory of the microscope, which predicted that the ultimate limit of the light microscope was set by the finite wavelength of light rather than by the accuracy of current optical manufacturing techniques. As the microscopes, made to Abbé's design, at Carl Zeiss Jena had aberration-free objectives, it meant that further improvement in resolving power was possible. The resolution, i.e., the distance  $d$  between two small adjacent particles that are just distinguishable from each other, is given from optical theory by

$$d = \frac{0.61\lambda}{n \sin \alpha} \quad [1]$$

where  $\lambda$  is the wavelength of the light,  $\alpha$  is the semi-angle of the cone of rays entering the objective lens, and  $n$  is the refractive index of the region between the specimen and the near surface of the objective. In an optical microscope, the semiangle can, in principle, be close to  $\pi/2$  radians and hence  $\sin \alpha$  is close to unity. Since the wavelength of visible light is  $\sim 550$  nm, the theoretical resolution of the microscope is  $\sim 340$  nm in air ( $n = 1$ ). This can be pushed down to 227 nm by oil immersion ( $n = 1.5$ ). In nonvisible ultraviolet illumination, which precludes the use of glass lenses, it might be possible, at a wavelength of 220 nm, to achieve a resolution of  $\sim 100$  nm with quartz lenses. This is still inadequate for the needs of researchers in the fields of materials science and life sciences. The atomic spacing of gold, for example, is 0.2 nm. There was therefore no hope, using the light microscope, of seeing fine detail in

bacteria, the structure of viruses, or the molecular and atomic structure of the cell or of metals.

Abbé himself, along with many others, did not believe that it would be possible to harness alternative radiation, such as X-rays or electrons, to the task of forming high-resolution images. Nevertheless, his wave theory of the light microscope applies remarkably well to the modern analytical electron microscope, in which the wavelength itself is not a limitation. The (dualistic) wavelength of an electron depends on its momentum, decreasing rapidly as the momentum increases. In many charged-particle devices, the electron velocity approaches that of light, so that relativistic theory is needed to estimate the wavelength. It may be readily calculated from the expression

$$\lambda = \frac{1.23}{V_r^{1/2}} \quad [2]$$

where  $\lambda$  is the wavelength (in nanometers) and  $V_r$  is the relativistically corrected acceleration voltage (in volts). Table 1 shows the electron wavelength, in nanometers, for the range of accelerating voltages  $V$  for electron microscopes now in commercial production. In the range of accelerating voltages from thousand to a million volts, the wavelength itself is therefore not a limiting factor in resolving any known structure.

This background information became available to scientists in 1924. Electron microscopes evolved in a similar empirical way to that of the light microscope, but over a brief 30-year period, compared with the many centuries needed to develop the light microscope.

The development of essential components of electron microscopes – such as electron sources, high-acceleration voltage generators, fluorescent screens, photographic plates, and beam scanning devices – had begun for use in cathode-ray oscillographs around the end of the nineteenth century. It was also known that a magnetic solenoid or ‘converging coil’ could focus a diverging beam of cathode rays (electrons) into a small spot. In 1924, Hans Busch in Jena showed theoretically that there is a close analogy between this action and that of a glass lens on light. However, it was Ernst Ruska and his Ph.D.

supervisor, Max Knoll, at the Technische Hochschule in Berlin, who, inspired by Busch’s theory, actually designed and built the first two-stage transmission electron microscope. In April 1931, they took the first TEM micrographs, at a magnification of some  $13\times$ , quite low by light microscope standards, but the design principle was established. In 1933, Ernst Ruska, working alone, designed and built a transmission electron microscope with a resolution better than that of the light microscope.

In 1935, Knoll devised and constructed the scanning electron microscope at Telefunken Berlin. In this microscope, a focused beam (‘electron probe’) is scanned in TV raster fashion across a solid specimen. In an SEM, the TEM column is in effect inverted to produce a demagnified image of ‘electron probe’ from the source. Electrons scattered by the specimen are detected and modulate the intensity of the beam of a cathode-ray oscilloscope scanned in synchronism with the electron probe. In Knoll’s SEM, the magnification, which is simply the ratio of the length of the scan on the screen to that on the specimen, was hardly more than unity, but the principle of imaging the surface of solid specimens was established. Its practical realization was beyond the technology of the time, but 50 or so years later it became a strong competitor of the TEM.

In 1938, again in Berlin, Manfred von Ardenne introduced a hybrid microscope. His newly devised scanning transmission electron microscope accepted a thin TEM specimen, which was scanned by a fine electron probe as in the SEM. The electrons leaving the specimen, having interacted with the atomic structure of the specimen, were collected and allowed to modulate a recorder placed immediately underneath the specimen and scanned in synchronism with the primary scanning beam, as in the SEM. As with the SEM, the principle in the STEM was demonstrated but the electronic technology at the time, and the limitations of the thermionic gun, were inadequate for commercial production. Ardenne’s motivation for the STEM was that since there were no lenses between specimen and detector, the chromatic aberration produced by the TEM objective lens would be absent. This is true, but inelastic scattering in a thick specimen can itself lead to delocalizing effects and hence loss of resolution.

The main disadvantage of scanning methods is that information is acquired serially, pixel by pixel, whereas in TEM all pixels are imaged simultaneously, leading to a much shorter exposure time. High-brightness field-emission guns (FEGs) now eliminate this difficulty. Although FEGs were being developed at the time by W.O. Mueller elsewhere in Berlin, they operated in all-glass high-vacuum systems that were

**Table 1** Electron wavelength ( $\lambda$ ) at various accelerating voltages in commercial electron microscopes

Voltage (kV)	$\lambda$ (nm)
1	0.04
100	0.004
1000	0.001

very inconvenient. Continuously pumped high-vacuum systems did not become generally available for STEMs and SEMs until the 1960s, when they were introduced into the STEM by A.V. Crewe in Chicago.

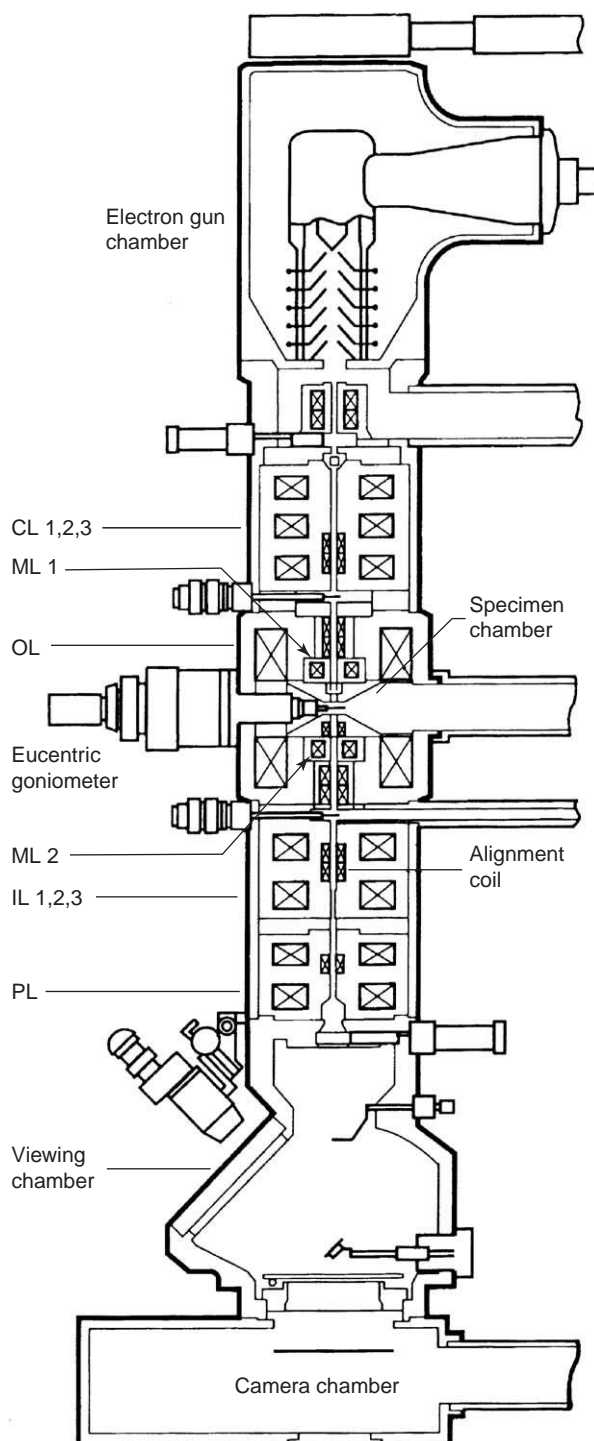
TEMs itself took off suddenly in 1939, with the installation of the Siemens Uebermikroskop (Super-microscope), the first serially produced transmission electron microscope, with a remarkable resolution of some 7 nm at IG-Farben in Germany. Thus, in a period of 8–9 years, the principal electron microscopes had been conceived and built in experimental form, and a commercially produced transmission electron microscope had begun to be used routinely in manufacturing industry. Today's electron microscopes offer the possibility of producing images with atomic resolution over a wide range of specimens. This is in itself a tremendous achievement. However, it should be noted that this is necessary but not sufficient condition for deducing the atomic structure of an unknown specimen. Although it is possible to calculate the image of a given structure, it is in general not possible to reverse this procedure. In practice, therefore, electron microscopists must learn how to interpret their images, using the many simulation techniques and other sources of information now available to them.

## Theory of Electron Microscopy

Fortunately, many ideas needed in electron microscopy can be taken over from optical microscopy theory. This applies both to wave optics and geometrical optics. Wave optics is particularly helpful in clarifying operation at high resolution, where phase-contrast effects predominate. Geometrical optics is important for understanding the action of the lenses in the column.

### The Electron Optical Column

A standard TEM is shown in Figure 1. Its general construction resembles that of the light microscope. The electron source at the top of the column emits electrons, which are accelerated, in this case, to 200 kV. The effective source size is small, some 25  $\mu\text{m}$  in diameter for a thermionic source, down to 2.5 nm for a field emitter. This source is further demagnified by the multiple lens system forming the illuminating beam. An illuminated area ranging from 3 mm, the size of a specimen grid, down to 0.5 nm, can be produced at the specimen plane located at the center of the air gap of the objective lens (OL) as shown in Figure 1. The objective lens lies at the heart of the instrument, as it alone determines the resolution that can be achieved. The objective is provided



**Figure 1** Cross-section of the Topcon EM-002B 200 kV column. CL, condenser lens; ML, minilens; OL, objective lens, with specimen in the center of the air gap; IL, intermediate lens; PL, projector lens. (The Topcon Corporation, Japan.)

with a set of apertures that enable the operator to alter image contrast and attainable resolution. The subsequent projector lenses project the image produced by the objective, or its associated diffraction pattern in the back focal plane, onto the fluorescent

screen, photographic plate, charge-coupled device (CCD) camera, or image intensifier recording system. The operator's task is thus to select the field of view, focus the image, correct any astigmatism, and record the image.

### Resolution

The wave optics of the light microscope is generally applicable. Because of the smaller working aperture of the electron microscope, for a TEM, the limiting resolution  $d$  as given by eqn [1] reduces to

$$d = 0.6\lambda/\alpha \quad [3]$$

where  $\lambda$  is the wavelength and  $\alpha$  is the semiangle of the TEM objective or the corresponding angle in the probe-forming lens of an SEM or a TEM. For example, for a resolution of 1 nm in an aberration-free TEM operating at 100 kV,  $\alpha = 2.2 \times 10^{-3}$  rad. Such an angle is minute compared with that of a light microscope; this is a consequence of the remarkably small wavelength of the electron. Fortunately, electron sources are many orders brighter than the light sources used on optical microscopes, so that the intensity of the illumination in such a small solid angle is adequate to provide a satisfactory image at the  $100\,000\times$  or so magnification needed to resolve this detail on the fluorescent screen. It is quite feasible with modern image recording systems to work, in ordinary room lighting, at magnifications of several million times. In fact, such magnifications are often essential in the critical alignment and adjustment of a TEM and an STEM.

The resolution  $d$  in eqn [3] is analogous to the 'half-width' of the Airy disk that is observed in astronomical telescopes when viewing point objects such as stars. The Airy disk arises from the diffraction of rays from the finite aperture of the optical system; its diameter can be reduced by increasing the aperture of the lens. In an electron microscope, this cannot be done because of the severe spherical aberration of electron lenses (correctors for spherical aberration are being developed now, so this situation could change in the near future). The effect of spherical aberration is to broaden the Airy disk and reduce its height, thereby reducing image contrast and worsening the resolution. A simple way of visualizing this broadening is to note that the disk diameter of spherical aberration in the Gaussian image plane is given by

$$d = 2C_s\alpha^3 \quad [4]$$

where  $C_s$  is known as the coefficient of spherical aberration. In an electron microscope,  $C_s$  is typically of the order of 1 mm or so for the objective lens. A point

object will thus appear as a disk whose diameter increases rapidly with  $\alpha$ . The diffraction error decreases, but more slowly, with increase in  $\alpha$ .

There is an optimum angle, as indicated by wave optics. For example, the effect of spherical aberration of the objective is to distort the wave leaving the specimen from the spherical shape produced by a perfect lens. This unwanted extra pathlength PD inserted into the wave is given by

$$PD = \frac{C_s\alpha^4}{4} \quad [5]$$

Lord Rayleigh considered this problem and concluded that in any optical system an unwanted pathlength of  $\lambda/4$  or less can be tolerated, as having an extremely small effect on the size of the Airy disk.

Hence, if  $C_s\alpha^4/4 = \lambda/4$  is set, then

$$\alpha_{\text{opt}} = \left(\frac{\lambda}{C_s}\right)^{1/4} \quad [6]$$

is obtained for the optimum angle  $\alpha_{\text{opt}}$ . Substituting in eqn [3] gives the optimum resolution  $d$  as

$$d = 0.6(C_s\lambda^3)^{1/4} \quad [7]$$

In an STEM, with a Rutherford angular back-scatter detector, the imaging may be incoherent, i.e., largely free from diffraction effects. Under such conditions, the above constant is reduced to 0.43, a notable gain in resolution compared with a conventional TEM. Equation [7] is especially important for evaluating the capabilities of different TEM instruments. Notice that the resolution depends more strongly on  $\lambda$  than  $C_s$ , and this encourages the use of high-accelerating voltages (small  $\lambda$ ).

It is clear from eqns [6] and [7] that the objective aperture must be large enough to allow the finest detail required to pass through. The smallest aperture that satisfies this condition should be chosen. Otherwise, chromatic aberration may unnecessarily obscure the image. Chromatic aberration arises mainly from the fact that electrons passing through the specimen suffer energy losses. Plasmon losses occur that are associated with inelastically scattered electrons that suffered collective electron excitations giving rise to broad peaks in the range 0–50 eV. The emission of X-rays by various elements in the specimen also leads to corresponding discrete energy losses in the imaging beam. Moreover, in all electron lenses, the refractive power varies with the energy of the electrons traversing the lens. From a geometrical point of view, the diameter of the disk of confusion  $d_c$  in the Gaussian image plane is given by

$$d_c = 2C_c\alpha\frac{dV}{V} \quad [8]$$

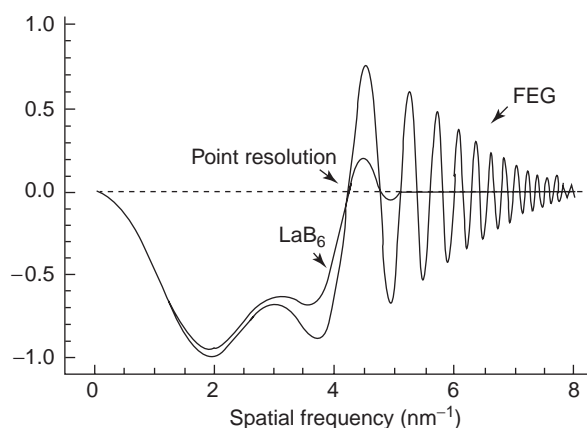
where  $C_s$  is the coefficient of chromatic aberration and  $\alpha$  is the semiangle of the cone of rays between specimen and focusing lens;  $V$  is the accelerating voltage and  $dV$  is the energy spread in the electrons. If the accelerating voltage of the lens excitation changes, further chromatic aberration will appear in the image; this may be avoided by well-designed electronic supplies having a stability of a few parts in a million for subnanometer resolution. In the limit, variations in the energy of the electron source, 0.1 eV for an FEG, can also destroy the contrast of fine detail in the image.

### The TEM Image

The TEM image is not simple to understand at high magnification, but the light microscope can provide some useful insights into the reasons for this. In the light microscope, there is no spherical or chromatic aberration. This means that the magnified image faithfully represents, in amplitude and phase, the wave leaving the specimen. The photographic plate or the CCD camera does not respond to phase and so this information is unfortunately lost. Such images are referred to as 'amplitude' contrast images. Zernike found that by placing, in the objective lens, a quarter-wave plate with a central hole to transmit the unscattered rays, an image containing phase information (phase-contrast image) could be produced.

Unfortunately, all attempts to create such a plate for the TEM have failed, causing a serious problem, since the chief contrast mechanism in the TEM at high resolution is phase contrast caused by elastic scattering at individual atoms. In the absence of a phase plate, some other means must be found to transform this into an amplitude contrast that can be recorded on the fluorescent screen or CCD camera. Spherical aberration itself (see eqn [5]) can insert a phase change into the wave leaving the specimen. It can also be arranged to provide a rough-and-ready phase plate. The spherical aberration coefficient of the lens is fixed, but by slightly decreasing the excitation current of the objective lens and hence its refractive power, to the so-called 'Scherzer defocus', a Zernike plate of sorts can be created that produces a phase-contrast image over a limited but useful range of spatial frequencies.

A typical phase-contrast transfer function is shown in Figure 2. This representation of TEM performance is analogous to the audio frequency response of a hi-fi amplifier. The units of spatial frequency here are reciprocal nanometers ( $\text{nm}^{-1}$ ), high values of spatial frequency corresponding to fine detail. The curve shows that the response of a conventional TEM in high-resolution mode is complicated and initially



**Figure 2** Typical phase-contrast transfer function of a high-resolution 200 kV TEM for  $\text{LaB}_6$  and field-emission guns. (Reprinted with permission from Coene W, Janssen G, Op de Beeck M, and Van Dyck D (1992) Phase retrieval through focus variation for ultra-resolution in field-emission transmission electron microscopy. *Physical Review Letters* 69: 3743–3746. © American Physical Society.) (Courtesy of Dr. W. Coene, Philips Research Laboratory, Eindhoven.)

mystifying. Practically no large area (low spatial frequency) phase-contrast information is transmitted. There is then a flat portion of the transfer function transmitting detail over the range 1–0.25 nm after which the contrast falls to zero. This point is known as the 'interpretable resolution limit', in this case  $\sim 0.25 \text{ nm}$  ( $4 \text{ nm}^{-1}$ ). Beyond this point the curve oscillates rapidly in phase, making it impossible to interpret the image visually. Nevertheless, this part of the transfer function contains encoded information about the finer details in the specimen. Under coherent illumination, spherical aberration, unlike chromatic aberration, does not destroy this information and it can, therefore, in principle, be recovered.

The envelope of the transfer function tails off exponentially with spatial frequency as shown in Figure 2. This is caused partly by energy changes in the electron beam, in the limit from the energy spread at the gun cathode. Figure 2 shows the transfer function for a lanthanum hexaboride and for an FEG, which has a lower energy spread. The other factor is the finite spatial coherence, or angular spread, in the illuminating beam. Here, the FEG scores again because of its much higher beam brightness (current density per solid angle). As soon as the transfer falls to the level of random noise, information is irretrievably lost. This important point is known as the 'information limit', in this case  $\sim 0.125 \text{ nm}$  for an FEG.

An FEG is essential for reaching the information limit by means of image decoding methods, and similarly the mechanical and electronic stability have to be compatible with this limit. From the point of view



of an operator working beyond the interpretable limit, what is more serious about spherical aberration is that, although fine detail remains reasonably sharp, this aberration can encode image detail. Thus, amplitude contrast is wrongly converted to phase contrast and vice versa, making the task of image interpretation very difficult at the atomic level and necessitating extensive image-simulation checks.

Fortunately, it is now possible to record phase- and amplitude-contrast information independently by electron beam holography, invented in 1948 by D. Gabor, but only made possible at the atomic level by developments in digital computing in the early 1990s. In a form of this method, known as 'off-axis holography', which has been brought to success by H. Lichte and his colleagues at the University of Tübingen, the TEM specimen is illuminated by a coherent beam of electrons from an FEG. Part of the beam is prevented from passing through the specimen, but acts as a reference beam for determining the phase of the wave. The two beams are merged lower down in the column by a Fresnel electron biprism located in the plane of the selected area aperture above the first projector. An off-axis interferogram (hologram) is formed that contains the complete information, in amplitude and phase, about the wave leaving the specimen, including the distortion of the wavefront by electron-specimen interaction, spherical aberration, defocusing, etc. The hologram is digitized and fed into a computer, which corrects the spherical aberration and astigmatism and produces images in amplitude and in phase contrast, just as is possible in the light microscope, but at atomic resolution. Such images form an accurate starting point for the determination of the object structure. Currently, modern instruments can be easily adapted for off-axis holography.

In 1992, an alternative technique for the correction of spherical aberration was developed by W. Coene and his colleagues at Philips Research Laboratory, Eindhoven, and at the University of Antwerp (RUCA). This is more computer intensive but does not require a biprism or any modification of the TEM column. It makes use of the fact that with coherent illumination the individual transfer functions of a focal series of images overlap and therefore contain, in coded form, the complete information about the wave leaving the specimen, not just to interpretable resolution limit but right down to the information limit of the microscope. Using the terminology of radio communication, which gives a good insight into the methods of holography, the basic carrier wave has to be separated from the two sidebands carrying the information about the specimen. In 'in-line' holography, as originally proposed

by Gabor, all three are superposed, so that the sidebands can interfere with each other, causing serious artifacts in the reconstruction. By advances in programming techniques, it is now possible to isolate the offending nonlinear terms that arise in the analysis of the images. By inserting the known spherical aberration coefficient into the program and adopting a least-squares fit approach to a set of some six or seven images in the series, it has proved possible to achieve aberration-free atomic resolution in a commercially produced 300 kV TEM. This is a great step forward since this procedure can be carried out on any high-performance TEM equipped with a CCD camera.

Using phase-contrast imaging, correction of the spherical aberration constant, or electron holography, electron microscope resolution has been increased over the years in order to improve visual interpretation of the images. The state-of-the-art spatial (point) resolution in a TEM has been limited to  $\sim 0.15$  nm, although this value is now decreasing due to recent improvements in microscope design and computer processing. More common dedicated high-resolution TEMs have a spatial resolution of 0.17 nm, while general-purpose microscopes have spatial resolutions of slightly more than 0.2 nm.

Since the work of Crewe in the early 1970s and pioneered by Pennycook in 1988, the Z-contrast imaging technique represents a different approach to atomic resolution electron microscopy. It allows 'incoherent' imaging of materials and provides directly interpretable images maps of the specimen at the atomic scale, in which higher atomic numbers (Z) show brighter. The Z-contrast image is obtained by scanning an electron probe of atomic dimensions across the specimen and collecting incoherent electrons elastically scattered at high angles, using a high-angle annular dark-field detector. The scattered intensity is understood as a sum of independent scattering from individual atoms, so the incoherent images of the Z-contrast method are interpreted more directly in terms of atom types and positions, since each summation, or integration, has the effect of suppressing the interference contrast that tends to produce image artifacts in phase-contrast images. Effects of contamination and beam damage must be minimized. In 1993, following the investigations of Browning and Pennycook, the Oak Ridge National Laboratory (TN, USA) took delivery of the first 300 kV Z-contrast STEM, providing a resolution of 0.13 nm.

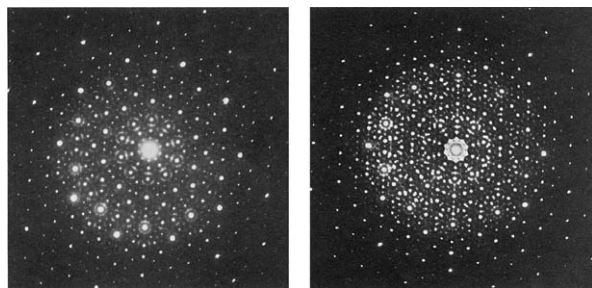
### Correction of Chromatic Aberration

It should be noted that the direct correction of spherical and chromatic aberrations in electron optical

lenses has been actively pursued since about 1936 by the insertion of quadrupoles, sextupoles, and octupoles into the column. In recent years, good progress has been made, leading the way to the commercial development of aberration-corrected TEMs. In 1998, Haider and co-workers, following the work of Rose, introduced the first successful aberration corrector for the imaging geometry used in a conventional TEM. Current studies on spherical aberration corrector open the door to a three-dimensional, sub-angstrom imaging of atomic arrangements.

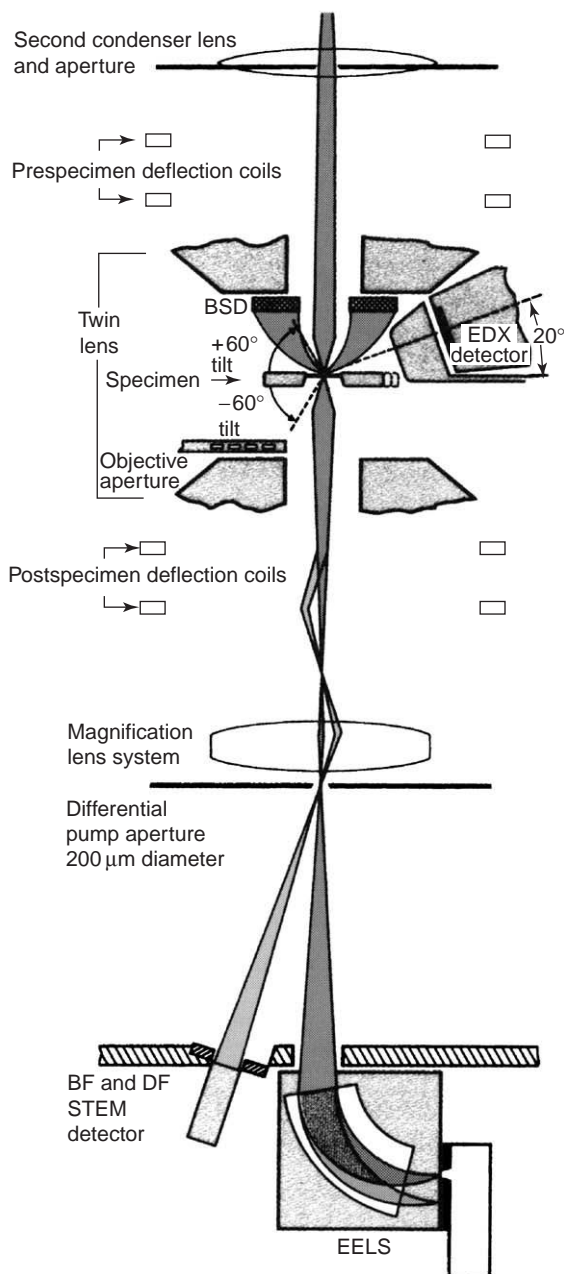
What firstly succeeded is electron spectroscopic imaging in the TEM, by means of focusing energy spectrometers. The basic idea here is to design an energy spectrometer that also has the properties of an axially symmetric (round) lens and so can take its place in the TEM column as an extra projector lens between the final and intermediate projectors. Such a device has several advantages and was available in certain Zeiss TEMs for sometime.

The electron diffraction pattern that appears in the back focal plane of the objective lens is also a powerful analytical tool. Unfortunately, this pattern is often overlaid by the diffuse background of inelastically scattered electrons, which do not contribute to the diffraction pattern but are defocused by the chromatic aberration of the objective lens. This causes loss of contrast in the pattern and also in a high-resolution image. These effects can be removed by allowing only elastically scattered electrons to reach the final screen. Figure 3 shows the results of this removal; it demonstrates that, with this designed filter using only magnetic focusing and correcting elements, the quality of the pattern is better than in unfiltered TEMs, even those working at more favorable, higher voltages.



**Figure 3** Diffraction pattern of a decagonal quasicrystal in CoNiAl alloy. Comparison between selected-area diffraction patterns of the same specimen, a CoNiAl quasicrystal, in a conventional (unfiltered) 200 kV TEM (left) and in a 125 kV EM912 Omega TEM with magnetic energy filter (right). Filtering enhances the visibility of low-intensity spots and fine structure close to the direct beam. Specimen and image: Max Planck Institute for Metal Research, Stuttgart, Germany. (Courtesy of Carl Zeiss, Germany.)

With a selective energy imaging filter, positive advantage can be taken, for analytical purposes, of the elemental information in electrons that have suffered a defined energy loss through ionization followed by X-ray emission (element mapping). Previously it was necessary to use the STEM mode for such work with a conventional energy spectrometer as



**Figure 4** Schematic arrangement of the electron probe micro-analysis detector and backscattered detector (BSD) in the twin-lens objective Philips CM30 TEM/STEM and the STEM and EELS (energy loss) detectors below the final projector lens. The postspecimen beam deflection system is used to deflect the electron beam to either detector. (Courtesy of Philips Ltd., Eindhoven.)

shown schematically in Figure 4. With an electron probe of  $\sim 2$  nm, one can obtain STEM images and electron energy-loss spectra (EELS), thereby obtaining an elemental map of the specimen. This can now be done in parallel operation with an imaging filter in the TEM. The first practical imaging energy filter was devised by Castaing and Henry in 1962 and eventually fitted in a Zeiss EM902A TEM between the first and second projector lenses as shown in Figure 5. The magnetic prism bends and focuses the incoming electrons into an electrostatic mirror that reflects the electron beam back onto the optical axis of the TEM. Because of the high degree of symmetry, aberrations are minimized and the system acts as a one-to-one magnification round lens for both images and diffraction patterns.

In the early 1990s, a magnetic filter (the Omega energy filter), capable of operating at higher accelerating voltages, was realized by Rose and Krahl in Germany and commercially available in the Zeiss (now LEO) EM912 Omega TEM. An important operational point here is that the imaging of all electrons is carried out at the same accelerating voltage as far as the column is concerned. For example, when

imaging electrons that have lost 200 eV, the computer increases the accelerating voltage by 200 eV. This brings these electrons to the nominal accelerating voltage of the microscope and avoids alignment difficulties and ensures high precision of measurement. Because of some delocalization of inelastically scattered electrons, the image resolution is  $\sim 2$  nm, similar to that for standard 200 kV TEMs equipped for X-ray probe analysis. Element mapping is therefore a very useful and rapid analytical feature in a TEM. In 1994, Uhlemann and Rose proposed a new imaging magnetic energy filter (the Mandoline energy filter, also developed by LEO), which even increased the energy resolution by one order of magnitude compared with then commercially available filter microscopes.

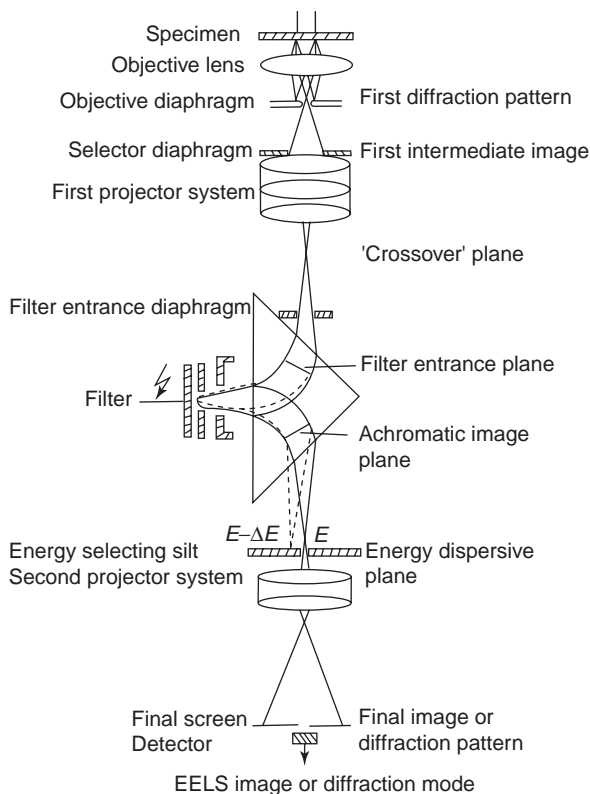
## Instrumentation and Results

Nowadays, there is a tendency to develop integrated multipurpose EM systems based on TEM/STEM or SEM instruments equipped with multiple detector systems, and using different electron and X-ray diffraction methods and observation modes. As a result, conventional TEM, STEM, and SEM instruments have evolved into a true AEM tool, which uses most of the signals produced during interaction between electron beam and sample.

### TEM and STEM

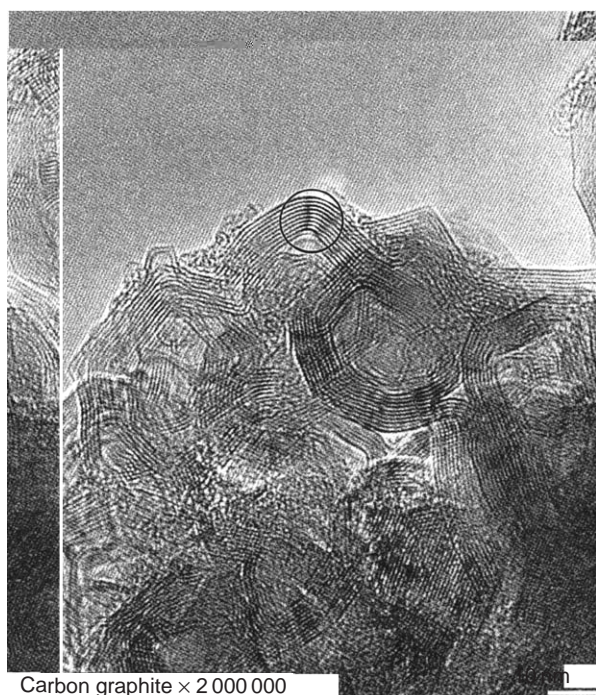
After 70 years of evolution, today's electron microscopes follow broadly similar electron optical designs, differing mainly in their orientation to either materials science or biological applications and in the special facilities that are required in different disciplines. It is not possible here to give a comprehensive survey but only to select some representative designs.

Referring again to the Topcon ABT EM-002B TEM shown in Figure 1, this design is interesting since it occupies an intermediate position between standard 100 kV TEM designs and 300 kV TEMs using techniques to produce aberration-free images. The EM-002B column is completely computer controlled allowing it to switch rapidly between TEM operation with a resolution of 0.18 nm, as shown in Figure 6, to electron probe X-ray microanalysis in the specimen stage with a probe size of 2.5 nm. The ringed area in Figure 6, in which the lattice planes in the specimen are clearly resolved, indicates the sort of microarea that becomes instantly available for selected area diffraction and X-ray microanalysis. The minimum probe size is 0.5 nm, a considerable design achievement in a TEM. In order to manage



**Figure 5** Schematic arrangement of the operating modes of the Castaing-Henry magnetic sector, electron mirror, imaging spectrometer as incorporated in the Zeiss EM902A TEM, showing path of monoenergetic electrons  $E$  through the system. (Courtesy of Carl Zeiss, Germany.)



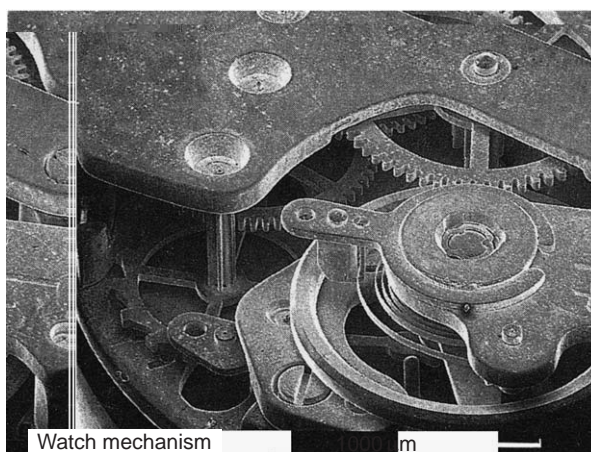


**Figure 6** Graphitized carbon specimen taken at high magnification in the 200 kV Topcon EM002B. Circle indicates the microregion from which selected-area electron diffraction and electron probe microanalysis can be obtained. (Topcon, Japan.)

this effectively, five condenser lenses, including the prefield of the objective, are needed, with close computer control of excitations to enable the probe to be focused without upsetting the focus of the objective. The correct setting of the objective lens for STEM and microdiffraction operation is aided by two minilenses placed on each side of the lens gap. An interesting feature is that the polepieces of the objective lenses may be exchanged, in a matter of 15 min, for example, in order to have a larger angle of tilt (with a slight loss of resolution), without dismantling other parts of the column.

### Scanning Electron Microscope

These instruments rely heavily on electron-optical developments from a TEM. The chief advantage of an SEM is that virtually no specimen preparation is necessary and large objects may be examined. For example, Figure 7 shows the escapement of a wristwatch, taken at a magnification of  $6 \times$ . The depth of focus is extremely high compared with that of a light microscope and it is even possible to take video pictures of the mechanism in operation. Figure 8 shows the cross-section of the column of the Topcon SM-700/700 SEM, an instrument with, unusually, two mechanical stages. The top stage is for high-resolution (2 nm) work, with the specimen up to 8 mm in



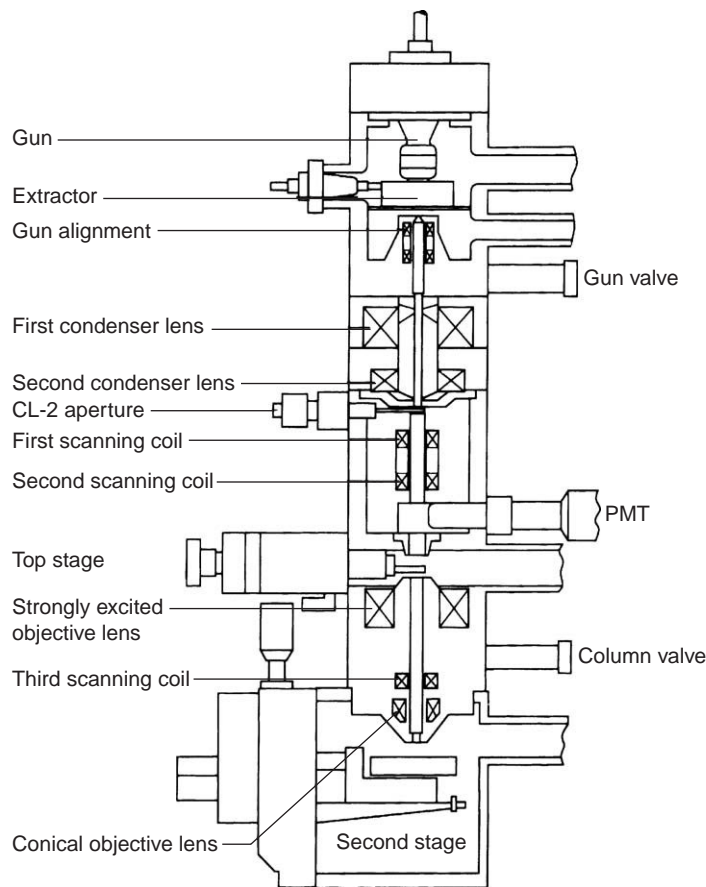
**Figure 7** SEM image of the escapement of a wristwatch, taken in a CamScan Series 4 SEM, illustrating the remarkable depth of focus compared with the light microscope. (Courtesy of Obducat CamScan Ltd.)

diameter; the lower stage is for a more modest resolution (5 nm) but with specimens up to 100 mm in diameter. The accelerating voltage may vary from 40 to 2 kV. At 2 kV the resolution is a useful 8 nm. A  $\text{LaB}_6$  cathode is needed to achieve these results. This design illustrates what trade-offs have to be made in the examination of a wide range of specimens. For even higher resolution, commercial SEMs are available with field-emission cathodes. The gain in resolution is typically by a factor of 2 or more, but greater care must be taken to ensure a pressure less than  $10^{-7}$  Pa in the gun region.

One may conclude this section by noting that TEM, STEM, and SEM are fast approaching the limits set by basic physics. State-of-the-art instruments rely increasingly on computers to carry out the complex adjustments that are needed for high performance, and also to record and produce the images and spectra that can be acquired.

### Limitations of Electron Beams in Microscopy

In the early days of electron microscopy the instruments were fairly daunting to use. Most of those early problems have now been brought under control by the manufacturers. In addition, computer control of alignment, autofocus, and automatic control of the intensity of illumination as a function of magnification are now common features of current microscopes. CCD cameras and other computer-based methods now produce high-quality images. For digital data acquisition, CCD cameras offer excellent sensitivity and linear response over a wide range of intensity. They provide image output in near real time, and



**Figure 8** Cross-section of the column of the Topcon SM-700/700 Dual-Stage SEM, showing the high-resolution objective in the top stage and the versatile conical objective lens in the second stage. (The Topcon Corporation, Japan.)

CCD cameras are often integrated into postcolumn energy filters, although these cameras do not provide the same resolution as the photographic film. The much superior sensitivity of CCD cameras at low electron exposures is a clear advantage for beam-sensitive materials, however. Many STEMs designed for research in surface science operate routinely at ultra-high vacuum, essential in critical work in this area.

With the great range of TEMs, STEMs, and SEMs now available, the range of acceptable specimens is extremely wide. Low-voltage ( $< 1$  kV) SEMs can give useful information on beam-sensitive materials such as biological specimens or very large scale integrated circuits. At the other end of the range the atomic perfection of a heterojunction in a semiconductor device can be demonstrated by a 300 kV TEM.

Specimen preparation methods have also improved enormously in the last decades and a wealth of documentation is now available. Specimen preparation methods such as ion beam thinning enable one to prepare thin transverse sections of an interface that can be imaged at atomic resolution in a 300 kV TEM. With so many techniques available, it is not

possible to say in advance what is or is not feasible in specimen preparation. A limitation in practice may well be the cost of preparing 'difficult' material to the high standard that is often demanded.

The ultimate limitation of electron beams in microscopy is, of course, the radiation damage they can inflict on the specimen. This is fundamental and unavoidable, although the damage extent depends on the electron dose needed to collect a correct signal, which in its turn is also influenced by the instrument performance. The atomic resolution that can be achieved in the electron microscope arises from the strong interaction between electrons and matter right down to the atomic level. This can cause rapid degradation in many crystal structures, but can often be alleviated by cooling the specimen to liquid-nitrogen or liquid-helium temperatures and carefully organizing the imaging system to minimize the number of electrons needed to take an exposure. The other side of the coin, of course, is that radiation damage in materials can often be studied more effectively and cheaply in a million-volt TEM than in a nuclear reactor!



See also: **Electron Energy Loss Spectrometry. Microscopy:** Overview. **Microscopy Techniques:** Light Microscopy; Scanning Electron Microscopy.

## Further Reading

- Amelinckx S, Van Dyck D, Van Landuyt J, and Van Tendeloo G (eds.) (1997) *Electron Microscopy, Principles and Fundamentals*. Weinheim: Wiley-VCH.
- Chapman SK (1986) *Maintaining and Monitoring the Transmission Electron Microscope*. Royal Microscopical Society Handbook 8. Oxford: Oxford University Press.
- Fultz B and Howe JM (2002) *Transmission Electron Microscopy and Diffractometry of Materials*, 2nd edn. Berlin: Springer.
- Hawkes PW (ed.) (1982) *Magnetic Electron Lenses*. Berlin: Springer.
- Hawkes PW and Kasper E (1989) *Principles of Electron Optics*; vol. 1, *Basic Geometrical Optics*; vol. 2, *Applied Geometrical Optics*. London: Academic Press.

- Kruit P (ed.) (1993–1994) *Ultramicroscopy*, vols 51–53. Amsterdam: North-Holland.
- Mulvey T and Sheppard CRJ (eds.) (1991–1994) *Advances in Optical and Electron Microscopy*, vols 12, 13, and 14. London: Academic Press.
- Reimer L (1998) *Scanning Electron Microscopy, Physics of Image Formation and Analysis*, 2nd edn. Springer Series in Optical Science, vol. 45. Berlin: Springer.
- Reimer L (ed.) (1995) *Energy-Filtering Transmission Electron Microscopy*. Springer Series in Optical Sciences, vol. 71. Berlin: Springer.
- Reimer L (1997) *Transmission Electron Microscopy, Physics of Image Formation and Analysis*, 4th edn. Springer Series in Optical Science, vol. 36. Berlin: Springer.
- Ruska E (1980) *The Early Development of Electron Lenses and Electron Microscopy* (translated Mulvey T). Stuttgart: Hirzel Verlag.
- Williams DB and Carter CB (1996) *Transmission Electron Microscopy, A Textbook for Materials Science*. New York: Plenum.

## Specimen Preparation for Electron Microscopy

**R Langford**, Trinity College Dublin, Dublin, Republic of Ireland

© 2005, Published by Elsevier Ltd.

This article is a revision of the previous-edition article by Dawn Chescocoe and Steve Chapman, pp. 3166–3174, © 1995, Elsevier Ltd.

## Introduction

Although it is common practice to separate materials science techniques from the life sciences, there is enough common ground in specimen preparation to combine them in one article.

However, the preparation for transmission electron microscopy (TEM), requiring thin samples (less than 100 nm) to allow the electron beam to penetrate the sample, will be discussed separately from scanning electron microscopy (SEM) methods since these normally require bulk preparation where sample stability and conductivity are the main criteria.

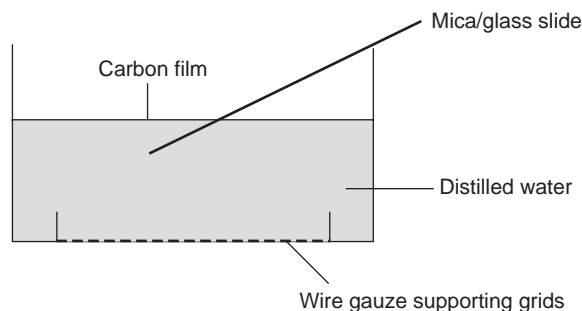
## Thin Sample Preparation Techniques

### Support Films

The most commonly used support films are either ultrathin carbon films or Formvar films stabilized with carbon. Essentially the carbon is vaporized in a vacuum chamber by passing a high current through two carbon rods. The carbon is deposited on to a

suitable substrate, normally freshly cleaved mica or a glass microscope slide. Thickness can be estimated by the color of the carbon deposit on a white surface; light brown represents ~20 nm. The film on the coated mica is then floated off on to a water surface (Fig. 1), collected on to 3 mm grids, normally made from copper, and finally dried at room temperature.

This type of film is a suitable substrate for collection of particulate material as a 'dry-dusted' dispersion; as its name suggests, dry powder is sprinkled on the coated grid and the excess is removed, leaving relatively few particles adhering to the carbon film, usually sufficient for a standard for X-ray microanalysis. When a better dispersion is required, ultrasonic agitation is employed in an ultrasonic bath; if there is difficulty dispersing the sample, an ultrasonic



**Figure 1** Method for preparing carbon-coated grids.

See also: **Electron Energy Loss Spectrometry. Microscopy:** Overview. **Microscopy Techniques:** Light Microscopy; Scanning Electron Microscopy.

## Further Reading

- Amelinckx S, Van Dyck D, Van Landuyt J, and Van Tendeloo G (eds.) (1997) *Electron Microscopy, Principles and Fundamentals*. Weinheim: Wiley-VCH.
- Chapman SK (1986) *Maintaining and Monitoring the Transmission Electron Microscope*. Royal Microscopical Society Handbook 8. Oxford: Oxford University Press.
- Fultz B and Howe JM (2002) *Transmission Electron Microscopy and Diffractometry of Materials*, 2nd edn. Berlin: Springer.
- Hawkes PW (ed.) (1982) *Magnetic Electron Lenses*. Berlin: Springer.
- Hawkes PW and Kasper E (1989) *Principles of Electron Optics*; vol. 1, *Basic Geometrical Optics*; vol. 2, *Applied Geometrical Optics*. London: Academic Press.

- Kruit P (ed.) (1993–1994) *Ultramicroscopy*, vols 51–53. Amsterdam: North-Holland.
- Mulvey T and Sheppard CRJ (eds.) (1991–1994) *Advances in Optical and Electron Microscopy*, vols 12, 13, and 14. London: Academic Press.
- Reimer L (1998) *Scanning Electron Microscopy, Physics of Image Formation and Analysis*, 2nd edn. Springer Series in Optical Science, vol. 45. Berlin: Springer.
- Reimer L (ed.) (1995) *Energy-Filtering Transmission Electron Microscopy*. Springer Series in Optical Sciences, vol. 71. Berlin: Springer.
- Reimer L (1997) *Transmission Electron Microscopy, Physics of Image Formation and Analysis*, 4th edn. Springer Series in Optical Science, vol. 36. Berlin: Springer.
- Ruska E (1980) *The Early Development of Electron Lenses and Electron Microscopy* (translated Mulvey T). Stuttgart: Hirzel Verlag.
- Williams DB and Carter CB (1996) *Transmission Electron Microscopy, A Textbook for Materials Science*. New York: Plenum.

## Specimen Preparation for Electron Microscopy

**R Langford**, Trinity College Dublin, Dublin, Republic of Ireland

© 2005, Published by Elsevier Ltd.

This article is a revision of the previous-edition article by Dawn Chescocoe and Steve Chapman, pp. 3166–3174, © 1995, Elsevier Ltd.

## Introduction

Although it is common practice to separate materials science techniques from the life sciences, there is enough common ground in specimen preparation to combine them in one article.

However, the preparation for transmission electron microscopy (TEM), requiring thin samples (less than 100 nm) to allow the electron beam to penetrate the sample, will be discussed separately from scanning electron microscopy (SEM) methods since these normally require bulk preparation where sample stability and conductivity are the main criteria.

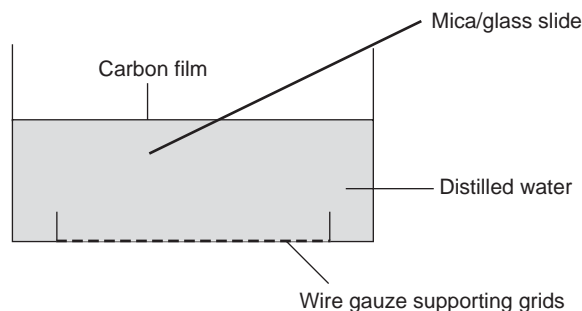
## Thin Sample Preparation Techniques

### Support Films

The most commonly used support films are either ultrathin carbon films or Formvar films stabilized with carbon. Essentially the carbon is vaporized in a vacuum chamber by passing a high current through two carbon rods. The carbon is deposited on to a

suitable substrate, normally freshly cleaved mica or a glass microscope slide. Thickness can be estimated by the color of the carbon deposit on a white surface; light brown represents ~20 nm. The film on the coated mica is then floated off on to a water surface (Figure 1), collected on to 3 mm grids, normally made from copper, and finally dried at room temperature.

This type of film is a suitable substrate for collection of particulate material as a 'dry-dusted' dispersion; as its name suggests, dry powder is sprinkled on the coated grid and the excess is removed, leaving relatively few particles adhering to the carbon film, usually sufficient for a standard for X-ray microanalysis. When a better dispersion is required, ultrasonic agitation is employed in an ultrasonic bath; if there is difficulty dispersing the sample, an ultrasonic



**Figure 1** Method for preparing carbon-coated grids.

probe immersed in the suspension is more likely to be successful. The choice of solution will depend on the material to be dispersed but could range from water to a solvent such as petroleum ether. The choice will depend on solubility and density.

Although carbon films are easily prepared, they need to be made hydrophilic by exposing them to a glow discharge or ultraviolet (UV) irradiation to ensure an even distribution of a water-based suspension such as a virus concentrate. The final preparation of the virus concentrate would have to be stained with a negative stain such as phosphotungstic acid; the stain surrounds the particles and these then appear bright against a dark background. Alternatively, Formvar films stabilized with carbon can be used. One of the methods for making these films is to prepare a 0.5% (w/v) solution of Formvar in chloroform, spread this evenly over a clean microscope slide, and cut round the edges of the film. The film is then floated off on to a distilled-water surface, collected on to copper grids, and finally coated with ~10 nm of carbon.

A modification to the above film, made by adding glycerin to a 0.25% Formvar–chloroform solution and dispersing this by ultrasonic agitation, will produce a network of holes. After collection of the film on to grids, methanol is used to remove the glycerin. The grids are then carbon coated and immersed in chloroform to remove the Formvar, thus producing an ideal support for material where the carbon structure is limiting the performance. Only regions across holes are examined. The concentration of glycerin will determine the size of hole. If the film is being used only as a test specimen, ~2% is adequate; when it is used purely as a supporting net, 10% will be more suitable.

### Replicas and Extraction Replicas

Although replication of surfaces is not required very frequently, there are occasions when it is not possible to view the entire surface in an SEM owing to size or perhaps when a timed sequence of events is involved. More frequently it is necessary to extract precipitates, inclusions, etc., from a matrix and to view these in a TEM without interference from the matrix.

There are two possible ways to prepare these replicas: one will not damage the sample and is ideal when the inclusions, etc., are easily removed; the other is a destructive technique that requires the matrix to be dissolved, leaving only particles of interest on the carbon film.

In the first case, cellulose acetate sheet is moistened with acetone and 'rolled' on to the surface, where it is left for ~30 min and then stripped off the surface,

attached to a glass slide and coated with ~20 nm of carbon. The cellulose acetate is then removed in acetone by placing grids on to a metal gauze with sufficient acetone to allow capillary action to wet the surface of the gauze. (This process is left in a covered environment for several hours.) Although this is a slow method of removing the acetate, it does prevent the carbon film breaking and the grid is ready for examination as an extraction replica. If the replica is for surface examination only, it will be necessary to shadow at an angle with some suitable metal (Au/Pd, Pt, or Pt/C) to improve image contrast.

For the second method, the surface is lightly etched to reveal the particulate to be removed, then coated with 20–40 nm of carbon. The carbon is scored into ~3 mm squares and sufficient matrix is removed to allow the film to float off on to the etching solution. The carbon film is then transferred to a clean water bath where it can be collected on to the grids.

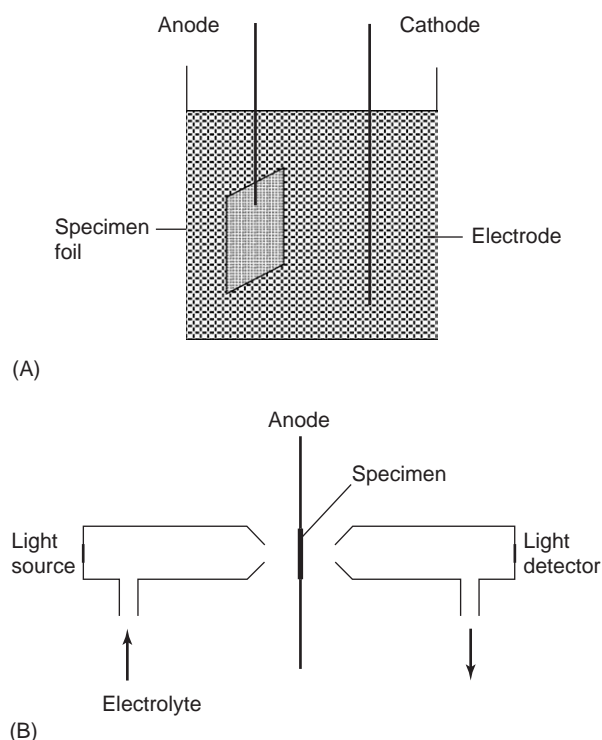
### Thin Samples for Materials Science Applications

There is a need to make a thin sample that represents the bulk without introducing defects into the thin sample. The sample is ideally less than 100 nm thick, of uniform thickness, and both mechanically stable and electrically conductive under the electron beam. Methods to be considered are electropolishing, ion beam thinning, and ultramicrotomy.

The choice of method depends on the sample; if it is conducting, e.g., a metal foil, then electropolishing will normally be the answer since this is the fastest method available. However, if the sample is non-conducting, or is multiphase with one phase thinning faster than the others, then ion beam thinning will be preferable. If the sample is a powder or a thin wire, then the ultramicrotome may provide a solution depending on the hardness of sample.

Before a description of the final thinning stage it is necessary to outline initial preparation. (1) Cut a slice as thin as possible, realistically 200–500  $\mu\text{m}$ . (2) This is mechanically polished to 50–100  $\mu\text{m}$ . (3) Punch/drill out a 3 mm disk. (4) Preferentially polish the center of the disk by 'dimpling' the sample (optional).

**Electropolishing** Essentially this is the reverse method of electroplating. Here the sample is held in an electrolyte (5% perchloric in methanol is a good starting point), with the sample as the anode and a stainless-steel cathode. When a suitable voltage is applied, material is removed from the sample, usually to perforation, allowing a thin area to be examined around the hole. In the method's simplest

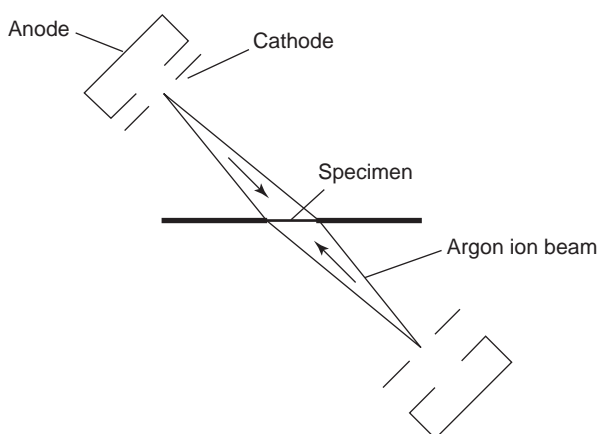


**Figure 2** (A) Simple electropolishing system with the specimen supported by forceps as the anode. (B) A schematic diagram of the electropolishing jet thinner (cathode not shown).

form a thin foil may be used as shown in **Figure 2A**, with an insulating layer painted around the contacts; this is usually 'Lacomite' the trade name for a solvent-based lacquer colored red to render it more visible. A DC power unit supplies a voltage from 0 to 120 V with typical values of 30 V and 0.1 A; perforation of the foil is decided by visual inspection. A more commonly used thinning system is one that is designed to produce a jet of electrolyte on either side and at the center of the disk, with a beam of light used to detect perforation (**Figure 2B**). Temperature, voltage, and flow rate are all variables that allow the sample to be polished rather than etched, with typical polishing times of a few minutes. The initial polishing action is to remove any surface asperities to provide a smoother starting point that facilitates a faster process and a better final result.

A comprehensive list of sample preparation techniques and electrolytes for different materials is given in the microscopy handbook by Goodhew (see Further Reading section).

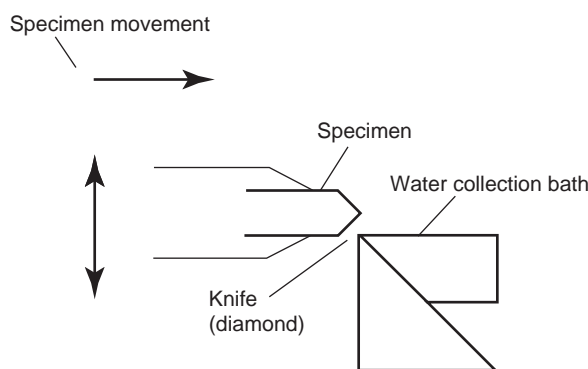
**Ion beam thinning** The idea of this technique is to bombard the sample with a beam of argon ions to remove surface atoms, thus thinning the specimen. The ion guns are designed to produce a fine beam of ions and/or neutral atoms. In its simplest form, an



**Figure 3** Schematic of an ion beam thinner showing the specimen mounted at 45° with respect to the ion guns.

ion gun is a hollow anode and cathode, with the applied voltage and resulting discharge ionizing the argon gas. The variable parameters that determine the thinning rate are current and voltage, the higher the current and voltage, the faster material is removed, but there is an upper limit as the sample will be damaged by the higher-energy argon ions. Specimen tilt will contribute to the size of thin area ultimately produced; the lower the glancing angle of the beam to the specimen, the larger the area (the geometry of the system will limit this to  $\sim 12^\circ$ ) but the longer the thinning time. It is therefore common practice to start with a higher angle and to reduce it part way through the sequence. A gun either side of the specimen allows the option of thinning from one or both sides (**Figure 3**). During the entire thinning process the sample is continually rotated. It is often necessary to reduce the specimen temperature to avoid heating effects and therefore a liquid-nitrogen stage is normally provided as an optional extra. Thinning times will vary for different materials; typical thinning times are 30 min for a 4  $\mu\text{m}$  copper foil and several days for a 100  $\mu\text{m}$  ceramic material. Even foils of the same thickness but different elemental composition have vastly different thinning rates. For example, the maximum sputtering yield expressed in atoms/ion (argon) will range from 2 for aluminum through 4 for elements such as iron and tungsten to 20 for zinc. Because of these differences, when thinning compound semiconductor material 'reactive' iodine gas is often used, since this limits the preferential thinning observed with argon ions by combining chemical and ion etching.

**Ultramicrotomy** This is normally associated with biological samples but can be used with great success on materials science samples. **Figure 4** outlines the principle of the technique: a knife, which has to be



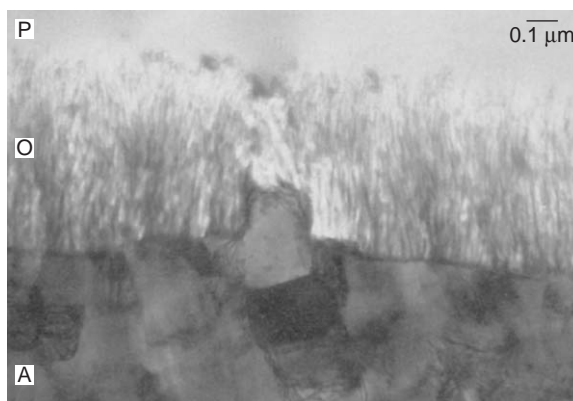
**Figure 4** Simplified diagram of an ultramicrotome demonstrating the movements of the specimen block.

diamond for most inorganic samples, is designed to move up and down against a sample advancing at rates of nanometers per step. Thin sections are produced and are normally collected on a water surface. The critical parameters are the knife angle, the shape of the specimen block, and the cutting speeds. The specimen block will have been produced by embedding the sample in a suitable resin that matches the hardness of the sample. For materials science use, a hydrophilic acrylic resin such as 'LR White', which offers grades from soft to hard, is ideal since it does not require mixing with hardener and cures at 60°C. Its main disadvantage is that air must be excluded during the curing process. This has been used successfully for embedding fine powders, thin wires, and multilayers of different material. A recent publication by Jacobs has shown that cyanoacrylate ester ('superglue') added before embedding can reduce the problem of the resin becoming detached from the material.

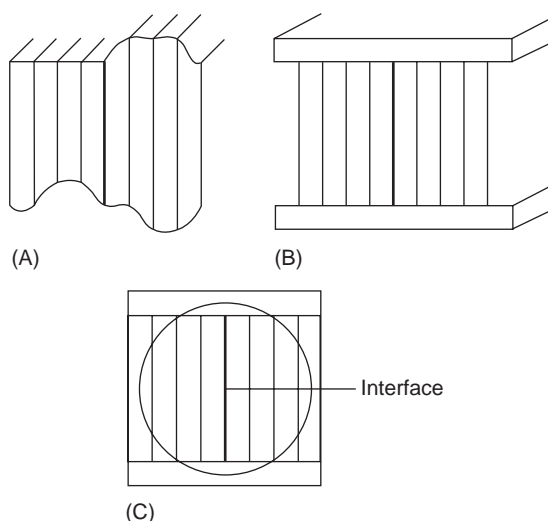
Samples of aluminum foil were prepared by each of the above techniques and, as expected, they produced essentially the same thinned specimen with the exception that the grain boundaries were less obvious in the sectioned sample (probably owing to the lack of any preferential thinning). The ion beam-thinned sample was more irregular since the initial surface was not perfectly flat. Since the sample is readily thinned by electropolishing, this would be the normal method selected.

An extension to normal ultramicrotomy is cryoultramicrotomy, in which, as the name suggests, the temperature of both the specimen and the knife is reduced. The temperature depends on the glass transition temperature of the material. A sample such as polymer film on aluminum substrate can be prepared only in this way, and **Figure 5** shows the aluminum foil, an oxide layer, and the polymer coating.

The techniques described so far have all produced plan-view samples, but occasionally it is of great



**Figure 5** Transmission electron micrograph of a polymer-coated aluminum foil: A = aluminum, O = oxide, P = polymer film.



**Figure 6** The initial stages in the preparation of a cross-section specimen: (A) bonded layers before polishing; (B) after the first polishing stage with a layer bonded to either end; (C) showing the position from where the 3 mm disk will be taken.

interest to examine layers such as GaAs and GaInAs deposited on a matrix. For this, it is necessary to produce thin 'cross-section' samples. Preparation of a cross-section is a tedious but necessary procedure as described by Brawman. A slight modification involves bonding several layers together (**Figure 6A**), and polishing top and bottom until flat. Two layers, one at either end, are bonded at 90° to make the structure more rigid as shown in **Figure 6B**. After mounting on a glass microscope slide, with cover slips either side for reference, the entire sample is polished to the level of the cover slips. A 3 mm disk is removed using a hollow drill (**Figure 6C**), dimpled and ion beam-thinned. With luck and careful alignment of the ion guns, the sample will perforate at the interface. With use of a variable speed of rotation during the thinning process (fast when the

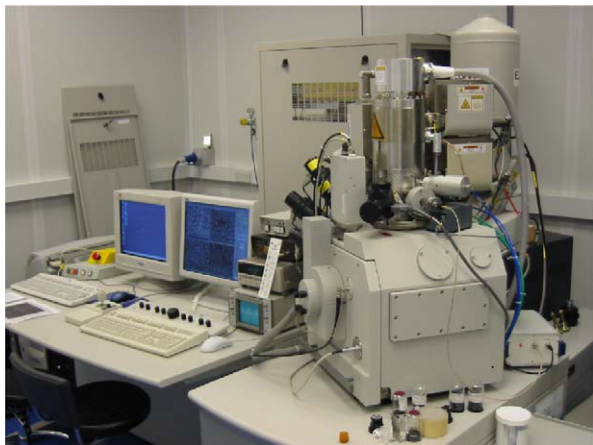


argon beam ‘sees’ the join), there is less chance of the specimen breaking at the vital point.

### Focused Ion Beam Sample Preparation

Two tools, which have become widely used during the 1990s for preparing specimens for material analysis, are focused ion beam (FIB) and FIB/SEM systems. As its name suggests, an FIB system uses a beam of focused ions (usually gallium) to sputter (cut) away material to prepare the specimens. Similarly to SEM, the FIB beam can be scanned over the sample and the generated secondary ions or secondary electrons (SE) used for imaging. The ions are field extracted from a liquid metal ion source and are accelerated (30–50 kV), collimated, and focused by a series of electrostatic lenses and apertures. A beam typically varies from 1 pA to 20 nA and the corresponding spot sizes range from 10 to 500 nm. An FIB/SEM has both an electron beam and an ion beam column, usually inclined at 52° to one another. **Figure 7** shows an FEI dual-beam FIB/SEM system.

FIB and FIB/SEM systems are routinely used to prepare TEM lamellae (both cross-section and plan views) and SEM specimens, which have been used for secondary ion mass spectrometry, electron backscatter diffraction, and Auger analysis. These systems offer numerous advantages relative to the other techniques described in the previous sections. The main advantage is that the specimens can be prepared to within 50 nm of a feature of interest (FOI). Other advantages are: (1) the lamella can be prepared with near parallel sidewalls enabling quantitative chemical analysis to be performed over the lamella; (2) the ion beam is perpendicular to the surface of the sample so that less preferential milling occurs than with ion beam thinning; and (3) the sample preparation time can be as short as 2 h.



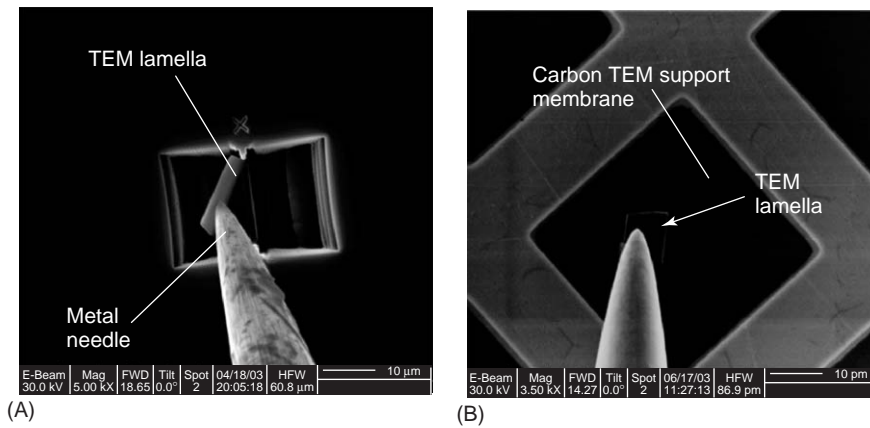
**Figure 7** A FEI dual-beam FIB/SEM system.

Currently, three techniques – the ‘lift-out’ technique, the trench technique, and the wedge technique – are used to prepare TEM lamellae. Typically, the prepared TEM lamellae are 10–15 μm wide, 8–10 μm deep, and 50–100 nm thick. In all three techniques, the first step involves depositing a 1 μm thick and 1 μm wide metal strap (usually Pt or W) using ion- or electron-beam-assisted deposition (a suitable precursor gas is bled into the chamber) over the FOI to protect it from being milled and to planarize the sample’s surface.

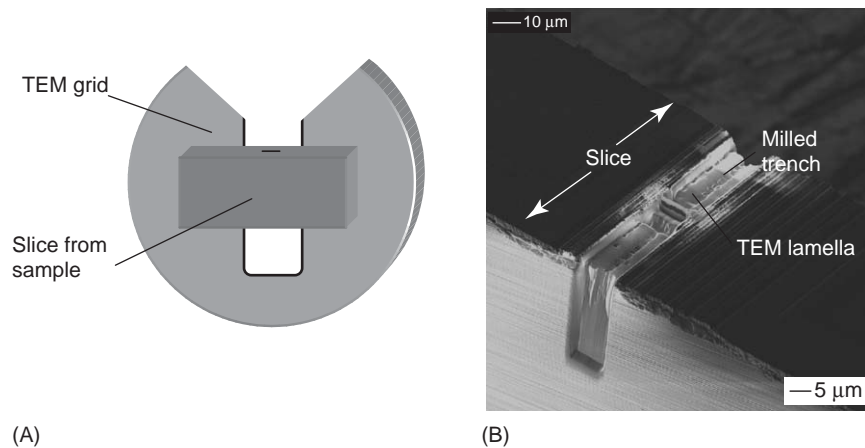
In the lift-out technique, a lamella is free from the sample and is lifted on to a carbon-coated TEM support grid using a needle and micromanipulator. Two cuts are milled on either side of the metal strap and then the material between the cuts is thinned to 300 nm by scanning the ion beam in lines (a similar procedure to slicing a loaf of bread). The sample is next tilted to 45° and the base and sidewall of the lamella are cut. After retilting to 0° the lamella is further FIB milled to electron transparency (50–100 nm) and its remaining sidewall is cut to free it from the sample. It is then lifted from the cuts by using either a glass or metal needle and a micromanipulator. This step is generally performed *ex situ* of the FIB chamber using an optical microscope but it can also be done in an SEM or FIB/SEM using secondary electron imaging. **Figure 8A** shows an SEM image of a free lamella that is being lifted out by a tungsten needle. The needle is repeatedly swept through the FIB cuts until the lamella attaches to it through electrostatic forces. A carbon-coated TEM grid then replaces the sample and the needle is swept across the carbon surface until the lamella attaches to the carbon membrane. **Figure 8B** shows a lamella that has just been placed on to the carbon support membrane by the needle. The success yield of this step is ~70%; the lamella may be ‘lost’ by falling off the needle during the lift-out step or when placing it on to the grid.

The trench technique has success yield of ~99% and involves preparing a slice of material from the sample (usually 5–100 μm wide) by dicing or polishing and fixing this on to a U-shaped TEM grid (**Figure 9A**). Trenches are then FIB milled from the edges of the slice to either side of the FOI to make the TEM lamella and a pathway for the electron beam in the TEM (**Figure 9B**). A drawback of this technique is that the trenches limit the angles to tilt in the TEM.

The wedge technique is a hybrid of the trench and the lift-out techniques, combining their relative advantages; it was developed using an FIB system and involves *in situ* micromanipulation. A wedged-shaped piece of material (typically 5–10 μm wide) is cut free from the sample by making four



**Figure 8** SE images of a lamella (A) being lift-out from the FIB cuts using a metal needle and (B) being placed on to a carbon support membrane.



**Figure 9** (A) Mounting of a slice on to a U-shaped TEM grid and (B) SE image of a finished trench cross-section showing the trenches and the TEM lamella.

interconnecting cuts. The end of a needle is then 'welded'/fixed, using *in situ* metal deposition, to the wedge (Figure 10A). The wedge is next lifted and positioned against the sidewalls of a TEM grid where it is fixed using metal deposition. At this stage the needle is cut from the wedge and moved away and the TEM lamella is FIB milled into the wedge. Figure 10B shows a wedge fixed to a TEM grid into which the lamella has been milled.

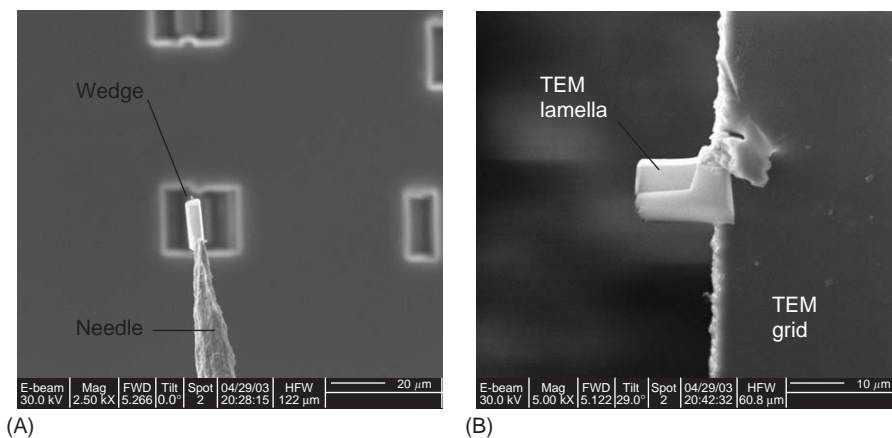
Variations of these three techniques are used to make plan-view lamellae. In the methods based on the trench and the lift-out techniques, the samples are cleaved or polished and placed end-on to the FIB. In the methods based on the wedge technique, the wedge is rotated by 90° before being fixed to the side of the TEM grid.

Site-specific SEM cross-sections are prepared in exactly the same way as TEM cross-section lamellae for the lift-out technique except the cuts are only made on one side of the Pt strap. This is shown schematically in Figure 11.

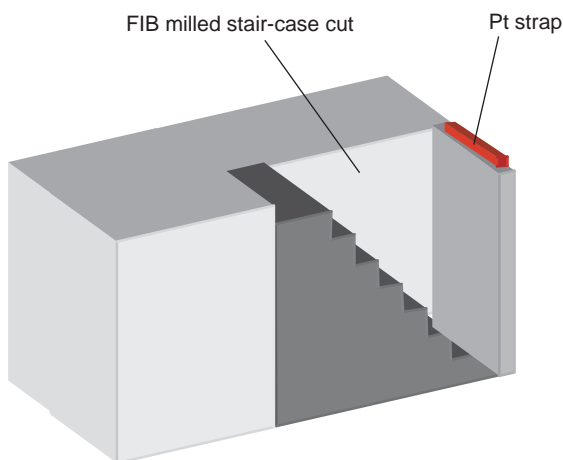
### Focused Ion Beam Artifacts

Unfortunately, the incident energy of the ion beam can create vacancies and interstitials at the sidewalls of the lamella, which, if of sufficient density, can result in dislocations or formation of an amorphous layer. For example, in silicon lamellae the amorphous layer at the sidewalls is 20 nm thick. However, there have been many reports of successful ways to reduce the depth of the damage at the sidewalls and to further reduce the thickness of the lamella. These include milling the sidewalls of the lamella using a 5–10 kV ion beam in the FIB chamber, using gas assisted etching, ion beam milling outside of the FIB chamber, and tilting the samples by large angles (5° and above) during the FIB milling to electron transparency.

However, despite these limitations, FIB and FIB/SEM systems have become very important tools for making site-specific specimens and preparing specimens from samples composed of materials with different sputter rates.



**Figure 10** SE images (A) of a needle attached to a wedge shaped piece of sample and (B) of a wedge that has been fixed to a TEM grid into which the TEM lamella has been milled.



**Figure 11** Schematic of a SEM cross-section.

### Biological Sample Preparation

Again the requirement is to produce a specimen representing the original live tissue but, since there have to be many chemical processes, doubt can be raised regarding the validity of the final result. Hence, controls are always run at the same time and under identical conditions, with changes from control being the most significant observation. If a new technique is employed, then comparison with previous established techniques is essential.

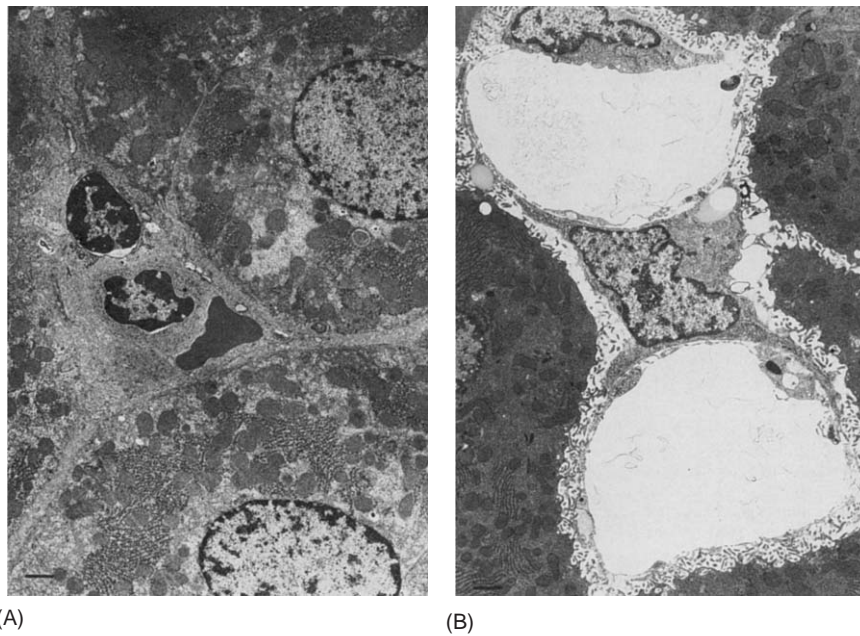
Which technique is used depends, to some extent, on what is required from the observation. For example, during processing by conventional methods, virtually all elements such as sodium and potassium are removed. Therefore, X-ray analysis for these elements would be useless; however, the morphology is totally acceptable. Other elements may survive the processing and the only consideration for X-ray analysis is interference of overlapping peaks due to the fixation and stain.

**Fixation** Speed is important since the tissue will deteriorate within minutes. The decision on immersion fixation, against per-fusion fixation, will depend on what else has to happen to the tissue – if biochemical analysis is required then obviously perfusion fixation is not an option. There is no doubt, however, that perfusion fixation can be an advantage, as shown in the comparison micrographs of liver in **Figure 12**. Here the sinusoid and associated cells are more readily seen in the perfusion-fixed sample.

Glutaraldehyde solution is used for the first stage, normally at ~4% in buffer, and the pieces of tissue are immersed for 4 h. The buffer is designed for correct osmolarity, ionic constitution, and pH of the living tissue. The tissue is then transferred to buffer and undergoes several washes before being postfixed with osmium tetroxide 1.2% for 2 h (these times depend on the size of the tissue sample). After several washes in buffer, the next stage is dehydration. Another common variation is to block-stain with uranyl acetate.

**Dehydration and embedding** It is essential to remove all water from the tissue. This is achieved by a graded series of water–alcohol mixes, starting with 50:50 water–alcohol to finally 100% absolute alcohol. The tissue is immersed in each wash for ~10 min, and the sample containers should be capped at all stages to avoid absorption of water from the atmosphere. After the final wash in absolute alcohol, depending on the next stage a secondary wash in a different solvent is required. If epoxy resin is to be used for embedding, then a change to propylene oxide is achieved by 50:50 dilutions and finally 100% propylene oxide.

Finally, the sample is embedded by immersing first in a 50:50 dilution of resin in propylene oxide, then



**Figure 12** Transmission electron micrographs of liver sections: (A) by immersion fixation; (B) by perfusion fixation.

to at least two changes of 100% resin, and curing at the appropriate temperature.

**Sectioning** The ultramicrotome described earlier is used to obtain sections that are  $\sim 40$  nm thick. Since biological tissue is easier to cut than inorganic samples, a glass knife may be used rather than diamond. However, the speed and performance advantages of diamond far outweigh the cost. The ribbon of section so produced can be picked up on a grid, which usually does not need to be coated with a support film. A thin (5 nm) carbon film evaporated on to the final section can, however, decrease charging in the TEM.

**Staining** An unstained section, even postfixed with osmium tetroxide, will result in a very low-contrast image. Poststaining with uranyl acetate and lead citrate, when specificity is not important, is the normal method for increasing the contrast for morphological examination. This is a positive staining technique in which the stain reacts chemically with some of the cell components to produce regions of high atomic number composition. The result is additional scattering of the electron beam and hence an increase in the image contrast.

If X-ray analysis is required, the addition of these heavy elements can interfere with the elements to be examined and it thus may be necessary to exclude one or both of the stains. It can be also an advantage to increase the thickness of the section so as to increase the volume of material to be analyzed.

Apart from the minor changes to the above procedure to compensate for different types of tissue, a significant change to fixation will normally be necessary to accommodate immunohistochemical techniques, since antigens may be destroyed by the fixation. This may be a change to 0.5% glutaraldehyde in paraformaldehyde and no postfixation in osmium, to removal of any conventional fixation chemical. In this case cryofixation is the solution, but subsequent sectioning is somewhat more involved.

**Cryotechniques** For TEM it is necessary to freeze the sample rapidly to avoid ice crystal damage, usually by 'slam freezing'. With this technique 10–20  $\mu$ m of the tissue surface will be suitable for morphological as well as chemical and biochemical analysis. The sample may be cut using a cryoultramicrotome and viewed either in its hydrated form or freeze-dried. This sample will be virtually chemically correct and certainly suitable for immunocytochemical analysis. The method suffers, however, from one disadvantage in that there is almost no contrast to determine the structure of the tissue in the TEM.

Freeze-substitution is perhaps of more use; here the sample is frozen rapidly and then held in an organic solvent such as ethanol for several days at  $-80^{\circ}\text{C}$  to remove the ice. After embedding in a resin and polymerization with a UV source, sections are cut and, if necessary, stained. The resultant section will be ideal for any examination and most types of analysis.

**Freeze–fracture** This technique is equally useful to TEM or SEM examination. Here the sample is rapidly frozen as previously described, and if a replica is required the sample is evacuated in a specialized coating unit. Once at a suitable vacuum, the sample is fractured, coated with carbon and shadowed with a heavy metal; after removal from the vacuum unit, the sample is thawed and the coating is removed to produce a carbon replica. The material will fracture at the weak points of the tissue, such as along membranes, revealing internal structure in this high-resolution image that can be examined in the TEM. It is also possible to freeze-etch the sample after fracture, but before replication, to reveal specific detail. As with any technique, considerable thought must be given to possible sources of artifacts.

For details of all the specimen preparation methods and related techniques see the Glauret series.

If SEM examination is required, there is no need to replicate the surface since it can be examined directly in the SEM, again with or without freeze-etching. (It will probably be necessary to coat the surface before examination.)

## Bulk Specimen Preparation

Almost any material can be observed within the SEM. Scientists routinely look at materials as diverse as contact lenses and engine parts, through to the fat particles in milk and crystals in diesel oil. The common factor between all types of specimens is the need to make them ready for viewing in the microscope.

The first question that a scientist should ask when considering a piece of material for SEM observation is ‘does it need to be prepared?’ Will the act of preparing the specimen damage the detail required to be imaged, or even remove the detail completely? If a specimen is being examined for the surface residue, almost any preparation may remove the information required. However, if the surface required resides beneath a contaminated surface, there is a possibility that this film will require removal. If the task is to determine elemental distribution by backscattered electrons, it is ideal to present a polished surface to the electron beam, as surface roughness will interfere with the signal produced. Similarly, a polished surface is also desired for most accurate X-ray analysis.

### Adhesives

The specimen will need to be mounted upon a metal disk, which is known as a stub. The disk may be the top surface of a short rod or the upper surface of a pin-shaped unit. Adhesion may be provided by a layer on the under surface of the specimen or, if the

specimen is transparent to the electron beam, by using the adhesive to tag the corners of the specimen to the stub. Proprietary clamping units may be used, but only if the specimen is sufficiently robust to tolerate the pressure exerted by the clamps.

**Proprietary adhesives** Many of the adhesives available for fixing metal to metal or porous material to metal will be suitable for use in the scanning electron microscope. The problems however are twofold: first, the adhesive may take a long time to give off its solvent; second, the adhesive may be nonconducting, insulating the specimen from earth. A nonconducting adhesive could create a surface charge that would destroy the image quality. The former may be overcome with patience, the latter either by careful control of the microscope parameters or by coating the specimen with carbon or a metal.

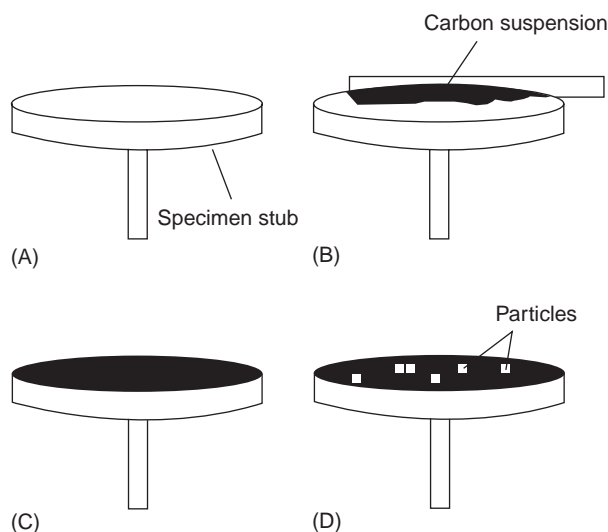
**Special adhesives** Silver Dag<sup>TM</sup> is the most commonly used adhesive in the scanning electron microscope. It not only dries quickly but also provides a conducting interface between the specimen and the mount. Carbon suspension, a carbon-based adhesive, shares with Silver Dag<sup>TM</sup> the properties of fairly rapid drying and conductivity. The main advantage of carbon suspensions is that with care they may be used with almost any specimen, no matter how fine. Carbon is a poor electron emitter and therefore does not ‘hide’ the specimen. Consideration must be given to any possible effects of the adhesive solvent on the sample.

For fine particles, a stub is cleaned by rubbing it on a clean piece of paper. A thin layer of carbon suspension is then wiped across the surface of the stub with one smooth movement. When the carbon dries and is almost totally matt, the specimen is deposited on to the surface. The excess is shaken off, leaving the material adhering to the stub (**Figure 13**). If one practices with sugar, the technique has failed if the sugar adheres but has turned black.

**Adhesive tapes and tabs** A very wide range of adhesive tapes is available in various forms: single-sided, double-sided, copper, aluminum, and now carbon. The main advantage with a tape or a tab is that very fine specimens may be mounted quickly. The main disadvantage, apart from outgassing of the tape in the microscope creating high levels of contamination, is that sticking the specimen on to a tape means a commitment to coating, unless the tape is in the form of carbon tape ‘tabs’, which seem to be an almost perfect mounting medium.

Another point worth noting is that particles on tapes may often become mechanically unstable in the





**Figure 13** (A) A typical specimen stub. (B) The initial stage of carbon suspension application. (C) The carbon covering the entire stub. (D) The particulate material adhering to the coated stub.

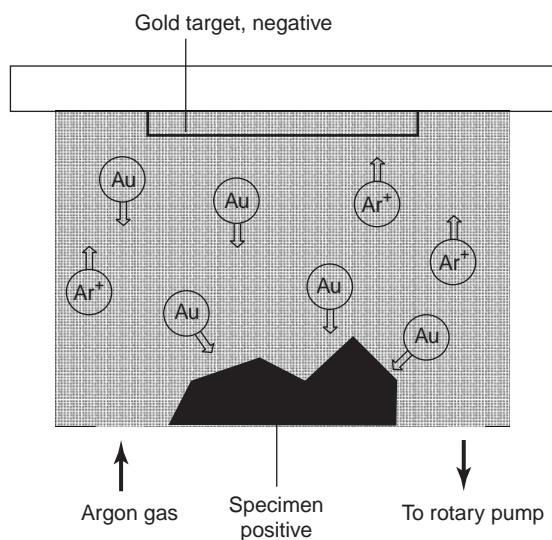
electron beam as they sit on small strands of material. Glass coverslips make an ideal clear background on which to sit a specimen that is in suspension. The coverslip is mounted on to a stub and is coated with a metal for  $\sim 2$  min before placing on to it a drop of suspension. The slip is then earthed by running a trace of silver dag from the top surface down to the specimen stub itself. If the sample is very irregular (or even very regular such as a large sphere), problems may arise in attaching the sample to the stub and in creating a conducting path, since very often there is only point contact with specimen 'overhang'. A somewhat larger quantity of adhesive will be required to ensure contact with the largest diameter of the specimen, and additional care is needed in forming the conducting path.

It is often useful to modify a standard stub to accommodate irregular samples, such as by drilling a small hole in the center or possibly adding a spike to impale the sample. Such modifications can increase the mechanical stability.

### Coating Techniques

Why is it necessary to coat specimens before inserting them into the SEM? In the first case, the specimen is coated because nonconducting specimens build up a surface charge, thus deflecting the major imaging signal, the SEs. Therefore, the image viewed may be distorted in both signal level and image form. In the second case, coating provides a surface layer that produces a higher SE yield than the normal specimen material.

The use of an SEM to observe as many specimens as possible in an uncoated form should be



**Figure 14** A schematic diagram demonstrating the principle of sputter coating.

encouraged, because only uncoated specimens are guaranteed to display the true surface structure. If a coating is required to make the surface more easy to handle in the microscope (less charge and distortion), the following sputter coating technique should be adopted.

In a low-vacuum environment (5–1 Pa; 0.05–0.01 mbar), a gold (gold–palladium or platinum) target is bombarded with argon ions. Erosion of the target takes place and the electrode arrangement causes the metal to deposit on all surfaces within the chamber, including the specimen (Figure 14). With a modern 'low-voltage' coater, metal may be deposited at up to  $50 \text{ nm min}^{-1}$ , which enables a sufficient coating to be put down in  $\sim 30$  s.

Sputtered metals are deposited in the form of islands, not a continuous coating. Although the coating will be relatively random, a coating unit should be tuned to obtain the optimum grain size, coating penetration, and coating thickness for the task in hand. Multiple coatings should be avoided as they produce small gold 'hills', adding features to the specimen surface. For use at the very low SEM magnifications, a coating that appears gold in color is desirable. This very thick coating will provide a high signal while protecting the specimen from the high beam currents used at these levels. When the specimen is to be investigated over a range of magnifications, it is important not to coat the surface with any more gold than will give a gray tint to the specimen ( $< 10 \text{ nm}$ ). This thickness is more than sufficient to obtain satisfactory results without forming an undesirable 'carpet' over the fine detail. For many electron microscopists, coating the specimen is a

standard procedure. It should be pointed out that specimen damage during sputter coating is one of the most common artifacts in scanning electron micrographs. There are very few specimens that, if correctly prepared, may not be viewed without coating, provided the SEM is operated at low accelerating voltages and with appropriate care.

If the sample contains water, this will have to be removed before examination by SEM. There are basically three methods: air drying, freeze drying, and critical-point drying (CPD). The one used will depend on the sample. Air drying is the quickest but the sample will distort and shrink unless it is protected by a rigid 'shell'. If the only interest is X-ray analysis, this may not be a problem. Freeze drying, i.e., rapid freezing and then removal of the ice, usually in a specialized vacuum chamber, is ideal for samples such as wood, but critical point drying is most commonly used. The object of this technique is to remove the water initially by solvent extraction (as in thin section), then after the 100% alcohol step to change to acetone, and finally in the CPD apparatus to substitute the acetone for liquid carbon dioxide. Conversion from the liquid phase to the gaseous phase is achieved at elevated temperature and pressure. The critical point is the change from liquid to gas 'without change of volume'. In reality, the volume change is minimized but it does occur, and this must be considered when finally viewing the sample.

It would be preferred that the specimen could be viewed without preparation, and for this reason cryostages and environmental chambers exist. As mentioned previously, it is possible to freeze-fracture tissue to reveal the internal structure, but it is

possible also to fracture the critical-point-dried sample, which will be very brittle. The resultant image is limited only by the fixation method.

Only the most commonly used techniques have been described, and it is inevitable that each specialized requirement will modify or create a new method.

**See also: Adhesives and Sealants. Microscopy Techniques: Specimen Preparation for Light; Scanning Electron Microscopy.**

## Further Reading

- Brawman JC and Sinclair R (1984) The preparation of cross-section specimens for transmission electron microscopy. *Journal of Electron Microscopy Techniques* 1: 53–61.
- Chapman SK (1986) *Working with a Scanning Electron Microscope*. Chislehurst, Kent: Lodgemark Press.
- Glauert AM.(ed.) (1974–1992). *Practical Methods in Electron Microscopy*. Amsterdam: Elsevier Science Publishers.
- Goodhew PJ (1984) *Specimen Preparation for Transmission Electron Microscopy of Materials*, Microscopy Handbooks no. 3. Oxford: Oxford University Press and Royal Microscopical Society.
- Jacobs EG, Forster IA, Wu Y, Wilson AR, and Pinizzotto RF (1993) Ultramicrotomy: A unique method for preparation of composite solder for transmission electron microscopy. *Journal of Material Research* 1: 1–20.
- Orloff J, Swanson L, and Utlaut MW (2002) *High Resolution Focused Ion Beams: Fib and its Applications: The Physics of Liquid Metal Ion Sources and Ion Optics and their Applications to Focused Ion Beam Technology*. New York: Plenum.

## Scanning Electron Microscopy

**M I Szynkowska**, Technical University of Lodz, Lodz, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by A F Otterloo, pp. 3151–3160, © 1995, Elsevier Ltd.

## Introduction

Scanning electron microscopy (SEM) is one of the most versatile and widely used of the surface analytical techniques as it allows both the morphology and composition of various materials in modern science to be studied. It is considered a relatively rapid,

inexpensive, and basically nondestructive approach to surface analysis. It covers a number of areas of application like biology, geology, metallurgy, semiconductor research, and catalysis. The first scanning electron microscopes-used to examine thick specimens was described by Zworykin and co-workers in 1942, but the first commercial scanning electron microscopes, based on the work of Oatley and co-workers at Cambridge University, became available only in 1965. Since then, many advances have been made owing to improvements in lens design, high-brightness electron sources, new detectors, and electronic signal processing. Nowadays there are many different types of scanning electron microscope

standard procedure. It should be pointed out that specimen damage during sputter coating is one of the most common artifacts in scanning electron micrographs. There are very few specimens that, if correctly prepared, may not be viewed without coating, provided the SEM is operated at low accelerating voltages and with appropriate care.

If the sample contains water, this will have to be removed before examination by SEM. There are basically three methods: air drying, freeze drying, and critical-point drying (CPD). The one used will depend on the sample. Air drying is the quickest but the sample will distort and shrink unless it is protected by a rigid 'shell'. If the only interest is X-ray analysis, this may not be a problem. Freeze drying, i.e., rapid freezing and then removal of the ice, usually in a specialized vacuum chamber, is ideal for samples such as wood, but critical point drying is most commonly used. The object of this technique is to remove the water initially by solvent extraction (as in thin section), then after the 100% alcohol step to change to acetone, and finally in the CPD apparatus to substitute the acetone for liquid carbon dioxide. Conversion from the liquid phase to the gaseous phase is achieved at elevated temperature and pressure. The critical point is the change from liquid to gas 'without change of volume'. In reality, the volume change is minimized but it does occur, and this must be considered when finally viewing the sample.

It would be preferred that the specimen could be viewed without preparation, and for this reason cryostages and environmental chambers exist. As mentioned previously, it is possible to freeze-fracture tissue to reveal the internal structure, but it is

possible also to fracture the critical-point-dried sample, which will be very brittle. The resultant image is limited only by the fixation method.

Only the most commonly used techniques have been described, and it is inevitable that each specialized requirement will modify or create a new method.

**See also: Adhesives and Sealants. Microscopy Techniques: Specimen Preparation for Light; Scanning Electron Microscopy.**

## Further Reading

- Brawman JC and Sinclair R (1984) The preparation of cross-section specimens for transmission electron microscopy. *Journal of Electron Microscopy Techniques* 1: 53–61.
- Chapman SK (1986) *Working with a Scanning Electron Microscope*. Chislehurst, Kent: Lodgemark Press.
- Glauert AM.(ed.) (1974–1992). *Practical Methods in Electron Microscopy*. Amsterdam: Elsevier Science Publishers.
- Goodhew PJ (1984) *Specimen Preparation for Transmission Electron Microscopy of Materials*, Microscopy Handbooks no. 3. Oxford: Oxford University Press and Royal Microscopical Society.
- Jacobs EG, Forster IA, Wu Y, Wilson AR, and Pinizzotto RF (1993) Ultramicrotomy: A unique method for preparation of composite solder for transmission electron microscopy. *Journal of Material Research* 1: 1–20.
- Orloff J, Swanson L, and Utlaut MW (2002) *High Resolution Focused Ion Beams: Fib and its Applications: The Physics of Liquid Metal Ion Sources and Ion Optics and their Applications to Focused Ion Beam Technology*. New York: Plenum.

## Scanning Electron Microscopy

**M I Szynkowska**, Technical University of Lodz, Lodz, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by A F Otterloo, pp. 3151–3160, © 1995, Elsevier Ltd.

## Introduction

Scanning electron microscopy (SEM) is one of the most versatile and widely used of the surface analytical techniques as it allows both the morphology and composition of various materials in modern science to be studied. It is considered a relatively rapid,

inexpensive, and basically nondestructive approach to surface analysis. It covers a number of areas of application like biology, geology, metallurgy, semiconductor research, and catalysis. The first scanning electron microscopes-used to examine thick specimens was described by Zworykin and co-workers in 1942, but the first commercial scanning electron microscopes, based on the work of Oatley and co-workers at Cambridge University, became available only in 1965. Since then, many advances have been made owing to improvements in lens design, high-brightness electron sources, new detectors, and electronic signal processing. Nowadays there are many different types of scanning electron microscope

designed for specific purposes ranging from routine morphological studies to high-speed compositional analyses or to the study of environment-sensitive materials (e.g., environmental scanning electron microscopes). The main advantages of SEM are the high lateral resolution (1–10 nm), large depth of focus (typically 100  $\mu\text{m}$  at  $\times 1000$  magnification), and the numerous types of electron–specimen interaction that can be used for imaging or chemical analyses purposes. In contrast to transmission electron microscopes (TEMs), specimens of largely varying sizes can be examined without elaborate specimen preparation.

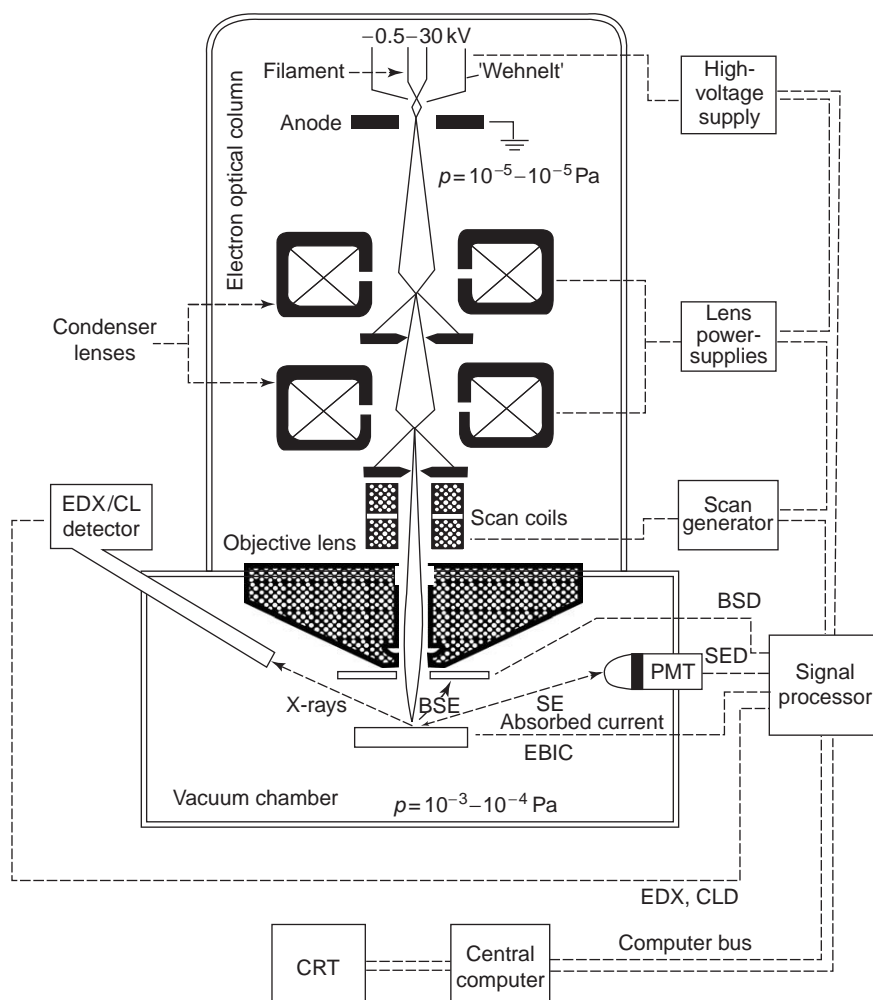
Owing to the rapid advances in computing power, SEM has become relatively easy. This enables development of microscopes to be focused on the analytical aspect of the problem. As a consequence, the latest generation of these instruments is increasingly being used for process and product control. This shift in the application of electron microscopes started in the semiconductor industry, where SEM is

used for automatic inspection of critical dimensions on semiconductor wafers. The user's task is to choose the optimum detector strategies and microscope settings for the problem to be analyzed. This article will discuss the different contrast formation mechanisms in SEM. The electron beam–specimen interactions are emphasized since these interactions form the basis of the signals generated.

## The Structure of the Scanning Electron Microscope

The scanning electron microscope consists of an electron–optical column mounted on a vacuum chamber (Figure 1).

An electron gun is placed on top of this column. This gun usually consists of a thermionic cathode made of tungsten or  $\text{LaB}_6$  in a triode configuration, i.e., an additional electrode (Wehnelt) between the filament and anode. The pressure in the specimen chamber is  $10^{-3}$ – $10^{-5}$  Pa. This pressure is much



**Figure 1** Schematic cross-section of a scanning electron microscope (see text for explanation of abbreviations).

lower than the saturation vapor pressure of water, requiring special preparation of water-containing biological samples. In modern instruments operating at low voltages and high magnifications, the thermionic sources are replaced by field-emission guns (FEGs), yielding a higher brightness and lower energy spread in the beam. Field-emission sources however require a better vacuum of  $10^{-7}$ – $10^{-8}$  Pa as opposed to  $10^{-4}$ – $10^{-6}$  Pa for thermionic emitters. The crossover formed in the electron gun is demagnified by a set of magnetic (or sometimes combined magnetic and electrostatic) lenses. The condenser lenses, often in combination with spray apertures, are used to change the beam divergence angle and thus the probe current. Changes in probe current in turn will result in changes in probe diameter. The objective lens is used to focus the electron beam into a fine spot on the sample surface. In state-of-the-art instruments the minimum beam diameter on the sample ranges from 1 to 10 nm in the energy range from 30 to 0.5 keV. A set of scan coils allows rastering of the focused beam across the surface. The signals generated in the specimen by the incident electron beam are monitored synchronously with the raster scan. The magnification and scan velocity are altered by changing the scan coil excitation.

Upon entering the sample, the electron beam interacts with the sample, and both electron and photon signals are generated (Figure 2).

These signals are collected by dedicated detectors, amplified, and displayed. The images can be stored on photographic material or in a computer memory (frame store). The latter option offers the possibility for further image processing and overlaying of various detector signals. In modern instruments, distributed processors, under the control of a personal

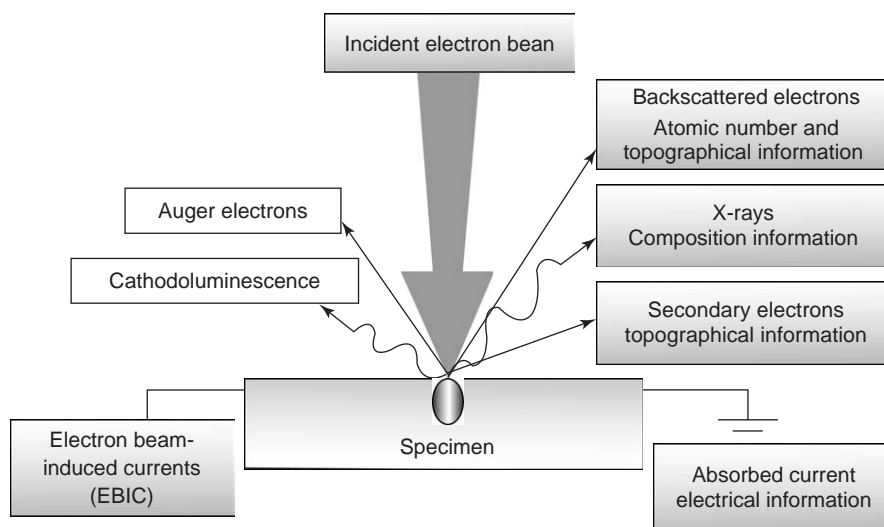
computer, operate the gun and lens power supplies, scan generator, and signal detectors.

## Electron–Optical Limits to Resolution

The lateral resolution obtainable in a scanning electron microscope is determined by two factors: (1) the incident beam diameter,  $d_p$ ; and (2) the generation volume of the signals used to obtain the image. The minimum usable beam size determined by the electron optics depends on the brightness of the gun and the electron–optical aberrations of the magnetic lenses used. The following aberrations are important in the formation of the electron beam: spherical aberration, aperture diffraction, chromatic aberration, aperture diffraction, and axial astigmatism.

### Spherical Aberration

If parallel rays are incident on an electron lens, those farther from the optical axis will be focused closer to the lens. The resulting diameter in the plane of least confusion is given by  $d_s = (1/2)C_s\alpha^3$ , where  $C_s$  is the spherical aberration coefficient, a lens parameter related to the focal length and the lens excitation, and  $\alpha$  is the aperture-limited semi-angle, which is typically  $10^{-2}$  rad and determined by a diaphragm in the polepiece gap of the objective lens. In conventional scanning electron microscopes, large working distances and weakly excited lenses are used. These lenses have large spherical aberration constants  $C_s \cong 10$ –20 mm. Small values of  $C_s \cong 1$ –2 mm can be attained by placing the specimen inside the lens field (immersion lens). These lenses are used in special high-resolution scanning electron microscopes and require in-lens detection of the secondary electrons



**Figure 2** The primary electron beam–specimen interaction in a scanning electron microscope.



since these will spiral around the magnetic field lines into the objective lens.

### Aperture Diffraction

The objective lens aperture diffracts the incident beam, enlarging the focal point into an Airy disk with half-width  $d_d = 0.6\lambda/\alpha$ . For this effect, the larger the value of  $\alpha$ , the smaller contribution of  $d_d$ . Thus, spherical aberration and aperture diffraction vary in opposite directions with  $\alpha$ . This leads to the need to find an optimum aperture angle,  $\alpha_{opt}$ , that is a balance between these two effects. For perfect lenses the diffraction error forms the physical limit to the minimum obtainable probe size. For field-emission microscopes this term becomes significant. The electron wavelength,  $\lambda$ , is given approximately by  $\lambda = 1.226/\sqrt{U}$  ( $U$  in V,  $\lambda$  in nm).

### Chromatic Aberration

The focal length of an electron lens depends on the accelerating voltage. Owing to the energy spread in the electron gun the result is a disk of least confusion with diameter  $d_c = C_c(\Delta E/E)\alpha$ , where  $C_c$  is the chromatic aberration coefficient and  $\Delta E/E$  is the fractional variation in the electron-beam energy. The value of  $C_c$  for weakly excited lenses is approximately equal to the focal length, typical 5–20 mm. Chromatic aberration, which is the largest aberration factor at low voltages ( $E < 5$  keV), can be minimized by the use of FEGs (small  $\Delta E$ ; 0.2–0.4 eV) and immersion lenses (small  $C_c$ ) and by reducing the convergence angle,  $\alpha$ . Also, very stable high-voltage and lens-current power supplies are necessary, a requirement that can be fulfilled with present-day electronic circuitry.

### Astigmatism

A final imperfection in lenses is astigmatism, which results from deviations from ideal rotational symmetry around the optical axis. Astigmatism can be caused by inhomogeneities in the polepieces, elliptical-shaped bores, charging effects in the bores, and aperture diaphragms. The effect is compensated by a stigmator, which consists of a pair of quadrupole lenses generally placed in the upper polepiece of the objective lens close to the scanning coils.

### The Minimal Probe Diameter

The geometric probe diameter,  $d_g$ , without considering lens aberrations, is limited by the apparent gun brightness,  $\beta$ . The brightness,  $\beta$ , in  $\text{A cm}^{-2} \text{sr}^{-1}$  is the most important performance indicator for electron guns. The geometrical probe diameter is related to the probe current,  $I_p$ , through  $d_g^2 = 4I_p/$

$(\pi^2\beta\alpha^2) = C_0\alpha^{-2}$ . For thermionic emitters the maximum brightness is given by  $\beta_{\max} = j_c E/(\pi k T_c)$ , where  $E$  is the accelerating (beam) energy in eV,  $k$  Boltzmann's constant ( $8.6 \times 10^{-5} \text{ eV K}^{-1}$ ),  $T_c$  the cathode temperature in K, and  $j_c$  the cathode current density in  $\text{A cm}^{-2}$ . According to this equation the brightness increases linearly with the accelerating voltage and inversely with the filament temperature. The maximum theoretical brightness of a field-emission gun is given by  $\beta_{\max} = j_c E/(\pi \Delta E)$ , where  $\Delta E$  is the energy spread of the beam. Typical values for the cathode parameters are compiled in Table 1.

Small values of  $d_g$  require small beam currents and therefore give poor signal-to-noise ratios in the images. The geometrical probe diameter,  $d_g$ , is broadened by the aforementioned lens aberrations of the objective lens, and the resulting probe diameter can be obtained from eqn [1]. For thermionic cathodes (tungsten and LaB<sub>6</sub>) the first and third terms in eqn [1] are the largest at  $E = 10$ –20 keV. For probe currents of 1–10 pA, used in high-resolution imaging, probe diameters of 5–10 nm can be attained using these guns:

$$d_p^2 = d_g^2 + d_d^2 + d_s^2 + d_c^2 \\ = [C_0^2 + (0.6\lambda)^2]\alpha^{-2} + \frac{1}{4}C_s^2\alpha^6 + \left(C_c \frac{\Delta E}{E}\right)^2 \alpha^2 \quad [1]$$

For low electron energies, however, the chromatic aberration,  $d_c$ , becomes the largest term owing to the relatively large energy spread,  $\Delta E$ . In low-voltage scanning electron microscopes, FEGs or Schottky-assisted (ZrO/W) cathodes are used. For these guns the second and last terms dominate the probe diameter, and at high voltages ( $E = 20$ –30 keV) the probe diameter becomes diffraction limited to 1–2 nm. The chromatic aberration term dominates at low voltages, limiting the probe size to 5–10 nm at  $E = 1$  kV. Chromatic aberration can be further reduced by

**Table 1** Properties of electron sources used in SEM

Parameter''	W	LaB <sub>6</sub>	Schottky ZrO/W	FEG
$T_c$ (K)	2500–3000	1400–2000	1800	300
$\phi$ (eV)	4.5	2.7	2.8	4.6
$j_c$ ( $\text{A cm}^{-2}$ )	1–3	20–50	500	$2 \times 10^5$
$\Delta E$ (eV)	1–3	1–2	0.5	0.2–0.4
$t$ (h)	40–100	200–1000	> 1000	1000–6000
$\beta$ ( $\text{A cm}^{-2} \text{sr}^{-1}$ )	$10^5$	$10^6$	$10^8$	$10^8$
$d_n$ (nm)				
at 20 kV		4–10		1–2
at 1 kV				5–10

$T_c$  = cathode temperature;  $\phi$  = work function;  $j_c$  = current density;  $t$  = lifetime;  $\beta$  = brightness.

placing the sample in the lens field or corrected by means of multipole correctors in combination with an electrostatic retarding field lens. Theoretical calculations and laboratory experiments show that probe diameters of 2–3 nm at 1 kV are feasible.

## Depth of Focus

The depth of focus,  $T$ , is determined by the electron probe aperture,  $\alpha$ , and the magnification,  $M$ . Assuming a resolution on the cathode-ray tube or micrograph of  $d_b = 0.1$  mm (e.g., 1000 resolved lines on a 10 cm display), the corresponding lateral resolution on the specimen,  $\delta$ , is given by  $\delta = d_b/M$ . A perfectly focused beam will be geometrically broadened to a radius  $\delta/2$  at a vertical distance  $L$  given by  $L = \frac{1}{2}\delta/\tan \alpha \approx \delta/2\beta$ . The depth of field  $T$  is the vertical distance from  $-L$  to  $L$ . Thus details at a depth  $T = \delta/\alpha = d_b/\alpha M$  will appear sharp. By changing the working distance and final lens diaphragm, the electron probe aperture,  $\alpha$ , can be varied. For a typical value of  $\alpha = 10$  mrad the depth of focus varies from 1 mm at  $\times 10$  magnification to  $0.1 \mu\text{m}$  at  $\times 100\,000$  magnification. The values for the depth of focus obtainable in a scanning electron microscope are a factor of 100–1000 larger than in a light-optical microscope. Owing to its large depth of focus, the scanning electron microscope is often preferred to light microscopes at low magnifications. This is particularly true when irregularly shaped specimens with large height differences have to be observed.

## Electron Beam–Solid Interactions

When a beam electron enters a solid, it is subject to electrostatic interactions with the nuclei and electrons of the target atoms. The angle of the incident electrons is changed in the solid owing to elastic scattering by the shielded target nuclei, and the electrons lose energy owing to inelastic scattering from the core and valence electrons of the target atoms. Owing to elastic scattering events in the solid, a fraction of the incident electrons is scattered back into the vacuum. These electrons are called backscattered electrons (BSEs), and the backscatter yield,  $\eta$ , depends on the target's atomic number,  $Z$ . By convention the term BSE applies to all electrons that escape into the vacuum with energy  $E > 50$  eV.

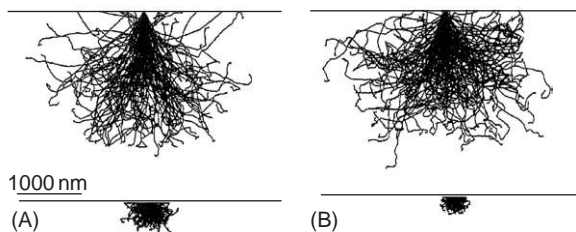
In inelastic scattering events a primary electron ionizes target atoms. The ionized atoms fall back into a lower energy state with the emission of Auger electrons or X-rays. The energy of the X-rays is characteristic of the atom involved and is used in most scanning electron microscopes for elemental

identification. Measurements of the energies or wavelengths of these X-rays gives information about the chemical composition of the specimen. Depending on whether an energy-dispersive or wavelength-dispersive X-ray analyzer is used, the analysis technique is called EDX or WDX spectrometry. When valence electrons are involved in the inelastic scattering process, the radiation is in the visible or ultraviolet (UV) region. The technique that uses this generated light for imaging purposes in a scanning electron microscope is called cathodoluminescence (CL). The excited valence electrons created in the inelastic scattering processes are called true secondary electrons (SEs). All electrons that leave the surface with energy  $E < 50$  eV are called SEs by convention. Owing to the small mean escape depth,  $\lambda$ , of an SE ( $\lambda = 2$ – $10$  nm), only those electrons that are generated close to the surface can escape into the vacuum and be detected. This offers the possibility of obtaining high-lateral resolution SE images.

## Monte Carlo Calculations

In the Monte Carlo electron trajectory simulation, the effects of elastic and inelastic scattering are calculated from appropriate models to determine reasonable scattering angles, distances between scattering sites, and the energy loss rate. It offers the possibility of calculating the influence of the electron scattering on the signals observed in the scanning electron microscope. The calculated values for the range and resolution presented in this article are based on Monte Carlo calculation, using Mott cross-sections for the elastic scattering and appropriate inelastic cross-sections for all relevant interactions.

In **Figure 3** electron trajectories are shown for four targets of different composition at  $E = 25$  keV. The average of many electron trajectories gives an estimate of the interaction volume involved. For the heavier materials the scattering volume is approximately half a sphere with a radius equal to the



**Figure 3** Monte Carlo electron trajectory simulation for four different materials. Beam energy, 25 keV; beam diameter, 1 nm; tilt,  $0^\circ$ ; 150 trajectories. Targets (density ( $\text{g cm}^{-3}$ )): (A) carbon (2.26); (B) silicon (2.33); (C) silver (10.50); and (D) gold (19.28).

electron range. From the graphs it is also clear that in heavy materials many large-angle deflections occur, resulting in a backscatter yield,  $\eta$ , that increases with atomic number.

## Electron Range

The generation volume of BSEs and X-rays and the fraction of SEs generated by BSEs are determined by the electron range (distance traveled by the beam electrons within the solid). Because of the complex nature of the interaction volume, a number of different definitions of the electron range exist in the literature: experimental range, maximum range, Bethe range, Kanaya–Okayama range. For energies between 10 and 30 keV, as used in conventional scanning electron microscopes, the empirical range formula of Kanaya and Okayama (eqn [2]) is often used.

$$R_{K-O} = \frac{0.0276AE_0^{1.67}}{Z^{0.89}\rho} (\mu\text{m}) \quad [2]$$

where  $A$  is the relative atomic mass ( $\text{g mol}^{-1}$ ),  $E_0$  is the beam energy (keV),  $Z$  is the atomic number, and  $\rho$  is the density ( $\text{g cm}^{-3}$ ). At lower energies ( $E = 1\text{--}10$  keV) the exponent,  $n$ , of  $E_0$  is somewhat lower,  $n = 1.3\text{--}1.38$ . In this energy range the electron range expressed in mass thickness ( $\text{g cm}^{-2}$ ) is independent of  $Z$ . The electron ranges for some selected materials are shown in Table 2.

## Contrast

Contrast is defined according to eqn [3]

$$C = \frac{S_2 - S_1}{S_2}, \quad S_2 > S_1 \quad [3]$$

where  $S_1$  and  $S_2$  are the signals detected at any two arbitrarily chosen points in the scan raster that defines the image field.  $C$  is always positive, in the range between 0 and 1. The observed contrast in most SEM images is largely determined by the electron scattering and detector characteristics. Emphasis will be placed on the most often used signals from SEs and BSEs. In discussions of contrast formation,

the total SE signal is usually divided into four separate contributions (SE<sub>1</sub>–SE<sub>4</sub>).

- SE<sub>1</sub>: Generated close to the beam impact point, these electrons carry high-lateral resolution information of the order of the mean escape depth,  $\lambda$ . The obtainable lateral resolution is typically 1–5 nm.
- SE<sub>2</sub>: Generated by (B)SEs that generate SEs while escaping from the surface area. The lateral resolution due to SE<sub>2</sub> is of the order of the backscattered electron range. Typical values of the SE<sub>2</sub> range, for medium and high atomic number materials at 10–30 keV beam voltages, vary between 0.1 and 1  $\mu\text{m}$ .
- SE<sub>3</sub>: Generated by BSEs that hit the final lens polepiece and specimen chamber walls.
- SE<sub>4</sub>: Electrons created in the optical column owing to scattering from the walls and transported down the column. Good design of the column, using well-placed spray apertures, generally suppresses this contribution.

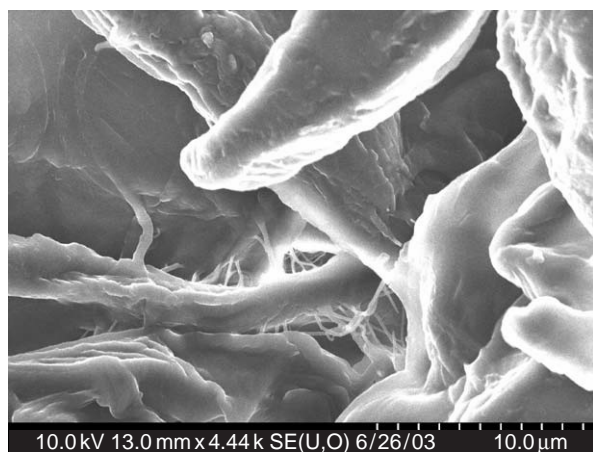
## SE Topographic Contrast

The topographic contrast observed in most SE micrographs gives these images a crisp 3D appearance (Figure 4). The observed contrast is due to a number of effects.

1. Tilt contrast. The SE yield,  $\delta$ , depends on the local tilt angle,  $\phi$ , of the surface. For voltages  $> 5$  keV the dependence of  $\delta$  on local tilt angle,  $\phi$ , is given approximately by  $\delta(\phi) = \delta_0/\cos(\phi)$ , where  $\delta_0$  is the SE coefficient at  $0^\circ$  tilt. At low voltages, when the electron range,  $R$ , becomes comparable with the mean escape depth,  $\lambda$ , of the SEs, this surface tilt contrast is greatly reduced.
2. Shadow contrast. This is the shadowing of the SE signal due to holes or protrusions. The amount of shadowing depends on the presence of electrostatic and magnetic extraction fields on the surface. Shadowing is less for detection through the lens as compared with that of the standard Everhart–Thornley detector.

**Table 2** Electron range ( $\mu\text{m}$ ) as a function of energy

$E$ (keV)	Polymer ( $C_nH_{2n+2}$ ) $\rho = 1.0 \text{ g cm}^{-3}$	$\text{SiO}_2$ $\rho = 2.20 \text{ g cm}^{-3}$	$\text{Cu}$ $\rho = 8.93 \text{ g cm}^{-3}$	$\text{Au}$ $\rho = 19.28 \text{ g cm}^{-3}$
1	0.043	0.034	0.013	0.010
3	0.28	0.18	0.056	0.041
5	0.71	0.38	0.13	0.078
10	2.2	1.3	0.39	0.20
30	15.7	8.9	2.5	1.4



**Figure 4** SE micrograph of the surface structure of cellulose.

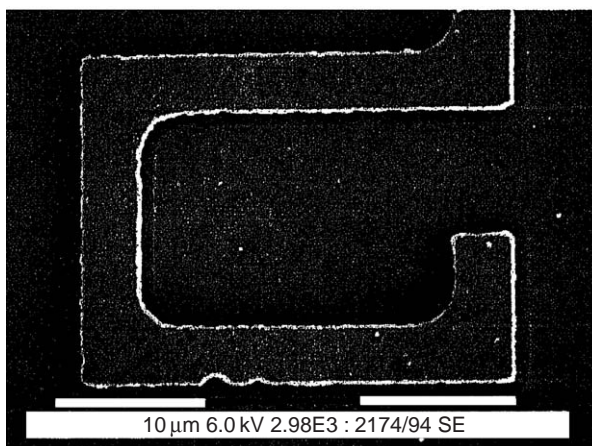
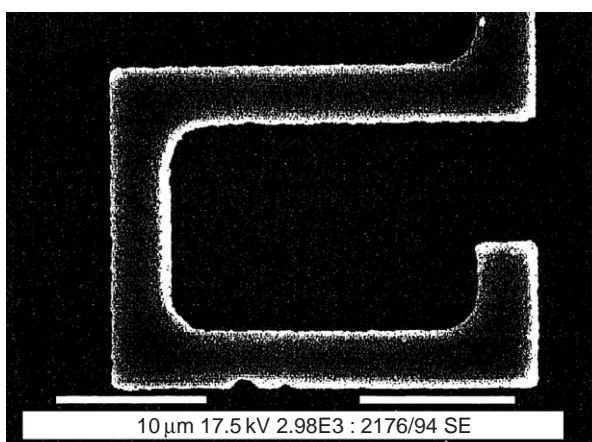
3. Edge effect contrast. This is due to SEs that penetrate through the edges of structures. These SEs give a large additional contribution to the SE signal. This contrast is observed in most micrographs as white edges extending over a distance of the order of the electron range. Therefore, the width and magnitude of the observed edge contrast depend on the beam energy and are greatly reduced at low voltages. The edge effect gives a large SE background on which small signal fluctuations on stepped surfaces can become invisible. This effect is most disturbing when the spatial variation of the background due to the edge effect is comparable with the size of the surface structures to be examined. At high voltages this situation occurs at medium magnifications  $M = 1000\text{--}50\,000$ . Low voltages are therefore preferred when observing cross-sections or details on small topographic structures, e.g., patterns on integrated circuits (ICs). **Figure 5** shows how the edge effect observed on an IC line pattern is reduced at low voltages.

### BSE Topographic Contrast

The BSE signal also carries topographical information. The tilt dependence of the backscatter yield,  $\eta$  (tilt contrast), is less pronounced and is given empirically by

$$\eta(Z, \phi) = (1 + \cos \phi)^{-9/\sqrt{Z}} \quad [4]$$

The shadow contrast in BSE images, however, is very pronounced because the fast BSEs are not influenced by electromagnetic stray fields. They have to move in a straight line to the detector in order to be detected. To collect enough electrons, the backscatter detector has to span a large solid angle. Therefore, backscatter detectors are usually mounted under the objective lens polepiece. The backscatter signals also show some edge effects due to scattered electrons



**Figure 5** SE image of a line pattern in an IC, showing a reduced edge effect at low voltages.

that leave the edge of structures or are captured under surface steps.

The topographic contrast in BSE images can be enhanced by the use of split backscatter detectors. Subtracting the signals collected by the right and left parts of a split detector enhances the signal due to shadow or local surface tilt effects. In contrast to the situation for SEs, the BSE tilt contrast does not disappear at low voltages. The maximum obtainable resolution in topographic BSE images is determined by the range of BSEs. This range is approximately one-third to one-half of the electron range. By selection of only the nearly elastically reflected electrons, the contrast is enhanced and the lateral resolution is increased. The use of solid-state (p-n junction) or scintillator detectors enhances the high-energy fraction in the BSE signal. At low energies the penetration depth of the primary beam is of the same order as the mean escape depth of the secondary electrons. Under this condition, the lateral resolutions that can be obtained with BSEs and SEs become comparable. The use of BSEs for high-lateral resolution imaging requires sensitive detection at low voltages.

### BSE Material Contrast (Z contrast)

BSEs carry atomic number or material contrast. This is due to the fact that large-angle elastic scattering events are more probable for the heavy nuclei. This results in a strong dependence of  $\eta$  on the atomic number,  $Z$ . According to August, the backscatter yield,  $\eta$ , in the energy range  $E = 5\text{--}30\text{ keV}$  may be expressed as a power series in  $Z$  by the empirical formula:

$$\eta(E, Z) = (0.1904 - 0.2236 \ln(Z) + 0.1292 [\ln(Z)]^2 - 0.01491 [\ln(Z)]^3)(0.0002167Z + 0.987) \times E^{(0.1382 - [0.9211Z^{-1/2}])} \quad [5]$$

At high energies ( $E > 5\text{ keV}$ ) the backscatter coefficient,  $\eta$ , hardly varies with  $E$  and shows a monotonic increase with  $Z$ . At low energies ( $E < 1\text{ keV}$ )  $Z$  contrast reversal occurs.

This is a consequence of the strongly decreased elastic scattering cross-section for the high-atomic number materials at low energies. As shown by Reimer, these effects can only be predicted by extensive quantum mechanical scattering cross-section calculations based on the partial wave expansion method. For a compound material the backscatter coefficient can be obtained by weighting the elemental backscatter coefficients with their mass fractions,  $c_i$ , according to eqn [6]:

$$\eta_{\text{compound}} = \sum_i c_i \eta_i \quad [6]$$

Atomic number ( $Z$ ) contrast is very useful in the analysis of foreign materials and catalysts, contamination detection, and metallurgical phase analysis. An example of atomic number contrast is shown in Figure 6.

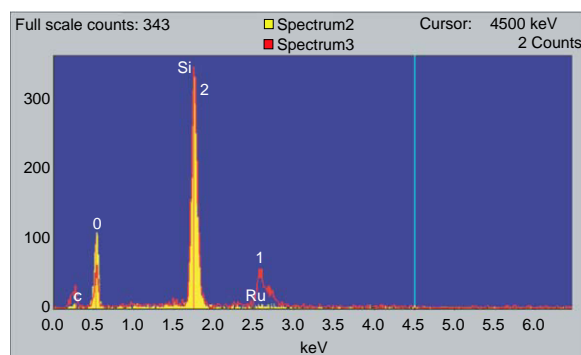
This micrograph shows the surface of a Ru/SiO<sub>2</sub> catalyst. The heavy ruthenium clusters appear bright, whereas the lighter silica spheres are darker in the image.

### SE Material Contrast

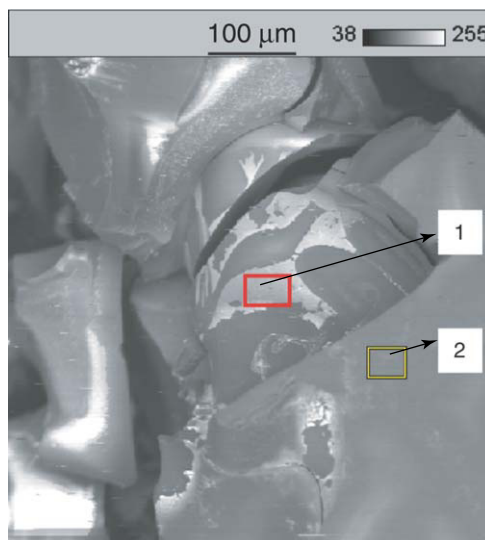
The SE yield,  $\delta$ , does not show a distinct atomic number dependence. Some material contrast can be observed in SE images owing to BSEs that generate SEs when leaving the specimen surface (SE<sub>2</sub>). There is also a contribution from BSEs that hit the polepiece and specimen chamber walls (SE<sub>3</sub>).

### Channeling Contrast

Through differences in crystallite orientation, grains in polycrystalline material appear with different contrast (channeling contrast). This contrast is very sensitive to small specimen tilts, and small electron probe apertures ( $\alpha = 1\text{--}10\text{ mrad}$ ) should be used. In



(A)



(B)

**Figure 6** (A) BSE micrograph of the surface of a Ru/SiO<sub>2</sub> catalyst. The backscatter image shows mainly atomic number contrast. The Ru clusters appear bright (point 1), while the lighter silica spheres are darker (point 2); (B) EDS spectrum from points 1 and 2.

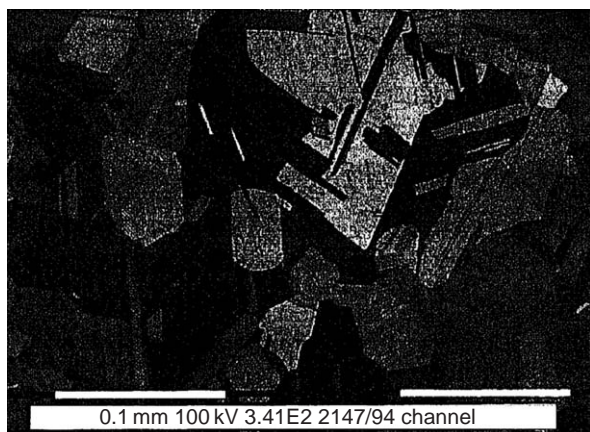
contrast to topographic and atomic number contrast, a different grain orientation can result in different channeling contrast, as shown in Figure 7 for a stainless steel sample reduced at 1050°C in hydrogen.

Channeling contrast is very sensitive to amorphous surface contamination or surface oxide layers. The contrast is most pronounced on clean (chemically etched or reduced) samples at low voltages ( $E < 10\text{ keV}$ ). Energy filtering of the BSEs further enhances this contrast. Although the channeling contrast is also present in the SE signal, it is best observed in the backscatter image.

### Cathodoluminescence

Many insulating and semiconducting materials emit electromagnetic radiation in the UV, visible, or



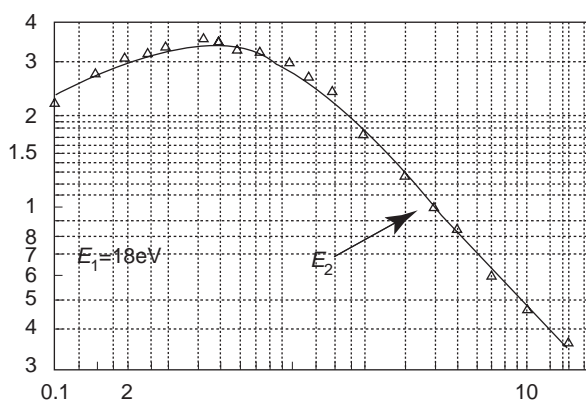


**Figure 7** BSE image showing strong channeling contrast in a stainless steel sample.

infrared wavelengths when excited by electrons. The incident electron beam generates electrons and holes in the luminescing material. These charge carriers diffuse through the material with a characteristic diffusion length  $L$ , before they recombine with or without the emission of light. The emitted light can be collected with a parabolic or elliptical mirror situated under the polepiece and coupled to a sensitive photomultiplier. Optionally the light can be led through a monochromator or a set of color filters for wavelength selection. The image has to be acquired under slow scan conditions owing to the relatively long decay times of most luminescing materials. The luminescence observed in most materials is faint, and relatively large beam currents have to be used. Owing to the sensitivity of the luminescence to electron-beam damage, low-dose techniques should be used on organic materials. The introduction of low-voltage scanning electron microscopes with digital image collection systems and sensitive CL detectors has led to a renewed interest in the CL technique for analysis of oxides, semiconducting materials, certain minerals, and biological specimens (Figure 8).

## Absorbed Current

The absorbed current, specimen current, or target current is equal to the probe current minus the total electron emission from the sample:  $I_{\text{specimen}} = I_p(1 - \eta - \delta) = I_p(1 - \sigma)$ , where  $\sigma$  is the total electron yield, which determines the charging behavior of insulating samples. Since the absorbed current signal depends on both the SE and BSE yield, the observed contrast will be a mix of these two contributions. For most samples at high voltages,  $\eta$  is the dominant factor, and the absorbed current images will resemble the inverted backscatter images.



**Figure 8** Dependence of the total electron yield  $\sigma = \eta + \delta$  on beam energy for  $\text{SiO}_2$  at normal incidence. Positive surface charging occurs for  $\sigma > 1$  and negative charging for  $\sigma < 1$ .

Absorbed current detection is often used to detect corrosion products or different phases in metallurgic samples at relatively low magnifications.

## Electron Beam-Induced Currents

Electron beam-induced currents (EBICs) can be observed in semiconductors, p-n junctions, and Schottky barriers. An incident electron beam generates electron-hole pairs. Within or near the depletion region these charge carriers are separated by the internal field in the diode, creating an increased reverse bias current. Electron-hole pairs created far from the junction recombine according to the lifetime of the material. The average generation energy for an electron-hole pair is a few electronvolts; thus, several thousand electron-hole pairs per primary electron are generated within the scattering volume. A low-impedance current amplifier amplifies the current flow through the p-n junction. The lateral resolution is determined by the scattering volume and the minority carrier diffusion length and is of the order of micrometers. Reducing the high voltage increases the lateral resolution. All crystal defects that influence the carrier recombination are visible in the EBIC current. Accordingly the EBIC signal can be used to study diffusion lengths, surface recombination, and p-n junctions; p-n junctions can be studied parallel to the surface, in which case some depth information can be obtained by changing the beam voltage and thus the electron range.

## Conductive Coatings

Conductive coatings are applied to prevent charging of insulator specimens. For organic samples the metal film plays an additional role in the stabilization of the structure and increases the heat conduction. For

low-resolution work up to  $\times 10\,000$  magnification, sputtered gold and gold/palladium films are used.

Gold cannot be used for EDX analysis because the Au L and M peaks may interfere with the elements to be analyzed. The second disadvantage of gold coatings in X-ray analysis is the attenuation of the low-energy X-ray lines due to absorption in the gold layer. Therefore, for EDX analyses evaporated carbon films are preferred. At high magnifications (above 10 000) the island structure of gold films becomes visible. Thin continuous chromium, tungsten, or platinum films, sputter-deposited under high-vacuum conditions, are used for high-resolution imaging with field-emission scanning electron microscopes.

**See also: Microscopy: Overview. Microscopy Applications: Biomedical; Proteins and Nucleic Acids. Microscopy Techniques: Electron Microscopy; Specimen Preparation for Electron Microscopy.**

## Further Reading

- Amelinckx S (1997) *Electron Microscopy*. New York: Wiley.
- Amelinckx S, van Dyck D, van Landuyt J, and van Tendeloo G (1997) *Handbook of Microscopy*. Weinheim: VCH.

- August HJ and Wemisch J (1990) Theoretical prediction of the electron backscattering coefficient for multilayer structures. *Journal of Microscopy* 157: 247–254.
- Feldman LC and Mayer JW (1986) *Fundamentals of Surface and Thin Film Analysis*. New York: Elsevier.
- Goldstein JI, Newbury DE, Echlin P, et al. (1992) *Scanning Electron Microscopy and X-ray Microanalysis*. New York: Plenum Press.
- Joy DC (1984) Beam interactions, contrast and resolution in the SEM. *Journal of Microscopy* 136: 241–258.
- Joy DC (1985) Resolution in low voltage SEM. *Journal of Microscopy* 140: 283–292.
- Lyman CE, Newbury DE, Goldstein JI, et al. (1990) *Scanning electron microscopy, X-Ray microanalysis and analytical electron microscopy. A Laboratory Workbook*. New York: Plenum.
- Newbury DE, Joy DC, Echlin P, Fiori CE, and Goldstein JI (1986) *Advanced Scanning Electron Microscopy and X-Ray Microanalysis*. New York: Plenum.
- Niemantsverdriet JW (2000) *Spectroscopy in Catalysis*. Weinheim: Wiley-VCH.
- Reimer L (1985) *Scanning Electron Microscopy: Physics of Image Formation and Microanalysis*. Berlin: Springer.
- Reimer L (1993) *Image Formation in Low Voltage Scanning Electron Microscopy*, vol. TT12. Bellingham, Washington DC: SPIE Optical Engineering Press.
- Seller H (1983) Secondary electron emission in the SEM. *Journal of Applied Physics* 54: R1–R18.

# Atomic Force and Scanning Tunneling Microscopy

**L A Bottomley, E D Gadsby, and M A Poggi**, Georgia Institute of Technology, Atlanta, GA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Scanning probe microscopy (SPM) is a family of related techniques that provide information about atomic scale structure and processes. The first of these to be developed was the scanning tunneling microscope (STM) by Gerd Binnig and Heinrich Rohrer at IBM Zurich in 1981. They were awarded the Nobel Prize in Physics in 1986 for their invention. Atomic force microscopy is another SPM technique invented by Gerd Binnig, Calvin Quate, and Christopher Gerber in 1986 to expand the types of surfaces on which atomic scale information can be obtained. Digital Instruments (Santa Barbara, California, USA) produced the first commercial atomic force microscope (AFM) in 1989.

These powerful techniques and others have been rapidly adopted by the scientific community and

applied in numerous fields ranging from biology to materials science. The number of publications and patents has increased exponentially over the last 20 years (**Figure 1**). SPM accounted for 8% of the microscopy market in 2002 worth US\$ 78 million and was the most rapidly growing segment.

The largest producer of SPM instruments is Veeco Instruments, Inc. (Santa Barbara, California, USA), which has acquired several AFM producers including Digital Instruments, Topometrix, and Park Scientific Instruments. Asylum Research, Pacific Nanotechnology, JEOL, Omicron NanoTechnology, and Nanonics Imaging are other major manufacturers (see **Table 1**).

## Scanning Tunneling Microscopy

The mode of operation of the STM is straightforward. An extremely sharp tip (ideally atomically sharp) is placed into close proximity with a conductive substrate. A voltage is applied between the probe tip and substrate. The gap between the two is decreased until electrons are passed from one to the

low-resolution work up to  $\times 10\,000$  magnification, sputtered gold and gold/palladium films are used.

Gold cannot be used for EDX analysis because the Au L and M peaks may interfere with the elements to be analyzed. The second disadvantage of gold coatings in X-ray analysis is the attenuation of the low-energy X-ray lines due to absorption in the gold layer. Therefore, for EDX analyses evaporated carbon films are preferred. At high magnifications (above 10 000) the island structure of gold films becomes visible. Thin continuous chromium, tungsten, or platinum films, sputter-deposited under high-vacuum conditions, are used for high-resolution imaging with field-emission scanning electron microscopes.

**See also: Microscopy: Overview. Microscopy Applications: Biomedical; Proteins and Nucleic Acids. Microscopy Techniques: Electron Microscopy; Specimen Preparation for Electron Microscopy.**

### Further Reading

- Amelinckx S (1997) *Electron Microscopy*. New York: Wiley.
- Amelinckx S, van Dyck D, van Landuyt J, and van Tendeloo G (1997) *Handbook of Microscopy*. Weinheim: VCH.

- August HJ and Wemisch J (1990) Theoretical prediction of the electron backscattering coefficient for multilayer structures. *Journal of Microscopy* 157: 247–254.
- Feldman LC and Mayer JW (1986) *Fundamentals of Surface and Thin Film Analysis*. New York: Elsevier.
- Goldstein JI, Newbury DE, Echlin P, et al. (1992) *Scanning Electron Microscopy and X-ray Microanalysis*. New York: Plenum Press.
- Joy DC (1984) Beam interactions, contrast and resolution in the SEM. *Journal of Microscopy* 136: 241–258.
- Joy DC (1985) Resolution in low voltage SEM. *Journal of Microscopy* 140: 283–292.
- Lyman CE, Newbury DE, Goldstein JI, et al. (1990) *Scanning electron microscopy, X-Ray microanalysis and analytical electron microscopy. A Laboratory Workbook*. New York: Plenum.
- Newbury DE, Joy DC, Echlin P, Fiori CE, and Goldstein JI (1986) *Advanced Scanning Electron Microscopy and X-Ray Microanalysis*. New York: Plenum.
- Niemantsverdriet JW (2000) *Spectroscopy in Catalysis*. Weinheim: Wiley-VCH.
- Reimer L (1985) *Scanning Electron Microscopy: Physics of Image Formation and Microanalysis*. Berlin: Springer.
- Reimer L (1993) *Image Formation in Low Voltage Scanning Electron Microscopy*, vol. TT12. Bellingham, Washington DC: SPIE Optical Engineering Press.
- Seller H (1983) Secondary electron emission in the SEM. *Journal of Applied Physics* 54: R1–R18.

## Atomic Force and Scanning Tunneling Microscopy

**L A Bottomley, E D Gadsby, and M A Poggi**, Georgia Institute of Technology, Atlanta, GA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Scanning probe microscopy (SPM) is a family of related techniques that provide information about atomic scale structure and processes. The first of these to be developed was the scanning tunneling microscope (STM) by Gerd Binnig and Heinrich Rohrer at IBM Zurich in 1981. They were awarded the Nobel Prize in Physics in 1986 for their invention. Atomic force microscopy is another SPM technique invented by Gerd Binnig, Calvin Quate, and Christopher Gerber in 1986 to expand the types of surfaces on which atomic scale information can be obtained. Digital Instruments (Santa Barbara, California, USA) produced the first commercial atomic force microscope (AFM) in 1989.

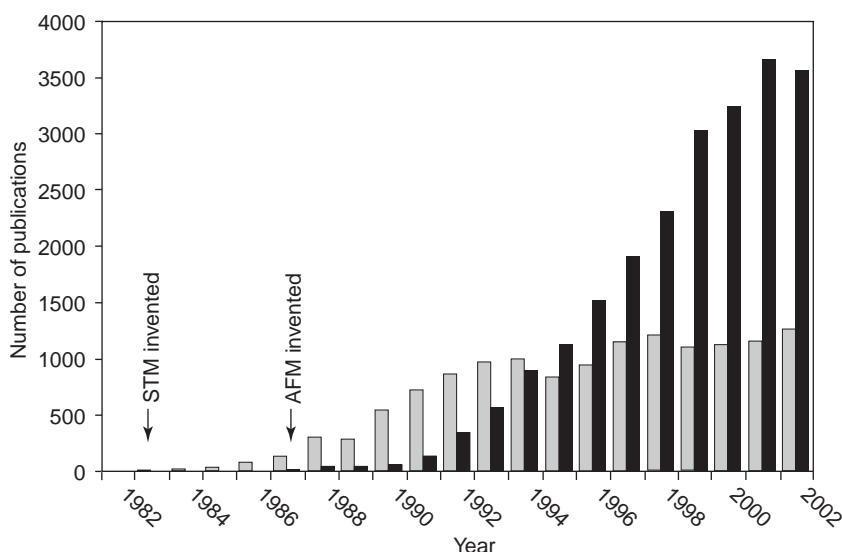
These powerful techniques and others have been rapidly adopted by the scientific community and

applied in numerous fields ranging from biology to materials science. The number of publications and patents has increased exponentially over the last 20 years (**Figure 1**). SPM accounted for 8% of the microscopy market in 2002 worth US\$ 78 million and was the most rapidly growing segment.

The largest producer of SPM instruments is Veeco Instruments, Inc. (Santa Barbara, California, USA), which has acquired several AFM producers including Digital Instruments, Topometrix, and Park Scientific Instruments. Asylum Research, Pacific Nanotechnology, JEOL, Omicron NanoTechnology, and Nanonics Imaging are other major manufacturers (see **Table 1**).

### Scanning Tunneling Microscopy

The mode of operation of the STM is straightforward. An extremely sharp tip (ideally atomically sharp) is placed into close proximity with a conductive substrate. A voltage is applied between the probe tip and substrate. The gap between the two is decreased until electrons are passed from one to the



**Figure 1** Number of publications of journal articles, books, and patents in the fields of scanning tunneling microscopy (gray bars) and atomic force microscopy (black bars) published per year.

**Table 1** Atomic force and scanning tunneling microscope instrument and cantilever manufacturers

Manufacturer name	Headquarters location	Website
Accurion	Menlo Park, California, USA	<a href="http://www.accurion.com">www.accurion.com</a>
Asylum Research Corporation	Santa Barbara, California, USA	<a href="http://www.AsylumResearch.com">www.AsylumResearch.com</a>
Independent Research Engineering	Moscow, Russia	<a href="http://www.mtu-net.ru/nanoscan">www.mtu-net.ru/nanoscan</a>
JEOL Europe Ltd./JEOL USA, Inc.	Japan	<a href="http://www.jeol.com">www.jeol.com</a>
JPK Instruments AG	Berlin, Germany	<a href="http://www.jpk.com">www.jpk.com</a>
Klocke Nanotechnik	Aachen, Germany	<a href="http://www.nanomoter.de">www.nanomoter.de</a>
Molecular Imaging	Tempe, Arizona, USA	<a href="http://www.molec.com">www.molec.com</a>
MikroMasch	Tallinn, Estonia	<a href="http://www.spmtips.com">www.spmtips.com</a>
Nanofactory Instruments	Goteborg, Sweden	<a href="http://www.nanofactory.com">www.nanofactory.com</a>
Nanonics Imaging Ltd.	Jerusalem, Israel	<a href="http://www.nanonics.co.il">www.nanonics.co.il</a>
Nanosurf AG	Basel, Switzerland	<a href="http://www.nanosurf.com">www.nanosurf.com</a>
Novascan Technologies	Ames, Iowa, USA	<a href="http://www.novascan.com">www.novascan.com</a>
NT-MDT Co.	Moscow, Russia	<a href="http://www.ntmdt.ru">www.ntmdt.ru</a>
Omicron NanoTechnology GmbH	Taunusstein, Germany	<a href="http://www.omicron-instruments.com">www.omicron-instruments.com</a>
Pacific Nanotechnology, Inc.	Santa Clara, California, USA	<a href="http://www.pacificnanotech.com">www.pacificnanotech.com</a>
PSIA	Sungnam, Korea	<a href="http://www.advancedspm.com">www.advancedspm.com</a>
Quesant Instruments Corp.	Agoura Hills, California, USA	<a href="http://www.quesant.com">www.quesant.com</a>
RHK Technology, Inc.	Troy, Michigan, USA	<a href="http://www.rhk-tech.com">www.rhk-tech.com</a>
Surface Imaging Systems GmbH	Herzogenrath, Germany	<a href="http://www.sis-gmbh.com">www.sis-gmbh.com</a>
Triple-O Microscopy GmbH	Potsdam, Germany	<a href="http://www.triple-o.de">www.triple-o.de</a>
Veeco Instruments, Inc.	Woodbury, New York, USA	<a href="http://www.veeco.com">www.veeco.com</a>
WITec	Ulm, Germany; Savoy, Illinois, USA	<a href="http://www.witec.de">www.witec.de</a> ; <a href="http://www.WITec-Instruments.com">www.WITec-Instruments.com</a>

Data based on search in SciFinder Scholar 2002; American Chemical Society.

other. Current magnitude is exponentially dependent on the distance between the tip and the sample whereas the direction of current flow depends upon the orientation of the applied electric field. Once the tunneling of electrons occurs, the tip is then scanned across the sample and current levels are recorded as a function of location. A topographical map of the surface results. Tunneling energies are significantly

lower than the electron energies commonly found in scanning electron microscopy. This reduces topological artifacts caused from high-energy electrons interacting with the substrate. Scanning tunneling microscopy can be performed under ambient, liquid, or vacuum environments. Image contrast is direct; no stains or contrast agents are required. However, the sample must be conductive to be imaged.

## Instrument Components

The essential components of an STM are shown schematically in (Figure 2). A computer/user interface permits the user to control parameters such as the scan size (the range that the probe tip will travel across the sample), scan rate of the tip across the surface, and feedback control to maintain a constant height or current between the tip and substrate. The commands from the user interface are then sent to a control module that converts these commands (such as a metric scan size) into a series of voltages to be sent to the microscope. The microscope contains the sharpened tip and a scanner that moves the tip or sample in  $x$ ,  $y$ , and  $z$  directions. A sample-mounting block holds the sample rigidly in place. Rigidity is important because all SPM techniques require minimization of all sources of electrical and mechanical noise. Typically, electrical wires in the microscope are fully shielded and all mechanical components are held tightly in place. The entire microscope is mounted on an antivibration apparatus. These can range from simple homemade antivibration units such as a concrete block hanging from bungee-cords (~US\$ 100) to antivibration tables that have pneumatically suspended tabletops (>US\$ 2500). If these types of noise are not controlled or reduced, visualizing atomic-scale features is impossible.

**Piezoelectric scanners** Piezoelectrics are materials whose dimensions deform in response to an applied electric field. If the voltages that are applied to a piezoelectric are precisely controlled, extremely precise movements can be performed. The geometry of piezoelectric positioning devices used in STM includes bars, bimorphs, or tubes; tubular scanners are the most commonly used. They operate at a high resonant frequency, enabling high scan rates. Thermal isolation of the piezoelectric elements is necessary since these ceramics are also sensitive to

temperature fluctuation. The sensitivity of a piezoelectric to temperature changes is defined as:

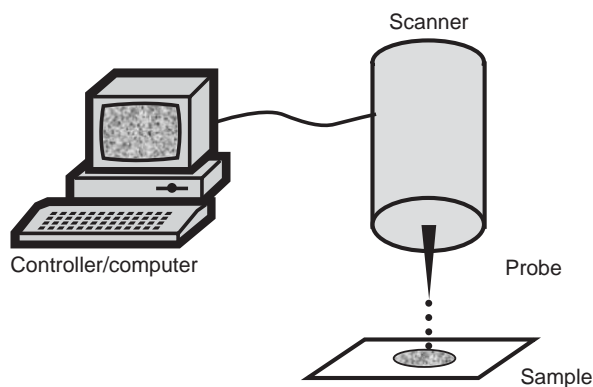
$$\frac{\Delta x}{x_0} = \Delta \alpha * \Delta T$$

where  $x_0$  is the original dimension of the piezoelectric,  $\alpha$  the thermal expansion coefficient, and  $T$  the temperature. Piezoelectric scanners exhibit both linear and nonlinear responses to an applied voltage. As a result, careful scanner calibration is required. Scanner dimensions must be carefully matched with the desired image domain.

**STM tips** The geometry and material of the STM probe tip are critical variables required for achieving atomic-scale images. The type of material used for the STM probe varies based upon the conditions of the experiment. If one is interested in achieving atomic-scale images under ambient conditions, inert metals such as platinum or gold will suffice. A drawback of these materials is that both are very malleable and if accidental contact is made between the probe tip and the sample, the apex of the tip can be severely blunted. Thus, tungsten or platinum-iridium alloy tips are used due to their high elastic modulus. For STM experiments carried out under ultrahigh vacuum (UHV) conditions, harder materials such as iridium, molybdenum, or tungsten are typically used.

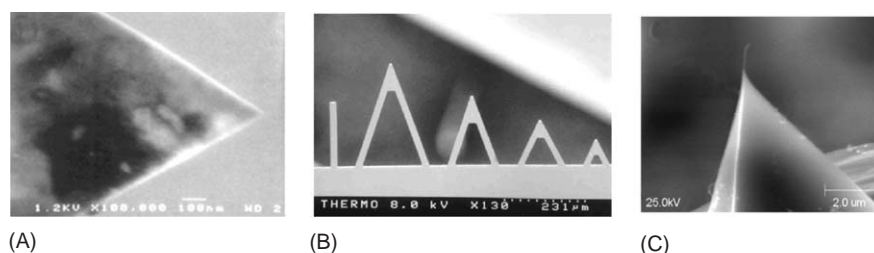
There are several methods for creating ultrasharp probes. For soft metals, tips can be fabricated mechanically by simple cutting of the probe wire (multiple attempts may be required) (Figure 3A). Probes can be sharpened by electrochemically etching the wire. The latter method will generate more consistent geometries but may also create an insulating oxide on the surface of the tip. Extremely sharp tips can be generated using a focused ion beam mill; however, the cost and availability of this tool currently precludes its widespread use. With all of these techniques, care must be taken to minimize the possibility of creating a double-tip where current is passed from multiple points between the probe and the sample. This can lead to tip-induced artifacts in the topographical image.

**Software** It is imperative that movement of the scanner be frequently calibrated. This is most often achieved by scanning a surface whose feature dimensions are accurately and precisely known. Calibration allows users to compare results from one instrument to the next and, more importantly, instills confidence in the length scales of the measured topographical features. The user must remain cognizant that they cannot visually see the sample. The image displayed is a pictorial representation of the

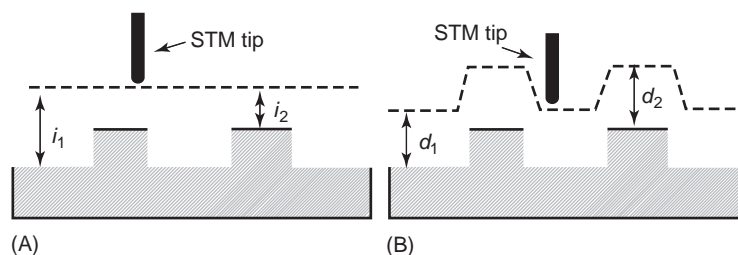


**Figure 2** Schematic representation of the essential components of a scanning tunneling microscope.





**Figure 3** Scanning electron micrograph of (A) mechanically sharpened platinum-iridium STM probe, (B) silicon nitride AFM cantilevers, and (C) carbon nanotube tip on silicon nitride AFM cantilever tip. (Images (A) and (B) are courtesy of Veeco Instruments; Santa Barbara, CA, USA, with permission.)



**Figure 4** A schematic depicting STM operation in (A) constant height and (B) constant current modes. In constant height mode, the tip moves horizontally over features without change in its vertical position; the change in current between the substrate and the tip ( $i_1 < i_2$ ) is recorded. In constant current mode, a feedback loop maintains a constant gap between the tip and substrate ( $d_1 = d_2$ ) by adjusting the vertical position of the tip as it moves horizontally over features.

changes in current as a function of movement in the scanner. If movement of the scanner is poorly calibrated then there is little reliability in image feature dimensions. For atomic-scale work, highly oriented pyrolytic graphite (HOPG) is a widely used calibration standard. This material is comprised of graphene sheets of carbon atoms that have hexagonal packing and well-defined carbon spacings (1.42 Å).

### Operation

Two modes are used for acquiring STM images: constant height and constant current. When the STM is operated in constant height mode, the STM probe is held at a fixed position over the sample and changes in current between the tip and the substrate (that occur when the tip travels over high or low areas) are plotted as a function of the  $x$  and  $y$  position of the scanner (Figure 4A). The constant height mode of operation can potentially lead to tip damage (also called tip crash) if the tip encounters a feature taller than the separation between the tip and the sample. When operating in constant current mode, the current between the STM tip and the conductive sample are kept constant. As the tip travels over high or low areas, changes in the scanner movement in the  $z$ -direction are plotted as a function of the  $x$  and  $y$  position of the scanner (Figure 4B). This is accomplished via the integration of a feedback loop

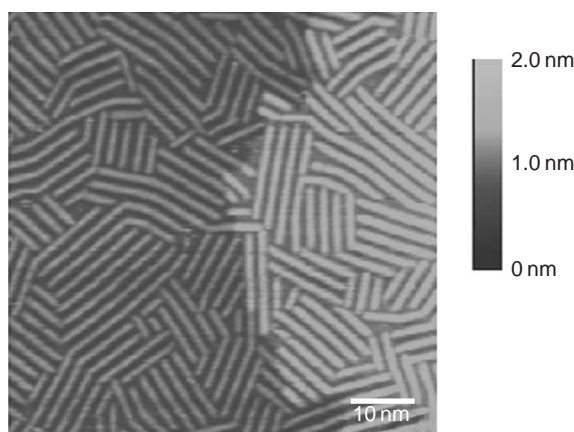
between the scanner and the controller. For example, when the current exceeds a preselected reference value, the controller increases the separation between the tip and the substrate until the measured current equals the reference.

### Applications

Scanning tunneling microscopy has been widely applied in research and manufacturing in fields spanning from biology to material science to microelectronics. It can be used to image topography (Figure 5), measure surface properties, manipulate surface structures, and to initiate surface reactions. The STM is an important tool in nanotechnology enabling accurate measurement of feature dimensions on the atomic scale, as well as moving and placing atomic-scale building blocks at specific locations on a surface. The latter capability makes possible the design of novel structures from single atoms or molecules.

### Atomic Force Microscopy

An AFM can achieve atomic resolution under certain conditions; however, surface features are typically resolved on the nanometer scale laterally and the angstrom scale vertically. This resolution is comparable to that of a transmission electron microscope (TEM). Resolution of the AFM depends on tip shape, scan

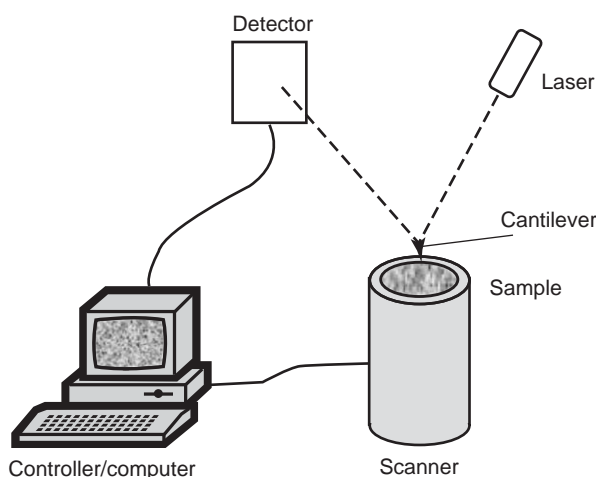


**Figure 5** An STM topographical image of self-assembled submonolayer of dodecanethiol chemisorbed onto an atomically flat gold surface. The 'herringbone' pattern results from the two-dimensional nucleation of the thiols on the surface (laying down of the carbon chains on the gold surface following Au–S chemical bond formation).

domain, and number of measurements per scan. Unlike many other forms of microscopy, atomic force microscopy does not require any stains, contrast agents, or conductive coatings that can cause partial obstruction of the actual sample. In addition, sample preparation is fairly simple and quick. It is a nondestructive method that allows the sample to be reused for additional analyses over time. The AFM can image a variety of materials (i.e., nonconductive, magnetic, elastic) under a variety of environmental conditions (i.e., ambient air, various gases, different humidity levels, and temperatures) and can be used to study both short- and long-range molecular and atomic forces.

## AFM Instrumentation

Atomic force microscopy measures the surface topography of a sample on a nanometer scale and the attractive and repulsive forces on a piconewton scale. An analogy can be made between AFM and a person using a cane to navigate along a landscape. The blind person strokes or taps his cane along the ground to feel where the ground is higher or lower, as well as softer or harder and rougher or smoother. In a similar fashion, the AFM measures surface topography and compliance of the sample. **Figure 6** shows a schematic of the key components of an AFM. It works by scanning a sample under an atomically sharp tip integrated on the end of a cantilever. The cantilever bends up or down as the tip contacts and is displaced by features on the surface. The most common method to measure these small movements is to use a laser beam focused on the end of the cantilever that is reflected onto a position sensitive detector. Changes in the angle of the



**Figure 6** Schematic representation of the essential components of an atomic force microscope.

cantilever induced by changes in sample topography result in different voltage levels out of the detector. These voltages are sent to a computer for processing and display of the topographic image. This scheme is the simplest and most commonly used in commercial instruments. Variations of this scheme include scanning the tip over the sample surface to allow analysis of larger samples and the use of different detectors to allow analysis of rougher samples.

## Cantilevers

Cantilevers are commercially produced from silicon and silicon nitride using microfabrication processes similar to those employed in integrated circuit manufacture. Cantilevers are often coated with metal on the topside to enhance laser reflection. There are two primary geometries of AFM cantilevers – triangular and rectangular (**Figure 3B**). Rectangular cantilevers are susceptible to torsional bending during contact mode imaging, whereas triangular cantilevers have been designed to reduce the lateral forces exerted on the cantilever.

Important cantilever specifications are their dimensions, flexibility, and resonance characteristics (resonance frequency and quality factor). The fundamental resonance frequency of the cantilever should be high in order to avoid interference from building vibrations and acoustic noise. Thermal noise from the cantilever can limit the sensitivity of the AFM and should be minimized by using shorter length cantilevers. A shorter length cantilever also provides greater angular displacement of the laser resulting in greater resolution. The spring constant ( $k$ ) of a cantilever is a measure of its flexibility. The spring constant of a rectangular cantilever is calculated from its geometry ( $w$ , width,  $l$ , length, and

$t$ , thickness) and material properties (Young's modulus,  $E$ ). The equation for the spring constant of a rectangular cantilever is

$$k = \frac{Ewt^3}{4l^3}$$

The spring constant increases with cantilever width and thickness but decreases with cantilever length. Typically stiffer cantilevers are used for dynamic imaging modes.

As in STM, the cantilever tip shape is critical to the quality of the AFM image. The overall tip geometry is often square pyramidal resulting from an etching process. The aspect ratio (height to width ratio) of the tip is important for imaging rough samples in order to fully contact recesses. The tip sharpness, or the radius of curvature at the apex, will produce artifacts in the final image. Recently, carbon nanotubes have become popular tips because of their high aspect ratio, small radius of curvature, and durability (Figure 3C). Tips are often functionalized in order to study tip to sample interactions or to conduct chemically sensitive measurements. Arrays of parallel cantilevers have been developed for use in sensor applications, data storage, and increasing scan sizes and imaging speed.

### Piezoelectric Scanner

The piezoelectric scanners used in AFM have the same characteristics as those mentioned in the STM section. The sample can be mounted directly onto the scanner and rastered underneath the cantilever tip, or the cantilever can be mounted to a scanner tube and rastered over a sample fixed below it. The former case is advantageous in imaging larger samples and increases the speed of imaging.

### Detectors

The original AFM used an STM to sense the movement of the cantilever in response to interactions with the sample. In most commercial AFM instruments, optical detectors have supplanted this type of electrical detector. The most widely used detection system uses laser beam reflection off the end of the cantilever into either a position sensitive detector or a quadrant photodiode. A change in the angle of the cantilever moves the spot on the detector, producing a change in the voltage out of the detector.

Interferometry is another detection system where a laser beam is reflected off the end of the cantilever. The phase change of the reflected laser beam is then compared to that of a standard to determine the change in cantilever position. This type of detection offers a better signal-to-noise ratio but it is more

difficult to set up and requires better vibrational and acoustic isolation.

Integration of the detector into the cantilever has also been accomplished using the piezoresistive properties of silica. This is beneficial for samples that are light sensitive and following the movements of cantilevers in an array. It can be used in conjunction with the integrated piezoelectric actuator; however, it can be difficult to decouple the signals.

## AFM Operational Modes

An AFM can be operated in several modes (contact, intermittent contact, and noncontact) to optimize topographical imaging of different types of samples, to explore different surface properties of samples, or to modify the surface.

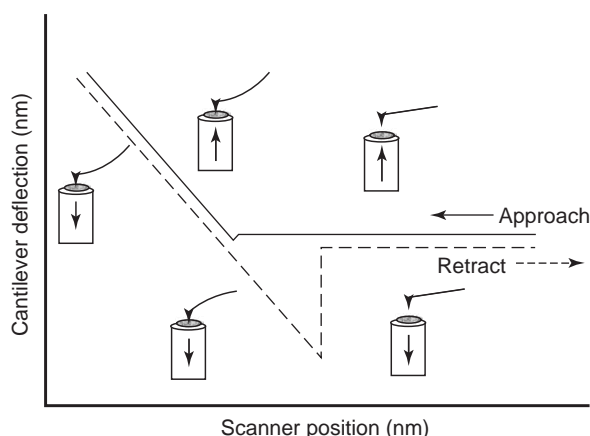
### Contact Mode

In this mode, the probe tip is maintained in contact with the sample surface. The image can be acquired under four different imaging methods: deflection, height, friction, and force modulation. In the 'deflection method', the image is a map of cantilever bending as a function of position and is acquired under variable vertical loading forces. To image under constant applied force, the 'height method' uses a closed-loop feedback system to raise or lower the sample (or tip) to maintain a fixed cantilever deflection. The user optimizes feedback system performance by adjustment of the proportional and integral gains. Proportional gain sets the magnitude, whereas, integral gain sets the temporal response of corrective action to return the cantilever to its reference position. If the gains are too high, feedback oscillations produce image artifacts. If the gains are too low, resolution is poor due to slow response of the feedback loop.

Force curves are used to correlate vertical movement of the scanner with cantilever deflection. A typical force curve is depicted in Figure 7. Force curves are obtained by disabling the scanner movement in the  $x$  and  $y$  directions and oscillating the scanner in the  $z$ -direction. As the tip approaches the sample, no cantilever bending is observed until the gap between the tip and the sample is extremely small. Then, the cantilever will be pulled down due to the van der Waals interactions between the tip and the substrate (or by the presence of a thin hydration layer on the surface). Once in contact, raising the sample produces an upward bend in the cantilever. The applied load is computed by multiplying the reference cantilever deflection value times its spring constant. The force curve can also be used to identify adhesive

interactions between the tip and the surface. On the downward movement of the scanner, the cantilever flattens. In the presence of attractive or adhesive forces, downward deflection of the cantilever may be observed. When the restoring force of the cantilever exceeds these forces, the tip releases from the substrate surface and the cantilever returns to its original position. The magnitude of the attractive or adhesive force is calculated by multiplying the downward deflection of the cantilever times its spring constant.

The 'friction force' method measures the local variations of adhesion that may exist between the cantilever tip and the substrate. Lateral forces on the tip cause torsional bending of the cantilever that is detected by horizontal movement of the laser spot on the detector.



**Figure 7** A typical force curve acquired by monitoring cantilever deflection as the piezoelectric raises and withdraws the sample surface from the tip. The approach and retract curves are dissimilar when strong chemical and/or physical attraction between the tip and the surface exists. Illustrations of the cantilever bending at various scanner positions are depicted above and below the force curve to assist the reader in interpreting the force curve.

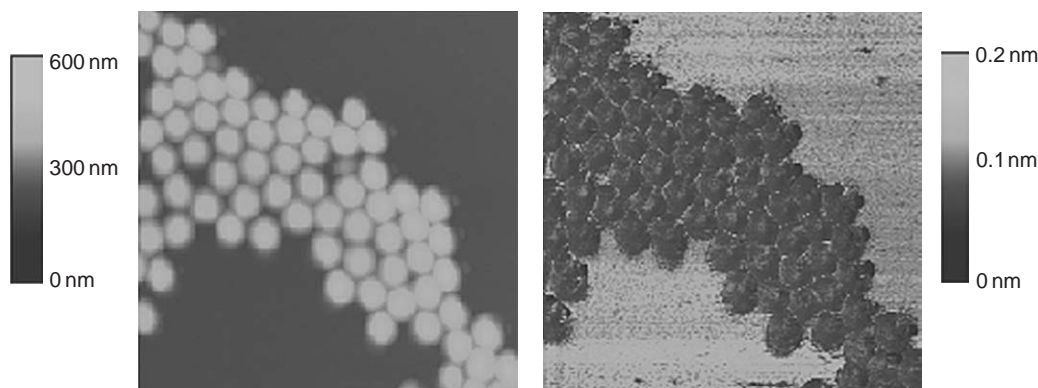
'Force modulation' is a contact imaging method that enables simultaneous measurement of the topology and compliance of the substrate. In force modulation, a piezoelectric actuator oscillates the cantilever chip causing the tip to slightly indent the substrate. Compliant areas on the sample will produce larger cantilever deflections compared to more rigid areas (see **Figure 8**). This is an invaluable method for characterizing the mechanical and topological properties of composite materials on the nanoscale.

### Noncontact Mode

Noncontact mode (or dynamic force) imaging relies on longer-range molecular forces to obtain surface topology. The cantilever is oscillated at its resonance frequency by a piezoactuator. The surface topography is measured by the shift in cantilever frequency based on tip/sample interactions. Specialized expertise is required to avoid image artifacts.

### Intermittent Contact Mode

Intermittent contact or TappingMode<sup>TM</sup> also oscillates the cantilever at its resonance frequency. Surface topology is measured with the height method while maintaining constant dampening of cantilever oscillation rather than cantilever deflection in contact mode. Cantilever oscillation amplitude is maintained at a fixed value using a closed-loop feedback system to raise or lower the sample (or tip). The major advantages of intermittent contact mode imaging are reduced vertical loads, lateral forces, and adhesive interactions between the tip and the substrate. This mode enables imaging of soft samples or molecules that are not strongly attached to the surface. Low drive amplitude is desired to minimize impact on soft molecules; however, it must be sufficient to overcome attractive capillary forces. This problem can be



**Figure 8** Height mode topographical (A) and force modulation (B) images of colloidal hydrogel nanoparticles immobilized on a silica substrate. Scan domain is  $12\mu\text{m} \times 12\mu\text{m}$ . Note that in the topographical image darker areas represent lower regions, whereas in the force modulation image the darker areas represent greater surface stiffness.

resolved by imaging under fluid. Imaging under fluid also provides the opportunity to examine interactions between the substrate and molecules in the liquid environment.

'Phase imaging' is an extension of intermittent contact mode that enables measurement of surface properties such as adhesion, friction, and elasticity. Phase imaging detects the phase shift of the cantilever oscillation, which is related to the surface rigidity. Phase shifts above  $90^\circ$  are due to attractive interactions, whereas phase shifts below  $90^\circ$  are due to repulsive interactions.

## AFM Applications

Atomic force microscopy has been applied to similar fields as scanning tunneling microscopy; however, AFM enables scientists to visualize non-conducting surfaces that were not possible with STM. This analytical tool enables researchers to characterize surface topology on the nanometer scale and visualize the orientation and spatial distribution of molecules adsorbed to these surfaces. Noteworthy biological applications of atomic force microscopy include the examination of drug-DNA and DNA-DNA interactions, protein conformations, enzyme reactions, and cell membrane structures (Figure 9). Applications in material science include the characterization of the surface of polymers, plastics, and coatings.

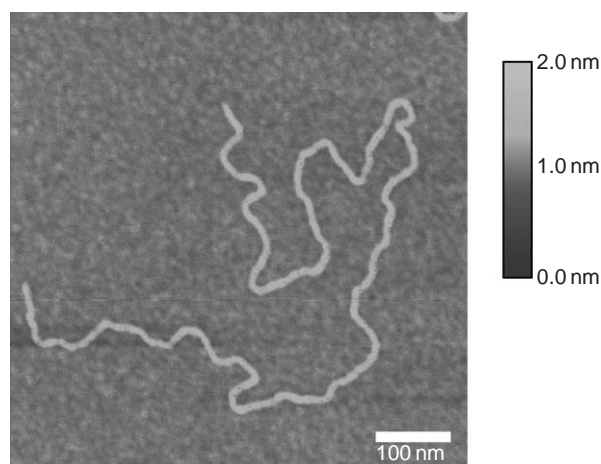
AFM has also been used to measure the mechanical properties of single molecules, molecular ensembles (membranes, cells), and surface structures. It has also enabled measurement of adhesion on the nanoscale that may exist between a chemically modified

tip and a substrate. This particular application is referred to as 'chemical force microscopy'. The experimental design for the single molecule mechanical testing is an elegant example of the chemist's ability to chemically tailor surfaces on the nanoscale. Generally these experiments are done by chemical tethering of a molecule on either a substrate or the end of the cantilever tip. The free end of the molecule possesses chemical functionality that reacts with the opposing surface. When the two surfaces are brought into close proximity, coupling of the free end of the molecule to a surface bound reactive group occurs. Pulling the two surfaces apart results in mechanical unfolding of the anchored molecule. Measurement of single molecule elasticity is afforded by analysis of the force curves generated during the pulling process. As an illustration, Figure 10 presents the repeated unfolding (tensile loading) of a polypeptide molecule using this approach.

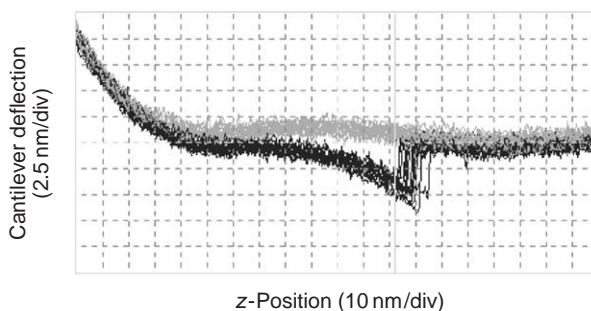
The AFM has been employed in manufacturing and quality control of microelectronics, including semiconductor silicon wafers, MEMS devices, CDs/DVDs, and computer hard disks. The AFM is capable of identifying defects that are too small to be seen using optical microscopy.

An exciting new and evolving application of the AFM involves the manipulation of matter on the atomic scale. The AFM tip can be used to place, move, or react chemical moieties at specific locations on a surface. This can be achieved by frictional wear, controlled surface oxidation, and/or deliberate material transport between two surfaces. Material transport is unrestricted; atoms, molecules, or even cells can be attached to the AFM tip and transferred to the opposing surface at a desired location. In addition, thermal patterning of substrates can be achieved by contacting a heated cantilever tip with the surface. Potential applications include preparation of ultrahigh density data storage devices, nanoscale sensors, and molecular machines.

In summary, SPM has revolutionized our understanding of chemical phenomena on the nanoscale.



**Figure 9** AFM topographic image of a linearized plasmid DNA electrostatically immobilized on mica. This image was acquired using intermittent contact mode.



**Figure 10** Sequential force curves acquired with an AFM depicting the unfolding of an elastin-mimetic peptide.



New applications of this important nanotechnological tool will make possible the manipulation of matter on the atomic scale and enable the construction of new materials and devices for the benefit of mankind.

**See also:** **Microscopy Applications:** Biomedical; Liquid Crystals. **Microscopy Techniques:** Electron Microscopy; Specimen Preparation for Electron Microscopy; Scanning Electron Microscopy.

## Further Reading

Bonnell DA (2000) *Scanning Tunneling Microscopy and Spectroscopy: Theory, Techniques, and Applications*, 2nd edn. New York: Wiley-VCH.

Hansma HG (2001) Surface biology of DNA by atomic force microscopy. *Annual Reviews of Physical Chemistry* 52: 71–92.

Hofer WA, Foster AS, and Shluger AL (2003) Theories of scanning probe microscopes at the atomic scale. *Reviews of Modern Physics* 75: 1287–1331.

Meyer E, Hug H-J, and Bennewitz R (2004) *Scanning Probe Microscopy – The Lab on a Tip*. Berlin: Springer.

Morita S, Wiesendanger R, and Meyer E (2002) *Noncontact Atomic Force Microscopy, Nanoscience and Technology*. Berlin: Springer.

Poggi MA, Bottomley LA, and Lillehei PT (2002) Scanning probe microscopy. *Analytical Chemistry* 74: 2851–2862.

Sarid D (1994) *Scanning Probe Microscopy: with Applications to Electric, Magnetic, and Atomic Forces*. New York: Oxford University Press.

Wiesendanger R (1994) *Scanning Probe Microscopy and Spectroscopy: Methods and Applications*. New York: Cambridge University Press.

## X-Ray Microscopy

**I Snigireva and A Snigireva**, European Synchrotron Radiation Facility, Grenoble, France

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The resolution of X-ray microscopy is between the visible light microscope on the one hand and the electron microscope on the other. Instruments available today deliver about five times better resolution than visible light microscopes on a routine basis and the best achievable resolution today is 20–30 nm. The ultimate resolution of X-ray microscopy is limited, in principle, only by the wavelength:

$$\lambda = \frac{hc}{E} = \frac{1.24 \text{ keV nm}}{\text{photon energy (keV)}}$$

which is typically in the range between a tenth of a nanometer and few nanometers.

X-rays are more penetrating than electrons; further, the penetrating power of X-rays varies much more strongly with photon energy and chemical composition. As a consequence, X-ray microscopes can image samples without sectioning in the natural state and in atmospheric-pressure air. X-ray absorption edges (jumps in the absorption cross-section curve that occur when the X-ray passes the threshold energy for removing an inner-shell electron) and near-edge resonances enable the mapping of both chemical elements and their chemical binding states.

Although the intrinsic advantages of X-rays have long been recognized, their full potential for imaging could not be realized until synchrotron X-ray sources, giving radiation of required wavelength, and high-performance X-ray microfocusing optics were developed.

The history of X-ray microscopy goes back to 1896, the year following to discovery of X-rays by Roentgen. The method used for studying the structural details of biological objects by enlargement of X-ray radiographs was called ‘microradiography’ by P. Goby in 1913. Beginning in the late 1940s, X-ray microscopy with grazing incidence mirror optics was proposed by P. Kirkpatrick to exceed the optical microscope in resolution.

As a branch of earlier developments in electron microscopy, projection microscopy was proposed by V.E. Cosslett and became very popular since the 1950s. In the early 1970s, several groups started new technological developments of X-ray optics, mainly Fresnel zone plates, and the modern era of X-ray microscopy started. In 1974, G. Schmahl and collaborators built a full-field transmission microscope at DESY (Deutsches Elektronen Synchrotron) in Germany. J. Kirz and colleagues at NSLS (National Synchrotron Light Source) at Brookhaven National Laboratory in USA built the first scanning transmission microscope using zone plate objective in 1982. Traditionally, this type of X-ray microscopy deals with rather soft X-ray energies (100–2000 eV), in particular in the so-called water window region between the K-shell X-ray absorption edges of carbon

New applications of this important nanotechnological tool will make possible the manipulation of matter on the atomic scale and enable the construction of new materials and devices for the benefit of mankind.

**See also:** **Microscopy Applications:** Biomedical; Liquid Crystals. **Microscopy Techniques:** Electron Microscopy; Specimen Preparation for Electron Microscopy; Scanning Electron Microscopy.

## Further Reading

Bonnell DA (2000) *Scanning Tunneling Microscopy and Spectroscopy: Theory, Techniques, and Applications*, 2nd edn. New York: Wiley-VCH.

Hansma HG (2001) Surface biology of DNA by atomic force microscopy. *Annual Reviews of Physical Chemistry* 52: 71–92.

Hofer WA, Foster AS, and Shluger AL (2003) Theories of scanning probe microscopes at the atomic scale. *Reviews of Modern Physics* 75: 1287–1331.

Meyer E, Hug H-J, and Bennewitz R (2004) *Scanning Probe Microscopy – The Lab on a Tip*. Berlin: Springer.

Morita S, Wiesendanger R, and Meyer E (2002) *Noncontact Atomic Force Microscopy, Nanoscience and Technology*. Berlin: Springer.

Poggi MA, Bottomley LA, and Lillehei PT (2002) Scanning probe microscopy. *Analytical Chemistry* 74: 2851–2862.

Sarid D (1994) *Scanning Probe Microscopy: with Applications to Electric, Magnetic, and Atomic Forces*. New York: Oxford University Press.

Wiesendanger R (1994) *Scanning Probe Microscopy and Spectroscopy: Methods and Applications*. New York: Cambridge University Press.

## X-Ray Microscopy

**I Snigireva and A Snigireva**, European Synchrotron Radiation Facility, Grenoble, France

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The resolution of X-ray microscopy is between the visible light microscope on the one hand and the electron microscope on the other. Instruments available today deliver about five times better resolution than visible light microscopes on a routine basis and the best achievable resolution today is 20–30 nm. The ultimate resolution of X-ray microscopy is limited, in principle, only by the wavelength:

$$\lambda = \frac{hc}{E} = \frac{1.24 \text{ keV nm}}{\text{photon energy (keV)}}$$

which is typically in the range between a tenth of a nanometer and few nanometers.

X-rays are more penetrating than electrons; further, the penetrating power of X-rays varies much more strongly with photon energy and chemical composition. As a consequence, X-ray microscopes can image samples without sectioning in the natural state and in atmospheric-pressure air. X-ray absorption edges (jumps in the absorption cross-section curve that occur when the X-ray passes the threshold energy for removing an inner-shell electron) and near-edge resonances enable the mapping of both chemical elements and their chemical binding states.

Although the intrinsic advantages of X-rays have long been recognized, their full potential for imaging could not be realized until synchrotron X-ray sources, giving radiation of required wavelength, and high-performance X-ray microfocusing optics were developed.

The history of X-ray microscopy goes back to 1896, the year following to discovery of X-rays by Roentgen. The method used for studying the structural details of biological objects by enlargement of X-ray radiographs was called ‘microradiography’ by P. Goby in 1913. Beginning in the late 1940s, X-ray microscopy with grazing incidence mirror optics was proposed by P. Kirkpatrick to exceed the optical microscope in resolution.

As a branch of earlier developments in electron microscopy, projection microscopy was proposed by V.E. Cosslett and became very popular since the 1950s. In the early 1970s, several groups started new technological developments of X-ray optics, mainly Fresnel zone plates, and the modern era of X-ray microscopy started. In 1974, G. Schmahl and collaborators built a full-field transmission microscope at DESY (Deutsches Elektronen Synchrotron) in Germany. J. Kirz and colleagues at NSLS (National Synchrotron Light Source) at Brookhaven National Laboratory in USA built the first scanning transmission microscope using zone plate objective in 1982. Traditionally, this type of X-ray microscopy deals with rather soft X-ray energies (100–2000 eV), in particular in the so-called water window region between the K-shell X-ray absorption edges of carbon

**Table 1** Comparison of the features of different types of X-ray microscopy

Type of microscopy	Source	Energy (keV)	Mode of contrast	Optics	Resolution (nm)
<i>Full-field microscopy</i>					
Microradiography					
Contact	SR, Lab	>0.5	Absorption	–	100–1000
Inline imaging	SR	>1	Absorption/phase, XANES/EXAFS	–	500–1000
Magnified projection (no optics)	Lab	>1	Absorption/phase	–	100–500
Soft X-ray microscopy	SR, Lab	<4	Amplitude/phase, XANES	FZP	20–50
Hard X-ray microscopy	SR, Lab	>4	Amplitude/phase, XANES/EXAFS possible	FZP, CRL	100–500
<i>Scanning microscopy</i>					
Soft X-ray scanning microscopy	SR	<4	Absorption, fluorescence, XANES/EXAFS, photoemission	FZP	20–50
Hard X-ray microprobe	SR, Lab	>4	Absorption, fluorescence, XANES/EXAFS, diffraction	FZP, CRL, KB	100–1000

SR – synchrotron radiation sources; Lab – laboratory based sources; XANES – X-ray absorption near edge spectroscopy; EXAFS – extended X-ray absorption fine spectroscopy.

and oxygen at 4.4 nm and 2.3 nm where organic materials show strong absorption and phase contrast, while water is relatively nonabsorbing. This enables imaging of specimens up to  $\sim 10\ \mu\text{m}$  thick with high intrinsic contrast using X-rays with a lateral resolution down to 25 nm.

During the last years considerable progress has been made in X-ray microscopy in the hard X-ray regime ( $E > 4\ \text{keV}$ ), which was enhanced by the simultaneous development of high-brilliance, high-energy X-ray sources coupled with advances in manufacturing technologies of focusing optics. One of the key strengths of hard X-ray microscopy is the large penetration depth of hard X-rays in matter (around 1 mm), allowing one to probe the inner structure of an object without the need for destructive sample preparation. Resolution in the order of 100 nm was reached with photon energies up to 30 keV.

X-ray microscopy can be divided in two classes: full-field microscopy and scanning microscopy. There are two versions of full-field X-ray imaging and microscopy. The first version is the microradiography involving lensless imaging with parallel and conical beams. The second version is the full-field transmission X-ray microscope, which uses the same optical arrangement as conventional light and transmission electron microscopes. Such types of microscopes use optical elements like Fresnel zone plates or refractive optics as objective lenses for high-resolution imaging. Scanning microscopes usually use a focusing optics to form a finely focused spot or microprobe through which the specimen is rastered. The microscope of

choice is generally determined by the specimen and observation to be made on it. **Table 1** summarizes different types of X-ray microscopes available today.

## Interaction of X-Rays with Matter

X-rays interact with matter in three ways – elastic scattering, inelastic scattering, and absorption via photoelectric effect. Elastic or coherent scattering is caused by two processes: Thomson scattering from single atomic electrons, and Rayleigh scattering, which occurs from strongly bound electrons acting cooperatively. Inelastic, incoherent, or Compton scattering occurs from loosely bound (essentially free) electrons and involves the transfer of a small fraction of the incident X-ray energy. Absorption occurs when the X-ray photon transfers all of its energy to an inner atomic electron, thereby ionizing the atom. We are mainly interested in coherent scattering and absorption; that is, the phase shift of an incoming wave and attenuation of its amplitude.

Let us consider a material of thickness  $t$  and refractive index

$$n = 1 - \delta - i\beta \quad [1]$$

illuminated by a plane wave with amplitude  $A_0$ . If  $A_1$  is the amplitude of the transmitted wave, the amplitude transmission is given by

$$T = \frac{A_1}{A_0} = \exp\left(-\frac{2\pi}{\lambda}\beta t\right) \exp\left(i\frac{2\pi}{\lambda}\delta t\right) \exp\left(-i\frac{2\pi}{\lambda}t\right) \quad [2]$$

The first factor describes the attenuation of the amplitude of the incoming wave; the second factor describes the phase shift of incident beam caused by the object.  $\beta$  and  $\gamma$  are related to the linear absorption coefficient  $\mu = (4\pi/\lambda)\beta$  and to the phase shift per unit length  $\eta = (2\pi/\lambda)\delta$ .

The macroscopic parameters  $\beta$ ,  $\mu$ , and  $\delta$  are related to the atomic scattering factor  $f = f_1 + if_2$  by

$$\delta = \frac{r_0 \lambda^2}{2\pi} n f_1 \quad \beta = \frac{r_0 \lambda^2}{2\pi} n f_2$$

where  $r_0$  is the classical electron radius and  $n$  is the number of atoms per unit volume. The atomic scattering factor has been tabulated for all elements.  $f_1$  and  $f_2$  vary strongly with photon energy and chemical composition, and so do the linear absorption coefficient  $\mu$  and phase shift coefficient  $\eta$ .

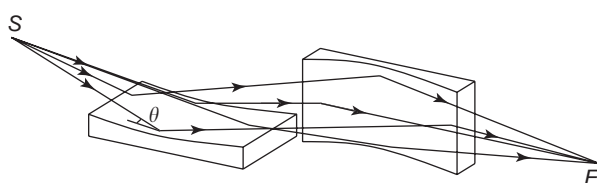
For X-rays,  $\delta$  is very small ( $10^{-3}$ – $10^{-6}$ ) and positive, i.e., the refractive index is slightly less than unity. This property of X-rays causes a near absence of reflection at interfaces and results in very clear images, even of thick specimens. X-ray microscopic samples are amplitude objects as well phase objects and X-ray microscopy can be performed in amplitude (absorption) contrast and phase contrast. Hence, at high photon energies for low-density materials it is much more favorable to image an object in phase contrast rather than in absorption contrast. In addition to natural contrast mechanisms, X-rays can provide chemical bond mapping or trace elements mapping of many elements.

## X-Ray Optics

Ever since the discovery of X-rays the possibility of using them to image the microscopic structure of biological and nonbiological materials has been discussed. However, the practical implementation of an X-ray microscope was hampered until recently by the lack of intense sources giving radiation of required wavelength, and by the severe difficulties in the manufacturing of suitable optical components.

### Reflective Optics: Mirrors and Multilayers

Conventional mirrors, used for visible wavelength, cannot be used, because the reflectivity at normal incidence is too low; typically, less than 100 000 X-ray photons will be reflected. High reflectivity can be obtained at grazing incidence angles ( $\theta < \sqrt{\delta}$ ). To reduce the astigmatism Kirkpatrick and Baez proposed to use two spherical or cylindrical mirrors in a crossed configuration. **Figure 1** shows a geometrical arrangement of the so-called Kirkpatrick–Baez system (KB mirrors). To enhance the reflectivity the



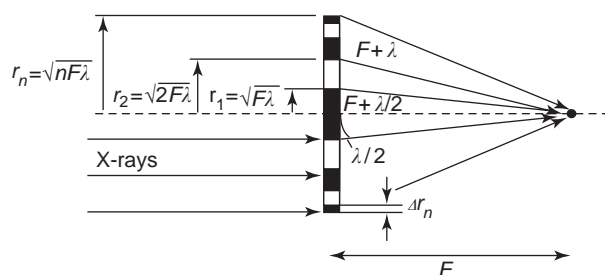
**Figure 1** Two-mirror Kirkpatrick–Baez system.

multilayer mirrors, where the refractive index varies periodically with depth, can be used. To build such mirrors, alternative layers are evaporated or sputtered onto a blank with a thickness  $\lambda/4$  and relatively high refractive index and relatively low refractive index, respectively. The weak radiation reflected at the interfaces of the multiple layer is superimposed coherently and in phase, and can give a considerably increased reflectivity. The reflecting focusing systems can be either static, with mirrors polished according to the proper figure optimized for a given incidence angle and focus, or it can be dynamic, with actuators bending flat mirrors into the elliptic shapes required by the experiment. Today, as a result of improved techniques for developing highly finished and perfect surfaces, ellipsoidal and paraboloidal mirrors can be manufactured to a high degree of perfection and focusing down to 100 nm spot size is achievable with KB systems.

The reflecting mirrors are able to focus X-rays over a large energy range, whereas Bragg reflecting mirrors only focus a small fraction of the radiation continuum. Within a given energy interval, as the optical properties for total reflecting optics are independent of the X-ray wavelength, the focal spot is retained while tuning the energy. Therefore, experiments requiring energy tuning, like spectroscopy, can be performed without any readjustment of the optics. However, the disadvantage of grazing incidence optics is low acceptance and it is mostly used for scanning microscopes.

### Fresnel Zone Plates

The focusing properties of zone plates were first discussed in the latter part of the nineteenth century, and Baez originally suggested their use as X-ray optical elements in 1952. In their most common form, this is circular diffraction grating that works as a lens for monochromatic light. A Fresnel zone plate consists of a series of concentric rings of radius  $r_n^2 = n\lambda F$  and the rings become narrower at larger radii until the last, finest zone of width  $\Delta r_n$  is reached (**Figure 2**). Linear, square, elliptical zone plates have also been considered, but only circular and, to a lesser extent, linear and elliptical forms have generally been used. The focusing capability is based on



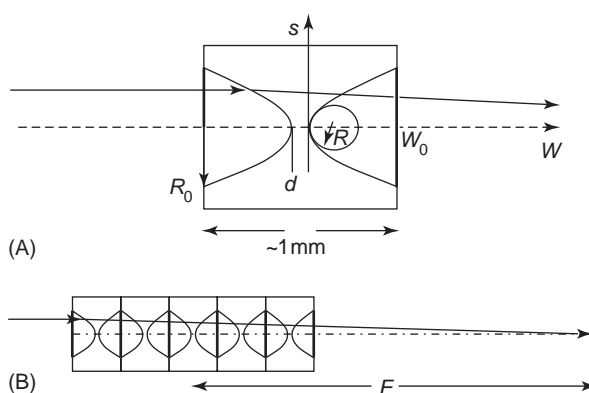
**Figure 2** Fresnel zone plate.

constructive interference of the wavefront modified by passage through the zone plate. The wavefront modification is obtained through the introduction of a relative change in amplitude or phase in the beams emerging from two neighboring zones. A zone plate is called an amplitude zone plate if the focusing results from different absorption between two neighboring zones. It is called a phase zone plate if the phase change upon transmission through a zone is the mechanism for focusing.

Just as the Rayleigh resolution of a light microscope lens is determined by its numerical aperture, NA, the Rayleigh resolution of a zone plate is determined by its maximum diffraction angle  $NA = \lambda / (2\Delta r_n)$ , so that the Rayleigh resolution is  $1.22\Delta r_n$ . Using state-of-the-art lithographic technologies, zone plates with last or outermost ring width of less than 20 nm can currently be fabricated. If illuminated with an X-ray beam whose spatial coherence length is equal to or greater than the diameter of the zone plate, a diffraction-limited focus can be obtained. In the soft X-ray region, zone plate efficiency is limited to  $\sim 15\%$  due to photoelectric absorption. For X-ray energies greater than 3 keV, it is, in principle, possible to produce phase zone plates with focusing efficiencies  $\sim 40\%$ .

### Refractive Lenses

Refractive lenses made of glass are among the most widely used optical components for visible light, with the wide spectrum of applications in focusing and imaging. Refractive lenses for X-rays were considered unfeasible for a long time due to the weak refraction and strong absorption. However, in 1996, it was shown that focusing by X-ray lenses is possible. Since  $(1 - \delta)$  in the index of refraction is smaller than 1, lenses must have a concave shape (Figure 3A). In order to obtain a focal length  $F$  in the range of 1 m many single lenses have to be stacked behind each other to form a compound refractive lens (CRL) as shown in Figure 3B. Fabricating the lenses from low-Z materials like Li, Be, B, C, and Al minimizes the problems associated with absorption. The focal



**Figure 3** Parabolic compound refractive lens (CRL). The individual lenses (A) are stacked to form a CRL (B).

length of such CRL with parabolic profile  $s^2 = 2Rw$  and  $N$  individual biconcave lenses is  $F = R/2N\delta$ , where  $R$  is the radius of curvature at the apex of parabola. A lens with thickness  $2w_0 + d$  has an aperture  $2R_0 = 2\sqrt{2Rw_0}$ .

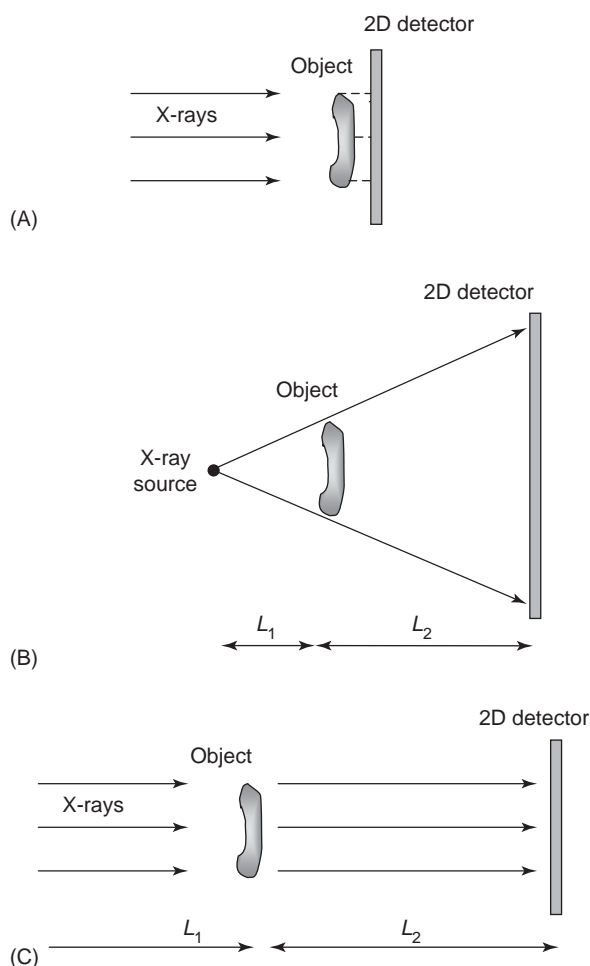
Refractive lenses act as a normal conventional lens and we can apply the Gauss lens formula, which relates the source distance  $L_1$ , the image distance  $L_2$ , and the focal distance  $F$  via  $L_2 = FL_1/(L_1 - F)$ . The diffraction-limited resolution of the lens  $\Delta$  is defined by an effective aperture:  $\Delta = 0.75\lambda/2NA$ , where the numerical aperture is  $NA = A_{\text{eff}}/2L_2$ .  $A_{\text{eff}}$  is the effective aperture of the lens, reduced by photon absorption and scattering, compared with the geometrical aperture  $2R_0$ . Nowadays, the parabolic refractive lenses, made of aluminum and beryllium, are available and widely used for hard X-ray microscopy applications providing resolution in the order of 300–500 nm.

## Full-Field Microscopy

### Microradiography

**X-ray contact microscopy** Soon after Roentgen's discovery of X-rays it was shown that a position-sensitive detector could detect the transmission map of the object with which it was in close contact (Figure 4A). Radiographs of small objects were produced and examined under the light microscope. It is obvious that, if a light microscope is used, the resolution of the resultant image is limited by the resolution of the optical microscope. Various improvements have been made including the use of higher-resolution crystalline or polymer resists instead of film, and the use of scanning electron microscope to examine or enlarge the image. With the development of X-ray lithography, a high-resolution recording medium such as poly(methyl methacrylate)





**Figure 4** X-ray microradiography: (A) contact microscopy, (B) projection microscopy, and (C) inline imaging.

(PMMA) was employed. In addition, atomic force microscopy is used for a more quantitative read-out of the PMMA. Since it is a two-step process – an image formation step on the photoresist followed by a magnification of the miniature contact print formed on the resist – it is difficult to use in the study of dynamic processes. The use of flash sources such as X-ray lasers or laser-produced plasmas in addition to the use of synchrotron radiation sources increased the activity of X-ray contact imaging, allowing the capture of the image on a nanosecond time scale.

Contact X-ray microradiography or X-ray contact microscopy (XCM) has been applied to the largest range of specimens due to its simple implementation and considerable versatility. The biological sciences are the most important field of applications for XCM. The nonlinearity of resists makes the resist-based contact microscopy unsuitable for quantitative imaging, which considerably limits its application to materials science. However, quite opposite to the

resist, the photographic plates as recording media with XCM have been largely used for elemental mapping obtained by the differential technique and quantitative measurements.

As an alternative to photoresist recording, a contact microscope was developed, where a photoemissive surface is used instead of a photoresist detector. The electrons from the photoemitter pass through an electron optical column to yield a magnified image. For most specimens, however, a large loss in the resolution arises from the fact that not all parts of the specimen of finite thickness are in close contact with the detector, leading to blurring of the microradiograph by diffraction.

**X-ray projection microscopy (XPM)** Conceptually, a projection microscope is quite simple and relies on shadow enlargement from a point source of X-rays (Figure 4B). Earlier an electron microscope type column produced a small and bright electron spot that in turn created an X-ray spot of sufficient brightness and small diameter (0.1–1  $\mu\text{m}$ ) to provide useful magnification of several hundred times. For soft X-rays, Fresnel diffraction limits the resolution; however, for hard X-rays, the image resolution is determined entirely by the X-ray emitting volume and the electron beam current. The usual resolution of these instruments was not less than 0.5  $\mu\text{m}$ .

To improve the resolution of XPM it was proposed to use as X-ray source a secondary source generated by optics able to strongly demagnify the image of the primary source. KB mirrors are often used for this purpose. A single ellipsoidal mirror can focus X-rays from one focus into the other focus. However, the requirements on the shape fidelity are high and difficult to achieve.

High-energy XPM (20–200 keV) with source of 10  $\mu\text{m}$  is widely used for industrial and medical applications. Industrial applications concern the control of products (castings, ceramics, electronic components) to detect microporosities and cracks. In materials science XPM is applied to study alloys, composite materials, and dynamic phenomena. In the medical sciences the applications of XPM concerns the study of cancer lesions, evaluation of bone mineral density, and teeth studies.

At medium and low energies (1–20 keV) projection microscopy was commercially established and used in materials and geological sciences and for biological applications. Nevertheless, the technique lost its popularity because of competition from the scanning electron microscope and electron probe analyzer, both of which offer X-ray elemental fluorescence maps superimposed on structural images. However, both of these instruments give information about the

near-surface layers, so they do not replace X-ray projection microscope's potential for three-dimensional imaging. Currently, active projection microscopes are based on simple modifications of a scanning electron microscope and are used in three-dimensional imaging experiments employing tomographic reconstruction techniques. Steady improvement in X-ray source technology over the last few years allows the manufacture of X-ray microtomographs (micro-CT) based on compact sealed X-ray microfocus tubes.

**Inline imaging** X-ray imaging at high energies such as contact radiography and projection microscopy and tomography has been used for many years to nondestructively discern the features of the internal structure of the objects in materials science, biology, and medicine. In so doing the main contrast formation is an absorption that makes some limitation in imaging of low-density materials. For imaging at micrometer and submicrometer resolutions there is no way to overcome these limits except by phase contrast imaging. At high X-ray energies the phase shift term can be up to 1000 times greater the absorption. Thus, it is possible to observe phase contrast when absorption contrast is undetectable. X-rays passing through regions of different  $\delta$  pick up different relative phases, which corresponds to being refracted and produces a distorted wave front. However, for imaging in phase contrast and to resolve phase variation across the beam a setup with a large lateral coherence length, which is favored by a small source size, is needed. For this reason, imaging in phase contrast was improved with the advent of the undulator sources at third-generation SR sources like ESRF, SPring 8, and APS. In the simple inline geometry (Figure 4C), an object is illuminated by a distant small X-ray source, which generates an almost parallel beam. A typical source-sample distance  $L_1$  is 30–50 m. The X-rays emerging from the sample at their various angles will propagate through the free space until they reach the detector. With the detector immediately behind the sample ( $L_2 = 0$ ), one will get a conventional absorption image. If the detector is placed at some distance behind the sample, we will observe the interference pattern as different components of the beam, which having been diffracted by the sample interfere with each other on further propagation through the space. Here, a combination of imaging and diffraction effects is found, typically involving interference fringes at the edges of features. These fringes improve edge visibility. The optimum positioning of the detector for best enhancement effects varies from sample to sample, depending on the X-ray wavelength and the size of

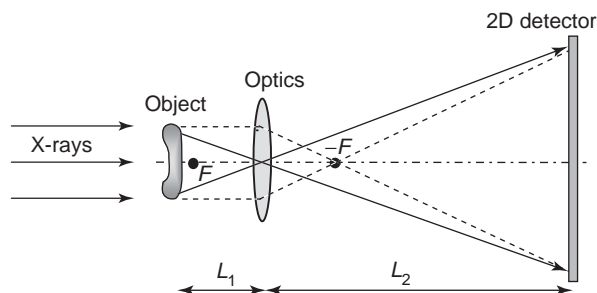
the features of interest. This inline phase-sensitive imaging is similar to the original Gabor inline holography techniques, and was first explored at the ESRF and now routinely used for imaging various kinds of samples. Since the beam is almost parallel, the resolution of this type of microscopy is determined by the detector resolution. Normally, a resolution is below  $1\mu\text{m}$ . However, it has the advantage of being simple since it needs no optics. When the two-dimensional projection microscopy is combined with tomography a three-dimensional structure determination can be achieved. For that purpose the sample has to be rotated in small steps around an axis perpendicular to the X-ray beam and for each step a micrograph is taken. Sophisticated tomography algorithms are available, which allow reconstructing the three-dimensional geometric structure of the sample.

The same spatial coherence available at synchrotrons can also be obtained with a microfocus X-ray tube with a very small source size similar to the projection microscope but with a significant source-sample distance ( $L_1 \sim 1\text{ m}$ ). It was shown that a monochromatic beam is not necessary and the polychromatic output of the X-ray tube can be used.

### Transmission X-Ray Microscope

The full-field transmission X-ray microscope (TXRM), illustrated schematically in Figure 5, uses the same optical arrangement as conventional light and transmission electron microscopes. The basic approach is to use a high-quality imaging optics as a microscope objective to create a magnified image of the object. When the object is placed slightly outside the focal distance ( $L_1$  is slightly larger than  $F$ ) then a strongly magnified image is generated at a distance  $L_2 = L_1 F / (L_1 - F)$  with magnification  $L_2 / L_1 = F / (L_1 - F)$ . This technique offers the possibility of dynamic imaging and can be adapted for phase contrast imaging.

The basic soft TXM approach is to use a condenser zone plate to monochromatize and to concentrate X-rays onto the sample. The microobjective zone



**Figure 5** Optical schematic of transmission X-ray microscope.

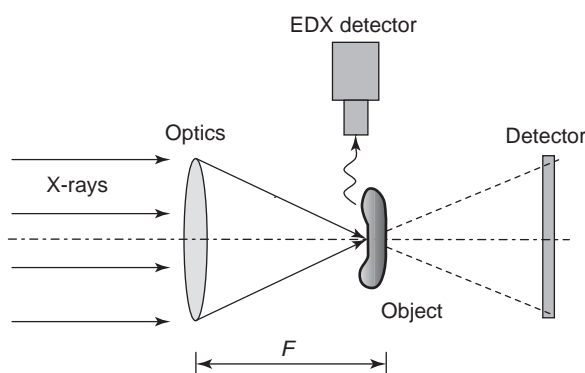
plate forms a magnified real image of the sample at roughly  $1000\times$  magnification. The enlarged image is recorded on a backside-thinned CCD camera. The sample is mounted in an environmental chamber where it can be prepared between two thin foils in its natural state, including when wet. The object chamber is in the atmospheric pressure, while the major parts of the X-ray optical path are in vacuum. Image exposure times range from under a second with dry specimens to a few seconds for wet specimens, depending on the zone plate used. Phase contrast in TXRM as a full analog of Zernike phase contrast in visible light microscopy has been realized. A removal ring aperture was placed near the condenser and a phase-ring was permanently mounted in the conjugate plane. The improved contrast allowed for reduced illumination and therefore reduced radiation dose to the specimen. A cryogenic object chamber has been developed and implemented at the TXRM. The cryogenic method allows imaging of chemically unfixed samples with resolution 20–30 nm. Furthermore, the high stability of frozen-hydrated samples allows taking multiple images for tomography.

Concerning hard X-rays there are two types of optics available today with sufficient quality to be used in the development of full-field imaging: CRLs and Fresnel zone plates. Both have in common that their focal length for X-rays with photon energies  $\sim 10$  keV and higher is of order of meters, and that their apertures are limited to several hundred micrometers. The length of the setup will increase as  $M$  increasingly differs from unity. The length of the entire microscope can be in the order of 10–25 m. Another implication of long focal length is a very small numerical aperture, which limits the resolution of the imaging setup. Resolution of  $\sim 100$ –300 nm is achievable.

### Scanning X-Ray Microscope

Finally, we consider X-ray microscopy in the scanning mode, illustrated in Figure 6. The basic idea in scanning microscopy is to form a microprobe across which the specimen is mechanically scanned. A proportional counter detects the transmitted X-rays and the image is built up pixel (picture element) by pixel. The focused X-ray probe can also be used to excite other process such as photoelectrons, fluorescent X-rays. The spatial resolution is limited by the focusing optics.

**Soft X-ray transmission scanning microscope** Scanning transmission X-ray microscopes (STXRM) for soft X-rays use Fresnel zone plates as high-resolution objectives to form a focused spot. Scanning microscopes require coherent illumination and are



**Figure 6** Schematic of scanning transmission X-ray microscope.

used with high brightness sources such as undulators at synchrotron storage rings. Scanning microscopes generally have exposure times of minutes; each pixel can be formed in a few milliseconds using synchrotron radiation. But such types of microscopes impact 5–10 times less radiation dose to the specimen and have energy resolution of 0.1 eV. STXRMs are better suited to spectromicroscopy. Two main types of scanning microscopes exist today: STXRMs equipped with photon detectors and scanning photoemission microscopes equipped with electron detectors.

**Hard X-ray scanning microprobe** In the case of hard X-rays a microfocus is generated by KB mirrors, Fresnel zone plate, or parabolic CRL. The high-energy microfocusing optics benefits from longer focal lengths and larger depth of field up to a few millimeters, which is advantageous for the use of specific sample environment like furnace or high-pressure cells. A shorter wavelength is favorable for diffraction studies including wide and small angle scattering. In microprobe the strategy is to scan the beam over the sample and to measure a signal in diffraction, fluorescence, or absorption (XANES, EXAFS) for each beam position. When combining scanning microscopy with tomographic techniques the inner structure of a sample can be reconstructed, including the distribution of different atomic species and even of the valence of atoms.

*See also:* **Microscopy:** Overview. **Microscopy Techniques:** Light Microscopy; Electron Microscopy; Scanning Electron Microscopy.

### Further Reading

Attwood D (1999) *Soft X-Rays and Extreme Ultraviolet Radiation. Principles and Applications*. Cambridge: Cambridge University Press.

- Janssens KHA, Adams FCV, and Rindby A (2000) *Microscopic X-Ray Fluorescence Analysis*. New York: Wiley.
- Larson BC and Lengeler B (eds.) (2004) High resolution three-dimensional X-Ray microscopy. *Materials Research Society Bulletin* 29(3): 152-181.
- Lengeler B, Schroer CG, Tümmeler J, *et al.* (2001) Refractive X-Ray optics. In: Bass M, Enoch JM, Van Stryland EW, and Wolfe L (eds.) *Handbook of Optics, Vol. III, Classical Optics, Vision Optics, X-Ray Optics*, ch. 20. New York: McGraw-Hill.
- Meyer-Ilse W, Warwick T, and Attwood D (eds.) (2000). *X-Ray Microscopy Proceedings of the Sixth International Conference*, Berkeley, CA, 1999, AIP Conference Proceedings 507. Melville, New York: American Institute of Physics.
- Michette A and Pfauntsch S (eds.) (1997) *X-Rays The First Hundred Years*. New York: Wiley.
- Michette AG (1986) *Optical Systems for Soft X-Rays*. New York: Plenum.
- Susini J, Joyeux D, and Polack F (eds.) (2003) *X-Ray Microscopy 2002, 7th International Conference on X-Ray Microscopy*, ESRF, Grenoble, France, Jul. 28–Aug. 2, 2002. EDP Sciences, *Journal de Physique IV*, vol. 104.
- Thieme J, Schmahl G, Rudolph D, and Umbach E (eds.) (1998) *X-Ray Microscopy and Spectroscopy*. Status Report from the Fifth International Conference, Würzburg, Aug. 19–23, 1996. Berlin: Springer.

## MICROWAVE-ASSISTED DIGESTION

See **SAMPLE DISSOLUTION FOR ELEMENTAL ANALYSIS: Microwave Digestion**

## MICROWAVE-ASSISTED SOLVENT EXTRACTION

See **EXTRACTION: Microwave-Assisted Solvent Extraction**

## MICROWAVE SPECTROSCOPY

**J G Baker**, The University of Manchester, Manchester, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Microwave spectroscopic techniques refer to the high-resolution absorption spectroscopy of rotational transitions of polar molecules in the gas phase. At sufficiently low pressures, of the order of a few microbars, individual transitions between quantum states become clearly separated, and their metrology using well-established radiofrequency and microwave counting methods is quite straightforward. Frequency measurements to a precision of 1 part in  $10^8$  are routine, and to several further orders of magnitude become possible with careful attention to instrumental design. As a consequence of this precision, molecular rotational spectroscopy (MRR) is highly selective, and it has been declared on many occasions that a sample compound and its isotopomers or conformers may be characterized by measurement of only a very few spectral lines, even

when the species is present in concentrations as low as nanomoles per mole, without destruction of the sample.

It will be seen in this article that the subject, written off by many as a mature technology only two decades ago, has undergone a dramatic resurgence thanks to the advent of cavity and jet spectrometry, and the ready availability of solid state and other sources together with computer control. Indeed, its characteristic use of coherent spectroscopic techniques has spread into the infrared and optical regions thanks to the production of frequency-stable diode lasers through the near- and mid-infrared, and the accessibility of data at optical wavelengths through the use of harmonic generation and the development of sensitive detectors of molecular ions.

### The Origin of Microwave Rotational Resonance Spectra

With the exception of diatomic hydrides, whose absorptions fall well into the THz region, the great majority of polar molecules show MRR spectra in

- Janssens KHA, Adams FCV, and Rindby A (2000) *Microscopic X-Ray Fluorescence Analysis*. New York: Wiley.
- Larson BC and Lengeler B (eds.) (2004) High resolution three-dimensional X-Ray microscopy. *Materials Research Society Bulletin* 29(3): 152-181.
- Lengeler B, Schroer CG, Tümmeler J, *et al.* (2001) Refractive X-Ray optics. In: Bass M, Enoch JM, Van Stryland EW, and Wolfe L (eds.) *Handbook of Optics, Vol. III, Classical Optics, Vision Optics, X-Ray Optics*, ch. 20. New York: McGraw-Hill.
- Meyer-Ilse W, Warwick T, and Attwood D (eds.) (2000). *X-Ray Microscopy Proceedings of the Sixth International Conference*, Berkeley, CA, 1999, AIP Conference Proceedings 507. Melville, New York: American Institute of Physics.
- Michette A and Pfauntsch S (eds.) (1997) *X-Rays The First Hundred Years*. New York: Wiley.
- Michette AG (1986) *Optical Systems for Soft X-Rays*. New York: Plenum.
- Susini J, Joyeux D, and Polack F (eds.) (2003) *X-Ray Microscopy 2002, 7th International Conference on X-Ray Microscopy*, ESRF, Grenoble, France, Jul. 28–Aug. 2, 2002. EDP Sciences, *Journal de Physique IV*, vol. 104.
- Thieme J, Schmahl G, Rudolph D, and Umbach E (eds.) (1998) *X-Ray Microscopy and Spectroscopy*. Status Report from the Fifth International Conference, Würzburg, Aug. 19–23, 1996. Berlin: Springer.

## MICROWAVE-ASSISTED DIGESTION

See **SAMPLE DISSOLUTION FOR ELEMENTAL ANALYSIS: Microwave Digestion**

## MICROWAVE-ASSISTED SOLVENT EXTRACTION

See **EXTRACTION: Microwave-Assisted Solvent Extraction**

## MICROWAVE SPECTROSCOPY

**J G Baker**, The University of Manchester, Manchester, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Microwave spectroscopic techniques refer to the high-resolution absorption spectroscopy of rotational transitions of polar molecules in the gas phase. At sufficiently low pressures, of the order of a few microbars, individual transitions between quantum states become clearly separated, and their metrology using well-established radiofrequency and microwave counting methods is quite straightforward. Frequency measurements to a precision of 1 part in  $10^8$  are routine, and to several further orders of magnitude become possible with careful attention to instrumental design. As a consequence of this precision, molecular rotational spectroscopy (MRR) is highly selective, and it has been declared on many occasions that a sample compound and its isotopomers or conformers may be characterized by measurement of only a very few spectral lines, even

when the species is present in concentrations as low as nanomoles per mole, without destruction of the sample.

It will be seen in this article that the subject, written off by many as a mature technology only two decades ago, has undergone a dramatic resurgence thanks to the advent of cavity and jet spectrometry, and the ready availability of solid state and other sources together with computer control. Indeed, its characteristic use of coherent spectroscopic techniques has spread into the infrared and optical regions thanks to the production of frequency-stable diode lasers through the near- and mid-infrared, and the accessibility of data at optical wavelengths through the use of harmonic generation and the development of sensitive detectors of molecular ions.

### The Origin of Microwave Rotational Resonance Spectra

With the exception of diatomic hydrides, whose absorptions fall well into the THz region, the great majority of polar molecules show MRR spectra in



the band from 3 GHz to 1 THz. The frequency of strongest absorption varies with molecular weight and temperature, but for many species of molecular weight up to 100 a useful observing range lies between 30 and 300 GHz. Nowadays this range is well covered by commercial computer controlled synthesized frequency generators of the requisite stability, though their power output falls and their cost escalates rapidly with increasing frequency.

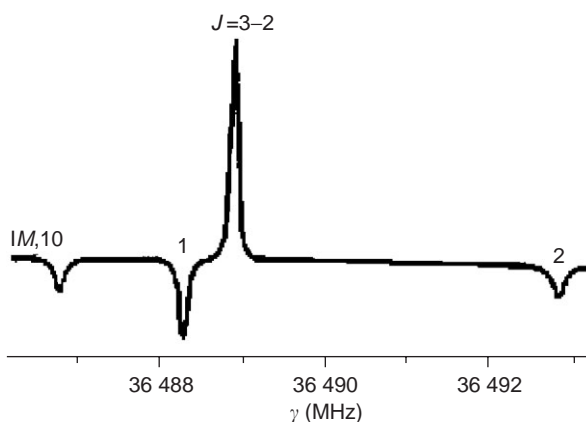
The intensities of typical absorptions for pure samples fall in the range  $1\text{--}10^{-8}\text{ m}^{-1}$ , much smaller than is the case for infrared or optical spectra. This is due to a number of reasons:

1. The population difference between the quantum states observed is markedly smaller than that of either alone.
2. Each state is split into numerous components by the effects of nuclear hyperfine structure, vibrational excitation, and molecular isomerism.
3. The inherent intensity of radiative transitions varies linearly with their frequency, so that at shorter wavelengths they are inevitably more intense.
4. The molecular density contributing to an absorption at microbar pressures is much less than that available for atmospheric pressure or condensed samples.

For these reasons, it is rarely practical to observe microwave or millimeterwave (mmwave) absorptions by simply following the variation of sample transmission with frequency. Rather, one seeks to modulate the absorption by switching it on and off at an audio- or radiofrequency, and to detect any resulting transmitted modulation by phase coherent detection.

In the earlier decades of microwave spectroscopy, and in all commercial spectrometers on the market at that time, the method of choice was that of Stark modulation. A uniform electric field was applied to the entire sample and switched off and on, typically at a 33 kHz rate. The resulting phase coherently detected spectral frequency scan (Figure 1) shows a spectral line in phase with the switching voltage and a number of electric-field shifted (Stark) components in antiphase with it.

Scans like these are very useful in enabling researchers to determine the molecular dipole moment as well as assigning the quantum states and positions of spectral lines. But, despite the value of such a spectrometer as a research tool, it has proved unpopular for industrial use because the high electric fields needed, up to  $0.4\text{ MV m}^{-1}$ , can produce sparking in the sample at millibar pressures leading both to sample deterioration and to an explosion hazard.



**Figure 1** A phase coherently detected Stark modulation spectrum of the  $J=3\leftarrow 2$  transition in OCS at 36488.8130 MHz. Three Stark components (negative) appear at varying distances from the unshifted spectral line (positive). (Reproduced with permission from Hollas JM (1992) *High Resolution Spectroscopy*, figure 4.13, p. 103. Chichester: Wiley; © John Wiley & Sons Ltd.)

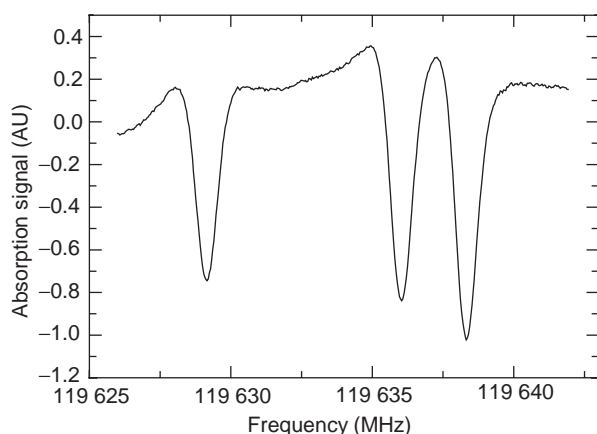
Furthermore, the specialist absorption cells needed cannot readily be extended to give longer transit paths through the sample, and the magnitude of the Stark shift produced by a given field falls off rapidly at shorter millimeter wavelengths, greatly reducing the value of such a spectrometer in detecting and resolving lines in this more sensitive band.

Most modern general purpose spectrometers use frequency modulation, in which the source frequency is driven back and forth at an audio or higher frequency with sufficient deviation to encompass several spectral line widths. In principle, this too should reveal a coherently detected spectral profile, though usually of first- or second-derivative form, as the source is tracked through the line width. In practice, there are many new problems due to nonuniform transmission of radiation through the sample caused by discontinuities and reflections in the radiation path, as shown in Figures 2 and 3. However, these are more easily dealt with and eliminated within a computer controlled system, whilst the advantages of being able to use a hazard-free setup in which radiation may be transmitted through the atmosphere or through sequential samples are very great.

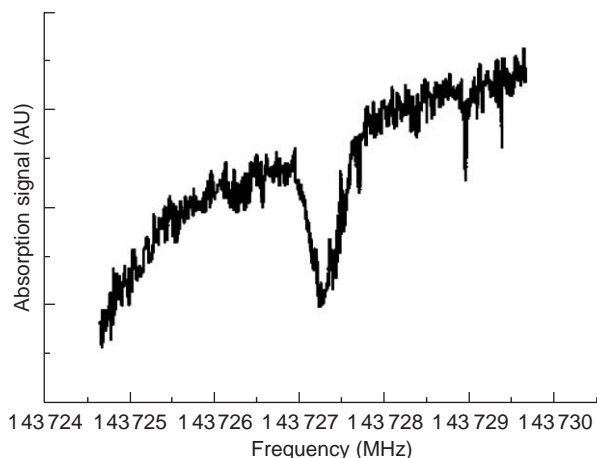
Other more specialist modulation techniques have been described in the literature; and will be discussed as appropriate in a later section.

## Appearance of Molecular Rotational Spectra

The simplest spectra are those generated from the ground vibrational state of a linear molecule. These



**Figure 2** A frequency modulation spectrum of  $K=2$ ,  $K=1$ , and  $K=0$  components of the  $J=7 \leftarrow 6$  transition of propyne,  $\text{CH}_3\text{CCH}$ , at 119.635 MHz, phase coherently detected at twice the modulation frequency, and obtained with a synchronously tuned Fabry–Perot cavity cell. (Courtesy of Wilks AT, with permission.)



**Figure 3** A frequency modulation spectrum of the  $4_{22} \leftarrow 4_{23}$  transition of HDO near 143 727 MHz, phase coherently detected at twice the modulation frequency and obtained with a synchronously tuned Fabry–Perot cavity cell. This species was present in natural abundance (270 ppm) in a sample of water vapor. The sloping background is due to the nonuniform coupling of the cavity, and may be eliminated by computer analysis. (From Alder JF and Baker JG (2002) *Quantitative Millimetre Wavelength Spectroscopy*. Cambridge: Royal Society of Chemistry; reproduced by permission of the Royal Society of Chemistry.)

consist of a ladder of almost harmonic frequencies of values  $\nu$  given by the formula:

$$\nu/(J+1) = 2B - 4D(J+1)^2$$

$J$  represents the rotational quantum number of the lower state of the absorption, and can take any integer value from zero upward.  $B$ , the rotational constant, is related to the molecular moment of

inertia  $I$  about an axis perpendicular to the linear frame by the formula:

$$B = h/8\pi^2 I \text{ (Hz)} = h/8\pi^2 I c \text{ (cm}^{-1}\text{)}$$

where  $c$  is the speed of light in  $\text{cm s}^{-1}$ .

Typical values for  $B$  can range through:

- CO 57.635 97 GHz
- HCN 44.315 80 GHz
- OCS 6.081 48 GHz

In addition, there is a centrifugal stretching contribution given by the term in  $D$ , which is some six orders of magnitude smaller than  $B$ , but which removes the exact harmonic relationship between the lines of differing  $J$  values. The practical result is that for all but the heaviest molecules only a few transitions will fall into a single conveniently accessible radiation frequency band.

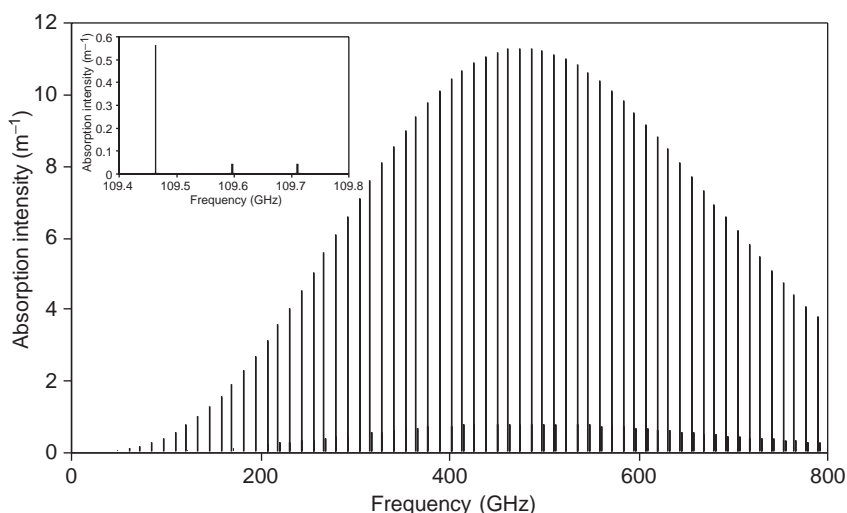
The peak absorption coefficient of the transition,  $\gamma_{\text{max}}$ , if Doppler broadening is neglected, is given by

$$\gamma_{\text{max}} = f\mu^2\nu^2(1 - \exp(-h\nu/k_B T))N/\Delta\nu$$

where  $f$  is the fraction of molecules in the lower state,  $\mu$  the molecular dipole moment,  $N$  the molecular density in the sample,  $\Delta\nu$  the half-line-width (HWHM) at this density,  $h$  is Planck's constant,  $k_B$  is Boltzmann's constant, and  $T$  the temperature. Because both  $N$  and  $\Delta\nu$  are separately proportional to the pressure, this peak absorption is independent of the sample pressure over a very wide range, from a few microbars to bars. Furthermore, because  $f$  depends both on the system degeneracy  $(2J+1)$  and the thermal Boltzmann factor for the lower state, it may readily be shown that this absorption becomes largest at a frequency  $\nu_{\text{max}} = (6Bck_B T)^{1/2}$ , as illustrated in a spectral stick diagram for OCS in **Figure 4**.

At a temperature of 290 K, the maximum absorption signals for CO, HCN, and OCS occur, respectively, at 1.5 THz, 1.2 THz, and 480 GHz. Only for much heavier species would these peak frequencies fall into a centimeter wavelength band.

However, this is not the only signal that may be observed in these bands. Linear triatomic and polyatomic molecules also show skeletal bending vibrations whose frequencies are sufficiently low that they are appreciably populated at room temperatures. They form l-type doublets centered round a frequency close to but not coincident with the ground state transition. So even in the simplest case, as shown in the inset in **Figure 4** for OCS, there will appear a minimum of three spectral lines, and possibly more,



**Figure 4** A stick diagram representing the positions and intensities of the spectral lines of OCS at 290 K through the GHz and THz bands. The strongest component occurs near 480 GHz. (Inset) An expanded version representing the  $11 \leftarrow 10$  transition, and showing a pair of vibrationally excited l-type doublets shifted to frequencies above the ground state. (Courtesy of Wilks AT, with permission.)

arising from further excited vibrational or electronic states.

In nonlinear molecules spectral patterns become both more profuse and more complicated due to the existence of molecular asymmetry and the possibility of dipole moment components along any of or all three inertial axes. The reader is recommended to an appropriate text book for details of these. However, databases tabulating transition frequencies and line strengths for most abundant species of atmospheric and astrochemical interest, their isotopomers, and their low-lying vibrational states are publicly available on the internet:

- NASA database: <http://spec.jpl.nasa.gov>
- CDMS database: <http://www.ph1.uni-koeln.de/vorhersagen>
- HITRAN database: <http://cfa-www.harvard.edu.-HITRAN/>

and it should only exceptionally be necessary to carry out such calculations for oneself rather than by consulting these databases and their custodians.

## Molecular Geometries from Rotational Spectra

It is nowadays quite unusual to need to relate a molecular geometry to its inertial constants from first principles. When this is essential, procedures using the Kraitichman isotopic substitution equations, described in the reference texts, tend to give best results. However, extensive material describing structures obtained by electron diffraction as well

as by rotational spectroscopy is collated in inexpensive reference databases now available on the internet:

- MOGADOC database: <http://www.uni-ulm.de/strudo/strudo.html>

## New Techniques and Instrumentation

### Quantitative Fabry–Perot Cavity Spectrometry

A popular method for extending the path traversed by radiation through the sample is to place it between the mirrors of a confocal Fabry–Perot interferometer. Such a device allied with high-reflectivity mirrors can deliver traverses of several kilometers at frequencies ranging from tens of gigahertz to terahertz. The presence of a sample increases the cavity half-width by an amount proportional to its attenuation factor and such behavior has been shown to be capable of monitoring over a very wide frequency band by jumping between successive resonances of the interferometer. This half-width is most usually measured by carrying out and fitting frequency scans over one or several of these resonances, but an alternative procedure has been demonstrated in which the interferometer is kept resonant by servo-locking the mirror spacing to match the source frequency, and sample data are taken by scanning the depth of a frequency modulation applied to the source. Such a procedure is well adapted to the use of compact solid-state mmwave sources by obviating the need for wideband frequency scans, and has been shown to be robust and effective in monitoring samples of acrylonitrile, propyne, sulfur dioxide, oxygen, and water

present in ambient air at parts per thousand concentration. The absolute absorption coefficient is determined at the chosen source frequency, so that spectral profiles may be obtained simply by scanning this value, and the system has been shown to be capable of routine and trouble free operation in the quantitative analysis of gas mixtures at ambient pressures up to and greater than atmospheric.

The versatility of the system described above has recently been enhanced by transmitting the exciting microwave radiation as modulation on an infrared laser communications link over a fiber-optic path up to 1 km, before demodulation and harmonic generation produces a mmwave signal at the far end that drives the Fabry–Perot cell. It then acts as a mobile remote analytical spectrometer that can be moved to any convenient location whilst remaining under full control from and returning results continuously to home base.

### **Microwave–Microwave and mmwave Double Resonance**

A powerful technique for identifying and assigning spectral lines is that of double resonance, in which a sample transition is pumped by an intense amplitude modulated source, at a low frequency, and a second probed by a weaker secondary source at a higher frequency. If the two transitions have a level in common, the population difference contributing to the intensity of the probe transmission will be modulated at the frequency of the pump modulation. This is an exceedingly useful tool for disentangling complex overlapping spectra, as shown in Figure 5.

Even though the induced population changes may be quite small, the resonant pump also generates a coherence splitting of the common level that drives the probe transition transparent. So 100% amplitude modulation of the pump may be transferred to the probe when both are resonant, clearly identifying the connected systems.

The potential of this technique for the unequivocal identification of the relative abundances of constituents of a complex mixture of isotopomers, conformers, and vibrationally excited states has been remarkably little exploited. It can and has been used extensively for identifying transitions by microwave–mmwave and microwave–infrared double resonance, but not in a systematic manner.

### **Fourier-Transform Spectrometry**

The technique of applying an intense but short pulse of radiation to a sample, and then monitoring the time response of the resulting emission, has become both successful and ubiquitous in microwave

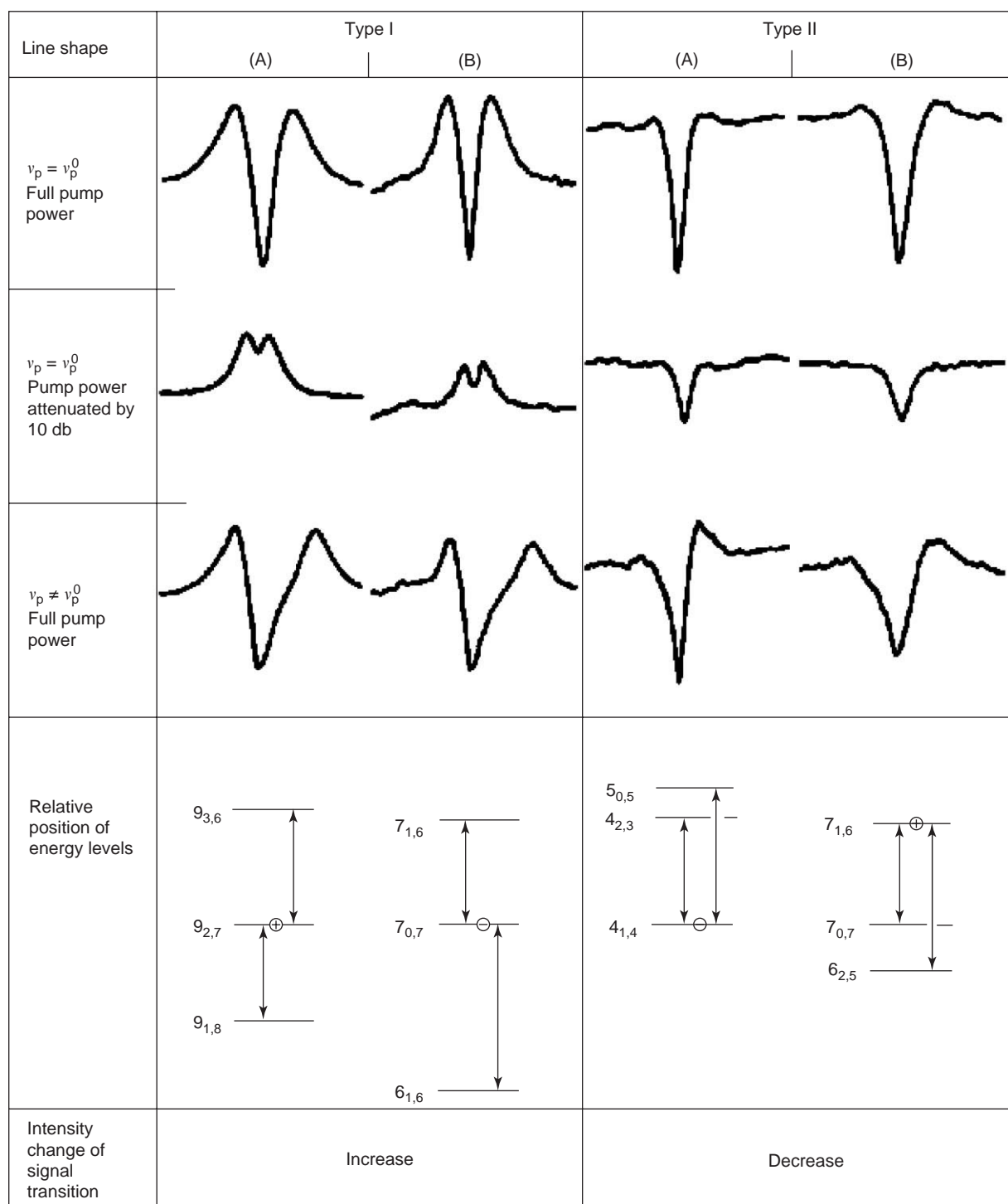
spectroscopy. The principle involved is that a short pulse of duration  $\tau$  excites all accessible transitions within a bandwidth  $1/2\pi\tau$  surrounding the pulse frequency, and that, when it is switched off, each excited state radiates coherently and separately at its own resonant frequency until its signal becomes destroyed through collision or inhomogeneous broadening effects. This is often detected by beating it with a much reduced version of the polarizing signal, after which the consequent time-resolved oscillating signal is digitized and Fourier transformed to yield the spectrum of the sample in the frequency domain.

As in the case of pulsed nuclear magnetic resonance spectrometry, this generates a signal that typically spans a range of megahertz, presents the multiplex advantage of covering many spectral features simultaneously, and can be repeated for as many shots as is necessary. But, because a typical signal decay rate lies in the region of 10 ps bar, one must work at low sample pressures using high-speed electronics. Nonetheless, the sensitivity and convenience of such machines has resulted in their steadily replacing more conventional instruments as laboratories have renewed their equipment over the past decade. Some recent published detection limits for a Fourier transform analytical spectrometer are listed in Table 1.

Extension of this technique into the more sensitive mmwave band has been foiled by a lack of broadband switches and isolators capable of functioning in this band. Attempts have been made to overcome this problem by means of two-dimensional pulsed double resonance, in which a low-frequency pump is isolated from the detector by a cutoff waveguide and the response of the probe to successive switches is digitized. Vogelsanger and Bauder have generated a two-dimensional frequency domain signal, from pump and probe, whilst an alternative procedure, in which the response to two precisely spaced pump pulses is digitized as a function of this spacing, has been described by Bowring *et al.* It enabled them to obtain Fourier transform spectra in C<sub>12</sub>N spanning as much as 75 MHz, and displaying many other nonlinear Raman beats between hyperfine components. Neither technique has seen development for chemical analysis, despite retaining the characteristic double resonance feature of distinguishing between components of complex mixtures.

### **Jet Spectroscopy**

A problem already alluded to and characteristic of rotational spectroscopy as a whole is the splitting of spectral features into many components originating in excited rotational and vibrational states. These



**Figure 5** Frequency-modulated double resonance profiles for various pumping schemes in *cis*-propionic acid. The sign displayed on the common energy level indicates whether the pump causes an increase or decrease in the population of that level. (Reproduced with permission from Stiefvater OL (1975) *Z Naturforsch* 30a: 1800.)

can largely be eliminated by expanding the sample, mixed with inert Ar, He, or Ne gas at bar pressures, through a nozzle into a high vacuum to form a highly directed supersonic jet. The expansion cools translational degrees of freedom to  $\sim 1$  K, thus greatly

reducing Doppler broadening, and rotational degrees of freedom to  $\sim 10$  K, with the result that higher  $J$  rotational transitions disappear, and lower  $J$  ones become greatly enhanced. Lower frequency vibrational states, too, tend to become depopulated.



**Table 1** Detection limits for various compounds using Fourier-transform millimeterwave spectroscopy

Compound	Formula	Mds <sup>a</sup> (nmol mol <sup>-1</sup> )
Acrolein	CH <sub>2</sub> CHCHO	0.5
Carbonyl sulfide	OCS	1
Sulfur dioxide	SO <sub>2</sub>	4
Propionaldehyde	CH <sub>3</sub> CH <sub>2</sub> CHO	100
Methyl- <i>t</i> -butyl ether	CH <sub>3</sub> OC(CH <sub>3</sub> ) <sub>3</sub>	65
Vinyl chloride	CH <sub>2</sub> CHCl	0.45
Ethyl chloride	CH <sub>3</sub> CH <sub>2</sub> Cl	2
Vinyl bromide	CH <sub>2</sub> CHBr	1
Toluene	CH <sub>3</sub> C <sub>6</sub> H <sub>5</sub>	130
Vinyl cyanide	CH <sub>2</sub> CHCN	0.28
Acetaldehyde	CH <sub>3</sub> CHO	1
Propylene oxide	CH <sub>3</sub> CHCH <sub>2</sub> O	11
<i>p</i> -Tolualdehyde	CH <sub>3</sub> C <sub>6</sub> H <sub>4</sub> CHO	150
Methanol	CH <sub>3</sub> OH	1000
Benzaldehyde	C <sub>6</sub> H <sub>5</sub> CHO	26

<sup>a</sup>Minimum detectable signal based on a 1 min average of accumulated pulses using Ne as the carrier gas.

From Suenram RG, Grabow J, Zuban A, and Leonov I (1999) A portable pulsed-molecular-beam Fourier-transform microwave spectrometer designed for chemical analysis. *Review of Scientific Instruments* 70: 2127–2135.

Because the nozzle produces a diverging beam, it is often the case that the spectra seen are split into two or more components, but this instrumental artifact is easily dealt with by computerized analysis.

Various types of nozzle have been deployed, from a simple hole to a shaped one that promotes cluster formation. It is also a common procedure to drive the nozzle through a switched magnetic valve to generate sample pulses that match the switching time of the microwave pulses. In this way one can attain higher sample densities whilst maintaining the same average pumping rate.

Although many van der Waals and cluster species of the utmost chemical interest have been characterized with such jet instruments, uncertainties in the sample density and the degree of cooling attained in an individual jet do not readily lend themselves to reproducible quantitative analysis without careful internal precalibration procedures.

### Spectroscopy of Biological Precursor Molecules

Until quite recently it was thought that such molecules were not amenable to gas-phase spectroscopic analysis, due to their involatility, their nonrigid bonding, and the presence of large numbers of overlapping conformational and vibrational states. Jet spectroscopy has changed all that. It has proved readily possible to vaporize such species at temperatures of a few hundred degree Celsius into a cool inert gas stream and to study them in a supersonic expansion. Pioneering work of this kind was carried out by Brown and Godfrey for a couple of decades

using a specially designed continuous beam Stark spectrometer, but in recent years many other groups have turned to using Fourier transform techniques with great success in determining both the nature of conformational isomers present and their detailed structures.

As an illustration, one may turn to a recent study of the alkaloid drug amphetamine. This is a phenyl- and amino-substituted propane with a total of nine possible configurations, listed in Figure 6.

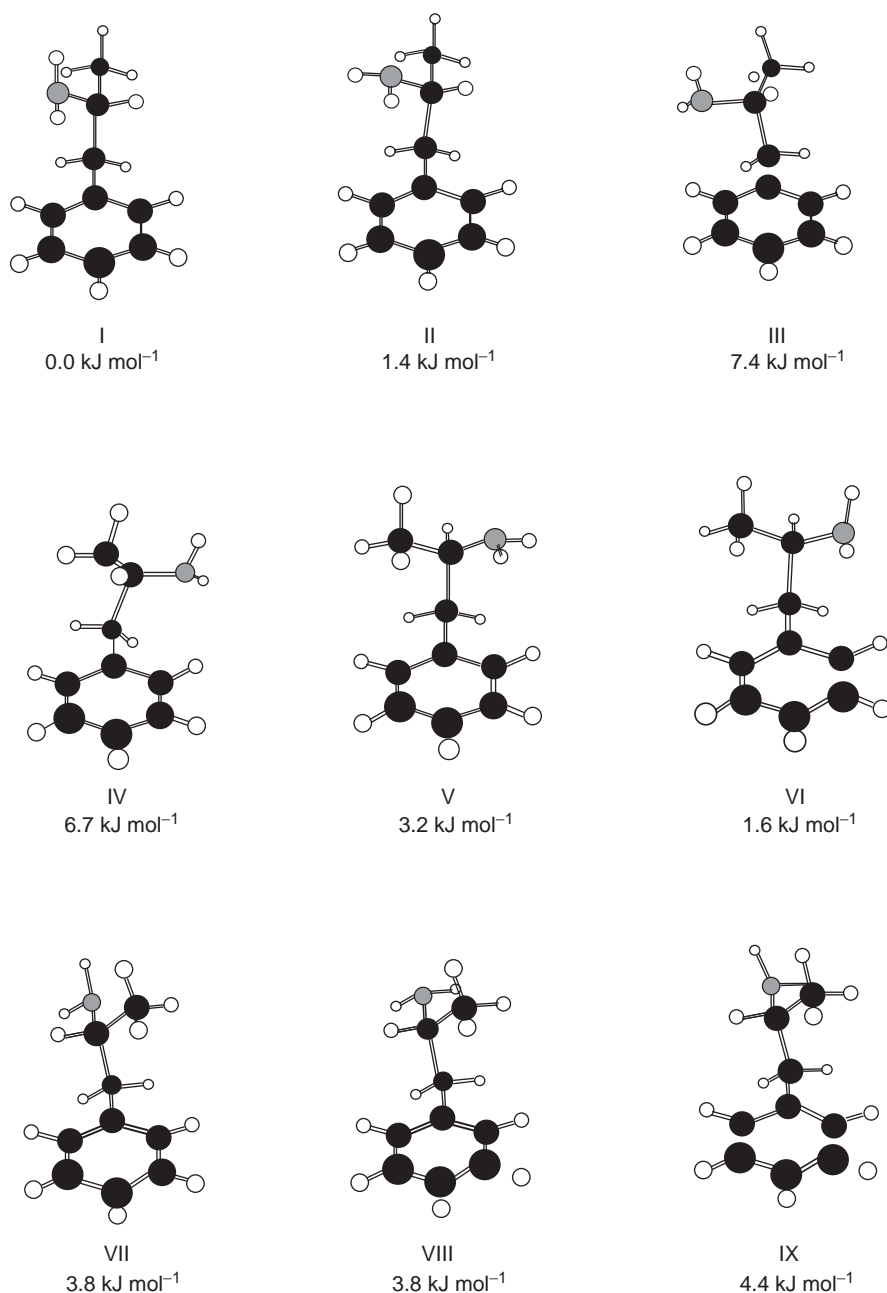
The study shows that only two of these are detectable in a jet expansion (I and II), and both of them correspond to an extended rather than a folded carbon backbone chain. In contrast, X-ray diffraction structures of the same molecule in the condensed phase predict a folded chain.

### Astrochemistry

One of the greatest triumphs for microwave spectroscopy has been the detection and characterization of over 130 molecules and ions present in interstellar gas clouds, most recently the simplest amino acid glycine. Indeed, the Nobel Prize winning discovery by Curl, Kroto, and Smalley of C<sub>60</sub>, buckminsterfullerene, arose directly from microwave studies of long chain interstellar acetylenes. In 2007, a new European satellite, HERSCHEL, sensitive to signals in the terahertz band, will be launched. Who can doubt that it will give rise to many more surprising revelations about the interstellar medium?

### Analytical Chemistry

Despite the dramatic successes of the techniques described in discovering new molecular species, they have seen little success and no commercial fallout in analytical spectroscopy. Amongst the reasons that may be adduced are the difficulties in obtaining reproducible lineshape integrals in cavity spectrometry and the uncertainties in sample concentrations and temperatures in jet spectrometry. Recent work in the groups of Alder and Baker and of Krupnov has, however, demonstrated that absolute line intensities may be determined both accurately and reproducibly in a cavity spectrometer by following the effect of a sample on the cavity quality factor. Their procedure is equivalent to that of 'cavity ringdown spectroscopy' now widely used in infrared spectrometry for spectral intensity measurements. Not only does this permit of simultaneous data collection for several species, but the hyperfine and vibrational structure may be 'smeared out' by operating at pressures of 1–100 mbar, thereby enhancing the sensitivity to levels comparable with those achieved by gas chromatography. This raises the real prospect of a



**Figure 6** The nine possible configurations and calculated energies of conformers of amphetamine. Only configurations I and II appear in the observed jet spectrum. (Reprinted with permission from Godfrey PD, McGlone SJ, and Brown RD (2001) The shapes of neurotransmitters by millimetrewave spectroscopy: Amphetamine. *Journal of Molecular Structure* 599: 139–152; © Elsevier.)

commercially viable instrument capable of online sample analysis at ambient pressures without any need for complex precalibration procedures.

**See also:** **Air Analysis:** Sampling; Outdoor Air; Work-place Air. **Fourier Transform Techniques.** **Nuclear Magnetic Resonance Spectroscopy Techniques:** Solid-State; Surface Coil; *In Vivo* Spectroscopy Using Localization Techniques. **Nuclear Magnetic Resonance Spectroscopy:** Instrumentation.

## Further Reading

- Baker JG (1979) Microwave–microwave double resonance. In: Chantry GW (ed.) *Modern Aspects of Microwave Spectroscopy*, pp. 65–122. London: Academic Press.
- Bowring NJ, Andrews DA, and Baker JG (1992) Novel coherence effects in microwave–microwave double resonance pulse Fourier transform spectroscopy. *Journal of Physics B: Atomic, Molecular and Optical Physics* 25: 667–677.

- Hollas JM (2002) *Basic Atomic and Molecular Spectroscopy*. Cambridge: Royal Society of Chemistry.
- Hollas JM (2003) *Modern Spectroscopy*, 4th edn. Chichester: Wiley.
- Krupnov AE, Tretyakov MY, Parshin VV, Shanin VN, and Myasnikova SE (2000) Modern millimeter-wave resonator spectroscopy of broad lines. *Journal of Molecular Spectroscopy* 202: 107–115.
- Suenram RG, Grabow JU, Zuban A, and Leonov I (1999) A portable, pulsed-molecular-beam, Fourier transform microwave spectrometer designed for

chemical analysis. *Review of Scientific Instruments* 70: 2127–2135.

- Vogelsanger B and Bauder A (1990) Two-dimensional microwave Fourier transform spectroscopy. *Journal of Chemical Physics* 92: 4101–4114.
- Wilks AT, Baker JG, Thompson M, and Alder JF (2002) Quantitative millimetre wavelength spectroscopy at pressures approaching atmospheric: I. Absolute determination of transition line strengths and sample concentrations with a Fabry–Perot cavity spectrometer. *Analytica Chimica Acta* 466: 57–67.

## MICROWAVE-INDUCED PLASMA

See ATOMIC EMISSION SPECTROMETRY: Microwave-Induced Plasma

## MILK

See FOOD AND NUTRITIONAL ANALYSIS: Dairy Products

## MIP

See ATOMIC EMISSION SPECTROMETRY: Microwave-Induced Plasma

## MIPs

See MOLECULARLY IMPRINTED POLYMERS

## MOLECULAR EMISSION CAVITY ANALYSIS

**A C Calokerinos**, University of Athens, Athens, Greece  
© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Molecular emission cavity analysis (MECA) is a flame chemiluminescence technique based on the generation of excited molecules, radicals, or atoms within a hydrogen diffusion flame. The excited species are formed by direct or indirect chemiluminescence mechanisms and are confined within the inner space of a small cavity, which is positioned at a pre-selected point of the flame environment. The emission is monitored at the characteristic wavelength of

each excited species and the intensity is proportional to the concentration of analyte in the sample. The cavity is responsible for the distinct characteristics of MECA over other emission techniques into which the sample is aspirated into the flame. Hence, the cavity is the cell for the chemiluminescence reactions and promotes excitation due to chemiluminescent rather than thermal processes.

### Basic Principles

#### Hydrogen Diffusion Flame

The hydrogen diffusion flame is normally used in MECA. The flame is formed by the combustion of

- Hollas JM (2002) *Basic Atomic and Molecular Spectroscopy*. Cambridge: Royal Society of Chemistry.
- Hollas JM (2003) *Modern Spectroscopy*, 4th edn. Chichester: Wiley.
- Krupnov AE, Tretyakov MY, Parshin VV, Shanin VN, and Myasnikova SE (2000) Modern millimeter-wave resonator spectroscopy of broad lines. *Journal of Molecular Spectroscopy* 202: 107–115.
- Suenram RG, Grabow JU, Zuban A, and Leonov I (1999) A portable, pulsed-molecular-beam, Fourier transform microwave spectrometer designed for

chemical analysis. *Review of Scientific Instruments* 70: 2127–2135.

- Vogelsanger B and Bauder A (1990) Two-dimensional microwave Fourier transform spectroscopy. *Journal of Chemical Physics* 92: 4101–4114.
- Wilks AT, Baker JG, Thompson M, and Alder JF (2002) Quantitative millimetre wavelength spectroscopy at pressures approaching atmospheric: I. Absolute determination of transition line strengths and sample concentrations with a Fabry–Perot cavity spectrometer. *Analytica Chimica Acta* 466: 57–67.

## MICROWAVE-INDUCED PLASMA

See ATOMIC EMISSION SPECTROMETRY: Microwave-Induced Plasma

## MILK

See FOOD AND NUTRITIONAL ANALYSIS: Dairy Products

## MIP

See ATOMIC EMISSION SPECTROMETRY: Microwave-Induced Plasma

## MIPs

See MOLECULARLY IMPRINTED POLYMERS

## MOLECULAR EMISSION CAVITY ANALYSIS

**A C Calokerinos**, University of Athens, Athens, Greece  
© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Molecular emission cavity analysis (MECA) is a flame chemiluminescence technique based on the generation of excited molecules, radicals, or atoms within a hydrogen diffusion flame. The excited species are formed by direct or indirect chemiluminescence mechanisms and are confined within the inner space of a small cavity, which is positioned at a pre-selected point of the flame environment. The emission is monitored at the characteristic wavelength of

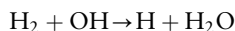
each excited species and the intensity is proportional to the concentration of analyte in the sample. The cavity is responsible for the distinct characteristics of MECA over other emission techniques into which the sample is aspirated into the flame. Hence, the cavity is the cell for the chemiluminescence reactions and promotes excitation due to chemiluminescent rather than thermal processes.

### Basic Principles

#### Hydrogen Diffusion Flame

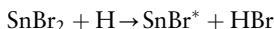
The hydrogen diffusion flame is normally used in MECA. The flame is formed by the combustion of

hydrogen with atmospheric oxygen. Hydrogen is diluted with nitrogen or argon. The main reaction occurring within the flame is



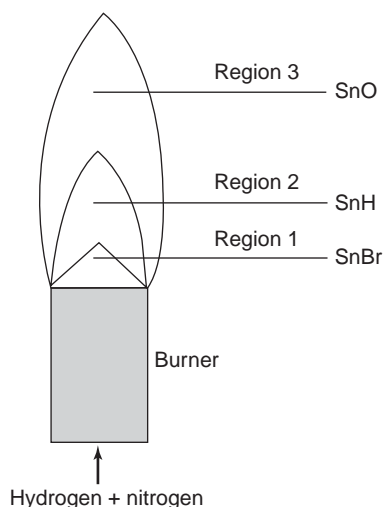
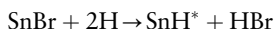
The maximum temperature of the hydrogen diffusion flame does not exceed  $1000^\circ\text{C}$  since mixing of hydrogen and atmospheric oxygen is accomplished by diffusion. Hence, the amount of oxygen, which penetrates from the edges to the center of the flame body, is drastically reduced. As a result of the difficulty in mixing of oxygen with hydrogen, the temperature in the center of the flame is lower than that at the edge. The gradient of temperature is accompanied by a change in nature and a gradient of concentration of flame particles. The edge of the flame is rich in oxygen and water vapors from the main chemical reaction of the flame. The inner part of the flame is rich in atomic and unburnt hydrogen.

The ability of the hydrogen diffusion flame to act as a temperature and concentration gradient across its body can be shown with a simple experiment. If tin(II) bromide ( $\text{SnBr}_2$ ) is aspirated into the flame, then three colored regions appear (Figure 1). In region 1, the concentration of atomic hydrogen is low and the excited particles (marked with an asterisk) are formed by the reaction



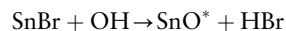
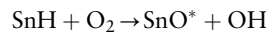
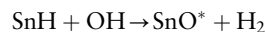
This part of the flame appears green.

After deexcitation,  $\text{SnBr}$  flows to higher parts of the flame where the concentration of atomic hydrogen is higher than in region 1. Thus, region 2 appears red due to the reaction:



**Figure 1** Schematic diagram of the distribution of predominating emitting species after aspiration of tin(II) bromide in a hydrogen diffusion flame.

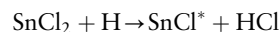
Unreacted  $\text{SnBr}$  from region 1 and  $\text{SnH}$  from region 2 flow upward and sideways toward the periphery of the flame. This part of the flame is rich in molecular oxygen and hydroxyl radicals and the following reactions occur:



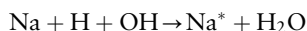
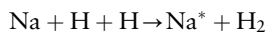
and region 3 appears blue.

### Excitation within the Hydrogen Diffusion Flame

The emitting species are formed and stabilized in the center of the flame where the temperature is minimized. These species are produced by chemiluminescence reactions via direct mechanisms, such as:



or indirect mechanisms in which energy is transferred by or during recombination reactions, such as:



### Sample Introduction into the Hydrogen Diffusion Flame

Direct aspiration of the sample into the hydrogen diffusion flame is always associated with certain problems. The low temperature of the flame may lead to formation of solid particles, which reduce the production of emitting species and decrease sensitivity. The solvent and other components of the analyte solution may alter the temperature and disturb the radical distribution and concentration within the flame body. The distribution of the sample vapors all over the flame may be responsible for generation of more than one emitting species, such as with tin(II) bromide. Production of excited species will also be affected by the difference of temperature at various points of the flame. Moreover, the emitting species will spread over a wide region and the intensity per unit area of flame facing the detector will be low. Finally, the residence time of the analyte within the flame is short and cannot be increased since it is mainly controlled by the flow rate of the support gas.

The problems associated with aspiration of sample solution into a hydrogen diffusion flame for chemiluminescent reactions are eliminated by using the cavity for sample introduction into the flame. The use of the cavity offers the following advantages:

1. The concentration of atoms, radicals, and molecules within the inner space of the cavity can be controlled and adapted to the demands of the analysis. Thus, if the cavity is positioned at the center of



the flame, it operates within an environment of low temperature and rich in hydrogen atoms, favoring reduction and the generation of excited molecules like disulfur ( $S_2$ ) and  $Se_2$ . When the cavity is positioned at the edge of the flame, then the environment is rich in molecular oxygen and hydroxyl radicals at high temperature, favoring oxidation and formation of oxygen-containing excited species. This region can also generate atomic emissions by thermal excitation due to elevated temperatures.

2. The cavity can be cooled by a constant flow of water through its body. This allows continuous operation of the cavity within the flame and promotes emissions from species like  $S_2$  and HPO (Salet phenomenon, see below).

3. The environment of the cavity can be altered by the introduction of a gas or from the material of construction in order to select and promote specific excitation species (Table 1).

**Table 1** Emissions generated within the cavity and the most commonly used wavelengths for analytical measurement

Element	Emitting species	Wavelength (nm)
<i>Normal cavity</i>		
Gallium	GaBr	350
	Gal	391
Germanium	GeCl	455
Phosphorus	HPO	528
Selenium	$Se_2$	411
Sulfur	$S_2$	384
Tellurium	$Te_2$	500
Tin	SnO	485
Lithium	Li	670.8
Cadmium	Cd	326.1
<i>Oxy-cavity<sup>a</sup></i>		
Antimony	SbO–O continuum	355
Arsenic	AsO–O continuum	400
Boron	$BO_2$	518
Carbon	CH	431.5
Nitrogen	NO–O continuum	500
Silicon	SiO	540
Thallium	Tl	377.5
<i>Copper cavity or indium lined cavity<sup>b</sup></i>		
Bromine	CuBr	494
	InBr	375
Chlorine	CuCl	532
	InCl	360
Iodine	CuI	510
	InI	410

<sup>a</sup>A slow flow of pure oxygen is supplied to the cavity in order to promote molecular emissions from elements which require the presence of oxygen in the excited molecule or radical (such as  $BO_2$ ) or due to oxygen atom combination reactions (such as  $AsO + O \rightarrow AsO_2 + h\nu$ ).

<sup>b</sup>Both cavities promote emissions from halide-containing molecules.

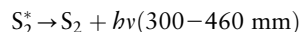
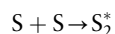
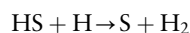
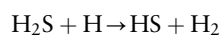
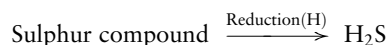
## Salet Phenomenon

In 1869, Salet observed that the blue emission from sulfur compounds and the green emission from phosphorous compounds within a hydrogen flame are intensified when a cool body is introduced into the hot environment of the flame. The importance of the discovery was realized during the development of the flame photometric detector for gas chromatography and is now known as Salet phenomenon. The phenomenon is due to the stabilization of the excited molecules on the cool surface.

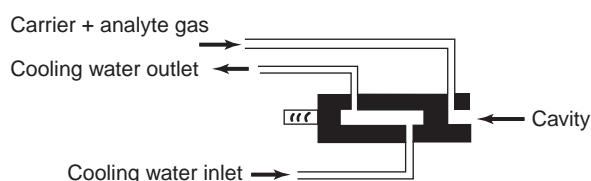
In MECA, the cavity acts as a cool surface and enhances the  $S_2$  emission. Nevertheless, this ability is gradually lost as the cavity heats up. This problem was overcome by designing the water-cooled cavity (Figure 2). The cavity is cooled by a constant flow of water through the body of the construction in order to allow operation at a temperature much lower than that of the flame. The temperature of the inner space of the water-cooled cavity can thus be maintained at a value of just above  $100^\circ\text{C}$ , to avoid condensation of water vapors produced by the flame reactions. This cavity design is used for continuous operation within the flame. The water-cooled cavity is the best design for making full use of the advantages of the Salet phenomenon.

## The Sulfur Emission

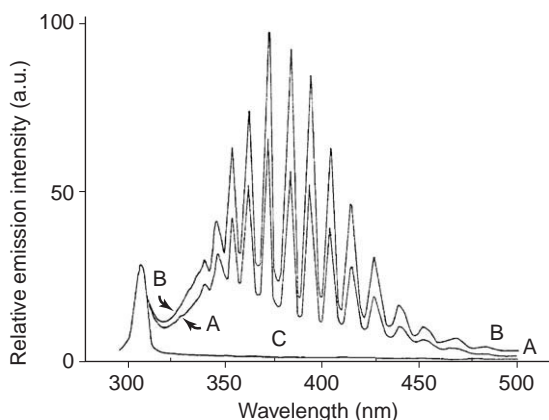
The generation of excited disulfur molecules by flame chemiluminescence provides an extremely sensitive way for the determination of sulfur compounds. The chemiluminogenic reactions are:



The  $S_2$  emission spectrum (Figure 3) is a typical band spectrum with maximum intensity at 384 and 394 nm. In fact, the appearance of the blue emission from a few milligrams per liter of sulfur visible by the human eye makes MECA the most sensitive technique available for the measurement of sulfur



**Figure 2** Schematic diagram of a water-cooled cavity.



**Figure 3**  $S_2$  emission spectrum from carbon disulfide vapors introduced into a water-cooled cavity into which cooling water is supplied at (A) low and (B) high flow rate, showing the effect of the Salet phenomenon on the emission (C: flame background emission). (Reproduced from Stiles DA, Calokerinos AC, and Townshend A (1994) *Flame Chemiluminescence Analysis by Molecular Emission Cavity Detection*. New York: Wiley.)

compounds. Sensitivity is further increased by the Salet phenomenon.

The emission intensity ( $I$ ) is proportional to the concentration of  $S_2$  molecules and consequently to the square of sulfur concentration  $[S]$  in the sulfur-containing analyte:

$$I \approx [S_2^*] = K[S]^2$$

where  $K$  is the reaction constant. In practice, ( $I$ ) is proportional to the  $n$ th power of the concentration of sulfur ( $I = K[S]^n$ ). The experimental values of  $n$  are within the range 1–2 and depend on the sulfur compound and measurement conditions. The exponential relation of emission intensity to sulfur concentration is responsible for the nonlinear shape of the calibration graph. The curve is linearized by plotting  $\log I$  versus  $\log C$  from which the value of  $n$  (slope) and  $K$  (intercept) can also be calculated:

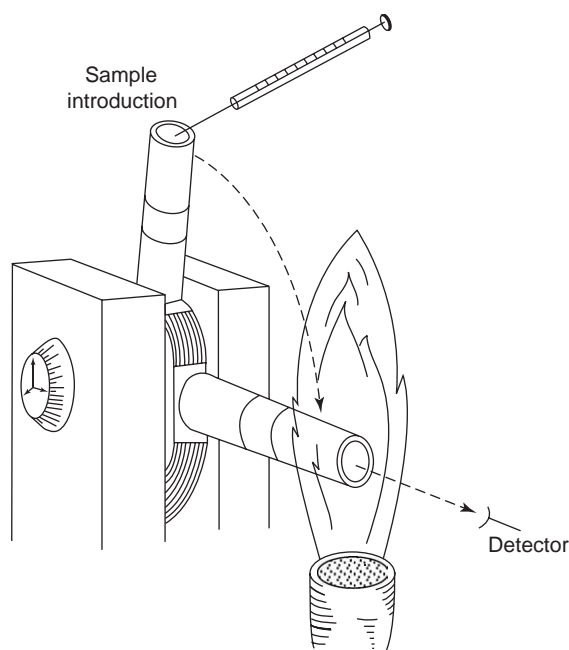
$$\log I = \log K + n \log C$$

## Instrumentation

The major components of a MECA spectrophotometer are the burner, the cavity probe and holder, the optical unit, and the readout system. The only component that requires special design and careful construction is the cavity probe.

### Cavity

Extensive research on MECA has led to the development of two different procedures for using the cavity as a means for introducing the sample into



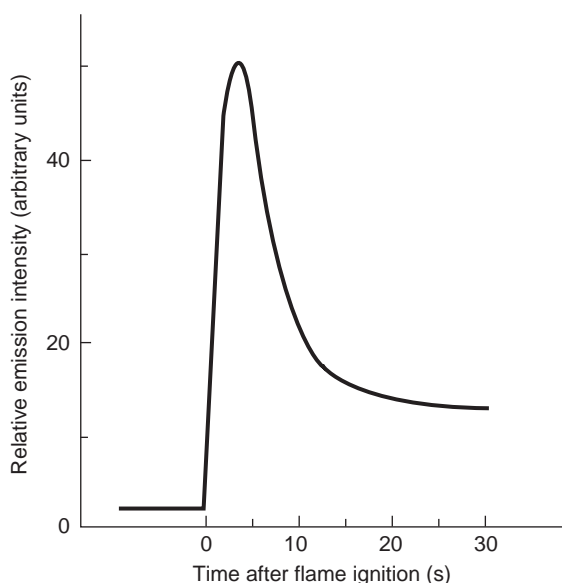
**Figure 4** Conventional MECA cavity introduction into the flame by manual rotation. (Reproduced with permission from Belcher R, Bogdanski SL, Knowles DJ, and Townshend (1975) *Analytica Chimica Acta* 77: 53–63; © Elsevier Science Publishers BV.)

the flame environment. These procedures define two different techniques: conventional MECA and gas generation MECA detection.

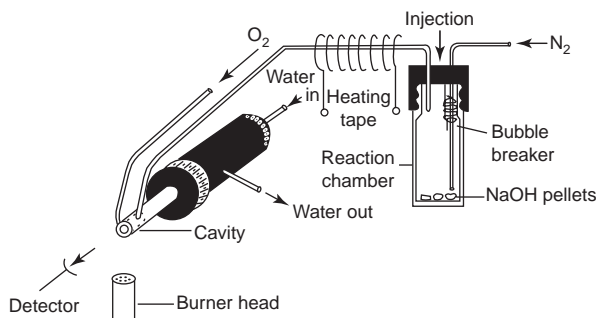
In conventional MECA, a liquid or solid sample is deposited into the cavity (typical dimensions: 8 mm diameter, 5 mm depth), which is then introduced rapidly and reproducibly into the flame (Figure 4).

After introduction into the flame, the cavity temperature increases from ambient to a maximum value, which depends on material and conditions. During the heat-up period of the cavity in the flame, a series of physical and chemical changes occur which are accompanied by generation of the characteristic molecular emission. A typical example of recorded response from thiourea, which generates the  $S_2$  emission, is shown in Figure 5. The cavity remains in the flame as long as required for recording the emission profile (emission intensity as a function of time). It is then removed and cooled before repetition of the process.

In gas generation MECA detection, the sample must be in the form of a gas or vapor, which is introduced into the cavity through a side duct (Figure 6). The emission is generated during passage of the gas or vapor through the cavity. The analyte is converted into a gas by a chemical reaction in a closed reactor, which is continuously purged by a carrier gas, such as nitrogen or argon.



**Figure 5**  $S_2$  emission profile at 384 nm from aqueous thiourea injected in a stainless steel cavity. (Reproduced from Stiles DA, Calokerinos AC, and Townshend A (1994) *Flame Chemiluminescence Analysis by Molecular Emission Cavity Detection*. New York: Wiley.)



**Figure 6** Schematic diagram of a gas generation MECA system for the determination of ammonium. The specific design incorporates a water-cooled cavity with a constant supply of oxygen (oxy-cavity) and is used for the generation of ammonia from ammonium samples. Ammonia is continuously purged into the cavity, reacts with oxygen and generates a bright white emission due to the NO–O reaction. (Reproduced with permission from Belcher R, Bogdanski SL, Calokerinos AC, and Townshend A (1981) *Analyst* 106: 625–635; © The Royal Society of Chemistry.)

## Conventional MECA

The processes, which occur when the cavity is positioned inside the flame by the conventional technique are:

1. Solvent evaporation.
2. Analyte evaporation by boiling, sublimation, thermal decomposition, reductive or catalytic breakdown.

3. Generation of excited species by direct or indirect chemiluminescence reactions.
4. Deexcitation and emission.

The vaporization step (2) determines the time for appearance of maximum emission intensity after the introduction of the cavity into the flame ( $t_m$ ).

The  $t_m$  value is a characteristic parameter for each compound (Table 2) but is also affected by the thermal characteristics of the cavity. The  $t_m$  value is an experimental parameter, which can be used for speciation.

A typical example of how the  $t_m$  value can be used to determine the components of a mixture is shown in Figure 7, where the  $S_2$  emission responses from a mixture containing sulfide, sulfite, and sulfate is shown. In the absence of air, sulfide and sulfite can be determined independently. The response from sulfate is then obtained by introducing air into the hydrogen diffusion flame to vaporize the refractory metal sulfate.

Typical examples of analytical applications of conventional MECA include the determination of saccharin (sulfur-containing sweetener) in soft drinks, total sulfur and sulfate in detergents, phosphate in detergents and rocks, sulfur dioxide in air, inorganic sulfate in urine, and halides in pesticides.

## Automated Conventional MECA Analyzers

Automation in conventional MECA can be achieved by positioning the cavity on a movable device. The sample is injected in the cavity automatically. The movable cavity device rotates and introduces the cavity into the flame and leads it out automatically after recording the emission.

The design can be used for the determination of phosphorus anions in inorganic samples and of sulfur compounds such as thiourea and promethazine.

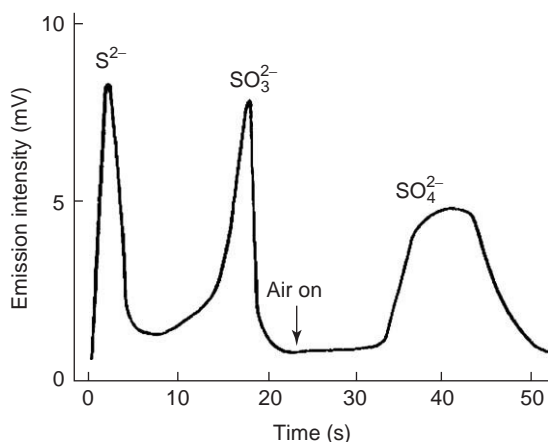
A conventional flow injection MECA analyzer has been used for delivering the liquid samples to the cavity. The design is based on a flow injection system with a slow flow of carrier stream into which sulfur- and phosphorus-containing analytes are injected. The analytes are delivered to the cavity for generation of the corresponding emissions.

## Gas Generation MECA

In gas generation MECA detection, a water-cooled cavity is positioned permanently within the flame. The processes that occur are the generation of excited species, deexcitation, and emission. Therefore, the analyte must be supplied to the cavity as a gas or vapor generated from a chemical reaction and the  $t_m$

**Table 2** Typical  $t_m$  values for some sulfur anions under the same experimental conditions

Anion	$S^{2-}$	$SO_3^{2-}$	$S_2O_3^{2-}$	$SCN^-$	$SO_4^{2-}$
$t_m$ (s)	2.0	2.4	10	18	40

**Figure 7**  $S_2$  emission profiles from 5  $\mu$ l of a mixture containing sulfide, sulfite, and sulfate (as sodium salts) injected into a stainless steel cavity. (Reproduced with permission from Al-Abachi MQ, Belcher R, Bogdanski SL, and Townshend (1976) *Analytica Chimica Acta* 86: 139–146; © Elsevier Science Publishers BV.)

concept is not valid. Hence, the components of a mixture can be determined simultaneously only if they are resolved prior to introduction into the cavity, i.e., by a series of chemical reactions or via a gas chromatographic column.

The major advantages of gas generation MECA detection are:

1. Water-cooled cavities can be employed for effective use of the Salet phenomenon.
2. A great variety of chemical reactions can be used, as shown in **Table 3**.
3. The technique can be very easily automated by flow injection or continuous flow manifolds.

### Automated Gas Generation Analyzers

In the continuous flow technique, the sample is injected into the flowing reagent stream. Intermixing of individual samples is eliminated by air bubbles placed into the stream. In the flow injection technique, the sample is injected into an unsegmented continuous stream of a liquid solvent, which is then mixed with the reagent stream by diffusion and conduction. Both techniques have been successfully used in conjunction with the MECA cavity, which in this

**Table 3** Gas evolution procedures used for gas generation MECA detection

Generation principle	Procedure	
Physical	Conversion of analyte into a vapor by an electrically heated tantalum filament or a modified Massmann furnace	
Gas chromatography	Use of the cavity as a detector	
Chemical	Conversion of analyte into a vapor by a chemical reaction, such as:	
Analyte	Reactant	Generated gas
<i>Hydride generation methods</i>		
As	$NaBH_4$	$AsH_3$
Sb	$NaBH_4$	$SbH_3$
Sn	$NaBH_4$	$SnH_4$
Se	$NaBH_4$	$SeH_2$
Te	$NaBH_4$	$TeH_2$
<i>Acidic generation methods</i>		
$S^{2-}$	$H^+$	$H_2S$
$SO_3^{2-}$	$H^+$	$SO_2$
Thiamine	Alkaline degradation to $S^{2-}$ and then acidification	$H_2S$
<i>Alkaline generation methods</i>		
$NH_4^+$	$OH^-$	$NH_3$
<i>Redox and miscellaneous methods</i>		
B	$CH_3OH + H_2SO_4$	$(CH_3O)_3B$
Si	$F^- + H_2SO_4$	$SiF_4$
$F^-$	$SiO_4^{4-} + H_2SO_4$	$SiF_4$
$NO_3^-$	Devarda's alloy	$NH_3$
$NO_2^-$	$H^+ + I^-$	$NO$
$SO_4^{2-}$	Tin + anhydrous phosphoric acid	$H_2S$
Organic sulfur compounds	Electrolysis	$H_2S$

case operates as a detector of a gas produced by a chemical reaction occurring within the flowing stream.

Typical examples of analytical applications of gas generation MECA detection include the determination of ammonium in fertilizers, fluorine in toothpaste, boron in steel samples, and thiamine and cephalosporins in pharmaceutical preparations.

### Gas and Liquid Chromatographic MECA Detection

The cavity can easily be used to detect volatile compounds as they elute from a gas chromatograph. The outlet of the column is connected to the cavity by a short stainless steel transfer tube, which is kept at the same temperature with the gas chromatographic oven.

Typical application of MECA detection after gas chromatographic separation include the separation and detection of chlorinated solvents and pesticides by using an indium-lined cavity and bromo- and iodo-compounds by using a copper water-cooled cavity.

The cavity can also be operated as a liquid chromatographic detector.

### Speciation

Since the discovery of MECA, the advantages and disadvantages of flame chemiluminescence have been

fully investigated and established. The sensitivity of MECA for sulfur-containing compounds is excellent and the selectivity can be drastically improved by using automated gas generation analyzers. The conventional technique can be used for speciation analysis based on the different  $t_m$  values of compounds, which generate the same emission within the cavity.

See also: **Chemiluminescence**: Overview.

### Further Reading

- Belcher R, Bogdanski SL, and Townshend A (1973) Molecular emission cavity analysis – a new flame analytical technique. Part I. Description of the technique and the development of a method for the determination of sulphur. *Analytica Chimica Acta* 67: 1–16.
- Burguera M, Bogdanski SL, and Townshend A (1980) Molecular emission cavity analysis. *CRC Critical Reviews in Analytical Chemistry* 10: 185–245.
- Calokerinos AC (1997) Molecular Emission Cavity Analysis: principles and applications. *Trends in Analytical Chemistry* 16: 78–84.
- Calokerinos AC and Townshend A (1982) Spectroscopic analysis using cool flames. Part II. A review of practical applications of molecular emission cavity analysis (MECA). *Progress in Analytical and Atomic Spectroscopy* 5: 63–77.

## MOLECULARLY IMPRINTED POLYMERS

**M E Díaz-García and A Fernández-González,**  
University of Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Molecular imprinting is a general strategy of synthesis that allows tailor-made polymeric materials to be obtained with memory. The concept of molecular imprinting emerged in 1940 following Linus Pauling's theory about the formation of antibodies. According to this theory, antigen molecules act as templates around which serum proteins assemble to form antibodies. Although this theory was demonstrated to be erroneous, chemists found this idea very suggestive for synthesizing materials with a certain memory for the template molecule. Following this

idea, Pauling's disciple Dickey imprinted silica gel with different dyes and obtained the first molecularly imprinted synthetic material. This experiment was repeated several years later by Sidney Bernhard.

In 1972, this approach for molecular imprinting was modified when Günter Wulff's research group developed a nonsiliceous cross-linked polymer synthesized around a template molecule. This route was also successfully followed by Shea and Thompson in 1978, who obtained a polymeric material with a predetermined spatial orientation of their functional moieties, a molecularly imprinted polymer (MIP). Both these syntheses were carried out by covalently binding the template to the functional monomer, followed by posterior polymerization of the derivatized monomer. Noncovalent approach for imprinted material synthesis was first carried out by Klaus Mosbach *et al.* in 1988. This technique of imprinting



case operates as a detector of a gas produced by a chemical reaction occurring within the flowing stream.

Typical examples of analytical applications of gas generation MECA detection include the determination of ammonium in fertilizers, fluorine in toothpaste, boron in steel samples, and thiamine and cephalosporins in pharmaceutical preparations.

### Gas and Liquid Chromatographic MECA Detection

The cavity can easily be used to detect volatile compounds as they elute from a gas chromatograph. The outlet of the column is connected to the cavity by a short stainless steel transfer tube, which is kept at the same temperature with the gas chromatographic oven.

Typical application of MECA detection after gas chromatographic separation include the separation and detection of chlorinated solvents and pesticides by using an indium-lined cavity and bromo- and iodo-compounds by using a copper water-cooled cavity.

The cavity can also be operated as a liquid chromatographic detector.

### Speciation

Since the discovery of MECA, the advantages and disadvantages of flame chemiluminescence have been

fully investigated and established. The sensitivity of MECA for sulfur-containing compounds is excellent and the selectivity can be drastically improved by using automated gas generation analyzers. The conventional technique can be used for speciation analysis based on the different  $t_m$  values of compounds, which generate the same emission within the cavity.

See also: **Chemiluminescence**: Overview.

### Further Reading

- Belcher R, Bogdanski SL, and Townshend A (1973) Molecular emission cavity analysis – a new flame analytical technique. Part I. Description of the technique and the development of a method for the determination of sulphur. *Analytica Chimica Acta* 67: 1–16.
- Burguera M, Bogdanski SL, and Townshend A (1980) Molecular emission cavity analysis. *CRC Critical Reviews in Analytical Chemistry* 10: 185–245.
- Calokerinos AC (1997) Molecular Emission Cavity Analysis: principles and applications. *Trends in Analytical Chemistry* 16: 78–84.
- Calokerinos AC and Townshend A (1982) Spectroscopic analysis using cool flames. Part II. A review of practical applications of molecular emission cavity analysis (MECA). *Progress in Analytical and Atomic Spectroscopy* 5: 63–77.

## MOLECULARLY IMPRINTED POLYMERS

**M E Díaz-García and A Fernández-González,**  
University of Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Molecular imprinting is a general strategy of synthesis that allows tailor-made polymeric materials to be obtained with memory. The concept of molecular imprinting emerged in 1940 following Linus Pauling's theory about the formation of antibodies. According to this theory, antigen molecules act as templates around which serum proteins assemble to form antibodies. Although this theory was demonstrated to be erroneous, chemists found this idea very suggestive for synthesizing materials with a certain memory for the template molecule. Following this

idea, Pauling's disciple Dickey imprinted silica gel with different dyes and obtained the first molecularly imprinted synthetic material. This experiment was repeated several years later by Sidney Bernhard.

In 1972, this approach for molecular imprinting was modified when Günter Wulff's research group developed a nonsiliceous cross-linked polymer synthesized around a template molecule. This route was also successfully followed by Shea and Thompson in 1978, who obtained a polymeric material with a predetermined spatial orientation of their functional moieties, a molecularly imprinted polymer (MIP). Both these syntheses were carried out by covalently binding the template to the functional monomer, followed by posterior polymerization of the derivatized monomer. Noncovalent approach for imprinted material synthesis was first carried out by Klaus Mosbach *et al.* in 1988. This technique of imprinting

is currently the one most used for the synthesis of such materials.

## The Molecular Imprinting Process

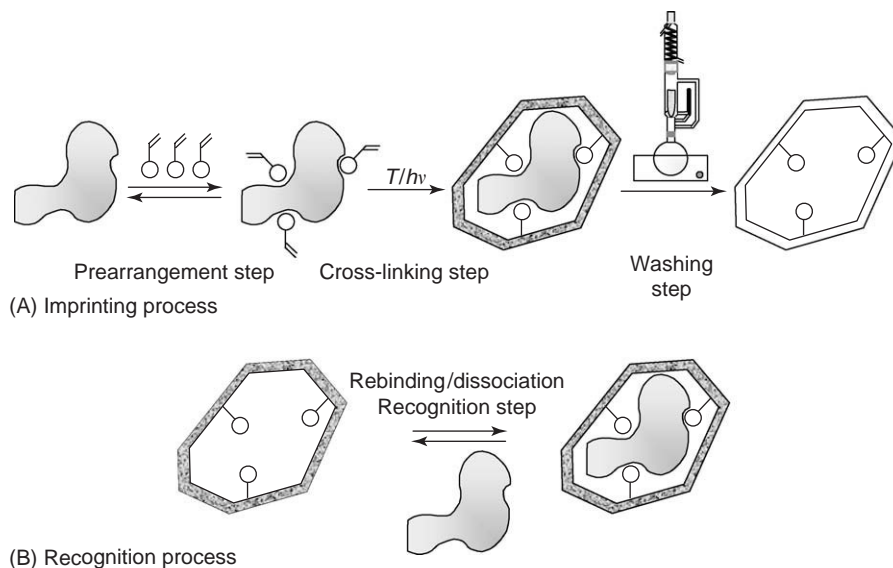
The memory that imprinted polymers exhibit for a particular molecule is created in the synthesis process as a result of interactions of the molecule with functional monomers. The process may be summarized in three main steps (Figure 1A):

1. Prearrangement step (complex formation), where the template associates with the functional monomer(s) forming an adduct or a polymerizable derivative.
2. Cross-linking step, in which the polymerizable derivative or the adduct copolymerizes or polymerizes with an excess of a cross-linking agent in an inert solvent, by thermal or photochemical radical initiation.
3. Washing step, where the template molecule is removed from the polymer by washing it with a

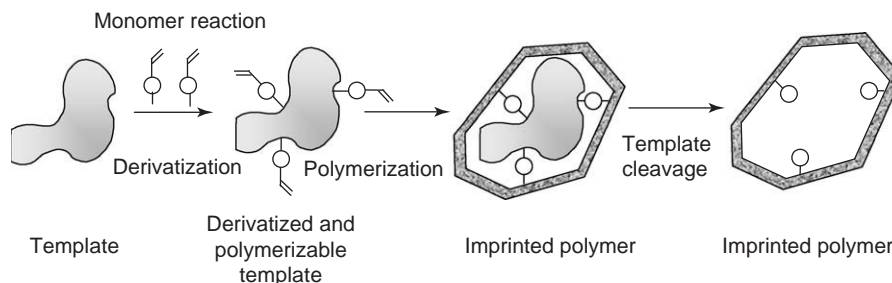
suitable solvent, usually with Soxhlet extraction, or by basic or acid hydrolysis.

The last step leaves a polymer matrix with cavities, or imprints, that are complementary in shape, size, position of the functional groups, and electronic environment to the template. Due to the rigid cross-linked structure of this imprinted material, selective rebinding and recognition of template is possible through a process similar to the lock-and-key fit (Figure 1B). There are two general imprinting strategies: the preorganized imprinting method (covalent approach) and the self-assembly imprinting method (noncovalent approach).

In the covalent approach (Figure 2), a polymerizable derivative of the template is obtained by linking the template with a vinyl functional monomer via a strong reversible covalent bond (e.g., boronate esters, Schiff bases, or ketal). The derivatized print molecule is then radical copolymerized with an excess amount of a cross-linking agent, using either thermal or photochemical radical initiation. Reaction conditions are



**Figure 1** Scheme of (A) molecular imprinting process and (B) recognition process.



**Figure 2** Covalent approach to the imprinting process.

chosen in order to stabilize the covalent bond during the polymerization. The template is finally cleaved from the imprinted sites using adequate chemical methods. Rebinding of the target molecule requires the original reversible covalent bond to reform under conditions that favor its uptake. The covalent approach results in relatively homogeneous binding sites distributed throughout the imprinted polymer. This is the result of the relatively strong covalent interactions that make possible the use of stoichiometric amounts of functional monomers with respect to the number of template molecules used. Another advantage of the covalent approach is that the polymer shows a weak dependence on the solvent used during the polymerization process. Shortcomings of the approach include the need of first preparing a monomer–template conjugate, the unavoidable chemical cleavage of the template from the polymer, and the slow rebinding kinetics. In addition, due to a lack of suitable functionality on the template, it is not always possible to prepare the polymerizable template.

The concept of noncovalent imprinting process relies on the self-assembly of functional monomers around the template in the prepolymerization mixture in a way that maximizes the binding interactions between the two species (Figure 3). These interactions between monomers and template may be of different nature: hydrogen bonding, electrostatic, hydrophobic, dipole–dipole, and/or  $\pi$ – $\pi$  interactions. After radical polymerization with a cross-linking monomer, the template is washed out with a solvent able to disrupt the noncovalent interactions between the template and the binding site. Noncovalent interactions (which are prevalent in biological systems) are exploited during both the polymerization and the rebinding processes. The main advantages of this approach are the simplicity of the synthesis, the relatively wide selection of functional monomers to choose from, the removal of the template through a simple solvent extraction process, and the rapid rebinding kinetics. MIPs prepared with a noncovalent monomer–template adduct usually contain a heterogeneous binding site population as a consequence of the low association constants, which limit the

formation of stable complexes. On the other hand, it is necessary to use an excess of functional monomer in order to shift the equilibrium toward the adduct formation that results in the random incorporation of functional monomers, many of which may not form a binding site. In spite of these limitations, at this time, most of the widely used MIPs are synthesized by radical polymerization of functional and cross-linking monomers using noncovalent interactions with the template.

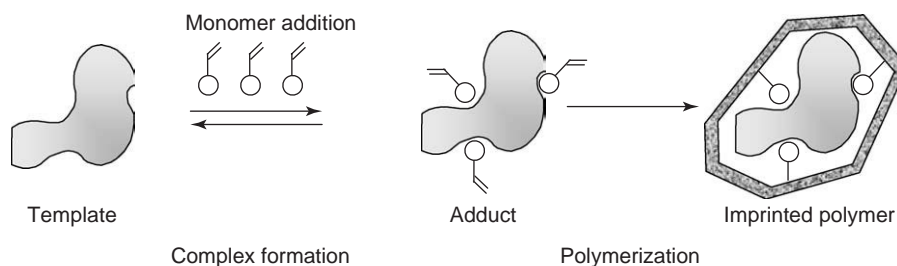
## Reagents to Create a Molecularly Imprinted Polymer

The efficiency of the molecular recognition in the molecular imprinting process is affected by many parameters. Some of the main characteristics of reagents that can be used for imprinting are listed in the following.

### Templates

Molecules offering multiple site interaction for the functional monomer to ensure stable complexation often yield binding sites of higher specificity and affinity in the polymer. Because different interactions are acting simultaneously during the imprinting process, in principle there is no limit as to which print molecules can be used as templates. Thermodynamic considerations indicate that the more rigid the template, the higher the selectivity and affinity of the recognition.

From a practical viewpoint, the structure and the chemical characteristics of the template seem to be the starting point to select the imprinting approach that should be followed. For example, when noncovalent imprinting is applicable and hydrophobic interactions are desired, the solvent can be selected to enhance this effect. If acid groups are present in the target molecule, basic functional monomers should provide strong ionic interactions between the charged template and the monomer. Other forces of attraction can also be used to produce strong interactions; for example, if the print molecule is able to form metal ion complexes, then metal chelating



**Figure 3** Noncovalent approach to the imprinting process.

functional monomers should be the choice. In some instances, a combination of two functional monomers can be used with good success.

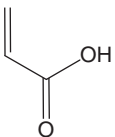
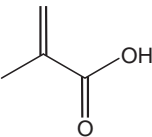
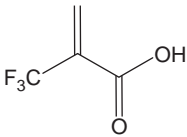
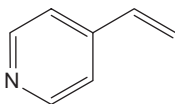
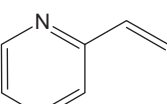
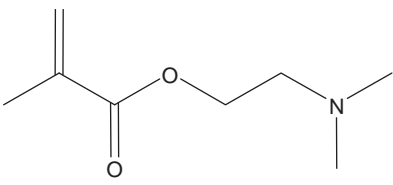
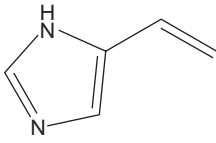
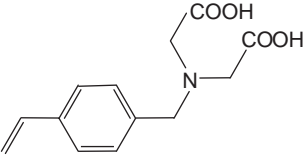
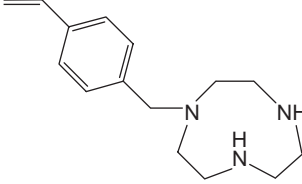
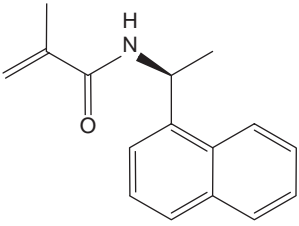
### Functional Monomers

The monomer is a key component for the synthesis of an analytically useful MIP. Its role is to provide points of electronic recognition for rebinding the template and, as stated above, its choice is often based on its capacity to form hydrogen bonds or an ion pair with the template. There is a selection of acidic, basic, neutral, metal chelating, and chiral compounds (Table 1) commercially available. Among them, methacrylic-acid-based monomers are frequently used to interact with templates having electron-donating functionalities, whereas vinylpyridine derivatives are useful as hydrogen acceptors. Chiral naphthyl monomers are expensive, but they are a good option to prepare chiral stationary phases based on imprinted polymers when no optically pure print molecules are available.

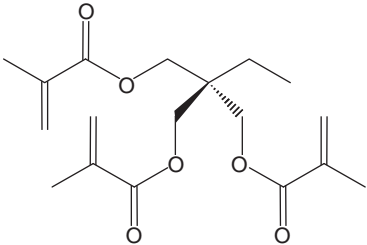
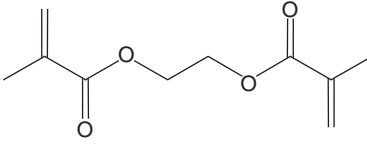
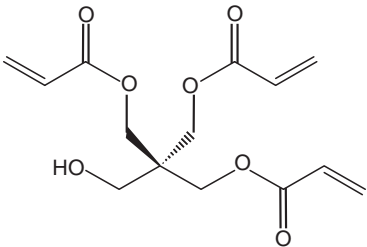
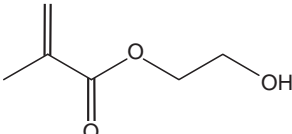
### Cross-Linker

A very high degree of cross-linking is needed in the imprinting process to achieve a very high specificity and selectivity, which is ensured by a rigid three-dimensional structure of the polymer. In order to achieve this, the functional monomer is copolymerized with a relatively large proportion of cross-linking monomer. However, the appropriate level of cross-linking agent must be optimized in order to maintain the binding specificity of the polymer. If the levels are too high, the high degree of cross-linking hinders the template diffusion into binding sites, while low levels of cross-linking may reduce rebinding because functional groups are not sufficiently fixed. On the other hand, hydrophobic/hydrophilic balance of the cross-linker determines the hydrophobic/hydrophilic nature of the MIP, which affects the extent of nonspecific binding. In general, the correct choice of the cross-linker and its proportion are often arrived at through trial and error.

**Table 1** Most common functional monomers in molecular imprinting, classified according to their functionality

Acid and hydrogen bond-forming monomers			
	Acrylic acid	Methacrylic acid	Trifluoromethyl acrylic acid
Basic functional monomers			
	4-Vinylpyridine	2-Vinylpyridine	<i>N,N</i> -Dimethylaminoethyl methacrylate
Metal chelating monomers			
	5-Vinylimidazole	4-Vinylbenzyliminodiacetic acid	1-(4'-Vinylbenzyl)-1,4,7-triazacyclononane
Chiral functional monomers			
	(S)-(-)- <i>N</i> -Methacryloyl-1-naphthylethylamine		

**Table 2** Most common cross-linkers in molecular imprinting, classified according to their nature

<i>Cross-linker</i>	<i>Name</i>	<i>Nature</i>
	Trimethylolpropane trimethacrylate	Hydrophobic
	Ethylene glycol dimethacrylate	Hydrophobic
	Pentaerythritol triacrylate	Hydrophilic
	2-Hydroxy-ethyl-methacrylate	Hydrophilic

The number of appropriate cross-linker agents is limited due to the solubility of the ‘monomerized’ template and the cross-linker itself. **Table 2** lists some cross-linker monomers. In early developments, isomers of divinylbenzene were used to cross-link styrene and other functional monomers. More recently, acrylic and methacrylic monomers have demonstrated provision for better selectivity. So, acrylic cross-linkers such as ethylene glycol dimethacrylate (EGDMA), trimethylolpropane trimethacrylate (TRIM), tri- and tetra-functional cross-linkers such as pentaerythritol triacrylate or pentaerythritol tetraacrylate are widely used. Among these EGDMA is the one most used for general imprinting purposes, and TRIM for developing imprinted stationary phases of great mechanical stability, which are useful for chromatographic systems.

#### Porogenic Solvent

The role of solvent in molecular imprinting is especially important in the prearrangement (noncovalent) approach. Solvent not only influences the polymer morphology (inner surface area and average pore

size) but also drives the strength of noncovalent interactions. Commonly, the more polar the solvent, the worse the resulting recognition, since the polarity of the solvent weakens the noncovalent interactions. In this way, the solvent should be as nonpolar as possible in order to maximize ionic interactions between the template and the functional monomer. Solvents with low hydrogen bonding basicity/acidity should be used when hydrogen bond interactions are expected to be involved. The solvent used also affects the physical characteristics of the MIP. Several studies seem to indicate that acetonitrile – a fairly polar solvent – leads to polymers with bigger pores than those prepared when chloroform is used. The solvent also influences the swelling of the polymer, but some authors have demonstrated that polymer morphology is not critical from the viewpoint of selectivity or strength in substrate rebinding.

#### Radical Initiators

Free radicals may be obtained by heating or by using UV radiation. The most common radical initiators used in molecular imprinting are the azo types and,



among them, azobisisobutyronitrile and azobisdime-thylvaleronitrile are the most important. If the template can be thermally decomposed, UV radiation at low temperature should be the reaction condition to produce free radicals. These conditions may also be desirable to improve molecular recognition when the noncovalent approach is used. However, in most cases, free radicals are produced thermally, at temperatures ranging between 60°C and 80°C.

Much of the work done on developing imprinted polymers is based on a trial and error system. New avenues are being opened to prepare these materials in a more rational way, for example, through the development of computer software programs, based on molecular dynamics and thermodynamics principles of imprinting, which are able to evaluate the most efficient functional monomer for the recognition process. In addition, the use of combinatorial chemistry approaches for synthesizing and testing a polymer library for a given template may result in high throughput systems to evaluate the best recognition material not only for the template but also for related compounds.

## MIP Configuration

Most MIP materials described to date have been prepared by the so-called monolith technique. However, it is possible to produce MIPs by different approaches, each with its advantages, shortcomings, and applications. **Figure 4** summarizes a typical synthetic procedure to obtain the polymerization

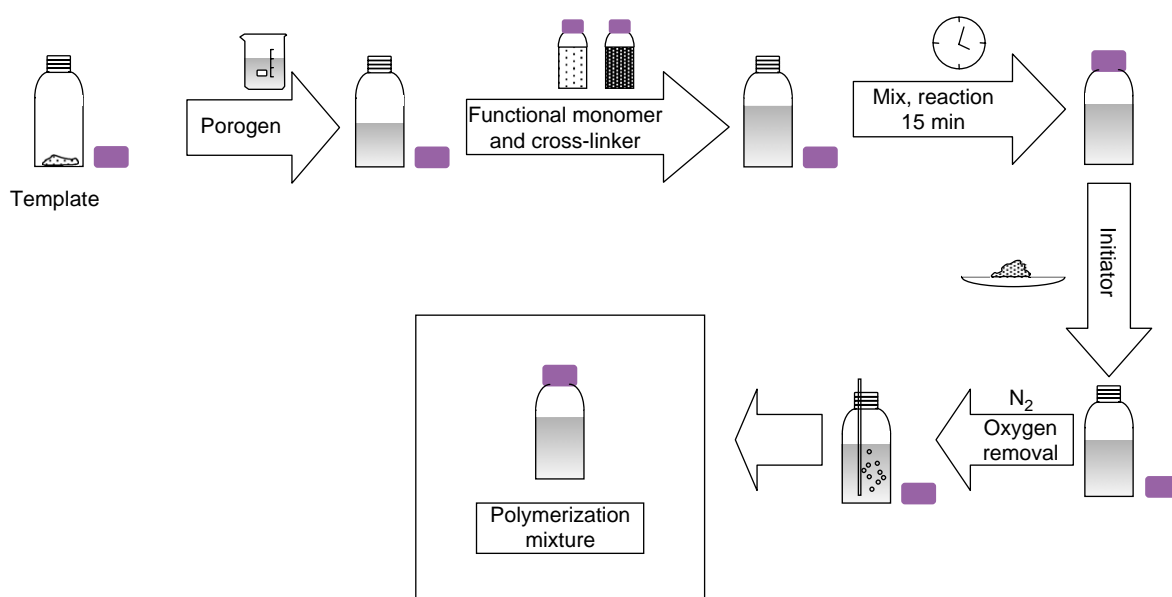
mixture. From this point, different techniques may be selected to develop the imprinted polymer.

### Monolithic Technique

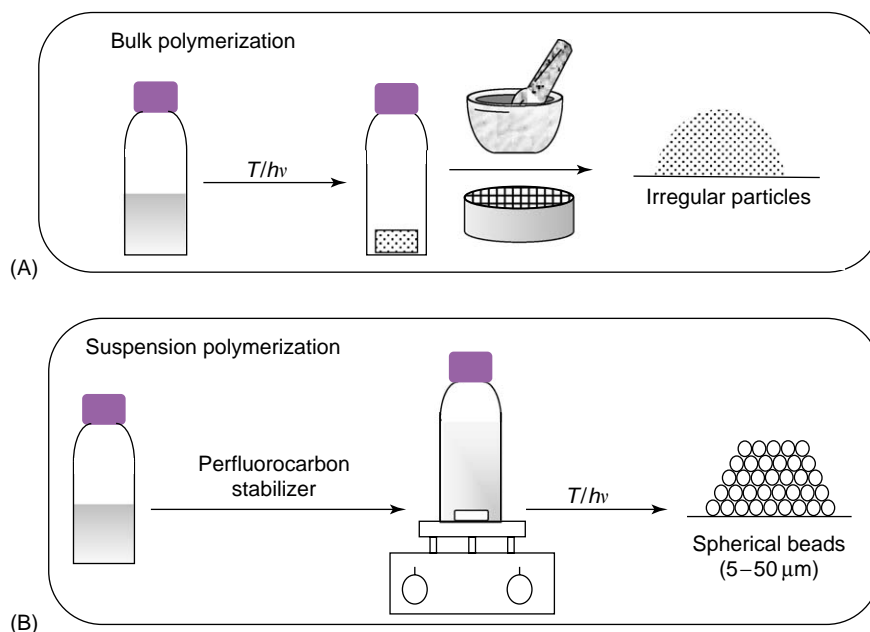
This technique refers to homogeneous solution/precipitation polymerization methods giving polymer monoliths that need to be grounded and sieved prior to use (**Figure 5A**). The main drawbacks of this approach are the loss of material in the sieving process, the large size distribution of particles, and the irregular particle shapes. Moreover, during the grinding process heat is generated, which may modify the physical characteristics of the MIP (e.g., surface area), thus creating new nonspecific binding sites. This technique is simple and most developments rely on it as the first step to demonstrate the imprinting effect and to evaluate potential applications.

### Suspension Polymerization

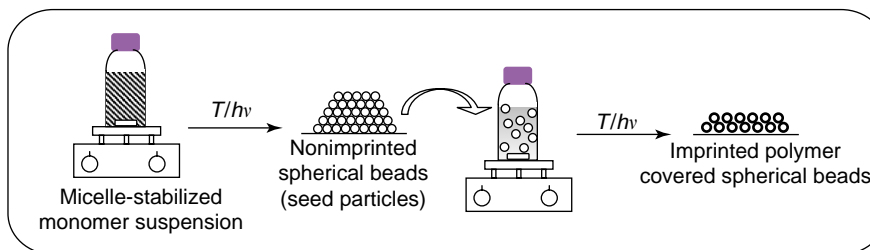
Some of the shortcomings of the monolithic technique, such as the irregular particle shape and the large size distribution, can be overcome by using the approach known as suspension polymerization, a fast method to synthesize spherical particles in the range of 5–50 µm. In a typical recipe, the polymerization of the emulsion of the reagents (monomer, template, cross-linker, initiator, porogenic solvent, and a suspension stabilizer) is thermally initiated upon heating and a chain reaction process takes place within the suspended globules (**Figure 5B**). Globules collide with each other, coalesce into larger ones, and



**Figure 4** Protocol for obtaining polymerization mixture, which will be polymerized later.



**Figure 5** Obtaining the monolithic and suspension presentations of the imprinting polymer.



**Figure 6** Obtaining grafted spherical particles of imprinted polymer.

rapidly break up again into smaller ones, so that their average size is the result of a dynamic process. The suspension stabilizer is necessary to avoid globule agglomeration, and perfluorocarbon liquid continuous phases have been reported to yield good-quality imprinted polymer beads with controlled particle sizes.

### Precipitation Polymerization

Particles in the submicrometer scale (typically 0.3–10  $\mu\text{m}$ ) can be obtained through this method. The synthesis is similar to the monolithic technique, but an excess of porogenic solvent is used. As particles grow, they become more insoluble in the organic solvent, and precipitation of nano- and microparticles is finally produced to form a polymer monolith. The main drawback is the larger amounts of template needed because of the dilution factor, but the particles work better than those obtained by grinding and sieving procedures. This method has been

successfully applied to prepare imprinted polymer beads for binding assays.

### Grafted Spherical Particles

In this approach, an imprinted polymer layer is grafted onto preformed spherical activated beads. Conventional supporting materials include silica particles, polystyrene and poly(TRIM) beads. While the surface of poly(TRIM) beads holds residual pendant vinyl groups, silica particles are needed to be derivatized with vinyl monomers (silane chemistry) (Figure 6). The surface groups provide the anchoring sites to incorporate the imprinted polymer layer. The technique has many advantages compared with bulk imprinted polymers: binding sites are more accessible and mass transfer and binding kinetics may be faster.

Takagi and co-workers have shown that this technique can be even performed using surfactant molecules as anchoring groups on the supporting material surface for the template molecule. Resins

imprinted by this surface approach have been prepared for metal chelation using oleic acid as a functional surfactant and oleyl-amine for catching inorganic anions such as hexacyanoferrate(II). The technique is less straightforward than one using pre-formed spherical activated beads, and its applicability to substrates of biological interest is limited due to the requirement for complexation at the aggregate-bulk solution interface.

### Imprinting at Surfaces

A number of standard techniques are available for production of MIP thin films at a surface such as a glass slide, an electrode surface, or a polymeric membrane. Spin coating, dip coating, spray coating, microcontact printing electropolymerization, chemical grafting, and physical adsorption can be chosen to synthesize thin films, depending on the nature of the surface and the specific application. For example, MIPs can be synthesized at a nonconducting surface by chemical grafting or at an electrode surface by electropolymerization. Piletsky and colleagues have described the *in situ* synthesis of a polymer, based on aminophenylboronic acid, in the presence of epinephrine (template) in the wells of a polystyrene microtiter plate. The process results in the grafting of a thin polymer layer onto the polystyrene surface and can be used in competitive enzyme-linked assays, using horseradish peroxidase-labeled norepinephrine. On the other hand, this *in situ* protocol is applicable to prepare combinatorial libraries of imprinted polymers as a robot can be easily used for automating the synthesis and evaluation of polymers.

### Sol-Gel Molecular Imprinting

The first synthetic materials with memory for a template were obtained by Dickey in 1949 using a silica gel matrix. Imprinted silica materials were produced by acid precipitation of aqueous solution of sodium silicate in the presence of dyes as templates (e.g., methyl orange). In the following years research on

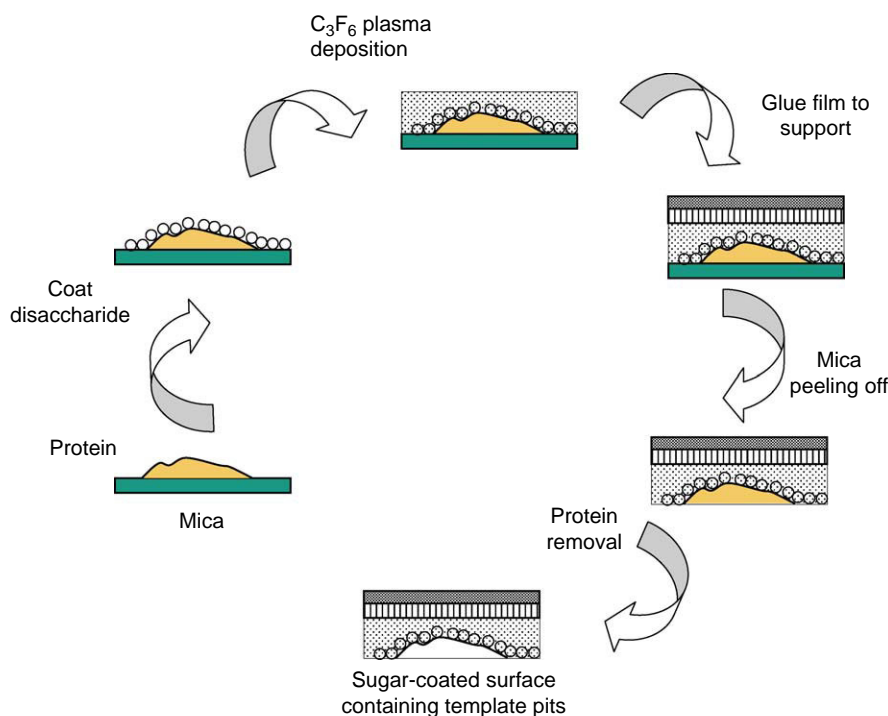
imprinted silicates and metal oxide sol-gel continued and simple amorphous silicates could be imprinted for different dyes, *N*-heterocycle aromatics, proteins, and for resolution of enantiomers (e.g., camphorsulphonic acid) and stereoisomers (e.g., *N*-methyl-3-methoxymorphine, nicotine, quinine, quinidine, cinchonine, and cinchonidine). The recognition properties of these materials are dictated by the gel structure and surface chemistry, both of which are influenced by a multitude of factors, such as catalysts, pH, solvents, aging, and nature of precursors. To date, sol-gel derived molecularly imprinted materials have not revealed the same level of success as imprinted polymers for separation, catalysis, and sensor applications. One of the most unique features of imprinted sol-gel materials is their high aqueous compatible nature, allowing incorporation of biologically based guests (e.g., saccharides, peptides, proteins, drugs, and metal ions).

### Bioimprinting

Bioimprinting is predominantly used for imprinting in biopolymers, mainly proteins. The basic imprinting process was first described by Klibanov *et al.* and consisted in equilibrating a protein (subtilisin) and the template molecule (a competitive inhibitor) in aqueous buffer solution. The protein surrounds the template, and a set of weak bonds is established between them. The resulting protein-template complex was frozen, dried, and the template was removed using an organic solvent under anhydrous conditions. The imprinted enzyme was found to be several times more active than the native enzyme in anhydrous medium. Since this pioneering work of Klibanov *et al.*, different approaches have been addressed to prepare imprinted proteins and other biopolymers with specific purposes. Table 3 lists some representative examples. Although the thrust of molecularly imprinting research is focused on the development of imprints for small molecules, a new direction to develop imprints against larger structures – such as proteins, cells, and viruses – is

**Table 3** Imprinting of biomolecules and applications

<i>Biomolecule</i>	<i>Template</i>	<i>Solvent for rebinding</i>	<i>Application</i>
Bovine serum albumin	L-Malic acid	Ethyl acetate	Chromatographic separation, molecular recognition
Subtilisin Carlsberg	<i>N</i> -Acetyl-L-phenylalanine amide	Cl <sub>4</sub> C	Catalysis (transesterification reaction)
Lipase	( <i>R</i> )-1-phenyl ethanol	Isooctane	Catalysis (lauric acid esterification)
Amylose (cross-linked with cyanuric chloride)	Methylene Blue	Aqueous	Molecular recognition
Chitosan	Metal ions (Cu(II), Cd(II), Zn(II), Hg(II))	Aqueous	Ion sequestering, ion recognition



**Figure 7** Novel imprinting process for large macromolecules.

receiving attention lately. An illustrative example of such advance is the design of stable molecular recognition elements for proteins developed by Ratner *et al.* **Figure 7** shows a schematic description of the technique. A template protein (e.g., bovine serum albumin) is bound to a mica surface, as proteins adsorbed to this atomically flat material with minimal denaturalization and conformational changes. A layer of a nonreducing disaccharide was then spin-coated on the protein surface, followed by glow-discharge plasma deposition of a fluoropolymer layer in order to bind covalently the sugar molecules. The film was epoxyglued to glass and the mica sheet peeled off. Finally, the protein is removed by using an aqueous NaOH/NaOCl solution, resulting in a sugar-coated surface containing imprinted nanocavities with the size and shape of the template protein. New nanotechnologies are emerging to generate imprints for large biomolecules, cells, and microorganisms to develop new recognition materials for biological applications.

## Applications of Molecularly Imprinted Materials

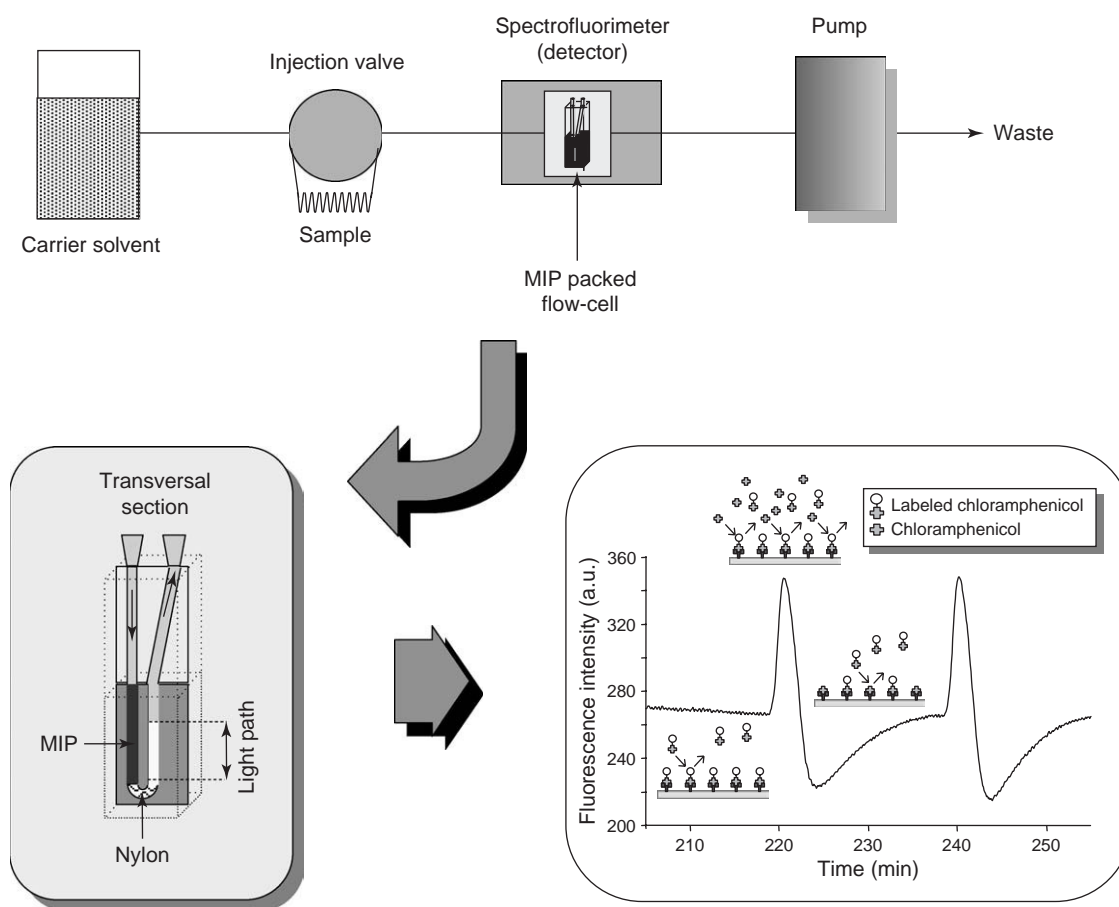
Molecularly imprinted materials can be used in many applications from synthetic chemistry as catalysts to pharmaceutical usages as drug deliverers. From an analytical point of view, imprinted polymers can be used in a variety of applications.

## Separations

MIPs are broadly used as stationary phases for high-performance liquid chromatography. An imprinted material may be *in situ* synthesized inside the chromatographic column or may be synthesized in a bulk way, grounded, sieved, and then packed into a column. At present, some companies are producing imprinted materials for solid-phase extraction (MISPE). For example, Affinity Chromatography Ltd., Ballasala, Isle of Man, is manufacturing an imprinted polymer for propranolol. MIPs as sorbents in MISPE are used to achieve selective cleanup of biological and environmental samples with high analyte enrichment factors. The MIPs can be directly used in extraction columns or cartridges in combination with existing instrumentation. Compared with immunoextraction sorbents, MIPs are stable, can be easily obtained, and have a higher adsorption capacity. A wide range of MIP materials has been examined as SPE sorbents for pesticides and compounds of biological and biomedical interest (e.g., atrazine, chlorotriazines, bentazone, propranolol, nicotine, cotinine, pentamidine,  $\beta$ -lactam antibiotics, and theophylline analogs).

## Biomimetic Assays

Due to their selectivity for the template, MIPs may be used as synthetic receptors, substituting antibodies, and biological receptors in many biochemical assays. **Figure 8** shows the principle of a continuous flow



**Figure 8** Flow injection analysis (FIA) based fluorescent competitive sensor based on molecularly imprinting acrylic polymer. (Reprinted with permission from Suárez-Rodríguez JL and Díaz-García ME (2001) Fluorescent competitive flow-through assay for chloramphenicol using molecularly imprinted polymers. *Biosensors and Bioelectronics* 16(9–12): 955–961; © Elsevier.)

**Table 4** Analytical applications of molecularly imprinted polymers

Template	Polymer composition	Polymerization type	Comments
$\beta$ -Lactam antibiotics	pABA, 1,2-DCE, SG	Silica gel	Liquid chromatography and preconcentration
Nafcillin	MTMOS, TMOS, PTMOS	Monolithic and film	Room temperature phosphorescence sensor based on imprinted xero-gels
Pyrimethamine	MAA, EGDMA	Thermal, monolithic	Bulk acoustic wave sensor
Diazepam	MAA, EGDMA	UV, monolithic	Radioligand binding assay
Propranolol	MAA, TRIM	UV, film	Enantioselective sensor based on a quartz-crystal microbalance
Creatine and creatinine	AM, o-PA, EGDMA, ACC	Thermal, monolithic	Computationally designed polymers worked better than trial-and-error ones
Ganglioside/sialic acid	pVB, TMAEMA, EGDMA, HEMA	UV, covalent, thin layer	Surface plasmon resonance sensor
Flavonol	MAA, EGDMA	Thermal, monolithic	Optical fluorescent sensor
3-Chloro-1,2-propanediol	4-VPB, EGDMA	Thermal, covalent, monolithic	Potentiometric chemosensor
Trichloroacetic acid	4-VP, EGDMA	Thermal	Conductometric sensor

pABA, *para*-aminobenzoic acid; 1,2-DCE, 1,2-dichloroethane; SG, silica gel; MTMOS, methyl trimethoxysilane; TMOS, tetramethoxysilane; PTMOS, phenyl trimethoxysilane; MAA, methacrylic acid; EGDMA, ethylene glycol dimethacrylate; TRIM, trimethylolpropane trimethacrylate; AM, allyl mercaptane; o-PA, *ortho*-phthalic dialdehyde; ACC, azobis (cyclohexane carbonitrile); pVB, *para*-vinylbenzene boronic acid; TMAEMA, *N,N,N*-trimethylaminoethyl methacrylate; HEMA, 2-hydroxyethyl methacrylate; 4-VPB, 4-vinylphenylboronic acid; 4-VP, 4-vinylpyridine.



biomimetic competitive assay for chloramphenicol. A fluorescent-labeled dummy template (a structurally related chloramphenicol molecule) competes with unlabeled template for binding to a limited amount of MIP. The chloramphenicol-imprinted polymer was placed into the flow cell, just before the optical path. A fluorescent-labeled dummy template was kept at a constant concentration in the mobile phase of the FIA system, so that recognition sites of the polymer were saturated with the fluorescent derivative. Injected samples were spiked with the labeled chloramphenicol in order to maintain its concentration constant all along the carrier stream. When samples reached the MIP, a fraction of the fluorescent derivative bound to the polymer was displaced by the chloramphenicol of the sample, yielding an increase in the fluorescence derivative concentration downstream, and a fluorescent peak was observed whose intensity was related to the total amount of chloramphenicol. Then, fluorescent-labeled dummy molecules from the carrier began to displace slowly the chloramphenicol bound to the polymer, so that the fluorescent signal decreased to a level corresponding to the amount of chloramphenicol. Finally, as the fluorescent derivative flow continued, the signal reached the baseline and the system was ready for the next assay.

### Sensors

MIPs can be used as recognition elements in sensors in the same way as other chemical or biochemical

recognition elements are used. There are different possibilities of transducing the binding event and different approaches to design the sensing system. **Table 4** depicts different transduction and configuration schemes for a variety of analytes.

*See also:* **Polymers:** Synthetic. **Sensors:** Overview.

### Further Reading

- Al-Kindy S, Badía R, Suárez-Rodríguez JL, and Díaz-García ME (2000) Molecularly imprinted polymers and optical sensing applications. *Critical Reviews in Analytical Chemistry* 30(4): 291–309.
- Andersson LI (2000) Molecular imprinting: developments and applications in the analytical chemistry field. *Journal of Chromatography B* 745(1): 3–13.
- Batra D and Shea KJ (2003) Combinatorial methods in molecular imprinting. *Current Opinion in Chemical Biology* 7: 434–442.
- Brüggemann O, Haupt K, Ye L, Yilmaz E, and Mosbach K (2000) New configurations and applications of molecularly imprinted polymers. *Journal of Chromatography A* 889(1–2): 15–24.
- Komiyama M, Takeuchi T, Mukawa T, and Asanuma H (2003) *Molecular Imprinting: From Fundamentals to Applications*. Chichester: Wiley-VCH.
- Sellergren B (ed.) (2003) *Molecularly Imprinted polymers, Man-made Mimics of Antibodies and their Applications in Analytical Chemistry*. Amsterdam: Elsevier.

## MULTIVARIATE CALIBRATION

*See* **CHEMOMETRICS AND STATISTICS: Multivariate Calibration Techniques**

## MULTIVARIATE CLASSIFICATION

*See* **CHEMOMETRICS AND STATISTICS: Multivariate Classification Techniques**

## MYCOTOXINS

*See* **FOOD AND NUTRITIONAL ANALYSIS: Mycotoxins. TOXINS: Mycotoxins**

# N

## NEAR-INFRARED SPECTROSCOPY

See **INFRARED SPECTROSCOPY: Near-Infrared**

## NEPHELOMETRY

See **SPECTROPHOTOMETRY: Turbidimetry and Nephelometry**

## NERVE AGENTS

See **CHEMICAL WARFARE AGENTS**

## NEUROTOXINS

See **TOXINS: Neurotoxins**

## NEUTRON ACTIVATION

See **ACTIVATION ANALYSIS: Neutron Activation**

## NITRIC OXIDE

**E E H van Faassen**, Utrecht University, Utrecht, The Netherlands

**A F Vanin**, Semenov Institute of Chemical Physics, Moscow, Russia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

In living organisms, the release of free radicals like superoxide or hydroxyl is usually highly cytotoxic and disruptive for normal physiological activity. The

volatile nitric oxide (NO) radical stands out as an important exception since its regulatory function in vascular tone was recognized in the late 1980s. Subsequent research has shown its regulatory and messenger function in a surprisingly wide variety of physiological processes. This small and uncharged radical has an unusually long lifetime of the order of hours in oxygenated water at physiological temperatures and undergoes very rapid diffusion. Its electrical neutrality allows NO to cross through cellular membranes easily and it partitions preferably in the

apolar lipid or protein compartment of a biological sample. In tissues, the NO lifetime can be orders of magnitude shorter due to the presence of scavenging compounds like oxyhemoglobin or superoxide and actual NO levels are determined by the continuous competition between production and consumption of NO. Under normal physiological conditions the lifetime is of the order of seconds, and the diffusion constant of  $\sim 3 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$  gives NO a diffusion range of  $\sim 170 \mu\text{m}$ . As such, it spans many cell diameters, which is essential for its role in physiological signaling to cellular targets at a distance. In living organisms, the main four functions of NO are vasodilatation, neuronal signal transduction, respiratory control, and immunologic defense. In addition, metabolites of NO have been implicated in a wide variety of human diseases.

The main sources of NO are enzymatic, although nonnegligible quantities of NO may result from nonenzymatic reactions involving nitrite ions, organic nitrates, and endogenous nitrosated or nitrosylated compounds. NO can have effects either by acting directly as NO or indirectly via reactive NO metabolites. At low NO concentrations, few metabolites are formed and direct action dominates. Higher NO levels lead to a considerable release of the more reactive metabolites and indirect mechanisms dominate the physiological response. The actual NO levels vary considerably with tissue type and physiological conditions. In unstimulated tissues, NO levels are difficult to measure and seem to be in the  $10\text{--}100 \text{ nmol l}^{-1}$  range. Upon stimulation with bradykinin, calcium ionophores, or cytokines, levels may rise to  $0.1\text{--}2 \mu\text{mol l}^{-1}$ . Particularly high levels of  $1\text{--}5 \mu\text{mol l}^{-1}$  are induced by ischemia in brain tissues. Still higher levels may result from septic shock or rejection of tissue transplants (Table 1).

## Endogenous Sources of NO

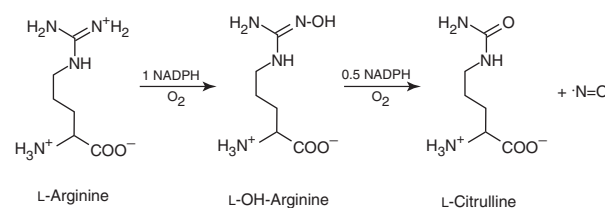
In mammals, NO regulates four crucial physiological functions: vasodilatation, neuronal signal transduction, immune response against xenobiotics, and mitochondrial respiration. For each of these four functions, mammals have developed specialized isoforms of NO synthase, which synthesizes NO from the substrate L-arginine (Figure 1).

Considerable homology in the genomic structure of the isoforms suggests that they evolved from a common ancestral gene (Figure 2 and Table 2).

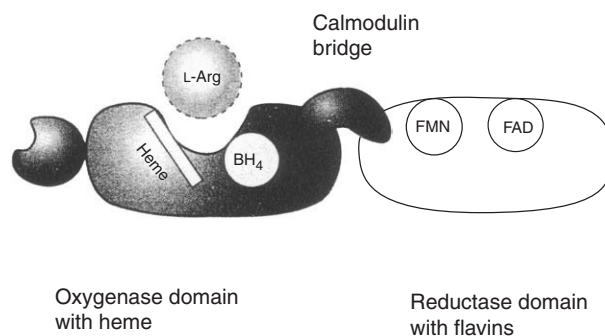
The endothelial (eNOS), neuronal (nNOS), and mitochondrial (mtNOS) isoforms are constitutively expressed and their NO release is controlled via the intracellular  $\text{Ca}^{2+}$  concentration and the enzyme cofactor calmodulin. Modest differences in reactivity

**Table 1** Tissue levels of NO after stimulus

Tissue	Stimulus	NO level ( $\mu\text{mol l}^{-1}$ )
Aortic endothelium	Bradykinin	0.3–1.0
Aortic smooth muscle	Bradykinin	0.1–0.8
Human veins	Bradykinin	0.01–0.1
Mesenteric arteries	Calcium ionophore	0.2–0.4
Brain	Electric	0.01–0.08
Brain	Ischemia	1.0–5.0
Mouse liver	Calcium ionophore	0.14
Mouse kidney	Calcium ionophore	0.20
Mouse placenta	Calcium ionophore	0.15



**Figure 1** Nitric oxide synthases synthesize NO by transforming L-arginine into L-citrulline in a two-step oxidation process at the heme domain. The reaction is driven by consumption of NADPH at the flavin reductase domain.

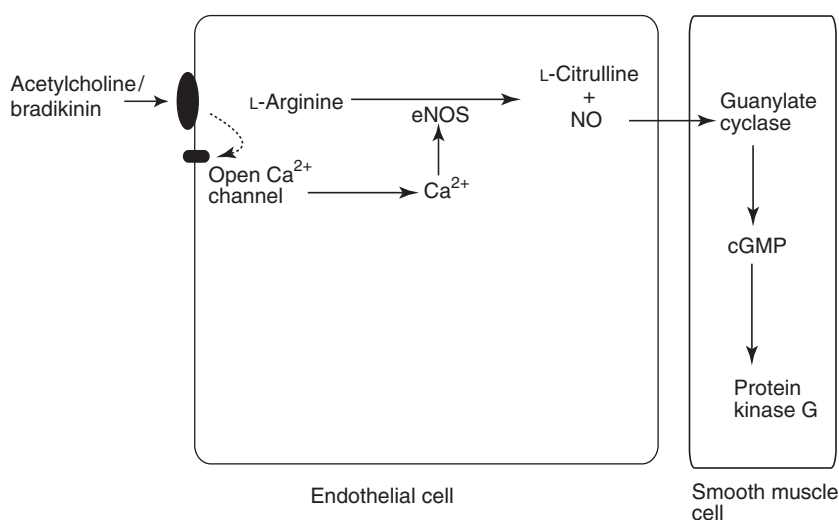


**Figure 2** Basic modular structure of all mammalian NOS isoforms. The calmodulin module is needed for electron transfer from the reductase to the heme domain and its adherence is regulated by the cytosolic  $\text{Ca}^{2+}$  concentration. The enzyme is active only in dimeric form.

to antibodies distinguish brain mtNOS from that in liver, thymus, and kidney. The gene for inducible iNOS is activated by various inflammatory cytokines, endotoxins, or lipopolysaccharides. Once formed, iNOS functions with unregulated high output and does not depend on the cytosolic  $\text{Ca}^{2+}$  concentration. Being located in mitochondria, mtNOS coexists with nNOS and eNOS in neuronal and endothelial cells, respectively. The NOS isoforms can synthesize NO only when arranged in a large protein cluster containing two NOS monomers and two calmodulin proteins and supplemented with the

**Table 2** Nomenclature and properties of the isoforms of the nitric oxide synthase enzyme

<i>Isoform</i>	<i>Acronyms</i>	<i>Activity</i>	<i>MW of human monomer (kDa)</i>	<i>Size (# amino acids)</i>	<i>Main location</i>
Endothelial NOS	eNOS, NOS III	Constitutive, $\text{Ca}^{2+}$ -regulated	133	1203	Endothelium, myocytes
Neuronal NOS	nNOS, NOS I	Constitutive, $\text{Ca}^{2+}$ -regulated	161	1434	Nerve tissue, adrenal glands, sympathetic ganglia
Mitochondrial NOS	mtNOS	Constitutive, $\text{Ca}^{2+}$ -regulated	130–145	Unknown	Mitochondria, varies with tissue type
Inducible NOS	iNOS, NOS II	Inducible, unregulated	131	1153	Macrophages, glomerulus, proximal tubule, cortical collecting duct



**Figure 3** The cascade for vasodilation in mammals. Acetylcholine or bradykinin activates endothelial membrane receptors and activates the eNOS enzyme by enhancing the intracellular calcium levels. The NO diffuses into the adjacent smooth muscle cells, where it stimulates the guanylate cyclase enzyme to produce cGMP. The latter activates protein kinase G, which stimulates outflux of  $\text{K}^+$  and  $\text{Ca}^{2+}$  from the cell. The drop in intracellular calcium triggers the relaxation of the muscle tissue. Alternatively, the guanylate cyclase in the smooth muscle cells may be activated by pharmaceutical NO donors in the bloodstream.

cofactor tetrahydrobiopterin. However, the enzyme may release substantial quantities of superoxide radicals under pathological conditions such as depletion of the substrate arginine, or after disruption of the [Zn-4S] cluster by peroxynitrite. This anomalous function is known as ‘uncoupling’ and is believed to be the initial step in a cascade leading to atherosclerosis and vascular dysfunction.

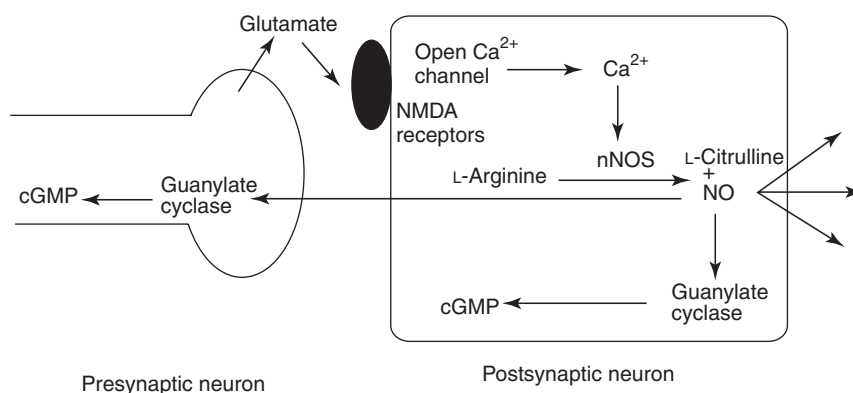
Nitrite and nitrosothiols have been identified as nonenzymatic sources of NO under physiological conditions, with a slow release of NO being catalyzed by spurious quantities of metal ions. Under normal conditions, this release is negligible with respect to enzymatic sources of NO. A possible physiological relevance for nitrosothiols as NO transporters is being discussed but controversial. Nitrate itself will release NO only under very unphysiological reductive conditions, but certain organic nitrates like

nitroglycerin may act as vasodilators in the presence of spurious metals and free cysteine.

## Physiological Targets of NO

The main function of eNOS is to trigger vasodilation by the activation of the guanylate cyclase enzyme in the smooth muscle cells of the contractile tissue surrounding the blood vessels (Figure 3). The main function of nNOS is the regulation of synaptic signal transduction between neuronal cells. The NO diffuses to adjacent neuronal cells and activates the synthesis of cGMP by the guanylate cyclase enzyme (Figure 4).

The main function of mtNOS is regulation of the mitochondrial respiration by inhibition of cytochrome C oxidase. This dimeric enzyme is located on the inner membrane of the mitochondrion and



**Figure 4** The cascade for synaptic signal transduction between neuronal cells. Release of the synaptic transmitter glutamate enhances the intracellular calcium levels and activates the nNOS enzyme. The NO can diffuse to adjacent neurons and activate the guanylate enzyme on both sides of the synaptic junction.

**Table 3** Reaction rates between NO and various physiologically relevant radical species, in aqueous environment and at room temperature

Radical	Reaction product	Reaction rate ( $\text{mol}^{-1} \text{s}^{-1}$ )	Remarks
$\text{O}_2^-$ , superoxide	$\text{ONOO}^-$	$6.7 \times 10^9$	Weak acid HOONO decomposes to nitrate
OH, hydroxyl	$\text{HNO}_2$	$1 \times 10^{10}$	Strong acid
$\text{NO}_2$ , nitrogen dioxide	$\text{N}_2\text{O}_3$	$2 \times 10^9$	Reacts with water to form nitrite
ROO, alkyl and alkoxy radicals	ROONO	$2 \times 10^9$	

The reaction products are not radicals

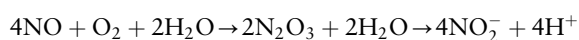
catalyzes the final step in the electron transport chain of mitochondrial respiration. It is inhibited by NO that competes with dioxygen for reversible binding to the catalytic binuclear  $\text{Cu}_B/\text{Heme}_{a3}$  binding site.

The main function of iNOS is the disabling of xenobiotics and tumor cells. NO released by macrophages and mast cells causes sufficient DNA damage in the attacked cells to induce apoptosis. The immunological potency of iNOS is further enhanced by the simultaneous release of substantial amounts of superoxide radicals, which cause oxidative damage.

The synthesis of prostaglandin and thromboxane hormones by cyclooxygenase enzymes is another important target of NO. With the exception of red blood cells, all mammalian cells release prostaglandins, which have an important role in inflammatory response, formation of blood clots, regulation of blood pressure, and muscle contraction.

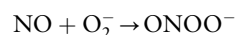
## Main Reaction Pathways for NO

Biological tissues provide many reaction pathways for the NO radical. It may react with ubiquitous oxygen to the unstable dinitrogen trioxide, which rapidly hydrolyzes to nitrite



The overall reaction has second-order dependence on the NO concentration. In water, the modest rate constant of  $\sim 3 \times 10^6 \text{ mol}^{-2} \text{s}^{-1}$  makes this pathway negligible at normal physiological levels of oxygen and NO (below  $1 \mu\text{mol l}^{-1}$ ). However, this reaction may be accelerated by orders of magnitude by the presence of a hydrophobic lipid compartment. The local NO concentration is increased 10-fold because the uncharged NO molecule preferentially partitions into the nonpolar compartment.

NO reacts very rapidly with other free radicals like superoxide, ethyl alcohol radicals, or organic peroxy radicals (Table 3). Physiological antioxidant defenses keep levels of such free radicals very low under normal conditions, but superoxide levels, in particular, can be dangerously increased after ischemia, acidosis, sepsis, or septic shock. The reaction leads to the formation of the noxious oxidant peroxynitrite:



Although not a free radical itself, peroxynitrite is strongly oxidizing and capable of nitrosating proteins. It also disrupts [Zn-4S] clusters as they exist in functional NOS dimers, for example. Additionally, with  $\text{pK}_a = 6.8$  at  $37^\circ\text{C}$ , a sizable fraction of peroxynitrite will be protonated to  $\text{ONOOH}$ , which is known to decompose via an uncertain mechanism yielding highly reactive OH radicals.



**Table 4** Nitric oxide and its physiologically relevant metabolites

Species	Name	Gibbs energy of formation (kJ mol <sup>-1</sup> )
NO	Nitric oxide radical	102
NO <sup>+</sup>	Nitrosyl cation	219
NO <sub>2</sub>	Nitrogen dioxide radical	63
NO <sub>2</sub> <sup>-</sup>	Nitrite	-32.2
NO <sub>3</sub>	Nitrogen trioxide radical	131
NO <sub>3</sub> <sup>-</sup>	Nitrate	-108.7
N <sub>2</sub> O	Dinitrogen monoxide	113.6
ONOO <sup>-</sup>	Peroxynitrite	42

In biological systems, the presence of NO may affect proteins via nitration, nitrosation, or nitrosylation. Nitration involves the addition of NO<sub>2</sub><sup>+</sup>. Although NO itself is not an effective nitrating agent, its metabolites like peroxynitrite or the NO<sub>2</sub> radical may transform tyrosine residues into nitrotyrosine (Table 4).

The formation of this nitrated residue is dramatically increased under pathological conditions, and tyrosine nitration is used as a diagnostic marker for the simultaneous release of reactive oxygen and nitrogen species. The tyrosine nitration can strongly affect the function of the parent protein as it inhibits the tyrosine phosphorylation. The latter is an important mechanism of enzymatic regulation. Nitration by NO<sub>2</sub> radicals may transform amines into *N*-nitrosamines. Nitrosation involves the addition of an equivalent NO<sup>+</sup>. Sulfhydryl groups or cysteine residues are the main targets for nitrosative transformation to *S*-nitrosothiols (RSNO) in a slow and reversible reaction. In the presence of an electron acceptor, e.g., a transition metal ion, the reaction proceeds as



The RSNO groups are rather unstable and may release NO under ultraviolet light or in the presence of trace metal ions like monovalent Cu<sup>+</sup>. The nitrosyl cation NO<sup>+</sup> is a potent nitrosating agent also, but cannot exist under aqueous physiological conditions. Finally, NO may form nitrosyl complexes by binding to catalytic or structural metal sites in proteins. This liganding often competes with small ligands like dioxygen or carbonmonoxide. Fast and reversible binding performs a crucial regulatory role in vascular relaxation (heme of guanylate cyclase) or mitochondrial respiration (the dinuclear Cu<sub>B</sub>/Heme<sub>a3</sub> center of cytochrome C oxidase). However, in most other proteins the binding of NO to iron is inhibitory as in cytochrome P<sub>450</sub>, or structurally disruptive as in aconitases.

NO easily binds to transition metal complexes, in particular iron(II) and iron(III) complexes. In many cases, the binding of NO is sufficiently strong to displace other ligands and rearrange the complex. The disruption of enzymatic [Fe-S] clusters ultimately leads to the formation of paramagnetic dinitrosyl-iron complexes (DNICs) of the form (RS)<sub>2</sub>Fe<sup>2+</sup>(NO)<sub>2</sub>. NO also disrupts the structure of noniron metal centers like the [Zn-S] clusters in zinc finger proteins, thereby disabling a class of proteins involved in reparation of oxidative DNA damage. In the vascular system, hemoglobin (Hb) in the erythrocytes is a very efficient scavenger of NO. In oxygenated blood, 99% of the hemoglobin carries a dioxygen ligand and reacts rapidly and irreversibly with NO to methemoglobin (metHb) and nitrate. In the absence of a dioxygen ligand, NO binds rapidly but reversibly to form a stable mononitrosyl-iron complex with the heme. These reactions significantly reduce the NO concentrations in the muscle tissue surrounding the blood vessel, where NO acts as a signaling molecule for the activation of cGMP synthesis by the guanylate cyclase enzyme.

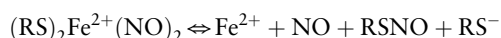
## Nitric Oxide in Clinical Pathology

Many hypertensive disorders reflect insufficient NO concentrations in the vascular system and are clinically treated by supplementation of NO. Pharmacological NO donors are effective against hypertension and angina pectoris, and pulmonary hypertension may be treated by low NO dosages in inhaled air. Macrophages kill bacteria and tumor cells by releasing large quantities NO from iNOS. However, extreme conditions of septic shock, sepsis, or cytokines may stimulate an excessive release of NO and cause an acute and profound drop in blood pressure (hypotension). Typical cases are induced by the endotoxins released after bacterial infections of abdominal traumas, by acute tissue rejection after organ transplantation or transfusion, or by chemotherapeutic treatment of neoplasias. Low-level inflammatory diseases like colitis, pancreatitis, and arthritis are associated with chronic overexpression of iNOS and can be treated by administration of specific iNOS inhibitors. On longer time-scales of years or decades, the proper function of eNOS is clearly correlated with good vascular function and low risk of hypertension, atherosclerosis, thrombosis, and stenosis. Cerebral NO production has been implicated in the brain damage caused by a temporary insufficiency of the oxygen supply (hypoxia-ischemia). Upon reoxygenation, the reactivation of the respiratory chain causes a massive release of superoxide radicals, which react with endogenous NO to form the harmful peroxynitrite. This reactive metabolite causes

subsequent oxidative and nitrosative damage to cerebral tissues. The formation of peroxynitrite may be reduced by supplementation of NOS inhibitors.

## Modulation of NO Levels *In Vivo*

Systemic NO levels in living animals may be raised in various ways: gaseous NO can be applied to exposed body surfaces like skin and lung tissue, and finds clinical application in accelerated healing of skin lesions and for treatment of pulmonary hypertension in newborn babies. Internal tissues can be reached with pharmaceutical NO donors like nitroglycerin, isoamyl nitrite, or NONO-ates that release NO in a controlled way. Paramagnetic DNICs with thiolate ligands  $(RS)_2Fe^{2+}(NO)_2$  decompose by a slow and reversible release of NO and a nitrosothiol:



At higher concentrations, nitrosothiols themselves may act as a donor via a metal-catalyzed formation of disulfide bonds:



Nitrosothiols are difficult to control as donors, since the release of NO is catalyzed by traces of free metal ions, in particular copper(I)  $Cu^+$ . The endogenous synthesis of NO may be enhanced by activating existing nitric oxide synthases with supplementation of co-factors like tetrahydrobiopterin, with folic acid or vitamin C, or by phosphorylation of certain serine groups of the enzyme. Alternatively, the expression of NOS enzyme in tissues or cell cultures may be artificially enhanced by gene therapy, i.e., genetic transfection with viral vectors. Finally, under favorable conditions NO levels may be increased by blocking the

main reaction channel for NO: removing superoxide radicals may dramatically enhance NO levels in animal tissues and facilitate the detection of NO.

Indiscriminate inhibition of all NOS isoforms is achieved by imidazole, *N*-methyl-L-arginine, L-nitro-arginine methyl ester, or by certain small proteins like PIN. Isoform-specific inhibition has only been achieved for iNOS, with inhibitors like aminoguanidine, aminoethyl-isothiourea, L-*N*-6-lysine or 1400W. In particular, the search for an effective and selective nNOS inhibitor has remained fruitless.

## Methods of NO Detection in Biological Systems

NO has an unusually long lifetime of the order of hours in oxygenated water at room temperature. However, in biological systems, scavengers like hemoglobin or superoxide may shorten this lifetime by orders of magnitude and make it very difficult to detect and quantify NO *in vivo* or *in vitro* directly. Thus, we can classify the detection techniques into three classes. The first class studies the presence of NO indirectly via its effect on the systemic response of a fully functional organism or tissue section. The second class detects reaction products or metabolites of NO. The third class involves direct detection of NO itself (Table 5).

### Detection via Systemic Response

**Vasorelaxation in vessel rings** NO may be detected through observation of vasorelaxation if the concentration exceeds the activation threshold of the guanylate cyclase enzyme. Quantification of NO is not possible and a vasoreactive response is also caused by impostors like Endothelium Derived Hyperpolarizing

**Table 5** Methods for the detection of nitric oxide in biological samples

Method	Material	Detection limit
Vasorelaxation in extracted tissues	Vessel rings ( <i>ex vivo</i> )	Nanomolar
Vasodilation	<i>In vivo</i> (human)	Nanomolar
Inhibition of platelet adhesion	Cultured cells	Nanomolar
Spectrophotometric with Griess reaction	Aqueous (urine, cerebrospinal fluid, plasma, cell culture medium)	Micromolar
Arginine to citrulline conversion	Aqueous, tissue	Nanomolar
Cyclic GMP	Aqueous, tissue	Nanomolar
Protein nitrosylation	Aqueous, tissue	Nanomolar
Histochemistry	Tissue (including imaging)	NA
Hemoglobin oxidation	Aqueous	Nanomolar
Fluorimetric	Aqueous, cultured cells (including imaging), tissue (including imaging)	Nanomolar
Chemiluminescence (gas phase)	Exhaled air, aqueous	Nanomolar
Chemiluminescence (liquid phase)	Tissue, aqueous	Picomolar
Electrochemical	Aqueous, <i>in vivo</i> (including humans), cultured cells	Nanomolar
Resistive NO sensors	Gas phase	10 ppm
Electron spin trapping	Aqueous, tissue, <i>in vivo</i> (animal)	Micromolar

NA, not applicable.

Factor, or substances that raise intracellular calcium levels and thereby artificially activate the constitutive NOS isoforms. The method may be applied to veins extracted from patients and laboratory animals but is impractical for diagnostic clinical use.

***In vivo* vasodilation** Two methods observe the release of NO by the endothelium by measuring changes in blood flow in humans. In flow-mediated dilatation, the diameter of the brachial artery is measured with ultrasound. The vasodilation is induced by the increase of arterial shear stress after the release of an inflatable cuff around the arm. Plethysmography measures the blood flow via the volume changes of the forearm after venous occlusion. Quantification of NO is not possible in either method and the diagnostic value is limited as individuals show considerable variations in systemic response.

**Inhibition of platelet adhesion** NO is a potent inhibitor of the aggregation of blood platelets, and the deceleration of the kinetics of platelet aggregation can be used as a qualitative but sensitive indicator for the presence of NO.

### Indirect Detection of NO

**Spectrophotometry with Griess reaction** One of the most widely used methods for NO detection is the spectrophotometric determination of the stable oxidation products of NO, in particular nitrite ( $\text{NO}_2^-$ ) and nitrate ( $\text{NO}_3^-$ ).  $\text{NO}_2^-$  is determined by the Griess coloring reaction and monitored via optical absorption spectroscopy at a wavelength of 540 nm.

**Arginine to citrulline conversion** The mammalian enzymes arginase and NOS both transform arginine into citrulline (see Figure 1). Still, the contribution of NOS to the arginine consumption may be unambiguously quantified via site-specific isotopic labeling of the arginine. The labeling site is chosen such that the label is removed from the citrulline if processed via the arginase pathway, but retained if the substrate is consumed by NOS. The isotopic labels may be unstable like  $^3\text{H}$  and  $^{14}\text{C}$  or stable like  $^{15}\text{N}$ . The citrulline yield may be quantified via column separation or liquid chromatography in combination with mass spectrometry. Alternative methods measure the radioactivity of the label using liquid scintillation counters. Radioactive isotopes have excellent detection sensitivity against a background of negligible natural abundance but require specialized and expensive lab facilities. In contrast, stable isotopes like  $^{15}\text{N}$  are simple and safe to use but suffer from high natural background levels.

**Levels of cGMP and activation of guanylate cyclase** NO regulates the activity of the cytosolic enzyme guanylate cyclase. This heterodimeric heme enzyme converts GTP into cGMP and pyrophosphate. The enzymatic yields of guanylate cyclase may be determined via isotopic labeling of the substrate. Quantification of the NO levels is not possible because the regulation of guanylate cyclase is not fully understood and because the activity of guanylate cyclase is affected by other compounds like C-type natriuretic peptide. Levels of cGMP may be determined in human body liquids like plasma or urine.

**Nitration or nitrosation of proteins** NO metabolites like  $\text{ONOO}^-$  or the  $\text{NO}_2$  radical can nitrate proteins by transforming tyrosine residues into characteristic nitrotyrosine groups. The degree of protein nitration can be determined with chromatographic and immunological techniques in tissue, plasma, and urine, and may serve as an indirect measure for the precursor molecule NO. The susceptibility for nitration varies with protein and requires the presence of other reactive oxygen species as NO itself is not a nitrating agent. Protein nitration is an accumulating process, as nitrotyrosine groups continue to be formed as long as reactive NO metabolites are released. Sulfhydryl groups or cysteine residues may be nitrosated reversibly to S-nitrosothiols. The latter are rather unstable and may be detected via their slow metal-catalyzed release of NO, or by biotinylation and subsequent gas chromatography-mass spectrometry.

**Histochemistry** Several NADPH consuming enzymes including NOS show an activity similar to NADPH diaphorase. The latter uses NADPH to convert soluble tetrazolium salts into insoluble visible formazan. Nitro blue tetrazolium is a widely used blue stain for NADPH diaphorase activity, and it has been observed that staining correlates well with NOS activity. The NOS contribution to the staining cannot be quantified, as several enzymes show diaphorase activity.

### Direct Detection of NO

**Hemoglobin oxidation** NO reacts rapidly with reduced oxyhemoglobin,  $\text{HbFe}^{2+}\text{O}_2$ , to form methemoglobin ( $\text{HbFe}^{3+}$ ) and  $\text{NO}_3^-$ . The oxidation state of the heme group is easily followed spectrometrically at wavelengths near 400 nm. A variation of this technique is the use of the iron-containing enzyme horseradish peroxidase or oxymyoglobin instead of oxyhemoglobin.

**Fluorimetric trapping** Fluorimetric trapping is based on the reaction of a suitable trap-molecule

with NO or nitrite with a fluorescent reaction product called 'adduct'. Suitable traps are those where the adduct has fluorescent properties very distinct from those of the original trap. Easiest to use are nonfluorescent traps with fluorescent adducts like 2,3-diaminonaphthalene, which can detect  $\text{NO}_2^-$  in cell culture medium, plasma, and urine but not in tissues. Fluorescent dyes like dihydrorhodamine and dichlorofluorescein (DCF) change their fluorescence spectra upon reaction with NO, but similar spectral changes are caused by reactive oxygen or nitrogen species other than NO. This lack of specificity makes these traps unsuitable for unambiguous determination of NO. The fluorescent trap diaminofluorescein (DAF) has been widely used in cultured cells, flow cytometry, and NO imaging. The diacetate variant DAF-2-DA is electrically neutral, cell permeable, and accumulates in viable cells as cellular esterases hydrolyze the acetate groups into negatively charged acid residues. DAF is not specific as it may react with NO imposters like peroxynitrite,  $\text{N}_2\text{O}_3$ , and even ascorbic acid. Additionally, catecholamines and certain antioxidants interfere in the reaction with NO. Good specificity for NO is reported for cheletropic nitric oxide traps (FNOCT). This red fluorescent trap reacts with NO to a blue-fluorescent adduct and has been used to determine NO levels in cell cultures and tissues (Figure 5).

**Chemiluminescence in the gas phase** In the gas phase, ozone will react with NO to form electronically excited nitrogen dioxide radicals. Upon relaxation some emit visible light, which may be easily

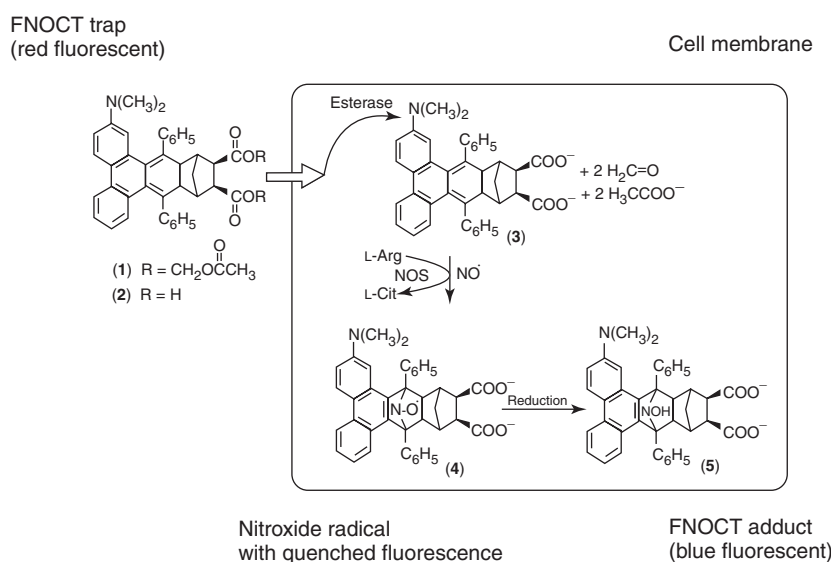
and sensitively detected. This method has been used to detect NO in exhaled air in chronic renal or hemodialytic patients. In aqueous media, dissolved NO may be released into the gaseous phase by bubbling with an inert gas like nitrogen or argon. NO metabolites like nitrate, nitrite, or nitrosothiols can be quantified by reductive reconversion into NO. For liquid samples, the sensitivity for NO detection has been reported at  $\sim 100 \text{ pmol l}^{-1}$ . The method can only be used on gaseous or aqueous samples like cell culture media, plasma, and urine, but NO levels inside cells or tissues remain inaccessible.

**Chemiluminescence in the liquid phase** Chemiluminescent markers may be used to detect radical molecules like superoxide or NO. Luminol in particular reacts with NO to give a metastable intermediate, which relaxes by light emission at 427 nm. The method is sensitive, but unspecific, as the same emission is also observed with a range of other reactive oxygen and nitrogen species. In addition, the method cannot be used with optically dense samples like tissue.

**Electrochemical detection with sensing electrodes** Electrodes can detect NO in the liquid phase via the redox couple:



The electrical current generated at the anode is directly proportional to the NO concentration in solution. Electrochemical electrodes have been used to detect NO released by cultured cells or isolated



**Figure 5** The mechanism of NO detection by cheletropic traps. Cellular esterases transform the membrane penetrating FNOCT ester into an acid that accumulates inside the cell. The red-fluorescent FNOCT trap reacts with NO to an unstable nonfluorescent nitroxide radical. The latter is hydroxylated to give a stable blue-fluorescent adduct.

tissues and even directly in the human vasculature. Electrochemical detection determines the NO concentrations and thus the NO bioavailability, not the actual NO production. The selectivity for NO may be greatly enhanced by modification of the electrode surface with polymeric ion exchangers. Spatial resolution is obtained with electrochemical micro-sensors where thin (5–10  $\mu\text{m}$ ) carbon fibers receive a surface cover of polymerized phthalocyanines or porphyrins and ion-selective Nafion layers. Concentration gradients of NO may be measured with oscillating microsensors. Interference from nitrite or catecholamines is problematic, the electrodes are difficult to handle, and the covering membranes are very sensitive to environmental influences and mechanical damage.

**Electron paramagnetic resonance (ESR)** Although NO is a paramagnetic radical with spin  $S=1/2$ , its degenerate ground state makes it impossible to detect directly with electron paramagnetic spin resonance spectroscopy. But its fairly long lifetime allows its capture and stabilization by diamagnetic spin trapping compounds. Certain iron complexes in particular have high affinity for NO ligands and form stable mono- or dinitrosyl-iron complexes, which are paramagnetic and may be quantified with ESR. This method is particularly valuable for NO detection in tissues.

*See also:* **Chemiluminescence:** Overview. **Chromatography:** Principles. **Clinical Analysis:** Overview. **Electron Spin Resonance Spectroscopy:** Biological Applications. **Fluorescence:** Overview. **Ion-Selective Electrodes:** Overview. **Mass Spectrometry:** Overview. **Microscopy:** Overview. **Nuclear Magnetic Resonance Spectroscopy:** Overview. **Ozone.** **Radiochemical Methods:** Overview. **Sensors:** Overview. **Spectrophotometry:** Overview.

## Further Reading

- Cadenas E and Packer L (eds.) (2002) *Nitric Oxide, part D: Nitric Oxide Detection, Mitochondria and Cell Functions, and Peroxynitrite Reactions. Methods in Enzymology*, vol. 359. London: Academic Press.
- Feelisch M and Stamler J (eds.) (1996) *Methods in Nitric Oxide Research*. New York: Wiley.
- Henry Y, Guissani A, and Ducastel D (1996) *Nitric Oxide Research from Chemistry to Biology: EPR Spectroscopy of Nitrosylated Compounds*. Austin: Landes Bioscience.
- Loscalzo J and Vita J (2000) *Nitric Oxide and the Cardiovascular System*. Totowa, NJ: Humana Press.
- Packer L (ed.) (1999) *Nitric Oxide, part C; Biological and Antioxidant Activities. Methods in Enzymology*, vol. 301. London: Academic Press.
- Rosen GM, Britigan BE, Halpern HJ, and Pou S (1999) *Free Radicals: Detection and Spin Trapping*. New York: Oxford University Press.

# NITROGEN

**G Hanrahan and G Chan**, California State University, Los Angeles, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Nitrogen is an essential element for all forms of life and is the structural component of amino acids from which animal and human tissues, enzymes, and many hormones are made. For plant growth, available (fixed) nitrogen is usually the limiting nutrient in natural systems. Nitrogen chemistry and overall cycling in the global environment are quite complex due to the number of oxidation states. Nitrogen itself has five valence electrons and can be found at oxidation states between  $-3$  and  $+5$ . Thus, numerous species can form from chemical, biochemical, geochemical, and biogeochemical processes. **Table 1** lists by oxidation state some of the most prevalent

nitrogen species (and important chemical data) that exist in the global environment.

## Global Nitrogen Distribution

The global distribution of nitrogen (both reservoirs and fluxes) is shown in **Table 2**. The atmosphere is the principal nitrogen reservoir, with over 99% of the total in the form of  $\text{N}_2$ . Nitrogen in terrestrial systems occurs mainly as a constituent in soil organic matter, with litter and soil inorganic nitrogen accounting for the majority (97%). Biomass accounts for just fewer than 3%; of this, 95% occurs in plant tissue. Dinitrogen, in dissolved form ( $\text{N}_{2,\text{aq}}$ ), is the most abundant nitrogen form in the world's oceans. Nitrogen also occurs in both inorganic forms (e.g., nitrate, nitrite, ammonia, hydrazine, nitrous oxide, and nitrogen dioxide) and organic forms (e.g., amino acids, amines, and amides). Amino acids are minor but important constituents of dissolved organic



tissues and even directly in the human vasculature. Electrochemical detection determines the NO concentrations and thus the NO bioavailability, not the actual NO production. The selectivity for NO may be greatly enhanced by modification of the electrode surface with polymeric ion exchangers. Spatial resolution is obtained with electrochemical micro-sensors where thin (5–10  $\mu\text{m}$ ) carbon fibers receive a surface cover of polymerized phthalocyanines or porphyrins and ion-selective Nafion layers. Concentration gradients of NO may be measured with oscillating microsensors. Interference from nitrite or catecholamines is problematic, the electrodes are difficult to handle, and the covering membranes are very sensitive to environmental influences and mechanical damage.

**Electron paramagnetic resonance (ESR)** Although NO is a paramagnetic radical with spin  $S=1/2$ , its degenerate ground state makes it impossible to detect directly with electron paramagnetic spin resonance spectroscopy. But its fairly long lifetime allows its capture and stabilization by diamagnetic spin trapping compounds. Certain iron complexes in particular have high affinity for NO ligands and form stable mono- or dinitrosyl-iron complexes, which are paramagnetic and may be quantified with ESR. This method is particularly valuable for NO detection in tissues.

*See also:* **Chemiluminescence:** Overview. **Chromatography:** Principles. **Clinical Analysis:** Overview. **Electron Spin Resonance Spectroscopy:** Biological Applications. **Fluorescence:** Overview. **Ion-Selective Electrodes:** Overview. **Mass Spectrometry:** Overview. **Microscopy:** Overview. **Nuclear Magnetic Resonance Spectroscopy:** Overview. **Ozone.** **Radiochemical Methods:** Overview. **Sensors:** Overview. **Spectrophotometry:** Overview.

## Further Reading

- Cadenas E and Packer L (eds.) (2002) *Nitric Oxide, part D: Nitric Oxide Detection, Mitochondria and Cell Functions, and Peroxynitrite Reactions. Methods in Enzymology*, vol. 359. London: Academic Press.
- Feelisch M and Stamler J (eds.) (1996) *Methods in Nitric Oxide Research*. New York: Wiley.
- Henry Y, Guissani A, and Ducastel D (1996) *Nitric Oxide Research from Chemistry to Biology: EPR Spectroscopy of Nitrosylated Compounds*. Austin: Landes Bioscience.
- Loscalzo J and Vita J (2000) *Nitric Oxide and the Cardiovascular System*. Totowa, NJ: Humana Press.
- Packer L (ed.) (1999) *Nitric Oxide, part C; Biological and Antioxidant Activities. Methods in Enzymology*, vol. 301. London: Academic Press.
- Rosen GM, Britigan BE, Halpern HJ, and Pou S (1999) *Free Radicals: Detection and Spin Trapping*. New York: Oxford University Press.

# NITROGEN

**G Hanrahan and G Chan**, California State University, Los Angeles, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Nitrogen is an essential element for all forms of life and is the structural component of amino acids from which animal and human tissues, enzymes, and many hormones are made. For plant growth, available (fixed) nitrogen is usually the limiting nutrient in natural systems. Nitrogen chemistry and overall cycling in the global environment are quite complex due to the number of oxidation states. Nitrogen itself has five valence electrons and can be found at oxidation states between  $-3$  and  $+5$ . Thus, numerous species can form from chemical, biochemical, geochemical, and biogeochemical processes. **Table 1** lists by oxidation state some of the most prevalent

nitrogen species (and important chemical data) that exist in the global environment.

## Global Nitrogen Distribution

The global distribution of nitrogen (both reservoirs and fluxes) is shown in **Table 2**. The atmosphere is the principal nitrogen reservoir, with over 99% of the total in the form of  $\text{N}_2$ . Nitrogen in terrestrial systems occurs mainly as a constituent in soil organic matter, with litter and soil inorganic nitrogen accounting for the majority (97%). Biomass accounts for just fewer than 3%; of this, 95% occurs in plant tissue. Dinitrogen, in dissolved form ( $\text{N}_{2,\text{aq}}$ ), is the most abundant nitrogen form in the world's oceans. Nitrogen also occurs in both inorganic forms (e.g., nitrate, nitrite, ammonia, hydrazine, nitrous oxide, and nitrogen dioxide) and organic forms (e.g., amino acids, amines, and amides). Amino acids are minor but important constituents of dissolved organic

**Table 1** Global nitrogen species and selected chemical data

Species	Oxidation state	b.p. (°C)	$\Delta H^{\circ}$ (f) (kJ mol <sup>-1</sup> , 298 K)	$\Delta G^{\circ}$ (f) (kJ mol <sup>-1</sup> , 298 K)
N <sub>2</sub> O <sub>5(g)</sub>	+5	11	115	
HNO <sub>3(g)</sub>	+5	83	-135	-75
HNO <sub>3(aq)</sub>	+5			
NO <sub>2(g)</sub>	+4	21	33	51
N <sub>2</sub> O <sub>4</sub>	+4		9	98
HNO <sub>2(g)</sub>	+3		-80	-46
HNO <sub>2(aq)</sub>	+3		-120	-55
NO(g)	+2	-152	90	87
N <sub>2</sub> O(g)	+1	-89	82	104
N <sub>2(g)</sub>	0	-196	0	0
NH <sub>3(g)</sub>	-3	-33	-46	-16.5
NH <sub>4</sub> <sup>+</sup> <sub>(aq)</sub>	-3		-72	-79
NH <sub>4</sub> Cl <sub>(s)</sub>	-3		-201	-203
CH <sub>3</sub> NH <sub>2(g)</sub>	-3		-28	28

**Table 2** The major global reservoirs and fluxes of nitrogen

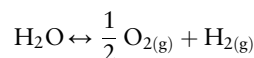
Reservoir	10 <sup>15</sup> g N
Lithosphere	190 000 000
Atmosphere	3 900 000
Oceanic	23 000
Terrestrial	500
Flux	10 <sup>15</sup> g N year <sup>-1</sup>
Dry and wet deposition (terrestrial + oceanic)	0.160–0.450
Denitrification (terrestrial)	0.043–0.390
Denitrification (oceanic)	0–0.330
Biological nitrogen fixation (terrestrial)	0.044–0.200
Biological nitrogen fixation (oceanic)	0.001–0.130
River runoff (terrestrial → oceanic)	0.013–0.040

nitrogen (DON), which are readily used by bacteria and some phytoplankton. Studies involving DON fractions isolated from seawater by solid-phase extraction and ultrafiltration, for example, have shown that D-enantiomers of some amino acids are present in significant amounts. In general, the organic fraction is not well characterized and estimated from knowledge of the carbon content and C:N ratios. Nitrous oxide (N<sub>2</sub>O) emissions arise from fertilized arable land, animal excreta, postclearing effects on soil emissions, fossil fuel, and industrial sources. N<sub>2</sub>O from animal excreta, in particular, form a significant (~50%) source of total emissions. Additional N<sub>2</sub>O is thought to arise in tropical soils through the process of denitrification, and currently, biogeochemists are devoting a large research effort to understand the true importance of this source of nitrogen.

## Activity Diagrams

Redox potential-pH diagrams are very useful as a way to visualize and summarize the aqueous speciation of redox-sensitive elements. pH is the negative base-10 logarithm of proton activity in aqueous solution,  $-\log a_{\text{H}^+}$ . Many variables are commonly employed to express redox conditions: Eh (electrode potential of a half-cell reaction with reference to standard hydrogen electrode),  $pe$  (electron activity in aqueous solution),  $f_{\text{O}_2}$  (fugacity of oxygen), and  $f_{\text{H}_2}$  (fugacity of hydrogen). They are all more or less useful depending on a specific problem, or an approach to measurements or calculations, and can be readily converted from one to the other given the appropriate conversion factors. Fugacity of hydrogen,  $f_{\text{H}_2}$ , is a convenient parameter because any redox reaction can be written to include hydrogen as a product or reactant. High values of  $\log f_{\text{H}_2}$  indicate reducing conditions whereas low  $\log f_{\text{H}_2}$  values indicate oxidizing conditions. An  $f_{\text{H}_2}$ -pH plot is a type of activity diagram that indicates both oxidation-reduction and acid-base conditions.

Figure 1 is an activity diagram for the N–O–H system at 25°C and 1 bar. The horizontal bold lines in the plot correspond to reducing (upper) and oxidizing (lower) boundaries in terms of  $\log f_{\text{H}_2}$  where H<sub>2</sub>O is stable. The dissociation reaction of water is:



where, at equilibrium

$$\log K = \log f_{\text{H}_2} + \frac{1}{2} \log f_{\text{O}_2}$$

and

$$\log f_{\text{H}_2} = \log K - \frac{1}{2} \log f_{\text{O}_2}$$

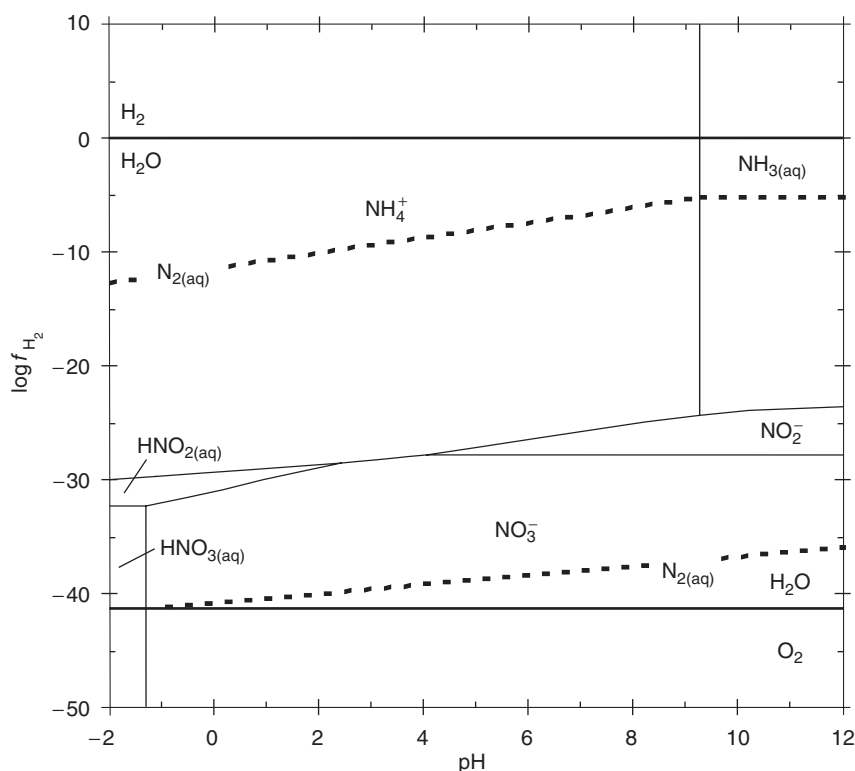
The usual atmospheric O<sub>2</sub> defines the oxidizing boundary of water stability, and we may use 0.21 bar (or  $\log f_{\text{O}_2} = -0.68$ ) since oxygen makes up ~21% of the atmosphere by volume. Combining with the equilibrium constant of reaction [14] in Table 3, this corresponds to  $\log f_{\text{H}_2} = -41.21$ . The reducing boundary occurs at 1 bar H<sub>2(g)</sub> since the pressure of hydrogen gas in surface environments cannot exceed the atmospheric pressure, and this sets the upper boundary at  $\log f_{\text{H}_2} = 0$ .

For the vertical line separating NH<sub>4</sub><sup>+</sup> and NH<sub>3(aq)</sub>,



$$\log K = -\text{pH} + \log a_{\text{NH}_3} - \log a_{\text{NH}_4^+}$$

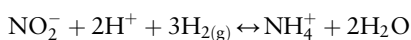
If we consider  $a_{\text{NH}_3} = a_{\text{NH}_3^+}$ , the equation reduces to  $\text{pH} = -\log K$ , or  $\text{pH} = 9.24$ . So this line corresponds



**Figure 1** Plot of  $\log f_{\text{H}_2}$  versus pH at 25°C and 1 bar total pressure showing fields of relative predominance of aqueous nitrogen species.  $\log f_{\text{H}_2}$  values represent numerical indices of the oxidation state of the system. Each line has a matching chemical reaction as listed in **Table 3**. Bold horizontal lines establish the stability limits for water. Dashed lines define the area of dissolved nitrogen in equilibrium with an atmosphere containing 78%  $\text{N}_2$ .

to the pH where the activity of  $\text{NH}_4^+$  equals the activity of  $\text{NH}_3(\text{aq})$ . At  $\text{pH} > 9.24$ , the ratio  $a_{\text{NH}_3}/a_{\text{NH}_4^+}$  is greater than 1; at any pH values left of this line, less than 1.

Consider next the redox couple between nitrite and ammonium ion,



This reaction includes both proton and hydrogen, and so the equilibrium constant expression gives:

$$\log K = \log a_{\text{NH}_4^+} - \log a_{\text{NO}_2^-} + 2\text{pH} - 3 \log f_{\text{H}_2}$$

Setting the activity of  $\text{NH}_4^+$  equal to the activity of  $\text{NO}_2^-$ , rearrangement gives  $\log f_{\text{H}_2}$  as a function of pH:

$$\log f_{\text{H}_2} = \frac{2}{3} \text{pH} - \frac{1}{3} \log K$$

This equation has a slope of 2/3 and its intercept can be evaluated from the  $\log K$  of reaction [6] in **Table 3**, and it is plotted in **Figure 1** as a line that separates the predominant areas for  $\text{NH}_4^+$  and  $\text{NO}_2^-$ . Using these methods, similar lines can be constructed for the other aqueous nitrogen species. The appropriate

reactions and their equilibrium constants are listed in **Table 3**. Also overlaid in the figure are dashed lines, derived from reactions [10] to [13], that illustrate the area of nitrogen,  $\text{N}_{2(\text{aq})}$ , dissolved in aqueous solutions in contact with the atmosphere having 78%  $\text{N}_2$ .

## Nitrogen Speciation

The activity diagram such as **Figure 1** is constructed for the N– $\text{H}_2\text{O}$  system, and in it, fields separated by lines essentially map out the predominant nitrogen species over ranges of pH and redox potential in  $f_{\text{H}_2}$ . Each field defines the redox and pH conditions in which a particular nitrogen species predominates. This does not mean that other species are unstable or nonexistent, but they are minor constituents. Moreover, the figure is drawn to show equilibrium relations tied to the equilibrium constants at a given temperature and pressure for a set of nitrogen reactions (**Table 3**). Although natural environments often do not attain equilibrium in reasonable time, an equilibrium diagram is useful to provide a backdrop against which compounds out-of-equilibrium are revealed. Such diagram can then be used as a ‘biogeochemical map’ to illustrate nitrogen speciation and a variety of biological and geochemical transformations

(see the section ‘Biogeochemical cycling processes’ below).

For example, under oxidizing conditions as in river runoffs and surface waters, we would expect to find  $\text{NO}_3^-$ . By contrast,  $\text{NH}_4^+$  or  $\text{NH}_3$  predominates in

reducing conditions; therefore, the occurrence of  $\text{NH}_4^+$  or  $\text{NH}_3$  in oxidizing environments represents species out of equilibrium, and this oxidation of  $\text{NH}_4^+$  or  $\text{NH}_3$  to  $\text{NO}_3^-$  is thermodynamically favorable. Where this oxidation reaction is hindered by slow reaction rate, microorganisms can tap into this nitrification reaction for energy (see below). The field of dissolved nitrogen  $\text{N}_{2(\text{aq})}$  occupies wide ranges of pH and redox conditions and represents many aqueous environments in contact with the atmosphere. Because  $\text{N}_{2(\text{aq})}$  is the predominant species, as shown in **Figure 1**, the conversion of  $\text{N}_2$  to other nitrogen compounds (as in biological fixation) requires the input of energy. This process is carried out by a wide variety of bacteria and algae via enzyme catalysis or photosynthesis.

**Table 3** Equilibrium constants at 25°C and 1 bar for nitrogen reactions and the dissociation reaction of water

Reaction	$\log K^a$
1 $\text{HNO}_{3(\text{aq})} \leftrightarrow \text{NO}_3^- + \text{H}^+$	1.303
2 $\text{NH}_4^+ \leftrightarrow \text{NH}_{3(\text{aq})} + \text{H}^+$	-9.241
3 $\text{NO}_3^- + \text{H}^+ + \text{H}_{2(\text{g})} \leftrightarrow \text{HNO}_{2(\text{aq})} + \text{H}_2\text{O}$	30.992
4 $\text{NO}_3^- + \text{H}_{2(\text{g})} \leftrightarrow \text{NO}_2^- + \text{H}_2\text{O}$	27.767
5 $\text{HNO}_{2(\text{aq})} + \text{H}^+ + 3\text{H}_{2(\text{g})} \leftrightarrow \text{NH}_4^+ + 2\text{H}_2\text{O}$	88.155
6 $\text{NO}_2^- + 2\text{H}^+ + 3\text{H}_{2(\text{g})} \leftrightarrow \text{NH}_4^+ + 2\text{H}_2\text{O}$	91.381
7 $\text{NO}_2^- + \text{H}^+ + 3\text{H}_{2(\text{g})} \leftrightarrow \text{NH}_{3(\text{aq})} + 2\text{H}_2\text{O}$	82.140
8 $\text{HNO}_{3(\text{aq})} + \text{H}_{2(\text{g})} \leftrightarrow \text{HNO}_{2(\text{aq})} + \text{H}_2\text{O}$	32.295
9 $\text{NO}_3^- + 2\text{H}^+ + 4\text{H}_{2(\text{g})} \leftrightarrow \text{NH}_4^+ + 3\text{H}_2\text{O}$	119.147
10 $\text{NO}_3^- + \text{H}^+ + \frac{5}{2}\text{H}_{2(\text{g})} + \frac{1}{2}\text{N}_{2(\text{g})} \leftrightarrow \text{N}_{2(\text{aq})} + 3\text{H}_2\text{O}$	102.041
11 $\text{N}_{2(\text{aq})} + \text{H}^+ + \frac{3}{2}\text{H}_{2(\text{g})} \leftrightarrow \text{NH}_4^+ + \frac{1}{2}\text{N}_{2(\text{g})}$	17.106
12 $\text{HNO}_{3(\text{aq})} + \frac{5}{2}\text{H}_{2(\text{g})} + \frac{1}{2}\text{N}_{2(\text{g})} \leftrightarrow \text{N}_{2(\text{aq})} + 3\text{H}_2\text{O}$	103.344
13 $\text{N}_{2(\text{aq})} + \frac{3}{2}\text{H}_{2(\text{g})} \leftrightarrow \text{NH}_{3(\text{aq})} + \frac{1}{2}\text{N}_{2(\text{g})}$	7.865
14 $\text{H}_2\text{O} \leftrightarrow \frac{1}{2}\text{O}_{2(\text{g})} + \text{H}_{2(\text{g})}$	-41.552

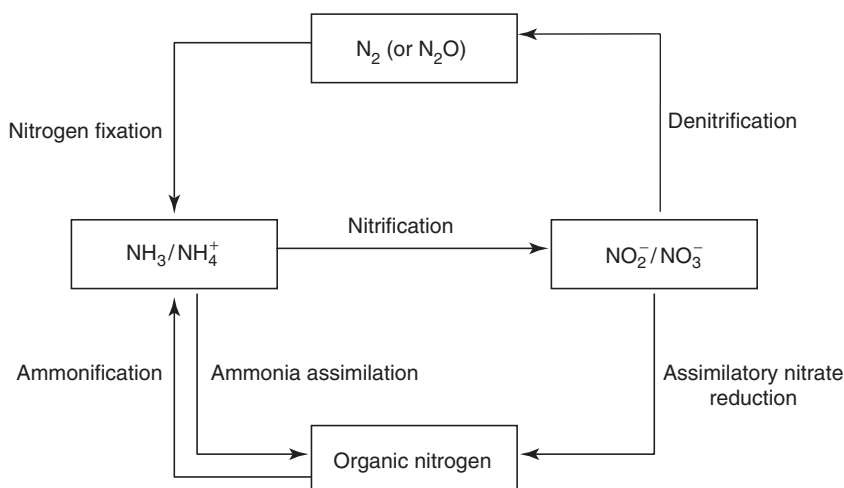
<sup>a</sup>All values of  $\log K$  were calculated with  $\Delta G_f^\circ$  data from Shock EL, Helgeson HC, and Sverjensky DA (1989) Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Standard partial molal properties of inorganic neutral species. *Geochimica et Cosmochimica Acta* 53: 2157–2183. Shock EL, Sassani DC, Willis M, and Sverjensky DA (1997) Inorganic species in geologic fluids: Correlations among standard molal thermodynamic properties of aqueous ions and hydroxide complexes. *Geochimica et Cosmochimica Acta* 61: 907–950. Wagman DD, Evans WH, Parker VB *et al.* (1982) The NBS tables of chemical thermodynamic properties: Selected values for inorganic and C1 and C2 organic substances in SI units. *Journal of Physical and Chemical Reference Data* 11 (supplement 2).

## Biogeochemical Cycling Processes

A schematic of the biological nitrogen transformations of nitrogen species is shown in **Figure 2** with the main processes listed below. All of these processes are mediated by various types of microorganisms with some processes being energy producing and others occurring symbiotically with other organisms.

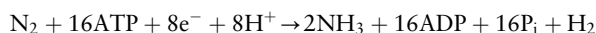
### Nitrogen Fixation

Nitrogen fixation is the process by which nitrogen is taken from its molecular form ( $\text{N}_2$ ) in the atmosphere and converted into nitrogen compounds useful for other biochemical processes. Fixation can occur through atmospheric (lightning), industrial, or biological processes. Biological nitrogen fixation can be represented by the following reaction, in which the enzyme-catalyzed reduction of  $\text{N}_2$  to  $\text{NH}_3$ ,  $\text{NH}_4^+$ , or



**Figure 2** Schematic representation of the biological transformations of nitrogen species.

organic nitrogen occurs:



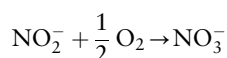
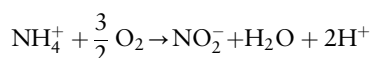
This process is performed by a variety of prokaryotes, both symbiotic and free living, using an enzyme complex termed nitrogenase that is composed of two separate protein components (dinitrogenase reductase and dinitrogenase). Dinitrogenase reductase donates two high potential electrons at a time to dinitrogenase and contains an Fe–S center that holds the electrons before donation. Dinitrogenase then catalyzes the reduction of  $\text{N}_2$ . Once nitrogen has been fixed, it can be oxidized to  $\text{NO}_2^-/\text{NO}_3^-$  or assimilated by organisms.

### Ammonia Assimilation

Ammonia assimilation is the process in which the fixed  $\text{NH}_3$  or  $\text{NH}_4^+$  is transformed to the organic form through assimilation by organisms to become part of their biomass. This process is of great importance for those organisms that can directly utilize nitrogen as  $\text{NH}_3$  or  $\text{NH}_4^+$ .

### Nitrification

This process describes the oxidation of  $\text{NH}_3$  or  $\text{NH}_4^+$  to  $\text{NO}_2^-$  or  $\text{NO}_3^-$  by organisms. Nitrification occurs in two energy-producing steps: (1) oxidation of ammonia to nitrite and (2) the subsequent oxidation of nitrite to nitrate. The following general reactions represent the two nitrification steps:



This process is used primarily by a group of aerobic chemotrophs called the nitrifiers. While some low levels of heterotrophic nitrification do occur, rates are low and quantities of nitrate produced are relatively small compared to those by the chemotrophs. The oxidation of ammonia into nitrite is done by *Nitrosomonas* species with the second step performed by *Nitrobacter* species.

### Assimilatory Nitrate Reduction

In the global environment, nitrogen in the form nitrate or nitrite is assimilated (immobilized) into biomass in the form of  $\text{NH}_3$ . This reduction is catalyzed by two assimilatory enzymes (nitrite reductase and nitrate reductase) and can be carried out by plants, fungi, and prokaryotes. This process is likely to dominate when reduced nitrogen is in low supply (e.g., during aerobic conditions).

**Table 4** Selected methods for the determination of nitrogen species in waters

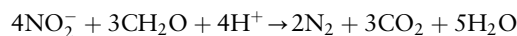
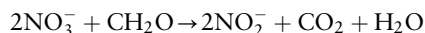
Species	Method and description <sup>a</sup>
Nitrate and nitrite	<p>Direct determination: In electrochemical procedures, ion-selective electrodes are used.</p> <p>In spectroscopic methods, ultraviolet absorbance can quantify nitrate and nitrite; nitrite may first require separation</p> <p>Indirect determination: Approach that is based on reduction of nitrate to nitrite (e.g., by cadmium) with nitrite measured directly or via derivatization</p> <p>Chromatography: Ions are separated using ion chromatography, ion-exchange, ion-exclusion, and ion-interaction techniques</p> <p>Flow injection analysis (FIA): This versatile method can be coupled with various modes of detection (e.g., spectrophotometric, luminescence, electrochemical)</p>
Ammonia	<p>Titrimetric determination: Volumetric titration (e.g., acid–base) is used to determine the amount of ammonia</p> <p>Indophenol blue colorimetry: Ammonia is reacted with phenolic agents to form blue indophenol dyes (absorbs at a wavelength range of 630–720 nm)</p> <p>Ammonia-selective electrode: These electrodes are electrochemical cells in which a potentiometric signal is correlated with the concentration of ammonia</p> <p>Chromatographic techniques: Methods that involve the separation of components in a mixture; e.g., gas chromatography, liquid chromatography, and ion chromatography</p> <p>FIA</p>
Organic nitrogen	<p>Kjeldahl method: Organic nitrogen is converted to ammonia using metals such as mercury and selenium as catalysts</p> <p>Photochemical oxidation: This method involves the use of ultraviolet radiation to decompose organic matter, during which organic nitrogen is oxidized to nitrate</p> <p>High-temperature combustion: During pyrolysis at 900–1100°C, organic nitrogen is transformed into NO, which can be quantified by chemiluminescence, or be further reduced to <math>\text{N}_2</math> and detected by thermal conductivity</p> <p>Koroleff method: Under strongly alkaline medium, total nitrogen is oxidized to nitrate using potassium peroxodisulfate</p> <p>FIA</p>

<sup>a</sup> Adapted from Robards K, McKelvie ID, Benson RL, *et al.* (1994) Determination of carbon, phosphorus, nitrogen, and silicon species in waters. *Analytica Chimica Acta* 287: 147–190. Cerdà A, Oms MT, and Cerdà V (2000) Determination of organic nitrogen. In: Nollé LML (ed.) *Handbook of Water Analysis*, pp. 261–271. New York: Dekker. Gibb SW (2000) Ammonia. In: Nollé LML (ed.) *Handbook of Water Analysis*, pp. 223–259. New York: Dekker.



## Denitrification

Denitrification (often referred to as dissimilatory nitrate reduction) is the microbial reduction of  $\text{NO}_3^-$  to  $\text{N}_2$ . The reaction steps require an organic carbon as electron donor ( $\text{CH}_2\text{O}$  generically used) and are shown below:



Many denitrifying bacteria consume organic matter while utilizing  $\text{NO}_3^-$  as their electron acceptor. Thus, denitrification can readily occur in the presence of organic matter, in soils devoid of oxygen, and with the help of denitrifying bacteria. These denitrifiers are represented by many diverse species: both the Gram-positive and Gram-negative bacteria, as well as archaea; some are thermophilic and others are halophilic.

## Ammonification

This process involves the breakdown of organic nitrogen compounds (from soil or aquatic organic compounds) into  $\text{NH}_3$  or  $\text{NH}_4^+$ . Heterotrophic bacteria are principally responsible for the process of ammonification, utilizing organic nitrogen compounds and leaving behind  $\text{NH}_3$  and  $\text{NH}_4^+$ . These nitrogen species can then be recycled back into the biosphere.

## Nitrogen Determination

A variety of methods has been described for the determination of nitrogen species (Table 4) but not all are routinely used. The cadmium reduction method is widely used in both batch and automated (continuous flow) spectrophotometric methods. In this procedure, nitrate is reduced to nitrite, which is then determined by diazotization with sulfanilamide and coupling with *N*-(1-naphthyl)ethylenediamine dihydrochloride (NED) to form an intensely pink-colored azo dye. This chemistry can be incorporated in a flow injection manifold to allow rapid, automated, *in situ* determinations in a robust and portable manner. Other common techniques for nitrogen determination are the nitrate ion-selective electrode and ion chromatography.

## Global Outlook

As discussed, the global nitrogen cycle comprises many important chemical, biochemical, geochemical, and biogeochemical processes. Nitrogen is also tightly coupled with other elements (carbon, phosphorus, sulfur, and trace metals) and understanding these relationships will help in determining the role of living matter in biogeochemical cycles. The effect of human activity on the global nitrogen cycle is also of interest. The realization that nitrogen is commonly a limiting nutrient in plant growth has led to the invention and large-scale usage of nitrogen fertilizers, which in turn accounts for more than half of the human perturbation to the global nitrogen cycle. In addition, the widespread use of fossil fuels has led to the increased production of nitrogen oxides, which ultimately contribute to photochemical smog and acid precipitation. However, it is difficult to assess the true impact of these inputs given the extent and magnitude of natural fluxes. Only through continued study of the important processes that occur in the nitrogen cycle, will we truly understand the relative impacts.

*See also:* **Elemental Speciation:** Waters, Sediments, and Soils. **Fertilizers.** **Flow Injection Analysis:** Principles. **Ion-Selective Electrodes:** Overview. **Water Analysis:** Potable Water. **Water Determination.**

## Further Reading

- Bashkin VN (2002) *Modern Biogeochemistry*. Dordrecht: Kluwer Academic Publishers.
- Brookins DG (1988) *Eh-pH Diagrams for Geochemistry*. New York: Springer.
- Burns RC and Hardy RWF (1975) *Nitrogen Fixation in Bacteria and Higher Plants*. New York: Springer.
- Hanrahan GH, Gardolinski P, Gledhill M, and Worsfold PJ (2002) Environmental monitoring of nutrients. In: Burden FR, McKelvie I, Förstner U, and Guenther A (eds.) *Environmental Monitoring Handbook*, pp. 8.1–8.16. New York: McGraw-Hill.
- Jaffe DA (2000) The nitrogen cycle. In: Jacobson MC, Charlson RJ, Rodhe H, and Orians GH (eds.) *Earth System Science: From Biogeochemical Cycles to Global Change, International Geophysics Series*, vol. 72, pp. 322–341. London: Academic Press.
- Ottone G (1997) *Principles of Geochemistry*. New York: Columbia University Press.
- Rosswall T (1983) The nitrogen cycle. In: Bolin B and Cook RB (eds.) *The Major Biogeochemical Cycles and Their Interactions*, SCOPE 21, pp. 46–50. Chichester: Wiley.

# NITROSAMINES

**A R Fernández-Alba and A Agüera**, University of Almería, Almería, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

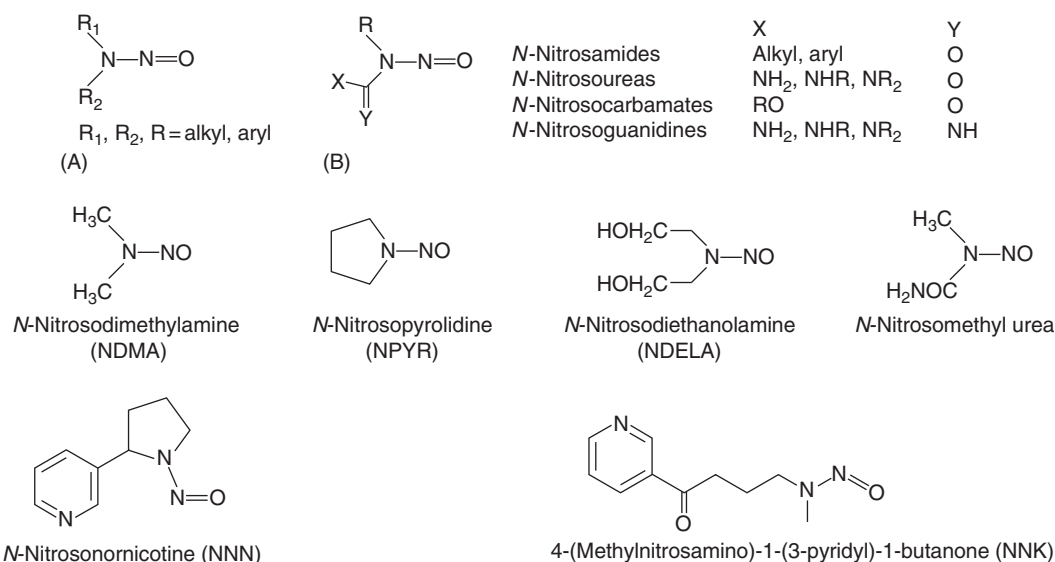
This article is a revision of the previous-edition article by W G de Ruig, pp. 3347–3352, © 1995, Elsevier Ltd.

## Introduction

Nitrosamines is a general term used to designate a vast group of *N*-nitroso compounds (NOCs), bearing a common functional  $>\text{N}-\text{N}=\text{O}$  group. The NOCs can be divided into two classes: *N*-nitrosamines and *N*-nitrosamides and related compounds. *N*-Nitrosamines are *N*-nitroso derivatives of secondary amines (Figure 1A); *N*-nitrosamides and its related compounds are those of substituted ureas, amides, carbamates, guanidines, and similar compounds (Figure 1B).

The interest for this group of compounds increased since the evidence of their mutagenic and carcinogenic properties. Most NOCs induce cancer in laboratory animals and may be involved in the etiology of several human cancers. The wide range of human exposure sources (exogenous and endogenous) have led to the necessity of regulatory and legislative action and to the search of advanced and sensitive methods of analysis (see Table 1).

Among the vast group of compounds included under the term of NOC some are of special interest. That is the case of the simplest *N*-nitroso compound, *N*-nitrosodimethylamine (NDMA), or the so-called tobacco-specific nitrosamines (TSNAs), which are among the most abundant carcinogens identified in tobacco and its smoke. Seven TSNAs have been identified both in tobacco filler and tobacco smoke. They are nonvolatile and formed by nitrosation of nicotine and related tobacco alkaloid precursors with



**Figure 1** Nitrosamines and nitrosamides. Some examples of interest.

**Table 1** Human exposure to *N*-nitroso compounds

Life style <sup>a</sup>	Occupational <sup>a</sup>	Intake of precursors <sup>b</sup>	Formation of precursors <sup>b</sup>
Food	Cutting oils	Nitrite	Nitrite from nitrate (saliva, gastric juice)
Tobacco	Hydraulic fluids	Nitrosable <i>N</i> -compounds	Nitrite formation in gut
Drugs	Air at working place		
Cosmetics			
Others			

<sup>a</sup> Exogenous exposure.

<sup>b</sup> Endogenous exposure.

nitrogen oxides. Some examples of the most representative NOCs are represented in Figure 1.

## Mutagenicity and Carcinogenicity of Nitrosamines

Experimental animal studies have established nitrosamines as a family of extremely potent carcinogens. Their cancer potencies are much higher than those of the trihalomethanes and it is for this reason that the International Agency for Research on Cancer (IARC) of the World Health Organization has focused great interest in this group of compounds.

Nitrosamines are not mutagenic *per se*, but if they are activated by liver enzymes such as cytochromes P50 (CYP2E1) and 2A6 (CYP2A6), they are able to react with DNA and the methylated DNA damage is then misrepaired resulting in modified transition of guanine–cytosine, adenine–thymine base pairs and finally in carcinogenicity. Many dialkyl nitrosamines are potent mutagens and carcinogens in rodents as well as powerful alkylating agents in humans.

Approximately 300 different *N*-nitrosamines can induce cancer in experimental animals. Some representative compounds of this class induce cancer in at least 40 different animal species, including higher primates. Experimental studies from animals can induce tumors in animals that resemble their human counterparts with respect to both morphological and biochemical properties. Nitrosamines are classified as suspect human carcinogens even though direct causal evidence is lacking. Literary data on the evaluation of carcinogenic hazard of nitroso compounds and in particular NDMA confirm their hazardous impact on humans due to chronic immunosuppression. Acute (short-term) exposure to NDMA may damage the liver in humans, with symptoms that include nausea, vomiting, headaches, and malaise. Chronic (long-term) exposure of humans to NDMA may cause liver damage and low platelet counts. Severe liver damage has been observed in animals. Animal studies have suggested that chronic ingestion and inhalation of NDMA may cause an increase in liver tumors and other types of tumors. EPA has classified NDMA as a Group B2, probable human carcinogen. Published data establish the total safety dose of NDMA at 0.1 mg per animal to determine NDMA safety levels in environment, i.e., atmospheric air and water reservoirs, at  $0.0001 \text{ mg m}^{-3}$  and  $0.0001 \text{ mg l}^{-1}$ , respectively, and to establish the total human safety level of NDMA at less than  $0.002 \text{ mg day}^{-1}$ .

Increasingly, extensive experimental and some epidemiological data suggest that humans are susceptible to carcinogenesis by *N*-nitroso compounds

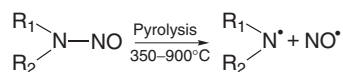
and that the presence of these compounds in some foods or the exposition to certain occupational settings may be regarded as an etiological risk factor, including cancers of the esophagus, stomach, nasopharynx, urinary bladder, childhood leukemia, brain, and colon. Esophageal cancer represents one of the most common and lethal cancers around the world and extensive data have showed that nitrosamines are the only carcinogens capable of inducing tumors in the esophagus of experimental animals, especially with the rat, which are the most susceptible species.

Of the seven TSNAs identified as components of tobacco and tobacco smoke, *N*-nitrosonornicotine (NNN), 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL) are considered the most carcinogenic. NNK and its primary metabolite, NNAL, are expected to be particularly important in the induction of adenocarcinoma, which is now the leading lung cancer type in the United States.

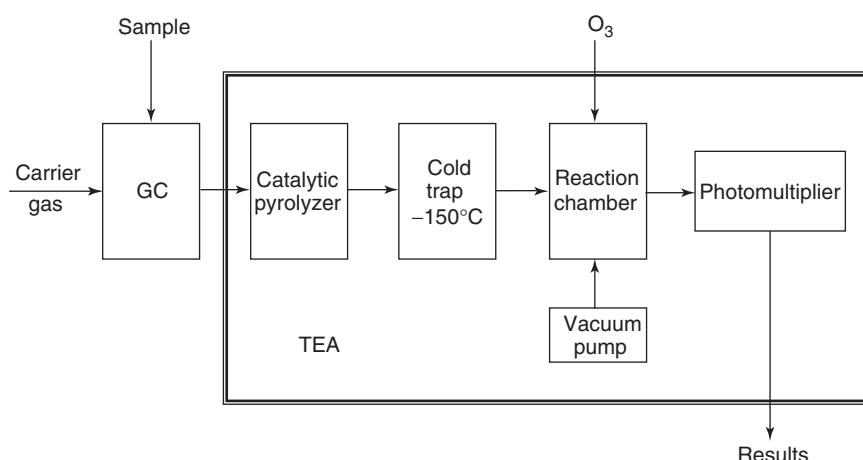
## Quantitative Analysis

Since the carcinogenic properties of nitrosamines were evidenced, the identification of these compounds in a wide range of foodstuffs represented a primary objective for many researchers. The analytical techniques initially applied to this purpose, such as thin-layer chromatography, gas chromatography (GC) with flame ionization or Hall detection, spectrophotometry, or polarography, lacked the necessary specificity and/or sensitivity to give reliable results and reports of false positive results were not uncommon. The development of more selective methodologies was required.

The only detection method that is recognized as specific for nitrosamines is the thermal energy analyser (TEA). This technique, combined with GC separation, has become the most commonly used detection method for volatile NOCs because of its specificity, speed, and accuracy. A functioning scheme is represented in Figure 2. The GC eluent is introduced into a catalytic pyrolyzer where the catalytic cleavage of the *N*-nitroso bond is produced, releasing the nitrosyl radical ( $\text{NO}^\bullet$ ):

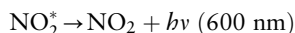
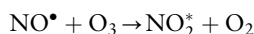


The reaction products are swept through a cold trap at  $-150^\circ\text{C}$  to remove potentially interfering products. In a reaction chamber, the nitrosyl radical is oxidized by ozone to give electronically excited nitrogen dioxide that rapidly decays back to the



**Figure 2** Schematic representation of the thermal energy analyzer (TEA).

ground state emitting red light at 600 nm:



Measurement of this radiation by a sensitive photomultiplier tube allows an accurate and specific determination of the NOCs present.

There are, however, a large number of polar nitroso compounds not generally amenable to direct analysis by GC, either because of their low volatility or their thermal instability. The determination of these nonvolatile NOCs was initially hampered by the incompatibility of the TEA with reversed-phase liquid chromatography (LC), because the polar mobile phase affects the stability of the detection signal. Interest in the analysis of nonvolatile *N*-nitrosamines has, however, been renewed due to the recent development of several new reversed-phase liquid chromatography (LC) interfaces to TEA or chemiluminescence detection. An efficient photolytic interface between LC and TEA detector has been developed. The chromatographic eluent is introduced into a glass coil irradiated with ultraviolet (UV) light with a purge stream of helium. Nitrogen oxide produced by photolysis is carried by helium through three cold traps to the reaction chamber of the TEA. This interface allows an efficient link between LC and TEA, providing good selectivity and sensitivity for nonvolatile *N*-nitrosamines, but it is not commercially available. An improved procedure using LC in combination with postcolumn photolysis has been developed. In this method, the LC effluent is transferred to a Teflon tube in a reaction chamber that is irradiated by a high-intensity discharge lamp emitting at 254 nm. The NOCs are cleaved by the UV radiation and the nitrosyl radicals formed are oxidized to

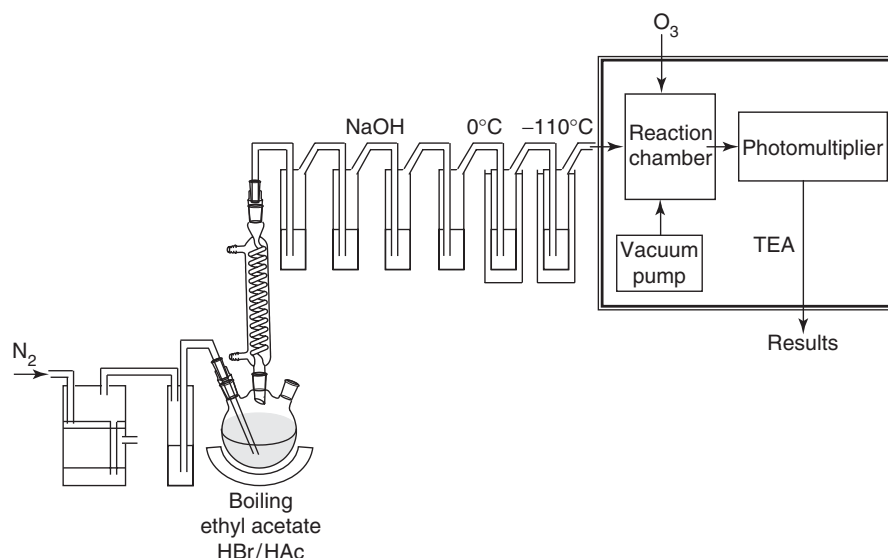
$\text{NO}_2$  radicals, which further react to give  $\text{N}_2\text{O}_3$  and  $\text{N}_2\text{O}_4$ . Both  $\text{N}_2\text{O}_3$  and  $\text{N}_2\text{O}_4$  are unstable in aqueous surroundings and will convert to nitrite. The effluent from the photohydrolysis coil is mixed with Griess reagent. The azo compound formed by reaction of nitrite with Griess reagent is measured spectrophotometrically at 546 nm.

A new application of a counter-flow gas diffusion cell (CFGDC)-based particle beam LC interface has been described for the LC-TEA analysis of nonvolatile *N*-nitrosodiethanolamine and *N*-nitroso-methyl-*p*-amino-2-ethylhexylbenzoate. The interface incorporates a thermospray vaporizer, desolvation chamber, and CFGDC to reduce the LC effluent to a dry aerosol and a single-stage momentum separator to form a particle beam of the nonvolatile analyte.

LC combined with postcolumn chemiluminescence detection has also been applied to the determination of trace *N*-nitrosamines in groundwater. The method combines solid-phase extraction (SPE) on to a mini-activated carbon column, followed by elution with acetone and concentration of the extracts by denitrosation and fluorogenic derivatization. Peroxyoxalate chemiluminescence detection was carried out by using bis(2-nitrophenyl) oxalate and hydrogen peroxide as chemiluminescent reagents. The method provides *N*-nitrosamines detection at picomole per liter levels.

Another strategy has been developed for the analysis of total NOCs, including the injection of the sample into boiling ethyl acetate containing HBr/HAc and further transference of the nitrosyl radicals formed into the reaction chamber of the TEA detector by a nitrogen stream; a scheme is represented in **Figure 3**. Only the total amount of nitrosamines can be determined since the method cannot distinguish between various nitrosamines.

The TEA has been the predominantly applied detection system for the analysis of NOCs, even though



**Figure 3** Schematic representation of quantitative determination of total NOCs.

this system is expensive and not available in most laboratories. For this reason there is the frequent use of GC–mass spectrometry (GC–MS) and LC–electrospray ionization (ESI)–MS. Thus, GC with tandem mass spectrometry in the chemical ionization mode (GC–CI–MS/MS) or GC with high-resolution mass spectrometry (GC–HRMS) have become the most common techniques currently used for analysis of low concentrations of NDMA. Recently, LC–MS/MS has been applied for the analysis of tobacco-specific *N*-nitrosamines in snuff or mainstream cigarette smoke, reporting detection limits ranging from 0.005 to 0.01 mg g<sup>-1</sup> and from 0.05 to 1.23 ng ml<sup>-1</sup> respectively. One of the main problems when using LC–MS/MS detection is signal suppression, or sometimes enhancement, caused by coeluting matrix components. To avoid these matrix effects, a thorough cleanup of the samples is often required. Alumina or silica gel column chromatography or SPE procedures with different sorbents (Oasis HLB) have been assayed.

*N*-Nitrosamines are directly determined as the free forms by GC–TEA. However, other GC- and LC-based methods have been referenced based on the application of derivatization reactions. Precolumn or postcolumn derivatization of the *N*-nitroso compounds very often provides unsatisfactory sensitivities. Better results can be obtained by previous denitrosation of the NOCs and subsequent detection of liberated secondary amines via derivatization. Denitrosation can be obtained by reaction with hydrobromic acid and the subsequent derivatization of the secondary amines formed by reaction with typical derivatization agents such as dansyl chloride or diethyl chlorothiophosphate.

The recovery of nitrosamines at ultratrace level requires the development of adequate preconcentration methods. Depending on the matrix to be analyzed different procedures have been used. Liquid–liquid extraction has widely been applied using different solvents such as dichloromethane, ethyl acetate, and buffer solution.

Supercritical fluid extraction (SFE) using supercritical carbon dioxide (SC–CO<sub>2</sub>) has been successfully used for isolation of volatile nitrosamines from different matrices such as tobacco and food products. This technique presents several advantages with respect to other extraction methods (e.g., mineral oil distillation or low-temperature vacuum distillation) currently used. Thus, SFE minimizes sample handling, provides fairly clean extracts, expedites sample preparation, and reduces the use of environmentally toxic solvents. Good results have also been obtained with the use of SPE in the analysis of food matrices combining extraction with Extrelut sorbent and purification with Florisil. This method is applicable for the analysis of a range of the most widely encountered volatile *N*-nitrosamines, including the poorly volatile NDBA, NDBzA, and *N*-nitroso-*N*-methylaniline in various food products. Active carbon is suitable for this preconcentration step due its low cost, versatility, and easy application.

A serious problem in NOC analysis is the risk of the formation of artifacts during storage, preparation, and analysis, due to the presence of relatively large amounts of precursors, and by the low concentration of NOCs, if present. Therefore, adequate measures have to be taken to prevent and detect the formation of artifacts. SPE methods are free of artifactual formation of *N*-nitrosamines, common when



a vacuum distillation procedure is used. In this procedure, during the distillation step, where nitrosamines precursors are present, there must be strict control.

## Human Exposure

*N*-Nitrosamines are widely distributed in various human environments. The concern was initially focused on their widespread occurrence in food and consumer products, as beer, meats cured with nitrite, smoked fish, tobacco and tobacco smoke, rubber products including baby bottle nipples and pacifiers, cosmetics, drug formulations, or herbicides formulations. Much data of their occurrence have been obtained by inadequate analytical methods and must await confirmation. Considerable progress has been made in the development of adequate and specific methods for trace analysis of nitrosamines, and reliable information is expected in the near future.

Recent results, which have been confirmed by MS, indicate that DMN does occur in certain food products at the  $5\text{--}10\text{ }\mu\text{g kg}^{-1}$  level. Investigations undertaken on the occurrence of nitrosamines (DMNA, DENA) in meat and meat products have assessed concentration levels of  $\sim 10\text{ }\mu\text{g kg}^{-1}$ , although large amounts can be found as a result of various transformation processes (in fried bacon or smoked meat). A precursor of DMNA is the amino acid glycine. It may also originate from quaternary amino salts such as choline, acetylcholine, and betaine, which are common in meat.

Evidence of human exposure from industrial sources has been shown in air of factories that use unsymmetrical dimethylhydrazine (UDMH)-based rocket fuel. *In situ* monitoring studies have been performed in working environments. The applicability of integrated chemical/biological approaches for monitoring genotoxic hazards in different rubber factories has been proved. In particular, chemical analyses combined with a battery of genotoxicity tests, which reveal different genetic endpoints (DNA damage, point mutation, and micronuclei) have provided more information on potential health hazards in working environments in the rubber industry.

Apart from particular locations where exposure may be high, the most widespread and potent form of exposure to NOCs is via tobacco. Tobacco contains a number of alkaloids, including nicotine and in much less extent nornicotine, anatabine, and anabasine. During the curing process, the tobacco alkaloids undergo nitrosation to produce TSNAs, which are known to be amongst the most significant cancer-causing constituents of tobacco.

Recently, the presence of high levels of the potent carcinogen NDMA has been observed in drinking water supplies, especially in locations where wastewater effluents are used for indirect potable reuse. It is proved that chlorination of water and wastewater leads to the formation of relatively high concentrations of NDMA. Human exposure to *N*-nitroso compounds is not limited to their intake or exogenous exposure. Endogenous formation of nitrosamines can occur in the acidic stomach through nitrosation reactions of amines when nitrosating agents such as nitrite and nitrate are present. Nitrite is readily protonated to nitrous acid ( $\text{HNO}_2$ ) and two molecules of nitrous acid form dinitrogen trioxide ( $\text{N}_2\text{O}_3$ ), which is the actual nitrosating species that reacts with unprotonated amines to form nitrosamines. Nitrate can also contribute to nitrosation because it can be reduced to nitrite by bacteria in the mouth. Human exposure to endogenously formed NOCs has been related to an increase risk of gastric, esophageal, nasopharyngeal, and bladder cancer.

Quantitation of human exposure to endogenous NOCs can be achieved by the application of the *N*-nitrosoproline (NPRO) test. This test is based on the excretion of NPRO and other *N*-nitrosoamino acids in the urine that are measured as an index of endogenous nitrosation, following ingestion of precursors. The NPRO test has been applied to human subjects in clinical and epidemiological studies, and the kinetics and dietary modifiers of endogenous nitrosation have been investigated. Many constituents present in the human diet have been shown to inhibit endogenous formation of NOCs; for example, in studies performed with human volunteer, ascorbic acid and green tea have been shown to reduce endogenous NDMA formation. Ascorbic acid reacts faster than the amine with the nitrosating agent reducing nitrous acid to nitric oxide (NO), which is not directly a nitrosating agent. Tea polyphenols also inhibit nitrosation by competing with secondary amines for nitrosating species; the phenolic groups are then oxidized to quinone moieties.

## Sources and Occurrence

The main route for the formation of NOCs is by nitrosation reactions that take place between a nitrosating agent like nitrite ( $\text{NO}_2^-$ ) in an acidic environment, nitrous acid ( $\text{HNO}_2$ ), nitrogen oxides ( $\text{NO}_2$  or  $\text{NO}_x$ ), nitrosyl chloride (NOCl), and, on the other hand, a nitrosable agent like some amines, amides, guanidines, urea, and urethane compounds.

The nitrosation mechanism is believed to be responsible for the formation of NOCs in some foods,

such as vegetables, fish, and especially meat products cured with nitrite to prevent the growth of *Clostridium botulinum*, the bacterium that generates botulism toxin. Reactions similar to nitrosation also can occur in the atmosphere. Combustion often results in the formation of nitrogen-containing species (i.e.,  $\text{NO}_x$  or nitroso radicals) that reacts with species such as dimethylamine, released by industrial sources, to produce nitrosamines. Gas-phase nitrosation may explain the occurrence of NDMA in cigarette smoke, malt beverages, dried foods, and rubber products.

Occupational exposures to some NOCs are high in the tyre and rubber industries where nitroso radicals in engine exhaust react with amine-containing accelerators used for vulcanization. Occurrence of volatile nitrosamines, mainly NDMA and *N*-nitrosomorpholine, in work room air in the rubber and tyre industries have been detected at trace levels, in the range of  $1\text{--}10\text{ }\mu\text{g m}^{-3}$ , and sometimes higher than the maximum exposure level for some European countries ( $1\text{ }\mu\text{g m}^{-3}$ ) and over the German target value for rubber vulcanization ( $2.5\text{ }\mu\text{g m}^{-3}$ ).

NOCs are also observed in wastewater effluents in a variety of industrial plants manufacturing amines, herbicides, pesticides, pharmaceuticals, and rubber. Concentrations up to  $2\text{ mg l}^{-1}$  have been detected in wastewater effluent of a tyre factory in Ontario.

Fertilization of arable lands and grassland soils with mineral nitrogen favors the formation of nitrosamines in soil. Formation of nitro and nitroso and other compounds has been observed, which are precursors of nitrosamines. They are also products of biotransformation of some pesticides and other precursors. Dynamics of the formation and development of nitrosamine precursors in soil depend on many physicochemical factors but soil acidification processes undoubtedly favor it. Once they are formed they remain in soils for between 90 and 128 days and in aquatic environment for 80–110 days.

Of recent interest has been the detection of NDMA in groundwater in Sacramento Country, CA, at concentrations as high as  $0.4\text{ mg l}^{-1}$ , which forced the closure of drinking water wells. Initially, the presence of NDMA at so high concentrations was associated to the proximity to facilities that used UDMH-based fuels. Oxidation of UDMH by different oxidants yields NDMA as by-product. Further survey studies have demonstrated, however, that the presence of NDMA in drinking water can be associated with chlorine disinfection of water and wastewater. The formation of NDMA during chlorination can be

explained by the slow reaction of monochloramine with dimethylamine to form UDMH as intermediate, which is rapidly oxidized by monochloramine or other oxidants to form NDMA. Other nitrogen organic compounds also present in the wastewaters such as tertiary amines with dimethylamine functional groups also serve as NDMA precursors.

**See also:** **Extraction:** Supercritical Fluid Extraction. **Food and Nutritional Analysis:** Overview. **Gas Chromatography:** Overview. **Liquid Chromatography:** Overview. **Mass Spectrometry:** Overview.

## Further Reading

- ECETOC Technical Report No. 42 (1991) Critical evaluation of methods for the determination of *N*-nitrosamines in personal care and household products. European Chemical Industry Ecology and Toxicology Centre, Brussels, Belgium.
- Egan H, Preussmann R, Eisenbrand G *et al.* (1983) *Environmental Carcinogens – Selected Methods of Analysis, vol. VI – N-Nitroso Compounds*. IARC Scientific Publications No. 45. Lyon: IARC.
- Fine DH, Ruffe F, Lieb D, and Rounbehler DP (1975) Description of the thermal energy analyser (TEA) for trace determination of volatile and non-volatile *N*-nitroso compounds. *Analytical Chemistry* 47: 1188–1190.
- Haorah J, Zhou L, Wang X, Xu G, and Mirvish SS (2001) Determination of total *N*-nitroso compounds and their precursors in Frankfurters, fresh meat, dried salted fish, sauces, tobacco and tobacco smoke particulates. *Journal of Agriculture and Food Chemistry* 49: 6068–6078.
- IACR Working Group on the Evaluation of the Carcinogenic Risk of Chemicals to Humans (1978) *N*-Nitrosodimethylamine. In: *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemical to Humans, vol. 17: Some N-Nitroso Compounds*, pp. 125–175. Switzerland: World Health Organization.
- Janson C, Paccou A, and Österdahl B-G (2003) Analysis of tobacco-specific *N*-nitrosamines in snuff by ethyl acetate extraction and liquid chromatography-tandem mass spectrometry. *Journal of Chromatography A* 1008: 135–143.
- Mirvish SS (1975) Formation of *N*-nitroso compounds: chemistry, kinetics, and *in vivo* occurrence. *Toxicology and Applied Pharmacology* 31: 325–351.
- Mirvish SS (1995) Role of *N*-nitroso compounds (NOC) in etiology of gastric, esophageal, nasopharyngeal and bladder cancer, and contribution to cancer of known exposures to NOC. *Cancer Letters* 93: 17–48.
- Mitch WA, Sharp JO, Trussell RR, *et al.* (2003) *N*-nitrosodimethylamine (NDMA) as a drinking water contaminant: a review. *Environmental Engineering Science* 20: 389–404.

# NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

Contents

**Overview**

**Principles**

**Instrumentation**

## Overview

**A Bainbridge**, University College London Hospitals Trust, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Nuclear magnetic resonance spectroscopy (NMRS) has become a useful tool for the nondestructive analysis of small samples and increasingly for non-invasive *in vivo* tissue analysis. In this article, an overview of the theory, equipment, and methodologies of NMRS are presented.

## Theory

The interaction of certain atomic nuclei with a strong magnetic field gives rise to the NMR signal and consequently NMR spectra. This section contains a brief overview of the physics of signal generation leading onto a discussion of the physics behind the specific structure of NMR spectra.

### The Physics of NMR

Atomic nuclei that have odd mass numbers and/or odd atomic numbers possess a nonzero quantum spin number,  $I$ , made up of the vector sum of the spins of its constituent protons and neutrons. The component of the vector  $I$  along an arbitrarily defined axis ( $z$  in this case) has a magnitude  $I_z$  given by

$$I_z = m_I \hbar \quad [1]$$

where  $\hbar$  is Planck's constant divided by  $2\pi$ . The quantum number  $m_I$  can take  $(2I + 1)$  values (such that  $m_I = -I, -(I - 1), \dots, [I - 1], I$ ); one can only measure a component of  $I$  in any one direction at any one time. The nuclei most commonly probed in NMR experiments ( $^1\text{H}$ ,  $^{31}\text{P}$ ,  $^{13}\text{C}$ , and  $^{19}\text{F}$ ) have  $I = 1/2$  and so  $m = \pm 1/2$ . These two states are normally energy

degenerate. However, in the presence of a strong magnetic field,  $B_0$ , applied along the same arbitrarily defined  $z$ -axis as in eqn [1], the strong field Zeeman effect causes the states to split into two energy levels separated by energy  $\Delta E$ :

$$\Delta E = \gamma \hbar B_0 \quad [2]$$

The gyromagnetic ratio,  $\gamma$ , is defined as the ratio of magnetic moment to angular momentum of the nucleus and impacts on the sensitivity of NMRS experiments. Assuming that the coupling between the individual spins is weak, the relative populations of the two levels are given by a Boltzmann distribution:

$$\frac{n_{\uparrow}}{n_{\downarrow}} = \exp \left[ \frac{\Delta E}{kT} \right] \quad [3]$$

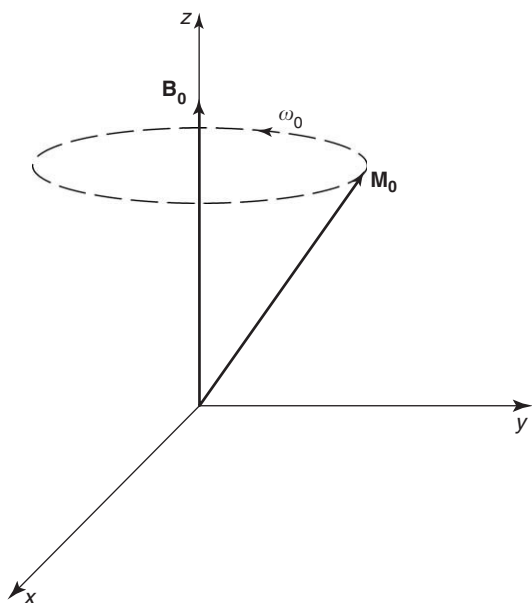
where  $k$  is the Boltzmann constant and  $T$  is the temperature of the spin system. A collection of spins will distribute themselves between the energy levels resulting in a small excess of spins in the lower ( $n_{\uparrow}$  – spin up) state at thermal equilibrium. This leads to a sample obtaining a net magnetization,  $M_0$ , lying parallel to the main field. This is the property that is detectable using NMR and its amplitude is dependent on  $\gamma^2$ ,  $B_0$ , and the number of equivalent nuclei per unit volume,  $n$ :

$$M_0 = \frac{n \gamma^2 \hbar^2 B_0}{4kT} \quad [4]$$

As  $M_0$  is a property of the bulk system, one can define all three components simultaneously and use classical mechanics to describe its evolution.  $M_0$  will precess around  $B_0$  at the Larmor frequency,  $\omega_0$  (see Figure 1). The Larmor frequency is equivalent to the frequency of photons required to induce transitions between the energy levels and is defined as:

$$\omega_0 = \gamma B_0 \quad [5]$$

Manipulation of the orientation of  $M_0$  relative to  $B_0$  using circularly polarized radio frequency (RF) at this resonant frequency leads to the generation of the NMR signal. To illustrate this it is commonplace to

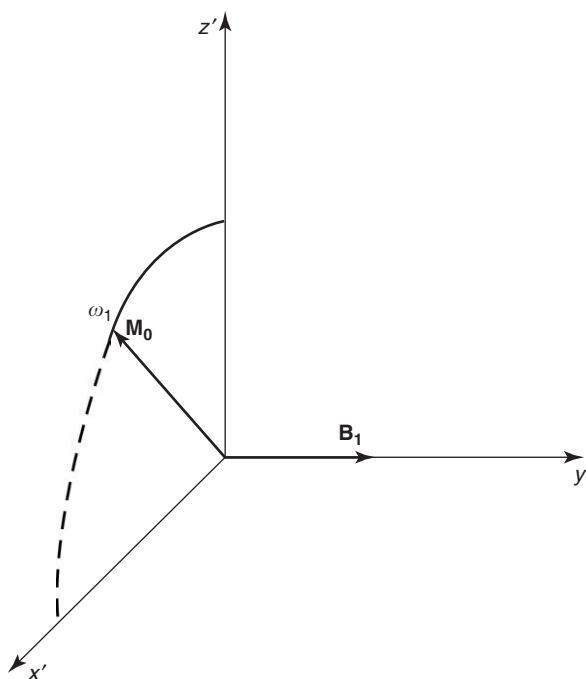


**Figure 1** The net magnetization of the sample  $M_0$  will precess around the applied magnetic field vector with an angular frequency  $\omega_0$ . This is equivalent to the frequency of RF required to stimulate a transition between the spin-up and spin-down energy states.

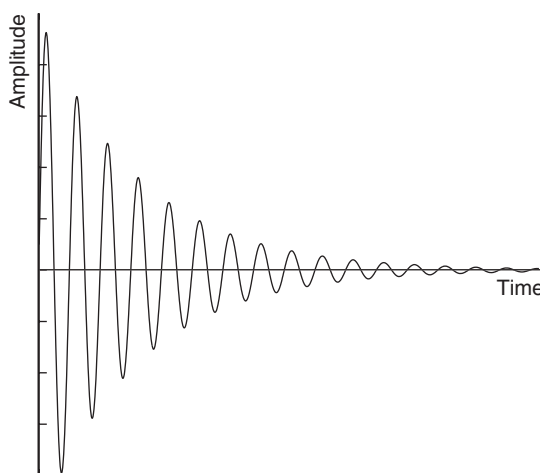
define a set of axes rotating about  $B_0$  at the Larmor frequency. With the introduction of RF at this frequency, right-hand circularly polarized along the  $z$ -axis, the equivalent magnetic field in this frame,  $B_1$ , becomes oriented in the  $x'$ - $y'$  plane and has an amplitude equal to the amplitude of the B-field component of the RF.  $M_0$  will precess around  $B_1$  with a frequency  $\omega_1$  equivalently to eqn [5] (see **Figure 2**). By adjusting the length and power of the RF pulse it is possible to flip  $M_0$  vector into the  $x'$ - $y'$  plane; this is equivalent to equalizing the populations of the quantum mechanical energy levels and is known as a  $90^\circ$  pulse. The amplitude of the NMR signal is proportional to the component of  $M_0$  in the  $x$ - $y$  plane and its precession about  $B_0$  causes an oscillating current to be induced in a suitably arranged RF coil (described below). This signal decays exponentially with a time constant  $T2^*$  due to field inhomogeneities causing spins to precess at different rates and dephase. The result is a free induction decay (FID) (see **Figure 3**).

### Chemical Shift

NMR spectra consist of a series of peaks identifiable to specific chemical structures on particular molecules rather than just a single peak resonating at  $\omega_0$ ; the position of a peak in the spectrum is termed its chemical shift. Peaks occur at different chemical shifts because the magnetic field at an atomic nucleus is not

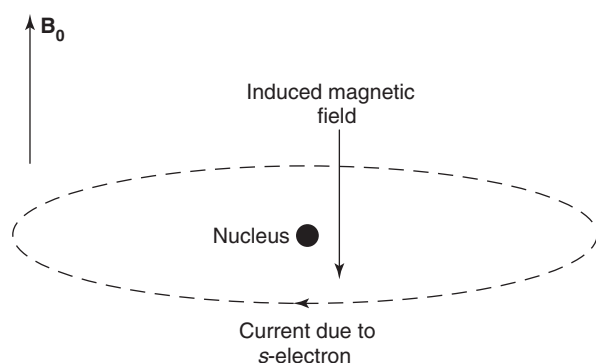


**Figure 2** The tilting of  $M_0$  into the  $x'$ - $y'$  frame due to the application of RF at the Larmor frequency  $\omega_0$ . This is represented in a frame of reference rotating about  $B_0$  at  $\omega_0$ . The equivalent field in the rotating frame is  $B_1$ , which appears stationary along  $y'$ .  $M_0$  precesses  $B_1$  with an angular frequency  $\omega_1$ .



**Figure 3** An example of FID consisting of a single frequency decaying exponentially.

equal to the applied magnetic field. The electrons in electronic orbitals shield the nucleus from the applied field, thus altering the resonant frequency. Electrons in atoms and molecules are described by a series of quantum numbers. Technically, the shape of an atomic orbital occupied by an electron with a given set of quantum numbers is best described by the square of its quantum mechanical wavefunction.

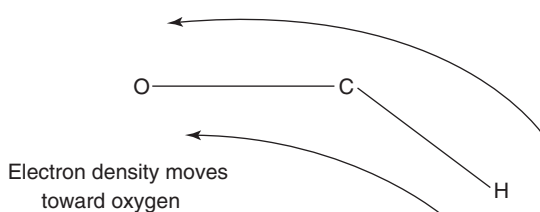


**Figure 4** Shielding of the nucleus from the applied magnetic field due to an orbital  $s$ -electron. The current due to the electron induces a magnetic field at the nucleus that acts to oppose the applied field.

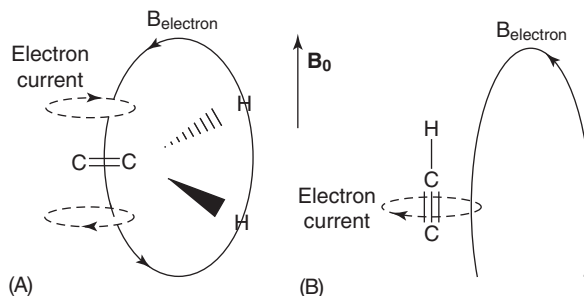
The principal quantum number,  $N$ , is dominant in determining the shape of an orbital; for a rigorous treatment of quantum mechanics and the physics of electronic orbitals, the reader can refer to the Further Reading. For the purpose of this discussion it is sufficient to recognize that  $s$ -orbitals ( $N=1$ ) have a spherical symmetry and higher orbitals ( $p$ ,  $N=2$ ;  $d$ ,  $N=3$ ; etc) have more complex shapes. Thus the simplest example of nuclear shielding is that arising from electrons in  $s$ -orbitals. A schematic of how an  $s$ -electron shields the nucleus is shown in **Figure 4**. In the presence of an applied field, the electron's orbital forms an effective current loop resulting in a magnetic field that opposes the applied field. Thus the nucleus experiences a reduced magnetic field and resonates at a lower frequency. The magnitude in hertz of the chemical shift is proportional to the size of the applied field, as this determines the magnitude of the effective currents produced by the electrons. In order to eliminate the dependence on  $B_0$ , chemical shifts are generally quoted in units of parts per million (ppm) of the static magnetic field strength.

### Proton Chemical Shifts

Only  $s$ -electrons have orbitals around hydrogen nuclei. However, the degree of shielding that they provide is further influenced by their chemical environment. Electronegative elements near to the proton withdraw electron density and reduce the extent of the shielding. An example of oxygen bonded to carbon influencing the chemical shifts of the protons also bonded to the carbon is shown in **Figure 5**. The larger the electronegativity of the directly bonded atom, the larger its influence on the chemical shift of its NMR visible neighbors. Double bonds ( $\pi$ -bonds) between carbon atoms in organic molecules also have a significant effect on proton chemical shift.



**Figure 5** Electronegative elements such as oxygen draw electron density toward them reducing the shielding of the proton.



**Figure 6** (A)  $\pi$ -Bonds between carbon atoms can lead to the magnetic field at a proton being enhanced when oriented in the plane perpendicular to the main field. (B) An exception is the terminal alkyne. In this example the shielding is increased.

For a portion of their time, these molecules will be oriented with their  $\pi$ -bonds perpendicular to applied field, as illustrated in **Figure 6**. The current distribution of the  $\pi$ -electrons is such that a magnetic field is produced that opposes the applied field above and below the bond, but reinforces the applied field at the attached proton. An exception to this is the terminal alkyne group (**Figure 6**), where the proton is oriented in the same plane as the  $\pi$ -bond. The major contribution comes when the  $\pi$ -bond is oriented parallel to the applied field. The current distribution from the  $\pi$ -electrons leads to an induced field that opposes the applied field at the proton. All of the above effects are additive in determining the final chemical shift of a proton and the position of resonances can be suggestive of specific function groups (see **Table 1**).

### Other Nuclei and Chemical Shift Mechanisms

Nuclei other than  $^1\text{H}$  that are typically used for NMRS, such as  $^{31}\text{P}$  and  $^{13}\text{C}$ , have electrons in higher orbitals that do not have spherical symmetry. These can produce comparatively large magnetic fields at the nucleus. Thus the range of chemical shifts is larger; typically, the range of chemical shifts for  $^{13}\text{C}$  (0–200 ppm) is 10–20 times the range of proton chemical shifts. The chemical shifts of some resonances are variable depending on the environment of



**Table 1** The chemical shift of a peak can indicate the type functional group producing the peak.

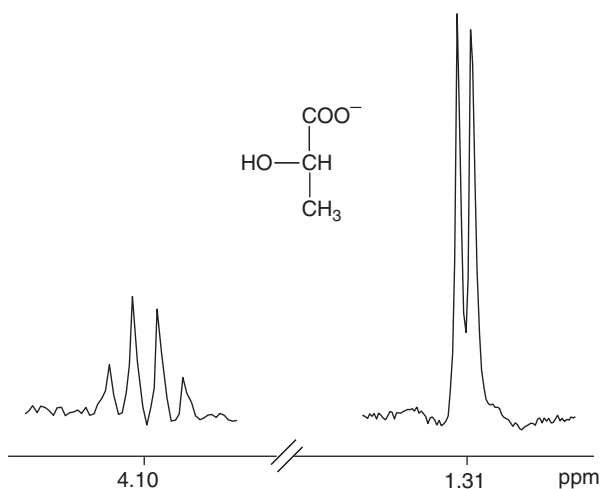
Chemical name	Functional group	Approximate chemical shifts relative to TMS (ppm)
Alkanes	$\text{R}-\text{CH}_3$	0.5–2
Alkenes	$\begin{array}{c} \diagup \\ -\text{C}=\text{C} \\ \diagdown \quad \diagup \\ \quad \quad \text{CH} \\ \quad \quad   \end{array}$	2–4
Alkynes	$-\text{C}\equiv\text{C}-\text{H}$	1.5–3
Alcohols	$\text{ROH}$	3–5
Aldehydes	$\begin{array}{c} \text{H} \\   \\ -\text{C} \\    \\ \text{O} \end{array}$	9.5–10
Carboxylic acids	$\begin{array}{c} \text{OH} \\   \\ -\text{C} \\    \\ \text{O} \end{array}$	10–12

Chemical shifts are expressed relative to that of tetramethylsilane (TMS) at 0 ppm

the sample. Hydrogen bound (acid) and unprotonated (base) chemicals have different chemical shifts. As the acid and base states are in rapid chemical exchange, a single resonance with the average chemical shift of the relative equilibrium populations is observed. The local concentration of hydrogen ions (and hence pH) determines the acid–base populations and so alters the chemical shift. Similarly, metal-ion concentration can influence chemical shift. Uncomplexed adenosine triphosphate (ATP) is observable using  $^{31}\text{P}$  spectroscopy. Free intracellular magnesium exists in equilibrium with complexed MgATP and so the observable ATP chemical shifts are sensitive to changes in free intracellular magnesium levels. Temperature is also important as increases in temperature lead to intensification of molecular motion. The chemical bonds weaken altering the degree of nuclear shielding. Thus the resonant frequency decreases. The  $^1\text{H}$  chemical shift of water has an almost linear dependence on temperature between  $0^\circ$  and  $40^\circ$  changing by approximately  $-0.01 \text{ ppm}^\circ\text{C}^{-1}$ .

### Spin–Spin Coupling

Spins of adjacent nuclei affect resonance energy levels, splitting the observed peaks into various multiplet structures. The coupling is not direct but is mediated by electrons in the chemical bonds between nuclei and allows a given nucleus to ‘sense’ the spin states of its neighbors. The coupling can be between



**Figure 7** The chemical structure and multiplets of lactate. The  $\text{CH}_3$  group produces the doublet at 1.31 ppm. The  $\text{CH}$  group produces the quartet at 4.1 ppm. Spectra were acquired on a 7 T Bruker Biospec spectrometer.

similar (homonuclear) or different nuclei (heteronuclear). However, chemically equivalent protons do not couple to each other. The chemical structure of lactate is shown in **Figure 7** with its corresponding  $^1\text{H}$  spectrum consisting of two resonances, a doublet and a quartet due to the  $-\text{CH}_3$  and  $-\text{CH}$  protons, respectively. The  $-\text{CH}$  proton has two possible spin orientations, parallel or antiparallel to  $\mathbf{B}_0$ , and is constantly changing between the spin states. Half of the time the  $-\text{CH}_3$  protons will sense the  $-\text{CH}$  proton parallel state and the remaining time the antiparallel state. The spin states of the  $-\text{CH}$  proton are at slightly different energy levels and due to electron mediation a slight change in the local magnetic field is experienced by the three equivalent  $-\text{CH}_3$  protons. Across the whole sample this results in a doublet with components of approximately equal amplitude. The  $-\text{CH}$  proton experiences eight possible combinations for the  $-\text{CH}_3$  protons, these are illustrated in **Figure 8**. However, some of these combinations are indistinguishable resulting in four effective combinations of spin states. The result is the quartet with the relative amplitudes of the components depending on the relative probabilities of the possible states,  $\sim 1:3:3:1$ . The total intensity of the doublet is three times that of the quartet as there are three times as many equivalent protons contributing to the resonance. The separation of multiplet components is known as the coupling constant  $J$ . This is measured in hertz and is independent of the applied field strength.

### Phase Modulation

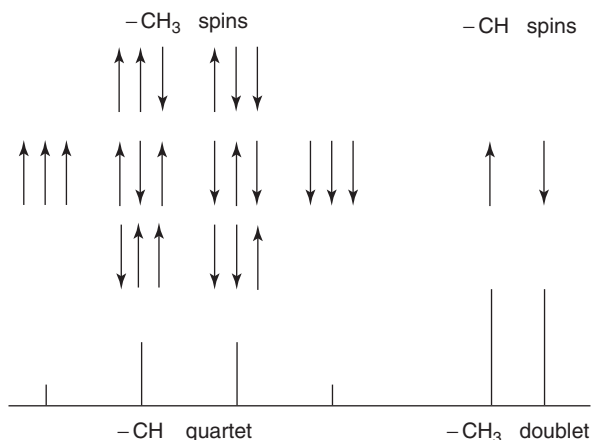
Where a spectrum is obtained using a sequence that involves acquiring an echo rather than direct

observation of the signal following excitation, spin-spin coupling affects the amplitudes and phases of spectral peaks. Consider the case of the lactate doublet observed using a spin-echo experiment with the resonant frequency,  $\omega_0$ , set half way between the two peaks for convenience (see Figure 9). In a frame of reference rotating at  $\omega_0$ , the spins in each peak of the doublet will disperse away from each other. Spins also dephase because of inhomogeneity in the main field. A  $180^\circ$  pulse applied along the  $y$ -axis has the

effect of flipping the spins about this axis. Subsequently, the dephasing due to applied field inhomogeneity is refocused. However, the dephasing due to spin coupling is not refocused and the components of the lactate doublet will fall out of phase with other uncoupled signals in the spectrum. The echo time, TE, of a sequence is the time between excitation of the spin system and formation of the echo. At TE the components of the doublet are separated in phase,  $\phi$ , by:

$$\phi = 2\pi/TE \quad [6]$$

In the case of the lactate doublet,  $J = 7$  Hz. If a spin echo is acquired at TE = 135 ms, the doublet will appear inverted relative to the prominently visible singlets (see Figure 10).



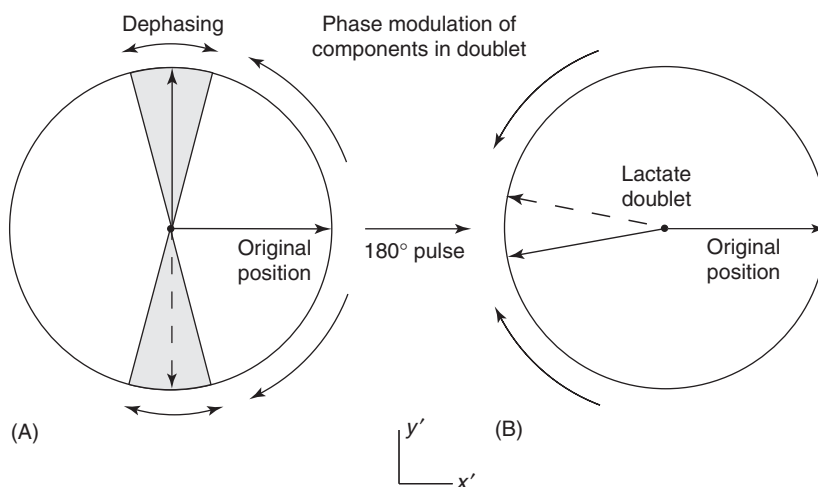
**Figure 8** The spin-spin coupling of lactate. The three protons in the  $CH_3$  group can arrange themselves in eight different combinations of orientations. Arrangements with two spins 'up' are all equivalent with respect to their effect on the CH proton. The result of spin coupling is that the peak from the CH group splits into a quartet with the middle two components being three times the amplitude of the outer two components. The  $CH_3$  peak splits into a doublet.

## Equipment

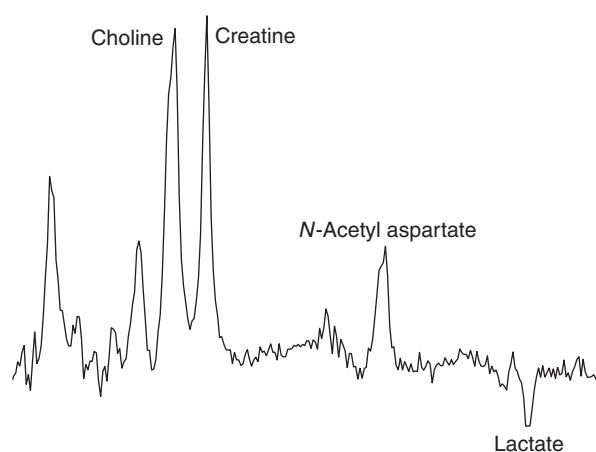
The specific hardware used for NMRS studies is dependent on the application but is generally very similar to that used for medical magnetic resonance imaging studies.

### Magnet and Shim Coils

The central piece of equipment is the magnet that generates the static magnetic field. The strength of the applied magnetic field is the main variable that can be exploited to improve the sensitivity of NMRS experiments. Persistent superconducting magnets are the most common type used. These have coil windings made of superconducting material through



**Figure 9** The phase modulation of the lactate doublet due to spin coupling. The view is of the  $x'$ - $y'$  rotating frame with the rotation frequency set halfway between the two spectral components. (A) The two components phase modulate away from each other and spins in each component de-phase because of field inhomogeneities. (B) Following a  $180^\circ$  pulse, which flips the spins about the  $x'$ -axis, the dephasing of the spins due to field inhomogeneities is refocused. However, the phase modulation due to spin coupling is not refocused. Eventually the doublet will become  $180^\circ$  out of phase with its original orientation.



**Figure 10** A  $^1\text{H}$  spectrum from a chick brain *in vivo* at 135 ms echo time. The major peaks are labeled and lactate is inverted relative to the uncoupled peaks in the spectrum. Spectrum taken using a 7 T Bruker Biospec spectrometer.

which the energizing current flows continuously. The materials that are used for these windings must be kept below their critical temperature for superconductivity and are immersed in liquid helium for this purpose. The presence of a magnetic field lowers the critical temperature placing a limit on the field achievable using a particular material. Niobium–titanium alloy is commonly used and can support a field up to 10 T in liquid helium; newer materials offer even better performance. Horizontal bore systems, suitable for *in vivo* use, are currently available up to  $\sim 11$  T. Vertical bore systems, suitable for high-resolution NMRS of samples, are currently available up to 21 T. The stability of superconducting magnets is very good with field drift typically less than  $0.05 \text{ ppm h}^{-1}$ . Equally as important as stability is the uniformity of the magnetic field. Spatial inhomogeneities in the magnetic field mean that a resonance at a specific ppm will have a different actual resonance (in hertz) depending on its location within the field. The result is that spectral lines are broadened and resolution between lines is lessened. Typically the nonuniformity over the region of interest must be less than the minimum chemical shift it is desired to resolve. A magnet may have a set of superconducting coils to shim (remove inhomogeneities from) the main field; currents in these coils are established that produce magnetic fields that oppose the inhomogeneities in field produced by the main coil. Introducing a sample into the magnet poses an additional problem, especially for *in vivo* samples. Boundaries between materials of different magnetic susceptibility generate field gradients that lead to broadening of spectral lines. These are corrected for by adjusting the currents in a set of

room temperature resistive shim coils. Typically the observation of the linewidth of a prominent spectral line is used as feedback for the iterative adjustment of these shim currents.

### Gradient Coils

In the case of a narrow bore vertical spectrometer designed for studying homogeneous samples, signals are typically collected from the whole of the sample. However, when studying heterogeneous samples, typically *in vivo* samples in a wide bore horizontal spectrometer, it is often necessary to localize the acquisition of signal to specific structures within the sample. The methods for localization are similar to those used in magnetic resonance imaging and involve the use of sequences of temporally applied field gradients along with RF. Spectrometers for medical applications include gradient coils that are designed to produce field gradients in three orthogonal directions. Gradient coils also allow for an imaging as well as a spectroscopy capability.

### RF Coil

RF coils are resonant circuits that are designed to produce oscillating magnetic fields with a suitable geometry to excite a sample and reciprocally to detect the subsequent NMR signal. The choice of coil is strongly influenced by the application for which it is to be used. In the simplest arrangement, a single coil is used to both transmit RF and to detect the signal. Alternatively, it may be advantageous to use one coil for excitation, typically a volume coil generating a uniform oscillating field, and one for signal detection whose geometry is matched to a region of interest in the sample, particularly *in vivo*. For a coil to work efficiently it must be tuned to the resonant frequency of the nuclei under investigation and impedance matched to the transmitter, receiver, or both depending on the mode of operation. Often a coil will be designed to tune to both the proton resonant frequency and that of the nuclei under investigation or be used in conjunction with a second coil tuned to protons. The larger NMR signal from protons is typically used to provide interactive feedback whilst shimming.

## Signal Acquisition and Processing

When the FID is acquired, it is mixed with an artificially generated reference signal so as to convert the RFs to the audio frequency range. The frequencies contained in the converted signal are now equivalent to the difference between the original signal and the reference. Thus if the original signal contains simply

a frequency  $\nu_1$  and was mixed with a reference  $\nu_r$ , the resulting signal would be  $(\nu_1 - \nu_r)$ . It is this converted signal that is then sampled and processed to produce the final spectrum. The signal is sampled into two channels that correspond to the projection of  $\mathbf{M}_0$  along orthogonal directions in the  $x$ - $y$  plane. If a quadrature coil is used the data in these channels are measured independently. If a single channel coil-receiver arrangement is used then the second channel is simulated. These two channels, termed real and imaginary, have a  $90^\circ$  phase difference and are combined to form a complex dataset. This is then Fourier transformed and, following a suitable phase adjustment, the real component is displayed as the final spectrum.

### Dwell Time and Sampling Time

The NMR signal is sampled continuously over the acquisition time,  $T$ , with a dwell time between samples  $\Delta t$ . For a given dwell time a signal can only be unambiguously identified if its frequency  $\nu$  is below a critical threshold termed the Nyquist frequency. Above the Nyquist frequency, signals will be aliased (see Figure 11). Aliasing occurs when a high frequency appears to look like a low frequency because the sampling rate is too low, thus the signal is misrepresented or aliased as the lower signal. If quadrature detection is used, then negative frequencies

can be distinguished from positive frequencies and the spectral width, SW - the range of frequencies that can be unambiguously detected, is centered round the reference frequency and covers the interval:

$$-\frac{1}{2\Delta t} < \nu_r < \frac{1}{2\Delta t} \quad [7]$$

For one-channel detection it is impossible to distinguish between positive and negative frequencies and so SW covers the interval  $0 < \nu_r < 1/2\Delta t$ . The acquisition time determines the spectral resolution of the final spectrum,  $\Delta\nu$ :

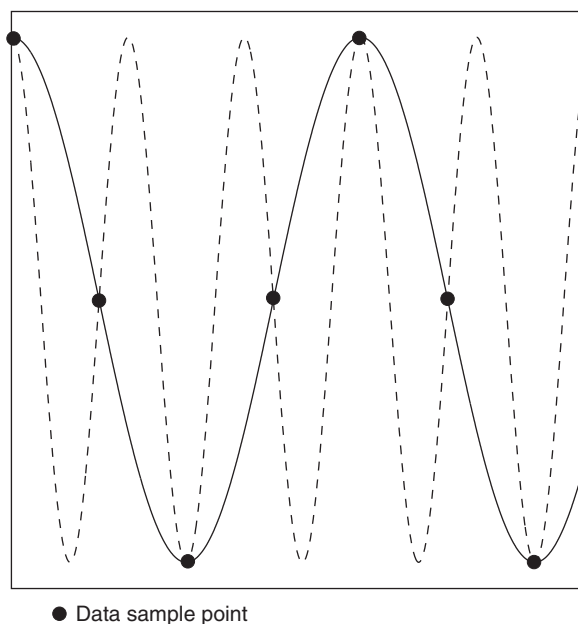
$$\Delta\nu = \frac{1}{T} \quad [8]$$

The acquisition parameters  $T$ ,  $\Delta t$ , and  $\nu_r$  are chosen so as to have an SW sufficiently large and properly centered to cover the chemical shift range of the spectrum with an appropriate spectral resolution.

### Filtering the FID

Often it is desirable to apply a filter to the digitized FID. Sometimes the sampling of the FID is completed before reaching the noise level and can be thought of as multiplying a complete FID by a step function. Fourier transforming these two functions would give a convolution of the spectrum with a sinc function producing lineshape distortions. A function can be applied to smooth the FID to zero within the acquisition period. Typically the FID is multiplied by a decaying exponential function. The resulting spectrum will be convoluted with Lorentzian function, the Fourier transform of the decaying exponential, resulting in the removal of the artifact but at the cost of broadening the spectral components. This technique is also used to improve the signal-to-noise ratio (SNR) of a spectrum but with the same costs.

Broad spectral components have large amplitudes in the first few digitized points of the FID, whereas narrow spectral components have longer signals. It is often advantageous to suppress these broad components. If the signal at the start of FID is reduced then the broad components are reduced relative to the narrow components. A suitable function to use is the linear ramp. The function increases from zero to one over the first few points and is used to multiply the FID. A further consequence is to apparently increase the decay constant  $T_2^*$ , resulting in narrower spectral lines. The cost is to increase the level of the noise relative to the signal in the FID resulting in a noisier spectrum. When filtering, there are always trade offs between spectral resolution and SNR; resolution enhancement leads to SNR degradation and vice versa.



**Figure 11** Data sampling. The solid line represents a wave that is sampled at a rate sufficient to identify its frequency. The dashed line represents a wave with a frequency that is above the Nyquist limit. The frequency of this wave cannot be properly identified and will be aliased as having a frequency equal to that of the solid line.

## Zero Filling

Once the FID has decayed below the noise level there is little more useful information that can be gained. However, increasing the number of data points at the end by filling with zeros can improve spectral resolution. The apparent increased sampling results, after Fourier transform, in a spectrum with more data points consequently improves the visibility of otherwise unresolved peaks. Filtering of the FID is usually applied before zero filling to avoid truncation artifacts.

## Analysis Methods

Analysis of NMR spectra may be performed on either the FID (in the time domain) or on the Fourier transformed spectrum (in the frequency domain). Determining the chemical shift and  $J$ -coupling of spectral peaks aids the identification of constituent molecules present in a sample, the measurement of physiological parameters such as pH, and also the determination of chemical structure of specific molecules. An analysis technique must be reliable, objective, and ideally fully automatic. It must also be able to cope with spectral imperfections and be accurate in accounting for lineshape distortions, phase modulation effects, and determining the baseline upon which the peaks sit. Spectra obtained *in vivo* typically do not have the resolution of *in vitro* biological or inorganic samples studied at high field and, for  $^1\text{H}$  NMRS in particular, peaks will overlap to a greater degree. The method of analysis chosen reflects both the application and the inherent limitations of the experiment. Below are brief descriptions of two commonly used analysis methods.

### Analysis in the Time Domain (Linear Combination of Decaying Exponentials)

Time domain analysis avoids the need for a Fourier transform prior to processing. The first few data points, in which signals from rapidly decaying immobile nuclei are mostly contained, and the last few that are at the noise level can be ignored in the analysis, leaving a section of FID that has good SNR and has a simplified baseline. Algorithms such as time domain analysis algorithm (VARPRO) and time domain analysis method (AMARES) are usually used to fit a sum of exponentially decaying sinusoids to the FID using nonlinear least squares. Prior knowledge is used to aid the fitting procedure. The precision of the fit may be significantly improved by fixing parameters whose values are known *a priori* such as the multiplicity,  $J$ -coupling, and relative intensities of a multiplet.

### Analysis in the Frequency Domain (LCModel)

LCModel is a widely used method of analysis, especially for *in vivo*  $^1\text{H}$  spectra. Spectra are analyzed as a linear combination of model spectra of metabolite solutions *in vitro*. The model spectra provide *a priori* information about chemical shifts, relative areas, number of equivalent protons, and phase modulation effects. However, these spectra must be obtained for each echo time at which *in vivo* spectra are to be obtained. A full basis set is required for the analysis at each specific echo time and consists of model spectra for all the metabolites that are likely to be observed *in vivo*. The method is model-free such that smooth lineshapes and baselines consistent with the data are determined rather than assumed.

**See also:** **Nuclear Magnetic Resonance Spectroscopy:** Principles; Instrumentation. **Nuclear Magnetic Resonance Spectroscopy-Applicable Elements:** Hydrogen Isotopes; Carbon-13; Fluorine-19; Nitrogen-15; Phosphorus-31; Organometallic Compounds. **Nuclear Magnetic Resonance Spectroscopy Applications:** Food; Forensic; Pharmaceutical; Proton NMR in Biological Objects Subjected to Magic Angle Spinning. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Nuclear Overhauser Effect; Multidimensional Proton; Solid-State; Surface Coil; *In Vivo* Spectroscopy Using Localization Techniques.

## Further Reading

- Bachelard H (ed.) (1997) *Magnetic Resonance Spectroscopy and Imaging in Neurochemistry*, Advances in Neurochemistry, vol. 8. New York: Plenum Press.
- Cady EB, D'Souza PC, Penrice J, and Lorek A (1995) The estimation of local brain temperature by *in vivo*  $^1\text{H}$  magnetic resonance spectroscopy. *Magnetic Resonance in Medicine* 33: 862–867.
- Danielsen ER and Henriksen O (1994) Absolute quantitative proton NMR spectroscopy based on the amplitude of the local water suppression pulse. Quantification of brain water and metabolites. *NMR in Biomedicine* 7: 311–318.
- Ernst T and Hennig J (1991) Coupling effects in volume-selective  $^1\text{H}$  spectroscopy of major brain metabolites. *Magnetic Resonance in Medicine* 21: 82–96.
- Govindaraju V, Young K, and Maudsley AA (2000) Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR in Biomedicine* 13: 129–153.
- [http://carbon.uab.es/mrui/mrui\\_Overview.shtml](http://carbon.uab.es/mrui/mrui_Overview.shtml). Web page for JMRUI spectroscopy analysis tool.
- Petroff OA and Prichard JW (1983) Cerebral pH by NMR. *Lancet* 2(8341): 105–106.
- Provencher SW (2001) Automatic quantitation of localized *in vivo*  $^1\text{H}$  spectra with LCModel. *NMR in Biomedicine* 14(4): 260–264.



Rae AIM (1992) *Quantum Mechanics*, 3rd edn. Bristol: Institute of Physics.

Vanhamme L, van den Boogaart A, and Van Huffel S (1997) Improved method for accurate and efficient

quantification of MRS data with use of prior-knowledge. *Journal of Magnetic Resonance* 129: 35–43.

Young IR (ed.) (2002) *Methods in Biomedical Magnetic Resonance Imaging and Spectroscopy*. Chichester: Wiley.

## Principles

**J B Grutzner**, Purdue University, West Lafayette, IN, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a rich source of molecular information. It is widely used for molecular structure analysis and for medical imaging under the name MRI (magnetic resonance imaging). NMR is a nondestructive quantitative method for the analysis of complex mixtures such as reactions, petroleum, materials, and foods. It finds major application in the determination of molecular structure and dynamics, especially of pharmaceuticals, polymers, and proteins. Diverse applications include measurement of internuclear distances, water content of food, characterization of oil deposits using the earth's magnetic field, *in vivo* studies of metabolic pathways in cells and humans, diffusion of adsorbates into porous materials, reaction dynamics, fluid flow, generation and measurement of temperatures in the millikelvin range, and online monitoring of manufacturing processes. Samples may be gases, liquids, or solids.

This article provides a description of the basic principles of NMR spectroscopy. The many applications of this technique, which make it indispensable as a tool for the analytical chemist, are described in subsequent articles, which also include discussion of advanced techniques such as the NMR of solids, multidimensional NMR, and MRI.

### Principles of NMR

NMR detects signals from nuclei in atoms and molecules when they are excited in a strong homogeneous magnetic field ( $B_0$  in tesla). NMR was invented independently by Bloch and Purcell to accurately measure nuclear magnetic moments ( $\mu$ ). It is now used to measure magnetic field strength with high precision based on known  $\mu$  values.

The nuclear magnets are ordered into  $2I + 1$  states determined by the nuclear spin quantum number  $I$ . The spin angular momentum is quantized in units of  $\hbar$  and the change in the  $z$  component is detected in the NMR experiment. For many common NMR nuclei  $I = 1/2$  ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{19}\text{F}$ ,  $^{29}\text{Si}$ ,  $^{31}\text{P}$ ) and there are two allowed states labeled spin up and spin down. Irradiation with radiofrequency (RF) electromagnetic radiation ( $\nu$  in Hz) causes spin flip transitions provided the resonance condition is satisfied:

$$h\nu = \hbar\gamma B_0 = \mu_z \beta B_0 \quad [1]$$

where  $h$  is Planck's constant ( $6.62608 \times 10^{-34}$  Js),  $\hbar = h/(2\pi)$ ,  $\beta$  is the nuclear Bohr magneton ( $eh/4\pi m_p = 5.0508 \times 10^{-27}$  J T $^{-1}$ ) and  $\mu_z$  is the  $z$ -component of the total moment  $= \sqrt{I(I+1)}\mu_z$ . The nuclear magnetogyric ratio ( $\gamma$  in rad s $^{-1}$  T $^{-1}$ ) is an exquisitely sensitive discriminator between elements and also of the immediate electronic environment of each nucleus. Tables list values of  $\mu$  in units of  $\beta$  according to different conventions. For protons, the values are 5.5854 with  $\mu = \gamma\hbar$ ; 2.7927 with  $\mu_z = \gamma\hbar I$  for the  $z$ -component of spin  $I$ ; or 4.8372 with  $\mu = \gamma\hbar\sqrt{I(I+1)}$  for the total magnetic moment.

The common information obtainable from an NMR spectrum is summarized in Table 1 and illustrated with the proton NMR spectrum of ethanol ( $\text{CH}_3\text{CH}_2\text{OH}$ ) (Figure 1). The horizontal axis, which gives the 'chemical shift' of the signal, is a linear frequency scale in hertz, which is almost always converted to the dimensionless  $\delta$  scale (ppm) with tetramethylsilane (TMS) as reference at  $\delta = 0.000$ . The  $\delta$  scale is independent of magnetic field strength and so is transferable from spectrometer to spectrometer. It is defined as

$$\delta_i = \Delta\nu_i/\nu_{\text{ref}} \quad [2]$$

where  $\nu_{\text{ref}}$  is the absolute resonance frequency of TMS (e.g., 200.010078 MHz in a field of 4.70 T) and  $\Delta\nu_i = \nu_i - \nu_{\text{ref}}$  is the difference between the frequency of the nucleus of interest ( $\nu_i$ ) and  $\nu_{\text{ref}}$ . The vertical axis is a linear relative intensity scale that is generally proportional to the number of nuclei in the resonant

Rae AIM (1992) *Quantum Mechanics*, 3rd edn. Bristol: Institute of Physics.

Vanhamme L, van den Boogaart A, and Van Huffel S (1997) Improved method for accurate and efficient

quantification of MRS data with use of prior-knowledge. *Journal of Magnetic Resonance* 129: 35–43.

Young IR (ed.) (2002) *Methods in Biomedical Magnetic Resonance Imaging and Spectroscopy*. Chichester: Wiley.

## Principles

**J B Grutzner**, Purdue University, West Lafayette, IN, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a rich source of molecular information. It is widely used for molecular structure analysis and for medical imaging under the name MRI (magnetic resonance imaging). NMR is a nondestructive quantitative method for the analysis of complex mixtures such as reactions, petroleum, materials, and foods. It finds major application in the determination of molecular structure and dynamics, especially of pharmaceuticals, polymers, and proteins. Diverse applications include measurement of internuclear distances, water content of food, characterization of oil deposits using the earth's magnetic field, *in vivo* studies of metabolic pathways in cells and humans, diffusion of adsorbates into porous materials, reaction dynamics, fluid flow, generation and measurement of temperatures in the millikelvin range, and online monitoring of manufacturing processes. Samples may be gases, liquids, or solids.

This article provides a description of the basic principles of NMR spectroscopy. The many applications of this technique, which make it indispensable as a tool for the analytical chemist, are described in subsequent articles, which also include discussion of advanced techniques such as the NMR of solids, multidimensional NMR, and MRI.

### Principles of NMR

NMR detects signals from nuclei in atoms and molecules when they are excited in a strong homogeneous magnetic field ( $B_0$  in tesla). NMR was invented independently by Bloch and Purcell to accurately measure nuclear magnetic moments ( $\mu$ ). It is now used to measure magnetic field strength with high precision based on known  $\mu$  values.

The nuclear magnets are ordered into  $2I + 1$  states determined by the nuclear spin quantum number  $I$ . The spin angular momentum is quantized in units of  $\hbar$  and the change in the  $z$  component is detected in the NMR experiment. For many common NMR nuclei  $I = 1/2$  ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{19}\text{F}$ ,  $^{29}\text{Si}$ ,  $^{31}\text{P}$ ) and there are two allowed states labeled spin up and spin down. Irradiation with radiofrequency (RF) electromagnetic radiation ( $\nu$  in Hz) causes spin flip transitions provided the resonance condition is satisfied:

$$h\nu = \hbar\gamma B_0 = \mu_z \beta B_0 \quad [1]$$

where  $h$  is Planck's constant ( $6.62608 \times 10^{-34}$  Js),  $\hbar = h/(2\pi)$ ,  $\beta$  is the nuclear Bohr magneton ( $eh/4\pi m_p = 5.0508 \times 10^{-27}$  J T $^{-1}$ ) and  $\mu_z$  is the  $z$ -component of the total moment  $= \sqrt{I(I+1)}\mu_z$ . The nuclear magnetogyric ratio ( $\gamma$  in rad s $^{-1}$  T $^{-1}$ ) is an exquisitely sensitive discriminator between elements and also of the immediate electronic environment of each nucleus. Tables list values of  $\mu$  in units of  $\beta$  according to different conventions. For protons, the values are 5.5854 with  $\mu = \gamma\hbar$ ; 2.7927 with  $\mu_z = \gamma\hbar I$  for the  $z$ -component of spin  $I$ ; or 4.8372 with  $\mu = \gamma\hbar\sqrt{I(I+1)}$  for the total magnetic moment.

The common information obtainable from an NMR spectrum is summarized in Table 1 and illustrated with the proton NMR spectrum of ethanol ( $\text{CH}_3\text{CH}_2\text{OH}$ ) (Figure 1). The horizontal axis, which gives the 'chemical shift' of the signal, is a linear frequency scale in hertz, which is almost always converted to the dimensionless  $\delta$  scale (ppm) with tetramethylsilane (TMS) as reference at  $\delta = 0.000$ . The  $\delta$  scale is independent of magnetic field strength and so is transferable from spectrometer to spectrometer. It is defined as

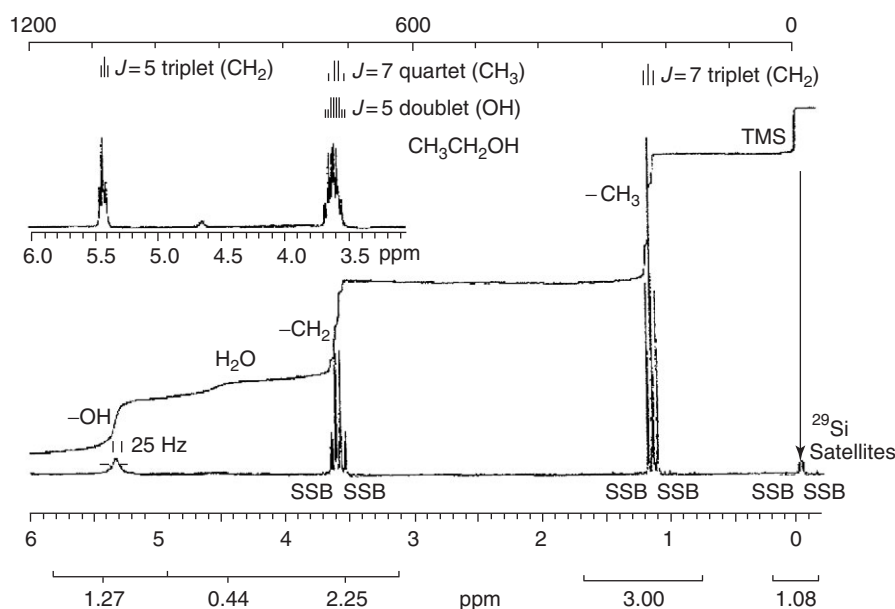
$$\delta_i = \Delta\nu_i/\nu_{\text{ref}} \quad [2]$$

where  $\nu_{\text{ref}}$  is the absolute resonance frequency of TMS (e.g., 200.010078 MHz in a field of 4.70 T) and  $\Delta\nu_i = \nu_i - \nu_{\text{ref}}$  is the difference between the frequency of the nucleus of interest ( $\nu_i$ ) and  $\nu_{\text{ref}}$ . The vertical axis is a linear relative intensity scale that is generally proportional to the number of nuclei in the resonant

**Table 1** Basic information in an NMR spectrum

Observable	NMR parameter	Symbol (units)	Chemical information
Peak position	Chemical shift	$\delta$ (ppm) (absolute Hz)	Functional group identification
Peak multiplicity	Coupling constant	$J$ (Hz)	Number and type of nuclei within three bonds of the nucleus being detected
Peak intensity	Integrated area		Count of nuclei of given type
Lineshape and linewidth	Spin–spin relaxation time	$T_2 = (1/\pi\Delta\nu)$ (s) Peak width at half-maximum intensity ( $\Delta\nu$ , Hz)	Lifetime of spin state determined by dynamic processes and local magnetic environment
Time dependence of intensity <sup>a</sup>	Spin–lattice relaxation time and nuclear Overhauser effect (nOe)	$T_1$ (s), nOe	Molecular dynamics and magnetic interactions

<sup>a</sup>Not shown in **Figure 1**. Requires repetitive experiments.

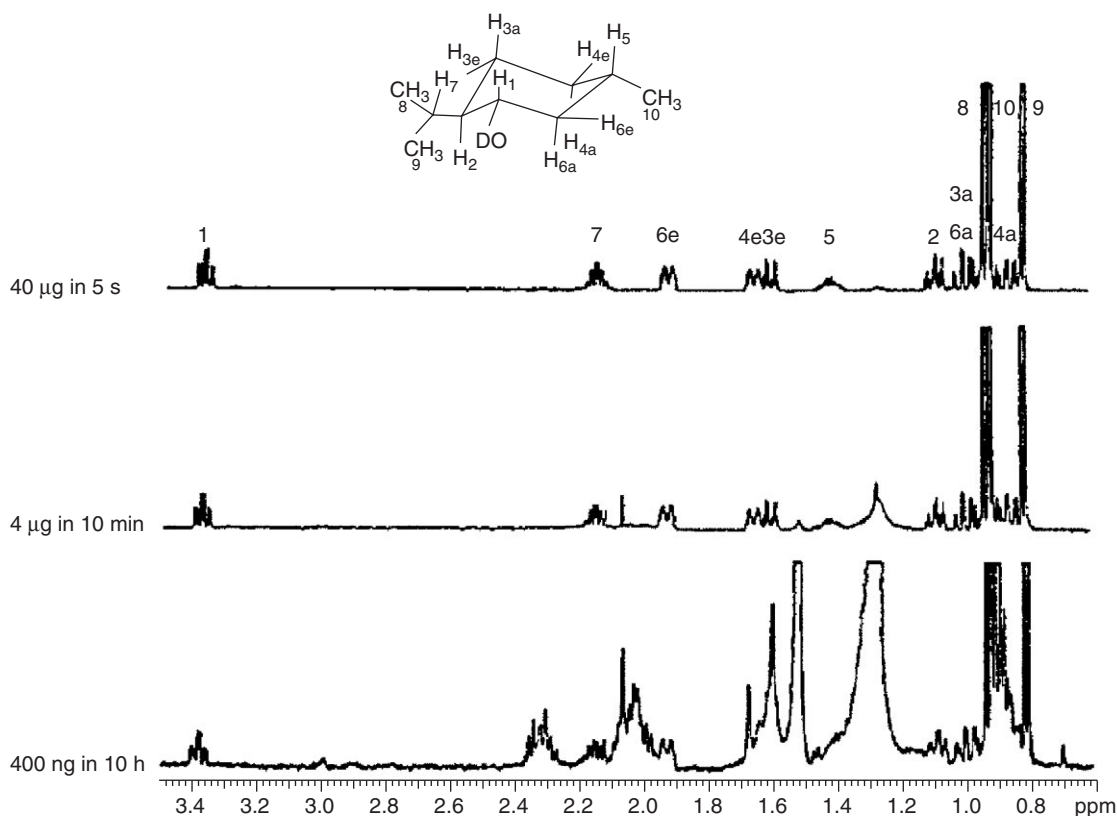


**Figure 1** 200 MHz proton NMR spectrum of ethanol containing water (13 mol.%) and tetramethylsilane (TMS) as reference. The inset shows dry ethanol with OH coupling. Splitting diagrams are shown over each multiplet. The upper scale is in Hz and lower scale in  $\delta$ /ppm. SSB indicates spinning side-bands, which are separated from the main peaks by the spinner frequency. The  $^{29}\text{Si}$  satellites on the TMS resonance arise from  $^1\text{H}$ – $^{29}\text{Si}$  coupling.  $^{29}\text{Si}$  ( $I = 1/2$ ) is present at 4% natural abundance. (Spectrum courtesy of Dr J. Hopkins.)

chemical group. More precisely, it is the relative magnitude of the magnetization precessing at the frequency of interest.

The spectrum of ethanol shown in **Figure 1** consists of five signals – a single sharp peak at  $\delta = 0.0$  from the TMS added as reference, a triplet at  $\delta = 1.16$  for the methyl group, a quartet at  $\delta = 3.61$  for the methylene group, and broad singlets at  $\delta = 5.41$  and  $4.62$ , which are the signals from the alcohol OH and the water present as a ‘contaminant’. An integral curve is shown above the spectrum giving the integrated area under each peak. Each

integral is normalized to the methyl peak intensity of 3.0 and gives the proton count in each group. The fine structure within each multiplet is generated by the  $n$  protons within three bonds of the observed proton. The multiplet intensity pattern is governed by the statistical probability of the  $2nI + 1$  possible spin arrangements of the neighboring nuclei. As  $I = 1/2$  for protons,  $n$  neighboring protons cause splitting into  $n + 1$  peaks with binomial intensity. The spacing within each multiplet is called the coupling constant,  $J$  (Hz), and the same spacing occurs in each multiplet for the pair of coupled



**Figure 2** 500 MHz  $^1\text{H}$  NMR spectrum of menthol (OD) in  $\text{CD}_2\text{Cl}_2$ . The spectra show the effects of sample size and signal averaging. Note the growing importance of solvent impurities as the sample size decreases. (Spectrum courtesy of Varian Associates.)

groups. For ethanol, the methyl group is split into a 1:2:1 triplet with  $J = 7\text{ Hz}$  by the neighboring  $\text{CH}_3$  group. Similarly, the methylene peak is split with the same spacing into a 1:3:3:1 quartet by the adjacent  $\text{CH}_3$  group. In the normal spectrum there is no coupling observed between the  $\text{CH}_2$  and OH groups because the OH bond is being broken several times per second by exchange with the water. In ultrapure ethanol (inset in **Figure 1**) when the OH exchange is slow, the  $\text{CH}_2$  group appears as a doublet (5 Hz) of quartets (7 Hz) and the OH signal is a 1:2:1 triplet with  $J = 5\text{ Hz}$ . The methyl triplet is unchanged as it is separated by four bonds from the OH.

Peak linewidths may be used to measure the spin life time. In **Figure 1**, the linewidths of the water and alcohol OH peaks are substantially broader than those from protons bound to carbon. The broadness shows that the water protons and alcohol OH are exchanging on the order of 10 times per second.

This example illustrates the power of NMR. About 0.5 ml of colorless liquid, contained in a glass tube, has been identified as ethanol containing  $\sim 15\%$  water. Less than 0.2% of other protonated impurities are present and the kinetics of proton

exchange was estimated. The whole experiment took a few seconds. Sensitivity is always an issue for NMR analysis. Millimolar solutions and milligram quantities are required for most applications. The effect of sample size on spectral quality is shown in **Figure 2**. Spectral analysis has been achieved on  $10\text{ }\mu\text{g}$  of a compound in  $0.3\text{ }\mu\text{l}$  of solution and spectra of 100 ng samples have been recorded and interpreted.

## NMR and the Periodic Table

The NMR periodic table (**Table 2**, *Bruker Almanac* (2003) Karlsruhe: Bruker Biospin) provides the nuclear properties for the main NMR active isotopes.

### Nuclear Spin Quantum Number ( $I$ )

In a magnetic field, a nucleus exists in  $2I + 1$  possible spin states. Protons and neutrons have  $I = 1/2$ . The vector sum of coupled proton and neutron spins gives the nuclear spin quantum number for each nucleus. Isotopes with an even number of both protons and neutrons may have zero spin and are NMR inactive (e.g.,  $^{12}\text{C}$ ,  $^{16}\text{O}$ ).

**Table 2** List of elements showing the nuclear properties of the main NMR-active isotopes

1	H										2	He
1 200 99.985 1.00 -	2 30.701 1.56×10 <sup>-2</sup> 9.65×10 <sup>-3</sup> 2.8×10 <sup>-3</sup>	3 213.317 -	Atomic number								3 152.355 1.3×10 <sup>-4</sup> 0.442 1/2	
1 200 99.985 1.00 -	H 30.701 1.56×10 <sup>-2</sup> 9.65×10 <sup>-3</sup> 2.8×10 <sup>-3</sup>	3 213.317 -	Symbol									
1 200 99.985 1.00 -	1 30.701 1.56×10 <sup>-2</sup> 9.65×10 <sup>-3</sup> 2.8×10 <sup>-3</sup>	3 213.317 -	isotope number									
1 200 99.985 1.00 -	1 30.701 1.56×10 <sup>-2</sup> 9.65×10 <sup>-3</sup> 2.8×10 <sup>-3</sup>	3 213.317 -	NMR frequency in 4.689 T (46.98 kG)									
1 200 99.985 1.00 -	1 30.701 1.56×10 <sup>-2</sup> 9.65×10 <sup>-3</sup> 2.8×10 <sup>-3</sup>	3 213.317 -	Natural isotopic abundance (%)									
1 200 99.985 1.00 -	1 30.701 1.56×10 <sup>-2</sup> 9.65×10 <sup>-3</sup> 2.8×10 <sup>-3</sup>	3 213.317 -	Relative sensitivity for equal number of nuclei									
1 200 99.985 1.00 -	1 30.701 1.56×10 <sup>-2</sup> 9.65×10 <sup>-3</sup> 2.8×10 <sup>-3</sup>	3 213.317 -	Electric quadrupole moment (10 <sup>-24</sup> cm <sup>2</sup> )									
1 200 99.985 1.00 -	1 30.701 1.56×10 <sup>-2</sup> 9.65×10 <sup>-3</sup> 2.8×10 <sup>-3</sup>	3 213.317 -	Denotes occurrence of additional known magnetically active nuclear isotopes which are radioactive or of very low natural abundance									
1 200 99.985 1.00 -	1 30.701 1.56×10 <sup>-2</sup> 9.65×10 <sup>-3</sup> 2.8×10 <sup>-3</sup>	3 213.317 -	Nuclear spin (h)									
3 6 29.431 7.42 8.50×10 <sup>-4</sup> 6.9×10 <sup>-4</sup> 1	Li 7 72.727 92.58 0.293 -3×10 <sup>-2</sup> 3/2 *	4 9 28.107 100 1.39×10 <sup>-2</sup> 5.12×10 <sup>-2</sup> -3/2									5 10 21.493 19.58 1.59×10 <sup>-2</sup> 7.4×10 <sup>-2</sup> 3	B 11 64.168 80.42 0.165 3.55×10 <sup>-2</sup> 3/2
11 23 52.903 100 9.25×10 <sup>-2</sup> 0.14 - 0.15	Na 23 12.239 10.13 2.67×10 <sup>-3</sup> -5/2	12 25 12.239 10.13 2.67×10 <sup>-3</sup> -5/2									13 27 52.115 100 0.206 0.149 5/2	Al 27 39.730 4.7 7.84×10 <sup>-3</sup> -1/2
19 39 9.333 93.08 5.08×10 <sup>-2</sup> 0.11 3/2	K 41 13.457 100 7.02 2.69×10 <sup>-3</sup> -7/2	20 43 13.457 100 7.02 2.69×10 <sup>-3</sup> -7/2									31 69 48.006 60.4 6.91×10 <sup>-2</sup> 0.178 3/2	Ga 71 60.990 39.6 0.142 0.112 3/2
37 85 19.310 72.15 1.05×10 <sup>-2</sup> 0.27 5/2	Rb 87 65.443 27.85 0.175 0.13 3/2 *	38 86 22.226 100 6.86×10 <sup>-3</sup> -9/2									49 107 43.733 42.8 0.345 1.14 9/2	In 113 44.366 12.28 0.347 1.16 9/2
55 133 26.234 100 4.74×10 <sup>-2</sup> -3×10 <sup>-3</sup> 7/2	Cs 135 26.386 6.59 4.5×10 <sup>-3</sup> 0.25 3/2 *	56 137 28.252 99.911 8.9×10 <sup>-2</sup> 6.38×10 <sup>-3</sup> 0.2 3/2									81 203 36.655 16.84 2.5×10 <sup>-5</sup> 0.59 3/2 *	Tl 205 115.416 70.5 0.182 0.5 1/2
87 141 58.582 100 0.293 -5.9×10 <sup>-2</sup> 5/2	Fr 223 22.703 100 0.293 -5.9×10 <sup>-2</sup> 5/2	88 226 22.703 100 0.293 -5.9×10 <sup>-2</sup> 5/2									103 209 32.139 100 0.137 -0.4 9/2	Lr 262 103.906 100 0.137 -0.4 9/2
91 147 58.582 100 0.293 -5.9×10 <sup>-2</sup> 5/2	Pa 231 22.703 100 0.293 -5.9×10 <sup>-2</sup> 5/2	92 235 22.703 100 0.293 -5.9×10 <sup>-2</sup> 5/2</										

\*Denotes occurrence of additional known magnetically active nuclear isotopes that are radioactive or of very low natural abundance.



### Quadrupole Moment ( $Q$ , $10^{-24} \text{ cm}^2$ )

Isotopes with  $I > 1/2$  are called quadrupolar nuclei. The nucleus responds to both magnetic and electric fields. Quadrupolar nuclei respond to the electric field gradient at the nucleus through the quadrupole coupling,  $\chi$ :

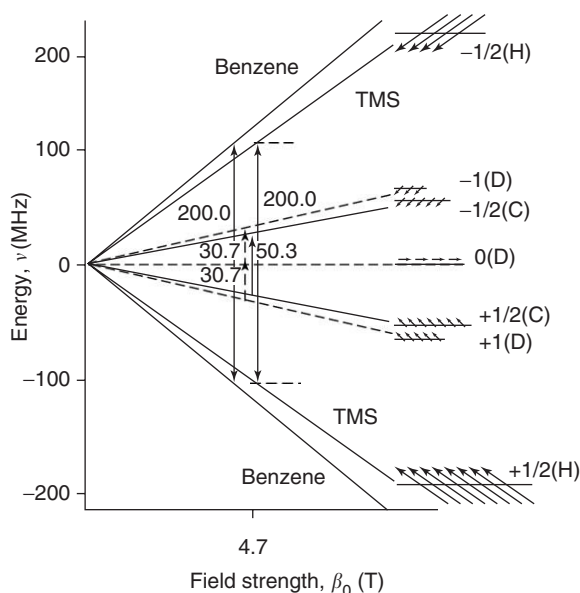
$$\chi = e^2 Q q_{zz} / h \text{ (Hz)} \quad [3]$$

where  $Q$  is the intrinsic electric quadrupole moment and  $eq_{zz}$  measures the effective electric field gradient at the nucleus induced by local bonding. In a symmetrical environment, e.g.,  $^{14}\text{N}$  in the  $\text{NH}_4$  cation,  $q$  and hence  $\chi$  may be zero. For monovalent elements, e.g.,  $^2\text{H}$  and  $^{79}\text{Br}$ , the electric field gradient is directed along the bond axis and is nonzero. The spectra of quadrupolar nuclei with large  $Q$  values are usually very broad since the electric field gradient is modulated by molecular motion and the nuclear relaxation time becomes very short. The linewidth is inversely proportional to the spin lifetime, and linewidths of the order of kilohertz to megahertz are common. The spectra of spin  $1/2$  nuclei are influenced by coupling to quadrupolar nuclei. The spectra of  $^{14}\text{N}$ -H protons are often broad because the rate of interchange of the  $^{14}\text{N}$  spin states is of the same order of magnitude as the  $^{14}\text{N}$ -H spin-spin coupling. This contrasts with O-H protons, where chemical bond breaking is responsible for line broadening. For  $\text{CDCl}_3$ , D has a small  $Q$  and the carbon is split into a 1:1:1 triplet, but  $^{35}\text{Cl}$  has a large  $Q$  and relaxes very rapidly, and so no coupling between  $^{13}\text{C}$  and  $^{35}\text{Cl}$  is observed.

In solid samples, the magnitude of  $\chi$  depends on the orientation of the electric field gradient with respect to  $B_0$ . As a consequence, deuterium NMR has been developed as a sensitive probe of molecular motion in materials over the 0–125 kHz range. A novel probe capable of rotating a sample about two different axes simultaneously has been developed to obtain narrow peaks for quadrupolar nuclei in solids. Quadrupolar nuclei give narrower signals at higher fields, and pulse sequences have been developed to further narrow lines.

### Nuclear Resonance Frequency ( $\nu$ , MHz)

Table 2 lists the resonance frequencies of each nucleus in a magnetic field of 4.698 T. Protons resonate at 200.0 MHz at this field. The frequency values are equivalent to  $\gamma$  as  $\gamma = 2\pi\nu/B_0$ . For protons  $\gamma = 26.7519 \times 10^7 \text{ rad T}^{-1} \text{ s}^{-1}$ . The resonance frequency in other fields is obtained by multiplying by the ratio of the fields. The proton resonance frequency is often used synonymously with magnetic field strength, as illustrated in Figure 3.



**Figure 3** The spin energy diagram for  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^2\text{H}$  ( $I=1$ ) as a function of field strength against energy ( $E=h\nu$ ) showing allowed transitions at 4.698 T. The energy difference between TMS and benzene protons has been exaggerated to show that, at fixed field, TMS resonates at lower frequency, while at fixed frequency, TMS resonates at higher field. The orientation of the nuclear moments and spin populations (eqn [4]) are shown on the right of the diagram.

### Natural Isotopic Abundance and Relative Sensitivity

The product of isotopic abundance and relative sensitivity determines the overall sensitivity to detection of the isotope relative to hydrogen with its 99.985% natural abundance and unit relative sensitivity. Low  $\gamma$  nuclei may be quite difficult to detect as sensitivity is proportional to  $\gamma^3$  (e.g.,  $^{57}\text{Fe}$  which is  $3 \times 10^{-5}$  times less sensitive than protons for equal numbers of nuclei). Sensitivity may be enhanced by repetitive spectral accumulation as the signal-to-noise (S/N) ratio is proportional to the square root of the number of accumulations. It requires  $10^9$  accumulations to give a  $^{57}\text{Fe}$  signal 1/50 the S/N of protons – a daunting task.

### The NMR Experiment

The essential components of an NMR spectrometer are an exceptionally stable magnet, a tunable source of RF, a high-power broadband amplifier, a high-sensitivity resonantly tuned coil surrounding the sample, and a computer for rapid data acquisition and analysis. Samples may be solids, liquids, or gases. Solutions (millimoles to moles per liter) are most widely used. There are four basic steps in an NMR experiment, as described below.

## Generation of Magnetization

When a sample is placed in a magnetic field the nuclei immediately experience a torque and begin to precess at their Larmor frequencies ( $\nu$ ) in the plane perpendicular to the magnetic field direction. The nuclear spins are oriented according to their magnetic moments. The energies of each of the  $2I + 1$  energy states are established very rapidly (the time-scale is  $1/\nu$ ). Initially, each spin state will be equally populated as the levels were degenerate without an applied field. Energy exchange with the surroundings occurs as spins flip. The spin temperature of the sample drops from infinity (equally populated states) until a Boltzmann distribution of spin populations is established at equilibrium:

$$n_i/n_0 = \exp(-\Delta E/kT) = \exp(-\hbar\gamma B_0/kT) \quad [4]$$

The approach to equilibrium is a first-order rate process and is exponential in time (cf. Newton's law of cooling). The time constant is the spin lattice relaxation time,  $T_1$  (s). The  $T_1$  of water at 200 MHz is  $\sim 15$  s. The resultant  $Z$  magnetization –  $M_z = \gamma\hbar(n_0 - n_i)$  arising from the excess population in the lower spin state – is then available for manipulation in the experiment (see Figure 3). As  $\Delta E$  for NMR is of the order of millijoules ( $\sim 2 \text{ mJ mol}^{-1}$  for protons at 4.7 T), the  $n_i/n_0$  ratio is very close to 1. However, it is this excess population in the lower state ( $n_0 - n_i \sim 3$  per  $10^5$  for protons) that generates the NMR signal. High  $\gamma$  nuclei, high fixed field, and low  $T$  all favor a larger population excess and hence higher sensitivity.

## Resonant Energy Absorption and Coherence Creation

The spins resonantly exchange energy with an oscillating RF electromagnetic field applied through a tuned coil. In pulse Fourier transform (FT) spectroscopy the RF field oscillating at  $\nu_0$  is applied as a large-amplitude, short pulse with magnetic field strength  $2\gamma B_1$ . For a typical spectrometer,  $\gamma B_1$  is of the order of 25 kHz and the pulse duration is 10  $\mu\text{s}$ . This establishes the width of the spectral range which can be excited by such a pulse at  $\sim \pm 12.5$  kHz ( $\pm \gamma B_1/2$ ).

Radiofrequency radiation induces up or down spin flips with equal probability. The number of quanta exchanged between the spins and the RF coil is enormous. Consequently, the overall behavior of the system is statistical and can be described classically. Energy is absorbed by the spins provided there is an excess population in the lower state at the time the RF is applied. This spin energy change is far too

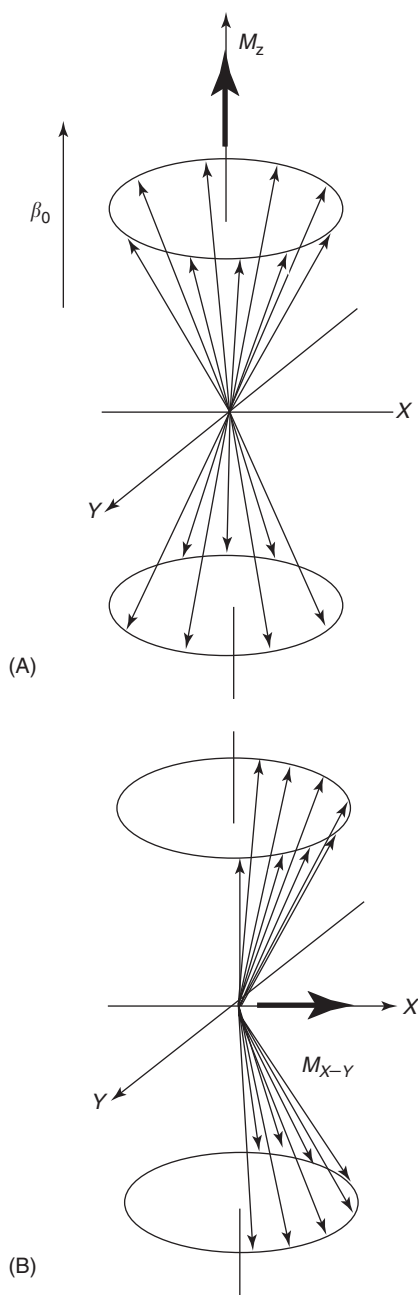
small to influence sample temperature. A high-energy RF pulse does induce substantial heating, especially if the solvent has a high dielectric constant or the sample is highly conducting. However, the heating is electrical, not magnetic. Most importantly, a new coherent spin motion is created in the  $x$ - $y$  plane perpendicular to  $B_0$ . At equilibrium there is no  $x$ - $y$  field and so no  $x$ - $y$  magnetization. When the  $B_{1x}$  field is applied, a new coherent precessional motion occurs about it and  $x$ - $y$  magnetization is detectable (see Figure 4). The spin magnetization rotates about the combined  $B_0$  static field and the  $B_1$  RF field oscillating at  $\nu_0$ . The description of the spin motion is simplified by transformation to the rotating frame where  $B_{1x'}$  is static since the rotating frame frequency is chosen to be the RF oscillator frequency. When the RF pulse is applied the magnetization rotates about  $B_{1\text{eff}}$  for the pulse duration ( $t_p$ , s). The effective  $B_1$  field is the vector sum of  $B_{1x'}$  and  $B_{z'} = (\nu_i - \nu_0)/\gamma$ . Pulses are usually described in terms of the angle  $\theta$  through which a resonant nucleus would rotate in time  $t_p$ :

$$\theta = \gamma B_1 t_p \quad [5]$$

There are two important pulses: a  $90^\circ$  pulse, which creates maximum magnetization in the  $x$ - $y$  plane and hence maximum signal; and a  $180^\circ$  pulse, which inverts the spin population present immediately prior to the pulse (see Figure 5). The  $180^\circ$  pulse is particularly important for the accurate determination of  $T_1$  and  $T_2$ . The  $180^\circ$  pulse initiates time reversal for the  $x$ - $y$  magnetization resulting in a Hahn spin echo.

## Detection

The  $x$ - $y$  magnetization created in the excitation step is minute in absolute terms. Thus it is essential that detection is separated in either time or space from excitation. The magnetization precessing at each nuclear resonance frequency induces an alternating current in the receiver coil, which is closely wound around the sample. In Bloch's original work separate transmitter and receiver coils were mounted orthogonally to one another to provide spatial isolation of the detector from the exciting RF field, which was applied continuously. The magnetic field was swept to bring signals into resonance with the fixed frequency detector in this continuous-wave (CW) NMR experiment. In a modern pulse spectrometer, the same coil is used for excitation and detection. With pulse excitation, the detected signals are separated in time. A few microseconds after the high-power excitation has been turned off, the detector is gated on.



**Figure 4** The precessing spins and net magnetization (A) at equilibrium and (B) following a  $90^\circ$  pulse. Note the equalization of populations and generation of  $x$ - $y$  coherence. (Reproduced with permission from Campbell ID and Dwek RA (1984) *Biological Spectroscopy*. Menlo Park, CA: Benjamin/Cummings.)

The multiple frequencies present are resolved by signal processing.

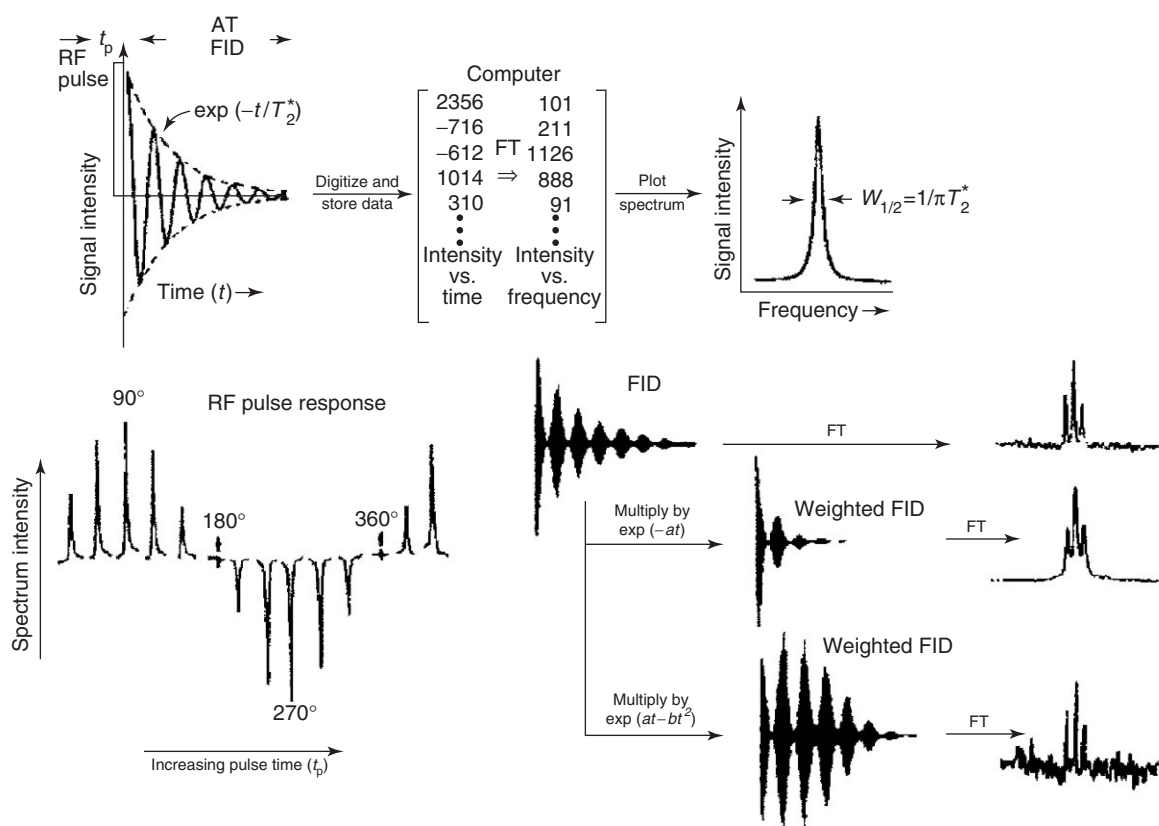
### Signal Processing

The oscillating output current from the receiver coil is amplified and mixed with the original excitation

signal. This phase-sensitive detection of the signal greatly improves the S/N ratio. The resulting signal oscillates at the difference between the transmitter frequency and the NMR resonance and is now at audiofrequencies, usually  $<20$  kHz. The multicomponent time-decaying analog signal (free induction decay, FID, or Bloch decay) is digitized into amplitude and time values in a computer (Figure 5). The process of Fourier transformation converts the FID into its frequency-equivalent spectrum. The spectrum is computed with  $NP$  points as a sum of cosine waves (real) and a sum of sine waves (imaginary) in the frequency domain, each containing  $NP/2$  points. The Nyquist requirement for a minimum of two data points to define a wave sets the frequency range. The best digital resolution – smallest resolvable frequency difference between adjacent peaks – is set by the total acquisition time,  $AT$  ( $\Delta\nu_{\min} = 1/AT$ ). This corresponds to resolving two waves whose frequencies differ by a single data point during the acquisition. There is a constant sampling time for each data point in the FID called the dwell time (DT). The finite time for signal sampling sets the highest frequency ( $f_{\max}$ ), which can be accurately represented such that  $f_{\max} = 1/2DT$ . The amplitudes of any higher frequencies (including noise!) are digitized identically to frequencies in the window and appear at  $\nu = \nu_i - \nu_{\max}$  (fold back). Ubiquitous white noise contains all possible frequencies. For realistic S/N, electronic filters with defined frequency bandwidth are essential to limit fold-back of noise from frequencies outside the spectral window.

Prior to Fourier transformation, a variety of digital weighting functions may be applied to improve the spectral appearance (Figure 5). For example, a matched filter – an exponentially decaying function with a time constant equal to the  $T_2$  decay time of the NMR signal – provides S/N optimization. Resolution enhancement is achieved at the expense of S/N by intensifying the long time tail of the FID. In addition to the standard fast FT algorithm, other data processing methods may be used for the time-to-frequency conversion. These include linear prediction – especially useful for truncated FIDs with few data points; Bayesian statistics; and maximum entropy, which are advantageous in special circumstances but with the cost of increased calculation time and memory. More elaborate acquisition schemes, which employ unequal time step functions, may be used to shorten spectral acquisition times for multidimensional experiments but at the expense of specialty data processing routines.

The absolute phase of the real and imaginary halves of the FT output cannot be established *a priori*. The experimental phase depends on the phase lags that occur inevitably in electronic circuitry. The real



**Figure 5** The basic pulsed NMR experiment showing essential processes and times. The spectrum is excited with a short resonant RF pulse of length  $t_p$ . The FID oscillates at  $\nu_1$  and decays with time constant  $T_2^*$ , which determines the linewidth and frequency following digital data acquisition and Fourier transformation. The influence of increasing pulse angle (pulse duration  $t_p$ ) is shown at lower left. For steady-state spectral acquisition, the pulse length is set for the Ernst angle ( $\cos \alpha_E = \exp(-AT/T_1)$ ). The effect of digital filtering functions applied to the FID prior to FT is shown in the lower right. The exponential filter,  $\exp(-at)$ , discriminates against high-frequency noise at the expense of line broadening; the shifted Gaussian,  $\exp(+at - bt^2)$ , generates sharp lines at the expense of the signal-to-noise ratio. (Reproduced with permission from Field LD and Sternhell S (1989) *Analytical NMR*. Chichester: Wiley. 1989; © John Wiley & Sons Ltd.)

and imaginary components must be empirically phase shifted so that the real series is in pure absorption (cosine) and the imaginary series is in pure dispersion (sine). This phase correction is optimized by maximizing the peak integral that corresponds to pure absorption. The integral for pure dispersion is zero. Furthermore, if data acquisition is delayed for a set of waves that begin with a common phase, they will dephase at a rate that is proportional to their frequency difference. Because acquisition delays appear regularly in FT-NMR experiments, a second frequency-dependent phase correction is usually required. The frequency independent phase correction is always in the range 0–360°. Much larger values may be found for the frequency-dependent correction, which is directly proportional to the acquisition time delay and the frequency offset. Finally, all the intensity information in an FID is contained in the first data point as no new magnetization is created

during acquisition. This intensity is resolved into its individual frequency contributions during the acquisition time.

## Chemical Shift

The chemical shift of a nucleus measures its immediate electromagnetic environment. The local circulation of electrons around the nucleus, induced when the sample is placed in the external magnetic field, modifies the local magnetic environment of the nucleus, resulting in the chemical shift. Its absolute magnitude is directly proportional to  $B_0$  but is only a tiny fraction of it. The range of chemical shifts varies enormously from  $\sim 1$  ppm for  $^3\text{He}$  through 20 ppm for  $^1\text{H}$  and  $^2\text{H}$  (isotopes have the same shifts) and 300–1500 ppm for main group elements  $^{13}\text{C}$ ,  $^{11}\text{B}$ ,  $^{15}\text{N}$ , and  $^{19}\text{F}$  to over 10 000 ppm for transition metals such as  $^{59}\text{Co}$  and  $^{119}\text{Sn}$  and the rare-earth elements.

Detailed correlations of chemical shifts with functional groups may be found in articles dealing with particular nuclei.

The chemical shift literature represents an interesting blend of philosophical approaches. Since frequencies are measured with great accuracy and linewidths are narrow in solution, chemical shifts are exquisitely sensitive reporters of local molecular environment. For example, the relative stereochemistry of the methyl groups, separated by up to five monomer units (13 bonds) along the chain, in polypropylene can be distinguished by carbon NMR (see Figure 6).

Empirical group additivity rules have proved to be effective predictors of chemical shifts for many classes of compounds. Extensive spectral tables are available in many texts and are very useful for structure identification provided representative model compounds are available (e.g., tripeptides for proteins, additivity rules for  $^{13}\text{C}$  in organics). The *ad hoc* explanations that accompany such assignments often have little basis in reality. For accurate interpretation of chemical shift trends, the individual components of the chemical shift tensor must be considered. Molecular orbital calculations on systems of up to

50 first-row atoms may now be done with close to experimental accuracy.

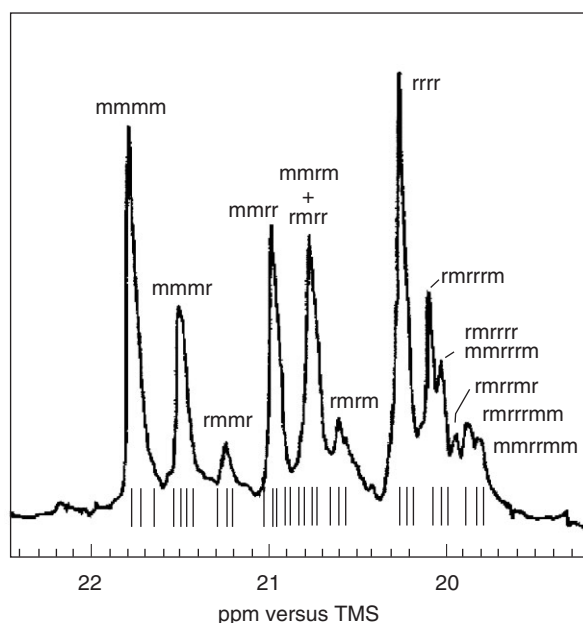
Small solvent and related effects may be useful for resolving spectral overlap, particularly when magnetically anisotropic solvents are used such as pyridine and acetone. Common NMR solvents are listed in Table 3. Anisotropic circulation of electrons in molecular groups, e.g., aromatic rings, ring currents, and shift reagents, may induce chemical shifts. The relative orientation and distance to the anisotropic group control the magnitude. This anisotropy induced shift is most important for protons as its magnitude is independent of the nucleus observed.

Chemical shift is measured on a ppm scale relative to a standard of known frequency. TMS is the practical frequency standard for  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{29}\text{Si}$ . In principle, the absolute frequency reference is the electron-free nucleus. For theoretical work, the shift of the neutral diamagnetic atom is used as it may be calculated accurately. Alternatively, the absolute scale may be established independently of the NMR experiment from the rotational splitting constant measured by microwave spectroscopy on a reference gas, e.g., CO for  $^{13}\text{C}$  and  $^{17}\text{O}$ . In Figure 7, the chemical shift/shielding scale for  $^{13}\text{C}$  is shown together with the terms used to describe changes in shift.

Chemical shift is a tensor quantity whose three principal elements can be determined for molecules with a small number of unique resonances from the turning points of the shielding pattern in solid samples. Multidimensional methods are required for resolution in more elaborate spectra. The isotropic chemical shift observed in solution is the simple average of the principal components measured in the solid. Thus,  $\sigma_{\text{iso}} = (\sigma_{xx} + \sigma_{yy} + \sigma_{zz})/3$ . Chemical shift is a highly localized interaction. Shielding is induced by electron circulation in the plane perpendicular to the applied magnetic field direction and so changes with molecular orientation. The original Ramsey perturbation theory of chemical shift defines the essential chemical shift components:

$$\begin{aligned} \sigma_{zz} = \sigma_{zz}^{\text{d}} + \sigma_{zz}^{\text{p}} = & \frac{e^2}{2mc^2} \left\langle 0 \left| \frac{x^2 + y^2}{r^3} \right| 0 \right\rangle \\ & + \left( \frac{eh}{2mc} \right)^2 \sum_n \left\{ \frac{\langle 0 | L_z | n \rangle \langle n | 2L_z / r^3 | 0 \rangle}{E_n - E_0} \right. \\ & \left. + \frac{\langle 0 | 2L_z / r^3 | n \rangle \langle n | L_z | 0 \rangle}{E_n - E_0} \right\} \end{aligned} \quad [6]$$

The first term is the diamagnetic contribution ( $\sigma^{\text{d}}$ , upfield, more shielded), which arises predominantly from the induced circulation of the inner-shell electrons generating a magnetic field opposed to the applied field. It increases monotonically with atomic



**Figure 6** The 90 MHz methyl  $^{13}\text{C}$  spectrum of atactic polypropylene in 1,2,4-trichlorobenzene at  $100^\circ$ . Empirical shift predictions for different stereoisomers reflecting the meso (m) or racemic (r) relative orientation of neighboring methyl groups. (Reprinted with permission from Schilling FC and Tonelli AE (1980) Carbon-13 nuclear magnetic resonance of atactic polypropylene. *Macromolecules* 13: 270; © American Chemical Society.)



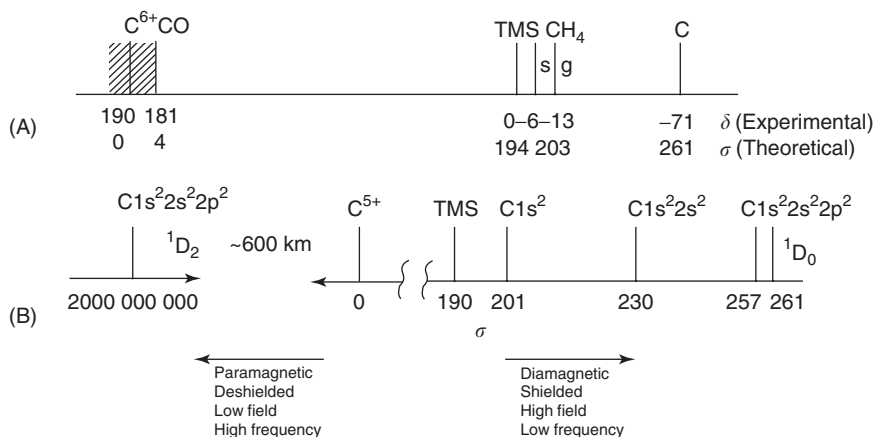
**Table 3** Useful properties of NMR solvents

Solvent	<sup>1</sup> H chemical shift <sup>a</sup> (ppm from TMS) (multiplicity)	<i>J</i> <sub>HD</sub> (Hz)	<sup>13</sup> C chemical shift <sup>a</sup> (ppm from TMS) (multiplicity)	<i>J</i> <sub>CD</sub> (Hz)	<sup>1</sup> H chemical shift of HOD <sup>b</sup> (ppm from TMS)	Physical properties				
						Density at 20 °C	Melting point (°C) <sup>c</sup>	Boiling point (°C) <sup>c</sup>	Dielectric constant	Molecular weight
Acetic acid-d <sub>4</sub>	11.65 (1) 2.04 (5)	2.2	178.99 (1) 20.00 (7)	20	11.5	1.12	17	118	6.1	64.08
Acetone-d <sub>6</sub>	2.05 (5)		206.68 (13) 29.92 (7)							
Acetonitrile-d <sub>3</sub>	1.94 (5)	2.5	118.69 (1) 1.39 (7)	21	2.1	0.84	– 45	82	37.5	44.07
Benzene-d <sub>6</sub>	7.16 (1)		128.39 (3)							
Chloroform-d	7.27 (1)	1.9	77.23 (3)	32.0	1.5	1.50	– 64	62	4.8	120.38
Cyclohexane-d <sub>12</sub>	1.38 (1)		26.43 (5)							
Deuterium oxide	4.80 (DSS) 4.81 (TSP)	1.9	163.15 (3) 34.89 (7)	29.4 21.0	4.8	1.11	3.8	101.4	78.5	20.03
<i>N,N</i> -Dimethylformamide-d <sub>2</sub>	8.03 (1) 2.92 (5) 2.75 (5)									
Dimethyl sulfoxide-d <sub>6</sub>	2.50 (5)	1.9	39.51 (7)	21.0	3.3	1.18	18	189	46.7	84.17
<i>p</i> -Dioxane-d <sub>8</sub>	3.53 (m)		66.66 (5)							
Ethanol-d <sub>6</sub>	5.29 (1) 3.56 (1) 1.11 (m)	1.7	56.96 (5) 17.31 (7)	22 19	5.3	0.91	< – 130	79	24.5	52.11
Methanol-d <sub>4</sub>	4.87 (1) 3.31 (5)		49.15 (7)	21.4	4.9	0.89	– 98	65	32.7	36.07
Methylene chloride-d <sub>2</sub>	5.32 (3)	1.1	54.00 (5)		1.5	1.35	– 95	40		86.95
Pyridine-d <sub>5</sub>	8.74 (1) 7.58 (1) 7.22 (1)		150.35 (3) 135.91 (3) 123.87 (5)	27.5 24.5 25	5	1.05	– 42	116	12.4	84.13
Tetrahydrofuran-d <sub>8</sub>	3.58 (1) 1.73 (1)	2.3	67.57 (5) 25.37 (1)	22.2 20.2						
Toluene-d <sub>8</sub>	7.09 (m) 7.00 (1) 6.98 (m) 2.09 (5)		137.86 (1) 129.24 (3) 128.33 (3) 125.49 (3) 20.4 (7)	23 24 24 19	0.4	0.94	– 95	111	2.4	100.19
Trifluoroacetic acid-d	11.50 (1)	2 (9)	164.2 (4) 116.6 (4)	22	11.5	1.50	– 15	72		115.03
Trifluoroethanol-d <sub>3</sub>	5.02 (1) 3.88 (4 × 3)		126.3 (4) 61.5 (4 × 5)							

<sup>a</sup>The <sup>1</sup>H spectra of the residual protons and <sup>13</sup>C spectra were obtained on a Varian Gemini 200 spectrometer at 295 K. The samples for the <sup>1</sup>H and <sup>13</sup>C spectra contain a maximum of 0.05 and 1.0% TMS (v/v), respectively. Since deuterium has a spin of 1, triplets arising from coupling to deuterium have the intensity ratio of 1:1:1. (m) denotes a broad peak with some fine structure. It should be noted that the chemical shifts, in particular, can be dependent on solute, concentration, and temperature.

<sup>b</sup>Approximate values only: may vary with pH, concentration, and temperature.

<sup>c</sup>Melting and boiling points are those of the corresponding compound (except for D<sub>2</sub>O). These temperature limits can be used as a guide to determine the useful liquid range of the solvents. Courtesy of Cambridge Isotope Labs, Andover, MA, USA, used by permission.



**Figure 7** The carbon chemical shift scale. (A) The experimental shift ( $\delta_{\text{TMS}} = 0.0$ ) and theoretical shielding ( $\sigma = 0$ ) scales for standard reference compounds. (B) The theoretical shielding scale for atomic carbon together with common shielding synonyms.

number. The second, paramagnetic, term is dominant for atoms containing electrons with orbital angular momentum (p, d, f, etc., but not s). It determines the induced paramagnetic circulation of electrons as the orbital angular momentum in the atom is quenched on bond formation. Linear molecules have particularly large chemical shift anisotropy, as only diamagnetic shielding is possible when the field and molecular axes are collinear. In the perpendicular orientation, large paramagnetic contributions are the rule.

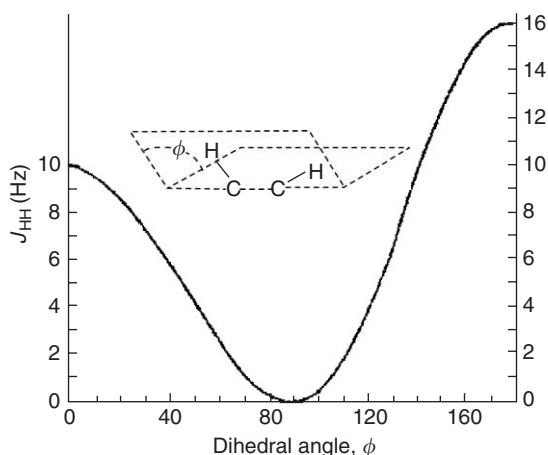
There are three essential components of eqn [6]. The  $r^{-3}$  distance term accounts for the common and generally valid correlation of chemical shift with charge density (e.g., 160 ppm/e for  $^{13}\text{C}$  in  $\pi$  systems). However, other terms can dramatically alter this simple correlation. For example, cations are known that appear upfield of anions, contradicting the simple charge prediction of cations at low field. The  $\langle 0|L_z|n\rangle$  orbital angular momentum terms select for orbital angular momentum contributions such as  $\pi_x \rightarrow \pi_y^*$ . In symmetric situations, this term can go to zero as it does in linear molecules leaving only the diamagnetic shielding. Finally, there is the excitation energy for electron promotion ( $E_n - E_0$ ), a consequence of the Ramsey perturbation treatment. This is best illustrated with transition metal shifts where a linear relationship is found between the 'forbidden'  $t_{2g} \rightarrow e_g$  optical transition energy and the chemical shift. Note that this energy term is generally not the HOMO-LUMO gap (HOMO is the acronym for highest occupied molecular orbital, and LUMO the acronym for lowest unoccupied molecular orbital). Induced electron circulation generates the chemical shift and so the perturbation terms must represent an electronic excitation that changes the electron angular momentum in direct proportion to the applied magnetic field.

Proton chemical shifts of O-H and N-H groups are strongly dependent on solvent and temperature because of changes in hydrogen bonding. The shift difference between alkyl and hydroxyl protons is used as a thermometer (methanol below room temperature, ethylene glycol above). The  $^{59}\text{Co}$  shift in  $\text{K}_3\text{Co}(\text{CN})_6$  is an even more sensitive thermometer. For solids, the  $^{207}\text{Pb}$  shift in lead nitrate is the standard thermometer.

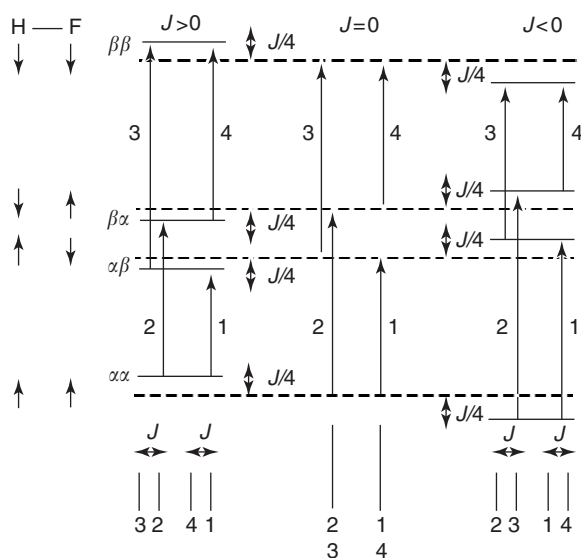
In systems with unpaired electrons (e.g., metals, radicals, paramagnetic transition metal complexes), a much larger range of chemical shifts is possible. Now the major magnetic interaction is between the nucleus of interest and the unpaired electron(s). The observed shift depends on the excess electron spin population and the coupling constant to the nucleus. The induced contact shifts or Knight shifts (in metals and conductors) often exceed 1000 ppm.

## Spin-Spin Coupling

The presence of coupling between a pair of nuclei establishes molecular connectivity usually over one, two, or three bonds and is a very powerful tool for molecular structure determination. The internuclear coupling constant  $J$  is transmitted by the bonding electrons. Coupling through as many as 10 bonds has been detected in delocalized  $\pi$  systems. By tracing coupling paths (e.g., in COSY spectra), molecular connectivity can be established uniquely and unequivocally. In a sophisticated experiment (INADEQUATE), one-bond  $^{13}\text{C}$ - $^{13}\text{C}$  coupling in natural abundance is used to trace carbon connectivity in organic molecules. The magnitude of the three-bond coupling constant often provides stereochemical information through the Karplus relationship relating  $J$  to the internuclear dihedral angle (see Figure 8).



**Figure 8** The Karplus relationship for three-bond proton-proton coupling relating dihedral angle and  $J$ . (Adapted from Campbell ID and Dwek RA (1984) *Biological Spectroscopy*. Menlo Park, CA: Benjamin/Cummings.)



**Figure 9** Energy diagram and spectrum for a two-spin coupled system showing allowed transitions and effect of  $J$  coupling. The energy scale is severely distorted as the energy gaps are in the megahertz range and the coupling energies are in hertz.

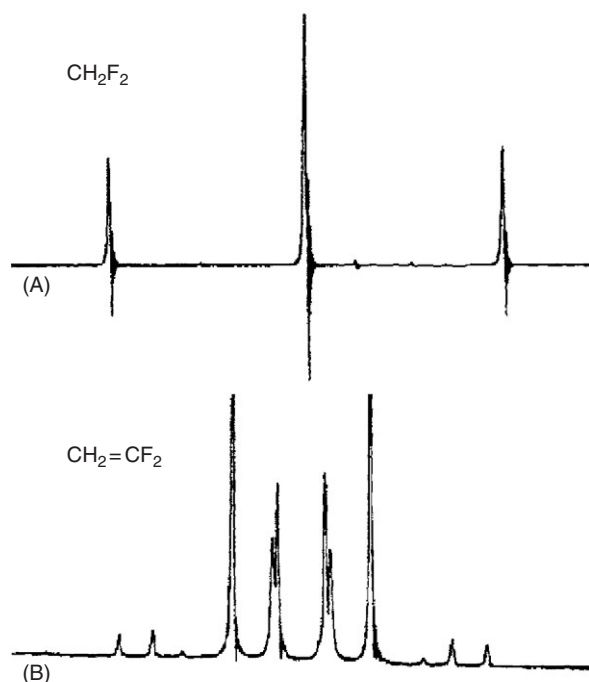
The coupling constant measures the energy of interaction between two nuclei and is independent of the applied magnetic field. When the spin dipoles are opposed, a positive coupling constant denotes that the energy is lowered, and vice versa. Positive coupling is the norm, but negative coupling is routinely found between protons in  $\text{CH}_2$  groups. For two spins, such as molecular HF, there are four possible spin combinations each with a specified energy – both up ( $\alpha\alpha$ ), both down ( $\beta\beta$ ), and two with one up and one down ( $\alpha\beta$  and  $\beta\alpha$ ) (see Figure 9).

				1		1					
				1		2		1			
			1		3		3		1		
		1		4		6		4		1	
	1		5		10		10		5	1	
	1	6		15		20		15	6	1	
	1	7	21		35		35	21	7	1	
	1	8	28	56		70		56	28	8	1
	1	9	36	84	126		126	84	36	9	1
1	10	45	120	210	252		210	120	45	10	1

**Figure 10** Pascal's triangle.

The four-line spectrum arises from the four allowed transitions in which there is a single spin flip. They have equal probability and so equal intensity. In the absence of coupling the pair of transitions for a given nucleus is of equal energy and appears as single lines. When the nuclei are coupled, the transitions differ in energy by  $J$  Hz. The splitting is the same for each nucleus. If a nucleus is coupled to  $n$  other spin  $1/2$  nuclei, there are  $2^n$  possible lines for the observed nucleus. For  $n$  nuclei coupled with equal  $J$  (e.g., the three hydrogens of a methyl group) to the observed nucleus, the peak multiplicity is  $2nI + 1$ . They have binomial intensities corresponding to the statistical probability of the spins having a given total spin quantum number – the familiar Pascal's triangle shown in Figure 10. Splitting patterns are multiplicative as shown by the  $\text{CH}_2$  group of dry ethanol which is a doublet of quartets (see Figure 1).

Spectral splitting patterns reveal the quantum character of the nuclear coupling interactions. Two energy levels of a coupled spin system may not be closer in energy than the  $J$  coupling between the nuclei. The spectral consequences of these quantum effects include distorted line intensities (haystacking), line spacings that no longer equal  $J$ , and the appearance of a finite number of extra lines and in some cases fewer lines than predicted by first-order rules. The single line spectrum observed for benzene is one such example. Even though each proton is coupled to the five other protons on the ring, no splitting is observed. Equivalent nuclei do not split each other. The proton spectra of 1,1-difluoromethane and 1,1-difluoroethene are dramatically different (see Figure 11). In  $\text{CH}_2\text{F}_2$ , the pairs of protons are indistinguishable and the fluorines are split into a simple triplet. For  $\text{CH}_2\text{CF}_2$  the protons are magnetically nonequivalent. They are distinguished by their different couplings to the *cis* and *trans* orientated fluorines. Despite the complexity of the pattern, the spectral symmetry reflects the symmetry in the



**Figure 11** 60 MHz  $^1\text{H}$  NMR spectra of (A)  $\text{CH}_2\text{F}_2$  and (B)  $\text{CH}_2=\text{CF}_2$  showing the effect of magnetic nonequivalence. The pattern is identical in the  $^{19}\text{F}$  spectrum and is unchanged at higher  $B_0$ . These spectra were obtained by sweeping the field at constant frequency as shown by the ringing distortion. (Reproduced with permission from Becker ED (1969) *High Resolution NMR*. Orlando, FL: Academic Press.)

coupled groups. Related effects are commonly observed in equivalently *ortho* and *para* disubstituted benzenes, and in  $\text{X}-\text{CH}_2\text{CH}_2$  systems. All these spectra may be computed with high accuracy in programs such as LAOCOON. The input information is the chemical shifts and coupling constants. The line positions and intensities are then computed in a quantum mechanical calculation without approximation.

The Fermi contact interaction is the major mechanism for  $J$  coupling. It is governed by the electron density at the nucleus ( $S_i$ ) and the nuclear gyromagnetic ratio according to the expression

$$J(\text{A}, \text{X}) = \left(\frac{\mu_0}{3\pi}\right)^2 \beta^2 h \gamma_{\text{A}} \gamma_{\text{X}} |S_{\text{A}}|^2 |S_{\text{X}}|^2 P_{\text{AX}}^2 / (E_{\text{T}} - E_{\text{S}}) \quad [7]$$

$P_{\text{AX}}$  is the A–X bond order and  $E_{\text{T}} - E_{\text{S}}$  the electronic excitation energy for triplet formation. The magnitude of the  $^{13}\text{C}-^1\text{H}$  one-bond  $J$  value increases with  $S$  character at carbon ( $J_{\text{C-H}}$  ethane = 125 Hz, acetylene = 210 Hz). Isotopic substitution is generally a minor perturbation on the electronic coupling and the observed  $J$  coupling is proportional to  $\gamma$ .

For example,

$$J_{\text{X-H}} = \frac{\gamma_{\text{H}}}{\gamma_{\text{D}}} J_{\text{X-D}} \quad [8]$$

Some notable exceptions to the generalization have been observed in systems with three hydrogen ligands on a common metal center where tunneling dominates the coupling surface. In its full form  $J$  is also a tensor quantity and additional orientation dependent contributions involving orbital and spin–dipolar coupling interactions have been established, especially with multielectron atoms.

In orientated samples (e.g., liquid crystals, solids), the through-space dipole–dipole coupling is combined with the  $J$  coupling. Because the dipolar coupling is often in the kilohertz range,  $J$  coupling is usually too small to detect in anything except fluid media. Residual dipole couplings measured in slightly oriented fluid medium find use in structure determination.

## Relaxation and Molecular Rotational Dynamics

Without relaxation, NMR spectra would be unobservable! NMR relaxation is a major tool for the determination of molecular dynamics on the microsecond to picosecond timescale. The nucleus is the ultimate molecular level sensor and is sensitive to motion on the scale of molecular fragments and small groups. The nuclear relaxation probe is most effective on timescales comparable to the inverse of the resonance frequency ( $1/\omega_0$ ), i.e., in the nanosecond to picosecond regime, which occurs most frequently in fluid solution. The energy absorbed by the spins from the RF excitation must be dissipated to the surroundings prior to repetition of the experiment. If not, the populations of each level will become equal (saturation) and there is no net magnetization to detect. Nuclei with  $I = 1/2$  are unaffected by electric fields and are only magnetically coupled to their surroundings. Thus, the broadening observed in electronic spectra from collisional processes is absent. As molecular magnetism is weak, spin relaxation times are usually long, ranging from milliseconds to kiloseconds. Relaxation is assumed to be a random dissipative process and follows an exponential decay with time constant  $T_1$ . Just as for spin excitation, the dissipative relaxation process requires an oscillatory magnetic field at the resonance frequency (i.e., an energy match). A change in nuclear spin state is limited to unit change in resonant spin quantum energy and requires a change in angular momentum by one unit. The nuclear spin quantum is the smallest quantum of energy. The relaxation rate  $R_1 = 1/T_1$  is specified by the expression  $R_1 = CB^2J(\omega)$  where  $C$  is a constant,  $B$  is the mean local magnetic field at the nucleus, and  $J(\omega)$  is the spectral density

**Table 4** Equations for spin–lattice relaxation rates,  $R_1$ , of nucleus A, for various mechanisms in the extreme narrowing approximation

Mechanism	$R_1$	Notes	Diagnostic	Examples
DD(intra) (homo)	$\left(\frac{\mu_o}{4\pi}\right)^2 \frac{3}{2} \gamma_A^2 h^2 \tau_c / r^6$	For a single pair of spin-1/2 nuclei of separation $r$	Overhauser effect	Protons at low conc. In deuterated solvent
DD(intra) (hetero)	$\left(\frac{\mu_o}{4\pi}\right)^2 \gamma_A^2 \gamma_X^2 h^2 \tau_c / r^6$	For a single pair of spin-1/2 nuclei AX of separation $r$	Overhauser effect	$^{13}\text{C}$ in cholesterol, most protonated carbons
DD(inter) (hetero)	$\left(\frac{\mu_o}{4\pi}\right)^2 \frac{2}{15} N_X \gamma_A^2 \gamma_X^2 h^2 / D a$	For paramagnetic relaxation of a spin-1/2 nucleus by particle X, when X has unpaired spins	Proportional to $N_X$	Relaxation of water protons by dissolved $\text{O}_2$ or $\text{Cr}(\text{AcAc})_3$
Unpaired electron (intra) (dipolar)	$\left(\frac{\mu_o}{4\pi}\right)^2 \frac{4}{3} \gamma_A^2 \gamma_e^2 h^2 S(S+1) \tau_c / r$	For relaxation by unpaired electrons of total spin $S$ at distance $r$	EPR	Most radicals: relaxation agent $\text{Gd}(\text{DPM})_3$
Spin rotation	$2I_r k T C^2 \tau_{sr} / h^2$	For isotropic molecular moment of inertia ( $I_r$ )	Proportional to $T$	Gaseous $\text{CS}_2$ : most gases
CSA	$\frac{2}{15} \gamma_A^2 B_0^2 \Delta \sigma^2 \tau_c$	For nuclei with large CSA	Proportional to $B_0^2$	Quaternary $^{13}\text{C}$ and $^{31}\text{P}$ in phosphates at high field $^{195}\text{Pt}$ in square planar complexes
Scalar coupling	$\frac{8\pi^2 J_{AX}^2 I_X(I_X+1) \tau_{sc}}{3[1 + (v_X - v_A)^2 \tau_{sc}^2]}$	Relaxation by coupling to spin, X, of quantum number $I_X$ . Requires $\gamma_A \approx \gamma_X$	Inversely proportional to $B_0^2$	$^{13}\text{C}$ in cyclohexylbromide 'Unique' to $^{13}\text{C}$ – $^{79}\text{Br}$
Quadrupole	$(eqQ/h)^2 \tau_c$	Relaxation of quadrupole by electric field gradient	$I > 1/2$	Almost all quadrupolar nuclei $^{17}\text{O}$ in compounds; halides except $^{19}\text{F}$
Raman scattering	$\frac{81\pi}{10\omega_D} (F_2)^2 \left(\frac{T}{\theta_D}\right)^2$	Relaxation of high Z spin 1/2 nuclei in solids	Proportional to $T^2$ Independent of $B_0$	$^{207}\text{Pb}$ in lead salts

AcAc, acetylacetonate; EPR, electron paramagnetic resonance; DPM, dipivaloylmethane;  $\tau_c$ , Correlation time for molecular tumbling;  $N_X$ , concentration of spins X (per unit volume);  $D$ , mutual translational self-diffusion coefficient of the molecules containing A and X;  $a$ , distance of closest approach of A and X;  $\gamma_e$ , magnetogyric ratio for the electron;  $C$ , spin–rotation interaction constant (assumed to be isotropic);  $\Delta\sigma$ , shielding anisotropy ( $\sigma_{\parallel} - \sigma_{\perp}$ );  $\omega_D$ , Debye frequency;  $\theta_D$ , the corresponding Debye temperature;  $F_2$ , spin–phonon coupling constant.

Adapted from Harris RK (1983) Nuclear Magnetic Resonance Spectroscopy. A Physicochemical View. London: Pitman Press.

function. The major local fields contributing to relaxation are shown in Table 4, together with representative examples and a method for confirming each contribution. For molecules in solution with  $I = 1/2$ ,  $T_1$  values are usually in the range 0.1–10 s. In the presence of unpaired electrons, and for quadrupolar nuclei, values of milliseconds to microseconds are found. For immobilized systems, such as solids, and spins isolated from other magnets, times of minutes to days have been measured.

The probability of finding the molecular fragment oscillating at the resonance frequency  $\omega$  is determined by  $J(\omega)$ . The oscillatory motion arises from random Brownian motion. The frequency distribution of Brownian motion is Gaussian and may cover

12 orders of magnitude for a mobile liquid:

$$J(\omega) = \int_{-\infty}^{\infty} G(\tau) e^{i\omega\tau} d\tau \quad [9]$$

with

$$G(\tau) = \langle f(t+\tau)^* f(t) \rangle \quad [10]$$

being the autocorrelation function. It measures the persistence of local fluctuations,  $f(t)$ , and is assumed to decay exponentially with a time constant  $\tau_c$  known as the correlation time. This gives

$$J(\omega) = \frac{2\tau_c}{1 + \omega^2 \tau_c^2} \quad [11]$$

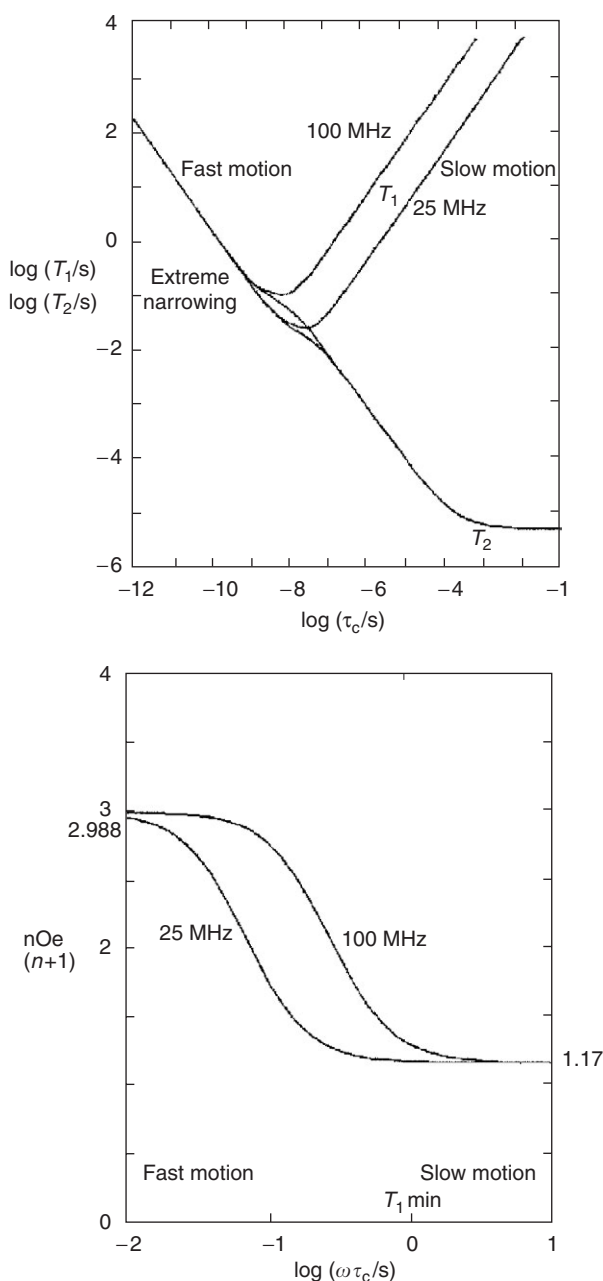


The correlation time can be interpreted as the mean time for the molecular fragment to rotate through 1 radian. Alternatively, it is the inflexion point on the Gaussian distribution function describing the motion. For moderate molecular weight, approximately spherical molecules of volume  $V$ , in a solvent of viscosity  $\eta$ , the Stokes–Einstein relationship  $\tau_c = V\eta/kT$  applies surprisingly well. For small molecules, a semiempirical correction factor may also be needed when dealing with solvents whose volume is comparable to the solute. For anisotropically rotating groups (e.g., methyl groups, side chains in polymers, and substituted phenyl groups rotating about their C2 axis) a more detailed motional analysis is required.

The relationship between the correlation time and relaxation time is shown in the log–log plot of **Figure 12**. On the short correlation time, fast motion side of the diagram where  $\omega\tau_c \ll 1$ , an increase in fluidity makes relaxation less efficient, resulting in a longer  $T_1$ . This is known as the ‘extreme narrowing’ regime and  $T_1$  is inversely proportional to  $\tau_c$ . In this regime, fast molecular motion averages dipole–dipole coupling and chemical shift anisotropy to zero, giving narrow spectral lines. On the slow motion side of the relaxation maximum,  $\omega\tau_c \gg 1$ , increased motion generates faster relaxation and  $T_1$  is proportional to  $\tau_c$ . On the slow motion side molecular systems are usually solids or high molecular mass polymers and broad lines are the rule. A treatment in terms of random Brownian motion is often too simplistic for such materials. A broad distribution of correlation times and differential segmental motion descriptions are usually needed to account for the relaxation behavior. In solids it is not uncommon to find that nuclei in all molecular fragments relax with a single correlation time, which corresponds to the motion of the mobile portion of the molecule such as a rotating methyl group. The maximum relaxation rate occurs when  $\omega\tau_c = 1$  and  $R_1 = CB^2/2\omega_0$ . Relaxation efficiency decreases with increasing resonance frequency as  $B^2$  is usually independent of field. Chemical shift anisotropy relaxation is an exception as the magnetic interaction increases as the square of the magnetic field.

NMR relaxation may be used to probe molecular motion in the kilohertz range, using the relaxation time in the rotating frame,  $T_{1\rho}$ . Here the relaxation is determined in the presence of the RF magnetic field. In this case there is a nonzero  $x$ – $y$  field given by the amplitude of the RF field ( $\gamma B_1$ ). The relaxation rate maximum ( $1/T_{1\rho}$ ) occurs at a frequency  $\omega = \gamma B_1/\pi$ .

The  $T_1$  relaxation time is best measured in an inversion–recovery pulse experiment. A  $180^\circ$  pulse is applied to the signal of interest that inverts the spin



**Figure 12** The relationship between carbon  $T_1$  and  $T_2$ , relaxation times, and the correlation time for rotational diffusion,  $\tau_c$ , for  $^{13}\text{C}$  spin being dipolar relaxed by a bonded proton. Relaxation is less efficient at higher  $B_0$  field (indicated by carbon resonance frequencies). Curves for other relaxation mechanisms may be generated by appropriate scaling. The lower curve shows the nOe in the region of the  $T_1$  minimum where  $\omega\tau_c \sim 1$  and differential cross-relaxation is effective. Curves for other dipolar pairs may be constructed by scaling the limiting nOes according to eqn [15].

population.  $z$ -Magnetization is perturbed as far as possible from equilibrium. A composite pulse is recommended to ensure full inversion and that there is no residual  $x$ – $y$  magnetization. A variable time delay

is introduced during which the  $z$  magnetization relaxes back toward equilibrium. A  $90^\circ$  pulse is then used to sample the residual  $z$  magnetization. The  $T_1$  is determined by fitting the recovery curve

$$M(t) = M(\infty)[1 - 2 \exp(-t/T_1)] \quad [12]$$

where  $t$  is the variable time delay and  $M(t)$  is the magnetization at time  $t$ . It is necessary to wait at least  $5 \times T_1$  between  $180^\circ$  pulses to ensure that spin equilibrium is restored and  $M(\infty)$  is at its maximum value. This is a general requirement for repetition of experiments. It applies to any experiment that requires equilibrium spin populations initially.

## Dynamics, Lineshapes, and Homogeneity

Frequency and time are inversely correlated in any NMR experiment. In order to observe an interaction in the frequency spectrum, the nuclei must experience that interaction for a time that is long with respect to the inverse of the interaction frequency. Fast processes result in frequency averaging. Examples include isotropic molecular tumbling in solution which averages dipolar coupling to zero with the resulting narrow lines; magic angle spinning in solid-state NMR to remove chemical shift anisotropy; and molecular dynamics which provides conformationally averaged coupling constants.

A solution NMR peak commonly has a Lorentzian shape. A Lorentzian line in the frequency domain results from a first-order exponential decay process with a time constant,  $T_2$ . It is specified by the lineshape function,  $f(\nu)$ , where

$$f(\nu) = \frac{2T_2}{1 + 4\pi^2 T_2^2 (\nu_i - \nu)^2} \quad [13]$$

Here  $T_2$  (s) is the spin-spin relaxation time.  $T_2$  can be estimated from the full linewidth at half-height  $\Delta\nu = 1/\pi T_2$ . For accurate work  $T_2$  is determined in a spin echo experiment, which cancels linewidth contributions from magnetic field inhomogeneities. In a highly homogeneous magnet the linewidths of small molecules, such as TMS, may be as narrow as 0.1 Hz.

The linewidth specification for commercial solution spectrometers is commonly  $\leq 0.3$  Hz across a 5 mm sample, independent of absolute magnetic field. Note that this requires a magnetic field that is stable (see eqn [1]), and uniform (homogeneous) to better than 1 part in  $10^9$  across the sample – a remarkable engineering feat. The sharpest line in a spectrum provides a measure of the actual instrument homogeneity. In practice, ideal Lorentzian lineshapes are

common, while Gaussian and other more complex shapes are also found in poorly shimmed (inhomogeneous) magnets. Deviation from magnet ideality contributes to the shape of all peaks in the spectrum.

Magnetic interactions also shorten spin lifetimes and contribute to line broadening. Paramagnetic impurities such as ferromagnetic particulates, dissolved oxygen, and transition metals are a common source of broadening. For large molecules – greater than 10 kDa – and in highly viscous media dipolar interactions between neighboring nuclei cause broadening when the slow tumbling in solution is insufficient to average the dipolar interaction.

## Chemical Exchange and NMR Timescales

The intrinsic linewidth of a peak measures the lifetime of a spin in a given configuration. Lifetimes may depend on chemical processes such as the proton exchange between water and the ethanol OH shown in Figure 1. In such cases, an order of magnitude estimate of the kinetics of the exchange process can be made using the uncertainty principle

$$\Delta E \Delta t \sim h/2\pi \quad \text{or} \quad \Delta\nu \Delta t \sim 1 \quad [14]$$

For the OH exchange,  $\Delta\nu_{1/2}$  for water was measured to be 25 Hz and the exchange rate is  $\sim 75 \text{ s}^{-1}$  for this OH group. The hydroxyl proton line of ethanol is significantly narrower. This is because the lifetime of the proton bound to alcohol is longer than its lifetime when bound to water, as required for chemical equilibrium. The exchange lifetimes are directly proportional to the population at each site. If the rate is very much faster than the chemical shift difference in hertz, then a single sharp line is observed at the population-weighted average shift. Accurate kinetics may be determined from exact lineshape calculations. NMR has a unique ability to reveal ‘invisible’ chemically degenerate processes. Examples include proton exchange rates in water, bond rotations, and ligand interchanges in complexes. Intramolecular processes may be distinguished from intermolecular processes by the presence of averaged coupling constants. Chemical kinetics may also be measured by saturation transfer. Spin population information is transferred from site to site by chemical exchange. Saturation at one site will appear at a new site provided exchange is faster than  $T_1$  relaxation.

## Diffusion

Molecular diffusion (and flow) may be measured in the presence of magnetic field gradients using the Stejskal–Tanner spin-echo technique. Diffusion in a gradient dephases the magnetization between pulses. The technique may be incorporated into

multidimensional pulse sequences, e.g., DOSY, to discriminate signals in mixtures on the basis of diffusion coefficient/molecular size. Combined nOe and diffusion experiments provide information on binding small molecules to large molecules, which finds application in drug discovery.

### Sample Spinning, Field/Frequency Lock, and Field Gradients

Sample spinning about the  $z$ -axis averages the static  $x$ - $y$  magnet inhomogeneities. In poorly shimmed magnets, spinning side-bands are found separated from the main peak at multiples of the spinning frequency. Ideal lineshapes may be recovered by computer deconvolution to remove the effects of magnetic inhomogeneity if an ideal reference line is available. Most spectrometers also employ a field/frequency lock to overcome effects of drifting fields. A signal from a reference – commonly deuterium in a deuterated solvent – is detected in dispersion mode. Magnet drift produces a signal which is used in a feedback loop to compensate for the drifting field.

Pulsed field gradients are used to shorten experimental times in multidimensional NMR by removing the  $T_2$  limit on pulse repetition times. As spin frequency is proportional to the applied field, field gradients are used to discriminate between multiple quantum spin orders in multipulse experiments, e.g., DQ filtered COSY experiments that discriminate against the strong diagonal signals arising from single quantum excitation.

### Double Resonance and Decoupling

Coupled spectra contain a wealth of bonding information. However, in many situations, the information content is more readily interpreted if this spectral complexity is removed by decoupling. Decoupling is achieved by applying a second RF field at the resonance of the nucleus whose coupling is to be removed. The decoupling may be selective, involving a single group or even a single line within a group, or may be broadbanded and cover the full spectrum. The amplitude of the decoupler field must be larger than the coupling constant to the nucleus to be decoupled. If lower fields are used, or the frequency is offset, additional lines and line broadening may be observed (see **Figure 13**). A low-power RF pulse applied selectively may be used for specific inversion of populations.

Broadband proton decoupling is used routinely when carbon spectra are obtained. For complete decoupling, the observed spin and the decoupled spin must be orthogonal.  $J$  coupling is a scalar, and the

coupling contribution to the energy is determined by the dot product of the spin angular momenta  $\mathbf{J} \cdot \mathbf{I} \cdot \mathbf{S}$ . Alternatively, but not equivalently, spins may be decoupled by randomizing their orientation relative to one another in a time which is fast with respect to  $1/J$ . Many processes randomize spin orientation: rapid  $T_1$  relaxation which decouples quadrupolar nuclei, chemical exchange which breaks bonds, and by noise decoupling in which the frequency of the irradiation is rapidly and randomly varied around the resonance frequencies to average  $\mathbf{J} \cdot \mathbf{I} \cdot \mathbf{S}$  to zero. More sophisticated multiple pulse versions for broadband decoupling exist, which minimize the power levels required (e.g., WALTZ, GARP, MLEV 17, etc.).

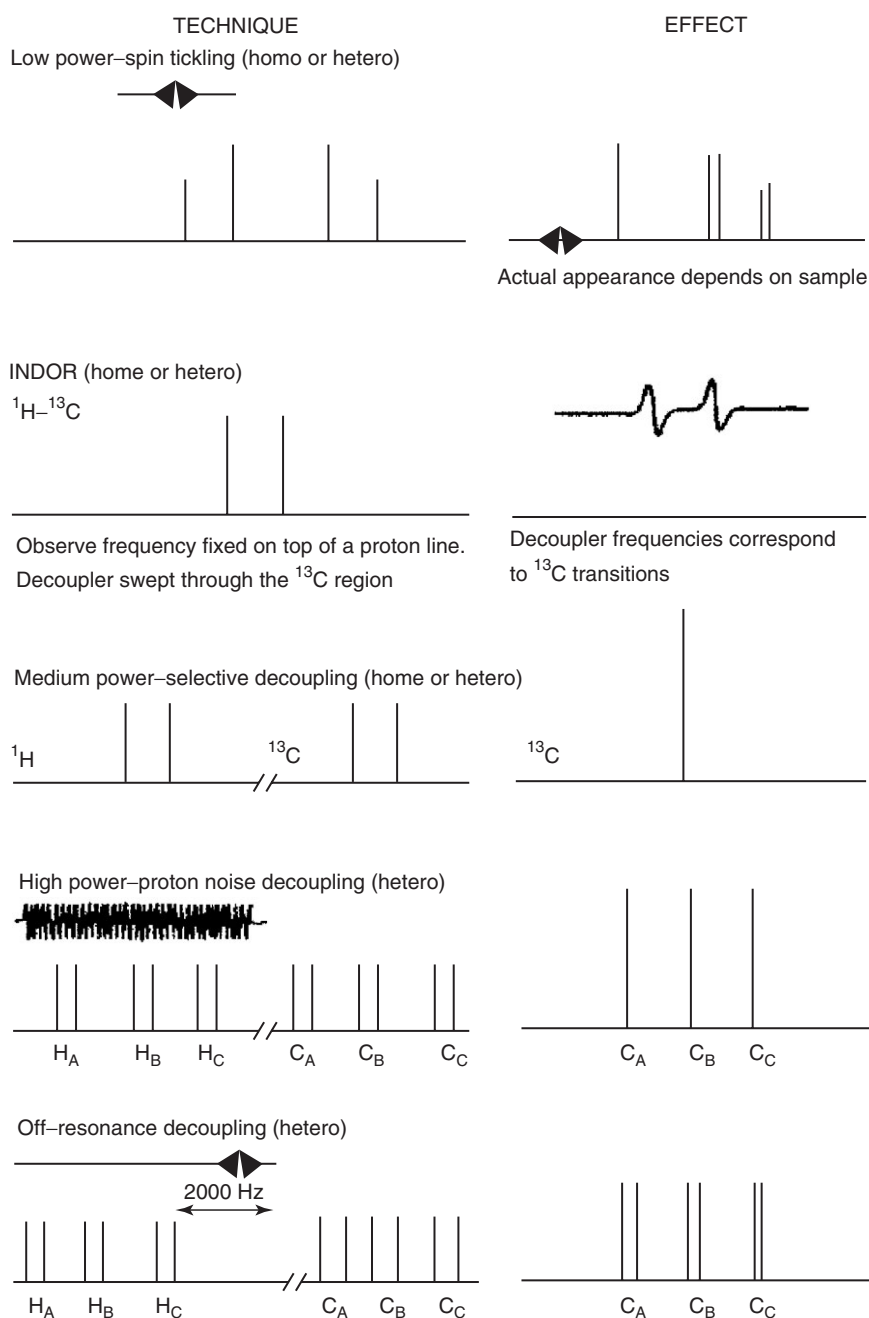
Since an additional oscillating magnetic field is present while the decoupler is on, the irradiated spins have new coherent motions and new energy states. The decoupler is also an energy source and so spin populations are perturbed. While the new energy states and line positions are generated immediately, spin population changes still require times comparable to  $T_1$  and  $T_{1\rho}$ . The most important population changes occur in dipolar coupled systems and give rise to the nuclear Overhauser effect (nOe). Pulse sequences have been developed to exploit the time differentials to obtain decoupled spectra without nOe for spin counting or coupled spectra with nOe to enhance sensitivity.

### Nuclear Overhauser Effect

The dipolar relaxation mechanism is unique because the dipolar coupling interaction contains two spin terms involving mutual spin flips – zero quantum  $\uparrow\downarrow \rightleftharpoons \downarrow\uparrow$  and double quantum  $\uparrow\uparrow \rightleftharpoons \downarrow\downarrow$  transitions. All other relaxation processes are limited to single spin interchange. While the decoupler is on, the spin population differences between the levels irradiated are equalized as the rate of energy input from the decoupler greatly exceeds the outflow by relaxation. In dipolar coupled systems, the availability of double and zero quantum relaxation pathways (cross-relaxation) produces non-Boltzmann spin populations in the energy levels of the observed nucleus. These perturbed populations are measured as the nuclear Overhauser effect (nOe) effect ( $\eta$ ). The relationship is

$$\eta = \frac{R_c}{R_1 + R_x} \frac{\gamma_S}{\gamma_I} = \frac{M_{\text{irr}}}{M_0} - 1 \quad [15]$$

where  $R_c$  is the cross-relaxation rate between spins  $I$  and  $S$ .  $I$  is the observed and  $S$  the irradiated nucleus.  $R_1$  is the dipolar relaxation rate and  $R_x$  is the relaxation rate for all other relaxations (**Table 4**).  $M_{\text{irr}}$  and  $M_0$  are the integrated intensity of  $I$  in the presence



**Figure 13** Summary of double resonance techniques and their influence on the spectrum. (Reproduced with permission from Abraham RJ, Fisher J, and Loftus P (1988) *Introduction to NMR Spectroscopy*. Chichester: Wiley; © Wiley.)

and absence of irradiation of  $S$ . For extreme narrowing when zero, single, and double quantum dipolar relaxations contribute,  $R_c/R_1 = 1/2$ . In the slow motion limit when only zero quantum terms apply,  $R_c/R_1 = -1$ . The  $n\text{Oe}$  dependence on correlation time was illustrated in Figure 12. The detection of an  $n\text{Oe}$  is diagnostic of dipolar relaxation and usually implies that nuclei are within 0.5 nm of each other. The relative rate of alternative relaxation processes sets the distance cutoff. If differential  $n\text{Oe}$  values or

the rate of change of  $n\text{Oe}$  can be measured, relative distance information may be obtained. A full relaxation analysis is necessary for accurate distance measurement.

### Polarization Transfer

The ability to transfer population information via dipolar coupling and double resonance is particularly useful for enhancing the detectability of insensitive

nuclei. The low-sensitivity nuclear resonance is irradiated, the information transferred via the coupling, and the sensitive nucleus is detected. Over 10% of the total spins in gaseous helium and xenon can be polarized through dipole coupling with rubidium atoms that have been optically pumped with a laser. The spin relaxation time for rare gases at low pressure can be of the order of hours. This enormous signal enhancement and the great sensitivity of xenon chemical shifts to local environment has been used to probe surfaces and also for MRI. Remarkable polarization transfers approaching 50% have been achieved by microwave irradiation of organic radicals. This technique allowed detection of  $^{13}\text{C}$  labeled urea by MRI.

## Quantitative Analysis

At first sight, quantitative analysis by NMR might be expected to be highly accurate and broadly applicable since the 'extinction coefficient' for NMR is unity. In practice, it is found that the best accuracy achievable in extensive studies of proton NMR of known mixtures is  $\pm 1\%$ . Results at this level of accuracy require very careful attention to the selection of the experimental parameters such as pulse width, pulse repetition time, and nOe effects and the availability of calibration mixtures. For  $^{13}\text{C}$  analysis the accuracy is commonly poorer,  $\sim 1\text{--}5\%$ , because of the reduced S/N ratio and the variation of nOe contributions with molecular structure and field strength. Errors of  $\pm 10\%$  or more are not uncommon in quantitative analyses especially when conducted by inexperienced operators. In proton NMR, linewidth variation is an underappreciated issue. Careful control of experimental parameters, and the use of calibration mixtures are minimum requirements for accurate analyses and accuracies of  $\pm 0.5\%$  have been achieved by experts.

A fundamental characteristic of Lorentzian lines is their slow return to baseline as a function of frequency offset from resonance. Thus, broad lines such as the OH peak in ethanol, and especially intense solvent peaks, contribute intensity throughout a spectrum. The decay of the signal to the baseline is linearly dependent on the linewidth at half-height. For example, to obtain an integral corresponding to 99.0% of the total peak intensity of a Lorentzian line, one must integrate over 63.6 linewidths. This value is obtained from  $f(v)$  – the Lorentzian function – eqn [13]:

$$\text{Int} = K \int_{-v}^v f(v) dv \quad [16]$$

where  $\pm v$  are the integration limits and  $K$  is the instrument response factor. If the frequency scale is normalized in units of the full-peak linewidth at half-height, the integral takes the form

$$\text{Int} = 2K \int_0^{x_L} f'(x) dx = 2(K/\pi) \arctan(x_L) \quad [17]$$

where  $x_L$  is the integration cutoff in units of linewidths. Precise integrals require that the cutoff limit be equal for every peak.

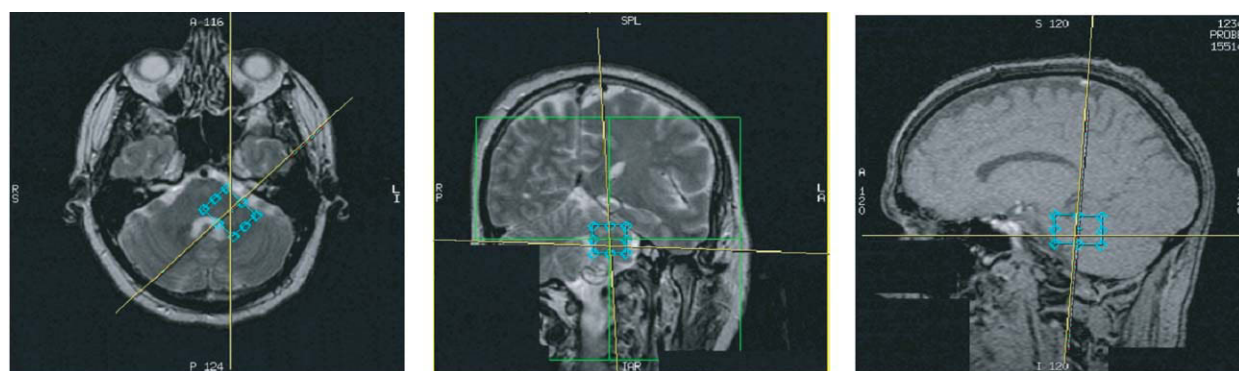
Further, because of the variation of long range  $^{13}\text{C}$ –H couplings and multiplicities, the extent of incorporation of satellite contributions within the linewidth cutoff ranges will vary from group to group. Solvents often contain trace amounts of impurities, especially water, and can cause overlap problems, as can their isotopic satellites.

## Magnetic Resonance Imaging, Field Gradients, and Diffusion

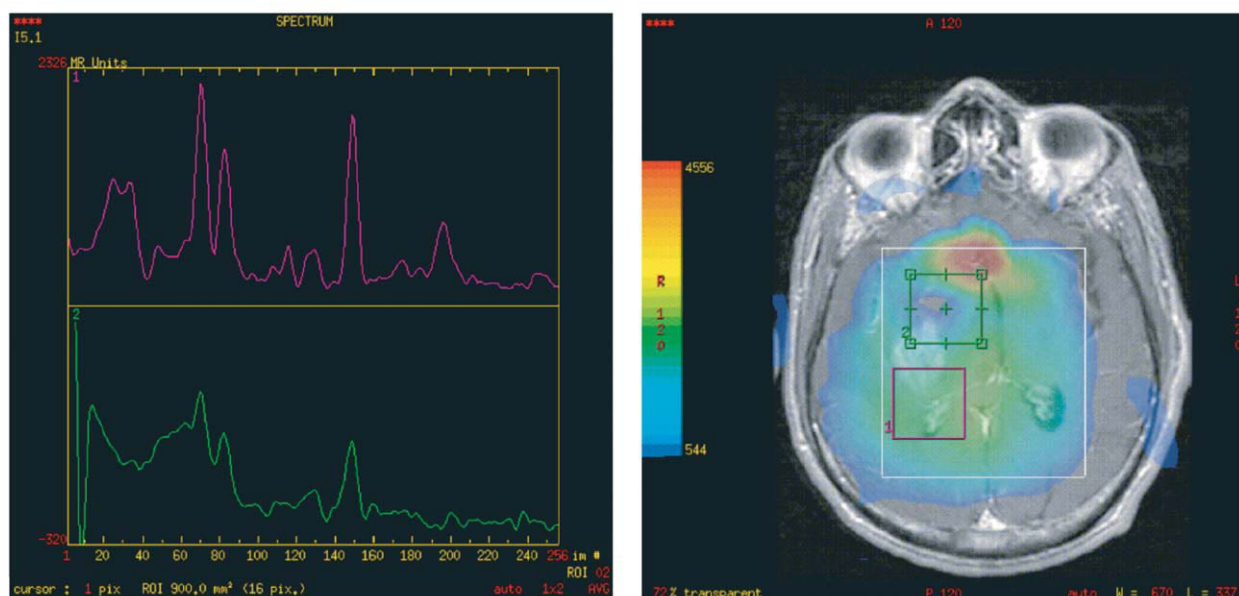
Application of magnetic field gradients adds spatial dimensions to the standard NMR experiment. The spatial location of spins – often the protons of water – is determined by making  $B_0$  spatially dependent across the sample with additional defined field gradients. The local field, which varies from point to point, determines the resonance frequency. This is the basis of MRI. Field gradients may be applied in three dimensions and generate sliced 3D images of objects such as the brain (Figure 14) and physical processes such as tablet dissolution (Figure 15). Data processing allows slicing through any desired plane. In early experiments static field gradients were used. In the simplest experiment the object was orientated at specific angles with respect to the field and the image intensity was projected onto the gradient axis. The image was reconstructed by a back projection computation. More recently pulsed field gradients are used and the image encoded in a frequency- and phase-modulated 2D array with intensity providing the third dimension. It is now possible to obtain useable images in a few seconds with such methods.

The resolution in an image is determined by the number of data points per dimension. The volume elements are called voxels. Their number is usually limited by the total acquisition time and the S/N ratio desired. The intrinsic limit for imaging of living systems is the water diffusion distance  $\sim 50\text{ }\mu\text{m}$ . In special situations  $5\text{ }\mu\text{m}$  resolution has been achieved. Practical images of flow, including blood flow, have been obtained. Polarized  $^3\text{He}$  provides images of breathing function in lungs.



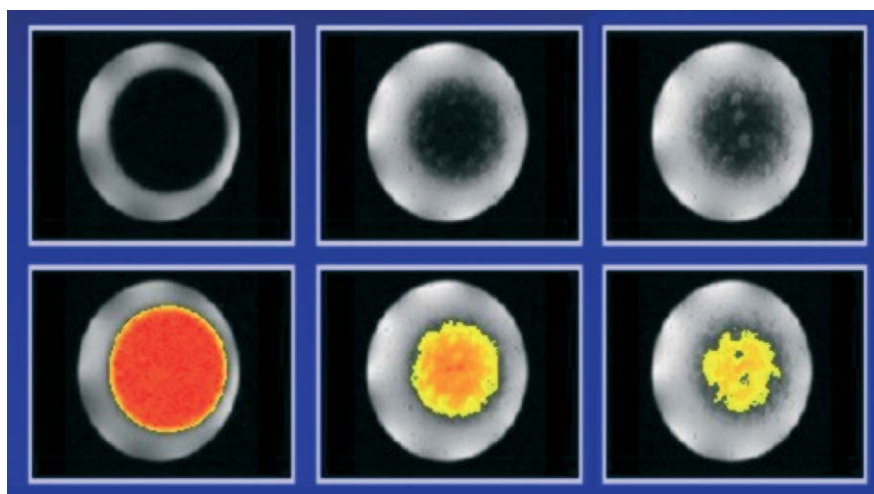


(A)



(B)

**Figure 14** (A) Planning images for magnetic resonance spectroscopy (MRS) of the brain: top and middle: spin-echo image TE = 96 ms, TR = 3 s. Bottom gradient echo image TE = 1.9 ms, TR = 116 ms. (B) Chemical shift metabolic image (CSI) of the brain overlaid on the anatomic image. Spectra are displayed for the selected regions. Chemical shift range is from 4.3 to 0.5 ppm. (Spectra were obtained by Dr JA Hopkins, GE Medical Systems and reproduced by permission of GE Medical Systems, Milwaukee, WI.)



**Figure 15** NMR microscopy: dissolving of a tablet in 0.1 mol l<sup>-1</sup> HCl solution. (Spectra were reproduced by permission of Bruker Biospin.)

Magnetic field gradients may also be generated with RF in homogeneous static fields. Surface coils are tuned to resonate spins at specified depths within a material. As the spins are in a homogeneous field, it is possible to resolve chemical shift information as a function of distance. Metabolic studies *in vivo*, especially of phosphate materials, provide unique information. Clinical MRI studies of disease states are becoming practical. Functional MRI is used to identify sites of neurological activity in the brain.

**See also:** Nuclear Magnetic Resonance Spectroscopy: Overview; Instrumentation. **Nuclear Magnetic Resonance Spectroscopy-Applicable Elements:** Hydrogen Isotopes; Carbon-13; Fluorine-19; Nitrogen-15; Phosphorus-31. **Nuclear Magnetic Resonance Spectroscopy Applications:** Proton NMR in Biological Objects Subjected to Magic Angle Spinning. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Nuclear Overhauser Effect; Multidimensional Proton; Solid-State.

## Further Reading

- Abragam A (1963) *The Principles of Nuclear Magnetism*. Oxford: Clarendon.
- Akitt JW (1992) *NMR and Chemistry*. London: Chapman and Hall.
- Braun S, Kalinowski H-O, and Berger S (1998) *150 and More Basic NMR Experiments*. Weinheim: Wiley-VCH.
- Brevard C and Granger P (1981) *Handbook of High Resolution Multinuclear NMR*. New York: Wiley.
- Callaghan PT (1991) *Principles of Nuclear Magnetic Resonance Microscopy*. Oxford: Clarendon Press.
- Carrington A and McLachlan AD (1967) *Introduction to Magnetic Resonance*. New York: Harper & Row.
- Chujo R, Hatada K, Kitamura R, *et al.* (1987) NMR measurement of identical polymer samples by round-Robin method. 1. Reliability of Chemical-shift and signal intensity measurements. *Polymer Journal* 19: 413–424.
- Derome AE (1987) *Modern NMR Techniques for Chemistry Research*. Oxford: Pergamon.
- Freeman R (1988) *A Handbook of NMR*. New York: Wiley.
- Frydman L, Lupulescu A, and Scherf T (2003) Principles and features of single-scan two-dimensional NMR spectroscopy. *Journal of the American Chemical Society* 125: 9204–9217.
- Fukushima E (1989). *NMR in Biomedicine*. American Institute of Physics, New York.
- Fukushima E and Roeder SBW (1981) *Experimental Pulse NMR*. Reading, MA: Addison-Wesley.
- Fyfe CA (1983) *Solid State NMR for Chemists*. Guelph, Ontario: CFC Press.
- Grant DM and Harris RK (eds.) (1996–2002) *Encyclopedia of Nuclear Magnetic Resonance*. New York: Wiley.
- Hore PJ (1995) *Nuclear Magnetic Resonance*. Oxford: Oxford University Press.
- Jackman LM and Sternhell S (1969) *Applications of Nuclear Magnetic Resonance Spectroscopy in Organic Chemistry*. Oxford: Pergamon.
- Johnson CS Jr. (1999) Diffusion ordered nuclear magnetic resonance spectroscopy: Principles and applications. *Progress in NMR Spectroscopy* 34: 203.
- Mason J (1987) *Multinuclear NMR*. New York: Plenum Press.
- MacNaughtan MA, Hou T, Xu J, and Raftery D (2003) High-throughput nuclear magnetic resonance analysis using a multiple coil flow probe. *Analytical Chemistry* 75: 5116–5123.
- Oslo D, Norcross J, O'Neil-Jounson M, *et al.* (2004) Microflow NMR: Concepts and capabilities. *Anal. Chem.* 76: 2966–2974.
- Popov AI and Hallenga K (1991) *Modern NMR Techniques and their Application in Chemistry*. New York: Dekker.
- Randall JC (1977) *Polymer Sequence Determination*. New York: Academic Press.
- Singleton D and Thomas A (1995) High-precision simultaneous determination of multiple small kinetic isotope effects at natural abundance. *Journal of the American Chemical Society* 117: 9357–9358.
- Slichter CP (1990) *Principles of Magnetic Resonance*, 3rd edn. Berlin: Springer.
- Van de Ven FJM (1995) *Multidimensional NMR in Liquids: Basic Principles and Experimental Methods*. New York: VCH.

## Glossary

### Anisotropy

Unequally distributed in space, a common situation for electrons in bonds where symmetry is rare. Anisotropy may be spatial, magnetic, or molecular in origin. Magnetic anisotropy is created by anisotropic electron motion induced by the applied magnetic field in groups such as phenyl. The anisotropic magnetic effects generate differential chemical shifts parallel and perpendicular to  $B_0$  for groups in fixed locations close to the anisotropic group.

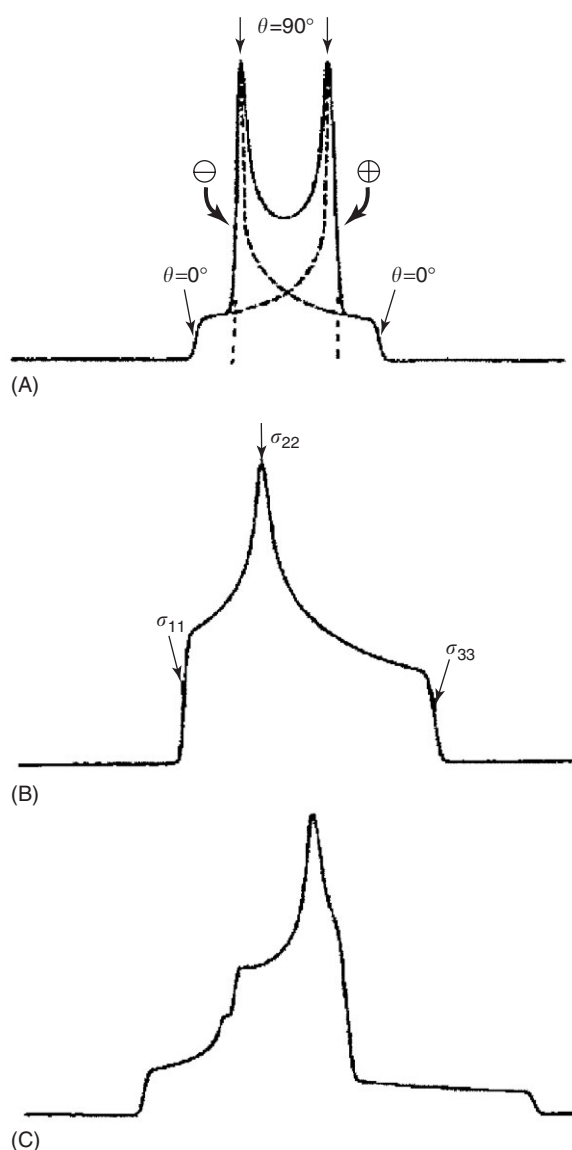
### $B_0$ , $B_1$ , $B_2$

The magnetic fields used in the nuclear magnetic resonance experiment quantified by their strengths.  $B_0$  is the large static field generated by the instrument magnet. Its direction defines the  $z$ -axis.  $B_1$  is the RF electromagnetic field oscillating close to the resonance frequency of the observed nucleus. The strength of  $B_1$  is determined by its amplitude. Its direction is usually perpendicular to  $B_0$ .  $B_1$  is applied as an oscillating current in a tuned coil.  $B_2$  is similar to  $B_1$  but is tuned to the nucleus to be decoupled and is often of higher amplitude.

### Bloch decay

This is also known as the free-induction decay (FID). The summed nuclear response detected by the receiver as the magnetization induced by an excitation

	<p>pulse decays to steady state with time constant <math>T_2^*</math>. The response is digitized and Fourier transformed to produce the observed spectrum.</p>
Chemical shift	<p>The position of a nuclear resonance in the spectrum. The chemical shift is specified on the <math>\delta</math> scale relative to a reference resonance. The chemical shift is characteristic of the local electronic environment of the nucleus and is used to identify functional groups.</p>
Chemical shift anisotropy (CSA)	<p>Chemical shift is a tensor quantity and depends on the molecular orientation with respect to the applied magnetic field. Different electron circulations induced by the applied field generate magnetic fields that differentially shield the nucleus depending on molecular orientation. The CSA is the difference between the extreme values of the chemical shift observable in solids but averaged to zero in solution (see Figure 16).</p>
Coalescence	<p>The point at which two separate peaks merge to a single unresolved peak. Coalescence usually refers to a system undergoing chemical exchange and exhibiting exchange-line broadening.</p>
Composite pulse	<p>A sequence of RF pulses designed to generate a desired spectral response. The individual pulses within a composite pulse may vary in time, frequency, and orientation to achieve the desired response. Composite pulses are commonly used to invert spins over a broad frequency range where a single pulse gives imperfect inversion far from resonance. Composite pulses are also used for excitation of selected frequencies.</p>
Coherence	<p>Processes are said to be coherent when they maintain a fixed-time or phase relationship with respect to each other. The Larmor precession frequencies are brought to coherence by an RF pulse, for example, generating <math>x</math>-<math>y</math> magnetization. Phase coherence is used in pulsed field gradient experiments to selectively detect different quantum spin states.</p>
Continuous wave (CW)	<p>NMR spectroscopy performed with time-independent fixed-frequency excitation and/or detection. From 1950 to 1975, most spectrometers operated in CW mode with nuclei being brought to the fixed resonance frequency by sweeping the <math>B_0</math> magnetic field. Fourier</p>



**Figure 16** Spectral lineshapes for powdered solids. (A) Peak doublet produced by dipole coupling between two spin  $1/2$  nuclei. The doublet is composed of two parts (shown dotted). They correspond to the observed proton flip occurring when its neighbor is spin up (left) or spin down (right). The indicated turning points correspond to the angle between the internuclear vector and  $B_0$ . (B) Chemical shift anisotropy pattern with shielding tensor components  $\sigma_{11}$ ,  $\sigma_{22}$ , and  $\sigma_{33}$ . (C) Combined DD and CSA spectrum. Note that this is not simply (A) + (B). (Reproduced with permission from Power WP and Wasylishen RE (1991) In: Webb GA (ed.) *Annual Reports in NMR Spectroscopy*, vol. 23, p. 17. London: Academic Press.)

transform spectrometers by contrast operate at fixed field and use broadband frequency excitation and detection. Continuous wave may also refer to fixed-frequency continuous decoupling experiments.

**Correlation time ( $\tau_c$ )** The mean time for a molecule or molecular fragment to rotationally diffuse

	through one radian. May also refer to the inverse of the most probable angular frequency for rotational motion.		
Coupling constant	Nuclei are coupled when their energies depend on their mutual spin orientations. The magnitude of the interaction is the coupling constant, which is independent of magnetic field. Coupling may be through bond ( <i>J</i> coupling), or through space (dipole coupling). Dipole coupling is inversely proportional to the third power of the internuclear distance and depends on the orientation of the spin pair to the magnetic-field direction. <i>J</i> coupling is transmitted by the bonding electrons and attenuates with the number of intervening bonds. <i>J</i> coupling confirms nuclear connectivity.	Dipole–dipole coupling	The dipole–dipole coupling between two nuclei is given by: $D_{AB} = (\mu_0/4\pi)\gamma_A\gamma_B\hbar(3\cos^2\theta_{AB} - 1)/r_{AB}^3 \cdot 2\pi$ where A is the observed nucleus, B is neighboring nuclear dipole, $\theta_{AB}$ is the angle the A–B internuclear vector makes with the applied field direction, and $r_{AB}$ is the through-space internuclear distance. The two possible spin states of one nucleus differ in energy by the coupling to the magnetic field of the second spin dipole (see Figure 16).
		Extreme narrowing	The fast motion regime where molecular motion is sufficiently rapid that $\omega_0\tau_c \ll 1$ . Under these conditions, chemical shift anisotropy and dipole–dipole interactions are averaged to zero.
Cross polarization with magic angle spinning (CPMAS)	A combination of techniques to narrow signals and facilitate observation of high-resolution spectra of solids. Cross-polarization enhances the sensitivity of low abundance and low $\gamma$ -nuclei by transferring polarization from high abundance, high $\gamma$ -nuclei such as protons. See also magic angle spinning.	Exchange broadening	When two resonance frequencies interchange on a timescale that is of the order of the inverse of the frequency difference between them, exchange broadening occurs. When the exchange rate matches the frequency difference, the linewidth broadens to a maximum comparable to the frequency difference. Exchange may involve change in chemical shift and/or coupling constant, bond breaking, conformational change, etc.
Cross polarization	The exchange of magnetic polarization from one nuclear type to another, e.g., $^1\text{H}$ to $^{13}\text{C}$ . Exchange requires an energy match and a coupling interaction for polarization to be transferred. Usually employed to increase the sensitivity of low-frequency low-abundance nuclei using the polarization generated by high-frequency, high-abundance nuclei.	Free induction decay	See Bloch decay.
		Fermi contact	The major magnetic interaction between the electron and the nucleus. It is the major contribution to hyperfine coupling and is maximized for S electrons because of their high spin density at the nucleus.
Decoupling and double resonance (DD)	The use of high-power resonant RF to collapse coupling interactions to zero. Versions include homonuclear (observed and decoupled nuclei have the same $\gamma$ ), heteronuclear (observed and decoupled nuclei with different $\gamma$ ), and noise (for wide frequency coverage). With low-power selective irradiation it is possible to map coupled-energy diagrams and transfer polarization.	Field gradient	A magnetic field applied to make the resonance frequencies depend – usually linearly – on the location of the spins in the sample. Specifically oriented coils in the probe generate fields ( $\sim 100\text{ mTm}^{-1}$ ) in the <i>x</i> , <i>y</i> , and <i>z</i> directions. More complex gradients can be created by appropriate coil design (e.g., shim coils). Pulsed field gradients are used to measure diffusion and to select multiple quantum coherences. See also MRI.
Diamagnetic	Repelled from a magnetic field – opposite of paramagnetic; raised in energy in the presence of a magnetic field. Most molecules with all electrons paired are diamagnetic. Diamagnetic shielding contributions reduce the static magnetic field at the nucleus causing it to resonate at lower frequency.	Fourier transform (FT)	The mathematical operation for interconverting two related variables such as time and frequency. Fourier showed that any time-dependent function could be represented equivalently by the sum of a
Dipole	A nucleus with $I = 1/2$ acts as a magnetic dipole having a north and south pole. Each dipole is associated with a local		

	series of frequency functions such as sin and cos waves. The process of Fourier transformation converts the digitized Bloch decay into a frequency spectrum.		
Hahn spin echo	The recovery of signal following a $90^\circ$ – $t$ – $180^\circ$ – $t$ – pulse sequence. For example, a FID decays rapidly in an inhomogeneous field because different parts of the sample resonate at different frequencies. When a $180^\circ$ pulse is applied, at time $t$ after the excitation pulse, the signal reappears after an additional time $t$ – the Hahn echo. Alternatively, if the FID decays because of spin exchange processes ( $T_2$ ) no echo will occur.	Lock or field/frequency lock	A method for accurately maintaining (locking) the static magnetic field at a fixed value. The frequency of a nuclear magnetic resonance signal – commonly deuterium from a deuterated solvent – is held at constant value by a field/frequency feedback loop to compensate for any magnetic field drift during an experiment.
Hartmann–Hahn match	A method for allowing energy exchange between dipolar coupled nuclei with different $\gamma$ . No exchange is possible in the static field alone as the spin quanta differ in energy at common $B_0$ fields. However, the precession frequencies can be made equal about the $B_1$ RF field. The match requires that amplitudes be adjusted so that $\gamma_A B_{1A} = \gamma_X B_{1X}$ . Population and energy transfer is now possible.	Magic angle spinning (MAS)	The process of spinning a sample, usually solid, rapidly about an axis set at $54.7^\circ$ ( $\arccos 1/\sqrt{3}$ ) with respect to the static field. The magic angle corresponds to the body diagonal of a cube. Spinning about this axis averages over the three spatial dimensions equally. MAS is used to average CSA.
Homospoil	The application of a brief field gradient pulse to the shim coils to vary $B_0$ across the sample. This destroys any residual $x$ – $y$ magnetization because the magnet homogeneity has been spoiled and frequencies dispersed proportionally.	Magnetic moment ( $\mu$ , $M$ )	Each nuclear spin has a magnetic moment, $\mu$ quantified in units of its spin angular momentum. Values of $\mu$ in units of the Bohr magneton, $\beta$ , are recorded according to different conventions. For protons, the values are 5.5854 with $\mu = \gamma\hbar$ ; 2.7927 with $\mu_z = \gamma\hbar I$ for the $z$ component of spin $I$ , or 4.8372 with $\mu = \gamma\hbar\sqrt{I(I+1)}$ where the total magnetic moment is the scaling factor. The observable component $p$ is given by $m_I = \mu/\sqrt{I(I+1)}$ where $m_I$ is the magnetic quantum number of an individual spin which has allowed values $-I, -I+1, \dots, I-1, I$ . The vector sum of the individual spin moments gives the bulk observable magnetic moment which is denoted by $M$ .
Knight shift	The large chemical shifts observed in the nuclear magnetic resonance of molecules with unpaired electrons – especially metals. It arises because the magnetic environment of the nucleus is dominated by the electron paramagnetism since $\gamma_e \gg \gamma_N$ . Shifts to high or low frequency are observed depending whether the $\alpha$ or $\beta$ electron spin state is low energy and hence more populated.	Magnetogyric ratio ( $\gamma$ )	The proportionality constant, $\gamma$ , between the resonance frequency of a particular nucleus and the magnetic field strength.
Larmor frequency	The precession frequency of the nuclear spin dipole when torqued by a magnetic field. The Larmor frequency is equal to the nuclear resonance frequency.	Magnetic resonance imaging (MRI)	The use of nuclear magnetic resonance to measure the spatial distribution of nuclei, particularly water protons. Magnetic field gradients generate spatial dependence of the resonance frequency. Three-dimensional images are constructed from the signals.
Lineshape, linewidth	A nuclear magnetic resonance line is usually found to have one of two ideal lineshapes – Gaussian, or more often, Lorentzian. A Gaussian line is found when there is a random distribution of static fields within the sample. A Lorentzian line by contrast arises because the spin lifetime follows a first-order decay law. Weighting functions can be applied to a free-induction decay to generate	Nuclear magnetic resonance (NMR) timescale	The NMR timescale most commonly refers to spin lifetimes of milliseconds to seconds and chemical exchange kinetic processes. The actual time is in the range $0.1/\Delta\omega - 10/\Delta\omega$ where $\Delta\omega$ is the frequency difference in $\text{s}^{-1}$ between exchanging signals. The NMR timescale



	may also refer to the nanosecond regime when relaxation processes are being discussed. This timescale relates to the correlation time, $\tau_c$ , for molecular motion involved in relaxation.		
Nuclear Overhauser effect (nOe)	The change in signal intensity of a nucleus when a transition of an adjacent spin is irradiated. The observed and irradiated spins must be dipole coupled to one another. The observation of an nOe implies the spins are within 0.1–1 nm of each other.	Quadrupole	A nucleus with $I \geq 1$ . The nucleus responds to both magnetic fields and electric-field gradients caused by the nonspherical distribution of charge induced by bonding.
Nyquist criteria/frequency	A set of fundamental criteria establishing the limits for digital data acquisition. For example, a minimum of two points taken in a specified time are required to define a wave and its frequency. The Nyquist frequency is the maximum frequency represented accurately in a spectrum and is set by the dwell time in the digital acquisition.	Radiofrequency (RF)	The energy difference between nuclear spin states in the magnetic field used for nuclear magnetic resonance falls in the RF region (kHz–GHz) of the electromagnetic spectrum. Thus, a source of RF electromagnetic radiation is used to excite nuclear spins.
Paramagnetic	Attracted into a magnetic field. Molecular paramagnetism is generated by unpaired electron spins. A paramagnetic interaction is one where energy is lowered by increasing the magnetic field. Electron orbital angular momentum may also generate paramagnetism. A paramagnetic chemical shift is to higher frequency.	Relaxation	The process where nuclei dissipate excess spin energy into other forms of molecular energy especially rotation and translation. $T_1$ (spin–lattice, longitudinal) and $T_2$ (spin–spin, latitudinal) relaxation are the independent time constants for the dissipation of $z$ and $x$ – $y$ magnetization, respectively. Spin–lattice relaxation – time constant $T_1$ – involves a change in the total $z$ magnetization of the sample as quanta of spin energy are exchanged with the surroundings. Spin–spin relaxation, $T_2$ , is the time constant for decay of spin order and coherence commonly represented as $x$ – $y$ magnetization. There is no energy exchange with the surroundings and spin order is randomized. Spin order is commonly randomized by mutual spin flips between dipolar coupled pairs. The $x$ – $y$ magnetization decays with time constant $T_2$ to its equilibrium value of zero. As there is no magnetic field in the $x$ – $y$ plane there is no $x$ – $y$ magnetization at equilibrium. When a $B_1$ or $B_2$ RF field is present, the $T_2$ relaxation time constant is replaced with $T_{1\rho}$ – relaxation in the rotating frame.
Polarization	The total net magnetization of spins created by unequal spin populations. When nuclear spins are placed in a magnetic field their random motion becomes orientated with respect to the field direction. For $I = 1/2$ nuclei, there are two spin states orientated with or against the field. At equilibrium the spin states are populated according to the Boltzmann distribution and so there are more spins in the lower energy state. This population imbalance creates an overall magnetization of the sample – the spins are polarized. Pulses generate nonequilibrium polarizations and polarization may be transferred between spins.	Residual dipolar coupling	In samples containing large anisotropic molecules such as phages and other additives, the motion of the small and large molecules becomes anisotropic. As a result the $(1 - 3 \cos^2 \theta)$ term in the dipole coupling expression does not average to zero. Residual dipole couplings between closely spaced nuclei may then be detected and correlated with molecular geometry.
Pulse	A short, usually microsecond, burst of RF electromagnetic radiation. The fixed RF is chosen to be close to the nuclear resonance frequency to be excited. The shorter the pulse, the broader the frequency range that may be excited. The pulse is applied through a conducting coil wound closely around the sample. Pulse nuclear magnetic resonance involves spectral excitation with a short	Resolution	Two peaks may only be resolved if their linewidth is smaller than the frequency difference between them. In nuclear magnetic resonance, resolution

	is usually governed by experimental factors such as magnetic field homogeneity and digital resolution. Digital resolution is set by the number of data points in a spectral window which in turn depends on total acquisition time. Absolute resolution is set by the spin lifetime which establishes the linewidth at half height. See also $T_2$ .		
Resonance	When the RF equals the energy gap between spin states that differ by one in spin quantum number, the spins are said to be in resonance.	Second-order coupling	This refers to the patterns observed for $J$ -coupled multiplets. In a first-order spectrum, the lines for spins coupled to groups containing $n$ equivalent atoms of spin $I$ are split into $2nI + 1$ lines. The spacing on both partners are equal, and equal to $J$ . This requires that the chemical shift difference between coupled spins is at least ten times $J$ . Second-order effects occur when $\Delta\delta \sim J$ , so called strong coupling. As $\Delta\delta$ approaches $J$ , at first line intensities become unequal within the multiplet and are canted towards the mean chemical shift (haystacking). Next, line spacings become unequal and differ from $J$ . When $\Delta\delta \sim J$ extra lines appear in the spectrum, intensities vary widely, and line spacings do not match $J$ .
Rigid lattice	The slow motion regime where $\omega_0\tau_c \gg 1$ and chemical shift anisotropy and dipole-dipole effects broaden spectra.		
Ring current	A magnetic dipole may be represented equivalently as a current loop. A magnetically anisotropic group such as a benzene ring acts as a magnetic dipole when placed in a magnetic field. The magnetic influence of such rings is often modeled as a current loop with electrons circulating around the $\pi$ system at a rate proportional to $B_0$ which is termed a ring current. Groups that are at fixed orientations relative to the magnetic dipole, ring current, have their resonances shifted by the induced ring current.	Shielding tensor	The electron motion around the nucleus induced by the applied magnetic field generates an internal magnetic field. The induced field is said to shield or screen the nucleus from the applied field and is observed as the chemical shift. The shielding tensor defines the principal values and orientation of the chemical shift in the molecular frame. The maximum difference between the principal values defines the CSA.
Rotating frame	The simplest Cartesian frame of reference to envisage for the nuclear magnetic resonance experiment is the laboratory frame with the magnetic-field direction defining the $z$ -axis. In the presence of an oscillating electromagnetic field (RF) the description of the motion of the nuclear spins and the net magnetization is complex in the laboratory frame. A mathematical transformation from the laboratory frame to a frame rotating at the RF oscillation frequency greatly simplifies the description of the magnetic field and the motion. (An analogy may be useful. The description of the motion of the moon in a solar coordinate system is greatly simplified by transforming to a coordinate system that rotates with the earth.) In the laboratory frame, the $x$ and $y$ coordinates are time independent, in the rotating frame the $B_1$ field is static in the $x'-y'$ plane which is rotating at the RF in the laboratory frame.	Shift reagent	A reagent which binds rapidly and reversibly to a molecule and induces a chemical shift. The common shift reagents are organometallic lanthanide complexes such as $\text{Eu(DPM)}_3$ that act as Lewis acids in nonpolar solvents. The reagents have a large magnetic anisotropy and induce substantial changes in chemical shift. The changes in shift depend on the binding constant, proximity, and orientation with respect to the binding site. The magnetic anisotropy effect is independent of the nucleus observed, so shift reagents are most effective in proton spectra.
		Shim	The adjustment of the magnet to achieve maximum homogeneity. In the early days of nuclear magnetic resonance, shimming involved physically moving the magnet pole faces relative to one another with pieces of wood (shims). On a modern spectrometer shimming is done by varying the current in carefully designed shim coils placed in specific orientations within the probe.
Saturation	The spins are said to be saturated when the spin populations are equalized by the absorption of RF energy. There is no residual $z$ magnetization and no signal can be excited.	Spectral window	The limited frequency range observed in a pulse nuclear magnetic resonance

	experiment. The spectral window is set by the dwell time for data point acquisition in the computer (see Nyquist). The strength and duration of the radio-frequency pulse also limit the frequency window that may be excited.		
Spin lock	The application of an on-resonance RF field to maintain the $x$ - $y$ magnetization in the direction of the $B_1$ field. The spin-lock pulse is usually applied immediately following a $90^\circ$ pulse in the $x$ - $y$ plane and perpendicular to the original $90^\circ$ pulse. It has the effect of fixing (locking) the $x$ - $y$ magnetization created by the initial pulse for a time of the order of $T_{1\rho}$ rather than $T_2$ .		$n$ times. During each acquisition time a free-induction decay is acquired in time $t_2$ . Fourier transformation with respect to $t_2$ will generate $n$ spectral sets with a standard frequency axis ( $F_2$ ) containing peaks modulated according to $t_1$ . A new time series is obtained by selecting each of the common frequency points in $F_2$ and following its magnitude as $t_1$ is incremented. A second Fourier transform with respect to $t_1$ and orthogonal to $t_2$ will generate new spectra with a new frequency dimension ( $F_1$ ). A 2D plot can be made with intensity providing the third dimension. New peaks are found if the $t_1$ and $t_2$ processes are correlated and occur at $\nu_{12}$ and $\nu_{21}$ in addition to the 1D frequencies at $\nu_{11}$ and $\nu_{22}$ . Pulse sequences have been developed to provide chemical shift information on one axis and coupling constant information on the other, i.e., 2D- $J$ spectroscopy. COSY (HETCOR, COLOC, HOMCOR, HMBC) spectra correlate chemical shifts of $J$ coupled nuclei, a very useful method for structure analysis. NOESY, ROESY, and TROESY spectra correlate chemical shifts of nuclei which are dipolar coupled. The higher dimension examples (3D, etc.) involve adding new timing sequences to separate variables of interest for a particular problem. Multidimensional spectra also provide enhanced peak separation in highly overlapped spectra of large proteins. 4D spectra may be created which show data with $^1\text{H}$ , $^{13}\text{C}$ , and $^{15}\text{N}$ frequency axes with intensity given by the NOESY spatial coupling information.
Spinning sidebands	Peaks in a spectrum that are separated from the true peaks by integer multiples of the sample spinning frequency. Spinning sidebands occur as sample spinning modulates the effective field at the nucleus when the sample rotates through different field strengths caused by magnetic inhomogeneities (solution) or spatially dependent fields (e.g., chemical shift anisotropy, dipole coupling, quadrupole coupling) in solids.		
Spin temperature	For a spin system at equilibrium with its surroundings the spin energy levels are populated according to the Boltzmann distribution: $n_i/n_0 = \exp(-\Delta E/kT) = \exp(-\hbar\gamma B_0/kT)$ (4). The energy gap is fixed for a given spectrometer and nucleus and so the spin population ratio is inversely proportional to absolute temperature with a calculable proportionality constant. The spin temperature of a sample is then defined by this spin population ratio. For example, following a $90^\circ$ pulse when the spin populations are equal, the temperature is said to be infinite. It is possible to create a negative spin temperature by inverting the spin population with a $180^\circ$ pulse.	$x$ - $y$ Magnetization ( $M_{xy}$ )	The instantaneous value of the magnetic moment of the sample along the axis perpendicular to the $B_0$ field. In the laboratory frame, the $x$ - $y$ magnetization is rotating with the Larmor frequency. It varies with time constant $T_2$ and has an equilibrium value of zero.
$T_1$ and $T_2$	The spin-lattice and spin-spin relaxation times. See Relaxation.	Zeeman energy	The energy of the nuclear spin states with different spin quantum numbers in a magnetic field. Often used to describe phenomena where the energy is proportional to the strength of the magnetic field.
2D, 3D, 4D, etc., NMR	Nuclear spins may be perturbed by an enormous variety of pulse sequences. If the response of the spin system is detected as a function of independent time variables, subsequent Fourier transformation with respect to each time set will provide an independent frequency dimension for each time sequence. For example, a set of $n$ free-induction decays could be obtained with the sequence Pulse- $t_1$ -Pulse- $t_1$ -Acquire with the time delay $t_1$ incremented	Zeugmatography	The original name given to MRI by its inventor, Paul Lauterbur.
		$z$ -Magnetization ( $M_z$ )	The instantaneous value of the magnetic moment of the sample along the axis of the applied field ( $z$ -axis). It varies with time constant $T_1$ and has an equilibrium value denoted $M_0$ .

## Instrumentation

J C Lindon, Imperial College London, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

After the first observation of bulk phase nuclear magnetic resonance (NMR) in 1945 and the realization that it would be useful for chemical characterization that came with the discovery of the chemical shift in 1951, it was only a few years before the first commercial spectrometers were being produced. By the end of the 1950s, a considerable number of publications on the application of NMR to chemical structuring and analysis problems had appeared and then during the 1960s and later, it became clear that useful answers could be obtained in biological systems. Since then the applications and the consequential instrument developments have diversified and now NMR spectroscopy is one of the most widely used techniques in chemical and biological analyses. The very high specificity, the exploratory nature of the technique without the need to preselect analytes,

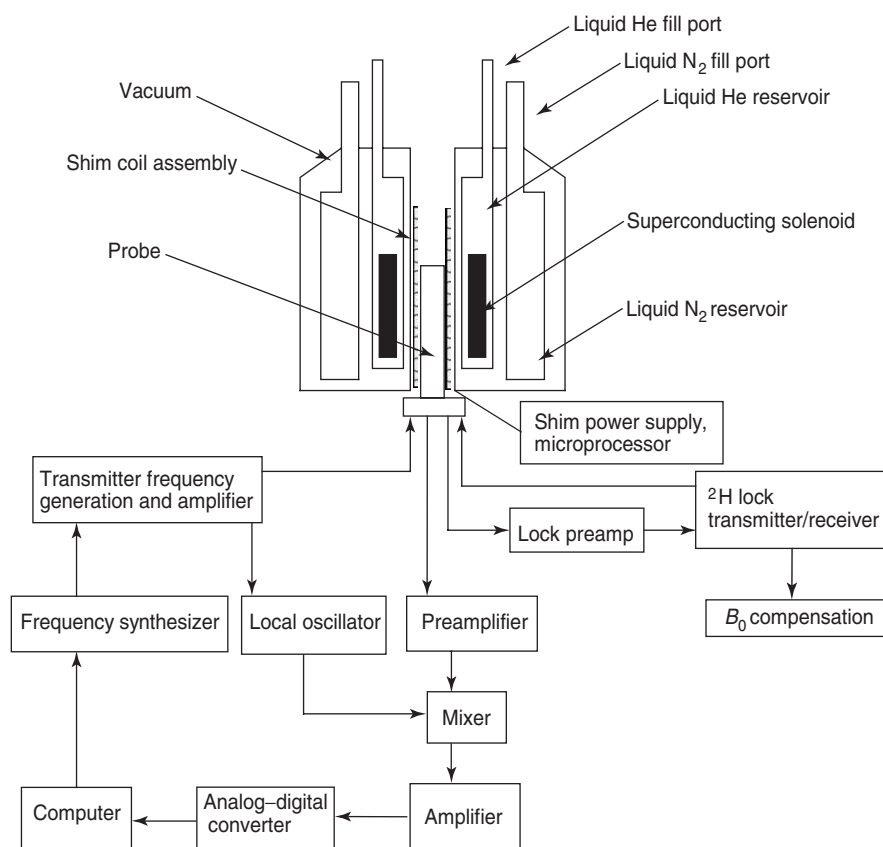
and its nondestructive nature have made it very useful despite its lower sensitivity compared to some other spectroscopic methods.

A general description is given here of the way in which a modern NMR spectrometer operates, of the various components that go into making a complete system, and the particular role that they play. Today, with the level of computer control present in modern spectrometers, this naturally includes a description of both the hardware and software. An overview diagram of the components of a high-resolution NMR spectrometer is given in Figure 1.

### Components and Principles of Operation of NMR Spectrometers

#### Continuous Wave and Fourier Transform Operation

For many years, all commercial NMR spectrometers operated in continuous wave (CW) mode. This type of operation required a sweep of the NMR frequency or the magnetic field over a fixed range to bring each nucleus into resonance one at a time. These scans for



**Figure 1** A block diagram of the principal components of a modern NMR spectrometer.

$^1\text{H}$  would take typically 500 s to avoid signal distortion. Since most NMR spectra consist of a few sharp peaks interspersed with long regions of noise, this was a very inefficient process. A fundamental paper by Ernst and Anderson in 1966 pointed out the favorable gain in efficiency that could be obtained by simultaneously detecting all signals. This is achieved by the application of a short intense pulse of radio-frequency (RF) radiation to excite the nuclei followed by the detection of the induced magnetization in an RF detector coil as the nuclei relax. The decaying, time-dependent signal, known as a free induction decay (FID) is then converted to the usual frequency domain spectrum by the process known as Fourier transformation. The efficient calculation of the digital Fourier transform (FT) requires the number of data values to be a power of two, typically perhaps 16 K points for modest spectral widths, up to 128 K or even 256 K points for wide spectral widths on high-field spectrometers (1 K is 1024 or  $2^{10}$  points). Acquisition of a  $^1\text{H}$  FID requires typically a few seconds and opens up the possibility of adding together multiple FID scans to improve the spectrum signal-to-noise ratio (S/N) since for perfectly registered spectra, the signals will co-add but the noise will only increase in proportion to the square root of the number of scans. The S/N gain therefore is proportional to the square root of the number of scans. This for the first time made routine the efficient and feasible acquisition of NMR spectra of less sensitive or less abundant nuclei such as  $^{13}\text{C}$ .

### The Magnet and Associated Components

The most fundamental component of an NMR spectrometer is the magnet. Originally, this would have been a permanent or electromagnet and these provided the usual configurations for field strengths up to 1.41 T (the unit of magnetic field strength is the tesla (T) equivalent to 10 000 gauss (G)), corresponding to a  $^1\text{H}$  observation frequency of 60 MHz. Because the sensitivity of the NMR experiment is proportional to about the  $3/2$  power of the field strength, denoted  $B_0$ , the drive to higher and higher magnetic fields has been remorseless. This led the commercial NMR manufacturers to develop stronger electromagnets for NMR spectroscopy. This took the highest field strengths to 100 MHz for  $^1\text{H}$  observation, i.e., 2.35 T. Materials suitable for electromagnets have maximum saturation field strengths at about this value. Because of the continued need for even higher field strengths, the development of magnets based on superconducting solenoids was commenced. In this case, the magnetic field is generated by a current circulating in a coil of superconducting wire

immersed in a liquid helium Dewar at 4.2 K. This bath is shielded from ambient temperature by layers of vacuum and a jacket of liquid nitrogen at 77 K, which is usually topped up at a weekly interval. A liquid helium refill is carried out at  $\sim 2$ –3-month intervals depending on the age and field strength of the magnet. The initial development of superconducting magnets was at 5.17 T corresponding to 220 MHz for  $^1\text{H}$  and operated in CW mode. Until  $\sim 1972$ , this represented the highest field strength, but then at regular intervals the available field strength gradually increased along with the emergence of wider bore magnets enabling the incorporation of larger samples. Thus, a 270 MHz spectrometer was produced along with a wide-bore 180 MHz machine and then the field was increased to allow  $^1\text{H}$  observation at 360, 400, 500, 600, 700, 750, and 800 MHz, and the present limit of any machine yet delivered to a customer is at 900 MHz (2003). For instruments that operate for  $^1\text{H}$  NMR at 700 MHz or greater, the liquid helium bath is kept at  $\sim 2$  K by a pumped refrigeration system so that the higher currents needed for the higher fields can be achieved. A modern high-field NMR spectrometer using this type of superconducting magnet is shown in Figure 2. One major achievement has been the introduction of actively shielded magnets in which the stray magnetic field from the main solenoid is reduced in extent by opposing magnetic field coils. This brings the stray field (e.g., the 5 G limit, outside of which magnetic metal objects, and people with cardiac pacemakers



**Figure 2** A modern high-resolution NMR spectrometer. The superconducting magnet is shown at the center, in this case providing a field of 21.2 T corresponding to a  $^1\text{H}$  observation frequency of 900 MHz. On the right is the console containing the RF, other electronics, and the temperature control unit (photograph reproduced by permission of Bruker Biospin GmbH, Rheinstetten, Germany).



should be kept), to  $<1.5$  m for a 600 MHz NMR system. This facilitates the adjacent positioning of other equipment and even allows NMR spectrometers to be much closer to each other than before. Nowadays, apart from very basic routine low-field spectrometers used, for example, for monitoring chemical reactions, all NMR spectrometers are based on superconducting magnets.

At present, 900 MHz for  $^1\text{H}$  NMR spectroscopy is the current commercial limit and the first machines at this field have been produced. Higher fields must be on the way and clearly an emotive figure would be the 1 GHz  $^1\text{H}$  NMR spectrometer. This development will require both the design of transmitter and detection technology working at or beyond the limit of RF methods and developments in superconducting wire technology. Although the higher field strengths provide greater spectral dispersion and yield better sensitivity, it may be that some applications involving heavier nuclei are less suited to such high fields because of the field dependence of certain mechanisms of nuclear spin relaxation that could cause an increased line broadening and hence lower peak heights and detectability.

Inserted into the magnet is the NMR detector system or probe. High-resolution NMR spectra are usually measured in the solution state in glass tubes of standard external diameters, 5 mm being the most common, but larger ones (10 mm) are used where improved sensitivity is required and sample is not limited. Also, a range of narrow tubes is available for limited sample studies (4 mm, 3 mm, and even smaller specially shaped cavities such as capillaries or spherical bulbs, etc.). Very small sample sizes can be accommodated in specialized micropubes with sample volumes in the microliter range. As described below, it is now possible to measure NMR spectra using special probes in a flow-injection mode avoiding the use of sample tubes completely. The probe contains tuneable RF coils for excitation of the nuclear spins and detection of the resultant signals as the induced magnetization decays away. A capability exists for measuring NMR spectra over a range of temperatures, typically  $-150^\circ\text{C}$  to  $+200^\circ\text{C}$ .

A major advance in detection of NMR signals has been the development of probes in which the RF coil and the preamplifier are cooled close to the temperature of liquid helium, but with the sample remaining at ambient temperature. These so-called cryoprobes have a S/N ratio improvement of  $\sim 500\%$  over conventional probes of the same sample diameter. This is because the thermal noise level in the circuitry scales approximately as the square root of the ratio of the absolute temperatures. There are some limitations to this improvement for highly conducting

solutions such as strongly ionic solutions, but these can be minimized by the use of smaller sample diameters thereby reducing dielectric losses.

Even though modern high-resolution magnets have very high field stability and homogeneity, this is not sufficient for chemical analysis, in that it is necessary to resolve lines to about a width of 0.2 Hz and, at a typical common operating frequency of 600 MHz, this represents a stability of one part in  $3 \times 10^9$ . This performance is achieved in two ways. Usually, deuterated solvents are used for NMR spectroscopy to avoid the appearance of solvent peaks in the  $^1\text{H}$  spectrum. Deuterium is an NMR active nucleus and the spectrometer will contain a  $^2\text{H}$  channel for exciting and detecting the solvent resonance. Circuitry exists in the spectrometer for maintaining this  $^2\text{H}$  signal exactly on resonance at all times by detecting any drift from resonance caused by inherent magnet drift or room temperature fluctuations and for providing an error signal to bring the magnet field back on resonance by applying small voltages through subsidiary coils in the magnet bore. This is known as a 'field-frequency lock' and it means that successive scans in a signal accumulation run are exactly coregistered. To improve the homogeneity of the magnet an assembly of coils is inserted into the magnet bore (shim coils). These consist of  $\sim 20$ –40 coils specially designed so that adjustable currents can be fed through them to provide corrections to the magnetic field in any combination of axes to remove the effects of field inhomogeneities. The criterion of the best homogeneity is based upon the fact that when the  $^2\text{H}$  lock signal is sharpest (i.e., the most homogeneous field) the signal will be at its highest. The currents in the shim coils are usually adjusted (called shimming), therefore, to give the highest lock signal. Alternatively, it is possible, although less common, to shim on the  $^1\text{H}$  NMR signal. This whole process is now largely computer-controlled in modern spectrometers. A form of shimming the magnetic field to obtain high homogeneity has arisen from concepts from magnetic resonance imaging (MRI). Here, the inhomogeneous magnetic field is mapped in three dimensions initially using the application of field gradients applied along the three orthogonal axes. Then small currents can be applied to the shim coils in a computer-optimized fashion over several iterations of the process. Often for subsequent operation only the homogeneity along the magnetic field direction needs to be optimized using a single gradient map.

NMR spectra were usually measured with the sample tube spinning at  $\sim 20$  Hz to further improve the NMR resolution. This can introduce signal sidebands at the spinning speed and its harmonics, and

on modern high-field machines with improved resolution, this is becoming less necessary and is undesirable in some cases.

In analytical laboratories where large numbers of samples have to be processed, it is accepted that automatic sample changers can play a large part in improving efficient use of the magnet time. These devices allow the measurement of up to  $\sim 120$  samples in an unattended fashion with the insertion and ejection of samples from the magnet under computer control. Automatic lock detection and optimization, sample spinning, NMR receiver gain, and shimming are also standard. The data are acquired automatically and can be plotted and stored on backing devices. As an additional aid in routine work, it is possible to purchase an automated work bench that will produce the samples dissolved in the appropriate solvent in an NMR tube starting from a solid specimen in a screw-capped bottle; it will also dispose of samples safely and wash the NMR tube.

If it is necessary to measure NMR spectra on large numbers of samples, e.g., from combinatorial chemical synthesis or from large biofluid studies, then flow-injection probes are now in general use. Samples can be made up in 96-well plates using a specialized robotic system; the plate is then transferred to a second robotic system in which the contents of a well can be extracted and flowed into the NMR probe where the sample is stopped and any NMR experiments carried out. After measurement, the sample is then sent back to the same well, to a well in a different plate, or to waste as desired.

### Excitation, Detection, and Computer Processing of NMR Signals

The RF signal generation is derived ultimately from a digital frequency synthesizer that is gated and amplified to provide a short intense pulse. Pulses have to be of short duration because of the need to tip the macroscopic nuclear magnetization by  $90^\circ$  or  $180^\circ$  and at the same time to provide uniform excitation over the whole of the spectral range appropriate for the nucleus under study. Thus, for  $^{13}\text{C}$  NMR, for example, where chemical shifts can cover  $>200$  ppm this requires a 25 KHz spectral width on a spectrometer operating at 500 MHz for  $^1\text{H}$ , which corresponds to 125 MHz for  $^{13}\text{C}$ . To cover this range uniformly requires a  $90^\circ$  pulse to be  $<10\ \mu\text{s}$ . The RF pulse is fed to the NMR probe which contains one or more coils which can be tuned and matched to the required frequency, this tuning changing from sample to sample because of the different properties of the samples such as the solution dielectric constant. The receiver is blanked off during the pulse and for a

short period afterwards to allow the pulse amplifier to ring down. The receiver is then turned on to accept the NMR signal that is induced in the coil as the nuclei precess about the field and decay through their relaxation processes. The detection coil is wound on a former as close as possible to the sample to avoid signal losses and is oriented with its axis perpendicular to the magnetic field. In a superconducting magnet the sample tube is aligned along the field, and a simple solenoid coil around the sample that would provide the best S/N is not possible. Consequently, most detector coils are of the saddle type. However, in a new generation of microsample coils using the flow-injection principle it is possible to have the sample chamber horizontal and hence a solenoid coil perpendicular to the magnetic field can then be used. The weak NMR signal is amplified using a preamplifier situated as close to the probe as possible, and then also in the main receiver unit where it is mixed with a reference frequency and demodulated in several stages leaving the free induction decay (FID) as an oscillating voltage in the kilohertz range. This signal is then fed to an analog-to-digital converter (ADC) and at this point the analog voltage from the probe is converted into a digital signal for data processing. ADCs are described in terms of their resolution, usually in terms of the number of bits of resolution, i.e., a typical high-field NMR FID is digitized to a resolution of 16 bits or one part in  $2^{16}$  or 65 536. If only one ADC is used to collect the NMR FID, it is not possible to distinguish frequencies that are positive from negative with respect to the pulse frequency. For this reason, the carrier frequency was set to one edge of the spectral region of interest to make sure that none of the NMR peaks were 'aliased', i.e., folded into the spectral region from outside it. This had the disadvantage of allowing all of the noise on the unwanted side of the carrier being aliased onto the noise in the desired spectral region; hence, reducing the final S/N by  $\sqrt{2}$ . To overcome this problem it is general now to collect two FIDs separated in phase by  $90^\circ$  using either two ADCs or to multiplex one ADC to two channels. This approach allows the distinction of positive and negative frequencies and means that the carrier can be set in the middle of the spectrum, the hardware filters correspondingly reduced in width by a factor of 2 and an increase in S/N by  $\sqrt{2}$ . This process is termed quadrature detection. The digital signals can then be manipulated to improve the S/N ratio or the resolution by multiplying the FID by an appropriate weighting function before the calculation of the digital FT.

Modern NMR spectrometers usually have two separate computer systems. One is dedicated to the

acquisition of the NMR FID and operates in background so that all necessary accurate timing requirements can be met. The FID is transferred, either at the end of the acquisition or periodically throughout it to enable inspection of the data, to the host computer for manipulation by the operator. These computers are based on modern operating systems such as UNIX or Windows. The computer software can be very complex and can, like any modern package, take advantage of networks, printers, and plotters. Typical operations include manipulations of the signal-averaged FID by baseline correction to remove DC offset, multiplication by continuous functions to enhance S/N or resolution, Fourier transformation, phase correction, baseline correction of the frequency spectrum, calculation and output of peak lists, calculation and output of peak areas (integrals), and plotting or printing of spectra. NMR data processing software can also be purchased from a number of companies other than the instrument manufacturers, and these often have links to document production software or provide output of NMR parameters for input into other packages such as those for molecular modeling. A number of approaches alternative to the use of FID weighting functions for improving the quality of the NMR data have been developed and are available from software suppliers. These include such methods as maximum entropy and it is possible to purchase these as supplementary items from some NMR manufacturers.

### Multiple Pulse Experiments and Multidimensional NMR

So far everything described applies to the basic one-dimensional (1D) NMR experiment when the nuclear spin system is subjected typically to a  $90^\circ$  pulse and the FID is collected. A wide variety of experiments exist in the literature and are routinely applied to measure NMR properties such as relaxation times  $T_1$ ,  $T_2$ , and  $T_{1\rho}$ , which can be related in some cases to molecular dynamics. These experiments involve the use of several pulses separated by timed variable delays and are controlled by pulse programs written in a high-level language for ease of understanding and modification. The computer system will have software to interpret the data and calculate the relaxation times using least-squares fitting routines.

Such pulse programs are also used to enable other special 1D experiments such as saturation or non-excitation of a large solvent resonance (these are different in that the former method will also saturate NH or OH protons in the molecules under study through the mechanism of chemical exchange), or the measurement of nuclear Overhauser enhancement effects

that are often used to provide distinction between isomeric structures or to provide estimates of internuclear distances. Pulse programs are also used for measuring NMR spectra of nuclei other than  $^1\text{H}$  and sometimes in order to probe connectivity between protons and the heteronucleus. In this case, pulses or irradiation can be applied on both the heteronucleus and  $^1\text{H}$  channels in the same experiment. The commonest use is in  $^{13}\text{C}$  NMR where all spin-couplings between the  $^{13}\text{C}$  nuclei and  $^1\text{H}$  nuclei are removed by 'decoupling'. This involves irradiation of all of the  $^1\text{H}$  frequencies whilst observing the  $^{13}\text{C}$  spectrum. In order to cover all of the  $^1\text{H}$  frequencies, the irradiation is provided as band of frequencies covering the  $^1\text{H}$  spectral width; this is consequently termed noise decoupling or broadband decoupling. Alternatively, it is possible to obtain the effect of broadband decoupling more efficiently by applying a train of pulses to the  $^1\text{H}$  system, this being known as composite pulse decoupling.

A whole family of experiments have been developed that detect low-sensitivity nuclei such as  $^{13}\text{C}$  or  $^{15}\text{N}$  indirectly by their spin coupling connectivity to protons in the molecule. This involves a series of pulses on both  $^1\text{H}$  and the heteronucleus but allows detection at the much superior sensitivity of  $^1\text{H}$  NMR. Special probes have been developed for such 'indirect detection' experiments in which the  $^1\text{H}$  coil is placed close to the sample, and the heteronucleus coil is placed outside it, the opposite or 'inverse geometry' to a standard heteronuclear detection probe.

The 1D NMR experiment is derived from measuring the FID as a function of time. If the pulse program also contains a second time period that is incremented, then a second frequency axis can be derived from a second FT. This is the basis for '2D' NMR and its extension to three or even four dimensions. For example, a simple sequence such as

$$90^\circ(^1\text{H}) - t_1 - 90^\circ(^1\text{H}) - \text{collect FID for time } t_2$$

where  $t_1$  is an incremented delay, results after double Fourier transformation, with respect to  $t_1$  and  $t_2$ , in a spectrum with two axes each corresponding to the  $^1\text{H}$  chemical shifts. This is usually viewed as a contour plot with the normal 1D spectrum appearing along the diagonal and any two protons that are spin coupled to each other giving rise to an off-diagonal contour peak at their chemical shift coordinates. This simple experiment is one of a large family of such correlation experiments involving either protons alone or heteronuclei. The extension to higher dimensions has already been exploited to decrease the amount of overlap by allowing spectral editing and the spreading of the peaks into more

than one dimension. Hardware and software in modern NMR spectrometers allows this wide variety of experiments.

The increasingly complex pulse sequences used today rely on the ability of the equipment to produce exactly  $90^\circ$  or  $180^\circ$  or any other angle pulses. One way to do this is to provide trains of pulses that have the desired net effect of, for example, a  $180^\circ$  tip, but which are compensated for any mis-setting. An example of such a 'composite pulse' is  $90^\circ_x-180^\circ_y-90^\circ_x$ , which provides a better inversion pulse than a single  $180^\circ$  pulse. Many complex schemes have been invented both for observation and decoupling (especially for low-power approaches which avoid heating the sample). A universal approach to removing artifacts caused by electronic imperfections and which is also used to simplify spectra by editing out undesired components of magnetization is the use of 'phase cycling'. This allows the operator to choose the phase of any RF pulse and of the receiver and cycling these in a regular fashion gives control over the exact appearance of the final spectrum.

For spectral editing purposes or to prove some NMR spin connectivity, it can be very convenient to excite only part of a spectrum, possibly only that corresponding to a given chemical shift or even one transition in a multiplet. This approach termed 'selective excitation' is achieved by using lower-power pulses applied for a longer period of time (e.g., a 10 ms  $90^\circ$  pulse will only cover 25 Hz). Such selective pulses are often not rectangular as for hard pulses but can be synthesized in a variety of shapes such as sine or Gaussian because of their desirable excitation frequency profiles. Modern research spectrometers can include such selective, shaped pulses in pulse programs.

Another method for achieving selective detection of certain types of spin systems is through the application of pulsed magnetic field gradients. This can be used to select particular coherence pathways for spins in a multiple pulse experiment, for example, in multi-dimensional NMR spectroscopy, to crush unwanted magnetization such as solvent resonance, to measure molecular diffusion coefficients, and, hence, to edit spectra on the basis of the diffusion coefficients of the molecules giving rise to individual peaks.

## Use of NMR Instrumentation

The vast majority of NMR spectrometers are used by chemists for molecular identification purposes in solution, mainly for small organic molecules. However, NMR spectroscopy can be used in a wide variety of application areas including macromolecule structure studies (e.g., proteins, RNA, DNA, polysaccharides),

molecular interaction studies (such as drug-protein binding), and, of course, in inorganic chemistry where a wide range of nuclides have been investigated. However, for some applications specialized instrumentation and methods are necessary and these areas are summarized here.

### NMR of Macromolecules

The determination of the 3D structure of proteins in the solution state is carried out by molecular modeling based on distance and angle constraints derived from NMR spectroscopy. Through the use of extensive  $^{13}\text{C}$ ,  $^{15}\text{N}$ , and  $^2\text{H}$  labeling procedures, a considerable number of structures have been obtained. However, even with the use of such techniques the NMR linewidths increase as the protein molecular weight increases and it has not been possible to study large proteins. The main cause of the linewidth increases is fast  $T_2$  relaxation caused by dipolar coupling and chemical shift anisotropy combined with slow molecular motion. However, a technique called transverse-relaxation optimized spectroscopy (TROSY) overcomes this limitation by making use of cancellation of the transverse relaxation effects between the dipolar coupling and chemical shift anisotropy. This is manifested, for example, in a 2D  $^1\text{H}$ - $^{15}\text{N}$  NMR spectrum that gives four lines for each  $^1\text{H}$ - $^{15}\text{N}$  correlation and TROSY selectively observes only the one component that has the cancelled effect and hence a narrower line. This improvement in resolution allows much larger proteins to be studied. Theory predicts that the effect is dependent on the magnetic field  $B_0$ , and should be complete at 900–1000 MHz, but that advantage already exists at 750 MHz. The TROSY technique has enabled structural studies on proteins of molecular weight  $>100\,000$  Da.

### NMR of Solids

Although  $^1\text{H}$  high-resolution NMR spectroscopy is possible in the solid, most applications to organic molecules have focused on heteronuclei such as  $^{13}\text{C}$ , although there are many published studies of inorganic systems that use NMR spectra of other nuclei such as  $^{29}\text{Si}$  and  $^{27}\text{Al}$ . High-resolution studies rely on very short RF pulses, so high-power amplifiers are necessary. Similarly, because of the need to decouple  $^1\text{H}$  from an observed nucleus such as  $^{13}\text{C}$  and thereby to remove dipolar interactions not seen in liquid state, high-power decoupling is required. However, the major difference between solution and solid-state high-resolution NMR studies lies in the use of 'magic-angle-spinning' in the latter case. This involves spinning the solid sample packed into a special rotor at an angle of  $54.7^\circ$  to the magnetic

field. This removes broadening due to any chemical shift anisotropies that are manifested in the solid-state spectrum and any residual dipolar coupling not removed by high-power decoupling. Typical spinning speeds are 2–6 kHz or 120 000–720 000 rpm, although higher speeds up to 25 kHz, where the rotor rim is moving at supersonic velocity, are possible and necessary in some cases.

### Separations Directly Coupled to NMR Spectroscopy

This has become a major use of modern NMR spectrometers, since the technology became commercially available. Many types of separation have been coupled directly to NMR spectrometers, and also to mass spectrometers in parallel, for simultaneous NMR and mass spectrometry analysis and identification of eluting materials. Published approaches where separations have been directly coupled to NMR spectrometers include solid-phase extraction, liquid chromatography (LC), capillary LC, supercritical fluid chromatography, capillary electrophoresis, and capillary electrochromatography. There is now a large literature on this subject. Application areas include impurity profiles and their identity for fine chemicals, identification of drug metabolites, and characterization of natural products including those of potential medicinal use.

### NMR Imaging

A whole new specialized subdivision of NMR has arisen in the allied disciplines of NMR imaging (MRI) and NMR spectroscopy from localized regions of a larger object. Applications range from the analysis of water and oil in rock obtained from oil exploration drilling to medical and clinical studies. Thus, spectroscopic applications include the possibility of measuring the  $^1\text{H}$  or  $^{31}\text{P}$  NMR spectrum from a particular volume element in the brain of a living human being and relating the levels of metabolites seen in a diseased condition. Some experiments on smaller samples can be carried out in the usual vertical bore superconducting magnets but studies are more often performed in specially designed horizontal bore magnets with a large clear bore capable of taking samples up to the size of adult human beings. Because of the large bore, they operate at lower field strengths compared to analytical chemical applications, and typical configurations would be 2.35 T with a 40 cm bore or 7.0 T with a 21 cm bore. Clinical imagers utilize magnetic fields up to 3 T with a 1 m bore. Imaging relies upon the application of magnetic field gradients to extra coils located inside the magnet bore in all three orthogonal

axes including that of  $B_0$  and excitation using selective RF pulses. Virtually all clinical applications of MRI use detection of the  $^1\text{H}$  NMR signal of water in the subject with the image contrast coming from variation of the amount of water, or its NMR relaxation or diffusion properties in the different organs or compartments being imaged. Very fast imaging techniques have been developed that allow movies to be constructed of the beating heart or studies of changes in brain activity as a result of light or aural stimulation to be conducted.

### Industrial Analysis

Specialist tabletop machines can be purchased and these are used for routine analysis in the food and chemical industries. They operate automatically at typically 20 MHz for  $^1\text{H}$  NMR using internally programmed pulse sequences and are designed to give automatic printouts of analytical results such as the proportion of fat to water in margarine or the oil content of seeds.

There have been a number of NMR-based devices that are used to measure properties of materials that are located remotely from the NMR magnet. Examples include the use of specialized devices for oil-well logging where measurements of NMR parameters can be made down an oil well, giving information on the presence of oil and water. Similarly, a device named the NMR mobile universal surface explorer (NMR MOUSE) has been developed for applications in materials science. This can measure relaxation and diffusion properties of elastomers, industrial coatings, and objects such as rubber vehicle tyres for quality control purposes.

**See also:** **Chromatography:** Overview. **Liquid Chromatography:** Instrumentation; Liquid Chromatography–Nuclear Magnetic Resonance Spectrometry. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Principles; Multidimensional Proton; Solid-State; *In Vivo* Spectroscopy Using Localization Techniques.

### Further Reading

- Albert K (ed.) (2002) *On-line LC-NMR and Related Techniques*. Chichester: Wiley.
- Ernst RR and Anderson WA (1966) Applications of Fourier transform spectroscopy to magnetic resonance. *Review of Scientific Instruments* 37: 93–102.
- Friebolin H (1998) *Basic One- and Two-Dimensional NMR Spectroscopy*, 3rd edn. Weinheim: Wiley-VCH.
- Lindon JC and Ferrige AG (1980) Digitisation and data processing in Fourier transform NMR. *Progress in NMR Spectroscopy* 14: 27–66.
- Sanders JKM and Hunter BK (1993) *Modern NMR Spectroscopy. A Guide for Chemists*, 2nd edn. Oxford: Oxford University Press.



# NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY-APPLICABLE ELEMENTS

Contents

**Hydrogen Isotopes**

**Carbon-13**

**Fluorine-19**

**Nitrogen-15**

**Phosphorus-31**

**Organometallic Compounds**

## Hydrogen Isotopes

**J R Jones**, University of Surrey, Guildford, UK

**S-Y Lu**, National Institute of Mental Health, Bethesda, MD, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Nuclear magnetic resonance (NMR) spectroscopy, arguably the most powerful technique in chemistry, is a relatively insensitive method, but in the short time interval between the appearance of the first and second editions of the *Encyclopedia of Analytical Sciences*, considerable improvements have been made. Consequently, the range of applications that now become possible is greatly expanded as is illustrated here for the two isotopes of hydrogen, tritium ( $^3\text{H}$ ) and deuterium ( $^2\text{H}$ ); further developments are also anticipated.

## Tritium NMR Spectroscopy

Tritium ( $^3\text{H}$  or T) possesses ideal NMR properties – spin of  $\frac{1}{2}$ , high sensitivity (21% better than for  $^1\text{H}$ ,  $\gamma_{^3\text{H}}/\gamma_{^1\text{H}} = 1.06664$ , where  $\gamma_{^3\text{H}}/\gamma_{^1\text{H}}$  is the gyromagnetic ratio) and extremely low natural abundance, but because it is radioactive (a weak  $\beta^-$ -emitter with a half-life of 12.3 years, Table 1) there is the need for associated radiochemical facilities and trained

personnel and this is the main reason why, outside the companies that supply tritiated compounds, and the pharmaceutical companies that use the products, together with a small number of universities/research institutions, the technique is not widely used.

Two recent developments are likely to improve the attractions of this particular isotope. First, a whole range of microwave-enhanced procedures have been developed which allow tritiations to be carried out more rapidly, sometimes more selectively, and, finally, with the production of less radioactive waste. The latter can frequently be converted to tritiated water that can then be reused. This represents a good example of ‘Green Chemistry’.

The second development relates to NMR sensitivity. Improvements in spectrometer technology over the last 20 years of the twentieth century led to an approximate threefold increase in NMR sensitivity. Now with the development of cryoprobes there is the possibility of being able to reduce the measurement time by an order of magnitude, or more. This kind of benefit will be maximized for nuclei such as tritium that have an extremely low natural abundance.

Cooling the RF coils of a probe to cryogenic temperature improves the RF efficiency and reduces the noise generated by the coils. Further improvements can be achieved if the preamplifier is also cooled to cryogenic temperatures, as in this way the noise generated in the circuit is reduced. These objectives, despite the formidable challenges presented by the need to keep the sample temperature stable at close

**Table 1** Some of the important properties of the hydrogen isotopes

Nucleus	Natural abundance (%)	Nuclear spin	NMR sensitivity	Radioactivity
$^1\text{H}$	99.985	1/2	1	No
$^2\text{H}$	0.015	1	$9.65 \times 10^{-3}$	No
$^3\text{H}$	$< 10^{-16}$	1/2	1.21	Weak $\beta^-$ , $t_{1/2} = 12.3$ years, $E_{\text{max}} = 18$ keV, specific activity = $29 \text{ Ci mmol}^{-1}$

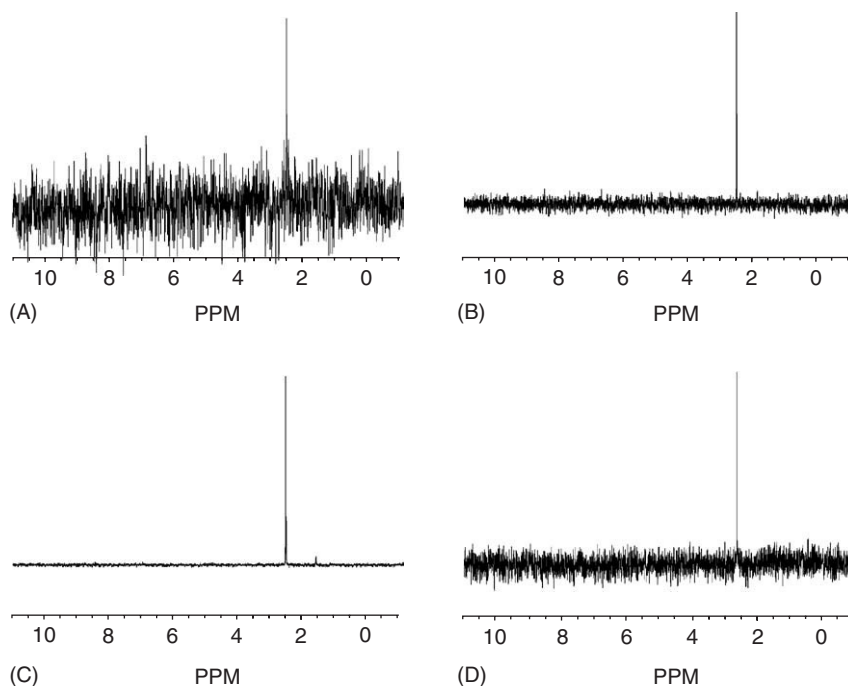
to room temperature, while the RF coils nearby are cooled to below 35 K, have now been achieved. The first tritium results have recently been published (Figure 1) and the observed improvement in signal-to-noise ratio is 4.3, equating to close on a 20-fold saving in time when acquiring a  $^3\text{H}$  NMR spectrum. In the early days of  $^3\text{H}$  NMR spectroscopy it was customary practice to use samples containing 20 mCi (740 MBq; MBq (megabecquerel) is the unit of radioactivity) of radioactivity. Nowadays, with a 500 MHz NMR spectrometer and a cryoprobe 20  $\mu\text{Ci}$  (0.74 MBq) samples are sufficient. Such an improvement is comparable to the performance expected from a 900 MHz NMR spectrometer but at a considerably lower cost. This 1000-fold improvement, if repeated in the next 35 years, would take the detection limit down to that of liquid scintillation counting. Added to the possibility of tritium magnetic resonance imaging it is clear that these techniques will continue to grow in importance.

### Applications

The major use of  $^3\text{H}$  NMR spectroscopy continues to be in the analysis of a tritiated compound produced through one of the customary routes,

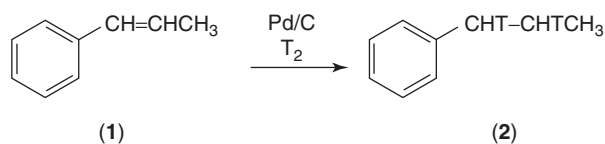
e.g., hydrogenation with  $\text{T}_2$  gas of an unsaturated precursor in the presence of a homogeneous or heterogeneous catalyst, aromatic dehalogenation, again in the presence of a catalyst, borotritide-type reduction or methylation using  $\text{CT}_3\text{I}$ , and the subsequent specific activity determination. At low specific activity ( $<1 \text{ Ci mmol}^{-1}$ ),  $^3\text{H}$ - $^3\text{H}$  couplings will be absent and the  $^3\text{H}$  NMR ( $^1\text{H}$  decoupled) spectra will consist of a series of single lines, integration of which gives the relative incorporation of  $^3\text{H}$  at each site. Nuclear Overhauser effects (NOEs) are small and differential effects even smaller.  $^3\text{H}$  chemical shifts can be related to the  $^1\text{H}$  chemical shift via the Lamor frequency ratio. At high specific activity,  $^3\text{H}$ - $^3\text{H}$  coupling constants are related to the corresponding  $^1\text{H}$ - $^1\text{H}$  coupling constants, with small isotope effects, and can be used to determine the nature and proportion of different isotopomers.

There is a growing realization of the power of the technique in the catalysis area, especially when combined with the use of modern multipulse procedures such as distortionless enhancement by polarization transfer (DEPT), 2D tritium/proton correlation, 2D  $J$ -resolved NOE difference spectroscopy (NOESY), and 2D exchange spectrum (EXSY). Thus, in the



**Figure 1**  $^3\text{H}$  NMR spectra of tritiated *o*-methoxyacetophenone ( $o\text{-MeO-C}_6\text{H}_4\text{COCH}_2\text{T}$ , 68  $\mu\text{Ci}$ ) observed (A) at 320.13 MHz with a 5 mm dual-proton/tritium probe, (B) at 533.5 MHz with a 5 mm leak-proof selective excitation proton/tritium probe, (C) at 533.5 MHz with the 5 mm selective excitation proton/tritium cryoprobe, and (D) the same compound at lower activity (11  $\mu\text{Ci}$ ) using the 5 mm selective excitation proton/tritium cryoprobe at 533.5 MHz. (Reproduced with permission from Bloxsidge JP, Garman RN, Gillies DG, Jones JR, and Lu SY (2004) Development and application of a tritium cryo-probe:  $^3\text{H}$  NMR studies at the microcurie (megabecquerel) level of radioactivity. In: Dean DC, Filer CN, and McCarthy KE (eds.) *Synthesis and Application of Isotopically Labelled Compounds*, vol. 8. Chichester: John Wiley; © Wiley.)

catalytic hydrogenation of  $\beta$ -methylstyrene (1) using 5% Pd/C catalyst and  $T_2$  gas the product formed is as expected (2) but there is also clear evidence for the existence of in excess of 30 other isotopomers.



This is made possible through a combination of primary and secondary tritium isotope effects on the two methylene and one methyl groups. What is now required is a combination of developments that allows one to measure the concentration of the isotopomers as the reaction proceeds.

Although the number of research papers reporting the use of  $^3\text{H}$  NMR spectroscopy, especially by comparison to  $^2\text{H}$  and  $^{13}\text{C}$ , is small the range of applications is extensive (and is described in more detail in the book by Evans *et al.*). For example, the measurement of specific activity, optical purity, stereochemistry, solution conformation, 'kinetic' acidities, hydrogen isotope exchange, hydrogen bonding, fractionation factors, and radiolysis.

In addition, there are several reported instances of  $^3\text{H}$  NMR being used to study: reaction mechanisms, chiral methyl applications, biosynthesis, and protein-ligand interactions.

These latter studies stand to benefit most from the emergence of ultrahigh sensitive NMR spectrometers.

## Deuterium NMR Spectroscopy

Initially,  $^2\text{H}$  NMR spectroscopy studies were difficult to perform as the nuclide is inherently less sensitive ( $\sim 100$ -fold) than  $^1\text{H}$  (Table 1). Furthermore, with a nuclear spin of unity the signals were broader and resolution problems were common. In addition, the 'dynamic range' ( $100\% \rightarrow$  natural abundance) is much smaller than is the case for  $^3\text{H}$ . However, by now, the tremendous improvements that have been achieved in spectrometer performance are such that  $^2\text{H}$  is now an attractive nucleus for NMR studies. Thus, spectra can be obtained in a short time interval at the natural abundance level so that for many applications it is not necessary to spend time synthesizing deuteriated compounds. Second, the small differences that exist in the natural abundance levels of different compounds, and of even different hydrogens within the same compound, can be exploited. Finally, the development of magic angle spinning (MAS), which reduces dipolar and other anisotropic interactions, makes it possible to produce  $^2\text{H}$  NMR

spectra in the solid state that, in terms of resolution and sensitivity, rival those from liquids.

In general, the preparation of deuteriated compounds follows the same route as is used in the synthesis of tritiated organic compounds. Specifically labeled compounds are usually prepared by catalytic hydrogenation, aromatic dehalogenation, borodeuteride reduction, and methyl iodide methylation. In addition, there is a range of acid, base, and metal-catalyzed hydrogen-isotope-exchange reactions that provide routes to more generally labeled compounds. However, for preparing labeled macromolecules deuterium has a distinct advantage over tritium – deuteriated water ( $\text{D}_2\text{O}$ ) can be used in large excess thereby ensuring high deuterium incorporation, whereas with tritium, because of its radioactivity, tritiated water (HTO) has to be used in small quantities and at low specific activity ( $5 \text{ Ci ml}^{-1}$  corresponds to  $\sim 0.2\%$  isotope abundance). There are, of course, other biosynthetic routes where, e.g., a suitably tritiated amino acid can be used but on balance  $^2\text{H}$  NMR spectroscopy seems to have greater potential than  $^3\text{H}$  NMR spectroscopy in macromolecules (protein) studies.

## Applications

Hydrogen-deuterium exchange is a widely studied reaction that can be catalyzed by acids, bases, and metals. In the past,  $^1\text{H}$  NMR spectroscopy has been used to follow the decrease in one or more of the  $^1\text{H}$  integrals as a function of time. Now with the improvements in NMR magnet performance and sensitivity it becomes possible to use  $^2\text{H}$  NMR spectroscopy and follow the increase in the concentration of any deuteriated species as a function of time. In this way the kind of reactions that can be investigated can be widened to include, e.g., epoxidations, etc. (Table 2). A further advantage is that it is possible to detect reaction intermediates that are undetectable using other spectroscopic techniques. Ionic liquids, with attractive environmental features, are good solvents for many of the above-mentioned reactions.

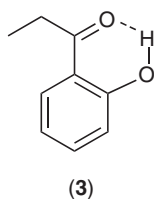
Being able to measure  $^2\text{H}$  NMR spectra of compounds at natural abundance is the basis of an important method – site-specific natural isotope fractionation NMR for determining the authentication (and adulteration) of wines and fruit juices. The deuterium content of all hydrogen-containing sites within a compound can be determined and whilst, for example, the  $^1\text{H}$  NMR spectrum of a compound containing a  $-\text{C}_2\text{H}_5$  group is unable to recognize the very small changes in the  $^2\text{H}$  content, these can be readily achieved from the  $^2\text{H}$  NMR spectrum. The first generation of spectrometers dedicated to this work had rather low sensitivity – a small chemical

**Table 2** Reactions followed by  $^2\text{H}$  NMR spectroscopy

Reaction	Reactant	Product
Epoxidation		
Hydroxylation		
Methylation		
Oxidation		

shift discrimination at the available field and poor stability of the decoupling channel. Now with 9.4 and 11.6 T NMR spectrometers fitted with dedicated probes and state-of-the-art software good signal-to-noise ratios are obtained in a reasonable time and minute variations in natural abundance levels can be quoted with confidence. With these benefits the principle of the technique can be extended to other nuclides such as  $^{13}\text{C}$  and  $^{17}\text{O}$  and also to other problems relating to tobacco, aroma components, and polysaturated fatty acids.

Measurement of the temperature dependence of primary deuterium isotope effects ( $\delta_{\text{XH}} - \delta_{\text{XD}}$ ) on chemical shifts can provide useful information for those compounds containing an intramolecular hydrogen bond such as that depicted in (3). Varying the donor and/or acceptor provides a route to the study of compounds containing localized hydrogen bonds as well as those that display tautomerism. Larger temperature variations are observed for the latter, this being a consequence of the change in the tautomeric equilibrium upon change of temperature.



Enantiomeric differentiation of chiral compounds by NMR spectroscopy usually relies on derivatization

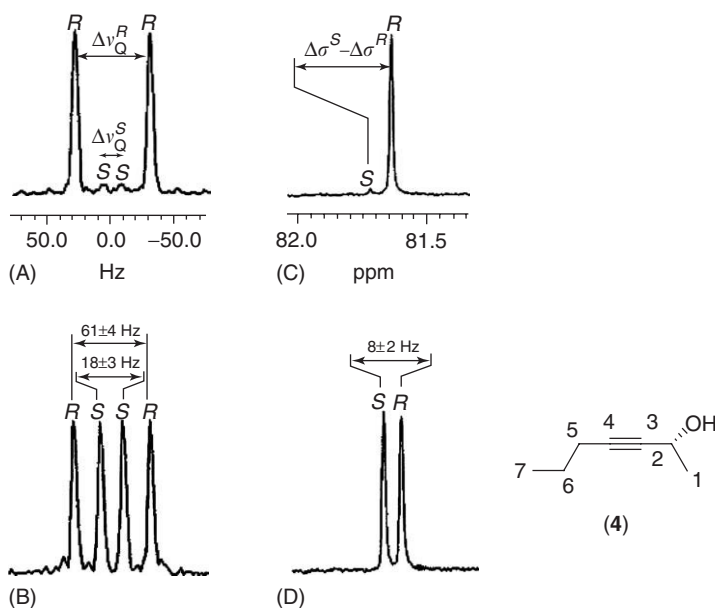
or the preparation of specific complexation agents *in situ*. Not surprisingly, therefore, the task is at its most demanding for chiral hydrocarbons. In a chiral polypeptide liquid-crystalline solvent composed of poly- $\gamma$ -benzyl-L-glutamate in chloroform it has been shown that the geometry of solutes plays an important role in the differential orientation of two enantiomers. Whilst the differential ordering effect is likely to be small in the case of *R* and *S* isomers if it could be coupled to a very sensitive order-dependent NMR interaction, such as the quadrupolar interaction for deuterium ( $I=1$ ) nuclei the basis for a differentiation exists. The large magnitude of the deuterium quadrupolar coupling constant for a  $\text{C}-^2\text{H}$  bond is  $\sim 170$  kHz, thereby ensuring sufficient difference in the residual deuterium quadrupolar splittings. This, in conjunction with the ability to obtain high-class  $^2\text{H}$  NMR spectra at natural abundance level, means that this difficult task has now been accomplished. All possible deuteriated sites within the molecule can be simultaneously probed and this without the need to synthesize any deuteriated compound. This important development can now be applied to a wide range of saturated and unsaturated chiral building blocks (Figure 2).

High-resolution NMR spectroscopy of solids only became possible with the development of MAS in which the sample is spun in a rotor at high speed (2–15 kHz), creating conditions that resemble the isotropic tumbling of small molecules in solution. At the magic angle of  $54.7^\circ$  the orientational dependence of the magnetic interactions is averaged to zero and NMR spectra can be obtained, which in terms of resolution and sensitivity rival those from liquids. The spinning speeds themselves are related to the nuclear spin coupling frequencies; consequently, by reducing the speed spinning side-bands appear in the spectra. These then provide details of dipolar couplings from which inter- and intramolecular distances can be obtained.

For nuclei such as  $^2\text{H}$  with a spin of unity the strong quadrupolar interaction ( $\sim 100$  kHz) dominates over the dipolar and chemical shift interactions and helps to simplify the analysis of the lineshapes. Selective deuterium labeling of ligands and/or proteins therefore offers considerable potential, not only for molecular distance information but also for studying the dynamics of the interactions.

## Recent/Future Developments

The importance of NMR spectroscopy has been recognized in many ways, not least by the award of various Nobel prizes over the years to Rabi, Bloch, Purcell, Ernst, Wüthrich, Lauterbur, and Mansfield. What will the future hold for deuterium and tritium



**Figure 2** (A) Natural abundance  $^2\text{H}\{-^1\text{H}\}$  signals (proton-decoupled deuterium spectrum) associated with deuterons of the methyl group 1 and (C)  $^{13}\text{C}\{-^1\text{H}\}$  signals (proton-decoupled carbon-13 spectrum) of C-4 recorded for  $R$ -(+)-**4**; (B) and (D) are obtained using the same experimental conditions as in (A) and (C), but using (±)-**4**. (Reprinted with permission from Parenty A, Campagne J-M, Aroulanda C, and Lesot P (2002) *Organic Letters* 4: 1663–1666; © American Chemical Society.)

NMR spectroscopy? As far as the latter is concerned we are well away from being able to make measurements at natural abundance. Consequently, there is much room for improvement. Some 12 years have elapsed since the first ultrahigh field NMR magnets became commercially available and prior to this a whole new magnet technology had to be developed. The stringent demands on field strength, homogeneity, and stability require the construction of long lengths (typically 100 km) of high-quality superconducting wires. The next generation system with field strength of 23.5 T, corresponding to 1000 MHz or 1 GHz spectrometer, will require that the present  $\text{Nb}_3\text{Sn}$  alloys be replaced by the emerging high-temperature superconductors, e.g.,  $\text{BiCoSrCrO}$ , which themselves will have to have the right physical properties for the construction of the high-resolution NMR magnets.

In the short time interval (5 years) since cryoprobes were commercialized they have become an indispensable accessory for NMR. Initially, they were available on 500 MHz instruments but now they have been installed on 700 and 800 MHz spectrometers. The more recent cryostats, thanks to ingenious design, are now less sensitive to disturbances such as changes in room temperature and helium evaporation rates. All of these factors will ensure considerable improvements in the performance of these ultrastabilized magnet systems.

There is at least one other area that may be of benefit to the tritium NMR spectroscopist. In recent years there has been a trend toward miniaturization

in chemical synthesis and screening and this places an increased requirement on analytical sensitivity. Such developments have relevance to tritium chemistry as work in this area is usually undertaken at the milligram to microgram levels. Coupling techniques such as liquid chromatography–nuclear magnetic resonance (LC–NMR) have become very useful in analytical chemistry but there is one problem that relates to the maximum amount of sample (typically a few milligrams) that can be efficiently submitted for chromatographic separation. This has now been solved through the introduction of solid-phase extraction at the interface between the chromatography instrument and the NMR spectrometer. Now the entire chromatographic peak can be concentrated in the NMR detection volume. Furthermore miniaturization of the chromatographic equipment (use of micro and capillary LC separation) in conjunction with miniaturization of the radiofrequency coils and reduction of the NMR detection volume leads to considerable improvements in sensitivity. NMR spectra are usually acquired under static or stopped-flow conditions but with the necessary software multiple eluted peaks in a chromatogram can be analyzed with just one sample injection. Such applications, when transferred to the tritium area, will greatly increase its attractions as a tracer. However, there will always be the need to synthesize tritiated compounds as it is unrealistic to expect the above-mentioned improvements to lead to a situation in which  $^3\text{H}$  NMR spectra can be obtained at natural abundance level.



As far as  $^2\text{H}$  NMR spectroscopy is concerned, the application of a number of new solid-state methods is in its infancy. Molecular biology is providing opportunities for producing new proteins; ingenious biosynthetic methods can be devised for producing deuterium labeled versions and one or more of the NMR methods can be used to increase our knowledge of protein structure and dynamics and the way in which they are linked to function – one of the central problems of the life sciences. Recent work on membrane proteins, which perform essential roles in the cells of all living organisms, be it as surface receptors or transporters that mediate the uptake of nutrients or the expulsion of waste products/toxins, highlight the potential.

All of the above advances have come about without researchers addressing the important issue of what can be done to isolate and identify ultraweak signals embedded in complex ‘background noise’. Now this may be about to change, as following pioneering work in quantum physics, active signal processing technology has been developed that can increase the sensitivity of standard detection systems by up to 10 000 times. The versatility of the technology is likely to create opportunities to develop products in a number of areas, of which NMR spectroscopy is likely to be one.

**See also:** **Extraction:** Solid-Phase Extraction. **Food and Nutritional Analysis:** Oils and Fats; Fruits and Fruit Products. **Lab-on-a-Chip Technologies.** **Liquid Chromatography:** Liquid Chromatography–Nuclear Magnetic Resonance Spectrometry. **Nuclear Magnetic Resonance Spectroscopy:** Overview; Principles; Instrumentation. **Nuclear Magnetic Resonance Spectroscopy Applications:** Food. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Solid-State. **Peptides.** **Radiochemical Methods:** Radiotracers; Pharmaceutical Applications.

## Further Reading

Bryce DL, Bernard GM, Gee M, *et al.* (2001) Practical aspects of modern routine solid-state multinuclear magnetic resonance spectroscopy: one-dimensional experiments. *Canadian Journal of Analytical Sciences and Spectroscopy* 46: 46–82.

- Cheng HN and English AD (eds.) (2003) *NMR Spectroscopy of Polymers in Solution and in the Solid State*. ACS Symposium Series 834. Washington DC: American Chemical Society.
- Dong RY (2002) Relaxation and dynamics of molecules in the liquid crystal phases. *Progress in Nuclear Magnetic Resonance Spectroscopy* 41: 115–151.
- Elander N, Jones JR, Lu SY, and Stone-Elander S (2000) Microwave-enhanced radiochemistry. *Chemical Society Reviews* 29: 239–249.
- Evans EA, Warrell DC, Elvidge JA, and Jones JR (1985) *Handbook of Tritium NMR Spectroscopy and Applications*. Chichester: Wiley.
- Lesot P, Sarfati M, and Courtieu J (2003) Natural abundance deuterium NMR spectroscopy in polypeptide liquid crystals as a new and incisive means for the enantiodifferentiation of chiral hydrocarbons. *Chemistry – A European Journal* 9: 1724–1745.
- Lian LY and Middleton DA (2001) Labeling approaches for protein structural studies by solution-state and solid-state NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* 39: 171–190.
- Loupy A (ed.) (2002) *Microwaves in Organic Synthesis*. Weinheim: Wiley-VCH.
- MacDonald PM (1997) Deuterium NMR and the topography of surface electrostatic charge. *Accounts of Chemical Research* 30: 196–203.
- Martin GJ and Martin ML (1991) Deuterium NMR in the study of site-specific natural isotope fractionation (SNIF-NMR). In: Diehl P, Fluck E, Gunther H, Kosfeld R, and Seelig J (eds.) *NMR – Basic Principles and Progress* 23 – *Deuterium and Shift Calculation*, pp. 1–61. Berlin: Springer.
- Saljoughian M (2002) Synthetic tritium labeling: reagents and methodology. *Synthesis – Stuttgart* 13: 1781–1801.
- Siminovitch DL (1998) Solid-state NMR studies of proteins: the view from static  $^2\text{H}$  NMR experiments. *Biochemistry and Cell Biology* 76: 411–422.
- Watts A (1998) Solid-state NMR approaches for studying the interaction of peptides and proteins with membranes. *Biochimica Biophysica Acta* 1376: 297–318.
- Wemmer DE and Williams PG (1994) Use of nuclear-magnetic-resonance in probing ligands in macromolecule interaction. In: James TL and Oppenheimer NJ (eds.) *Methods in Enzymology* 239, *Nuclear Magnetic Resonance Part C*, pp. 739–767. San Diego: Academic Press.

## Carbon-13

**S Simova**, Bulgarian Academy of Sciences, Sofia, Bulgaria

## Introduction

Carbon-13 nuclear magnetic resonance ( $^{13}\text{C}$  NMR) undoubtedly represents one of the most significant

As far as  $^2\text{H}$  NMR spectroscopy is concerned, the application of a number of new solid-state methods is in its infancy. Molecular biology is providing opportunities for producing new proteins; ingenious biosynthetic methods can be devised for producing deuterium labeled versions and one or more of the NMR methods can be used to increase our knowledge of protein structure and dynamics and the way in which they are linked to function – one of the central problems of the life sciences. Recent work on membrane proteins, which perform essential roles in the cells of all living organisms, be it as surface receptors or transporters that mediate the uptake of nutrients or the expulsion of waste products/toxins, highlight the potential.

All of the above advances have come about without researchers addressing the important issue of what can be done to isolate and identify ultraweak signals embedded in complex ‘background noise’. Now this may be about to change, as following pioneering work in quantum physics, active signal processing technology has been developed that can increase the sensitivity of standard detection systems by up to 10 000 times. The versatility of the technology is likely to create opportunities to develop products in a number of areas, of which NMR spectroscopy is likely to be one.

**See also:** **Extraction:** Solid-Phase Extraction. **Food and Nutritional Analysis:** Oils and Fats; Fruits and Fruit Products. **Lab-on-a-Chip Technologies.** **Liquid Chromatography:** Liquid Chromatography–Nuclear Magnetic Resonance Spectrometry. **Nuclear Magnetic Resonance Spectroscopy:** Overview; Principles; Instrumentation. **Nuclear Magnetic Resonance Spectroscopy Applications:** Food. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Solid-State. **Peptides.** **Radiochemical Methods:** Radiotracers; Pharmaceutical Applications.

## Further Reading

Bryce DL, Bernard GM, Gee M, *et al.* (2001) Practical aspects of modern routine solid-state multinuclear magnetic resonance spectroscopy: one-dimensional experiments. *Canadian Journal of Analytical Sciences and Spectroscopy* 46: 46–82.

- Cheng HN and English AD (eds.) (2003) *NMR Spectroscopy of Polymers in Solution and in the Solid State*. ACS Symposium Series 834. Washington DC: American Chemical Society.
- Dong RY (2002) Relaxation and dynamics of molecules in the liquid crystal phases. *Progress in Nuclear Magnetic Resonance Spectroscopy* 41: 115–151.
- Elander N, Jones JR, Lu SY, and Stone-Elander S (2000) Microwave-enhanced radiochemistry. *Chemical Society Reviews* 29: 239–249.
- Evans EA, Warrell DC, Elvidge JA, and Jones JR (1985) *Handbook of Tritium NMR Spectroscopy and Applications*. Chichester: Wiley.
- Lesot P, Sarfati M, and Courtieu J (2003) Natural abundance deuterium NMR spectroscopy in polypeptide liquid crystals as a new and incisive means for the enantiodifferentiation of chiral hydrocarbons. *Chemistry – A European Journal* 9: 1724–1745.
- Lian LY and Middleton DA (2001) Labeling approaches for protein structural studies by solution-state and solid-state NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* 39: 171–190.
- Loupy A (ed.) (2002) *Microwaves in Organic Synthesis*. Weinheim: Wiley-VCH.
- MacDonald PM (1997) Deuterium NMR and the topography of surface electrostatic charge. *Accounts of Chemical Research* 30: 196–203.
- Martin GJ and Martin ML (1991) Deuterium NMR in the study of site-specific natural isotope fractionation (SNIF-NMR). In: Diehl P, Fluck E, Gunther H, Kosfeld R, and Seelig J (eds.) *NMR – Basic Principles and Progress* 23 – *Deuterium and Shift Calculation*, pp. 1–61. Berlin: Springer.
- Saljoughian M (2002) Synthetic tritium labeling: reagents and methodology. *Synthesis – Stuttgart* 13: 1781–1801.
- Siminovitch DL (1998) Solid-state NMR studies of proteins: the view from static  $^2\text{H}$  NMR experiments. *Biochemistry and Cell Biology* 76: 411–422.
- Watts A (1998) Solid-state NMR approaches for studying the interaction of peptides and proteins with membranes. *Biochimica Biophysica Acta* 1376: 297–318.
- Wemmer DE and Williams PG (1994) Use of nuclear-magnetic-resonance in probing ligands in macromolecule interaction. In: James TL and Oppenheimer NJ (eds.) *Methods in Enzymology* 239, *Nuclear Magnetic Resonance Part C*, pp. 739–767. San Diego: Academic Press.

## Carbon-13

**S Simova**, Bulgarian Academy of Sciences, Sofia, Bulgaria

## Introduction

Carbon-13 nuclear magnetic resonance ( $^{13}\text{C}$  NMR) undoubtedly represents one of the most significant

developments in organic structural analysis. Although its potential as an analytical tool was recognized soon after the first experiments had been carried out by Lauterbur in 1957, on molecules containing the isotope in natural abundance, it took more than a decade until the technique could be applied routinely. Today,  $^{13}\text{C}$  NMR spectroscopy has a very wide application profile and is mandatory for studying many structure, conformational, dynamic, and mechanistic problems.

Early  $^{13}\text{C}$  spectroscopists were mostly struggling with problems associated in some way with spectrometer sensitivity. NMR is an insensitive technique when compared with optical or mass spectroscopy due to the small energy difference between the ground and excited states and the long lifetimes of the excited states (typically in the order of  $10^{-3}$ – $10^3$  s). Additionally, for the carbon atoms the low natural abundance (1.108%) and the approximately four times smaller gyromagnetic ratio ( $6.728286 \times 10^7 \text{ rad s}^{-1} \text{ T}^{-1}$ ) lead to an overall sensitivity of only 1/5681 of that of the protons. The intrinsic sensitivity problems have been overcome only after the introduction of Fourier transformation (FT) methods, proton noise decoupling (Pnd), and signal averaging techniques. Additional signal enhancement could be achieved by the use of nuclear Overhauser enhancement (NOE) or polarization transfer (PT) in liquids and cross-polarization (CP) transfer in solids. Further considerable sensitivity gain is obtained by increasing the magnetic field strength. By the end of 2003, 21 T NMR spectrometers are accessible, corresponding to 226 MHz Larmor frequency for the carbon atom. Many other nuclei are more sensitive than  $^{13}\text{C}$ , e.g.,  $^{14}\text{N}$ ,  $^{23}\text{Na}$ ,  $^{31}\text{P}$ ,  $^{19}\text{F}$ , and  $^{29}\text{Si}$ ; nevertheless,  $^{13}\text{C}$  NMR is second in importance to  $^1\text{H}$  NMR and is often the only approach suitable for complex organic molecules, especially for polymers and solids. The special features that make  $^{13}\text{C}$  NMR an attractive alternative to  $^1\text{H}$  NMR or a valuable extension for the solution of structural and other analytical problems include:

- The chemical shift scale spans usually 250 ppm ( $\sim 20$  times that for protons), which allows small differences in electronic and/or steric environments around the carbon nuclei to be distinguished.
- The  $^{13}\text{C}$  NMR spectrum gives information about the carbon skeleton of the molecule, including functional groups that do not contain protons, e.g., carbonyl, cyano, etc., groups.
- The  $^{13}\text{C}$  chemical shifts are less sensitive to medium and anisotropy effects.
- $^{13}\text{C}$  peak assignment is simpler than for  $^1\text{H}$  NMR due to the lack of couplings to nearby carbons in

the broadband-decoupled spectra. Heteronuclear coupling constants could be easily distinguished.

- Determination of the number of directly bound hydrogen atoms, called multiplicity assignment, is easy through application of editing pulse sequences, e.g., distortionless enhancement by polarization transfer (DEPT) or attached proton test (APT).
- Information about molecular shape and motion can be derived from experimental spin–lattice relaxation times and measured NOE factors.
- Detailed structure elucidation is greatly facilitated through the application of various multidimensional correlation techniques for establishing through bond and through space connectivities between carbon, hydrogen, and other nuclei.
- $^{13}\text{C}$  labeling allows determination of reaction pathways in biosynthesis as well as of carbon coupling constants.
- Many analytical applications are known for solid as well as for liquid samples.

The term  $^{13}\text{C}$  NMR needs some clarification. Until  $\sim 1980$  it simply meant that the NMR spectrometer was equipped to detect  $^{13}\text{C}$  resonance lines and to ascertain  $^{13}\text{C}$  NMR parameters. An indirect approach for the latter task was  $^1\text{H}$  detection combined with radio frequency (RF) irradiation at  $^{13}\text{C}$  resonance lines. Such  $^1\text{H}$ – $^{13}\text{C}$  double-resonance studies were, however, of limited use for the solution of structural problems. Today, NMR instrumentation provides three, four, or more RF irradiation channels with interlocked frequencies, e.g., for  $^1\text{H}$ ,  $^2\text{H}$ ,  $^{13}\text{C}$ , and ( $^{15}\text{N}$  or  $^{31}\text{P}$ ), computer control for complex multipulse sequences and software for multidimensional data processing. The heteroatom resonance frequencies ( $X = ^{13}\text{C}$ ,  $^{15}\text{N}$ , etc.) could then be derived indirectly from the appropriate multidimensional spectrum. For  $^{13}\text{C}$  this produces a wealth of structural information without acquiring a single  $^{13}\text{C}$  free induction decay (FID) signal. Decision on whether to use direct or indirect detection depends on different factors, e.g., the problem under investigation, the amount of substance available, the spectrometer configuration, etc. Traditionally, direct  $^{13}\text{C}$  observation is most often used. It is this direct observation that is meant by the term  $^{13}\text{C}$  NMR.

## Experimental Requirements for $^{13}\text{C}$ NMR

### Spectrometer

All commercially produced FT-NMR spectrometers, irrespective of the magnetic field strength (25–226 MHz  $^{13}\text{C}$  at 100–900 MHz for  $^1\text{H}$ ), generally

meet the basic requirements for one- and multidimensional  $^{13}\text{C}$  NMR in liquids. The main interest in higher magnetic fields for  $^{13}\text{C}$  NMR spectra lies in the sensitivity gain that theoretically obeys  $B_0^{3/2}$  dependence. However, the actual improvement at very high fields may be less than expected for certain carbon atoms, e.g., carbonyl because of nonuniform excitation, decrease of  $T_2$  relaxation times, etc. A large variety of probes could be used for  $^{13}\text{C}$  detection – single, dual, and broadband probeheads, QNP probes (e.g.,  $^1\text{H}$ – $^{19}\text{F}$ – $^{31}\text{P}$ – $^{13}\text{C}$ ) with computer controlled switching among four nuclei at high sensitivity, etc. Sample tubes of 5 and 10 mm outer diameter are mainly used for direct detection. The best  $^{13}\text{C}$  sensitivity is reached with 10 mm probes with inner coil optimized for the  $^{13}\text{C}$  resonance frequency. Their use results in approximately threefold increase in signal-to-noise ratio (S/N) relative to the 5 mm probes, at the expense of somewhat lower resolution. However, for economic and practical reasons, 5 mm probes are most often used. For many experiments actively shielded Z-gradient coils are used. It should be, however, noted that most often they are optimized for the  $^1\text{H}$  resonance frequency and have a low sensitivity for direct  $^{13}\text{C}$  detection. In recent years, the so-called cold- or cryoprobes have been designed by the use of cryogenically cooled RF coil and preamplifier with up to four times higher sensitivity as compared to conventional room temperature coil probes. Multinuclear magic angle spinning (MAS) probes for both CP and high power decoupling experiments are applied for solid-state  $^{13}\text{C}$  NMR. Depending on the particular hardware a variety of different experiments could be performed. Most often used in the practice are 1D (dimensional) or 2D multipulse sequences, comprising proton and carbon pulses. The NMR spectrometers possess central

software packages for spectrometer control, data acquisition, and data processing. The latter could also be achieved on a separate computer, usually connected with the spectrometer through the network.

### Sample

Sample preparation for  $^{13}\text{C}$  NMR is very similar to that for  $^1\text{H}$  NMR. The standard sample volume is 0.4–0.6 ml for 5 mm, ~2.5 ml for 10 mm probes but may be as little as 0.04 ml in optional nanoprobe. Owing to the low sensitivity of  $^{13}\text{C}$ , the major requirement is to obtain as concentrated solution as possible in order to obtain a spectrum in as short time as possible. Obviously, if sample amount is unrestricted or solubility is low the use of larger sample tubes and correspondingly sample volumes becomes profitable. The minimum concentration requirements for a given sample and experiment depend mainly on the magnetic field strength and on the quality of the probe. For a relative molecular mass of less than 1000, 1D  $^{13}\text{C}$  NMR is routinely feasible for solutions at  $\sim 0.01 \text{ mol l}^{-1}$  in an overnight experiment. Approximately 1 h is needed for the measurement of the same sample on a 500 MHz spectrometer equipped with a cryogenically cooled probe. The suitability of the solvent to be used is dictated by the sample solubility at the measurement temperature. Recommended solvents are  $\text{CDCl}_3$ ,  $\text{DMSO-d}_6$ , and  $\text{D}_2\text{O}$ . A list of common solvents used in NMR work with their chemical shifts, multiplicities, dielectric constants, and liquid ranges are given in Table 1.

For  $^{13}\text{C}$  detection nondeuterated solvents could alternatively be used, especially for rare or very expensive solvents. In this case, a small capillary containing a deuterated substance, e.g., acetone- $\text{d}_6$ , is put into the sample tube. It provides the lock substance

**Table 1** Properties of some important solvents for NMR<sup>a</sup>

Solvent	Chemical shift $\delta_{\text{C}}$ (ppm)	Multiplicity due to $^1J_{\text{CD}}$	Liquid temperature range ( $^{\circ}\text{C}$ )	Dielectric constant
$\text{CCl}_4$	96.0	1	– 23 to 77	2.2
$\text{CS}_2$	192.3	1	– 112 to 46	2.6
$\text{CDCl}_3$	77.0	3	– 64 to 61	4.8
$\text{CD}_2\text{Cl}_2$	53.6	5	– 95 to 40	8.9
$\text{CDCl}_2\text{CDCl}_2$	75.5	3	– 44 to 146	8.2
THF- $\text{d}_8$	25.8, 67.8	5	– 108 to 66	7.6
Benzene- $\text{d}_6$	128.4	3	6 to 80	2.3
Pyridine- $\text{d}_5$	123.9, 135.9, 150.2	3	– 42 to 115	12.4
Acetone- $\text{d}_6$	30.0, 206.0	7	– 95 to 56	20.7
$\text{DMSO-d}_6$	39.6	7	19 to 189	46.7
Methanol- $\text{d}_4$	49.3	7	– 98 to 65	32.7
$\text{D}_2\text{O}$	–	–	0 to 100	78.5
TFA- $\text{d}$	114.5, 161.5 <sup>b</sup>	–	– 15 to 72	8.6

<sup>a</sup>The physical constants of the protonated solvents are given.

<sup>b</sup> $^1J_{\text{CF}} = 283 \text{ Hz}$ ;  $^2J_{\text{CF}} = 43 \text{ Hz}$ .

and assures the field-frequency stabilization. Usually, a small amount (0.03–0.05%) of tetramethylsilane (TMS) or of the less volatile hexamethyldisiloxane (HMDS, 2 ppm downfield from TMS), when working at elevated temperature, is added to the samples as reference. Often the signal of the deuterated solvent is used for referencing purposes (Table 1).

## Basic Characteristics of $^{13}\text{C}$ NMR

### Natural Abundance $^{13}\text{C}$ Spectra

Samples with natural  $^{13}\text{C}$  abundance are the usual case in routine  $^{13}\text{C}$  application. For a molecule with  $N$  carbon atoms, the  $^{13}\text{C}$  NMR spectrum usually appears as a mixture of  $N$  different spin systems  $\text{CH}_d\text{H}_l\text{H}_m, \dots, \text{H}_r$ , where the subscripts  $d, l, m, \dots, r$  refer to protons directly attached to, or separated by two or more bonds from, the carbon, respectively. Accordingly, the  $^{13}\text{C}$  spectrum is a superposition of  $N$  subspectra centered at the chemical shift  $\delta_i$  of the carbon nucleus  $i$  with  $i = 1, \dots, N$ . In case of molecular symmetry, two or more spin systems are alike and the intensities of the corresponding subspectra increase by a factor of two or more. In single-scan FT spectra, starting from thermal equilibrium (no NOE), all subspectra have the same overall intensity because the  $N$  spin systems are equally abundant. In multiscan FT spectra with coherent averaging (accumulation), intensities are generally different because the various carbons relax differently toward thermal equilibrium so that the  $z$ -magnetizations differ from carbon to carbon prior to each pulse. This is particularly pronounced for quaternary carbons that have comparatively long  $T_1$  relaxation times so that their signals are less intense or even lost in the noise compared to proton-connected ones.

### Coupled $^{13}\text{C}$ NMR Spectra

Coupled  $^{13}\text{C}$  NMR spectra are usually very complex and difficult to interpret. For a simple one-pulse experiment, each subspectrum is a symmetric multiplet centered at the chemical shift of the carbon nucleus (see Figure 1A). The multiplet structure is governed by the large one-bond  $^1J_{\text{CH}}$  value (100–320 Hz total range) and a number of small longer-range  $J_{\text{CH}}$  values mainly for  $^{13}\text{C}$ – $^{12}\text{C}$ –H and  $^{13}\text{C}$ – $^{12}\text{C}$ – $^{12}\text{C}$ –H fragments. Primary ( $\text{CH}_3$ ), secondary ( $\text{CH}_2$ ), and tertiary (CH) atoms can be distinguished by their multiplicities, since according to the first order- $(n+1)$  splitting rule they give rise to wide spaced quartets ( $n=3$ ), triplets ( $n=2$ ), and doublets ( $n=1$ ). Quaternary (C) carbons present singlets ( $n=0$ ) with low intensity. Each multiplet line is further split or broadened due to the longer-range couplings. Many

of these multiplets represent higher-order spectra despite which they often appear deceptively simple. Then a full spin-quantum mechanical analysis (e.g., with a computer program like LAOCOON) must be performed in order to derive the true  $\delta$  and  $J$  values. In many cases not only the  $^{13}\text{C}$  multiplet, but also the  $^1\text{H}$  NMR  $^{13}\text{C}$  satellite resonances must be analyzed. Further complications arise from the severe overlap of multiplets even in simple organic molecules.

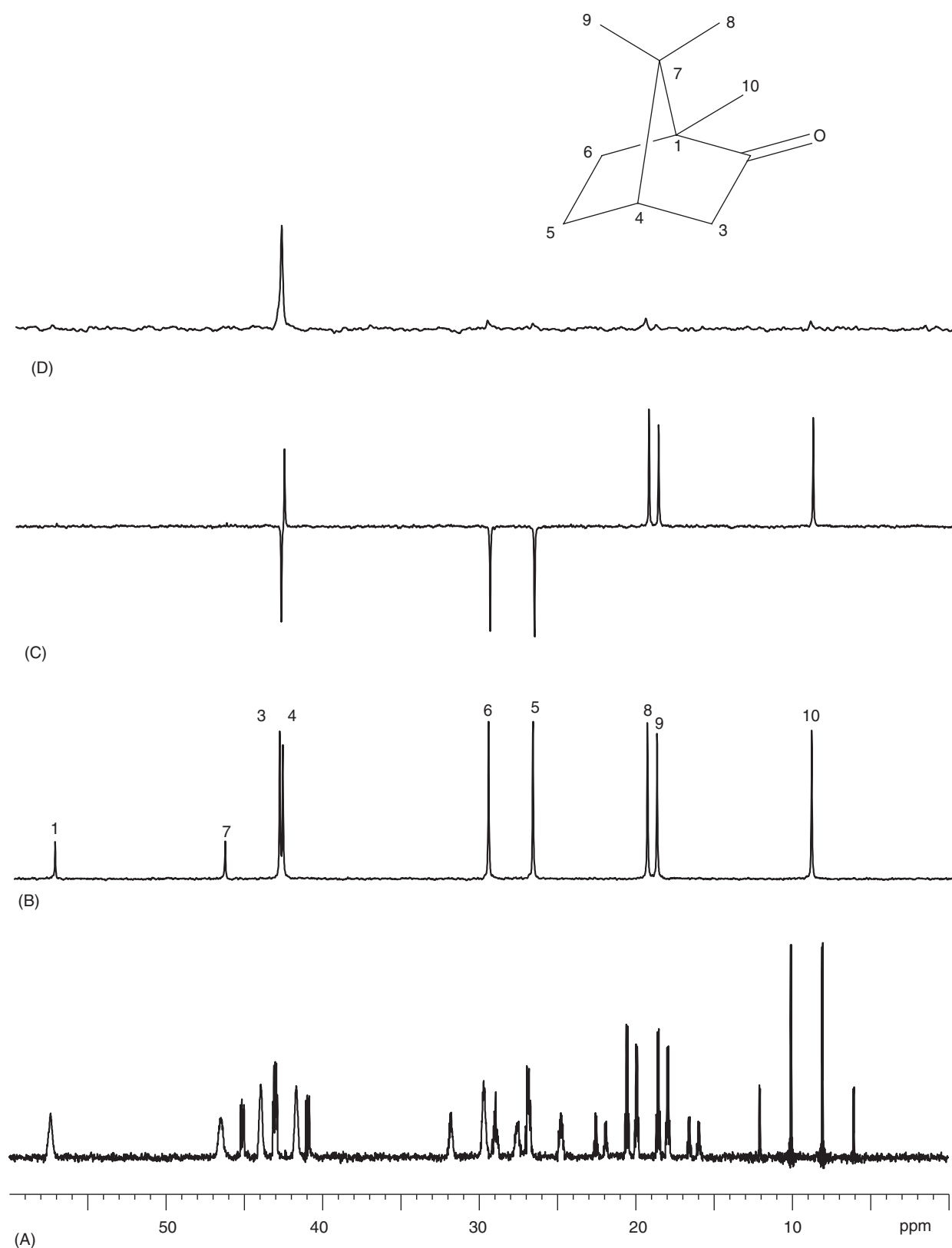
Another serious drawback of the  $^{13}\text{C}$ -coupled spectra is the intrinsic low S/N ratio. The splitting in many components leads to a substantial decrease of the maximum line heights as compared with uncoupled lines. This leads to a dramatic increase in measuring time of at least a factor of 10, but usually much more. The gated-decoupling technique is most often used for direct measurement of coupled  $^{13}\text{C}$  NMR spectra since applying it up to a threefold gain in carbon sensitivity from the build-up of NOE could be achieved.

In summary, coupled  $^{13}\text{C}$  NMR spectra are only occasionally directly measured. The tendency is to measure decoupled spectra and to retrieve the coupling information by well-designed more or less sophisticated 1D or 2D pulse sequences in a more efficient and timesaving way.

### $^1\text{H}$ -Decoupled $^{13}\text{C}$ Spectra

A considerable simplification of the appearance of the  $^{13}\text{C}$  spectra (see Figure 1B), is achieved by application of broadband (BB) decoupling – the multitude of lines concentrate to a number of single sharp peaks with high intensities, each at the chemical shift  $\delta$  of the chemically nonequivalent  $^{13}\text{C}$  atoms. The technique effectively eliminates all coupling constants between the proton and the carbon nuclei. The spectra are usually denoted as  $^{13}\text{C}\text{--}\{^1\text{H}\}$ , the nucleus outside the brackets being observed while that inside the brackets is simultaneously decoupled. Apart from the increase in S/N due to the collapse of the multiplet structure, proton decoupling additionally adds to the signal intensity due to the nuclear Overhauser effect (NOE), which in favorable cases can produce an almost threefold increase over and above that produced by the collapse of the fine structure. The analysis of these spectra is relatively easy. Determination of the values of the chemical shifts only is in striking contrast to the embarrassment by the examination of the complex structure of  $^1\text{H}$ -coupled spectra. The high sensitivity and the easy interpretation are the two main reasons why decoupled  $^{13}\text{C}$  NMR spectra are most widely used. Furthermore, signal separation in  $^{13}\text{C}$  NMR is much better than in  $^1\text{H}$  NMR owing to the considerably





**Figure 1**  $^{13}\text{C}$  NMR spectra of camphor – aliphatic region with assignment (the carbonyl resonance at 220 ppm is not shown): (A)  $^1\text{H}$ -coupled  $^{13}\text{C}$  NMR spectrum; (B)  $^1\text{H}$  BB-decoupled  $^{13}\text{C}$  NMR spectrum: nine aliphatic carbons can be readily determined from the observed nine single-line resonances; (C) DEPT-135: CH<sub>3</sub> and CH pointing up, CH<sub>2</sub> – down, quaternary are missing; (D) DEPT-90: only CH signals with significant intensity.

larger  $^{13}\text{C}$  shift range, as accidental signal coincidence is seen less often.

Different experimental schemes could be used for heteronuclear BB decoupling. Older instruments used to apply Pnd – the decoupler is set in the center of the proton region and modulated using a ‘noise generator’ with a bandwidth wide enough to cover the complete proton region (e.g., 1000 Hz at 2.4 T). This is equivalent to simultaneous irradiation of every proton frequency. Composite-pulse decoupling (CPD) is the method of choice in advanced instruments since it requires less power and sample heating could be kept to a minimum. Important characteristics of the CPD sequences are the bandwidth – the range of proton offsets that can be decoupled for a given  $^1\text{H}$  carrier frequency – and the CPD quality, indicated by the presence or absence of weak side band signals, residual broadening, or fine structure in the decoupled  $^{13}\text{C}$  peaks. WALTZ-16 scheme (wide-band alternating-phase low-power technique for zero-residual splitting) with composite pulse element  $90_x^\circ\text{--}180_x^\circ\text{--}270_x^\circ$  is commonly employed for proton-decoupled  $^{13}\text{C}$  spectra because of its optimum performance – sufficient bandwidth and good quality. For decoupling of the full carbon range,  $^1\text{H}\text{--}\{^{13}\text{C}\}$ , as, e.g., needed in inverse 2D methods, the GARP sequence (globally optimized alternating-phase rectangular pulses) is used because of its large bandwidth, with some compromise of the decoupling quality. For more special applications many more different CPD sequences have been designed.

### Multiplicity-Edited $^{13}\text{C}$ NMR Spectra

Against the obvious benefits in decoupled spectra – simplicity and sensitivity – is the loss of all coupling information that is often needed for assignment purposes. Especially valuable is the multiplicity of a carbon atom, indicating the number  $n$  of its directly connected protons. The most efficient methods to recover this coupling information rely on encoding of the carbon multiplicity in the signal sign and intensity. Two pulse sequences and their modifications are most often applied. Historically first was the so-called APT, based on the use of a  $J$ -modulated heteronuclear spin-echo. More advanced is the DEPT sequence, characterized by editing properties with negligible dependence on the  $^1J_{\text{CH}}$  value and signal enhancement by PT. The observed  $^{13}\text{C}$  signal intensities are proportional to  $\cos^n \theta$ , where  $n$  is the number of the attached protons and  $\theta$  the third proton pulse in the DEPT sequence. The collection of three spectra with  $\theta = 45^\circ$ ,  $90^\circ$ , and  $135^\circ$  is the starting point for the determination of the carbon multiplicities. Depending on their multiplicities the carbon atoms

**Table 2** Signs of multiplet resonances in DEPT spectra with different  $\theta$

Carbon type	DEPT-45	DEPT-90	DEPT-135
$\text{CH}_3$	+	0	+
$\text{CH}_2$	+	0	–
CH	+	+	+
C	0	0	0

produce positive, negative, or zero signal intensities for these angles, the type of the responses is summarized in Table 2. In principle, decomposition into  $\text{CH}$ ,  $\text{CH}_2$ , and  $\text{CH}_3$  subspectra can be produced by a linear combination of the three spectra. More often, however, the multiplicities are determined mainly on base of the signal signs in the DEPT-135 spectrum (see Figure 1C). Methyl and methine carbons show positive signals, methylene atoms give negative signals, whereas quaternary carbons do not appear.  $\text{CH}_3$  and CH signals can be usually distinguished by chemical shift arguments, but sure proof is obtained from the DEPT-90 spectrum (see Figure 1D), containing signals with considerable intensity only for the methine atoms.

### NOE Enhancement

Under conditions of proton decoupling the enhancement of the carbon signals is normally greater than it would be expected from the collapse of the multiplet structure into a single line due to perturbation of the spin state populations from their equilibrium states. This additional enhancement is known as the nuclear Overhauser effect or the NOE factor  $\eta$ . It is quantitatively expressed by  $I = (1 + \eta)I_0$ , where  $I$  and  $I_0$  are the total  $^{13}\text{C}$  line intensities with and without  $^1\text{H}$  decoupling. The intensity gain  $\eta$  approaches one-half the quotient of the  $^1\text{H}$  and  $^{13}\text{C}$  magnetogyric ratios,  $\gamma_{\text{H}}/2\gamma_{\text{C}} = 1.98$  for isotropically reorienting small organic molecules in nonviscous solvents under C–H internuclear dipole–dipole relaxation. It should be noted that  $\eta$  could be quite different for each individual carbon atom and is an additional reason for their different intensities in decoupled spectra. As a relaxational phenomenon the NOE builds up and disappears with a time constant equal to the spin-lattice relaxation time  $T_1$ . This behavior is opposite to the process of decoupling that disappears instantly when the decoupler is turned off and provides the basis of the so-called gated decoupling technique. By using switching of the decoupler power the two effects – coupling and NOE – can be separately observed. Two schemes are more widely used, which are:

- Gated decoupling: coupled spectra with NOE for sensitivity enhancement – with the decoupler turned on only during the relaxation time.

- Inversed gated decoupling: decoupled spectra without the NOE for determination of  $\eta$  or quantitative analysis – with the decoupler turned on only during the acquisition time.

## Spectrum–Structure Relationship

### $^{13}\text{C}$ Chemical Shifts

The ‘normal’ range of  $^{13}\text{C}$  chemical shifts is  $-30$  to  $230\delta$ , so that the spread of chemical shifts is  $\sim 20$  times that of the proton. Figure 2 shows typical chemical shift ranges for a number of common carbon environments. There is an overall similarity to the proton chemical shifts. What is similar for both nuclei is the pronounced dependence on the hybridization and the order of shielding – aldehydes, olefins and aromatics, substituted alkanes to alkanes. However, the analogy must not be carried too far. For example, although shielding caused by exchange with first row atoms follows the order of the substituent electronegativities, this is not the case for chlorine, bromine, and iodine, in which upfield shifts have been found. This ‘heavy atom’ effect has no

counterpart in proton chemical shifts. Furthermore, the effect of the substituents on  $^{13}\text{C}$  shifts is not confined to the nearest atom ( $\alpha$ ), as in proton chemical shifts, but must be evaluated for substituents two ( $\beta$ ), three ( $\gamma$ ), and four ( $\delta$ ) bonds from the carbon atom considered. It should be noted that the substitution in  $\alpha$ -,  $\beta$ -, and  $\gamma$ -position generally deshields the carbon but one in  $\gamma$ -position shields. The  $\gamma$ -effect has been interpreted as a steric shift and is of importance for distinguishing *Z/E* and *gauche/anti* stereoisomers. The effects of  $\delta$ - or further positions are usually small (less than 3 ppm), but these long-range effects ( $\gamma$ ,  $\delta$ , etc.) may be large in cyclic compounds, depending on the distance of the substituent from the respective carbon. Additionally, many other shift effects such as electric field, isotope, solvent may reach, in special cases, a magnitude of several ppm.

The various factors contributing to the shielding generally combine in a complex manner. An outstanding feature of  $^{13}\text{C}$  chemical shifts is the finding that substituent contributions are reasonably additive, especially when they do not directly interact. This allows substituent parameters for different classes of organic compounds to be determined and

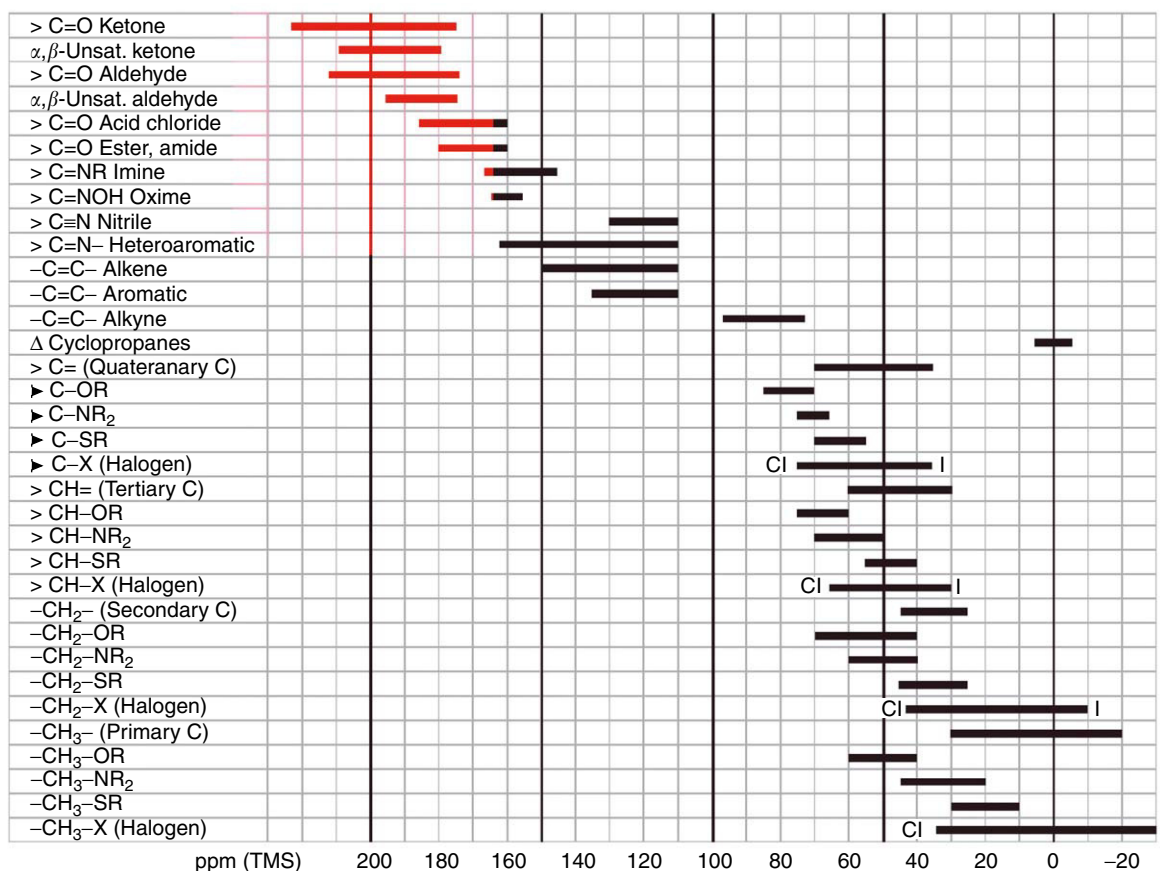


Figure 2  $^{13}\text{C}$  chemical shift correlation chart for common organic compounds.

**Table 3**  $\alpha$ - and  $\beta$ -Substituent-induced chemical shifts ( $\delta_{\text{SCS}} = \delta_{\text{C}} - \delta_{\text{H}}$ ) for  $sp^3$  carbons in alkyl derivatives RX

$-R$	$-CH_3$	$-CH_2$	$CH_3$	$-CH$	$(CH_3)_2$
X-	$\alpha$	$\alpha$	$\beta$	$\alpha$	$\beta$
$-H^a$	-2.3	7.3	7.3	15.9	15.4
$-CH=CH_2$	16.4	20.1	6.1	16.4	6.7
$-C_6H_5$	19.1	21.8	8.5	18.4	8.6
$-F$	70.5	72.4	8.1	71.6	7.5
$-Cl$	23.3	32.6	11.6	37.8	11.9
$-Br$	7.3	20.2	12	28.6	13.6
$-I$	-24.6	-8.4	13.2	15.2	4.9
$-OH$	47.9	50.5	10.9	48.1	9.9
$-OCH_3$	58.6	60.4	7.4	56.7	6.0
$-OCOCH_3$	49.2	53.1	7.1	51.6	6.5
$-COCH_3$	28.4	27.9	-0.3	25.7	2.8
$-COOCH_3$	18.3	19.9	1.9	18.2	3.7
$-NH_2$	26	29.6	10.5	27.1	10.6
$-N(CH_3)_2$	45.8	46.9	5.5	40.4	5.6
$-NHCOCH_3$	23.8	26.8	7.3	24.6	6.9
$-CN$	-0.54	3.5	3.3	3.9	4.5
$-NO_2$	-61.1	63.5	5.0	62.9	5.4

<sup>a</sup>Basic chemical shifts for the carbon atoms in the parent hydrocarbon.

**Table 4**  $^{13}\text{C}$  Chemical shifts ( $\delta$ ) for  $sp^2$  carbons in vinyl derivatives  $-CH_2=CH-X$ 

X	$CH_2=$	$=CH-$
$-H$	123.3	123.3
$-CH_3$	114.7	135.0
$-C_6H_5$	112.3	135.8
$-CH=CH_2$	116.3	136.9
$-COCH_3$	128.0	137.1
$-COOCH_3$	130.3	129.6
$-Cl$	116.0	124.9
$-Br$	120.8	114.3
$-I$	129.2	84.1
$-CN$	137.5	108.2
$-N(CH_3)$	91.3	151.3

used for signal assignments. As an example empirically determined effects (substituent-induced chemical shifts, SCS) for alkanes are presented in Table 3. Unsaturated carbon shifts are generally easier to interpret because they primarily depend on the local p-electron density, which is mainly governed by mesomeric effects (see Table 4). This is especially straightforward for monosubstituted benzene derivatives (see Table 5). The SCS values for carbons in *para*-positions of benzene rings can be used for determination of Hammett  $\sigma$ -constants of rare substituents. A number of empirical parameters and additivity schemes can be found in the literature for different carbon environments. Recently, many standalone or web-based computer programs offer different schemes for prediction of carbon chemical shifts, which rely on empirically based increments,

**Table 5** SCS values for monosubstituted benzene derivatives

$-X$	$C_i$			
	<i>Ips</i>	<i>Ortho</i>	<i>Meta</i>	<i>Para</i>
$-H^a$	0	0	0	0
$-CH_3$	+9.3	+0.8	0	-2.9
$-CH_2CH_3$	+15.6	-0.4	0	-2.6
$-CH=CH_2$	+9.5	-2.0	+0.2	-0.5
$-C_6H_5$	+13.0	-1.0	+0.4	-1.0
$-F$	+34.8	-12.9	+1.4	-4.5
$-Cl$	+6.0	+0.4	+1.3	-1.9
$-Br$	-5.4	+3.3	+2.2	-1.0
$-I$	-32.0	+10.0	+3.0	+1.0
$-OH$	+26.9	-12.7	+1.4	-7.3
$-OCH_3$	+32.4	-14.4	+1.0	-7.7
$-OCOCH_3$	+22.7	-6.8	+0.7	-2.6
$-COCH_3$	+9.0	+0.1	-0.1	+4.3
$-COOCH_3$	+2.1	+1.2	0	+4.4
$-NH_2$	+18.1	-13.4	+0.9	-9.9
$-N(CH_3)_2$	+23.1	-16.2	+1.0	-12.0
$-NHCOCH_3$	+11.1	-9.9	0	-6.0
$-CN$	-15.4	+3.6	+0.6	+3.9
$-NO_2$	+20.0	-4.9	+0.9	+5.9

<sup>a</sup>Chemical shift for benzene carbons - 128.5 ppm.

on quantum chemistry calculations, on library search with or without use of artificial neural networks, etc.

Not only the isotropic chemical shift but also the anisotropy of the chemical shielding (CSA) tensor of the  $^{13}\text{C}$  nucleus is another unique NMR parameter that could be generally measured. Such a tensor is a probe to the electron density distribution around the nucleus and provides structural information that is very difficult to obtain by other means.

### $^{13}\text{C}$ Coupling Constants

Coupling constants to  $^{13}\text{C}$  are generally tedious to measure because of the low natural abundance of the carbon nucleus. Depending on the subject under investigation as well as on the availability of substance and instrumentation different experimental schemes could be most practical. Often, the presence or absence of a coupling above certain threshold (e.g., 2 Hz) could be sufficient for solving the problem. Commonly used methods are the following:

- inspection of the  $^{13}\text{C}$  satellites in the  $^1\text{H}$  spectrum;
- $^{13}\text{C}$  NMR spectra under gated decoupling conditions;
- various selective and nonselective 2D methods with direct or inverse detection (CH-*J*-resolved, HSQC, HMBC, etc.)

Coupling to abundant nuclei ( $^1\text{H}$ ,  $^{31}\text{P}$ ,  $^{19}\text{F}$ ) is most easily detected. Measurement of C-C and of heteronuclear coupling constants to low abundant nuclei is still a time consuming task, even in modern

spectrometers with high sensitivity unless  $^{13}\text{C}$  or heteroatom enrichment is carried out.

In dealing with heteronuclear couplings the general relationship  ${}^nJ_{\text{CX}} = \gamma_{\text{C}} \cdot \gamma_{\text{X}} \cdot {}^nK_{\text{CX}}$  can be used.  ${}^nK_{\text{CX}}$  depends on the electron density distribution along the  $n$  bonds. This provides the basis for an estimation of  $J$  values in situations where bonding is expected to change little with variation in X, e.g., when D is substituted for H. Tables 6–8 show an overview of representative carbon couplings.

One-bond  ${}^1J_{\text{CH}}$  couplings are of some theoretical interest, as they seem to be directly proportional to the fraction of s-character ( $\rho$ ) of the C–H bond ( ${}^1J_{\text{CH}} \sim 500\rho$ ). The couplings in methane ( $\text{sp}^3$ ),

ethylene ( $\text{sp}^2$ ), and acetylene ( $\text{sp}$ ) agree well with this prediction. Observed coupling of 160 Hz strongly supports the trigonally hybridized model for cyclopropane. This simple and very useful relationship is, however, strictly valid only for hydrocarbons and compounds without strongly perturbing substituents. In substituted methanes the C–H coupling changes considerably with the introduction of polar substituents and this is not primarily due to any change in the carbon atom hybridization. The substituent effects are reasonably additive and may be used to estimate C–H couplings in any multisubstituted compound. In summary, these  ${}^1J_{\text{CH}}$  couplings are of diagnostic use and can be utilized as routinely as HH couplings to provide structural information. Additionally, they are relatively easily measured.

Geminal and vicinal couplings have a similar general pattern to the analogous H–H coupling, being somewhat smaller in magnitude – the relation  ${}^{2,3}J_{\text{CH}} = (0.6 \pm 0.1){}^{2,3}J_{\text{HH}}$  usually holds for cases of comparable binding. Especially useful in aliphatic compounds is the Karplus-like dependence of the vicinal couplings. For carbon and proton in synclinal or antiperiplanar position one measures couplings  $\sim 2$  and 9 Hz, respectively, whereas in freely rotating fragments a value in the range of 4 Hz should be expected. Of special diagnostic value in alkenes and aromatics is the rule that  ${}^3J_{\text{CH}} > {}^2J_{\text{CH}}$ .

One-bond C–C coupling, as  ${}^1J_{\text{CH}}$ , correlates with the s-character of both carbons according to:  ${}^1J_{\text{CC}} \sim 730s_1 \cdot s_2 - 17$ .  ${}^2J_{\text{CC}}$  values are usually smaller than  ${}^3J_{\text{CC}}$ . Larger  ${}^2J_{\text{CC}}$  values are observed if the carbons have high s-character (e.g., ethynes) or when an electron deficient carbon is involved. C–C vicinal coupling constants behave in a similar way as  ${}^3J_{\text{CH}}$  and  ${}^3J_{\text{HH}}$ . In fact, available data indicate that substitution of a hydrogen by carbon reduces  ${}^3J$  by a factor of  $c = 0.6\text{--}0.8$ , i.e.,  ${}^3J_{\text{CC}} = c \cdot {}^3J_{\text{CH}} = c^2 \cdot {}^3J_{\text{HH}}$ .

$^{13}\text{C}$ – $^{15}\text{N}$  coupling constants have smaller magnitude than their C–H counterparts because of the low magnetogyric ratio of  $^{15}\text{N}$ . One-bond couplings may be estimated according to  $|{}^1J_{\text{CN}}| = 125s_{\text{C}}s_{\text{N}}$ , where  $s_{\text{C}}$

**Table 6** Ranges of  $^{13}\text{C}$  and  $^1\text{H}$  coupling constants  ${}^nJ$

No. of bonds	${}^nJ_{\text{CC}}$	${}^nJ_{\text{CH}}^{\text{a}}$	${}^nJ_{\text{HH}}$
1	+30, +180	+98, +320	+276( $\text{H}_2$ )
1 <sup>b</sup>	{35,50}, (60,70)	{120,130}, (170,220)	–
2	$ J  < 20$	–10, +20	$ J  < 30$
2 <sup>b</sup>		$ J  < 8$	$12 <  J  < 18$
3	0, +15	0, +15	0, +18
3 <sup>b</sup>		0, +14	0, +10
>3 <sup>b</sup>	$ J  < 1$	$ J  < 1$	$ J  < 7$

<sup>a</sup>Division by 6.515 yields instead of hydrogen the corresponding deuterium constant.

<sup>b</sup>Coupling constant ranges in peptides and proteins; aliphatic in { }, aromatic in ( );  ${}^1J(\text{C}^{\alpha}, \text{C}') = 50\text{--}60\text{ Hz}$ ;  ${}^1J(\text{C}^{\alpha}, \text{N}) = 4\text{--}6\text{ Hz}$ ;  ${}^1J(\text{C}', \text{N}) = 12\text{--}15\text{ Hz}$ .

**Table 7** Representative examples of carbon coupling constants  ${}^nJ$

Compounds	${}^1J_{\text{CH}}$	${}^1J_{\text{CC}}$	${}^2J_{\text{CH}}$	${}^3J_{\text{CH}}$
$\text{H}_3\text{C}-\text{CH}_3$	125	35	–4	–
$\text{H}_2\text{C}=\text{CH}_2$	156	68	–2	–
$\text{H}_3\text{C}-\text{CH}=\text{CH}_2$	–	–	5	8 Z 12 E
$\text{H}_3\text{C}-\text{CH}=\text{O}$	172 <sup>a</sup>	–	+27 <sup>b</sup>	–
$\text{C}_6\text{H}_6$	158	–	+1	7
$\text{HC}\equiv\text{CH}$	249	171	+50	–

<sup>a</sup>‘a’ and ‘b’ over carbon atoms indicate target nuclei.

**Table 8** Representative examples of  ${}^1J_{\text{CH}}$  coupling constants

$\text{CH}_3\text{X}$	X =	F	Cl	OH	$\text{NH}_2$	H	$\text{CH}_3$	Li
		149	150.0	141	133	125	125	98
$\text{C}_n\text{H}_{2n}$	$n =$	3	4	5	6	10		
		160	134	128	125	124		
$\text{C}_n\text{H}_{2n-2}$	$n =$	3	4	5	6	10		
		228	169	162	158	156		
$(\text{CH}_3)_2-\text{CH}_2$			$(\text{CH}_3)_3-\text{CH}$			$\text{CH}_2\text{Cl}_2$	$\text{CHCl}_3$	$\text{HC}\equiv\text{N}$
119			114			178	209	269



and  $s_N$  represent the  $s$ -character of the bonding partners. The signs of  $^nJ_{CN}$  are usually opposite to the signs of the C–H and H–H couplings because of the negative magnetogyric ratio of  $^{15}\text{N}$ .

$^{13}\text{C}$ – $^{19}\text{F}$  coupling constants are larger than their C–H counterparts, e.g., 271 and 235 for  $\text{CH}_3\text{CF}_3$  and  $\text{CF}_2\text{H}_2$ .

An illustrative example of  $^{13}\text{C}$ – $^{31}\text{P}$  couplings is provided for  $\text{CH}_3\text{CH}_2\text{PO}(\text{OCH}_2\text{CH}_3)_2$ :  $^1J_{\text{CP}} = 143\text{ Hz}$ ;  $^2J_{\text{CP}}$  (CCP) = 7.1 Hz;  $^2J_{\text{CP}}$  (COP) = 6.9 Hz;  $^3J_{\text{CP}}$  (CCOP) = 6.2 Hz.

### Spin–Lattice Relaxation Times $T_1$

$^{13}\text{C}$   $T_1$  values are often measured under conditions of BB decoupling using the inversion–recovery (IR) pulse sequence. A single line for each carbon atom is collected with an intensity, following the law:  $I = I_0[1 - 2\exp(-\tau/T_1)]$  for ideal pulses and sufficiently long relaxation delays. In practice, intensities are fitted to the expression  $A - B\exp(-\tau/T_1)$  in order to avoid systematic errors due to imprecise  $I_0$  values, pulse imperfections, and offset effects. Determination of  $I$  is often more accurate if lines are artificially broadened during FT data processing. In medium-sized organic molecules (10–50 carbon atoms)  $T_1$  values are typically in the range 0.1–15 s for proton-bearing carbon atoms. The  $T_1$  values of quaternary carbons are much longer.

A number of mechanisms are known to contribute to spin–lattice relaxation in a molecule: dipole–dipole, spin–rotation, quadrupolar, scalar, and CSA, each associated with a corresponding relaxation rate. Analytical interest focuses only on the internuclear C–H dipole–dipole relaxation with a rate  $R_{\text{IDD}} = 1/T_{\text{IDD}} = (\eta/1.988)(1/T_1)$ , which is proportional to the inverse sixth power of the carbon–proton distance  $r_{\text{CH}}$ , and depends on the rotational and internal molecular motion. For medium-sized molecules of nearly spherical shape (isotropic molecular reorientation) in solvents of low viscosity the condition of extreme narrowing holds, i.e.  $\omega_C\tau_c \ll 1$  for all Larmor frequencies  $\omega_C$ , and therefore  $R_1 = AN\tau_c$ . The correlation time  $\tau_c$  is a measure of the velocity of the molecule's rotational diffusion jumps. For  $N$  nearest protons within a distance  $r_{\text{CH}}$  the quantity  $A$  takes the form  $A = (b\gamma_C\gamma_H/2\pi)^2/r_{\text{CH}}^6$  and becomes  $3.56 \times 10^{10}\text{ s}^{-2}$  for  $r_{\text{CH}} = 109\text{ pm}$ .

If proton-bearing carbons have  $T_1$  values less than 20 s the dipole–dipole term predominates and  $T_{\text{IDD}}$  can be set equal to  $T_1$ . The  $N$  dependence strictly holds only if CH and  $\text{CH}_2$  are part of the rigid backbone of the molecule, then their motional characteristics are alike.  $\text{CH}_3$  groups have usually additional rotation about their symmetry axis with little or no hindrance, so that their overall motion is quite

different and the simple  $N$  law given above does not hold.  $T_1$  values in the rigid backbone of a molecule generally decrease with molecular size since  $\tau_c$  increases with size. The rule of thumb is that  $\tau_c$  (in ps) is equal to their molecular weight (in Da) for globular molecules in water. In practice, nonspherical molecules with conformational flexibilities are more commonly met than rigid spheres. For instance, segmental mobility in  $n$ -alkyl side chains produce pronounced differences in relaxation rates. A classic example is found in 1-substituted  $n$ -alkanes, where the substituent ‘anchors’ the end of the molecule due to a high moment of inertia in the case of Br or to hydrogen bonding for substitution with a hydroxyl group. This reflects in the  $T_1$  values, measured for  $n$ -decanes that are shorter the slower the effective motion of each individual  $\text{CH}_2$  moiety occurs (in s):

	X - $\text{CH}_2$ - $\text{CH}_2$ - $(\text{CH}_2)_3$ - $\text{CH}_2$ - $\text{CH}_2$ - $\text{CH}_2$ - $\text{CH}_2$ - $\text{CH}_2$							
X = H	8.7	6.6		4.4	5.0	5.7	6.6	8.7
X = Br	2.8	2.7	2.0	2.0	2.0	3.1	3.9	5.3
X = OH	0.7	0.8	0.8	0.8	1.1	1.6	2.1	3.1

Rotational diffusion jumps around a parallel (long) axis in rod-shaped rigid molecules turn up faster than those with respect to any vertical axis. Consequently, the effective correlation time for the C–H vector reorientation, responsible for the relaxation rate of the target  $^{13}\text{C}$  nucleus, depends on the angle between the C–H vector and the parallel axis. The detailed treatment is quite complex but the trends become evident from the  $T_1$  values in molecules like phenylethyne and its shape analogous derivatives (Table 9). Depending on the solvent the observed  $T_1$  values differ up to a factor of 2.5 that is quite substantial for molecules of that size. The *para*- and the ethyne-carbons have shorter  $T_1$  values because their C–H reorientation is mainly caused by slow rotational jumps about the vertical axis.  $T_1$  values for *ortho*- and *meta*-carbons are larger due to the further fast rotational diffusion around the parallel axis. The pronounced dependence of the relaxation rates on solvent viscosity is clearly revealed from the data in Table 9. Additionally, the effects of intermolecular interaction in phenylpropylene, of hydrogen bonding in phenylethyne, and of stacking due to electric dipole–dipole interactions in benzonitrile could be followed.

### $^{13}\text{C}$ Line Widths

The  $^{13}\text{C}$  signal, as generated by a single pulse, nearly follows the exponential decay law  $\exp(-t/T_2^*)$  as is

**Table 9** Solvent dependence of  $^{13}\text{C}$   $T_1$  relaxation times (in s)<sup>a</sup>

Solvent	$\text{Ph}-\text{C}\equiv\text{CH}$				$\text{Ph}-\text{C}\equiv\text{C}-\text{CH}_3$			$\text{Ph}-\text{C}\equiv\text{N}$		
	Ortho	Meta	Para	$\equiv\text{CH}$	Ortho	Meta	Para	Ortho	Meta	Para
$\text{Ph}-\text{C}\equiv\text{CH}$	12.8	12.8	8.7	8.8	8.4	8.4	4.2	10.4	11.6	6.6
			(neat)		(6.4 mol.%)			(infin. dil.)		
$\text{C}_6\text{H}_{12}$	13.1	12.5 (5.5 mol.%)	8.5		9.9 (6.4 mol.%)	10.1	5.0	13.0	12.9 (8.8 mol.%)	8.4
$\text{OP}(\text{N}(\text{CH}_3)_2)_3$	2.4 <sup>c</sup>	1.9 <sup>b</sup> (9.7 mol.%)	1.9 <sup>b</sup>	0.95	3.2 (7.5 mol.%)	3.2	1.6	2.9	2.9 (8.2 mol.%)	2.7

<sup>a</sup> 22.6 MHz  $^{13}\text{C}$  measurements at 303 K. Concentrations are given in parentheses; 'infin. dil.' stands for extrapolation to infinite dilution.

<sup>b</sup>  $^{13}\text{C}$  lines not resolved for *meta*- and *para*-carbons.

<sup>c</sup> Solvent viscosities: 0.93 cP for phenylethyne and 4.05 cP for  $\text{OP}(\text{N}(\text{CH}_3)_2)_3$  at 303 K; 0.9 cP for cyclohexane at 298 K.

common in NMR. Accordingly, the  $^{13}\text{C}$  resonance line approximates the Lorentzian shape with a full half-width of  $\Delta\nu = 1/(\pi T_2^*)$ . The broadening parameter  $T_2^*$  is a sum of both natural, namely spin-spin relaxation and anisotropy broadening, and instrumental contributions. The latter usually dominates, most notably due to magnetic field inhomogeneity, unresolved long-range couplings, and/or residual linewidth effects from broadband decoupling. It is often overlooked that molecular motion may be not fast enough, e.g., solutions of large biomolecules or polymers, for complete averaging of the large  $^{13}\text{C}$  chemical shift anisotropies to be achieved. In such a case anisotropy broadening may become substantial. This has in fact been observed and may be expected to be more important for  $^{13}\text{C}$  NMR at very high static magnetic fields.

In a multiscan pulse experiment with pulse spacing,  $\tau$ , the first pulse may produce maximum signal intensity  $I_0$  of each  $^{13}\text{C}$  resonance line. The intensity  $I$  after the second pulse is a function of  $\tau$  and generally comes out smaller if  $\tau < 5T_1^*$ , the ratio  $I/I_0$  being equal to  $1 - \exp(-\tau/T_1)$ , where  $T_1$  is the spin-lattice relaxation time. For pulse intervals greater than  $5T_1$ , the ratio  $I/I_0$  is essentially unity. Experimental  $^{13}\text{C}$   $T_1$  values vary from  $\sim 0.1$  to 300 s for organic molecules. Large variations for each individual carbon site in a given molecule could occur. In neat phenylethyne this varies from  $\sim 10$  s for the proton-bearing carbons to 100 s for quaternary carbons, while in *n*-decanol values from 3.1 s at the aliphatic-end carbon down to 0.7 s for the O-end carbon are measured. An important analytical aspect of the large  $T_1$  variations in a given molecule is that line intensities do not strictly reflect the number of chemically equivalent carbon atoms. These differences can be usually analytically explored for distinguishing of proton-bearing and quaternary carbons,

provided the latter have intensities above the noise level. For assuring their appearance, excitation by the so-called Ernst angle, an average optimum pulse angle, instead of the  $90^\circ$  pulse, is usually exerted. For routine applications  $30^\circ$  pulses with an interpulse delay of 2 s are most common.

### $^{13}\text{C}$ Spectra for Structure Elucidation

Routinely 1D decoupled and multiplicity-edited spectra are the primary source of  $^{13}\text{C}$  NMR spectral information. Combined with the available  $^1\text{H}$  NMR spectral information, an overall picture of the number of structural elements in the molecule is obtained. It could be used for structure elucidation and/or spectral assignments. Further spectroscopic information (mass and/or optical spectra) or other knowledge (compound class, synthesis routes, etc.) facilitates generation of guess structures. More or less luckily this could be done from NMR data alone, even in absence of additional data. The number of the carbon and proton atoms in the molecule could often be reliably assumed. Equivalence by accident can be discarded by inspection of the proton and carbon spectra, which are usually unlike in contrast to cases with equivalence by symmetry. Use of special information, temperature dependence of spectral parameters, exchange line broadening, segmental motion and/or anisotropic rotation of molecules, solvent and dilution effects, etc., can have additional diagnostic value. Knowledge of established relationships between NMR parameters – chemical shifts, coupling constants, relaxation times – and molecular structure are of prime importance in the next step – unambiguous assignment of as many as possible carbon and proton signals based on shift and proton coupling arguments. Because of the great detail in the spectra, this can be quite laborious. Fortunately,

extensive support by an increasing number of powerful spectroscopic tools is provided, offered both by instrument producers and chemistry software houses. These can easily be used on- or offline since virtually every NMR spectrometer is now linked to a workstation and hence to local and far-reaching computer networks. Assignments may be achieved on the bases of one or more of the following approaches:

- $^{13}\text{C}$  spectral similarity search in libraries – printed collection or a computer database with software for storage, retrieval, and search. Libraries may be homemade or commercially available and accessed in home or via the public network.
- Generation of trial structures and substructures possibly constrained by known substructures.
- Comparison of observed  $^{13}\text{C}$  chemical shifts with predicted parameters in trial (sub)structures. Predictions can be made using correlation charts, empirically derived additivity relationships (e.g., HOSE code), or theoretical calculations at different levels of precision (including *ab initio*). In recent years, artificial neural networks are more and more applied for prediction.

At the end of this step typically one or more trial structures survive, some of the assignments still remain uncertain. This necessitates additional experiments for establishing and verification of C–H connectivities along the carbon chain. More often a combination of proton-detected gradient accelerated 2D correlation experiments (or their selective 1D versions) is applied because of their higher sensitivity and shorter measuring times. Most important sequences are: HSQC (HMQC) and HMBC – for establishing one- and multiple-bond C–H correlations. They assist structure unfolding and afford sure signal assignment of one nucleus provided the assignment of the other is certain. Complete NMR analysis generally includes additional COSY, NOESY (for establishing through-bond and through-space H–H connectivities) and/or TOCSY (as a proof of a common H–H coupling pathway) experiments.

Due to its inherent insensitivity the important direct 2D experiment for tracing of the carbon skeleton – INADEQUATE (incredible natural abundance double quantum transfer experiment), based on measurement of doubly  $^{13}\text{C}$ -labeled species needs, unless  $^{13}\text{C}$  enriched, a large quantity of the compound under investigation. In favorable cases the proton-detected analogue ADEQUATE could be applied.

Despite the fact that a totally automated structure elucidation tool still remains a vision for chemists and spectroscopists and is a declared object for

software developers, in recent years tremendous progress toward automation of that process has been achieved. By analogy to the intuitive approach so far applied, a few software tools have been designed that allow interactive structure elucidation of an unknown compound, e.g., using  $^1\text{H}$ ,  $^{13}\text{C}$ , DEPT, COSY, HSQC, and HMBC spectra.

### Isotopic Enrichment

Along with the improvement in  $^{13}\text{C}$  detection techniques,  $^{13}\text{C}$  labeling developed into an excellent tool for structure analysis not only of small organic molecules (determination of  $^{13}\text{C}$  coupling constants and connectivities) but also of proteins and oligonucleotides and for elucidation of reaction pathways in biosynthesis. Today, a huge variety of labeling patterns are achievable, among them complete, block, site, stochastic, exchange, metabolic, and residue-specific labeling. A common approach is to grow microorganisms on a nutrient medium containing  $^{13}\text{C}$ -labeling precursors. In order to ascertain where the precursor has been incorporated into the target molecule it is mandatory to assign each  $^{13}\text{C}$  NMR signal. Depending on the system studied and on the acceptable degree of spectral complexity, a choice of the type of precursor, of the degree of enrichment, and of the number and position of the  $^{13}\text{C}$  label(s) in the precursor molecule has to be made. For example, experiments with doubly labeled compounds such as  $^{13}\text{CH}_3^{13}\text{COONa}$ , a nutrient of frequent use, allows the follow-up of the fate of a bond rather than of an individual carbon as for singly labeled acetates. However, use of highly enriched species may lead to complicated  $^{13}\text{C}$  spectra due to multiple C–C coupling and/or strong coupling effects. A simple remedy in practice is the dilution of one part in two to five parts of unlabelled sodium acetate that renders C–C couplings between the doubly labeled acetate groups negligible.

### Solid-State $^{13}\text{C}$ NMR

$^{13}\text{C}$  NMR studies of polycrystalline or glassy organic materials are now possible on a routine basis. Here, the low natural  $^{13}\text{C}$  abundance becomes an advantage rather than a drawback since dipolar broadening of the target  $^{13}\text{C}$  nuclei due to protons can be removed by high-power  $^1\text{H}$  decoupling while that due to other carbons is negligible because of the underlying inverse-cubic distance dependence of dipolar interactions. Another major obstacle, namely the long  $^{13}\text{C}$   $T_1$  values in solids and hence the long relaxation delays needed for signal averaging, is overcome with the H–C CP technique. As a result,  $^{13}\text{C}$

signal enhancement of 3.98 ( $\gamma_{\text{H}}/\gamma_{\text{C}}$ ) could be gained for each scan, so that more FIDs can be collected in a given time because the faster relaxing protons govern the recovery to equilibrium. This applies both to proton bearing and to quaternary carbons since the underlying dipolar interactions have through-space character in contrast to that mediated by the electrons of the C–H bond scalar interactions.

$^{13}\text{C}$  CP spectra with high-power  $^1\text{H}$  decoupling exhibit broad resonance lines which reflect the CSA of each  $^{13}\text{C}$  nucleus. For example, in poly hydroxybutyrate ( $-\text{O}-\text{CH}(\text{CH}_3)-\text{CH}_2-\text{CO}-$ ) $_n$ , the spectrum at 50.3 MHz consists of two broad lines that span over the range from  $-50$  to  $280$  ppm. The CO resonance shows a CSA pattern and extends from  $\sim 120$  to  $260$  ppm. The patterns completely overlap for the aliphatic carbons. These resonances can be made to collapse into relatively sharp single lines by rapid rotation of the sample container about an axis being tilted by  $54^\circ 44'$  with respect to the  $B_0$  direction (MAS). For spinning at  $\sim 4$  kHz, a sharp  $^{13}\text{C}$  resonance line is observed for each of the four carbon types. For spinning rates of  $\sim 600$  Hz, many sharp spinning side bands show up. Their intensities map the CSA line profile. Today, CPMAS  $^{13}\text{C}$  NMR is a standard tool for the study of polymers and other organic materials because high-resolution  $^{13}\text{C}$  spectra can be obtained for insoluble substances. A widely used application is the differentiation of polymorphic forms in the pharmaceutical industry (e.g., aspirin). When spinning cannot be set to a high enough rate for complete collapse, e.g., at very high fields, suppression techniques such as TOSS (total suppression of side bands) can be applied for spectrum simplification. In recent years many new  $^{13}\text{C}$  solid-state applications have evolved.

## Quantitative Analysis

Quantitative NMR determination ( $^{13}\text{C}$  at the order of 1%) is an analytically important though troublesome topic in NMR in general. In principle, the free induction decay (FID) height  $F_0$  immediately after a single RF pulse is a faithful relative measure of the molecular concentration. It can easily be determined in the recorded FID for single-line spectra. For analytical purposes, the single-line  $F_0$  values need to be compared for several samples that require the changing of sample tubes. Care has to be taken that instrumental conditions are not affected, e.g., the tuning of the receiver coil as caused by a different dielectric constant of the bulk sample. In multiline spectra, this usually poses no problem since relative intensities are measured in the same sample. Here,

however, individual  $F_0$  values are not accessible in the FID so that the FT or a method for nonlinear spectral analysis such as linear prediction (LP), maximum entropy method (MEM), or Bayesian analysis must be applied. Then, intensities  $F_0$  are obtained from the areas (not from the height!) under the resonance lines. Practical spectral analysis, unfortunately, is prone to error, mostly because of FID truncation effects and inaccurate digital integration (bad baseline, insufficient number of data points per line, poor integration routine).

In the case of  $^{13}\text{C}$  NMR there are specific advantages and disadvantages. Advantageous are the  $^1\text{H}$ -decoupled spectra since there is only one line for each carbon. In not too crowded spectral regions, Lorentzian line-broadening can be applied during data processing that alleviates truncation and integration artifacts. Regions with closely spaced lines may be selectively subjected to nonlinear spectral analysis (LP, MEM, Bayesian). The main disadvantages are the low sensitivity and the pronounced differences in  $^{13}\text{C}$  spin-lattice relaxation times and NOE factors within the same molecule and their different susceptibility to the molecular environment, e.g., when the temperature or solvent is changed. Common solutions comprise careful and precise adjustment of the relaxation delays and acquisition using the inverse gated decoupling scheme for suppression of the NOE. Additionally, the use of paramagnetic relaxation reagents (e.g.,  $\text{Co}(\text{acac})_2$ ) reduces the measurement times due to sizably shorter  $^{13}\text{C}$  spin-lattice relaxation times and considerable quench of the NOE. In cases of sensitivity problems the use of the highest possible magnetic field strength could be of general help.

*See also:* **Nuclear Magnetic Resonance Spectroscopy:** Overview; Principles; Instrumentation. **Nuclear Magnetic Resonance Spectroscopy-Applicable Elements:** Fluorine-19; Nitrogen-15; Phosphorus-31; Organometallic Compounds. **Nuclear Magnetic Resonance Spectroscopy Applications:** Food; Pharmaceutical. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Nuclear Overhauser Effect; Multidimensional Proton; Solid-State.

## Further Reading

- Abraham RJ, Fisher J, and Loftus P (1996) *Introduction to NMR Spectroscopy*. Chichester: Wiley.
- Atta-ur-Rahman and Ahmad VU (1992) *C-13 NMR of Natural Products; vol. 1, Monoterpenes and Sesquiterpenes; vol. 2, Diterpenes*. New York: Plenum.
- Breitmaier E and Voelter W (1987) *Carbon-13 NMR spectroscopy – high resolution methods and applications in organic chemistry and biochemistry*. Weinheim: VCH.

- Bremser W, Franke B, and Wagner H (1986) *Chemical shift ranges in  $c$ -13 nmr spectroscopy*. Weinheim: VCH.
- Claridge T (1999) *High resolution nmr techniques in organic chemistry*. Amsterdam: Pergamon.
- Field LD, Sternhell S, and Kalman JR (2002) *Organic Structures from Spectra*. New York: John Wiley.
- Kalinowski HO, Berger S, and Braun S (1996) *Carbon-13 NMR spectroscopy*. New York: John Wiley.
- Pihlaja K and Kleinpeter E (1994) *Carbon-13 NMR Chemical Shifts in Structural and Stereochemical Analysis*. Weinheim: VCH.
- Tonelli AE (1989) *NMR Spectroscopy and Polymer Microstructure: The Conformational Connection*. Weinheim: VCH.
- Wehrli FW and Wirthlin T (1976) *Interpretation of Carbon-13 NMR Spectra*. London: Heyden.

## Fluorine-19

J C Lindon, Imperial College London, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Fluorine has a nuclear magnetic resonance (NMR)-active nucleus with one of the highest sensitivities, third only to proton ( $^1\text{H}$ ) and the radioactive tritium ( $^3\text{H}$ ). As it has favorable NMR properties, fluorine NMR has found widespread application in both chemistry and biology. This has resulted in a large number of reviews and books and there have been extensive compilations of solution-state  $^{19}\text{F}$  NMR chemical shifts and coupling constants in organic, organometallic, and inorganic compounds, these being very useful for confirmation of molecular structures. As well as studies on small molecules, a significant number of fluorinated polymers are important and  $^{19}\text{F}$  NMR spectroscopy can be used in their characterization, particularly the determination of polymer chain microstructure. In addition, the application of high-resolution techniques in solids, such as magic-angle spinning (MAS) and combined rotation and multiple pulse spectroscopy (CRAMPS), has expanded with a variety of applications in the literature. There are an increasing number of studies using  $^{19}\text{F}$  NMR spectroscopy in biology including the study of binding of fluorinated molecules to macromolecules, applications to membrane structure and flexibility, the elucidation of drug metabolites *in vitro*, and the study of the distribution and metabolism of compounds *in vivo*.

### Properties of the Nucleus

Fluorine possesses only one isotope,  $^{19}\text{F}$ , which has a spin  $I = 1/2$  and a sensitivity relative to  $^1\text{H}$  for an equal number of nuclei of 0.83. The magnetogyric ratio has a value of 0.94 that of  $^1\text{H}$ , so the basic resonance frequency for  $^{19}\text{F}$  is, for example, 282.2 MHz on a spectrometer that operates at a

field strength of 7.05 T with  $^1\text{H}$  NMR observation at 300 MHz.

### Sample Preparation and Spectrum Measurement

Samples for solution study are prepared in the usual manner as for  $^1\text{H}$  NMR spectroscopy. Degassing the solutions can be important because of the effect of paramagnetic oxygen on the  $^{19}\text{F}$  spin relaxation times. A number of substances have been proposed for chemical shift referencing, and for a time an external sample of trifluoroacetic acid was commonly used. Now, however, the standard reference compound is internal trichlorofluoromethane,  $\text{CFCl}_3$ , added at a few mole percent and with a fluorine chemical shift ( $\delta_{\text{F}}$ ) taken to be at 0.0 ppm. This compound is very volatile and quite unreactive, and therefore easy to remove from samples if necessary. At high-field strengths, the  $^{19}\text{F}$  NMR chemical shift is perturbed by the combined isotope effects of the various possible combinations of  $^{35}\text{Cl}$  and  $^{37}\text{Cl}$  containing molecules. This means that at best the signal is broadened and at worst an asymmetric multiplet structure from slightly different chemical shifts can be resolved. The current sign convention for fluorine chemical shifts is for negative values (denoted by  $\delta$  or ppm) to low frequency of  $\text{CFCl}_3$  (to the so-called high field) and increasing positive values to high frequency of  $\text{CFCl}_3$ . Early literature values of chemical shifts generally used the reverse sign convention and so care needs to be taken in relating recent data to older data. Some measurements have also used a number of secondary reference compounds and the conversion factors are given in Table 1.

The wide range of  $^{19}\text{F}$  NMR chemical shifts can lead to problems in data acquisition, in particular nonuniform excitation and poor digitization. For example, peaks from a typical organic compound sample containing  $\text{CFCl}_3$  and both aliphatic and aromatic fluorines can cover a spectral range of



- Bremser W, Franke B, and Wagner H (1986) *Chemical shift ranges in  $c$ -13 nmr spectroscopy*. Weinheim: VCH.
- Claridge T (1999) *High resolution nmr techniques in organic chemistry*. Amsterdam: Pergamon.
- Field LD, Sternhell S, and Kalman JR (2002) *Organic Structures from Spectra*. New York: John Wiley.
- Kalinowski HO, Berger S, and Braun S (1996) *Carbon-13 NMR spectroscopy*. New York: John Wiley.
- Pihlaja K and Kleinpeter E (1994) *Carbon-13 NMR Chemical Shifts in Structural and Stereochemical Analysis*. Weinheim: VCH.
- Tonelli AE (1989) *NMR Spectroscopy and Polymer Microstructure: The Conformational Connection*. Weinheim: VCH.
- Wehrli FW and Wirthlin T (1976) *Interpretation of Carbon-13 NMR Spectra*. London: Heyden.

## Fluorine-19

J C Lindon, Imperial College London, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Fluorine has a nuclear magnetic resonance (NMR)-active nucleus with one of the highest sensitivities, third only to proton ( $^1\text{H}$ ) and the radioactive tritium ( $^3\text{H}$ ). As it has favorable NMR properties, fluorine NMR has found widespread application in both chemistry and biology. This has resulted in a large number of reviews and books and there have been extensive compilations of solution-state  $^{19}\text{F}$  NMR chemical shifts and coupling constants in organic, organometallic, and inorganic compounds, these being very useful for confirmation of molecular structures. As well as studies on small molecules, a significant number of fluorinated polymers are important and  $^{19}\text{F}$  NMR spectroscopy can be used in their characterization, particularly the determination of polymer chain microstructure. In addition, the application of high-resolution techniques in solids, such as magic-angle spinning (MAS) and combined rotation and multiple pulse spectroscopy (CRAMPS), has expanded with a variety of applications in the literature. There are an increasing number of studies using  $^{19}\text{F}$  NMR spectroscopy in biology including the study of binding of fluorinated molecules to macromolecules, applications to membrane structure and flexibility, the elucidation of drug metabolites *in vitro*, and the study of the distribution and metabolism of compounds *in vivo*.

### Properties of the Nucleus

Fluorine possesses only one isotope,  $^{19}\text{F}$ , which has a spin  $I = 1/2$  and a sensitivity relative to  $^1\text{H}$  for an equal number of nuclei of 0.83. The magnetogyric ratio has a value of 0.94 that of  $^1\text{H}$ , so the basic resonance frequency for  $^{19}\text{F}$  is, for example, 282.2 MHz on a spectrometer that operates at a

field strength of 7.05 T with  $^1\text{H}$  NMR observation at 300 MHz.

### Sample Preparation and Spectrum Measurement

Samples for solution study are prepared in the usual manner as for  $^1\text{H}$  NMR spectroscopy. Degassing the solutions can be important because of the effect of paramagnetic oxygen on the  $^{19}\text{F}$  spin relaxation times. A number of substances have been proposed for chemical shift referencing, and for a time an external sample of trifluoroacetic acid was commonly used. Now, however, the standard reference compound is internal trichlorofluoromethane,  $\text{CFCl}_3$ , added at a few mole percent and with a fluorine chemical shift ( $\delta_{\text{F}}$ ) taken to be at 0.0 ppm. This compound is very volatile and quite unreactive, and therefore easy to remove from samples if necessary. At high-field strengths, the  $^{19}\text{F}$  NMR chemical shift is perturbed by the combined isotope effects of the various possible combinations of  $^{35}\text{Cl}$  and  $^{37}\text{Cl}$  containing molecules. This means that at best the signal is broadened and at worst an asymmetric multiplet structure from slightly different chemical shifts can be resolved. The current sign convention for fluorine chemical shifts is for negative values (denoted by  $\delta$  or ppm) to low frequency of  $\text{CFCl}_3$  (to the so-called high field) and increasing positive values to high frequency of  $\text{CFCl}_3$ . Early literature values of chemical shifts generally used the reverse sign convention and so care needs to be taken in relating recent data to older data. Some measurements have also used a number of secondary reference compounds and the conversion factors are given in Table 1.

The wide range of  $^{19}\text{F}$  NMR chemical shifts can lead to problems in data acquisition, in particular nonuniform excitation and poor digitization. For example, peaks from a typical organic compound sample containing  $\text{CFCl}_3$  and both aliphatic and aromatic fluorines can cover a spectral range of

250 ppm, which at a field strength of 11.7 T (i.e., 500 MHz for  $^1\text{H}$  NMR) corresponds to a spectral width of 117 500 Hz. To cover this frequency uniformly with a  $90^\circ$  pulse requires a pulse width of just over  $2\ \mu\text{s}$ , a value that solution-state commercial NMR spectrometers do not usually achieve. Fortunately, the need to do this can be avoided as measurements can often be focused on one region of the spectrum either by allowing signals to be aliased or by the use of digital filtering. The corresponding need for high digital resolution given the sharp  $^{19}\text{F}$  NMR lines and often highly resolved complex spin-coupled multiplets is also obvious and a 117 500 Hz spectral width with a digital resolution of only 0.5 Hz requires a Fourier transform size of 256 K words.

Particularly in organic molecules,  $^{19}\text{F}$  nuclei often show spin-spin coupling over many bonds to other nuclei. To simplify spectra when most of the couplings are between  $^{19}\text{F}$  and  $^1\text{H}$ , it is common to measure  $^1\text{H}$ -broadband-decoupled spectra. This can be quite difficult given the closeness of  $^{19}\text{F}$  and  $^1\text{H}$  NMR resonance frequencies and careful filtering of the fluorine observation channel is usually necessary.

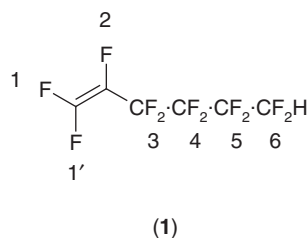
Increasingly, two-dimensional (2D) NMR methods are playing a part in spectral assignment and in polyfluoro compounds  $^{19}\text{F}$ - $^{19}\text{F}$  COSY can be very useful for resonance assignment. Similarly,  $^1\text{H}$ - $^{19}\text{F}$  correlation spectra can be used to assign couplings between  $^1\text{H}$  and  $^{19}\text{F}$ . For polyfluoro compounds,

$^{13}\text{C}$ - $^{19}\text{F}$  correlation experiments can be useful but decoupling  $^{19}\text{F}$  from  $^{13}\text{C}$  over the full range of fluorine chemical shifts is generally only feasible using techniques such as globally optimized alternating-phase rectangular pulse decoupling. On the other hand, with the advent of inverse geometry probes and inverse detection pulse sequences, detection of  $^{13}\text{C}$ - $^{19}\text{F}$  correlations at the  $^{19}\text{F}$  NMR frequency will likely prove to be very useful especially in conjunction with the use of magnetic field gradients to eliminate fluorine signals from the  $^{12}\text{C}$  containing molecules that obscure the  $^{13}\text{C}$ -coupled  $^{19}\text{F}$  NMR signals.

## $^{19}\text{F}$ NMR Chemical Shifts

$^{19}\text{F}$  NMR chemical shifts cover a wide range, if organic and inorganic substances are considered. The extensive range of  $^{19}\text{F}$  NMR chemical shifts is exemplified by the series of fluorides shown in Table 2, ranging from  $-448$  ppm for ClF to  $+865$  ppm for  $\text{F}_2\text{O}_2$ .

For organic compounds, as is well known for  $^1\text{H}$  NMR chemical shifts,  $^{19}\text{F}$  NMR chemical shifts fall into diagnostic ranges with generally negative  $\delta_{\text{F}}$  values. Major compilations of chemical shifts exist in the reviews given in the Further Reading section, most concentrating on organic compounds. To illustrate the wide range of  $^{19}\text{F}$  NMR chemical shifts in an organic compound, Figure 1 demonstrates the range ( $>100$  ppm) observed for such a relatively simple molecule. This shows the 564 MHz  $^{19}\text{F}$  NMR spectrum of 6H-perfluorohex-1-ene (1) in solution in  $\text{CDCl}_3$ .

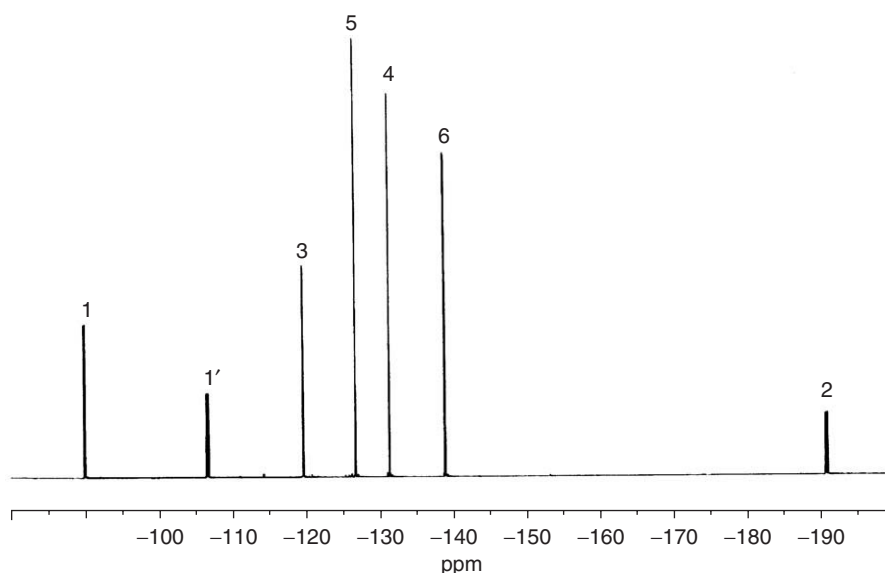


**Table 1**  $^{19}\text{F}$  chemical shifts of reference compounds, relative to  $\text{CFCl}_3$  at  $\delta_{\text{F}} = 0.0$  ppm

Compound	$\delta_{\text{F}}$ (ppm)	Compound	$\delta_{\text{F}}$ (ppm)
$\text{CF}_3\text{COOH}$	$-78.5$	<i>Cyclo</i> - $\text{C}_4\text{F}_8$	$-136.0$
$\text{C}_6\text{F}_6$	$-162.9$	<i>p</i> - $\text{C}_6\text{H}_4\text{F}_2$	$-120.0$
$\text{C}_6\text{H}_5\text{CF}_3$	$-63.9$	$\text{CF}_3\text{COCH}_3$	$-74.2$
$\text{CFCl}_2\text{CFCl}_2$	$-67.3$	$\text{CF}_3\text{C}(\text{OH})_2\text{CF}_3$	$-92.7$
<i>cyclo</i> - $\text{CF}_2\text{CF}_2\text{CCl}_2\text{CCl}_2$	$-114.1$	$\text{CF}_3\text{CCl}_3$	$-82.2$
$\text{F}_2$	$+422.9$	1-Fluoronaphthalene	$-123.0$
$\text{CF}_3\text{COCF}_3$	$-84.6$	2-Fluoronaphthalene	$-117.0$
$\text{C}_6\text{H}_5\text{F}$	$-113.2$	$\text{C}_6\text{H}_5\text{SO}_2\text{F}$	$+65.5$
$\text{CF}_2\text{Cl}_2$	$-6.9$		

**Table 2** Some representative  $^{19}\text{F}$  NMR chemical shifts in fluorides

Fluoride	$\delta_{\text{F}}$ (ppm)	Fluoride	$\delta_{\text{F}}$ (ppm)	Fluoride	$\delta_{\text{F}}$ (ppm)
$\text{AsF}_5$	$-66$	$\text{C}_6\text{F}_6$	$-163$	$\text{ReF}_7$	$345$
$[\text{AsF}_6]^-$	$-70$	$\text{CHF}=\text{CH}_2$	$-114$	$\text{SbF}_5$	$-108$
$\text{BF}_3$	$-131$	$\text{CF}_2=\text{CH}_2$	$-81$	$[\text{SbF}_6]^-$	$-109$
$[\text{BeF}_4]^-$	$-163$	$\text{CF}_2=\text{CF}_2$	$-135$	$\text{SeF}_6$	$55$
$\text{CH}_3\text{F}$	$-272$	ClF	$-448$	$[\text{SiF}_6]^{2-}$	$-127$
$\text{CH}_2\text{F}_2$	$-144$	$\text{ClF}_3$	$116; -4$	$\text{TeF}_6$	$-57$
$\text{CF}_2\text{Cl}_2$	$-8$	$\text{ClF}_5$	$247; 412$	$\text{WF}_6$	$166$
$\text{CF}_2\text{Br}_2$	$7$	$\text{F}_2\text{O}_2$	$865$	$\text{XeF}_2$	$258$
$\text{CFBr}_3$	$7$	$\text{IF}_7$	$170$	$\text{XeF}_4$	$438$
$\text{CFH}_2\text{Ph}$	$-207$	$\text{MoF}_6$	$-278$	$\text{XeF}_6$	$550$



**Figure 1** 564 MHz  $^{19}\text{F}$  NMR spectrum of 6H-perfluorohex-1-ene (1) in  $\text{CDCl}_3$  solution. Chemical shifts are quoted relative to  $\text{CFCl}_3$  and assignments are as shown.

Assignment of the chemical shifts can be made based on diagnostic coupling constants and the use of 2D homonuclear correlation spectroscopy (COSY) and these are also indicated in **Figure 1**. The chemical shifts, in ppm from  $\text{CFCl}_3$ , are  $-90.02$  (F1),  $-106.70$  (F1'),  $-190.71$  (F2),  $-119.69$  (F3),  $-131.38$  (F4),  $-126.78$  (F5), and  $-138.95$  (F6). It is notable that  $-\text{CF}_2-$  and  $=\text{CF}$  resonances cover similar chemical shift ranges, unlike the  $^1\text{H}$  NMR chemical shifts of corresponding  $-\text{CH}_2-$  and  $=\text{CH}$  groups.

Much effort has been given to evaluating substituent effects on fluorine chemical shifts especially in fluoroaromatic molecules. Additionally, because of the high sensitivity of fluorine NMR, it has been used extensively to study bonding in, and reactions of, a wide range of inorganic and organometallic species and the reader is referred to the Further Reading section for details of individual studies.

#### Solvent and Temperature Dependence of $^{19}\text{F}$ NMR Chemical Shifts

Significant solvent effects, in particular the effect of solvent dielectric constant, are known to occur for  $^{19}\text{F}$  NMR chemical shifts and this has led to significant differences between literature values for a given species and also has been the subject of considerable theoretical effort.

$^{19}\text{F}$  NMR chemical shifts are also very sensitive to temperature. Although their temperature dependence has been used extensively to evaluate free energies of activation for conversion of rotamers and ring conformers in substituted ethanes and other

conformationally mobile species, it is necessary to realize that these chemical shifts can be intrinsically temperature dependent, including the resonance of any internal reference compound.

#### Anisotropy of $^{19}\text{F}$ NMR Chemical Shifts

Although only the trace of the chemical shift tensor is measured in isotropic solution, in general  $^{19}\text{F}$  nuclei have large chemical shift anisotropies and in anisotropic media such as solids or liquid crystal solutions, which can include biological matrices such as membranes, the spectrum can be severely dependent on such parameters in that the observed chemical shift is dependent on the orientation of the molecule relative to the magnetic field. The definition of the chemical shift anisotropy in terms of molecule-based axes depends on the symmetry of the molecule under study but it can easily have a magnitude of tens or hundreds of ppm, for example, being 158 ppm in hexafluorobenzene, which has an isotropic shift of  $-163$  ppm.

#### Spin Coupling to Fluorine

##### Solvent and Temperature Dependence

Spin-spin couplings involving  $^{19}\text{F}$  can also be rather solvent dependent. For example, in simple substituted fluoroethenes, the  $^3J_{\text{HF}}$  in fluoroethene itself has values of 19.6 Hz for the *cis* coupling and 51.8 Hz for the *trans* coupling in cyclohexane solution, and these vary over a range of 2.1 and 4.6 Hz, respectively, in

different solvents. Similarly, the  $^2J_{\text{FF}}$  in 1,1-difluoroethene has a value of 30.7 Hz in cyclohexane but can vary by up to 8.8 Hz in other solvents. Coupling constants can be temperature dependent because of various aspects of molecular conformational changes, but apart from this, couplings to fluorine can have marked intrinsic temperature dependence. For example, the  $^2J_{\text{FF}}$  in 1,1-difluoroethene has a temperature coefficient of  $1.39 \text{ Hz K}^{-1}$  in hexane but only  $0.16 \text{ Hz K}^{-1}$  in dimethylformamide. Thus, care must be taken in relating changes in coupling constants to changes in molecular conformations.

### Anisotropy of Coupling Constants

In principle, NMR spectroscopy in liquid crystal solutions, or in other situations where the molecule can be oriented, allows the measurement of elements of the spin-coupling tensor other than the simple isotropic trace. Unfortunately, this anisotropy enters the NMR Hamiltonian in exactly the same fashion as the corresponding dipolar couplings that are also observed in anisotropic media and hence cannot be measured separately. Thus, estimation of this anisotropy relies on any deviation between the observed dipolar coupling and that calculated from geometrical considerations. In principle, the coupling tensor can also be asymmetric.

### $^1\text{H}$ – $^{19}\text{F}$ Spin Coupling

Couplings between  $^1\text{H}$  and  $^{19}\text{F}$  nuclei are, in general, very dependent on substitution and on electronegativity. For example, for  $^2J_{\text{HF}}$  in methanes,  $\text{CHFXY}$ , it is possible to predict the coupling to  $\sim 2 \text{ Hz}$  using the relationship:

$$^2J_{\text{HF}} = 78.76 + 8.45E_{\text{X}}E_{\text{Y}} - 16.73(E_{\text{X}} + E_{\text{Y}}) \quad [1]$$

where  $E_{\text{X}}$  and  $E_{\text{Y}}$  are the electronegativities of the substituents. A typical value where X and Y are both carbon would be  $\sim 50 \text{ Hz}$ . Vicinal  $^1\text{H}$ – $^{19}\text{F}$  spin coupling in aliphatic systems is dependent on the dihedral angle between the CH and CF bonds in a Karplus-type relationship, which has been the subject of much theoretical prediction. Long-range  $^1\text{H}$ – $^{19}\text{F}$  coupling is often observed, values being usually larger than the corresponding  $^1\text{H}$ – $^1\text{H}$  coupling.

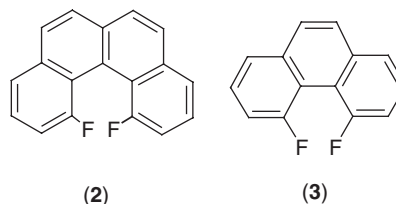
### $^{13}\text{C}$ – $^{19}\text{F}$ Spin Coupling

Carbon–fluorine coupling also shows large variations with substituents, including the one-bond coupling that has a range centered  $\sim 300 \text{ Hz}$ . Two-bond couplings are much smaller and have values up to  $\sim 10 \text{ Hz}$ . Long-range coupling is often observed in fluoroaromatic compounds but the effect of

substitution is difficult to ascertain because very few relative signs of couplings have been measured, although it is probable that  $^1J_{\text{CF}}$  is always negative and  $^2J_{\text{CF}}$  and  $^3J_{\text{CF}}$  are positive.

### $^{19}\text{F}$ – $^{19}\text{F}$ Spin Coupling

The data on coupling between  $^{19}\text{F}$  nuclei are very widespread and these can exist over many intervening bonds. Again extensive compilations of data are presented in the Further Reading section. Two-bond couplings vary widely according to geometry and structure; for instance, being 27 Hz in 1,1-difluoroethene and 87 Hz in trifluoroethene, and over 200 Hz in fluorinated cyclobutanes. Three-bond  $^{19}\text{F}$ – $^{19}\text{F}$  spin coupling shows good consistency in aromatic systems allowing structural diagnosis, but in aliphatic systems large variations can occur making its use as a structural and conformational tool questionable because substituent effects can compete with any dihedral angle effects. Four-bond couplings are sometimes larger than three-bond couplings and particularly in cyclohexanes are diagnostic in that the coupling between pairs of axial  $^{19}\text{F}$  nuclei is  $\sim 26 \text{ Hz}$ , that between pairs of equatorial  $^{19}\text{F}$  nuclei is  $\sim -6$  to  $-9 \text{ Hz}$  and the axial–equatorial coupling is near zero. The axial–axial situation has the two carbon–fluorine bonds eclipsed and, in other situations like this a large positive  $^4J_{\text{FF}}$  can be measured, such as 1,8-difluoronaphthalene where the coupling is 58.8 Hz. These large couplings have been used to postulate a ‘through-space’ mechanism of coupling as opposed to the usual recognized mechanism of through-bond coupling. Indeed, when two  $^{19}\text{F}$  nuclei are close in space but separated by many bonds, large couplings can result, for example, in (2) the  $J_{\text{FF}}$  is 43.2 Hz, whilst in (3)  $J_{\text{FF}}$  has the large value of 174 Hz.



$^{19}\text{F}$ – $^{19}\text{F}$  spin coupling in aromatic systems is useful in determining substitution patterns and structure in that  $^3J_{\text{FF}}$  is remarkably insensitive to substitution being close to  $-20 \text{ Hz}$  whilst  $^4J_{\text{FF}}$  and  $^5J_{\text{FF}}$  are very dependent on substituents. For  $^5J_{\text{FF}}$  only one substituent parameter is necessary while for  $^4J_{\text{FF}}$  three parameters are needed. The coupling constants are calculated using the equations below (where *o*, *m*, and *p* correspond to *ortho*, *meta*, and *para*,

respectively) and the parameters in Table 3 are used:

$${}^4J_{\text{FF}} = 5.8 + {}^4X(o-o) + \Sigma {}^4X(o-p) + {}^4X(m-m) \quad [2]$$

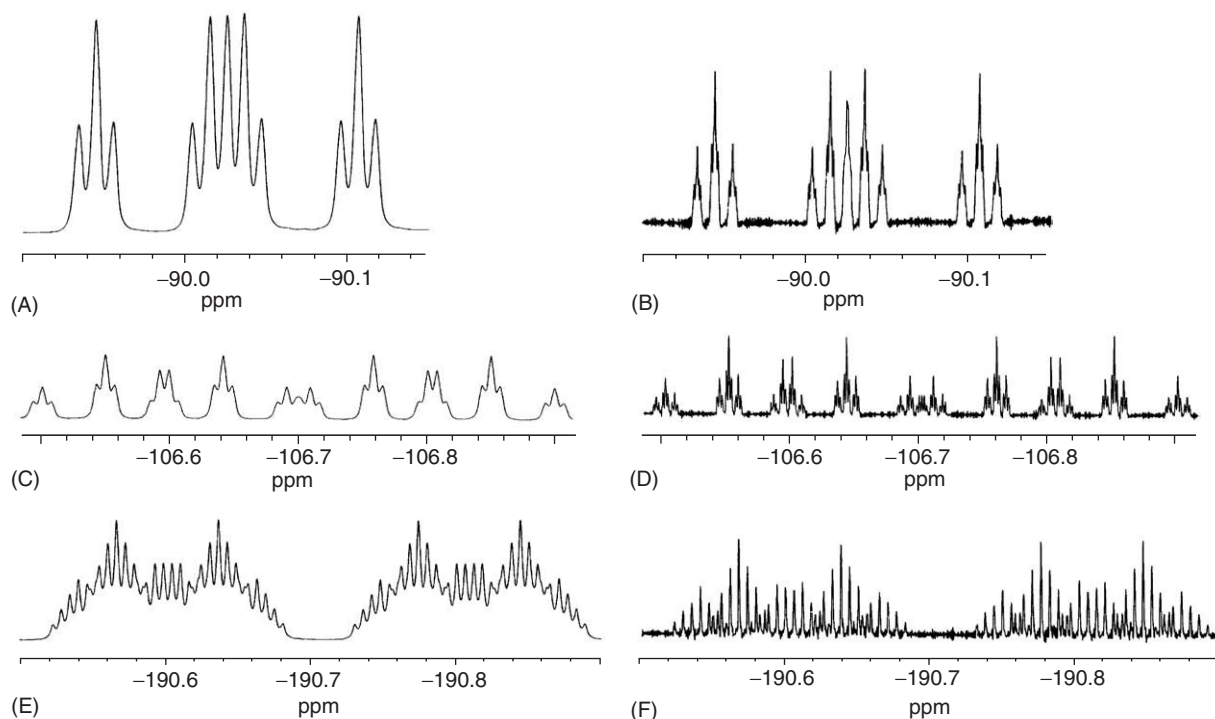
$${}^5J_{\text{FF}} = 18.1 + \Sigma {}^5X \quad [3]$$

As an example of the complexity of the  ${}^{19}\text{F}$  NMR spectra of polyfluorocompounds, Figure 2 shows expansions of the individual resonances of the olefinic fluorines from the spectrum of 6H-perfluorohex-1-ene (1)

**Table 3** Substituent parameters for *meta* and *para*  ${}^{19}\text{F}$ – ${}^{19}\text{F}$  spin coupling

Substituent	${}^4X$			${}^5X$
	Ortho-ortho	Ortho-para	Meta-meta	
NH <sub>2</sub>	7.2	−8.7	0.5	−2.9
OH	4.7	−7.7	−0.8	−3.0
CH <sub>3</sub>	0.7	−2.0	0.1	−0.3
Ph	0.6	−1.6	0.4	−0.3
H	0.0	0.0	0.0	0.0
F	−0.1	−3.8	−0.3	−3.2
Cl	−3.2	−0.4	0.7	−2.0
Br	−3.4	−0.5	0.1	−2.1
I	−3.1	0.5	0.7	−1.4
CF <sub>3</sub>	−5.5	4.2	2.0	0.0
CN	−5.7	4.3	1.9	−0.3
NO <sub>2</sub>	−8.2	5.6	2.3	−0.6
Pyridyl N	−	12.8	2.3	15.7

given in Figure 1. On the left is given the result obtained by Fourier transformation of the raw data and on the right is shown the result of resolution enhancement of the acquired data using the Lorentzian–Gaussian transformation method. The complexity of each of the bands indicates a wealth of spin–spin coupling over two to six bonds. The signal arising from F1 shows two doublet splittings corresponding to the two-bond geminal interaction of 52 Hz and a three-bond *cis* interaction of 40 Hz, together with a series of triplet splittings caused by coupling to the  $-\text{CF}_2\text{CF}_2\text{CF}_2-$  fragment over four, five, and six bonds. The band due to F1' also exhibits many couplings and the major splittings comprise doublets of 118 Hz (the three-bond *trans* coupling, which is known to be of opposite sign to the geminal and *cis* couplings between olefinic fluorines) and 52 Hz (the two-bond geminal coupling seen on the F1 resonance). In addition, there are three sets of triplets arising from coupling to the  $-\text{CF}_2\text{CF}_2\text{CF}_2-$  fragment, with the four-bond coupling being  $\sim 28$  Hz. The resonance arising from F2 is also complex showing the *trans* and *cis* couplings as doublets together with triplets from coupling to the  $-\text{CF}_2\text{CF}_2\text{CF}_2-$  fragment (four- to six-bond couplings). The exact assignment of these smaller couplings could be achieved using selective spin-decoupling experiments.



**Figure 2** Expansions of the individual resonances of the olefinic fluorines in the 564 MHz  ${}^{19}\text{F}$  NMR spectrum of 6H-perfluorohex-1-ene (1) in  $\text{CDCl}_3$  solution. On the left is shown the result of Fourier transformation of the raw data and on the right is given the result of resolution enhancement. Assignments are F1 (top), F1' (center), and F2 (bottom).



### **$^{19}\text{F}$ Spin Coupling to Other Nuclei**

The coupling between  $^{19}\text{F}$  and some heteronuclei can be very large and diagnostic of bonding situations. To give some examples of the wide range of values, the coupling to tungsten in alkoxy tungsten fluorides is  $\sim 60\text{--}70\text{ Hz}$ . In platinum complexes, the coupling between platinum and directly bonded  $-\text{CF}_3$  groups (i.e., over two bonds) can be as large as  $750\text{ Hz}$ . The one-bond coupling between phosphorus and fluorine in trifluorophosphoranes, for example,  $\text{MeS}(\text{Me})\text{PF}_3$ , and tetrafluorophosphoranes can be as large as  $1000\text{ Hz}$  whilst difluorophosphoranes generally have values between  $600$  and  $800\text{ Hz}$ . Some directly bonded phosphorus–fluorine coupling constants can be even bigger going as high as  $1284\text{ Hz}$  in  $\text{EtOPF}_2$ .  $^{29}\text{Si}$ – $^{19}\text{F}$  directly bonded coupling constants are more modest being  $\sim 200\text{--}300\text{ Hz}$  in fluorosilanes.

### **$^{19}\text{F}$ NMR Relaxation Times**

Few studies of  $^{19}\text{F}$  NMR relaxation times have been reported considering the high sensitivity of the nucleus, and this is probably due to the extensive spin coupling observed in fluorine NMR leading to extensive cross-relaxation effects and difficulty in interpreting the results. Whilst the relaxation times of  $^1\text{H}$  nuclei are usually dominated by dipole–dipole terms, and this is often also true for  $^{13}\text{C}$  relaxation in  $\text{CH}$ ,  $\text{CH}_2$ , and  $\text{CH}_3$  groups, it is quite feasible for the chemical shift anisotropy term to be a major factor in relaxation in the case of  $^{19}\text{F}$ . This term is very field dependent, the relaxation rate scaling as the square of the field strength. Some use has been made of  $^{19}\text{F}$  relaxation times ( $T_1$ ,  $T_2$ , and  $T_{1\rho}$ ) in determining dynamics in a co-polymer of tetrafluoroethene and hexafluoropropene. The chemical shift anisotropy contribution to  $^{19}\text{F}$  NMR relaxation can be a disadvantage because it can cause a field-dependent broadening, and hence loss of resolution in studies of binding between small molecules and proteins, but it can also be used to advantage in some circumstances to extract values of the chemical shift anisotropy that can be used in situations of partial ordering in order to obtain information on the averaging process.

## **Analytical Applications of $^{19}\text{F}$ NMR**

### **Chemicals and Polymers**

Because fluorine possesses a spin- $1/2$  nucleus with high NMR sensitivity, analytical applications of fluorine NMR to small molecules follow closely those using  $^1\text{H}$  NMR spectroscopy. Recently, the use of

high throughput screening for biological activity of small molecules in pharmaceutical research has entailed the use of new approaches and  $^{19}\text{F}$  NMR spectroscopy has been used for detection of target-ligand binding.

The parallel with  $^1\text{H}$  NMR spectroscopy also applies to studies in the solution state of macromolecules such as polymers where fluorinated species are commercially important.

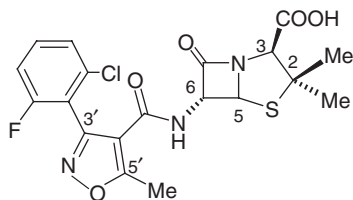
However, the difficulty of studying polymers in solution and the need to understand their solid-state properties has led to developments in solid-state NMR. To obtain high-resolution  $^{19}\text{F}$  NMR spectra of solids it is necessary to use MAS to remove the chemical shift anisotropic contribution to the line-widths, and as this alone is not sufficient to remove  $^1\text{H}$ – $^{19}\text{F}$  heteronuclear and  $^{19}\text{F}$ – $^{19}\text{F}$  homonuclear dipolar couplings, it has to be combined with other methods. Such pulse sequences exist and have been used to sharpen solid-state  $^{19}\text{F}$  NMR spectra for a number of fluorine-containing polymers. If the polymer has flexibility, then MAS alone may be sufficient and this approach has been successfully applied in fluoro-polymers of the type  $-\text{CF}_2-\text{O}-\text{CF}_2-\text{O}-\text{CF}_2-\text{CF}_2-\text{O}-$  in order to analyze for polymer sequence information and end-group effects. Similar methods have been used on dental materials such as fluoro-apatite and its derivatives in order to investigate the various fluorine environments. If it is necessary to study rigid solid materials with  $^{19}\text{F}$ – $^{19}\text{F}$  dipolar coupling, there is a requirement for very fast spinning if multiple pulse line narrowing techniques are not used and speeds in excess of  $20\text{ kHz}$  have been used on highly fluorinated co-polymers. The use of additional line narrowing methods based on multiple pulse sequences when combined with MAS is termed CRAMPS and has been applied to simple fluoro-organic molecules in order to measure shielding tensors and molecular mobility in the solid state. In addition, studies on polymers have probed solid-state morphology and molecular orientation. These techniques cannot be regarded as routine in the same way that solution-state NMR has become, but the information obtainable is difficult to acquire by other means and makes the effort worthwhile.

A number of studies have introduced fluorine atoms into amino acids and then used these to obtain fluorine-labeled proteins.  $^{19}\text{F}$  NMR spectroscopy of such proteins can then be carried out in conjunction with other multidimensional NMR experiments to obtain their 3D structures. The fluorine label can also be used as a probe to study molecular dynamics, protein structural changes caused by the introduction of the label and intermolecular interactions.

### Drug Metabolism and Cellular Applications

The use of indicator molecules containing fluorine that can be transported into cells has proved useful for monitoring intracellular environments. The  $^{19}\text{F}$  NMR shifts of various specifically designed complexant molecules are strongly dependent on pH,  $\text{Na}^+$ , or  $\text{Ca}^{2+}$  and provide a sensitive method of monitoring intracellular  $\text{Na}^+$  or  $\text{Ca}^{2+}$  levels or pH and pH gradients, and to distinguish intracellular pH from extracellular pH because of the different chemical shifts.  $^{19}\text{F}$  NMR spectroscopy has also been used to follow the metabolism and distribution of 2-fluoro-2-deoxyglucose and 3-fluoro-3-deoxyglucose used for the investigation of brain and cardiac glucose metabolism and the employment of fluorinated polyamines has been used with  $^{19}\text{F}$  NMR spectroscopy for detecting hypoxic cells in tumors.

A considerable proportion of drugs under development contain one or more fluorine atoms that act as useful tracer nuclei for NMR studies of drug metabolism when measuring NMR spectra of cell extracts or biological fluids.  $^{19}\text{F}$  NMR spectroscopy is particularly useful because of the high sensitivity of the  $^{19}\text{F}$  nucleus, its wide chemical shift range that allows resolvable chemical shifts for metabolic or chemical changes up to  $\sim 10$  bonds away from the fluorine, and, because of the negligible background from other fluorine-containing species in the environment, there are few if any interfering resonances. Also, there are no dynamic range problems for the NMR acquisition. For example, for flucloxacillin it has been shown that the parent molecule (4) and its metabolites can be detected and quantified in rat urine using a combination of 1D and 2D NMR methods. The presence of the parent drug and the major metabolite, the 5'-hydroxymethyl derivative, was confirmed by the study, and the detection of two previously unidentified metabolites, the C-5 isomers (*R* and *S*) of flucloxacillin penicilloic acid were demonstrated.



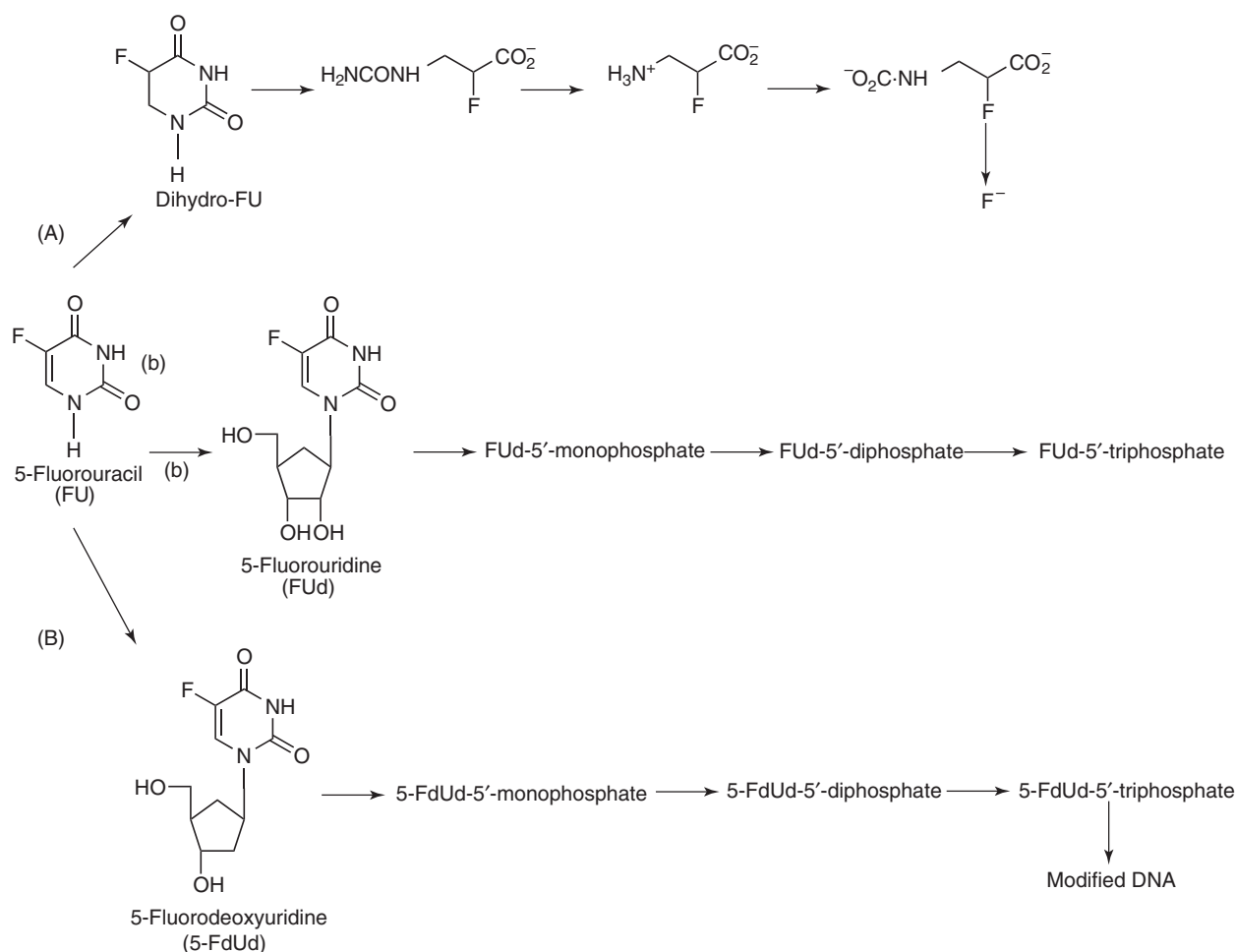
(4)

Much effort has been expended in studying the metabolism of fluorinated pyrimidines because the fluorine in the pyrimidine ring is rarely lost during metabolism and also because of the importance of this class of compounds as anticancer drugs. These

compounds, in particular 5-fluorouracil (FU), are metabolized to produce a variety of products. The number of studies published reflects the importance of this drug in the clinic and the effectiveness of  $^{19}\text{F}$  NMR spectroscopy. Thus, studies have ranged from those on body fluids such as plasma and urine, through tissue extracts to full *in vivo* investigations. FU remains an important therapy mainly for colon and breast cancers and the metabolism of FU is shown in Figure 3. Two types of process are possible, namely catabolic metabolism or anabolic processes that ultimately lead to incorporation of fluorinated species into nucleic acids. The balance between anabolism and catabolism may be different in different tissues and subjects and may be a significant factor in therapeutic benefit versus toxic side-effects. The cytotoxicity results from the anabolic route because of the inhibition by 5-fluoro-2'-deoxyuridine-5'-monophosphate of thymidylate synthetase and hence DNA synthesis and by incorporation of 5-fluorouracil components into RNA and DNA.  $^{19}\text{F}$  NMR spectra from plasma from patients on FU infusion therapy showed the resonances from FU itself and peaks from three expected metabolites (dihydro-FU, 2-fluoro-3-ureidopropionic acid, and 2-fluoro-3-aminopropionate) and in some cases fluoride ion was also detected. Detailed pharmacokinetics was also possible. In urine, a similar picture emerges with over 97% of the observed  $^{19}\text{F}$  NMR spectral intensity due to the same species, although *N*-carboxy-2-fluoro-3-aminopropionate and two other unidentified species were observed. Human biliary excretion has also been studied and led to the characterization of 2-fluoro-3-aminopropionate complexed with bile acids. FU metabolism has also been studied in *Escherichia coli* bacteria, murine, and human tumor cells, excised tissue samples from patients receiving FU therapy where it was shown that uptake and anabolic metabolism of FU was lower in liver metastases compared to normal liver. Finally, *in vivo* studies have been conducted using both rodents and humans, and in particular in the clinical situation, surface coil NMR spectroscopy has been used to study the catabolism of FU to 2-fluoro-3-aminopropionate in the liver with a time resolution of 9 min.

### Membranes and Macromolecule Binding Studies

$^{19}\text{F}$  NMR spectroscopy can be used to study membrane structure and dynamics by replacing hydrogens on a phospholipid molecule by fluorines. Because of the favorable NMR properties of fluorine, it is an ideal probe nucleus as an adjunct to using  $^{13}\text{C}$  or  $^2\text{H}$  incorporation for this application. In fact, it is preferable to replace a  $\text{CH}_2$  group by a  $\text{CF}_2$  group rather



**Figure 3** The fluorinated derivatives of 5-fluorouracil involved in (A) catabolic and (B) anabolic metabolism.

than a CHF group because it is then often possible to observe  $^{19}\text{F}$ – $^{19}\text{F}$  dipolar coupling and also no additional chirality is incorporated. Application have included sonicated model membranes, oriented synthetic bilayers, liposomes that give powder-type spectra, and biosynthetically derived bacterial membranes grown using media containing fluorinated fatty acids. It is possible to use the derived NMR parameters to yield information on order parameters along the hydrocarbon chains, to investigate the phase diagram for the heterogeneous systems, and to study the effect of added cholesterol or proteins, for example, on the bilayer dynamics and order. In oriented systems it is possible to use the  $^{19}\text{F}$  chemical shift anisotropy to advantage to yield ordering information and it can also be feasible to measure the  $^{19}\text{F}$ – $^{19}\text{F}$  dipolar coupling, which leads directly to the degree of order along the inter-fluorine axis. It has been possible to synthesize fluorinated fatty acids where the fluorine probes are placed at different parts of the aliphatic chains and incorporate these into membranes. It has to be recognized that these

fluorinated probe molecules are not natural phospholipid components and the question has to be asked how great a perturbation they introduce. The C–F bond is  $\sim 0.02$ – $0.03$  nm longer than a C–H bond and as well as any electronic effects of the substitution there have to be steric consequences. For example, relative to dimyristoylglycerophosphorylcholine, the compounds which have 4,4- $\text{F}_2$ , 8,8- $\text{F}_2$ , and 12,12- $\text{F}_2$  substituents have different liquid crystal to isotropic transition temperatures by +5.0, –7.4, and –3.8 K, respectively.

Fluorine-containing ligands have proved popular for probing the binding of small molecules to proteins. There are the usual advantages of using  $^{19}\text{F}$  NMR spectroscopy but also the disadvantages of the potential steric and electronic changes to the local binding structure by introducing the fluorine substitution and the difficulty of interpreting the  $^{19}\text{F}$  NMR relaxation times because of the significant chemical shift anisotropy contribution. The effect of this on  $T_2$  means a rapidly increasing linewidth at higher magnetic field strengths. Studies have been carried

out on ligands that are in exchange with a binding site on the protein (fast, slow, or intermediate on the NMR timescale) or ligands that are covalently bonded or introduced into the structure of the protein, for example, as fluorinated amino acids. The parameters that can be measured and are influenced by binding are chemical shifts, integrals,  $T_1$ ,  $T_2$ , NOEs, molecular diffusion coefficients measured by special NMR pulse sequences, and possibly spin coupling constants, if the linewidths are narrow enough to resolve them. In addition, exchange rates can be calculated based on full lineshape analysis, transfer of saturation experiments or the Carr–Purcell–Meiboom–Gill spin-echo sequence. There is the usual difficulty of interpreting results quantitatively if there is more than one binding site and if weak nonspecific binding also takes place. Through the use of  $^{19}\text{F}$ – $^1\text{H}$  NOE experiments it is possible to provide evidence of the conformation of the bound molecule and by using a 2D NOE sequence with a short enough mixing time and following the intensity of NMR cross-peaks, spin diffusion effects can be minimized.

It has been observed that models for dynamics of a ligand at a binding site which can explain  $T_1$  and  $^{19}\text{F}$ – $^1\text{H}$  NOE data often underestimate the linewidths or  $T_2$ . This additional line-broadening is thought to arise because the ligand can exchange between several conformational states in the ligand–protein structure and the  $^{19}\text{F}$  label can experience several different chemical shifts that are not averaged by the exchange.

The  $^{19}\text{F}$  NMR chemical shift changes induced by protein binding are usually to high frequency by up to 9 ppm although larger shifts and shifts to low frequency are known. This shift can occur as a result of electrostatic effects, van der Waals interactions, dipolar-induced electric fields, or magnetic anisotropy effects caused by adjacent carbonyl groups or aromatic rings. The shifts are temperature dependent and any deviation from a linear high-frequency shift or a shift greater than  $0.01\text{ ppm K}^{-1}$  can indicate a protein conformational change or the onset of an exchange process. If the solvent is changed from  $\text{H}_2\text{O}$  to  $\text{D}_2\text{O}$ , a fluorine probe nucleus can experience an isotope shift, typically  $0.2$ – $0.3\text{ ppm}$  to low frequency. This effect can be used to probe whether the protein-bound ligand is exposed to solvent or solvent-accessible exchangeable protons.

The consequences of introducing the fluorine probe have to be considered. Replacement of a C–H by a C–F makes the ligand more hydrophobic and hence this can affect binding properties. Evidence so far accumulated indicates that fluorine substitution does not, in general, have a significant influence on the tertiary structure of proteins.

### ***In Vivo* $^{19}\text{F}$ NMR Spectroscopy**

Localized *in vivo*  $^{19}\text{F}$  NMR spectroscopy has been used to monitor the distribution of drugs. Unlike  $^1\text{H}$  NMR spectroscopy, there is no need to suppress the large water resonance, and the relatively high sensitivity of fluorine NMR makes this an attractive approach. The advantage of spectroscopy is that it yields both spatial and chemical information at the same time. For example, the methods have been applied to the neuroleptic drug trifluoroperazine in the rat brain and extensive studies have concentrated on the uptake and elimination of fluorinated anesthetics because of the relatively high concentrations that can be achieved *in vivo*. Detailed pharmacokinetic studies have also been undertaken using  $^{19}\text{F}$  NMR *in vivo*, especially for 5-FU.

The anesthetic compound halothane,  $\text{CF}_3\text{CHBrCl}$ , is metabolized in the liver and this process has been monitored using  $^{19}\text{F}$  NMR spectroscopy, both *in vivo* and in dissected rat livers. Initially, several resonances were observed *in vitro* which did not correspond to trifluoroacetate or fluoride, two known metabolic products. Using surface coil NMR experiments *in vivo*, it has been possible to show that the halothane has a first-order decay in the liver with a half-life of 2.5 h, that trifluoroacetate was observed 8 h after dosing, and that induction of liver enzymes shortened this period considerably. No other resonances were observed *in vivo* but in liver extracts other molecules were also detected, namely, fluoride,  $\text{CF}_3\text{CH}_2\text{Cl}$ , and  $\text{CHF}_2\text{CH}_2\text{Cl}$ .

The  $^{19}\text{F}$  relaxation times of the fluorine nuclei in perfluorocarbons have been shown to be linearly inversely dependent on oxygen concentration due to the paramagnetic effect of oxygen. This has been shown to be of potential use for monitoring the level of  $p\text{O}_2$  in tissues and organs.

### **$^{19}\text{F}$ Magnetic Resonance Imaging**

The potential of using the fluorine resonances of perfluoro-aliphatic compounds to carry out NMR imaging has been explored. One material, a mixture of perfluorodecalin and perfluorotripropylamine, was developed as a blood substitute and has been undergoing clinical trials as a radiotherapy potentiators. Using these molecules, the time-course of the distribution and retention of the compounds in tumor-bearing mice has been explored and the results correlated with  $^1\text{H}$  magnetic resonance imaging (MRI) results. The  $^{19}\text{F}$  MRI measurements demonstrated the preferential location of the substances in the spleen and liver and a combination of imaging and *in vivo* spectroscopy showed that although the compounds left the blood stream within hours, they were retained in tissues at later times and the

compounds accumulated in tumors in a dose-dependent fashion, at levels 10–100 times lower than those in the spleen or liver. Other studies include one that followed the gastrointestinal transport in the rat of a nondisintegrating pellet containing a fluorinated drug derivative. Also, the detailed distribution of a fluorinated drug has been attempted in that  $^{19}\text{F}$  NMR images have been obtained showing the distribution of an antifolate anticancer agent in the mouse. This compound, an inhibitor of thymidylate synthetase, was administered intravenously and images obtained with a time resolution of 4 and 20 min, respectively, for 2D and 3D imaging. The images showed high concentrations of the drug in the gall bladder, urinary bladder, and small intestine. The methods will probably be limited by the acceptable concentrations of the drugs that can be achieved in the body but with technology and methods continuing to develop, the field promises to become very useful.

See also: **Drug Metabolism:** Metabolite Isolation and Identification. **Nuclear Magnetic Resonance Spectroscopy:** Instrumentation. **Nuclear Magnetic Resonance Spectroscopy-Applicable Elements:** Hydrogen Isotopes. **Nuclear Magnetic Resonance Spectroscopy Applications:** Pharmaceutical. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Multidimensional Proton; Surface Coil; *In Vivo* Spectroscopy Using Localization Techniques. **Polymers:** Synthetic.

## Further Reading

Bachert P (1998) Pharmacokinetics using fluorine NMR *in vivo*. *Progress in NMR Spectroscopy* 33: 1–56.

Dalvit C, Fagerness PE, Hadden DTA, Sarver RW, and Stockman BJ (2003) Fluorine NMR experiments for high-throughput screening: Theoretical aspects, practical considerations, and range of applicability. *Journal of the American Chemical Society* 125: 7696–7703.

Emsley JW and Phillips L (1971) Fluorine chemical shifts. *Progress in NMR Spectroscopy* 7: 1–523.

Emsley JW and Lindon JC (1975) *NMR Spectroscopy in Liquid Crystal Solvents*. Oxford: Pergamon.

Emsley JW, Phillips L, and Wray V (1976) Fluorine coupling constants. *Progress in NMR Spectroscopy* 10: 83–756.

Fields R (1972) Fluorine-19 nuclear magnetic resonance spectroscopy. *Annual Reports on NMR Spectroscopy* 5A: 99–304.

Gerig JT (1994) Fluorine NMR of proteins. *Progress in NMR Spectroscopy* 26: 293–370.

Harris RK and Jackson P (1991) High-resolution fluorine-19 magnetic resonance of solids. *Chemical Reviews* 91: 1427–1440.

Ho C, Dowd SR, and Post JFM (1985) Fluorine-19 NMR investigations of membranes. *Current Topics in Bioenergetics* 14: 53–95.

Jenkins BG (1991) Detection of site-specific binding and co-binding of ligands to macromolecules using fluorine-19 NMR. *Life Sciences* 48: 1227–1240.

Malet-Martino MC and Martino R (1989) The application of nuclear magnetic resonance spectroscopy to drug metabolism studies. *Xenobiotica* 19: 583–607.

Wray V (1980) Fluorine-19 nuclear magnetic resonance spectroscopy (1976–1978). *Annual Reports on NMR Spectroscopy* 10B: 1–507.

Wray V (1983) Fluorine-onenine nuclear magnetic resonance spectroscopy (1979–1981). *Annual Reports on NMR Spectroscopy* 14: 1–406.

## Nitrogen-15

**A Lyčka**, Research Institute for Organic Syntheses, Pardubice, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Nitrogen is a very important element in organic, inorganic, and organometallic chemistry as well as in biochemistry. It combines with practically all elements of the periodic table. Nitrogen has oxidation numbers from  $-3$  to  $5$ , coordination numbers from  $1$  to  $6$ , and bond order from  $1$  to  $3$ . All these properties can be studied by means of nitrogen NMR spectroscopy.

There are two naturally occurring isotopes of nitrogen,  $^{14}\text{N}$  and  $^{15}\text{N}$ . The  $^{14}\text{N}$  isotope has almost

100% natural abundance (Table 1) and its theoretical relative sensitivity is much higher than that of  $^{15}\text{N}$ . However,  $^{14}\text{N}$  NMR spectroscopy suffers from the fact that  $^{14}\text{N}$  is a quadrupolar nucleus ( $I = 1$ ). Despite the relatively small quadrupole moment ( $Q = 0.0017 \times 10^{-28} \text{ m}^2$ ),  $^{14}\text{N}$  NMR signals are commonly broad ( $w_{1/2}$  are tens to thousands of Hz) except for highly symmetrical environments (e.g.,  $\text{NR}_4^+$ ,  $\text{NO}_3^-$ , etc.) and some special cases (e.g., isocyanates R-NC) in which the signals are reasonably narrow.

In  $^{15}\text{N}$  NMR spectra, the line widths are comparable to those for  $^1\text{H}$  and  $^{13}\text{C}$ . Thus, in spite of very low sensitivity, low natural abundance, negative magnetogyric ratio, and the often long relaxation times of the  $^{15}\text{N}$  nucleus (Table 1),  $^{15}\text{N}$  spectroscopy has been the technique of choice.



compounds accumulated in tumors in a dose-dependent fashion, at levels 10–100 times lower than those in the spleen or liver. Other studies include one that followed the gastrointestinal transport in the rat of a nondisintegrating pellet containing a fluorinated drug derivative. Also, the detailed distribution of a fluorinated drug has been attempted in that  $^{19}\text{F}$  NMR images have been obtained showing the distribution of an antifolate anticancer agent in the mouse. This compound, an inhibitor of thymidylate synthetase, was administered intravenously and images obtained with a time resolution of 4 and 20 min, respectively, for 2D and 3D imaging. The images showed high concentrations of the drug in the gall bladder, urinary bladder, and small intestine. The methods will probably be limited by the acceptable concentrations of the drugs that can be achieved in the body but with technology and methods continuing to develop, the field promises to become very useful.

See also: **Drug Metabolism:** Metabolite Isolation and Identification. **Nuclear Magnetic Resonance Spectroscopy:** Instrumentation. **Nuclear Magnetic Resonance Spectroscopy-Applicable Elements:** Hydrogen Isotopes. **Nuclear Magnetic Resonance Spectroscopy Applications:** Pharmaceutical. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Multidimensional Proton; Surface Coil; *In Vivo* Spectroscopy Using Localization Techniques. **Polymers:** Synthetic.

## Further Reading

Bachert P (1998) Pharmacokinetics using fluorine NMR *in vivo*. *Progress in NMR Spectroscopy* 33: 1–56.

- Dalvit C, Fagerness PE, Hadden DTA, Sarver RW, and Stockman BJ (2003) Fluorine NMR experiments for high-throughput screening: Theoretical aspects, practical considerations, and range of applicability. *Journal of the American Chemical Society* 125: 7696–7703.
- Emsley JW and Phillips L (1971) Fluorine chemical shifts. *Progress in NMR Spectroscopy* 7: 1–523.
- Emsley JW and Lindon JC (1975) *NMR Spectroscopy in Liquid Crystal Solvents*. Oxford: Pergamon.
- Emsley JW, Phillips L, and Wray V (1976) Fluorine coupling constants. *Progress in NMR Spectroscopy* 10: 83–756.
- Fields R (1972) Fluorine-19 nuclear magnetic resonance spectroscopy. *Annual Reports on NMR Spectroscopy* 5A: 99–304.
- Gerig JT (1994) Fluorine NMR of proteins. *Progress in NMR Spectroscopy* 26: 293–370.
- Harris RK and Jackson P (1991) High-resolution fluorine-19 magnetic resonance of solids. *Chemical Reviews* 91: 1427–1440.
- Ho C, Dowd SR, and Post JFM (1985) Fluorine-19 NMR investigations of membranes. *Current Topics in Bioenergetics* 14: 53–95.
- Jenkins BG (1991) Detection of site-specific binding and co-binding of ligands to macromolecules using fluorine-19 NMR. *Life Sciences* 48: 1227–1240.
- Malet-Martino MC and Martino R (1989) The application of nuclear magnetic resonance spectroscopy to drug metabolism studies. *Xenobiotica* 19: 583–607.
- Wray V (1980) Fluorine-19 nuclear magnetic resonance spectroscopy (1976–1978). *Annual Reports on NMR Spectroscopy* 10B: 1–507.
- Wray V (1983) Fluorine-onenine nuclear magnetic resonance spectroscopy (1979–1981). *Annual Reports on NMR Spectroscopy* 14: 1–406.

## Nitrogen-15

**A Lyčka**, Research Institute for Organic Syntheses, Pardubice, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Nitrogen is a very important element in organic, inorganic, and organometallic chemistry as well as in biochemistry. It combines with practically all elements of the periodic table. Nitrogen has oxidation numbers from  $-3$  to  $5$ , coordination numbers from  $1$  to  $6$ , and bond order from  $1$  to  $3$ . All these properties can be studied by means of nitrogen NMR spectroscopy.

There are two naturally occurring isotopes of nitrogen,  $^{14}\text{N}$  and  $^{15}\text{N}$ . The  $^{14}\text{N}$  isotope has almost

100% natural abundance (Table 1) and its theoretical relative sensitivity is much higher than that of  $^{15}\text{N}$ . However,  $^{14}\text{N}$  NMR spectroscopy suffers from the fact that  $^{14}\text{N}$  is a quadrupolar nucleus ( $I = 1$ ). Despite the relatively small quadrupole moment ( $Q = 0.0017 \times 10^{-28} \text{ m}^2$ ),  $^{14}\text{N}$  NMR signals are commonly broad ( $w_{1/2}$  are tens to thousands of Hz) except for highly symmetrical environments (e.g.,  $\text{NR}_4^+$ ,  $\text{NO}_3^-$ , etc.) and some special cases (e.g., isocyanates R-NC) in which the signals are reasonably narrow.

In  $^{15}\text{N}$  NMR spectra, the line widths are comparable to those for  $^1\text{H}$  and  $^{13}\text{C}$ . Thus, in spite of very low sensitivity, low natural abundance, negative magnetogyric ratio, and the often long relaxation times of the  $^{15}\text{N}$  nucleus (Table 1),  $^{15}\text{N}$  spectroscopy has been the technique of choice.

**Table 1** Basic properties important from the viewpoint of NMR spectroscopy of selected nuclei

Isotope	Natural abundance (%)	Spin number <i>I</i>	Magnetogyric ratio ( $\gamma/10^7$ ) <sup>a</sup>	RS <sup>b</sup>	RF <sup>c</sup> (MHz)
<sup>1</sup> H	99.99	1/2	26.751	5680	500.0
<sup>13</sup> C	1.11	1/2	6.726	1	125.8
<sup>14</sup> N	99.635	1	1.932	5.69	36.1
<sup>15</sup> N	0.365	1/2	−2.711	0.022	50.6

<sup>a</sup>Magnetogyric ratio  $\gamma$  (rad T<sup>−1</sup> s<sup>−1</sup>).<sup>b</sup>RS, relative sensitivity at natural isotopic abundance level to <sup>13</sup>C.<sup>c</sup>RF, resonance frequency in MHz at a field of 11.7 T.

## Instrumental and Experimental Requirements for <sup>15</sup>N NMR

Nitrogen-15 NMR spectra are obtained almost exclusively by the pulsed Fourier transform technique. An NMR spectrometer must be equipped with either a single <sup>15</sup>N probe or a multinuclear tuneable probe and appropriate frequency synthesizer and detector. Instruments with magnetic field strengths as high as possible and wide NMR tube diameters are recommended for direct NMR measurements because of the low sensitivity of <sup>15</sup>N NMR measurements. Small and negative magnetogyric ratios can lead to strong enhancements of signals corresponding to proton-bearing nitrogens due to the nuclear Overhauser effect (NOE). A theoretical NOE of −4.93 is obtained (see equation below) for the case where the mechanism of spin–lattice relaxation *T*<sub>1</sub> is solely dipole–dipole:

$$\text{NOE} = 0.5 \frac{\gamma_{\text{H}} T_1}{\gamma_{\text{N}} T_1^{\text{DD}}} \quad [1]$$

where  $\gamma_{\text{H}}$  and  $\gamma_{\text{N}}$  are the magnetogyric ratios of hydrogen and nitrogen-15, respectively, and *T*<sub>1</sub><sup>DD</sup> is the dipole–dipole relaxation time. Partial dipolar relaxation for tertiary nitrogens may result in null signals (NOE = −1), as follows from the equation:

$$\frac{M_z}{M_z^0} = 1 + 0.5 \frac{\gamma_{\text{H}} T_1}{\gamma_{\text{N}} T_1^{\text{DD}}} \quad [2]$$

where *M*<sub>z</sub> is the *z*-magnetization and *M*<sub>z</sub><sup>0</sup> is the *z*-magnetization at thermal equilibrium. In the latter case, an application of gated proton decoupling (decoupler ‘on’ during acquisition) is recommended. Chromium(III) acetylacetonate is often used as the so-called relaxation reagent to shorten relaxation times and suppress the dipole–dipole relaxation mechanism. Obtaining signals of all nitrogen atoms (differing considerably in relaxation mechanisms) in <sup>15</sup>N NMR spectra can be achieved in some cases only by using several measurement techniques.

Nitrogen-15 is an ideal nucleus for the application of methods based on polarization transfer such as

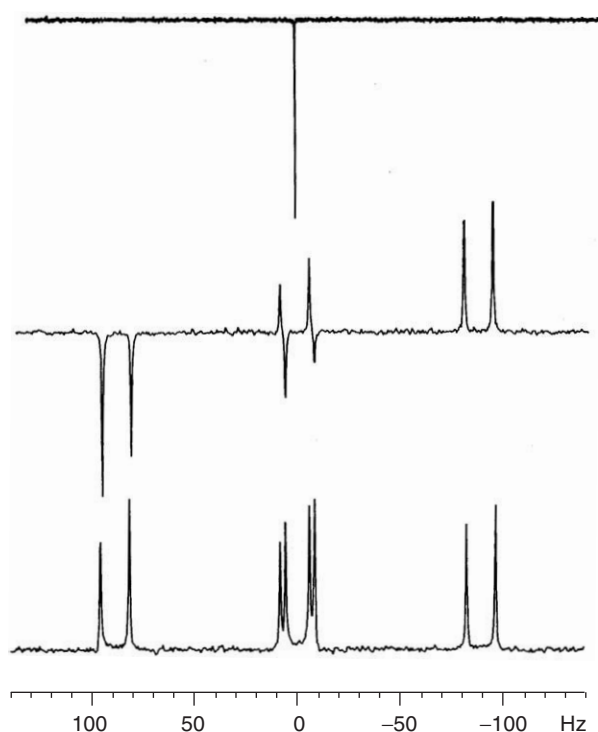
**Table 2** Comparison of theoretical sensitivities of various NMR experiments as a function of different excitation and detections<sup>a</sup>

Experiment	S/N <sup>b</sup>	Amplification factor	<sup>15</sup> N <sup>c</sup>
(a) Direct observation	$\gamma_{\text{N}}(\gamma_{\text{N}})^{3/2}$		1.0
(b) INEPT	$\gamma_{\text{H}}(\gamma_{\text{N}})^{3/2}$		
Compared with (a)		$(\gamma_{\text{H}}/\gamma_{\text{N}})^{3/2}$	9.9
(c) Inverse INEPT	$\gamma_{\text{N}}(\gamma_{\text{H}})^{3/2}$		
Compared with (a)		$(\gamma_{\text{H}}/\gamma_{\text{N}})^{3/2}$	31.0
Compared with (b)		$(\gamma_{\text{H}}/\gamma_{\text{N}})^{1/2}$	3.1
(d) Inverse shift correlation	$\gamma_{\text{H}}(\gamma_{\text{H}})^{3/2}$		
Compared with (a)		$(\gamma_{\text{H}}/\gamma_{\text{N}})^{5/2}$	306.0
Compared with (b)		$(\gamma_{\text{H}}/\gamma_{\text{N}})^{3/2}$	31.0

<sup>a</sup>From Kessler H, Gehrke M, and Griesinger C (1988) Two-dimensional NMR spectroscopy: background and overview of the experiments. *Angewandte Chemie (International Edition in English)* 27: 490–536.<sup>b</sup>Signal-to-noise ratio.<sup>c</sup>Relative sensitivity.

intensive nuclei enhanced by polarization transfer (INEPT) or distortionless enhancement by polarization transfer (DEPT). The INEPT technique is preferred, especially when polarization transfer is based in small long-range coupling constants, because the sequence is shorter and the loss of magnetization during the evolution and refocusing period is lower. Table 2 shows the amplification factors that can be achieved with respect to the standard experiment. The pulse repetition time in these one-dimensional pulse sequences is governed by much shorter proton relaxation times than those of <sup>15</sup>N, which leads to an additional improvement in the signal-to-noise ratio in a given time compared with standard experiments. Both proton-coupled and proton-decoupled spectra can be obtained using the above-mentioned techniques, depending on a modification of sequences (Figure 1).

Two-dimensional (2D) <sup>1</sup>H–<sup>15</sup>N shift-correlated spectra are used for the assignment of <sup>15</sup>N resonance based on known <sup>1</sup>H chemical shifts, particularly in complicated synthetic and natural products. Measurements via both one-bond and long-range coupling constants *J* (<sup>15</sup>N, <sup>1</sup>H) are possible. Classical 2D



**Figure 1**  $^{15}\text{N}$  NMR (50, 76 MHz) spectra of formamide: normal spectrum with proton noise decoupling (upper trace), INEPT spectrum (middle trace), and DEPT spectrum (bottom trace) ( $^1J(^{15}\text{N}, ^1\text{H}) = 90.3$  and  $88.0$  Hz,  $^2J(^{15}\text{N}, ^1\text{H}) = 13.6$  Hz).

$^1\text{H}$ - $^{15}\text{N}$  shift-correlated experiments suffer from the low sensitivity of the  $^{15}\text{N}$  nucleus, but this can be overcome by the use of inverse detection. A special probe is required for this type of experiment with the orientation of the coils in the probe reversed, i.e., the proton coil is closer to the sample. Inverse detection measurements give the highest amplification factor (Table 2) and very dilute solutions of complex molecules can be measured even at the natural abundance level of  $^{15}\text{N}$ .

The problem of  $^1\text{H}$ - $^{15}\text{N}$  shift-correlated spectra measurements using inverse detection consists in the fact that 99.635% of signal intensity corresponding to the  $^1\text{H}$ - $^{14}\text{N}$  pairs must be very efficiently suppressed and only 0.365% of original proton signal intensity corresponding to the  $^1\text{H}$ - $^{15}\text{N}$  pairs is observed. Magnetic field gradients are successfully used for this purpose. There are three very important reasons for their application: (1) gradients are used to suppress  $t_1$  noise and are highly desirable for  $^1\text{H}$ - $^{15}\text{N}$  shift-correlated spectra measurements at the natural abundance of the  $^{15}\text{N}$  isotope; (2) they can be used to select coherence pathways; and (3) magnetic field gradients are used for solvent suppression, which is especially important when biological samples at low concentrations (typically polypeptides, etc.) are measured in a mixture of 90%  $\text{H}_2\text{O}$ /10%  $\text{D}_2\text{O}$ .

Probably the most important benefit obtained from an application of magnetic field gradients comes from  $t_1$ -noise suppression. Magnetic field gradients are used very frequently in practically all pulse sequences starting from simple  $^1\text{H}$ - $^{15}\text{N}$  shift-correlated experiments to very sophisticated two- and especially three-dimensional measurements.

From the viewpoint of solid-state NMR,  $^{15}\text{N}$  is one among the 'dilute' nuclear spin systems. The  $^{15}\text{N}$  cross-polarization-magic angle spinning (CP-MAS) NMR technique is recommended so that the sensitivity of measurement may be increased and the signals narrowed.

The use of  $^{15}\text{N}$  enrichment (from the natural abundance of 0.365% up to ~100% of  $^{15}\text{N}$ ) should be mentioned as a tool for (1) a considerably shortening experimental time and (2) unequivocal signal assignment (provided selective labeling has been performed).

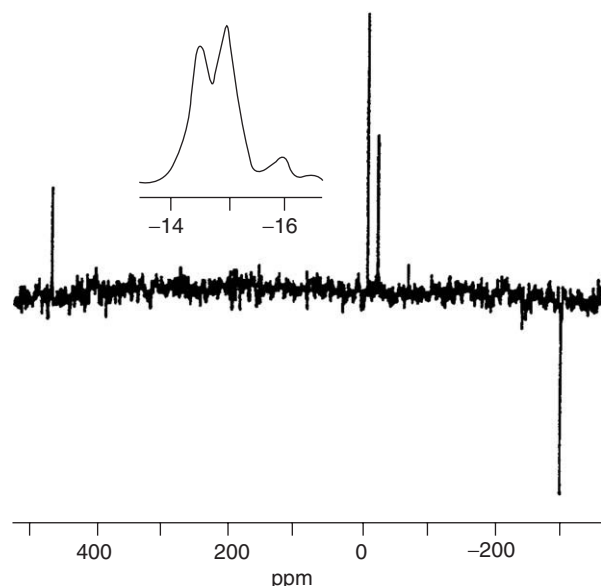
Nitrogen-15 chemical shifts are mostly referenced to an external standard, usually nitromethane, placed in a coaxial capillary. The position of nitromethane resonance produces mostly negative values of  $^{15}\text{N}$  chemical shifts. A wide variety of other standards have been used ( $\text{NH}_3$ ,  $\text{NH}_4\text{Cl}$ ,  $\text{NH}_4\text{NO}_3$ ,  $\text{HNO}_3$ , etc.). Liquid ammonia is frequently used, as it has the advantage that practically all  $^{15}\text{N}$  chemical shifts are positive. The difference in  $^{15}\text{N}$  chemical shifts of nitromethane (being shifted to the higher frequency) and liquid ammonia is 381.7 ppm. In the solid state, the  $^{15}\text{N}$  chemical shifts are referenced to a standard solid compound by sample replacement. Two sign conventions have been used. High-frequency shifts with respect to a standard are taken as (1) positive in agreement with the IUPAC recommendation or (2) negative in line with shielding constants.

### Illustrative Examples of the Use of $^{15}\text{N}$ NMR in Structure Determination

The  $^{15}\text{N}$  NMR spectrum of 4,6-dinitro-2-nitrosoaniline (Figure 2) is a good example of the use of the technique as the compound shows nearly the whole range of  $^{15}\text{N}$  chemical shifts (~900 ppm for diamagnetic compounds). Typical ranges of  $^{15}\text{N}$  NMR chemical shifts in some nitrogen-containing compounds are collected in Table 3. Half-line widths in  $^{15}\text{N}$  NMR spectra obtained from compounds in solution are usually very narrow (nearly several hundredths or tenths of ppm) and many types of functional groups can be distinguished directly.

The protonation shifts are typical of various compounds: slight deshielding in alkyl amines (<20 ppm), usually slight shielding in arylamines

and strong shielding ( $>100$  ppm) in compounds containing the arrangements  $\text{--N=}$  and  $\equiv\text{N}$  have been observed. In compounds containing several



**Figure 2**  $^{15}\text{N}$  NMR (10.17 MHz) spectrum of 4,6-dinitro-3-nitrosoaniline in dimethyl sulfoxide ( $\delta(^{15}\text{N}) = -473.9$  (NO), 0.0) ( $\text{CH}_3\text{NO}_2$ , reference compound,  $-14.4$  and  $-14.9$  ( $2 \times \text{NO}_2$ ) and  $-293.9$  ( $\text{NH}_2$ )). (Macháček V, Hassanien MMM, Štěrba V, and Lyčka A (1987) Formation of 4,6-dinitro-2-nitrosoaniline by intermolecular redox reaction of esters and amides of 2-(2,4,6-trinitroaniline) carboxylic acids. *Journal of the Chemical Society, Perkin Transaction II*, p. 869; reproduced by permission of The Royal Society of Chemistry.)

nitrogen atoms it is possible to determine site(s) of protonation and  $\text{pK}_a$  from the pH dependence of the  $^{15}\text{N}$  chemical shifts.

Nitrogen-15 chemical shifts are an ideal tool for tautomeric equilibria studies. Nitrogen atoms are directly involved in various tautomeric systems and  $^{15}\text{N}$  chemical shifts reflect the changes in equilibria in a much better way than either  $^1\text{H}$  or  $^{13}\text{C}$  chemical shifts. Differences in  $\delta^{15}\text{N}$  of pure tautomeric forms can reach several hundreds of ppm (nitroso-isonitroso, azo-hydrazone tautomerism, etc.). Similarly, large changes in  $^{15}\text{N}$  chemical shifts with respect to those in parent compounds have been observed in metal complexes, these changes being both positive and negative.

A set of typical coupling constants  $^nJ(^{15}\text{N}, ^1\text{H})$  and  $^nJ(^{15}\text{N}, \text{X})$ , which can be used both for assignment purpose as well as for polarization transfer techniques, are shown in Table 4.  $^{15}\text{N}$  enrichment is usually necessary for measurements of  $^nJ(^{15}\text{N}, \text{X})$  coupling constants when X nuclei are of low natural abundance. Some coupling constants  $^nJ(^{15}\text{N}, \text{X})$  describe the 's character' of bonds, reflect coordination, site of protonation, tautomeric equilibria, or solvent effects, and are very useful in *syn-anti* and *cis-trans* stereochemistry and conformational analysis. In many cases, the absolute values of the  $^nJ(^{15}\text{N}, \text{X})$  coupling constants for characteristics to be distinguished represent the lowest and highest values of a typical range and the differentiation and/or relative quantification of a property is easy.

**Table 3** Characteristic ranges of  $^{15}\text{N}$  chemical shifts in various types of compounds<sup>a</sup>

Compound	$\delta(^{15}\text{N})$	Compound	$\delta(^{15}\text{N})$
Alkylamines	$-310$ to $-380$	Triazenes $\text{RN}=\text{N}-\text{NR}'_2$	
Arylamines	$-280$ to $-330$	$\text{RN}=\text{}$	$-20$ to $-40$
Hydrazines	$-260$ to $-340$	$=\text{N}-$	$-60$ to $-80$
Hydrazones		$-\text{NR}'_2$	$-210$ to $-240$
$=\text{N}-$	$5$ to $-60$	Azides $\text{RN}=\text{N}\equiv\text{N}$	
$-\text{NHR}$	$-190$ to $-250$	$\text{RN}=\text{}$	$-240$ to $-330$
Amides	$-210$ to $-300$	$=\text{N}\equiv$	$-120$ to $-150$
Azoles	$-230$ to $-260^b$	$\equiv\text{N}$	$-110$ to $-280$
Diazoles	$-160$ to $-230^b$	$\text{RCN}$	$-110$ to $-140$
Triazoles	$-120$ to $-220^b$	$\text{RNC}$	$-180$ to $-220$
Tetrazoles	$-100$ to $-150^b$	$\text{RCNO}$	$-160$ to $-180$
Azines	$-50$ to $-150$	$\text{ROCN}$	$\sim -220$
Diazines	$-25$ to $-160$	$\text{RSCN}$	$\sim -100$
Triazines	$70$ to $-160$	$\text{RN}=\text{C}=\text{O}$	$-325$ to $-365$
Nitro compounds		$\text{RN}=\text{C}=\text{S}$	$-265$ to $-290$
Aliphatic	$30$ to $-50$	Diazonium salts	
Aromatic	$-5$ to $-40$	$\text{R}-\text{N}^+$	$-130$ to $-190$
$\text{R}_2\text{N}-\text{NO}$		$\text{N}$	$-10$ to $-80$
$\text{R}_2\text{N}-$	$-100$ to $-170$	$\text{R}-\text{N}=\text{N}-\text{R}'$	$50$ to $200$
$-\text{NO}$	$130$ to $180$	$\text{R}-\text{NO}$	$430$ to $580$
$\text{R}_2\text{N}-\text{NO}_2$		$\text{R}_2\text{C}=\text{NOH}$	$0$ to $-55$
$\text{R}_2\text{N}-$	$-100$ to $-220$	$\text{M}-\text{NO}$ (linear)	$100$ to $-80$
$-\text{NO}_2$	$-20$ to $-40$	$\text{M}-\text{NO}$ (bent)	$850$ to $380$

<sup>a</sup> Referred to nitromethane ( $\delta=0.0$ ), positive values denote shifts to higher frequency.

<sup>b</sup> Pyrrole-type of nitrogen,  $\delta(^{15}\text{N})$  of pyridine-type of nitrogen:  $0$  to  $-140$ .

Both  $^{15}\text{N}$  chemical shifts and  $^nJ(^{15}\text{N}, \text{X})$  coupling constants are very sensitive to the presence of a lone pair on nitrogen, to its geometrical orientation, and to  $\pi$ -delocalization. Correlation analyses of  $\delta(^{15}\text{N})$  and  $^nJ(^{15}\text{N}, \text{X})$  and various physical, topological, and semiempirical parameters have been computed.

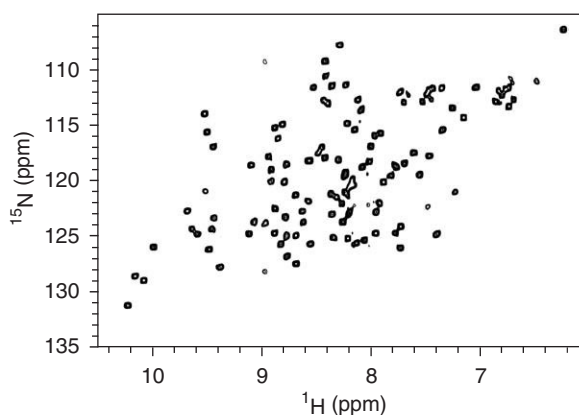
Two-dimensional  $^1\text{H}$ - $^{15}\text{N}$  shift-correlated spectra have been used in the assignment of polypeptides and similar compounds. The inverse detection technique (Figure 3) enables the measurements of very dilute solutions of compounds having high relative molecular mass. The experiment was optimized for the  $^1J(^{15}\text{N}, \text{H})$  coupling constant ( $\sim 90$  Hz) and magnetic field gradients were used in the pulse sequence. Application of magnetic field gradients is of key

importance in  $^1\text{H}$ - $^{15}\text{N}$  correlation experiments optimized for long-range  $^nJ(^{15}\text{N}, \text{H})$  coupling constants ( $n \geq 2$ ,  $^nJ(^{15}\text{N}, ^1\text{H}) \leq 16$  Hz). Figure 4 shows a 2D gradient selected  $^1\text{H}$ - $^{15}\text{N}$  HMBC spectrum of 2-benzyl-1-butyl-2,3-dihydro-imidazo[1,5-*c*]quinazoline-3,5-dione optimized for the long-range coupling constant  $^nJ(^{15}\text{N}, ^1\text{H}) = 7$  Hz. In addition to the values of  $^{15}\text{N}$  chemical shifts, correlations of an appropriate nitrogen with several protons can be

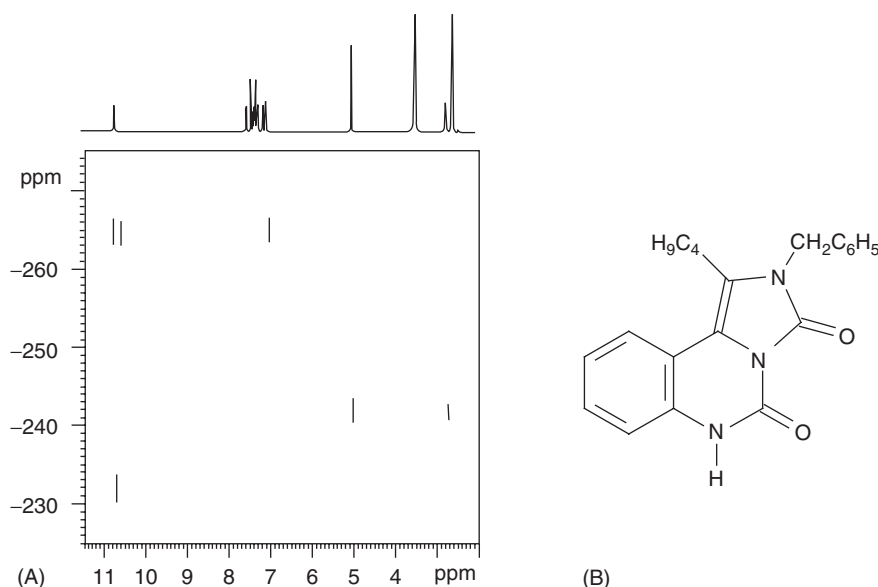
**Table 4** Absolute values of coupling constants  $^nJ(^{15}\text{N}, \text{X})$  (Hz)

X	$^nJ(^{15}\text{N}, \text{X})$ (Hz)			
	$n=1$	$n=2$	$n=3$	$n>3$
$^1\text{H}$	50–140	1–16	0–6	0–3 (8.2) <sup>a</sup>
$^{13}\text{C}$	0–35 (78) <sup>a</sup>	0–15 (37) <sup>a</sup>	0–6	<1 (3.9) <sup>a</sup>
$^{15}\text{N}$	5–25	1–8.5	<4	
$^{19}\text{F}$	150–460	5–105	0–12	0–3
$^{29}\text{Si}$	1–20			
$^{31}\text{P}$	0–95	0–10 (91) <sup>a</sup>	<5	
$^{119}\text{Sn}$	2–180			
$^{195}\text{Pt}$	100–600	<65	25–50	
$^{207}\text{Pb}$	19–240			

<sup>a</sup>Values in parentheses correspond to special cases.



**Figure 3** Two-dimensional gradient selected  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of unspecific  $^{15}\text{N}$  labeled triple mutant D26N/C7A/C106A of the 12 kDa form of Mason-Pfizer monkey virus protease (107 amino acids residues). (Spectrum taken from Veverka V, Bauerová H, Zábranský A, Pichová I, and Hrabal R (2001). *Journal of Biomolecular NMR* 20: 291–292, with permission.)



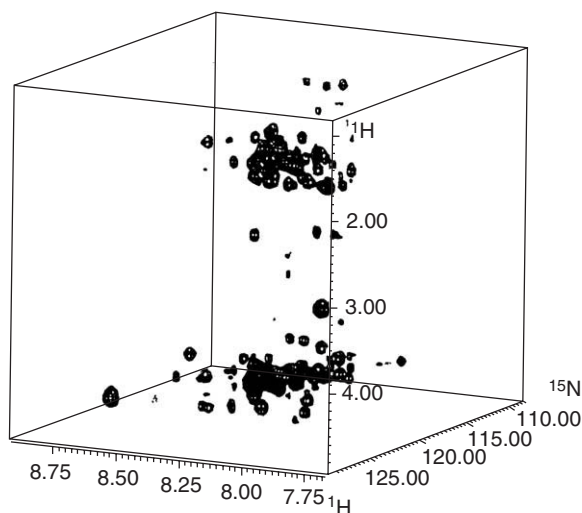
**Figure 4** (A) Two-dimensional gradient selected  $^1\text{H}$ - $^{15}\text{N}$  HMBC NMR spectrum of 2-benzyl-1-butyl-2,3-dihydro-imidazo[1,5-*c*]quinazoline-3,5-dione optimized for a long-range coupling constant  $^nJ(^{15}\text{N}, ^1\text{H}) = 7$  Hz. The NH proton correlates with both nitrogens of NH group (giving a clearly visible doublet due to  $^1J(^{15}\text{N}, ^1\text{H})$ ,  $\delta(^{15}\text{NH}) = -264.6$ , second cross-peak belongs to the correlation with aromatic proton from *peri* position) and nitrogen via  $^3J(^{15}\text{N}, ^1\text{H})$  from  $\text{HN-CO-N}$  ( $\delta(^{15}\text{N}) = -232.0$ ). Third nitrogen correlates with methylene protons of both butyl and benzyl groups ( $\delta(^{15}\text{N}) = -241.8$ ). (B) Structure of 2-benzyl-1-butyl-2,3-dihydro-imidazo[1,5-*c*]quinazoline-3,5-dione. (Spectrum taken from Klásek A, Kořístek K, Lyčka A, and Holčápek M (2003) *Tetrahedron* 59: 1283–1288, with permission.)



used very successfully in structure elucidation as well as in unequivocal assignment of  $^{15}\text{N}$  chemical shifts.

Three-dimensional spectra allow the spatial structure of complex molecules, e.g., proteins, to be studied (Figure 5).

Single crystals, crystalline, or amorphous powders and whole biological systems have been studied in the solid state by means of  $^{15}\text{N}$  NMR spectroscopy. The proton-decoupled  $^{15}\text{N}$  NMR powder pattern, measured in the solid state without rotation of a



**Figure 5** A part of the 3D NOESY  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR spectrum of uniformly labeled protein showing NOE connectivities of the NH to aliphatic protons;  $1\text{ mmol}^{-1}$  sample of the 79 amino acids fragment of transcriptional activator Jun in 90%  $\text{H}_2\text{O}/10\%$   $\text{D}_2\text{O}$  that formed a stable dimer (relative mass  $\approx 18000$ ) under given conditions. Total measuring time on a Bruker AMX 500 spectrometer equipped with a triple resonance probe and a gradient accessory was 60 h. (Reproduced with permission from Sklenář V, Piotto M, Lepik R, and Saudek V (1984); unpublished results.)

sample, arises from chemical shift anisotropy and covers a range of  $10^3$  to  $10^4$  Hz. Relatively very narrow signals can be obtained in  $^{15}\text{N}$  CP-MAS NMR spectra (Figure 6). This allows the study in the solid state of problems that are also studied in solutions. The isotropic peaks are flanked by spinning sidebands, the intensities of these side-band resonances correspond to the MAS rate. Great progress in improvement of instrumentation for solid state, very short pulses, effective decouplers, very high MAS rates, and pulse sequences development allow studying rather complex molecules including biological systems to be studied.

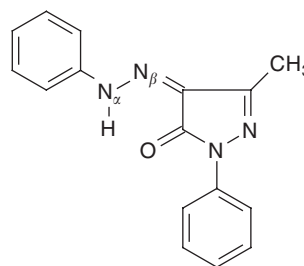
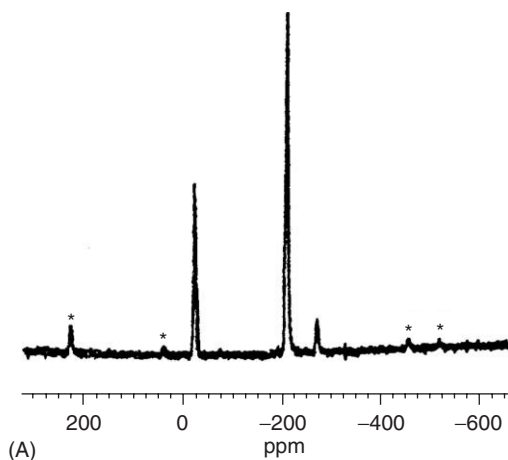
Dynamic solid-state  $^{15}\text{N}$  NMR studies have been performed to study proton transfer in, e.g., porphine derivatives. Line-shape analysis has provided rates of proton exchange that differ greatly from case to case and that are often very different when comparing solution and solid-state data.

Temperature changes of  $^{15}\text{N}$  chemical shifts in the solid state have been used to characterize tautomeric equilibria.

High-resolution 2D dipolar chemical shift experiments in the solid state permit the determination of bond distances  $^{15}\text{N}$ - $^1\text{H}$  and  $^{15}\text{N}$ - $^{13}\text{C}$ .

## Quantitative Analysis

The application of  $^{15}\text{N}$  NMR spectroscopy in quantitative analysis is limited to special cases because of the complicated relaxation mechanisms of the  $^{15}\text{N}$  nucleus and its negative magnetogyric ratio. It is extremely difficult to find experimental conditions providing intensities of  $^{15}\text{N}$  NMR signals exactly proportional to the relative abundance of compounds containing various types of nitrogens in their molecules.



**Figure 6** (A)  $^{15}\text{N}$  CP-MAS NMR (20.28 MHz, magic-angle spinning frequency 5 kHz) spectrum of  $^{15}\text{N}_\alpha^{15}\text{N}_\beta$  doubly enriched 3-methyl-1-phenyl-4,5-dione 4-phenylhydrazine.  $\delta(^{15}\text{N}_\alpha) = -222.9$ ,  $\delta(^{15}\text{N}_\beta) = -37.0$ . The asterisks denote spinning sidebands. (B) Structure of 3-methyl-1-phenyl-4,5-dione 4-phenylhydrazine.

Probably the best approach makes use of gated decoupling in order to suppress the NOE effect, long relaxation times, and an effective relaxation reagent with a nonspecific effect for all the nitrogens present. Even traces of paramagnetic impurities in sample, which act preferentially on certain molecular structures, can make quantitative analysis impossible.

On the other hand,  $^{15}\text{N}$  chemical shifts and  $^nJ(^{15}\text{N}, \text{X})$  (especially  $^nJ(^{15}\text{N}, ^1\text{H})$ ) can be used successfully for relative quantification, e.g., tautomeric equilibria, in cases of fast exchange on the NMR time scale when average values of the above-mentioned characteristics are observed experimentally.

**See also:** Nuclear Magnetic Resonance Spectroscopy: Instrumentation. Nuclear Magnetic Resonance Spectroscopy Techniques: Multidimensional Proton; Solid-State. Peptides.

## Further Reading

- Buchanan GW (1989) Applications of  $^{15}\text{N}$  NMR spectroscopy to the study of molecular structure, stereochemistry and binding phenomena. *Tetrahedron* 45: 581–604.
- Levy GC and Lichter RL (1979) *Nitrogen NMR*. New York: Wiley.
- Marek R and Lyčka A (2002)  $^{15}\text{N}$  NMR spectroscopy in structural analysis. *Current Organic Chemistry* 6: 35–66.

- Martin GE and Hadden CE (2000) Long-range  $^1\text{H}$ – $^{15}\text{N}$  heteronuclear shift correlation at natural abundance. *Journal of Natural Products* 63: 543–585.
- Martin GJ, Martin ML, and Gouesnard JP (1981)  $^{15}\text{N}$ -NMR spectroscopy. *NMR Basic Principles and Progress* 18: 1.
- Mason J (1981) Nitrogen nuclear magnetic resonance spectroscopy in inorganic, organometallic and bio-inorganic chemistry. *Chemical Reviews* 81: 205–227.
- Mason J (ed.) (1987) Nitrogen. In: *Multinuclear NMR*, ch. 12. New York: Plenum Press.
- von Philipsborn W and Müller R (1986)  $^{15}\text{N}$ -NMR spectroscopy – New methods and applications. *Angewandte Chemie (International Edition in English)* 25: 383–413.
- Witanowski M and Webb GA (eds.) (1973) *Nitrogen NMR*. London: Plenum Press.
- Witanowski M, Stefaniak L, and Webb GA (1977) Nitrogen NMR spectroscopy. *Annual Reports on NMR Spectroscopy* 7: 117–291.
- Witanowski M, Stefaniak L, and Webb GA (1981) Nitrogen NMR spectroscopy. *Annual Reports on NMR Spectroscopy* 11B: 1–502.
- Witanowski M, Stefaniak L, and Webb GA (1986) Nitrogen NMR spectroscopy. *Annual Reports on NMR Spectroscopy* 18: 3–761.
- Witanowski M, Stefaniak L, and Webb GA (1993) Nitrogen NMR spectroscopy. *Annual Reports on NMR Spectroscopy* 25: 1–4801.

## Phosphorus-31

R M Twyman, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Janice K Gard, William B Wise, and John C Burquin, pp. 3462–3470, © 1995, Elsevier Ltd.

## Introduction

Phosphorus occurs predominantly as the isotope  $^{31}\text{P}$ , which has a nuclear spin value of 1/2 and is therefore visible in nuclear magnetic resonance (NMR) spectrometry. Phosphorus NMR ( $^{31}\text{P}$  NMR) has been in use since the development of multinuclear, high-field Fourier-transform instruments in the late 1970s. The  $^{31}\text{P}$  NMR method is widely used, with both one-dimensional (1D) and multidimensional techniques, in such diverse areas as the characterization of organic and inorganic molecular structures, the analysis of biological fluids, the determination of intracellular pH, the noninvasive study of intact tissues and organs, and quantitative assays of industrial

products. In many routine applications,  $^{31}\text{P}$  NMR can now be carried out in a fully automated manner. Used in conjunction with other NMR-visible nuclei, the quantitation and full structure elucidation of a host of disparate species is possible. This article provides an overview of  $^{31}\text{P}$  NMR technology and some of its major applications.

## General Properties and Chemical Shifts

The chemical shift ( $\delta$ ) is a sensitive indicator of the chemical environment around the phosphorus atom. It relates closely to the molecular structure, and under some circumstances even facilitates stereochemical identification. The unique chemistry of phosphorus-containing materials is partially attributable to the variable oxidation state of  $^{31}\text{P}$ , the participation of d orbitals in bonding, and the ability of the phosphorus atom to vary its coordination number between 1 and 6.

Probably the best approach makes use of gated decoupling in order to suppress the NOE effect, long relaxation times, and an effective relaxation reagent with a nonspecific effect for all the nitrogens present. Even traces of paramagnetic impurities in sample, which act preferentially on certain molecular structures, can make quantitative analysis impossible.

On the other hand,  $^{15}\text{N}$  chemical shifts and  $^nJ(^{15}\text{N}, \text{X})$  (especially  $^nJ(^{15}\text{N}, ^1\text{H})$ ) can be used successfully for relative quantification, e.g., tautomeric equilibria, in cases of fast exchange on the NMR time scale when average values of the above-mentioned characteristics are observed experimentally.

**See also:** Nuclear Magnetic Resonance Spectroscopy: Instrumentation. Nuclear Magnetic Resonance Spectroscopy Techniques: Multidimensional Proton; Solid-State. Peptides.

## Further Reading

- Buchanan GW (1989) Applications of  $^{15}\text{N}$  NMR spectroscopy to the study of molecular structure, stereochemistry and binding phenomena. *Tetrahedron* 45: 581–604.
- Levy GC and Lichter RL (1979) *Nitrogen NMR*. New York: Wiley.
- Marek R and Lyčka A (2002)  $^{15}\text{N}$  NMR spectroscopy in structural analysis. *Current Organic Chemistry* 6: 35–66.

- Martin GE and Hadden CE (2000) Long-range  $^1\text{H}$ – $^{15}\text{N}$  heteronuclear shift correlation at natural abundance. *Journal of Natural Products* 63: 543–585.
- Martin GJ, Martin ML, and Gouesnard JP (1981)  $^{15}\text{N}$ -NMR spectroscopy. *NMR Basic Principles and Progress* 18: 1.
- Mason J (1981) Nitrogen nuclear magnetic resonance spectroscopy in inorganic, organometallic and bio-inorganic chemistry. *Chemical Reviews* 81: 205–227.
- Mason J (ed.) (1987) Nitrogen. In: *Multinuclear NMR*, ch. 12. New York: Plenum Press.
- von Philipsborn W and Müller R (1986)  $^{15}\text{N}$ -NMR spectroscopy – New methods and applications. *Angewandte Chemie (International Edition in English)* 25: 383–413.
- Witanowski M and Webb GA (eds.) (1973) *Nitrogen NMR*. London: Plenum Press.
- Witanowski M, Stefaniak L, and Webb GA (1977) Nitrogen NMR spectroscopy. *Annual Reports on NMR Spectroscopy* 7: 117–291.
- Witanowski M, Stefaniak L, and Webb GA (1981) Nitrogen NMR spectroscopy. *Annual Reports on NMR Spectroscopy* 11B: 1–502.
- Witanowski M, Stefaniak L, and Webb GA (1986) Nitrogen NMR spectroscopy. *Annual Reports on NMR Spectroscopy* 18: 3–761.
- Witanowski M, Stefaniak L, and Webb GA (1993) Nitrogen NMR spectroscopy. *Annual Reports on NMR Spectroscopy* 25: 1–4801.

## Phosphorus-31

R M Twyman, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Janice K Gard, William B Wise, and John C Burquin, pp. 3462–3470, © 1995, Elsevier Ltd.

## Introduction

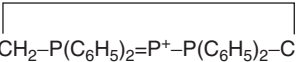
Phosphorus occurs predominantly as the isotope  $^{31}\text{P}$ , which has a nuclear spin value of 1/2 and is therefore visible in nuclear magnetic resonance (NMR) spectrometry. Phosphorus NMR ( $^{31}\text{P}$  NMR) has been in use since the development of multinuclear, high-field Fourier-transform instruments in the late 1970s. The  $^{31}\text{P}$  NMR method is widely used, with both one-dimensional (1D) and multidimensional techniques, in such diverse areas as the characterization of organic and inorganic molecular structures, the analysis of biological fluids, the determination of intracellular pH, the noninvasive study of intact tissues and organs, and quantitative assays of industrial

products. In many routine applications,  $^{31}\text{P}$  NMR can now be carried out in a fully automated manner. Used in conjunction with other NMR-visible nuclei, the quantitation and full structure elucidation of a host of disparate species is possible. This article provides an overview of  $^{31}\text{P}$  NMR technology and some of its major applications.

## General Properties and Chemical Shifts

The chemical shift ( $\delta$ ) is a sensitive indicator of the chemical environment around the phosphorus atom. It relates closely to the molecular structure, and under some circumstances even facilitates stereochemical identification. The unique chemistry of phosphorus-containing materials is partially attributable to the variable oxidation state of  $^{31}\text{P}$ , the participation of d orbitals in bonding, and the ability of the phosphorus atom to vary its coordination number between 1 and 6.

**Table 1** Chemical shift ranges of selected phosphorus compounds, referenced to 85% H<sub>3</sub>PO<sub>4</sub> with downfield shifts defined as positive. Biological compounds are degassed perchloric acid muscle extracts at 37.7°C, pH 7.2, K<sup>+</sup> counterion (from Gorenstein)

Compound/class	General structure	$\delta$ (ppm)	Comments
<i>Inorganic compounds</i>			
Orthophosphate		0.0	85% H <sub>3</sub> PO <sub>4</sub>
Protonated sodium tripolyphosphate		$\alpha = -11.5$	
[O <sub>3</sub> P <sub><math>\alpha</math></sub> OP <sub><math>\beta</math></sub> (O <sub>2</sub> )OP <sub><math>\gamma</math></sub> O <sub>3</sub> ] <sup>5-</sup> + 5 H <sup>+</sup>		$\beta = -23.9$	
Protonated tetrapolyphosphate		$\alpha = -11.5$	
[O <sub>3</sub> P <sub><math>\alpha</math></sub> OP <sub><math>\beta</math></sub> (O <sub>2</sub> )OP <sub><math>\gamma</math></sub> (O <sub>2</sub> )OP <sub><math>\delta</math></sub> O <sub>3</sub> ] <sup>6-</sup> + 6 H <sup>+</sup>		$\beta = -23.9$	
Phosphorus halides	PX <sub>3</sub>	97.0	X = F
		220	X = Cl
		178	X = I
H <sup>+</sup> + PF <sub>6</sub> <sup>-</sup>		-144	
[(CH <sub>3</sub> ) <sub>3</sub> Si] <sub>3</sub> C-P=As-C[Si(CH <sub>3</sub> ) <sub>3</sub> ] <sub>3</sub>		688	Coord 2
(CH <sub>3</sub> ) <sub>3</sub> CP[Cr(CO) <sub>5</sub> ] <sub>2</sub>		1362	
P(NCO) <sub>6</sub> <sup>-</sup>		-388.4	Coord 6
Gaseous P <sub>4</sub>		-552	
(CH <sub>3</sub> ) <sub>3</sub> SiC≡P		96	Coord 1
FC≡P		-207	Coord 1
<i>Organophosphorus compounds</i>			
Aliphatic phosphate triesters	(RO) <sub>3</sub> P=O	2.1	R = CH <sub>3</sub>
		1.0	R = C <sub>2</sub> H <sub>5</sub>
		-0.7	R = <i>n</i> -C <sub>3</sub> H <sub>7</sub>
		-3.3	R = <i>i</i> -C <sub>3</sub> H <sub>7</sub>
		-1.0	R = <i>n</i> -C <sub>4</sub> H <sub>9</sub>
		-13.3	R = <i>i</i> -C <sub>4</sub> H <sub>9</sub>
Acyclic phosphonates	(RO) <sub>2</sub> RP=O	323.4	R = OCH <sub>3</sub>
		30	R = OC <sub>2</sub> H <sub>5</sub>
		27.4	R = C <sub>3</sub> H <sub>7</sub>
Acyclic phosphites	P-(OR) <sub>3</sub>	139.7	R = CH <sub>3</sub>
		137.6	R = C <sub>2</sub> H <sub>5</sub>
		137.5	R = <i>i</i> -C <sub>3</sub> H <sub>7</sub>
		138.1	R = <i>t</i> -C <sub>4</sub> H <sub>9</sub>
		-232	Coord 2
	R <sub>3</sub> P	-240	R = H
		-62	R = CH <sub>3</sub>
		-20.1	R = C <sub>2</sub> H <sub>5</sub>
		20.1	R = <i>i</i> -C <sub>3</sub> H <sub>7</sub>
		61.9	R = <i>t</i> -C <sub>4</sub> H <sub>9</sub>
<i>Biological compounds</i>			
Triose phosphates		4.13	
Adenosine 5-monophosphate (AMP)		3.44	
Orthophosphate		2.20	
Glycerol 3-phosphorylcholine		-0.13	
Phosphocreatine		-2.89	
Adenosine triphosphate (ATP)		$\alpha = -10.87$	
		$\beta = -20.50$	
		$\gamma = -6.19$	
Adenosine diphosphate (ADP)		$\alpha = -10.61$	
		$\beta = -6.33$	

Phosphorus compounds give resonances in characteristic ranges depending primarily on the oxidation state and coordination number of the phosphorus atoms present (Table 1). Chemical shifts also depend on the pH, concentration, and salt content of the solution, solvent effects, and the electronegativity of any substituents. Since the phosphorus atom is a comparatively reactive center, it can usually

be found at or near an active site of interest. In general, phosphorus chemical shifts are referenced to an external standard. This requires minor corrections to be made because of differences in the bulk properties of the samples. The primary <sup>31</sup>P reference is 85% phosphoric acid placed in a sealed spherical container or cylindrical capillary tube. A number of secondary external and internal standards have been used

for specialized applications, including  $0.2 \text{ mol l}^{-1}$  phosphoric acid in 14% aqueous perchloric acid, aqueous disodium methylenediphosphonate, deuterated phosphoric acid ( $\text{D}_3\text{PO}_4$ ), trimethyl phosphine oxide, trimethylphosphonate,  $\text{P}_4\text{O}_6$ , phosphocreatine, and hexachlorocyclotriphosphazene in  $\text{C}_6\text{D}_6$ .

The chemical shift is sensitive to changes in the chemical environment, as is spin-spin coupling ( $J$ ). Examples of this sensitivity are listed in Table 1, which also gives examples of the wide variety of coordination numbers and oxidation states of phosphorus.  $^{31}\text{P}$  chemical shifts cover a wide range. Dissolved or gaseous diamagnetic phosphorus compounds span  $\sim 1900$  ppm. For dissolved paramagnetic phosphorus compounds, this range extends to  $\sim 2600$  ppm, with  $\text{OsCl}_4\text{P}[(n\text{-C}_4\text{H}_9)_2\text{C}_6\text{H}_5]_2$  observed at  $\sim 1219$  ppm. While it is beyond the scope of this article to provide a comprehensive list of  $^{31}\text{P}$  chemical shifts, further details can be found in the books and articles listed in the bibliography. The chemical shift is not the only parameter available for probing the molecular environment of a phosphorus atom. Pairs of phosphorus atoms can couple to each other, and individual phosphorus atoms can couple to  $^{13}\text{C}$  and  $^1\text{H}$  atoms.  $^{31}\text{P}$ - $^{13}\text{C}$  and  $^{31}\text{P}$ - $^1\text{H}$  couplings are sensitive to the stereochemical relationship between the coupling nuclei. Used together with carbon chemical shifts, the conformation and stereochemical features of many chemical systems can be deduced.  $^1\text{H}$ - $^{31}\text{P}$  couplings have been studied through one to four bonds and exhibit a dependence on the molecular conformation and geometry. Examples of the different types of couplings are shown in Table 2. The use of homonuclear and heteronuclear coupling constants and multinuclear (e.g.,  $^{31}\text{P}$ ,  $^1\text{H}$ , and  $^{13}\text{C}$ ) NMR data allows the structure of monophosphorus

and polyphosphorus compounds to be determined fully.

## Instrumentation and Sample Requirements for $^{31}\text{P}$ NMR

Many contemporary NMR spectrometers are capable of operating in multinuclear mode, and  $^{31}\text{P}$  NMR can be performed in most cases simply by tuning the NMR probe to the phosphorus frequency and carrying out the necessary calibrations. Common  $^{31}\text{P}$  resonance frequencies (MHz) and magnetic field strengths (T) are 80.96/4.7, 121.44/7.05, 161.92/9.4, and 202.4/11.75.

The signal-to-noise ratio is dependent on many variables, not all of which can be controlled by the operator. These include the volume of sample, the magnetic field strength ( $B_0$ ), the magnetic probe design, spectral accumulation time, relaxation times  $T_1$  and  $T_2$ , and sample concentration. With the wide range of available instruments,  $^{31}\text{P}$  NMR can be performed on samples ranging in size from a few microliters, through a few tenths of a milliliter to several milliliters, using respectively a microprobe, a 5 mm probe, and a 10 mm probe. When the operator is limited to a given sample and a single type of instrument, the most sensitive probe for  $^{31}\text{P}$  measurements should be chosen, and the sample should be concentrated as much as possible, since this helps to decrease accumulation time and improve sample throughput. Nevertheless, reasonable spectra can be obtained with sample concentrations in the range  $1\text{--}5 \text{ mmol l}^{-1}$ .

Faced with dilute samples, proper presentation of the sample and accurate spectrometer calibrations are particularly important. The sample should be mobile rather than bound, deoxygenated, and free of particulates. The removal of particulates by filtration makes a significant improvement to spectral quality. High-resolution solution work usually requires dissolution in a deuterated solvent for instrumental locking. However, excellent spectra can often be obtained in protonated solvents in the presence of a concentrically placed sealed capillary tube containing the deuterated solvent. This approach is unsuitable for intact biological samples where the NMR probe and coil assembly is designed around the specimen.

As with other nuclei, the highest  $^{31}\text{P}$  signal-to-noise ratio is obtained when the following operations are performed:

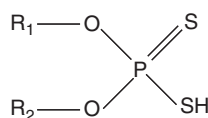
- Optimization of magnetic field homogeneity (shimming).
- Calibration of the observe coil  $90^\circ$  pulse width ( $\pi/2$ ).

**Table 2** Typical  $^{31}\text{P}$  coupling constants

Compound	Phosphorus atom coupling constant (Hz, absolute value)
$\text{PH}_3$	$186 J_{\text{PH}}$
$\text{P}(\text{CH}_3)_3$	$2.7 J_{\text{PH}}$
$\text{F-C}\equiv\text{P}$	$182 J_{\text{PF}}$
$\text{K}^+ \text{PF}_6^-$	$710 J_{\text{PF}}$
$\text{PF}_5$	$\sim 1000 J_{\text{PF}}$
$\text{OP}(\text{CH}_3)_3$	$68 J_{\text{PC}}$
$\text{P}(\text{CH}_2\text{CH}_3)_2$	$13.7 J_{\text{PCH}}$
	$0.5 J_{\text{PCCH}}$
$[\text{O}_3\text{P}_\alpha\text{OP}_\beta(\text{O}_2)\text{OP}_\alpha\text{O}_3]^{5-} + 5\text{H}^+$	$16.7 J_{\text{P}_\alpha\text{P}_\beta}$
$[\text{O}_3\text{P}_\alpha\text{OP}_\beta(\text{O}_2)\text{OP}_\beta(\text{O}_2)\text{OP}_\alpha\text{O}_3]^{6-} + 6\text{Na}^+$	$19.9 J_{\text{P}_\alpha\text{P}_\beta}$
	$16.7 J_{\text{P}_\alpha\text{P}_\beta}$
$\text{Mo}(\text{CO})_5\text{P}(\text{OCH}_3)_3$	$217 J_{\text{PMo}}$
$(\text{EtO})_2\text{P}(\text{O})\text{-SeP}^+(\text{NEt}_2)_3\text{-O}(\text{Se})\text{P}(\text{OEt})_2$	$17 J_{\text{PP}}$
	$420 J_{\text{PSe}}$



- Determination of the inherent relaxation times of the sample ( $T_1$ ) and the ensuing pulse repetition rate.
- Choice of sufficient numbers of data points to digitize the data adequately.
- Collection of sufficient time-averaged transients.



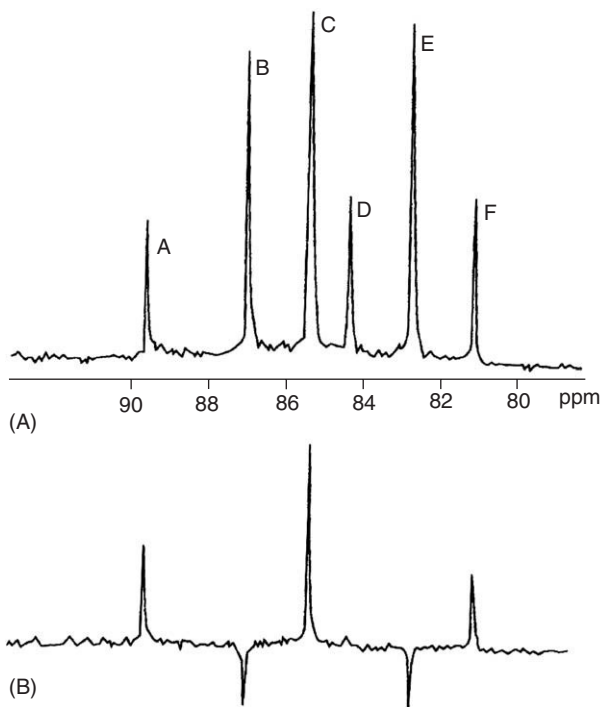
- A  $R_1 = R_2 = \text{Me}$   
 B  $R_1 = \text{Me}$   $R_2 = \text{Et}$   
 C  $R_1 = \text{Me}$   $R_2 = i\text{-Pr}$   
 D  $R_1 = R_2 = \text{Et}$   
 E  $R_1 = \text{Et}$   $R_2 = i\text{-Pr}$   
 F  $R_1 = R_2 = i\text{-Pr}$

## Applications of $^{31}\text{P}$ NMR Spectroscopy

### One-Dimensional Methods

Useful information can often be obtained by collecting a normal 1D survey (semiquantitative) spectrum employing gated proton decoupling and rapid repetition conditions (i.e., a flip angle  $< 90^\circ$  and a relatively short recycle time). Species can be identified by the observation of chemical shifts and relative concentrations from integration of the resonances.

In many cases, the 1D spectra are sufficiently complicated to prohibit a straightforward analysis. For example, **Figure 1A** shows the reaction products from the alcoholysis of  $\text{P}_4\text{S}_{10}$  with methanol, ethanol, and isopropanol. In this case, the selective use of techniques commonly employed in  $^{13}\text{C}$  NMR, such as the attached proton test (APT), distortionless enhancement through polarization transfer (DEPT), and insensitive nucleus enhancement through polarization transfer (INEPT), can yield simplified results. These methods facilitate the identification of appropriate functional groups surrounding the  $^{31}\text{P}$  atom. Coding of the heteronuclear coupling into an observable intensity modulation of the  $^{31}\text{P}$  signals is achieved using appropriately chosen delays in the pulse sequences and decoupler gating. The intensity of the observed  $^{31}\text{P}$  signal in these experiments depends on the number of protons, the heteronuclear coupling constant, and pulse delays (often represented by  $\tau$ ) as is shown in the APT spectrum in **Figure 1B**. Appropriate analysis of the intensity profiles as a function of the delays then results in a complete identification of the complex mixture. Further verification can be achieved by collecting the  $^1\text{H}$ -coupled  $^{31}\text{P}$  spectrum. Little use has been made of the INEPT method in  $^{31}\text{P}$  NMR experiments because optimization of the correct timing intervals is difficult due to the presence of numerous  $^1\text{H}$ - $^{31}\text{P}$  couplings. However, a greater than 10-fold enhancement can result from the use of polarization transfer procedures. This translates into a factor of 100 or more in time savings during data accumulation. However, where small (i.e., long-range) coupling constants are involved, longer delays are necessary in the pulse sequence. During these delays, the observable signal diminishes due to relaxation processes, resulting in lower signal-to-noise ratios. In



**Figure 1** (A) 32.44 MHz  $^{31}\text{P}$  spectrum of the alcoholysis products of  $\text{P}_4\text{S}_{10}$  with methanol, ethanol, and isopropanol. The sample is dissolved in toluene, with external  $\text{C}_6\text{D}_6$  as the field frequency lock. (B) The  $^{31}\text{P}$  APT spectrum of the same mixture, showing the encoding of functional group and multiplicity information into the intensity and phase of the resonances. (Reprinted from Jancke H, Radeglia R, Neels J, and Porzel A (1984) Application of the attached proton test technique in  $^{31}\text{P}$  NMR spectroscopy. *Organic Magnetic Resonance* 22: 376–378; © Wiley.)

such cases, multinuclear multidimensional techniques become more suitable.

### Two-Dimensional Methods

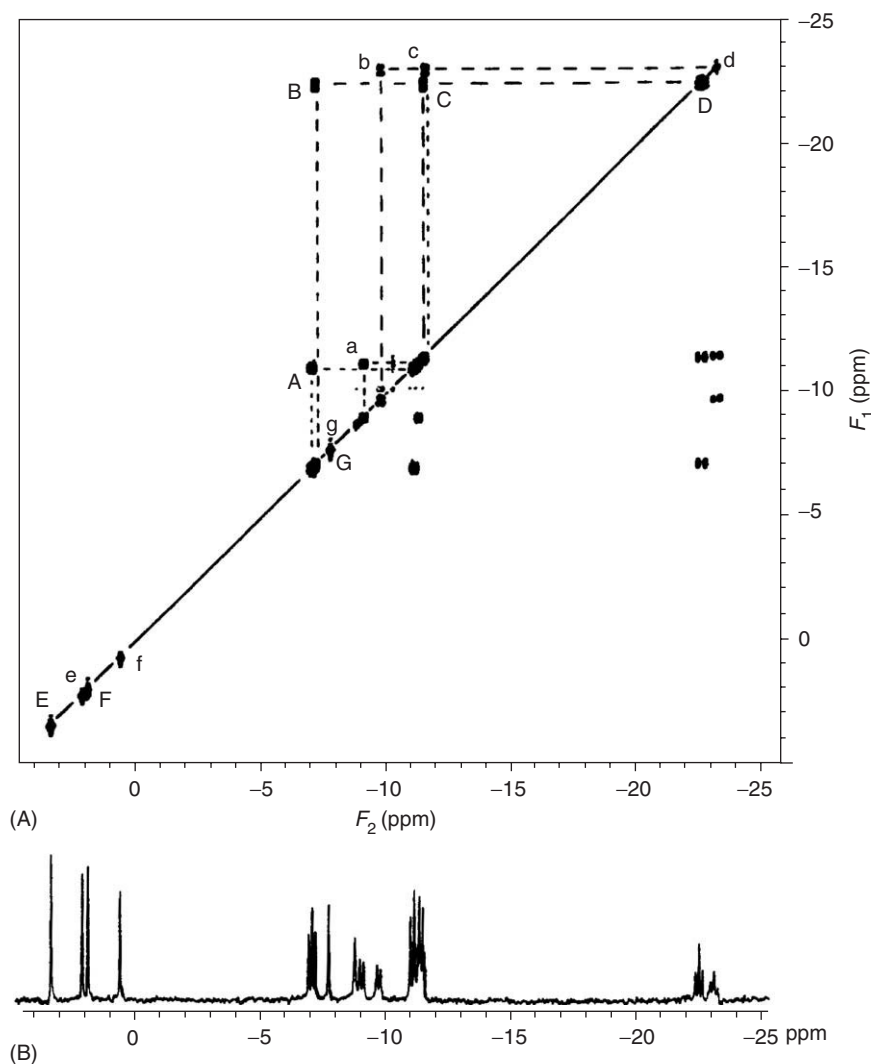
**Homonuclear NMR** Structural elucidation becomes much more difficult as spectral complexity increases. Under these circumstances, proton NMR spectroscopy has benefited considerably from the use of 2D NMR techniques. For example, homonuclear correlation spectroscopy (COSY) identifies spin-coupled pairs of nuclei as well as spin-coupled networks of nuclei in a molecule, even without prior structural information. The 2D  $J$ -resolved spectroscopy method permits even highly overlapping resonances to be resolved into readily interpretable multiplets. This enables chemical shift assignments to be made in a very straightforward manner. Both of

these techniques can easily be applied to  $^{31}\text{P}$  NMR, with enormous gains in the detail of spectral information.

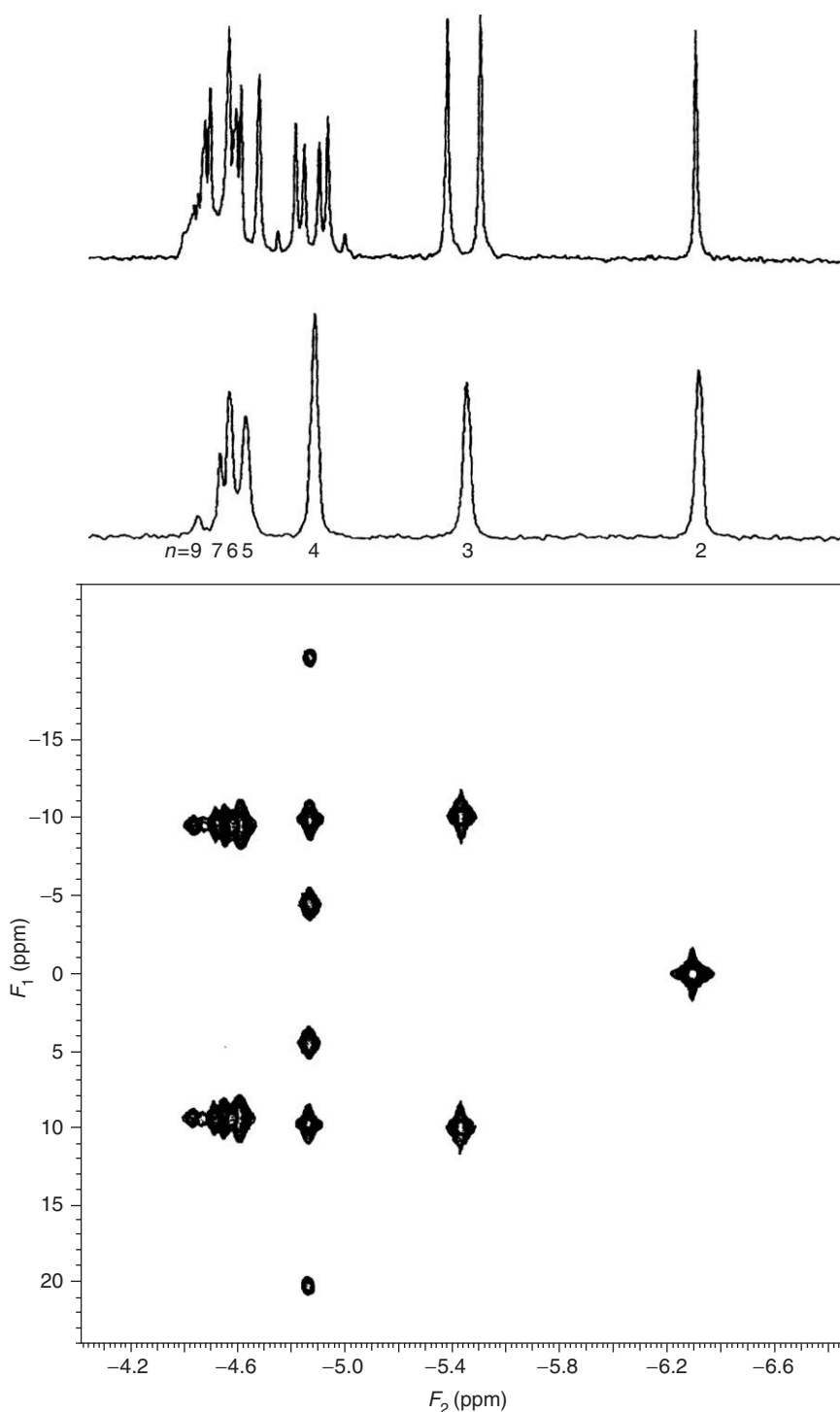
Figure 2A shows a  $^{31}\text{P}$ - $^{31}\text{P}$  COSY spectrum of a mixture of phosphate metabolites (ATP, AMP, or orthophosphate, and pyrophosphate, at two different pH values). The 1D spectrum shown in Figure 2B is also present along the diagonal of the contour plot. The off-diagonal elements, or cross-peaks, connect pairs of spin-coupled phosphorus resonances to each other, allowing the entire network to be rapidly assigned. Such a straightforward analysis is not

possible from the congested 1D spectrum. Along with connectivity information, the pH of the solution can be measured noninvasively from the chemical shift of the inorganic phosphate resonance.

Similarly, homonuclear  $^{31}\text{P}$ - $^{31}\text{P}$  2D  $J$ -resolved spectroscopy helps simplify even the most crowded spectra. The 1D spectrum of a dissolved phosphate glass of average chain length  $n=4.1$  is shown along the top of Figure 3. Even at 161.9 MHz, the peaks overlap significantly. Carrying out a 2D  $J$ -resolved experiment on this sample resolves the individual species clearly, with a monotonic increase in



**Figure 2** Two-dimensional homonuclear correlated spectrum (COSY) for a mixture of phosphate metabolites (A) in separate pH compartments. The 1D experiment (B) is shown below the  $^{31}\text{P}$ - $^{31}\text{P}$  COSY spectrum. The sample consisted of concentric NMR tubes with the inside pH at 7.5 (designated by capital letters) and the outside at pH 6.5 (lower-case letters). The diagonal peaks labelled D-d, E-e, F-f and G-g, represent the phosphate resonances of ATP ( $\beta$ -phosphate), AMP, orthophosphate, and pyrophosphate, respectively. The cross-peak A-a correlates the  $\alpha$ - and  $\beta$ -phosphorous resonances of ADP, B-b correlates the  $\gamma$ - and  $\beta$ -resonances of ATP, and C-c correlates the  $\alpha$ - and  $\beta$ -resonances of ATP. The resonances were broadened by the addition of  $\text{GdCl}_3$ . (Reprinted from Van Divender JM and Hutton WC (1982) The application of homonuclear two-dimensional chemical shift correlation maps to the  $^{31}\text{P}$  nuclear magnetic resonances of mixtures of nucleotides involved in cellular bioenergetic process. *Journal of Magnetic Resonances* 48: 272-279.)



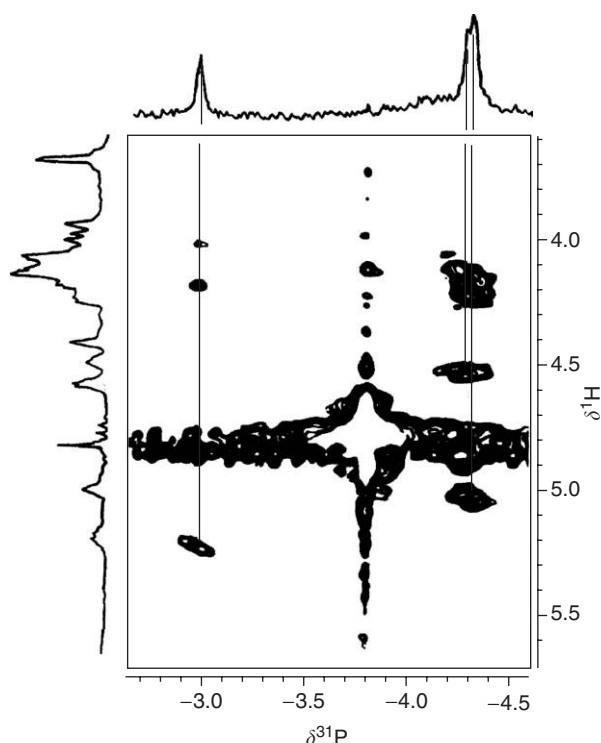
**Figure 3** A  $^{31}\text{P}$ – $^{31}\text{P}$  homonuclear 2D  $J$ -resolved NMR spectrum of a dissolved sodium phosphate glass with  $\bar{n} = 4.1$  (shown at the bottom). The characteristic  $J$ -coupling for each species is observed, causing extensive overlap in the corresponding 1D spectrum shown at the top. The vast simplification obtained by projecting the 2D  $J$ -resolved spectrum vertically onto the chemical shift axis (middle spectrum) clearly delineates the effectively broadband decoupled resonances. All spectra were accumulated at 161.9 MHz using a 5% (m/m) phosphate solution in  $\text{H}_2\text{O}$  with a  $\text{D}_2\text{O}$  insert for locking purposes.

chemical shift observed with increasing chain length,  $P_n$ . The  $F_1$  axis displays the coupling constant information, while the  $F_2$  axis shows the chemical shift of each species. Further simplification is achieved by

projecting the spectrum onto the chemical shift axis, which is shown above the contour plot. This effectively decouples all the resonances from each other, yielding individually resolved singlets.

**Heteronuclear NMR** As discussed above, complex spectral information can be simplified greatly by resolving  $^{31}\text{P}$ – $^{31}\text{P}$  homonuclear interactions into two dimensions while simultaneously increasing the information content of the spectra. Using additional 2D modes available on the majority of current spectrometers, further information can be obtained by correlating individual phosphorus resonance alignments with their respective protons. In these heteronuclear shift-correlated experiments, the nucleus that is detected can be either  $^1\text{H}$  or  $^{31}\text{P}$ .

Much information can be gained by observing the phosphorus nucleus in these experiments. One approach known as  $^1\text{H}$ -detected heteronuclear multiple quantum coherence (HMQC) is shown in Figure 4, applied to the detection of  $^{31}\text{P}$  resonances in d(TGGT)Pt(en), an oligodeoxyribonucleotide adduct of an anticancer drug. In this example, each  $^{31}\text{P}$  resonance was uniquely correlated with the corresponding  $\text{H}_3'$  resonance in the preceding nucleotide residue; the  $\text{H}_5'$ ,  $\text{H}_5''$  resonances were also resolved



**Figure 4** HMQC contour plot and corresponding  $^{31}\text{P}$  (top) and  $^1\text{H}$  (side) traces for a  $5\text{ mmol l}^{-1}$   $\text{D}_2\text{O}$  solution of d(TGGT)Pt(en) at 400 MHz ( $^1\text{H}$ ), 20°C and pH 7.1. (Reprinted from Byrd RA, Summers MF, Zon G, Fouts CS, and Marzilli LG (1986) A new approach for assigning  $^{31}\text{P}$  NMR signals and correlating adjacent nucleotide deoxyribose moieties via  $^1\text{H}$ -detected multiple-quantum NMR. Application to the adduct of d(TGGT) with the anticancer agent (ethylenediamine)dichloroplatinum. *Journal of the American Chemical Society* 108: 504–505. © 1986, American Chemical Society.)

for the  $^{31}\text{P}$  resonances at  $-2.96$  ppm. Subsequent comparison with 2D nuclear Overhauser effect (NOE) studies and variable temperature work provided unambiguous assignments of all the  $^{31}\text{P}$  resonances.

Heteronuclear shift correlation is not restricted to  $^1\text{H}$  and  $^{31}\text{P}$  alone.  $^{31}\text{P}$ -detected experiments permit ready observation of rare nuclei, such as  $^{183}\text{W}$  and  $^{57}\text{Fe}$ , which are normally difficult to observe. This approach has been used with success for the observation of organometallic phosphorus tungsten(IV), rhodium(I), iron(0), iron(II), and nickel(0) complexes. Indirect observation of insensitive spin 1/2 nuclei offers tremendous gains in sensitivity over direct observation.

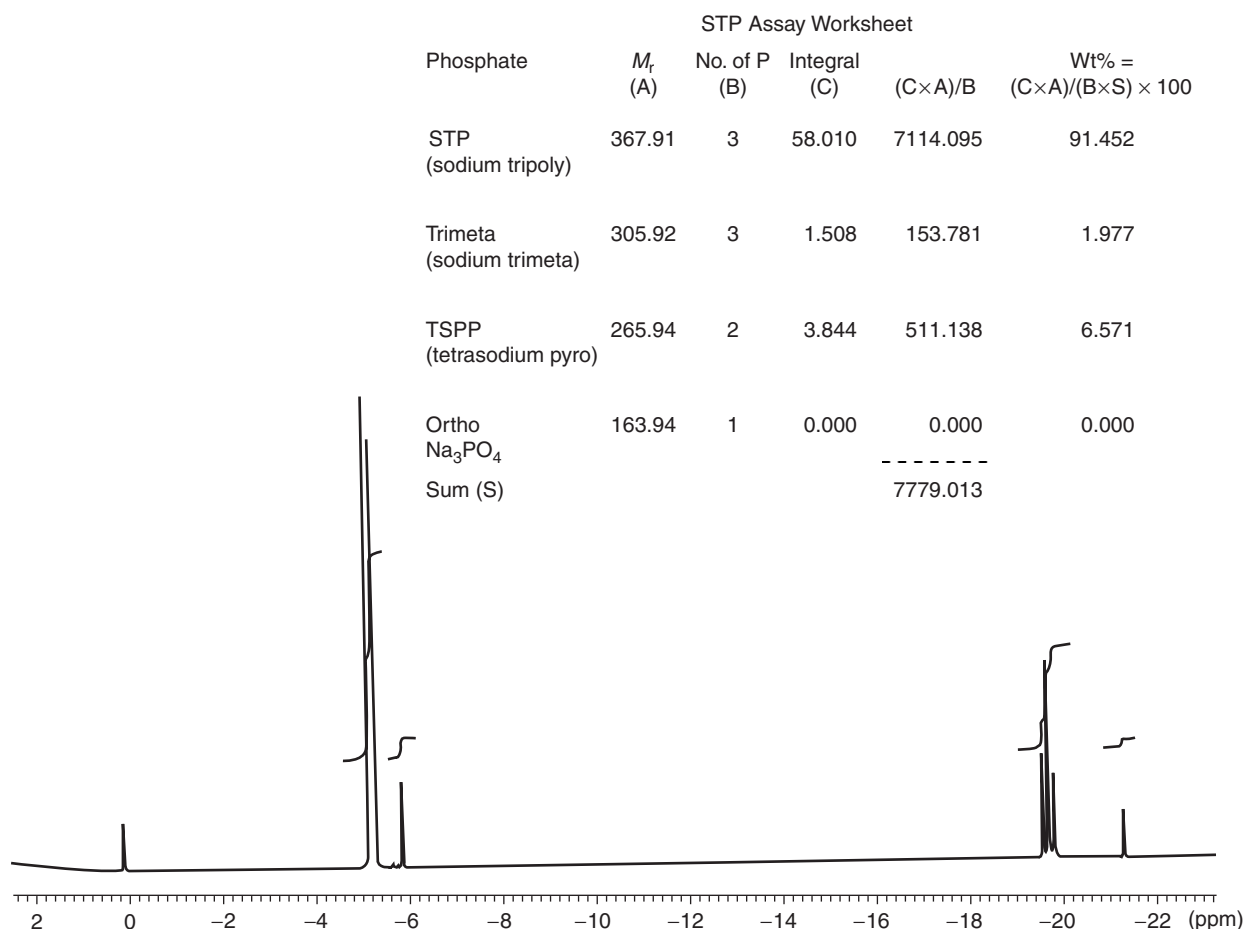
## Analytical Aspects

Using many of the techniques often applied to the study of  $^1\text{H}$  and  $^{13}\text{C}$ ,  $^{31}\text{P}$  NMR is a powerful tool for structure elucidation. Observation of the  $^{31}\text{P}$  nucleus opens yet another realm of study – quantitation of the individual species that were previously identified using the methods described above. Quantitation requires the optimization of observation conditions such as pulse width, recycle delay and decoupler gating, as well as determination of the inherent relaxation time of the nucleus,  $T_1$ . Relative quantitation is then easily achieved, with the area under any peak directly proportional to the concentration of the species represented by that resonance. Comparison of the resulting integrals with that of integral standards yields absolute quantities. The precision of phosphate assays by  $^{31}\text{P}$  NMR is consistently within 0.2–0.6%, comparable with results obtained using chromatographic methods.

Indeed,  $^{31}\text{P}$  NMR is widely used in a quantitative fashion. Computer-controlled and automated instruments, used in combination with robotic sample changers, make the large number of analytical samples much more manageable. With customized software, the spectrometer can automatically accumulate and store data, perform the Fourier transform and integration, calculate the relative distribution of species present and generate the analytical report. As an example, Figure 5 shows the output from a routine oligophosphate assay.

## Biological and *In Vivo* Applications

$^{31}\text{P}$  NMR has also been widely used in biological investigations. High-resolution NMR has been used increasingly for the analysis of biological samples and tissues, allowing the identification and quantitation of all  $^{31}\text{P}$ -containing species without further



**Figure 5** 121 MHz  $^{31}\text{P}$  spectrum of a 5% (m/m)  $\text{H}_2\text{O}$  solution of oligophosphates, with a  $\text{H}_3\text{PO}_4\text{-D}_2\text{O}$  capillary inserted for locking purposes. The spectrum and analytical report were automatically generated after making only a few menu choices and placing the sample into the robot sample tray.

purification. NMR has been used to study  $^{31}\text{P}$ -containing molecules such as oligonucleotides, nucleotides, proteins, and phosphosugars. Positive resonance assignments are obtained even in cases of high spectral complexity, such as the phosphate metabolite mixture shown in Figure 2.

The invention of the surface coil and the development of NMR probes containing life-support systems have made it possible to observe high-energy phosphates in intact tissues and whole organisms.  $^{31}\text{P}$  NMR has been used for the safe, noninvasive detection of disease states in tissues without risk to the patient. Metabolic products soluble in the cytoplasm of skeletal muscle, heart, kidney, brain, and eye have all been qualitatively observed and identified. With the appropriate use of standards, determination of relaxation times, and performance of the necessary instrument calibrations, quantitative results can be comparable to those obtained in wet-chemical assays. The protonation/deprotonation equilibrium of phosphates influences the  $^{31}\text{P}$  chemical shift, facilitating the noninvasive determination of tissue pH.

See also: **Nuclear Magnetic Resonance Spectroscopy:** Overview; Principles; Instrumentation. **Nuclear Magnetic Resonance Spectroscopy Applications:** Proton NMR in Biological Objects Subjected to Magic Angle Spinning.

## Further Reading

- Chemical Concepts (2001)  *$^{31}\text{P}$ -NMR Spectral Collection*. New York: Wiley.
- Crutchfield MM, Dungan CH, Letcher JH, Mark V, and Van Wazer JR (1967)  *$^{31}\text{P}$  Nuclear Magnetic Resonance, Topics in Phosphorus Chemistry*, vol. 5. New York: Interscience Publishers.
- Gard JK, Gard DR, and Callis CF (1992) Quantitative analysis of inorganic phosphates using  $^{31}\text{P}$  NMR spectroscopy. In: Walsh EN, Griffith EJ, Parry RW, and Quin LD (eds.) *Phosphorus Chemistry, Developments in American Science*. Washington, DC: American Chemical Society.
- Gard JK, Gard DR, and Callis CF (1992) Quantitative analysis of inorganic phosphates using P-31 NMR spectroscopy. *ACS Symposium Series* 486: 41–55.



- Gorenstein DG (1984) *Phosphorus-31 NMR Principles and Applications*. New York: Academic Press.
- Gorenstein DG (1989) P-31 nuclear magnetic resonance of enzyme complexes – bound ligand structure, dynamics, and environment. *Methods in Enzymology* 177: 295–316.
- Gorenstein DG (1992) P-31 NMR of DNA. *Methods in Enzymology* 211: 254–286.
- Lopez-Ortiz F and Carbajo RJ (1998) Applications of polarization transfer and indirect detection NMR spectroscopic methods based on phosphorus-31 in organic and organometallic chemistry. *Current Organic Chemistry* 2: 97–130.
- Mitchell T and Costisella B (2004) *NMR – From Spectra to Structures: An Experimental Approach*. New York: Springer.
- Tebby JC (ed.) (1991) *Handbook of Phosphorus-31 Nuclear Magnetic Resonance Data*. New York: CRC Press.
- Verkade JG and Quin LD (eds.) (1987) *Phosphorus-31 NMR Spectroscopy in Stereochemical Analysis*. Deerfield Beach: VCH Publishers.

## Organometallic Compounds

**C Pettinari**, Università degli Studi, Camerino, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Nuclear magnetic resonance (NMR) spectroscopy is one of the most powerful instruments in determining the structure and properties of an organometallic compound because of the sensitivity of the chemical shift to difference in geometry. Also, the term ‘organometallic’ seems suitable only for compounds having at least one carbon atom directly linked to a metal, and the scientific community uses this term also for compounds in which organic groups are bonded to nonmetallic elements such as boron, silicon, and phosphorus. In the first edition of the *Encyclopedia of Analytical Science*, the discussion of the application of NMR spectroscopy to organometallic chemistry has been mainly directed to the  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra, here the discussion will be addressed mainly on multinuclear NMR spectroscopy.

NMR spectra of nuclei having a spin quantum number  $I = 1/2$  are generally more easy to investigate due to a coupling behavior similar to that of  $^1\text{H}$ . However, some factors, that could make it difficult to record and assign an NMR spectrum, need to be considered: (1) isotopic abundance, (2) sensitivity, (3) nuclear quadrupole, and (4) relaxation time. For example, small molecules having nuclei  $1/2$  exhibit generally sharp lines ( $W_{1/2} = 0\text{--}10\text{ Hz}$ ), but if the nuclei interacts weakly with environments, long relaxation times, that make the detection of signals very difficult, can occur (the  $T_1$  for  $^{109}\text{Ag}$  is up to 1000 s). Nuclei having spin with  $I$  in the range  $1\text{--}9/2$  are also suitable for an NMR investigation; however, these nuclei possess a quadrupole moment (deviation from spherical charge distribution) which produces extremely short relaxation time and extremely large linewidth  $W_{1/2}$  (up to 50 Hz). If the compounds have

low molecular weight (small  $\tau_c$ ) or if the nuclei are embedded in ligand field of cubic (tetrahedral, octahedral) symmetry ( $q_{zz}$  blocked), narrow lines have been generally observed (eqn [1]).

$$W_{1/2} = \frac{(2I + 3)Q^2 q_{zz}^2 \tau_c}{I^2(2I - 1)} \quad [1]$$

where  $Q$  is the quadrupole moment;  $q_{zz}$  the electric field gradient;  $\tau_c$  the correlation time; and  $I$  the spin quantum number.

It should also be taken into account that while the  $^1\text{H}$  NMR chemical shift is dominated by the diamagnetic term, heavier nuclei chemical shifts are generally dominated by the paramagnetic term. In **Tables 1** and **2** a summary of the NMR properties of most relevant spins  $1/2$  nuclei and spins quadrupolar nuclei, respectively, are reported.

Much of the chemical information in NMR spectra arises from chemical shifts. A decrease in atomic electron density increases the deshielding, i.e., an increase in shielding results in a decrease in the chemical shift,  $\delta$ , i.e., in a shift upfield, i.e., to less positive/more negative  $\delta$  values. Several factors, such as coordination number, oxidation state, geometry, and electronegativity of substituents, influence the shielding of the nucleus and hence the chemical shift in organometallic complexes (**Table 3**). It is very difficult to find general rules for the chemical shifts of nuclei in organometallic compounds since many interdependent factors are often involved. Shifts can be very large, up to 20 000 ppm, in metal complexes, often due to strong interaction of metal electrons with the ligands. The geometric factors can make otherwise equivalent nuclei distinct: in  $\text{Fe}_3(\text{CO})_{12}$  (**Figure 1**) the two bridging carbons will be distinct from the 10 terminal CO ligands and the terminal carbonyls on  $\text{Fe}^1$  will be different from those on  $\text{Fe}^2$ . So the  $^{13}\text{C}$  NMR spectrum will exhibit different signals for the distinct terminal CO

- Gorenstein DG (1984) *Phosphorus-31 NMR Principles and Applications*. New York: Academic Press.
- Gorenstein DG (1989) P-31 nuclear magnetic resonance of enzyme complexes – bound ligand structure, dynamics, and environment. *Methods in Enzymology* 177: 295–316.
- Gorenstein DG (1992) P-31 NMR of DNA. *Methods in Enzymology* 211: 254–286.
- Lopez-Ortiz F and Carbajo RJ (1998) Applications of polarization transfer and indirect detection NMR spectroscopic methods based on phosphorus-31 in organic and organometallic chemistry. *Current Organic Chemistry* 2: 97–130.
- Mitchell T and Costisella B (2004) *NMR – From Spectra to Structures: An Experimental Approach*. New York: Springer.
- Tebby JC (ed.) (1991) *Handbook of Phosphorus-31 Nuclear Magnetic Resonance Data*. New York: CRC Press.
- Verkade JG and Quin LD (eds.) (1987) *Phosphorus-31 NMR Spectroscopy in Stereochemical Analysis*. Deerfield Beach: VCH Publishers.

## Organometallic Compounds

**C Pettinari**, Università degli Studi, Camerino, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Nuclear magnetic resonance (NMR) spectroscopy is one of the most powerful instruments in determining the structure and properties of an organometallic compound because of the sensitivity of the chemical shift to difference in geometry. Also, the term ‘organometallic’ seems suitable only for compounds having at least one carbon atom directly linked to a metal, and the scientific community uses this term also for compounds in which organic groups are bonded to nonmetallic elements such as boron, silicon, and phosphorus. In the first edition of the *Encyclopedia of Analytical Science*, the discussion of the application of NMR spectroscopy to organometallic chemistry has been mainly directed to the  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra, here the discussion will be addressed mainly on multinuclear NMR spectroscopy.

NMR spectra of nuclei having a spin quantum number  $I = 1/2$  are generally more easy to investigate due to a coupling behavior similar to that of  $^1\text{H}$ . However, some factors, that could make it difficult to record and assign an NMR spectrum, need to be considered: (1) isotopic abundance, (2) sensitivity, (3) nuclear quadrupole, and (4) relaxation time. For example, small molecules having nuclei  $1/2$  exhibit generally sharp lines ( $W_{1/2} = 0\text{--}10\text{ Hz}$ ), but if the nuclei interacts weakly with environments, long relaxation times, that make the detection of signals very difficult, can occur (the  $T_1$  for  $^{109}\text{Ag}$  is up to 1000 s). Nuclei having spin with  $I$  in the range  $1\text{--}9/2$  are also suitable for an NMR investigation; however, these nuclei possess a quadrupole moment (deviation from spherical charge distribution) which produces extremely short relaxation time and extremely large linewidth  $W_{1/2}$  (up to 50 Hz). If the compounds have

low molecular weight (small  $\tau_c$ ) or if the nuclei are embedded in ligand field of cubic (tetrahedral, octahedral) symmetry ( $q_{zz}$  blocked), narrow lines have been generally observed (eqn [1]).

$$W_{1/2} = \frac{(2I + 3)Q^2 q_{zz}^2 \tau_c}{I^2(2I - 1)} \quad [1]$$

where  $Q$  is the quadrupole moment;  $q_{zz}$  the electric field gradient;  $\tau_c$  the correlation time; and  $I$  the spin quantum number.

It should also be taken into account that while the  $^1\text{H}$  NMR chemical shift is dominated by the diamagnetic term, heavier nuclei chemical shifts are generally dominated by the paramagnetic term. In **Tables 1** and **2** a summary of the NMR properties of most relevant spins  $1/2$  nuclei and spins quadrupolar nuclei, respectively, are reported.

Much of the chemical information in NMR spectra arises from chemical shifts. A decrease in atomic electron density increases the deshielding, i.e., an increase in shielding results in a decrease in the chemical shift,  $\delta$ , i.e., in a shift upfield, i.e., to less positive/more negative  $\delta$  values. Several factors, such as coordination number, oxidation state, geometry, and electronegativity of substituents, influence the shielding of the nucleus and hence the chemical shift in organometallic complexes (**Table 3**). It is very difficult to find general rules for the chemical shifts of nuclei in organometallic compounds since many interdependent factors are often involved. Shifts can be very large, up to 20 000 ppm, in metal complexes, often due to strong interaction of metal electrons with the ligands. The geometric factors can make otherwise equivalent nuclei distinct: in  $\text{Fe}_3(\text{CO})_{12}$  (**Figure 1**) the two bridging carbons will be distinct from the 10 terminal CO ligands and the terminal carbonyls on  $\text{Fe}^1$  will be different from those on  $\text{Fe}^2$ . So the  $^{13}\text{C}$  NMR spectrum will exhibit different signals for the distinct terminal CO

**Table 1** Properties of some spin 1/2 nuclei

Isotope	Natural abundance (%)	$(\gamma)^a 10^7 \text{ rad T}^{-1} \text{ s}^{-1}$	Frequency (MHz)	Relative receptivity
$^1\text{H}$	99.985	26.7519	100.0	1.0
$^{13}\text{C}$	1.11	6.7283	25.1	$1.8 \times 10^{-4}$
$^{15}\text{N}$	0.37	-2.712	10.1	$3.9 \times 10^{-6}$
$^{19}\text{F}$	100.0	25.181	94.1	$8.3 \times 10^{-1}$
$^{29}\text{Si}$	4.7	-5.3188	19.9	$3.7 \times 10^{-4}$
$^{31}\text{P}$	100.0	10.841	40.5	$6.6 \times 10^{-2}$
$^{57}\text{Fe}$	2.2	0.8661	3.2	$7.4 \times 10^{-7}$
$^{77}\text{Se}$	7.6	5.12	19.1	$5.3 \times 10^{-4}$
$^{103}\text{Rh}$	100.0	-0.846	3.2	$3.2 \times 10^{-5}$
$^{109}\text{Ag}$	48.2	-1.250	4.7	$4.9 \times 10^{-5}$
$^{113}\text{Cd}$	12.3	-5.9550	22.2	$1.3 \times 10^{-3}$
$^{119}\text{Sn}$	8.6	-10.021	37.3	$4.5 \times 10^{-3}$
$^{125}\text{Te}$	7.0	-8.498	31.5	$2.2 \times 10^{-3}$
$^{183}\text{W}$	14.4	1.120	4.2	$1.1 \times 10^{-5}$
$^{187}\text{Os}$	1.6	0.616	2.3	$2.0 \times 10^{-7}$
$^{195}\text{Pt}$	33.8	5.768	21.4	$3.4 \times 10^{-3}$
$^{199}\text{Hg}$	16.8	4.8154	17.9	$9.8 \times 10^{-4}$
$^{205}\text{Tl}$	70.5	15.589	57.6	$1.4 \times 10^{-1}$
$^{207}\text{Pb}$	22.6	5.540	20.9	$2.0 \times 10^{-3}$

<sup>a</sup> Fuller GH (1976) *Journal of Physical and Chemical Reference Data*-1976-5: 835-1092: Nuclear spins and moments.

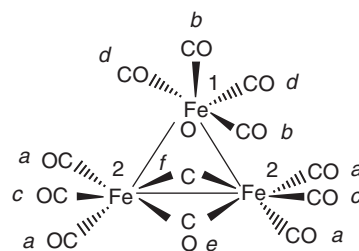
**Table 2** NMR properties of some quadrupolar nuclei

Isotope	Spin	Natural abundance (%)	$(\gamma)^a 10^7 \text{ rad T}^{-1} \text{ s}^{-1}$	NMR frequency (MHz)	Relative receptivity	Quadrupole moment $10^{-28} \text{ m}^2$
$^6\text{Li}$	1	7.4	3.9371	14.7	$6.3 \times 10^{-4}$	$-8 \times 10^{-4}$
$^7\text{Li}$	3/2	92.6	10.3975	38.9	$2.7 \times 10^{-1}$	$-4 \times 10^{-2}$
$^9\text{Be}$	3/2	100.0	-3.7596	14.1	$1.4 \times 10^{-2}$	$5 \times 10^{-2}$
$^{10}\text{B}$	3	19.6	2.8746	10.7	$3.9 \times 10^{-3}$	$8.5 \times 10^{-2}$
$^{11}\text{B}$	3/2	80.4	8.5843	32.1	$1.3 \times 10^{-1}$	$4.1 \times 10^{-2}$
$^{25}\text{Mg}$	5/2	10.1	-1.639	6.1	$2.7 \times 10^{-4}$	$2.2 \times 10^{-1}$
$^{27}\text{Al}$	5/2	100.0	6.9760	26.1	$2.1 \times 10^{-1}$	$1.5 \times 10^{-1}$
$^{55}\text{Mn}$	5/2	100.0	6.608	24.7	$1.8 \times 10^{-1}$	$4 \times 10^{-1}$
$^{73}\text{Ge}$	9/2	7.8	-0.9357	3.5	$1.1 \times 10^{-4}$	$-1.8 \times 10^{-1}$

<sup>a</sup> Fuller GH (1976) *Journal of Physical and Chemical Reference Data*-1976-5: 835-1092: Nuclear spins and moments.

**Table 3**  $^1\text{H}$  NMR chemical shift for selected hydride compounds

Compound	$\delta$ (ppm)		
	M=C	M=Si	M=Ge
$\text{MH}_4$	0.1	3.2	3.1
$\text{MH}_3\text{I}$	2.0	3.4	3.5
$\text{MH}_3\text{Br}$	2.5	4.2	4.5
$\text{MH}_3\text{Cl}$	2.8	4.6	5.1
$(\text{MH}_3)_2\text{O}$	3.2	4.6	5.3
$\text{MH}_3\text{F}$	4.1	4.8	5.7

**Figure 1** Different CO groups in  $\text{Fe}_3(\text{CO})_{12}$ .

environments and the bridging CO groups in  $\text{Fe}_3(\text{CO})_{12}$ . The effects of electronegativity, charge, and oxidation state can be seen in the  $^{31}\text{P}$  chemical shift of some typical phosphino-compounds. In  $\text{PPh}_3$  the  $\delta$  is 5.6 ppm, in  $\text{P}(\text{OPh})_3$  (more electronegative substituents) is 128 ppm,  $\text{PPh}_4^+$  exhibits a chemical

shift of 20.8 ppm (positive charge, whereas for  $\text{O}=\text{PPh}_3$  (+5 oxidation state)  $\delta$  is  $\sim 29$  ppm). The nature of the metal bonded to an organic group can be also an important factor in the shielding; in the series  $\text{M}(\text{CO})_6$  the shift of the carbonyl carbon decreases from 212.3 to 204.1 to 191.9 ppm on going

from  $M = \text{Cr}$  to  $\text{Mo}$  to  $\text{W}$ . These variations are due to a decreasing paramagnetic contribution to the shielding from low-lying excited states as the atomic number of  $M$  increases.

A number of factors also influences the magnitude of the coupling constants: formal hybridization of the nuclei, coordination number,  $s$ -character in the bond, electronegativity of the substituents, presence of lone pairs on the elements, oxidation state, interbond angles, *trans*-influence, etc. The magnitude of the reduced coupling constant generally increases down a group, for example,  $K_{\text{MC}}$  increases in the series  $\text{SiMe}_4$ ,  $\text{GeMe}_4$ ,  $\text{SnMe}_4$ ,  $\text{PbMe}_4$  (8.3, 18.0, 30.0, and  $39.3 \times 10^{20} \text{ NA}^{-2} \text{ m}^3$ , respectively).  $J$  ( $^{195}\text{Pt}$ ,  $^{13}\text{C}$ ) coupling constants generally increase from 450–700 Hz in platinum alkyl compounds ( $\text{sp}^3$ ) to 750–850 Hz in carbene complexes ( $\text{sp}^2$ ) to 990–2250 Hz in terminal platinum carbonyls. The coupling between platinum and phosphorus falls in the range 1050–2075 Hz for  $\text{Pt}^{\text{IV}}(\text{PR}_3)$ , 1675–5850 Hz for  $\text{Pt}^{\text{II}}(\text{PR}_3)$ , and finally 2820–9150 Hz for  $\text{Pt}^0(\text{PR}_3)$  species.

It is worth mentioning that one-bond coupling depends on  $s$ -orbital character of the bond and on the electronegativity of the nuclei, and it can be used to determine coordination number and to distinguish between axial and equatorial orientation of P-donor and fluoro groups around a metal. Two-bond couplings can give structural information, and so a relationship between  $^2J$  and bond angles is being determined for a number of compounds. The range of two-bond couplings passes through zero value, therefore the sign of the couplings must be determined in order to evaluate the magnitude of the interaction between nuclei. Finally, the three-bond couplings depend on the dihedral angles in accordance with the following equation:

$$^3J(\text{E}, \text{X}) = A \cos 2\phi + B \cos \phi + C \quad [2]$$

( $A, B, C$  = empirical constants)

NMR spectroscopy can be easily used to assign a rigid molecular structure. But not all compounds however are rigid, for example rotation can occur about the bonds in a molecule, rings can spin or change conformation and ligands can migrate between different sites. These intramolecular processes can be easily investigated by NMR and are called fluxional process. A well-investigated example in organometallic chemistry is constituted by  $(\text{Cp})_2\text{Fe}_2(\text{CO})_4$  (Figure 2) that exists as a mixture of *cis* and *trans* isomers interconverting in solution. At temperature down to  $-48^\circ\text{C}$  one single resonance is observed. At lower temperatures the signal splits into two, assignable to the *trans* and *cis* isomers (Figure 3). Analysis of changes in the NMR lineshapes with

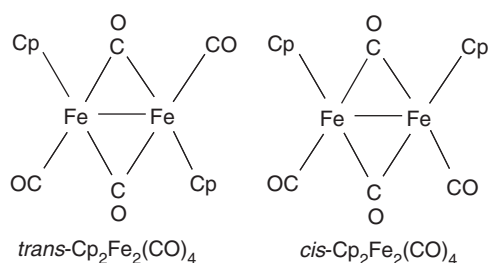


Figure 2 *cis* and *trans* isomers of  $\text{Cp}_2\text{Fe}_2(\text{CO})_4$ .

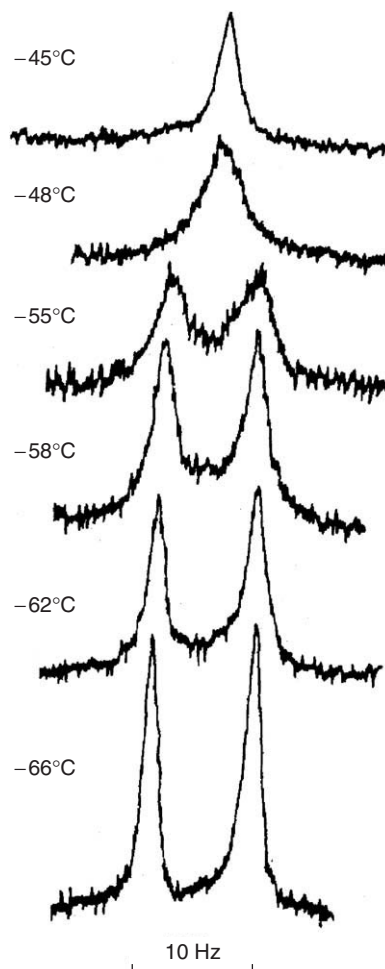
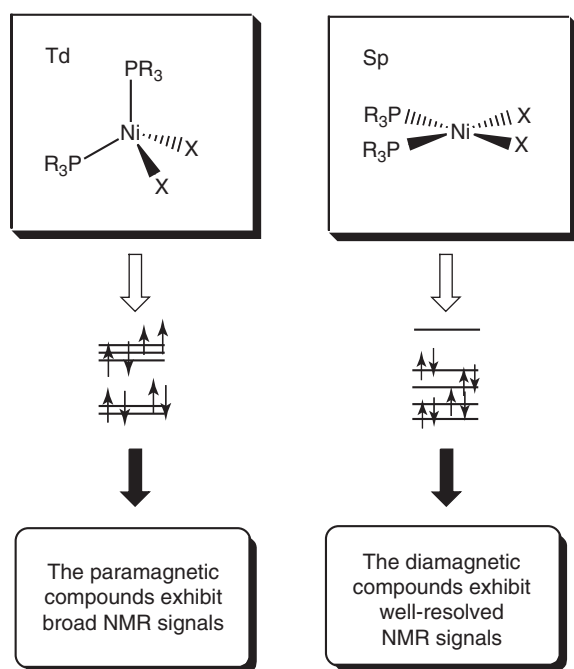
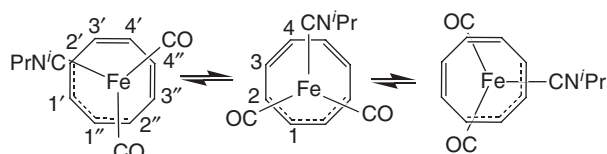


Figure 3 The signal splitting observed at low temperature for  $\text{Cp}_2\text{Fe}_2(\text{CO})_4$ .

temperature allowed to extract the rates of the fluxional processes, and in addition,  $\Delta G^\ddagger$ ,  $\Delta H^\ddagger$ , and  $\Delta S^\ddagger$  for both the *cis*–*trans* isomerization and for the equivalencing of the carbonyl ligands and then to individuate the (concerted) mechanism. Coordination number isomerization can be also investigated: for example, two geometries are possible for species such as  $\text{Ni}(\text{PR}_3)_2\text{Cl}_2$ , tetrahedral, or square-planar

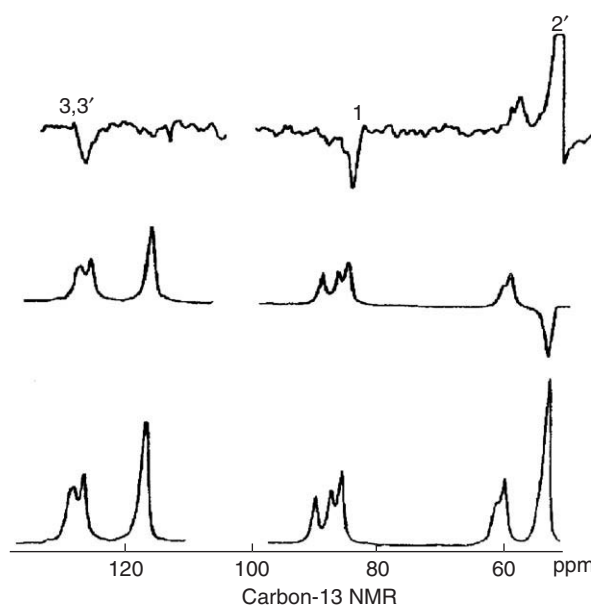


**Figure 4** Diamagnetic and paramagnetic  $\text{Ni}(\text{PR}_3)_2\text{Cl}_2$  compounds.



**Figure 5** The migration of the metal fragment in  $(\eta^4\text{-cyclooctatetraenyl})\text{Fe}(\text{CO})_2(\text{CN}'\text{Pr})$ .

(Figure 4). In a tetrahedral coordination environment the compound is paramagnetic and exhibits broad NMR signals, whereas the square-planar species are diamagnetic and exhibit well-resolved NMR signals. Dynamic processes involving polyene and polyenyl ligands such as cyclopentadiene, benzene, cycloheptatriene have been widely investigated since these molecules may spin about the metal to ligand axis or may contain free olefinic bonds allowing the possibility of exchange between the free and bound carbon-carbon double bonds. We can, for example, consider the migration of the metal fragment around the  $\eta^4\text{-cyclooctatetraenyl}$  (cot) group in  $[(\eta^4\text{-cot})\text{Fe}(\text{CO})_2(\text{CN}'\text{Pr})]$  (Figure 5). Several migration pathways are possible, for example, a stepwise migration by [1,2] shifts or migration by [1,3], [1,4], [1,5] shifts, the latter being predicted by the Woodward-Hoffmann rules. This has been confirmed by saturating the resonance of a specific carbon atom in the cot ligand. As the metal fragment migrates around



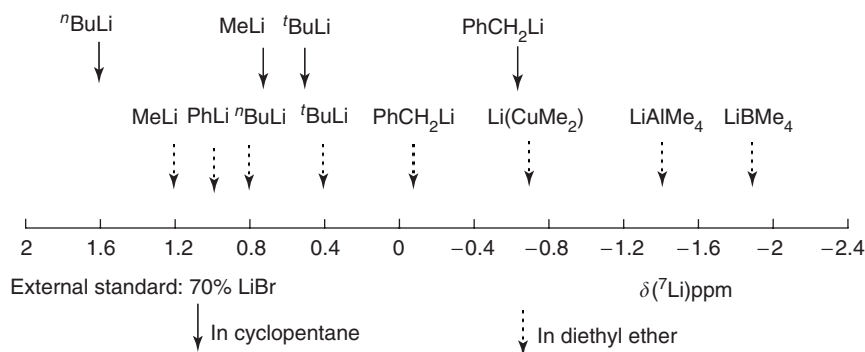
**Figure 6** Excitation at carbon 2' in  $[\eta^4\text{-cyclooctatetraenyl}]\text{Fe}(\text{CO})_2(\text{CN}'\text{Pr})$  (bottom spectrum) is transferred to carbons 1, 3, and 3' ([1,5] sigmatropic shifts). (Reproduced with permission from The Royal Society of Chemistry.)

the polyenyl ligand, we can identify the sites visited by the excited carbon by following the appearance of the excitation (Figure 6). The initial saturation, which appears as negative NMR intensity at  $\text{C}^1$  moves initially to  $\text{C}^2$ ,  $\text{C}^3$ , and  $\text{C}^4$ , consistent only with the occurrence of a [1,5] migration pathway.

## The Alkali and Alkaline Earth Metals

Lithium, beryllium, and magnesium are the only members exhibiting a covalent chemistry. All the elements of this group have nuclei with quadrupole moments, and for this reason high-resolution NMR spectra of most of these nuclei exhibit severe line broadening. Notable exceptions are  $^9\text{Be}$ ,  $^{133}\text{Cs}$ , and the lithium isotopes  $^6\text{Li}$  and  $^7\text{Li}$ . The former has been the nuclide of choice for the study of organolithium compounds. The lithium chemical shifts of alkyl-lithium compounds seem to be determined completely by neighbor anisotropic effects.  $^{13}\text{C}$  and  $^7\text{Li}$  NMR are ideal methods for studying the degree of aggregation of alkyl-lithium compounds and also for the investigation of exchanges occurring between aggregates. Solvent effects are generally important: in fact solvating power affects the polarity of  $\text{Li-C}$  bond and then influences the association degree. In Figure 7 lithium chemical shifts are summarized, whereas selected coupling constant values are reported in Table 4. The  $\delta(\text{Li})$  covers a small range





**Figure 7** Selected  $\delta(^7\text{Li})$  chemical shifts for organolithium compounds.

**Table 4** Selected coupling constants involving lithium nuclei

Compound	$^1J(^6\text{Li}-^{13}\text{C})$ (Hz)	$^1J(^7\text{Li}-^{13}\text{C})$ (Hz)
(MeLi) <sub>4</sub>		14.5–15
( <sup>n</sup> BuLi) <sub>4</sub>		14
( <sup>n</sup> BuLi) <sub>4</sub>		10–14
(Br <sub>3</sub> CLi) <sub>4</sub>		43
<sup>13</sup> CH <sub>3</sub> Br <sub>2</sub> <sup>13</sup> CLi		45
( <sup>n</sup> PrLi) <sub>6</sub>	3.35	
( <sup>n</sup> PrLi) <sub>8</sub>	2.48	
( <sup>n</sup> PrLi) <sub>9</sub>	2.22	

(10 ppm) and in particular lithium covalent compounds appear at low field (2 ppm range). The detection of coupling constants between carbon and lithium indicate the presence of a covalent bond. For example the <sup>13</sup>C NMR spectrum of (<sup>n</sup>BuLi)<sub>4</sub> in ether indicates coupling of <sup>13</sup>C with three equivalent lithiums, the  $^1J(^{13}\text{C}, ^7\text{Li})$  values being 14 Hz at  $-70^\circ\text{C}$ , whereas in the case of (<sup>t</sup>BuLi)<sub>4</sub> in hydrocarbon solvents, the <sup>13</sup>C NMR spectrum indicates coupling of the  $\alpha$ -carbon with four equivalent lithiums [ $^1J(^{13}\text{C}, ^7\text{Li}) = 10\text{--}11$  Hz].

Very recently an isotopic finger print method which rests on deuterium-induced isotope shifts for <sup>6</sup>Li resonances has been proposed by Günther and exemplified with applications from the aggregation behavior of cyclopropyllithium systems and mixed aggregates formation between methyllithium and lithium salts. By using one- and two-dimensional NMR experiments based on scalar spin–spin coupling and nuclear overhauser effects the structural aspects associated with benzyllithium and the formation of polylithium systems by lithium reduction of biphenylenes have been discussed.

NMR studies on organoberyllium compounds are more limited, cyclopentadienyl (Cp) beryllium species being generally the most investigated. For example <sup>9</sup>Be resonance for CpBeMe and CpBeCl have been reported and compared with those found

**Table 5** <sup>9</sup>Be NMR parameters of selected organometallic compounds

Compound	Solvent	$\delta(^9\text{Be})^a$ (ppm)	$W_{1/2}$ (Hz)
CpBeMe	Cyclohexane	+ 20.1	7
	Benzene	+ 20.6	10
	Toluene	+ 20.5	9
CpBeCl	Benzene	+ 18.8	3
CpBeBr	Benzene	+ 19.5	4
Cp <sub>2</sub> Be	Toluene	+ 18.5	
	Methylcyclohexane	+ 18.3	
MeBeCl.2Me <sub>2</sub> S	Me <sub>2</sub> S	– 4.2	
Me <sub>2</sub> Be.Me <sub>2</sub> S	Me <sub>2</sub> S	– 11.6	
Me <sub>2</sub> Be.2PMe <sub>3</sub>	PMe <sub>3</sub>	– 3.6	
Me <sub>2</sub> Be.2NMe <sub>3</sub>	NMe <sub>3</sub>	– 12.0	
Me <sub>2</sub> Be.NMe <sub>3</sub>	Cyclohexane	– 19.9	
Me <sub>2</sub> Be.OEt <sub>2</sub>	Et <sub>2</sub> O	– 20.8	

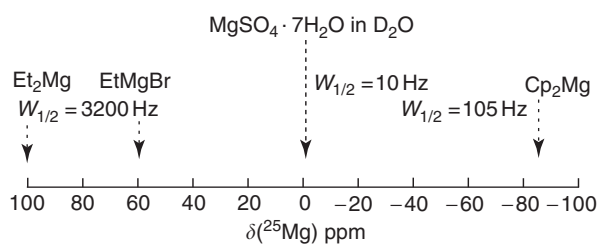
<sup>a</sup>With reference to  $\text{Be}(\text{H}_2\text{O})_4^{2+}$ .

for beryllium complexes. Selected values of <sup>9</sup>Be chemical shifts are reported in Table 5. The great shielding of the beryllium nucleus found in Cp species has been assigned to strong ring current effects. Some beryllboranes have been recently investigated with <sup>9</sup>Be. In CpBeBH<sub>4</sub>, coupling is observed of beryllium to boron and hydrogen of the BH<sub>4</sub> group. It has been also proposed that coordination number has great effect on  $\delta(^9\text{Be})$ , four-coordinate beryllium species being generally at higher field with respect to three-coordinate compounds. A variable NMR temperature study of Cp<sub>2</sub>Be has measured the barrier height between the two equivalent potential wells for a pure pentahapto and a slip structure. The Cp multiplet fine structure found in the NMR spectrum of CpBePh suggests a  $\pi$ -complex structure or a fast migration of the PhBe moiety around the Cp ring. It is worth noting that <sup>13</sup>C–<sup>9</sup>Be coupling was not observed for CpBePh, it was for CpBeBr.

The <sup>25</sup>Mg NMR area has been scarcely explored, however, some <sup>25</sup>Mg NMR parameters (Table 6)

**Table 6**  $^{25}\text{Mg}$  parameters of selected magnesium compounds<sup>a</sup>

Compound	Solvent <sup>b</sup>	Concentration ( $\text{mol l}^{-1}$ )	$\delta(^{25}\text{Mg})$	$W_{1/2}$ (Hz)	$\delta(^{13}\text{C})^b$
$\text{Et}_2\text{Mg}$	THF	1.78	99.2	3200	
$(\text{C}_4\text{H}_7)_2\text{Mg}$	THF	1.20	70.4	2000	
$\text{MeMgBr}$	THF	1.63	67.8	1900	
$\text{EtMgBr}$	THF	1.26	56.2	1100	
$(\text{C}_3\text{H}_5)_2\text{MgBr}$	THF	1.00	29.7	1000	
$\text{Cp}_2\text{Mg} \cdot \text{TMED}$	TMED	0.03	-25.8	50	103.9
$\text{CpMgBr}$	THF	1.51	-26.8	60	105.7
$\text{Cp}_2\text{Mg} \cdot \text{TMBD}$	Toluene	0.65	-31.0	140	105.3/108.5
$\text{Cp}_2\text{Mg} \cdot n\text{ THF}$	THF	1.34	-33.8	90	105.8
$(\text{CpMgBr} \cdot \text{OEt}_2)_2$	Toluene	0.81	-37.9	140	106.5
$\text{Cp}_2\text{Mg} \cdot \text{dioxane}$	Dioxane	0.01	-48.5	~300	105.8
$\text{Cp}_2\text{Mg} \cdot \text{DME}$	DME	0.30	-49.4	75	105.2
$\text{Cp}_2\text{Mg} \cdot n\text{ PMe}_3$	$\text{PMe}_3$	0.13	-60.9	220	105.7
$\text{Cp}_2\text{Mg} \cdot n\text{ OEt}_2$	$\text{Et}_2\text{O}$	1.45	-75.3	150	106.9
$\text{Cp}_2\text{Mg}$	Toluene	0.75	-85.4	105	107.7

<sup>a</sup> At 24.49 MHz and 309 K. External standard:  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  in  $\text{D}_2\text{O}$ .<sup>b</sup> THF, tetrahydrofuran; TMED, 1,2-bis(dimethylamino)ethane; DME, 1,2-dimethoxyethane; TMBD, 1,3-bis(dimethylamino)butane.**Figure 8** Selected  $\delta(^{25}\text{Mg})$  chemical shifts for organo-magnesium compounds.

have been detected for a number of organomagnesium compounds  $\text{R}_2\text{Mg}$ ,  $\text{RMgX}$  ( $\text{R}$  = alkyl, aryl, allyl, Cp,  $\text{X}$  = halide) and for their adducts with Lewis bases. Ranges of  $^{25}\text{Mg}$  chemical shifts for a number of organomagnesium compounds are reported in Figure 8. It has been shown that the  $^{25}\text{Mg}$  NMR spectra for Grignard reagents depend on concentration and temperature.

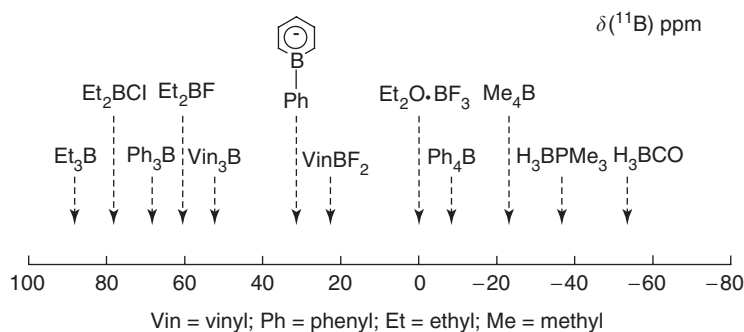
## Boron

In the case of boron we have the choice between two nuclei,  $^{10}\text{B}$  and  $^{11}\text{B}$ , the former having higher sensitivity and showing no severe line-broadening problems, although having quadrupolar nucleus. The currently accepted chemical shift standard is  $[\text{BF}_3(\text{OEt}_2)]$ . Ranges of  $^{11}\text{B}$  chemical shift for selected bonding environments are given in Figure 9. The total  $^{11}\text{B}$  shielding range covers ~200 ppm. It has been found that the shielding range for four-coordinate  $\text{sp}^3$  species are generally several tens of parts per million to higher field of the nearest equivalent three-coordinate species which have generally an  $\text{sp}^2$

trigonal boron bonding environment. This fact can be employed for the assignment of structures and coordination number to nonpolyhedra boron compounds. This also means that the addition of a ligand or base to the empty p-orbital on boron results in an upfield shift as compared to the tricoordinate borane. The chemical shift is dependent on the strength of the coordination complex, with the stronger complexes shifted to higher field. Borane dimers are also found upfield from the free monomer and may show a more complex coupled spectra due to the presence of both terminal and bridging hydrogens coupling to the boron.

Selected couplings for trigonal and tetrahedral compounds are given in Table 7. One-bond couplings are dominated by the Fermi contact mechanism that means a dependence on hybridization at boron. The coupled  $^{11}\text{B}$  NMR spectra generally show a B-H coupling, unless there is a fast exchange of the hydrogens on the NMR timescale. The chemical shifts of the trivalent organoboranes having a B-H bond are more variable than that of the trialkylboranes, and are dependent on the structure of the organic group present. The trialkylboranes are found in a narrow low-field region, 83–93 ppm, the chemical shifts being widely independent of the structure of the alkyl group. Substitution at the  $\alpha$ -carbon,  $\text{R}_2\text{B}-\text{CH}_2-\text{X}$  with  $\text{X} = \text{N}_3$ , OH,  $\text{NH}_2$ , Cl, Br, Ph, Vi (Vi = vinyl) results in an upfield shift of the borane, an effect that can be assigned to inter- or intramolecular effects of the lone electron pair with the empty p-orbital on the boron or a neighboring group anisotropic effect.

Polyhedral boron-containing compounds are the most investigated species: one-bond couplings to exo-terminal protons  $^1J(^{11}\text{B}, ^1\text{H})$  are typically in the



**Figure 9** Selected  $\delta(^{11}\text{B})$  chemical shifts for organoboron compounds.

**Table 7** Selected coupling constants  $^1J(^{11}\text{B}, \text{X})$  for trigonal and tetrahedral boron compounds

Compound	X	$^1J(^{11}\text{B}, \text{X})$ (Hz)
$\text{Me}_3\text{B}$	$^{13}\text{C}$	47
$\text{MeB}(\text{OMe})_2$	$^{13}\text{C}$	76
$(\text{Me}_2\text{N})_2\text{BH}$	$^1\text{H}$	126
$(\text{MeO})_2\text{BH}$	$^1\text{H}$	141
$\text{Me}_2\text{BF}$	$^{19}\text{F}$	119
$\text{MeBF}_2$	$^{19}\text{F}$	76
$\text{Me}_3\text{SnBH}_3$	$^{119}\text{Sn}$	-554
$\text{B}(\text{NHMe}_3)$	$^{15}\text{N}$	-45
$\text{Me}_2\text{P-B}(\text{NMe}_2)_2$	$^{31}\text{P}$	50
$\text{H}_3\text{B} \cdot \text{CO}$	$^1\text{H}$	105
$\text{HCl}_2\text{B} \cdot \text{PH}_3$	$^1\text{H}$	154
$\text{H}_3\text{B} \cdot \text{PH}_3$	$^1\text{H}$	104
$\text{H}_3\text{B} \cdot \text{AsMe}_3$	$^1\text{H}$	100
$\text{H}_3\text{B} \cdot \text{NH}_3$	$^1\text{H}$	94
$\text{BH}_4^-$	$^1\text{H}$	80
$[\text{BMe}_4]^-$	$^{13}\text{C}$	22
$[\text{BPh}_4]^-$	$^{13}\text{C}$	49
$\text{Et}_2\text{FB} \cdot \text{NEt}_3$	$^{19}\text{F}$	67
$\text{F}_3\text{B} \cdot \text{NEt}_3$	$^{19}\text{F}$	16
$\text{F}_3\text{B} \cdot \text{CH}_2 \cdot \text{PMe}_3$	$^{19}\text{F}$	53
$\text{H}_3\text{B} \cdot \text{PF}_3$	$^{31}\text{P}$	39

range 120–170 Hz, consistent with  $sp$  hybridization of an exo-bond at boron on the basis of  $sp^3$  hybridization for  $[\text{BH}_4]^-$  which has  $^1J(^{11}\text{B}, ^1\text{H}) \sim 80$  Hz.

## Aluminum, Gallium, and Thallium

$^{27}\text{Al}$  is a high sensitivity nucleus that yields broad lines (quadrupole moment of 5/2) over a moderate chemical shift range. The main use is to determine the presence of aluminum or the number of different sites that it occupies. Hexacoordinate aluminum resonates between 0 and 100 ppm, while tetracoordinate aluminum can resonate between 200 and -200 ppm. The standard usually used is  $\text{AlCl}_3$  ( $1 \text{ mol l}^{-1}$  in  $\text{D}_2\text{O}$ ).  $^{27}\text{Al}$  NMR can be used to indicate the aggregation degree of organoaluminum compounds. The NMR spectrum of the 1:1 complex

of  $\text{Me}_3\text{Al}$  and  $N,N$ -dimethylethylene diamine in toluene exhibits a single resonance downfield shifted (177.1) with respect to that of  $\text{Me}_3\text{Al}$  (154 ppm). It has been proposed that the 1:1 adduct may exist as dimeric or oligomeric complex, as usually observed with many organoaluminum complexes. A range of  $^{27}\text{Al}$  NMR spectra of alkylaluminum compounds and their oxo-derivatives, recently measured exhibit narrow linewidths, the shift of which depend only upon the coordination number of the Al, though the alkyl ligands cause a marked downfield shift. The dimer  $[\text{Me}_2\text{Al} \dots \gamma \text{ picoline}]_2$  gives a resonance at 44.4 ppm, remarkably to high field of monomeric  $\text{Me}_3\text{AlNC}_5\text{H}_4\text{Me}$ , which is found at 174.7 ppm.  $[\text{AlH}_4]^-$  has been studied in a variety of systems, the Al-H coupling constant was found to be somewhat larger than early measurements had indicated, at  $\sim 170$  Hz. The  $\delta(^{27}\text{Al})$  in  $[\text{AlH}_4]^-$  is slightly dependent on the nature of the cation.

The coupling constants have been determined between  $^{27}\text{Al}$  and  $^{13}\text{C}$  or  $^1\text{H}$  in  $[\text{AlMe}_4]^-$  and in  $[\text{AlEt}_4]^-$  and ion-pairing detected. A few further studies have described the aluminum borohydride system. Shifts and coupling constants are recorded for mixed alkylhydride-borohydrides and their adducts. A chart showing the main features of  $^{27}\text{Al}$  chemical shift variation is reported in Figure 10.

Gallium has two magnetically active isotopes,  $^{69}\text{Ga}$  and  $^{71}\text{Ga}$ . The least abundant  $^{71}\text{Ga}$  is the one preferred for NMR because of its higher receptivity and slightly less efficient quadrupolar relaxation. Gallium-hydrogen spin-spin coupling has been observed in  $[\text{GaH}_4]^-$  anion, the chemical shift being 110 ppm. The chemical shift range for hydride compounds is  $\sim 5.7$  times greater for  $^{71}\text{Ga}$  than for  $^{27}\text{Al}$ . The influence of additional ligands is also different for the two elements in the two groups of compounds: in octahedral complexes the smaller  $\text{Al}^{3+}$  brings about greater steric crowding of the ligand atoms, whereas there is no such crowding in the tetrahedral complexes.

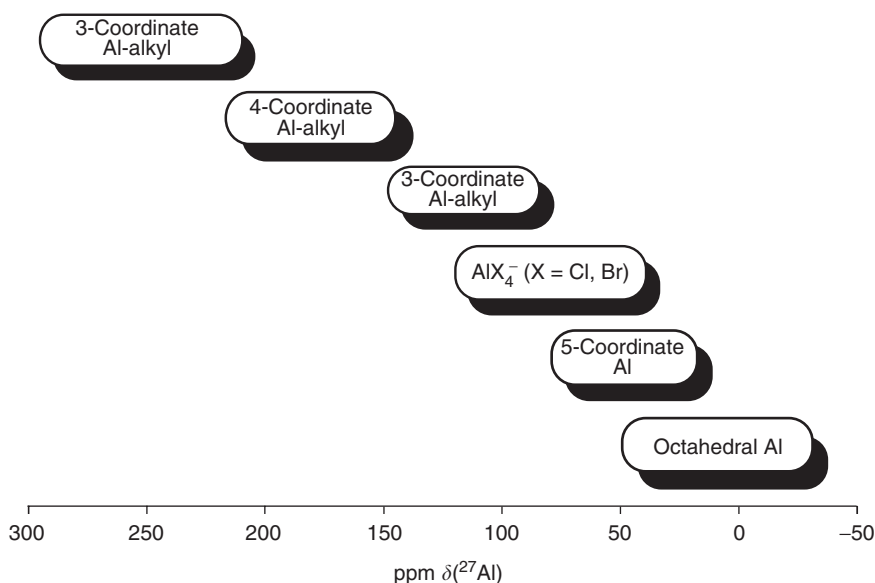


Figure 10 Trend of  $\delta(^{27}\text{Al})$  chemical shifts.

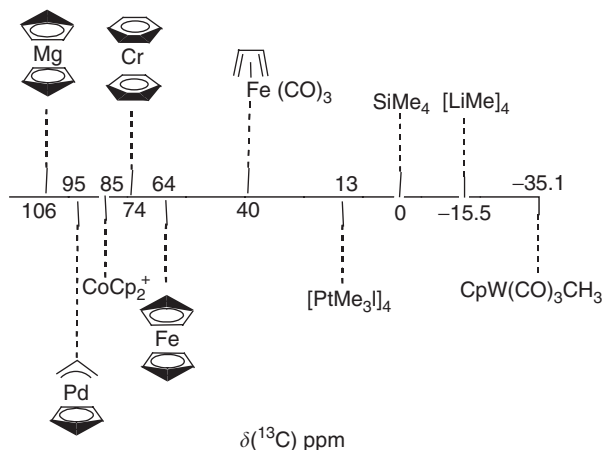


Figure 11 Selected examples of  $\delta(^{13}\text{C})$  chemical shifts.

The  $^{205}\text{Tl}$  chemical shifts of organothallium (III) compounds cover a range of over 2000 ppm and they display solvent dependence, although not as striking as that of Tl (I). Solvent basicity and ion pairing contribute to the solvent dependence. It has been found that temperature produces greater changes on the  $^{205}\text{Tl}$  chemical shifts of  $\text{Me}_2\text{TlX}$  salts with respect to other factors such as solvent concentration and nature of X. Substituent changes produce large chemical shifts, for example, the difference in chemical shift between  $\text{Me}_3\text{Tl}$  and  $\text{Et}_3\text{Tl}$  is 130 ppm. The relaxation behavior of  $^{205}\text{Tl}$  has been investigated for  $\text{Me}_2\text{Tl}^+$  in solution as a function of temperature and magnetic field strength and the dominant relaxation mechanism was found to be chemical shift anisotropy.

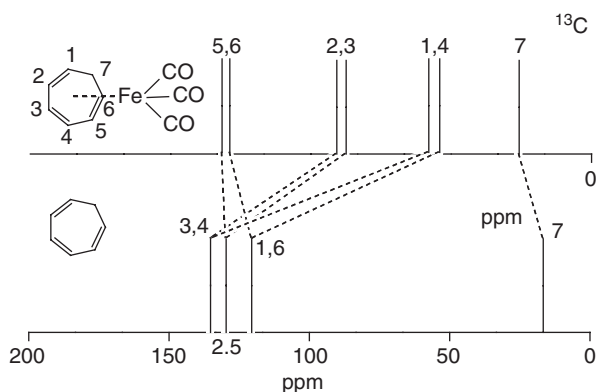


Figure 12  $\delta(^{13}\text{C})$  chemical shift values of the complex  $(\text{COT})\text{Fe}(\text{CO})_3$  compared with those of the free COT.

## Carbon

$^{13}\text{C}$  NMR spectroscopy is an extremely useful technique in organometallic chemistry. The following trends are generally found: saturated carbons appear in the range 0–100 ppm, electronegative substituents on them increasing the shift. Chemical shifts for aromatic carbons are in the range 110–170 ppm, but  $\pi$ -bonded metal alkene may be shifted up to 100 ppm: the shift is strongly dependent on the mode of coordination. Metal carbonyls span from 170 to 290 ppm (they show long relaxation times that make their detection difficult). Metal carbene exhibits resonances in the range 250–370 ppm. A trend of  $\delta(^{13}\text{C})$  chemical shift is reported in Figure 11, whereas in Figure 12 a simulated  $^{13}\text{C}$  NMR spectrum for free cyclohepta- triene has been compared with the data reported for

the coordinated cycloheptatriene in ( $\eta^4$ -cycloheptatriene)Fe(CO)<sub>3</sub>.

## Silicon

<sup>29</sup>Si NMR is a low sensitivity nucleus, which generally yields sharp lines. It has a wide chemical shift range useful for the determination of chemical environment in silicon compounds. It is worth noting that the spectra always contain a background signal from the glass in the tube and probe at 90 ppm. The commonly used reference compound is SiMe<sub>4</sub>. Although several researchers will not find a general need for this technique, as a solution technique it is only slightly more difficult than a <sup>13</sup>C experiment in most cases (solid-state <sup>29</sup>Si NMR is a valuable tool for zeolites and solid-state chemistry). The inductive effect of Si typically moves <sup>1</sup>H NMR aliphatic resonances upfield to ~0–0.5 ppm, making assignment of Si-containing groups rather easy. In addition, both carbon and proton spectra display Si satellites comprising 4.7% of the signal intensity.

Although the important earlier work on <sup>29</sup>Si used indirect <sup>1</sup>H–(Si) double resonance method of detection, the preferred method is now the direct Fourier transform technique by using special pulse sequences as INEPT and DEPT to effect spin polarization transfer from protons and hence achieving dramatic increases in the signal-to-noise ratio with corresponding reductions of spectral accumulation times. Much work has been done on <sup>29</sup>Si NMR in the solid state using magic angle spinning (MASS) and cross-polarization (CP) methods. The overall chemical shift range known is now ~600 ppm for silicon and the pattern is shown in Figure 13. An increase in coordination number from four to five or six produces a significant increase in the shielding of tetravalent silicon. It has been found that in cationic and neutral hexacoordinate silicon complexes such as those reported in Figure 14 the  $\delta(^{29}\text{Si})$  ranges from –198 to –188 ppm, i.e., to lower frequencies with respect to four-coordinate species and show little dependence on the nature of R groups or indeed upon the formal charge on silicon. Five coordinate species as that in Figure 15 exhibit  $\delta(^{29}\text{Si})$  in the range –180 to –160, again at low frequency with respect to four-coordinate species. Silicon can participate in  $p_\pi$ – $p_\pi$  bonding and this fact can lead to significant changes in the chemical shift: for example the silene species containing an  $sp^2$ -hybridized silicon atom exhibit a  $\delta(^{29}\text{Si}) = +41.4$  ppm. The existence of  $d_\pi$ – $d_\pi$  interaction when it is bonded to transition metals does not produce sensible variation on the

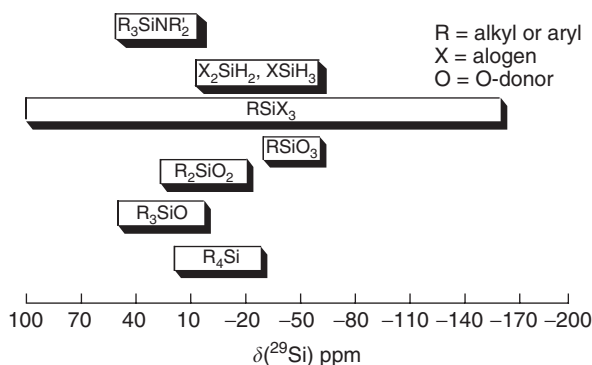


Figure 13 Trend of  $\delta(^{29}\text{Si})$  chemical shifts.

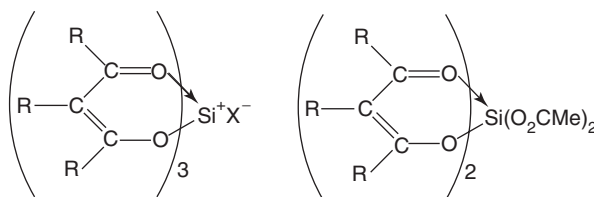


Figure 14 Diketonate complexes of Si.

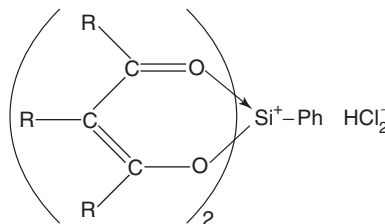


Figure 15 Five-coordinate organosilicon compound.

chemical shifts values. Also polymeric siloxanes can be easily investigated by NMR. The silicon atoms in terminal  $\text{R}_3\text{SiO}-$ , chain  $\text{R}_2\text{Si}(\text{O}-)_2$ , branch  $\text{R}-\text{Si}(\text{O}-)_3$  and quaternary  $\text{Si}(\text{O}-)_4$  groups have very different chemical shifts that makes it possible to study both the polymerization degree and cross-linking. Good correlations have been found between the coupling constants and the s character and/or effective nuclear charge of the central atom: in  $\text{Me}_3\text{SiX}$ , the  $^1J(^{29}\text{Si}, ^{13}\text{C})$  is linearly related to the electronegativity of X. Analogous correlations have been found for  $\text{Me}_n\text{Si}(\text{OEt})_{4-n}$ ,  $(\text{Me}_3\text{Si})_3\text{COR}$ , silyl acetylenes, and tetraorganosilanes.

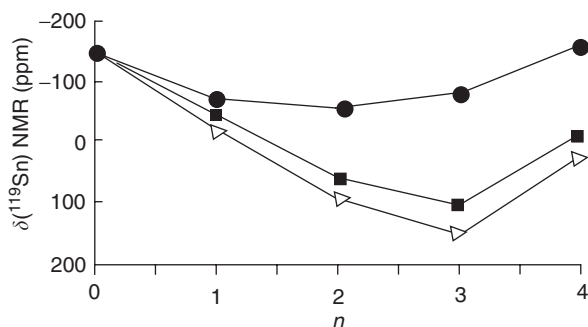
## Tin, Germanium, and Lead

Group 14 of the periodic table has generally favorable NMR properties: with the exception of Ge, all elements have at least one spin 1/2 isotope of good sensitivity. Moreover, in the past years, one-dimensional (1D) and two-dimensional (2D) proton detected heteronuclear correlation NMR techniques



involving  $^{119}\text{Sn}$  and  $^{207}\text{Pb}$  nuclei have been used for applications in organometallic NMR. Germanium possesses only one magnetic isotope,  $^{73}\text{Ge}$ , with spin  $9/2$ . Tin possesses three magnetically active tin isotopes with spin  $I = 1/2$  ( $^{115}\text{Sn}$ ,  $^{117}\text{Sn}$ ,  $^{119}\text{Sn}$ ). Two of them,  $^{117}\text{Sn}$  and  $^{119}\text{Sn}$  have an appreciable natural abundance and receptivities  $\sim 20$  and  $25$  times greater, respectively, than that of  $^{13}\text{C}$ .  $^{207}\text{Pb}$ , the only lead isotope with spin  $1/2$  (the natural abundance is  $\sim 22\%$ ), has a positive gyromagnetic ratio. Germanium, tin, and lead chemical shifts cover a very large range if inorganic and organometallic compounds are examined: more than  $1200$  ppm for  $^{73}\text{Ge}$ ,  $5000$  ppm for  $^{119}\text{Sn}$ , and  $12\,000$  ppm for  $^{207}\text{Pb}$ . Table 8 exhibits selected  $^{73}\text{Ge}$ ,  $^{119}\text{Sn}$ , and  $^{207}\text{Pb}$  data, all referred to  $(\text{Me})_4\text{E}$  ( $\text{E} = \text{Ge}, \text{Sn}, \text{or Pb}$ ), a negative sign indicating a higher shielding or a shift to lower frequencies. Tin, germanium, and lead chemical shifts are not very sensitive to temperature in the absence of chemical changes (association, dissociation, changes in the coordination number). Shielding of the  $^{119}\text{Sn}$  nucleus generally increases roughly linearly with increasing temperature. Change in the solvent, in the absence of solvent-solute chemical interactions, autoassociation or dissociation affects the tin, germanium, and lead chemical shifts only slightly with respect to their chemical shift ranges. The knowledge of  $\delta(^{119}\text{Sn})$  chemical shifts is very important for predicting the geometry of coordination polyhedra of tin compounds in solution. In the case of tetravalent tin compounds, an increase in tin coordination number from four to five corresponds to an increase in tin

shielding by  $\sim 150$ – $200$  ppm. The magnitude of the chemical shift can permit determination of association constants between the tin compound and the solvent. In several organometal derivatives, changes in the coordination number can arise from self-association, as in the case of organotin alkoxides. Several  $\text{R}_4-n\text{Sn}(\text{OR}')_n$  compounds show values of chemical shifts suggesting self-association as neat liquid or in noncoordinating solvents. For example butyltin(IV) trialkoxides with an unbranched alkoxy group (such as  $\text{BuSn}(\text{OMe})_3$ ) are six-coordinate at room temperature, the association presumably involving octahedral structures with both bridging and not bridging alkoxy oxygen atom. The effect of ligand electronegativity can be demonstrated by the changes of  $^{119}\text{Sn}$  chemical shifts along the series  $\text{R}_n\text{SnCl}_{4-n}$  ( $\text{R} = \text{alkyl or aryl}$ ) as the values of  $n$  changes from 2 to 0. However, the plot (Figure 16) of the  $\delta(^{119}\text{Sn})$  for the compounds  $\text{R}_n\text{SnCl}_{4-n}$  indicates clearly that



**Figure 16** Plot of  $\delta(^{119}\text{Sn})$  against  $n$  for a series of monomeric organotin(IV) compounds:  $\Delta$ , Me;  $\blacksquare$ , Bu; and  $\bullet$ , Ph.

**Table 8** Selected  $^{73}\text{Ge}$ ,  $^{119}\text{Sn}$ , and  $^{207}\text{Pb}$  NMR chemical shifts in parts per million relative to  $\text{Me}_4\text{E}$  ( $\text{E} = \text{Ge}, \text{Sn}, \text{and Pb}$ )

Compound	$\delta$ (ppm)	Notes	Compound	$\delta$ (ppm)	Notes
$\text{Me}_4\text{Ge}$	0		$\text{Me}_3\text{SnPh}$	-28.6	$\text{CH}_2\text{Cl}_2$
$\text{Et}_4\text{Ge}$	17.3		$\text{MeSnPh}_3$	-98	$\text{CH}_2\text{Cl}_2$
$\text{Ph}_4\text{Ge}$	-31.6		$\text{Me}_2\text{SnPh}_2$	-60	Neat liquid
$\text{GeH}_4$	-283.7		$\text{Me}_3\text{Sn}(\text{C}\equiv\text{CH})$	-68.1	$\text{CH}_2\text{Cl}_2$
$\text{EtGeH}_3$	-186.4		$\text{MeSnCl}_3$	+18.7	Benzene
$\text{Et}_2\text{GeH}_2$	-88.4		$\text{PhSnCl}_3$	-64	$\text{CDCl}_3$
$\text{Et}_3\text{GeH}$	-15.7		$\text{Me}_3\text{SnCl}$	+168.9	$\text{CH}_2\text{Cl}_2$
$\text{Ge}(\text{CH}=\text{CH}_2)_4$	-58.7		$\text{Ph}_3\text{SnCl}$	-44.7	$\text{CDCl}_3$
$\text{Me}_3\text{Ge}(\text{CH}=\text{CH}_2)$	-15.6		$\text{Me}_2\text{SnCl}_2$	+137	$\text{CH}_2\text{Cl}_2$
$\text{Me}_2\text{Ge}(\text{CH}=\text{CH}_2)_2$	-30.6		$\text{Ph}_3\text{SnOH}$	-86	$\text{CH}_2\text{Cl}_2$
$\text{MeGe}(\text{CH}=\text{CH}_2)_3$	-44.9		$\text{Cp}_2\text{Sn}$	-2199	$\text{C}_6\text{H}_{12}$
$\text{Ge}(\text{C}\equiv\text{CH})_4$	-173		$\text{Me}_3\text{SnH}$	-104.5	Benzene
$\text{MeGe}(\text{C}\equiv\text{CH})_3$	-118		$\text{Me}_2\text{SnH}_2$	-229	Neat liquid
$\text{Me}_2\text{Ge}(\text{C}\equiv\text{CH})_2$	-77		$\text{Me}_4\text{Pb}$	0	15% toluene
Germacyclohexane	-34		$\text{Et}_4\text{Pb}$	+73.3	Neat liquid
$\text{Me}_4\text{Sn}$	0	Neat liquid	$\text{Vi}_4\text{Pb}$	-251	Neat liquid
$\text{Et}_4\text{Sn}$	+1.4	$\text{CCl}_4$	$\text{Pb}_2\text{Ph}_6$	-79.8	$\text{CDCl}_3$
$\text{Ph}_4\text{Sn}$	-137	$\text{C}_2\text{H}_5\text{Cl}_3$	$\text{Me}_3\text{PbCl}$	+432	$\text{CH}_2\text{Cl}_2$
$\text{Me}_3\text{SnEt}$	+3.0	$\text{CCl}_4$	$\text{Me}_2\text{PbCl}_2$	-222	DMSO

Me, methyl; Et, ethyl; Vi, vinyl; Ph, phenyl.

there are no obvious general relationships between  $\delta(^{119}\text{Sn})$  and the ligand electronegativity. The U-shaped trend is usually assigned to change in the paramagnetic term due to p electron imbalance, generally maximum at  $n=2$ . In a series of methylvinylgermanes, the  $^{73}\text{Ge}$  chemical shifts were compared with those of the corresponding methylvinylsilanes, and the correlation  $\delta(^{73}\text{Ge}) = 1.96 \delta(^{29}\text{Si}) + 2.37$  was found. It has been found that the magnitude of the chemical shift cannot be simply assigned to the electron densities, but relative contributions of the diamagnetic and paramagnetic terms to the  $^{73}\text{Ge}$  chemical shift should be considered.

MO calculations indicate in monomeric tin(II) compounds such as stannocene,  $(\text{C}_5\text{H}_5)_2\text{Sn}$ , a highly shielded tin atom ( $-2199$  ppm in  $\text{C}_6\text{H}_{12}$ ) as a consequence of an inefficient paramagnetic charge circulation due to the high LUMO energy of this molecule. The values of the chemical shift ( $\delta(^{119}\text{Sn})$ ) of diphenyl-, dibenzyl-, and di-n-butyl-tin(IV) compounds exhibit a good mutual correlation which can be expressed by eqn [3]:

$$\delta(^{119}\text{Sn})[\text{R}'_2\text{Sn}] = a \times \delta(^{119}\text{Sn})[\text{R}''_2\text{Sn}] + b \quad [3]$$

where  $[\text{R}'_2\text{Sn}]$  and  $[\text{R}''_2\text{Sn}]$  represent pairs of diorganotin(IV) compounds with different substituents, and  $a$  and  $b$  represent the slope and the intercept of the linear correlations.

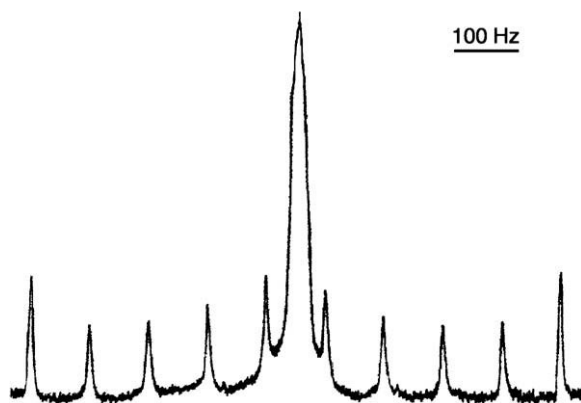
The valence electron-mediated nuclear spin-spin couplings,  $J(^{73}\text{Ge}, \text{X})$ ,  $J(^{117}\text{Sn}, \text{X})$ ,  $J(^{119}\text{Sn}, \text{X})$ , and  $J(^{207}\text{Pb}, \text{X})$ , are dominated by the 'FERMI contact' term arising from the interaction of the nuclear spin magnetic dipole with the electron spin, which has a finite density at the nucleus. In the case of  $\text{X} = ^1\text{H}$ ,  $^{19}\text{F}$ ,  $^{31}\text{P}$ , highly abundant spin 1/2 nuclei, the  $^nJ(\text{E}, \text{X})$  data are obtained conveniently from the X atom NMR spectra, whereas in the case of couplings with rare spin 1/2 nuclei ( $\text{X} = ^{13}\text{C}$ ,  $^{29}\text{Si}$ ,  $^{15}\text{N}$ ), the  $^nJ(\text{E}, \text{X})$  can be obtained also from E NMR spectra, by using Hahn-echo extended pulse sequences. Table 9 gives a selection of the values of the coupling constants in germanium, tin, and lead compounds. Spin couplings involving  $^{73}\text{Ge}$  have been generally reported only in highly symmetrical compounds. In Figure 17 the  $^1\text{H}$  NMR spectrum of  $\text{GeH}_4$  is reported. The 10 evenly spaced lines are due to 8% of the molecules, which contain  $^{73}\text{Ge}$  ( $I=9/2$ ), whereas the intense central line arises from all other isotopic species. A simple linear relationship between  $J(^{119}\text{Sn}, ^{13}\text{C})$  coupling constants of a group of dimethyltin(IV) compounds and their C-Sn-C angles ( $\theta$ ) has been found on the basis of the analysis of a large set of experimental data (eqn [4])

$$|J(^{119}\text{Sn}, ^{13}\text{C})| = (10.7 \pm 0.5)\theta - (778 \pm 64) \quad [4]$$

**Table 9** Some representative spin couplings, involving  $^{73}\text{Ge}$ ,  $^{119}\text{Sn}$ , and  $^{207}\text{Pb}$

Compound	X	E	$^1J(\text{E}, \text{X})$ values (Hz)
$\text{GeH}_4$	$^1\text{H}$	$^{73}\text{Ge}$	-97.6
$\text{GeMe}_4^a$	$^{13}\text{C}$	$^{73}\text{Ge}$	-18.7
$\text{SnH}_4$	$^1\text{H}$	$^{119}\text{Sn}$	-1930
$\text{MeSnH}_3$	$^1\text{H}$	$^{119}\text{Sn}$	-1850
$\text{Me}_2\text{SnH}_2$	$^1\text{H}$	$^{119}\text{Sn}$	-1797
$\text{Me}_3\text{SnH}$	$^1\text{H}$	$^{119}\text{Sn}$	-1744
$\text{Ph}_3\text{SnH}$	$^1\text{H}$	$^{119}\text{Sn}$	-1935.8
$\text{Me}_4\text{Sn}$	$^{13}\text{C}$	$^{119}\text{Sn}$	-340
$\text{Me}_3\text{SnCl}$	$^{13}\text{C}$	$^{119}\text{Sn}$	-380
$\text{Me}_2\text{SnCl}_2$	$^{13}\text{C}$	$^{119}\text{Sn}$	-566
$\text{Ph}_2\text{SnCl}_2$	$^{13}\text{C}$	$^{119}\text{Sn}$	-785
$\text{Me}_2\text{Sn}(\text{acac})_2$	$^{13}\text{C}$	$^{119}\text{Sn}$	-966
$(\text{Me}_3\text{Sn})_2$	$^{13}\text{C}$	$^{119}\text{Sn}$	-240
$(\text{Me}_3\text{Sn})_3\text{P}$	$^{31}\text{P}$	$^{119}\text{Sn}$	+832.5
$\text{Me}_3\text{SnSnMe}_3$	$^{119}\text{Sn}$	$^{119}\text{Sn}$	+4240
$\text{Me}_3\text{SnPH}_2$	$^{31}\text{P}$	$^{119}\text{Sn}$	+463
$\text{Me}_2\text{PbH}_2$	$^1\text{H}$	$^{207}\text{Pb}$	+2457
$\text{Me}_3\text{PbH}$	$^1\text{H}$	$^{207}\text{Pb}$	+2295
$\text{Ph}_4\text{Pb}$	$^{13}\text{C}$	$^{207}\text{Pb}$	+481
$\text{Me}_4\text{Pb}$	$^{13}\text{C}$	$^{207}\text{Pb}$	+320
$\text{Me}_3\text{PbPbMe}_3$	$^{207}\text{Pb}$	$^{207}\text{Pb}$	+290
$\text{Me}_3\text{PbSeMe}$	$^{13}\text{C}$	$^{207}\text{Pb}$	-1170
$\text{Ph}_2\text{Pb}(\text{acetate})_2$	$^{13}\text{C}$	$^{207}\text{Pb}$	1203
$\text{Me}_3\text{PbCH}=\text{CHMe}$	$^{13}\text{C}$	$^{207}\text{Pb}$	268

<sup>a</sup>  $^2J(^{73}\text{Ge}, ^1\text{H}) = +3.0$  Hz.

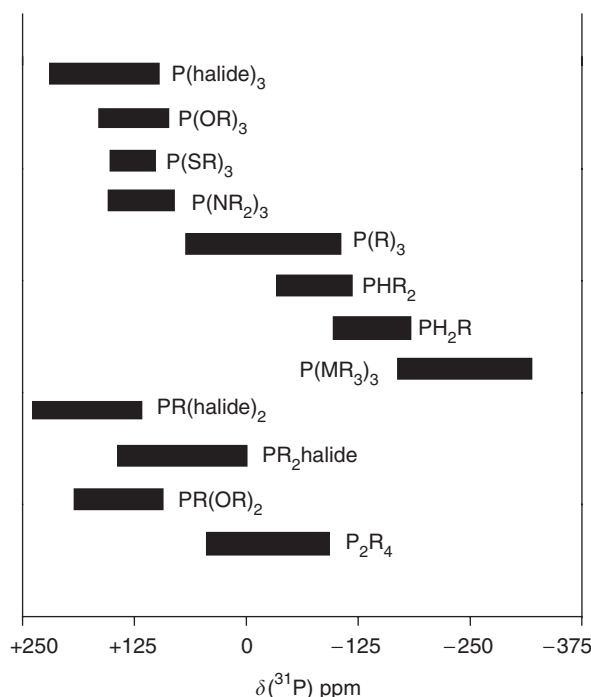


**Figure 17**  $^1\text{H}$  NMR spectrum of  $\text{GeH}_4$ .

The replacement of alkyl groups on  $\text{R}_4\text{M}$  with electro-negative substituents causes an increase in the magnitude of  $^1J(\text{M}, ^{13}\text{C})$  as the result of the greater  $\sigma$  character in the Sn-C orbital.

## Phosphorus

The  $^{31}\text{P}$  NMR technique is very useful due to the fact that it is easier than  $^{13}\text{C}$ . Although the chemical shift range is not as diagnostic as with other nuclei, the



**Figure 18** Trend of  $\delta(^{31}\text{P})$  chemical shifts.

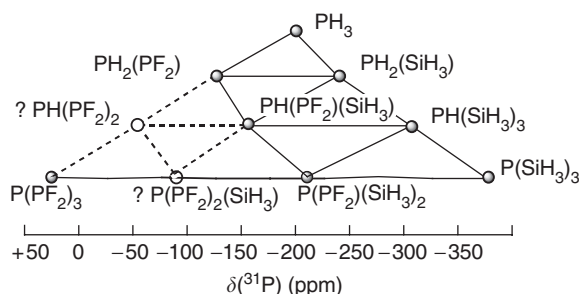
magnitude of  $^1J(^{31}\text{P}, \text{X})$  coupling constants is useful for the assignment of structures.

$^{31}\text{P}$  is spin 1/2, 100% abundant, and has a receptivity relative to the  $^1\text{H}$  of 0.0063. The range of chemical shifts is *c.*  $\pm 250$  ppm from  $\text{H}_3\text{PO}_4$ , however, very different values corresponding to highly shielded or deshielded molecule have been reported in literature, as the two extremes  $\text{P}_4$  ( $-460$  ppm) and the phosphinidene complex  $[\text{tBuP}[\text{Cr}(\text{CO})_5]_2$  ( $+1362$  ppm). Interpretation of the shift is not easy, there seems to be many contributing factors. The  $\delta(^{31}\text{P})$  at a first approximation is determined by the number and type of directly bound atoms and is relatively little affected by consideration of charge and structure. A pattern of chemical shifts for the three-coordinate P compounds is depicted in **Figure 18**. P(III) covers the whole normal range and depends strongly on the substituents. Coordination of a P(III) compound to a metal results in a coordination shift from the original resonance position of the uncoordinated ligand. In **Table 10** selected values of coordination shifts for some representative compounds are reported. With a restricted number of ligands an approximate linear correlation was found between the coordination shift,  $\Delta\delta$  and the chemical shift. The coordination shifts were mostly in the direction of decreased shielding relative to the free ligand and the shifts were larger (more positive) with highly shielded ligands. Unknown values can be predicted by extrapolation or interpolation as indicated in

**Table 10**  $\delta(^{31}\text{P})$  ligand chemical shifts (ppm) and coordination shifts (ppm)<sup>a</sup> in Cr, Mo, and W complexes containing P-donor molecules

Ligand	Ligand shift	$\text{Cr}(\text{CO})_5 \text{L}$ ( $\delta$ )	$\text{Mo}(\text{CO})_5 \text{L}$ ( $\Delta\delta$ )	$\text{W}(\text{CO})_5 \text{L}$ ( $\Delta\delta$ )
$\text{PH}_3$	-239	109	74	51
$\text{PMe}_3$	-62	69	45	26
$\text{PMePh}_2$	-28	63	43	24
$\text{P}^i\text{Bu}_3$	-33	63	45	26
$\text{PEt}_3$	-20		40	
$\text{PPh}_3$	-6	61	44	27
$\text{PClPh}_2$	80			25
$\text{PF}_3$	97		50	
$\text{PEtPh}_2$	-12	60	42	24
$\text{P}(\text{NMe}_2)_3$	123		23	
$\text{P}(\text{OEt})_3$	138		18	
$\text{P}(\text{OPh})_3$	128		26	
$\text{P}(\text{OMe})_3$	141	39	21	-4

<sup>a</sup>Coordination shift ( $\Delta\delta$ ) =  $\delta(\text{complex}) - \delta(\text{ligand})$ .



**Figure 19** Interpolation can help in the prediction of  $\delta(^{31}\text{P})$  values.

**Figure 19.** P(V) compounds exhibit a narrower range  $\delta(^{31}\text{P})$  spanning from  $-50$  to  $+100$  ppm.

Coupling constants involving  $^{31}\text{P}$  have been measured for most of the magnetically active nuclei. Their magnitudes can be very large, reaching 12 kHz in some Hg complexes. The couplings are very sensitive to the phosphorus substituents, hybridization, and oxidation state, and generally increase in magnitude with substituents' electronegativity and show an algebraic increase in reducing coupling constant from P(III) to P(V).

The measurements of  $^2J(^{31}\text{P}-\text{M}-^{31}\text{P})$  where the intervening atom is a metal has been widely investigated due to a special interest for its value in making stereochemical assignments. The large variations in values may be usefully discussed under these headings: (1) the complex geometry; (2) the nature of the metal; and (3) the nature of the substituents on phosphorus. It has been found that coupling between phosphorus atoms in mutually *trans* relationship are large and positive, whereas *cis* phosphorus lead to small and often negative values. There is a general tendency for  $^2J(^{31}\text{P}-\text{M}-^{31}\text{P})$  to become more positive

as one descends a transition metal triad, and finally the  $^2J(^{31}\text{P}-\text{M}-^{31}\text{P})$  generally increases in magnitude with increasing electronegativity of substituents on P.

The Karplus angle relationship works quite well for organometallic phosphine complexes. For example in the compound  $[\text{Ir}(\text{PMe}_3)_3\text{HCl}(\text{C}_5\text{H}_5\text{N})]$ , the  $^2J(^{31}\text{P}, ^1\text{H})$  is 153.5 Hz for the  $\text{PMe}_3$  *trans* to the hydride, but only 19.8 Hz to the (chemically equivalent) *cis* phosphines.

## Arsenic, Antimony, and Bismuth

All the elements of group 15 of the periodic table possess at least one isotope suitable for investigation by NMR spectroscopy. The magnetically active isotopes of arsenic, antimony, and bismuth ( $^{75}\text{As}$ ,  $I = 3/2$ ;  $^{121}\text{Sb}$ ,  $I = 5/2$ ;  $^{123}\text{Sb}$ ,  $I = 7/2$ ;  $^{209}\text{Bi}$ ,  $I = 9/2$ ) have high abundance and relatively high sensitivity; however, they have been scarcely employed in multinuclear NMR spectroscopy owing to their high spin quantum numbers, which produce large quadrupolar moments. The Group 15 nuclides can be studied by NMR spectroscopy only when the compounds containing these elements are highly symmetric in the oxidation state +5. These nuclides are spectroscopically silent in nonsymmetric compounds or in the +3 oxidation state. The chemical shift standards (generally used at room temperature) are  $0.1 \text{ mol l}^{-1}$   $[(\text{Et}_4\text{N})^+(\text{AsF}_6)^-]$  in  $\text{CH}_3\text{CN}$  for  $^{75}\text{As}$ ,  $0.3 \text{ mol l}^{-1}$   $[(\text{Et}_4\text{N})^+(\text{SbCl}_6)^-]$  in  $\text{CH}_3\text{CN}$  for  $^{121,123}\text{Sb}$ , and a saturated solution of  $[(\text{Me}_4\text{N})^+(\text{BiF}_6)^-]$  in  $\text{CH}_3\text{CN}$  for  $^{209}\text{Bi}$ .  $^{75}\text{As}$ ,  $^{121,123}\text{Sb}$ , and  $^{209}\text{Bi}$  chemical shift ranges are very large (selected values are reported in Table 11). A positive sign indicates a chemical shift

to higher frequency (lower shielding) with respect to the reference compound. In the case of arsenic (+5) derivatives, the chemical shift range spans  $\sim 700$  ppm, with the tetrahedral oxoanion  $(\text{AsO}_4)^{3-}$  and the cation  $(\text{AsH}_4)^+$  occupying the deshielded and the shielded ends of the range, respectively. Even if there seems to be no regularity among the shifts of the  $\text{AsR}_4^+$  compounds, the ordering in terms of alkyl substituent is consistent with that found for group 14 alkyls, labeling the atoms of the structural unit  $[\text{As}(\text{alkyl})_4]^+$  as  $\text{As}_\alpha-\text{C}_\beta-\text{C}_\gamma-\text{C}_\delta$ , methyl substitution for one hydrogen at the  $\beta$  position ( $\Delta\beta$ ) deshields by 11 ppm,  $\Delta\gamma$  shields by 5 ppm, and  $\Delta\delta$  shields by 1 ppm. This trend in shielding effect as one proceeds down the chain is characteristic of methyl substitution. The antimony chemical shift range spans  $\sim 3500$  ppm. To date  $(\text{SbS}_4)^{3-}$  and  $\text{SbBr}_6^-$  occupy the deshielded ends of the range. In  $(\text{SbF}_6)^-$  and  $(\text{SbCl}_6)^-$  the small difference between the antimony shifts ( $\sim 87$  ppm) compared with those for analogous  $^{31}\text{P}$  and  $^{75}\text{As}$  species (152 and 392 ppm, respectively) is probably due to the electronic excitation energy that in both antimony, and, more likely, arsenic is responsible for the anomalous shielding term. In the case of the heaviest bismuth nuclide, only a few  $^{209}\text{Bi}$  NMR data are available yet.

In the case of  $^{75}\text{As}$ , only one  $J(^{75}\text{As}-^1\text{H})$  (555 Hz) value related to the  $(\text{AsH}_4)^+$  cation, has been reported. The  $\Delta\nu_{1/2}$  values for arsenic compounds, unlike  $^{75}\text{As}$  chemical shifts, are generally strongly dependent on the type of solvent employed, partly surely owing to ion pairing; in fact, the linewidths for  $(\text{AsEt}_4)^+\text{Br}^-$  are reported to be 168, 455, 670, and 700 Hz in water, acetonitrile, dimethylformamide, and dimethylsulfoxide, respectively.

**Table 11**  $^{75}\text{As}$  and  $^{121}\text{Sb}$  NMR data for selected group 15 compounds

Compound	Symmetry	Solvent	$\Delta$ (ppm)
$[\text{AsH}_4]^+$	$T_d$	HF	-291
$[\text{AsMe}_4]^+$	$T_d$	$\text{H}_2\text{O}$	206
$[\text{AsEt}_4]^+$	$T_d$	$\text{H}_2\text{O}$	249
$[\text{As}^n\text{Pr}_4]^+$	$T_d$	$\text{H}_2\text{O}$	230
$[\text{As}^n\text{Bu}_4]^+$	$T_d$	$\text{H}_2\text{O}$	234
$[\text{AsEt}_3\text{Me}]^+$	$T_d$	$\text{H}_2\text{O}$	242
$[\text{AsEt}_3^i\text{Pr}]^+$	$T_d$	$\text{H}_2\text{O}$	243
$[\text{AsEt}_3^t\text{Bu}]^+$	$T_d$	$\text{H}_2\text{O}$	245
$[\text{As}^n\text{Pr}_3\text{Me}]^+$	$T_d$	$\text{H}_2\text{O}$	225
$[\text{As}^n\text{Pr}_3\text{Et}]^+$	$T_d$	$\text{H}_2\text{O}$	234
$[\text{As}^n\text{Pr}_3^i\text{Bu}]^+$	$T_d$	$\text{H}_2\text{O}$	231
$[\text{As}^i\text{Pr}_3\text{Et}]^+$	$T_d$	$\text{H}_2\text{O}$	258
$[\text{AsPh}_4]^+$	$T_d$	$\text{H}_2\text{O}$	217
$[\text{SbS}_4]^{3-}$	$T_d$	$\text{H}_2\text{O}$	1031
$[\text{SbMe}_4]^+$	$T_d$	$\text{H}_2\text{O}$	780
$[\text{SbF}_6]^-$	$O_h$	$\text{CH}_3\text{CN}$	86.6
$[\text{Sb}(\text{OTeF}_5)_6]^-$	$O_h$	$\text{SO}_2\text{ClF}$	73.1

Pr, propyl; Bu, butyl.

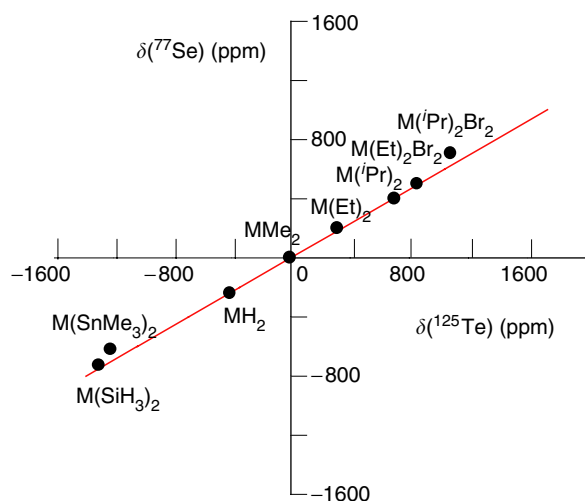
## Selenium and Tellurium

$^{77}\text{Se}$  and  $^{125}\text{Te}$  are of moderate and good NMR receptivity, respectively. Coupling constants to nuclei such as the  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{31}\text{P}$  are normally obtained from the positions of the satellites in the spectra of these nuclei. In Table 12 selected  $^{77}\text{Se}$  and  $^{125}\text{Te}$  chemical shifts relative to  $\text{Me}_2\text{Se}$  and  $\text{Me}_2\text{Te}$ , respectively, are reported. It has been generally found that variations up to  $\sim 10$  ppm with solvent and temperature can occur for  $^{77}\text{Se}$ . A study carried out on  $\text{Me}_2\text{Se}$  and  $\text{Me}_2\text{Te}$  in 30 different solvents indicates a degree of correlation between  $V_{\text{max}}$  and the solvent. The shifts in the series  $\text{SeR}_2$  increase in the order  $\text{R} = \text{Me} < \text{Et} < ^i\text{Pr} < ^t\text{Bu}$ .

Selenium and tellurium shielding generally runs closely parallel in equivalent compounds, plot of  $\delta(^{125}\text{Te})$  against  $\delta(^{77}\text{Se})$  being linear with a slope of  $\sim 1.8$ . An investigation on closely correlated

**Table 12** Selected  $\delta(^{77}\text{Se})$  and  $\delta(^{125}\text{Te})$  chemical shifts (ppm) for organoselenium and organotellurium compounds

Compound	$\delta(^{77}\text{Se})$	Notes	Compound	$\delta(^{125}\text{Te})$	Notes
$\text{Me}_2\text{Se}$	0	Liquid	$\text{Me}_2\text{Te}$	0	Liquid
$\text{MeSeH}$	-116	Liquid	$\text{Me}_3\text{Te}^+\text{I}^-$	443	In DMSO
$\text{MeSe}^-\text{Na}^+$	-332	In $\text{H}_2\text{O}$	$\text{Me}_2\text{TeCl}_2$	749	In $\text{CH}_2\text{Cl}_2$
$\text{Me}_3\text{Se}^+\text{I}^-$	253	In $\text{H}_2\text{O}$	$\text{Me}_2\text{TeCl}_2$	1218	In toluene
$\text{Me}_2\text{Se}_2$	275	Liquid	$\text{Me}_2\text{Te}_2$	688	In $\text{CH}_2\text{Cl}_2$
$\text{Me}_2\text{SeCl}_2$	448		$\text{MeTeCl}_3$	758	In benzene
$\text{MeSeCl}_3$	890		$\text{MeTeBr}_3$	647	In benzene
$\text{Me}_2\text{SeO}$	812	In $\text{H}_2\text{O}$	$\text{Et}_2\text{Te}$	380	In benzene
$\text{Et}_2\text{Se}$	233		$\text{Et}_3\text{Te}^+\text{Br}^-$	573	In $\text{D}_2\text{O}$
$\text{EtSe}^-\text{Na}^+$	-150	In $\text{H}_2\text{O}$	$\text{Ph}_2\text{Te}$	688	In $\text{CH}_2\text{Cl}_2$
$\text{PhSeH}$	152	Liquid	$\text{Ph}_2\text{Te}_2$	422	In $\text{CH}_2\text{Cl}_2$
$\text{Ph}_2\text{SeO}$	738		$\text{PhTeCl}_3$	1238	In DMSO
$\text{Ph}_2\text{Se}$	402	Liquid	$\text{Ph}_2\text{TeCl}_2$	981	In $\text{CH}_2\text{Cl}_2$
$\text{PhSeCH}_2\text{SePh}$	346		$\text{PhTeBr}_3$	1193	In DMSO/ $\text{CH}_2\text{Cl}_2$
$(\text{PhSe})_3\text{Sn}^-$	164		$\text{PhTeI}_3$	1101	In DMSO/ $\text{CH}_2\text{Cl}_2$
$(\text{PhSe})_3\text{As}$	266		Tellurophene	782	In acetone
Selenophene	605	Proposed ref	$(\text{Me}_3\text{Sn})_2\text{Te}$	-1214	In $\text{CH}_2\text{Cl}_2$
<i>cis</i> -( $\text{Me}_2\text{Se}$ ) $_2\text{PtCl}_2$	120		$\text{Fe}_3\text{Te}_2(\text{CO})_9$	1123	In $\text{CDCl}_3$ (nido cluster)
<i>trans</i> -( $\text{Me}_2\text{Se}$ ) $_2\text{PtCl}_2$	135		$\text{Fe}_3\text{Te}_2(\text{CO})_6$	-753	In $\text{CDCl}_3$ (Arachno cluster)

**Figure 20** Correlation between chemical shifts of analogous selenium and tellurium compounds.

compounds  $\text{R}^1\text{R}^2\text{E}$  ( $\text{E} = \text{Se}$  or  $\text{Te}$ ,  $\text{R}^1, \text{R}^2 = \text{unsaturated groups}$ ) reported  $\delta(^{125}\text{Te}) = 1.61 \delta(^{77}\text{Se})$  (Figure 20). It has been established that electron withdrawal from selenium or tellurium will lead to decreased shielding, as emerges from the chemical shifts of  $\text{MeSe}^-$ ,  $\text{Me}_2\text{Se}$ , and  $\text{Me}_3\text{Se}^+$  and from the low frequency shift of  $\text{R}_3\text{PE}$  ( $\text{E} = \text{Se}$  or  $\text{Te}$ ). It has been found that in the series  $\text{CF}_3\text{SeX}$   $\delta(^{77}\text{Se})$  correlates well with the electronegativity of  $\text{X}$  ( $\text{Cl}$ ,  $\text{Br}$ ,  $\text{CN}$ ,  $\text{H}$ ,  $\text{Ag}$ ). In a number of aromatic derivatives there are usually correlations between  $\delta(^{77}\text{Se})$  and  $\delta(^{125}\text{Te})$  with Hammett  $\sigma$  and Swain and Lupton substituent parameters. Also the role of stereochemistry in determining  $\text{Se}$  and  $\text{Te}$  chemical shifts has been

investigated. It has been reported that the  $\delta(^{77}\text{Se})$  differs by 2.1 ppm in the two diastereomers of the ester of 2-octanol and 2-phenylselenopropionic acid. On the other hand the difference for the meso and DL forms of  $[\text{PhMeP}(\text{Se})]_2\text{CH}_2$  is  $\sim 17.0$  ppm. In selenoethers, complex selenides, 2-substituted tellurophenes the usual dependence of  $^1J(^{77}\text{Se}, ^{13}\text{C})$  and  $^1J(^{125}\text{Te}, ^{13}\text{C})$  upon the hybridization and/or effective nuclear charge of the participating nuclei has been found.

## Manganese and Tungsten

The relatively high receptivity of  $^{55}\text{Mn}$  and its medium size quadrupole moment (that produces broad lines) make this nucleus a favorable one for NMR detection.  $^{55}\text{Mn}$  exhibits a shift range of  $\sim 3000$  ppm, with lines  $\sim 10$  Hz large in the case of symmetrical environment such as  $\text{MnO}_4^-$  and  $10000$  Hz in the case of unsymmetrical carbonyl-manganese complexes. There is a relationship between the oxidation state and chemical shielding as is noted for other metals. The reference compound is an  $\text{Mn}(\text{VII})$  species:  $\text{MnO}_4^-$ . The range of chemical shift for  $\text{Mn}^{\text{I}}$  species spans from  $-1000$  to  $-1500$ , whereas for  $\text{Mn}^{-\text{I}}$  complexes the range goes from  $-1500$  to  $-3000$  ppm. Only a limited number of diamagnetic organometallic Mn complexes that include isonitriles and carbonyl species have been investigated. The carbonyl-manganese complexes cover the whole  $^{55}\text{Mn}$  shift range with the hydride species generally at the low frequency margin (Table 13).

The first shift value for solid  $[\text{Mn}(\text{CO})_6]^+$  was  $-1500$  ppm. While in the vanadium, niobium, and



**Table 13**  $^{55}\text{Mn}$  NQR and NMR data on selected manganese carbonyl complexes

Compound	(NQCC) <sup>a</sup> (MHz)	$W_{1/2}$ <sup>b</sup> (Hz)	$\delta$ (ppm)
$[\text{Mn}_2(\text{CO})_{10}]$	3.05	83	– 2227 to – 2331
$[\text{MnCp}(\text{CO})_3]$	64.28	10 040	– 2225 to – 2280
$[\text{Na}[\text{Mn}(\text{CO})_5]]$	65.0	10 585	– 2780
$[\text{MnH}(\text{CO})_5]$	45.66	4340	– 2630/ – 2578
$[\text{MnCl}(\text{CO})_5]$	13.86	331	– 1005
$[\text{MnBr}(\text{CO})_5]$	17.46	688	– 1160
$[\text{MnI}(\text{CO})_5]$	19.85	1013	– 1485
$[\text{Mn}(\text{SnPh}_3)(\text{CO})_5]$	18.33	3200/2610	– 2510/ – 2610
$[\text{Mn}(\text{CO})_5\text{Re}(\text{CO})_5]$	8.67		

<sup>a</sup>Nuclear quadrupole coupling constants derived from single crystal studies.

<sup>b</sup>Linewidths and chemical shifts are given for solution of the respective complexes.

molybdenum hexacarbonyls replacement of CO results in a decrease of metal shielding, the  $[\text{Mn}(\text{CO})_6]^+$  derivatives may exhibit both substantial increase and decrease of  $^{55}\text{Mn}$  shielding. The complexes  $[\text{Mn}(\text{CO})_{4-n}(\text{PR}_3)_n(\text{S,P-donor})]$  ( $\text{R} = \text{OMe}$ ,  $\text{Ph}$ ,  $\text{Me}$ ,  $\text{Et}$ ,  $\text{Ph}$ ,  $n = 1$  or  $2$ ) exhibit deshielding of the  $^{55}\text{Mn}$  nucleus with an increasing number of phosphines coordinated to Mn. Phosphines are efficient  $\pi$  acceptors and then influence the paramagnetic deshielding term via the  $\pi$ -interacting  $e$  level ( $d_{xz}$ ,  $d_{yz}$ ) of the Mn-3d set under  $C_{4v}$  symmetry. If the ligand–Mn interaction does not allow for much  $\pi$  contribution,  $\sigma$  influences upon the HOMO–LUMO separation may become predominant for variations of  $\delta$  as in  $[\text{MnH}(\text{CO})_{5-n}(\text{PF}_3)_n]$ ,  $[\text{MnX}(\text{CO})_5]$ , and  $[\text{MnSnR}_3(\text{CO})_5]$ .

The diamagnetic and paramagnetic terms for  $[\text{MnX}(\text{CO})_5]$  ( $\text{X} = \text{halide}$ ,  $\text{H}$ ,  $\text{CN}$ ,  $\text{Me}$ ),  $[\text{Mn}(\text{CO})_6]^+$ , and  $[\text{MnO}_4]^-$  have been measured: a qualitative agreement is found between the calculated paramagnetic  $\sigma(\text{para})$  term and the observed shielding. It seems that the main contribution to  $\sigma(\text{para})$  arises from the  $d$ – $d$  excitation. Overall shielding increases with increasing  $\pi$ -donating ability, and via  $\delta_\sigma$ , with increasing hardness, the two partly counteracting effects giving rise to an ordering  $\text{H} < \text{CN} < \text{Me} < \text{Cl} < \text{Br} < \text{I}$ .

The only tungsten isotope with a nonzero value of  $I$  is  $^{183}\text{W}$ ,  $I = 1/2$  (natural abundance 14.4%), a nuclide with low receptivity relative to  $^1\text{H}$  ( $1.06 \times 10^{-5}$ ). This nucleus is radioactive, an  $\alpha$ -particle emitter with a long half-life ( $> 1 \times 10^{17}$  years). Satellite peaks from coupling with  $^{183}\text{W}$  are easily found in the  $^{31}\text{P}$  NMR spectra of tertiary phosphine complexes, in the  $^1\text{H}$  NMR spectra of hydrides, in the  $^{19}\text{F}$  NMR spectra of fluoro complexes, and finally in the  $^{13}\text{C}$  NMR spectra of exclusively carbonyl complexes. A solution of  $\text{Na}_2[\text{WO}_4]$  ( $1 \text{ mol l}^{-1}$  in  $\text{D}_2\text{O}$ , pD 9)

**Table 14**  $^{183}\text{W}$  chemical shifts for selected organotungsten compounds

Compound	$\delta_W$ (ppm)
$\text{W}(\text{CO})_6$	– 3505
$\text{W}(\text{CO})_5(\text{PMePh}_2)$	– 3324
$[\{\text{W}(\text{CO})_3(\text{Cp})\}_2]$	– 4040
$[\text{W}(\text{CO})_3(\text{Cp})\text{Cl}]$	– 2406
$[\text{W}(\text{CO})_3(\text{Cp})\text{Br}]$	– 2584
$[\text{W}(\text{CO})_3(\text{Cp})\text{I}]$	– 2996
$[\text{W}(\text{CO})_3(\text{Cp})\text{Me}]$	– 3549
$[\text{W}(\text{CO})_3(\text{Cp})\text{H}]$	– 4017
$[\text{W}(\text{Cp})_2\text{H}_2]$	– 4671

has been now accepted as the standard reference for  $^{183}\text{W}$  NMR. Relaxation times for  $^{183}\text{W}$  are generally long. Since a major relaxation mechanism is through chemical shift anisotropy (CSA), shorter, more favorable relaxation times are obtained when the environment about the tungsten nucleus is less symmetric. Addition of a paramagnetic substance, such as  $[\text{Cr}(\text{acac})_3]$  can be used to shorten relaxation times which are inconveniently long. The tungsten chemical shift is very sensitive to the environment about the nucleus. The overall range is large ( $\sim 6000$  ppm). Chemical shifts for selected organometallic compounds are reported in Table 14. Variations in chemical shifts for heavy atom are dominated by changes in  $\sigma_p$ , the paramagnetic contribution to total nuclear screening transition metal complex, for which a simplified expression is the following eqn [5]

$$\sigma_p \propto - (r^{-3}) \sum_j \sum_k (E_k - E_j)^{-1} C \quad [5]$$

where  $r$  is the average distance of valence d-electrons from the nucleus,  $E_k$  and  $E_j$  the energies of unoccupied ( $k$ ) and occupied molecular orbitals ( $j$ ), and  $C$  the sum containing the coefficients of the metals s, p, and d orbitals used in the summation to develop the various molecular orbitals.

Three classes of low-valent carbonyl tungsten compounds have been widely investigated:  $[\text{W}(\text{CO})_{6-n}\text{L}_n]$  (where L usually is a  $\text{PR}_3$ ),  $[\text{WCp}(\text{CO})_3\text{R}]$  ( $\text{R} = \text{halide}$ ,  $\text{H}$ ,  $\text{D}$ ,  $\text{alkyl}$ ,  $\text{SnMe}_3$ ),  $[\text{WCp}(\text{CO})_3]$ , and  $[\text{WCp}(\text{PR}_3)(\text{CO})_3]$ . Shielding decreases as CO is successively replaced by a weaker  $\pi$ -acceptor ligand. The  $^{183}\text{W}$  nucleus is more shielded in *trans*- than in *cis*- $[\text{W}(\text{CO})_4\text{L}_2]$  analogously to that found in  $[\text{Mn}(\text{CO})_4\text{L}_2]$  and  $[\text{Mo}(\text{CO})_4\text{L}_2]$ , but opposite trend to that in  $[\text{V}(\text{CO})_4\text{L}_2]^-$ . It will be noted that the tungsten nucleus in the series  $[\text{W}(\text{CO})_3(\text{Cp})\text{X}]$  becomes progressively more shielded as the halide is changed from chloride to bromide to iodide, which is the ‘normal halide dependence’ order for transition metal complexes. This is opposite to the

order expected from the effect of  $X^-$  on excitation energies.

## Cobalt, Rhodium, and Platinum

There is a scarcity of data and results to consider for palladium, osmium, iridium, iron, nickel, and ruthenium, the overall NMR behavior of these six metals being closely similar, although there are variations strictly connected to their chemical features. All group VIII transition metals exhibit a large range of the values of chemical shifts, over 18 000 ppm for  $^{59}\text{Co}$ . The following trends have been found: (1) species with the same donor atoms or with the same groups bonded to the metal are grouped together; (2) complexes of strong field donors resonate at low frequency; (3) donor atoms from lower rows of the periodic table cause lower frequency shifts; and (4) lower oxidation states give lower frequency shifts, but there is considerable overlap between the ranges covered by different oxidation states. The common reference compound for  $^{59}\text{Co}$  NMR is  $\text{K}_3[\text{Co}(\text{CN})_6]$  in  $\text{D}_2\text{O}$ . A large temperature dependence of  $^{59}\text{Co}$  chemical shift was found from the early observations. This behavior, associated with variation of  $\Delta E$ , has been firstly ascribed to the changes in the mean vibrational energies of the ground and excited states and then alternatively, owing to the anharmonicity of vibrations, to an increase of the mean M–L distance and so  $\Delta E^{-1}$  with increasing temperature. The resonance of  $^{59}\text{Co}$  has been proposed as NMR thermometers.

The  $^{59}\text{Co}$  chemical shifts of Co(III) complexes are clearly grouped together according to the identity of the donor atom. In Table 15 selected  $\delta(^{59}\text{Co})$  have been reported. In view of the large line-widths found, often large variations in shift values have been found. Small anisotropy has been found in the low-valent compounds  $[\text{Co}_2(\text{CO})_8]$  and  $[\text{Co}(\text{CO})_4\text{X}]$  where X is  $\text{Mn}(\text{CO})_5$  as also in  $[\text{Co}(\text{NH}_3)_5(\text{CN})]\text{Cl}_2$ ,  $[\text{Co}(\text{NH}_3)_4(\text{CO}_3)]\text{Br}$ , whereas

large anisotropy is in  $[\text{Co}(\text{Cp})_2]\text{NO}_3 \cdot \text{H}_2\text{O}$ . The resonances of  $[\text{Co}(\text{CN})_6]^{3-}$ , and  $[\text{Co}(\text{NO}_2)_6]^{3-}$  moves 22.3 and 46.4 ppm, respectively, to low frequency when a pressure of  $1000 \text{ kg cm}^{-2}$  was applied. The effect of the pressure seems to reduce the metal–ligand distance and then the value of  $\Delta E^{-1}$ .

$^{103}\text{Rh}$  and  $^{195}\text{Pt}$  chemical shifts (Table 16) are used in the investigation of the coordination mode adopted by ambidentate ligands. Replacement of the methyl group in  $[\text{PtCl}_2(\text{Me})\text{CO}]^-$  by larger alkyls causes little changes in  $\delta(^{195}\text{Pt})$ . Table 17 gives typical ranges for some coupling constants to group VIII nuclei. If the bonding terms in the usual equation for one-bond coupling constants are the same, then the coupling constants to two different metals should be in the ratio of their values of  $|\gamma|\psi_s(0)|^2$ .  $^1J(^{195}\text{Pt}, \text{X})$  is 16–22 times  $^1J(^{103}\text{Rh}, \text{X})$ . Relationships between structural features and  $^1J(^{195}\text{Pt}, ^1\text{H})$  and  $^1J(^{195}\text{Pt}, ^{13}\text{C})$

**Table 16**  $^{103}\text{Rh}$  and  $^{195}\text{Pt}$  chemical shifts for selected organometallic derivatives

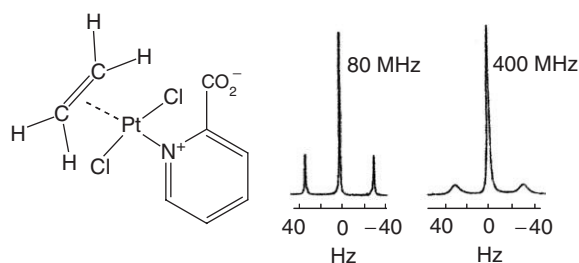
Compound	$\delta_M$ (ppm)
$[\text{Rh}(\text{CO})_4]^-$	– 644
$[\text{Rh}(\eta^5\text{-Cp}^*)\text{Cl}_2]_2$	2303
$[\text{Rh}_3\text{H}_3(\eta^5\text{-Cp}^*)_3\text{O}]^+$	– 136
$[\text{Rh}(\text{Cp})_2\text{Cl}]$	2736
$[\text{Rh}_4(\text{CO})_{12-m}\{\text{PR}_3\}_m]$ ( $m=1-4$ )	– 351 to – 792
$[\text{Rh}_5(\text{CO})_{15}]^-$ , eq, ax	– 149, – 949
$[\text{Rh}_6\text{H}(\text{CO})_{15}]^-$	– 374 to – 538
$[\text{Rh}(\text{acac})(\text{CO})_2]$	290
$[\text{Rh}(\text{acac})(\text{diene})]$	1760–1184
$[\text{Rh}(\text{N-donor})_2(\text{CO})_2]$	46, – 54
$[\text{RhCp}(\text{diene})]$	– 348 to – 1272
$[\text{Rh}(\text{Cp})(\mu\text{-polyene})]$	– 676 to – 1802
$[\text{Pt}(\text{CN})_6]^{2-}$	667 ( $\text{D}_2\text{O}$ ), 796 (acetone)
$[\text{PtMe}_3(\mu\text{-OH})]_4$	3062
$[\text{PtMe}_3(\mu\text{-X})]_4$ ( $\text{X} = \text{SR}, \text{N}_3, \text{Cl}, \text{Br}$ )	2552–2280
$[\text{PtMe}_3(\mu\text{-I})]_4$	1757
$\text{fac-}[\text{PtMe}_3\text{X}_3]^{n-}$ ( $\text{X} = \text{OH}_2, \text{py}, \text{NO}_2$ )	2534–2275
$[\text{PtMe}_2\text{X}_4]^{n-}$ ( $\text{X} = \text{OH}, \mu\text{-OH}, \text{OH}_2, \text{glycinate}$ )	4427–3820

**Table 17** Ranges of some spin–spin coupling constants for Rh and Pt organometallic compounds

$J$	Complex type/ligands	Range (Hz)
$^1J(^{103}\text{Rh}, ^1\text{H})$	Rh(III) hydrides	15–30
$^2J(^{103}\text{Rh}, ^1\text{H})$	Methyl derivatives	~ 2
$^1J(^{103}\text{Rh}, ^{13}\text{C})$	Olefin, allyl	5–16
$^1J(^{103}\text{Rh}, ^{13}\text{C})$	CO	55–85
$^1J(^{103}\text{Rh}, ^{13}\text{C})$	Alkyl, aryl	11–43
$^1J(^{195}\text{Pt}, ^1\text{H})$	Pt hydrides	700–1370
$^2J(^{195}\text{Pt}, ^1\text{H})$	Methyl derivatives	40–90
$^1J(^{195}\text{Pt}, ^{13}\text{C})$	Olefin, allyl	38–429
$^1J(^{195}\text{Pt}, ^{13}\text{C})$	CO, CNR	884–2459
$^1J(^{195}\text{Pt}, ^{13}\text{C})$	Alkyl, aryl, CN, CR	38–429
$^2J(^{195}\text{Pt}, ^{13}\text{C})$	Alkyl, aryl, vinyl, CR	0–362

**Table 15**  $^{59}\text{Co}$  chemical shifts for selected carbonyl, nitrosyl, and cyclopentadienyl derivatives of cobalt

Compound	$\delta_{\text{Co}}$ (ppm)
$[\text{CoL}_4]^-$ ( $\text{L} = \text{CO}, \text{PF}_3$ )	– 3090, – 4200
$[\text{CoHL}_4]$ ( $\text{L} = \text{CO}, \text{PF}_3$ )	– 3700, – 3900
$[\text{Co}_2(\text{CO})_8]$	– 2100
$[\text{Co}(\text{CO})_4\text{E}]$ ( $\text{E} = \text{metalloids}$ )	– 1800 to – 3270
$[\text{Co}_4(\text{CO})_{12}]$	– 300, – 1500
$[\text{CoCp}(\text{CO})_2]$	– 2670
$[\text{Co}(\text{CO})_3(\text{NO})]$	– 1322
$[\text{CoX}(\text{NO})_2]_2$ ( $\text{X} = \text{Cl}, \text{Br}, \text{or I}$ )	3640–2790
$[\text{CoCp}_2]^+$	– 2400



**Figure 21**  $^1\text{H}$  NMR spectrum of *trans*-[Pt(ethylene)(2-carboxypyridine) $\text{Cl}_2$ ] complex showing CSA broadening.

in CO,  $\text{CH}_3$ , and COD (cyclooctadiene) derivatives have been reported. It has been shown that  $^1J(^{195}\text{Pt}, ^{13}\text{C})$  in one isomer of  $[\text{PtCl}(\text{Ph})\text{CO}(\text{PR}_3)]$  increases with the temperature.

CSA relaxation on  $^{195}\text{Pt}$  can have unexpected influence on proton satellites. CSA relaxation increases with the square of the applied field. Olefinic  $^1\text{H}$  NMR signal of *trans*-Pt(ethylene)(2-carboxypyridine) $\text{Cl}_2$  at 80 and 400 MHz showed severe CSA broadening of the  $^{195}\text{Pt}$  satellites at the higher field (Figure 21).

Coupling to the  $I = 1/2$  Rh nucleus is often a powerful tool, particularly in Rh-phosphines and Rh-carbonyl complexes. In the  $^{13}\text{C}$  NMR spectrum of  $[\text{Rh}_2(\text{CO})_3(\text{CH}_2(\text{Cp}))]$ , the bridging carbonyl is observed at  $\delta$  232.53 and is a triplet with  $^1J(^{103}\text{Rh}, ^{13}\text{C}) = 46$  Hz. The equivalent terminal carbonyl occurs as a doublet at  $\delta$  190.18 with  $^1J(^{103}\text{Rh}, ^{13}\text{C}) = 84$  Hz.

## Cadmium and Mercury

Direct observations are relatively easy for both isotopes of cadmium with  $^{113}\text{Cd}$  having a small advantage on receptivity. In recent years, direct measurements of  $^{199}\text{Hg}$  using multinuclear FT systems have produced a wealth of data, some of which are summarized in Table 18. The most relevant problem with  $^{199}\text{Hg}$  observation is excessive line-width. It has been found that  $^{199}\text{Hg}$  resonances at 4.7 T are severely broadened by shielding anisotropy relaxation. Many researchers have used external neat dimethylmercury as a chemical shift reference or as a concentrated sample for the initial search for the  $^{199}\text{Hg}$  reference. In fact dimethylmercury has several intrinsic properties that are well suited for  $^{199}\text{Hg}$  experiments (rapid and accurate measurements, strong coupling with proton), but is very toxic ( $\text{LD}_{50}$  MeHg, 0.1 ml). For this reason mercury perchlorate has been also reported in literature as reference. At a concentration of  $0.1 \text{ mol l}^{-1}$  in a  $0.1 \text{ mol l}^{-1}$  perchloric acid solution the chemical shift is  $-2250$  ppm, relative to dimethylmercury.

**Table 18**  $^{113}\text{Cd}$  and  $^{199}\text{Hg}$  chemical shifts (solution) for selected organometallic complexes

Complex	$\delta_M$ (ppm)
$\text{Cd}(\text{alkyl})_2$	641–434
$\text{CdPh}_2$	329
$\text{CdMe}(\text{OR})$	382–258
$\text{CdMe}(\text{SR})$	610, 597
$[\text{Cd}(\text{CN})_4]^{2-}$	510
$\text{HgMe}_2$	2255 (neat liquid)
$\text{HgEt}_2$	1964 (neat liquid)
$\text{HgPh}_2$	1504 ( $\text{CHCl}_3$ )
$\text{Hg}(\text{CH}_2\text{COR})_2$	1485
$\text{Hg}(\text{Vi})_2$	1612 (neat)
$\text{Hg}(\text{C}_6\text{F}_5)_2$	1333 (acetone)
$\text{Hg}(\text{CF}_3)_2$	577 (THF)

Vi = vinyl.

Linear organomercurials provide a convenient basis for the investigation of the effect of substitution in the organic moiety on the metal chemical shift. Increase in the shielding of  $\sim 150$  ppm are observed for each substitution of Me for a H in  $\text{Hg}-\text{CH}_3$  derivatives, while smaller decreases in shielding ( $\sim 30$  ppm) occur on substitution of  $\beta$ -Hs. Cadmium alkyls show a similar behavior, but for silyl mercury compounds, replacements at both  $\alpha$  and  $\beta$  position results in deshielding.

The first ionization potential of  $[\text{HgCl}(\text{R})]$  which relates predominately to the  $\sigma$ -HgC bond shows a correlation with  $\delta(^{199}\text{Hg})$ . However, increasing the number of  $\beta$ -carbon atoms, which decreases the ionization potential, should increase the electron density on mercury and decrease  $\Delta E$ , both of which would cause deshielding, not shielding. The electron-withdrawing substituent generally increases the shielding. When the substituent effect is transmitted by  $-\text{CC}-$  in  $\text{Hg}(\text{CCR})_2$ , the shielding is roughly in agreement with inductive and mesomeric effects, the electron rich groups such as *tert*-butyl giving the most deshielded resonances.

Of the post-transition metals,  $^{199}\text{Hg}$  has been the subject of the majority of spin-spin coupling reports.

In Table 19 the typical ranges for some coupling constants involving Cd and Hg are reported. In alkylmercury compounds,  $^3J(^{199}\text{Hg}, ^{13}\text{C})$  shows a Karplus type of dependency on the dihedral angle. For a series of  $[\text{Hg}(\text{L})\text{CH}_2\text{C}(\text{OMe})\text{Me}_2]$  the variations in  $^1J(^{199}\text{Hg}, ^{13}\text{C})$  and  $^2J(^{199}\text{Hg}, ^1\text{H})$  closely follow the calculated mutual polarizabilities and it is suggested that changes in the 6s contribution to the Hg-C bond are the main cause of the changes in the magnitude of  $^1J(^{199}\text{Hg}, ^{13}\text{C})$ . The increase in the magnitude of  $^2J(^{199}\text{Hg}, ^1\text{H})$  and  $^1J(^{199}\text{Hg}, ^{13}\text{C})$  in organomercurials with the donor ability of the solvent is generally interpreted in terms of coordination. The same order is observed for both poor and

**Table 19** Ranges of some spin–spin coupling constants for Cd and Hg organometallic compounds

<i>J</i>	Complex type/ligands	Range (Hz)
$^2J(^{113}\text{Cd}, ^1\text{H})$	Alkyls	49–90
$^1J(^{113}\text{Cd}, ^{13}\text{C})$	Alkyls, CN	271–1060
$^3J(^{113}\text{Cd}, ^1\text{H})$	Alkyls	6–70
$^nJ(^{113}\text{Cd}, ^{13}\text{C})$	Alkyls	17–45
$^2J(^{199}\text{Hg}, ^1\text{H})$	Alkyls	94–265
$^3J(^{199}\text{Hg}, ^1\text{H})$	Organomercurials	38–311
$^1J(^{199}\text{Hg}, ^{13}\text{C})$	Alkyls, aryls, CN	264–3875
$^2J(^{199}\text{Hg}, ^{13}\text{C})$	Alkyls, aryls	24–920
$^3J(^{199}\text{Hg}, ^{13}\text{C})$	Alkyls, aryls	93–205
$^2J(^{199}\text{Hg}, ^{19}\text{F})$	Vinyls, alkyls	338–1911
$^1J(^{199}\text{Hg}, ^{31}\text{P})$	PR <sub>3</sub>	1980–12 970

intermediate mercury acceptors, and so far the latter does not correspond to the behavior of  $\delta(^{199}\text{Hg})$ . In bis-fluoroalkyl mercury derivatives  $^2J(^{199}\text{Hg}, ^{19}\text{F})$  decreases as the donor ability of the solvent increases.

In  $\text{CdMe}_2$ ,  $^2J(^{113}\text{Cd}^1\text{H})$  is positive, but  $^3J(^{113}\text{Cd}^1\text{H})$  has the opposite sign. Although  $^2J(^{199}\text{Hg}^1\text{H})$  is negative in alkyl derivatives of mercury it is positive in vinyl derivatives. Couplings to  $^{19}\text{F}$  in  $\text{Hg}-\text{CF}_3$  compounds have the opposite sign to those to  $^1\text{H}$  in the analogous  $\text{Hg}-\text{Me}$  system.

See also: **Nuclear Magnetic Resonance Spectroscopy:** Overview; Principles. **Nuclear Magnetic Resonance Spectroscopy-Applicable Elements:** Hydrogen Isotopes; Carbon-13; Fluorine-19; Nitrogen-15; Phosphorus-31. **Nuclear Magnetic Resonance Spectroscopy Applications:** Proton NMR in Biological Objects Subjected to Magic Angle Spinning. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Solid-State; Surface Coil.

## Further Reading

- Gielen M, Willem R, and Wrackmeyer B (eds.) (1997) *Advanced Applications of NMR to organometallic Chemistry*, Physical Organometallic Chemistry, vol. 1. New York: Wiley.
- Kidd RG (1971) NMR spectroscopy of organometallic chemistry. In: Tsutsui (ed.) *Characterization of Organometallic Compounds*, part II, ch. 8. New York: Wiley-Interscience.
- Mann BE (1974)  $^{13}\text{C}$  NMR chemical shifts and coupling constant of organometallic compounds. *Advances in Organometallic Chemistry* 12: 135–213.
- Mann BE (1974)  $^{13}\text{C}$  NMR spectroscopy of organo-transition metal complexes. In: Axenrod T and Webb GA (eds.) *NMR spectroscopy of Nuclei other than protons*, ch. 11: 153–156. New York: Wiley-Interscience.
- Mann BE (1988) Recent developments in NMR spectroscopy of organometallic compounds. *Advances in Organometallic Chemistry* 28: 397–457.
- Mann BE, et al. (1985) *Journal of the Chemical Society, Dalton Transactions*: 693.
- Orrell K (1999) Dynamic NMR spectroscopy in inorganic and organometallic chemistry. *Annual Reports on NMR Spectroscopy* 37: 1–74.
- Roc (1987) CD spectroscopic characterization of inorganic and organometallic complexes by metal and high-pressure NMR. In: Wayda AL and Darensbourg MV (eds.) *Experimental Organometallic Chemistry*, ACS Symposium Series No. 357, ch. 8. Washington, DC: American Chemical Society.
- Vrieze K and Van Leeuwen PWNM (1971) Studies of dynamic organometallic compounds of the transition metals by means of NMR. *Progress in Inorganic Chemistry* 14: 1–63. New York: Wiley Interscience.

# NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY APPLICATIONS

Contents

**Food**

**Forensic**

**Pharmaceutical**

**Proton NMR in Biological Objects Subjected to Magic Angle Spinning**

## Food

**C Deleanu**, National NMR Laboratory and Institute of Organic Chemistry, Bucharest, Romania

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Food materials exhibit a wide range of states of complexity. These range from homogeneous and relatively simple mixtures, as in edible oils, to quite complex heterogeneous systems, as in meat or bread (in the latter case both air and heterogeneous solid

**Table 19** Ranges of some spin–spin coupling constants for Cd and Hg organometallic compounds

$J$	Complex type/ligands	Range (Hz)
$^2J(^{113}\text{Cd}, ^1\text{H})$	Alkyls	49–90
$^1J(^{113}\text{Cd}, ^{13}\text{C})$	Alkyls, CN	271–1060
$^3J(^{113}\text{Cd}, ^1\text{H})$	Alkyls	6–70
$^nJ(^{113}\text{Cd}, ^{13}\text{C})$	Alkyls	17–45
$^2J(^{199}\text{Hg}, ^1\text{H})$	Alkyls	94–265
$^3J(^{199}\text{Hg}, ^1\text{H})$	Organomercurials	38–311
$^1J(^{199}\text{Hg}, ^{13}\text{C})$	Alkyls, aryls, CN	264–3875
$^2J(^{199}\text{Hg}, ^{13}\text{C})$	Alkyls, aryls	24–920
$^3J(^{199}\text{Hg}, ^{13}\text{C})$	Alkyls, aryls	93–205
$^2J(^{199}\text{Hg}, ^{19}\text{F})$	Vinyls, alkyls	338–1911
$^1J(^{199}\text{Hg}, ^{31}\text{P})$	$\text{PR}_3$	1980–12 970

intermediate mercury acceptors, and so far the latter does not correspond to the behavior of  $\delta(^{199}\text{Hg})$ . In bis-fluoroalkyl mercury derivatives  $^2J(^{199}\text{Hg}, ^{19}\text{F})$  decreases as the donor ability of the solvent increases.

In  $\text{CdMe}_2$ ,  $^2J(^{113}\text{Cd}, ^1\text{H})$  is positive, but  $^3J(^{113}\text{Cd}, ^1\text{H})$  has the opposite sign. Although  $^2J(^{199}\text{Hg}, ^1\text{H})$  is negative in alkyl derivatives of mercury it is positive in vinyl derivatives. Couplings to  $^{19}\text{F}$  in  $\text{Hg}-\text{CF}_3$  compounds have the opposite sign to those to  $^1\text{H}$  in the analogous  $\text{Hg}-\text{Me}$  system.

See also: **Nuclear Magnetic Resonance Spectroscopy:** Overview; Principles. **Nuclear Magnetic Resonance Spectroscopy-Applicable Elements:** Hydrogen Isotopes; Carbon-13; Fluorine-19; Nitrogen-15; Phosphorus-31. **Nuclear Magnetic Resonance Spectroscopy Applications:** Proton NMR in Biological Objects Subjected to Magic Angle Spinning. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Solid-State; Surface Coil.

## Further Reading

- Gielen M, Willem R, and Wrackmeyer B (eds.) (1997) *Advanced Applications of NMR to organometallic Chemistry*, Physical Organometallic Chemistry, vol. 1. New York: Wiley.
- Kidd RG (1971) NMR spectroscopy of organometallic chemistry. In: Tsutsui (ed.) *Characterization of Organometallic Compounds*, part II, ch. 8. New York: Wiley-Interscience.
- Mann BE (1974)  $^{13}\text{C}$  NMR chemical shifts and coupling constant of organometallic compounds. *Advances in Organometallic Chemistry* 12: 135–213.
- Mann BE (1974)  $^{13}\text{C}$  NMR spectroscopy of organo-transition metal complexes. In: Axenrod T and Webb GA (eds.) *NMR spectroscopy of Nuclei other than protons*, ch. 11: 153–156. New York: Wiley-Interscience.
- Mann BE (1988) Recent developments in NMR spectroscopy of organometallic compounds. *Advances in Organometallic Chemistry* 28: 397–457.
- Mann BE, et al. (1985) *Journal of the Chemical Society, Dalton Transactions*: 693.
- Orrell K (1999) Dynamic NMR spectroscopy in inorganic and organometallic chemistry. *Annual Reports on NMR Spectroscopy* 37: 1–74.
- Roc (1987) CD spectroscopic characterization of inorganic and organometallic complexes by metal and high-pressure NMR. In: Wayda AL and Darensbourg MV (eds.) *Experimental Organometallic Chemistry*, ACS Symposium Series No. 357, ch. 8. Washington, DC: American Chemical Society.
- Vrieze K and Van Leeuwen PWNM (1971) Studies of dynamic organometallic compounds of the transition metals by means of NMR. *Progress in Inorganic Chemistry* 14: 1–63. New York: Wiley Interscience.

# NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY APPLICATIONS

Contents

**Food**

**Forensic**

**Pharmaceutical**

**Proton NMR in Biological Objects Subjected to Magic Angle Spinning**

## Food

**C Deleanu**, National NMR Laboratory and Institute of Organic Chemistry, Bucharest, Romania

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Food materials exhibit a wide range of states of complexity. These range from homogeneous and relatively simple mixtures, as in edible oils, to quite complex heterogeneous systems, as in meat or bread (in the latter case both air and heterogeneous solid



phases are present). This range of physical states and chemical compositions requires a wide palette of nuclear magnetic resonance (NMR) techniques and approaches for obtaining in-depth information and characterization of various foodstuffs. NMR techniques are versatile and can provide different types of data and information on the same sample, depending on the experimental parameters selected. The applications include qualitative and quantitative analysis monitoring of processes or biochemical reactions determination of the structure of isolated compounds, the mobility and state of fat and water, the microstructure and aggregation state of various components in foodstuffs, as well as macroscopic imaging with all the advantages of obtaining cross sections without the need for cutting the sample. In terms of quantifying various compounds, NMR spectroscopy has the ability to provide the global concentration of the sample and not only the surface concentration. The disadvantages of the early days of NMR spectroscopy such as 'low sensitivity' or 'expensiveness' of the instrumentation do not hold true in the early twenty-first century. Thus, the modern hardware, software, automation, and hyphenation have reduced the experimental time to a few minutes even for two-dimensional (2D) experiments and increased the sensitivity, allowing easy detection of compounds at levels of micrograms and even nanograms. Moreover, there are arguments supporting the fact that sometimes well established techniques like gas chromatography (GC)–mass spectrometry (MS) produce false results, several of the 'new' compounds reported based on these techniques being in fact experimental artifacts resulting from the decomposition of some original labile compounds in foodstuffs. In such cases NMR has the great advantage of providing information on the studied compounds under ambient conditions of temperature and pressure.

## High Resolution (HR) NMR (Frequency Domain NMR)

Most HR NMR food sciences analyses are done in solution. However, the information provided by HR solid-state NMR is also very valuable for food sciences. Thus it is very likely that in the near future we will see a more balanced number of liquid- and solid-state NMR studies in food sciences.

### Elucidating the Structure of Isolated Compounds

HR NMR is currently the most powerful technique for elucidating the structure of isolated compounds in solution. There are numerous papers describing

the determination of the structure of various compounds isolated from all types of foodstuff. This application is an essential tool in advanced food research and requires prior isolation and purification of the compound under study.

As this approach is identical to that employed in synthetic chemistry, we would not cover it in this article. The interested reader will find extensive coverage of NMR techniques for structure determination in other articles of this encyclopedia.

The following sections will provide examples of applications of various NMR techniques to mixtures of compounds in various food matrices or fluids.

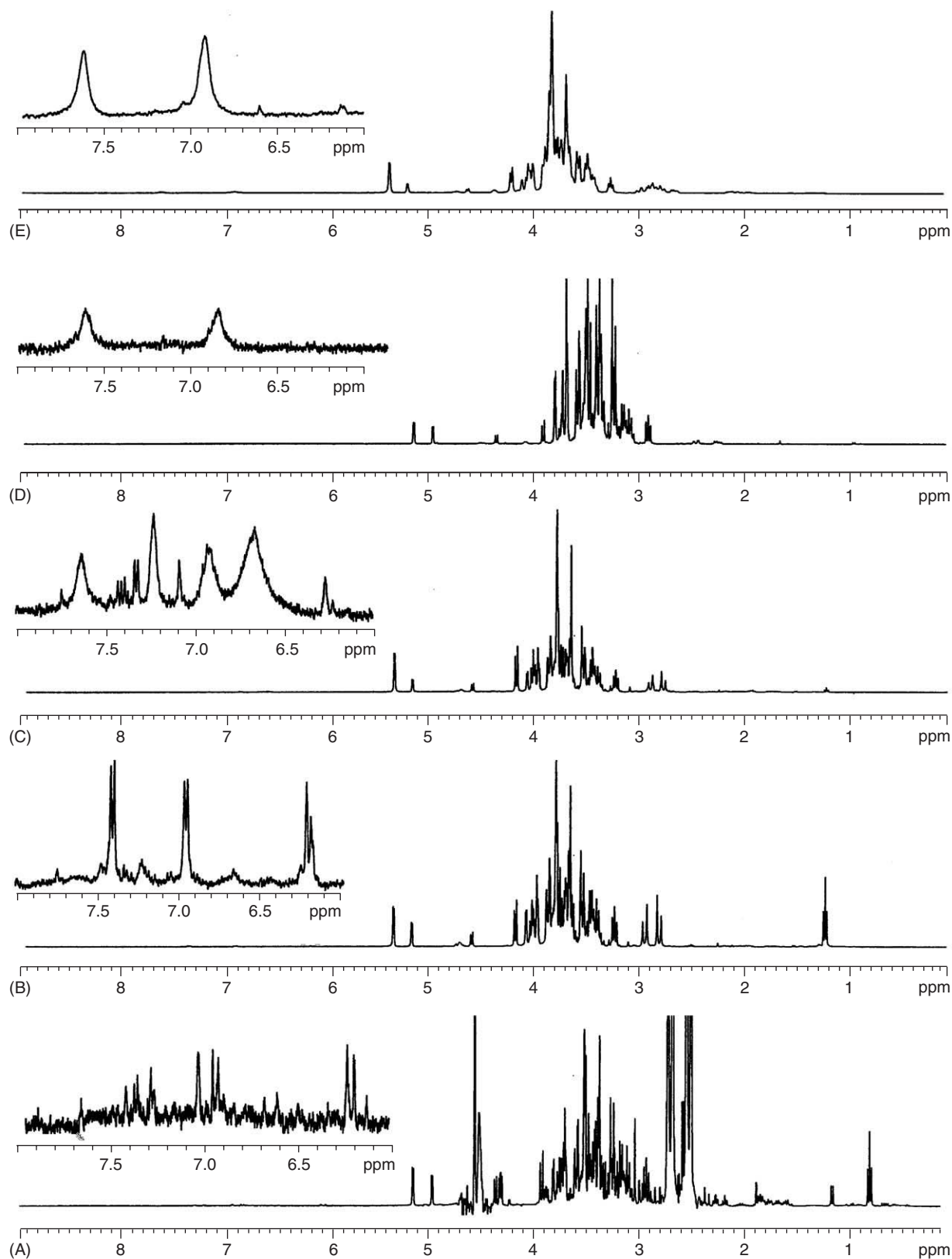
### Fruits and Fruit Juices

$^1\text{H}$ -NMR spectra for various fruit juices have been published, and signals have been assigned to sugars, amino acids, and other compounds like alcohols, acids, and polyphenols. Furthermore,  $^1\text{H}$ -NMR spectra have been used in combination with chemometric techniques for discriminating between apple varieties and for detecting adulteration of orange and apple juices. There are several non-NMR methods classically employed for determination of fruit juice adulteration (density, total amino acids, flavonoid and carotenoid composition). Some of these methods (e.g., liquid chromatography (LC)) require however standards of the suspected markers that are not always available. Moreover, with the classical methods one should preselect the most likely markers for a particular sample based on an educated guess on the expected type of adulteration. HR NMR has the advantage of offering a global profile. Thus, even without a total assignment of signals, in conjunction with chemometric techniques the method can discriminate between various patterns.

Figure 1 shows examples of  $^1\text{H}$ -NMR spectra for some common fresh fruit juices.

The typical profile of the  $^1\text{H}$ -NMR spectrum of fruits and fruit juices could be rationalized by considering three regions. Between 0.5 and 3 ppm there are various signals generated by amino acids, other carboxylic acids, and alcohols. Between 3 and 5.5 ppm the NMR spectrum is dominated by sugars as well as by the residual water signal (4.8 ppm). Although other compounds give rise to some signals in this region, in most cases these signals are of little practical use as they are overlapped by the sugar signals. The region between 5.5 and 10 ppm exhibits signals generated by aromatic and heterocyclic compounds, formic acid, and aldehydes.

Solid-state  $^{13}\text{C}$ -NMR has also been used to study fruits, for instance to follow changes in the cell wall polymers during ripening.



**Figure 1**  $^1\text{H}$ -NMR spectra (400 MHz) for some common fresh fruit juices. (A) lemon juice; (B) grapefruit juice; (C) orange juice; (D) apple juice; and (E) apricot pure.

## Vegetables

The assignment of many components in the  $^1\text{H}$ -NMR spectrum of tomato juice has been published recently. The  $^1\text{H}$ -NMR profile of tomato juice resembles that of fruit juices, with the same distinct zones, i.e., aminoacids, sacharides, and aromatic compounds.

$^1\text{H}$ -NMR spectra of methanolic extracts of control and genetically modified varieties of tomatoes have been analyzed using chemometric techniques. It was possible to assess variations in several metabolites.  $^1\text{H}$ ,  $^2\text{H}$ , and  $^{13}\text{C}$  solid-state magic angle spinning (MAS) NMR spectra have been recorded for tomato skin and potato tissue, providing information on the structure and dynamics of the cuticle polyesters.

It has been demonstrated that for the case of some flavor compounds present in garlic, NMR is superior to GC-MS as the latter produces false results due to the decomposition of some labile compounds.

Large differences in the ratio between proteins and polysaccharides in various mushroom strains have been observed using  $^{13}\text{C}$  solid-state MAS NMR spectroscopy.

## Wine

Surprisingly, in contrast with the early success of  $^2\text{H}$ -NMR spectroscopy in wine analysis (see Wine section of SNIF-NMR),  $^1\text{H}$ -NMR has not been used on whole wine samples until very recently. Although the possibility of quantifying methanol in wine was proved in the early 1990s, it was only a decade later that the field took off, with many minor compounds identified in whole wine samples and with chemometrics helping to rationalize the global profile of the  $^1\text{H}$ -NMR spectrum of wine samples. The same trend is paralleled with spectra for other alcoholic beverages like beer.

Recently an interesting application used a home-made spectrometer based on a 310 mm bore magnetic resonance imaging (MRI) instrument operating at 2 T to record  $^1\text{H}$ - and  $^{13}\text{C}$ -NMR spectra of intact unopened wine bottles. Thus, NMR was able to detect spoiled bottles with acetic acid developed in them without opening them. The application could be very valuable for old and expensive wines.

## Sugars (Carbohydrates)

There are many studies involving  $^1\text{H}$ - and  $^{13}\text{C}$ -NMR spectroscopy of carrageenans, a family of polysaccharides with the structure of linear sulfated galactans. They are extracted from certain species of red seaweed and have been used for a long time as natural texturing ingredients in the food industry.

The polysaccharide  $\beta$ -glucan has been proved to induce beneficial effects such as immunobiological,

hypcholesterolemic, and hypoglycemic effects.  $\beta$ -Glucan is isolated from yeast, and the established wet analytical methods are quite tedious.  $^1\text{H}$ -NMR has been proved to be a rapid alternative method.

Sugars have been determined using both liquid- and solid-state NMR in fruits, vegetables, and their juices, as well as in honey.

Starch products and glycogen have been studied extensively using  $^1\text{H}$ -,  $^{13}\text{C}$ -, and  $^{31}\text{P}$ -NMR spectroscopy. NMR spectroscopy has been used for determining the degree of polymerization, the average number of glucose units, the branching degree, and the anomer distribution of the reducing sugars. The interaction of starch with water and the mobility of polysaccharide chains have been investigated through one-dimensional (1D) and 2D solid-state NMR experiments. The mobility of 'freezable' and 'unfreezable' water in starch was studied using  $^1\text{H}$  and  $^2\text{H}$  solid-state NMR. The aging (retrogradation) of starch has been studied using cross polarization magic angle spinning (CPMAS)  $^{13}\text{C}$ -NMR.

Figure 2 presents the  $^1\text{H}$ -NMR spectrum of glucose. As always, in solution glucose is formed by a mixture of  $\alpha$ - and  $\beta$ -anomers. The spectrum emphasizes the region between 3.2 and 4.0 ppm, which is very crowded in the case of saccharides, glucose being only one of the saccharides present in various foodstuffs. Thus, many food products (including fruits, vegetables, honey, and wines) exhibit an NMR spectrum crowded with signals in the saccharide region.

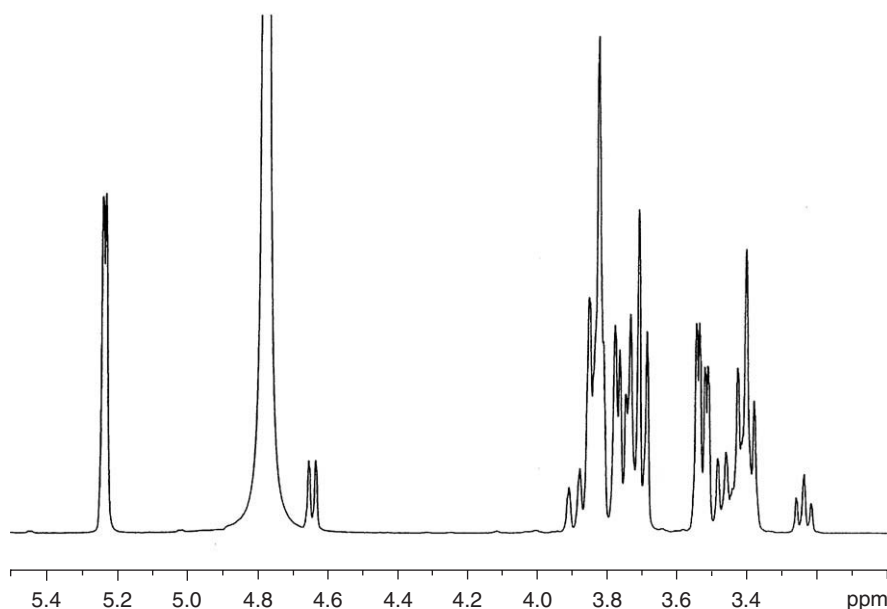
## Tea

Several flavonoids isolated from tea have been analyzed and their structure determined using NMR. There are several problems with the classical methods of analysis of flavonoids in tea. Due to the presence of complex mixtures of flavonoids in tea, they are often characterized as 'total polyphenols'. The colorimetric method for analysis of total phenols can interfere with other reducing compounds. LC can well resolve peaks for individual flavonoids; however, there are only a few standards available commercially, making the assignment of peaks uncertain in many cases. Thus, the structure of flavonoids giving rise to peaks in LC is often determined using various 1D- and 2D-NMR experiments.

## Coffee

The early NMR studies of coffee were limited to elucidating the structure of isolated compounds.

Several of the major constituents of espresso coffee have been identified in the  $^1\text{H}$ -NMR spectrum.  $^1\text{H}$ -NMR data in combination with chemometrics



**Figure 2** The  $^1\text{H}$ -NMR spectrum of a mixture of  $\alpha$ - and  $\beta$ -glucose in  $\text{D}_2\text{O}$  recorded at 400 MHz. The high peak at 4.8 ppm is the residual water signal.

have been recently used for verifying the authenticity of instant coffees from various producers.

### Milk and Dairy Products

The water-methanol extracts of cheese have been analyzed using  $^1\text{H}$ -NMR, and assignments of several amino acids were done using 2D spectra. Variations in the content of amino acids were observed as a function of ripening time and distance from the center of the cheese wheel. The fatty acid triglycerides from milk have been analyzed using both  $^1\text{H}$ - and  $^{13}\text{C}$ -NMR.

Liquid state  $^{31}\text{P}$ -NMR has been employed since a long time back for studying milk and casein. Solid-state  $^{31}\text{P}$ -NMR has been used for studying the nature of micelles and the interaction between compounds that form them.

### Cereals and Bread

Solution-state  $^{13}\text{C}$ -NMR and  $^{31}\text{P}$ -NMR have been employed in the study of lipids from corn suspension cells. Corn cells have also been grown in the presence of  $^{13}\text{C}$ -enriched acetate for following its incorporation into oleates and linoleates.

Solid-state CPMAS  $^{13}\text{C}$ -NMR has been used for investigating changes in maize and sorghum proteins on wet cooking and popping.

The hydration of proteins from wheat seeds has been studied using solid-state  $^{13}\text{C}$  CPMAS and  $^1\text{H}$

high resolution magic angle spinning (HRMAS) NMR spectroscopy.

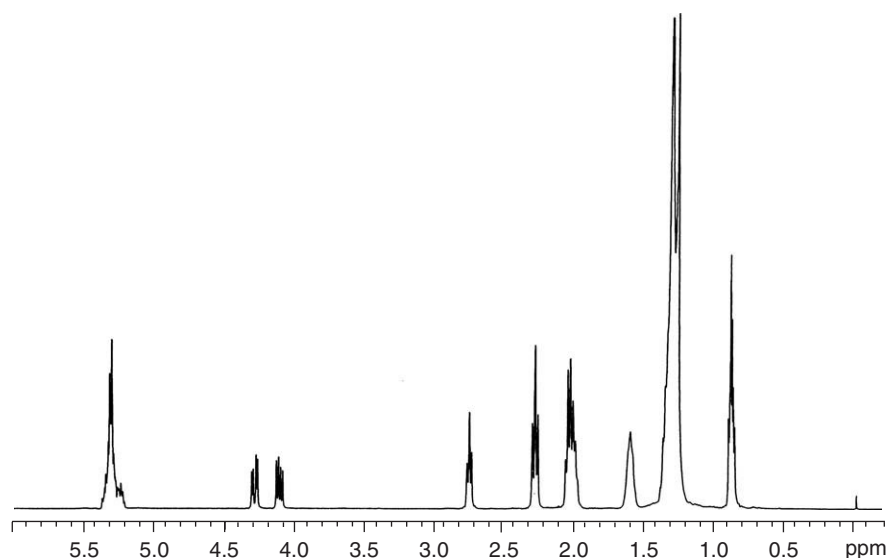
### Fats and Oils

Edible fats and oils are ideal candidates for HR NMR. Their good solubility in organic solvents makes the NMR analysis fast, easy, and accurate. Issues such as the composition of fatty acids in oils and fats, adulteration with cheaper oils (especially in the case of extra virgin olive oil), degradation of oils and fats in time and at elevated temperatures, and content of fatty acids in single intact seeds (for breeding purposes) have been studied using NMR. NMR method has some advantages over other methods. Thus, there is no need for sample manipulation before the analysis, information on the distribution of the fatty acid chains in the glycerol moiety (position 1/3 versus 2) is available, and the analysis has a relatively high speed (even for  $^{13}\text{C}$ -NMR, results being obtained within minutes).

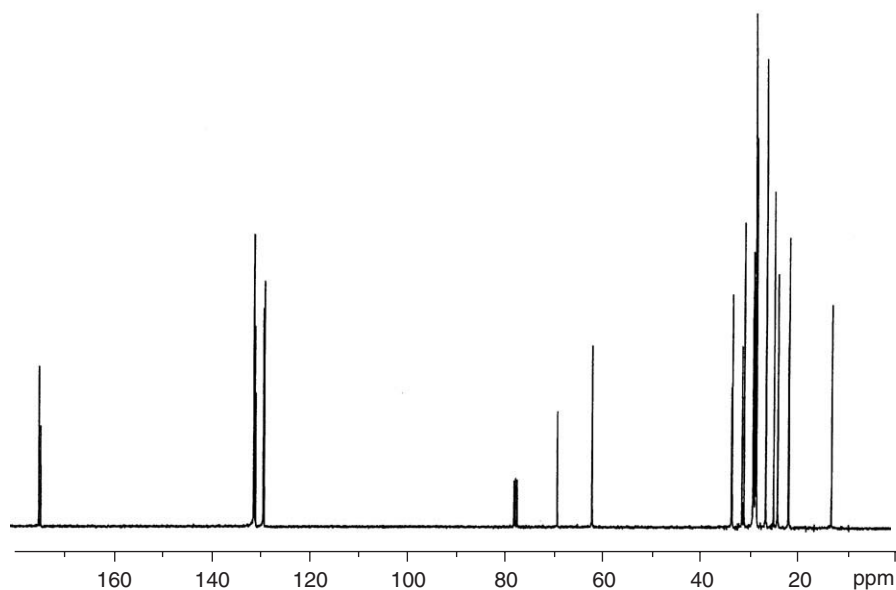
A typical  $^1\text{H}$ -NMR spectrum for sunflower oil is presented in **Figure 3**.

A typical  $^{13}\text{C}$ -NMR spectrum for the same sunflower oil is presented in **Figure 4**.

**Figure 5** presents details of the region 172–174 ppm in the  $^{13}\text{C}$ -NMR spectrum (100 MHz), illustrating the power of the method in assigning the distribution of fatty acids in the positions 1/3 and 2 of the glycerol moiety, based on the signals of the CO groups.



**Figure 3**  $^1\text{H}$ -NMR spectrum (400 MHz,  $\text{CDCl}_3$ ) of a sample of sunflower oil.



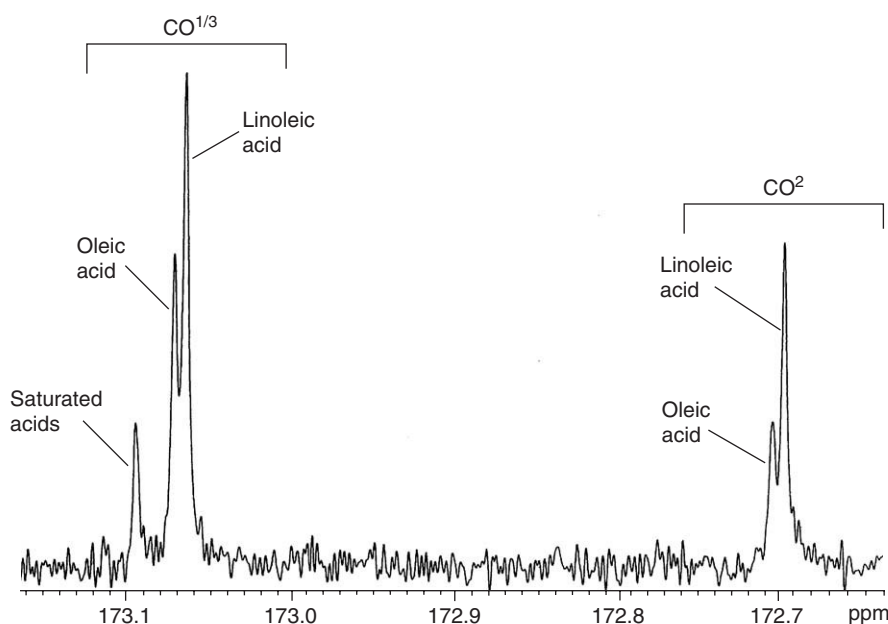
**Figure 4**  $^{13}\text{C}$ -NMR spectrum (100 MHz,  $\text{CDCl}_3$ ) of a sample of sunflower oil.

If for most food applications, recording the NMR spectra at the highest available magnetic field strength is a great advantage, for edible oils and fats there is not much improvement on passing from 300 to 600 MHz. In this case, the same types of groups from different long chain fatty acids produce similar signals in the  $^1\text{H}$ -NMR spectrum. This fact makes the routine medium-field (300–400 MHz) NMR spectrometers suitable and very competitive (in terms of price-to-quality ratio) for research and quality control of edible oils.

Individual fatty acids and esters have been used as model compounds for assignment of signals in the

NMR spectra of edible oils and fats. The concentration dependence on chemical shifts of model triglycerides has also been explored. Several papers deal with authentic edible oil samples, with reference to both assignment of NMR signals and authentication issues. As expected, many papers deal with NMR of olive oils due to the high market value of the virgin and extra virgin qualities. A lot of effort is put in authentication and determination of the origin of olive oils in connection with the Denomination of Protected Origin. Papers also deal with the identification of mono- and diglycerides in edible oils, as well as other compounds like phenolic derivatives.





**Figure 5** Enlargement of the carbonyl region in the  $^{13}\text{C}$ -NMR spectrum (100 MHz,  $\text{CDCl}_3$ ) of a sample of sunflower oil.

Polyunsaturated fatty acids (PUFAs), characteristic of fish lipids but also present in small quantities in vegetable oils, have been also investigated using NMR. Apart from the composition in fatty acids or minor compounds, the CH region in  $^1\text{H}$ -NMR, and the  $\text{CH}_2$  region in  $^{13}\text{C}$ -NMR have been used for determining the iodine value. Several studies deal with issues like the changes induced in the composition of oils by oxidation, heating, or hydrogenation.

### Seeds

$^{13}\text{C}$ -NMR spectra of intact seeds allow the identification and quantitation of the main fatty acids.

Solid-state  $^{13}\text{C}$  HRMAS and distortionless enhancement by polarization transfer (DEPT) MAS NMR also have been used for studying integral seeds. Thus, the three major components, fats, carbohydrates, and proteins (including the position of the fatty acid in glycerol) could be identified and quantified.

*In vivo*  $^{31}\text{P}$ -NMR has been used for studying the metabolites during the ripening and drying of seeds.

### Fish

Fish oils have been extensively investigated using  $^1\text{H}$ - and  $^{13}\text{C}$ -NMR. PUFAs, important in diet, are believed to reduce the risk of cardiovascular diseases. NMR analysis of PUFA is an interesting alternative to the widely used GC and GC-MS techniques when artifacts may arise either during transesterification or within the GC-MS equipment due to elevated

temperatures. PUFAs are markers for the fish species, age, freshness, storage, and processing conditions.

### Meat

$^{13}\text{C}$  solid-state HRMAS NMR studies have been carried out on cartilages.

$^{31}\text{P}$ -NMR spectroscopy studies of energy-related metabolites have been carried both in HR NMR instruments as biopsies and in MRI instruments as entire muscle pieces. The phosphorus metabolites are directly linked to meat quality. The influence of the stunning method on the postmortem energy metabolism and meat quality has been explored using  $^{31}\text{P}$ -NMR. The pH has been determined using  $^{31}\text{P}$ -NMR and associated with the meat quality. Although the pH is measured routinely through voltammetry, the NMR method has the advantage of measuring the pH in an inner part of the sample, being able to assess it both intra- and extracellularly, as well as *in vivo*.

### Eggs

The oxidation of cholesterol in egg powder has been detected using  $^1\text{H}$ -NMR.  $^{31}\text{P}$ -NMR spectroscopy has been used to quantitate phospholipids in egg lecithin.

### Essential Oils

Essential oils are important flavors and aromas. Many individual compounds have been isolated and characterized. Unseparated mixtures forming essential oils have also been analyzed using  $^{13}\text{C}$ -NMR.

### Agricultural Chemistry

It has been shown that even with medium-field NMR spectrometers (300 MHz), quantitative  $^1\text{H}$ - and  $^{31}\text{P}$ -NMR can rival with chromatography as an analytical method for agricultural chemicals. The advantage of the NMR is that it does not require a reference standard as is the case with chromatography. The possibility of detecting organophosphorus insecticide residues in crops using  $^{31}\text{P}$ -NMR has been demonstrated. Soil samples have been analyzed using  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{31}\text{P}$  solid-state and solution-state NMR experiments.

### Plant and Cell Metabolism and Other Biotransformation

The metabolism of plants has been followed extensively using various NMR techniques, employing  $^1\text{H}$ ,  $^{31}\text{P}$ , and  $^{13}\text{C}$  nuclei. All kinds of samples, cells, tissues, and extracts, both *in vivo* and *in vitro*, have been investigated. A special hardware design is used for following biotransformations *in vivo*. Thus the living material is suspended in a culture medium in the NMR tube and air is bubbled through a capillary tube to maintain life.

### $^2\text{H}$ -NMR (Site Specific Natural Isotope Fractionation Using NMR (SNIF-NMR))

$^2\text{H}$  SNIF-NMR is a high-resolution technique. However, the success of this technique in food sciences, together with the large number of published papers, justifies treating it under a separate section.

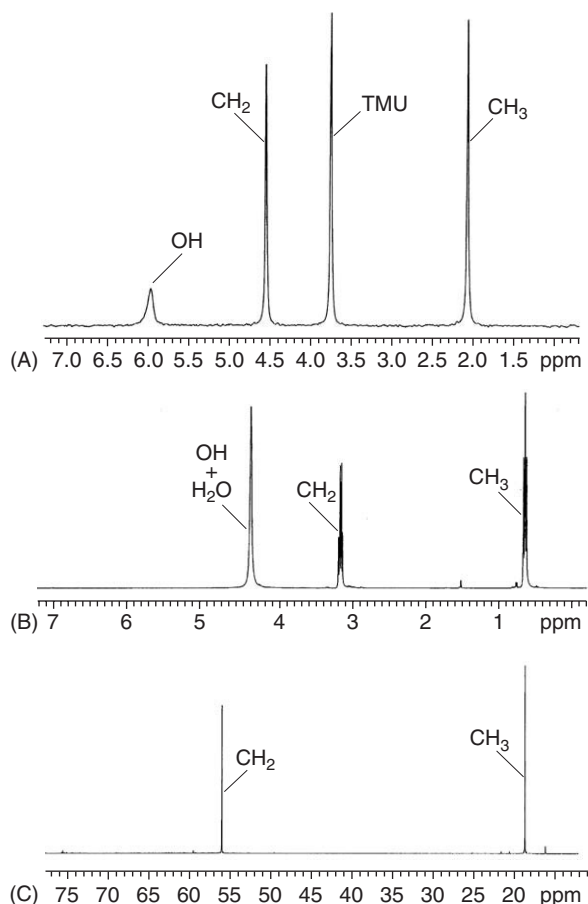
### Wine

Wine analysis using  $^2\text{H}$ -NMR is one of the major applications of NMR in food sciences, probably the most important application of HR NMR.

The basis for authentication of the geographic origin of wine and for spotting its adulteration is deuterium ( $^2\text{H}$ ) NMR spectroscopy. The method developed in the early 1980s by Gerard J. Martin is known as SNIF-NMR. The method was adopted by the French government and later as an official method by the EU and by the Office International de la Vigne et du Vin. SNIF-NMR was registered as a trademark by Eurofins Scientific (Nantes, France).

The principle of the method consists in comparing the ratios of the signals from the  $\text{CH}_3$  and  $\text{CH}_2$  groups of the ethanol molecule in the  $^2\text{H}$ -NMR spectrum. The method requires prior distillation of the ethanol from wine. The power of the method relies on the fact that the  $^2\text{H}/^1\text{H}$  ratio in various

positions of an organic molecule depends on factors such as the plant metabolism (thus bearing information on the type of sugar that by fermentation produces the ethanol) and the isotopic composition of the ground and rain water (thus bearing information on the geographic position). The isotopic ratio  $^2\text{H}/^1\text{H}$  in water varies on Earth between 90 and 160 ppm, depending on the latitude, with the highest values at the equator. It has been proven that deuterium in the  $\text{CH}_3$  site of the ethanol molecule comes from sugars, whereas deuterium in the  $\text{CH}_2$  site comes mainly from water. Thus, in the end, the isotopic ratio  $^2\text{H}/^1\text{H}$  in different sites of the distilled ethanol bears complex information related to the grape type, geographical origin, climate, production year, added sugar, or added water. In order to improve the power of the method, the isotopic composition in the two sites is no longer expressed as a relative ratio but as an absolute ratio, calibrated against a standard of known isotopic composition. The currently official standard is *N,N*-tetramethylurea (TMU). Typically fields of



**Figure 6** HR NMR spectra of ethanol recorded at 9.4 T. (A)  $^2\text{H}$ -NMR (61 MHz); (B)  $^1\text{H}$ -NMR (400 MHz); and (C)  $^{13}\text{C}$ -NMR (100 MHz).

400 or 500 MHz are used in routine analysis; however, owing to the good separation of deuterium signals, 300 MHz instruments can equally provide reliable results.

Figure 6 presents the  $^2\text{H}$ -NMR spectrum of ethanol, employed in the SNIF-NMR method, and for comparison the  $^1\text{H}$ - and  $^{13}\text{C}$ -NMR spectra.

SNIF-NMR is also applied to a whole range of alcoholic beverages, including: cognac, whisky, beer, and grape must.

### Fruit Juices

SNIF-NMR is already a well established technique for fruit juices, being suitable for tracing both their geographical origin and adulteration. The method was originally based exclusively on  $^2\text{H}/^1\text{H}$  analysis, but later its discrimination power was greatly increased by using it in combination with  $^{13}\text{C}/^{12}\text{C}$  analysis through isotopic ratio mass spectrometry (IRMS). In the case of fruit juices SNIF-NMR is mostly used for detecting adulteration with cheaper sugars, being accepted as an Association of Official Analytical Chemists (AOAC) method.

### Miscellaneous Applications of SNIF-NMR

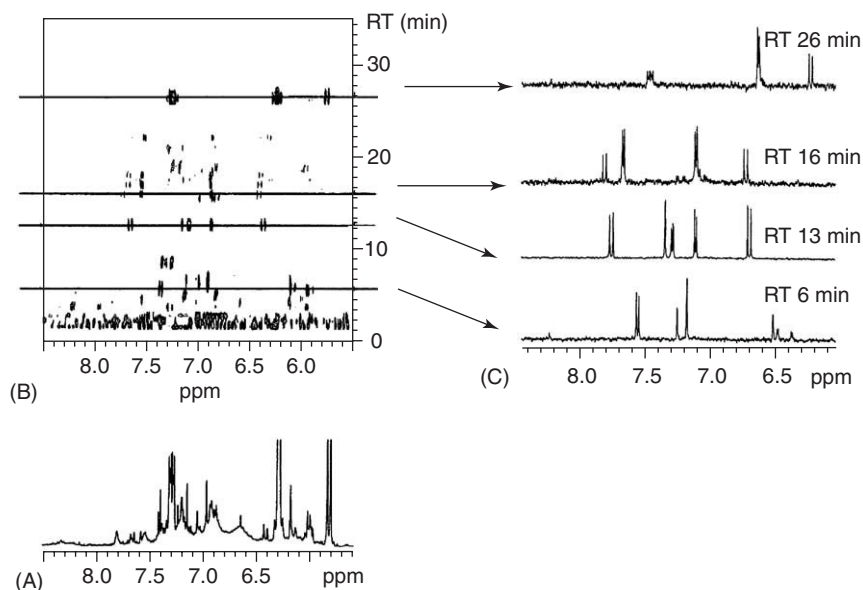
Apart from wine and juices the method has been successfully applied to issues such as the authentication of sugars of various origins, origin of glycerol (sugar fermentation versus edible oil transesterification),

natural versus synthetic origin of aromas, various essential oils, terpenoids, and others. There is a clear legislative discrimination between 'chemically identical' and 'natural' food aromas and flavors, the 'natural' ones being premium products. Owing to the great difference in price between such natural and synthetic aromas, the adulteration techniques became more and more sophisticated. Thus, vanilla adulteration techniques included even the addition of  $^{13}\text{C}$ - and  $^2\text{H}$ -enriched vanillin. Consequently, the analytical methods have had to be improved over the years, and presently combinations of  $^2\text{H}$  SNIF-NMR,  $^{13}\text{C}$  IRMS, and chemometrics applied to several marker compounds in the flavor can successfully cope with the most sophisticated types of adulteration used to date.

### Hyphenated Techniques

Emerging hyphenated techniques have already proved their analytical value and in the near future will certainly become common techniques in food sciences. Commercially available equipment include LC-NMR in both on-flow and stop-flow versions, LC-NMR-MS, LC-NMR-UV-MS, and more recently hyphenation with solid-phase extraction (LC-solid phase extraction-NMR).

Figure 7 illustrates the power of the on-flow liquid chromatography (LC)-NMR technique in the analysis of grape juice.



**Figure 7** On-flow LC-NMR spectrum of a sample of grape juice. (A) Detail of the aromatic region in the normal  $^1\text{H}$ -NMR spectrum; (B) on-flow LC-NMR spectrum; (C) rows at different retention times (RTs) correspond to NMR spectra of individual compounds in the mixture. (Reproduced with permission from Bruker BioSpin, Rheinstetten, Germany.)

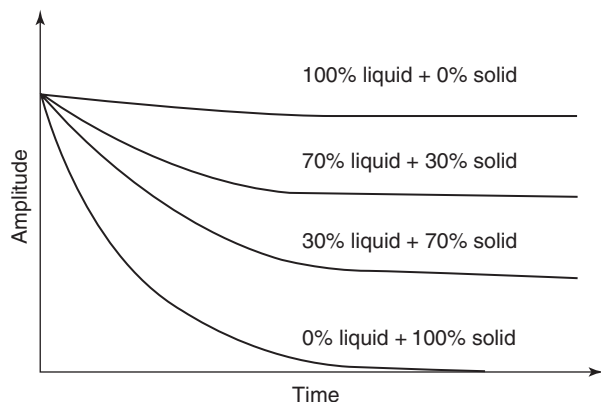
## Low-Resolution (LR) NMR (Time Domain NMR)

LR-NMR is the most commonly used NMR technique to date for quality control in food science and industry, with several official quality control methods in force. The success of the technique is due to several factors, including the power of the method (in terms of information and speed), the early application of the method to foodstuffs, the ecological appeal (no longer needing polluting chemicals), and the relatively low cost of the equipment, making it a very attractive alternative to the tedious wet chemical methods.

### Fats and Oils

Solid fat content (SFC) analysis is probably the most used LR-NMR application in the food industry. The initial success of the method prompted Unilever (manufacturer of margarines and related products) and Bruker (manufacturer of NMR instruments) to start a joint venture with the goal of building a tabletop LR-NMR spectrometer for solid-to-liquid ratio analysis in the fat industry. The method was developed in the early 1970s and over the years led to various quality control protocols for fat and oils that are by now adopted as official methods by various international and national organizations. The success story of SFC analysis opened wide the door for LR NMR methods to penetrate as routine techniques in the food industry.

As the relaxation time depends on the mobility of a particular system (with solids having the shortest relaxation times), it is easy to quantify mixtures using NMR. The bulk magnetization decay (free induction decay (FID)) for samples containing various solid-to-liquid ratios is presented in Figure 8.



**Figure 8** Bulk NMR signal amplitude for samples with various solid-to-liquid ratios. The signal decays exponentially with a constant  $T_2$  relaxation time.

The SFC is a critical parameter for the fats and oils industry. The official American Oil Chemists' Society (AOCS) wet method is dilatometry. Alternative wet methods are differential thermal analysis and differential scanning calorimetry (DSC). LR NMR was proved to be an alternative method for SFC determination in late 1950s. The early continuous wave LR NMR spectrometers rapidly found their way into the fats and oils industry, the method being accepted by the Instrumental Techniques Committee of the AOCS as early as in 1972. Presently the technical choice is radio frequency (RF) pulsed LR NMR. Pulse NMR spectrometers are more compact, very efficient, and relatively cheap. They have the advantage of exciting the protons in the whole sample at once.

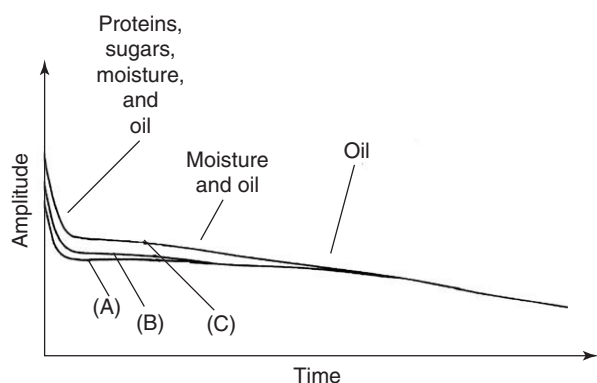
The determination of SFC has been extended to oil-in-water emulsions, where additional experimental parameters have to be considered in order to distinguish between oil and water protons. The droplet size of oil-in-water emulsions is usually measured using optical diffraction techniques, but it has also been determined through diffusion NMR experiments.

The degradation of fats and oils during frying has also been determined using LR NMR. During frying oils undergo complex chemical reactions such as oxidation, polymerization, hydrolysis, and isomerization. During these processes polar groups accumulate. The official International Union of Pure and Applied Chemistry and AOAC method for assessing the frying fat quality is the determination of total polar groups through preparative column chromatography. Based on the fact that the relaxation time is decreasing in direct relationship with the amount of polar groups present in the sample, it has been proved that LR NMR data correlate well with the values obtained using the official column chromatography method. The crystallization kinetics of fats is best studied using LR NMR, and as a result most of the published studies follow the crystallization using either LR NMR or DSC.

### Seeds

The wet method for determining fats in seeds is using extraction with organic solvents. LR NMR revolutionized the analysis of seeds, allowing rapid and nondestructive analysis without the need for weighing and drying them. Thus the seeds could be further used for selection in order to improve the genetic capacity of oil production.

The principle of separating and quantifying various components in a system based on the  $T_2$  relaxation time is the following. Various components have various relaxation times mainly due to differences in



**Figure 9** Effect of different drying methods on FID from sunflower seeds. The final moisture content is different when drying takes place at 100°C (A), 60°C (B), or room temperature (C).

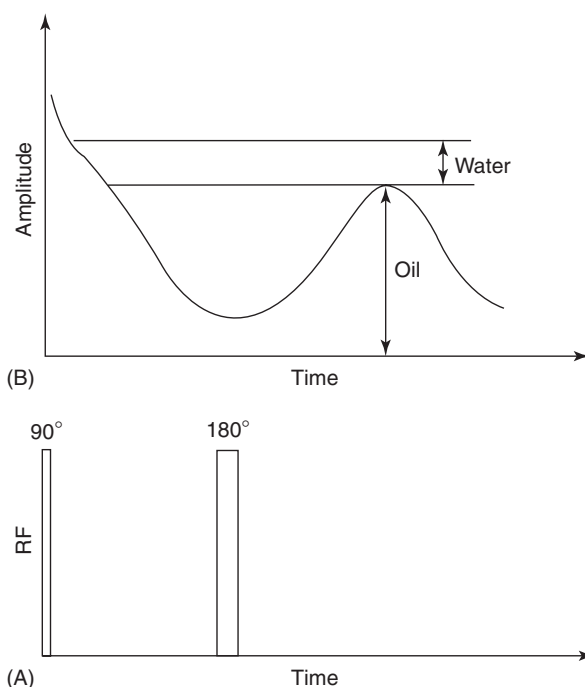
mobility (with solids having the shortest relaxation times). For instance, in seeds proteins and carbohydrates relax fast, and they mainly contribute to the first part of the NMR signal, the free induction decay (FID). Water has an intermediate relaxation time, whereas oil has the longest relaxation time. Thus, by carefully selecting the region of the FID where the component of interest is making the major contribution, one can quantify this component relative to other components. Meanwhile the intensity of the signal is proportional to the number of nuclei (protons) generating that signal. Thus by comparing the intensity of FID in the region of the component of interest for different samples or in comparison with a calibrated standard, one can obtain quantitative results. **Figure 9** illustrates how water in various seeds can be quantified with LR NMR when analyzing the middle part of the FID signal. Thus, by drying seeds at room temperature, the quantity of water that remains is higher (and the NMR signal's amplitude is higher) than by drying the same seeds at 60 or 100°C.

An almost identical curve was obtained on drying various seeds (peanut kernels, mustard, sunflower, and soybean) by sun and at 105°C.

When it is desired to quantify the slowly relaxing oil component, a more suitable pulse sequence scheme would be the spin-echo. As illustrated in **Figure 10**, the time at which the second refocusing pulse (of 180°) is applied can be selected so as to obtain an NMR signal generated exclusively by oil. The oil content can be accurately quantified when the amplitude of the signal is compared with a calibration curve generated by standard samples.

### Fruits

LR NMR has been used to determine the changes in water mobility in dehydrated fruits, and the sugar



**Figure 10** The spin-echo RF pulse sequence (A) and its result on the amplitude of the NMR signal (B) when applied to corn samples.

content in intact fruits. The technique is suitable for following the ripening process.

### Sugars (Carbohydrates) and Starch

LR NMR is a routinely used technique for determination of water content in sugars.

Starch hydration, water mobility, and the effect of chemical modification on molecular mobility of starch have also been studied using LR NMR.

Alteration of starch over time (retrogradation) has been studied using LR NMR. Most of the studies employed the  $T_2$  relaxation curves of  $^1\text{H}$ -NMR. Thus it was shown that the number of protons in the solid phase increases with aging, this being explained by the recrystallization of starch.

### Milk and Dairy Products

The state of water in milk, milk powder, casein, whey proteins, and cheese has been investigated using LR NMR.

### Chocolate

LR techniques are routinely used for determination of the fat content and solid-to-liquid ratio in chocolate and cocoa products. Analysis of the fluidification of cocoa butter using LR NMR enabled the identification of cocoas according to the process and the type of roaster used.



## Meat

LR NMR has been widely used in meat studies for quite a long time. There are several advantages associated with NMR, for instance the fact that NMR measures the whole volume of the sample, being less affected by surface effects. The total water and its state in meat (e.g., bounded versus free) can be estimated using LR  $^1\text{H}$ -NMR. This difference is evident on cooling when the free water is freezing, leading to a shortening of  $T_1$  (longitudinal relaxation time). The water holding capacity is an important parameter of meat that could be estimated using NMR. LR NMR has also been correlated with pH and cooking loss using chemometrics. LR NMR is currently a standard method for estimating the total fat content in meat. This way the meat needs not to be dried before the NMR analysis. Types of tissue in calf and cow have been differentiated using  $T_2$  LR NMR of water based on different types of collagen present in various types of tissue. Another important parameter of meat is the development of flavor associated with a long storage time and storage temperature. The so-called warmed-over flavor (WOF) is produced by autoxidation of membrane phospholipids and degradation of proteins and heteroatomic compounds. Some WOFs have been predicted using LR NMR and chemometrics.

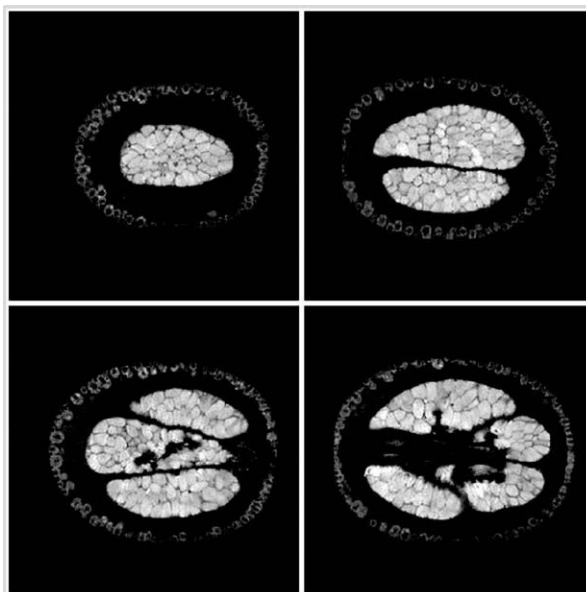
## Bread

LR NMR can monitor the distribution and mobility of water during bread making (dough, baking to bread, staling).

The baking process has been performed inside the NMR magnet and followed using the  $^1\text{H}$   $T_2$  relaxation time. The water mobility during storage of bread has been studied using the  $T_1$  and  $T_2$  relaxation curves. The texture parameters (elasticity and firmness) have correlated well with  $T_2$  curves of bread crumbs.

## Potatoes

Prediction of the sensory texture properties of cooked potatoes, based on raw potato analysis, is of major importance to the food industry. The sensory texture quality of potatoes has been predicted using  $T_2$  LR NMR. Correlation of the  $T_2$  LR NMR data with the chemical composition of potatoes has also been performed using chemometrics. Differentiation between potato varieties and determination of dry matter content has also been done using  $T_1$  and  $T_2$  LR NMR.



**Figure 11** MRI cross-sections through a kumquat fruit at 7 T. (Reproduced with permission from Bruker BioSpin, Rheinstetten, Germany.)

## NMR Imaging (Spatial Domain NMR)

The current MRI techniques provide information not only on the internal structure of various foodstuffs but also on the distribution of various compounds (chemical shift imaging), mobility and diffusion of various compounds (like water, oil, or compounds from various packing materials), or dynamic changes like crystallization.

The technique is increasingly being used in food sciences but still less than its true potential. The explanation for the delay in widespread use of NMR imaging techniques in food science and industry is the high cost of the equipment.

**Figure 11** shows four MRI cross-sections through a kumquat fruit. They start from near the surface and advance toward the center of the fruit. The fruit flesh in the centre is surrounded by the peel. The tissue between the fruit flesh and the peel is not visible because of a very short  $T_2$  relaxation time. Spherical structures in the peel are visible, and they contain the typical aromatic contributions of the fruit.

MRI has been applied to various foodstuffs such as meat, cereals, seeds, fruits, vegetables, cheese, or chocolate. The distribution and mobility of various constituents including water, sugars, and lipids have been monitored.

*See also:* **Carbohydrates:** Overview; Sugars – Spectrophotometric Methods. **Food and Nutritional Analysis:** Overview; Coffee, Cocoa, and Tea; Alcoholic Beverages;

Meat and Meat Products; Dairy Products; Vegetables and Legumes; Oils and Fats; Fruits and Fruit Products. **Lipids:** Overview; Fatty Acids. **Liquid Chromatography:** Liquid Chromatography–Nuclear Magnetic Resonance Spectrometry. **Mass Spectrometry:** Overview. **Nuclear Magnetic Resonance Spectroscopy:** Overview; Principles; Instrumentation. **Nuclear Magnetic Resonance Spectroscopy-Applicable Elements:** Hydrogen Isotopes; Carbon-13. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Solid-State; *In Vivo* Spectroscopy Using Localization Techniques.

## Further Reading

- Beauvallet C and Renou JP (1992) Applications of NMR spectroscopy in meat research. *Trends in Food Science & Technology* 3: 241–246.
- Belloque J and Ramos M (1999) Application of NMR spectroscopy to milk and dairy products. *Trends in Food Science & Technology* 10: 313–320.
- Belton PS, Delgadillo I, Gil AM, and Webb GA (eds.) (1994) *Magnetic Resonance in Food Science*. Cambridge: Royal Society of Chemistry.
- Belton PS, Delgadillo I, Holmes E, *et al.* (1996) Use of high-field  $^1\text{H}$ -NMR spectroscopy for the analysis of liquid foods. *Journal of Agricultural Food Chemistry* 44: 1483–1487.
- Belton PS, Hills B, and Webb GA (eds.) (1999) *Advances in Magnetic Resonance in Food Sciences*. Cambridge: The Royal Society of Chemistry.
- Colquhoun IJ and Lees M (1998) Nuclear magnetic resonance spectroscopy. In: Ashurst PR and Dennis MJ (eds.) *Analytical Methods of Food Authentication*, pp. 36–75. London: Blackie Academic and Professional.
- Deleanu C and Paré JRJ (1997) Nuclear magnetic resonance. Principles and applications. In: Paré JRJ and Bélanger JMR (eds.) *Instrumental Methods in Food Sciences*, pp. 179–237. Amsterdam: Elsevier.
- Guillen MD and Ruiz A (2001) High resolution  $^1\text{H}$  nuclear magnetic resonance in the study of edible oils and fats. *Trends in Food Science & Technology* 12: 328–338.
- Guillou C, Remaud G, and Martin GJ (1991) Application of deuterium NMR and isotopic analysis to the characterization of foods and beverages. *Trends in Food Science & Technology* 2: 85–89.
- McCarthy MJ (1994) *Magnetic Resonance Imaging in Foods*. New York: Chapman & Hall.
- Rutledge DN (2001) Characterization of water in agro-food products by time domain-NMR. *Food Control* 12: 437–445.
- Sacchi R, Addeo F, and Paolillo L (1997)  $^1\text{H}$  and  $^{13}\text{C}$  NMR of virgin olive oil. An overview. *Magnetic Resonance in Chemistry* 35: S133–S145.
- Tiwari PN and Gambhir PN (1995) Seed oil determination without weighing and drying the seeds by combined free induction decay and spin-echo nuclear magnetic resonance signals. *Journal of the American Oil Chemists' Society* 72: 1017–1020.
- Todt H, Burk W, Guthausen G, *et al.* (2001) Quality control with time-domain NMR. *European Journal of Lipid Science and Technology* 103: 835–840.
- Vlahov G (1999) Application of NMR to the study of olive oils. *Progress in Nuclear Magnetic Resonance Spectroscopy* 35: 341–357.

## Forensic

**B Dawson**, Health Canada, Ottawa, ON, Canada

Canadian Crown Copyright © 2005. Published by Elsevier Ltd. All rights reserved.

## Introduction

Since conclusions reached in forensic laboratories are used in the criminal justice system, they must leave no room for doubt. Thus, analytical methods must meet strict criteria. They must be extremely selective, reproducible, sufficiently sensitive, and suitable for qualitative and quantitative analysis. It is also highly desirable for the method to call for the minimum number of pretreatment steps and to be applicable to compound mixtures without preliminary separation of their components. Nuclear magnetic resonance (NMR) spectroscopy meets these criteria. It is well

known to be a powerful tool for the elucidation of chemical structures and the identification of organic compounds. It has been used in various types of forensic analysis for many years. It should be stressed that NMR is not used as a stand-alone technique in the forensic laboratory, but in conjunction with many others, including mass spectrometry, infrared (IR) spectroscopy, and various chromatographic techniques.

Some early reports on investigations using NMR spectroscopy for forensic analysis were pessimistic about its application as a routine method. NMR was viewed as an insensitive technique when compared with some of the other methods of analysis; the cost of the instrument was comparatively high and the results obtained required considerable expertise to interpret. This situation no longer exists. During the last several years, the amounts of material required

Meat and Meat Products; Dairy Products; Vegetables and Legumes; Oils and Fats; Fruits and Fruit Products. **Lipids:** Overview; Fatty Acids. **Liquid Chromatography:** Liquid Chromatography–Nuclear Magnetic Resonance Spectrometry. **Mass Spectrometry:** Overview. **Nuclear Magnetic Resonance Spectroscopy:** Overview; Principles; Instrumentation. **Nuclear Magnetic Resonance Spectroscopy-Applicable Elements:** Hydrogen Isotopes; Carbon-13. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Solid-State; *In Vivo* Spectroscopy Using Localization Techniques.

## Further Reading

- Beauvallet C and Renou JP (1992) Applications of NMR spectroscopy in meat research. *Trends in Food Science & Technology* 3: 241–246.
- Belloque J and Ramos M (1999) Application of NMR spectroscopy to milk and dairy products. *Trends in Food Science & Technology* 10: 313–320.
- Belton PS, Delgadillo I, Gil AM, and Webb GA (eds.) (1994) *Magnetic Resonance in Food Science*. Cambridge: Royal Society of Chemistry.
- Belton PS, Delgadillo I, Holmes E, *et al.* (1996) Use of high-field  $^1\text{H}$ -NMR spectroscopy for the analysis of liquid foods. *Journal of Agricultural Food Chemistry* 44: 1483–1487.
- Belton PS, Hills B, and Webb GA (eds.) (1999) *Advances in Magnetic Resonance in Food Sciences*. Cambridge: The Royal Society of Chemistry.
- Colquhoun IJ and Lees M (1998) Nuclear magnetic resonance spectroscopy. In: Ashurst PR and Dennis MJ (eds.) *Analytical Methods of Food Authentication*, pp. 36–75. London: Blackie Academic and Professional.
- Deleanu C and Paré JRJ (1997) Nuclear magnetic resonance. Principles and applications. In: Paré JRJ and Bélanger JMR (eds.) *Instrumental Methods in Food Sciences*, pp. 179–237. Amsterdam: Elsevier.
- Guillen MD and Ruiz A (2001) High resolution  $^1\text{H}$  nuclear magnetic resonance in the study of edible oils and fats. *Trends in Food Science & Technology* 12: 328–338.
- Guillou C, Remaud G, and Martin GJ (1991) Application of deuterium NMR and isotopic analysis to the characterization of foods and beverages. *Trends in Food Science & Technology* 2: 85–89.
- McCarthy MJ (1994) *Magnetic Resonance Imaging in Foods*. New York: Chapman & Hall.
- Rutledge DN (2001) Characterization of water in agro-food products by time domain-NMR. *Food Control* 12: 437–445.
- Sacchi R, Addeo F, and Paolillo L (1997)  $^1\text{H}$  and  $^{13}\text{C}$  NMR of virgin olive oil. An overview. *Magnetic Resonance in Chemistry* 35: S133–S145.
- Tiwari PN and Gambhir PN (1995) Seed oil determination without weighing and drying the seeds by combined free induction decay and spin-echo nuclear magnetic resonance signals. *Journal of the American Oil Chemists' Society* 72: 1017–1020.
- Todt H, Burk W, Guthausen G, *et al.* (2001) Quality control with time-domain NMR. *European Journal of Lipid Science and Technology* 103: 835–840.
- Vlahov G (1999) Application of NMR to the study of olive oils. *Progress in Nuclear Magnetic Resonance Spectroscopy* 35: 341–357.

## Forensic

**B Dawson**, Health Canada, Ottawa, ON, Canada

Canadian Crown Copyright © 2005. Published by Elsevier Ltd. All rights reserved.

## Introduction

Since conclusions reached in forensic laboratories are used in the criminal justice system, they must leave no room for doubt. Thus, analytical methods must meet strict criteria. They must be extremely selective, reproducible, sufficiently sensitive, and suitable for qualitative and quantitative analysis. It is also highly desirable for the method to call for the minimum number of pretreatment steps and to be applicable to compound mixtures without preliminary separation of their components. Nuclear magnetic resonance (NMR) spectroscopy meets these criteria. It is well

known to be a powerful tool for the elucidation of chemical structures and the identification of organic compounds. It has been used in various types of forensic analysis for many years. It should be stressed that NMR is not used as a stand-alone technique in the forensic laboratory, but in conjunction with many others, including mass spectrometry, infrared (IR) spectroscopy, and various chromatographic techniques.

Some early reports on investigations using NMR spectroscopy for forensic analysis were pessimistic about its application as a routine method. NMR was viewed as an insensitive technique when compared with some of the other methods of analysis; the cost of the instrument was comparatively high and the results obtained required considerable expertise to interpret. This situation no longer exists. During the last several years, the amounts of material required

for NMR studies have dropped drastically so that most experiments on state-of-the-art equipment require a few tens of micrograms of small molecule organic sample. The advances in NMR technology which gave rise to these huge increases in sensitivity have been driven by the desire to determine the structures of larger molecules, such as proteins. These advances include the development of higher-field magnets (currently up to 900 MHz) and improved probe technology (including the new cryoprobes that give very large sensitivity enhancement by eliminating much of the electrical noise present in normal probes). The costs of the modern mid-field strength instruments (400–600 MHz), which are more than adequate for essentially all forensic work, are comparable to those of other equipment such as mass spectrometers. Most modern spectrometers can be purchased with software which makes the acquisition of data completely automatic. The amount of training required by the forensic analyst for most routine experiments is minimal.

## Sample Types

Because of the diverse areas of investigation carried out by forensic analysts, the types of samples vary greatly. They can be in any form – solid, liquid, or gas. Samples can originate from any source and may be organic, inorganic, natural, or synthetic. They may be present in large quantities or as a minute residue. Illicit drug seizures usually consist of powders, plant materials (or their extracts), or occasionally dosage forms. Sometimes it is only necessary to identify the major constituent of a sample, while for other samples the minor components may be important. The information obtained in the analysis may be used, not only to support a case, but also to provide valuable intelligence information.

Some applications for which NMR has been used in forensic analysis include the identification of compounds by comparison of spectra with those of authentic materials, the determination of structures of unknown compounds such as ‘designer’ drugs, the identification of impurities in illicit drugs, the determination of optical purity of drugs, and the characterization of explosives, accelerants, fire residues, and various body fluids or tissue extracts. This article will briefly review each of these applications.

## Sample Preparation

The procedure used to prepare a forensic sample for NMR analysis depends on its nature and the information required. Because of the wide variety of types

of samples, a number of different procedures are used. In the case of an exhibit that is a ‘pure’ compound, or for which data are required for the total sample, the analyst need only select a suitable deuterated solvent and dissolve an appropriate amount of material for acquisition of required spectra. For any other kind of samples, decisions concerning the information required from the material will determine the method of sample preparation.

The general methodology used in the isolation and determination of illicit drugs from component mixtures depends on the molecular structure of these compounds. It is often possible to select a solvent that effectively dissolves the drug of interest but does not dissolve the other components to any significant degree. Thus, spectral interference from the excipients is eliminated. For example, deuteriochloroform may be used to extract many drugs from formulations in which the excipients consist of sugars, since the sugars will remain in the aqueous phase. When this simple procedure is not applicable, others, which take advantage of the characteristics of the drugs, must be used. Drugs may be divided into four general categories (acidic, basic, amphoteric, and neutral), which have different partition coefficients between immiscible organic and aqueous phases. This occurs because the hydrogen ion concentration (a function of pH) in the aqueous phase often affects the ionic state or polarity of molecules, depending on the functional groups present. Adjustment of the aqueous phase pH allows the separation of drugs from the four different categories. For example, acidic and neutral drugs would be extracted into the organic solvent if the aqueous solution is made acid, while basic and neutral drugs would be extracted from aqueous base. For amphoteric drugs, such as morphine or phenylephrine, the pH must be carefully controlled to be at their respective isoelectric points because the drugs exhibit both acidic and alkaline properties. At the isoelectric point, the molecules exist predominantly in the nonpolar or nonionized form. Neutral drugs may be isolated from the others by extracting an acidic solution with an organic solvent and then washing this solvent with aqueous base.

Other forensic samples require different methods of preparation. Biological fluids often may be analyzed directly, or after the addition of a small amount of D<sub>2</sub>O for signal frequency locking. Plant material, tissue samples, and biological fluids may be put through extraction procedures to obtain samples suitable for NMR analysis. Explosives and accelerant samples may be prepared by normal chemical procedures. Postexplosion residues and fire debris may be extracted, and samples from these extracts prepared in the normal way.

## Applications

### Analysis of Drugs and Related Materials

**Identification of 'known' compounds** The most common application of NMR in forensic analysis is its use in the identification of materials in drug seizures. For this purpose the analyst compares the NMR spectra for the exhibit with those for authentic materials, in much the same way as for IR spectra, allowing the rapid identification of illicit drugs. When the spectra being compared are from spectrometers of different magnetic field strengths, or if they are determined using different solvents or at different concentrations, the analyst must be aware that the appearance of  $^1\text{H}$  spectra may appear to be very different. Because these kinds of differences are usually minor for  $^{13}\text{C}$  NMR spectra (which are also much simpler in appearance than  $^1\text{H}$  spectra), most commercial spectral libraries are for  $^{13}\text{C}$  spectra. However, modern spectrometers allow users to create, maintain, and search their own proton spectral libraries. Also, a number of books are available that contain collections of proton spectra of chemicals of forensic interest.

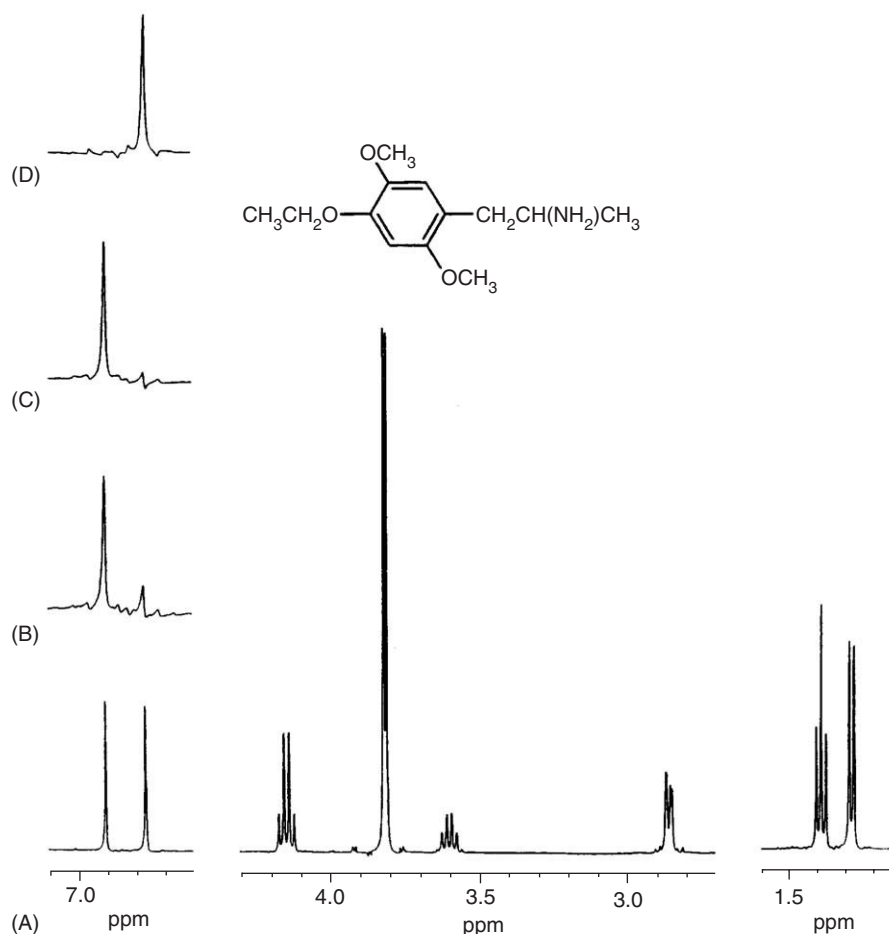
It is often difficult to obtain matching IR spectra for drugs that possess hygroscopic or polymorphic properties. Also, many drugs fragment in such a manner that their mass spectra are ambiguous and difficult to interpret. It is often difficult to distinguish between positional isomers using mass spectrometry. NMR provides an attractive alternative method of analysis since it avoids these problems, while at the same time allows identification and quantification of drugs in dosage forms.

**Identification of unknown compounds** NMR spectroscopy provides the forensic analyst with one of the most powerful techniques for identification of unknown compounds. The full range of structural elucidation techniques of modern spectrometers is available. First, the analyst obtains a high-resolution proton (NMR) spectrum in an appropriate deuterated solvent. The chemical shifts and integration in the spectrum give an indication of the types (aliphatic, olefinic, aromatic, etc.) and relative numbers of protons present in the molecule. The appearance of the coupling patterns in the molecule often provides very useful structural information. If the identity of the unknown cannot be determined from the results of the  $^1\text{H}$  NMR study alone, the analyst next obtains  $^{13}\text{C}$  information. The  $^{13}\text{C}$  NMR spectrum gives a count of the number of nonequivalent carbon atoms, as well as the types of carbon (aliphatic, aromatic, carbonyl, etc.) present in the unknown. The number of protons attached to each carbon may

be determined by using pulse sequences such as distortionless enhancement by polarization transfer, which are available on all modern spectrometers. If the structure of the unknown is still not determined from these results, the analyst may then turn to more sophisticated techniques such as nuclear Overhauser enhancement (NOE) experiments or various two-dimensional (2D) experiments.

It is crucial for the analyst to determine the exact structure of chemical substances obtained in police seizures. In a number of countries, only the compounds on a specific list are illicit. This has led to the appearance of 'designer' drugs. These are compounds that differ from illicit drugs in the position of substituents or by having had one substituent replaced by another, having an ethoxy group instead of a methoxy, for example. If no authentic samples of these materials are available, it is sometimes difficult to determine which positional isomer of a particular drug is being analyzed. Consider, for example, that for something as simple as ethoxydimethoxyamphetamine (analogs of trimethoxyamphetamines) there are 16 possible isomers. The  $^1\text{H}$  NMR spectrum of each of these compounds would be expected to be distinct. The aliphatic portion of the spectrum should be similar for all isomers. Simple examination of the aromatic portion of the spectra of ethoxydimethoxyamphetamines allows them to be sorted into three sets of isomers – those in which the two aromatic protons are *ortho* (six isomers), *meta* (seven isomers), or *para* (three isomers) to each other. These sets can be distinguished by the magnitude of the observed coupling constants for the aromatic protons,  $J_{\text{ortho}} = 8 \text{ Hz}$ ,  $J_{\text{meta}} = 2 \text{ Hz}$ , and  $J_{\text{para}} = 0 \text{ Hz}$ . All of the isomers in these sets should be easily differentiated by the use of NOE difference spectra. **Figure 1** shows the proton spectrum of the hydrochloride salt of one of the ethoxydimethoxyamphetamines, as well as the aromatic region of some representative NOE difference spectra. The aromatic protons show no coupling ( $J = 0 \text{ Hz}$ ), indicating that they are *para* to each other. This means that the substituents are in the 2-, 4-, and 5-positions. The problem is to determine in which of these positions the ethoxy is located. Irradiation of the high-field methoxy signal produced an NOE for the low-field aromatic signal (**Figure 1B**), while the low-field methoxy gave a similar effect for the high-field aromatic resonance (not shown). Only substituents in the 1- or 5-positions should produce an enhancement for the proton on the 6-position. Since both aromatic protons are affected by irradiation of the two methoxy resonances, the 5-substituent must be a methoxy. Irradiation of the signal for the alkyl methylene (at 2.86 ppm) produced NOE effects for the low-field





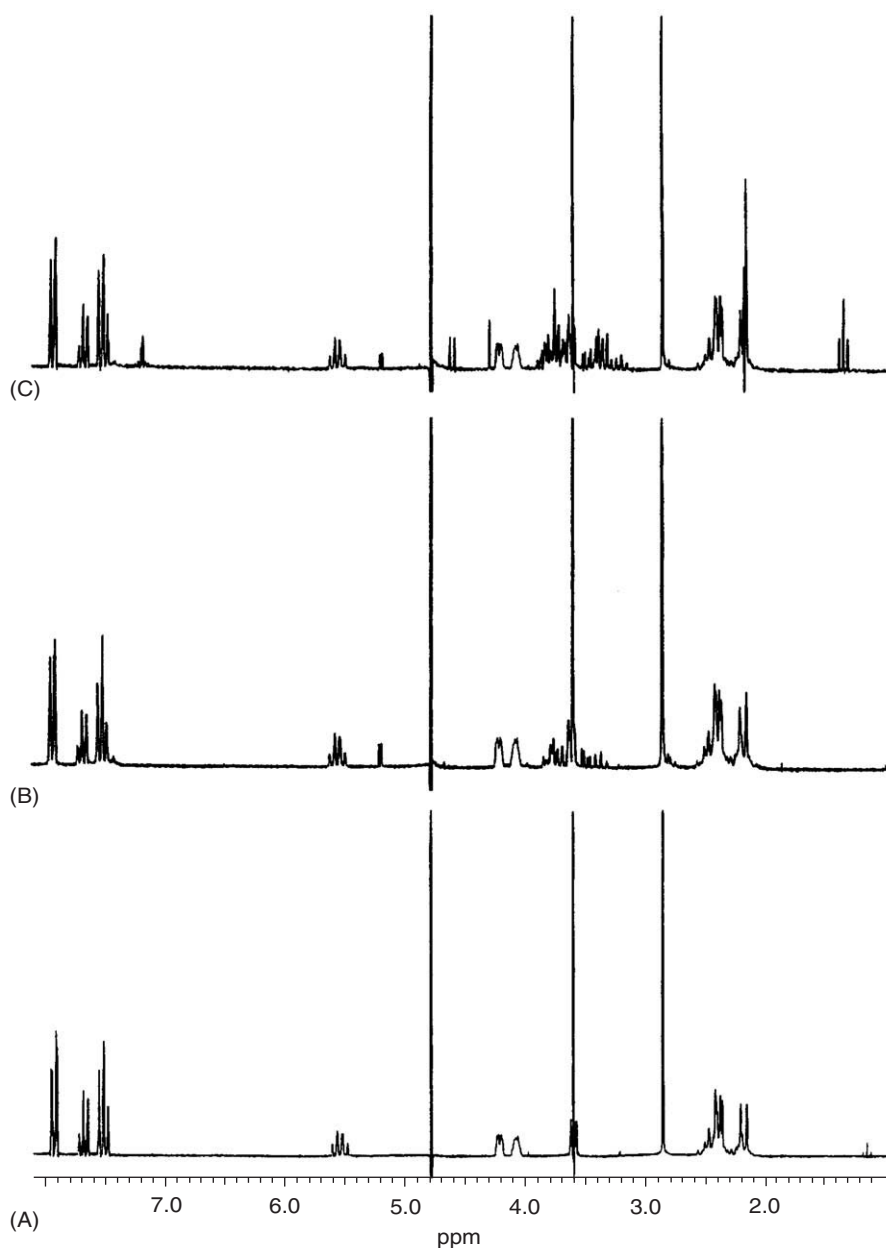
**Figure 1** 400 MHz  $^1\text{H}$  NMR spectrum of 4-ethoxy-2,5-dimethoxyamphetamine hydrochloride in  $\text{D}_2\text{O}$  (A) and the aromatic portions of the NOE difference spectra obtained by irradiating the high-field methoxy (B), the alkyl methylene (C), and the ethoxy methylene (D).

aromatic proton, which therefore must be the proton on 6-position (**Figure 1C**). It also gave an effect for the low-field methoxy (not shown), indicating that this methoxy must be in the 2-position. Irradiation of the ethoxy methylene gave rise to an NOE at the high-field aromatic proton (**Figure 1D**). This is the same proton that was enhanced by the methoxy in the 2-position, providing further proof that the ethoxy is in the 4-position. These results clearly illustrate how the structure of this chemical substance can be established as 4-ethoxy-2,5-dimethoxyamphetamine, whose structure is shown in **Figure 1**.

**Quantification** After identifying an illicit substance in a police exhibit, the forensic analyst may also be called upon to provide quantitative results. In many seizures, the drug is 'cut' with other substances. These range from starch or sugars to other drugs. For example, cocaine exhibits are often found to contain drugs such as lidocaine. NMR spectroscopy provides the analyst with a quick method for quantification of illicit substances in their exhibits. This is

done in a straightforward manner using either  $^1\text{H}$  or  $^{13}\text{C}$  spectra acquired under the proper conditions. Using these procedures the analyst can obtain absolute quantitative results for each substance present in the sample, even for fairly complex mixtures. The basic requirement for quantification is that each substance to be studied gives rise to at least one signal that is not overlapped with resonances for other materials.  $^1\text{H}$  NMR provides the quickest method for quantification because of its relative, inherent sensitivity and shorter relaxation times compared to  $^{13}\text{C}$ . However, the much wider chemical shift range and the simpler appearance of the carbon spectrum often make it the method of choice for complex mixtures.

**Impurity profiling** When an NMR spectrum is taken for a seized material, it shows peaks for the impurities as well as for the drug itself. These 'extra' peaks will often allow the identification of one or more of the impurities in the sample. This information can be used for police intelligence operations.



**Figure 2** 200 MHz <sup>1</sup>H NMR spectra of cocaine hydrochloride standard (A) and two police exhibits (B and C).

Knowledge of which impurities, including solvent residues, are present in an exhibit can often be used to determine the synthetic route used to prepare the drug or solvent used for recrystallization. This information may be used to show whether or not various exhibits were from the same source. **Figure 2** shows spectra for cocaine hydrochloride standard (**Figure 2A**) and some police exhibits (**Figures 2B** and **2C**). Note the extra peaks between 3 and 4 ppm and at ~5.2 ppm, which are due to sugars in the exhibits. The presence of lidocaine can be detected in one sample (**Figure 2C**) by the presence of singlet

resonances at ~7.1 and 2.2 ppm and a triplet at 1.4 ppm.

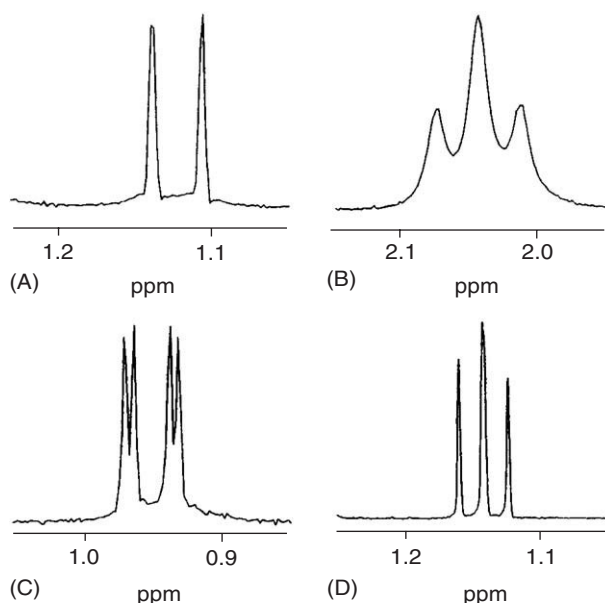
**Optical purity** There are a number of drugs for which only certain optical isomers are illicit, depending on the country. For example, methamphetamine is an illicit drug in Canada, whereas in the USA only the (+)-isomer is controlled. Other common drugs for which only specific isomers are restricted are the methorphans (levomethorphan is an internationally recognized narcotic, whereas the dextromethorphan is a common ingredient in over-the-counter cough

medications) and propoxyphenes (dextropropoxyphene is internationally recognized as a narcotic, while levopropoxyphene, an antitussive, is not). NMR spectroscopy provides a quick and easy method for determining the optical state of drug substances. It should be noted that even in cases where optical isomers are not legislated against, information on the drugs' optical states can still provide some useful information. For example, if a seizure of methamphetamine is found to be racemic or optically pure, this gives valuable intelligence information regarding the starting material and the synthetic route employed in its production.

There are a variety of NMR methods that may be used to determine the optical purity of drugs. These include the use of chiral solvating agents (CSAs), chiral lanthanide shift reagents (CLSRs), and chiral derivatization reagents (CDRs). Each of these reagents has its advantages and disadvantages, depending on the drug being studied and the field strength of the spectrometer available to the analyst. The CLSRs work best at lower fields because the line-broadening caused by these reagents increases as the square of the field strength. Normally, CSAs and CDRs do not produce as much resolution of optical isomer signals as the CLSRs. However, the CSAs and CDRs do not cause line-broadening and may be used with high-field spectrometers where better resolution is obtainable. Figure 3 shows some typical spectra for the methyl of a racemic amphetamine sample alone (Figure 3A), in the presence of tris[3-(trifluoromethylhydroxymethylene)-(+)-camphorato] europium(III) derivative (Figure 3B) and in the presence of *R*(+)-1,1'-bi-2-naphthol (Figure 3C). Also shown is the methyl resonance for the 2,3,4,5-tetra-*O*-acetyl-beta-D-glucopyranosyl isothiocyanate (GITC) derivative of amphetamine (Figure 3D). These spectra illustrate typical separation of signals for optical isomers obtained by each of these methods.

### Novel Applications of NMR in Forensic Analysis

**Arson investigations** The use of  $^1\text{H}$  NMR spectroscopy has been reported to complement the results of other methods for analyzing debris from suspected arson cases. The type of accelerant used to start the fire can be identified by the relative sizes and shapes of the peaks in different regions of the NMR spectrum – the strictly aliphatic portion from 0 to 2 ppm, the aliphatic moieties attached to aromatic components from 2 to 3 ppm, and the aromatic region from 6.5 to 7.5 ppm. Thus, paint thinner, lighter fluid, kerosene, diesel fuel, and various brands of gasoline all give their own unique patterns. These patterns



**Figure 3** 200 MHz  $^1\text{H}$  NMR spectra of the methyl signals for racemic amphetamine alone (A) and in the presence of the tris[3-(trifluoromethylhydroxymethylene)-(+)-camphorato] europium(III) derivative (B) and *R*(+)-1,1'-bi-2-naphthol (C). 400 MHz spectrum of the methyl group of the 2,3,4,5-tetra-*O*-acetyl-beta-D-glucopyranosyl isocyanate derivative of racemic amphetamine (D).

were found to be relatively unchanged when the petroleum distillates were weathered, burned, or steam distilled. Even with the low-field instruments used for these studies, distinctions between accelerants within a class could be accomplished. Certainly, especially with the increased sensitivity and resolution of the more modern high-field instruments, NMR spectroscopy could provide the arson analyst with a powerful analytical tool that could supplement other current methods of analysis.

**Analysis of explosives**  $^1\text{H}$  NMR spectroscopy has been used to study both unexploded and postexplosion samples for a variety of single explosives and mixtures. It provides a simple, fast, and reliable method of identifying explosive materials, even for mixtures without preseparation. The postexplosion samples were prepared by extracting debris with acetone, removing the solvent, and redissolving the residue in acetone- $d_6$ . These results fully or partially confirmed the results obtained by thin-layer chromatography (TLC) or gas chromatography-mass spectrometry in more than half of the cases studied. The sensitivity of the 250 MHz instrument used was insufficient for detection of the explosive residues in the remainder of the samples. With more modern higher-field instruments this should be less of a problem. In some cases 2,4-DNT and

2,6-DNT (impurities in TNT) were detected by NMR when they had been missed by analysts using TLC. Qualitative and quantitative determinations of small amounts of additives or impurities have been used to differentiate between homogeneous explosives originating from different sources. Characteristic impurities may allow the analyst to determine the synthetic route used to produce the explosive.

**Analysis of body fluids and tissues** Although the use of NMR spectroscopy for the study of body fluids and tissues is now common, it has not been employed frequently for forensic analysis. There have been a few studies of autopsy material extracts that clearly show the potential utility of NMR.  $^1\text{H}$  NMR has been used to identify a number of drugs from autopsy material, whereas  $^{31}\text{P}$  NMR was used quantitatively for postmortem blood, liver, and urine specimens taken from victims who committed suicide using the herbicide glyphosphate. It has also been shown that  $^1\text{H}$  NMR may be used, at least with the rat, to provide information on postmortem biochemical changes in skeletal muscle extracts that may be used to establish the time of death. Forensic analysts have also used NMR in some cases of lethal and nonlethal drug poisoning for identifying and quantitatively analyzing specimens obtained from blood, urine, and stomach contents, even in the presence of drug mixtures.  $^{31}\text{P}$  NMR spectra of gastric content, urine, and serum from poisoned people have been used as a screening tool for a number of organophosphorus pesticides. With modern spectrometers, these compounds are readily detectable in submicrogram quantities in short periods of time.  $^1\text{H}$  NMR has also been used to detect salicylate poisoning by examining the aromatic portion of spectra of urine. It has also been used for identification of persons who have been drinking by quantification of the ethanol content of their saliva.

## Trends in the Use of NMR in Forensic Analysis

NMR spectroscopy will probably continue to play a supporting role in forensic analysis. It will not be used as the primary source of analysis for routine samples, but rather as a powerful tool for solving problems in more unusual cases. It will certainly continue to play a vital role in the identification of unknown compounds such as 'designer' drugs. The ease with which NMR may be used to determine isomeric composition of drugs could make it an attractive method for determining the optical state of illicit drugs where this is a consideration, but the recent advances in chiral chromatographic methods may limit its use in this area. As the number of forensic analysts with knowledge of the enhanced sensitivity, resolution, and capabilities of modern spectrometers grows, there will likely be many new uses for NMR in forensic analysis. The use of NMR in some of the areas discussed above will undoubtedly increase.

**See also:** Chiroptical Analysis. Clinical Analysis: Sample Handling. Forensic Sciences: Blood Analysis; Explosives; Systematic Drug Identification. Nuclear Magnetic Resonance Spectroscopy Techniques: Nuclear Overhauser Effect. Pharmaceutical Analysis: Drug Purity Determination; Sample Preparation.

## Further Reading

- Dawson BA (1991) The use of nuclear magnetic resonance spectroscopy for detection and quantitation of abused drugs. In: Gough TA (ed.) *The Analysis of Drugs of Abuse*. Chichester: Wiley.
- Mills T III, Roberson JC, McCurdy HH, and Wall WH (1993) *Instrumental Data for Drug Analysis*, 2nd edn, vols 1–5. Boca Raton, FL: CRC Press.
- Sanders JKM and Hunter BK (1993) *Modern NMR Spectroscopy – A Guide for Chemists*, 2nd edn. Oxford: Oxford University Press.

## Pharmaceutical

J R Everett, Pfizer Global R&D, Sandwich, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Nuclear magnetic resonance (NMR) spectroscopy has been widely used in the pharmaceutical industry for over a quarter of a century for the structure

elucidation of chemical samples, and this remains the main use up to this day. The preeminent position of NMR in this area reflects the fact that it is *the* most powerful tool for the elucidation of molecular structure in solution. Over the past two decades, the use of NMR spectroscopy has grown enormously as a result of the development of: (1) high-field ( $>14.1\text{ T}$ ,  $\approx 600\text{ MHz}$  for  $^1\text{H}$  NMR) superconducting magnets, giving better sensitivity and signal

2,6-DNT (impurities in TNT) were detected by NMR when they had been missed by analysts using TLC. Qualitative and quantitative determinations of small amounts of additives or impurities have been used to differentiate between homogeneous explosives originating from different sources. Characteristic impurities may allow the analyst to determine the synthetic route used to produce the explosive.

**Analysis of body fluids and tissues** Although the use of NMR spectroscopy for the study of body fluids and tissues is now common, it has not been employed frequently for forensic analysis. There have been a few studies of autopsy material extracts that clearly show the potential utility of NMR.  $^1\text{H}$  NMR has been used to identify a number of drugs from autopsy material, whereas  $^{31}\text{P}$  NMR was used quantitatively for postmortem blood, liver, and urine specimens taken from victims who committed suicide using the herbicide glyphosphate. It has also been shown that  $^1\text{H}$  NMR may be used, at least with the rat, to provide information on postmortem biochemical changes in skeletal muscle extracts that may be used to establish the time of death. Forensic analysts have also used NMR in some cases of lethal and nonlethal drug poisoning for identifying and quantitatively analyzing specimens obtained from blood, urine, and stomach contents, even in the presence of drug mixtures.  $^{31}\text{P}$  NMR spectra of gastric content, urine, and serum from poisoned people have been used as a screening tool for a number of organophosphorus pesticides. With modern spectrometers, these compounds are readily detectable in submicrogram quantities in short periods of time.  $^1\text{H}$  NMR has also been used to detect salicylate poisoning by examining the aromatic portion of spectra of urine. It has also been used for identification of persons who have been drinking by quantification of the ethanol content of their saliva.

## Trends in the Use of NMR in Forensic Analysis

NMR spectroscopy will probably continue to play a supporting role in forensic analysis. It will not be used as the primary source of analysis for routine samples, but rather as a powerful tool for solving problems in more unusual cases. It will certainly continue to play a vital role in the identification of unknown compounds such as 'designer' drugs. The ease with which NMR may be used to determine isomeric composition of drugs could make it an attractive method for determining the optical state of illicit drugs where this is a consideration, but the recent advances in chiral chromatographic methods may limit its use in this area. As the number of forensic analysts with knowledge of the enhanced sensitivity, resolution, and capabilities of modern spectrometers grows, there will likely be many new uses for NMR in forensic analysis. The use of NMR in some of the areas discussed above will undoubtedly increase.

**See also:** Chiroptical Analysis. Clinical Analysis: Sample Handling. Forensic Sciences: Blood Analysis; Explosives; Systematic Drug Identification. Nuclear Magnetic Resonance Spectroscopy Techniques: Nuclear Overhauser Effect. Pharmaceutical Analysis: Drug Purity Determination; Sample Preparation.

## Further Reading

- Dawson BA (1991) The use of nuclear magnetic resonance spectroscopy for detection and quantitation of abused drugs. In: Gough TA (ed.) *The Analysis of Drugs of Abuse*. Chichester: Wiley.
- Mills T III, Roberson JC, McCurdy HH, and Wall WH (1993) *Instrumental Data for Drug Analysis*, 2nd edn, vols 1–5. Boca Raton, FL: CRC Press.
- Sanders JKM and Hunter BK (1993) *Modern NMR Spectroscopy – A Guide for Chemists*, 2nd edn. Oxford: Oxford University Press.

## Pharmaceutical

J R Everett, Pfizer Global R&D, Sandwich, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Nuclear magnetic resonance (NMR) spectroscopy has been widely used in the pharmaceutical industry for over a quarter of a century for the structure

elucidation of chemical samples, and this remains the main use up to this day. The preeminent position of NMR in this area reflects the fact that it is *the* most powerful tool for the elucidation of molecular structure in solution. Over the past two decades, the use of NMR spectroscopy has grown enormously as a result of the development of: (1) high-field ( $>14.1\text{ T}$ ,  $\approx 600\text{ MHz}$  for  $^1\text{H}$  NMR) superconducting magnets, giving better sensitivity and signal



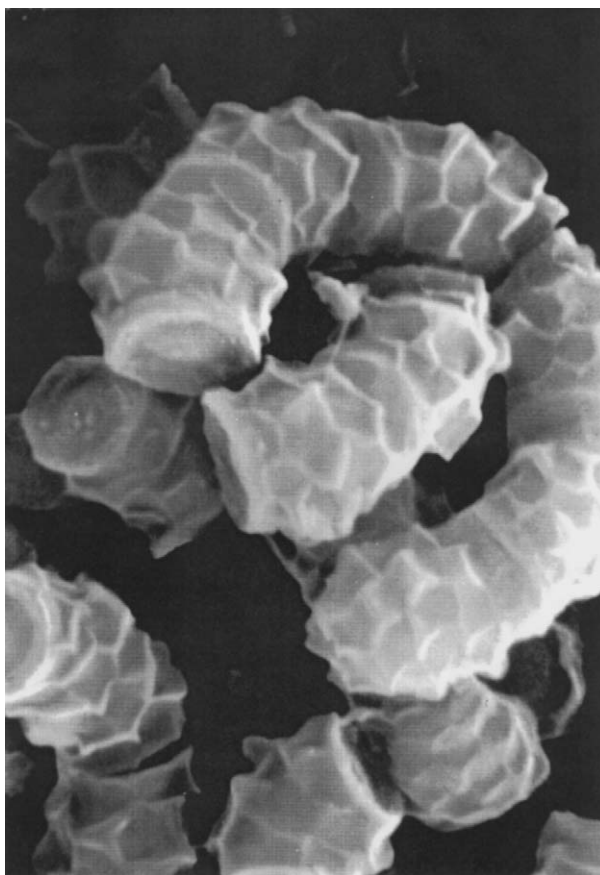
dispersion (spreading out of the signals) and (2) two-dimensional (2-D), 3-D, and 4-D NMR spectroscopy, giving better signal dispersion and phenomenal capability for determining specific atom-to-atom connectivities, either through space or through bonds, especially in large and complex molecules such as natural products and proteins. As the power of NMR has increased, so has its range of applications within the pharmaceutical industry, especially within the R&D environment. Two particularly important new applications of NMR spectroscopy are: (1) hit-finding against the protein targets that are involved in various disease states, and (2) the study of diseases, their treatment, and of drug safety via the new science of metabolomics. Indeed, nowadays, NMR is increasingly used to solve biological problems, and the technologies used range from biofluid NMR to *in vivo* NMR spectroscopy of perfused organs and whole animals, right through to magnetic resonance imaging (MRI) of whole animals and human beings. This article focuses on the key areas in this huge field of endeavor and is necessarily selective.

## Chemical Samples

### Molecular Structure Elucidation

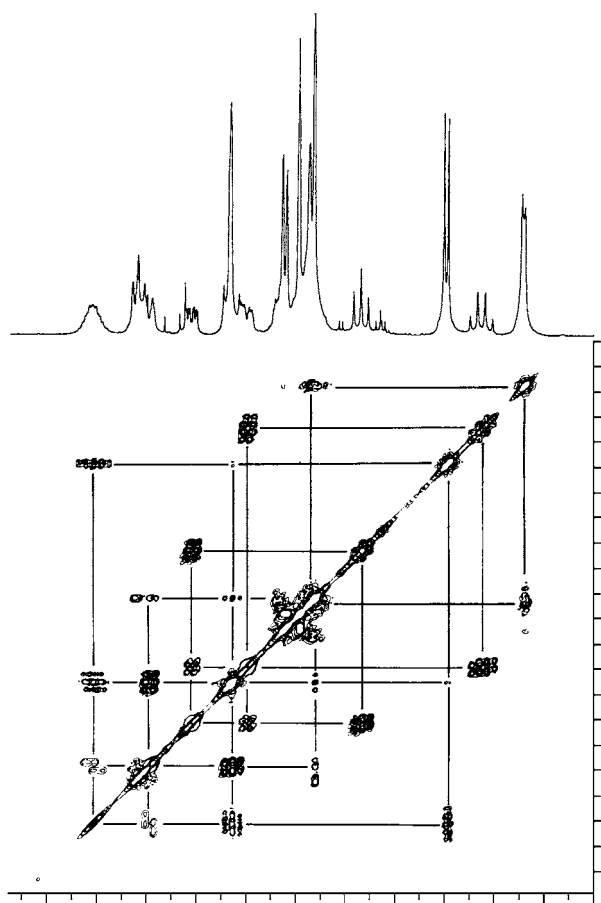
Structure elucidation is an exercise in which a large number of techniques may be used: NMR spectroscopy on its own will often not be sufficient to solve a problem, at least in an efficient manner. Ultraviolet (UV) spectroscopy gives information on chromophores in molecules, whilst mass spectroscopy (MS) is used to determine the molecular mass and, for compounds with molecular masses less than  $\sim 1000$  Da, the molecular formula, via high-resolution measurements on the molecular ion. In addition, fragmentation patterns in the mass spectrum can often be interpreted in terms of structural features in the molecule. Infrared (IR) spectroscopy provides information on functional groups such as ketone, unsaturated ester groups in the molecule, but NMR spectroscopy remains the only method of structure elucidation in solution that is capable of unambiguously piecing together structural fragments into the final molecular structure. NMR is uniquely powerful in its ability to: (1) elucidate molecular structure, (2) distinguish between isomers, (3) provide information on molecular conformations, and (4) give information on molecular dynamics. These points are best illustrated with examples.

**Natural products** The structure determination of a natural product is often more demanding than that



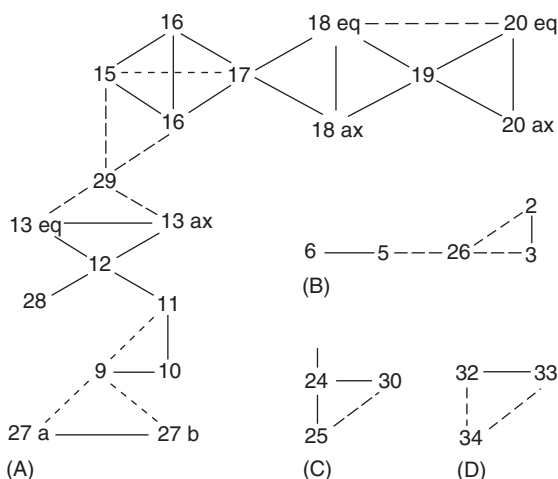
**Figure 1** Scanning electron micrograph of a novel *Streptomyces* species from which compound (I) and many other related milbemycins have been isolated. (See Baker GH, Dorgan RJJ, Everett JR, Hood JD, and Poulton ME (1990) A novel series of milbemycin antibiotics from *Streptomyces* strain E225. II. Isolation, characterisation, structure elucidation and solution conformations. *Journal of Antibiotics* 43: 1069–1076.)

of a synthetic compound, for which the precursor and synthetic protocol will usually be well defined. **Figure 1** shows a scanning electron micrograph of a novel *Streptomyces* species that produced a series of active anthelmintic (antiworm) metabolites. After isolation, the structure of metabolite (I) was determined unambiguously using a variety of spectroscopic techniques, but principally NMR. Compound (I) possesses a UV maximum at 224 nm, suggesting the presence of a diene chromophore. The high-resolution electron impact mass spectrum shows a molecular ion at  $m/z$  568.3412, which indicates that the molecular formula is  $C_{34}H_{48}O_7$  (theoretical mass 568.3400). The IR spectrum shows the presence of a saturated ester function and alcohol groups. This was the sum total of the information prior to the application of NMR. The 2-D  $^1H$ ,  $^1H$  correlation spectroscopy (COSY)-45 NMR spectrum of (I) (**Figure 2**) gives information on which protons in



**Figure 2** A contour plot of a portion of the 400 MHz 2-D  $^1\text{H}$  COSY-45 NMR spectrum of (I) in  $\text{CDCl}_3$ :TMS. Connectivities between the signals of  $J$ -coupled protons are traced out with horizontal and vertical lines and are labeled. (See Baker GH, Dorgan RJJ, Everett JR, Hood JD, and Poulton ME (1990) A novel series of milbemycin antibiotics from *Streptomyces* strain E225. II. Isolation, characterisation, structure elucidation and solution conformations. *Journal of Antibiotics* 43: 1069–1076.)

the molecule are spin-spin coupled over (usually) two or three bonds. In this case, the experiment was tuned to pick up four- and five-bond couplings as well. The large number of through-bond, atom-to-atom connectivities (Figure 3) for the hydrogen atoms enabled a considerable portion of the structure to be assembled by inspection. However, as often occurs, the fragments of protonated carbon structure that can be assembled (see Figures 3A–3D) are separated from one another by the so-called spectroscopically silent centers: quaternary carbons and heteroatoms (all oxygen in this case). These silent centers possess no hydrogens and thus the chain of proton-to-proton connectivity can be broken. In some cases, including this one, long-range (four- and five-bond) proton-to-proton connectivities are helpful; in others the problem is overcome by

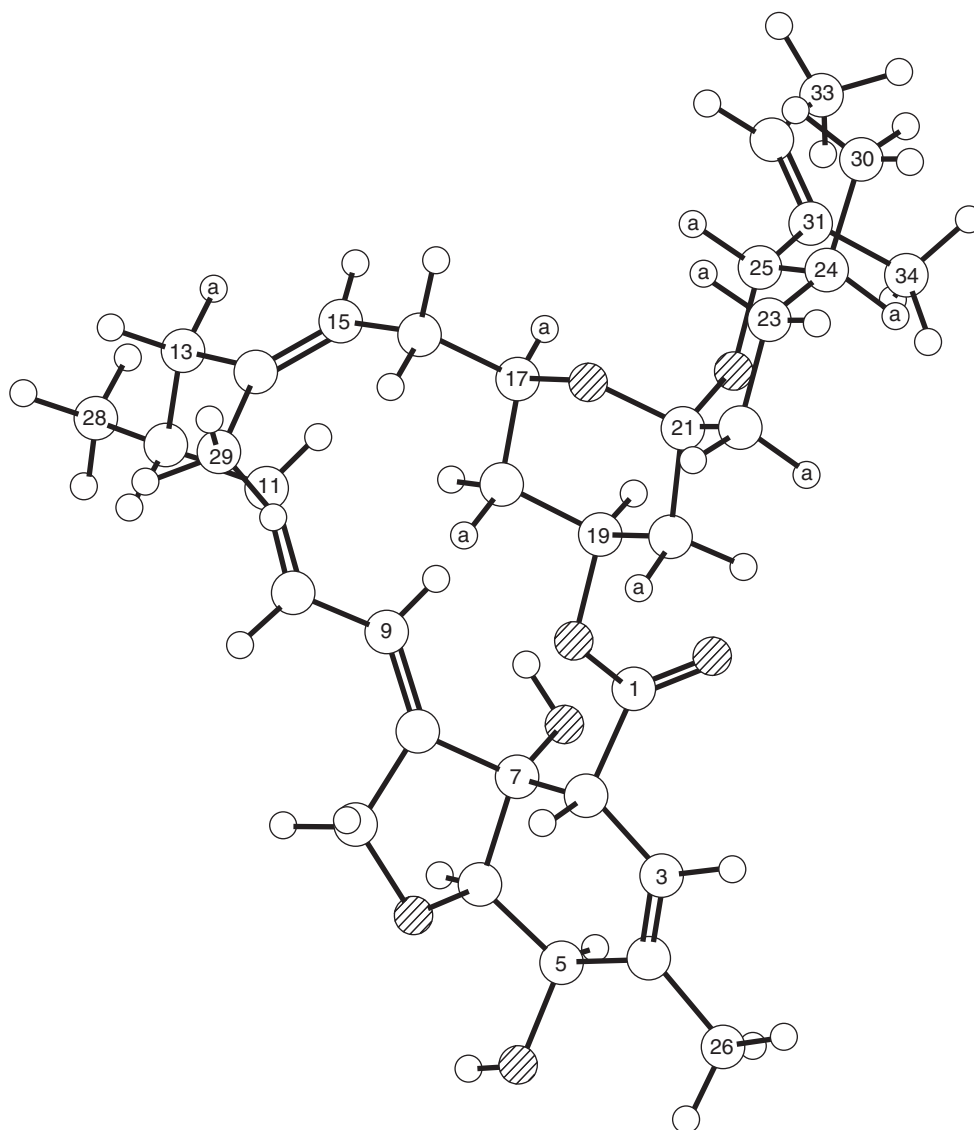


**Figure 3** Map of the four fragments (A)–(D) of proton-to-proton connectivities found for (I) in the 2-D  $^1\text{H}$  COSY-45 spectrum. Key: —  $^2J$ ,  $^3J$ ; - - -  $^4J$ ,  $^5J$  proton-to-proton through-bond connectivities. (See Baker GH, Dorgan RJJ, Everett JR, Hood JD, and Poulton ME (1990) A novel series of milbemycin antibiotics from *Streptomyces* strain E225. II. Isolation, characterisation, structure elucidation and solution conformations. *Journal of Antibiotics* 43: 1069–1076.)

obtaining long-range (two- and three-bond) proton-to-carbon connectivities from 2-D correlation experiments such as heteronuclear multiple bond correlation experiments. Thus, using this long-range connectivity information, compound (I) was shown to be a milbemycin antibiotic with structure (I) as shown in Figure 4. The relative configuration at the chiral centers of (I) was determined by the measurement of stereospecific spin-spin coupling constants ( $J$ -values) and from nuclear Overhauser effect (NOE) experiments. For instance, the large value of the three-bond coupling between H24 and H25 ( $^3J_{24,25} \sim 9.3$  Hz) indicates that the O21–C25 tetrahydropyran ring is in a chair conformation with H24 and H25 *trans*-diaxial or *anti* to one another. The observation of NOEs between H17 and H25 confirmed that these two hydrogens were close in space and thus the stereochemistry of the *spiro* ring fusion. On a modern NMR spectrometer operating at 500 MHz or higher, the entire NMR data collection and analysis exercise could take less than 2 days, with <1 mg of reasonably pure material.

### Conformational Analysis in the Solution State and in the Solid State (Polymorphism)

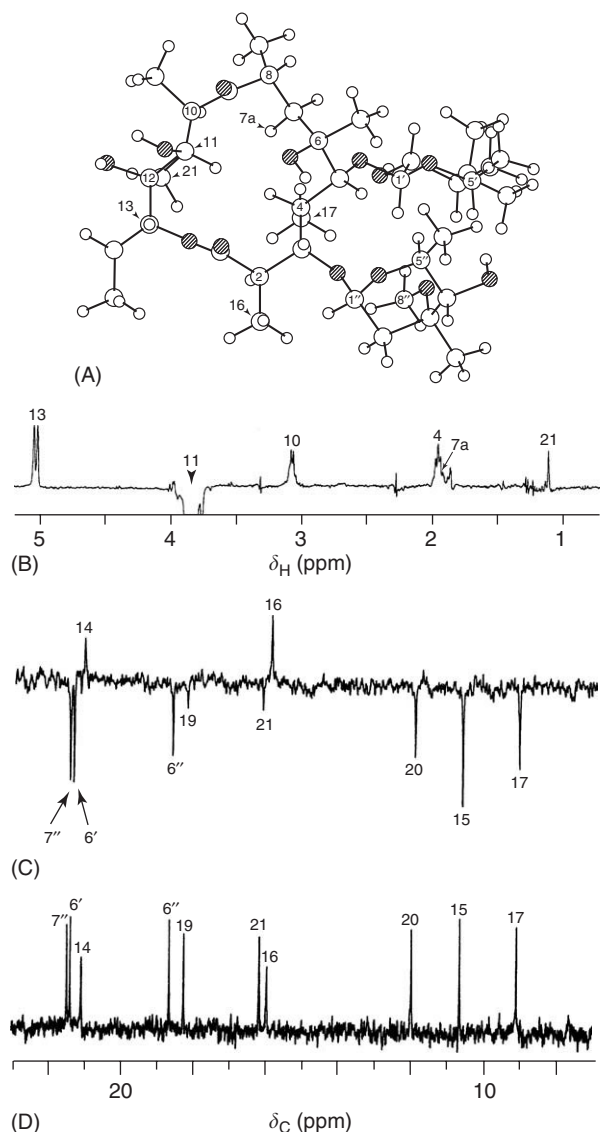
**Solution state** One goal of medicinal chemistry research is to understand the structural features responsible for the biological activity of drugs. The structure elucidation of a drug substance is an important first step in this process, but equally



**Figure 4** The molecular structure of (I) as determined by NMR: oxygen atoms are shaded and pseudoaxial protons are labeled 'a'.

important is the understanding of the conformation (shape) and dynamics of the molecule. These can be important determinants of drug–receptor interactions and hence of drug activity. NMR spectroscopy is uniquely powerful in its ability to determine detailed aspects of molecular shape and dynamics in both solution and solid. In the solution state,  $J$ -couplings can be analyzed to give information on the conformational populations about single bonds, NOEs can be used to probe close through-space interactions between pairs of protons (or protons and hetero atoms), and, finally, relaxation time measurements are able to give detailed information on molecular dynamics. It is often useful to combine the NMR studies with X-ray crystallographic studies and with molecular modeling calculations. The

macrolide antibiotic erythromycin A provides a good example. **Figure 5A** shows the crystal structure of a salt of erythromycin A. Several features are of interest, including the close cross-ring approach of H4 and H11 in the macrolide ring and the very close approach of the hydrogens of CH<sub>3</sub>-16 to H1'' (<1.5 Å (0.15 nm)). NMR was used to determine whether these conformational features are also present in solution. **Figure 5B** shows a portion of an NOE difference spectrum of erythromycin A after the irradiation of H11. The clear NOE response at H4 shows that these two protons are close together in solution as well as in the solid state. Calculations based on the crystal structure predicted that the rotation of CH<sub>3</sub>-16 would be very restricted due to its proximity to H1''. This feature is also found in

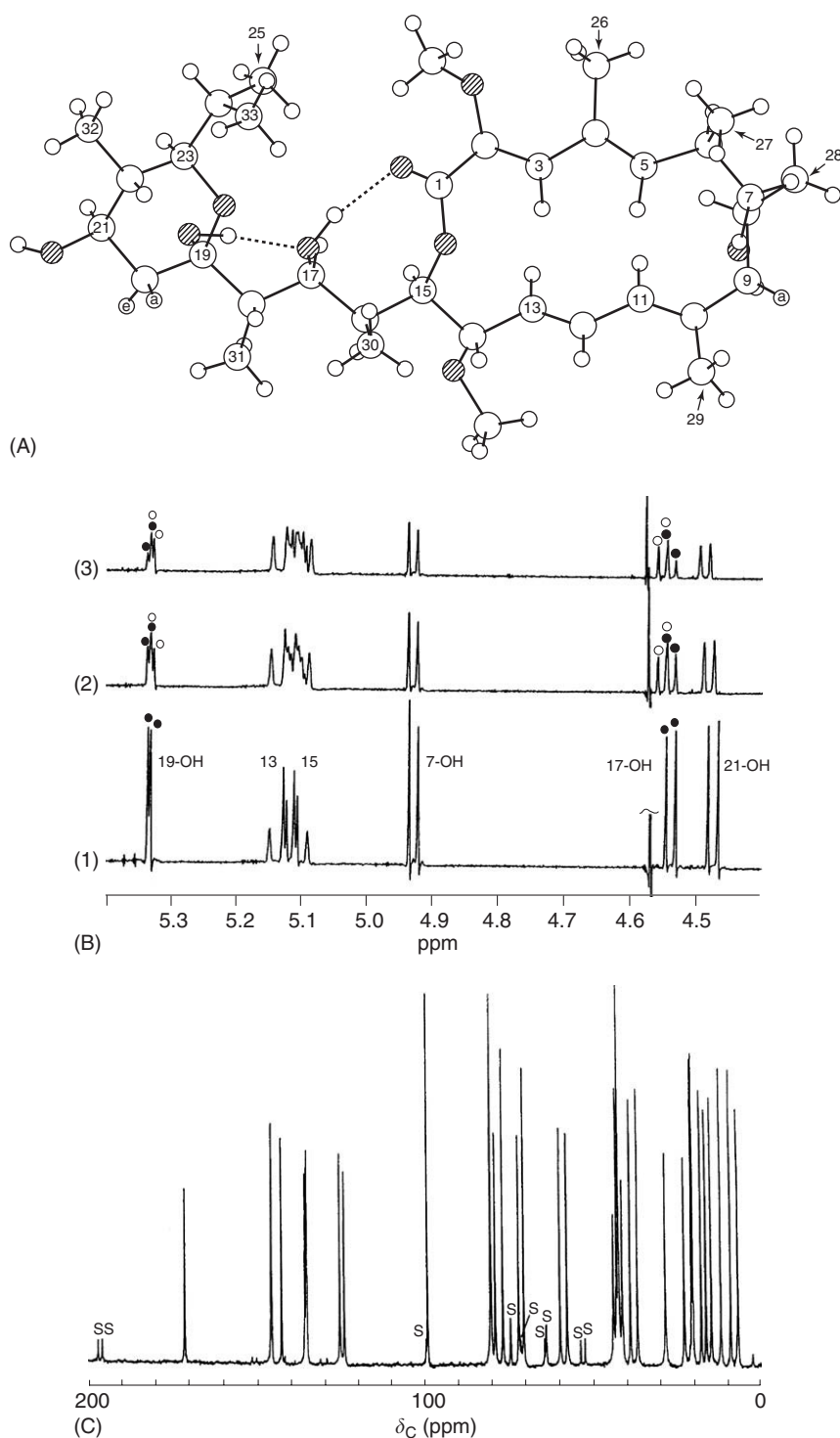


**Figure 5** (A) Crystal structure of erythromycin A hydroiodide dihydrate. The oxygen atoms are shaded; (B) 400 MHz  $^1\text{H}$  NOE difference spectrum obtained by subtraction of a control spectrum from one in which the resonance of H11 was irradiated. The only signals appearing in the NOE difference spectrum are those from protons close in space to H11. (C, D) Expansions of the low-frequency region of the 100 MHz  $^{13}\text{C}$  NMR spectrum of erythromycin A in  $\text{CDCl}_3$ :TMS; (C) Inversion-recovery spectrum with delay  $\tau = 0.25$  s and (D) Normal  $^1\text{H}$  broadband-decoupled spectrum. In an inversion recovery experiment the magnetizations of each  $^{13}\text{C}$  nucleus are inverted with a nonselective  $180^\circ$  pulse and then the relaxation rates of the magnetizations are monitored with a nonselective  $90^\circ$  pulse after a delay  $\tau$ ; the faster the relaxation of the particular nucleus, the more quickly that signal changes from an inverted state back to the relaxed, upright orientation. It is clear that the relaxation rate of methyl-16 is much faster than that of any other methyl group, as predicted for a very hindered environment. The signal for C-14 relaxes rapidly as it arises from a relatively immobile *methylene* carbon. (Reproduced with permission from Everett JR and Tyler JW (1987) The conformational analysis of erythromycin A. *Journal of the Chemical Society, Perkin Transactions II*, 1659–1667; reproduced by permission of The Royal Society of Chemistry.)

solution, as inversion-recovery NMR relaxation experiments show that the relaxation rate of  $\text{CH}_3$ -16 after perturbation is very much faster than that of any other methyl group in the molecule (see Figures 5C and 5D). On the basis of this and much other evidence, it was concluded that the major conformer of erythromycin A in  $\text{CDCl}_3$  solution has a shape very similar to that of the crystal structure.

Hydrogen bonding is often an important determinant of molecular conformation, both in solution and in the solid state. A useful NMR method for probing hydrogen-bonding patterns in solution is SIMPLE NMR, in which  $^1\text{H}$  or  $^{13}\text{C}$  NMR spectra of a compound are taken before and after the addition of a small amount of  $\text{D}_2\text{O}$  to the sample (preferably in a slow-exchanging solvent such as dimethyl sulfoxide). In the case of the hygrolyde antibiotic bafilomycin  $\text{A}_1$  (X-ray structure shown in Figure 6A), the compound was expected to adopt very similar conformations in the solution and crystalline states, on the basis of extensive NMR studies. Figure 6B shows a series of  $^1\text{H}$  NMR spectra of bafilomycin  $\text{A}_1$  both before and after the addition of small amounts of  $\text{D}_2\text{O}$ , resulting in the partial exchange of OH for OD in the hydroxy groups. The resonances of hydroxyl protons involved in a hydrogen-bonding network, such as  $-\text{O}-\text{H} \cdots \text{O}-\text{H}$ , are expected to split into two on the addition of  $\text{D}_2\text{O}$ , as each of the hydroxyl protons experiences a different chemical shift according to whether or not its partner hydroxyl group carries an H or D. The splitting observed for the C19-OH and C17-OH protons confirms the existence of the C19-O-H  $\cdots$  O(17)-H  $\cdots$  O(1)=C (1) hydrogen-bonding network observed in the crystal and inferred in solution from other NMR experiments. The equivalent  $^{13}\text{C}$  SIMPLE experiments are not only useful in studies of hydrogen bonding but are invaluable in structure determination exercises in distinguishing between carbons bearing hydroxyl groups and those bearing substituted oxygen atoms: the former group of carbons will suffer large, low-frequency, two-bond isotope effects (C-OD versus C-OH  $\sim -0.1$  ppm) whereas the latter group will not. Smaller three-bond and longer-range isotope effects are also often observed and are similarly useful in spectral assignment and structure elucidation.

**Solid state** Materials that have the same chemical composition but a different solid-state form are known as polymorphs. Polymorphs can arise if QJ; recrystallizations are carried out under different conditions and they can have different physical properties such as melting point, rate of dissolution, and rate of decomposition. Thus, a clear understanding of polymorphism is crucial to efficient drug



**Figure 6** (A) The crystal structure of the hygrolyde antibiotic bafilomycin A<sub>1</sub>: the two hydrogen bonds between C19–OH, C17–OH, and O1 are shown with dotted lines, carbon atoms 1, 2, 3, 4, 5, 10, 11, 12, and 13 are sp<sup>2</sup> hybridized. (B) A series of 400 MHz SIMPLE <sup>1</sup>H NMR spectra of bafilomycin A<sub>1</sub> in d<sub>6</sub>-dimethylsulfoxide in the region of the hydroxyl proton signals: (1) normal; (2) after the addition of 1.0 μl D<sub>2</sub>O; and (3) after the addition of 2.0 μl D<sub>2</sub>O. The solid and open circles indicate the signals of OH protons hydrogen-bonded to OH and OD partners, respectively; (C) The CPMAS <sup>13</sup>C NMR spectrum of solid, crystalline bafilomycin A<sub>1</sub>. Peaks marked s are due to spinning sidebands. (Reproduced with permission from Everett JR (1987) Novel long-range <sup>1</sup>H and <sup>13</sup>C NMR isotope effects transmitted via hydrogen bonds in a macrolide antibiotic: Bafilomycin A<sub>1</sub> and Baker GH, Brown PJ, Dorgan RJJ, and Everett JR (1989) *Journal of the Chemical Society, Perkin Transactions II* 1073–1079. The conformational analysis of bafilomycin A <sup>13</sup>C CPMAS spectrum courtesy of Dr. Mark Twyman.)



development, as one polymorph of a drug may be much more readily absorbed in the body, or easier to formulate into the final dosage form than another. Solid-state  $^{13}\text{C}$  NMR spectroscopy is a very powerful method of analyzing polymorphism and is widely used in conjunction with FTIR, powder X-ray diffraction, differential scanning calorimetry, and increasingly, single-crystal X-ray crystallography. Different polymorphs, that is, solid-state forms, of the same compound will have different NMR chemical shifts (of the order of a few tenths of ppm) in their solid-state  $^{13}\text{C}$  NMR spectra. The signals in solid-state  $^{13}\text{C}$  NMR spectra are much broader than those in the corresponding solution-state spectra, but this effect is less pronounced for crystalline samples. For instance, the  $^{13}\text{C}$  NMR spectrum of solid, crystalline bafilomycin  $\text{A}_1$  (Figure 6C, see Figure 6A for the structure) exhibits excellent resolution. This cross-polarization magic angle spinning (CPMAS)  $^{13}\text{C}$  NMR spectrum was acquired with magic angle spinning (MAS) at 3600 Hz in order to eliminate broadening due to chemical shift anisotropy, with cross-polarization (CP), i.e., transfer of the magnetization from the protons to the carbon-13 nuclei to improve sensitivity, and with high-power  $^1\text{H}$  decoupling to eliminate dipolar and spin-spin  $^1\text{H}$ ,  $^{13}\text{C}$  couplings.

### Analysis of Purity and of Impurities

NMR spectroscopy can be a useful tool for the quantitative analysis of chemical samples since, given certain provisos, the area of a signal observed in the spectrum is directly proportional to the number of nuclei giving rise to that signal, and in principle no calibrations or extinction coefficients need to be used. The provisos necessary include: (1) complete relaxation of the spin systems of interest; (2) absence of differential NOEs; (3) uniform radiofrequency excitation across the entire spectral width; (4) adequate analog-to-digital converter resolution in situations of high sample dynamic range, i.e., the need to measure very small peaks in the presence of big ones; (5) absence of overlap of the signals of interest; (6) lack of distortion of the data by any processing routines; and (7) adequate correction of any baseline distortions prior to the measurement of signal areas by integration or peak fitting. Given these provisos, it is possible to precisely measure an impurity in a sample down to the level of a few tenths of a mole per cent. However, it is sometimes found that NMR measurements of absolute purity levels of drug substances in relation to a reference standard of known purity are less precise than those obtained by separation methods, which are generally preferred.

A major new advance in this area is the development of online, liquid chromatography (LC)–NMR for both isocratic and gradient elution LC. Using a 600 MHz NMR spectrometer with a purpose-designed probe, it is now possible to obtain stop-flow  $^1\text{H}$  NMR spectra of impurities down to the  $\sim 25$  ng level. LC–NMR offers the prospect of solving the structures of components in mixtures (impurities, natural products, and biofluid components) without further separation. Use of this technology is bound to increase in the future, especially with the advent of integrated LC–NMR–MS systems that allow the integrated acquisition of UV, NMR, and MS data on the same sample.

## Biological Samples

### Molecular Structure Elucidation

**Structural biology – the molecular structure elucidation of biological macromolecules** With the advent of NMR spectrometers operating at 600 MHz and above for protons and using sophisticated 2-D, 3-D, and 4-D NMR experiments, often in conjunction with  $^{15}\text{N}$  and  $^{13}\text{C}$  isotopic labeling, it has been possible for sometime to solve the very complex  $^1\text{H}$  NMR spectra of biomacromolecules such as proteins with molecular masses up to 25 000 Da, and to determine the 3-D structures and dynamics of these molecules in solution. However, many proteins of interest to drug discovery are significantly higher in molecular weight. These large molecules had been intractable to structure elucidation by NMR spectroscopy because of the broad lines that they exhibit in their NMR spectra because of slow tumbling in solution. A new class of Transverse Relaxation Optimized Spectroscopy (TROSY) experiments has helped to break through this molecular weight barrier. TROSY exploits constructive interference between two relaxation mechanisms (chemical shift anisotropy and dipole–dipole coupling) in order to suppress the rapid relaxation that gives rise to the broad lines. However, whilst this has been a significant breakthrough, it is clear that X-ray crystallography remains the preeminent technology for protein (and nucleic acid) structure elucidation, especially in the pharmaceutical industry. The reasons for this preference are many-fold and include: (1) increased success of protein crystallization technologies; (2) lower sample mass requirements than NMR; (3) ability to determine structure from ever smaller crystals due to increased X-ray beam flux and new detector technologies; (4) faster structure solution, especially for complexes of drugs with a known protein; and (5) more precise structures due

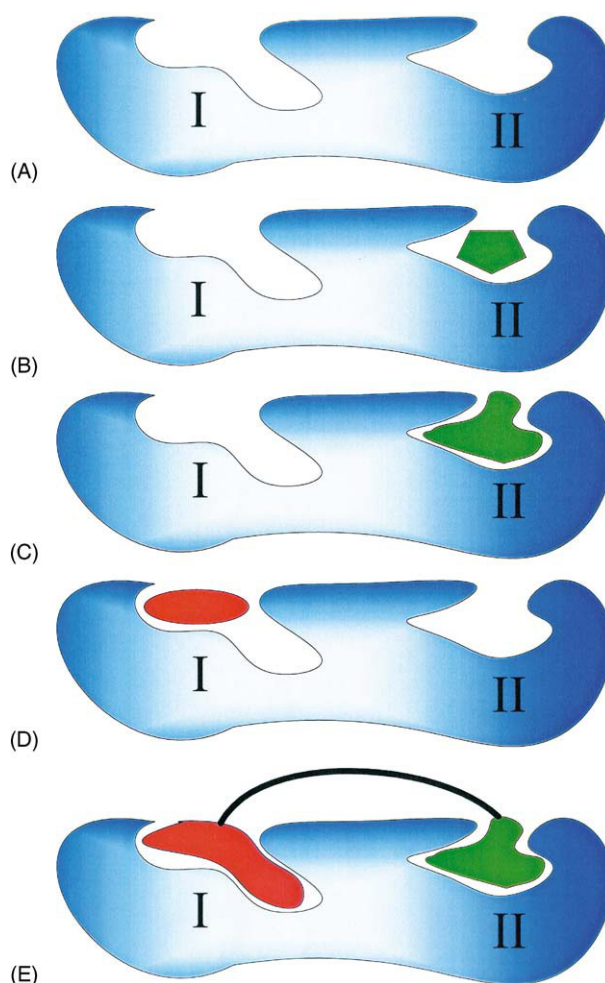
to the innate advantage of having molecular structural constraints over large distances. There is no question that NMR spectroscopy retains the advantage of solving the protein structures in the solution state and providing information on molecular dynamics, but at the time of writing nearly all pharmaceutical industry structural biology is exploited using X-ray crystallography. This position could change if new NMR technologies appear that allow faster structure solution on small samples, especially if NMR could tackle the current Achilles heel of X-ray: its inability to solve membrane protein structures easily because of reagent provision and crystallization difficulties. Because of the issues mentioned above, many NMR spectroscopists working with proteins in pharmaceutical R&D have turned their skills and the technology to a new application; hit- and lead-finding.

### Hit- and Lead-Finding Using NMR Technologies

The traditional pharmaceutical drug discovery paradigm involves the following steps:

1. Selection of a protein (usually) target on the basis of its linkage to a disease of importance and its amenability to intervention with (usually) small molecule therapeutics.
2. Generation of the protein reagent and design and development of a robust assay for the reagent that can be used in high-throughput screening (HTS) mode.
3. Execution of the HTS against a screening file that could comprise hundreds of thousand or even millions of compounds in order to generate 'hits' against that target that can then be optimized into leads and eventually drug candidates in development.

Recently, there has been a surge of interest in the application of various NMR technologies that provide an alternative hit-finding technology. The technologies rely on detecting ligand-macromolecule binding through observations of changes in the NMR spectra of either the ligand or the macromolecular target (see the reviews by Pellechia *et al.* and by Stockman and Dalvit). One of the first methods was the 'SAR by NMR' method developed by Stephen Fesik and co-workers at Abbott laboratories (Figure 7). In this method, low-molecular-weight drug fragments are screened against a  $^{15}\text{N}$ -labeled protein. Drug fragments that cause changes to the  $^{15}\text{N}$  or  $^1\text{H}$  chemical shifts of the amide groups in the  $^{15}\text{N}$  HSQC NMR spectra of the protein molecule are judged to be binding; the region of binding can be determined if the spectral data have been assigned. Iterative



**Figure 7** In SAR by NMR, amide NH chemical shift changes in a protein are used to infer binding of a ligand to a particular region. This figure represents the process (as a cartoon) for constructing a potent bidentate compound for a protein target (blue) with two adjacent binding sites I and II. (A) native protein, both binding sites empty; (B) small molecule ligand (green) is found that binds to binding site II by monitoring chemical shift mapping; (C) further screening leads to the discovery of another ligand (green) with optimized binding, i.e., optimized fit to binding site II; (D) and (E) the process is repeated for binding site I (red ligands); the two optimized ligands are then joined by a linker (black); this can rapidly lead to nanomolar binding affinity bidentate compounds, based on tethering two micromolar ligands together.

screening with related fragments can then optimize the binding of the ligand. In SAR by NMR the process is then repeated to discover an optimized binding fragment for a second, adjacent binding site. The locations of the two binding fragments in the protein structure are then determined by NMR or X-ray crystallography, and a linker is designed to join them together, often resulting in a more potent hit molecule. Many variations on this theme have been developed, including methods that monitor the

effects of binding upon ligand NMR relaxation times or diffusion properties. The advantages of these various NMR methods relative to conventional HTS technologies include:

- no need to know the natural ligand for the protein target;
- no need to know the function of the target;
- can find weak binding hits that may be interesting to optimize and may be missed by conventional methods;
- no requirement for extensive assay development and validation; and
- capable of finding hits against allosteric binding sites.

### NMR Spectroscopy of Biological Fluids: Metabonomics

NMR spectroscopy of biofluids, especially of urine, serum, and plasma, is being increasingly used in the pharmaceutical industry in studies of drug metabolism, drug safety, and clinical efficacy. The great advantage of NMR in this area is that it is a non-selective detector and will report on the levels of all low molecular mass biofluid components above the detection threshold of the particular experiment (generally better than 100 nmol for  $^1\text{H}$  NMR at 600 MHz). By contrast, many chromatographic techniques require preselection of conditions (column, eluent, etc.), which can result in the selective detection of only certain components within a complex sample such as a biofluid. This nonselective detector attribute has resulted in biofluid NMR analyses discovering previously unknown drug metabolites and biomarkers for toxic lesions and disease states.

**Metabonomics** The study of genomics and proteomics has become well established over the past decade. A new complementary science, metabonomics, has also emerged recently, which is less familiar but will become increasingly important in the future for pharmaceutical R&D. Metabonomics is defined as ‘The quantitative measurement of the multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification’. Thus, metabonomics is an *in vivo* or systems biology approach to the study of the metabolite pools (metabonomes) in living systems and how those pools reflect health and disease, aging, drug treatment, and environmental influences. Metabonomics has many application areas including:

- response of an organism to the administration of a toxin; and
- development and application of metabolic biomarkers for drug safety and efficacy.

NMR plays a key role in metabonomics as the science is typically studied by measuring the metabolite pools in various biofluids or tissues using NMR spectroscopy, although other technologies such as MS or LC–MS can be employed.

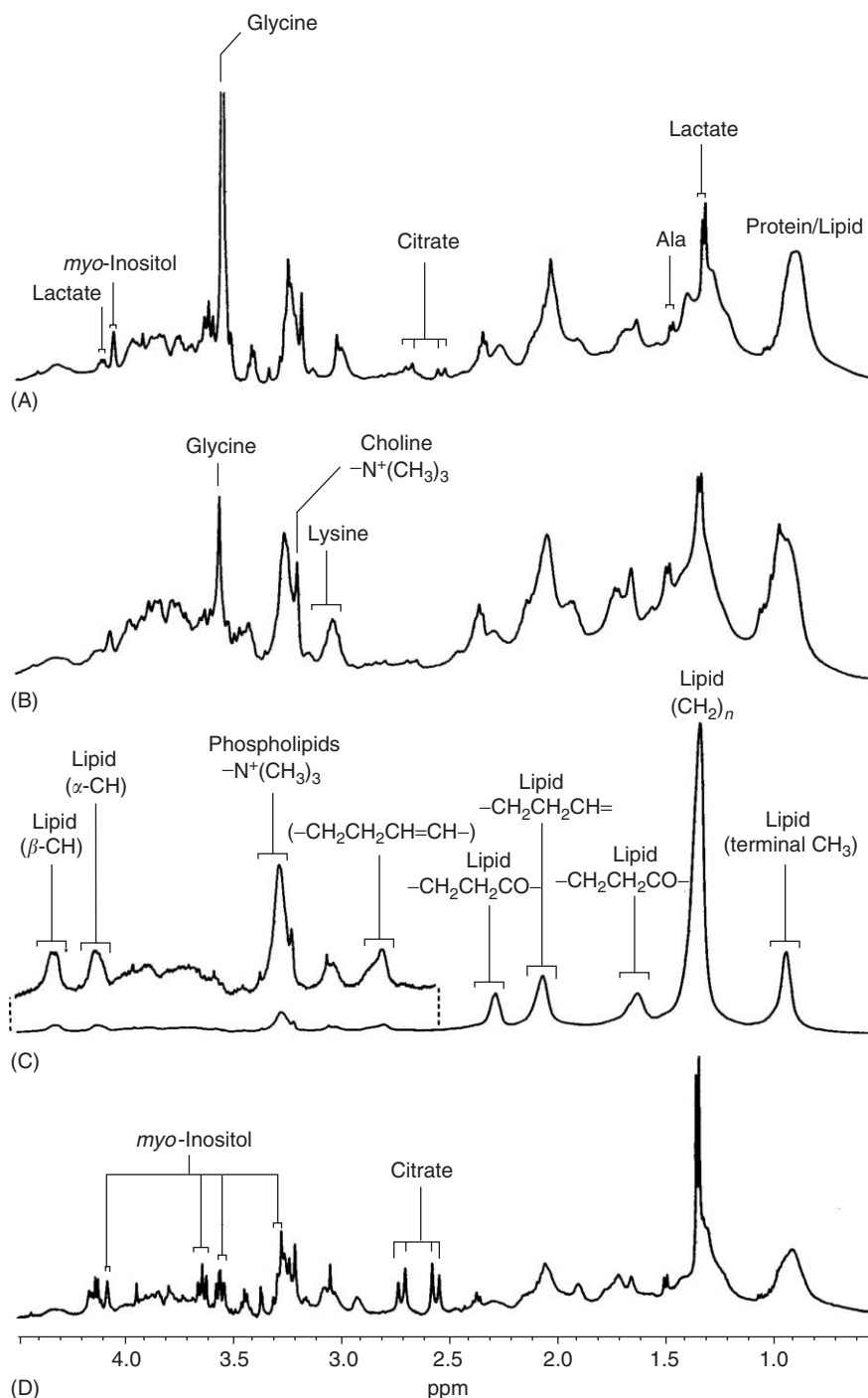
Figure 8 shows expansions from magic angle spinning  $^1\text{H}$  NMR spectrum of human prostate tissue samples. Differences were observed between the spectra of intact prostate tissue from benign prostatic hyperplasia and cancer patients. This example demonstrates that NMR methods can be applied to semisolid tissues as well as to biofluids, and points to the useful information that NMR can provide in these disease states.

In contrast to the tissue spectra, which have relatively broad signals, the 900 MHz  $^1\text{H}$  NMR spectrum of human urine exhibits thousands of (mostly) sharp lines due to the signals of hundreds of endogenous metabolites. Except in cases of dramatic or simple changes, it can be very difficult to interpret changes in such a complex spectrum and therefore computer-aided pattern recognition (PR) technologies have come to the fore to aid spectral analysis. PR methods may be unsupervised; that is, the algorithm is not informed about the group to which the subject belongs, e.g., treated versus controls, or they may be supervised, where the algorithm is informed about group membership. In each case, the purpose of the analysis is to achieve complexity reduction and reduce the dimensionality of the analysis problem. Each NMR spectrum can be thought of as a multidimensional object in chemical shift space. Indeed, most PR methods split each spectrum into  $\sim 200$  ‘buckets’ or windows, each  $\sim 0.04$  ppm wide. The spectrum is thus a 200 dimensional object with an intensity value for each dimension. In one popular unsupervised method, principal components analysis, the complex high-dimensional space is reduced to two or three dimensions, so that clustering of samples of the same type, or significant changes to the spectra, can be easily visualized.

### NMR Spectroscopy and Imaging of Living Systems

This area covers a wide range of applications from studies of biosynthesis, metabolism, and toxicity in living cells, to studies of metabolism in perfused whole organs and tissues, right through to imaging and metabolism studies of human beings and animals.

- response of an organism to disease and the treatment of that disease with medicines;



**Figure 8** 500 MHz  $^1\text{H}$  NMR magic angle spinning spectra of prostatic tissue from humans with malignant prostate cancer sampled by transurethral radical prostatectomy (TURP) (A), or needle biopsies (B, C) contrasted with that of a benign prostatic hyperplasia (BPH) patient sampled via radical prostatectomy (D). Spectrum (C) is different, as the needle has sampled a different region of the prostate. The spectrum of the prostate tissue from the BPH patient (D) displays higher levels of citrate but lower lipid levels relative to spectra (A) and (C) from the malignant cancer patients. (See Tomlins AM, Foxall PJD, Lindon JC, Lynch MJ, Spraul M, Everett JR, and Nicholson JK (1998). *Analytical Communications* 35: 113–115).

NMR studies of isolated organs have been widely used to investigate metabolic processes *in vivo* and their responses to various experimental interventions. In particular, the perfused isolated heart has

been extensively studied by  $^{31}\text{P}$  NMR spectroscopy as an aid to the understanding of the action of cardiovascular drugs.  $^{31}\text{P}$  spectra of the intact beating heart exhibit distinct resonances for ATP,

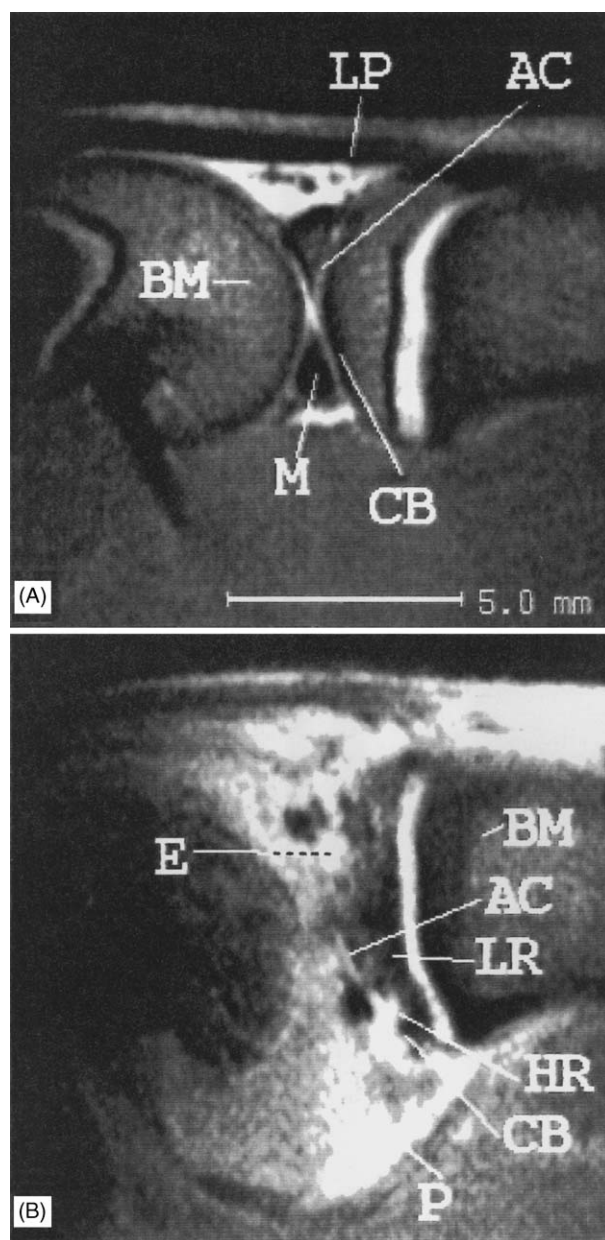
phosphocreatinine, inorganic phosphate, and phosphoesters, some of which play key roles in energy metabolism. The technology is well suited to the study of the efficacy of, for instance, cardio-protective agents in mitigating the effects of ischemia on the heart tissue. However, we will now focus on NMR applications to whole animals and human beings.

One major goal of modern medical research is to develop noninvasive methods of diagnosing diseases and monitoring their responses to therapy. Magnetic resonance imaging (MRI) and the complementary technique of magnetic resonance spectroscopy (MRS, i.e., NMR spectroscopy on living humans and animals) have themselves developed to such an extent that they are now used routinely in this role, both in the clinic and to a lesser extent in pharmaceutical research and development. In pharmaceutical research it is still necessary, or obligatory, to test experimental drugs on small laboratory animals, especially rodents, to establish both efficacy and safety, prior to their first administration to humans. The application of MRI and MRS to the study of disease models in rodents presents great technical problems relative to the clinical situation, because the small size of the animal results in poorer sensitivity and/or resolution. However, there are great benefits to be gained from the use of noninvasive imaging procedures to monitor diseases and drug therapy in experimental animals, because the same animals can be followed over long time periods, reducing the numbers of animals that need to be used. The biological significance of the results obtained is also improved relative to those obtained with invasive methods, since each animal can serve as its own control.

There are many areas of application of MRI and MRS in pharmaceutical research, including arthritis, cancer, cardiovascular diseases, central nervous system diseases, renal diseases, respiratory tract diseases, and various toxicological applications. In general, MRI is better established than MRS, although very rapid progress is being made in the latter area. For instance, it is now possible to obtain spatially localized  $^1\text{H}$  or  $^{31}\text{P}$  NMR spectra on small volumes of interest within animals or human beings.

**Magnetic resonance imaging** Arthritis is the largest single cause of disability in the UK. Rheumatoid arthritis (RA) is a severe form of the disease that is characterized by a primary inflammation in the joints, followed by later erosion and remodeling of the bones: no effective treatments exist for RA. Whilst X-radiography is well suited to monitoring changes in the bones, detection of the earlier,

soft-tissue phase of RA is difficult. MRI has emerged as an excellent tool with which to monitor the early inflammatory and later erosive stages of arthritis in both animals and man. This is well illustrated in Figure 9, which shows two  $^1\text{H}$  MR images of the knee joints of a rat with arthritis. The image of the normal knee (Figure 9A) clearly visualizes both the bony and soft tissues, including the cortical



**Figure 9**  $^1\text{H}$  magnetic resonance images in a sagittal plane through the knee of a rat with arthritis: (A) normal, (B) at day 12 of arthritis. See the text for an explanation of the labeling. (Reproduced with permission from Carpenter TA, Everett JR, Hall LD, *et al.* (1994) *Skeletal Radiology* 23: 429–437; © Springer-Verlag.)



bone (CB), the calcified menisci (M), and the patellar ligament (LP) (dark regions), the synovial fluid and epiphyseal cartilage (bright), and the articular cartilage (AC), bone marrow (BM), and deep muscle (intermediate intensity). By contrast, the image of the rat at a later stage of arthritis shows the presence of the inflammatory pannus (P), erosion of the bone outline (E), and both high-intensity region (HR) and low-intensity region (LR) within the bone, which are associated with erosion and remodeling, respectively. This methodology allows the time course of the disease and its response to drug treatments to be monitored over long periods of time in a single group of animals, thus reducing the use of experimental animals, decreasing biological variability, and expediting the evaluation of novel anti-arthritic agents.

## Emerging Techniques

In 10 years time, the face of pharmaceutical NMR is likely to have changed in a number of significant ways. Metabonomics is emerging as a technology of central importance to pharmaceutical R&D as the industry struggles with the twin imperatives of increasing the confidence in safety of the drugs in its R&D portfolio and the confidence in the biological rationale of the targets that its drugs are directed against; metabonomics will impact upon both these imperatives. The power of computing systems will drive ever more sophisticated and more automated NMR data analysis for molecular structure elucidation and for biological sample analysis. A totally new approach to multidimensional NMR has just emerged, which promises to reduce by orders of magnitude the data acquisition times for the demanding 3D, 4D, and higher-dimensional NMR experiments required for protein structure elucidation. GFT NMR (G-matrix and Fourier transform NMR) is predicted to revolutionize the use of NMR in structural biology. NMR methods are also likely to play a more important role in kick-starting the drug discovery process by providing alternative hit-finding methods to conventional HTS. An exciting era lies ahead of us.

See also: **Bioassays:** Overview. **Drug Metabolism:** Metabolite Isolation and Identification. **Infrared Spectroscopy:** Overview. **Liquid Chromatography:** Liquid Chromatography–Nuclear Magnetic Resonance Spectrometry; Pharmaceutical Applications. **Mass Spectrometry:**

Overview. **Nuclear Magnetic Resonance Spectroscopy:** Overview; Principles. **Nuclear Magnetic Resonance Spectroscopy-Applicable Elements:** Phosphorus-31. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Nuclear Overhauser Effect; Multidimensional Proton; Solid-State; Surface Coil; *In Vivo* Spectroscopy Using Localization Techniques; **Proteins:** Overview. **Spectrophotometry:** Organic Compounds. **Structural Elucidation. X-Ray Absorption and Diffraction:** X-Ray Diffraction – Single Crystal.

## Further Reading

- Evans JNS (1995) *Biomolecular NMR Spectroscopy*. Oxford: Oxford University Press.
- Gadian DG (1995) *NMR and Its Applications to Living Systems*. Oxford: Oxford University Press.
- Grant DM and Harris RK (eds.) *Encyclopedia of Nuclear Magnetic Resonance* (8 volumes). New York: Wiley.
- Günther H (1995) *NMR Spectroscopy, 2nd edn. Basic Principles, Concepts and Applications in Chemistry*. Chichester: Wiley.
- Holzgrabe U, Wawer I, and Diehl B (1999) *NMR Spectroscopy in Drug Development and Analysis*. New York: Wiley.
- Kim S and Szyperki T (2003) GFT NMR, a new approach to rapidly obtain precise high dimensional NMR spectral information. *Journal of the American Chemical Society* 125: 1385–1393.
- Lindon JC, Nicholson JK, Holmes E, and Everett JR (2000) Metabonomics: metabolic processes studied by NMR spectroscopy of biofluids combined with pattern recognition. *Concepts in Magnetic Resonance* 12(5): 289–320.
- Pellecchia M, Sem DS, and Wüthrich K (2002) NMR in drug discovery. *Nature Reviews: Drug Discovery* 1: 211–219.
- Rabenstein DL and Keire DA (1991) Quantitative chemical analysis by NMR. In: Popov AI and Hallenga K (eds.) *Modern NMR Techniques and Their Application in Chemistry*, pp. 323–369. New York: Dekker.
- Stockman BJ and Dalvit C (2002) NMR screening techniques in drug discovery and drug design. *Progress in Nuclear Magnetic Resonance Spectroscopy* 41(3–4): 187–231.
- Tishmack PA, Bugay DE, and Byrn SR (2003) Solid-state nuclear magnetic resonance spectroscopy – pharmaceutical applications. *Journal of Pharmaceutical Sciences* 92: 577–610.
- Watts A (1999) NMR of drugs and ligands bound to membrane receptors. *Current Opinion in Biotechnology* 10: 48–53.
- Wüthrich K (1986) *NMR of Proteins and Nucleic Acids*. New York: Wiley.

## Proton NMR in Biological Objects Subjected to Magic Angle Spinning

R A Wind and J Z Hu, Pacific Northwest National Laboratory, Richland, WA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

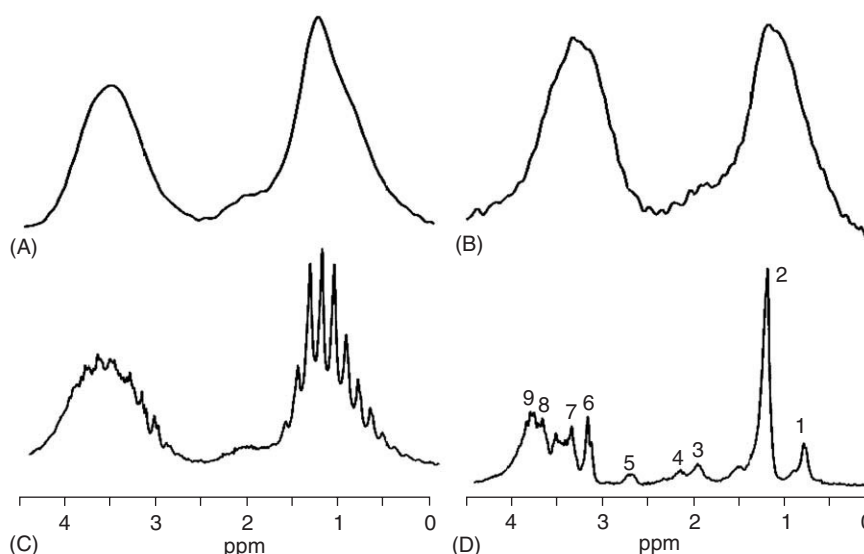
Magnetic resonance imaging (MRI) of the spatial distribution of water in biological objects has developed as one of the major tools to diagnose lesions and diseases and follow the therapy response in animals and patients in a minimally invasive way. In addition, *in vitro* and *in vivo* localized magnetic resonance spectroscopy (MRS) and spectroscopic imaging or chemical shift imaging (CSI) are increasingly used in biochemical and biomedical studies in cells, tissues, animals, and humans. With these techniques the resonance lines of several relatively small molecular weight chemical compounds such as amino acids, lipids, and other key mobile metabolites are measured, and their presence and intensities have been linked to tumor phenotype, tumor formation, tumor size, increased cell proliferation, and cell death pathways. However, a major problem associated with nuclear magnetic resonance (NMR) spectroscopy in intact biological tissues is that relatively large resonance line widths are observed, often one to two orders of magnitude larger than the widths measured in liquids using established NMR techniques. This is especially a problem for  $^1\text{H}$  NMR, which is the most widely used nucleus because of its relatively large NMR sensitivity, but has a relatively small chemical shift range of  $\sim 10$  parts per million (ppm). As usually many metabolites contribute to the NMR signal, the result is a spectrum with severely overlapping spectral lines, which seriously hampers a quantitative analysis of the spectra, and sometimes even makes it impossible to assign the spectral lines unambiguously.

In biological samples, the main mechanisms for this broadening are the local magnetic field gradients arising from variations in the isotropic bulk magnetic susceptibility near boundaries of intra- and extracellular structures, such as the various intracellular compartments, air-tissue interfaces near the lungs and sinuses, and bone-tissue interfaces. Then, chemically equivalent nuclei experience different local magnetic fields, depending on their spatial localization, giving rise to line broadening. Using cell extracts can eliminate this broadening, but this

procedure is time consuming, introduces spectral artifacts, and cannot be applied to study intact tissues and organs. Another method of improving the spectral resolution is to increase the external magnetic field  $B_0$ , as the separation between the lines in an NMR spectrum increases linearly proportional with  $B_0$ . However, the susceptibility-induced broadening is linearly proportional to  $B_0$  as well, and in many cases zero or only marginal improvements in the spectral resolution have been reported when higher field strengths were employed.

In principle, the susceptibility broadening can be averaged to zero by the technique of magic angle spinning (MAS), where the sample is rotated about an axis making an angle of  $54^\circ 44'$  relative to the external magnetic field. In a standard MAS experiment, where the NMR signal is observed after a single  $90^\circ$  radio frequency (RF) pulse (to be called single-pulse MAS or SP-MAS hereafter), the spinning frequency must be larger than the broadening in order to avoid the occurrence of spectral spinning sidebands (SSBs) surrounding the various resonance peaks, which can overlap with other resonance lines, rendering the interpretation of the spectra difficult again. In fact, in practice often the spinning frequency is chosen larger than the spectral width, i.e., a kilohertz or more, in order to avoid SSBs arising from the water signal, which occurs in biological objects in a concentration of  $\sim 30 \text{ mol l}^{-1}$ , and which is often much stronger than the metabolite signals, arising from compounds with two to three orders of magnitude smaller concentrations, even when water suppression is applied. **Figure 1** shows  $^1\text{H}$  water-suppressed NMR metabolite spectra of excised rat liver, obtained in a 7T external field (i.e., 300 MHz proton frequency) on a stationary sample and with SP-MAS at different spinning speeds, illustrating the appearance of SSBs at low spinning speeds. In fact, at 1 Hz spinning the side bands are so dense that the spectrum is virtually the same as the static spectrum. In these experiments, water suppression was achieved by preceding the  $90^\circ$  pulse by a DANTE (delays alternating with nutations for tailored excitation) water suppression sequence, consisting of a train of equally spaced small-tip-angle hard pulses.

A serious problem associated with fast SP-MAS is the large centrifugal force,  $F_c$ , induced in the sample by the spinning, which destroys tissue structures and even individual cells at high spinning rates.  $F_c$  is



**Figure 1** 300 MHz  $^1\text{H}$  NMR spectra of freshly excised rat liver samples obtained with different methods: (A) static sample; (B) 1 Hz SP-MAS; (C) 40 Hz SP-MAS; (D) 4 kHz SP-MAS. The external field was 7 T. Line assignments: 1 ( $\sim 0.88$  ppm), triglycerides  $\text{CH}_3$  terminal, or neutral amino acid methyl, valine, leucine, isoleucine methyl; 2 ( $\sim 1.28$  ppm), triglycerides  $-(\text{CH}_2)_n$ , lactate methyl, threonine methyl; 3 ( $\sim 2.04$  ppm), triglycerides  $\text{CH}=\text{CH}-\text{CH}_2-\text{CH}_2$ ; 4 ( $\sim 2.24$  ppm), triglycerides  $\text{CH}_2-\text{CH}_2-\text{CO}$ ; 5 ( $\sim 2.8$  ppm), triglycerides  $\text{CH}=\text{CH}-\text{CH}_2-\text{CH}=\text{CH}$ ; 6 ( $\sim 3.2$  ppm), choline methyl, phosphocholine methyl,  $\beta$ -glucose, trimethylamine-*N*-oxide methyl; 7 (3.4 ppm), 8 (3.6 ppm), glucose, glycogen; and 9 (3.8–4.0 ppm), glucose, glycogen, amino acids. (Reproduced with permission from Wind RA and Hu JZ (2003) Magnetic susceptibility effects in nuclear magnetic resonance spectroscopy of biological objects. In: *Recent Research Developments in Magnetism and Magnetic Materials*, vol. 1, pp. 147–169.)

given by  $F_c = m\omega^2 r$ , where  $m$  is the mass,  $\omega = 2\pi F$ ,  $F$  being the spinning frequency, and  $r$  the distance from the rotation axis to the point of interest. For example, when  $F = 2$  kHz and  $r = 1$  cm,  $F_c = 1.6 \times 10^5$  times the gravitational force  $F_g$ . Therefore, the SP-MAS method is not viable for MRS or CSI in large intact biological samples or *in vivo* studies, and methods are needed that yield high-resolution, SSB-free spectra at reduced MAS frequencies. In the remainder of this article, two such methods and some applications will be discussed, following short background introductions about magnetic susceptibility and MAS.

## Magnetic Susceptibility

The magnetic susceptibility factor  $\chi$  arises from the magnetization  $\mathbf{M}$  induced in a material exposed to an external magnetic field  $\mathbf{H}_0$ , corresponding to a magnetic induction  $\mathbf{B}_0 = \mu_0 \mathbf{H}_0$ , in the absence of the material, where  $\mu_0 = 4\pi \times 10^{-7}$  (T m A $^{-1}$ ) is the magnetic permeability in vacuum. When the magnetization is oriented in the same direction as  $\mathbf{B}_0$ ,  $\chi$  is a scalar and the susceptibility is called isotropic.  $\chi$  is called the volumetric susceptibility and is a unitless quantity defined as  $\chi = \mathbf{M}/\mathbf{H}$ , where  $\mathbf{M}$  is the magnetic dipole moment per unit volume and  $\mathbf{H}$  is the total magnetic field strength.  $\chi$  is equal to 0 in free space and is practically 0 for air. In biological systems, the

additional magnetic field strength induced by  $\mathbf{M}$  is much less than  $\mathbf{H}_0$  so that  $\chi \approx \mathbf{M}/\mathbf{H}_0$ . With this definition, in the material the net magnetic field induction  $\mathbf{B}$ , which is the fundamental magnetic field responsible for NMR, is given by  $\mathbf{B} = \mu_0 \mathbf{H}_0 + \mu_0 \mathbf{M} \approx (1 + \chi)\mathbf{B}_0 = \mu_r \mathbf{B}_0$ , where  $\mu_r$  is the relative magnetic permeability.  $\mathbf{M}$  can arise from several sources: (1) The nuclear magnetization. Although this magnetization is responsible for the MR signal, its contribution to  $\mathbf{M}$  can usually be neglected. (2) The magnetization associated with slight changes in the angular velocities of paired electrons in their orbitals, induced by  $\mathbf{B}_0$ . According to Lenz's law this will cause a magnetic field at the center of the orbital that opposes  $\mathbf{B}_0$ , resulting in the well-known chemical shift. Outside the orbital at a distance large compared with the dimensions of the orbital, the effect of  $\mathbf{B}_0$  on an electron orbital can be approximated by a local magnetic field arising from a magnetic dipole associated with a ring current in the orbital. This results in a negative value of  $\chi$ , the diamagnetism. (3) Paramagnetic susceptibility arises in materials containing unpaired electrons, resulting in a positive value of  $\chi$ . (4) Ferromagnetism occurs in materials possessing permanent magnetic moments aligned in Weiss domains.

Most biological objects are diamagnetic. Water and organic compounds possess susceptibility factors  $\chi$  of the order of  $-4\pi \times 10^{-6}$ . In liquids, this

susceptibility gives rise to an induced additional homogeneous magnetic field in the sample, resulting in a shift of the resonance lines rather than a line broadening. However, in biological samples, which contain many intra- and intercellular structures, the various compounds are distributed heterogeneously, and often close to a boundary of a material with a different susceptibility factor. Then, the susceptibility differences between the various compartments induce magnetic field gradients in the sample. As a result the resonance shifts become space dependent, resulting in a line broadening. Often, line widths of the order of 0.5 ppm are observed, at least one order of magnitude larger than their intrinsic line widths. These values are in accordance with estimations based on the susceptibility differences one can encounter at interfaces in biological objects.

## Magic Angle Spinning

In order to illustrate the effect of MAS we consider the magnetic moment  $\mathbf{m}$  arising from a sphere of a stationary diamagnetic material (see Figure 2A). At a distance  $R$  equal to or larger than the radius  $a$ , the magnetic field generated by this moment has two components  $B_r$  and  $B_\gamma$ , oriented in the direction and perpendicular to  $R$ , respectively. These components are given by

$$B_r = \frac{\mu_0}{2\pi} \frac{m \cos \gamma}{R^3} \quad \text{and} \quad B_\gamma = -\frac{\mu_0}{4\pi} \frac{m \sin \gamma}{R^3}$$

where  $\gamma$  is the angle between  $R$  and the direction of the magnetic moment, defined as the  $z$  direction. It follows that the dipolar field component along the  $z$ -axis,  $B_z$ , is given by

$$B_z = B_{rz} + B_{\gamma z} = \frac{\mu_0 m}{4\pi R^3} (3 \cos^2 \gamma - 1) \quad [1]$$

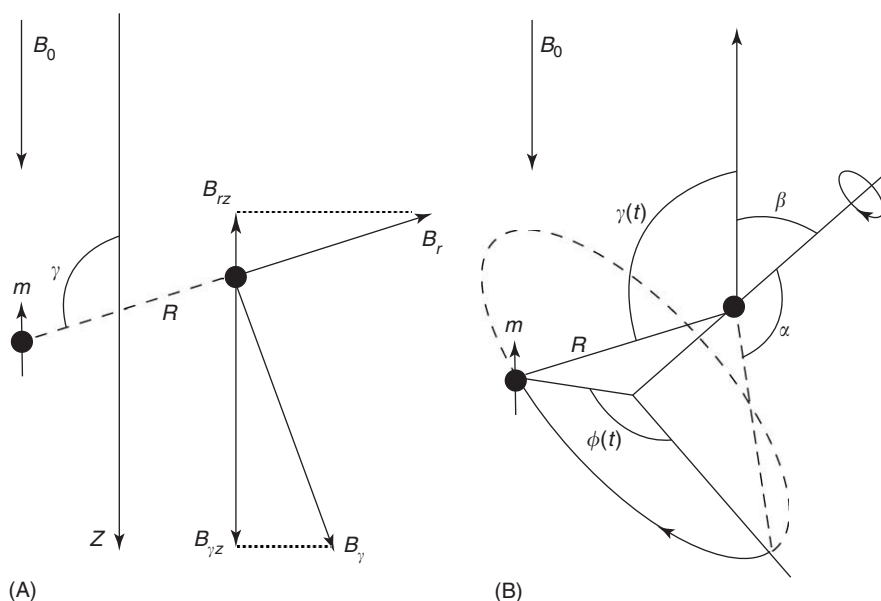
Figure 2B shows the case that the sample is rotated with an angular frequency  $\omega_r$  about an axis making an angle  $\beta$  with the external field. Then, the angle  $\gamma$  and the azimuth angle  $\phi$  become time dependent, and  $\cos \gamma(t)$  is given by

$$\begin{aligned} \cos \gamma(t) &= \cos \alpha \cos \beta + \sin \alpha \sin \beta \cos[\phi(t)] \\ &= \cos \alpha \cos \beta + \sin \alpha \sin \beta \cos(\omega_r t + \phi_0) \end{aligned} \quad [2]$$

$\alpha$  is the angle between the rotation axis and  $R$ . Equation [1] becomes

$$\begin{aligned} B_z &= \frac{\mu_0 m}{8\pi R^3} \{ (3 \cos^2 \alpha - 1)(3 \cos^2 \beta - 1) \\ &\quad + 3 \sin 2\alpha \sin 2\beta \cos(\omega_r t + \phi_0) \\ &\quad + 3 \sin^2 \alpha \sin^2 \beta \cos[2(\omega_r t + \phi_0)] \} \end{aligned} \quad [3]$$

Hence,  $B_z$  contains a static term and two time-dependent terms. Then, in a standard SP-MAS experiment after Fourier transformation the NMR spectrum arising from the interactions between the nuclear spins and these local fields consists of a center-band line, located at the frequency determined by the static term, and SSBs, located at frequency



**Figure 2** Magnetic dipole field outside a sphere of a diamagnetic material with dipolar moment  $\mathbf{m}$  induced by the external field  $\mathbf{B}_0$ : (A) static sample; (B) rotating sample. (Reproduced with permission from Wind RA and Hu JZ (2003) Magnetic susceptibility effects in nuclear magnetic resonance spectroscopy of biological objects. In: *Recent Research Developments in Magnetism and Magnetic Materials*, vol. 1, pp. 147–169.)

distances  $\pm n\omega_r$ ,  $n = 1, 2, 3, \dots$ , from the center-band line and with amplitudes depending on the static line width and  $\omega_r$ . Hence, by choosing the angle  $\beta = \cos^{-1}\sqrt{3} = 54.74^\circ$ , the effect of the static term becomes zero, irrespective of the angle  $\alpha$ , and for spinning speeds much larger than the spectral line widths the SSBs have very small amplitudes and can often be neglected. As a result, the susceptibility shift in a homogeneous liquid and the susceptibility line broadening in heterogeneous samples are eliminated. Moreover, MAS eliminates other line broadenings as well, arising from other interactions with a similar angular dependence as eqn [1]. Table 1 summarizes the various interactions that can occur in a material and the effectiveness of MAS in averaging out these interactions. As already mentioned above,

**Table 1** The impact of magic angle spinning on the various spin interactions playing a role in NMR

Interaction	Impact of MAS <sup>a</sup>
Indirect spin–spin or <i>J</i> -coupling	N
Static spin–spin dipolar coupling	Y
Isotropic chemical shift	N
Static anisotropic chemical shift	Y
First-order quadrupolar coupling	Y
Second-order quadrupolar coupling	P
Isotropic susceptibility	Y
Anisotropic susceptibility	P
Spin–lattice relaxation time $T_1^b$	N
Intrinsic spin–spin relaxation time $T_2^b$	N

<sup>a</sup>Y: MAS averages the interaction to zero, N: MAS does not affect the interaction, P: MAS partially averages the interaction.

<sup>b</sup>If determined by interactions rendered time-dependent by molecular motions.

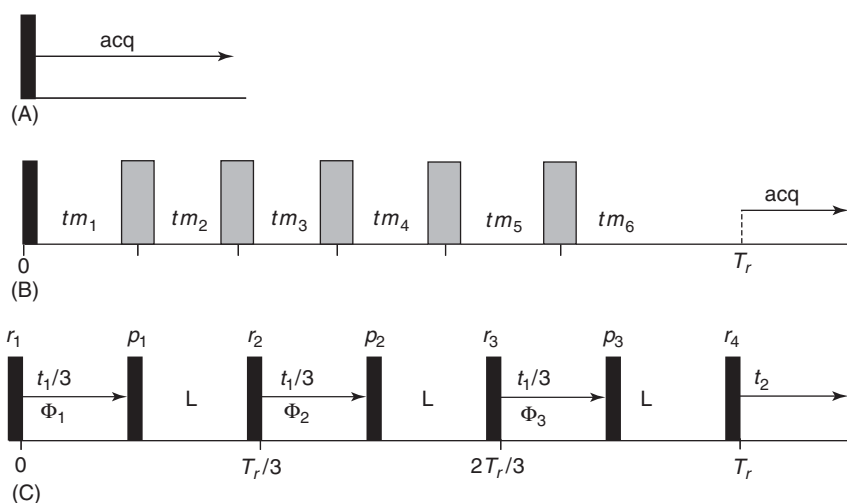
in biological materials the (dia)magnetic susceptibility is the main source of the line broadening observed in stationary samples.

## Slow-MAS Techniques

In solid-state NMR, several methods have been developed where slow MAS is combined with special RF pulse sequences to suppress the spinning sidebands or to separate them from the isotropic spectrum so that a sideband-free high-resolution isotropic spectrum is obtained. Examples of such methodologies are total suppression of spinning sidebands (TOSS), phase-adjusted spinning sidebands (PASS), and phase-corrected magic angle turning (PHORMAT). While TOSS cannot be used at low spinning speeds because of serious spectral distortions, it was found that PASS and PHORMAT can be modified successfully for studies of biological samples at low speeds, an order of magnitude or more lower than the speeds typically used in solid-state NMR experiments. The basic RF pulse sequences used in SP-MAS, PASS, and PHORMAT are shown in Figure 3. In the following paragraphs PASS and PHORMAT will be described briefly (SP-MAS is self-explanatory).

### PASS

PASS is a one-rotor-period ( $T_r$ ), constant evolution time 2D experiment, during which five  $\pi$  pulses are applied, with time intervals  $tm_1$ – $tm_6$  (Figure 3B). In PASS, the center-band spectrum and the SSB spectra are separated by order. This is achieved by acquiring



**Figure 3** Three RF pulse sequences used in combination with MAS. The  $90^\circ$  pulses are black while the  $180^\circ$  pulses are gray: (A) SP-MAS; (B) PASS; (C) the prototype PHORMAT. The various timing parameters are explained in the text. (Reprinted with permission from Hu JZ and Wind RA (2002) The evaluation of different MAS techniques at low spinning rates in aqueous samples and in the presence of magnetic susceptibility gradients. *Journal of Magnetic Resonance* 159: 92–100; © Elsevier)

the signal after a series of PASS experiments with different values of the time intervals  $tm_1$ – $tm_6$ . Each combination of time intervals has been chosen in such a way that the contribution of the signal in the observed free induction decay (FID) in the acquisition dimension ( $t_2$ ), arising from the center band and SSBs, is proportional to a phase factor given by  $\exp(-ik\Theta)$ , where  $k$  denotes the sideband order and  $\theta$  is a variable called ‘pitch’. Then after 2D Fourier transform with respect to  $t_2$  and  $\theta$  a series of spectra is obtained that separates the contributions for each  $k$  value, i.e., it separates the center-band and sideband spectra. In practice, it suffices to use  $n$  discrete values of  $\theta$ , varying from 0 and  $2\pi$  in steps of  $2\pi/n$ , where  $n$  denotes the total number of center-band and sideband spectra that have to be resolved. **Figure 4A** shows the stacked  $^1\text{H}$  PASS water-suppressed spectra obtained on excised rat liver in a 7 T field using a spinning speed of 40 Hz. Water suppression was achieved by preceding the PASS sequence by a DANTE sequence. In this PASS experiment, 16 different combinations of delay times  $tm_1$ – $tm_6$  are used, which makes it possible to separate the center-band and 15 sideband spectra without spectral aliasing. **Figure 4B** shows the center-band spectrum separately. It follows that the spectral resolution is at least the same as that obtained in the standard fast SP-MAS experiment shown in **Figure 1D**, and similar results have been obtained in other excised tissues and organs.

## PHORMAT

PHORMAT is a regular 2D experiment, with a variable evolution time and a (fixed) acquisition time (**Figure 3C**). The PHORMAT methodology is based on the so-called magic angle hopping (MAH) experiment, where the sample is hopped over angles of  $120^\circ$  about an axis at the magic angle. In PHORMAT, the sample is spun slowly and continuously instead of hopped. The prototype PHORMAT sequence is shown in **Figure 3C**. The parameter  $T_r$  denotes the rotation period of the sample,  $t_1$  is the variable evolution time, and  $t_2$  is the acquisition time. The  $90^\circ$  pulses labeled  $r_1$ ,  $r_2$ ,  $r_3$ , and  $r_4$  are synchronized to  $1/3$  of the rotor period, and rotate the magnetization into the transverse plane. Then during the evolution periods,  $t_1/3$ , the magnetization precesses through angles  $\Phi_1$ ,  $\Phi_2$ , and  $\Phi_3$ , respectively. The  $90^\circ$  pulses labeled  $p_1$ ,  $p_2$ , and  $p_3$  are the storage pulses, which project a component of the precessing magnetization after the corresponding  $t_1/3$  period to the  $z$ -axis, where it remains during the storage periods labeled L. A FID is acquired following the last  $90^\circ$  pulse ( $r_4$ ). For local fields arising from second-rank interactions, such as interactions with the

magnetic susceptibility fields, the sum of the precession angles  $\Phi_1$ ,  $\Phi_2$ , and  $\Phi_3$  averages to the isotropic values of the interactions, i.e.,  $\Phi_1 + \Phi_2 + \Phi_3 = \omega_{\text{iso}}t_1$ . With a proper phase cycling of the projection pulses,  $p_1$ ,  $p_2$ ,  $p_3$  and the receiver, the FID can be expressed as

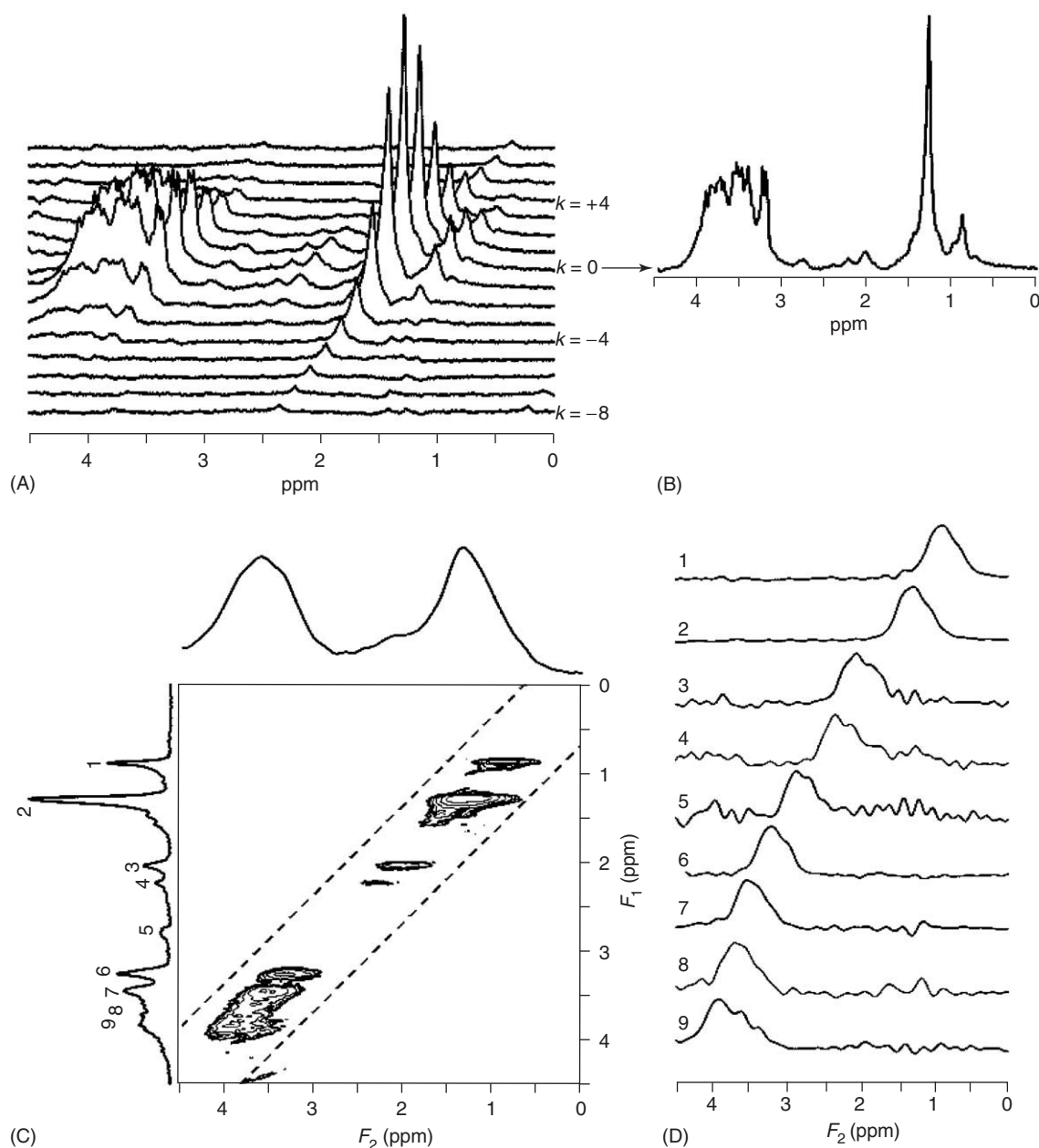
$$\text{FID}(t_2, t_1) = \exp(-i\omega_{\text{iso}}t_1)\text{FID}(t_2) \quad [4]$$

After Fourier transformation as a function of  $t_2$  and then as a function of  $t_1$  a pure absorption-mode 2D spectrum is obtained. **Figure 4C** shows an example of water-suppressed  $^1\text{H}$  PHORMAT metabolite spectra of excised rat liver tissue, obtained at a MAS spinning frequency of 1 Hz and in a 7 T field. **Figure 4C** displays the 2D plot together with the projections along the isotropic  $F_1$  ( $t_1$ ) and anisotropic  $F_2$  ( $t_2$ ) dimensions. By making slices parallel to the  $F_2$  axis, the anisotropic line shapes of each isotropic peak can be determined separately, nine of which are plotted in **Figure 4D**. In this way the susceptibility gradients surrounding individual metabolites can be determined, which could be of diagnostic value. The spectral line widths in the isotropic projection are comparable to, albeit slightly larger than, the widths obtained with 4 kHz SP-MAS (**Figure 1D**) and 40 Hz PASS (**Figure 4B**). As will be discussed in the next section, this slight increase (a few hertz) is attributed to the diffusion of the metabolites in the susceptibility gradients.

## Limitations of SP-MAS, PASS, and PHORMAT

The question arises to what values the MAS frequency can be reduced with the various MAS methodologies, as this will determine the size and type of biological object that can be studied with these techniques. The answer depends on the MAS methodology used and the NMR properties of the magnetic nuclei under investigation. Moreover, the NMR sensitivity per unit measuring time and the total measuring time itself are different for the various techniques. **Figure 5** shows 300 MHz  $^1\text{H}$  SP-MAS, PASS center-band, and PHORMAT isotropic spectra of a mixture of water and spherical glass beads with diameters of  $\sim 230 \mu\text{m}$  as a function of the spinning frequency  $F$ . The susceptibility difference between the beads and water, broadened the static water line to 3.7 kHz (12.5 ppm), at least a factor 20 larger than the line widths observed in biological systems. For SP-MAS it follows that  $F$  has to be at least of the order of the static line width in order to reduce the SSBs. In contrast, it follows from **Figure 5** that PASS produces a nearly sideband-free isotropic chemical

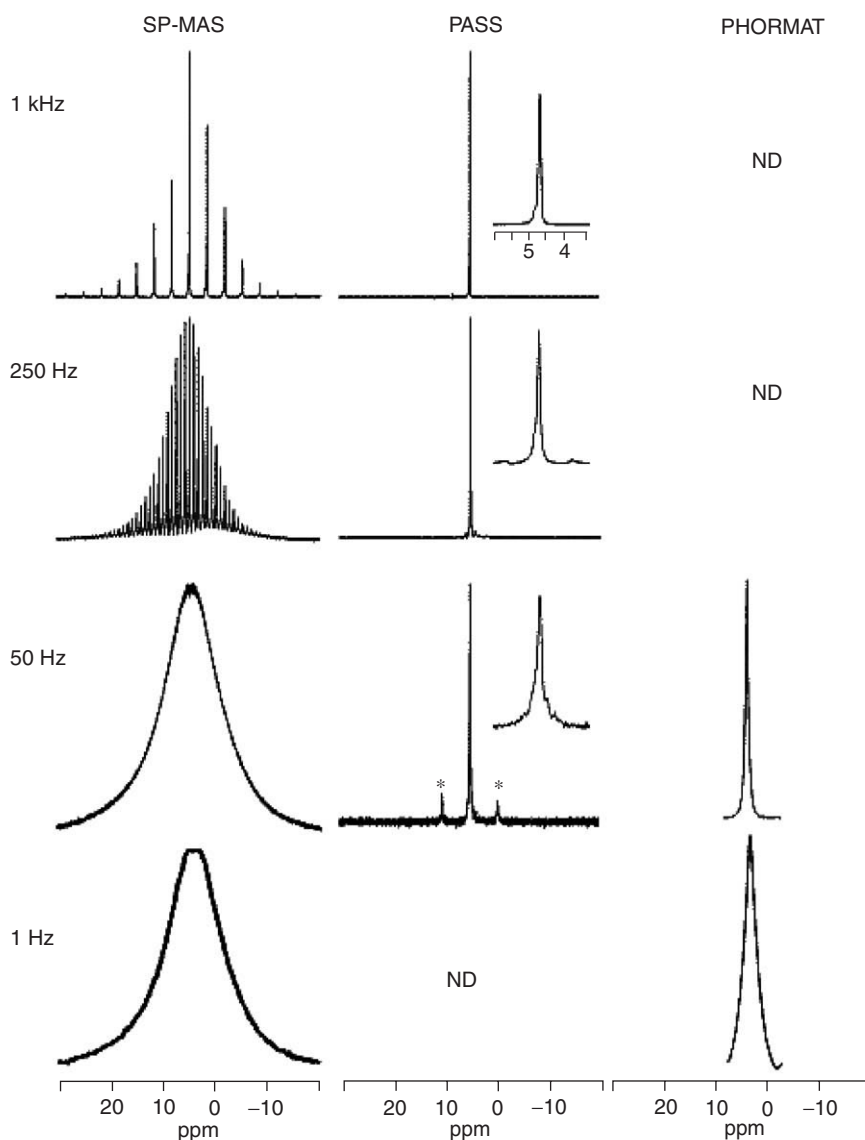




**Figure 4** Top: The water suppressed <sup>1</sup>H 2D-PASS spectra of freshly excised rat liver acquired at a spinning rate of 40 Hz. The external field was 7 T. (A) The stacked 2D plot. The parameter  $k$  denotes the  $k$ th sideband;  $k = 0$  corresponds to the center band; (B) The center-band spectrum. Bottom: The water suppressed <sup>1</sup>H PHORMAT spectra of a freshly excised rat liver sample, acquired at a sample spinning rate of 1 Hz. The external field was 7 T; (C) The contour plot of the 2D PHORMAT spectrum along with its anisotropic ( $F_2$ ) and isotropic ( $F_1$ ) projections. The  $F_1$  projection was obtained by summing only the data inside the dotted box; (D) The stacked plot of the anisotropic line shapes corresponding to nine isotropic peaks, which were obtained by taking a slice parallel to the  $F_2$  axis at the center of each isotropic peak ( $F_1$ ). ((A) and (B) are reproduced with permission from Wind RA and Hu JZ (2003) Magnetic susceptibility effects in nuclear magnetic resonance spectroscopy of biological objects. In: *Recent Research Developments in Magnetism and Magnetic Materials*, vol. 1, pp. 147–169. (C) and (D) are reproduced with permission from Hu JZ, Rommereim DN, and Wind RA (2002) High-resolution <sup>1</sup>H NMR spectroscopy in rat liver using magic angle turning at a 1 Hz spinning rate. *Magnetic Resonance in Medicine* 47: 829–836.)

shift spectrum at a spinning frequency as low as 50 Hz. (The peaks marked by the symbol “\*” in Figure 5 are aliased sidebands, arising because only 32 evolution increments were used to acquire the

PASS data, whereas at this spinning rate the spectrum contained  $\sim 90$  visible sidebands, requiring at least 90 increments.) Moreover, the isotropic line width observed at 50 Hz is similar to that measured at



**Figure 5**  $^1\text{H}$  MAS spectra obtained at different spinning rates by SP-MAS, PASS, and PHORMAT on a mixture of  $\text{H}_2\text{O}$  and spherical glass beads with diameters of  $230 \pm 20 \mu\text{m}$ . The PASS spectra are the center-band spectra, the PHORMAT spectra are the isotropic projections. (Reprinted with permission from Hu JZ and Wind RA (2002) The evaluation of different MAS techniques at low spinning rates in aqueous samples and in the presence of magnetic susceptibility gradients. *Journal of Magnetic Resonance* 159: 92–100; © Elsevier.)

higher spinning frequencies. However, at this speed the signal is seriously attenuated. This is caused by the decay of the magnetization during the rotor period  $T_r$  prior to the acquisition (cf. Figure 3B). During this period the magnetization dephases in the transverse plane, and is partially refocused by the  $180^\circ$  pulses. Hence, the time constant governing the decay is approximately given by the intrinsic spin-spin relaxation time  $T_2$ , which means that the minimum spinning frequency should be larger than  $(T_2)^{-1}$ , i.e., the isotropic line width, in order to avoid serious signal losses. This has as a consequence that for biological samples, where the minimum

metabolite  $T_2$  values are of the order of 30 ms, with PASS spinning frequencies of  $\sim 40$  Hz or larger should be used.

The  $T_2$  attenuation is avoided in a PHORMAT experiment. Here, the magnetization is stored parallel to the main field direction with a maximum duration of  $2/3$  times the rotor period, which means that the spinning frequency has to be large compared with the spin-lattice relaxation rate  $(T_1)^{-1}$  of the spins rather than  $(T_2)^{-1}$  in order to avoid signal attenuation. Hence, for the water/bead mixture, where the water  $T_1$  is several seconds, the spinning speed can be made as low as 1 Hz without causing

serious signal attenuation, and the same is true for biological samples, where the metabolite  $T_1$  values are at least an order of magnitude larger than the  $T_2$  values. However, it follows from Figure 5 that at ultralow spinning speeds the isotropic line width increases. This is due to the diffusion of the water molecules in the susceptibility gradients. This diffusion-induced broadening, which cannot be eliminated with MAS techniques, is proportional to  $G\sqrt{D/F}$ , where  $G$  is the susceptibility gradient,  $D$  is the diffusion coefficient, and  $F$  is the MAS frequency. It follows that this broadening is much less in biological samples, where both  $G$  and  $D$  are at least an order of magnitude less than in the water/bead mixture. Comparing the SP-MAS, PASS, and PHORMAT experiments of the excised rat liver shown above, it was estimated that the diffusion-induced line broadening in a 1 Hz PHORMAT experiment is  $\sim 2$  Hz at 7 T, considerably less than the intrinsic line width.

Finally, it is worth noting that compared to PASS, PHORMAT has a considerably lower NMR sensitivity and often requires a longer measuring time. The sensitivity loss is mainly due to an intrinsic loss of a factor of 4 resulting from the use of two storage pulses  $p_1$  and  $p_2$  in the PHORMAT sequence, cf. Figure 3C (the storage pulse  $p_3$  is omitted in a regular PHORMAT experiment). Also, PHORMAT often requires a relatively large number of evolution steps, resulting in a long measuring time, an hour or more. Although the performance of PHORMAT can be improved, e.g., by applying multiple-echo acquisition, PASS should be considered as the method of choice if the sample can tolerate spinning speeds of tens of hertz. PHORMAT should be reserved for research in larger biological samples, including *in vivo* applications.

### In Vivo PHORMAT

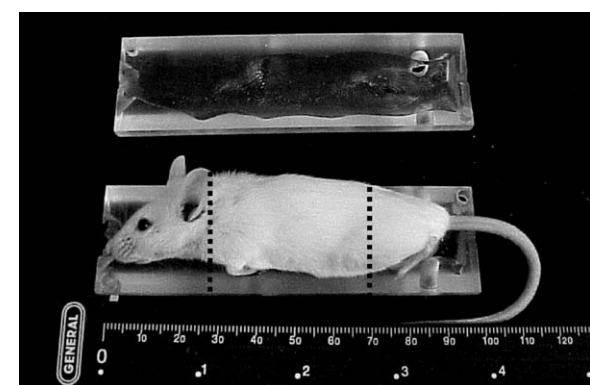
Two effects associated with spinning a live animal can cause harm: the centrifugal forces induced in the animal, and the effects induced by the external magnetic field, rendered partially time-dependent from the rotating animal's point of view when spinning inside the magnet. This time-dependent field  $dB/dt$  can cause nerve and cardiac stimulations in a similar way as pulsed field gradients. However, both effects can be small in a PHORMAT experiment. For instance, when a mouse is placed in a cylinder with a 1 cm radius, and rotated around the magic angle in a 2 T field and with a spinning speed of 1 Hz, the maximum centrifugal force at the periphery of the animal is 0.04 times the gravitational force  $F_g$ , about two orders of magnitude below the forces applied in

chronic centrifuging experiments, and  $dB/dt$  is  $\sim 10 \text{ T s}^{-1}$ , well below the threshold of  $\sim 90 \text{ T s}^{-1}$  for which nerve stimulations have been reported. In fact, 15 mice were spun in a 2 T field at frequencies up to 8 Hz and for durations up to 60 min without causing any apparent short-term or long-term health effects. Hence, it is possible to use ultraslow PHORMAT for *in vivo* studies on live animals. Figure 6 shows the first result of such an experiment, obtained on the middle section of the body of an (anesthetized) female BALBc mouse between the dotted lines shown in Figure 6A and the arrows shown in Figure 6B. Figures 6C and 6D shows the  $^1\text{H}$  85 MHz water-suppressed metabolite spectra obtained in a stationary mouse and a mouse subjected to 1.5 Hz MAS, respectively. Even in this relatively low field a significant increase in spectral resolution is obtained.

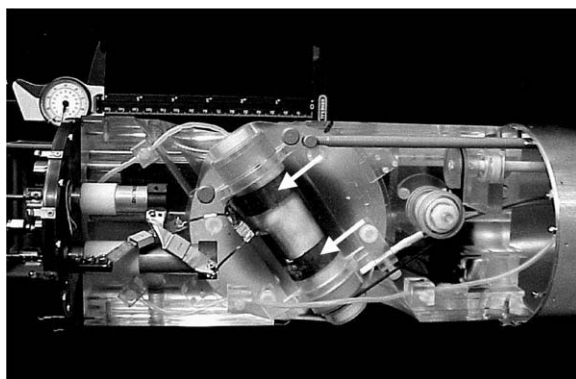
### Future Perspectives

It can be concluded that it is possible to significantly increase the resolution in the proton NMR metabolite spectra in intact biological samples by PASS and PHORMAT. It was found that for small samples, with sizes of a few millimeter or less, where spinning speeds of 40 Hz or more can be tolerated, PASS is the method of choice because of its superior sensitivity and short measuring time. For larger biological samples, including animals, ultraslow-MAS PHORMAT, allowing spinning speeds as low as 1 Hz, has to be used. In a 7 T field with PASS the spectral line widths are reduced by an order of magnitude or more to values determined by the intrinsic  $T_2$ , originating from the various spin-spin dipolar interactions, rendered time-dependent by the molecular motions. With 1 Hz PHORMAT slightly larger isotropic line widths are observed, resulting from the molecular diffusion in the susceptibility gradients. Increasing the spinning speed, if allowed, will reduce this contribution. In external fields larger than 7 T the line width reductions are expected to become even larger, as the residual isotropic line widths are essentially field-independent, whereas the susceptibility broadening increases linearly proportional to the field. Hence, with PASS or PHORMAT the full benefits of high-field NMR are obtained.

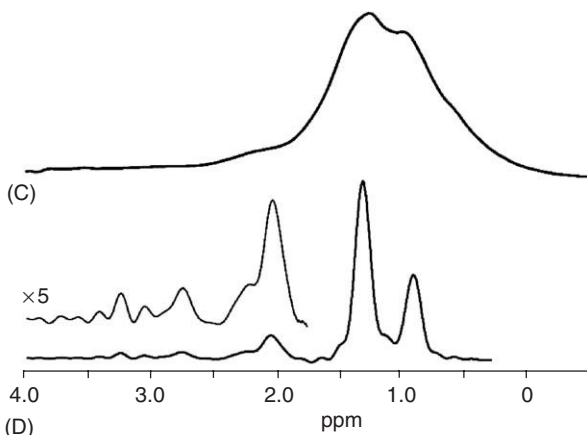
It is worth noting that the slow-MAS methodology is a new area of research, and that several further improvements are necessary to make PASS and PHORMAT more viable methods for biochemical and biomedical research. Particularly, specific combinations of RF and pulsed-field-gradient sequences need to be developed so that PASS or PHORMAT spectra can be obtained of a select volume in the sample or the animal rather than the whole sample or



(A)



(B)



**Figure 6** (A, B) The mouse-MAS NMR probe: (A) an anesthetized female BALBc mouse placed in a mouse-shaped cavity in one-half of the rotor; (B) top part of the probe with the mouse and the mold mounted in place. The area between the arrows is the NMR-sensitive area of the NMR coil; (C) The anisotropic ( $F_2$ ) projection of the 2D PHORMAT spectrum, obtained on the part of the mouse body between the dotted lines shown in (A); (D) The isotropic ( $F_1$ ) projection of the 2D PHORMAT spectrum, obtained on the same part of the mouse body. ((A) and (B) are reproduced with permission from Wind RA, Hu JZ, and Rommereim DN (2003) High resolution  $^1\text{H}$  NMR spectroscopy in a live mouse subjected to 1.5 Hz magic angle spinning. *Magnetic Resonance in Medicine* 50: 1113–1119. (C) and (D) are reproduced with permission from Wind RA and Hu JZ (2003) Magnetic susceptibility effects in nuclear magnetic resonance spectroscopy of biological objects. In: *Recent Research Developments in Magnetism and Magnetic Materials* vol. 1, pp. 147–169.)

the main part of the body. If these implementations are successful, it can be expected that PASS and PHORMAT will significantly increase the utility of proton metabolite NMR spectroscopy for biochemical and biomedical studies in cells, tissues, organs, and animals.

Finally, the question arises whether the slow-MAS approach will ultimately reach the hospital to investigate patients. Obviously in this case PHORMAT is the only candidate. Although spinning a patient at a frequency of (e.g., 1 Hz) may cause no physical harm, it is likely that it will induce unacceptable stress in many patients. An alternative approach might be to rotate the external magnetic field instead of the patient, either mechanically or electronically, or by rotating both the magnetic field and the patient in opposite directions. In this way the spinning speed of the patient can be reduced. The future will tell whether this approach will become a viable option.

**See also:** Nuclear Magnetic Resonance Spectroscopy: Overview; Principles. Nuclear Magnetic Resonance Spectroscopy Applications: Pharmaceutical. Nuclear Magnetic Resonance Spectroscopy Techniques: Solid-State.

## Further Reading

- Andrew ER (1996) Magic angle spinning. In: Grant DM and Haris RK (eds.) *Encyclopedia of Magnetic Resonance*, pp. 2891–2901. New York: Wiley.
- Antzakin ON (1999) Sideband manipulation in magic-angle-spinning nuclear magnetic resonance. *Progress in NMR Spectroscopy* 35: 203–266.
- Boxerman JL, Weisskopf RM, and Rosen BR (2000) Susceptibility effects in whole body experiments. In: Young IR (ed.) *Biomedical Magnetic Resonance Imaging and Spectroscopy*, pp. 654–661. New York: Wiley.
- Callaghan PT (1993) *Principles of Nuclear Magnetic Resonance Microscopy*. Oxford: Clarendon Press.
- Gadian DG (2000) Animal methods in MRS. In: Young IR (ed.) *Biomedical Magnetic Resonance Imaging and Spectroscopy*, pp. 898–904. New York: Wiley.
- Hu JZ, Wang W, and Pugmire RJ (1996) Magic angle turning and hopping. In: Grant DM and Haris RK (eds.) *Encyclopedia of Magnetic Resonance*, pp. 2914–2921. New York: Wiley.
- Mukherji SK (ed.) (1998) *Clinical Applications of MR Spectroscopy*. New York: Wiley.
- Slichter CP (1990) *Principles of Magnetic Resonance*, pp. 392–406. Berlin: Springer.
- Smith ICP and Bezabeh T (2000) Tissue NMR *ex vivo*. In: Young IR (ed.) *Biomedical Magnetic Resonance Imaging and Spectroscopy*, pp. 891–897. New York: Wiley.
- VanderHart DL (1996) Magnetic susceptibility & high resolution NMR of liquids & solids. In: Grant DM and Haris RK (eds.) *Encyclopedia of Magnetic Resonance*, pp. 2938–2946. New York: Wiley.

# NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY TECHNIQUES

Contents

**Nuclear Overhauser Effect**

**Multidimensional Proton**

**Solid-State**

**Surface Coil**

**In Vivo Spectroscopy Using Localization Techniques**

## Nuclear Overhauser Effect

**M P Williamson**, The University of Sheffield, Sheffield, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The nuclear Overhauser effect (NOE) is a powerful nuclear magnetic resonance (NMR) tool for structure elucidation because it can be used to identify pairs of nuclei that are close together, typically within 0.5 nm (5 Å). The size of the effect is strongly dependent upon the internuclear separation  $r$ ; in optimal cases, it is proportional to  $r^{-6}$ . However, as outlined below, complicating factors often reduce this dependence. We therefore stress throughout that structural analyses that rely on exact quantitative measurements should be avoided, and the NOE is best used in a semiquantitative manner.

The NOE is manifested as a change in the intensity of one resonance when a nearby resonance is partially or fully saturated. The behavior of the NOE differs markedly depending on the size of the molecule studied. In small molecules, the NOE builds up slowly, and is positive (i.e., saturation of one resonance leads to an increase in the intensity of a neighboring one), whereas in large molecules, the NOE builds up much more quickly and is negative. Moreover, in large molecules the NOE can be passed on from one nucleus to its neighbor in a process known as spin diffusion, in a way that does not happen in small molecules. The way that the NOE is applied therefore differs depending on the size of the molecule studied.

### Theory

Each NMR-active nucleus is a dipole: that is, it behaves like a small magnet with north and south

poles. A nuclear dipole has a strong effect on the local magnetic field seen by its neighbor. This effect is orientation dependent and is known as dipolar coupling, and is responsible for the very large linewidths seen in NMR spectra of solids. In solution, dipolar coupling is normally averaged out by molecular tumbling. (There is, however, an increasingly important field of research in which molecules are persuaded to adopt a small degree of residual order, in order to keep some residual dipolar couplings, which can be used very effectively to give orientational information.) The molecular tumbling means that dipolar fields fluctuate widely at rates governed by the tumbling rate of the molecule. The important consequence is that molecular tumbling is a source of nuclear spin relaxation, and is indeed the major source of relaxation in many cases.

The NOE is best understood as a competition between different relaxation pathways. In a system consisting of two spins  $I$  and  $S$ , perturbation of the  $z$  magnetization of  $S$  ( $S_z$ ) away from its equilibrium value  $S_z^0$  results in a change in the  $I$  spin magnetization given by

$$dI_z/dt = -\rho_{IS}(I_z - I_z^0) - \sigma_{IS}(S_z - S_z^0) \quad [1]$$

This equation shows that perturbation of  $S_z$  affects the  $z$  magnetization of  $I$  at a rate  $\sigma_{IS}$ , known as the cross-relaxation rate. In contrast to this,  $I_z$  relaxes back to its equilibrium magnetization at a rate  $\rho_{IS}$ , known as the dipolar longitudinal relaxation rate (which is similar to the spin-lattice relaxation rate  $1/T_1$ ).

The rate constants  $\sigma$  and  $\rho$  are given by

$$\sigma_{IS} = (1/20)K^2[6J(\omega_I + \omega_S) - J(\omega_I - \omega_S)] \quad [2]$$

$$\rho_{IS} = (1/20)K^2[J(\omega_I - \omega_S) + 3J(\omega_I) + 6J(\omega_I + \omega_S)] \quad [3]$$

where  $K = (\mu_0/4\pi)\hbar\gamma_I\gamma_S r_{IS}^{-3}$  and  $J(\omega) = 2\tau_c/(1 + \omega^2\tau_c^2)$ ,  $\mu_0$  is the permeability constant in a vacuum, and  $(\mu_0/4\pi)$  is simply a normalization constant for SI

units;  $\hbar$  is Planck's constant divided by  $2\pi$ ;  $\gamma_I$  and  $\gamma_S$  are the gyromagnetic ratios of  $I$  and  $S$ ;  $r_{IS}$  is the internuclear distance;  $J(\omega)$  is called the spectral density function, and characterizes the power available for relaxation as a function of frequency;  $\omega$  is the frequency of the transition; and  $\tau_c$  is the rotational correlation time, which is approximately the time it takes for a molecule to rotate by one radian.

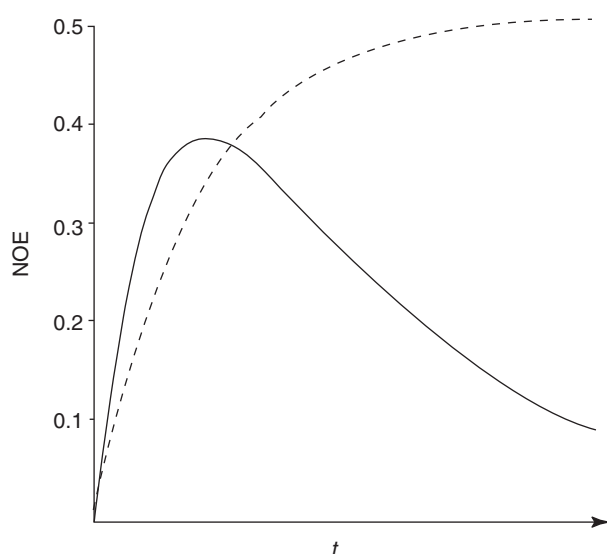
### Time Dependence of the NOE in a Two-Spin System

Equations [1]–[3] allow us to calculate how the NOE behaves in a two-spin system. As explained in more detail below, there are two main ways of observing NOEs. In the first, one spin  $S$  is perturbed, for example, by a selective  $180^\circ$  pulse: almost all two-dimensional (2D) methods work in a similar way. This results in the buildup of an NOE at adjacent spins  $I$  (Figure 1), with an initial rate given simply by

$$f_I\{S\}(t) = 2\sigma_{IS}t = \zeta r_{IS}^{-6}t \quad [4]$$

where the factor of 2 comes from the inversion of  $S$ , and  $\zeta$  is a constant containing all the nondistance-dependent terms in eqn [2]. (A similar equation, but without the factor of 2, applies for cross-peaks in 2D NOESY experiments.) This allows us to write down a very simple way to measure distances. If there is a known fixed distance in the molecule  $r_{\text{ref}}$ , which has an NOE intensity  $f_{\text{ref}}(t)$ , then any unknown distance can be calculated as

$$r_{IS} = r_{\text{ref}}[f_I\{S\}(t)/f_{\text{ref}}(t)]^{-6} \quad [5]$$



**Figure 1** The time dependence of NOE build-up for an ideal two-spin system with  $\omega\tau_c \ll 1$ , following inversion of  $S$  (solid line) or continuous saturation of  $S$  (dashed line).

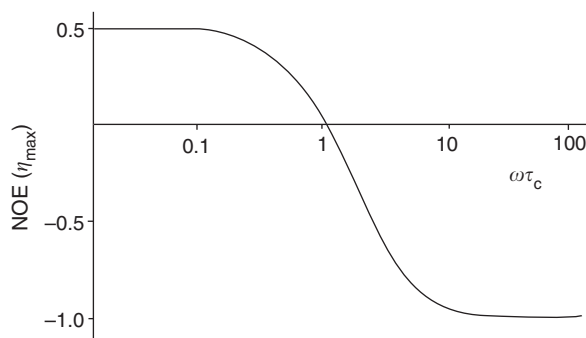
This method is widely used to measure internuclear distances, particularly in macromolecules. It is simple to apply, and does not require very accurate measurement of NOE intensities, because of the sixth root, which reduces the error in the calculated distance. However, it is only accurate at very short times after the initial perturbation of  $S$ , as described below, and so is usually applied with wide error margins, by imposing distance restraints as ranges rather than exact distances.

The second method for measuring NOEs, which is only used for studies in small molecules, is to saturate spin  $S$  continuously. The NOE at  $I$  builds up to a steady state (Figure 1), given by

$$f_I\{S\} = (I_z - I_z^0)/I_z^0 = (\gamma_S/\gamma_I)\sigma_{IS}/\rho_{IS} \quad [6]$$

### Dependence on Molecular Tumbling Rate

Equations [2] and [3] show that dipolar relaxation depends strongly on the molecular tumbling rate, because of its dependence on  $J(\omega)$ .  $\rho$  increases steadily with  $\tau_c$ . The magnitude of  $\sigma$  also increases with  $\tau_c$ , implying that for small molecules, where  $\tau_c$  is small, the rate at which the NOE builds up is very slow, while NOEs build up faster in large molecules. However, the sign of  $\sigma$  changes as  $\tau_c$  increases, because  $J(\omega_I + \omega_S)$  becomes less efficient as molecular tumbling slows, whereas  $J(\omega_I - \omega_S)$  becomes more efficient. The maximum achievable steady-state NOE is calculated by substituting eqns [2] and [3] into eqn [6], and is shown in Figure 2, which indicates that the steady-state NOE is positive for small molecules, goes through zero at  $\omega\tau_c \sim 1.12$ , and then becomes negative for large molecules. Note that for very large molecules, where  $\omega\tau_c \gg 1$ , the steady-state NOE is  $-1$ . This implies that saturation of one spin  $S$  will eventually lead to an indiscriminate loss of signal from all the other spins in very large molecules, by a process known as spin diffusion. Although spin diffusion can sometimes be useful, for example, in the



**Figure 2** The dependence of NOE intensity on  $\omega\tau_c$  for an ideal two-spin system following steady-state irradiation of  $S$ .



saturation transfer experiment for mapping intermolecular interfaces, it means that steady-state NOE experiments in large molecules give no useful distance information. This is one major reason why transient experiments such as 2D NOESY and ROESY are almost universally used for large molecules (the other being that selective saturation is nearly impossible in large molecules with crowded spectra).

In real situations, other spins and other sources of relaxation are present, which will increase spin-lattice relaxation rates, and therefore reduce the size of the NOE seen. It is therefore necessary to modify the equations: for example, eqn [6] becomes

$$f_I\{S\} = (\gamma_S/\gamma_I)\sigma_{IS}/(\rho_{IS} + \rho_I^*) \quad [7]$$

where  $\rho_I^*$  represents any other sources of relaxation, which are usually unknown. A common source of relaxation is provided by unpaired electrons, found in transition metal ions and also in dissolved molecular oxygen.  $\rho_I^*$  is often independent of molecular size (whereas  $\rho_{IS}$  is small in small molecules). Consequently, it is a more severe problem when studying NOEs in small molecules, in which cases the size of the NOE can often be increased by removing unpaired electrons, for example, by degassing the solution. Even with such precautions, the NOEs seen in small molecules are often much smaller than predicted by 'ideal' equations, such as eqn [6]. It also means that, for small molecules, with slow NOE build-up rates and more significant external relaxation pathways, transient NOE experiments are slow and give only small NOEs, and it is therefore common to use steady-state experiments. As the molecule gets larger, cross-relaxation rates increase, which makes 2D experiments more attractive.

### The Rotating Frame NOE

Figure 2 shows that there is a molecular tumbling rate ( $\omega\tau_c \sim 1$ ) for which the NOE is very small. Typically, this will occur for molecules with relative molecular masses of  $\sim 1000$ , although the relationship between molecular mass and correlation time is not easily predictable, and depends on other factors such as solvent viscosity and temperature. The rotating frame NOE experiment was developed in order to avoid this difficulty, since in this experiment the NOE is always positive, in fact becoming larger as the correlation time increases. It has a further advantage in that spin diffusion is much less of a problem than in conventional NOE experiments, and it is therefore used widely for intermediate sized molecules, often in its 2D form, known as ROESY or HOHAHA. For large molecules, it is slightly less sensitive and more prone to artifacts than NOESY,

but can still be a useful alternative particularly where spin diffusion is thought to be important.

### Multispin Systems

So far we have only considered two-spin systems. We now consider the effect of additional spins, and again it is convenient to divide the discussion into (1) large molecules, with  $\omega\tau_c \gg 1$ , studied using NOESY, and (2) small molecules, with  $\omega\tau_c < 1$ , studied using 1D steady-state techniques.

**Large molecules: spin diffusion** We consider the effect of a third spin ( $X$ ) on the buildup of the NOE between two protons  $I$  and  $S$ , as would be observed in a transient or 2D NOE experiment. The buildup can be calculated by adding an extra term to eqn [1], which is then solved either by numerical integration of the equation, or by matrix algebra, the so-called 'relaxation matrix' approach. The results show that when  $S$  is inverted, and  $X$  is distant from  $I$ , no effect is seen on the buildup. In a linear arrangement  $S-I-X$ , as  $X$  gets closer to  $I$ , it helps to relax  $I$ , and so the NOE at  $I$  decays sooner than it would otherwise do. At the same time, an NOE is seen building up at  $X$ . On a naive 'two-spin' interpretation, this would appear to imply that  $X$  is closer to  $S$  than it really is. This is the effect commonly known as spin diffusion, in which the NOE 'diffuses' outwards from the site of irradiation; it makes the interpretation of NOE intensities complicated, especially at longer NOE build-up times when the NOE has had more time to diffuse, and is a major problem in the quantitative interpretation of NOE intensities. For a known geometry, the effect can be calculated exactly, but for an unknown geometry, it is hard to spot. The effects of spin diffusion are normally minimized by keeping the NOE build-up time as short as practicable.

**Small molecules** In small molecules (and in ROESY for any size of molecule) spin diffusion also occurs, but its consequences are altered because the third spin ( $X$  in the linear  $S-I-X$  example) now receives a negative NOE, by contrast to the normal positive NOE seen at  $I$ . This is known as a three-spin effect, and is much easier to spot than spin diffusion in large molecules, because of the change in sign of the NOE. In addition, because the maximum positive NOE is  $+0.5$  (Figure 2), the magnitude of the three-spin effect can never be bigger than  $0.5^2 = 0.25$  and is usually much smaller, and the subsequent (positive) four-spin effect (e.g.,  $S$  to  $Y$  in  $S-I-X-Y$ ) is very rarely observed. By contrast, four-spin spin diffusion pathways in large molecules are readily observed.

If the  $S-I$  distance is kept fixed, and  $X$  is rotated around  $I$  to become closer to  $S$ , it starts to receive a

direct (positive) NOE in addition to the indirect negative NOE. For certain geometries, it is thus possible for the  $X\{S\}$  NOE to be very small, even if the  $S-X$  distance is quite short. Once again, distance calculations using the NOE are seen to be not straightforward, even though they may be strongly dependent on distance! It should be noted, however, that NOEs can, in principle, be calculated exactly, for any number of spins (given all the appropriate motional and relaxation parameters, which in practice are not experimentally accessible). The difficulties inherent in using the NOE to derive distances have been emphasized here because of the widespread over-interpretation of NOEs to produce quite erroneous and unrealistically precise distances.

### Distance Calculation

**Large molecules** As outlined above, in principle the NOE at short build-up time is directly proportional to  $r^{-6}$ . However, at times short enough to be certain of the proportionality (i.e., such that spin diffusion should not have got very far), the NOE may be so small as to be almost unobservable. One must therefore work at slightly longer build-up times (of the order of 100 ms), and accept some error in the derived distance, particularly for weaker NOEs. This is achieved by specifying NOE-derived distances as ranges, rather than single distances, with the size of the range larger at longer distance. The reference distance is best obtained from known intramolecular distances of similar distances to the target: for example, by calibrating the strongest sequential  $C\alpha H-NH$  NOE in a protein to 0.23 nm, the shortest possible distance for such an NOE.

**Small molecules** Small molecules are generally studied using steady-state techniques. The resulting intensity effects can therefore involve several spins, and can be quite complex. For example, steady-state NOEs are not symmetrical – in other words, the NOE at A on saturation of B is not equal to the NOE at B on saturation of A, since the NOEs depend on the other neighbors of the spins observed. It is still usually true that NOEs to nearby protons are large, while NOEs to distant protons are small, but this is not universally true; for example, an NOE to a distant but isolated proton can be very large, as it has no competing relaxation pathways to reduce the NOE. Analytical equations can be written down for any given spin system, however complex, but rapidly become unwieldy. In practice, a complete analysis is seldom attempted, and the normal practice is to limit analysis to a qualitative survey, as described below.

A general problem with the derivation of distances from NOEs is intramolecular motion. Motion can

both alter the local correlation time and also alter internuclear distances. Because the cross-relaxation rate is  $r^{-6}$ -dependent, motions that allow two nuclei to become close together some of the time give rise to a disproportionately large NOE. Indeed, local motions can result in patterns of NOEs that are incompatible with any single static structure. This is another reason for caution in the quantitative interpretation of NOEs.

### Heteronuclear NOEs

The dipole–dipole cross-relaxation mechanisms that give rise to homonuclear NOEs can just as easily produce heteronuclear NOEs. The heteronuclear NOE is often used to increase the intensity of the heteronuclear signal. For example, saturation of protons during acquisition of  $^{13}C$  spectra gives rise to an NOE on carbon of just under 2 (eqn [6]). Thus, the net intensity of carbon signals in small molecules is up to three times greater than it would be without the NOE.

As in the homonuclear case, the NOE is reduced by competing relaxation mechanisms. If a carbon atom has a proton attached, its relaxation is dominated by dipole–dipole relaxation from the proton, whereas for quaternary carbons, dipolar relaxation is slower, and competes less effectively with other relaxation pathways, often producing a reduced NOE. The size of the NOE also depends on the correlation time, and, like the homonuclear NOE, changes markedly around  $\omega\tau_c \approx 1$ . Most heteronuclear NOEs become very small for large molecules; the only common exception is the  $^{19}F\{^1H\}$  NOE, which behaves very similarly to the  $^1H\{^1H\}$  NOE. Nuclei with negative gyromagnetic ratio, such as  $^{15}N$ , have negative NOEs. This can lead to loss of signal if the NOE is only partial and is close to  $-1$ . The properties of the more common nuclei are listed in Table 1.

The heteronuclear NOE can be used as a structural tool, by irradiating specific protons individually and measuring NOEs to nearby heteronuclei. Because

**Table 1** NOE characteristics of common nuclei

Nucleus	Gyromagnetic ratio, $\gamma$ ( $10^7 \text{ rad T}^{-1} \text{ s}^{-1}$ )	Maximum $f_X\{^1H\}^a$	
		$\omega\tau_c \ll 1$	$\omega\tau_c \gg 1$
$^1H$	26.75	50	–100
$^{13}C$	6.73	199	15.4
$^{15}N$	–2.71	–494	–21.7
$^{19}F$	25.18	53	–104
$^{31}P$	10.84	123	2.2

<sup>a</sup>Maximum % NOE obtainable in a steady-state experiment, listed for the extremes of small and large molecules.

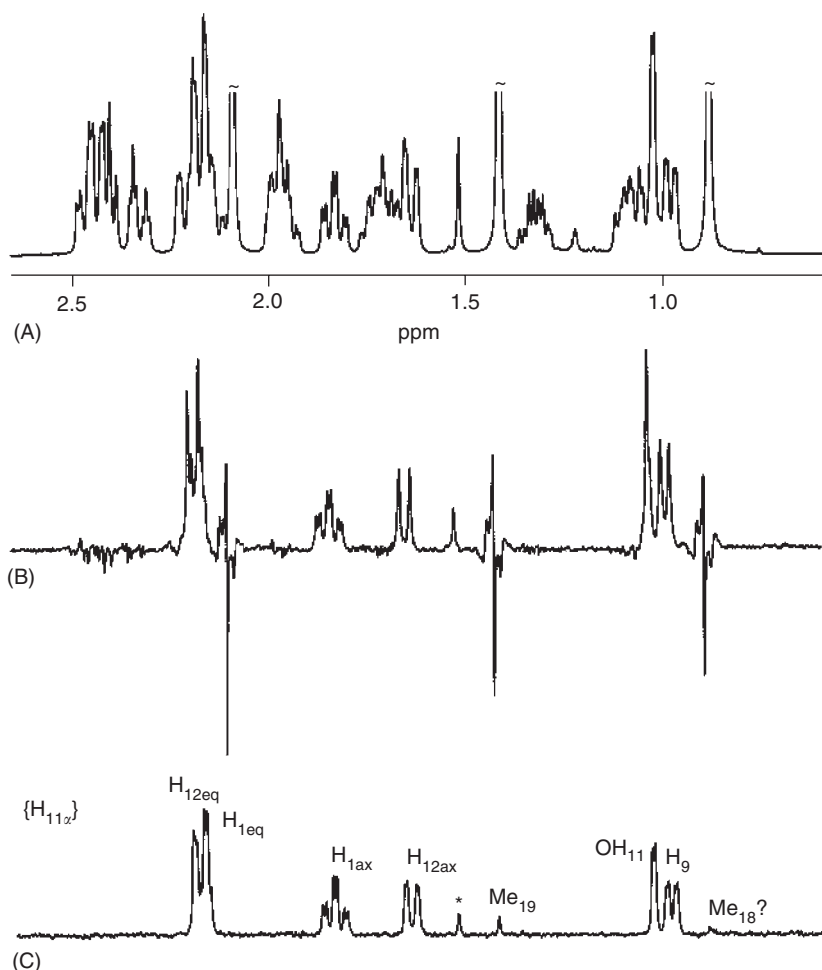
heteronuclear relaxation is dominated by directly bonded protons, the only structurally useful heteronuclear NOEs are to atoms with no protons attached.

Heteronuclear NOEs are also widely used to characterize mobility. In particular, the  $^{15}\text{N}\{^1\text{H}\}$  NOE is very small for  $\omega\tau_c \gg 1$ , but goes to almost  $-5$  for fast tumbling. (The literature for heteronuclear NOEs is confusing, in that it is common to describe the NOE as relative intensity rather than relative enhancement, which results in NOE values larger by one than the enhancements used normally.  $^{15}\text{N}\{^1\text{H}\}$  NOEs on this definition range from  $-3.9$  to  $0.8$ , rather than  $-4.9$  to  $-0.2$ .) Proteins generally have  $\omega\tau_c \gg 1$  for all backbone N-H vectors, giving rise to small NOE enhancements. However, if the backbone is locally mobile, the NOE becomes

more negative. Measurement of the NOE provides a simple measure of local mobility, and is often combined with measurements of  $^{15}\text{N}$   $T_1$  and  $T_2$  for a more quantitative assessment.

## NOE Experiments

For large molecules, the 2D NOE experiment NOESY is universally used. It is easy to set up and analyze, and very reliable. The only parameter that requires thought is the mixing time, during which the NOE builds up. Shorter values (e.g., 50 ms) reduce spin diffusion, but also reduce sensitivity. Times longer than  $\sim 200$  ms are not useful for large molecules, because spin diffusion becomes too severe, and sensitivity gain is not significant.



**Figure 3** NOE experiments recorded using a  $20\text{ mmol l}^{-1}$  sample of  $11\beta$ -hydroxyprogesterone in  $\text{CDCl}_3$  using a Bruker DRX500 spectrometer.  $H_{11z}$  was irradiated. (A) Conventional 1D spectrum. (B) Standard steady-state difference spectrum, 5 s saturation time. Note the subtraction artifacts, particularly at the intense methyl singlets at 0.8, 1.4, and 2.1 ppm. (C) Double pulsed field gradient spin echo NOE spectrum, 1 s NOE evolution period. The spectrum is essentially free of artifacts, and permits the observation of an enhancement at  $Me_{19}$  and possibly  $Me_{18}$  that could not reliably be detected in the conventional experiment. The signal marked \* is an intermolecular enhancement to  $\text{H}_2\text{O}$  dissolved in the  $\text{CDCl}_3$  presumably transmitted via exchange from  $OH_{11}$ . (Reproduced with permission from Stott K, Ph.D. Thesis. University of Cambridge, UK.)

For medium-sized molecules (molecular weight 500–1000), the ROESY experiment is most common. This again is easy to set up, though it does require slightly more care because the continuous power during the NOE build-up time leads to sample heating and can damage the hardware if applied at too high a power.

The choice of experiments is most difficult with small molecules. The standard experiment is the steady-state NOE, in which specific low-power saturation is applied for 5 s or more, and a control spectrum is subtracted. The experiment can be speeded up with little loss in sensitivity (but a reduction in quantitative information) by reducing the saturation time to 1–2 s. This is a difference experiment, and always contains experimental artifacts, resulting from incomplete subtraction of reference signals. A few years ago, a gradient-enhanced experiment was proposed, which has lower absolute sensitivity but much smaller artifacts, and is therefore gradually taking over as the experiment of choice. It requires a spectrometer with reliable pulsed field gradient hardware, but once set up, is as simple and reliable as steady-state experiments, and produces exceptionally clean NOE spectra, allowing the observation of very small enhancements (Figure 3).

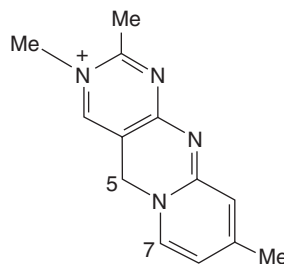
## Use in Structure Determination

The widespread use of the NOE in structure determination is a consequence of its strong dependence on distance; that is, NOEs are used to define approximate distances, and the distances are then used to generate structures compatible with these distances. In large molecules, the relationship between NOE intensity and distance is reasonably strong, facilitating such methods as distance geometry and simulated annealing, which directly use distance information to restrain structures. In small molecules, the NOE is more effective when giving qualitative rather than quantitative information. Thus, if the mere presence of an NOE can be used to distinguish between two structures, the NOE is a reliable tool, whereas if the size of an NOE is being used to fix the conformation of a molecule, great care is necessary, and use of other information, for example, coupling constants, is preferred.

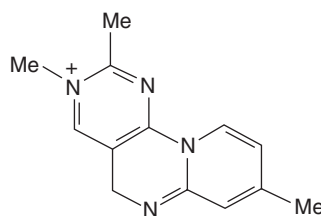
### Small Molecule Structure Determination

Substitution patterns around aromatic rings form a good subject for analysis using the NOE, since the information required is almost entirely of a qualitative nature. For example, a condensation reaction produced a compound that was expected to have

structure [I] or [II]. Saturation of the H<sub>5</sub> protons produced a strong (24%) NOE at H<sub>7</sub>, thereby identifying the compound as [I].



[I]



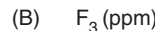
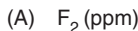
[II]

Stereochemical problems can often be solved very simply using the NOE, but only in more or less rigid systems. Thus, fused ring systems can often be analyzed in a straightforward manner, because such systems are frequently rigid enough for study using the NOE (e.g., Figure 3). The fungal neurotoxin penitrem A [III] was characterized by <sup>1</sup>H and <sup>13</sup>C NMR, and most of the covalent structure was pieced together using (<sup>13</sup>C, <sup>1</sup>H) and (<sup>13</sup>C, <sup>13</sup>C) spin-spin coupling connectivities. At this point, the NOE was used to obtain stereochemical information. This molecule is much larger than [I], but in organic solvents such as perdeuterated acetone it still tumbles rapidly enough to give positive NOEs, and has a spectrum with enough spectral dispersion that all protons can be selectively irradiated. The result of a large series of steady-state experiments is shown in the structure of [III]. The NOEs unequivocally give the relative stereochemistries of all groups in the molecule, which is only possible because of the rigid ring junctions that fix most of the internuclear distances.

Conformational details in more mobile systems are harder to derive, as they rely on accurate quantification, and require a number of assumptions to be made. A study was made of the conformation of the dienyl side chain of [IV]. On the basis of the exanomeric effect, the diene fragment would be expected to adopt one or both of the conformations [IVA] or [IVB]. Note that there is no direct evidence from NMR for these conformations; because NMR



signals are time-averaged, only a single resonance is observed for each nucleus, and all the observed NMR parameters are weighted averages. Irradiation of H<sub>1'</sub> produces a 13% enhancement at H<sub>1</sub> and a 4% enhancement at H<sub>2</sub>. This suggests that both rotamers are present, and that the major rotamer is [IVA]. Any more quantitative assessment would rely on the 'model' conformations [IVA] and [IVB] being correct, and would also require detailed measurements of relaxation rates for all the protons involved, as well as assumptions as to rotation rates and relaxation mechanisms.



**Figure 4** Part of the NOESY spectrum of the C-terminal half of rat testis calmodulin. (A) 2D NOESY spectrum, mixing time 200 ms. (B) A plane from the 3D  $^{15}\text{N}$ -resolved NOESY spectrum, taken at a  $^{15}\text{N}$  frequency of 127.04 ppm. Note the greatly improved resolution and the high signal-to-noise ratio of spectrum (B). (*Natural Product Reports* (1993) 10: 207–232; reproduced by permission of The Royal Society of Chemistry.)

$r^{-6}$  dependence of the NOE at short mixing time, and protocols for the application of distance constraints in distance geometry or simulated annealing calculations are becoming standardized. Often the biggest problem in analyzing NOESY spectra of proteins is the chemical shift degeneracy of protons, which makes it impossible to provide confident assignments for the cross-peaks. It is therefore common to label proteins with  $^{15}\text{N}$  and  $^{13}\text{C}$ , and to use the chemical shifts of the labels to help to resolve degenerate  $^1\text{H}$  signals, using 3D and 4D  $^{15}\text{N}$ - and  $^{13}\text{C}$ -resolved spectra (Figure 4). As noted in the previous section, motion is one of the hardest problems to deal with, because the observed NOEs are highly nonlinear averages.

Intermolecular NOEs have been very useful for defining the geometry of complexes, including those of protein dimers. Differential heteronuclear labeling of the two components of a dimer is often

used to distinguish intramolecular from intermolecular NOEs.

### The Transferred NOE

The transferred NOE is a technique used to obtain structural information on the conformation of a small ligand bound to a much larger receptor, such as a protein. It relies on the fact that the cross-relaxation rate for a free ligand is slow, whereas for the bound ligand it is much faster. If the ligand is in exchange between free and bound states, any NOE built up in the bound state will be transferred to the free state by exchange processes. The observed average cross-relaxation rate  $\langle\sigma_{IS}\rangle$  is

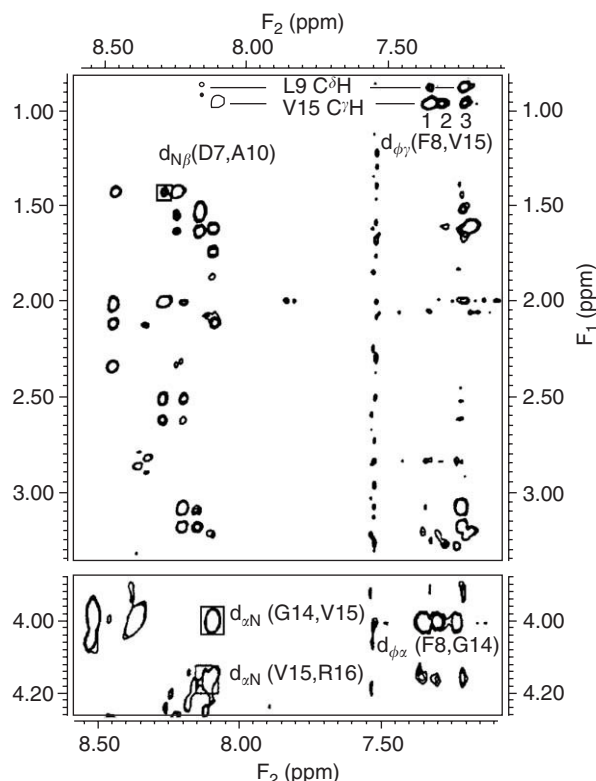
$$\langle\sigma_{IS}\rangle = P^F\sigma_{IS}^F + P^B\sigma_{IS}^B \quad [8]$$

where  $P^F$  and  $P^B$  are the relative populations of the free and bound states, and  $\sigma_{IS}^F$  and  $\sigma_{IS}^B$  are their cross-relaxation rates. Because  $\sigma_{IS}^B$  is much greater than  $\sigma_{IS}^F$ , it is possible (and usual) to have a 5- to 20-fold excess of free over bound states ( $P^F/P^B = 5\text{--}20$ ), and still observe only the cross-relaxation in the bound state. The experiment is most commonly carried out as a 2D NOESY experiment, where the intensities of NOESY cross-peaks, observed at the frequencies of the free resonances, give information on distances in the bound state (Figure 5).

**See also:** Nuclear Magnetic Resonance Spectroscopy: Principles. Nuclear Magnetic Resonance Spectroscopy-Applicable Elements: Carbon-13; Nitrogen-15. Nuclear Magnetic Resonance Spectroscopy Applications: Pharmaceutical. Peptides. Structural Elucidation.

### Further Reading

- Claridge TDW (1999) *High-resolution NMR Techniques in Organic Chemistry*. Oxford: Pergamon.
- Günther H (1995) *NMR Spectroscopy*, 2nd edn, pp. 397–407. Chichester: Wiley.
- Neuhaus D and Williamson MP (2000) *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, 2nd edn. New York: Wiley-VCH.
- Ni F (1994) Recent developments in transferred NOE methods. *Progress in NMR Spectroscopy* 26: 517–606.
- Sanders JKM and Meresh JD (1982) Nuclear magnetic double resonance: the use of difference spectroscopy. *Progress in NMR Spectroscopy* 15: 353–400.
- Takahashi H, Nakanishi T, Kami K, Arata Y, and Shimada I (2000) A novel NMR method for determining the interfaces of large protein–protein complexes. *Nature Structural Biology* 7: 220–223.
- Wüthrich K (1986) *NMR of Proteins and Nucleic Acids*. New York: Wiley.



**Figure 5** A 2D transferred NOE experiment, obtained from a mixture of thrombin and a peptide Ac-Asp<sup>7</sup>-Phe-Leu-Ala-Glu-Gly-Gly-Gly-Val-Arg<sup>16</sup>, in a molar ratio of 1:12, using a mixing time of 200 ms. Most of the cross peaks are intraresidue or sequential; some longer range NOEs are indicated, which show that the bound peptide is bent, with a hydrophobic cluster including Phe<sup>8</sup>, Leu<sup>9</sup>, Gly<sup>14</sup>, and Val<sup>15</sup>. The peptide alone gives very weak NOEs under these conditions, confirming that the NOEs seen are transferred from the thrombin-bound peptide. (Reprinted with permission from *Biochemistry* (1989) 28: 3094; © ACS.)



## Multidimensional Proton

J C Lindon, Imperial College London, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Modern nuclear magnetic resonance (NMR) spectroscopy is carried out in the pulse-Fourier transform mode. This is achieved by applying a short intense pulse of radiofrequency (RF) power to a sample and monitoring the decaying voltage induced in a receiver coil as nuclei excited by this RF pulse relax back to equilibrium. This signal, known as free induction decay (FID), is related to the conventional NMR spectrum by the Fourier transform. This type of NMR spectroscopy is termed one-dimensional (1D) NMR because the signal is collected for a single acquisition time and this signal, after Fourier transformation, results in a spectrum of signal intensity as a function of a single frequency axis.

In many molecular systems such as proteins and other macromolecules, or in multicomponent mixtures such as biofluids, the spectral complexity and signal overlap in 1D NMR spectra can be too great for simple assignment of NMR resonances. Under these circumstances, it would be desirable to improve the dispersion of the signals. One approach is to simplify the complex 1D NMR spectrum to leave only resonances that are amenable to interpretation by the use of special pulse sequences. These can edit the original spectrum into subsets of data, or excite or detect only specific types of resonances; for example, only those that are spin-spin coupled to a given resonance. These latter experiments use longer, lower power pulses termed selective pulses. An alternative approach is to achieve this dispersion is through the use of 2D or higher dimensional ( $n$ D) methods. The pulse sequences that form the basis of such multidimensional NMR approaches not only enable increased spectral dispersion, but can also have the added advantage of giving information on relationships between the various resonances in the spectrum; for example, showing which ones are coupled to each other by spin-spin coupling.

There are numerous examples of the application of multidimensional NMR spectroscopy in disparate areas of study, and some idea of the breadth of such NMR experiments can be gleaned from the Specialist Periodical Report on NMR spectroscopy from the UK's Royal Society of Chemistry.

### Principles of 2D NMR

As an alternative to the specific solutions that are available in 1D NMR, the general approach adopted in 2D NMR is to apply a series of RF pulses to the sample such that there are two independent variable time intervals in the pulse sequence. One of these is the acquisition time, denoted by  $t_2$ , and the other is some incremental delay denoted  $t_1$ . If an NMR FID is acquired for a period  $t_2$  for each of a set of  $t_1$  values, the digital NMR signal intensity ( $S$ ) will be a function of both  $t_1$  and  $t_2$  giving a matrix of data  $S(t_1, t_2)$ . If Fourier transformation is carried out with respect to both  $t_1$  and  $t_2$  a matrix of NMR intensity as a function of two frequencies will result. This is now a 2D NMR spectrum as it represents signal intensity as a function of two frequency axes. The delay  $t_1$  and the actual nature of the pulse sequence will, as may be seen later, define exactly what the two frequency axes will represent. As a consequence of this process, there are two immediate benefits: increased signal dispersion and hence less overlap, and if the experiment is conducted in certain ways, it is possible to evaluate connectivities between the various NMR resonances. On the other hand, the 2D approach usually requires increased experiment time and increased computer data storage requirements.

Adding more independent incremental delays to the pulse sequence followed by corresponding Fourier transformation steps leads to further increases in frequency dimensionality to yield 3D NMR, 4D NMR, and so on. A number of 3D and 4D experiments have found use, particularly in the assignment of the complex spectra that arise from proteins and carbohydrates.

### Definition of a 2D NMR Experiment

A 2D NMR experiment contains up to five separate elements. These are:

1. A preparation period, consisting of RF pulses and possibly delays, after which the nuclei are in some defined nonequilibrium spin state.
2. An evolution period of length  $t_1$  during which the properties of the nuclei such as chemical shifts or coupling constants are allowed to evolve at their appropriate frequencies. It is this period that is incremented in the 2D NMR experiment.
3. A mixing period that can comprise further RF pulses and delays during which information is

transferred between the nuclei in specific ways. This mixing period will have a different effect on the final detected signal according to the value of  $t_1$ . Not all 2D NMR experiments require a mixing period.

4. A detection period of length  $t_2$  during which the FID is acquired.
5. A relaxation period to ensure that the spin system is at equilibrium before the next cycle of acquisition.

The actual procedure is to obtain successive FIDs each for a time period of  $t_2$  using  $n_2$  data points for each increment in the value of  $t_1$ , with the first value being very short and increasing in equal steps for a total of  $n_1$  steps. Thus, the data matrix consists of ( $n_1 \times n_2$ ) data points. After data acquisition, the  $n_1$  FIDs are Fourier transformed in the normal way to give  $n_1$  spectra in a data matrix  $S(t_1, \omega_2)$ , one for each value of the evolution period  $t_1$ . The data are then transposed by taking corresponding points in each spectrum (i.e., all of the first points in the original  $n_1$  spectra are put into a new representation to give a spectrum of  $n_1$  points, this process being repeated until there will be  $n_2$  such spectra). These transposed spectra are called interferograms and the data matrix is now  $S(\omega_2, t_1)$ . A Fourier transformation with respect to  $t_1$  gives the final 2D NMR spectrum  $S(\omega_2, \omega_1)$ .

The FIDs and interferograms can be multiplied by appropriate mathematical functions before Fourier transformation in order to improve sensitivity, resolution, or line shape exactly as in 1D NMR. The NMR data are usually obtained as two separate components  $90^\circ$  out of phase from each other in a mode called quadrature detection to yield two spectral components denoted real and imaginary. A suitable linear combination of these two components allows the generation of a spectrum with a pure absorption phase. In 2D NMR, this phase-sensitive detection sometimes results in line shapes and phases that do not allow a pure absorption spectrum to be obtained and then the compromise of a magnitude mode presentation is used. This comprises the square root of the sum of the squares of the real and imaginary components and is wholly positive. Usually, 2D NMR spectra are plotted as contour maps as though the 2D spectral peaks are a series of mountains viewed from above relative to the orthogonal  $\omega_1$  and  $\omega_2$  axes.

There are basically two types of 2D NMR experiments. These are termed (1) resolved and (2) correlated, according to the type of dispersion in the second frequency dimension  $\omega_1$ . Resolved experiments are a way of improving spectral dispersion

by rotating the appearance of one NMR parameter at right-angles to another so that the  $\omega_1$  and  $\omega_2$  axes correspond to different parameters. The commonest version of this is to cause the rotation of all spin-spin coupled multiplets by  $90^\circ$ , thus producing only chemical shifts along the  $\omega_2$  axis and at each chemical shift the spin-spin coupled multiplet is spread along the  $\omega_1$  axis. Generally, no information is available from a resolved experiment on the spin-spin coupling connectivity between resonances. The other class of proton 2D NMR experiment, the correlated type, provides information on connectivity between resonances such as those that have a common spin-spin coupling or arise from nuclei that are close together in space. In these cases, for identical nuclei, the  $\omega_1$  and  $\omega_2$  axes are identical.

There have been many multidimensional NMR experiments published for homonuclear  $^1\text{H}$ - $^1\text{H}$  studies and for  $^1\text{H}$ -heteronuclear studies as well as others involving heteronuclei only. Many are variations of general schemes for specific applications and only a few 2D NMR experiments have become widely used. The following sections discuss these principal 2D NMR experiments and the reader is referred to the Further Reading section for specific details of the theory and practical aspects of multidimensional NMR.

There are further classes of experiment that result in pseudo-2D NMR spectra. These do not have a second frequency axis resulting from Fourier transformation of a variable time, but the second axis is some other parameter. One example is provided by continuous-flow directly coupled HPLC-NMR spectra where the second axis in the pseudo-2D plot is the chromatographic retention time. Another example is diffusion-ordered NMR spectroscopy where the second axis plots the molecular diffusion coefficient associated with each NMR peak, this parameter being derived from the dependence of peak intensities on the square of an applied magnetic field gradient.

## Resolved Experiments

The principal pulse sequence in the resolved experiment category for high-resolution solution-state NMR is the  $J$ -resolved experiment (JRES). It is based on the spin-echo pulse sequence in which the delay between the  $90^\circ$  and  $180^\circ$  pulses is incremented as the evolution period  $t_1$ , and so the pulse sequence is

$$-\text{RD} - 90^\circ - t_1 - 180^\circ - \text{ACQ}(t_2) -$$

where RD is a relaxation delay and ACQ( $t_2$ ) is the data acquisition period. The spin-echo sequence leads to the elimination of the effects of chemical shifts during the period  $t_1$ , but spin-spin couplings cause a phase modulation of the NMR signals to a degree that is dependent on the length of  $t_1$ . The two Fourier transform steps therefore produce a 2D NMR spectrum with both chemical shifts and spin-spin couplings on the  $\omega_2$  axis at their normal frequency positions, but only the couplings appear on the  $\omega_1$  axis. In fact, the pulse sequence results in multiplets that appear at an angle of  $45^\circ$  relative to the axes and a representation is usually shown in which these multiplets have been further tilted by  $45^\circ$  to make them orthogonal to the  $\omega_2$  axis. Under these circumstances, the projection of the spectrum on to the  $\omega_2$  axis produces a  $^1\text{H}$  NMR spectrum consisting of singlets at each  $^1\text{H}$  chemical shift, i.e., a fully proton-decoupled proton NMR spectrum. The results are not so clear-cut if the spin system is not first order; in this case additional lines appear at the average  $\omega_2$  position of the strongly coupled nuclei. However, the JRES experiment can be a very useful aid to the assignment of resonances in the complex spectra that arise from mixtures of small molecules such as biofluids where many of the individual components give first-order spectra. Also, because the JRES experiment is based on the spin-echo sequence, any nuclei with short  $T_2$  values will relax within the  $t_1$  evolution period and not contribute to the final spectrum. This has proved useful for the analysis of blood plasma by proton NMR as the broad, short  $T_2$  resonances from proteins and lipoproteins are greatly suppressed. Any heteronuclear couplings such as those between protons and fluorine nuclei will not be rotated into the  $\omega_1$  axis as the lines act as though they are chemically shifted. These couplings therefore appear separated on the  $\omega_2$  axis. In the IRES experiment, the spectrum is calculated as the magnitude presentation after the second Fourier transform because the tilted phase-sensitive spectrum has zero intensity on the  $\omega_1$  axis. In order to overcome the adverse line shape that results in magnitude mode spectra, appropriate resolution enhancement is usually carried out in the time domain before Fourier transformation. There is a corresponding heteronuclear experiment, usually for  $^{13}\text{C}$  NMR, where the  $^{13}\text{C}$ - $^1\text{H}$  coupling patterns for each resonance are rotated into the  $\omega_1$  dimension.

The various proton 2D NMR experiments are illustrated using dexamethasone dissolved in DMSO- $d_6$  solution and measured at an observation frequency of 600 MHz. Dexamethasone has the molecular structure and numbering system given below.

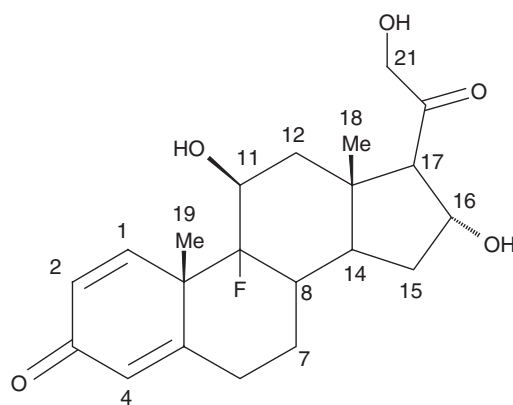


Figure 1 shows the partial 600 MHz  $^1\text{H}$  JRES NMR spectrum of dexamethasone in the region from  $\sim\delta 0.7$  to  $\delta 3.0$ . This expansion excludes the resonances from the olefinic hydrogens H1, H2, and H4, the signals from the deshielded aliphatic hydrogens H11a, H21, and H21', and the resonances from OH protons. In the 2D NMR contour plot, the multiplet nature of each resonance is resolved at right angles to the  $\omega_2$  axis. Some artifacts can be seen, particularly close to the large peaks. These probably arise from the tilting and symmetrization manipulation of the data. Plotted above the 2D spectrum is the projection of the peaks on to the  $\omega_2$  axis to give a resonance at the chemical shift of each proton. It should be noted that these projections show a wide variation in intensity that cannot be directly related to the number of protons contributing to the multiplet.

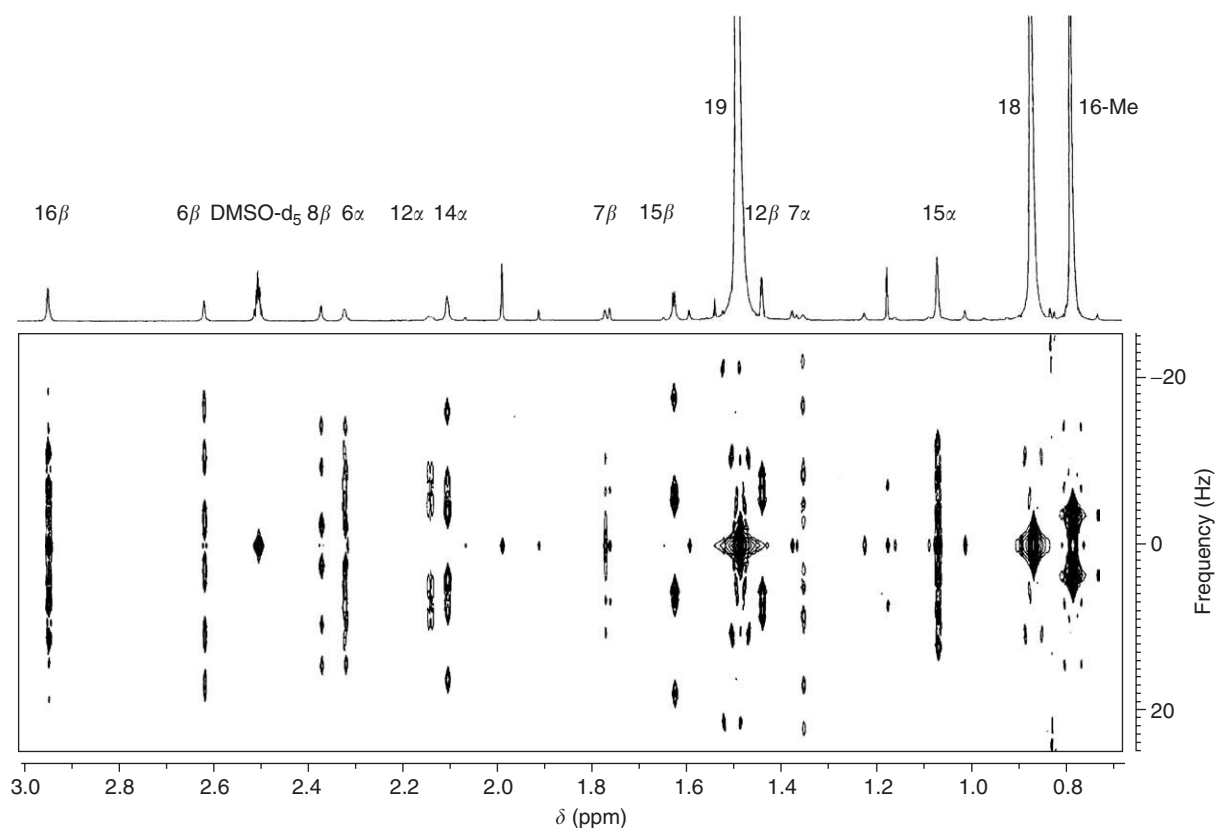
## Correlation Experiments

### Resonance Correlation via Spin-Spin Coupling

The proton homonuclear experiment, termed COSY (COReletion Spectroscopy), serves to show which resonances in a proton spectrum have mutual spin-spin couplings. The method works for a wide range of coupling constant values and can be carried out in magnitude mode or phase-sensitive mode. It has largely replaced the conventional 1D NMR single-frequency spin-decoupling approach because it gives all connectivities in one experiment, albeit at the expense of longer acquisition times and a requirement for large data matrices. The pulse sequence is

$$-\text{RD} - 90^\circ - t_1 - 90^\circ - \text{ACQ}(t_2) -$$

and the resulting spectrum has the conventional 1D NMR spectrum along the diagonal and off-diagonal cross-peaks at chemical shifts corresponding to pairs of coupled nuclei. The phases of the RF pulses are changed in a systematic fashion in concert with the phase of the receiver in order to suppress unwanted



**Figure 1** Partial 600 MHz  $^1\text{H}$   $J$ -resolved NMR spectrum of dexamethasone in  $\text{DMSO-d}_6$ . The region from  $\delta 0.7$  to  $\delta 3.0$  is shown. This expansion excludes the resonances from the olefinic hydrogens H1, H2, and H4, the deshielded aliphatic hydrogens H11a, H21, and H21', and the resonances from OH protons. The assignments are as marked.

artifacts and features in the spectrum. As stated above, the experiment is sensitive to a wide range of spin couplings and may result in cross-peaks even in situations where the mutual coupling, although present, cannot be resolved. It is also useful for discovering connectivities to resonances buried in a complex overlap of many bands. To overcome problems of line shape in the 2D NMR spectrum, especially on older spectrometers, it was conventional to produce a magnitude mode spectrum because the phase-sensitive detected signals had a phase-twist line shape from which in-phase absorption signals could not be obtained. This magnitude calculation results in an all-positive spectrum. In order to overcome the adverse line shape effects on resonances near the diagonal peaks, the COSY-45 variation was introduced:

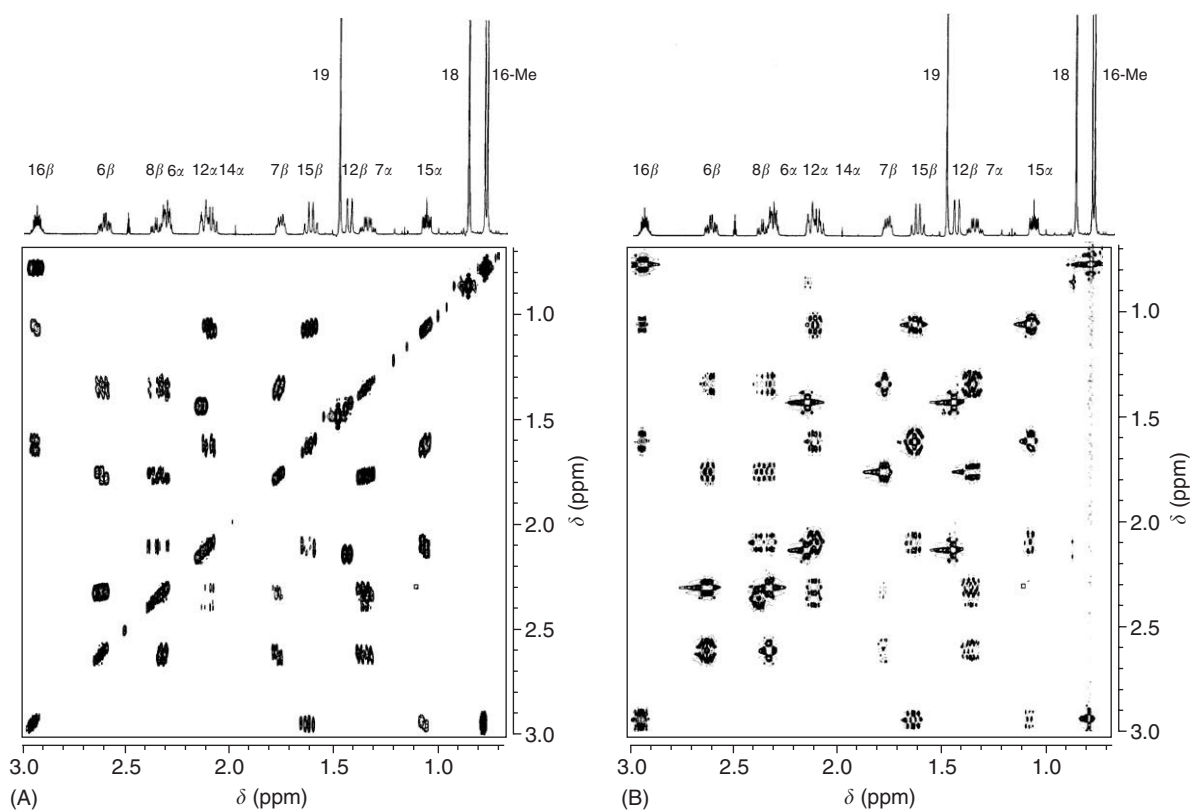
$$-\text{RD} - 90^\circ - t_1 - 45^\circ - \text{ACQ}(t_2) -$$

The reduction in the second pulse width has the principal effect of reducing the intensity of the diagonal relative to the cross-peaks and leaving the cross-peaks with a simpler structure (for three-spin systems or greater). This means that assignments

in crowded regions of the spectrum or close to the diagonal become easier.

However, on modern spectrometers, the COSY experiment is now carried out in phase-sensitive mode through the use of one of two modifications introduced to achieve quadrature detection in the  $t_1$  axis, as well as in  $t_2$  which is done conventionally. These two approaches are called the States-Haberkorn method and time proportional phase incrementation; readers are referred to the Further Reading section for details.

**Figure 2A** shows the COSY-45 2D NMR spectrum of dexamethasone acquired using magnetic field gradients rather than phase cycling (see later). This spectrum is plotted using a magnitude-mode presentation. Again only the region between  $\delta 0.7$  and  $\delta 3.0$  is shown and the assignments are given on the 1D NMR spectrum plotted across the top at the same expansion. A cross-peak can be observed at the chemical shifts of each pair of coupled protons. A modification to the COSY sequence called COSYLR is possible, which is better suited to elucidating connectivities through small- or long-range spin couplings. This is achieved by introducing an additional fixed delay within the evolution period  $t_1$ .



**Figure 2** (A) Partial 600 MHz  $^1\text{H}$ – $^1\text{H}$  COSY-45 NMR spectrum of dexamethasone in  $\text{DMSO-d}_6$  acquired using magnetic field gradients and presented in magnitude mode. The region from  $\delta 0.7$  to  $\delta 3.0$  is shown. (B) Partial 600 MHz  $^1\text{H}$ – $^1\text{H}$  DQF-COSY NMR spectrum of dexamethasone in  $\text{DMSO-d}_6$  acquired using magnetic field gradients and presented in phase-sensitive mode. The region from  $\delta 0.7$  to  $\delta 3.0$  is shown. The vertical bands are common artifacts (called  $t_1$  noise) which appear on high intensity peaks and arise as a consequence of the method of data acquisition.

### Double Quantum Filtered COSY

The double quantum filtered COSY (DQF-COSY) approach causes the suppression of diagonal peaks arising from singlets (i.e., resonances without any spin couplings, or nuclei not part of a multiple spin or multiple quantum system) and allows the cross-peaks to be in pure absorption mode. This simplifies complex spectra and improves resolution. The pulse sequence is modified from the COSY sequence by the inclusion of a short fixed time delay  $\Delta$  as shown:

$$-\text{RD} - 90^\circ - t_1 - 90^\circ(\phi_1) - \Delta - 90^\circ(\phi_2) - \text{ACQ}(t_2) -$$

The third  $90^\circ$  pulse filters out signals that are not spin–spin coupled, i.e., those that cannot be part of a double or higher quantum system. The term  $\phi$  is used to denote that the phases of the  $90^\circ$  pulses are changed in a systematic way to produce this singlet resonance suppression by appropriate addition and subtraction of the received signal on successive scans and the fixed delay  $\Delta$  (10–20  $\mu\text{s}$ ) is included to allow time for the necessary phase switching.

Figure 2B shows the 600 MHz DQF-COSY spectrum of dexamethasone, now acquired and presented

in phase-sensitive mode, with positive contours given full lines and negative contours given dotted lines. Again cross-peaks are observed between coupled protons, but the singlets from the methyl groups are now absent from the spectrum and the coupling patterns of the cross-peaks can be used to determine which is the actual coupling (i.e., that between the protons giving rise to the cross-peak) and passive couplings (i.e., from either of the coupled protons to other protons).

Refinements of the technique can be used which eliminate all resonances from one- and two-spin systems, i.e., singlets and doublets (triple-quantum filtered COSY), and so on for higher-spin systems. The DQF-COSY experiment is not as sensitive as the normal COSY approach and the higher quantum filtering is usually achieved using pulse phase cycling with increments in phase angles of less than  $90^\circ$ , a facility not always available on older spectrometers.

### Total Correlation Spectroscopy

This very powerful experiment, called TOCSY, provides information on unbroken chains of coupled

protons in one 2D NMR spectrum. The experiment is sometimes also called the homonuclear Hartmann–Hahn experiment. The results are not very dependent on the magnitude of the spin–spin couplings involved in the spectrum and all of the peaks are in phase, in absorption mode. The pulse sequence is

$$-\text{RD} - 90^\circ - t_1 - [\text{spin lock}] - \text{ACQ}(t_2) -$$

The method relies on the application of a spin-lock field, which causes the nuclear magnetizations to precess about this RF field, i.e., the spins are said to be locked to this field. During this period, transfer of magnetization occurs between coupled spins. The longer the spin-lock period is applied, the further the magnetization will be transferred down a chain of coupled nuclei. The most widely used spin-lock sequence is based on a train of pulses called MLEV-17. The pulse sequence is identical to that for the ROESY experiment for measuring nuclear Overhauser effects (NOEs) (see later) and it is possible for NOE effects to appear as cross-peaks, which may cause confusion. However, there are means of minimizing these effects.

## 2D Nuclear Overhauser Effect NMR

This is a very useful 2D NMR approach that provides connectivities based on the NOE. The NOE is an alteration in signal intensity based on a direct through-space mechanism that is distance dependent, and thus the experiment can be used to probe internuclear distances. The 2D NMR NOE experiment is called NOESY and has the pulse sequence:

$$-\text{RD} - 90^\circ - t_1 - 90^\circ - t_m - 90^\circ - \text{ACQ}(t_2) -$$

In addition to the usual evolution time,  $t_1$ , this sequence contains a fixed mixing time  $t_m$  during which the NOE information is generated. The 2D NOESY spectrum consists of a diagonal, plus cross-peaks at the chemical shifts which demonstrate an NOE.

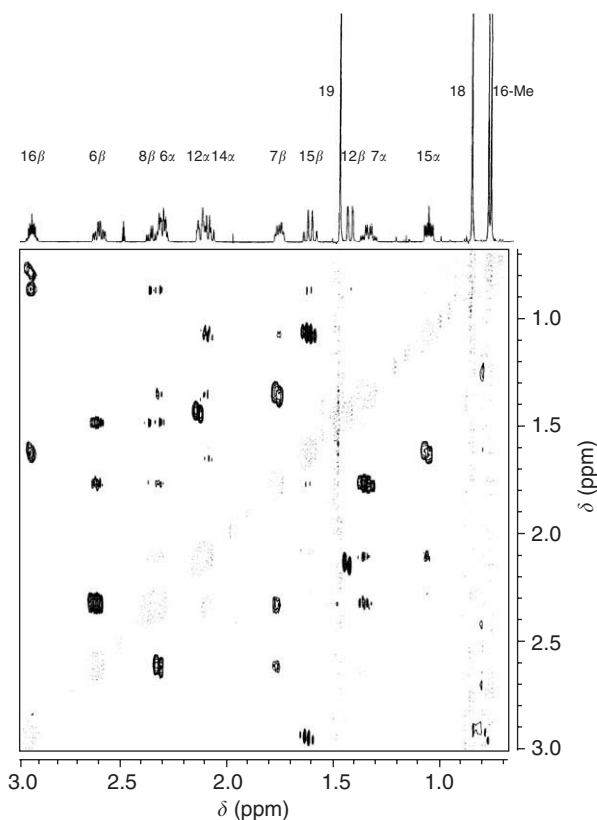
One of the problems with this experiment is that both chemical exchange and NOEs cause transfer of magnetization, and so any peaks involved in chemical exchange that is slow on the NMR timescale (i.e., those components that give rise to separate peaks) will also show cross-peaks in the 2D NOESY spectrum. On the other hand, this can be useful for studying complex exchanging systems. One way of partially separating NOE from chemical exchange cross-peaks is to carry out the experiment in phase-sensitive mode. In this mode chemical exchange effects give negative peaks and NOEs can give rise to both positive and negative peaks according to the tumbling rate of the molecule in solution and the Larmor frequency of the nuclei. The 2D NOESY

experiment is the major tool in the determination of the detailed 3D structure of proteins.

By varying the value of the mixing time  $t_m$ , it is possible to derive information on the rate at which the NOE builds up. This can give more detailed information on the internuclear distances.

The magnitude and sign of a proton–proton NOE is dependent on both the molecular tumbling rate and the NMR observation frequency. When using high-field NMR spectrometers, for molecules in the molecular mass range  $\sim 1000$  Da, NOEs can be close to zero, even if the nuclei are close together in space. One way of overcoming this limitation is to carry out a ROESY experiment, named because the NOEs are generated in the rotating frame. The pulse sequence is identical to the TOCSY sequence given earlier for connectivity via spin–spin coupling with the mixing time  $t_m$  now defined as the spin-lock period. This is usually applied for longer but at a lower power than for TOCSY studies. The distinction of NOE cross-peaks from other effects in ROESY spectra is not always easy, but complicating effects can sometimes be distinguished by the different peak phases that result.

Figure 3 shows the 600 MHz 2D ROESY NMR spectrum of dexamethasone. In this presentation, the



**Figure 3** Partial 600 MHz  $^1\text{H}$ – $^1\text{H}$  ROESY NMR spectrum of dexamethasone in  $\text{DMSO-d}_6$  acquired with a mixing time of 200 ms. The region from  $\delta 0.7$  to  $\delta 3.0$  is shown.

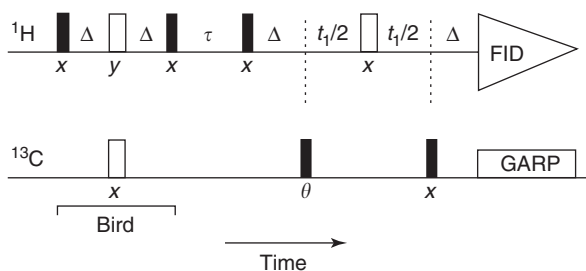


negative diagonal is shown with dotted contour lines and the positive cross-peaks are depicted with full line contours. A mixing time of 200 ms was used. Cross-peaks are observed at the chemical shifts of protons that experience a direct through-space dipolar interaction that is usually strongly distance dependent, thus indicating which hydrogens are close in space.

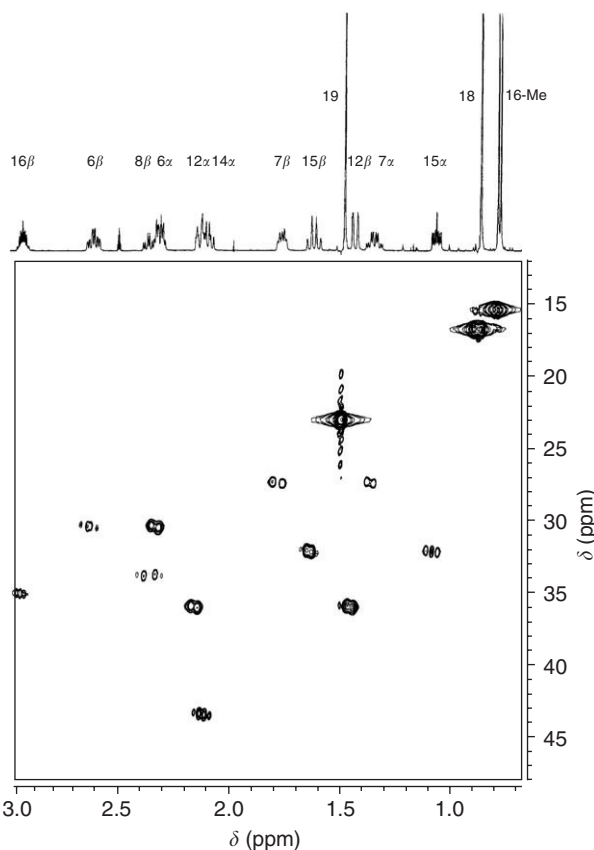
### Proton–Heteronuclear Chemical Shift Correlation

Correlation between  $^1\text{H}$  chemical shifts and heteronucleus chemical shifts such as  $^{13}\text{C}$  or  $^{15}\text{N}$  can be achieved through a number of 2D NMR experiments in which the heteronucleus is detected directly such as INEPT or DEPT. However, it is also possible to obtain the same type of correlation but with the advantage of the detection (i.e., the  $t_2$  dimension) at the much more sensitive  $^1\text{H}$  nucleus. This type of approach is called inverse detection. One problem is that  $^1\text{H}$  NMR signals from protons attached to the 99% of naturally abundant  $^{12}\text{C}$  nuclei have to be suppressed, leaving only the  $^1\text{H}$  signals from protons attached to the 1% of carbon nuclei which are  $^{13}\text{C}$  at natural abundance. This can be readily achieved, and if the pulse sequence also includes broadband heteronucleus decoupling (e.g., covering the  $^{13}\text{C}$  chemical shift range) then a 2D NMR presentation is possible in which each  $^1\text{H}$ – $^{13}\text{C}$  correlation peak appears as a singlet, the connectivity being based on  $^1\text{H}$ – $^{13}\text{C}$  spin coupling. This presentation, unlike the homonuclear correlation experiments, has no diagonal peak and the axes correspond to the appropriate chemical shift ranges of the  $^1\text{H}$  and  $^{13}\text{C}$  nuclei.

One main experimental approach is termed heteronuclear multiple quantum coherence, or HMQC, and the result is achieved by making use of the fact that coupled  $^1\text{H}$  and  $^{13}\text{C}$  nuclei can experience magnetization effects involving both spins, i.e., multiple quantum effects, which evolve during the evolution period  $t_1$ . The pulse sequence is given in Figure 4. An explanation of how this pulse sequence works is beyond the scope of this article but is available in standard texts. The scheme labeled BIRD is called a bilinear rotation decoupling sequence and has the effect, by optimizing the delay  $\tau$ , of removing  $^1\text{H}$  resonances of protons attached to  $^{12}\text{C}$  nuclei. The delay  $\Delta$  is set to the reciprocal of twice the one-bond  $^1\text{H}$ – $^{13}\text{C}$  spin coupling constant. To efficiently decouple all of the  $^{13}\text{C}$  nuclei from the protons, special pulsed decoupling schemes have been developed to cover the wide range of frequencies required (called WALTZ and GARP). Another related pulse sequence is termed heteronuclear single quantum coherence and this gives analogous information, but with better



**Figure 4** The NMR pulse sequence for the  $^1\text{H}$ -detected (inverse detection),  $^1\text{H}$ – $^{13}\text{C}$  correlation 2D NMR spectrum (HMQC). The initial part of the sequence, labeled BIRD, is a pulse scheme for removing from the spectrum the  $^1\text{H}$  resonances of protons attached to  $^{12}\text{C}$  nuclei. The filled symbols denote  $90^\circ$  pulses and the open symbols denote  $180^\circ$  pulses.  $x$ ,  $y$ , and  $\theta$  refer to RF pulse phases. The delay  $\Delta$  is set to  $1/(2 \times J_{\text{CH}})$ . GARP denotes a pulse method of spin decoupling all  $^{13}\text{C}$  nuclei.



**Figure 5** Partial 600 MHz  $^1\text{H}$ – $^{13}\text{C}$  HMQC NMR spectrum of dexamethasone in  $\text{DMSO-d}_6$  acquired using magnetic field gradients. Only the  $^1\text{H}$  region from  $\delta 0.7$  to  $\delta 3.0$  and the  $^{13}\text{C}$  region from 12 to 48 ppm are shown.

sensitivity. Analogous pulse sequences exist based on long-range  $^1\text{H}$ – $^{13}\text{C}$  couplings can then give connectivity information for quaternary carbons.

Figure 5 shows the 600 MHz  $^1\text{H}$ – $^{13}\text{C}$  HMQC spectrum of dexamethasone. Again an expansion of the spectrum is shown which corresponds to the earlier figures. In this case, the horizontal ( $\omega_2$ ) axis

represents the  $^1\text{H}$  NMR chemical shifts and the vertical ( $\omega_1$ ) axis corresponds to the  $^{13}\text{C}$  chemical shifts. A cross-peak is observed where the chemical shifts of directly bonded C–H nuclei intersect. It should be noted that where a  $\text{CH}_2$  group has two protons with different chemical shifts, then two peaks appear on the  $^1\text{H}$  axis at the same position on the  $^{13}\text{C}$  axis.

## Incorporation of Magnetic Field Gradients

Changing the phase of the RF pulses and of the receiver in an NMR experiment in a systematic fashion (phase cycling) has been the method used to select only signals of interest, to eliminate artifacts, and to achieve quadrature detection in the  $t_1$  dimension. However, often 4, 8, or even 16 steps are needed for every  $t_1$  increment in a 2D NMR pulse sequence in order to achieve the correct level of suppression of undesired peaks. If the sensitivity of the experiment is such that one scan for each  $t_1$  increment would have been sufficient, then significant time is wasted in the data acquisition.

A new alternative method has been introduced for achieving all of the above ends. This is accomplished using pulsed magnetic field gradients along the main magnetic field axis. At least two such gradients are inserted into a pulse sequence, the first to cause dephasing of transverse magnetization and the second, some time later, to refocus only the components of magnetization that are desired. This can all be done in one scan, so removing the need for multiple scans for each  $t_1$  increment. This results in much more efficient data acquisition. Some of the disadvantages of using gradients include the fact that the method is inherently less sensitive than phase cycling, plus any molecular diffusion process that occurs during the pulse sequence will cause loss of signal intensity if the timing and strength of the gradients are not considered carefully. Field gradients are now used routinely for selecting exactly which part of the magnetization is detected.

## Three-Dimensional NMR Spectroscopy

Extension to 3D NMR spectroscopy is achieved by inserting a further evolution and mixing period. This means that a third time axis is introduced. Thus, a 3D experiment can be considered as two tandem 2D experiments, and 3D NMR spectra can be visualized as contour peaks in a cube defined by three axes,  $\omega_1$ ,  $\omega_2$ , and  $\omega_3$  derived by three Fourier transformation steps from the three time axes  $t_1$ ,  $t_2$ , and  $t_3$  (the

acquisition time). One example would be a  $^1\text{H}$  NOESY–TOCSY spectrum where  $t_1$  is the NOESY evolution time and  $t_2$  is the TOCSY mixing time. In this case, the  $\omega_1$ – $\omega_2$  plane is equivalent to a 2D TOCSY spectrum, the  $\omega_2$ – $\omega_3$  plane is a 2D NOESY spectrum, and the  $\omega_1$ – $\omega_3$  plane, sometimes referred to as the back-transfer plane, contains peaks from spins that are correlated through both NOE and  $J$  coupling. The 3D display is viewed as a stack of 2D planes that are parallel to one of the axes, usually  $\omega_3$  since this normally has the highest digital resolution.

Following the characterization of the human genome sequence and the identification of specific proteins coded for by the genes, this has led to a resurgence of interest in protein structure determination, or structural genomics as it is sometimes called. There are a number of important 3D NMR experiments that are used for assigning the peaks in NMR spectra of macromolecules such as proteins, which have been obtained fully labeled with  $^{13}\text{C}$  and  $^{15}\text{N}$ . Additionally, 3D NMR spectroscopy has been used to investigate the structures of synthetic polymers.

**See also:** **Liquid Chromatography:** Liquid Chromatography–Nuclear Magnetic Resonance Spectrometry. **Nuclear Magnetic Resonance Spectroscopy:** Principles; Instrumentation. **Nuclear Magnetic Resonance Spectroscopy–Applicable Elements:** Carbon-13; Nitrogen-15. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Nuclear Overhauser Effect.

## Further Reading

- Derome AE (1987) *Modern NMR Techniques for Chemistry Research*. Oxford: Pergamon Press.
- Ernst RR, Bodenhausen G, and Wokaun A (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Oxford: Clarendon Press.
- Martin GE and Zektzer AS (1988) *Two-dimensional NMR Methods for Establishing Connectivity. A Chemist's Guide to Experiment Selection, Performance and Interpretation*. New York: VCH Publishers Inc.
- Neuhaus D and Williamson MP (1989) *The Nuclear Overhauser Effect in Structural and Conformational Analysis*. New York: VCH Publishers Inc.
- Rinaldi PL (2003) Polymer characterization by 3D solution NMR. *NMR Spectroscopy of Polymers in Solution and in the Solid State: ACS Symposium Series 834*: 94–122.
- Sanders JKM and Hunter BK (1993) *Modern NMR Spectroscopy: A Guide for Chemists*, 2nd edn. Oxford: Oxford University Press.
- Staunton D, Owen J, and Campbell ID (2003) NMR and structural genomics. *Accounts of Chemical Research* 36: 207–214.
- Webb GA (Senior reporter) (2003) *Nuclear Magnetic Resonance. Specialist Periodical Reports*, vol. 32. London: Royal Society of Chemistry.

## Solid-State

**J W Zwanziger**, Dalhousie University, Halifax, NS, Canada

**H-W Spiess**, Max-Planck Institut fuer Polymerforschung, Mainz, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

As in nuclear magnetic resonance (NMR) of liquid samples, solid-state NMR probes the magnetic interactions of atomic nuclei. These interactions yield detailed information about the local structure and dynamics of the sample, including the bonding types and geometry, the site-site connectivity patterns, and the spatial characteristics and time scales of atomic and molecular motions. All kinds of solids can be studied with NMR, including single crystals and powders, disordered materials such as glass and rubber, and metals and superconductors. Spectral resolution, although not as high as in liquid-state NMR, is extraordinary (parts per million or better) but sensitivity is not. Sample volumes of order 50  $\mu\text{l}$  are typical.

The magnetic interactions probed in solid-state NMR include those studied in the liquid state, beginning with the Zeeman interaction between the nuclear spin and the applied magnetic field. This induces precession at the Larmor frequency  $\omega_0$ , which is defined by the nucleus and the strength of the external field. Fields as high as 21.2 T are in use, yielding proton Larmor frequencies of 900 MHz. Internal interactions observed include the chemical shift and, in favorable cases, scalar couplings. Additionally, magnetic dipole and electric quadrupole interactions, which are observable only indirectly in liquid-state NMR spectra, can be detected as frequency shifts in the solid state. In metals, the major spectral observable is the Knight shift. Because all these interactions depend sensitively on the local bonding geometry, they can be used to measure dynamic properties of the sample, either directly through spectral changes as a function of experimental parameters, or indirectly through the nuclear spin relaxation time.

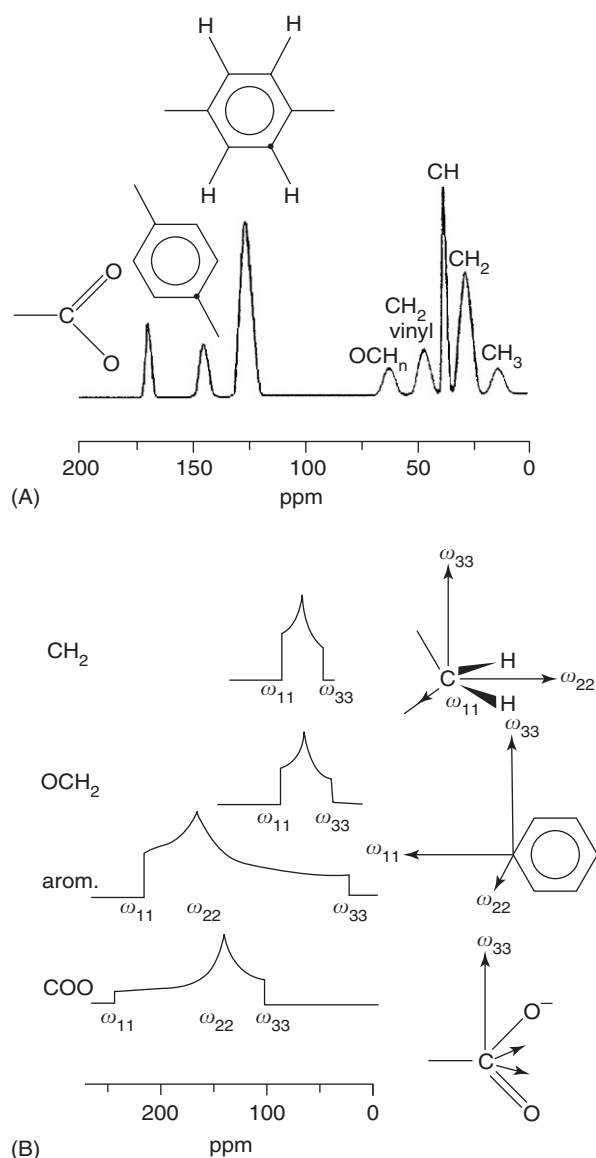
The primary difference between solid-state and liquid-state NMR is one of time scale. The atomic dynamics of the sample define a natural internal time scale, denoted  $\tau$ . The motion of interest might be, for example, the rotational tumbling of molecules in a liquid, the reorientation of segments in a polymer, or the hopping of ions in a solid electrolyte. Clearly,  $\tau$  can range from picoseconds to seconds or more. As

the observed nucleus moves to different locations or orientations, its NMR spectrum changes. This occurs both because the different sites may differ chemically, and also because the observable interactions are orientation dependent, the sense of orientation being defined by the external magnetic field. The different orientations define a range of frequencies  $\Delta\omega$  centered on the Larmor frequency. If  $\Delta\omega\tau \gg 1$ , a given nucleus samples many local environments during the NMR experiment. All interactions that depend on molecular orientation are in this way averaged to zero, and the spectrum is 'liquid-like'. 'Solid-like' spectra result when  $\Delta\omega\tau < 1$  for then the anisotropic portions of the interactions remain. These include for example the above-mentioned magnetic dipole and electric quadrupole terms. Typical examples of both regimes are shown in **Figure 1**.

**Figure 1** also shows the principal difficulty encountered in solid-state NMR spectra: the additional information provided by the anisotropic interactions can seriously congest the spectrum, making interpretation difficult. Our aim in this article is to outline the current principle methods by which solid-state NMR spectra can be acquired in interpretable form. Rather than giving an exhaustive account of the current developments in the field, we present the most important techniques in the context of the physical and chemical problems that they can help to solve.

### Resolving Chemically Distinct Sites

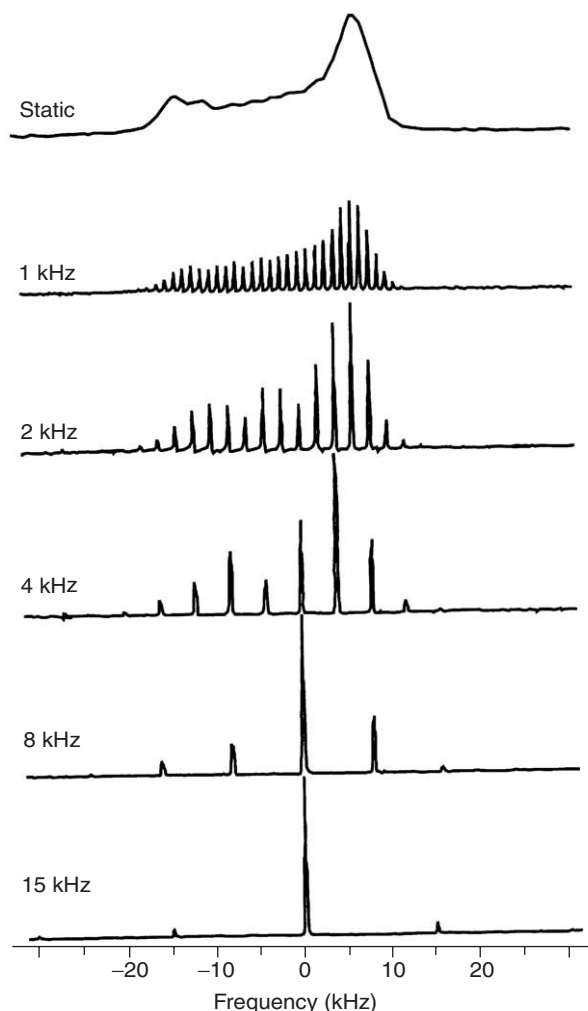
The most frequent application of solid-state NMR, as in the liquid state, is resolution of chemically distinct sites in a material. However, as **Figure 1** shows, the anisotropy observed in solid spectra typically creates so much spectral congestion that assignment is difficult. Moreover, as **Figure 1** also shows, the anisotropically broadened lines exhibit a variety of step and singularity features which, while informative in their own right, further obstruct a rapid assessment of the types and relative concentrations of distinct sites. The most important method for improving resolution in solid-state NMR is magic angle spinning (MAS). This method, so called because the sample is rotated about an axis inclined at the magic angle of  $54.74^\circ$  with respect to the magnetic field, can enhance the resolution by more than two orders of magnitude. It arises because the dominant anisotropies transform under rotations as second-rank spherical tensors, that is, like d-orbitals. Recall that the  $d_{22}$  orbital has an angular node; this is in fact at  $54.74^\circ$ . Thus, an interaction that transforms in the



**Figure 1** (A) Typical chemical shifts of carbon in different functional groups. The linewidths indicate the ranges observed; in a liquid sample, the actual linewidth will typically be much smaller. (B) Chemical shift powder patterns of carbons in the same functional groups. Such shapes are observed in powdered solids. The complex shapes, with steps and singularities, arise from the nontrivial orientation dependence of the chemical shift interaction, which is averaged to zero in a liquid but is observable in solids. The powder patterns are shown separately here for convenience; in real samples they overlap, making interpretation difficult. This problem is addressed by techniques such as magic angle spinning. (Reproduced with permission from Schmidt-Rohr K and Spiess HW (1994) *Multidimensional Solid-State NMR and Polymers*. London: Academic Press.)

same way can be averaged to zero by spinning about an axis located at this node.

To achieve effective averaging, the rotation frequency  $\omega_r$  in MAS must be of the order of, or greater than, the spread of interaction frequencies:



**Figure 2** The effect of magic angle spinning. The figure shows  $^{31}\text{P}$  spectra of  $\text{Na}_4\text{P}_2\text{O}_7 \cdot 10\text{H}_2\text{O}$ , as a function of rotor frequency. Note the extreme line-narrowing achieved, while still in a powdered solid. This illustrates that the symmetry of the chemical shift interaction is such that the full isotropic averaging of the liquid state is more than necessary to suppress the anisotropy; rotation about a single axis is in this case sufficient. The small splittings show that crystallographically, as well as magnetically, different sites can be resolved. (Adapted from Schnell I (1996) Diploma Thesis, Johannes-Gutenberg-Universität Mainz; see also and reproduced with permission from Kubo A and McDowell CA (1990) *Journal of Chemical Physics* 92: 715; © American Institute of Physics.)

$\omega_r/\Delta\omega > 1$ . The resulting resolution enhancement is dramatic, as shown in **Figure 2**.

For nuclei like  $^{13}\text{C}$ ,  $^{31}\text{P}$ , and  $^{29}\text{Si}$ , which have modest chemical shift ranges, spinning frequencies of 5–10 kHz are often sufficient, and are well within the range of typical commercial MAS probes.  $^1\text{H}$  spectroscopy is particularly challenging in solid-state NMR, in contrast to liquids, because of the strong magnetic dipole coupling between protons. This interaction gives a proton linewidth typically in the

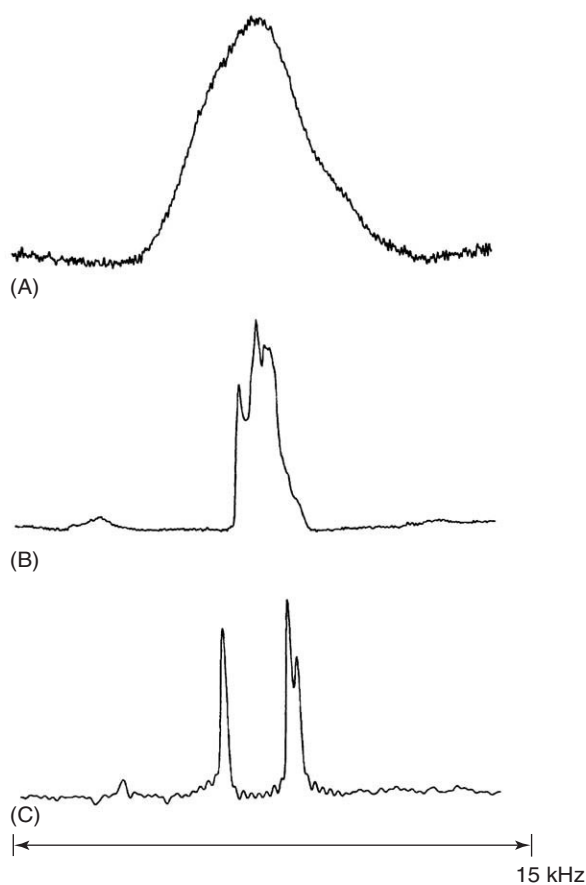
range of 20–50 kHz. Commercial MAS NMR probes are now available with spinning frequencies above 35 kHz, and so in many cases even  $^1\text{H}$  solid-state NMR can be accomplished with the MAS technique.

While MAS can provide significant resolution enhancement, it enhances sensitivity only insofar as the signal from broad resonances is concentrated into narrower resonances. For naturally low-abundance nuclei like  $^{13}\text{C}$  (1% naturally occurring), this increase may be insufficient. For dilute spins in the presence of an abundant species with good sensitivity (such as protons), which is to say, nearly all organic solids, double resonance methods may be used to achieve an additional gain in sensitivity. Coupled with MAS, these techniques are collectively referred to as cross-polarization MAS (CP-MAS). In CP-MAS, magnetization is first excited using the abundant species (typically  $^1\text{H}$ ), and then transferred to the dilute species ( $^{13}\text{C}$  or  $^{15}\text{N}$ , say) by simultaneously irradiating both nuclei, at their respective Larmor frequencies. Then, the dilute spin is detected, often with decoupling of the abundant spin. All this is carried out in the presence of MAS, to obtain good resolution of the resulting spectrum. The theoretical sensitivity gain is the ratio of the Larmor frequencies of the two species, for example, 4 for the  $^1\text{H}$ – $^{13}\text{C}$  pair. In practice, protons often have significantly shorter relaxation times than their CP partners, so this method also allows for shorter recycle delays in pulsed NMR, and thus more rapid acquisition of the spectrum. A disadvantage of CP-MAS is that the CP efficiency is a function of proximity of the abundant and dilute species, so that CP-MAS spectra cannot be assumed to be quantitative reflections of the abundances of the resolved sites. This limits the use of CP-MAS-NMR for analytical applications.

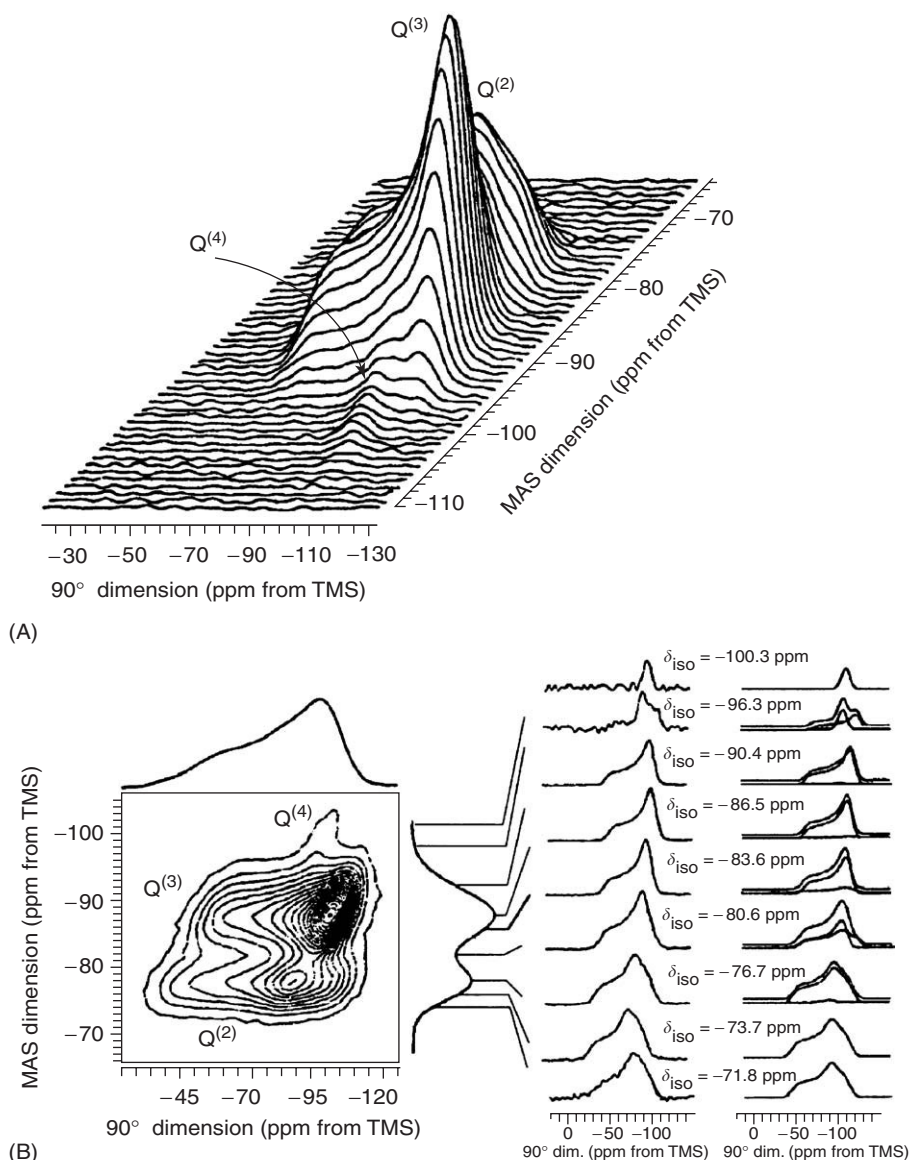
While MAS and CP-MAS are often sufficient to resolve chemical sites for nuclei like  $^{13}\text{C}$ ,  $^{31}\text{P}$ , and  $^{29}\text{Si}$ , this is not the case for other nuclei such as  $^{27}\text{Al}$ ,  $^{17}\text{O}$ , and  $^{11}\text{B}$ . The reason is that the first group of nuclei has spin 1/2, while the second have higher spin. Nuclei with spin greater than 1/2 are subject to electric quadrupole effects, in addition to the chemical shift, and these effects present a significant additional source of line broadening. Because the quadrupole anisotropy in the presence of a strong magnetic field does not transform simply like a second-rank tensor, it cannot be removed completely by MAS alone. During the last 15 years significant progress has been made in devising methods to average quadrupole interactions in addition to chemical shift anisotropy. In particular, multiple-quantum magic angle spinning (MQ-MAS) has become popular, which uses not just an MAS probe, but also a complex pulse sequence to excite and detect

triple- and higher-order coherences during the MAS experiment. As in CP-MAS, the technique is not analytically quantitative as uniform excitation of different chemical sites is difficult with existing pulse sequences. Despite such limitations impressive resolution advances for quadrupolar nuclei (**Figure 3**) have been achieved.

The methods described above yield greatly enhanced resolution, and as such help to determine the types and amounts of distinct sites in a material. This resolution gain comes at the price of discarding the information available from the interaction anisotropies. This information typically relates to the local site symmetry, for example, distortions in the bond angles, number of nearest neighbors, and so forth.



**Figure 3**  $^{87}\text{Rb}$  spectra of  $\text{RbNO}_3$ : (A) shows the static spectrum, while (B) shows the effect of magic angle spinning. The symmetry of strong quadrupole interactions is such that spinning about a single axis alone is not sufficient to remove all the anisotropy broadening. (C) Shows the results of multiple-quantum magic angle spinning, one of several methods currently available to obtain high-resolution spectra of quadrupolar nuclei like  $^{87}\text{Rb}$ . In this spectrum the three crystallographically distinct rubidium sites are resolved, and the total line narrowing is comparable to that achieved by MAS alone for spin-1/2 nuclei like  $^{13}\text{C}$  and  $^{31}\text{P}$  (see **Figure 2**). (Reproduced with permission from Brown S (1998) DPhil Thesis, Oxford University; and Brown SP and Wimperis S (1997) *Journal of Magnetic Resonance* 128: 42–61.)



**Figure 4** A separation of interactions-type spectrum of  $^{29}\text{Si}$  in a glass. (A) By spinning the sample off the magic angle during the first part of the experiment, and on the magic angle in the second, a two-dimensional spectrum is generated that has a high-resolution dimension correlated with the anisotropies of the individual sites. (B) Here, in a glass, each site itself shows a distribution of environments, which can be mapped out quantitatively by taking slices through the two-dimensional spectrum. In this way, for example, bond angle distributions, even in complex materials, can be determined, often with superior precision as compared to diffraction-based methods. (Reproduced with permission from Zhang P and Spiess HW (1996) *Journal of Non-Crystalline Solids* 204: 294.)

Both high-resolution and interaction anisotropies may be obtained by taking advantage of a second spectral dimension, using the so-called 'separation of interactions' experiments. An example is shown in **Figure 4**, where it is seen that the anisotropically broadened resonances are sorted according to their isotropic shift. In this way each resonance may be examined in isolation, and the anisotropy parameters determined with no congestion from neighboring bands. There are many ways to implement such experiments; an elegant approach for simple chemical shift correlations is to spin the sample at an angle

other than the magic angle during the first part of the experiment, followed by a hop to the magic angle and subsequent signal acquisition. In this way the anisotropic interactions 'label' the detected signal, and a double Fourier transform gives the type of spectrum shown in **Figure 4**.

### Determining the Connectivity between Sites

Once the types of sites in the material have been determined, using, for example, the techniques



discussed above, the second step in determining the material structure can be considered, namely, what the connectivity between these sites is. In liquid-state NMR, connectivities are primarily determined through scalar couplings, a through-bond interaction mediated by the electrons. However, this interaction is very small compared to the anisotropies of the magnetic interactions, and so is hard to probe in solids in any but well-ordered samples. On the other hand, magnetic dipole interactions, that is, the through-space effects of the nuclear magnetic moments on each other, can be substantial. This interaction varies with distance as  $r^{-3}$ , and so is of particular use in determining local structure. As noted above, dipole-dipole couplings are averaged to zero in 'liquid-like' spectra, but in solids they have easily observable effects on the resonance line shapes. Such interactions are also observed *indirectly* in liquids, through the nuclear Overhauser effect, where they appear as second-order interactions and thereby survive the averaging due to the molecular motion. Knowledge of the through-space connectivities does not give directly a map of the bonding network, but when combined with knowledge of the material composition and chemistry, can yield much about the bonding pattern.

For example, the CP-MAS experiment described above can already be used to obtain some degree of through-space information. The magnetization transfer, from  $^1\text{H}$  to  $^{13}\text{C}$  say, is mediated by the magnetic dipole interactions. By varying the duration of the transfer time (usually called the contact pulse) sites can be distinguished by their transfer efficiency. For example, primary and secondary carbons can be selectively excited, relative to tertiary and quaternary carbons, due to their greater proximity to protons. This is done simply by using a short contact pulse. Much more elaborate spectral editing schemes yield more accurate results, and are more flexible.

While variants of CP-MAS are particularly suitable for exploring proximities in heteronuclear systems with protons as one partner; other experiments can be performed to probe other heteronuclear systems and homonuclear couplings. All make use of the dipole-dipole coupling as the mechanism for encoding distance information.

In general, one wants to combine the distance measurement with some kind of resolution enhancement, in order to determine which sites are close to which. This is not always possible. For example, in many inhomogeneous solids, only one (broad) resonance will be observed, even with MAS or similar techniques. A well-studied example is sodium in a glass. Because techniques like MAS average the dipole-dipole coupling to zero, if they do not provide

sufficient resolution enhancement, they should not be used. Then, one studies a static sample. The spatial distribution of species in a static sample can be estimated by measuring the decay properties of spin echoes.

In spin echo experiments, an excitation pulse is followed at some time  $\tau$  later by a refocusing pulse. At time  $\tau$  after the second pulse, an echo will typically form. A classic use of this experiment for measuring distances is to estimate the so-called second moment ( $M_2$ ) of the resonance line. Interactions on a local scale, for example, the chemical shift and quadrupole interactions, and interactions involving isolated pairs of spins, can be refocused by using suitably chosen pulses. However, dipole coupling to a bath of partners cannot. Therefore, the echo cannot be refocused indefinitely, but only up to a characteristic time which is a measure of properties of the bath of nuclei coupled to the studied species. The decay constant of the spin-echo envelope is proportional to  $M_2$ , which is given essentially by summing over  $r_{ij}^{-6}$ , where the  $r_{ij}$  are internuclear distances. Because of the exponent  $-6$ , this experiment gives short-range information. It is valuable in assessing qualitative features of the distribution of species in inhomogeneous materials. An important extension of this experiment is called SEDOR, for spin-echo double resonance, in which an additional refocusing pulse is applied to a second nuclear species, and the echo signals with and without this secondary pulse are compared. In this way the mixing of different species in an inhomogeneous solid may be assessed.

Similar to SEDOR, but appropriate for isolated pairs of spins is the rotational echo double resonance method, or REDOR. In REDOR, the combined dynamics of the isolated two-spin system and the sample rotation serve to generate echoes at the rotor period. These echoes can be dephased by application of a pulse to one of the coupled partners. The amount of dephasing caused by this additional pulse is a measure of the coupling strength, and hence proximity, of the spins in the pair. This experiment is mostly applied to doubly or multiply labeled samples, for example, biopolymers enriched with  $^{13}\text{C}$  and  $^{15}\text{N}$ .

When the spectrum of the material consists of resolved resonances, much more detailed information on the nuclear distances can be derived than is possible with the spin-echo techniques outlined above. Magnetic dipole coupling is still the interaction to probe, but when the various sites are resolved, experiments can be used that give signals only if two distinct sites are near enough to each other to have a significant interaction. Clearly this yields much more detailed information than when the sites

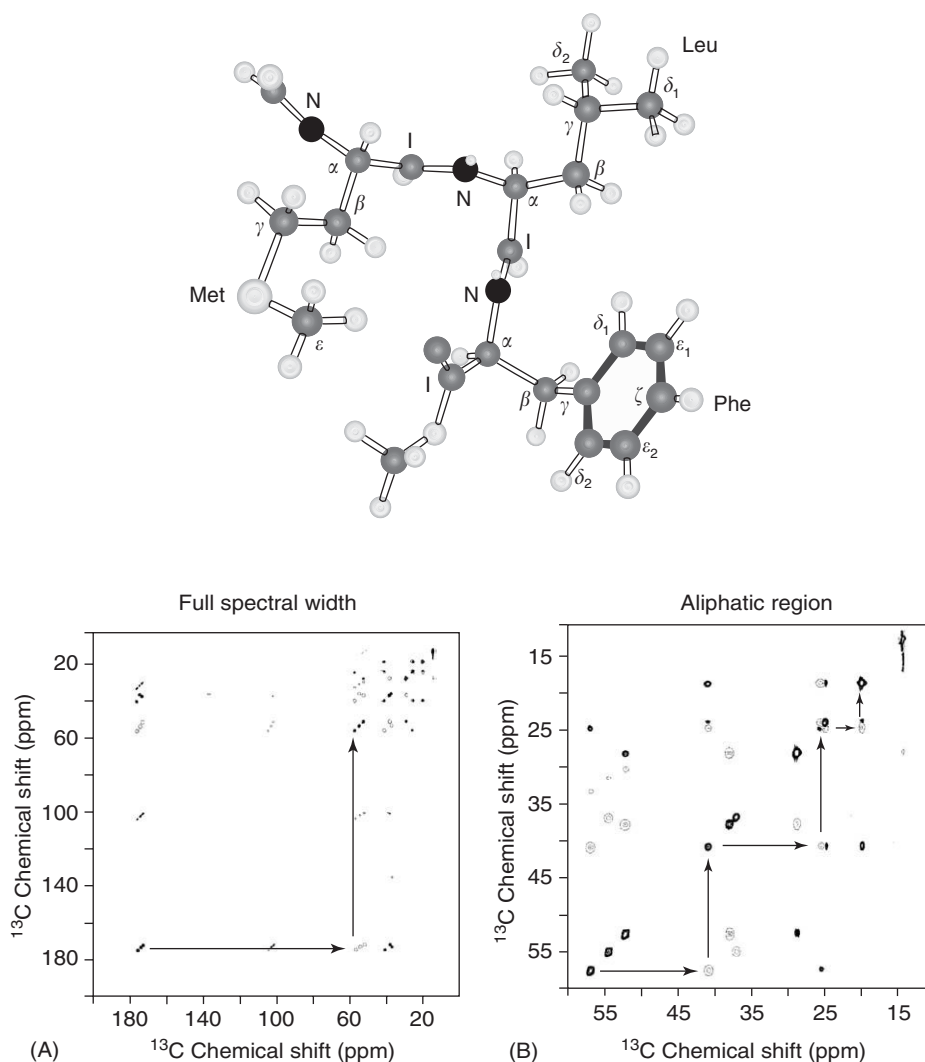
serve primarily to generate a background bath. Several types of signals can be generated and measured in this context, but the most precise are the so-called double quantum coherences. Isolated nuclei (here we have in mind only spin-1/2, such as  $^1\text{H}$  or  $^{13}\text{C}$ ) do not have enough energy levels to support quantum number changes greater than one, and therefore also cannot support coherences greater than one. If two such spins are coupled, however, the composite system can support two-quantum coherence. Experiments can be designed that are selective only for two-quantum coherence, thus yielding a connectivity map of the sites that are close enough spatially to couple in this way.

Generating such a connectivity map for a solid requires additionally a resolution-enhancement

technique, such as MAS. Such techniques, as discussed previously, suppress precisely the spin-spin interactions that the connectivity map is meant to reveal. Therefore, to combine multiple-quantum experiments with MAS, a pulse sequence that counteracts the averaging effect of MAS, thereby restoring the dipole-dipole coupling, must be implemented during excitation and evolution of the two-quantum coherence. A variety of such dipolar recoupling sequences currently exist, of varying levels of performance and complexity.

The resulting two-dimensional spectra give a connectivity map that can be traced in much the same way as is routinely done for liquid samples (Figure 5).

Such approaches are particularly important in biological applications to elucidate, e.g., the structure



**Figure 5** Double-quantum  $^{13}\text{C}$ – $^{13}\text{C}$  correlation spectra of the solid chemotactic tripeptide formyl-Met-Leu-Phe-OH. Such two-dimensional spectra give signals due to spatially proximal carbons, and thus provide means for determining the conformation of insoluble peptides. Together with other two- and three-dimensional experiments, solid-state NMR structures of proteins are now within reach, as is done routinely in liquid NMR using scalar couplings (a through-bond interaction). (Reproduced with permission from Rienstra C *et al.* (2000) *Journal of the American Chemical Society* 122: 10979–10990; © American Chemical Society.)

of membrane proteins, a rapidly developing field. It must be remembered, of course, that the signals observed reflect coupling through space, not through chemical bonds, so additional information about the chemistry must be used to interpret such spectra. Nevertheless, this method is fast becoming routine, as it requires only standard solids NMR instrumentation.

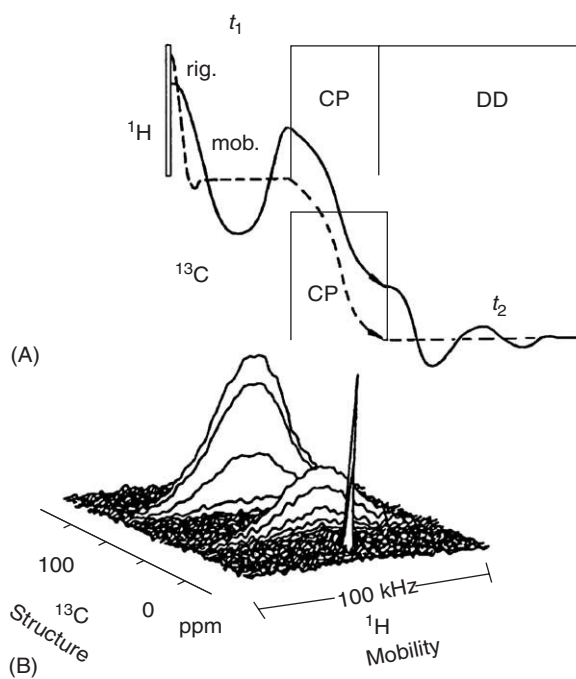
## Dynamics in Solids

The dynamics of atoms in solids may be probed directly, through their effects on the NMR spectra, and indirectly, through the nuclear spin relaxation. Because the NMR signal is observed only after the nuclear magnetization has been perturbed from its equilibrium state, relaxation is a standard feature of all NMR experiments. Two primary relaxation processes are usually identifiable. The first is the relaxation of the total magnetization back to its thermal equilibrium value; this occurs on a time scale denoted  $T_1$ . The second is the time scale for relaxation of quantum coherences in the spins, and is denoted  $T_2$ . In liquids, due to the strong decoupling resulting from the molecular motion, these two processes occur on similar time scales. In solids, however, they are usually very different, with  $T_2$  ranging typically from  $10^{-4}$  to  $10^{-2}$  s, and  $T_1$  from  $10^{-3}$  to  $10^{+3}$  s. Relaxation occurs because fluctuations in the surroundings of a spin induce transitions within the spin quantum states. Therefore, measurements of relaxation times are indirect probes of the dynamics in the solid. However, identifying what sort of fluctuation is operative is usually very difficult, unless one type of interaction is clearly dominant (conduction electrons in a metal or superconductor is a good example). Otherwise, the best that can be done is to estimate the temperature and magnetic field dependence of the relaxation to be expected from candidate fluctuation modes, and to compare the predictions with the data.

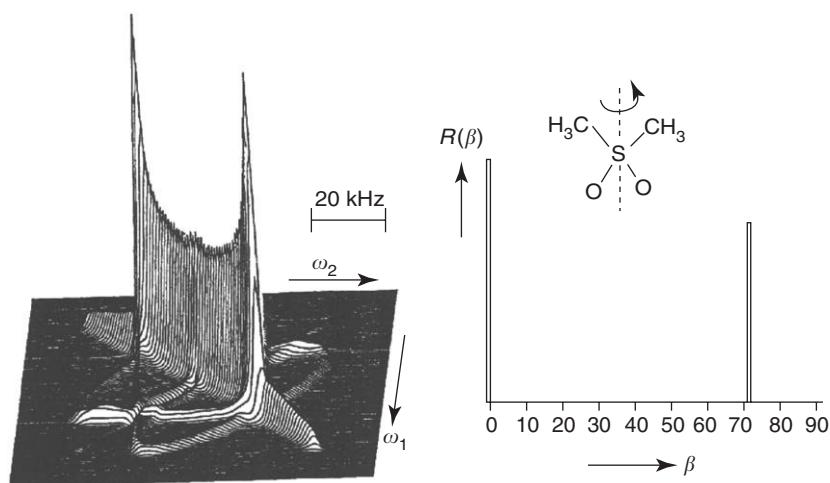
An easier qualitative assessment of dynamics can often be obtained from resonance line-shapes. As noted above, the key distinction between solid-like and liquid-like NMR spectra is the time scale of the atomic motions, compared to the frequency spread of the detected interactions. As the rate of a dynamical process increases, say as a function of temperature, it can be followed through changes in the linewidths of the nuclei involved. These changes can be substantial, as a resonance goes from solid-like at low temperatures to liquid-like at high temperatures, with the linewidth decreasing by orders of magnitude.

The linewidth strategy is particularly effective if no additional line narrowing is needed to interpret the low-temperature spectra; however, it is often the case

that these spectra will be so congested that additional techniques such as MAS must be applied. Then the line narrowing upon heating is much less dramatic. In this case, two-dimensional spectroscopy can again be very helpful. For organic solids and polymers the wideline separation of interactions (WISE) experiment is a convenient qualitative measure of the relative site dynamics. This experiment combines the good resolution found in  $^{13}\text{C}$  spectra under MAS, with the strong internuclear coupling of protons. The latter feature makes proton spectra particularly good indicators of motional narrowing due to dynamics: broad proton resonances (30–50 kHz) are seen in static samples, and narrow (<1 kHz) for mobile sites. The WISE experiment works by adding an additional evolution time to the CP-MAS sequence, between the proton excitation and the contact pulse



**Figure 6** Two-dimensional wideline separation of interactions spectrum of polystyrene–poly(dimethyl siloxane) di-block co-polymer (PS-*b*-PDMS). In the first part of the experiment (A) proton magnetization is allowed to evolve, and then transferred to carbon in the second part of the experiment. In this way the proton spectra of individual carbon sites are sorted by the shift of each site, and the result is a wideline proton spectrum in one dimension, and a high-resolution, MAS carbon spectrum in the other. In this example (B) the PDMS is seen to be quite mobile: it gives the carbon signal near 0 ppm, and the proton spectrum for this site is very sharp, indicating significant motional narrowing. The PS peaks, on the other hand, give very broad proton resonances, showing that the PS part of this block copolymer is essentially static at this temperature. With this experiment, quick qualitative assessments of relative local mobility in organic solids and polymers can be made. (Reproduced with permission from Schmidt-Rohr K and Spiess HW (1994) *Multidimensional Solid-State NMR and Polymers*. London: Academic Press.)



**Figure 7** Two-dimensional  $^2\text{H}$  exchange spectrum of deuterated dimethylsulfoxide. The strong diagonal ridge reflects molecules that have not reoriented during the mixing time, while the pattern of ellipses off the diagonal shows those that have. The ellipses arise due to the orientational dependence of the  $^2\text{H}$  quadrupole interaction, the dominant anisotropy here. The distribution of jump angles is shown on the right, and sharply peaked at zero (static molecules) and  $72^\circ$ , the included angle of the C–D bonds as the entire molecule executes hops about its symmetry axis. With this type of experiment, slow to moderate dynamics of molecules and polymers can be followed in atomic level detail. (Reproduced with permission from Schmidt-Rohr K and Spiess HW (1994) *Multidimensional Solid-State NMR and Polymers*. London: Academic Press.)

to the carbons, and using relatively slow sample spinning. In this way a two-dimensional spectrum is obtained, with proton resonances sorted by the carbon sites to which they are bonded (Figure 6).

One can immediately see, therefore, which carbon sites are mobile (narrow associated proton resonances) and which are static (broad proton resonances).

More detailed information on dynamics is available from the so-called exchange experiments. This class of two-dimensional techniques provides a correlation between spectral components, which exchange during a mixing period. The exchange may occur because of a chemical transformation during the mixing time, resulting in a new frequency due to a new chemical environment, or because of site re-orientation. Since the nuclear spin interactions are orientation dependent, if the molecular unit changes its orientation during the mixing time, the involved nuclear spins will exhibit altered NMR frequencies that can be correlated to their initial values. Because the orientation dependences of NMR interactions are well known, it is often a straightforward matter to relate the observed correlation spectrum to the underlying molecular motion that gave rise to it. In this way very detailed information on microscopic molecular dynamics can be obtained.

Deuteron NMR spectra provide particularly clear examples of the above approach. For deuterons ( $^2\text{H}$ ), the quadrupole interaction is by far dominant, and in a C–D bond is aligned with the C–D bond itself. Therefore, changes in time of the  $^2\text{H}$  quadrupole

orientation give a direct reflection of the orientational dynamics of the C–D bond itself. Figure 7 shows the  $^2\text{H}$  two-dimensional exchange spectrum of deuterated dimethylsulfone ( $(\text{CD}_3)_2\text{SO}_2$ ).

The strong diagonal ridge is a typical  $^2\text{H}$  NMR spectrum, and arises from molecules that did not exchange during the mixing time. The pattern of ellipses off the diagonal arises due to  $^2\text{H}$  nuclei with one orientation, and hence one frequency, at the start of the experiment, and a second orientation, hence frequency, after the mixing time. This pattern is consistent with  $180^\circ$  jumps of the molecules about their symmetry axis. The exchange experiment can be applied to other nuclei as well, such as  $^{13}\text{C}$ , although it can be harder to relate the spin interaction orientation to a molecular frame of reference.

**See also:** Nuclear Magnetic Resonance Spectroscopy: Overview; Principles; Instrumentation. Nuclear Magnetic Resonance Spectroscopy-Applicable Elements: Carbon-13; Phosphorus-31; Organometallic Compounds.

## Further Reading

- Abragam A (1961) *Principles of Nuclear Magnetism*. Oxford: Clarendon Press.
- Blümich B (ed.) (1994) Solid state NMR I–IV. In: Diehl P, Fluck E, Günther H, Kosfeld R, and Seelig J (eds.) *NMR Basic Principles and Progress*, vols. 30–33. Berlin: Springer.
- Fukushima E and Roeder SBW (1981) *Experimental Pulse NMR: A Nuts and Bolts Approach*. London: Addison-Wesley.

- Fyfe C (1983) *Solid State NMR for Chemists*. Guelph: CFC Press.
- Harris RK and Grant DM (1996) *Encyclopedia of Nuclear Magnetic Resonance*, vols. 1–8; vol. 9 (2002). Chichester: Wiley.
- Levitt MH (2001) *Spin Dynamics*. Chichester: Wiley.

- Mehring M (1983) *High Resolution NMR in Solids*, 2nd edn. Berlin: Springer.
- Slichter CP (1983) *Principles of Magnetic Resonance*, 3rd edn. Berlin: Springer.
- Traficante DD (ed.) *Concepts in Magnetic Resonance, An Educational Journal*. New York: Wiley.

## Surface Coil

**A N Priest and E B Cady**, University College London Hospitals Trust, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Nuclear magnetic resonance (NMR) spectroscopy can be used as a powerful noninvasive tool for clinical and other investigations of tissue biochemistry *in vivo*. There are many techniques for obtaining spectra from only a localized region of tissue, the simplest being to use a surface coil probe as both radio-frequency (RF) transmitter and receiver. This article reviews briefly surface coil spectroscopy, covering design, spectroscopic techniques, and clinical applications.

*In vivo* NMR spectroscopy has enabled the investigation and monitoring of metabolic changes in living tissues *in situ*. Unlike many other biochemical techniques, it is noninvasive, facilitating studies on previously inaccessible tissues such as the brain.

Surface coils can provide superior signal-to-noise ratios (SNRs) compared with volume coils. They provide some degree of spatial localization because as RF transmitters they excite magnetization in only a relatively small volume of the subject, while as RF receivers they only detect signals from a similar volume. However, the sensitive volume is poorly defined – in order to acquire data only from the tissue of interest, careful attention must be paid to contaminating signals. Further techniques, using magnetic field gradients, can give a second degree of localization, but this article focuses primarily on using the surface coil alone.

*In vivo* spectra often contain many overlapping resonances, due to both the relatively poor static magnetic field ( $B_0$ ) homogeneity and the vast number of metabolites in biological systems; hence, peak overlap can present serious problems for interpretation of results. This effect is more severe at low field strengths (e.g., 2.4 T, see **Figure 1A**) but can still be significant at e.g., 7.0 T (see **Figure 1B**).

Phosphorus ( $^{31}\text{P}$ ) surface coil spectroscopy has produced fruitful results in clinical applications. This nucleus acts as a probe of energy metabolism, via phosphocreatine (PCr), inorganic phosphate (Pi) and adenosine triphosphate (ATP). It also allows monitoring of intracellular pH ( $\text{pH}_i$ ), using the Pi chemical shift, and membrane metabolism, via phosphomonoesters (PMEs) and phosphodiesteres (PDEs). Carbon ( $^{13}\text{C}$ ) spectroscopy is also a powerful probe of energy metabolism; however, in many cases the low natural abundance of this isotope and its poor intrinsic NMR sensitivity make it difficult to obtain adequate signal except by infusion of  $^{13}\text{C}$ -labeled compounds. It is harder to apply proton ( $^1\text{H}$ ) surface coil spectroscopy *in vivo*, both because of strong, contaminating signals from superficial tissues (e.g., subcutaneous fat) and because extremely good  $B_0$  homogeneity is needed since numerous  $^1\text{H}$  resonances appear within a narrow bandwidth. However,  $^1\text{H}$  surface coil spectroscopy has been applied successfully to *in vivo* studies of human and other mammalian brains (see **Figure 2**). Recent  $^1\text{H}$  studies usually combine surface coil localization with gradient-localized pulse sequences.

### Theory

A surface coil consists of one or more coplanar loops of conductor, capacitively tuned to the NMR RF of the nucleus of interest and impedance-matched to the spectrometer's receiver. The surface coil is positioned flatly against the body surface adjacent to the tissue under investigation. Both the RF magnetic field ( $B_1$ ) generated during pulse transmission and the spatial sensitivity during reception are highly inhomogeneous. Magnetization close to the coil windings experiences a greater flip angle during transmission and induces a larger signal during reception. Indeed, the well-known 'principle of reciprocity' states that, for each point in space, the detection sensitivity is directly proportional to the magnetic field generated by the same coil (per unit current) when used as a transmitter.

- Fyfe C (1983) *Solid State NMR for Chemists*. Guelph: CFC Press.
- Harris RK and Grant DM (1996) *Encyclopedia of Nuclear Magnetic Resonance*, vols. 1–8; vol. 9 (2002). Chichester: Wiley.
- Levitt MH (2001) *Spin Dynamics*. Chichester: Wiley.

- Mehring M (1983) *High Resolution NMR in Solids*, 2nd edn. Berlin: Springer.
- Slichter CP (1983) *Principles of Magnetic Resonance*, 3rd edn. Berlin: Springer.
- Traficante DD (ed.) *Concepts in Magnetic Resonance, An Educational Journal*. New York: Wiley.

## Surface Coil

**A N Priest and E B Cady**, University College London Hospitals Trust, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Nuclear magnetic resonance (NMR) spectroscopy can be used as a powerful noninvasive tool for clinical and other investigations of tissue biochemistry *in vivo*. There are many techniques for obtaining spectra from only a localized region of tissue, the simplest being to use a surface coil probe as both radio-frequency (RF) transmitter and receiver. This article reviews briefly surface coil spectroscopy, covering design, spectroscopic techniques, and clinical applications.

*In vivo* NMR spectroscopy has enabled the investigation and monitoring of metabolic changes in living tissues *in situ*. Unlike many other biochemical techniques, it is noninvasive, facilitating studies on previously inaccessible tissues such as the brain.

Surface coils can provide superior signal-to-noise ratios (SNRs) compared with volume coils. They provide some degree of spatial localization because as RF transmitters they excite magnetization in only a relatively small volume of the subject, while as RF receivers they only detect signals from a similar volume. However, the sensitive volume is poorly defined – in order to acquire data only from the tissue of interest, careful attention must be paid to contaminating signals. Further techniques, using magnetic field gradients, can give a second degree of localization, but this article focuses primarily on using the surface coil alone.

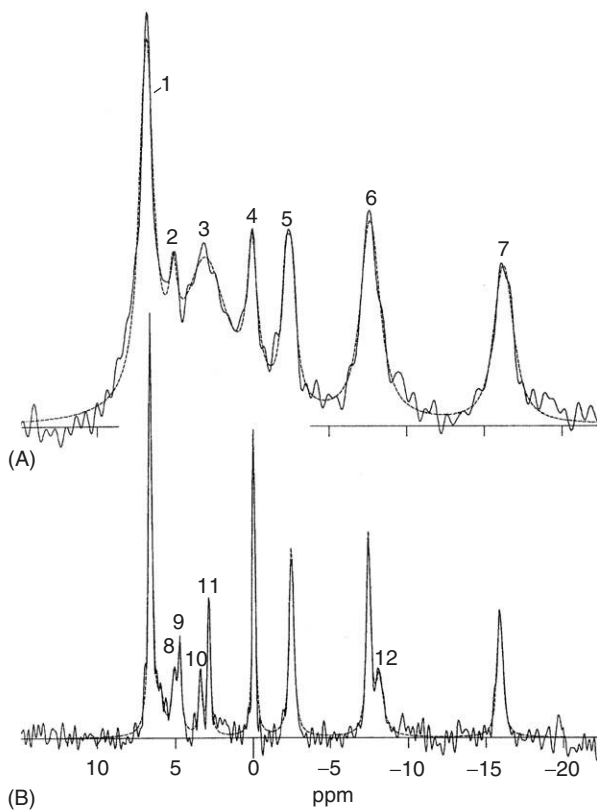
*In vivo* spectra often contain many overlapping resonances, due to both the relatively poor static magnetic field ( $B_0$ ) homogeneity and the vast number of metabolites in biological systems; hence, peak overlap can present serious problems for interpretation of results. This effect is more severe at low field strengths (e.g., 2.4 T, see **Figure 1A**) but can still be significant at e.g., 7.0 T (see **Figure 1B**).

Phosphorus ( $^{31}\text{P}$ ) surface coil spectroscopy has produced fruitful results in clinical applications. This nucleus acts as a probe of energy metabolism, via phosphocreatine (PCr), inorganic phosphate (Pi) and adenosine triphosphate (ATP). It also allows monitoring of intracellular pH ( $\text{pH}_i$ ), using the Pi chemical shift, and membrane metabolism, via phosphomonoesters (PMEs) and phosphodiesteres (PDEs). Carbon ( $^{13}\text{C}$ ) spectroscopy is also a powerful probe of energy metabolism; however, in many cases the low natural abundance of this isotope and its poor intrinsic NMR sensitivity make it difficult to obtain adequate signal except by infusion of  $^{13}\text{C}$ -labeled compounds. It is harder to apply proton ( $^1\text{H}$ ) surface coil spectroscopy *in vivo*, both because of strong, contaminating signals from superficial tissues (e.g., subcutaneous fat) and because extremely good  $B_0$  homogeneity is needed since numerous  $^1\text{H}$  resonances appear within a narrow bandwidth. However,  $^1\text{H}$  surface coil spectroscopy has been applied successfully to *in vivo* studies of human and other mammalian brains (see **Figure 2**). Recent  $^1\text{H}$  studies usually combine surface coil localization with gradient-localized pulse sequences.

### Theory

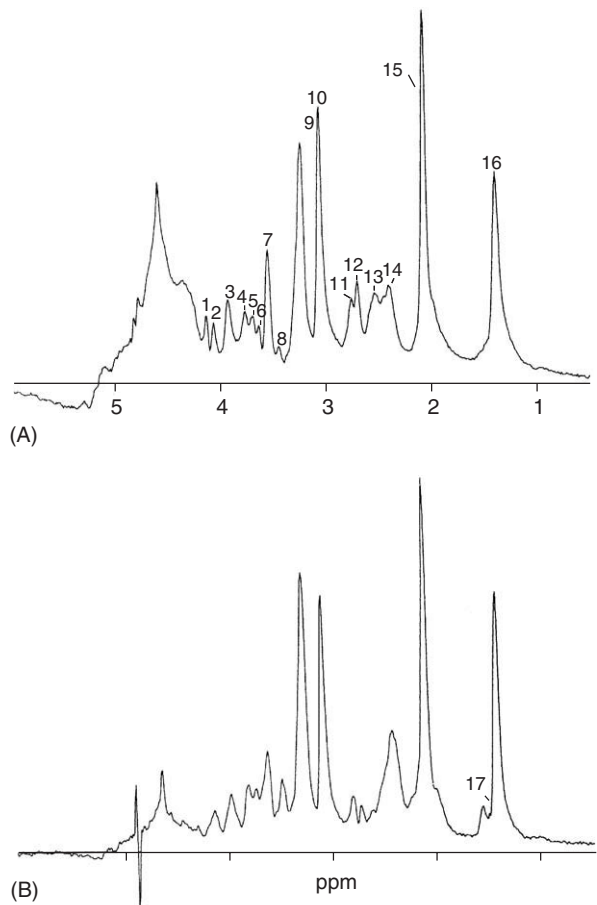
A surface coil consists of one or more coplanar loops of conductor, capacitively tuned to the NMR RF of the nucleus of interest and impedance-matched to the spectrometer's receiver. The surface coil is positioned flatly against the body surface adjacent to the tissue under investigation. Both the RF magnetic field ( $B_1$ ) generated during pulse transmission and the spatial sensitivity during reception are highly inhomogeneous. Magnetization close to the coil windings experiences a greater flip angle during transmission and induces a larger signal during reception. Indeed, the well-known 'principle of reciprocity' states that, for each point in space, the detection sensitivity is directly proportional to the magnetic field generated by the same coil (per unit current) when used as a transmitter.





**Figure 1**  $^{31}\text{P}$  surface coil spectra acquired noninvasively from mammalian brain *in vivo* at field strengths of (A) 2.4 T (40.6 MHz) and (B) 7.0 T (121.6 MHz). The 7.0 T spectrum (B) shows much greater resolution. (A) From the temporo-parietal cortex of a normal newborn human infant. Acquisition conditions: 6 cm single-turn coil;  $90^\circ$  flip angle at coil center;  $T_R = 2.256$  s; 256 summed free induction decays. Peak assignments: 1, PME (mainly phosphoethanolamine); 2, Pi; 3, PDE (mainly cross-linked PDEs); 4, PCr; 5–7,  $\gamma$ -,  $\alpha$ -, and  $\beta$ -NTP, respectively. The dashed line is a simulated spectrum generated from Lorentzian peaks, fitted by  $\chi^2$  minimization. The NTP peaks were fitted as doublets ( $\gamma$  and  $\alpha$ ) and a triplet ( $\beta$ ) with J-couplings of 16 Hz. (B) From normal neonatal porcine brain. Acquisition conditions: 2.5 cm two-turn, balanced-matched (inductively coupled), series-tuned coil;  $180^\circ$  flip angle at coil center;  $T_R = 10$  s; 384 free induction decays averaged. A DANTE pulse sequence (500 ten microsecond pulses, 200  $\mu\text{s}$  repetition time) was used prior to the acquisition pulse in order to reduce the broad bone and phospholipid signal. Additional resonance assignments: 8,  $\text{Pi}_1$  (probably extracellular and mitochondrial); 9,  $\text{Pi}_2$  (probably intracellular cytosolic); 10, glycerolphosphorylethanolamine; 11, glycerolphosphorylcholine; 12, nicotinamide dinucleotides. The simulated spectrum (dashed) was fitted in the manner described for spectrum (A). At 7 T the enhanced spectral resolution enables detection of additional resonances. Pi is resolved into two components (peaks 8 and 9), and, whereas the broad PDE signal (peak 3) is almost undetectable at 7 T (because of the effects of chemical shift anisotropy), two narrow resonances from mobile PDEs (peaks 10 and 11) are seen. Furthermore, nicotinamide dinucleotide is now clearly resolved from the adjacent  $\alpha$ -NTP doublet.

For any surface coil design, we can calculate  $B_1$  at any point in space by integrating the Biot–Savart law around the coil. For a single loop, planar surface coil of radius  $r_c$  in the  $y$ - $z$  plane of an orthogonal frame



**Figure 2**  $^1\text{H}$  surface coil spin-echo spectra obtained at 7.0 T (300 MHz) from *in situ* neonatal porcine brain post mortem. Data were collected using an EXORCYCLED  $1\bar{1}2 - \tau - 2 - \tau$ -acquire echo sequence with  $\tau = 66$  ms (a) and  $\tau = 135$  ms (b). The Hore  $1\bar{1}2$  and  $1\bar{1}2$  binomial solvent suppression sequences were used to reduce the signal from brain water. The  $1\bar{1}2$  and  $2\bar{2}$  interpulse delay was set to maximize the sensitivity for NAA  $\text{CH}_3$ - at  $\sim 2.02$  ppm. Acquisition conditions: coil as for **Figure 1B**;  $T_R = 1$  s; 256 summed echoes. Major resonance assignments: 1, lactate ( $\text{H}_\alpha$ ); 2, myo-inositol ( $\text{H}_2$ ); 3, creatine ( $\text{H}_\alpha\text{H}_\alpha'$ ); 4–6, glutamate and glutamine ( $\text{H}_\alpha$ ) and myo-inositol ( $\text{H}_4$  and  $\text{H}_6$ ); 7, glycine ( $\text{H}_\alpha$ ) and myo-inositol ( $\text{H}_1$  and  $\text{H}_3$ ); 8, taurine ( $\text{H}_\beta\text{H}_\beta'$ ); 9, phosphorylcholine and glycerolphosphorylcholine ( $-(\text{CH}_3)_3$ ) and taurine ( $\text{H}_\alpha\text{H}_\alpha'$ ); 10, creatine ( $\text{CH}_3$ -); 11 and 12, NAA ( $\text{H}_\beta$ ) and aspartate ( $\text{H}_\beta$  and  $\text{H}_\beta'$ ); 13 and 14, glutamate and glutamine ( $\text{H}_\gamma\text{H}_\gamma'$ ), NAA ( $\text{H}_\beta'$ ), and  $\gamma$ -aminobutyrate (GABA) ( $\text{H}_\alpha\text{H}_\alpha'$ ); 15, NAA ( $\text{CH}_3$ -) and small contributions from *N*-acetylaspartylglutamate ( $\text{CH}_3$ -), glutamate and glutamine ( $\text{H}_\beta\text{H}_\beta'$ ), and GABA ( $\text{H}_\beta\text{H}_\beta'$ ); 16, lactate ( $\text{CH}_3$ -); and 17, alanine ( $\text{CH}_3$ -). Residual unsuppressed water signals are seen at  $\sim 4.8$  ppm.

of reference  $xyz$  (see **Figure 3**), the radial ( $r$ ) and axial ( $x$ ) components of the  $B_1$  field at  $P(x, y, z)$  generated by the flow of unit current are given by

$$B_r = \frac{\mu_0}{2\pi} \frac{x}{\rho[(r_c + \rho)^2 + x^2]^{1/2}} \times \left[ -K(k^2) + \frac{r_c^2 + \rho^2 + x^2}{(r_c - \rho)^2 + x^2} E(k^2) \right] \quad [1]$$

$$B_x = \frac{\mu_0}{2\pi} \frac{1}{\rho[(r_c + \rho)^2 + x^2]^{1/2}} \times \left[ K(k^2) + \frac{r_c^2 + \rho^2 + x^2}{(r_c - \rho)^2 + x^2} E(k^2) \right] \quad [2]$$

where  $\mu_0$  is the permeability of free space, the off-axis distance of  $P$  is  $\rho = (y^2 + z^2)^{1/2}$ ,  $k^2 = 4r_c\rho/((r_c + \rho)^2 + x^2)$ , and  $E(k^2)$  and  $K(k^2)$  are complete elliptical integrals of the first and second kind, respectively. If points on the  $x$ -axis only are considered ( $y = 0, z = 0$ , and so  $\rho = 0, k^2 = 0$ ), both  $E(0)$  and  $K(0) = \pi/2$ , and

eqn [2] reduces to

$$B_x = \frac{\mu_0}{2} \frac{r_c^2}{(r_c^2 + x^2)^{3/2}} \quad [3]$$

Because NMR signals are weak, it may be desirable to use a pulse-sequence repetition time ( $T_R$ ) much less than five times the longitudinal relaxation time ( $T_1$ ) in order to maximize the SNR per unit acquisition time. The strong spatial dependence of the flip angle also causes the relaxation state of the nuclei to vary strongly with position. **Figure 4** shows the spatial sensitivity of a surface coil for various  $T_{RS}$ , demonstrating the imprecise boundary of the sensitive volume.

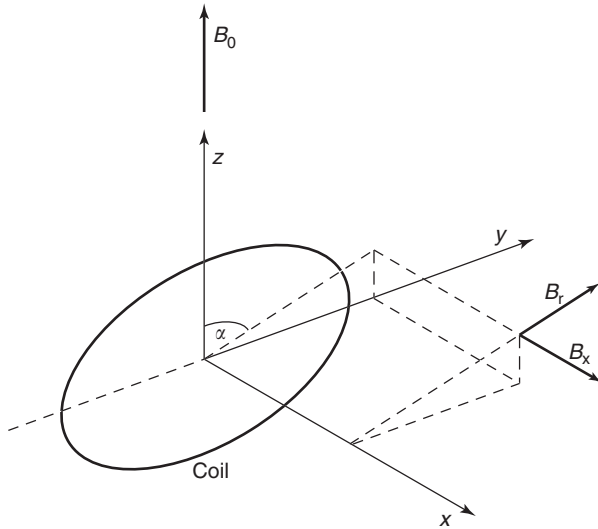
## Instrumentation

The specific requirements of many studies often require dedicated surface coils, and so it is useful to be able construct one's own.

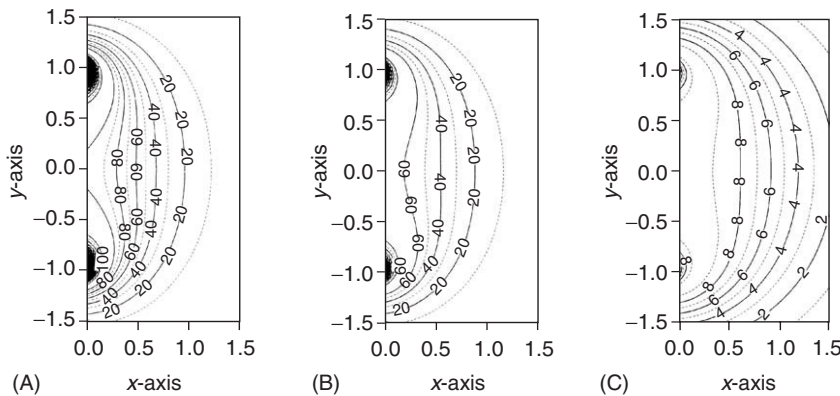
### Surface Coil Design

A surface coil is a resonant circuit, with an inductive part (the conducting loops) and tuning capacitors. The simplest design (see **Figure 5**) consists of one or more loops of conductor (e.g., high-purity copper wire) tuned with a parallel-mounted, high-efficiency 'fixed' tuning capacitor ( $C_f$ ). A high-efficiency variable capacitor ( $C_t$ ) in parallel with this enables fine-tuning of the resonant frequency. The coil is coupled to the spectrometer by a further variable capacitance ( $C_m$ ) that matches the coil impedance to that of the spectrometer (usually  $50 \Omega$ ). The inductance,  $L_c$ , is given by

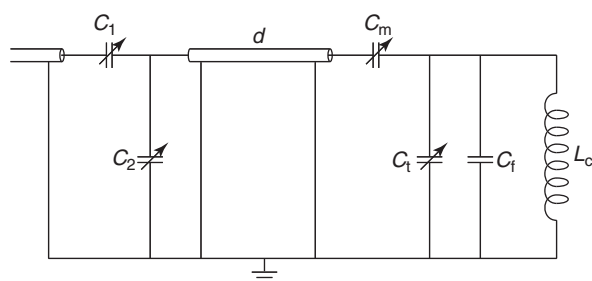
$$L_c = r_c n^2 \{ \mu_0 [\ln(8r_c/r_w) - 2] \mu' / 4 \} \quad [4]$$



**Figure 3** The radial ( $r$ ) and axial ( $x$ )  $B_1$  components at the point  $P(x, y, z)$ , generated by a unit current flowing in a single-turn surface coil.  $\alpha$  is the angle between  $B_r$  and  $B_0$ . (Reproduced with permission from Cady EB (1990). *Clinical Magnetic Resonance Spectroscopy*. New York: Plenum; © Plenum.)



**Figure 4** Signal amplitude maps in the  $x$ - $y$  plane, at  $z = 0$ , calculated for a single pulse under steady-state conditions with various  $T_{RS}$ . (A) Slow pulse repetition ( $T_R/T_1 = 5$ ). (B)  $T_R = T_1$ . (C) Rapid pulsing ( $T_R/T_1 = 0.05$ ). The signal amplitudes are normalized to 100 for a single  $90^\circ$  pulse at the coil center. (Reprinted with permission from Evelhoch JL, Crowley MG, and Ackerman JJH (1984) Signal-to-noise optimization and observed volume localization with circular surface coils. *Journal of Magnetic Resonance* 56: 110–124; © Elsevier.



**Figure 5** A circuit for a surface coil tuned for a single nucleus (e.g.,  $^{31}\text{P}$ ) with an add-on circuit that improves the efficiency at the  $^1\text{H}$  frequency (e.g., for shimming). The surface coil circuit (right-hand side) comprises the coil inductance,  $L_c$ , a fixed tuning capacitor,  $C_t$ , a variable tuning capacitor,  $C_t$ , and a variable matching capacitor,  $C_m$ . The add-on circuit (left-hand side) consists of a balanced length (approximately a quarter wavelength at the  $^1\text{H}$  frequency) of coaxial cable,  $d$ , and additional variable capacitors,  $C_1$  and  $C_2$ . (Reproduced with permission from Cady EB (1990). *Magnetic Resonance Spectroscopy*. New York: Plenum; © Plenum.)

where  $n$  is the number of turns of conductor,  $r_w$  is the radius of the wire, and  $\mu'$  is the conductor permeability. The required tuning capacitance ( $C_{\text{tune}}$ ) is given by

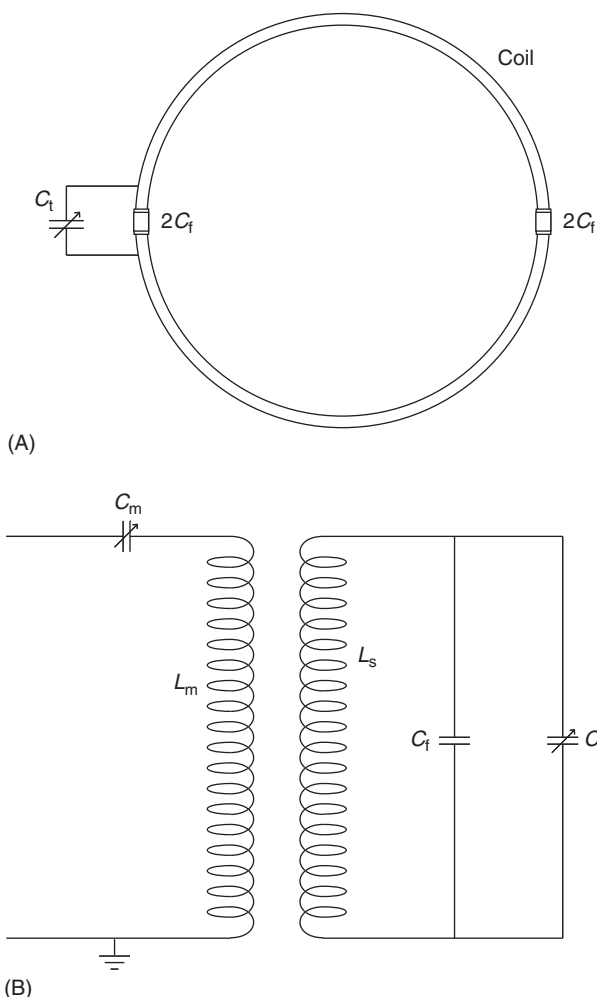
$$C_{\text{tune}} = \frac{1}{4\pi^2\nu^2 L_c} \quad [5]$$

where  $\nu$  is the resonant frequency in hertz. For a coil with this simple design, the duration,  $\tau$ , of the rectangular (hard) RF pulse required to produce a flip angle  $\theta$  on the  $x$ -axis is given by

$$\tau = \frac{40ZC_m r_c (r_c^2 + x^2)^{3/2}}{CV_t \gamma \mu_0 n r_c^3} \quad [6]$$

where  $Z$  is the spectrometer impedance,  $C$  is the total capacitance (i.e.,  $C_m + C_t + C_f$ ),  $V_t$  is the peak-to-peak transmitter voltage, and  $\gamma$  is the gyromagnetic ratio of the nucleus concerned.

When designing a surface coil, it is important to consider the study to be undertaken. The size of coil should match the sample dimensions as most spectroscopic information will be obtained from depths within about one coil radius. For too small a diameter, superficial tissues (i.e., skin and fat) occupy most of the sensitive volume, while for too large a diameter, other deeper or surrounding tissues may contribute a significant contaminating signal. Typical surface coil diameters are 3–10 cm for many human studies, or e.g. 1–2 cm for studying the brain of a small mammal. When the target tissue is skin, still smaller coils may be used or unusual coil geometries (such as the ‘crossover coil’ or the ‘zig-zag coil’) that detect strong superficial signals while having greatly reduced sensitivity further from the coil.



**Figure 6** (A) A series-tuned, balanced-matched, inductively coupled sample (primary) coil. This example consists of a single loop of conductor with two symmetrically positioned fixed tuning capacitors,  $2C_t$ , and a variable capacitor,  $C_t$ , for fine adjustment of the resonance frequency. The series tuning capacitors lower the coil voltage and should reduce dielectric losses. A secondary (impedance-matching) coil is required for coupling the primary to the spectrometer. (B) A circuit for a balanced-matched, inductively coupled surface coil.  $L_s$  is the inductance of the sample (primary) coil, which is tuned by  $C_t$  and  $C_t$ .  $L_m$  is the inductance of the matching (secondary) coil. The impedance matching can be fine-adjusted using  $C_m$ . (Reproduced with permission from Cady EB (1990). *Magnetic Resonance Spectroscopy*. New York: Plenum; © Plenum.)

It is important to position the tuning capacitances as close to the loops of the conductor as possible, minimizing unwanted inductances and capacitances – the ‘fixed’ capacitors should be almost integral components of the coil (see Figure 6A). Because the capacitances are quite close to the sample, they must be completely nonmagnetic; otherwise the  $B_0$  homogeneity will be compromised significantly. The coil components must be capable of handling the RF power to be used.

The optimum number of coil loops is usually  $<3$  and depends on the coil diameter and the NMR frequency of the nucleus concerned. Equation [5] implies that in order to tune the coil; the higher the frequency, the smaller the inductance (and hence the number of turns) or the capacitance. It is usually best to use the maximum number of turns compatible with tuning the coil to minimize the effects of stray inductances. The coil turns should be separated by a distance of  $\sim r_w$  to reduce interloop inductive effects. The loops must be insulated from each other and from the sample, both to ensure subject safety and to prevent coil earthing.

For good spectral quality, the coil's SNR should be optimized. As suggested by the principle of reciprocity, SNR losses as a receiver are closely related to energy losses as a transmitter. Both are quantified in terms of the surface-coil  $Q$ -factor, the proportion of energy lost per cycle. There are three sources of loss – the coil, the sample, and radiation losses. Coil losses result from both resistance and capacitor dielectric losses. Sample losses (loading) result from currents induced in an electrically conductive subject. Radiative losses refer to electromagnetic energy leaving the system.

Resistive losses in the coil are usually small compared with sample (loading) effects. However, at very low field strengths or for very small samples, resistive effects can dominate over other loss mechanisms. They can be reduced by exercising care in designing the coil geometry, by cooling the coil wires, or by making the coil from a high-temperature superconductor such as YBaCuO.

Dielectric losses in the capacitors may be more substantial, especially in heavily loaded coils where large transmitter voltages are needed. To reduce dielectric losses, various design improvements have been suggested that lower coil voltages. **Figure 6** shows one such design, known as a 'balanced-matched' coil. This probe consists of two coils forming an air-cored RF transformer: a primary coil (the sample coil that is positioned adjacent to the tissue of interest); and a secondary coil (used to impedance-match and couple the sample coil to the spectrometer). The mutual inductance is adjusted to match the primary coil impedance to that of the spectrometer, by fine-tuning either the physical position or a capacitance in series with the secondary coil. Dielectric losses are reduced in this design because voltages are distributed symmetrically about the tuning capacitances and therefore are smaller. In the example in **Figure 6**, the primary coil voltages are reduced still further using two tuning capacitances in series with the inductance (also split into two equal parts).

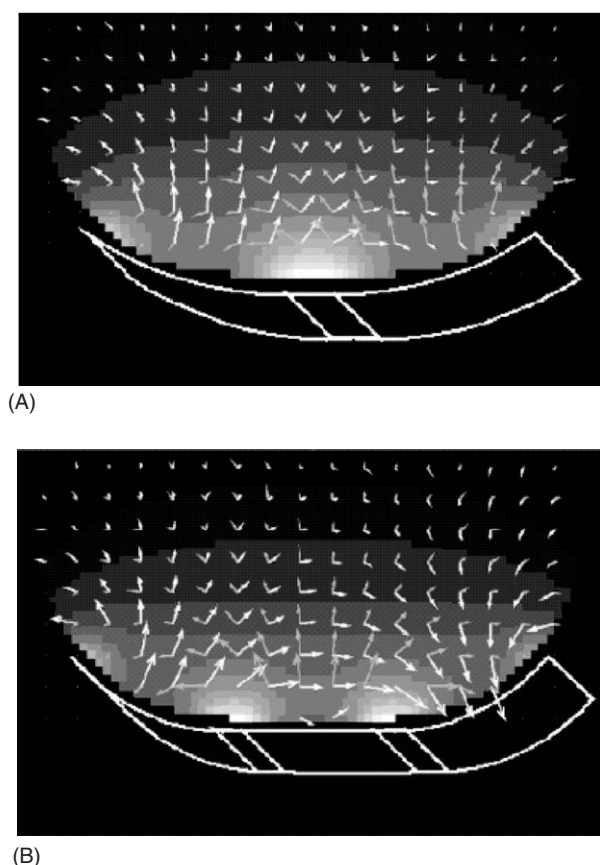
When a conductive sample (e.g., part of the human body) is brought near the probe, energy is dissipated by currents induced in the sample, thereby reducing the coil's efficiency. This effect, known as 'loading', increases with the RF, the sensitive volume of the coil, and the conductivity of the subject. *In vivo* loading nearly always dominates over other losses, most of which are reduced by good coil design. *In vivo* loading leads to a large reduction in  $Q$ . One consequence of this is the so-called 'lift-off effect' – a higher SNR may sometimes be achieved by displacing the coil a small distance from the subject because the reduced loading outweighs the drop in sensitivity due to increased distance.

Radiative losses are usually minimized by ensuring that all winding lengths are substantially less than the wavelength, and may be reduced further using radiation shields. Innovative designs such as 'microstrip' coils can also reduce radiative losses by combining the function of a shield with that of a distributed capacitance.

In addition to the nucleus of interest (e.g.,  $^{31}\text{P}$ ), the surface coil may need to detect the  $^1\text{H}$  signal from tissue water for sample shimming (making  $B_0$  uniform throughout the sample). Even for a coil tuned to  $^{31}\text{P}$ , a sufficient water signal may normally be detected, and its amplitude can be further increased by adding an extra tune-and-match network and a selected length of coaxial cable in series (see the left-hand side of **Figure 5**). Alternatively, a double-tuned coil may be used (see later) that may also be used for spectroscopy with the second nucleus.

Quadrature surface coils can improve the SNR by a factor up to  $\sqrt{2}$ . These consist of two separate coils that produce approximately mutually orthogonal  $B_1$  fields, with similar amplitudes, over much of the sensitive region (**Figure 7**). This allows the two circular polarizations to be distinguished. However, for a quadrature coil pair to function, there must be minimal coupling between the component coils – no net magnetic flux should be captured in one coil that originates from the other. One common design is to use two partially overlapping coils; ideally they will be tilted at  $90^\circ$ , but in practice the sample geometry usually requires a smaller angle. Another common design is a single loop overlapping a figure-of-eight double loop (butterfly) configuration, wrapped around a half circle.

In many applications, multinuclear coils are necessary, so that data may be acquired from more than one nucleus in the same experiment (e.g., both  $^1\text{H}$  and  $^{31}\text{P}$ ). The simplest design, allowing only sequential study of the different nuclei, changes the tuning capacitance, using a nonmagnetic microswitch to add a further capacitor in parallel to  $C_f$ . In more



**Figure 7** Two designs for quadrature detection surface coils: (A) two planar loops and (B) butterfly coil and single loop coil. The image grayscale represents the computed  $B_1$  amplitude in the  $x$ - $y$  plane, while the arrows show the direction of the  $B_1$  fields produced by the two coils. Good quadrature detection occurs in the central region where the fields from the two coils are perpendicular and of similar magnitude. (Reproduced from Haase A, Odoj F, Von Kienlin M, *et al.* (2000) NMR probeheads for *in vivo* applications. *Concepts in Magnetic Resonance* 12: 361–388, with permission of John Wiley & Sons, Inc. Copyright © 2000 Wiley Periodicals, Inc. A Wiley Company.)

complicated designs, the coil may be ‘double-tuned’ to two frequencies simultaneously. It is a challenge to design double-resonant circuits that operate optimally at both frequencies.

### Techniques

Despite a very inhomogeneous spatial distribution of flip angles, many *in vitro* NMR spectroscopy methods can be used with surface coils. For instance, the large  $^1\text{H}$  signal from tissue water may be reduced using the Hore binomial water-suppression sequences (e.g., the  $1\bar{3}3\bar{1}$  or  $1\bar{3}3\bar{1} - 2\bar{6}6\bar{2}$  pulse sequences). Each  $+\theta$  and  $+3\theta$  flip angle in the sequence (denoted by 1 and 3, respectively) has a corresponding  $-\theta$  and  $-3\theta$  flip angle (denoted by  $\bar{1}$  and  $\bar{3}$ ) so that the water magnetization (on resonance) is returned to its initial

orientation parallel to  $B_0$  but the metabolite magnetizations (off resonance) finally have components in the transverse plane and are detected. Other frequency-selective saturation techniques (e.g., DANTE sequences) can be used to suppress unwanted resonances partially (see **Figure 1B**).

Spin-echo techniques ideally require exact  $90^\circ$  or  $180^\circ$  flip angles throughout the sample but may still be used with a nonuniform flip angle, despite some difficulties. Phase-cycling sequences such as EXOR-CYCLE can eliminate disadvantages such as phase and amplitude distortions. Spin-echoes allow  $^1\text{H}$  spectra to be edited and simplified, thereby facilitating measurements of resonances that would otherwise be unresolved owing to peak overlap. Using a double-resonant coil (see above) in  $^{13}\text{C}$  or  $^{31}\text{P}$  studies, applications such as  $^1\text{H}$  decoupling, nuclear Overhauser enhancement, and polarization transfer can all help increase the SNR for  $^1\text{H}$ -coupled nuclei.

Although a surface coil alone can provide sufficient localization for many studies, a second level of localization may be useful for others. Magnetic field gradient coils, commonly used for imaging the subject, also allow collection of spectra from much more specifically defined volumes (voxels). ‘Nonecho localization’ methods such as image-selected *in vivo* spectroscopy (ISIS) and outer-volume suppression (OVS; such as single-shot inversion-recovery based nonecho) are especially useful with surface coils, but the echo techniques point-resolved spectroscopy (PRESS) and stimulated echo acquisition mode (STEAM) may also be used. Selecting a voxel with a smaller volume reduces the SNR of the spectrum, but this is compensated partially by narrower linewidths due to the superior  $B_0$  field homogeneity obtainable with additional localization. With additional localization, the spectroscopic information acquired is much more specific to the target tissue; therefore contaminating signals, e.g., from superficial lipids, are greatly reduced. These gradient-localized methods are especially useful for  $^1\text{H}$  spectroscopy, which has a high intrinsic sensitivity and reveals many overlapping resonances within a relatively small chemical shift range. By adding ‘phase-encoding’ gradients to any of these methods, surface coils may also be used for chemical shift imaging experiments.

PRESS, based on a double spin-echo, requires exact  $90^\circ$  and  $180^\circ$  pulses ideally. However PRESS too can be applied in the inhomogeneous surface coil  $B_1$  field, provided the out-of-voxel signal from non-ideal pulses is removed using phase-cycling schemes and/or large spoiler gradient pulses. The same applies to the stimulated echo sequence STEAM. More complicated versions of PRESS may be implemented

using pairs of adiabatic pulses, such as the hyperbolic secant or frequency offset corrected inversion pulses (FOC1), to obtain a uniform flip angle throughout the voxel. For ISIS, such adiabatic pulses can give good inversions almost independent of  $B_1$ . For OVS methods,  $B_1$ -independent saturation outside the voxel of interest may be achieved using either inversion recovery methods or 'noise' pulses. For  $^1\text{H}$  spectroscopy, these techniques may be combined with water suppression using chemical shift selective (CHESS) pulses. Multiple CHESS pulse schemes have been optimized to give excellent suppression even in inhomogeneous  $B_1$  fields, for a wide range of  $T_1$  relaxation times.

Another consequence of the nonuniform sensitivity of the coil is that quantification of absolute metabolite concentrations is only possible if it is assumed that the underlying metabolite distribution is uniform. Nevertheless, with appropriate handling of relaxation effects, acceptably accurate absolute quantification has been demonstrated. With a coil tuned for a single nucleus, a homonuclear external concentration reference can be employed. Using a double-tuned coil, the signal from the nucleus of interest (e.g.,  $^{31}\text{P}$ ) can be referenced to the  $^1\text{H}$  signal from water, either in the subject or in an external standard. It is also possible to calculate the absolute sensitivity of the coil at a given point in space by using the principle of reciprocity combined with a knowledge of the transmitter pulse amplitude required to produce a  $90^\circ$  flip angle.

## Spectral Data Analysis

The poor  $B_0$  homogeneity encountered *in vivo* may lead to substantial overlap of neighboring peaks, especially at the lower field strengths used commonly for clinical investigations. Rigorous data analysis methods are needed for getting accurate measurements of both peak area and chemical shift despite such overlaps. Ideally, the analysis should be automated and user-independent. Computers have allowed the use of very sophisticated analysis techniques, covering both analysis of spectra and interpretation of the results through pattern recognition approaches like principal component analysis, linear discriminant analysis, or neural network approaches.

A spectrum may be analyzed either in the frequency domain (i.e., the Fourier transform spectrum) or in the time domain (i.e., the raw signal). The simplest methods in both approaches attempt to fit model functions to the different resonances in the spectrum. Typically, the models might be Lorentzian or Gaussian lineshapes in the frequency domain or exponentially decaying sinusoids in the time domain.

The chemical shift, amplitude, phase, and width (or damping factor) are estimated for each peak. In the frequency domain, it is common to 'phase' the spectrum by applying constant and linear phase corrections, determined either automatically or manually.

There have been many improvements on these basic principles, reflected in many sophisticated software packages. Lineshape distortions may be compensated by correcting either the spectrum or the model functions. Broad background features may be fitted using suitable slowly varying functions in the frequency domain or suppressed in the time domain by excluding or damping the first few raw data points. For compounds that produce several resonances or have multiplet resonances, the fits may be constrained using prior knowledge about the relative amplitudes, frequencies, widths, J-couplings (multiplet splittings), and phases of the peaks. Extending this idea, the chosen model functions may be the entire spectrum for each metabolite (either measured from sample solutions or simulated) instead of the individual resonances – this is the basis of the linear combination model (LC model), which analyzes NMR data in the frequency domain. A similar approach may also be applied to time domain data, for example within the readily available magnetic resonance user interface (MRUI) package. Such approaches can however be problematic when the resonance frequencies shift due to changes in  $\text{pH}_i$  or metal ion concentrations.

## Applications

Surface coils have been applied to studies of many tissues in a wide variety of situations, using several different nuclei including  $^{31}\text{P}$ ,  $^1\text{H}$ ,  $^{13}\text{C}$ , and fluorine ( $^{19}\text{F}$ ). Two example applications discussed here are studies of the brain of the newborn infant and cancer.

For clinical applications there are safety guidelines from official national organizations limiting the exposure of staff and patients to static and varying magnetic fields, RF fields, and acoustic noise (from gradient coils). In nonclinical applications these guidelines may often be relaxed unless human volunteers are involved.

### Neonatal Brain, Birth Asphyxia, and Energy Metabolism

In the newborn, the cranial musculature is poorly developed, allowing  $^{31}\text{P}$  surface coil spectroscopy of neonatal brain without additional localization. An extremely broad peak underlies the spectrum, from relatively immobile  $^{31}\text{P}$  nuclei in cranial bone, but this can be removed in postacquisition processing. As



shown in **Figure 1A**, in neonatal brain, strong signals are detected from PME (mainly phosphoethanolamine (PEt)), PDEs (a fairly broad peak originating mainly from cross-linked PDEs in cellular membrane bilayers), and nucleotide triphosphates (NTPs) – mainly ATP. Smaller, clearly resolved peaks due to Pi and PCr are seen also. Several studies of the brains of human infants have demonstrated significant developmental changes in the relative levels of the phosphorylated metabolites detectable in  $^{31}\text{P}$  spectra. As infants develop, the cerebral PCr/Pi and PDE/ $P_{\text{tot}}$  ratios increase, whereas PME/ $P_{\text{tot}}$  decreases ( $P_{\text{tot}}$  is the total mobile phosphate:  $\text{PME} + \text{Pi} + \text{PDE} + \text{PCr} + 3 \times \beta\text{-NTP}$ ). The PME/PDE ratio has been used as an index of brain development.

In infants who have suffered perinatal hypoxic–ischaemic brain injury (birth asphyxia),  $^{31}\text{P}$  spectra obtained within a few hours of birth show no abnormalities. However, in many cases a delayed ‘secondary energy failure’ (SEF) develops within  $\sim 24$  h. The Pi signal increases, accompanied by reduced PCr and, in severe cases, low NTP. Furthermore, in contrast to the profound acidosis seen during acute hypoxia–ischemia, an intracellular alkalosis may be detected. Additionally, localized  $^1\text{H}$  spectroscopy shows elevated lactate levels for several weeks following SEF. Both PCr/Pi and lactate/N-acetylaspartate (Lac/NAA) ratios have proved to be extremely useful indices of the severity of hypoxic–ischaemic injury, with strong prognostic capabilities.

### Cancer

NMR spectroscopy of tumors has become an area of great interest, with  $^{31}\text{P}$  studies in particular showing potential as a noninvasive means of understanding tumor metabolism, to grade tumor malignancy or stage disease progression and to predict the efficacy of particular forms of therapy. Many different types of tumor have been studied, with *ex vivo* and *in vitro* NMR as well as *in vivo*. Most tumor spectra contain strong PME signals, consisting of both PEt and phosphocholine (PCh), with some indications that high levels of PCh may correlate with higher-grade tumors. Pi, PDE, and NTP are also seen, but the levels of PCr are usually much smaller than in many other tissues. Usually, PME increases as the disease progresses, whereas a decrease in PME often gives a clearer and earlier indication of a positive treatment response than reduced tumor volume and could help determine clinical management. The levels of PDE, NTP, and total phosphates can also change significantly. In some sarcomas, the pretreatment pH<sub>i</sub>, measured from the Pi chemical shift, correlates with posttreatment tumor necrosis.

Other nuclei may also be useful for improving our understanding of cancer. Gradient-localized  $^1\text{H}$  spectra have been used in tumor classification, sometimes in conjunction with computer-based ‘pattern recognition’ methods such as principal component analysis. Also,  $^{19}\text{F}$  NMR has been used to assess the uptake of particular cytotoxics such as 5-fluorouracil. Patients who show tumor uptake of  $^{19}\text{F}$ -labeled 5-fluorouracil are more likely to respond to chemotherapy with this agent.

### Other Applications

Skeletal muscle is a classic early example of the use of  $^{31}\text{P}$  surface coil spectroscopy for studying energy metabolism. With a surface coil, muscle metabolism can be monitored easily, not only in unfatigued, resting muscle but also during contraction or ischemia. During exercise, the normally strong PCr signal falls, with a concomitant increase in Pi. If the exercise is severe, ATP may decline. Lactic acidosis is revealed by a reduced Pi chemical shift. In contrast, an alkaline pH change is observed in McArdle’s syndrome, in which glycolysis is inhibited, thereby stopping lactate production.

Despite the difficulties of obtaining an adequate signal from  $^{13}\text{C}$ , which has a natural abundance of only  $\sim 1.1\%$ , it is still possible to use this nucleus for investigating energy storage compounds such as fatty acids and glycogen, e.g., in liver, muscle, or adipose tissue. This can distinguish patients with glycogen storage diseases from normal subjects. Many more compounds can be detected after infusion with  $^{13}\text{C}$ -labeled compounds such as  $^{13}\text{C}$ -glucose, which has been used to investigate glucose storage and metabolism in some detail, especially in patients such as diabetics.

Gradient-localized  $^1\text{H}$  spectroscopy is applied widely in the brain since spectra with good line-widths may be obtained from a specific region. Many pathologies have been investigated, including epilepsy, dementia, multiple sclerosis, stroke, Canavan’s disease, and Huntington’s disease. A common feature of many brain pathologies is a reduced level of NAA (primarily a neuronal marker). Changes in the levels of glutamate, glutamine, myo-inositol, and gamma-aminobutyric acid (GABA) can also be seen in certain pathologies. Spectroscopy has even discovered new syndromes, in particular a deficiency of creatine.

### Trends

Surface coils continue to be used for specific applications in studies of living tissue. In particular, they are proving of great utility for experimental studies in high-field, small-bore spectrometers. Recent and

ongoing developments of spectral editing, data analysis, localization, and RF pulse design enhance the utility of these NMR probes, provided by their high SNR.

*See also:* **Nuclear Magnetic Resonance Spectroscopy: Principles; Instrumentation. Nuclear Magnetic Resonance Spectroscopy-Applicable Elements:** Phosphorus-31. **Nuclear Magnetic Resonance Spectroscopy Techniques:** *In Vivo* Spectroscopy Using Localization Techniques.

## Further Reading

Choi I-Y, Tkáč I, and Gruetter R (2000) Single-shot, three-dimensional “non-echo” localisation method for *in-vivo* NMR spectroscopy. *Magnetic Resonance in Medicine* 44: 387–394.

Evelhoch JL, Crowley MG, and Ackerman JJH (1984) Signal-to-noise optimization and observed volume

localization with circular surface coils. *Journal of Magnetic Resonance* 56: 110–124.

Gadian DG (1995) *Nuclear Magnetic Resonance and its Applications to Living Systems*, 2nd edn. Oxford: Oxford University Press.

Garwood M and De la Barre L (2001) The return of the frequency sweep: Designing adiabatic pulses for contemporary NMR. *Journal of Magnetic Resonance* 153: 155–177.

Igata A, Asakura T, and Fujimoto T (2000) *MRS of the Brain and Neurological Disorders*. Boca Raton, FL: CRC Press.

Medical Devices Agency (UK) (2000) Guidelines for Magnetic Resonance Equipment in Clinical Use, 2nd edn.

Mukherji SR (ed.) (1998) *Clinical Applications of Magnetic Resonance Spectroscopy*. New York: Wiley.

Young IR (ed.) (2000) *Methods in Biomedical Magnetic Resonance Imaging and Spectroscopy*, vols. 1 and 2. New York: Wiley.

Van Hecke P and Van Huffel S (eds.) (2001) NMR spectroscopy quantitation, special issue NMR in Biomedicine 14(4).

## *In Vivo* Spectroscopy Using Localization Techniques

**A Bainbridge and E B Cady**, University College London Hospitals Trust, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

*In vivo* nuclear magnetic resonance (NMR) spectroscopy (abbreviated to magnetic resonance spectroscopy (MRS) for clinical purposes) allows the noninvasive assay of metabolite concentrations and provides information about their biophysical/biochemical environment (mainly intracellular). For many applications it is desirable that spectra should be obtained from a well-defined volume of tissue, with minimum contamination from signal originating elsewhere. There are a number of established localization techniques that are used clinically for *in vivo* MRS. Spectroscopic imaging, where spectra are obtained from a matrix of voxels (voxel is an element within a matrix containing spectroscopic information in MRSI), has also become increasingly common. Mainly proton ( $^1\text{H}$ ) and phosphorus ( $^{31}\text{P}$ ) studies are used in the clinical environment. This article describes the localization techniques that are commonly used, the factors affecting spectral quality, and gives some example of their use in *in vivo* MRS.

## Localization Techniques

### Image-Selected *In vivo* Spectroscopy

Image-selected *in vivo* spectroscopy (ISIS) is a multi-shot technique for localization; that is, localization is achieved by the application of several successive pulse sequences. Selective radiofrequency (RF) pulses are used to invert the magnetization in up to three mutually orthogonal slices. Following each inversion sequence a nonselective excitation pulse is applied followed by acquisition of the free induction decay (FID) (Figure 1). Individual FIDs are not localized to the volume of interest (VOI). Instead, the initial inversions modify the eventual phase of the magnetization in the transverse plane. Localization is achieved by acquiring data from eight separate experiments using different combinations of selective inversions. These data are then added or subtracted to each other in accordance with the protocol outlined in Table 1. This protocol ensures that signal from within the VOI adds constructively whereas signal from outside the VOI cancels; thus, localization is achieved. The FID is collected directly following the excitation pulse and so spin-spin relaxation and phase modulation effects are negligible. As a result, this technique has found particular application in  $^{31}\text{P}$  spectroscopy where phase modulation complicates observation of adenosine

ongoing developments of spectral editing, data analysis, localization, and RF pulse design enhance the utility of these NMR probes, provided by their high SNR.

*See also:* **Nuclear Magnetic Resonance Spectroscopy: Principles; Instrumentation. Nuclear Magnetic Resonance Spectroscopy-Applicable Elements:** Phosphorus-31. **Nuclear Magnetic Resonance Spectroscopy Techniques:** *In Vivo* Spectroscopy Using Localization Techniques.

## Further Reading

- Choi I-Y, Tkáč I, and Gruetter R (2000) Single-shot, three-dimensional “non-echo” localisation method for *in-vivo* NMR spectroscopy. *Magnetic Resonance in Medicine* 44: 387–394.
- Evelhoch JL, Crowley MG, and Ackerman JJH (1984) Signal-to-noise optimization and observed volume

localization with circular surface coils. *Journal of Magnetic Resonance* 56: 110–124.

- Gadian DG (1995) *Nuclear Magnetic Resonance and its Applications to Living Systems*, 2nd edn. Oxford: Oxford University Press.
- Garwood M and De la Barre L (2001) The return of the frequency sweep: Designing adiabatic pulses for contemporary NMR. *Journal of Magnetic Resonance* 153: 155–177.
- Igata A, Asakura T, and Fujimoto T (2000) *MRS of the Brain and Neurological Disorders*. Boca Raton, FL: CRC Press.
- Medical Devices Agency (UK) (2000) Guidelines for Magnetic Resonance Equipment in Clinical Use, 2nd edn.
- Mukherji SR (ed.) (1998) *Clinical Applications of Magnetic Resonance Spectroscopy*. New York: Wiley.
- Young IR (ed.) (2000) *Methods in Biomedical Magnetic Resonance Imaging and Spectroscopy*, vols. 1 and 2. New York: Wiley.
- Van Hecke P and Van Huffel S (eds.) (2001) NMR spectroscopy quantitation, special issue NMR in Biomedicine 14(4).

## *In Vivo* Spectroscopy Using Localization Techniques

**A Bainbridge and E B Cady**, University College London Hospitals Trust, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

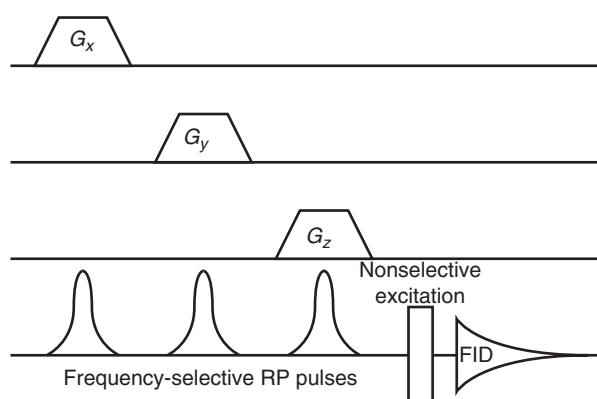
## Introduction

*In vivo* nuclear magnetic resonance (NMR) spectroscopy (abbreviated to magnetic resonance spectroscopy (MRS) for clinical purposes) allows the noninvasive assay of metabolite concentrations and provides information about their biophysical/biochemical environment (mainly intracellular). For many applications it is desirable that spectra should be obtained from a well-defined volume of tissue, with minimum contamination from signal originating elsewhere. There are a number of established localization techniques that are used clinically for *in vivo* MRS. Spectroscopic imaging, where spectra are obtained from a matrix of voxels (voxel is an element within a matrix containing spectroscopic information in MRSI), has also become increasingly common. Mainly proton ( $^1\text{H}$ ) and phosphorus ( $^{31}\text{P}$ ) studies are used in the clinical environment. This article describes the localization techniques that are commonly used, the factors affecting spectral quality, and gives some example of their use in *in vivo* MRS.

## Localization Techniques

### Image-Selected *In vivo* Spectroscopy

Image-selected *in vivo* spectroscopy (ISIS) is a multi-shot technique for localization; that is, localization is achieved by the application of several successive pulse sequences. Selective radiofrequency (RF) pulses are used to invert the magnetization in up to three mutually orthogonal slices. Following each inversion sequence a nonselective excitation pulse is applied followed by acquisition of the free induction decay (FID) (Figure 1). Individual FIDs are not localized to the volume of interest (VOI). Instead, the initial inversions modify the eventual phase of the magnetization in the transverse plane. Localization is achieved by acquiring data from eight separate experiments using different combinations of selective inversions. These data are then added or subtracted to each other in accordance with the protocol outlined in Table 1. This protocol ensures that signal from within the VOI adds constructively whereas signal from outside the VOI cancels; thus, localization is achieved. The FID is collected directly following the excitation pulse and so spin-spin relaxation and phase modulation effects are negligible. As a result, this technique has found particular application in  $^{31}\text{P}$  spectroscopy where phase modulation complicates observation of adenosine



**Figure 1** ISIS. In eight separate acquisitions combinations of up to three frequency-selective RF pulses invert the longitudinal magnetization in three orthogonal slices prior to a nonselective excitation pulse and collection of the free induction decay. The inversion pulses modify the resultant phase of the transverse magnetization existing after excitation. For three-dimensional localization, the inversion pulses are applied and the data added or subtracted according to the protocol in **Table 1**.

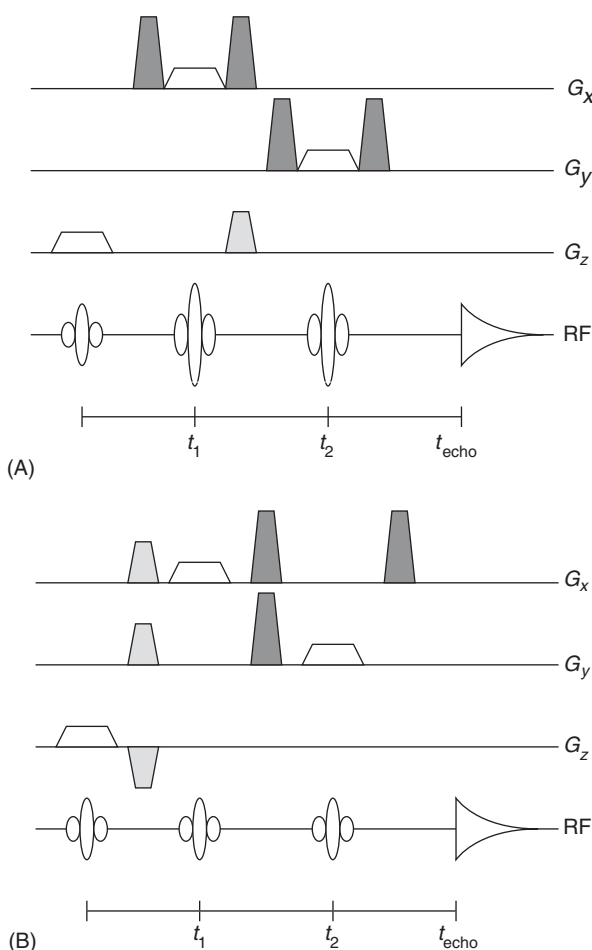
**Table 1** Details of the slice inversion scheme and combination of data for a three-dimensional ISIS localization

Experiment	X-slice inversion	Y-slice inversion	Z-slice inversion	Add/subtract data
1	Yes	No	Yes	Add
2	No	Yes	No	Subtract
3	No	Yes	Yes	Add
4	Yes	No	No	Subtract
5	No	No	Yes	Subtract
6	Yes	Yes	No	Add
7	Yes	Yes	Yes	Subtract
8	No	No	No	Add

triphosphate (ATP). However, because ISIS is a multishot technique, subject motion or receiver instability can degrade localization efficiency resulting in contamination from signal originating outside the VOI.

### Point Resolved Spectroscopy and Stimulated-Echo Acquisition Mode

Point resolved spectroscopy (PRESS) and stimulated-echo acquisition mode (STEAM) use three selective RF pulses in combination with linear field gradients to yield localization of the VOI in a single shot. Thus, the localization provided by these techniques is not degraded by subject motion or system instability. Generally, the use of three RF pulses leads to the formation of five coherent pathways for echo formation. Both PRESS and STEAM are designed to yield a particular echo that is localized to the VOI; the other echoes are suppressed by the use of spoiler gradients and phase cycling of the RF pulses. The



**Figure 2** PRESS (A) and STEAM (B). Both methods use a combination of three frequency-selective RF pulses with linear field gradients in three orthogonal directions to yield VOI localization. Slice selection gradients are unshaded; trim gradients, which ensure zero net de-phasing of transverse magnetization due to the selection gradients, are lightly shaded. A VOI echo is collected during data acquisition (Acq); other echoes are suppressed by phase cycling the RF pulses and spoiler gradients (heavy shading). (A) PRESS uses selective  $180^\circ$  refocusing pulses at times  $t_1$  and  $t_2$ , following an initial excitation, to yield a spin-echo at time  $2t_2 - t_1$ . (B) STEAM uses selective excitation pulses at times  $t_1$  and  $t_2$ , following an initial excitation, to yield a stimulated echo at time  $t_1 + t_2$ . No spin-spin relaxation occurs during the mixing time (TM) – hence  $TE = 2t_1$ .

PRESS sequence consists of an initial excitation pulse (optimally  $90^\circ$  if the magnetization is allowed to fully relax between acquisitions), two refocusing ( $180^\circ$ ) pulses at times  $t_1$  and  $t_2$  after the excitation pulse, and the acquisition of a spin echo at echo time  $(TE) = 2t_2 - t_1$  (**Figure 2A**). The STEAM sequence consists of an initial excitation pulse, two  $90^\circ$  pulses at times  $t_1$  and  $t_2$  after the excitation pulse and the acquisition of a stimulated echo at  $t_1 + t_2$  (**Figure 2B**). In STEAM, the period between the second and third pulses is often termed TM – the ‘mixing’ time

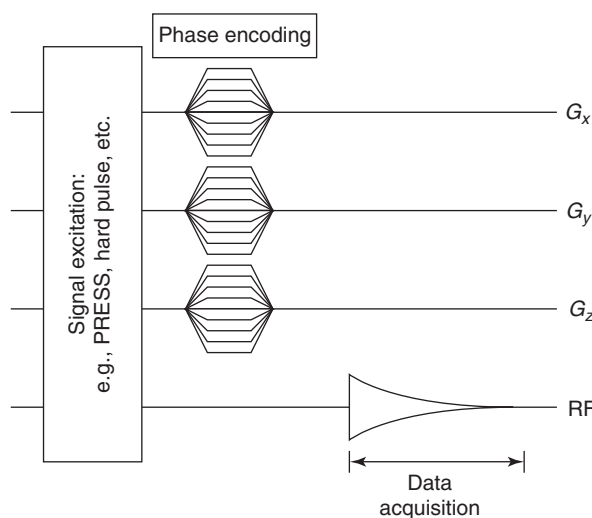
(see Figure 2B); spin-spin relaxation does not occur during TM and hence,  $TE = 2t_1$ . Both PRESS and STEAM are subject to spin-spin relaxation of signal and phase modulation effects. Typically, these sequences are used for  $^1\text{H}$  MRS because the  $^1\text{H}$  resonances of mobile metabolites have relatively long  $T_2$ s. However, short  $TE$ s of  $\sim 10$  ms are easily achievable with PRESS and with STEAM even shorter effective  $TE$ s are achievable. Because phase modulation effects are different for PRESS and STEAM the metabolites of interest are leading determinants of which sequence and  $TE$  to use. Short  $TE$   $^1\text{H}$  spectra have a better signal-to-noise ratio (SNR) and less phase modulation effects than long  $TE$  spectra; the former are most useful for studies where absolute quantification of metabolite concentrations is required. However, signals from macromolecules and lipids, which have short  $T_2$ s, produce an undulating baseline in short  $TE$  *in vivo* spectra, which can greatly complicate the spectral analysis. For this reason, it is often desirable to use longer  $TE$ s in order to simplify spectra. PRESS has an intrinsic two-to-one advantage in SNR over STEAM because of the collection of a spin echo. However, at a given  $TE$  and for a particular multiplet resonance of interest, phase modulation effects may reduce the SNR advantage of PRESS.

### Outer-Volume Suppression

Instead of exciting only the spins in the target VOI, a signal can be localized by selective suppression of the signal from the rest of the sample. Selective RF pulses are used to flip longitudinal magnetization in the regions outside the VOI into the transverse plane. This transverse magnetization is then de-phased by the application of spoiler gradients so that no net magnetization remains outside the VOI. After the application of six appropriately positioned and orientated suppression slices longitudinal magnetization remains unaffected within the VOI and localized spectroscopic information can be acquired by using a simple nonselective excitation followed by collection of the FID. Like PRESS and STEAM, outer-volume suppression (OVS) is a single-shot technique and is, therefore, resilient to subject movement and system instability. However, because data are acquired with a single nonselective excitation pulse, OVS, like ISIS, has advantages for the collection of data from short  $T_2$  metabolites. Suppression of longitudinal magnetization outside the VOI must be extremely efficient to avoid significant contamination of the resulting spectrum. OVS can also be used in combination with most other localization methods (e.g., OSIRIS combines ISIS with OVS) thereby greatly improving localization efficiency.

### Magnetic Resonance Spectroscopic Imaging

The techniques described above are generally used to obtain data from a single VOI. Magnetic resonance spectroscopic imaging (MRSI) has the advantage of generating a matrix of voxels, each containing spectroscopic information. Thus, the spatial distribution of any metabolic abnormalities can be assessed from a single MRSI data acquisition. Furthermore, post-processing of the spectra can yield maps of the concentration of individual metabolites, metabolite ratios, or any other measurable (e.g., intracellular pH ( $\text{pH}_i$ )) and so morphological magnetic resonance imaging (MRI) can be correlated to maps of tissue metabolism. MRSI data are acquired by using a single-shot pulse sequence with the addition of spatial phase-encoding gradients. A simple pulse-acquire acquisition scheme is easily modified with the addition of phase-encoding gradients in three orthogonal directions – three-dimensional MRSI (Figure 3). Fourier transforming, the  $k$ -space data along the phase-encoding dimensions, leaves a three-dimensional matrix of FIDs; a further Fourier transform yields the final matrix of spectra. The data acquisition can be restricted to a VOI defined by using slice selection or a voxel localization technique, such as PRESS or STEAM, in place of the nonselective excitation pulse in the above example. The matrix of voxels can then be restricted to one or two dimensions by phase encoding in only one or two directions if required and thus be tailored to suit the morphology of the tissue being studied. Reducing the



**Figure 3** MRSI. Phase-encoding gradients localize spectra to a matrix of voxels within a volume of tissue. A nonselective hard pulse may be used to excite the entire sample prior to phase encoding or a localization technique, such as PRESS, can restrict data acquisition to a VOI to reduce lipid contamination. Phase encoding may be restricted to one or two dimensions if required.

number of phase-encoding steps also reduces the total acquisition time.

The acquisition details of an MRSI experiment are predominantly governed by the sensitivity of the imaging apparatus to the nucleus of interest and the practical limitations of time imposed on the experiment due to the tolerance of the subject or the time-scale of the dynamic metabolic processes that are being monitored. The low intrinsic sensitivity of the  $^{31}\text{P}$  nucleus and the fact that only single nuclei contribute to most resonances means that  $^{31}\text{P}$  voxels must be substantially larger than can be used for  $^1\text{H}$ . A typical  $^{31}\text{P}$  matrix would consist of  $16 \times 16 \times 1$  voxels each with a volume of  $\sim 8\text{--}27\text{ cm}^3$ . The greater sensitivity of  $^1\text{H}$  allows better spatial resolution; typically  $32 \times 32 \times 1$  voxels each with a volume  $\sim 1\text{--}8\text{ cm}^3$ . However,  $^1\text{H}$  matrix sizes are still small, and the spatial resolution consequently worse, compared to standard MRI techniques. As a result of the small number of phase-encoding steps the size of the point spread function (PSF) is increased; the PSF is a sinc function and widened side-bands can result in serious voxel contamination. In order to minimize this effect, a filter (e.g., Hanning) is usually applied to the data, although at the cost of decreasing the effective spatial resolution.  $^1\text{H}$  MRSI presents further challenges: water suppression and the avoidance of contributions from subcutaneous fat. Good water suppression depends upon good shimming (i.e., homogenization of the static magnetic field) throughout the matrix; this consideration is dealt with in the next section. Lipid signal can be avoided by restricting the data acquisition to regions within the organ of interest – for example, by phase encoding within a PRESS VOI.

## Factors Affecting Localization and Spectral Quality

### Sensitivity – Field Strength and Coil Performance

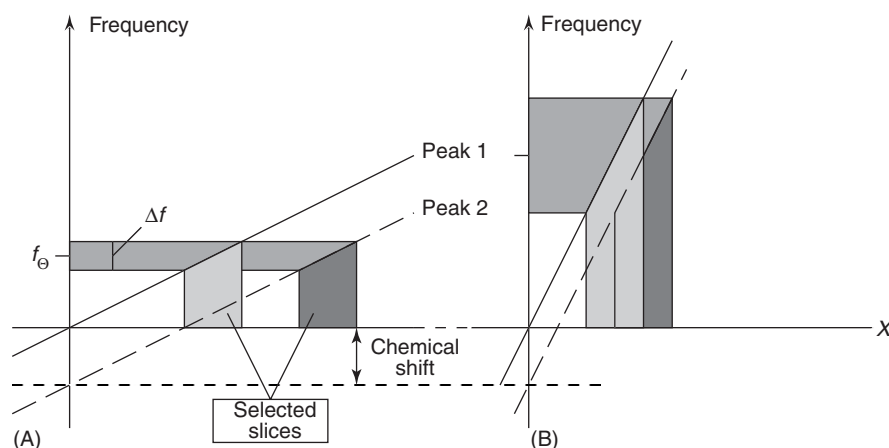
One of the most significant factors affecting spectral quality is the sensitivity to the nucleus of interest. The maximum available signal is proportional to almost the square of the main magnetic field strength and it is this factor that principally governs the sensitivity of the experiment. Increasing the field strength from 1.5 to 3 T allows for the same level of signal to be obtained either from about a quarter the volume of tissue or in substantially less scanning time. Increasing field strength generally requires a new magnet and, of course, this is an expensive option. Contributions to the RF noise component of the detected signal are important. These come from various sources, but the dominant contributor is the inductive coupling between the coil and the sample.

The whole of the sample within the sensitive volume of the coil contributes to the noise in the spectrum whereas localized signal originates only from the VOI. An appropriate RF coil design is important to the success of the investigation. Among the factors that influence the optimal coil design are the location and size of the VOI within the sample and the choice of localization methodology. Volume coils that can accommodate large parts of the human body have the advantage of a relatively homogeneous  $B_1$  field over their sensitive volume. Surface coils, in comparison, have a limited sensitive volume, significant  $B_1$  gradients, and are most useful for peripheral applications. However, in certain appropriate applications, surface coils typically yield localized spectra with a greater SNR than that obtained using volume coils – but to achieve this the target VOI must be close to the surface. The main disadvantage of a surface coil is the highly inhomogeneous  $B_1$  field. Unless a VOI is small compared to the dimensions of the coil, there will be a large variation in both flip angle and sensitivity over the VOI. As a consequence, phase modulation effects and signal strength will vary leading to difficulties in absolute quantitation. Furthermore, surface-coil  $B_1$  inhomogeneity may contraindicate the use of larger VOIs because a particular localization technique may require exact inversion pulses. Water suppression is essential for  $^1\text{H}$  MRS in order to digitize the low molarity metabolite signals –  $B_1$  inhomogeneity can drastically reduce the efficiency of water suppression schemes. However, these problems can be addressed by the use of adiabatic RF pulses, which are discussed in a later section.

### Chemical Shift Displacement Artifact

Localization of the NMR signal is achieved through the use of magnetic field gradients to encode spatial information in the frequency of the RF signal. Signals from nuclei in various metabolites resonate at different RFs: these are termed ‘chemical shifts’ and for simplicity are expressed as parts per million of the static magnetic field strength (ppm). A consequence of metabolite signals occurring at discrete chemical shifts is that each resonance will originate from its own discrete VOI, which will be displaced spatially with respect to VOI for resonances at other chemical shifts (Figure 4). This is called the ‘chemical shift displacement artifact’ (CSDA) and this effect has important repercussions for most localized spectroscopy methods. Spectrum analysis often involves calculating the ratio of two particular metabolites (e.g., lactate (Lac)/N-acetylaspartate (Naa)). Information is acquired from two VOIs, one for each



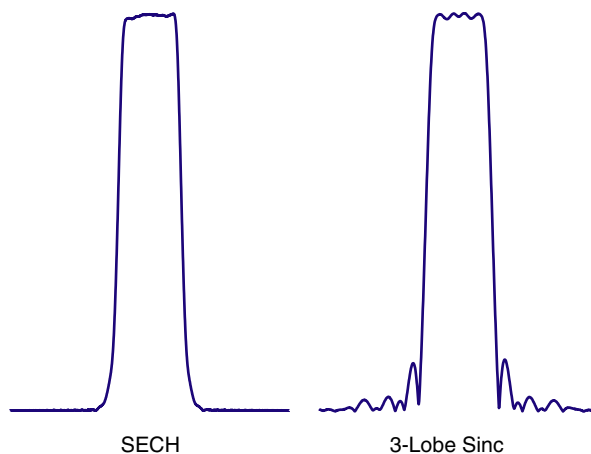


**Figure 4** The chemical shift displacement artifact. (A) An applied linear magnetic gradient encodes spatial position in the resonant frequencies of two particular spectrum peaks (represented by the sloping lines). Peak 1 has a different chemical shift to peak 2. A selective RF pulse, centered on frequency  $f_0$  and with bandwidth  $\Delta f$ , will excite a slice at a different position for each peak as shown. (B) Increasing the strength of the linear magnetic gradient reduces the difference in slice position – however, to achieve slices with the same spatial width as in (A), a larger bandwidth RF pulse must be used.

metabolite, which occupy different spatial positions and the measured ratio may be unrepresentative of the true value in the nominal VOI. This problem is accentuated for  $^{31}\text{P}$  spectra where the chemical shift range across the spectrum is  $\sim 25$  ppm resulting in much greater spatial displacements of the VOI. CSDA is not a problem in three-dimensional MRSI unless spatially selective pulses are used. However, whenever slice or volume selection is employed for MRSI, the position of the slice or volume is chemical-shift dependent. CSDA can inhibit the effects of phase modulation that will only occur where the discrete VOIs for the two coupled resonances intersect. A further consequence of CSDA when used with a surface coil is that, due to inhomogeneous  $B_1$ , signal amplitude will change with chemical shift because of the variable VOI location.

### Design of RF Pulses

Some of the problems of  $B_1$  inhomogeneity and CSDA can be addressed with the use of specially designed RF pulses. Adiabatic pulses, and in particular a subset known as hyperbolic-secant (SECH) pulses, have desirable properties for many spectroscopic applications. A ‘full-passage’ SECH pulse modulates the frequency as well as the amplitude of the RF such that longitudinal magnetization is fully inverted at the end of the pulse, regardless of  $B_1$  inhomogeneity within certain thresholds. In order for these pulses to work, an adiabatic condition must be met, which places a lower limit on the RF power required. However, with careful design SECH pulses can be used *in vivo* without exceeding RF power



**Figure 5** Examples of the frequency profiles of a SECH pulse and a three-lobe sinc pulse. The SECH pulse has far less prominent side-bands than the sinc. The profiles were generated on a 2.4 T Bruker Avance Scanner at University College London.

deposition limits. The ability to perform an adiabatic inversion with an inhomogeneous  $B_1$  field is beneficial when using ISIS with a surface coil; good inversion across the whole VOI greatly improves the SNR. The frequency bandwidth of an SECH pulse is modulated by the frequency sweep within the pulse and has a sharply defined profile with minimal side-bands compared to conventional sinc or Gaussian shaped pulses (Figure 5). Pulses that have very large bandwidths can be designed (10 kHz or more). The larger the bandwidth of the pulse, the larger the magnetic field gradient required to localize a VOI of a given size. Large bandwidth RF pulses require large magnetic field gradients and the latter are advantageous

in reducing CSDA. The main limitation in using adiabatic pulses is not generally the maximum available magnetic gradient strength but rather the maximum RF power that can be safely applied to the subject; pulses designed with larger frequency sweeps and consequently broader bandwidths require higher RF powers to achieve the adiabatic condition. A solution is to use frequency offset corrected inversion (FOCI) pulses. FOCI pulses utilize time-dependent shaped magnetic gradient waveforms during application of an SECH pulse. Spin inversion occurs at different times for different frequencies. The gradient waveform is shaped so that it is maximum when slice edges are being inverted. Thus, pulses require smaller bandwidths and hence lower powers. As a result of the use of adiabatic pulses with surface coils the effects of many pulse sequences are largely CSDA-independent and leading to less chemical shift dependence of signal strength.

### Localized Shimming and Water Suppression

*In vivo* spectrum quality is strongly influenced by the homogeneity of the static magnetic field within the VOI. Shim coils are used to produce shaped magnetic fields that cancel inhomogeneities in the main field. This process of shimming can greatly improve the chemical-shift resolution of *in vivo* spectra. Spectral resolution is optimized if the actual VOI is shimmed. For single-shot techniques such as PRESS and STEAM, it is only necessary to monitor the  $^1\text{H}$  water signal whilst adjusting the shim-coil currents. For shimming ISIS VOIs, it is necessary to use either OVS, to restrict the signal to the VOI, or to use a single-shot technique. Shimming can be performed manually, where the operator adjusts the currents in the shim coils whilst monitoring the  $^1\text{H}$  water signal; the aim is to minimize the rate of exponential decay of the FID and, hence, reduce static magnetic field inhomogeneities. An alternative is to use an automated technique. A commonly used technique is FASTMAP, which maps the magnetic field along six projections within the defined VOI. Using these data, the required current amplitudes in the shim coils are calculated analytically.

A good VOI shim is a prerequisite to addressing a further problem in  $^1\text{H}$  spectroscopy: the strong signal from tissue water: it is difficult to adequately digitize the signals from low molarity ( $\sim 10 \text{ mmol l}^{-1}$ ) metabolites in the presence of a large signal from tissue water ( $\sim 45 \text{ mol l}^{-1}$ ). Localization sequences are often preceded by chemical-shift-selective (CHESS) pulses, which are used in conjunction with spoiling gradients to selectively saturate the water signal.  $B_1$ -insensitive techniques for water suppression such as

BISTRO, which uses specially designed adiabatic pulses, are also available and are of particular use in surface coil studies.

### Line-Shape Corrections

Gradient-pulse-induced eddy currents can distort the line shapes of spectral peaks, particularly in short TE spectra. This problem can make spectra difficult to analyze because analytical fitting functions may not be appropriate. The use of high-performance gradient coils with very short rise and fall times can minimize eddy currents. However, postprocessing of spectra, using a phase-correction derived from an unsuppressed water peak, can improve the quality of spectra where eddy-current-induced distortion is apparent.

### Absolute Quantification of Metabolite Concentrations

*In vivo* spectra are often analyzed to give metabolite peak-area ratios: these can be simply and even automatically obtained from available analysis packages. However, rigorous interpretation of an abnormal peak-area ratio is difficult if the observer is unaware which metabolite concentration has changed – or maybe both have changed but by differing amounts. Furthermore, unless data are acquired fully relaxed (repetition time  $> 5T_1$ ) with very short TE (or by ISIS or OVS) peak-area ratios can be modulated by changes in  $T_1$  or  $T_2$  relaxation ( $T_1$ : spin-lattice relaxation time – characteristic time for recovery of longitudinal magnetization and  $T_2$ : spin-spin relaxation time – characteristic time for de-phasing of magnetization in the transverse plane). Absolute quantification of metabolite concentrations has specific requirements in order to be successful: the *in vivo* signal amplitude must be normalized for VOI size and numbers of resonating nuclei per resonance, a stable concentration reference (external or internal) must be available, differences in signal amplitude due to variations in coil loading must be correctable, and, with surface coils, variations in signal amplitude resulting from  $B_1$  inhomogeneity and CSDA must be addressed.

In brain studies a significant part of the VOI may be occupied by cerebrospinal fluid (CSF), which is largely water and contains only small quantities of detectable metabolites. In order to determine the VOI fraction occupied by brain tissue, detailed three-dimensional MR imaging datasets may be used. Alternatively, since CSF  $T_2$  is long compared to that of brain water, measuring the un-suppressed water signal as a function of TE allows separation of the brain-water and CSF components and determination of their individual VOI fractions.

A commonly used internal concentration reference (ICR) is the tissue water itself. Reference spectra (without water suppression) are acquired from the same VOI as the metabolite spectra; factors such as the sensitivity profile across the VOI and the degree of coil loading are thus the same for both spectra – hence, there is mutual cancellation of many factors that would otherwise need to be taken into account. However, the concentration of the ICR must be known *a priori* and pathological variations can lead to inaccuracy. If an external concentration reference (ECR) is used, this can be positioned adjacent to the subject and within the sensitive volume of the RF coil. In this configuration, coil loading for the ECR and subject are identical. However, RF coil sensitivity may be significantly different at the ECR and the VOI. A further approach is to acquire reference data from the ECR in a separate experiment. The coil loading with just the ECR in the RF coil will often be different to the *in vivo* situation and a correction may be necessary. A loading phantom (typically containing a high concentration of sodium chloride) can be introduced in order to match the loading conditions between the two experiments. An alternative is to use the principle of reciprocity, which implies that for a given spin density within a particular VOI the product of the pulse amplitude for a given excitation flip angle and the consequent signal amplitude is a constant. This principle can be used to correct for differences in coil loading between *in vivo* and ECR experiments. The optimum voltage of a water suppression CHESS pulse can be defined much more precisely than, e.g., an optimal excitation pulse and the former is thus a better index of coil loading. When an optimal water-suppression pulse voltage,  $V_1$ , is applied, the water magnetization is rotated through an angle  $\phi_1 = 90^\circ + x$ . A second optimal pulse voltage for suppression,  $V_2$ , exists corresponding to a flip angle of  $\phi_2 = 270^\circ - x$ . Since  $V_1 = k\phi_1$  and  $V_2 = k\phi_2$  where  $k$  is the proportionality constant for the VOI concerned:

$$\phi_1 = \frac{360^\circ}{(1 + (V_2/V_1))}$$

$V_1$  and  $V_2$ , and hence  $\phi_1$ , can be accurately determined during *in vivo* and ECR experiments, thereby enabling precise correction of the effects of coil loading on the signal amplitude. The metabolite concentration,  $n_m$ , can then be calculated by comparing the metabolite peak area,  $A_m$ , to the ECR peak area,  $A_c$  (concentration  $n_c$ ):

$$n_m = n_c \cdot \frac{A_m}{A_c} \cdot \frac{f_c}{f_m} \cdot \frac{T_m}{T_c} \cdot C_{\text{loading}}$$

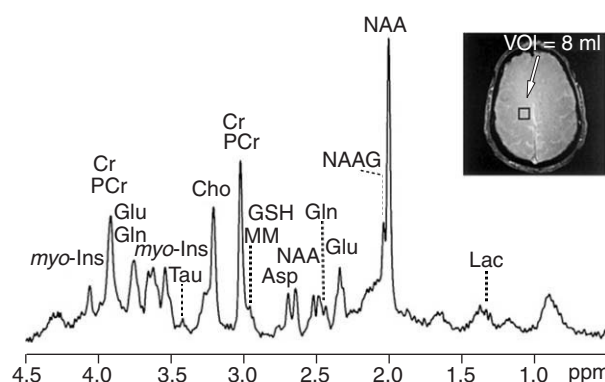
where  $C_{\text{loading}}$  is the correction for coil loading:

$$C_{\text{loading}} = \frac{V_{1m}}{V_{1c}} \cdot \frac{\phi_{1c}}{\phi_{1m}}$$

The subscripts c and m correspond to the concentration reference and the *in vivo* metabolite respectively,  $f$  is the number of equivalent nuclei contributing to the ECR/metabolite signal, and  $T$  is the *in vivo*/ECR temperature. The *in vivo* and phantom VOI volumes are assumed equal.

## Applications

Because of its noninvasive nature, MRS is especially suited to *in vivo* studies of brain. Numerous cerebral metabolites are detectable, particularly with  $^1\text{H}$  MRS (Figure 6). A few metabolites contribute prominent peaks and these are most easily quantifiable. Naa is predominantly found in the central nervous system, especially in neurons, although its functional role is not fully understood. Dynamic changes in brain Naa concentration have been observed; reduced concentration may indicate neuronal dysfunction as well as loss. A prominent peak due to choline (Cho) has contributions from several metabolites including phosphorylcholine and glycerophosphorylcholine. Phosphatidylcholine is synthesized from choline and is a major constituent of cell membranes; being bound, this component is NMR invisible due to a very short  $T_2$ . Pathologically increased membrane



**Figure 6** A  $^1\text{H}$  MRS spectrum from normal human brain at 7 T. An 8 ml VOI was localized in the parietal white matter using STEAM with effective TE 6 ms, repetition time 5 s, TM 32 ms, and the collection of 160 averaged echoes. Outer volume suppression, employing SECH pulses, was used and FASTMAP was used to shim the VOI. Peaks undefined in the main text are: *myo*-inositol – *myo*-Ins; taurine – Tau; glutathione – GSH; aspartate – ASP; *N*-acetylaspartylglutamate – NAAAG; macromolecule – MM. (Reprinted with permission from Tkáč I, Anderson P, Adriany G, *et al.* (2001) *In vivo*  $^1\text{H}$  NMR spectroscopy of the human brain at 7 Tesla. *Magnetic Resonance in Medicine* 46: 451–456; © Wiley-Liss, a subsidiary of Wiley.)

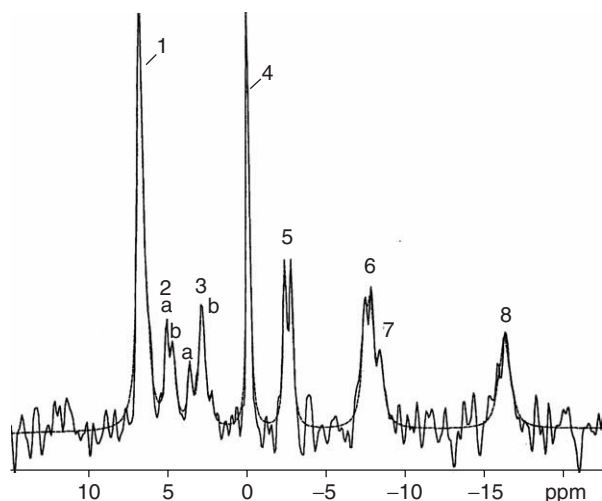
turnover is thought to result in raised levels of NMR visible Cho. Creatine (Cr) is synthesized in the liver and is phosphorylated to phosphocreatine (PCr) by creatine kinase. In  $^1\text{H}$  MRS a co-resonating Cr and PCr signal is seen and total Cr (i.e.,  $\text{Cr}_{\text{total}} = \text{Cr} + \text{PCr}$ ) has often been used as a reference for the comparison of peak-area ratios such as  $\text{Naa}/\text{Cr}_{\text{total}}$ . However, care must be taken when interpreting such data since neither the  $\text{Cr}_{\text{total}}$  concentration nor the effective  $T_2$  of the co-resonating Cr and PCr components may remain constant with respect to pathology. The Lac concentration is too low to be easily detected in healthy adult brain tissue but Lac is detectable in term and preterm neonates. Lac appears in increased concentrations in pathologies that lead to failed oxidative metabolism or increased glycolysis. Measurements of the concentrations of neurotransmitters such as glutamate, glutamine, and  $\gamma$ -aminobutyric acid (GABA) also have use in studies of various neurological conditions.

$^{31}\text{P}$  *in vivo* spectra are less complex (Figure 7). ATP is the principal donor of free energy in biological systems: the Na/K pump, responsible for maintaining intracellular water content/nerve signal transmission, and muscle contraction are all dependent on ATP. When involved in these vital processes ATP is hydrolyzed to adenosine diphosphate (ADP) and inorganic phosphate (Pi) during which a large

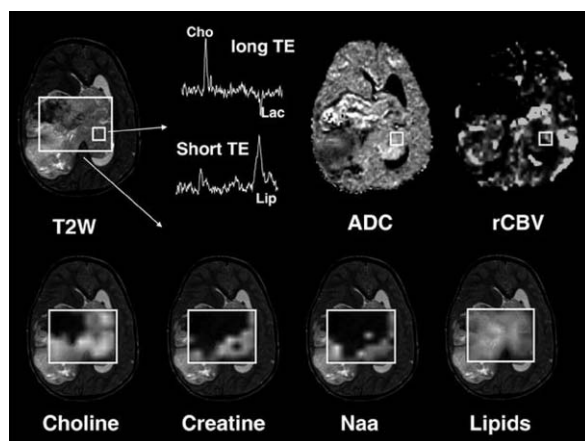
amount of free energy is liberated. In addition to generation by other metabolic pathways, if levels fall ATP can be rapidly regenerated by rephosphorylation of ADP by PCr. Three prominent ATP multiplet resonances are seen:  $\gamma$ -,  $\alpha$ -, and  $\beta$ -ATP (two doublets and a triplet, respectively). PCr and Pi are detected as prominent singlet peak: the Pi chemical shift is pH dependent.  $^{31}\text{P}$  MRS thus provides a window on tissue energy metabolism. Phosphomonoesters (PME) and phosphodiesteres (PDE) are also visible and are related to membrane synthesis.

A good example of the prognostic and diagnostic utility of both  $^1\text{H}$  and  $^{31}\text{P}$  MRS is their use in perinatal asphyxia. A significantly increased Lac/Naa peak-area ratio and/or reduced PCr/Pi in the first few days after birth is associated with impaired neurological development with disability at age 1 year.

Further applications of MRS allow the identification and grading of cancer tissue; brain tumors often display increased Cho, reduced Naa, and the presence of Lac and mobile lipids. The use of MRSI is particularly advantageous; tumors are often heterogeneous in structure and MRSI offers the ability to analyze different parts of a lesion separately (Figure 8). Differentiation between certain brain tumors and between infiltrative and circumscribed lesions can be made based on changes in metabolite ratios across the lesion and in surrounding tissue. Voxels may also be added together to match the anatomical shape of the lesion.



**Figure 7** A  $^{31}\text{P}$  MRS spectrum from neonatal brain at 2.4 T. A 125 ml VOI, localized using PRESS, was centered on the thalami. Acquisition parameters were: TE 10 ms; repetition time 12 s; 160 averaged echoes. The numbered peaks are: 1 PME; 2 extracellular and intracellular Pi; 3 glycerolphosphorylethanolamine and GPC; 4 PCr; 5, 6, and 8  $\gamma$ -,  $\alpha$ -, and  $\beta$ -NTP; 7 nicotinamide dinucleotides and uridine diphosphosugars. (Reproduced with permission from Cady EB, Wylezinska M, Penrice J, Lorek A, and Amess P (1996) Quantitation of phosphorus metabolites in newborn human brain using internal water as a reference standard. *Magnetic Resonance Imaging* 14(3): 293–304; © Elsevier.)



**Figure 8**  $^1\text{H}$  MRSI of brain. The T2-weighted image shows a large brain tumor in a 4-year-old boy. The large white rectangle denotes the MRSI VOI preselected using PRESS. Metabolite maps of Cho, Cr, Naa, and lipids are shown. Necrotic regions are characterized by a lack of metabolites; Cho and/or Cr are seen in regions of viable tumor. Also shown are maps of apparent diffusion coefficient (ADC) and cerebral blood flow (rCBF). (Reproduced with permission from Tzika AA, Astrakas LG, Zarifi MK, et al. (2003) Multiparametric MR assessment of pediatric brain tumours. *Neuroradiology* 45: 1–10; © Springer.)

Resonance attributes can be modulated by intracellular biophysical conditions. For example,  $^{31}\text{P}$  MRS can be used to measure  $\text{pH}_i$  by measuring the chemical shifts of Pi and other peaks (e.g., ATP). Alkaline  $\text{pH}_i$  is indicative of a poor prognosis following perinatal asphyxia.  $^1\text{H}$  spectra can be used to measure local brain temperature by comparing the chemical shift of the un-suppressed water peak, which has a relatively strong temperature dependence due to the hydrogen bonding between molecules, with those of Naa, Cho, and Cr. This property may be combined with MRSI to yield temperature maps of the brain.

*See also:* **Nuclear Magnetic Resonance Spectroscopy: Principles. Nuclear Magnetic Resonance Spectroscopy-Applicable Elements: Phosphorus-31. Nuclear Magnetic Resonance Spectroscopy Techniques: Surface Coil.**

### Further Reading

- Bachelard H (ed.) (1997) *Magnetic Resonance Spectroscopy and Imaging in Neurochemistry*, Advances in neurochemistry, vol. 8. New York: Plenum Press.
- Cady EB, D'Souza PC, Penrice J, and Lorek A (1995) The estimation of local brain temperature by *in vivo*  $^1\text{H}$  magnetic resonance spectroscopy. *Magnetic Resonance in Medicine* 33: 862–867.
- Choi I-Y, Tkáč I, and Gruetter R (2000) Single-shot, three-dimensional 'non-echo' localisation method for *in-vivo* NMR spectroscopy. *Magnetic Resonance in Medicine* 44: 387–394.
- Danielsen ER and Henriksen O (1994) Absolute quantitative proton NMR spectroscopy based on the amplitude of the local water suppression pulse. Quantification of brain water and metabolites. *NMR in Biomedicine* 7: 311–318.
- Ernst T and Hennig J (1991) Coupling effects in volume-selective  $^1\text{H}$  spectroscopy of major brain metabolites. *Magnetic Resonance in Medicine* 21: 82–96.
- Govindaraju V, Young K, and Maudsley AA (2000) Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR in Biomedicine* 13: 129–153.
- Keevil SF, Barbiroli B, Brooks JCW, *et al.* (1998) Absolute metabolite quantification by *in vivo* NMR spectroscopy: II. A multicentre trial of protocols for *in vivo* localised proton studies of human brain. *Magnetic Resonance Imaging* 16: 1093–1106.
- Klose U (1990) *In-vivo* proton spectroscopy in the presence of eddy currents. *Magnetic Resonance in Medicine* 14: 26–30.
- Nelson SJ, Vigneron DB, Star-Lack J, and Kurhanewicz J (1997) High spatial resolution and speed in MRSI. *NMR in Biomedicine* 10: 411–422.
- Tkac I, Anderson P, Andriany G, *et al.* (1995) Quality Assessment in *in vivo* NMR spectroscopy: Results of a concerted research project of the European Economic Community. *Magnetic Resonance Imaging* 13: 117–176.
- Young IR (ed.) (2002) *Methods in Biochemical Magnetic Resonance Imaging and Spectroscopy*. Chichester: Wiley.

## NUCLEIC ACIDS

### Contents

**Chromatographic and Electrophoretic Methods**  
**Spectroscopic Methods**  
**Electrochemical Methods**  
**Immunoassays**  
**mRNA Identification and Quantification**

## Chromatographic and Electrophoretic Methods

**F Foret and K Klepárník**, Institute of Analytical Chemistry, Brno, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Michael Thompson and Hongbo Su, pp. 3609–3615, © 1995, Elsevier Ltd.

### Introduction

Techniques for separation of DNA and RNA provide powerful tools for the analysis of genetic material and transcription products. Modern methods of analysis can answer most of the questions about the sequence, nucleotide substitutions (mutations, polymorphisms, or single nucleotide polymorphisms – SNPs), and structure of nucleic acids. Crucial techniques in the analysis of DNA or RNA are the

Resonance attributes can be modulated by intracellular biophysical conditions. For example,  $^{31}\text{P}$  MRS can be used to measure  $\text{pH}_i$  by measuring the chemical shifts of Pi and other peaks (e.g., ATP). Alkaline  $\text{pH}_i$  is indicative of a poor prognosis following perinatal asphyxia.  $^1\text{H}$  spectra can be used to measure local brain temperature by comparing the chemical shift of the un-suppressed water peak, which has a relatively strong temperature dependence due to the hydrogen bonding between molecules, with those of Naa, Cho, and Cr. This property may be combined with MRSI to yield temperature maps of the brain.

*See also:* **Nuclear Magnetic Resonance Spectroscopy: Principles. Nuclear Magnetic Resonance Spectroscopy-Applicable Elements: Phosphorus-31. Nuclear Magnetic Resonance Spectroscopy Techniques: Surface Coil.**

### Further Reading

- Bachelard H (ed.) (1997) *Magnetic Resonance Spectroscopy and Imaging in Neurochemistry*, Advances in neurochemistry, vol. 8. New York: Plenum Press.
- Cady EB, D'Souza PC, Penrice J, and Lorek A (1995) The estimation of local brain temperature by *in vivo*  $^1\text{H}$  magnetic resonance spectroscopy. *Magnetic Resonance in Medicine* 33: 862–867.
- Choi I-Y, Tkáč I, and Gruetter R (2000) Single-shot, three-dimensional 'non-echo' localisation method for *in-vivo* NMR spectroscopy. *Magnetic Resonance in Medicine* 44: 387–394.
- Danielsen ER and Henriksen O (1994) Absolute quantitative proton NMR spectroscopy based on the amplitude of the local water suppression pulse. Quantification of brain water and metabolites. *NMR in Biomedicine* 7: 311–318.
- Ernst T and Hennig J (1991) Coupling effects in volume-selective  $^1\text{H}$  spectroscopy of major brain metabolites. *Magnetic Resonance in Medicine* 21: 82–96.
- Govindaraju V, Young K, and Maudsley AA (2000) Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR in Biomedicine* 13: 129–153.
- Keevil SF, Barbiroli B, Brooks JCW, *et al.* (1998) Absolute metabolite quantification by *in vivo* NMR spectroscopy: II. A multicentre trial of protocols for *in vivo* localised proton studies of human brain. *Magnetic Resonance Imaging* 16: 1093–1106.
- Klose U (1990) *In-vivo* proton spectroscopy in the presence of eddy currents. *Magnetic Resonance in Medicine* 14: 26–30.
- Nelson SJ, Vigneron DB, Star-Lack J, and Kurhanewicz J (1997) High spatial resolution and speed in MRSI. *NMR in Biomedicine* 10: 411–422.
- Tkac I, Anderson P, Andriany G, *et al.* (1995) Quality Assessment in *in vivo* NMR spectroscopy: Results of a concerted research project of the European Economic Community. *Magnetic Resonance Imaging* 13: 117–176.
- Young IR (ed.) (2002) *Methods in Biochemical Magnetic Resonance Imaging and Spectroscopy*. Chichester: Wiley.

## NUCLEIC ACIDS

### Contents

**Chromatographic and Electrophoretic Methods**  
**Spectroscopic Methods**  
**Electrochemical Methods**  
**Immunoassays**  
**mRNA Identification and Quantification**

### Chromatographic and Electrophoretic Methods

**F Foret and K Klepárník**, Institute of Analytical Chemistry, Brno, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Michael Thompson and Hongbo Su, pp. 3609–3615, © 1995, Elsevier Ltd.

### Introduction

Techniques for separation of DNA and RNA provide powerful tools for the analysis of genetic material and transcription products. Modern methods of analysis can answer most of the questions about the sequence, nucleotide substitutions (mutations, polymorphisms, or single nucleotide polymorphisms – SNPs), and structure of nucleic acids. Crucial techniques in the analysis of DNA or RNA are the



**Table 1** Chromatographic techniques used in nucleic acid analysis

<i>Chromatographic technique</i>	<i>Mobile phase</i>	<i>Stationary phase</i>	<i>Principle</i>
Size exclusion (SEC)	DNA fragments in a buffer	Silica or polymer particles containing a suitable size distribution of pores	Separation according to DNA fragment effective size
Ion exchange (IEC)	DNA fragments in a buffer	Positively charged groups of the anion-exchange resin	Differences in DNA charge characteristics, which depend on molecular mass
Reversed phase and hydrophobic interaction	Hydrophobic alkyl alkylammonium chain bound to the negative phosphates of DNA	Hydrophobic matrix surface	Hydrophobic interactions
Hydroxyapatite chromatography	Single- and double-stranded DNA	Hydroxyapatite solid surface	Single- and double-stranded DNA have different interaction affinities with hydroxyapatite
Liquid-liquid partition (LLPC)	DNA fragments in aqueous poly(ethylene glycol) (PEG)	Aqueous dextran (DX) bound to the solid support surfaces	Different partition of DNA between aqueous PEG and DX

traditional separation methods – chromatography and electrophoresis – and, more recently, microfluidics and chip-based DNA arrays and mass spectrometric methods.

Chromatographic methods are commonly used in a variety of single-use kits for the isolation of genomic DNA from debris in a cell lysate. Electrophoresis, on the other hand, plays an essential role in the analysis of the structure, sequence, and function of DNA and RNA because of the high resolution achievable. Recently, however, new protocols were developed for chromatographic DNA separations suitable also for analytical separations, especially for fast and highly efficient analysis of DNA mutations.

## Chromatographic Methods

The chromatographic methods are based on a variety of possible interactions of single- and double-stranded DNA chains with the stationary phase of the chromatographic column. Various forms of chromatographic interactions can be used for the purification of the DNA preparations. A number of single-use cartridges for desalting of the DNA fragments prepared by polymerase chain reaction (PCR) or DNA sequencing reactions are currently available in the market. Measurement of the ultraviolet (UV) absorbance at 260 nm can be employed to determine concentrations of nucleic acids. The ratio of absorbance at 260 and 280 nm is used as an estimate of DNA content relative to the protein concentrations. Usually, at a ratio of >1.9 the DNA preparation is considered to be >99% pure.

The most commonly employed chromatographic methods in nucleic acid analysis are based on size exclusion, ion exchange, reversed phase and

hydrophobic interaction, hydroxyapatite (for the purification and characterization of single- and double-stranded DNA), and liquid-liquid partitioning. The various types of chromatographic separation used for nucleic acids are summarized in **Table 1**.

While some of the separation principles listed in the **Table 1** have only a limited analytical use now and are employed mainly for sample preparation prior to consecutive electrophoretic analysis, the reverse-phase-based separations still represent an important analytical separation tool.

Reversed-phase and hydrophobic-interaction chromatographies are based on very similar types of interaction; in both techniques the nucleic acid is adsorbed to a hydrophobic surface of the chromatographic stationary phase. However, different hydrophobic parts of the nucleic acid are involved in the interaction. Reversed-phase chromatography is an extension of partition chromatography, and in the case of nucleic acids an alkylammonium ion is added to the mobile phase in order to mask the negative phosphates of the nucleic acid, and, thus, obtain retention on the column. Owing to 'ion pair' formation, reversed-phase chromatography of nucleic acids is often called ion pair chromatography. Thus, the nonpolar nucleic acid ion pairs in solution are partitioned on to the bonded hydrophobic matrix phase through the interaction of nonpolar functional groups of both the sample and the sorbent. The solubility of the DNA solute in the mobile phase versus the stationary phase determines the elution order, with the most hydrophobic species eluting last. The adsorbed nucleic acids are generally eluted using an increasing gradient of organic solvents, typically acetonitrile or isopropanol. With the use of nonporous polystyrene-divinylbenzene particles the technique is

capable of analyzing PCR products according to their chain length up to a size of 500 base pairs within a few minutes.

In hydrophobic-interaction chromatography, nucleic acids are adsorbed to the hydrophobic surface by salting out from the aqueous mobile phase. The adsorbed nucleic acids are eluted by a decreasing salt gradient that redissolves the nucleic acids in the aqueous mobile phase. The hydrophobic interaction between a DNA molecule and the matrix is enhanced by high ionic strengths. This makes hydrophobic interactions an ideal tool for purification of DNA molecules that have been prepared in a high salt concentration.

While most of the older chromatographic methods for DNA analysis have been substituted by the more efficient electrophoretic techniques (see later), some newly developed liquid chromatography methods are regaining popularity, especially for rapid analysis of the double-stranded DNA. Thus, denaturing high-performance liquid chromatography (DHPLC) has emerged as one of the most versatile technologies for the analysis of genetic variations. DHPLC has been under development since the mid-1990s and has become a method of choice for the screening of mutations and SNPs. With the development of novel polymer chemistries used for separation, the accuracy, sensitivity, and throughput of DHPLC for DNA and RNA analysis have greatly improved. In addition, the high sensitivity of DHPLC combined with the accuracy of heteroduplex analysis has allowed the development of applications beyond the scope of traditional sequencing or genotyping, e.g., the early detection of cancer. The separation is performed under partial denaturation conditions, which amplifies the differences in the nucleotide hydrogen bonding. This allows separation of long DNA chains differing by only a single nucleotide. Besides requirements on the gradient elution reproducibility, a stringent temperature control is critical for achieving good performance. An example of the DHPLC analysis of 563 base pairs (bp) long PCR products obtained from different ecotypes of *Arabidopsis thaliana* is shown in Figure 1.

An important feature of the DHPLC is the ease of collecting fractions for consecutive DNA sequencing. High-throughput applications (e.g., for screening purposes) are assured not only by the inherent speed of the technique, but also by the recent development of multicolumn automated instrumentation.

## Electrophoretic Methods

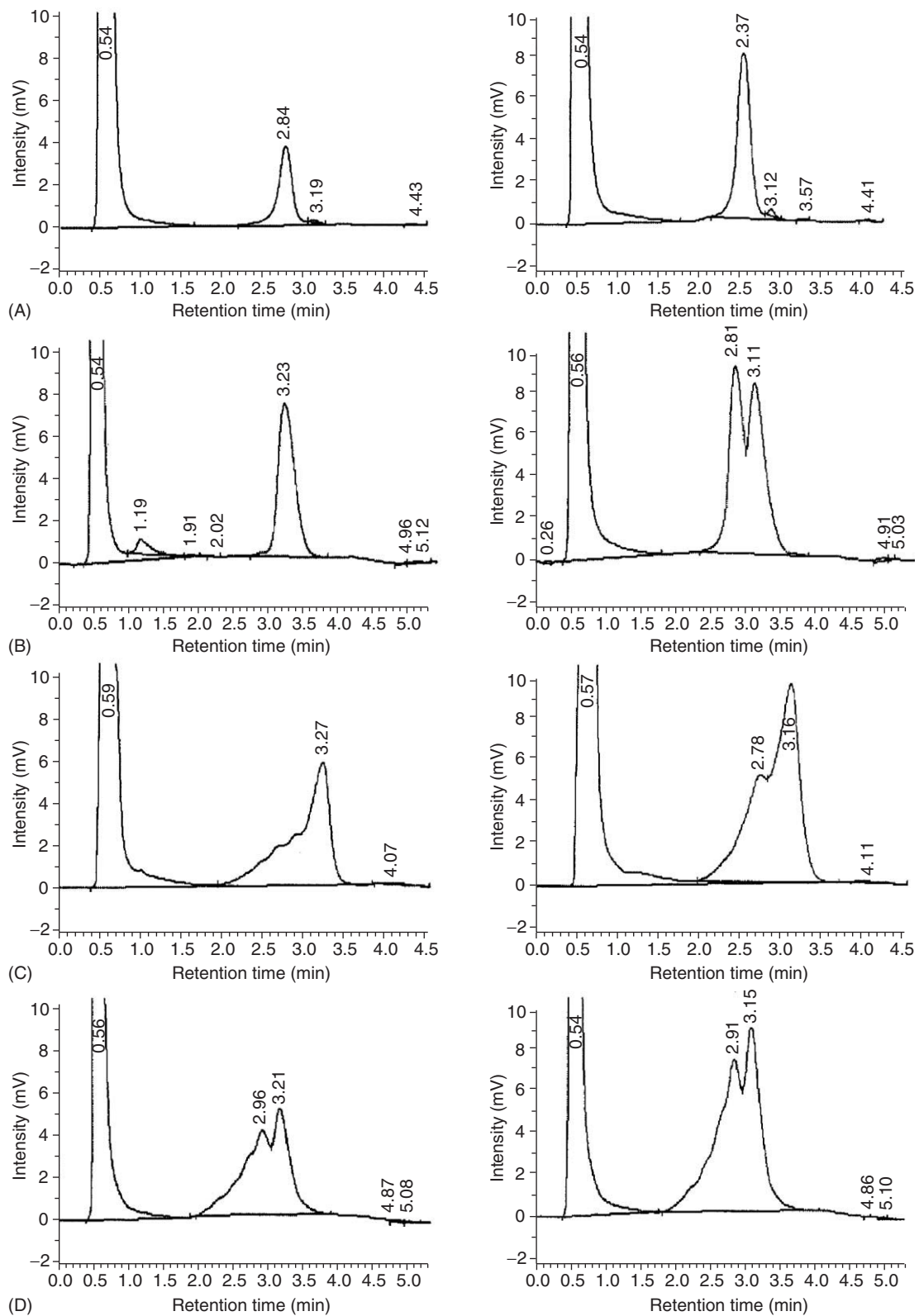
Electrophoresis is a key separation technique for the analysis of nucleic acids. Typically performed

in a bed of hydrophilic gel or in a channel filled with a viscous polymer solution, the separation is based on differential migration of the nucleic acid strands through the pores of the gel. The driving force is the electric field applied across the separation path. Although successfully practiced in the slab gel form for over 50 years, the most successful development of the capillary electrophoresis (CE) mode of the technique was triggered by the sequencing of the human genome. Up to 384 samples can now be analyzed in parallel in an array of the separation columns to achieve the required throughput. Additional improvements of the technique are under way in the form of micromachined microfluidic systems.

The most important feature of CE is the analysis speed and separation efficiency. Electric field strengths as high as  $1 \text{ kV cm}^{-1}$  (as compared to  $\sim 10 \text{ V cm}^{-1}$  typical in the slab gel) can be used in narrow capillaries without overheating. Consequently, analysis time of CE is typically more than an order of magnitude shorter than in the conventional slab gel electrophoresis (SGE). Filling of the separation capillaries with fresh separation matrix prior to each analysis provides defined and reproducible conditions for each run. Fully automated instruments are available. With the help of the capillary sequencers, the goal of the Human Genome Project, to sequence  $\sim 3$  billion base pairs of human DNA was achieved well before the schedule. The combination of the ultrasensitive laser-induced fluorescence detection (LIF) with CE opened enormous possibilities for molecular clinical diagnostics. The total separation volume of a few microliters, sample volumes as low as a few nanoliters together with the high-sensitivity LIF detection enable the analysis of the content of a single cell or even the detection of a single DNA molecule. The substantial improvement in the throughput of analyses in the combination with multiplex PCR or sequencing reactions make the technique suitable for large-scale clinical screening, sequencing, and research applications.

## Gel Electrophoresis

Since the mass-to-charge ratio of DNA molecules is size independent, the electrophoretic velocities of DNA fragments in a free solution are also the same. Therefore, a sieving medium must be used to provide the required separation. The term sieving medium relates not only to physical or chemical gels, but also to a variety of solutions of linear polymers. During the migration the chains of the nucleic acids are selectively retarded by a network of macromolecular obstacles due to the differences in their sizes and

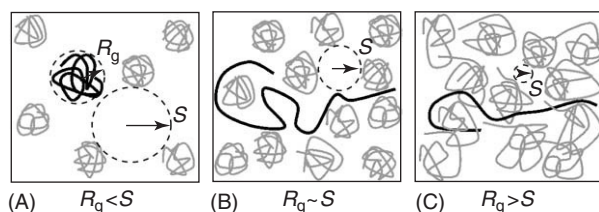


**Figure 1** Characteristic monomorphic (A and D) and polymorphic (B and C) DHPLC profiles obtained upon the analysis of mixtures of denatured and renatured Columbia and Landsberg erecta products (right column). The left column shows profiles of only Columbia fragments that served as monomorphic controls. Parallel analysis of monomorphic controls facilitates the recognition of subtle changes in chromatographic profiles (row C) indicating the presence of mutation(s) and minimizes false positives that are caused mostly by runs of bases or short variable tandem repeats (row D). Analyses were performed at column oven temperatures of 51°C (A), 46°C (B), and 58°C (C and D). (Reproduced with permission from Spiegelman JI, Mindrinos MN, and Oefner PJ (2000) High-accuracy DNA sequence variation screening by DHPLC. *BioTechniques* 29: 1084–1092; © Eaton publishing.)

conformation. To describe the retardation in terms of the dependence of effective mobility on the gel concentration, three migration regimes may be distinguished. If a DNA polymer coil adopts a conformation with the radius of hydrodynamic equivalent sphere,  $R_g$ , smaller than the average mesh size,  $S$ , of the network, the migration is controlled by accidental interactions of the coil with obstacles of a sieving medium. Here, the polymer coil can be considered to be a rigid sphere undergoing biased Brownian movement. The separation under this condition has been explained by a sieving model, where the mobility is regarded as inversely proportional to the probability of the molecule's interaction with the fibers forming the porous network. In this regime, called also Ogston migration, a spherical particle collides with the network at only one place at a time, i.e., the polymer chain is not entangled with the sieving matrix.

Rather different is the migration of a polymer chain of a length comparable with the mesh size. From time to time, the molecule gets hooked and unravels like a rope sliding over a pulley. The velocity of such a migration, which resembles reptile motion, is strongly dependent on the chain size. This mechanism of electromigration is called biased reptation. In this migration regime, the most pronounced drop of the electrophoretic mobility with the molecular mass is observed. In other words, the separation selectivity is the highest here. According to this model, the mobility of a flexible molecule migrating by the reptation mechanism is inversely proportional to its size.

When a DNA molecule is much longer than the average mesh spacing of the fibrous network, then a strong entanglement may be expected and the molecule migrates stretched nearly all the time. The tendency to the permanent stretching increases with the length of a molecule and even shorter molecules can be permanently stretched in high electric field strength. The separation selectivity under these conditions decreases to zero. The scheme of DNA migration in a network of pores of the separation matrix is shown in Figure 2.



**Figure 2** Schematic view of DNA migration regimes: (A) Ogston sieving; (B) biased reptation; and (C) reptation with permanent stretching.  $R_g$  – radius of hydrodynamic equivalent sphere of DNA coil;  $S$  – average mesh size.

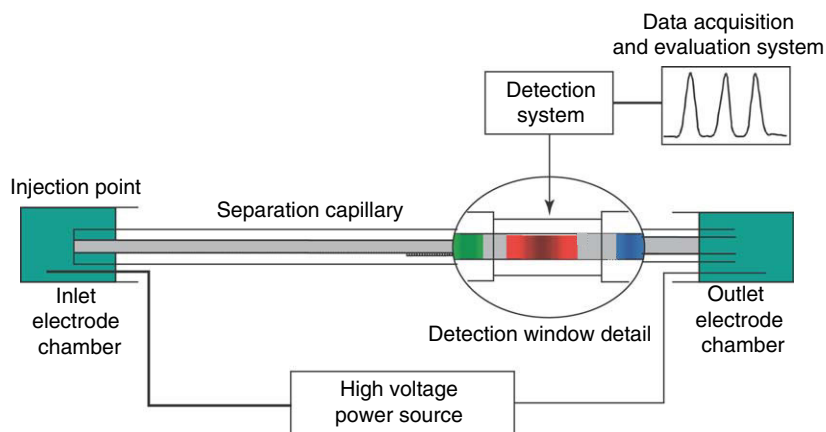
The model of DNA electrophoretic migration explains the difficulties with separations of long DNA fragments (longer than few tens of kilobases). This effect can be overcome by using pulsed electric fields for the separation of the very long fragments. Besides the simplest pulsed field arrangement a number of two-dimensional electrode arrangements and pulse programs were developed and commercial instrumentation is now available for the separation of chromosomal DNA with the length in the range of megabases.

SGE is one of the analytical techniques most often used in molecular biology, forensic science, and biotechnology. The instrumentation typically consists of an electrolyte tank for the placement of the gel equipped with electrodes for connection of the power supply. The gels, mostly of polyacrylamide or agarose, can be prepared in the laboratory or purchased precast. In the case of polyacrylamide the pore size can be varied by selecting the total concentration of the monomer acrylamide taken for the polymerization (typically 3–20%, denoted as T) and the amount of the bisacrylamide cross-linker (typically 0.5–5%, denoted as C). The agarose gels are mostly prepared in the concentrations of 0.5–2%. The detection of the DNA bands in SGE is most often performed either using radiography with radiolabeled DNA fragments or, after staining with an intercalating dye (e.g., ethidium bromide), by observing visible fluorescence under a UV transilluminator.

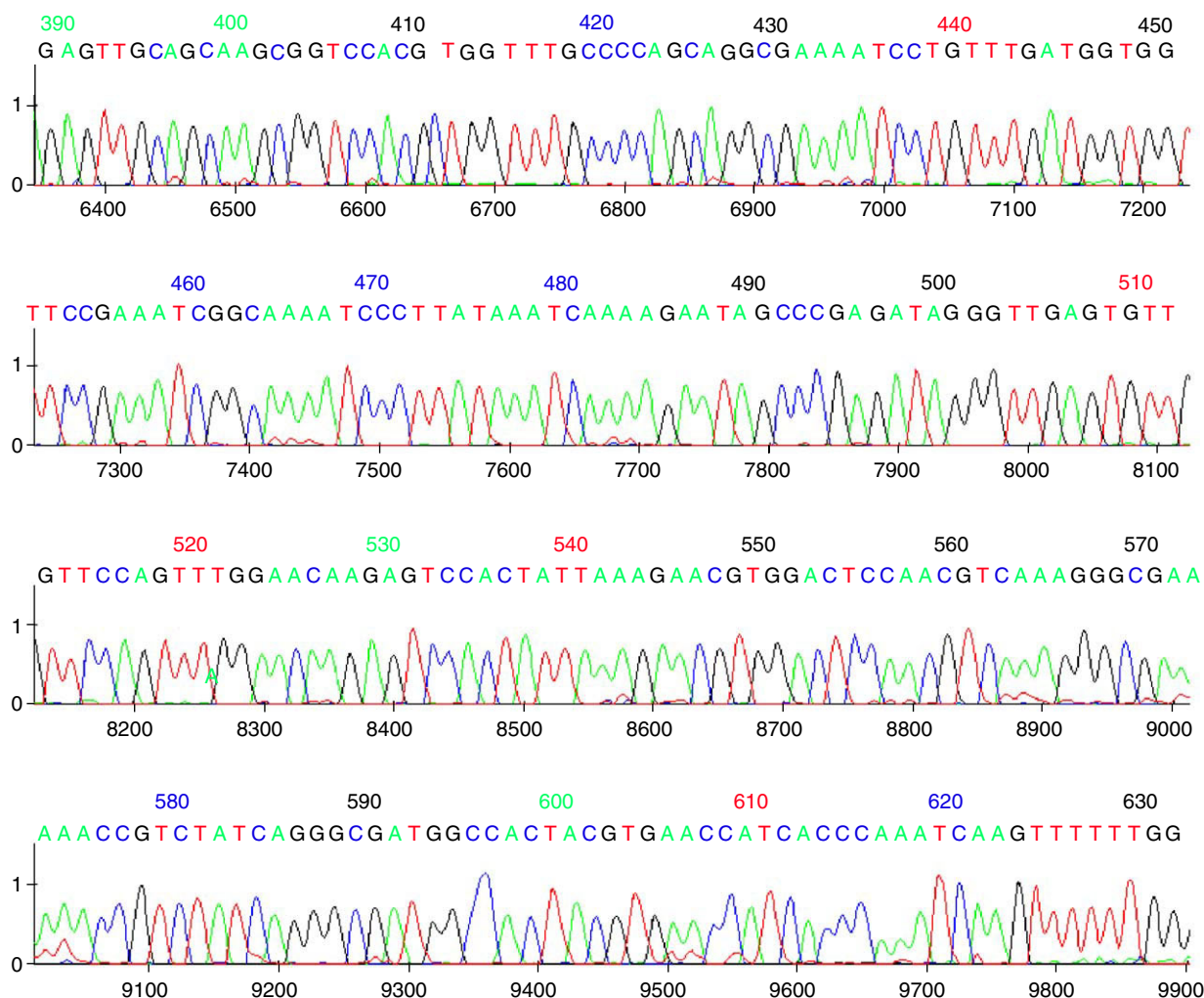
## Capillary Electrophoresis

The basic scheme of CE instrumentation is shown in Figure 3. The separation capillary is filled with the separation matrix (solution of hydrophilic linear polymers, e.g., linear polyacrylamide or its derivative) from a common pressurized matrix reservoir and is replaced after every analysis run. Besides the linear polymer the matrix contains an electrolyte buffer also (mostly good buffers at pH 8–10) to maintain constant ionic strength and pH. The samples are loaded from microtiter well plates by applying voltage for predefined periods of time. Finally, the separation voltage is applied via the electrode reservoirs at the ends of the capillary array. Fused silica capillaries with inner diameters of 50–100  $\mu\text{m}$  are typically used as the separation columns. The detection systems with direct data acquisition are based on LIF. The separation voltage ranges from 2 to 30 kV.

In analogy to SGE, where multiple samples can be separated in parallel, capillary array electrophoresis (CAE) systems were developed to enhance the



**Figure 3** Scheme of a capillary electrophoresis system.



**Figure 4** Typical record of a separation of DNA sequencing fragments. Four colors indicate four terminating nucleotides specifically labeled by four fluorescent tags. Thus, the individual bases are called and sequence determined (row of letters). The segment of the separation spectrum between 6400 and 9900 time points corresponds to the migration time 60–95 min. The numbers over the sequence are sizes of fragments in number of nucleotides or position of a nucleotide in the sequence.

throughput of genetic mapping and DNA sequencing. Present fully automated CAE systems operate up to 384 capillaries simultaneously.

## Separation Modes in Electrophoresis of Nucleic Acids

There are different modes of conducting the electrophoresis to achieve the required separation of DNA or RNA fragments. For example, the mutation detection or gene-polymorphism analysis can be based on restriction fragment length polymorphism (RFLP), length polymorphism of PCR amplified fragments (AFLP), single-strand conformation polymorphism (SSCP), heteroduplex analysis, DNA-DNA hybridization, or denaturing gradient gel electrophoresis (DGGE). Varieties of these techniques can be performed in both the slab gel and in the capillary formats. Besides using the network of pores of the separation (sieving) matrix, there are additional ways of influencing the resolution of nucleic acids. One of the most important factors is the use of denaturants and elevated temperature. Typically, the addition of urea (up to  $11\text{ mol l}^{-1}$ ) and/or formamide and elevation of temperature up to  $90^{\circ}\text{C}$  provides conditions for partial or complete denaturation of the DNA strands. While the use of native (no denaturant added) or partly denaturing conditions may be useful for the analysis of double-stranded DNA and/or sequence-dependent separation of the fragments of the same size, the highly denaturing conditions are required for the size-dependent separations such as DNA sequencing.

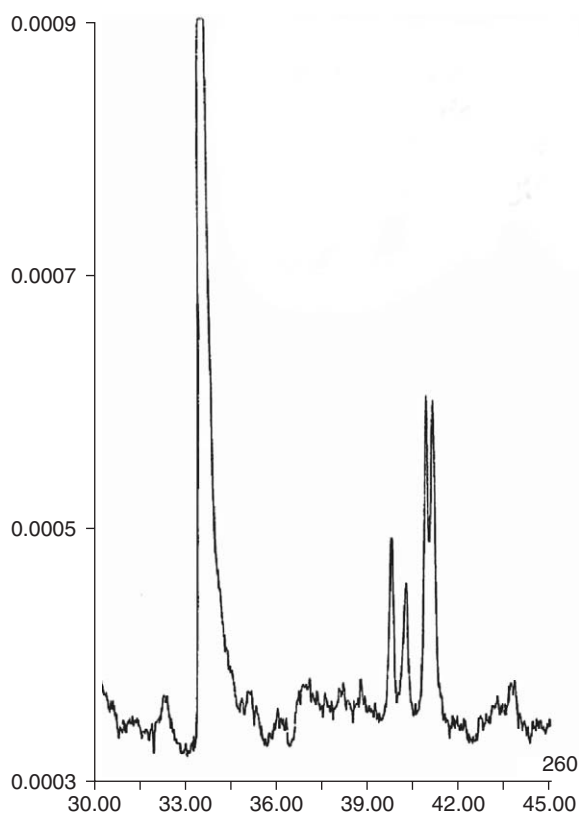
## DNA Sequencing

DNA sequencing was the most successful application of CE, both scientifically and commercially. Since the procedure relies on the capability to resolve fragments differing by a single nucleotide, the conditions providing the highest separation efficiency must be selected. Thus, relatively long separation distances (tens of centimeters) are used together with highly denaturing sieving matrices operated at elevated temperature. Such conditions prevent self-hybridization of the DNA, which could result in compression of the bands of the G- and C-rich DNA fragments. The sequencing read length well over 1000 bases is possible in a 1 h with the current capillary array sequencers. Fully automated instrumentation including LIF detection and sophisticated software for real-time base calling is commercially available. An example of the sequencing record is given in Figure 4.

## Analysis of DNA Fragment Length Polymorphism

One of the most frequent CE techniques of DNA diagnostics is the monitoring of a length polymorphism of fragments prepared by PCR amplification of a genomic DNA (AFLP) or enzymatic cleavage (RFLP). The mutations caused by deletions or insertions in a genome can be detected as zones of fragments with different sizes.

The RFLP method is used not only for the mutation detection but also as a basis for construction of restriction (or physical) maps of various genomes. Such maps can be used for the molecular identification of microorganisms. Similar application of the size-based separations in diagnostics is the analysis of tandem repeat polymorphism. High-resolution AFLP methodologies are developed for the sizing of repetitive DNA sequences that are arrayed as tandem repeat units in noncoding regions of a genome. The separation of the double-stranded DNA fragments in



**Figure 5** Typical record of an SSCP analysis. Detection of point mutation (substitution of thymine to cytosine) causing phenylketonuria in a heterozygote. Four completely dissociated DNA strands with different sequences (two couples of complementary strands with and without the point mutation) are resolved as four conformers in a native separation environment. The large peak represents a portion of undissociated double-stranded DNA.

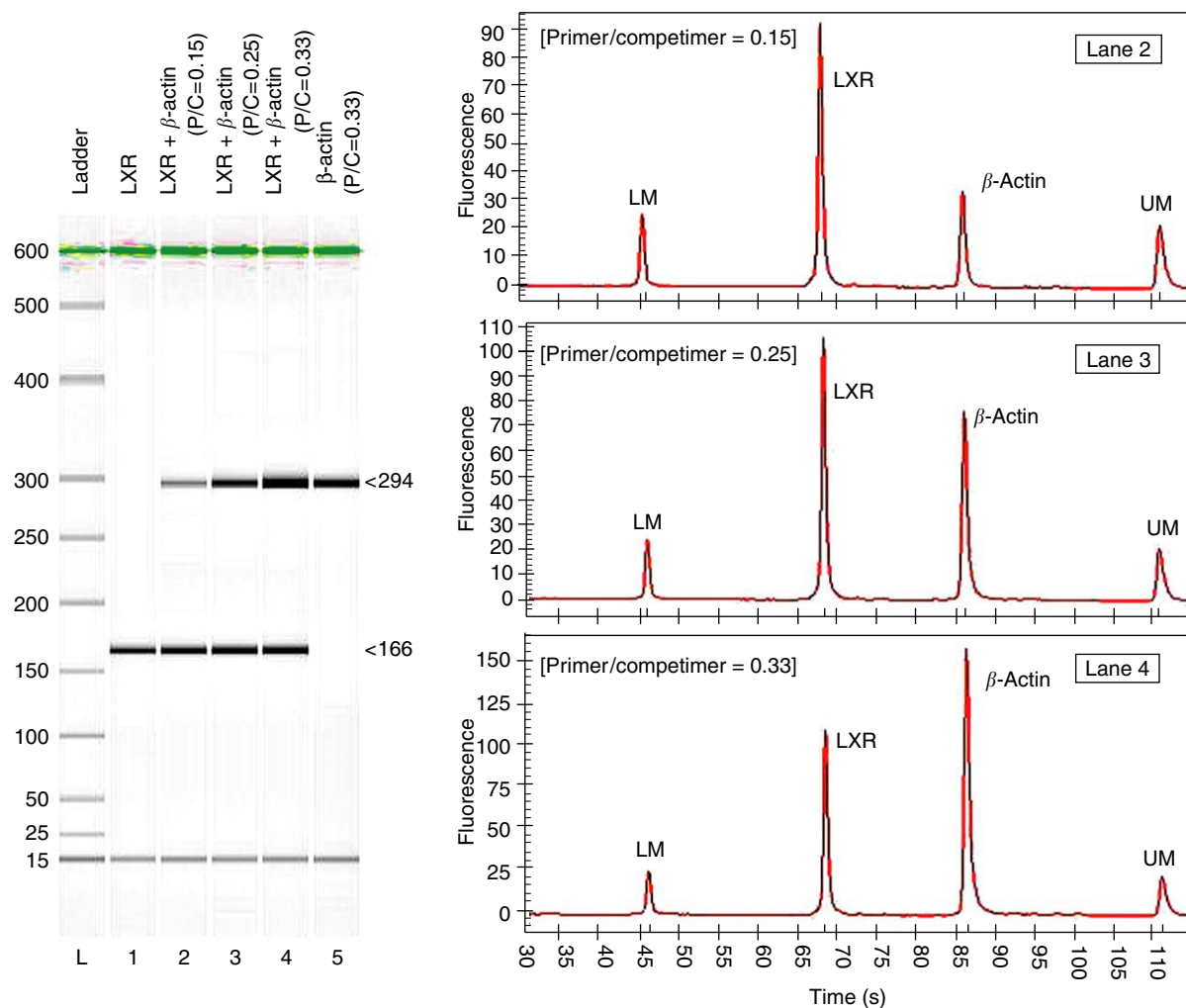


the size range of 50 bp to several kilobase pairs is performed under nondenaturing conditions.

## Single-Strand Conformation Polymorphism

SSCP is a popular technique for the screening of mutations or polymorphisms in short segments of DNA. Mutations are detected by monitoring of mobility shifts of individual DNA strands caused by their conformation changes. The main advantage of the SSCP technique is the ability to detect mutations (substitutions, small deletions, and insertions) at a variety of positions in DNA fragments with good

sensitivity. Even point mutations often give rise to conformation changes in nondenaturing environment and result in a mobility shift when molecules of the same number of nucleotides migrate in a sieving medium. Under nondenaturing conditions, the single DNA strands (ssDNA) adopt a folded structure, which is stabilized by intramolecular duplex between complementary regions and/or base stacking within each strand. Consequently, the folding and structural conformations depend not only on the presence of the mutations but also on the nucleotide sequence. Full theoretical understanding of all the factors affecting the three-dimensional folding of ssDNA fragments is still not available. Thus, the finding of the optimum conditions for SSCP analysis is highly



**Figure 6** On-chip analysis of the PCR amplification of LXR $\alpha$  gene from control human colon carcinoma cell line (CaCo-2) cells cDNA with 27 cycles. Lanes: 1, gene amplified with 5 pmol each of primers for LXR $\alpha$  gene; lanes 2–4, co-amplification of LXR $\alpha$  gene with  $\beta$ -actin at different ratios between primer and competitor; lane 5, amplification of  $\beta$ -actin gene with the primer-to-competitor being 0.42. The densitograms show the exact comparison of areas of LXR $\alpha$  gene and internal reference. LM, lower marker; UM, upper marker. (Reproduced with permission from Cantàfora A, Blotta I, Rivabene R, Pisciotto L, and Bertolini S (2003) Evaluation of RNA messengers involved in lipid trafficking of human intestinal cells by reverse-transcription polymerase chain reaction with competitor technology and microchip electrophoresis. *Electrophoresis* 24: 3748–3754; © Wiley-VCH.)

empirical. In practice, the sample is denatured prior to analysis, the dissociated ssDNA fragments are applied to a native environment of a sieving medium and separated by electrophoresis. The denaturing can be accomplished chemically (e.g., by formamide) under highly alkaline conditions, physically at an elevated temperature, or by combination of both. The initial sample denaturing is accomplished by heating of the sample up to over 90°C for 1–5 min followed by immediate chilling on ice. An example of the application of the SSCP analysis performed in CE is shown in Figure 5.

## Separation Based on Partial Denaturing Conditions

Several variants of electrophoresis performed under partially denaturing conditions were developed in both the slab and capillary formats for detection of mutations and DNA variants. In contrast to the SSCP method, where the completely denatured ssDNA samples are analyzed in a nondenaturing sieving matrix, here the analyses proceed in a partially denaturing environment. Thus, the DNA strands are partially melted resulting in the stronger retardation of their migration through the sieving matrix. The separation selectivity of the method is based not only on the character and position of a mutation, but also on the denaturing parameters during the migration through a separation column. When the separation is performed using a gradient of the denaturant along the separation gel (DGGE), the migrating DNA bands individually reach a point of maximum retardation without the need for laborious searching for the optimum denaturant concentration. Besides the use of denaturants like urea and formamide, the precise control of the temperature during the run is important to achieve reproducible results. The capillary versions of the technique often keep the concentration of the denaturant constant and rely on the temperature control to optimize the separation, for example constant denaturant capillary electrophoresis.

## Electrophoresis on a Microfluidic Device

The development of the microfluidic systems for DNA analysis is a logical extension of the CE technology. The ‘chips’ manufactured from glass or plastic materials contain a number of microfabricated

channels and reservoirs integrating the sample loading and separation with postcolumn detection. The separation channels prepared by photolithography/wet chemical etching have typically semi-circular cross-section with the radius of  $\sim 50\ \mu\text{m}$  and depth  $\sim 20\ \mu\text{m}$ . The microfabricated systems in general use shorter separation distances than typical in capillaries resulting in further increase in the separation speed. Monitoring of the LIF signal at the end of the separation channel provides a means for detecting the very sharp zones migrating close to each other. Arrays of separation channels can be integrated with other components into a single device – the lab-on-a-chip concept. Thus, the miniaturized electrophoresis can be coupled with an automated online reaction chamber for DNA cleavage, PCR amplification, ligase chain reaction, etc. Several kits for specific applications in the DNA and RNA analysis with disposable separation chips are currently available commercially. An example of the analysis performed on a commercial microfabricated system is shown in Figure 6.

**See also:** **Electrophoresis:** Overview; Nucleic Acids. **Liquid Chromatography:** Ion Pair.

## Further Reading

- Ausubel FM, Brent R, Kingston RE *et al.* (eds.) *Current Protocols in Molecular Biology*. New York: Wiley.
- Birren B and Lai E (eds.) (1993) *Pulsed Field Gel Electrophoresis: A Practical Guide*. London: Academic Press.
- Gjerde DT, Hanna CP, and Hornby D (eds.) *DNA Chromatography*, 1st edn. New York: Wiley-VCH.
- Heller C (ed.) (1997) *Analysis of Nucleic Acids by Capillary Electrophoresis*. Braunschweig: Vieweg.
- Medintz IL, Paegel BM, Blazej RG, *et al.* (2001) High-performance genetic analysis using microfabricated capillary array electrophoresis microplates. *Electrophoresis* 22: 3845–3856.
- Mitchelson KR and Cheng J (eds.) (2001) *Capillary Electrophoresis of Nucleic Acids, vol. I: Introduction to the Capillary Electrophoresis*. Totowa, NJ: Humana press.
- Mitchelson KR and Cheng J (eds.) (2001) *Capillary Electrophoresis of Nucleic Acids vol. II: Practical Applications of Capillary Electrophoresis*. Totowa, NJ: Humana Press.
- Paegel BM, Emrich CA, Wedemayer GJ, Scherer JR, and Mathies RA (2002) High throughput DNA sequencing with a microfabricated 96-lane capillary array electrophoresis bioprocessor. *Proceedings of the National Academy of Sciences of the USA* 99: 574–579.

## Spectroscopic Methods

**M Vorlíčková and J Kypr**, Academy of the Sciences of Czech Republic, Brno, Czech Republic

**V Sklenář**, Masaryk University, Brno, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Spectroscopic methods can be divided into three main groups:

1. Electronic spectroscopy, including ultraviolet (UV) and visible absorption spectroscopy, fluorescence spectroscopy, circular dichroism (CD) and linear dichroism spectroscopy.
2. Vibrational spectroscopy, including infrared (IR) absorption spectroscopy, Raman spectroscopy, and vibrational circular dichroism spectroscopy.
3. Nuclear magnetic resonance (NMR) spectroscopy.

The basic principles and technical aspects of these methods are given elsewhere in this encyclopedia. Here, we discuss briefly the applications of CD, NMR, UV absorption, IR, and Raman spectroscopies to nucleic acid samples and the interpretation of such measurements. Other spectroscopic methods have been used much less to study nucleic acids during the last 5 years and we therefore omit them in this article.

The frequency domains where these three kinds of spectroscopic methods operate differ, thus determining the instances where these methods can be used and the concentrations (amounts) of the material to be measured. Electronic spectroscopies work with chromophores, usually of the nucleic acid bases or ligands bound to the nucleic acids. The chromophores absorb UV or visible light to make electronic spectroscopies relatively very sensitive compared to IR and NMR methods. Concentrations of the samples used for electronic spectroscopies can be very low (down to  $10 \mu\text{g ml}^{-1}$ ), which is important to avoid nucleic acid aggregations. Electronic spectroscopies essentially reflect orientation of the chromophores and hence the global nucleic acid conformation. Vibrational spectroscopies operate in the IR spectral region. They work with photon energies 1–2 orders of magnitude lesser than that used in electronic spectroscopies, which makes it necessary to use samples having concentrations of the order of milligrams per milliliter. IR methods reflect bond vibrations, i.e., they provide local information about the molecules of interest. The vibrations can be calculated quite reliably and the calculations can be compared to

experiments, thus contributing to the interpretation of the experimental data.

NMR spectroscopy measures transitions between the energy levels of nuclear spins in the radiofrequency range.  $^1\text{H}$ ,  $^{31}\text{P}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  are the most commonly used isotopes to study nucleic acids. A strong external magnetic field splits the energy levels of nuclear spins in proportion to the magnetogyric ratio  $\gamma$ . This constant, characteristic for each isotope, is generally very small causing low sensitivity of NMR even when the strongest available magnetic fields are applied. Some isotopes, e.g.,  $^{12}\text{C}$  (spin quantum number  $I = 0$ ) are inactive in NMR; others, e.g.,  $^{14}\text{N}$  ( $I = 1$ ), have undesirable NMR properties. The natural abundance is  $\sim 100\%$  for  $^1\text{H}$  (99.985%) and  $^{31}\text{P}$  (100%), but much lower in the cases of  $^{13}\text{C}$  (1.108%) and  $^{15}\text{N}$  (0.370%). This stimulates the preparation of nucleic acid samples enriched with  $^{13}\text{C}$  and/or  $^{15}\text{N}$  isotopes.

### UV Absorption Spectroscopy

Nucleic acid bases absorb UV light. Maximum absorption is observed  $\sim 260 \text{ nm}$ . The absorption can be used for basic nucleic acid sample characterization before it is subjected to a more sophisticated technique. Besides the band maximum  $\sim 260 \text{ nm}$ , nucleic acids display an absorption minimum at  $\sim 230 \text{ nm}$  and another maximum within  $170\text{--}200 \text{ nm}$ . Beyond  $300 \text{ nm}$ , nucleic acids absorb no light. Hence, a nonzero absorption signal beyond  $300 \text{ nm}$  indicates a loss of light caused by light scattering owing to nucleic acid aggregates or impurities. The  $A_{260}/A_{280}$  absorbance ratio should be  $\sim 1.8$  and the  $A_{260}/A_{230}$  ratio should be 2 or higher. Values of  $A_{260}/A_{280}$  lower than 1.8 or values of  $A_{260}/A_{230}$  lower than 2 indicate protein or other impurities.

UV absorption spectroscopy is further used to determine nucleic acid concentration. The average molecular weight of a nucleotide is 320 and  $50 \mu\text{g ml}^{-1}$  concentration of a random sequence DNA provides the  $A_{260}$  absorbance of  $\sim 1$  in a  $1 \text{ cm}$  pathlength cell. In practice, the absorption coefficient of DNA is precisely determined by enzymatic hydrolysis of DNA to monomers and calculation using the absorption coefficients of the monomers, or by measurement of thermally denatured DNA. The absorption of the denatured DNA sample is calculated on the basis of absorption coefficients of the constituent mono- and dinucleotides.

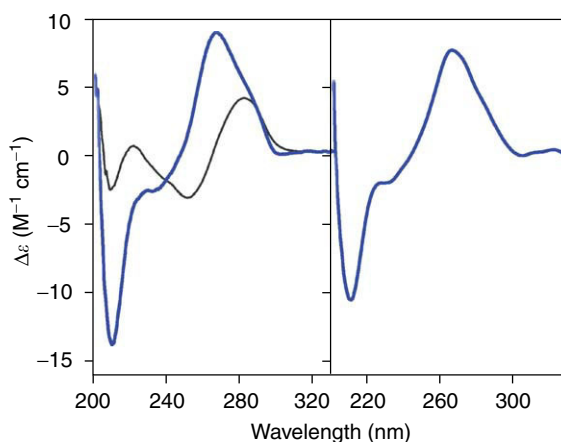
The absorption spectra of DNA and RNA are hypochromic, which means that the absorbance of

the polymeric or oligomeric chains is lower than that of the respective monomers. The absorbance of DNA and RNA decreases when going from single-stranded to double-stranded forms. Heating of nucleic acid solution causes a sharp UV absorption increase due to a partial loss of base stacking accompanying the strand dissociation. Melting point ( $T_m$ , midpoint of the transition) measurements have been widely used for the characterization of DNA samples of different origins to determine their (G + C) content having a taxonomic meaning, e.g., in bacteria. The sharpness (cooperativity) of the melting is an important characteristic of the DNA sample studied. It reflects internal heterogeneity of the primary structure of the DNA. Absorption spectroscopy also reflects conformational transitions, e.g., the B–Z transition, or transitions to tetraplex arrangements (see below).

## Circular Dichroism Spectroscopy

Circular dichroism (CD) arises owing to different absorptions by the analyzed nucleic acid sample of right-handed and left-handed circularly polarized light. Hence, it is a very small difference (of the order of thousandth) of two large numbers. This is one of the reasons why calculations mostly do not reproduce the experimental CD spectra of nucleic acids even qualitatively. Another reason is that the calculations are complex and the accurate dipolar and magnetic moments of the nucleic acid bases are not known. This is why CD is mostly used for empirical comparative studies of nucleic acids. On the other hand, the instrumentation is relatively cheap, the recording of spectra is easy and fast, and the samples can be titrated to change pH, salt concentration, or ligand concentration without any substantial problem. Moreover, CD spectroscopy is extremely sensitive to even slight changes in the mutual orientation of chromophores of nucleic acid bases. Therefore, it is the most frequently cited method in papers describing conformational transitions of nucleic acids and nucleic acid interactions.

Nucleic acids contain chiral nucleosides because of the C1' carbon atom of the sugar ring. However, the main contribution to CD follows from the asymmetric stacking of nucleosides into helical arrangements. These arrangements include, e.g., the B-form, A-form (a constitutive conformer of double-stranded RNA) that the DNA adopts upon dehydration, guanine tetraplexes, cytosine tetraplexes, and other conformers different from the common B-form DNA. CD spectra can not only be measured with nucleic acid solutions but also with films where the overall conformation can be assessed by X-ray diffraction. The CD spectra of the basic double helix types A and

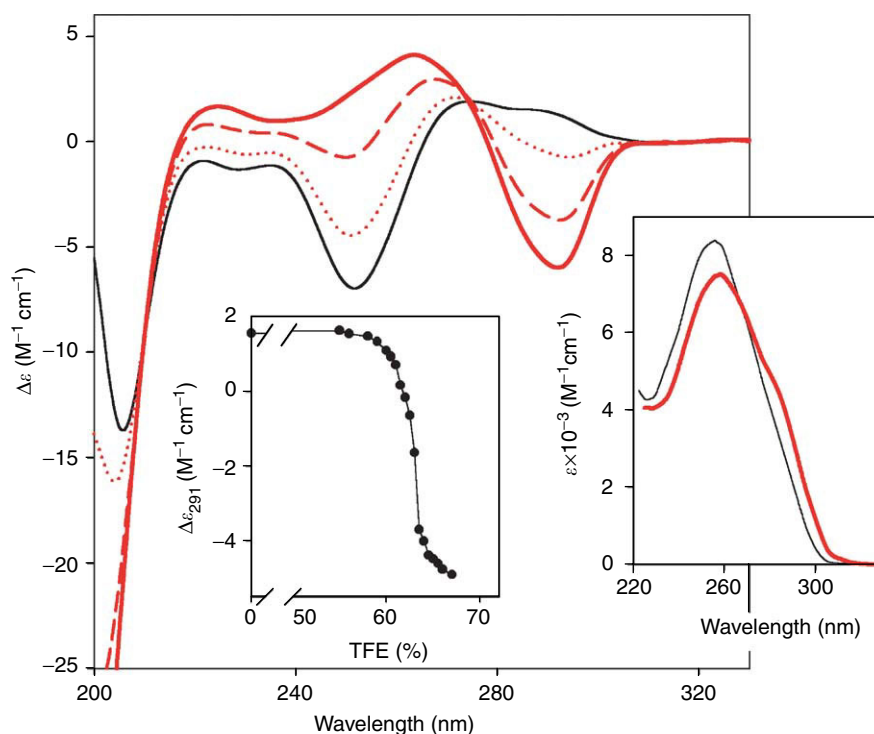


**Figure 1** CD spectra of (left) B-DNA and A-DNA of d(GCGGCGACTGGTGAGTACGC) duplexed with its complementary sequence, and (right) A-RNA of the duplex with the same nucleotide sequence (with U instead of T). The NA samples were dissolved in 1 mmol l<sup>-1</sup> Na phosphate + 0.3 mmol l<sup>-1</sup> EDTA, pH 7. Trifluoroethanol was added to DNA up to 80% to induce the A-form. Spectra were taken at 0°C.

B are shown in **Figure 1**. The B-form CD spectrum is rather weak in the 200–300 nm spectral region and the amplitudes of the positive and negative bands are similar. The CD spectrum of the A-form is much stronger and is dominated by a positive band ~260 nm and a negative one at 210 nm. The A-form is adopted by RNA, and by many molecules of DNA in 80% aqueous ethanol.

Varying temperature, concentration of salt, pH, or other agents causes changes in the CD spectrum. The changes can be either noncooperative or cooperative (**Figure 2**). The noncooperative changes reflect gradual modifications of nucleic acid geometry within a single type of conformation, e.g., within B-DNA. In some situations, however, the molecule cooperatively switches from one type of geometry to a distinct geometry, e.g., from B to A. The cooperative changes can be identified by CD spectroscopy because the spectra are then linearly dependent, i.e., they can be reproduced by a linear combination of the spectra of the participating conformers. The cooperative changes are accompanied by isodichroic points, which are a consequence of intersections of the spectra of the two participating conformers. All the spectra recorded in the course of the transition should pass through the isodichroic point as well (**Figure 2**). Cooperative phenomena are characteristic by the S-shaped course of changes, monitored by the ellipticity at a selected wavelength (**Figure 2**, inset).

A synthetic DNA poly(dG–dC) · poly(dG–dC) is composed of a strictly alternating sequence G and C. It can be prepared *in vitro* by enzymatic methods. High salt or alcohol concentrations induce



**Figure 2** CD and (inset right) UV absorption spectra of poly(dG-dC) reflecting its B-Z transition induced by trifluoroethanol: (inset left) the transition monitored by  $\Delta\epsilon_{291}$ . The alcohol was added to the polynucleotide dissolved in 1 mmol l<sup>-1</sup> Na phosphate, 0.3 mmol l<sup>-1</sup> EDTA, pH 7 up to (thin black trace) 59.0, (dots) 62.5, (dashes) 63.5, and (thick red trace) 67.0%. The spectra were measured at 0°C.

a cooperative transition of poly(dG-dC) · poly(dG-dC) into the left-handed Z-DNA double helix. This transition was discovered by CD spectroscopy (Figure 2). Another synthetic polynucleotide poly(dA-dT) · poly(dA-dT) also undergoes salt- or alcohol-induced transitions. For some of the structures there is still no convincing conformational explanation. Alternating sequences of the purine and pyrimidine residues introduce a conformational polymorphism into DNA that can be studied by CD spectroscopy under a wide range of the solution conditions. In contrast, blocks of purine nucleotides bound to the complementary strand fix DNA in a single type of double-stranded conformation. For example, poly(dA) · poly(dT) is a heteronomous B-like duplex in which the strand of poly(dA) has a rather different conformation than the strand of poly(dT). The duplex of poly(dA) · poly(dT) cannot switch into the A-form at all, and only can become a part of a three-stranded conformation under favorable conditions. Also, strands of poly(dG) and poly(dC) form triplexes at slightly acid pH. Their duplex poly(dG) · poly(dC) adopts an unusual A-like conformation even in aqueous solution. It easily switches into the A-form in ethanol solutions. Most of these facts were first found using CD spectroscopy.

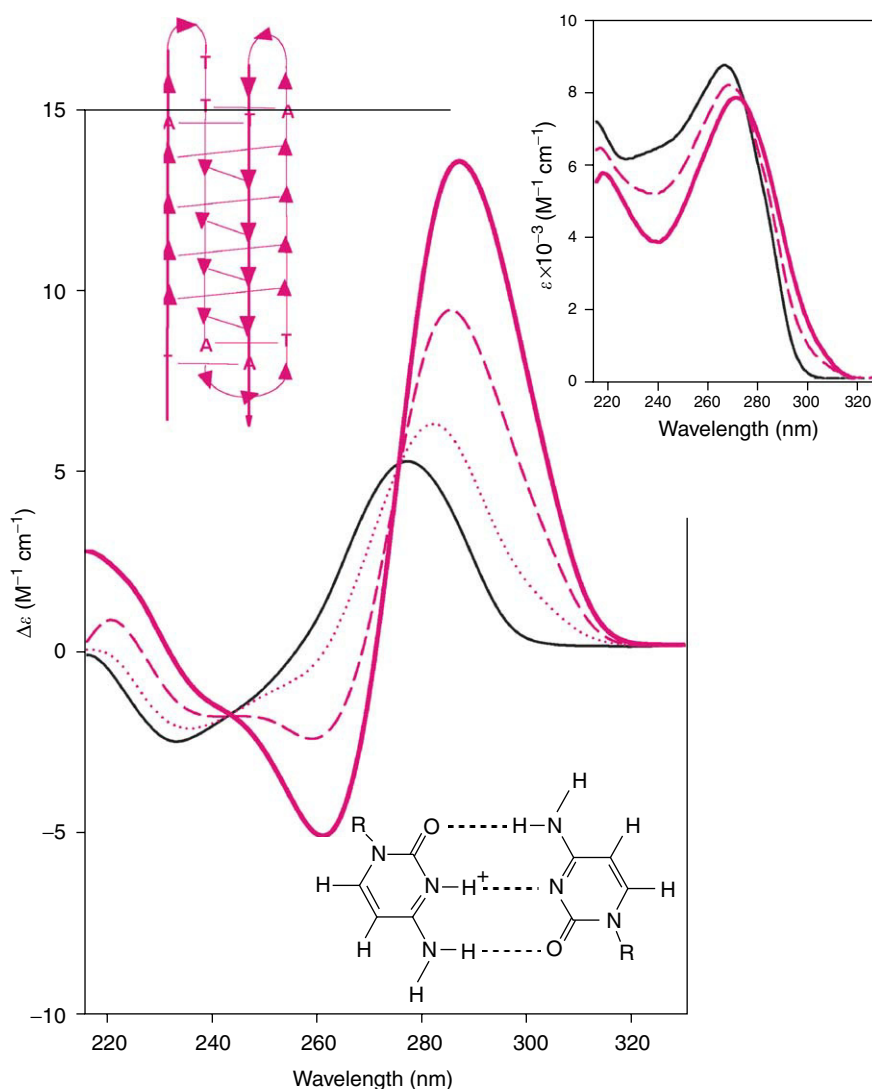
Cytosine-rich DNA strands associate into intercalated tetraplexes (i-tetraplexes). They consist of two parallel-stranded duplexes bound by hemiprotonated

C<sup>+</sup> · C<sup>+</sup> pairs that are mutually intercalated in anti-parallel direction (Figure 3). The intercalated tetraplexes can be formed by a single, two, or four molecules, but irrespective of the number of molecules involved the i-tetraplex always has the same strand topology and provides the same characteristic type of CD spectrum (Figure 3). Intercalated tetraplexes are mostly induced by slightly acid pH to make the cytosines hemiprotonated.

Guanine-rich DNA strands also associate into tetraplexes. Guanine tetraplexes are much more complex and difficult to work with because they can have various topologies (Figure 4). The type and concentration of cations, length and number of the guanine blocks in the molecule, and the primary structure of the loops determine the preferred type of guanine tetraplex topology. The various topologies provide different CD spectra (Figure 4). Transitions involving intermolecular tetraplexes are very slow (days).

(dG-dA)<sub>n</sub> strands associate into parallel-stranded homoduplexes whose CD spectra are very similar to the CD spectra of intermolecular guanine tetraplexes. The homoduplex formation is not restricted to the strictly alternating sequence of G and A, but a presence of guanine doublets or longer guanine runs leads to the formation of G-tetraplex.

CD spectroscopy was used in many instances to study conformational properties of DNA strands



**Figure 3** CD and (inset) UV absorption spectra reflecting the acid-induced transition of a DNA fragment d(TCCCCACCTTCCC-CACCCTCCCCACCTCCCCA) of a c-myc human oncogene into an i-tetraplex. Spectra were recorded at room temperature in Robinson–Britton buffer, pH (thin black trace) 9.2, (dots) 7.2, (dashes) 6.7, and (thick violet trace) 5.0. The triangles in the sketch stand for cytosines. They point in the 5' → 3' direction. The C<sup>+</sup>·C pair is shown at the bottom.

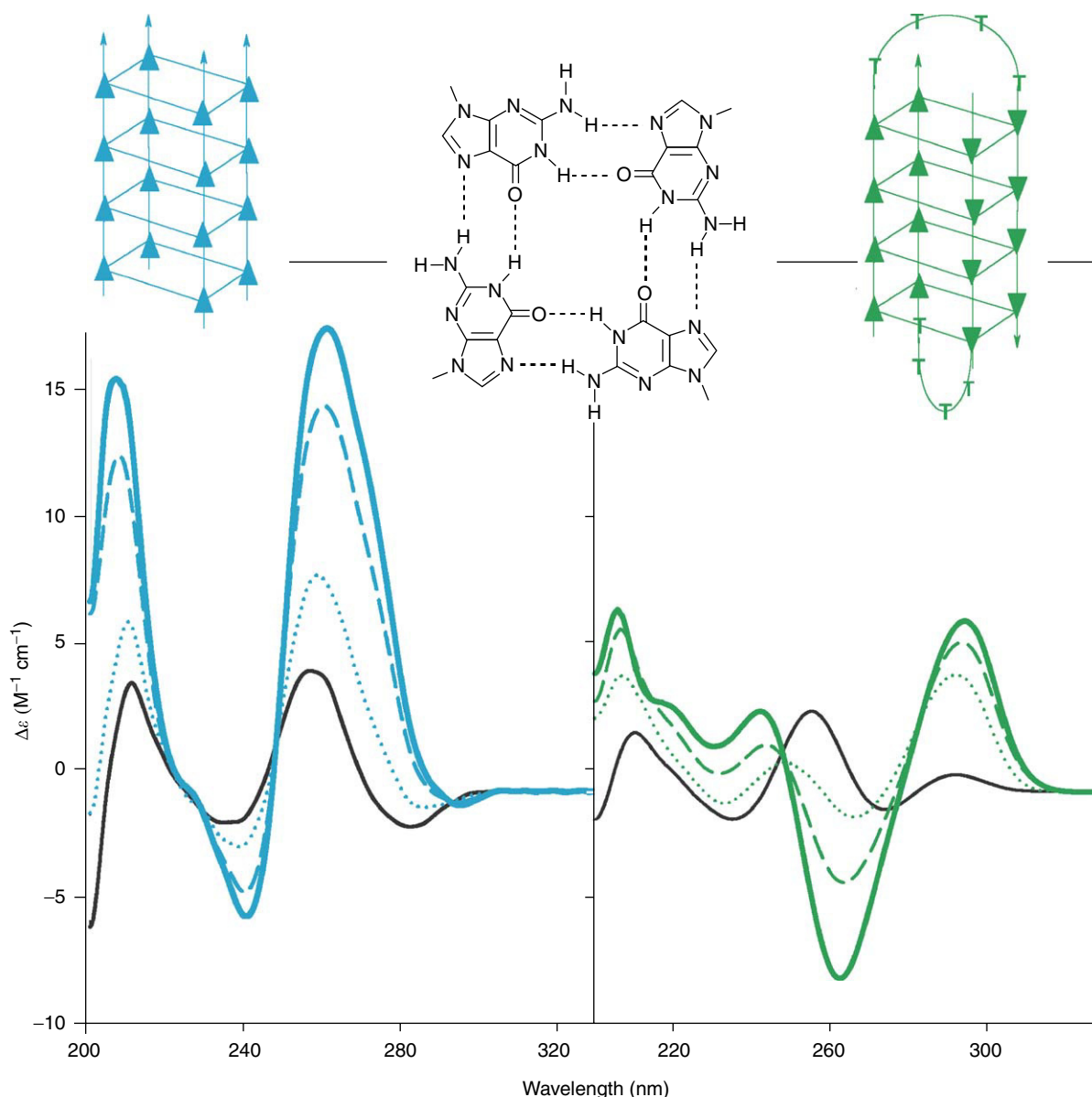
containing trinucleotide repeats whose expansion correlates with serious, mostly neurodegenerative human diseases. The strands include d(CAG)<sub>n</sub> and d(CTG)<sub>n</sub> (Huntington chorea and other diseases), d(GAC)<sub>n</sub> (multiple epiphyseal dysplasia), d(CCG)<sub>n</sub> and d(CGG)<sub>n</sub> (fragile X-chromosome syndrome), and d(GAA)<sub>n</sub> (Friedreich ataxia). CD spectroscopy contributed to a demonstration that all of the trinucleotide repeats generated stable intrastrand conformations, competing with the heteroduplex, with the complementary strand. Just these intrastrand structures are considered to be responsible for the repeat expansions. The strands of d(CAG)<sub>n</sub> fold into a stable hairpin, d(GAC)<sub>n</sub> form a hairpin, parallel homoduplex, and the left-handed Z-form, d(CCG)<sub>n</sub> generate a hairpin and two kinds of tetraplexes in

which C plays a dominant role, d(CGG)<sub>n</sub> form a hairpin and G-tetraplex for  $n < 5$ , and d(GAA)<sub>n</sub> form a homoduplex, ordered single strand, and a triplex in the presence of the complementary sequence. All of these unusual structures, arising as an alternative to classical DNA heteroduplex, yield characteristic CD spectra. Thus, CD spectroscopy can easily monitor the formation and the conditions of the stability of these structures.

## IR and Raman Spectroscopy

IR spectra can be obtained from solutions, gels, fibers, or films as well as crystals of nucleic acids. This is an important advantage because crystal conformations of DNA correspond to conformations that DNA





**Figure 4** CD spectra reflecting the formation of (left)  $\text{K}^+$ -induced parallel-stranded tetramolecular tetraplex of  $\text{G}_4$  and (right)  $\text{Na}^+$ -induced antiparallel bimolecular tetraplex of  $\text{G}_4\text{T}_4\text{G}_4$ . Both oligonucleotides were dissolved in  $1 \text{ mmol l}^{-1}$  Na phosphate +  $0.3 \text{ mmol l}^{-1}$  EDTA, pH 7, and thermally denatured before starting measurements. CD spectra were measured at room temperature (left): (thin black trace) immediately and (from dots to thick cyan line) 10', 3.5 and 24 h after adding  $10 \text{ mmol l}^{-1}$  K phosphate, pH 7; (right): (from the thin black to the thick green line) 1, 10, 100  $\text{mmol l}^{-1}$  Na phosphate, and +  $0.5 \text{ mol l}^{-1}$  NaCl, pH 7. The triangles in the sketches stand for guanines and point in the  $5' \rightarrow 3'$  direction. The G-tetrad is shown in the middle.

adopts in concentrated aqueous alcohols rather than dilute aqueous solutions. IR spectroscopy puts no limits on the size of the molecules. It can be used with high molecular weight native molecules of DNA as well as short oligonucleotides. Furthermore, it is a nondestructive method. The samples can be used for other purposes after the IR spectroscopic measurements.

The use of IR spectroscopy is based on the existence of marker bands in the spectral region between

$1800$  and  $700 \text{ cm}^{-1}$ , which are sensitive to nucleic acid conformation. The marker bands arise from vibrations in various parts of the molecule. Stretching double bond in plane vibrations of the bases provide bands within  $1800\text{--}1500 \text{ cm}^{-1}$ , sugar-base interactions (depending on the glycosidic torsion angle) are reflected within  $1500\text{--}1250 \text{ cm}^{-1}$ . Phosphate and sugar vibrations provide bands within  $1250\text{--}1000 \text{ cm}^{-1}$ , and the phosphodiester bonds vibrate below  $1000 \text{ cm}^{-1}$ . Solvent  $\text{H}_2\text{O}$  makes it

difficult to obtain data  $\sim 1600$  and below  $1000\text{ cm}^{-1}$ . This is why  $\text{D}_2\text{O}$  is used instead of  $\text{H}_2\text{O}$  in many IR spectroscopy studies of nucleic acids.

Photons are scattered by molecules and lose energy during the scattering process. The lost energy corresponds to the vibrational energy of the molecule. Hence, a Raman spectrum fingerprints vibrational states of the molecule that depend on its conformation and interactions. Raman spectroscopy complements IR spectroscopy. The major advantage of Raman spectroscopy lies in a very weak signal of water, whereas the  $\text{H}_2\text{O}$  signal is strong in the IR spectra and masks the signal of the nucleic acid dissolved in the water solution. Raman spectroscopy shares the advantage of IR spectroscopy to be able to work with crystals, solutions, gels, fibers, as well as amorphous solids. Combination of the X-ray data with Raman studies of the corresponding crystals permitted the identification of several spectral bands (i.e., their frequencies and intensities) that are characteristic, for example, including nucleoside sugar pucker (C3'-endo or C2'-endo) or orientation around the glycosidic bond (*syn* or *anti*). These spectral lines are frequently well separated from one another to detect two or more conformers in the same molecule. Overlapping bands should be deconvoluted.

## NMR Spectroscopy

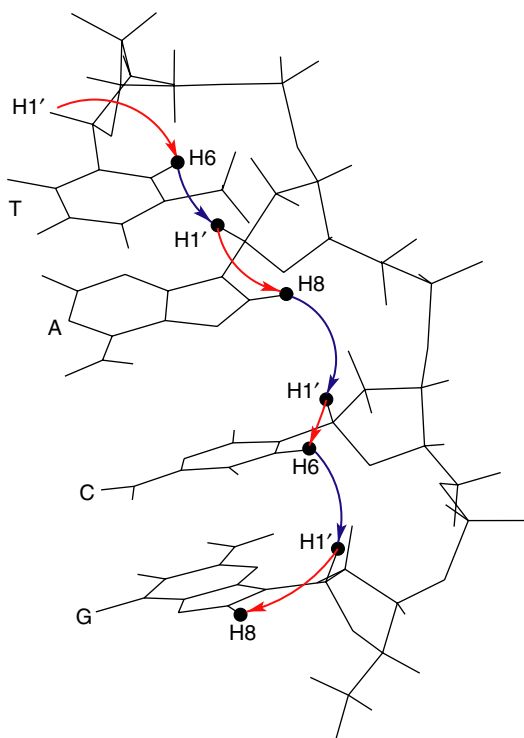
NMR spectroscopy belongs to the group of radio-frequency spectroscopic techniques. Energy levels of atomic nuclei with nonzero spin quantum number  $I$  are split by the presence of a strong magnetic field  $B_0$  in proportion to their magnetogyric ratios  $\gamma$ . The energy difference  $\Delta E = h\gamma B_0/2\pi$  ( $h$  – Planck constant), and consequently the difference  $\Delta n$  in Boltzman populations of individual energy levels are very small. Typically,  $\Delta E \sim 10^{-25}$ – $10^{-27}$  J and  $\Delta n \sim 10^{-5}$ – $10^{-7}$ . As a result, NMR spectroscopy is a highly insensitive technique requiring very high magnetic fields (11.7–21.4 T) and relatively high concentration of samples ( $c \sim 0.5$ – $1\text{ mmol l}^{-1}$ , vol.  $\sim 200$ – $500\text{ }\mu\text{l}$ ) when studying DNA and RNA oligonucleotides. Resonance frequencies corresponding to  $\Delta E = h\nu$  depend on  $\gamma$  and  $B_0$ . For example, at the  $B_0 = 11.7\text{ T}$  the most sensitive natural abundant isotope  $^1\text{H}$  resonates at 500 MHz while the resonance frequencies for  $^{13}\text{C}$ ,  $^{15}\text{N}$ , and  $^{31}\text{P}$  are 125.8, 50.7, and 202.2 MHz, respectively. Due to a large difference of resonance frequencies the high-resolution spectra of individual isotopes have to be measured separately. The differences in the electronic environment of atomic nuclei in different parts of the same molecule are slightly modifying the resonance frequencies and give rise to chemical shift dispersion

of the observed signals. The chemical shift ranges for  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ , and  $^{31}\text{P}$  isotopes in nucleic acids are  $\sim 1$ – $16$ ,  $10$ – $160$ ,  $70$ – $230$ , and  $0$ – $5$  ppm on their respective chemical shift scales. As standards for the chemical shift referencing, the signals of  $^1\text{H}$  and  $^{13}\text{C}$  of tetramethylsilane,  $^{15}\text{N}$  of liquid ammonia, and  $^{31}\text{P}$  of trimethylphosphate are typically used.

High-resolution NMR spectroscopy of samples in liquid state provides a wealth of structural information since the linewidth of the resonance signals is much smaller ( $\sim 10^{-1}$ – $10^1$  Hz) than their chemical shift dispersion. For example, the chemical shifts of  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ , and  $^{31}\text{P}$  isotopes in nucleic acids at  $B_0 = 11.7\text{ T}$  ( $\nu(^1\text{H}) \sim 500\text{ MHz}$ ) cover the frequency ranges of 7.5, 19, 8.2, and 1.1 kHz, respectively. In addition, a hyperfine splitting and multiplet structure of resonance signals can be observed as a result of electron-mediated nuclear spin–spin scalar interactions. Both chemical shifts and scalar interactions reflect the chemical environment and provide a wealth of structure-related data. Since specific information about the individual atoms is available, NMR data can be successfully used to calculate a three-dimensional structure of RNA and DNA molecules. Fuelled by methodological developments during the past 10 years, NMR became an indispensable tool for structure elucidation of nucleic acids. At present (April 2003), almost 600 NMR structures of nucleic acids and/or protein/nucleic acid complexes can be found in the PDB database.

The process of DNA/RNA structure determination can be divided in several steps. First, a sufficient amount of the sample has to be prepared. Although basic structural information can be obtained using solely  $^1\text{H}$  NMR spectroscopy, high-resolution structure determination of larger oligonucleotides requires isotope labeling by  $^{13}\text{C}$  and/or  $^{15}\text{N}$ . For preparation of samples with natural level of  $^{13}\text{C}$  and  $^{15}\text{N}$  isotopes, the chemical synthesis of DNAs by the phosphoramidite method and the enzymatic synthesis of RNAs using T7-polymerase are typically used. For isotope labeling with high levels of enrichments ( $\sim 99\%$ ), enzymatic synthesis both for RNAs and DNAs is a method of choice. In the next step, the signals in NMR spectrum are assigned to individual atoms. A vast number of homonuclear and heteronuclear NMR techniques have been designed to assist the assignments procedure. The base pairing is established by measuring the NMR spectra of exchangeable protons in  $\text{H}_2\text{O}$  solution. The observation of imino and amino signals is indicative for base pairing and hydrogen bond formation. Otherwise, these protons exchange rapidly with the solvent and their signals cannot be observed. Two-dimensional nuclear Overhauser effect (NOE) spectroscopy,

which identifies neighboring protons in the vicinity of  $\sim 5$  Å, enables the detection of both intraresidual and sequential imino-imino and imino-amino connectivities and the determination of the specific type of base pairing (Watson-Crick, non-Watson-Crick, or Hoogsteen). Then, NOE contacts of the imino and amino protons with the base (H6/H8, H5(methyl), adenine H2) and H1' protons are examined to obtain partial assignments. Assignments of other nonexchangeable protons are obtained from spectra measured in deuterated water mostly through H6/8-H1' and H6/8-H2'/2'' sequential NOE connectivities (Figure 5). The spin systems of sugar (H1', H2'/H2'', H3', H4', H5'/H5'') and pyrimidine bases (H5, H6) are identified using two-dimensional correlation spectroscopy or total correlation spectroscopy based on the existing network of spin-spin scalar interactions. Due to the increasing overlap of signals, the assignment of protons described above gets more complicated for larger systems with more than 20–30 nucleotides. The isotope labeling in combination with multidimensional data acquisition allows one to reduce the overlap observed in proton NMR spectra by using the advantage of much larger chemical shift dispersion of  $^{13}\text{C}$  and  $^{15}\text{N}$  isotopes. A large number of heteronuclear two- and three-dimensional techniques have been proposed to aid this process.

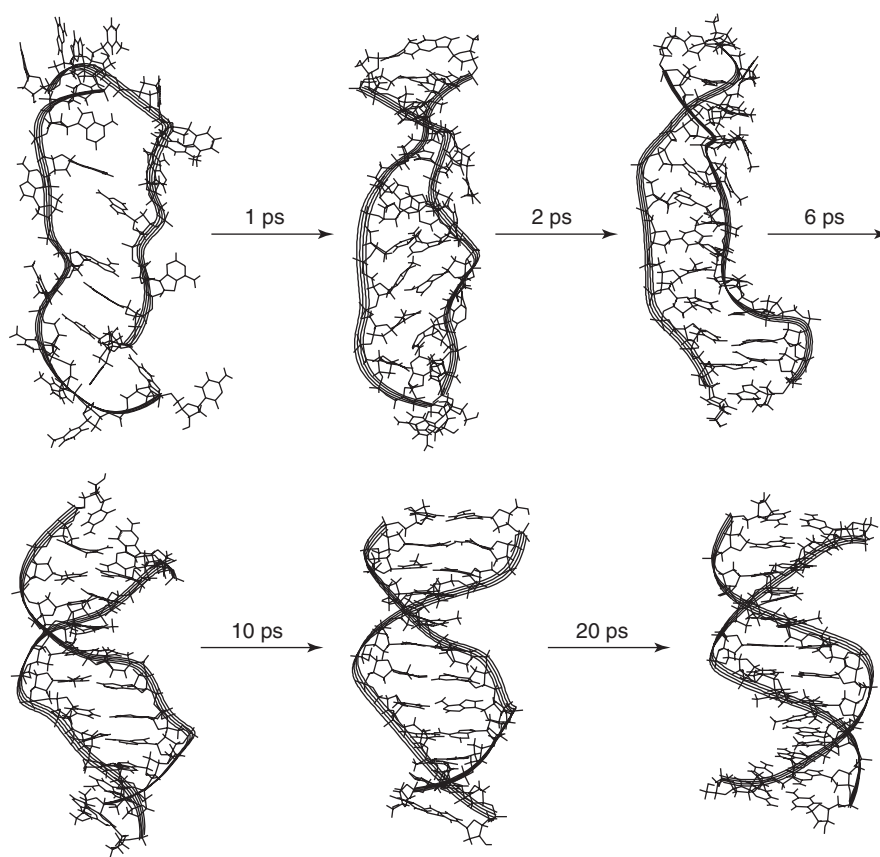


**Figure 5** Sequential NOE connectivities in DNA between sugar H1' and base protons (H6 in pyrimidines and H8 in purines). The intra- and inter-residual connectivities are marked in red and blue, respectively.

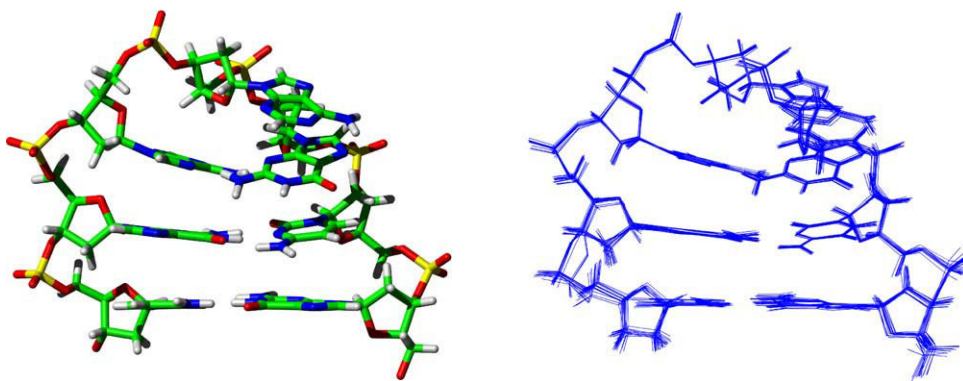
After the signals have been assigned the structural parameters are extracted. The basic information is obtained from the chemical shifts. Their values depend on the electron densities around the nuclei and are affected both by intra- and inter-residual interactions. The proton chemical shifts are strongly influenced by stacking interactions and their changes can be readily used to follow conformational changes such as melting of the elements of secondary and tertiary structures. The phosphorus chemical shifts depend on the conformation of the sugar-phosphate backbone. For example, their values display characteristic changes during B<sub>I</sub>-B<sub>II</sub>, B-Z, and B-X transitions. Quantitative structure-related information is obtained from the measurements of NOE intensities in multidimensional spectra. NOE reflects the dipole-dipole interaction between two nuclei and its intensity is indirectly proportional to the sixth power of the internuclear distance ( $\sim r^{-6}$ ). Distances up to 5 Å are determined with sufficiently high precision and accuracy. A number of torsion angles along the sugar-phosphate backbone, as well as the glycosidic torsion angle  $\chi$ , can be estimated from the values of homonuclear ( $^1\text{H}$ - $^1\text{H}$ ) and heteronuclear ( $^1\text{H}$ - $^{13}\text{C}$ ,  $^1\text{H}$ - $^{31}\text{P}$ ,  $^{13}\text{C}$ - $^{31}\text{P}$ ) three-bond scalar couplings using suitably parametrized Karplus equations. Proton-proton scalar interactions H1'-H2', H1'-H2'', H1'-H3', H2'-H3', and H2''-H3' are used to obtain information about the sugar pucker conformation both in  $\beta$ -d-deoxyribose (DNA) and  $\beta$ -d-ribose (RNA) and to estimate the populations of N- and S-type conformers.

The distances and the torsion angles extracted from NMR spectra as described above, together with the basic information about the primary sequence, bond length, and bond angles are subsequently used to calculate a three-dimensional structure of nucleic acids by molecular dynamics. First, the distance geometry approach or molecular dynamics in torsion space is applied to obtain the basic fold, which fulfills the experimental data used as restraints. Calculations are repeated for a set of completely randomized structures (Figure 6). In the next step, based on the energy and violations of the input restraints, a subset of converging structures is selected and subsequently refined using more sophisticated restrained molecular dynamics protocols.

Structure determination of nucleic acids faces several problems, which stem mostly from low proton density, lack of long-range restraints due to an elongated shape of majority structures, and a small number of NOE restraints between the distant elements of secondary structure. These difficulties can be circumvented by applying recently introduced technology for obtaining additional structural



**Figure 6** Folding of a DNA double helix during the 20 ps molecular dynamics refinement using NMR restraints.



**Figure 7** Three-dimensional structure of DNA hairpin d(GCGAAGC) determined using NMR data obtained on a  $^{13}\text{C}$ - and  $^{15}\text{N}$ -labeled molecule. NOE, dihedral, and RDC restraints were applied to refine the structure using molecular dynamics simulations using AMBER 7.0 software package.

restraints using a small degree of molecular alignment with the magnetic field. Newly introduced methodology provides the direct way to acquire the lacking long-range restraints. A weak alignment of biological macromolecules is easily obtained in solutions of bicells, filamentous phages, or other compounds. In such media, dipole–dipole and quadrupolar interactions, as well as chemical shift anisotropy, normally averaged out by an isotropic

tumbling, give rise to observable changes in the NMR spectra. Residual dipolar couplings or changes in the chemical shifts supply information about the mutual orientation of remote parts of the molecules. For example, in nucleic acids, mutual position of distant ends of double-helical DNA structures or of individual domains in RNA molecules can be accurately defined. To name just two from a large number of recently published results, high-resolution

structures of the Drew–Dickerson dodecamer (dCGCGAATTCGCG)<sub>2</sub> and the DNA hairpin d(GCGAAGC) (Figure 7) have been determined with very high precision using the isotope labeling and the alignment methodology.

Current state-of-the-art NMR technology allows studying the fragments of nucleic acids of the size up to 70 nucleotides. For molecules with more than 40 residues, the uniform <sup>13</sup>C, <sup>15</sup>N isotope labeling, which, in general, is more complicated than in proteins, has to be complemented also by residue- or site-specific labeling. For oligonucleotides with more than 50 residues partial deuteration seems to be mandatory. Despite these obstacles, the recent improvement of labeling strategies, development of new and more sophisticated NMR methods, and progress in measurements and applications of residual dipolar couplings for structure refinement have paved the way to improve the accuracy and precision of structures obtained by NMR and to study systems with larger number of nucleotides.

**See also:** **Infrared Spectroscopy:** Overview; Sample Presentation. **Nuclear Magnetic Resonance Spectroscopy:** Overview; Principles. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Nuclear Overhauser

Effect; Multidimensional Proton. **Raman Spectroscopy:** Instrumentation.

## Further Reading

- Guschlbauer W (1976) *Nucleic Acid Structure*. Heidelberg: Springer.
- Johnson WC (1996) Spectroscopic methods in nucleic acid studies. In: Fasman GD (ed.) *Circular Dichroism and the Conformational Analysis of Biomolecules*, pp. 433–468. New York: Plenum.
- Neidle S (ed.) (1999) *Oxford Handbook of Nucleic Acid Structure*. New York: Oxford University Press.
- Peticolas WL and Evertsz E (1992) Conformation of DNA *in vitro* and *in vivo* from laser Raman scattering. In: Lilley DMJ and Dahlberg JE (eds.) *Methods in Enzymology*, vol. 211, pp. 335–352. San Diego: Academic Press.
- Sanger W (1984) *Principles of Nucleic Acid Structure*. New York: Springer.
- Taillandier E and Liquier J (1992) Infrared spectroscopy of DNA. In: Lilley DMJ and Dahlberg JE (eds.) *Methods in Enzymology*, vol. 211, pp. 307–335. San Diego: Academic Press.
- Wijmenga SS and van Buuren BNM (1998) The use of NMR methods for conformational studies of nucleic acids. *Progress in NMR Spectroscopy* 32: 287–387.
- Wüthrich K (1986) *NMR of Proteins and Nucleic Acids*. New York: Wiley.

## Electrochemical Methods

**E Paleček**, Institute of Biophysics, Brno, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

DNA and RNA are complex biomacromolecules whose structures and interactions are intensively studied mainly in relation to their biological roles. The main interest of scientists is not oriented to the determination of nucleic acids in biological materials but to analysis of their nucleotide sequence, secondary and tertiary structures, specific interactions with proteins, etc. A number of diverse techniques are necessary such as absorption spectroscopy, chromatography, gel electrophoresis, and radioactive and nonradioactive labeling, which can be found in almost every nucleic acid laboratory, while other less universal techniques are applied in relation to the specific aims of the given laboratory. Electrochemical methods briefly reviewed in this article belong to the latter category. They can be extremely useful for a number of purposes, such as determination of traces

of proteins, RNA and single-stranded DNA (ssDNA) in double-stranded DNA (dsDNA) samples, detection of DNA damage and nucleotide sequencing by DNA hybridization, etc. They do not require expensive equipment and give a good sensitivity, which, in some cases, can reach levels comparable to those of radioactive tracing. In recent years, electrochemistry has gained importance in the development of the DNA sensors.

Electroactivity of nucleic acids was discovered more than 40 years ago. Nucleic acid bases undergo redox processes at electrodes and in long-chain DNA molecules the resulting electrochemical signals can be significantly influenced by their ordered higher structures. This makes it possible to study structural transitions in nucleic acids and to detect minor damages to the DNA double helix. Interactions of nucleic acids with electroactive substances can be manifested by changes both in the signals of the given electroactive substance and in the signals of the nucleic acid. In addition to the natural electroactivity of nucleic acids, noncovalently (DNA groove binders and intercalators) and covalently binding electroactive

structures of the Drew–Dickerson dodecamer (dCGCGAATTCGCG)<sub>2</sub> and the DNA hairpin d(GCGAAGC) (Figure 7) have been determined with very high precision using the isotope labeling and the alignment methodology.

Current state-of-the-art NMR technology allows studying the fragments of nucleic acids of the size up to 70 nucleotides. For molecules with more than 40 residues, the uniform <sup>13</sup>C, <sup>15</sup>N isotope labeling, which, in general, is more complicated than in proteins, has to be complemented also by residue- or site-specific labeling. For oligonucleotides with more than 50 residues partial deuteration seems to be mandatory. Despite these obstacles, the recent improvement of labeling strategies, development of new and more sophisticated NMR methods, and progress in measurements and applications of residual dipolar couplings for structure refinement have paved the way to improve the accuracy and precision of structures obtained by NMR and to study systems with larger number of nucleotides.

*See also:* **Infrared Spectroscopy:** Overview; Sample Presentation. **Nuclear Magnetic Resonance Spectroscopy:** Overview; Principles. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Nuclear Overhauser

Effect; Multidimensional Proton. **Raman Spectroscopy:** Instrumentation.

## Further Reading

- Guschlbauer W (1976) *Nucleic Acid Structure*. Heidelberg: Springer.
- Johnson WC (1996) Spectroscopic methods in nucleic acid studies. In: Fasman GD (ed.) *Circular Dichroism and the Conformational Analysis of Biomolecules*, pp. 433–468. New York: Plenum.
- Neidle S (ed.) (1999) *Oxford Handbook of Nucleic Acid Structure*. New York: Oxford University Press.
- Peticolas WL and Evertsz E (1992) Conformation of DNA *in vitro* and *in vivo* from laser Raman scattering. In: Lilley DMJ and Dahlberg JE (eds.) *Methods in Enzymology*, vol. 211, pp. 335–352. San Diego: Academic Press.
- Sanger W (1984) *Principles of Nucleic Acid Structure*. New York: Springer.
- Taillandier E and Liquier J (1992) Infrared spectroscopy of DNA. In: Lilley DMJ and Dahlberg JE (eds.) *Methods in Enzymology*, vol. 211, pp. 307–335. San Diego: Academic Press.
- Wijmenga SS and van Buuren BNM (1998) The use of NMR methods for conformational studies of nucleic acids. *Progress in NMR Spectroscopy* 32: 287–387.
- Wüthrich K (1986) *NMR of Proteins and Nucleic Acids*. New York: Wiley.

## Electrochemical Methods

**E Paleček**, Institute of Biophysics, Brno, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

DNA and RNA are complex biomacromolecules whose structures and interactions are intensively studied mainly in relation to their biological roles. The main interest of scientists is not oriented to the determination of nucleic acids in biological materials but to analysis of their nucleotide sequence, secondary and tertiary structures, specific interactions with proteins, etc. A number of diverse techniques are necessary such as absorption spectroscopy, chromatography, gel electrophoresis, and radioactive and nonradioactive labeling, which can be found in almost every nucleic acid laboratory, while other less universal techniques are applied in relation to the specific aims of the given laboratory. Electrochemical methods briefly reviewed in this article belong to the latter category. They can be extremely useful for a number of purposes, such as determination of traces

of proteins, RNA and single-stranded DNA (ssDNA) in double-stranded DNA (dsDNA) samples, detection of DNA damage and nucleotide sequencing by DNA hybridization, etc. They do not require expensive equipment and give a good sensitivity, which, in some cases, can reach levels comparable to those of radioactive tracing. In recent years, electrochemistry has gained importance in the development of the DNA sensors.

Electroactivity of nucleic acids was discovered more than 40 years ago. Nucleic acid bases undergo redox processes at electrodes and in long-chain DNA molecules the resulting electrochemical signals can be significantly influenced by their ordered higher structures. This makes it possible to study structural transitions in nucleic acids and to detect minor damages to the DNA double helix. Interactions of nucleic acids with electroactive substances can be manifested by changes both in the signals of the given electroactive substance and in the signals of the nucleic acid. In addition to the natural electroactivity of nucleic acids, noncovalently (DNA groove binders and intercalators) and covalently binding electroactive



markers can be introduced into DNA and used in their analysis. In the last decade, the above principles were intensively utilized in the development of sensors for DNA damage and hybridization.

### Working (Indicator) Electrodes

Nucleic acids have been analyzed primarily with liquid mercury and solid carbon electrodes; some work has also been done with other electrodes such as platinum, gold, indium tin oxide (ITO), silver, copper, as well as solid amalgam electrodes. The potential windows in which most solid electrodes can be used is shifted by  $\sim 1$  V to more positive values when comparing with mercury and some solid amalgam electrodes (operating between 0 and  $-2$  V against the saturated calomel electrode at neutral and alkaline pHs). Therefore, solid (not containing mercury) electrodes are better suited for studying nucleic acid oxidation, while mercury electrodes (mainly hanging mercury drop electrode (HMDE) and solid amalgam electrodes) are better for investigating nucleic acid reduction. Moreover, the atomically smooth and highly reproducible surfaces of liquid mercury are

well suited for a.c. impedance measurements, providing information about DNA interfacial properties.

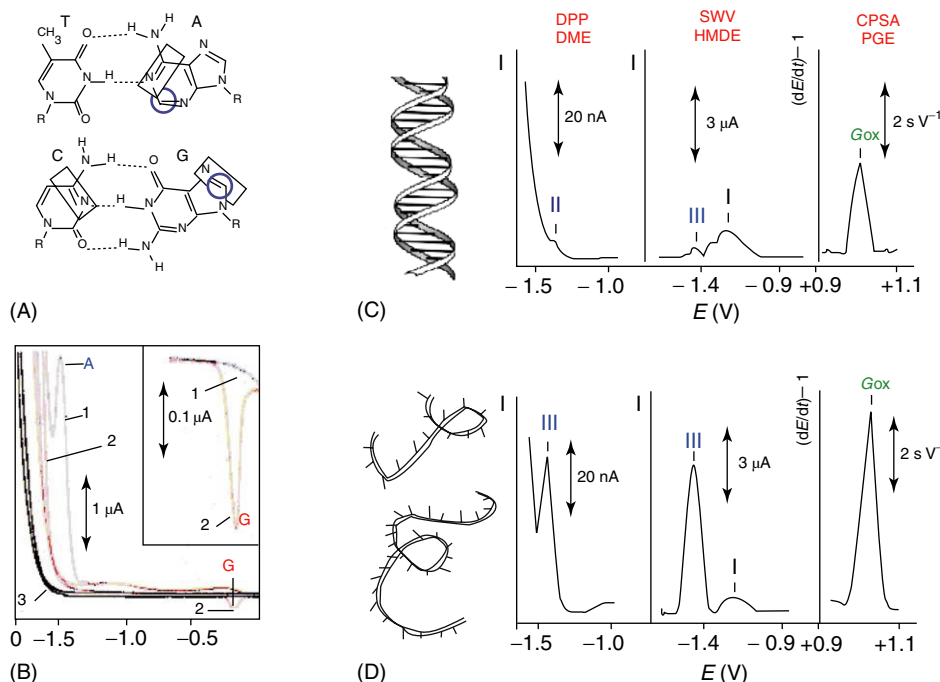
## Redox Groups in Nucleic Acids

Among nucleic acid components purine and pyrimidine bases are electroactive; deoxyribose and ribose in nucleosides and nucleotides as well as phosphate groups are inactive at most types of electrodes.

### Reduction and Oxidation of Bases

Primary reduction sites of adenine, cytosine, and guanine and oxidation sites of adenine and guanine are shown in the Watson–Crick base pairing scheme (Figure 1A).

The reduction sites of adenine and cytosine are involved in the hydrogen bond system and hidden in the double helix interior. Neither the reduction site of guanine nor the oxidation sites of adenine and guanine are involved in this system. Signals caused by reduction of adenine and cytosine are strongly influenced by the DNA secondary structure in agreement with the location of the respective reduction sites.



**Figure 1** Reduction and oxidation of single-stranded and double-stranded nucleic acids at electrodes. (A) Schematic representation of Watson–Crick base pairs and electroactive groups. Primary reduction and oxidation sites at mercury (rectangles) and carbon electrodes (circles) are shown. (B) Cyclic voltammograms of synthetic polynucleotides ( $100 \mu\text{g ml}^{-1}$ ) measured at the hanging mercury drop electrode (HMDE). (1) Polyadenylic acid; (2) Polyguanylic acid; (3) background electrolyte; *inset*, detail of peak G. (C, D) Redox signals obtained with (C)  $100 \mu\text{g ml}^{-1}$  of native (double-stranded) or (D)  $50 \mu\text{g ml}^{-1}$  of denatured (single-stranded) ssDNA by differential pulse polarography (DPP) at the static dropping mercury electrode (DME), by adsorptive stripping square-wave voltammetry (SWV) at the hanging mercury drop electrode (HMDE), and by constant current chronopotentiometric stripping analysis (CPSA) at the pyrolytic graphite electrode (PGE). (Reprinted with permission from Paleček E and Fojta M (2001) Detecting DNA hybridization and damage. *Analytical Chemistry* 73: 75A–83A; © American Chemical Society.)

**Table 1** Abilities of (A) nucleic acid components and (B) some other purine and pyrimidine derivatives to form sparingly soluble compounds with the electrode mercury<sup>a</sup>

<i>Pyrimidine derivatives</i>				<i>Purine derivatives</i>			
<i>Bases</i>		<i>Nucleosides</i>		<i>Bases</i>		<i>Nucleosides</i>	
(A)							
Cytosine <sup>b</sup>	+	Cytidine	–	Adenine	+	Adenosine	+
Thymine	+	Thymidine	–	Guanine	+	Guanosine	+
Uracil	+	Uridine	–	Xanthine	+	Xanthosine	+
5-Methylcytosine	+	Pseudouridine	+	Hypoxanthine	+	Inosine	+
5-Hydroxymethyl-cytosine	+	1-Methyladenine	+	3-Methyladenine	+	6-Methyladenine	+
(B)							
5-Acetyluracil	+	5-Formyluracil	+	2-Aminopurine	+		
5-Azauracil	+	5-Fluorouracil	+	6-Benzyladenine	+		
6-Azauracil	+	6-Methylthymine	+	8-Oxyadenine	+		
5-Bromouracil	+	5-Nitrouracil	+	Uric acid	+		
5-Chlorouracil	+	Orotic acid	+				
6-Chlorouracil	+	2-Thiouracil	+				

<sup>a</sup>Data from: Paleček E *et al.* (1981) Reaction of the purine and pyrimidine derivatives with the electrode mercury. *Bioelectrochemistry and Bioenergetics* 8: 621–631. Paleček E, Osteryoung J, and Osteryoung RA (1982) Interaction of methylated adenine derivatives with the mercury electrode. *Analytical Chemistry* 54: 1389–1394. Bouzid B and Macdonald AMG (1988) Polarographic studies of uracil derivatives. *Analytica Chimica Acta* 211: 155–173.

<sup>b</sup>Sparingly soluble compound with the electrode mercury +, is produced, –, is not produced. In CSV, the detection limit of nucleic acid bases is  $\sim 10^{-9} \text{ mol l}^{-1}$  for purines and  $10^{-8} \text{ mol l}^{-1}$  for pyrimidines with the exception of 5-methyl cytosine and 5-hydroxymethyl cytosine, which can be determined only at higher concentrations. Sparingly soluble compounds of copper(I) with bases accumulate at electrodes. Liquid and solid copper amalgam electrodes can be used to determine purine bases at concentrations down to  $10^{-11} \text{ mol l}^{-1}$  (Jelen *et al.* (2002) Determination of picogram quantities of DNA by stripping voltammetry with solid copper amalgam or mercury electrodes in the presence of copper. *Analytical Chemistry* 74: 4788–4793 and references therein).

The small difference in the intensities of the ds- and ssDNA oxidation signals (obtained with carbon electrodes) was explained by the difference in flexibility of these two DNA forms rather than by the direct influence of the secondary structure.

### Compounds with the Electrode Mercury

Groups responsible for the formation of the mercury compounds may involve exocyclic and ring nitrogens depending on the nature of the base and experimental conditions. At alkaline pH the mercury binding site of adenine is the 6-aminogroup. Purine nucleosides behave similarly to their parent bases while the pyrimidine analogs are inactive due to substitution of the pyrimidine ring by a sugar residue at N1, which is involved in the binding of mercury.

### Analysis of Nucleic Acid Components

Adenine and guanine can be analyzed at a carbon electrode and adenine and cytosine at a mercury electrode by differential pulse voltammetry (polarography) (DPV, DPP) at micromolar concentrations. Determination of bases by cathodic stripping voltammetry (CSV) (see above) gives substantially better sensitivities. In addition to the nucleic acid

components, a large number of purine and pyrimidine derivatives have been investigated (Table 1).

Most of them form sparingly soluble compounds with the electrode mercury and can be determined by CSV at nanomolar concentrations. At these concentrations 100-fold excess of various substances including proteins and nucleic acids have little influence on the CSV peak height of the nucleic acid bases. This makes it possible to analyze mixtures of bases with nucleic acids, the analysis of which by optical methods without separation is impossible. Sensitivity of the determination of purine and pyrimidine bases can be increased by one to three orders of magnitude if the analysis is performed in alkaline media in the presence of copper ions or at solid copper amalgam electrodes.

### Analysis of Nucleic Acids

Various electrochemical methods have been applied for the analysis of nucleic acids, including DPP and DPV, cyclic voltammetry, a.c. voltammetry, and constant current chronopotentiometry, which are particularly useful. Guanine (G) and adenine (A) residues in DNA and RNA molecules are oxidized at carbon electrodes; however, their voltammetric peaks are poorly developed, providing insufficient sensitivity in

the DNA analysis. Recent application of sophisticated baseline correction techniques, mainly in connection with constant current chronopotentiometry, produced well-developed oxidation peaks of DNA and RNA at carbon electrodes (Figures 1C and 1D). Cytosine (C) and A residues in ssDNA and RNA are reduced at mercury electrodes, producing reduction signals close to  $-1.4$  V at acid and neutral pH. In cyclic modes, G produces an anodic signal close to  $-0.3$  V, which is due to oxidation of the G reduction product forming at the highly negative potentials of the background discharge (Figure 1).

### Influence of DNA Structure

The primary reduction sites of A and C form a part of the Watson–Crick hydrogen bonding system (Figure 1A), being located in the interior of the DNA double helix. Reduction signals of A and C are strongly influenced by DNA structure; the reduction peak of single-stranded (denatured) calf thymus DNA can be almost 100-times higher than that of the parental duplex (native) calf thymus DNA. Similarly, the non-Faradaic signals of DNA produced by a.c. voltammetry are highly sensitive to changes in DNA structure. These signals can provide information about accessibility of bases for the interaction with the electrode. Such a strong influence of DNA structure on electrochemical signals makes the mercury electrodes well suited for detection of local conformational changes and DNA structural transitions. Mercury electrodes sensitively reflect single-strand interruptions in linear and circular DNA molecules, as well as superhelix density-dependent structural transitions in DNA. Mercury electrodes are able to discriminate between RNA and DNA making possible the most sensitive label-free determination of traces of RNA in a large excess of DNA. Oxidation peaks of A and G at carbon electrodes and anodic signal of G at mercury electrodes are less sensitive to changes in DNA structure.

### Changes in DNA Structure at the Electrode Surface

At a wide range of potentials, dsDNA is adsorbed at the electrode without any detectable unwinding. However, at certain potentials relatively slow unwinding of dsDNA at mercury was observed. No such changes were observed in covalently closed circular DNAs. At neutral and weakly alkaline pHs prolonged contact of dsDNA with the surface of the mercury electrode charged to potentials around  $-1.2$  V results in profound changes in the DNA electrochemical responses. These changes are caused by the opening of the DNA double helix at the electrode surface. They are relatively slow (tens to

hundreds of seconds) and can be observed by various techniques if dsDNA is exposed to potentials close to  $-1.2$  V for a sufficient period of time. To avoid these changes methods working with small voltage excursions during the (DME) drop lifetime should be applied. In experiments with HMDE the surface changes in the DNA structure can be minimized by proper voltage scanning. A d.c. field induced denaturation of short dsDNA molecules was recently detected by surface plasmon resonance at gold electrodes.

## Electroactive Labels and Chemical Modification of Nucleic Acids

Reduction and oxidation of natural nucleic acids is electrochemically irreversible and occurs at highly positive or highly negative potentials. To increase the sensitivity of the analysis electroactive markers were introduced into DNA which either undergo reversible electrode reaction at less extreme potentials or produce high electron yield catalytic signals. Here, only a few examples of nucleic acid electroactive labels will be mentioned, and the electrochemical properties of peptide nucleic acid (PNA) will be summarized.

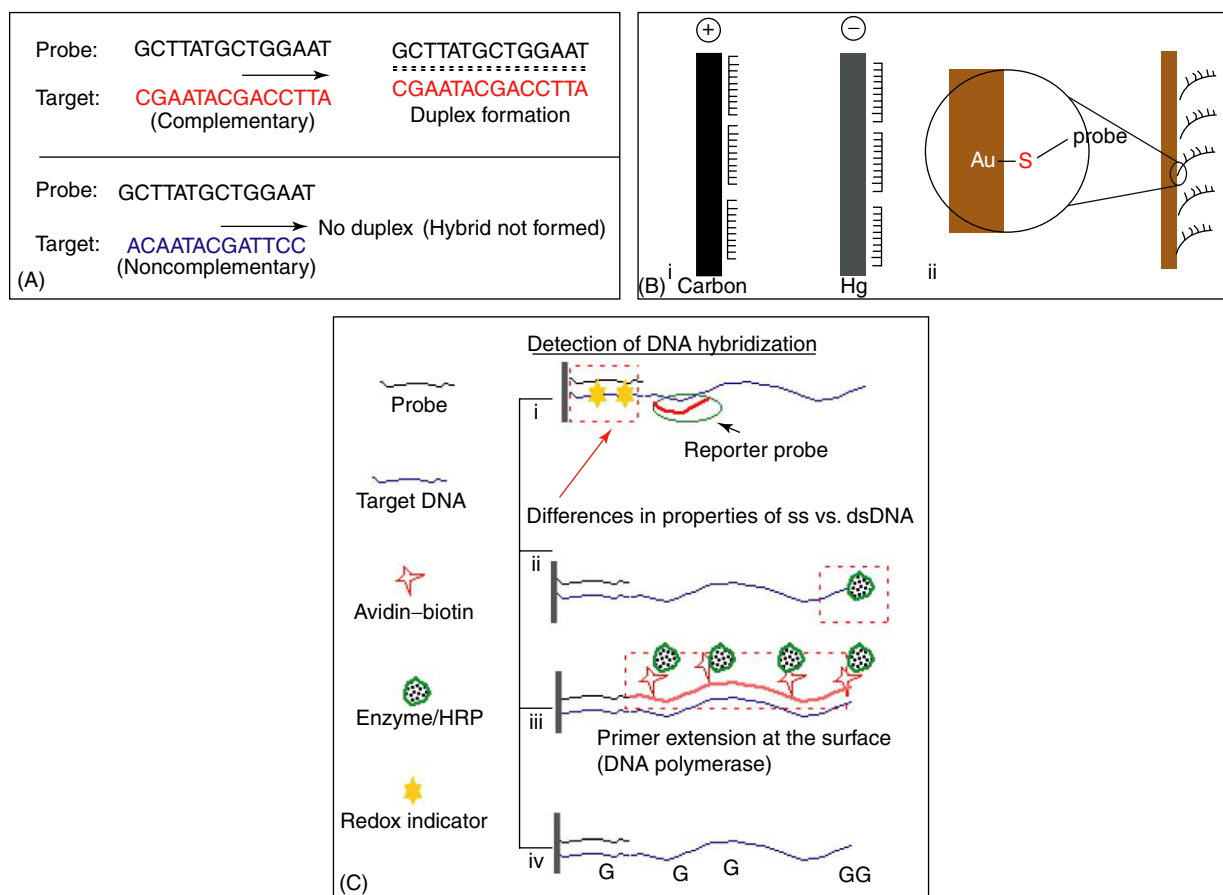
### Noncovalently Bound Labels

Molecules and ions interact with DNA in three basic ways: (1) nonspecific electrostatic binding along the exterior of the DNA double helix; (2) groove binding involving direct interaction of the bound molecule with the edges of base pairs in the DNA grooves; (3) intercalation of planar aromatic ring systems between DNA base pairs. A large number of the substances capable of interacting with DNA, particularly intercalators and groove binders, are electroactive.

Intercalating and DNA surface binding of metal chelates were studied by voltammetric methods. On addition of dsDNA to intercalators (e.g.,  $\text{Co(phen)}^{3+}$ ) a shift to more positive values was observed (suggesting stronger hydrophobic interactions) while with electrostatic binders such as  $\text{tris(2,2'-bipyridine)-osmium(II)}$ , the cathodic and anodic peak potentials shifted to more negative values. Electrogenated chemiluminescence as well as DNA redox active intercalators and groove binders were utilized in the development of the DNA hybridization sensors (see Figure 2 and below).

### Chemically Modified Nucleic Acids

Osmium tetroxide complexes were probably the first electroactive labels covalently bound to DNA. Some of these complexes, such as those with osmium tetroxide, and 2,2'-bipyridine, bind preferentially to



**Figure 2** Principles of DNA hybridization and its electrochemical detection (single-surface techniques). (A) Scheme of the formation of duplex DNA (DNA hybridization) from two complementary ss oligodeoxynucleotides (ODNs). Duplex DNA is not formed when the target DNA sequence is noncomplementary. In an electrochemical hybridization detector, one of the two strands is used as a probe which is immobilized at the electrode surface. (B) Scheme of immobilization of the probe at the electrode surface. (i) at positively charged carbon electrodes DNA is adsorbed electrostatically via the negatively charged sugar-phosphate backbone. At the hydrophobic mercury electrodes, hydrophobic bases are strongly bound to the surface. (ii) Covalent binding of thiolated ODN at gold electrodes via the sulfur atom securing oriented ODN immobilization. (C) Detection of the DNA hybridization: (i) based on the formation of the hybrid duplex DNA using either (a) a redox indicator that binds preferentially to the hybrid duplex DNA and produces an oxidation (or reduction) signal: (simple intercalators and groove binders do not provide best resolution; threading intercalators or bis-intercalators with high DNA binding constant appear to be a better choice) or (b) direct measurements of changes in DNA properties (such as conductivity) due to the duplex formation. Alternatively, a reporter (signaling) probe can be used as shown in **Figure 3B**; (ii and iii) based on the determination of the target DNA at the electrode surface; (ii) a suitable enzyme such as horseradish peroxidase (HRP) is coupled to the target DNA. Upon hybridization, an electrocatalytic current due to hydrogen peroxide reduction is measured. (iii) After DNA hybridization the DNA probe serves as a primer, which is extended at the electrode surface by DNA polymerase incorporating biotinylated nucleotides in the newly synthesized strand. HRP is then attached to the DNA using streptavidin-biotin chemistry followed by measurements as in (ii); (iv) DNA probe not containing guanine is used and presence of guanine in target DNA is detected after the hybridization using guanine oxidation signals either on indium tin-oxide or carbon electrodes. (Reprinted with permission from Paleček E and Fojta M (2001) Detecting DNA hybridization and damage. *Analytical Chemistry* 73: 75A–83A; © American Chemical Society.)

thymine residues in ssDNA, while others (e.g., Os, 1,10-phenanthroline) bind both ds- and ssDNA. They produce a high electron yield catalytic signal at about  $-1.2$  V at liquid mercury and solid amalgam electrodes and quasireversible signals between  $-0.2$  and  $-0.7$  V at both carbon and mercury electrodes. Other electroactive markers such as viologen derivatives, daunomycin, and ferrocene were later coupled to nucleic acids primarily for their potential use in biosensing.

PNA is currently investigated as a possible candidate for diagnostic and therapeutic applications in medicine in the twenty-first century. In PNA, the entire sugar-phosphate backbone of DNA or RNA is replaced by *N*-(2-aminoethyl)glycine units. This structural change produced a DNA mimetic with higher binding affinity to complementary strands as compared to unmodified DNA. In contrast to DNA and RNA with negatively charged backbone, the backbone of PNA is electrically neutral. It produces

electrochemical responses of PNA at carbon and mercury electrodes similar to DNA, i.e., A and G residues in PNA are oxidized at carbon electrodes, while at mercury electrodes reduction of A, C, and G takes place. Differences in electric charges of PNA and DNA backbones are manifested by the different adsorption behavior of these two compounds as detected by a.c. impedance at mercury electrodes. PNA has been used in the development of the DNA hybridization sensors.

## DNA-Modified Electrodes

Up to the middle of the 1980s electrochemical analysis of nucleic acids was usually performed with the electrode immersed into the nucleic acid solution. To decrease the volume of the analyzed sample a simple preparation of a DNA-modified electrode was proposed in 1986. Instead of performing the voltammetric measurements with the electrode immersed into the DNA sample, the DNA-modified electrode was first prepared by immersing the electrode into a small drop of a DNA solution (10–3  $\mu$ l). Both ss- and dsDNAs were irreversibly adsorbed either at mercury or at carbon electrodes, resisting subsequent washing. At present, nucleic acid-modified electrodes prevail in electrochemical analysis of nucleic acids; covalent (e.g., via –SH groups at the end DNA molecules to gold or mercury electrodes) and noncovalent binding (e.g., strong hydrophobic interactions of the nucleic acid bases with the mercury electrodes or electrostatic binding to positively charged carbon electrodes) is used for immobilization of nucleic acids.

## DNA and RNA Sensors

In recent years, attention of scientists in nucleic acid electrochemistry has greatly increased, not only because of high sensitivity of electrochemical methods for small changes in the DNA structure, but primarily because of expectations that electrochemical DNA sensors may become important and inexpensive tools for use in many areas of human life, including decentralized detection of DNA damage and rapid diagnostics of various diseases in the physician's office. Recent progress in the development of sensors for DNA hybridization and damage is encouraging.

### DNA Hybridization

Sequencing of individual eukaryotic genomes is still too laborious and expensive. New technologies are therefore sought and among them sequencing by DNA hybridization appears most promising. At

present, DNA sequencing can be performed by means of the DNA hybridization sensors with optical detection. Recently, it has been shown that this detection can be complemented by voltammetric and chronopotentiometric methods, which are simpler and should be less expensive. DNA hybridization sensors are well suited for decentralized diagnostics and other purposes, including determination of bacteria and viruses in biological warfare.

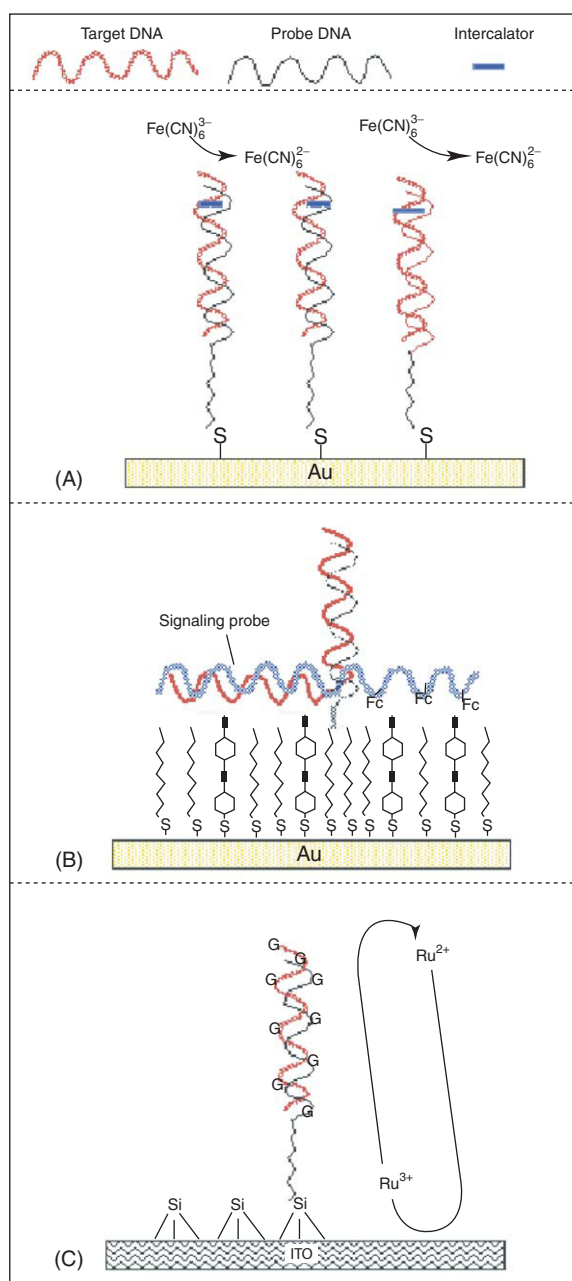
### Single-Surface Techniques

Up to 2001, DNA hybridization sensors used single-stranded, short (15–25 nucleotides) oligodeoxynucleotide (ODN, probe DNA) immobilized on an electrode (see **Figures 2** and **3**), using PNA as a probe, appeared advantageous in the detection of point mutations. The ODN-modified electrode was immersed in target DNA solution to test its nucleotide sequence. When the sequence of target DNA matched that of the probe (based on the complementary Watson–Crick pairing), a probe–target (hybrid) duplex DNA was formed at the electrode surface. The hybridization event (DNA duplex formation) was detected electrochemically in various ways (**Figure 2**).

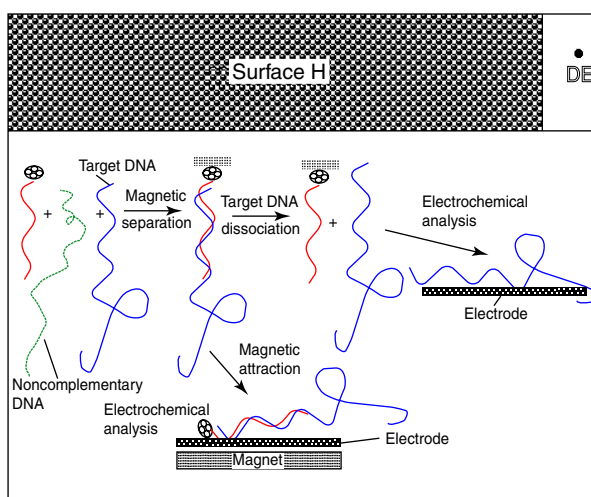
This system worked quite well with synthetic ODNs when probe and target DNAs were of about the same lengths. In a real DNA sequence analysis with longer polymerase chain reaction (PCR) products, viral or chromosomal DNAs, the target DNAs can be substantially longer than the probe. With longer target DNAs difficulties connected with the nonspecific DNA adsorption frequently arose, resulting in a loss of specificity and decreased sensitivity. Nevertheless, several promising single-surface techniques shown in **Figure 3** have been developed and recently reviewed by Popovich and Thorp. The eSensor™ DNA Biochip whose principles are shown in **Figure 3B** is already on the market. It is produced by Motorola. The critical components (capture probe, target probe, and signaling probe) are in the cartridge and the current produced by this system is measured and interpreted by the eSensor™ DNA Detection Reader and Software. Various eSensor™ assays are being developed, such as cytochrome P450 and cystic fibrosis assays. More details can be obtained from eSensor@motorola.com. Due to the fast progress in nucleic acid electrochemistry and sensor development, new variants of electrochemical DNA sensors may appear in the market very soon.

### Double-Surface Techniques

Optimum properties of the hybridization surface greatly differ from those of the (electrode) surface



**Figure 3** Recent single-surface DNA hybridization sensors. (A) Approach of Barton *et al.*, as described in Boon EM *et al.*, (2000) *Nature Biotechnology* 18: 1096. Alkanethiol-modified probes are attached to gold electrodes and hybridized to target DNA. In the presence of redox-active intercalator, reduction of ferricyanide in solution is observed. (B) Approach of Kayem *et al.*, as described in Umek RM *et al.* (2001) *Journal of Molecular Diagnosis* 3: 74. Alkanethiol-modified capture probes are attached to gold surfaces in the presence of phenylacetylene thiols. Hybridization of target DNA and signaling probes position ferrocene labels (Fc) adjacent to the phenylacetylene thiols, which communicate with the gold electrode. Diluent thiols are terminated with polyethylene glycol (not shown). (C) Approach of Thorp *et al.* (1998). Silane-modified capture probes are attached to ITO electrodes. Hybridization of target allows for electrocatalytic oxidation of guanine by  $\text{Ru}(\text{bpy})_3^{3+/2+}$ . Adapted from Popovich ND and Thorp HH (2002), with permission.



**Figure 4** Principles of double-surface (DS) DNA hybridization sensors. In difference to the single-surface techniques (Figures 2 and 3) in the DS sensors the DNA hybridization takes place at one surface and electrochemical detection of the hybridization event at another surface. Hybridization is performed at a relatively large surface H best suited for this purpose (e.g., magnetic beads with covalently (or avidin–biotin) bound DNA probe). Then the target DNA is released from the surface H and determined at a small detection electrode (DE, micro-, or nanoelectrodes can be used for this purpose). Alternatively, the magnetic beads with the bound target DNA can be attracted to the electrode (under which a magnet is placed) and DNA can be determined without dissociating it from the beads. In contrast to the single-surface techniques, in which the reporter (signaling) probe (RP) (see Figure 3) has to bind close to the capture probe (to communicate with the electrode). In the DS techniques the RPs can bind to any site at the target DNA. Label-free determination of long target DNA at subfemtomole level was proposed (Jelen F *et al.* (2002) *Analytical Chemistry* 74: 4788–4793). Various electroactive DNA labels were designed, such as metal nanoparticles, osmium complexes, enzymes, and enzyme-linked immunoassays. Carbon, solid dental amalgam, and mercury electrodes were mainly applied.

used for the electrochemical detection, thus making optimization of both hybridization and detection at a single surface difficult. Recently, a new method has been proposed in which DNA hybridization is performed at one surface (surface H) and electrochemical detection at the detection electrode (DE) (Figure 4).

Commercially available magnetic beads with attached probe ODN served as surface H; due to minimum nonspecific DNA adsorption at the beads, very high specificity of the DNA hybridization was achieved. Optimum DE was chosen only with respect to the given electrode process. Detection of relatively long target DNAs was possible (1) by enzyme-linked immunoassay of target DNA modified by osmium tetroxide, 2,2'-bipyridine (Os,bipy) at carbon DEs (direct determination of Os,bipy at mercury and carbon electrodes was also performed) and (2) by using CSV at mercury or solid mercury



amalgam DEs for the determination of purine bases, released from DNA by acid treatment. CSV in the presence of copper allowed determination of sub-nanomolar concentrations of DNA. Long target DNAs in the presence of a large excess of nonspecific DNA were easily analyzed by techniques mentioned under points (1) and (2). These techniques are well suited also for analysis of RNAs, which have been analyzed to a lesser extent.

Many advantages of the double-surface technologies over the single-surface ones were discussed in the literature. In addition to higher specificity and easy work with long target DNAs, it is the larger choice of electrodes (such as mercury and solid amalgam ones) and methods for the DNA determination that are not limited by requirements for easy immobilization of the DNA probe at the electrode surface. This greatly increases possibilities in DNA detection, including use of reporter probes (which can bind to target DNA regardless of the distance from the capture probe at the surface), application of well-established techniques, such as avidin–biotin, immunoassays, nanoparticles, to detect the target sequence. Recent application of the double-surface technology has already resulted in an increase in sensitivity of DNA detection by several orders of magnitude, simultaneous detection of multiple DNA targets, determination of lengths of DNA repetitive sequences such as those characteristic for some neurodegenerative genetic diseases. Magnetic beads can be incorporated into microfluidic systems and are well suited for the development of the lab-on-chip suitable for analysis of DNA hybridization and DNA damage starting with a drop of blood. Such a module-composed device is being developed by November AG in collaboration with Siemens Medical Solutions. The central element of this device is an exchangeable credit card-sized cartridge with a microchannel structure that integrates all the steps including nucleic acid isolation, PCR, and electrochemical detection of the hybridization event. More details can be found at <http://www.november.de>.

With both single- and double-surface techniques detection of single-base mismatches (point mutations) is possible. At present principles of the electrochemical detection of the DNA hybridization event and determination of various genomic nucleotide sequences, after the PCR amplification of DNA, are known. On the other hand, a lot of work has to be done to develop and optimize inexpensive and simple electrochemical DNA sensors with performance comparable to that of DNA arrays based on optical detection. Determination of nucleotide sequences without DNA amplification represents a great challenge. To reach this goal it will be necessary to

increase further the sensitivity of the electrochemical detection and to solve the problem of an extremely high excess of nonspecific DNA that interferes with the analysis of unamplified genomic sequences.

In recent years, substantial progress in studies of the DNA conductivity was achieved but many questions still remain to be answered. One of the most exciting discoveries is the change in the electrical properties of dsDNA upon introduction of a base mismatch, demonstrated by a number of authors using different approaches. Elucidation of the DNA conductivity may help not only to understand better many biological processes, including mutagenesis, carcinogenesis, and DNA repair, but it can be also utilized for the development of DNA sensors.

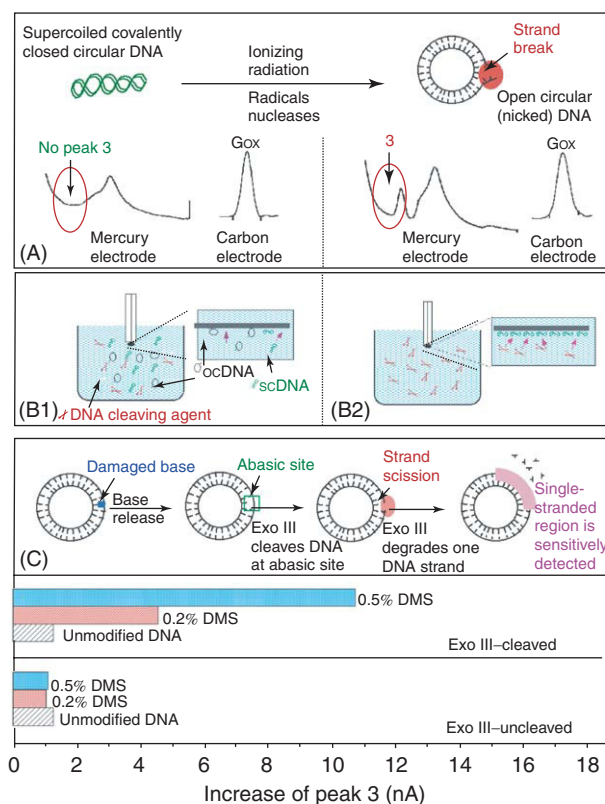
### DNA Damage

DNA in living organisms is a critical target of various agents, including oxidative metabolism, photoexcited dyes, radiation, and other agents that induce damage of bases and/or ruptures of the sugar–phosphate backbone. Products of DNA damage may be involved in mutagenesis, carcinogenesis, and cell lethality. Development of sensitive and fast assays of DNA damage has recently attracted increasing attention among analytical chemists. The sensitivity of these assays must be high enough to allow detection of one damaged nucleotide in  $10^4$ – $10^7$  intact nucleotide residues in microgram amounts of DNA.

A number of methods were developed involving separation of DNA; damaged entities are separated by liquid chromatography, gas chromatography, or capillary electrophoresis and detected online by various techniques such as mass spectrometry, amperometry, fluorimetry, and radioactivity measurement. Alternatively, DNA lesions can be measured without hydrolyzing DNA by immunoassays, sedimentation, or gel electrophoresis that quantify the number of strand breaks in unhydrolyzed DNA samples. These techniques are useful, but not one of them can serve as a fast response sensor providing information about the DNA damage in seconds or minutes.

Recently, electrochemical methods for the determination of DNA strand breaks at an electrode have been developed. With these methods, one single-strand break (ssb) among  $>2 \times 10^5$  intact phosphodiesteric bonds can be determined in submicrogram amounts of DNA. Circular covalently closed supercoiled (sc) DNA is attached to HMDE. If there is a strand break a new electrochemical signal appears (Figure 5A).

This signal is due to bases at the strand interruption interacting with the electrode and can be obtained by



**Figure 5** Detection of DNA damage with supercoiled (sc) DNA-modified electrodes. (A) AC voltammetric peak 3 is due to desorption of DNA segments adsorbed at mercury electrodes via bases. In scDNA bases are not available for the interaction with the electrode surface and peak 3 is not observed. Introduction of a strand break (sb) in scDNA results in a better accessibility of bases in the vicinity of the sb and in formation of peak 3. Peak 3 can be thus used for the detection of DNA cleavage by different agents, including radiations and free radicals. In difference to the mercury electrodes, carbon electrodes are little sensitive to sb in DNAs. (B1) scDNA is treated with a DNA-cleaving agent in solution followed by DNA adsorption at the electrode. After washing, the electrode is transferred into a cell and DNA voltammetric response is measured. (B2) Intact scDNA adsorbed at the electrode serves as a sensitive layer of a biodetector for DNA damage. (C) Damage to DNA bases can be transformed to DNA strand breaks and/or single-stranded DNA regions by DNA repair enzymes. For example, exonuclease III (Exo III) cleaves the DNA sugar-phosphate backbone at abasic sites resulting from spontaneous release of dimethylsulfate (DMS)-methylated guanine. This strand cleavage is followed by exonucleolytic degradation of one DNA strand creating segments of ssDNA which are sensitively detected through peak 3. DNA lacking damaged bases are not cleaved showing an electrochemical response similar to that of undamaged DNA. (Reprinted with permission from Paleček E and Fojta M (2001) Detecting DNA hybridization and damage. *Analytical Chemistry* 73: 75A–83A; © American Chemical Society.)

different methods such as a.c. voltammetry and constant current chronopotentiometry. Either the DNA which was damaged in solution can be immobilized at the electrode surface (to determine the strand breaks) or the electrode modified with intact scDNA

can be used as a recognition layer to detect DNA-damaging agents present in the medium into which the electrode is immersed. Primary strand breaks and/or damage to bases, transformed by the activity of enzymes or chemicals into breaks or single-stranded regions (Figure 5), can be determined. To overcome the difficulties associated with liquid mercury electrode in the field, the HMDE can be replaced by a mercury film carbon or solid dental amalgam electrodes. This electrochemical DNA detector responds in seconds, being much faster than the currently used electrophoretic methods.

## DNA-Protein Interactions

Interactions of DNA with proteins and peptides are of crucial importance in many life processes. These interactions are intensively studied by various methods but electrochemical analysis has been little utilized for this purpose. This is surprising because not only DNA is electroactive but also proteins produce various kinds of electrochemical signals. Recently, a great increase in sensitivity in protein and peptide analysis has been reached by means of constant current chronopotentiometric stripping analysis (CPSA). This method appears very attractive for protein analysis offering sensitivity by one to three orders of magnitude higher than that of previously used polarographic or voltammetric methods. CPSA signals of proteins and peptides at carbon electrodes are based on the oxidation of tyrosine and tryptophan residues and take advantage of the sophisticated baseline correction available in CPSA. At the mercury electrodes peptides and proteins yield catalytic peak H at highly negative potentials (close to  $-1.7$  V against SCE) allowing determination of some peptides and proteins at femto- and subfemtomole levels. Such a high sensitivity may find use in proteomics, e.g., in the detection of proteins expressed at low level. As a result of the recent progress in the analysis of proteins and nucleic acids the electrochemical techniques appear ready for studies of sequence-specific and nonspecific DNA-protein interactions. The number of proteins known to interact with DNA is large, including transcription factors, antibodies, oncoproteins. Electrochemical studies may involve not only DNA-protein interactions in solution but also interactions with one component immobilized at the electrode surface and the other in solution; introducing an electric charge to the electrode (resembling the situation at living cell interfaces) during the DNA-protein interactions might provide a new biophysical dimension to these studies. Recently, some reports about electrochemical analysis of the tumor suppressor protein p53 and its DNA binding

have appeared. Also, changes in the DNA conductivity resulting from specific interactions of some proteins with DNA have been reported.

## Acknowledgments

The author is indebted to Dr Z. Pechan for his help with the preparation of the manuscript. This work was supported by grants of the Grant Agency of the Czech Republic 204/03/0566 and of the Grant Agency of the Academy of Sciences of the Czech Republic S5004355 and KSK4055109.

*See also:* **DNA Sequencing. Enzymes:** Enzyme-Based Electrodes. **Forensic Sciences:** Blood Analysis. **Immunoassays, Techniques:** Enzyme Immunoassays. **Microelectrodes. Polarography:** Techniques; Organic Applications. **Purines, Pyrimidines, and Nucleotides. Sensors:** Chemically Modified Electrodes. **Voltammetry:** Organic Compounds.

## Further Reading

- Brabec V, Vetterl V, and Vrána O (1994) Electroanalysis of biomacromolecules. In: Milazzo G (ed.) *Treatise on Bioelectrochemistry*, vol. 2, pp. 287–359. Basel: Birghäuser.
- Fojta M (2002) Electrochemical sensors for DNA interactions and damage. *Electroanalysis* 14: 1449–1463.
- Paleček E (1983) Modern polarographic (voltammetric) techniques in biochemistry and molecular biology. II. Analysis of macromolecules. In: Milazzo G (ed.) *Topics in Bioelectrochemistry and Bioenergetics*, vol. 5, pp. 65–151. Chichester: Wiley.
- Paleček E (1996) From polarography of DNA to microanalysis with nucleic acid-modified electrodes. *Electroanalysis* 8: 7–14.
- Paleček E (2002) Past, presence and future of nucleic acids electrochemistry. *Talanta* 56: 809–821.
- Paleček E and Jelen F (2002) Electrochemistry of nucleic acids and development of DNA sensors. *Critical Reviews of Analytical Chemistry* 32: 261–270.
- Paleček E, Fojta M, and Jelen F (2002) New approaches in the development of DNA sensors. Hybridization and electrochemical detection of DNA and RNA at two different surfaces. *Bioelectrochemistry* 56: 85–90.
- Paleček E, Fojta M, Jelen F, and Vetterl V (2002) Electrochemical analysis of nucleic acids. In: Bard AJ and Stratsmann M (eds.) *The Encyclopedia of Electrochemistry*, vol. 9 (Wilson G (ed.)), pp. 365–430. Weinheim: Wiley-VCH.
- Popovich ND and Thorp HH (2002) New strategies for electrochemical nucleic acid detection. *Interface* 11(4): 30–34.
- Sequaris JM (1992) Analytical voltammetry of biological molecules. In: Smyth MR and Vos JG (eds.) *Comprehensive Analytical Chemistry*, vol. XXVII, Analytical Voltammetry, pp. 115–157. Amsterdam: Elsevier.
- Talanta (2002) Special issue on ‘Electrochemistry of nucleic acids and development of electrochemical DNA sensors’. vol. 56: 805–976.
- Thorp HH (1998) Cutting out the middleman: DNA biosensors based on electrochemical oxidation. *Trends in Biotechnology* 16: 117–121.
- Wang J (1999) Towards genoelectronics: electrochemical biosensing of DNA hybridization. *Chemistry—A European Journal* 5: 1681–1685.
- Wang J (2002) Electrochemical nucleic acid biosensors. *Analytica Chimica Acta* 469: 63–71.

## Immunoassays

**J M Wages, Jr**, Palmetto Consulting and Research, Tupelo, MS, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Nucleic acid analysis involves the determination of particular sequences of bases. Because of the low concentrations of individual sequences in most samples, enzymatic amplification processes such as polymerase chain reaction (PCR) must be integrated into the assay to enrich the sample for the sequences of interest. These amplification processes are never completely specific, and unless the assay is to rely solely on sequence selection during PCR, postamplification analysis is required to confirm the identity of

the amplified sequences. Hybridization to an oligonucleotide probe is the most common means to confirm specificity of deoxyribonucleic acid (DNA).

Enzyme-linked immunosorbent assays (ELISAs) are sensitive enough to detect relevant concentrations of small molecules and proteins. Their detection limits are inadequate for direct analysis of DNA in most cases. For this reason, it was not until the development of PCR that nucleic acid analysis became routine. Prior to PCR, the large quantities of nucleic acid required for hybridization precluded routine use and mandated the use of radioactive probes. After PCR, the abundance of amplified sequences allowed many methods of detection. In this environment, it was natural that immunochemical technologies first developed for immunoassay would be applied to analysis of amplified DNA.

have appeared. Also, changes in the DNA conductivity resulting from specific interactions of some proteins with DNA have been reported.

## Acknowledgments

The author is indebted to Dr Z. Pechan for his help with the preparation of the manuscript. This work was supported by grants of the Grant Agency of the Czech Republic 204/03/0566 and of the Grant Agency of the Academy of Sciences of the Czech Republic S5004355 and KSK4055109.

*See also:* **DNA Sequencing. Enzymes:** Enzyme-Based Electrodes. **Forensic Sciences:** Blood Analysis. **Immunoassays, Techniques:** Enzyme Immunoassays. **Microelectrodes. Polarography:** Techniques; Organic Applications. **Purines, Pyrimidines, and Nucleotides. Sensors:** Chemically Modified Electrodes. **Voltammetry:** Organic Compounds.

## Further Reading

- Brabec V, Vetterl V, and Vrána O (1994) Electroanalysis of biomacromolecules. In: Milazzo G (ed.) *Treatise on Bioelectrochemistry*, vol. 2, pp. 287–359. Basel: Birghäuser.
- Fojta M (2002) Electrochemical sensors for DNA interactions and damage. *Electroanalysis* 14: 1449–1463.
- Paleček E (1983) Modern polarographic (voltammetric) techniques in biochemistry and molecular biology. II. Analysis of macromolecules. In: Milazzo G (ed.) *Topics in Bioelectrochemistry and Bioenergetics*, vol. 5, pp. 65–151. Chichester: Wiley.
- Paleček E (1996) From polarography of DNA to microanalysis with nucleic acid-modified electrodes. *Electroanalysis* 8: 7–14.
- Paleček E (2002) Past, presence and future of nucleic acids electrochemistry. *Talanta* 56: 809–821.
- Paleček E and Jelen F (2002) Electrochemistry of nucleic acids and development of DNA sensors. *Critical Reviews of Analytical Chemistry* 32: 261–270.
- Paleček E, Fojta M, and Jelen F (2002) New approaches in the development of DNA sensors. Hybridization and electrochemical detection of DNA and RNA at two different surfaces. *Bioelectrochemistry* 56: 85–90.
- Paleček E, Fojta M, Jelen F, and Vetterl V (2002) Electrochemical analysis of nucleic acids. In: Bard AJ and Stratsmann M (eds.) *The Encyclopedia of Electrochemistry*, vol. 9 (Wilson G (ed.)), pp. 365–430. Weinheim: Wiley-VCH.
- Popovich ND and Thorp HH (2002) New strategies for electrochemical nucleic acid detection. *Interface* 11(4): 30–34.
- Sequaris JM (1992) Analytical voltammetry of biological molecules. In: Smyth MR and Vos JG (eds.) *Comprehensive Analytical Chemistry*, vol. XXVII, Analytical Voltammetry, pp. 115–157. Amsterdam: Elsevier.
- Talanta (2002) Special issue on ‘Electrochemistry of nucleic acids and development of electrochemical DNA sensors’. vol. 56: 805–976.
- Thorp HH (1998) Cutting out the middleman: DNA biosensors based on electrochemical oxidation. *Trends in Biotechnology* 16: 117–121.
- Wang J (1999) Towards genoelectronics: electrochemical biosensing of DNA hybridization. *Chemistry—A European Journal* 5: 1681–1685.
- Wang J (2002) Electrochemical nucleic acid biosensors. *Analytica Chimica Acta* 469: 63–71.

## Immunoassays

**J M Wages, Jr**, Palmetto Consulting and Research, Tupelo, MS, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Nucleic acid analysis involves the determination of particular sequences of bases. Because of the low concentrations of individual sequences in most samples, enzymatic amplification processes such as polymerase chain reaction (PCR) must be integrated into the assay to enrich the sample for the sequences of interest. These amplification processes are never completely specific, and unless the assay is to rely solely on sequence selection during PCR, postamplification analysis is required to confirm the identity of

the amplified sequences. Hybridization to an oligonucleotide probe is the most common means to confirm specificity of deoxyribonucleic acid (DNA).

Enzyme-linked immunosorbent assays (ELISAs) are sensitive enough to detect relevant concentrations of small molecules and proteins. Their detection limits are inadequate for direct analysis of DNA in most cases. For this reason, it was not until the development of PCR that nucleic acid analysis became routine. Prior to PCR, the large quantities of nucleic acid required for hybridization precluded routine use and mandated the use of radioactive probes. After PCR, the abundance of amplified sequences allowed many methods of detection. In this environment, it was natural that immunochemical technologies first developed for immunoassay would be applied to analysis of amplified DNA.

Only limited development of new methodologies has taken place for immunochemical analysis of nucleic acids. Most published methods rely on modifications to classical DNA probe hybridization or immunoassay methods, with considerable blending of the two. For example, some methods employ immobilized oligonucleotide probes to capture the analyte DNA followed by immunoenzymatic detection. Other methods use immunocapture followed by detection with an enzyme-labeled DNA probe. Distinctly new methodologies mostly impact on assay formats (e.g., DNA microarrays and *in situ* hybridization) and detection reagents (e.g., chemiluminescent enzyme substrates).

## Immunoreagents for Nucleic Acid Detection

Antibodies are immunoreagents. Considering the ligand–receptor basis for immunochemistry, avidin and biotin are often considered immunochemical reagents. The literature is replete with assays that employ immunological methods to increase sensitivity or specificity of nucleic acid analysis.

### Antibodies

Monoclonal or polyclonal antibodies comprising the complete immunoglobulin structure or recombinant single-chain antibodies, even incorporating only the variable regions, can be used in two modes: as immunocapture or immunodetection reagents.

A significant limitation on the use of antibodies for DNA analysis is the relatively low affinity reflected in the dissociation constant ( $K_d$ ).  $K_d$  for an antibody is defined as the antigen concentration required to saturate one-half the binding sites on the antibody. Hence, one-half  $K_d$  can be considered the effective detection limit. Dissociation constants for antibody–antigen interactions usually fall within the range of  $10^{-7}$ – $10^{-11}$  mol l<sup>-1</sup>. Yet, the concentration of analyte DNA is typically much lower. For example, 1 µg of DNA in 100 µl contains  $\sim 10^5$  copies of a single-gene sequence or  $10^{-15}$  mol l<sup>-1</sup>. Rare targets can be even more dilute, approaching zero in the clinically asymptomatic stage. For example, 1000 virions ml<sup>-1</sup> of plasma represents an analyte concentration of  $10^{-18}$  mol l<sup>-1</sup>.

### Avidin and Biotin

Avidin is a tetrameric glycoprotein isolated from egg white. Bacterial streptavidin lacks carbohydrate side chains. Both proteins bind four biotin molecules per tetramer with dissociation constants approaching  $10^{-15}$  mol l<sup>-1</sup> for avidin and  $\sim 10^{-14}$  mol l<sup>-1</sup> for

streptavidin. Affinity is greater than that of most monoclonal antibodies for their haptens. Streptavidin has gained a reputation for lower backgrounds due to the lack of carbohydrate side chains that promote nonspecific interactions, although design of an efficient blocking reagent is an advantageous way to approach the problem, while maintaining the tighter binding and lower cost of avidin.

### Nucleic Acid Probes

Cooperativity in base-pairing between two strands of DNA results in tight binding for perfectly matched oligonucleotides that depends on sequence and length of the probe and on ionic strength, ionic composition, and temperature. The feasibility of using oligonucleotides for hybridization to targets approaching infinite dilution is amply demonstrated by PCR primers.

## Assay Considerations

### Qualitative Assays

Immunochemical methods perform well to rapidly screen hundreds or thousands of samples for the presence of the target gene sequence. Immunoreagents in combination with high-performance PCR can enable low-cost, high-throughput detection of very few copies of analyte.

### Quantitative Assays

Immunoreagents are well suited for quantitation of amplified DNA. However, the exponential nature of PCR amplification makes it difficult to extrapolate from the amount of amplified DNA to the amount of starting material. The most reliable quantitative PCR methods involve real-time assay of amplified DNA during the exponential phase of the amplification reaction. Because the analyte of interest is usually the DNA template, not the amplification product, immunochemical methods at their present stage of development are generally best applied to an unamplified template where concentrations permit direct analysis or for qualitative assays.

### Sequence Discrimination

Molecular genetics assays can be designed to amplify both normal and mutant alleles, then to determine which allele is present in a postamplification phase of the analysis. For detection of single base changes, a generic amplification is conducted with primers that flank the site of the mutation. Following PCR, oligonucleotide probes for wild type or mutant are

hybridized under the requisite stringency followed by immunoenzymatic detection.

### Assay Components

Immunoassays for nucleic acids usually employ a solid phase on which the analyte is immobilized either before or after hybridization. Two general configurations, immunocapture or hybridization capture (enzyme-linked oligonucleotide-sorbent assay or ELOSA), are possible (Figure 1). Careful design and optimization of each component can yield a high-performance assay. Typical components for a micro-well PCR immunoassay are provided in Table 1.

### The Solid Phase

The ideal solid phase would permit specific, directional attachment of binding moieties, then complete blocking of the remaining surface to analyte, detection reagents, and visualization reagents. Such a solid

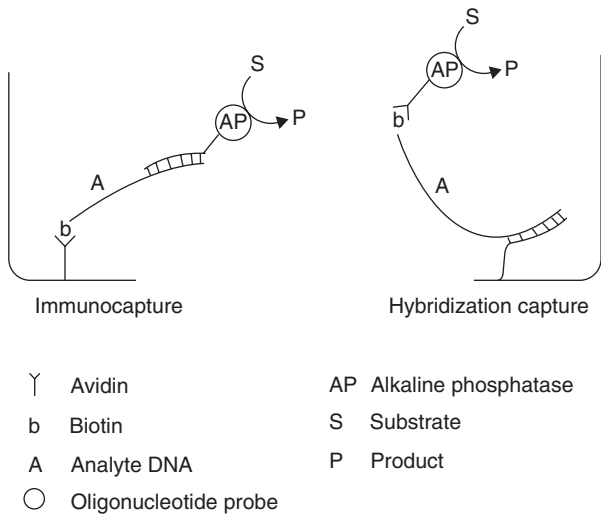
support would be highly hydrated and nonreactive. No such ideal solid phase has yet been described.

In general, any material that binds nucleic acids can be used as the solid phase. The ideal solid phase is one with high selectivity for nucleic acids over proteins. The solid phase must not react with the enzyme substrate itself. A final criterion is that the solid phase must bind sufficiently tightly so that nucleic acids will not be lost during blocking, washing, or detection steps. Solid phases that allow covalent immobilization of DNA are preferable but are not required.

Original ELISA plates were polystyrene and were not suitable for hybridization at temperatures much above 50°C due to deformation of the plastic. Plates of polycarbonate, polyvinyl, and other materials and with proprietary surface treatments to reduce non-specific binding are now available. Novel surfaces, including nonplastic solid supports, are incorporated into commercial products. Polypropylene exhibits very low binding of DNA and is not suitable as a solid phase without surface modification.

Classic methods for nucleic acid detection relied on electrophoresis followed by blotting to filter membranes. Membranes that have been used for this purpose include nitrocellulose, nylon (either unmodified, positively charge modified, or negatively charge modified), and poly(vinylidene fluoride). Nucleic acids bind very tightly to such membranes regardless of the net charge on the membrane. The use of negatively charged nylons may be advantageous for background reduction due to charge repulsion of protein reagents. DNA microarray chips use silica-based solid phases to which DNA is bound either covalently or by adsorption.

Polystyrene particles, sometimes called 'latex' microparticles (not to be confused with latex rubber material), and paramagnetic microparticles are widely used.



**Figure 1** Immune complexes formed in typical PCR immunoassay. The two general configurations for DNA immunoassay that are shown differ in the means of capture. Analyte DNA is assumed to have been amplified with a 5'-biotinylated primer. In the immunocapture approach, biotin provides the means for capture onto an avidin-coated multiwell plate. In the hybridization approach, the analyte is captured by hybridization to an immobilized probe followed by immunoenzymatic detection with an avidin-alkaline phosphatase conjugate.

### The Capture Reagent

With the exception of materials to which nucleic acids bind directly (e.g., nylon, silica), an intermediary is required to bind the DNA or ribonucleic acid (RNA) to the solid phase. Two general approaches have been used: immunocapture and hybridization capture (ELOSA) (Figure 1).

**Table 1** Components of a typical PCR immunocapture assay

Reagent	A typical composition
Plate coating solution	50 µg ml <sup>-1</sup> avidin, 50 mmol l <sup>-1</sup> sodium hydrogencarbonate, 150 mmol l <sup>-1</sup> NaCl, pH 8.5
Wash solution	Phosphate-buffered saline (0.02 mol l <sup>-1</sup> sodium phosphate, 0.85% NaCl, pH 7.4), 1% Tween 20
Blocking solution	Wash solution + 5% fish skin gelatin
Hybridization solution	Alkaline phosphatase labeled probe diluted to 10 nmol l <sup>-1</sup> in blocking solution



**Immunocapture** Immunocapture uses a solid phase coated with avidin or streptavidin to capture nucleic acids containing biotin. Biotin is introduced into the nucleic acids by the use of a biotinylated deoxyribonucleoside triphosphate (dNTP) or a 5'-biotinylated primer during PCR amplification. Haptens such as digoxigenin or fluorescein are also widely used in combination with monoclonal antibodies.

Immunocapture requires adequate loading of the solid phase to bind not only the biotinylated analyte DNA but also the biotinylated dNTP or primer that is present in large excess. dNTPs in a typical amplification reaction are each present at  $200\ \mu\text{mol l}^{-1}$ . Modified dNTP is added at  $1\text{--}10\ \mu\text{mol l}^{-1}$  to avoid inhibiting amplification. Variable incorporation of modified dNTPs by DNA polymerase is affected by DNA sequence such that the use of modified dNTPs is problematic as a general approach. Biotinylated primers are typically used at  $0.1\text{--}0.5\ \mu\text{mol l}^{-1}$ . High-purity primers are routinely synthesized using biotin phosphoramidites on automated DNA synthesizers and can be purified by HPLC if necessary. A label on the 5'-end of a PCR primer has no practical effect on amplification, unlike the presence of a modified dNTP. For these reasons, modified primers are the preferred approach.

Adsorption of an immunocapture reagent to polystyrene wells in a microtiter plate is as simple as antibody coating of an ELISA plate. Coating for a DNA probe assay should use a plate that is qualified for high protein binding capacity and low well-to-well variation. Coating solution should contain at least  $100\ \mu\text{g ml}^{-1}$  avidin for maximum loading. Because passive adsorption depends on hydrophobic interactions, somewhat higher loadings may be obtained with pH at the isoelectric point, and coating solutions typically contain  $0.1\text{--}0.2\ \text{mol l}^{-1}$  salt. Coated plates should be washed with the same 'wash solution' that will be used in the assay to remove incompletely adsorbed proteins.

**Hybridization capture** The alternative to immunocapture is hybridization capture (ELOSA). Oligonucleotide probes are coupled to the solid phase with

a suitable spacer to minimize steric hindrance and electrostatic repulsion from the surface. Analyte nucleic acid is captured by hybridization to the immobilized probe. Many approaches have been described for chemically binding oligonucleotides to polystyrene surfaces, and a few have used passive adsorption. Surface-modified plates suitable for binding amine-modified oligonucleotides are commercially available.

Assays that employ an oligonucleotide capture probe are termed ELOSA by analogy to ELISA.

### Reporter Groups

Because of the need for sensitivity, most labels for DNA analysis are enzymes or highly efficient fluorophores such as phycobiliproteins. The most commonly used enzyme labels are calf intestinal alkaline phosphatase and horseradish peroxidase (Table 2). Beta-galactosidase is used primarily for gene expression monitoring. Other enzymes such as glucose-6-phosphate dehydrogenase have not found as wide utility as phosphatase and peroxidase.

A reporter enzyme may be directly coupled to the DNA probe such that no secondary binding step is required. Alternately, the enzyme is conjugated to avidin or a monoclonal antibody that binds to a hapten label on the DNA probe. Assays that use direct enzyme-labeled probes are capable of extremely low backgrounds and can be operationally simpler and hence less expensive, faster, and more amenable to high-throughput screening.

**Enzyme substrates** Considerable development has taken place in the field of new enzyme substrates, most notably the addition of several stable chemiluminescent substrates for alkaline phosphatase and peroxidase (Table 3).

Colorimetric reagents generate a color change upon reaction with the reporter enzyme. Color changes are advantageous for rapid tests because they can be detected visually. Precise quantitation or assay automation requires absorbance (optical density, OD) measurement, and a wide variety of spectrophotometric and filter colorimeters are commercially

**Table 2** Enzymatic parameters for alkaline phosphatase and horseradish peroxidase

	<i>Alkaline phosphatase (calf intestinal)</i>	<i>Horseradish peroxidase</i>
Molecular weight	140 000	40 000
Reaction	Cleavage of phosphate, release of monophosphate	Oxidation by $\text{H}_2\text{O}_2$ of substrate
Optimum pH	9.8	7.0
Reaction temperature	$37^\circ\text{C}$	$25^\circ\text{C}$
Cofactor requirements	$\text{Zn}^{2+}$ , $\text{Mg}^{2+}$	None
Isoelectric point	5.7	7.2
Inhibitors	EDTA, other chelators, pH <6	Cyanide, sulfide, azide

**Table 3** Substrates for alkaline phosphatase and horseradish peroxidase

Enzyme	Substrate	Product	Readout
Alkaline phosphatase	<i>p</i> -nitrophenyl phosphate (PNPP)	Soluble, yellow	
	Nitroblue tetrazolium (NBT) in combination with 5-bromo-4-chloro-3-indolylphosphate (BCIP)	Insoluble, purple	
	4-Methylumbelliferyl phosphate (4-MUP)	Fluorescent	358 nm excitation, 450 nm emission
	Fluorescein diphosphate	Fluorescent	490 nm excitation, 520 nm emission
	2'-[2-Benzothiazoyl]-6'-hydroxybenzothiazole phosphate [BbTP] (AttoPhos)	Fluorescent	435 nm excitation, 555 nm emission
	1,2-Dioxetanes (LumiPhos, CSPD)	chemiluminescent	
Horseradish peroxidase	3-Amino-9-ethylcarbazole (AEC)	Insoluble, red	
	3,3'-Diaminobenzidine (DAB)	Insoluble, brown	
	4-Chloro-1-naphthol	Insoluble, blue-black	
	3,3',5,5'-Tetramethylbenzidine (TMB)	Soluble, yellow	OD at 450 nm
	2,2'-Azinobis(3-ethylbenzthiazoline-6-sulfonic acid) (ABTS)	Soluble, green	OD at 410 nm and 650 nm
	<i>o</i> -phenylenediamine (OPD)	Soluble, brown	OD at 492 nm
	10-Acetyl-3-dihydroxyphenoxazine (Amplex Red)	Resorufin	OD at 563 nm; or 530 nm excitation, 590 nm emission
	Luminol, enhanced luminol	Chemiluminescent	
	Lumigen PS substrates	Chemiluminescent	

available for this purpose. Despite the development of improved light emitting substrates, continued interest in colorimetric reagents is due to two factors. First, their cost is generally lower than the cost of more sensitive substrates. In particular, the cost of instrumentation to read OD is low relative to fluorescence and chemiluminescence detectors. Second, adequate sensitivity is achievable by adjusting either incubation time of the enzymatic visualization reaction or the DNA amplification step prior to hybridization analysis. Because of its higher turnover number, peroxidase is the preferred label for colorimetric detection. Phosphatase assays will require longer incubation times, up to 1 h, whereas peroxidase assays are complete within 10 min. Although signal can be accumulated over time to enhance weak signals, background also increases.

Fluorescent substrates are nonfluorescent until a quenching group is cleaved by the reporter enzyme, or enzyme action results in a shift in fluorescence emission. Generally, fluorescence detection can be more sensitive than OD methods. Whether a theoretical increase in sensitivity is realized depends on the choice of enzyme and substrate as well as many other assay parameters.

Chemiluminescent substrates offer maximum sensitivity for reporter enzyme detection with a dynamic range of five orders of magnitude. Historically, their use has been limited by poor stability to heat and light. Developments over the past ~20 years,

however, have yielded stable chemiluminescent substrates for alkaline phosphatase and beta-galactosidase. With these materials, less than  $10^{-19}$  mol of alkaline phosphatase can be measured. Parallel development of peroxidase substrates has resulted in stable luminophores with sustained light yields to take advantage of the more rapid kinetics of the peroxidase reaction. The most developed of these reagents enables detection of peroxidase with similar detection limits and the same instrumentation as for the alkaline phosphatase substrates. With glow-type reagents that produce sustained light output over many minutes, signal accumulation is possible by instrumental integration. Chemiluminescence readers are widely available in single-tube and multi-well formats. Instrumentation can be programmed to read in either steady-state or kinetic mode.

Further enhancement in detection sensitivity of reporter enzymes is achievable by enzyme amplification or cascade reactions often termed 'enzyme cycling' assays. One approach is to use alkaline phosphatase as the reporter enzyme. Phosphatase cleavage of NADP forms NAD, which enters cyclic reactions catalyzed by alcohol dehydrogenase and diaphorase. Each turnover of phosphatase substrate initiates a cascade resulting in numerous detectable product molecules. Such approaches, perhaps incorporating chemiluminophores as the terminal product, hold promise for further extension of the sensitivity of immunoenzymatic methods.

## Blocking Solutions

Complete coating of the solid phase will eclipse sites on the solid phase where antibody conjugates or reporter enzymes might bind. In addition, a 'blocking solution' containing proteins and nonionic detergents is typically incubated with the solid phase prior to addition of analyte.

Early blocking reagents tended to be those that were familiar from Southern blotting: salts, polymers (dextran sulfate, polyvinylpyrrolidone), and non-homologous DNA. Later, protein-based blocking reagents such as bovine serum albumin or gelatin came into use. Low-lipid casein (Hammarsten grade) proved advantageous for use with alkaline phosphatase. Gelatin from coldwater fish provides substantial improvement over previous reagents.

Detergents are components of many blocking reagents. Classical Southern blot methodology using sodium dodecyl sulfate (SDS) is typically modified by substitution of more enzyme-compatible nonionic detergents such as polyoxyethylene (20) sorbitan monolaurate (Tween 20) or octylphenolpoly(ethyleneglycolether) (Triton X-100). SDS at low concentrations may preclude enzyme activity but may not irreversibly inactivate enzymes. Therefore, it and other strong detergents may be used during hybridization and subsequent detection steps, provided concentrations are reduced by washing to allow activity by the detection reagents. Because SDS denatures proteins, the concentrations at which it can be used depend on temperature. Generally speaking,

higher concentrations can be tolerated if the assay temperature is 42°C than if it is 60°C. Given that both alkaline phosphatase and horseradish peroxidase for immunoassay are purified from natural sources, the actual concentration of SDS that can be used may also depend on stability of the particular enzyme preparation.

Several important considerations in the design of a blocking reagent are listed in **Table 4**. It is not possible to list acceptable ranges for each of these parameters because they depend on the reagents, the solid phase, and assay specifications.

Most of the background signal in immunoassay of amplified DNA comes from nonspecific binding of enzymatic detection reagents to the solid phase. Therefore, attention should be paid to 'blocking reagent' design to block nonspecific binding by protein components. The nonionic detergent Tween 20 has been reported to block both nonspecific binding sites on the solid phase and nonspecific interactions between assay reagents.

## Wash Solutions

After immunocapture and immunodetection, the solid phase must be washed to remove weakly or nonspecifically adsorbed reagents. A pH and salt concentration should be chosen to maintain specific interactions between assay reagents and analyte. Detergents, preferably enzyme-compatible nonionic detergents, are included to facilitate desorption of loosely bound reagents. Similar considerations

**Table 4** Parameters for blocking, hybridization, and immunobinding reagent design

pH		Avoid inactivation of reporter enzyme Near isoelectric point, proteins may tend to bind to hydrophobic surfaces The pH must support hybridization and/or immunochemical reactions to be performed in the blocking solution
Detergents	0.1–2.0%, Triton X-100, Tween 20	Avoid inactivation of enzyme reagents Nonionic detergents are preferable Low concentrations of anionic detergents may be used with adequate washing
Protein blockers	Bovine serum albumin, gelatin, fish skin gelatin	For maximum sensitivity, must be free of proteases and nucleases Fish skin gelatin is preferred
Nucleic acid blockers	Herring sperm DNA, salmon sperm DNA	Not required
Polymers	Ficoll, dextran sulfate, polyvinylpyrrolidone, poly(ethylene glycol)	Not required
Salts	0.1–0.2 mol l <sup>-1</sup> NaCl, saline sodium citrate (SSC), saline sodium phosphate–EDTA (SSPE)	Salt concentration must support immunocapture or immunodetection reactions
Chelators	Ethylenediaminetetraacetic acid (EDTA)	Compatible with peroxidase not compatible with alkaline phosphatase

pertain to the ‘wash solution’ as for the ‘blocking solution’.

While the ideal wash regimen for hybridization with oligonucleotide probes on the solid phase is undoubtedly different from that for detection with enzyme conjugates, in practice, the same ‘wash solution’ is used. Most commonly, hybridization and capture are performed simultaneously or in series with no intervening wash step. Finally, a single ‘wash solution’ is used to wash the solid phase at least thrice before addition of substrate.

Assay Configurations

Direct Assays for Unamplified DNA

Assays for low-copy targets must amplify before the immunochemical step or must configure the assay in such a way that antigen concentration is close to the  $K_d$  for antibody binding. Some amplification of antibody-binding sites can be achieved by the use of long probes incorporating multiple haptens. Hybrid capture assays are commercially available for hepatitis B virus, human cytomegalovirus, and human papillomavirus. All of these pathogens express DNA at a high level during active infection. Viral DNA is captured onto a solid phase by immobilized RNA probes. RNA–DNA hybrids are detected using alkaline phosphatase-conjugated monoclonal antibodies to RNA–DNA duplex. Through the use of long RNA probes, multiple sites are generated for enzyme conjugate binding. Despite their success in specific cases, such approaches remain problematic for low-copy targets.

Immunoassays are ideal for measuring the extent of DNA damage. Many mutagens chemically react with DNA to form adducts. Oxidative damage to DNA produces several modified bases including 8-oxoguanine. Adducts, modified bases, and abasic sites resulting from excision repair of damaged bases can be assayed using typical immunoassay formats with antibodies against the modified base and a fluorescent or chemiluminescent label. Immunohistochemical approaches have been more common, but competitive and noncompetitive ELISA assays for modified bases have been described. Recently described methods have employed avidin to detect sites of DNA damage that have been tagged with biotin.

Hybridization Analysis of Amplified DNA

While immunoenzymatic detection can increase sensitivity of amplified DNA detection (e.g., by detection of label incorporated during amplification), virtually all applications of immunoreagents in PCR have utilized probe hybridization for increased

specificity. Double-stranded PCR products must be denatured by heat, organic solvents, or other methods, prior to hybridization.

Hybridization Combined with Separation

Genomic DNA is separated according to size by electrophoresis, then transferred to a filter membrane where it is immobilized. The resulting Southern blot (or Northern blot, in the case of RNA) can be hybridized to a probe, then stained to visualize the particular DNA fragments that hybridize to the probe. The combination of hybridization with a size criterion for identification increases specificity but reduces sensitivity.

Hybridization without Separation

**Dot blots** The popularity of size-fractionated DNA immobilized on filters led to the application of similar detection methods for DNA simply spotted on filters without prior electrophoresis. These methods are generally referred to as dot blots or spot blots. In general, identical detection methods may be employed as for Southern or Northern blots. Without size separation, particular attention must be paid to hybridization stringency so that high backgrounds do not confound data interpretation. Dot blots are best suited to high-throughput screening assays. Reverse dot blots involve immobilization of probe rather than analyte.

**Microtiter plate** Configuration for parallel assay is an important consideration if high-throughput analysis is contemplated. While filter plates and manifolds that can be disassembled and cleaned after each dot blot experiment are widely used, most ‘homebrew’ DNA immunoassays have used commercially available plates for ELISA. Microtiter (96-well) plates allow assays to be conducted with multichannel pipettors and automated robotic workstations. Higher-order multiplexing with 384-well or 1024-well plates is possible. A typical assay protocol is provided in Table 5.

**DNA microarrays** DNA microarrays provide a means for parallel analysis of thousands of sequences in a single experiment. The original DNA arrays

Table 5 General procedure for PCR immunocapture assay

Denature the nucleic acid analyte
Add analyte to hybridization solution in coated, blocked microwell
Incubate for 20 min at 42°C
Discard hybridization solution
Wash with wash solution
Add substrate and detect as appropriate to the substrate

used immobilized oligonucleotides tethered to silica chips. Analyte RNA was tagged with fluorescein, hybridized to the chip, and detected by fluorescence. Subsequent developments have led to a wide variety of methods for preparing the microarrays as well as performing the labeling and hybridization. Immunochemical techniques involving reporter enzymes are not well suited for these methods due to the short distances between various probes on the chip surface. High-sensitivity detection is currently achieved by incorporation of biotin into the analyte followed by hybridization on the chip and detection with streptavidin-conjugated phycoerythrin.

***In situ* hybridization and *in situ* PCR** *In situ* hybridization provides information on spatial localization, within a cell or tissue, of a particular nucleic acid, usually a messenger RNA (mRNA) or a pathogen's DNA or RNA. Similar considerations apply to assay formats and labels as apply for the analysis of purified nucleic acids. The main difference is that the specimen must first be fixed prior to immunochemical and hybridization analysis. Insoluble substrates are used to prevent diffusion of signal from one point to another. Methods have been devised for PCR amplification in fixed cells (*in situ* PCR). Immunochemical detection by fluorescence (fluorescence *in situ* hybridization) using digoxigenin and fluorescent anti-digoxigenin antibodies has been reported, as have techniques incorporating antibodies conjugated to colloidal gold (ImmunoGold) with or without enhancement by silver deposition.

## Introducing Labels

Unless hybridization to two probes is contemplated (sandwich hybridization), nucleic acids to be analyzed must be labeled for capture or detection.

### Labeling the Analyte

When the analyte is amplified prior to analysis, labeling is greatly simplified. Several methods for incorporating detectable labels have been described. Labeled dNTPs allow more labels per analyte, but in practice the increase in sensitivity is offset by cost and background. A labeled primer allows efficient, uniform labeling during amplification. The label can be used for detection or for capture.

**Modified dNTP approach** Either capture labels or detection labels can be incorporated during amplification or primer extension. The amount of signal enhancement, however, is limited by steric hindrance. Labeled nucleotides must be stable to the conditions employed in amplification. To minimize steric

hindrance during polymerase incorporation, long spacer arms generally are recommended. Even so, no available modified dNTP is incorporated as efficiently as the unmodified dNTP. A variety of labeled dNTPs are available, including biotin, digoxigenin, and various fluorophores. In principle, antibodies could be raised against any of these labels and employed as either capture or detection moieties.

The cardinal disadvantage of the labeled dNTP approach is a nonspecific signal that arises from incorporation of the label into primer oligomer and other nontarget-derived amplification products. A second disadvantage is the excess label that must be somehow washed off the solid phase to avoid increased assay background. A third disadvantage is expense.

**Labeled primers** Oligonucleotide primers are easily labeled on the 5'-phosphate group with biotin, digoxigenin, dinitrophenol, other haptens, fluorescein, other fluorophores, and many other reagents. These labels can be used for either capture or detection. To increase specificity, the primer label is generally used for capture, rather than for detection. Thus, hybridization to a labeled oligonucleotide probe can be performed in solution where kinetics are more rapid, and followed by capture on the solid phase.

For example, a solid phase coated with avidin or streptavidin can be used to capture hybrids containing a biotinylated primer. Subsequent detection reactions involve the label that is attached to the probe (Figure 1). Solid phases also can be coated with monoclonal or polyclonal antibodies. Cost and availability are considerations that have prompted many researchers to use avidin.

Labeled primers can be used for both capture and detection. The sense-strand primer can be labeled with a capture reagent such as biotin, and the anti-sense-strand primer can be labeled with a detectable reagent such as digoxigenin. This combination is capable of high sensitivity but suffers from background due to primer oligomer and nonspecific products unless PCR is well designed and optimized.

**Labeling unamplified nucleic acids** Chemical methods to label unamplified nucleic acids include photoactivatable biotin reagents and sulfonation of cytosines. Enzymatic labeling using random primer extension is also possible. Extension of oligodeoxythymidine primers hybridized to the 3'-polyadenylate sequence in mRNA can be used to generate labeled antisense RNA for analysis.

### Labeling the Probe

**DNA probes** Long DNA probes can be generated using recombinant DNA techniques as inserts in

plasmids. Linearization of plasmid DNA yields a DNA probe of several hundred to several thousand base pairs in length. A standard method of random priming or nick translation is used to introduce labels into this probe. DNA probes can be labeled with photoactivatable biotin or by any methods used to label analyte DNA.

Shorter DNA probes (less than ~50 bases) are usually synthesized. A wide variety of labels are available as phosphoramidites for automated incorporation at 5', 3', or internal positions in the oligonucleotide. Alternately, amino or thiol groups can be introduced during synthesis. These reactive groups are available for linking to essentially any fluorophore or enzyme using crosslinking reagents.

**RNA probes** RNA probes bind tighter to their complementary strands than do DNA probes. Poor stability due to ubiquitous ribonucleases has hampered more widespread use of short RNA probes, as has the difficulty of efficient chemical synthesis of long RNA oligomers. Recent advances in RNA synthetic chemistry have solved the latter problem.

Long RNA probes are generated by *in vitro* transcription from linearized plasmid DNA containing a promoter sequence for a DNA-dependent RNA polymerase such as SP3, T3, or T7 polymerases. Commercially available kits for high-yield transcription with label incorporation are available. It is also technically trivial to introduce a T7 promoter sequence as a 5'-extension of a PCR primer. Amplification introduces the promoter sequence into the amplicon. Transcription of the amplicon with T7 RNA polymerase and a modified NTP yields the labeled RNA probe. Shorter RNA probes (less than ~50 bases) are chemically synthesized, and label is most conveniently introduced during synthesis.

**Labeling immunoenzyme conjugates** Extensive publications describe methods to covalently link avidin or immunoglobulins to reporter enzymes. A wide range of bifunctional linkers are available commercially. Prepared, stabilized conjugates are available from immunological reagent suppliers.

## Limitations and Future Developments

The most significant limitation of immunochemical methods is inadequate sensitivity such that preamplification by PCR is required. Further improvements in solid phases, reporter enzymes, substrates, and hybridization probes will extend the range of analytes amenable to assay without amplification. Solid phases designed specifically for DNA immunoassays are needed. Enzyme engineering for thermostability,

lower background binding, and higher activity is possible with current technology. These advances could yield a new generation of reporter labels.

Triplex, quadruplex, and nucleic acid analog probes hold promise for assays of improved sensitivity and specificity. To date, limited use has been made of nucleic acid analogs as probes even though the advantages of some structural types for antisense therapeutics may extend to diagnostic applications. The use of such probes will increase specificity and sensitivity of probe hybridization assays.

Ultimately, immunochemical reagents of relatively low cost hold the key to rapid tests for DNA analysis without amplification.

*See also:* **Chemiluminescence:** Overview. **Derivatization of Analytes.** **Electrophoresis:** Overview. **Enzymes:** Overview; Immobilized Enzymes; Enzyme-Based Electrodes; Enzymes in Physiological Samples; Industrial Products and Processes; Enzyme-Based Assays. **Fluorescence:** Clinical and Drug Applications. **Immunoassays:** Overview; Production of Antibodies. **Immunoassays, Applications:** Clinical; Food; Forensic. **Immunoassays, Techniques:** Radioimmunoassays; Enzyme Immunoassays; Luminescence Immunoassays. **Mass Spectrometry:** Polymerase Chain Reaction Products. **Nucleic Acids:** Chromatographic and Electrophoretic Methods; Electrochemical Methods. **Polymerase Chain Reaction.**

## Further Reading

- Akhavan-Tafti H, De Silva R, Thakur K, Eickholt RA, Chrovian C, Lauwers K, Handley RS, Siripurapu S, Endreszl C, and Schaap AP (2000) Recent advances in chemiluminescent enzyme substrates. In: Case JE, Herring PJ, Robison BH, Haddock SH, Kricka LJ, and Stanley PE (eds.) *Bioluminescence and Chemiluminescence*, pp. 215–218. Chichester: Wiley.
- Anderson MLM (ed.) (1999) *Nucleic Acid Hybridization*. New York: Springer.
- Andrus A (1995) Chemical methods for 5'; non-isotopic labeling of PCR probes and primers. In: McPherson MJ, Hames BD, and Taylor GR (eds.) *PCR 2: A Practical Approach*, pp. 39–54. Oxford: Oxford University Press.
- Butler JE (1996) Solid phases in immunoassay. In: Diamandis EP and Christopoulos TK (eds.) *Immunoassay*, pp. 205–225. San Diego: Academic Press.
- Chehab FF, Wall J, and Cai S (1995) Analysis of PCR products by covalent reverse dot blot hybridization. In: Innis MA, Gelfand DH, and Sninsky JJ (eds.) *PCR Strategies*, pp. 130–139. San Diego: Academic Press.
- Englebienne P (2000) Analytical signal: Generation and detection. *Englebienne P Immune and Receptor Assays in Theory and Practice*, pp. 227–289. Boca Raton, FL: CRC Press.
- Gold B (2003) Origin and utility of the reverse dot-blot. *Expert Review of Molecular Diagnostics* 3(2): 3–52.



- Helmuth R (1990) Nonisotopic detection of PCR products. In: Innis MA, Gelfand DH, Sninsky JJ, and White T (eds.) *PCR Protocols: A Guide to Methods and Applications*, pp. 119–128. San Diego: Academic Press.
- Hermanson GT (1996) *Bioconjugate Techniques*. San Diego: Academic Press.
- Holodniy M, Winters MA, and Merigan TC (1992) Detection and quantification of gene amplification products by a nonisotopic automated system. *Biotechniques* 12(1): 36–39.
- Kessler C (1995) Methods for nonradioactive labeling of nucleic acids. In: Kricka LJ (ed.) *Nonisotopic Probing, Blotting, and Sequencing*, 2nd edn., pp. 42–109. San Diego: Academic Press.
- Knudsen S (2002) *A Biologist's Guide to Analysis of DNA Microarray Data*, pp. 1–13. New York: Wiley-Interscience.
- Kow YW and Dare A (2000) Detection of abasic sites and oxidative DNA base damage using an ELISA-like assay. *Methods* 22(2): 164–169.
- Martin CS, Voyta JC, and Bronstein I (2000) Quantitative polymerase chain reaction and solid phase capture nucleic acid detection. In: Ziegler MM and Baldwin TO (eds.) *Bioluminescence and Chemiluminescence, Part C. Methods in Enzymology*, pp. 466–476. San Diego: Academic Press.
- Nuovo GJ (1995) PCR *in situ* hybridization. In: Wiedbrauk DL and Farkas DH (eds.) *Molecular Methods for Virus Detection*, pp. 237–259. San Diego: Academic Press.
- Santella RM (1999) Immunological methods for detection of carcinogen-DNA damage in humans. *Cancer Epidemiology Biomarkers & Prevention* 8: 733–739.
- Vary CPH (1996) Nonisotopic DNA labeling strategies. In: Dangler CA (ed.) *Nucleic Acid Analysis: Principles and Bioapplications*, pp. 47–65. New York: Wiley-Liss.
- Vlieger AM, Medenblik AM, van Gijlswijk RP, Tanke HJ, van der Ploeg M, Gratama JW, and Raap AK (1992) Quantitation of polymerase chain reaction products by hybridization-based assays with fluorescent, colorimetric, or chemiluminescent detection. *Analytical Biochemistry* 205(1): 1–7.
- Wilchek M and Bayer EA (1990) Introduction to avidin-biotin technology. In: Wilchek M and Bayer EA (eds.) *Avidin-Biotin Technology, Methods in Enzymology*, vol. 184, pp. 5–13. San Diego: Academic Press.

## mRNA Identification and Quantification

**M Pfaffl**, Technische Universitaet Muenchen, Freising-Weihenstephan, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Molecular technologies are currently evolving rapidly the biological sciences. This results in an immense progress in accumulation of new data potentially useful for molecular diagnostics. One of these highly sophisticated methodologies is the quantitative assessment of target nucleic acids, mostly performed as quantitative polymerase chain reaction (PCR) on DNA level or combined with reverse transcription PCR (RT-PCR) to investigate the transcriptome on RNA level. To quantify local tissue-specific expression even in tissues with low abundances, very sensitive methods are required, which allow reliable mRNA quantification. The aim of a full quantitative method is to estimate, as exact and reliable as possible, the number of target molecules in the sample. For this exact quantification of low abundant gene expression only a few PCR methods allow reliable mRNA quantification.

(A) Internally standardized competitive RT-PCR measured by high-performance liquid chromatography separation and ultraviolet detection or high-resolution

gel electrophoresis followed by densitometric analysis: in a competitive RT-PCR, a reference RNA mutant is reverse transcribed and coamplified in the same reaction tube with the native mRNA sequence of interest. Internally standardized RT-PCR is a very time consuming and laborious technique. It is generally believed to yield the most precise results, because all parameters throughout RT-PCR act on both the analyte and reference mutant.

(B) Externally standardized RT-PCR with online-detection using SYBR Green I technology: real-time RT-PCR with SYBR Green I detection produces sensitive and reliable results. Due to the use of an external standard curve, the amplification efficiencies for the calibration curve and the analyte must be equal for accurate quantification.

(C) Externally standardized RT-PCR with online-detection using specific hybridization probes: this detection format is based on various fluorescence detection formats, e.g., fluorescence resonance energy transfer (FRET).

(D) Array technologies in combination with real-time RT-PCR: microarray technology is a powerful technique in order to analyze the gene expression of thousands of genes (up to 20 000) in a short time. Problems encountered include inconsistent fidelity, high variability, sensitivity not sufficient for low abundant expressed genes, discrepancy in fold-changes calculation, and lack of specificity for different isoforms

- Helmuth R (1990) Nonisotopic detection of PCR products. In: Innis MA, Gelfand DH, Sninsky JJ, and White T (eds.) *PCR Protocols: A Guide to Methods and Applications*, pp. 119–128. San Diego: Academic Press.
- Hermanson GT (1996) *Bioconjugate Techniques*. San Diego: Academic Press.
- Holodniy M, Winters MA, and Merigan TC (1992) Detection and quantification of gene amplification products by a nonisotopic automated system. *Biotechniques* 12(1): 36–39.
- Kessler C (1995) Methods for nonradioactive labeling of nucleic acids. In: Kricka LJ (ed.) *Nonisotopic Probing, Blotting, and Sequencing*, 2nd edn., pp. 42–109. San Diego: Academic Press.
- Knudsen S (2002) *A Biologist's Guide to Analysis of DNA Microarray Data*, pp. 1–13. New York: Wiley-Interscience.
- Kow YW and Dare A (2000) Detection of abasic sites and oxidative DNA base damage using an ELISA-like assay. *Methods* 22(2): 164–169.
- Martin CS, Voyta JC, and Bronstein I (2000) Quantitative polymerase chain reaction and solid phase capture nucleic acid detection. In: Ziegler MM and Baldwin TO (eds.) *Bioluminescence and Chemiluminescence, Part C. Methods in Enzymology*, pp. 466–476. San Diego: Academic Press.
- Nuovo GJ (1995) PCR *in situ* hybridization. In: Wiedbrauk DL and Farkas DH (eds.) *Molecular Methods for Virus Detection*, pp. 237–259. San Diego: Academic Press.
- Santella RM (1999) Immunological methods for detection of carcinogen-DNA damage in humans. *Cancer Epidemiology Biomarkers & Prevention* 8: 733–739.
- Vary CPH (1996) Nonisotopic DNA labeling strategies. In: Dangler CA (ed.) *Nucleic Acid Analysis: Principles and Bioapplications*, pp. 47–65. New York: Wiley-Liss.
- Vlieger AM, Medenblik AM, van Gijlswijk RP, Tanke HJ, van der Ploeg M, Gratama JW, and Raap AK (1992) Quantitation of polymerase chain reaction products by hybridization-based assays with fluorescent, colorimetric, or chemiluminescent detection. *Analytical Biochemistry* 205(1): 1–7.
- Wilchek M and Bayer EA (1990) Introduction to avidin-biotin technology. In: Wilchek M and Bayer EA (eds.) *Avidin-Biotin Technology, Methods in Enzymology*, vol. 184, pp. 5–13. San Diego: Academic Press.

## mRNA Identification and Quantification

**M Pfaffl**, Technische Universitaet Muenchen, Freising-Weihenstephan, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Molecular technologies are currently evolving rapidly the biological sciences. This results in an immense progress in accumulation of new data potentially useful for molecular diagnostics. One of these highly sophisticated methodologies is the quantitative assessment of target nucleic acids, mostly performed as quantitative polymerase chain reaction (PCR) on DNA level or combined with reverse transcription PCR (RT-PCR) to investigate the transcriptome on RNA level. To quantify local tissue-specific expression even in tissues with low abundances, very sensitive methods are required, which allow reliable mRNA quantification. The aim of a full quantitative method is to estimate, as exact and reliable as possible, the number of target molecules in the sample. For this exact quantification of low abundant gene expression only a few PCR methods allow reliable mRNA quantification.

(A) Internally standardized competitive RT-PCR measured by high-performance liquid chromatography separation and ultraviolet detection or high-resolution

gel electrophoresis followed by densitometric analysis: in a competitive RT-PCR, a reference RNA mutant is reverse transcribed and coamplified in the same reaction tube with the native mRNA sequence of interest. Internally standardized RT-PCR is a very time consuming and laborious technique. It is generally believed to yield the most precise results, because all parameters throughout RT-PCR act on both the analyte and reference mutant.

(B) Externally standardized RT-PCR with online-detection using SYBR Green I technology: real-time RT-PCR with SYBR Green I detection produces sensitive and reliable results. Due to the use of an external standard curve, the amplification efficiencies for the calibration curve and the analyte must be equal for accurate quantification.

(C) Externally standardized RT-PCR with online-detection using specific hybridization probes: this detection format is based on various fluorescence detection formats, e.g., fluorescence resonance energy transfer (FRET).

(D) Array technologies in combination with real-time RT-PCR: microarray technology is a powerful technique in order to analyze the gene expression of thousands of genes (up to 20 000) in a short time. Problems encountered include inconsistent fidelity, high variability, sensitivity not sufficient for low abundant expressed genes, discrepancy in fold-changes calculation, and lack of specificity for different isoforms

or differentially expressed genes. The DNA microarray based screening of tissue-specific gene expression and confirmation of putative candidate target genes by kinetic RT-PCR represents a powerful and optimal combination. Hereby, the advantages of both quantification systems can be added – the high throughput capacity of the microarray as well as sensitivity and specificity of the real-time RT-PCR.

Because of its high sensitivity, RT-PCR is being increasingly applied to quantify physiologically relevant changes in gene expression. RT-PCR has a detection limit 10- to 100-fold better than other methods, e.g., RNA-protection-assay or Northern-hybridization. It offers a new dimension in the detection of rare RNA by amplifying a single-stranded cDNA after reverse transcription. Today real-time RT-PCR (qRT-PCR or kinetic RT-PCR) is increasingly used because of its high sensitivity, good reproducibility, and wide dynamic quantification range. The first practical kinetic PCR technology, the 5'-nuclease assay, was established in 1993 and combines the exponential PCR amplification of a specific transcript with the monitoring of newly synthesized DNA in each performed PCR cycle. It is the most sensitive method for the detection and quantification of gene expression levels, in particular for low abundant transcripts in tissues with low RNA concentrations, from limited tissue samples, and for the elucidation of small changes in mRNA expression levels. While kinetic RT-PCR has a tremendous potential for analytical and quantitative applications, a comprehensive understanding of its underlying

principles is important. Fidelity of real-time RT-PCR is associated with its true specificity, sensitivity, reproducibility, robustness, and as a fully reliable quantitative method. It suffers from the problems inherent in RT and PCR. This review analyzes the possibilities and problems in mRNA quantification analytics.

RNA Isolation and cDNA Synthesis

The integrity of purified RNA is critical to all gene expression analysis techniques. For successful and reliable diagnostic use, real-time RT-PCR needs high quality, DNA-free, and undegraded RNA. Accurate quantification and quality assessment of the starting RNA sample is particularly important for absolute quantification methods that normalize specific mRNA expression levels against total RNA ('molecules per gram total RNA' or 'transcript concentrations per gram total RNA'). The source of RNA, sampling techniques (biopsy material, single cell sampling, laser microdissection) as well as RNA isolation techniques (either total RNA or poly-adenylated RNA isolation techniques) often vary significantly between processing laboratories. In Figure 1 the transcriptome of various tissue total RNA is characterized and all subfractions are described. As shown, most genes are transcribed at very low abundant mRNA levels under 20 copies per cell. Therefore, highly sensitive quantification methods are necessary to perform a precise and reliable quantification of low abundant genes, e.g., cytokines, growth factors, hormones, their corresponding receptors, and enzymes.

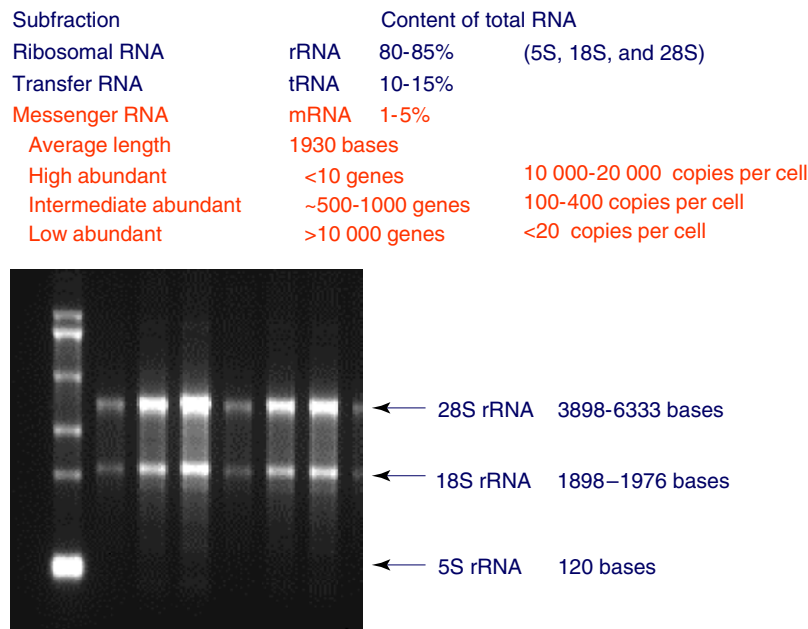


Figure 1 Characterization of the transcriptome isolated from numerous tissue samples.

Isolated RNA may contain tissue enzyme inhibitors and genomic DNA that result in reduced RT and PCR reaction efficiencies and generate unreliable and wrong quantification results. While this is not a problem for some applications, the tremendous amplification power of kinetic PCR may result in even the smallest amount of DNA contamination interfering with the desired specific amplification. To confirm the absence of residual DNA, a minus-RT should be included. Additionally it may be necessary to treat the RNA sample with commercially available RNase-free DNase, to get rid of residual DNA. Furthermore, the design of the PCR product should incorporate at least one exon to exon splice junction to allow a product obtained from the cDNA to be distinguished from genomic DNA contamination.

A further important step in quantitative RT-PCR is the production of a single-stranded (ss) complementary DNA copy (cDNA) of the RNA through the reverse transcriptase (RT). Its dynamic range, sensitivity, and specificity are of prime consideration for a successful kinetic RT-PCR assay. For many quantitative applications, moloney murine leukaemia virus (MMLV) H<sup>-</sup> RT is the enzyme of choice, as its cDNA synthesis rate is up to 50-fold greater than that of avian myeloblastosis virus (AMV). Newly available thermostable enzymes maintain their activity up to 70°C, thus permitting increased specificity and efficiency of first primer annealing. However, this enzyme type may be less robust than more conventional ones as it appears to be more sensitive to inhibitors present in RNA preparation.

The RT step is the source of most of the variability in a kinetic RT-PCR experiment and for each enzyme, the specific reaction conditions has to be optimized. It cannot be assumed that different reactions have the same cDNA synthesis efficiency. Therefore, the result can be high variability during multiple RT reactions. To circumvent this high interassay variations in RT, target gene unspecific primers, e.g., random-hexamer, -octamer, or -decamer primers, can be used and a cDNA pool can be synthesized. Similarly, poly-T oligo-nucleotides can anneal to the polyadenylated 3'-(poly-A) tail found on most mRNA. Therefore, the gene expression results are directly comparable between the applied assays, at least within one RT pool. In conclusion, a rank order of RT efficiency can be shown for the applied different primers for one specific gene: random hexamer primers > poly-dt primer > gene specific primer. Importantly, not only RNA quantity and quality, but also yield and quality of cDNA can be highly variable. Certainly, there is evidence that cDNA yield from sequences near the 5'-end of partially degraded mRNA is significantly less than from sequences near

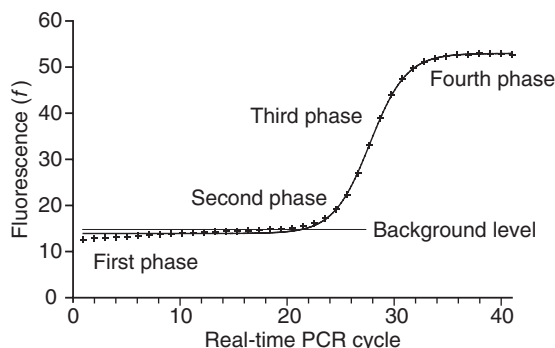
the poly-A tail, and assays aimed at identifying RNA degradation are being developed. Thus, reliable internal quality control of cDNA synthesis is essential. Controls are generally performed by PCR amplification of reference genes, mostly common housekeeping genes (GAPDH, albumin, actins, tubulins, cyclophilin, microglobulins, 18S or 28S rRNA). The reference genes used as well as the expression levels vary between different laboratories, and only few of them have been critically evaluated.

## Real-Time RT-PCR Reaction Kinetics

The efficacy of kinetic RT-PCR is measured by its specificity, low background fluorescence, steep fluorescence increase and high amplification efficiency, and high level plateau. Typically the PCR can be divided into four characteristic phases (Figure 2):

- first phase is hidden under the background fluorescence, where an exponential amplification is expected;
- second phase with exponential amplification, that can be detected above the background;
- third phase with linear amplification efficiency and a steep increase of fluorescence;
- fourth phase or plateau phase, defined as the attenuation in the rate of exponential product accumulation, which is seen concomitantly in later cycles;

The amount of amplified target is directly proportional to the input amount of target, only during the exponential phase of PCR amplification. Hence, the key factor in the quantitative ability of kinetic RT-PCR is that it measures the product of the target gene within that phase. Since data acquisition and analysis are performed in the same tube, this increases sample throughput, reduces carryover contamination, and removes post-PCR processing as a potential source of error. In contrast, during the plateau phase of the



**Figure 2** The four characteristic phases of PCR, evaluated by real-time PCR fluorescence acquisition.

PCR there is no direct relation of DNA input to amplified target; hence classical RT-PCR assays have to be stopped at least in linear phase. The exponential range of amplification has to be determined for each transcript empirically by amplifying equivalent amounts of cDNA over various cycles of the PCR or by amplifying dilutions of cDNA over the same number of PCR cycles.

## Chemistry Developed and Real-Time RT-PCR Platforms

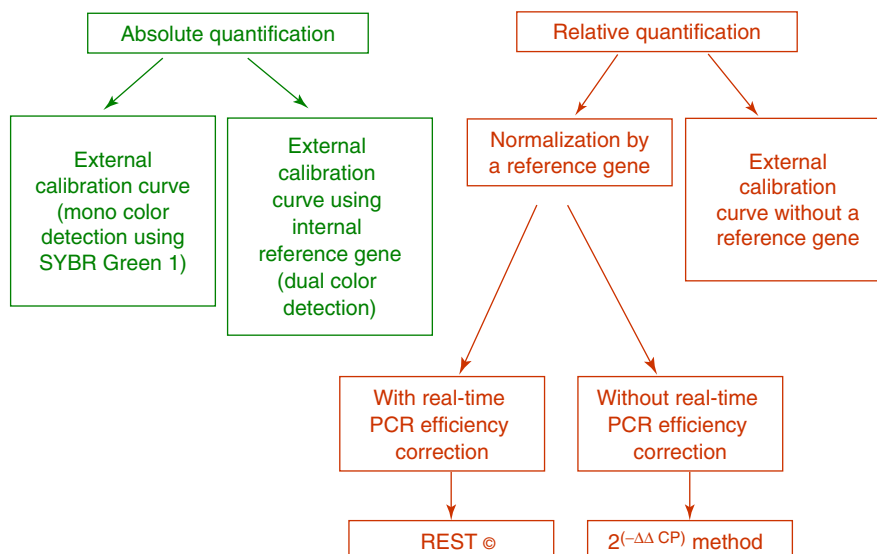
Two general methods for the quantitative detection of the amplicon became established: specific double-strand (ds) DNA binding agents or gene specific fluorescent probes, e.g., based on FRET. The best-known probe-based system is ABI TaqMan, which makes use of the 5'-exonuclease activity of Taq polymerase to quantify target sequences in the samples. Probe hydrolysis separates fluorophore and quencher and results in an increased fluorescence signal, called FRET. The alternative is a nonsequence specific fluorescent intercalating ds DNA binding dye, e.g., SYBR Green I (Molecular Probes). For single PCR product reactions with well-designed primers, SYBR Green I can work extremely well, with spurious nonspecific background only showing up in very late cycles. Among the real-time detection chemistry, SYBR Green I and TaqMan assays produced comparable dynamic range and sensitivity, while SYBR Green I detection was more precise and produced a more linear decay plot than the TaqMan probe detection.

The PCR machines differ in sample capacity, up to 96- and 384-well standard format, others process 72

samples or require 32 specialized glass capillaries, excitation method (lasers, others broad spectrum light sources with various filters), and fluorescence acquisition channels. There are also platform-specific differences in how the software processes data with focus on absolute or relative quantification strategies. For at least two systems and chemistries, the ABI PRISM 7700 using TaqMan Probes and Roche's LightCycler using Hybridisation Probes, there is little difference in accuracy and performance.

## Quantification Strategies in qRT-PCR

Two general quantification strategies can be performed in qRT-PCR: the levels of expressed genes may be measured by absolute or relative quantitative real-time RT-PCR (Figure 3). Absolute quantification relates the PCR signal to input copy number using a calibration curve, while relative quantification measures the relative change in mRNA expression levels. The reliability of an absolute real-time RT-PCR assay depends on the condition of identical amplification efficiency (see the section 'Real-Time PCR Amplification Efficiency') for both the native target and the calibration curve in RT reaction and in following kinetic PCR. Nowadays, a relative quantification is easier to perform than absolute quantification because there is no need for a calibration curve. It is based on the expression levels of a target gene versus a housekeeping gene and in theory is adequate for most purposes in animal sciences to investigate physiological changes in gene expression levels. The units used to express relative quantities are irrelevant, and the relative quantities can be compared across multiple real-time RT-PCR experiments.



**Figure 3** Quantification strategies in real-time PCR (<http://www.wzw.tum.de/gene-quantification/optimization.html>).

## Absolute Quantification

Calibration curves allow the generation of highly specific, sensitive, and reproducible data. However, the external calibration curve model has to be thoroughly validated as the accuracy of absolute quantification in real-time RT-PCR depends entirely on the accuracy of the standard material. Standard design, production, determination of the exact standard concentration, and stability over long storage time is not straightforward and can be problematic. The dynamic range of the calibration curve performed can be up to nine orders of magnitude from  $<10^1$  to  $>10^{10}$  start molecules, depending on the applied purity of the standard material. The calibration curves used in absolute quantification can be based on known concentrations of DNA standard molecules, e.g., recombinant plasmid DNA (recDNA), *in vitro* transcribed recombinant RNA (recRNA), genomic DNA, RT-PCR product, and commercially synthesized big oligo-nucleotide. Stability and reproducibility in kinetic RT-PCR depends on the type of standard used. Cloned recDNA and genomic DNA are very stable and generate highly reproducible standard curves even after a long storage time, in comparison to freshly synthesized recRNA standard material. DNA based calibration curves are subject to the PCR step only, unlike the unknown mRNA samples that must first be reverse transcribed. This increases the potential for variability of the RT-PCR results and the amplification results may not be strictly comparable with the results from the unknown samples. However, the problem of sensitivity of the RT-PCR to small variations in the reaction setup is always lurking in the background as a potential drawback to this simple procedure. Therefore, quantification with external standards requires careful optimization of its precision (replicates in the same kinetic RT-PCR run – intraassay variation) and reproducibility (replicates in separate kinetic RT-PCR runs – interassay variation) in order to understand the limitations within the given application.

A recRNA standard that was synthesized *in vitro* from a cloned RT-PCR fragment in plasmid DNA is one option. However, identical RT efficiency, as well as real-time PCR amplification efficiencies for calibration curve and target cDNA will have to be tested and confirmed, if the recDNA is to provide a valid standard for mRNA quantification. This is because only the specific recRNA molecules are present during RT and the kinetics of cDNA synthesis are not like those in native RNA (the unknown sample) that also contain a high percentage of natural occurring subfractions (Figure 1). These missing RNA

subfractions can influence the cDNA synthesis rate and in consequence RT efficiency rises and calibration curves are then overestimated in gene quantification. To compensate for background effects and mimic a natural RNA distribution, like in native total RNA, total RNA isolated from bacterial, insect cell lines, poly-A RNA or tRNA can be added.

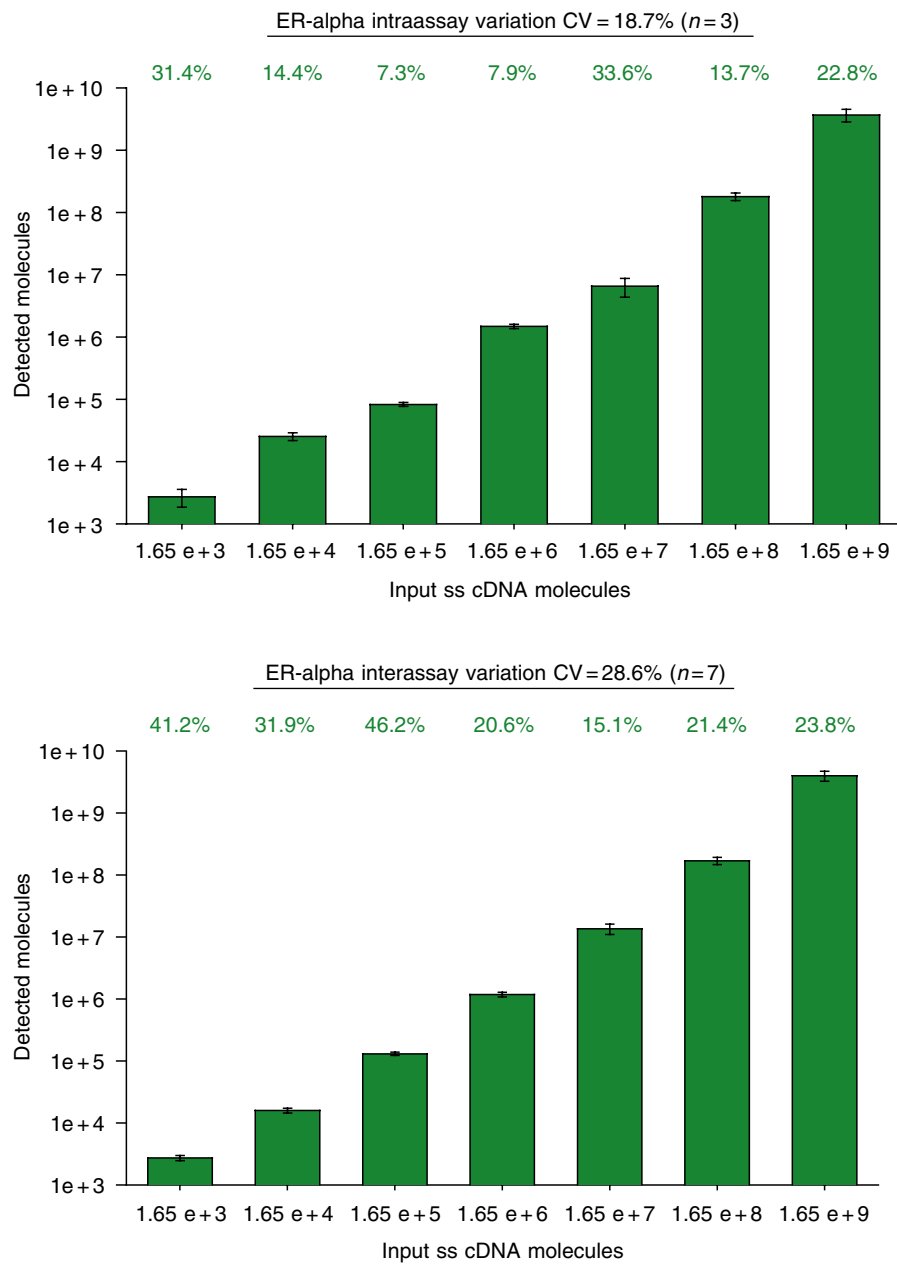
No matter how accurately the concentration of the standard material is known, the final result is always reported relatively compared to a defined unit of interest: e.g., copies per defined nanograms of total RNA, copies per genome (6.4 pg DNA), copies per cell, copies per gram of tissue, copies per milliliter of blood, etc. If absolute changes in copy number are important then the denominator still must be shown to be invariable across the comparison. The quality of gene quantification data cannot be better than the quality of the denominator itself. Any denominator variation will obscure real changes and produce artificial changes and wrong quantification results. Careful use of controls is critical to demonstrate that the choice of denominator was a wise one.

External standard quantification is the method of choice for the nucleic acid quantification, independent of any hardware platform used. The specificity, sensitivity, linearity, and reproducibility allow absolute and accurate quantification of molecules even in tissues with low mRNA abundance ( $<100$  mRNA molecules) and a detection down to a few molecules ( $<10$  mRNA molecules). The dynamic range of an optimally validated and optimized external standardized real-time RT-PCR assay can accurately detect target mRNA up to nine orders of magnitude or a billion-fold range with high assay linearity ( $r > 0.99$ ). In general a mean intraassay variation of 10–20% and a mean interassay variation of 15–30% on molecule basis (maximal 2–4% variability on crossing point (CP) basis) is realistic over the wide dynamic range, as shown for estrogen receptor (ER) kinetic RT-PCR (Figure 4). At high ( $>10^7$ ) and low ( $<10^3$ ) template copy input levels, the assay variability is higher than in the range between the two. At very low copy numbers, under 20 copies per tube, the random variation due to sampling error (Poisson's error law) becomes significant.

## Relative Quantification

Relative quantification determines the changes in steady-state mRNA levels of a gene across multiple samples and gives a result relative to the levels of an internal control RNA. This reference gene is often a housekeeping gene and can be coamplified in the same tube in a multiplex assay or can be amplified in a separate tube (Figure 3). Therefore, relative





**Figure 4** Mean intraassay ( $n = 7 \times 3$ ) and interassay variation ( $n = 7 \times 7$ ) of estrogen receptor (ER) alpha real-time RT-PCR assay, using externally standardized recDNA calibration curve.

quantification does not require standards with known concentrations and the reference can be any transcript, as long as its sequence is known. Relative quantification is based on the expression levels of a target gene versus a reference gene and in many experiments is adequate for investigating physiological changes in gene expression levels. To calculate the expression of a target gene in relation to an adequate reference gene, various mathematical models are established. Calculations are based on the comparison of the distinct cycle determined by various methods, e.g., CP or threshold values (Ct) at a constant level of

fluorescence; or CP acquisition according to established mathematic algorithm. To date several mathematical models have been developed, which calculate the relative expression ratio. Relative quantification models without efficiency correction are available and published (eqns [1] and [2]).

$$R = 2^{-(\Delta CP_{\text{sample}} - \Delta CP_{\text{control}})} \quad [1]$$

$$R = 2^{-\Delta \Delta CP} \quad [2]$$

More recently developed mathematical models with kinetic PCR efficiency correction allow for a more

precise quantification of the relative expression levels (eqns [3] and [4]).

$$\text{ratio} = \frac{(E_{\text{target}})^{\Delta\text{CP}_{\text{target}}(\text{control}-\text{sample})}}{(E_{\text{ref}})^{\Delta\text{CP}_{\text{ref}}(\text{control}-\text{sample})}} \quad [3]$$

$$\text{ratio} = \frac{(E_{\text{target}})^{\Delta\text{CP}_{\text{target}}(\text{mean control}-\text{mean sample})}}{(E_{\text{ref}})^{\Delta\text{CP}_{\text{ref}}(\text{mean control}-\text{mean sample})}} \quad [4]$$

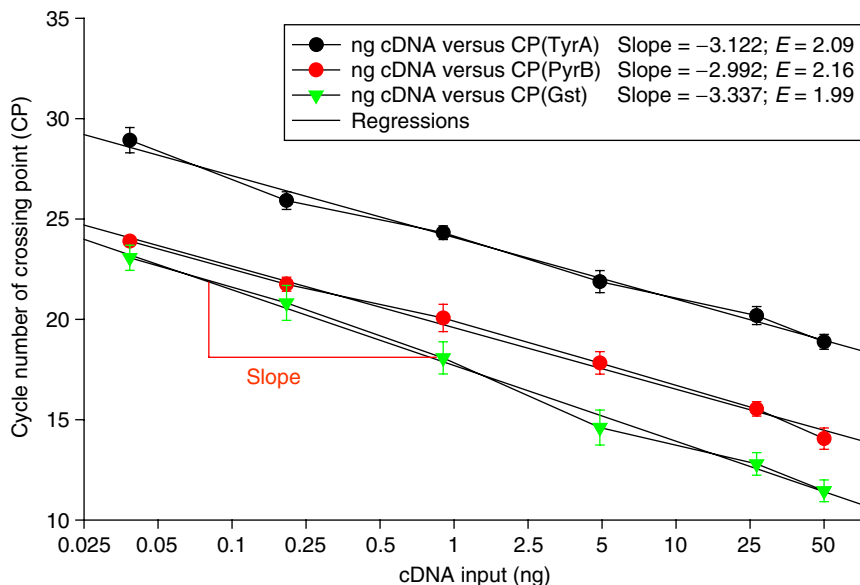
The relative expression ratio of a target gene is computed, based on its calculated real-time PCR efficiencies ( $E$ ) and the CP difference ( $\Delta\text{CP}$ ) of one unknown sample (treatment) versus one control ( $\Delta\text{CP}_{\text{control}-\text{treatment}}$ ). Using relative expression software tool (REST) the relative calculation procedure is based on the mean CP of the experimental groups (eqn [4]). In these models, the target gene expression is normalized by a nonregulated reference gene expression, e.g., derived from classical and frequently described housekeeping genes. The crucial problem in this relative approach is the most common reference gene transcripts from the so-called housekeeping genes, whose mRNA expression can be regulated and whose levels vary significantly with treatment or between individuals. However, relative quantification can generate useful and biologically relevant information when used appropriately.

### Real-Time PCR Amplification Efficiency

Individual samples generate different and individual fluorescence histories in kinetic RT-PCR. The shapes

of amplification curves differ in the steepness of any fluorescence increase and in the absolute fluorescence levels at plateau depending on background fluorescence levels. The PCR efficiency has a major impact on the fluorescence history and the accuracy of the calculated expression result and is critically influenced by PCR reaction components. A correction for efficiency, as performed in efficiency corrected mathematical models, is strongly recommended, which results in a more reliable estimation of the real expression ratio compared to no efficiency correction. Small efficiency differences between target and reference gene generate false expression ratio, resulting in over- or underestimation of the real initial mRNA amount. Therefore, efficiency corrected quantification corrections should be included in the automation and calculation procedure in relative quantification models. Several methods are described in the literature or were developed to calculate real-time PCR efficiency.

(A) Efficiency calculation from the slopes of the calibration curve, according to the equation:  $E = 10^{(-1/\text{slope})}$  (Figure 5). Efficiency vary with high linearity ( $r > 0.95$ ) from  $E = 1.60$  to maximal values up to  $E = 2.10$ . Typically, the relationship between CP and the logarithm of the starting copy number of the target sequence should remain linear for up to four orders of magnitude. This calculation method results, in some cases, in efficiencies higher than ( $E > 2.0$ ), which is practically impossible in the PCR



**Figure 5** Determination of real-time PCR efficiencies from the slopes of the calibration curve (method A), according to the equation:  $E = 10^{(-1/\text{slope})}$  of reference gene glutathione transferase (Gst), and two 'target genes': tryptophan operon (TyrA) and aspartate transcarbamylase (PyrB). CP cycles versus cDNA concentration input (log scale) were plotted to calculate the slope (mean  $\pm$  SD;  $n = 3$ ).

amplification theory. This probably indicates that this efficiency calculation method is not optimal and overestimates the real efficiency.

(B) Efficiency calculation from the fluorescence increases in third linear phase (Figure 2 or at first derivate maximum (FDM) Figure 6) of each logarithmic fluorescence history plot. The investigator has to decide which cycle number to include in the analysis and plot a linear regression, where the slope of the regression line represents the PCR efficiency. Here efficiencies between  $E = 1.35$  and  $E = 1.60$  are found, which differ dramatically from the results above. This efficiency calculation method might underestimate the real efficiency, because data evaluation is made in linear phase near the plateau, where reaction efficiency gets restrictive.

(C) Efficiency calculation on the basis of all fluorescence data points (starting at first cycle to the last cycle), according to a sigmoidal or logistic curve fit model. The advantage of such model is that all data points will be included in the calculation process. No background subtraction is necessary. Slope value is only measured at the point of inflexion at absolute maximum fluorescence increase of FDM and therefore nearly identical to method B ( $1.35 < E < 1.60$ ). This method is easy to perform and a good estimator for the maximum curve slope with high correlation coefficient ( $r > 0.99$ ) and level of significance ( $p < 0.001$ ).

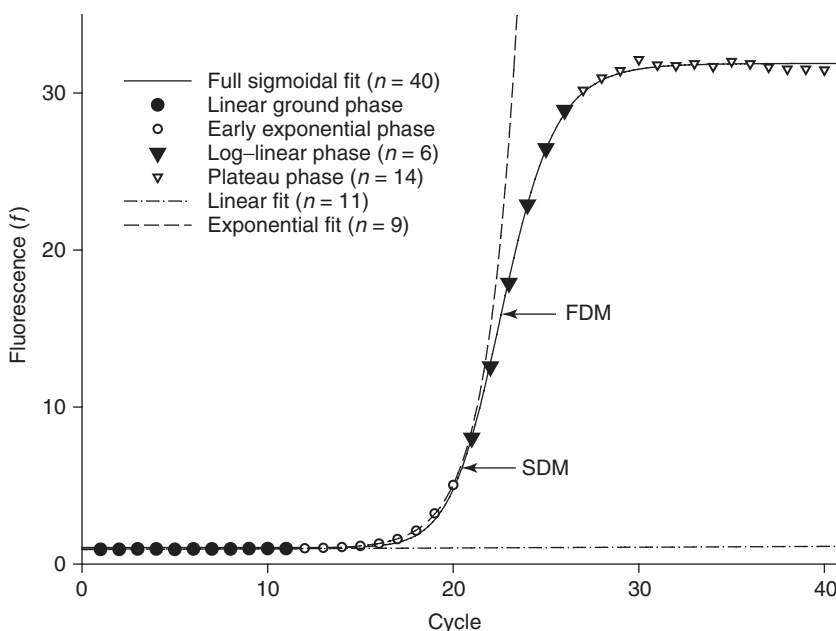
(D) Efficiency calculation from the fluorescence increase only in the second real exponential phase,

according to a polynomial curve fit, as described earlier  $Y_n = Y_0(E)^n$ , where  $Y_n$  is fluorescence acquired at cycle  $n$ , and  $Y_0$  initial fluorescence, so called ground fluorescence. This phase around the second derivate maximum (SDM) exhibit a real exponential amplification behavior (Figure 6). Here, in the exponential part, the PCR reaction kinetic is still under full enzymatic activity with no restrictions. In this method the calculation is performed on each reaction kinetic plot and the amplification efficiency can be determined exactly. They range from  $E = 1.75$  to  $E = 1.90$ , hence are between the other methods.

Which efficiency calculation method is the best or correct one and which one shows the realistic real-time PCR kinetic, and thereby is highly reproducible, has to be evaluated in further trials.

## Normalization of Expression Results

The reliability of any relative RT-PCR experiment can be improved by including an invariant endogenous control in the assay to correct for sample to sample variations in RT-PCR efficiency and errors in sample quantification. A biologically meaningful reporting of target mRNA copy numbers requires accurate and relevant normalization to some standard and is strongly recommended. But the quality of normalized quantitative expression data cannot be better than the quality of the normalizer itself. Any variation in the normalizer will obscure real changes



**Figure 6** Plot of fluorescence observations from LightCycler (Roche Diagnostics). Forty observations give a sigmoid trajectory that can be described by full data fit (four-parametric logistic model). Ground phase can be linearly well regressed. FDM and SDM denote position of FDM and SDM within full data fit.

and produce artificial changes. Real-time RT-PCR-specific errors in the quantification of mRNA transcripts are easily compounded with any variation in the amount of starting material between the samples, e.g., caused by sample-to-sample variation, variation in RNA integrity, RT efficiency differences, and cDNA sample loading variation. This is especially relevant when the samples have been obtained from different individuals, different tissues, and different time courses, and will result in the misinterpretation of the derived expression profile of the target genes. Therefore, normalization of target gene expression levels must be performed to compensate intra- and interkinetic RT-PCR variations. Data normalization can be carried out against an endogenous unregulated reference gene transcript or against total cellular DNA or RNA content (molecules per gram total DNA/RNA and concentrations per gram total DNA/RNA). Normalization according to the total cellular RNA content is increasingly used, but little is known about the total RNA content of cells or even about the mRNA concentrations. The content per cell or per gram tissue may vary in different tissues (*in vivo*), in cell culture (*in vitro*), between individuals, and under different experimental conditions. Nevertheless, it has been shown that normalization to total cellular RNA is the least unreliable method. It requires an accurate quantification of the isolated total RNA or mRNA fraction. Alternatively, the rRNA content has been proposed as an optimal and stable basis for normalization, despite reservations concerning its expression levels, transcription by a different RNA polymerase, and possible imbalances in rRNA and mRNA fractions between different samples.

Commonly used housekeeping genes may be suitable for reference genes, since they are present in all nucleated cell types and are necessary for basic cell survival. The mRNA synthesis of housekeeping genes is considered to be stable in various tissues, even under experimental treatments. However, numerous treatments and studies have already shown that the above-mentioned housekeeping genes are regulated and vary under experimental conditions. It remains up to the individual investigator to choose a reference gene that is best for reliable normalization in their particular experimental setting. It has to be emphasized again, that the choice of housekeeping or lineage specific genes is critical.

Today, there is increasing appreciation of a more reliable normalization in relative quantification. It is vital to develop universal, artificial, and stable, internal standard materials that can be added prior to the RNA preparation to monitor the efficiency of RT as well as the kinetic PCR. Usually more than one housekeeping gene should be tested in a multiple

correlation analysis and its behavior summarized to a housekeeping gene index. According to this index, which is based on the weighted expression of at least three housekeeping genes, a more reliable basis of normalization in relative quantification can be postulated. Recently the software tool GeNorm was established for the evaluation of housekeeping genes expression levels. GeNorm allows for an accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes (<http://allserv.rug.ac.be/~jvdesomp/genorm/>). The normalization strategy used in GeNorm is a prerequisite for accurate kinetic RT-PCR expression profiling, which opens up the possibility of studying the biological relevance of small expression differences.

## Data Processing

After performing qRT-PCR, a reliable data evaluation is next. The calculation unit in real-time PCR is a sample specific and characteristic CP. For CP determination, various fluorescence acquisition methodologies are possible. The Fit Point Method and Threshold Cycle Method measure the CP at a constant fluorescence level. These constant threshold methods assume that all samples have the same cDNA concentration at the point where the fluorescence signal significantly increases in second to third phase over the background fluorescence (Figure 2). Measuring the level of background fluorescence can be a challenge in real-time PCR reactions with significant background fluorescence variations, caused by drift-ups and drift-downs over the course of the reaction. Averaging over a drifting background will give an overestimation of variance and thus increase the threshold level. The threshold level can be calculated by fitting the intersecting line upon the 10-times value of ground fluorescence standard deviation. This acquisition mode can be easily automated and is very robust. In the Fit Point Method, the user discards the uninformative background points, excludes the plateau values by entering the number of log-linear points, and then fits a log line to the linear portion of the amplification curves. These log lines are extrapolated back to a common threshold line and the intersection of the two lines provides the CP value. The strength of this method is its extreme robustness. The weakness is that it cannot be easily automated and requires a lot of user input. Fit Point Method or Threshold Cycle Method can be used on all available platforms with different evaluation of background variability.

The problems of defining a constant background for all samples within one run, sample-to-sample

differences in variance and absolute fluorescence values lead to develop a new acquisition modulus according to mathematical algorithms. In the LightCycler or in Rotor-Gene software packages, the SDM are performed where CP is automatically identified and measured at the maximum acceleration of fluorescence. The amplification reaction and the kinetic fluorescence history over various cycles is obviously not a smooth and easy function. It is possible to fit sigmoidal- and polynomial-curve models, with high significance ( $p < 0.001$ ) and coefficient of correlation ( $r > 0.99$ ), which can be differentiated and the SDM can be estimated. This increase in the rate of fluorescence increase, or better called the acceleration of the fluorescence signal, slows down at the beginning of the third linear phase. Therefore, the cycle where the SDM is always between the second exponential and third linear phase.

Nevertheless, successful application of real-time RT-PCR depends on a clear understanding of the practical problems. Therefore, a coherent experimental design, application, and validation of the individual real-time RT-PCR assay remain essential for accurate and complete quantitative measurement of mRNA transcripts.

**See also:** **Electrophoresis:** Nucleic Acids. **Microscopy Applications:** Proteins and Nucleic Acids. **Nucleic Acids:** Chromatographic and Electrophoretic Methods; Spectroscopic Methods; Electrochemical Methods; Immunoassays. **Polymerase Chain Reaction.**

## Further Reading

- Bustin SA (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology* 25: 169–193.
- Bustin SA (2002) Quantification of mRNA using real-time RT-PCR: trends and problems. *Journal of Molecular Endocrinology* 29(1): 23–39.
- Freeman WM, Walker SJ, and Vrana KE (1999) Quantitative RT-PCR: pitfalls and potential. *Biotechniques* 26(1): 112–125.
- Livak KJ (1997 and 2001) ABI Prism 7700 Sequence detection System User Bulletin #2 Relative quantification of gene expression. <http://docs.appliedbiosystems.com/pe-biodocs/04303859.pdf>.
- Livak KJ and Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* 25(4): 402–408.
- Mannhalter C, Koizar D, and Mitterbauer G (2000) Evaluation of RNA isolation methods and reference genes for RT-PCR analyses of rare target RNA. *Clinical Chemistry and Laboratory Medicine* 38: 171–177.
- Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* 29(9): e45.
- Pfaffl MW (2002) Gene Quantification. Home Page: <http://www.wzw.tum.de/gene-quantification/>.
- Pfaffl MW and Hageleit M (2001) Validities of mRNA quantification using recombinant RNA and recombinant DNA external calibration curves in real-time RT-PCR. *Biotechnology Letters* 23: 275–282.
- Pfaffl MW, Horgan GW, and Dempfle L (2002) Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Research* 30(9): e36.
- Rasmussen R (2001) Quantification on the LightCycler. In: Meuer S, Wittwer C, and Nakagawara K (eds.) *Rapid Cycle Real-Time PCR, Methods and Applications*, pp. 21–34. Heidelberg: Springer Press.
- Schnell S and Mendoza C (1997) Theoretical description of the polymerase chain reaction. *Journal of Theoretical Biology* 188(3): 313–318.
- Thellin O, Zorzi W, Lakaye B, *et al.* (1999) Housekeeping genes as internal standards: use and limits. *Journal of Biotechnology* 75: 291–295.
- Vandesompele J, De Preter K, Pattyn F, *et al.* (2002) Accurate normalisation of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* 3(7): 0034.1–0034.11.

# NUCLEOTIDES

**See** PURINES, PYRIMIDINES, AND NUCLEOTIDES



## OIL-BASED FUELS

See **FUELS: Oil-Based**

## OPTICAL SPECTROSCOPY

### Contents

**Radiation Sources**

**Wavelength Selection Devices**

**Detection Devices**

**Stray Light**

**Spectroscopic Materials**

**Refractometry and Reflectometry**

### Radiation Sources

**C Burgess**, Burgess Consultancy Co., Durham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 3627–3631, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

### Introduction

Optical spectroscopy may be defined as covering that portion of the electromagnetic spectrum (EMS) from 100 nm in the vacuum ultraviolet (UV) to 20 000 nm in the infrared. The relative position of optical spectroscopy within the entirety of the EMS is shown in **Figure 1**.

Essentially the spectroscopic information in this region arises from atomic and molecular transitions involving both inner and outer electrons together with vibrational transitions and rotational fine structure. In principle, the spectroscopic processes of emission, absorption, luminescence, and scattering should all be covered. Within the scope of these four topics, discussion will be confined to the UV–visible (absorption and fluorescence), the near-infrared

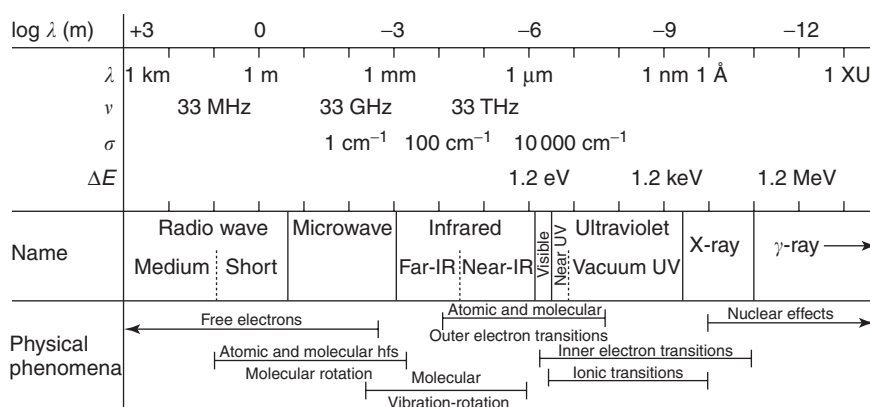
(NIR), and infrared (IR). Aspects of atomic spectroscopy and scattering are covered elsewhere, as are polarization and related effects.

### Overview of Radiation Sources

Ideal spectroscopic sources should have a uniform high-energy output over a wide wavelength range and be stable over long periods of time. In practice, sources show marked wavelength dependence of intensity although most sources are intrinsically stable provided that their power sources are stable.

Broadly speaking spectroscopic sources may be classified into two types: broad band or continuum sources for absorption work and narrow band or line sources for emission work. The latter category is dealt with under atomic emission spectrometry. Laser sources find their greatest application in Raman spectroscopy and will not be covered here. The continuum sources subdivide into two further categories: thermal and arc sources. The arc sources are of higher energy and are primarily used in the UV range and for fluorescence work. The thermal sources are blackbody radiators and cover the majority of the spectral region 300–20 000 nm. For this reason, the underlying theory to blackbody radiation will be dealt with in more detail later.





**Figure 1** The electromagnetic spectrum. (Reproduced with permission from Thorne AP (1998) *Spectrophysics* 2nd edn. London: Chapman and Hall.)

## Arc Sources

These sources operate on the principle that an electric discharge through a suitable ionized gas results in a continuum spectrum. The majority of this energy lies within the UV region of the spectrum,  $\sim 100\text{--}300\text{ nm}$ .

### High-Pressure Arc Sources

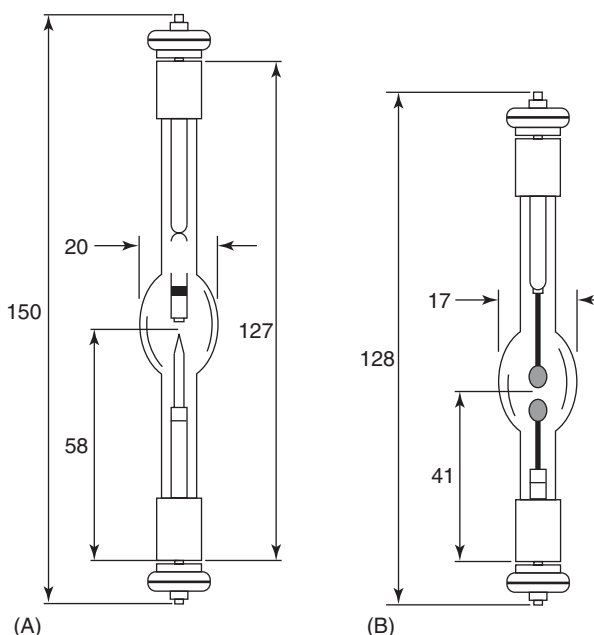
These sources are constructed in the following way. A quartz envelope, containing an inert gas, usually xenon at pressures up to several atmospheres, has an arc struck between a tungsten anode, and cathode. The arc requires high powers of 500 W or more. These sources find the majority of their application in fluorescence work where the signal output is directly proportional to input intensity of the source. Pulsed sources have been used for the vacuum UV although high-intensity a.c. arcs are available in more modern instruments (Figure 2).

### Low-Pressure Arcs

Much lower powers are required in absorption work where the transmittance is a ratio of the input intensity to the transmitted intensity of the beam. For this reason, the much lower powered deuterium arc is the source of choice in the UV.

Its power requirements are much more modest,  $\sim 30\text{ W}$ , and a continuum from  $\sim 180\text{--}350\text{ nm}$  is available (Figure 3). An added advantage is that at lower energies in the visible region, the lamp becomes a line source yielding two strong lines at the 486.1 and 656.3 nm, which are useful for wavelength calibration purposes.

A more recent innovation of low-pressure xenon arcs which use pulsed power supplies allows much increased spectral output. The advantage is that the peak power output can be as high as 40 kW but since the pulse width is  $\sim 1\text{ }\mu\text{s}$ , the average power consumption

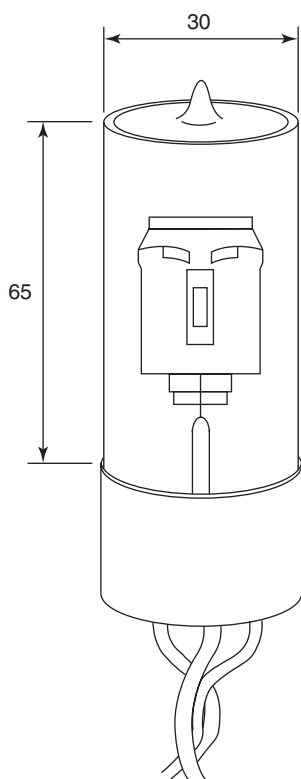


**Figure 2** Typical high-pressure arc lamps. (A) 150 W xenon lamp, (B) 200 W mercury lamp. Dimensions in mm. (Reproduced with permission from Knowles A and Burgess C (1984) *Practical absorption spectrometry*. Ultraviolet Spectrometry Group. London: Chapman and Hall.)

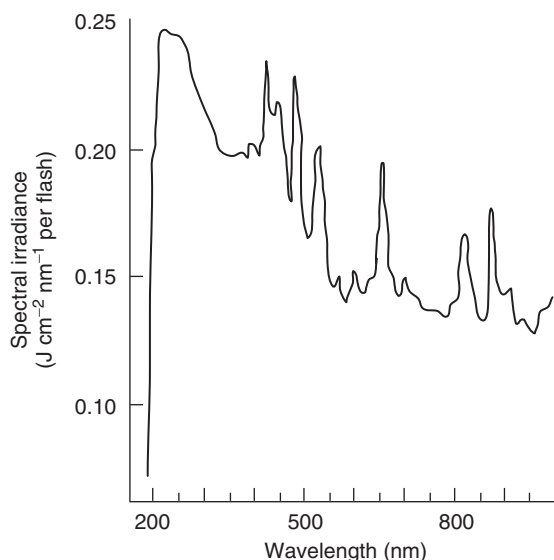
may be less than 5 W. Ozone health and safety issues with high-pressure arcs are absent. The spectral distributions of such lamps approximate to blackbody emitters of 10 000 K or more. A typical spectral output from such a source is shown in Figure 4.

## Blackbody Radiation

An ideal blackbody exhibits perfect emission and absorption of energy at all wavelengths and at all angles. The total radiant energy of such a body is a function only of its temperature. This ideal is shown in Figure 5.

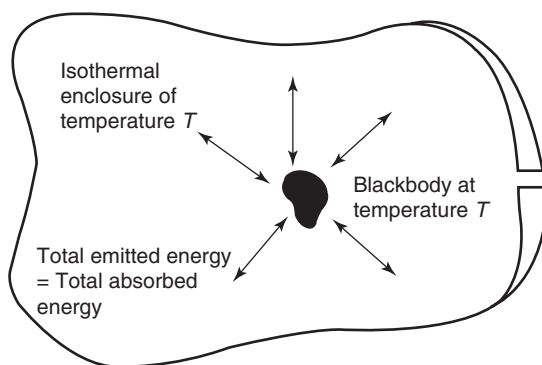


**Figure 3** Typical deuterium low-pressure arc lamp, dimensions in mm. (Reproduced with permission from Knowles A and Burgess C (1984) *Practical absorption spectrometry*. *Ultraviolet Spectrometry Group*. London: Chapman and Hall.)



**Figure 4** Typical spectral output from a low-pressure pulsed xenon source. (Reproduced with permission from Knowles A and Burgess C (1984) *Practical absorption spectrometry*. *Ultraviolet Spectrometry Group*. London: Chapman and Hall.)

In practice, few surfaces closely approach blackbody behavior, e.g., carbon black, platinum black, gold black, and carborundum. However, a small



**Figure 5** Idealized blackbody cavity radiator. From Grum and Becherer (1979).

aperture in a blackbody cavity can closely approximate to a true blackbody source. The differences in behavior of any thermal radiator and ideal blackbody behavior can be described in terms of emissivity, absorptance, and reflectance parameters (Kirchhoff's Law).

The following is a short description of blackbody behavior. For a more detailed explanation see Grum and Becherer (1979). Throughout this section the following symbols will be used:

$c$  = speed of light ( $2.998 \times 10^{-8} \text{ m s}^{-1}$ ),  
 $\nu$  = frequency of the radiation in Hz,  
 $\lambda$  = wavelength of the radiation in m,  
 $h$  = Planck's constant ( $6.626 \times 10^{-34} \text{ J s}$ ),  
 $K$  = temperature in K, and  
 $k$  = Boltzmann's constant ( $1.381 \times 10^{-23} \text{ J K}^{-1}$ )

The average spectral energy density in a blackbody cavity radiator,  $\bar{\omega}_\nu$ , is given in eqn [1].

$$\bar{\omega}_\nu = \left[ \frac{8\pi\nu^2}{c^3} \right] \left[ \frac{h\nu}{e^{h\nu/kT} - 1} \right] \text{ J m}^{-3} \text{ Hz}^{-1} \quad [1]$$

This equation is expressed in terms of frequency as opposed to wavelength. Using the following relationships:

$$\begin{aligned} \bar{\omega}_\nu d\nu &= \bar{\omega}_\lambda d\lambda \\ c &= \nu\lambda \\ d\lambda &= -\frac{c}{\nu^2} d\nu \end{aligned} \quad [2]$$

it can be shown that the equivalent average spectral energy density in terms of wavelength,  $\bar{\omega}_\lambda$ , is given by

$$\bar{\omega}_\lambda = \frac{8\pi hc}{\lambda^5} \frac{1}{e^{h\nu/kT} - 1} \text{ J m}^{-3} \text{ nm}^{-1} \quad [3]$$

It is important to note that  $\bar{\omega}_\lambda \neq \bar{\omega}_\nu$ .

Given a very small aperture in the side of the cavity which does not change the energy distribution within

it, we can now find the radiant power,  $M_\lambda$  from

$$M_\lambda = \pi \frac{c}{4\pi} \bar{\omega}_\lambda \quad [4]$$

where the factor  $c/4\pi$  represents the integrated radiance from  $4\pi$  steradians at any point in the cavity.

We can re-express this in terms of eqn [4] as

$$M_\lambda = \frac{2\pi hc^2}{\lambda^5} \frac{1}{e^{h\nu/kT} - 1} \text{ W m}^{-3} \text{ nm}^{-1} \quad [5]$$

This is the key relationship and is the Planck black-body radiation equation. Integration of this equation over all wavelengths yields the total radiant exitance of a blackbody.

Rearranging eqn [5] and integrating we obtain

$$M_\lambda = 2\pi hc^2 \int_0^\infty \frac{\lambda^{-5}}{e^{h\nu/kT} - 1} d\lambda \quad [6]$$

which may be solved by substitution of

$$x = \frac{hc}{\lambda kT}$$

so that

$$\lambda = \frac{hc}{xkT} \quad \text{and} \quad d\lambda = -\frac{hc}{x^2 kT} dx \quad [7]$$

and eqn [6] becomes

$$M_\lambda = \left(\frac{2\pi h}{c^2}\right) \left(\frac{kT}{h}\right)^4 \int_0^\infty \frac{x^3}{e^x - 1} dx \quad [8]$$

Given that

$$\int_0^\infty \frac{x^3}{e^x - 1} dx = 6 \left(1 + \frac{1}{2^4} + \frac{1}{3^4} + \frac{1}{4^4} + \dots\right) = 6a \quad [9]$$

where  $a = 1.0823\dots$  this simplifies to

$$M_\lambda = \frac{12\pi^5 k^4}{15 h^3 c^2} T^4 = \sigma T^4 \quad [10]$$

This is the Stefan–Boltzmann law where the constant term  $\sigma$  is equal to  $5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$ .

If, on the other hand, we differentiate eqn [5] with respect to wavelength and set to zero we obtain

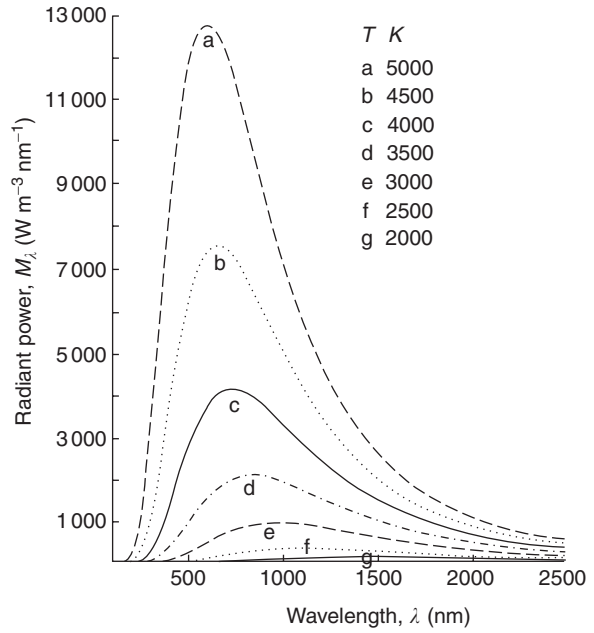
$$\lambda_{\max} T \approx \frac{hc}{5k} \approx 2.88 \times 10^6 \text{ nm K} \quad [11]$$

which enables the wavelength at which maximum spectral energy occurs to be calculated from the absolute temperature. Note that the exact solution is

$$\lambda_{\max} T = \frac{hc}{4.96511\dots k}$$

This is the Wein displacement law.

The output of a blackbody for a given temperature is shown graphically in **Figure 6** by direct substitution in eqn [5]. Note that the position of maximum



**Figure 6** Emission from a perfect blackbody radiator calculated from eqn [5].

spectral output  $\lambda_{\max}$ , is given by eqn [11] and increases with decreasing temperature.

For practical situations where the system is not *in vacuo* eqn [11] is modified to be

$$n\lambda_{\max} T \approx 2.88 \times 10^6 \text{ nm K} \quad [12]$$

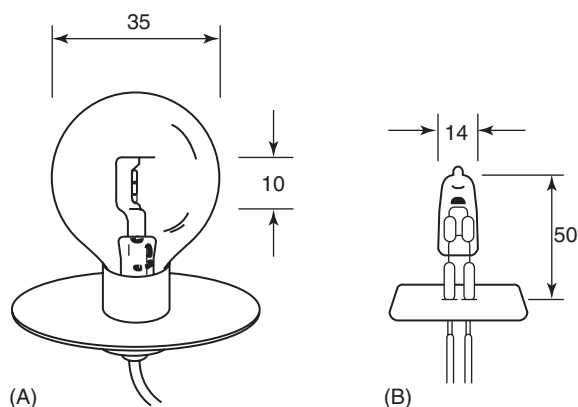
where  $n$  is the refractive index of the medium. Similarly the Stefan–Boltzmann law [10] becomes

$$M_\lambda = n^2 \sigma T^4 \quad [13]$$

This arises from the fact that the speed of light is a function of the density of the medium through which it travels.

## Thermal Sources for the Visible and NIR

The most commonly used thermal source for the visible and NIR is the tungsten filament lamp and its derivative the quartz halogen lamp (**Figure 7**). The tungsten lamp approximates to a 3000 K blackbody radiator below 1000 nm but shows considerable deviation above. Strictly speaking, it is more correct to use the term graybody radiator as in the visible region it exhibits an emissivity of 0.422 that is wavelength independent but the power output is not Planckian. Typical spectral outputs are shown in **Figure 8**.

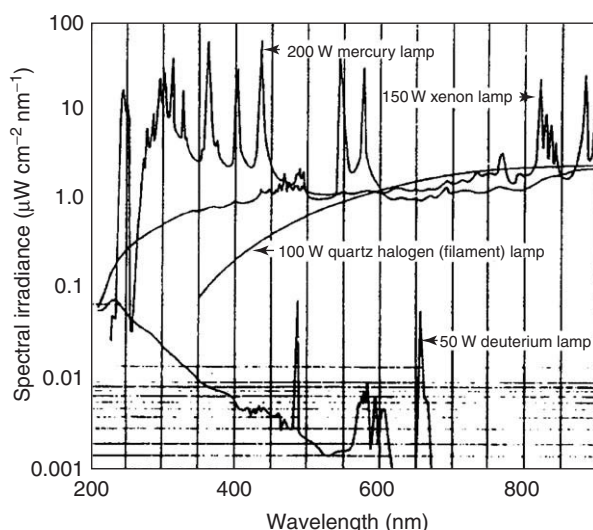


**Figure 7** Tungsten sources. (A) Tungsten filament lamp, and (B) tungsten-halogen lamp, dimensions in mm. Reproduced with permission from Knowles A and Burgess C (1984) *practical absorption spectrometry*. Ultraviolet Spectrometry Group. London: Chapman and Hall.

## Thermal Sources for the IR and Far-IR

One of the earliest sources was the Nernst filament. This is a mixture of rare earth oxides, primarily of zirconium, made into rods 20–50 mm in length and 1–2.5 mm in diameter. These are fired at 1800°C and have platinum wire connections at both ends. The resultant output at 1800–2000°C is approximately blackbody. However, while relatively cheap to produce these have a very unpredictable life and are very sensitive to environmental factors including draughts. Refractory coated platinum-rhodium wire has been used as a more robust solution but is very expensive.

The Globar is a carborundum rod which gives excellent performance at ~1200°C and is much more reliable than the Nernst glower.



**Figure 8** Typical spectral outputs from a variety of sources. (Reproduced with permission from Knowles A and Burgess C (1984) *Practical absorption spectrometry*. Ultraviolet Spectrometry Group. London: Chapman and Hall.)

See also: **Atomic Emission Spectrometry:** Principles and Instrumentation.

## Further Reading

- Grum F and Becherer RJ (1979) *Radiometry*. New York: Academic Press.
- Knowles A and Burgess C (eds.) (1984) *Practical Absorption Spectrometry*. London: Chapman and Hall.
- Norwicka-Jankowska T, Gorczyńska K, Michalik A, and Wieteska E (1986) Analytical visible and ultraviolet spectrometry, vol. XIX. In: Wilson and Wilson (eds.) *Comprehensive Analytical Chemistry*. Amsterdam: Elsevier.
- Thorne AP, Litzén U, and Johansson S (1999) *Spectrophysics*. Berlin: Springer.

## Wavelength Selection Devices

**C Burgess**, Burgess Consultancy Co., Durham, UK

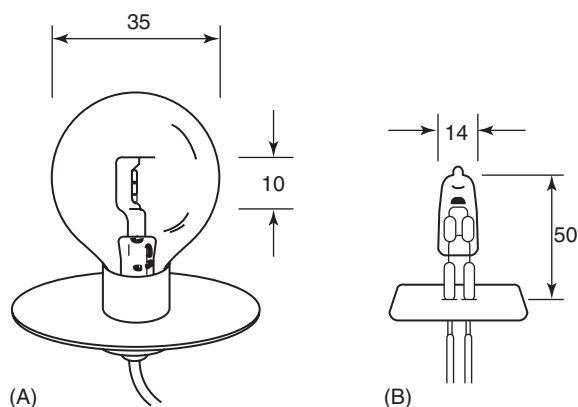
© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 3631–3638, © 1995, Elsevier Ltd.

## Introduction

A fundamental distinction should be drawn between two main types of instrument employed in analytical spectrometry; namely nondispersive and dispersive optical arrangements. In the former instance, the instruments are called photometers and operate

primarily in the visible region and wavelength selection is by the use of a filter of some type. Dispersive instruments, on the other hand, are generally called spectrophotometers and employ a dispersive element, usually a prism or diffraction grating (plane or curved), to effect wavelength selection. The generic term for such an optical arrangement is a monochromator. Exceptionally, the dispersive element may be a curved diffraction grating arranged in conjunction with a solid-state detection system in the focal plane. This latter system is usually referred to as a polychromator. The whole subject of wavelength selection is very large indeed and this article will be

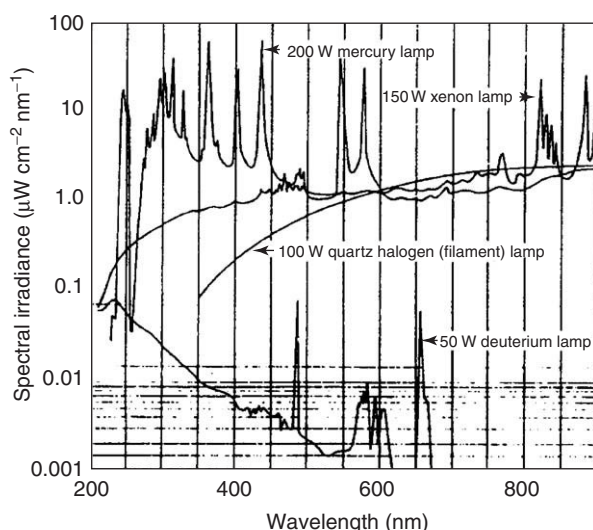


**Figure 7** Tungsten sources. (A) Tungsten filament lamp, and (B) tungsten-halogen lamp, dimensions in mm. Reproduced with permission from Knowles A and Burgess C (1984) *practical absorption spectrometry*. Ultraviolet Spectrometry Group. London: Chapman and Hall.

## Thermal Sources for the IR and Far-IR

One of the earliest sources was the Nernst filament. This is a mixture of rare earth oxides, primarily of zirconium, made into rods 20–50 mm in length and 1–2.5 mm in diameter. These are fired at 1800°C and have platinum wire connections at both ends. The resultant output at 1800–2000°C is approximately blackbody. However, while relatively cheap to produce these have a very unpredictable life and are very sensitive to environmental factors including draughts. Refractory coated platinum-rhodium wire has been used as a more robust solution but is very expensive.

The Globar is a carborundum rod which gives excellent performance at ~1200°C and is much more reliable than the Nernst glower.



**Figure 8** Typical spectral outputs from a variety of sources. (Reproduced with permission from Knowles A and Burgess C (1984) *Practical absorption spectrometry*. Ultraviolet Spectrometry Group. London: Chapman and Hall.)

See also: **Atomic Emission Spectrometry:** Principles and Instrumentation.

## Further Reading

- Grum F and Becherer RJ (1979) *Radiometry*. New York: Academic Press.
- Knowles A and Burgess C (eds.) (1984) *Practical Absorption Spectrometry*. London: Chapman and Hall.
- Norwicka-Jankowska T, Gorczyńska K, Michalik A, and Wieteska E (1986) Analytical visible and ultraviolet spectrometry, vol. XIX. In: Wilson and Wilson (eds.) *Comprehensive Analytical Chemistry*. Amsterdam: Elsevier.
- Thorne AP, Litzén U, and Johansson S (1999) *Spectrophysics*. Berlin: Springer.

## Wavelength Selection Devices

**C Burgess**, Burgess Consultancy Co., Durham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 3631–3638, © 1995, Elsevier Ltd.

## Introduction

A fundamental distinction should be drawn between two main types of instrument employed in analytical spectrometry; namely nondispersive and dispersive optical arrangements. In the former instance, the instruments are called photometers and operate

primarily in the visible region and wavelength selection is by the use of a filter of some type. Dispersive instruments, on the other hand, are generally called spectrophotometers and employ a dispersive element, usually a prism or diffraction grating (plane or curved), to effect wavelength selection. The generic term for such an optical arrangement is a monochromator. Exceptionally, the dispersive element may be a curved diffraction grating arranged in conjunction with a solid-state detection system in the focal plane. This latter system is usually referred to as a polychromator. The whole subject of wavelength selection is very large indeed and this article will be



able to present only a basic outline. A selection of more detailed works is given in Further Reading.

## Filters

Filters are categorized into three main groups: bandpass filters, edge filters, and neutral density filters.

The properties of each of the three main groups are summarized in **Table 1**. Strictly speaking, only bandpass filters are of direct relevance although blocking and attenuation filters are commonly used in optical spectroscopic instruments and are included for the sake of completeness. These properties are more easily assembled in diagrammatic form and typical transmittance–wavelength characteristics are shown in **Figure 1**. An example of a nominal 500 nm bandpass filter is shown in **Figure 1A**.

Edge filters are characterized by an abrupt change between the region of transmittance and the region of rejection. Long wavepass (LWP) filters reject shorter wavelengths and transmit longer wavelengths. The converse is true for short wavepass filters (SWP). By placing an LWP and SWP filter in series, the user can construct a variety of bandpass filters of varying bandwidths. However, except for specialized work, it is more usual to purchase a tailor-made bandpass filter. Commercial bandpass filters are available for wavelengths between 200 and 15 000 nm with bandwidths of between 0.1 nm and several thousand nm. Applications for bandpass filters include spectral radiometry, medical diagnostics, chemical analyses, colorimetry, astronomy, and laser physics.

## Interference Filters

The most common and efficient type of bandpass filter is the interference filter. The basic principles of operation of such filters are illustrated by considering the simple Fabry-Perot filter. This is made by placing a spacer of dielectric medium between two partially transmitting metal films supported on a suitable substrate. These substrates are usually glass or quartz for the visible and ultraviolet whilst in the infrared they include germanium, silicon, IRTRAN II, IV and V, sapphire, indium arsenide, indium antimonide, and arsenic trisulphide. The basic principles of operation are shown in **Figure 2**.

The optical thickness of the dielectric spacer of refractive index,  $n$ , is chosen such that it is half the desired wavelength of maximum transmission. As can be seen from **Figure 2**, the components of the transmitted beam will be in phase if the optical path length difference is an integral number of wavelengths or when eqn [1]

$$2nd \cos \theta' = k\lambda \quad [1]$$

is satisfied,  $\theta'$  is the angle in the dielectric and  $k$  is the order of the interference. The longest wavelength at which a peak occurs is given by

$$\lambda = 2nd \quad [2]$$

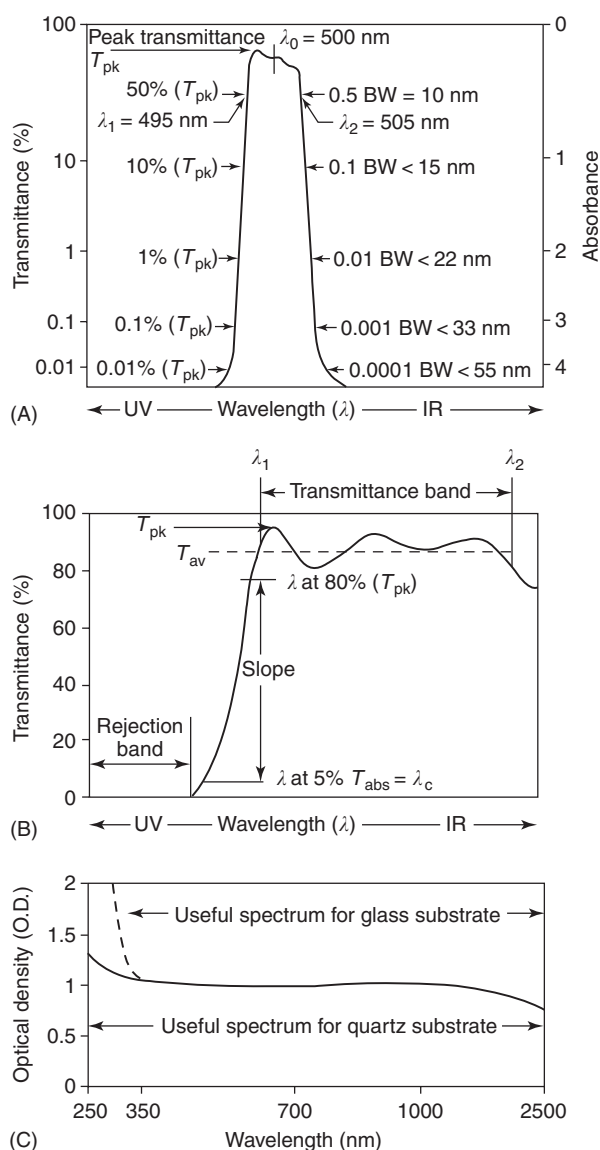
but harmonics will occur also at

$$\lambda = 2nd/k \quad [3]$$

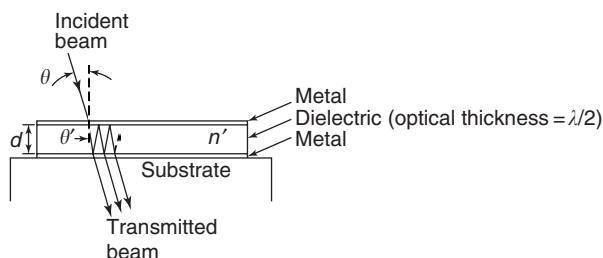
**Table 1** Properties of the main groups of optical filters

Filter	Type	Key parameters	Use
Bandpass ( <b>Figure 1A</b> )	Interference or absorption	$T_{pk}$ -maximum transmittance of filter $\lambda_1$ and $\lambda_2$ – wavelength at 50% $T_{pk}$ at the lower and higher wavelength values Bandwidth (BW) = $\lambda_2 - \lambda_1$ Central wavelength: $\lambda_0 = \lambda_1 + \frac{\lambda_2 - \lambda_1}{2}$	Spectral isolation
Edge ( <b>Figure 1B</b> )	Interference or absorption	Transmittance band $\lambda_2 - \lambda_1$ % $T_{av}$ -average transmittance over the range $\lambda_1$ to $\lambda_2$ Cuton/cutoff wavelength at either 50% $T_{pk}$ or 5% $T_{abs}$ Slope is a measure of the sharpness of the cutoff Slope = $\frac{\lambda_{80\%T_{pk}} - \lambda_{5\%T_{abs}}}{\lambda_{5\%T_{abs}}}$	Blocking filter to remove unwanted spectral radiation
Neutral density ( <b>Figure 1C</b> )	Absorbing glass or metal coated glass or quartz	% Transmittance Spectral neutrality	Even attenuation of radiation over a selected wavelength range





**Figure 1** (A) Bandpass filter, (B) edge filter, and (C) neutral density filter. (Adapted from Corion Corporation Catalogue (1988) pp. 17, 19, 20.)



**Figure 2** Principles of Fabry-Perot interference filter. (Adapted from Grum F and Becherer RJ (1979) *Radiometry*. New York: Academic Press.)

This is the simplest type of filter and has the disadvantage that the value of  $T_{pk}$  is limited by absorption in the metallic films. More sophisticated designs are available involving only dielectric layers which give much greater transmittance but the basic principles are the same.

## Prisms

For many years, the prism formed the basis of most commercial monochromators. The spectral purity of the refracted radiation is determined by the dispersion characteristics of the prism material. The principle of operation is illustrated in Figure 3.

The collimated beam of incident radiation emerges at an angle  $\theta$  to the normal of the second surface. For each value of  $\lambda$ , the refraction is different and hence  $\theta$  is different. The rate of change of the angle  $\theta$  with wavelength  $\lambda$  is known as the angular dispersion  $d\theta/d\lambda$  and can be expressed as

$$\frac{d\theta}{d\lambda} = \frac{d\theta}{dn'} \frac{dn'}{d\lambda} \quad [4]$$

where  $d\theta/dn'$  is dependent only on the geometry of the prism and  $dn'/d\lambda$ , the dispersion, is a characteristic of the material of which the prism is made. Typical dispersion characteristics are shown in Figure 4 for some common spectroscopic materials.

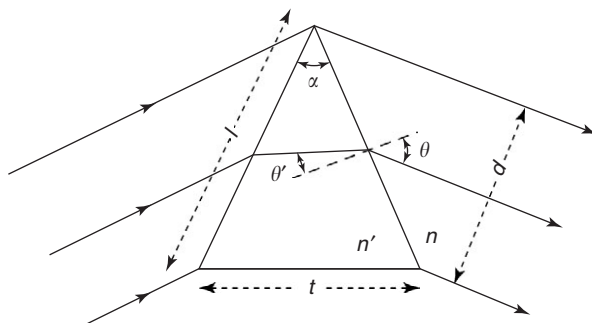
From Snell's Law of refraction

$$n \sin \theta = n' \sin \theta' \quad [5]$$

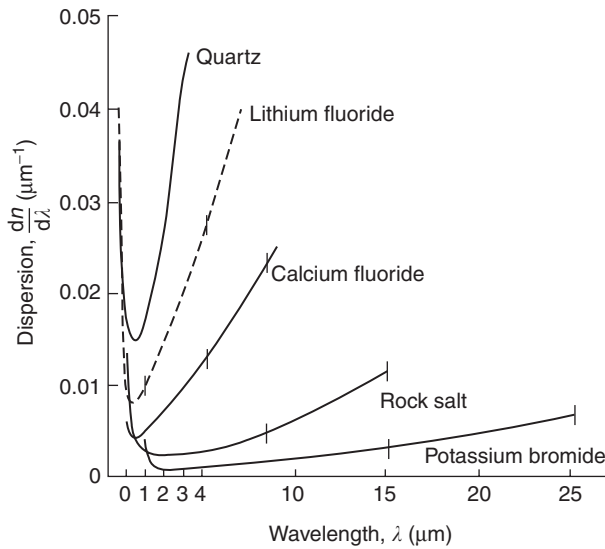
In air  $n = 1$  and, at minimum deviation in the angle,  $\theta'$ , is a constant. The total deviation for both faces is  $2 \sin \theta' / \cos \theta$

Since  $\alpha = 2\theta'$

$$\frac{d\theta}{dn'} = \frac{2l \sin(\alpha/2)}{l \cos \theta} = \frac{t}{a} \quad [6]$$



**Figure 3** Principles of the minimum deviation prism. (Adapted from Grum F and Becherer RJ (1979) *Radiometry*. New York: Academic Press.)



**Figure 4** Dispersion properties of spectroscopic materials. (Adapted from Grum F and Becherer RJ (1979) *Radiometry*. New York: Academic Press.)

therefore showing that  $d\theta/dn'$  is a function only of the geometry of the prism, i.e., the ratio of the base width to the emergent beam width. This relationship leads directly to a method for calculating the resolving power of the prism.

The resolving power,  $R$ , is defined as the mean wavelength divided by the wavelength of two spectral lines which are just resolved

$$R = \frac{\lambda}{d\lambda} \quad [7]$$

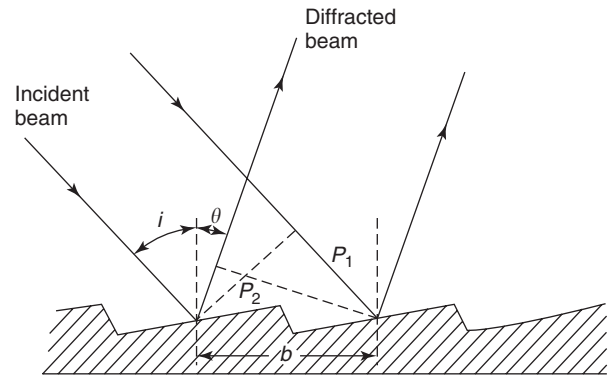
Assuming a rectangular slit, then the refraction limited angular separation of two lines,  $d\theta$ , is approximately  $\lambda/a$  hence the maximum resolving power,  $R_0$  is given by

$$R_0 = t \frac{dn'}{d\lambda} \quad [8]$$

and hence the importance of the prism material.

## Gratings

A diffraction grating is a very simple optical device that operates on the principle of constructive interference. In its simplest form it consists of an optical surface upon which are cut, by means of a ruling engine, many equally spaced grooves. Gratings produced in this way are called master gratings. Replica gratings are made from the master by a technique not unlike the production of gramophone records. The line spacings have to be both small and very precise. For a UV-visible grating the line density is



**Figure 5** Principles of the diffraction grating. (Adapted from Grum F and Becherer RJ (1979) *Radiometry*. New York: Academic Press.)

of the order of 1200 lines per mm over an area of 400 mm<sup>2</sup>.

The development of gratings has taken over 100 years from the simple ruled grating to the modern holographic aspherical type. Holographic gratings are produced using a photographic process and are effectively master grating quality.

There are two fundamental types, the transmission grating and the reflection grating. The former type is not widely used in optical spectroscopy probably because the application range is limited to the absorption characteristics of the dispersion material, and will not be discussed further. Reflection gratings are subdivided into plane and concave gratings. The latter category is further subdivided into spherical and aspherical types. The following simplified discussion refers primarily to plane reflection gratings. The basic physics of reflection gratings is illustrated in Figure 5.

A collimated incident beam of radiation is diffracted by the groove in the grating surface. The angle of the diffracted beam is determined by the necessity of their being in phase (constructive interference). This will happen when the optical path difference between adjacent grooves is an integer multiple of the wavelength (cf. interference filters). This is to say when the difference between  $P_1$  and  $P_2$  (in Figure 5) is equal to an integer,  $k$ , times the wavelength  $\lambda$ .

Using basic trigonometry,

$$P_1 = d \sin i$$

and

$$P_2 = \pm d \sin \theta \quad [9]$$

where  $d$  is the groove spacing.

Note the convention to use a minus sign to indicate when  $i$  and  $\theta$  are on opposite sides of the normal and plus when on the same side. Hence

$$k\lambda = d(\sin i \pm \sin \theta) \quad [10]$$

which is the basic grating equation.

Integer  $k$  is called the order. When  $k = 0$ , i.e. zero order, then the device acts merely as a mirror and all wavelengths are superimposed on one another. When  $k \neq 0$ , eqn [10] shows that for an incident angle  $i$  then the diffracted angle  $\theta$  is dependent on wavelength  $\lambda$ . The angular dispersion is obtained by differentiation of eqn [10]; substituting for  $k$  and  $d$  yields

$$\frac{d\theta}{d\lambda} = \frac{1}{\lambda} \frac{\sin i \pm \sin \theta}{\cos \theta} \quad [11]$$

The important conclusion is that the angular dispersion of a grating depends only on the angles of incidence and diffraction. If it is arranged so that  $i = \theta$  and they are both on the same side of the normal (the Littrow mount) then

$$\frac{d\theta}{d\lambda} = \frac{2}{\lambda} \tan \theta \quad [12]$$

The theoretical resolving power of a grating monochromator is identical to that derived for a prism. Hence the smaller the intergroove distance, the greater the resolution.

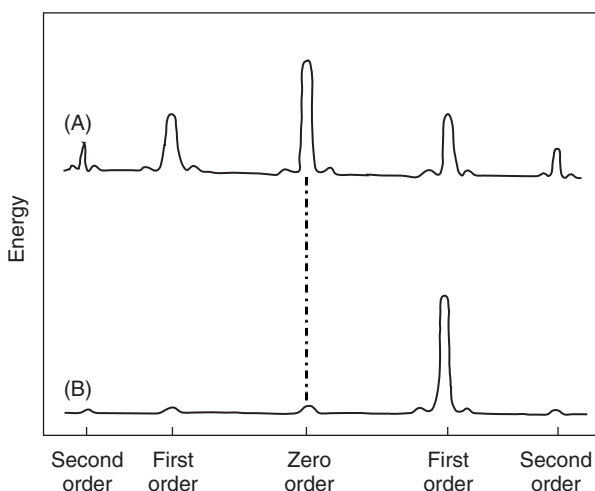
## Blaze

Consider the basic diffraction grating equation for the Littrow mount, which is most commonly used in instrumentation:

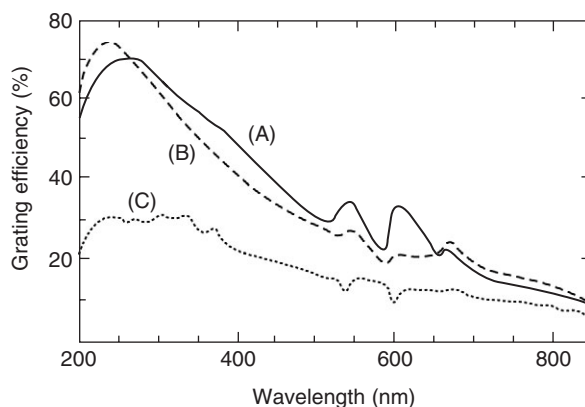
$$k\lambda = d(\sin i + \sin \theta) \quad [13]$$

The light energy distribution diffracted by a grating amongst the various orders ( $k$  in eqn [13]) is illustrated in **Figure 6A**. From an optical design viewpoint, the energy should be concentrated in the first-order spectrum. This may be achieved by use of the blazing technique whereby the angle of the reflecting surfaces on the grating is changed with respect to the plane of the grating surface. The blazing is wavelength-dependent and the blazing wavelength is the wavelength at which the grating efficiency (energy throughput) is at a maximum. A detailed discussion of blazing and grating efficiencies is beyond this short overview and readers are referred to Hutley (1982).

In addition the efficiency of the grating depends upon the shape of the grooves. Basically the sharper the groove the greater is the optical efficiency. Ruled gratings have a sharper V-shaped groove and high



**Figure 6** Energy profiles of monochromatic light from a grating: (A) nonblazed; (B) blazed with improved energy profile.



**Figure 7** Effect of grating type on the efficiency: (A) ruled grating blazed at 240 nm; (B) holographic grating blazed at 260 nm; (C) sinusoidal holographic grating.

efficiency (**Figure 7A**) but poor stray light characteristics. Holographic gratings have U-shaped grooves (sinusoidal) and are much less efficient but have excellent stray light characteristics (**Figure 7C**). However, blazing of holographic gratings restores optical efficiency (**Figure 7B**) and concentrates the energy in the first-order spectrum (**Figure 6B**). Stray light is clearly an issue particularly with gratings but this topic will be dealt with in a subsequent article.

However, one of the problems with gratings is the presence of 'ghosts' between the orders due to spurious reflections. This effect is almost completely eliminated when master holographic gratings are used.

The development of gratings has been both long and complex. The main features are summarized in **Table 2**.

**Table 2** The main features of some types of gratings

Grating type	Key points	References
Simple ruled	Diamond points on speculum metal or glass	Harrison GR (1949) <i>Journal of the Optical Society of America</i> 39: 413 (historical review)
	Diamond points onto copper	Wood RJ (1944) <i>Journal of the Optical Society of America</i> 34: 509.
	Diamond points onto aluminum sputtered onto optically flat glass	Wood RJ (1937) <i>Nature</i> 140: 723.
Transmission echelon	Very high resolution use in spectroscopes. Spectrally limited	Michaelson A (1898) <i>Astrophysics Journal</i> 8: 37.
Reflecting echelon	Wider spectral range than the transmission echelon but difficult to manufacture	Williams W (1933) <i>Proceedings of the Physics Society. London</i> 45: 699.
Echelle	Saw tooth groove structure with large $d$ spacing used in the infrared	Wood RW and Trowbridge A (1910) <i>Philosophical Magazine</i> 20: 886, 898.
Echelle	Reflection grating similar properties to the reflecting echelon but easier to make. Normally used in conjunction with a prism	Harrison GR (1949) <i>Journal of the Optical Society of America</i> 39: 522.
Concave	Spherical grating in which the slit, the detector, and the radius of the grating all lie on the same circle (Rowland Circle)	Rowland HA (1883) <i>Philosophical Magazine</i> 16: 197.
Holographic	Very low stray light, high efficiency, low cost	Hayat GS and Pienchard G (1968) <i>Handbook of Diffraction Gratings. Ruled and Holographic.</i> Jobin-Yvon.

**Table 3** Choice of grating type by spectral region

Spectral region	Type	Comments
Vacuum UV	Concave grating	One reflection minimizes energy losses
UV-visible–near-IR	Plane holographic gratings up to 1400 grooves per mm, or spherical or aspherical types with diode array detection	Aspherical types correct for optical aberrations
IR and far-IR	Echelette plane gratings in Ebert-Fastie or Czerny–Turner configurations	Reflections are very high in the IR. However, elimination of high order reflections is a problem.

The application of gratings by spectral region is summarized in Table 3.

### Comparison of Gratings and Prisms

Gratings offer significant advantages over prisms in that

1. dispersion is essentially linear over their wavelength range making easier monochromator design;
2. temperature sensitivity is not nearly so great a problem as with prisms;
3. reflection gratings are available over the whole range of optical spectroscopy (plane and curved); and
4. optical systems containing gratings can be made much more compactly.

Many of the problems of the older ruled gratings, i.e. overlap of spectral orders, ghosts, high stray light

etc., have been virtually eliminated with modern holographic methods.

### Optical Configurations

The whole subject of optical mounting design occupies many volumes. The discussion herein will be limited to an overview of modern grating instruments. Of these, four designs dominate the plane grating applications namely, Littrow, Ebert–Fastie, Czerny–Turner, and Monk–Gillieson.

The object of an ideal optical configuration is that there are no optical aberrations, no stray light, and no energy loss through the system. In practice, various compromises have to be made to minimize the effect of the various variables. Within a practical optical system, five types of optical aberration are possible which give rise to image distortion, namely primary spherical aberration, coma, astigmatism (anamorphism), field curvature, and distortion.

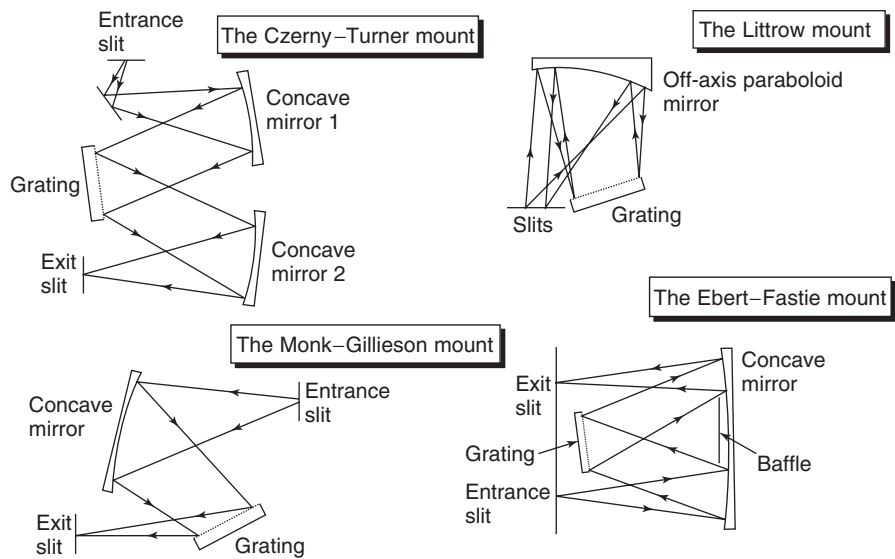


Figure 8 Main types of optical mount.

Table 4 Comparison of strengths and weaknesses of plane grating optical mountings

Mounting type	Strengths	Weaknesses
Littrow	Simple arrangement 1 mirror 3 reflections Slit positions allow compact system and minimize astigmatism	Susceptible to coma aberration Needs paraboloid mirror Slits are close together The curvature of the slit image is wavelength dependent
Ebert-Fastie	1 mirror 3 reflections Coma compensation	Requires large mirror Needs curved slits to minimize effects of astigmatism High energy losses
Czerny-Turner	Coma compensation 2 mirrors 3 reflections	Needs curved slits to reduce effects of astigmatism High energy losses
Monk-Gilleson	Coma compensation and astigmatism correction 2 reflections 1 mirror High energy throughput	Requires large grating Subject to linear shifts in aberration

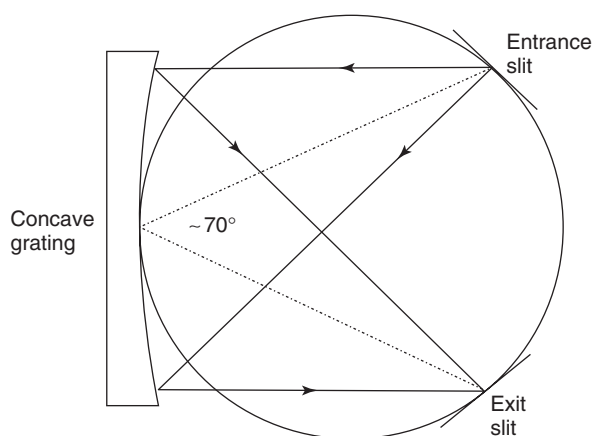
Stray light is minimized using master holographically produced gratings and energy loss is minimized by reducing the number of reflections within the optical mounting. This is particularly important in the vacuum UV. The four main optical mounts are shown in Figure 8. A comparison of strengths and weaknesses is given in Table 4.

For energy-intensive applications, concave gratings are ideal as only one reflection is required. These are based on the Rowland circle. This is illustrated in Figure 9. The optimum angle between the entrance and exit slits is  $\sim 70^\circ$ . The use of only one optical

element combines the functions of collimating, dispersing and focusing optics. The design does suffer from inherent astigmatism although in modern aspherical (ellipsoidal or paraboloidal) holographic gratings these problems can be overcome. However, they are expensive to make.

Slit Function

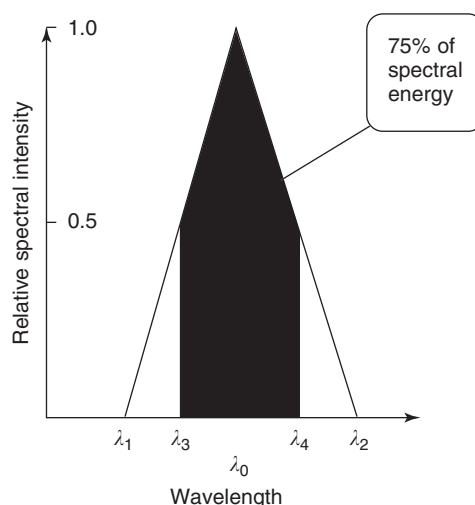
The slit function of a grating or prism monochromator is the image of the slit in light of the appropriate



**Figure 9** Seya-Namioka design based on the Rowland circle.

wavelength. Ideally this slit function is triangular when the monochromatic image of the entrance slit exactly fills the exit slit. The triangular slit function arises simply out of the convolution of the two identical rectangular images of the entrance and exit slits of a monochromator. If this condition is not met, i.e. different entrance and exits slit widths, then the profile is trapezoidal. The ideal situation is illustrated in **Figure 10** for equal entrance and exit slits which are larger than the diffraction limit.

See also: **Optical Spectroscopy**: Stray Light.



**Figure 10** Idealized slit function. For a nominal wavelength  $\lambda_0$  the spectral slit width is  $\lambda_2 - \lambda_1$  and  $\lambda_4 - \lambda_3$  is the effective spectral slit width (spectral bandwidth).

## Further Reading

- Hutley MC (1982) *Diffraction Gratings*. London: Academic Press.
- Knowles A and Burgess C (eds.) (1984) *Practical Absorption Spectrometry*. London: Chapman and Hall.
- Longhurst RS (1973) *Geometrical and Physical Optics*. London: Longman.
- Marr IL (1975) Instrumentation for spectroscopy. In: *Wilson and Wilson's Comprehensive Analytical Chemistry*, vol. IV Amsterdam: Elsevier.

## Detection Devices

**C Burgess**, Burgess Consultancy Co., Durham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 3638–3643, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

## Introduction

The ideal characteristics for a detector are high sensitivity, low noise, linearity of response, and uniform response with respect to wavelength.

Detection devices for optical spectroscopy fall into two distinct groups: photon detectors and thermal detectors. The fundamental difference is illustrated in an idealized way in **Figure 1**. For a photon detector, the spectral responsivity is a function of wavelength

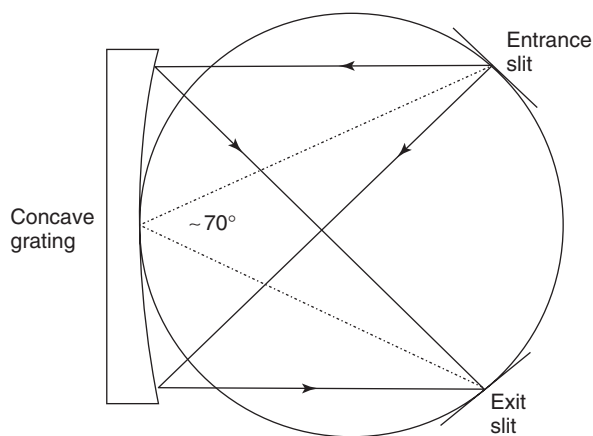
and has a finite upper bound (long wavelength cut-off) – whereas the thermal detector's response is essentially flat.

Photon detection is generally used from the vacuum ultraviolet to the near infrared and thermal detection from the mid-infrared onwards although there is some overlap. **Figure 2** shows detector types for optical spectroscopy and detectors.

The criteria for choosing a detector are:

1. wavelength responsivity,
2. sensitivity,
3. rate of change of the optical signal to be measured (e.g., chopping frequency),
4. ruggedness, and
5. physical factors, e.g., size of beam, maximum acceptance, solid angle, uniformity of response over the detector surface etc.

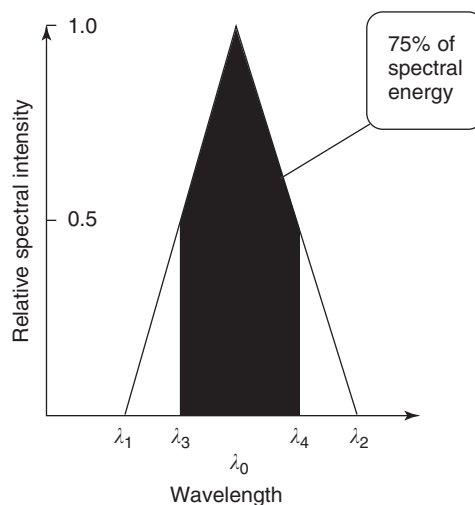




**Figure 9** Seya-Namioka design based on the Rowland circle.

wavelength. Ideally this slit function is triangular when the monochromatic image of the entrance slit exactly fills the exit slit. The triangular slit function arises simply out of the convolution of the two identical rectangular images of the entrance and exit slits of a monochromator. If this condition is not met, i.e. different entrance and exits slit widths, then the profile is trapezoidal. The ideal situation is illustrated in **Figure 10** for equal entrance and exit slits which are larger than the diffraction limit.

See also: **Optical Spectroscopy**: Stray Light.



**Figure 10** Idealized slit function. For a nominal wavelength  $\lambda_0$  the spectral slit width is  $\lambda_2 - \lambda_1$  and  $\lambda_4 - \lambda_3$  is the effective spectral slit width (spectral bandwidth).

## Further Reading

- Hutley MC (1982) *Diffraction Gratings*. London: Academic Press.
- Knowles A and Burgess C (eds.) (1984) *Practical Absorption Spectrometry*. London: Chapman and Hall.
- Longhurst RS (1973) *Geometrical and Physical Optics*. London: Longman.
- Marr IL (1975) Instrumentation for spectroscopy. In: *Wilson and Wilson's Comprehensive Analytical Chemistry*, vol. IV Amsterdam: Elsevier.

## Detection Devices

**C Burgess**, Burgess Consultancy Co., Durham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 3638–3643, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

## Introduction

The ideal characteristics for a detector are high sensitivity, low noise, linearity of response, and uniform response with respect to wavelength.

Detection devices for optical spectroscopy fall into two distinct groups: photon detectors and thermal detectors. The fundamental difference is illustrated in an idealized way in **Figure 1**. For a photon detector, the spectral responsivity is a function of wavelength

and has a finite upper bound (long wavelength cut-off) – whereas the thermal detector's response is essentially flat.

Photon detection is generally used from the vacuum ultraviolet to the near infrared and thermal detection from the mid-infrared onwards although there is some overlap. **Figure 2** shows detector types for optical spectroscopy and detectors.

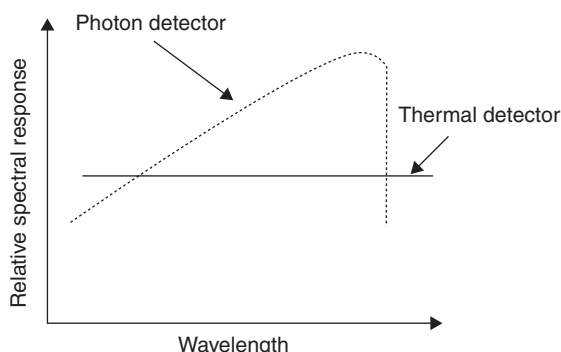
The criteria for choosing a detector are:

1. wavelength responsivity,
2. sensitivity,
3. rate of change of the optical signal to be measured (e.g., chopping frequency),
4. ruggedness, and
5. physical factors, e.g., size of beam, maximum acceptance, solid angle, uniformity of response over the detector surface etc.

Photon detectors may have the first three characteristics but not the last. Thermal detectors do not enjoy high sensitivity and may be quite noisy but are generally less sensitive to wavelength.

## Photoemissive Detectors

The mode of action of this type of detector is based on the release of electrons from a photosensitive surface by absorption of photons of incident radiation. The released electrons' flow is either measured directly or amplified to provide a measure of the



**Figure 1** Idealized detector characteristics.

incident flux. If the current is measured directly then the device is known as a photocell or phototube and if amplified this is known as a photomultiplier.

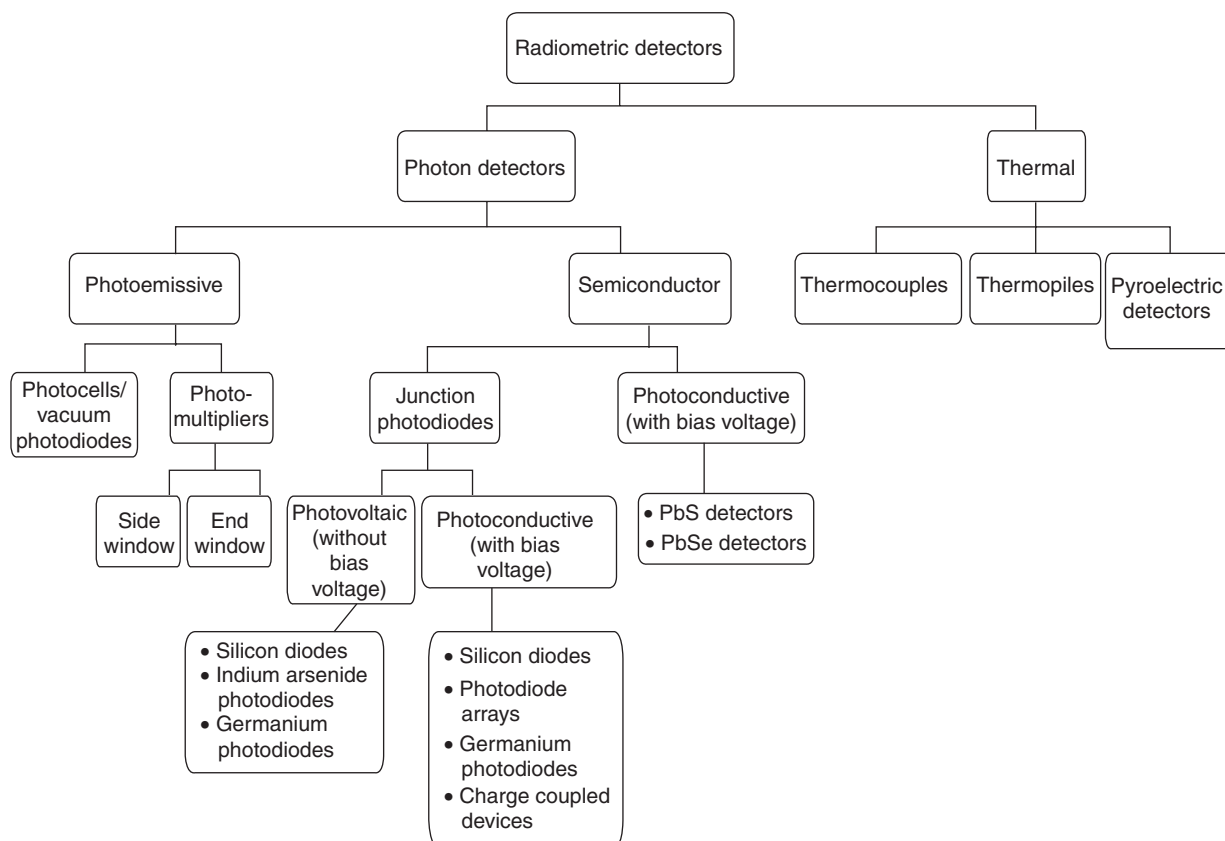
## Photocells/Vacuum Photodiodes

These are essentially simple devices consisting of a photocathode coated with an alkali metal and an anode sealed under a low-pressure argon atmosphere in a glass or quartz envelope. The coating material for the photocathode needs to be chosen for the spectral range of interest. These detectors were common in early instruments but have been largely superseded by silicon photodiodes. A schematic diagram of their construction is shown in **Figure 3**.

The quantum efficiency of these devices is somewhat low and they suffer from both short-term cathode fatigue (i.e., decrease in overall sensitivity) which is reversible after storage in the dark and irreversible cathode aging. However, they have very small dark currents of between  $10^{-9}$  and  $10^{-11}$  A.

## Photomultipliers

In effect, these are photocells with multiple amplification stages. The emitted electron from the photocathode is accelerated by a voltage drop towards an



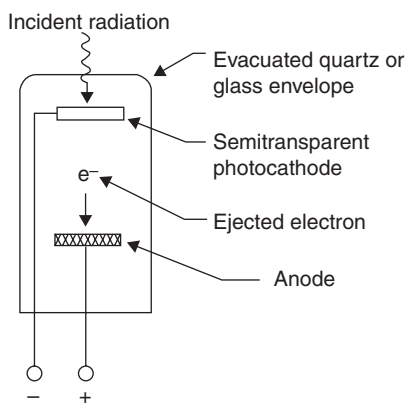
**Figure 2** Detector types for optical spectroscopy.

electrode (the first dynode) which on impact produces a secondary electron shower. The process is continued usually up to 9 dynodes (or stages) until the anode is reached. By this stage the original photocathode current  $I_c$  has been amplified to a final current equal to  $I_c E^n$  where  $E$  is the secondary electron emission coefficient and  $n$  is the number of dynodes. The quantum efficiency of the photocathode does not usually exceed 40%.

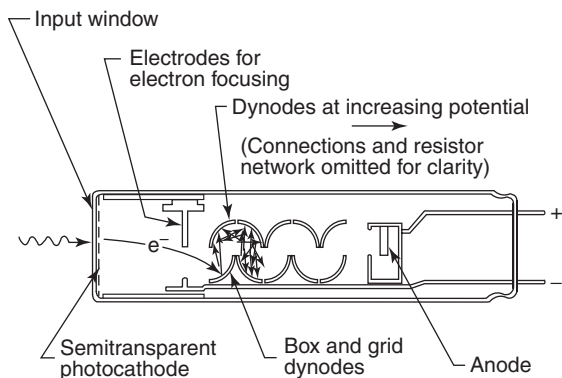
There are two types of photomultipliers – end window and side window. The end-window design is illustrated in Figure 4.

Photomultiplier tubes (PMTs) are commercially available from 120 to 1200 nm (albeit not in one device) and have:

- a fast rise time (1–2 ns side window, 10–15 ns end window);
- excellent current amplification;
- a wide linear dynamic range (typically  $10^6$  or better); and
- a low dark current (typically 0.1–30 nA);



**Figure 3** Schematic representation of a vacuum photodiode. (Courtesy of Newport Corporation.)



**Figure 4** Schematic representation of an end-window photomultiplier tube. (Courtesy of Newport Corporation.)

but they require or exhibit:

- a high voltage stabilized d.c. power supply,
- detector cooling for photon counting applications (temperature affects the dark current),
- shielding from overexposure to light to prevent burnout,
- shielding from magnetic fields, and
- spatial sensitivity.

It is the last problem, in common with photocells, that is the greatest potential problem particularly in double beam spectrometers where image shifts can occur.

## Semiconductor Detectors

All semiconductor detectors operate on the photoconductive principle, namely the inner photoelectric effect. However, it is normal to use the term photoconductive detector for those that rely on the incident photons to change the conductivity in the bulk of the photoconductive layer. When a p–n junction is present these are called junction detectors or photodiodes. Photodiodes are generally subdivided further into photovoltaic (i.e., operating without bias voltage) and photoconductive (i.e., operating with a bias voltage). The mode of operation usually depends on the requirements for the detector electronics.

### Photoconductive Detectors

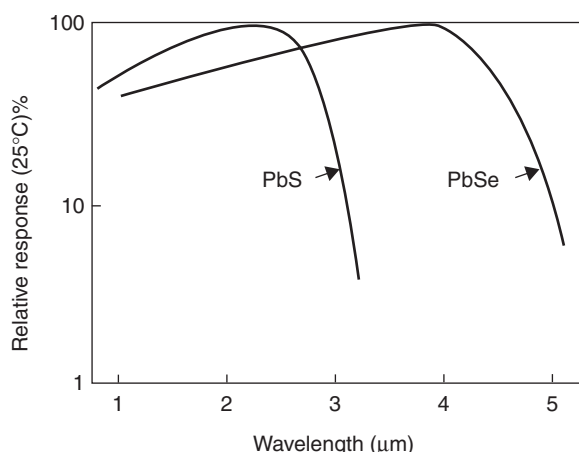
**Lead sulfide (PbS) detectors** Before the development of red enhanced PMTs, PbS photoconductive cells were widely used for detection in the red and infrared (IR) regions. They were made by depositing a sensitized microcrystalline layer of PbS on glass or quartz and attaching two electrodes before sealing *in vacuo* in glass envelope. Under dark conditions, an applied voltage produces a small current. Illumination gives increases in current proportional to the incident flux.

However, PbS detectors suffer from a number of problems including:

- variable response characteristics dependent on the manufacturing process variables;
- temperature effects on spectral responsivity, dark current, and signal-to-noise (S/N) ratio;
- the response time at room temperature is of the order of a millisecond; and
- they are very sensitive to exposure to ultraviolet (UV) or visible radiation.

A typical response curve is shown in Figure 5.

However, modern versions are much more robust and give excellent response from 800 to 3000 nm without cooling and are much more compact.



**Figure 5** Relative responses of lead sulfide and lead selenide detectors. (Courtesy of Newport Corporation.)

**Lead selenide (PbSe) detectors** These are similar in construction and mode of operation to lead sulfide detectors but have an enhanced infrared response to 4500 nm. The relative response curve shown in **Figure 5**. However, these responses are normalized in sensitivity terms. The PbSe detectors are more than 100 times less sensitive than their PbS counterparts.

### Junction Photodiodes

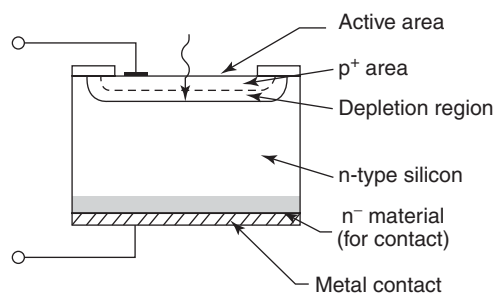
A schematic diagram of a junction photodiode is shown in **Figure 6**.

Incident photons absorbed in the region of the p-n junction produce hole-electron pairs. When operating in the photovoltaic mode, these hole-electron pairs migrate to opposite sides of the junction producing a voltage and consequent current flow in the detector electronics.

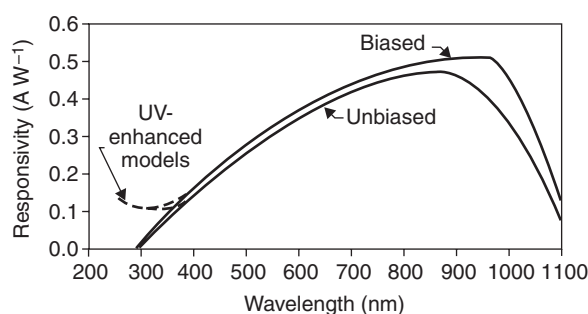
If, however, a reverse bias is applied across the junction, then the conductance is greatly increased thereby producing a current which is proportional to the light flux. This mode of operation is called the photoconductive mode.

**Silicon photodiodes** These are the most commonly used detectors in the region 100–1100 nm. However, UV-enhanced models extend the range into the UV to ~180 nm. They may be separated with or without bias. **Figure 7** shows some typical response characteristics.

**Indium arsenide and germanium photodiodes** Indium arsenide photodiodes are used for the near-infrared region and typically cover 1800–3600 nm. They are destroyed by bias voltages in excess of 1 V and are only separated in the photoconductive mode. They have a fast rise time (~100 ns) and give excellent



**Figure 6** Schematic representation of a junction photodiode. (Courtesy of Newport Corporation.)



**Figure 7** Spectral response for silicon photodiodes. (Courtesy of Newport Corporation.)

performance at room temperature. The spectral response region of germanium photodiodes is smaller, typically 600–1900 nm, and have longer rise times (~5 μs), but are three orders of magnitude more sensitive than indium arsenide. The output from a germanium photodiode is usually linear over a  $10^6$  intensity scale. They are best used unbiased but, unlike indium arsenide, are capable of being operated up to a 10 V bias.

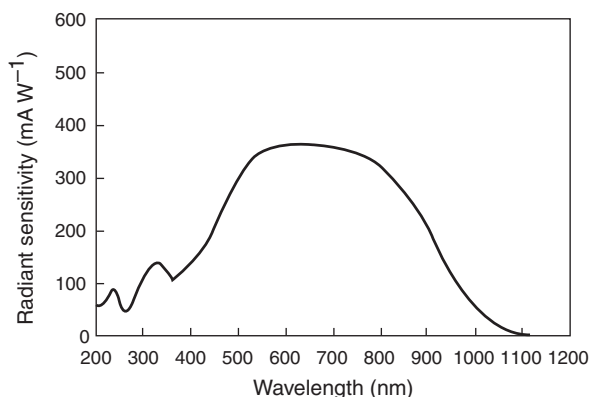
**Photodiode arrays** Basically, these arrays are made up of silicon photodiodes linearly set up within an integrated circuit. The numbers in each array are typically 512 or 1024. The photocurrent from each diode causes a capacitance build-up proportional to the incident light flux, which is read by sequentially discharging the array elements via an analog-to-digital converter for data processing. These arrays have a spectral response similar to that of the silicon photodiode, i.e., 180–1100 nm. A typical sensitivity plot is shown in **Figure 8**.

The advantage of such a multichannel device is that data at many wavelengths may be collected in a single measurement cycle which, for many purposes, may be regarded as instantaneous (readout time is in the order of milliseconds). However, photodiode arrays are much less sensitive than a photomultiplier tube and if low light levels are to be observed then

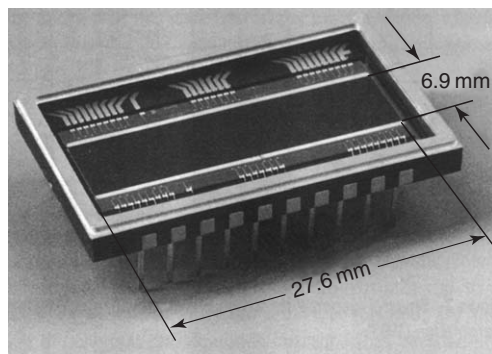
cooling of the array is required. The application of diode array detectors for liquid chromatography and for UV-visible spectrophotometry has been extensive and the arrays are generally not cooled. Individual diodes are small and are typically  $25\ \mu\text{m}$  in width and  $2.5\ \text{mm}$  high.

**Charge coupled detectors** These devices are not yet commonly available in commercial instrumentation for analytical spectrophotometry although they are used in applications in inductively coupled plasma atomic emission spectrometry. However, they have found extensive application in imaging and astronomical applications. Essentially they are two-dimensional photodiode arrays which allow many spectra to be acquired in one readout. A typical array sensor is shown in Figure 9.

This has a diode pixel density of  $1024 \times 256$  giving an effective size of  $27\ \mu\text{m}$  by  $27\ \mu\text{m}$ . This device is capable of generating 74 500 spectra per second and is up to 100 times more sensitive than conventional linear diode array detectors.



**Figure 8** Spectral response for a photodiode array. (Courtesy of Newport Corporation.)



**Figure 9** Example of a CCD detector. (Courtesy of Newport Corporation.)

## Thermal Detectors

The principle of operation of a thermal detector is that the absorption of radiation leads to one of four effects: a change in its resistance, production of a thermoelectric voltage, altered capacitance, or a rise in the temperature.

In the first category, the detectors are known as bolometers which consist of blackened thin metal strips or, if based on semiconductors, thermistors. Neither of these types is used widely nowadays in analytical spectrophotometry although bolometers have been used in infrared spectrometers.

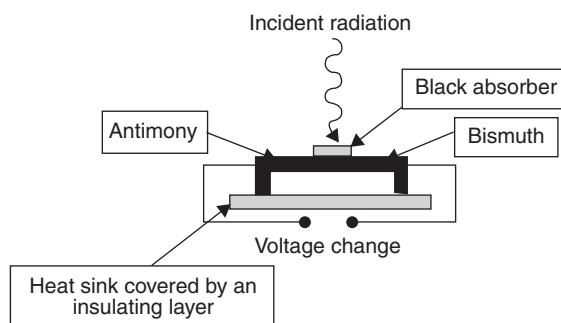
The thermoelectric voltage effect is found in thermo couples and thermopiles. A temperature dependence of capacitance in certain materials gives rise to the pyro-electric effect.

The fourth and final effect is essentially calorimetry which is used in radiometry but does not find application in analytical spectrophotometry. However, one particular detector, the Golay detector, was formerly used in the infrared. This consisted of a chamber containing xenon gas to which was attached a mirrored membrane. Deformations in the mirrored surface caused by heat absorption were measured optically. These detectors were very expensive to make and were temperamental, and have largely been superseded.

## Thermocouples and Thermopiles

A thermocouple generates a voltage at the junction of two dissimilar metals when irradiation produces a temperature change. The metals most commonly employed are antimony and bismuth. The generation of a voltage in this way is known as the Seebeck effect. Figure 10 shows an idealized thermocouple function.

The absorption of radiation causes the temperature of the junction to rise to  $T^\circ\text{C}$  and a voltage is generated proportional to the temperature change  $\Delta T = T - T_0$ . The ratio  $\Delta V / \Delta T$  is a characteristic of the two metals and is known as the Seebeck



**Figure 10** Thermocouple, idealized diagram.



coefficient. If a number of thermocouples are placed in series (20–120 is typical), they are known as thermopiles. Thermopiles have a very wide wavelength range (200–50 000 nm), are very rugged, and no bias is required. However, they have very slow response times especially as they become larger (100–1000 ms). Commercial thermopile detectors tend to be packaged in cylindrical housings with a window. Window materials include fused silica, calcium fluoride, potassium bromide, KRS-5, Irtran, and silicon. The transmittance characteristics of this window material defines the usable spectral range. In addition, the physical robustness of the window material to humidity will limit its applicability. Spectroscopic materials are discussed in a later article.

### Pyroelectric Detectors

The pyroelectric effect is a change in surface charge (capacitance) in response to a change in temperature. Pyroelectric materials are usually crystalline and are coated with a black absorbing material. The thermal mass of these devices has to be kept small. Since it is a change in temperature which generates the current, the incident radiation must be chopped. However, the response times of pyroelectric detectors are much higher than for thermocouples or thermopiles. As with these detectors, the spectral range is largely typically 200–40 000 nm. They are used

primarily for the measurement of radiant power. These detectors have excellent stability and can usually operate between 0°C and 70°C with a relatively small temperature coefficient ( $\approx 0.2\%$  change per °C).

*See also:* **Optical Spectroscopy:** Radiation Sources; Wavelength Selection Devices; Stray Light; Spectroscopic Materials; Refractometry and Reflectometry.

### Further Reading

- Budde W (1982) *Physical Detectors of Optical Radiation*. New York: Academic Press.
- Dereniak EL and Crowe DG (1984) *Optical Radiation Detectors*. New York: Wiley.
- Grum F and Becherer RJ (1979) *Radiometry*. New York: Academic Press.
- Keyes RJ (ed.) (1976) *Optical and InfraRed Detectors*. Berlin: Springer.
- Kingston RH (1978) *Detection of Optical and InfraRed Radiation*. Berlin: Springer.
- Martin AE (1966) *Infra-red Instrumentation and Techniques*. Amsterdam: Elsevier.
- Oriel Corporation (1993) *Light Sources, Monochromators and Detection Systems*.
- Smith RA, Jones FE, and Chasmar RP (1968) *Detection and Measurement of Optical Radiation*. Oxford: Oxford University Press.

## Stray Light

**C Burgess**, Burgess Consultancy Co., Durham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 3643–3647, © 1995, Elsevier Ltd., with revisions made by the Editor.

### Introduction

Essentially, stray light is radiant energy to which the detector is sensitive, and which should not be there. More exactly it may be described as light that lies outside the spectral bandwidth of the spectrometer. More correctly it is to be called stray radiant energy (SRE). Over the years it has caused many a spectroscopist grief by causing bias to absorbance readings, nonlinearity of response as well as false absorbance maxima.

In 1966, Edisbury was so moved as to liken the effect of stray light to a quotation from Shakespeare's

*Henry IV* (Part I iii 10) "Company, villainous company, hath been the spoil of me".

Indeed, the definition and measurement of stray light has bedevilled practical spectroscopists for more than 50 years. As far as the analytical scientist is concerned, it is instrumental stray light (ISL) that is the problem. However, optical designers are as concerned with monochromator stray light (MSL) which arises from the imperfections of optical surfaces, principally the grating, within the monochromator itself.

While, in theory, stray light is a problem throughout optical spectroscopy, most attention has been paid to the ultraviolet (UV) and visible regions and latterly to the infrared as the regions in which most problems occur.

### Instrumental Stray Light

The actual amount of instrumental stray light (ISL) is governed by the characteristics of the sample. This



coefficient. If a number of thermocouples are placed in series (20–120 is typical), they are known as thermopiles. Thermopiles have a very wide wavelength range (200–50 000 nm), are very rugged, and no bias is required. However, they have very slow response times especially as they become larger (100–1000 ms). Commercial thermopile detectors tend to be packaged in cylindrical housings with a window. Window materials include fused silica, calcium fluoride, potassium bromide, KRS-5, Irtran, and silicon. The transmittance characteristics of this window material defines the usable spectral range. In addition, the physical robustness of the window material to humidity will limit its applicability. Spectroscopic materials are discussed in a later article.

### Pyroelectric Detectors

The pyroelectric effect is a change in surface charge (capacitance) in response to a change in temperature. Pyroelectric materials are usually crystalline and are coated with a black absorbing material. The thermal mass of these devices has to be kept small. Since it is a change in temperature which generates the current, the incident radiation must be chopped. However, the response times of pyroelectric detectors are much higher than for thermocouples or thermopiles. As with these detectors, the spectral range is largely typically 200–40 000 nm. They are used

primarily for the measurement of radiant power. These detectors have excellent stability and can usually operate between 0°C and 70°C with a relatively small temperature coefficient ( $\approx 0.2\%$  change per °C).

**See also:** **Optical Spectroscopy:** Radiation Sources; Wavelength Selection Devices; Stray Light; Spectroscopic Materials; Refractometry and Reflectometry.

### Further Reading

- Budde W (1982) *Physical Detectors of Optical Radiation*. New York: Academic Press.
- Dereniak EL and Crowe DG (1984) *Optical Radiation Detectors*. New York: Wiley.
- Grum F and Becherer RJ (1979) *Radiometry*. New York: Academic Press.
- Keyes RJ (ed.) (1976) *Optical and InfraRed Detectors*. Berlin: Springer.
- Kingston RH (1978) *Detection of Optical and InfraRed Radiation*. Berlin: Springer.
- Martin AE (1966) *Infra-red Instrumentation and Techniques*. Amsterdam: Elsevier.
- Oriel Corporation (1993) *Light Sources, Monochromators and Detection Systems*.
- Smith RA, Jones FE, and Chasmar RP (1968) *Detection and Measurement of Optical Radiation*. Oxford: Oxford University Press.

## Stray Light

**C Burgess**, Burgess Consultancy Co., Durham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 3643–3647, © 1995, Elsevier Ltd., with revisions made by the Editor.

### Introduction

Essentially, stray light is radiant energy to which the detector is sensitive, and which should not be there. More exactly it may be described as light that lies outside the spectral bandwidth of the spectrometer. More correctly it is to be called stray radiant energy (SRE). Over the years it has caused many a spectroscopist grief by causing bias to absorbance readings, nonlinearity of response as well as false absorbance maxima.

In 1966, Edisbury was so moved as to liken the effect of stray light to a quotation from Shakespeare's

*Henry IV* (Part I iii 10) "Company, villainous company, hath been the spoil of me".

Indeed, the definition and measurement of stray light has bedevilled practical spectroscopists for more than 50 years. As far as the analytical scientist is concerned, it is instrumental stray light (ISL) that is the problem. However, optical designers are as concerned with monochromator stray light (MSL) which arises from the imperfections of optical surfaces, principally the grating, within the monochromator itself.

While, in theory, stray light is a problem throughout optical spectroscopy, most attention has been paid to the ultraviolet (UV) and visible regions and latterly to the infrared as the regions in which most problems occur.

### Instrumental Stray Light

The actual amount of instrumental stray light (ISL) is governed by the characteristics of the sample. This

can be elegantly demonstrated for direct reading spectrometers. For a given wavelength setting,  $\lambda$ , let the total detector signal be denoted by  $P_t$  and  $P_\lambda$  be that contribution to  $P_t$  made by radiation fully within the bandpass of the monochromator. An additional combination,  $P_s$ , to the total detector signal is caused by ISL outside the bandpass of the monochromator. Hence,

$$P_t = P_\lambda + P_s \quad [1]$$

If we assume a linear response for the detector, the total response in the presence of sample,  $P'_t$ , is given by

$$P'_t = P_\lambda T_\lambda + P_s T_s \quad [2]$$

where  $T_\lambda$  is the true transmittance of the sample at wavelength  $\lambda$  and  $T_s$  is the transmittance of the sample to ISL,  $s$ , at wavelength  $\lambda$ .

The apparent transmittance

$$T'_\lambda = \frac{P_\lambda T_\lambda + P_s T_s}{P_t} \quad [3]$$

but since  $P_s = sP_t$  and  $P_\lambda = (1-s)P_t$  eqn [3] becomes

$$T'_\lambda = (1-s)T_\lambda + sT_s = T_\lambda + s(T_s - T_\lambda) \quad [4]$$

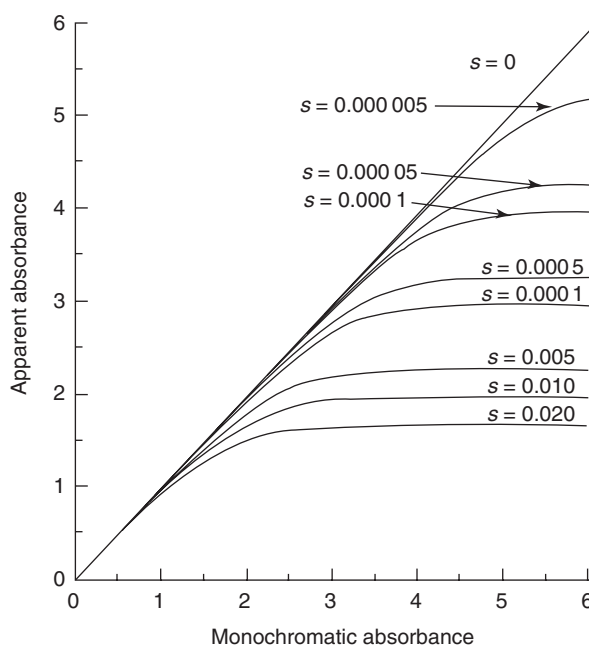
Note that the sample here includes both cell and solvent. The important feature of eqn [4] is that it clearly indicates that the apparent transmittance may be greater or less than the true transmittance depending on the magnitudes of  $T_s$  and  $T_\lambda$ . One important point to be made here is that the only ISL considered is the one that causes a detector signal. This may appear pedantic but as photon detectors show a marked upper cutoff in response, so stray light of wavelengths longer than that will not matter.

The most common occurrence is that  $T_s > T_\lambda$  and hence  $T'_\lambda > T_\lambda$ , which means that the effect is an apparent low absorbance. This situation occurs primarily near source or solvent cutoffs, e.g., below 220 nm in the UV.

The effect for a variety of values of  $s$  and  $T_s = 1$  is shown in Figure 1. The limiting values of these curves are  $-\log_{10}s$ . If  $T_s < T_\lambda$ , then the opposite effect is observed namely a greater apparent absorbance. However, this is rare but may be encountered when nearing absorbance minima.

## The Measurement of Instrumental Stray Light

The concept of stray light applies to all of optical spectroscopy. However, the problems encountered are primarily in the UV portion of the spectrum. In



**Figure 1** Effect of stray light on absorbance assuming  $T_s > T_\lambda$  (eqn [4]).

reality, the problem is just as great at the upper limit of the wavelength scale but there is no easy way of measuring the effect. As far as analytical science is concerned, the practical impact of ISL is an important consideration. In concept, the amount of stray light is readily determined using a low wavelength cutoff filter.

Rewriting eqn [4] in terms of  $s$  we obtain

$$s = \frac{T'_\lambda - T_\lambda}{T_s - T_\lambda} \quad [5]$$

If the cutoff filter is effective then  $T_\lambda = 0$  and hence

$$s = \frac{T'_\lambda}{T_s} \quad [6]$$

If we assume that the filter is transparent to the stray light then

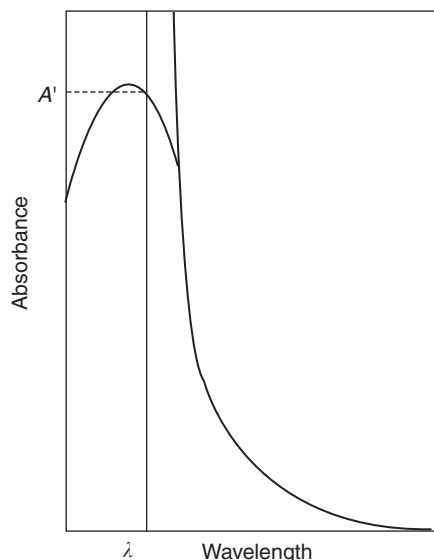
$$s = T'_\lambda = 10^{-A'} \quad [7]$$

where  $A'$  is the apparent absorbance measured at a wavelength  $\lambda$  (Figure 2). This is essentially the method defined by the ASTM. However, in practice, the apparent absorbance plateau lies beyond the scale range of most spectrophotometers. It is therefore necessary to attenuate the energy in the reference beam in order to bring it on scale.

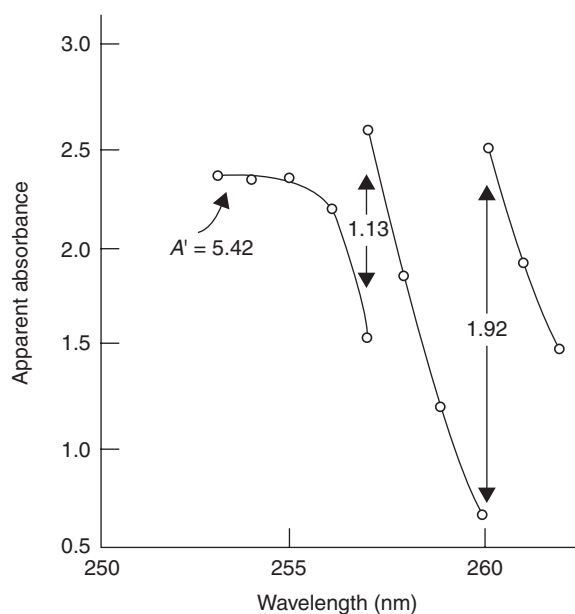
An example is shown in Figure 3, whereby the cutoff filter used is a  $10 \text{ g l}^{-1}$  solution of potassium iodide in a 10 mm cuvette. The instrument under test was a double beam recording instrument with a

readout scale of 0–3 absorbance. Two attenuation filters (1.92 and 1.13 absorbance, respectively) were needed to observe the plateau. Hence the apparent absorbance,  $A'$ , was 5.42 and the stray light fraction calculated from eqn [7] was  $3.8 \times 10^{-6}$ .

An easier method has been proposed involving attenuation of the reference beam with the same cutoff



**Figure 2** Effect of stray light on the apparent absorbance (---) of a cutoff filter. (Reproduced with permission from Buist GJ (1981) In: Burgess C and Knowles A (eds.) *Standards in Absorption Spectrometry*, ch. VI. London: Chapman and Hall.)

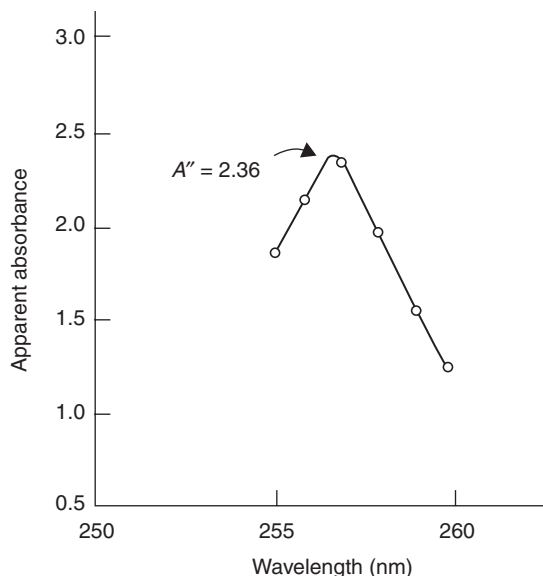


**Figure 3** ASTM method. (Reproduced with permission from Mielenz KD, Weidener VR, and Burke RW (1982) *Heterochromatic stray light in UV absorption spectrometry: A new test method. Applied Optics* 21(18): 3354–3356.)

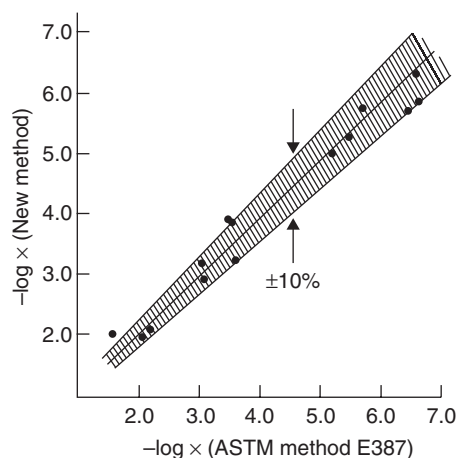
filter solution but in a 5 mm cuvette and has shown that the apparent absorbance observed under these conditions.  $A''$ , is related to the stray light ratio by

$$s \approx \frac{1}{4} 10^{-2A''} \quad [8]$$

The equivalent result for the spectrometer shown in **Figure 3** is  $4.6 \times 10^{-6}$  absorbance. The resultant trace is shown in **Figure 4**. Comparative data for 15 instruments showed good agreement between the methods (**Figure 5**).



**Figure 4** Mielenz method. (Reproduced with permission from Mielenz KD, Weidener VR, and Burke RW (1982) *Heterochromatic stray light in UV absorption spectrometry: A new test method. Applied Optics* 21(18): 3354–3356.)



**Figure 5** Comparison of the ASTM method with the Mielenz method. (Reproduced with permission from Mielenz KD, Weidener VR, and Burke RW (1982) *Heterochromatic stray light in UV absorption spectrometry: A new test method. Applied Optics* 21(18): 3354–3356.)

The slope of the line was  $1.01 \pm 0.1$  and the authors conclude that both methods give equivalent results. An alternative approach has also been developed. Experimental observations have shown that for a geometrical series of solutions, the transmittance ratio forms a minimum as the reference absorbance increases (Figure 6). The theoretical curve, for zero stray light, or true monochromatic absorbance, is curve A. Curve B shows the experimentally determined result.

These plots are determined using the following procedure:

1. For a stock solution having a monochromatic absorbance of between 4 and 5, prepare dilutions at a constant geometric ratio (1.25 is the most practical being a dilution of 20–25 ml).
2. Using the least absorbing (highest transmitting) solution first as the solution in the reference beam, measure the apparent transmittance (transmittance ratio) of the next most concentrated solution and so on.
3. Plot the transmittance ratio against the reference absorbance.

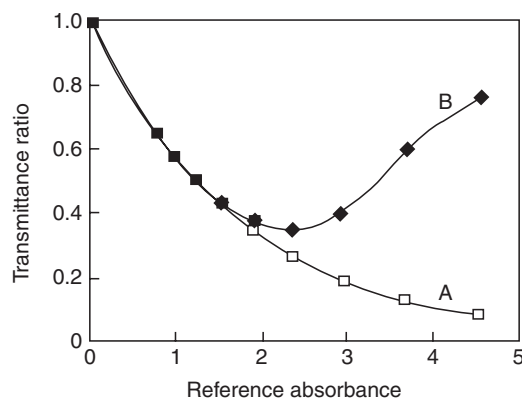
It has been shown that the stray light fraction,  $s$ , was given by

$$s = \left[ 1 + \frac{(1 - \rho)}{(\alpha - 1)} \cdot \left( \frac{\rho}{\alpha} \right)^{\frac{\alpha}{1-\alpha}} \right]^{-1} \quad [9]$$

where  $\rho$  = minimum of the transmittance ratio curve and  $\alpha$  is the geometrical concentration ratio.

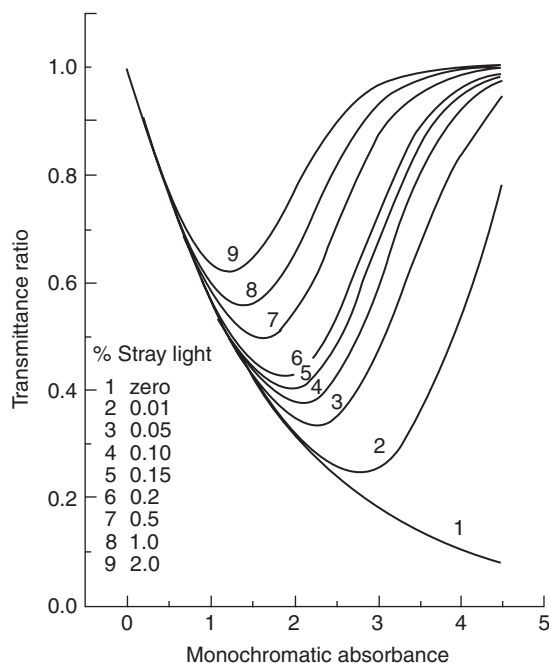
For  $\alpha = 1.25$  eqn [9] simplifies to

$$s = \left[ 1 + \frac{(1 - \rho)}{0.25} \cdot \left( \frac{\rho}{1.25} \right)^{-5} \right]^{-1} \quad [10]$$

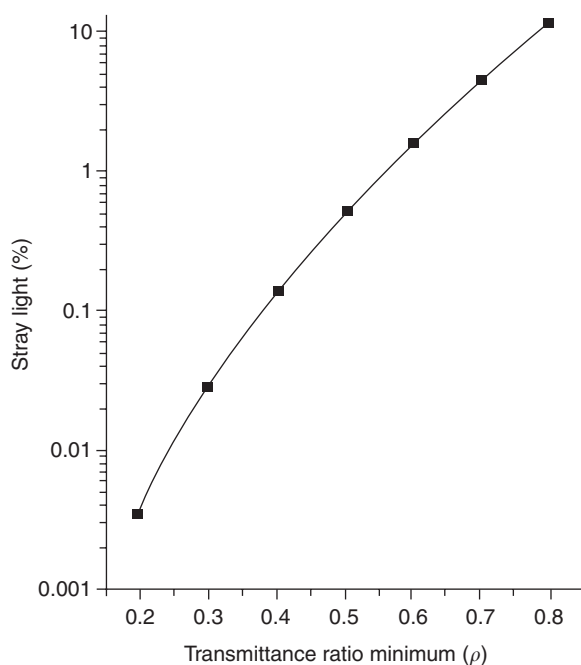


**Figure 6** Experimental observation of the transmittance ratio minimum (eqn [9]). A, theoretical curve assuming monochromatic radiation; B, experimental values. (Fleming P (1990) Transmittance ratio spectrometry as a stray radiant energy test method in spectrophotometry. *Analyst* 115: 375–378; reproduced by permission of The Royal Society of Chemistry.)

Figure 7 shows the effect of various values of  $s$  on the transmittance ratio observed as calculated by eqn [9]. Hence for any value of the transmittance ratio minimum,  $\rho$ , it is possible to calculate the corresponding value of  $s$  (Figure 8).



**Figure 7** The effect of stray light on the transmittance ratio minimum,  $\rho$ , calculated from eqn [9].



**Figure 8** Relationship between the transmittance ratio minimum,  $\rho$ , and the percentage of ISL.

It has been further shown that this approach and the one used to derive eqn (8) are theoretically equivalent.

Both these approaches suffer from the problem that values obtained are likely to underestimate the amount of stray light because of sample absorption (i.e.  $T_s < 1$ ). For conventional transmittance instruments, a method has been developed for determining the amount of stray light absorbed by the sample.

Restating eqn [4] for values of  $s \ll 1$ , it becomes

$$T'_\lambda = (1 - s)T_\lambda + s \approx T_\lambda + s \quad [11]$$

if it is assumed that  $T_s$ , the transmittance of the stray light is unity. However, if this is not the case, then a factor  $T_s^A$ , needs to be included to compensate for the attenuation of the stray light by the sample.

$$T'_\lambda = T_\lambda + sT_s^A \quad [12]$$

The contribution by the stray light to high transmittance (i.e. low concentrations of chromophore) is negligible since  $s \ll T_\lambda$  and eqn [12] becomes

$$\begin{aligned} T'_\lambda &= T_\lambda \\ \text{or} \\ \log_{10} T'_\lambda &= -A \end{aligned} \quad [13]$$

However, the contribution of stray light predominates at low transmittances (i.e. high chromophore concentrations where  $s > 50 T_\lambda$  and eqn [12] becomes

$$\begin{aligned} T'_\lambda &= sT_s^A \\ \text{or} \\ \log_{10} T'_\lambda &= \log_{10} s + A \log_{10} T_s \end{aligned} \quad [14]$$

If the observed transmittance,  $T'_\lambda$ , is plotted on semi-log axes against the monochromatic absorbance,  $A$ , then curves of the type shown in Figure 9 result;

where  $s \leq 0.02 T_\lambda$ , eqn [13] holds,

where  $T_\lambda \leq 0.02 s$ , eqn [14] holds,

where  $T_\lambda \approx s$  the 'elbow' region is observed.

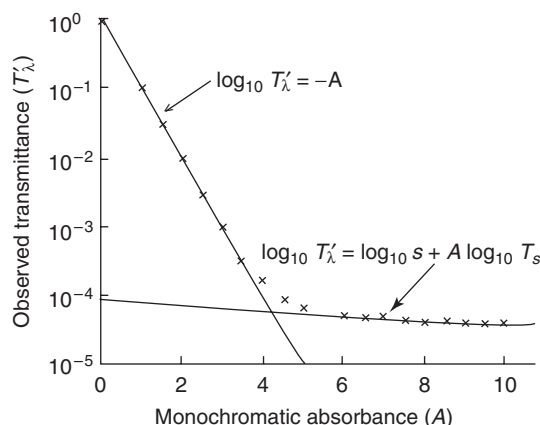
The true stray light level,  $s$ , is readily available from the extrapolation of eqn [14] to  $A = 0$  and the slope gives the fraction of the stray light transmitted by the sample.

In practice, the method requires the use of a reference beam attenuation solution as in the ASTM method to allow the scale reading to be observed. The monochromatic absorbances are determined using 1 mm cuvettes and 10 mm cuvettes for the observed absorbances. Note that the term differential absorbance and transmittance ratio are equivalent.

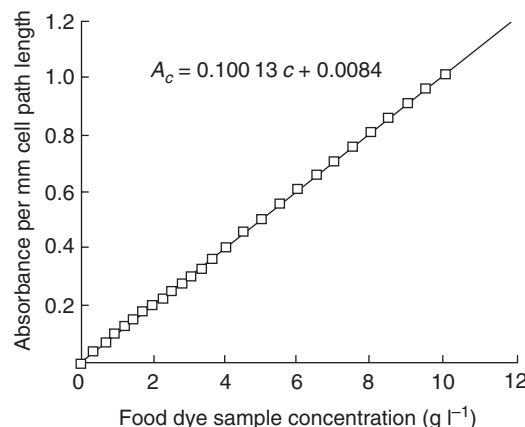
A food dye, Orleans Blue, in aqueous solution was used to provide an arithmetic series of solutions in

the range 0–1 absorbance per mm at 630 nm. The best fit linear regression line (Figure 10) was found to be  $A_c = 0.10013c + 0.0084$  ( $\pm 0.0005$ ) $c + 0.0084$  ( $\pm 0.0026$ ). Hence monochromatic absorbances could be calculated for all solutions in 10 mm cuvettes. The measurements were repeated using 10 mm cuvettes using a 5 mm slitwidth and the results plotted (Figure 9).

To observe the second linear region, a solution having a monochromatic absorbance of 2.003 was used in the reference beam and the solutions remeasured. As the differential absorbance readings suffered fluctuation, repeated measurements were needed to estimate the mean value. The interpolation yielded a value for  $s$  of 0.0000893 ( $\pm 0.0000015$ ) and a value for  $T_s$  of 0.907 ( $\pm 0.005$ ).



**Figure 9** Observed transmittance versus monochromatic absorbance. (Fleming P and O'Dea J (1991) Stray radiant energy test method in spectrometry based on direct transmittance spectrometry. *Analyst* 116: 195–198; reproduced by permission of The Royal Society of Chemistry.)



**Figure 10** Absorbance linearity of Orleans Blue food dye solution (in 1 mm cuvettes). (Fleming P and O'Dea J (1991) Stray radiant energy test method in spectrometry based on direct transmittance spectrometry. *Analyst* 116: 195–198; reproduced by permission of The Royal Society of Chemistry.)

See also: **Optical Spectroscopy:** Radiation Sources; Wavelength Selection Devices; Detection Devices; Spectroscopic Materials; Refractometry and Reflectometry.

## Further Reading

- ASTM E 387–84 (1995) Standard Test Method for Stray Radiant Power Ratio of Spectrophotometers by the Opaque Filter Method. (This standard is likely to be revised in 2004.)
- Edisbury JR (1966) *Practical Hints on Absorption Spectrometry*. London: Adam Hilger.
- Fleming P (1999) In: Burgess C and Frost T (eds.) *Standards and Best Practice in Absorption Spectrometry*. Oxford: Blackwell Science.

- Kaye WI (1987) In: Burgess and Mielenz KD (eds.) *Advances in Standards and Methodology in Spectrophotometry*. Amsterdam.
- Neal WTL (1956) *Photoelectric Spectrometry Group Bulletin* 9: 204.
- Palmer EW, Hutley MC, Franks A, Verril JF, and Gale B (1975) Diffraction Gratings. *Reports on Progress in Physics* 38: 975–1048.
- Sharpe MR (1984) Stray Light in UV Spectrophotometers. *Analytical Chemistry* 56(2): 339A–356A.
- Slavin W (1963) Stray Light in Ultraviolet, Visible, and Near-infrared Spectrophotometry. *Analytical Chemistry* 35(4): 561–566.
- Tarrant AWS (1978) Optical techniques for studying stray light in spectrophotometers. *Optica Acta* 25: 1167.

## Spectroscopic Materials

**C Burgess**, Burgess Consultancy Co., Durham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 3647–3651, © 1995, Elsevier Ltd.

materials used to construct optical components and some consideration of solvents for spectroscopy.

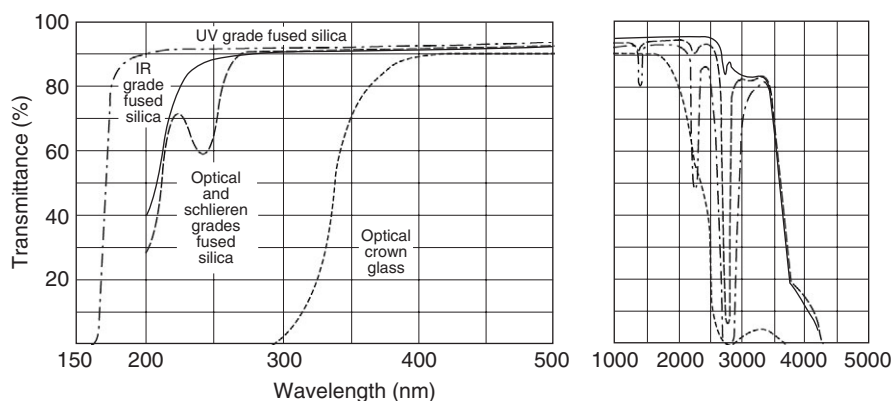
## Introduction

The physical chemical and optical properties of materials involved in optical spectroscopy are of crucial importance. The quality of these materials effectively determines the quality of the data derived from spectral measurements. The majority of measurements are carried out in solution and solvent quality is a major factor in obtaining accurate data. This section will cover the optical and physical properties of

## Transmitting Materials

Ideal quality considerations for transmitting materials include:

- homogeneity of refractive index;
- freedom from bubbles, inclusions, or striae;
- absence of birefringence;
- absence of fluorescence;
- wide wavelength range for high transmittance;
- very hard;
- durable to atmospheric attack particularly moisture;



**Figure 1** Typical transmittance curves for 10 mm samples of fused silica and optical glass. (Adapted from the Oriel catalogue (1993), volume 3.)



See also: **Optical Spectroscopy:** Radiation Sources; Wavelength Selection Devices; Detection Devices; Spectroscopic Materials; Refractometry and Reflectometry.

## Further Reading

- ASTM E 387–84 (1995) Standard Test Method for Stray Radiant Power Ratio of Spectrophotometers by the Opaque Filter Method. (This standard is likely to be revised in 2004.)
- Edisbury JR (1966) *Practical Hints on Absorption Spectrometry*. London: Adam Hilger.
- Fleming P (1999) In: Burgess C and Frost T (eds.) *Standards and Best Practice in Absorption Spectrometry*. Oxford: Blackwell Science.
- Kaye WI (1987) In: Burgess and Mielenz KD (eds.) *Advances in Standards and Methodology in Spectrophotometry*. Amsterdam.
- Neal WTL (1956) *Photoelectric Spectrometry Group Bulletin* 9: 204.
- Palmer EW, Hutley MC, Franks A, Verril JF, and Gale B (1975) Diffraction Gratings. *Reports on Progress in Physics* 38: 975–1048.
- Sharpe MR (1984) Stray Light in UV Spectrophotometers. *Analytical Chemistry* 56(2): 339A–356A.
- Slavin W (1963) Stray Light in Ultraviolet, Visible, and Near-infrared Spectrophotometry. *Analytical Chemistry* 35(4): 561–566.
- Tarrant AWS (1978) Optical techniques for studying stray light in spectrophotometers. *Optica Acta* 25: 1167.

## Spectroscopic Materials

**C Burgess**, Burgess Consultancy Co., Durham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 3647–3651, © 1995, Elsevier Ltd.

## Introduction

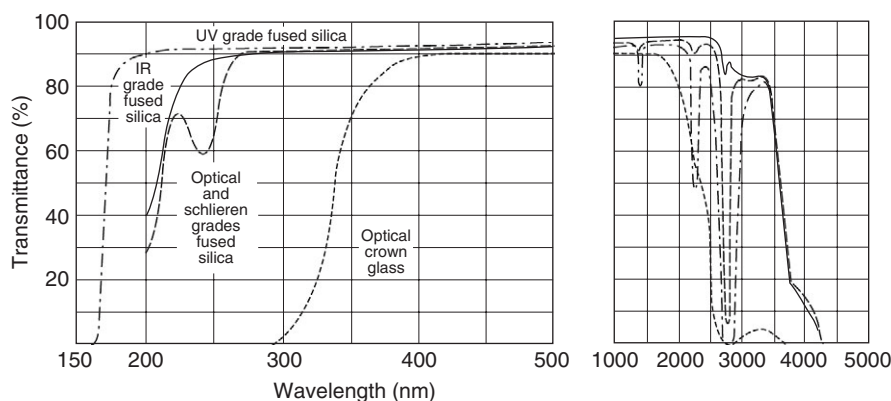
The physical chemical and optical properties of materials involved in optical spectroscopy are of crucial importance. The quality of these materials effectively determines the quality of the data derived from spectral measurements. The majority of measurements are carried out in solution and solvent quality is a major factor in obtaining accurate data. This section will cover the optical and physical properties of

materials used to construct optical components and some consideration of solvents for spectroscopy.

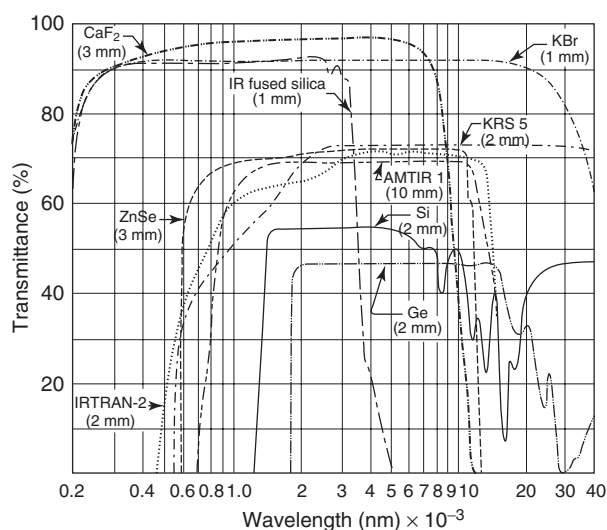
## Transmitting Materials

Ideal quality considerations for transmitting materials include:

- homogeneity of refractive index;
- freedom from bubbles, inclusions, or striae;
- absence of birefringence;
- absence of fluorescence;
- wide wavelength range for high transmittance;
- very hard;
- durable to atmospheric attack particularly moisture;



**Figure 1** Typical transmittance curves for 10 mm samples of fused silica and optical glass. (Adapted from the Oriel catalogue (1993), volume 3.)



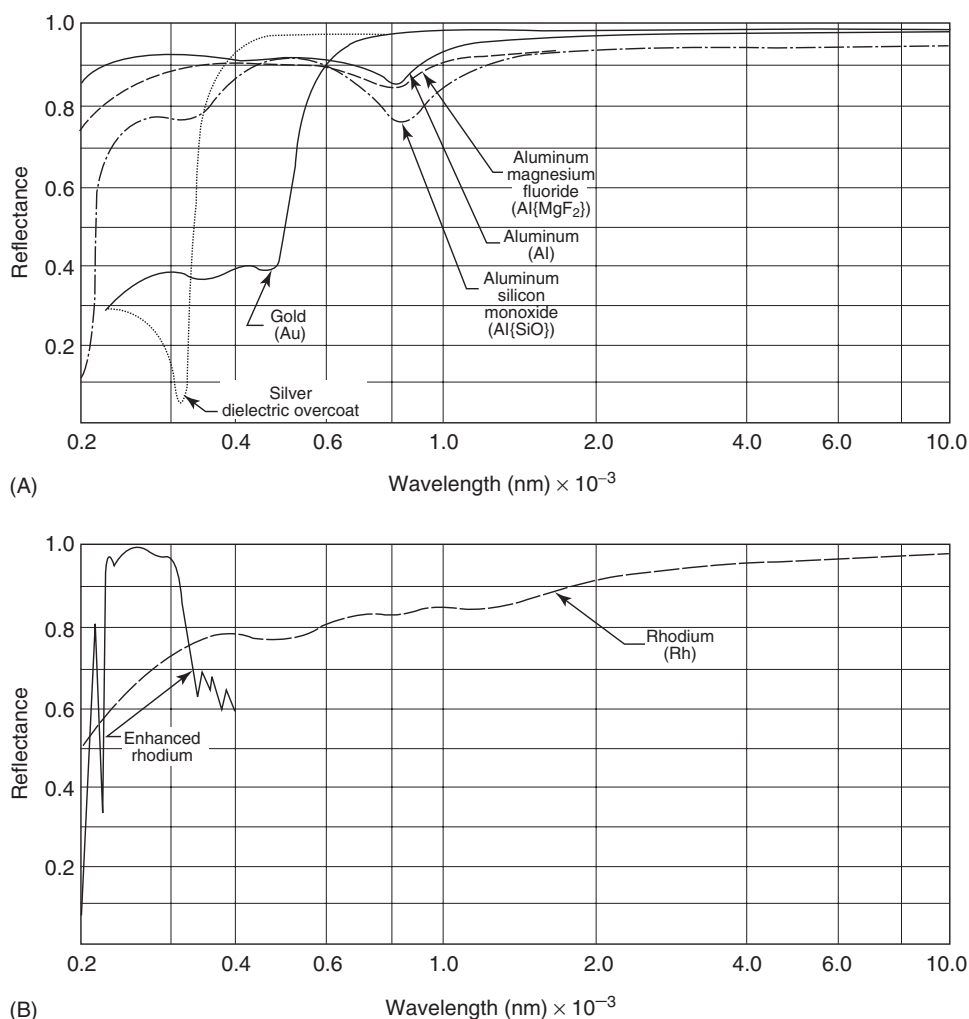
**Figure 2** Typical transmittance curves for materials commonly used in the infrared. (Adapted from the Oriel catalogue (1993), volume 2.)

- small temperature coefficients for all properties; and
- low cost and easily worked.

Not surprisingly, no material meets all these criteria.

Fused silica (UV grade) has excellent transmittance properties and is both durable and thermally stable. It finds uses as the material of choice for prisms, lenses, windows and cuvettes in the region 160–2000 nm. However, infrared grade fused silica may be used from 250 to 3500 nm. Optical crown glasses are more restricted in wavelength range from 300 to 2500 nm.

The optical transmittance properties of materials used in the ultraviolet through to the infrared are shown in **Figure 1**. A similar transmittance plot for optical materials used in the infrared is shown in **Figure 2**. However, optical properties are only one of the factors to be considered, particularly in the



**Figure 3** Reflectance characteristics of optical materials used for mirrors. (Adapted from the Oriel catalogue (1993), volume 2.)

**Table 1** Properties of commonly used infrared materials

<i>Material</i>	<i>Hardness (Knoop) (kg mm<sup>-2</sup>)</i>	<i>Density (g cm<sup>-3</sup>)</i>	<i>Softening point (°C)</i>	<i>Refractive index at midrange</i>	<i>Solubility in water</i>	<i>Birefringent</i>	<i><math>\frac{dn}{dT}</math> at 589 nm at 20°C</i>	<i>Useful wavelength range (nm)</i>	<i>Comments</i>
Potassium bromide	–	2.75	734	1.52	Very soluble	No	–	400–20 000	Wide transmittance range but soft and attacked by water
AMTIR-1	150	4.7	300 (maximum useful 200°C)	2.51	None	No	$7.5 \times 10^{-5}$ at 5000 nm	900–12 000	Amorphous glass material of Ge, As and Se. Good homogeneity
Calcium fluoride	–	3.18	1423	1.4	Small	No	– 9	200–8 000 21 000–38 000	Hard material slowly attacked by moisture
KRS-5				2.37	Small	No	–	600–35 000	Thallium bromiodide. Very wide transmission but soft material
Germanium	–	5.35	937	4	None	No	–	2000–16 000	Very durable but nontransmitting above 150°C
Silicon	–	2.33	1410	3.42	None	No	–	1300–11 000	Similar to germanium; nontransmitting above 350°C
IR fused silica	530	2.209	1585	1.43	None	No	10	2600–4 000	Durable material unaffected by water
Zinc selenide	120	5.27	1100	2.41	None	No	61 at 10 500 nm	550–16 000	Very expensive but used extensively in laser applications
Sapphire	1370	3.97	2040	1.70	None	Slightly	–14	150–5 000	Durable material unaffected by water
Zinc sulfide (IRTRAN 2)	160	4.09	1020	2.23	Very small	No	39 at 33 900 nm	400–12 000	Very soft material but better visible- region properties than zinc selenide
NaCl	18	2.16	1074	1.51	Small	No	$-3 \times 10^{-5}$	250–16 000	Similar to potassium bromide but less susceptible to water and extends to the UV

infrared. The ability to work the material and its durability are also key factors. Some of the factors are summarized for commonly used materials in Table 1. The materials are used in the construction of lenses, cuvettes, prisms and sample cell windows.

## Reflecting Materials

Mirrors were for many years primarily aluminum-coated glass. Recently, however, coatings of magnesium fluoride or silica onto aluminum have increased their durability and lasting reflectance properties. All reflectors deteriorate with time, atmospheric attack and radiation damage.

Very specialized coatings, e.g., rhodium, are very expensive but have excellent reflectance characteristics when optimized for the ultraviolet.

Above 2000 nm most of the materials show high reflectance. Figures 3A and 3B shows the reflectance characteristics of some commonly used materials.

Gold has such excellent properties above 800 nm that it is the material of choice for near-infrared and mid-infrared integrating spheres.

## Solvents

The quality of the spectroscopic solvent is sometimes overlooked. The analytical scientist needs to consider:

- the transmittance properties of the solvent;
- the transmittance properties of the sample of the solvent being used;
- the chemical stability of the solvent;
- the temperature coefficient at expansion and the volatility of the solvent; and
- health and safety aspects associated with its use.

As a rule of thumb, do not operate using a solvent with less than 50% transmittance at the analytical

**Table 2** Lower wavelength values for 50% transmittance of common spectroscopic grade solvents

<i>Solvent</i>	<i>λ at 50% T</i>
Hexafluoropropan-2-ol	< 190 <sup>a</sup>
Hexane	208
Heptane	209
Isooctane	210
Isopropanol	218
Ethanol	218
Methanol	210
Cyclohexane	220 <sup>b</sup>
Diethyl ether	232
Dioxan	235
Dichloromethane	237
Chloroform	250
Tetrahydrofuran	250
Dimethylformamide	273
Dimethylsulphoxide	282
Benzene	283
Pyridine	305
Acetone	350
Water	190 <sup>c</sup>

<sup>a</sup>Hexafluoropropan-2-ol has 90% transmittance at 200 nm.

<sup>b</sup>The removal of oxygen by nitrogen purging of hydrocarbon solvents decreases the transmittance markedly, e.g., for cyclohexane  $\lambda$  at 50%  $T$  decreases to 205 nm after purging.

<sup>c</sup>Doubly distilled water (glass apparatus) may be used down to 180 nm if purged of dissolved oxygen. However, contact with plastics and polyethylene dramatically reduces its transmittance below 200 nm. Perkampus HH (1992) *UV Atlas of Organic Compounds*, 2nd edn. Weinheim: VCH.

wavelength unless forced and then only if it can be ascertained that stray light is not a problem.

All solvents should be checked before use even if spectroscopic grade solvents are chosen. The storage of spectroscopic solvents needs particular care. Store in glass, in amber bottles, preferably under nitrogen for work of the highest quality.

Table 2 contains information relating to commonly used spectroscopic solvents.

## Refractometry and Reflectometry

**J Pawliszyn**, University of Waterloo, Waterloo, ON, Canada

**R E Hummel**, University of Florida, Gainesville, FL, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Refractometry

Refractometry is an analytical technique that characterizes structure and composition of materials by the velocity of light transmitted through them. The result of the measurement is usually expressed as the ratio of the velocity of light through a vacuum ( $c$ ) to that of the sample ( $v$ ), and is referred to as the refractive index:

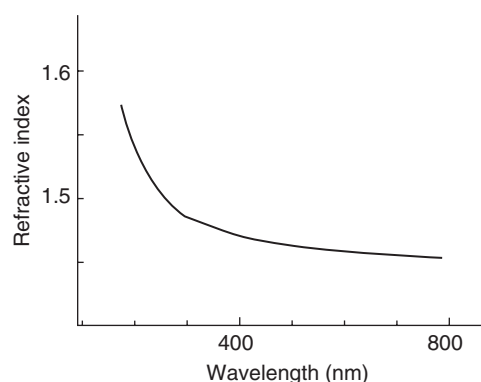
$$n = \frac{c}{v} \quad [1]$$

Table 1 shows the refractive indices for several materials. The variation of refractive index as a function of electromagnetic radiation wavelength (dispersion spectrum) provides simultaneous information about the polarizability and absorption properties of the sample. The rapid development of modern optical and imaging technologies such as lasers and photodiode arrays creates opportunities for application of refractometry to small analytical systems and multi-dimensional monitoring.

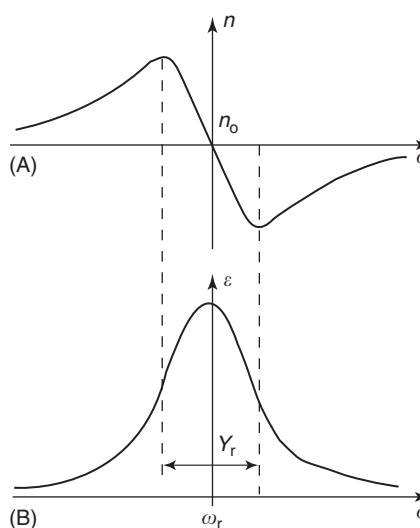
### Theory

The decrease of light velocity in the sample compared to vacuum is caused by the interaction of electromagnetic radiation with electron clouds present in the specimen and therefore corresponds to electron density and their state of binding. Light velocity and refractive index vary as a function of wavelength and generally increase with an increase in

frequency for nonabsorbing materials. Figure 1 shows the variation of refractive index for fused quartz. For absorbing species, the situation is more complex since significant variations in the refractive index occur near the absorption bands, as shown in Figure 2. This phenomenon is produced as the result of major changes in the charge configuration associated with electronic transitions. These changes will induce polarizability and therefore increase the refractivity of the absorbing species. With decreasing frequency near the absorption band the refractive



**Figure 1** The refractive index of fused quartz as a function of light wavelength.



**Figure 2** (A) Refractive index,  $n$ , and (B) dielectric constant,  $\epsilon = (n - ik)^2$ , related to absorption coefficient,  $k$ , around a resonance frequency,  $\omega_r$ , for a Lorentzian oscillator. Half-width of the absorption curve is  $Y_r$ .

**Table 1** Selected refractive indices for  $\lambda = 589$  nm (yellow sodium light), 1 atm, and 20°

Air	1.000293
Methanol	1.326
Water	1.333
Sucrose solution (5%)	1.340
Ethanol	1.359
Fused quartz	1.458
Zinc crown	1.539
Sodium chloride	1.544
Polystyrene	1.55
Carbon disulfide	1.628
Sapphire	1.77
Diamond	2.42

index first falls to a minimum, then increases through  $n_0$  at the resonance frequency to a maximum, and finally levels off to the initial value  $n_0$  (see **Figure 2A**). A dramatic decrease of refractive index near the absorption band is referred to as ‘anomalous dispersion’. Variation of the refractive index, as a function of electromagnetic wavelength, is called the dispersion spectrum. At low frequencies (at the radio and microwave ranges) the presence of ions and polar molecules in the sample affects the electromagnetic wave propagation through the media. However, at high frequencies, characteristic of visible light, only the polarizability of low mass valence electrons determines the refractive index of the sample. Near the resonance frequency another effect associated with loss of light intensity will occur due to absorption (see **Figure 2B**). This is expressed as an imaginary part of the dispersion spectra and corresponds to the absorption curve.

The degree to which an electromagnetic wave is slowed down upon entering a given medium depends upon the characteristics of the electronic environment it encounters. This is a function of the individual molecular electron clouds, as well as the number of molecules per unit volume,  $N$  (particle density). If the medium contains  $N$  molecules per unit volume, the magnitude of the charge distortion in the molecules by the electromagnetic field of the radiation is limited by their polarizability,  $\alpha$ , and the dielectric constant,  $\varepsilon$ , of the medium. The relationship between these parameters is expressed in the Clausius–Mosotti equation:

$$\frac{\varepsilon - 1}{\varepsilon + 2} = \frac{4}{3} \pi N \alpha \quad [2]$$

Moreover, according to Maxwell’s theory of the electromagnetic field:

$$\varepsilon = n_\infty^2 \quad [3]$$

where  $n_\infty$  is the refractive index corresponding to an infinitely long wavelength. Since  $N$  is proportional to the density,  $d$ , eqn [2] can be rewritten to yield the Lorentz–Lorenz expression for specific refraction,  $r$ :

$$r = \frac{n_\infty^2 - 1}{n_\infty^2 + 2} \frac{1}{d} \quad [4]$$

Specific refraction is a measure of the polarizability of the system involved, corrected for the variation in the number of molecules per unit volume. It refers to the volume of the electron clouds themselves, and does not include the empty spaces in the medium. Therefore, it is a measure of the electronic properties of the molecules. If the electronic

structures of the molecules are not changed due to the intermolecular forces, the specific refraction is independent of temperature, pressure, or state of aggregation.

Based on the above discussion, it is expected that the refractive index of a sample varies with temperature, pressure, and its composition. The decrease in refractive index with increasing temperature and decreasing pressure is primarily the result of the expansion of the media volume and decrease in density, causing the light beam to encounter fewer molecules per unit distance. These effects can be calculated by considering thermal expansion or compressibility parameters.

The refractive index for multicomponent mixtures is not an additive function. In the first approximation, the change in the refractive index of the medium,  $\Delta n$ , can be calculated by assuming that the densities of the medium and the solute are similar:

$$\Delta n = \frac{M_s}{m_M} (n_s - n_M) \Delta C \quad [5]$$

where  $M_s$  is the molecular weight of a given solute,  $C$  the solute concentration,  $m_M$  the mass of 1 l of medium with solute, while  $n_s$  and  $n_M$  are refractive indices of solute and medium, respectively. A more precise relationship can be derived from the Clausius–Mosotti relation (eqn [2]), which leads to the following relationship:

$$\Delta n = \frac{M_s C_s}{m_M K_M} (F_s - F_M) \quad [6]$$

where

$$F_s = (n_s^2 - 1)(n_s^2 + 2)$$

$$F_M = (n_M^2 - 1)(n_M^2 + 2)$$

$$K_M = 6n_M / (n_M^2 + 2)^2$$

## Principles of Refractometric Instrumentation

Differences in refractive indices among naturally occurring materials often result in refraction, that is, deflection of the light path. This optical phenomenon is frequently used to measure the refractive index or its changes. **Figure 3** illustrates the refraction of the light at the boundary between two materials with different refractive indices ( $n_1 > n_2$ ). The refraction of the light at the boundary can be explained by considering the light as a wave as shown in **Figure 3B**. As the wave crosses the interface it speeds up when it encounters the lower refractive index, resulting in a change in the direction of propagation. The



relationship between the angle of incidence  $\theta_1$  and the angle of propagation  $\theta_2$  can be derived from Figure 3B and is defined by Snell's Law:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \quad [7]$$

A portion of the light is also reflected from the surface (Figure 3A). The intensity of reflected light is dependent on its polarization and is defined by Fresnel's law of reflection. It is a function of  $n = n_2/n_1$

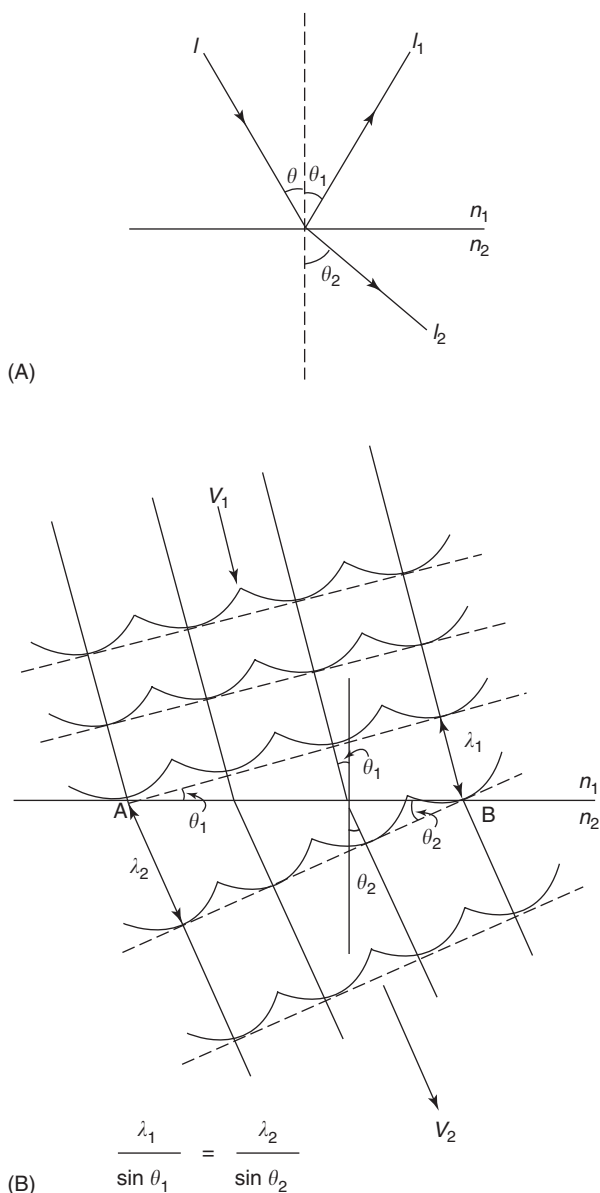
and angle of incidence  $\theta$ :

$$I_2(\parallel) = I \left( \frac{-n^2 \cos \theta + \sqrt{n^2 - \sin^2 \theta}}{n^2 \cos \theta + \sqrt{n^2 - \sin^2 \theta}} \right)^2 \quad [8]$$

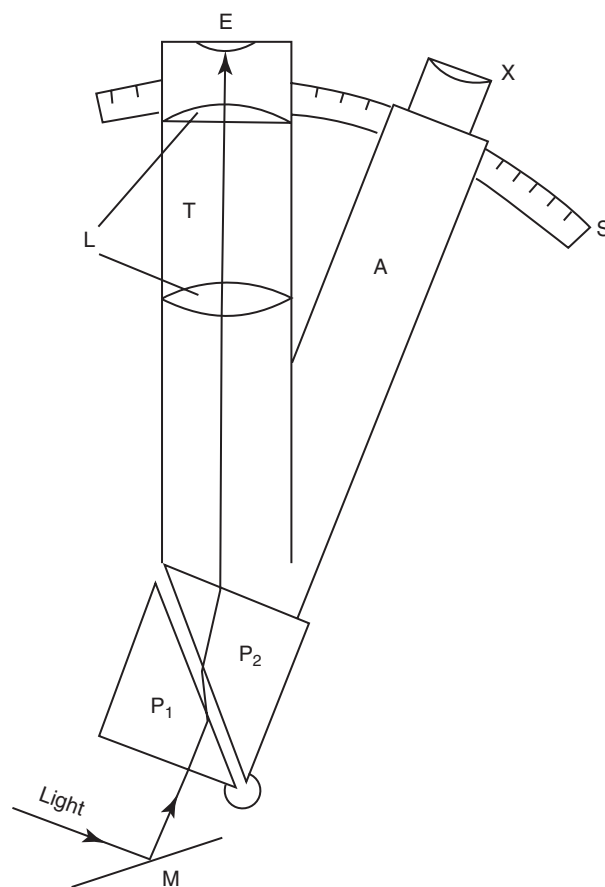
$$I_2(\perp) = I \left( \frac{\cos \theta - \sqrt{n^2 - \sin^2 \theta}}{n^2 \cos \theta + \sqrt{n^2 - \sin^2 \theta}} \right)^2 \quad [9]$$

where the  $I_2(\parallel)$  is the intensity of the light parallel to the surface and  $I_2(\perp)$  the intensity of reflected light polarized perpendicular to the surface.

For  $\sin \theta > n_2/n_1$ , total reflection of the light occurs at the boundary. This phenomenon is the principle of the popular Abbé refractometer (see Figure 4). A small amount of liquid sample is introduced between prisms  $P_1$  and  $P_2$ . Usually, the prisms are kept at constant temperature using thermostated water. The light beam illuminates the prisms from below with the help of the mirror  $M$ . The prisms  $P_1$  and  $P_2$  are attached to a rotatable arm,  $A$ , which carries a marker and magnifier to facilitate accurate readings,



**Figure 3** Graphic illustration of Snell's law of refraction. (A) Refracted and reflected light beam at the refractive index boundary for  $n_1 > n_2$ . (B) The refraction of the plane wave. Reflected wave is not shown.



**Figure 4** Schematic view of an Abbé refractometer. A, rotatable arm; S, sector scale;  $P_1$ , auxiliary prism;  $P_2$ , refracting prism; T, telescope; L, telescope lenses; E, eyepiece; X, magnifier; M, illumination mirror.

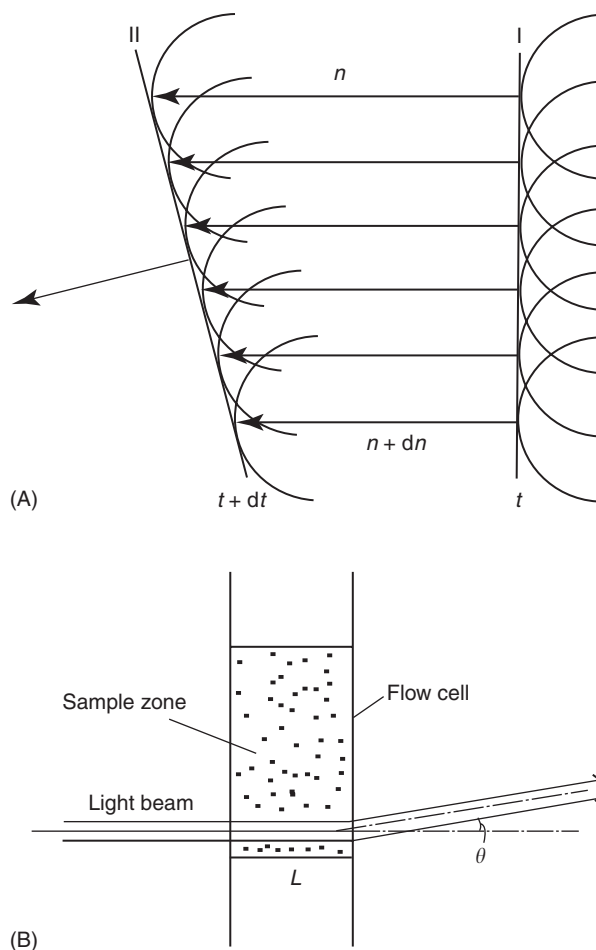
using a fixed scale S. The reading of the scale is a measure of the angle between normal to the surface and the telescope axis. The telescope, T, is equipped with cross hairs, which are aligned with the observed critical boundary. The resulting light emergence angle,  $r_c$ , corresponds to the characteristic of the sample and the prism. The refractive index of the sample,  $n$ , can be calculated using the following equation:

$$n = \sin r_c \cos \beta + \sin \beta (n_p^2 - \sin^2 r_c)^{1/2} \quad [10]$$

where  $\beta$  is the prism angle and  $n_p$  its refractive index. Typically, the instrument is designed to use white light, but readings can be made for a sodium line (589.3 nm) by applying optical filters based on Amici prisms. The standard Abbé refractometer allows the estimation of the refractive index to  $\pm 0.0001$ . There are several other types of refractometers based upon various optical phenomena such as interferometry, light diffraction, Snell's law of refraction, and Fresnel's law of reflection.

### Schlieren Optics

The refraction of light is also caused by a non-homogeneous refractive index within the sample. For example, the nonuniform distribution of the analyte in a medium will generate a corresponding refractive index gradient:  $dn/dx = (dn/dC)(dC/dx)$ . The light propagated through the medium with such a gradient will produce a 'streak' ('*schliere*' in German; '*mirage*' in French). Such an effect can be easily observed when dissolving sugar in water. A similar effect can be produced by temperature gradients, for example, above a hot surface ( $dn/dx = (dn/dT)(dT/dx)$ , where  $T$  is temperature), which is a principle of indirect absorption methods based on photothermal deflection techniques including thermal lens spectrometry. The physical principle of this phenomenon is illustrated in **Figure 5**. The wavefront, I, of the light beam encounters the refractive index gradient field at time  $t$ . The different sections of this wavefront experience different refractive indices. Therefore, various parts of the probe beam will propagate through the gradient in the medium with different velocities since the velocity is inversely proportional to the refractive index (eqn [1]). After a time increment,  $dt$ , the parts of the beam that experience a lower refractive index will have propagated further than those in other regions, resulting in a tilted wavefront II, formed at time  $t + dt$  with respect to wavefront I. The net effect is a deflection of the light beam toward higher refractive indices since light, similar to other types of waves, propagates perpendicular to its wavefront. The quantitative relationship



**Figure 5** Principle of Schlieren optics. (A) Tilting of the light wavefront propagating through the medium with a nonuniform refractive index. (B) Refraction of the light beam passing through the detector cell with the concentration gradient produced by solute.

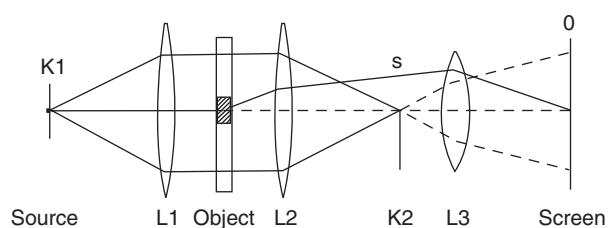
between the deflection angle,  $\theta$ , and the detector cell dimension,  $L$ , assuming a uniform refractive index gradient,  $dn/dx$ , normal to the probe beam direction (**Figure 5B**), can be derived using Fermat's principle; it states that the light path through a medium is such that the time necessary for its traversal is a minimum, which results in

$$\tan \theta = \sinh \frac{L}{n} \frac{dn}{dx} \quad [11]$$

where  $n$  is the refractive index of the medium. For most practical analytical cases, where  $L$  and  $\theta$  are small, we can approximate  $\theta$  by

$$\theta = \frac{L}{n} \frac{dn}{dx} \quad [12]$$

Many different optical systems have been developed to investigate refractive index inhomogeneities using the light deflection phenomenon described above. These methods are usually referred to as Schlieren



**Figure 6** Example of Toepler–Schlieren optical system. L1, collimated lens; L2, Schlieren head; L3, objective lens; K1, first knife edge (aperture); K2, second knife edge.

optical techniques. In these methods, the light beam is passed through an object with an inhomogeneous refractive index. After some optical signal processing, the beam is recorded on a photographic plate or more recently on photodiode array sensors. The change in the intensity distribution across the probe beam reflects the refractive index gradient fields present in the object. An example of such a system is shown in **Figure 6**. In this arrangement, the angle of deflection is transformed into a light intensity variation in the image of the object, on the photographic plate or screen. The region of high gradients is indicated by darker or lighter regions in the image. In the Toepler–Schlieren (**Figure 6**) system, an image of the light source is produced at ‘S’ and an image of the object is produced at point ‘O’. The knife edge K2 intercepts the image of the source, sharpened by aperture K1. The deflection of the beam produced by the upward directed gradient will cause the light to miss the K2 edge. If no gradient is present, or its orientation is downwards, then the appropriate light will be blocked. Therefore, the net effect will be that the lighter areas in the image of the object indicate regions in which gradients are directed upwards. The rest of the image will be dark.

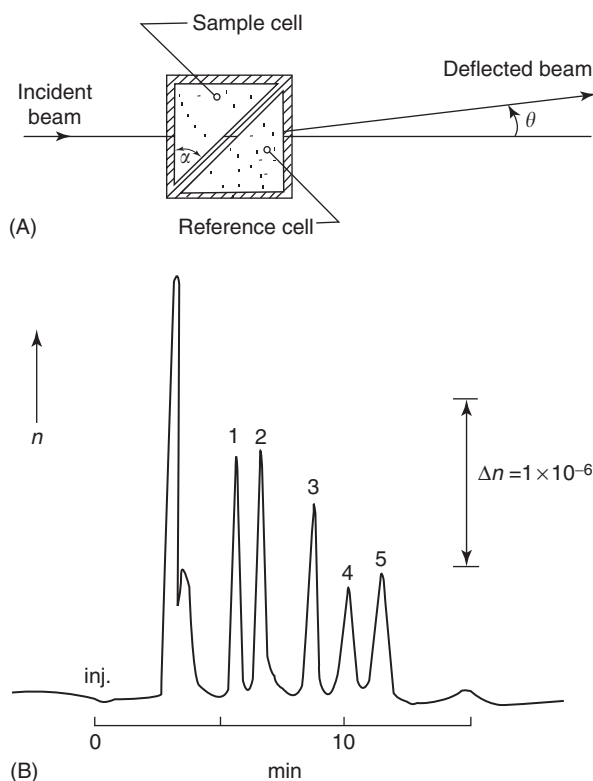
The magnitude and direction of the gradient can be studied by moving the knife edge deeper into the image of the light source. Starting from the non-intercepting position, the areas of the object with highest gradients directed downwards will be darkened first, followed by areas with no gradient, and then finally areas with gradients directed upwards. Scanning the knife edge allows only for the analysis of the gradients directed perpendicular to the edge. Simultaneous analysis in both directions can be accomplished by using a centrosymmetric aperture made, for example, of shaded or colored centrosymmetric rings. The magnitude of the deflection is indicated by the darkness or the color of the image. In addition, the direction of the deflection can be determined using an aperture made up of colored regions.

The velocity of light through the sample is not identical in all directions in anisotropic materials, where the polarizability of the electron clouds varies in relation to optical axis. Similarly, propagation of circularly polarized light can differ depending on its orientation if the sample contains optically active material. This effect can be monitored by measuring the rotation angle of the linearly polarized light. This phenomenon is the principle of the analytical technique called polarimetry.

## Applications

In principle, absolute values of the refractive index can be used to identify chemical species in the same fashion as boiling and melting points, and the density of light absorption properties. Specific refraction and dispersion spectra are used to characterize the chemical structure of analytes. However, it is very difficult to apply refractometry for this purpose because of small differences in refractive index between substances, and relatively large dependence of the index with temperature, dissolved gas, and impurities. This method has been used more successfully in combination with microscopy to differentiate between components of inhomogeneous solid samples.

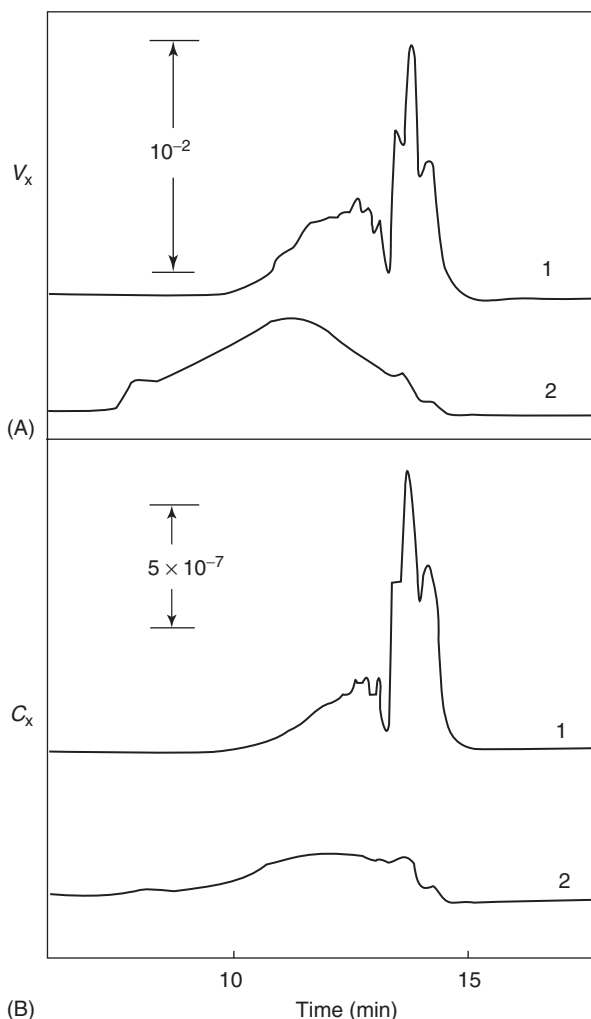
A more important application of this method is to quantify the composition of well-defined binary systems in which pure components differ appreciably in refractive index and an approximately linear relationship exists between  $n$  and the concentration of analyte. An example is the determination of protein or sugar content in aqueous solutions. One of the most important applications of refractometry is the measurement of refractive index in flowing liquids, particularly in high-resolution liquid chromatography. The refractive index measurement is done in a difference mode as shown in **Figure 7A** for the beam deflection type of the detector. In a typical design, two adjacent triangular cross-section cells are used (**Figure 7A**). When the refractive index of the effluent in one cell is the same as in the reference cell, which contains a mobile phase, the probing light is not deflected. As the composition of the liquid changes in the first cell, because of elution of sample components, the refractive index of the mixture changes, resulting in light deflection. The major advantage of the refractive index detector over other methods is its universal response. Usually, the solute differs from the eluting solvent in refractive index by  $\sim 0.1$  units. This enables the detection of analytes at only semi-trace ( $\text{mg l}^{-1}$ ) levels. However, in some cases this is the most sensitive detector available. For example, in the analysis of sugars, absorption or fluorescence



**Figure 7** Refractive index detection in liquid chromatography. (A) Dual-triangular section flow cell for detector based on Snell's law. (B) Separation and detection of carbohydrates with LC and refractive index detection. Sample: 4  $\mu\text{g}$  each of (1) xylose, (2) fructose, (3) sucrose, (4) maltose hydrate, and (5) lactose. (Reproduced with permission from Munk M (1993) *Refractive index detection*. In: Parriott D (ed.) *A Practical Guide to HPLC Detection*. San Diego: Academic Press.)

detection cannot be used effectively. **Figure 7B** shows an example of the separation and detection of a sugar mixture.

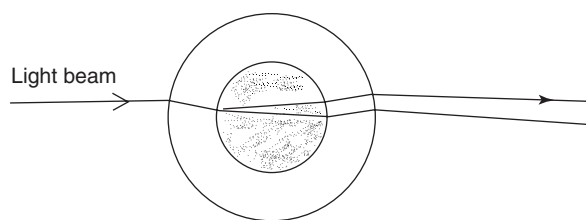
The refractive index detector also plays an important role in the characterization of polymeric mixtures and natural materials such as crude oils after exclusion chromatographic analysis. The interesting feature of this method is its ability to quantify multicomponent samples without standards. As shown in eqn [5], the magnitude of the signal for the two unknown quantities is determined by their concentrations and refractive indices. However, in the case where the same sample is eluted by different solvents, characterized by different refractive indices but having identical elution characteristics to the first one, different peak magnitudes will be observed. The two signals give rise to two independent equations, which can be solved to determine both the concentration and the refractive index of the analyte. **Figure 8** illustrates the separation and quantification of components present in two types of crude oil.



**Figure 8** Concentration of materials in crude oils separated by size-exclusion chromatography. (A) Volume fraction eluted each second; (B) moles eluted each second. 1, Arun crude oil; 2, North Slope crude oil. (Reprinted with permission Yeung ES (1986) *Detectors for Liquid Chromatography*. New York: Wiley; © John Wiley & Sons, Inc.)

### Trends in the Use of Refractometry

**Miniaturization** Application of lasers as light sources resulted in a reduction in the size of refractive index detectors and made them compatible with in-line monitoring systems that are needed in capillary separation technology. The refractive index of minute amounts of solution may be determined by the measurement of the diffraction phenomenon produced in the interaction between the laser beam and a fluid-filled capillary. The laser beam passing through an off-center capillary produces a fan of scattered light in the plane perpendicular to the tube axis. The position of a single line corresponds to the refractive index of the fluid within the tube



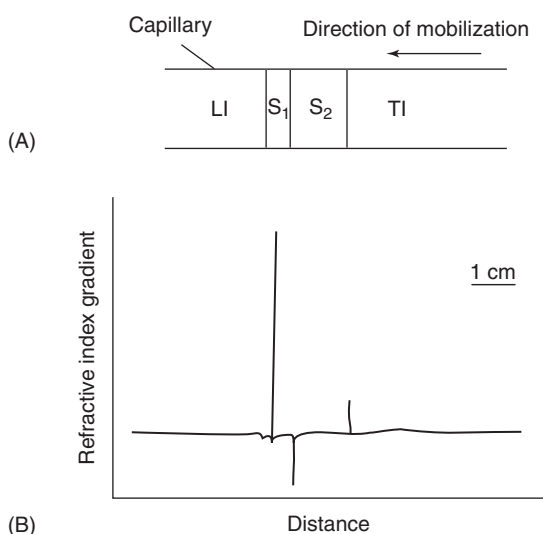
**Figure 9** Refraction of a light ray by a fluid-filled tube. (Reprinted with permission from Bornhop DJ and Dovichi NJ (1986) *Analytical Chemistry* 58: 504–505; © American Chemical Society.)

(Figure 9). With micrometer diameter tubes, such a refractive index detector can probe volumes approaching a few picoliters and result in femtogram detection limits. This feature makes this detection scheme very useful in small channel separation techniques, such as capillary and chip technology.

**Concentration gradient detection** In many modern analytical methods, such as electrochemical and high efficiency separation techniques, high concentration gradients are generated, which produce a highly inhomogeneous refractive index field in the system and can be effectively detected by the Schlieren optics methods. These investigations have traditionally been used only for qualitative observations. However, the development of new optical technology such as lasers and imaging systems facilitate quantitative applications.

**Figure 10** illustrates the application of a Schlieren optics detector for capillary isotachopheresis (cITF). In cITF, ionic analytes are separated into pure distinct zones that are stacked according to their mobility. The zones then migrate with constant velocity toward the common electrode (Figure 10A). In this technique, the sample is introduced between leading and tailing electrolytes that have higher and lower mobilities, respectively, than the components of the analytical mixture. The boundary between the zones formed in this technique is very narrow due to the sharpening effect. Also, the zones contain high concentrations of analytes because of the concentration effect. These properties of the method ensure the formation of high concentration gradients at the zone boundaries that can be readily detected by the gradient methods (Figure 10B). The high spatial resolution and universal response of the laser-based Schlieren optics method (see Figure 5B) allows sensitive and quantitative detection of all zones produced during electrophoretic separation.

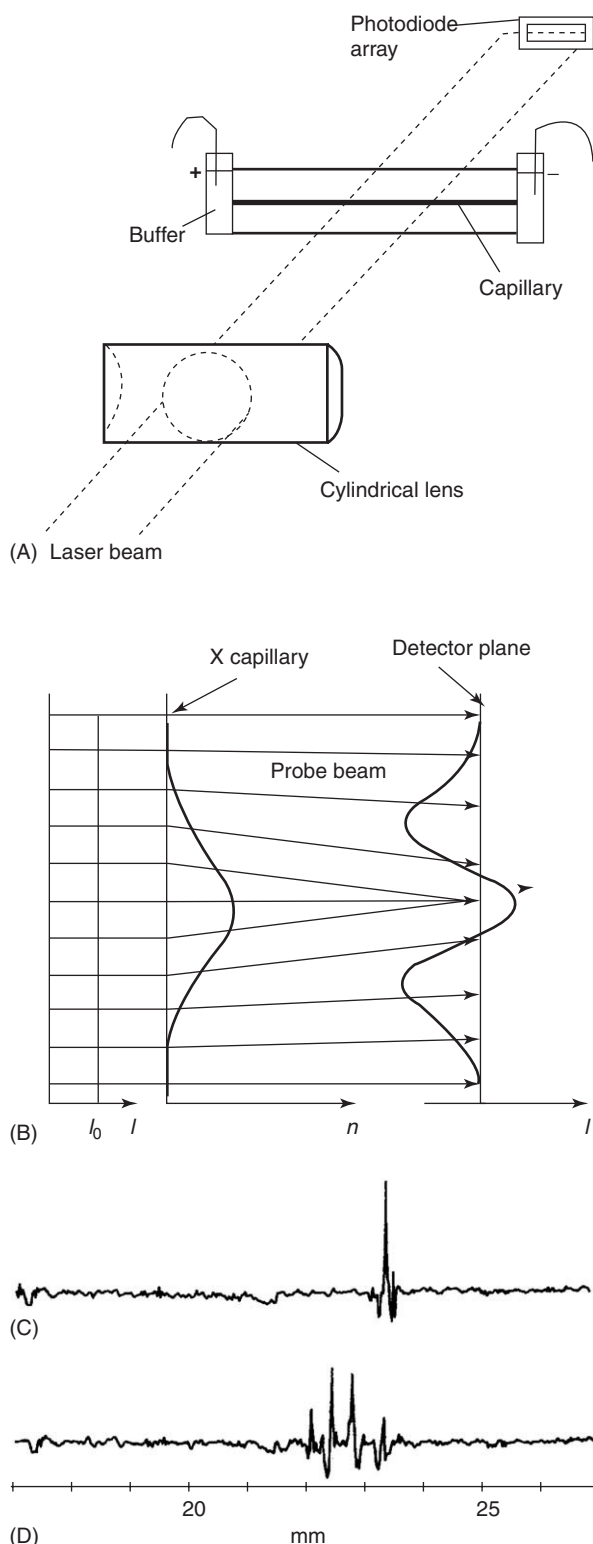
Another capillary electrophoretic method that possesses both concentrating and focusing properties is capillary isoelectric focusing. In this technique, a pH gradient is formed inside the capillary using synthetic



**Figure 10** (A) Capillary isotachopheretic separation and (B) concentration gradient detection of two cations. Sample: S<sub>1</sub>, hexadecylammonium and S<sub>2</sub>, triphenylphosphonium ions. Leading ion (LI) is ammonium ion and tailing ion (TI) is tetraoctylammonium ion. (Reprinted with permission from McDonnell T and Pawliszyn J (1991) *Analytical Chemistry* 63: 1884–1889; © American Chemical Society.)

carrier ampholytes. The ampholytic samples, such as proteins, are concentrated at their isoelectric points inside the capillaries. The analytes are then mobilized toward one end of the capillary for detection. The analysis time can be substantially reduced (from ~30 to 2 min) by eliminating the mobilization step when applying imaging methods of detection. In the Schlieren shadowgraph system shown in Figure 11, the laser beam is expanded and focused into the separation capillary using a cylindrical lens in between (Figure 11A). The diode array detects changes in the light intensity distribution caused by the presence of the analyte bands. For the optical alignment shown in Figure 11A, in theoretical treatment, the probe beam can be decomposed into a bundle of infinitesimal light filaments. The individual light filaments are refracted and deflected from their original path upon encountering a refractive index gradient. This gradient is produced by the concentration gradients generated inside the capillary by concentrated analytes. As suggested in Figure 11B, the relative changes of the probe beam intensity on the detector plane are proportional to the second derivative of the refractive index inside the capillary. Figures 11C and 11D show an application of this method to separation and quantification of protein samples and the investigation of interaction between biological species. Schlieren imaging is the best possible choice for *in vitro* monitoring of formation and dissociation rates of biological complexes since it detects all





**Figure 11** Concentration gradient imaging detection in capillary isoelectric focusing. (A) Experimental arrangement and (B) principle of the imaging system. (C) Images of focused transferrin (iron poor form) and (D) after reaction with iron(III). (Reprinted with permission from Wu J and Pawliszyn J (1992) *Analytical Chemistry* 64: 224–227 and 2934–2941; © American Chemical Society.)

species present in the capillary and does not disrupt the system by introducing energy through absorption processes. Toepler–Schlieren optical arrangement (see Figure 6) can be used in place of the shadowgraph system (Figure 11) to enhance sensitivity of the imaging detection.

## Reflectometry and Differential Reflectometry

Refractometry and reflectometry both use visible (or near-visible) light to investigate a given material. Refractometry is generally utilized to study transparent materials such as glasses, transparent polymers, or a number of chemical species described in the above sections. On the other hand, dense materials such as metals, alloys, semiconductors, etc., are generally studied in reflection because of the short penetration depth of light into these materials. The ratio between the reflected intensity  $I_R$  and the impinging intensity  $I_0$  of the light serves as a definition for the reflectivity (see Figure 3A):

$$R = \frac{I_R}{I_0} \quad [13]$$

The spectral reflectivity of dense materials is generally a complicated function of the index of refraction (see eqn [1]) and a parameter  $k$  that describes the amount of damping of the light within the material. (The damping constant  $k$  is often neglected in glasses, etc., because of its extremely small value for these materials.)

As occurs in the refraction ‘spectrum’ shown in Figure 1, the reflectivity spectrum is often relatively featureless. Thus, a first derivative technique is generally applied which enhances the ‘structure’ in a conventional reflection spectrum. This enhancement technique, which allows one to emphasize small pieces of information in a relatively featureless reflection spectrum, is called differential reflection spectrometry. Refractometry and differential reflection spectrometry complement each other for their respective fields of applications.

## Principles of Reflectometric Instrumentation

The differential reflection spectrometer (also called the differential reflectometer) measures the normalized difference between the reflectivities of two specimens (or two slightly different parts of a specimen) as a function of photon energy. Unpolarized, monochromatic light (having the possibility to vary the wavelength continuously) is alternately deflected to one or the other sample by means of a vibrating



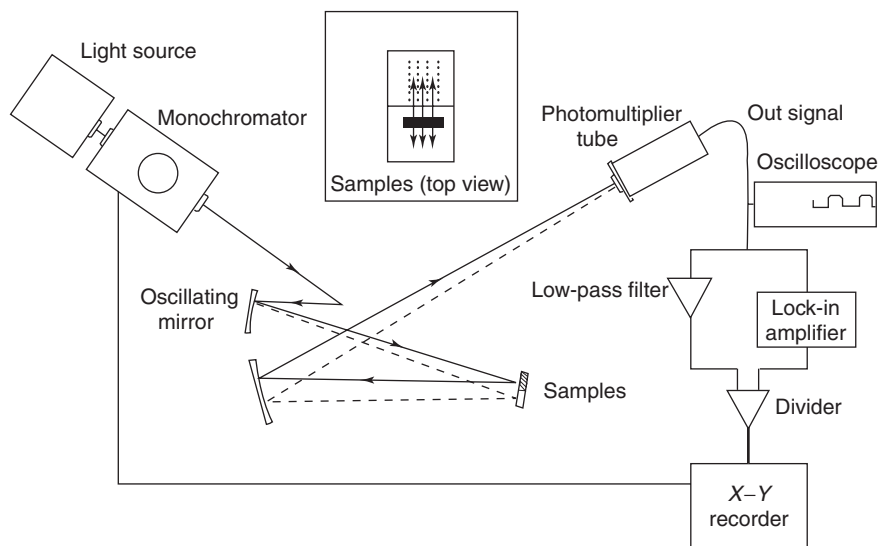
mirror that oscillates at a convenient frequency such as 50 or 60 Hz (Figure 12). The cross-section of the light beam is rectangular and is parallel to the boundary that separates the two samples. This beam sweeps over the two specimens perpendicular to the boundary and at near normal incidence to the surface (see the insert of Figure 12). The total area scanned is generally  $2 \times 4 \text{ mm}^2$ , but can be varied in size by regulating the voltage to the vibrator coils driving the mirrors. The light that is reflected by the two samples eventually reaches a light-sensitive device. The signal thus obtained is electronically processed to yield  $\Delta R/\bar{R}$ , where  $\Delta R = R_1 - R_2$ , is the difference in reflectivity between the two specimens and  $\bar{R} = (R_1 + R_2)/2$  is the average reflectivity. Measuring  $R_1$  and  $R_2$  at the same time and forming the ratio between  $\Delta R$  and  $\bar{R}$  eliminates possible influences from fluctuations of the line voltage. It also eradicates intensity variations in the spectral output of the light source, the spectral sensitivity of the detector, and the spectral reflectivities of the mirrors. Further, owing to the difference-forming technique, all disturbances of the surface (such as roughness, oxidation, contamination) are eliminated as long as they are common to both sample parts.

The output signal from the divider circuit of the differential reflectometer is fed to the Y-axis of an X-Y recorder (Figure 12). A potentiometer attached to the scanning gear of the monochromator delivers a d.c. voltage proportional to the wavelength that is fed to the X-axis of the X-Y recorder. In this way, a differential reflectogram, i.e., a continuous plot of  $\Delta R/\bar{R}$  versus wavelength (or photon energy), is

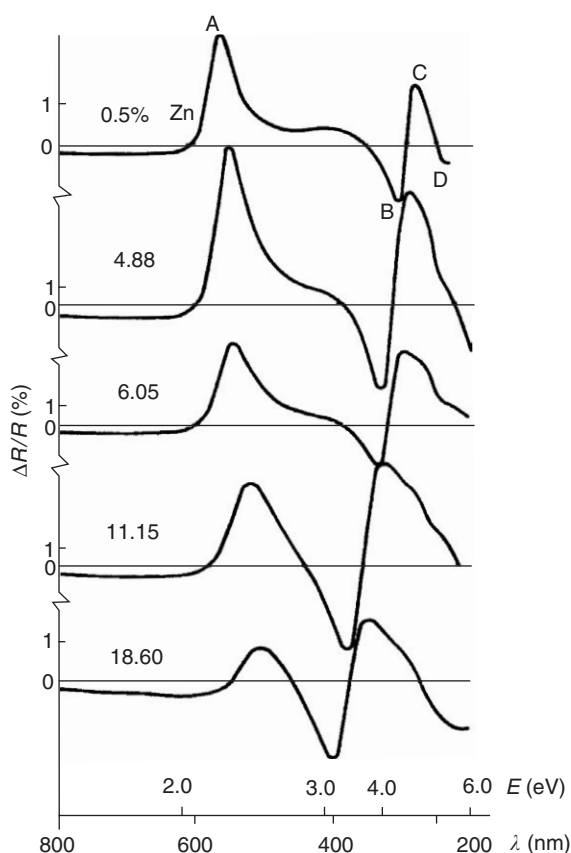
generated automatically without necessitating individual measurement and computational stops at narrowly spaced wavelength increments, as required by spectroscopic ellipsometry. A scan between 200 and 800 nm (1.6–6.0 eV) requires from 1 to 3 min, depending on the scanning speed selected. The sensitivity in observing the normalized difference in reflectivity is better than one-hundredth of a percent.

### Applications

The differential reflectometer is successfully used in a number of research fields such as the study of the electronic structure of alloys, order-disorder transformations, investigations of surface films, corrosion, or the effects of ion implantation in semiconducting materials. The differential reflectometer is also helpful when the nature and the thickness of one or several active layers in an electronic device need to be quickly and nondestructively determined. The peaks in the ultraviolet (UV) region of differential reflectograms indicate whether or not a change in the electronic structure of the top layer ( $> 1 \text{ nm}$ ) of the host material (caused by doping, lattice damage, chemical reactions, etc.) has taken place. Further, the UV peaks provide information concerning whether or not a disturbed layer is submerged below an essentially unperturbed area. On the other hand, the interference peaks in the infrared region yield the thickness of a specific layer. Further, differential reflectometry is well suited for a quick and nondestructive check of semiconductor wafers, stemming from different vendors or lots, and for monitoring



**Figure 12** Schematic diagram of the differential reflection spectrometer. For clarity, the angles of the incident and reflected light beams on the sample are shown larger than they are in reality.



**Figure 13** Differential reflectograms for various copper-zinc alloys. The information is obtained from a surface layer of approximately 100 nm thickness. The parameter on the curves is the average zinc concentration of the two alloys given in at.%.

possible deviations from a reference wafer that has been shown to have required properties.

The usefulness of differential reflection measurements will be illustrated here on a characteristic sample. Copper is a reddish metal which changes its color toward yellow when zinc is added to it (to form brass). This change in color, which has its origin in an increase in the average number of electrons when bivalent zinc is added to monovalent copper, can be followed in a series of differential reflectograms.

Figure 13 depicts the normalized difference in reflectivity between various copper-zinc alloys as a function of the wavelength of the light (or the photon energy). A number of peaks designated A–D can be observed. The peak marked ‘A’, which is situated in the red part of the spectrum, relates to the color of this alloy. It is observed in Figure 13 that with increasing zinc content peak A moves to higher photon energies, that is, from red to yellow. In other words, differential reflectometry can provide nondestructively and rapidly a qualitative account of the chemical composition of solid materials.

See also: **Carbohydrates:** Overview. **Electrophoresis:** Isotachopheresis; Isoelectric Focusing. **Laser-Based Techniques.** **Liquid Chromatography:** Overview; Size Exclusion.

## Further Reading

- Bottcher CJF (1973) *Theory of Electric Polarisation*. New York: Elsevier.
- Hummel RE (1983) Differential reflectometry and its application to the study of alloys, ordering, corrosion, and surface properties. *Physica Status Solidi (a)* 76: 11–44.
- Levin SZ and Bauer N (1965) Refractometry and dispersometry. In: Kolthoff IM and Elving PJ (eds.) *Treatise on Analytical Chemistry*, part 1, vol. 6. New York: Wiley-Interscience.
- Mao Q and Pawliszyn J (1999) Capillary isoelectric focusing with whole column imaging detection for analysis of proteins and peptides. *Journal of Biochemical and Biophysical Methods* 39: 93–110.
- Niemczyk TM (1978) Refractive index measurement. In: Kuwana T (ed.) *Physical Methods in Modern Chemical Analysis*, vol. 2. New York: Academic Press.
- Pawliszyn J (1990) Concentration gradient detection based on Schlieren optics. *Spectrochimica Acta Reviews* 13: 311–354.
- Sorensen HS, Pranov H, Larsen NB, Bornhop DJ, and Andersen PE (2003) Absolute refractive index determination by microinterferometric backscatter detection. *Analytical Chemistry* 75: 1946–1953.
- Vasil’ev LA (1971) *Schlieren Methods*. New York: Keter Inc. (Israel Program for Scientific Translations Ltd.).

# OSCILLOMETRY

See CONDUCTIMETRY AND OSCILLOMETRY

## OSMOSIS, REVERSE

See MEMBRANE TECHNIQUES: Dialysis and Reverse Osmosis

## OUTDOOR AIR

See AIR ANALYSIS: Outdoor Air

## OXYGEN FLASK COMBUSTION

See SAMPLE DISSOLUTION FOR ELEMENTAL ANALYSIS: Oxygen Flask Combustion

## OZONE

**K Takeuchi**, National Institute for Resources and Environment, Tsukuba, Ibraki, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 3689–3697, © 1995, Elsevier Ltd.

### Origin and Nature

Ozone ( $O_3$ ) is a colorless gas at ambient temperature and pressure and has a characteristic odor even at very low concentrations. It has a strong oxidizing ability that is hazardous to plants and animals, and is unstable especially at higher concentrations. Electric discharge in oxygen (or air) and irradiation of oxygen (air) with short-wavelength ultraviolet (UV) rays produce ozone.

Ozone plays important roles in atmospheric chemistry and physics. Up to  $5 \times 10^{12}$  molecules  $cm^{-3}$  or 4–5 ppmv ( $10^{-6}$  mixing ratio by volume) of ozone is found in the stratosphere where the solar radiation intensity of shorter wavelengths is high. Ozone in the stratosphere has made it possible for animals to live on land by preventing harmful 200–300 nm UV radiation from reaching the Earth's surface. The absorption of UV radiation by ozone, however, heats the air, disturbing air convection and creating a layered structure in the stratosphere. As some useful and stable chemicals like chlorofluorocarbons (Freons) and other substances were found to decompose in the stratospheric environment and to destroy the

ozone layer by chain reactions, ozone levels in the stratosphere have been monitored with great concern.

In the lower troposphere, photochemical reactions of air pollutants (hydrocarbons and nitrogen oxides) produce ozone as well as other atmospheric oxidants such as hydrogen peroxide ( $H_2O_2$ ), organic peroxides, and peroxyacyl nitrates that result in what is called photochemical smog. Daytime ozone concentrations often exceed 0.1 ppmv in polluted air. Ozone and other oxidants promote oxidation of sulfur dioxide ( $SO_2$ ) and nitrogen oxides into sulfuric and nitric acids, respectively, bringing about the occurrence of acid rain. In addition, ozone has strong absorption bands in the infrared region near  $9.6 \mu m$ , and hence is one of the so-called greenhouse gases. Monitoring of ozone concentration is thus indispensable for protection of both the local and global environment.

Ozone is also emitted from germicide lamps, copy machines, printers, welding, and other industrial processes. The working environment criteria for ozone are set at 0.05–0.1 ppmv in most countries. Monitoring of ozone is therefore important from the viewpoint of workplace health and hygiene.

Ozone is being used in the field of water treatment, disinfection, decolorization, deodorization, organic synthesis, materials testing, dry etching and cleaning processes in semiconductor industries, and other industrial areas. The advantages of using ozone over other chemicals are: its strong oxidizing power, its

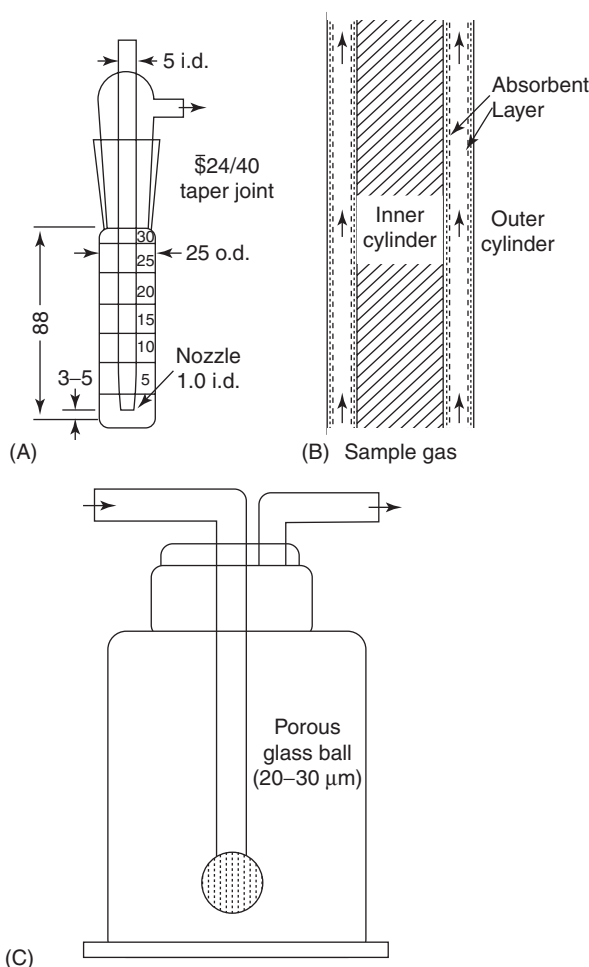
clean nature leaving only oxygen after the treatment, and electrical generation at the site. Monitoring of ozone dose and residual levels is necessary in all these processes.

## Sampling and Calibration

Since ozone is very reactive and easily decomposes on contact with other substances, use of inert (and clean) materials such as fluoropolymers (e.g., polytetrafluoroethene (PTFE)) or glass is mandatory for sampling tubes and other parts in contact with the sample gas and water. Sample flow rate and tube diameter should be large enough to minimize decomposition loss of ozone. Particulate matter in the sample should be removed with a PTFE filter. For applications of a high ozone concentration, durable material such as stainless steel should be used.

For atmospheric ozone measurements, the sampling site should be carefully chosen to avoid any influence from the ground, plants, trees, highways, etc. Use of thermal mass-flow controllers is recommended for sampling at reduced and changing pressures (i.e., aircraft measurements). In manual wet analyses, glass midjet impingers (30 ml) are often used to strip ozone from sample air as shown in **Figure 1A**, while other types of air scrubbers with a sintered-glass diffuser have been devised. Annular denuders (**Figure 1B**), in which sample gas is passed between the narrow gap made by two concentric glass cylinders where an absorbing agent is coated, can strip ozone by diffusion without being affected by particulate matter in the sample. Air samples are generally sampled with a flow system and are analyzed at the site. For batch water sampling, the sample water should gently fill and be tightly sealed in a sampling bottle and be analyzed as quickly as possible to prevent decomposition and evaporation of ozone from the sample. Glass bottles with a sintered-glass ball are used for aeration and reabsorption of ozone as shown in **Figure 1C**. Aqueous ozone can be stabilized by cooling and maintaining pH values below 4.

While some analytical systems based on absolute methods can be calibrated statically without ozone (i.e., with standard iodine solution for iodometry), dynamic calibration using a standard ozone sample is necessary in most cases. Since ozone cannot be stored in a cylinder, an ozone generator (UV or discharge type), which is frequently standardized by iodimetry or another method, is used. Discharge-type ozone generators may produce small amounts of nitrogen oxides when air is used as a source gas. Use of purified oxygen is recommended for contamination-free and efficient ozone generation.



**Figure 1** Typical sampling apparatus for determination of ozone: (A) midjet impinger (dimensions in mm), (B) annular denuder (cross-sectional view), and (C) gas scrubbing/aeration bottle (500 or 1000 ml).

## Gaseous Ozone

Analytical methods for ozone must meet various requirements and a wide range of concentrations:  $10^{-3}$ – $10^0$  ppmv for environmental monitoring,  $10^{-2}$ – $10^2$  ppmv for biological exposure and testing, and  $10^1$ – $10^5$  ppmv for industrial applications. Automatic and dry methods are preferred rather than manual and wet methods, because of the requirements of continuous and unattended monitoring. Such instruments can provide reliable data as long as they are used and maintained properly. As there are few perfect analytical methods for ozone, an appropriate method should be chosen with a good understanding of its performance and limitations. The characteristics of major analytical methods for gaseous ozone are summarized in **Table 1**.

**Table 1** Analytical methods for gaseous ozone determination<sup>a</sup>

<i>Method</i>	<i>Analytical characteristics</i>	<i>Reagent (principle)</i>	<i>Sampling rate (1 min<sup>-1</sup>)</i>	<i>Response time<sup>b</sup> (s)</i>	<i>Detection limit (ppbv)</i>	<i>RSD<sup>c</sup> (%)</i>	<i>Remarks/interference</i>
Iodimetry (neutral buffered; photometry)	Wet, batch	1% KI in 0.2 mol l <sup>-1</sup> phosphate (pH 7.0); 352 nm	1–3	(1800)	10	2	ASTM D2912; JIS B7957; WHO. Not specific to O <sub>3</sub> . Not always accurate/ SO <sub>2</sub> , etc.
CL (gas phase)	Dry, continuous	Ethene (25 ml min <sup>-1</sup> )	1	60	1	2	EPA; JIS; WHO
CL (gas phase)	Dry, continuous	NO (168 ml min <sup>-1</sup> )	3.7	0.05–0.08	0.1	n.a.	For flux sensor
CL (aqueous phase)	Wet, continuous	10 mg l <sup>-1</sup> IDS in 2 mmol l <sup>-1</sup> phosphate (pH 7.2)	0.015	10–100	0.4	3.3	Higher sensitivity with higher sample flow rate/SO <sub>2</sub>
CL (aqueous phase)	Wet, continuous	1 g l <sup>-1</sup> eosin Y in ethylene glycol	1–7	0.14–0.7	0.2	n.a.	For aircraft measurement
UV photometry	Dry, continuous	(254 nm)	0.5–1.5	20–120	1–2	1	EPA; JIS
IR (FTIR)	Dry, continuous	(1055 cm <sup>-1</sup> )	19 800	< 300	10	n.a.	900 m path, 0.5 cm <sup>-1</sup> resolution
IR (TDLAS)	Dry, continuous	(1050 cm <sup>-1</sup> )	n.a.	60	0.3	n.a.	100 m path
Spectrophotometry	Wet, batch	10 mg l <sup>-1</sup> IDS in 0.1 mol l <sup>-1</sup> phosphate (pH 6.8)	0.4–2.0	(600–3600)	4	< 5	610 nm/NO <sub>2</sub> (+ 6%)
Semiconductor sensor	Dry, continuous	(adsorption)	0.3	< 180	10	30	Portable, battery operation
Detector tube	Dry, batch	Indigo	0.1	(30–360)	50	15	Disposable, with no electric power/NO <sub>2</sub> , Cl <sub>2</sub>

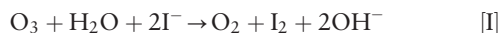
<sup>a</sup>Key: CL, chemiluminescence; UV, ultraviolet; IR, infrared; FTIR, Fourier-transform infrared spectroscopy; TDLAS, tunable diode laser absorption spectroscopy; IDS, indigo-5,5'-disulfonate; ASTM, American Society for Testing and Materials; EPA, US Environmental Protection Agency approved methods; JIS, Japanese Industrial Standard; WHO, World Health Organization selected methods; n.a., not available.

<sup>b</sup>For batch methods, sampling time (s) is shown in parentheses.

<sup>c</sup>Precision indicated as relative standard deviation.

## Iodimetry

One of the most obvious techniques is based upon iodimetry because it has been used as a standard or reference method in spite of its drawbacks. Ozone oxidizes iodide ion in solution:



The triiodide ion formed is determined by spectrophotometry (352 nm), volumetric titration (against sodium thiosulfate standard solution and starch as an indicator), or coulometry.

It should be noted that the reaction is not specific to ozone but proceeds with  $\text{H}_2\text{O}_2$ , nitrogen dioxide ( $\text{NO}_2$ ), peroxyacyl nitrates, and other oxidizing species, with different relative responses (this may be convenient in some cases, i.e., for measurement of the total or photochemical oxidant concentration in air). Furthermore, the ozone to iodine stoichiometry of the reaction deviates from unity depending on the pH of the solution, iodide concentration, etc., and the iodine formed may evaporate from the solution. For atmospheric applications, a neutral-buffered potassium iodide method (1% KI in  $0.2 \text{ mol l}^{-1}$  phosphate buffer,  $\text{pH } 7.0 \pm 0.2$ ) is often used amongst many variants to minimize the above instabilities. The typical detection limit is 10 ppbv (ppbv =  $10^{-9}$  mixing ratio by volume) with an air sample of 30–60 l (sampling at  $1\text{--}2 \text{ l min}^{-1}$  for 30 min). Automated instruments for continuous air monitoring are available with a countercurrent ozone absorber.

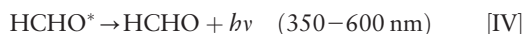
Sulfur dioxide in air dissolves in the stripping solution and gives a negative error. A chromium(VI) oxide-impregnated filter can remove  $\text{SO}_2$  from sample air. The filter is prepared as follows: a  $400 \text{ cm}^2$  portion of a glass-fiber filter is dipped in a mixture of purified water (15 ml), concentrated sulfuric acid (0.7 ml), and  $\text{CrO}_3$  (2.5 g), and is dried at  $80\text{--}90^\circ\text{C}$  for 1 h. This filter also oxidizes nitric oxide (NO) into  $\text{NO}_2$  to some extent. Although the interference from  $\text{NO}_2$  is not serious, corrections for  $\text{NO}_2$  and NO are necessary to obtain a photochemical oxidant concentration (= total oxidant –  $\text{NO}_2$ ). Nitrogen dioxide and NO give 3–5% response of equivalent amount of ozone. The relative response should be determined experimentally.

## Chemiluminescence

In chemiluminescence (CL), excess energy of a certain chemical reaction involving the species to be determined is released as light. The chemical reaction makes the CL method selective. Unlike other

photometric techniques, no incident light is necessary to obtain CL signals. This enables a high baseline stability to be attained, and hence a very low detection limit, especially for gas-phase CL methods in which background light emission is negligible.

Ozone selectively attacks olefinic double bonds in organic compounds (ozonolysis reactions). When reacted with ethene (ethylene) (Nederbragt's method), ozone gives an electronically excited methanol (formaldehyde) that emits visible light:



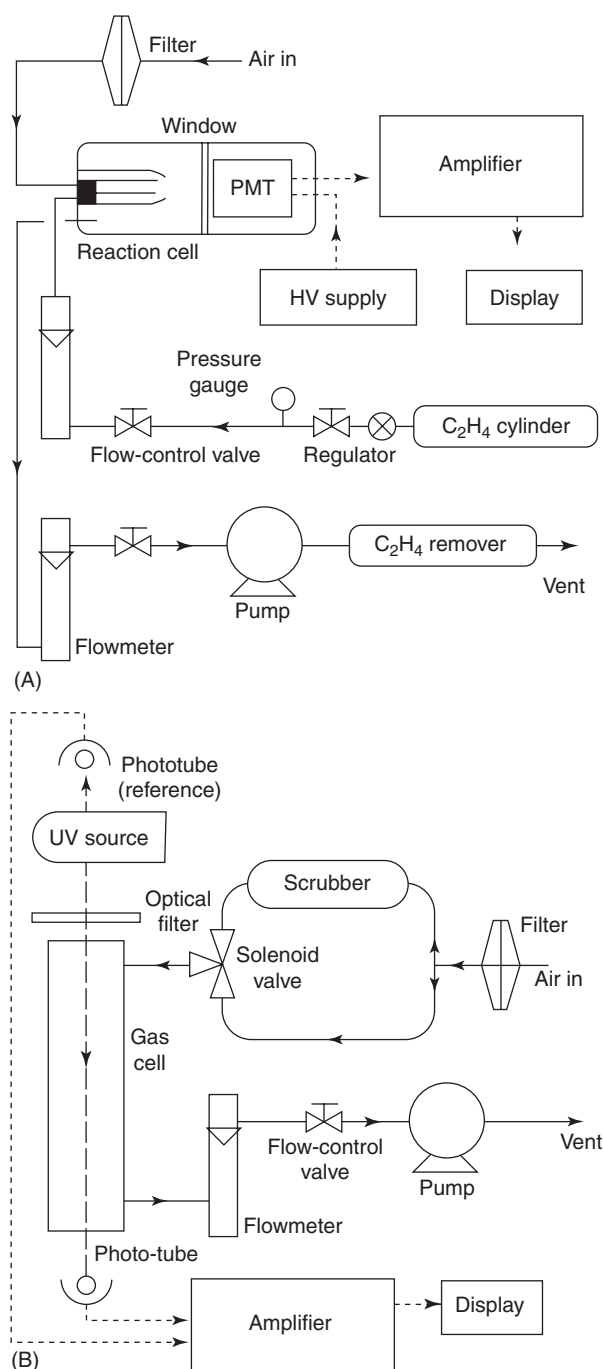
The CL (maximum at 450 nm) is conveniently detected with a photomultiplier tube (PMT). As shown in **Figure 2A**, a commercially available analyzer consists of a reaction chamber, flow control valves, and flowmeters for sample air and ethene. A coaxial double tube is used as a mixing jet in the reaction chamber. Sample air and ethene flow out from the inner tube at  $0.3\text{--}1.5 \text{ l min}^{-1}$  and the outer tube at  $20\text{--}30 \text{ ml min}^{-1}$ , respectively. Ethene is supplied from a high-pressure cylinder and unreacted ethene is removed with a catalytic converter to prevent it from polluting the environment. The ethene CL method has good selectivity and fast response, and is convenient for continuous air monitoring. However, fluctuations in the sample flow rate directly affect the measurement and care must be taken in the handling of ethene. Some instruments use ethene at a lower concentration than the flammability limit (<2.75%) to prevent possible accidents.

Another gas-phase CL method is that utilizing a reaction with NO:



This is carried out by operating a chemiluminescent  $\text{NO}_x$  analyzer (cf.  $\text{NO}_x$  analysis) at a NO excess condition. As the reaction of ozone with NO is faster and gives more intense CL than that with ethene, this method is preferred for some purposes in which a fast response is required (i.e., aircraft flux measurements). However, this CL method is subject to interference from water vapor and may contaminate the environment with NO. The ozone–NO reaction is also known as a gas-phase titration reaction, in which sample gas is mixed with a low-concentration standard NO gas and the  $\text{NO}_2$  formed in quantities equivalent to ozone is determined by an appropriate method. While CL reactions of ozone with other alkenes, alkyl sulfides, phosphine, arsine, and stibine





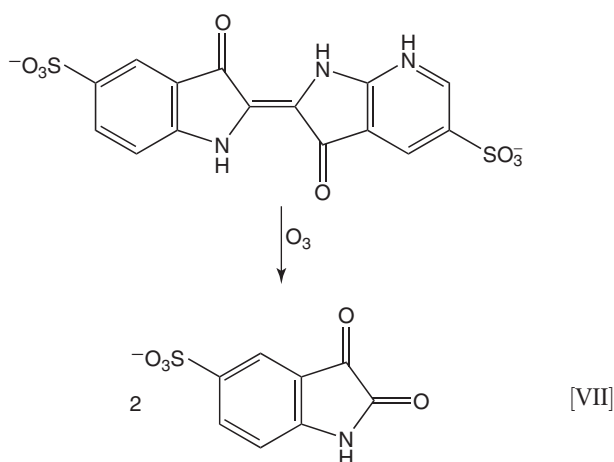
**Figure 2** Schematic diagrams of (A) chemiluminescent ozone analyzer and (B) UV photometric ozone analyzer.

have been reported, none of them have been utilized as analytical methods to date.

There are some liquid-phase CL methods using dye compounds as a reagent. Ozone can be measured either by passing sample air into an ethanolic solution of rhodamine B and gallic acid or using a rhodamine B impregnated silica disk. CL methods with eosin Y and with coumarin 47 are also reported for rapid

response. So far these methods have been used for research purposes only because of insufficient selectivity and long-term stability. The mechanism involved in ozone-dye CL is not yet clear.

Some ozonolysis reactions that take place in the liquid phase also give CL. An aqueous solution of indigo-5,5'-disulfonate (IDS, indigo carmine), which is used in the spectrophotometric determination of ozone as described below, emits light at 430 nm on contact with ozone reaction [VII]. This method is more selective than other liquid-phase CL methods. Some reducing gases like  $SO_2$  interfere, but that is tolerable for usual atmospheric measurements. The reagent solution can be recycled for long time measurements. The IDS-CL method gives a sensitivity three orders of magnitude higher than the IDS spectrophotometric method on the basis of sample amounts of equal portions.



### Direct UV Spectrophotometry

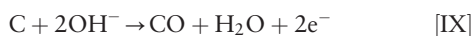
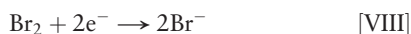
Ozone exhibits a strong absorption band centered at  $\sim 255$  nm (Hartley band). The maximum absorption coefficient (base 10) is  $135 \text{ cm}^{-1}$ . This absorption is conveniently measured with a low-pressure mercury lamp that has an intense emission line at 253.7 nm. Using a mercury lamp, bandpass filter, gas cell 40–60 cm long, and phototube, detection of ozone down to 1 ppbv is possible. Another phototube equipped in the analyzer corrects the instability and aging of the light source. Commercially available instruments (Figure 2B) avoid interferences from other UV-absorbing species by generating reference (zero) gas internally: a solenoid valve switches the sample to a metal oxide catalyst layer at a certain time interval and the catalyst selectively decomposes ozone in the sample gas. However, ppmv levels of aromatic hydrocarbons give positive responses to the instrument.

The UV photometric ozone monitor is relatively maintenance free except for periodic calibration by iodimetry and changing of the inlet filter and catalyst. Hence, UV photometry is suitable for continuous monitoring of atmospheric ozone and has been adopted as a standard method along with the ethene-CL method in major counters. UV ozone monitors with higher concentration ranges (up to 10 vol%) are also available. In high-concentration ozone measurements, ozone-free reference should be prepared separately for calibration, because of insufficient activity of the catalyst for ozone destruction. For a humid air sample, a water trap (or humidifier) is necessary.

### Electrochemical Methods

Although ozone at high concentrations can be determined by direct electrochemical reduction, reactions [I] and [II] ( $\text{Br}^-$  in place of  $\text{I}^-$ ) are generally employed to increase sensitivity and to eliminate interference from oxygen. Electrochemical methods are sensitive enough for atmospheric ozone and can be operated continuously. However, these methods have disadvantages similar to iodimetry: they are not specific to ozone and are subject to interference from other oxidizing and reducing substances. Coulometric methods are described here, although a number of potentiometric and amperometric methods have been reported.

A galvanic cell-type (Hersch cell) coulometer consists of a platinum (Pt) cathode, KBr (or KI) electrolyte (stripping solution), and active carbon anode. When air containing ozone is bubbled into the solution, the bromine (or iodine) liberated by the ozone is reduced at the Pt cathode and hydroxide ion discharges at the carbon anode:



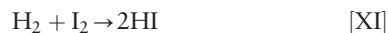
The current between the electrodes directly represents the concentration of ozone in the solution. Commercial instruments have internal  $\text{Br}_2$ -generating electrodes for calibration without ozone.

In Brewer–Milford-type coulometry, a potential of 0.2–0.3 V (lower than the decomposition potential of iodide solution) is applied between two Pt electrodes. Polarization current is observed until the surface of the cathode is completely covered with hydrogen:



When iodine is formed in the solution by reaction with ozone, some of the hydrogen film on the

electrode is removed as:



Then, the current again from reaction [X] is observed, which corresponds to the ozone concentration. The solution can be circulated after passing an adsorbent column. The air-monitoring data obtained with these coulometric ozone analyzers do not always coincide with those with iodimetry–photometric detection analyzers. This may be attributed to the difference in responses to other oxidants such as  $\text{NO}_2$ .

### Infrared Spectrophotometry

Infrared spectrophotometry of ozone based on the principal absorption band near  $9.5\text{ }\mu\text{m}$  is relatively free from interference by the bands of other atmospheric constituents. However, a long optical path is necessary for detection of atmospheric ozone. A White cell (multiple reflection cell, 10–1000 m path) combined with a Fourier-transform infrared (FTIR) spectrometer (spectral resolution of  $1\text{ cm}^{-1}$  or better) with a HgCdTe detector is often used in multi-component air-monitoring and smog chamber experiments.

Recently, very sensitive and selective measurements became possible by tunable diode laser absorption spectroscopy (TDLAS). Diode lasers that lase in the mid-infrared region give extremely high resolution ( $3 \times 10^{-4}\text{ cm}^{-1}$ ) and can tune the emission line to one of many vibration–rotation bands by changing the laser temperature and current (10–100 K, 0.1–2.0 A). The TDLAS measurement is usually carried out at reduced pressure to avoid band broadening due to molecular collision. Practically interference-free measurements are possible with a typical detection limit of sub-ppbv levels (100 m path, at 25 Torr), although the TDLAS system is still expensive and under development.

### Spectrophotometric Methods

Indigo-5,5'-disulfonate (or trisulfonate) reacts with ozone as described in the CL section. The blue color of IDS disappears on reaction with ozone and the decrease in absorbance at 610 nm is measured. As the molar absorptivity ( $\sim 2 \times 10^4\text{ mol l}^{-1}\text{ cm}^{-1}$ ) depends on the brand and batch of commercially available IDS probably due to impurity, the apparent reaction stoichiometry of IDS to ozone is not exactly unity. However, the stoichiometric ratio is reproducible when the same IDS reagent is used throughout. Nitrogen dioxide interferes positively by 6% of the mass of  $\text{NO}_2$ , but  $\text{SO}_2$ ,  $\text{NO}$ ,  $\text{HF}$ ,  $\text{H}_2\text{S}$ ,  $\text{H}_2\text{O}_2$ , and peroxyacetyl nitrate do not interfere. Sampling time

required can be estimated during sampling by noting decolorization.

In annular denuder sampling, potassium iodide (20 g KI in 10 ml of deionized water + 10 ml of 10% glycerol in methanol), phenoxazine (0.4 g in 20 ml of acetone), and 4-allyl-2-methoxyphenol (eugenol, 10% in methanol) are used as absorbents. The third absorbent releases formaldehyde by ozonolysis, which is collected with another downstream denuder coated with 2,4-dinitrophenylhydrazine (0.5% in acetonitrile containing 1% phosphoric acid).

### Thermometry

A large amount of heat is released when ozone is decomposed into oxygen ( $\Delta H = 143 \text{ kJ mol}^{-1}$ ). The sensing device has two thermistors that form a part of a bridge circuit and some catalyst (e.g., hopcalite, a mixture of manganese and copper oxides) coated on one of the thermistors to decompose ozone. When sample air containing ozone is passed through the cell, ozone heats catalyst-coated thermistor, and thus a signal related to ozone concentration is obtained. Devices for both atmospheric and industrial ozone have been designed.

### Sensors

There is some confusion in defining the term 'sensor', and only small semiconductor sensors are described here. Membrane electrode-type analyzers are discussed in the Aqueous Ozone section. Semiconductor ozone sensors based on field effect transistor (FET) are now commercially available. In this sensor, the gate of a FET is replaced with n-type semiconductor metal oxide (e.g.,  $\text{SnO}_2$ ) thin film. When electrophilic ozone is adsorbed on the film, an electric field is induced in the device, resulting in a change in the current between the source and drain terminals. The device is maintained at a high temperature so that water vapor and other substances do not affect the measurement. Sensitivity and selectivity can be optimized by changing the film thickness, the dopant in the film as the donor and its concentration, operating temperature, etc. The characteristics of the sensor can be precisely controlled by semiconductor production technologies. Moreover, producing different sensors on a tip, it is possible to provide a multisensing device. This sensor is small but has fairly good sensitivity and accuracy for quick measurements.

### Other Methods

In the early stage of air pollution studies, rubber cracking due to ozone was utilized for semiquantitative determination of ozone. Manometric methods based on the pressure increase after catalytic decomposition

as described in thermometry are also proposed for high ozone concentrations.

Redox reactions with ozone other than iodimetry are reported with different detection methods. Iron(II) reacts with ozone stoichiometrically in an acidic medium (pH 2.0). The resultant iron(III) forms a red complex with potassium thiocyanate, or the excess iron(II) is determined with permanganate, 1,10-phenanthroline derivatives, etc. Arsenic(III), manganese(II), and copper(II) can also be used as reagents for ozone determination. Analytical methods based on ozonolysis reactions with 1,2-bis(4-pyridyl)ethene, 1,1-diphenylethene, indigo, dimethoxystilbene, and diacetyldihydrolutidine have been reported. Other organic reagents for ozone include phenolphthalein, *N*-phenyl-2-naphthylamine, aromatic isocyanates, diphenylaminesulfonate, and *o*-tolidine.

Disposable detector tubes are available for simple and easy detection of ozone. It is based on the indigo-ozone decolorization reaction. In the glass tube is an adsorbent layer containing indigo. Ozone is determined from the length of the layer decolorized. Sampling is done with a special hand pump. Tubes for several concentration ranges are supplied: 0.05–3, 2.5–100, and 25–1000 ppmv, and 0.02–4  $\text{mg l}^{-1}$  (aqueous ozone after purging dissolved ozone).

## Remote Sensing

With a growing interest in the global atmospheric environment, remote sensing techniques are regarded as a most powerful and efficient way of monitoring stratospheric ozone. Some of the methods described can be applied to tropospheric ozone measurements as well.

### Dobson Spectrophotometer

This instrument is a differential spectrophotometer observing two closely separated wavelengths with different absorption coefficients for ozone in order to eliminate the scattering effect and interference from other atmospheric constituents. When the sun is used as a light source, the Dobson spectrophotometer gives the total ozone concentration in that air column. One of the following pair of wavelengths (nm) is selected from the structured Huggins bands: (325.4, 305.5), (329.1, 308.8), (332.4, 311.4), and (339.8, 317.8). Results are presented by the Dobson unit (equivalent to  $10^{-3} \text{ cm}$  thickness of pure ozone at standard temperature and pressure). Accuracy is generally better than 2% (as a standard error).

### Differential Absorption Lidar

In the differential absorption lidar (DIAL) technique, the average gas concentration over some range is determined by monitoring the laser backscatter light intensity for wavelengths tuned on and off an absorption peak of the gas of interest. The backscattered light is received with a telescope and PMT. The concentration can be determined from the ratio of the signals at the on/off wavelengths.

In NASA's airborne DIAL system, for example, these wavelengths are tuned to 286 and 300 nm for ozone ranging using frequency-doubled tunable dye lasers pumped with neodymium: yttrium–aluminum–garnet lasers or excimer lasers. Light of longer wavelength is also transmitted to correct the data for aerosol distribution. Typical precision and accuracy of the *in situ* measurements are 2 and 10%, respectively. Airborne DIAL systems can operate in both nadir and zenith modes with a vertical resolution of 50–200 m through the entire troposphere. They have been applied to the observations of the antarctic ozone hole, and other areas. An infrared DIAL system for ozone has also been developed using transverse excited atmospheric carbon dioxide lasers.

### Instruments Carried by Satellites

The solar backscatter UV instrument used by NASA called TOMS (total ozone mapping spectrometer) on the Nimbus-7 satellite provides total ozone maps by nadir viewing. The striking pictures of the Antarctic ozone hole have proved its power. An earth-limb scanning UV–visible grating spectrometer, SAGE (stratospheric aerosol and gas experiment) also operated by the NASA, observes the sun to the side through the atmosphere. This instrument can measure vertical profiles of ozone and other species with 1 km resolution and an accuracy of 5%. Instruments based on FTIR are also being developed for satellite measurements.

## Aqueous Ozone

Most of the analytical methods for gaseous ozone can be applied to aqueous ozone. Wet analytical methods readily accept water samples. After aeration of water samples, aqueous ozone can be determined either by dry methods or by wet methods that are subject to interference from coexisting solutes. However, the transmission efficiency of ozone from one solution to another must be considered. Analytical methods for aqueous ozone are summarized in Table 2.

### Redox Titration

Iodometry is also used as a standard method for aqueous ozone. Ozone in a water sample (800 ml) is transferred to a 400-ml portion of 2% KI absorbing solution by aeration with  $0.2\text{--}1.01\text{ min}^{-1}$  purified air (or  $\text{N}_2$ ) stream. The solution is acidified with  $0.5\text{ mol l}^{-1}$  sulfuric acid (20 ml) and is then titrated with  $2.5\text{ mmol l}^{-1}$  standard sodium thiosulfate solution to a starch endpoint. A blank test is necessary for accurate measurements.

In addition to iodometry, an arsenite solution (pH 6.5–7.0) is used as titrant for determination of aqueous ozone. The excess arsenic(III) is back-titrated with a standard iodine solution. This method is reported to be more selective for ozone than iodometry.

### UV–Visible Spectrophotometry

Direct measurement of UV absorbance is applicable to aqueous ozone. Unfortunately, the molar absorptivity at 260 nm has not yet been exactly established:  $3290\text{ mol l}^{-1}\text{ cm}^{-1}$  is considered to be reasonable out of the reported values ranging from 2900 to  $3600\text{ mol l}^{-1}\text{ cm}^{-1}$ . For real aqueous ozone samples, special care should be taken to avoid interferences from coexisting substances. Background absorbance can be determined after destruction of the residual ozone with an excess of nitrite or butenol.

Indigosulfonates are used as reagents for spectrophotometric determination of ozone. The decrease in absorbance at 610 or 600 nm after mixing with the aqueous sample is measured with disodium indigo-5,5'-disulfonate (IDS) or tripotassium indigo-5,5',7-trisulfonate (ITS). The residual absorbance of the mixture (at pH 2.0) is stable over several hours. The indigo method is not subject to interference from chlorite and chlorate ions and  $\text{H}_2\text{O}_2$ , while chlorine, bromine, iodine, and permanganate interfere to some extent. The interference from chlorine can be masked with malonic acid ( $0.5\text{ g l}^{-1}$ ).

A flow injection analysis based on the indigo method with a gas diffusion cell has been reported for interference-free automatic measurement. A microporous PTFE membrane is used as a septum in the cell. Silicone membranes with higher permeability give irreproducible results. When the aqueous sample is switched to one compartment of the gas diffusion cell, the reagent solution containing ITS remains in the other compartment to give higher sensitivity. After this sampling period (10–30 s), the reagent solution is sent to a spectrophotometer. The detection limit and sample throughput are  $30\text{ ng ml}^{-1}$  and  $65\text{ h}^{-1}$ , respectively. It can be said that the gas diffusion cell performs the aeration–reabsorption procedure on a microscale.

**Table 2** Analytical methods for aqueous ozone determination<sup>a</sup>

<i>Method</i>	<i>Analytical characteristics</i>	<i>Reagent (principle)</i>	<i>Sample size or flow rate</i>	<i>Response time<sup>b</sup> (s)</i>	<i>Det. limit (ng ml<sup>-1</sup>)</i>	<i>RSD<sup>c</sup> (%)</i>	<i>Remarks/interference</i>
Iodometry (aeration–reabsorption; titration)	Batch	2% KI; 0.5 mol l <sup>-1</sup> H <sub>2</sub> SO <sub>4</sub> ; 2.5 mmol l <sup>-1</sup> Na <sub>2</sub> S <sub>2</sub> O <sub>3</sub>	800 ml	(> 300)	30	1	Standard method (Am Public Health Assoc.), blank test necessary
Spectrophotometry (600 nm)	Batch	1 mmol l <sup>-1</sup> ITS (IDS) in 0.05 mol l <sup>-1</sup> phosphate (pH 2)	1–80 ml	–	5	2	Stable residual color/ I <sub>2</sub> , Br <sub>2</sub> , Cl <sub>2</sub> ; Cl <sub>2</sub> can be masked with malonic acid
Flow injection–spectrophotometry	Continuous	0.13 mmol l <sup>-1</sup> ITS in 1 mol l <sup>-1</sup> H <sub>3</sub> PO <sub>4</sub> ; gas diffusion cell	1.3 ml min <sup>-1</sup>	55	30	2	Diffusion cell remarkably reduced interferences/Cl <sub>2</sub> (+ 0.8%)
Spectrophotometry (548 nm)	Batch	0.2 mmol l <sup>-1</sup> ACVK in 1 mol l <sup>-1</sup> NH <sub>3</sub> buffer (pH 8.3)	180 ml	–	20	n.a.	Less sensitive than indigo method/ClO <sub>2</sub>
CL (aqueous phase)	Continuous	10 mg l <sup>-1</sup> IDS in 2 mmol l <sup>-1</sup> phosphate (pH 7.2)	2.5 ml min <sup>-1</sup>	10–100	0.006	6.2	High sensitivity for environmental studies/halide ions
UV photometry	Continuous	(254 nm)	> 08.1 min <sup>-1</sup>	20–60	1	1	Needs reference sample
Voltammetry with membrane electrode	Continuous	Au cathode, 0.1 mol l <sup>-1</sup> HClO <sub>4</sub> membrane, Pt, Ag/AgCl	n.a.	33	1 (pulse mode)	n.a.	Sensor-like measurements <i>in situ</i> for process control

<sup>a</sup> Key: CL, chemiluminescence; UV, ultraviolet; IDS, indigo-5,5'-disulfonate; ITS, indigo-5,5',7-trisulfonate; ACVK, Acid Chrome Violet K (Alizarin Violet 3R); n.a., not available.

<sup>b</sup> For batch methods, sampling time (s) is shown in parentheses.

<sup>c</sup> Precision indicated as relative standard deviation.



As in the case of indigosulfonates, Alizarin Violet 3R (Acid Chrome Violet K) loses its color on reaction with ozone and chlorine dioxide. Ozone can be determined from the decrease in absorbance at 548 nm after mixing with a sample solution. This method is less sensitive than the indigo methods. Chlorine, hypochlorite, chlorite, and chlorate do not interfere. While chlorate(III) can also be determined by this method, it does not coexist with ozone in treated waters.

Ozone oxidizes bis(terpyridine)iron(II) into the corresponding iron(III) complex, as well as the bipyridine and 1,10-phenanthroline complexes of iron(II). The first complex is more stable than the latter complexes in a slightly acidic medium. Judging from the molar absorptivity of the iron(II) complex ( $1.1 \times 10^4 \text{ mol l}^{-1} \text{ cm}^{-1}$ ) and reaction stoichiometry (complex/ $\text{O}_3 = 2$ ), this method may be as sensitive as the indigo method.

Spectrophotometric methods using *o*-tolidine and *N,N*-diethyl-1,4-phenylenediamine are not specific for ozone and the colors developed are unstable with time.

### Electrochemical Methods

Amperometric methods using a gas-permeable membrane provide selective and sensitive measurement for aqueous ozone. Such an electrode system consists of a microporous PTFE (or polypropylene) membrane (25–100  $\mu\text{m}$  thick), gold cathode, supporting electrolyte, reference electrode (in a three-electrode system), and counterelectrode. The membrane covers the gold cathode tip and a capillary film of electrolyte. An applied voltage of +0.6 V (versus Ag/AgCl) is adopted for selective reduction of ozone. Use of the membrane reduces the current sensitivity to ozone, but improves selectivity greatly. Relative responses to possible interfering species such as  $\text{Br}_2$ ,  $\text{ClO}_2$ ,  $\text{HOBr}$ ,

$\text{HOCl}$ ,  $\text{NCl}_3$ , and  $\text{H}_2\text{O}_2$  were less than 2%. Three-electrode-type instruments with pulsed operation (100 ms pulse duration and 5 s interval) provide much higher sensitivity, because the ozone concentration builds up in the membrane layer during the interval between pulses and a transient current is obtained consuming the ozone in the layer when a pulse is applied. These sensor-like instruments are now commercially available.

*See also:* **Air Analysis:** Sampling; Outdoor Air; Workplace Air. **Amperometry.** **Chemiluminescence:** Liquid-Phase; Gas-Phase. **Coulometry.** **Remote Gas Sensing:** Overview. **Spectrophotometry:** Inorganic Compounds.

### Further Reading

- Browell EV (1989) Differential absorption lidar sensing of ozone. *Proceedings of the Institute of Electrical and Electronics Engineers* 77: 419–432.
- Ebara Jitsugyo Co. (1989) *Measurement of Ozone Concentrations in Gas and Liquid Phases* (in Japanese). Japan.
- Finlayson-Pitts BJ and Pitts JN Jr. (1986) *Atmospheric Chemistry: Fundamentals and Experimental Techniques*. New York: Wiley.
- Gordon G, Pacey GE, Cooper WJ, and Rice RG (1988) Current state-of-the-art measurements of ozone in the gas phase and in solution. *Ozone: Science and Engineering* 10: 353–366.
- Horváth M, Bilitzky L, and Hüttner J (1985) *Ozone*. Amsterdam: Elsevier.
- Rice RG and Netzer A (eds.) (1982) *Handbook of Ozone Technology and Applications*. Ann Arbor, MI: Ann Arbor Science.
- Takeuchi K, Kutsuna S, and Ibusuki T (1990) Continuous measurement of ozone in air by chemiluminescence using indigo-5,5'-disulphonate. *Analytica Chimica Acta* 230: 183–187.



# P

## PACKAGING MATERIALS

See **FOOD AND NUTRITIONAL ANALYSIS: Packaging Materials**

## PAHs

See **POLYCYCLIC AROMATIC HYDROCARBONS: Determination; Environmental Applications**

## PAINTS

Contents

**Water-Based**

**Organic Solvent-Based**

### Water-Based

**C Carr**, The Paint Research Association, Teddington, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 3715–3722, © 1995, Elsevier Ltd.

Water-borne coatings are hardly new. Emulsion paints, mostly used in domestic decorating, have been in use since the 1950s; today they account for around a third of the United Kingdom's total paint production. Water-based coatings can be formulations made from water-soluble polymers or dispersions of polymer particles in water. The vast majority of water-based coatings used for domestic and industrial purposes today are of the dispersion type because it is easier to obtain high solids at application viscosities and emulsion polymers dry quickly to give tough films.

This article will concentrate on the synthesis and characterization of water-based dispersions. Analytical and test methods that find particular use with water-based coatings will be detailed.

### Production

Dispersions are sometimes called emulsions because of the method of synthesis, but this does not signify that the final product is a liquid-in-liquid dispersion. Most commercial formulations are solid polymer dispersions and should more accurately be called lattices (or latexes).

A fully formulated water-borne dispersion paint is a much more complex mixture than a solvent-borne coating. A selection of materials that may be present is listed below:

- polymer particles,
- ionic/nonionic surfactant,
- coalescing solvent,
- corrosion inhibitor,
- antifreeze,
- biocide,
- freeze/thaw stabilizer,
- water,
- pigment/extender,
- thickening agent,
- pigment dispersant,
- antifoaming agent,
- preservative,

- neutralizing agent, and
- surface tension depressant.

Emulsion polymerization is the process in which vinyl and acrylic monomers undergo addition polymerization in water in the presence of surfactant and water-soluble initiators. Stable polymer dispersions are formed with particle sizes in the range of 0.1 to 0.5  $\mu\text{m}$ .

There are two stages involved in a typical emulsion polymerization. In the seed stage, a mixture of water, surfactant, and colloid is first heated to the reaction temperature (85–90°C). Next, 5–10% of the monomer mixture with a portion of the initiator is added. At this point the reaction mixture contains monomer droplets stabilized by surfactant, some dissolved monomer, the initiator, and surfactant (in solution and in micelles). The initiator breaks down to produce radicals, when heated and these initiate the polymerization of the dissolved monomers. Growing polymer chains eventually enter a micelle, initiating reaction of the monomer inside. If a second growing polymer enters the micelle, termination can occur.

During the seed stage, micelles of monomer are converted to latex particles, which are stabilized by surfactant. Most of the particles present in the final latex are formed in the seed stage and polymerization will have proceeded to 10% or 15% completion.

In the feed stage, the remainder of the monomer and initiator are added to the reaction mixture. Monomer diffuses from droplets through the aqueous phase to the growing polymer particles and propagates the polymerization. Radicals can also enter the particles, causing termination or re-initiation of polymerization. As the particles become larger, more surfactant from the aqueous phase is absorbed onto the surface to stabilize the dispersion.

Before cooling, a further shot of initiator may be added to ensure that all monomer has been converted to polymer.

It is more difficult to control emulsion polymerization than normal solution acrylic polymerization. Vigorous stirring is required to disperse the reactants and later to allow efficient extraction of heat through the reactor vessel walls. However, the mixture may be shear-sensitive, so excessive agitation must be avoided. Too much heating can have adverse effects, so it is usual to heat the reactants in the initial stages but apply cooling in the later stages to maintain a constant temperature.

Alkyd emulsions are made by dispersing a liquid alkyd resin in water by mechanical means. Alkyds are polyester resins modified with fatty acids. They are produced by reacting fatty acids with a polyol such as glycerol (1,2,3-trihydroxypropane) and a dibasic acid such as orthophthalic acid

(benzene-1,2-dicarboxylic acid). To facilitate emulsification, the acid content is increased by using some tribasic acid such as trimellitic (1,2,4-benzene tricarboxylic acid). Alkali, ammonia, or amine is then added to partially neutralize the resin.

Alkyd resins undergo autoxidation after application, which results in a buildup of molecular mass, improving the coating's protective properties. Autoxidation is catalyzed by the addition of small amounts (0.5%) of metal driers, such as cobalt, in the form of a metal naphthanate or octoate. The activity of the metal drier is inhibited by the presence of amine and by hydration in aqueous systems.

The latest alkyd emulsions contain only resin, water, and nonionic surfactant. High shear rates are used to disperse the resin so that high acid values and amines are not required. Drying is still inhibited to some extent as some of the cobalt resides in the aqueous phase. For this reason more dryer is required than for solvent-based systems.

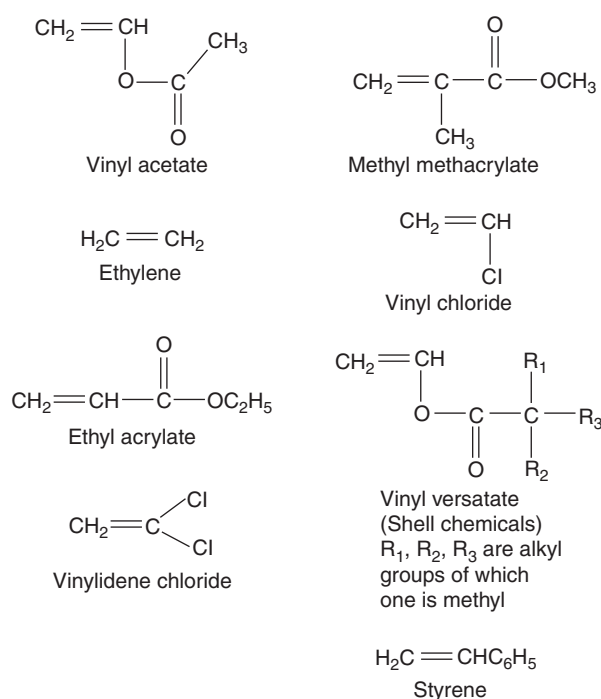
## Chemical Composition

### Monomers

The monomer composition must be chosen carefully as it will affect the properties of the final coating and the film-forming ability of the latex. The minimum film-forming temperature (MFFT) is defined as the lowest temperature at which the latex particles will coalesce to give a clear film. It is related to the glass transition temperature ( $T_g$ ) of the polymer but is also influenced by the presence of other materials such as coalescing solvents, plasticizers, and surfactants. In order that the latex particles can coalesce and form a continuous film on application, the MFFT must be below the temperature at which film formation will occur. For domestic paints an MFFT of below 10°C is required.

Emulsions can be formulated with only hard monomer and plasticized at a later stage, but it is normal to internally plasticize the emulsion by copolymerizing some soft monomer. Vinyl acetate plasticized with an acrylate such as butyl or ethylhexyl acrylate, or a dialkyl maleate, are common combinations. Other hard (high  $T_g$ ) monomers include methyl methacrylate, styrene, and vinyl chloride. Soft monomers include Vinyl Versatate<sup>TM</sup> (Shell Chemicals), ethylene, and vinylidene chloride (**Figure 1**).

The exact choice of monomers is determined primarily by the properties required in the final coating. Acrylic polymers tend to give high-quality coatings, whereas formulations based on vinyl acetate plasticized with butyl acrylate have good



**Figure 1** Common monomers used in emulsion polymerization.

chalking resistance but poor resistance to alkalis and hydrolysis.

Exterior durability is enhanced by inclusion of methacrylates and acrylates. Styrene, methacrylic acid, and acrylic acid increase hardness in the coating, while flexibility is conferred by ethyl, butyl, or ethylhexyl acrylate. Stain resistance is improved when short-chain (meth)-acrylates are used and acrylics also impart improved resistance to water. Solvent and grease resistance can be improved by copolymerizing acrylonitrile, methacrylamide, and methacrylic acid.

Thermosetting latices contain functional groups, which can be used to cross-link the coating after film formation usually by heating. These polymers normally contain hydroxy monomers that can be cross-linked with materials such as water-soluble melamine derivatives, urea-formaldehyde, or phenolic resins. Alternatively, epoxy groups in the latex can be cross-linked with amines.

Emulsion systems for anticorrosive coatings are usually copolymers of vinylidene chloride, vinyl chloride, and an alkyl acrylate or methacrylate. Relatively small amounts of surfactant and colloid are used in order to minimize water sensitivity in the dried coating. Flash rusting inhibitors such as ammonium chromate/dichromate or alkali metal nitrate/nitrites may also be added.

Paints for cathodic electrodeposition are usually emulsions of amine-modified epoxy resins with a

positively charged particle surface. The particles are destabilized when an electric current is applied using a pair of electrodes dipped into the emulsion. The particles are deposited on the cathode, which is formed by the substrate to be coated.

### Initiators and Activators

Common initiators for emulsion polymerization are compounds containing a peroxide bond, such as hydrogen peroxide and sodium peroxodisulfate. The most commonly used are the latter type, which contribute significantly to latex stabilization by introducing terminal sulfate groups on each polymer chain formed in the initial solution polymerization.

The rate of dissociation into radicals is accelerated by addition of reducing agents such as sodium hydrogensulfate (NaHSO<sub>3</sub>), sodium disulfite (metabisulfite) (Na<sub>2</sub>S<sub>2</sub>O<sub>5</sub>), sodium thiosulfate (Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub>) or sodium formaldehyde sulfoxylate (NaHSO<sub>2</sub> · CH<sub>2</sub>O · 2H<sub>2</sub>O). Activators are also added in the form of iron salts such as iron(II) sulfate (FeSO<sub>4</sub> · 7H<sub>2</sub>O) or ammonium iron(II) sulfate (Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub>).

### Surfactants

The van der Waals attraction between two atoms is very short range, acting over a few tenths of a nanometer. For latex particles, however, each atom or molecule of one particle attracts every atom or molecule in the other particle. The net attraction is long range and of considerable strength. At practical particle concentrations, dispersions would be stable for only a few seconds if they were not stabilized. The attractive force between latex particles can be overcome by electrostatic or steric stabilization.

‘Electrostatic stabilization’ is achieved by producing a charged surface on the particles. The electrostatic component can be due to charge on the particle surface or charge associated with the stabilizing moieties. This produces a layer of opposite charge in the dispersion medium and this electric double layer causes repulsion of the particles. ‘Steric stabilization’ is achieved by attaching macromolecules to the surface of the particle either by grafting or by physical adsorption.

The most common method of stabilizing latex particles is by addition of mixed anionic and non-ionic surfactants. The key property of a surfactant is that it possesses chemically dissimilar groups: one hydrophobic and one hydrophilic. The hydrophobic group is physically adsorbed onto the polymer latex particle while the hydrophilic portion extends into the aqueous phase. It is the hydrophilic groups that provide the stabilization.

Anionic surfactants are typically alkali metal salts of long-chain carboxylic or sulfonic acids, for example sodium di-2-ethylhexyl sulfosuccinate (Aerosol OT).

The best steric stabilizers are amphiphathic block or graft copolymers such as poly(oxyethylene lauryl ether) ( $M_r \approx 1200$ ). Commercial nonionic surfactants are classified according to the hydrophilic-lipophilic balance (HLB), which scales the relative solubilities of the two components in aqueous and nonaqueous media. The need for the anchor part of the stabilizing molecule can be eliminated if the stabilizing moieties can be covalently bonded to the latex particles.

### Coalescing Aids

In most cases low-volatility organic coalescing agents such as Texanol<sup>TM</sup> (2,2,4-trimethyl-1,3-pentanediol monoisobutyrate, Eastman Chemical Co.), ethylene/propylene glycols, or benzyl alcohol are added to cause high  $T_g$  polymers to coalesce at room temperature. Hydrocarbons are effective when the copolymer contains a substantial fraction of higher alkyl monomer such as ethylhexyl acrylate.

For domestic paints the MFFT is typically in the range 0–10°C but lower MFFTs are required with highly pigmented and extended finishes. For industrial applications the MFFT may be above room temperature but the coating will require heating after application to produce fully coalesced films.

To provide good film properties the  $T_g$  of the final film must be raised above ambient temperatures by evaporation of coalescing solvent for thermoplastic systems or cross-linking in thermosetting formulations.

### Colloids

Colloids or water-soluble polymers are added to the formulation to assist with particle size control and to modify the rheology of the final paint. They can be water-soluble polymers such as poly((meth)acrylic acid) and its copolymers, poly(vinyl alcohol) or substituted celluloses such as hydroxyethylcellulose.

Associative thickeners are low-molecular-mass water-soluble polymers with at least two hydrophobes such as hydrophobically modified ethylene oxide-urethane block copolymers (HEUR) or hydrophobically modified hydroxyethylcellulose (HMHEC). The hydrophobes can associate with themselves or with hydrophobes on surfactant, cosolvent, latex, and pigment. This sets up a loose network that is sensitive to mechanical disturbance but re-forms quickly. The result is that pigment settling and film sag are reduced because the network structure increases the low shear viscosity, but the formulations show easy

brushing and improved leveling because the network breaks down under high shear.

### Pigments and Extenders

A paint is made by mixing the binder, in this case the latex, with pigment. The pigment is initially dispersed in water with a surfactant and the slurry added to the latex and mixed further. The amount of pigment used is characterized by the pigment volume concentration (PVC), which is defined as

$$\text{PVC} = \frac{\text{volume of pigment}}{\text{volume of pigment and binder}} \times 100\% \quad [1]$$

For a gloss emulsion paint the PVC will be  $\sim 20\%$ , but for a matt emulsion finish the PVC will be much higher.

The most common white pigment used in coatings is titanium dioxide ( $\text{TiO}_2$ ), though small quantities of antimony oxide ( $\text{Sb}_2\text{O}_3$ ), zinc oxide ( $\text{ZnO}$ ), and lithopone ( $\text{ZnS} + \text{BaSO}_4$ ) may still be used. The surface of  $\text{TiO}_2$  pigments is photoactive and can initiate reactions that would lead to the breakdown of the organic binder. For this reason the surface is coated with small amounts of oxides of aluminum and/or silicon and/or zirconium. A typical  $\text{TiO}_2$  pigment for a water-based coating would contain 1–4% aluminum oxides, 0–6% silicon oxides, and 0–1% zirconium oxides.

Extenders may be added to reduce the price of paint or to modify the appearance by reducing gloss. Some common extenders are barium sulfate, calcium carbonate, silica, talc ( $\text{Mg}_3\text{H}_2(\text{SiO}_3)_4$ ), and china clay ( $\text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2 \cdot 2\text{H}_2\text{O}$ ).

It is more difficult to achieve good pigment dispersion in water-based than in organic solvent-based systems. Good dispersion is important to obtain high gloss and opacity and good barrier properties in the final coating. Pigment flocculation can be caused by soluble polymers or ionic impurities in the aqueous phase. Some anticorrosive pigments must be slightly soluble to be effective.

Even if the dispersion is stable, flocculation can occur during drying of the film as the concentration increases. For these reasons the choice of dispersing agent (surfactant) for the pigment is critical. There are several types of commercial dispersants available including anionic polyphosphate (Calgon), sodium salts of polycarboxylic or acrylic acids, or amino alcohols.

Typical formulations for two types of domestic water-based coatings are shown in Table 1.

Coatings made from water-based dispersions suffer several deficiencies when compared to

**Table 1** Typical formulations for domestic water-based coatings

	<i>Gloss emulsion weight (%)</i>	<i>Water-based undercoat weight (%)</i>
TiO <sub>2</sub>	21.0	25.0
Extender		14.5
Acrylic latex (50% solids)	56.0	26.0
Antifoam	0.2	0.2
Pigment dispersant	0.2	0.25
Coalescing solvent	10.0	1.0
Biocide	0.5	0.5
Ammonia	1.0	1.0
Thickener	0.2	0.25
Water	10.9	31.3

solvent-borne coatings. Incomplete coalescence and pigment flocculation affect the film's barrier properties and reduce gloss. Temperature and humidity strongly influence the drying characteristics. At low humidity, drying can be so rapid as to cause craters to appear in the final film; at high humidity, drying can be so slow that severe sagging occurs.

Because of water's high latent heat of vaporization, as much as four times more energy is required to force-dry or stove water-borne coatings compared to conventional coatings.

On ferrous substrates, flash rusting may be a problem and penetration into porous substrates is poorer than with solvent-borne coatings, which can lead to premature failure especially on a chalky surface.

Finally, biocides may be required because many components of a water-borne coating are nutrients for microorganisms.

## Analytical Techniques

Routine methods of chemical analysis can be applied to water-based coatings. For example, the nature of the binder can be determined by infrared spectroscopy. Water-resistant plates, such as KRS5 or zinc selenide, must be used if capillary films are cast from the aqueous dispersion. Dried films can be analyzed by surface-sensitive techniques such as attenuated total reflection (ATR) or photoacoustic spectroscopy (PAS). Both these techniques require the use of Fourier transform infrared (FTIR) instruments to obtain spectra in a reasonable time.

The nature and amount of volatile organic components, such as coalescing solvents or plasticizers and remaining monomer, can be measured by gas chromatography coupled with mass spectrometry (GC-MS).

## Particle Size

A fundamental property of water-based dispersion paints is the particle size of the latex and pigment particles. There are several methods available for measuring particle dimensions and size distributions but the methods used are determined by the size ranges involved. Latex particles are submicrometer particles with diameters typically within the range 100–500 nm. For the pigment titanium dioxide, particles are always ~250 nm as this size gives the maximum light-scattering and hence opacity.

Sedimentation techniques are amongst the oldest methods of particle size measurement. Sizes are derived from the rate of settling of particles under the action of gravitational or centrifugal forces. One sedimentation technique is the disk centrifuge, which is used to measure particles in the range 0.1–50 µm.

The disk centrifuge consists of a transparent disk-shaped container that spins at a speed of 500–15 000 rpm. The particles to be measured are introduced onto the spinning disk using a so-called buffered line start method in which a spin fluid is made up so that a density and viscosity gradient is created within it. This prevents 'streaming' or 'settling' of the particles in the spin fluid.

Measurements are made using X-rays rather than visible radiation because X-rays are absorbed in direct proportion to the mass of particles between the source and detector. The diameter of a particle can be calculated by measuring the time  $t$  for the particle to reach a position  $r_t$  from a starting position  $r_i$  using a modified form of Stokes' law:

$$t = \frac{18\eta \ln(r_t/r_i)}{w^2 d^2 \Delta p} \quad [2]$$

where  $\eta$  = the viscosity of the spin fluid,  $w$  = the centrifugal speed of the disk, and  $\Delta p$  = the difference in density between the particle and spin fluid.

The technique is very sensitive to differences in particle size as the rate of sedimentation is proportional to the square of the particle diameter.

Direct observation of particles can be achieved using scanning electron microscopy (SEM) or transmission electron microscopy (TEM). This can give very accurate measurements as monodisperse particles can be used as standards but the process can be time-consuming. It is important that the measured particles are representative of the whole sample, so many particles, preferably from different micrographs, must be measured. This is possible by automating the measurements using image-analysis systems that are now available for personal computers.

Light-scattering methods can be used to measure the size of dispersed latex and pigment particles.



Scattering of radiation originates in the oscillating dipoles of polarizable particles induced by the oscillating electric field of an incident light beam. The intensity of scattered radiation is dependent upon the ratio of the particle size to the wavelength of the incident radiation and the refractive index difference between the particle and the suspending medium.

In photon correlation spectroscopy (PCS), light from a low-power helium–neon laser is focused on a temperature-controlled sample cell and light scattered at a known angle to the cell is detected by a photomultiplier. The random motion of particles in the laser beam causes fluctuations in the intensity of the scattered radiation that can be analyzed with a digital correlator. The smaller the particle the more rapid the fluctuations due to more rapid motion. The time dependence of the fluctuations is used to generate a correlation function, which is the sum of fluctuations caused by all particles. Autocorrelation theory can then be used to determine the diffusion coefficient,  $D$ , for the particles and hence the particle's hydrodynamic diameter,  $S$ , from the Stokes–Einstein equation:

$$D = \frac{kT}{3\pi\eta S} \quad [3]$$

where  $k$  is Boltzmann's constant,  $T$  is temperature, and  $\eta$  is the viscosity of the suspending medium.

The information is presented in the form of an intensity distribution, with the mean particle size and a measure of the polydispersity that is related to the variance in the average diameter (Figure 2).

PCS gives very quick, reliable measures of particle sizes for monodisperse systems but problems can arise in systems containing two distinct sizes of

particles. Separation into two peaks is only possible when the size ratio approaches 2:1.

The practical size range covered by PCS is 10–3000 nm. For larger particles, laser diffraction can be used up to sizes of 600  $\mu\text{m}$ . The principle behind laser diffraction is that diffraction from small particles is greater than from large particles. The diffraction of a laser beam through small angles is measured, from which the particle size can be calculated. The scattering angle is inversely proportional to the particle size and the intensity at each angle is the sum of scattering from particles of a given size.

### Zeta Potential

Photon correlation spectroscopy can also be used to measure electrophoretic mobility and zeta potentials of suspended particles. The sample is subjected to an electric field that causes charged particles to migrate to one of the electrodes. Two coherent laser beams intersect within the sample, giving rise to a series of interference fringes. The fluctuations in the signal intensity are analyzed in the manner described above to calculate the mobility distribution of the particles and then the potential difference across the diffuse part of the double layer (zeta potential).

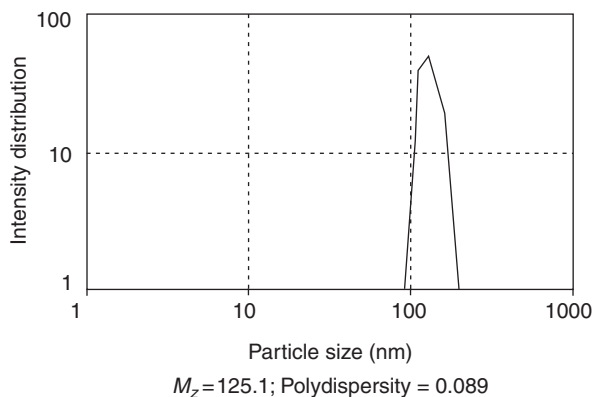
Typically, the zeta potentials for latex or pigment particles are measured at different pH values. The zeta potential and hence the magnitude of the electric double-layer surface charge vary with the pH of the aqueous phase. The point at which the net charge is zero is termed the isoelectric point.

The stability of emulsion paints will be affected by the charge surrounding the latex pigment and extender particles. Flocculation will occur if the electric double layer is insufficient to maintain interparticle separation. Therefore, zeta potentials of the individual particles in the paint can give a measure of the tendency to flocculate and hence of the long-term stability of the paint. However, emulsion paints cannot rely solely on electrostatic stabilization because they need to withstand dilution with hard water. This means that some steric stabilization will invariably be used.

### Degree of Flocculation

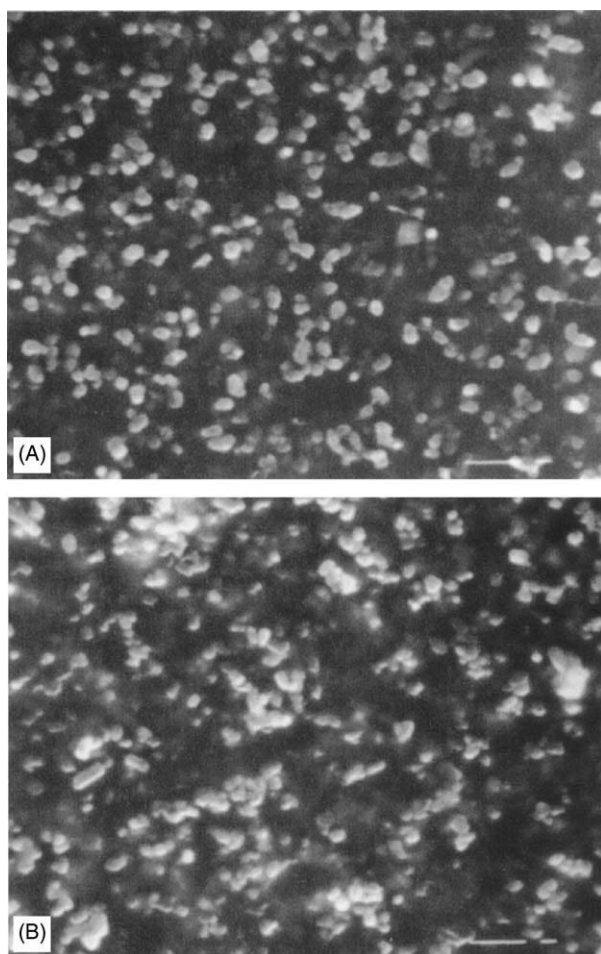
Flocculation of pigment is more difficult to control in water-based coatings than in solvent-borne ones. Good pigment dispersion is important to achieve good gloss, opacity, and barrier properties.

The scattering efficiency of electromagnetic radiation by spherical particles is, among other factors, a function of their diameter and of the wavelength of the incident radiation. To give good opacity and whiteness,



**Figure 2** Particle size distribution for an acrylic latex by PCS.  $M_z$  is an expression of mean particle size that depends upon the scattering power per particle ( $P_i$ ) of size  $S_i$  where there are  $N_i$  particles of size  $S_i$ .  $M_z = \sum_i (S_i^3 P_i N_i) / \sum_i (P_i N_i)$



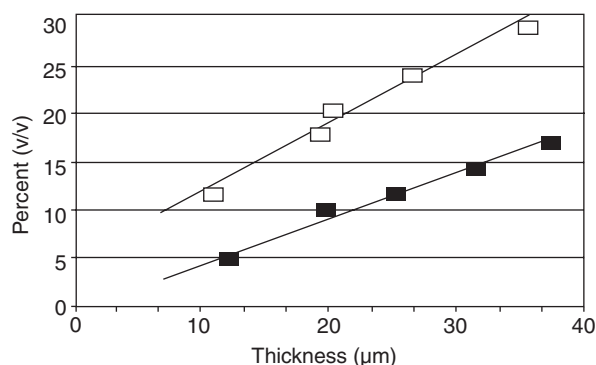


**Figure 3** Scanning electron micrographs showing pigment distribution in commercial solvent-based (A) and water-based (B) white gloss paints (bar = 1  $\mu\text{m}$ ).

the optimum diameter of  $\text{TiO}_2$  particles is 0.24  $\mu\text{m}$ , which gives maximum scattering of radiation at a wavelength of 0.55  $\mu\text{m}$ , the center of the visible spectrum. Flocculation of pigment means that the effective particle size of the  $\text{TiO}_2$  is increased and longer wavelengths of light will be preferentially scattered.

Pigment distribution in dried films can be observed directly by SEM. The surface of the film must first be etched away to reveal the pigment particles underneath (Figure 3). This can be achieved using a cold oxygen plasma generated by a commercial radiofrequency plasma barrel etcher. Image analyzers can be used to determine the dimension and area of the clumps of pigment.

The dependence of the scattered radiation on the degree of flocculation is the basis of instruments designed to measure flocculation. The Flocculation Monitor developed by Tioxide Europe Ltd measures the flocculation of pigment in wet and dry films. For



**Figure 4** Flocculation gradients for commercial solvent-based (solid symbols) and water-based (open symbols) white gloss paints.

dry films, samples of different thicknesses are irradiated with light of 2.5  $\mu\text{m}$  wavelength. The percentage of light that is back-scattered at each of the thicknesses is measured. A plot of reflectance versus film thickness produces a straight-line graph with a slope equal to the flocculation gradient. Low values indicate good pigment dispersion and high values indicate flocculation. This is illustrated in Figure 4 for commercial solvent-borne and water-borne glosses. The flocculation gradient for the water-borne system is 0.71 compared to 0.47 for the solvent-borne system.

For wet samples, a single value is obtained, and the lower the value the better the degree of dispersion.

### Standard Test Methods

There are a large number of standard test methods available for paints to ensure that specific performance criteria are met. The most widely used standards originate from the American Society for Testing and Materials (ASTM), British Standards Institution (BSI), Deutsche Institut für Normung (DIN), and the International Organization for Standardization (ISO). The methods used will depend on the paint type and intended application. A selection of methods is described below that are likely to be used for water-based coatings.

'Minimum film-forming temperature' is measured according to ASTM D2354. It entails the visual observation of cracking or whitening in films that have dried over a metal bar having a controlled temperature gradient. Temperatures between  $-5$  and  $90^\circ\text{C}$  can be covered with commercial equipment such as the Rhopoint MFFT-90.

At points on the bar above the MFFT, an applied latex will coalesce to produce a clear film if no pigment or other opaque materials are present.

Below the MFFT, a white, powdery, cracked film is obtained. Temperature sensors are placed at regular intervals under the bar and are used to control the temperature and to indicate the temperature at different points along the bar.

The gloss or 'sheen' (BS3900: Part D5) of a coating is determined on a meter that measures the proportion of light reflected at a fixed angle (normally 20°, 60°, or 85° depending on the gloss level).

'Opacity or contrast ratio' (BS3900: Part D4) is the paint's hiding power. It is determined by measuring percentage reflectance of light from films drawn down over black and white substrates. For equal hiding power over black and white, the contrast ratio is 100%.

'Porosity or water permeability' (ASTM D1653) is determined by measuring the rate at which water escapes from a cup sealed by a sample of the test coating. Alternatively, the cup can be filled with desiccant and the rate at which water migrates into the cup through the coating is measured.

A paint's 'stain resistance' is determined according to BS3900: Part G5 and entails applying typical stains to a coating and assessing the ease of removal using various cleaning solutions.

'Scrub resistance' is the ability of the paint to withstand repeated cleaning with detergent. Test methods such as ASTM D2486 require the use of mechanical scrub machines, brushes, or pads. Resistance can be quoted as weight loss for a given number of cycles or the number of cycles to erode a given film thickness.

'Blocking' is the undesirable sticking together of two painted surfaces when they are pressed together under normal conditions or under specified conditions of temperature, pressure, and relative humidity. The blocking point is the lowest temperature at which two coated surfaces in mutual contact will stick together sufficiently to injure the surfaces permanently and/or prevent easy separation. This test is important for coatings that could be applied to window frames.

There are several standard methods for determining block resistance (such as ISO 4622, ASTM 3003-71 and ASTM 4946-89). In ISO 4622 the pressure at which damage or sticking occurs to painted surfaces in contact at a given temperature and humidity is measured.

'Freeze-thaw stability' is determined according to ASTM D2243 by freezing the paint at -18 °C for 17 h then equilibrating at 23 °C and measuring any change in consistency on thawing.

'Leveling' of a paint describes the flow-out of brush marks, etc. It can be estimated by drawing the paint down with a crenellated applicator and

comparing undulations on the dry film surface with a set of standard photographs. The result is quoted on a scale of 1 to 10, where 1 signifies poor leveling.

'Surface dry time' (BS3900: Part C2) is estimated allowing a series of coated panels to dry and observing whether ballotini powder sticks to the coating at different time periods.

'Hard dry time' (BS3900: Part C3) is determined using a mechanical thumb that rotates a rubber pad in contact with the film under a specified force through 90° and looking for damage to the film. The film is hard dry when no damage can be seen.

'Salt spray resistance' is a common test applied to anticorrosive paints even though the validity of this test, especially for water-based coatings, has been questioned. BS3900: Part F4 describes a continuous spray method the result of which is a pass or fail. The coating is applied to a 4 inch by 6 inch (100 mm by 150 mm) steel plate and allowed to dry for 24 h. A scalpel is used to score a line through to the metal, starting 2.5 cm from the top of the panel to 2.5 cm from the bottom. The panel is then placed in a salt spray cabinet where it is subjected to a salt mist produced by spraying a salt solution through an atomizer. The panel is observed after 48 h, 1, 2, and 3 weeks and any blistering or adhesion problems are noted as well as the distance corrosion has extended from the cut.

'Fungal resistance' (BS3900: Part G6) is measured by inoculating the coating with a spray containing spores of several species of fungus. The test sample is then incubated for 14 days and the percentage coverage of the panel by fungal growth is estimated. The result is quoted on a scale of 0 to 5, where 5 is 70% coverage.

**See also:** **Gas Chromatography:** Mass Spectrometry. **Paints:** Organic Solvent-Based. **Surfactants and Detergents.**

## Further Reading

- Gardner HA and Sword GG (1972) *Paint Testing Manual*, 13th edn. American Society for Testing and Materials.
- Karsa DR (ed.) (1990) *Additives for Waterbased Coatings*. Royal Society of Chemistry Special Publication no. 76.
- Lambourne R (ed.) (1987) *Paint and Surface Coatings*. Chichester: Ellis Horwood.
- Napper DH (1983) *Polymeric Stabilization of Colloidal Dispersions*. London: Academic Press.
- Oil and Colour Chemist's Association (1976) *Introduction to Paint Technology*, 4th edn. Middlesex: Oil and Colour Chemist's Association.

Oil and Colour Chemist's Association, Australia (1983) *Surface Coatings*, vols. 1 and 2, 2nd edn. London: Chapman and Hall.

Turner GPA (1988) *Introduction to Paint Chemistry and Principles of Paint Technology*, 3rd edn. London: Chapman & Hall.

Wilson AD, Nicholson JW, and Prosser HJ (eds.) (1987, 1988) *Surface Coatings, I and II*. Amsterdam: Elsevier.

Woodbridge R (1990) *Principles of Paint Formulation*. Edinburgh: Blackie and Son.

## Organic Solvent-Based

**S Garrigues Mateo**, University of Valencia, Burjassot, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Paints are, in general, fluid materials that, applied as a thin film on a surface and after a drying or curing process, provide a solid, protective, decorative, or functional film.

From the first patent of a paint formulation filed by Averill Paint Co., New York, in 1867, and the initial production of a series of varnishes, dyes, and enamels in 1880 by Sherwin-Williams Co., the development of synthetic resins and the industrial production of a wide range of solvents and new pigments are the principal reasons for the tremendous development of the paint industry until today.

In spite of the development of environmentally friendly coating systems, based on the use of low-solvent content paints, free-solvent powder coatings, and waterborne-based products, many of the traditional organic solvent-based paints are difficult to replace without affecting the quality of the finished products.

In this article, a basic scheme for the complete analysis of organic solvent-based paints is described and the main analytical techniques available for this purpose are discussed.

### Chemical Composition

**Figure 1** summarizes the composition of paints. In general, a paint is composed of a binder, a pigment, and a solvent or mixture of solvents. However, some paints do not contain some of these products (e.g., varnishes do not contain pigments and solid paints do not include any solvent). Additionally, some additives are included in paint formulations during their manufacture in order to control the dispersion of the

different components or to facilitate the application or preservation of the paint.

### Binders

The binder is the most important component of a paint formulation. It binds the pigment particles in a homogeneous dispersion and fixes these particles to the painted surface, providing a final continuous film. Binders determine the application method, drying and hardening behavior, adhesion to the substrate, mechanical properties, and chemical and weathering resistance of the paints. The binder is dissolved in an appropriate solvent, and both binder and solvent constitute the vehicle of the paint.

Binders are, in general, organic resins (of natural or synthetic origin) but in some cases inorganic salts, for example, silicates, can be used as binders in fire-resistant paints.

Drying or curing of the paint film occurs by simple evaporation of the solvent in which the binder is dissolved (carried out at room temperature using air or accelerated by increasing the temperature), or by chemical reactions between the binder components (oxidation, addition, and condensation reactions have been employed in paint curing).

**Table 1** summarizes the different types of resins commonly employed in paint manufacture.

### Pigments and Extenders

Pigments are fine solid particles dispersed in the binder. Pigments provide color, gloss, and opacity required of the paint film. Some pigments also act as protective agents against corrosion by atmospheric agents.

On the basis of their chemical composition, pigments can be classified as inorganic (carbonates, oxides, sulfates, sulfides, silicates, chromates, molybdates, hexacyanoferrates, metals, and synthetic ceramic pigments) and organic (carbon black, aniline

Oil and Colour Chemist's Association, Australia (1983) *Surface Coatings*, vols. 1 and 2, 2nd edn. London: Chapman and Hall.

Turner GPA (1988) *Introduction to Paint Chemistry and Principles of Paint Technology*, 3rd edn. London: Chapman & Hall.

Wilson AD, Nicholson JW, and Prosser HJ (eds.) (1987, 1988) *Surface Coatings, I and II*. Amsterdam: Elsevier.

Woodbridge R (1990) *Principles of Paint Formulation*. Edinburgh: Blackie and Son.

## Organic Solvent-Based

**S Garrigues Mateo**, University of Valencia, Burjassot, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Paints are, in general, fluid materials that, applied as a thin film on a surface and after a drying or curing process, provide a solid, protective, decorative, or functional film.

From the first patent of a paint formulation filed by Averill Paint Co., New York, in 1867, and the initial production of a series of varnishes, dyes, and enamels in 1880 by Sherwin-Williams Co., the development of synthetic resins and the industrial production of a wide range of solvents and new pigments are the principal reasons for the tremendous development of the paint industry until today.

In spite of the development of environmentally friendly coating systems, based on the use of low-solvent content paints, free-solvent powder coatings, and waterborne-based products, many of the traditional organic solvent-based paints are difficult to replace without affecting the quality of the finished products.

In this article, a basic scheme for the complete analysis of organic solvent-based paints is described and the main analytical techniques available for this purpose are discussed.

### Chemical Composition

**Figure 1** summarizes the composition of paints. In general, a paint is composed of a binder, a pigment, and a solvent or mixture of solvents. However, some paints do not contain some of these products (e.g., varnishes do not contain pigments and solid paints do not include any solvent). Additionally, some additives are included in paint formulations during their manufacture in order to control the dispersion of the

different components or to facilitate the application or preservation of the paint.

### Binders

The binder is the most important component of a paint formulation. It binds the pigment particles in a homogeneous dispersion and fixes these particles to the painted surface, providing a final continuous film. Binders determine the application method, drying and hardening behavior, adhesion to the substrate, mechanical properties, and chemical and weathering resistance of the paints. The binder is dissolved in an appropriate solvent, and both binder and solvent constitute the vehicle of the paint.

Binders are, in general, organic resins (of natural or synthetic origin) but in some cases inorganic salts, for example, silicates, can be used as binders in fire-resistant paints.

Drying or curing of the paint film occurs by simple evaporation of the solvent in which the binder is dissolved (carried out at room temperature using air or accelerated by increasing the temperature), or by chemical reactions between the binder components (oxidation, addition, and condensation reactions have been employed in paint curing).

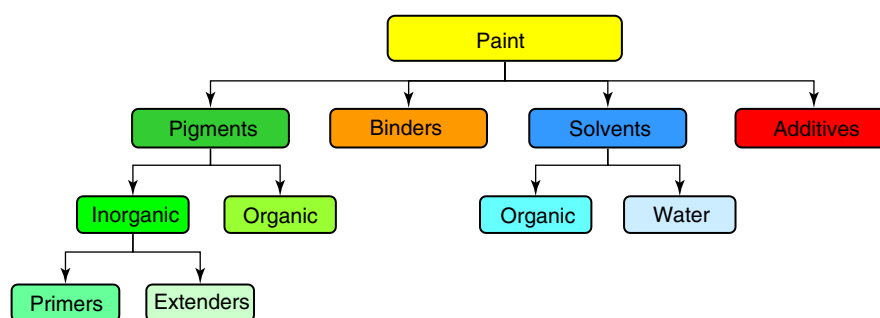
**Table 1** summarizes the different types of resins commonly employed in paint manufacture.

### Pigments and Extenders

Pigments are fine solid particles dispersed in the binder. Pigments provide color, gloss, and opacity required of the paint film. Some pigments also act as protective agents against corrosion by atmospheric agents.

On the basis of their chemical composition, pigments can be classified as inorganic (carbonates, oxides, sulfates, sulfides, silicates, chromates, molybdates, hexacyanoferrates, metals, and synthetic ceramic pigments) and organic (carbon black, aniline





**Figure 1** Main components of liquid paint formulations.

**Table 1** Synthetic resins commonly employed in paint manufacture

Resin type	Description
Alkyd	Condensation products of polybasic acids or anhydrides and polyhydric alcohols. Contains monobasic fatty acids
Polyester	Condensation products composed mainly of co-reacted di- or polyhydric alcohols and di- or tribasic acids or anhydrides
Polyurethane	Condensation products of diisocyanate and diol
Phenolic	Condensation products of phenol and formaldehyde
Epoxy	Polyesters based on the reactivity of the epoxy group (i.e., condensation product of bisphenol A and epichlorohydrin)
Polystyrene	Co-polymer based on styrene monomer
Acrylic	Co-polymers based on multifunctional monomers, such acrylates and methacrylates
Vinyl	Contains unsaturated vinyl groups
Amino	Condensation products of an amide and formaldehyde
Silicone	Produced by intermolecular condensation of silanols
Cellulosic	Polymer based on cellulose
Fluorocarbons	Co-polymer of fluorocarbon monomers

black, phthalocyanines, monoazo compounds, azo dyes and vat dyes, quinacridones, and other substances such as metal complexes and basic dyestuff complexes).

Organic pigments have more intense colors and are more brilliant than inorganic ones. However, the latter provide higher opacity and hiding power.

In addition to pigments a series of extenders or charges are introduced in the paint manufacturing process. Extenders have a poor hiding power and are more economical than pigments, and these products can be used as fillers to introduce modifications in the final characteristics of the paint, such as to modify the thixotropic behavior of the matrix or to reduce

the cost of the paint. The extenders most usually employed are carbonates, sulfates, silicates, phosphates, borates, silica, and alumina.

### Solvents

Solvents are introduced into the paint formulation to dissolve the binder and to provide adequate fluidity of the paint to allow it to be applied. The most important characteristics of solvents are their capacity to dissolve the resins and their volatility, which controls the speed of evaporation. Solvents employed in solvent-based products are hydrocarbons (both aliphatic and aromatic compounds), oxygenated solvents (frequently alcohols, esters, ethers, or ketones), and terpenic solvents; in general, mixtures of solvents are used in paint manufacture.

The present trend in paint manufacture is to avoid high concentrations of low-boiling solvents in order to prevent the emission of pollutants into the environment. However, the worldwide production of solvent-based paints for industrial purposes continues to be very important.

### Additives

Additives are minor components in the paint composition and, in general, each one of these products represents no more than 2% of the total volume of the formulation. However, they are important in order to achieve specific properties of the paint.

The most frequently used paint additives are wetting, dispersant, and antissettling agents, used to prevent pigment settling; antiskinning agents, employed to prevent surface drying; defoamers and antifoams, necessary in the manufacture of aqueous and some solvent-based paints; driers; preservatives and fungicides, employed to prevent growth of microorganisms; thickening agents to control the rheological properties; ultraviolet absorbers, which reduce the degradation of the coating by sunlight;

catalysts, employed for the curing of paints; anti-floating and antiflooding agents; and matting agents. Paint additives can be classified into two groups as a function of their solubility or insolubility in the paint vehicle; accordingly, in paint analysis they can be found in the pigment or in the binder fraction after treatment of the sample with an appropriate solvent.

## Production of Paints

The manufacture of paints is based on adequate dispersion of the pigment into the vehicle. The dispersion of the pigment is carried out by mechanical stirring and by means of a series of additives (humidifiers, surfactants, antifoaming products), after which the mixture is ground or milled, using ball mills, to achieve adequate fineness of the dispersion. After this process the physical parameters, such as color and viscosity, are adjusted to obtain the desired properties. Filtering, filling, and shipping are the final steps in paint manufacture.

The most important characteristics of paint are determined by the amount of pigment and binder, and so an important parameter, which can be used to evaluate paint films, is the volume of pigment and charges, termed the pigment volume concentration (PVC). This can be defined as

$$\text{PVC}(\%) = \left( \frac{\text{volume of pigments and extenders}}{\text{volume of nonvolatile components}} \right) \times 100$$

The properties of paints, such as water permeability, corrosion resistance, gloss, hiding power, or the tendency to blister, are dependent on the PVC. There is a critical value of this parameter, called the critical pigment value concentration, at which the practical properties of the paint undergo important changes.

The PVC of a paint varies as a function of its use as a primer (in this case values of PVC between 30% and 50% are recommended), as an intermediate film (a PVC between 30% and 35% is most appropriate), or as a finishing paint (in which case the PVC must be lower than 20%). Only in the case of paints with a high content of zinc is PVC higher than 95% expected.

## Analysis of Paints

In the paint manufacturing industry, chemical analysis is essential for investigating and developing new products, solving problems during the production process, and controlling the quality of both the raw materials and the final products. The analysis of the

manufactured products can provide valuable information for the technical development of other industries and can help to solve customer complaints.

The analysis of a liquid paint can be performed on a well-characterized sample, during its production, or on a finished unknown sample. In the first case it is possible to develop simple and rapid analytical procedures in order to control various parameters.

The analysis of unknown samples is the most difficult problem because of the lack of prior information that must be compensated by the development of a rigorous analytical scheme. So it is necessary to carry out a series of pretreatments of the sample in order to simplify the application of instrumental techniques to the identification and quantitative analysis of the components. In this case, prior separation of the paint components is one of the most important steps.

On the other hand, a series of technical norms have been developed to guarantee the quality and characteristics and/or properties of paints and coating materials; these are provided by international organizations such as the ASTM (American Society for Testing and Materials) and ISO (International Standard Organization). Concerning the control tests that are currently applied to paint samples, **Table 2** summarizes the most important properties of paints and the techniques employed in their evaluation.

When paints are exposed outdoors they generally become less glossy and have more tendency toward cracking and flaking, also showing color changes or fading. Paints that are indoors and exposed to sunlight that has passed through window glass generally degrade less rapidly as they are not exposed to rainfall or to the ultraviolet (UV) radiation from the sun. Permanence or fastness refers to the chemical stability of the paint pigments in relation to any chemical or environmental factor, including light, heat, water, acids, alkalis, or mold.

Different standard test methods have been proposed in order to evaluate lightfastness and weather fastness of paints. A painted test panel, generally with controlled conditions of temperature and humidity, is exposed to a light source (sun, artificial daylight fluorescent lamp, or xenon arc) directly or behind glass. After a specified period of exposure, any change in color in the paint film is determined by comparing the sample with a reference specimen that has been stored away from the light. Color difference units can be calculated by the CIE 1976  $L^*a^*b$  color difference equation and according to the relative resistance to change paint pigments can be assigned to categories of relative lightfastness. ASTM defines a five-point rating system for paints lightfastness, from I (excellent) to V (very poor).



**Table 2** Test employed for the study and control of the properties of liquid-solvent based paints

Property	Assay
Viscosity	Standardized flow cups for Newtonian fluids Brookfield viscometer for non-Newtonian materials
Density	Measurement using a pycnometer
Degree of fineness of dispersion	Using a fineness gauge
Opacity or hiding power	For liquid film using cryptometer  By measurement of the reflectivity of a dry film applied on a black-and-white hiding-power chart
Gloss	Using a reflectometer by measurement of the specular reflectance for different angles of incidence
Flash point	Measurement of the lowest temperature at which the application of a test flame ignites the vapor generated by the heated sample
Volatile content	
Nonvolatile matter	By difference of weight after heating the sample at 105°C in an oven for ~3 h
Pigment content	By weight difference after extraction of all the vehicle using a suitable solvent mixture and centrifugation By low-temperature furnace ashing (some organic pigments may be lost by this method)
Color measurement	By measuring the spectral reflection using a spectrophotometer and establishing the color coordinates

Another lightfastness category is the Blue Wool Scale, which includes eight levels: from 1 (poor) to 8 (extremely lightfast).

### Sampling of Paints

Sampling of paints can be carried out during manufacture, in the filling step, or from a finished consignment. In the first case, the sample must be previously homogenized, or several samples must be taken at different levels using an adequate probe. In this case, it is necessary to carry out a filtration step before the analysis.

When the sampling is carried out during filling, the laboratory sample is obtained from different samples taken at regular intervals, but when the paint is already filled it is necessary to carry out random sampling of a representative sample comprising  $\sqrt{n/2}$  packages,  $n$  being the total number of packages in the lot. Adequate sample containers must be selected to avoid losses of volatile solvents through their walls.

In general, it is necessary to ensure that the laboratory sample is representative of the whole consignment.

To weigh liquid paint sample it is convenient to use a glass syringe and to determine the mass of the sample by difference. This process avoids losses of volatile compounds during sample handling.

In the analysis of coatings of dry paint film, the components can be characterized using microscopic techniques that can be applied to the study of the homogeneity of samples and require only small amounts of material.

### Separation of the Paint Components

A prior separation of the binder, pigments, and solvents from the whole paint is recommended for their complete analysis by the different techniques available.

The separation of the pigment from the vehicle is carried out by centrifugation of the sample. For this purpose an adequate quantity of paint (of the order of 20 ml) is placed in a centrifuge at 5000 rpm (or up to 20 000 rpm for samples containing finely divided pigments and low-density and viscous solvents) for 30 min or until complete separation is achieved. To avoid sedimentation of the pigment dragging significant quantities of the ligand, especially with high-viscosity samples, it is recommended to dilute the sample with an appropriate solvent, such as tetrahydrofuran or methyl ethyl ketone. Samples are decanted to separate the vehicle and the pigment is washed several times with the diluent employed.

When it is necessary to determine quantitatively the amount of pigment in a paint sample, an accurate weight of sample is extracted and re-centrifuged thrice with a solvent (a toluene–acetone, 2:1 v:v, mixture or tetrahydrofuran), and after a final extraction with ethyl ether or petroleum ether, the pigment is dried in an oven at 105°C for 2 h and its amount established by weighing.

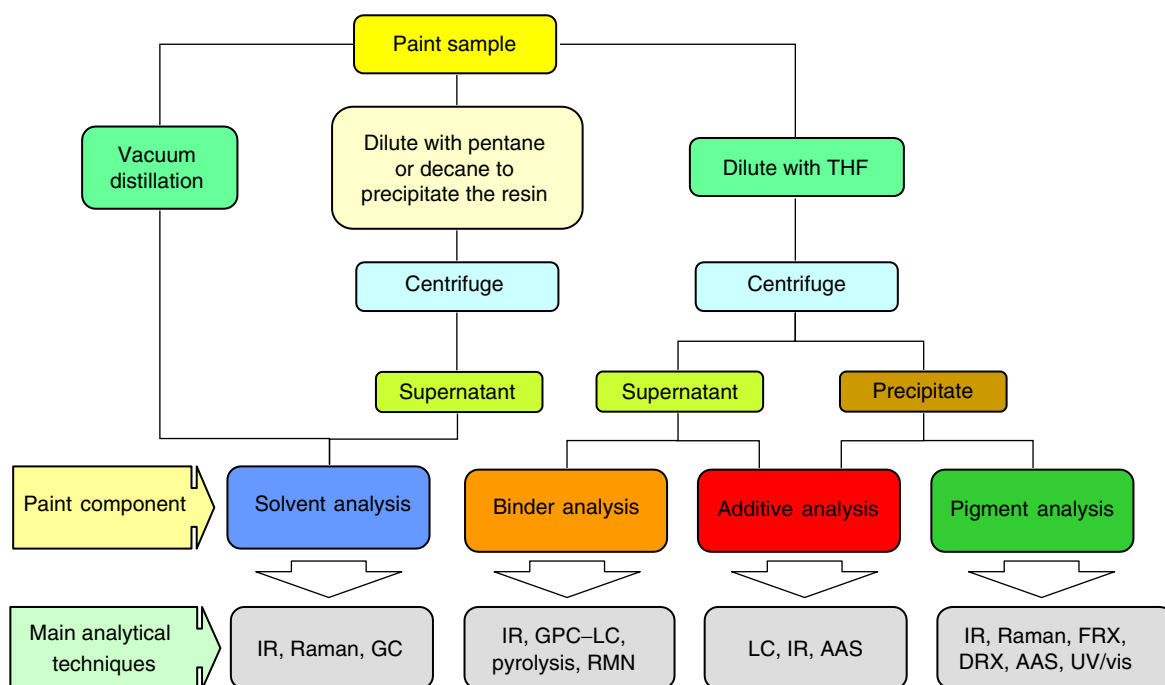
The evaporation of the solvent, added to the vehicle for the separation of the pigment by centrifugation, permits isolation of the binder for subsequent identification.

The separation of the solvent from the other components of a paint sample can be carried out by vacuum distillation. For this purpose tricresyl phosphate is added to the sample and the mixture is heated (at 155°C or 160°C) in a silicone bath under vacuum.

Figure 2 summarizes the procedures employed in the separation of solvent-based paint components.

### Analytical Techniques Employed in Paint Analysis

At present, the majority of procedures employed for the identification and quantification of paint



**Figure 2** Separation scheme for solvent-based paint components and analytical techniques employed for their analysis.

components are based on the use of instrumental techniques. However, some chemical methods can be applied to determine the presence of cations, anions, or functional groups in a paint sample and from these results information about pigments and binders can be obtained. The elemental analysis of paint, after classical fusion with sodium, permits the establishment of the presence of nitrogen, sulfur, halogens, or phosphorus, and from that it information about the type of resin that constitutes the binder can be obtained.

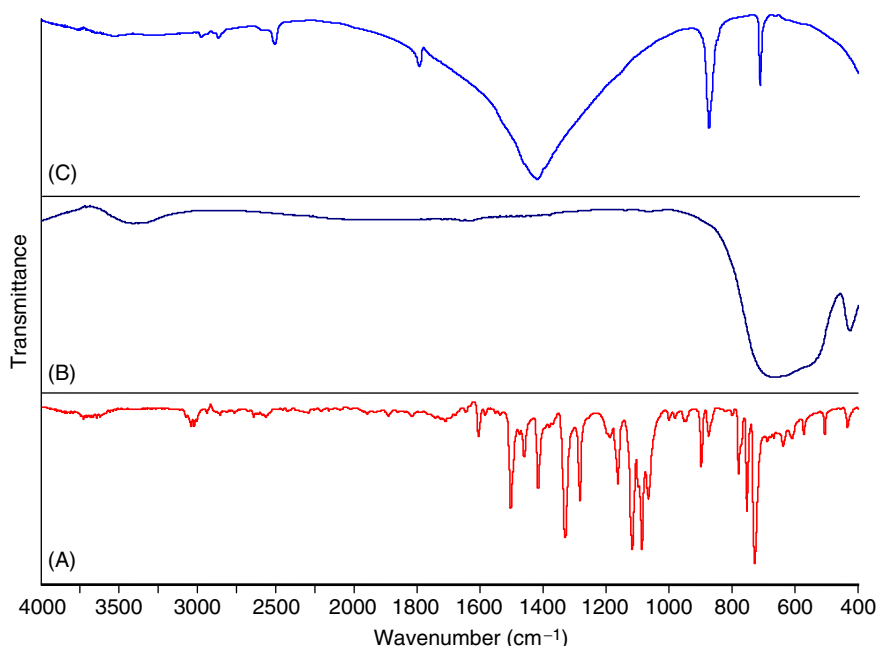
The use of selected volumetric methods permits quantitative determination of some functional groups such as carboxylic acid, hydroxyl or isocyanate groups by means of acid–base titrimetric, saponification, or condensation reactions. From these data the content of polymer in the binder can be established.

Among techniques employed in paint analysis, spectrophotometric methods (infrared (IR), Raman, and UV–visible), chromatography (both gas and liquid), X-ray diffraction and fluorescence, atomic spectrometry, nuclear magnetic resonance, and pyrolysis are the most useful and provide more selective or more sensitive information than the classical procedures about the main components of paints. In the following sections the main merits and applications of these techniques will be described.

**Infrared spectroscopy** IR spectroscopy provides the most valuable information on the qualitative composition of paints. Binders, pigments, solvents, and some additives can be characterized by their corresponding IR spectra. However, the identification of components can be facilitated if the IR spectra are obtained after separation.

For the characterization of the binder, several drops of the vehicle, previously separated from the pigment, can be dried on the surface of an appropriate crystal and the corresponding spectrum recorded. From the presence or absence of typical bands of the different functional groups in this spectrum, the type of resin employed can be identified. From the IR spectra it is also possible to analyze quantitatively mixtures of resins, such as those of the polyester and epoxy type.

To obtain the IR spectrum, the pigment can be prepared as a potassium bromide pellet using small amount of pigments previously separated by centrifugation. **Figure 3** shows the IR spectra of organic and inorganic pigments compared with an extender. In the case of organic pigments a well-defined spectrum can be seen that provides much information; for inorganic pigments only a small number of bands are found. Inorganic pigments with polynuclear anions present very similar spectra in spite of the presence of different cations. However, IR spectrometry can be used for the quantification of several

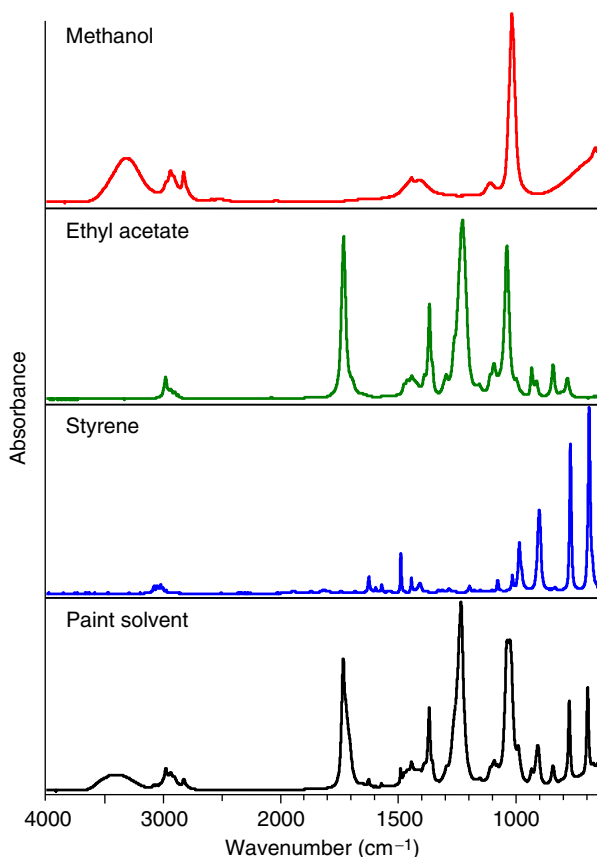


**Figure 3** FTIR spectra of pigments and charges in potassium bromide pellets: (A) organic pigment (copper phthalocyanine type  $\beta$ ); (B) inorganic pigment (titanium dioxide); and (C) charge (calcite).

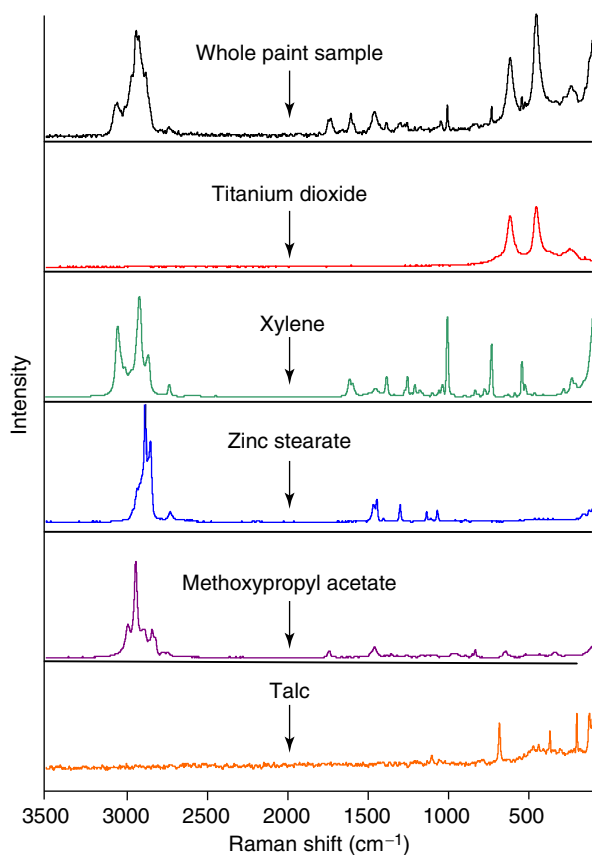
inorganic charges such as barium sulfate, calcium carbonate, and talc, with the use of a suitable internal standard.

Solvents employed in paint manufacture can be qualitatively and quantitatively determined using liquid cells or from the spectra obtained from a few drops of the solvent, previously distilled, deposited between two crystals of an alkyl halide (in this case, for quantitative purposes an internal standard or the bands quotient approach must be employed). From these spectra the presence of several solvents in the formulation can be established and, if the mixture is not too complex, quantitative determination of all solvents can be carried out without chromatographic separation. Attenuated total reflectance (ATR) measurements offer a simple and fast alternative to the use of transmittance cells for qualitative and quantitative analysis. For this purpose, the solvent fraction obtained from the paint is placed over the surface of an ATR crystal (usually ZnSe). **Figure 4** shows the spectra obtained for pure solvents and the solvent fraction of one polyester type paint.

**Raman spectroscopy** Raman spectroscopy offers information complementary to that obtained by IR spectroscopy. Problems related to sample fluorescence and poor sensitivity are the main drawbacks, but Raman spectroscopy requires a minimum



**Figure 4** FTIR spectra of pure solvents and a typical solvent mixture for polyester-type paint. Measurement mode: attenuated total reflectance.



**Figure 5** FT-Raman spectra of a whole solvent paint and different components.

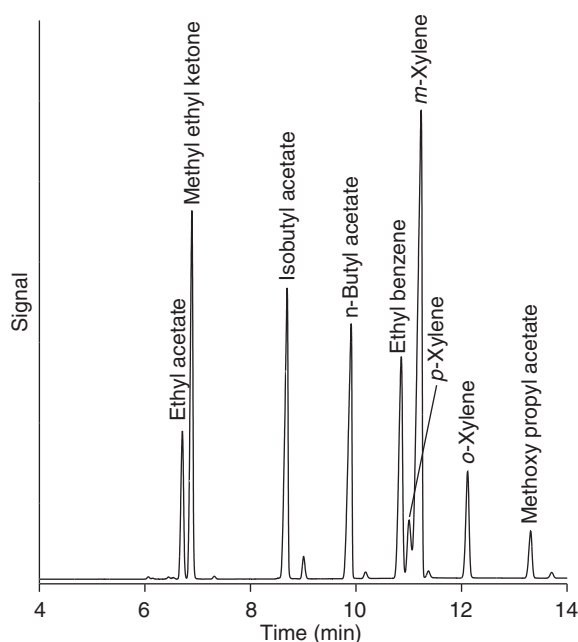
manipulation of samples and the spectra can be obtained directly from liquid paint samples contained in glass vials or using remote optical fibers.

The use of near-FT-Raman spectrometers, usually with a YAG laser ( $\lambda$  excitation of 1064 nm), reduces fluorescence enhances signal-to-noise ratio as compared with dispersive spectrometers by increasing the number of cumulated scans per sample. **Figure 5** shows the FT-Raman spectra of a paint sample and different components, all directly obtained in the back-scattering mode.

**Chromatographic methods** Gas chromatography (GC) is the most useful technique for quantitative analysis of solvent mixtures in paint samples. The solvent, after isolation from the other components of the paint, is injected directly into the chromatograph.

Either capillary or packed columns can be used. However, when samples contain naphtha solvent (a complex mixture of hydrocarbons) the use of capillary columns is required to obtain adequate separation of all the solvents.

Headspace GC permits direct analysis of paint samples without prior separation, providing a rapid



**Figure 6** Headspace gas chromatogram obtained from a solvent-based paint. Instrumental conditions: fused-silica column (0.20 mm ID, 50 m length) with Carbowax 20 M stationary phase; carrier gas helium (at  $0.6 \text{ ml min}^{-1}$ ); FID detector.

method for quantitative analysis of solvent (see **Figure 6**).

Detection systems usually employed in GC for the analysis of solvents in paints are the flame ionization detector (FID) and the electron-capture detector, but coupled GC and Fourier transform infrared spectroscopy (FTIR) or GC with mass spectrometry (GC-MS) offer improved peak identification capability.

Liquid chromatography (LC) can effectively replace GC in the analysis of paint components with low volatility. Among LC techniques the most interesting in paint analysis is gel permeation chromatography (GPC), which provides an interesting tool for the characterization of binders from the study of the composition and molecular mass distribution of the resins.

The analysis of binders by GPC requires the prior isolation of the binders and their dissolution in an appropriate solvent (tetrahydrofuran is one of the most frequently used). The separation of different oligomers of the resins takes place in a series of columns arranged in increasing order of porosity. A refractometer is commonly used as the detection system. FTIR can be used as a selective detector but it has some limitations due to the nature of the mobile phase employed. Open columns can be used for the GPC analysis of polymers with a high molecular

mass, but for low molecular mass polymers open columns do not provide sufficient resolution and high-performance GPC is absolutely necessary. Typical gel permeation chromatograms of polyester and an epoxy resin are shown in Figure 7.

Thin-layer chromatography can also be applied to the separation and qualitative characterization of organic pigments, but meager information is provided by this technique compared with spectrometric techniques and GC or LC.

**Pyrolysis** Pyrolysis is really a sample preparation technique and not an analytical technique. However, the interest in pyrolysis for the characterization of polymeric materials justifies some comments.

Pyrolysis is based on the rapid heating of the sample under controlled conditions and online identification of the decomposition products. GC, FTIR, and MS are used for the analysis of the pyrolyzed samples.

For the determination of binders by pyrolysis a small amount of sample, prepared by removing the pigment by centrifugation and evaporating the solvent, is introduced into a pyrolysis chamber, in which it is heated under a carrier gas flow. Volatile products obtained from pyrolysis are introduced

into a gas chromatograph to separate and identify them. Polymers with well-characterized composition are used to compare the chromatogram of the sample with the different types of resins employed as binders.

This procedure provides poor reproducibility, because the pyrolysis products depend strongly on the experimental conditions, but provides a useful analytical tool in the identification of acrylic and vinylic polymers or monomer units from polyester, polyurethane, and phenolic and polyacrylate resins.

#### X-ray fluorescence and X-ray diffraction techniques

X-ray fluorescence (XRF) spectroscopy is useful for qualitative elemental analysis of paint samples. It does not require dissolution of the sample and can be applied to dry films. When an energy-dispersive instrument is employed, XRF provides rapid information on the presence of elements of atomic number higher than or equal to 12 (e.g., above magnesium). However, from a quantitative point of view, the sensitivity, accuracy, and reproducibility of XRF measurements is lower than that of flame, electrothermal, or plasma atomic spectrometry.

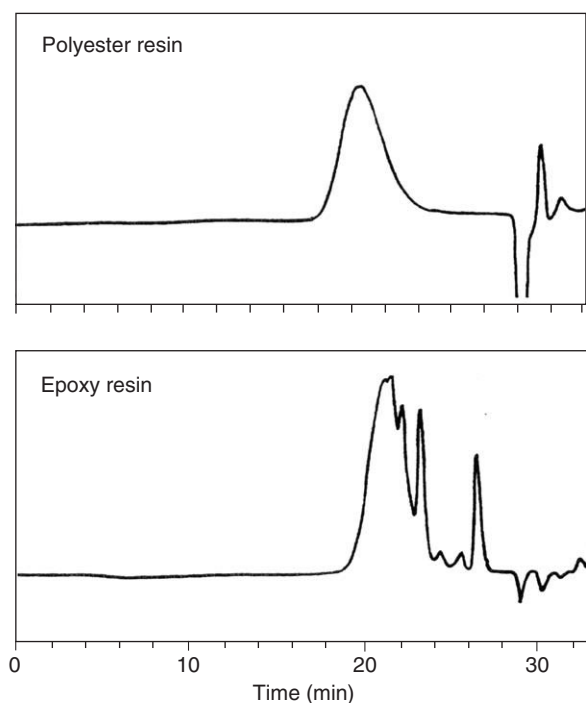
Figure 8 shows the XRF spectra of different solvent-based paints obtained as dry films.

X-ray diffraction is a powerful technique for discriminating inorganic pigments and extenders some specific applications are very interesting in the field of paints. Sample preparation is simple and the compounds can be identified directly and not indirectly through their elements.

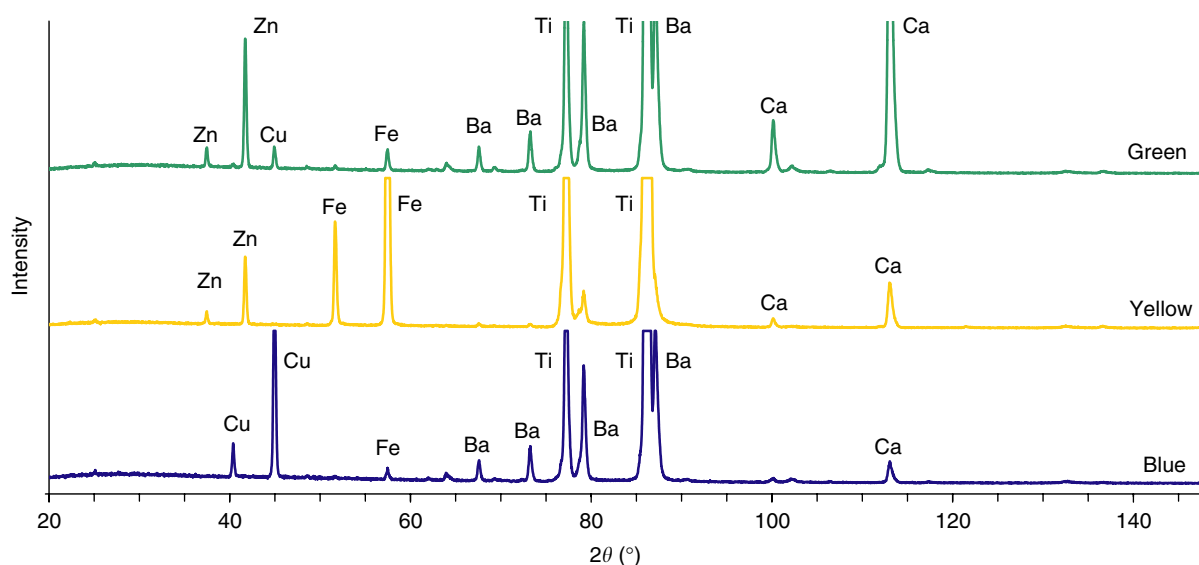
**Atomic spectrometry** The use of modern analytical techniques based on atomic absorption and emission spectrometry provides accurate and precise quantitative determination of inorganic species in paints.

Atomic absorption spectrometry (both using flame and electrothermal atomization) and plasma emission spectrometry, especially inductively coupled plasma-atomic emission spectrometry (ICP-AES) provide selective methods for quantification of inorganic species at different concentrations. These techniques are important for the analysis of inorganic pigments, charges (especially inorganic extenders), and additives such as catalysts, driers, and antifouling compounds.

Atomic spectrometry generally requires prior dissolution of the sample, which can be carried out with either acids or organics solvents, but in some cases necessitates destroying the matrix by means of a wet acid treatment or a dry digestion. This can be a serious drawback, but the new strategies for sample preparation, based on the use of microwave-assisted digestion procedures for sample dissolution and



**Figure 7** Liquid gel permeation chromatogram of polyester and epoxy resins. Columns: Waters Ultrastayage  $10^4$  Å, 500 Å, and 100 Å. Solvent carrier: tetrahydrofuran at  $1 \text{ ml min}^{-1}$ . Oven temperature:  $36^\circ\text{C}$ . Detector: differential refractometer.



**Figure 8** Dispersive X-ray fluorescence spectra of dry films of different lacquers. Experimental conditions: rhodium anticathode; stainless steel holder; LiF200 analyzer crystal; 150  $\mu\text{m}$  collimator; and flow detector.

direct introduction of solid samples as slurries or emulsions, can reduce problems involved in sample preparation.

**Nuclear magnetic resonance spectroscopy**  $^1\text{H}$  and  $^{13}\text{C}$  nuclear magnetic resonance (NMR) are effective for the characterization of resins in the binders of paints.

Binders for NMR analysis must be isolated from the sample and dissolved in deuterated chloroform (a very good solvent for many resins as it produces only small proton impurity peaks and is less expensive than other deuterated solvents). The NMR spectra of binders provide information about the functional groups present.

Using flow charts for paints the type of resin employed can be identified, and in some cases, for example, alkyl and phenolic resins, quantitative determination of resins can be made from the relative areas of the spectrum peaks.

$^{13}\text{C}$  NMR provides better results than  $^1\text{H}$  NMR for the resolution of peaks obtained from high molecular mass and very viscous polymers, but the instrumentation required is more expensive and a larger amount of sample and longer analysis times are involved.

**UV-visible spectrophotometry** UV-visible spectrophotometry can be employed for identification of pigments, especially for the characterization of organic compounds or the determination of light

filters. The technique is based on the possibilities offered by an appropriate combination of solvent extraction or the selective solubility of each pigment and its characteristic absorbance spectra. The use of mathematical procedures for the deconvolution of spectra and the use of derivatives can improve the selectivity of these determinations.

### Recent Developments for Quality Control of Paints

Quality assurance in the paint industry demands not only quality control of final products, but also a careful control of raw materials and manufacturing processes. Paint manufacturers require simple and fast analytical procedures for online control. It is important to obtain measurements directly from whole samples, without any prior treatment, and the incorporation of sample preparation techniques, such as supercritical fluid extraction or accelerated solvent extraction, which facilitate the separation of paint components prior to their analysis.

For process control, analytical techniques such as vibrational spectroscopies, which provide information about paint composition, are important based on: (1) the popularization of FT instruments, with a better signal-to-noise ratio and fast data acquisition speed; (2) reflectance techniques (total attenuated and diffuse), photoacoustic spectroscopy, and the development of optical fiber-based devices that provide easy spectrometric measurements on crude samples; (3) vapor-phase generation coupled with FTIR for fast analysis of the volatile paint fractions;



and (4) the improvement of database and library tools.

Finally, it is necessary to mark the importance of chemometrics for the exploitation of instrumental data from paint samples. From the treatment of vibrational spectra (near-IR, IR, and Raman) or chromatographic signals and using calibration models, established from well-known paint samples, chemometrics provides a powerful tool for the prediction and control of the properties of paints throughout their manufacturing.

See also: **Atomic Absorption Spectrometry:** Flame; Electrothermal. **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Color Measurement.** **Forensic Sciences:** Paints, Varnishes, and Lacquers. **Gas Chromatography:** Pyrolysis. **Infrared Spectroscopy:** Industrial Applications. **Liquid Chromatography:** Size-Exclusion. **Paints:** Water-Based. **Spectrophotometry:** Organic Compounds. **X-Ray Absorption and Diffraction:** X-Ray Diffraction – Powder. **X-Ray Fluorescence and Emission:** Wavelength Dispersive X-Ray Fluorescence; Energy Dispersive X-Ray Fluorescence.

## Further Reading

Brandau AH (1990) *Introduction to Coating Technology*. Philadelphia: Federation of Societies for Coating Technology.

Brezinski DR (ed.) (1991) *An Infrared Spectroscopy Atlas for the Coatings Industry*. Blue Bell: Federation of Societies for Coating Technology.

Davies IL (1985) Instrumental analysis on the coating industry. *Journal of the Oil and Colour Chemists Association* 65(5): 109–115.

Freitag W and Stoye D (eds.) (1998) *Paints, Coatings, and Solvents*. Weinheim: Wiley-VCH.

Golton WC (ed.) (1992) *Analysis of Paints and Related Materials: Current Techniques for Solving Coating Problems*. West Conshohocken: ASTM International.

Koleske JV (ed.) (1995) *Paint and Coating Testing Manual: Garner-Sward Handbook*, 14th edn. West Conshohocken: ASTM International.

Lambourne R and Strivens TA (eds.) (1999) *Paint and Surface Coatings: Theory and Practice*. Cambridge: Woodhead Publishing Limited.

Lear JB (1981) Analysis of paint. *Journal of Coating Technology* 53(674): 51–57.

López-Anreus E, Garrigues S, and de la Guardia M (1998) Simultaneous vapour phase Fourier transform infrared spectrometric determination of butyl acetate, toluene, and methyl ethyl ketone in paint solvents. *Analyst* 123: 1247–1252.

Marshall M (1983) NMR analysis of paint media. *Journal of the Oil and Colour Chemists Association* 63(10): 285–293.

Morgans VM (1990) *Outlines of Paint Technology*. Chichester: Wiley.

Wichs ZW, Jones FN, and Pappas SP (1999) *Organic Coating: Science and Technology*. Chichester: Wiley.

# PARTICLE SIZE ANALYSIS

M Wedd, Lower Broadheath, Worcs, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Determining particle size distributions are one of the key elements in ensuring product performance and consistency. The range of materials and processes for which particle size distributions are important is vast. Many materials cross national boundaries and thus it is necessary to declare protocols of measurement in order to avoid unnecessary disputes.

A range of ISO standards (see Appendix A) to ensure international agreement are being created.

The materials range in size from nanometers to millimeters and many techniques have been invented to perform the particle size measurement.

The more commonly encountered methods will be described in this section. The modern equipments available for particle size measurements rely upon sophisticated software. However, it should be remembered that most of these do not measure particle size directly but rely upon the interpretation of some other parameter.

How this interpretation is conducted together with any attendant assumptions bears considerably upon the final result. With this in mind each measurement method covered will contain a description of the operational method.

Crushed road stone and gravels are made available in millimeter size fractions created by screening. Sands for building and sands for molds used in metal casting foundries are in the fractions of a millimeter range. Portland cement is ground to achieve a broad range of sizes from submicrometer to  $\sim 200\ \mu\text{m}$ . This is to ensure a fairly rapid onset of gelling whilst maintaining a desirable 28-day strength.

and (4) the improvement of database and library tools.

Finally, it is necessary to mark the importance of chemometrics for the exploitation of instrumental data from paint samples. From the treatment of vibrational spectra (near-IR, IR, and Raman) or chromatographic signals and using calibration models, established from well-known paint samples, chemometrics provides a powerful tool for the prediction and control of the properties of paints throughout their manufacturing.

See also: **Atomic Absorption Spectrometry:** Flame; Electrothermal. **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Color Measurement.** **Forensic Sciences:** Paints, Varnishes, and Lacquers. **Gas Chromatography:** Pyrolysis. **Infrared Spectroscopy:** Industrial Applications. **Liquid Chromatography:** Size-Exclusion. **Paints:** Water-Based. **Spectrophotometry:** Organic Compounds. **X-Ray Absorption and Diffraction:** X-Ray Diffraction – Powder. **X-Ray Fluorescence and Emission:** Wavelength Dispersive X-Ray Fluorescence; Energy Dispersive X-Ray Fluorescence.

## Further Reading

Brandau AH (1990) *Introduction to Coating Technology*. Philadelphia: Federation of Societies for Coating Technology.

Brezinski DR (ed.) (1991) *An Infrared Spectroscopy Atlas for the Coatings Industry*. Blue Bell: Federation of Societies for Coating Technology.

Davies IL (1985) Instrumental analysis on the coating industry. *Journal of the Oil and Colour Chemists Association* 65(5): 109–115.

Freitag W and Stoye D (eds.) (1998) *Paints, Coatings, and Solvents*. Weinheim: Wiley-VCH.

Golton WC (ed.) (1992) *Analysis of Paints and Related Materials: Current Techniques for Solving Coating Problems*. West Conshohocken: ASTM International.

Koleske JV (ed.) (1995) *Paint and Coating Testing Manual: Garner-Sward Handbook*, 14th edn. West Conshohocken: ASTM International.

Lambourne R and Strivens TA (eds.) (1999) *Paint and Surface Coatings: Theory and Practice*. Cambridge: Woodhead Publishing Limited.

Lear JB (1981) Analysis of paint. *Journal of Coating Technology* 53(674): 51–57.

López-Anreus E, Garrigues S, and de la Guardia M (1998) Simultaneous vapour phase Fourier transform infrared spectrometric determination of butyl acetate, toluene, and methyl ethyl ketone in paint solvents. *Analyst* 123: 1247–1252.

Marshall M (1983) NMR analysis of paint media. *Journal of the Oil and Colour Chemists Association* 63(10): 285–293.

Morgans VM (1990) *Outlines of Paint Technology*. Chichester: Wiley.

Wicks ZW, Jones FN, and Pappas SP (1999) *Organic Coating: Science and Technology*. Chichester: Wiley.

# PARTICLE SIZE ANALYSIS

M Wedd, Lower Broadheath, Worcs, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Determining particle size distributions are one of the key elements in ensuring product performance and consistency. The range of materials and processes for which particle size distributions are important is vast. Many materials cross national boundaries and thus it is necessary to declare protocols of measurement in order to avoid unnecessary disputes.

A range of ISO standards (see Appendix A) to ensure international agreement are being created.

The materials range in size from nanometers to millimeters and many techniques have been invented to perform the particle size measurement.

The more commonly encountered methods will be described in this section. The modern equipments available for particle size measurements rely upon sophisticated software. However, it should be remembered that most of these do not measure particle size directly but rely upon the interpretation of some other parameter.

How this interpretation is conducted together with any attendant assumptions bears considerably upon the final result. With this in mind each measurement method covered will contain a description of the operational method.

Crushed road stone and gravels are made available in millimeter size fractions created by screening. Sands for building and sands for molds used in metal casting foundries are in the fractions of a millimeter range. Portland cement is ground to achieve a broad range of sizes from submicrometer to  $\sim 200\ \mu\text{m}$ . This is to ensure a fairly rapid onset of gelling whilst maintaining a desirable 28-day strength.

Powdered drugs suitable for the treatment of asthma are 'micronized' to 2–3  $\mu\text{m}$ . Much finer particles would be exhaled and thus wasted whilst much larger particles would remain in the bronchial tubes and not reach the deep lung where the drug is most effective.

Particles destined for use in sun-block creams are created to have a size of 70–80 nm.

Studies of proteins require sizing instruments capable of measuring down to a few nanometers.

Agricultural sprays are required to be atomized within size limits. Very large droplets are wasteful of the chemical, whilst very fine droplets can be carried away by the wind and cause damage or destruction of neighboring crops.

The fuel efficiency and emission levels of diesel and petrol engines are intimately dependant upon fuel atomization.

Products are also supplied in the form of emulsions, e.g., paints, adhesives, bitumens, and some medicines whilst others are created as suspensions, e.g., casting slips and coal slurries for burning in electricity power plants.

## Particle Size Distributions

Particle sizing presents a challenge in as much it attempts to provide a single value of size for what are usually three-dimensional objects. The sphere is the only shape for which the single value of diameter fully characterizes its size.

The quantity of particles that have a specific size is also required (Figure 1).

A person with a headache who takes an analgesic medicine in powdered form receives a defined volume or weight of the drug. The powder is likely to have a distribution of sizes. The rate of dissolution

will be greatly influenced by volume of drug that has the finer sizes due to the larger surface area presented by these sizes. For dissolution studies a quantity axis in terms of surface area against size is important.

In polishing a surface, one or two rogue particles in the polishing material can ruin the whole process with scratches. The number of objects having the size of the grit is important here.

Instruments and techniques providing particle size distributions have different value judgments in terms of both the particle size reported and in terms of the quantity axis.

Instruments and techniques will only provide common estimates for these values in the case of homogeneous spheres having defined physical properties.

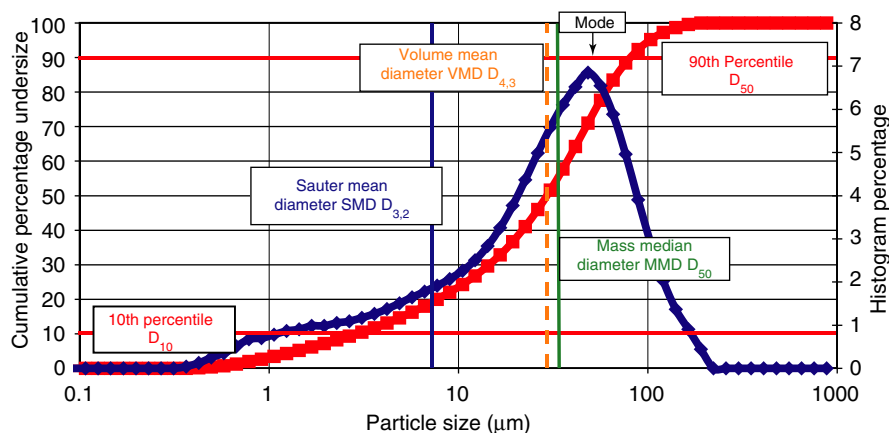
Therefore, any published particle size distribution should also contain a full description of the method of preparation of the material and the method by which the size distribution was determined.

The ISO standards (Appendix A) not only define methods of validation of the units described but also provide excellent advice on their general use and on the preparation of the material to be measured.

The relationship between the number of particles that have a specific size and the equivalent volume is proportional to the inverse of the particle diameter, cubed ( $1/D^3$ ). Therefore, if knowledge of the particle size distribution by number is required, a technique that examines each particle is chosen and the value added to a specific size class.

If a size distribution by volume is required, a technique that responds principally to the volume of material that has a specific size is chosen.

Attempts to numerically transform from number to volume or vice versa are to be discouraged for any size distribution whose dynamic range of sizes exceeds 8:1.



**Figure 1** A typical particle size distribution. The quantity axis is by volume.

Modern particle size instruments have been designed to be easy to use. However, considerable additional skills and understanding is required before reliable results can be obtained. These skills include subsampling and dispersion of aggregates. The understanding of how to transport all the sizes of particles to the measurement zone without introducing bias and the avoidance of introducing additional, unwanted, artifacts such as dirt and air bubbles is also needed.

## Particle Size Instruments

Table 1 provides a brief synopsis of the instruments and their properties that are in more common use.

Both image analysis and electrical zone sensing observe and respond to each particle. Laser diffraction, sedimentation, and dynamic light scattering data are obtained from ensembles of dispersed particles in suspension.

## Laser Diffraction

Introduced in the late 1970s, laser diffraction is now the most widely accepted method of general-purpose particle sizing. The technique has wide application, is nonintrusive, easy to use, and does not require calibration.

The ISO standard ISO 13320-1, Particle size analysis – Laser diffraction methods – Part 1: General principles, covers this method.

Particle sizing from 0.05 to 2000  $\mu\text{m}$  is available within a single instrument, in most modern implementations, for liquid suspensions of particles, dry powders, and droplet sprays. Robust versions of the technology are now to be found online in hostile environments, typical of which is the manufacture of cement.

## Theory

A dilute suspension of particles or droplets is caused to flow through a measurement zone (see Figure 2).

**Table 1** Principal features of particle sizing instruments in common use

	<i>Laser diffraction</i>	<i>Dynamic light scattering</i>	<i>Sedimentation</i>	<i>Electrozone sensing</i>	<i>Image analysis</i>	<i>Sieving</i>
Size range ( $\mu\text{m}$ )	0.05–2000	0.01–3, limited by density	0.1–100	0.2–200	Visible, 0.5–500; electron, 0.01–10	10 $\mu\text{m}$ to several millimeters
Dynamic range	1000:1	10:1	20:1	40:1	10:1 per image	100:1
Principal reported quantity	Volume	Intensity weighted distribution	Mass	Number and summed volume	Number	Mass
Property measured	Angular light scattering	Light intensity fluctuations	Settling speed	Electrical pulse	Image size	Mass
Time required for one completed measurement	1–2 min	5–10 min	6–60 min, depending upon size range	10–15 min	10–30 min	Dry 10–30 min; wet 1–2 h
Resolution	Medium to high	Low	Medium	High	High	Medium
Calibration required	No	No	No	Yes	Yes	No
Validation method	Certified reference materials	Certified reference materials	Certified reference materials	Certified reference materials	Certified reticle	Optical verification of mesh size
Suspension media	Liquids or gases	Liquids	Liquids	Liquids	Particles deposited on a substrate	Air or liquid
Dry powder measurement	Yes	No	No	No	Yes	Yes
Droplet sprays	Yes	No	No	No	Possible	No
Concentration	Volume fraction <0.01, limited by multiple scattering	Very dilute	Upper volume fraction limited to avoid hindered settling	Dilute to avoid coincidence	Limited number per slide	Several grams of sample required
Density limitations	None	Upper size density limit that induces sedimentation	Upper size density limit when settling induces turbulence	Limited	None	None

A laser beam is passed through the measurement cell and brought to a waist of focus at the plane of the multielement detector. This is the reverse Fourier optical setup. Conventional Fourier optics arrangements are also used.

It is required that measurements of the angular dependence of scattering be conducted over as large a range of angles as possible. The laser beam may be polarized facilitating measurements of the polarization dependence of the scattered light.

The extinction or obscuration due to the presence of the particles is additionally measured. Finally, the scattering dependence with wavelength can be established using additional wavelength sources.

Gustav Mie formulated a rigorous solution, predicting the scattering from a homogeneous spherical particle, by solving Maxwell's equation. With the advent of modern PCs, Mie theory is the preferred model for predicting how particles scatter light. Knowledge of the refractive index of the particles, both real and imaginary, together with the refractive index of the suspending medium enables a matrix predicting the anticipated signals from each detector

for all size classes to be precalculated and stored in memory.

A typical energy response from latex beads in water is shown in Figure 3.

We can see from Figure 3 that large particles scatter light at small angles and small particles scatter light to greater angles. We also note that strong amplitude dependence is evident below  $\sim 10\ \mu\text{m}$ .

Below  $\sim 0.3\ \mu\text{m}$  additional data in the form of information on wavelength, polarization, and extinction dependence is needed.

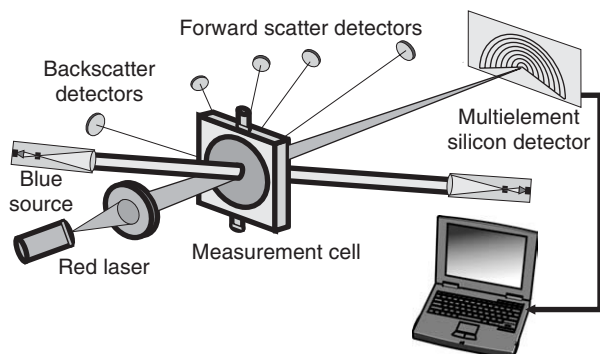
The approximation of Fraunhofer has been employed in some implementations. However, it does not include amplitude and polarization dependence and also ignores light transmitted through the particle. Thus, the Fraunhofer approximation is prone to predicting incorrect quantities of small particles having a specific size. A matrix of the predicted light scattering data is precalculated and stored in memory.

The ensemble of scattering from the sample presented in the measurement cell is sampled many times to form a stable average of the light contribution to each detector.

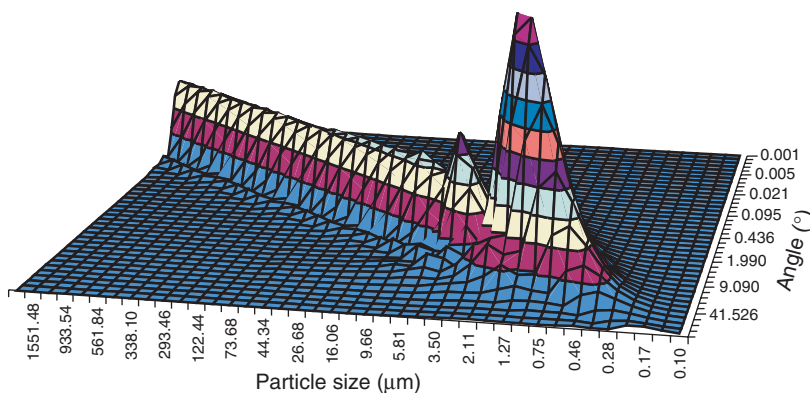
A direct matrix inversion, with a constraint to obtain a particle size distribution, is adopted in some implementations. Other solutions to the inversion problem are achieved by iteration of the particle size distribution until a 'best fit' between observation and prediction light energy is achieved. This necessary constraint limits the resolution.

Designers of laser diffraction units strive to achieve a design that results in the predicted energy matrix being optimal and consistent with the light energy contribution from the presence of each size class.

Suspension concentrations are produced such as to ensure single scattering conditions. Some progress has been made in order to accommodate the influence of multiple scattering.



**Figure 2** An example of the arrangement of components in a laser diffraction unit.



**Figure 3** Light energy predicted over  $140^\circ$  of log-spaced angle. Plotted with log-spaced size range  $0.1\text{--}2000\ \mu\text{m}$ . The vertical axis represents the energy recorded by an idealized detector.



## Applications

Almost any form of finely divided powder, droplet, or emulsion can be measured using laser diffraction. This in turn leads to the technique having penetrated into almost every branch of industry.

Applications include the following material systems: powdered materials either dry or in suspensions; continuous or pulsed droplet sprays, emulsions, and some solids in solids provided that the suspending medium is transparent.

## Photon Correlation Spectroscopy (Submicrometer Particle Sizing)

### Introduction

As in so many other fields, advances in measurement, in this case submicrometer particle sizing using laser light scattering, only became a reality when suitable technology became available to exploit the basic theories and mathematics that were already available.

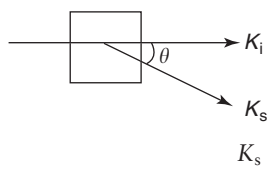
By 1968, the laser was an established light source as was single photon detection with the photomultiplier tube.

The digital correlator for processing single photon detection from a photomultiplier tube proved invaluable in retrieving useful information from the very weak scattering encountered with very dilute systems of nanometer-sized particles.

The International Standard ISO 13321:1996 'Particle size analysis – Photon correlation spectroscopy (PCS)', provides for methods of validation as well as good general advice on the use of this method. This standard is being amended to include frequency analysis techniques as well as addressing measurements at concentrations greater than very dilute.

### Theory

Consider a dilute suspension of nanometer size molecules or particles illuminated by a monochromatic coherent light source of defined wavelength. Provided the particles have a refractive index that is different from the suspending medium light scattering will be observed.  $K_s$  is the modulus of the scattering vector,  $n$  the refractive index of the medium,  $\lambda$  the wavelength of illumination, and  $\theta$  the angle of observation.



The wavelength of the scattered light from each particle is identical to the incident beam and its phase

is initially coincident with that of the incident beam. The amplitude of the light scattered by each particle depends upon its scattering cross-section and the difference in refractive index from that of the suspending medium. If the particle size is much less than the wavelength, the scattering is assumed to emanate from a point source (i.e., isotropic radiation).

If the viewing screen is placed some distance from the measurement cell and positioned at a small angle from the incident beam, a speckle pattern will be observed due to the interference pattern created from the scattering of the ensemble of particles within the measurement cell.

The size of each speckle is determined by the wavelength, observation angle, and the geometry of the incident beam. The aperture of any single photon detector should ideally be equal to the size of a single speckle. This is defined as coherent detection; that is, particles will be moving (diffusing) under the influence of Brownian motion. This means that the speckle pattern is moving about upon the screen at a rate determined by the rate of particle diffusion, the wavelength of illumination, and the angle of observation.

The output of the coherent detector, positioned in the plane of the screen, will now take the form of a fluctuating intensity with respect to time.

The output of the single photon detection system electronics may be such as to form equal height, equal width pulses for each photon detected, whose output rate is high for high-intensity illumination and low at low-light levels. This form of signal is an ideal input for the digital autocorrelation process.

Within the Physics group of RSRE (now QinetiQ), in Malvern, Worcestershire, England, Dr Roy Pike, FRS, and his team created a Digital Correlator to process the photon output of the photomultiplier such that submicrometer particle sizing, using a laser, became a reality.

The autocorrelation function has the form of

$$G_2(\tau) = [I(t) \cdot I(t + \tau)]$$

The square brackets refer to an average value of the product  $I(t) \cdot I(t + \tau)$  where  $\tau$  is a time difference independent of the arbitrary time  $t$ .

For the case of a large number of monosized-dispersed particles in the measurement cell,  $G_2(\tau)$  has the form of a single exponential decaying function of time difference  $\tau$ .

$$G_2(\tau) = A[1 + B \exp(-2\Gamma\tau)]$$

The decay constant  $\Gamma$  relates to the rate of the translational diffusion coefficient  $D$

$$\Gamma = DK_s^2$$

The diffusion coefficient of noninteracting spherical particles of diameter  $x$  is provided by the



Stokes–Einstein relationship:

$$D = \frac{kT}{3\pi\eta x}$$

where  $k$  is Boltzmann's constant,  $T$  the absolute temperature, and  $\eta$  the viscosity of the medium.

Early designs of correlator were built having linear time difference  $\tau$ . However, it was demonstrated that it was better to provide logarithmically spaced time intervals, as this offered the optimum information for subsequent data analysis.

A process of normalization of the measured autocorrelation function is applied. Monitor channels of the input count and total number of samples taken provide the data for the normalization process.

A further refinement, referred to as symmetrical normalization, is beneficially applied.

In practice, samples presented for measurement will have a distribution of sizes.

The simplest form of postcorrelation analysis is called 'the method of cumulants' and provides for a mean particle size together with an index of polydispersity. If a defined *a priori* form of a distribution (as recommended in ISO 13321) is now assumed, the cumulants' result can now be expressed as a mean size  $x_{\text{pcs}}$  and standard deviation of a size distribution.

At this point it is necessary to point out that for particles substantially smaller than the wavelength of the illumination, the quantity of light scattered reduces as radius to the sixth power.

For a distribution of particles in this size region, any autocorrelation function is biased due to the higher photon contribution from the large particle fraction compared with that of the small sizes. This bias can be unweighted within the *a priori* distribution to provide a revised estimate of particle size  $x_{\text{vol}}$ .

For very small particles  $< \lambda/10$ , light scattering becomes proportional to  $x^6$ , and thus distributions having a range of sizes beyond 10:1 are problematic.

In order to obtain further information about the nature of the distribution, without an *a priori* assumption of its shape, more sophisticated data analysis schemes are implemented. The most popular method is called CONTIN. This is a freely available program.

Further routines follow a route of nonlinear least squares, whilst an exponential sampling method and singular value analysis routine have also been tried.

### Operational Limitations of PCS

The smallest particle size that can be measured is determined by the cleanliness of the sample preparation, the input laser power, and/or detector

quantum efficiency (scattering  $\sim x^6$ ). The cholesterol molecule, with a diameter of approximately 0.6 nm with a molecular weight of 387 Da, has been measured.

The largest particle that can be measured depends upon its density. The technique observes Brownian motion, and is perturbed by the influence of sedimentation.

The resolution of the technique should be considered as low to medium. A bi-modal mixture whose sizes are different by a factor of 2 can be resolved under favorable conditions.

An important assumption for fundamental particle sizing by PCS is that the particles be noninteracting. It has been previously stated that particles are assumed to be in a dilute suspension. In order to ensure that this condition applies, a series of measurements is recommended at increasing dilutions to demonstrate that particle interaction be absent.

To be able to conduct particle size determinations at high concentration is very advantageous. Measurements are now conducted at high concentration using a variation of the PCS technique and others. These can be conducted with particle systems having concentrations of up to 40% by weight.

It is important to realize that such particle size determinations at high concentration may or may not be influenced by particle–particle interactions and the influence of multiple scattering. It is also possible that the dynamic viscosity of the suspending medium is also hindering the particle motion leading to an oversize estimation.

The value of viscosity used to make the particle size calculation may not be the true viscosity. Used in this way the value entered for viscosity should be considered a calibration factor used to report the particle size that has been previously determined by a fully dilute experiment.

### Applications

These have thus far included studies of the following systems: proteins, microemulsions, colloids, copolymers, micelles, liposomes, fibrinogen, internal molecular motions, liquid interfaces, fatty acids, viruses, bacteria, vesicles, viscosity, lipids, motile cells, enzymes, lipoprotein, polyelectrolytes, spores, liquid crystals, glass transmissions, sols, microgels, soot, blood plasma, nanoparticles, swelling latex, gene delivery, and intravenous fat emulsions.

### Sedimentation

This method of measuring particle size distributions exploits the way that objects settle very slowly in a fluid under gravity or centrifugal force.

The International Standard ISO 13317, parts 1–3, inclusive cover this method.

Particles of sand or mud are transported in flowing water and settle in strata by much the same method. The sedimentation method, although not as quick or convenient as other techniques, finds favor with geologists and other mineral processors.

Sedimentation assumes a spherical particle assumption to relate settling speed with particle size that does result in a bias of the size reported for nonspherical particles.

Natural sediments consist of many differing minerals having differing shapes and densities. Despite this, considerable experience of natural sediments has been gained using sedimentation size distributions to create empirical predictive models of the movement and deposition of silts and mud.

### Theory

The illustration in Figure 4 is of a spherical particle settling very slowly in a fluid.

A particle having a higher density to that of the suspending fluid accelerates under the influence of gravity. It reaches a terminal velocity as a result of the drag force exerted by the liquid's viscosity operating upon the total wetted surface of the particle. The resulting terminal velocity is described by the Stokes equation provided the limits of creep flow ( $Re < 0.25$ ) are obeyed:

$$u_{st} = \frac{(\rho_s - \rho_l)gD^2}{18\eta}$$

where  $U_{st}$  is the Stokes settling velocity,  $\rho_s$  the particle density,  $\rho_l$  the liquid density,  $D$  the particle diameter,  $\eta$  the liquid viscosity, and  $Re$  the Reynolds number.

During the derivation of the Stokes equation the simple relationship between the mass of a

homogeneous sphere being proportional to  $D^3$ , whilst the drag due to the surface area is proportional to  $D^2$ , results in this simple equation. For irregular particles the solution is more complex and thus a spherical equivalent particle diameter is reported.

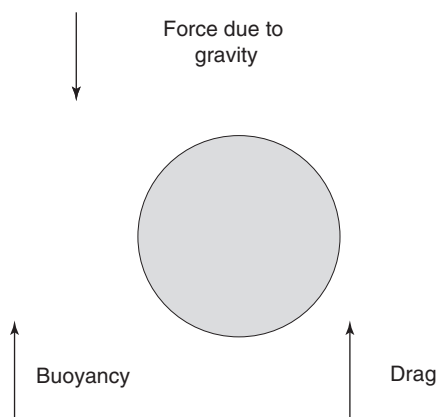
With a thin disk-shaped particle, the surface area will have a much greater influence than that of its mass. A disk-like particle will settle very much more slowly due to it experiencing a higher drag relative to its mass and thus may be undersized.

### Instrumental Implementations

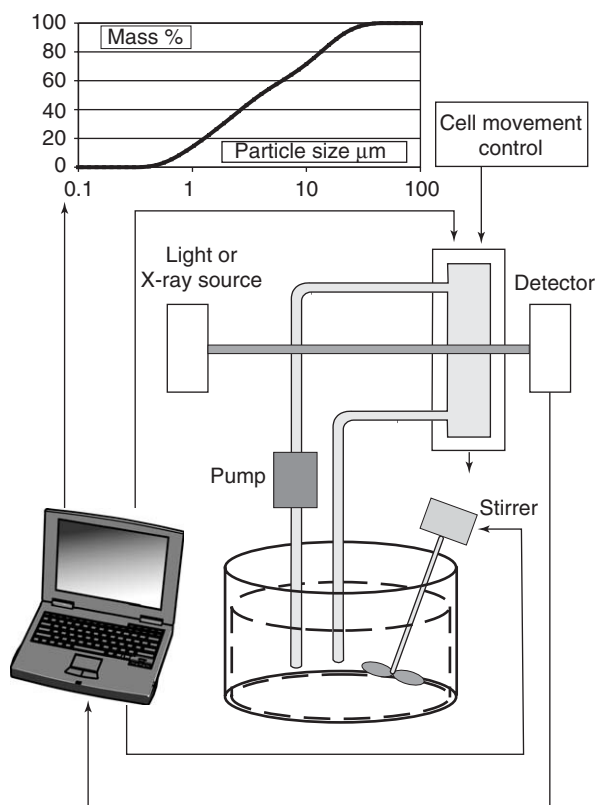
A patent was granted in 1969 for a gravity, X-ray, sedimentation design. A predispersed sample is introduced into a temperature controlled measurement cell. Observations are taken of the incremental change in concentration or density, with respect to time, of the sample at a specific point within the cell. These observations are made by either X-ray or light scattering extinction techniques (Figure 5).

There are two main modes of operation.

**Line start** Here the sample is introduced as a thin layer onto the surface of the liquid suspending medium. The line start method is more often



**Figure 4** Forces acting upon a settling spherical particle.



**Figure 5** Block diagram of a typical sedimentation unit.

applicable to centrifugal sedimentation. For some materials, where the line start can be applied, higher-resolution results may be possible.

This is due to the particles being segregated by size making the quantity determination more precise.

As the largest size particles experience the gravitational or centrifugal force the greatest they not only settle in the fluid but also are separated from their neighbors.

**Homogeneous start** Here the sample is added to the suspending medium and made homogeneous by stirring prior to the commencement of the measurement. The sample preparation may be carried out in a beaker, the suspension then being pumped into the measurement cell and the pump stopped prior to the measurement sequence.

### Practical Considerations

Natural sediments and minerals often contain heterogeneous populations of sizes, shapes, and densities. The average value of density to be employed for particle size calculations should be the specific gravity and not the absolute density. The use of specific gravity for the value of density avoids additional and unnecessary errors in the reporting of size due to the presence of surface and internal unwetted porosity.

Any gas trapped within the pores of a particle will contribute to additional buoyancy. If the creep flow requirements are to be met in the presence of samples having a distribution of sizes, then care must be taken to ensure that the largest particles settle sufficiently slowly. This is often achieved by increasing the viscosity of the suspending fluid. More viscous suspensions may well hinder the creation of reliable dispersions.

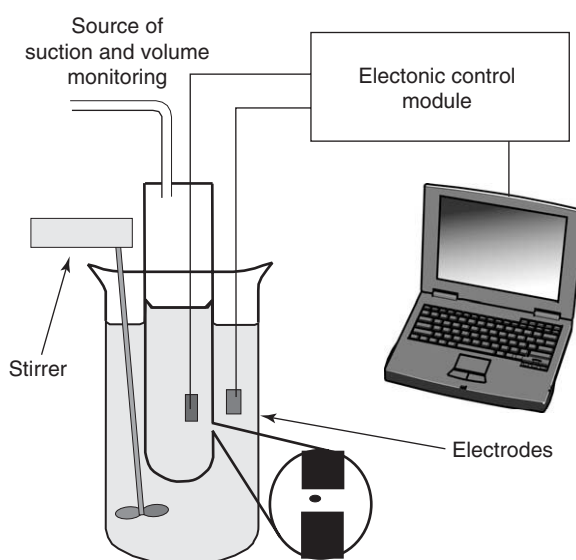
The temperature of the dispersed sample needs to be controlled to within 1°C and known in order that an appropriate value of viscosity is employed.

The duration of the measurement depends upon the width of the size distribution being evaluated. In general, the measurement time is proportional to  $1/D^2$ . In some implementations the cell is moved slowly during the data collection phase. This can reduce the overall measurement time but at some risk to the final resolution.

### Electrical Sensing Zone Method

In this method the change in conductivity of a liquid by the presence of a particle is exploited. The origins of the technique lie in a method of counting fish passing through a salmon ladder.

Particles are homogeneously suspended in an electrically conducting fluid and caused to flow through a



**Figure 6** Principal components of an electrical sensing zone unit.

small aperture in an insulating wall. The small current passing through this aperture changes when a particle is present within the aperture.

The International Standard ISO 13319:2000, Determination of particle size distributions – Electrical sensing zone method, covers this method.

The experimental setup is shown in Figure 6.

As the particle passes through the aperture, it displaces an equivalent volume of electrolyte, so the change in current or a voltage pulse derivative responds directly to the volume of the particle. The electrical response of the aperture is calibrated, either by passing a range of certified particles of known size through the instrument, or by a mass integration technique.

The use of certified reference particles provides a means of validation.

As particles are examined one by one a highly resolved distribution by number can be established. The number of particles recorded in each size class can be readily totaled to form a distribution by volume. If the total volume of suspension fluid passing through the aperture is also recorded then the number of particles per unit volume of suspension can also be obtained. This feature has provided an important tool in medicine in being able to count red, white, and platelet cells from blood samples.

Apertures having different diameters may be fitted into the instrument to accommodate samples of different size ranges. The dynamic range of particle size that can be handled by a single aperture is ~40:1.

The concentration of particles within the suspending fluid must be restricted to very dilute levels in order that coincidence of particles appearing in the aperture is kept to a very low level such that a distorted result is avoided.

If a distribution by volume is desired then further care must be exercised to ensure that an adequate number of particles are counted. This requires sufficient experiment time whilst a substantial quantity of suspension flows through the aperture.

The lower size limit is stated to be just less than  $1\text{ }\mu\text{m}$ . This restriction is imposed by the noise limit of pulse detection.

Care must be exercised when dealing with a particle that is either conducting, or that has continuous pore structures, if additional errors are to be avoided.

A substantial bibliography exists on suitable electrolyte mixtures, many of which are cited in the ISO standard.

### Image Analysis

The rapid technological advance in small computer systems, imaging cameras, and light sources is driving a major advance in this technique for particles of a few micrometers and above. Dynamic image collection and rapid data storage for immediate or subsequent analysis is rapidly replacing fixed image analysis from particles deposited upon a substrate.

Both static and dynamic image analyses are supported by ISO 13322 parts 1 and 2 currently in preparation.

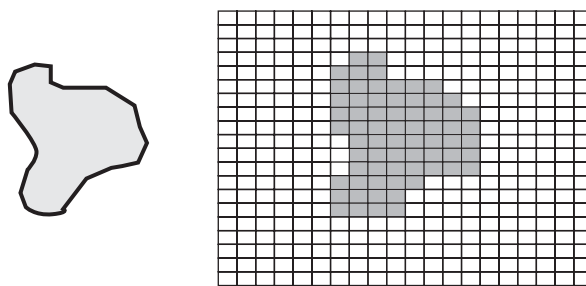
For submicrometer particles, where electron microscopy is employed, advances will be more modest. Here fixed image analysis remains the norm.

Image analysis provides a method where size distribution is obtained together with valuable shape data from essentially two-dimensional objects. Dynamic image analysis from randomly orientating particles of large numbers of two-dimensional objects offsets the influence of having only a two-dimensional observation, to some degree.

By far the majority of industrial particle size distributions are required to be presented as distributions by volume. It should be remembered that a single  $100\text{ }\mu\text{m}$  particle has the equivalent volume of 1 million,  $1\text{ }\mu\text{m}$  particles.

In the past, the image analysis technique has been found tedious and expensive on preparation time to provide a sufficiently high count of particles. This is necessary to support statistically justified distributions by volume for polydisperse samples.

The contrast of the particle image depends upon its optical properties with respect to the suspending medium and the method of illumination.



**Figure 7** Image of a particle and pixel representation.

Software mechanisms have been developed in order that a representation of the true boundaries of the particle is established. Images of touching particles are either software segregated or rejected from the analysis. Out-of-focus objects are also rejected by software techniques. A binary array, where the shadowed pixels are 1 in a background of 0, is often employed to permit rapid data storage and image analysis (Figure 7).

The image of a two-dimensional irregular shaped object can be reduced to a dimension of size by an array of methods. The choice of method will influence the final reported size distribution and varies from unit to unit. It is therefore necessary to report the method chosen if confusion of results from different machines is to be avoided.

Calibration is achieved by the use of a reference reticle.

The size range covered is set by the imaging optics together with the resolution and pixel area of the camera employed. As the particle image approaches the size of a single pixel the resolution of size degrades rapidly. This, coupled with the optical arrangement, defines a lower limit.

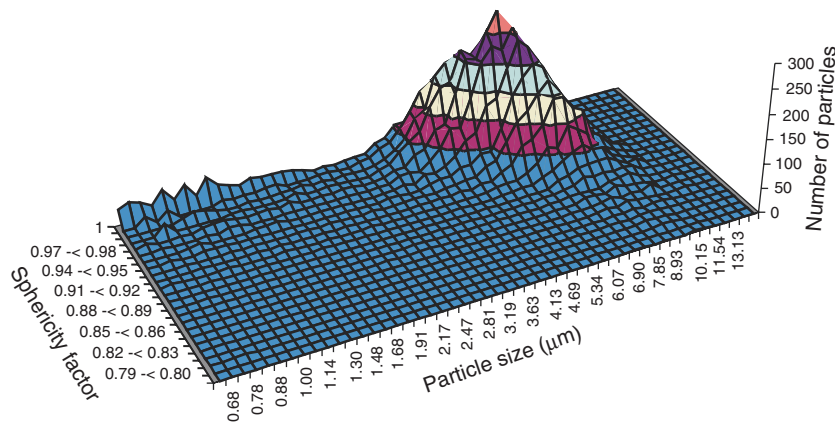
The upper size is determined as to whether the image of such a large particle is wholly captured within the frame.

Some implementations employ a line-scan camera observing a stream of dispersed particles falling at a constant velocity. The above size range restrictions continue to apply.

Particle size results are sometimes portrayed in three-dimensional form where a sphericity or shape-determining factor is displayed as the third axis. An example is shown in Figure 8. Images of the particles are often made available for visual examination.

### Sieves and Sieving

The segregation of various sizes of particles using woven cloth screens can be traced back as far as the ancient Egyptians where the method was used to separate seeds from other detritus.



**Figure 8** Three-dimensional graph of sphericity versus particle size.

Many materials such as sands, gravels, road stone are sold in specified size fractions, prepared by screening.

ISO 2591-1:1998 (Test sieving – Part 1: Methods using test sieves of woven wire cloth and perforated metal plate) covers the sieving method.

Sieves are either made from woven wire, punched plate, or are electroetched, as illustrated in Figure 9.

A substantial range of apertures is available, classified as coarse (4–100 mm), medium (0.2–4 mm), and fine (less than 0.2 mm). The sieve sizes often follow a size progression based upon the fourth root of 2. Other progressions such as a 10th root of 10 have also been employed. The lower size of woven wire sieves is  $\sim 37 \mu\text{m}$ . A specialized range of micromesh sieves is produced with sizes down to 1–2  $\mu\text{m}$ .

Aperture shapes approximate to squares for woven sieves. Punched or etched sieves can be of any shape in principle but are either square or round in practice.

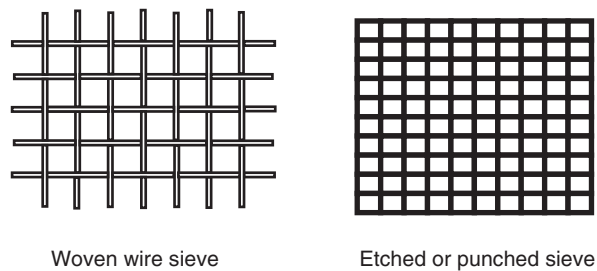
The sieving process may be carried out on either dry or wet samples, shaking by hand or by machine. Sieving is either carried out for a fixed time or until the sample passes through the sieve at a low but constant rate. This latter approach is not usually adopted for stacked sieves.

The weight of sample retained upon the sieve is required. This is sometimes achieved by the following method.

Accurate weights of the clean sieves prior to a measurement are made. A second weighing of the sieve and the retained sample fraction remaining on the sieve at the conclusion of the sieving process is also made.

The results from a sieving operation are influenced by the following factors:

- the shape of the sieve aperture;
- the load of powder to be sieved;



**Figure 9** Illustration of woven wire and punched plate sieves.

- the duration of sieving;
- the cohesiveness of the powder sample;
- the presence of fine particles;
- the particle shape;
- the friability of the sample;
- the agitation method.

The time of sieving is closely related to the sieve loading that in turn is influenced by the precision required for the retained weight of sample on each sieve. Some trial runs will be necessary before an optimum strategy for the material in question is achieved.

### Dry Sieving

The size distribution, particularly the presence of fines, and the cohesive properties of the powder are key elements to consider if sieving is to be adopted. Aggregates may be reduced by drying or by the addition of a small quantity of flow enhancing agents such as silica or stearic acid.

Air-jet sieving apparatus enhances the process of sieving by fluidizing the powder in order that it may more readily pass through a bank of sieves. Sonic sifting is also applied in some implementations.



**Wet Sieving**

This method is applied to particle systems that are already in suspension or that are prone to aggregate. The sample is washed or rinsed through a nest of sieves whilst agitation is being applied. Additionally, sonication and vibration may also be employed.

One method of analysis requires that on completion of the sieving process the retained sample on each sieve be rinsed off into a collecting dish of known tare weight and dried in an oven. The final retained weight of sample is then deduced.

**Sieve Calibration**

Spherical particles such as glass beads are often to be found retained in the mesh of woven sieves at the conclusion of a sieving process. If the size distribution of these retained particles is carefully measured using a calibrated microscope then an estimate of the range of mesh sizes for a particular sieve can be obtained.

**Sieve Cleaning**

It is during the process of cleaning that sieves often become damaged. Loose dry powder may be gently removed with a very soft camel haired brush. In order to ensure a reasonable open area, fine mesh sieves are made with finer gauges of wire and are thus more vulnerable to damage. Cleaning by gentle rinsing may be sufficient. Cleaning by immersion in an ultrasonic bath may be used; again the finer mesh sieves are vulnerable to ultrasound damage. Particles trapped within the mesh may be dissolved with some materials. The temptation with trapped particles is to poke them out with a sharp point. This practice is to be avoided at all costs as it invariably widens one or more mesh holes.

See also: **Quality Assurance:** Reference Materials; Production of Reference Materials. **Sensors:** Overview.

**Appendix A**

<i>ISO reference</i>	<i>Title</i>
ISO 2194:1991	Industrial screens – Woven wire cloth, perforated plate, and electroformed sheet – Designation and nominal sizes of openings
ISO 2395:1990	Test sieves and test sieving – Vocabulary
ISO 9045:1990	Industrial screens and screening – Vocabulary
ISO 2591-1:1990	Test sieving – Part 1: Methods using test sieves of woven wire cloth and perforated metal plates

ISO:16270	Particle size analysis by methods other than sieving – Terminology (in preparation)
ISO 9276-1:1998	Representation of results of particle size analysis – Part 1: Graphical representation
ISO 9276-2:2001	Representation of results of particle size analysis – Part 2: Calculation of average particle sizes/diameters and moments from particle size distributions
ISO 9276-3	Representation of results of particle size analysis – Part 3: Fitting of an experimental curve to a reference model (in preparation)
ISO 9276-4	Representation of results of particle size analysis – Part 4: Characterization of a classification process (in preparation)
ISO 9276-5	Representation of results of particle size analysis – Part 5: Validation of calculations relating to particle size analysis using logarithmic normal probability distribution (in preparation)
ISO 13320:2000	Particle size analysis – Laser diffraction methods – Part 1: General principles
ISO 13321:1996	Particle size analysis – Photon correlation spectroscopy
ISO 13322	Particle size analysis – Image Analysis methods – Part 1: Static image analysis methods (in preparation)
ISO 13322	Particle size analysis – Image Analysis methods – Part 2: Dynamic image analysis methods (in preparation)
ISO 13317-1:2001	Determination of particle size distribution by gravitational liquid sedimentation – Part 1: General principles and guidelines
ISO 13317-2:2002	Determination of particle size distribution by gravitational liquid sedimentation methods – Part 2: Fixed pipette method
ISO 13317-3:2001	Determination of particle size distribution by gravitational liquid sedimentation methods – Part 3: X-ray gravitational technique
ISO 13318-1:2001	Determination of particle size distribution by centrifugal liquid sedimentation methods – Part 1: General principles and guidelines
ISO 13318-3	Determination of particle size distribution by centrifugal liquid



	sedimentation methods – Part 3: Centrifugal X-ray method (in preparation)
ISO 13319:2000	Determination of particle size distributions – Electrical sensing zone method
ISO 14887:2000	Sample preparation – Dispersing procedures for powders in liquids
ISO 14488	Sample preparation – Sample splitting (in preparation)

### Further Reading

Allen T (1990) *Particle Size Measurement*. London: Chapman & Hall.

Bernhardt C (1994) *Particle Size Analysis, Classification and Sedimentation Methods*. London: Chapman & Hall.

Born M and Wolf E (1980) *Principles of Optics*. Oxford: Pergamon Press.

Cummins HZ and Pike ER (1977) *Photon Correlation and Light Beating Spectroscopy*. New York: Plenum Press.

Heiskanen K (1993) *Particle Classification*. London: Chapman & Hall.

Murphy CH (1984) *Handbook of Particle Sampling and Analysis Methods*. Berlin: Verlag Chemie.

Nelson RD (1995) *Dispersing Powders in Liquids*. Amsterdam: Elsevier.

Van De Hulst HC (1957) *Light Scattering by Small Particles*. London: Chapman & Hall.

Xu R (2000) *Particle Characterisation: Light Scattering Methods*. Amsterdam: Kluwer Academic Publishers.

## PAS

See PHOTOACOUSTIC SPECTROSCOPY

## PCBs

See POLYCHLORINATED BIPHENYLS

## PCR

See POLYMERASE CHAIN REACTION

# PEPTIDES

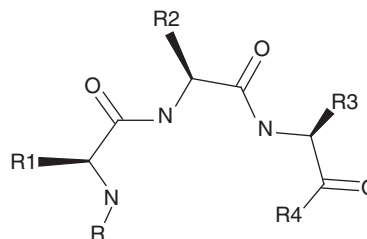
**M-I Aguilar**, Monash University, Clayton, VIC, Australia

**A W Purcell**, University of Melbourne, Parkville, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

A peptide is a molecule formed by the condensation of a small number of amino acids, of general formula



They are often obtained from the breakdown of proteins.

	sedimentation methods – Part 3: Centrifugal X-ray method (in preparation)
ISO 13319:2000	Determination of particle size distributions – Electrical sensing zone method
ISO 14887:2000	Sample preparation – Dispersing procedures for powders in liquids
ISO 14488	Sample preparation – Sample splitting (in preparation)

### Further Reading

Allen T (1990) *Particle Size Measurement*. London: Chapman & Hall.

Bernhardt C (1994) *Particle Size Analysis, Classification and Sedimentation Methods*. London: Chapman & Hall.

Born M and Wolf E (1980) *Principles of Optics*. Oxford: Pergamon Press.

Cummins HZ and Pike ER (1977) *Photon Correlation and Light Beating Spectroscopy*. New York: Plenum Press.

Heiskanen K (1993) *Particle Classification*. London: Chapman & Hall.

Murphy CH (1984) *Handbook of Particle Sampling and Analysis Methods*. Berlin: Verlag Chemie.

Nelson RD (1995) *Dispersing Powders in Liquids*. Amsterdam: Elsevier.

Van De Hulst HC (1957) *Light Scattering by Small Particles*. London: Chapman & Hall.

Xu R (2000) *Particle Characterisation: Light Scattering Methods*. Amsterdam: Kluwer Academic Publishers.

## PAS

See PHOTOACOUSTIC SPECTROSCOPY

## PCBs

See POLYCHLORINATED BIPHENYLS

## PCR

See POLYMERASE CHAIN REACTION

# PEPTIDES

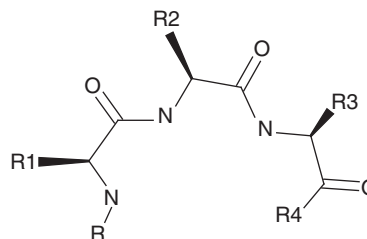
**M-I Aguilar**, Monash University, Clayton, VIC, Australia

**A W Purcell**, University of Melbourne, Parkville, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

A peptide is a molecule formed by the condensation of a small number of amino acids, of general formula



They are often obtained from the breakdown of proteins.

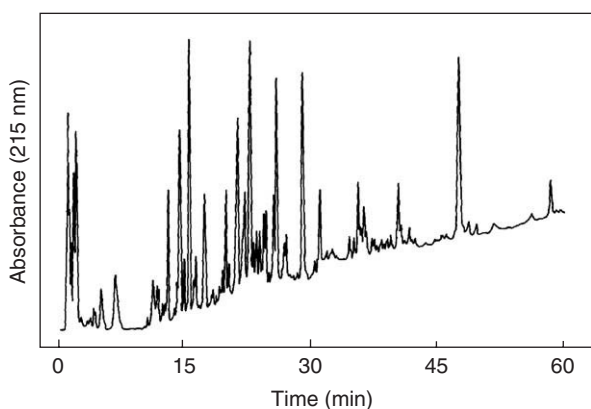
The analysis of peptide structure is central to the enormous advances made in the last 20 years in all aspects of biomedical sciences. The advances in peptide and protein chemistry have been accompanied by the development of a range of analytical procedures that in turn have accelerated the progress made in the understanding of, for example, hormone–receptor interactions and antigen–antibody interactions.

The vast majority of peptides currently produced are generated by solid-phase peptide synthesis (synthetic peptides) or from the enzymatic or chemical digestion of proteins. The procedures that are commonly used in the analysis of peptides can be divided into four stages: (1) purification, (2) composition and sequence analysis, (3) conformational analysis, and (4) biological analysis. For synthetic peptides, purification procedures are generally of routine nature and amino acid composition and sequence analysis are used to confirm the identity of the synthetic product. The purified peptide samples are then subjected to conformational analysis and biological evaluation to establish the structure–function relationships in the particular biological system. Peptides derived from the enzymatic or chemical digestion of protein (i.e., peptide mapping) are generally prepared in order to derive partial sequence information of a newly isolated protein or, in the case of a recombinant protein, the combined sequence analysis of all peptide fragments can be used to verify the structure of that protein. A wide variety of experimental techniques are now routinely used in the determination of peptide structure. This article deals only with the first three stages of peptide analysis, as a detailed overview of the range of biological assays commonly used to determine the activity of a peptide is outside its scope.

## Purification

### Liquid Chromatography

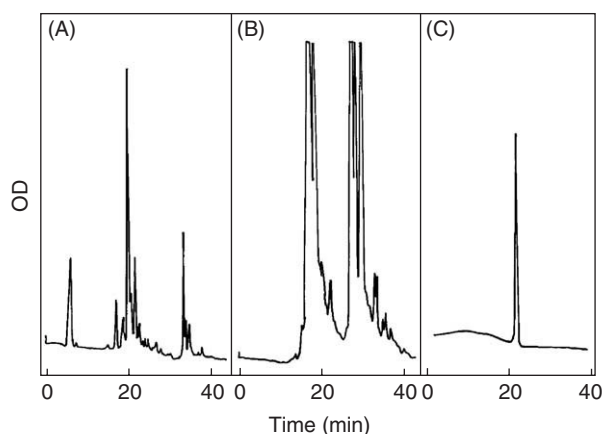
The most commonly used method for the analysis and purification of peptides mixtures is reversed-phase liquid chromatography (RPLC). The experimental system usually comprises an *n*-alkylsilica-based stationary phase material from which peptides are eluted with gradients of increasing concentration of acetonitrile in the presence of ionic modifier, e.g., trifluoroacetic acid (TFA). With modern instrumentation and columns, complex mixtures of peptides can be separated and low picomole amounts of resolved components can be collected. Separations can be easily manipulated by changing the gradient slope



**Figure 1** Reversed-phase liquid chromatographic profile of a tryptic digest of bovine growth hormone on a C18 silica column, particle diameter 5  $\mu$ m, average pore size 30 nm packed into 25 cm  $\times$  4.6 mm ID columns. Gradient elution was carried out from 0% to 50% acetonitrile in 0.1% TFA over 60 min at a flow rate of 1 ml min<sup>-1</sup>. Detection was by absorbance at 215 nm.

or organic solvent composition. The technique is equally applicable to the analysis of enzymatically derived mixtures of peptides and also for the analysis of synthetically derived peptides. An example of the analysis of a tryptic digest of bovine growth hormone is shown in **Figure 1**. This figure demonstrates the rapid and highly selective separation of tryptic peptides of proteins that can be carried out by RPLC. The chromatographic separation shown in **Figure 1** was obtained with an octadecylsilica (C18) stationary phase. Separations can also be carried out with other reversed-phased sorbents, e.g., octylsilica (C8) or butylsilica (C4). Fractions collected during the separation can be subjected to further analysis as described below. For the purification of synthetically derived peptides, the crude synthetic product is separated initially on an analytical scale to assess the complexity of the mixture. This is usually followed by large-scale purification and collection of the product, and an aliquot of the purified sample is then subjected to further chromatography or mass spectrometry to check for homogeneity. An example of this procedure is illustrated in **Figure 2** for the purification of a synthetic analog of a human growth hormone-related peptide.

A major requirement in the purification of peptides generated from previously uncharacterized proteins is to maximize the detection sensitivity due to limited supplies of the proteins. Separations by RPLC are now carried out with narrow-bore columns of 1–2 mm internal diameter which increases the solute concentration thereby increasing the sensitivity. Further miniaturization of RPLC systems has also been achieved with the use of capillary columns of internal diameters of 0.32–0.075 mm.

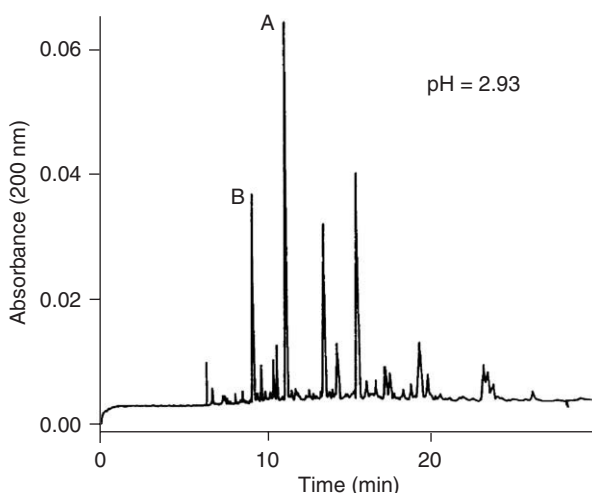


**Figure 2** Reversed-phase elution profiles illustrating the purification of a synthetic peptide. (A) Analytical profile (100 µg) of a crude peptide mixture from solid-phase peptide synthesis. Column: C18 sorbent 25 cm × 4.6 mm ID, gradient 0–50% acetonitrile in 0.1%TFA over 30 min at 1 ml min<sup>-1</sup> monitored at 214 nm. (B) Semipreparative profile (60 mg). Column: C18 sorbent 15 cm × 10 mm ID, gradient 0–50% acetonitrile in 0.1%TFA over 120 min at 1 ml min<sup>-1</sup> monitored at 254 nm (C) Analytical profile of purified peptide. Chromatographic conditions as in (A).

## Capillary Electrophoresis and Electrochromatography

Capillary electrophoresis (CE) is a technique that is particularly suitable for the analysis of peptides that differ in net charge. As such CE is therefore a technique that is complementary to RPLC. It involves the separation of peptides by application to a fused silica capillary (typically 100 cm × 100 µm). Electrophoretic mobility is controlled by an external electric field and selectivity can be manipulated by a number of factors including solvent pH, ionic strength, and other additives. CE eliminates the need for organic solvents and extends the experimental working pH range above that of most silica-bonded stationary phases used in RPLC. Before the development of CE, very high sensitivity electrophoresis of peptides in the nanogram range was not generally performed. However, CE can be used to rapidly assess (typically in less than 15 min) the purity of synthetic peptides. CE also provides alternate selectivity to RPLC for the separation of peptides. For example, **Figure 3** shows a typical CE analysis of tryptic digest fragments of salmon calcitonin. The use of various ionic additives, e.g., hexylcetyldimethylammonium salts, provides a useful extension of CE procedures with polar peptides.

High-performance capillary electrochromatography (HP-CEC) is a tool that is complementary to capillary LC and CE and has been recently



**Figure 3** Analysis by CE of the tryptic fragments of salmon calcitonin. (Reprinted with permission from Camillieri P, Okafo GN, Southan C, and Brown R (1991) Analytical and micropreparative capillary electrophoresis of the peptides from calcitonin. *Analytical Biochemistry* 198: 36–42; © Elsevier.)

applied to the analytical separation and analysis of peptides. The technique involves the electrokinetic separation of components through a capillary packed with microparticulate stationary phase.

## Composition, Sequence, and Molecular Weight

### Amino Acid Analysis

Amino acid analysis has played a crucial role in the elucidation of peptide and protein structure. However, with the advent of mass spectrometric sequencing, it is now used less frequently. Following purification, peptide samples are hydrolyzed to the constituent amino acids in the presence of strong acid and heat. After hydrolysis, the amino acids are separated by RPLC and detected by precolumn derivatization with a fluorogenic reagent or by postcolumn derivatization with ninhydrin. The composition of the resulting amino acids should agree with the predicted integer number for the constituent amino acid residues. **Table 1** lists the amino acid composition of two of the chromatographic peaks in **Figure 1** derived from the tryptic digestion of bovine growth hormone. The data presented in **Table 1** generally provide unambiguous identification and assignment of tryptic fragments. However, deviations from the true integer value can be observed for some amino acids such as threonine, serine, tryptophan, and methionine.

**Table 1** Amino acid composition data for two tryptic peptides derived from bovine growth hormone

Amino acid	T5		T9	
	Theor	Exp	Theor	Exp
Aspartic acid	1	1.11	2	2.07
Glutamic acid	3	3.12	—	—
Serine	2	1.91	2	1.93
Glycine	1	1.36	1	1.19
Histidine	—	—	—	—
Arginine	—	—	1	1.04
Threonine	3	3.02	2	1.95
Alanine	2	1.99	—	—
Proline	2	2.20	—	—
Tyrosine	1	1.10	—	—
Valine	—	—	2	1.89
Methionine	—	—	—	—
Cysteine	—	—	—	—
Isoleucine	2	2.24	—	—
Leucine	—	—	1	1.16
Phenylalanine	2	2.04	2	1.99
Lysine	1	1.05	—	—

### N-Terminal Sequencing

Peptide sequence information is crucial for a number of applications including (1) fingerprinting recombinantly derived proteins, (2) complete protein structure determinations, (3) disulfide bond assignment, (4) identification of post-translational modification sites, (5) epitope mapping, and (6) construction of oligonucleotide probes for molecular cloning. The N-terminal sequencing procedure developed by Edman and Begg has been routinely used to determine the sequence of a peptide. While mass spectrometry (MS) has also significantly impacted on the use of Edman degradation for peptide sequencing, it is often used to sequence peptides that cannot be sequenced by MS, including hydrophobic peptides and samples containing high levels of detergent. Briefly, the procedure entails coupling of the Edman reagent phenylisothiocyanate (PITC) with the amino group of the terminal amino acid residue. The PITC-derivatized amino acids are then cleaved from the protein (exposing the next amino acid derivative and converted to the more stable phenylthiohydantoin amino acid derivative). The identity of the amino acid derivative is then determined by RPLC. The sequencing cycle is then continued with the PITC coupling step with the second amino acid, and so on. The major constraint in determining the amino acid sequence of subfemtomole quantities of a peptide is the ability to purify and manipulate the samples at these very low levels. The generation and subsequent purification of peptides following protein fragmentation often requires

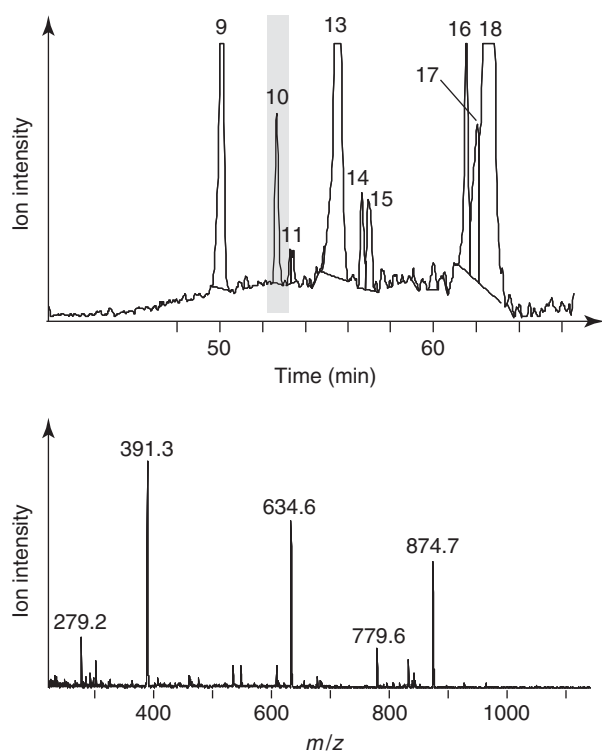
a number of micromanipulation steps such as desalting after reduction and alkylation, concentration, and buffer exchange into solvents used for further protein fragmentation. Narrow-bore or capillary chromatography is therefore now routinely used for the isolation of peptide fragments for sequence analysis.

### Mass Spectrometry

MS has emerged as a central tool for peptide structure analysis because of a number of important features. Identification in MS is based on the precise measurement of the mass-to-charge ratio from which the molecular mass of the peptide can be derived. The application of MS to the determination of large biomolecules is possible due to the development of new techniques for the ionization and detection of large polar nonvolatile molecules. The two ionization methods typically used for peptide analysis include electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI). Accurate molecular weight information can now routinely be determined on femtomolar and in some cases attomolar amounts of peptide. MS has largely superseded Edman sequencing procedures for the determination of amino acid sequence. In addition, the mass resolution and fragmentation allow post-translational modifications and other chemical modifications to be detected such as phosphorylation, sulfation, deamidation, and aspartic acid isomerization.

### Liquid Chromatography–Mass Spectrometry

With the widespread availability of sensitive mass spectrometric detectors suitable for online analysis of liquid chromatographic separations, LC–MS-based approaches are now the preferred option for chromatographic detection. The informational content of the ion chromatogram (ion intensity versus retention time) far supersedes the content of spectrophotometric detection. Moreover, the ability to resolve the peaks of the ion chromatogram into individual mass components gives instant compositional analysis and with instruments capable of performing MS/MS analysis sequence information of peptides during the separation can provide absolute identification. LC–MS is also used to quantitate species from relatively complex mixtures following appropriate standardization. An example of LC–MS analysis of a tryptic digest of cytochrome *c* is shown in **Figure 4**. Separation of tryptic fragments is easily achieved and individual species identified based on their characteristic masses, which can be predicted from the primary sequence of the protein. In addition to

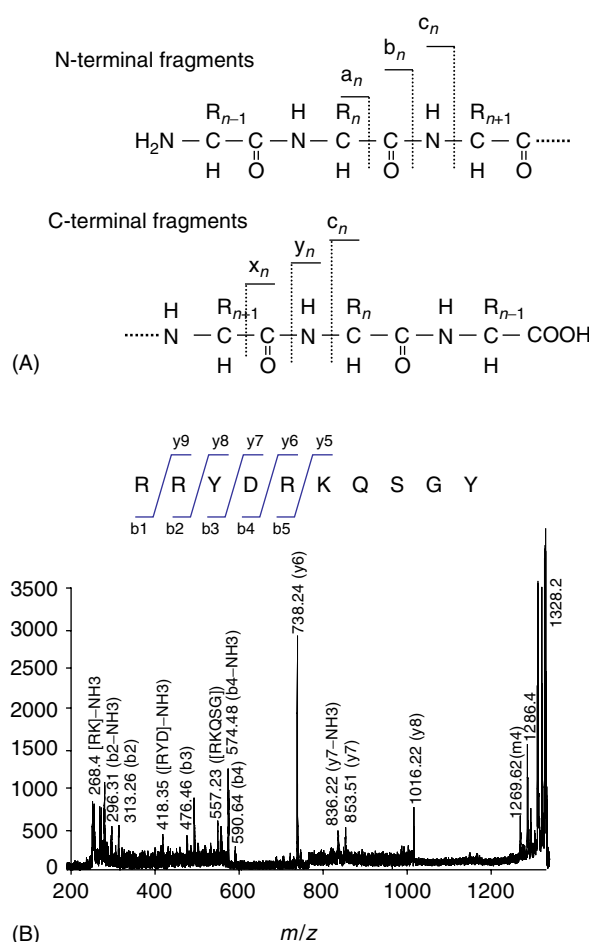


**Figure 4** LC-MS of a cytochrome *c* tryptic digest. About 400 fmol of cytochrome *c* was digested with trypsin and the resulting peptides separated on a LC Packings PepMap C18 column (300 ×  $\mu$ m ID × 5 mm length). The upper panel shows the total ion chromatogram recorded on an Agilent 1100 MSD ion trap fitted with a nanospray source. The ion chromatogram approximates the UV chromatogram, but is significantly more sensitive, no signal was observed in this instance by absorbance at 214 nm. The highlighted region of the chromatogram was displayed as a mass spectrum (lower panel), revealing the presence of three major tryptic fragments in the single peak and revealing the enhanced informational content of this form of analysis.

quantitation and identification, such mass spectral detection can be used in split flow systems to switch from fraction collection to automate peptide purification postsynthesis or from natural sources.

### Peptide Sequence Determination Using Mass Spectrometry

Modulating the conditions of ion formation and subsequent manipulation of these ions can induce fragmentation during both MALDI and ESI. In the case of peptides, nested sets of ion fragments are formed from which amino acid sequence information can be derived (see **Figure 5**). For ESI-based techniques, tandem mass spectrometry or MS/MS instrumentation can be used to generate and analyze fragments. These instruments consist of two mass analyzers arranged in tandem and separated by a collision cell. Thus, in an MS/MS instrument the



**Figure 5** Fragmentation of peptides during postsource decay and collision induced dissociation. The peptide bond is sensitive to cleavage producing N-terminal a-, b-, and c-series ion fragments and C-terminal z-, y-, and x-series ion fragments. These cleavage patterns provide sequence information for oligopeptides, which yield a nested set of N- and C-terminal fragments from which sequence information can be readily derived. PSD fragments observed for a HLA B27-bound peptide isolated by immunoaffinity chromatography and RP-LC (5), highlight sequence information gleaned by fragmentation in MS. (Reprinted with permission from Purcell AW and Gorman JJ (2001) The use of post-source decay in matrix-assisted laser desorption/ionisation mass spectrometry to delineate T cell determinants. *Journal of Immunological Methods* 249(1): 17–31; © Elsevier.)

first mass analyzer can be used to select a particular ionic species of defined mass. The selected ion is subsequently channeled into the collision cell that usually contains an inert gas. Collision with the gas molecules induces fragmentation and ion fragments are analyzed in the second mass analyzer. The most common instruments of this type of collision-induced dissociation (CID) MS/MS experiment are the triple quadrupole instruments. Other types of mass analyzers may be used in tandem



such as those used in hybrid instruments such as electrospray ionization–quadrupole–quadrupole–time-of-flight MS and four-sector MS/MS instruments. Another type of mass analyzer that is useful for MS/MS experiments using ESI sources are the ion-trap instruments.

Fragmentation of ions in MALDI need not involve CID, since the conditions of MALDI and ion acceleration may be changed to induce fragmentation in the ion source (in-source decay or ISD) or after leaving the ion source (postsources decay, PSD). Laser intensity is varied to induce both ISD and PSD, which are both mediated by collisions with the matrix cloud formed by MALDI. An inert gas may also be used in the source to induce CID fragmentation during MALDI. ISD of peptides and proteins is induced by performing MALDI at very high laser irradiance and retaining the ion in the source to decay by using a longer than usual period of delayed extraction, it is therefore essential to have pure samples for such analysis. In contrast, short delay times are used to be able to observe PSD ions. Both CID and PSD fragment ions form after leaving the ion source and can be only analyzed in a TOF instrument fitted with a reflectron. The reflectron potential is changed systematically to disperse the PSD fragment and parent ions. Spectra are collected at different reflector potentials if the instrument is fitted with a multistage or linear reflector and the spectra stitched onto a single continuous mass range, or as a single spectrum using a curved field reflectron.

## Conformational Analysis

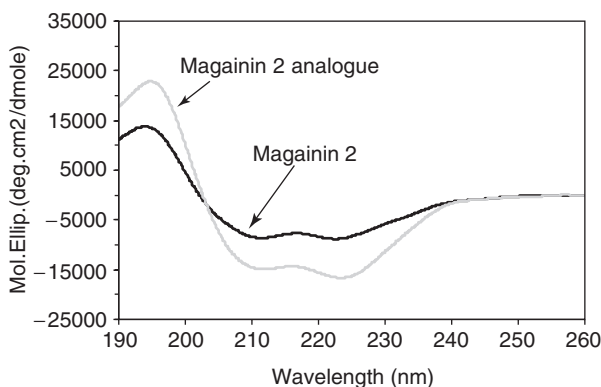
Central to the understanding of peptide structure–function relationships is the characterization of peptide conformation. The amino acid sequence predisposes the peptide chain to adopt several conformations in aqueous and lipid environments. The ultimate goal of the investigator is therefore to determine the relationship of these solution structures to the biologically active conformation. This knowledge provides information on the molecular factors involved in the action of the peptide. Ultimately, conformational analysis is an integral part of the rational design and synthesis of peptide analogs and has led to the development of improved therapeutic substances.

### Circular Dichroism Spectroscopy

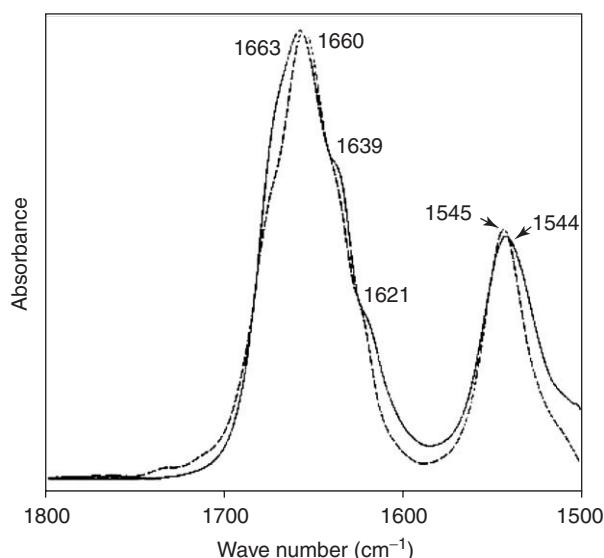
Circular dichroism (CD) spectroscopy is the most extensively used technique for the rapid evaluation of peptide secondary structure. The far-UV region

of the CD spectrum, which corresponds to the amide bond region, is used to interpret peptide secondary structure. Deconvolution of the CD spectra between 190 and 240 nm provides an indication of the relative contributions of  $\alpha$ -helix,  $\beta$ -sheet,  $\beta$ -turn and random coil structures to the overall spectrum. The diagnostic signal of an  $\alpha$ -helix is a strong negative ellipticity (Cotton effect) at 208 and 222 nm. The CD spectrum for the antimicrobial peptide magainin II and a potent analog is shown in **Figure 6**; a negative ellipticity between 208 and 222 nm, which is characteristic for an  $\alpha$ -helix, is clearly evident. The interpretation of these spectra depends on the presence of a significant population of helical conformations. For short peptides with less stabilized structure and with highly flexible N- and C-termini, the presence of  $\alpha$ -helix may be difficult to detect. The presence of significant populations of  $\beta$ -sheet structure is indicated by a negative ellipticity at 215 nm. The presence of  $\beta$ -turns in a peptide can be determined by the positive ellipticity at 200–210 nm for type II turns while type I turns exhibit similar spectra to  $\alpha$ -helices.  $\beta$ -Turns are most easily detected in the CD spectra of small peptides as the measured  $\beta$ -turn ellipticity in larger peptides is masked by the signals from the rest of the peptide. The recent development of synchrotron radiation CD has resulted in significant enhancement of spectral quality and will undoubtedly lead to improved structural information.

While CD provides a procedure for the rapid determination of secondary structure, it does not provide information in the precise location of the secondary structure within the peptide. This information is more readily obtained from nuclear magnetic resonance (NMR) spectroscopy.



**Figure 6** Circular dichroism spectra of magainin 2 and ala<sup>8,13,18</sup>-magainin analog measured in 0.02 mol l<sup>-1</sup> phosphate buffer (pH 7.4) containing dimyristylphosphatidylcholine/dimyristylphosphatidylglycerol/cholesterol (70:15:15) liposomes.



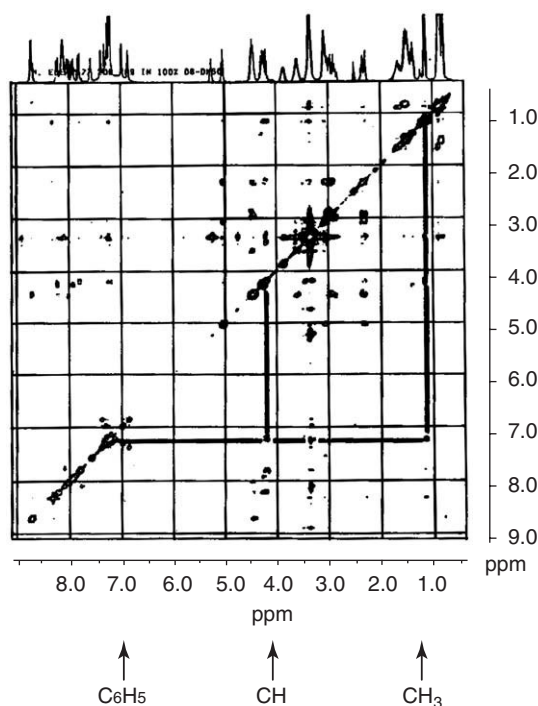
**Figure 7** Infrared spectra in the amide I and amide II regions of alamethicin (solid line) and an alamethicin analog (dotted line) in methanol. (From Haris PU, Molle G, and Duclohier H (2004) *Biophysical Journal* 86: 248–253; © Biophysical Society.)

### Fourier-Transform Infrared Spectroscopy

Fourier-transform infrared spectroscopy (FTIR) provides significant value in the analysis of peptide conformation. The most commonly used spectral band is the amide I stretching vibrations of the backbone carbonyl groups. The vibration frequency is related to the degree of hydrogen bonding that is determined by the local secondary structure. **Figure 7** shows the infrared spectra (in methanol) in the amide I region of the potent antimicrobial agent alamethicin and an analog in which all eight amino-isobutyric acids have been replaced. While the spectra in the amide I and amide II regions are similar for both peptides, the amide I bands at 1660 and 1663  $\text{cm}^{-1}$  indicate differences in peptide conformation. Spectra can also be obtained in a range of solvent conditions including the presence of phospholipids. Vibrational spectroscopy has the potential to provide information on the relative conformations of a particular peptide. These analytical techniques thus complement NMR spectroscopy which gives site-specific information.

### NMR Spectroscopy

NMR spectroscopy now represents the major tool for the precise determination of the conformation of peptides in different solution environments. Detailed information on the molecular geometry of peptides in terms of bond angles can be determined to a high degree of accuracy. However, owing to the long time-scale of the measurements, the



**Figure 8** Two-dimensional nuclear Overhauser enhancement spectroscopy  $^1\text{H}$  NMR spectrum of hGH[6–13] related peptide showing the nuclear Overhauser effect connectivities between a phenylalanine ring ( $\text{C}_6\text{H}_5$ ) and an alanine methyl ( $\text{CH}_3$ ) and methine (CH) protons).

structural information is a population-weighted average over all structures in the conformational repertoire of the peptide.

The most important structural probe is the nuclear Overhauser effect (NOE), which provides valuable information on the structure of linear peptides. Briefly, the observation of a direct NOE between a pair of protons indicates the presence of a significant population of conformers in which the distance between these two protons is relatively short. The overall pattern of connectivity therefore corresponds to a particular conformation. Constraints on both backbone and side chain dihedral angles are obtained from coupling contacts. An example of a two-dimensional NMR spectrum for a synthetic peptide related to residues 6–13 of human growth hormone is shown in **Figure 8**.

**See also:** **Capillary Electrophoresis:** Overview. **Chiroptical Analysis.** **Liquid Chromatography:** Column Technology; Mobile Phase Selection; Reversed Phase; Instrumentation; Amino Acids. **Mass Spectrometry:** Peptides and Proteins. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Nuclear Overhauser Effect. **Proteins:** Traditional Methods of Sequence Determination; Foods.

## Further Reading

- Aguilar MI (ed.) (2004) *Methods in Molecular Biology. Vol 251: HPLC of Peptides and Proteins: Methods and Protocols*. Totowa NJ: Humana Press.
- Benedetti E and Pedone C (eds.) (2002) *Peptides 2002. Proceedings of the 27th European Peptide Symposium*, Edizioni Ziino, Napoli.
- Cohen SA (2003) Amino acid analysis using pre-column derivatization with 6-aminoquinoly1-N-hydroxysuccinimidy1 carbamate. Analysis of hydrolyzed proteins and electroblotted samples. *Methods in Molecular Biology* 211: 143–154.
- Cordero MM, Cornish TJ, Cotter RJ, and Lys IA (1995) Sequencing peptides without scanning the reflectron: post-source decay with a curved-field reflectron time-of-flight mass spectrometer. *Rapid Communications in Mass Spectrometry* 9: 1356–1361.
- de Jong A (1998) Contribution of mass spectrometry to contemporary immunology. *Mass Spectrometry Reviews* 17: 311–335.
- Hicks RP (2001) Recent advances in NMR: Expanding its role in rational drug design. *Current Medicinal Chemistry* 8: 627–650.
- Keiderling TA and Xu Q (2002) Unfolded peptides and proteins studied with infrared absorption and vibrational circular dichroism spectra. *Advances in Protein Chemistry* 62: 111–161.
- Lennon JJ and Walsh KA (1999) Locating and identifying posttranslational modifications by in-source decay during MALDI-TOF mass spectrometry. *Protein Science* 8: 2487–2493.
- Popa TV, Mant CT, and Hodges RS (2003) Capillary electrophoresis of synthetic peptide standards varying in charge and hydrophobicity. *Electrophoresis* 24: 4197–4208.
- Shively JE (2000) The chemistry of protein sequence analysis. *EXS* 88: 99–117.
- Spengler B (1997) Post-source decay analysis in matrix-assisted laser desorption/ionisation mass spectrometry of biomolecules. *Journal of Mass Spectrometry* 32: 1019–1036.
- Wallace BA and Janes RW (2001) Synchrotron radiation circular dichroism spectroscopy of proteins: Secondary structure, fold recognition and structural genomics. *Current Opinion in Chemical Biology* 5: 567–571.
- Woody RW (1995) Circular dichroism. *Methods in Enzymology* 246: 34–71.

# PERFUMES

**A Salvador-Carreno and A Chisvert**, Universidad de Valencia, Valencia, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The use of perfumes goes back thousands of years. The Egyptians used plants, gums, and resins in religious rites. As the years went by, scented substances were used to enhance body attractiveness and to make homes and public places more pleasant.

Fragrances are considered normal components of our everyday lives. Many people feel the need to wear a fragrance in order to feel good: this is probably because there is a connection between scent and emotion as well as between scent and memory; moreover, studies have shown that some fragrances can alter moods and even alleviate anxiety and stress.

Perfumes can be defined as substances that emit and diffuse a pleasant and fragrant odor. They consist of manmade mixtures of aromatic chemicals and essential oils. Until the nineteenth century perfumes were usually composed of natural aromatic oils.

Nowadays, most perfumes are synthetic and may contain many components.

This article deals with the different types of products containing perfumes, raw materials, legislation and safety aspects, analytes of interest, and analytical techniques.

## Types of Products Containing Perfumes

Perfumes can be used in different types of products (Figure 1) such as cosmetics and toiletries (fine fragrances, shampoos, conditioners, hair sprays, shaving creams, makeup, baby care products, deodorants, soaps, feminine products, etc.) and household products (cleaners, air fresheners, bleaches, laundry detergents, fabric softeners, etc.).

### Cosmetics and Toiletries

Fragrance compositions used in cosmetics are complex mixtures of hundreds of ingredients, which together give a compound its unique scent.

Fine fragrances are hydroethanolic solutions prepared from aromatic chemicals and essential oils. They are named according to the fragrance

## Further Reading

- Aguilar MI (ed.) (2004) *Methods in Molecular Biology. Vol 251: HPLC of Peptides and Proteins: Methods and Protocols*. Totowa NJ: Humana Press.
- Benedetti E and Pedone C (eds.) (2002) *Peptides 2002. Proceedings of the 27th European Peptide Symposium*, Edizioni Ziino, Napoli.
- Cohen SA (2003) Amino acid analysis using pre-column derivatization with 6-aminoquinoly1-N-hydroxysuccinimidy1 carbamate. Analysis of hydrolyzed proteins and electroblotted samples. *Methods in Molecular Biology* 211: 143–154.
- Cordero MM, Cornish TJ, Cotter RJ, and Lys IA (1995) Sequencing peptides without scanning the reflectron: post-source decay with a curved-field reflectron time-of-flight mass spectrometer. *Rapid Communications in Mass Spectrometry* 9: 1356–1361.
- de Jong A (1998) Contribution of mass spectrometry to contemporary immunology. *Mass Spectrometry Reviews* 17: 311–335.
- Hicks RP (2001) Recent advances in NMR: Expanding its role in rational drug design. *Current Medicinal Chemistry* 8: 627–650.
- Keiderling TA and Xu Q (2002) Unfolded peptides and proteins studied with infrared absorption and vibrational circular dichroism spectra. *Advances in Protein Chemistry* 62: 111–161.
- Lennon JJ and Walsh KA (1999) Locating and identifying posttranslational modifications by in-source decay during MALDI-TOF mass spectrometry. *Protein Science* 8: 2487–2493.
- Popa TV, Mant CT, and Hodges RS (2003) Capillary electrophoresis of synthetic peptide standards varying in charge and hydrophobicity. *Electrophoresis* 24: 4197–4208.
- Shively JE (2000) The chemistry of protein sequence analysis. *EXS* 88: 99–117.
- Spengler B (1997) Post-source decay analysis in matrix-assisted laser desorption/ionisation mass spectrometry of biomolecules. *Journal of Mass Spectrometry* 32: 1019–1036.
- Wallace BA and Janes RW (2001) Synchrotron radiation circular dichroism spectroscopy of proteins: Secondary structure, fold recognition and structural genomics. *Current Opinion in Chemical Biology* 5: 567–571.
- Woody RW (1995) Circular dichroism. *Methods in Enzymology* 246: 34–71.

# PERFUMES

**A Salvador-Carreno and A Chisvert**, Universidad de Valencia, Valencia, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The use of perfumes goes back thousands of years. The Egyptians used plants, gums, and resins in religious rites. As the years went by, scented substances were used to enhance body attractiveness and to make homes and public places more pleasant.

Fragrances are considered normal components of our everyday lives. Many people feel the need to wear a fragrance in order to feel good: this is probably because there is a connection between scent and emotion as well as between scent and memory; moreover, studies have shown that some fragrances can alter moods and even alleviate anxiety and stress.

Perfumes can be defined as substances that emit and diffuse a pleasant and fragrant odor. They consist of manmade mixtures of aromatic chemicals and essential oils. Until the nineteenth century perfumes were usually composed of natural aromatic oils.

Nowadays, most perfumes are synthetic and may contain many components.

This article deals with the different types of products containing perfumes, raw materials, legislation and safety aspects, analytes of interest, and analytical techniques.

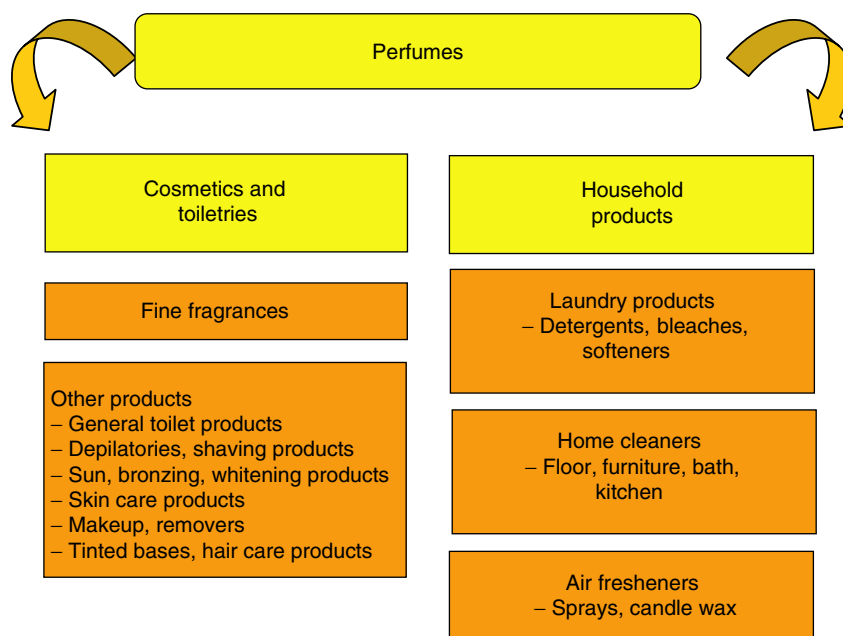
## Types of Products Containing Perfumes

Perfumes can be used in different types of products (Figure 1) such as cosmetics and toiletries (fine fragrances, shampoos, conditioners, hair sprays, shaving creams, makeup, baby care products, deodorants, soaps, feminine products, etc.) and household products (cleaners, air fresheners, bleaches, laundry detergents, fabric softeners, etc.).

### Cosmetics and Toiletries

Fragrance compositions used in cosmetics are complex mixtures of hundreds of ingredients, which together give a compound its unique scent.

Fine fragrances are hydroethanolic solutions prepared from aromatic chemicals and essential oils. They are named according to the fragrance



**Figure 1** Types of products usually containing perfumes.

Parfum	12–18%
Eau de Parfum	8–12%
Eau de Toilet	5–8%
Eau Fraiche	4–5%
Eau de Cologne	3–4%
Cologne	2–3%
Baby Cologne	1–2%

**Figure 2** Approximate concentration of perfumes in different types of fine fragrances.

content (Figure 2): Parfum, Eau de Parfum, Eau de Toilette, Eau Fraiche, Eau de Cologne, Cologne, Baby Cologne. The last has a special formulation, with a very low content of fragrance and some hydrating ingredients added to give softness to the baby skin.

Perfumes are used in practically all cosmetics and toiletries, sometimes with a function secondary to the main one of the product but often solely to provide fragrance. For example, aftershave lotions and splash colognes usually contain 0.5–2% perfume oil.

### Household Products

A wide range of perfumes, predominantly based on synthetic ingredients, are used in household products. They are used to make cleaners smell more pleasant than their chemical ingredients. Some of the cleaners that usually contain fragrances in their compositions are softeners and laundry detergents, dish-washing products, floor cleaning products, furniture polish, and bathroom cleaners. Air fresheners also contain perfumes in their formulations.

Special problems can occur in the formulation of cosmetics and toiletries or household products containing perfumes (odors may be changed or may become unstable in the new medium, the perfume may change the physical properties of the cosmetic, etc.). Therefore, the fragrance must be formulated individually for each product.

### Types of Raw Materials

Raw materials used as perfumes include natural products (of plant or animal origin) and synthetic materials.

#### Natural Ingredients: Methods of Obtaining

Natural ingredients originate from materials found in nature. These are obtained through physical processes.

Vegetable products are obtained from different parts of plants – flowers (jasmine, rose, gardenia),



fruits (lemon, orange), roots (vetiver, cistus), leaves (violela, patchouli), wood (vetiver, sandalwood, cedarwood), bark (cinnamon, nutmeg), resin (benjui, tolu, galbanum), seeds (angelica, apio) – and whole plants (lavender, geranio).

Some animal secretions contain odoriferous substances that increase the diffusion and qualities of perfumes. Some of them act as fixatives, preventing the rapid evaporation of the most volatile ingredients of the perfume.

The most widely used animal products are ambergris (from the sperm whale), castor or castoreo (from the beaver), civet (from the civet cat), and musk (from the musk deer).

Different procedures can be used to obtain raw materials for fragrances from plants and animals.

Hydrodistillation is used to obtain essential oils (lavender, rose) from plants. Plants are boiled in water, and the essential oils are evaporated and dragged by the water vapor.

Solvent extraction is used for delicate flowers (e.g. jasmine). Thus, the damage caused by the high temperatures needed to boil water when using the hydrodistillation method is avoided. The concentrated flower oil is called absolute.

Soxhlet extraction can facilitate the extraction of essential oils in some cases. Supercritical fluid extraction (SFE) allows faster extraction, and moreover, use of supercritical fluids as extraction solvents allows, in some cases, a green extraction procedure.

In the extraction method called enfleurage, petals are placed between layers of purified animal fat, which become saturated with flower oil; alcohol is then used to obtain the absolute. This technique is carried out at low temperature; it is very costly and is rarely used today.

Maceration is similar to enfleurage. The flowers are steeped in vats of oil until the scented parts dissolve. The oil may be heated to speed up the process. Maceration is, however, very time consuming.

The expression method is used to obtain essential oils from fresh fruit peels (lemon, orange), it can be done either by using the traditional procedure of pressing with sponges or by mechanical maceration: as there is no heat involved, the smell of the oil is very close to that of the original plant.

Apart from these techniques there are various other operations such as rectification, fractional distillation, terpene removal, decolorization, etc., that improve and refine the numerous raw materials used for the blending and making of perfumes.

Natural ingredients coming from animal products are usually extracted using the alcohol maceration procedure.

## Synthetic Ingredients

Synthetic ingredients are manufactured through chemical processes. Diverse types of synthetic products are used as perfumes. The components of a great number of natural fragrances have been determined and then chemically synthesized to imitate natural perfumes. Artificial fragrances that do not imitate nature have also been used to create new perfumes.

Different organic groups have been used in synthetic perfumes: hydrocarbons, terpenes, sesquiterpenes, alcohols (citronelol, geraniol, cinnamic alcohol), fenols (*p*-cresol, thymol), aldehydes (cuminaldehyde, vanillin), ketones (carvone, muscone, ionones, metilionones), lactones (nonalactone, undecalactone), esters (acetates, formates, butyrates), nitrogen compounds (indol, methyl anthranilate, musk xylol, quinolines).

The main advantage of using synthetic products is that their costs are lower than those of natural products and they can always be obtained consistently without any problems related to poor crop quality or lack of supply or to difficulties using animal extracts, etc. In other words, artificial products cost less and represent fewer market variations.

The selection of raw materials depends on different questions such as the type of the desired odor, the users focused on, the particular application, the stability of the finished product, the approximate price desired for the final product, etc.

## Legislation and Safety

Both EU and US legislations stipulate that the ingredients of a cosmetic product must be listed on the label in order of predominance. There are a few exceptions to the labeling requirements, such as fragrance formulas, which are considered trade secrets: therefore the ingredients in fragrances are not required to be revealed. The word 'fragrance' must be included on any label of products in which ingredients have been added to give off an odor. This word on the label may represent many ingredients.

Many of the ingredients used in fragrances have had little or no safety testing done on them. Most of the safety testing that has been done revolved around the dermatological effects of fragrance chemicals. The effects on the respiratory system, the brain, and other organs of the body have not been determined for individual chemicals, much less in the combinations in which they are used.

Fragrance chemicals can be inhaled through the nose, ingested through the mouth, and absorbed through the skin. Once in the body they are absorbed into the bloodstream and transported throughout the body. Individual sensitivities to the effects of



fragrance chemicals vary widely from no effect at all to severe reactions. Different symptoms have been found in different people: skin sensitivity, rashes, dermatitis, coughing, asthma attacks, migraine, etc. The incidence of these effects can be studied in relation to the method of application: leave on skin, rinse off skin, non-skin contact, handling household products.

Different organizations such as the Research Institute for Fragrance Materials and the International Fragrance Association (IFRA) have been formed by the industry and have an important role in consumer protection. They carry out research on ingredients used in fragrances in order to ensure the safety of perfumery materials. Recent studies on criteria for safety evaluation are particularly interesting.

At the moment, EU legislation forbids the use of 36 fragrances as cosmetic ingredients; therefore, any other fragrance could be used. Since fragrance materials do not have to be declared on the label, it is impossible to know exactly which materials are in the product. This makes it difficult to ascertain which potentially problematic material is in any given product. Fortunately, a recent EU normative obliges manufacturers to declare on the labels the name of 26 ingredients that have been proven to be allergenic. This and similar initiatives in other countries should contribute to the consumer being better informed.

On the other hand, terms such as 'unscented' or 'fragrance free' have no legal definitions. A product labeled with these terms still contains fragrance chemicals, even if the scent cannot be easily detected, because if a fragrance is added to a product to mask or cover up the unpleasant odor of some ingredients, it need not be mentioned on the label.

In the same way, the term 'hypoallergenic' is sometimes used for some cosmetics. Perfumes are the substances mainly responsible for producing allergies, although some preservatives, dyes, ultraviolet filters, detergents, solvents, etc. can also be allergenic. Hypoallergenic cosmetics are products that manufacturers claim produce fewer allergic reactions than other cosmetic products. However the term has no legal definition, and consequently the use of a hypoallergenic product does not guarantee immunity to any allergic reaction.

## Analytes of Interest

As mentioned before, a perfume is a mixture of a lot of compounds that are responsible for an aroma; perfume manufacturers use ~1500 compounds. These compounds are chemically very different and can be grouped into different families according their chemical structure. **Table 1** shows some examples of these compounds.

**Table 1** Examples of compounds that could be present in a perfume composition, ordered according to chemical structure

Group	Compound
Alcohol	$\beta$ -Citronellal, eugenol, geraniol, linalool, menthol
Aldehyde	Benzaldehyde, citral, hexyl cinnamal, anisaldehyde, vanillin
Hydrocarbon	Nonane, tetradecane, camphene
Terpene	Limonene, myrcene, $\alpha$ -pinene
Sesquiterpene	Caryophyllene
Ester	Methyl salicylate, menthyl lactate, terpineol acetate
Ether	Anethole, eucalyptol
Ketone	Cyclohexanone, camphor
Lactone	Coumarin, 6-methyl-coumarin, $\gamma$ -undecalactone

Analytical control of the components in perfume formulations ensures that the final product has the desired fragrance.

Other compounds that are permitted but can still cause health problems and must be declared on the label have a special analytical interest. For instance, **Table 2** shows 26 potentially allergenic substances whose presence in the formulation has to be advertised according to legislation. Of these 26 substances, two are natural extracts (oak moss and tree moss), and the rest are volatile chemicals. Recently, the IFRA has developed an analytical procedure that allows the determination of these 24 volatile chemicals using gas chromatography (GC) coupled with mass spectrometry (MS).

## Preparation of Samples for Analysis

Sample preparation depends not only on the type of sample but also on the analytical technique used for analysis.

Sometimes sample pretreatment is not necessary due to the simplicity of some of these samples, and they can be injected directly in a GC system. Due to the volatile nature of the compounds present in a perfume, the headspace (HS) sampling mode is also often employed to introduce the compounds to be determined in the GC system.

However, a previous distillation is sometimes needed to separate the target compounds from a complex matrix sample. After they are condensed, they can then be either injected directly or through the HS mode. In other cases, the use of extraction techniques like solvent-solvent extraction, solid-phase extraction, and solid-phase microextraction not only achieves the isolation of the target compounds from the matrix but also produces a preconcentration of

**Table 2** Potentially allergenic fragrances that have to be declared on the labels of cosmetic formulations according to EU legislation*Compound (Chemical Abstracts Service Registry Number)*

Amyl cinnamal (122–40–7)	Anisyl alcohol (105–13–5)
Benzyl alcohol (100–51–6)	Benzyl cinnamate (103–41–3)
Cinnamyl alcohol (104–54–1)	Farnesol (4602–84–0)
Citral (5392–40–5)	2-(4- <i>t</i> -Butylbenzyl) propionaldehyde (80–54–6)
Eugenol (97–53–0)	Linalool (78–70–6)
Hydroxycitronellal (107–75–5)	Benzyl benzoate (120–51–4)
Isoeugenol (97–54–1)	Citronellal (106–22–9)
Amyl cinnamyl alcohol (101–85–9)	Hexyl cinnamaldehyde (101–86–0)
Benzyl salicylate (118–58–1)	Limonene (5989–27–5)
Cinnamal (104–55–2)	Methyl heptin carbonate (111–12–6)
Coumarin (91–64–5)	3-Methyl-4-(2,6,6-trimethyl-2-cyclohexen-1-yl)-3-buten-2-one (127–51–5)
Geraniol (106–24–1)	Oak moss extract (90028–68–5)
Hydroxymethylpentyl- cyclohexenecarboxaldehyde (31906–04–4)	Tree moss extract (90028–67–4)

the analytes. (HS-solid-phase microextraction–GC has been applied successfully to the determination of perfume components in different household products such as scented candle wax, liquid detergents, and fabric softeners.)

## Analytical Techniques

In the past, quality control of perfumes was done by measuring physical properties like the refractive index, density, and optical rotation, in addition to very simple chemical tests like acidimetry and measurement of the saponification and carbonyl indexes. However, these tests were only useful for checking the raw materials or the final product, and they did not offer an actual determination of the compounds that were present in an unknown perfume.

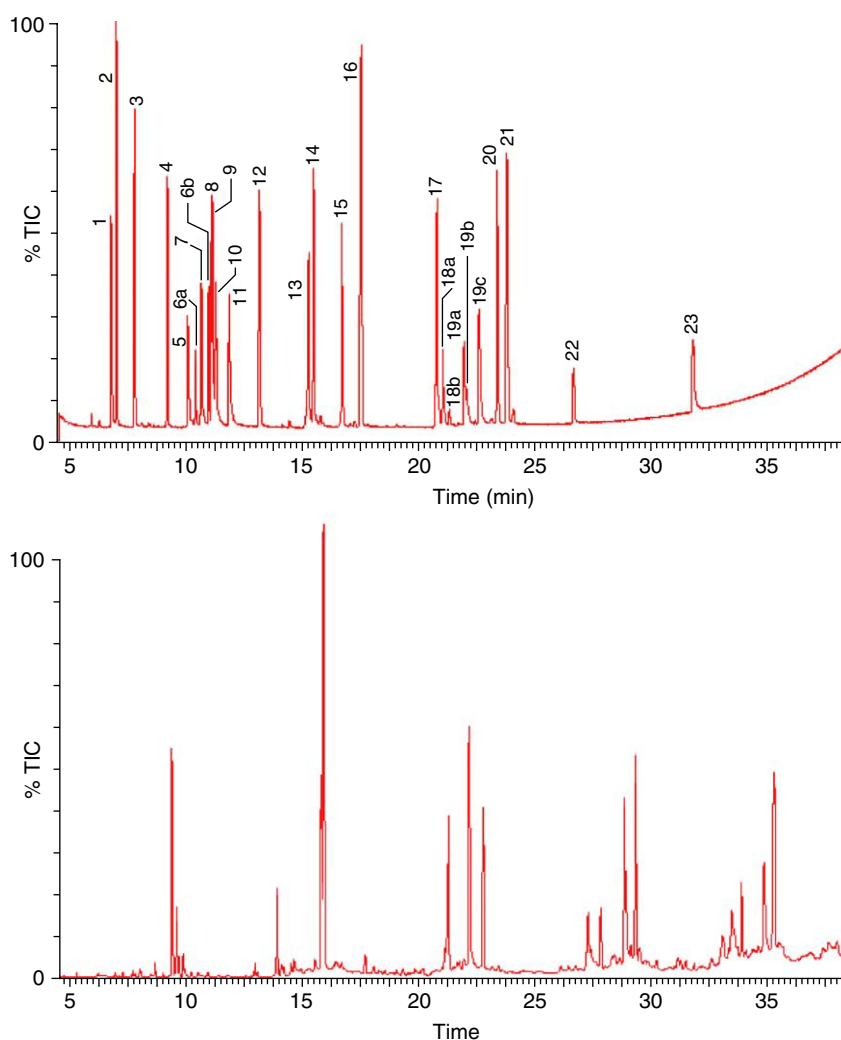
Taking into account the fact that many compounds are present in a perfume, the determination of the composition of an unknown perfume is not an easy task. As an example, fine fragrances can contain over 50 ingredients, and although the expertise of perfumers is sometimes useful, it is not enough to identify (and much less to quantify) all the compounds involved; powerful analytical techniques are also required. Separation techniques like chromatographic and related techniques are the most suitable for individual separation and determination of these compounds.

Bearing in mind that the compounds forming a perfume have a low boiling point, GC has been the most widely used technique in the perfume industry. In this sense, a perfume, after appropriate sample pretreatment, is analyzed using GC with a flame ionization detector (FID), and the Kovats index (KI) for each peak is experimentally determined. The identification is carried out by comparing the

experimentally determined KI with values kept in databases, although sometimes a composition is not completely resolved because there can be peaks with very similar KI values. Moreover, although the databases try to update frequently, new ingredients can be present, and the software library will not recognize these novel compounds. In these cases, an MS detector coupled with GC, instead of an FID, plays a crucial role since the chemical structure can be elucidated by studying the mass spectra of the compound. The use of GC-MS has been applied satisfactorily to analysis of different cosmetic products such as shampoos, creams, lotions, lipsticks, and face powder and to fine fragrances. In this sense, current databases besides the KI database also contain the mass spectra of compounds, making GC-MS a powerful analytical technique.

As has been mentioned above, recent EU norms require manufacturers to declare on the labels of cosmetic products the name of several ingredients that have been proven to be allergenic. However, no initiative has been taken yet to regulate the use of fragrance chemicals in cleaning products and other consumer products. The reason for this may be that knowledge about exposure estimates is not sufficient. **Figure 3** shows the feasibility of using GC-MS to determine 24 allergenic volatile chemicals in household and other consumer products.

On the other hand, perfumes contains compounds in very low concentrations that can be detected by the human nose, which can be extremely sensitive to these compounds. In these cases an FID is not appropriate because it is highly sensitive; however, an MS detector can settle the problem. Moreover, the MS detector can work in the selected ion-monitoring mode, making it more sensitive and selective.



**Figure 3** GC-MS chromatograms obtained for a standard mixture of potentially allergenic target compounds (upper) and for a commercial dishwasher product containing some of these compounds (bottom). 1, benzyl alcohol; 2, limonene; 3, linalool; 4, methyl heptin carbonate; 5, citronellal; 6, citral; 7, geraniol; 8, cinnamal; 9, hydroxycitronellal; 10, anisyl alcohol; 11, cinnamyl alcohol; 12, eugenol; 13, coumarin; 14, isoeugenol; 15, 3-methyl-4-(2,6,6-trimethyl-2-cyclohexen-1-yl)-3-buten-2-one; 16, 2-(4-*t*-butylbenzyl)propionaldehyde; 17, amyl cinnamal; 18, hydroxy-methylpentyl cyclohexenecarboxaldehyde; 19, farnesol; 20, benzyl benzoate; 21, hexyl cinnamaldehyde; 22, benzyl salicylate; 23, benzyl cinnamate. (Reproduced with permission from Rastogi SC (2002) *Survey of Chemical Compounds in Consumer Products, Survey No. 8: Contents of Selected Fragrance Materials in Cleaning Products and Other Consumer Products*. Copenhagen: Danish Environmental Protection Agency.)

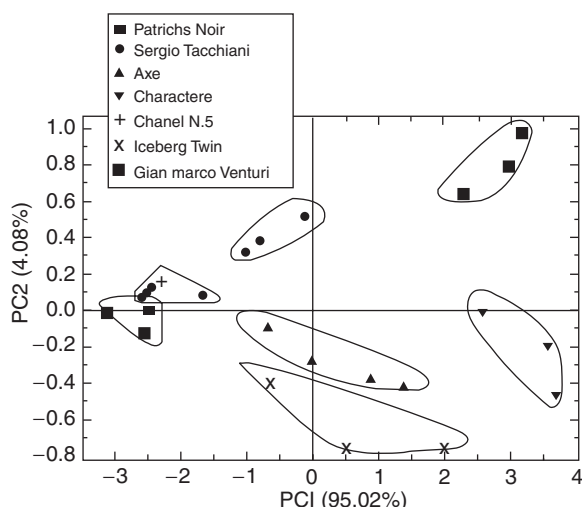
The internal standard method is often used for quantitative requirements.

Other chromatographic techniques, like liquid chromatography (LC) and thin layer chromatography (TLC), have been less widely used in perfume analysis. Different compounds such as benzylideneacetone, cinnamyl alcohol, isoeugenol, cinnamyl anthranilate, cinnamaldehyde, bergapten, citropten, coumarin, 6-methyl coumarin, and saffrole have been determined using either LC or TLC.

Many of the compounds present in perfumes are chiral, and sometimes the chiral ratio of a mixture can affect the odor of the perfume since the olfactory

receptor in the human nose is protein based and thus enantiospecific. Chiral columns can achieve separation and quantification of the different enantiomers of a target compound.

Nowadays, electronic noses have an increasingly prominent role in the field of perfume analysis. These instruments are devices composed of an array of nonselective gas sensors that can act in a manner similar to that of real biological noses. In this way, after exposure to analyte vapors, the analyte molecules diffuse and pass over the detectors, producing characteristic signal patterns, which are conveniently processed using multivariate data analysis (principal



**Figure 4** Classification of seven cosmetic products according to odor pattern recognition obtained using an electronic nose and analyzing the data using PCA. (Reprinted with permission from Van Asten (2002) The importance of GC and GC-MS in perfume analysis. *Trends in Analytical Chemistry* 21: 698–708; © Elsevier.)

component analysis (PCA), partial least squares, artificial neuronal networks). Electronic noses should be able to provide a fingerprint for each odor without any information about chemical composition. In this context, they have been applied to the identification of different compounds. Another application of the electronic nose is differentiation between various commercial fine fragrances, as can be seen in Figure 4.

See also: **Chemometrics and Statistics:** Multivariate Classification Techniques. **Cosmetics and Toiletries.**

**Distillation. Essential Oils. Extraction:** Solvent Extraction Principles; Solid-Phase Extraction; Solid-Phase Microextraction. **Gas Chromatography:** Detectors; Mass Spectrometry; Chiral Separations. **Headspace Analysis:** Static; Purge and Trap. **Mass Spectrometry:** Principles; Selected Ion Monitoring. **Quality Assurance:** Quality Control. **Sensors:** Overview.

## Further Reading

- Cadby PA, Troy WR, and Vey MGH (2002) Consumer exposure to fragrance ingredients: Providing estimates for safety evaluation. *Regulatory Toxicology and Pharmacology* 36: 246–252.
- Directive 2002/34/EC of 15 April 2002 adapting to technical progress Annexes II, III, and VII to Council Directive 76/768/EEC on the approximation of the laws of the Member States relating to cosmetic products; <http://europa.eu.int>.
- Directive 2003/15/EC of 27 February 2003 amending Council Directive 76/768/EEC on the approximation of the laws of the Member States relating to cosmetic products; <http://europa.eu.int>.
- Ford RA, Domeyer B, Easterday O, Maier K, and Middleton J (2000) Criteria for development of a database for safety evaluation of fragrance ingredients. *Regulatory Toxicology and Pharmacology* 31: 166–181.
- Pybus DH and Sell C (1999) *The Chemistry of Fragrances*. Cambridge: The Royal Society of Chemistry.
- Rastogi SC (2000) Allergens in perfumes: Gas chromatography mass spectrometry. In: Wilson ID, Adlard TR, Cooke M, and Poole CF (eds.) *Encyclopedia of Separation Science*, vol. III, pp. 1974–1981. New York: Academic Press.

# PERSONAL MONITORING

Contents

**Active**

**Passive**

## Active

**K Kuwata**, Air Pollution Control Division, Environmental Bureau, Osaka, Japan

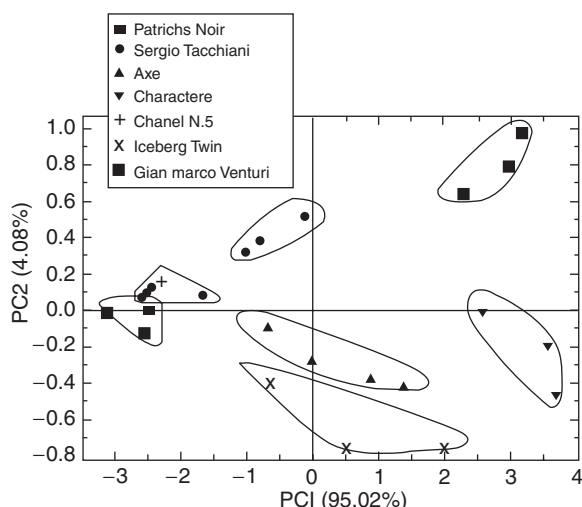
**Y Nishikawa**, Environmental Pollution Control Centre, Osaka, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 3770–3776, © 1995, Elsevier Ltd.

## Introduction

Atmospheric monitoring was very effective in the days of air pollution episodes (such as the London smog incident of 1952) for investigating acute effects of highly concentrated atmospheric pollutants on human health because the outdoor concentrations of air pollutants were related to personal exposure to the pollutants. In recent years, such air conditions have been improved in many cities, air pollution criteria have been established for the protection of



**Figure 4** Classification of seven cosmetic products according to odor pattern recognition obtained using an electronic nose and analyzing the data using PCA. (Reprinted with permission from Van Asten (2002) The importance of GC and GC-MS in perfume analysis. *Trends in Analytical Chemistry* 21: 698–708; © Elsevier.)

component analysis (PCA), partial least squares, artificial neuronal networks). Electronic noses should be able to provide a fingerprint for each odor without any information about chemical composition. In this context, they have been applied to the identification of different compounds. Another application of the electronic nose is differentiation between various commercial fine fragrances, as can be seen in Figure 4.

See also: **Chemometrics and Statistics:** Multivariate Classification Techniques. **Cosmetics and Toiletries.**

**Distillation. Essential Oils. Extraction:** Solvent Extraction Principles; Solid-Phase Extraction; Solid-Phase Microextraction. **Gas Chromatography:** Detectors; Mass Spectrometry; Chiral Separations. **Headspace Analysis:** Static; Purge and Trap. **Mass Spectrometry:** Principles; Selected Ion Monitoring. **Quality Assurance:** Quality Control. **Sensors:** Overview.

## Further Reading

- Cadby PA, Troy WR, and Vey MGH (2002) Consumer exposure to fragrance ingredients: Providing estimates for safety evaluation. *Regulatory Toxicology and Pharmacology* 36: 246–252.
- Directive 2002/34/EC of 15 April 2002 adapting to technical progress Annexes II, III, and VII to Council Directive 76/768/EEC on the approximation of the laws of the Member States relating to cosmetic products; <http://europa.eu.int>.
- Directive 2003/15/EC of 27 February 2003 amending Council Directive 76/768/EEC on the approximation of the laws of the Member States relating to cosmetic products; <http://europa.eu.int>.
- Ford RA, Domeyer B, Easterday O, Maier K, and Middleton J (2000) Criteria for development of a database for safety evaluation of fragrance ingredients. *Regulatory Toxicology and Pharmacology* 31: 166–181.
- Pybus DH and Sell C (1999) *The Chemistry of Fragrances*. Cambridge: The Royal Society of Chemistry.
- Rastogi SC (2000) Allergens in perfumes: Gas chromatography mass spectrometry. In: Wilson ID, Adlard TR, Cooke M, and Poole CF (eds.) *Encyclopedia of Separation Science*, vol. III, pp. 1974–1981. New York: Academic Press.

# PERSONAL MONITORING

Contents

**Active**

**Passive**

## Active

**K Kuwata**, Air Pollution Control Division, Environmental Bureau, Osaka, Japan

**Y Nishikawa**, Environmental Pollution Control Centre, Osaka, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 3770–3776, © 1995, Elsevier Ltd.

## Introduction

Atmospheric monitoring was very effective in the days of air pollution episodes (such as the London smog incident of 1952) for investigating acute effects of highly concentrated atmospheric pollutants on human health because the outdoor concentrations of air pollutants were related to personal exposure to the pollutants. In recent years, such air conditions have been improved in many cities, air pollution criteria have been established for the protection of



human health, and various programs have been implemented effectively for cutting down emission of airborne pollutants to maintain or attain the criteria set. On the other hand, indoor pollutant levels may originate from indoor sources such as home cooking, some types of home heating, tobacco smoking, and others. The concentrations of such compounds as sulfur dioxide (SO<sub>2</sub>), nitrogen oxides (NO<sub>x</sub>), formaldehyde, and carbon monoxide (CO) might frequently be higher indoors than outdoors. Outdoor concentration data are no longer primary indicators of personal exposure to pollutants in the living environment as these days urban dwellers might spend as much as 90% of their time indoors. Instruments and/or devices are thus essential for monitoring either residential or personal exposure to hazardous air pollutants, so that the effects of the pollutants on human health may be estimated in epidemiological surveys.

Personal samplers are those for continuously monitoring exposure of persons to hazardous air-borne substances in the living and/or occupational environment. These samplers are often used in surveys to measure personal exposure to air pollutants and thus to calculate subjects' doses from their respiration and living patterns. Personal samplers may be categorized as either active or passive.

Passive samplers consist of a film or cartridge containing a collection medium and a small holder with a clasp and utilize the principles of permeation or diffusion. They do not use a pump to draw air through the collection medium but are directly exposed to air by being worn on a person's shoulder or chest for short or longer periods. Average exposures can be calculated for the relevant period from the amount of gas absorbed on the sampler. This type of device is inexpensive, lightweight, and easy to handle and wear. It is useful for monitoring personal exposure to gaseous substances for periods from several days up to a month. However, because the gas collection rates are low, indication of hourly changes of exposure or peaks of concentration may not be possible using them. In addition, the samplers are not applicable to suspended particulate matter (SPM) monitoring.

In contrast, active samplers are designed to draw a large volume of air through a collection medium using an air pump. They can detect air pollutants sensitively so that hourly variations of the components or high-concentration peaks can be monitored, and they are also applicable to SPM monitoring. As active devices are much larger than passive samplers, it may be somewhat difficult for a person to always wear the samplers in the living and/or occupational environment. Accordingly, most active samplers

should be located where people live and work, so that exposures can be determined through knowledge of how long a person remains in the monitoring place.

The following sections describe active personal samplers in use and some that are anticipated in the near future.

## Requirements for Active Personal Samplers

### Collection Media and Samplers for Air Pollutants

Table 1 shows active samplers that have actually been used in the fields of industrial hygiene and air pollution analysis. Midget impingers and bubblers using absorptive solutions have been mainly used to monitor hazardous substances at occupational settings. Recently, filters impregnated with collecting agents and minicartridges packed with effective adsorbents have been used to sample air pollutants in ambient air or industrial emissions. These can be convenient to handle and useful as active personal samplers. In particular, Sep-Pak<sup>®</sup> cartridges packed with silica particles, silica particles chemically bonded with octadecylsilane (ODS), or a mixture of silica particles and magnesium oxide particles, are conveniently used to collect various air pollutants. A Sep-Pak<sup>®</sup> cartridge coated with a derivatization reagent is useful for monitoring reactive or unstable hazardous substances such as aldehydes, thiols, and amines. Annular denuders can be useful for separately sampling gas and particulate substances either indoors or outdoors.

It is said that an ideal personal sampler is one that can continuously record any sampling place, time, and exposure to substances in the living place of a person. It is important for active samplers to monitor hourly exposures to the substances and to detect the numbers of hours or days when the exposure exceeds given concentrations or exhibits high-concentration peaks. For this purpose, the samplers must be able to monitor hazardous substances sensitively over a long period, with subsequent rapid and simple analysis. Furthermore, they must be lightweight, small, safe, and rugged enough to be carried or worn by a person or placed in any living or occupational setting. They must be inexpensive and require little maintenance since a number of devices are required simultaneously in an epidemiological survey. Unfortunately, there are no such ideal samplers: the most suitable samplers should be selected and used according to the purpose of the survey and the environment monitored.



**Table 1** Portable samplers for monitoring air pollutants in the active sampling mode<sup>a</sup>

<i>Sampler</i>	<i>Typical substances</i>	<i>Description</i>
Impingers and bubblers	NO <sub>x</sub> , SO <sub>x</sub> , NH <sub>3</sub> , inorganic anions, organic vapours	Collect substances by bubbling a sample through an absorbent solution. Unsuitable as personal samplers
Filters	Particulate matter, inorganic anions, PAHs, heavy metals, asbestos, inorganic gases, organic vapors	Mainly collect particulate matter in gas or air sample. Ionic component could be determined by IC, PAHs by LC after purification of the sample, and heavy metals by PIXE or INAA
Impactors	Particulate matter, inorganic anions, heavy metals	Classify and collect particulate matter in different sizes
Solid adsorbent columns (charcoal, silica gel, Tenax GC, etc.)	Solvent vapors, hydrocarbons, halogenated carbons, pesticides	Collect substances by passing an air sample through a sampling column packed with an adsorbent and desorbed by heating or eluted with CS <sub>2</sub> followed by GC analysis
Minicartridges impregnated with collecting reagents	Low-molecular-mass aldehydes, amines, thiols, carboxylic acids, phenols, NO <sub>2</sub> , HNO <sub>3</sub> , NH <sub>3</sub> , SO <sub>2</sub> , pesticides	Collect substances by passing a sample through a cartridge. A reactive or unstable substance may be converted to a stable compound using a derivatization reagent followed by LC analysis. Many compounds may be determined by LC or IC. Pesticides may be determined by GC-MS in secondary ion monitoring mode
Denuders	NH <sub>3</sub> , NO <sub>2</sub> , HNO <sub>3</sub> , SO <sub>x</sub> , HCl, carboxylic acid vapors	Separately collect gaseous and particulate matter using a diffusion tube followed by a particulate collector. Ionic components may be determined by IC after extraction

<sup>a</sup> See text for abbreviations.

Substances sampled using a sampler are usually determined directly or after sample preparation using an instrumental analysis technique such as gas chromatography (GC), liquid chromatography (LC), ion chromatography (IC), gas chromatography-mass spectrometry (GC-MS), particle-induced X-ray emission analysis (PIXE), or instrumental neutron activation analysis (INAA), among others.

### Sampling Pumps

A portable small pump that draws air efficiently is essential for an active personal sampler. A suitable pump should be compatible with SPM samplers and gas samplers, though the performance required may differ somewhat for different samplers. The pump should be small and lightweight and should be able to draw air for long periods using either small batteries or a household power supply. Several commercially available diaphragm air pumps, operated with either batteries or household power supplies, can draw air at precise flow rates in the range 0.2–6.0 l min<sup>-1</sup>. A sampler, an air pump, a flow controller, and batteries may be integrated into a small sound-proof case so as to be easily carried or worn and to alleviate noise pollution. The system should be

operated for more than 8 h. A thermometer and hygrometer may be necessary if there is no other way to measure temperature and humidity at the sampling site.

## Applications

A few simple personal sampling systems are shown in **Figure 1** for sampling SPM and gaseous and/or vapor components. The use of active samplers is described in the following sections according to the air pollutants.

### SPM and Metal Elements in SPM

SPM is usually defined as comprising particles less than 10 µm in diameter suspended in the atmospheric environment. In particular, air particulates less than 2 µm in diameter, which mainly originate from such sources as incinerators, boilers, and automobiles, may reach deeply into human lungs during respiration and thus cause respiratory diseases. SPM can be also produced by photochemical reactions of gaseous substances in the atmosphere.

An SPM sampler should sample a substantial volume of air (~1000 l) to collect size-classified SPM to the sensitivity level of a chemical balance.

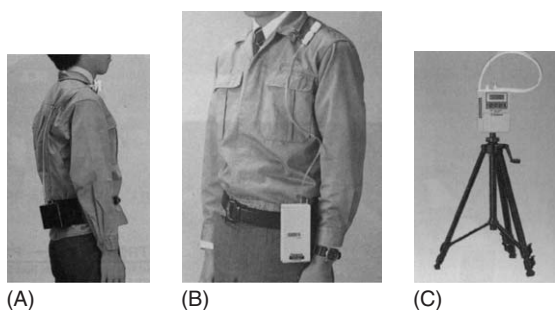
The sampling system is somewhat large and heavy compared with a gas sampling system. Sampling is usually made with two- to three-stage size classification using an impactor and filters with different pore sizes. **Figure 2A** shows an NBS model portable ambient aerosol sampler for SPM that classifies 15–7  $\mu\text{m}$  particulates using an impactor, 5–3  $\mu\text{m}$  particulates on a 6.8  $\mu\text{m}$  nuclear pore filter coated with Apiezon grease, and particles less than 3  $\mu\text{m}$  on a 3  $\mu\text{m}$  polytetrafluoroethylene (PTFE) fiber filter.

Another sampler for industrial hygiene applications removes particles larger than 10  $\mu\text{m}$  and collects

2–10  $\mu\text{m}$  particles with an impactor and particles less than 2  $\mu\text{m}$  on a glass fiber filter on a PTFE net support. Air is sampled at 2.5  $\text{l min}^{-1}$  for 8 h or more. The classified SPM samples are weighed separately using a chemical balance, and the average SPM concentration can be calculated during the sampling period.

A small SPM sampler has been developed by improving a  $\beta$ -ray attenuation mass monitor often used at monitoring stations (**Figure 2B**). A magnetic card with a 13 mm diameter filter is inserted into the sample holder. Air is sampled through a 7 mm diameter circle of the filter at 0.5–1.0  $\text{l min}^{-1}$  for 8–24 h. Particulates larger than 10  $\mu\text{m}$  can be removed using a preimpactor. The card is removed from the holder and the SPM concentration is measured using the  $\beta$ -ray absorption method at a laboratory and recorded on the magnetic card with other necessary information.

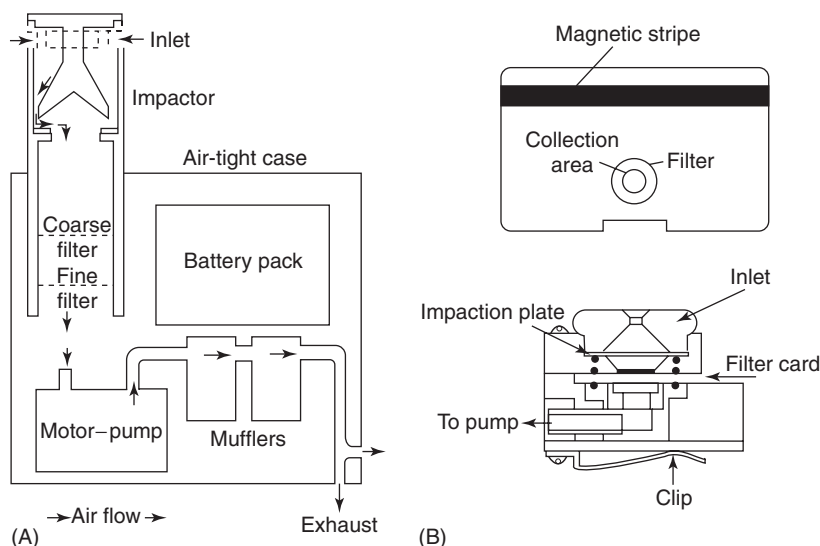
SPM samples can be subjected to direct determination of metallic elements at ultratrace levels using PIXE or INAA. If such analytical instruments are available in a laboratory, personal monitoring of metallic elements in SPM is possible for short or long periods.



**Figure 1** Personal sampling devices. (A) Personal sampler for particulate matter. (B) Personal sampler for gaseous or vapor compounds. (C) Personal sampler for placement at a living or occupational site. (Courtesy of Sibata Scientific Technology, Tokyo, Japan, used by permission.)

### Organic Vapors

There are many sources of emission of organic vapors, including petroleum refining, chemical, painting,



**Figure 2** SPM samplers. (A) NBS model sampler. (Reprinted with permission from Bright DS and Fletcher RA (1983) New portable ambient aerosol sampler. *American Industrial Hygiene Association Journal* 1983: 528–536. Copyright 1983 American Industrial Hygiene Association, reprinted with permission.) (B) SPM sampler for  $\beta$ -ray absorption analysis. (Reproduced with permission from Masuda S and Koichiro Takahashi (1990) *Aerosols: Science, Industry, Health & Environment: Proceedings of the Third International Aerosol Conference, Kyoto, Japan, 24–27 September 1990* set 2 Vols. Oxford: Pergamon Press.)

cleaning and printing processes, automobiles, gasoline stations, and others.

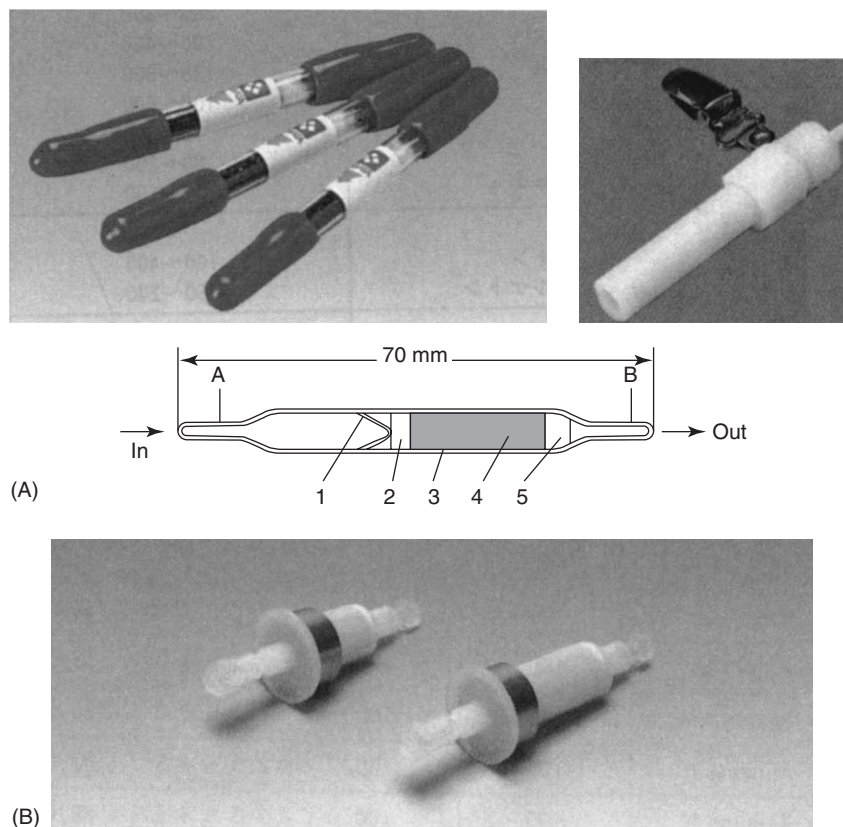
A 70 mm long  $\times$  4 mm inner diameter (ID) glass tube, packed with 100–200 mg of 20–40 mesh charcoal, silica gel, or a mixture of charcoal and silica gel and sealed by melting the tube ends, can be used as a personal sampler for organic solvent vapors and petroleum vapors in the living or occupational place (Figure 3A). Air is sampled at 0.5–1.5 l min<sup>-1</sup> through the cartridge after the tube ends are broken. For analysis, the packing is placed into a test tube and extracted with 5 ml of carbon disulfide in an ultrasonic bath. The solution is centrifuged and analyzed for organic components using GC.

### Low-Molecular Mass Aldehydes and Phenols

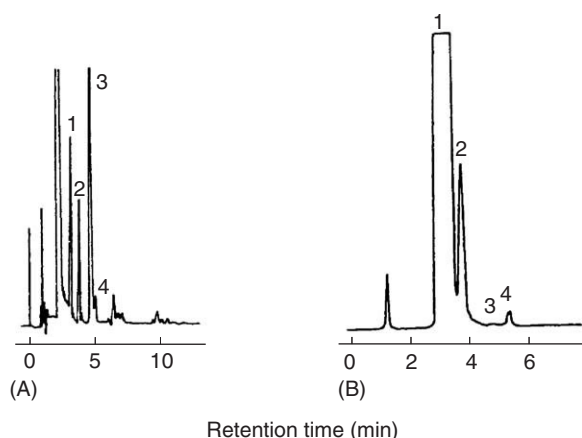
Low-molecular mass aldehydes and phenols known to be hazardous and/or odorous substances may be emitted from incinerators, chemical plants, baking finish plants, and automobiles. They may be produced as secondary products of organic air-borne pollutants

by photochemical reaction. They might even be emitted indoors from tobacco smoking and items of furniture or utensils made from plastic materials.

A sampler used for C<sub>1</sub>–C<sub>4</sub> aldehydes is a minicartidge packed with 55–105  $\mu$ m silica particles or 55–105  $\mu$ m ODS chemically bonded silica particles (Sep-Pak<sup>®</sup> Silica or Sep-Pak<sup>®</sup> C<sub>18</sub>) coated with 1.0 mg of 2,4-dinitrophenylhydrazine (DNPH) and 20 mg of phosphoric acid (Figure 3B). A 10–50 l volume of air is sampled through the cartridge at 0.2–1.5 l min<sup>-1</sup> for 0.5–12 h. The substances trapped are eluted with 2 ml of acetonitrile. The aldehydes are derivatized quantitatively with DNPH to stable hydrazones on the trap and/or in the elution media. The derivatization products are analyzed for the aldehydes using LC (Figure 4A). When 50 l of air is sampled, aldehydes can be detected down to the 0.2 ng l<sup>-1</sup> level. Sep-Pak<sup>®</sup> C<sub>18</sub> coated with 900  $\mu$ g of sodium hydroxide is used as a sampler for phenols. A 2–100 l volume of air is sampled at 0.2–1.0 l min<sup>-1</sup> through the cartridge. The adsorbed substances are eluted with 2 ml of methanol and derivatized at pH 11.5 by



**Figure 3** Minicartridges for use as personal samplers for gases or vapors. (A) Charcoal or silica gel tube for sampling organic vapors: 1, spring stopper; 2, quartz wool; 3, glass tube; 4, 20–40 mesh charcoal or silica gel (200 mg); 5, polyurethane foam plug. (Courtesy of Sibata Scientific Technology, used by permission.) (B) Sep-Pak<sup>®</sup> cartridges. (Courtesy of Waters Corporation, Milford, MA, USA, used by permission.)



**Figure 4** Liquid chromatograms in the determination of hazardous and odorous substances. (A) Aldehydes in an urban sample. LC conditions: 15 cm  $\times$  4.0 mm ID Develosil ODS-3 (3  $\mu$ m); mobile phase 65:35 (v/v) acetonitrile–water 1.0 ml min<sup>-1</sup>. Peaks: 1, formaldehyde 4.3 ng l<sup>-1</sup>; 2, acetaldehyde 4.4 ng l<sup>-1</sup>; 3, acetone; 4, propionaldehyde 2.0 ng l<sup>-1</sup>. (From Kuwata K, Uebori M, Yamasaki H, Kuge Y, and Kiso Y (1983) Determination of aliphatic aldehydes in air by liquid chromatography. *Analytical Chemistry* 55: 2013) (B) Alkyl amines emitted from a night soil container. LC conditions: 20 cm  $\times$  4.0 mm ID LiChrosorb RP-18 (5  $\mu$ m); mobile phase 50:5:45 (v/v) methanol–acetonitrile–water 1.0 ml min<sup>-1</sup>. Peaks: 1, ammonia; 2, methylamine 5.9 ng l<sup>-1</sup>; 3, ethylamine (below detection limit); 4, dimethylamine 7.5 ng l<sup>-1</sup>. (From Nishikawa Y and Kuwata K (1984) Liquid chromatographic determination of low molecular weight aliphatic amines in air via derivatization with 7-chloro-4-nitro-2,1,3-benzoxadiazole. *Analytical Chemistry* 56: 1790.)

adding 300  $\mu$ g of *p*-nitrobenzenediazonium tetrafluoroborate. The azo derivatives produced are assessed for the phenols using LC with ultraviolet (UV) detection. A large volume of the sample solution can be introduced into the LC column by using a minicolumn to detect the phenols at 0.01 ng l<sup>-1</sup> levels.

#### Low-Molecular Mass Alkylamines, Alkylthiols, and Fatty Acids

These compounds may be odorous substances produced by anaerobic decomposition of organic compounds containing nitrogen, oxygen, and/or sulfur. The vapors can be sampled in a manner similar way to that for the aldehydes using the following minicartridges (**Figure 3B**).

A Sep-Pak<sup>®</sup> C<sub>18</sub> cartridge coated with 900  $\mu$ g of phosphoric acid is used for sampling C<sub>1</sub>–C<sub>4</sub> alkyl primary and secondary amines. A 5–100 l volume of air is sampled through the cartridge at 0.8–1.2 l min<sup>-1</sup>. The trapped substances are eluted with 5 ml of methanol and derivatized by mixing with 4 mg of 7-chloro-4-nitro-1,2,3-benzoxadiazole (NBD-Cl) and 2 mg of potassium hydroxide. The

derivatization products are determined for the amines using LC (**Figure 4B**).

A Sep-Pak<sup>®</sup> Fluorisil (20 mm  $\times$  10 mm ID minicartridge packed with 0.14 g of 50–200  $\mu$ m silica particles and 0.76 g of 50–200  $\mu$ m magnesium oxide) is coated with 2 mg of NBD-Cl for sampling C<sub>1</sub>–C<sub>4</sub> alkylthiols. A 1–100 l volume of air is sampled at 0.8–1.2 l min<sup>-1</sup> through the cartridge. The trapped components are eluted with 2 ml of methanol. The derivatization may be completed by the catalytic effects of magnesium oxide in the methanol media. The derivatization products are assessed for the alkylthiols using LC with fluorescence detection.

Sep-Pak<sup>®</sup> C<sub>18</sub> coated with 300  $\mu$ g of sodium hydroxide is used for sampling C<sub>2</sub>–C<sub>4</sub> fatty acids. A 1–100 l volume of air is sampled at 0.5–1.5 l min<sup>-1</sup> through the cartridge. The adsorbed substances are eluted with 2 ml of acetonitrile and derivatized by mixing with 0.4 mg of 18 *p*-bromophenacylbromide and 0.04 mg of 18-crown-6-ether. The phenacyl derivatives produced under catalytic effect of the crown ether are assessed for the fatty acids using LC with UV absorbance detection.

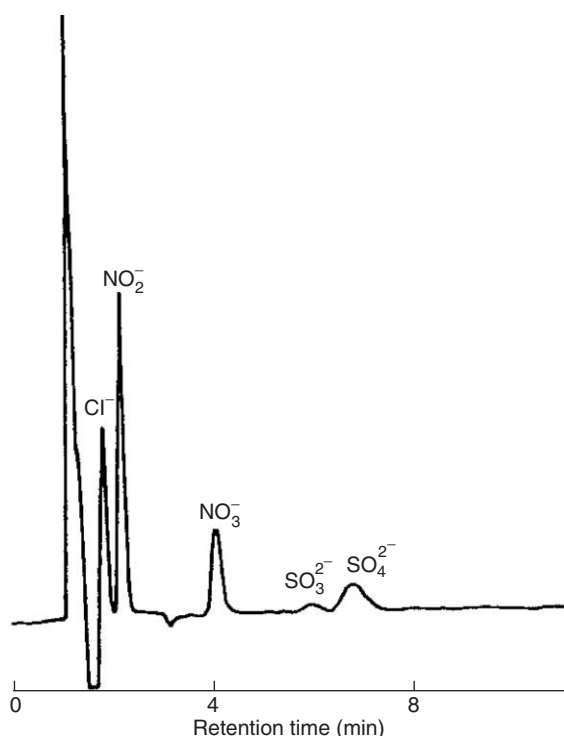
Any of the methods described above can be used to detect the substances down to 0.1 ng l<sup>-1</sup> when 100 l of air is sampled.

#### Nitrogen Oxides, Sulfur Oxides, and Hydrochloric Acid

These compounds are mainly emitted from incinerators, boilers, and chemical plants. In particular, high levels of nitrogen oxides may often be detected indoors as they may arise from some types of home-heating stove, kitchens, and even tobacco smoking.

Various types of monitoring instruments are used for these gases at monitoring stations or at emission sources, but they may be difficult to use as personal samplers owing to their large size and high cost.

Minicartridge samplers have been developed so that the substances can be sensitively and simultaneously determined using IC. A Sep-Pak<sup>®</sup> C<sub>18</sub> cartridge impregnated with 600  $\mu$ g of triethanolamine and 600  $\mu$ g of potassium hydroxide may be convenient as a personal gas sampler for these components, while an absorption filter or a combined filter and annular denuder can be used for separate collection of the gas-phase and particulate-phase substances. A 10–100 l volume of air is sampled at 0.8–1.2 l min<sup>-1</sup> through the cartridge. The trapped ionic substances are eluted with 5 ml of an IC eluent and determined simultaneously using IC. **Figure 5** shows a chromatogram of ionic compounds from the atmosphere. The minicartridge can be used as a passive sampler after some modification.



**Figure 5** Ion chromatogram for determining ionic substances in the atmosphere. Active sampling volume 201 l;  $\text{NO}_2^-$  118 ng;  $\text{NO}_3^-$  72 ng;  $\text{SO}_3^{2-}$  22 ng;  $\text{SO}_4^{2-}$  47 ng. (From Nishikawa Y and Taguchi K (1987) Ion chromatographic determination of nitrogen dioxide and sulphur dioxide in the atmosphere using triethanolamine-potassium hydroxide-coated cartridges. *Journal of Chromatography* 396: 251.)

### Ammonia

Ammonia is mainly produced by anaerobic decomposition of nitrogen-containing organic compounds and is emitted into the atmosphere. It is the major nitrogen compound released on the surface of the earth. It is odorous and stimulative.

A minicartridge packed with a weak cation exchanger (Sep-Pak<sup>®</sup> ACCELL CM) has been used for sampling ammonia. A 1–100 l volume of air is sampled at 0.2–1.0 l min<sup>-1</sup> through the cartridge. The trapped substance is eluted with 2 ml of 0.02 mol l<sup>-1</sup> hydrogen chloride solution and determined using IC. The recovery of ammonia from the cartridge is 98% and the relative standard deviation ( $n = 6$ ) of the determined values is 2%. Gaseous ammonia and ammonium compounds can be sampled separately in combination with a small annular denuder coated with citric acid and a filter.

### Polynuclear Aromatic Hydrocarbons

Polynuclear aromatic hydrocarbons (PAHs) may be produced in the incineration of organic materials and may be adsorbed on particulate matter in the

atmosphere. They may be important factors in risk assessment, because some congeners such as benzo(a)pyrene may show high carcinogenicity.

The SPM sample, collected using the methods described above ( $\sim 1 \text{ m}^3$  of air sampled), is extracted with 4 ml of benzene–ethanol (3:1 v/v) in an ultrasonic bath for 10 min. A 2 ml volume of the sample is dried at 60°C by slowly flushing nitrogen over it, dissolved in 1 ml of hexane, and introduced into a Sep-Pak<sup>®</sup> Silica cartridge.

The cartridge is rinsed with 5 ml of hexane, and the PAHs are eluted with 4 ml of dichloromethane–*n*-hexane (50:50 v/v). The eluate is again dried in a similar way and dissolved in an LC eluent (acetonitrile–methanol–water (50:35:15 v/v), followed by reversed-phase LC with fluorescence detection. The PAHs are detectable at 0.1 ng m<sup>-3</sup> levels when 1 m<sup>3</sup> of air is sampled.

### Pesticides

A person might be exposed to pesticides at a golf course or in a garden or even indoors as well as in agricultural or forestry settings.

A minicartridge, such as a Sep-Pak<sup>®</sup> C<sub>18</sub> cartridge, may be convenient for personal monitoring. A volume of air up to 1000 l is sampled at 0.7 l min<sup>-1</sup> through the cartridge. The trapped components are eluted with 5 ml of hexane. The sample is mixed with 5–50 ng of [<sup>13</sup>C<sub>6</sub>]hexachlorobenzene as an internal standard and reduced to 0.1 ml by gently flushing nitrogen over it. A 2 µl volume of the sample is analyzed in the selected-ion monitoring mode using GC–MS. Pesticide components can be determined at 0.1 ng m<sup>-3</sup>.

**See also:** Air Analysis: Sampling; Workplace Air. **Lipids:** Fatty Acids. **Personal Monitoring:** Passive. **Pesticides:** Polycyclic Aromatic Hydrocarbons: Determination.

### Further Reading

- Ando M and Tamura K (1987) Study on the suspended particulate matter around a crossroad, outdoor and indoor air environment. *Nihon Eiseigaku Zasshi (Japanese Journal of Hygiene)* 42: 939.
- Baek SO, Goldstone ME, Kirk PW, Lester JN, and Perry R (1991) Methodological aspects of measuring polycyclic aromatic hydrocarbons in the urban atmosphere. *Environmental Technology* 12: 107.
- Brauer M, Koutrakis P, and Spengler JD (1989) Personal exposures to acidic aerosols and gases. *Environmental Science and Technology* 23: 1408.
- Dams R, Robbing JA, Rahn KA, and Winchester JW (1970) Nondestructive neutron activation analysis of air pollution particulates. *Analytical Chemistry* 42: 861.



- Environmental Protection Agency (1988) *Determination of Formaldehyde in Ambient Air Using Adsorbent Cartridge Followed by High Performance Liquid Chromatography*. Method TO-11. USA: Environmental Protection Agency.
- Fujimori K, Oki N, Nakano T, Tsuji M, and Okuno T (1991) Analysis of pesticides in air by collection method using ODS cartridge. *Kankyo Kagaku (Journal of Environmental Chemistry, Japan)* 1: 575.
- Koutrakis P, Fasano AM, Slater JL, *et al.* (1989) Design of a personal annular denuder sampler to measure atmospheric aerosols and gases. *Atmospheric Environment* 23: 2767.
- Kunugi M, Matsumoto Y, Shindo J, Koyama H, and Tani T (1990) A new personal sampling system. *Science, Industry, Health and Environment* 1: 648.
- Kuwata K and Tanaka S (1988) Liquid chromatographic determination of traces of phenols in air. *Journal of Chromatography* 442: 407.
- Kuwata K and Tanaka S (1988) Liquid chromatographic determination of trace amounts of low-molecular-weight fatty acid vapours in industrial emissions. *Journal of Chromatography* 455: 425.
- Nishikawa M and Kayahara H (1993) Determination of polycyclic aromatic hydrocarbons in airborne suspended particulate by high-performance liquid chromatography. *Kankyo Gijutsu (Environmental Conservation Engineering, Japan)* 22: 53.
- Nishikawa Y and Kuwata K (1985) Liquid chromatographic determination of low molecular weight alkylthiols in air via derivatization with 7-chloro-4-nitro-2,1,3-benzoxadiazole. *Analytical Chemistry* 57: 1864.
- Risner CH and Conner JM (1991) Collection of ammonia in indoor air by means of weak cation exchange cartridge. *Environmental Toxicology and Chemistry* 10: 1417.
- Sickles JE II, Grohse PM, Hodson LL, *et al.* (1990) Development of a method for the sampling and analysis of sulfur dioxide and nitrogen dioxide from ambient air. *Analytical Chemistry* 62: 338.
- Tejada SB (1986) Evaluation of silica gel cartridges coated *in situ* with acidified 2,4-dinitrophenylhydrazine for sampling aldehydes and ketones in air. *International Journal of Environmental Analytical Chemistry* 26: 167.

## Passive

**M Harper**, National Institute for Occupational Safety & Health, Morgantown, WV, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 3776–3782, © 1995, Elsevier Ltd.

## Basic Theory

Acute or chronic exposure to the thousands of minor components that can be present in the air we breathe may cause adverse health effects. Such exposure may occur in the workplace, home, or outdoor environment. Monitoring of these components is required to ensure compliance with recommended or regulated air quality standards. The preferred monitoring method would give a continuous and accurate measurement of instantaneous changes in the concentration of each harmful species present, in a manner that would model human exposure. The compromise most often applied in practice is to extract the components from a large volume (to provide sensitivity) of breathing zone air together with subsequent laboratory analysis (to provide selectivity).

In the 'active' method of extracting the gas and vapor components of air, a known volume is passed through an apparatus containing liquids in which the

components can dissolve, porous solids on which the components can be adsorbed, or liquids or solids containing chemicals with which the components can react. The total volume of air passed is metered by the pump, and both pump and extraction apparatus are placed appropriately about the person to obtain a representative sample. The pumps have been criticized as bulky, heavy, and noisy, and therefore an encumbrance to the wearer, and also as expensive.

An alternative method known as 'passive' or diffusive sampling, which does not require a pump or air mover, has gained popularity in recent years. Diffusive samplers operate by allowing the gas and vapor molecules to make their own way to the collection medium by diffusion along a carefully controlled path. The rate of movement, which is a function of the diffusion coefficient of the gas or vapor in air and the path geometry (Figure 1), can be derived from Fick's first law of diffusion:

$$W = -DA \frac{dc}{dx}$$

where  $W$  is the rate of mass transfer ( $\text{ng s}^{-1}$ ),  $D$  is the diffusion coefficient ( $\text{cm}^2 \text{s}^{-1}$ ),  $A$  is the cross-sectional area of the diffusion path ( $\text{cm}^2$ ), and  $dc/dx$  is the rate of change in concentration over the



- Environmental Protection Agency (1988) *Determination of Formaldehyde in Ambient Air Using Adsorbent Cartridge Followed by High Performance Liquid Chromatography*. Method TO-11. USA: Environmental Protection Agency.
- Fujimori K, Oki N, Nakano T, Tsuji M, and Okuno T (1991) Analysis of pesticides in air by collection method using ODS cartridge. *Kankyo Kagaku (Journal of Environmental Chemistry, Japan)* 1: 575.
- Koutrakis P, Fasano AM, Slater JL, *et al.* (1989) Design of a personal annular denuder sampler to measure atmospheric aerosols and gases. *Atmospheric Environment* 23: 2767.
- Kunugi M, Matsumoto Y, Shindo J, Koyama H, and Tani T (1990) A new personal sampling system. *Science, Industry, Health and Environment* 1: 648.
- Kuwata K and Tanaka S (1988) Liquid chromatographic determination of traces of phenols in air. *Journal of Chromatography* 442: 407.
- Kuwata K and Tanaka S (1988) Liquid chromatographic determination of trace amounts of low-molecular-weight fatty acid vapours in industrial emissions. *Journal of Chromatography* 455: 425.
- Nishikawa M and Kayahara H (1993) Determination of polycyclic aromatic hydrocarbons in airborne suspended particulate by high-performance liquid chromatography. *Kankyo Gijutsu (Environmental Conservation Engineering, Japan)* 22: 53.
- Nishikawa Y and Kuwata K (1985) Liquid chromatographic determination of low molecular weight alkylthiols in air via derivatization with 7-chloro-4-nitro-2,1,3-benzoxadiazole. *Analytical Chemistry* 57: 1864.
- Risner CH and Conner JM (1991) Collection of ammonia in indoor air by means of weak cation exchange cartridge. *Environmental Toxicology and Chemistry* 10: 1417.
- Sickles JE II, Grohse PM, Hodson LL, *et al.* (1990) Development of a method for the sampling and analysis of sulfur dioxide and nitrogen dioxide from ambient air. *Analytical Chemistry* 62: 338.
- Tejada SB (1986) Evaluation of silica gel cartridges coated *in situ* with acidified 2,4-dinitrophenylhydrazine for sampling aldehydes and ketones in air. *International Journal of Environmental Analytical Chemistry* 26: 167.

## Passive

**M Harper**, National Institute for Occupational Safety & Health, Morgantown, WV, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 3776–3782, © 1995, Elsevier Ltd.

## Basic Theory

Acute or chronic exposure to the thousands of minor components that can be present in the air we breathe may cause adverse health effects. Such exposure may occur in the workplace, home, or outdoor environment. Monitoring of these components is required to ensure compliance with recommended or regulated air quality standards. The preferred monitoring method would give a continuous and accurate measurement of instantaneous changes in the concentration of each harmful species present, in a manner that would model human exposure. The compromise most often applied in practice is to extract the components from a large volume (to provide sensitivity) of breathing zone air together with subsequent laboratory analysis (to provide selectivity).

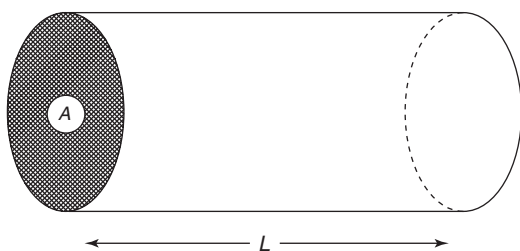
In the 'active' method of extracting the gas and vapor components of air, a known volume is passed through an apparatus containing liquids in which the

components can dissolve, porous solids on which the components can be adsorbed, or liquids or solids containing chemicals with which the components can react. The total volume of air passed is metered by the pump, and both pump and extraction apparatus are placed appropriately about the person to obtain a representative sample. The pumps have been criticized as bulky, heavy, and noisy, and therefore an encumbrance to the wearer, and also as expensive.

An alternative method known as 'passive' or diffusive sampling, which does not require a pump or air mover, has gained popularity in recent years. Diffusive samplers operate by allowing the gas and vapor molecules to make their own way to the collection medium by diffusion along a carefully controlled path. The rate of movement, which is a function of the diffusion coefficient of the gas or vapor in air and the path geometry (Figure 1), can be derived from Fick's first law of diffusion:

$$W = -DA \frac{dc}{dx}$$

where  $W$  is the rate of mass transfer ( $\text{ng s}^{-1}$ ),  $D$  is the diffusion coefficient ( $\text{cm}^2 \text{s}^{-1}$ ),  $A$  is the cross-sectional area of the diffusion path ( $\text{cm}^2$ ), and  $dc/dx$  is the rate of change in concentration over the



**Figure 1** The geometry of the diffusion path.

diffusion path ( $\text{ng cm}^{-3} \text{ cm}^{-1}$ ). If the diffusion path is considered to be the length  $L$  (cm) of the air gap in the sampler, and the concentration at the sorbent is considered to be negligible (the so-called ‘zero-sink’ condition), then the mass  $M$  (ng) adsorbed by the sampler from an atmosphere of concentration  $C$  ( $\text{ng cm}^{-3}$ ) in time  $t$  (s) is given by

$$M = D \frac{A}{L} C t$$

The uptake rate of the sampler is given by the expression  $D(A/L)$ , which has units ( $\text{cm}^3 \text{ s}^{-1}$ ) dimensionally equivalent to a volume flow rate. The geometry of the sampler is known from measurement, but very few diffusion coefficients have been measured. They can be calculated, assuming certain rules for calculating molecular size, but these need to be correlated with experimental determinations of uptake rates from atmospheres of known concentration. This gives a correction factor that depends on the chemical makeup of the molecule, and which can then be used to adjust the calculated values of other, similar molecules.

Fick’s second law of diffusion can be used to derive an expression for the time required to set up steady-state conditions within the sampler once the diffusion path is exposed to an ambient vapor concentration, which can also be applied to the behavior of molecules within the diffusion path after the sampler is closed. The time taken for a molecule to traverse the diffusion path is given by  $L^2/D$ , which for benzene is 11 s when the diffusion path is 1 cm. When diffusive samplers, especially those where the diffusion path is long, are used to sample for short periods of time, these time-related effects can be significant sources of error. Sampling transient and rapidly fluctuating peaks of concentration provides another source of time-related errors. In this case, however, the finite bounds on the height and width of real peaks are a finite bound on the maximum error. These types of errors have been carefully considered and discussed in the literature. Experimental validations often provide evidence

of a higher uptake rate during short ( $<15$  min) exposure periods.

The diffusion coefficient is affected by temperature and pressure, but so too is the concentration. The mass collected by the sampler is therefore proportional to the square root of the absolute temperature and independent of pressure. The temperature effect should be less than  $0.2\% ^\circ\text{C}^{-1}$ , and is normally ignored, since the variation from the measured uptake rate should be less than 5% from room temperature to  $0^\circ\text{C}$  and from room temperature to  $40^\circ\text{C}$ .

Results from diffusive sampling are often provided in terms of parts per million by volume ppm under the sampling site conditions of temperature and pressure. When the limit values to which these concentrations will be compared are stated in the same terms, then the statements in the preceding paragraph hold true. However, the origin of many limit values used around the world are the TLV<sup>®</sup>s of the American Conference of Governmental Industrial Hygienists (ACGIH). These values are stated in units of mass per unit volume (e.g.,  $\text{mg m}^{-3}$ ), while TLVs for gases and vapors are also stated in units of ppm at NTP (or  $25^\circ\text{C}$ , 1 atm, in the USA). In order to compare a result to these values, the diffusion coefficient is corrected for the difference between NTP and the sampling site pressure and temperature, but the concentration is not. This procedure is to be used, for example, for comparing results to the US Occupational Safety and Health Administration (OSHA) permissible exposure limits (PELs). The potential effect of this calculation is significant and could be as much as 20%, given extremes of temperatures or altitude.

The diffusion pathlength is affected by ambient air movement. Strong air currents around the face of the sampler may cause turbulent effects within the diffusion gap, effectively reducing the diffusion path, while weak air currents may not be sufficient to replenish molecules that have migrated to the sampler, effectively extending the diffusion path beyond the sampler face. The diffusion pathlength is also affected if the concentration at the sorbent face is not zero; therefore, the two principal factors that need to be considered in the application of a diffusive sampler are the geometry of the sampler and the type of sorbent. These two factors have the greatest influence on sampler performance.

## Types of Sampler

### ‘Tubes’ versus ‘Badges’

The diffusion path to the sorbent can take many forms. The simplest version is an open tube. This was

the construction used for sampling nitrogen dioxide in the original investigations into diffusive sampling by Palmes, and it is still in use today. An open tube may be affected by air currents disturbing the still air conditions required in the diffusion path, so that tubes without a windscreen must be long with respect to their diameter, with a ratio in excess of 3:1. Unfortunately, a long diffusion path results in a lower uptake rate (typically  $\sim 0.3\text{--}3.0\text{ ml min}^{-1}$ ) and much increased response time, reducing the application for short-term exposure determinations. However, the low uptake rate also reduces the possibility of starvation at the open face, so that the tube design can be used in areas of reduced air movement; and the low uptake rate also decreases the possibility of sorbent saturation, so that much longer sampling periods, of the order of weeks or months, can be used.

One major advantage of a tube design is the potential for gas-flushing organic vapors from open-ended tubes in a thermal desorption system directly into an analytical device, usually a gas chromatograph with flame-ionization detector or mass spectrometer. An example is shown in Figure 2 (second from left). The convenience of this procedure makes it very popular, although the analysis is limited to a single shot.

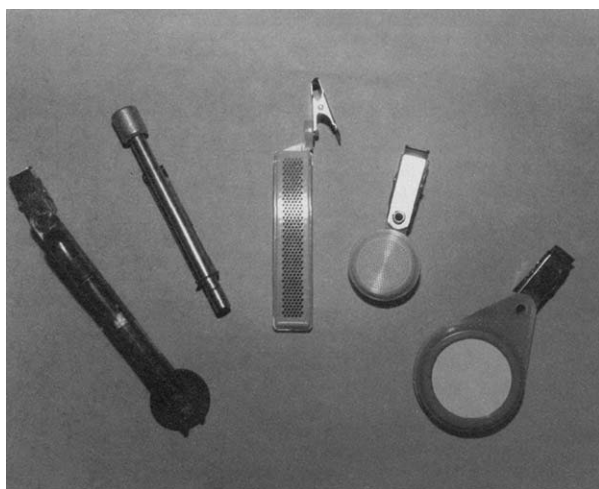
A second form of construction is to use a number of shorter, narrower tubes, which at the extreme can be reduced to a thin, porous membrane. The short

diffusion path results in a faster response time, and also combines with the larger overall sorbent surface area to give much higher uptake rates. This situation is better for short-term sampling, but may lead to the possibility of overloading at high exposure doses. The higher uptake rate of these samplers (typically  $6\text{--}60\text{ ml min}^{-1}$ ) requires more air movement, normally at least  $10\text{--}20\text{ cm s}^{-1}$ , around the open face in order to prevent starvation. With personal monitoring, natural movements of the body are assumed to give rise to sufficient air movement, but problems can still arise where workers are in sedentary occupations or where their movements are severely restricted. While badge-type samplers are not prone to effects from air movement parallel to the face, since each individual diffusion element normally has a length-to-diameter ratio of at least 3:1, perpendicular air movement has been shown to be able to cause a 'scooping' effect if there is communication between the bases of the tubes. A windscreen made from a porous membrane, sited above or below the diffusion path, will prevent this. Examples of these 'badge-type' samplers are shown in Figure 2 (third, fourth, and fifth from left).

### The Choice of Sorbent

The best choice of sorbent is one that will react with the component of interest to produce a nonvolatile, easily analyzable derivative. Indeed, if the derivative is colored, the analysis can be of extremely low cost and can be performed onsite. This was a very exciting possibility that was investigated at an early stage in the development of diffusive samplers, and a number of samplers have been produced commercially. Many use water-soluble reagents in aqueous solution as the sorbent, held in place by hydrophobic or high-water-content polymer membranes. These membranes can act as the control over mass transfer, or they can be used together with a diffusive air gap. Analysis requires an inexpensive spectrophotometer. Reagents coated onto an inert substrate can also be used and can be analyzed by reflectance spectrophotometry, although cheap and accurate field instrumentation has only recently become available.

There are a number of drawbacks to this technique, such that it is generally employed for certain specific gases and vapors. The reagents and products must be stable, and the analysis must be sensitive and free from interference. Finally, the method must satisfy stringent precision and accuracy requirements for use in hygiene investigations where compliance with government regulated exposure guidelines must be demonstrated; otherwise, the technique still has value in screening tests. Greater precision and accuracy are normally attainable with a laboratory



**Figure 2** Types of diffusive samplers. From the left: a colorimetric length-of-stain sampler; a tube-type sampler for thermal desorption; a badge-type sampler with multiple channels and no windscreen, containing a removable sorbent pad for solvent desorption; a badge-type sampler with multiple channels and an internal windscreen, containing granular sorbent for *in situ* solvent desorption; a badge-type sampler with an external windscreen, behind which is an open diffusion path, and containing a sorbent pad for *in situ* solvent desorption.

analytical technique, such as liquid chromatography (LC), ion chromatography, or polarography. The gases and vapors most frequently applicable to this method are inorganic compounds; for example, chlorine, ammonia, sulfur dioxide, and hydrofluoric acid. Formaldehyde is the compound most frequently determined by this method.

The use of colorimetric detector tubes in a diffusive mode is an extension of this type of method. In this case, reaction with a substrate contained in a tube produces a color change, the extent of which is related to the applied dose. In addition to the drawbacks of colorimetric analysis mentioned above, there is a further inaccuracy involved in reading the extent of the color change, which is complicated by being nonlinear with respect to exposure dose (concentration multiplied by time). An example of this type of sampler is shown in **Figure 2** (far left).

The preferred technique for sampling organic vapors is collection on a porous sorbent by the van der Waals forces of adsorption (particularly the London dispersion force), with later desorption and instrumental analysis. The rate of adsorption at the sorbent surface is extremely fast and is not normally considered a rate-controlling step, and the adsorption equilibrium is normally shifted far enough in the direction of the adsorbed phase that the concentration at the sorbent surface can be regarded as insignificant. All molecules that arrive at the surface are therefore adsorbed. The adsorbed vapors are extracted from the sorbent by means of a solvent or heat, and analysis is normally carried out by gas chromatography (GC), or less often by LC.

Sorbents differ in their affinity for different vapors, and sometimes this will cause a particular sorbent-vapor combination to deviate from the ideal behavior described above. Weak sorbent-sorbate interactions can cause the concentration of the vapor to be significantly greater than zero at the sorbent surface, and to increase with increasing exposure dose. This will result in saturation where the sorbent is a thin layer, as in badge-type samplers, or a progressive increase in the diffusion path through the sorbent where it is disposed in a thick bed as in many tube-type samplers. Both effects produce the same result, a progressive reduction in the sampling rate with time, leading to an underestimation of the applied concentration. In addition, if the concentration at the open face of the sampler is reduced to zero, adsorbed sample can re-equilibrate with the vapor phase and diffuse back out of the sampler. This is known as reverse diffusion, and it too results in an underestimate of the true exposure dose.

In general, the sorbents used in the tube-type samplers for thermal desorption tend to be porous

aromatic polymers, or graphitized carbon blacks that adsorb little water to interfere with the GC analysis, and where the van der Waals forces of adsorption are not so strong as to require excessive temperatures for desorption. Because of these weak interactions, saturation and reverse diffusion must be expected and guarded against in the method evaluation. Carbon and zeolite molecular sieves have been used to adsorb the most volatile molecules, but water adsorption makes analysis by thermal desorption more complex.

Conversely, the preferred sorbent for the badge-type samplers is activated charcoal, which has a greater capacity for organic vapors. Saturation and reverse diffusion are less likely to occur except with the most volatile compounds, e.g., acetone or methylene chloride. One method of counteracting these effects is by means of two layers of sorbent, the second acting as a backup analogous to the backup section in pumped sorbent sample tubes. Unfortunately, research has shown that reverse diffusion takes place at the same time as migration to the second layer, and sample loss is slowed but not avoided completely. Extra sorbent sections, therefore, are not a completely reliable guarantee of sampling accuracy.

Desorption of these samplers is by means of a solvent, which often can be added directly to the closed sampler. Solvent desorption of active charcoal is not usually as efficient as thermal desorption with porous polymers for a number of reasons. First, desorption efficiency depends on the partition equilibrium of the compound between the solvent and the charcoal. Second, a certain proportion of the sample may be lost irretrievably within bottle-necked pores, by reaction with surface-active sites, or by partition into an aqueous phase resulting from the simultaneous desorption of adsorbed water. This latter can be retrieved by desorption using a more polar solvent mixture. In general, these losses are somewhat balanced by loss of the solvent through adsorption on the charcoal. The resulting recovery can range from 80% to 110% of the applied dose, and this needs to be carefully characterized. Recoveries of polar and reactive compounds tend to fall further at lower loadings. Losses may also occur during storage of the sampler after use, and this too must be carefully characterized. Porous polymers are not immune to sample losses during thermal desorption or storage, although it occurs less often. Perhaps the biggest drawback of solvent desorption is the reduced analytical sensitivity caused by the solvent dilution of the sample. Thermal desorption without focusing also results in sample dilution because the amount of carrier gas required for sample desorption and correction cannot all be passed into a capillary GC column.



A splitter valve may be required to dispose off much of the sample.

For most general purposes the choice of tube versus badge revolves practically around the preferred desorption technique, provided the correct sorbent is selected for the tube-type sampler. Tenax<sup>®</sup> TA is often used, even though it has a small surface area and low capacity for volatile compounds. Porapak N is sometimes employed for polar compounds, but the capacity is also quite low and is compromised by high relative humidities. Chromosorb<sup>®</sup> 106 is a good all-round medium, with a surface area  $\sim 700\text{--}800\text{ m}^2\text{ g}^{-1}$  and hydrophobic character, but there is a gradually increasing background caused by slow polymer degradation or trapping of adsorbed material that is released on subsequent uses. This can be a problem in sampling very low concentrations, such as in indoor air-quality surveys.

One important consideration is the possibility of reuse of thermal desorption tubes, since they are cleaned in the desorption step. Charcoal badge-type samplers are of single-use nature but cheaper. Solvent desorption also allows multiple analyses from a single sample to give greater accuracy.

## Validation of the Sampling Method

The precision and accuracy of any measurement method must fall within acceptable bounds. Precision is normally denoted by the bias, or deviation of an average measurement from the 'true' measurement as recorded by an unbiased reference method. Accuracy is normally denoted by the spread of results about the mean, and is defined as twice the pooled coefficient of variation of the method. According to the National Institute of Occupational Safety and Health (NIOSH) in the United States, a method is acceptable if 95% of the results fall within 25% of twice the pooled coefficient of variation plus the absolute bias (regardless of sign). A lower level of overall accuracy for measurements taken at concentrations below regulated limit values is often acceptable and the Committee for European Normalization (CEN) suggests a relaxation to 50% at concentrations below one-half the limit value, or over the full range if the measurements are made for screening purposes only.

Experimental tests are required to confirm acceptable sampler performance. NIOSH, and the Health and Safety Executive in the United Kingdom, have published diffusive sampler evaluation protocols that cover the effects of exposure, time, concentration, desorption efficiency, storage stability, reverse diffusion, temperature, humidity, presence of interfering compounds, face velocity, and orientation. Many of

these factors are varied together in a multifactorial study to determine the possibility of interactions. Precision and accuracy are determined from the combined results of these experiments. A working group of CEN is currently drafting a similar protocol. All of these protocols require laboratory exposure studies in atmospheres of closely controlled composition. The expense of producing good-quality atmosphere generation rigs together with the time and costs of running the full set of experiments (equivalent to about three technician-months per validated compound) have been a barrier to the widespread adoption of these protocols. A useful suggestion for reducing the number of required experiments is the concept of bilevel validation, which involves performing the full set of validation tests on one member of a homologous series, together with a reduced set of tests on the higher members of the series if the first member passes. This allows the validation of about one series (e.g., aliphatic hydrocarbons, aromatic hydrocarbons, chlorinated aliphatic hydrocarbons, ketones, acrylates, aliphatic alcohols, etc.) per year.

In addition to laboratory tests, field trials are recommended or required by the protocols. Two types of field trials are suggested, one being a large set of diffusive samplers and reference methods arranged in a static array, the other consisting of a number of pairs of diffusive samplers and reference methods as static or personal monitors. The results of these tests require cautious analysis to be sure they are effective. For instance, the effects of environmental conditions on the reference methods need to be known and accounted for, and the statistical tests most appropriate to the situation must be applied. It is unfortunate that to date very few samplers have been evaluated according to these protocols, since only through such tests can confidence in the measurements be assured. Table 1 gives a summary of the compounds validated using one particular badge-type sampler containing one or another of two different sorbents. All of the compounds tested according to the full protocol had acceptable overall accuracies. It is normally the case that well-designed diffusive samplers have precision and accuracy characteristics comparable to other accepted methods.

## Active versus Passive Sampling

As mentioned in the introduction, the chief drawback to the pumped system has been the pump itself, which is normally described as bulky, heavy, noisy, and expensive, and which requires frequent calibration and periodic maintenance and repair. However, manufacturers have addressed these criticisms over

**Table 1** Compounds validated using the SKC Inc 575 series badge-type sampler with (A) charcoal and (B) Anasorb<sup>®</sup> 747 sorbents

(A) Charcoal sorbent		
Aliphatics	Aromatics	Chlorinated
Pentane <sup>a</sup>	Benzene <sup>a</sup>	Methylene chloride <sup>a</sup>
Hexane	Toluene	<i>cis</i> -1,2-Dichloroethylene <sup>a</sup>
Heptane	Ethylbenzene	Trichloroethylene <sup>a</sup>
Octane	<i>o</i> -Xylene	Tetrachloroethylene <sup>a</sup>
Nonane	<i>m</i> -Xylene	Vinylidene chloride
Cyclohexane	<i>p</i> -Xylene	Chloroform
		Methylcyclohexane
Cumene	Carbon tetrachloride	
	<i>t</i> -Butyltoluene	1,2-Dichloroethane
		1,2-Dichloropropane
		1,1,1-Trichloroethane
		1,1,2-Trichloroethane
		1,2,3-Trichloropropane
		1,1,2,2-Tetrachloroethane
(B) Anasorb <sup>®</sup> 747 sorbent		
Ketones	Acrylates	
Acetone <sup>a</sup>	Methyl acrylate <sup>a</sup>	
2-Butanone (MEK)	Methyl methacrylate	
Methyl isobutyl ketone	Ethyl acrylate	
Diisobutyl ketone	Butyl acrylate	

<sup>a</sup>Full validation according to NIOSH protocol, all others partially validated according to bi-level validation theory.

the previous decade. Today's models are smaller, lighter, and quieter, and miniaturized versions with total volumes of 50 cm<sup>3</sup> or less will be available in the near future. Onboard computer calibration is possible, and pumps are generally of more rugged construction than before. In addition, the cost of a pump is spread over a very large number of individual samples, and the major cost per sample is the analysis. A pump provides the advantage of a series of samples, such as in a backup sorbent layer, to detect breakthrough from the primary collection layer. To ensure the same safety factor in diffusive sampling, two samplers with different uptake rates can be employed, the ratio between the collected samples being the critical factor in proving a good sample. One definite advantage of the diffusive sampler is that it can be left unattended during operation, while the operation of the sampling pumps must be checked regularly during use. In addition, it is not necessary to account for pump error.

Often the cost factor is represented as a contest between diffusive samplers with thermal desorption

on the one hand and pumped samplers with solvent desorption on the other. This is not quite an accurate picture, since there are plenty of examples of alternative techniques; that is, diffusive sampling with solvent extraction and pumped sampling with thermal desorption. Both thermal and solvent desorption can be automated and run overnight. The tubes used in pumped sample trains are less expensive than any diffusive sampler, but they are of single-use type, unlike thermal desorption tubes. Overall, the costs are probably not very different, and the differences are minimized further when including other costs of the sampling exercise, such as the costs of taking the sample, the analysis, and the interpretation of the result.

One very important difference between pumped and diffusive sampling occurs when aerosols or dusts containing the compounds of interest are present. Pumps are able to draw these species into an adsorbent tube or impinger, where the compounds of interest will be collected, although not necessarily with 100% efficiency. Diffusive samplers cannot collect anything other than gases or vapors, but if dust or aerosol becomes attached to the face it can outgas volatile components even after the sampler is closed. A heavy loading of dust, grease, or condensation on the sampler face can block the diffusive uptake, and tears or perforations of delicate membranes can also affect the uptake rate. Many problems associated with the use of diffusive samplers have been traced to careless handling and contamination.

Finally, it is important to remember that the sampling rate of a diffusive sampler is different for individual chemicals and it is not possible to provide uptake rates for complex mixtures, and that the sampling rate is fixed, while that of a pump can be adjusted up to 1000 ml min<sup>-1</sup> or more for additional sensitivity.

## Permeation Samplers

Diffusion through materials other than air is possible, and a number of polymer materials have been tried, either alone, or in combination with an air gap. It is important to distinguish these membranes from those that are porous, where the primary control of mass transfer is still diffusion in air. Diffusion through polymers is much slower than through air and is normally referred to as permeation. Permeation samplers have advantages (for instance, organic vapors may permeate through a hydrophobic membrane while interfering water vapor does not) but also have disadvantages, mainly as a result of variations in the thickness or composition of the membranes, and the lower uptake rate.



See also: **Personal Monitoring:** Active.

## Further Reading

- Berlin A, Brown RH, and Saunders KJ (eds.) (1987) Diffusive sampling: An alternative approach to workplace air monitoring. *Proceedings of the International Symposium*, Luxembourg, 1986. London: Royal Society of Chemistry (includes the NIOSH validation protocol).
- Brown RH, Harvey RP, Purnell CJ, and Saunders KJ (1984) A diffusive sampler evaluation protocol. *American Industrial Hygiene Association Journal* 45: 67–75.
- Guild LV, Myrmet KH, Myers G, and Dietrich DF (1992) Bi-level passive monitor validation – a reliable way of assuring sampling accuracy for a larger number of related chemical hazards. *Applied Occupational and Environmental Hygiene* 7: 310–317.
- Harper M and Purnell CJ (1987) Diffusive sampling – a review. *American Industrial Hygiene Association Journal* 48: 214–218.
- Harper M (2000) Sorbent trapping of volatile organic compounds from air. *Journal of Chromatography A* 885: 129–151.
- Harper M (2004) Assessing exposure to workplace chemicals: the role of exposure monitoring. *Journal of Environmental Monitoring* 6: 404–412.
- Palmes ED and Gunnison AF (1973) Personal monitoring device for gaseous contaminants. *American Industrial Hygiene Association Journal* 34: 78–81.
- Rose VE and Perkins JL (1982) Passive dosimetry – state of the art review. *American Industrial Hygiene Association Journal* 43: 605–621.

# PERVAPORATION

See **MEMBRANE TECHNIQUES:** Pervaporation

# PESTICIDES

**M N Vasilescu**, Institute of Public Health, Bucharest, Romania

**A V Medvedovici**, University of Bucharest, Bucharest, Romania

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

An exhaustive overview on pesticides should include aspects referring to their synthesis and formulation (and related quality control features), mechanisms of action, degradation and/or metabolism patterns, and their impact on environment and living organisms, including on human health. Analytical chemistry is directly involved in all these domains and takes into account at least three major difficulties: extreme diversity of the target compounds, increased sample matrix complexity, and imposed levels of detectability, which are continuously decreasing. The article starts by defining pesticides and related topics together with their classification criteria and class exemplification. Thus, attention is focused on the

analytical contributions to the topic. The sample preparation approach reviews analytical techniques involved in the isolation, concentration, and derivatization of pesticides and pesticide residues from real samples of different origin (environmental, food, clinical, forensic, etc.). Recent developments such as focused microwaves Soxhlet extraction, solid-phase extraction and microextraction on immunosorbents, shielded materials and molecular imprints, stir-bar sorptive extraction, supported liquid membrane extraction, and supercritical fluid and pressurized fluid extractions are discussed. The main achievements related to pesticide separation by means of thin layer, gas, liquid, supercritical fluid, micellar electrokinetic, and capillary electro-driven chromatographic methods are also discussed. Special emphasis is given to mass spectrometric detection for its inherent selectivity and high sensitivity. In some circumstances, other analytical methods such as spectrometric, immunochemical, and electrometric (including bio- and immunosensors) methods can be successfully used for pesticide analysis. Due to the complexity of the subject, more often only recent advances have been detailed at some length.

**S Głąb and A Hulanicki**, University of Warsaw, Warsaw, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The concept of pH is one of the most fundamental in chemistry, but has a broad significance in many other disciplines such as biochemistry, biology, physics, medicine, agriculture, environmental pollution, and is important in numerous practical fields, especially in chemical industry, environmental protection, food production, and others. This article discusses the basic definitions of pH, its effect on some fundamental chemical reactions of importance in analytical chemistry. Although the concept of pH is mainly used for aqueous solutions its extension to other solvents is mentioned, as well as the main methods of pH evaluation.

## General Concept

The concept of pH arose from the idea that the acidity of a medium was related to the concentration of hydrogen ions. This has been shown to be valid for water and water-like media, where the species responsible for acidity is the ion assumed by convention to be  $H^+$ . In fact, this is a solvated ion and the number of solvating molecules of the solvent is not strictly defined, but for dilute solutions it is assumed to be constant. Nevertheless, this solvated ion is sometimes shown as  $H_3O^+$  in water,  $CH_3OH_2^+$  in methanol,  $CH_3COOH_2^+$  in acetic acid, etc.

The first definition of pH, given by S P L Sørensen in 1909, defined pH as the negative logarithm of the hydrogen ion concentration:

$$pH = -\log_{10}[H^+] \quad [1]$$

The introduction of the concept of activity ( $a_H$ ) as a driving force in chemical and biochemical processes has modified this definition such that

$$pH = -\log_{10} a_H = -\log_{10}(\gamma_{m,H} m_H) \quad [2]$$

where  $\gamma_{m,H}$  is the single molal activity coefficient of the hydrogen ion and  $m_H$  is the molality of the hydrogen ion. As formally the logarithmic term is dimensionless, the correct full expression should be

$$pH = -\log_{10}(\gamma_{m,H} m_H / m^0) \quad [3]$$

where  $m^0$  is an arbitrary constant, representing the standard state condition, numerically equal to

$1 \text{ mol kg}^{-1}$ . For most analytical purposes, this pH scale is equivalent to the scale based on molarity, where

$$pH = -\log_{10}(c_H \gamma_{c,H} / c^0) \quad [4]$$

where  $c_H$  is the molar concentration, expressed in  $\text{mol dm}^{-3}$  or  $\text{mol l}^{-1}$ , and  $\gamma_{c,H}$  is the molar activity coefficient. The difference between the two scales depends on the density of water; at 298.15 K this difference is  $\sim 0.001$  pH. As in most practical applications the uncertainty of pH measurement is usually  $< 0.01$  pH, such a difference is insignificant. The definitions given above are only notational definitions because it is not possible to measure single ion ( $H^+$ ) activity.

## Operational Definition

For practical purposes, pH is defined by measuring the difference between the electromotive forces of two cells, both of which have identical indicator electrodes and identical reference electrodes. One of the cells contains the standard reference solution (S) and the other cell contains the solution whose pH is to be determined (X). Denoting the corresponding electromotive forces as  $E(S)$  and  $E(X)$ , and the pH values as  $pH(S)$  and  $pH(X)$ , the pH value of the solution X is defined as

$$pH(X) = pH(S) + \frac{F[E(S) - E(X)]}{RT \ln 10} \quad [5]$$

where  $R$ ,  $T$ , and  $F$  are the gas constant, the thermodynamic temperature, and the Faraday constant, respectively. Hydrogen electrodes are generally used as the indicator electrodes in fundamental studies, but in practical measurements they can be replaced by any hydrogen-ion responsive electrode that works properly in the pH range studied. Among them mostly are used pH-sensitive glass electrodes. The reference electrode is connected through a potassium chloride bridge, which has KCl concentration of not less than  $3.5 \text{ mol kg}^{-1}$ .

The pH scale is usually based on primary buffer solutions (PS). A number of substances have been chosen as primary pH standards and their pH values are tabulated in Table 1. Their pH values have been accurately determined using a cell without liquid junction potential composed of the standard hydrogen electrode and a silver–silver chloride electrode



**Table 1** Values of pH(PS) for primary standard reference solutions at 25°C

Standard reference solution	pH
0.0341 mol kg <sup>-1</sup> potassium hydrogen tartrate	3.557
0.05 mol kg <sup>-1</sup> potassium dihydrogen citrate	3.776
0.05 mol kg <sup>-1</sup> potassium hydrogen phthalate	4.005
0.025 mol kg <sup>-1</sup> disodium hydrogenphosphate + 0.025 mol kg <sup>-1</sup> potassium dihydrogenphosphate	6.865
0.03043 mol kg <sup>-1</sup> disodium hydrogenphosphate + 0.008695 mol kg <sup>-1</sup> potassium dihydrogenphosphate	7.413
0.01 mol kg <sup>-1</sup> sodium tetraborate	9.180
0.025 mol kg <sup>-1</sup> sodium hydrogencarbonate + 0.025 mol kg <sup>-1</sup> disodium carbonate	10.012

Knowing the standard potential of the Ag–AgCl electrode, and the mean ion activity coefficient  $\gamma_{\pm}$  of 0.01 mol kg<sup>-1</sup> HCl from independent measurements the pH(PS) can be calculated as

$$\text{pH(PS)} = \left[ -\log_{10} \left( \frac{m_{\text{H}^+} \gamma_{\text{m,H}^+} \gamma_{\text{m,Cl}}}{m^0} \right) \right] + \log_{10} \gamma_{\text{Cl}} \quad [6]$$

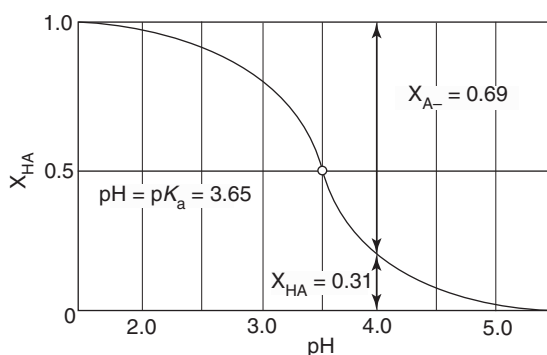
where  $\gamma_{\text{Cl}}$  is obtained from the Bates–Guggenheim convention. The primary buffer solutions should be characterized by large buffer capacity, small temperature coefficient, small residual liquid junction potential, ionic strength  $\leq 0.1$  mol kg<sup>-1</sup>, and must have a certificate of the national metrological institute. The best results (uncertainty  $\pm 0.002$  pH) are obtained when the standard used as reference solution is similar (in terms of pH, composition, and ionic strength) to the solution being measured. Measurements within the pH range 3–10 can have an uncertainty of pH measurements  $< 0.02$  pH.

At constant ionic strength the activity coefficients may be assumed to be constant and therefore the difference between the activity  $\text{pH} = -\log_{10}(c_{\text{H}^+} \gamma_{\text{c,H}}/c^0)$  and concentration  $\text{pH}_c = -\log_{10}(c_{\text{H}^+}/c^0)$  is constant. If the standardization is carried out not by standard buffer solutions but by using solutions containing known hydrogen ion concentration, e.g., dilute HClO<sub>4</sub>, then measurements of the electromotive force allow the determination of  $\text{pH}_c$ , assuming that all solutions have the same ionic strength. Such a procedure is used in the study of protonation or complexation equilibria.

## Influence of pH on Chemical Equilibria

### Simple Protolytic Equilibria

pH is the parameter most often measured in chemistry, particularly in analytical chemistry. This is

**Figure 1** Distribution diagram (mole fraction versus pH) for formic acid ( $\text{p}K_a = 3.65$ ).

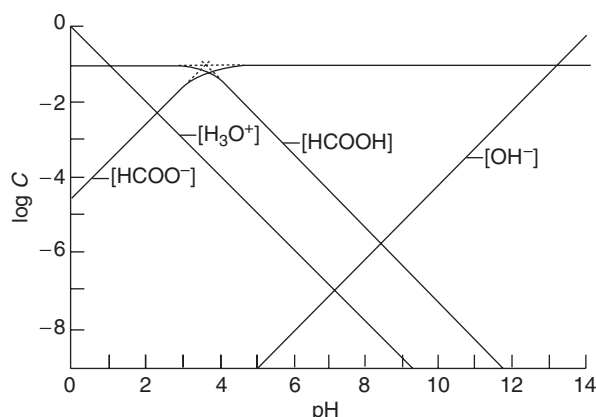
because the pH influences the position of chemical equilibrium of the majority of chemical reactions in aqueous solutions, and this can also be extended to other media. In several instances pH also influences the reaction kinetics, changing the rate and acting as a catalyst.

The simplest case involves the equilibrium of an acid–base reaction (in the Brønsted–Lowry sense). In typical acid–base titrations, the pH value is the analytical signal observed during the titration, but knowledge of the pH is essential for understanding the process and for choosing the optimal conditions for endpoint evaluation. The equilibrium of the reaction  $\text{HA} \rightleftharpoons \text{H}^+ + \text{A}^-$  is determined by the values of the acid dissociation constant,  $K_a = [\text{H}^+][\text{A}^-]/[\text{HA}]$  or the protonation constant  $K_H = [\text{HA}]/[\text{H}^+][\text{A}^-]$ . The influence of pH on the position of the equilibrium can be shown graphically by means of a mole fraction versus pH diagram (Figure 1), or by a plot of log concentration of a given species versus pH (Figure 2). Such diagrams are useful in predicting the pH at the endpoint of a titration,  $\text{pH}_{\text{end}}$ .

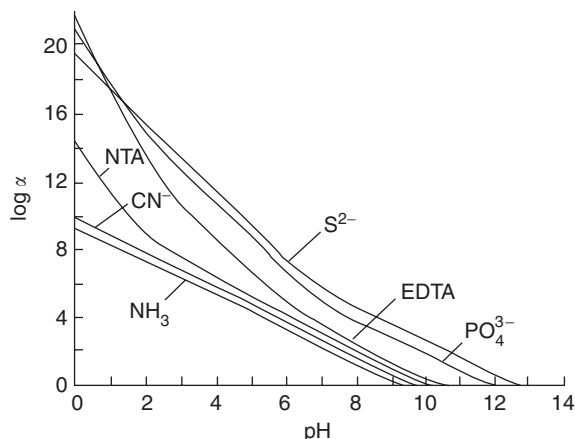
### Complex Equilibria

The concept of side reaction coefficients, introduced by A. Ringbom, is useful in describing the effect of pH on various equilibria. These coefficients combine the total (analytical) concentration,  $[\text{X}']$ , of the species of interest, with the concentration of its form  $[\text{X}]$ , which undergoes reaction. For example, for ethylenediaminetetraacetic acid (EDTA) the reacted form is the  $\text{EDTA}^{4-}$  anion, but the total concentration also includes all protonated forms. Thus, the side reaction coefficient,  $\alpha_{\text{EDTA(H)}}$ , is given by

$$\begin{aligned} \alpha_{\text{EDTA(H)}} &= [\text{EDTA}']/[\text{EDTA}] \\ &= 1 + K_{\text{H1}}[\text{H}^+] + K_{\text{H2}}K_{\text{H1}}[\text{H}^+]^2 \\ &\quad + K_{\text{H3}}K_{\text{H2}}K_{\text{H1}}[\text{H}^+]^3 \\ &\quad + K_{\text{H4}}K_{\text{H3}}K_{\text{H2}}K_{\text{H1}}[\text{H}^+]^4 \end{aligned} \quad [7]$$



**Figure 2** Logarithmic diagram for the formic acid–formate ion system for total concentration  $C = 0.1 \text{ mol l}^{-1}$ .



**Figure 3** Dependence of side-reaction coefficients, as  $\log \alpha$ , on pH for some bases (EDTA, anion of ethylenediaminetetraacetic acid; NTA, anion of nitrilotriacetic acid).

For various common ligands used in analytical chemistry, such side reaction coefficients can be presented graphically (**Figure 3**).

A similar procedure can be used in the case of such acids (according to the Brønsted–Lowry theory) as metal ions forming hydroxy complexes at elevated pH. Considering the reaction



as acid dissociation, the coefficient  $\alpha$  can be written as

$$\alpha_{M(OH)} = [M]/[M'] \\ = 1 + K_{Hn}[H^+] + K_{Hn}K_{H(n-1)}[H^+]^2 + \dots \quad [8]$$

or considering this reaction as the formation of hydroxo complexes with successive stability

constants  $K_1$ ,  $K_2$ , etc. of metal– $OH^-$  complexes:

$$\alpha_{M(OH)} = [M']/[M] \\ = 1 + K_1[OH^-] + K_1K_2[OH^-]^2 + \dots \quad [9]$$

The functions of  $\log \alpha_{M(OH)}$  versus pH are tabulated for a number of ions in **Table 2** and are shown graphically for some typical metal ions in **Figure 4**. The accuracy of such presentation is not very high. The accuracy of the calculations is also limited because of the dependence of equilibrium constants on the ionic strength and temperature, the discrepancies between data presented by different authors, and in particular because of the uncertainties about data in the range where polynuclear species may be formed. Nevertheless, in many analytical predictions and calculations such approach is very useful.

The pH value influences the solubility of precipitates of hydroxides of metals, for which the coefficient  $\alpha$  is  $>1$ . Using the concept of side reaction coefficients the conditional solubility products,  $K'_{s0}$ , can be calculated:

$$K'_{s0} = K_{s0}/\alpha_{M(OH)} \quad [10]$$

The  $\alpha_{M(OH)}$  coefficients should take into account the formation of polynuclear species.

The solubility of hydroxides can be also expressed in terms of the dependence of the total concentration of the soluble species of metal ion  $[M']$  on pH (**Figure 5**):

$$[M'] = K_{s0}/\alpha_{M(OH)}[OH^-]^n \quad [11]$$

For the salts of weak acids this equation can be extended by additionally taking into account the  $\alpha$  coefficient of the anion. Thus,

$$K'_{s0} = K_{s0}/\alpha_{M(OH)}\alpha_{L(H)} \quad [12]$$

Diagrams for several insoluble systems are shown in **Figure 6**.

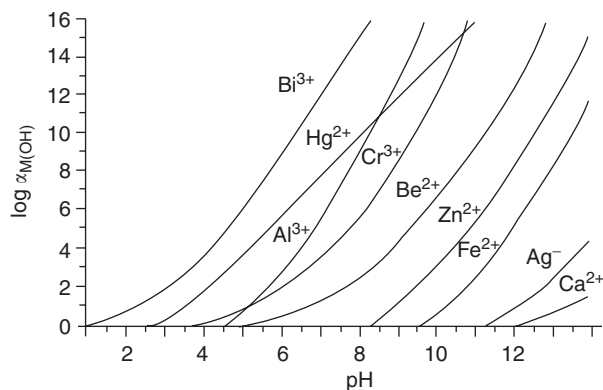
The concept of conditional stability constants can also be applied to complexation reactions. The constant can be evaluated in the same way as shown in eqn [12] only instead of solubility products the complex stability constants,  $\beta'_{ML}$ , are considered. Thus,

$$\beta'_{ML} = \beta_{ML}/\alpha_{M(OH)}\alpha_{L(H)} \quad [13]$$

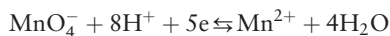
The function  $\beta'_{ML}$  versus pH for a number of EDTA complexes of metals is shown in **Figure 7**. Such plots are of direct use in optimizing conditions for compleximetric titrations. They can be further combined with  $\alpha$  coefficients for masking reagents, as shown by the dashed line in **Figure 7** for  $1 \text{ mol l}^{-1}$  ammonia buffer.

**Table 2** The  $\alpha_{M(H)}$  coefficients for metal ions in aqueous solutions for ionic strength 0.1 and temperature 20°C

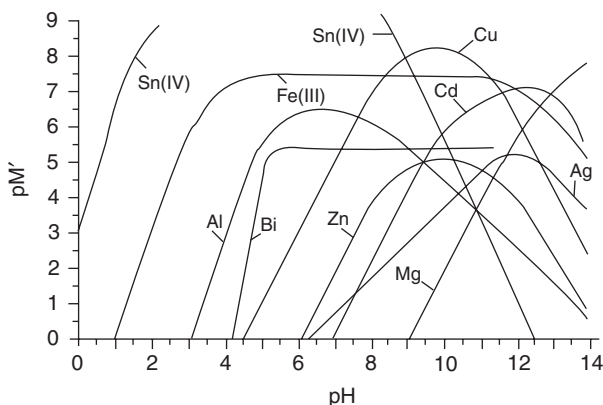
Ion	pH											
	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	
Ag <sup>+</sup>										0.1	0.6	
Al <sup>3+</sup>				0.5	2.6	5.6	8.8	12.7	16.5	20.5	24.5	
Bi <sup>3+</sup>	0.6	2.8	3.6	5.9	8.8	11.8	14.5	17.5	20.5	23.5	27.5	
Cd <sup>2+</sup>									0.2	1.5	3.4	
Co <sup>2+</sup>								0.1	1.4	3.3	5.3	
Cr <sup>3+</sup>			0.2	0.8	1.9	3.7	5.9	8.6	12.2	16.2	20.2	
Cu <sup>2+</sup>							0.2	1.1	2.7	5.5	8.9	
Fe <sup>2+</sup>								0.1	0.5	2.1	5.0	
Fe <sup>3+</sup>	0.1	0.7	2.4	5.1	8.0	10.9	14.6	18.6	22.6			
Hg <sup>2+</sup>		0.3	1.9	3.9	5.9	7.9	9.9	11.9	14.3	17.2	20.2	
Mg <sup>2+</sup>											0.2	
Mn <sup>2+</sup>										0.4	1.9	
Ni <sup>2+</sup>									0.9	3.3	6.1	
Pb <sup>2+</sup>							0.4	1.2	2.9	5.2	8.0	
Pd <sup>2+</sup>	0.4	1.3	5.2	7.2	9.2	11.2	13.2					
Sn <sup>2+</sup>		0.1	0.9	2.7	4.7	6.7	8.7	11.0	13.4	17.4	21.4	
TiO <sup>2+</sup>	0.2	1.1	3.0	5.0	7.0	9.0	11.0	13.0	15.0			
Tl <sup>3+</sup>	2.3	5.2	8.1	11.1	14.1	17.1	20.1	23.1				
U <sup>4+</sup>	1.1	3.1	5.3	8.5	12.9	17.9	22.9					
UO <sub>2</sub> <sup>2+</sup>				0.4	1.1	2.9	4.9	6.9	8.9	10.9		
VO <sub>2</sub> <sup>2+</sup>				0.4	1.1	2.9	4.8	6.8	8.8	10.8		
Zn <sup>2+</sup>							0.1	0.9	2.9	5.1	7.8	
Zr <sup>4+</sup>	1.9	3.9	6.1	9.1	13.4	18.0	23.0					

**Figure 4** Dependence of side-reaction coefficients, as  $\log \alpha_{M(OH)}$ , for some metal ions (cf. **Table 2**).

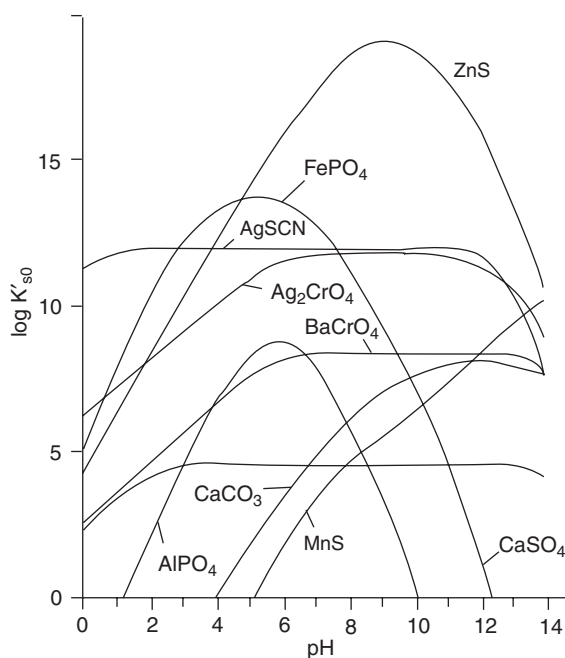
The coupled effect of pH change and redox potential on the equilibrium position of analytical reactions is very important. For some reactions the pH effect is evident from the equation of the chemical reaction considered, e.g., in the case of the redox reaction in the  $MnO_4^-/Mn^{2+}$  system:



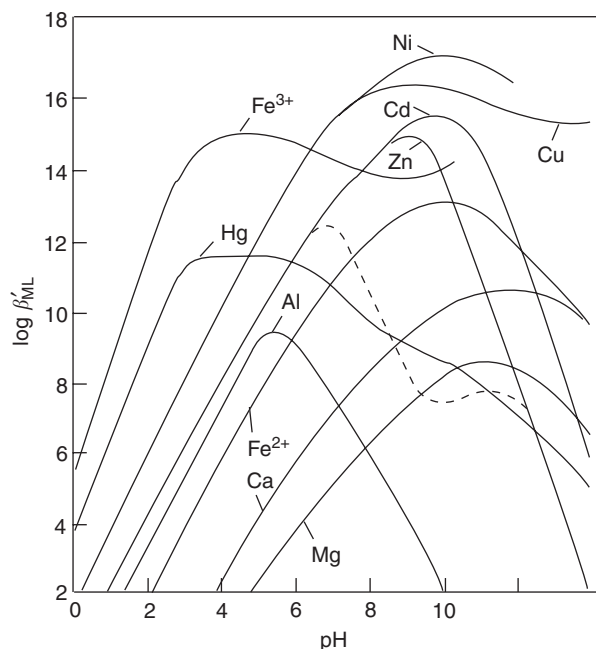
Here a pH decrease shifts the equilibrium to the formation of  $Mn^{2+}$ . Other types of reactions do not

**Figure 5** Dependence of solubility of some metal hydroxides on pH. Solubility is expressed as  $pM'$ , where  $M'$  is the total concentration of all soluble metal species in equilibrium with the insoluble hydroxide.

show direct evidence of a pH effect, e.g., the  $Fe^{3+}/Fe^{2+}$  system. However, in this system both participants are pH sensitive through their side reactions (the formation of hydroxo complexes) and undergo changes in their redox potentials. This complicated situation is usually presented in the form of a logarithmic diagram called a Pourbaix diagram. Such diagrams can be plotted for reactions that occur in

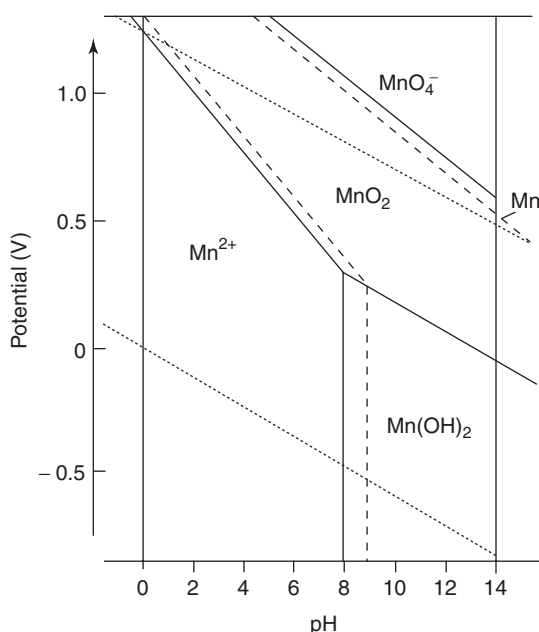


**Figure 6** Dependence of the conditional solubility product,  $K'_{s0}$ , on pH for some selected slightly soluble precipitates.



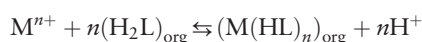
**Figure 7** pH dependence of conditional stability constants,  $\beta'_{ML}$ , of some analytically important EDTA-metal complexes. The dashed line indicates the effect of  $1 \text{ mol l}^{-1}$  ammonia/ammonium ion buffer on the conditional constants of the Zn-EDTA complex.

the absence of other reactants except those that are present in the aqueous media (Figure 8) or in the additional presence of, e.g., complexing substances.



**Figure 8** The predominance diagram (Pourbaix diagram) indicating the pH-redox potential interaction for manganese species in aqueous solution. The dotted lines indicate the thermodynamic water stability region. The full lines correspond to  $1 \text{ mol l}^{-1}$  concentration of species in the liquid phase, the dashed lines to  $10^{-2} \text{ mol l}^{-1}$  concentration.

All protolytic reactions (which include many separation and determination procedures) are influenced by pH. An example is the extraction of certain metal chelates, which is the basis for separation, preconcentration, and determination of metal ions. For the extraction reaction:

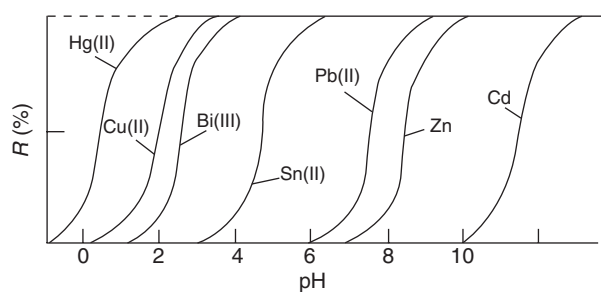


the distribution ratio,  $D$ , is related to the extraction constant,  $K_{\text{ex}}$ , the reagent concentration,  $[\text{H}_2\text{L}]$ , and the pH of the aqueous phase:

$$\log D = \log K_{\text{ex}} + n\text{pH} + n \log [\text{H}_2\text{L}]_{\text{org}} \quad [14]$$

The pH value for which 50% of metal as its chelate is extracted, i.e.,  $D = 1$ , and for  $[\text{H}_2\text{L}] = 1$ , is equal to  $(\log K_{\text{ex}})/n$ . This is termed  $\text{pH}_{1/2}$ , and has a value characteristic for the particular reagent, solvent, and the metal ion being determined. For example, in the case of metal extraction in the form of dithizonates using carbon tetrachloride as a solvent, the  $\text{pH}_{1/2}$  values for Pb(II), Zn(II), In(III), and Bi(III) are 8.0, 5.0, 3.8, and 2.0, respectively. Similar considerations apply to other chelating reagents as, e.g., 8-hydroxyquinoline, acetylacetone, and thenoyltrifluoroacetone. Such relationships are often presented in graphical form (Figure 9).





**Figure 9** The extraction percentage versus pH diagram for some metal dithizonates extracted by carbon tetrachloride.

A number of examples of the pH effect on the position of the half-wave potentials in polarography can be found when pH directly influences the  $\alpha$  coefficients of the metal ion as well as the protonation of the buffer components in particular complexing ligands, such as polyamine polycarboxylic acids. These two effects often shift the wave toward more negative potentials. However, strictly thermodynamic considerations do not always make prediction possible as kinetic factors (reversibility of the electrode process) may also contribute to the final result.

## pH in Nonaqueous and Mixed Solvents

The pH concept is most commonly used for dilute aqueous media; however, a similar formalism can be extended to other systems. The extent of the pH scale, which in aqueous media can be described as 14 units, depends on the autoprotolysis constant of the amphiprotic solvent, so that the equivalent range, e.g., in methanol, equals 16.7 units, in sulfuric acid 2.9 units, and in acetic acid 14.5 units. In such solvents, as in water, the pH of neutrality corresponds to the middle of this range. Such reasoning cannot be extended to protophilic (e.g., pyridine, ethers), and aprotic (e.g., hydrocarbons) solvents, for which the  $\log a_{H^+}$  scale is from one or both sides, respectively, unlimited.

The mutual correlation of pH scales in different solvents creates more problems as a number of effects must be considered, such as electric permittivity and the solvation of ions. Potentiometric measurements of the hydrogen ion activity can be made using the hydrogen electrode in many solvents; however, more difficulties are concerned with the reference electrode. When using an aqueous reference electrode the liquid junction potential on the boundary of two solvents is large. For a reference electrode in a nonaqueous solvent the interactions of various ions must be considered. Some good approximation can be reached when the reference electrode uses a

system that is not influenced by solvation processes. At least, in principle, such a system may be based on the redox couple ferrocene/ferricinium ion. The potential of the cell



should, at least, in principle, be independent of the solvent, and give the measure of the actual proton activity in those solvents. The activity of the proton in one solvent can be correlated to that in another solvent. Denoting the activity of the solvated proton  $(a_{H^+})_S$  in the solvent S, and  $(a_{H^+})_W$  in water, then

$$(a_{H^+})_W = (a_{H^+})_S \gamma_{t(H^+)} \quad [15]$$

where  $\gamma_{t(H^+)}$  is the activity coefficient for transfer of the  $H^+$  ion from solvent S into water (at infinite dilution). On such a basis it is possible to correlate the pH scales for different solvents. Similar considerations can be applied to mixed organic–aqueous systems.

In each individual solvent system the practical use of pH is based on standard buffer solutions, which have been proposed, e.g., for water–methanol and water–ethanol systems.

## Experimental Determination of pH

The oldest method of pH determination is based on the color change of some natural or synthetic dyes, either in the form of a solution or as indicator papers. Even in the best case the accuracy of pH determination is not better than 1 pH unit. Somewhat more accurate (uncertainty 0.1 pH) may be achieved by the spectrophotometric measurement of absorbance of a solution containing such indicators. The pH range for a single protonation equilibrium of the indicator spans not more than 2 pH units around the pH value corresponding to the  $pK_H$  value of the indicator. The disadvantage of optical pH sensing is the effect of other colored substances present in the sample and the influence of macromolecular species such as proteins. The change of the color of the indicator is also the basis of optodes sensitive to pH. The indicator is immobilized in a polymeric film at the sensor end that is in contact with the sample. The dissociation characteristic of the indicator in those conditions may be significantly different from that known for dilute aqueous media. Besides, the optode senses the pH value at the solution–sensor interface, which may differ from that in the bulk sample solution. On the other hand, absorbance measurements do not need a reference sensor, contrary to the electrochemical measurements.

The most common electrochemical procedures of pH evaluation are based on potentiometric measurements. The theoretically best described is the hydrogen electrode, consisting of platinum metal in contact with hydrogen gas but its use, however, is restricted to specialized laboratories. The various types of glass electrodes offer a wide pH range of dynamic response, which only at the ends of the pH scale may be subject to acid or alkaline errors. The glass electrodes are of various constructions, both macro and micro down to the micrometer scale. Most convenient for routine measurements are glass electrodes mounted in one body with the reference electrode. The uncertainty of electrochemical pH measurements is usually below 0.02 pH unit but may depend on the construction details of the cell. For special applications the glass electrode can be replaced by some metal/metal oxide electrodes, of which the antimony electrode is most commonly used. However, such sensing electrodes are more sensitive to the presence of redox species in the sample solution. Sensors that recently became more popular are those with polymeric membranes, containing a pH sensitive carrier such as tridodecylamine. Their use in nonaqueous media is often restricted, and their lifetime is often shorter than that of glass electrodes.

**See also:** **Buffer Solutions.** **Extraction:** Solvent Extraction Principles. **Indicators:** Acid–Base. **Ion-Selective Electrodes:** Glass. **Sensors:** Overview. **Quality Assurance:** Internal Standards. **Titrimetry:** Potentiometric.

## Further Reading

- Bates R (1973) *Determination of pH. Theory and Practice*. New York: Wiley.
- Buck RP, Rondinini S, Covington AK, *et al.* (2001) *The Measurement of pH – Definition, Standards, and Procedures*. Report of the Working Party on pH, IUPAC Provisional Recommendations, IUPAC.
- Hulanicki A (1987) *Reactions of Acids and Bases in Analytical Chemistry*. Chichester: Ellis Horwood.
- Inczyedy J (1973) *Analytical Applications of Complex Equilibria*. Chichester: Ellis Horwood.
- IUPAC (1998) In: Inczyedy J, Lengyel T, and Ure AM (eds.) *Compendium of Analytical Nomenclature*. Oxford: Blackwell Science.
- Kragten J (1978) *Atlas of Metal–Ligand Equilibria in Aqueous Solutions*. Chichester: Ellis Horwood.
- Minczewski J, Chwastowska J, and Dybczyński R (1982) *Separation and Preconcentration Methods in Inorganic Trace Analysis*. Chichester: Ellis Horwood.
- Pourbaix MJN (1949) *Thermodynamics of Dilute Aqueous Solutions*. London: Edward Arnold.
- Ringbom A (1963) *Complexation in Analytical Chemistry*. New York: Wiley.

# PHARMACEUTICAL ANALYSIS

Contents

**Overview**

**Drug Purity Determination**

**Stability Testing**

**Dissolution Testing**

**Sample Preparation**

**Plant Extracts**

## Overview

**A-M Tivert and J Vessman**, AstraZeneca R&D Mölndal, Mölndal, Sweden

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Pharmaceutical analysis is traditionally defined as analytical chemistry dealing with drugs both as bulk

drug substances and as pharmaceutical products (formulations). However, in academia, as well as in the pharmaceutical industry, other branches of analytical chemistry are also involved, viz. bioanalytical chemistry, drug metabolism studies, and analytical biotechnology. The development of drugs in the pharmaceutical industry is a long-term process, often taking more than a decade from the start of the research project to appearance of a drug on the market. That process involves several decision points, such as the choice of the candidate drug after the discovery phase, the application to the authorities before

See also: **Personal Monitoring:** Active.

## Further Reading

- Berlin A, Brown RH, and Saunders KJ (eds.) (1987) Diffusive sampling: An alternative approach to workplace air monitoring. *Proceedings of the International Symposium*, Luxembourg, 1986. London: Royal Society of Chemistry (includes the NIOSH validation protocol).
- Brown RH, Harvey RP, Purnell CJ, and Saunders KJ (1984) A diffusive sampler evaluation protocol. *American Industrial Hygiene Association Journal* 45: 67–75.
- Guild LV, Myrmet KH, Myers G, and Dietrich DF (1992) Bi-level passive monitor validation – a reliable way of assuring sampling accuracy for a larger number of related chemical hazards. *Applied Occupational and Environmental Hygiene* 7: 310–317.
- Harper M and Purnell CJ (1987) Diffusive sampling – a review. *American Industrial Hygiene Association Journal* 48: 214–218.
- Harper M (2000) Sorbent trapping of volatile organic compounds from air. *Journal of Chromatography A* 885: 129–151.
- Harper M (2004) Assessing exposure to workplace chemicals: the role of exposure monitoring. *Journal of Environmental Monitoring* 6: 404–412.
- Palmes ED and Gunnison AF (1973) Personal monitoring device for gaseous contaminants. *American Industrial Hygiene Association Journal* 34: 78–81.
- Rose VE and Perkins JL (1982) Passive dosimetry – state of the art review. *American Industrial Hygiene Association Journal* 43: 605–621.

# PERVAPORATION

See **MEMBRANE TECHNIQUES:** Pervaporation

# PESTICIDES

**M N Vasilescu**, Institute of Public Health, Bucharest, Romania

**A V Medvedovici**, University of Bucharest, Bucharest, Romania

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

An exhaustive overview on pesticides should include aspects referring to their synthesis and formulation (and related quality control features), mechanisms of action, degradation and/or metabolism patterns, and their impact on environment and living organisms, including on human health. Analytical chemistry is directly involved in all these domains and takes into account at least three major difficulties: extreme diversity of the target compounds, increased sample matrix complexity, and imposed levels of detectability, which are continuously decreasing. The article starts by defining pesticides and related topics together with their classification criteria and class exemplification. Thus, attention is focused on the

analytical contributions to the topic. The sample preparation approach reviews analytical techniques involved in the isolation, concentration, and derivatization of pesticides and pesticide residues from real samples of different origin (environmental, food, clinical, forensic, etc.). Recent developments such as focused microwaves Soxhlet extraction, solid-phase extraction and microextraction on immunosorbents, shielded materials and molecular imprints, stir-bar sorptive extraction, supported liquid membrane extraction, and supercritical fluid and pressurized fluid extractions are discussed. The main achievements related to pesticide separation by means of thin layer, gas, liquid, supercritical fluid, micellar electrokinetic, and capillary electro-driven chromatographic methods are also discussed. Special emphasis is given to mass spectrometric detection for its inherent selectivity and high sensitivity. In some circumstances, other analytical methods such as spectrometric, immunochemical, and electrometric (including bio- and immunosensors) methods can be successfully used for pesticide analysis. Due to the complexity of the subject, more often only recent advances have been detailed at some length.



## Classification

Pesticides are classified according to the following criteria:

1. the target pest they are intended to destroy (see Table 1);
2. their mode of action (see Table 2);
3. their nature (see Table 3);
4. type of formulation in which a pesticide is included (see Table 4); and
5. their impact on health (see Table 5).

In order to illustrate the large variety of compounds used as pesticides, the main chemical classes belonging to fungicides and insecticides are listed in Table 6.

## Sample Preparation

As shown in Figure 3, sample preparation methods applied to pesticide analysis are widely different. The

sophistication of the procedure varies depending on the type of sample to be analysed, its size, and corresponding concentration levels of the target compounds. More often, two or three successive sample preparation techniques need to be applied to a sample in order to achieve analyte concentration and to eliminate the matrix interferences efficiently.

For practical reasons, classical procedures are continuously upgraded to get improvements in terms of reliability and performance. Such an approach relates, for example, to the conventional Soxhlet extraction assisted by focused microwaves (FMSE). The main factors affecting extraction efficiency are microwave power, irradiation time, volume of the extractant, and number of cycles. The result, in terms of extracted mass and repeatability is similar to the classical method. In addition, there is a significant reduction in the total time taken for the process and far less amount of extractant is consumed. Organochlorine pesticides (OCPs) and related residues have been successfully isolated from sunflower seeds by means of FMSE. Comparison with the ISO 659-1988

**Table 1** Classification of pesticides according to the target pest they are intended to destroy

No.	Pesticide class	Synonyms	Target
1	Acaricides	Miticides	Mites feeding on plants and/or animals
2	Algicides	—	Algae from aquatic media (lakes, canals, swimming pools, water tanks)
3	Antifouling agents	—	Living organisms, attaching to underwater exposed surfaces (e.g., boat bottoms)
4	Avicides	—	Birds
5	Biocides	Bactericides	Microorganisms
6	Fungicides	—	Fungi (including blights, mildews, and rusts)
7	Herbicides	—	Weeds
8	Insecticides	—	Insects or other arthropods
9	Molluscicides	—	Snails and slugs
10	Nematicides	—	Nematodes (microscopic worm-like organisms feeding on plant roots)
11	Ovicides	—	Eggs of insects and mites
12	Rodenticides	—	Mice and other rodents
13	Virucides	—	Viruses

**Table 2** Classification of pesticides according to the mode of action

No.	Pesticide class	Function
1	Antifeedant	Inhibits feeding due to inherent toxicity or taste
2	Attractant	Attracts pests to a trap
3	Chemosterilant	Generates sterility of an organism on ingestion
4	Defoliant	Determines dropping of leaves or other foliage from plants
5	Desiccant	Promotes drying of living tissues
6	Fumigant	Releases gas or vapors intended to destroy pests
7	Growth regulator	Alters the expected growth, flowering, or reproduction rate of plants; Disrupts molting, maturity from pupal stage to adult or other life processes of insects
8	Mating disrupter	Acts on mating behavior of insects (e.g., pheromones)
9	Plant incorporated protectant (PIP)	Exogenically induced genetic material determining self-production of substances destroying pests
10	Repellent	Repels pests (avian, insect, mammal)
11	Safener	Increases tolerance of crops to the toxic effects of herbicidal compounds
12	Synergist	Amplifies the specific action of a pesticide

**Table 3** Classification of pesticides according to their nature

No.	Pesticide class	Characterization
1	Inorganics	Synthetic or naturally occurring inorganic compounds exhibiting pesticidal action
2	Organics	Synthetic organic compounds exhibiting pesticidal action
3	Biochemicals	Naturally occurring substances controlling pests by means of nontoxic mechanisms
4	Microbials	Microorganisms (bacteria, fungi, viruses, protozoans, nematodes) affecting pest metabolism

**Table 4** Classification of pesticides according to the type of formulation

No.	Pesticide class	Type of formulation
1	Granulars	Inert carrier impregnated with the active agent or active mixture
2	Dusts	Undiluted active agents representing solid materials with low granulometry
3	Baits	Target-oriented active agents (generally acting after ingestion) with practical no impact on environment
4	Coatings	Active agent deposition in a film covering a surface (e.g., leaves, seeds)
5	Aerosols	Air-suspended solid microparticles or droplets of active agents as such or in solution
6	Slow release formulations	Microencapsulated active agents, paint-in lacquers or resin strips for indoor use

**Table 5** Classification of pesticides according to their impact on health*Classification of pesticides by hazard**(according to the World Health Organization – WHO)*

No.	Class	Hazard	LD <sub>50</sub> (mg per kg) in rats			
			Oral		Dermal	
			Solids	Liquids	Solids	Liquids
1	Ia	Extreme	≤5	≤20	≤10	≤40
2	Ib	High	5–50	20–200	10–100	40–400
3	II	Moderate	50–500	200–2000	100–1000	400–4000
4	III	Slight	> 500	> 2000	> 1000	> 4000

*Classification of pesticides according to their carcinogenic risk on humans**(according to the International Agency for Research on Cancer – IARC)*

No.	Class	Carcinogenicity	Evidence on humans	Evidence on animals	Other supporting evidence
5	1	Carcinogenic	Sufficient Insufficient	Sufficient	Mechanism of carcinogenicity possible in humans
6	2a	Probably carcinogenic	Limited Inadequate	Sufficient Sufficient	Mechanism of carcinogenicity possible in humans
7	2b	Possibly carcinogenic	Limited Inadequate Inadequate	Less than sufficient Sufficient Limited	Other evidence
8	3	Not classified as carcinogenic	Inadequate Inadequate Inadequate	Inadequate Limited Sufficient	Mechanism of carcinogenicity not possible in humans
9	4	Probably not carcinogenic	Evidence of the lack of carcinogenicity Inadequate	Evidence of the lack of carcinogenicity Evidence of the lack of carcinogenicity	Other evidence



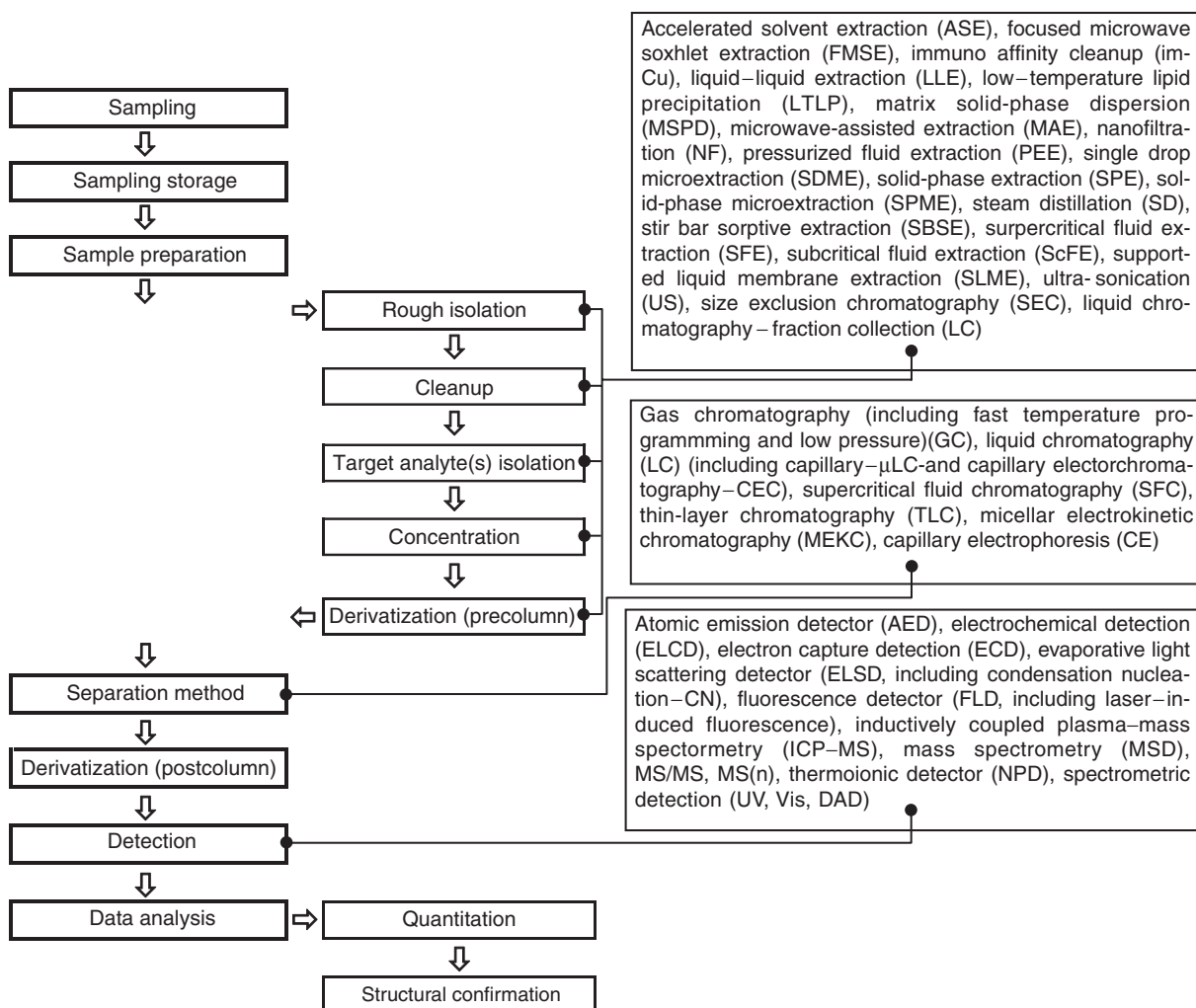
**Table 6** Chemical classes included in the fungicides/insecticides groups

No.	Fungicides		Insecticides	
	Main class	Related classes	Main class	Related classes
1	Acylaminoacids	Acylalanine	Antibiotics	Milbemycin
2	Amides	Benzamides, furamides, valinamides	Carbamates	Oximecarbamates, Phenylmethyl carbamates, Dimethyl carbamates
3	Anilides	Benzanilides, furanilides, sulfonanilides	Cyclodienes	—
4	Antibiotics	—	Dinitrophenols	—
5	Aromatics	Bridged biphenyls	Formamides	—
6	Carbamates	Dithiocarbamates, cyclic dithiocarbamates, thiocarbamates, benzimidazolyl carbamates	Macrocyclic lactones	—
7	Carbanilates	—	Nereistoxines	—
8	Conazoles	—	Nicotinoids	—
9	Dicarboximides	Dichlorophenyl dicarboximides	Nitroguanidines	—
10	Dinitrophenols	—	Organochlorine	—
11	Imidazoles	Benzimidazoles	Organophosphate	—
12	Organometallics	Mercury, tin	Organophosphorus	—
13	Organophosphorus	—	Organothiophosphates	Aliphatic, heterocyclic isoindole, isoxazole, oxime, phenyl, pyrazolopyrimidine, quinoxaline, thiadiazole, triazole derivatives
14	Oxathiin	—	Oxazidine	—
15	Oxazoles	—	Phosphonates	—
16	Phenylsulfamides	—	Phosphonothioates	—
17	Pyridine	Quinolines	Phosphoroamidothioates	—
18	Pyrimidine	—	Phosphorodiamides	—
19	Pyrroles	—	Phthalimides	—
20	Quinones	—	Pyrethroids	Esters, ethers
21	Quinoxalines	—	Pyridylmethylamines	—
22	Strobilurin	—	Pyrimidineamines	—
23	Thiazoles	Benzthiazoles	Pyrroles	—
24	Thiophenes	—	Tetronic acids	—
25	Triazines	—	Ureas	—
26	Triazoles	—	Inorganic arsenic salts	—
27	Ureas	—	Inorganic fluorine salts	—
28	Inorganic copper salts	—		
29	Inorganic mercury salts	—		

reference extraction method confirmed that FMSE is a more efficient process.

In the field of solid-phase extraction (SPE) methods, research is oriented toward the identification and synthesis of new sorbents with increased selectivity. It has already been shown that polar adsorbents such as Oasis HLB (vinylpyrrolidone-divinylbenzene copolymer) can isolate organophosphorous pesticides from blood or serum, or triazines, acetamides and phenoxy acids from surface and waste water. Recently, a new polymeric material, polyaniline, was introduced for the extraction of chlorophenolic pesticides in water samples. Immunosorbents produced by covalent immobilization of the antibody generated against the target analyte on

inert materials (silica gel or activated agarose gel) were successfully used for the isolation of analytes from complex matrices. Thifluzamide in peanut samples and imazalil in citrus fruits have been rapidly isolated using immunosorbents. Target-tailored materials have also been developed by molecular imprinting technologies. 2,4,5-Trichlorophenoxy acetic acid has been used as a template in a system of 4-vinylpyridine as monomer, ethylenedimethacrylate as reticulation agent, and methanol/water mixture as porogen. A molecularly imprinted polymer for the isolation of pirimicarb is based on methacrylic acid as functional monomer and ethyleneglycoldimethacrylate as cross-linker. Some other materials have especially been developed for specific



**Figure 3** General analytical process for pesticide-containing samples.

matrix elimination. Restricted access or shielded materials selectively eliminate macromolecular compounds from pesticide-containing samples (e.g. soil extracts, serum, plasma). Basically, the hydrophilic character of the external particle's surface and the lipophilic character of the inner pore surfaces in these materials are responsible for their intrinsic capabilities of discrimination, combining size exclusion and polarity based interaction mechanisms.

The quest for new suitable materials represents a similar trend in the development of the solid-phase microextraction (SPME) techniques. Polypyrrole and poly *N*-phenylpyrrole films exhibit better selectivity toward polar, aromatic, and basic pesticides, due to their inherent multifunctional properties. Their high surface areas (also revealed by scanning electron microscopy) are responsible for the high extraction efficiencies shown by these materials. Mixtures of divinylbenzene, carboxen, and polydimethylsiloxane-based

fibers have been used for the isolation of cyprodinil and fludioxonil in white wines, while carbowax-templated resins exhibit selectivity towards chlorophenols in environmental samples.

The stir bar sorptive extraction (SBSE) technique, which serves as a more robust alternative to SPME, has been applied for the determination of dicarboximide fungicides in wines.

Microporous membrane liquid-liquid extraction of lipophilic pesticides in biological fluids combines the size exclusion properties of the membrane, reducing lipid coextraction, with the stirring capabilities of the sample phase together with the permanent pumping of the organic phase.

Hollow fiber membranes are used for preconcentration of nitrophenolic pesticides in water samples. The mechanism is based upon the pH difference inside and outside the hollow fiber, while the organic solvent is immobilized within the pores of the

membrane. Such liquid–liquid–liquid microextraction device is suitable for interfacing with micro-column liquid chromatography.

Supported liquid membrane extraction (SLME) is emerging as a fast and efficient sample preparation technique. Aromatic aminophosphonate isolation from water samples based on SLME allowed the identification and study of the operational parameters (pH and ionic strength of the aqueous phase, composition of the membrane phase, and concentration of analytes) as well as the structure–extraction efficiency relationship.

Given the increasing complexity of samples, it becomes imperative that the preparation techniques are highly reliable. This necessitates minimal effort and duration combined with enhanced accuracy and precision. Supercritical fluid extraction (SFE), pressurized fluid extraction (PFE), and assisted solvent extraction (ASE) procedures have been developed as reliable alternatives to traditional methods of sample preparation.

Combining the properties of supercritical fluids and/or extraction solvents with pressure/temperature/density effects and automatically controlled static/dynamic setup results in an increased flexibility and efficiency of the previously cited methods. However, two practices related to SFE/PFE/ASE processes should be mentioned. The first relates to fixation effects in the extraction thimble. Such practice was discovered during development of SFE procedures for pesticide determination in fruit juices (diatomaceous earth was ‘solidifying’ liquid samples). Addition of silica onto solid vegetal materials (such as tobacco leaves) proved to induce a major matrix removal effect during analysis of OCPs. All pesticides enlisted in the 608 EPA method were 100% recovered by means of such a practice at a concentration level of  $10 \text{ pg l}^{-1}$ . The other practice relates to *in situ* derivatization of target analytes during SFE/PFE/ASE procedures. Pesticides containing carboxyl or phenol groups were readily derivatized simultaneously using PFE by using acetic anhydride for acetylation, *N,O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA) for silylation and borontrifluoride/methanol, phenyltrimethylammonium hydroxide, and trimethylsulfonium hydroxide for methylation. Subcritical extraction conditions are sometimes preferred (especially when using water as the extraction agent for pesticides from soil samples). Under optimized conditions, most of the analytes are quantitatively extracted within 90 min. *In situ* degradation of the target compounds in order to generate class markers has also been practised for nonpersistent pesticide analysis in biological samples. 3,4-Dichloroaniline and 3,5-dichloroaniline

as markers for the exposure to fungicides such as vinclozolin, procymidone, iprodione, chlozolate and herbicides such as diuron, linuron, neburon, and propanil, respectively, are generated by the *in situ* hydrolysis simultaneously carried out with steam distillation (SD) removal, followed by liquid–liquid extraction, derivatization with heptafluorobutyric anhydride, and gas chromatography/electron-capture detection.

Last but not least, nanofiltration, a technique derived from technologies used for removing hazardous organic micropollutants, should be considered as a promising tool for sample preparation methods.

Some examples of sample preparation procedures used for pesticide isolation and concentration are given in Table 7.

## Gas Chromatography

Gas chromatography should be considered as the principal primary screening method for pesticide analysis in food, environmental, and clinical samples.

Due to the broad range of polarity characterizing the analytes, separations are achieved on stationary phases that possess different polar characters. Consequently, the open tubular columns with silicone polymer bonded phases that are used range from 100% methyl silicone or 95% methyl and 5% phenyl (apolar) to 50% methyl and 50% phenyl, 50% methyl and 50% trifluoromethylpropyl, 86% methyl and 14% phenylcyanopropyl, or 50% methyl and 50% phenylcyanopropyl (medium polar or highly polar). The primary choice with respect to the stationary phase for screening residues of pesticides is a low polarity one, either 100% methyl silicone or 95% methyl and 5% phenyl.

Considering the close relationship between the deactivation degree of the column wall and the deposited layer of the stationary phase, use of films thinner than  $0.1 \text{ }\mu\text{m}$  should be avoided, especially for the separation of labile pesticides. Fused silica open tubular columns of 25–30 m length, 0.20–0.25 mm internal diameter, and 0.15–0.33  $\mu\text{m}$  film thickness should be considered as the first choice.

Low-pressure gas chromatography (LPGC) is also used for the fast analysis of multiple pesticide residues, mainly for food applications. The LPGC/mass spectrometric detection (MSD) technique uses an analytical 0.53 mm internal diameter column and 1  $\mu\text{m}$  thick film coupled to a 0.15 mm inner diameter restriction capillary at the inlet side. While the injection conditions are similar to conventional GC methods, subatmospheric pressures occur through the column because of the MS vacuum source.

**Table 7** Sample preparation techniques used for isolation of some pesticides from various matrices (before separation and detection)

No.	Name	Class	Chemical class	US	LLE	CU+ ImCU	SPE	SPME	MSPD	SBSE	SD	FMSE	PFE+ ASE	ScFE	SFE	SEC	LC	LTLP
1	Aldrin	I	Cyclodiene		X		X	X						X		X		
2	Bromophos (ethyl)	A, I	Organothiophosphate (phenyl)		X		X	X								X		
3	Bupirimate	F	Pyrimidine				X					X						
4	Captan	F	Phthalimide				X									X	X	
5	Carbaryl	A, I, GR	Carbamate		X		X									X	X	
6	Carbofuran	A, I, N	Carbamate (benzofuran)		X		X							X		X	X	
7	Carbophenothion	A, I	Organothiophosphate (phenyl)				X										X	
8	Chlormequat	GR	Quaternary ammonium		X		X									X		
9	Chlorpyrifos	A, I, N	Organothiophosphate		X		X	X				X		X		X	X	
10	Cyhalothrin	A, I	Pyrethroid	X		X												
11	Deltamethrin	I	Pyrethroid	X	X	X										X	X	
12	DDD ( <i>p,p'</i> )	I	Organochlorine, bridged biphenyl	X	X		X	X				X			X	X	X	
13	DDT ( <i>p,p'</i> )	A, I	Organochlorine, bridged biphenyl	X	X		X	X				X	X	X	X	X	X	
14	Diphacinone	R	Indandione		X		X											
15	Endosulfan	A, I	Organochlorine, cyclodiene	X	X	X	X	X				X			X	X	X	
16	Endrin	AV	Cyclodiene		X		X	X						X		X		
17	Ethion	A, I	Organothiophosphate (aliphatic)		X		X	X								X		
18	Ethirimol	F	Pyrimidine				X					X						
19	Fenthion	AV, I	Organothiophosphate (phenyl)				X	X								X		
20	Fenvalerate	A, I	Pyrethroid ester	X		X	X									X		
21	Heptachlor	I	Cyclodiene		X		X	X						X		X		
22	Imazalil	F	Conazole			X										X		
23	Lindane	A, I, R	Organochlorine	X	X		X	X						X		X		
24	Malathion	A, I	Organothiophosphate (aliphatic)		X		X	X				X				X	X	
25	Parathion	A, I	Organothiophosphate (phenyl)		X		X	X						X		X	X	
26	Permethrin	A, I	Pyrethroid ester	X	X	X	X									X		
27	Prochloraz	F	Conazole		X				X									
28	Procymidon	F	Dicarboximide				X			X	X					X		
29	Propham	H, GR	Carbanilate		X		X											
30	Propoxur	A, I	Carbamate (methylphenyl)		X		X									X		X
31	Tolclofos (methyl)	F	Organophosphorus	X	X				X							X		X
32	Vinclozolin	F	Oxazole, dicarboximide		X					X	X					X		
33	Warfarin	R	Coumarine	X	X		X											

A, acaricide; AV, avicide; I, insecticide; F, fungicide; H, herbicide; GR, growth regulator; N, nematocide; R, rodenticide; US, ultrasonication; LLE, liquid-liquid extraction; CU, cleanup; ImCU, immuno cleanup; SPE, solid-phase extraction; MSPD, matrix solid-phase dispersion; SBSE, stir bar sorptive extraction; SD, steam distillation; FMSE, focused microwave Soxhlet extraction; PFE, pressurized fluid extraction; ASE, assisted solvent extraction; ScFE, subcritical fluid extraction; SFE, supercritical fluid extraction; SEC, size-exclusion chromatography; LC, liquid chromatography (fraction collection); LTLP, low temperature lipid precipitation.

Optimized conditions allow faster separations (three-fold reduction), substantial increased injection volume capacity, reduced degradation of thermally labile compounds, and lower detection limits due to increased sample loadability.

Nitrogen, helium, and hydrogen are the carrier gases commonly used in GC. Hydrogen is not suitable for some selective GC detectors, while nitrogen exhibits less flat Van Deemter curves for gas velocities higher than optimal. Thus, helium remains the best choice.

As injectors for capillary GC separation of pesticides and related residues, cold on-column (OCI), hot splitless injector (HSI), and programmed temperature vaporizer (PTV) are mainly used. Injection of thermally unstable analytes can be controlled via pressure pulses (based on electronic pressure controlled systems – or EPC systems), by precise temperature control during injection (achieved by PTV), and by the presence of suitable chemical additives. As an example, phenylureas in conventional hot splitless injectors lead to extensive and irreproducible formation of isocyanates and amines. The presence of acetic acid, low molecular mass amines, and organic anhydrides during injection result in minimization of the thermal decomposition effect or reproducible conversion to the corresponding isocyanates. Studies carried out on organophosphorous and carbamate pesticides, in parallel, on PTV splitless, PTV solvent split, pulsed splitless, and on-column injectors lead to the following hierarchical tolerance: OCI < pulsed HSI < PTV splitless < PTV solvent split. Some other phenomena are associated with injection systems. Tralomethrin, for example, is quantitatively transformed to deltamethrin in HSIs. The latter, when vaporized in packed liner-fitted PTVs, readily isomerizes, leading to splitting of the chromatographic peak into two equally intense peaks that are baseline separated, corresponding to the *cis* and *trans* isomers. Derivatization may also enhance thermal stability. Phenylurea pesticides are alkylated with iodoethane and sodium hydride to yield thermostable compounds. When using MSD and iodoethane instead of iodomethane, differentiation between the parent compounds and *N*-demethylated metabolites is possible.

The use of a retention gap between the injection port and the capillary column has been widely discussed in the literature, especially for OCI. The influences of the solvent, the gap geometry, the injection volume, and the analyte concentration on peak efficiency and symmetry do not lead to a general accepted recommendation considering the use of retention gaps. Checking for each particular case is more advisable.

Peak tailing processes due to undesired interactions of the analytes with the active sites in the inlet and the chromatographic column occur with a relative high incidence in GC analysis of pesticides. For susceptible analytes, significant peakshape improvements are obtained when matrix components are present because of their specific competitive interactions with the active sites (the so-called matrix-induced chromatographic response enhancement). Some of the reported applications deal with ‘analyte protectants’ for providing a convenient and effective solution for the enhancement of chromatographic response. Most of such compounds spiked to real samples have hydrogen bonding capabilities in order to preferentially interact with the active sites (sugar, sugar derivatives, and gulonolactones are used).

Detection systems used for GC analysis of pesticides belong to the following categories: (1) classical selective and sensitive detectors such as electron-capture detector (ECD – including the pulse modulated setup and the microdesign); nitrogen–phosphorous detector (NPD – including the separate electrical heating of the alkali bed design); flame photometric detector (FPD, including the sealed chimney, the double burner, and the pulsed flame designs); and (2) structural information detectors, mainly the MSD and the atomic emission detector using the microwave-induced plasma (AED-MIP).

Both electron impact (EI) and chemical ionization (CI) are currently used as ionization techniques in MSD, in addition to the positive or negative ion monitoring modes. The use of the selective ion monitoring (SIM) feature improves the sensitivity greatly, although some structural confirmation is lost. In Table 8, the ion fragments frequently monitored in MSD for some pesticides separated by GC are listed.

AED provides molecular confirmation by means of its ability to calculate the molecular formula by means of the responses obtained on the simultaneously monitored specific element channels. It is worth noting that sensitivity is tuned according to the chemical element monitored (e.g., acephate is characterized by a 5 ng detection limit when monitoring the nitrogen channel, 3 ng for phosphorous, 10 ng for oxygen, and 0.3 ng for sulfur).

## Liquid Chromatography

The main reasons for the failure of GC methods in pesticide analysis are thermal lability, low volatility, and high polarity of some target analytes. There are a relatively reduced number of classes of pesticides that are not suitable for GC separation



**Table 8** Major *m/e* signals in the mass spectra of some pesticides separated in gas chromatography with mass spectrometric detection (ionization by electron impact)

No.	Pesticide	Class	Chemical classification	Major <i>m/e</i> signals in the mass spectrum
1	Azinfos ethyl	A, I	Organothiophosphate	77, 132, 160
2	Chlorfenvifos	A, I	Organophosphate	267, 323
3	Chlorpyrifos	A, I, N	Organothiophosphate (pyridine)	197, 242, 270, 298
4	Chlordane	I	Cyclodiene	373
5	Coumafos	A, I	Organothiophosphate (heterocyclic)	226, 362
6	Dichlorvos	A, I	Organophosphate	109, 185
7	Dicrotofos	I	Organophosphate	127
8	DDD ( <i>p,p'</i> )	I	Organochlorine, bridged biphenyl	235
9	DDT ( <i>p,p'</i> )	A, I	Organochlorine, bridged biphenyl	235
10	Endosulfan	A, I	Cyclodiene, Organochlorine	195, 241, 265, 339
11	Endrin	AV, I	Cyclodiene	263, 281
12	Ethion	A, I	Organothiophosphate (aliphatic)	97, 125, 153, 231, 384
13	Ethoprop	I, N	Organothiophosphate (aliphatic)	158, 200, 242
14	Fenchlorfos	I	Organothiophosphate (phenyl)	125, 285
15	Fenitrothion	I	Organothiophosphate (phenyl)	109, 125, 260, 277
16	Heptachlor	I	Cyclodiene, Organochlorine	100, 272
17	Lindane	A, I, R	Organochlorine	109, 181, 219
18	Leptophos	I	Phosphonothioate	171, 377
19	Merphos	D, GR	Phosphonothioate	209, 298
20	Methoprene	I	Ester	73
21	Methoxychlor	I	Organochlorine	227
22	Parathion	A, I	Organothiophosphate (phenyl)	291
23	Parathion (methyl)	I	Organothiophosphate (phenyl)	109, 125, 263
24	Phenothrin	I	Pyrethroid ester	123, 183
25	Resmethrin	I	Pyrethroid ester	123, 143, 171
26	Stirofos	A, I	Organophosphate	109, 329

D, defoliant.

and which therefore require the use of liquid chromatography (LC). Phenylurea, carbamates, dinitrophenols, benzimidazoles, azoles, benzoylureas, some organophosphorous, pyrethroid, and quaternary ammonium derivatives are some examples for such pesticides.

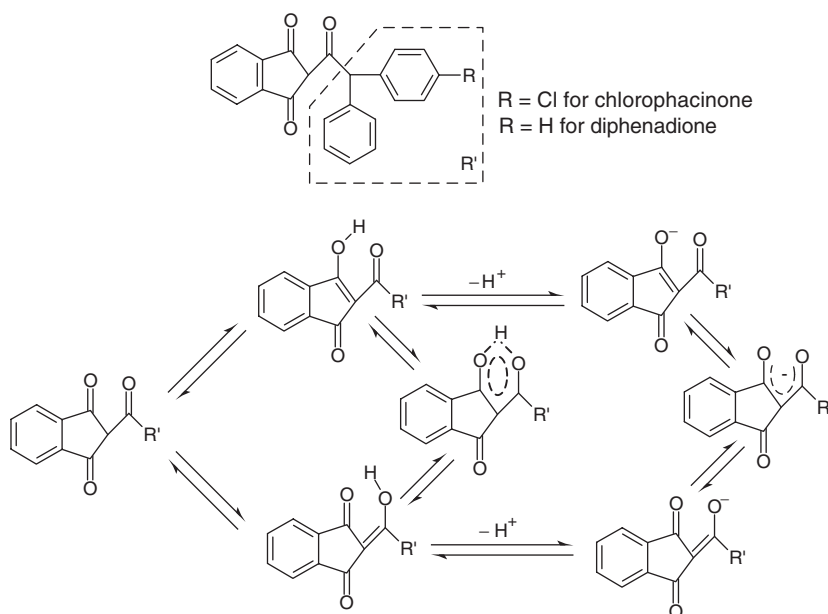
In LC, the separation mechanism favored for the separation of pesticides is undoubtedly the reversed phase (RP) chromatography. However, ion exchange (IE) and ion pair (IP) mechanisms are also used, especially for quaternary ammonium derivatives (quats). In RP applications, polar end-capped octadecyl silica gel stationary phases are especially recommended for separations requiring high content or even 100% aqueous mobile phases (e.g. separation of water unextractable organophosphorous pesticides such as acephate, methamidophos, monocrotophos, omethoate, oxydemeton-methyl, and vamidothion). More often, the addition to mobile phases of concentrated buffers or strong acids is needed in order to stabilize the analytes in a single form. When tautomeric equilibrium is competitive to chromatographic partition, the peakshape and symmetry are strongly affected. Sometimes, the inherent selectivity of LC affords separation of the tautomeric structures, leading to serious peak splitting, mainly

at low concentration of the target compounds. **Figure 4** gives an example showing tautomerism of chlorophacinon and diphacinon rodenticides, whose stabilization requires addition of strong acid to the mobile phase. Normal phase (NP) mechanism is used only for achieving enantioselective separations. Enantiomers of the organophosphorous pesticides crotoxyphos, dialifor, fonofos, fenamiphos, fensulfothion, isofenphos, malathion, methamidophos, profenophos, crufomate, prothiophos, and trichloronate have been separated successfully on polysaccharide-based chiral stationary phases (ChiralPak AD, AS, OD and OJ) under NP conditions.

Micro-LC columns are used for better compatibility with specific sample preparation methods (e.g. on-column focusing) or interfaces for MSD. The use of microcolumns also enables temperature gradient as an additional selectivity/efficiency tuning factor. Applications on triazines use temperature gradients ranging from subambient conditions up to 70°C.

It is worth noting that robust online configurations between SPE or SPME sample preparation techniques and LC separations have been reported, which have lead even to commercially available dedicated devices (e.g. SAMOS, OSP2A).





**Figure 4** Chlorophacinone and diphenadione's tautomerism in aqueous media.

For pesticide separations by LC, different calibration approaches were studied. Among 'external solvent solubilized standard', 'internal standard', 'external matrix-matched standard', and 'echo peak internal standard' approaches, the last two procedures generate better results. The 'echo peak internal standard' is a novel technique based on the possibility of producing a delayed injection of an internal standard having the same identity as the target compound.

Detection systems used for LC analysis of pesticides are: UV spectrometric detection (especially diode array detection – DAD – allowing peak confirmation by means of spectra comparison), fluorescence detection (FLD), electrochemical detection (ED), evaporative light scattering detection (ELSD), and MSD. Their characteristic sensitivities can be considered to vary in the following order: ELSD < UV (DAD) < ED ~ FLD < MSD. Particle condensation nucleation practices are also expected to enhance ELSD sensitivity. Among the detection systems mentioned above, only ELS and MS are universal detectors.

UV or FL detection of the target compounds is sometimes not achievable due to the lack of suitable chromophoric/fluorophoric molecular sites. This could be overcome by derivatization. Two postcolumn derivatization modes are noticeable for pesticide detection. One relates to the fluorescent detection of carbamates based on their reaction with *ortho*-phthalaldehyde (OPA), and the other deals with the photoirradiation of benzoylurea insecticides (namely diflubenzuron, triflumuron, hexaflumuron, lufenuron,

and flufenoxuron) or pyrethroids (namely phenpropathrin, cyfluthrin, deltamethrin, fenvalerate, acrinathrin,  $\tau$ -fluvalinate, and bifenthrin) with production of highly fluorescent photodegradation products.

Detection according to (geno)toxicity has been reported recently. Such system is based on the collection over microtiter plates of 1 min eluate fractions. After evaporation and redissolution in suitable solvents, the (geno)toxicity of each individual fraction is determined before and after enzyme activation, by means of the umu test. (Geno)toxicity is measured against 4-nitroquinoline-*N*-oxide and 2-aminoanthracene standards, exhibiting detection limits of the order of  $0.1 \mu\text{g l}^{-1}$ .

MS is undoubtedly the solution of the near future for LC detection. Improvements made to interfacing devices together with a continuous and sensible diminution of instrumentation costs promote MS as a universal/selective tunable detection system. Atmospheric pressure electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) are the most robust and popular devices for interfacing MS to LC systems. In Table 9, LC–MS data for some pesticides are given. Although ESI and APCI are more often used, other LC–MS interfaces produce reliable results in pesticide applications: thermospray (TSI), particle beam (PBI) and matrix-assisted post-source decay laser desorption/ionization (CID-PSD-MALDI).

Multidimensional techniques (MS/MS or (MS)<sup>n</sup>) enable finding the right ratio between selectivity and sensitivity of the detection process.

**Table 9** LC–MS data for some pesticides

No.	Compound	Class	Chemical classification	ESI/MS				APCI/MS		
				Mode	m/e signal	LOD (pg l <sup>-1</sup> )	Detection	m/e signal	LOD (ng)	
									TIC	SIM
1	Alachlor	H	Chloracetanilide	–	–	–	–	270, 243, 226	8.7	0.43
2	Aldicarb	A, I, N	Oxime carbamate	(+)	191	1.0	MS/MS	208, 222, 191, 148	0.2	–
3	Azinphos-methyl	A, I	Organothiophosphate (benzotriazine)	(+)	318	2.0	MS/MS	–	–	–
4	Bendiocarb	I	Carbamate	(+)	224	7.5	MS/MS	–	–	–
5	Benfuresate	H	Benzofuranyllalkylsulfonates	(+)	342	28.0	MS	–	–	–
6	Buminafos	GR	Phosphoroamidothioate	–	–	–	–	212, 195	13.1	–
7	Buprofezin	I	Thiadiazin	(+)	306	0.1	MS	–	–	–
8	Carbaryl	A, I, GR	Carbamates	(+)	202	9.0	MS/MS	219, 202, 145	1.0	0.08
9	Carbofuran	A, I, N	Carbamate (benzofuranyl)	(+)	222	12.5	MS/MS	222, 239, 182	0.3	0.06
10	Chlorfenvinphos	A, I	Organophosphonate	–	–	–	–	377, 360	23.9	–
11	Chlorflurazuron	I	Benzamide	(–)	538	16.0	MS	–	–	–
12	Chlorpyrifos	A, I, N	Organothiophosphate (pyridine)	(+)	351	9.0	MS/MS	–	–	–
13	Diazinon	A, I,	Organothiophosphate (pyrimidine)	(+)	305	2.0	MS/MS	–	–	–
14	Diflubenzuron	I	Benzamide	(–)	309	148, 0	MS	–	–	–
15	Dimethoate	A, I, N	Organothiophosphate (aliphatic)	(+)	230	2.5	MS/MS	230, 247	14.9	–
16	Diquat	H	Quaternary ammonium	(+)	157, 183	0.1	MS	–	–	–
17	Dymrom	H	Phenyl urea	(+)	269	0.1	MS	–	–	–
18	Fenpyroximate	A	Pyrazole	(+)	422	0.1	MS	–	–	–
19	Flufenoxuron	A, I	Benzamide	(–)	487	0.4	MS	–	–	–
20	Furametpyr	F	Pyrazole	(+)	334	12, 0	MS	–	–	–
21	Hexaflumuron	I	Benzamide	(–)	459	0.1	MS	–	–	–
22	Imazalil	F	Conazole	(+)	297	0.1	MS	–	–	–
23	Inabenfide	GR	Carboxydimide	(+)	339	3.6	MS	–	–	–
24	Iprodione	F	Imidazole	(+)	297, 299	0.2	MS	–	–	–
25	Malathion	A, I	Organothiophosphate (aliphatic)	(+)	331	19.0	MS/MS	–	–	–
26	Merphos	GR	Phosphorothioate	–	–	–	–	315, 299	20.6	–
27	Methabenzthiazuron	H	Urea	(+)	222	32.0	MS	222, 165	2.0	0.16
28	Myclobutanil	F	Conazole	(+)	289	24.0	MS	–	–	–
29	Oxamyl	A, I, N	Carbamate (oxime)	–	–	–	–	237, 163, 180	13.2	0.34
30	Paraquat	H	Quaternary ammonium	(+)	171, 185	0.2	MS	–	–	–
31	Parathion	A, I	Organophosphorothioate (phenyl)	(+)	292	5.0	MS/MS	–	–	–
32	Pencycuron	F	Urea	(+)	329	0.1	MS	–	–	–
33	Phorate	A, I, N	Organothiophosphate (aliphatic)	(+)	261	0.5	MS/MS	–	–	–
34	Phosalone	A, I	Organothiophosphate (heterocyclic)	–	–	–	–	385, 368	24.3	–
35	Phosmet	A, I	Organothiophosphate (phtalimide)	–	–	–	–	335, 318, 237, 209	29.7	–
36	Pirimicarb	I	Carbamate (pyrimidyl)	–	–	–	–	239	1.7	0.04
37	Promecarb	I	Carbamate (phenyl)	–	–	–	–	225, 208	0.9	–
38	Prosulfocarb	H	Thiocarbamate	–	–	–	–	252	1.4	0.10
39	Pyrazoxyfen	H	Pyrazole	(+)	403	0.1	MS	–	–	–
40	Tebufozozide	I	Hydrazide	(+)	353	112.0	MS	–	–	–
41	Temephos	I	Organothiophosphate	(+)	467	28.0	MS/MS	–	–	–
42	Terbacil	H	Uracil	–	–	–	–	178, 161	4.9	–
43	Terbufos	I, N	Organothiophosphate (aliphatic)	(+)	289	1.0	MS/MS	–	–	–
44	Triallate	H	Thiocarbamate	(+)	306	23.0	MS/MS	162, 304	1.5	0.09
45	Triflumizole	F	Conazole	(+)	346	24.0	MS	–	–	–

## Supercritical Fluid Chromatography

Packed column supercritical fluid chromatography (P-SFC) offers an interesting alternative for pesticide separations based on the normal phase partition mechanism, even when using apolar modified stationary phases. Reproducible retention, high selectivity, and good peak symmetry are obtained on applying both density/pressure and modifier gradients. Online coupling of SFC with SPE or SPME (desorption is made by the supercritical mobile phase) results in high extraction yields and sample throughput. Interfacing with MSD (both ESI and APCI) does not require any modification of the commercialized interface designs but only an additional solvent flow, generally introduced before the automated pressure relief valve (nozzle) of a downstream configuration. A similar setup based on the introduction of an additional suitable solvent flow through the nozzle leads to a versatile SFC semi-preparative or even preparative fraction collection/isolation alternative, allowing fast (automated) off-line bidimensional separation techniques. Extremely selective, reproducible, and fast SFC applications relate to chiral separation of pesticide racemics on chemically modified polysaccharides or to antibiotic modified stationary phases.

## Thin-Layer Chromatography

Thin-layer chromatography (TLC) is more often associated with the screening of pesticides, especially from water samples. The automated multiple development (AMD) feature significantly increases the separation capacity and the versatility of the method. The AMD approach uses more than 25 different development steps (1–3 mm migration distance each) with intermediate drying steps, the mobile phase composition or nature changing continuously from step to step. Both mobile phase change and drying are done in an automated way, AMD-TLC equipments are now commercially available. In fact, the entire procedure represents the application of three or even more different step composition gradients during a single run over TLC silica plates having thickness higher than 100  $\mu\text{m}$ . Because of intermediary drying steps, each new development generates focusing of the spots, finally leading to extremely sharp lines. Sample throughput of the method should be considered as satisfactory since 12–24 sample applications are feasible on a single TLC plate. By associating densitometry detection, quantitative determinations are possible, and currently research efforts are underway to produce UV-Vis reflectance spectra of the separated spots. Structural confirmation

and sensitive quantitation are made possible by scratching spots from plates followed by analyte extraction from the solid material in appropriate solvents and subsequent introduction in a mass spectrometer. Standardization of AMD-TLC as a DIN method for pesticide determination in ground and drinking water dates back to 1993.

## Micellar Electrokinetic Chromatography and Capillary Electrochromatography

Micellar electrokinetic chromatography (MEKC) and capillary electrochromatography (CEC) are powerful techniques for achieving efficient pesticide and pesticide residue separations, although accepted standardized methods are still missing. The *N*-methylcarbamates, carbendazin, imazalil, methylthiophanate, prochloraz, procimidon, thiabendazole, triadimefon, metribuzin, lenacil, ethofumesate, atrazine, terbuthrin, isoproturon, chlorotoluron, linuron, desethyl atrazine, 2-hydroxyatrazine, desethyl 2-hydroxyatrazine, and 3-chloro-4-methylphenyl urea are only some of the pesticides that have been reported to be separated by means of MEKC and CEC. Online coupling of MEKC with ESI-MS was recently found as achievable using partial filling (PF) or reverse migrating micelles (RMM) techniques. Quantitative migration-toxicity relationship for phenoxy acid herbicides has been obtained with MEKC on using micellar Brij 35 based migration media. Online stacking procedure carried out on injection acts as a preconcentration step and together with SPE for sample preparation shifts the detection limits to the low  $\mu\text{g l}^{-1}$  range, even on use of the classical on-column diode array detection.

## Capillary Electrophoresis

Capillary electrophoresis (CE) represents an important tool with several advantages for determining pesticides and pesticide residues in environmental samples due to its intrinsic simplicity and high efficiency. Target analytes should be ionic or easily ionized. Some limitations arise from a lack of selective detectors, although CE-MS interfacing has become more affordable.

## Immunochemical Assays

Immunoassays (IAs) are based upon the selective interaction between a specific antibody (Ab) and a hapten (H or antigen – Ag). Pesticides are not ordinarily antigenic; consequently they have to be

**Table 10** Immunochemical assays for some pesticides

No.	Pesticide	Class	Chemical class	Test format	Ag Type	Detection limit ( $\mu\text{g l}^{-1}$ )	Sample type
1	Aldrin	I	Cyclodiene	RIA	p	0.7 (ng) <sup>a</sup>	Biological fluids
2	Atrazine	H	Chlorotriazine	CLIA	p	0.025	Environmental
3	Benomyl	A, F, N	Carbamate, benzimidazole	PFIA	p	0.100	Environmental
				RIA	p	1.250	Food, crops
4	Bioallethrin	I	Pyrethroid ester	RIA	p	0.03	–
5	Captan	F	Phthalimide	EIA	p	1.000	Food
6	Carbofuran	A, I, N	Carbamate	EIA	p	0.060	Soil
7	Chlorothalonil	F	Aromatic	EIA	p	0.070	Water
8	Chlorpyrifos	A, I, N	Organothiophosphate, pyridine	EIA	p	0.050	Soil, water
9	2,4-D	H, GR	Phenoxy acetic acid	RIA	p	5.000	Biological fluids
				PFIA	m	0.600	–
10	Dichlorprop	H, GR	Phenoxy propionic acid	PFIA	p	10.000	–
11	Dieldrin	I	Cyclodiene	RIA	p	0.08 (ng) <sup>a</sup>	–
12	Heptachlor	I	Cyclodiene	EIA	m	10.000	Food
13	Hexazinone	H	Triazinone	EIA	p	0.220	Soil
14	Imazamethabenz	H	Imidazoline	EIA	p	0.500	Cereal grains
15	Paraquat	H	Quaternary ammonium	RIA	m	0.500	Biological fluids
				FIA	p	20.000	Biological fluids
				RIA	p	50.000	Biological fluids
				EIA	p	0.020	Fruits and vegetables
16	Parathion	A, I	Organothiophosphate, phenyl	RIA	p	100.000	Agricultural
				EIA	p	0.100	Ground and surface water
17	Permethrin	A, I	Pyrethroid ester	EIA	m	15.000	Meat
				EIA	m	1.500	Grains
18	Picloram	H	Pyridine	RIA	p	50.000	Water, urine
19	Pyrimiphos methyl	A, I	Organothiophosphate, pyrimidine	EIA	p	30.000	Wheat grains, flour
20	Simazine	Al, H	Chlorotriazine	PFIA	p	3.000	–
21	2,4,5-T	H, GR	Phenoxy acetic acid	RIA	p	1.000	Surface water
22	Thiabendazole	F	Thiazole, benzimidazole	EIA	m	0.500	Liver
23	Triadimefon	F	Conazole	EIA	p	2.000	Food

<sup>a</sup> Referred to as absolute quantity.

A, acaricide; Al, algicide; F, fungicide; GR, growth regulator; H, herbicide; I, insecticide; N, nematocide; m, monoclonal antibody; p, polyclonal antibody.

conjugated to a carrier molecule (usually a protein) in order to induce an immune response. The result of the binding reaction between the Ab and H can be measured by means of enzymatic methods (EIAs), radiometric methods (RIAs), fluorescence methods (FIAs, including the polarized PFIA approach), or chemiluminescence methods (CLIAs). Some of the immunoassays designed for pesticides are given in Table 10, together with the corresponding detection limits.

## Spectrometric Methods

Spectrometric methods are more often used for pesticide analysis in formulations. However, some spectrometry-based screening methods for environmental and biological samples have also been reported. Powerful structural elucidation spectrometric techniques (such as nuclear magnetic resonance spectrometry – NMR) have been also extensively used for the investigation of degradation

patterns of parent pesticides, mainly in biological systems. Derivatization is often needed for introduction of chromophores, fluorophores, or luminophores in the target molecules. A few recent examples of pesticide analysis by means of spectrometric methods are given in Table 11.

## Bio- and Immunosensors

Bio- and immunosensors are systems based on measurement of the results generated by means of binding to a specific receptor of the target analyte. The basic construction of a biosensor for pesticide analysis is schematically given in Figure 5. The main advantages related to bio- and immunosensors is because of the high sample throughput, any or limited sample preparation procedures, *in situ* measurement capabilities, and selectivity. The main drawbacks are related to inherent matrix interferences (mainly on biosensors), limited shelf-life, difficulties for coupling

**Table 11** Spectrometric techniques used in pesticide analysis – few recent applications

No.	Pesticide (class or compound)	Spectrometric technique	Sample	Details on experiments
1	Isothiocyanates and dithiocarbamates	UV	Biological fluids	Cyclocondensation with benzenedithiol
2	Carbaryl	UV	Formulations	Coupling/complexing with 2-amino-4-nitrophenol and 2,4-dimethoxy aniline; detection in the 510–520 nm interval
3	Carbofuran	UV	Formulations	Hydrolysis followed by derivatization with sodium nitroprusside; detection 530 nm
4	Organochlorinated	Tunable dye laser absorption spectrometry	Water	Detection limits $\sim 10^{-3}$ – $10^{-4}$ $\mu\text{g g}^{-1}$
5	Organochlorinated	Laser induced fluorescence (LIF)	Water	Detection limits $\sim 10^{-1}$ – $10^{-2}$ $\mu\text{g g}^{-1}$
6	Carbaryl	LIF	Water	Detection limit at femtogram level
7	Benzoylurea and fenfenoxuron, lufenuron, hexaflumuron, triflumuron	Induced fluorescence (IF)	River water and formulations	UV irradiation to produce highly fluorescent degradation products
8	Morestan	Solid phase phosphorescence	Natural, tap, sea, river, waste water	Excitation 363 nm; emission 520 nm
9	Dichlorvos	Chemiluminescence (CL)	Vegetables	Chemiluminescent system with luminal, perhydrole, and cyltrimethylammonium bromide
10	Folpet and metalaxyl	FTIR	Formulations	Extraction in chloroform; detection limits $\sim 15$ $\mu\text{g g}^{-1}$
11	Chlorinated aromatic amines	IR	Water	Extraction on hydrophobic polyacrylonitrile-co-butadiene material
12	Ziram	IR (vapor phase)	Formulations	Evolution of $\text{CS}_2$ as degradation product
13	Cyromazine	Surface enhanced Raman scattering spectrometry (SERS)	Formulations	Study on surface adsorption mechanisms
14	Glyphosate and metabolites (AMPA, MAMPA and MPA)	$^{31}\text{P}$ -NMR	Water	–
15	2-Aminobenzothiazoles	$^1\text{H}$ -NMR, $^1\text{H}$ - $^{15}\text{N}$ hetero nuclear shifts	–	Studies on biodegradation
16	Sulfotep and propoxur	Ion mobility spectrometry (IMS)	Liquid matrices	–
17	Organophosphorous	IMS	Liquid matrices	Degradation studies
18	Pesticides	Thermal lens spectrometry	Food	–

receptors to supports, identification of sensitive transducers, and a relative delay in generating the response.

The measuring principle of a biosensor designed for pesticide determination is mainly enzyme inhibition. Cyanobacteria, thylakoid membranes, protoplasts, immobilized enzymes, and labeled enzymes are receptor components. The analyte blocks the enzyme, enzyme systems, or electron transport systems of intact cells.

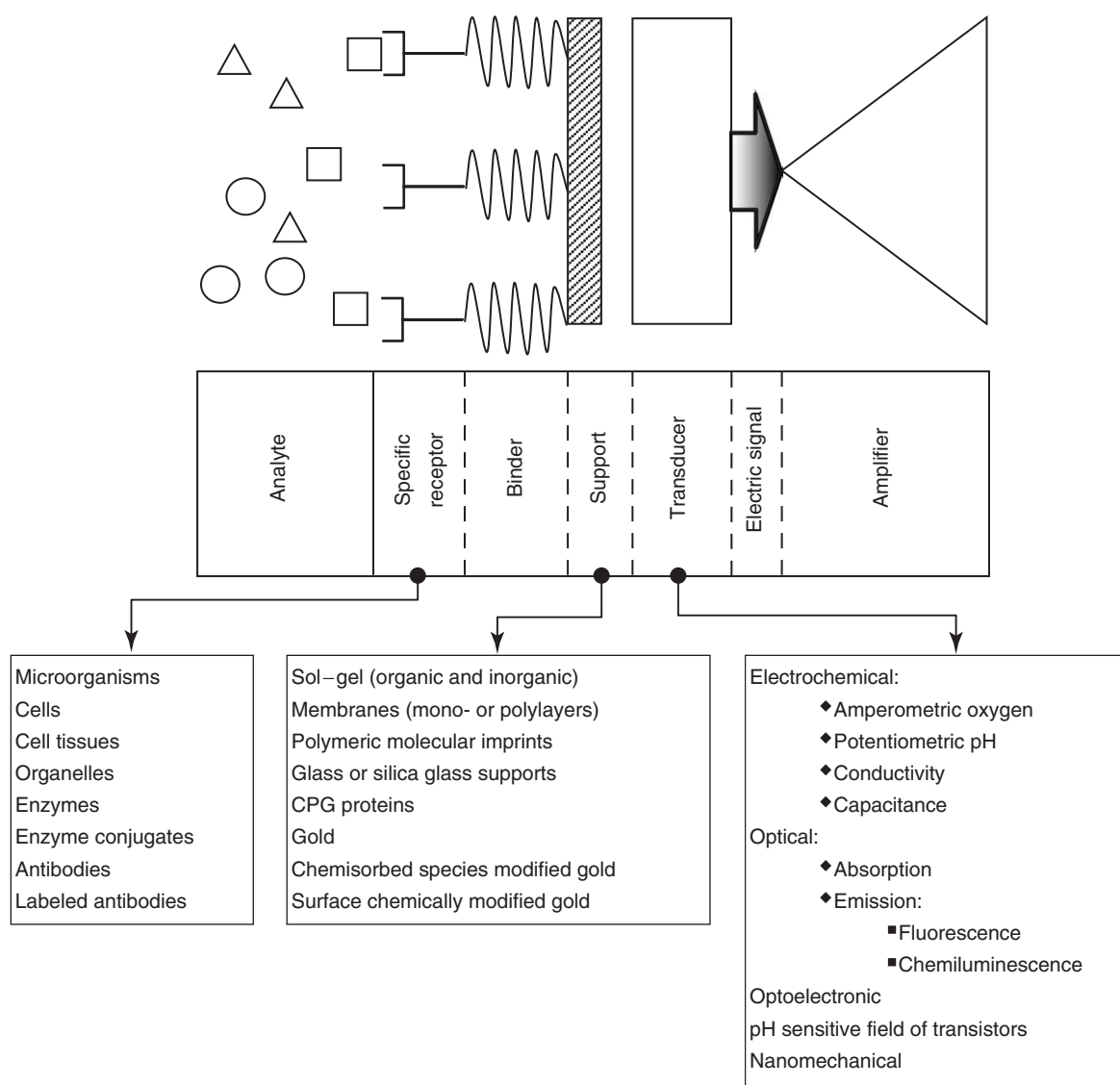
In an immunosensor, the antigen–antibody selective interaction is monitored. In order to improve the sensitivity, the indirect monitoring of the Ag–Ab interaction should be preferred, using competitive tracers or markers. Determination of alachlor, for example, can be achieved with increased sensitivity by using liposome encapsulated markers in a competitive binding reaction instead of enzymes.

The conjugation of a lipid to a pesticide allows its incorporation into a liposome structure, leading to competitive liposome IA-based sensor. The use of piezoelectric crystals as physical sensors is a continuously developing research direction. Other physical sensors are optical, such as surface plasmon resonance (SPR), interferometric, or grating couplers.

Bio- and immuno-sensors are complementary to high-resolution techniques and represent a stimulating border research area between chemical analysis and microelectronics.

## Electroanalytical Techniques

Electroanalytical techniques are often used for the determination of a target pesticide in a given matrix.



**Figure 5** Basic construction of a biosensor designed for pesticide analysis.

In order to obtain the required sensitivity levels, preconcentration techniques are needed. Adsorptive stripping voltammetry (ASV) is, in this respect, a method of choice. Specific accumulation potentials applied for 30–60 s before the measurement have shifted the detection levels to the  $10^{-7}$ – $10^{-9}$  mol l<sup>-1</sup> interval (e.g. buprofezin requires an accumulation potential of  $-0.8$  V over a period of 60 s on the hanging mercury electrode before quantitation by means cyclic voltammetry or differential pulse voltammetry, desmetryn needs an accumulation potential of  $-0.7$  V over a period of 50 s before differential pulse voltammetry determination from micellar or emulsified media). Attempts for simultaneous determination of complex mixtures of

pesticides in the same sample can generate serious overlapping of the polarographic signals. In such cases (e.g., captan, captafol, and folpet together, in river water samples), different chemometric methods such as artificial neuronal networks (ANNs) have been successfully used.

## Indirect Tests

Sometimes it is beneficial to avoid determination of the target compounds in complex samples requiring tedious sample preparation steps, as their presence, once confirmed, leads to the qualitative rejection of a product or an alert reaction. For example, existence



of quaternary ammonium pesticides in baby fruit pastes is reflected by the determination of the corresponding saccharide profile (the presence of maltose confirms adulteration of the products with sugars of cereal origin, related to potential presence of 'quats' at trace levels, and subsequently conducted to batch rejection). Ethephon (used as growth regulator and ripening accelerator in fruit and vegetable growing) is thermally hydrolyzed to ethylene, whose presence is determined by headspace GC.

## Quality Assurance in Pesticide Analysis

Analysis of pesticides in environmental, food, clinical, and forensic samples is undoubtedly a complex task because of the matrix complexity and low concentration values of the target compounds.

In order to produce accurate results, sampling, sample preparation methods, analytical methods, analytical equipments, data processing should be validated and the working personnel should be qualified. Sampling and sample storage should be, at the least, studied for representativeness and stability. In this direction, a knowledge of the sample history is essential. Sample preparation techniques should be checked for selectivity, precision, and recovery. Separation techniques and detection should be validated for selectivity, precision (repeatability and intermediate reproducibility), linearity (statistical determination of quantitation and detection limits – LOQ and LOD – included), accuracy, and robustness. Analytical instrumentation should be attentively and periodically tested (analytical column(s) should be considered separately). The use of certified reference materials during method validation remains an imperative. Working personnel should be periodically trained. Well-trained and science-oriented analysts should be able to critically review and confirm their own results. Repetitive and sensitive analysis steps should be described clearly and in detail by means of standard operation procedures. One of the main objectives of the laboratory should be data traceability. In this respect, description of the applied methodologies with the mention of all relevant details, equipment status, and related information on its calibration and calculation procedures should be available. Error propagation studies are highly advisable for the entire analytical procedure, from sampling to results calculation. Participation in collaborative studies should be

considered in order to evaluate the overall quality of a given laboratory. Good laboratory practice guidelines should be interpreted in a positive, science-oriented way, focused on the awareness of the critical daily working problems. Whole activity should be oriented to ensure complete equivalence between good laboratory practices and good analytical practices, which do not necessarily mean the same thing.

*See also:* **Capillary Electrochromatography.** **Capillary Electrophoresis:** Environmental Applications. **Gas Chromatography:** Overview. **Immunoassays:** Overview. **Liquid Chromatography:** Overview. **Mass Spectrometry:** Overview. **Micellar Electrokinetic Chromatography.** **Sensors:** Overview. **Spectrophotometry:** Overview. **Supercritical Fluid Chromatography:** Applications. **Thin-Layer Chromatography:** Overview.

## Further Reading

- Aprea C, Colosio C, Mammone T, Minoia C, and Maroni M (2002) Biological monitoring of pesticide exposure: A review of analytical methods. *Journal of Chromatography B* 769: 191–219.
- Chuang JC, Hart K, Chang JS, *et al.* (2001) Evaluation of analytical methods for determining pesticides in baby foods and adult duplicate-diet samples. *Analytica Chimica Acta* 444: 87–95.
- Geerdink RB, Niessen WMA, and Brinkman UATh (2002) Trace level determination of pesticides in water by means of liquid and gas chromatography. *Journal of Chromatography A* 970: 65–93.
- Kataoka H (2003) New trends in sample preparation for clinical and pharmaceutical analysis. *Trends in Analytical Chemistry* 22: 232–244.
- Keith LH (ed.) (1996) *Compilation of EPA's Sampling and Analysis Methods*, 2nd edn. Boca Raton: CRC Press.
- Luque-Garcia JL and Luque de Castro MD (2003) Ultrasound: A powerful tool for leaching. *Trends in Analytical Chemistry* 22: 41–47.
- Mallat E and Barceló D (2001) Immunosensors for pesticide determination in natural waters. *Trends in Analytical Chemistry* 20: 124–132.
- Stan HJ (ed.) (1995) *Analysis of Pesticides in Ground and Surface Water II*. Berlin: Springer.
- Tomlin CDS (ed.) (2000) *The Pesticide Manual*, 12th edn. Hampshire: British Crop Protection Council Publications.
- Turner C, Eskilsson CS, and Björklund E (2002) Collection in analytical-scale supercritical fluid extraction. *Journal of Chromatography A* 947: 1–22.
- Warren N, Allan IJ, Carter JE, House WA, and Parker A (2003) Pesticides and other micro-organic contaminants in freshwater sedimentary environments – a review. *Applied Geochemistry* 18: 159–194.

The most common electrochemical procedures of pH evaluation are based on potentiometric measurements. The theoretically best described is the hydrogen electrode, consisting of platinum metal in contact with hydrogen gas but its use, however, is restricted to specialized laboratories. The various types of glass electrodes offer a wide pH range of dynamic response, which only at the ends of the pH scale may be subject to acid or alkaline errors. The glass electrodes are of various constructions, both macro and micro down to the micrometer scale. Most convenient for routine measurements are glass electrodes mounted in one body with the reference electrode. The uncertainty of electrochemical pH measurements is usually below 0.02 pH unit but may depend on the construction details of the cell. For special applications the glass electrode can be replaced by some metal/metal oxide electrodes, of which the antimony electrode is most commonly used. However, such sensing electrodes are more sensitive to the presence of redox species in the sample solution. Sensors that recently became more popular are those with polymeric membranes, containing a pH sensitive carrier such as tridodecylamine. Their use in nonaqueous media is often restricted, and their lifetime is often shorter than that of glass electrodes.

**See also:** Buffer Solutions. **Extraction:** Solvent Extraction Principles. **Indicators:** Acid–Base. **Ion-Selective Electrodes:** Glass. **Sensors:** Overview. **Quality Assurance:** Internal Standards. **Titrimetry:** Potentiometric.

## Further Reading

- Bates R (1973) *Determination of pH. Theory and Practice*. New York: Wiley.
- Buck RP, Rondinini S, Covington AK, *et al.* (2001) *The Measurement of pH – Definition, Standards, and Procedures*. Report of the Working Party on pH, IUPAC Provisional Recommendations, IUPAC.
- Hulanicki A (1987) *Reactions of Acids and Bases in Analytical Chemistry*. Chichester: Ellis Horwood.
- Inczyedy J (1973) *Analytical Applications of Complex Equilibria*. Chichester: Ellis Horwood.
- IUPAC (1998) In: Inczyedy J, Lengyel T, and Ure AM (eds.) *Compendium of Analytical Nomenclature*. Oxford: Blackwell Science.
- Kragten J (1978) *Atlas of Metal–Ligand Equilibria in Aqueous Solutions*. Chichester: Ellis Horwood.
- Minczewski J, Chwastowska J, and Dybczyński R (1982) *Separation and Preconcentration Methods in Inorganic Trace Analysis*. Chichester: Ellis Horwood.
- Pourbaix MJN (1949) *Thermodynamics of Dilute Aqueous Solutions*. London: Edward Arnold.
- Ringbom A (1963) *Complexation in Analytical Chemistry*. New York: Wiley.

# PHARMACEUTICAL ANALYSIS

Contents

**Overview**

**Drug Purity Determination**

**Stability Testing**

**Dissolution Testing**

**Sample Preparation**

**Plant Extracts**

## Overview

**A-M Tivert and J Vessman**, AstraZeneca R&D Mölndal, Mölndal, Sweden

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Pharmaceutical analysis is traditionally defined as analytical chemistry dealing with drugs both as bulk

drug substances and as pharmaceutical products (formulations). However, in academia, as well as in the pharmaceutical industry, other branches of analytical chemistry are also involved, viz. bioanalytical chemistry, drug metabolism studies, and analytical biotechnology. The development of drugs in the pharmaceutical industry is a long-term process, often taking more than a decade from the start of the research project to appearance of a drug on the market. That process involves several decision points, such as the choice of the candidate drug after the discovery phase, the application to the authorities before

testing the compound for the first time in humans, and finally the new drug application for marketing, which summarizes the data obtained from all the studies needed for approval of the drug as a medicine. In all these steps the amount of data generated is enormous. Analytical chemists are involved in many of the studies that constitute this documentation. Substance quality and its specifications are based on substance analysis, and that knowledge is later used for quality control (QC) of the substance during full-scale production. Product analysis involves dealing with the various formulations used for toxicological studies, clinical studies, and marketing. The results from such work lead to specifications that form the basis for the QC of the product. For both substances and formulations there is an increasing interest in the introduction of process analytical chemistry.

Biomolecules, i.e., macromolecules such as proteins or hormones, either produced by isolation from biological sources or by means of biotechnology, must also be subjected to careful analytical control. Thus, while the analytical tasks required for biomolecules are somewhat different from those of ordinary pharmaceuticals, when it comes to regulation and documentation of their quality and properties they definitely belong to the same group.

There are a number of regulations that have to be followed in the development of pharmaceuticals as well as in their production. Regulatory approval is required prior to each clinical trial and before marketing is licensed.

An important part of the development process is safety evaluation, primarily the toxicology tests, which run from 1 to 24 months in different species. During this time bioanalytical studies are performed as well as control of the formulations used in the tests. After approval for marketing the authorities exercise control of products on the market and require postproduction stability data. Public interest in the quality of drugs is also reflected in the compilation of substance monographs in compendia that are known as pharmacopoeias. In addition to collections of substance monographs, these pharmacopoeias contain general analytical methods and some also contain monographic requirements on the formulation of the substances.

This article provides an overview of mainly substance and product analysis (traditional pharmaceutical analysis), as used in the pharmaceutical industry. The support of other branches of analytical chemistry will be mentioned.

## Bulk Drug and Pharmaceutical Products

### Common Features

**Identity testing** Identity testing is used to verify that the drug substance is what it is stated to be, that a raw material is qualified, or that the formulation contains the correct compounds. Infrared (IR) spectra are used quite extensively in industry, whereas the pharmacopoeias often have a set of alternative tests. These can be color reactions, melting point of a compound or derivative, optical rotation values, or ultraviolet (UV) spectral data such as maximum wavelength and absorptivity. Today, chromatographic data are also used to support identity tests. Near-infrared (NIR) spectrometric methods are used increasingly for identity testing as no sample preparation is needed and it can be used in a noninvasive way, i.e., the container is left intact if made of glass or plastic.

**Impurities** Impurities or degradates require separation methods and are usually studied at the level from 0.05% to 2% (purity patterns) or from 0.1% to 5% (stability profiles). This means that the analytes have to be quantified in up to a 2000-fold excess of the major compound. In practice, qualitative work is performed at still lower levels. This sometimes creates problems in the chromatographic methods as minute amounts of related substances may be hidden under the peak of the drug itself. This was the background for the interest in peak purity tests, which, however, did not really get momentum, one reason being that a peak impurity present at below 1% may be difficult to detect. Also, peak purity tests are inferior to the use of complementary separation systems. Alternative liquid chromatography (LC) methods, gradient elution, or capillary electrophoresis (CE) are advisable instead. Today, sensitive nuclear magnetic resonance (NMR) detection online is a reality and can be used when other means fail.

**Selected analytes** Selected analytes sometimes have to be analyzed at parts per million levels. Typical examples are aromatic amines, nitrosamines, reactive intermediates left from the synthesis, or certain solvent residues, i.e., all components that are known to be noxious and thus must be controlled separately.

**Compendial analysis** Pharmacopoeias are the official collections of drug standards. They all include requirements for drug substances but only a few have monographs for products (formulations). Harmonization between the compendia for the three big

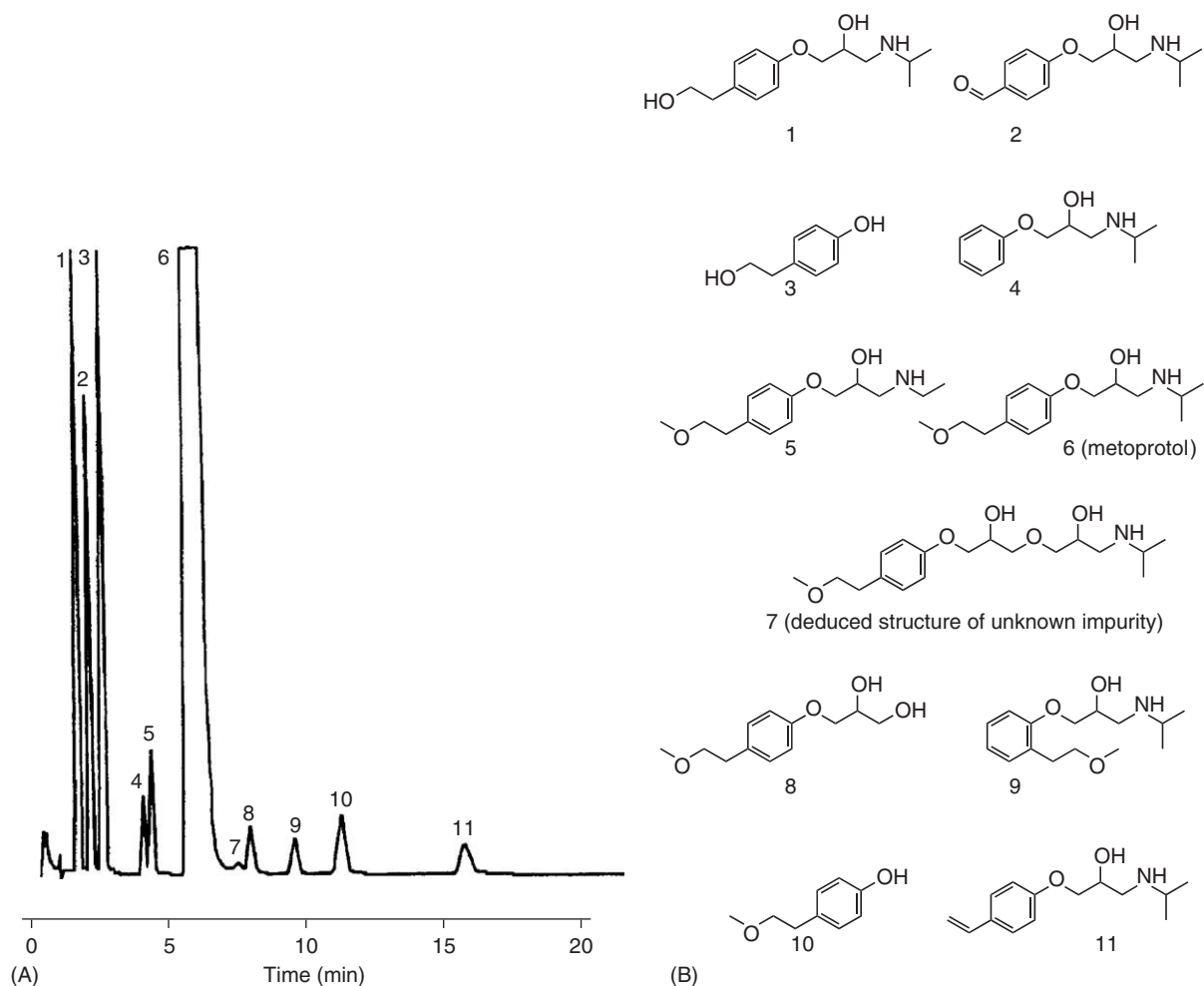
markets, Europe, Japan, and the US, is under way. There is an acceptance now that the major pharmacopoeias are, in principle, intended for the pharmaceutical industry and the authorities and not for community pharmacies as was the case earlier.

### Bulk Drug Analysis

**Physicochemical characterization** Physicochemical characterization yields a number of important parameters that can be used in the control of the quality of a substance. Typical properties are melting point and other thermal data, solubility, acid–base behavior with  $pK_a$  values, redox potentials, polymorphism, and spectral information. Other property-influencing parameters are choice of counterion and salification studies.

**Purity tests** Purity tests are, in particular, focused on related substances such as homologs, analogs,

by-products from the synthesis, or degradates. Enantiomeric purity is now a regular test requirement for chiral compounds. Chromatography is currently the most important check, giving essentially the fingerprint of a synthesis. An example of a separation of some potential impurities that have been added to a metoprolol sample is given in **Figure 1**. For qualitative studies LC–mass spectrometry (MS) is now established as the primary online analytical tool for the elucidation of unknown structures among the impurities. From a toxicological point of view, the impurity profile of the substance batch used in safety studies should form a reference for the full-scale production material. This means that in later batches impurities in amounts that exceed those found in the batches used for toxicology should be avoided. The high standard and good reproducibility required of the purity profile method is clearly



**Figure 1** (A) Separation by LC of metoprolol and some of its potential impurities that have been added in amounts from 0.01% to 0.4% to a pure sample. Peak 6, Metoprolol; peak 7, an unknown impurity. Peaks 1–11, see structures in (B). Column: 125 mm long and 4 mm ID filled with 5  $\mu$ m C-8 particles (Li Chrospher RP-Select B). Mobile phase: acetonitrile, 17% in ammonium phosphate buffer 0.05 mol  $l^{-1}$ , pH 3.2. Detection at 280 nm. (Reproduced with permission from Lars A. Svensson, Astra Hässle/AstraZeneca, Mölndal, Sweden.)



evident from that perspective. There are some other tests that also contribute to the general impression of the quality of a substance, i.e., tests for protolytic impurities, content of chloride, sulfated ash or residue upon ignition that gives the inorganic content. These tests reflect the performance of the purification process in general, but their importance will probably diminish in the future. For inorganic ionic analytes the older methods can be replaced by ion chromatography.

**Heavy metals and arsenic** Heavy metals are routinely determined, often with one or other form of sulfide precipitation. These compendial tests are performed from the viewpoint of safety and the general limits ( $1\text{--}30\ \mu\text{g ml}^{-1}\ \text{g}^{-1}$ ) are now more often related to the dose. For metals such as mercury, lead, cadmium, or nickel, atomic absorption spectrometry or other instrumental methods are often prescribed. Copper and other transition metals can act as catalysts in certain degradation reactions and thus require special attention. Arsenic tests that were important at the beginning of the previous century are now being phased out.

**Potency assay** Common for all bulk drugs is an assay of potency. This can be an aqueous or nonaqueous titration based on protolytic properties or on some other property. Many compounds lack functional groups suitable for titration and here separation methods (LC in particular) are often used. However, titrations are preferable as their precision is, in general, superior.

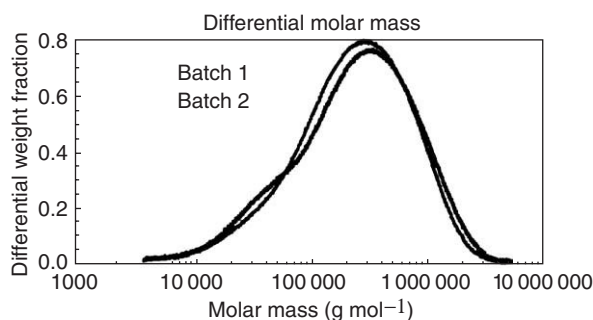
**Biomolecules** Biomolecules such as proteins represent a special type of bulk drug. Depending on the source, various chromatographic tests are used to show the absence of contaminating proteins remaining from the purification process. The separation methods used have a more biochemical character and differ from those for compounds of synthetic origin. Immunochemical techniques are very often used as complements. Typical tests include separation of dimers, trimers, etc., from the biomolecule itself. Molar mass determination by size-exclusion chromatography (SEC) is common and is strongly supplemented by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS).

Biotechnological products have some advantages over products from human or animal origin with respect to the risk of transferred diseases, but they also have some special requirements. In particular, when parenteral use is intended, the absence of host proteins has to be guaranteed as well as DNA residues

from the vector used for expression. Such matters often need the attention of specialist laboratories.

**Bioassay** Many biomolecules have, over the years, been assayed using methods where biological activity in an animal, organ, or receptor is assessed, i.e., a bioassay. The traditional opinion has been that such bioassays cannot be replaced by physicochemical methods as the latter do not reflect the biological activity. Bioassays, no matter how well characterized they are, have certain disadvantages with respect to precision, time, and cost compared to instrumental methods. The pharmaceutical industry has therefore been able to show that, for several of its biotechnological products, e.g., insulin and human growth hormone, chromatographic methods (LC in particular) can give the same information. Moreover, proteins from different species can be chromatographically separated and degradation products well quantified. This is not possible in quantitative bioassays where only the sum of activities is obtained. A similar paradigm shift has also taken place for antibiotics, where LC methods give information not available from microbiological assays. However, it should be noted that for many biomolecules there are, as yet, no alternatives to bioassay.

**Excipients** There are a great number of materials that are used to transform a substance from an active compound to a medicine useful for a patient in a therapeutic situation. These materials, or excipients, are becoming more important in the construction of modern drug delivery systems. Many excipients are macromolecules and have been used for decades in traditional remedies. With modern drug delivery systems the old requirements may not always fit those necessary for the technologically advanced products of today. This has become increasingly evident in recent years. Moreover, the requirements on excipients in the various pharmacopoeias are not always consistent with each other and are often rather vague. This has been recognized at an international level and efforts at harmonization are proceeding. Polymeric excipients are generally characterized by some average physicochemical property such as viscosity. Studies of the distribution of molar mass have seldom been performed on such excipients due to the lack of suitable methods. However, with SEC in online combination with light scattering (e.g., MALS, multi-angle light scattering) and refractometric detection it is now possible to determine the molar mass distribution directly without any standard calibration procedure. A nice example is given in **Figure 2**. The majority of the excipients used today, however, have molar masses



**Figure 2** The molar mass distributions of two hydroxypropyl-methyl cellulose (HPMC) batches (Metolose 60 SH 10 000 cps, Shin-Etsu, Tokyo, Japan) determined with SEC connected online to multi-angle scattering (MALS) and refractometric (RI) detectors. The broad distributions contain HPMC chains ranging from a few thousands up to several millions in molar mass. Despite the fact that the two batches belong to the same viscosity grade, significant differences are seen in the molar mass distributions. This emphasizes the importance of proper characterization of the complete molar mass distribution instead of only estimations via viscosity averages. (Reproduced with permission from Bengt Wittgren, Analytical Development, Astra Hässle/AstraZeneca, Mölndal, R&D Mölndal.)

in the same range as the active compounds. The tests for safety and purity are then similar to those of the drugs.

Excipients could usefully be classified or tested according to their properties at three levels, viz. molecular, particulate, and bulk properties. Those are tested for by the manufacturer of a dosage form. It is not clear which of those properties should be covered by the official compendia. Testing of functionality, i.e., at particulate or bulk level, does not seem to be possible yet. Typical tests are bulk density, specific surface area, flowability, and particle size distribution. However, the standardization of methodology in compendia, without specification limits, will probably be of help for both vendor and buyer. Therefore, functionality related tests are now being proposed in the pharmacopoeias. As excipients are getting more complex, their analytical characterization is very important. Interesting opportunities lie ahead, particularly with macromolecular separation, MALDI-TOF-MS, and spectrometric methods such as NIR.

### Pharmaceutical Products

A medicine is much more than simply a drug substance, and huge efforts are put into the development of biopharmaceutically optimized drug delivery systems. Analytical chemists contribute to that process by analyzing the experimental formulations with respect to various properties such as homogeneity, content, stability, and release of the active agent in dissolution testing but also through bioanalysis to

create data for *in vitro*–*in vivo* correlations. In particular, stability has to be monitored over the entire test period.

**Solid dosage forms** Solid dosage forms, e.g., tablets and capsules, are by far the most common for several reasons. The production of relevant doses is easy to accomplish, and scale-up is usually a standard technological process. All divided dosage forms have strict requirements for uniformity of content, i.e., a statistical sampling of the batch should show a uniform distribution of the active component. This requirement is especially important for units with very small amounts of the active component, i.e., from a few micrograms per dose to 50 mg. This has often led to automated analytical methods to cope with the large number of samples.

The pharmacokinetic performance of a drug influences the construction of a formulation. This has nowadays led to a dominance of drug delivery systems that provide a controlled or modified release of the drugs, defined as extended or delayed release. Release-controlling polymers are used to build up a barrier that prevents immediate release of a compound. In this way high peak plasma concentrations of drugs are avoided and usually only one dose per day is necessary. The characteristic properties of the formula are evaluated in *in vivo* tests where blood samples are analyzed often by extremely sensitive bioanalytical methods.

However, for routine QC it is usual to rely on *in vitro* models, which obviously should be correlated with the *in vivo* data. Dissolution testing has been standardized in the pharmacopoeias for a long time with respect to release media, apparatus, and other conditions. However, in modified release formulations the prescribed conditions might have to be changed. This is the responsibility of the analytical chemist, whilst still having the routine testing conditions in mind, i.e., QC methods should also be practically feasible.

**Parenterals** Parenterals are dosage forms intended for injection into the body. Water is normally used as the solvent. The special tests for parenterals include sterility and absence of particles as well as endotoxins that can give fever reactions. The old test that was performed on rabbits is nowadays often replaced by a test based on the reaction of endotoxins with a lysate from *Limulus amoebocyte* (LAL test), which is less time consuming and more exact. Sterility testing is complicated from a sampling point of view because of the random appearance of microbial contaminants. A thorough in-process validation is the best way to ensure that the products are sterile. In large



volume, parenterals' requirements on limits for particle contamination have created a need to analyze for particles down to the size of a few micrometers, usually by light blocking or by conductivity techniques. For compounds that are sparingly soluble in water other solvents or cosolvents can be used creating problems of quite a different kind for both formulator and analyst, such as evaluation of precipitation phenomena and interactions with packaging materials. For freeze-dried (lyophilized) products the reconstitution properties have to be controlled.

**Other dosage forms** Other dosage forms are needed for topical administration such as creams, ointments, and patches. The latter should usually deliver the drug over an extended period of time and thus require reliable *in vitro* release testing and as a consequence also *in vivo* data.

Sometimes other routes of administration are necessary. Thus, suppositories are used for rectal delivery and sprays for the nasal route. Sublingual delivery can be advantageous, e.g., for nitrate esters. For asthma the use of inhalators has increased considerably as reliable hi-tech delivery systems have been developed. Here, the uniformity of the inhaled dose (usually a few micrograms), as well as the narrowness of the particle size distribution, must be safeguarded.

The analytical problems of these dosage forms have to do both with the type of excipients used and their characterization and quality as well as the function of the delivery system.

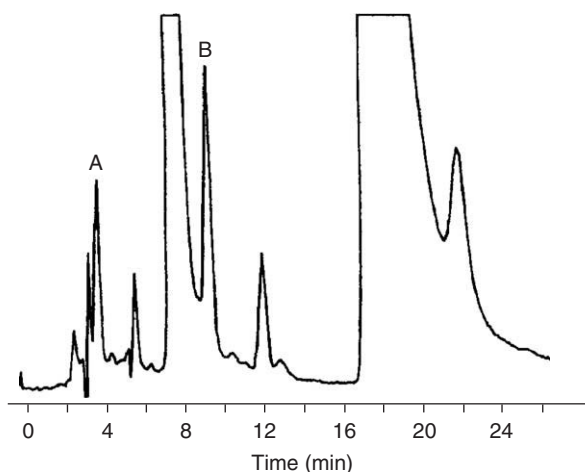
**Toxicological formulations** Toxicological formulations appear in the early project work in short-term toxicology tests performed before testing in humans and later on during long-term carcinogenicity studies before approaching the authorities for marketing approval. The reason for mentioning them here is that at this early stage a full understanding of the properties of a drug is not always available, and, most important, the animal feed into which the drug may be blended, is a very difficult matrix in which to analyze a drug. Yet, the analyst has to determine the homogeneity and the stability of such formulated animal feeds during the use of the material. Here the rules of Good Laboratory Practice (GLP) are emphasized including validation of analytical methods.

**Packaging materials** Packaging materials are also part of the medicine presented to the patient. They are usually polymers that have properties chosen to give the drug product protection during storage and handling. Usually multilayer materials are used, as in blister packages for tablets, with some layers being a barrier toward moisture penetration.

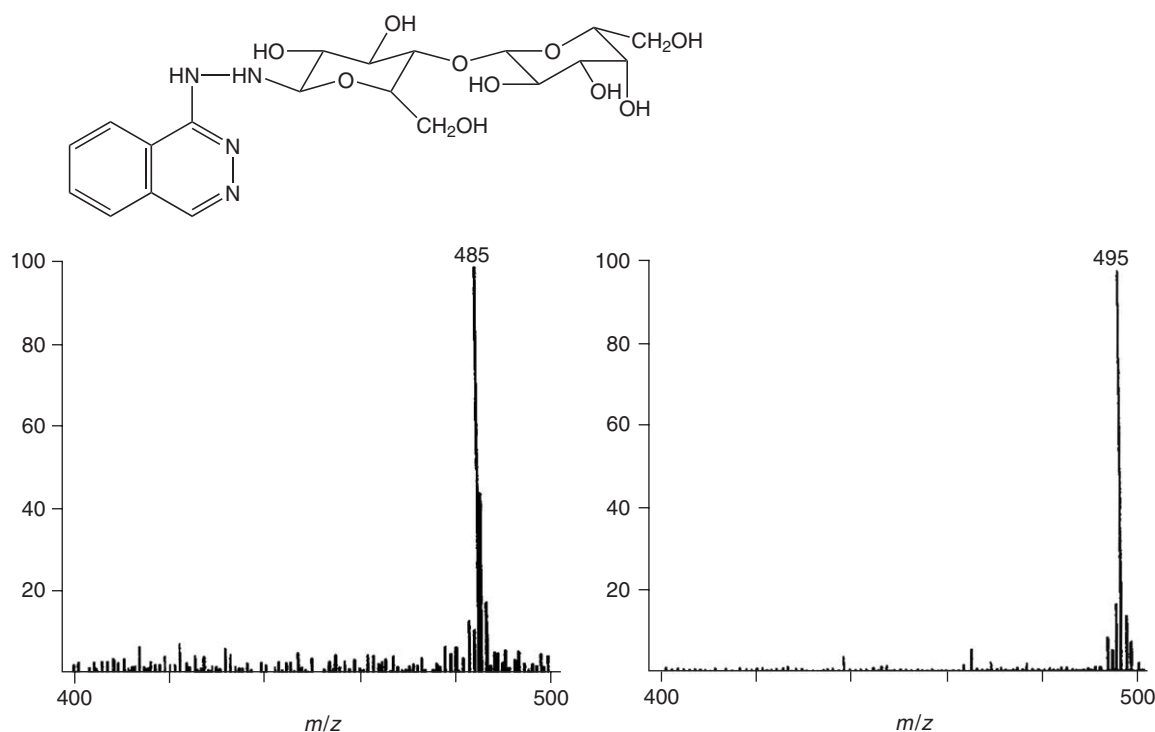
Plastic bottles are becoming more common now. For many products, particularly parenterals, there is a risk of interaction between the polymeric material and the active component, especially if the latter has a lipophilic character. Rubber stoppers in parenterals are likewise prone to trap organic molecules from, and also leak stabilizers into, the solution. Biomolecules in solution may adsorb onto these surfaces and this may be especially significant if the amounts in solution are minute. From these comments it is clear that stability testing must also be performed in the consumer package under appropriate conditions of temperature, humidity, and light.

**Stability studies** Stability studies constitute a major task for the analytical chemical laboratory. The aims of initial physicochemical studies and investigations of incompatibilities, i.e., preformulation studies, are to identify the weak points of a compound, in order to avoid vulnerable conditions in the formulation. An example of an incompatibility reaction in a formulation containing three active ingredients and stored under accelerated conditions is given in Figure 3. The structure of the degradate was elucidated with electrospray LC-MS and verified as shown in Figure 4. Heat, moisture, oxygen, and light may all influence stability.

Tests have to be performed under standardized conditions with respect to temperature and humidity. Accelerated tests can be used to identify the weak points, but in the documentation submitted to the authorities long-term stability data under normal conditions are required. In these studies stability



**Figure 3** Separation by LC of degradates from a formulation containing hydralazine. After storage at accelerated conditions components A and B started to grow the peaks. They were both the result of an incompatibility with lactose. The structure of B was elucidated by LC-MS as shown in Figure 4. (Reproduced with permission from Rose-Marie Janson, Astra Hässle/AstraZeneca, Mölndal, Sweden.)



**Figure 4** Mass spectra of degradate B in **Figure 3**. Electrospray LC–MS was used with a 0.5 m fused silica capillary column, ID 250  $\mu\text{m}$ . The packing material was 5  $\mu\text{m}$  C-18 Chromasil and the mobile phase was 60% acetonitrile in 5  $\text{mmol l}^{-1}$  ammonium acetate with a flow rate of 15  $\mu\text{l min}^{-1}$ . The left-hand panel mass spectrum was obtained with the mobile phase consisting of water and the right-hand panel with the use of heavy water (deuterium oxide) instead. The mass number difference is due to nine hydrogen atoms exchanged with deuterium plus the ion charge  $\text{H}^+$  being changed to  $\text{D}^+$ . (Reproduced with permission from Karl-Erik Karlsson, Astra Hässle/AstraZeneca, Mölndal, Sweden.)

indicating assays are important and usually LC is the method of choice as the measurement of the drug content should not be disturbed by interfering components. Early degradation is, however, more reliably monitored through the degradates as the precision in the determination of a degradate is less demanding than the measurement of the corresponding decrease in the parent compound.

**Full-scale production and process analytical chemistry** Full-scale production is the final step in the research process. Scaling up from the pharmaceutical development laboratory via the pilot plant to full-scale production is not without problems. This is one of the reasons for the requirement by the authorities that stability studies should be performed and reported for the first three production batches. The transfer (scaling up) can be facilitated by thorough process controls that give an understanding of those parameters that must be controlled. Such efforts should start at least at the pilot plant scale and pave the way for the transfer to full scale. As pharmaceutical formulations are more complicated today than ever before, it is clear that process analytical chemistry/process analytical technology can

play an important role in the scaling-up procedure. In this way, it is possible to control the process by feedback reactions before severe deviations occur. Noninvasive techniques such as NIR and Raman spectrometry will be important. Probes with laser-based spectrometry also provide interesting possibilities in this context. Multivariate analysis for evaluation is a prerequisite for the success. If process analysis is utilized from the early formulation batches, many surprises will be avoided during scale-up and the production will be more reliable in quality and hence will be economical.

## Specifications and Quality Control

### Specifications

The quality requirements of a substance in bulk and those of a pharmaceutical formulation are compiled in specifications. In these documents the requirements on the various quality parameters are given as minimum or maximum limits or ranges. Formulations often have requirements on technical properties such as dissolution rate, disintegration, and hardness for tablets. All those requirements are the result of

comprehensive studies in the R&D phase, where the knowledge of the properties of the drug is gathered, resulting in the optimized formulation of the active component. The analytical and technical test methods that are linked to the specifications will have evolved during the R&D process and can, before the new drug application stage, be transferred to QC laboratories.

It is important to bear in mind that what is used in clinical trials should be reproduced in full-scale production. Of special importance is the particular batch, which is studied *in vivo* and compared with *in vitro* properties as in dissolution testing to establish an *in vivo*–*in vitro* correlation. This batch is of critical importance for the future and its documentation has to be thorough. The bioavailability, as documented in this batch, is the foundation for the coming market presentation. Equally important are the substance batches used in safety studies, and requirements on the knowledge of their quality are very exacting. The requirements at release can be somewhat tighter than those at a control performed any time during the entire lifetime of a medicine.

### Reference Substances

Most methods require some form of chemical reference substances (CRS), which have been characterized and presented in a regulatory submission more thoroughly than according to normal specifications, for example, by adding thermoanalytical data and spectrometric data for structural evidence, e.g., IR, NMR, MS, and UV. Compendial methods have official CRS, which are available for customers all over the world. It is important to remember that a reference substance can be used for different purposes, not all of them requiring extensive testing. So a CRS for identity testing is less demanding from a purity point of view. For daily work a less expensive working standard can be calibrated versus a CRS.

### Quality Control

Full-scale production is checked in the QC laboratories according to the specifications and test methods approved by the authorities. Decisions by QC management cannot be overruled by any person in the organization, which puts a particular onus on the competence and judgment of the person in charge of QC. In addition to the chemical and biological tests that comprise QC the organization can introduce preventive measure to avoid quality impairment. This activity is defined as integrated or total QC and is further outlined in Good Manufacturing Practice (GMP). It cannot be emphasized enough that chemical control at the end of a process can never replace

high standards in the process itself. This is sometimes so evident that release in certain cases can be given based on the documentation and control in the process steps. This is called parametric release. Parametric release has been accepted by the authorities in those cases where end control does not fully reflect failures in the production. This has been most evident in many biotechnological processes, where the absence of host proteins, DNA residues, or virus particles has been approved through a thorough validation procedure where the purification step is challenged. The analytical methodology is not adequate in this situation at the end control. Process analytical chemistry will in the near future play a similar role in showing that a process does not run beyond prescribed limits. This not only gives better quality in products, but also fewer failures and thus reduced costs.

Verification of identity is required at several stages during the process, not only at the end. This can be done in many ways and it is important to remember that the sum of the tests performed during a process also contributes to that verification. Biotechnological products are a special case in that the identity of the recombinant protein with that of the native one has to be established. Thus, it is important that the correct order of the sequence of amino acids is verified. Peptide mapping has been one way of showing this by comparing the chromatographic pattern of peptide fragments obtained after enzymatic cleavage. CE complements LC in this role. In this area, MS techniques will complement peptide mapping as MS will allow molecular ions to be determined up to and above 200 kDa.

### Regulatory Aspects

GMP was mentioned above and has been a cornerstone of pharmaceutical production and control since the 1970s. These regulations state clearly what has to be done to safeguard quality from the beginning of the production process to the end, viz. documentation, staff qualifications, standard of facilities, technical standards, handling of material, labeling, equipment, etc. GMP is valid for production and R&D, such as in validation studies, technology transfer, stability studies, clinical trial material release. The rules also state that all documentation of analytical methods should fulfill certain performance criteria, that instruments should have maintenance records, and that their performance and those of the method should be documented. GLP is the term used in safety studies for this kind of documentation and validation and is similar to GMP but differs in details.

### International Harmonization

Harmonization of requirements has successfully been made by the International Conference on Harmonisation (ICH), e.g., in guidelines on stability documentation, method validation, and investigation of impurities. Regarding impurities information is given on how the limits should be set and at what level identification is needed. Similar efforts are seen for dissolution testing and how to correlate *in vivo* and *in vitro* data. The question of bioavailability and bioequivalence is an important one and has been dealt with as well. As a consequence of the rapid development in chiral separations we can now see a harmonization of requirements also for chiral drugs.

### Validation of Analytical Methods

Validation of an analytical method establishes in laboratory studies that the performance characteristics of the method meet the requirements for the intended application, thus the method does what it is expected to do. The following items are listed by ICH: precision, accuracy, limit of detection, limit of quantification, specificity, range, linearity, and ruggedness. Of these, accuracy is probably the most difficult to document or obtain, at least for solid formulations. This has to do with the fact that recovery experiments are difficult to design in such a way that they resemble the process conditions. The reactions there can create interactions that are not obtained in an experiment where the analyte has only been mixed or spiked to the sample.

Specificity, or selectivity as recommended by IUPAC, is the factor in the validation process that is most often discussed. Selective methods are required in a product release specification, which means that a chromatographic procedure often has to be carried out but sometimes a simple UV method is sufficiently selective and thus free from interferences. Here cost-effectiveness should be the guide and analysts should therefore use their scientific arguments to justify using the simpler method. Validation procedures are equally important in the documentation of bioanalytical methods.

### Emerging Techniques in Pharmaceutical Analysis

Analytical chemistry is one of the disciplines most frequently involved in the R&D work performed in the pharmaceutical industry. This makes the industry very analysis intensive, which explains the high level of interest in testing new methods in order to get further information. It is also clear that this interest in learning, and using the latest techniques, stems

from a desire to obtain reliable information more quickly, or to obtain complementary data.

The research process is a long-term commitment and requires high standards in the results from the very beginning. The quality of the results relies on competent analytical chemists, but also on the availability of good instrumentation, which is often evident in the laboratories of the research intensive pharmaceutical industry. Separation techniques, especially LC, have had an enormous impact on pharmaceutical analysis. The combination of LC with MS has now extended the possibilities of the techniques as qualitative and quantitative data can be obtained online. The availability of less-expensive bench-top instruments operated by nonspecialists has changed the situation considerably.

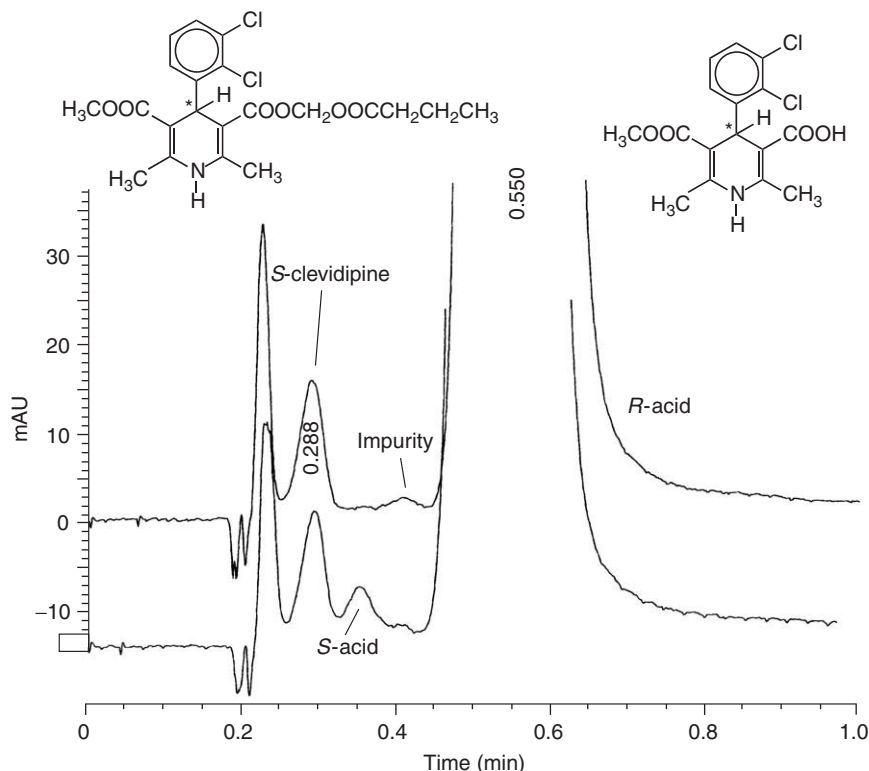
The development of new techniques usually proceeds in a stepwise fashion and, at present, mass spectrometric techniques are taking a giant step toward the analysis of macromolecules in a reliable way. This revolutionary process now shows promising results with molar masses of more than 200 kDa being determined using TOF techniques. Possibilities to combine them online with separation methods are at hand. These MS extensions have great impact on biotechnology products and macromolecular compounds used as excipients. Another area where rapid development is taking place is Raman spectrometry. New technology has opened interesting possibilities for this old technique. Many opportunities lie ahead, not least in imaging techniques, both in regular and in in-process analysis.

Separation methods will see continued growth, particularly the capillary techniques. Capillary chromatography is particularly useful when expensive mobile phases are used. A good example of LC-MS using packed fused silica columns 250  $\mu\text{m}$  in diameter is given in **Figure 4**. Here, heavy water was used interchangeably with an aqueous phase. Hydrogens bound to heteroatoms (O, N, S) were then replaced with deuterium, which was easily revealed in the mass spectrometer. This tells the analyst how many labile hydrogens the molecule has (in the case of the example this was 10). For macromolecules up to particles, the use of field-flow fractionation techniques is approaching broader acceptance. Further development of matrix-forming agents will take place that will facilitate the use of MALDI-TOF-MS for other classes than proteins, i.e., carbohydrate-based polymers. CE is an excellent complement to LC, not least for biotech applications and the technique can be used in QC laboratories as well. Capillary electrochromatography has not come that far yet but can deliver separations with impressive efficiency. Supercritical fluid chromatography (SFC)

with packed columns has mainly been used for chiral separations, one reason being the easy transfer of conditions to the semipreparative scale for fast isolation of enantiomers. An example of a rapid chiral separation is given in Figure 5. One limiting factor with SFC is that CO<sub>2</sub> can react with primary and secondary amines. However, it has now been shown with SFC–NMR that with a little steric hindrance at the nitrogen this reaction will not take place. NMR-based methods will be more common, with and without previous separations. LC–NMR is now a frequently used technique. With solid-state NMR valuable information on powder properties can be obtained. Circular dichroism in detection of chiral compounds can provide information on which isomer is at hand.

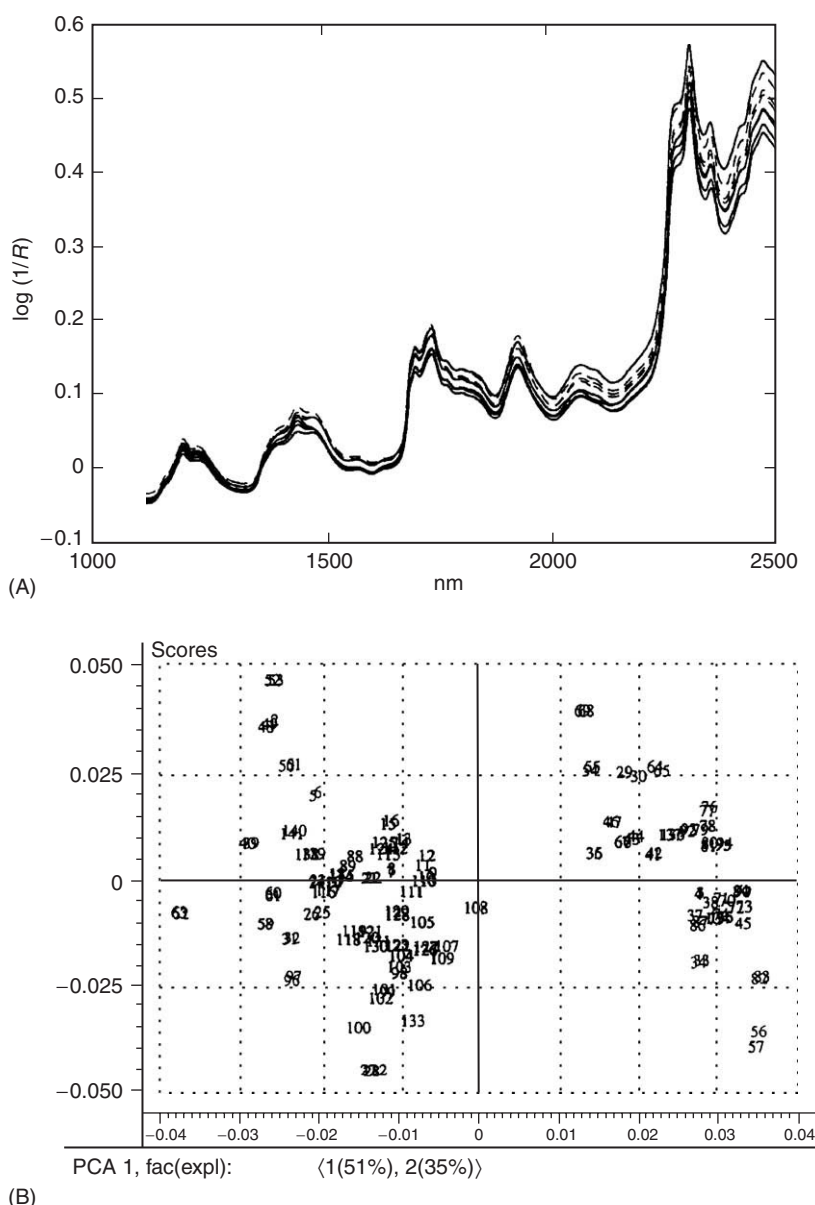
Chemometric methods, in particular multivariate analysis, have an impact on many types of techniques, especially when large amounts of spectral data are collected, and in areas where specific properties that are not easily measurable (such as taste, texture) have to be correlated with physicochemical

parameters. In this context, NIR has gained prominence within the pharmaceutical industry. An example from the analysis of a number of cellulose ether samples with reflective NIR is shown in Figure 6. The individual spectra do not show much in the way of differences. The spectra were corrected for different light penetration depths by multiplicative signal correction. Then, the entire spectra were projected by principal components analysis as points on a plane. The samples could be grouped into two clusters in this plane, one for each vendor. The calibration of this technique is entirely dependent on multivariate analysis or chemometrics and this combination grows in use. In the same way process analytical chemistry is heavily based on multivariate data. Many noninvasive approaches rely on NIR and multivariate calibration. New approaches combined with techniques that will stand harsh process conditions are also introduced, e.g., in process analytical applications. In the near future we will probably see applications in the pharmaceutical analytical field based on nanotechnology or lab on a chip and



**Figure 5** Fast SFC analysis of the *S*-enantiomer of clevidipine after hydrolysis into its corresponding acid. Column and conditions: Chiralpak AD 50 × 4.6 mm ID at 30°C, mobile phase carbon dioxide with 28% of 2-propanol, flow rate 4.0 ml min<sup>-1</sup>, backpressure 150 bar, UV detection at 240 nm. Sample preparation and work-up: 2 mg of *S*-clevidipine substance was dissolved in 0.5 ml of methanol followed by 50 μl of 1 mol l<sup>-1</sup> sodium hydroxide. After 10 min, 1 ml of water and 50 μl of sulfuric acid, 1 mol l<sup>-1</sup>, was added and 0.5 ml of dichloromethane as extraction medium. After brief vortexing and centrifugation 5 μl of the lower organic phase was loaded manually and injected. Upper trace sample, lower trace sample with ~0.1% of *S*-acid added. (Reproduced with permission from Gyllenhaal O (2001) Fast enantioselective separation of clevidipine and a dihydropyridine substituted acid by SFC on Chiral pak AD. *Fresenius' Journal of Analytical Chemistry* 369: 54–56; Springer.)





**Figure 6** (A) Reflectance spectra in the near-IR of a number of ethyl celluloses obtained from two vendors. (B) Principal component projection of the spectra in (A) after multiplicative signal correction. The left cluster is from vendor A and the right one from vendor B. (With thanks to Mats Josefson, Astra Hässle/AstraZeneca, Mölndal, Sweden.)

laser-based spectrometry that will open new possibilities to reveal conditions in the products during various stages of production. Clearly, in the future new principles of measurement may also be necessary and the analytical chemist will have much to contribute here.

See also: **Bioassays:** Overview. **Chemometrics and Statistics:** Multivariate Calibration Techniques. **Infrared Spectroscopy:** Near-Infrared. **Liquid Chromatography:** Size-Exclusion; Liquid Chromatography–Mass Spectrometry; Liquid Chromatography–Nuclear Magnetic Resonance Spectrometry. **Pharmaceutical Analysis:** Drug

Purity Determination; Stability Testing. **Process Analysis:** Overview. **Quality Assurance:** Quality Control; Reference Materials; Production of Reference Materials. **Supercritical Fluid Chromatography:** Applications.

### Further Reading

- Beckett AH and Stenlake JB (1988) *Practical Pharmaceutical Chemistry*, 4th edn, parts 1 and 2. London: Athlone Press.
- Connors KA (1982) *A Textbook of Pharmaceutical Analysis*. Chichester: Wiley.



- Fischer H, Gyllenhaal O, Vessman J, and Albert K (2002) *Analytical Chemistry* 75: 622–626.
- Folestad S and Johansson J (2003) Raman Spectroscopy. *European Pharmaceutical Review* 4: 36–42.
- Munson JW (1981) *Pharmaceutical Analysis: Modern Methods*, part A. New York: Dekker.
- Schill G, Ehrsson H, Vessman J, and Westerlund D (1984) *Separation Methods for Drugs and Related Compounds*,

2nd edn, pp. 187–198. Stockholm: Swedish Pharmaceutical Press.

Schirmer RE (1991) *Modern Methods of Pharmaceutical Analysis*, 2nd edn, vols. I and II. Boca Raton, FL: CRC Press.

Vessman J, Stefan RI, van Staden JF, *et al.* (2001) IUPAC Recommendations 2001. *Pure and Applied Chemistry* 73(8): 1381–1386.

## Drug Purity Determination

**P M Lacroix**, HPFBI Laboratory Programme, Ottawa, ON, Canada

Canadian Crown Copyright © 2005, Published by Elsevier Ltd. All Rights Reserved.

### Introduction

The purity of a drug and the determination of the nature and quantity of any impurities present are key questions for the modern pharmaceutical analyst. The nature and quantity of impurities in a drug product depend on the route of synthesis used to produce the drug raw material, purification procedures, dosage form, manufacturing and formulation, and the conditions under which it has been stored.

Impurities in reagents or starting materials may be carried through the synthesis intact or react to produce new impurities. By-products may be formed at various stages of the synthesis resulting in compounds with substitution at other sites or the formation of alternate ring systems. These by-products, in turn, may react in subsequent steps. Reaction residues such as reagents, solvents, catalysts, or unreacted precursors of the drug may remain as impurities. Nonstereospecific reactions may give rise to drug isomers. Undesirable polymorphs of the drug may be formed during recrystallization. Impurities may also result from drug degradation, particularly in formulations, or under certain conditions of light, heat, humidity, or acidity.

The terms process contaminants and related substances are commonly used when referring to impurities. A related substance is a compound that is chemically related to the drug, arises from the manufacturing process, or forms upon storage. It may be a starting material, synthetic intermediate, or degradation product. A process contaminant is an identified or unidentified impurity that may be introduced during the manufacturing or handling, but excludes related substances and water. Examples of process contaminants include inorganic compounds (heavy metals, and ions such as chloride or sulfate), reagents, and solvents.

### Types of Impurities and Purity Specifications

To ensure the quality, safety, and efficacy of pharmaceutical products, most nations have enacted laws and regulations governing the production, manufacturing, import, and sale of drugs. Many countries, the Council of Europe, and the World Health Organization (WHO) have established pharmacopoeias containing specifications for drugs, including purity, and the methods of analysis that must be used to verify that these specifications are met. Examples of pharmacopoeias include the Japanese Pharmacopoeia (JP), the British Pharmacopoeia (BP), the United States Pharmacopoeia (USP), the European Pharmacopoeia (EP), and the International Pharmacopoeia (IP).

The USP defines five types of impurities: foreign substances, toxic impurities, concomitant components, signal impurities, and ordinary impurities.

Foreign substances are adulterants or contaminants that do not result from the synthesis or manufacturing process. Depending on the nature and concentration of the foreign substance, it may or may not be detected using the pharmacopoeial method. A drug found to contain a foreign substance would not meet compendial standards.

Toxic impurities are those that have a significant undesirable biological activity even at low concentrations and require identification and quantification by specific tests. They may arise from the synthesis, preparation, or degradation of the drug. An example of a toxic impurity is amphetamine hydrochloride in phenylpropanolamine hydrochloride, which the USP limits to 0.001%.

Concomitant components are not regarded as impurities but merely as a normal component of the drug. They may be compounds extracted along with a drug that is a natural product, for example, 17- $\alpha$ -hydroequilin, 17- $\alpha$ -oestradiol, and 17- $\beta$ -dihydroequilin in conjugated estrogens derived from equine urine. Concomitant components also include the inactive isomer of a drug administered as a racemic mixture

- Fischer H, Gyllenhaal O, Vessman J, and Albert K (2002) *Analytical Chemistry* 75: 622–626.
- Folestad S and Johansson J (2003) Raman Spectroscopy. *European Pharmaceutical Review* 4: 36–42.
- Munson JW (1981) *Pharmaceutical Analysis: Modern Methods*, part A. New York: Dekker.
- Schill G, Ehrsson H, Vessman J, and Westerlund D (1984) *Separation Methods for Drugs and Related Compounds*,

2nd edn, pp. 187–198. Stockholm: Swedish Pharmaceutical Press.

Schirmer RE (1991) *Modern Methods of Pharmaceutical Analysis*, 2nd edn, vols. I and II. Boca Raton, FL: CRC Press.

Vessman J, Stefan RI, van Staden JF, *et al.* (2001) IUPAC Recommendations 2001. *Pure and Applied Chemistry* 73(8): 1381–1386.

## Drug Purity Determination

**P M Lacroix**, HPFBI Laboratory Programme, Ottawa, ON, Canada

Canadian Crown Copyright © 2005, Published by Elsevier Ltd. All Rights Reserved.

### Introduction

The purity of a drug and the determination of the nature and quantity of any impurities present are key questions for the modern pharmaceutical analyst. The nature and quantity of impurities in a drug product depend on the route of synthesis used to produce the drug raw material, purification procedures, dosage form, manufacturing and formulation, and the conditions under which it has been stored.

Impurities in reagents or starting materials may be carried through the synthesis intact or react to produce new impurities. By-products may be formed at various stages of the synthesis resulting in compounds with substitution at other sites or the formation of alternate ring systems. These by-products, in turn, may react in subsequent steps. Reaction residues such as reagents, solvents, catalysts, or unreacted precursors of the drug may remain as impurities. Nonstereospecific reactions may give rise to drug isomers. Undesirable polymorphs of the drug may be formed during recrystallization. Impurities may also result from drug degradation, particularly in formulations, or under certain conditions of light, heat, humidity, or acidity.

The terms process contaminants and related substances are commonly used when referring to impurities. A related substance is a compound that is chemically related to the drug, arises from the manufacturing process, or forms upon storage. It may be a starting material, synthetic intermediate, or degradation product. A process contaminant is an identified or unidentified impurity that may be introduced during the manufacturing or handling, but excludes related substances and water. Examples of process contaminants include inorganic compounds (heavy metals, and ions such as chloride or sulfate), reagents, and solvents.

### Types of Impurities and Purity Specifications

To ensure the quality, safety, and efficacy of pharmaceutical products, most nations have enacted laws and regulations governing the production, manufacturing, import, and sale of drugs. Many countries, the Council of Europe, and the World Health Organization (WHO) have established pharmacopoeias containing specifications for drugs, including purity, and the methods of analysis that must be used to verify that these specifications are met. Examples of pharmacopoeias include the Japanese Pharmacopoeia (JP), the British Pharmacopoeia (BP), the United States Pharmacopoeia (USP), the European Pharmacopoeia (EP), and the International Pharmacopoeia (IP).

The USP defines five types of impurities: foreign substances, toxic impurities, concomitant components, signal impurities, and ordinary impurities.

Foreign substances are adulterants or contaminants that do not result from the synthesis or manufacturing process. Depending on the nature and concentration of the foreign substance, it may or may not be detected using the pharmacopoeial method. A drug found to contain a foreign substance would not meet compendial standards.

Toxic impurities are those that have a significant undesirable biological activity even at low concentrations and require identification and quantification by specific tests. They may arise from the synthesis, preparation, or degradation of the drug. An example of a toxic impurity is amphetamine hydrochloride in phenylpropanolamine hydrochloride, which the USP limits to 0.001%.

Concomitant components are not regarded as impurities but merely as a normal component of the drug. They may be compounds extracted along with a drug that is a natural product, for example, 17- $\alpha$ -hydroequilin, 17- $\alpha$ -oestradiol, and 17- $\beta$ -dihydroequilin in conjugated estrogens derived from equine urine. Concomitant components also include the inactive isomer of a drug administered as a racemic mixture

but for which only one optical isomer has biological activity. An example of such a drug is atenolol for which the  $\beta$ -adrenoreceptor blocking activity resides mainly in the *S*-(–)-enantiomer.

Signal impurities are compounds that provide key information on the synthesis or degradation of the drug and require identification and quantification by specific tests. 4-Epianhydrotetracycline is a signal impurity that provides information on the degradation of tetracycline. The USP has set a limit of 2.0% on this impurity in tetracycline raw material.

Ordinary impurities are considered innocuous in that they have no significant undesirable biological activity at or below the specified limits. These impurities arise from the synthesis, formulation, or degradation of the drug. The USP monograph for glycopyrrolate contains a thin-layer chromatography (TLC) test for ordinary impurities for which the limits are 0.5% for individual impurities and 2.0% for the total of all impurities.

Impurities are typically resolved from the drug by a chromatographic procedure and quantified by comparison to an external standard, often the drug itself, rather than by comparison to standards of the individual impurities. Pharmacopoeial methods must be suitable for pharmaceuticals produced by different manufacturers that may contain different types and amounts of impurities. Pharmacopoeias typically view 1.0–2.0% as the general limit for total ordinary impurities, unless there is documentation to support a higher or lower level. Concomitant components, toxic or signal impurities are not included in the estimate of ordinary impurities and separate limits may be set for these, as necessary.

Individual ordinary impurities in drug raw material are typically limited to 0.5% or less. Many regulatory agencies require manufacturers to identify impurities present at levels of 0.1% or greater. If the structure is similar to that of a known carcinogen, mutagen, or other hazardous compound, toxicological and pharmacological data will be required prior to approval. Factors that are considered when establishing impurity limits are: toxicity, the costs involved in the cleanup procedure, and what levels can be expected when Good Manufacturing Practices are in place.

Pharmacopoeial methods reflect the state of knowledge of a particular drug at the time of publication. As more knowledge is gained, the need to revise existing methods or develop new ones may arise. For example, the existence of new impurities formed as a result of a different route of synthesis may come to light or, alternatively, better analytical techniques may be developed. Thus, modern pharmacopoeias have ongoing revision procedures.

*Pharmacopoeial Forum* (PF) and *Pharmeuropa* are USP and EP publications, respectively, and contain proposed monograph revisions, new monographs, and articles on other compendial matters. The purpose of these publications is to make proposed changes available for public comment before they become official.

In 1990, the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) initiated a process involving both regulators and industry as partners in scientific and technical discussions on harmonization of testing procedures required to ensure and assess the safety, quality, and efficacy of medicinal products containing new drugs.

The six parties that are the founding members of ICH and directly involved in the decision-making process represent the regulatory bodies and the research-based industry in the European Union, Japan, and the USA, the three regions where the majority of new drugs are developed. They are: the Committee for Proprietary Medicinal Products of the European Agency for the Evaluation of Medicinal Products and the European Federation of Pharmaceutical Industries and Associations in the EU, the Ministry of Health, Labour and Welfare and Japan Pharmaceutical Manufacturers Association in Japan, the Food and Drug Administration and Pharmaceutical Research and Manufacturers of America in the USA.

A number of ICH guidelines have been published and are now at step 5, which means that the final draft has been signed by the three regulatory agencies and it is incorporated into regional guidelines or regulations. These include: Impurities in new drug substances (1995), Impurities in new drug products (1996), Impurities: Guideline for residual solvents (1997), Text on validation of analytical procedures (1994), and Validation of analytical procedures: Methodology (1996).

The ICH guidelines on impurities apply to new drugs only and several classes of drugs are excluded, for example, biological or biotechnological products, radiopharmaceuticals, herbal products, and crude products of plant or animal origin. Drug substance impurities are classified as organic, inorganic, and residual solvents and drug product impurities are defined as degradation products. Polymorphs, enantiomeric impurities, and foreign contaminants are not covered in the impurity guidelines. Organic impurities are generally related to the synthesis of the drug substance or may arise during storage and include starting materials, by-products, intermediates, degradation products, and materials used in the synthesis, such as reagents, ligands, and catalysts. Inorganic impurities are generally related to the

synthesis and may include heavy metals, inorganic salts, reagents, ligands, catalysts, and process materials such as filter aids or charcoal that remain in the final drug substance. Residual solvents are organic or inorganic liquids that are used in the manufacturing process and remain in the final drug substance.

The ICH guidelines on impurities set requirements for the reporting, identification, qualification, and control of impurities. Qualification is the process of acquiring and evaluating data that establishes the biological safety of an individual impurity or a given impurity profile at the level or levels specified. Specified impurities are identified and unidentified impurities expected to be present in the new drug substance and new drug product over the period of intended use and under recommended storage conditions. Specified impurities should be individually listed in the specifications.

## Methods and Techniques for Drug Purity Determination

### Methods

Factors to consider in the choice of an analytical method for a pharmaceutical product include the chemical and physical nature of the drug and its impurities, the matrix in which it exists, the purpose of the analysis, and the analytical performance required. Different methods may be necessary to quantify impurities with different physicochemical properties.

Impurity levels are frequently controlled by limit tests based on TLC. A standard, either the drug itself or an impurity, is applied to a TLC plate at a level corresponding to the limit for individual impurities, and used for comparison purposes to impurity spots observed in the chromatogram of a test sample developed on the same plate. There is usually a requirement that no spot in the test chromatogram be more intense in color than the spot corresponding to the standard.

For semiquantitative tests, the level of each impurity in the test solution is estimated by comparison to a series of concentration levels of standard solution applied to the same plate.

Quantitative tests provide an estimate of the concentration of a specific impurity. This type of test is normally required for the determination of signal or toxic impurities and is used with increasing frequency for ordinary impurities. Quantitative tests are based on techniques that provide a high degree of accuracy and precision such as gas chromatography (GC) and liquid chromatography (LC).

### Method Development

The development of methods for purity determination requires knowledge of the structure and other properties of compounds likely to be present as impurities. Potential impurities may be postulated from a review of the routes of synthesis of the drug. These compounds, and others, may be isolated from mother liquors or drug raw materials that have been degraded under stressed conditions. In some cases, impurities may be commercially available. Information on the chemical and physical properties of impurities is helpful in selecting the technique and type of method for detection and quantification. Some of the properties examined include: solubility, stability, absorbance in the ultraviolet (UV) or visible range, pH, and melting and boiling points.

The performance of a method is expressed in terms of its specificity, sensitivity, accuracy, precision, detection and quantification limits, linearity, range, robustness, and ruggedness. In addition to the parameters listed above, other factors to be considered are simplicity and cost. The latter may be of less importance for methods used at the product development stage or for regulatory analysis but may be of considerable importance for routine quality control methods.

During method development, the analyst must try to build safeguards into the method to ensure reproducibility. This can be achieved, for example, by using well-defined reagents or columns for chromatography, using suitably qualified equipment, and designing a proper system suitability test. System suitability tests are usually included in all pharmacopoeial LC methods. They may include requirements for the resolution of two compounds, retention times or capacity factors, tailing factor, column efficiency, or precision of the system. They are a measure of the suitability of the entire system, hardware, electronics, and solvents, to carry out the analysis. An analysis should be carried out only when the system suitability requirements can be met. It provides a means of ensuring good results without having to validate the method again.

### Method Validation

Validation is the process by which a method is tested and shown to be effective for its intended purpose. The type of method will determine which analytical parameters are to be evaluated and how these can be measured. The definitions provided below correspond to those contained in the ICH guidelines on method validation.

Specificity, according to the ICH definition, is the ability of the method to measure the analyte



accurately in the presence of other compounds expected to be present in the sample matrix. This corresponds to the International Union of Pure and Applied Chemistry (IUPAC) definition of selectivity; according to the IUPAC, specificity is the ultimate in selectivity, meaning that no interferences are supposed to occur. For related substances methods, the individual impurities must all produce a response within a reasonable time frame and be resolved from the drug and the solvent front. For nontoxic impurities, it is preferable, though not absolutely necessary, that they be resolved from each other. In order to be able to demonstrate the specificity of a method, the analyst normally requires standards of each of the possible related substances. Should a new impurity become available after a method is developed, it would be necessary to revalidate the method by showing that this impurity is resolved from the drug and can be suitably quantified.

Accuracy is the closeness of a measured value to its true value. Although it may not always be possible to determine the accuracy of a method, results from different types of methods can sometimes be compared. When making these comparisons, it is important to keep in mind the limitations of the methods; for example, some methods are less specific than others and others involve the use of standards that are assumed to assay at 100.0%. Accuracy can be estimated to some extent in the case of formulations: known amounts of impurities can be added to a placebo and the percentage of each impurity recovered determined.

The detection limit is the smallest amount of analyte that can be detected but not necessarily quantitated as an exact value in a sample. For instrumental methods, the investigator may compare test results from blanks and samples of increasingly low concentrations and establish at which concentration the analyte signal is distinguishable above the noise. Usually a signal-to-noise ratio of 2:1 or 3:1 is acceptable.

The quantitation limit is the lowest amount of analyte in a sample that can be measured with a suitable degree of precision or accuracy. Several approaches for determining it are acceptable. Many investigators determine the concentration of analyte where the signal-to-noise ratio is 10:1; others may define it as the concentration at which the relative standard deviation (RSD) on a specified number of replicates is of a certain magnitude, for example, 10% for 5 replicates.

Precision expresses the degree of agreement between individual test results on the same sample. To determine the precision, a number of weighings of a particular sample are analyzed and the mean, standard

deviation, and RSD of the results are determined. Repeatability or intra-assay precision expresses the precision under the same operating conditions over a short interval of time. This is done by determining the mean response and RSD of replicate injections of a standard solution to obtain an estimate of the error that is attributable to the operating system only, excluding sample preparation. Generally, lower precision is required for related substances methods than for assay methods. Intermediate precision expresses within-laboratory variations such as different days, different analysts, different equipment.

The robustness of a method is a measure of its capacity to remain unaffected by small but deliberate variations in method parameters and provides an indication of its reliability during normal usage. It can be demonstrated by showing that operational or environmental influences have no significant effect on test results.

Ruggedness is the ability of the method to produce similar results for a given sample when it is used in a variety of normal test conditions, such as different equipment, different analysts, different laboratories, or different days. Having the method evaluated by another analyst, preferably in another laboratory, provides a means to evaluate the ruggedness of the method. Such an external evaluation may reveal that the instructions are ambiguous or incomplete, or that differences in results can arise due to the use of different brands of equipment or chemicals. Collaborative studies, such as those organized by AOAC International, in which a given set of samples is analyzed by a number of laboratories, are probably the best way of demonstrating the ruggedness and reproducibility of a method, but are seldom done for drug standard methods.

Linearity is the ability of the method to provide a measurable response to the analyte that is directly, or by a well-defined mathematical transformation, proportional to its concentration over a given range. It is usually determined from regression analysis of a plot of response versus concentration. The variance of the slope of the regression line, calculated by the method of least squares, provides a measure of linearity. Linearity is usually defined over a given range, e.g., for a related substances method, it is usually necessary to demonstrate linearity from the minimum quantifiable amount to an amount exceeding the proposed limit on individual impurities.

Sensitivity, though not defined in the ICH Validation Guideline, is a useful parameter to evaluate during method validation. It is defined as the slope of the analytical response curve. It is the difference in response for a given difference in concentration. When several compounds are quantified against a

single standard, it is necessary to show whether the system is more or less sensitive to these compounds, relative to the standard.

In order to develop monographs for new drugs the EP, USP, and JP require validation data and analytical results for production batches of drug substance or drug product that are consistent with ICH guidelines. The EP has adopted a policy of transparency with respect to validation data; this means that key aspects of the validation may be published or made available upon request. Key data may include the brands of column packing materials that have been evaluated or the names of the related compounds for which the method has been validated. Such a policy allows users to better understand the scope and limitations of particular methods.

### Techniques

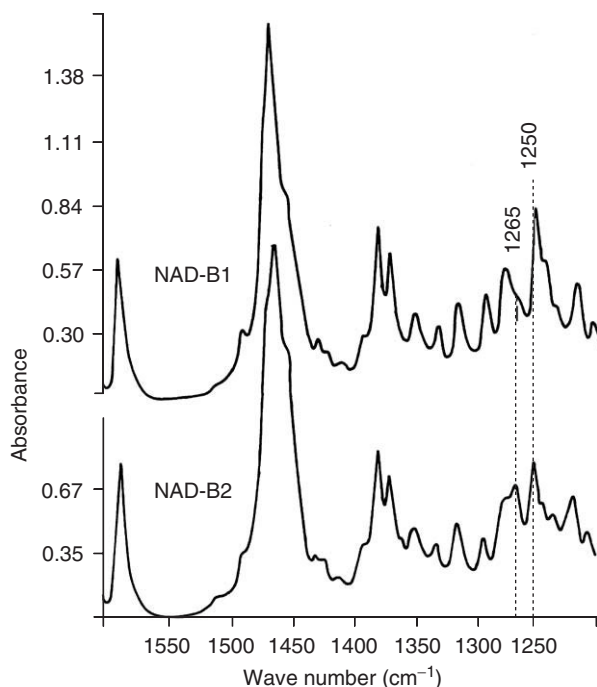
Analytical techniques are based on a particular chemical or physical property. For example, spectrophotometry may be used for compounds with a UV chromophore or it may be possible to devise a titrimetric method based on the acid-base properties of the analyte. The solubility and stability of the drug and impurities in various media and under various conditions may also play a role in the choice of technique.

In 1990, chromatographic methods accounted for ~95% of the tests used to assess purity in USP monographs. Spectrophotometric methods accounted for less than 4% of methods. The remaining 1% included titrimetric, nuclear magnetic resonance, colorimetric, and other methods.

Spectroscopic methods include both UV and infrared (IR) methods. Both the BP and USP contain a UV method for adrenalone in adrenaline (epinephrine). Other drugs that are produced from the reduction of aromatic ketones or undergo oxidation to aromatic ketones have analogous tests; these include isoprenaline (isoproterenol) and metaproterenol (orciprenaline).

IR spectroscopy has been applied to determining the presence of an excess of one racemate in nadolol, defined as a 1:1 mixture of racemate A and racemate B. The USP test is based on the fact that a 1:1 mixture has a value of 0.9 for the ratio of the absorbance at  $7.9\ \mu\text{m}$  ( $1265\ \text{cm}^{-1}$ ), corresponding to racemate A, over that at  $8.00\ \mu\text{m}$  ( $1250\ \text{cm}^{-1}$ ), corresponding to racemate B. The USP limit for racemate composition is 40–60% racemate A. **Figure 1** shows the partial IR spectra of two nadolol samples from the same manufacturer. Samples B1 and B2 contained ~30% and 48% racemate A, respectively.

Melting point determination is a test that is included in most pharmacopoeias. Other thermal



**Figure 1** Infrared absorbance spectra obtained for the evaluation of racemate composition of two samples of nadolol from the same manufacturer. Sample NAD-B1 contained 30% racemate A; sample NAD-B2, 48% (USP limits 40–60% racemate A).

analysis methods such as differential scanning calorimetry and thermogravimetric analysis, although useful for purity determination and polymorphism studies, are not used on a routine basis.

Polarimetry can be used to monitor the presence of an unwanted enantiomer in some drugs. However, the determination of optical rotation is not a very specific test; in addition to the unwanted enantiomer, other chiral drug related impurities can affect the results.

As indicated above, chromatography is almost invariably used for impurity determination. Thus, many pharmacopoeial methods employ TLC to resolve the impurities from the drug with subsequent detection under visible or UV irradiation, or reaction to produce colored spots. Tests are based on a comparison of the intensity of impurity spots with that produced by a drug standard at a concentration corresponding to the limit for individual impurities specified in the monograph. A densitometric evaluation of the chromatogram may provide a more quantitative result than visual inspection. In some cases, impurity spots are scraped from the plate, extracted and quantified using a spectrophotometer. The main advantages of TLC are its simplicity, low cost, high resolving power, and relatively short analysis times.



LC provides good reproducibility and high-quality results; the cost is reasonable for routine applications and it is fairly easy to operate the equipment. UV-visible detectors are most commonly used for routine impurity determinations by LC. Other detectors that are useful for the identification of impurities or for specialized purposes are refractive index, fluorescence, electrothermal detectors, polarimeters, Fourier transform infrared spectrometers, and mass spectrometers.

The resolving power and sensitivity of LC are often greater than TLC. However, quantification is not possible unless the compounds are eluted from the column whereas in TLC all the components are on the plate and can be measured, if they are resolved, even if they do not move from the origin.

Figure 2 shows two samples of verapamil hydrochloride, VER-B1 and VER-D, analyzed using an LC method for chromatographic purity. This example illustrates a difference in impurity profile for lots originating from different manufacturers. Each sample contained less than 0.5% impurities.

GC is sometimes used for drug-related compounds and is the technique of choice for the determination of residual organic solvents in drug raw materials. Packed GC columns and wide-bore wall-coated open tubular columns are used in pharmacopoeial methods. Thermal

conductivity, flame ionization, and electron capture are the types of detectors most commonly used.

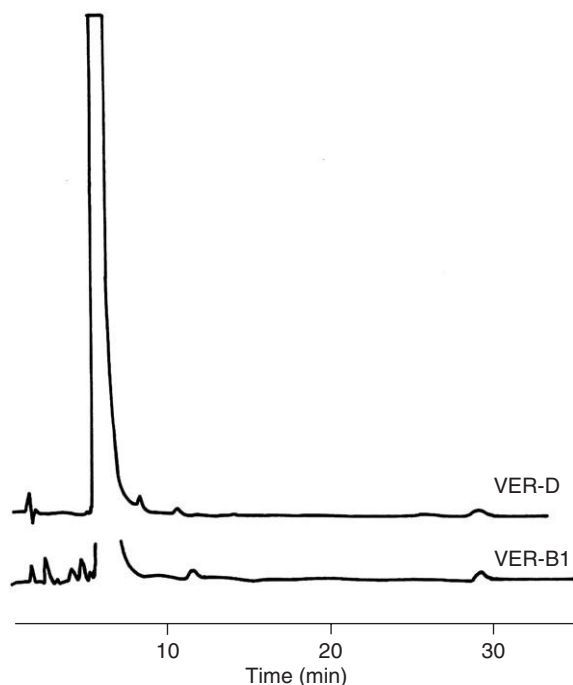
For residual organic solvents, the USP sets limits of 380  $\mu\text{g per g}$  for 1,4-dioxane, 600  $\mu\text{g per g}$  for methylene chloride, 80  $\mu\text{g per g}$  for trichloroethylene, and 60  $\mu\text{g per g}$  for chloroform. When warranted, some monographs contain an additional test for ethylene oxide for which the limit is 10  $\mu\text{g per g}$ . Several methods are described in USP general chapter (no. 467, *Organic Volatile Impurities*) and the one to be used for a particular drug substance is specified in the individual monograph. These methods are based on direct injection or dynamic headspace techniques.

LC has several advantages over GC: separations are usually carried out at ambient temperature and therefore thermal degradation is less likely to occur, derivatization is not normally required, and sample preparation is usually simpler, even for formulations. There are, however, some disadvantages when compared to GC: the resolving power of LC is generally lower, particularly when compared to capillary GC; and there is the problem of solvent waste disposal.

## Emerging Techniques in Drug Purity Determination

An increase in the variety of commercially available GC and LC columns has spawned an interest in developing methods to directly quantify optical isomers present as impurities in chiral drugs. These methods are viewed as generally more specific than tests for specific rotation and may soon be replacing such tests in pharmacopoeias.

Biotechnologically derived pharmaceuticals such as proteins from recombinant DNA (r-DNA) techniques or monoclonal antibodies pose new challenges. The absence of DNA at the picogram per dose level in r-DNA derived proteins must be demonstrated because such DNA could be incorporated into the human genome and become a potential oncogene. Host cell or host cell components (e.g., proteins, viruses) and substances used during production or purification (e.g., media, antibiotics, and reagents) are examples of potential impurities in these products. In addition, the organisms used during production must be monitored for mutations, which could result in new undesirable impurities. Well-established pharmacopoeial procedures suitable for detecting impurities in biotechnologically derived products include LC, the limulus amoebocyte lysate test for bacterial endotoxins, the rabbit pyrogen test, and gel electrophoresis. Newer techniques include peptide mapping, isoelectric focusing, capillary gel electrophoresis, and immunoassays.



**Figure 2** Chromatograms for two lots of verapamil hydrochloride analyzed by LC. Column:  $150 \times 4.6$  mm i.d.  $3 \mu\text{m}$  Spherisorb ODS-2; mobile phase: acetonitrile,  $0.01 \text{ mol l}^{-1}$  sodium acetate containing  $33 \text{ ml l}^{-1}$  glacial acetic acid, 2-aminoheptane (45:55:0.5 v/v/v); flow rate  $0.9 \text{ ml min}^{-1}$ ; UV detector 278 nm.

See also: **Chiroptical Analysis**. **Gas Chromatography**: Detectors. **Infrared Spectroscopy**: Overview. **Liquid Chromatography**: Instrumentation; Pharmaceutical Applications. **Thermal Analysis**: Overview; Temperature-Modulated Techniques. **Thin-Layer Chromatography**: Overview.

## Further Reading

- Ahuja S (1998) *Impurity Evaluation of Pharmaceuticals*. New York: Dekker.
- British Pharmacopoeia Commission Office (2002) *British Pharmacopoeia 2002*. London: The Stationery Office.
- Carr GP and Wahlich JC (1990) A practical approach to method validation in pharmaceutical analysis. *Journal of Pharmaceutical and Biomedical Analysis* 8: 613–618.
- Council of Europe (2002) *European Pharmacopoeia*, 4th edn. Strasbourg, France: EDQM.
- Grieg DGT (1980) Chromatography and quality assurance in the pharmaceutical industry. In Knapman CEH (ed.) *Developments in Chromatography*, vol. II. London: Applied Science.
- Görög S (ed.) (2000) *Identification and Determination of Impurities in Drugs*. Amsterdam: Elsevier.
- Hubert BB (1990) Rational limits for drug impurities. *Drug Safety* 5(Supplement 1): 88–94.
- International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. See *ICH Harmonized Tripartite Guidelines* at <http://www.ich.org/ich5q.html>
- Kirshbaum JJ (1989) Inter-laboratory transfer of HPLC methods: problems and solutions. *Journal of Pharmaceutical and Biomedical Analysis* 7: 813–833.
- Martin-Smith M and Rudd DR (1990) The importance of proper validation of the analytical methods employed in the quality control of pharmaceuticals. *Acta Pharmaceutica Yugoslavia* 40: 7–19.
- Phillips GF (1989) The need for impurity control: an overview. *Pharmaceutical Industry* 51: 1282–1286.
- The Society of Japanese Pharmacopoeia (1996) *The Japanese Pharmacopoeia*, 13th edn. Tokyo: The Society of Japanese Pharmacopoeia.
- United States Pharmacopoeial Convention (2003) *United States Pharmacopoeia 26/National Formulary 21*. Rockville, MD: USPC.
- Wilson TD (1990) Liquid chromatographic methods validation for pharmaceutical products. *Journal of Pharmaceutical and Biomedical Analysis* 8: 389–400.
- World Health Organization (1979) *The International Pharmacopoeia*, 3rd edn. Geneva: World Health Organization.

## Stability Testing

**T Coomber**, Pharmaceutical Development Consultant, Gravesend, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A key component in a pharmaceutical development program for a new drug or formulation is the stability testing of the drug substance often termed active pharmaceutical ingredient (API) and the dosage forms used throughout the development of the marketed product or products. Stability tests have to be carried out so that recommended storage conditions and shelf life can be included on the label to ensure that the medicine is safe and effective throughout its shelf life.

The detailed analytical science behind many of the tools used in this activity is covered in various sections of this volume. No attempt has been made to go into the details of the methods for the identification of impurities or quantitative analysis of the main component and impurities. The exciting aspect for many involved with this process is to bring the various scientific threads together in the practical

determination of the storage characteristics of drugs and their formulations, the process of doing this in the limited time frame of the development of a new product, and the challenge of providing and influencing the regulatory information that has to be submitted.

Understanding the science behind the degradation of the drug substance in the solid state and more practically in solution enables a suite of tests to be chosen to monitor the crucial stability characteristics of the drug substance and gives a guide to the testing that should be carried out on the dosage forms.

In order to underwrite a shelf life, the stability of the dosage form should be understood in terms of appearance, degradation, and physical stability. Changes in appearance give rise to an instant query from the patient: ‘the medicine has changed, is this alright?’ Increased levels of impurities due to degradation have to be strictly controlled because of possible toxicity effects. Degradation also leads to a loss of active ingredients that will decrease the effectiveness of the treatment. The impact of physical change needs to be understood. For instance, moisture loss from a solution through a semipermeable container could lead to overdosing of the drug.

See also: **Chiroptical Analysis**. **Gas Chromatography**: Detectors. **Infrared Spectroscopy**: Overview. **Liquid Chromatography**: Instrumentation; Pharmaceutical Applications. **Thermal Analysis**: Overview; Temperature-Modulated Techniques. **Thin-Layer Chromatography**: Overview.

## Further Reading

- Ahuja S (1998) *Impurity Evaluation of Pharmaceuticals*. New York: Dekker.
- British Pharmacopoeia Commission Office (2002) *British Pharmacopoeia 2002*. London: The Stationery Office.
- Carr GP and Wahlich JC (1990) A practical approach to method validation in pharmaceutical analysis. *Journal of Pharmaceutical and Biomedical Analysis* 8: 613–618.
- Council of Europe (2002) *European Pharmacopoeia*, 4th edn. Strasbourg, France: EDQM.
- Grieg DGT (1980) Chromatography and quality assurance in the pharmaceutical industry. In Knapman CEH (ed.) *Developments in Chromatography*, vol. II. London: Applied Science.
- Görög S (ed.) (2000) *Identification and Determination of Impurities in Drugs*. Amsterdam: Elsevier.
- Hubert BB (1990) Rational limits for drug impurities. *Drug Safety* 5(Supplement 1): 88–94.
- International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. See *ICH Harmonized Tripartite Guidelines* at <http://www.ich.org/ich5q.html>
- Kirshbaum JJ (1989) Inter-laboratory transfer of HPLC methods: problems and solutions. *Journal of Pharmaceutical and Biomedical Analysis* 7: 813–833.
- Martin-Smith M and Rudd DR (1990) The importance of proper validation of the analytical methods employed in the quality control of pharmaceuticals. *Acta Pharmaceutica Yugoslavia* 40: 7–19.
- Phillips GF (1989) The need for impurity control: an overview. *Pharmaceutical Industry* 51: 1282–1286.
- The Society of Japanese Pharmacopoeia (1996) *The Japanese Pharmacopoeia*, 13th edn. Tokyo: The Society of Japanese Pharmacopoeia.
- United States Pharmacopoeial Convention (2003) *United States Pharmacopoeia 26/National Formulary 21*. Rockville, MD: USPC.
- Wilson TD (1990) Liquid chromatographic methods validation for pharmaceutical products. *Journal of Pharmaceutical and Biomedical Analysis* 8: 389–400.
- World Health Organization (1979) *The International Pharmacopoeia*, 3rd edn. Geneva: World Health Organization.

## Stability Testing

**T Coomber**, Pharmaceutical Development Consultant, Gravesend, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A key component in a pharmaceutical development program for a new drug or formulation is the stability testing of the drug substance often termed active pharmaceutical ingredient (API) and the dosage forms used throughout the development of the marketed product or products. Stability tests have to be carried out so that recommended storage conditions and shelf life can be included on the label to ensure that the medicine is safe and effective throughout its shelf life.

The detailed analytical science behind many of the tools used in this activity is covered in various sections of this volume. No attempt has been made to go into the details of the methods for the identification of impurities or quantitative analysis of the main component and impurities. The exciting aspect for many involved with this process is to bring the various scientific threads together in the practical

determination of the storage characteristics of drugs and their formulations, the process of doing this in the limited time frame of the development of a new product, and the challenge of providing and influencing the regulatory information that has to be submitted.

Understanding the science behind the degradation of the drug substance in the solid state and more practically in solution enables a suite of tests to be chosen to monitor the crucial stability characteristics of the drug substance and gives a guide to the testing that should be carried out on the dosage forms.

In order to underwrite a shelf life, the stability of the dosage form should be understood in terms of appearance, degradation, and physical stability. Changes in appearance give rise to an instant query from the patient: ‘the medicine has changed, is this alright?’ Increased levels of impurities due to degradation have to be strictly controlled because of possible toxicity effects. Degradation also leads to a loss of active ingredients that will decrease the effectiveness of the treatment. The impact of physical change needs to be understood. For instance, moisture loss from a solution through a semipermeable container could lead to overdosing of the drug.

The 'in use' stability of the drug product is also tested. This is highly dependent on the product, but once the seal has been broken on a tablet container or a freeze-dried powder has been reconstituted then the stability of each under the less controllable conditions of a patient's home or a clinical setting should be established over the recommended period of dosing.

This scientific activity is controlled by government regulations that define the amount and scope of testing and the quality of the data that has to be provided in all major markets of the world. There are also guidelines on testing that have been produced by an umbrella organization called 'The International Committee for Harmonisation (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use' and agreed between the authorities in Europe, Japan, and the United States and the manufacturer's organizations in those areas. Input from the World Health Organization (WHO) has also been included. These guidelines are now well accepted as an appropriate level of testing worldwide.

This testing is carried out as part of the development process of a new drug product. So the work has to be managed in the overall timescale and process of development and final data package must meet requirements on data quality and content.

## Drug Substance

For any new drug, the potential and, if possible, actual degradation processes need to be understood in order to develop appropriate analytical methods. Perversely, from a technical viewpoint, it is more straightforward if the drug degrades to a small extent. The mechanism also needs to be understood in order to develop a product with optimum stability (this may be a compromise between shelf life and storage conditions). The identification of degradation products and the monitoring of their formation and the level of the drug substance are, therefore, key components of the program. For drugs having isomeric forms, the likelihood of those forms interconverting needs to be understood as the activity and the toxicity of the forms may be different and they are viewed as different substances.

The understanding of the physical nature of the drug substance in terms of crystal form can also be crucial. This understanding will enable changes in moisture content to be avoided to enable accurate measurements of the drug substance throughout its use and to avoid having to take onerous packaging precautions. Similarly, especially for poorly soluble drugs, the most stable crystal form should be chosen as efficacy may change if the crystal form changes.

The potential for degradation should be considered in terms of the effect of heat, light, and oxidation in the solid state and also in the presence of acids and bases in solution or suspension. In preliminary studies, the drug substance will be stressed, usually at temperatures of 50–80°C in the solid and solution, at extremes of pH (about 1 and 13), and in the presence or absence of oxygen.

Solid-state degradation is not straightforward to model and predict as the first step is some sort of disruption of the crystal that has large activation energy. When this is true, although high-temperature studies may show degradation, the rate will decrease rapidly with decreased temperature, perhaps ~10-fold per 10°C decrease in temperature.

In solution the rate of degradation will often follow regular kinetics. It has been found useful to describe the rate of degradation as the time to 90% drug remaining,  $t_{90\%}$ , rather than use rate constants. Activation energies ranging from 50 to 100 kJ mol<sup>-1</sup> have been reported. With activation energies of this order,  $t_{90\%}$  will increase by a factor of ~2–3.5 for every 10°C rise in temperature. If the reaction undergoes specific acid or base catalysis in solution, the rate will change by a factor of 10 for every pH unit change in the appropriate pH region. This emphasizes the potential importance of adequate pH control. Under the controlled conditions of formulation, the degradation of the major product will often follow first-order degradation kinetics, which means  $t_{90\%}$  will stay constant over a range of concentrations of the drug.

In contrast, photolysis is a zero-order process and the  $t_{90\%}$  of a weaker solution will decrease in proportion to concentration.

Oxidation reactions will be controlled by the amount of oxygen present. A genuine oxidation of the drug is a major problem, but often oxidation reactions that effect formulations are of a secondary nature, i.e., often in conjunction with photolysis or catalyzed by the presence of trace metals so that degradation can be controlled by addressing those factors.

The limits on manufacture and shelf life for degradation products will be set by the knowledge of the amount present in the drug substance used in toxicity testing, by knowledge of the material itself, or by knowledge of compounds of a similar structure. (This is called qualification.) In the absence of such information, ICH sets pragmatic lower limits for unqualified degradation impurities in drug substance and drug product.

Several batches of the drug will be stored under the defined conditions to establish storage conditions and a retest date or shelf life.



## Dosage Form

Prior to dosage form, development compatibility studies with proposed excipients are carried out. This is an extension of the testing carried out on the initial investigation of the drug substance and should build on and expand the knowledge generated. For solid products, these studies are often carried out under high humidity with simple mixtures. A storage condition of 40°C and 75% relative humidity is often used. Solution products are tested at 50–70°C in a narrow range of pH compatible with the drug and the patient.

Some considerations on the testing of individual formulations are given below. However, the following tests are common to all formulations: appearance, strength (usually expressed as percent labeled strength), and related (degradation) impurities. A selection of other tests that are used is given in **Table 1**. Each formulation needs to be considered on its merits and there is also significant detailed advice from the regulatory bodies that should be followed.

### Tablets

Tablets are a mixture of the drug substance and excipients that enable a homogeneous mixture of the drug to be made and then a solid tablet to be compressed from that mixture. Tablets are often film coated to mask taste and/or to give a consistent appearance. Water is often present in the excipients and can be used as part of the manufacturing process in wet granulation, often providing a medium for potential degradation on manufacture and storage.

To obtain a precise assay, the ratio of the drug substance to excipients in the tablets needs to be considered when deciding on the number of tablets taken. There are pharmacopeial standards for content and/or weight uniformity of the tablet, and the variability of the batch will be established using this

test. In general, a sample of 5–10 tablets needs to be taken to obtain sufficient accuracy. The sample preparation is best achieved by using whole tablets. It is rare for the excipients to dissolve completely, so physical disruption of the tablets in an ultrasonic bath or with a high sheer mixer is often employed. A filtration or centrifugation step may be required, although it is not always necessary if subsequent dilutions are made. Changes in the tablet on storage (e.g., increased hardness) can lead to difficulties in sample preparation. As high-performance liquid chromatography (HPLC) analysis is often the final step, the final diluent should be chosen to minimize any affect on the chromatography.

Tablets will normally have a dissolution test and this should remain within specification throughout the shelf life of the product. This is particularly important for compounds that are poorly absorbed and of low solubility, where changes in dissolution characteristics may indicate potential delivery problems.

### Capsules

A capsule is made from either a hard or soft shell using gelatin or other materials that will dissolve rapidly in the stomach. Hard capsules are usually filled with a mixture of the drug substance and an inert bulking agent such as lactose. Soft capsules are filled with an oily solution or suspension. Sample preparation can either be by opening the capsules and transferring the contents or by using gentle warming in an aqueous medium to dissolve the shell. The impact of moisture on the contents should be considered as the shell will transmit water. Hardening of the shell can also occur on storage, which can effect the sample preparation.

### Liquids

Liquid formulations will have defined pH, and testing will be carried out on formulations manufactured

**Table 1** Summary of other tests by product

<i>Product</i>	<i>Test requirements</i>
Tablets	Color, odor, dissolution (disintegration), moisture, friability
Capsules	Color, odor of contents, dissolution, moisture, microbial limits
Emulsions	pH, color, odor, viscosity, microbial limits, preservative content, dispersed phase sizing
Oral solutions and suspensions	pH, color, odor, preservative content, microbial limits, suspension characteristics
Metered dose inhalers	Delivered dose, number of doses, particle size distribution
Topical ophthalmic and otic preparations	pH, color, odor, homogeneity, preservative content, antioxidant content
Parenterals	pH, color, particulate matter sterility and pyrogenicity, preservative content, other tests as in type of formulation

This list is indicative of the types of test and is not complete. The appropriate regulations should be consulted and applied to the particular formulation and justified by the science.

at the target pH and at the expected limits. They will also, in general, include preservatives to prevent microbial contamination. The stability of these preservatives will need to be monitored and they will often be determined by HPLC simultaneously with the drug content. It will need to be confirmed that the product is adequately preserved at the end of the shelf life.

### Oral Solutions and Suspensions

These may contain levels of sugars or artificial sweeteners, coloring and flavoring agents, which make the HPLC determination more complex. Many of these components will be detected by HPLC and may interfere with the components that are being determined. The container should be stored such that the contents are in contact with the closure system.

### Parenterals (Injections)

These include solutions, suspensions, and powders for reconstitution and are often produced as sterile single-dose or preserved multiple-dose products. In terms of testing the general considerations for liquids apply. Particle size monitoring of suspensions should be carried out. Interactions with any closure system should be tested. Attention also has to be paid to sterility testing during the stability study.

### Topical Creams and Ointments

These cover a wide range of formulation types from ointments to nonmetered aerosols. The physical status of an emulsion can be critical. This is tested by temperature cycling. Where appropriate the homogeneity of the content is also monitored by sampling the top, middle, and bottom of the container.

### Metered Dose Inhalers

As well as the chemical testing, inhalers require specific characteristics to be determined. Dose delivery and number of doses and particle size, which is crucial for delivery to the lung, should be determined. The testing for inhalers is a labor-intensive activity. The robustness of the delivery, number of doses and particle size, system, the container and valve seal, and propellant losses are all determined.

## Analytical Considerations

The common tests for stability testing are appearance, strength (content), and degradation impurities. In order to monitor stability, a quantitative test is best as it will demonstrate a trend. Thus, a description of color, for instance, may be best made by an instrumental approach. Pharmacopeial requirements

for quantitative tests such as HPLC should be followed.

For the determination of strength and degradation, the primary tool for the analytical chemist involved in stability testing is HPLC with ultraviolet (UV) detection.

In the early stages of development, this is used firstly as a separation tool to establish what degradation products are produced. In conjunction with a mass detector, this will often enable the identity of the impurity to be proposed. Confirmation of structure is often achieved by the use of nuclear magnetic resonance spectroscopy and mass spectroscopy or by synthesis.

In routine storage tests, HPLC will also be used to monitor the formation of degradation impurities in both the drug substance and the formulated product, and generally to determine the amount of the active material present. HPLC is not the only tool that can be used. Capillary electrophoresis has great separating power and is often useful in separating optical isomers. The choice of method is a question of understanding the scientific and practical scope of the different tools that are available. UV spectroscopy can be used to measure the active material in a product in dissolution and homogeneity studies, for instance.

A method of analysis for degradation (and synthetic) impurities may need to have a limit of quantitation of  $\sim 0.03\%$  as this is the lowest reporting threshold required. The lowest limit for unqualified impurities can be  $0.05\%$ . In the analysis of products, the limit for unqualified impurities ranges from  $0.05$  to  $1.0\%$  depending on the dose. During routine testing, a combination of a suitability test using a mixture of the drug substance and a critical (from a separation standpoint) impurity and relative retention time will be used to identify a degradation product. Usually, this determination will be carried out using the drug substance as a quantitative reference material and an experimentally measured response factor. Responses between  $0.8$  and  $1.2$  can be assumed to be  $1.0$  for practical purposes. These determinations are well within the scope of current automated HPLC with an integration system.

In drug product analysis, the amount of the drug substance remaining is commonly determined simultaneously with that of impurities using an external standard of a drug substance reference compound produced as part of the development process for new drugs or from a Pharmacopoeia, when available, for established drug substances. Using automated injection and data handling systems, an accurate and precise assay can be achieved. Simultaneous determinations of active impurities provide the challenge



of quantification over a large dynamic range, but this is within the scope of the art and should be carefully considered during method development.

The most critical stage of any study will be the appearance of the primary degradation product and this will usually be in excess of the limit well before there is a detectable change in the main component assay. The assay is, however, important to ensure there is mass balance (i.e., we are not losing drug through an unexpected degradation) and to demonstrate that there have been other changes in the formulation (i.e., loss of water or other solvent through a semipermeable container leading to an excess dose being delivered) or loss of the active component to the walls of a container.

## Process and Timing

The development of a new drug runs through a series of four phases. The determination of stability characteristics is an ongoing activity during that process.

Phase I is taken to cover the development of the drug from the proof of concept (i.e., some information in a disease model that the compound is active and worth the investment in developing) through acute toxicity studies until it is dosed to human volunteers. Stability data are generated to ensure the model has actually been challenged by the drug and not a degradation product. The requirements of good laboratory practice (GLP) require that the formulation used in animal studies is adequately stable for the duration of use of the formulation. Similarly, the stability of the formulation used in the Phase I clinical studies needs to be underwritten. This assurance will be required by the investigators, the ethics committee of the organization carrying out the study, and in some countries the regulatory body of the country concerned.

Phase II studies are those carried out in the patient population to establish whether the drug is active in the patient and also establish the optimum dose. Longer-term toxicity testing will also be carried out. As before, GLP will require the formulations used for toxicity testing to have their stability assessed. The clinical formulations will also have to be tested and will usually be made under the control of good manufacturing practice (GMP) regulations. These data are required to be generated to a protocol that covers the expected shelf life of the formulations. The regulatory bodies will require protocols and/or data summaries and an ongoing commitment to testing. The conclusion from 3 months data is a common minimum requirement in a submission for approval to carry out a clinical trial.

Phase III studies are those carried out on the formulation proposed for marketing. This is where the

testing requirements expand in terms of batches tested and amount of data and where the ICH requirements operate. One year data on three batches of defined sizes in comparison to production scale are required for submission in the dossier for approval for marketing in the EEC (Marketing Authorization Application, MAA) or in the USA (New Drug Application, NDA).

Phase IV studies are those carried out after the drug is first released and will include additional formulations that have to meet the same requirements as phase III. There is also an ongoing commitment to complete the studies commenced in phase III until the end of shelf life, and a requirement for ongoing testing of a subset of production batches of all marketed formulations under GMP regulations.

## Regulation and Guidelines

The stability testing of pharmaceuticals is regulated by national and international agencies. Four major bodies, the Food and Drug Administration (FDA) (through the Centre for Drug Evaluation Research, CDER), the European Agency for the Evaluation of Medicinal Products (EMA) (via the Committee for Proprietary Medicinal Products (CPMP)), Ministry of Health, Labor and Welfare (MHLW), Japan, and the WHO, publish requirements on the storage conditions, testing times, and amount of data needed to establish the shelf life required for drug products to be licensed for use. For new products, these stability testing requirements have been the subject to over 10 years of worldwide harmonization through the ICH process and are now available from individual registration bodies and from the ICH. In particular, the storage temperature and humidity conditions required to cover storage in normal climatic conditions have been defined for four different climate areas covering most of the world's population.

The data, testing requirements, and standards are laid out by each body and include extra guidance in addition to that provided by the ICH for new products and guidance on existing products that are not formally covered by the ICH process.

Laboratory standards required are also controlled by quality programs. In particular, the stability testing of materials for safety testing in animals is controlled by the GLP regulations. The testing of materials for registration of final products is performed under the GMP regulations that apply to materials made in full-scale production.

ICH stability testing guidelines are available on the Internet, see **Table 2**, and there is a process by which they are introduced eventually into official status in the USA, EEC, and Japan.

**Table 2** ICH guidelines on stability testing

Number	Title
Q1A(R2)	Stability Testing of New Drugs and Products (Revised Guideline)
Q1B	Photostability Testing
Q1C	Stability Testing for New Dosage Forms
Q1D	Bracketing and Matrixing Designs for Stability Testing of Drug Substances and Drug Products
Q1E	Evaluation for Stability Data
Q1F	Stability Data Package for Registration Applications in Climatic Zones III and IV
Q2A	Text on Validations of Analytical Procedures
Q2B	Methodology
Q3A(R)	Impurities in New Drug Substances (Revised Guideline)
Q3B(R)	Impurities in New Drug Products (Revised Guideline)
Q5C	Stability of Products (Biotechnology Products)

Q1A gives recommendations on the stability testing protocols that should be followed to assess the stability of new drug substances and products. Recommendations are given on temperature and humidity levels as well as the duration of trials. Q1B forms an annex to the main stability guideline, and gives guidance on the basic testing protocol required to evaluate the light sensitivity and stability of new drugs and products. Q1C extends the main stability guideline and defines the circumstances under which reduced stability data can be accepted, at the time of filing an application, in the case of new formulations of already approved medicines. Q1D gives guidance on specific experimental designs to optimize the amount of testing. Q1E amplifies the guidance on extrapolating data for shelf life determinations, and Q1F amplifies the requirements for the more extreme climate zones.

All analytical methods used should be validated appropriately as laid out by the regulatory bodies and ICH guidelines available in Q2. The guidelines on impurities in Q3 are also relevant to setting the limits for degradation impurities. The Q5 guidelines give information on the testing of biotechnology products.

There are additional guidelines from the CPMP in Europe (see Table 3).

The FDA has also issued Draft Guidance on stability testing of drug substances and drug products, which includes and extends the scope of the ICH Guidelines and includes guidance on phases I and II and testing for postapproval changes.

These documents are called guidelines although they eventually have legal status in the USA. The principles behind them should be considered and applied where the guideline is not directly applicable

**Table 3** CPMP additional guidelines for stability testing

Number	Title
CPMP/QWP/609/96	Note for guidance on Declaration of Storage Conditions for Medicinal Products Particulars and Active Substances
CPMP/QWP/122/02	Note for guidance on Stability Testing of Existing Active Substances and Related Finished Products
CPMP/QWP/072/96	Note for guidance on Start of Shelf-Life of the Finished Dosage Form
CPMP/QWP/2934/99	Note for guidance for In-Use Stability Testing of Human Medicinal Products
CPMP/QWP/576/96	Note for guidance on Stability Testing for a Type II Variation to a Marketing Authorization
CPMP/QWP/556/96	Note for Guidance on Stability Testing of Existing Active Substances and Related Finished Products
CPMP/QWP/159/96	Note for guidance on Maximum Shelf-Life for Sterile Products after First Opening or following Reconstitution

and the reasons for a departure explained by a careful scientific rationale.

## The Future

The critical area of stability testing is the amount of data generated and the number of tests required. There will be a continued drive to reduce this where patient safety is not compromised.

The ICH process led to a commonality of requirements in the major markets. The ICH do have a proven mechanism to introduce change into guidelines and the FDA through their recent initiative, A Risk-Based Approach to Pharmaceutical GMP, are looking to embrace changes in testing procedures and approaches. However, the regulatory process is perceived as a barrier to change and industry tends to take a conservative approach to minimize the very expensive risk of the rejection of a data package.

In the early stages of development, there are opportunities to predict routes and rates of degradation. There are a number of published studies of degradation routes, so simple prediction for new molecules is possible and an intelligent database could be constructed to assist that process. A properly conducted scientific determination of the routes of degradation would still have to be carried out but this could be more focused and less time consuming. Advances in understanding the degradation behavior of the solid state for both pure substances and mixtures would be welcome.

The kinetics of reaction is a little more difficult to predict but the activation energy of many

degradation reactions in solution is in the order of 50–100 kJ mol<sup>-1</sup>. This implies that the  $t_{90\%}$  will change by a factor of 2–3.5-fold per 10°C change in temperature. Genuine solid-state reactions have far higher activation energy of up to 300 kJ mol<sup>-1</sup>, leaving the rate change of a factor of ~10-fold per 10°C change in temperature. So ball park predictions from a single accelerated temperature determination of rate are possible and an early risk assessment could be made.

The analytical science of stability testing includes the identification of degradation routes and the quantitative assessment of these in a highly regulated environment. Advances in the availability of the identification techniques especially with coupled HPLC determinations will continue to be made and utilized.

In the quantitative field, HPLC analysis and data handling are extensively automated. Sample preparation is often the rate-determining step. Automated sample preparation equipment is available and is of value for large sample numbers, but requires extensive validation studies. The analytical challenge of quantitative mixture determination is adequately met by automated HPLC. *In situ* analysis of whole tablets using near-IR chemometric approaches has been successfully demonstrated although this has not been applied to stability testing. It has also been demonstrated that low levels of an impurity can be determined in drug substances without separation using a chemometric approach. The regulatory environment has been skeptical of the use of chemometrics and the practical technology would require investment in time and development for a marginal short-term return although undoubted long-term rewards. Although the determination of content should be carried out by a stability indicating method, the product will generally go out of specification from the degradation impurities increasing before the assay shows a significant change. There is potential,

therefore, to use faster and less specific methods for the assay when this can be justified scientifically.

The use of experimental design to reduce the amount of testing has been accepted and is enshrined in the ICH guideline on bracketing and matrixing, so there is certainly some scope to further explore the options in this area.

It is expected that the rate of change of approaches will be slow and likely to be incrementally based on current technology.

*See also: Pharmaceutical Analysis: Overview; Drug Purity Determination; Dissolution Testing.*

## Further Reading

- Albini A and Fasani E (1998) *Drugs Photochemistry and Photostability*. London: Royal Society of Chemistry.
- Carstensen JT and Rhodes CT (2000) *Drug Stability Principles and Practices*, 3rd edn. New York: Marcel Dekker.
- Doelker E (1999) Physicochemical behaviour of active substances. Consequences for the feasibility and stability of pharmaceutical forms. *STP Pharma Pratiques* 9(5): 399–409.
- EMA Quality Approved guidelines. In: Guidance Documents – Human Medicines. Available at <http://www.emea.eu.int>
- FDA (1998) Guidance for Industry Stability Testing of Drug Substances and Drug Products Draft Guidance. In: *Guidances Chemistry (Draft)*. Available at <http://www.fda.gov/cder>
- ICH (1996–2003) Quality Topics. Q1: Stability Testing. In: *Publications – Guidelines*. Available at <http://www.ich.org>
- Reynolds DW, Facchine KL, Mullaney, *et al.* (2002) Conducting forced degradation studies. *Pharmaceutical Technology* 26(2): 48–56.
- Wessels P, Holz M, Erni F, Krummen K, and Ogorka J (1997) Statistical evaluation of stability data of pharmaceutical products for specification setting. *Drug Development and Industrial Pharmacy* 23(5): 427–439.

## Dissolution Testing

**S Bateman**, Longfield, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Dissolution testing is a means of monitoring the rate of release of a drug substance from an oral pharmaceutical dosage form, and is widely used within the

pharmaceutical industry during formulation development, for quality control, stability testing, the prediction of bioavailability, and to assist in the determination of bioequivalence.

In early drug development, dissolution testing may yield important information when comparing different formulations. For some drugs, dissolution rate may provide a useful indication of the bioavailability of a preparation, provided a correlation between

degradation reactions in solution is in the order of 50–100 kJ mol<sup>-1</sup>. This implies that the  $t_{90\%}$  will change by a factor of 2–3.5-fold per 10°C change in temperature. Genuine solid-state reactions have far higher activation energy of up to 300 kJ mol<sup>-1</sup>, leaving the rate change of a factor of ~10-fold per 10°C change in temperature. So ball park predictions from a single accelerated temperature determination of rate are possible and an early risk assessment could be made.

The analytical science of stability testing includes the identification of degradation routes and the quantitative assessment of these in a highly regulated environment. Advances in the availability of the identification techniques especially with coupled HPLC determinations will continue to be made and utilized.

In the quantitative field, HPLC analysis and data handling are extensively automated. Sample preparation is often the rate-determining step. Automated sample preparation equipment is available and is of value for large sample numbers, but requires extensive validation studies. The analytical challenge of quantitative mixture determination is adequately met by automated HPLC. *In situ* analysis of whole tablets using near-IR chemometric approaches has been successfully demonstrated although this has not been applied to stability testing. It has also been demonstrated that low levels of an impurity can be determined in drug substances without separation using a chemometric approach. The regulatory environment has been skeptical of the use of chemometrics and the practical technology would require investment in time and development for a marginal short-term return although undoubted long-term rewards. Although the determination of content should be carried out by a stability indicating method, the product will generally go out of specification from the degradation impurities increasing before the assay shows a significant change. There is potential,

therefore, to use faster and less specific methods for the assay when this can be justified scientifically.

The use of experimental design to reduce the amount of testing has been accepted and is enshrined in the ICH guideline on bracketing and matrixing, so there is certainly some scope to further explore the options in this area.

It is expected that the rate of change of approaches will be slow and likely to be incrementally based on current technology.

*See also: Pharmaceutical Analysis: Overview; Drug Purity Determination; Dissolution Testing.*

## Further Reading

- Albini A and Fasani E (1998) *Drugs Photochemistry and Photostability*. London: Royal Society of Chemistry.
- Carstensen JT and Rhodes CT (2000) *Drug Stability Principles and Practices*, 3rd edn. New York: Marcel Dekker.
- Doelker E (1999) Physicochemical behaviour of active substances. Consequences for the feasibility and stability of pharmaceutical forms. *STP Pharma Pratiques* 9(5): 399–409.
- EMA Quality Approved guidelines. In: Guidance Documents – Human Medicines. Available at <http://www.emea.eu.int>
- FDA (1998) Guidance for Industry Stability Testing of Drug Substances and Drug Products Draft Guidance. In: *Guidances Chemistry (Draft)*. Available at <http://www.fda.gov/cder>
- ICH (1996–2003) Quality Topics. Q1: Stability Testing. In: *Publications – Guidelines*. Available at <http://www.ich.org>
- Reynolds DW, Facchine KL, Mullaney, *et al.* (2002) Conducting forced degradation studies. *Pharmaceutical Technology* 26(2): 48–56.
- Wessels P, Holz M, Erni F, Krummen K, and Ogorka J (1997) Statistical evaluation of stability data of pharmaceutical products for specification setting. *Drug Development and Industrial Pharmacy* 23(5): 427–439.

## Dissolution Testing

**S Bateman**, Longfield, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Dissolution testing is a means of monitoring the rate of release of a drug substance from an oral pharmaceutical dosage form, and is widely used within the

pharmaceutical industry during formulation development, for quality control, stability testing, the prediction of bioavailability, and to assist in the determination of bioequivalence.

In early drug development, dissolution testing may yield important information when comparing different formulations. For some drugs, dissolution rate may provide a useful indication of the bioavailability of a preparation, provided a correlation between

*in vitro* dissolution and *in vivo* bioavailability has been established. For example, successful *in vitro/in vivo* (IV/IV) correlation has been established for digoxin tablets containing different particle sizes of the drug substance.

For quality control purposes, dissolution testing is a means of ensuring batch-to-batch consistency and compliance with specifications. Dissolution rate may be able to highlight whether any serious changes or faults have occurred in the formulation, manufacture, or storage of a product.

Increased interest by regulatory authorities has emphasized the importance of dissolution testing, and has led to significant modifications and improvements in equipment. Comparison of dissolution profiles is being widely used in assessing product uniformity, particularly in the regulation of postapproval changes to formulation, site of manufacture, scale of manufacture, and/or the manufacturing process and equipment.

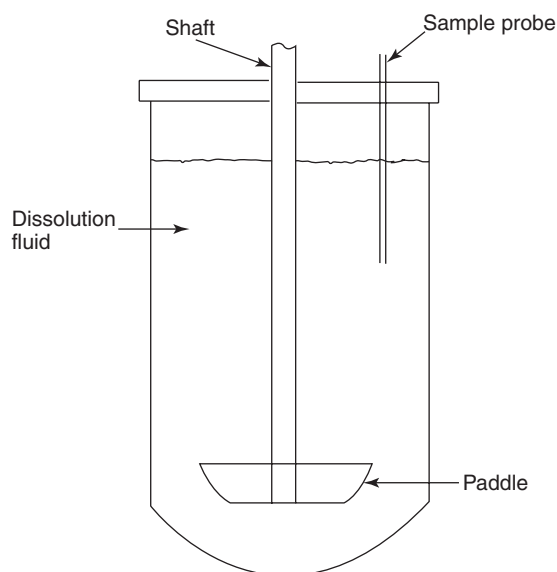
## Apparatus

A comparison of compendial dissolution techniques is given in Table 1.

### Paddle and Basket Apparatus

The United States Pharmacopeia/European Pharmacopoeia/Japanese Pharmacopoeia (USP/EP/JP) rotating paddle apparatus (Figure 1) is recommended for

the vast majority of oral dosage forms. It is the simplest system to operate and fully automated equipment is available from a number of manufacturers. The dosage form is placed directly in the dissolution medium (maintained at 37°C) and the paddle is rotated. Samples of the fluid are removed at predetermined intervals, filtered immediately to remove undissolved drug particles, and analyzed for drug content.



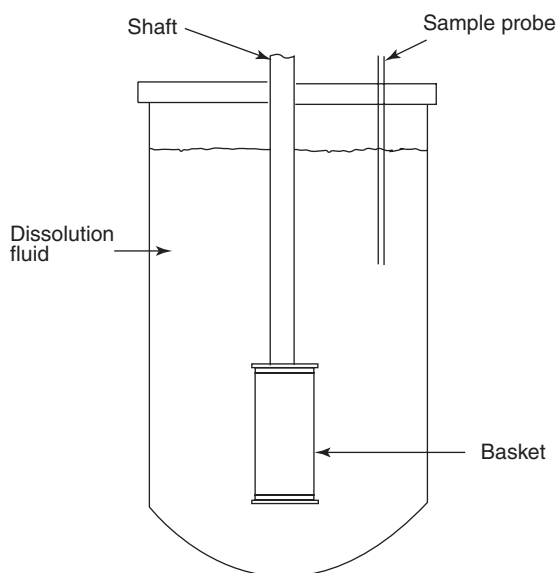
**Figure 1** Rotating paddle apparatus.

**Table 1** Compendial dissolution techniques

Apparatus	Compendial status	Agitation method	Advantages	Disadvantages	Application
Basket	USP 1 EP basket JP method 1 BP I	Rotating stirrer	Dosage form confined leads to consistent solid/liquid interface Floating products kept immersed	Limited sink conditions Trapped bubbles in basket can slow dissolution rate Inadequate mixing at slow speed	Floating dosage forms Dosage forms that tend to adhere to the vessel surfaces
Paddle	USP 2 EP paddle JP method 2 BP II	Rotating stirrer	Simple to use and automate Widely available	Limited sink conditions Sinkers needed for floating dosage forms Coning <sup>a</sup> can occur with large dosage forms	Majority of solid dosage forms Suspensions
Reciprocating cylinder	USP 3	Reciprocating	pH changes are easy Easy to automate	Carryover and incomplete drainage between rows Usually only applicable to nondisintegrating dosage forms	Delayed/extended release products Bead products
Flow-through	USP 4 EP flow-through JP method 3 BP III	Continuous flow	Sink conditions maintained for low solubility drugs pH changes are easy	Large volumes of dissolution medium required Tendency of filters to clog Thorough deaeration of medium necessary	Low solubility drug products Extended release products

<sup>a</sup> The accumulation of material at the bottom of the vessel.





**Figure 2** Rotating basket apparatus.

For some products the USP/EP/JP rotating basket apparatus (Figure 2) may be more appropriate; for example, in the case of floating capsules and granules. Operation of the basket apparatus is similar to that of the paddle apparatus, except that the basket containing the dosage form is immersed in the dissolution medium prior to starting the rotation.

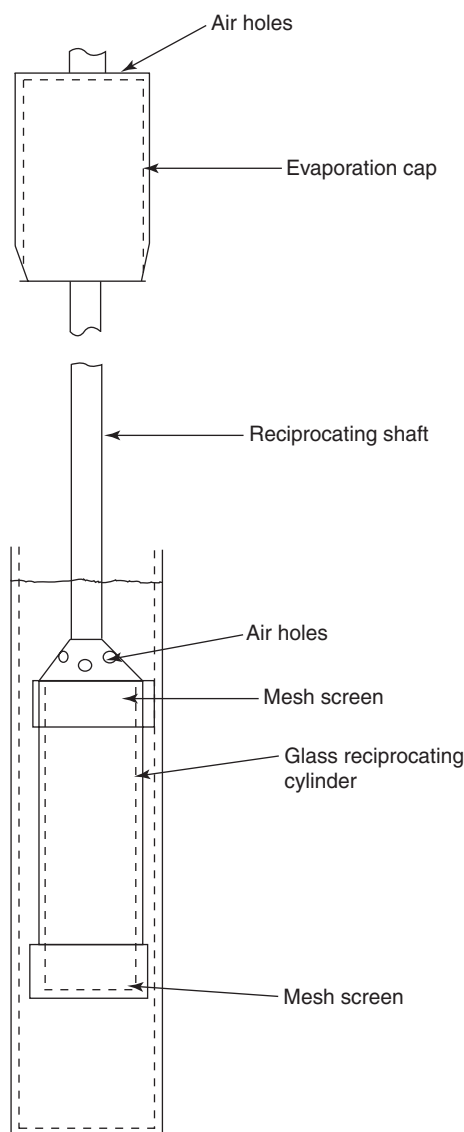
Rotation speeds of 50 or 75 rpm for the paddle apparatus and 50 or 100 rpm for the basket apparatus are recommended for tablets, pellets, and capsules. The paddle apparatus at 25 rpm is recommended for suspensions. In general, mild rather than vigorous agitation conditions will allow for greater discriminating power to detect potentially significant changes in dosage form performance.

The volume of dissolution medium is generally within the range 500–1000 ml, and for most applications the pharmacopoeial norm of 900 ml is used.

### Flow-Through Apparatus

The USP/EP/JP flow-through apparatus (Figure 3) should be considered if the drug substance exhibits low solubility or if pH changes to the dissolution medium are required during the dissolution test, as in the case of controlled or delayed release products.

The dosage form is placed in a cell on a bed of glass beads or on a wire tablet holder. The pump forces the dissolution medium upwards through the cell. Filters prevent undissolved particles from escaping from the top of the cell. Analysis of the eluate may be performed directly online or by the collection of fractions at each specified time point.



**Figure 3** Reciprocating cylinder apparatus.

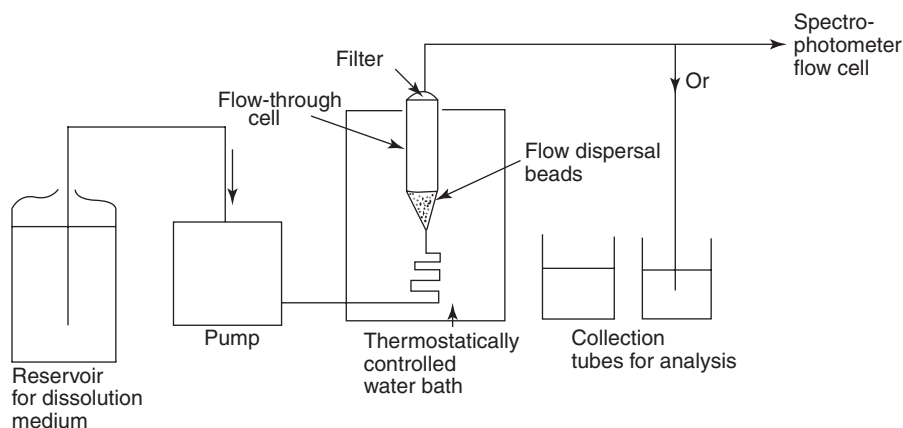
Typical flow rates are 4, 8, 16, and 32 ml min<sup>-1</sup> and standard cell diameters are 12 and 22.6 mm. For most formulations a flow rate of 16 ml min<sup>-1</sup> using the 22.6 mm cell are appropriate.

The bed of glass beads normally used to ensure laminar flow within the cell may be omitted (combined with a low flow rate) when investigating IV/IV relationships to more accurately reproduce conditions found within the small intestine.

### Reciprocating Cylinder

The reciprocating cylinder (USP apparatus 3) (Figure 4) is useful for nondisintegrating extended-release products, particularly where accurate IV/IV correlation is needed.





**Figure 4** Flow-through apparatus.

The apparatus consists of a set of transparent cylinders, capped at each end with a screen, in which the dosage form is enclosed. The cylinders are gently moved up and down (the dip rate) in dissolution medium contained in glass tubes held in a water bath. The medium enters the cylinders allowing dissolution to occur. At the specified times the reciprocating cylinders are raised and a portion of the solution is removed for analysis.

### Other Apparatus

The USP also lists three other dissolution apparatus for transdermal delivery systems, apparatus 5 (paddle over disk), apparatus 6 (cylinder), and apparatus 7 (reciprocating holder).

## Dissolution Media

The medium selected should be based on the physicochemical properties of the active ingredient(s) and the relevance of the medium to the performance of the dosage form in the body. This particularly applies to the pH and ionic strength of the medium.

In the case of drugs with high aqueous solubility in the range pH 1.2–6.8, the adoption of test conditions without regard to discriminating power (ability to distinguish between batches) is acceptable, provided dissolution rate does not affect bioavailability and is not influenced by formulation or manufacturing variables.

If it is known that the dissolution rate has a significant effect on bioavailability, or if formulation or manufacturing variables affect dissolution rate, then an attempt should be made to develop test conditions that can distinguish between these changes.

Generally, the objective is to select a medium in which sink conditions are maintained so that the drug already in solution does not exert a modifying effect on the dissolution of the remaining drug. For low solubility drugs it may not be possible to achieve sink conditions, and so it may be preferable to test formulations of these drugs under nonsink conditions that accurately reflect the *in vivo* situation.

The pH of the medium is usually within the range pH 1.2–8.0. Water, traditionally a first choice dissolution medium, is now considered less than ideal for most new dissolution tests. Reasons for this include lack of *in vivo* relevance and lack of buffering capacity.

Hydrochloric acid (0.01 or 0.1 mol l<sup>-1</sup>) is the preferred medium for drugs that are readily soluble at gastric pH. Aqueous buffer solutions are used when solubility, dissolution rate, or drug stability are pH dependant. The use of simulated gastrointestinal fluids (from the USP), with or without enzymes, may be justified in some circumstances.

Surfactant solutions are recommended for drugs that are poorly soluble in aqueous media but have increased solubility in surfactant solutions. The addition of low concentrations of surfactants such as polysorbate, sodium lauryl sulfate, and hexadecyltrimethylammonium bromide may allow sink conditions to be obtained, although foaming can be a problem.

A summary table for selecting the most suitable media for immediate release products is given in **Table 2**.

Delayed-release dosage forms (enteric coated) require a two-stage dissolution test, first in simulated gastric fluid to ensure drug is not released in the stomach, and then in simulated intestinal fluid to ensure drug release in the intestine.

**Table 2** Dissolution medium options

<i>Dissolution medium</i>	<i>Comments</i>
0.01 mol l <sup>-1</sup> Hydrochloric acid	First choice for a low pH medium, and now the recommended medium for gastric soluble drugs. Recent USP guidance suggests 0.01 mol l <sup>-1</sup> is more appropriate than traditionally used concentration of 0.1 mol l <sup>-1</sup>
0.1 mol l <sup>-1</sup> Hydrochloric acid	Recommended as the simulated gastric fluid for the first phase of testing for enteric-coated products. Widely used for existing products
Water	No longer a recommended medium for most new dissolution tests. Still widely used for existing products
Buffer solutions	Useful when solubility/dissolution rate is pH dependant or when chemical stability is pH dependant. Recommended buffered media are pH 1.2, 1.5 (HCl media), 4.5, 5.5, 5.8, 6.8, 7.2, and 7.5 (phosphate buffers)
Surfactant solutions	Recommended for poorly soluble drugs in combination with the paddle apparatus
Simulated gastro and intestinal fluids with enzymes	Recommended for aged gelatine capsules (may be used as the second method)
Physiologically based or biorelevant media	Useful for predicting food effects during formulation development of poorly soluble drugs and for IV/IV correlation. Suggested media are fasted state simulated intestinal fluid and fed state simulated intestinal fluid
Nonaqueous	Last resort for insoluble drugs. Not recommended

For extended-release dosage forms it is beneficial to establish an IV/IV correlation. Media that mimic the *in vivo* conditions are more likely to give reliable IV/IV correlation. If the majority of drug release is in the intestine, media in the pH range 6.8–7.5 should be considered. If the release of drug from the dosage form is pH, bile salt, or osmotic potential dependant, pH changes to the dissolution medium during the test may improve the chances of a reliable IV/IV correlation.

## Methods of Detection

Ultraviolet (UV)/visible spectrophotometry is the most widely used technique as it is simpler and easier to automate. High-performance liquid chromatography or other methods are usually only considered when it has been established that UV/visible spectrophotometry (including derivative UV and multi-component analysis) is not feasible or when the drug product contains more than one active ingredient.

If continuous monitoring of the dissolution medium is adopted through the use of flow-cells or fiber optics,

**Table 3** Typical dissolution data for instant and controlled release products

<i>Time (min)</i>	<i>% labeled strength dissolved (mean of 6)</i>		
	<i>Tablet formulation of a high solubility drug</i>	<i>Tablet formulation of a low solubility drug</i>	<i>Controlled release capsule formulation</i>
15	93	69	—
30	98	82	—
45	100	89	—
60	—	91	24
240	—	—	53
480	—	—	87

then the method is normally validated against manual sampling.

## Dissolution Limits and Interpretation of Results

Dissolution limits are established to ensure batch-to-batch consistency and to signal potential problems with bioavailability. The most widely used acceptance criteria are those of the USP but acceptance criteria are also given in the BP and JP (check against the current editions). IV/IV correlation may be used to establish these criteria if bioavailability data are available for batches showing different release rates.

Generally, the dissolution test is required to be performed on a total of six dosage units. A dissolution profile may be generated by sampling at 15 min intervals. For rapidly dissolving products, sampling at 5 or 10 min intervals may be necessary to produce an adequate profile. For immediate-release formulations a single time point limit is usually specified to ensure that most of the active ingredient is released within the preset time (typically 70–80% released after 15, 30, or 45 min). For extended-release dosage forms, limits are normally set at a minimum of three time points. The first limit is set to prevent dose dumping (rapid release if the drug) and should therefore correspond to ~20–30% released. The second limit should define the dissolution pattern and thus be set ~50% release. The final limit is specified to ensure almost complete release of drug, i.e., > 80%. Examples of various dissolution data are given in Table 3.

## Calibration of Dissolution Apparatus

To ensure reproducibility of dissolution data from one apparatus to another it is essential that the apparatus complies with the dimensions and tolerances

**Table 4** Parameters influencing calibration

Parameter	Comments
Location	Strong, level bench required
Vibration	Minimize vibration from external sources, e.g., heaters, pumps
Dimensional verification	Dimensions of equipment must comply with pharmacopoeias
Head plate co-planarity	Head plate (holding the shafts) must be level with the base plate (holding the vessels)
Shaft perpendicularity	Basket and paddle shafts must be perpendicular to the base plate
Centering vessels to the stirrer shaft	Distance between shaft axis and vertical axis of the vessel should be $\leq 2$ mm
Stirrer height	Distance of stirrer from bottom of vessel should be $25 \pm 2$ mm
Rotation speed	Speed of rotation of paddle or basket stirrer should be within $\pm 4\%$ of the specified value
Eccentricity (wobble or run out)	Each paddle or basket shaft must be straight along its entire length to ensure smooth stirring characteristics
Temperature of dissolution medium	Medium temperature should be controlled to $37 \pm 0.5^\circ\text{C}$
Flow rate (flow-through apparatus only)	Flow rate of dissolution medium through cells should be within $\pm 5\%$ of the set value

specified in the pharmacopoeias. There are a number of parameters that can lead to a calibration failure and these are summarized in **Table 4**, although it is possible for other factors to have an effect.

Reference tablets of known dissolution rate are available from the USP to determine apparatus suitability. Prednisone tablets (disintegrating) and salicylic acid tablets (nondisintegrating) may be used to check the calibration of the basket, paddle, and flow-through apparatuses. Chlorpheniramine extended-release tablets and theophylline extended-release beads have recently become available for checking the calibration of the reciprocating cylinder apparatus.

## Intrinsic Dissolution

During early formulation development it is useful to have a constant, defined surface area exposed to the dissolution medium in order to compare dissolution rates of drug substances. Typically, intrinsic dissolution is carried out using a compressed disk of material with only one side exposed to the dissolution fluid. Comparison of intrinsic dissolution rate may be especially useful when characterizing different salts,

polymorphs, or milling methods of the same drug substance.

**See also:** **Pharmaceutical Analysis:** Stability Testing. **Pharmacokinetics:** Absorption, Distribution, and Elimination; **Pharmacodynamics.**

## Further Reading

- Abdou HM (1989) *Dissolution, Bioavailability and Bioequivalence*. Mack Publishing Corp.
- Amidon GL, Lennernas H, Shah VP, and Crison JR (1995.) A theoretical basis for a biopharmaceutical drug classification system: the correlation of *in vivo* drug product dissolution and *in vivo* bioavailability. *Pharmaceutical Research* 12(3).
- Banakar UV (1992) Pharmaceutical dissolution testing. In: Swarbrick J (ed.) *Drug and the Pharmaceutical Sciences*, vol. 49. New York: Dekker.
- Dissolution Technologies* (Quarterly journal).
- Dissolution Testing of Solid Oral Products (1997) Joint Report of the Section for Official Laboratories and Medicines Control Services and the Section of Industrial Pharmacists of the F.I.P. *Dissolution Technologies* 4: 5–14.
- Dressman JB, Amidon GL, Reppas C, and Shah VP (1998.) Dissolution testing as a prognostic tool for oral drug absorption: immediate release dosage forms. *Pharmaceutical Research* 15(1).
- FDA Guidance for Industry, Dissolution Testing of Immediate Release Solid Oral Dosage Forms, FDA (CDER), August 1997.
- Galia E, Nicholoides E, Horter D, *et al.* (1998.) Evaluation of various dissolution media for predicting *in vivo* performance of class I and II drugs. *Pharmaceutical Research* 15(5).
- Guidance for Industry – Immediate Release Solid Oral Dosage Forms (SUPAC-IR): Chemistry, Manufacturing and Controls, *In vitro* Dissolution Testing and *In vivo* Bioequivalence Documentation, Food and Drug Administration (CDER), November 1995.
- Guidance for Industry – Modified Release Solid Oral Dosage Forms (SUPAC-MR): Chemistry, Manufacturing and Controls, *In vitro* Dissolution Testing and *In Vivo* Bioequivalence Documentation, Food and Drug Administration (CDER), September 1997.
- Hanson WA (1991) *Handbook of Dissolution Testing*, 2nd edn. Eugene, OR: Aster Publishing Corp.
- ICH Guideline Q6A Pharmeuropa 12 (2) April 2000, pp. 188–189.
- Leeson LJ (2000) Some observations on rethinking the use of water as a dissolution medium. *Dissolution Technologies* 7(2): 16–17.
- Moore JW and Flanner HH (1996) Mathematical comparison of dissolution profiles. *Pharmaceutical Technology* 20(6): 64–74.
- Noory C, Tran N, Ouderkirk L, *et al.* (1999) Rethinking the use of water as a dissolution medium. *Dissolution Technologies* 6(4): 6–7.

## Sample Preparation

R Whelpton, University of London, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

No matter how sophisticated the analytical equipment, the results will be rendered worthless if the sampling, sample storage, and sample preparation are not performed correctly. Thus, it is important to be familiar with the chemistry of the analyte and the nature of the sample matrix. Knowledge of the history of the sample is desirable; this entails knowing from where it originates and how it has been obtained, stored, and transported. Whenever possible, the analyst should be involved in the design of any protocols relating to drug dosing and sample collection. There is little point in assaying plasma samples that have been collected through an indwelling cannula that was used for administration of the drug (or samples collected from the same limb for that matter). The phlebotomist should be warned not to swab with alcohol if the sample is to be assayed for ethanol. It may sound fatuous, but such things occur; a problem during a study with chlorhexidine was traced to sampling with apparatus that had been sterilized with the compound being assayed.

### Sample Types and Storage

The range of samples submitted to analysis extends from pure solutions of a pharmaceutical ingredient to a putrefying piece of tissue for forensic analysis. Liquids, such as blood, plasma, serum, saliva, urine, and cerebrospinal fluid, are generally easier to sample than solids and semi-solids. However, they are not necessarily homogeneous with regard to drug concentration. Feces are not homogeneous, and determination of 24 h drug excretion rates will require that the entire sample be taken for analysis; something to remember when the experiment is being planned. Tissues need homogenization or digestion prior to sampling. With soft tissues, such as brain and liver, this can be done with a polytetrafluorethylene (PTFE)-in-glass homogenizer, but lung and muscle generally require something with a cutting or mincing action. Alternatively the tissue may be digested with a proteolytic enzyme – a technique frequently used in forensic and toxicological analyses. Enzymes can be used to ‘release’ plasma-bound drugs, but this may also release interfering substances.

### Blood and Blood Fractions

A decision should be taken as to whether whole blood, plasma, or serum is to be assayed. For drugs that are not distributed into erythrocytes, the concentration in plasma will be higher than that in blood, and there is little point in contending with red cells if plasma or serum can be used. Collection of plasma requires an anticoagulant, and an appropriate choice has to be made. Some drugs have been reported to bind to heparin, resulting in a loss of extraction efficiency. Plasma, as opposed to serum, may be chosen when it is thought that enzymic degradation of the drug requires  $\text{Ca}^{2+}$  or  $\text{Mg}^{2+}$ , the idea being that the chelating agents will bind these ions. The concentration of fluoride in fluoride oxalate tubes is unlikely to be sufficient to inhibit plasma esterases, which may not be inhibited completely by fluoride even at very high concentrations. The anticoagulant in citrate tubes is in aqueous solution, and the sample will be correspondingly diluted, a fact that has not always been recognized.

For some applications serum may be preferred to plasma. Problems with anticoagulants, including use of lithium heparin tubes for samples to be assayed for lithium and fibrinogen precipitation on freezing and thawing are avoided, but the delay while the clot forms and binding of analyte to the clot must be considered. Drugs may be metabolized differently in blood compared with in plasma. For example, at therapeutic concentrations pyridostigmine is lost rapidly from human blood but only slowly from plasma. Analyte stability in blood is rarely considered when the analysis is to be performed on plasma or serum but may be important, particularly if the plasma is not separated immediately after the blood has been taken.

For drugs that are extensively ‘bound’ to red cells, e.g., chlorthalidone concentrations can be 40 times higher in red cells than in plasma due to binding to carbonic anhydrase, an imperceptible degree of hemolysis will result in elevated plasma or serum concentrations. Addition of stabilizers to blood could affect the plasma:red cell distribution of a drug, as could temperature and pH (see later), and so for some drugs it may be better to assay whole blood. The problem of disruption of the cells on freezing and subsequent sampling of blood can be overcome by having known volumes of blood pipetted into extraction tubes while it is fresh.

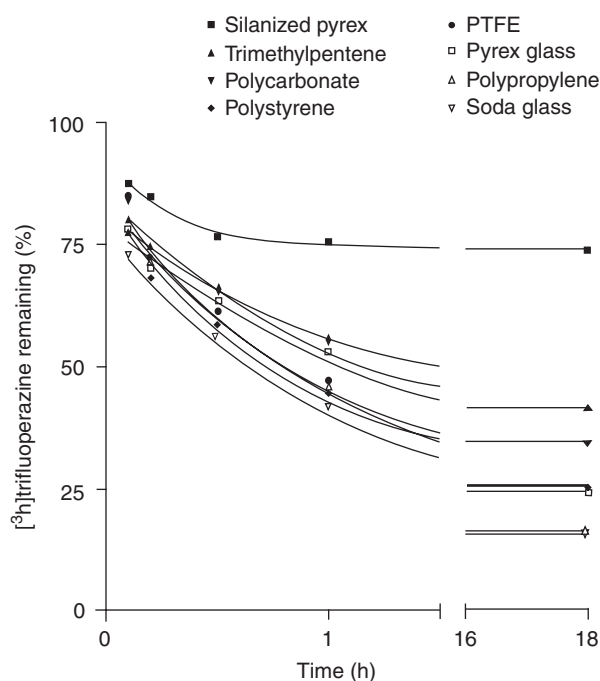
Having chosen the fluid to be assayed, the problem of preparing calibration standards should be

considered. Commercially obtained plasma (human and animal) frequently contains high concentrations of lignocaine (lidocaine) and may be diluted with citrate solution. The pH of stored plasma may have to be adjusted as it tends to be higher than physiological pH, presumably due to the loss of CO<sub>2</sub>. Blood for calibration standards is problematic; it cannot be frozen without disrupting the red cells, and when drug is added to fresh blood it may take several hours for equilibration between plasma and erythrocytes.

### Containers

The choice of containers, their labeling, transportation, and storage are no less important to the success of an assay than the choice of detection system or chromatographic conditions. Drugs may be lost from, or contaminated by, inappropriately chosen containers. For volatile compounds such as halothane, plastic containers and stoppers should not be used. Glass tubes with aluminum-lined screw caps are generally suitable for such analytes, but to reduce loss on sample transfer, the tube should be filled so as to avoid air spaces. Cooling samples to 4°C before transfer may help reduce losses. An alternative, if headspace analysis is to be used, is to sample directly into preweighed headspace vials. Loss of drugs onto glass and into plastics is well known, but it must be remembered that different types of glass and plastic behave differently, and what may be suitable for one analyte may not necessarily be suitable for another. PTFE-lined caps are preferred by some as being inert, but it is not necessarily 'nonstick' as far as drugs are concerned (Figure 1). Adsorption of basic drugs onto glass is sometimes reduced by addition of amine modifiers (triethylamine, ethanolamine, etc.), in the same way that they are used to reduce peak tailing on high-performance liquid chromatography (HPLC) columns, or high concentrations of a related drug may be introduced to compete for the binding sites. This approach was adopted in a gas chromatography electron-capture detection (GC-ECD) assay of chlorpromazine, high concentrations of promazine, which has weak electron-capturing properties, being added at high concentrations to compete for adsorptive sites on the extraction tubes.

Interferences from plastic containers are often plasticizers that leach into the sample. If they produce strong responses in the electron capture detector, then even small quantities can have a deleterious effect on an assay. Phosphate-based plasticizers have caused problems with nitrogen-phosphorous detection. Furthermore, it has been shown that tris (2-butoxyethyl)phosphate, a plasticizer once used in Vacutainer stoppers, can displace basic drugs



**Figure 1** Loss of radiolabeled trifluoperazine from aqueous solution to different materials, as a function of time.

from protein-binding sites and that in the case of nortriptyline this resulted in reduced plasma concentrations as the displaced drug partitioned into erythrocytes. The plasticizer ethylhexylphthalate has been shown to increase the liquid extraction of zimelidine by adduct formation.

The container should not crack, nor should its stopper come off, during expansion of the contents due to freezing and thawing and other pressure changes, e.g., in aircraft holds. Generally, screw caps are preferable to push-fit stoppers. It is sensible to arrange samples in logical groups and to pack them in outer containers that minimize the risk of cross-contamination should a number of containers fail. In pharmacokinetic investigations, is it better to lose samples from one subject or all the samples at one time point? Tubes containing frozen samples should be inspected for damage before thawing. Naturally, the shape and size of the container must be suitable for its purpose. Psychiatric in-patients have been known to produce in excess of 2.5 l of urine a day. Human feces may be collected in polythene bags lining a suitable vessel to aid collection.

Irrespective of the nature of the container, the labels must remain firmly attached even when wet or during cold storage. The information must still be legible, and special marker pens may have to be issued. Preprinted labels can be used, but then one is reliant on the sample being put in the correct tube.



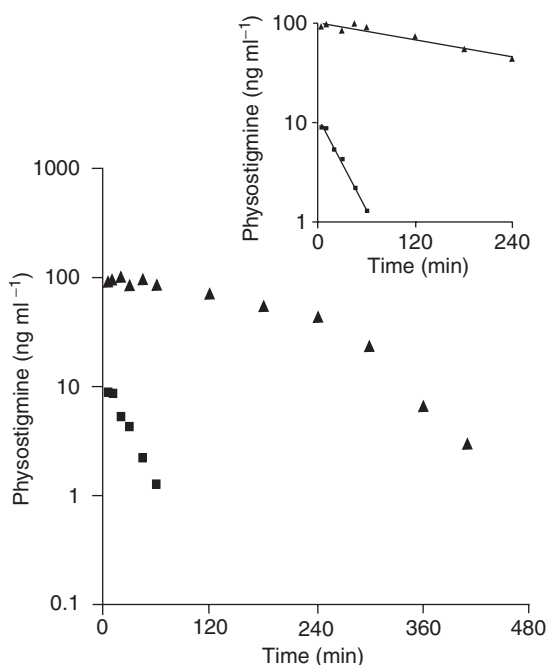
## Temperature

Generally samples are stored frozen at  $-20^{\circ}\text{C}$ , but lower temperatures may be needed. Plasma esterases may still be active at this temperature, and  $-70^{\circ}\text{C}$  may be required. The effect of freezing and thawing must be considered. Obviously, red cells will be hemolyzed and fibrinogen will separate from plasma. These solids will have to be separated before sampling, and it is important to prove that the analyte does not bind to them. Protein binding is temperature dependent, and centrifuging blood at  $4^{\circ}\text{C}$  rather than  $24^{\circ}\text{C}$  has been shown to result in higher plasma phenytoin concentrations. The protein binding was reduced at the higher temperature and the nonbound form partitioned into the red cells. Stratification occurs when samples are frozen, and so, after thawing, the contents of tubes should be agitated before sampling. Obviously, liquid should not be withdrawn until the sample is completely thawed to avoid 'apple-jacking', but when dealing with a large batch of tubes the presence of ice in some tubes may be overlooked.

## Preservatives

If the analyte is known to be unstable, then it may be appropriate to add some form of stabilizer to prevent decomposition. For example, tubes for blood alcohol determination should contain sufficient sodium fluoride to produce a final concentration  $>1.5\%$  (m/v), as otherwise colonies of opportune bacteria and fungi may produce, or utilize, ethanol.

For compounds that are unstable at certain pH values, adjustment of the pH before storage would be advisable. Esters and, to some extent, amides are subject to enzyme hydrolysis and may be rapidly destroyed by plasma enzymes. Stability testing must be conducted at appropriate concentrations. The *in vitro* half-life of physostigmine in plasma at  $37^{\circ}\text{C}$  has been reported as 190 or  $\sim 15$  min. This discrepancy is easily explained as the studies used different initial concentrations of the drug, which exhibits pseudo first-order metabolism at concentrations  $<10\text{ ng ml}^{-1}$  (Figure 2). Since the therapeutic concentrations of physostigmine are generally below  $2\text{ ng ml}^{-1}$ , the lower concentration was more appropriate for the stability study. The loss of physostigmine can be prevented by collecting blood in tubes containing neostigmine or pyridostigmine to produce a final concentration of  $10\text{ mg l}^{-1}$ . An advantage of using the quaternary ammonium compounds is that they are not extracted into organic solvents. Cocaine is similarly lost from plasma, and excess physostigmine has to be used to stabilize samples. Silver



**Figure 2** Effect of initial concentration on the kinetics of *in vitro* hydrolysis of physostigmine from human plasma at  $37^{\circ}\text{C}$ . From initial concentrations of  $10\text{ ng ml}^{-1}$  (■) and  $1\text{ ng ml}^{-1}$  (data not shown), the kinetics are pseudo-first-order ( $t_{1/2} \approx 19\text{ min}$ ), whereas the kinetics of decay from  $100\text{ ng ml}^{-1}$  (▲) are best described by the Michaelis–Menten equation. Inset: erroneously fitting the initial part (0–240 min) of the  $100\text{ ng ml}^{-1}$  data to a single exponential gives an apparent half-life of 210 min. Therapeutic concentrations of physostigmine are generally below  $2\text{ ng ml}^{-1}$ . (Reproduced with permission from Hurst PR and Whelpton R (1989) Solid phase extraction for an improved assay of physostigmine in biological fluids *Biomedical Chromatography* 3: 226–232; © John Wiley & Sons Ltd.)

nitrate can be used to prevent enzyme hydrolysis of glyceryl trinitrate and its dinitro metabolites. If the nature of the reaction is known, then specific inhibitors may be chosen, for example the use of the monoamine oxidase inhibitor pargyline during catecholamine assays.

Antioxidants such as ascorbate and metabisulfite have their place, but should not be added indiscriminately. Many tertiary amine drugs are metabolized to amine oxides that will be reduced by such antioxidants, not only elevating the concentration of the parent compound but often also resulting in some dealkylation to secondary amines, thus producing a spurious pattern of compounds. It should be noted that N-oxides and other oxides of nitrogen tend to be labile in alkaline solutions and prolonged exposure to high pH during sample preparation may cause problems, particularly in the presence of biological materials.



## Strategies for Sample Preparation

The methods chosen for sample preparation depend to a large extent on the overall strategy. The nature of the analyte, the required sensitivity, the type of sample matrix, the number of samples to be assayed, and the time and resources available will influence the choice. Before deciding on an approach, several questions need to be answered. Is the drug stable (chemical, thermal, light), volatile, reactive, lipophilic, or ionic? Is it bound to proteins and does it partition into erythrocytes? How is it to be quantified? What is the required sensitivity and how difficult will it be to meet that requirement? If the drug is metabolized, will the metabolites be in the sample, and are they to be assayed? Might metabolites interfere with the quantification? Questions concerning the matrix include 'Is it liquid or solid?', 'Is it homogeneous?', 'Will it contain other drugs or interferences?', and 'Will the sample volume be restricted?' Answers to questions concerning the number of samples and from where they originate will influence whether samples are to be assayed almost immediately or whether they will go into storage. If samples are to be stored, then the effects of storage time and temperature will have to be investigated.

The answers will exclude certain assay procedures, e.g., if the analyte is heat labile, then gas chromatography is usually inappropriate, as is evaporating an extraction solvent at elevated temperatures. If the concentration is high, or a particular assay is sensitive, then sample preparation may be minimal. On the other hand, trace amounts of analyte may require a complex assay procedure with multiple concentration and cleanup steps.

A method should be designed to be as simple as possible. Analysts should ask themselves why they are doing the assay. If the samples are the result of a multicenter trial that has generated hundreds, if not thousands, of samples, then throughput and method stability are paramount, and time spent in method optimization is justified. If a drug is to be measured only occasionally, then one should consider if it could be measured using a method currently in use, e.g., a gas chromatography (GC) method with a temperature change or an HPLC method with a change in detector wavelength. For the occasional sample, long retention times may be tolerable. If a method includes gas or thin-layer chromatography, then liquid extraction should be considered. It may be possible to inject a sample of the organic phase, or it may be necessary to evaporate it. Liquid-liquid extraction is usually appropriate if the analyte is to be derivatized. Solid-phase extractions are suitable for HPLC methods as the drug is usually eluted from

the cartridge with an eluent compatible with HPLC mobile phases.

### Direct Injection

For samples at high concentration or in relatively pure solution, it may be possible to assay the samples with minimal work-up. Some samples may simply need dissolving in a suitable solvent or diluting before analysis. Sometimes blood and plasma samples can be treated this way. Blood ethanol concentrations have been determined by adding blood to an aqueous solution of an internal standard (*n*-propanol) and injecting onto a GC column packed with Porapak-Q. Similarly, volumes of plasma containing sulfadimidine can be mixed with an equal volume of internal standard solution and centrifuged and a portion injected onto a C-18 HPLC column. By preparing the internal standard solution with twice the concentration of organic modifier as in the HPLC phase, the injection solution is compatible with the HPLC eluent, thus reducing the risk on precipitation of the column.

### Protein Precipitation

For plasma or blood samples in which the analyte concentration is high but which cannot be sampled directly, protein precipitation may be a suitable approach. Several agents have been used, the more efficient being acids those such as trichloroacetic, perchloric, metaphosphoric, and tungstic acids, which can be added as concentrated solutions so as not to dilute the sample unduly. The analyte must be stable at low pH values. As the supernatant is aqueous, this approach is best used with HPLC. Acetonitrile, methanol, or acetone may be more suitable for precipitating proteins when the analyte is acid labile. Also, these organic solvents may be preferable prior to GC analysis as they are easily evaporated and generally compatible with GC phases. Whatever agent is used, the risk of co-precipitation of the analyte must be assessed. For some drugs precipitation may be the first step, followed by purification and concentration by liquid-liquid extraction or solid-phase extraction (SPE).

### Liquid-Liquid Extraction

Although liquid-liquid extraction is an old technique, correctly applied, it remains one of the most useful. Generally it is best to use the least polar solvent that will effectively extract the drug. The nonpolar solvents *n*-heptane, cyclohexane, and toluene should be examined first, and if these prove unsuitable, then diethyl ether, ethyl acetate, and

chloroform can be considered. There is usually no advantage in selecting a solvent with too much 'extracting power' as this may reduce the selectivity of the assay by promoting the extraction of interfering compounds. The density of the solvent may be important; if large volumes of organic solvents are to be used and the phases separated in a separating funnel, then chlorinated solvents with densities  $> 1$  might be chosen. (Chlorinated solvents may be used with phase separating filter paper, although centrifugation to break any emulsion is required prior to use.) If the extraction is performed in tubes, it is usually better to have the solvent as the upper layer to facilitate removal. By mixing solvents, e.g., ether/chloroform, a polar extraction solvent that 'floats' can be obtained. An alternative is to add sufficient salt to the aqueous phase so that it is denser than the organic one. Potassium carbonate has been added to an aqueous solution containing acetonitrile to 'force' an organic layer to form.

If the organic solvent is to be evaporated, then a low boiling point may be an advantage, and hexane may be preferable to heptane, for example. However, very volatile solvents present handling difficulties. Evaporation of diethyl ether during sample transfers can result in extraction efficiencies that are apparently  $> 100\%$ , if appropriate corrections are not made. Under these circumstances the judicious use of an appropriate internal standard is recommended.

A further consideration is whether the solvent is likely to interfere with the analysis. Toluene is a poor first choice for a method that is to quantify the drug by ultraviolet detection. It is difficult to remove all traces of the solvent, and even a small residual amount can affect the detection limit seriously. Similarly, chlorinated solvents are best avoided if radioactive scintillation counting or electron-capture detection is to be used. Some solvents may not be suitable because they would react with the analyte, for example a ketone such as ethyl methyl ketone would react with primary amines. Solvent impurities and additives may be unknown to the analyst. Antioxidants such as hydroquinone and pyrogallol are added to diethyl ether to limit peroxide formation. These highly electroactive molecules can affect electrochemical detection methods adversely, particularly if they have been concentrated by solvent evaporation. Freshly distilled diethyl ether may be used, but it should not be stored as this is not only potentially dangerous, the peroxides that form may decompose the analyte. Methyl *t*-butyl ether (boiling point  $55^\circ\text{C}$ ) is supplied without antioxidants and is a useful alternative to diethyl ether. Chloroform and dichloromethane may be stabilized with ethanol although pentene is used by at least one manufacturer and may

be preferred for some applications. Impurities are a similar cause for concern: even with the addition of stabilizers, low concentrations of peroxides and phosgene have been known to present problems. Isovaleraldehyde, an impurity in isoamyl alcohol, has been shown to react with primary amine metabolites of chlorpromazine.

### pH-Controlled Extraction

The extraction of a weak acid or base into an organic phase is a function of the pH of the aqueous solution, the  $pK_a$  of the analyte, and the partition coefficient describing the distribution of the analyte between the particular solvent and water. Under ideal conditions, the fraction extracted,  $F$ , is given by

$$F = [1 + V_{\text{aq}}/(V_{\text{org}}\text{APC})]^{-1},$$

where  $V_{\text{aq}}$  and  $V_{\text{org}}$  are the volumes of the aqueous and organic phases, respectively, and APC is the apparent partition coefficient of the analyte. APC is a function of the true partition coefficient (TPC) (i.e., the partition coefficient of the nonionized form that is extracted),

$$\text{TPC} = [1 + 10^{(pK_a - \text{pH})}]\text{APC}$$

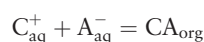
for a base or

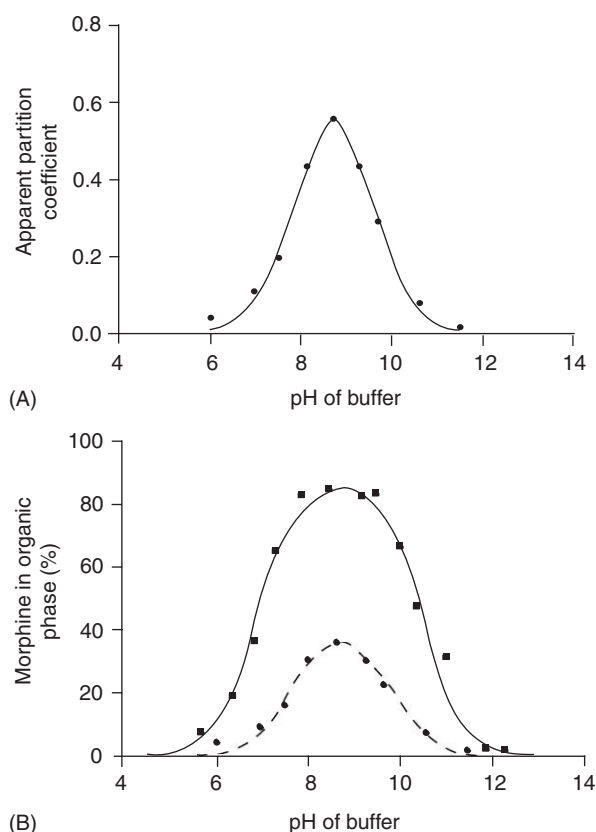
$$\text{TPC} = [1 + 10^{(\text{pH} - pK_a)}]\text{APC}$$

for an acid. Thus, the fraction extracted can be calculated from knowledge of pH,  $pK_a$ , and TPC. Usually, TPC for the particular solvent system is unknown, and the  $pK_a$  may be unknown or in doubt. However, these can be derived from iterative curve fitting of the experimental data, plotted as APC versus pH (Figure 3A). The advantage of this is that once the parameters are known, the extraction recoveries can be predicated for any organic:aqueous volume ratios for any pH value (Figure 3B). A day's experimentation allows optimization of extraction pH, solvent volume, and pH required for back-extraction into a predefined volume of aqueous phase. If a method is to be modified, the effect of a change in experimental conditions can be calculated in a few minutes.

### Ion-Pair Extraction

Ionized drugs can be extracted into organic solvents, provided an ion of the opposite charge (counterion) is extracted to maintain electrostatic neutrality. For a cation,  $C^+$ , being extracted as an ion-pair with a suitable anion,  $A^-$ ,





**Figure 3** pH-partition characteristics of morphine. (A) Partitioning of [ $^3\text{H}$ ]morphine between buffers and an equal volume of toluene–butanol (9:1 v/v). The data were fitted to the equations described in the text, using Graph Pad Prism v3, to give  $\text{TPC} = 0.62$ ,  $\text{p}K_{\text{a}1} = 7.77$ ,  $\text{p}K_{\text{a}2} = 9.69$  (solid line). (B) Comparison of predicted (solid line) and actual extraction of morphine assayed by liquid chromatography (■) for a 10:1 volume ratio of organic solvent:aqueous phase. The data of A plotted as percent extracted (—●—).

The distribution ratio,  $D_{\text{C}}$ , is given by

$$D_{\text{C}} = [\text{C}_{\text{org}}]/[\text{C}_{\text{aq}}] = E_{\text{CA}}[\text{A}_{\text{aq}}^-]$$

where  $E_{\text{CA}}$  is the equilibrium constant. Thus, the distribution is a function not only of the nature of the analyte and the extraction solvent but also the nature and concentration of the counterion. The partitioning can be further affected by introducing a substance with which the analyte is able to complex. This makes ion-pair extraction versatile and also complicated. Ion-pair extraction may be used with SPE columns – although interactions with cationic or anionic solid phases may result in a simpler approach.

It is worth noting that aberrant partitioning in liquid–liquid extraction can sometimes be explained by unintentional ion-pair extraction, the counterions arising from the buffer or biological matrix. Therefore,

unless ion-pair extraction is required, solvents that promote ion-pair extraction are best avoided.

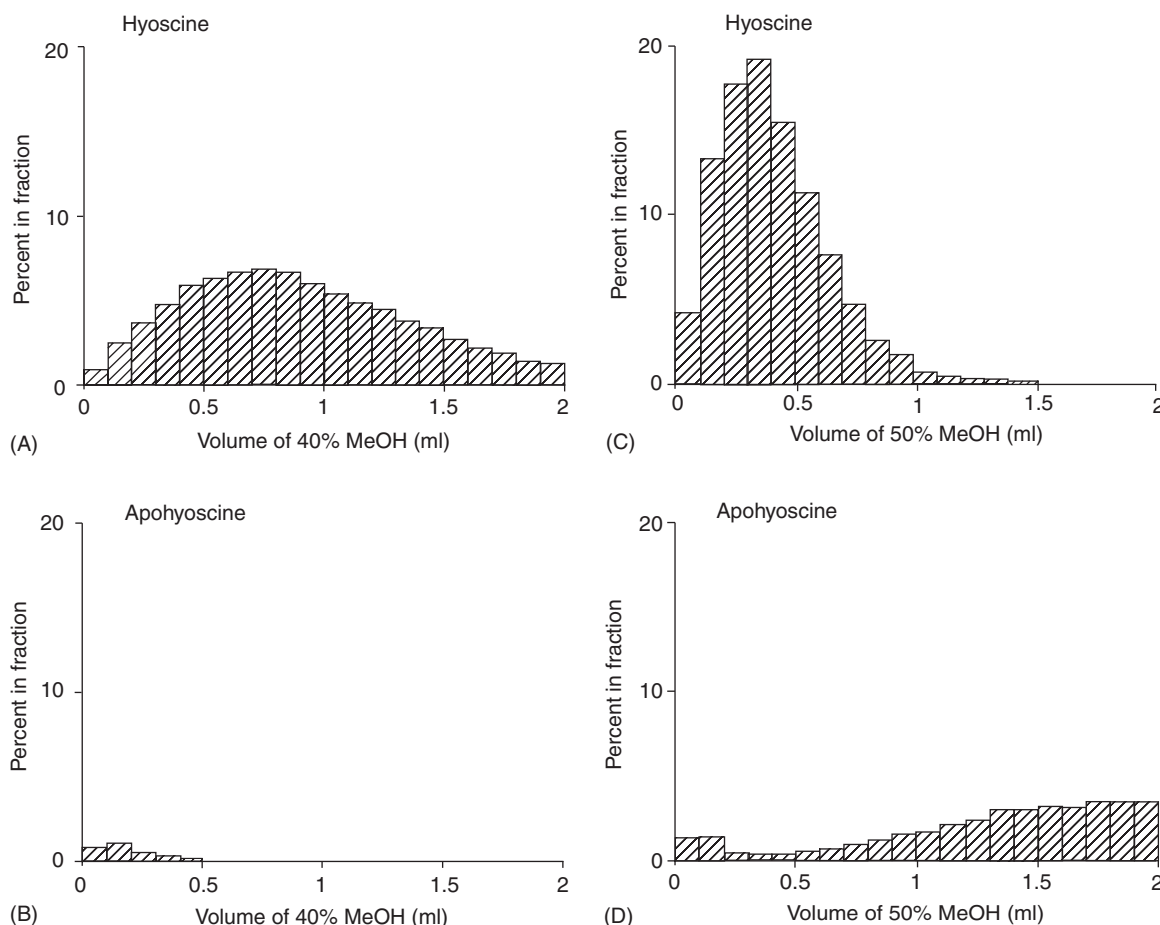
### Solid-Phase Extraction

SPE continues to be popular for sample preparation prior to analysis. Recoveries exceeding 95% with correspondingly low relative standard deviations are possible, and subnanogram quantities of analyte have been assayed without difficulty. It is particularly useful for drugs that are difficult to extract using liquid extraction, for example quaternary ammonium compounds such as pyridostigmine and (+)-tubocurarine and other hydrophilic compounds such as glucuronide metabolites. Because ionic forms can be extracted, it is possible to develop SPE methods that do not require such extreme pH values as required by liquid extraction, which makes SPE suitable for pH-sensitive drugs. Furthermore, it should be possible to perform SPEs in an inert atmosphere, by using syringes to load and elute samples or nitrogen to provide positive pressure to SPE columns.

Trace analysis, the ability to concentrate a drug and its metabolites from a large volume of fluid, is easily accomplished with SPE, and frequently, the compounds of interest can be eluted in a small volume of fluid. Even if the resulting eluate contains interfering substances, the reduced volume allows easier sample handling, e.g., smaller extraction volumes, during subsequent cleanup stages.

Once the analyte has been retained, a range of washing procedures is possible. Water may be used to remove residual hydrophilic materials, including proteins and buffer. Water–organic solvent mixtures may be applied to remove interfering substances and finally a suitable eluent to recover the analyte(s). If the minimum volume of eluent is used, then further concentration may be unnecessary. By careful choice of wash solutions, it may be possible to separate the drug, not only from endogenous interferences but from its metabolites. **Figure 4** shows the elution of hyoscyne and a metabolite, apohyoscyne, from a C-18 SPE column using 50% methanol. Over 90% of the hyoscyne was eluted in the first milliliter, whereas less than 5% of the metabolite was in this fraction. This example is a little unusual in that the metabolite is less polar than the parent and so is retained in preference to the parent drug. Generally metabolites would be expected to elute first from reversed-phase materials.

The continued development of SPE has led to an almost bewildering choice of column geometries and chemistries. The realization that ‘secondary’ interactions with residual silanol groups play an important, and sometime subtle, role in many SPE methods has

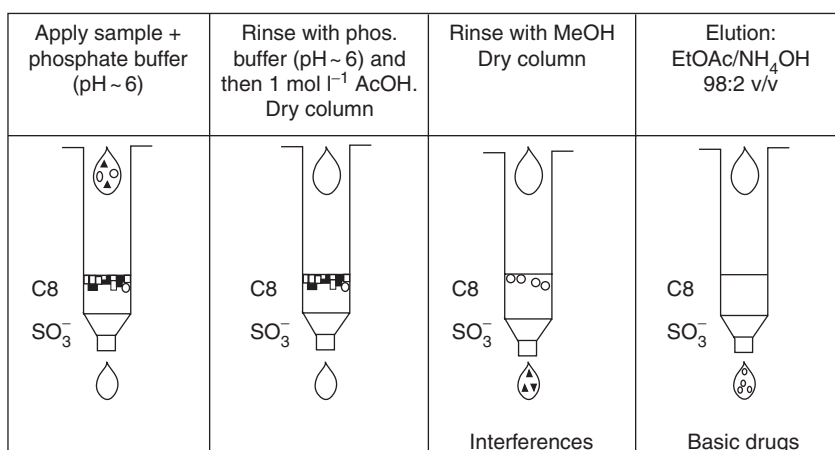


**Figure 4** Elution patterns of hyoscine (scopolamine) and its dehydrated metabolite, apohyoscine, from  $C_{18}$  SPE columns. Hyoscine (A) or apohyoscine (B) eluted with 40% (v/v) methanol:water and hyoscine (C) or apohyoscine (D) eluted with 50% (v/v) methanol:water. The latter eluent was chosen for the assay as >90% of hyoscine was recovered in the first milliliter, whereas <5% of the apohyoscine was eluted in this volume. Washing with 20% (v/v) methanol removed the majority of the urinary pigments but not the hyoscine. (Reproduced with permission from Whelpton R, Hurst PR, Metcalfe RF, and Saunders SA (1992) Liquid chromatographic determination of hyoscine (scopolamine) in urine using solid phase extraction *Biomedical Chromatography* 6: 198–204; © John Wiley & Sons Ltd.)

resulted in manufacturers offering a greater choice of phases, with popular columns being available in end-capped and non-endcapped versions, as well as columns with monofunctional bonding of the phase. Extending the concept of more than one type of interaction, mixed-mode columns utilize two or more primary retention mechanisms. The combination of nonpolar and ion exchange materials allows efficient washing to remove nonpolar impurities, without the risk of eluting the compounds of interest prematurely. This is exemplified in Figure 5, which illustrates a 'generic' method for bases, including quaternary ammonium compounds. Such an approach will appeal to those involved in drug development, where several candidate molecules have to be assayed and it is clearly not cost-effective to produce specific methods for each one. To increase throughput even more, the

numbers of samples that can be handled has been increased by adoption of 96-well plate technology (Figure 6) and robotics.

SPE is also moving in the other direction, with moves to make columns much more specific, with the development of antibodies bound to columns, and molecularly imprinted polymers (MIPs), an approach referred to as predetermined selectivity. Once polymerized, the template molecule must be removed from the polymer by extensive washing if high sensitivities are to be achieved; otherwise, there is the danger of the residual template molecule interfering in the assay. It has been suggested that a homolog of the drug to be analyzed should be used as the template molecule, and providing drug and homolog can be separated in a subsequent chromatographic procedure, then leaching of traces of template from the



**Figure 5** A 'generic' method for extraction of bases using a mixed-mode phase. The use of the strong cation exchanger allows rigorous washing with methanol to remove interferences, prior to elution of the analytes with alkaline ethyl acetate. (Reproduced with permission from *General Approach to the Extraction of Basic Drugs from Biological Fluids Using ISOLUTE<sup>®</sup> Confirm HCX Mixed Mode SPE Columns*, Application note – TN113 rev 1.1 (2001); © Argonaut Technologies. ISOLUTE is a trademark of Argonaut Technologies.)



**Figure 6** Photograph of a 96-well ISOLUTE Array<sup>®</sup> SPE system (ISOLUTE is a trademark of Argonaut Technologies). (Courtesy of Argonaut Technologies, used by permission.)

column will not be a problem. However, if such an approach works, then one must question the specificity of the MIP. A further problem is that retention by MIPs is affected by the nature of the incubation medium. It is possible to have MIPs custom synthesized, and although this requires an initial high expense, the added specificity, when coupled with a large numbers of samples to be analyzed, may make this approach cost-effective. Other specialized approaches include use of antibodies bound to SPE phases (immunoaffinity SPE), for example.

SPE is not without its difficulties. An obvious disadvantage is that particulate material in the sample will block the column frits so that sample types are

limited. Tissue homogenates are unsuitable unless the analytes are in the supernatant liquid after centrifugation. It is advisable to filter or centrifuge samples (after any pH adjustment) before they are applied to SPE columns. A white gelatinous precipitate is frequently formed when urine is made alkaline, and so methods in which buffer and sample are added to the column reservoirs are best avoided. Blood can be assayed provided the erythrocytes are disrupted, e.g., by freezing. There is little cell debris, and so after centrifugation, ~90% of the sample can be added to the columns.

The recoveries from biological materials may differ, not only between sample types but within a sample type. This may mean that a method optimized using water or a different sample type will have to be modified, possibly by the use of a more retentive column at the expense of selectivity. With some drugs, recoveries from urine appear to be related to the concentration of the sample: deep yellow samples give lower recoveries than paler ones. Reduced recoveries with plasma or serum may be indicative of plasma protein binding competing with binding to the solid phase. For example, the retention of buprenorphine from plasma on C-18 extraction columns was 33%, 90%, and 95% at flow rates of 10, 1, and 0.1 ml min<sup>-1</sup>, respectively. When the buprenorphine was applied in water, the retention was >97%, irrespective of the flow rate through the column. Increasing the capacity of the column, reducing the flow rate to allow equilibration with the solid phase, and disrupting the binding, e.g., with a small amount of ethanol or diluting the sample, say, 10 times with water, may be necessary. An alternative



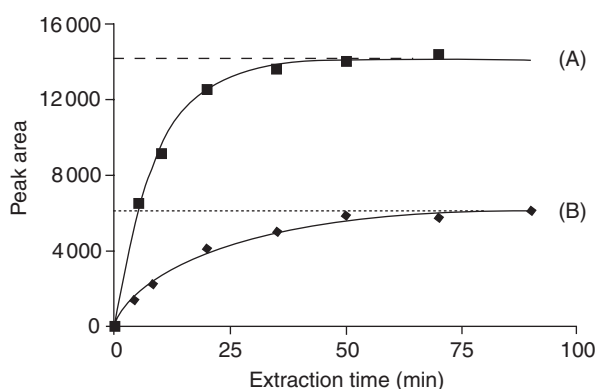
approach to avoiding variable recoveries due to protein binding is to combine LLE and SPE.

Note that because of the trace enrichment capabilities of SPE, it is only a small inconvenience to dilute the sample if need be. However, impurities in water and buffer solutions may be retained, and so it is important to ensure they are pure – even to the extent of passing them down suitable SPE columns prior to use.

### Solid-Phase Microextraction

Solid-phase microextraction (SPME), initially introduced for analysis of environmental samples, has been increasingly used for extraction of drugs from biological matrices. In its simplest form, a solid probe (originally a fused silica fiber) coated with a polymer film is inserted into the sample containing the analyte, which is subsequently thermally desorbed in a hot injection port, in the case of GC or GC–mass spectrometry quantification, or in a desorption interface, in the case of HPLC analysis. Thus, the drug partitions between the sample, usually aqueous, and the liquid phase supported on the fiber, usually polydimethylsiloxane, polydimethylsiloxane–divinylbenzene, or polymethacrylate. Unlike liquid–liquid extraction, described above, the aim is not to extract all the analyte from the sample as this is an equilibrium extraction; indeed the concentration of the sample may be little affected if only a very small proportion of drug is removed. Obvious attractions of SPME are that no solvents are required and all the material that is extracted by the probe is injected for analysis.

When acids or bases are extracted, it is normally necessary to buffer the pH of the sample to increase the proportion of unionized drug to optimize the extraction. Salting out may be employed to increase the proportion of drug extracted. Differences in the ionic strength of urine samples may lead to variable recovery unless the samples are brought to similar ionic strength by addition of excess salt. Increasing the temperature usually reduces the time to reach equilibrium but will also reduce the amount extracted. Potential disadvantages include competition between drug and endogenous compounds for the fiber, particularly when the mechanism is adsorption rather than partitioning. A well-chosen internal standard (preferably isotopically labeled) may be required to ensure that the method is sufficiently robust. It is important to define the time required for equilibrium, and protein binding is likely to have more of an effect on recovery than when practically all the analyte is extracted into an organic phase. Plasma protein binding not only reduces the amount of



**Figure 7** Time-course for the extraction of lignocaine (lidocaine) from plasma samples after buffering 1:1 to pH 9.5 (A) with and (B) without deproteinization. The solid lines represent the least-squares fit to Peak Area = Maximum Peak Area  $(1 - \exp(-kt))$ . Note how protein binding affects both the rate of attainment of equilibration and the maximum amount that can be extracted (broken lines). (Reprinted with permission from Koster EH, Wemes C, Morsink JB, and de Jong GJ (2000) Determination of lidocaine in plasma by direct solid-phase microextraction combined with gas chromatography. *Journal of Chromatography B* 739: 175–182; © Elsevier.)

analyte extracted, it also increases the time for equilibration. When the data of Figure 7 are fitted to exponential curves,

$$X_t = X_m[1 - \exp(-kt)]$$

where  $X_t$  is the amount extracted at time  $t$ ,  $X_m$  is the maximum amount that can be extracted at equilibrium, and  $k$  is the first-order rate constant, then the half-lives (derived from  $t_{1/2} = 0.693/k$ ) with and without deproteinization are 6.1 and 13.8 min, respectively. Because  $>95\%$   $X_m$  is attained in five half-lives, with deproteinization equilibration was reached in  $\sim 30$  min, whereas to reach the same degree of equilibration in the nondeproteinized sample would take 70 min. If the time for desorption and analysis is similar to, or longer than, the time required for extraction, then one sample can be extracted while the previous one is being analyzed.

The high initial cost may deter some laboratories from automation of sample preparation, but whether automated or not, liquid–liquid extraction and SPE are almost certainly going to remain the most widely applied sample preparation techniques for the foreseeable future. A good understanding of the principles involved will ensure that they are applied in the most appropriate way.

**See also:** Blood and Plasma. Pharmaceutical Analysis: Overview; Drug Purity Determination; Stability Testing.



## Further Reading

- Andersson LI (2000) Molecular imprinting for drug bioanalysis. A review on the application of imprinted polymers to solid-phase extraction and binding assay. *Journal of Chromatography B* 739: 163–171.
- Kataoka H (2003) New trends in sample preparation for clinical and pharmaceutical analysis. *Trends in Analytical Chemistry* 22: 232–244.
- Liu RH and Gadzala (eds.) (1997) *Handbook of Drug Analysis: Applications in Forensic and Clinical Laboratory*. Washington, DC: American Chemical Society.
- McDowall RB (1989) Sample preparation for biomedical analysis. *Journal of Chromatography* 492: 3–58.
- Moffat T, Osselton D, and Widdop B (eds.) (2003) *Clarke's Isolation and Identification of Drugs*. London: Royal Pharmaceutical Society.
- Pawliszyn J (ed.) (2002) *Sampling and Sample Preparation for Field and Laboratory*. Amsterdam: Elsevier.
- Poole CF and Wilson ID (2000) Foreword. *Journal of Chromatography A* 885: 1. (Special volume devoted to solid-phase extraction.)
- Schill G (1978) *Separation Methods for Drugs and Related Organic Compounds*. Stockholm: Apotekarsocieteten.
- Stevenson D (2000) Immuno-affinity solid-phase extraction. *Journal of Chromatography B* 745: 39–48.
- Whelpton R (1989) Iterative least squares fitting of pH-partition data. *Trends in Pharmaceutical Sciences* 10: 182–183.
- Whelpton R and Hurst PR (1988) Solid-phase extraction: personal experiences. In: Reid E, Robinson JD, and Wilson ID (eds.) *Bioanalysis of Drugs and Metabolites Especially Anti-inflammatory and Cardiovascular*, pp. 289–294. New York: Plenum.
- Whelpton R, Dudson P, Cannell H, and Webster K (1990) Determination of prilocaine in human plasma samples using high-performance liquid chromatography with dual-electrode electrochemical detection. *Journal of Chromatography B* 526: 215–222.

## Plant Extracts

A R Bilia, University of Florence, Florence, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Beat Meier, pp. 3882–3892, © 1995, Elsevier Ltd.

## Pharmaceutical Plant Extracts

The role of herbal medicinal products (HMPs) is growing not only in self-medication but also in medical practice, and the expectations related to their quality, safety, and efficacy, which represent the precondition for obtaining an authorization for marketing as medicinal products according to Directive 2001/83/EC, have increased hugely.

The first 'Guidelines for the assessment of herbal medicines' were published by the World Health Organization (WHO) in 1991, handing down the basic principles for the analysis of plant preparations (herbal drug preparations, HDPs) and the finished products prepared from these (HMPs): "A method for identification and, where possible, quantification of the plant material in the finished product should be defined. If the identification of an active principle is not possible, it should be sufficient to identify a characteristic substance or mixture of substances (e.g., 'chromatographic fingerprint') to ensure consistent quality of the preparation."

Subsequently, the European Commission and the European Agency for the Evaluation of Medicines (EMA) pushed forward harmonization of the

assessment criteria for herbal medicinal products. The Herbal Medicinal Products Working Party (HMPWP) of the EMA developed the Note for Guidance 'Quality of Herbal Medicinal Products', adopted by the Committee for Proprietary Medicinal Products (CPMP) in 2001, which includes the specifications for the profile and stability of the constituents.

Furthermore, the European Pharmacopoeia has up to now published 150 monographs on herbal drugs (HDs) and HDPs, including general monographs on extracts. They are classified:

- In terms of their physical state (liquid extracts including fluid extracts, tinctures, oil macerates, semisolid preparations called soft extracts, and dry extracts).
- In terms of the question of whether they only contain herbal extractable matter (native extracts) or additional excipients (not-native extracts = extract preparations).
- In terms of the knowledge relating to their therapeutically active constituents (adjusted/standardized, quantified, other extracts). Refined extracts are also reported and defined by the production process.

Standardized extracts contain constituents (single or groups) that are solely responsible for the acknowledged and documented therapeutic activity and the adjustment to a defined content is acceptable (e.g., standardized Senna leaf dry extract, Aloe dry extract, Belladonna leaf dry extract, Liquorice

## Further Reading

- Andersson LI (2000) Molecular imprinting for drug bioanalysis. A review on the application of imprinted polymers to solid-phase extraction and binding assay. *Journal of Chromatography B* 739: 163–171.
- Kataoka H (2003) New trends in sample preparation for clinical and pharmaceutical analysis. *Trends in Analytical Chemistry* 22: 232–244.
- Liu RH and Gadzala (eds.) (1997) *Handbook of Drug Analysis: Applications in Forensic and Clinical Laboratory*. Washington, DC: American Chemical Society.
- McDowall RB (1989) Sample preparation for biomedical analysis. *Journal of Chromatography* 492: 3–58.
- Moffat T, Osselton D, and Widdop B (eds.) (2003) *Clarke's Isolation and Identification of Drugs*. London: Royal Pharmaceutical Society.
- Pawliszyn J (ed.) (2002) *Sampling and Sample Preparation for Field and Laboratory*. Amsterdam: Elsevier.
- Poole CF and Wilson ID (2000) Foreword. *Journal of Chromatography A* 885: 1. (Special volume devoted to solid-phase extraction.)
- Schill G (1978) *Separation Methods for Drugs and Related Organic Compounds*. Stockholm: Apotekarsocieteten.
- Stevenson D (2000) Immuno-affinity solid-phase extraction. *Journal of Chromatography B* 745: 39–48.
- Whelpton R (1989) Iterative least squares fitting of pH-partition data. *Trends in Pharmaceutical Sciences* 10: 182–183.
- Whelpton R and Hurst PR (1988) Solid-phase extraction: personal experiences. In: Reid E, Robinson JD, and Wilson ID (eds.) *Bioanalysis of Drugs and Metabolites Especially Anti-inflammatory and Cardiovascular*, pp. 289–294. New York: Plenum.
- Whelpton R, Dudson P, Cannell H, and Webster K (1990) Determination of prilocaine in human plasma samples using high-performance liquid chromatography with dual-electrode electrochemical detection. *Journal of Chromatography B* 526: 215–222.

## Plant Extracts

A R Bilia, University of Florence, Florence, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Beat Meier, pp. 3882–3892, © 1995, Elsevier Ltd.

## Pharmaceutical Plant Extracts

The role of herbal medicinal products (HMPs) is growing not only in self-medication but also in medical practice, and the expectations related to their quality, safety, and efficacy, which represent the precondition for obtaining an authorization for marketing as medicinal products according to Directive 2001/83/EC, have increased hugely.

The first 'Guidelines for the assessment of herbal medicines' were published by the World Health Organization (WHO) in 1991, handing down the basic principles for the analysis of plant preparations (herbal drug preparations, HDPs) and the finished products prepared from these (HMPs): "A method for identification and, where possible, quantification of the plant material in the finished product should be defined. If the identification of an active principle is not possible, it should be sufficient to identify a characteristic substance or mixture of substances (e.g., 'chromatographic fingerprint') to ensure consistent quality of the preparation."

Subsequently, the European Commission and the European Agency for the Evaluation of Medicines (EMA) pushed forward harmonization of the

assessment criteria for herbal medicinal products. The Herbal Medicinal Products Working Party (HMPWP) of the EMA developed the Note for Guidance 'Quality of Herbal Medicinal Products', adopted by the Committee for Proprietary Medicinal Products (CPMP) in 2001, which includes the specifications for the profile and stability of the constituents.

Furthermore, the European Pharmacopoeia has up to now published 150 monographs on herbal drugs (HDs) and HDPs, including general monographs on extracts. They are classified:

- In terms of their physical state (liquid extracts including fluid extracts, tinctures, oil macerates, semisolid preparations called soft extracts, and dry extracts).
- In terms of the question of whether they only contain herbal extractable matter (native extracts) or additional excipients (not-native extracts = extract preparations).
- In terms of the knowledge relating to their therapeutically active constituents (adjusted/standardized, quantified, other extracts). Refined extracts are also reported and defined by the production process.

Standardized extracts contain constituents (single or groups) that are solely responsible for the acknowledged and documented therapeutic activity and the adjustment to a defined content is acceptable (e.g., standardized Senna leaf dry extract, Aloe dry extract, Belladonna leaf dry extract, Liquorice

ethanolic liquid extract, Ipecacuanha tincture, etc.) using inert excipients or preparations with a higher or lower content. Quantified extracts are those containing chemically defined constituents (single or groups) possessing relevant pharmacological properties (active markers, e.g., the extracts of ginkgo and St John's wort). 'Other extracts' are those containing no constituents documented as being determinant or relevant for efficacy, or as having pharmacological or clinical relevance. They may be used for control purposes (e.g., *Matricaria* liquid extract, Valerian or Stinging nettle root dried extract).

## Pharmaceutical Aspects

The constituents of the extracts are generally represented by a complex mixture of compounds and it is practically impossible within quality assurance to take account of all constituents of an herbal extract, apart from the fact that in most cases the complete spectrum of constituents is far from being known or analytically detectable. Reproduction and control of a defined spectrum of constituents can only be based on those substances or groups of substances that, according to the state of scientific knowledge, are pharmaceutically, pharmacologically, or toxicologically relevant.

As a result, obtaining reliable chromatographic fingerprints that chemically represent pharmacologically active and characteristic components is not trivial work and determining their pharmaceutical quality can be a difficult and complicated process. The quality of the chromatographic fingerprint obtained is closely dependent on the degree of chromatographic separation and concentration distribution of all chemical components in the investigated extract.

If the active principle or the group of active substances in the extract has been identified – such as reserpine in *rauwolfiae* extract or anthraquinones in senna extract – then the other substances contained are, at least from the point of view of the analyst, considered to be of secondary importance. These extracts have a permitted variability of the defined range of constituents and it should not exceed a numeric difference of  $\pm 5\%$  (more than  $\pm 5\%$  in rare cases).

Very often, however, the situation is more complicated in that many unrelated substances (single or groups) may be involved in the pharmacological efficacy of a particular extract. Flavonoids, sesquiterpenes (bilobalide), and diterpenes (ginkgolides) are all endowed with the biological activity of ginkgo extracts, while flavonols, naphthodianthrone, and phloroglucinols are related to the activity of St John's

wort extract. These extracts should generally be quantified by specifying a relatively wide spectrum of qualitative and quantitative aspects of the contents. The ratio of the weight of the processed drug and the weight of the resulting extract is of particular importance; the dose of the extract is frequently determined with a view to the efficacy of the original HD. Therefore, the standardization of extracts is not merely an analytical tool, but also involves the whole production process.

As plant extracts are pharmaceutical substances, they are subject to the same strict controls as other medications. Particularly high demands are placed on the validation of analytical methods.

Further problems are caused by the need for precision and accuracy. A method for determining accuracy would be to add a known quantity of the natural leading principle, but often this is simply not possible as the compound is not available in sufficient quantities.

Today, analytical chemistry offers the opportunity to investigate these problems in different ways using: chromatographic techniques, such as thin-layer chromatography (TLC), high-performance liquid chromatography (HPLC), or gas chromatography (GC); electrophoretic techniques, such as capillary electrophoresis (CE); or spectroscopic methods, such as mass spectrometry (MS), nuclear magnetic resonance (NMR), ultraviolet (UV), infrared (IR), or near-infrared (NIR).

However, the complete analysis of all the constituents of an extract such as the case of the flavonol glycosides in ginkgo extract would require the evaluation of more than 30 peaks in the chromatogram; thus, it is used in practice only for the qualitative verification of stability using comparative chromatograms. Simpler methods have been developed for day-to-day quality control to quantify the compounds. Thus, glycosides of ginkgo extract are hydrolyzed to form the three aglycones kaempferol, quercetin, and isorhamnetin, which are quantified by HPLC–UV detection at 370 nm. This procedure is also reported in the draft monographs on ginkgo leaf for the United States and European Pharmacopoeias.

## Extraction and Cleanup Procedures

Extraction, sample cleanup, separation, and detection of constituents should be described in detail and thoroughly validated, but unfortunately analytical procedures are rarely published as the manufacturers usually only make their methods known to the pharmaceutical licensing authorities.

The extraction procedures used most frequently in phytoanalysis are solid-liquid extraction (SLE),

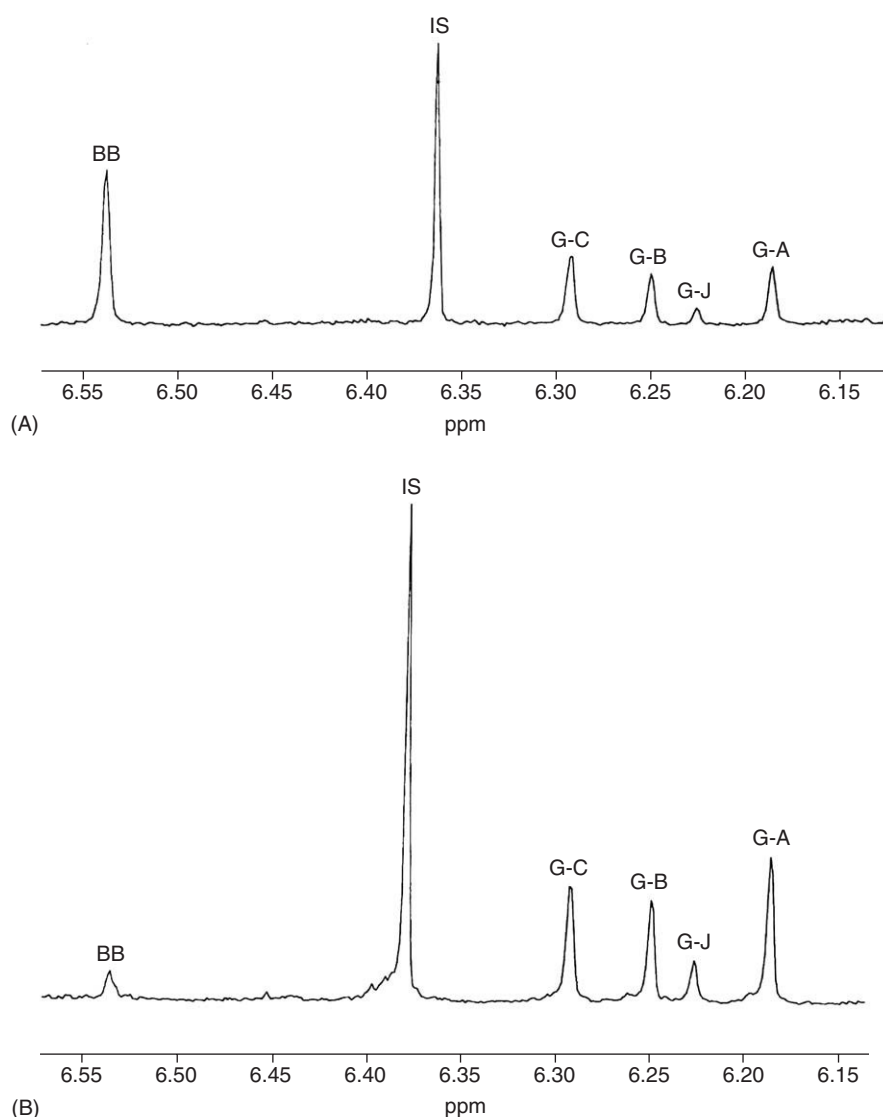
solid-phase extraction (SPE), liquid–liquid extraction (LLE), and rarely reported supercritical fluid extraction (SFE). SLE solvents include methanol, ethanol, acetone, water, ethyl acetate, propanol, dimethylformamide, dichloromethane, and their combinations, while LLE needs two immiscible phases, such as ethylacetate/water or hexane/water.

SPE is used both for preconcentration and cleanup of analytical samples. It offers a great variety of different sorbents, which are based on silica, polystyrene, and pyrrolidone polymers and copolymers. SFE could be useful for the cleanup of essential oils before GC analysis or for less polar constituents such as ginkgo terpenes and kavalactones from kava-kava.

The advantages of this purification are reproducibility and automation.

### Spectroscopic and Spectrometric Methods of Analysis

UV–visible spectrophotometry has long been used in the analysis of extracts and is the preferred technique described in pharmacopoeias. The main problem is its lack of selectivity, but this can be turned into an advantage, especially when the identification of the chemical group to which the substance belongs makes it – in rare cases – possible to evaluate efficacy as in the cases of anthraquinones in the extracts



**Figure 1** Relevant parts of 200 MHz NMR spectra of (A) a sample of ginkgo leaves and (B) a sample of a pharmaceutical product. Solvent: acetone- $d_6$ –pyridine- $d_5$ –methanol- $d_4$  (18:6:1); internal standard: maleic acid; 128 scans. (Reproduced with permission from van Beek TA, von Veldhuizen A, Lelyveld GP, Piron I, and Lankhorst PP (1993) *Phytochemical Analysis* 4: 261–268; © Wiley.)

of laxative HDs or anthocyanins in bilberry extracts. This method avoids time-consuming separation and at the same time gives a significant spectrum of the whole fingerprint.

UV-visible spectrophotometry is still important for LC and CE detection of numerous important naturally occurring groups of substances, such as flavonoids, phenolic acids, anthraquinones, and coumarins, because they have very characteristic UV spectra.

NIR can also play an important role in phytoanalysis. There are some reports on the use of NIR for the quantitative analysis of water content, residual solvents of dry extracts, as well as for the analysis of constituents. It has been used in the analysis of polyphenols (wine, soy) and a quantitative NIR reflectance spectroscopy method was established for the determination of two major constituents of St John's wort (hyperforin and biapigenin) using HPLC as reference method. It was also successfully used for the rapid evaluation and quantitative analysis of essential oils using GC as reference method. It can be considered a rapid and highly effective alternative method to conventional quantitative analysis of plant extracts.

NMR is not frequently used for the direct analysis of plant extracts.  $^{13}\text{C}$  NMR spectroscopy, for example, is sometimes used in the identification of essential oils. Quantitative  $^1\text{H}$  NMR procedures for the determination of constituents are rarely reported but

could be truly useful because the selectivity, reproducibility, and sensitivity are comparable with HPLC, besides the great advantage that no reference substances are needed. Quantification of bilobalide and ginkgolides in ginkgo extracts and commercial phytopharmaceuticals (as reported in Figure 1), and studies on extracts of kava-kava, arnica, and St John's wort have been performed.

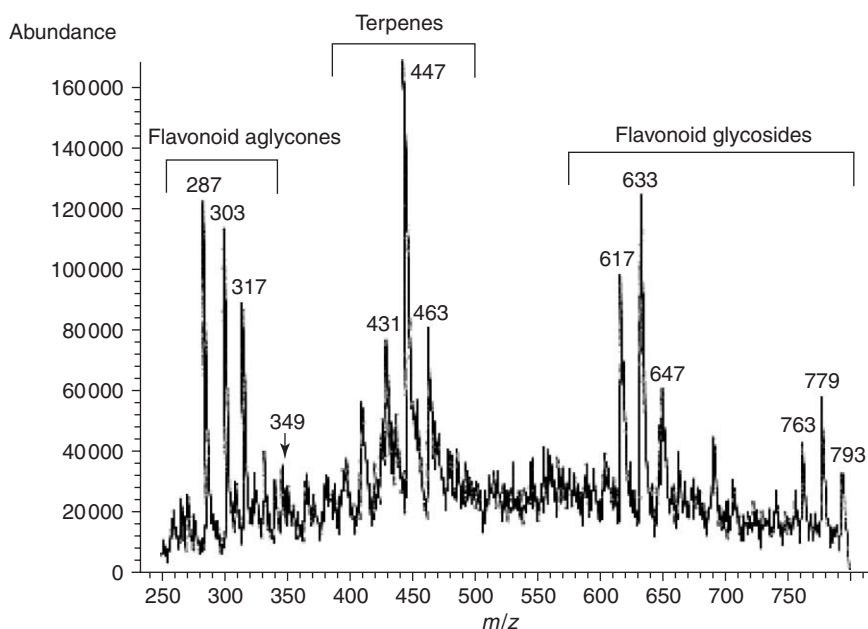
MS (electrospray ionization (ESI)-MS and matrix-assisted laser desorption ionization time-of-flight) can also be used as a qualitative analytical tool for the analysis of plant extracts. A typical sample of ESI-MS analysis of ginkgo extract is reported in Figure 2.

## Chromatographic Methods of Analysis

The development of chromatographic methods and especially of techniques coupled with different detectors during the last two decades has greatly aided the analysis of extracts.

### Capillary Gas Chromatography

Capillary gas chromatography has been found to be the most selective technique for the analysis of essential oils. This technique is now recommended in the European Pharmacopoeia as the standard method to analyze essential oils. The recommended stationary phase is poly(ethylene glycol) 20000. GC



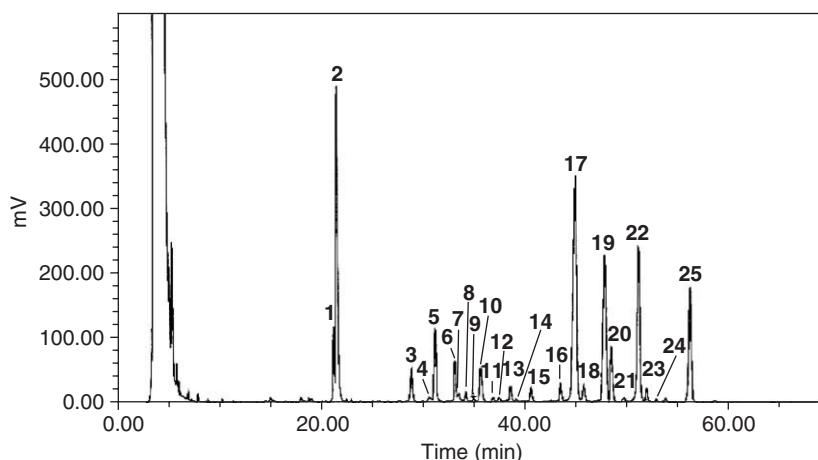
**Figure 2** Positive ion mass spectrum of a ginkgo standardized extract introduced by direct infusion into an ESI-MS. Peaks at  $m/z$  349, 431, 447, and 463 correspond with  $[\text{M} + \text{Na}]$  for bilobalide, G-A, G-B, G-J, and G-C, respectively. (Reproduced with permission from Mauri P, Migliazza B, and Pietta P (1999) *Journal of Mass Spectrometry* 34: 1361–1367; © Wiley.)



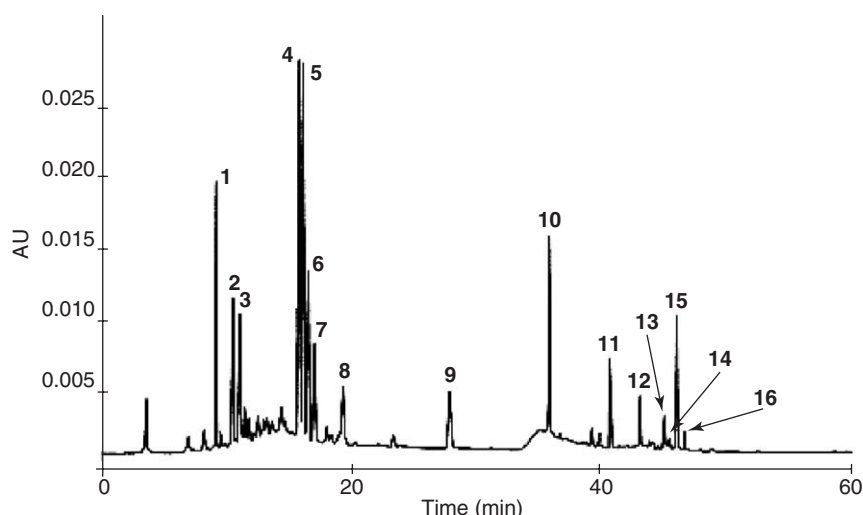
is still underutilized for the determination of polar substances because of the necessity to form derivatives of the constituents. The phenolglycosides in willow bark can be detected as trimethylsilyl derivatives (reagent: trimethylsilylimidazole in pyridine) on an OV-1 fused-silica capillary column with high separation power and selectivity. Ginkgolide derivatives obtained with 1% trimethylchlorosilane in

*N,O*-bis(trimethylsilyl)trifluoroacetamide are separated with a 30-m capillary column coated with 100% dimethyl polysiloxane phase.

However, silylation of nonvolatile constituents represents the main disadvantage, in comparison to HPLC, while separation is at least similar. Detection by flame ionization (FID) surpasses any available LC detector in reproducibility, baseline stability,



**Figure 3** HPLC-ELSD profile of both acidic and neutral ginsenosides. Separation is performed on a Hypersil BDS C18 column using a step gradient of (A)  $8\text{ mmol l}^{-1}$  ammonium acetate, pH 7 with ammonium hydroxide and (B) MeCN. Peaks 1–25 are identified as neutral and malonyl ginsenosides, respectively. Rg<sub>1</sub> (1), Re (2), Ro (3), malonyl-Ra<sub>1</sub>/Ra<sub>2</sub> (4), malonyl-Rb<sub>1</sub> (5), malonyl-Rb<sub>2</sub>/Rb<sub>3</sub>/Rc (6), malonyl-Rb<sub>1</sub> (7), malonyl-Ra<sub>1</sub>/Ra<sub>2</sub> (8), malonyl-Rb<sub>2</sub>/Rb<sub>3</sub>/Rc (9), malonyl-Rb<sub>2</sub>/Rb<sub>3</sub>/Rc (10), Ra<sub>3</sub> (11), malonyl-Rb<sub>2</sub>/Rb<sub>3</sub>/Rc (12), Rf (13), malonyl-Rb<sub>2</sub>/Rb<sub>3</sub>/Rc (14), malonyl-Rd Notoginsenoside R<sub>2</sub>/F3 (15), malonyl-Rd Ra<sub>1</sub>/Ra<sub>2</sub>/isomer (16), Rb<sub>1</sub> (17), Rg<sub>2</sub> (18), Rc (19), Ra<sub>1</sub>/Ra<sub>2</sub>/isomer (20), Ra<sub>1</sub>/Ra<sub>2</sub>/isomer (21), Rb<sub>2</sub> (22), Rb<sub>3</sub> (23), Ra<sub>1</sub>/Ra<sub>2</sub>/isomer (24), Rd (25). (Reproduced with permission from Fuzzati N, Gabetta B, Jayakar K, Pace R, and Peterlongo F (1999) *Journal of Chromatography A* 854: 69–79; © Elsevier.)



**Figure 4** HPLC-DAD profile of the constituents of St John's wort extract. Separation is performed on a 201 TP 54 RP-18 Vydac column using three solvents (A = water–85% phosphoric acid (99.7:0.3 v/v); B = AcCN; C = methanol) in a linear gradient program. Peaks are detected at 270 nm and identified as caffeoyl quinic derivatives (1–3), rutin (4), hyperoside (5), isoquercitrin (6), 3,3',4',5,7-pentahydroxyflavanone 7-*O*-rhamnopyranoside (7), quercitrin (8), quercetin (9), I3, I18 biapigenin (10), pseudohypericin (11), hypericin (12), hyperforin analogs (13, 14), hyperforin (15), adhyperforin (16). (Reproduced with permission from Brolis M, Gabetta B, Fuzzati N, Pace R, Panzeri F, and Peterlongo F (1998) *Journal of Chromatography A* 825: 9–16; © Elsevier.)



ruggedness, cost, and sensitivity. Instead of FID, two other detectors can be used: electron capture (ECD) and MS. However, the slight increase in selectivity and sensitivity of ECD compared with FID is offset by the problems associated with this detector: not generally available, limited linear range, and its inherent radioactivity. GC-MS is only necessary when the concentrations are very low such as in the analysis of biological fluids.

### Liquid Chromatography

The development of numerous derivatives of silica gel with an apolar surface (reversed phase (RP)) as a stationary phase has revolutionized extract analysis and is still considered the most straightforward technique for high-melting, nonvolatile compounds.

In most cases, gradient elution is vital, as even small structural variations will cause considerable differences in retention times.

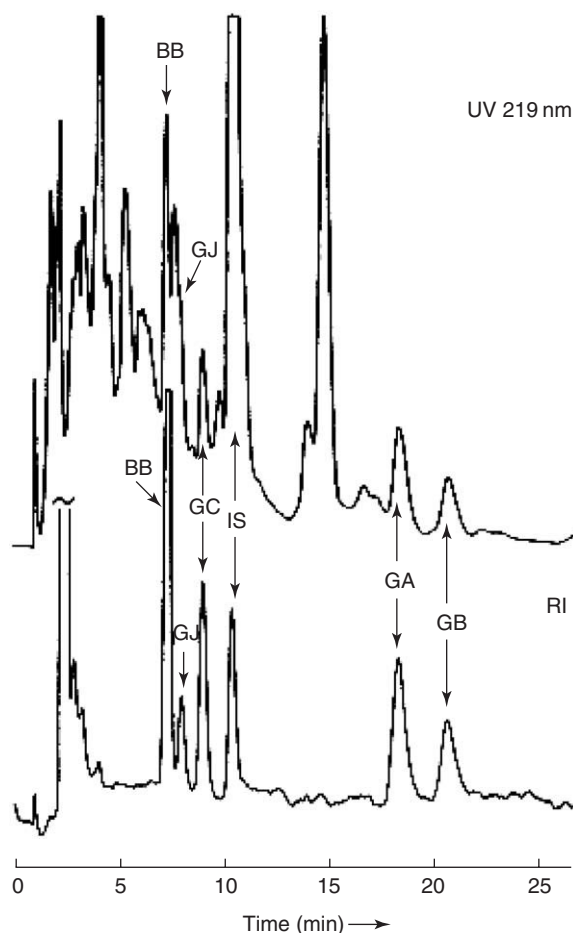
Spectrophotometric detection is generally used and the choice of solvents is limited to those with low absorption properties. Especially, acetonitrile, tetrahydrofuran, methanol, or isopropanol and, less commonly, dioxan (which can be used for the separation of cardiac glycosides and derivatives of amino acids) have been found to be sufficiently selective in most cases. Water is used to adjust solvent strength and the addition of acid (preferably orthophosphoric acid, although acetic and formic acids can be used, especially if MS detection is hyphenated) or buffer in RPLC will decrease peak tail. A typical sample of the use of an ammonium buffer is reported for the HPLC analysis of both neutral and acid ginsenosides (Figure 3).

Phenolic compounds and constituents with conjugate systems can be easily detected by UV because they have a maximum wavelength in the range 250–350 nm and some of them (anthocyanins) reach values in the range 450–550 nm. Problems caused by differences in the wavelength for maximum UV absorption by individual constituents such as in St John's wort can be solved by using diode array detection (DAD). In the developed HPLC-DAD method, the identification and quantification of 16 compounds of four different classes of constituents – flavonoids, phenolic acids, naphthodianthrones, and phloroglucinols – can be quantified (Figure 4).

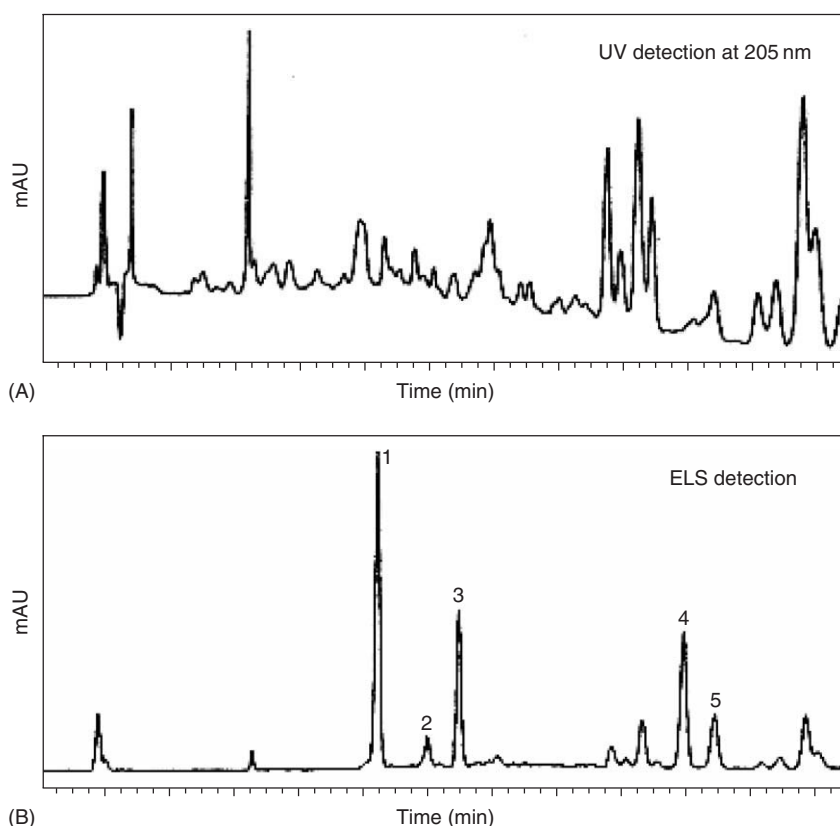
UV detection of compounds having very low values around the nonselective wavelength of 210–220 nm is a poor choice and a much better approach is, therefore, a detector that shows less variation in response factors, e.g., refractive index (RI) detection or evaporative light scattering (ELS) detection. Examples of UV/RI and UV/ELS detection of ginkgolides are given in Figures 5 and 6, respectively.

Advantages of ELS detection over RI are better baseline stability, compatibility with THF and gradients, small solvent peak, and greater sensitivity. Advantages of RI over ELSD are larger linear range, lower costs, and its broader availability.

The only other LC detection technique used is MS. Atmospheric pressure chemical ionization and ESI interfaces are also employed in many extract analyses. These are highly sensitive and stable. However, although the MS detectors offer a high selectivity and sensitivity, due to their high cost and more complicated operation and maintenance they remain a method of choice for clinical studies of the metabolites of natural constituents.



**Figure 5** HPLC profiles of a purified ginkgo preparation with (upper trace) UV detection at 219 nm and (lower trace) RI detection. Separation is performed on a C-18 SPE column using a mixture of methanol–water (33:67) as eluent. BB is biolbalide, GJ, GC, GA, and GB are, respectively, ginkgolide J, ginkgolide C, ginkgolide A, and ginkgolide B. Internal standard (IS) is benzyl alcohol. (Reprinted with permission from van Beek TA, Scheeren HA, Rantio T, Melger WC, and Lelyveld GP (1991) *Journal of Chromatography A* 543: 375–387; © Elsevier.)



**Figure 6** HPLC–UV and HPLC–ELSD profiles of a filtered methanolic extract of a ginkgo drug containing standardized extract. Separation is performed on a Phenomenex Synergi Max-RP 80 column using two solvents (A = 10 mmol l<sup>-1</sup> NH<sub>4</sub>OAc adjusted to pH 5; B = MeOH–iBuOH (9:1)) in a linear gradient program. Peaks 1–5 were identified as bilobalide, ginkgolide J, ginkgolide C, ginkgolide A, and ginkgolide B, respectively. (Reproduced with permission from Ganzera M, Zhao J, and Khan IA (2001) *Chemical Pharmaceutical Bulletin* 49: 1170–1173, The Pharmaceutical Society of Japan.)

### Supercritical Fluid Chromatography

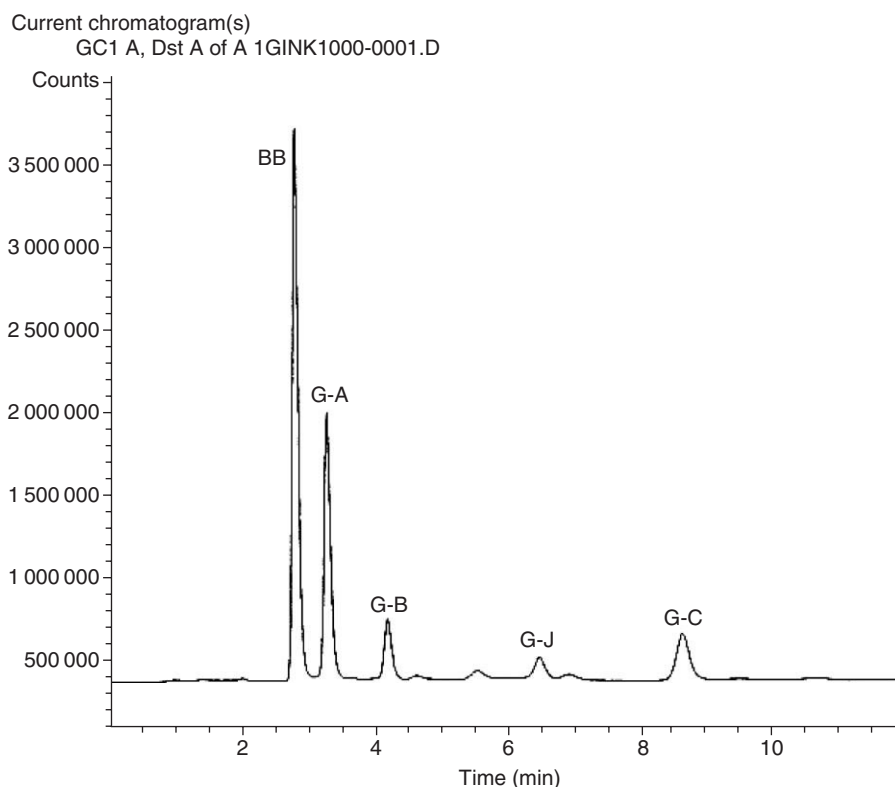
Supercritical fluid chromatography is also capable of separating constituents of extracts, often with a higher selectivity than RPLC. There is a recent report on the analysis of ginkgolides on a deactivated aminopropyl silica column using 12% methanol in carbon dioxide as fluid with a baseline separation of bilobalide and all four ginkgolides achieved within 9 min. Detection was facilitated by an ELS detector and a detection limit of ~10 ng was reported without sample cleanup (Figure 7). Thus, SFC can be used as an alternative for the analysis of extracts even if in the near future it is unlikely to replace either HPLC or GC in the average quality control laboratory.

### Thin-Layer Chromatography

TLC is an essential aid to extract analysis and is commonly used for qualitative assays. As the planar method gives ‘two-dimensional’ results, it is ideal for the comparison of the raw material and the end-products. With a few notable exceptions, the main

goal of extract manufacturing is to ensure that substances extracted from the biological matrix reoccur in unchanged form in the resultant extract (see Figure 8). TLC is thus used primarily for identification purposes. G60 silica gel on glass or aluminum plates is widely recommended as the stationary phase.

It is instructive to concentrate on the flavonoids, as these are of primary importance in qualitative assays of several HDs. Silica gel as stationary phase and ethyl acetate–ethyl methyl ketone–formic acid–water (50:30:10:10 by volume) as mobile phase is used in the analysis of flavonoids with C-glycosides (hawthorn, passion flower), and a mobile phase of ethyl acetate–formic acid–acetic acid–water (100:11:11:27 by volume) is used as the basic solvent for O-glycosides (arnica, calendula, chamomille). Following spraying of the plates with 2-aminoethyldiphenylborinate (NTS, 1% in methanol) and subsequent treatment with poly(ethylene glycol) 4000 (PEG, 5% in ethanol), flavonoids display yellow to green fluorescence under UV irradiation at 365 nm. The frequently observed blue spots represent phenol



**Figure 7** SFC-ELSD profile of a methanol solution of a standardized ginkgo extract without any prior sample cleanup. Separation is performed on a Deltabond PEG column using CO<sub>2</sub> with 12% modifier (MeOH) at 280 atm and 40°C. BB is biolbalide, GJ, GC, GA, and GB are, respectively, ginkgolide J, ginkgolide C, ginkgolide A, and ginkgolide B. (Reprinted with permission from Thompson J, Strode JBT III, Taylor LT, and van Beek TA (1996) *Journal of Chromatography A* 738: 115–122; © Elsevier.)

carboxylic acids, which often occur in HDs that are rich in flavonoids.

Quantitative analysis with the aid of densitometry is not yet in general use in extract analysis. Various problems are present as the separation potential of the thin layer is limited.

### Capillary Electrophoresis

CE is a relatively new separation technique compared to GC and HPLC. It offers the advantages of rapid, high-resolution separation. The instrumentation format of CE is similar to HPLC and therefore most of the detectors used for HPLC can be adapted to CE.

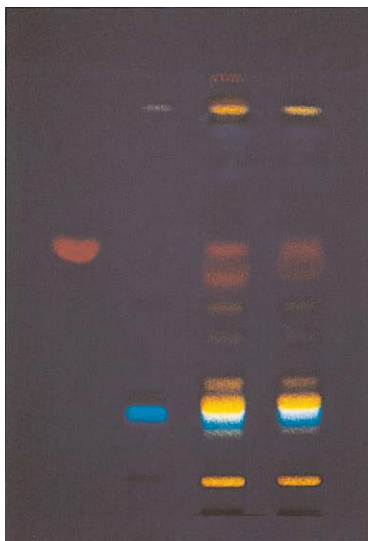
Owing to the rising cost of solvent disposal, CE is becoming a more attractive method for extract analysis. The method makes it easy to separate ionizable substances using minimal amounts of solvents. On-column detection removes the coupling problems, and the separation power is comparable to that obtained in capillary gas chromatography. If alkaloids have to be analyzed, this method is currently the most suitable.

*Uncaria tomentosa* alkaloids can be separated using a phosphate buffer with a pH of 5–6 using 10 kV. Fused-silica capillary tubes with an internal diameter of 50–100 µm and a length of approximately 1 m are positioned in the electric field and loaded with the sample. Recently, CE has also been used for quantitative determination of glycyrrhizinic acid in pharmaceutical preparations using borate buffer at a pH of 10 using 25 kV. The method has also been used for the analysis of flavonoids and phenolic acids of echinacea extracts (see **Figure 9**), where the charge required for transport in the electrical field was produced using a micelle (micellar electrokinetic chromatography). Sodium dodecyl sulfate micelles are prepared in a sodium borate buffer for the separation of the various polyphenols. High-resolution separation and versatility of CE are reported in all studies.

### Stability Problems

It is impossible to make generalized statements about the stability of plant extracts: each has its own individual properties in this regard and, above all,

stability is related to chemical, microbiological, and physical stability. However, according to HMPWP guidelines, 'Quality for HMPs', the acceptable ranges of the constituents, related to the initial value are  $\pm 5\%$  if the product contains an HD or HDP with

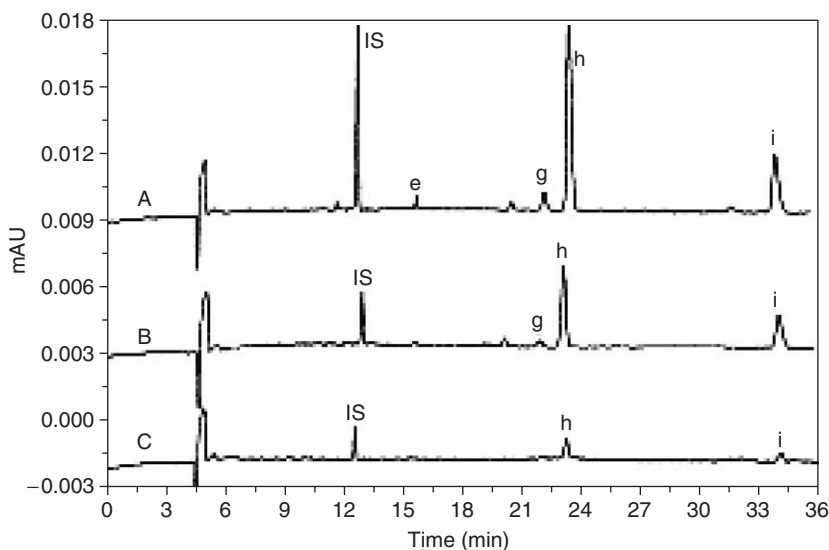


**Figure 8** Thin-layer chromatography of *Hypericum perforatum*. The chromatograms show no important differences between the herbal drug and the extract except the red spots in front. Both spots are the result of chlorophyll, which is eliminated by the extraction procedure. Red spots: Hypericin and pseudohypericin; yellow and orange spots: flavonoids (glycosides R, 0.5; aglycones R, 0.8); blue spot: ubiquitous plant acids. Reference traces on the left: Hypericin; chlorogenic acid; and quercetin. (Reproduced with permission from Beat Meier, Zeller AG, Herbal Remedies, Romanshorn, Switzerland.)

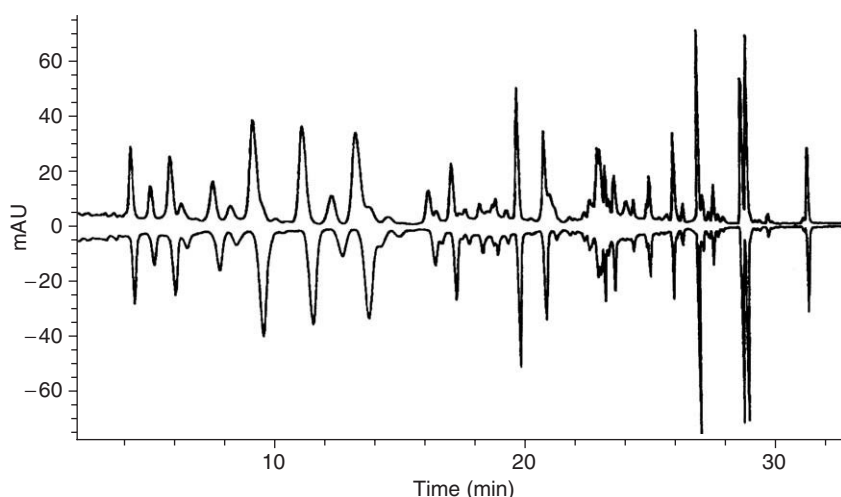
known active constituents, and  $\pm 10\%$  if the active constituents are not known.

Generally, glycosides undergo hydrolysis to become the corresponding aglycones, so it is possible to evidence the transformation over time by appropriate fingerprinting making a comparison between the stored finished product and the dry extract. A typical example is represented by the complex mixture of flavonoids of ginkgo preparations (Figure 10).

The stability of natural compounds is quite variable. Salicin, in drug form and also in the form of its esters salicortin and tremulacin, is stable over several years in willow bark and extracts. Kavalactones from kava-kava extracts and preparations are stable for several years as well. The fact that essential oils are very volatile means that they are never major constituents in dry extracts unless they are added later in a stabilized form, for example, as microcapsules. Some important plant constituents are very sensitive to temperature. The content of naphodianthrones and phoroglucinols in St John's wort extracts and preparations are very sensitive to light, humidity, and temperature. Alliin in garlic will fall dramatically if the extract is stored at temperatures of  $30^{\circ}\text{C}$ . The second main problem with dry extracts is not so much their chemical stability as their actual physical stability: most dry extracts are hygroscopic and tend to absorb water. Thus, it is essential that they are stored away from heat and moisture. The stability of fluid extracts is somewhat more problematic than that of dry forms; it should also be kept in mind that not only chemical but also enzymatic degradation is



**Figure 9** Representative electropherograms obtained from: A, collected *E. purpurea* roots; B, commercial dried extract of *E. purpurea*; and C, commercial product. MEKC conditions:  $40\text{ mmol l}^{-1}$  sodium tetraborate pH 9.2 with  $70\text{ mmol l}^{-1}$  sodium deoxycholate. Peak identity: *p*-nitrobenzoic acid (IS), vanillic acid (e), caffeic acid (g), cichoric acid (h), and caftaric acid (i). (Reproduced with permission from Pomponio R, Gotti R, Hudaib M, and Cavrini V (2002) *Journal of Chromatography A* 945: 239–247; © Elsevier.)



**Figure 10** Stability test of a ginkgo full extract with fingerprint chromatography. Separation is performed on a Phenomenex Synergi Max-RP 80 column using three solvents (A = water–85% phosphoric acid (95:5 v/v); B = *i*-propanol–THF (27.8:72.2 v/v); C = AcCN) in a linear gradient program. Peaks are detected at 350 nm. After storage for 6 months (at room temperature) no differences in the upper trace (liquid preparation) and the lower trace (standardized extract), with permission from Elsevier Science.

possible. There are enzymes, such as the glycosidases, that are surprisingly robust and can suddenly become active under the most disadvantageous conditions. If stability improves significantly following heat treatment, it is more than likely that enzymes are involved.

However, according to the HMPWP guideline for the evaluation of extract stability “it must be shown, as far as possible, e.g., by means of appropriate fingerprint chromatograms, that other substances present in the herbal drug or in the herbal drug preparation are likewise stable and that their proportional content remains constant.” Thus, unknown substances that very often represent the main part of the extract can undergo degradation so that specific chromatographic methods of analysis may not suffice. For this reason NMR, both using one-dimensional experiments, such as  $^1\text{H}$  and  $^{13}\text{C}$ , and two-dimensional experiments, could be useful to have a general overview of the whole extract and evaluate the stability of the preparation.

## Pharmacological Interpretation of Data

The chromatographic profile should feature the fundamental attributions of ‘sameness’ and ‘differences’ of the constituents of the investigated extract and define their quality in order to get reproducible biological data in terms of safety and efficacy. The analytical methods should be so precise that different batches of the extract in question will have the same efficacy. As described above, it has now become routine to define a ginkgo extract in terms of the content of the two groups, the flavonoid glycosides and

ginkgolides. The quality of optimized chamomille extracts should be defined at least in terms of the flavone apigenin, of chamazulen, and of  $\alpha$ -bisabolol contents, if the preparation is mentioned for the therapeutic activities of *Matricaria chamomilla*.

A great deal of research into potential synergistic effects still needs to emerge. These studies can only be conducted in living organisms and require considerable resources, while results themselves are difficult to interpret. However, there is plenty of evidence to suggest that synergistic reactions do occur. Standard biological *in vitro* and *in vivo* methods (e.g., using isolated cells or microorganisms) for defining plant extract quality are therefore still in the process of being developed. Some current methods use, for example, antiviral activity, inhibition of platelet-activating factor (which induces thrombocyte aggregation in rabbit plasma rich in platelets), inhibition of cyclooxygenase in sheep seminal vesicles and of 5-lipoxygenase in porcine leukocytes, or inhibition of cholesterol biosynthesis in cultured human and rat hepatocytes. There is often a correlation between the characteristics of the active principle as defined by analysis and the postulated pharmacological effects of the HDs. However, as many active agents extracted from plants are prodrugs, and thus first become active following metabolic transformation, it means that experimentation and evaluation of results are particularly difficult in this context.

**See also:** Derivatization of Analytes. Electrophoresis: Overview. Extraction: Solid-Phase Extraction. Infrared Spectroscopy: Near-Infrared. Liquid Chromatography:



Pharmaceutical Applications. **Nuclear Magnetic Resonance Spectroscopy Applications:** Pharmaceutical. **Spectrophotometry:** Pharmaceutical Applications. **Thin-Layer Chromatography:** Overview.

## Further Reading

- Bauer R (1998) Quality criteria and standardization of phytopharmaceuticals: can acceptable drug standards be achieved? *Drug Information Journal* 32: 101–110.
- van Beek TA, von Veldhuizen A, Lelyveld GP, Piron I, and Lankhorst PP (1993) Quantitation of bilobalide and ginkgolides A, B, C and J by means of NMR spectroscopy. *Phytochemical Analysis* 4: 261–268.
- Bilia AR, Bergonzi MC, Morgenni F, Mazzi G, and Vincieri FF (2001) Evaluation of stability of St. John's wort commercial extract and some preparations. *International Journal of Pharmaceutics* 213: 199–208.
- Bilia AR, Bergonzi MC, Lazari D, and Vincieri, FF (2002). Characterisation of commercial kava-kava herbal drug and herbal drug preparations by means of nuclear magnetic resonance spectroscopy. *Journal of Agricultural and Food Chemistry* 50: 5016–5025.
- Bilia AR, Bergonzi MC, Mazzi G, and Vincieri FF (2002) Analysis of plant complex matrices by use of nuclear magnetic resonance spectroscopy: supercritical carbon dioxide arnica extract *Journal of Pharmaceutical and Biomedical Analysis* 30: 321–330.
- CPMP/CVMP Note for guidance on specifications: Test Procedures and Acceptance Criteria for Herbal Drugs, Herbal Drug Preparations and Herbal Medicinal Products (CPMP/QWP/2800/00).
- CPMP/CVPM Note for guidance: Quality of Herbal Medicinal Products. European Agency for the Evaluation of Medicinal Products (CPMP/QWP/2819/00).
- Franz G and Koehler H (1992) *Drogen und Naturstoffe-Grundlagen und Praxis der chemischen Analyse*. New York: Springer-Verlag.
- European Pharmacopoeia (2002) European Department for the Quality of Medicines, Strasbourg, 4th edn. Council of Europe.
- Gaedcke F and Steinhoff B (2003) *Herbal Medicinal Products – Scientific and Regulatory Basis for Development, Quality Assurance and Marketing Authorisation*. Stuttgart: Medpharm Scientific Publishers.
- Stecher G, Huck CW, Stöggel WM, and Bonn GK (2003) Phytoanalysis: a challenge in phytomics. *Trends in Analytical Chemistry* 22: 1–14.
- United States Pharmacopoeia (2003) 26, United States Pharmacopoeial Convention, Rockville, MD. Pharmacopoeial Convention.
- Wagner H and Bladt S (1996) *Drug Analysis – A Thin Layer Chromatography Atlas*. New York: Springer-Verlag.
- WHO (1998) *Quality Control Methods for Herbal Medicinal Plant Material*. Geneva: World Health Organization.

# PHARMACOKINETICS

Contents

**Absorption, Distribution, and Elimination**  
**Pharmacodynamics**

## Absorption, Distribution, and Elimination

I P Nnane, Temple University, Philadelphia, PA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The therapeutic outcome of a drug is influenced by several factors, including the pharmacokinetic and pharmacodynamic profile of the agent in animals and humans. In order for a drug to produce a pharmacological effect, it must come into contact and interact

with a biological system. Consequently, after administration of a drug, it undergoes absorption, distribution, metabolism, and excretion (ADME) in the body. Although absorption is bypassed when a drug is administered intravenously, distribution and elimination of all drugs must occur in the body irrespective of the route of administration. The quantitative study of the kinetics of drug absorption, distribution, and elimination is known as pharmacokinetics. The related discipline of pharmacodynamics evaluates the relationship between the concentration of the drug at the site of action and the pharmacological effect. Increasingly, pharmacokinetic and pharmacodynamic principles are used to relate the concentration of a drug in plasma to the pharmacological response in order to establish the time



Pharmaceutical Applications. **Nuclear Magnetic Resonance Spectroscopy Applications:** Pharmaceutical. **Spectrophotometry:** Pharmaceutical Applications. **Thin-Layer Chromatography:** Overview.

## Further Reading

- Bauer R (1998) Quality criteria and standardization of phytopharmaceuticals: can acceptable drug standards be achieved? *Drug Information Journal* 32: 101–110.
- van Beek TA, von Veldhuizen A, Lelyveld GP, Piron I, and Lankhorst PP (1993) Quantitation of bilobalide and ginkgolides A, B, C and J by means of NMR spectroscopy. *Phytochemical Analysis* 4: 261–268.
- Bilia AR, Bergonzi MC, Morgenni F, Mazzi G, and Vincieri FF (2001) Evaluation of stability of St. John's wort commercial extract and some preparations. *International Journal of Pharmaceutics* 213: 199–208.
- Bilia AR, Bergonzi MC, Lazari D, and Vincieri, FF (2002). Characterisation of commercial kava-kava herbal drug and herbal drug preparations by means of nuclear magnetic resonance spectroscopy. *Journal of Agricultural and Food Chemistry* 50: 5016–5025.
- Bilia AR, Bergonzi MC, Mazzi G, and Vincieri FF (2002) Analysis of plant complex matrices by use of nuclear magnetic resonance spectroscopy: supercritical carbon dioxide arnica extract *Journal of Pharmaceutical and Biomedical Analysis* 30: 321–330.
- CPMP/CVMP Note for guidance on specifications: Test Procedures and Acceptance Criteria for Herbal Drugs, Herbal Drug Preparations and Herbal Medicinal Products (CPMP/QWP/2800/00).
- CPMP/CVPM Note for guidance: Quality of Herbal Medicinal Products. European Agency for the Evaluation of Medicinal Products (CPMP/QWP/2819/00).
- Franz G and Koehler H (1992) *Drogen und Naturstoffe-Grundlagen und Praxis der chemischen Analyse*. New York: Springer-Verlag.
- European Pharmacopoeia (2002) European Department for the Quality of Medicines, Strasbourg, 4th edn. Council of Europe.
- Gaedcke F and Steinhoff B (2003) *Herbal Medicinal Products – Scientific and Regulatory Basis for Development, Quality Assurance and Marketing Authorisation*. Stuttgart: Medpharm Scientific Publishers.
- Stecher G, Huck CW, Stöggel WM, and Bonn GK (2003) Phytoanalysis: a challenge in phytomics. *Trends in Analytical Chemistry* 22: 1–14.
- United States Pharmacopoeia (2003) 26, United States Pharmacopoeial Convention, Rockville, MD. Pharmacopoeial Convention.
- Wagner H and Bladt S (1996) *Drug Analysis – A Thin Layer Chromatography Atlas*. New York: Springer-Verlag.
- WHO (1998) *Quality Control Methods for Herbal Medicinal Plant Material*. Geneva: World Health Organization.

# PHARMACOKINETICS

Contents

**Absorption, Distribution, and Elimination**  
**Pharmacodynamics**

## Absorption, Distribution, and Elimination

I P Nnane, Temple University, Philadelphia, PA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The therapeutic outcome of a drug is influenced by several factors, including the pharmacokinetic and pharmacodynamic profile of the agent in animals and humans. In order for a drug to produce a pharmacological effect, it must come into contact and interact

with a biological system. Consequently, after administration of a drug, it undergoes absorption, distribution, metabolism, and excretion (ADME) in the body. Although absorption is bypassed when a drug is administered intravenously, distribution and elimination of all drugs must occur in the body irrespective of the route of administration. The quantitative study of the kinetics of drug absorption, distribution, and elimination is known as pharmacokinetics. The related discipline of pharmacodynamics evaluates the relationship between the concentration of the drug at the site of action and the pharmacological effect. Increasingly, pharmacokinetic and pharmacodynamic principles are used to relate the concentration of a drug in plasma to the pharmacological response in order to establish the time

course of the drug effect. An understanding of the ADME and pharmacodynamics is essential for a rationale approach to drug discovery and development, and optimization of dosage forms in the clinical setting. In this article, the relevance of ADME in drug discovery and clinical practice is discussed. The application of the open one-compartment pharmacokinetic model to simplify our understanding of the plasma concentration–time profile and current *in vitro* approaches used to study the ADME of drugs are also highlighted.

## Importance of ADME

Successful drug candidates must have the right ADME profiles in addition to desirable pharmacological and toxicological profiles *in vitro* and *in vivo*. Unfavorable ADME profiles contribute significantly to the failure rates of drug candidates during drug discovery. Additionally, ADME studies provide supportive information to facilitate the interpretation of pharmacology and toxicology findings during drug discovery. As a result, pharmaceutical companies are allocating a considerable amount of resources for the evaluation of the ADME of drug candidates in the early stages in order to find the right ADME balance and reduce the failure rates of drug candidates in the later stages of drug discovery. ADME studies allow the identification of drugs that would be absorbed after administration but are also eliminated at acceptable rates, so that a desirable steady-state plasma concentration is attained in a target population when a suitable dosage regimen is given. Furthermore, regulatory authorities are demanding information on ADME profiles and biological activity before new drug applications are approved. Although ADME studies are routinely performed *in vitro* and in animals, increasingly these types of investigations are conducted in humans as well, and the information that is generated is critical in all phases of drug discovery and development programs. Clearly, a full consideration of the factors that influence the plasma concentration–time profiles of drugs *in vivo* is pivotal in the discovery and use of drugs.

## Routes of Drug Administration

Several routes of drug administration, including extravascular (oral, intramuscular, subcutaneous, transdermal, inhalation, etc.) and intravascular routes (intravenous and intra-arterial), are used to deliver drugs into the body. Oral administration of drugs is the most common route because it is

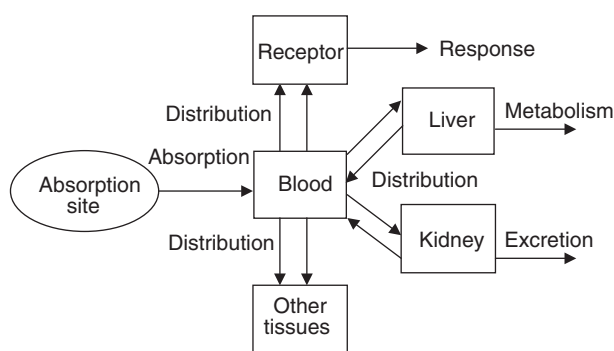
relatively convenient and safe. Many factors – including the physicochemical properties of drugs, type of formulation, physiology of the gastrointestinal tract (GIT), other drugs, and food – affect the way in which drugs are absorbed after oral administration. Although drug absorption occurs in the mouth and stomach, it takes place mostly in the small intestine due to the large surface area. After oral administration, the drug must pass through the intestinal wall and liver in order to reach the general circulation. The drug metabolism enzymes in the intestinal wall and liver metabolize many drugs, thereby decreasing the amount of drug that is bioavailable. This phenomenon is often referred to as first-pass metabolism. The acidic environment and digestive enzymes in the stomach may also chemically degrade some drugs, resulting in erratic absorption. In contrast, drugs administered intravenously do not undergo absorption and therefore the entire dose reaches the general circulation intact.

## Fate of Drugs in the Body

When a drug is administered to a human or an animal, a sequence of events must occur in order for the drug to be effective; the drug must be liberated from the dosage form and dissolved in the fluids at the site of absorption. Several factors influence the rate and extent of dissolution of the drug at the site of absorption, including the solubility of the drug, particle size, and surface area of a solid dosage form. The transfer of the unchanged drug molecules from the site of absorption into the systemic circulation is referred to as absorption. The drug molecules that arrive in the general circulation are reversibly transferred between the blood and other tissues (distribution). At the same time, the drug is eliminated from the body by excretory organs such as the kidney (excretion) or by metabolism to metabolites that are readily excreted. At the site of action, drug molecules interact with receptors to initiate a cascade of events that elicit a pharmacological response (pharmacodynamics). In essence, after entering the body, the drug acts on the body to produce a response, and simultaneously the body acts on the drug to remove it from the system (Figure 1). The processes of ADME directly impact the plasma concentration–time profile of a drug and ultimately its pharmacodynamic effects.

## Absorption

After extravascular administration, a drug must be transferred from the site of absorption, through a cell



**Figure 1** Schematic representation of the fate of a drug in the body.

membrane, in order to enter the systemic circulation. The physicochemical properties of the drug and the structure of the cell membrane influence the passage of the drug from the site of administration into the general circulation. Although there is considerable variation in the structure and function of biological membranes, the basic structure of the cell membrane is a bimolecular lipid layer, which is associated with proteins and small aqueous pores. The transport of drugs across cell membranes is mediated by several mechanisms, including passive diffusion, facilitated transport, active transport, endocytosis, exocytosis, and ion-pair transport. Accumulated evidence indicates that passive diffusion is the most common mechanism of drug absorption. Passive diffusion is driven by lipid solubility and the magnitude of the drug concentration gradient across the cell membrane and governed by Fick's first law according to eqn [1]:

$$\text{flux} = -DA \frac{\Delta C}{h} \quad [1]$$

The flux refers to the flow of drug molecules across an area ( $A$ ) per unit time;  $\Delta C/h$  is the concentration gradient divided by the membrane thickness ( $h$ ), and  $D$  is the diffusion coefficient of the drug. The negative sign indicates that drug transfer occurs from an area of high concentration to one of low concentration. In general, the membrane lipid bilayer is highly permeable to lipid-soluble drugs and poorly permeable to water-soluble drugs. However, some embedded proteins in membranes can form channels, carriers, or pumps that facilitate the movement of polar drugs across membranes. The rate of diffusion of a drug also depends on its molecular size, although the variation with molecular weight is modest for drugs with molecular weights of 200–1000 Da. The degree of ionization of a drug also plays a significant role in the diffusion of the drug across membranes. Since most drugs are either weak acids or bases, their

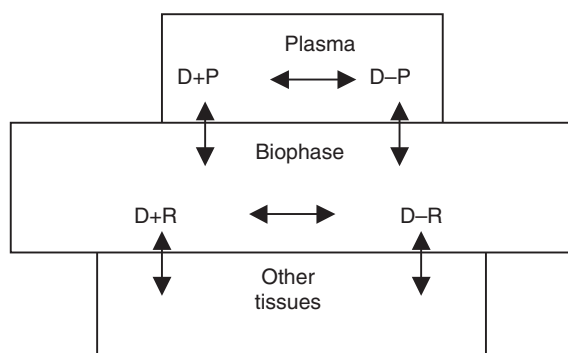
**Table 1** Ionization constants ( $pK_a$ ) of some drugs

<i>Weak bases</i>		<i>Weak acids</i>	
<i>Drug</i>	$pK_a$	<i>Drug</i>	$pK_a$
Chloroquine	10.7	Levodopa	2.2
Amphetamine	9.8	Penicillins	2.5–2.7
Orphenadrine	8.4	Aspirin	3.5
5-Fluorouracil	8.1	Ibuprofen	4.4
Codeine	6.0	Warfarin	5.0
Aminopyrine	5.0	Chlorothiazide	6.8
Diazepam	3.3	Phenobarbital	8.1
Caffeine	0.8	Phenytoin	8.3

degree of ionization will influence their ability to cross biological membranes. The proportion of the drug present in either the ionized or nonionized form depends on the ionization constant,  $pK_a$ , and pH at the absorption site. The  $pK_a$  values of some drugs are given in Table 1. Acidic drugs are predominantly in the nonionized form when the pH value of the environment in which they are present is below their  $pK_a$  values, while basic compounds are predominantly unionized when the pH value of the environment in which they are present is above their  $pK_a$  value. For many drugs, the ionized species exhibits low lipid solubility and poor membrane permeability except when specific carrier mechanisms are in operation. Conversely, the unionized species of most drugs are lipid soluble and readily cross biological membranes. For example, a weak acid such as aspirin will be concentrated in plasma (pH 7.4) with respect to the gastric juice (pH 3.0) because the pH of the stomach favors its absorption. However, it is important to note that the large surface area of the ileum compared to the stomach plays a far more significant role than the pH partition in drug absorption. The factors that influence drug absorption have a significant impact on the overall pharmacokinetics of the drug in question.

## Distribution

After a drug is absorbed into the systemic circulation, it mixes rapidly in the blood and is transferred reversibly into the various tissues and site of action in the body. The rate and extent of distribution of drugs into different tissues depends on their physicochemical properties such as ability to cross cell membranes, affinity of drugs to certain tissues, pH partition, and fat/water partition. In each of the body compartments, drugs exist in both free and bound form (Figure 2). Many acidic drugs that are tightly bound to plasma proteins such as albumin reside



**Figure 2** Schematic representation of drug (D) binding to plasma proteins (P) and receptors (R) in body compartments.

mainly in the bloodstream. For many drugs, it is the free form that is pharmacologically active, because it can readily cross capillary membranes. Therefore, changes in drug binding due to drug–drug interactions or disease conditions may cause significant changes in redistribution and pharmacological effect of a drug in the body. Several drugs do not exhibit a uniform pattern of distribution in the body. For example, thiopental tends to accumulate in body fat, chloroquine concentrates in melanin-containing tissues such as the liver, while small polar drugs are distributed in total body water. Physiologic factors such as blood flow to the tissues also influence the distribution of drugs in the body. Since lipophilic drugs cross cell membranes more readily than hydrophilic ones, their distribution in highly perfused tissues is often significant. Conversely, accumulation of drugs in adipose tissue is limited by its blood flow, which is less than 2% of the cardiac output. The pattern of distribution of a given drug in different body compartments may have important implications on its pharmacokinetics and pharmacodynamics in humans and animals.

## Elimination

Drugs are irreversibly eliminated from the body mainly by the processes of renal excretion and hepatic metabolism. Drugs may also be excreted from the body via the bile, feces, lungs, sweat, and breast milk. Excretion refers to the processes by which the body removes the unchanged drugs and their metabolites through the organs of excretion such as the kidneys. Through the processes of filtration, tubular reabsorption, and secretion, the kidneys remove mostly water-soluble drugs and their metabolites from the blood and excrete them into the urine. Several factors – including plasma protein binding, urinary pH, the dissociation constant ( $pK_a$ ) of the

drug, urine flow, and renal blood flow – can affect the efficiency of the kidneys to excrete drugs and metabolites. For example, drugs that are tightly bound to plasma proteins are not readily filtered by the glomerulus. At low glomerular filtrate pH, weak acidic drugs are unionized and reabsorbed in the tubules of the nephron, whereas at high pH they are ionized in the filtrate and excreted into the urine. The efficiency of the kidney in excreting drugs may be impaired by several factors, including age and diseases states, such as high blood pressure and diabetes. In such cases, the drug dosage may be adjusted especially if the drug is eliminated primarily by the kidneys.

Drug metabolism or biotransformation is the process by which a drug is chemically altered by enzymes in the body to produce a metabolite. Although biotransformation can occur in several organs, the liver is the principal organ of drug metabolism. The cytochrome P450 family of enzymes in the liver mediate phase I biotransformation reactions such as oxidation, reduction, hydroxylation, and hydrolysis of drugs. The cytochrome P450 has broad substrate specificity, and the structural features of a particular drug may help to predict potential metabolic pathways. Cytochrome P450 3A (CYP3A) is estimated to metabolize 50–70% of drugs that are in current use. For example, it has been demonstrated that lopinavir, a protease inhibitor, produces inadequate therapeutic concentrations, when administered alone, due to extensive metabolism by CYP3A isoenzymes. Phase II biotransformations such as glucuronidation, sulfation, acetylation, and glycine conjugation are carried out by transferases that attach endogenous macromolecules to drugs or their metabolites to produce highly polar metabolites. Ezetimibe, a lipid-lowering drug, is extensively conjugated to ezetimibe-glucuronide, a pharmacologically active metabolite. In addition, some drugs and metabolites are excreted into the bile. For most drugs, phase II metabolism must occur before excretion into the bile. The drugs and metabolites in the bile are dumped into the GIT and excreted in the feces. Interestingly, drugs and metabolites in the bile that enter the GIT are available for reabsorption into the bloodstream if they escape decomposition and excretion. This phenomenon is referred to as enterohepatic cycling.

A number of factors such as age, sex, fasting, disease, genetics, and pregnancy influence the metabolism of drugs. For example, drug metabolism capacity is lower in infants because metabolic enzyme systems are only partially developed at birth. Other factors such as physicochemical properties and physiologic variables affect drug metabolism. For example, lipid-soluble compounds are primarily



metabolized in the liver. Moreover, the binding characteristics of drugs also affect their intrinsic metabolic clearance. Extensive plasma protein binding tends to retard hepatic metabolism since bound drugs generally do not have access to metabolic sites, such as hepatocytes in the liver. Although drug metabolism is essentially a deactivation process, it is important to note that metabolites may possess similar or different degrees of pharmacological or toxicological activity as the parent drug. Prodrugs, for example, produce the desired therapeutic effects after biotransformation of an inactive form of a chemical entity to active metabolites in the body.

Several *in vitro* systems and animal models are available to evaluate metabolic stability of drugs, identify the enzyme systems that are involved in their metabolism, and predict drug–drug interactions. The Michaelis–Menten expression (eqn [2]) is commonly used to describe the relationship between the rate of metabolism ( $V$ ) and the concentration of the substrate ( $C$ ). This equation is a fundamental interpretation of how an enzyme interacts with a drug, where  $K_m$  is defined as the substrate concentration at half the maximum rate of metabolism ( $V_{max}$ ). Interestingly, the rates of metabolism obtained from human *in vitro* systems such as liver microsomes may be used to predict drug hepatic clearance in humans when the drug concentrations are low relative to  $K_m$ . The *in vitro* intrinsic clearance ( $CL_{int}$ ), a measure of the innate enzyme activity to metabolize a drug, is determined from the ratio of  $V_{max}$  to  $K_m$  of the major metabolic pathways and forms the basis for the prediction of *in vivo* hepatic clearance. Using this method, the intrinsic clearance obtained from human liver microsomal experiments is scaled up to *in vivo* hepatic clearance using appropriate scaling factors such as microsomal protein content, protein binding, liver weight, and blood flow. This approach has been demonstrated to provide good predictions of hepatic clearance in intact animals and humans:

$$V = \frac{V_{max}C}{K_m + C} \quad [2]$$

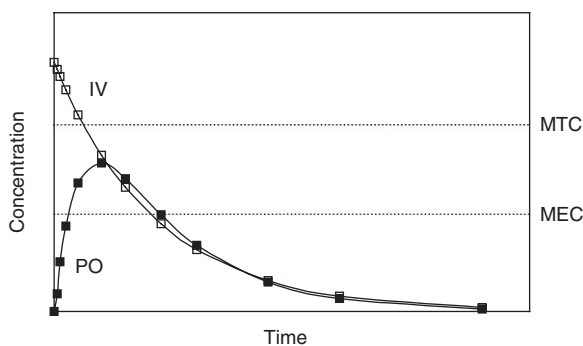
## Drug Transporters

In recent years, it has become apparent that transport proteins play a major role in regulating the absorption, distribution, and excretion of several drugs. Briefly, accumulated evidence has demonstrated that P-glycoprotein (P-gp), a transporter that is embedded in several biological membranes including the blood–brain barrier (BBB) and the GIT, plays an important role in the ADME of many drugs. A major function of P-gp is the energy-dependent cellular efflux of

endogenous substrates and the excretion of drugs and metabolites into urine, bile, and the intestinal lumen. Some studies have shown that digoxin, a P-gp substrate, has significantly higher bioavailability in subjects with reduced P-gp function. At the BBB, P-gp limits the accumulation of many drugs – including digoxin, vinblastine, dexamethasone, and cyclosporine – in the brain. Although P-gp transports a wide range of substances with diverse chemical structures, lipophilic and amphipathic drugs appear to be particularly good P-gp substrates. Interestingly, P-gp and cytochrome P450 3A4 (CYP3A4) are both localized in tissues with major function for drug absorption and disposition, such as the small intestine and liver; it also appears that the substrate specificities of P-gp and CYP3A4 show significant overlap. This observation is of particular importance because P-gp and CYP3A4 may work in a coordinate fashion to influence drug absorption and disposition in the body. The role of drug transporters in modulating drug absorption, distribution, and excretion may contribute to variability in drug pharmacokinetics and pharmacodynamics.

## Plasma Concentration–Time Profile

A plot of the concentration of the drug in plasma against time is referred to as the plasma concentration–time profile. The processes of ADME influence the concentrations of drugs in blood with respect to time. For drugs that are rapidly distributed in the body, changes in the concentrations of the drug in blood reflect changes in the concentration of the drug in other tissues. In order to achieve a desirable therapeutic effect, the drug concentrations in the blood are maintained within a desirable concentration range (Figure 3). This is especially true when the drug concentration in plasma is rapidly equilibrated with the site of action. However, the relationship



**Figure 3** Plasma concentration–time curve after intravenous (IV) and oral (PO) administration of a single dose of a hypothetical drug showing the therapeutic range, minimum effective concentration (MEC), and minimum toxic concentration (MTC).

between drug concentration in plasma and therapeutic effect has not been delineated for many drugs.

## Pharmacokinetic Models

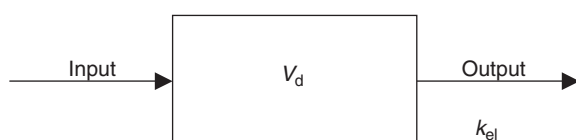
Various pharmacokinetic models are commonly used for the quantitative description of the plasma concentration–time profile. In spite of the inherent assumptions and limitations of these models, they are useful in predicting drug concentration in plasma, urine, and other tissues at different times after drug administration. Pharmacokinetic models are also used to design the optimum dosage regimen, estimate tissue accumulation of drugs and their metabolites, and explain drug–drug interactions and effects of disease on the ADME of drugs in the body. In this article, the open one-compartment models will be highlighted to illustrate how the main pharmacokinetic parameters are determined from a set of drug concentrations in plasma versus time data.

### One-Compartment Model – Instantaneous Input

After intravenous administration of a single bolus dose of a drug that is rapidly distributed in the body, it bypasses the process of absorption and may be removed from the body by a single first-order elimination pathway (Figure 4). The equation that describes the rate of change of drug concentrations in plasma ( $dC_p/dt$ ) for a drug that is rapidly distributed in the body after intravenous administration and follows first-order elimination is shown in eqn [3]. The corresponding integrated form of eqn [3] that describes the concentration of the drug in plasma ( $C_p$ ) at any time ( $t$ ) is depicted in eqn [4]. In this case, the elimination rate constant ( $k_{el}$ ) controls the amount of the drug eliminated per unit time. The pharmacokinetic parameters such as  $k_{el}$ , elimination half-life ( $t_{1/2}$ ), volume of distribution ( $V_d$ ), and clearance (CL) that are determined from the blood concentration–time profile of a drug are useful in drug discovery and for the design of suitable dosage regimens.

$$\frac{dC_p}{dt} = -k_{el}C_p \quad [3]$$

$$C_p = C_0 e^{-k_{el}t} \quad [4]$$



**Figure 4** Schematic representation of the one-compartment model showing drug input, a single first-order elimination step, and associated pharmacokinetic parameters.

### Elimination Half-Life

The elimination half-life ( $t_{1/2}$ ) is the time required for the amount of drug in the body to decrease by half (eqn [5]). The half-life is independent of the dose of the drug administered and is a useful indicator of how fast a drug is removed from the body. It takes five elimination half-lives for  $\sim 97\%$  of the bioavailable dose to be eliminated from the body. The half-life is useful to estimate the dosing interval, the duration of action of a drug, and the time required to attain steady-state plasma concentrations during a multiple dosage regimen. The volume of distribution and clearance of a drug influence the half-life (eqn [6]). For example, the antimalarial drug chloroquine has a high clearance but a very long half-life due to its large volume of distribution. In some renal and hepatic diseases, the half-life may remain unchanged although the clearance and volume of distribution may change by the same proportion. In such situations, the half-life is not a good indicator of drug elimination from the body:

$$t_{1/2} = \frac{0.693}{k_{el}} \quad [5]$$

$$t_{1/2} = \frac{0.693 V_d}{CL} \quad [6]$$

### Volume of Distribution

The apparent volume of distribution ( $V_d$ ) is a virtual volume required to relate the plasma or blood concentration ( $C_p$ ) to the amount ( $A_p$ ) of a drug in the body. In fact,  $V_d$  is the proportionality constant when the plasma or blood concentration of a drug is proportional to the amount of drug in the body. The amount of drug in the body at zero time after administration of a single intravenous dose is equivalent to the dose ( $D$ ). In this case, knowledge of the initial drug concentration in plasma ( $C_0$ ) provides a simple approach for estimating the  $V_d$  of a drug that distributes rapidly in a single homogeneous compartment:

$$V_d = \frac{D}{C_0} \quad [7]$$

$V_d$  is an important pharmacokinetic parameter that is used to calculate the priming dose of a drug that will achieve a required steady-state plasma concentration. In addition, it is a useful indicator of the extent of drug distribution from the blood to the tissues. Drugs that are highly bound to plasma proteins reside mainly in blood, resulting in a high blood concentration and a  $V_d$  that is approximately the volume of blood (5 l) in the average adult human.



**Table 2** Apparent volume of distribution for some drugs in humans

Drug	$V$ (l, 70 kg)
Heparin	5
Sulfisoxazole	11.2
Phenobarbital	38.5
Ethosuximide	49
Diazepam	168
Digoxin	490
Haloperidol	1400
Chloroquine	12 950

On the other hand, drugs that are extensively distributed in the body exhibit a large apparent  $V_d$ . In this case, the apparent  $V_d$  is likely to be greater than the true body volume. In general, the apparent  $V_d$  for drugs ranges from the plasma volume to values that are much larger than the body volume (Table 2).

### Clearance

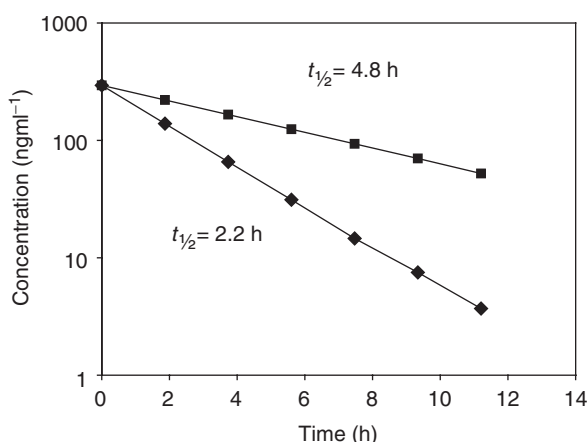
The total body clearance (CL) is an important pharmacokinetic parameter that measures the efficiency with which a drug is irreversibly removed from the body by all organs of elimination. Thus, the sum of the clearance values of individual organs of elimination is equal to the systemic clearance. For most drugs, the systemic clearance of a drug is influenced by the capacity of the liver to metabolize the drug and the inherent ability of the kidney to excrete the unchanged drug. The total clearance, defined as the volume of blood cleared of the drug per unit time, can be calculated from the intravenous bolus dose ( $D$ ) and the area under the plasma concentration–time profile extrapolated to infinity ( $AUC_{(0 \rightarrow \infty)}$ ) as shown in eqn [8]:

$$CL = \frac{D}{AUC_{(0 \rightarrow \infty)}} \quad [8]$$

The total clearance is also a function of blood flow to specific organs of elimination and the extraction ratio. The extraction ratio is an empirical measure of the efficiency of an eliminating organ. Thus, disease states that affect the condition of organs of elimination may alter blood flow, extraction ratio, and ultimately the systemic clearance of drugs.

### Bioavailability

Several factors – including the route of administration, the physicochemical properties, dosage form, and the physiological state of an individual – can affect the amount of the administered dose of a drug that reaches the systemic circulation after extravascular administration. The term bioavailability ( $F$ )



**Figure 5** Nifedipine concentrations in plasma after a single intravenous dose of 25 mg to a patient. The data show a mono-exponential decline with a  $t_{1/2}$  of 2.2 h and elevated concentrations of the drug in plasma when elimination rate is reduced ( $t_{1/2} = 4.8$  h).

refers to the rate and extent to which a drug is absorbed into the bloodstream. The fraction of the administered dose that reaches the systemic circulation after administration by any extravascular route is often calculated from the ratio of the area under the plasma concentration–time profile extrapolated to infinity ( $AUC_{(ext)}$ ) after extravascular administration to the area under the plasma concentration–time profile extrapolated to infinity ( $AUC_{(iv)}$ ) after intravenous administration of the same dose as shown in eqn [9]:

$$F = \frac{AUC_{(ext)}}{AUC_{(iv)}} \quad [9]$$

The area under the plasma concentration–time curve is a useful indicator of the amount of drug that reaches the systemic circulation and of the duration resides there before it is removed. The processes of ADME influence the plasma concentration–time profile of an administered drug and the associated parameters. For example, a reduction in the rate of elimination of nifedipine will result in elevated and prolonged drug levels in plasma (Figure 5).

### New Technologies for Studying ADME

The introduction of new technologies in recent years is increasing the throughput of ADME studies during the drug discovery process. The advent of cell culture techniques in recent years has facilitated the assessment of intestinal permeability for many drugs. A notable example is the use of CaCO-2 cells, an immortalized human colon adenocarcinoma cell line

that contains microvilli, expresses functional transport proteins and metabolic enzymes, and retains many characteristics of the intestinal brush border to predict oral bioavailability of drugs based on the capacity of the drugs to cross the CaCO-2 cell monolayer. Although it has been demonstrated that bioavailability can be predicted with reasonable accuracy using this model for some drugs, *in vitro* approaches are still inferior to animal studies.

In drug discovery, several *in vitro* and animal models are available to evaluate drug distribution. For example, the bovine brain micro-endothelial capillary model (BBMEC), cultured from blood vessels of the BBB, can be used to predict the transport of drugs across the BBB. Plasma protein binding studies, readily performed using standard techniques, augment our understanding of the relationship between total plasma drug concentrations and the unbound drug concentration that is available to interact with the receptors. Traditional *in vivo* techniques such as whole-body autoradiography are also still used to characterize drug distribution in animals and humans.

Several *in vitro* methods such as microsomes and recombinant enzymes are available for evaluating drug metabolism. The high volume of information gathered from *in vitro* drug metabolism studies is useful in choosing drug candidates for future development and for predicting *in vivo* intrinsic clearance of drugs. Although some insight into the various aspects of ADME can be obtained from *in vitro* methods, there are no robust alternatives to intact animals to predict accurately the pharmacokinetic profiles of drug candidates after administration to a whole animal or person at the present time. However, the prediction of *in vivo* pharmacokinetics from *in vitro* data will continue to improve as our understanding of the roles of transporters and cytochrome P450 enzymes in ADME increases. Furthermore,

the advent of *in silico* approaches to predict ADME profiles of drugs based on chemical structure, physicochemical properties, crystal structure of biotransformation enzymes, and ADME profiles of a library of known drugs will increase the efficiency of predicting ADME properties and ultimately the plasma concentration–time profiles of drugs in the body.

*See also:* **Pharmacokinetics:** Pharmacodynamics.

## Further Reading

- Birkett DJ (1991) Volume of distribution. *Australian Prescriber* 11(21988): 36–37.
- Ekins S (2003) *In silico* approaches to predicting drug metabolism, toxicology and beyond. *Biochemical Society Transactions* 31(Part 3): 611–614.
- Gunaratna C (2001) Drug metabolism and pharmacokinetics in drug discovery: a primer for bioanalytical chemists, Part II. *Current Separations* 19(3): 87–92.
- Guttendorf RJ (1996) *The Emerging Role of ADME in Optimizing drug Discovery and Design*. <http://www.net-sci.org/Science/Special/feature06.html>.
- Nnane IP, Damani LA, and Hutt AJ (1995) Drug metabolism studies: metabolite isolation and identification. In: Townsend A, Worsfold PJ, Haswell SJ (eds.) *Encyclopedia of Analytical Science*, vol. 2, pp. 944–951. London: Academic Press.
- Obach RS (2001) The prediction of human clearance from hepatic microsomal metabolism data. *Current Opinions in Drug Discovery Development* 4(1): 36–44.
- Shargel L and Yu ABC (1993) *Applied Biopharmaceutics and Pharmacokinetics*, 4th edn. East Norwalk: Appleton-Century-Crofts.
- Smith DA, van de Waterbeemd H, and Walker DK (2001) *Pharmacokinetics and Metabolism in Drug Design. Methods and Principles in Medicinal Chemistry*, vol. 13, 1st edn. Weinheim: Wiley-VCH.
- Welling PG (1997) *Pharmacokinetics, Processes, Mathematics, and Applications*, 2nd edn. Washington: American Chemical Society.

## Pharmacodynamics

I P Nnane, Temple University, Philadelphia, PA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The fate of a drug in the body involves the processes of absorption, distribution, metabolism, and excretion (ADME), which control the time course of a drug *in vivo* and the interaction of a drug with

receptors in order to produce a biologic effect (Figure 1). While pharmacokinetics examines how the body handles a drug after administration, pharmacodynamics deals with the relationship between the concentrations of a drug at the site of action and the response elicited, including the mechanisms of interaction between a drug and receptors in the body. The interaction of a drug with specific receptors or enzymes initiates a cascade of molecular events that trigger a pharmacodynamic response. In many cases,

that contains microvilli, expresses functional transport proteins and metabolic enzymes, and retains many characteristics of the intestinal brush border to predict oral bioavailability of drugs based on the capacity of the drugs to cross the CaCO-2 cell monolayer. Although it has been demonstrated that bioavailability can be predicted with reasonable accuracy using this model for some drugs, *in vitro* approaches are still inferior to animal studies.

In drug discovery, several *in vitro* and animal models are available to evaluate drug distribution. For example, the bovine brain micro-endothelial capillary model (BBMEC), cultured from blood vessels of the BBB, can be used to predict the transport of drugs across the BBB. Plasma protein binding studies, readily performed using standard techniques, augment our understanding of the relationship between total plasma drug concentrations and the unbound drug concentration that is available to interact with the receptors. Traditional *in vivo* techniques such as whole-body autoradiography are also still used to characterize drug distribution in animals and humans.

Several *in vitro* methods such as microsomes and recombinant enzymes are available for evaluating drug metabolism. The high volume of information gathered from *in vitro* drug metabolism studies is useful in choosing drug candidates for future development and for predicting *in vivo* intrinsic clearance of drugs. Although some insight into the various aspects of ADME can be obtained from *in vitro* methods, there are no robust alternatives to intact animals to predict accurately the pharmacokinetic profiles of drug candidates after administration to a whole animal or person at the present time. However, the prediction of *in vivo* pharmacokinetics from *in vitro* data will continue to improve as our understanding of the roles of transporters and cytochrome P450 enzymes in ADME increases. Furthermore,

the advent of *in silico* approaches to predict ADME profiles of drugs based on chemical structure, physicochemical properties, crystal structure of biotransformation enzymes, and ADME profiles of a library of known drugs will increase the efficiency of predicting ADME properties and ultimately the plasma concentration–time profiles of drugs in the body.

*See also:* **Pharmacokinetics:** Pharmacodynamics.

## Further Reading

- Birkett DJ (1991) Volume of distribution. *Australian Prescriber* 11(21988): 36–37.
- Ekins S (2003) *In silico* approaches to predicting drug metabolism, toxicology and beyond. *Biochemical Society Transactions* 31(Part 3): 611–614.
- Gunaratna C (2001) Drug metabolism and pharmacokinetics in drug discovery: a primer for bioanalytical chemists, Part II. *Current Separations* 19(3): 87–92.
- Guttendorf RJ (1996) *The Emerging Role of ADME in Optimizing drug Discovery and Design*. <http://www.net-sci.org/Science/Special/feature06.html>.
- Nnane IP, Damani LA, and Hutt AJ (1995) Drug metabolism studies: metabolite isolation and identification. In: Townsend A, Worsfold PJ, Haswell SJ (eds.) *Encyclopedia of Analytical Science*, vol. 2, pp. 944–951. London: Academic Press.
- Obach RS (2001) The prediction of human clearance from hepatic microsomal metabolism data. *Current Opinions in Drug Discovery Development* 4(1): 36–44.
- Shargel L and Yu ABC (1993) *Applied Biopharmaceutics and Pharmacokinetics*, 4th edn. East Norwalk: Appleton-Century-Crofts.
- Smith DA, van de Waterbeemd H, and Walker DK (2001) *Pharmacokinetics and Metabolism in Drug Design. Methods and Principles in Medicinal Chemistry*, vol. 13, 1st edn. Weinheim: Wiley-VCH.
- Welling PG (1997) *Pharmacokinetics, Processes, Mathematics, and Applications*, 2nd edn. Washington: American Chemical Society.

## Pharmacodynamics

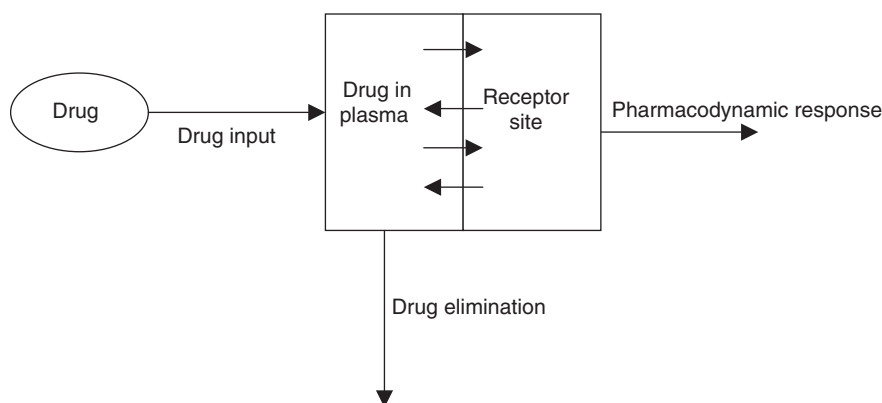
I P Nnane, Temple University, Philadelphia, PA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The fate of a drug in the body involves the processes of absorption, distribution, metabolism, and excretion (ADME), which control the time course of a drug *in vivo* and the interaction of a drug with

receptors in order to produce a biologic effect (Figure 1). While pharmacokinetics examines how the body handles a drug after administration, pharmacodynamics deals with the relationship between the concentrations of a drug at the site of action and the response elicited, including the mechanisms of interaction between a drug and receptors in the body. The interaction of a drug with specific receptors or enzymes initiates a cascade of molecular events that trigger a pharmacodynamic response. In many cases,

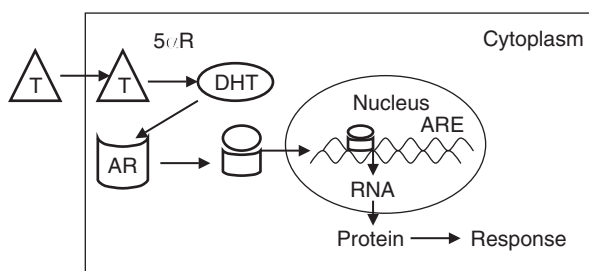


**Figure 1** Schematic representation of the time course of drug action and the relationship between pharmacokinetics and pharmacodynamics.

the effect produced by a drug is correlated to the concentration of the drug at the site of action. Consequently, factors that influence the circulating concentration–time profile of a particular drug are expected to influence the onset, duration, and intensity of the pharmacodynamic effect. During drug discovery and development, pharmacodynamic studies provide an understanding of the mechanism of action of a drug and other critical information that are used for the selection of drug candidates. In addition, the information generated in pharmacodynamic investigations is useful for optimizing drug therapy in the clinical setting. In this article, the basic principles of pharmacodynamics and the models that are used to characterize the drug concentration–effect relationships are discussed. The application of the pharmacokinetic–pharmacodynamic link modeling to increase our understanding of the time course of drug effects is also highlighted.

## Basic Principles of Drug Action

It has been widely demonstrated that drugs interact with specific classes of receptors or interfere with specific signaling pathways in order to induce a pharmacodynamic effect. Receptors denote macromolecules with which a drug interacts in order to initiate biochemical and physiological changes in an organism. Several classes of drug receptors, including receptors for androgens (**Figure 2**) and other endogenous hormones, have been described. Incidentally, many drugs with specific chemical structures also selectively interact with receptors for endogenous ligands. Drug–receptor interactions may involve ionic, hydrogen, van der Waals, and covalent bonding. Generally, many drugs rapidly interact with free receptors in a reversible fashion, and the resulting drug–receptor complex triggers a rapid and direct



**Figure 2** Conversion of testosterone (T) to dihydrotestosterone (DHT) by 5 $\alpha$ -reductase (5 $\alpha$ R), interaction of DHT with the androgen receptor (AR), and binding of the DHT–AR complex to the androgen response elements (ARE) in the androgen signaling pathway.

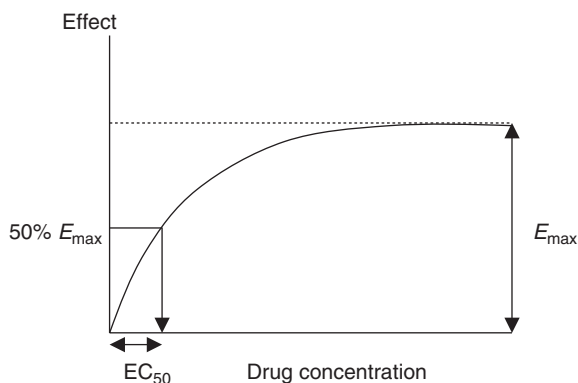
response as shown schematically below:



The scheme illustrates that the drug, receptor, and drug–receptor complex are at equilibrium and assumes that a single drug molecule interacts with a single receptor site. The law of mass action is commonly used to describe the interaction of a drug molecule with a receptor site. According to this approach, the rate of production of a pharmacodynamic effect is proportional to the molar concentrations of the drug at the receptor site [Drug] and the concentration of the receptor [Receptor] in the same way that the rate of an enzyme reaction is proportional to the molar concentrations of the substrate and the enzyme. A greater pharmacological effect is produced, as the drug molecules occupy more receptors, until a maximal effect is attained. The receptor occupancy theory considers a drug that binds to a receptor, mimics the effects of the endogenous ligand, and produces a maximal response as a full agonist; a drug that binds to a receptor and elicits less than a

maximal response is referred to as a partial agonist. Inverse agonists constitute another category of ligands, which produce an effect opposite to that of the agonist by occupying the same receptors. The term antagonist is used to describe a drug that binds to the receptor but does not activate it. Therefore, antagonists do not produce a response because they lack intrinsic activity. Antagonists also interfere with the interaction of an agonist with the receptor. For example, flutamide, a drug that is used to treat prostate cancer, works by blocking the androgen receptors on the surface of the prostate cancer cells and prevents the interaction of the natural ligands, testosterone or its derivative, dihydrotestosterone (DHT), with the androgen receptors (Figure 2). Incidentally, DHT binds to the androgen receptor with about five times greater affinity than testosterone. Since most prostate cancers depend on the supplies of these hormones to grow, flutamide interferes with the action of these androgens and retards or stops the proliferation of prostate cancer cells. Flutamide interacts with the androgen receptor in a specific manner since it has a structure similar to testosterone.

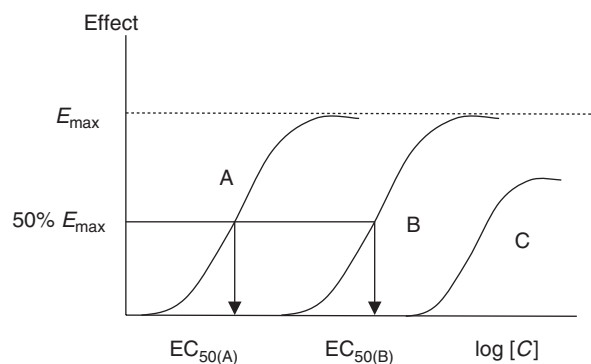
Drugs that bind to the same receptors are also differentiated in terms of efficacy and potency. The intrinsic ability of a drug to produce a specific response is referred to as efficacy. If the relationship between the concentration of a drug and the response is described by a simple hyperbolic profile, the height of any point on the concentration–effect curve that represents a desired response is a measure of the efficacy of a drug (Figure 3). In contrast, potency is a measure of the concentration of a drug needed to elicit a specific response. The width of a point from left to right of the concentration–effect curve is a measure of potency (Figure 3). Therefore, agonists that produce the same pharmacodynamic effect may



**Figure 3** Hyperbolic relationship between the drug concentration and pharmacodynamic effect.  $E_{\max}$  is the maximal effect and  $EC_{50}$  is the drug concentration required to produce 50% of the maximal effect.

exhibit different efficacy and potency. The relationship between the concentrations of epinephrine and norepinephrine, catecholamines with  $\alpha$ - and  $\beta$ -adrenergic receptor activity, and increase in blood pressure are described by graded concentration–effect curves as shown in Figure 4. Epinephrine and norepinephrine interact with the same receptors and produce the same maximal effect, indicating that they are equally effective and may be described as full agonists. Additionally, epinephrine and norepinephrine activate the same number of receptors since efficacy is proportional to the number of receptors activated. However, the concentration of drug required to activate a similar number of receptors is not the same for the two agonists. The concentration of each drug, which produces a half-maximal response, is useful for comparing drugs that act on the same receptors. The potency of epinephrine is greater than that of norepinephrine since the concentration required to produce a half-maximal response is lower. Although epinephrine and norepinephrine are equally effective in eliciting a maximal increase in blood pressure, ephedrine produces a submaximal increase in blood pressure (Figure 4). Thus, ephedrine activates fewer receptors than epinephrine or norepinephrine and is described as a partial agonist.

The enzymes of important signaling pathways are also exploited as drug targets. Inhibitors of enzymes are useful as drugs by selectively blocking specific metabolic pathways, decreasing the concentration of metabolic products or increasing the concentration of enzymatic substrates. In fact, about half of the top



**Figure 4** Sigmoidal relationship between the logarithm of drug concentration ( $\log C$ ) and pharmacodynamic effect.  $E_{\max}$  is the maximal effect and  $EC_{50}$  is the drug concentration to give 50% of the maximal effect. Drugs A (e.g., epinephrine) and B (e.g., norepinephrine) are full agonists and have the same efficacy ( $E_{\max}$  of drug A is the same as  $E_{\max}$  of drug B), but the potency of drug A is higher than that of drug B ( $EC_{50}$  of drug A is lower than  $EC_{50}$  of drug B). Drug C (e.g., ephedrine) is a partial agonist and less potent.



20 drugs sold worldwide are enzyme inhibitors. For example, the aromatase inhibitors such as letrozole and anastrozole reduce the synthesis of estrogens by inhibition of the aromatase enzyme system, which catalyzes the rate-limiting step in the biosynthesis of estrogens from androgenic precursors. Several aromatase inhibitors have been developed in the last few years for breast cancer treatment: the new aromatase inhibitors are considerably more selective, highly potent, and effective in decreasing circulating estrogen concentrations. For example, the efficacy of letrozole for aromatase inhibition has been reported to be greater than 98% at the  $0.5 \text{ mg day}^{-1}$  dose in breast cancer patients.

## Structure–Activity Relationships

The connection between the chemical structure of a drug and its intrinsic activity has been well established. In fact, the molecular structure of a drug influences the nature of drug–receptor interaction and the response that is produced. In drug discovery, the chemical structure of new chemical entities is usually exploited in designing new drugs. Traditionally, new chemical entities are synthesized, tested for pharmacological activity, and structure–activity trends established. Increasingly, mathematical models are used to predict structure–activity relationships. In this case, physicochemical properties such as lipophilicity, electronics, and sterics are used to predict biological activity of potential new drugs. Some therapeutically useful inhibitors of androgen signaling, for instance, have been developed by chemical modification of the chemical structure of the endogenous substrates. The design of steroidal inhibitors of  $17\alpha$ -hydroxylase/ $C_{17,20}$ -lyase, a key enzyme in the biosynthesis of androgens that converts the  $C_{21}$  steroid precursors (pregnenolone and progesterone) to the corresponding  $C_{19}$  androgens, dehydroepiandrosterone, 5-androstenediol, testosterone, and androstenedione, is based on the chemical modification of the endogenous ligands. The anabolic steroid stanozolol has been shown to be a potent competitive inhibitor of  $17\alpha$ -hydroxylase/ $C_{17,20}$ -lyase. Various structural modifications of new chemical entities have been shown to have identifiable effects on drug–receptor interaction, ADME properties, and ultimately on the pharmacodynamic activity of the agent.

## Concentration–Effect Relationships

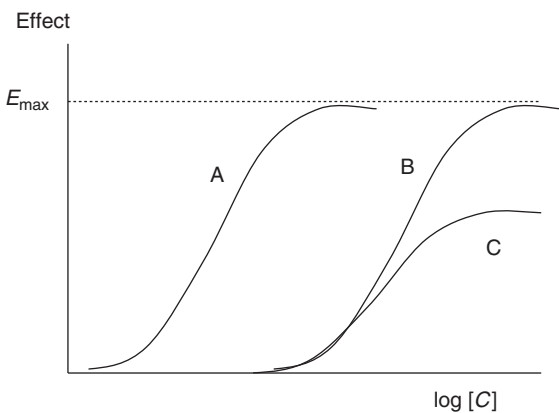
An important consideration of pharmacodynamics is the concentration–effect relationship, which allows one to evaluate how much a response changes when

the drug concentration is increased. The concentration–effect relationship also allows us to determine the potency and efficacy of a drug in order to compare its effects with that of other agonists producing the same response and to determine the safety margin of the drug. It has been demonstrated on several occasions that the drug effect is better correlated to the circulating drug concentration than the administered dose. Therefore, the drug concentration–effect relationship must be properly characterized for optimal drug therapy. The drug concentration at the receptor site depends on the dose; typically the drug concentrations increase as the dose is increased. For most drugs, as the concentration at the receptor site increases, the pharmacological effect increases in a hyperbolic manner, approaching a plateau (Figure 3). Although several mathematical approaches have been used to evaluate the concentration–effect relationship, the Hill expression (eqn [1]) is commonly used to describe the relationship between the concentration of the drug at the site of action ( $C$ ) and the pharmacodynamic effect ( $E$ ). This equation is a basic interpretation of how a drug interacts with a receptor; the affinity of the drug for the receptor is described by  $EC_{50}$ , the concentration of the drug required to produce 50% of the maximal effect ( $E_{\max}$ ). The Hill coefficient ( $s$ ), the number of drug molecules that combine with each receptor molecule, influences the slope and shape of the concentration–effect curve:

$$E = \frac{E_{\max} C^s}{EC_{50}^s + C^s} \quad [1]$$

According to the Hill equation, the maximal effect is attained when the receptor is saturated with the drug. When the effect is plotted against the logarithm of concentration, the concentration–effect curve becomes sigmoidal in shape and shows an approximately linear range between 20% and 80% of the maximal effect (Figure 4). An approximately log-linear relationship between 20% and 80% of the maximal effect is commonly observed for most drugs used at therapeutic doses. Above 80% of the maximal effect, very little increase in response occurs by a given increase in drug concentration, but the risk of adverse effects increases. In addition, some studies have defined the therapeutic range for some drugs: the plasma concentration range that is safe and effective in treating specific diseases in patients. The desired therapeutic effects of a drug are seen within the therapeutic range; subtherapeutic effects are seen below the therapeutic range, and above the therapeutic range there is a greater probability that toxicity may occur.





**Figure 5** Illustration of competitive and noncompetitive antagonism. Competitive antagonism shifts the concentration–effect curve of an agonist to the right without altering the  $E_{\max}$ . Noncompetitive antagonism lowers the observed  $E_{\max}$  of an agonist. Curve A represents the concentration–effect curve of an agonist by itself. Curve B represents the concentration–effect curve of an agonist in the presence of a competitive antagonist. Curve C represents the concentration–effect curve of an agonist in the presence of a noncompetitive antagonist.

Antagonists are designed to block the responses elicited by endogenous agonists. Although antagonists have no effects by themselves, they affect the concentration–effect curves for agonists. An antagonist may prevent the action of an agonist by competitive or noncompetitive inhibition (Figure 5). In competitive antagonism, the reversible interaction between an antagonist and a receptor can be overcome by increasing the concentration of the agonist. In contrast, increasing the concentration of the agonist does not restore its maximal effect in noncompetitive (irreversible) antagonism. For example, the potency of epinephrine in increasing blood pressure is decreased significantly by prazosin since the receptors occupied by prazosin are not activated by epinephrine.

## Pharmacodynamic Models

Pharmacodynamic models are used to characterize the relationship between the drug concentration and the pharmacological response elicited. The most common pharmacodynamic models that relate the drug concentration at the site of action (or the steady-state drug concentrations in plasma) to a rapid and direct response are based on the Hill equation (vide supra). These models are useful for most drugs that exhibit rapid equilibration between the plasma and receptor compartments, and reversible binding of the drug to the receptors. In this case, there is a predictable relationship between measurable drug

concentrations in plasma and drug concentrations at inaccessible receptor sites. It is the concentration of a drug at the biophase that will ultimately determine the intensity of the pharmacodynamic effect. Pharmacodynamic models that account for irreversible drug–receptor interaction, delayed drug distribution, and production of indirect response are beyond the scope of this article.

### The Maximal Effect Model

The maximal effect ( $E_{\max}$ ) pharmacodynamic model relates drug concentrations and the pharmacodynamic effect ( $E$ ) according to eqn [2]. The  $E_{\max}$  model defines a maximal effect ( $E_{\max}$ ) and shows that a sufficiently large increase in drug concentrations near the  $E_{\max}$  elicits a disproportionately small increase in response. The measured baseline effect ( $E_0$ ) occurs when the drug concentration is zero.  $EC_{50}$  is the concentration of the drug required to produce 50%  $E_{\max}$ . The simple  $E_{\max}$  model has been used to describe the relationship between the concentrations of theophylline in plasma and the bronchodilator activity (forced expiratory volume) in asthma patients. The calculated  $EC_{50}$  and  $E_{\max}$  values for theophylline were  $10 \text{ mg l}^{-1}$  and 63% of normal forced expiratory volume, respectively:

$$E = E_0 + \frac{E_{\max} C}{EC_{50} + C} \quad [2]$$

### The Sigmoidal Maximal Effect Model

The sigmoidal maximal effect ( $E_{\max}$ ) model relates the drug concentration ( $C$ ) and the pharmacodynamic effect ( $E$ ) as depicted in eqn [3]. The sigmoidal  $E_{\max}$  model also defines a maximal effect ( $E_{\max}$ ) and shows that a large increase in drug concentrations near the  $E_{\max}$  produces a disproportionately small increase in response. The sigmoidal  $E_{\max}$  model incorporates a slope factor ( $s$ ) that accounts for differences in the steepness of the concentration–effect relationship. The larger the magnitude of the slope factor, the smaller the change in concentration required to produce a change in response from 20% to 80% of  $E_{\max}$ . The sigmoidal  $E_{\max}$  model reduces to the simple  $E_{\max}$  model when the exponent  $s = 1$ . The simple sigmoidal  $E_{\max}$  model has been used to describe the relationship between the concentrations of propranolol in plasma and the antiarrhythmic effects in patients. The calculated  $EC_{50}$  and  $E_{\max}$  values for propranolol were  $\sim 70 \text{ ng ml}^{-1}$  and 75% success rate, respectively. The  $E_{\max}$  model also accommodates a quantal response when the values of  $s$  are greater than 5. In this case, no pharmacodynamic effect is observed below a certain threshold

concentration and a maximal response is observed at concentrations just above the threshold concentration of the drug:

$$E = E_0 + \frac{E_{\max} C^s}{EC_{50}^s + C^s} \quad [3]$$

### The Log-Linear Model

The log-linear pharmacodynamic model describes the linear relationship between the logarithm of drug concentrations ( $\log C$ ) and the pharmacodynamic response ( $E$ ) between 20% and 80% of the maximal effect as shown in eqn [4], where  $I$  is the intercept of the logarithm of drug concentration versus effect plot and  $m$  is the slope of the regression line. For example, the relationship between the logarithm of the concentration of propranolol in plasma and the antiarrhythmic effects in patients has been shown to be approximately linear between 20% and 80% of the maximal effect:

$$E = m(\log C) + I \quad [4]$$

### The Linear Model

The linear pharmacodynamic model describes a linear relationship between the drug concentration ( $C$ ) and the pharmacodynamic response ( $E$ ) when drug concentrations are significantly below the  $EC_{50}$  as shown in eqn [5].  $E_0$  is the baseline effect when the drug concentration is zero and the ratio of  $E_{\max}$  to  $EC_{50}$  is the slope of the regression line. The simple linear model has been used to describe the relationship between the tissue concentrations of *d*-lysergic acid diethylamide (LSD) and behavioral response (performance scores in arithmetic tests):

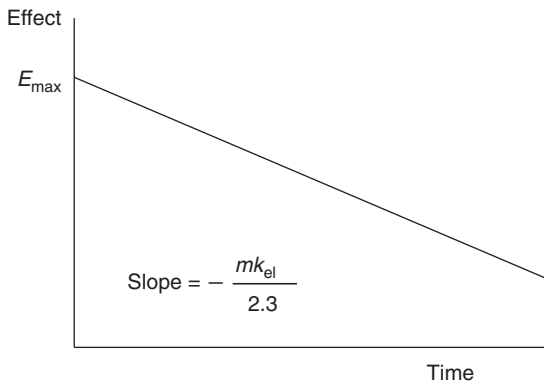
$$E = E_0 + \frac{E_{\max}}{EC_{50}} C \quad [5]$$

## Pharmacokinetic–Pharmacodynamic Models

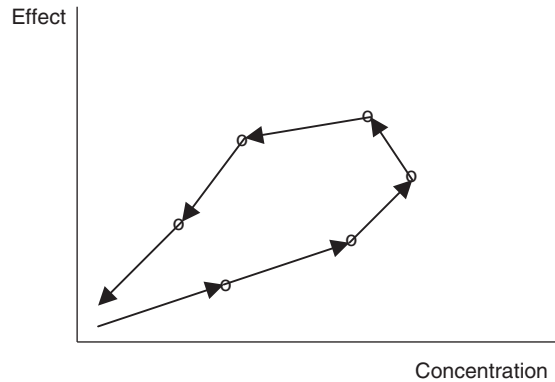
Although the pharmacodynamic effect is primarily a function of the concentration of a drug at the site of action and the affinity of the drug for the receptors, time-dependent processes such as ADME also influence the observed effect. Consequently, the pharmacodynamic effect is also a function of time, especially under *in vivo* conditions where drug molecules may not equilibrate rapidly with the biophase. Recently, relationships between plasma concentrations of drugs and pharmacodynamic effects have received considerable attention. The mathematical analysis of the relationships between pharmacokinetics and pharmacodynamics is often termed

pharmacokinetic–pharmacodynamic (PK–PD) modeling. Although PK–PD models oversimplify drug–receptor interaction, they provide a basis for predicting the time course of drug action and designing dosing regimens. The PK–PD approach is also useful to obtain realistic estimates of the potency and intrinsic activity of the direct effects for drugs exhibiting agonistic properties and the potency of competitive antagonists on the basis of their pharmacodynamic interaction with a full agonist. PK–PD modeling has been applied in studies with many drugs, including benzodiazepines and antiepileptic drugs. Although a detailed description of the various PK–PD models is beyond the scope of this article, we will briefly discuss the time course of a rapid and direct pharmacodynamic effect after a single intravenous dose of a drug that obeys the one-compartment open pharmacokinetic model, and the response is described by the simple  $E_{\max}$  model.

After a single intravenous dose of a drug that obeys the one-compartment open pharmacokinetic model, drug concentrations in plasma fall in a mono-exponential manner with respect to time. Furthermore, a plot of the logarithm of drug concentration in plasma with respect to time is linear. The relationship between the logarithm of drug concentration in plasma and a direct response is also approximately linear within the 20–80% of maximal response range for several therapeutic agents. Therefore, the time course of the pharmacodynamic effect can be established from simultaneous analysis of the plasma concentration versus time and plasma concentration versus effect data obtained after administration of a single intravenous dose of a drug. Using this PK–PD approach, it is apparent that the pharmacodynamic effect falls at a constant rate (zero order) with respect to time in the range of 20–80% of maximal response for drugs that obey the one-compartment model after a single intravenous dose, and the drug effect is described by the simple  $E_{\max}$  model (eqn [6] and Figure 6). The slope of the logarithm of concentration versus effect plot is  $m$  and  $k_{el}$  is the first-order elimination rate constant of the drug;  $E$  is the pharmacodynamic effect at any time ( $t$ ) and  $E_{\max}$  is the maximal response elicited by a single dose. The stimulant effect of amphetamine, for example, declines in a zero-order fashion after administration. If the value of  $m$  or  $k_{el}$  of the drug is very large, the response will decline rapidly after a single intravenous dose. Therefore, a multiple dosage regimen has to be given at small intervals in order to maintain the therapeutic response in such situations. Although the pharmacodynamic effects of many drugs decline at a constant rate after intravenous administration, a first-order decline in response has



**Figure 6** Time course of drug effect after a single intravenous dose of a drug that obeys the one-compartment open pharmacokinetic model, and the response is described by the simple  $E_{\max}$  model.



**Figure 7** Counterclockwise hysteresis of drug concentrations vs. effect.

been observed in some situations:

$$E = E_{\max} - \frac{mk_{el}}{2.3} t \quad [6]$$

#### Duration of Pharmacodynamic Response

Frequently, the duration of action ( $t_d$ ) of a drug is influenced by the amount of drug in the body and the rate of drug elimination. The duration of action may be defined as the time the drug concentration stays within the therapeutic range of a drug. The duration of action for a drug that obeys the one-compartment model after intravenous bolus administration may be calculated using eqn [7], where  $A_0$  is the amount of drug administered;  $A_{\min}$  is the minimum effective dose and  $k_{el}$  is the first-order elimination rate constant of the drug. In this case,  $t_d$  is the time required for the initial amount of drug in the body to decline to the minimum effective amount. This expression demonstrates that there is a linear relationship between the duration of response and the logarithm of the amount of drug in the body. Thus, increasing the dose prolongs  $t_d$  but there is a risk of producing adverse effects if the therapeutic index of the drug is small. Furthermore, the equations show that the duration of action is inversely proportional to the first-order elimination rate constant of the drug. Therefore,  $t_d$  is expected to be prolonged in patients with renal or hepatic failure since impaired efficiency of drug elimination will lead to retention of drug in the body. Notwithstanding, the duration of action is also influenced by pharmacodynamic parameters of a drug in question. Consequently, some drugs such as  $\beta$ -blockers are usually administered once or twice daily, despite having short elimination half-lives, because the doses used are sufficient to produce a maximal effect for a significant part of the dosing

interval:

$$t_d = \frac{2.3}{k_{el}} \left( \log \frac{A_0}{A_{\min}} \right) \quad [7]$$

#### Hysteresis of Response

It has been shown that the drug concentration is not a good indicator of the pharmacodynamic effect in several situations. A looped profile or hysteresis (Figure 7) has been obtained for some drugs when drug concentrations are plotted against response. The loop suggests a time dependency of the response. Additionally, the direction of the loop, which may be clockwise or counterclockwise, may suggest a mechanism for the observed time-dependent response. A clockwise loop may indicate development of tolerance or formation of metabolite with antagonistic effect. In contrast, a counterclockwise loop may indicate slow equilibration of drug concentrations with the receptor site or increased receptor sensitivity. In this case, the response will tend to increase with decreasing concentration of the drug.

In order to obtain a predictive relationship between drug concentrations and pharmacodynamic response, two questions must be addressed: (1) Is drug concentration directly or indirectly related to response? (2) Does the drug interact with the receptors in a reversible or irreversible manner? The quantitative pharmacodynamic approaches described in this article are useful for characterizing rapid and direct responses resulting from reversible drug-receptor interactions. However, drug-receptor interaction initiates a cascade of events, such as formation of second messengers, that ultimately trigger a pharmacodynamic response. These events take time and may affect the time course of drug action *in vivo*. Other

factors, including delayed drug distribution, production of active metabolites, irreversible drug–receptor interactions (for example, the effect of aspirin on cyclooxygenase in platelets), and production of indirect therapeutic effects, complicate the elucidation of the time course of drug effect *in vivo*. In these cases, other appropriate mathematical models are available to establish a good correlation between drug concentration and pharmacodynamic effect and predict the time course of drug action under steady-state conditions *in vivo*. Since the kinetics of a pharmacodynamic effect is influenced by several factors, especially *in vivo*, it is imperative to account for both pharmacokinetic and pharmacodynamic factors when developing realistic PK–PD models with predictive value. The development of robust PK–PD models provides useful information on the nature of the pharmacological properties of drugs and kinetics of drug action *in vivo* both in preclinical and clinical investigations. At the preclinical stage, PK–PD concepts are applicable in the evaluation of *in vivo* potency and intrinsic activity, the identification of suitable markers of therapeutic effect, and optimization of dosage regimens. In the clinical setting, PK–PD concepts are useful in the characterization of the effect versus time relationship and identification of the sources of variability in drug response such as food, age, disease, gender, drug–drug interactions, tolerance development, and genetic polymorphism. Thus, PK–PD concepts provide a promising tool for

selecting new chemical entities for development and optimizing how drugs should be administered.

**See also:** **Pharmacokinetics:** Absorption, Distribution, and Elimination.

## Further Reading

- Birkett DJ (1995) Pharmacokinetics Made Easy 10: Pharmacodynamics – the concentration – effect relationship. *Australian Prescriber* 18: 102–104.
- Bouroujerdi M (2002) *Pharmacokinetics: Principles and Applications*. New York: McGraw-Hill.
- Danhof M, Mandema JW, Hoogerkamp A, and Mathot RA (1993) Pharmacokinetic–pharmacodynamic modeling in pre-clinical investigations: principles and perspectives. *European Journal of Drug Metabolism and Pharmacokinetics* 18(1): 41–47.
- Gabrielsson J and Weiner D (2001) *Pharmacokinetic and Pharmacodynamic Data Analysis: concepts & applications*, 3rd edn, pp. 175–3195. Kristianstads Boktryckeri AB, Stockholm: Taylor and Francis.
- Gibaldi M and Perrier D (1982) *Pharmacokinetics*, 2nd edn., pp. 221–269. New York: Dekker.
- Ross EM (1995) Pharmacodynamics. In: Hardman JG and Limbird LE (eds.) *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, 9th edn, p. 29. New York: McGraw-Hill.
- Shargel L and Yu ABC (1993) *Applied Biopharmaceutics and Pharmacokinetics*, 4th edn. East Norwalk: Appleton-Century-Crofts.

# PHEROMONES

**G R Jones and J E Parker**, Keele University, Staffordshire, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Pheromones are chemicals that transmit information between individuals of the same species evoking physiological or behavioral responses in the individual receiving the signal. Pheromones fall under the broader category of semiochemicals, from the Greek *semeon* meaning signal or mark, which are simply chemicals through which organisms communicate. Chemical communication is ubiquitous in the animal kingdom, with more interactions being mediated by pheromones than any other form of signal. This article introduces the diversity of chemicals that are

pheromones and outlines the general methodology used in their discovery. It then concentrates on the role of gas chromatography techniques in the analysis of volatile pheromones, comprehensively reviewing all the current methods of sampling separation, structural elucidation, and detection as applied to pheromones. The use of high-performance liquid chromatography (HPLC) in the elucidation of nonvolatile pheromones is then highlighted, followed by a brief forward look to future developments.

## The Chemical Diversity of Pheromones

Pheromones can be individual chemicals or complex mixtures of chemicals and the chemical properties of the pheromone can be directly related to its function. A sex pheromone released by a female moth to

factors, including delayed drug distribution, production of active metabolites, irreversible drug–receptor interactions (for example, the effect of aspirin on cyclooxygenase in platelets), and production of indirect therapeutic effects, complicate the elucidation of the time course of drug effect *in vivo*. In these cases, other appropriate mathematical models are available to establish a good correlation between drug concentration and pharmacodynamic effect and predict the time course of drug action under steady-state conditions *in vivo*. Since the kinetics of a pharmacodynamic effect is influenced by several factors, especially *in vivo*, it is imperative to account for both pharmacokinetic and pharmacodynamic factors when developing realistic PK–PD models with predictive value. The development of robust PK–PD models provides useful information on the nature of the pharmacological properties of drugs and kinetics of drug action *in vivo* both in preclinical and clinical investigations. At the preclinical stage, PK–PD concepts are applicable in the evaluation of *in vivo* potency and intrinsic activity, the identification of suitable markers of therapeutic effect, and optimization of dosage regimens. In the clinical setting, PK–PD concepts are useful in the characterization of the effect versus time relationship and identification of the sources of variability in drug response such as food, age, disease, gender, drug–drug interactions, tolerance development, and genetic polymorphism. Thus, PK–PD concepts provide a promising tool for

selecting new chemical entities for development and optimizing how drugs should be administered.

**See also:** **Pharmacokinetics:** Absorption, Distribution, and Elimination.

## Further Reading

- Birkett DJ (1995) Pharmacokinetics Made Easy 10: Pharmacodynamics – the concentration – effect relationship. *Australian Prescriber* 18: 102–104.
- Bouroujerdi M (2002) *Pharmacokinetics: Principles and Applications*. New York: McGraw-Hill.
- Danhof M, Mandema JW, Hoogerkamp A, and Mathot RA (1993) Pharmacokinetic–pharmacodynamic modeling in pre-clinical investigations: principles and perspectives. *European Journal of Drug Metabolism and Pharmacokinetics* 18(1): 41–47.
- Gabrielsson J and Weiner D (2001) *Pharmacokinetic and Pharmacodynamic Data Analysis: concepts & applications*, 3rd edn, pp. 175–3195. Kristianstads Boktryckeri AB, Stockholm: Taylor and Francis.
- Gibaldi M and Perrier D (1982) *Pharmacokinetics*, 2nd edn., pp. 221–269. New York: Dekker.
- Ross EM (1995) Pharmacodynamics. In: Hardman JG and Limbird LE (eds.) *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, 9th edn, p. 29. New York: McGraw-Hill.
- Shargel L and Yu ABC (1993) *Applied Biopharmaceutics and Pharmacokinetics*, 4th edn. East Norwalk: Appleton-Century-Crofts.

# PHEROMONES

**G R Jones and J E Parker**, Keele University, Staffordshire, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Pheromones are chemicals that transmit information between individuals of the same species evoking physiological or behavioral responses in the individual receiving the signal. Pheromones fall under the broader category of semiochemicals, from the Greek *semeon* meaning signal or mark, which are simply chemicals through which organisms communicate. Chemical communication is ubiquitous in the animal kingdom, with more interactions being mediated by pheromones than any other form of signal. This article introduces the diversity of chemicals that are

pheromones and outlines the general methodology used in their discovery. It then concentrates on the role of gas chromatography techniques in the analysis of volatile pheromones, comprehensively reviewing all the current methods of sampling separation, structural elucidation, and detection as applied to pheromones. The use of high-performance liquid chromatography (HPLC) in the elucidation of nonvolatile pheromones is then highlighted, followed by a brief forward look to future developments.

## The Chemical Diversity of Pheromones

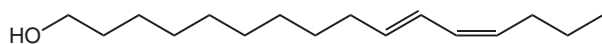
Pheromones can be individual chemicals or complex mixtures of chemicals and the chemical properties of the pheromone can be directly related to its function. A sex pheromone released by a female moth to



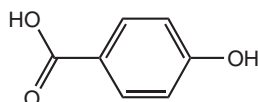
attract a male several kilometers away must be highly volatile and of relatively low molecular weight (less than 300 amu), whereas an aquatic pheromone is often less volatile, of higher molecular weight, and at least partially water soluble. Clearly very different techniques are required to study volatile insect pheromones than water-soluble peptides of marine animals.

The wide variety of pheromones means that the properties and hence the structures of the chemicals involved also vary widely including alkanes, alkenes, alcohols, carboxylic acids, esters, ketones, aldehydes, peptides, and steroids. Therefore, the techniques used to study pheromones must be capable of analyzing a wide range of compounds. These analytical techniques must also be able to detect compounds in very low concentrations, as the amount of pheromone released is very low, usually nanograms or picograms. A male tree frog can attract a female from 1 m away within minutes by the release of just 40 ng of pheromone into the water. In addition, biologists often desire to determine the pheromonal composition of individual organisms, which is challenging for the analyst if the individual is an ant less than 2 mm long.

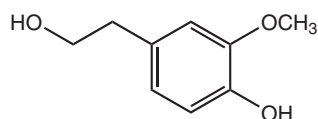
The first pheromone identified was the sex attractant of the silkworm moth *Bombyx mori*, (*E*, *Z*)-10,12-hexadecadien-1-ol or Bombykol (Figure 1). The identification of Bombykol was completed in 1959 by Adolp Butenandt after the extraction of 12 mg of the pheromone from 400 000 individuals and its structure determined by 20 years of chemical degradation reactions. In contrast, the first aquatic pheromone, the steroid 17 $\alpha$ ,20 $\beta$ -dihydroxy-4-pregnen-3-one of the goldfish, was not identified until 1987 using gas chromatography-mass spectrometry (GC-MS) techniques.



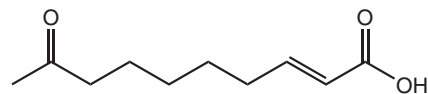
**Figure 1** 'Bombykol' the sex attractant of the silkworm moth.



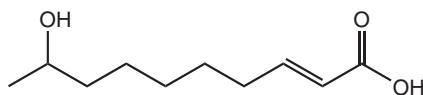
4-Hydroxybenzoic acid



4-(2-Hydroxyethyl)-2-methoxyphenol



(*E*)-9-Oxo-dec-2-enoic acid



(*E*)-9-Hydroxy-dec-2-enoic acid

**Figure 2** Components of the 'queen mandibular pheromone' of the honey bee.

Pheromone elucidation often starts with the observation by a biologist that a pheromone may be involved in a specific behavior of an organism. This leads to bioassays being developed utilizing crude extracts of glands or secretions which if successful are then analyzed usually after further purification. Once the chemical structure of a potential pheromone(s) has been identified, purified extracts or synthetic compounds are then used in the bioassay in order to confirm the pheromonal activity of the specific compounds.

The most common type of pheromones, termed releaser pheromones, change the behavior of the recipient and are used for a wide range of purposes including mating, alarm, attraction, recognition, and trail following. In addition, there are primer pheromones of which only one has been chemically identified, the queen mandibular pheromone in the honey bee, which leads to a physiological change in the recipient (Figure 2).

Although pheromones are known to exist throughout the animal kingdom, most of the information known about them is through the study of insects. The relative ease of studying the behavior of insects and the practical incentive of the development of pest management makes this an appealing area of research.

## Gas Chromatography

Many pheromones are multicomponent mixtures and can occur in nanogram or even picogram amounts. As a result, pheromone research demands techniques that have high-resolution and high-sensitivity capabilities. GC perhaps fulfills these requirements better than any other technique, and as a result has revolutionized pheromone research. As the majority of insect and mammalian pheromones are thermally stable, i.e., volatilize below 300°C, they are ideal for GC analysis.



As with all chromatographic techniques, GC involves the distribution of a compound between two phases, one stationary and one mobile. In GC, the mobile phase is an inert carrier gas, usually helium or hydrogen, and the stationary phase is a capillary column usually coated with a polysiloxane-bonded phase. Samples are volatilized in the injector port and directed onto the capillary column, which is heated in an oven. As the system is heated, compounds are carried through the column, and separated by volatility, polarity, and molecular weight. The resolving power of the system can be altered by the temperature program used, or by the use of different column coatings or column lengths. The capabilities of GC are further increased by the ability to connect the column to a number of different detection systems.

### Sampling and Extraction Techniques

**Solvent extraction** The lipid nature of many insect and mammalian pheromones means that they readily dissolve in pentane or hexane, or dichloromethane if a more polar solvent is required. Solvents should have a low boiling point so that they are rapidly eluted from the GC column and hence do not cover compounds present in the extract. Whole insects, insect parts, or excised glands can be soaked or crushed in the solvent, and the solution then concentrated down ready for injection. The high sensitivity of GC means that all solvents must be of the highest purity possible and any glassware or equipment used must be clean. As contamination is possible it is not uncommon for samples to be purified on silica columns prior to injection onto the GC.

**Solid sampling** A number of 'solvent-less' sampling techniques have been developed over the years in order to minimize the problems of contamination encountered with solvent extraction. Solid sampling is one such method, which allows direct insertion of biological material into the injector port of the GC. A number of solid sampling techniques have been used, one of which, the 'Keele injector', was developed by Morgan and Wadhams in 1972. In this technique, the dissected tissue, gland, or body part is sealed in a soft glass capillary. The solid sampler (Figure 3) is inserted into the injector port of the GC and the capillary is placed inside the injector and subsequently crushed, releasing the volatiles onto the column. This particular technique was used in the identification of the recruitment pheromones of the ants *Aphenogaster albisetosus* and *A. cockerelli*.

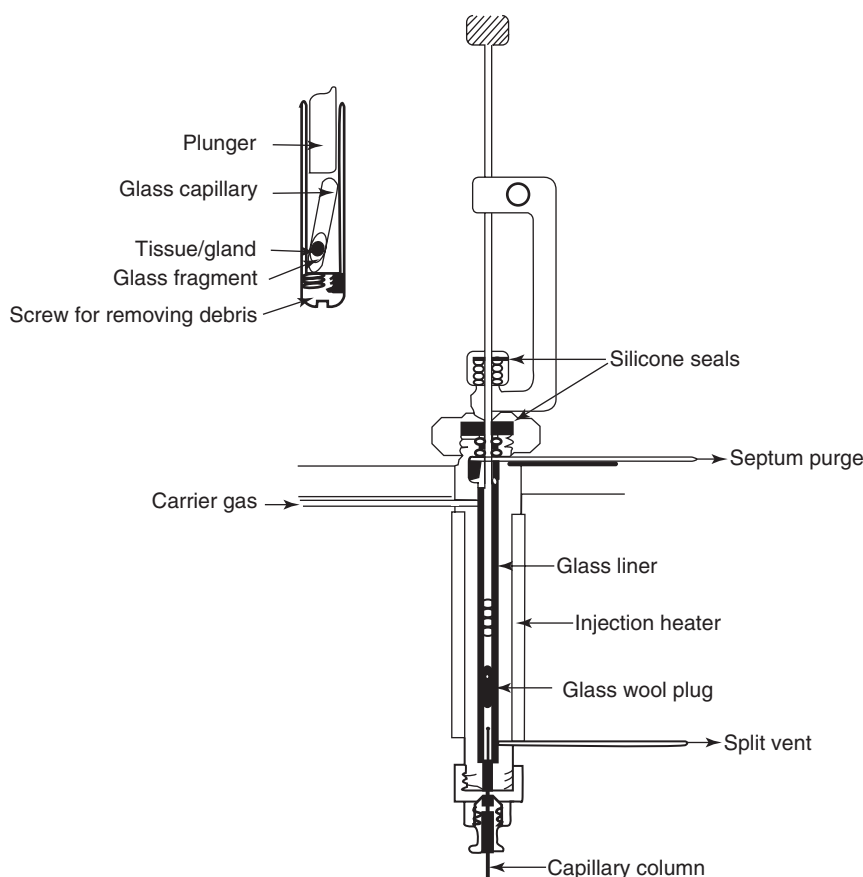
There are many benefits to the solid sampling technique including the fact that the entire sample is used, there is no dilution or contamination with

solvent, and there is no loss of quickly eluting compounds in the solvent peak. The technique also means that samples collected in the field can immediately be sealed in the capillaries, allowing samples to be easily stored and kept free from contamination.

**Volatile trapping** There are a variety of entrapment methods that can be employed in pheromone research including cold trapping, closed loop stripping, and solid adsorbent trapping. A classical way of extracting insect pheromones involves trapping airborne volatiles onto an adsorbent matrix such as activated charcoal or organic polymers (e.g., Porapak-Q and Tenax) and either eluting the compounds with solvent or through thermal desorption directly into a GC. Closed loop stripping has been adapted to sample pheromones from water such as (5R)-5-methyl-3-heptanone and the (3E,5E) and (3E,5Z) isomers of 3,5-octadien-2-one, which induce egg release in the marine polychaete *Platynereis dumerilli*.

Solid-phase microextraction (SPME) is a relatively new method of volatile trapping, and uses a length of fused silica fiber coated with a polymer material or solid adsorbent. The use of SPME in pheromone research was first reported in 1995, and since then has rapidly grown in popularity. Compounds are adsorbed onto the fiber from the headspace above the organism or simply by rubbing the fiber on the surface of the tissue. Pheromones in solution, either aqueous or organic, are adsorbed on to the fiber simply by dipping it into the solution containing the compounds being analyzed. Desorption of the compounds occurs when the fiber is heated in the injector port of a gas chromatogram. SPME provides many advantages over other sampling techniques as it is very simple, takes only a few minutes, and uses no solvent. In studies on the male pheromone of the asparagus fly, the fiber trapped the emission of a single fly after exposure for only 1 min. The other major advantage of SPME is its nondestructive nature allowing repeat sampling of the same organism, which was previously not possible. It also means that samples can be taken from living organisms with few disturbances to the organism and its environment.

There are a wide variety of SPME fiber coatings available. Polymethylsiloxane fibers are useful for low to medium polarity compounds and are therefore suitable for most pheromones. Mixed coated fibers are available and these are possibly more suitable for a starting point as they offer complimentary properties and are therefore suitable for a wide range of compounds. The mixed coated polydimethylsiloxane-divinylbenzene fiber offers the broadest range of adsorption.



**Figure 3** Cross-section of the Keele injector (permission from Prof. E D Morgan).

SPME has been used to sample a wide variety of compounds and organisms, including (Z)-7-dodecenyl acetate, from the female Asian elephant *Elephas maximus*, 4,8-dimethyldecanal, from the flour beetle *Tribolium castaneum*, and 2-butanol, 3-hydroxybutanone, and 2,3-butanediol, from the Rhinoceros beetle *Scapanes australis* (Figure 4).

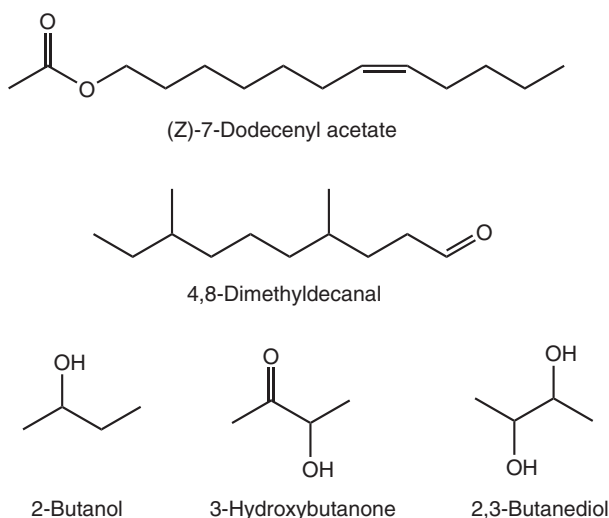
**Microreactions** Microreactions have found widespread use in pheromone research as derivatization and functional group modification can aid chromatographic properties and MS structural determination.

Pheromones can differ simply by the position or geometry of a double bond as is seen with the female sex pheromone of the Asian elephant (Z)-7-dodecenyl acetate and the sex pheromone of the citrus fruit borer (E)-8-dodecenyl acetate (Figure 5). This therefore illustrates the need to know the exact structure of pheromones, including position and geometry of double bonds and other functional groups.

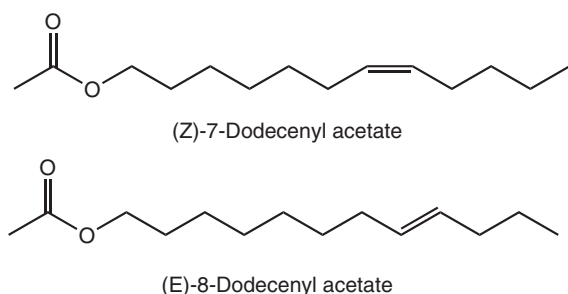
There are a number of methods used for the determination of the number and position of double bonds in unsaturated molecules. Reduction using hydrogen over palladium on carbon, or diimide

reduction can be used to determine the degree of unsaturation. Epoxidation using *m*-chloroperoxybenzoic acid has been used in the study of multi-unsaturated compounds, and was used to aid the identification of (3Z, 6Z, 9Z)-nonadecatriene, the sex pheromone of the fall cankerworm *Alsophila pometaria*. Monoepoxidation and catalytic hydrogenation produced a mixture of epoxide isomers, which allowed the double bond positions to be elucidated by the mass spectral fragmentation patterns. Cleavage in the MS occurs preferentially alpha to the epoxide. This therefore produces fragments that indicate position of the epoxide and hence the double bonds (Figure 6).

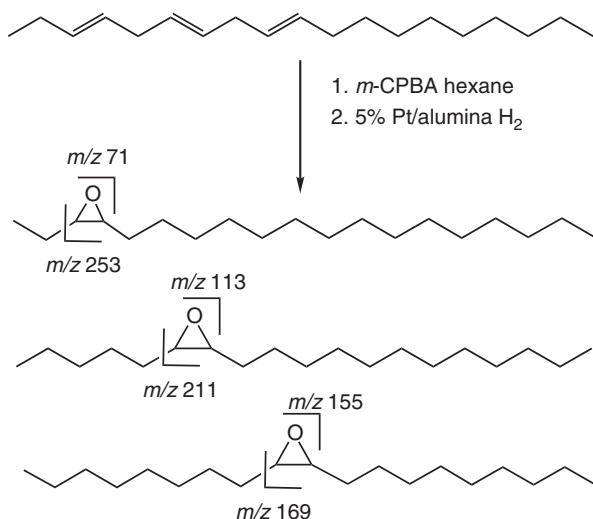
Dimethyldisulfide derivatization is the most common method used for double bond position determination. Reaction of the alkene in hexane with dimethyl disulfide (DMDS) and iodine under an inert atmosphere at 60°C produces the DMDS adduct. MS fragmentation of the DMDS derivatives occurs between the methylsulfide groups, thus locating the original double bond position (Figure 7). DMDS derivatization was used for the determination of the double bond position in (E)-8-dodecenyl acetate, the sex pheromone of the citrus fruit borer.



**Figure 4** Examples of compounds sampled by SPME.



**Figure 5** Double bond positions in the sex pheromones of the Asian elephant and the citrus fruit borer.



**Figure 6** Determination of double bond positions in multi-unsaturated compounds by epoxidation (*m*-CPBA) and reduction (Pt/alumina and  $H_2$ ).

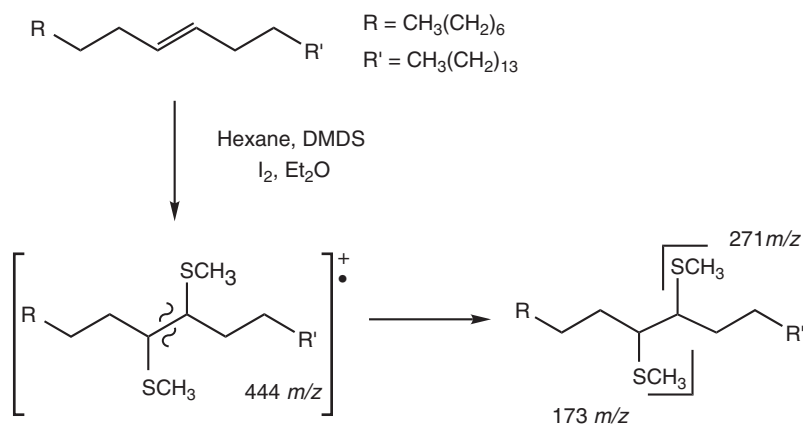
Carboxylic acids are highly polar and reactive substances and do not lend themselves well to gas chromatography, although they can be analyzed on a polar column. Analysis of acids is usually achieved through conversion to methyl esters, or diphenyl esters, which are excellent subjects for GC. Diazomethane is a common methylating reagent resulting in clean and highly selective reactions. The disadvantage of using diazomethane is that it does have explosive and toxic properties. Boron trifluoride in methanol is also used to methylate carboxylic acids and is a very straightforward procedure without the explosive nature of diazomethane.

### Separation and Columns

The majority of insect and mammalian pheromones are small and relatively simple molecules with low polarity. Nonpolar polysiloxane phases such as 100% polysiloxane or 5% polyphenylmethyl siloxane are the preferred column for pheromone research due to a broad range, thermal stability, and long lifetime. Five per cent (phenyl)methylpolysiloxane columns were used in the identification of dodecenyl acetate and dodecenol, the sex pheromones of the citrus fruit borer, and anisole, the sex pheromone of the scarab beetle. A 100% polysiloxane column was used in the analysis of dodecenyl acetate, the female sex pheromone of the Asian elephant.

Mixtures of pheromones can often be complex and demanding to separate, particularly when isomers of molecules are present. This is particularly acute in the study of social insect cuticular hydrocarbons and which are often complex mixtures of 50 or more different hydrocarbons of increasing chain length (Figure 8). For these mixtures the separation of isomeric hydrocarbons such as 9-, 11-, and 13-methylheptacosane is not possible, which has a consequent impact on peak integrations and statistical correlation of these data with colony function.

If better separation is required for more polar compounds a column with a polyethylene glycol phase can be used, as used in the studies of the termite trail pheromone dodec-3-en-1-ol. These phases are also particularly suitable for the separation of carboxylic acids, oxo-acids, lactones, ketones, and polar heterocyclic molecules as seen in the analysis of the metapleural gland secretion of the leaf cutting ant *Acromyrmex insinuator* (Figure 9). It may also be necessary to use more than one column in order to fully separate and identify compounds. In studies on the aggregation pheromone of the sugar cane weevil, both a 5% phenyl methylpolysiloxane column and a polyethylene glycol column were used for identification purposes.



**Figure 7** General scheme for dimethyl disulfide (DMS) derivatization of double bonds.

Chiral columns are also employed in pheromone research for the determination of the absolute configuration of stereoisomers. The configuration of 4-methylheptan-3-ol and 4-methylheptan-3-one, sex pheromones of the oak bark beetle, were determined using a fused silica column coated with octakis [6-O-methyl-2,3-di-O-pentyl]- $\gamma$ -cyclodextrin. The elution order of the ketone stereoisomers was determined according to the literature and the natural compound was found to be (*S*)-methylheptanone. The absolute configuration of the alcohol present in the insects was determined by comparison with authentic standards and was found to be (3*R*, 4*S*)-methylheptanol.

## Detection

**Flame ionization detection (FID)** The most simple detection technique used with GC is FID, which can detect as little as 1 ng of compound. As the analyte enters the detector it passes through a hydrogen flame, which creates a current between two electrodes, producing a signal. Retention indices such as Kovats indices or equivalent chain length can be used to determine chain length and possible identification of compounds (Figure 10). The major disadvantage with FID is that there is no structural information provided, which means that identification can never be definite.

**Mass spectrometry** The combination of GC and MS provides an extremely powerful tool for the study of pheromones. The mass spectrometer is a highly sensitive instrument, normally being able to detect as little as nanogram or picogram amounts. The most important feature of MS is that it provides important structural information including a characteristic fragmentation or 'fingerprint'. Fragmentation patterns are specific for groups of compounds and therefore spectra can be compared to spectral

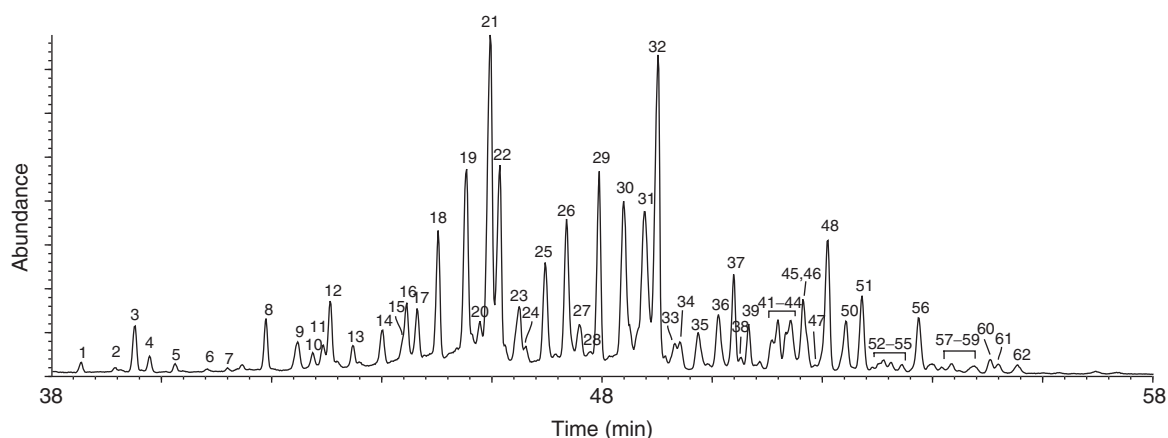
libraries and standard samples. Gas chromatographs are easily connected to mass spectrometers with the capillary column passing through a heated 'transfer line' into the ion source of the mass spectrometer.

The mass spectrometer ionizes molecules causing them to fragment, creating a characteristic group of ions of different masses. Electron ionization (EI-MS) is most commonly used, employing a stream of electrons, which bombards the sample molecules causing radical ions to be formed, and the molecule to fragment. Fragmentation patterns from EI-MS provide a great deal of structural information and often sufficient to allow the identification of compounds. GC-MS has therefore been used in the identification of countless pheromones in both insects and mammals.

The trail pheromone of the ant *Tetramorium meridionale* was successfully identified by EI-MS. The pheromone was found to be made up of four compounds: 2,5-dimethylpyrazine, trimethylpyrazine, 3-ethyl-2,5-dimethylpyrazine, and indole (Figure 11). The compounds were only present in picogram amounts, and all four were required to elicit a response in biological testing thus illustrating both the detection and separation powers of GC-MS.

Chemical ionization (CI-MS) employs a reagent gas, which is introduced into the ion source creating ions to react with the sample molecules. Although abundant molecular ions are usually seen, only simple fragmentation patterns and limited structural information is obtained. CI-MS, therefore, is generally used in addition to EI-MS to give accurate molecular ions or to elucidate specific structural features. For example, ionization of mono-enes using an ion-trap mass spectrometer and acetonitrile as the reagent gas gives rise to two characteristic fragment ions from which the position of a double bond can be located.

Finally, gas chromatography-time-of-flight mass spectrometry (GC-TOF-MS) is now finding its way



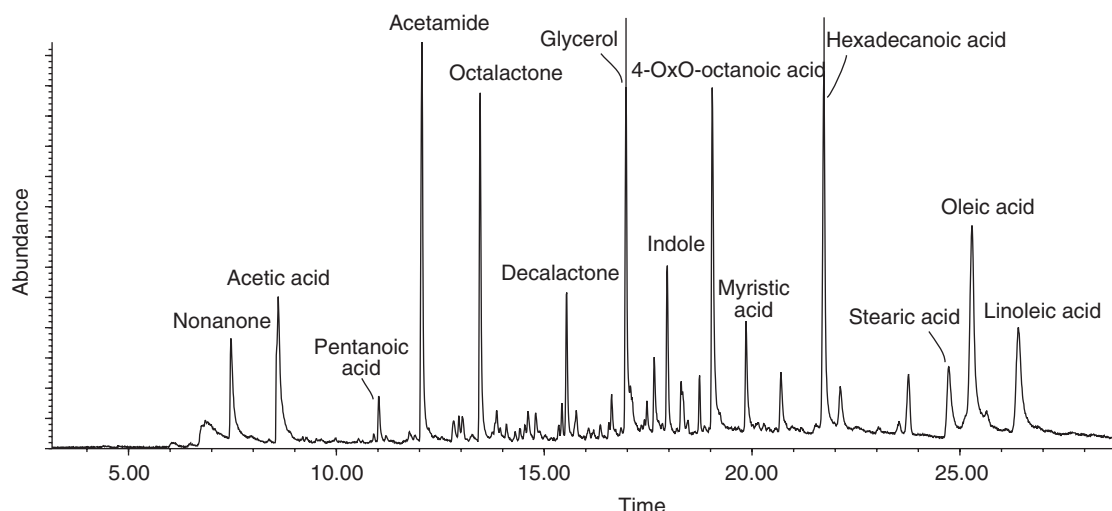
Peak No	Compound	ECL	Peak No	Compound	ECL
1	C23 alkane		32	3-Me C29	29.75
2	11-Me & 13-Me C23	23.35	33	C30 alkane	
3	Unknown		34	X,Y- DiMeC29	30.08
4	3-Methyl C32	23.72	35	Central Me-C30	30.35
5	C24 alkane		36	4-Me C30	30.6
6	12-Me & 14-Me C24	24.35	37	C31alkene	30.8
7	4-Me C24	24.6	38	4, 8 DiMe C30	30.92
8	C25 alkane		39	C31 alkane	
9	9-Me, 11-Me, & 13-Me C25	25.35	40	Branched alkane	31.15
10	5-Me C25	25.43	41	9, 11, 15-Me C31	31.38
11	11,15-DiMe & 9,13-DiMe C25	25.52	42	7-Me & 9-Me C31	31.39
12	3-Me C25	25.72	43	5-Me C31	31.5
13	C26 alkane		44	Branched alkane	31.58
14	10-Me,12-Me, 14-Me, 16-Me C26	26.34	45	3-Me C31	31.76
15	4- Me (70)	26.6	46	5, 9 DiMe C31	31.8
16	10, 14-DiMe & 12,16-DiMe C26	26.63	47	C32 Alkane	
17	Alkene	26.76	48	X,Y- DiMeC31	32.1
18	C27 Alkane		49	Multi branched alkane	32.27
19	11-Me & 13-Me C27	27.36	50	Central Me-C31	32.35
20	5-Me C27	27.53	51	4-Me C32	32.57
21	11, 15-DiMe C27	27.65	52	4,8-DiMe	32.87
22	3-Me C27	27.76	53	C33 alkane	
23	C28 alkane		54	Branched alkane	33.15
24	X,Y-DiMe C27	28.08	55	Central Me	33.34
25	12-Me, 14-Me & 16-Me C28	28.32	56	9- & 7-Me C33	33.46
26	12,16-DiMe, 10,14-DiMe, 14,18-DiMe C28	28.61	57	5-Me	33.64
27	C29 alkene	28.78	58	5, 9-DiMe C33	33.87
28	4, 8-DiMe C28	28.9	59	2,X-DiMe C33	34.08
29	C29 alkane		60	Multi branched alkane	34.27
30	11-Me, 13-Me, & 15-Me C29	29.35	61	Central-Me C34	34.36
31	13,17-DiMe & 11,15-DiMe C29	29.62	62	Central DiMe C34	34.63

**Figure 8** Cuticular hydrocarbon profile of a *Camponotus floridanus* queen (data from J E Parker).

into pheromone research with the added advantages of higher sensitivity and the ability to obtain accurate mass data.

**Electroantennographic detection (EAD)** EAD is an electrophysiological technique that is specifically

used in the chromatography of insect pheromones involving the use of an excised antenna or whole insect. EAD measures the change in potential between the tip and base of the antenna as odors are passed over it giving a relative measurement of the number of chemoreceptors stimulated.



**Figure 9** Gas chromatogram of a metapleural gland of *Acromyrmex insinuator* using a 'carbowax' polyethylene glycol column (data from J E Parker).

#### Kovats indices

$$KI_x = 100 \times \left( \frac{\log(rt_x) - \log(rt_n)}{\log(rt_{n+1}) - \log(rt_n)} \right)$$

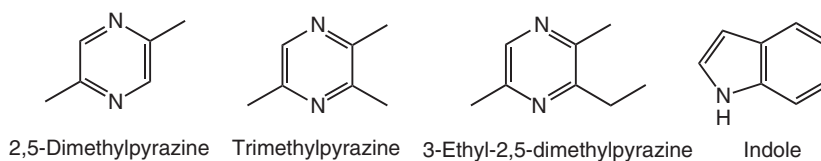
$KI_x$  = Kovats indices of unknown peak,  $x$   
 $rt_x$  = Retention time of peak,  $x$   
 $rt_n$  = Retention time of preceding  $n$ -alkane  
 $rt_{n+1}$  = Retention time of next  $n$ -alkane

#### Equivalent chain length

$$ECL_x = n + \left( \frac{(rt_x) - (rt_n)}{rt_{n+1} - (rt_n)} \right)$$

$ECL_x$  = Equivalent chain length of unknown peak,  $x$   
 $n$  = Chain length of preceding  $n$ -alkane  
 $rt_x$  = Retention time of peak,  $x$   
 $rt_n$  = Retention time of preceding  $n$ -alkane  
 $rt_{n+1}$  = Retention time of next  $n$ -alkane

**Figure 10** Equations for Kovats indices and equivalent chain length.



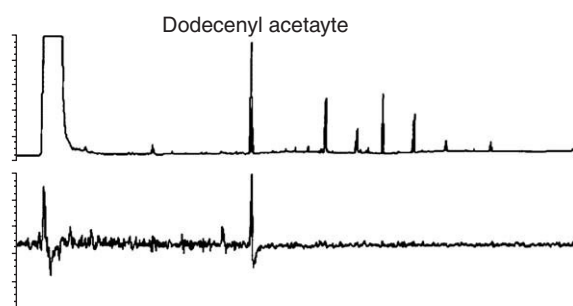
**Figure 11** Components of the trail pheromone of the ant *Tetramorium meridionale*.

Coupling of EAD to GC was first achieved in 1969, and further developed in 1975 in studies on the sex pheromone of the European grape vine moth. The use of capillary columns allowed a continuous stream of the GC effluent on to the antenna, which had previously not been possible. The GC effluent can be split between the EAD and an FID allowing the correlation of biological response to identified compounds (**Figure 12**). Recent studies on the Egyptian armyworm found that the use of the whole insect provides better results than the use of the excised head, or the antenna, and showed that as little as 15 pg of compound could be detected.

The use of GC-EAD allows the rapid screening of compounds and extracts in order to identify those eliciting a response and eliminate those not detected by the insect. GC-EAD takes full advantage of the separation capabilities of GC together with the detection capabilities of the EAD. Stereochemical determination can also be achieved using GC-EAD, and the use of chiral stationary phases allowed the stereochemical identification of the sex pheromone of the brown-banded cockroach.

A further extension of GC-EAD is the use of single cell recording when the response of an individual olfactory cell is monitored. The basic set-up is the





**Figure 12** GC-FID trace (top) and GC-EAD responses of a male citrus fruit borer (bottom) to hexane extracts of female abdominal tips. (Reproduced with permission from (2001) *Journal of Chemical Ecology* 27(10): 2041–2045 (figure 2); © Kluwer Academic Publishers.)

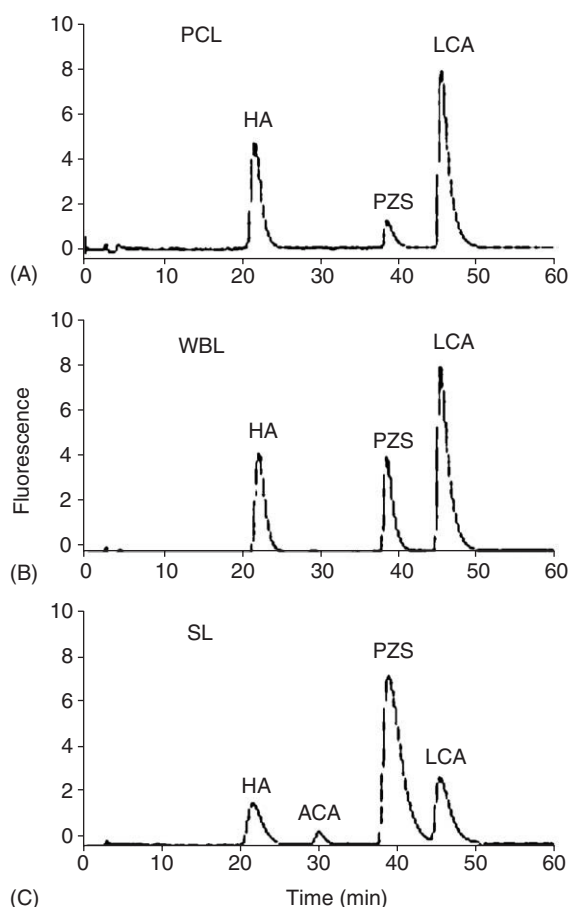
same as in EAD; however, fine tungsten electrodes and micromanipulators are used in order to position a recording electrode at the base of an antennal sensilla.

## High-Performance Liquid Chromatography

Insects have been ideal organisms on which to carry out semiochemical studies and as a majority of their pheromones are volatile or semivolatile they are particularly suited to GC–MS analysis. This, coupled with the slower introduction of LC–MS explains why the use of HPLC in pheromone identification is more limited; however, this is changing. A recent report used the greater resolving power of chiral LC compared to chiral GC to directly measure the natural ratio of enantiomers of (*Z,Z*)-*cis*-3,4-epoxy-6,9-nonadiene, the female sex pheromone of the Japanese giant looper moth. This article also suggests that with the use of microcolumns LC–MS has the potential to rival GC–MS for sensitivity, thus taking HPLC out of its previous use of prefraction prior to GC–MS analysis and the purification and separation of synthetic pheromones.

A wide range of reversed phase (RP) stationary phases are available for use and for preliminary investigations use of a standard octadecyl (C-18, ODS) phase is recommended along with running a standard gradient of water/acetonitrile. A number of different detectors can be incorporated into the HPLC system, the most common being ultraviolet – visible, fluorometric, amperometric, and mass spectrometric. Generally, mass spectrometers have an electrospray ionization (ESI) source, which is good for molecules of a molecular weight above 300 amu and can be used in both negative and positive modes.

There has been biological evidence for the existence of pheromones in marine species since the 1960s



**Figure 13** Chromatograms of gall bladder extracts from the Pacific lamprey (A), western brook lamprey (B), and sea lamprey (C). Petromyzonol sulfate (PZS) and allocholic acid (ACA) cleanly separate from each other and hyocholic acid (HA) and lithocholic acid (LCA) are the internal standards. (Reprinted with permission from *Steroids*, 68 (2003) 515–523; © Elsevier.)

and from the late 1980s. HPLC and LC–MS have been applied to the elucidation of their chemical structure. For example, steroidal bile acids have been identified in a number of *Lampetra* species using RP-HPLC. Petromyzonol sulfate and allocholic acid have been confirmed as a migratory pheromone in the larval sea lamprey and the related 3-ketopetromyzonol sulfate as a male sex pheromone. To aid detection the HPLC eluent was mixed with buffered NAD, dithiothreitol, and ethylenediaminetetraacetic acid using a tee and passed down a column of immobilized 3 $\alpha$ -hydroxy-steroid dehydrogenase column and the resulting NADH detected fluorometrically (**Figure 13**). Similarly, RP-HPLC has been used in the purification of a number of peptide pheromones such as attractin, a 58-residue protein sex pheromone produced by the female mollusk *Aplysia californica*.

Amphibian species have also been found to use peptide and protein pheromones. For example, the structure of the decapeptide sodefrin, the first peptide

pheromone identified in a vertebrate, was isolated from male red-bellied newts in 1995 using RP-HPLC purification. Similarly, a number of pheromones from male and female tree frogs have been separated and identified using RP-HPLC and ESI-MS/MS.

## Future Developments

The analysis and identification of pheromones is a diverse and complex area of research. The use of GC and GC-MS is well established and provides sensitive and reliable methods of analysis for a majority of volatile pheromones. The development of fast GC techniques and two-dimensional separations are likely to have an impact in the near future. The increased sensitivity of mass spectrometers will further improve the limits of detection and GC-TOF-MS provides a fast and efficient means of obtaining accurate mass data. The coupling of GC to EAD allows the fast screening of compounds for biological activity, and the introduction of SPME provides a reliable method for sampling live individual organisms with the minimum of disturbance. Increasingly, LC-MS techniques are now finding uses in the field of pheromone research, particularly in the identification of nonvolatile pheromones. This trend is likely to continue and over the next decade it is anticipated that the diversity of species in which pheromones are known will further expand.

See also: **Gas Chromatography**: Overview. **Perfumes**.

## Further Reading

Agosta WC (1992) *Chemical Communication: The Language of Pheromones*. New York: Scientific American Library.

- Acree TE and Soderlund DM (1985) *Semiochemistry: flavours and pheromones*. Berlin: Walter de Gruyter.
- Howse P, Stevens I, and Jones O (1998) *Insect Pheromones and their Use in Pest Management*. London: Chapman & Hall.
- Hummel HE and Miller TA (eds.) (1984) *Techniques in Pheromone Research*. Berlin: Springer-Verlag.
- Jones GR and Oldham NJ (1999) Pheromone analysis using capillary gas chromatographic techniques. *Journal of Chromatography A* 843: 199–236.
- Kikuyama S, Yamamoto K, Iwata T, and Toyoda F (2002) Peptide and protein pheromones in amphibians. *Comparative Biochemistry and Physiology Part B* 132: 69–74.
- Leonhardt BA and Beroza M (1982) *Insect Pheromone Technology: chemistry and applications*. Washington, DC: American Chemical Society.
- McCaffery AR and Wilson ID (eds.) *Chromatography and Isolation of Insect Hormones and Pheromones*. New York: Plenum Press.
- Merfort I (2002) Review of the analytical techniques for sesquiterpenes and sesquiterpene lactones. *Journal of Chromatography A* 967: 115–130.
- Millar JG and Haynes KF (eds.) (1998) *Methods in Chemical Ecology. Volume 1, Chemical Methods*. Dordrecht: Kluwer Academic Publishers.
- Ryan MF (2002) *Insect Chemoreception: Fundamental and Applied*. Dordrecht: Kluwer Academic Publishers.
- Stanley-Samuelson DW and Nelson DR (eds.) *Insect Lipids: Chemistry, Biochemistry, Biology*. Lincoln: University of Nebraska Press.
- Wyatt TD (2003) *Pheromones and Animal Behaviour. Communications by Smell and Taste*. Cambridge: Cambridge University Press.
- Yamazawa H, Yamamoto M, Karasawa KI, Pu G-Q, and Ando T (2003) Characterisation of geometrid sex pheromones by electrospray ionization time-of-flight mass spectrometry. *Journal of Mass Spectrometry* 38: 328–332.

# PHOSPHORESCENCE

Contents

**Principles and Instrumentation**

**Room-Temperature**

## Principles and Instrumentation

**A Sanz-Medel and J M Costa-Fernandez**, University of Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The emission of visible or ultraviolet (UV) radiation by matter is due to spontaneous radiative decay of electronically excited atoms or molecules. Thus, such emission entails that the emitting matter loses energy in the form of electromagnetic radiation and therefore some form of energy must be supplied for a

pheromone identified in a vertebrate, was isolated from male red-bellied newts in 1995 using RP-HPLC purification. Similarly, a number of pheromones from male and female tree frogs have been separated and identified using RP-HPLC and ESI-MS/MS.

## Future Developments

The analysis and identification of pheromones is a diverse and complex area of research. The use of GC and GC-MS is well established and provides sensitive and reliable methods of analysis for a majority of volatile pheromones. The development of fast GC techniques and two-dimensional separations are likely to have an impact in the near future. The increased sensitivity of mass spectrometers will further improve the limits of detection and GC-TOF-MS provides a fast and efficient means of obtaining accurate mass data. The coupling of GC to EAD allows the fast screening of compounds for biological activity, and the introduction of SPME provides a reliable method for sampling live individual organisms with the minimum of disturbance. Increasingly, LC-MS techniques are now finding uses in the field of pheromone research, particularly in the identification of nonvolatile pheromones. This trend is likely to continue and over the next decade it is anticipated that the diversity of species in which pheromones are known will further expand.

See also: **Gas Chromatography**: Overview. **Perfumes**.

## Further Reading

Agosta WC (1992) *Chemical Communication: The Language of Pheromones*. New York: Scientific American Library.

- Acree TE and Soderlund DM (1985) *Semiochemistry: flavours and pheromones*. Berlin: Walter de Gruyter.
- Howse P, Stevens I, and Jones O (1998) *Insect Pheromones and their Use in Pest Management*. London: Chapman & Hall.
- Hummel HE and Miller TA (eds.) (1984) *Techniques in Pheromone Research*. Berlin: Springer-Verlag.
- Jones GR and Oldham NJ (1999) Pheromone analysis using capillary gas chromatographic techniques. *Journal of Chromatography A* 843: 199–236.
- Kikuyama S, Yamamoto K, Iwata T, and Toyoda F (2002) Peptide and protein pheromones in amphibians. *Comparative Biochemistry and Physiology Part B* 132: 69–74.
- Leonhardt BA and Beroza M (1982) *Insect Pheromone Technology: chemistry and applications*. Washington, DC: American Chemical Society.
- McCaffery AR and Wilson ID (eds.) *Chromatography and Isolation of Insect Hormones and Pheromones*. New York: Plenum Press.
- Merfort I (2002) Review of the analytical techniques for sesquiterpenes and sesquiterpene lactones. *Journal of Chromatography A* 967: 115–130.
- Millar JG and Haynes KF (eds.) (1998) *Methods in Chemical Ecology. Volume 1, Chemical Methods*. Dordrecht: Kluwer Academic Publishers.
- Ryan MF (2002) *Insect Chemoreception: Fundamental and Applied*. Dordrecht: Kluwer Academic Publishers.
- Stanley-Samuelson DW and Nelson DR (eds.) *Insect Lipids: Chemistry, Biochemistry, Biology*. Lincoln: University of Nebraska Press.
- Wyatt TD (2003) *Pheromones and Animal Behaviour. Communications by Smell and Taste*. Cambridge: Cambridge University Press.
- Yamazawa H, Yamamoto M, Karasawa KI, Pu G-Q, and Ando T (2003) Characterisation of geometrid sex pheromones by electrospray ionization time-of-flight mass spectrometry. *Journal of Mass Spectrometry* 38: 328–332.

# PHOSPHORESCENCE

Contents

**Principles and Instrumentation**

**Room-Temperature**

## Principles and Instrumentation

**A Sanz-Medel and J M Costa-Fernandez**, University of Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The emission of visible or ultraviolet (UV) radiation by matter is due to spontaneous radiative decay of electronically excited atoms or molecules. Thus, such emission entails that the emitting matter loses energy in the form of electromagnetic radiation and therefore some form of energy must be supplied for a

continuous emission of radiation. 'Incandescence' is the process in which the emission of radiation is sustained by simple heating, e.g., a hot body emits light. All other forms of light emission by matter can be included under the more general term of 'luminescence' and the most common, and analytically most useful, type of luminescence is 'photoluminescence', in which the necessary energy is supplied by an external exciting light. The two sister techniques, fluorescence and phosphorescence, derived from photoluminescence phenomenon are the result of two basic processes: absorption (excitation) of light and spontaneous emission of light from the matter studied.

Historically, one of the first known references on luminescence dates from the Chinese literature of the years 1500–1000 BC, where the phenomenon of light coming out from different types of glowworms is described. However, it appears that the first reported observation of photoluminescence can be traced back to the sixteenth century when a Spanish physician and botanist, Nicolás Monardes, reported a blue tint that appeared in the water contained in cups of a special type of wood. Edmond Becquerel (1820–91) was, during the nineteenth century, the most important researcher on phosphorescence phenomenon: he not only obtained the excitation and emission spectra from different phosphors, but also performed many different fundamental studies including the effect of temperature or the measurement of the emission time of the phosphorescence phenomenon and established a exponential law to explain the decay of the phosphorescence emission. For such purposes he built in 1858 the first phosphoroscope that was able to measure lifetimes as short as  $10^{-4}$  s.

The term 'photoluminescence' and the first report of phosphorescence, however, was made by Eilhardt Wiedemann in 1887, who observed the phosphorescence of aniline dyes in solid solutions and in gelatin.

However, the satisfactory explanation of phosphorescence as a radiative transition from the lowest triplet state had to wait till the decisive paper of G.N. Lewis and M. Kasha in 1944. Moreover, considering that the characteristics of the phosphorescence phenomenon, in terms of wavelengths, triplet lifetime, quantum yield, etc., are specific of the emitting molecule, these authors also suggested the potential of phosphorescence for the individual identification of organic compounds in complex mixtures.

Analytical low-temperature phosphorimetry was mainly developed during 1960s by the group of J.D. Winefordner, after the pioneering paper by Kiers and co-workers that appeared in 1957.

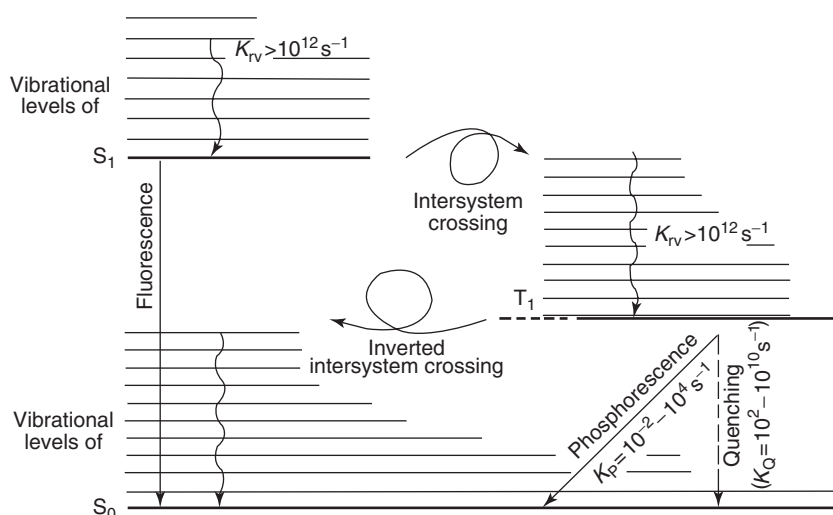
We will concentrate here on the basic principles, instrumentation, and experimental factors affecting

the phosphorescence signals observed. A brief survey of low-temperature phosphorimetry (LTP) and some selected applications will also be included. The analytical potential and scope of room-temperature phosphorescence (RTP), not fully disclosed until the 1970s, are dealt with elsewhere in the encyclopedia.

## Photophysical Basis of Phosphorescence

The general theory of molecular photoluminescence is discussed elsewhere in this encyclopedia. Thus, only the principles pertinent to phosphorescence will be revised here in the light of the basic processes that can take place in molecular luminescence, and these are shown schematically in **Figure 1**. A phosphorescent emission originates as a radiative transition from the lowest triplet state to the ground state, according to the following mechanism: the molecule is originally in one of the vibrational levels of the electronic ground state. After the absorption of electromagnetic radiation of appropriate frequency in the UV or visible region, the molecule is 'excited' to an upper energy level of the same multiplicity; that is, as the ground state is usually a 'singlet' state (all the electron spins are paired and so in opposing directions) the most probable excitation by electromagnetic radiation is that promoting the molecule to a vibrational level of some of the excited electronic singlet states,  $S_1$ ,  $S_2$ , etc. Direct excitation to triplet states,  $T_1$ ,  $T_2$ , etc. (where the electron spins are parallel) is most unlikely because the transition  $S_0 \rightarrow T_1$ ,  $T_2$ , etc., involves a change in the electron spin and so is a spin-forbidden transition. Therefore, excitation of the molecule always takes place via singlet–singlet transitions. These are of such high probability that the time to pass from the ground state to the excited singlet states is  $\sim 10^{-15}$  s (a period so short that, according to the Franck–Condon principle, such electronic transition do not essentially modify the relative positions of the atomic nuclei of the molecule). But if phosphorescence is a triplet→singlet transition, how is the triplet state populated?

Once excited by electromagnetic radiation, the molecule relaxes very quickly, via vibrational relaxation and internal conversion processes, down to the lowest vibrational level of the lowest excited singlet state,  $S_1$ , from where fluorescence originates. If there is spin–orbit coupling between an excited singlet state and an excited triplet state, the vibrational levels of  $S_1$  overlap with the higher vibrational levels of the triplet  $T_1$  and a mixing of states results that removes the original spin-forbidden character of such



**Figure 1** Simplified energy level diagram of an organic luminescent molecule showing basic photophysical processes.  $K_{rv}$ , rate of vibrational relaxation processes,  $K_p$ , rate of phosphorescence,  $K_q$ , rate of quenching.

$S_1 \rightarrow T_1$  transitions, favoring the process of ‘intersystem crossing’ between similar energy levels (the time for intersystem crossing is  $10^{-4}$ – $10^{-12}$  s), which can therefore compete favorably with fluorescence (decay time,  $10^{-9}$ – $10^{-7}$  s). Therefore, the process of ‘intersystem crossing’ is crucial to populating the triplet states (and so to observe phosphorescence). The molecule will then undergo rapid vibrational relaxation and perhaps internal conversions to arrive at the lowest vibrational level of the lowest excited triplet state. Once there, a most peculiar spin-forbidden radiative transition  $T_1 \rightarrow S_0$  (lifetime  $10^{-4}$ – $10^2$  s) can take place, bringing about ‘phosphorescence’, which is the emission of electromagnetic radiation of energy  $E = h\nu$ , corresponding to the  $T_1 \rightarrow S_0$  energy gap.

As shown in **Figure 1**, there are two processes to be considered and minimized for observation of strong phosphorescence: ‘reverse (or inverted) intersystem crossing’ and quenching of the phosphorescence. As a rule the intrinsic rate constants of intersystem crossing  $S_1 \rightarrow T_1$  are orders of magnitude greater than for the  $T_1 \rightarrow S_0$  transitions because the energy gap is smaller and the probability of vibrational levels overlapping is greater in the former case.

In the light of this ‘pumping’ or ‘populating’ of the triplet mechanism, the main distinguishing features of phosphorescence versus fluorescence become clear:

1. As the triplet  $T_1$  lies lower than  $S_1$ , the phosphorescence emissions will appear at longer wavelengths than those of fluorescence.
2. Any factor or manipulation that favors the  $S_1 \rightarrow T_1$  rate constant, e.g., heavy atoms, will

decrease the fluorescence intensity while, as a rule, it will increase the phosphorescence intensity.

3. Phosphorescence originates as a result of a spin-forbidden transition. Therefore, decay times of phosphorescence are orders of magnitude longer than those of fluorescence (lifetimes of  $10^{-4}$ – $10^{-2}$  s as compared to  $10^{-9}$ – $10^{-7}$  s).

The long lifetimes in phosphorescence are most useful for achieving improved analytical selectivity for three main reasons. First, the measurement of phosphorescence sufficiently delayed from the excitation process permits the elimination of possible spectral interferences caused by fast processes such as Rayleigh or Raman scattering and fluorescence emissions. Second, it is easy to differentiate among several phosphorescent molecules or even to identify them just by measuring their lifetimes, which are characteristic in given environments (time-resolved phosphorimetry). Third, phosphorescence provides multidimensional information (excitation and emission spectra, intensity of emissions, lifetimes, and even polarization characteristics) in a straightforward manner. Such information can be used conveniently, with the aid of chemometrics, to resolve the analysis of difficult mixtures without previous separation.

The long lifetimes of phosphorescence are also responsible for the main drawback of the technique: the emission is much more critical and less intense than conventional fluorescence. Because the rate of phosphorescence decay is so slow, the probability of occurrence of other competing processes increases



substantially. As well as reverse intersystem crossing  $T_1 \rightarrow S_1$  (see Figure 1), in particular collisional deactivation of the excited triplet state by solvent molecules and impurities, especially dissolved oxygen, will become most effective. In fact, the probability of triplet deactivation by collisions (quenching) is so high for an excited molecule in solution and at room temperature that phosphorescence in such conditions is very rare (observed only under special circumstances). For this reason, it was believed initially that the only way to obtain analytically useful phosphorescent signals was by dissolving the phosphor molecule in a solvent that was then frozen at the temperature of liquid nitrogen (77 K) to form a rigid glass in which motion, and so external collisions (quenching), would be hindered.

Practical analytical work with such cryogenic techniques is complex and cumbersome and this has deterred general acceptance of the technique by the analytical community (except for some special applications that are reviewed later). However, several advances have made it possible to obtain strong phosphorescence signals at room temperature, even in oxygenated liquid solutions, opening new possibilities for analytical spectrometry applications.

Apart from the peculiarities derived from the triplet state deactivations, the main basic parameters (concepts) that describe a phosphorescent emission are virtually the same as those that describe fluorescence emissions. The most important of those parameters, their concepts, and dependences on intrinsic and extrinsic factors (via simple formulae) are summarized in Table 1.

## Factors Affecting the Phosphorescence Signal: Basis of Quantitative Analysis

As in fluorescence, phosphorescence signals depend upon both structural and environmental factors. Under appropriate experimental conditions, the intensity of the phosphorescent emission measured is linearly related to the concentration of the phosphor. This relationship, which constitutes the theoretical basis of quantitative phosphorescence analysis, is very similar to that obtained for fluorescence, except that the efficiency of 'intersystem crossing', which populates the triplet, should also be taken into account. Therefore, at very low concentrations of the analyte (trace analysis), the basic equation is

$$\Phi_P = 2.303(\Phi_T \Phi_{PP}) I_0 \epsilon l c \quad [1]$$

(see Table 1 for definition of terms).

Although theoretical expressions like eqn [1] under experimental conditions can become rather complex, or have not been fully developed, the basic equation (eqn [1]) is most useful for simple systems. It demonstrates that for a given phosphor ( $\epsilon$  is constant at a given wavelength) under constant experimental conditions in dilute solutions, there is a linear relationship between the property measured,  $I_P$ , and concentration,  $c$ , of the phosphor. It shows us that one way to increase the analytical signal is by resorting to instrumental parameters optimization. In fact, increasing the excitation slitwidth or lamp intensity ( $I_0$ ) or increasing the pathlength would result in an increase of  $I_P$  for a given concentration of the phosphor. Equation [1] also shows that the analytical

**Table 1** Important photophysical parameters<sup>a</sup> in phosphorescence

Phosphorescence parameter	Concept	Basic formula
Quantum yield, $\Phi_P$	Photons emitted as phosphorescence/photons absorbed from excitation	$\Phi_P = \Phi_T \frac{K_P}{K_P + K_{UD} + K_{PQ}}$ $= \Phi_T \Phi_{PP}$
Intensity (continuous emission measurement), $I_P$	The analytical signal measured at a given wavelength bandpass, $\Delta\lambda$	$I_P = 2.3\Phi_P I_0 \epsilon l c \quad (\epsilon l c < 0.01)$
Lifetime, $\tau_P$	Time required for the emission to decrease to $1/e$ of its initial ( $t=0$ ) intensity at given conditions	$\tau_P = \frac{1}{K_P + K_{UD} + K_{PQ}}$
Intrinsic lifetime, $\tau_P^*$	Lifetime of the triplet in the absence of radiationless deactivation processes ( $\Phi_P = 1$ )	$\tau_P^* = \frac{\tau_P}{\Phi_P} \Phi_T = \frac{1}{K_P}$

<sup>a</sup>  $\Phi_T$  = quantum efficiency for triplet,  $T_1$ , formation;  $K_P$  = rate of phosphorescence;  $K_{UD}$  = rate of radiationless unimolecular  $T_1$  deactivating processes;  $K_{PQ}$  = rate of phosphorescence bimolecular quenching;  $I_0$  = intensity of the excitation radiation;  $\epsilon l c$  = absorbance of the sample solution;  $c$  = analyte concentration;  $l$  = pathlength through sample for absorption;  $\epsilon$  = molecular absorptivity.



signal can be increased by optimization of the microenvironment that surrounds the phosphor. Any manipulation of the environment that favors the process of intersystem crossing,  $\Phi_T$  (e.g., a change of solvent or by addition of heavy atoms such as  $\text{Pb}^{2+}$  or  $\text{Tl}^+$ ), will increase the analytical signal  $I_P$ .

According to eqn [1],  $I_P$  would also be increased by optimizing the experimental conditions to protect the populated triplet from radiationless deactivation by collisions with impurities of the solvent itself. Phosphorescence demands radiational deactivation of the triplet. Unfortunately, however, owing to its long lifetime, the triplet is prone to many deactivation processes possible in fluid solution at room temperature (where collisions are very probable). In fact, in such experimental conditions the value of

$$\Phi_{PP} = \frac{K_P}{K_P + K_{PQ}}$$

(see Figure 1) is close to zero because the velocity constant of the quenching processes,  $K_{PQ}$ , is much higher than the corresponding constant of the phosphorescence process ( $K_{PQ} \gg K_P$ ). If we wish to increase the phosphorescence quantum efficiency,  $K_{PQ}$  should be minimized. This can be achieved by resorting to protecting the phosphor in a more or less rigid medium, as is the case in rigid matrices (low-temperature rigid glasses, high-viscosity solvents, fluid-organized media, and solid substrates) where collisional deactivations are thus hindered. In this way,  $K_{PQ}$  may become smaller than  $K_P$  and so analytically useful phosphorescence can be obtained.

Other environmental factors that should be controlled are the presence of possible impurities with quenching properties (e.g., oxygen), the solvent used, and the pH value if the luminescent molecule exhibits acid-base properties. It is interesting to note that heavy atoms, but particularly paramagnetic species (e.g., oxygen and nitric oxide) can enhance both intersystem crossing and quenching of the triplet state, rendering it very difficult to predict the overall effect on phosphorescence signals observed.

For a complete picture, we should briefly refer to the influence of the chemical structure of the emitting molecule. As with fluorescence, aromatic compounds are most important because they may yield phosphorescence. A shift of this emission toward longer wavelengths with an increase in the number of condensed benzene rings is observed. Introduction of a heteroatom into the aromatic electron system (e.g., in a polyaromatic hydrocarbon) or substitution by atoms of high atomic number (e.g., the internal 'heavy-atom' effect of bromine or iodine) usually

increases the phosphorescence observed as both factors tend to enhance the rate of intersystem crossing and so of  $\Phi_T$ .

Phosphorescence of aliphatic compounds is rarer, with the exception of biacetyl ( $\text{CH}_3\text{CO}-\text{COCH}_3$ ), which gives strong phosphorescence, and some aliphatic ketones that have been shown to phosphorescence with lifetimes on the order of a few milliseconds.

It should be stressed, finally, that phosphorescence of an analyte, or its derivatized compound, is a rather complex process that can be affected drastically by slight modifications of the chemical structure or small variations of the chemical nature of the environment.

## Basic Instrumentation for Phosphorescence

The detection and measurement of phosphorescence signals requires a photoluminescence spectrophotometer. The basic instrument for phosphorescence is essentially a fluorimeter with five basic components:

1. A light source (e.g., a xenon lamp, a laser, a light-emitting diode).
2. Two dispersive devices (optical filters of various types, monochromators) that will allow selection of the appropriate excitation and emission wavelengths.
3. A sample compartment, where the sample cell is positioned for measurements (alternatively the implementation of optical fiber devices, which have gained recently an increased popularity in commercial instrumentation, allow to carry the excitation light from the source to the sample, and to collect the emission).
4. A transducer, such as a photomultiplier tube (or a multichannel charge-coupled device, CCD) that will convert the light intensity into electric current magnitude.
5. A data-readout device (an analog or, more usually today, digital recoding system).

Such an instrument measures total luminescence emitted, and so there is a need to distinguish between fluorescence emission and the desired phosphorescence signals. This is usually accomplished by taking advantage of the orders of magnitude differences between the characteristic lifetimes of the processes. The long-lived phosphorescence can be isolated from the shorter-lived fluorescence by resorting to a measuring system allowing out-of-phase delayed detection relative to the excitation of the sample.

The types of approaches available to do this are described below.

### Mechanically Chopped Systems: Phosphoroscopes

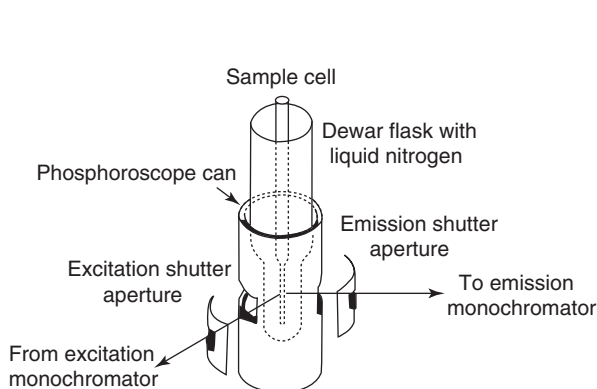
The 'phosphoroscope' is the simplest device for measuring phosphorescence, rejecting fast emissions such as scattered light and fluorescence. The exciting radiation is periodically interrupted by some type of mechanical 'chopper' and the emission is observed only after a time delay following the excitation cycle. Only the long-lived phosphorescence persists after a given delay time and can be measured selectively. Phosphoroscope design and mathematical expressions for the measured and instantaneous intensities of the phosphorescence were investigated in detail by O'Haver and Winefordner (see the legend of Figure 2). Two main designs of phosphoroscopes were investigated: the 'rotating can' and the 'rotating disk' or Becquerel type of phosphoroscope. These approaches, although very simple and commonly employed in the originally developed systems for phosphorescence measurements, are practically out of use today. However, considering their historical importance and didactic value a brief description of both approaches is given.

As shown in Figure 2, in the 'rotating-can' approach the phosphorescence of the sample, placed in the center of a rotating cylinder with two opposite apertures, is collected in the conventional mode at right angles to the exciting radiation. This radiation passes through the excitation shutter aperture and strikes the rotating can in such a way that one of its holes is aligned with the excitation shutter aperture the primary radiation reaches the sample cell, exciting the analyte. The holes (apertures) of the can are located in such a way that at the same moment the emission is blocked by the can wall and so no radiation

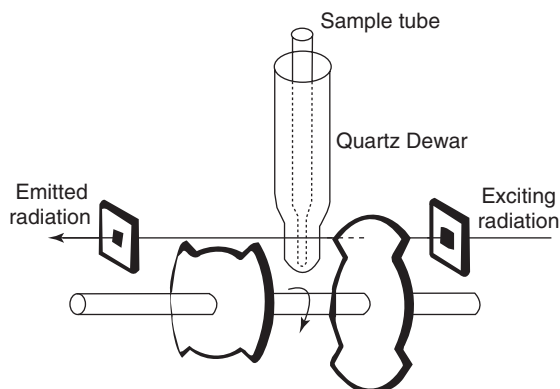
will reach the detector. As the phosphoroscope can rotate, an instant later, the excitation radiation is blocked by the wall while the other hole of the can now becomes aligned with the emission shutter aperture. In this way only phosphorescence remains while fast fluorescent and scattered radiation emissions will have decayed to negligible values after the cessation of excitation. Thus, if there is no 'delayed fluorescence' (a rare phenomenon, with longer lifetimes) only phosphorescence emissions will reach the emission monochromator and detector to be measured. For right-angled light observation, excitation takes place out of phase with the observed phosphorescent emission.

The phosphoroscope acts as an optical gating device or chopper of radiation so that the signal reaching the detector consists of a periodic series of optical pulses that are integrated in the averaged current finally measured. Therefore, this measured signal may be significantly decreased compared to that observed without the rotating phosphoroscope. The maximum rotational speed of most commercial phosphoroscopes is  $\sim 6 \times 10^3$  rpm and so the resulting cycling time of the phosphoroscope,  $t_c$ , is  $10^{-2}$  s. The ratio of the averaged DC currents observed with ( $I_P$ ), and without ( $I_{P'}$ ), the phosphoroscope is given by  $a = I_P/I_{P'}$  and depends not only on the value of  $t_c$  but also on the value of the phosphorescence lifetime,  $\tau$ , the value of the delay time,  $t_d$  (time elapsed between the end of the excitation and the beginning of the observation period) and, of course, the value of the observation time,  $t_p$ .

A less common type of phosphoroscope, not using right-angled excitation-emission geometry, is the 'rotating disk'. As shown in Figure 3, it consists of



**Figure 2** A conventional rotating-can phosphoroscope. (Reprinted with permission from O'Haver TC and Winefordner JD (1966) *Analytical Chemistry* 38: 603. © 1966 American Chemical Society.)



**Figure 3** A conventional rotating-disk (Becquerel type) phosphoroscope. (Reprinted with permission from Fischer RP and Winefordner JD (1972) *Analytical Chemistry* 44: 949. © American Chemical Society.)

two disks mounted on a common axis with excitation and emission slots aligned along the optical path. The sample tube, immersed in liquid nitrogen in a Dewar flask, as is usual in low-temperature phosphorescence (LTP), is inserted between these two disks, which have strategic openings (see Figure 3) that allow periodic delayed excitation and detection with continuous rotation of the axis.

### Electronic Systems: Pulsed-Source Gated Detector

At the beginning of the 1970s, Winefordner's group also developed the principles and instrumentation of an electronic approach for achieving time-resolved phosphorimetric measurements by resorting to the use of pulsed sources for excitation. A flash lamp, which provides pulses of short duration,  $t_f$  (e.g., 10  $\mu$ s), is operated in a synchronized manner with a specially operated photomultiplier (the 'gated-detection' system) able to 'read' the emissions from the sample that are delayed with respect to the excitation pulses. In fact, the gated detector can be turned off and on at different moments of consecutive excitation cycles: after a convenient 'delay time',  $t_d$ , has elapsed to allow for the light pulse to decay to a negligible value (see Figure 4), the detector is turned on and the phosphorescent emission is measured during a given 'gating time',  $t_p$ . The principle of operation of this electronic version of a phosphoroscope can be better explained by following a complete cycle of excitation/measurement as schematically shown in Figure 4. With an initial pulse of

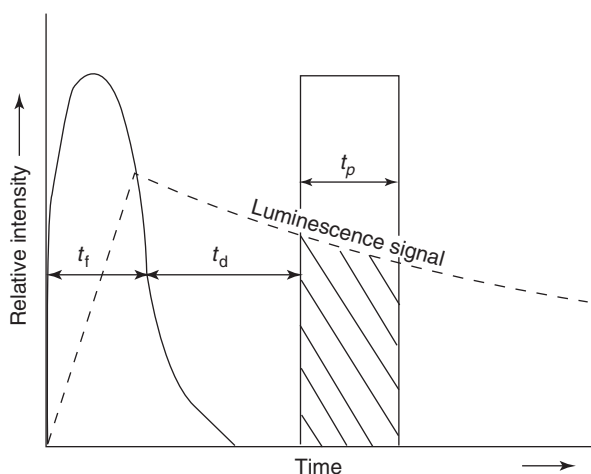
exciting light, lasting  $t_f$ , the phosphorescence signal increases to a maximum and then decays exponentially. While this happens the detector is 'closed', and only after a delay time,  $t_d$  in Figure 4, is it turned on, so that only the longer-lived emissions (phosphorescence) are thus monitored for a predetermined gating time,  $t_p$ . This cycle is repeated for each pulse of the lamp and integrated photoluminescence intensity observed is measured. The delay between excitation and observation of luminescence is thus secured in order to discriminate between phosphorescence and other shorter-lived emission phenomena.

Although there are still a few commercially available instruments for phosphorescence measurements based on mechanically chopped systems, nowadays most of the commercial luminometers used for phosphorescence measurements are based on pulsed-source/gated detections. These pulsed instruments are typically equipped with a high-power xenon flash-lamp and a photomultiplier detector.

A significant instrumental development is the use of lasers for excitation. In particular, a pulsed  $N_2$  laser constitutes a most adequate source both because it is a 'pulsed' source (and this excludes the need for mechanical phosphoroscopes) and because it provides a very intense coherent radiation in the UV region at 337.1 nm that is able to excite many organic molecules. Pulsed-source time-resolved phosphorimetry is clearly advantageous compared to mechanical chopping of excitation to obtain time discrimination because (1) better signal-to-noise ratios are achieved and so lower detection limits are possible; (2) as  $t_d$  and  $t_p$  can be varied within quite broad limits, better discrimination between phosphors of different lifetimes is possible (increased selectivity); (3) they provide the ability to measure shorter lifetimes (e.g.,  $\sim 10 \mu$ s, as compared to 100  $\mu$ s for a typical rotating-can phosphoroscope); and (4) one can straightforwardly obtain the corresponding phosphorescence decay curve and thus decide on the presence of simple or complex phosphorescence systems.

In this vein, these systems also allow to perform the phosphorescence lifetime decay evaluation. After excitation of the phosphorescent molecule by the pulsed excitation source the phosphorescence decay is collected with the light detector. The signal observed in the time domain is the phosphorescence emission at various times after the excitation pulse (a multipoint decay curve is obtained from which it is possible to obtain the phosphorescence lifetime after an appropriate fitting of the obtained curve to an exponential decay function by a least squares model).

However, phosphorescence lifetime detection can be also performed in the frequency domain. In this



**Figure 4** Schematic diagram of events occurring during one cycle of sample excitation and observation in a pulsed-source gated-detector phosphorimeter. (Reprinted with permission from Fisher RP and Winefordner JD (1972) *Analytical Chemistry* 44: 950. © 1972 American Chemical Society.)

latter approach, which is becoming increasingly popular, a sinusoidally modulated excitation source (typically a laser) provides the excitation energy for the phosphorescent molecule. The phosphorescence emission from the molecule will be then also modulated but is time delayed or phase shifted, relative to the excitation signal. For a typical single exponential decay of the phosphorescence, the relationship between the lifetime,  $\tau$ , and the corresponding phase shift,  $\phi$ , is

$$\tau = \frac{\tan \phi}{2\pi f}$$

where  $f$  is the modulation frequency used.

These systems are nowadays being preferred to the most common instruments based on pulsed excitation sources for lifetime evaluation of the phosphorescence phenomenon. However, an important disadvantage of phase phosphorimetry is that the signal-to-noise ratio decreases with increasing modulation frequency, and since the phase sensitivity increases with modulation frequency, an optimal frequency has to be selected.

## Selected Applications

Most of the past analytical work in phosphorimetry has been carried out by Winefordner's research group using the technique of LTP; that is, using a solution of the sample in an appropriate solvent that is frozen at liquid nitrogen temperatures to avoid collisional quenching. In a review from this group, it is stated that, in spite of the potential of LTP in organic trace analysis, this technique has failed to attract widespread acceptance for routine applications mainly because of two major practical drawbacks. One is the need for working at cryogenic temperatures and the other is the great inconvenience involved in introducing samples in such cryogenic conditions. Different techniques of RTP have emerged during the last decade in an attempt to overcome such limitations. The characteristics and applications of such RTP techniques are discussed elsewhere in this encyclopedia.

At this point, it should be stressed that conventional LTP, notwithstanding its limitations, has been demonstrated to be an interesting tool for the analysis of many biologically and environmentally important molecules. It offers low detection limits (in the microgram per liter region), great selectivity arising from its time-resolution capability, and wide ranging analytical calibration curves.

Some low-temperature applications involving adsorption of the liquid sample on to a solid surface have been reported. However, low-temperature

solution phosphorescence, in which a long tube containing the sample solution is immersed into a special Dewar that contains liquid nitrogen (Figures 2 and 3) to cool the analyte to 77 K, is by far the most common LTP technique. Sample is dissolved in an adequate solvent (e.g., *iso*-pentane) that forms a transparent glass by cooling where the analyte is trapped in such a rigid medium. Problems associated with adequate solvent selection, immersion of the sample tube, reproducibility of the signals, etc., are serious drawbacks.

Nevertheless, analytical problems can be found where such disadvantages can be outweighed by the convenient final resolution of the problem based on LTP data. On other occasions, both fluorescence and phosphorescence data can easily be obtained from a given sample, providing complementary information. One example is the use of phosphorescence measurements at low temperature to examine mixtures of aromatic compounds dissolved in *n*-alkanes (Shpol'skii spectrometry). The narrow-band fluorescence and phosphorescence spectra complement one another and have proved most useful in the analysis of complex samples such as multicomponent mixtures of polycyclic aromatic hydrocarbons (PAHs) in fuel materials.

The first quantitative application of phosphorimetry was proposed by Winefordner's group in the 1960s for the determination of aspirin in blood by LTP. It was the first evidence that phosphorimetry could be a useful methodology for quantitative analysis in 'real' samples. Since then, LTP has proved especially useful in the determination of low levels of well-known organic pollutants. PAHs and other dangerous substances (including nitrated PAHs, polychlorinated biphenyls, a wide range of other common pesticides, and even dibenzo-*p*-dioxins) have been determined at the microgram per liter level by LTP, with precision ranging from 2% to 10%.

Different methodological advances in phosphorimetry have been developed by using LTP, including the elucidation of the effect of the presence of heavy atoms on the phosphorescence signal or the demonstration of the advantages of time-resolved phosphorimetry for the determination of individual components with different phosphorescence lifetimes in a complex mixture. In fact, a great number of biochemically important compounds and metabolites have been determined in rather complex matrices by LTP (e.g., pesticides, sulfamides, anticoagulants, carcinogenic compounds, different metabolites of triptofan, vitamins, nucleosides, DNA, amino acids, purines, cytosine, cytidine, chlorophylls, etc.), although the vast majority of published applications



**Table 2** Analytical characteristics of low-temperature phosphorimetry for some selected drug and pharmaceutical compounds

Compound	Solvent	Wavelength (nm)		Lifetime, $\tau$ (s)	Detection limit ( $\mu\text{g ml}^{-1}$ )
		Ex	Em		
Caffeine	EtOH	285	440	2.0	0.2
Cocaine-HCl	EtOH	240	400	2.7	0.01
Lidocaine	EtOH	265	400	1.1	1.2
Morphine	EtOH	285	500	0.3	0.01
Phenobarbital	EtOH	240	380	1.8	0.1
Aspirin	EPA	240	380	2.1	0.1
Oestradiol	EtOH	292	403	2.0	0.3
Chlorpromazine-HCl	EtOH	320	490	0.07	0.03
Papaverine-HCl	EtOH	260	480	1.5	0.0005
Sulphacetamide	EtOH	280	410	1.3	0.0001
Carbazole	EtOH	341	436	7.8	0.001
Yohimbine-HCl	EtOH	290	410	7.4	0.01

EtOH, ethanol; EPA, eicosapentanoic acid.

refer to the analysis and control of drugs and pharmaceutical compounds. **Table 2** illustrates this, also giving analytical LTP characteristics for some selected drugs and pharmaceutical compounds.

It should be noted that although the sensitivity and selectivity of LTP are excellent, the technique was never considered as a routine technique, mainly due to the difficulties to introduce the sample for measurement and the drawback of the low temperatures required. In fact, the number of applications of LTP started to decrease from 1975. Emerging developments during the last decade came from room-temperature phosphorescence methodologies, which finally have displaced the LTP techniques.

See also: **Fluorescence**: Overview; Instrumentation. **Pesticides**. **Phosphorescence**: Room-Temperature. **Polycyclic Aromatic Hydrocarbons**: Determination; Environmental Applications.

## Further Reading

- Agbaria RA, Oldham PB, McCarroll M, McGown LB, and Warner IM (2002) Molecular fluorescence, phosphorescence, and chemiluminescence spectrometry. *Analytical Chemistry* 74: 3952–3962.
- Hurtubise RJ (1990) *Phosphorimetry: Theory, Instrumentation and Applications*. New York: VCH.
- Rao RP (1992) *Luminescence: Phenomena, Materials and Devices*. Commack, NY: Nova Science Pub.
- Schulman SG (1977) *Fluorescence and Phosphorescence Spectroscopy: Physicochemical Principles and Practice*. Elmsford, NY: Pergamon.
- Vo-Dinh T (1984) *Room Temperature Phosphorimetry for Chemical Analysis*. New York: Wiley.
- Ward JL, Walden GL, and Winefordner JD (1981) A review of recent uses of phosphorimetry for organic analysis. *Talanta* 28: 201.
- Winefordner JD, Schulman SG, and O'Haver TC (1972) *Luminescence Spectrometry in Analytical Chemistry*. New York: Wiley-Interscience.

## Room-Temperature

**M E Díaz García and R Badía Laíño**, University of Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by M E Díaz García and A Sanz-Medel, pp. 3942–3950, © 1995, Elsevier Ltd.

## Introduction

The long timescales associated with phosphorescence increases the likelihood that intermolecular quenching collisions, along with possible intramolecular

vibrational–rotational relaxations, take place at room temperature, so bringing about radiationless decay of triplet states. These facts explain why for some time it was believed that only cryogenic conditions (liquid nitrogen at 77 K), hindering bimolecular quenching and intramolecular relaxations, were suitable for analytically useful phosphorescence measurements.

It has been shown, however, that phosphorescence signals can be obtained at room temperature by restricting competing deactivation mechanisms, e.g., in the gas phase at sufficiently low pressure the

**Table 2** Analytical characteristics of low-temperature phosphorimetry for some selected drug and pharmaceutical compounds

Compound	Solvent	Wavelength (nm)		Lifetime, $\tau$ (s)	Detection limit ( $\mu\text{g ml}^{-1}$ )
		Ex	Em		
Caffeine	EtOH	285	440	2.0	0.2
Cocaine-HCl	EtOH	240	400	2.7	0.01
Lidocaine	EtOH	265	400	1.1	1.2
Morphine	EtOH	285	500	0.3	0.01
Phenobarbital	EtOH	240	380	1.8	0.1
Aspirin	EPA	240	380	2.1	0.1
Oestradiol	EtOH	292	403	2.0	0.3
Chlorpromazine-HCl	EtOH	320	490	0.07	0.03
Papaverine-HCl	EtOH	260	480	1.5	0.0005
Sulphacetamide	EtOH	280	410	1.3	0.0001
Carbazole	EtOH	341	436	7.8	0.001
Yohimbine-HCl	EtOH	290	410	7.4	0.01

EtOH, ethanol; EPA, eicosapentanoic acid.

refer to the analysis and control of drugs and pharmaceutical compounds. **Table 2** illustrates this, also giving analytical LTP characteristics for some selected drugs and pharmaceutical compounds.

It should be noted that although the sensitivity and selectivity of LTP are excellent, the technique was never considered as a routine technique, mainly due to the difficulties to introduce the sample for measurement and the drawback of the low temperatures required. In fact, the number of applications of LTP started to decrease from 1975. Emerging developments during the last decade came from room-temperature phosphorescence methodologies, which finally have displaced the LTP techniques.

See also: **Fluorescence**: Overview; Instrumentation. **Pesticides**. **Phosphorescence**: Room-Temperature. **Polycyclic Aromatic Hydrocarbons**: Determination; Environmental Applications.

## Further Reading

- Agbaria RA, Oldham PB, McCarroll M, McGown LB, and Warner IM (2002) Molecular fluorescence, phosphorescence, and chemiluminescence spectrometry. *Analytical Chemistry* 74: 3952–3962.
- Hurtubise RJ (1990) *Phosphorimetry: Theory, Instrumentation and Applications*. New York: VCH.
- Rao RP (1992) *Luminescence: Phenomena, Materials and Devices*. Commack, NY: Nova Science Pub.
- Schulman SG (1977) *Fluorescence and Phosphorescence Spectroscopy: Physicochemical Principles and Practice*. Elmsford, NY: Pergamon.
- Vo-Dinh T (1984) *Room Temperature Phosphorimetry for Chemical Analysis*. New York: Wiley.
- Ward JL, Walden GL, and Winefordner JD (1981) A review of recent uses of phosphorimetry for organic analysis. *Talanta* 28: 201.
- Winefordner JD, Schulman SG, and O'Haver TC (1972) *Luminescence Spectrometry in Analytical Chemistry*. New York: Wiley-Interscience.

## Room-Temperature

**M E Díaz García and R Badía Laíño**, University of Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by M E Díaz García and A Sanz-Medel, pp. 3942–3950, © 1995, Elsevier Ltd.

## Introduction

The long timescales associated with phosphorescence increases the likelihood that intermolecular quenching collisions, along with possible intramolecular

vibrational–rotational relaxations, take place at room temperature, so bringing about radiationless decay of triplet states. These facts explain why for some time it was believed that only cryogenic conditions (liquid nitrogen at 77 K), hindering bimolecular quenching and intramolecular relaxations, were suitable for analytically useful phosphorescence measurements.

It has been shown, however, that phosphorescence signals can be obtained at room temperature by restricting competing deactivation mechanisms, e.g., in the gas phase at sufficiently low pressure the



frequency of molecular collisions is reduced. Quenching of the triplet state by oxygen in the vapor state can be decreased by using an inert gas and, thus, low phosphorescence intensity of some compounds can be observed. An alternative route to prevent radiationless deactivation of the triplet at room temperature is by adsorbing the phosphors on solid supports i.e., solid supports room temperature phosphorescence (SS-RTP). The use of SS-RTP for sensitive detection of 18 organic compounds was first described in 1967 by Roth. Since its beginning, the technique has been extended to numerous other species and has grown in scope, applicability, and utilization.

However, in spite of the large interest of RTP, only in the early 1980s, after the pioneering work of Cline Love, did systematic research start on the development of RTP in liquid aqueous solution when the emitting triplet molecules were incorporated into a conventional micelle under deoxygenated conditions. This phenomenon, generally referred to as micelle-stabilized room-temperature phosphorescence (MS-RTP), attracted considerable attention and rapidly found its place as a new analytical technique, allowing the use of RTP as a detection mode in flow techniques, e.g., flow injection analysis (FIA) or liquid chromatography (LC). The successful application of MS-RTP was the starting point for a large number of studies in the area and it opened up a new world of stabilized room-temperature phosphorescence procedures; thus, cyclodextrin-induced (CD-RTP), vesicle-stabilized, microemulsion-stabilized, and bio-macromolecular-stabilized room-temperature phosphorescence experimental procedures have been developed subsequently.

Room-temperature phosphorescence in liquid solutions (LS-RTP) is a further step toward practical applications of phosphorescence. The observation of RTP of biacetyl in liquid solution or that of eosin in glycerol is commonly viewed as odd exceptions to the striking and persistent failure to observe phosphorescence from organic molecules at room temperature in liquid solution. Only after several scattered reports of the observation of long-lived emission from a variety of organic dyes in deoxygenated solutions did Parker and Joyce demonstrate that the use of a carefully purified and deoxygenated special solvent (such as perfluorocarbons) allowed observation of weak LS-RTP. Although this type of LS-RTP has failed to catch on as a popular method for routine analysis, Donkerbroed *et al.* applied the technique of sensitized phosphorescence (SP) in liquid solutions as a useful alternative to direct solution RTP. The knowledge of the spectroscopic behavior of lumiphors under various conditions is crucial for the design of new RTP-based assays. As an example, the

applicability of LS-RTP is expected to change with the introduction of modern nonprotected room-temperature phosphorescence (NP-RTP), developed around the use of adequate heavy atoms in solution. An excellent review covering the analytical applications of room-temperature phosphorescence in the liquid state has been produced by Gooijer *et al.*

The sections that follow will focus on the most useful of these new analytical RTP approaches, starting with fluid solution techniques (RTP in ordered media, NP-RTP, sensitized RTP) and ending with the principles and recent applications of SS-RTP techniques.

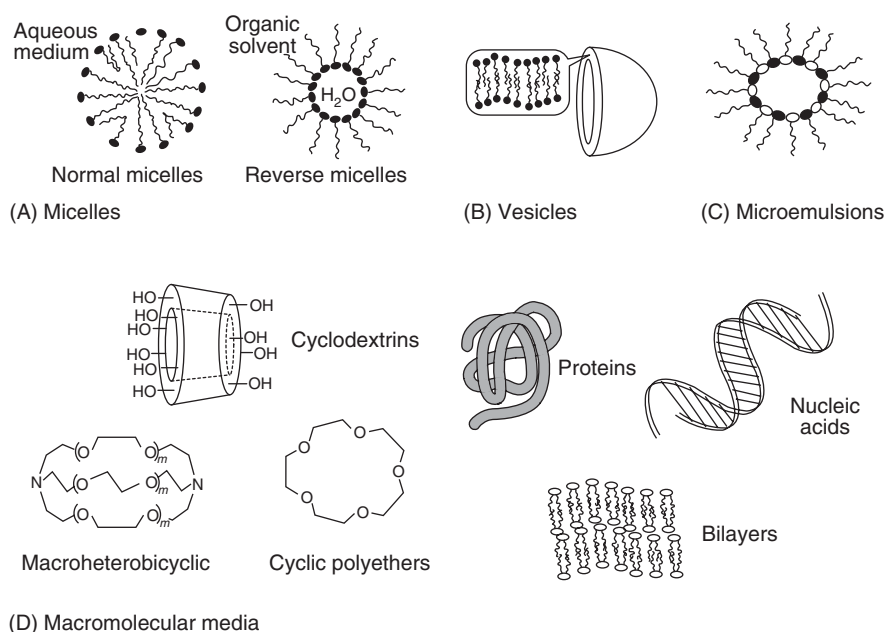
## Room-Temperature Phosphorescence in Ordered Media

### Micelle Stabilized Room-Temperature Phosphorescence

Surface-active agents are amphiphilic molecules in which a polar 'head' group is attached to a long nonpolar chain. The nonpolar moiety is usually a hydrocarbon tail and the polar region can be either an ionic or a polar neutral group. Above a certain concentration, the critical micelle concentration (CMC), these amphiphilic molecules associate dynamically in solution to form relatively well-defined aggregates termed micelles (see **Figure 1**).

Micelle characteristics depend upon the chemical structure of the surfactants and the nature of the solvent. In aqueous solution, amphiphiles assemble in such a way that the tails are packed together, forming a nonpolar core, while the polar head groups form a boundary surface between the core and the isotropic aqueous solution (normal micelles). Surfactants can also form micelles in nonaqueous solvents of low polarity. The monomers in these reverse micelles are oriented in the opposite sense to those in aqueous solutions (**Figure 1**). Anionic or cationic long-chain dialkyl surfactants generate bilayers and vesicles either by sonic dispersal in water or by gentle injection of their nonaqueous solution into pure water. Synthetic surfactant vesicles are sealed membrane capsules (**Figure 1**) that enclose aqueous volumes of 1–1000  $\mu\text{m}^3$  and compared to micellar media are much less dynamic and more stable.

These organized media possess many unique properties including their ability to solubilize, concentrate, organize, and localize solutes, to modify the spectral parameters and the effective microenvironment about solubilized molecules, to alter chemical and photochemical pathways and rates, as well as the position of chemical equilibrium processes, among other effects.



**Figure 1** Schematic representation of different ordered macromolecular systems.

Cline Love and colleagues showed that there are three fundamental requirements for observing MS-RTP for most lumophores: (1) securing the presence of micellar aggregates (using a surfactant concentration above its CMC); (2) the presence of a heavy atom or a heavy species; and (3) ensuring oxygen scavenging. The protective screening effect of micelles against external quenchers by compartmentalization of solubilized phosphor molecules greatly reduces collisional quenching of triplet molecules by quenchers. Only at surfactant concentrations well above the CMC is the protective effect on phosphorescence obvious (micellar effect). Relative phosphor-micelle charge-charge interactions are found to be also important. Negatively charged molecules in cationic micellar media and negatively charged phosphors in cationic micellar experience attractive electrostatic interactions that may concurrently act with hydrophobic ones to provide the 'rigidity' needed to observe strong RTP signals. On other hand, if micelles are 'functionalized' with a heavy atom, e.g., anionic micelles with  $\text{I}^-$ , its 'effective local concentration' within the micelle is considerably larger than its actual molar concentration. In this way the rate of intersystem crossing,  $K_{\text{isc}}$ , is greatly enhanced. At the same time, sheltered by the micelle, the nonradiative deactivation of the triplet state is prevented. Both effects result in an enhanced quantum yield of phosphorescence. By prudent choice of heavy atom and surfactant assemblies, it is possible to induce MS-RTP in a variety of compounds.

Oxygen is a most potent quenching agent of the excited triplet state. The use of an inert gas (e.g., nitrogen) to remove dissolved oxygen used to be the most common technique for oxygen scavenging, but this generates a significant amount of foam. Díaz-García and Sanz-Medel studied in great detail the use of sodium sulfite as a chemical oxygen scavenger in micellar media for MS-RTP. The technique based on the reaction  $2\text{SO}_3^{2-} + \text{O}_2 \rightarrow 2\text{SO}_4^{2-}$  is simple and reliable and is now widely accepted to develop different RTP approaches.

Among the different possibilities, normal micelles have been the organized media more commonly used to obtain RTP in fluid solution: in a classic experiment, the lumophore is dissolved in a sodium lauryl sulfate (SLS) solution and then diluted with thallium lauryl sulfate-SLS solution to give a total detergent concentration of  $0\text{--}10\text{ mol l}^{-1}$  with a 30:70 ratio of  $\text{TI}/\text{Na}$ . The solution is deoxygenated using sodium sulfite as oxygen scavenger and the sample can be analyzed using time discrimination. Most of pioneering analytical work done on MS-RTP refers to the determination of various organic compounds of significance in clinical (drugs) and environmental (polycyclic aromatic hydrocarbons, PAHs) chemistry.

Because MS-RTP permits the use of fluid solutions at room temperature, a further possibility offered by MS-RTP is that of detector in LC. This point is more interesting for analysis of body fluids, where the sample matrix produces intense background that severely reduces the signal-to-background ratio.

MS-RTP techniques can be applied to LC in two different ways:

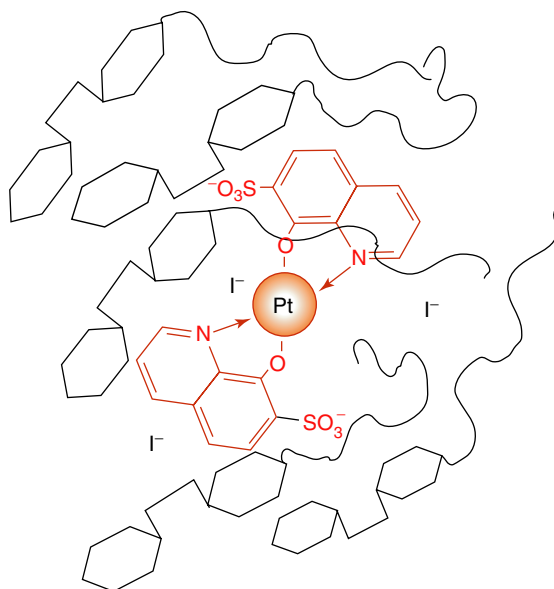
1. The micellar phase is used as a mobile phase. In this context a typical LC/MS-RTP procedure for PAHs including naphthalene, biphenyl, and phenanthrene using an SLS/thallium laurylsulfate (70%/30%) micellar mobile phase exhibits the following features: working ranges covered more than three orders of magnitude and the detection limits were in the low nanogram range for the PAHs investigated in coal liquids and wastewater products.
2. In a postcolumn addition mode, the micellar solution is added to the mobile phase after chromatographic separation, as demonstrated also for aromatic compounds (phenanthrene, biphenyl, 2-naphthol).

Micellar eluents have been used with reversed-phase LC columns to analyze a variety of phosphor solutes by direct sample injection (e.g., serum samples). Avoidance of the sample preparation step entirely is a considerable advantage since it simplifies and significantly shortens an assay, thus improving precision. However, some factors have delayed the application of MS-RTP as detector in chromatographic separations: the foaming problems generated by nitrogen purging (which may be solved by using alternative deoxygenation schemes), the preparation of the micellar mobile phase must be carefully controlled in order to avoid eluent conditions that are unfavorable for micelle formation or favor micelle destruction. For this reason surfactant concentration must be maintained well above the CMC and the organic solvent composition should be kept relatively low (especially in the postcolumn mode).

Metal-ion MS-RTP determinations have also been demonstrated. For example, platinum(II) chelates with 8-quinolinol derivatives were found to form highly phosphorescent complexes at room temperature in the sugar-based *n*-dodecyl- $\beta$ -D-maltoside micellar medium, using iodine as heavy atom (see Figure 2).

The applicability of MS-RTP can be extended as a detection mode for metal ion determinations in flowing systems. The simplest scenario involves the use of a typical FIA configuration and a sodium sulfite deoxygenated micellar stream containing the metal–ligand lumophor, into which the deoxygenated metal ion solution is injected. After mixing, MS-RTP peaks are registered. Sulfite deoxygenation of micellar solutions greatly simplifies FIA–MS-RTP techniques.

Time-resolved phosphorimetry has also been proposed as a means of increasing the selectivity of



**Figure 2** Simplified overview of the palladium–ferron complex in sugar-based anionic micellar media.

MS-RTP metal determination mixtures. In this approach, a pulsed radiation source excites the sample and the decay profile is monitored. Thus, it provides a measure of the time dependence of phosphorescence intensity after the excitation pulse, which can also be made as a function of emission wavelength. For a binary system of components, the MS-RTP signal from a rapidly decaying species (e.g., Zn(II)–Ferron chelate in cetyltrimethylammonium bromide (CTAB) micelles) and that from a slowly decaying species (e.g., Al(III)–Ferron complex in CTAB micelles) are related through the expressions:

$$I_t^{\text{Zn}} = I_0^{\text{Zn}} e^{-t/\tau}$$

$$I_t^{\text{Al}} = I_0^{\text{Al}} e^{-t/\tau}$$

$$I_t^{(\text{Zn}+\text{Al})} = I_t^{\text{Zn}} + I_t^{\text{Al}}$$

where  $I_t$  is the total MS-RTP intensity of the mixture at a delay time  $t$ , and  $I_0$  the MS-RTP intensity at time  $t=0$ . For multicomponent systems similar expressions can be derived. Selectivity of MS-RTP for metal-ion determinations in mixtures can be further enhanced through multidimensional and mathematical approaches. Table 1 provides an overview of selected miscellaneous applications of MS-RTP.

### Cyclodextrin Room-Temperature Phosphorescence

Cyclodextrins (CDs) are macrocyclic oligosaccharides consisting of  $\alpha$ -(1,4)linkages of D-(+)-glucopyranose units arranged in a torus (see Figure 1).

**Table 1** Analytical figures for MS-RTP determination of several analytes

Analyte	Reaction medium	$\lambda_{\text{exc}}$ (nm)	$\lambda_{\text{em}}$ (nm)	pH	DL	RSD%
<i>Pharmaceutical/drug</i>						
Naproxen (anti-inflammatory)	TI (I)/SDS	330	551	7.1	$0.03 \mu\text{g ml}^{-1}$	1.2–2.1
Allopurinol (inhibitor)	TI (I)/SDS	286	420	7.0	$0.014 \mu\text{g ml}^{-1}$	0.36–0.41
Nafronyl (vasodilator)	TI (I)/SDS	288	491	10.5	$6.1 \text{ ng ml}^{-1}$	2.4
Nafcillin (antibiotic)	TI (I)/SDS	284	540	7.2	$0.18 \text{ ng ml}^{-1}$	3.8
<i>Environmental compound</i>						
1-Naphthalenacetamide	TI(I)/SDS	292	524	—	$25 \text{ ng ml}^{-1}$	2.55
$\beta$ -Naphthoxyacetic acid	TI(I)/Triton-X-100	326	540	—	$0.13 \mu\text{g ml}^{-1}$	3.55
Pyrene	TI (I)/SDS	N <sub>2</sub> laser	550/590	6.82	$1 \times 10^{-10} \text{ mol l}^{-1}$	6.8
PAHs (Phenanthrene, pyrene, fluoranthrene, benzo[a]anthracene, 1,5-dimethylnaphthalene)	TI (I)/SDS	—	—	7.04–7.4	$(1.7\text{--}3.4) \times 10^{-7} \text{ mol l}^{-1}$	0.51–3.2
Carbazole	TI (I)/SDS	338	440	—	$1 \times 10^{-7} \text{ mol l}^{-1}$	—
<i>Metals</i>						
Eu(III)	Tetracyclines/ Triton-X100 or CPCI	394	594/618	8.0	$2 \times 10^{-9} \text{ mol l}^{-1}$	—
Ga(III), Al(III), Nb(IV)	Ferron/CTAB	396, 386, 363	615, 586, 572	5.5–5.7	$4\text{--}5.4 \text{ ng ml}^{-1}$	2–4.5
Pd(II)	Coproporphyrin IV/ Triton X-100	395	666	1.0	$0.18 \text{ ng ml}^{-1}$	2

SDS, sodium laurylsulfate; CTAB, cetyltrimethylammonium bromide; CPCI, cetylpyridinium chloride; Ferron, 7-iodide-8-hydroxyquinoline-5-sulphonic acid.

Three types of CDs are commercially available, the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -CDs, composed of six, seven, and eight glucose units, respectively. CDs are cylinder-shaped molecules with an axial void cavity. Their outer surface is hydrophilic that makes them soluble in water, while the interior of the toroidal cavity is highly hydrophobic and constitutes a relatively isolated microenvironment that can selectively ‘encapsulate’ lumophores of appropriate molecular dimensions. CDs are chemically and structurally quite stable, in contrast with micellar aggregates formed by monomer surfactant units in dynamic equilibrium with the micelle.

As a result of the host–guest complexes formed by CDs with a variety of lumophores, characteristic luminescence properties of the included lumophore, sensitive to viscosity/polarity or dielectric, may change. In fact, the inclusion of a lumophore in a CD cavity has proved to provide a means of triplet-state stabilization, luminescence intensity enhancement, optical activity induction, improved solubility of poorly soluble analytes in water, etc.

As in MS-RTP, a heavy atom (external or intramolecular) or a heavy atom molecule should be present to populate the excited triplet state. In contrast to MS-RTP, an exciting feature of CDs is their seeming ability to partially protect the phosphor’s triplet state from quenching by dissolved oxygen

in solution. Nonetheless, deoxygenation of samples results in increased phosphorescence intensity.

Many interesting applications of CD-RTP have concentrated on the determination of PAHs and nitrogen-containing heterocycles. Successful applications of CD-RTP based on host size-dependent selectivity (molecular recognition) can be envisaged. For example, in  $\beta$ -CD solutions, intense RTP is observed for the smaller naphthalene (which fits neatly into the  $\beta$ -CD cavity) and no phosphorescence is detectable from a larger molecule such as 1-phenylnaphthalene.

In a typical experiment for CD-RTP analysis of PAHs an aliquot of the compound of interest is added to a flask and the solvent is evaporated gently on a hot plate. An aliquot of 1,2-dibromomethane or 2-bromoethanol (heavy-atom species) is then added, followed by an aliquot of  $0.1 \text{ mol l}^{-1}$  sodium sulfite (oxygen scavenger), and final dilution with  $0.01 \text{ mol l}^{-1}$  aqueous CD solution. The solution is shaken vigorously by hand. Some precipitation, usually due to inclusion of excess heavy atom by CD, may cause cloudiness in the solution, which does not affect the reproducibility and quality of phosphorescence spectra. Sample preparation time is less than 5 min in this method. Cline Love and colleagues reported RTP detection limits for several polynuclear aromatic compounds (naphthalene, biphenyl, phenanthrene, etc.) in CD solution in the range  $10^{-6}\text{--}10^{-7} \text{ mol l}^{-1}$ .

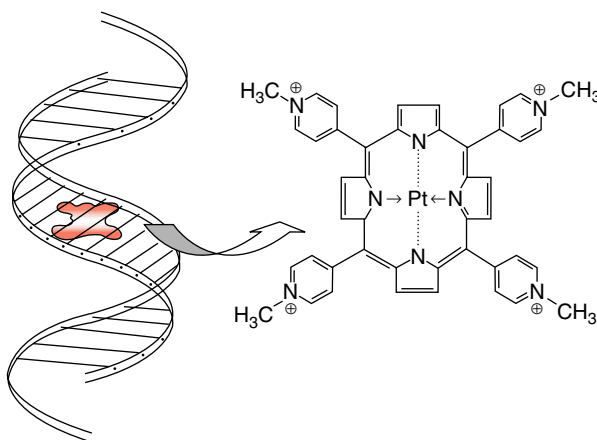
As in the case of MS-RTP, CD-RTP has been used as detector in LC. Unless the problems of low efficiency with CD mobile phases in chromatography are resolved, CD-RTP will not be a competitive alternative to the MS-RTP detection approach.

### Biomacromolecular Stabilized Room-Temperature Phosphorescence

Biological supramolecular assemblies such as proteins, nucleic acids, or membranes as well as single cells and tissues provide particularly useful microenvironments where lumophors (metal complexes, molecules, metal ions) can be protected or isolated from contact with neighboring quenchers. In these microdomains, molecular motions are highly restricted so that luminescence properties of those lumophors may be modified. The binding of the Pd-meso-4-(tetracarboxyphenyl)porphyrin complex to bovine serum albumin reduces the phosphorescence quenching constant by an order of magnitude. This effect can be explained as due to the protection experienced by the porphyrin complex buried inside the macromolecular structure.

Many metal complexes have been used in biological studies, most of them based on Pt- or Pd-porphyrin complexes. They exhibit strong phosphorescence in the near-infrared region (760–800 nm) and can be excited either at Soret (380–430 nm) or Q bands (500–560 nm). In order to observe RTP, oxygen must be removed from the samples. Deoxygenation schemes using chemical oxygen scavengers may result in damage of some biological structures, so enzymatic systems are recommended. To deplete oxygen in biological media a mixture of glucose and catalytic amounts of glucose oxidase/catalase can be used. The system ascorbate/ascorbate oxidase is also useful as it reduces oxygen to water without generation of hydrogen peroxide (which also may damage biological structures).

Applications of RTP probes in biological systems are typically linked to their interesting photophysical and photochemical properties. For example, the oxygen-dependent quenching of Pd-porphine complexes has been used to measure the  $O_2$  dependence of respiration by suspensions of mitochondria and cells and to develop two-dimensional maps of  $O_2$  pressure in tissues. Selective determinations of DNA in the presence of RNA can be achieved using the RTP enhancement of a Pd-porphine complex when bound to DNA. This enhancement arises from the hindered rotational motion of the DNA bound complex (Figure 3). A linear calibration graph was obtained up to  $6 \times 10^{-5} \mu\text{mol l}^{-1}$  DNA, the limit of detection being  $5 \times 10^{-8} \mu\text{mol l}^{-1}$ . The nucleic acid



**Figure 3** Pd-meso-tetra(4-crobyphenyl)porphine in a nucleic acid microdomain.

was assayed with 2.6% RSD and the method applied to DNA determination in tissues.

Pt- and Pd-coproporphyrin complexes may also be used as labels for antibodies to develop phosphorescence immunoassays. Also, RTP provides a sensitive tool for diagnosis of vascular disorders and image oxygen distribution in cancerous tissues (porphyrins are known to concentrate in tumors). Many other potential uses of RTP in biological systems include the study of processes such as rotational motion, distances, excited state reactions. RTP measurements have an important role in the study of biological systems and areas for future development are apparent.

### Sensitized and Quenched Room-Temperature Phosphorescence

'Sensitized-RTP' in solution is a method of quenching based on detection of the phosphorescence emitted by an 'acceptor'. The donor compound acts as a sensitizer, i.e., after being excited and before radiationless decay, transfers its energy to the acceptor, thus producing an acceptor triplet with a high phosphorescent yield. Sensitized-RTP has been observed for a variety of organic species (polynuclear aromatics, drugs, etc.) and compounds that do not luminescence in liquid solutions. Biacetyl and 1,4-dibromonaphthalene appear to be appropriate acceptors. Only ~10–15% of the molecules that fulfill the energy requirements are found to undergo SP.

Energy transfer between the donor and the acceptor is a probabilistic event. The ability of ordered media to bring the donor and acceptor molecules into close proximity for effective energy transfer and the partial protection of the excited phosphor provide further evidence of how the potential applications of sensitized-RTP in organized media might be. So, on dissolving the donor and the acceptor in



micellar or CD media, an enhancement of SP intensity and selectivity may occur. For benz[*a*]anthracene and anthracene determination in the presence of pyrene, for example, this approach improved the selectivity by factors as high as 200–220, when using an SLS micellar solution, thallium(I) as heavy atom, sodium sulfite as oxygen scavenger, and triphenylamine and acridine orange as triple energy donors.

A complementary technique to sensitized-RTP is quenched-RTP in solution, in which the presence of the analyte leads to a decrease of direct RTP intensity of biacetyl or other phosphorescent probes in ordered media. Energy transfer and electron transfer are the main pathways that deactivate the photoexcited state of the selected phosphors. For example, oxygen can be determined by quenched-RTP of palladium and platinum derivatives of porphyrins dissolved in ordered biological media (proteins). This approach also holds considerable potential for use in continuous-flow detector systems, including LC and capillary electrophoresis (CE), as it can be used for species that may have inherently poor detection properties by other techniques. So, it can be used to detect and quantify non-ultraviolet-absorbing compounds in ion chromatography (nitrite, sulfite, *cis*-platin, carboplatin, etc.). Also, quenched phosphorescence has been used as a highly sensitive detection mode for CE- and CD-based electrokinetic chromatography. Using conventional, lamp-based detection systems, detection limits have been reported to be in the  $10^{-7}$ – $10^{-8}$  mol l<sup>-1</sup> range and absolute detection limits  $\sim 1$  fmol. These figures are beyond the possibilities of many other techniques, especially for compounds with poor chromophores.

## Nonprotected Room-Temperature Phosphorescence

In 1997, a new concept was introduced in room-temperature phosphorescence in liquid solution: non-protected or heavy atom induced room-temperature phosphorescence, in which the presence of a protecting medium is not a necessary condition for RTP emission. The term NP-RTP seems to be more appropriate for the technique as the heavy atom perturber is a necessary condition for all RTP emission (exception made for biacetyl, which does not meet this requirement). This approach is an important technical advance, since numerous organic molecules can be determined through a very simple technique. As in all RTP methods, the following are necessary: (1) the presence of heavy atoms such as KI or TiNO<sub>3</sub> and (2) removal of oxygen (sodium sulfite has been found to be a suitable chemical oxygen scavenger).

Since its introduction, determination of several analytes and their NP-RTP characteristics have been reported, among which plant growth regulators ( $\alpha$ - and  $\beta$ -naphthoxyacetic acids), drugs, pharmaceuticals and pesticides (naphazoline, naproxen, thiazobenzodazole, carbaryl, acetamide), PAHs (fluorene, naphthalene), and other compounds (trypsin, tryptophan, indole-3-butyric acid, acridine derivatives) have been studied.

The full sensitivity potential of NP-RTP can be exploited as detector in flowing systems, as illustrated for the sensitive determination of the  $\beta$ -lactamic antibiotic nafcillin using a flow-injection approach. The NP-RTP of nafcillin was developed using potassium iodide as heavy atom promoter and sodium sulfite as oxygen scavenger. The system is capable to detect micromolar levels of nafcillin (detection limit 0.53  $\mu$ mol l<sup>-1</sup> nafcillin) and a linear calibration graph was obtained for concentrations of the antibiotic up to 85  $\mu$ mol l<sup>-1</sup>. On the other hand, and by contrast with MS- and CD-RTP, the coupling of NP-RTP with separation methods is expected to be quick and simple to perform as no micelles or CDs are necessary. However, it is known that some common organic solvents used in chromatographic separations may reduce phosphorescence in several NP-RTP systems. So, the technique may be limited to aqueous-phase separations.

NP-RTP is in its infancy and basic research on the origin of the NP-RTP phenomenon as well as the basic factors governing it have not yet been established. Looking further into the future, the technique could find many applications for both the monitoring and/or mapping of environmental or biologically important molecules.

## Solid-Surface Room-Temperature Phosphorescence

Since its introduction in 1967, the field of SS-RTP has expanded considerably. A number of organic compounds, polar and nonpolar, exhibit RTP when adsorbed on a suitable solid support. Because many of the solid supports used to develop RTP give background phosphorescence, only certain materials are useful for the task, among which sodium acetate shows the lower background emission. The phosphor molecules can be held tightly to the solid support through hydrogen bonding of polar or ionic organic compounds to hydroxyl groups of paper or to the surface silanol groups of silica gel as well as through hydrophobic interactions between the phosphor molecule and the solid network. This restriction of intramolecular rotation of the phosphor is believed

to be responsible for enhancement of the radiative decay of the triplet state ('rigidly held mechanism').

In order to observe analytically useful emissions from compounds adsorbed on paper and other solid supports, thorough drying of the sample has been found to be necessary. Moisture competes with the phosphor molecule for hydrogen-bonding sites on the adsorbent, thus increasing the mobility of the phosphor and the chances of collisional deactivation. In most cases, RTP is so sensitive to moisture that this phenomenon may be used in the determination of humidity.

It has been suggested that accessibility of oxygen to a solid surface-immobilized phosphor in a dry atmosphere is less favorable than for a wet atmosphere. Oxygen quenching efficiency is greatly enhanced in wet atmospheres as water molecules weaken the support-phosphor interactions and the integrity of the sample substrate, thus allowing oxygen to diffuse and quench the phosphor. For maximum SS-RTP intensity dry and inert environments are recommended.

The addition of external heavy-atom perturbers (such as thallium, lead, thorium, and many others) to RTP samples generally increases the radiative rate of triplet-state decay. The heavy atom effect is most striking amongst the nonionic organic compounds, where phosphorescence appears only if the heavy atom is present. The RTP intensity is highly dependent on the concentration of heavy atom perturber and significant enhancement of SS-RTP is observed when using halide salts as external heavy atom perturbers. For halides the enhancement follows the expected trend of  $I^- > Br^- > Cl^- > F^-$  with virtually no enhancement from fluoride or chloride ions and  $Tl^+ > Ag^+ > Pb^{2+} > Hg^{2+}$  for cations. Although SS-RTP spectral features are similar to those observed at low temperatures, RTP spectra generally provide less resolution owing to increased vibrational freedom of the molecules at the higher temperature. The nature of the solid support employed in SS-RTP usually has only a marginal effect on spectral features, but RTP intensity and lifetimes can be significantly affected. A considerable amount of SS-RTP research involves collecting analytical data on model compounds adsorbed on solid surfaces in the hope of developing a general working mechanism to explain the interactions required for RTP.

Significant enhancements in RTP intensity and reduction in lifetimes can be observed when the microenvironment experienced by a phosphor adsorbed onto the solid support is manipulated by the use of ordered media as solid substrate modifiers. For example, enhancements of the paper RTP sensitivity of several PAHs ranging from factors of

2 to 9 were found when an anionic surfactant was added or when the phosphors were spotted from micellar solutions. SS-RTP exhibits in all cases poorer detection limits than low-temperature phosphorescence.

Although for those nonexperts SS-RTP would hardly be the technique of choice if fluorescence is a good alternative, SS-RTP is applicable to a wide variety of compounds and samples in many areas, ranging from clinical analysis, biomedical research, pharmaceutical, agricultural chemistry, and energy technology to biomedical, environmental, and industrial control. The technique is inexpensive and small sample sizes can be employed, which is particularly advantageous for clinical and environmental samples. An excellent example of this is the determination of PAHs that are widely dispersed in the environment, often in small quantities. A striking personal dosimeter badge, based on molecular diffusion and direct detection by paper RTP of PAH vapors, at the ng per g level, with precision ranging from ~3% to 10%, has been described by Vo-Dinh, demonstrating the versatility of the technique. Hurtubise has studied in detail the best conditions to obtain solidmatrix room-temperature phosphorescence from femtomole levels of pollutants, specially from PAHs, their DNA adducts, as well as from their metabolic carcinogens (e.g., tetrols, which are products obtained from the hydrolysis of benzo[a]pyrene-DNA adducts). The limit of detection of the benzo[a]pyrene-DNA adducts was reported to be two adducts in 107 bases using both Whatman 1PS paper and a 30%  $TiNO_3/NaOAc$ . It should be mentioned that there are very few methods for the direct determination of benzo[a]pyrene-DNA adducts and SS-RTP offers the advantages of sensitivity, selectivity, no radioactive materials are used and it is developed in mild room-temperature conditions.

Another example of the power that SS-RTP affords can be seen in the use of time-resolved phosphorimetry to distinguish several phosphors with overlapping spectra but with different lifetimes such as PAHs. Two major revisions on the SS-RTP field have been produced by Hurtubise and Vo-Dinh.

SS-RTP exhibits in all cases poorer detection limits than low-temperature phosphorescence. However, it is potentially useful as a routine analytical method because (1) no cryogenic equipment, expensive and rather cumbersome to use, is needed; (2) no time-consuming degassing of the solvent is mandatory; and (3) chromatographic separations can be performed on the substrate before the analysis. These features make SS-RTP particularly suitable to exploit new detection schemes. So, SS-RTP is a convenient means of observing delayed fluorescence in those

compounds with close enough spacing between  $S_1$  and  $T_1$  levels since adequate thermal energy is available to the molecule to promote reverse inter-system crossing in this situation.

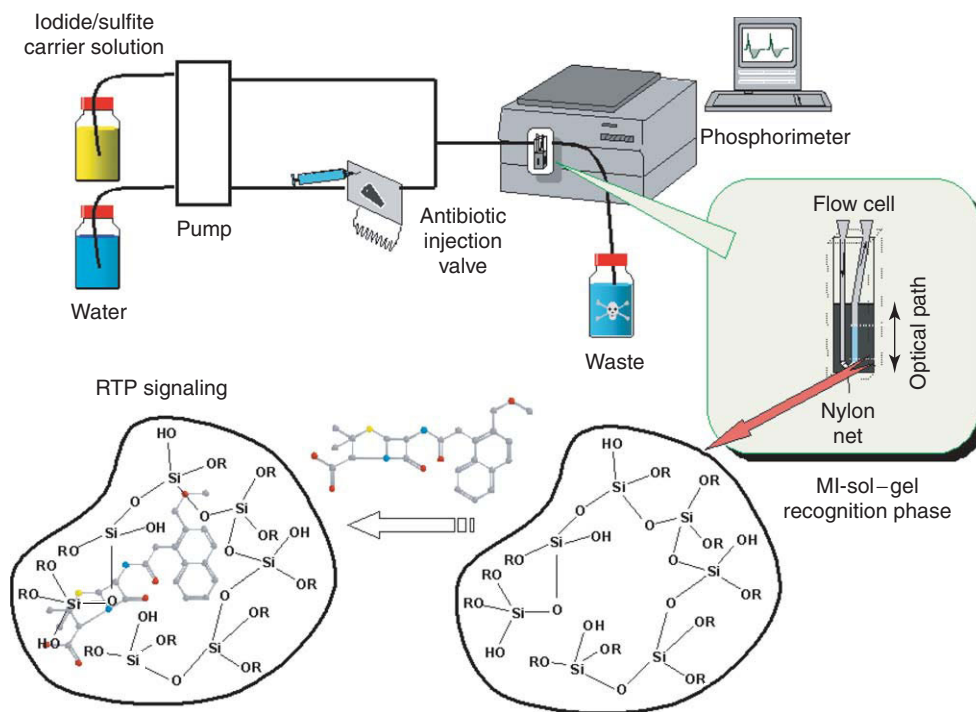
The need for real-time measurements and the development of fiber-optic technology gave the basis for designing optical sensors by monitoring changes in the RTP of phosphors immobilized on a solid support on the tip of an optical fiber. Limits of detection at the ng per ml level were estimated for pyrene, coronene, 2,3-benzofluorene, 1,2:3,4-dibenzanthracene, benzo[e]pyrene, and benzo[ghi]perylene, and the feasibility of monitoring PAHs in aqueous media was demonstrated.

In its conventional form, SS-RTP detection is not well suited for aqueous flow techniques (in contrast to homogeneous solution RTP) because of the strong quenching by moisture and oxygen and the intrinsically discontinuous character of the technique. Although the use of flow injection for continuous sample introduction in SS-RTP has been proposed, this methodology is not continuous in the flow because it involves sample nebulization on a paper strip followed by a drying step before the final measurement of the RTP signal.

In response to the demand for developing new SS-RTP methodologies for the continuous and *in situ* monitoring of different analytes, the coupling of SS-RTP to flow-injection systems was successfully

demonstrated. The approach involves the integration of the basic reaction/retention/detection process at the solid support packed into a flow cell. This detection principle in flow analysis is known as optosensing at active surfaces. To illustrate the concept, the molecular recognition of nafcillin, a  $\beta$ -lactamic antibiotic, is based on the transient immobilization/ RTP detection of nafcillin on an imprinted sol-gel material packed in a flow cell (Figure 4). Chemical deoxygenation by adding of sodium sulfite was employed. The immobilization process onto the imprinted material (insert in Figure 4) is produced by hydrogen bonding and electrostatic interactions between the template (nafcillin) and the selective recognition points inside the sol-gel. The phosphorescence properties of nafcillin are used for RTP signaling purposes. The analytical performance of this SS-RTP optosensing scheme compares favorably with other luminescent methodologies developed for nafcillin in micellar media.

There are very exciting alternatives for extending this transduction RTP approach to other analytical applications (particularly for luminescence optosensing). For example, Erythrosine B immobilized onto sol-gel materials proved to be an excellent RTP sensor for oxygen in nonaqueous media. Taking advantage of the fact that oxygen quenches this RTP emission very efficiently, a flow-cell-type sensing device can be used in SS-RTP optosensing of oxygen



**Figure 4** Flow injection setup for RTP optosensing of nafcillin using a molecularly imprinted sol-gel material as molecular recognition element.

in organic solvents using a simple flow setup as well as in gas mixtures (using a special flow manifold). These approaches hold promise for novel biosensing in oxygen-mediated biochemical reactions both in organic solvents (e.g., cholesterol determination) and at the solid–gas interface (e.g., ethanol or methanol vapor determinations). The applicability and convenience of use (easy automation) of the SS-RTP optosensing approach should encourage active research in this field in the future for the development of new (bio)chemical sensors. Analytical data for RTP oxygen determination in different scenarios are shown in Table 2.

Another application of particular note involving SS-RTP is the technology named phosphorescence barometry, which provides an inexpensive method for continuous pressure mapping of aerodynamic and automotive surfaces. The technology makes use of a luminescent paint (pressure-sensitive paint) that consists of an oxygen permeable polymer (e.g., silicone resins) in which a phosphorescent compound, a platinum porphyrin derivative, is dispersed. A temperature-sensitive paint would be similar except that an oxygen impermeable support should be used (e.g., an epoxy resin).

In this approach, pressure sensitivity is based on the oxygen quenching of the RTP of the dispersed platinum porphyrin complex. In practice, the surface

to be studied is coated with the paint and illuminated with light of the appropriate energy to excite the coating-entrapped probe molecules. The luminescence intensity is found to be proportional to the pressure at the surface as air contains a constant fraction of oxygen, independent of the total pressure. RTP images are captured with a charge-coupled device (CCD) camera with computer interface (Figure 5). Calibration is accomplished by obtaining a reference image (wind tunnel fans off,  $I_{\text{ref}}$ ) for which the pressure distribution is constant and equal to atmospheric conditions. In the run conditions (wind-on,  $I$ ), the pressure distribution is unknown:

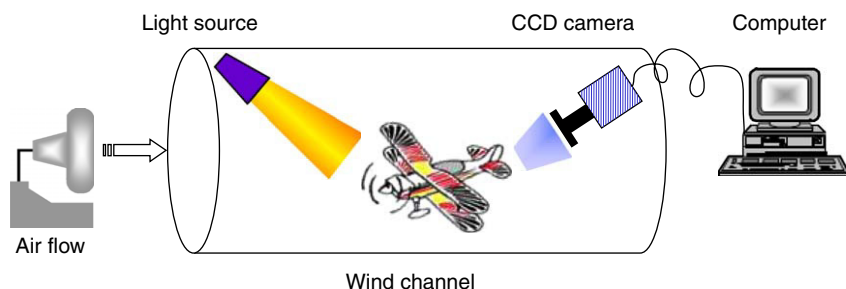
$$\frac{I_{\text{ref}}}{I} = A + B \frac{P}{P_{\text{ref}}} = A + B \frac{\tau_{\text{ref}}}{\tau}$$

The coefficients  $A$  and  $B$ , which are temperature dependent, are experimentally determined. Calibration of the intensity ratio ( $I_{\text{ref}}/I$ ) or lifetime ( $\tau$ ) can be correlated with the output of the CCD. The entire test object can be sampled simultaneously and the output of the array can be visually represented as an image with unequaled spatial resolution, thus providing a convenient tool for the generation of a spatially continuous pressure map. Compared to conventional pressure tap methods, RTP barometry does not require drilling holes in the studied surface,

**Table 2** Analytical figures for RTP oxygen sensing in different environments

RTP sensing phase	Dye immobilization	Medium	DL ( $\mu\text{g ml}^{-1}$ )	RSD%
EB-Sol-gel	Doped	Gaseous	6.0	0.06
		Aqueous	0.004	0.3
		Organic (hexane/chloroform)	0.49	1.4
EB-Amberlite XAD2	Adsorbed	Gaseous	5.0	0.20
		Aqueous	n.s.	n.s.
		Organic (hexane/chloroform)	0.65	6.3
EB-SCN-Bondesil-NH <sub>2</sub>	Covalent bonded	Gaseous	0.6	0.25
		Aqueous	n.s.	n.s.
		Organic (hexane/chloroform)	1.2	1.7

EB, erythrosine B; EB-SCN, erythrosine B isothiocyanate; Bondesil-NH<sub>2</sub>, commercial chromatographic support; n.s., no RTP signal observed.



**Figure 5** Oversimplified arrangement for phosphorescence barometry using a model aircraft in a wind tunnel.



provides a much faster response, and allows to map the pressure over the entire surface. This technique is currently being applied to the challenging problem of insect flight.

## Room-Temperature Phosphorescence in Microcrystalline/Colloidal Media

A microcrystalline or a colloidal-like suspension generated by syringe injection of a dilute organic solution of a PAH into pure water fulfils all the criteria for colloidal dispersion, except that the particles are larger. In this crystalline state, molecular diffusion is negligible and intermolecular photophysical events depend upon the crystal-formation step. A major spectral feature observed for many colloidal PAHs is intense RTP. This phenomenon is quite similar to SS-RTP. Generally, all colloidal RTP spectra exhibit large blue shifts compared to cryogenic temperature conditions, while the spectral resolution is much better than in MS-RTP spectra and only slightly worse than in low-temperature phosphorescence. It should be noted that colloidal RTP is completely insensitive to oxygen quenching. These features are observed in RTP spectra of fluorene, naphthalene, dibenzofuran, biphenyl, and others PAHs. However, the analytical potential of this technique is rather limited.

*See also:* **Bioluminescence.** **Chemiluminescence:** Overview. **Flow Injection Analysis:** Principles; Instrumentation. **Fluorescence:** Overview; Instrumentation. **Liquid Chromatography:** Overview; Principles. **Luminescence:** Overview; Solid Phase. **Phosphorescence:** Principles and Instrumentation. **Polycyclic Aromatic Hydrocarbons:** Environmental Applications.

## Further Reading

- Bruzzzone L, Badía R, and Díaz-García ME (2000) Room-temperature phosphorescent complexes with macromolecular assemblies and their (bio)chemical applications. *Critical Reviews in Analytical Chemistry* 30(2–3): 163–178.
- Cline Love LJ, Skrilec M, and Habarta JG (1980) Analysis by micelle stabilized room temperature phosphorescence in solution. *Analytical Chemistry* 52: 754–759.
- Díaz-García ME and Sanz-Medel A (1986) Facile chemical deoxygenation of micellar solutions for room temperature phosphorescence. *Analytical Chemistry* 58: 1436–1440.
- Donkerbroek JJ, Veltkamp AC, Gooijer C, Velthorst NH, and Frei RW (1983) Quenched room temperature phosphorescence detection for flow injection analysis and liquid chromatography. *Analytical Chemistry* 55: 1886–1893.
- Hurtubise RJ (1990) *Phosphorimetry: Theory, Instrumentation and Applications*. New York: Wiley-VCH.
- Kuijt J, Ariese F, Brinkman UATH, and Gooijer C (2003) Room temperature phosphorescence in the liquid state as a tool in analytical chemistry. *Analytica Chimica Acta* 488: 135–171.
- Li L-D, Chen X-K, Mou L, Long W-Q, and Tong A-J (2000) Non-protected fluid room temperature phosphorescence properties of  $\alpha$ -naphthylloxyacetic acid and  $\beta$ -naphthylloxyacetic acid. *Analytica Chimica Acta* 424: 177–183.
- Oglesby DM, Puram CK, and Upchurch BT (1995) Optimization of measurements with pressure sensitive paints. *NASA Technological Memorandum* 4695 (available electronically at <http://techreports.larc.nasa.gov/ltrs/ltrs.html>).
- Vo-Dinh T (1984) *Room Temperature Phosphorimetry for Chemical Analysis*. New York: Wiley.

# PHOSPHORUS

**I D McKelvie and A Lyddy-Meaney**, Monash University, Churchill, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Phosphorus in the hydrosphere may originate from natural diffuse sources such as the weathering of phosphate minerals, the decay of algae, plants, runoff from grazing and agricultural land, or it may be derived from anthropogenic point sources such as sewage and industrial effluent discharges. Phosphorus

plays a critical role in the process of eutrophication, because in many aquatic systems it is the nutrient that limits the growth of phytoplankton.

Phosphorus occurs in a variety of physically and functionally different inorganic and organic forms in aquatic systems (**Figure 1**). An understanding of these forms is important because it is the speciation that controls their physical and chemical behavior and their biological availability.

The simplest definition of phosphorus species involves separation of the dissolved and particulate components of a sample by filtration (**Figure 1**). The dissolved component is operationally defined by the filter pore size; for this reason, the term 'filterable' is



provides a much faster response, and allows to map the pressure over the entire surface. This technique is currently being applied to the challenging problem of insect flight.

## Room-Temperature Phosphorescence in Microcrystalline/Colloidal Media

A microcrystalline or a colloidal-like suspension generated by syringe injection of a dilute organic solution of a PAH into pure water fulfils all the criteria for colloidal dispersion, except that the particles are larger. In this crystalline state, molecular diffusion is negligible and intermolecular photophysical events depend upon the crystal-formation step. A major spectral feature observed for many colloidal PAHs is intense RTP. This phenomenon is quite similar to SS-RTP. Generally, all colloidal RTP spectra exhibit large blue shifts compared to cryogenic temperature conditions, while the spectral resolution is much better than in MS-RTP spectra and only slightly worse than in low-temperature phosphorescence. It should be noted that colloidal RTP is completely insensitive to oxygen quenching. These features are observed in RTP spectra of fluorene, naphthalene, dibenzofuran, biphenyl, and others PAHs. However, the analytical potential of this technique is rather limited.

See also: **Bioluminescence**. **Chemiluminescence**: Overview. **Flow Injection Analysis**: Principles; Instrumentation. **Fluorescence**: Overview; Instrumentation. **Liquid Chromatography**: Overview; Principles. **Luminescence**: Overview; Solid Phase. **Phosphorescence**: Principles and Instrumentation. **Polycyclic Aromatic Hydrocarbons**: Environmental Applications.

## Further Reading

- Bruzzzone L, Badía R, and Díaz-García ME (2000) Room-temperature phosphorescent complexes with macromolecular assemblies and their (bio)chemical applications. *Critical Reviews in Analytical Chemistry* 30(2–3): 163–178.
- Cline Love LJ, Skrilec M, and Habarta JG (1980) Analysis by micelle stabilized room temperature phosphorescence in solution. *Analytical Chemistry* 52: 754–759.
- Díaz-García ME and Sanz-Medel A (1986) Facile chemical deoxygenation of micellar solutions for room temperature phosphorescence. *Analytical Chemistry* 58: 1436–1440.
- Donkerbroek JJ, Veltkamp AC, Gooijer C, Velthorst NH, and Frei RW (1983) Quenched room temperature phosphorescence detection for flow injection analysis and liquid chromatography. *Analytical Chemistry* 55: 1886–1893.
- Hurtubise RJ (1990) *Phosphorimetry: Theory, Instrumentation and Applications*. New York: Wiley-VCH.
- Kuijt J, Ariese F, Brinkman UATH, and Gooijer C (2003) Room temperature phosphorescence in the liquid state as a tool in analytical chemistry. *Analytica Chimica Acta* 488: 135–171.
- Li L-D, Chen X-K, Mou L, Long W-Q, and Tong A-J (2000) Non-protected fluid room temperature phosphorescence properties of  $\alpha$ -naphthylloxyacetic acid and  $\beta$ -naphthylloxyacetic acid. *Analytica Chimica Acta* 424: 177–183.
- Oglesby DM, Puram CK, and Upchurch BT (1995) Optimization of measurements with pressure sensitive paints. *NASA Technological Memorandum* 4695 (available electronically at <http://techreports.larc.nasa.gov/ltrs/ltrs.html>).
- Vo-Dinh T (1984) *Room Temperature Phosphorimetry for Chemical Analysis*. New York: Wiley.

# PHOSPHORUS

I D McKelvie and A Lyddy-Meaney, Monash University, Churchill, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Phosphorus in the hydrosphere may originate from natural diffuse sources such as the weathering of phosphate minerals, the decay of algae, plants, runoff from grazing and agricultural land, or it may be derived from anthropogenic point sources such as sewage and industrial effluent discharges. Phosphorus

plays a critical role in the process of eutrophication, because in many aquatic systems it is the nutrient that limits the growth of phytoplankton.

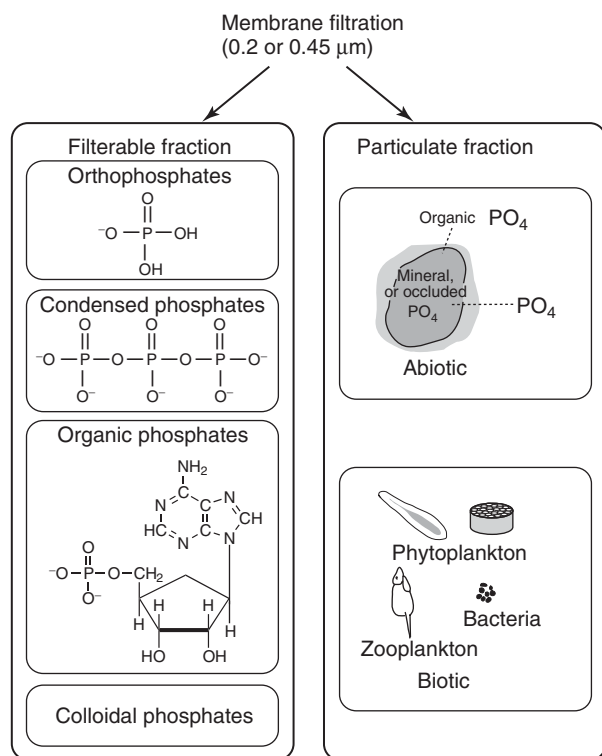
Phosphorus occurs in a variety of physically and functionally different inorganic and organic forms in aquatic systems (**Figure 1**). An understanding of these forms is important because it is the speciation that controls their physical and chemical behavior and their biological availability.

The simplest definition of phosphorus species involves separation of the dissolved and particulate components of a sample by filtration (**Figure 1**). The dissolved component is operationally defined by the filter pore size; for this reason, the term 'filterable' is

preferred to either 'dissolved' or 'soluble', both of which are used extensively and interchangeably in the literature. The filterable total phosphorus (FTP) component is comprised of the filterable reactive (FRP), condensed (FCP), and organic (FOP), fractions. Of these, the FRP consists of inorganic orthophosphates ( $\text{H}_2\text{PO}_4^-$ ,  $\text{HPO}_4^{2-}$ ,  $\text{PO}_4^{3-}$ ) and some labile organic and colloidal phosphates that will react with acidic molybdate to form the phosphomolybdate complex that is the basis for most phosphorus analysis. FCP consists of inorganic polyphosphates, metaphosphates, and branched ring structures, and the

FOP fraction is composed of compounds such as nucleic acids, phospholipids, inositol phosphates, phosphoamides, phosphoproteins, sugar phosphates, aminophosphonic acids, phosphorus-containing pesticides, and organic condensed phosphates.

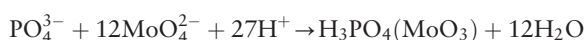
The particulate phosphorus (PP) fraction refers to the fraction retained by the filter (usually 0.45 or 0.2  $\mu\text{m}$  pore size membrane). PP may consist of biological material (animal, plant, bacterial), weathering products (primary and secondary minerals), precipitates (authigenic minerals), organic and inorganic coprecipitates and aggregates, in addition to phosphorus associated with aggregates through metal binding or adsorbed to the surface of clay and mineral particles. Determination of FTP, FOP, FCP, and PP in natural waters requires a preliminary digestion step to convert the various phosphorus species to the detectable orthophosphate form (Table 1).



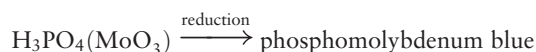
**Figure 1** Representation of phosphorus species in aquatic systems operationally defined by filtration. (McKelvie ID, Peat D, and Worsfold PJ (1995) Techniques for the quantification and speciation of phosphorus in natural waters. *Analytical Proceedings* 32: 437–445; reproduced by permission of The Royal Society of Chemistry.)

## Spectrophotometric Detection

The analysis of phosphorus in waters has historically been based on the photometric measurement of 12-phosphomolybdate or the phosphomolybdenum blue species, which are produced when phosphomolybdate is reduced. The majority of manual and automated methods of phosphate determination are based on the spectrophotometric determination of phosphorus as phosphomolybdenum blue, i.e.,



followed by



Phosphorus species that are determined in this manner are referred to as molybdate reactive, or simply reactive, and much of the nomenclature of phosphorus speciation derives from this origin.

There have been numerous variations on this method, usually involving different reductants (tin(II) chloride, ascorbic acid, 1-amino-2-naphthol-4-sulfonic

**Table 1** Specification scheme for phosphorus (P) based on determination of molybdate reactive (R), phosphorus in total (T), filterable (F), and particulate (P) forms; O and C designate organic and condensed forms, respectively

	Total sample (T)	Filterable fraction (F)	Particulate fraction (P)
Total P determination: digestion + colorimetry	TP (TOP+TCP+TRP)	– FTP (FOP+FCP+FRP)	= PTP (POP+PCP+PRP)
Condensed P determination: hydrolysis + colorimetry	TCP+TRP	– FCP+FRP	= PCP+PRP
Organic P determination (by subtraction)	TOP	– FOP	= POP
Colorimetry (reaction with molybdate)	TRP	– FRP	= PRP

acid, sodium sulfite, hydrazine sulfate, or combinations thereof) or different acid concentrations, in attempts to improve the selectivity and stability of the chromophore produced. The most commonly used methods for both manual and automated analyses are based on the ascorbic acid reduction with potassium antimonyl tartrate as a catalyst, which is used because it is more tolerant of temperature and salinity variations, and is less susceptible to interference from silica.

## Speciation of Phosphorus Based on Digestion and Molybdate Reactivity

Major differences in size (filterable, particulate) and chemical reactivity (condensed, organic) of phosphorus forms in samples can be used as the basis for speciation, as shown in Table 1. Of these fractions, total phosphorus (TP) and FRP are perhaps the most commonly measured, although it is arguable that the understanding of the aquatic phosphorus cycle is somewhat lopsided because of that bias. For example, wastewater discharge licenses often specify a maximum permissible concentration of TP, and provide an indication of the maximum potentially bioavailable phosphorus discharged. However, FRP, comprising mostly orthophosphate, is a measure of the amount of most readily bioavailable phosphorus.

The determination of TP, condensed and organic phosphorus all require predigestion and/or hydrolysis of the water sample prior to detection of the orthophosphate produced. Complete conversion of particulate and filterable components requires conditions that are conducive to dissolution of phosphate mineral phases, hydrolysis of condensed phosphates, and oxidation of organic phosphorus species. Numerous methods have been proposed, but whichever procedure is selected for the determination of TP or FTP, the digestion efficiency should be assessed using appropriate certified reference materials and a range of organic or condensed phosphorus model compounds. The latter should include compounds such as tripolyphosphate, inositol hexakisphosphate, and 2-aminoethylphosphonic acid that contain P–O–P, C–O–P, or C–P chemical bonds, respectively, and that are known to be more refractory.

## Condensed Phosphorus Hydrolysis

Determination of the condensed phosphorus component requires hydrolytic conditions without oxidation, and is typically achieved by boiling samples with 0.05 M sulfuric acid for 90 min or autoclaving for 30 min, prior to determination as molybdate reactive phosphorus.

## TP Digestion

### Thermal Digestion Methods

Wet chemical digestions with either alkaline or acid peroxydisulfate are the most common digestion methods for determining TP. If the alkaline peroxydisulfate digestion is used, sufficient time must be allowed for complete breakdown of peroxydisulfate to sulfuric acid if the condensed component is to be hydrolyzed. Similarly, if nitric–sulfuric acid or nitric–sulfuric–perchloric acid is used, it must be established that the conditions used are sufficiently oxidizing to digest the most refractory organic compounds, like aminoethyl phosphonates. Digestion may be performed at ambient pressure, or, for greater speed, at elevated pressure and temperature using a pressure cooker or autoclave. Recovery tests for autoclave methods involving peroxydisulfate and nitric–sulfuric acid digestion reagents for determination of TP in waters with high suspended load have shown that recovery of TP at  $>100 \mu\text{g P l}^{-1}$  is incomplete using the peroxydisulfate reagent. Dilution to  $\leq 100 \mu\text{g P l}^{-1}$  is recommended to overcome this problem.

A number of workers have also reported the use of microwave ovens for thermal digestion of samples for TP analysis. Digestion times can be reduced significantly by the use of microwave heating, which can be performed both in batch and online flow injection modes.

As alternatives to the wet chemical methods described above, high-temperature combustion with magnesium sulfate followed by acid leaching or high-temperature fusion with magnesium nitrate have been proposed. The latter method has been shown to decompose phosphonates that are quite refractory.

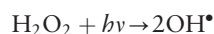
### Photochemical Methods

Ultraviolet (UV) photooxidation may be employed to convert organic phosphorus to phosphate prior to detection, and can be performed either in batch mode using a high-intensity, ventilated UV source and a quartz reactor vessel, or in continuous-flow mode using either quartz or polytetrafluoroethylene (PTFE) photoreactors. Because batch UV radiation systems usually involve the use of high wattage UV lamps (e.g., 1000 W) and extended irradiation times, condensed phosphates are hydrolyzed due to the elevated temperature and gradual acidification of the sample as peroxydisulfate degrades to form sulfuric acid. Thus, UV photooxidation alone is insufficient to convert condensed phosphates to orthophosphate, and the use of UV photooxidation with alkaline peroxydisulfate may provide a basis for discrimination

between the organic and condensed phosphorus fractions.

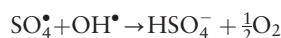
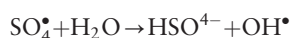
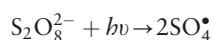
Photooxidation of organic phosphorus may be performed by UV irradiation on the untreated sample, and relying on the dissolved oxygen present in the sample to provide an adequate source of oxygen or hydroxyl radicals. However, it is more common that hydrogen peroxide, peroxydisulfate, ozone, or other oxidizing agents are added to enhance the completeness of the oxidation process.

When  $\text{H}_2\text{O}_2$  is exposed to UV light, it forms hydroxyl radicals:



This radical is among the strongest oxidizing agents found in aqueous systems and initiates a series of radical chain reactions with organic substances, resulting in mineralization of the sample to bicarbonate and orthophosphate.

Photooxidation using peroxydisulfate also produces hydroxyl radicals and oxygen by the following route:



Titanium dioxide-mediated oxidation of organic phosphorus can also be achieved using long-wavelength UV. Excitation of an electron from the valence band (v) into the conduction band (c) creates an electron-hole pair, which may then react with, for example, oxygen adsorbed to the  $\text{TiO}_2$  surface to produce radicals such as  $\text{O}_2^-$  and  $\text{OH}^\bullet$ .

In order to determine the TP concentration in water, the digestion process must involve both oxidative and hydrolytic processes in order to hydrolyze P–O–P linkages (e.g., polyphosphates) and oxidize phosphoesters and C–P compounds to inorganic phosphate. For example, in an online TP digestion system, which involves both thermal digestion and UV photooxidation, it is necessary to use a mixture of sulfuric acid and peroxydisulfate in order to obtain high recoveries of both organic and condensed phosphorus.

## Bioavailable Phosphorus

Analysis of phosphorus species in natural waters is driven largely by a need to assess the likely potential for eutrophication, and a number of phosphorus

analysis parameters have been employed as estimators of bioavailable phosphorus (BAP).

### Algal Bioassay

BAP has traditionally been determined using algal bioassays (e.g., the algal assay bottle test). However, these take 7–21 days to perform, are labor intensive because they require daily measurement of the growth rate, are susceptible to large statistical variability, and are relatively insensitive. Furthermore, caution should be exercised in the interpretation of the BAP concentration obtained, because native algae may have a much wider range of phosphorus substrate activity than the axenic algal monocultures that are used for this purpose. Consequently, there has been a great deal of research effort into replacing algal bioassay by some chemical parameter that will provide a rapid and more convenient means of determining the BAP concentration.

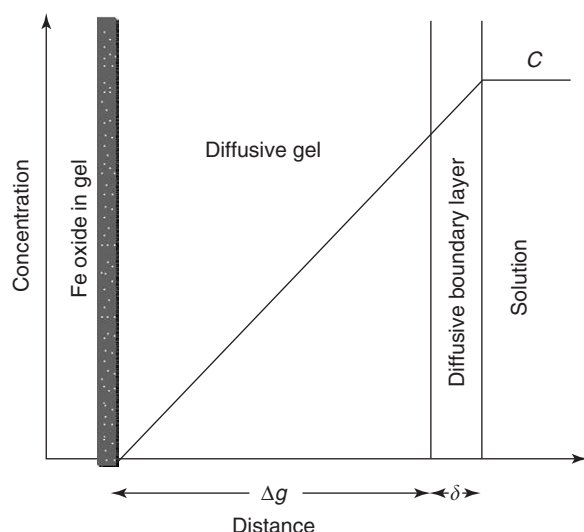
### TP and FRP

Because of the time and labor involved in algal assays, FRP (0.45 or 0.2  $\mu\text{m}$ ) has often been used as a surrogate for readily BAP, and TP (which includes the particulate, condensed, and organic phosphorus components) as a measure of potentially BAP. However, a number of studies have shown that neither FRP nor TP correlate well with algal assay-measured BAP. This may be because FRP tends to overestimate the true orthophosphate due to hydrolysis of labile organic and condensed species. Use of ultrafiltration through a low molecular mass filter prior to measurement of reactive phosphorus has been proposed as better estimator of BAP than other available chemical methods.

### Iron Strip and DGT Methods

A proposed alternative for estimating the amount of BAP in waters involves the use of iron strip adsorption techniques used in soil science to measure plant-available phosphorus. This approach involves the equilibration of dissolved and particulate-bound phosphorus with an iron-oxide-impregnated paper strip for a period of hours to days, and this is followed by acid leaching and spectrophotometric analysis for FRP. These exchangeable phosphorus tests have been shown to be well correlated with BAP determined by algal bioassay in waters. However, co-adsorbed organic phosphorus or adsorbed particulate material may also be hydrolyzed at the iron-oxide surface and during the acid leaching and colorimetry stages of the process, leading to an overestimation of the amount of BAP.





**Figure 2** Representation of a steady-state concentration gradient in a DGT device.  $\Delta g$  is the thickness of the ion-permeable gel membrane,  $\delta$  is the thickness of the diffusive boundary layer, and  $C$  is the concentration of reactive phosphorus species in the bulk of solution. (Adapted from Zhang H, Davison W, Gadi R, and Kobayashi T (1998) *In-situ* measurement of dissolved phosphorus in natural waters using DGT. *Analytica Chimica Acta* 370: 29–38, with permission from Elsevier.)

An alternative approach is the use of diffusive gradients in thin films (DGT). Phosphate diffuses through a thin polyacrylamide gel and is bound in a second gel layer containing iron-oxide. *In situ* deployment is followed by sectioning, leaching, and analysis for molybdate reactive phosphorus. From a knowledge of the diffusion coefficient of phosphate in the gel and the measured phosphate concentration, the bulk solution phosphate concentration can be calculated from Fick's first law of diffusion (Figure 2). This enables a time-integrated measurement of FRP concentration, and can also be used to determine sediment–water fluxes of phosphate.

## Enzymatic Methods

Algae and bacteria are known to release extracellular alkaline phosphatase, which facilitates utilization of otherwise unavailable dissolved and particulate organic phosphorus species. This behavior has been used as the basis for a speciation technique aimed at estimating BAP in the dissolved fraction. Alkaline phosphatase has been used in soluble and immobilized forms to hydrolyze FOP in natural waters, and the resultant orthophosphate is detected as FRP. Use of the soluble enzyme technique has essentially been discarded because of product inhibition of the alkaline phosphatase by orthophosphate already present

in the water. However, if an immobilized alkaline phosphatase reactor is used in a flow injection configuration, the problem is obviated because the hydrolysis product is transported away from the active sites of the enzyme. The technique has been successfully applied to the determination of alkaline phosphatase hydrolyzable phosphorus (APHP) in a range of natural and wastewaters. The enzyme phytase has also been used in a similar manner to hydrolyze organic phosphates in waters. The use of these enzyme techniques thus provides additional information as to the functionality of the phosphorus species present.

## Selective Extraction Techniques for Phosphorus in Sediments and Soils

In soils and sediments, chemical fractionation or sequential extraction methods offer another approach to speciation of the forms of phosphorus. Traditionally, there has been more emphasis on bioavailable or plant available inorganic forms, e.g., Olsen's extraction used for soils.

The SEDEX procedure has emerged as the benchmark extraction scheme for sediments. It uses only magnesium chloride, acetate buffer, and citrate/dithionite–bicarbonate reagents at pH values between 4 and 8 to leach sediments of the inorganic associated phosphorus fractions before ashing at 550°C and a final extraction with 1 M HCl to determine the so-called residual organic phosphorus.

However, extraction of organic phosphorus from sediments and soils should be carried out in a manner that, as far as possible, avoids hydrolysis or oxidation to orthophosphate, and this is often incompatible with the procedures used for inorganic phosphorus extraction. Some authors have suggested that strongly acidic and basic conditions are intrinsically unsuitable for extraction because of the likely hydrolytic breakdown that they cause, and have suggested instead that chelating extractants such as NTA or ethylenediamine tetraacetic acid (EDTA) at near-neutral pH should be used as an alternative (Table 2).

## Determination of Specific Phosphorus Compounds

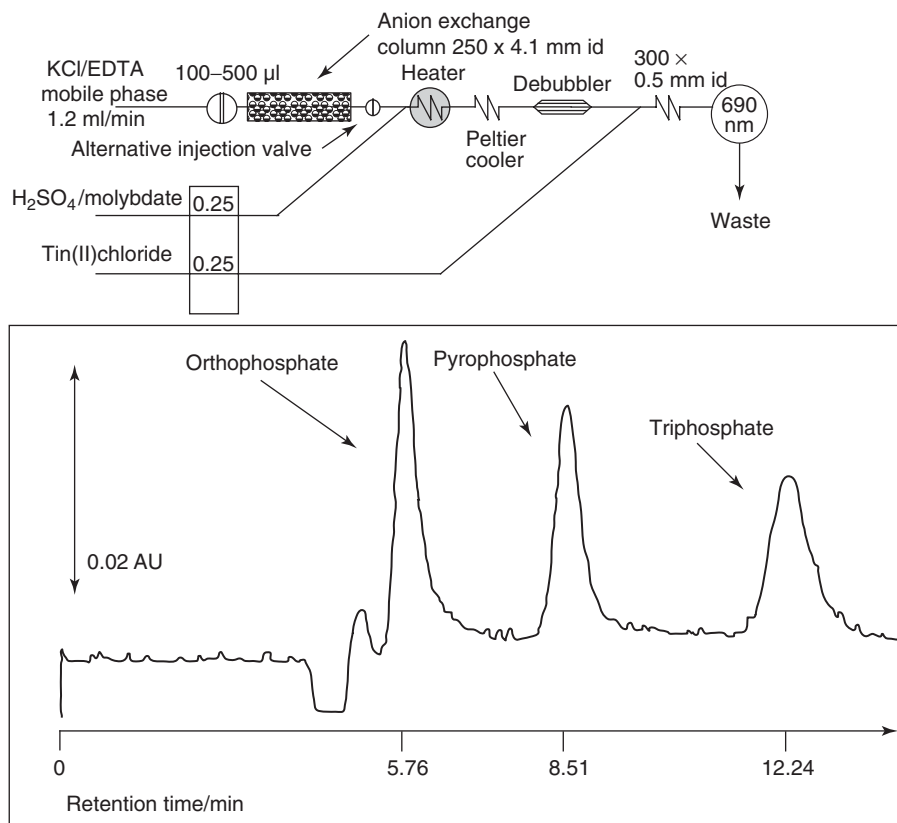
### Chromatographic Methods

**High-performance liquid and ion chromatography** Phosphate is commonly determined by ion chromatography, both in suppressed and unsuppressed modes. The sensitivity of the conductivity detection techniques commonly used may be inadequate for direct application to the analysis of pristine waters,



**Table 2** Sequential chemical extraction for isolation of various phosphate fractions involving use of chelating agents at near-neutral pH

Fraction designation	Abbreviation	Extractant used
Iron-bound phosphate	$\text{Fe}(\text{OOH}) \approx \text{P}$	0.02 M Ca-NTA/dithionite, pH 7.8–8.0
Calcium-bound phosphate	$\text{CaCO}_3 \approx \text{P}$	0.05 M Na-EDTA, pH ~8.0
Acid-soluble organic phosphate	ASOP	0.5 M HCl or 0.025 M $\text{H}_2\text{SO}_4$ (30 min)
NaOH-extractible phosphate	$\text{NaOH}_{\text{extr}}\text{-P}$	2.0 M NaOH (90°C, 30 min)

**Figure 3** (A) Ion-exchange chromatography–FIA system for separation of polyphosphates. (B) Optimized separation of orthophosphate ( $50 \mu\text{g PI}^{-1}$ ), diphosphate ( $50 \mu\text{g PI}^{-1}$ ), and triphosphate ( $50 \mu\text{g PI}^{-1}$ ) with an injection volume of  $500 \mu\text{l}$ . (Halliwell DJ, McKelvie ID, Hart BT, and Dunhill RH (1996) Separation and detection of condensed phosphates in waste waters by ion chromatography coupled with flow injection. *Analyst* (Cambridge, UK) 121: 1089–1093; reproduced by permission of The Royal Society of Chemistry.)

and some form of preconcentration is required. For analysis of seawaters, extensive dilution or chloride removal is necessary. The advantage of this technique is that it is true orthophosphate rather than FRP that is determined.

Anion exchange chromatography and ion exchange chromatography have been used extensively for the separation and quantitation of condensed phosphates. Because phosphate is a poor UV chromophore, direct UV detection cannot be used, and fraction collection for subsequent acid hydrolysis and detection as molybdate reactive phosphorus (MRP)

has been the norm. High-performance liquid chromatography (HPLC)-ion chromatography with post-separation hydrolysis and detection using flow injection analysis has advantages of speed of analysis, sensitivity, and selectivity. A schematic diagram of this instrument setup and typical separations obtained from a hyphenated system of this type is shown in **Figure 3**.

Interest in characterization of organic phosphorus present in natural waters has also prompted the development of ion chromatographic separation systems for compounds such as inositol phosphates.

Online UV photooxidation has been utilized for oxidation and subsequent detection of these organic phosphate species. While ion exchange HPLC is used extensively for separation of FOP species, a reversed-phase partition HPLC has a definite role in the separation of even quite charged or polar organic phosphorus species. Phospholipids are commonly separated and quantified using either reversed- or normal-phase HPLC. Reversed-phase HPLC has also been used in the study of inositol phosphates. Use of ion-pair reversed-phase HPLC is also advantageous for the separation of highly charged species and may avoid some of the adsorption problems that are encountered using gel filtration or ion exchange chromatography, especially of highly charged species. Reversed-phase ion-pair chromatographic separations have been described for inositol phosphates, and some common nucleotides and sugar phosphate phytate.

**Gel filtration/size exclusion chromatography** Separation using gel filtration gained popularity as a means of separating high and low molecular mass phosphorus fractions, and as a means of estimating BAP concentrations, i.e., the reactive low molecular mass phosphorus. Most involve the use of large separation columns, long elution times, and the use of fraction collection and off-line digestion/digestion to measure total or reactive phosphorus, and as such are unsuitable for routine monitoring applications. However, an on-line postcolumn flow injection detection system for detection of organic phosphorus species has also been described. Potential difficulties associated with the use of gel filtration include early elution due to anion exclusion and late elution due to hydrophobic interactions, and specific adsorption.

### Capillary Electrophoresis

Capillary electrophoresis (CE) techniques have been used for selective orthophosphate analysis. CE separations of anions in waters are much faster than ion chromatography, but detection of phosphate by UV absorbance is very insensitive. This may be overcome by on-capillary preconcentration using isotachopheresis, which enables  $\text{sub-}\mu\text{g l}^{-1}$  detection limits to be achieved in high ionic strength matrices.

### Gas Chromatography

Gas chromatography (GC) with either flame photometric or mass spectrometric detection is commonly used for determination of organophosphate pesticide. In some cases, GC may be used for separation of organic phosphorus species if they or their degradation products are derivatized, e.g., inositol

phosphates have been characterized by GC separation of their acetylated dephosphorylation products.

### Other Spectroscopic Methods

Nuclear magnetic resonance, especially  $^{31}\text{P}$  NMR, offers a powerful means of characterizing and identifying individual phosphorus species in soils and sediments. In the case of waters, there is usually the need for preconcentration, and interferences may arise because of the corresponding increase in concentration of paramagnetic substances. However, if these difficulties are resolved,  $^{31}\text{P}$  NMR provides a qualitative means of detecting the presence of functionally different phosphorus groups, such as phosphonates, orthophosphates, orthophosphate mono- and diesters, and pyrophosphates, within a sample.

GC–Mass spectrometry combines high-efficiency separation with sensitive and highly selective detection, and is a powerful tool for characterization of organic phosphorus.

*See also: Nuclear Magnetic Resonance Spectroscopy-Applicable Elements: Phosphorus-31.*

### Further Reading

- APHA-AWWA-WEF (1998) *Standard Methods for the Examination of Water and Wastewater*. Washington, DC: American Public Health Association.
- Broberg O and Persson G (1988) Particulate and dissolved phosphorus forms in freshwater: composition and analysis. *Hydrobiologia* 170: 61–90.
- De Groot CJ and Golterman HL (1993) On the presence of organic phosphate in some Camargue sediments: evidence for the importance of phytate. *Hydrobiologia* 252: 117–126.
- Kérroul R and Aminot A (1996) Model compounds for the determination of organic and total phosphorus dissolved in natural waters. *Analytica Chimica Acta* 318: 385–390.
- McKelvie ID, Peat D, and Worsfold PJ (1995) Techniques for the quantification and speciation of phosphorus in natural waters. *Analytical Proceedings* 32: 437–445.
- Robards K, McKelvie ID, Benson RL, *et al.* (1994) Determination of carbon, phosphorus, nitrogen and silicon species in waters. *Analytica Chimica Acta* 287: 147–190.
- Ruttenberg KC (1992) Development of a sequential extraction method for different forms of phosphorus in marine sediments. *Limnology and Oceanography* 37: 1460–1482.
- Sharpley AN (1993) An innovative approach to estimate bioavailable phosphorus in agricultural runoff using iron oxide-impregnated paper. *Journal of Environmental Quality* 22: 597–601.
- Zhang H, Davison W, Gadi R, and Kobayashi T (1998) *In-situ* measurement of dissolved phosphorus in natural waters using DGT. *Analytica Chimica Acta* 370: 29–38.

# PHOTOACOUSTIC SPECTROSCOPY

**R D Snook and L Mitchem**, UMIST, Manchester, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

In this article, the key principles of photoacoustic spectrometry (PAS) will be described together with a simple discussion of the instrumentation required to perform PAS. Some important applications will be discussed and future scope of the technique's application outlined. The introduction brings together some historical generalities of the technique followed by sections on its theoretical aspects, instrumentation, and analytical applications. The emphasis is on chemical analysis throughout.

Photoacoustic detection is one of a class of photothermal detection techniques that can be used to measure the optical absorption of a sample by monitoring the absorption of modulated ultraviolet (UV)/visible or infrared radiation and its subsequent conversion to heat by nonradiative processes to produce a periodic thermal signal. The signal is normally measured by the effect that the periodic heat flow has on the absorbing medium, i.e., the sample, or any medium in contact with the sample. For example, measurement of the change in refractive index that occurs on absorption of a modulated laser beam in a transparent or semitransparent medium is the basis of thermal lens detection, photothermal interferometry, and photothermal refraction. The use of these detection techniques and quantitation of absorption at different excitation wavelengths gives rise to photothermal spectrometry.

Most of these photothermal techniques, including PAS, have been exploited in the last 30 years for materials characterization and chemical analysis. However, the fundamental phenomenon of the photoacoustic effect was first observed by Alexander Graham Bell in 1880, who noted that when modulated sunlight was allowed to fall on absorbing materials in a transparent tube, sound was heard and that the more strongly absorbing materials produced louder sounds. This led to the construction of Bell's photo-phone. The principles by which sound was produced were later quantified and utilized by Rose-ncwaig and Gershov, and Kirkbright and Adams to provide a powerful new class of spectroscopic techniques originally known as optoacoustic spectrometry but now known as photoacoustic spectrometry or PAS.

In classical PAS, as demonstrated by Kirkbright and Adams, modulated continuum radiation from a high-power UV/visible continuum light source such as a high-pressure Xe or Hg lamp is passed through a monochromator to isolate particular wavelength bands. Monochromatic light thus selected then falls upon a gas-tight cell that contains the sample, and in the gas space above the sample, a sensitive microphone. The output of the microphone can be recorded as amplitude or connected to a phase-sensitive lock-in amplifier.

If the sample absorbs at the incident wavelength excitation takes place and subsequent relaxation of excited states via nonradiative mechanisms, such as collisional relaxation, generates heat. As the incident radiation is modulated the sample becomes periodically heated, which results in a periodic thermal wave propagating through the sample. When this periodic heat flow reaches the surface of the sample it propagates through the contacting gas as an acoustic wave, caused by the compression and rarefaction of the gas, which can be detected as sound by the microphone. It follows that the higher the absorption coefficient at the incident wavelength then the louder (higher amplitude) the signal. If the monochromator is a wavelength-scanning device then a complete optical absorption spectrum in the range of the spectrometer can be obtained simply by measuring the PAS signal at each wavelength. A great advantage of measuring the absorption spectrum in this way is that solid phases can be measured directly even if there are irregularities present that causes scattering at the surface. Thus, the analysis of powders and solids becomes feasible, which would not be when using conventional transmission spectrometry because of variable scattering.

Other information is also available when studying samples using PAS that derives from the processes of optical absorption and heat generation in the surface layers of the sample. For example, under certain conditions the depth from which a thermal signal can be detected in a sample depends upon the inverse of the modulation frequency. Thus, by changing the modulation frequency depth profiling can be achieved. Alternatively, depth information can be obtained by measuring the phase lag of the signal determined using lock-in detection. This dependency on modulation frequency is particularly important when using Fourier transform infrared (FTIR) spectrometers in conjunction with PAS cells to obtain IR spectra of solids because in FTIR each

wavelength is modulated in the interferometer output beam at a different frequency thus allowing the recovery of comprehensive depth information. This can, of course, also be a complicating factor if the depth information is not required. Alternatively, step scanning can be employed to impose a fixed modulation frequency on the interferometer output beam thus maintaining a constant thermal diffusion length. The principles of PAS with special reference to FTIR PAS are described in the following section.

## Theory and Principles

### Basic Principles of PAS

PAS involves the analysis of acoustic waves, generated as a result of the absorption of modulated light. The acoustic waves are generated in the following manner:

1. Excitation of a sample by UV/IR light modulated at a frequency,  $f$ , increases the sample's internal energy. Providing luminescence or photochemical degradation does not occur, the internal energy is converted into heat by nonradiative deexcitation of the excited states.
2. The heat propagates from the region of absorption back toward the sample surface in the form of a thermal wave.
3. At the sample–boundary gas interface the thermal wave heats the adjacent boundary gas, which expands and contracts periodically, creating acoustic waves modulated at the same frequency as the incident radiation. (The gas within the photoacoustic cell is assumed to behave like a perfect gas, where temperature variations induce equivalent pressure variations.)
4. The acoustic waves propagate through the cell to a microphone, where they are detected and the acoustic signal is converted into an electrical signal. Ideally, the PAS signal is directly proportional to the intensity of the incident beam.

The amplitude of the PAS signal at any wavelength is dependent on the ability of the sample to absorb at that wavelength; therefore, the microphone output is equivalent to the absorbance spectrum of the sample. Only photons absorbed by the sample produce a photoacoustic signal; therefore, unlike transmission and reflectance techniques PAS is unaffected by scattering and can be used to analyze highly scattering materials, e.g., fine powders, crystalline and biological samples. Although PAS is unaffected by scatter signal loss does occur. A fraction of the incident beam is lost due to reflection at the sample surface during sample excitation. This quantity is dependent upon the sample under analysis. Additionally, on

propagation of the thermal wave from the sample to the boundary gas a large fraction of the thermal wave is reflected back into the sample. This loss greatly affects the magnitude of the photoacoustic signal.

In addition to the generation of a photoacoustic signal, excitation of a sample with modulated light induces a number of additional responses, for example, IR emission is produced and a refractive index gradient is generated within and above the sample. Each response can be used to analyze the sample. The response used, however, is dependent upon the sample and the environment it is in.

### Rosencwaig Gersho Theory (RG Theory)

Rosencwaig and Gersho published their 'Thermal Piston' theory describing photoacoustic generation in solids in 1976. Prior to this publication the photoacoustic effect was principally used to quantitatively estimate gas concentrations within mixtures. The RG theory, the first to describe the photoacoustic effect mathematically, led to the rediscovery and rapid development of the photoacoustic technique. It was developed by solving one-dimensional, thermal diffusion equations defining heat propagation within a sample, its gaseous surroundings, and its backing material, following the absorption of light. By assuming the pressure at the sample surface to be dependent on the periodic temperature variation at the sample–boundary gas interface, and the gas to respond adiabatically to the pressure, Rosencwaig and Gersho solved the thermal diffusion equations, using appropriate experimental boundary conditions, and defined equations for the amplitude and phase of the induced photoacoustic signal. Many modifications to the RG theory followed, for example, to account for signal generation in multilayered samples and three-dimensional systems. These modifications have resulted in the high development of the photoacoustic technique.

### Depth-Profiling Capabilities

When analyzing a sample using PAS spectroscopy, the depth beneath the surface from which a photoacoustic signal originates is dependent upon the interaction between the sample's optical absorption length,  $\mu_i$ , and thermal diffusion length,  $\mu_s$ .

The incident beam decays exponentially as it propagates through the sample following excitation at its surface. The optical absorption length,  $\mu_i = 1/\beta$ , where  $\beta$  is the optical absorption coefficient of the sample, defines the maximum depth the beam penetrates the sample. Thermal waves also decay exponentially as they propagate from the region of absorbance where they are generated, to the

sample surface. The thermal diffusion length, which defines the maximum depth from which a photoacoustic signal can be detected before significant dampening makes it undetectable, is defined by eqn [1]:

$$\mu_s = (\alpha/\pi f)^{1/2} \quad [1]$$

where  $\alpha$  is the thermal diffusivity of the sample and  $f$  the modulation frequency of the incident light.

The magnitude of the photoacoustic signal is directly proportional to the heat generated within this depth. The thermal diffusion length tends to range from several to over a hundred micrometers.

The interaction between the thermal diffusion length, optical absorption length, and sample thickness ( $l$ ) is used to describe the photoacoustic properties of a sample. Thus, a sample is described as opaque if the optical absorption length is less than the sample thickness, and light cannot be transmitted through the sample. However, when  $\mu_l$  and  $\mu_s$  are greater than  $l$  the sample is described as optically and thermally thin, respectively, and when  $\mu_l$  and  $\mu_s$  are less than  $l$  the sample is described as optically and thermally thick, respectively.

The thermal diffusion length is dependent upon the modulation frequency of the incident radiation used during sample excitation (see eqn [1]). By changing the modulation frequency of the incident light the thermal diffusion length is varied, a property used to depth profile samples. The effect of modulation frequency on the depth from which a photoacoustic signal is detected, using a thermal diffusivity of  $0.12 \text{ cm}^2 \text{ s}^{-1}$ , is illustrated approximately in Figure 1. The depth-profiling capabilities of PAS have resulted in its frequent use in the structural determination of layered heterogeneous samples (transmission and reflection techniques do not provide

depth-related information, only the average over the thickness analyzed).

If PAS is to be used for its depth-profiling capabilities the thermal diffusion length must not exceed the optical absorption length of the sample. If  $\mu_l < \mu_s$  all incident radiation is absorbed within the thermal diffusion length, resulting in photoacoustic saturation of the signal. Photoacoustic saturation results in the truncation and broadening of spectral bands. Additionally, the sample acts as a blackbody making depth profiling impossible. For the photoacoustic spectrum to accurately represent the absorption spectrum the sample's optical absorption length must exceed the thermal diffusion length, a condition met by selecting the appropriate modulation frequency.

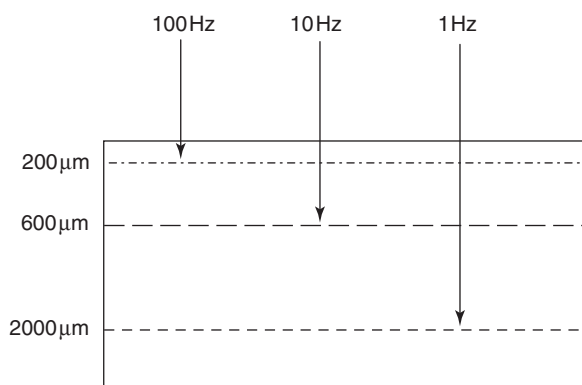
PAS does not require the incident beam to pass through the sample under analysis like conventional spectroscopic techniques; the depth probed is defined by the thermal diffusion length. By selecting an appropriate modulation frequency PAS can be used to analyze opaque solids. This is advantageous when compared to transmission techniques that require dilution before analysis can be performed.

## Instrumentation

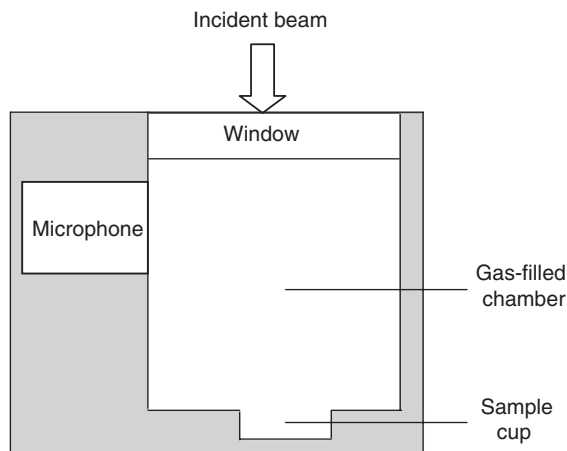
The components essential to perform PAS spectroscopy are: a source that determines the PAS spectral range, a spectrometer or interferometer for wavelength selection, a method of modulation of the incident beam, and a photoacoustic cell.

### The Photoacoustic Cell

PAS is used to analyze solids, liquids, and gases with little or no sample preparation required. A simplified photoacoustic cell, used to analyze solids, is illustrated in Figure 2. This cell can also be used to analyze



**Figure 1** Effect of modulation frequency on thermal diffusion length.



**Figure 2** Diagram of a simplified photoacoustic cell.



liquids and gases. PAS tends to be used to analyze weakly absorbing, highly scattering or highly viscous liquids. These sample types are difficult to analyze by conventional techniques.

The photoacoustic cell consists of a small volumed, vibrationally isolated, closed chamber. The microphone that detects pressure changes within the cell chamber forms part of cell wall. The incident excitation beam enters the PAS cell via the window. The transmission range of this window defines the spectral range of the photoacoustic cell. When working in the IR region a KBr window is used. The greatest limitation of PAS is the background signal, generated by the cell window and walls absorbing the incident beam. By using optically transparent windows and polished cell walls this absorbance is reduced.

Following excitation and subsequent nonradiative de-excitation of the sample pressure waves are generated at the sample-boundary gas interface. These waves propagate through the gas-filled chamber to the acoustic microphone where they are detected and converted into an electrical signal. The gas through which the thermal waves travel must be nonabsorbing and transparent to the spectral range of interest. Usually air or helium is used. Helium is favored because its thermal coupling properties increase the sensitivity of the technique by two to three times, owing to its decreased heat capacity compared to air. Additionally, helium purging eliminates water and carbon dioxide, which, unless removed, are present in the resultant spectra. The ability to remove air and moisture allow the analysis of unstable samples in an inert atmosphere, an advantage over conventional spectroscopic methods.

When using the PAS cell illustrated in **Figure 2**, solids and liquids are simply placed in the sample cup. Spectral analysis of a gaseous compound is performed by filling the PAS cell with the gas. All spectra are normalized by division with a spectrum of carbon black, a procedure that removes spectral variations due to the source, optics, and detector.

Solids should be dried in an oven/desiccator prior to analysis to ensure no water is generated upon irradiation and the incident beam should be attenuated when analyzing liquids to reduce evaporation. These precautions are essential in FTIR-PAS because the KBr window is easily damaged by water and gaseous compounds condensing upon it. When analyzing solids, peak intensities can be dependent upon the particle size and packing density, with the PAS signal increasing as the particle size decreases, and the surface area increases.

Spectral analysis of a gaseous compound is performed by filling the PAS cell with the gas. A cylindrical

cell with windows at each end is favored for gas studies. The cell is filled with the gas and the incident beam passes directly through the cell. This system reduces reflections of the incident beam on the cell walls, which would occur when using the cell above, and affect the PAS signal. Photoacoustic signal generation in a gas varies slightly with solids and liquids. When analyzing solids and liquids the photoacoustic signal is detected indirectly by monitoring periodic changes in the boundary gas surrounding the sample. When analyzing gases the pressure variations are measured directly within the gas. The gas absorbs the modulated light upon excitation increasing its pressure, and the microphone detects the resulting pressure fluctuations that occur at the same frequency as the modulated incident light.

### Modulation of the Incident Beam

The incident beam must be modulated prior to photoacoustic excitation of the sample. When the beam is modulated it oscillates in intensity. Modulation of the beam is essential for the periodic production of the heat. Conventionally modulation is performed using a rotating sector (chopper) that interrupts the beam periodically, resulting in the periodic change of the pressure within the cell. Depth profiling is performed by changing the chopping frequency, and hence modulation frequency of the incident radiation. High chopping frequencies correspond to shallow depths and low frequencies correspond to depths deep within the sample. This method of modulation is successful but can introduce significant noise contributions to the PAS signal especially when probing shallow depths when high chopping frequencies are required. A chopper is often used to modulate the incident beam (amplitude modulation) when using a dispersive spectrometer for PAS studies. This can introduce additional problems because the PAS signal is directly proportional to the intensity of the beam, but dispersive instruments experience power loss at the input and exit slits reducing the incident beam intensity and the PAS signal.

Currently, in the IR region, modulation is performed using interferometric methods. The interferometer has a greater optical throughput, signal-to-noise ratio, resolution, speed, and spectral range, than the dispersive instrument. (The greater signal-to-noise ratio arises from the interferometer's multiplex and throughput advantages.) Consequently, by combining interferometry and photoacoustic spectroscopy, problems associated with the dispersive spectrometer and chopper setup are avoided, resulting in the significant development of the technique. Modulation by interferometry is the preferred method

for the majority of photoacoustic studies performed today.

An idealized Michelson interferometer is illustrated in **Figure 3**. The interferometer modulates the incident beam by changing the optical path difference of the interferometer. The optical path difference can be changed continuously or in increments, methods called rapid scanning and step scanning, respectively.

In 'rapid scan mode' the motion of the interferometer's moving mirror modulates the incident beam sinusoidally. The mirror travels at a constant velocity,  $v$ , along its trajectory, continuously varying the optical path difference, modulating the beam at a frequency  $f = 2v/\lambda$ , where  $\lambda$  is the wavelength of incident beam and  $v$  the velocity of the moving mirror. The modulation frequency is dependent upon wavenumber; therefore, as defined by eqn [1], the thermal diffusion length (PAS sampling depth) varies linearly across the spectral range, increasing with decreasing wavelength. This dependence of thermal diffusion length on modulation frequency can be used to depth profile materials.

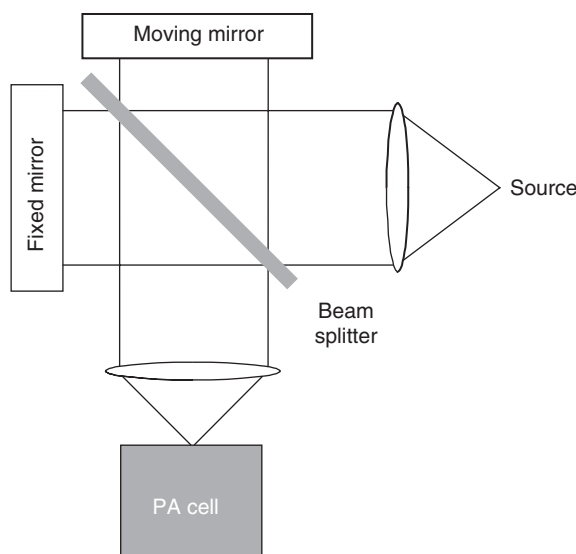
In 'step scan mode' the optical path difference is changed incrementally, the mirror does not scan continuously. It steps to a position, collects data, and moves to the next position. Data are collected at each step with the path difference held constant. At each step the mirror is jittered at a constant frequency, a procedure that modulates the incident beam. The ability to select a constant modulation frequency, independent of the mirror velocity and wavenumber, results in a constant probed depth over the entire spectral range. This is advantageous over the varying sampling depth observed when using rapid scanning,

because it simplifies depth-profile analysis. Step scanning also permits the use of phase modulation techniques to obtain higher signal-to-noise ratios than can be obtained by amplitude (chopper) modulation.

## Applications of PAS

Since the early days of PAS two of its properties have been important in defining its role as a complementary technique to other molecular spectrometric techniques in the visible near-IR and IR regions of the electromagnetic spectrum. These are the freedom of effects of scattering particles, which derives from the fact that a thermal signal is measured via the acoustic wave rather than attenuation or reflection of an optical beam, and the depth-profiling capability as described in the previous section. Thus, the technique has proved extremely useful for making direct measurements of solid materials that would be rendered impossible by optical techniques because of their opacity. In addition, as the PAS technique utilizes measurement of an acoustic wave in gases it has also found a further limited use in atmospheric pollution measurements. There are few reviews available that describe applications of the technique, so in the remainder of this article we have selected a few recent publications for discussion to illustrate the range of applications of the PAS technique.

The predominant trend in recent years has been the growth of FTIR-PAS driven by the ease with which depth information can be obtained through either rapid scanning or step scanning/phase modulation techniques. The use of FTIR-PAS has also opened up the possibility of recording time-resolved depth measurements to facilitate chemical mass transport and diffusion studies, which is of particular interest for drug diffusion studies. For example, Nakamura *et al.* employed a rapid scanning technique and diffusion model to determine the diffusion coefficient of the angina prophylactic nitroglycerin from drug delivery patches through a skin mimetic. The model utilized four different IR wavelengths of the nitroglycerin IR-PAS spectrum to monitor its mass transport rate through the mimetic and to compute the diffusion coefficient. More recent work has refined this model and the diffusion rate thus determined has been shown to compare favorably with that obtained using attenuated total reflection (ATR) IR spectrometry. In a previously published paper Hahn *et al.* used step scan FTIR-PAS with phase modulation to determine penetration rates of dithranol and methoxsalen from a semisolid Vaseline<sup>®</sup> into an artificial dodecanol-collodion membrane by monitoring appropriate IR absorption bands of the active drugs in the formulation. The work



**Figure 3** Diagram of an idealized Michelson interferometer.

demonstrated that FTIR-PAS was most suitable for obtaining the drug penetration profile in the membrane whereas the ATR technique was favored for surface penetration measurements in accordance with the relative depths that can be probed with the techniques, PAS providing deeper measurement depths than ATR.

The depth-profiling capability of step-scan mode FTIR-PAS was also exploited in the measurement of coated papers by Halttunen *et al.* Coatings of cationized starch and sodium oleate (a surfactant) were characterized at depths of 20–60  $\mu\text{m}$ . The study, using a phase modulation technique, was able to demonstrate that the position of the surfactant carbonyl band is influenced by its chemical environment and the presence of the paper, and also that the surfactant is enriched in the outermost surface of the coating. This ability to discriminate chemical composition is perhaps one of the most significant uses to which PAS can be applied and is taken to its extreme in an elegant piece of work published by Jiang in which the microsampling capabilities of FTIR-PAS are explored. The technique was applied to single particles and fibers coated with oils, greases, and gels in order to determine their heterogeneity. Thus, depth information concerning oil-coated polymer beads of 150  $\mu\text{m}$  diameter and gel-coated human hair was obtained. The ability to study such discrete and small entities may well have significance in the drive to apply spectroscopic techniques to single biochemical cell analysis and Lab-on-a-chip applications. Fiber analysis has also been developed by Jurdana *et al.* in a study of nondestructive evaluation of keratin to determine spectroscopic differences between the cortex and cuticle of wool and hair fibers attribute to differences in protein composition between these domains. In that study both rapid and step scans were used to depth profile the samples.

Turning to gas analysis, there has been interest in the use of PAS to study systems of environmental importance. A particularly useful study has been published by Sigrist, who used a tuneable CO IR-laser PA system to study multicomponent gas mixtures, in particular volatile organic compounds in automobile exhaust systems. A separate CO<sub>2</sub> laser PAS system was also employed for air monitoring in urban and rural locations for the selective determination of species such as H<sub>2</sub>O, CO<sub>2</sub>, NH<sub>3</sub>, O<sub>3</sub>, and C<sub>2</sub>H<sub>2</sub> in the parts per billion to parts per million concentration range. Similarly, a CO<sub>2</sub> laser PAS system has been applied to the analysis of gas samples emitted by a diesel engine left running in a closed room, by Schramm *et al.* The species of interest, SO<sub>2</sub> and NO<sub>2</sub>, were detected in the parts per million range. The results demonstrated the importance of

considering the contribution of diesel exhaust from generating equipment to atmospheric pollution.

There are many other applications of PAS including the analysis of industrial samples not amenable to conventional spectroscopic analysis using UV/visible and IR spectrometry, especially those solid sample that are opaque. We have only outlined a few of the more interesting and as some would say academic examples of the use of PAS. However, this can be justified in pointing the way forward to the requirement for research in the analysis of specialist materials and in specialist scientific areas that require information that is difficult or tedious to achieve by conventional means. It is perhaps here that the niche for the still relatively uncommon technique of PAS will be best applied.

**See also:** **Infrared Spectroscopy:** Sample Presentation; Photothermal.

## Further Reading

- Adams MJ, King AA, and Kirkbright GF (1976) Analytical optoacoustic spectrometry. Part I: Instrument assembly and performance characteristics. *Analyst* 101: 73–85.
- Adams MJ, Beadle BC, King AA, and Kirkbright GF (1976) Analytical optoacoustic spectrometry. Part II: Ultraviolet and visible optoacoustic spectra of some inorganic, biochemical and phytochemical samples. *Analyst* 101: 553–561.
- Bialkowski SE (1996) Photothermal spectroscopy methods for chemical analysis. In: Winefordner J (ed.) *Chemical Analysis*, vol. 134. New York: Wiley.
- Halttunen M, Tenhunen J, Saarinen T, and Stenius P (1999) Applicability of FTIR/PAS depth profiling for the study of coated papers. *Vibrational Spectroscopy* 19: 261–269.
- Hanh BD, Neubert RHH, Wartewig S, Christ A, and Hentzsch C (2000) Drug penetration as studied by non-invasive methods: Fourier transform infrared-attenuated total reflection, Fourier transform infrared, and ultraviolet photoacoustic spectroscopy. *Journal of Pharmaceutical Sciences* 89: 1106–1113.
- Jiang EY (1999) Heterogeneity studies of a single particle/fibre by using Fourier transform infrared micro-sampling photoacoustic spectroscopy. *Applied Spectroscopy* 53: 583–587.
- Jurdana LE, Ghiggino KP, Leaver IH, and Cole-Clarke P (1995) Application of FT-IR step-scan photoacoustic phase modulation methods to keratin fibres. *Applied Spectroscopy* 49: 361–366.
- McClelland JF, Bajic SJ, Jones RW, and Seaverson LM (1998) Photoacoustic spectroscopy. In: Mirabella FM (ed.) *Modern Techniques in Applied Molecular Spectroscopy*, pp. 221–265. New York: Wiley.
- Nakamura O, Lowe RD, Mitchem L, and Snook RD (2001) Diffusion of nitroglycerin from drug delivery patches through micro-fibre filters using Fourier transform

- infrared photoacoustic spectrometry. *Analytica Chimica Acta* 427: 63–73.
- Rosencwaig A and Gersho A (1976) Theory of the photoacoustic effect with solids. *Journal of Applied Physics* 47: 64–69.
- Rosencwaig A (1980) *Photoacoustics and Photoacoustic Spectroscopy*. New York: Wiley.

- Schramm DU, Sthel MS, da Silva MG, *et al.* (2003) Application of laser photoacoustic spectroscopy for the analysis of gas samples emitted by diesel engines. *Infrared Physics and Technology* 44: 263–269.
- Sigrist MW (1995) Trace gas monitoring by laser-photoacoustic spectroscopy. *Infrared Physics and Technology* 36: 415–425.

## PIEZOELECTRIC SENSORS

See **SENSORS: Piezoelectric Resonators**

## PIXE

See **SURFACE ANALYSIS: Particle-Induced X-Ray Emission**

## PLASTICS

**V Pacáková**, Charles University, Prague, Czech Republic

**J Virt**, Dekonta, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

A plastic is a material that contains as a major component one or more organic polymeric substances of high molecular weight. It is a solid in its final state, and, at some stage in its manufacture or processing into final products, can be shaped by flow. One of the most important classes of materials in which plastics are contained is synthetic resins. These resins vary widely in their chemical composition and their physical properties. The number of synthetic resins that can be synthesized is large; however, relatively few have achieved commercial importance. Well over 90% of all synthetic resins made today comprise no more than 20 different types, although there are certain variations within each type. Plastics also have other uses (in the manufacture of surface coatings, glues, synthetic textile fibers, etc.).

Synthetic resins can be divided into two classes: thermosetting and thermoplastic resins, each class differing in its behavior on heating. Both types are composed of macromolecules; the difference in their

thermal behavior is due to differences in the structure of the resins. A thermoplastic polymer is capable of repeatedly being softened by heating and hardened by cooling through a characteristic temperature range and in the softened state can be shaped by flow into articles, mostly by injection molding or extrusion. A thermosetting polymer, having been cured by heat or other means, is substantially infusible and insoluble. The macromolecules of a thermosetting resin often consist of extensively branched chains that are chemically cross-linked, thus forming a complex network. On heating, there is less possibility of free movement, so that the material remains rigid.

Plastics can be divided according to their character into amorphous and crystalline. Crystallization is never complete and the so-called crystalline polymers are virtually semicrystalline ones. Examples of amorphous plastics are polystyrene, acrylonitrile–butadiene–styrene copolymers, styrene–acrylonitrile copolymers, polymethylmethacrylate, poly(vinyl chloride), cellulose acetates, phenylene oxide-based resins, polycarbonates, etc. Amorphous polymers are characterized by their glass transition temperature, semicrystalline polymers by both melting and glass transition temperatures.

Most commercially produced plastics are complex mixtures of many components, involving the

- infrared photoacoustic spectrometry. *Analytica Chimica Acta* 427: 63–73.
- Rosencwaig A and Gersho A (1976) Theory of the photoacoustic effect with solids. *Journal of Applied Physics* 47: 64–69.
- Rosencwaig A (1980) *Photoacoustics and Photoacoustic Spectroscopy*. New York: Wiley.

- Schramm DU, Sthel MS, da Silva MG, *et al.* (2003) Application of laser photoacoustic spectroscopy for the analysis of gas samples emitted by diesel engines. *Infrared Physics and Technology* 44: 263–269.
- Sigrist MW (1995) Trace gas monitoring by laser-photoacoustic spectroscopy. *Infrared Physics and Technology* 36: 415–425.

## PIEZOELECTRIC SENSORS

See **SENSORS: Piezoelectric Resonators**

## PIXE

See **SURFACE ANALYSIS: Particle-Induced X-Ray Emission**

## PLASTICS

**V Pacáková**, Charles University, Prague, Czech Republic

**J Virt**, Dekonta, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

A plastic is a material that contains as a major component one or more organic polymeric substances of high molecular weight. It is a solid in its final state, and, at some stage in its manufacture or processing into final products, can be shaped by flow. One of the most important classes of materials in which plastics are contained is synthetic resins. These resins vary widely in their chemical composition and their physical properties. The number of synthetic resins that can be synthesized is large; however, relatively few have achieved commercial importance. Well over 90% of all synthetic resins made today comprise no more than 20 different types, although there are certain variations within each type. Plastics also have other uses (in the manufacture of surface coatings, glues, synthetic textile fibers, etc.).

Synthetic resins can be divided into two classes: thermosetting and thermoplastic resins, each class differing in its behavior on heating. Both types are composed of macromolecules; the difference in their

thermal behavior is due to differences in the structure of the resins. A thermoplastic polymer is capable of repeatedly being softened by heating and hardened by cooling through a characteristic temperature range and in the softened state can be shaped by flow into articles, mostly by injection molding or extrusion. A thermosetting polymer, having been cured by heat or other means, is substantially infusible and insoluble. The macromolecules of a thermosetting resin often consist of extensively branched chains that are chemically cross-linked, thus forming a complex network. On heating, there is less possibility of free movement, so that the material remains rigid.

Plastics can be divided according to their character into amorphous and crystalline. Crystallization is never complete and the so-called crystalline polymers are virtually semicrystalline ones. Examples of amorphous plastics are polystyrene, acrylonitrile–butadiene–styrene copolymers, styrene–acrylonitrile copolymers, polymethylmethacrylate, poly(vinyl chloride), cellulose acetates, phenylene oxide-based resins, polycarbonates, etc. Amorphous polymers are characterized by their glass transition temperature, semicrystalline polymers by both melting and glass transition temperatures.

Most commercially produced plastics are complex mixtures of many components, involving the



polymer matrix, oligomers, residual monomers, and additives, such as plasticizers, antioxidants, lubricants, and flame retarders. All these components determine the properties of plastics. The analysis of plastics is more complicated than other organic analyses. For example, functional group analysis is limited by the poor solubility of polymers and often their degradation products (pyrolytic or hydrolytic) must be investigated. Physicochemical methods, primarily spectroscopic, thermoanalytic, and chromatographic, make the analysis of plastics substantially easier.

## Analytical Techniques

Analysis of plastics is a complex task and involves preliminary tests, determination of nonmetallic and metallic elements, analysis of functional groups and double bonds, molecular weight determinations, chemical compositional analysis, sequence length distribution in copolymers, determination of tacticity and branching in polymers, and analysis of additives. Because of the great variety in structures in commercially produced plastics, the number of methods that can be applied for their analysis is considerable.

## Preliminary Tests

An unknown plastic from which additives or admixtures have been removed can be subjected to the following tests:

1. Evaluation of the plastic properties, e.g., coloration, tactile properties, smell, elasticity, hardness or fragility, transparency, translucence, etc. The manner in which the plastic is applied can also be revealing.
2. Flame tests (orientative) provide rough information on the test polymer.
3. Pyrolysis test. Thermoplastic and thermosetting polymers can be differentiated by this test.
4. Solubility test. Solubility in acetone, benzene, dimethylformamide, ethanol, chloroform, water, dioxan, acetic acid, acetates, formic acid, phenols, ethylene glycol, tetrahydrofuran, etc., is tested.
5. Chemical characteristics. These include hydroxyl and acid values and saponification values.
6. Physical characteristic. Melting temperatures  $T_m$  (melting ranges) and glass transition temperature  $T_g$  are characteristic constants of plastics and thus can be used for their identification and quality control. Table 1 lists  $T_g$  and  $T_m$  values for selected polymers.

**Table 1** Glass transition temperatures  $T_g$  and melting points  $T_m$  of selected polymers

Polymers	$T_g$ ( $^{\circ}\text{C}$ )	$T_m$ ( $^{\circ}\text{C}$ )
Polyacrylonitrile	100	320
Polyamide 6	40	220
Polyamide 6,6	50	255
Polyamide 6,10	46	226
Polyamide 11		186
Polybutadiene	-86	-20
Polycarbonate	155	235
Polyisobutylene	-73	44
Polyethylene, low density	-100	120
Polyethylene, high density	-70	135
Polypropylene	-30	165
Poly(butylene terephthalate)	65	220
Poly(ethylene terephthalate)	69	256
Poly(ethylene vinyl acetate)	-20 to +20 <sup>a</sup>	40-100 <sup>a</sup>
Poly(fluoroethylene propylene)		280
Poly(methylmethacrylate)	105	
Polytetrafluoroethylene	-20	327
Poly(vinyl acetate)	30	
Poly(vinyl alcohol)	85	
Poly(vinyl chloride)	85	190
Poly(vinylidene chloride)	-17	
Poly(vinylidene fluoride)		178

<sup>a</sup> Depending on ethylene content.

Adapted from Widman G (1987) Thermal analysis of plastics. *American Laboratory* 19: 98-103.

## Determination of Metallic and Nonmetallic Elements

Polymers contain various elements, metallic and nonmetallic. Some elements are a constituent part of the monomers, such as nitrogen in acrylonitrile or chlorine in poly(vinyl chloride), while other elements occur as impurities or are part of some additives (e.g., zinc stearate). Their concentrations range from a  $\mu\text{g}$  per kg level to several percent. Analysis of the element content is especially important for manufacturing process control. Elements can be determined after chemical or physical destruction of polymer, or directly by nondestructive methods.

### Destructive Methods

**Dry ashing with or without ashing aid** This involves combustion of the polymer under controlled conditions in a platinum crucible (possibly with an ashing aid, e.g., magnesium carbonate, to prevent losses). The residue is then treated with dilute nitric acid or by fusion with potassium peroxodisulfate.

**Wet ashing with sulfuric acid** This is a very common technique. It involves oxidative degradation of plastic with sulfuric acid.

**Microwave-assisted digestion in a closed vessel** This method is based on degradation of plastic with

concentrated  $\text{HNO}_3/\text{HF}$  in a microwave furnace. This method gives the best results.

#### **Fusion with sodium carbonate or sodium peroxide**

The polymer is mixed with sodium carbonate or a mixture of sodium peroxide and sucrose or ethylene glycol, and ashed in a muffle furnace at  $500^\circ\text{C}$  (for sodium carbonate) or in an autoclave. The method is suitable for determination of zinc, chlorine, and bromine.

**Oxygen flask combustion** Combustion takes place at atmospheric pressure in oxygen; water is used as the absorbent. It is suitable for determining sulfur, phosphorus, chlorine, and bromine.

**Micro-Kjeldahl digestion method** The classical method has been adapted for determination of nitrogen in plastics.

**Atomic absorption spectrometry and atomic emission spectrometry** These are very useful techniques for determination of metals at low concentrations.

#### **Nondestructive Methods**

**X-ray fluorescence spectrometry** This technique is extensively used in the industrial analysis of plastics for routine determination of traces of metals and nonmetal elements, i.e., iron, cobalt, nickel, chromium, copper, zinc, chlorine, bromine, titanium, aluminum, sodium, potassium, calcium, magnesium, vanadium, cadmium, and selenium. The main advantages are simple sample preparation and independence on the element state in chemical combination.

**Neutron activation analysis** This is a very sensitive technique applicable to a wide range of elements present in plastics.

**Electron probe microanalysis and scanning electron microscopy** These are suitable techniques for identifying the nature of metallic inclusions in polymer films and sheets.

### **Determination of Molecular Weight**

Polymer consists of macromolecules with different chain lengths. The average molecular weight of a plastic, i.e., number average, mass average, and viscosity average, can be determined by osmometric, light scattering, and viscometric measurements and by mass spectrometry, respectively. Average molecular weight of plastics can also be calculated from the results of functional analysis of the end groups (i.e.,  $-\text{COOH}$ ,  $-\text{NH}_2$ ,  $-\text{SH}$ ,  $-\text{OH}$ ). The distribution of molecular weights in polymers can be determined firstly by size-exclusion chromatography (see below),

and also by solubility fractionation and turbidimetric titration.

### **Determination of the Structure of Plastics**

The functional groups and the degree of unsaturation are the main structural components to be determined in plastics. Functional groups and double bonds can be present in plastics, over a wide range of concentration, from a few  $\mu\text{g}$  per kg to several percent. On the basis of their determination, the polymer can be identified. Functional analysis of the end groups gives the average molecular weight of a plastic. Techniques for their determination can be divided into chemical and instrumental methods.

#### **Microchemical Methods**

Many of the techniques that are applicable to the determination of functional groups and double bonds in organic compounds are also applicable to polymers. Typical examples are:

- a titration procedure for the determination of carboxylic groups in acrylic copolymers;
- acetylation and phthalation procedures for determination of hydroxyl groups (determination of primary and secondary hydroxyl groups using reaction with toluenephenyl isocyanate and/or reaction with toluene diisocyanate, respectively);
- a saponification procedure for the determination of ester groups;
- derivatization methods, for example, the reaction of 2,4-dinitrophenylhydrazine with carbonyl groups present at low concentrations, followed by ultraviolet (UV)–visible spectrophotometry;
- iodine monochloride and bromination procedures for determination of the unsaturation degree (e.g., in styrene–butadiene copolymers and polymethylacrylate, respectively); and
- hydrogenation methods for total unsaturation in polymers.

#### **Instrumental Methods**

Spectral methods based on UV–visible spectrophotometry, laser-induced breakdown spectroscopy (LIBS), infrared (IR), Raman, nuclear magnetic resonance (NMR) spectroscopy, mass spectrometry (MS), thermoanalytical and chromatographic methods, especially liquid chromatography (LC) or gas chromatography (GC) combined with pyrolysis are most common.

## Spectroscopic Methods

**Ultraviolet spectrophotometry** This technique has been used extensively for polymers of the polystyrene or polyacrylate types. It has also been used for the estimation of optical properties of polymers (transparency, haze, color, and color stability) and quantitative analysis of additives (antioxidants, UV stabilizers, etc.).

**Infrared spectrometry** This spectrometry is one of the most frequently used methods in analysis of plastics. IR spectra in the region of up to  $4000\text{ cm}^{-1}$  are characteristic of the polymer. IR spectra of plastics can be measured by the thin-film, KBr-disc, or attenuated total reflectance (ATR) techniques. The ATR technique is very useful when normal transmission methods fail, e.g., for analyzing surfaces of elastomeric materials with low glass transition temperatures (polyethylene, polypropylene, poly(vinyl chloride), urethanes, polyester-melamine, and silicon-containing polymers), determining crystallinity, surface modification, and defect structures in materials. Fourier transform infrared (FTIR) spectrometers are well suited to all spectrometric purposes, especially to ATR and near-IR (NIR) techniques. Compared with dispersive IR spectrometers, they have a greater energy throughput, a higher resolution, a better wavelength calibration, and a higher signal-to-noise ratio. Hundreds of scans can be run and co-added in a few minutes. Plastics are identified on the basis of a comparison of measured IR spectra with computer-retrieved spectra of standard polymers (fingerprinting). Identity of a plastic can be confirmed by other methods, e.g., pyrolysis-gas chromatography. Near-IR spectroscopy has become widely used in analysis of plastics. The main reasons are that NIR instruments are generally easy to operate, results are obtained very fast, and there is little to no sample preparation required.

**Laser-induced breakdown spectroscopy** This is based on the emission spectrometry from the plume induced by a focused laser pulse. It enables both on-line and *in situ* trace determination of elements.

**Mass spectrometry** This is often used for analysis of plastics. Thermal degradation can be carried out using an ion source, or in a pyrolyzer connected directly to a mass spectrometer or via a chromatographic column. Two soft ionization techniques, electrospray ionization and matrix-assisted laser desorption ionization (MALDI), are the most common in the analyses of large molecules. MALDI-MS is usually used offline but online connection with a

separation method such as size-exclusion chromatography is possible. MS methods offer a very high resolution. They have been applied to the qualitative and quantitative analysis of plastics, oligomers, and additives. Search in a mass spectral library and pattern recognition analysis are usually employed for the identification of plastics. Information on number-average molecular weight, molecular weight distribution, end group composition, unsaturation, branching, etc., can be obtained by these techniques.

**Laser-induced mass analysis (LIMA)** This provides information on elements and various species in a single particle.

**Raman spectrometry** This is one of the nondestructive methods of polymer analysis; polymer samples can be measured directly, without any pretreatment. While Raman spectra primarily reveal the vibrations of the molecular skeleton (e.g., carbon atom chains), IR spectra are better suited for observing the vibrations of polar groups. An advantage of Raman spectrometry is the possibility of measurement over the whole wavelength range, from 5 to  $4000\text{ cm}^{-1}$ . Raman spectra provide information and limits of detection similar to those of IR spectra, but IR spectra are easier to obtain and cheaper. Therefore, Raman spectrometry is less common in industrial analyses of plastics.

**Nuclear magnetic resonance spectroscopy** NMR, especially  $^{13}\text{C}$  NMR, is well suited for structural analysis of polymers. It gives information on chemical compositional distribution, branching, and cross-linking in polymers and polymer tacticity. The sequence distribution in copolymers can be calculated from NMR data, e.g., for vinyl copolymers from chemical shift of the olefinic carbons of vinyl monomers. Heterogeneity, e.g., due to fluctuation in parameters involved in the reaction process during polymerization, can be followed by solid-state  $^{13}\text{C}$  and  $^3\text{H}$  NMR, expressed in terms of two-dimensional presentation. Curing of polymers can be followed by  $^2\text{H}$  NMR or  $^{13}\text{C}$  CP-MAS NMR. NMR has been extensively used for determination of degree of unsaturation and for functional analysis of carboxyl, ester, and carbonyl groups in plastics.

## Chromatographic Methods

After spectroscopic methods (IR, UV, NMR), chromatographic methods are the most suitable for separation, identification, and determination of plastics. They are used for determination of both volatile and

nonvolatile components, e.g., monomers, residual solvents, additives, oligomers, and polymer matrix.

**Gas chromatography** This is best suited for the analysis of volatile components, i.e., residual monomers and solvents. There are two possible methods:

1. headspace analysis; and
2. direct injection of a polymer solution or extract.

In headspace analysis, the plastic is placed in a vial (at a raised temperature) and the volatiles formed are stripped by a flow of carrier gas. The stripped volatiles are trapped in a suitable sorbent (e.g., using a solid-phase microextraction device) and subsequently thermally desorbed into a gas chromatograph. Process gas chromatographs are used in industrial analysis of volatiles in plastics. An example of this technique is the determination of residual vinyl chloride monomer in plastics in the range of 5–50 g per kg. With direct injection of a polymer solution, there is a danger of side-effects (a loss of reactive monomers due to polymerization in the injection port or an increase in its content due to depolymerization at a high injection temperature).

GC with thermostable stationary phases can be used for the analysis of substances with relatively high boiling temperatures. The possibilities of GC have been expanded using specific detectors, i.e., the electron-capture detector, alkali flame ionization detector, detection with an inductively coupled plasma, or by coupling GC with MS or FTIR (GC–MS or GC–FTIR). GC is also used for analysis of additives (slip agents, stabilizers, antioxidants) either directly or after conversion into volatile derivatives. It can also be utilized on a preparative scale for isolation of polymer additives for further identification, e.g., by using NMR. Inverse GC utilizes plastics as the GC stationary phases. The melting temperature, glass transition temperature, etc., can also be determined by this method.

**Thin-layer chromatography** This is often used for semiquantitative determination of stabilizers in plastics. The method is simple, cheap, easy to carry out, and can also be utilized for preparative purposes for identification by more sophisticated techniques (IR, NMR, MS).

**Liquid chromatography** This is a very efficient and versatile separation method. Depending on the separation mode (adsorption, separation on chemically bonded phases, or size exclusion), it can be used for analysis of stabilizers, monomers in polymers, oligomers, and high molecular weight polymers.

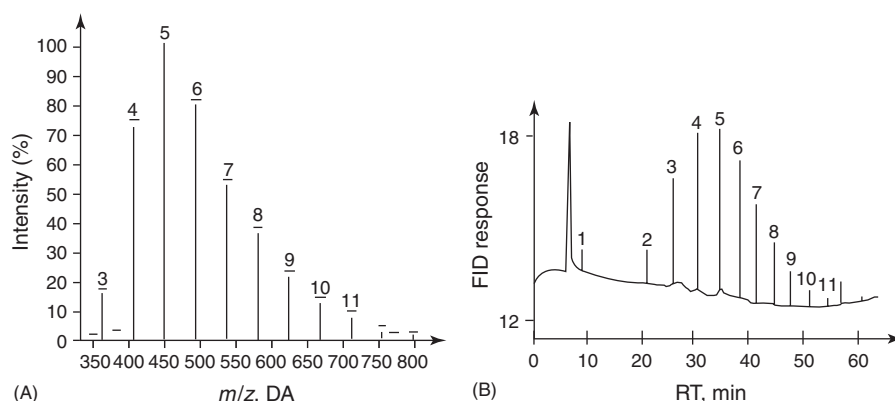
**Size-exclusion chromatography (SEC)** This is a powerful tool for evaluating size distributions in oligomers and polymers. The retention in SEC depends on the relative size of the solute molecules (Stokes radius) and the size and shape of the pores of the chromatographic stationary phase. Within an optimal linear separation range, the elution order is proportional to the molecular weight. In combination with UV photometric or refractive index detection, SEC is therefore suitable for determination of the relative molecular weight distribution of polymers; calibration with standard compounds is necessary. Depending on the size-exclusion limit of the stationary phase, either low molecular weight oligomers or high molecular weight polymers can be analyzed. Since the size of macromolecules may be affected by their composition and structure (e.g., long-chain branching), molecular-weight-sensitive detectors, e.g., low-angle or multi-angle laser light scattering detectors, viscometers, and MALDI-MS have been applied for the determination of absolute molecular weights. Several other detectors are useful in polymer analysis: diode-array UV detectors are important in determination of heterogeneity of copolymers and complex polymer systems; functional group distribution can be measured with an FTIR detector.

**Adsorption (liquid–solid) chromatography (LSC), normal- or reversed-phase liquid chromatography (NPLC, RPLC)** As polymers cannot be separated by SEC on the basis of their composition, other LC modes, e.g., LSC, NPLC, or RPLC, must be employed. Since the retention in these separation modes is proportional to the length of the polymer units, a gradient elution is necessary for polymer separation. Various detectors, similar to those applied in SEC, are used. High resolution can be obtained in RPLC. This is demonstrated in an example of separation of phenoxy polyethylene oxide oligomers (**Figure 1**). Analysis of oligomers is important for process control, as the oligomer's profile (oligomer formation, chain defects, and differences in the end groups) can be used in the modification of processing and synthesis procedures.

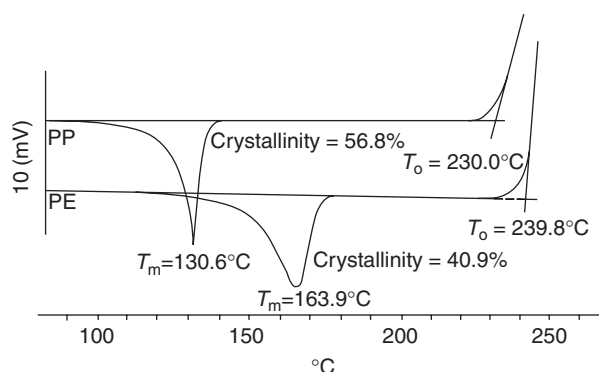
**Multidimensional chromatography** This combines two or more different separation modes in one experiment, e.g., SEC and RPLC, with different mobile phases, and is a very useful technique in analysis of polymers. The SEC column serves for a pre-separation step according to the molecular weight and the RPLC column then separates the substances in the fractions obtained from the SEC column according to their chemical composition. A scheme of the







**Figure 3** SFC analysis with MALDI-MS of the main oligomers from **Figure 1**. (Reproduced with permission from Pasch H and Trathnigg B (1999) *HPLC of Polymers*. Berlin: Springer; © Springer-Verlag.)



**Figure 4** Thermal analysis of polypropylene (PP) and polyethylene (PE). Crystallinities, melting temperatures,  $T_m$ , and oxidation induction temperatures,  $T_o$ , are indicated. (Reproduced with permission from Widman G (1987) *Thermal analysis of plastics*. American Laboratory 19: 98–103.)

decomposed and the changes connected with the thermal degradation process are recorded by thermogravimetry (TG) or DSC. Often, differential thermogravimetry (DTG) curves are characteristic of certain plastics and can, therefore, be used for identification purposes. The behavior of plastics during oxidation is important for their practical use and for their recycling or environmental decomposition. The method permits the determination of melting temperature  $T_m$ , crystallinity, and the oxidation induction temperature as shown in **Figure 4**.

**Glass transition and melting behavior** Glass transition temperature  $T_g$  is an important parameter used for identification of plastics. The  $T_g$  value is the temperature at which amorphous polymers change from hard to soft. The concentration of crystalline regions in amorphous (semicrystalline) polymers affects the rigidity of the polymer. The crystalline melting

temperature  $T_m$  can be used for quality control and for identification of polymers. Glass transition temperatures ( $T_g$ ) and crystalline melting temperatures ( $T_m$ ) of selected polymers are listed in **Table 1**. Additives, solvent residues, moisture, etc. can lower the  $T_g$ , which is also influenced by the degree of curing; therefore,  $T_g$  allows the curing level to be monitored. Melting and softening processes are also influenced by the composition of copolymers and thus facilitate the determination of the composition of copolymers.

**Composition** The height of TG steps determines the quantity, while the temperature of DTG peaks determines the quality. Direct coupling of a thermal analyzer to a mass spectrometer enables the identification of degradation products, such as water, inorganic gases, volatile organic compounds.

**Curing reactions, degree of curing** Curing of thermosets can be monitored using thermoanalytical methods. Cross-linking and glass transition temperature,  $T_g$ , increase with increasing degree of curing. Differential photocalorimetry is a very effective method for process control, as both light and heat flow can be measured during curing.

## Analysis of Additives in Plastics

The properties of plastics are usually improved by adding different types of additives such as antioxidants, plasticizers (e.g., fatty acids, alcohols, esters, or hydrocarbons), UV stabilizers, flame retardants, slip agents, etc. Analytical problems arise from the following factors: the presence of additives in a more or less insoluble polymer matrix, the high reactivity of many types of additives, especially antioxidants, and the low concentrations of additives often present in the polymer matrix. Most methods for the

determination of additives in plastics fall in one of the following categories:

1. Analysis of compounds released upon heating the plastic (only applicable to volatile additives); gas chromatography is usually preferred.
2. Analysis of solvent extracts of plastics. Acetone, carbon disulfide, chloroform, cyclohexane, diethyl ether, ethanol, hexane, toluene, and water are used as solvents, extracts are further analyzed by, e.g., UV-visible spectrophotometry, fluorescence and phosphorescence methods, GC, LC, electrochemical methods, etc.
3. Direct examination of plastic, i.e., nondestructive testing of polymer films by IR or UV spectroscopy or of thicker sections of polymer by ATR IR spectroscopy to estimate antioxidants at levels ranging from 0.002% to 1.00%.

See also: **Activation Analysis:** Neutron Activation. **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Principles and Instrumentation. **Chromatography:** Overview; Principles. **Gas Chromatography:** Pyrolysis; Mass Spectrometry. **Headspace Analysis:** Static; Purge and Trap. **Infrared Spectroscopy:** Near-Infrared; Industrial Applications. **Liquid Chromatography:** Normal Phase; Reversed Phase; Size-Exclusion. **Microscopy Techniques:** Scanning Electron Microscopy. **Polymers:** Natural Rubber; Synthetic. **Process Analysis:** Chromatography. **Sample Dissolution for Elemental Analysis:** Dry

Ashing; Oxygen Flask Combustion. **Supercritical Fluid Chromatography:** Applications. **Thermal Analysis:** Overview. **X-Ray Fluorescence and Emission:** Wavelength Dispersive X-Ray Fluorescence; Energy Dispersive X-Ray Fluorescence.

## Further Reading

- Barth HG and Mays JW (eds.) (1991) *Modern Methods of Polymer Characterization*. New York: Wiley.
- Campbell D and White JR (1989) *Polymer Characterization*. London: Chapman and Hall.
- Crompton TR (1989) *Analysis of Polymers: An introduction*. Oxford: Pergamon.
- Dechant J (1972) *Ultrarotspektroskopische Untersuchungen an Polymeren*. Berlin: Akademie-Verlag.
- Glöckner G (1986) *Polymer Characterization by Liquid Chromatography*. Amsterdam: Elsevier.
- Kraft M (1973) *Struktur und Absorptionsspektroskopie der Kustostoffe*. Weinheim: Verlag Chemie.
- Prowder T (1999) *Chromatography of polymers: Hyphenated and Multidimensional Techniques*. Washington: ACS.
- Schoenmakers PJ and de Koster CG (2002) LC-MS of synthetic polymers: the parable of the lame and the blind. *Recent Applications in LC-MS* 14: 2-7.
- Schröder E, Müller G, and Arndt KF (1989) *Polymer Characterization*. Berlin: Akademie-Verlag.
- Urban MW (1996) *ATR Spectroscopy of Polymers: Theory and Practice*. Washington: ACS.

# PLATINUM METALS

See **PRECIOUS METALS**

# POLAROGRAPHY

Contents

**Overview**

**Techniques**

**Inorganic Applications**

**Organic Applications**

## Overview

**V Cerdà**, University of the Balearic Islands, Palma de Mallorca, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Polarography was first developed by Heyrovsky in 1922. The polarographic method was defined by the inventor as electrolysis with a polarizable dropping mercury electrode (DME) and it gained rapid acceptance not only in the investigation of electrode

determination of additives in plastics fall in one of the following categories:

1. Analysis of compounds released upon heating the plastic (only applicable to volatile additives); gas chromatography is usually preferred.
2. Analysis of solvent extracts of plastics. Acetone, carbon disulfide, chloroform, cyclohexane, diethyl ether, ethanol, hexane, toluene, and water are used as solvents, extracts are further analyzed by, e.g., UV-visible spectrophotometry, fluorescence and phosphorescence methods, GC, LC, electrochemical methods, etc.
3. Direct examination of plastic, i.e., nondestructive testing of polymer films by IR or UV spectroscopy or of thicker sections of polymer by ATR IR spectroscopy to estimate antioxidants at levels ranging from 0.002% to 1.00%.

See also: **Activation Analysis:** Neutron Activation. **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Principles and Instrumentation. **Chromatography:** Overview; Principles. **Gas Chromatography:** Pyrolysis; Mass Spectrometry. **Headspace Analysis:** Static; Purge and Trap. **Infrared Spectroscopy:** Near-Infrared; Industrial Applications. **Liquid Chromatography:** Normal Phase; Reversed Phase; Size-Exclusion. **Microscopy Techniques:** Scanning Electron Microscopy. **Polymers:** Natural Rubber; Synthetic. **Process Analysis:** Chromatography. **Sample Dissolution for Elemental Analysis:** Dry

Ashing; Oxygen Flask Combustion. **Supercritical Fluid Chromatography:** Applications. **Thermal Analysis:** Overview. **X-Ray Fluorescence and Emission:** Wavelength Dispersive X-Ray Fluorescence; Energy Dispersive X-Ray Fluorescence.

## Further Reading

- Barth HG and Mays JW (eds.) (1991) *Modern Methods of Polymer Characterization*. New York: Wiley.
- Campbell D and White JR (1989) *Polymer Characterization*. London: Chapman and Hall.
- Crompton TR (1989) *Analysis of Polymers: An introduction*. Oxford: Pergamon.
- Dechant J (1972) *Ultrarotspektroskopische Untersuchungen an Polymeren*. Berlin: Akademie-Verlag.
- Glöckner G (1986) *Polymer Characterization by Liquid Chromatography*. Amsterdam: Elsevier.
- Kraft M (1973) *Struktur und Absorptionsspektroskopie der Kustostoffe*. Weinheim: Verlag Chemie.
- Prowder T (1999) *Chromatography of polymers: Hyphenated and Multidimensional Techniques*. Washington: ACS.
- Schoenmakers PJ and de Koster CG (2002) LC-MS of synthetic polymers: the parable of the lame and the blind. *Recent Applications in LC-MS* 14: 2-7.
- Schröder E, Müller G, and Arndt KF (1989) *Polymer Characterization*. Berlin: Akademie-Verlag.
- Urban MW (1996) *ATR Spectroscopy of Polymers: Theory and Practice*. Washington: ACS.

# PLATINUM METALS

See **PRECIOUS METALS**

# POLAROGRAPHY

Contents

**Overview**

**Techniques**

**Inorganic Applications**

**Organic Applications**

## Overview

**V Cerdà**, University of the Balearic Islands, Palma de Mallorca, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

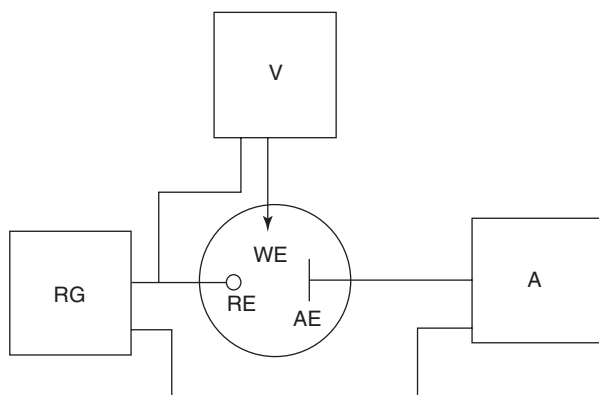
## Introduction

Polarography was first developed by Heyrovsky in 1922. The polarographic method was defined by the inventor as electrolysis with a polarizable dropping mercury electrode (DME) and it gained rapid acceptance not only in the investigation of electrode

processes but, particularly, as a very important analytical technique. In 1925, Heyrovsky and Shikata constructed the polarograph, a device that automatically records current–potential curves. The polarograph was probably the first automatic apparatus to be applied in analytical chemistry. The theoretical basis of the method was laid by Ilkovič in 1934 (the value of the diffusion current) and by Heyrovsky and Ilkovič in 1935 (the shape of the current–potential curve, i.e., the so-called polarogram).

Polarography uses potentiostatic control (Figure 1) of the working electrode polarized under conditions of convective diffusion; the current response  $i$  of the electrode system is measured depending on the polarizing voltage  $E$  whose time rate of change,  $dE/dt$ , is negligible with respect to the time constant of the mass transport toward the electrode surface. This means that measurable  $i$  is independent from the time rate of change of the polarizing voltage. The International Union of Pure and Applied Chemists definition limits polarography to the use of liquid electrodes whose surfaces are periodically or continuously renewed (e.g., dropping and streaming mercury electrodes).

Methods derived from polarography increase analytical sensitivity (or yield new information concerning the mechanism of the electrode process). Pulse polarographic techniques apply discontinuous polarization potential and measure the current response in given time intervals. The alternating current (AC) polarographic techniques are derived from direct current (DC) polarography as described above by imposing a small perturbation voltage of various frequencies upon the main DC polarizing voltage. These methods chiefly detect the adsorption of the participants of the electrode processes.



**Figure 1** Fundamental scheme of a three-electrode polarographic circuit for  $iR$  drop correction. RG, ramp generator; A, current measurement device; V, voltmeter; WE, working electrode; RE, reference electrode; AE, auxiliary electrode.

## Theory

The measurement of electrolytic currents as a function of applied potential imposes certain requirements on the experimental conditions that are determined by properties of the electrode–solution interface. The electrode in the solution of an electrolyte is surrounded by an electrode double layer, with capacity  $C$ , which must be charged when the electrode comes into contact with the solution. The resulting current is termed the charging or capacity current. In order to ensure low variation of the double-layer structure and to obtain sufficient sample conductivity, a suitable strong electrolyte is added. This strong electrolyte is known as an indifferent, supporting, or base electrolyte, which must not interfere with the system to be investigated. As polarographic electrodes, a DME and exceptionally a rotating disk electrode are used. The mean area of the drop in the DME is given by the following expression:

$$\bar{A} = 0.51(mt_1)^{2/3} \quad [1]$$

The instantaneous value at time  $t$  is

$$\bar{A} = 0.85(mt)^{2/3} \quad [2]$$

where  $m$  is the mercury flow rate and  $t_1$  the drop time.

The value  $i_c$  of the charging current depends on the time change of the electrode charge density  $Q$

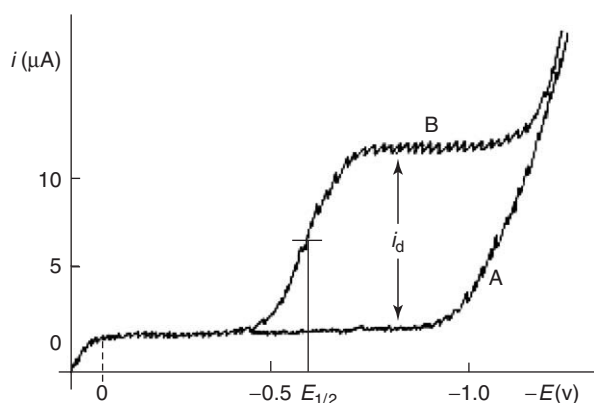
$$i_c = \frac{dQ}{dt} \quad [3]$$

Consequently, the time change of the electrode surface brings about the charging current that would rapidly decay if the electrode were stationary,

$$i_c = (E - E_z)C \frac{dA}{dt} = 2/3 \times 0.85(E - E_z)Cm^{2/3}t^{-1/3} \quad [4]$$

$E$  is the electrode potential and  $E_z$  is the zero charge potential. This shows that the charging current is infinitely large at the beginning of the drop life and decays with time.

When the electrode potential in the presence of an electroactive substance reaches a value characteristic of the nature of the substance being reduced or oxidized, the current starts to increase exponentially, with the potential thus causing a depletion of the electroactive species (the so-called depolarizer) at the electrode surface. At sufficiently high potentials the current increase slows down and finally levels off forming a constant plateau that is usually called the limiting current ( $i_d$ ) (Figure 2). Under these conditions the concentration of the depolarizer at the electrode is equal to zero and



**Figure 2** Polarograms of (A) 1.0 mol l<sup>-1</sup> HCl, (B) 1.0 mol l<sup>-1</sup> Cd(II) 5 × 10<sup>-4</sup> mol l<sup>-1</sup>.

only the particles reaching the electrode surface by mass transport undergo the electrode reaction.

The current caused by the electron exchange between the solid phase of the electrode and the solution particles results in a change of the overall number of electrons of the depolarizer and is called the Faradaic current

$$i = nF \frac{dN}{dt} \quad [5]$$

Here,  $n$  is the number of electrons exchanged by a single particle,  $F$  equals 96 500 C, and  $dN/dt$  is the number of electroactive particles reaching the electrode surface per unit time.

### Mass-Transfer Contributions

The mass transport toward the electrode has three components: diffusion, migration, and convection.

Migration currents are undesirable in the electrolysis of charged species and are mostly suppressed by adding a 20–100 times surplus of indifferent electrolyte over the concentration of electroactive ions.

The convective component arises, e.g., from the growth of the DME.

Diffusion currents are the electrolytic currents controlled by the diffusion rate of the depolarizer from the bulk solution to the electrode surface and are caused by the concentration gradient due to the depletion of the active substance in the close vicinity of the electrode surface. Consequently, the rate of the electrode reaction must be high enough to obtain currents that are controlled by diffusion only. Therefore, the first and second laws of Fick may be applied, and taking into account the initial and boundary concentration conditions the final Ilkovič equation may be obtained:

$$i_d = 0.732nFcD^{1/2}m^{2/3}t^{1/6} \quad [6]$$

If the Faraday constant is included in the numerical value the instantaneous diffusion current is

$$i_d = 0.706ncD^{1/2}m^{2/3}t^{1/6} \quad [7]$$

By using the mean surface area during the drop life the mean limiting diffusion current can be derived

$$\bar{i}_d = 0.627nFcD^{1/2}m^{2/3}t^{1/6} \quad [8]$$

### Current–Potential Curves for a Reversible Electrode Reaction

Along the rising part of the polarographic wave the Nernst equation is valid:

$$E = E^0 + \frac{RT}{nF} \ln \left( \frac{[\text{Red}_0]}{[\text{Ox}_0]} \right) \quad [9]$$

where  $E^0$  is the standard redox potential,  $n$  is the number of exchanged electrons, and the subscript 0 denotes the concentrations at the electrode surface. Three different cases should be taken into account depending on the species that are present in the bulk of the solution:

1. only the oxidized form is reduced;
2. only the reduced form is oxidized (also in the case of amalgams); and
3. both the oxidized and the reduced forms are present in the solution.

The resulting curves are denoted as cathodic, anodic, and cathodic–anodic curves, respectively. At the potentials where the surface concentration is nonzero the product must diffuse away from the electrode due to the resulting concentration gradient. The net current is expressed by the concentration of the oxidized species

$$i = -\kappa([\text{Ox}] - [\text{Ox}]_0) \quad [10]$$

or of the reduced species

$$i = -\kappa'([\text{Red}] - [\text{Red}]_0) \quad [11]$$

The parameters are denoted as

$$\kappa = 0.627nFD_{\text{ox}}^{1/2}m^{2/3}t^{1/6} \quad [12]$$

$$\kappa' = 0.627nFD_{\text{red}}^{1/2}m^{2/3}t^{1/6} \quad [13]$$

and are often denoted as the Ilkovič constants. Simple arithmetic operations lead to the equation of a reversible polarographic wave:

$$E = E_0 \mp \frac{RT}{nF} \ln \sqrt{\frac{D_{\text{ox}}}{D_{\text{red}}}} \mp \frac{RT}{nF} \ln \left( \frac{\bar{i}}{\bar{i}_d - i} \right) \quad [14]$$

where minus and plus signs denote cathodic and anodic waves, respectively, and  $i_d$  is again the limiting diffusion current. For the cathodic-anodic wave

$$E = E_0 - \frac{RT}{nF} \ln \sqrt{\frac{D_{\text{ox}}}{D_{\text{red}}}} - \frac{RT}{nF} \ln \left( \frac{\bar{i} - \bar{i}_{\text{da}}}{\bar{i}_{\text{dc}} - \bar{i}} \right) \quad [15]$$

where  $i_{\text{da}}$  and  $i_{\text{dc}}$  are anodic and cathodic limiting currents, respectively. The potential at which the current is  $\bar{i} = 1/2 i_d$  is denoted as the half-wave potential  $E_{1/2}$  and, assuming  $D_{\text{ox}} = D_{\text{red}}$  for a reversible system is equal to  $E^0$ :

$$E = E_{1/2} - \frac{RT}{nF} \ln \left( \frac{\bar{i} - \bar{i}_{\text{da}}}{\bar{i}_{\text{dc}} - \bar{i}} \right) \approx E_{1/2} - \frac{RT}{nF} \ln \left( \frac{\bar{i} - \bar{i}_{\text{da}}}{\bar{i}_{\text{dc}} - \bar{i}} \right) \quad [16]$$

The properties of diffusion controlled currents are as follows:

1. Current at a given potential is linearly proportional to the bulk concentration of the redox form undergoing the electrode process.
2. Current is proportional to the square root of the diffusion coefficient, which may be thus evaluated.
3. Mean limiting diffusion current is proportional to the square root of the mercury height  $h^{1/2}$ .
4. The shape of the  $i$ - $E$  curve determines the dependence of  $\ln\{i/(i_d - i)\}$  versus  $E$ , which should be linear with a slope of  $\pm RT/nF$ . (This so-called log-plot analysis is often used for testing the reversibility of a reaction or for estimating the number of electrons ( $n$ ) involved in a given electrode process.)
5. The half-wave potential  $E_{1/2}$  is independent of the concentration and for  $D_{\text{ox}} = D_{\text{red}}$  equals the standard redox potential  $E^0$  of a corresponding redox couple.

#### Finite Charge-Transfer Rate in Electrode Processes (Irreversible or Slow Waves)

If the electron-transfer rate is not high enough it becomes the controlling factor and the observed electrode processes are called irreversible or slow. The Nernst equation is not obeyed in this situation.

Reversibility is not an absolute property in electrochemistry. It depends on the mutual ratio of the electrochemical time constant and the charge-transfer reaction rate. In direct current polarography the method's time constant is given by the drop time, the time available for the establishment of equilibrium, which can be varied only over a limited interval. Electrode reactions with charge-transfer control



are described by the rate constants  $k_{+e}$  and  $k_{-e}$  for reduction and oxidation, respectively. Both constants are potential dependent:

$$k_{-e} = k^0 \exp \left( -\frac{(1-\alpha)nF}{RT} (E - E^0) \right)$$

$$k_{+e} = k^0 \exp \left( -\frac{\alpha nF}{RT} (E - E^0) \right) \quad [18]$$

where  $k^0$  is the standard heterogeneous rate constant corresponding to the rate at  $E^0$  and is expressed in centimeters per second and  $\alpha$  is the transfer coefficient. At potentials far from  $E^0$  the charge-transfer rate becomes so high that the diffusion rate assumes control and a limiting diffusion current is observed.

The half-wave potential is as follows:

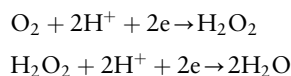
$$E_{1/2} = E^0 + \frac{2.3RT}{\alpha nF} \log \left( 0.886 k^0 \sqrt{\frac{t_1}{D}} \right) \quad [19]$$

The following features of irreversible polarographic waves are relevant:

1. The half-wave potential differs from  $E^0$ , depends on  $t_1$ , and the anodic and cathodic waves of the same redox couple have different  $E_{1/2}$  values.
2. The irreversible polarographic wave is more protracted when compared with a reversible wave.
3. The log-plot analysis does not yield the approximate  $n$  value because the slope also includes the  $\alpha$  transfer coefficient.
4. The upper limit of the heterogeneous charge-transfer rate, distinguishable from the diffusion control, is about  $k^0 \ll 0.02 \text{ cm s}^{-1}$ . Larger values of  $k^0$  must be measured by faster techniques (AC or pulse polarography).
5. Irreversible electrode reactions with the participation of complex ions or of organic compounds cannot be analyzed by means of expressions for reversible compounds: more complicated expressions need to be applied.
6. The  $E_{1/2}$  value may also depend on the concentration and on the type of the indifferent electrolyte.

#### Oxygen Behavior

Usually it is necessary to work in the absence of oxygen, since in the potential working zone of the mercury it is active, the following reduction reactions taking place:



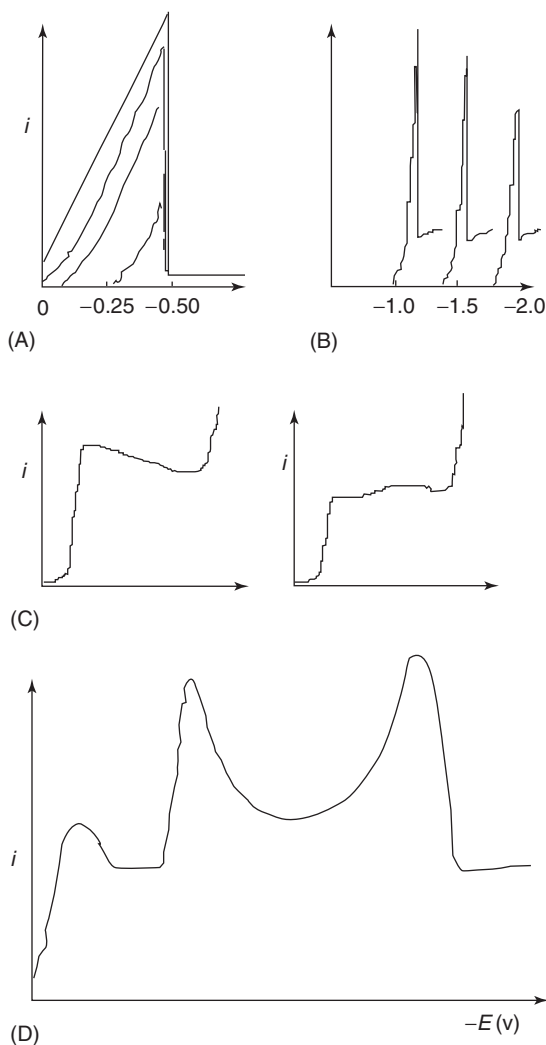
which give two reduction waves for potentials  $\sim -0.1$  and  $-0.9 \text{ V}$  (versus SCE), which may interfere



with the polarographic waves of the analytes to be determined. The nonremoved oxygen traces (a nitrogen flow of sulfite is used) defines, together with the residual current, the detection limit of the technique. These reduction waves of oxygen are, on the other hand, the basis of the electrochemical Clark probe for dissolved oxygen in waters.

### Polarographic Maxima

Certain anomalies related to polarographic curves are known by this name and consist in sharp intensity increases on the rising zone of the curve or on that in which the diffusion limit is attained. Depending on both their morphology and origin they are classified as first, second, and third type maxima. Besides, those corresponding to the first kind can be positive and negative (Figure 3).



**Figure 3** Polarographic maxima: A, positive first kind maxima; B, negative first kind maxima; C, second kind maxima; D, third kind maxima.

The high increase of intensity regarding the different types of maxima is always due to the internal movements of the mercury drop, which in turn induces movements in the adjacent zone of the solution. Consequently, the flow rate of the electroactive species toward the electrode is higher than that which would take place by diffusion.

Polarographic maxima may be eliminated by addition to the solution of tensioactive substances. At present, Triton X-100 is the most widely used compound owing to the higher stability of its solutions with time, together with the fact that since it is a nonionic tensioactive agent, it is, thus, absorbed within a wide range of potentials and is unlikely to chemically interact with other compounds of the solution.

The use of the static mercury electrode enhances the elimination of polarographic maxima, since measurements are carried out with the mercury not flowing.

### Quantification

Regardless of the degree of reversibility of the electrode process, both reversible and irreversible processes can be used for qualitative identification (using  $E_{1/2}$ ), and the quantitative determination based on the Ilkovič equation. If the classical DC polarographic method is applied the lower limit of determination is in the region of  $2 \times 10^{-5} \text{ mol l}^{-1}$ . This sensitivity is not sufficient in comparison with the more advanced analytical methods. The evaluation is carried out by calibration curves or by standard addition.

### Instrumentation

The experimental arrangement in DC polarography is one of the simplest in chemical instrumentation: in essence it consists of the polarizable working (measuring) electrode, the unpolarizable reference electrode, the voltage source, the current measuring device, and the electrolytic vessel. The indicator electrode, i.e., a DME in polarography, is polarized by the voltage source of a low-output resistance (e.g., a ramp generator), which polarizes the electrode with continuously increasing voltage. An unpolarizable electrode with a defined potential (like standard calomel electrode) is used as a reference electrode. This electrode system is immersed into an electrolytic vessel containing the solution to be examined. The current flowing through the electrode system is measured by the current-measuring device and depends on the voltage applied to the working (measuring) electrode.

In recent polarographs, potentiostatic systems with three electrodes are employed (an auxiliary platinum electrode in addition to both the dropping mercury

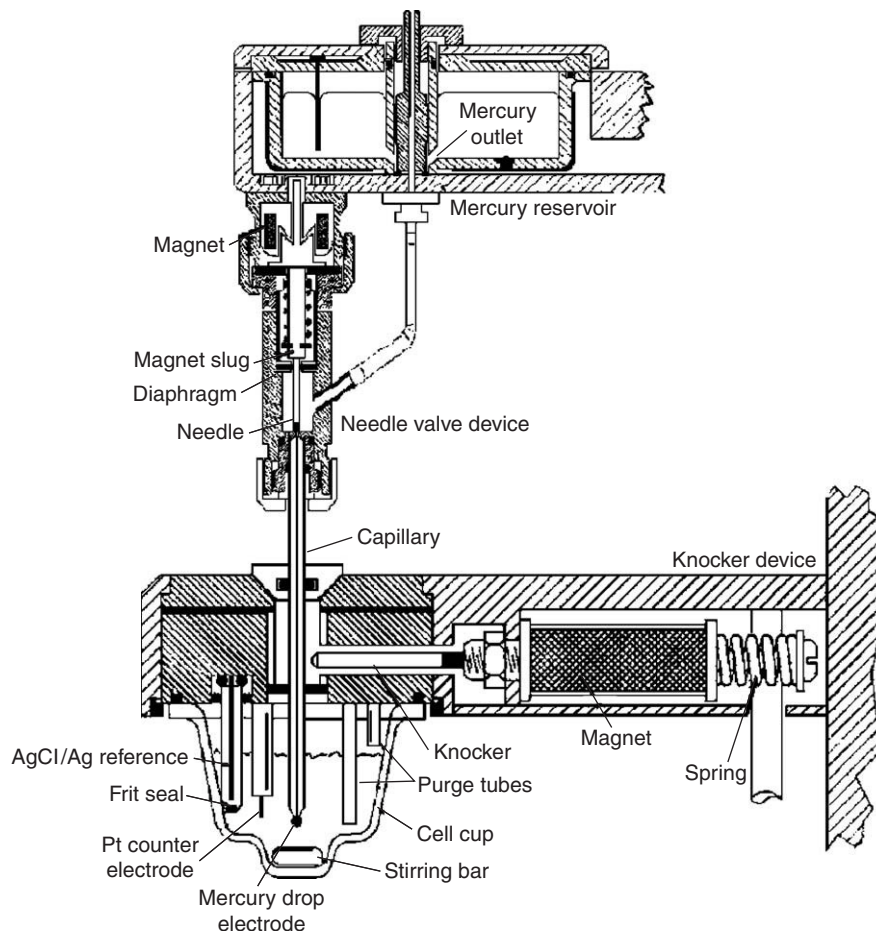
and the reference electrodes) since they allow to compensate the ohmic drop of the solutions when working in nonexclusively aqueous media. With the former system, thus, the distortion of the polarographic wave is avoided.

The recent introduction of digital systems controlled by microprocessor allow the replacement of the systems of mechanical voltage scanning by digital/analog converters, which offer a higher reproducibility. Voltage variations as a function of time are not completely linear any longer, showing a stepping appearance instead, the voltage remaining fixed in DC polarography during the drop life. The knocker's blow implies not only the beginning of the reading synchronism, but it is also used to increase the voltage to be applied in the next drop.

### Mercury Electrode

The DME consists of mercury droplets flowing out under constant pressure from a mercury reservoir-connected capillary (inner diameter 0.05–0.07 mm;

length 10–15 cm). The natural drop time is within the range 3–10 s. The DME is characterized by its drop time  $t_1$  and by the flow rate of mercury  $m$ , which is in the range  $1\text{--}4\text{ mg s}^{-1}$ . Drop-time measurements must take place under reproducible conditions, e.g., the mercury column height, the supporting electrolyte, the same electrode potential. The drop time may be externally controlled by a device (knocker) that periodically dislodges the mercury drop from the tip of the capillary as in sampled (tast) polarography and in advanced methods. Since  $t_1$  and  $m$  depend on column height,  $mt_1$  and the drop size  $A$  are constant and independent of the column height. The size of the drop changes only with changes in the surface tension of mercury. When the surface tension is changed, e.g., by changing the potential, the solution, the surfactant, etc., the drop time changes but the flow rate remains constant. In newly developed capillary types the tip of the capillary has a spindle-shaped inner space; these spindle-shaped capillaries have a highly reproducible drop. An improvement over the classical DME is the static mercury drop



**Figure 4** Modern polarograph, constituted by a potentiostatic system with three electrodes (static mercury, AgCl/Ag reference, and Pt auxiliary), knocker for synchronism and needle valve allowing to control the size of the hanging drop. With permission of AMEL.

electrode in which a needle or other kind of valve controls the mercury flow through the capillary and the drop growth (Figure 4). The current measurement is carried out when the drop area does not change with time. This kind of electrode is used as a hanging mercury drop electrode applied in stripping analysis or as a periodically renewed hanging electrode that may replace the DME. The surface after the drop formation remains constant. The DM is sensitive to mishandling. The capillary should not be left in the solution to be examined when the mercury is not flowing.

The advantages of the static electrode over the dynamic electrode are several. On the one hand, the current measurement is carried out without varying the electrode surface at all, which enhances the decrease of the capacitive current. On the other hand, since the mercury flow is nonexistent the undesirable polarographic maxima tend to disappear; thus, achieving more neat polarograms for the solution, since the solution that is in contact with the electrode remains still. Finally, it is possible to easily obtain voltammograms instead of polarograms, since the whole scanning can be performed with only one mercury drop hanging from the capillary. The drop is renewed in this latter case several times just before the voltage scanning takes place; thus, the advantage of the easiness of reproducibility regarding the mercury electrode remains against solid electrodes, which are more likely to present the memory phenomenon.

### Reference Electrode

The potential of the working electrode is referred to nonpolarizable reference electrode, the potential of

which does not depend on the current flowing through the solution. The current is usually very low so that the last term in the following equation can be neglected:

$$E_{\text{appl}} = E_{\text{WE}} - E_{\text{RE}} + iR \quad [20]$$

See also: **Voltammetry:** Anodic Stripping; Cathodic Stripping.

### Further Reading

- Bard AJ and Faulkner LR (1980) *Electrochemical Methods*. New York: Wiley.
- Bond AM (1980) *Modern Polarographic Methods in Analytical Chemistry*. New York: Dekker.
- Carabias R and Hernández J (2001) Polarografía de corriente continua. In: Blanco M, Cerdà V, and López-Cueto G (eds.) *Métodos Electroanalíticos I*. Palma de Mallorca: Universitat de les Illes Balears.
- Heyrovsky J and Kůta J (1965) *Principles of Polarography*. Prague: Publishing House of the Czechoslovak Academy Sciences.
- Kalvoda R and Volke J (1991) Differential pulse polarography and voltammetry. In: Zýka J (ed.) *Instrumentation in Analytical Chemistry*, vol. 1, p. 39. New York: Ellis Horwood.
- Thomas FG and Henze G (2001) *Introduction to Voltammetric Analysis*. CSIRO Publishing.
- Vlček AA, Volke J, Pospíšil L, and Kalvoda R (1986) Polarography. In: Rossiter BW and Hamilton JE (eds.) *Physical Methods of Chemistry*, vol. II, *Electrochemical Methods*, 2nd edn, ch. 9, p. 39. New York: Wiley.

## Techniques

V Cerdà, University of the Balearic Islands, Palma de Mallorca, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by J. Volke, pp. 4016–4020, © 1995, Elsevier Ltd.

### Introduction

Ever since the development of direct current (DC) polarography, much efforts have been undertaken to compensate for the following three main limitations:

1. The drawback of DC polarography with regards to sensitivity is the current required to change the potential of the dropping mercury

electrode (DME) to the required value: at a depolarizer concentration of  $10^{-5} \text{ mol l}^{-1}$ , this time-averaged charging current is comparable to the Faradaic current.

2. Lack of resolution of the sigmoid-shaped waves (e.g., DC polarography, normal pulse polarography (NPP)). In order to sufficiently distinguish two polarographic waves so as to be able to measure their corresponding half-wave potentials and limiting current values, it is necessary that the former half-wave potentials should differ in more than 200 mV. All subsequent polarographic techniques, in which the obtained response is in the form of a peak (derivative polarography, square-wave polarography (SWP), alternating current (AC) polarography, differential

electrode in which a needle or other kind of valve controls the mercury flow through the capillary and the drop growth (Figure 4). The current measurement is carried out when the drop area does not change with time. This kind of electrode is used as a hanging mercury drop electrode applied in stripping analysis or as a periodically renewed hanging electrode that may replace the DME. The surface after the drop formation remains constant. The DM is sensitive to mishandling. The capillary should not be left in the solution to be examined when the mercury is not flowing.

The advantages of the static electrode over the dynamic electrode are several. On the one hand, the current measurement is carried out without varying the electrode surface at all, which enhances the decrease of the capacitive current. On the other hand, since the mercury flow is nonexistent the undesirable polarographic maxima tend to disappear; thus, achieving more neat polarograms for the solution, since the solution that is in contact with the electrode remains still. Finally, it is possible to easily obtain voltammograms instead of polarograms, since the whole scanning can be performed with only one mercury drop hanging from the capillary. The drop is renewed in this latter case several times just before the voltage scanning takes place; thus, the advantage of the easiness of reproducibility regarding the mercury electrode remains against solid electrodes, which are more likely to present the memory phenomenon.

### Reference Electrode

The potential of the working electrode is referred to nonpolarizable reference electrode, the potential of

which does not depend on the current flowing through the solution. The current is usually very low so that the last term in the following equation can be neglected:

$$E_{\text{appl}} = E_{\text{WE}} - E_{\text{RE}} + iR \quad [20]$$

See also: **Voltammetry:** Anodic Stripping; Cathodic Stripping.

### Further Reading

- Bard AJ and Faulkner LR (1980) *Electrochemical Methods*. New York: Wiley.
- Bond AM (1980) *Modern Polarographic Methods in Analytical Chemistry*. New York: Dekker.
- Carabias R and Hernández J (2001) Polarografía de corriente continua. In: Blanco M, Cerdà V, and López-Cueto G (eds.) *Métodos Electroanalíticos I*. Palma de Mallorca: Universitat de les Illes Balears.
- Heyrovsky J and Kůta J (1965) *Principles of Polarography*. Prague: Publishing House of the Czechoslovak Academy Sciences.
- Kalvoda R and Volke J (1991) Differential pulse polarography and voltammetry. In: Zýka J (ed.) *Instrumentation in Analytical Chemistry*, vol. 1, p. 39. New York: Ellis Horwood.
- Thomas FG and Henze G (2001) *Introduction to Voltammetric Analysis*. CSIRO Publishing.
- Vlček AA, Volke J, Pospíšil L, and Kalvoda R (1986) Polarography. In: Rossiter BW and Hamilton JE (eds.) *Physical Methods of Chemistry*, vol. II, *Electrochemical Methods*, 2nd edn, ch. 9, p. 39. New York: Wiley.

## Techniques

V Cerdà, University of the Balearic Islands, Palma de Mallorca, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by J. Volke, pp. 4016–4020, © 1995, Elsevier Ltd.

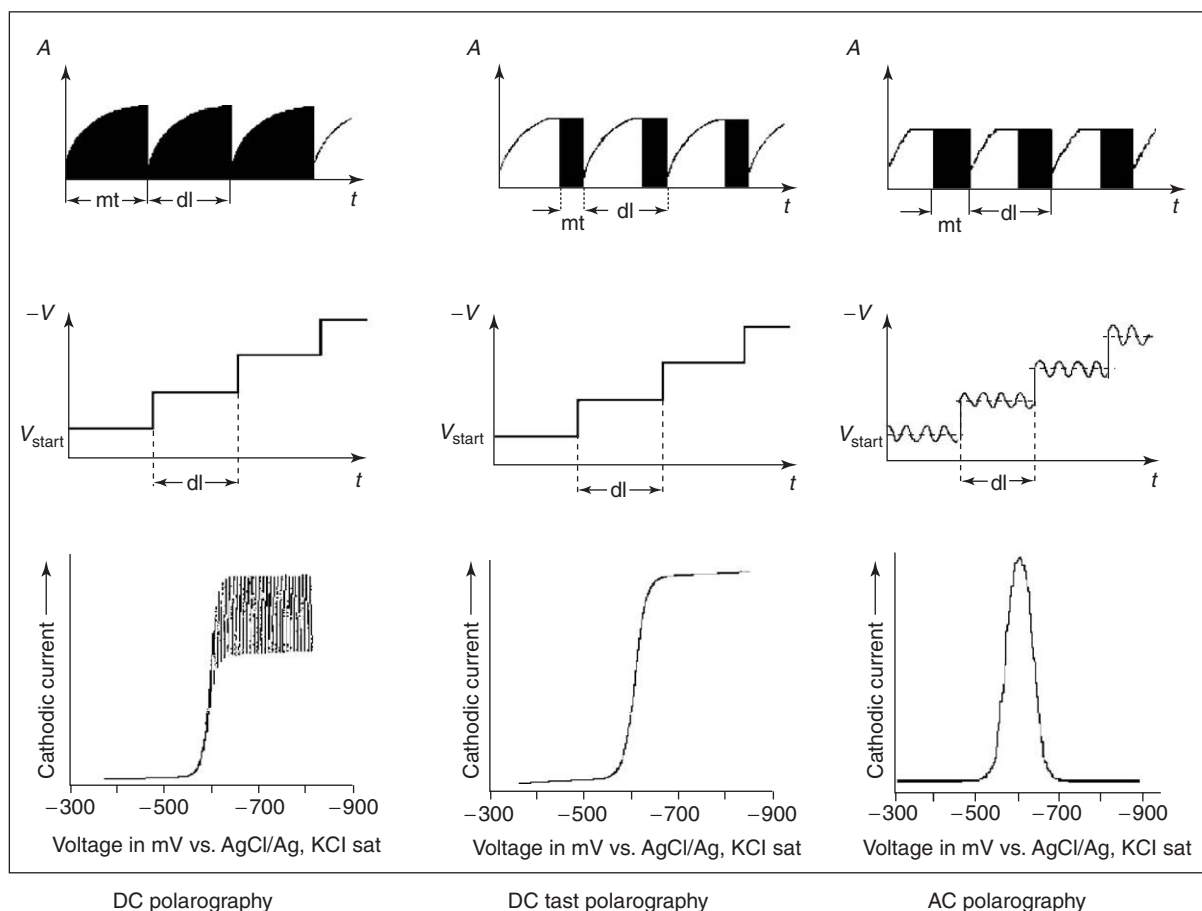
### Introduction

Ever since the development of direct current (DC) polarography, much efforts have been undertaken to compensate for the following three main limitations:

1. The drawback of DC polarography with regards to sensitivity is the current required to change the potential of the dropping mercury

electrode (DME) to the required value: at a depolarizer concentration of  $10^{-5} \text{ mol l}^{-1}$ , this time-averaged charging current is comparable to the Faradaic current.

2. Lack of resolution of the sigmoid-shaped waves (e.g., DC polarography, normal pulse polarography (NPP)). In order to sufficiently distinguish two polarographic waves so as to be able to measure their corresponding half-wave potentials and limiting current values, it is necessary that the former half-wave potentials should differ in more than 200 mV. All subsequent polarographic techniques, in which the obtained response is in the form of a peak (derivative polarography, square-wave polarography (SWP), alternating current (AC) polarography, differential



**Figure 1** Several polarographic techniques obtained using a digital polarograph. The DC technique has been obtained with a dynamic mercury electrode, whereas those of DC fast and AC have been attained with a static electrode and current integration. mt = measuring time, dl = drop life,  $t$  = time,  $A$  = electrode surface,  $V$  = voltage applied to the working electrode. The scanning are cathodic, i.e., toward more negative voltages, giving rise to cathodic currents.

pulse polarography (DPP), etc.) present a higher resolution, a difference of only  $\sim 50$  mV being sufficient to be able to discriminate two substances presenting close half-wave potentials.

3. As observed in **Figure 1** (DC polarography), the noise of the oscillations increases considerably after the reduction of a substance. Thus, making the calculation of the limiting current of any other compound, which is being reduced after the first substance, is difficult, especially when this compound is present at low concentrations. This difficulty decreases when peak-shaped polarograms are attained, for which the signal returns to the baseline.

## Evolution of Polarographic Techniques

In order to solve the above-mentioned drawbacks different solutions have been proposed, although nowadays several can be considered obsolete. In

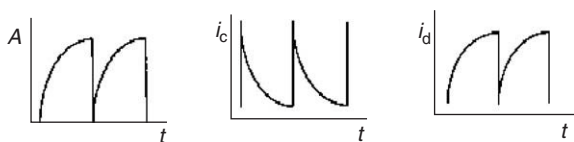
**Table 1** Polarographic techniques

Technique	Resolution	Limit of determination (for lead) ( $\text{mol l}^{-1}$ )
<i>Polarograms with sigmoid shape</i>		
DC	200 mV	$10^{-4}$
Rapid	200 mV	$10^{-4}$
DC fast	200 mV	$10^{-5}$
Differential	50 mV	$10^{-5}$
NPP	200 mV	$10^{-6}$ – $10^{-7}$
<i>Polarograms with peak shape</i>		
Derivative	50 mV	$10^{-5}$
AC	50 mV	$10^{-6}$ – $10^{-7}$
SWP	50 mV	$10^{-7}$ – $10^{-8}$
DPP	50 mV	$10^{-7}$ – $10^{-8}$

DC, direct current; NPP, normal pulse polarography; AC, alternating current; SWP, square wave polarography; DPP, differential pulse polarography.

**Table 1**, an account of the most significant contributions is presented.

One of the first interesting contributions was directed toward the elimination of the noise produced



**Figure 2** Variation of the drop surface, and the intensities of the capacitive ( $i_c$ ) and diffusion ( $i_d$ ) currents with time.

by the capacitive current, which contributes, especially in the beginning, to the growth of the mercury drop. This is due to the fact that for a constant mass flow rate of mercury, the variation rate of the drop's surface is very fast, which is comparable to a condenser whose plates are changing their surface very rapidly and, therefore, the charging current will be initially very high (see **Figure 2**).

The equations giving the intensity variations of both the diffusion and capacitive currents, together with their corresponding ratio, are

$$i_c = km^{2/3}t^{-1/3} \quad [1]$$

$$i_d = k'm^{2/3}t^{1/6} \quad [2]$$

$$i_d/i_c = k''t^{1/2} \quad [3]$$

which reveal that the signal-to-noise ratio improves with time. Thus, in rapid polarography the first 20 ms of the drop's life are neglected in order to subsequently integrate the current during the rest of the drop's life. This option implied the introduction of a synchronization element, the knocker, nowadays present in all modern polarographs: a blow of the knocker releases the previous drop and defines the starting time of the new drop (from which the first 20 ms are discarded before measuring the current). The period of time between two successive blows of the knocker will define the drop's life. Fast polarography can be considered as an evolution of rapid polarography, in which the current measurement is carried out only at the end of the drop's life and where the drop's surface remains almost practically constant with time.

## Alternating Current Polarography

This method is based on superimposition of a small amplitude sine-wave voltage  $E_{\sim}$  over the linear voltage ramp (see **Figure 1**).

The  $E_{\sim}$  frequency ranges from 10 Hz to  $\sim 10$  kHz. The resulting current contains, in addition to the filtered-off DC component, the AC component  $i$ , which is proportional to the total cell admittance

$$Y = Y' + jY'' \quad [4]$$

where  $Y'$  and  $Y''$  are the real and the imaginary admittance components, respectively.

After amplification and rectification, the total AC magnitude was originally recorded as a function of DC voltage. The cell impedance in the simplest case is composed of an ohmic solution resistance  $R_s$  in series with the parallel combination of the electrode double-layer capacity  $C$  and the Faradaic impedance component  $Z_F$ ; the alternating current is a vector with a definite phase in relation to the applied AC perturbation voltage. Measurement of only total alternating current offers no particular advantage during a detailed electrochemical investigation. AC polarography became increasingly attractive after the introduction of more sophisticated instrumentation based on phase-sensitive detection with automatic solution resistance compensation. This procedure consists of measuring the two curves: one current component, also known as the real or the resistive component, which is in phase with  $E_{\sim}$ , and the other, the imaginary or the quadrature component, which is phase-shifted  $90^\circ$  with respect to the  $E_{\sim}$  vector. If the amplitude of  $E_{\sim}$  is precisely known, cell admittance  $Y(\omega)^0$  or impedance  $Z(\omega)$  at a given frequency  $\omega$  may be calculated as

$$i_{\sim} = \frac{E_{\sim}}{Z(\omega)} = Y(\omega)E_{\sim} \quad [5]$$

This information may be analogously obtained by finding the absolute value of the AC vector and the phase shift of  $i_{\sim}$  in reference to  $E$ . Appropriate  $R_s$  compensation is crucial in phase-sensitive AC polarography because the  $IR$  drop causes a phase error that is eliminated only after laborious vector analysis. The in-phase component of a phase-sensitive AC polarogram of a blank solution without any electroactive compounds should be a zero line since

$$Y'\omega = \frac{1}{R_s} \rightarrow 0 \quad [6]$$

and the double-layer capacitance for the imaginary component

$$Y''(\omega) = \omega C(E) \quad [7]$$

The double-layer structure and adsorption of various surface-active substances may be examined by this method, which replaces cumbersome balancing bridge measurements.

The Faradaic peak-shaped contributions in both real and imaginary AC curves, denoted as  $Y'_F$  and  $Y''_F$ , respectively, are found in the potential region of the polarographic wave. A general case with a finite charge-transfer rate relates these Faradaic components



to the kinetic parameters by

$$Y_F' = nFA \frac{\xi + 1}{2\xi^2 + 2\xi + 1} \left[ c_{\text{Red}}' \frac{\delta k_{+e}}{\delta E} - c_{\text{Ox}}' \frac{\delta k_{-e}}{\delta E} \right] \quad [8]$$

$$Y_F'' = nFA \frac{\xi}{2\xi^2 + 2\xi + 1} \left[ c_{\text{Red}}'' \frac{\delta k_{+e}}{\delta E} - c_{\text{Ox}}'' \frac{\delta k_{-e}}{\delta E} \right] \quad [9]$$

where  $c'$  is surface concentration and

$$\xi = \frac{k_{+e}}{\sqrt{2\omega D_{\text{Ox}}}} + \frac{k_{-e}}{\sqrt{2\omega D_{\text{Red}}}} \quad [10]$$

Both Faradaic components are proportional to bulk concentration and  $Y_F''$  is a signal completely without charging components, with a zero line ideally suited to analytical applications. Furthermore, phase-sensitive AC polarography gives instantaneous information about reversibility at a given  $\omega$  by comparing the ratio of  $Y_F'$  and  $Y_F''$ .

A reversible electrode reaction corresponds to the limit  $\xi \rightarrow 0$ , hence  $Y_F' = Y_F''$ . In contrast, the totally irreversible process is characterized by  $\xi \rightarrow 0$  and  $Y_F'' \rightarrow 0$ . This AC technique also determines the double-layer capacity in the presence of a quasireversible electrode reaction by utilizing sufficiently high frequency, since  $\omega \rightarrow \infty$ , and hence  $Y' \rightarrow 0$  and  $Y'' \rightarrow \omega c$ .

The kinetic parameters  $k^0$  and  $\alpha$  are evaluated by plotting  $\xi$  versus  $E$ :

$$\xi = \frac{Y_F''}{Y_F' - Y_F''} = \frac{k_{+e}}{\sqrt{2\omega D_{\text{Ox}}}} = \frac{k_{-e}}{\sqrt{2\omega D_{\text{Red}}}} \quad [11]$$

yielding two linear asymptotes at far positive and negative potentials with respect to  $E_{1/2}$ .

The transfer coefficient is calculated from the slope and  $k^0$  and  $E^0$  from the coordinates of the asymptotes' intercept. The appropriate method for the determination is based on the Faradaic phase angle

$$\tan \varphi = \frac{Y_F''}{Y_F'} \quad [12]$$

which is plotted versus  $\sqrt{\omega}$  for the  $k^0$  determination. The potential of the maximum on a cotan versus  $E$  plot yields the  $\alpha$  value. Admittance or impedance data displayed as complex plane plots are often used.

Electroactive species adsorption is sensitively indicated by this method, which thus determines whether redox system properties are obscured by a specific type of interaction with the electrode. Strongly absorbable electroinactive compounds yield peaks on the AC curves located at potentials of the adsorption-desorption process. However, they are much narrower and their frequency dependence differs markedly from the Faradaic peaks. Such peaks were also used for determination of surface-active

compounds. In general, AC polarography is not very sensitive, in particular in irreversible processes without adsorption.

## Pulse Polarography

In pulse methods, the procedures are based on the application of pulse changes of potential and the current response is measured at a suitable time relative to the time of the pulse. The concept includes three methods: NPP, DPP, and SWP.

Pulse techniques improve detection limits since they benefit from the different variation of diffusion and capacitive current intensities with time: when carrying out measurements at the pulse end, the capacitive current is practicably negligible, the value of the Faradaic current still being significant (see eqn [3] and Figure 3).

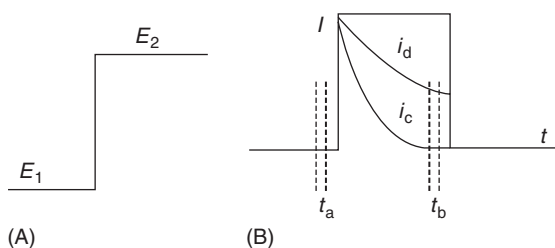
NPP keeps the DME at a constant potential before the start of the Faradaic current; then, almost at the end of the drop's life, a voltage pulse is applied, whose amplitude gradually increases from drop to drop. The duration of the pulse is  $\sim 200$  ms. The current is usually measured shortly and before the pulse end ( $\sim 20$  ms after the pulse edge) and is recorded versus the pulse amplitude. The current-potential relationship for a reversible process is

$$i = nFcA \sqrt{\frac{D}{\pi\tau} \left( \frac{1}{1+P} \right)} \quad [13]$$

where  $P = \exp(nF/RT)(E - E_{1/2})$  and  $\tau$  is the pulse duration. As the pulse voltage becomes more negative than  $E_{1/2}$ , the value of  $P$  approaches zero, yielding a limiting current given by the Cottrell equation:

$$i = nFcA \sqrt{\frac{D}{\pi\tau}} \quad [14]$$

This mode is about seven times more sensitive than classical DC polarography.



**Figure 3** (A) Sudden variation of the potential, (B) behavior of the diffusion ( $i_d$ ), and capacitive ( $i_c$ ) currents with time.  $t_a$  and  $t_b$  are the times used in the measurement of the current.

The shape of the curve is that of a normal DC polarogram. In irreversible processes, NPP may also determine the kinetic parameter  $k^0$ . Only the upper limit for  $k^0$  is higher than in DC polarography since the time is now of the order of milliseconds in contrast with the drop time  $t_1$  in DC polarography.

DPP (Figure 4) differs from NPP when prior to the pulse application the potential is not constant but is replaced by a ramp voltage with DC polarographic characteristics. The pulse amplitude is constant and the measured current is displayed as the difference between the current sampled closely before pulse application and the current sampled at the pulse end.

If the polarographic  $i$ - $E$  equation is now differentiated as

$$i = i_d \frac{1}{1 + P} \quad [15]$$

and the Cottrell equation is substituted for  $i_d$ , the relationship for the DPP with a reversible process is

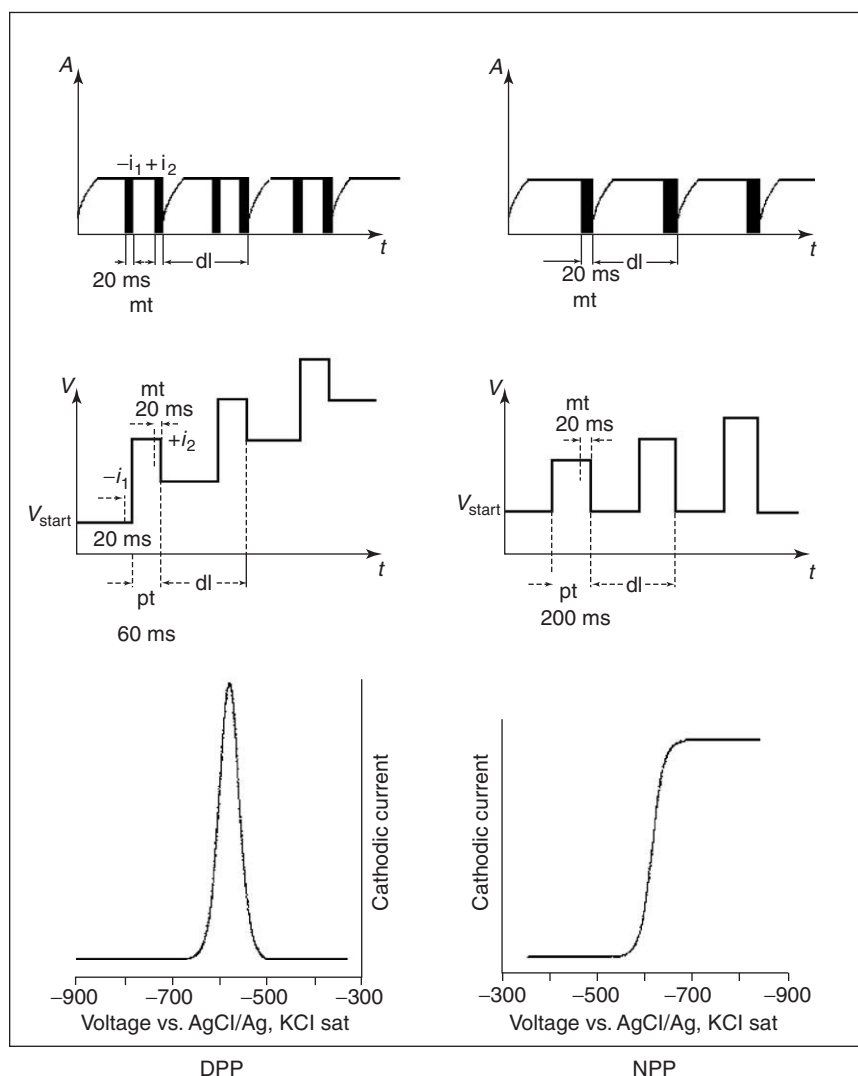
$$i = \frac{n^2 F^2}{RT} A c \Delta E \sqrt{\frac{D}{\pi \tau}} \cdot \frac{P}{(1 + P)^2} \quad [16]$$

This equation is only valid for small values of pulse amplitude  $E < RT/nF$ . The  $i$ - $E$  curve is a derivative of the classical DC polarographic wave. The peak current is

$$i = \frac{n^2 F^2}{4RT} A c \Delta E \sqrt{\frac{D}{\pi \tau}} \quad [17]$$

The peak potential is related to  $E_{1/2}$  by

$$E_p = E_{1/2} - \frac{\Delta E}{2} \quad [18]$$



**Figure 4** DPP and NPP polarographies. Variation of the mercury drop area and voltage with time using the static mercury electrode. dl = drop life, mt = measuring time; pt = pulse time; A = electrode area,  $t$  = time;  $V$  = working electrode voltage.

The solution for  $i$ - $E$  curves is also available for a slow electrode reaction or for considering electrode sphericity.

In SWP, the polarization potential is a linear varying voltage ramp over which a small amplitude square-wave signal, usually 225 Hz and 1–50 mV amplitude, is superimposed. This differs from NPP and DPP in that perturbation continues during the whole life of the drop and the response readout is positioned at a time of negligible time change of the electrode surface. The signal, measured as a function of the linear varying potential, is the difference of current samples obtained at the end of each half-cycle of the square-wave signal; the DC signal component is filtered off. The resulting  $i$ - $E$  curve is a peak described for a reversible reaction by the equation

$$i = \frac{n^2 F^2}{RT} c \Delta E \sqrt{\frac{D}{\pi \tau}} \cdot \frac{P}{(1+P)^2} L \quad [19]$$

In this equation,  $L$  is the constant of the instrument defined as

$$L = \sum_{m=0}^{\infty} \frac{(-1)^m}{(m + t/\tau)^{1/2}} \quad [20]$$

$\Delta E$  is the square wave amplitude,  $t$  is the time measured from the beginning of the half-cycle and is the half-period of the square wave. The peak half-width,  $W_{1/2}$ , for small  $\Delta E$  is given by

$$W_{1/2} = 3.52 \frac{RT}{nF} = \frac{90.5}{n} \text{ mV at } 25^\circ\text{C} \quad [21]$$

The theory of slow charge-transfer reactions reveals that the peak height strongly decreases in the range of rate constants  $k^0 = 10^{-1}$ – $10^{-4} \text{ cm s}^{-1}$ ; below this value it is almost completely insensitive toward the  $k^0$  change and yields peaks  $\sim 20$  times smaller than those in the reversible case.

See also: **Polarography:** Overview; Inorganic Applications; Organic Applications.

## Further Reading

- Bard AJ and Faulkner LR (1980) *Electrochemical Methods*. New York: Wiley.
- Bond AM (1984) *Modern Polarographic Methods in Analytical Chemistry*. New York: Dekker.
- Esteban M and Casassas E (2001) Técnicas polarográficas Modernas. In: Blanco M, Cerdà V, and López-Cueto G (eds.) *Métodos Electroanalíticos I*. Palma de Mallorca: Universitat de les Illes Balears.
- Henze G and Neeb R (1986) *Elektrochemische Analytik*. Berlin: Springer.
- Heyrovsky J and Kůta J (1965) *Principles of Polarography*. Prague: Publishing House of the Czechoslovak Academy Sciences.
- Kalvoda R and Volke J (1991) Differential pulse polarography and voltammetry. In: Zyka J (ed.) *Instrumentation in Analytical Chemistry*, vol. 1, p. 39. New York: Ellis Horwood.
- Vlček AA, Volke J, Pospíšil L, and Kalvoda R (1986) Polarography. In: Rossiter BW and Hamilton JE (eds.) *Physical Methods of Chemistry*, vol. II, *Electrochemical Methods*, 2nd edn, ch. 9, p. 39. New York: Wiley.
- Volke J and Liška F (1994) *Electrochemistry in Organic Synthesis*. Berlin: Springer.

## Inorganic Applications

**M Nimmo**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Polarography is a voltammetric technique in which chemical species (ions or molecules) undergo oxidation (lose electrons) or reduction (gain electrons) at the surface of a dropping mercury electrode (DME) at an applied potential. Polarography only applies to the DME. The resultant reduction or oxidation current is recorded against the applied potential to yield the analytical parameter. Polarography may be used simply as an analytical tool or as a diagnostic method to interpret mechanisms and rates of electrode

chemical reactions. Electrode surface redox reactions may occur with both inorganic and organic chemical species, although this article will deal solely with the inorganic applications of polarography.

Polarography was first introduced in 1922 by Heyrovsky; however, by the mid-1950s its usefulness as an analytical technique was superseded by spectroscopic techniques such as atomic absorption spectroscopy (AAS). For example their popularity for dissolved trace metal analysis was due to their greater sensitivities (ca.  $<0.1 \text{ mmol l}^{-1}$  compared to typically  $>0.1 \text{ mmol l}^{-1}$  for polarography), elemental diversity, flexibility, and ease of use. However, during the 1970s and 1980s polarographic techniques were more widely adopted owing to advances in instrumental design, particularly in enhancement in

The solution for  $i$ - $E$  curves is also available for a slow electrode reaction or for considering electrode sphericity.

In SWP, the polarization potential is a linear varying voltage ramp over which a small amplitude square-wave signal, usually 225 Hz and 1–50 mV amplitude, is superimposed. This differs from NPP and DPP in that perturbation continues during the whole life of the drop and the response readout is positioned at a time of negligible time change of the electrode surface. The signal, measured as a function of the linear varying potential, is the difference of current samples obtained at the end of each half-cycle of the square-wave signal; the DC signal component is filtered off. The resulting  $i$ - $E$  curve is a peak described for a reversible reaction by the equation

$$i = \frac{n^2 F^2}{RT} c \Delta E \sqrt{\frac{D}{\pi \tau}} \cdot \frac{P}{(1+P)^2} L \quad [19]$$

In this equation,  $L$  is the constant of the instrument defined as

$$L = \sum_{m=0}^{\infty} \frac{(-1)^m}{(m + t/\tau)^{1/2}} \quad [20]$$

$\Delta E$  is the square wave amplitude,  $t$  is the time measured from the beginning of the half-cycle and is the half-period of the square wave. The peak half-width,  $W_{1/2}$ , for small  $\Delta E$  is given by

$$W_{1/2} = 3.52 \frac{RT}{nF} = \frac{90.5}{n} \text{ mV at } 25^\circ\text{C} \quad [21]$$

The theory of slow charge-transfer reactions reveals that the peak height strongly decreases in the range of rate constants  $k^0 = 10^{-1}$ – $10^{-4} \text{ cm s}^{-1}$ ; below this value it is almost completely insensitive toward the  $k^0$  change and yields peaks  $\sim 20$  times smaller than those in the reversible case.

See also: **Polarography:** Overview; Inorganic Applications; Organic Applications.

## Further Reading

- Bard AJ and Faulkner LR (1980) *Electrochemical Methods*. New York: Wiley.
- Bond AM (1984) *Modern Polarographic Methods in Analytical Chemistry*. New York: Dekker.
- Esteban M and Casassas E (2001) Técnicas polarográficas Modernas. In: Blanco M, Cerdà V, and López-Cueto G (eds.) *Métodos Electroanalíticos I*. Palma de Mallorca: Universitat de les Illes Balears.
- Henze G and Neeb R (1986) *Elektrochemische Analytik*. Berlin: Springer.
- Heyrovsky J and Kůta J (1965) *Principles of Polarography*. Prague: Publishing House of the Czechoslovak Academy Sciences.
- Kalvoda R and Volke J (1991) Differential pulse polarography and voltammetry. In: Zyka J (ed.) *Instrumentation in Analytical Chemistry*, vol. 1, p. 39. New York: Ellis Horwood.
- Vlček AA, Volke J, Pospíšil L, and Kalvoda R (1986) Polarography. In: Rossiter BW and Hamilton JE (eds.) *Physical Methods of Chemistry*, vol. II, *Electrochemical Methods*, 2nd edn, ch. 9, p. 39. New York: Wiley.
- Volke J and Liška F (1994) *Electrochemistry in Organic Synthesis*. Berlin: Springer.

## Inorganic Applications

**M Nimmo**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Polarography is a voltammetric technique in which chemical species (ions or molecules) undergo oxidation (lose electrons) or reduction (gain electrons) at the surface of a dropping mercury electrode (DME) at an applied potential. Polarography only applies to the DME. The resultant reduction or oxidation current is recorded against the applied potential to yield the analytical parameter. Polarography may be used simply as an analytical tool or as a diagnostic method to interpret mechanisms and rates of electrode

chemical reactions. Electrode surface redox reactions may occur with both inorganic and organic chemical species, although this article will deal solely with the inorganic applications of polarography.

Polarography was first introduced in 1922 by Heyrovsky; however, by the mid-1950s its usefulness as an analytical technique was superseded by spectroscopic techniques such as atomic absorption spectroscopy (AAS). For example their popularity for dissolved trace metal analysis was due to their greater sensitivities (ca.  $<0.1 \text{ mmol l}^{-1}$  compared to typically  $>0.1 \text{ mmol l}^{-1}$  for polarography), elemental diversity, flexibility, and ease of use. However, during the 1970s and 1980s polarographic techniques were more widely adopted owing to advances in instrumental design, particularly in enhancement in

solid-state electronic circuitry, which led to enhanced sensitivities. In addition, the applications of polarography widened, along with the technique's relatively low capital and running costs. More recently advances in the technique may be attributed to the design of the DME and interfacing with information technology (enabling quicker analyses; the electrode has minimum interference, consuming very little mercury and is sealed and simple to handle, as well as economical to use).

Polarographic analysis has been applied to a wide variety of samples including natural waters, biological, mineral, and environmental samples. A summary of the applications is shown in **Table 1**. In addition, **Table 1** illustrates the typical detection limits encountered for each element, although this will vary depending upon the sample solution conditions. All the applications quoted in **Table 1** are for differential pulse polarography (DPP; discussed later). Lower detection limits may be achieved using techniques

such as stripping voltammetry. Stripping voltammetry, both anodic and cathodic, has advanced over the last decade enabling the detection of trace levels of organic and inorganic chemical species in aqueous solutions (typical limits of detection  $<10^{-9} \text{ mol l}^{-1}$ ), surpassing the detection limits of polarography for certain analyte determinations, and indeed advanced spectroscopic techniques.

Essentially, direct current (DC) polarography yields detection limits in the orders of  $10^{-5}$ – $10^{-6} \text{ mol l}^{-1}$ . However, if the applied potential scan is either in the DPP or the alternating current (AC) mode then the detection limits may be enhanced to  $10^{-7}$ – $10^{-8} \text{ mol l}^{-1}$ , widening their analytical applications.

The major advantages using polarography for inorganic analysis may be summarized: (1) comparatively inexpensive equipment is required, (2) ability of the technique to distinguish between elemental oxidation states (i.e., Cr, As), (3) ability of the technique to establish the chemical form of elements

**Table 1** Summary of DP polarographic inorganic applications

<i>Element</i>	<i>Application</i>	<i>DL (<math>\mu\text{mol l}^{-1}</math>)</i>	<i>Comments<sup>a</sup></i>
Ti	j, k, o	0.05	
V	m, i	0.6	Catalytic enhancement of sensitivity
Cr	e, c, i, j, o, r	1	Discrimination between Cr(III) and Cr(VI)
Mo	d, e, j, k, o, p	0.1	Catalytic enhancement of sensitivity
W	k, o	5	
Mn	a, b, h, o, p	0.1	
Fe	c, b, h, i, j, k, n, o, p	1	Discrimination between Fe(III) and Fe(II) possible
Co	c, g, i, j, o, p	0.1	Adsorptive process
Ni	c, g, i, k, m, o, p	0.1	Adsorptive process
Cu	a, g, h, i, j, l, m, n, o, p, r	0.1	
Zn	g, h, i, j, o, p	0.5	
Cd	a, c, g, h, i, j, k, l, p	0.01	
Al	i, o	0.1	Derivatization followed by adsorption onto the DME
In		1	Determination after complex preconcentration
CN <sup>-</sup>	c	0.04	
Si	b, d,	0.1	Derivatization followed by adsorption
Sn	a, c, g, i, o	1	
Pb	a, c, g, h, i, j, l, m, o, p	0.1	
NO <sub>3</sub> <sup>-</sup>	a, b, c, e, j, p	1	
NO <sub>2</sub> <sup>-</sup>	a, b, c, e, j, p	0.1	Derivatization followed by adsorption
P (as PO <sub>4</sub> <sup>3-</sup> )	a, b, c, d, e	0.05	Derivatization followed by adsorption
As	c, h, n, o	0.1	Differentiation between As(V) and As(III) possible
Sb	a, p	0.1	Differentiation between Sb(V) and Sb(III) possible
Bi	o	0.5	
O <sub>2</sub>	a, b, c, d, e, f		
S	b, c, h, i, l, m, s		
Se	a	0.01	Catalytic enhancement of sensitivity
Br (as BrO <sub>4</sub> <sup>-</sup> )	c, f	1	
Cl (ClO <sub>4</sub> <sup>-</sup> )	c, f	1000	Catalytic enhancement of sensitivity
I (as IO <sub>4</sub> <sup>-</sup> )	a, d	0.02	
Ce, Pr, Nd, Sm, Eu		0.1	
U(VI)	a, c, k, o	1.2	After adsorption of ion-associated complex

<sup>a</sup> See text.

a, biological fluids; b, freshwater; c, wastewater; d, seawater; DL, detection limit; e, drinking water; f, groundwater; g, foods; h, beverages; i, ceramics and glass; j, soil; k, minerals; l, aerosol; m, fuel and mineral oils; n, electronics; o, alloys, steel, bronzes; p, plant material; q, sediments; r, cements.

(e.g., characterization and quantification of metal-organic complexation in natural waters), (4) simultaneous elemental analysis (e.g., Cu, Pb, Cd, Zn), and (5) sensitivity (down to  $10^{-8} \text{ mol l}^{-1}$ ).

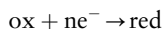
As with all analytical techniques polarography has a number of limitations; these are (1) technique is not sensitive enough for direct ultratrace elemental analysis (in these circumstances the use of stripping voltammetry may be considered), (2) long analysis times (minutes compared to seconds for other analytical techniques such as AAS, ICP-AES) are required, (3) operator experience more critical (understanding of electrode processes is essential), in data interpretation, (4) limited elemental range owing to the constraints of the voltage window of the DME, and (5) toxic nature of mercury has recently limited its use.

However, the present discussion is to highlight the extensive capabilities of polarography for inorganic analysis. The discussion is by no means complete in both analytes detectable and sample application. However, it does serve to indicate the range of elements, which may be determined, along with the approximate detection limits for each application.

## Basic Concepts

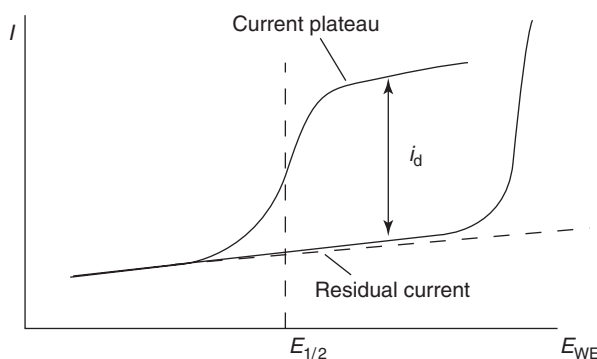
Polarography is just one of many electrochemical techniques available to the analytical chemist. Polarography is restricted to applications using the DME. This is termed the working electrode. The DME guaranteed its popularity amongst electrochemists because (1) the electrode surface is continuously renewed and (2) it has a large hydrogen overpotential leading to a wide electrode potential window.

During the determination a changing negative voltage (cathodic) is applied at the electrode, reducing the analyte of interest (or conversely oxidizing the analyte if an anodic potential scan is applied at the electrode surface):



Under such conditions transfer of the analyte from the bulk solution to the electrode surface is principally by diffusion. A typical DC polarogram is presented in **Figure 1**. The half wave potential ( $E_{1/2}$ ) is characteristic of the analyte and hence yields qualitative analytical data. The analyte reduction current, produced at the electrode surface, is the limiting diffusion current ( $i_d$ ) which is proportional to the concentration of the analyte in the solution, hence providing the quantitative information. Calibration is performed by internal standard additions or the construction of a calibration curve.

Ilkovic defined the theoretical relationship between the analyte concentration in the bulk solution



**Figure 1** Typical polarogram of the reduction of an electro-active species.

to the diffusion current such that,

$$i_d = 607nD_{\text{ox}}^{1/2}m^{2/3}t^{1/6}c_{\text{ox}}$$

where  $i_d$  is the diffusion-controlled current (A),  $n$  the number of Faradays of electricity per molar unit of electrode reaction,  $D_{\text{ox}}$  the diffusion coefficient of analyte ( $\text{cm}^2 \text{ s}^{-1}$ ),  $m$  the flow rate of mercury through DME ( $\text{mg s}^{-1}$ ),  $t$  the drop time (s), and  $c_{\text{ox}}$  the concentration of analyte ( $\text{mol l}^{-1}$ ).

Knowledge of whether an electrode process is reversible or irreversible is crucial in interpreting polarograms. The reversibility of the electrode reaction depends upon the rate of reaction of the following stages in an electrode process (1) the charge transfer step, (2) the chemical reaction, (3) diffusion of analyte to the electrode, and (4) adsorption processes.

For a more extensive treatment of electrode reversibility and the theory of polarography the reader can refer to the Further Reading section.

## Instrumentation

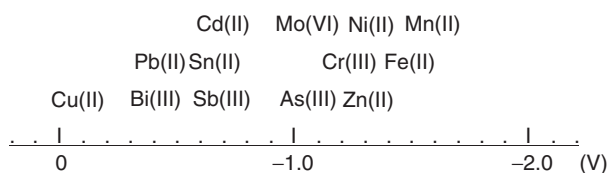
Polarographic analysis requires a three-electrode circuit containing the working electrode (DME), counter-electrode (usually a platinum wire or graphite rod) and finally the reference electrode, generally being either a standard calomel electrode (SCE) or a silver/silver chloride reference electrode.

The three-electrode system is connected to the polarographic analyzer which maintains the working electrode potential using a potentiostat which also controls the potential scan applied at the DME whilst monitoring reduction/oxidizing currents.

## Limitations of Polarography

A chemical species will be responsive at the DME provided that it undergoes oxidation or reduction at a potential that lies within the voltage window of a





**Figure 2** Typical reduction potentials (vs. SCE) of a number of metallic species within the potential window of a DME (acetate buffer, pH=5).

DME (an analyte which is not responsive within these limits may be derivatized to a chemical species which is, discussed later). The limits of the voltage window are set by (1) the oxidation of mercury ( $>0.1$  V – SCE) and (2) the reduction of the hydronium ion ( $-0.8$  to  $-2.0$  V depending on the solution pH, increasing the pH will shift the reduction potential cathodically).

**Figure 2** illustrates just some of the analytes whose redox potential lie within the voltage window of a DME.

In addition dissolved oxygen is electroactive and will give rise to two reduction waves, potentially masking the analyte of interest, therefore oxygen must be stripped from solution before analysis can proceed, usually by purging the sample with argon or pure nitrogen.

Inorganic applications are generally carried out in aqueous solution, which may require the addition of a supporting electrolyte (e.g.,  $0.1 \text{ mol l}^{-1}$  KCl) to lower the solution resistance. However, care must be taken to use the highest purity of reagent available to minimize sample contamination.

Careful choice of supporting electrolyte has to be made owing to possible complexation of the analyte with the components of the supporting electrolyte, which may lead to a shift in the reduction potential. If there is a lowering of the free energy, resulting from complexation, then the  $E_{1/2}$  will shift to more negative potentials. This may be advantageous, as interference from overlapping analyte reduction peaks may thus be eliminated or reduced (see the section ‘Applications’; Ni and Co). Complex formation constants can also be evaluated from this shift in the  $E_{1/2}$ .

The solution pH may significantly alter the speciation and hence the electroactive behavior of analytes involved in acid/base reactions. For example, a protonated form may be electroactive whereas another form of the same analyte may not, or may appear at a different reduction potential. Therefore, analyte solutions need to be buffered at an optimum pH during analysis, to maintain a constant analyte current response.

**Table 2** Examples of electrolytes used in inorganic polarographic analysis

$0.1 \text{ mol l}^{-1}$ KCl
$0.1 \text{ mol l}^{-1}$ $\text{KNO}_3$
$0.1 \text{ mol l}^{-1}$ $\text{NH}_3/\text{NH}_4\text{Cl}$
$0.1 \text{ mol l}^{-1}$ KCNS
$0.1 \text{ mol l}^{-1}$ EDTA (pH = 7)
$1 \text{ mol l}^{-1}$ HCl
$1 \text{ mol l}^{-1}$ NaOH/KOH
$0.2 \text{ mol l}^{-1}$ tartrate/ $\text{NH}_3$ (pH = 9)
$0.1 \text{ mol l}^{-1}$ citrate/ $\text{NH}_3$ (pH = 3)
$0.25 \text{ mol l}^{-1}$ oxalic acid + $(\text{NH}_4)_2$ -oxalate (pH = 4)

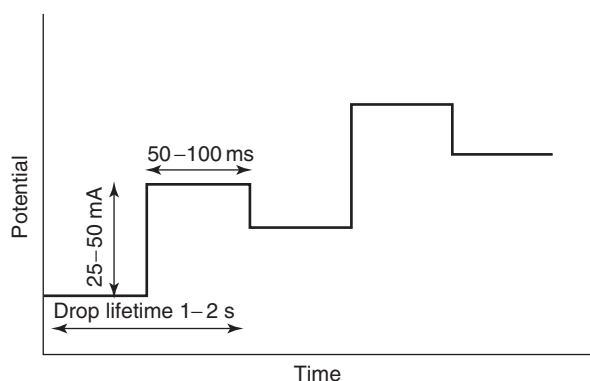
**Table 2** illustrates some of the reported supporting electrolytes used in inorganic polarographic analysis.

## Mode of Potential Scans

During DC polarographic analysis a linearly changing potential with time is applied at the working electrode. Historically DC was the first mode of potential scan, but its downfall was its limited sensitivity (detection limits;  $\sim 10^{-5} \text{ mol l}^{-1}$ ). At such concentrations the charging or capacitance current ( $i_c$ ) is similar to the faradaic current ( $i_f$ ) of the current resulting from the electrode surface redox reaction. Even sampling the current at the end of the drop lifetime only increases the sensitivity slightly due to the depletion of analyte at the surface of the drop.

This limitation led to the development of normal pulse polarography (NPP) introduced by G.C. Barker, which minimized background responses (principally the capacitance current of the growing mercury drop) and maximized the analytical response. NPP requires the application of a series of potential pulses at the working electrode for each drop. The pulse increases in potential with every drop. The current is sampled just before the end of the drops lifetime, minimizing the effect of the capacitance current on the faradaic current. As the potential pulse for each drop starts at a potential lower than the redox potential, no depletion of the analyte would have occurred.

The limit of detection may be lowered further by one or two orders of magnitude using this potential scan mode. Improvement on NPP has occurred with the development of DPP. The difference between NPP and DPP is that the pulse does not return to a constant potential, instead the applied pulses (amplitude being typically 25–50 mV) are superimposed on a linear voltage profile (**Figure 3**), for between 50 and 100 ms. The pulse is applied toward the end of the drop's lifetime with the analyte redox current being measured at the DME before and after the pulse. This differential current is then plotted against the average potential to obtain the differential pulse



**Figure 3** Diagram to illustrate the characteristics of a differential pulse potential scan.

polarogram. Generally, a peak-shaped output is obtained (where  $E_p$  is similar to  $E_{1/2}$ ) during the analytical stage. The peak current is proportional to the electroactive analyte in solution. Limits of detection with DPP are of the order  $10^{-8}$ – $10^{-9}$ .

Another form of applied potential scan involves the superimposition of an AC voltage on the applied potential ramp, which has been shown to enhance sensitivity over DPP. For a more comprehensive description of pulse and AC techniques describing subtle variations, the reader can refer to the Further reading section.

All applications described in the present article use DPP unless otherwise stated.

## Applications

Principal inorganic components detected by polarography are metallic species and certain anions. These applications will be described with respect to periodic elemental groupings. Determination of such analytes have been carried out in a wide variety of sample types such as water (waste and potable), foods (drinks, oils, meat, fruits, vegetables, and cereals), mineral ores, metallurgy, biological fluids (blood, urine), and environmental samples (sediments, soils, aerosols, and natural waters).

The aim of the present article is to highlight which elements are detectable by polarography and the typical detection limits for the applied solution conditions. Enhancement of sensitivity by adsorption (leading to greater analyte concentration at the DME surface, e.g., Si), derivatization (formation of a different electroactive species, e.g.,  $\text{NO}_3^-$ ), catalytic processes (where the product of the electrode process chemically reacts with a nonelectroactive species to regenerate the analyte, e.g., Mo), or combinations of these are illustrated by examples. Where applicable interfering species will be mentioned.

## Groups IA and IIA

Due to the very negative reduction potentials of the elements in these two groups (e.g.,  $E_{\text{red}}$  for potassium lies at around  $-2.9$  V versus SCE) polarographic analysis is not possible in acidic aqueous solutions due to interference by the hydrogen reduction current. In alkaline conditions there is a negative shift of the hydrogen reduction potential and a positive shift in the metal-amalgam reduction potential such that determination is possible. The supporting electrolyte should not react at these extreme potentials; the tetraalkyl ammonium halides generally being the chosen supporting electrolytes. Major cations ( $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ , and  $\text{Mg}^{2+}$ ) in seawater have been determined by polarography following chemical pretreatment of the sample. However, flame emission spectroscopy would be the preferred technique for alkali and alkali earth metals.

## Group IIIB

Y(III) and Sc(III) yield double reduction currents representing the reduction to the bivalent form and subsequently to the metallic state, although there is little reported polarographic analysis for elements from group IIIB.

## Group IVB

Titanium is reducible from the (IV) oxidation state to the (III) oxidation state, producing a reversible reduction current at  $-0.81$  V versus SCE in  $0.1 \text{ mol l}^{-1}$  HCl, although the reduction potential is shifted greatly with pH changes and addition to the supporting electrolyte of complexing electrolytes such as EDTA, citrate, and tartrate. Titanium has been successfully measured in alloys, steel, minerals, and foodstuffs, having a typical limit of detection of  $5 \times 10^{-8} \text{ mol l}^{-1}$ . Polarographic detection of zirconium, hafnium is not possible.

## Group VB

Vanadium(V) may be determined in ammoniacal electrolyte producing two reduction waves at  $-0.1$  V versus SCE and at  $-1.26$  V versus SCE. Detection limits down to  $10^{-6} \text{ mol l}^{-1}$  have been achieved for vanadium present in petroleum fluids and organic material. Detection limits may be lowered by up to an order of magnitude by catalytic oxidation of the reduced vanadium(V).

Niobium, is electrochemically active in acidic solutions, however, there are few reported polarographic applications. Tantalum has not been determined by polarography.

## Group VIB

Chromium oxidation state discrimination is of major importance, principally owing to the much greater environmental toxicological threat of Cr(VI) compared to Cr(III). Polarography offers the analytical chemist/toxicologist techniques enabling the detection of these two different oxidation states. Cr(VI) and Cr(III) may both be determined electrochemically. Cr(VI) will yield a reduction current in a NaOH solution at around  $-1.5$  V versus SCE. Cr(III), under these solution conditions, is not responsive. However, Cr(III) can be determined after oxidation to Cr(VI), using, for example,  $\text{H}_2\text{O}_2$ , to give the total Cr concentration. Cr(III) is determined by the difference before and after treatment with  $\text{H}_2\text{O}_2$ . However, such an approach may lead to interference from the peroxide reduction current. Alternatively, Cr(III) is electroactive in an acetic acid solution producing a reduction current at about  $-0.8$  V versus SCE. It is therefore possible to measure Cr(VI) after Cr(III) by the chemical reduction of Cr(VI) to Cr(III). The latter approach is subject to interference from the proximity of the Ni and Zn reduction waves. Owing to the toxic nature of Cr(VI) applications have concentrated on biological and environmental samples.

Mo has been determined in soils, rocks, plant materials, and steels. Generally a catalytic reduction wave of Mo(VI), in the presence of nitrate in acidic solution, is monitored (approximately  $-0.13$  V versus Ag/AgCl).

W produces a reduction current at  $-0.2$  V versus Ag/AgCl in acid media representing W(VI) reduction. Detection is possible to  $10^{-6} \text{ mol l}^{-1}$ , below which the reduction of U(VI) may become a source of interference.

## Group VIIB

Mn(II) produces well defined reduction waves in most sample media and has been successfully determined in a wide range of sample types from natural waters to steel samples. Determination is usually carried out in alkaline conditions ( $\text{pH} \sim 9.5$ ), the reduction current appearing at  $-1.55$  V versus SCE. Detection limits are generally of the order of  $10^{-7} \text{ mol l}^{-1}$ . Cr(III), Cr(II), and Fe(II) reduction peaks may, however, be a source of interference. The application of anodic stripping voltammetry (ASV) will lower the detection limits further for Mn.

## Group VIIIB

The determination of Fe(II) at near neutral pH (to prevent hydrogen wave masking) is possible with the

reduction of Fe(II) to Fe metal, being detected at  $-1.4$  V versus SCE (detection limits;  $\sim 10^{-6} \text{ mol l}^{-1}$ ). Fe(III) normally would be too close to the mercury reduction potential, however, with the addition of a complexing agent (e.g., EDTA) and solution  $\text{pH} < 6$ , the reduction potential would be cathodically shifted, enabling differentiation of the oxidation states of iron. Recently Fe(III) has been detected via the reversible reduction of Fe(III) oxinate, preceded by a solvent extraction of the complex in chloroform. Samples analyzed for iron encompass wastewaters, anoxic lake waters, rocks, solar grade silicon, nonferrous alloys, and plant material.

Ni(II) and Co(II) both produce reduction waves ( $-0.8$  and  $-1.2$  V versus SCE, respectively, in ammoniacal solution) and have been determined in chromites, reactor waste, and natural waters. Interferences such as Zn can be minimized by the addition of complexing reagents (e.g., crown ethers), which selectively complex with Ni and Co producing a greater cathodic potential shift compared to the Zn reduction potential. Detection limits using this approach would amount to  $10^{-6} \text{ mol l}^{-1}$ .

Using an ammoniacal/dimethylglyoxime (DMG) media both Ni and Co reduction waves are well resolved from the Zn reduction wave. Additionally, owing to the adsorptive properties of the Ni/Co DMG complex, the limits of detection would be lowered by up to an order of magnitude. Adsorptive stripping voltammetry (anodic and cathodic) would reduce the detection limits even further.

## Group IB

The reduction of Cu(II) occurs typically at  $-0.2$  V versus SCE. Samples analyzed have mainly been potable waters, foodstuffs, soils, and biological samples. Typical limits of detection are of the order of  $10^{-7} \text{ mol l}^{-1}$ . Although the sensitivity may conceivably be enhanced by the presence in the sample solution of an organic ligand, which would form an adsorptive copper complex (e.g., catechol or 8-hydroxyquinoline) analogous to that observed for Ni and Co.

## Group IIB

Zinc and cadmium both exhibit well-defined reduction waves in almost all sample media ( $-1.00$  and  $-0.64$  V versus SCE, respectively, in  $1 \text{ mol l}^{-1}$  KCl solution). Cadmium being perhaps the most quoted example of a reversible electrode process, a detection limit for Cd of  $10^{-8} \text{ mol l}^{-1}$  and  $5 \times 10^{-7} \text{ mol l}^{-1}$  for Zn would be expected. Ni and

Co interference for Zn determinations has already been discussed.

### Group IIIA

The reduction potential of Al(III) is similar to those of the alkali and alkali earth metals making its determination difficult. However, Al may be determined indirectly by measuring an anodically shifted reduction current of an adsorbed di-*o*-hydroxyazo dye complexed with aluminum. The complexed dye would have a different reduction potential compared to the uncomplexed dye.

Recently indium has been detected by DPP following adsorption of the In 1-(2-pyridylazo)-2-naphthol complex on naphthalene in the pH range of 6.5–11.5. The quoted detection limit being  $2 \times 10^{-6} \text{ mol l}^{-1}$ .

### Group IVA

Si has been measured by the formation of the heteropoly acid with Sb(III) and Mo(VI) followed by adsorption onto the DME and is reducible within the DME potential window ( $-0.3 \text{ V}$  versus SCE), with detection limits in the region of  $\sim 10^{-7} \text{ mol l}^{-1}$ .

Reduction of  $\text{Pb}^{2+}$  at the DME is well documented having a characteristic peak potential at  $-0.4 \text{ V}$  versus SCE in  $\text{KNO}_3$  solution (detection limit being  $10^{-7} \text{ mol l}^{-1}$ ). Application of the technique encompasses samples such as rocks, soils, foods, ceramics glass, and biological samples. Recently an alternative approach has been adopted, whereby Pb is complexed with 1-(2-thiazolyazo)-2-naphthol-tetraphenylborate, which is quantitatively retained on microcrystalline naphthalene. Extraction of the Pb complex is carried out using an  $\text{HCl/NaBH}_4$  solution followed by polarographic analysis.

Sn illustrates two reduction processes in a chloride-supporting electrolyte. The first, representing the reduction of Sn(IV) to Sn(II) occurs at  $-0.25 \text{ V}$  versus SCE, followed by the reduction of Sn(II) to the metal (Hg) at  $-0.5 \text{ V}$  versus SCE. Detection limits down to  $10^{-6} \text{ mol l}^{-1}$  would be expected. Organo forms of tin (i.e., dimethyl; trimethyl) have recently been detected in 20% (v/v) methanol/water, tetraethylammonium perchlorate at pH 2.5 by DPP. Detection limits of  $6 \times 10^{-7}$  and  $4 \times 10^{-6} \text{ mol l}^{-1}$  were calculated for dimethyltin and trimethyltin, respectively.

### Group VA

Nitrate and nitrite may be determined by polarography, both anions having significance in the

monitoring of waste and potable waters and for soil analysis. Polarography has a direct advantage over ion-selective electrode (ISE) determination of nitrate as ISE is prone to interferences in high chloride media.

The nitrite and nitrate reduction currents are enhanced in the presence of certain heavy elements such as La(III), Yb(III), U(IV), and Ce(III). Simultaneous determination of nitrate and nitrite is achievable with careful choice of supporting electrolyte. Typical detection limits of  $1 \mu\text{mol l}^{-1}$  are encountered. Interference may arise from sulfate, phosphate, and oxalate.

Nitrite alternatively may be determined via the diazotization with sulfanilic acid followed by the coupling with *N*-(1-naphthyl)ethylenediamine dihydrochloride, detection limits being  $\sim 10^{-7} \text{ mol l}^{-1}$ . Possible enhancement of the sensitivity could be achieved by adsorptive cathodic stripping of this complex.

More recently nitrate has been determined in vegetables by its reduction to nitrous oxide then allowed to react in solution with cobalt(II) and thiocyanate ions, forming an electroactive complex that is reduced at the DME at  $-0.5 \text{ V}$  (versus SCE). The calculated detection limit was 39 mg of nitrate per kilogram of fresh vegetables. The performance of this new approach has been found to be comparable to reference spectrophotometric methods.

Similar to Si, P as orthophosphate may be detected after derivatization with Sb(III) and Mo(VI) of the ternary heteropoly acid which is adsorbed onto the DME (detection limit being  $5 \times 10^{-8} \text{ mol l}^{-1}$ ). Phosphorus determination, like nitrite and nitrate determinations, is of importance for the evaluation of natural, waste, and potable water quality.

Arsenic may occur as different oxidation states, principally as the +5 arsenate and the more toxic +3 arsenite. Arsenite is usually determined in acidic ( $1 \text{ mol l}^{-1} \text{ HCl}$ ) conditions by DPP, two reduction waves appearing at  $-0.4$  and  $-0.85 \text{ V}$  versus SCE, respectively. However, under such conditions the arsenate is not electrochemically active, therefore, arsenate must be reduced to arsenite and then measured polarographically as arsenite (thereby, giving the total arsenic concentration). The detection limit for this approach is of the order of the  $3 \times 10^{-9} \text{ mol l}^{-1}$ . Due to the toxic nature of As compounds, detection is principally carried out in biological and environmental samples.

Sb compounds also exhibit toxic properties, the trivalent form being more toxic than the pentavalent state. Sb(V) is detectable only in strongly acidic solutions in the presence of excess chloride or bromide leading to a double reduction step of Sb(V) to Sb(III), followed by Sb(III) to Sb(Hg). In less acidic solutions

only Sb(III) is electroactive, enabling the possible differentiation of the two oxidation states of Sb (approximate detection limit is  $10^{-7} \text{ mol l}^{-1}$ ).

Recently Bi(III) has been determined in alloy samples by reduction of Bi(III)-oxinate, preceded by its extraction in chloroform (detection limit being  $4.7 \times 10^{-7} \text{ mol l}^{-1}$ ).

## Group VIA

Oxygen is very electrochemically active, such that it is a serious electrochemical interference during the determination of other chemical ion aqueous solutions, hence the necessity for its removal. Principally dissolved oxygen will produce two reduction waves (in alkaline solutions) at the DME. One at around,  $-0.05 \text{ V}$  versus SCE with the production of  $\text{H}_2\text{O}_2$  and the second reduction peak is observable at approximately  $-0.9 \text{ V}$  versus SCE. This property has been capitalized on for the determination of dissolved oxygen, particularly in water (all types), sewage, and physiological applications. However, the traditional DME, although perfectly usable in the laboratory, is not practical when applied in field conditions. Therefore, a membrane-covered sensor is commonly used for fieldwork, encompassing the indicator and reference electrode, separated from the test solution by a gas permeable hydrophobic membrane.

The various oxidation states of sulfur have been determined by polarography applying such techniques to sample types including drinks, wastewaters, petroleum, and environmental samples (anoxic lake waters, sediments/sediment pore waters). Species detected have included  $\text{S}^{2-}$ ,  $\text{SO}_3^{2-}$ ,  $\text{S}_2\text{O}_3^{2-}$ .  $\text{S}^{2-}$  in NaOH solution yields a peak potential around  $-0.75 \text{ V}$  versus SCE. Additions to the solution of acetic acid with subsequent purging will remove  $\text{H}_2\text{S}$ , enabling the detection of  $\text{S}_2\text{O}_3^{2-}$  and  $\text{SO}_3^{2-}$  at respective potentials at  $-0.15$  and  $-0.63 \text{ V}$  versus SCE.

Recently the determination of elemental sulfur and hydrogen sulfide at sub-parts per million levels in petroleum has been achieved by preliminary extraction followed by DPP.

The importance of selenium in biological systems has initiated the application of polarography for its study. Selenium(IV) in acid media will produce two reduction currents one at about  $-0.01 \text{ V}$ , and the other at approximately  $-0.54 \text{ V}$  versus SCE, the first corresponding to the reduction of Se(IV) and the formation of  $\text{HgSe}$ , the second being the reduction of  $\text{HgSe}$  leading to the formation  $\text{Hg}$  and  $\text{Se}^{2-}$ . Using a nitric acid media, detection limits of  $10^{-6} \text{ mol l}^{-1}$  may be achieved. The limit of detection for Se analysis by DPP may be enhanced by the measurement

of the hydrogen catalytic wave. The reduction peak occurs at  $-1.1 \text{ V}$  versus SCE, in the presence of Mo (similar to 100–200 times higher concentrations than Se) at a pH of 1–4.

Te(IV), is reducible in nitric acid ( $\sim 0.25 \text{ mol l}^{-1}$ ) with detection limit of  $\sim 10^{-7} \text{ mol l}^{-1}$ .

## Halogens

The halogens may be determined polarographically by reduction of their halate form (except for chlorate). Bromide, for example, may be determined by DPP after oxidation to the bromate species yielding a cathodic current at  $\sim -1.53 \text{ V}$  versus SCE. The detection limit is  $\sim 1 \times 10^{-6} \text{ mol l}^{-1}$ . Possible interferences would arise from iodate reduction.

Chlorate is not reducible by polarography except in the presence of titanium and molybdenum salts producing a catalytic reduction current, although this technique has a low sensitivity with detection limits being in the order of  $1 \text{ mmol l}^{-1}$ .

Iodate ( $\text{IO}_3^-$ ) may be determined by DPP with detection limits down to  $\sim 2 \times 10^{-8} \text{ mol l}^{-1}$ . Oxidation of the  $\text{I}^-$  in the sample to  $\text{IO}_4^-$  followed by polarographic analysis will allow the determination of the chemical speciation of iodine.

Halides may also be determined by the formation of insoluble mercuric halide salts as a film at the electrode surface followed by the reduction of the film to the metal and the halide.

## Lanthanide Series

Simultaneous determination of Ce, Pr, Nd, Sm, and Eu has recently been carried out in a supporting electrolyte of xylenol orange at pH=4, leading to calculated detection limits in the region of  $10^{-7} \text{ mol l}^{-1}$ .

## Actinide Series

U(VI) has been recently determined by adsorption of the uranium trifluoethylxanthate–cetyltrimethylammonium ion pair complex on microcrystalline naphthalene, in the pH range of 4.2–7.0. A reduction peak would be observed at  $-0.2 \text{ V}$  versus SCE (detection limit being  $1.2 \times 10^{-6} \text{ mol l}^{-1}$ ). The method has been optimized to determine U(VI) in alloys, coal fly ash, biological, synthetic, and wastewater samples.

The above discussion has provided the reader with an overview of polarographic inorganic analyses, emphasizing the flexibility, sensitivity, and limitations of the technique. For more specific information of inorganic applications and recent developments, the reader is referred to references listed in the 'Further Reading' section.

See also: **Polarography**: Overview; Techniques; Organic Applications. **Voltammetry**: Overview; Linear Sweep and Cyclic; Anodic Stripping; Cathodic Stripping; Inorganic Compounds.

## Further Reading

Bard AJ and Faulkner LR (2000) *Electrochemical Methods: Fundamentals and Applications*, 2nd edn. New York: Wiley.

Bond AM (1980) *Modern Polarographic Methods in Analytical Chemistry*. New York: Dekker.

Plambeck J (1980) *Electroanalytical Chemistry: Basic Principles and Applications*. New York: Wiley.

Scholz F (ed.) (2002) *Electroanalytical Methods: Guide to Experiments and Applications*. Berlin: Springer.

Thomas F and Henze G (2001) *Introduction to Voltammetric Analysis: Theory and Practice*. Collingwood/Victoria: CSIRO Publishing.

Wang J (1994) *Analytical Electrochemistry*, 2nd edn. New York: Wiley.

## Organic Applications

C E Efstathiou, University of Athens, Athens, Greece

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Polarography is probably the only dynamic (electrolytic) electroanalytical technique that can be conveniently and reliably applied for the determination of a wide spectrum of organic compounds. Candidate analytes must typically contain at least one electroactive functional group, which can participate in electrode reactions at potentials within the polarization range of the dropping-mercury electrode (DME). The  $E_{1/2}$  values depend mainly on the nature of the particular electroactive functional group(s), while the rest of the molecule (the size and/or the overall conformation) has only a minor effect on  $E_{1/2}$ . Therefore, it should be stressed that as polarography cannot discriminate between similar organic compounds, such as members of the same homologous series, chromatographic techniques are the methods of choice in these cases.

Organic polarographic analysis is primarily used for the microdetermination of small amounts of a particular electroactive compound in the presence of excess of other nonelectroactive compounds and matrices. It can also be used for the simultaneous determination of mixtures of electroactive compounds with different electroactive groups, under conditions (carrier electrolyte, solvent, and appropriate pH) ensuring wide separation of the corresponding polarographic waves. In the latter case, there must also be no interaction between the product(s) of the electrode reaction of one analyte with the other analyte(s). This article reviews the practical aspects of organic polarography and typical areas of applications.

### Polarographically Active Functional Groups

A functional group of an organic compound is considered to be 'polarographically active' if it can participate in an electrode reaction (thus acting as a depolarizer) at potential values within the polarization range of the DME. The electrode reaction must be fast enough to yield well-developed polarographic waves. In most cases the analyte is reduced during the electrode reaction, resulting in cathodic polarographic waves during a cathodic potential scan. Oxidation reactions are not uncommon creating anodic polarographic waves during anodic potential scans.

Electrochemical oxidation of mercury is also facilitated, taking place at more negative potentials, in the presence of certain organic compounds because of the formation of either soluble complexes or insoluble compounds with mercury. In both cases anodic polarographic waves are obtained during an anodic potential scan allowing the indirect determination of these nonelectroactive compounds.

Typical polarographically active functional groups and the corresponding electrode reactions are shown in Table 1.

### Polarographic Determination of Nonelectroactive Organic Compounds

Organic compounds not containing electroactive functionalities may also be determined polarographically after their quantitative conversion into electroactive species through a suitable reaction. There must also be a convenient route for the separation of the formed species from the excess of the electroactive reagent usually used for this purpose. This conversion may take place in the following ways.



See also: **Polarography**: Overview; Techniques; Organic Applications. **Voltammetry**: Overview; Linear Sweep and Cyclic; Anodic Stripping; Cathodic Stripping; Inorganic Compounds.

## Further Reading

Bard AJ and Faulkner LR (2000) *Electrochemical Methods: Fundamentals and Applications*, 2nd edn. New York: Wiley.

Bond AM (1980) *Modern Polarographic Methods in Analytical Chemistry*. New York: Dekker.

Plambeck J (1980) *Electroanalytical Chemistry: Basic Principles and Applications*. New York: Wiley.

Scholz F (ed.) (2002) *Electroanalytical Methods: Guide to Experiments and Applications*. Berlin: Springer.

Thomas F and Henze G (2001) *Introduction to Voltammetric Analysis: Theory and Practice*. Collingwood/Victoria: CSIRO Publishing.

Wang J (1994) *Analytical Electrochemistry*, 2nd edn. New York: Wiley.

## Organic Applications

C E Efstathiou, University of Athens, Athens, Greece

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Polarography is probably the only dynamic (electrolytic) electroanalytical technique that can be conveniently and reliably applied for the determination of a wide spectrum of organic compounds. Candidate analytes must typically contain at least one electroactive functional group, which can participate in electrode reactions at potentials within the polarization range of the dropping-mercury electrode (DME). The  $E_{1/2}$  values depend mainly on the nature of the particular electroactive functional group(s), while the rest of the molecule (the size and/or the overall conformation) has only a minor effect on  $E_{1/2}$ . Therefore, it should be stressed that as polarography cannot discriminate between similar organic compounds, such as members of the same homologous series, chromatographic techniques are the methods of choice in these cases.

Organic polarographic analysis is primarily used for the microdetermination of small amounts of a particular electroactive compound in the presence of excess of other nonelectroactive compounds and matrices. It can also be used for the simultaneous determination of mixtures of electroactive compounds with different electroactive groups, under conditions (carrier electrolyte, solvent, and appropriate pH) ensuring wide separation of the corresponding polarographic waves. In the latter case, there must also be no interaction between the product(s) of the electrode reaction of one analyte with the other analyte(s). This article reviews the practical aspects of organic polarography and typical areas of applications.

### Polarographically Active Functional Groups

A functional group of an organic compound is considered to be 'polarographically active' if it can participate in an electrode reaction (thus acting as a depolarizer) at potential values within the polarization range of the DME. The electrode reaction must be fast enough to yield well-developed polarographic waves. In most cases the analyte is reduced during the electrode reaction, resulting in cathodic polarographic waves during a cathodic potential scan. Oxidation reactions are not uncommon creating anodic polarographic waves during anodic potential scans.

Electrochemical oxidation of mercury is also facilitated, taking place at more negative potentials, in the presence of certain organic compounds because of the formation of either soluble complexes or insoluble compounds with mercury. In both cases anodic polarographic waves are obtained during an anodic potential scan allowing the indirect determination of these nonelectroactive compounds.

Typical polarographically active functional groups and the corresponding electrode reactions are shown in Table 1.

### Polarographic Determination of Nonelectroactive Organic Compounds

Organic compounds not containing electroactive functionalities may also be determined polarographically after their quantitative conversion into electroactive species through a suitable reaction. There must also be a convenient route for the separation of the formed species from the excess of the electroactive reagent usually used for this purpose. This conversion may take place in the following ways.

**Table 1** Typical polarographically active organic compounds and the associated electrode reactions

<i>Class of organic compounds</i>	<i>Electrode reaction</i>
<i>Compounds resulting in cathodic waves</i>	
Olefins <sup>a</sup> and polyunsaturated compounds	$\text{RCH}=\text{CHR}' + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{RCH}_2\text{CH}_2\text{R}'$
Halogenated compounds	$\text{RX} + \text{H}^+ + 2\text{e}^- \rightarrow \text{RH} + \text{X}^- \quad (\text{X} = \text{Cl}, \text{Br}, \text{I})$
Aldehydes	$\text{RCHO} + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{RCH}_2\text{OH}$
Ketones	$\text{RCOR}' + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{RCH}(\text{OH})\text{R}'$
Peroxides	$\text{ROOR}' + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{ROH} + \text{R}'\text{OH}$
Epoxides	$\text{RCH}-\underset{\text{O}}{\text{CHR}'} + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{RCH}(\text{OH})\text{CH}_2\text{R}'$
Disulfides	$\text{RSSR}' + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{RSH} + \text{R}'\text{SH}$
Sulfoxides	$\text{RSO}_2\text{R}' + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{RH} + \text{R}'\text{SO}_2\text{H}$
Sulfones	$\text{RSOR}' + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{RSR}' + \text{H}_2\text{O}$
Nitriles	$\text{RCN} + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{RCH}=\text{NH}^{\text{b}}$
Nitroso compounds	$\text{RNO} + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{RNHOH}^{\text{b}}$
Nitro compounds	$\text{RNO}_2 + 4\text{H}^+ + 4\text{e}^- \rightarrow \text{RNHOH} + 2\text{H}_2\text{O}^{\text{b}}$
Hydroxylamines	$\text{RNHOH} + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{RNH}_2 + \text{H}_2\text{O}$
Oximes	$\text{RC}=\text{NOH} + 3\text{H}^+ + 3\text{e}^- \rightarrow \text{RCH}_2\text{NH}_2 + \text{H}_2\text{O}$
Imines	$\text{RCH}=\text{NH} + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{RCH}_2\text{NH}_2$
Diazo compounds	$\text{RN}=\text{NR}' + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{RNHNHR}'$
Aminoxides	$\text{R}_2\text{NO} + 3\text{H}^+ + 3\text{e}^- \rightarrow \text{RNHR} + \text{H}_2\text{O}$
Quinones	$\text{O}=\text{Ar}=\text{O} + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{HO}-\text{Ar}-\text{OH}$
Aryl isothiocyanates	$\text{ArNCS} + 4\text{H}^+ + \text{H}_2\text{O} + 4\text{e}^- \rightarrow \text{ArNH}_2 + \text{HCHO} + \text{H}_2\text{S}$
<i>Compounds resulting in anodic waves</i>	
Amines (aromatic)	$\text{ArNH}_2 \rightarrow \text{oxidation products} + x\text{H}^+ + xe^-^{\text{c}}$
Phenols	$\text{ArOH} \rightarrow \text{oxidation products} + x\text{H}^+ + xe^-^{\text{c}}$
<i>Compounds facilitating oxidation of Hg and resulting in anodic waves</i>	
Thiols (mercaptans)	$2\text{RSH} + \text{Hg} \rightarrow (\text{RS})_2\text{Hg} + 2\text{H}^+ + 2\text{e}^-$
Barbiturates, uracil	$-\text{NHCONH}- + \text{Hg} \rightarrow \text{Hg-complexes} + x\text{H}^+ + xe^-$
Thioureas, thiobarbiturates	$\text{RNHCSNHR}' + \text{Hg} \rightarrow \text{Hg-complexes} + x\text{H}^+ + xe^-$
Complexing reagents L (L = ammonia, amines, aminopolycarboxylates, etc.)	$x\text{L}^{b-} + \text{Hg} \rightarrow [\text{HgL}_x]^{(n-xb)+} + ne^-$

<sup>a</sup> Simple aliphatic olefins are practically nonelectroactive unless activated by conjugation. Fused aromatic rings (naphthalene, biphenyl, anthracene) are also electroactive.

<sup>b</sup> The reaction represents the first reduction step and (depending on the conditions) it can be followed by a second reduction (of the product) step.

<sup>c</sup> The nature of the products depends on the conditions, the number of electroactive groups and their positions. Imines and quinone types of compounds are usually the initial products but polymerization may also take place through free-radical mechanisms.

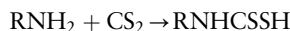
## Direct Derivatization

A typical example is the nitration reaction:

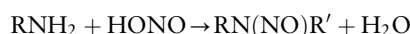


Nonelectroactive benzene or toluene can be quantitatively converted to electroactive nitroderivatives. The method can be extended to any polarographically inactive but easily nitrated aromatic compounds such as phenols, phenyl-bearing amino acids, and aniline derivatives. Interesting discriminations between similar compounds based on their ability to be nitrated or not can be achieved. Thus, the readily nitrated phenylbarbituric acid (phenobarbital) can be polarographically determined in the presence of a 100-fold excess of diethylbarbituric acid (barbital). The nitration of morphine also provides an excellent route for the polarographic determination of this otherwise practically nonelectroactive alkaloid.

Secondary amines can be converted to electroactive nitrosamines by the action of an excess of nitrous acid:



The polarographically inactive carbon disulfide in alkaline solutions reacts with amines to form electroactive dithiocarbamates:

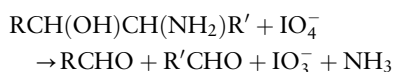
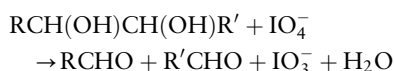


The same reaction has been exploited for the determination of nonelectroactive amines and amino acids.

## Decomposition

A typical example is the determination of *vic*-glycols or  $\alpha$ -aminoalcohols after their reaction with an excess of periodate (Malaprade reactions) to yield

electroactive aldehydes that are separated from the reaction mixture by distillation:



### Precipitation

A typical example is the determination of organic nitrogen bases (mainly of alkaloids, dyestuffs, and of several pharmaceuticals, e.g., phenothiazine derivatives, antihistaminics, and spasmolytics) after their precipitation with electroactive inorganic anions (such as silicotungstates or  $\text{BiI}_4^-$ ) and appropriate dissolving of the precipitates.

### Practical Considerations

Although the same general principles of polarographic measurements of inorganic species also apply to the polarographic determination of organic compounds, the nature of the electrode reactions involved imposes certain practical problems. These differences and their practical consequences are summarized below.

#### Rate of the Electrode Reactions

Most electrode reactions of inorganic species (e.g., of metal cations) are fast, as far as the electron transfer step is concerned. Hence they are reversible, the limiting currents measured are diffusion controlled and the polarographic waves are well developed, being adequately described by the Heyrovsky–Ilkovic equation. Electrode reactions involving organic species are generally irreversible, noticeable exceptions being those involving quinone–hydroquinone, imino–phenol–aminophenol, and azo–hydrazo redox pairs. Irreversible electrode reactions yield ill-shaped (drawn out) polarographic waves, thus making a wider separation from adjacent waves of other co-existing electroactive species necessary. With irreversible waves  $E_{1/2}$  values vary with the mercury drop time. Hence, it is not generally advisable to use tabulated half-wave potentials for identification purposes.

The rate of electrode reaction is of primary concern when modern pulsed (normal or differential pulse (DP) polarography) a.c. or fast sweep polarographic techniques are used. These techniques are really advantageous when fast electrode reactions are taking place and the observed currents are mainly diffusion limited.

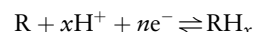
Kinetic waves are common in organic polarography. Carbonyl compounds and carbohydrates result in

kinetic waves because of the slowly established equilibrium between their main form (in water solution) and their electroactive form. In this case temperature effects are more pronounced and the use of thermostated cells is generally recommended for accurate work.

Catalytic waves are of special importance in organic polarography because they result in unusually low detection limits (down to  $10^{-8} \text{ mol l}^{-1}$ ). Many nitrogen-containing compounds (pyridine and quinoline derivatives, many alkaloids) give rise to well-pronounced catalytic waves. Similarly, in the presence of ammoniacal cobalt and nickel solutions, catalytic waves are obtained with sulfur-containing organic compounds (cysteine, dithiopyrimidine, and proteins containing sulfur amino acids).

#### Effect of pH

Contrary to the situation in inorganic polarography, where the complex-formation properties of the supporting electrolyte are of primary concern, in organic polarography the interest is shifted to pH-buffering properties. As a general rule, in all electrode reactions of organic species,  $\text{H}^+$  participates either as a reactant (cathodic reactions) or a product (anodic reactions). Hence,  $E_{1/2}$  values are pH dependent. In the case of the reversible electrode reaction of the general scheme

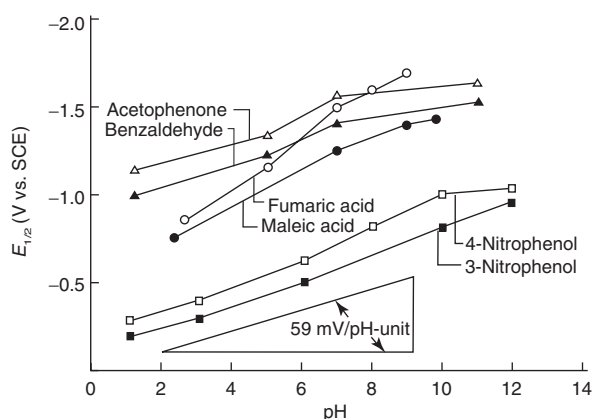


the dependence of  $E_{1/2}$  on pH is given by the equation

$$E_{1/2} = E^0 - \frac{2.303RT}{n} x\text{pH}$$

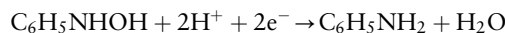
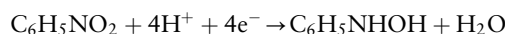
Due to the pronounced lack of reversibility of most electrode reactions of organic species, the pH effect is not always predictable, although one should always expect a cathodic (toward more negative values) shift of  $E_{1/2}$  values upon increase of pH. The degree of reversibility of many electrode reactions is also affected by pH. In cases of total irreversibility (e.g., reduction of organic halides)  $E_{1/2}$  values are more or less pH independent. The effect of pH on  $E_{1/2}$  of three pairs of closely related compounds is shown in **Figure 1**.

By controlling pH more efficient separations of the  $E_{1/2}$  values of closely related compounds can be occasionally achieved. Thus, at pH 3 the reduction waves of maleic acid and fumaric acid practically coincide, but at pH 7 the waves are separated sufficiently enough (**Figure 1**) to allow the polarographic determination of maleic acid in the presence of much higher concentrations of fumaric acid.



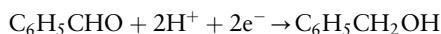
**Figure 1** Effect of pH on  $E_{1/2}$  values of three pairs of closely related electroactive organic compounds.

Because of the multistep nature of many electrode reactions of organic species, a single analyte may give rise to more than one polarographic wave with different degrees of reversibility. Nitrobenzene in acidic solutions ( $\text{pH} < 4$ ) is reduced in two steps corresponding to the electrode reactions

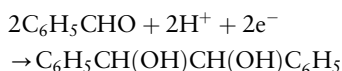


resulting in two well-separated cathodic waves, the first being about twice as high as the second, indicating the ratio of electrons participating in each half-reaction. In neutral or alkaline solutions a single cathodic wave is obtained, corresponding to the first electrode reaction.

In many cases, the nature of the final product is also pH dependent. Benzaldehyde in alkaline solutions is reduced in a single step according to the electrode reaction



whereas at  $\text{pH} < 2$  a single cathodic wave of half the height also occurs corresponding to the electrode reaction



In intermediate pH values two waves appear with their height ratio dependent on the actual pH.

In many cases the electrode reaction is favored only in a specific pH range, e.g., carbonyl-containing compounds are preferably determined in alkaline solutions because their enolic form is generally reduced more easily than their carbonyl form.

Since  $\text{H}^+$  is consumed during cathodic electrode reactions, or produced during anodic electrode reactions, the pH of the measured solution in the vicinity of the DME (diffusion layer) is expected to be much higher than that of the main bulk of the solution. Therefore, in order to obtain well-developed and reproducible polarographic waves the supporting electrolyte must contain a strong pH-buffer system. Weak buffering may cause drawn out polarographic waves with  $E_{1/2}$  dependent on the analyte concentration.

Interactions of buffer constituents with the analytes are often undesirable, e.g., borate buffers should be avoided with compounds containing vicinal hydroxyl groups, whereas such interactions may be exploited for masking interfering electroactive species in the samples. On the other hand, carbonyl-containing compounds are preferably determined in the presence of excess of ammonia because their imino form yields better shaped polarographic waves.

### Supporting (Carrier) Electrolyte Solvent

Depending on the nature of the analyte or sample, organic polarography may be carried out in aqueous solutions, water-organic solvent mixtures, or pure organic solvents of a protic nature. Commonly used solvents are lower alcohols, dioxane, tetrahydrofuran, dimethylsulfoxide, dimethylformamide, and acetonitrile.

The organic solvent serves the following purposes:

1. It ensures complete dissolution of water-insoluble samples. At concentrations below  $10^{-3} \text{ mol l}^{-1}$ , the analyte may be soluble in water but its matrix components may be water insoluble.
2. By maintaining a low water concentration the DME polarization range increases in the cathodic direction (e.g., up to  $-2.9 \text{ V}$  versus SCE), so the range of possible candidate analytes is enlarged and compounds with reduction waves preceded by the reduction of  $\text{H}^+$  in aqueous solutions can also be determined in purely organic media.
3. The presence of a less protic organic solvent may have a suppressive effect on reducing adsorption phenomena, and hence adsorption prewaves and waves, which in many cases unpredictably complicate the polarograms of organic compounds in real samples.
4. The reversibility of one-electron reductions of compounds with  $\pi$ -electron systems is often enhanced due to stabilization of anion radicals.

The supporting electrolyte must be soluble in the solvent used. Lithium salts ( $\text{LiCl}$ ,  $\text{LiNO}_3$ ) and quaternary ammonium salts ( $\text{R}_4\text{N}^+\text{X}^-$ , where  $\text{R} = \text{CH}_3$  to  $\text{C}_4\text{H}_9$ , and  $\text{X} = \text{Cl}^-$ ,  $\text{Br}^-$ ,  $\text{I}^-$ ,  $\text{ClO}_4^-$ ,

$\text{BF}_4^-$ ,  $\text{PF}_6^-$ ) are extensively used for this purpose. Relatively concentrated solutions are preferable to reduce the pronounced ohmic resistance of the organic solvent and minimize the  $iR$  shifts imposed on observed  $E_{1/2}$  values. Attention should also be paid to the compatibility of the reference electrode used (reference electrodes with nonaqueous internal solutions are available). An electrolyte bridge compatible with both junctions of the electrolyte–solvent solution must separate the reference electrode and the sample solution to prevent clogging of the reference electrode glass frit by precipitation of the internal solution salt.

Deoxygenation of the supporting electrolyte–sample solution must be carried out for relatively extended time periods compared to those required for pure aqueous solutions (e.g., 10–15 min compared to 3–5 min), due to the increased solubility of oxygen in many organic solvents. The inert gas serving this purpose must be saturated in solvent vapor (by passing through washing bottles containing the same supporting electrolyte solution) to prevent evaporation of volatile solvents, which eventually lead to positive analytical errors.

Polarographic data of some representative classes of simple organic compounds are given in **Table 2**. The proximity of  $E_{1/2}$  values of closely related compounds and their dependence on pH are obvious. It should be emphasized that the tabulated  $E_{1/2}$  values are only indicative since in many cases they are strongly dependent on the actual experimental conditions and the DME characteristics.

### Quantification Limit

The quantification limit depends on factors such as: (1) the number of electrons participating in the electrode reaction; (2) the reversibility of the electrode reaction and hence the applicability of modern pulse techniques; (3) the separation of the polarographic wave of the analyte from the waves of other polarographically active compounds – generally a separation of the order of 0.10–0.15 V is adequate provided that well-shaped waves are obtained (a separation of 0.05 V would be enough when DP polarography is used); and (4) factors such as the occurrence of adsorption waves that may distort or totally obscure the polarographic waves.

As a general rule, quantification limits of  $10^{-4} \text{ mol l}^{-1}$  can be readily obtained with conventional (d.c.) polarography, whereas these limits may be reduced to  $10^{-7} \text{ mol l}^{-1}$  or lower by using modern pulse polarographic techniques and particularly square wave (Osteryoung) polarography.

**Figure 2** shows typical polarograms of tetracycline obtained by DP and conventional (d.c.) polarography.

Lower quantification limits (by one to two orders) can be achieved in a limited number of cases by exploiting catalytic polarographic waves, where the concentration of the catalytically active compound is related to the height of the polarographic wave of another species. A typical example is the determination of cystine at concentrations down to  $10^{-7} \text{ mol l}^{-1}$  by its catalytic effect on the reduction wave of  $\text{Co}^{2+}$  (Brdička wave).

### Interferences

Apart from interferences caused by the presence of other electroactive compounds with  $E_{1/2}$  values close to that of the analyte, the most serious problems occur when the samples contain surface-active compounds (e.g., detergents, proteins, water-soluble polymeric compounds used as excipients in drug formulations). Generally, the adsorption of these compounds on the mercury electrode decreases its active surface, hence reducing the electron transfer rate and suppressing the height of the polarographic waves or peaks (in DP polarography). Cases are also known where the polarographic waves are increased. Cathodic shifts of  $E_{1/2}$  values have also been observed upon increasing the concentration of surfactants. This more or less unpredictable action of surfactants makes the use of the standard addition (single or multiple) technique absolutely necessary for quantification purposes.

### Applications

Due to the large number of published applications of organic polarography, only a selected number of determined compounds and of other applications are presented here.

#### Applications in Pharmaceutical Analysis

Organic polarography has its widest range of applications in the analysis and quality control of pharmaceuticals. Literally hundreds of compounds of pharmaceutical importance have been determined polarographically in a broad spectrum of matrices such as pharmaceutical formulations (injections, tablets, ointments, creams, dragées), bulk materials, biological fluids, and extracts. Whereas polarography has been superseded by liquid chromatography as an official analytical technique in the quality control of pharmaceuticals, it is often considered as a complementary technique to ultraviolet–visible spectrophotometry where it can be used as a secondary control.

**Table 2** Polarographic data of some typical simple organic analytes

Compound	Supporting electrolyte	pH	$E_{1/2}$ (V vs. SCE) <sup>a</sup>	$n^b$
CH <sub>3</sub> Cl	0.05 mol l <sup>-1</sup> Et <sub>4</sub> NBr/75% dioxane		-2.23	2
CH <sub>2</sub> Cl <sub>2</sub>	0.05 mol l <sup>-1</sup> Et <sub>4</sub> NBr/75% dioxane		-2.23	
CHCl <sub>3</sub>	0.05 mol l <sup>-1</sup> Et <sub>4</sub> NBr/75% dioxane		-1.67/?	2/(2)
C <sub>2</sub> H <sub>5</sub> Cl	0.05 mol l <sup>-1</sup> Et <sub>4</sub> NBr/75% dioxane		Not reduced	
CH <sub>3</sub> Br	0.05 mol l <sup>-1</sup> Et <sub>4</sub> NBr/75% dioxane		-1.63	2
CH <sub>2</sub> Br <sub>2</sub>	0.05 mol l <sup>-1</sup> Et <sub>4</sub> NBr/75% dioxane		-1.48	4(?)
CHBr <sub>3</sub>	0.05 mol l <sup>-1</sup> Et <sub>4</sub> NBr/75% dioxane		-0.64/-1.51	2/4
C <sub>2</sub> H <sub>5</sub> Br	0.05 mol l <sup>-1</sup> Et <sub>4</sub> NBr/75% dioxane		-2.08	2
<i>n</i> -C <sub>3</sub> H <sub>7</sub> Br	0.01 mol l <sup>-1</sup> Et <sub>4</sub> NBr/DMF		-2.20	2
<i>n</i> -C <sub>4</sub> H <sub>7</sub> Br	0.05 mol l <sup>-1</sup> Et <sub>4</sub> NBr/75% dioxane		-2.23	2
<i>n</i> -C <sub>5</sub> H <sub>11</sub> Br	0.01 mol l <sup>-1</sup> Et <sub>4</sub> NBr/DMF		-2.26	2
CH <sub>3</sub> I	0.05 mol l <sup>-1</sup> Et <sub>4</sub> NBr/75% dioxane		-1.63	2
CH <sub>2</sub> I <sub>2</sub>	0.05 mol l <sup>-1</sup> Et <sub>4</sub> NBr/75% dioxane		-1.12/-1.53	2/2
CHI <sub>3</sub>	0.05 mol l <sup>-1</sup> Et <sub>4</sub> NBr/75% dioxane		-0.49/-1.09/-1.50	2/2/2
C <sub>2</sub> H <sub>5</sub> I	0.05 mol l <sup>-1</sup> Et <sub>4</sub> NBr/75% dioxane		-1.67	2
CH <sub>3</sub> NO <sub>2</sub>	BR <sup>c</sup> /3% MeOH	4.6	-0.83	4
C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub>	BR <sup>c</sup> /3% MeOH	4.6	-0.83	4
<i>n</i> -C <sub>3</sub> H <sub>7</sub> NO <sub>2</sub>	BR <sup>c</sup> /3% MeOH	4.6	-0.84	4
Nitrobenzene	Various buffers/10% EtOH	1.0	-0.22/-0.68	4/2
Nitrobenzene	Various buffers/10% EtOH	3.0	-0.34x-0.86	4/2
Nitrobenzene	Various buffers/10% EtOH	6.0	-0.54	4
Nitrobenzene	Various buffers/10% EtOH	12.0	-0.83	4
2-Chloro-	Various buffers/50% EtOH	7.4	-0.91	
3-Chloro-	Various buffers/50% EtOH	7.4	-0.86	
4-Chloro-	Various buffers/50% EtOH	7.4	-0.87	
2-Nitrotoluene	Borate buffer	9.2	-0.75	4
3-Nitrotoluene	Borate buffer	9.2	-0.71	4
4-Nitrotoluene	Borate buffer	9.2	-0.69	4
1-Nitronaphthalene	Various buffers	2.1	-0.30/-0.65	4/2
1-Nitronaphthalene	Various buffers	6.4	-0.49	4/-
1-Nitronaphthalene	Various buffers	11.0	-0.82	4
2-Nitronaphthalene	Various buffers	2.1	-0.30/-0.65	4/2
2-Nitronaphthalene	Various buffers	6.4	-0.49	4/-
2-Nitronaphthalene	Various buffers	11.0	-0.82	4
Azobenzene	0.01 mol l <sup>-1</sup> HCl + 0.02 mol l <sup>-1</sup> KCl/30% MeOH		-0.06/-0.81	2/2
Azoxybenzene	0.01 mol l <sup>-1</sup> HCl + 0.02 mol l <sup>-1</sup> KCl/30% MeOH		-0.25/-0.83	4/2
Benzenediazonium chloride	Acetate buffer	4.0	-0.19	
HCHO	0.1 mol l <sup>-1</sup> LiOH		-1.71	2
CH <sub>3</sub> CHO	0.1 mol l <sup>-1</sup> LiOH		-1.89	2
C <sub>2</sub> H <sub>5</sub> CHO	0.1 mol l <sup>-1</sup> LiOH		-1.92	2
CH <sub>3</sub> COCH <sub>3</sub>	0.05 mol l <sup>-1</sup> ( <i>n</i> -Bu) <sub>4</sub> NCl/90% EtOH		-2.57	2
C <sub>2</sub> H <sub>5</sub> COCH <sub>3</sub>	0.05 mol l <sup>-1</sup> ( <i>n</i> -Bu) <sub>4</sub> NCl/90% EtOH		-2.59	2
Cyclohexanone	0.05 mol l <sup>-1</sup> ( <i>n</i> -Bu) <sub>4</sub> NCl/90% EtOH		-2.40	2
1,4-Benzoquinone	Phosphate buffer	7.0	+0.04	2
1,4-Naphthoquinone	Phosphate buffer	7.0	-0.13	2
Anthraquinone	Phosphate buffer	7.0	-0.55	2
C <sub>2</sub> H <sub>5</sub> SH	0.01 mol l <sup>-1</sup> H <sub>2</sub> SO <sub>4</sub>		(+0.02)	-2
Cystine	Acetate buffer	3.8	-0.68	2
Diethyldisulfide	0.025 mol l <sup>-1</sup> Bu <sub>4</sub> NOH/2-PrOH-MeOH-H <sub>2</sub> O(2:2:1)		-1.78	2
Methacrylic acid	0.1 mol l <sup>-1</sup> LiCl/50% EtOH		-1.65	
Methacrylic acid methyl ester	0.1 mol l <sup>-1</sup> LiCl/30% EtOH		-1.96	
Acrylonitrile	0.05 mol l <sup>-1</sup> Me <sub>4</sub> Nl/30% EtOH		-1.96	2
Methacrylonitrile	0.1 mol l <sup>-1</sup> Me <sub>4</sub> NBr		-2.07	
Styrene	0.175 mol l <sup>-1</sup> ( <i>n</i> -Bu) <sub>4</sub> Nl/75% dioxane		-2.35	2
β-methyl-	0.175 mol l <sup>-1</sup> ( <i>n</i> -Bu) <sub>4</sub> Nl/75% dioxane		-2.54	2
Naphthalene	0.175 mol l <sup>-1</sup> ( <i>n</i> -Bu) <sub>4</sub> Nl/75% dioxane		-2.49	2
Anthracene	0.175 mol l <sup>-1</sup> ( <i>n</i> -Bu) <sub>4</sub> Nl/75% dioxane		-1.94	2
Phenanthrene	0.175 mol l <sup>-1</sup> ( <i>n</i> -Bu) <sub>4</sub> Nl/75% dioxane		-2.44	2
Fructose	0.1 mol l <sup>-1</sup> LiCl		-1.76 <sup>d</sup>	
Glucose	Phosphate buffer	7.0	-1.54 <sup>d</sup>	

Continued



**Table 2** Continued

Compound	Supporting electrolyte	pH	$E_{1/2}$ (V vs. SCE) <sup>a</sup>	$n^b$
Glucose	Phosphate buffer	10.8	-1.68 <sup>d</sup>	
Mannose		7.0	-1.51 <sup>d</sup>	

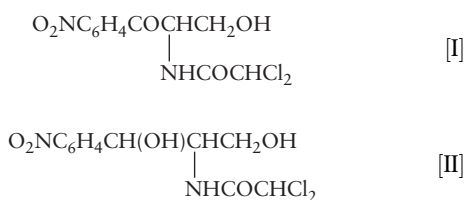
<sup>a</sup>  $E_{1/2}$  corresponding to two different polarographic waves are separated by a solidus.

<sup>b</sup> Number of electrons participating in each reduction step. Negative number of electrons indicates anodic electrode reactions (anodic polarographic waves).

<sup>c</sup> Britton–Robinson buffer (acetate–phosphate–borate composite).

<sup>d</sup> The acyclic tautomeric form of the carbohydrate is polarographically active, hence relatively weak kinetic waves only are obtained.

Generally, polarographic determinations, which have an analytical accuracy within the range 1–5%, are more appropriate for measurements of minor constituents of pharmaceutical formulations (typical w/w content <5%). Polarography can also be used for the determination of impurities (typically less than 1%) in bulk pharmaceutical materials. A typical example is the polarographic determination of the by-product ketocompound [I] in commercial lots of chloramphenicol [II]. This determination is of primary importance because of some unwanted physiological side effects of [I]. The polarographic determination of [I] is favored because its cathodic polarographic wave precedes that of chloramphenicol ( $E_{1/2} = -0.75$  V versus SCE at pH 10) by  $\sim 0.15$  V.



Typical compounds of pharmaceutical importance determined polarographically are listed in Table 3.

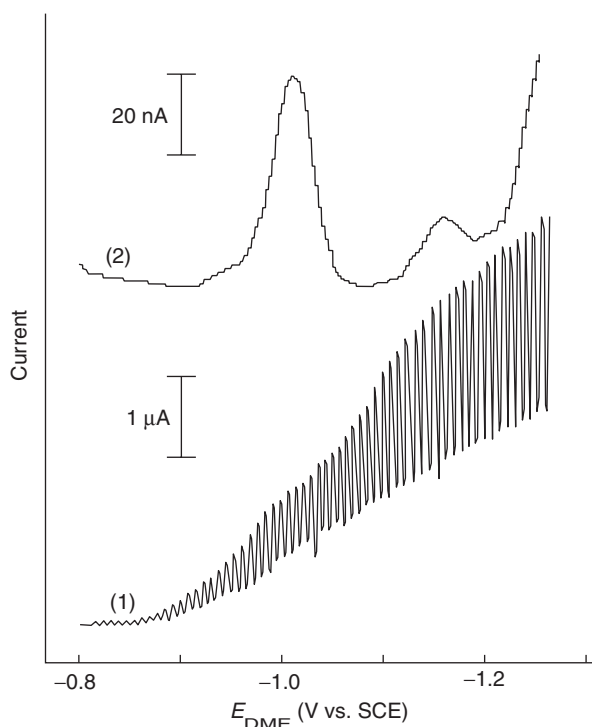
### Biochemical and Biomedical Applications

Polarographic determinations have been mainly used in clinical chemistry, toxicology, and studies on pharmacokinetics and therapeutic drug monitoring.

Among the very few biochemical applications of polarography the most famous was the 'Brdička protein test'. This test was based on the catalytic double wave (attributed to the –S–S– bonds in peptides and proteins) in ammoniacal solutions containing  $\text{Co}^{2+}$  or  $\text{Co}^{3+}$  salts. This (now obsolete) test provided a fast discrimination between normal sera and those obtained from patients suffering from cancer, and displayed a good correlation between the tumor growth state and the observed polarographic response.

Generally, polarography is not considered as a practical technique for routine clinical analysis.

Typical compounds of biochemical and biomedical importance determined polarographically (most of



**Figure 2** Polarograms of tetracycline in acetate buffer. (1) Conventional (d.c.) polarogram of a  $150\text{ mg l}^{-1}$  solution. (2) Differential pulse polarogram of a  $1\text{ mg l}^{-1}$  solution. (Reprinted from *Analytical Chemistry* September 1972, 44(11): 75A–87A. Published 1972, American Chemical Society.)

them in biological fluids) and other biochemical studies based on polarographic measurements are listed in Table 4.

### Industrial and Agricultural Applications

Organic polarography is extensively used for the determination of compounds of industrial and agricultural interest (particularly for process and quality control). A classic example is the determination of trace impurities (0.001–0.1%) of nitrobenzene in technical-grade aniline.

Polarography has been advantageously used in the dye industry. A variety of intermediates (anthraquinone

**Table 3** Representative compounds of pharmaceutical importance determined polarographically

Alkaloids	Berberine, codeine, dihydrocodeinone, dihydromorphine, heroin, hydroxycodeinone, morphine, quinine, quinone, jervine, lobeline, narceine, pelletierine, piperine, reserpine, strychnine, thebaine, veratramine, veratrosine
Antibiotics	Adriamycin, chloramphenicol, rubomycin, streptomycin, tetracyclines
Barbiturates	Barbituric acid, Phenobarbital, etc.
Benzodiazepines	Chlordiazepoxide, diazepam, flurazepam, lorazepam, medazepam, nitrazepam, oxazepam
Hormones and steroids	Bufagin, convallotoxin, deoxycorticosterone, dienestrol, digitoxin, estrone, flucinolone acetonide, gitoxin, methyltestosterone, pregnenolone-17-ol-3-one, progesterone, strophanthidin, testosterone, thyroxine
Organic nitrates	Glyceryl trinitrate, isosorbite dinitrate, triethanolamine trinitrate (nitranol)
Vitamins:	Vitamin B <sub>1</sub> (thiamine), B <sub>2</sub> (riboflavin), B <sub>6</sub> (pyridoxine), B <sub>12</sub> (cyanocobalamin), C (ascorbic acid), E (tocopherols), folic acid, pantothenic acid, K <sub>1</sub> , K <sub>2</sub> , K <sub>3</sub> (menadione), PP factor (nicotinamide), biotin
Miscellaneous	Benzimidazole drugs, bromoisovalerylurea, cardiazol, chlorhexidine, dicoumarol, dulcin, isoniazid, 2-nitrofur derivatives, parabens (4-hydroxybenzoic acid esters), phenacetin, phenylbutazone (and its metabolite oxyphenbutazone), propiopocaine, quercetin, rutin, thiomersal (sodium ethylmercurithiosalicylate)

**Table 4** Representative biochemical and biomedical determinations and studies by organic polarography

Determinations	Nonprotein sulphhydryl compounds, purines, thiopurines, cystine in the presence of cysteine, cholinesterase and peroxidase activity in blood, nucleic acids, ascorbic acid after $\alpha$ -ray radiation experiments, albumin, heme, hemoglobin, 4-chlorophenylalanine, <i>N</i> -substituted derivatives of nicotinamide, riboflavin in fermentation broths, urea in process streams, activity of BCG vaccine, catecholamines in tissues
General studies	Investigations of the effect of various proteins and of sulfhydryl-type radioprotective drugs on the Brdička wave, interaction of vasopressin and its analogues on native and denatured DNA, transitions of DNA helix coil, electron-transfer studies of various compounds of biochemical interest and proteins, denaturation and renaturation of various proteins, respiration of plant mitochondria, binding of various azo dyes with proteins, radiolysis of ascorbic acid, oxygen uptake by oxidases, dissociation of oxyhemoglobin in blood

**Table 5** Representative applications of organic polarography in industry and agriculture

Dyes	Determination of intermediates in dye synthesis (anthraquinone and derivatives), azo dyes in intermediates, products, coloring solutions and industrial effluents. Determination of aggregation number of textile dyes. Determination of food colors (e.g., of tartrazine type) in soft drinks
Polymers and plastics	Determination of monomers (aldehydes, acrylamide, acrylonitrile, isocyanates, methyl methacrylate, styrene, vinyl acetate) in final products and industrial effluents. Determination of initiators (azodiisobutylnitrile, benzoyl peroxide, cyclohexylperoxydicarbonate, laurylperoxide), inhibitors (hydroquinones, butylated hydroxyanisole, 4- <i>t</i> -butyl phenolsulphide, <i>N</i> -phenyl- $\beta$ -naphthylamine) and organo-tin stabilizers in PVC plastics
Agrochemicals (herbicides, pesticides, fungicides)	Azomethine group ( $>C=N-$ ) containing compounds: cyolane, cytolane, chlordimerform, drazoxolon s-Triazines: atrazine Chlorinated compounds: 2,4-D (2,4-dichlorophenoxyacetic acid), propachlor (2-chloro- <i>N</i> -2-propylacetanilide), pentachlorophenol, picloram (4-amino-3,5,6-trichloropicolinic acid), hexachlorocyclohexane Nitro-group-containing compounds: metaphos ( $\alpha,\alpha$ -dimethyl- $\alpha$ -(4-nitrophenyl) thiophosphate, parathion, paraoxon

and other quinone-type compounds) and final products (azo dyes, e.g., in industrial effluents of dyestuff plants) have been determined polarographically. The aggregation numbers of various textile dyes have been also determined by organic polarography.

The determination of small amounts of monomers, initiators, inhibitors, and stabilizers in plastic products is another field of applications of organic polarography.

Organic polarographic analysis has also been used for the determination of a variety of agrochemicals in various formulations as well as in polluted natural

waters. Determination by DP polarography of metaphos in waters down to  $10^{-8}$ – $10^{-9}$  mol l<sup>-1</sup>, and of parathion and paraoxon in mixtures in the range  $5 \times 10^{-8}$  to  $1 \times 10^{-6}$  mol l<sup>-1</sup> have been reported.

Typical industrial and agricultural applications of organic polarography are listed in Table 5.

**See also: Buffer Solutions. Clinical Analysis:** Overview. **Derivatization of Analytes. Pesticides. pH. Pharmaceutical Analysis:** Drug Purity Determination.

## Further Reading

- Bersier PM (1985) Applied polarography and voltammetry of organic compounds in practical day-to-day analysis, part I. *Review Articles in Critical Reviews in Analytical Chemistry* 16(1): 15–79.
- Bersier PM (1985) Applied polarography and voltammetry of organic compounds in practical day-to-day analysis, part II. *Reviews in Analytical Chemistry* 16(2): 81–128.
- Heyrovsky J and Zuman P (1968) *Practical Polarography*: London: Academic Press.
- Meites L (1965) *Polarographic Techniques*, 2nd edn. New York: Wiley.
- Pietrzyk DJ (1970) Organic polarography (review article). *Analytical Chemistry* 42: 139R–152R.
- Pietrzyk DJ (1972) Organic polarography (review article). *Analytical Chemistry* 44: 457R–478R.
- Pietrzyk DJ (1974) Organic polarography (review article). *Analytical Chemistry* 46: 52R–73R.
- Sawyer DT and Roberts JL Jr. (1974) *Experimental Electrochemistry for Chemists*. New York: Wiley.
- Zuman P (1964) *Organic Polarographic Analysis*. Oxford: Pergamon.

# POLYCHLORINATED BIPHENYLS

**J de Boer**, Netherlands Institute for Fisheries Research, IJmuiden, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Polychlorinated biphenyls (PCBs) are derived from biphenyl by substitution of one to 10 hydrogen atoms by chlorine atoms. Each homolog group has a particular number of isomers: mono-chlorobiphenyl 3, di- 12, tri- 24, tetra- 42, penta- 46, hexa- 42, hepta- 24, octa- 12, nona- 3, and decachlorobiphenyl 1. In total there are 209 possible PCB congeners. A coding system with numbers from 1 (mono-Cl) to 209 (deca-Cl) is normally used for the PCB congeners (chlorobiphenyls or CBs).

PCBs have been produced and used worldwide in large quantities (c. 1.3 million metric tons in total) for many years as transformer oils, cutting oils, hydraulic oils, heat transfer fluids, additives in plastics, dyes and carbonless copying paper, and metal-casting release oils. The production of PCBs was terminated worldwide around the late 1970s to the early 1980s after authorities became aware of the adverse effects of PCBs to the environment, due to their persistency, bioaccumulative properties, and toxicity. Commercial PCB products are produced under the names of Aroclor (USA), Chlophen (Germany), Kanechlor (Japan), Fenclor (Italy), and others.

During the last decade much attention has been paid to the toxicology of PCBs, particularly to the congeners that show the same type of toxicity as polychlorinated dibenzo-*p*-dioxines (PCDDs) and dibenzofurans (PCDFs). Certain PCBs, which lack chlorine substituents in the *ortho*-position, show a particularly high 'dioxin-like' toxicity (enzyme induction through binding of CB to the Ah-receptor),

viz. PCB-77, PCB-126, and PCB-169. Also, mono-*ortho*-substituted CBs show such a type of dioxin-like toxicity, although to a minor extent. The toxicity of dioxin-like CBs is compared to that of the most toxic dioxin, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (2,3,7,8-TCDD) and a toxic equivalency factor (TEF) has been given to those CBs. The TEF value of 2,3,7,8-TCDD is defined as 1.0. These TEF values have been regularly updated according to the most recent information from toxicological studies. The most recent TEF values derived by the World Health Organization (WHO) are given in Table 1. The product of the concentration and the TEF is called the TCDD equivalent (TEQ). The TEQ of all dioxin-like CBs, dioxins and furans together is called total-TEQ.

## Sampling and Sample Pretreatment

PCBs are analyzed in all sorts of environmental matrices: air, seawater, freshwater, sediments, sewage sludge, benthic organisms, fish, marine mammals, human blood, adipose tissue, and others. All samples require a specific approach. Sampling of sediments and biota will be discussed because these belong to the most frequently analyzed matrices in environmental analysis. The analysis of PCBs in water is known to be extremely difficult because their solubility in water is low. Consequently, PCB concentrations in water are extremely low and the analysis is hindered by background contamination. PCB concentrations in air can be determined after sampling by filters or passive sampling systems.

Surface sediments can be collected intertidally or by means of a variety of grab samplers from a vessel. Because PCB concentrations in sediments can show a patchy distribution, several grabs from one location are normally combined to one pooled sediment

## Further Reading

- Bersier PM (1985) Applied polarography and voltammetry of organic compounds in practical day-to-day analysis, part I. *Review Articles in Critical Reviews in Analytical Chemistry* 16(1): 15–79.
- Bersier PM (1985) Applied polarography and voltammetry of organic compounds in practical day-to-day analysis, part II. *Reviews in Analytical Chemistry* 16(2): 81–128.
- Heyrovsky J and Zuman P (1968) *Practical Polarography*: London: Academic Press.
- Meites L (1965) *Polarographic Techniques*, 2nd edn. New York: Wiley.
- Pietrzyk DJ (1970) Organic polarography (review article). *Analytical Chemistry* 42: 139R–152R.
- Pietrzyk DJ (1972) Organic polarography (review article). *Analytical Chemistry* 44: 457R–478R.
- Pietrzyk DJ (1974) Organic polarography (review article). *Analytical Chemistry* 46: 52R–73R.
- Sawyer DT and Roberts JL Jr. (1974) *Experimental Electrochemistry for Chemists*. New York: Wiley.
- Zuman P (1964) *Organic Polarographic Analysis*. Oxford: Pergamon.

# POLYCHLORINATED BIPHENYLS

**J de Boer**, Netherlands Institute for Fisheries Research, IJmuiden, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Polychlorinated biphenyls (PCBs) are derived from biphenyl by substitution of one to 10 hydrogen atoms by chlorine atoms. Each homolog group has a particular number of isomers: mono-chlorobiphenyl 3, di- 12, tri- 24, tetra- 42, penta- 46, hexa- 42, hepta- 24, octa- 12, nona- 3, and decachlorobiphenyl 1. In total there are 209 possible PCB congeners. A coding system with numbers from 1 (mono-Cl) to 209 (deca-Cl) is normally used for the PCB congeners (chlorobiphenyls or CBs).

PCBs have been produced and used worldwide in large quantities (c. 1.3 million metric tons in total) for many years as transformer oils, cutting oils, hydraulic oils, heat transfer fluids, additives in plastics, dyes and carbonless copying paper, and metal-casting release oils. The production of PCBs was terminated worldwide around the late 1970s to the early 1980s after authorities became aware of the adverse effects of PCBs to the environment, due to their persistency, bioaccumulative properties, and toxicity. Commercial PCB products are produced under the names of Aroclor (USA), Chlophen (Germany), Kanechlor (Japan), Fenclor (Italy), and others.

During the last decade much attention has been paid to the toxicology of PCBs, particularly to the congeners that show the same type of toxicity as polychlorinated dibenzo-*p*-dioxines (PCDDs) and dibenzofurans (PCDFs). Certain PCBs, which lack chlorine substituents in the *ortho*-position, show a particularly high 'dioxin-like' toxicity (enzyme induction through binding of CB to the Ah-receptor),

viz. PCB-77, PCB-126, and PCB-169. Also, mono-*ortho*-substituted CBs show such a type of dioxin-like toxicity, although to a minor extent. The toxicity of dioxin-like CBs is compared to that of the most toxic dioxin, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (2,3,7,8-TCDD) and a toxic equivalency factor (TEF) has been given to those CBs. The TEF value of 2,3,7,8-TCDD is defined as 1.0. These TEF values have been regularly updated according to the most recent information from toxicological studies. The most recent TEF values derived by the World Health Organization (WHO) are given in Table 1. The product of the concentration and the TEF is called the TCDD equivalent (TEQ). The TEQ of all dioxin-like CBs, dioxins and furans together is called total-TEQ.

## Sampling and Sample Pretreatment

PCBs are analyzed in all sorts of environmental matrices: air, seawater, freshwater, sediments, sewage sludge, benthic organisms, fish, marine mammals, human blood, adipose tissue, and others. All samples require a specific approach. Sampling of sediments and biota will be discussed because these belong to the most frequently analyzed matrices in environmental analysis. The analysis of PCBs in water is known to be extremely difficult because their solubility in water is low. Consequently, PCB concentrations in water are extremely low and the analysis is hindered by background contamination. PCB concentrations in air can be determined after sampling by filters or passive sampling systems.

Surface sediments can be collected intertidally or by means of a variety of grab samplers from a vessel. Because PCB concentrations in sediments can show a patchy distribution, several grabs from one location are normally combined to one pooled sediment

**Table 1** Toxic equivalency factors for dioxin-like compounds for humans and wildlife derived at a WHO meeting in Stockholm, Sweden, 15–17 Jun. 1997

Congener	Toxic equivalency factor (TEF)		
	Humans/ mammals	Fish <sup>a</sup>	Birds <sup>a</sup>
1,2,3,7,8-PeCDD	1	1	1 <sup>f</sup>
1,2,3,4,7,8-HxCDD	0.1 <sup>a</sup>	0.5	0.05 <sup>f</sup>
1,2,3,6,7,8-HxCDD	0.1 <sup>a</sup>	0.01	0.01 <sup>f</sup>
1,2,3,7,8,9-HxCDD	0.1 <sup>a</sup>	0.01 <sup>e</sup>	0.1 <sup>f</sup>
1,2,3,4,6,7,8-HpCDD	0.01	0.001	<0.001 <sup>f</sup>
OCDD	0.0001 <sup>a</sup>	—	—
2,3,7,8-TCDF	0.1	0.05	1 <sup>f</sup>
1,2,3,7,8-PeCDF	0.05	0.05	0.1 <sup>f</sup>
2,3,4,7,8-PeCDF	0.5	0.5	1 <sup>f</sup>
1,2,3,4,7,8-HxCDF	0.1	0.1	0.1 <sup>c,f</sup>
1,2,3,6,7,8-HxCDF	0.1	0.1 <sup>c</sup>	0.1 <sup>c,f</sup>
1,2,3,7,8,9-HxCDF	0.1 <sup>a</sup>	0.1 <sup>c,e</sup>	0.1 <sup>c</sup>
2,3,4,6,7,8-HxCDF	0.1 <sup>a</sup>	0.1 <sup>c</sup>	0.1 <sup>c</sup>
1,2,3,4,6,7,8-HpCDF	0.01 <sup>a</sup>	0.01 <sup>b</sup>	0.01 <sup>b</sup>
1,2,3,4,7,8,9-HpCDF	0.01 <sup>a</sup>	0.01 <sup>b,e</sup>	0.01 <sup>b</sup>
OCDF	0.0001 <sup>a</sup>	0.0001 <sup>b,e</sup>	0.0001 <sup>b</sup>
3,3',4,4'-TCB(77)	0.0001	0.0001	0.05
3,4,4',5-TCB(81)	0.0001 <sup>a,b,c,e</sup>	0.0005	0.1 <sup>e</sup>
3,3',4,4',5-PeCB(126)	0.1	0.005	0.1
3,3',4,4',5,5'-HxCB(169)	0.01	0.0005	0.001
2,3,3',4,4'-PeCB(105)	0.0001	<0.000005	0.0001
2,3,4,4',5-PeCB(114)	0.0005 <sup>a,b,c,d</sup>	<0.000005 <sup>b</sup>	0.0001 <sup>g</sup>
2,3',4,4',5-PeCB(118)	0.0001	<0.000005	0.00001
2',3,4,4',5-PeCB(123)	0.0001 <sup>a,c,d</sup>	<0.000005 <sup>b</sup>	0.00001 <sup>g</sup>
2,3,4',4,4',5-HxCB(156)	0.0005 <sup>b,c</sup>	<0.000005	0.0001
2,3,4',4,4',5'-HxCB(157)	0.0005 <sup>b,c,d</sup>	<0.000005 <sup>b,c</sup>	0.0001
2,3',4,4',5,5'-HxCB(167)	0.00001 <sup>a,d</sup>	<0.000005 <sup>b</sup>	0.00001 <sup>g</sup>
2,3,3',4,4',5,5'-HpCB(189)	0.0001 <sup>a,c</sup>	<0.000005	0.00001 <sup>g</sup>

<sup>a</sup>Limited data set.<sup>b</sup>Structural similarity.<sup>c</sup>QSAR modeling prediction from CYP1A induction (monkey, pig, chicken, or fish).<sup>d</sup>No new data from 1993 review.<sup>e</sup>*In vitro* CYP1A induction.<sup>f</sup>*In vivo* CYP1A induction after *in ovo* exposure.<sup>g</sup>QSAR modeling prediction from class-specific TEFs.

The dash indicates no TEF because of lack of data.

sample. Stainless steel materials should be used to avoid contamination from plasticizers such as phthalates. These can interfere in the gas chromatographic (GC) analysis. The grab samplers such as Van Veen sampler normally collect the upper few centimeters of sediment layers. Passive sediment traps can be used to collect samples of material sedimenting through the water column or resuspended from the seabed. In depositional areas, the changing fluxes of PCBs to the

seabed over time can be reconstructed by analysis of dated sediment cores. For this purpose cores of more than a meter in length may be required.

PCB concentrations in biota can vary substantially, particularly at higher levels of the food chain. It is therefore difficult to obtain a representative sample from a certain population. Benthic organisms, shellfish, or fishes are normally pooled to obtain a representative sample and to save expensive analyses of individual samples. PCBs are metabolized relatively slowly, which means that biota samples can be stored at  $-25^{\circ}\text{C}$  for relatively long periods (years) prior to analysis. However, changes in the fat composition may take place at those temperatures, but the influence of those processes on the final result of the PCB analysis is negligible compared to the analytical error. Biota samples can be taken in various ways: benthic organisms can be taken intertidally from rocks or beaches, or sampled by means of a grab dredge or trawl in deeper water. Fish can be taken by rod and line, or in a variety of trawls and other nets depending on the habitat and habits of the species under investigation. Electric fishery techniques are particularly useful for collecting eel samples. Guidelines describing the collection of biota and sediment samples for contaminant analysis have been issued by the Oslo and Paris Commissions.

Sediment samples are transferred to wide-mouthed, solvent-cleaned glass jars for storage in a freezer at  $-25^{\circ}\text{C}$  prior to analysis. The jars should not be overfilled because breakage can easily occur for samples with high-water content. The plastic lids of the jars should be covered inside with solvent-rinsed aluminum foil to avoid the introduction of plasticizers in the sample. Sediment samples can also be dried before analysis, but several procedures carry a risk for the integrity of the samples due to evaporation losses or introduction of interferences or cross-contamination.

Biota samples can be stored in a freezer at  $-25^{\circ}\text{C}$ . The tissues needed for analysis should be removed under appropriate conditions of cleanliness in the laboratory. The glass jars need to be treated in the same way as those used for sediment samples. The samples are normally extracted as wet samples, after drying with, e.g., sodium sulfate. Freeze-drying is not recommended because losses of lower chlorinated CBs can occur. In addition, freeze-drying carries a risk of cross-contamination.

## Extraction

### Soxhlet Extraction

Soxhlet extraction is a simple and effective method. It has been used for a wide range of samples like

soils, sediments, and animal and plant tissues. A wide variety of solvents like dichloromethane (DCM), pure or mixed with acetone or hexane, and acetone-hexane mixtures can be used. The use of nonpolar solvents only is not recommended. The minimum time needed for a regular Soxhlet extraction is normally ~8 hours. Sulfur present in the sediment and soil samples is also extracted, and must be removed by a later cleanup step.

### Solid-Phase Extraction

Solid-phase extraction (SPE) has been used for PCB containing air and water samples. XAD-2 in glass cartridges can be used to collect PCBs from air. XAD-2 can also be used to treat water samples. The amounts of water sampled are usually between a liter and hundreds of liters. This technique puts heavy demands on the cleaning procedure to ensure a low blank from the resin matrix. Column extraction can also be used for fish. After homogenization the sample is ground with anhydrous sodium sulfate ( $\text{Na}_2\text{SO}_4$ ), eventually together with some sand, packed in a chromatography column, and eluted with DCM.

### Supercritical Fluid Extraction

The use of supercritical fluid extraction (SFE) as an extraction technique is related to the unique properties of the supercritical fluid. These fluids have a low viscosity, high diffusion coefficients, low toxicity, and low flammability, all clearly superior to the organic solvents used in SPE extraction. The most common fluid used is carbon dioxide. SFE extractions of sediment samples have shown recoveries of >95% for all the individual PCBs. The separation of PCDDs from PCBs and chlorinated benzenes is difficult because of their similar solubility. An interesting development is the use of fat retainers. Samples, mixed in different weight ratios with, e.g., silica/silver nitrate 10% or basic alumina, can be placed in 7 ml extraction cells. The analytes are recovered by elution with 1.5–1.8 ml of hexane. With the correct fat-silica ratios and SFE conditions, no additional cleanup procedure is necessary for GC with an electron-capture detector (ECD). One drawback of SFE may be that the methods developed are valid for a specific matrix, but as soon as, e.g., the fat content of a biota sample or the type of lipids changes, the method has to be adapted. SFE is relatively complicated compared to other extraction techniques. In addition, the cell volumes are small, which limits the sample intake, and, with that, the detection limits. Finally, some reliable types of SFE equipment have recently been withdrawn from the market. This will have a substantial negative effect on the use of SFE in the near future.

### Pressurized Liquid Extraction

Pressurized liquid extraction (PLE), also called accelerated solvent extraction, has been applied more frequently during the last years. PLE can be used for extraction only, using organic solvents such as toluene or hexane/acetone. Good recoveries may be obtained, but the investment is relatively high (unless the apparatus are self-made). The gain in labor time compared to a simple system such as Soxhlet extraction is limited. However, the use of fat retainers offers the possibility of combining extraction and cleanup. This option leads to considerable savings in labor time, and is very attractive. The fat retainer ratios need to be optimized. Normally, an excess of fat retainer (e.g., silica) is required for achieving good recoveries and quantitative removal of the lipids. This has implications for the sample intake. Larger cells (up to 100 ml) have been introduced recently, which help to overcome possible problems with detection limits. Further developments may include the use of carbon for a selective extraction of dioxin-like PCBs, which would be even more attractive because direct injection of the dioxin-like PCB fraction into the gas chromatograph may be possible.

### Other Extraction Methods

Other extraction methods like (Ultra-Turrax) blending, ultrasonic, and solvent extractions are available. Most of these techniques show good recoveries for specific matrices, and after optimization of the extraction conditions. Microwave assisted extraction is another promising technique. Good recoveries for PCBs have been obtained. The extraction time is short, and samples can be extracted simultaneously. An overview of advantages and disadvantages of the various extraction techniques is given in Table 2.

### Cleanup

Because the amount of lipophilic material in the extract can affect the active surface of the stationary phase and degrade the resolving power of the GC columns, an effective cleanup procedure is essential. Cleanup of extracts can be performed in either a nondestructive or destructive way.

### Nondestructive Lipid Removal

Two methods are available for nondestructive lipid removal: gel permeation chromatography (GPC) and adsorption column chromatography. GPC works according to the principle of size exclusion. Smaller molecules such as PCBs are retained while larger molecules such as fat are eluted earlier. The PCBs in the sample can be isolated in this way from the lipids.



**Table 2** Evaluation of extraction techniques

	Ease of automation	Maximum sample intake	Extraction time	Cost – initial setup	Cost – running expenses	Ease of optimization	Labor time consumption
Soxhlet	Possible	Moderate/large	Moderate	Low	Low/moderate	Easy	Moderate
Blending/ultrasonic	Difficult	Large	Rapid	Moderate	Low	Moderate	High
Chemical modification	Difficult	Large	Slow	Very low	Very low	Moderate	High
Column extraction	Possible	Large	Slow	Low	Low	Difficult	Moderate
Microwave extraction	Difficult	Moderate/large	Rapid	Moderate	Low	Easy	Moderate
PLE	Easy	Moderate/large	Rapid	High	Low	Moderate	Low <sup>a</sup>
SFE	Easy	Small	Rapid	Very high	High	Difficult	Low <sup>a,b</sup>

<sup>a</sup>When combined with fat removal or in another selective way.<sup>b</sup>But high when analyzing samples with a varying composition or lipid content.

With GPC it is difficult to remove ‘all’ the lipids. The remaining traces of lipids have to be removed in a second cleanup procedure, e.g., on an additional silica column or by a second GPC step. Rigid Polymer Laboratories (PL) gels may offer the best result from the current choice of gels. GPC does not separate the PCBs from the other compounds in the same molecular range such as organochlorine pesticides. Therefore, an additional fractionation is often required.

Adsorption column cleanup procedures have been designed to cope with two types of interferences: the coextracted bulk components comprising predominantly lipids and coextracted organochlorine pesticides (described under fractionation). Alumina columns normally provide an excellent cleanup of PCBs from bulk components such as lipids.

### Destructive Lipid Removal

Treatment with sulfuric acid or concentrated base offers an alternative solution for the removal of lipids and other interferences. Because PCBs are generally resistant to sulfuric acid, concentrated sulfuric acid treatment is used for degradation of most aliphatic and many aromatic compounds in environmental samples. Sulfuric acid may affect other compounds such as some halogenated pesticides (e.g., dieldrin) that are determined together with PCBs in one extract.

Lipids can be saponified by heating the extract in a small volume of solvent with 20% ethanolic potassium hydroxide at ~70°C for 30 min. Saponification is not only used for lipids but is also used for the removal of sulfur from sediment extracts. The conditions of saponification are critical. Too high temperatures and too long saponification times can cause decomposition of higher chlorinated compounds such as hexa-deca PCBs, in particular when trace metals are present, e.g., in sediment samples. Metals can act as a catalyst. The chlorine in aromatic meta-ortho CB concentrations as these are normally ~100–1000-fold lower than those of the indicator CBs.

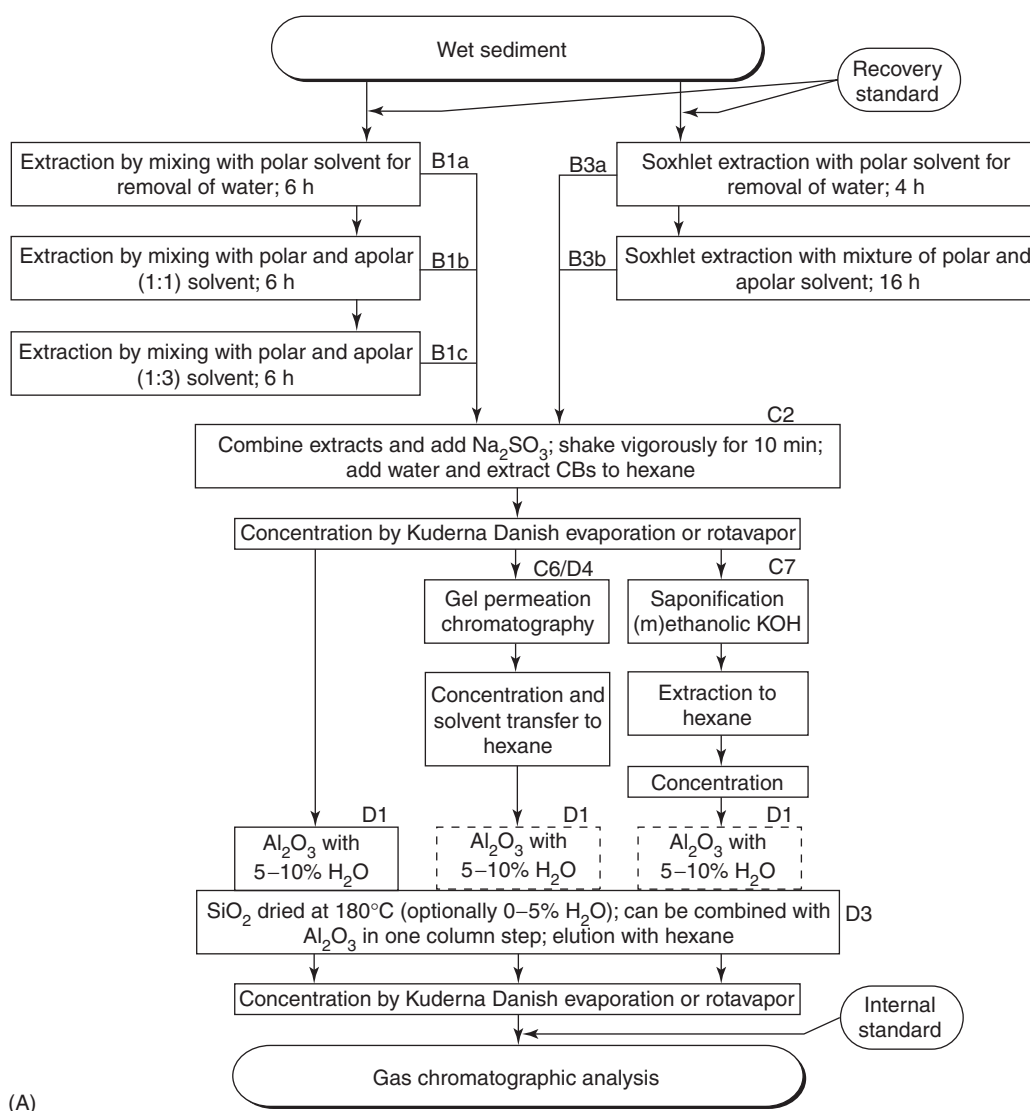
### Prefractionation

PCBs are normally present in the extract after the cleanup together with other halogenated compounds such organochlorine pesticides, brominated biphenyls, and diphenyl ethers. Therefore, it is often important to apply a prefractionation in order to avoid coelution problems. A prefractionation is also required to separate non-*ortho* CBs from the other CBs.

**Adsorption column chromatography** For the cleanup of PCBs in several matrices, alumina, silica, and Florisil columns, in different mesh sizes, levels of activity, and column sizes, can be used. Florisil, a

mixture of several inorganic oxides with  $\text{SiO}_2$  and  $\text{MgO}$  as the main substances, is one of the oldest materials used in the cleanup procedure for PCBs. Because Florisil is a mixture, the composition varies from batch to batch. Silica and alumina have been applied to clean extracts from air, water, sediments, rice bran oil, and animal or human tissue. Preheating and deactivation of the silica with a few per cent of water is necessary. Alumina does not require preheating, but 3–10% of water needs to be added. Disposable silica columns (SPE), precleaned with hexane, can also be used for the cleanup of PCB-containing samples. In Figure 1, a scheme of methods used for the isolation cleanup and fractionation of PCBs in wet and dry sediments is shown.

**Carbon column chromatography (CCC)** CCC exhibits excellent selectivity for those CBs for which a planar conformation is assumed. Different forms of carbon are being used to isolate these congeners. The sorbents used for CCC include activated carbon, activated carbon/polyurethane foam, activated carbon/glass fiber, activated carbon/silica gel, activated carbon/celite, activated carbon/Chromosorb W-HP, and porous graphitic carbon. Activated charcoal is a popular choice, but it has some serious drawbacks such as low efficiency, severe tailing of elution profiles, irreversible adsorption, and large batch-to-batch variations. The use of high-performance liquid chromatography (HPLC) with a porous graphitic carbon column may help to overcome these drawbacks.



**Figure 1** Methods used for the isolation, cleanup, and fractionation of PCBs in (A) wet sediment and (B) dry sediment. (Reprinted with permission from Smedes F and de Boer J (1997) Determination of chlorobiphenyls in sediments – analytical methods. *Trends in Analytical Chemistry* 16: 503–517; © Elsevier.)

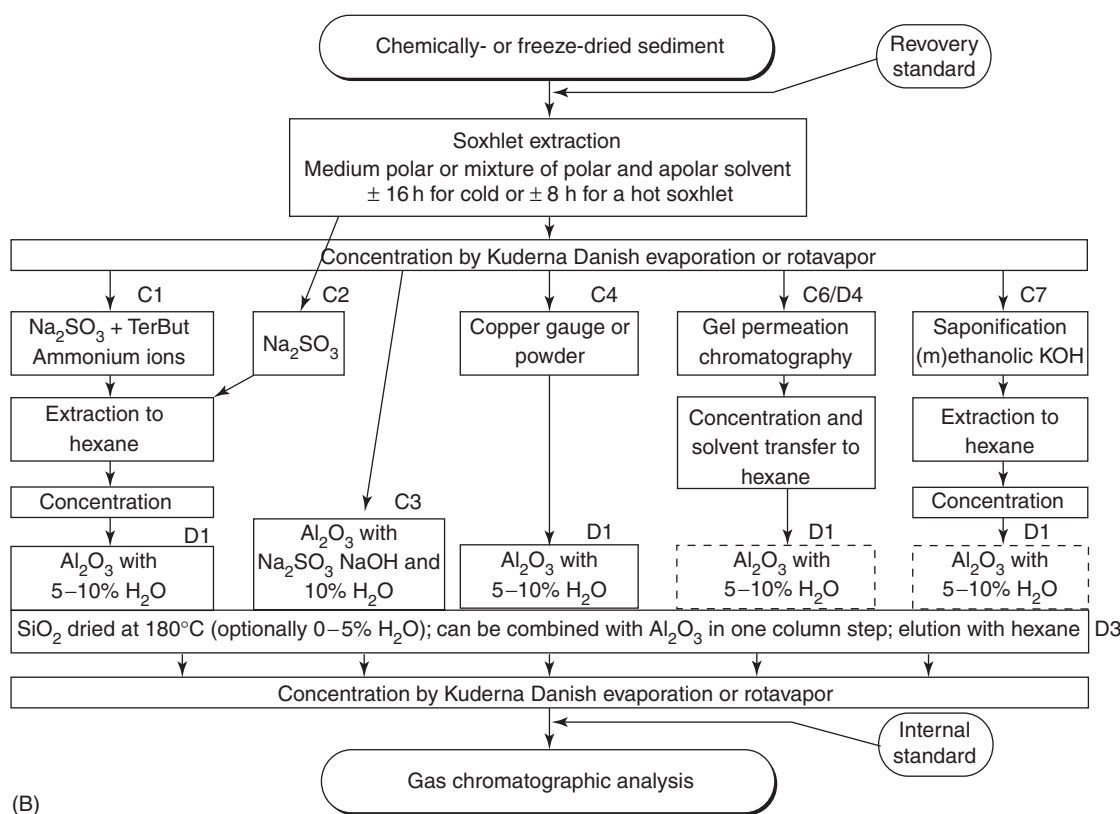


Figure 1 Continued

**Other HPLC columns** The separation of non-*ortho* CBs from *ortho* CBs is also possible by using an electron donor-acceptor HPLC method using a 2-(1-pyrenyl)ethyldimethylsilylated silica (PYE) column. The PYE column separates PCBs according to the number of chlorine atoms in the *ortho* position. The selectivity of the PYE phase may be explained by a charge transfer mechanism, in which electron-density acceptor (EDA) and donor regions of the CBs induce a change in the localization of the  $\pi$ -electron cloud of the pyrene moieties of the phase so that an EDA complex is formed. Separation of the non-*ortho* CBs on a PYE column takes less than 15 min. A 2,4-dinitroanilinopropyl silica (DNAP) phase for the isolation of the non-*ortho* CBs from the other CBs may provide another alternative.

## GC Analysis

PCBs are almost exclusively determined by GC. This is primarily due to the higher sensitivity of the GC detectors, in particular the ECD, as well as to the better separation that can be obtained by GC. Apart from a choice for injection or detection technique in GC, there are many possibilities for the type of columns used, both in single-column GC and in multicolumn approaches.

## Injection Techniques

The most frequently used injection techniques in PCB analysis are splitless and on-column injection. Split injection is not recommended because strong discrimination effects can occur. The presence of high-boiling compounds can increase the effect of discrimination. On-column injection is a technique in which the sample is deposited directly into the column with a syringe. On-column injection is well suited for high-boiling compounds. On-column injection yields better results, but is much more sensitive to contamination than splitless injection. Splitless injection is the most frequently used injection technique. In splitless injection a sample is injected in a hot liner while the splitter is closed. The sample evaporates and enters the GC column. During the time the splitter is closed, the oven is relatively cold, 80–100°C, dependent on the solvent. The ideal temperature is ~10°C below the boiling point of the solvent. After opening the splitter, the oven is rapidly heated and the chromatographic analysis begins. If the liner volume is not large enough, memory effects can occur due to contamination of the gas tubing attached to the injector. Large volume injection is becoming more popular. The increase of the injection volume is directly related to a

decrease of sample intake. Volumes up to 1 ml can be introduced at rates between 1 and 2000  $\mu\text{l min}^{-1}$ . Disadvantages of large volume injection are the stronger influence of solvent impurities in the chromatogram and contamination of the column and detector. Mass spectrometric (MS) detection with negative chemical ionization (NCI) is in particular sensitive for contamination due to large volume injection. A regular cleaning of the ion source may be necessary.

### Detection Techniques

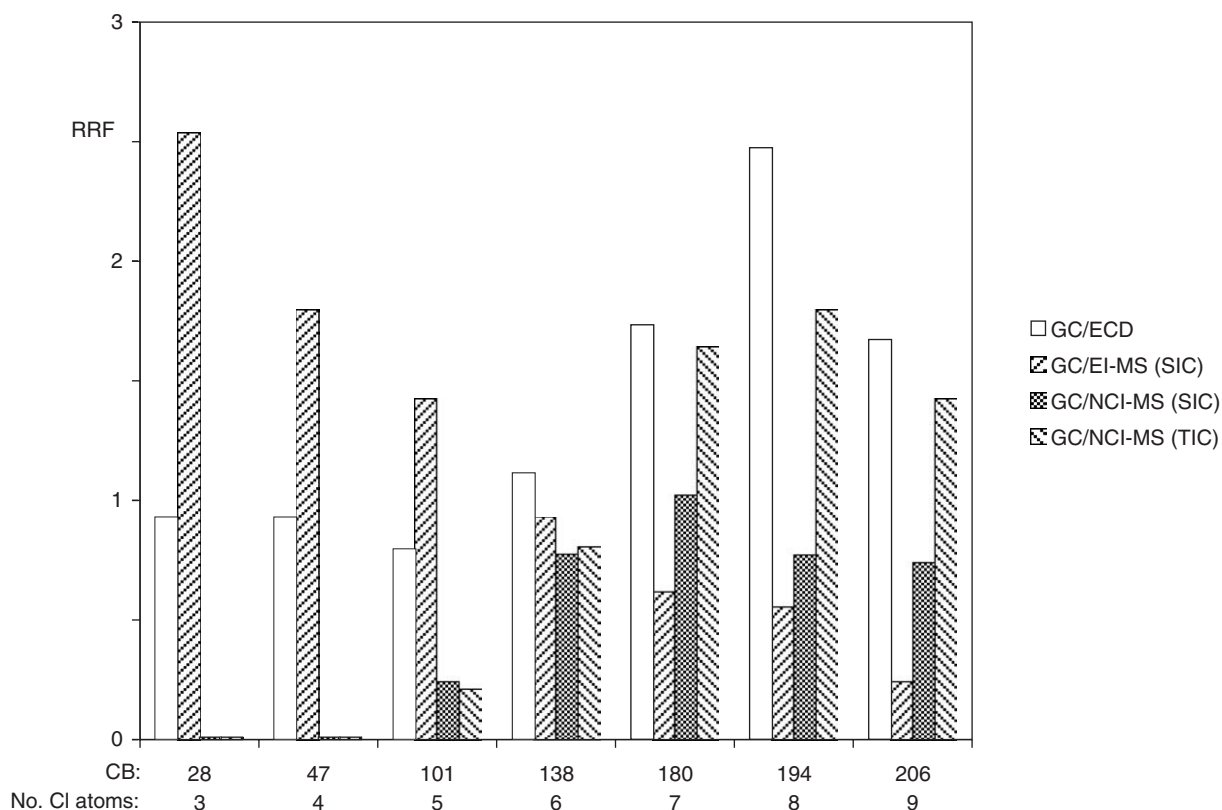
The most frequently used detection system for the analysis of PCBs in various matrices is the ECD. This detector is selective for halogenated compounds. The very high sensitivity of ECD for PCBs theoretically allows detection limits below 1  $\text{ng kg}^{-1}$ . Its extreme sensitivity, on the other hand, makes ECD vulnerable to dirt and overloading. The ECD response varies from one detector to another. The limited linearity is a particular problem. Although linear ranges of several decades are often claimed by manufacturers, in practice the linear range of all ECDs is substantially lower than one decade. Consequently, the use of multilevel calibrations is essential.

The second most frequently used detection system is MS. There are several modes in which the MS is employed to detect PCBs: electron impact (EI) ionization, chemical ionization with negative-ion detection (NCI), or chemical ionization with positive-ion detection (PCI). EI mass spectra are fairly reproducible. They show relatively intense molecular ions and the natural isotopic distribution of chlorine gives rise to typical clusters, which can easily be recognized. The response factors in EI-MS between isomers differ by no more than about twofold, which is much less than in ECD. It is therefore possible to use one surrogate standard within each isomer group. The sensitivity of CBs in NCI increases dramatically with an increase in number of chlorine atoms. The opposite occurs for the CBs determined by PCI. The choice, NCI or PCI, depends on the PCBs to be analyzed, but as most PCBs found in the environment are relatively highly chlorinated, NCI is more often used. Actually, PCI applications for PCB analysis are rather scarce. Absolute detection limits of GC-NCI-MS are  $\sim 100$ , 100, and 50 fg for the CBs 77, 126, and 169, respectively. Compared to GC-ECD, GC-NCI-MS has a more than 10-fold higher sensitivity for CBs with more than four chlorine atoms. EI-MS has a  $\sim 10$ -fold lower sensitivity than ECD. The low environmental concentrations of non- and mono-*ortho* CBs will in general not allow a proper determination by GC-EI-MS. **Figure 2** shows relative response factors (RRFs) of six CBs with a

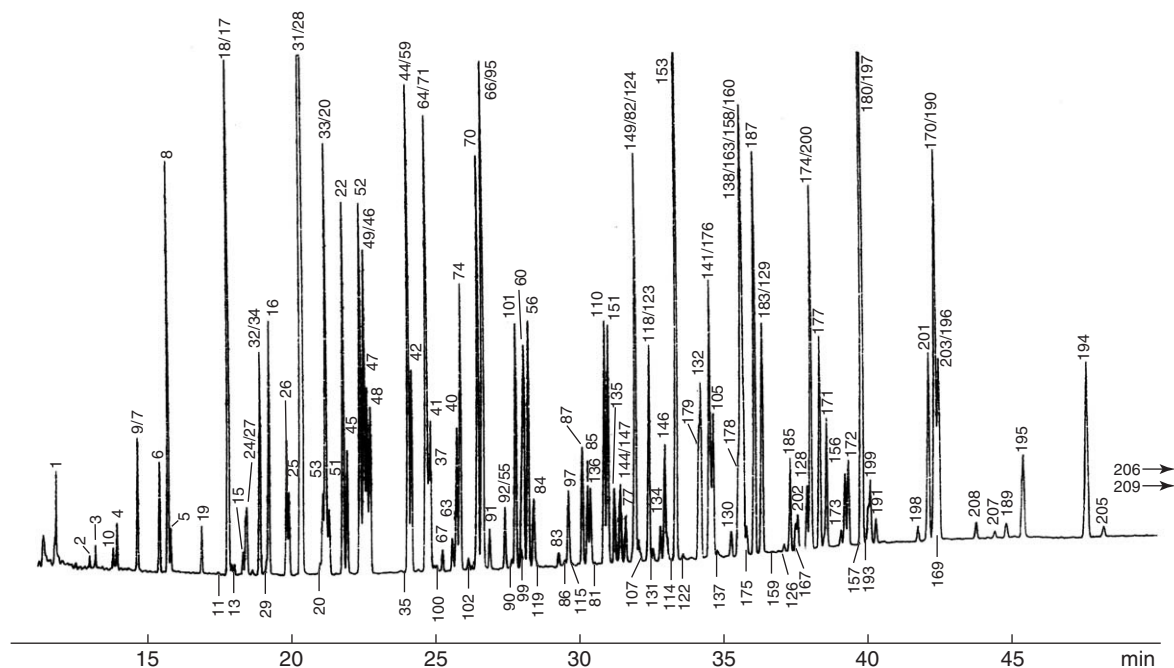
different number of chlorine atoms for GC-ECD, GC-EI-MS, GC-NCI-MS (total ion chromatogram), and GC-NCI-MS (single ion chromatograms). The highly different NCI signals of CBs with less than five chlorine atoms and higher chlorinated CBs are obvious.

### GC Separation

**Single-column GC** For the analysis of CBs several GC columns can be used. The most frequently used GC column is the SE-54 column (5% diphenyl-1% vinyl dimethyl siloxane). A complete set of retention time data on an SE-54 column is available for all 209 CBs. Other 5% phenyl phases comparable to the SE-54 phase are: CP-Sil-8, DB-5, HP-5, Ultra-2, SE-52, and OV-73. One disadvantage of the SE-54 phase is the number of coeluting PCB congeners: CBs 15/17/18, 28/31, 37/42/59, 47/48/75, 52/73, 66/95, 67/100, 77/110, 82/151, 84/90/101, 87/115, 105/132/153, 114/122/131, 118/123/148/149, 126/129/178, 137/176, 138/158/160/163, 156/171/202, 157/173/201, 170/190, 195/208, and 196/203. Longer columns and smaller diameters may help to resolve some of these congener pairs. A CP-Sil 19 (14% cyanopropyl phenyl-1% vinyl dimethyl siloxane) column offers better possibilities for the separation of the most important PCBs than a SE-54 type column. Comparable phases to CP-Sil-19 are: OV-1701, DB-1701, BP-10, and SFB-7. A single CP-Sil 19 column (85% methyl-7% phenyl-7% cyanopropyl-1% vinyl polysiloxane) offers accurate analyses of 84 congeners by GC-ECD (**Figure 3**). Coelution on a CP-Sil-19 column also occurs: CBs 7/9, 16/29, 17/18, 20/33, 24/27, 32/34, 37/41/64/71, 44/59, 46/49, 55/92, 60/99, 66/95, 81/87/97/115, 82/124/149, 84/119, 107/123, 114/146, 129/183, 132/179, 138/158/160/163/178, 141/176, 144/147, 157/180/197, 169/196/203, 170/190, and 174/200. Other stationary phases used for the analysis of PCB congeners are OV-210 (50% trifluoropropyl methylsiloxane) or comparable phases like QF-1, SP-2401, and RSL-400. Also, 25% phenyl-25% cyanopropyl (OV-225) and comparable phases such as CP-Sil-43, DB-225, BP-15, XE-60, and CS-5 have been used. The highly polar *bis*-cyanopropyl (CP Sil-88) column, commonly used in dioxin analysis has the drawback of a low maximum temperature of 240°C, which increases the analysis time. This phase offers good separation for the most toxic PCBs (77, 126, and 169). None of the stationary phases is able to resolve all 209 congeners in single-column GC. **Table 3** shows the number of separated PCBs on 11 different phases. This table shows that the low bleeding HT-8 column (8% phenyl polycarborane siloxane) offers the best resolution (**Figure 4**).



**Figure 2** Relative response factors (RRF) of six CBs with a different number of chlorine atoms for GC/ECD, GC/EI-MS, and GC/NCI-MS. (Reprinted with permission from Wester PG *et al.* (1995) *Environmental Science and Technology* 30: 473–480; © American Chemical Society.)



**Figure 3** ECD chromatogram of an Aroclor mixture (A1016, A1232, A1248, A1260, 1:1:1:1) on a 50 m x 0.22 mm x 0.26 μm CP Sil 19 column. (Reproduced with permission from Larsen B *et al.* (1992) *International Journal of Environment Analytical Chemistry* 47: 47–68; © Taylor & Francis Ltd. <http://www.tandf.co.uk/journals>)

It is possible to separate the (+) and (–) enantiomers of nine stable atropisomeric CBs (84, 91, 95, 132, 135, 136, 149, 174, and 176) on a Chirasil-Dex (per-methylated  $\beta$ -cyclodextrin) column.

**Multicolumn GC** Multicolumn GC techniques can be divided in parallel column, serial column, and multidimensional (MD) GC. Of these three techniques, MDGC is the most powerful and most frequently used technique.

**Table 3** Number of CBs that can be analyzed with less than 10% interference from any coeluting congener on 11 thoroughly characterized stationary phases<sup>a</sup>

Name	Indicator CBs		Priority CBs		Aroclor CBs	
	MS	ECD	MS	ECD	MS	ECD
ACTYL <sup>b</sup>	6	6	33	31	100	ND
SIL-5	6	5	29	27	102	84
SIL-8	5	5	24	22	81	65
SE-54 <sup>c</sup>	5	4	18	15	75	69
SIL-13	5	4	25	23	104	93
HT-5	6	5	24	22	91	78
HT-8	7	6	36	28	157	106
DB-17	5	4	22	19	93	86
SIL-19	5	4	25	20	100	84
BPX70	4	2	19	13	89	72
SIL-88	5	2	24	18	99	74

<sup>a</sup> Unless stated otherwise, the temperature program was optimized for a 1 h total run time.

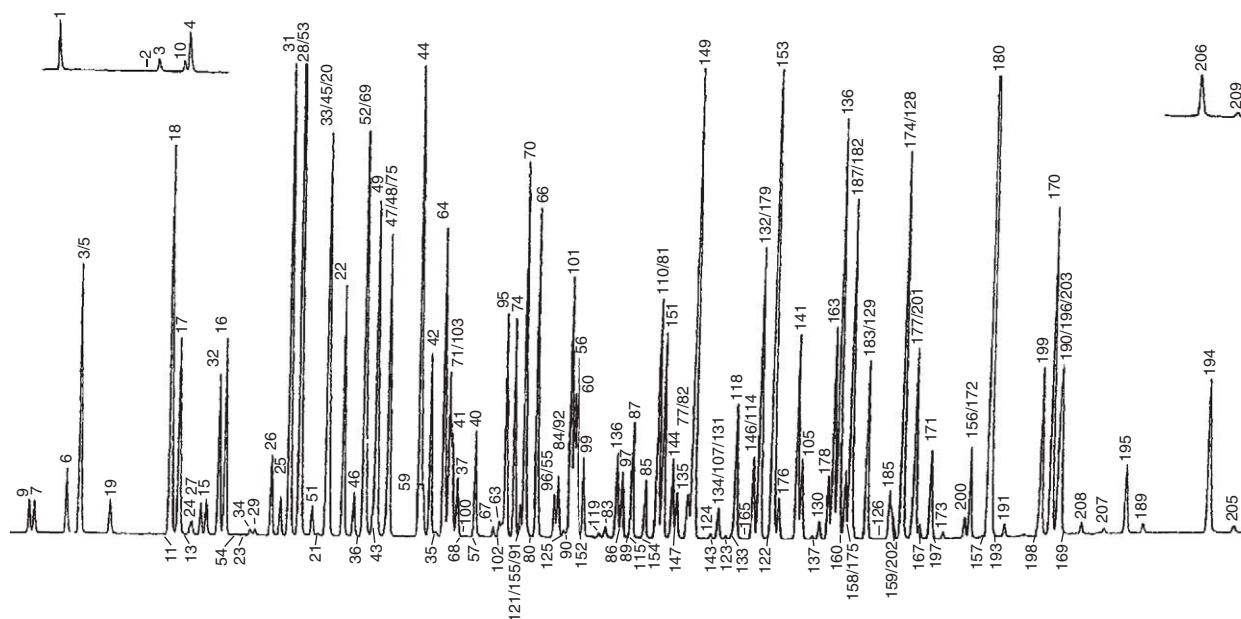
<sup>b</sup> 1.5 h total run time.

<sup>c</sup> 2 h total run time.

Parallel column GC is a method in which two columns are used in parallel and are connected via a glass T-split to a retention gap that is led into the injector. At the other end the columns are each connected to a detector. In this way the PCB analyses can be carried out on two different stationary phases at the same time. An advantage of dual-GC is that only one GC oven, one autosampler, and one injector are necessary.

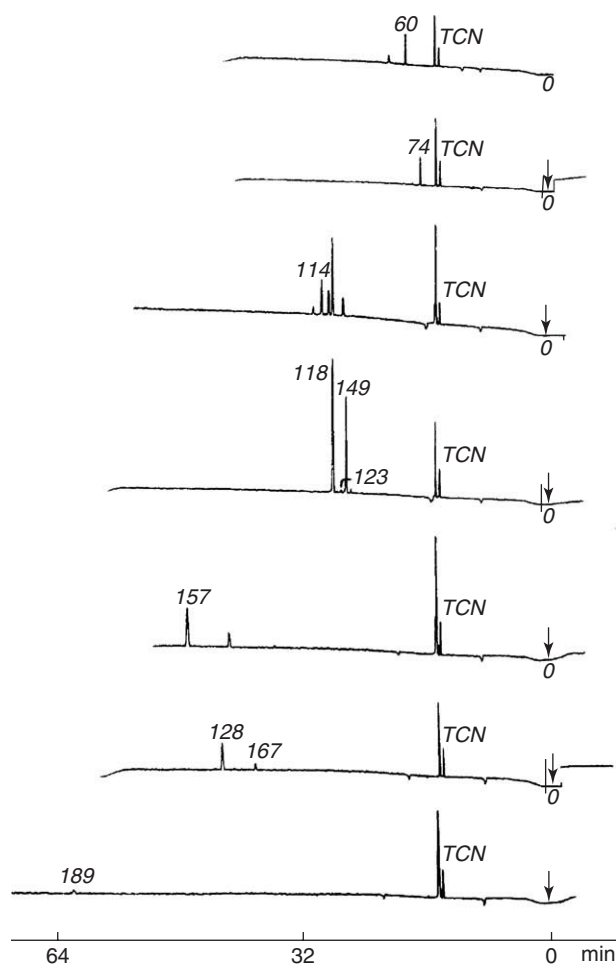
In serial GC two columns (CP-Sil-8 and HT-5) are coupled in series via a quick-seal glass tube. In this way more separation power can be obtained than with single-column GC, although this depends of the selection of the two columns. The second column may sometimes also destroy separations achieved at the first column. A combination of parallel- and serial-coupled columns is also possible.

The resolving power of capillary gas chromatography can substantially be increased by using MDGC. MDGC can be distinguished in two techniques, heart-cut MDGC and comprehensive MDGC. MDGC is one of the most effective techniques to separate PCBs and to avoid interference from other congeners or other compounds. In heart-cut MDGC normally a nonpolar phase such as SE-54 or CP-Sil-8 is used in the first column to make the initial, well-characterized separation. The column flow is switched by use of a valveless switching technique for the duration of the elution of the peak(s) of interest only to the second column of a different, usually more polar character, preferably installed in an independently controlled oven. This technique offers retention



**Figure 4** ECD chromatogram of an Aroclor mixture (A1016, A1232, A1248, A1260, 1:1:1:1) at an 60 m × 0.25 mm × 0.15 μm HT-8 (1,7-dicarba-closo-dodecarborane phenyl methyl siloxane) column. (Reprinted with permission from Larsen BR (1995) *Journal of High Resolution Chromatography* 18: 141; © Wiley-VCH.)





**Figure 5** Separate heart-cuts of PCBs 60, 74, 114, 123, 157, 167, and 189 from Aroclor 1254; TCN = tetrachloronaphthalene (internal standard). (Reprinted with permission from de Boer *J et al.* (1995) *Analytica Chimica Acta* 300: 155; © Elsevier.)

times with a precision of tenths of seconds. During one run multiple heart-cuts can be made. **Figure 5** shows the chromatograms of the heart-cuts of the seven mono-*ortho*-PCBs of interest present in Aroclor.

The use of a heart-cut technique is relatively laborious when more peaks per run need to be checked for coelution. When the speed of the secondary separation is high enough to separate a cut from the first separation while the next cut is being collected, it is possible to record a complete set of secondary chromatograms. From the secondary chromatograms the complete two-dimensional chromatogram can be constructed. A method capable of doing this is called comprehensive MDGC or GC  $\times$  GC. The heart of the comprehensive MDGC is the modulator of which two basic types are available: thermal modulators and cryogenic modulators. The latter type is meanwhile available in different versions, using either carbon dioxide or nitrogen as cooling agent. The peaks

produced by GC  $\times$  GC are extremely narrow, which improves the sensitivity, but which requires a detector that is able to respond very quickly to the analyte. Flame ionization detection (FID) would be suitable but obviously lacks the sensitivity required for PCB analysis. A micro-ECD is able to handle the GC  $\times$  GC peaks and has enough sensitivity for PCBs. However, an even smaller cell volume would be preferred, as still some peak tailing can be observed. A time-of-flight-MS may be the ideal detector as it couples a high selectivity to a rapid response, and, with the GC  $\times$  GC, with a high sensitivity. An example of a separation of PCBs is given in **Figure 6**. During the last years GC  $\times$  GC has become much more robust and has been used in many different applications. It is expected that software packages for integration will soon be available.

## Quality Assurance

A number of measurements should ensure a sufficient quality of the analysis. Four main areas can be identified: calibrants and calibration, system performance, control of extraction and cleanup, and control of long-term stability.

### Calibrants and Calibration

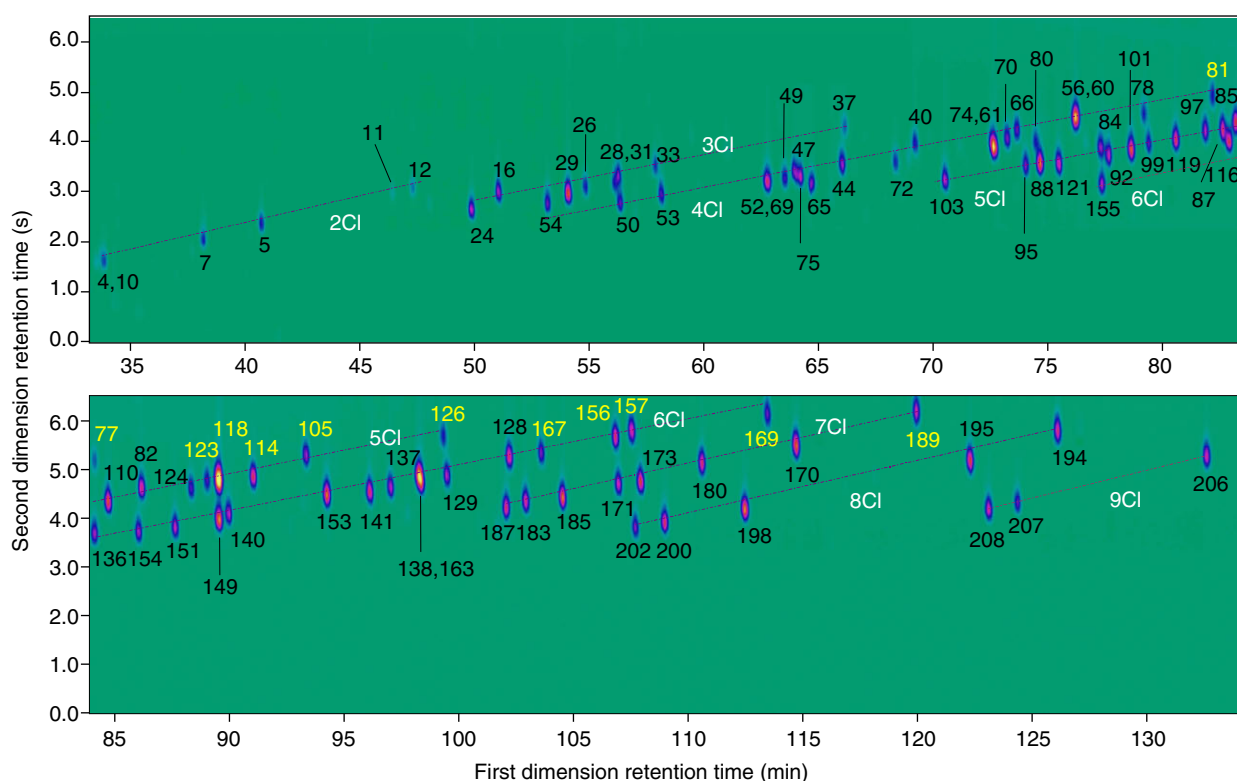
A CB determination should preferably be carried out using calibration solutions prepared from crystalline, if possible certified CBs. Utmost attention should be paid to the preparation of calibration solutions. Always two independent solutions should be prepared simultaneously. Calibration solutions should preferably be stored in ampoules.

### System Performance

By regularly measuring the resolution of two closely eluting CBs, the performance of the system can be monitored. A decreasing resolution points to deteriorating conditions. The signal-to-noise ratio gives information on the condition of the detector. A dirty ECD detector can be recognized from a higher signal together with a lower signal-to-noise ratio.

### Control of Extraction and Cleanup

For the control of extraction and cleanup it is recommended to let a standard solution pass the whole procedure, from the extraction to the final determination. This standard also gives an indication of the recovery of the sample series.  $^{13}\text{C}$  labeled CBs can be used as internal standards provided MS detection is used. In addition, to all samples and to the standard solution mentioned above, a recovery standard should be added to check whether in



**Figure 6** GC × GC- $\mu$ ECD chromatogram of a mixture of 90 CBs with a HP-1–HT-8 column combination. (Reprinted with permission from Korytar P *et al.* (2002) *Journal of Chromatography A* 958: 203; © Elsevier.)

individual samples losses have occurred. It is recommended that this recovery standard is not used to correct any data, but only for a control on the whole procedure. If major losses have occurred, results should not be reported. CB 29 is suggested as recovery standard. Because of its high volatility, losses due to evaporation are easily detected. CB 29 elutes relatively late from alumina and silica columns. Therefore, also losses due to cleanup may easily be detected with this CB. Small peaks that may be present in the chromatogram at the retention time of CB 29 do not hinder the use of this CB because the recovery standard only controls major errors in extraction or cleanup. It is recommended to use a syringe standard, added in a fixed volume or weight to all standards and sample extracts, to control the final concentration or dilution step and the injection. Appropriate syringe standards are the CBs 29, 112, 155, and 198.

### Long-Term Stability

One internal reference material should be included in each series of samples. This sample should be taken from a large, homogeneous batch of sediment or fish or other frequently analyzed material that can serve as an internal reference material over a long period.

A quality chart should be recorded. In case the warning limits are exceeded, the method used should be checked on possible errors. When the alarm limits are exceeded, the results obtained should not be reported. It is recommended that approximately once each month a certified reference material is analyzed. A laboratory analyzing CBs should take part on a regular basis in interlaboratory studies on the determination of CBs.

*See also:* **Extraction:** Solvent Extraction Principles. **Gas Chromatography:** Multidimensional Techniques; Mass Spectrometry. **Lipids:** Overview. **Quality Assurance:** Quality Control; Interlaboratory Studies; Reference Materials; Clinical Applications.

## Further Reading

- Bertsch W (2000) Two-dimensional gas chromatography. Concepts, instrumentation, and applications – part 2: comprehensive two-dimensional gas chromatography. *Journal of High Resolution Chromatography* 23: 167–181.
- de Boer J (2001) Polychlorinated biphenyls. In: Kleiböhmer W (ed.) *Environmental Analysis, Handbook of Analytical Separation*, vol. 3, chapter 7, pp. 237–267. Amsterdam: Elsevier.

- de Boer J and Law RJ (2003) Developments in the use of chromatographic techniques in marine laboratories for the determination of halogenated contaminants and polycyclic aromatic hydrocarbons. *Journal of Chromatography* 1000: 223–251.
- Dallüge J, Beens J, and Brinkman UATH (2003) Comprehensive two-dimensional gas chromatography: a powerful and versatile tool. *Journal of Chromatography* 1000: 69–108.
- Frame G (1997) A collaborative study of 209 PCB congeners and 6 Aroclors on 20 different HRGC columns. *Fresenius Journal of Analytical Chemistry* 357: 714–722.
- de Geus HJ, de Boer J, and Brinkman UATH (1998) Two-dimensional gas chromatography. In: Meyers RA (ed.) *Encyclopedia of Environmental Analysis Remediation*, pp. 4909–4921. New York: Wiley.
- Lang V (1992) Polychlorinated biphenyls in the environment. *Journal of Chromatography* 595: 1–43.
- Law RJ and de Boer J (1995) Quality assurance of analysis of organic compounds in marine matrices: application to analysis of chlorobiphenyls and polycyclic aromatic hydrocarbons. In: Quevauviller Ph (ed.) *Quality Assurance for Environmental Monitoring*, pp. 129–156. Weinheim: VCH Verlag.
- Mullin MD, Pochini C, McGrindle S, *et al.* (1984) High resolution PCB analysis: synthesis and chromatographic properties of all 209 PCB congeners. *Environmental Science and Technology* 18: 468–476.
- Pearson CR (1982) Halogenated aromatics. In: Hutzinger O (ed.) *The Handbook of Environmental Chemistry*, vol. 3, part B: *Anthropogenic Compounds*, p. 90. Heidelberg: Springer.
- Quevauviller Ph and Maier EA (eds.) (1999) *Interlaboratory Studies and Certified Reference Materials for Environmental Analysis – The BCR Approach*. Amsterdam: Elsevier.
- Wells DE and Hess Ph (2000) Methods for the determination and evaluation of chlorinated biphenyls (CBs) in environmental matrices. In: Barceló D (ed.) *Sample Handling and Trace Analysis of Pollutants*. Amsterdam: Elsevier.

# POLYCYCLIC AROMATIC HYDROCARBONS

Contents

**Determination**

**Environmental Applications**

## Determination

**A N Gachanja**, African Institute for Capacity Development, Nairobi, Kenya

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Polycyclic aromatic hydrocarbons (PAHs) are a complex class of condensed multinumbered benzenoid-ring compounds (three or more fused benzene rings in linear, angular, or cluster arrangement, which sometimes include a five-membered ring) containing only carbon and hydrogen atoms. Polycyclic aromatic compounds (PACs) include PAHs, and are compounds that have the basic PAH structure and substituent moieties such as alkyl, amino, chloro, cyano, hydroxyl, or thiol groups, and/or containing atoms such as nitrogen, oxygen, or sulfur in the aromatic structure. PACs are ubiquitous environmental pollutants found in air, water, and soil (including sediment). Structures of some PAHs and representative PACs are shown in **Figure 1**. This article will

briefly review the toxicology, sources, and modes of formation of PAHs and the analytical methods used for PAH analysis.

## Toxicology

The potential mutagenic and carcinogenic effects of PAHs are the basic reasons of concern for the levels of these compounds in the environment. The first cancers to be associated with PAH-containing substances were through skin contact that resulted in scrotal cancer of chimney sweeps and skin cancer among mule spinners in the cotton, wool, and jute industries.

Mammals and many lower organisms metabolize PAHs primarily by enzymatic oxygenation to epoxides, phenols, dihydrodiols, quinones, and water-soluble conjugates in an attempt to make them more soluble and thus facilitate their excretion from the organism. In this way, the precarcinogen is converted to the ultimate carcinogen that covalently binds to the information molecules (e.g., DNA, RNA) and leads to carcinogenesis (tumor formation). The biological activity of PAHs is dependent upon molecular structure and thus isomers show diverse carcinogenic activity as shown in **Table 1**. Studies

- de Boer J and Law RJ (2003) Developments in the use of chromatographic techniques in marine laboratories for the determination of halogenated contaminants and polycyclic aromatic hydrocarbons. *Journal of Chromatography* 1000: 223–251.
- Dallüge J, Beens J, and Brinkman UATH (2003) Comprehensive two-dimensional gas chromatography: a powerful and versatile tool. *Journal of Chromatography* 1000: 69–108.
- Frame G (1997) A collaborative study of 209 PCB congeners and 6 Aroclors on 20 different HRGC columns. *Fresenius Journal of Analytical Chemistry* 357: 714–722.
- de Geus HJ, de Boer J, and Brinkman UATH (1998) Two-dimensional gas chromatography. In: Meyers RA (ed.) *Encyclopedia of Environmental Analysis Remediation*, pp. 4909–4921. New York: Wiley.
- Lang V (1992) Polychlorinated biphenyls in the environment. *Journal of Chromatography* 595: 1–43.
- Law RJ and de Boer J (1995) Quality assurance of analysis of organic compounds in marine matrices: application to analysis of chlorobiphenyls and polycyclic aromatic hydrocarbons. In: Quevauviller Ph (ed.) *Quality Assurance for Environmental Monitoring*, pp. 129–156. Weinheim: VCH Verlag.
- Mullin MD, Pochini C, McGrindle S, *et al.* (1984) High resolution PCB analysis: synthesis and chromatographic properties of all 209 PCB congeners. *Environmental Science and Technology* 18: 468–476.
- Pearson CR (1982) Halogenated aromatics. In: Hutzinger O (ed.) *The Handbook of Environmental Chemistry*, vol. 3, part B: *Anthropogenic Compounds*, p. 90. Heidelberg: Springer.
- Quevauviller Ph and Maier EA (eds.) (1999) *Interlaboratory Studies and Certified Reference Materials for Environmental Analysis – The BCR Approach*. Amsterdam: Elsevier.
- Wells DE and Hess Ph (2000) Methods for the determination and evaluation of chlorinated biphenyls (CBs) in environmental matrices. In: Barceló D (ed.) *Sample Handling and Trace Analysis of Pollutants*. Amsterdam: Elsevier.

# POLYCYCLIC AROMATIC HYDROCARBONS

Contents

**Determination**

**Environmental Applications**

## Determination

**A N Gachanja**, African Institute for Capacity Development, Nairobi, Kenya

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

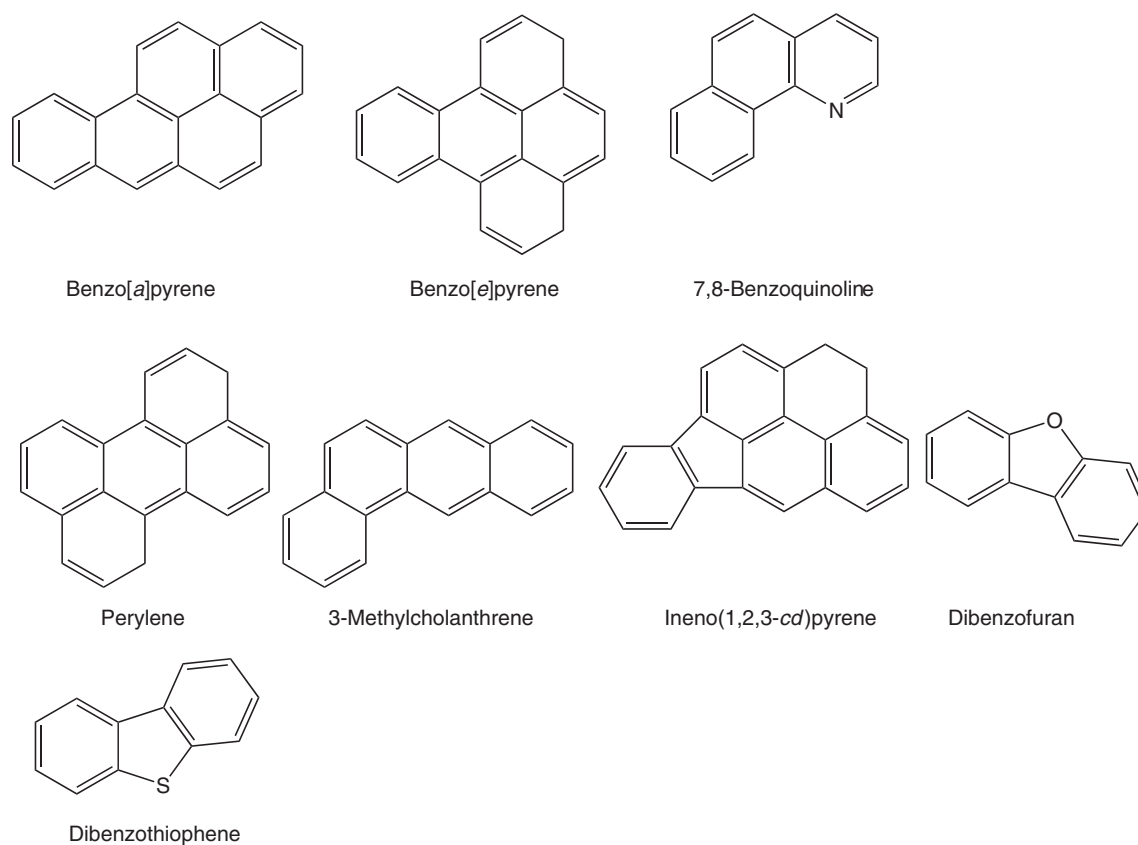
Polycyclic aromatic hydrocarbons (PAHs) are a complex class of condensed multinumbered benzenoid-ring compounds (three or more fused benzene rings in linear, angular, or cluster arrangement, which sometimes include a five-membered ring) containing only carbon and hydrogen atoms. Polycyclic aromatic compounds (PACs) include PAHs, and are compounds that have the basic PAH structure and substituent moieties such as alkyl, amino, chloro, cyano, hydroxyl, or thiol groups, and/or containing atoms such as nitrogen, oxygen, or sulfur in the aromatic structure. PACs are ubiquitous environmental pollutants found in air, water, and soil (including sediment). Structures of some PAHs and representative PACs are shown in **Figure 1**. This article will

briefly review the toxicology, sources, and modes of formation of PAHs and the analytical methods used for PAH analysis.

## Toxicology

The potential mutagenic and carcinogenic effects of PAHs are the basic reasons of concern for the levels of these compounds in the environment. The first cancers to be associated with PAH-containing substances were through skin contact that resulted in scrotal cancer of chimney sweeps and skin cancer among mule spinners in the cotton, wool, and jute industries.

Mammals and many lower organisms metabolize PAHs primarily by enzymatic oxygenation to epoxides, phenols, dihydrodiols, quinones, and water-soluble conjugates in an attempt to make them more soluble and thus facilitate their excretion from the organism. In this way, the precarcinogen is converted to the ultimate carcinogen that covalently binds to the information molecules (e.g., DNA, RNA) and leads to carcinogenesis (tumor formation). The biological activity of PAHs is dependent upon molecular structure and thus isomers show diverse carcinogenic activity as shown in **Table 1**. Studies



**Figure 1** Structures of some polycyclic aromatic hydrocarbons.

**Table 1** Toxic equivalency factors for polycyclic aromatic hydrocarbons

Compound	Nielsen <i>et al.</i> <sup>a</sup>	USEPA <sup>b</sup>	Nisbet and LaGoy <sup>c</sup>
Benzo[a]pyrene	1.0	1.0	1.0
Benz[a,h]anthracene	1.1	1.0	5.0
Benzo[ghi]perylene	0.02	Not done	0.01
Benz[a]anthracene	0.005	0.1	0.1
Chrysene	0.03	0.001	0.01
Benzo[b]fluoranthene	0.1	0.1	0.1
Benzo[k]fluoranthene	0.05	0.001	0.1

<sup>a</sup>Best estimates of carcinogenic potencies of various PAHs, relative to BaP, based on carcinogenicity studies in experimental animals using oral, pulmonary, and skin applications of PAHs. Nielsen T, Jorgensen HE, Poulsen M, Jensen AB, Schramm J, and Tonnesen J (1985) Traffic PAH and other mutagens in air in Denmark. Miljøprojekt 285, Danish Environmental Protection Agency, Copenhagen, Denmark.

<sup>b</sup>USEPA, Provisional guidance for Quantitative risk Assessment of PAH, EPA/600/R-93/089.

<sup>c</sup>Nisbet KT and LaGoy PK (1992) *Regulatory Toxicology and Pharmacology* 16: 290.

employing animal bioassays have shown that dibenzo[a,b]anthracene, dibenzo[a,b]pyrene, dibenzo[a,l]pyrene, benzo[a]pyrene, benzo[b]fluoranthene, 5-methylchrysene, 7H-dibenzo[c,g]carbazole, 6-nitro-

chrysene, and dinitropyrenes and dinitrofluoranthenes are the strongest carcinogenic PAHs while benzo[fluoranthene]s are moderately carcinogenic.

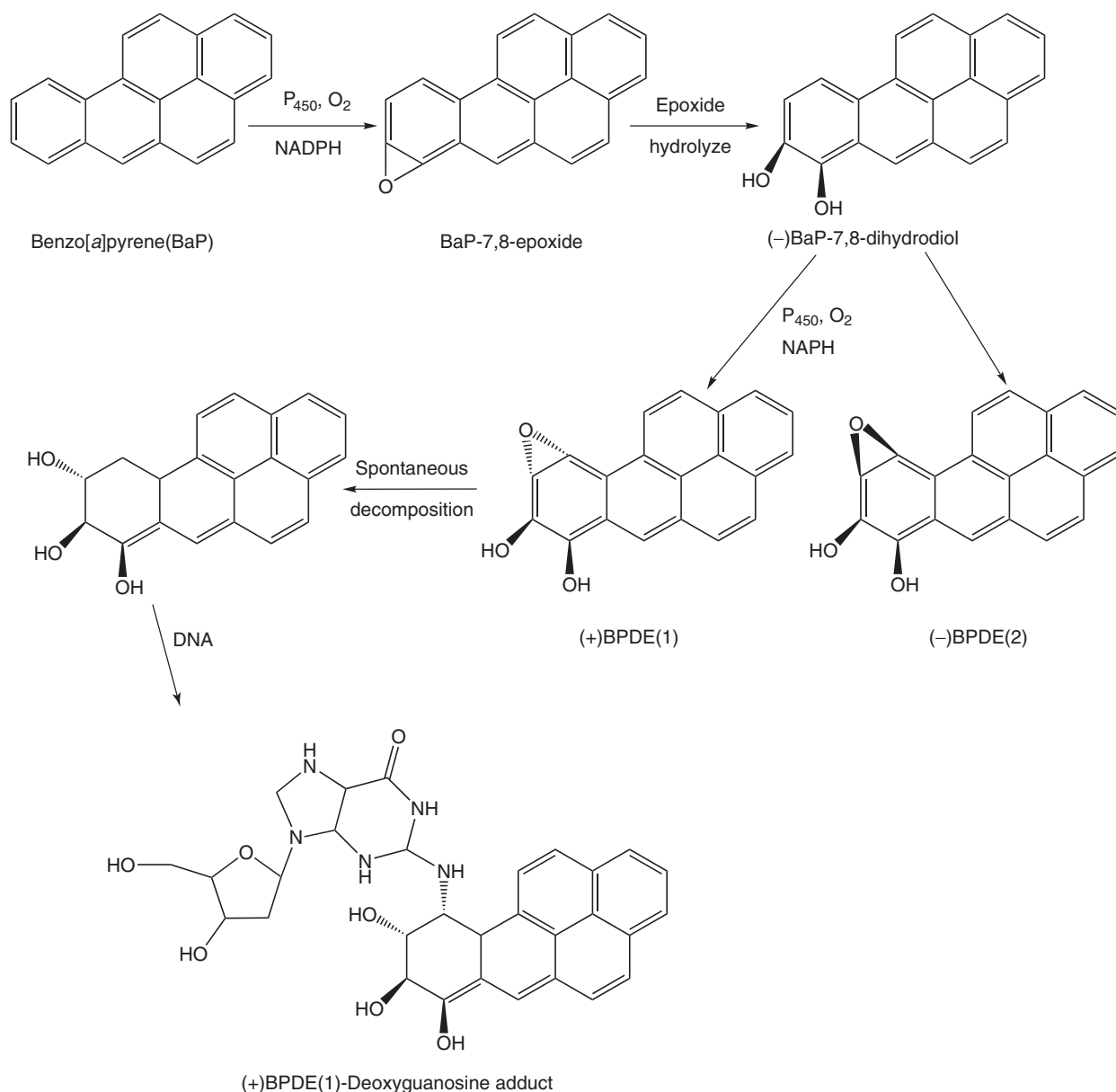
Though benzo(a)pyrene (BaP) accounts for only a small fraction of the total carcinogenicity of any material or environment (see Table 1), it is the most widely reported and studied PAH. This has been due to its ubiquity and abundance, high carcinogenicity, and relative ease of detectability compared with other carcinogenic PAHs, particularly with luminescence detection. BaP has also been used as an indicator compound for the presence of PAHs.

Relative to other tissues, unmetabolized BaP has been found localized extensively in the mammary gland and body fat after a single feeding of the carcinogen (10–30 mg) to test animals. Minimal tissue localization of BaP and/or its metabolites occurs in the spleen, kidney, lung, and stomach while maximum amounts are recovered from the bile and feces. The level of carcinogen detected in the tissues is also directly related to the dose administered and is dependent on the use of a lipid vehicle. However, owing to the rapid metabolism of PAHs, no significant accumulation takes place in body fat. The amount of PAH excreted in urine is ~20–25% of the amount excreted through feces.

PAHs are metabolized by microsomal mixed-function oxidase systems in conjunction with  $P_{450}$ -type cytochrome enzymes. Various forms of cytochrome  $P_{450}$ -type enzymes are preferentially induced by different PAHs. Isoenzymes belonging to  $P_{450}$  1A,  $P_{450}$  2A,  $P_{450}$  3A, and  $P_{450}$  2B superfamilies are thought to be involved in the metabolism and activation of PAHs. Individually these PAHs also induce enzymatic oxidation to varying degrees and specific inducers stimulate the metabolism of these PAHs differently. This means that some weak carcinogenic or noncarcinogenic PAHs may be potent tumor inducers and may lead to increased amounts of activated metabolites of carcinogenic PAHs. The highest metabolizing

capacity of PAHs is present in the liver followed by the lungs, intestinal mucosa, skin, and kidney.  $P_{450}$  1A related activities in human lung are  $\sim 1\text{--}4\%$  of those in the liver.

One of the activation pathways converting BaP to a highly reactive and probably ultimate carcinogenic metabolite, (+)-7,8-dihydroxy-9,10-oxy-7,8,9,10-tetrahydrobenzo(a)pyrene ((+)-BPDE(1)), is shown in Figure 2. The vicinal (bay-region) diol epoxides are the ultimate mutagenic and carcinogenic species of alternate PAHs. (–)-BPDE(2) and (+)-BPDE(1) bind only weakly with DNA while (+)-BPDE(1), the major product of the enzymatic oxidation reaction, readily decomposes to form a carbonium ion. This



**Figure 2** Enzymatic activation of benzo[a]pyrene.



reacts rapidly with nucleophilic sites in DNA and RNA. In the cellular environment, possible detoxification of (+)BPDE(1) is in competition with its covalent binding to DNA. Both (–)BPDE– and (+)BPDE–deoxyguanosine adducts have been identified in mice and rabbits that have been treated with BaP. The DNA adducts were found to persist in the lung and brain, which could result in their significant accumulation from long-term exposure to low levels of PAHs. Reactions with cellular DNA cause mutation and an eventual loss of the ability to undergo controlled replication.

Nonalternant PAHs, e.g., benzo[fluoranthene], exert their genotoxic effects through their reactive metabolites. Peroxidizing enzyme systems in various body tissues also can form free radicals. These free radicals produced bind to C8 and N7 in guanine and N7 of adenine in DNA. The significance of this pathway for PAHs carcinogenicity has not been evaluated.

Of particular importance and relevance is the presence of PAHs in the atmosphere. At room temperature (20°C), PAHs with more than three aromatic rings have low vapor pressure and therefore will be adsorbed onto particulates. The nature of the particulates either enhances or reduces the cellular uptake of PAH. The following processes are essential to the induction of respiratory tract and lung cancer by airborne PAHs:

1. Retention of the inhaled carcinogenic PAH by the respiratory tract. The PAHs are adsorbed on particulate matter of appropriate size that is inhaled, deposited, and efficiently retained to increase the total exposure level.
2. The deposited PAH is eluted from the particulates by body fluids and taken up in the cell tissues. This process can be facilitated by particles that readily release adsorbed PAHs relative to the rate they are removed from the lungs, or by the presence of those particles that are able to penetrate cell membranes. Particles with aerodynamic diameters ( $<3.5\ \mu\text{m}$ ) may readily reach the pulmonary region of the lungs while larger particles ( $3.5\text{--}10\ \mu\text{m}$ ) are deposited in the nose and tracheobronchial region. Exposure is a function not only of the PAH concentration but also of its distribution between gaseous and particulate phases and among the different sizes of particles with which it is associated.

In the environment, PAHs undergo reactions such as photooxidation, sulfonation, and nitration. These reactions are important not only because they represent a major pathway for the removal of PAHs but also because the reaction products are health hazards themselves and more toxic than the PAHs. Two

examples are the conversions of perylene and pyrene (nonmutagenic) to 3-nitroperylene (directly active mutagen) and 1-nitropyrene (highly mutagenic), respectively. These reactions are influenced by humidity, ultraviolet (UV) radiation, temperature, and the nature of particulates onto which the PAHs are adsorbed. Mutagens may occur in ambient particulates because of the reactions of PAHs with ozone, nitrous oxides, peroxyacetyl nitrate, and free radicals all present to varying degrees in polluted atmospheres.

## Sources

Conditions for the formation of PAHs are most favorable during the pyrolysis of organic matter in air-deficient environments at temperatures in the range 650–900°C. The aromatic compounds formed are more stable than their precursors that include acetylene, butadiene, and butyl benzene.

However, a high temperature is not necessarily required for the aromatization of organic matter. During the formation of crude oil, for example, which takes millions of years, aromatization proceeds in sediments at lower temperatures ( $<150^\circ\text{C}$ ). Also, certain rare organic minerals that fluoresce, e.g., idrialite, pendletonite, and curtisite contain PACs. Other natural sources of PAHs include grassland and forest fires together with volcanic eruptions. The PAH fraction found in young marine sediments is thought to originate largely from forest and grassland fires from which they are dispersed by prevailing winds. Thus, PAHs are found even in remote parts of the world.

PAH emissions from anthropogenic sources account for the largest percentage in terms of annual global input. These sources, which all involve the combustion of organic materials, can be classified as either stationary or mobile (transportation related) processes as illustrated in **Table 2**. The combustion of fossil fuels in the twentieth century is the single most important source of PAHs.

Though tobacco smoking is a minor source of PACs in terms of their total global input, it is of considerable importance with respect to human health. Recent estimates show that ~20% (250 million people) of those living in developed countries will eventually be killed by tobacco-related cancers, e.g., lung and mouth cancers.

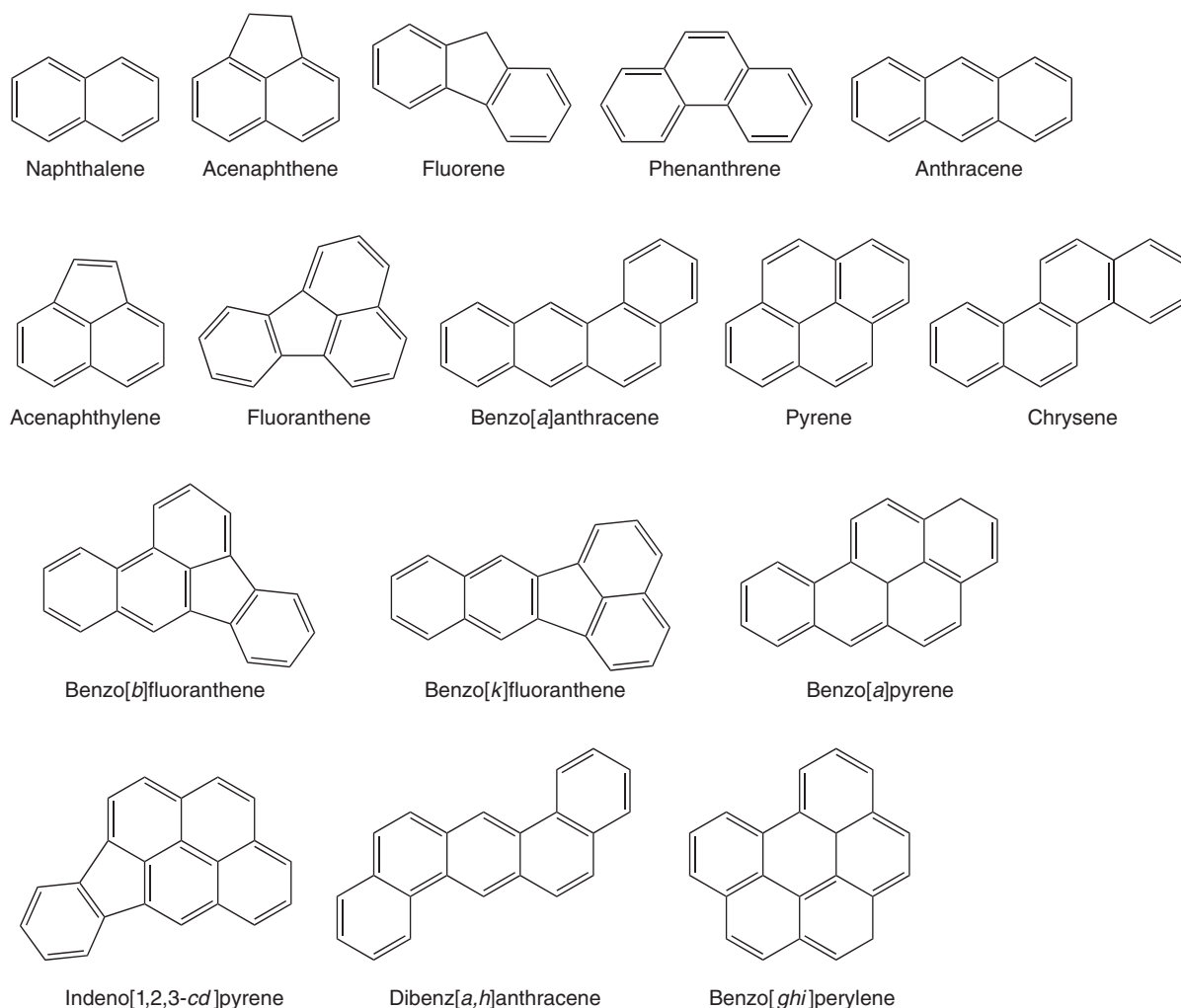
The historical diversity and lack of standardization of analytical methods used prior to the 1980s means that the PAH data obtained from diverse environments by different workers prior to the 1980s are difficult to compare. In the 1980s, the availability of

**Table 2** Anthropogenic sources of PAHs

<i>Stationary</i>	<i>Mobile</i>
Industrial processes, e.g., coke production Carbon black production Aluminum plants Asphalt production Petroleum cracking Industrial waste discharge	Gasoline and diesel engines  Land, air, and sea transport Oil spills Tyre wear on roads
Power plants and industrial boilers	
Municipal waste burning and incineration	
Open burning, e.g., coal refuse burning Agricultural waste burning Wild oil well fires Proscribed forest fires	
Residential heating and cooking using fossil and biomass fuels	
Tobacco smoking	

PAH reference materials, as discussed below, has led to standard and widely used methods for PAH analysis. The term 'total PAHs' as used by different researchers has been misleading since there was no universally accepted PAH list. In the 1990s, 16 PAHs listed in the priority pollutants of the US Environment Protection Agency (USEPA) list have been adopted for monitoring in many countries around the world. **Figure 3** shows the structures of the 16 PAHs listed in the priority pollutants of USEPA.

In many cases, reported emission factors have been obtained from a relatively small number of measurements, sometimes only one, and are thus not representative. Emission factors from a particular source will depend on the state of maintenance (e.g., for incinerators, how often the chimneys are scrubbed) and any special measures that have been taken to limit the extent of PAH emission. More PAH data than is presently available is therefore needed for > 90% of all anthropogenic PAH sources in order to

**Figure 3** Molecular structures of PAHs listed in the priority pollutants of USEPA.

accurately estimate the annual PAH global input and the level of human exposure to PAHs.

It is important to pinpoint the PAH source even in an environment where more than one pollutant source is evident. The following source-specific compounds have been used: cyclopenta(*cd*)pyrene for automobile exhaust; benzo(*b*)naphtho(2,1-*d*)thiophene for hard coal combustion; picene derivatives for brown coal combustion, and 1-methyl-7-isopropylphenanthrene (retene) for wood combustion.

## Analytical Methods

Because of the differences in biological activity of PAC isomers, the analytical challenge is to identify and quantify all of the PACs in a particular environment. The task of isomer differentiation is particularly difficult because of the large number of possible isomeric PACs, e.g., the number of possible isomers of unalkylated condensed four-ring PAHs is three, benzo(*a*)anthracene, chrysene, and triphenylene, whereas of unalkylated mononitrobenzo(*a*)anthracene is 12.

The analytical procedure for PAH analysis usually involves sampling, sample preparation techniques, and finally identification and quantification. In environmental samples, PAHs are found in mixtures with wide variation in the relative concentrations of individual components that are at low levels (approximately nanogram per liter). The PAHs are therefore usually concentrated during sampling.

Samples containing PAHs are obtained from air, soil (and sediment), or water so collection procedures to ensure the validity of the sample obtained are very important. For the sample analyzed to represent the 'true' composition of the environmental matrix, sampling must be quantitative, no artifacts should be formed during sample collection, and all sample treatment and storage before analysis must be free from interferences and losses.

During atmospheric PAH sampling, suspended particulate matter is collected on filters while vapor-phase PAHs are collected on solid adsorbents, impregnated filters, polyurethane foam plugs, or liquid impingers. Cryogenic traps and impingers are less popular since they are not sufficiently rugged for field applications.

Water samples are collected in clean glass bottles. If possible, the water should be sampled directly into the extraction vessel and PAHs extracted as soon as possible. If the sample will be stored, then subzero temperatures are used to minimize reactions that lead to loss of PAHs. In the aquatic environment, PAHs are found both dissolved in water and adsorbed on particulates. PAHs are sparingly soluble in water, their solubility decreasing with increasing relative molecular mass. Surface concentration is highest.

Sediment samples (10–100 g) are collected using mechanical grabs, coring devices, or vacuum suction. At low tide, samples are obtained by simply scooping the sediment while water divers use hand coring. Special traps are used for collecting newly deposited sedimentary material.

Birds and eggs are hand collected. Fish and sedentary organisms (e.g., shellfish, oysters, and plankton) are collected at low tide using commercial dredging equipment or nets.

## Sample Extraction and Cleanup

Before proceeding with the chemical analysis, PAHs in the collected samples must be extracted into solution. PAHs are extracted from adsorbents, filters, and solids (including soils and sediments) by Soxhlet extraction, ultrasonic vibration, microwave-assisted solvent extraction, and, more recently, supercritical fluid extraction using a variety of solvents. A disadvantage of Soxhlet extraction is the time taken (6–24 h) while ultrasonic or microwave methods take shorter times, use less solvent volumes, and do not require special glassware. Also pressurized hot solvent extraction in the static, dynamic, and static–dynamic modes give good extraction efficiencies for PAHs. The powerful solvent power and gas-like mass transfer properties of supercritical fluids provide the potential for a rapid and efficient extraction process that results in a concentrated extract (*c.* <1 ml). This may be compared with liquid solvent extractions that take many hours and give a dilute extract (usually more than 50 ml).

Liquid–liquid partition is employed to exhaustively extract PAHs from water using a series of organic solvents. Solid-phase extraction of PAHs is also gaining in usage. In this technique, the aqueous solution is passed through octadecyl- or octyl-bonded silica cartridges to retain organic compounds after which they are selectively eluted with an organic solvent. Fish and biota samples are freeze dried, then mechanically homogenized with an organic solvent such as methanol, acetonitrile, or hexane. Protein-rich samples are saponified with sodium or potassium hydroxide before extraction with an organic solvent.

The liquid extracts of environmental samples contain other organic compounds in addition to PAHs so cleanup procedures must be applied to yield a PAH fraction sufficiently free from extraneous compounds for instrumental analysis. The degree of sample handling and work-up is a function of the sample matrix, extraction method, and solvent, and the instrumental method to be used for the PAH analysis. Aliphatic, polar, and nonpolar fractions of the sample extract are separated by solvent partitioning.

Liquid–solid chromatography at ambient pressure on silica, alumina, or Florisil columns is also used. In this case the polarity of the eluting solvents is gradually changed so that the compounds in the sample are separated on the basis of their polarity. This method has the disadvantages of being slow, having poor reproducibility, requiring manual control, and a new column for every sample and necessitating the use of large quantities of solvents. It is also subject to irreversible adsorption that results in the loss of a percentage of all compounds or, in extreme cases, complete loss of some compounds. Reproducible elution characteristics have been obtained using lipophilic gels, e.g., Sephadex LH-20 and Bio-beads SX-12, in open column liquid chromatographic sample cleanup procedures. Samples have also been separated satisfactorily on thin layer chromatographic plates, the zone containing the PAH fraction being scraped from the plate with subsequent extraction into an organic solvent.

Reversed- and normal-phase preparative liquid chromatography (LC) have both been employed to give reproducible sample cleanup procedures. C<sub>18</sub> prepacked cartridges have gained in popularity because of their speed and reproducible elution properties. Normal-phase preparative LC on polar bonded silica is more popular than reversed-phase preparative LC because the PAH containing aliquots can easily be concentrated by removal of nonpolar solvents at low temperatures. In reversed-phase preparative LC, liquid–liquid extraction of the PAH has to be effected before concentration of the PAH aliquot.

The complexity of the sample dictates the number of isolation steps necessary before the final analysis. For instance, ambient air particulate extracts usually contain predominantly neutral organic compounds that greatly simplify the sample cleanup. In more complex samples, e.g., crude oil, the procedure involves removal of waxes, separation on a silica column, normal-phase preparative LC on an aminosilica column, and elution through a Sephadex LH-20 column.

## Analysis

Even after comprehensive separation and isolation of a PAH fraction from an environmental sample, although it consists mainly of parent PAH (and isomers) and alkyl substituents, it is still likely to be complex and may contain hundreds of compounds spread over wide volatility and concentration ranges. Gas and liquid chromatographic techniques are complementary in the characterization of PAHs in environmental samples. For high molecular mass and nonvolatile PAHs, LC is superior to gas chromatography (GC). High resolution capability of

capillary GC and easy compatibility with mass spectrometers and element specific detectors are strong points in favor of choosing capillary GC over LC for those PACs where both techniques are suitable. However, LC–mass spectrometry (MS) instrumentation has now become more readily available.

## Gas Chromatographic Analysis

Conventional gum phases, methylpolysiloxanes (SE-30, OV-1), 5% phenylmethylpolysiloxane (SE-52), 5% phenyl–95% dimethylpolysiloxane, and 5% phenyl–1% vinylmethylpolysiloxane (SE-54) are popular stationary phases in capillary column GC, by which separation of PAHs up to coronene (relative molecular mass = 300) may be achieved. The stationary phase giving the best separation for a maximum number of PAHs, however, is SE-54 bonded to fused silica capillary columns. It is a slightly more selective stationary phase than SE-30, its selectivity arising from the polarizability of its phenyl groups.

Gas chromatographic separation on gum phases results in the elution of solutes according to their boiling points, which in turn means that isomers having nearly the same boiling points are not separated. Good selectivity and separation of PAH isomers has been achieved using liquid crystal and mesogenic phases, in which the geometry of the molecule is also important. The elution order in an isomeric unsubstituted PAH series follows, in general, the solute length-to-breadth (*l/b*) ratio. The isomer with the larger (*l/b*) ratio is retained longer. Retention of solutes also depends on their strength of interaction with the stationary phase, i.e., the effective surface area of interaction. A planar molecule that can penetrate between the liquid crystals will have a larger interacting surface area than a similar nonplanar PAH that does not penetrate as well.

The flame ionization detector (FID) is the most frequently used detector for PAH analysis in GC. It has a general response character independent of the structure of the hydrocarbon, is reliable, easy to maintain, and has a wide linear range. The photoionization detector (PID) has been used to selectively detect PAHs in the presence of alkanes. The PID has a large linear dynamic range, high sensitivity (100 times that of FID for PAH), and a concentration-dependent response. The electron-capture detector (ECD) is rarely used for PAH detection, mainly because it is necessary to determine response factors for each PAH, it has a low linear range, and shows baseline irregularity when the oven temperature is programmed. The ratio of ECD to FID and FID to PID responses has been used to identify and/or distinguish between different groups of PACs.

The coupling of a mass spectrometer to capillary GC combines an efficient chromatographic separation with high-performance identification. The mass spectrometer can be used as a low-picogram range detector in the total-ion current mode, as a selective detector in the single monitoring mode, or for positive identification to support retention data in the scanning mode. GC-MS is mainly used for identification and confirmation purposes. However, using electron impact ionization, PAH isomers are not easily distinguishable. Positive chemical ionization (CI) with methane, a mixture of methane/argon, or dimethyl ether as reagent can be used to differentiate between PAH isomers. Ion generation using a variable energy ionizing laser in the GC-MS determination of PAHs has shown high sensitivity and selectivity for isomers. A quadrupole mass selective detector employing chemical ionization, electron impact, or fast atom bombardment techniques to give ionization is most commonly used. The amount of data provided by GC-MS during each run (approximately one spectrum every second) makes computerized data acquisition, storage, and processing essential. Computerized data reduction allows subtraction of any spectral background and direct comparison with stored library spectra for identification purposes.

Fourier-transform infrared spectroscopy (FTIR) can be carried out directly by passing the GC eluent through a gold-coated light pipe with IR-transparent windows, or the GC eluent can be frozen on a gold disc and FTIR performed later (matrix-isolation IR). The chromatogram is reconstructed and a plot of chemical-function groups against time produced. Though IR has low sensitivity and resolution, it has the advantage over GC-MS in being able to distinguish between isomers when standard compounds are unavailable. Because gas-phase fluorescence spectra of PAHs are broad and featureless, this method of detection is not very applicable in their analysis by GC.

### Liquid Chromatographic Analysis

Though LC does not approach the separation efficiency of capillary GC, it offers the following advantages:

1. Selectivity can be achieved by changing both the stationary and mobile phases.
2. It can be used for the analysis of compounds that are thermally unstable compounds, e.g., partially condensed molecular structures such as binaphthyl isomers that rearrange to form condensed molecular structures at elevated temperatures.
3. Ultraviolet and luminescence (fluorescence and chemiluminescence) spectrometry provides sensitive and selective detection of PAHs.

A study of a group of alkyl chemically bonded reversed-phase stationary liquids ( $C_6$ - $C_{22}$ ) concluded that  $C_{18}$  provided better PAH separations. With monomeric stationary phases, PAHs are usually separated according to the number of aromatic rings in their chemical structure while with polymeric  $C_{18}$  phases, selective retention of PAH isomers based on the molecular shapes ( $I/b$  ratio) is observed. Phase selectivity of PAHs on polymeric  $C_{18}$  in LC and on smectic liquid crystal phases in GC show close correspondence of retention behavior. PAH specialty columns for separations in LC are now readily available in the market.

Separations of PAH isomers by shape-selective phases are useful if the sample contains compounds within a limited number of aromatic rings, otherwise the sample becomes too complex to analyze because molecules with different ring sizes coelute. This makes identification of the components difficult even when selective fluorescence detectors are used. Multi-dimensional LC should be applied in cases where it is necessary to characterize completely all the components of a complex polyaromatic compound mixture. Normal-phase LC on polar bonded silica gives eluent fractions consisting of molecules with similar numbers of pi-electrons. These can then be further separated on a monomeric-type reversed-phase column into different isomer types. Finally, a shape-selective polymeric  $C_{18}$  reversed-phase column (or liquid crystal column in capillary GC) may be used to separate the specific PAH isomers that differ in planarity and ( $I/b$ ) ratios. Two-dimensional LC is a powerful tool for the separation of PAHs in complex environmental samples because of the higher peak capacity than single column LC.

A major advantage of LC in PAH analysis is the availability of detectors that are highly sensitive and selective to the polyaromatic molecular structure. PAH molecules display strong and characteristic UV absorption, are naturally and strongly fluorescent, and also give sensitized luminescence under peroxyoxalate chemiluminescence (PO-CL) conditions. The high selectivity of fluorescence detection permits the analysis of PAHs in the presence of nonfluorescing compounds. Maximum selectivity and sensitivity can be selected for each compound as it elutes by changing the excitation and emission wavelengths with time (wavelength programming) during the chromatographic run. Therefore, individual PAHs in a mixture can be determined even when complete LC resolution of components has not been achieved. Measurement of fluorescence lifetimes provides the additional information needed to confirm peak assignments based on retention times.

In CL detection, the absence of a light source results in a reduction of the background noise and therefore detection limits are governed by the chemical noise of the system. Usually, a compromise has to be made between the optimum mobile phase for best LC separation and optimum reaction conditions for maximum CL emission. Water, an important solvent in reversed-phase LC, strongly affects the solubility of peroxyoxalate reagents and the kinetics of the PO-CL reaction. However, PO-CL has the potential of being highly selective and sensitive in determining individual PACs as they are eluted in LC, since the CL reactions have high CL quantum yields (>20%) and are highly selective.

MS has been used with LC (moving belt interface), though not as extensively as with GC. Atmospheric pressure chemical ionization and pneumatically assisted electrospray mass spectrometry coupled to LC have been used for PAC analysis. Since it is easy to collect fractions continuously from LC, bioassays of the collected fractions can be performed and a mutachromatogram (a plot of mutagenic potency versus fraction number) obtained.

### Other Analytical Techniques

The potential of supercritical fluid chromatography (SFC) for the analysis of larger relative molecular mass PACs (more than six aromatic rings) using both packed and capillary columns has been realized. SFC provides a higher separation efficiency than LC and good solvation of the larger nonvolatile molecules. However, it has a lower efficiency than capillary GC. FID, UV, fluorescence, and MS have all found application in the detection of PAHs after supercritical fluid chromatographic separation.

Shpol'skii spectrofluorimetry has been used for the determination of PAHs in crude environmental sample extracts with minimum sample cleanup. This technique gives a vibrationally resolved fluorescence spectrum of samples dissolved in a suitable solvent (usually an *n*-alkane) at cryogenic temperatures, e.g., 26 K. It combines the selectivity of an infrared spectrum with the sensitivity of fluorimetry, though the sensitivity suffers considerably from the presence of large amounts of interfering substances such as fatty components in crude extracts since these give a poor-quality matrix with a high sample absorbance.

Recent research has intensified the search for portable instruments for real-time monitoring of airborne PAHs, particularly in the industrial environment. A derivative UV absorption spectrophotometer has been used for real-time monitoring of a small number of aromatic vapors. When synchronous fluorescence and room-temperature phosphorescence (using a

filter paper impregnated with heavy atom salts) are incorporated into automated, continuous air sampling devices, near real-time monitoring of a selected number of PAHs is achieved. Personnel exposure meters to be worn by industrial workers are available. These comprise a filter paper, pretreated with heavy atom salts that enhance the phosphorescence of a selected number of PAHs. The filter holder fits into a phosphorescence spectrometer, allowing fast and direct analysis. A photoelectric aerosol sensor, working on the basis of photoelectric ionization of less than 1  $\mu\text{m}$  induced by UV radiation of a KrBr excimer lamp is used to give the concentration of total PAHs.

### Validation

In GC, correspondence between the retention data of an unknown and a reference compound on two columns of different polarity is considered sufficient for positive identification. Retention indices of PAHs based on naphthalene, phenanthrene, chrysene, and picene are also reliable for identifications. Mass spectral data, though identical for isomers, are invaluable for identification purposes.

Internal standards (IS) are popularly used for the quantitative analysis of PAHs, particularly in methods using capillary GC with FID. This has the advantages of direct measurement of peak area (or height) of solute to peak area of IS, ease of sample cleanup (since what is lost in the sample component is also lost in the IS), and of no necessity to know the exact sample volumes injected. Since the boiling points of PAHs cover a wide range, more than one internal standard for GC analysis is required. Compounds that have been used include 3,6-dimethylphenanthrene for the low relative molecular mass PAHs and benzo(*b*)chrysene or 2,2'-binaphthyl for the higher relative molecular mass PAHs that have longer retention times. Perdeuterated perylene is also used as an internal standard for PAHs during analysis by GC.

Diode array detection of PAHs after LC allows acquisition and storage of each UV spectrum of the eluting peaks. The chromatogram can then be reconstructed at any defined wavelength after the analysis. The spectral output of a diode array is used not only for qualitative and quantitative analysis but also to augment the resolution of chromatographic separation, since poorly resolved components can be identified by deconvolution of the eluent spectra.

Reference standards are essential in the implementation of regulations where limits have been set for PAH concentration levels in various matrices. Many PAHs have been synthesized and highly purified



**Table 3** List of available certified reference materials for PAH analysis

Order no.	Material	Supplier
SRM 1469 (a)	Urban dust	NIST
SRM 1648	Urban particulate matter	NIST
N16-49	Urban dust (organics)	NIST
N16-50	Diesel particulate matter	NIST
SRM 1941(a)	Sediment	NIST
HS-3 to HS-6	Harbor marine sediments	NRC
SES-1	Estuarine sediment	NRC
BCR 088	Sewage sludge	BCR
SRM 1974	Mussel tissue	NIST
BCR 524	Industrial waste soil	BCR
BCR 535	Fresh water harbor sediment	BCR

(>99.9%) and are commercially available. The availability of reference standards helps greatly in the unambiguous identification and reproducible quantification of PAHs in environmental samples.

Certified reference materials are intended primarily for calibration and in quality control of analytical techniques. A certified reference material is a reference material, accompanied by a certificate, one or more of whose property values are certified by a procedure that establishes its traceability to an accurate realization of the unit in which the property values are expressed and for which each certified value is accompanied by an uncertainty at a stated level of confidence. They are used to test, validate, and optimize new analytical techniques as well as in quality control of routine laboratory work. **Table 3** gives a list of environmental samples that provide certified values of PAH content. These are available from various bodies as certified reference materials. In measuring the concentration of a substance for certification purposes, more than two independent and reliable analytical methods are used. Certified reference materials when used for standardization of analytical methods will make comparisons between PAH data obtained by a variety of workers using

different analytical methods possible. Thus, it will be possible to obtain a clear picture of the annual global PAH input and the level of human exposure to these compounds.

*See also:* **Air Analysis:** Sampling. **Environmental Analysis. Gas Chromatography:** Pyrolysis. **Liquid Chromatography:** Overview. **Mass Spectrometry:** Environmental Applications. **Polycyclic Aromatic Hydrocarbons:** Environmental Applications. **Quality Assurance:** Primary Standards; Reference Materials; Production of Reference Materials.

## Further Reading

- Bjorseth A (1983) *Handbook of Polycyclic Aromatic Hydrocarbons*. New York: Dekker.
- Bjorseth A and Ramdahl T (1985) *Handbook of Polycyclic Aromatic Hydrocarbons*, vol. 2. New York: Dekker.
- Cavelieri E and Rogan E (1985) In: Woo YT, Lai DY, Arcos JC, and Argus MF (eds.) *Chemical Induction of Cancer, Structural Bases and Biological Mechanisms*, vol. IIIB, pp. 533–569. London: Academic Press.
- Gammage RB, Hawthorne AR, Vo-Dinh T, and Schuresko DD (1983) In: Veriroglu TN (ed.) *Alternative Energy Resources*, vol. 3, pp. 209–228. Berlin: Springer-Verlag.
- Glatt H and Oesch F (1986) In: de Serres FJ (ed.) *Chemical Mutagens, Principles and Methods for their Detection*, vol. 10, pp. 73–127. London: Plenum.
- Hester RE and Harrison RM (1998) Air pollution and health. *Issues in Environmental Science and Technology*, RSC.
- Lee ML, Novotny MV, and Bartle KD (1981) *Analytical Chemistry of Polycyclic Aromatic Compounds*. New York: Academic Press.
- Nikolaou K, Masclet P, and Mouvier G (1984) Sources and chemical reactivity of polynuclear aromatic hydrocarbons in the atmosphere. *The Science of the Total Environment* 32: 103–132.
- Vo-Dinh T (1989) *Chemical Analysis of Polycyclic Aromatic Compounds*. New York: Wiley.

## Environmental Applications

**A N Gachanja**, African Institute for Capacity Development, Nairobi, Kenya

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Polycyclic aromatic hydrocarbons (PAHs) from anthropogenic sources are discharged into the atmosphere

either in the gaseous state or adsorbed onto particulates. These PAHs are subject to removal mechanisms such as oxidative and photolytic reactions and wet and dry deposition. Air currents and prevailing winds disperse and transport the PAHs over long distances and deposit them on soil and in water. PAHs on soil are carried during storms by water runoff to rivers and seas. In the aquatic environment, the PAHs enter marine plants, fish, and sedentary organisms. PAHs are

**Table 3** List of available certified reference materials for PAH analysis

Order no.	Material	Supplier
SRM 1469 (a)	Urban dust	NIST
SRM 1648	Urban particulate matter	NIST
N16-49	Urban dust (organics)	NIST
N16-50	Diesel particulate matter	NIST
SRM 1941(a)	Sediment	NIST
HS-3 to HS-6	Harbor marine sediments	NRC
SES-1	Estuarine sediment	NRC
BCR 088	Sewage sludge	BCR
SRM 1974	Mussel tissue	NIST
BCR 524	Industrial waste soil	BCR
BCR 535	Fresh water harbor sediment	BCR

(>99.9%) and are commercially available. The availability of reference standards helps greatly in the unambiguous identification and reproducible quantification of PAHs in environmental samples.

Certified reference materials are intended primarily for calibration and in quality control of analytical techniques. A certified reference material is a reference material, accompanied by a certificate, one or more of whose property values are certified by a procedure that establishes its traceability to an accurate realization of the unit in which the property values are expressed and for which each certified value is accompanied by an uncertainty at a stated level of confidence. They are used to test, validate, and optimize new analytical techniques as well as in quality control of routine laboratory work. **Table 3** gives a list of environmental samples that provide certified values of PAH content. These are available from various bodies as certified reference materials. In measuring the concentration of a substance for certification purposes, more than two independent and reliable analytical methods are used. Certified reference materials when used for standardization of analytical methods will make comparisons between PAH data obtained by a variety of workers using

different analytical methods possible. Thus, it will be possible to obtain a clear picture of the annual global PAH input and the level of human exposure to these compounds.

*See also:* **Air Analysis:** Sampling. **Environmental Analysis. Gas Chromatography:** Pyrolysis. **Liquid Chromatography:** Overview. **Mass Spectrometry:** Environmental Applications. **Polycyclic Aromatic Hydrocarbons:** Environmental Applications. **Quality Assurance:** Primary Standards; Reference Materials; Production of Reference Materials.

## Further Reading

- Bjorseth A (1983) *Handbook of Polycyclic Aromatic Hydrocarbons*. New York: Dekker.
- Bjorseth A and Ramdahl T (1985) *Handbook of Polycyclic Aromatic Hydrocarbons*, vol. 2. New York: Dekker.
- Cavelieri E and Rogan E (1985) In: Woo YT, Lai DY, Arcos JC, and Argus MF (eds.) *Chemical Induction of Cancer, Structural Bases and Biological Mechanisms*, vol. IIIB, pp. 533–569. London: Academic Press.
- Gammage RB, Hawthorne AR, Vo-Dinh T, and Schuresko DD (1983) In: Veriroglu TN (ed.) *Alternative Energy Resources*, vol. 3, pp. 209–228. Berlin: Springer-Verlag.
- Glatt H and Oesch F (1986) In: de Serres FJ (ed.) *Chemical Mutagens, Principles and Methods for their Detection*, vol. 10, pp. 73–127. London: Plenum.
- Hester RE and Harrison RM (1998) Air pollution and health. *Issues in Environmental Science and Technology*, RSC.
- Lee ML, Novotny MV, and Bartle KD (1981) *Analytical Chemistry of Polycyclic Aromatic Compounds*. New York: Academic Press.
- Nikolaou K, Masclet P, and Mouvier G (1984) Sources and chemical reactivity of polynuclear aromatic hydrocarbons in the atmosphere. *The Science of the Total Environment* 32: 103–132.
- Vo-Dinh T (1989) *Chemical Analysis of Polycyclic Aromatic Compounds*. New York: Wiley.

## Environmental Applications

**A N Gachanja**, African Institute for Capacity Development, Nairobi, Kenya

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Polycyclic aromatic hydrocarbons (PAHs) from anthropogenic sources are discharged into the atmosphere

either in the gaseous state or adsorbed onto particulates. These PAHs are subject to removal mechanisms such as oxidative and photolytic reactions and wet and dry deposition. Air currents and prevailing winds disperse and transport the PAHs over long distances and deposit them on soil and in water. PAHs on soil are carried during storms by water runoff to rivers and seas. In the aquatic environment, the PAHs enter marine plants, fish, and sedentary organisms. PAHs are

thus ubiquitous pollutants that have been detected even in remote parts of the earth, e.g., the average atmospheric concentration of the sum of 11 PAHs at Barrow, Alaska, was reported to be 1.2 and 0.16 ng m<sup>-3</sup> for Mar. and Aug. 1979, respectively, and PAH concentrations in sediments have ranged from undetectable in the Amazon river to greater than 10 000 µg l<sup>-1</sup> in the Charles river, Boston, USA. This article briefly reviews the analysis of PAHs in the atmosphere, water, food, and biological samples.

## Atmosphere

PAHs from all sources where organic matter is burnt (e.g., gasoline and diesel engines, industrial boilers, open burning, etc.) are discharged into the air. At ambient temperatures (20°C), PAHs with up to three rings exist in the gaseous state while those with six or more rings are found exclusively bound to particulates. Four- and five-ring PAHs are found in both gaseous and particulate phases.

The identification and quantitation of airborne PAHs in the ambient air represent a critical aspect in human health assessment. It is therefore necessary to collect both gaseous and particulate bound PAHs. Two sampling trains recommended for general purpose stack sampling of organic compounds with boiling points above 100°C are the modified method 5 and source assessment sampling system with sampling rates of 14–28 and 110–140 l min<sup>-1</sup>, respectively. Both trains have a filter to remove particulates, a cooled sorbent module, and a condensate collector. Samplers operating at flow rates greater than 1.0 m<sup>3</sup> min<sup>-1</sup> (>1000 l min<sup>-1</sup>) have also been used. Isokinetic sampling pumps should be used during both low and high volume samplings.

Glass fiber, quartz, polytetrafluoroethene (PTFE), cellulose acetate membrane, and silver membrane filters have been used for particulate sampling. PAH blanks and losses during sampling, and resistance to airflow should be considered when choosing the filter material. Solid adsorbents that have been investigated for sampling volatile PAHs include charcoal, silica gel, Florisil, molecular sieves, XAD resins (polystyrene-divinylbenzene polymers), Chromosorb 101 and 102, and Tenax GC. Charcoal exhibits poor recovery properties for PAHs while silica gel, Florisil, and molecular sieves show poor moisture tolerance. Chromosorb 101 has a low surface area while Chromosorb 102 has similar properties to XAD-2 but is only available in fine mesh sizes that cause high pressure drops during sampling. When thermal desorption is the chosen extraction method, Tenax GC is preferred because of its high thermal resistance (up

to 375°C). Polyurethane foam is used to trap medium volatile PAHs while XAD-2 resin (used when solvent extraction of PAHs is undertaken) traps the volatile fraction. The Environmental Protection Agency, USA (USEPA) method TO-13 recommends sampling for PAHs in the atmosphere at 230 l min<sup>-1</sup> using QMA quartz filter and XAD-2 resin.

Chemical reactions of organic material collected on filters with other pollutants such as NO<sub>2</sub>, O<sub>3</sub>, HNO<sub>3</sub>, SO<sub>2</sub>, H<sub>2</sub>SO<sub>4</sub>, photooxidation, volatilization during sampling, and sample storage are all ways in which the sample integrity can be violated. Several factors, e.g., sampling face velocity, relative humidity, filter type, and the nature of particulates, may affect volatilization rates and reactions of the trapped PAHs. As sampling times are increased, not only does oxidative degradation occur to PAHs already accumulated on the collection medium, but also sublimation and/or reevaporation processes may take place, resulting in lower quantities of PAHs being trapped. The presence of ultraviolet radiation and oxidants (e.g., sulfur oxides, ozone, nitrous oxides) increases the decomposition rate of PAHs.

Efficient extraction of PAHs from the sample matrix is a crucial step in the preparation of a sample for PAH determination. Knowledge of extraction recoveries allows corrections to be made for PAH quantities not extracted and losses in analytical measurements. Solvent extraction methods, employing Soxhlet, ultrasonic, and supercritical fluid extraction (SFE) procedures are widely used.

PAHs can be Soxhlet extracted from air particulates using solvents such as benzene, toluene, carbon disulfide, dichloromethane, cyclohexane, and methanol for up to 24 h. Methanol is not only efficient in extracting the organic compounds but also extracts a larger amount of inorganic compounds than benzene, toluene, and cyclohexane. Multiple batchwise ultrasonic extractions for 10–60 min are as effective as Soxhlet extraction of PAHs from particulates. Ultrasonic extraction is faster and gives more reproducible extraction than Soxhlet extraction, though a high powered ultrasonicator is more expensive than a Soxhlet apparatus. SFE of PAHs from solid matrices gives comparable PAH extraction efficiencies to Soxhlet extraction but SFE has the advantage of speed (10–60 min) and ease of sample handling. Off- and online techniques may be used to collect the extracted PAHs after depressurization of the supercritical fluid. Though offline SFE is simpler and the extracted sample may be used for various analyses, sample losses may occur due to aerosol formation during depressurization. During online SFE, the extracted analytes are directly transferred to a chromatographic system for analysis.

Because air particulate extracts contain predominantly neutral organic compounds, the sample clean-up procedure is greatly simplified. Polar compounds are removed by partitioning in dimethylformamide/water or in dimethylsulfoxide. Chromatography on a Sephadex LH-20 column separates the aromatic and polar fractions. Normal-phase preparative high-performance liquid chromatography on an amino-bonded silica column with a hexane or heptane mobile phase provides separation of the PAHs based on the number of aromatic carbon atoms in the molecule; aliphatic compounds elute first, then the PAHs in order of increasing ring size, the largest ring compounds eluting last.

The contribution of any PAH source to the ambient air PAH load depends on a number of factors that include emission rate of the source, geographical location, and climatic conditions (e.g., wind strength and direction). Stationary sources account for ~90% of the annual total PAH emissions (estimated total PAH atmospheric emissions for USA, Sweden, and Norway in 1983 were 8600, 500, and 290 tons, respectively). Each year in the USA, ~1.9 million hectares of vegetation land are burnt due to wildfire ignitions in addition to 1.2–2.0 million hectares that are intentionally ignited to achieve land management objectives. Quantitative data are lacking for some mobile sources of PAHs, e.g., from aircraft engines, ships and boats, motorcycles, and lawn mowers. A jet engine has been reported to emit 2–4 mg benzo(a)pyrene (BaP) per minute when cruising and up to 40 000 mg min<sup>-1</sup> during take off. Mobile sources are likely to be the major PAH contributors in areas where major stationary sources are absent. Biomass and coal combustion make an appreciable contribution in areas where they are the dominant energy source, particularly in developing countries. The ratio of methylphenanthrene to phenanthrene is an indicator of the origin of PAHs since a ratio >1 indicates fuel oil contamination, e.g., vehicular exhausts.

Table 1 shows the concentrations of selected PAHs in the ambient air of some cities. PAH emissions exhibit seasonal fluctuations, higher values occurring during the winter due to increase in space heating emissions, lower photochemical activity, and meteorological conditions (e.g., lower temperatures) compared with summer and air mass origin. The profiles of PAHs also differ between and within cities. The levels of PAH in some cities, e.g., Brazzaville (Congo) and Calcutta (India) are due to gasoline and diesel engine exhausts with contributions from biomass combustion sources. The PAH levels in London (and many other European cities) has been decreasing, since peaking (BaP concentration ~100 ng m<sup>-3</sup>) in

**Table 1** Concentrations (ng m<sup>-3</sup>) of some representative PAHs in the urban air of some cities

Location and year	Hong Kong, Jan.–Dec. 2000	Fuji, Japan, 1999, S/W <sup>a</sup>	London, UK, 1996	Chicago, USA, 1995, ± SD	Brazzaville, Congo, Aug.–Sep. 1989	Athens, Jul. 2000	Santiago, Chile, 1993	Salvador, Brazil, 1991	Oviedo, Spain, 2000
Pyrene	3.4	21.00/12.72	2.5–5	24.6±14.7	5.7	2.1	1.08	1.1	2.2
Benzo[a]A	0.4	0.90/4.31	0.2–0.9	2.1±1.3	0.5	0.27	1.55	3.5	1.5
Chrysene	0.6	1.66/5.45	0.5–2	3.6±1.5	1.1	1.01	1.53	14.2	2.1
Benzo[e]P	0.4	2.87/4.50	ND	ND	0.4	0.36	ND	ND	ND
Benzo[a]P	0.2	1.50/3.38	0.05–0.6	1.6±1.0	0.2	0.17	2.51	3.06	1.4
Benzo[ghi]P	0.43	7.18/4.36	0.3–10	1.1±0.7	0.7	0.44	3.32	3.86	1.2
Indeno[123-cd]P	0.3	7.27/4.30	ND	1.2±0.7	ND	0.45	4.47	2.41	1.4

<sup>a</sup> Summer and winter maximum value. Prevedouros K, Brorstrom-Lunden E, Halsall CJ, Jones KC, Lee RGM, and Sweetman AJ (2004) Seasonal and long-term trends in atmospheric PAH concentrations: Evidence and implications. *Environmental Pollution* 128: 17–27. Ohura T, Amagai T, Fusaga M, and Matsushita H (2004) Spatial distributions and profiles of atmospheric polycyclic aromatic hydrocarbons in two industrial cities in Japan. *Environmental Science and Technology* 38: 49–55. Odabasi M, Vardor N, Sofuoglu A, Tasdemir Y, and Holsen TM (1999) Polycyclic aromatic hydrocarbons (PAHs) in Chicago air. *The Science of the Total Environment* 227: 57–67.

the 1960s, through decreasing usage of coal (in both domestic and industrial sectors by replacement with oil and natural gas) and increasing usage of catalytic converters in gasoline engines (starting in the 1980s). Starting from the 1990s, the level of BaP in rural sites in Europe is below  $1.0 \text{ ng m}^{-3}$  while in the urban areas with heavy traffic it is in the range  $1\text{--}10 \text{ ng m}^{-3}$ . In Copenhagen, a mean BaP concentration (Jan.–Mar. 1992) at a petrol station in a busy street was found to be  $4.4 \text{ ng m}^{-3}$ . The reported average total concentration of PAHs (total of 17 PAHs) for two different sites in New York city in 2000 were 3.2 and  $2.9 \text{ ng m}^{-3}$ .

Exhaust emissions from gasoline and diesel engines account for the majority of PAH emissions in urban areas. The emission of PAHs from gasoline engines is affected by factors such as air/fuel ratio, fuel aromaticity, driving mode, oil consumption, cold start, engine design, and the use of alternative fuels, e.g., ethanol and catalytic converters (reduction of BaP emission from  $4.6 \mu\text{g}$  per kg without catalyst to  $0.36 \mu\text{g}$  per kg with catalyst). Diesel engine exhausts may vary in BaP emissions by two orders of magnitude depending on engine design and can give emissions of the same order of magnitude as those from gasoline engines with catalytic converters. Although diesel engine exhausts contain less PAHs than gasoline engine exhausts, diesel emissions are more mutagenic and carcinogenic. This results from the formation of oxygenated and nitro-derivatives of PAHs that occur through reactions between nitrous oxide-rich exhaust and PAH. There has been a worldwide trend to reduce the amounts of lead alkyl antiknock additives in gasoline in order to reduce lead emissions in the atmosphere. In most countries, the octane quality of gasoline with reduced lead content is maintained by increasing the aromatic content, leading to increased PAH emissions.

Currently, there are no national or international standard limits for PAH levels in the environment. The scientific community is in general agreement that it is impossible to formulate a dose (or concentration) of a proven carcinogen or mutagen that is without a residual risk to health. Even the smallest doses of a carcinogen produce some genotoxic damage though no tumor formation is observed during finite experimental conditions. This implies that occupational exposure standards cannot be established. However, industrial hygienists argue that some limit of exposure to carcinogens should be introduced, one reason being that a certain limit is better than none at all. The introduction of occupational exposure limits for carcinogens would have the advantage that the prevailing concentrations would be monitored, with the hope of reducing

exposures to the lowest feasible levels (based on technical, analytical, and economic criteria). Extrapolation of results from laboratory animal inhalation experiments lead to the conclusion that a lifetime exposure to carcinogens in air will result in increased cases of cancer. Human exposure (from ambient air) for BaP is in the range  $<10$  to  $>100 \text{ ng day}^{-1}$ . Additional exposure will result from tobacco smoking and use of unvented heating sources. In the Netherlands, the interim goal is to reduce the annual average air concentration of BaP to  $5 \text{ ng m}^{-3}$  while the USEPA proposes a  $10 \text{ ng m}^{-3}$  guideline level. Poland has set a hygiene standard for BaP in breathing air of  $150 \text{ ng m}^{-3}$ . The United Kingdom has proposed an annual average standard for BaP of  $0.25 \text{ ng m}^{-3}$ . The recommendation by the expert panel on air quality standards (EPAQS) for PAHs is  $0.25 \text{ ng m}^{-3}$ . Reduction of BaP in air will in effect bring down the levels of other PAHs.

## Water

PAHs enter the aquatic environment through offshore drilling activities, oil spills, runoff water, and atmospheric fallout. The water solubility of PAHs decreases with increasing relative molecular mass and molecules with a linear arrangement of fused rings are usually less soluble than those having angular or pericondensed ring systems. Alkyl substituents on the aromatic ring of a given PAH decrease its water solubility. Because PAHs lack polar substituents, they are only sparingly soluble in water and are found mostly adsorbed on suspended particulates and sediment. An equilibrium exists between the PAH dissolved in the water and that adsorbed on particulates, which results in some percentage of even large PAH molecules getting into solution. The presence of organic pollutants, e.g., organic industrial effluents and detergents, influences the solubility of PAHs such that solubility values as high as  $7\text{--}12 \mu\text{g l}^{-1}$  have been reported.

The concentrations of individual PAH in water systems range from less than 1 ppt (pg per g) in pure ground water supplies to greater than 1 ppm ( $\mu\text{g per g}$ ) in heavily contaminated sewage. Therefore, some preconcentration and extraction techniques are required to raise the concentrations to levels at which identification and quantitative analysis are possible. Because PAHs may only represent as little as 0.01% of the total organic fraction present in the water sample, the analytical scheme must be devised so that the PAHs can be analyzed without the interference from the other pollutants. Since the concentrations are so low, serious errors may occur from losses or contamination during sampling or the analytical

process. Any surface with which the sample comes into contact can be a source of contamination or a sink for the trace pollutant. Whenever possible, the sampling should be done directly into the extraction vessel and extracted as soon as possible.

PAH concentrations are highest in the surface film of water. In waters that exhibit turbidity, e.g., rivers and lakes, the PAH will largely be adsorbed on the suspended particulate matter. Analysis of filtered water samples gives lower PAH amounts. Where particulate loading is low, PAHs may be extracted from the unfiltered sample but when particulate loading is high, separate extraction of PAHs from the filtered water and suspended particulate material is more desirable. However, there is the additional problem that PAH loss may occur through adsorption on to the apparatus during filtration.

Water samples should be collected in scrupulously cleaned glass bottles that should be completely filled at the sampling site and stoppers inserted to leave no air space. The screw caps or bottle tops should be PTFE lined or wrapped in aluminum foil and sources of contamination, e.g., dust, motor vehicle exhaust, cigarette smoke avoided. Water samplers have been designed in which a spring-loaded PTFE stopper may be removed and resealed at the sampling depth before recovery of the bottle is undertaken, usually with a rope. During the sampling of surface waters, the use of sampling devices through which the sample flows before entering the sample bottle should be avoided, as they may contaminate or adsorb PAHs. When sampling tap water, the tap should be allowed to flush for at least 2 min unless the first flush samples are specifically required. Residual chlorine must be removed from sampled potable water using sodium thiosulfate. Sampling of the surface film is more involved because of water inclusion and the analytical results are expressed in terms of surface area collected rather than volume collected. The samples are stored in the dark at  $\sim 4^{\circ}\text{C}$  and should be extracted as soon as possible after sampling. An organic solvent, e.g., benzene, is added if the samples are to be stored for several days before extraction. The incorporation of micelles, e.g., polyoxyethylene(*a*)-*p*-nonylphenol, in water samples before storage reduces the rate of loss of PAHs through adsorption onto the container surface. Additionally, the increased solubility of PAHs in the aqueous phase results in the extraction of PAHs from the suspended particulates into the water.

Hydrocarbons are exhaustively extracted using liquid-liquid partition with an organic solvent. Alternatively, the water is passed through a sorbent cartridge, e.g., XAD-2 resin, where the trace organics are retained and the water passes through. The organic compounds are later recovered by elution with a small

volume of organic solvent. This method, compared with liquid-liquid extraction, offers reduced processing times, less labor, and significant saving in glassware and solvent usage. The extraction solvent is dried on sodium sulfate and concentrated in a rotary evaporator. Depending on the concentration levels of other organic compounds that may interfere with the final PAH analysis and the method to be used, it may be necessary to fractionate the sample. Sample cleanup is achieved using column chromatographic techniques.

The concentration levels of PAHs in groundwater and drinking water supplies are of primary concern because a dose-related increase in the incidence of stomach cancer tumors in mice fed on BaP has been reported. The European Community and World Health Organization (WHO) have specified the analysis of six indicator PAHs (fluoranthene, benzo(*b*)fluoranthene, benzo(*k*)fluoranthene, benzo(*a*)pyrene, benzo(*ghi*)perylene, and indeno(1,2,3-*cd*)pyrene) for assessment of the quality of water for human consumption, setting a maximum admissible concentration for the sum of the six PAHs in potable waters of  $200\text{ ng l}^{-1}$  and maximum admissible limit for concentration of BaP at  $0.010\text{ }\mu\text{g l}^{-1}$ . Analysis of water samples obtained from groundwater and drinking water supplies in Germany in 1974 showed that the total concentration of the six indicator PAHs exceeded  $0.11\text{ }\mu\text{g l}^{-1}$  in only 1% of samples and was between  $0.001$  and  $0.01\text{ }\mu\text{g l}^{-1}$  for 90% of the samples. Analysis of water for marine culture in China (Daya bay) for 16 PAHs gave a total of  $4228\text{--}29\,325\text{ ng l}^{-1}$  with BaP concentration at sites as high as  $>10\text{ }\mu\text{g l}^{-1}$ , which can cause acute toxicity. Chlorination of water can remove 50–60% of dissolved BaP while filtration through activated charcoal removes  $>99\%$ . Due to lack of sufficient information, WHO has been unable to set maximum admissible concentrations for individual PAHs except for BaP (maximum admissible concentration of  $0.01\text{ }\mu\text{g l}^{-1}$  was set in 1984). However, methods to reduce the concentration of BaP will result in the reduction of the concentrations of all PAHs. It has been estimated that food accounts for 99% of the daily oral intake of total PAHs with drinking water contributing only 0.1%.

During water distribution, contact with coal-tar-based pipe linings lead to marked increases in PAH concentrations in the water. For this reason, WHO has recommended discontinuation of coal-tar-based pipe lining for drinking water distribution.

## Sediments

Due to their low water solubility and hydrophobic nature, PAHs tend to associate with particulate



matter. Accumulation of PAHs is dependent on sediment type since those with high organic content and smaller particle size (large surface area to volume ratio) have a higher potential to accumulate PAHs compared to coarser sandy sediments. Mussels, which are sedentary, filter feeding, and low-metabolism aquatic organisms, are widely used as sentinels for pollution monitoring of the aquatic environment. They take up and concentrate contaminants to levels above those in the surrounding waters. Particle surface properties and organic matter content are key factors for assessing the bioavailability of PAHs to sedentary organisms. **Table 2** shows a comparison of the total priority USEPA list of PAHs analyzed in water, pore water, sediment, and mussel in different marine sites. In all cases, water has the lowest concentration of PAHs compared with sediments and edible mussel. There is good correlation between BaP and total PAHs ( $r=0.953$ ) for surface sediments. Attempts to replace living organism for measuring contaminants in marine waters with semipermeable membrane devices have been unsuccessful due to different PAH accumulation patterns.

**Table 2** Polycyclic aromatic hydrocarbons in water and sediment in various sites

Location	Matrix	Concentration range
Daya Bay, China <sup>a</sup>	Water	4228–29 325 ng l <sup>-1</sup>
	Sediment	Dry mass 115–1134 ng per g
Bay of Marmara sea, Izmit Bay <sup>b</sup>	Water	1.16–13.68 µg l <sup>-1</sup>
	Sediment	Dry mass 30.0–1670.0 µg per g
	Edible mussel	Wet mass 5.67–14.81 µg per g
Jiulong river estuary and western Xiamen sea, China <sup>c</sup>	Water	6.96–26.9 µg l <sup>-1</sup>
	Sediment	Dry mass 59–1177 ng per g Pore water 158–949 µg l <sup>-1</sup>
Pearl River Delta, China <sup>d</sup>	Sediment	Dry mass 156–9220 ng l <sup>-1</sup>

<sup>a</sup>Zhou JL and Maskaoui K (2003) Distribution of polycyclic aromatic hydrocarbons in water and surface sediments of Daya Bay, China. *Environmental Pollution* 121: 269–281.

<sup>b</sup>Telli-Karakoc F, Tolun L, Henkelman B, Klimm C, Okay O, and Schramm KW (2002) Polycyclic aromatic hydrocarbons (PAHs) and polychlorinated biphenyls (PCBs) distributions in bay of Marmara: Izmit Bay. *Environmental Pollution* 119: 383–397.

<sup>c</sup>Maskaoui K, Zhou JL, Hong HS, and Zhang ZL (2002) Contamination of polycyclic aromatic hydrocarbons in the Jiulong river estuary and Xiamen sea, China. *Environmental Pollution* 118: 109–122.

<sup>d</sup>Mai BX, Fu JM, Sheng GY, Kang YH, Lin Z, Zhang G, Min YS, and Zeug EY (2002) Chlorinated and polycyclic aromatic hydrocarbons in riverine and estuarine sediments of pearl river delta, China. *Environmental Pollution* 117: 457–474.

## Food

PAHs have been reported in various types of foods including root and leaf vegetables, vegetable oils, cereals, smoked or grilled fish and meats, roasted foods, and seafood. The source of PAHs in food could either be from the environment or from food preparation or preservation processes. Analysis of PAHs in a food product requires quantitative isolation of the PAH by saponification/digestion in alcoholic potassium hydroxide followed by liquid–liquid extractions, drying over anhydrous sodium sulfate, and cleanup using column chromatographic techniques. For highly insoluble samples, e.g., meats and fish, complete hydrolysis is an absolute necessity. Samples that are highly soluble in hexane or methanol, e.g., oils, are processed directly.

Atmospheric PAHs are deposited on grain foliage, leafy vegetables (e.g., spinach, kale), fruits, and oil plants while PAHs deposited on soil may be absorbed in the roots and translocated to other plant parts. Processing of PAH-contaminated vegetable raw materials gives vegetable oils containing PAHs. PAHs have been reported in soyabean, cottonseed, corn, olive, sunflower, rapeseed, palm, peanut, linseed, and coconut oils. A BaP concentration as high as 43.7 µg per kg in oil from smoke-dried coconut has been reported. Mineral oils and refined petroleum products, which potentially contain PAHs, are ingredients of many cosmetic preparations (e.g., sun-tan oils, cold creams) and medicinal products. To remove the PAHs, oils are treated with activated charcoal.

Smoking, grilling, and the direct heating of foods with air containing combustion gases increases their PAH levels. Products smoked for long periods to blackness have been found to contain high levels of PAHs. During the smoking of food, PAHs on the surface migrate to the interior of food, the extent of migration being dependent on the character of the food product, and its storage time. For example, in smoked bacon, at least 50% of PAHs will be found in the fat drippings. Research into smoked food packaging has shown that cotton fabric and cellophane reduce the PAH content of the smoked meats without removing the smoked flavor and preservative constituents.

PAHs are taken up by both leaves and roots, translocated to other parts of the plant, and may also be concentrated more in certain plant parts than others. Sorghum irrigated with PAH contaminated water has been reported to show an accumulation of fluoranthene and pyrene. The uptake and translocation rates vary with different plant species. The rate (amount per unit time) of PAH uptake by

plants is also dependent on PAH concentration, solubility, phase (vapor or particulate bound), relative molecular mass, and the nature of substrate in which the plant is growing. Analysis of carrots grown on soil containing high levels of PAHs has shown that carrot foliage PAH concentration was unaffected by PAH soil loadings. However, the root peel (<1 mm thick) PAH concentrations increased to a maximum ( $\sim 200 \mu\text{g}$  per kg dry mass for a sum of 14 PAHs) with increasing soil PAH levels, tending to plateau above  $500 \mu\text{g} \sum \text{PAH kg}^{-1}$  (sum of 14 PAHs) in the soil. The carrot core PAH concentrations were significantly lower than peel concentrations. A maximum value of  $38.9 \mu\text{g}$  per kg dry mass of carrot peel for the sum of benzo(*a*)anthracene, chrysene, benzo(*a*)pyrene, and benzo(*ghi*)perylene in the carrots grown on heavily PAH contaminated soils was observed, compared with  $10.7 \mu\text{g}$  per kg dry mass of carrot core for the sum of the same PAHs.

A very low percentage of PAHs in the environment is reported to arise from microbial and plant synthesis. Conversely, some PAH molecules can be catabolized by certain plant species, e.g., anthracene by soyabean. This catabolism is slow and depending on the plant species, transformations of 2–18% of assimilated BaP, mostly to organic acids, over a 14-day period have been reported.

Monitoring of PAHs in a diet is important because of the hypothesis that a correlation exists between a high incidence of gastrointestinal cancers in some populations and levels of PAHs in their diet. PAHs have been found in river and sea sediment, seaweed, and phytoplankton that serve as food for fish and marine filter-feeding bivalves. Values for BaP levels reported in nori, a Japanese seaweed, range from 7.4 to  $31.3 \mu\text{g}$  per kg and benzo(*a*)anthracene in clams purchased from Rhode Island stores (USA) range from 0.1 to  $0.8 \mu\text{g}$  per kg wet mass.

Food can be a major environmental source for PAH exposure to man depending on the geographical location, occupation, lifestyle, and dietary habits. Elevated PAH levels are found in smoked, grilled, and roasted foods. The amounts of these foods consumed vary between regions and therefore some segments of the world population are exposed to high levels of PAHs in the diet. The mean total exposure for a reference man of  $312 \mu\text{g day}^{-1}$  (total PAHs) for a nonsmoker, food contributed 96.2%, air 1.6%, water 0.2%, and soil 0.4%. Smokers get (one pack nonfiltered per day) an additional  $1\text{--}5 \mu\text{g day}^{-1}$ . Most countries do not have maximum acceptable levels of PAHs in foods. In Germany, Austria, and Poland, a limit of  $1 \mu\text{g}$  per kg for BaP in smoked foods must not be exceeded.

## Biological Matrices

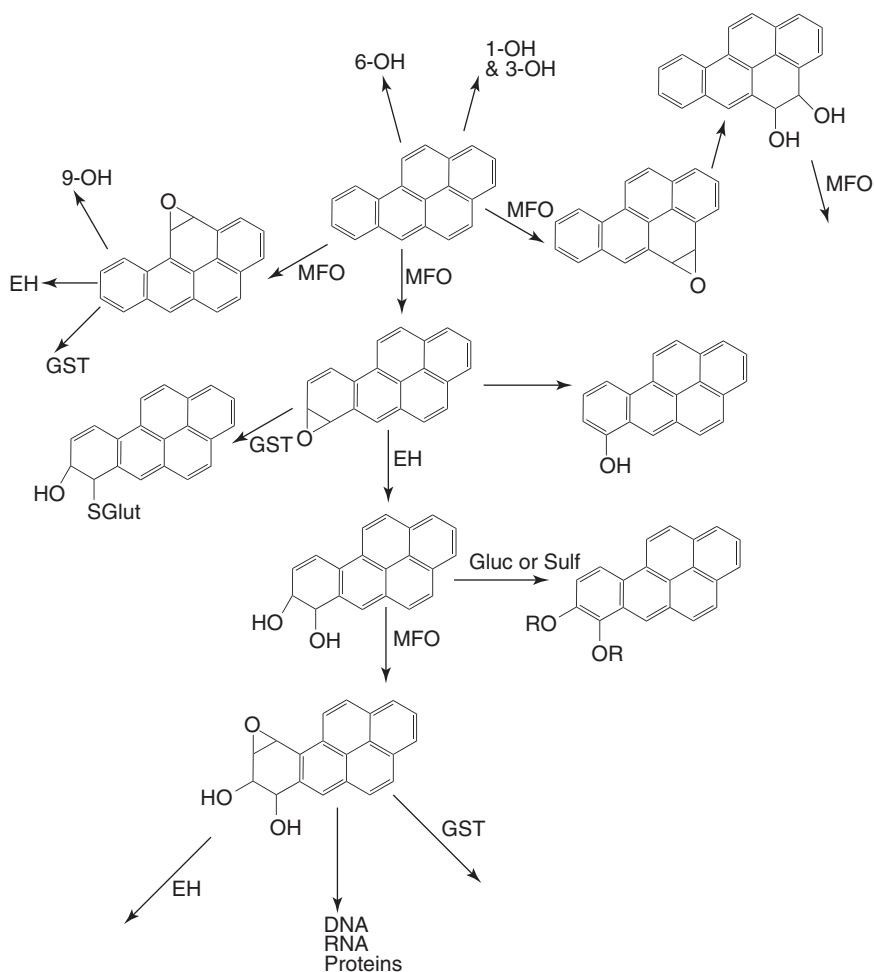
PAHs are enzymatically converted in mammalian cells to polar reactive intermediates, capable of covalently binding to cellular macromolecules. This metabolism is complex, involving many steps and enzymes. Most of these intermediate species are converted through secondary metabolic processes in the liver to form inactive products (e.g., glucuronides, sulfates, glutathione conjugates). These are excreted in urine and bile and it is only those products that escape these secondary reactions that react with nucleic acids and proteins, and probably lead to carcinogenesis.

BaP and other PAHs produce liver and lung tumors within half a year following intraperitoneal and subcutaneous injection to new born animals. Skin tumors after dermal application of PAHs to mice provides a background for the initiation/promotion theory in chemical carcinogenesis. BaP is the only PAH that has been tested following inhalation. Long-term inhalation of  $10 \text{ mg BaP per m}^3$  resulted in cancer of the respiratory tract in 35% of golden hamsters.

Many tissues in the human body, including liver, lung, colon, and lymphocytes, have the ability to metabolize BaP. The most important enzymes that mediate the initial oxidation of PAHs in the body are cytochrome P<sub>450</sub> mixed function oxidases (MFOs), which are found in microsomes, nuclei, and mitochondria of cells. MFO enzyme systems function together with NADPH-cytochrome P<sub>450</sub> reductase and require phosphatidylcholine for activity.

Figure 1 shows the principal metabolic pathways of BaP (a representative PAH) and the formation of some conjugates. In the interest of simplicity, further metabolic transformations, such as the P<sub>450</sub> catalyzed oxidation of phenols to phenol-epoxides, conversion of phenols to phenol-dihydrodiols, all the enzymatic steps, and the stereochemistry of the metabolites are omitted.

The principal metabolites formed are arene oxides, which are relatively unstable and are further modified by other enzymes (e.g., epoxide hydrolase (EH) and glutathione-S-transferase (GST)) present in the endoplasmic reticulum. The highest concentration of GST enzymes is found in the liver and most of the glutathione conjugates are preferentially excreted in the bile. In the kidney, glutathione conjugates are biochemically converted to mercapturic oxides, which are excreted in urine. Isomerization of arene oxides, which is assumed to be nonenzymatic (though it can be catalyzed by amino or sulfhydryl groups in proteins) also takes place, resulting in the production of phenols. Single phenolic isomers, in line with theoretical molecular orbital predictions based on the relative stability of the two possible cationic intermediates, are formed predominantly. Further



**Figure 1** Selected metabolic pathways of benzo(a)pyrene. Gluc, glucuronidase; Sulf, sulfatase; R, glucuronate or sulfate; MFO, mixed function oxidases; GST, glutathione-S-transferase; EH, epoxide hydrolase.

oxidation to quinones also takes place. Since the diol epoxides are highly active chemically, they are not normally found among urinary metabolites.

PAH *trans*-dihydrodiols undergo further oxidation, catalyzed by MFOs, to diastereomeric epoxide metabolites. These diol epoxides are poor substrates of EH and undergo nonenzymatic hydration in aqueous media to give tetrol products. The diol epoxide metabolites are principally detoxified through their reaction with glutathione. The phenols, dihydrodiols, and other polar metabolites of PAHs undergo conjugation (mediated by transferase enzymes) with glucuronic and sulfuric acids to form more water-soluble metabolites.

The enzymatic metabolism of PAHs proceeds with high stereoselectivity, giving dihydrodiols with exclusively *trans*-configuration. The principal metabolites isolated for BaP are 4,5-, 7,8-, 9,10-dihydrodiols, 1-, 3-, 7-, 9-phenols, and 1,3-, 1,6-, 6,12-quinones.

An ideal biological monitoring method would permit the direct measurement of PAHs in the target organs. Since routine sampling of tissue from the target organs is impractical, body fluids (mainly urine and blood) are generally used. Unmetabolized BaP has been extracted from urine with dichloromethane, separated by liquid chromatography (LC), and quantified by spectrofluorimetry. C<sub>18</sub> cartridges have also been used for the extraction and concentration of PAHs and metabolites from urine samples. Urinary excretion of BaP and benzo(*a*)anthracene have been found to be significantly higher (>60%) in persons living in highly industrialized areas compared with those living in rural environments. An alternative method is to analyze the urine for specific PAH metabolites that are consistently prominent in environmental samples. The urinary concentration of one metabolite, 1-hydroxypyrene, has been used as a biomarker of PAH exposure. Measurement of metabolites of several PAHs in urine would better

predict occupational exposure. For road paving workers, the level of 1-hydroxypyrene in postshift urine samples has been reported to vary from 0.1 to  $2.8 \mu\text{mol mol}^{-1}$  creatinine. There is a strong correlation between urinary 1-hydroxypyrene and the air concentration of pyrene. Significant correlations have also been observed between urinary 1-hydroxypyrene of coke-oven workers and city residents and levels of pyrene and BaP in ambient air. In another approach, the metabolized PAHs in the urine are reduced back to the parent PAHs by refluxing with hydriodic acid.

Hydrophobic interaction between PAHs and blood serum albumin plays an important role in PAH transportation in the body. Activated carcinogens bind covalently to nucleic acids and proteins and therefore quantitation of the PAH-DNA adduct formed in the body may be used as a measure of the biologically active dose of the metabolically activated PAH. Measurement of the extent of adduct formation of PAHs and their metabolites with haemoglobin in blood samples taken from exposed individuals has been proposed as a marker of the biologically effective dose. Haemoglobin serves as a convenient trapping agent and may be used for assessing the levels of activated carcinogens *in vivo* since it can be obtained in large amounts from red blood cells, and is readily purified and analyzed.

Sensitive immunoassays specific for PAH-DNA adducts allow the detection of one adduct per  $10^6$  nucleotides. Fluorimetry has also been used as an alternative detection method to immunoassays. In another method, isomeric tetrols of PAH are liberated by acid hydrolysis of the DNA-PAH adducts and analyzed by LC with fluorescence detection. Structural studies and detailed characterization of PAH metabolites and their conjugates has been performed by trapping the corresponding fractions at the exit of the LC system. As an example, the total characterization of the *in vitro*-formed benzo(a)pyrenetetrahydrodiol-epoxide-guanosine adduct has been achieved by a combination of nuclear magnetic resonance, circular dichroism, and mass spectroscopic techniques.

In the environment, bacteria that have the ability to degrade PAHs have been reported. Saprotrophic soil fungi *Fusarium solani*, and yeast *Rhodotorula glutinis* metabolize pyrene as a sole source of carbon. Mushroom compost is reported to be used directly to ameliorate phenanthrene contaminated soil. After 111 days incubation time,  $36.7 \pm 2.9\%$  loss of soil-associated phenanthrene was observed. Exposure of

PAHs absorbed on spruce (*Picea abies* (L.) Karst) needles to sunlight irradiation showed photolysis of 18 PAHs under the study to follow first-order kinetics, photolysis half-lives ranging from 15 h for Db(a,b)A to 75 h for phenanthrene.

By detecting and quantitating PAHs and their metabolites in body fluids, and measuring the effective dose of PAH at the cellular level, the amounts of PAH that has actually entered the body (through ingestion, inhalation, and/or skin absorption) can be measured. This will result in a better assessment of human risk associated with exposure to PAH from different environments and a further insight into chemical carcinogenesis.

**See also:** **Air Analysis:** Sampling; Outdoor Air; Workplace Air. **Environmental Analysis. Fluorescence:** Environmental Applications. **Mass Spectrometry:** Environmental Applications. **Polycyclic Aromatic Hydrocarbons:** Determination. **Sample Handling:** Sample Preservation; Automated Sample Preparation. **Sampling:** Practice.

## Further Reading

- Bjorseth A and Ramdahl T (1985) *Handbook of Polycyclic Aromatic Hydrocarbons*, vol. 2. New York: Dekker.
- Cooke M, Loening K, and Merritt K (1991) *Polynuclear Aromatic Hydrocarbons: Measurements, Means and Metabolism*. Columbus: Battelle Press.
- Davis CS, Fellin P, and Otson R (1987) A review of sampling methods for polyaromatic hydrocarbons in air. *International Journal of Air Pollution Control and Hazardous Waste Management* 37: 1397-1407.
- Edwards NT (1983) Polycyclic aromatic hydrocarbons in the terrestrial environment - a review. *Journal of Environmental Quality* 12: 427-441.
- Harvey RG (1991) *Polycyclic Aromatic Hydrocarbons: Chemistry and Metabolism*. Cambridge: Cambridge University Press.
- Hester RE and Harrison RM (1998) Air pollution and health. *Issues in Environmental Science and Technology*, RSC.
- Her Majesty Stationary Office (1988) The determination of 6 specific polynuclear aromatic hydrocarbons in waters (with notes on the determination of other PAH) 1985. In: *Methods for the Examination of Waters and Associated Materials*. London: HMSO.
- Wild SR and Jones KC (1992) Polynuclear aromatic hydrocarbon uptake by carrots grown in sludge-amended soil. *Journal of Environmental Quality* 21: 217-225.
- WHO (1984) *Health Criteria and Other Supporting Information*, vol. 2, GP. WHO/328. Geneva: WHO.

# POLYMERASE CHAIN REACTION

**J M Wages Jr.**, Palmetto Consulting & Research, Tupelo, MS, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Polymerase chain reaction (PCR) is a technology for exponential amplification of a fragment of DNA. (The PCR is covered by patents owned by Hoffman-La Roche. A license is required to use the PCR process.) The limit of its sensitivity is a single molecule, making PCR a superb qualitative tool for the specific detection of rare DNA sequences. Under proper conditions, the yield of amplified DNA is proportional to the initial number of target molecules, rendering it a quantitative analytical tool as well. Since its original description in 1985, PCR has evolved into an assemblage of varied methodologies almost universally used in basic biological research, biotechnology, clinical research, clinical diagnostics, forensics, food technology, environmental testing, archaeology and anthropology, and other fields. Even though other nucleic acid amplification technologies have been described, PCR remains by far the most widely used.

## Biochemical Basis of PCR

PCR involves the enzymatic synthesis of millions of copies of a specific DNA segment. The exponential amplification of a very small amount of template DNA is achieved using a heat-stable DNA polymerase and an automated heat block that is capable of rapid changes of temperature. The template DNA molecule is initially denatured to two single strands by heating to high temperature (typically 90–95°C) in the first stage of the PCR cycle (**Figure 1A**). Two small oligonucleotides that are complementary to sequences on opposite strands of the template molecule and that flank the DNA segment to be amplified are used as primers for the DNA polymerase. The second stage of the PCR cycle consists of cooling the reaction, which permits the annealing of the single-stranded oligonucleotide primers to the denatured template molecule (**Figure 1B**). The third stage of the reaction is the extension of the new strand from the annealed primer in a 5' to 3' direction by the heat-stable polymerase (**Figure 1C**). This is performed at the optimum temperature for the polymerase (68–72°C). The most commonly used polymerase is the enzyme isolated from *Thermus aquaticus* (*Taq* DNA polymerase). After the first cycle, each template

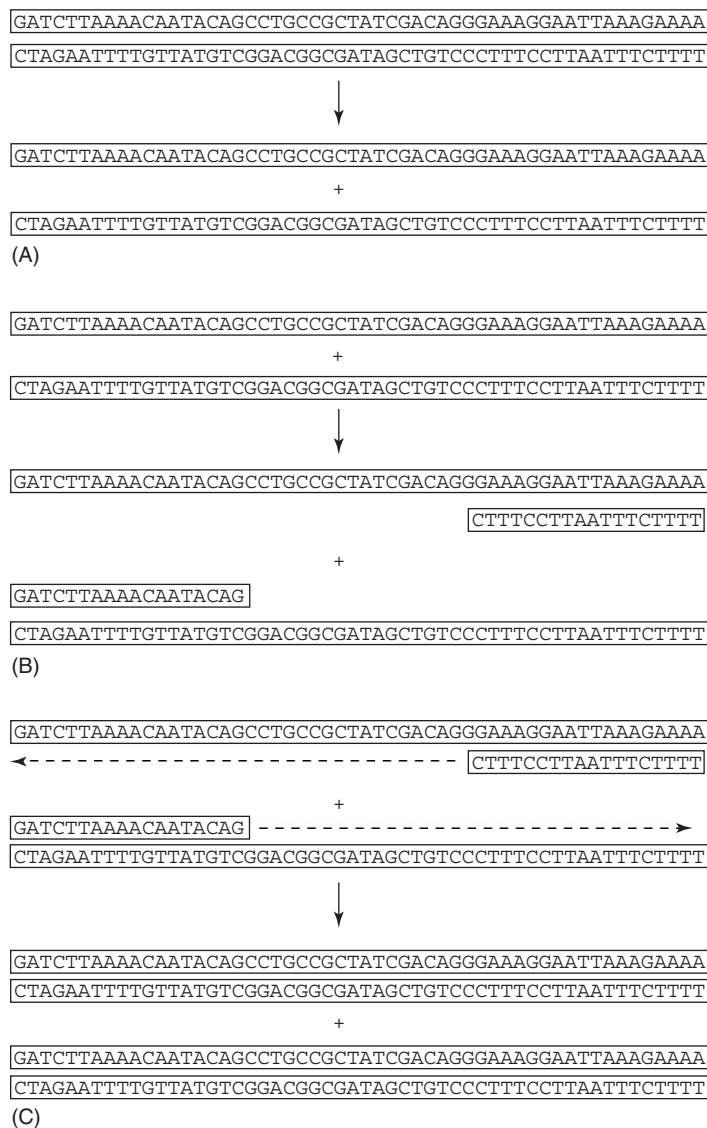
molecule has been amplified to two molecules. These in turn are denatured in the next cycle and amplified to produce four molecules. The four molecules are amplified to eight in the third cycle, and so on. Each successive cycle effectively doubles the amount of DNA product. The three-stage cycle of denaturation, annealing, and primer extension is repeated 25–40 times in a typical PCR procedure. The product of such a reaction is a large quantity ( $10^{-9}$  to  $10^{-8}$  mol l<sup>-1</sup> in a 10–100 µl volume) of a double-stranded DNA molecule whose length is determined by the distance between the primer sites on the original template molecule. Typically, PCR products or amplicons are 100–3000 bp in length, although much longer (up to ~50 000 bp) amplifications are possible under specific conditions. PCR products are visualized by electrophoresis followed by staining with fluorescent dyes or by hybridization to labeled oligonucleotide probes.

## Technical Aspects of PCR

One of the advantages of PCR technology is the speed and simplicity with which the technique can be performed. A PCR reaction can be set up in a short time, and multiple samples can be handled easily. Components of the amplification reaction (template DNA, DNA polymerase, oligonucleotide primers, buffers) are added to the amplification vessel. The most common containers for PCR are microcentrifuge tubes (0.5–0.2 ml) with thin walls to facilitate rapid heat transfer or multiwell plates (96- or 384-well plates being the most common). Some commercially available systems utilize specially designed vessels such as capillary tubes. The reaction tubes are then placed into an automated programmable thermal cycler for controlled heating and cooling for denaturation, annealing, and extension reactions. Following amplification, some type of DNA sequence analysis is typically performed, although technology for real-time monitoring of amplification obviates the need even to open the PCR tube after cycling is complete.

### Template DNA

One of the advantages of PCR is that DNA from relatively clean samples need not be highly purified to be used as a template for amplification. It is possible to simply heat samples to near boiling to release DNA, then to add the resulting lysate to the PCR reaction tube. Such simplicity is the hallmark of PCR. With some types of biological samples, inhibitors of



**Figure 1** One cycle of a PCR. (A) The template is shown as a double-stranded DNA molecule, which is denatured to two single strands. (B) The short oligonucleotide primers anneal to the complementary regions of the single strands. (C) The polymerase generates a new strand from each single-stranded molecule, creating two double-stranded molecules.

PCR such as heme or polyanionic mucopolysaccharides must be removed by DNA purification.

Virtually any form of DNA can be used as a template in a PCR reaction. Plasmids, cosmids, phagemids, lambda phage, M13 phage, genomic DNA, and many other sources of DNA have all been used successfully. A PCR can be performed directly on colonies or plaques. Although it is possible to amplify a single DNA molecule, a typical PCR reaction uses  $10^5$ – $10^6$  copies of the target DNA as template. In practice, this means adding  $\sim 1 \mu\text{g}$  of eukaryotic genomic DNA, 10 ng of yeast DNA, or 1 ng of bacterial DNA. The concentration of DNA is not critical for most applications, however, and a PCR reaction will work with a wide range of template concentrations.

Often the concentration of template DNA is unknown and does not need to be determined. If necessary, a series of control reactions, each containing a different amount of template, can be performed.

### Biochemical Components

**DNA polymerase** A wide range of DNA polymerases is available for PCR. A cloned heat-stable DNA polymerase from *Thermus aquaticus* (Taq DNA polymerase) is the most commonly used. Other enzymes such as DNA polymerase from *Pyrococcus furiosus* (Pfu DNA polymerase) or *Thermococcus litoralis* (Vent<sup>TM</sup> DNA polymerase, New England Biolabs) have been reported to have a lower error rate



than *Taq* polymerase and may be advantageous for some applications. Typically, 0.5–2.5 units of enzyme are used per 50  $\mu$ l reaction.

**Oligonucleotide primers** Oligonucleotide primers are usually used at a concentration of  $1 \mu\text{mol l}^{-1}$  in PCR reactions (50 pmol per 50  $\mu$ l reaction). Concentrations between 0.1 and  $1 \mu\text{mol l}^{-1}$  can be used. Higher concentrations may increase nonspecific annealing of primers and thus to nonspecific amplification products. PCR primers are normally between 18 and 30 nucleotides in length and should preferably have a guanine + cytosine (G + C) content of  $\sim 50\%$ . The temperature at which half the DNA molecules will be double-stranded,  $T_m$ , in degree celsius, for a primer is estimated by the rule-of-thumb equation:  $2 \times (\text{number of As and Ts}) + 4 \times (\text{number of Gs and Cs})$  (A = adenine, T = thymine). However, more accurate calculation of oligonucleotide  $T_m$  requires consideration of the effects of ionic strength and neighboring bases in the DNA strand (nearest neighbor  $T_m$  calculation methods). Software is available for performing these calculations. The  $T_m$  values for the two primers in a reaction should be similar, and the annealing temperature used is normally  $\sim 5^\circ\text{C}$  below the  $T_m$ . As annealing temperature approaches  $T_m$ , more specific amplifications are achieved, but overall yields may decrease. Despite careful calculations, empirical testing of annealing temperatures is essential for a well-optimized PCR assay. Complementarity at the 3' ends of primer pairs should be avoided as this promotes the formation of primer oligomers (primer dimers). Such artifactual products are themselves templates for PCR amplification and compete with the desired products for deoxynucleotide triphosphates (dNTPs), polymerase, and primers. This leads to a decreased yield of the desired product and the presence of nonspecific products that may complicate analysis.

A primer oligomer forms when two or more primers anneal to each other and are extended during PCR. The double-stranded product acts as a template during subsequent rounds of amplification and competes with the desired product. In general, low annealing temperatures, high enzyme concentrations, and high primer concentrations increase the probability of primer oligomerization. However, it is stringency during the initial cycle of PCR that is most critical for primer oligomer formation, as well as for nonspecific PCR products in general. Various techniques for maintaining high stringency up until the point when primers and enzyme are mixed together have been described (generally called Hot Start). Hot Start methods involve mixing of enzyme and primers only after reactions have reached a

stringent temperature. A simple technique for Hot Start is manual addition of a small volume (typically 5–10  $\mu$ l) of reaction buffer containing the enzyme to the other components maintained in the thermal cycler at  $\sim 80^\circ\text{C}$ , then continuing with standard PCR amplification. Another method involves sealing a portion of the amplification reaction under wax, then adding the remaining reactants above the wax barrier. Enzymatic methods to accomplish the same goal include the use of a modified DNA polymerase that is blocked with a thermolabile group or the use of uracil DNA glycosylase (uracil-N-glycosylase, UNG), deoxyuridine triphosphate (dUTP), and a brief initial incubation at  $50^\circ\text{C}$ . Typically, dUTP is substituted on an equimolar basis for deoxythymidine triphosphate (dTTP) ( $200 \mu\text{mol l}^{-1}$ ), but higher concentrations of dUTP ( $125\text{--}300 \mu\text{mol l}^{-1}$ ) may be beneficial in some assays. Uracil DNA glycosylase is used at 1–2 Units per reaction. This method ensures that any nonspecific product or primer oligomer that is generated during the initial low-stringency conditions will contain uracil (U) and therefore be susceptible to cleavage by UNG. Uracil is excised during the  $50^\circ\text{C}$  incubation, and strand breakage occurs at these abasic sites during the initial denaturation step. In addition to inactivating any contaminating amplicons, this method reduces nonspecific products and primer oligomer. Because residual UNG can degrade the U-containing amplicons, PCR products must be analyzed immediately or stored at  $4^\circ\text{C}$  for short periods of time or frozen for longer times. Alternately, UNG can be removed by extraction with organic solvents or digested with proteases.

The 3' end of a PCR primer should be perfectly complementary to the template DNA. Failure of the 3' end to hybridize results in inefficient amplification. Internal sequences are less critical. Provided a suitable annealing temperature is used, degenerate primers (primers that consist of a pool of different but closely related oligonucleotides) may be used to ensure that primers will anneal to highly variable sequences or sequences that are only partially known. These primers are designed from amino acid sequences or from a comparison of similar genes from other organisms. Sequences at the 5' end of the primer are least critical. Providing the hybridizing portion of the primer is long enough to ensure annealing to the template DNA, nonhomologous 5' extensions (regions that do not correspond to the sequence of the template) can be used to introduce restriction enzyme sites into PCR products to be cloned or to add other sequences such as phage RNA polymerase promoters.

Multiplex PCR enables the detection of multiple gene sequences within the same reaction. Because several sets of PCR primers must function in the

same reaction without interference, primer design and optimization of reaction conditions are critical.

Primers are often labeled with detectable groups to facilitate post-PCR analysis. Radioactive isotopes, haptens such as biotin, or fluorescent dyes are the most widely used. Labels attached to the 5' end of a primer have little or no effect on amplification.

Many heat-stable polymerases used for PCR exhibit a template-independent activity that adds deoxynucleosides (predominantly deoxyadenosine) to the 3' ends of all double-stranded molecules in the reaction. These A-overhangs must be filled in prior to blunt-end cloning of PCR products. Alternatively, special vectors are available that have single 3' T-overhangs at the cloning site ready for insertion of the PCR product. Choosing a polymerase without this activity or adjusting reaction conditions to either minimize or maximize the extent of A-overhangs may produce sharper peaks in high-resolution electrophoretic analysis.

**Deoxynucleoside triphosphates** dNTPs are typically used at a concentration of  $\sim 200 \mu\text{mol l}^{-1}$  each dNTP in a PCR reaction. Excessively high concentrations promote nonspecific product formation. Modified dNTPs are sometimes used to label PCR products with radioactive or fluorescent markers or with haptens such as biotin, fluorescein, or digoxigenin. The modified dNTP is typically used at a concentration ( $0.1\text{--}1 \mu\text{mol l}^{-1}$ ) much lower than the unmodified dNTP ( $\sim 200 \mu\text{mol l}^{-1}$ ). Probes can be easily generated using PCR amplification with a labeled nucleotide, followed by removal of the unincorporated label.

**Buffer components** Several reagents containing buffer ions, monovalent salts, and divalent cations required for polymerase activity, have been described for use in PCR amplification. The most widely used buffer consists of  $10 \text{ mmol l}^{-1}$  Tris-HCl (pH 8.3 at room temperature) and  $50 \text{ mmol l}^{-1}$  KCl. At the extension temperature ( $72^\circ\text{C}$ ), the pH of this Tris buffer falls to 7.2, near the optimum for *Taq* DNA polymerase. Sulfate-containing buffers such as  $20 \text{ mmol l}^{-1}$  Tris- $\text{SO}_4$  (pH 8.5–9.0 at room temperature) and  $20 \text{ mmol l}^{-1}$   $(\text{NH}_4)_2\text{SO}_4$  are also widely used. Monovalent cations ( $\text{K}^+$  or  $\text{NH}_4^+$ ) are included to adjust ionic strength. DNA polymerases require divalent cations for activity, and PCR reactions contain  $\text{MgCl}_2$  ( $1\text{--}5 \text{ mmol l}^{-1}$ ). In general, higher  $\text{MgCl}_2$  concentrations promote nonspecific primer annealing and nonspecific product amplification. Reaction buffers usually include a low percentage (0.1–0.5%) of nonionic detergent such as Igepal CA-630, Tween 20, or Triton X-100 to minimize adsorption of polymerase to the surfaces of the PCR tube.

Proteins (gelatin or bovine serum albumin) are sometimes included at similar concentrations as nonionic detergents for the same reason. Other components that have been reported to increase PCR product yields or specificity on specific templates include betaine, dimethylsulfoxide, formamide, and glycerol.

### Thermal Cycling Equipment

The PCR process involves thermal cycling between denaturation, annealing, and extension temperatures (Figure 1). Each incubation step is a segment, and a complete round of denaturation, annealing, and extension is a cycle. First, samples are heated to a temperature adequate to melt the double-stranded DNA template and any secondary structure in the primers. Denaturation is most commonly performed at  $94^\circ\text{C}$ , although temperatures of  $97\text{--}99^\circ\text{C}$  may be required for templates with high GC content. After the first several cycles, denaturation temperature can be lowered to  $90\text{--}92^\circ\text{C}$  to reduce loss of polymerase activity through heat denaturation during the course of PCR. Next, samples are cooled to an annealing temperature at which the primers will hybridize to their target sequences. The choice of annealing temperature depends on primer  $T_m$ , but is usually in the range of  $50\text{--}60^\circ\text{C}$ . Finally, the samples are heated to the extension temperature ( $68\text{--}72^\circ\text{C}$ ). If primers are designed to permit efficient annealing at  $60\text{--}68^\circ\text{C}$ , the annealing and extension steps can be combined into one step. Typical PCR amplification is performed over 20–30 cycles, although 40 or more cycles may be used with template DNA at very low initial concentration or with otherwise marginal amplifications involving, for example, degenerate primers or poor-quality template DNA.

Although it is possible to perform 20 or even 40 cycles of PCR manually using sandbaths, waterbaths, or heat blocks and physically moving the samples from one to the other at timed intervals, the tedium involved makes this approach impractical. Many different types of machines are commercially available to perform thermal cycling. Some cyclers use robotic arms to transfer a rack of PCR samples from one heat block to another. Other machines use resistive heating elements, compressors, water circulators, fan-forced air, high-intensity lamps, Peltier elements, or combinations of these heating and cooling devices. Microprocessors enable the user to program the instrument for time and temperature of each segment, and the number of cycles to be performed.

Thermal cyclers that can monitor progress of the amplification reaction while it is taking place are called real-time thermal cyclers. Real-time instruments are designed to monitor fluorescence from labels whose emission intensity is proportional to the

amount of amplified DNA. Flexibility varies greatly among the available machines, but all instruments can measure fluorescence at least once during the annealing segment of every cycle. Intercalating dyes such as ethidium bromide or SYBR Green that are selective for double-stranded DNA provide the simplest method. Probes that hybridize adjacent to one another on one strand of the amplicon can be designed to have a fluorescence energy transfer donor and acceptor that will be close enough for energy transfer if and only if hybridization occurs. Emission from the hybridized donor- and acceptor-labeled probes is monitored during each cycle. A third approach is to use the 5' nuclease activity of *Taq* DNA polymerase to cleave a probe labeled with a fluorophore and a quencher. When *Taq* DNA polymerase (and some, but not all, other heat-stable polymerases) encounters the probe, it cuts it, dissociating the fluorophore and quencher. As with energy transfer methods, amplification is detectable as an increase in fluorescence. In addition to monitoring fluorescence, real-time thermal cyclers provide data analysis for quantitation of initial template concentration. Real-time PCR has become the standard for quantitative PCR.

### Post-PCR Analysis

Unless real-time thermal cyclers are used, some form of manipulation or analysis of PCR products must be performed. In preparative applications, amplicons are inserted into a suitable vector and propagated by standard molecular biological methods. In analytical applications, some form of DNA size or sequence analysis is performed. In many cases, high-performance liquid chromatography or electrophoresis on agarose or polyacrylamide gels to confirm that the amplicon is the expected size is sufficient. In clinical testing, hybridization of an oligonucleotide probe complementary to a sequence between the PCR primers is more commonly used to confirm specificity of the amplified DNA. Probe hybridization can be combined with DNA size analysis by electrophoresis or can be performed using methods analogous to immunoassays.

## Considerations

### Fidelity

For preparative applications and the analysis of single-base changes, fidelity of the DNA polymerase is important. The fidelity of the DNA polymerase in a PCR reaction determines the similarity between the sequence of the PCR product and the original template. *Taq* DNA polymerase sometimes incorporates an incorrect base in the growing DNA strand, and this error is propagated during subsequent rounds of cycling. The higher the fidelity, the more closely the sequence

of the PCR product reflects the sequence of the original DNA template. A range of DNA polymerases is commercially available, and certain enzymes have demonstrably higher fidelity. In addition to intrinsic differences among the various heat-stable polymerases, fidelity is influenced by a number of factors. It is therefore possible to increase or decrease the number of mismatches between the product and initial template. Conditions of low fidelity may be chosen to introduce random point mutations into a PCR product. More general applications benefit from accurate amplification. Fidelity is improved by keeping polymerase, dNTP, and  $MgCl_2$  concentrations as low as practical, using the same concentrations of each of the four dNTPs, maintaining an annealing temperature near the  $T_m$  of the primers, and programming a short extension segment and as few cycles as practical.

### Cross-Contamination

Because PCR primers are incorporated into each molecule of PCR product, amplicons are themselves suitable templates for amplification. Re-amplification of previously amplified DNA is a major problem for PCR. To reduce the likelihood of contamination of reagents, separate laboratory work areas should be designated for reagent preparation, sample preparation, reaction assembly and thermal cycling, and post-PCR analysis. Depending on the analytical rigor required and the resources available, well-isolated rooms for each step of the process will be beneficial. In the research laboratory, however, it is more common to rely on separate work areas and equipment for pre- and postamplification work and gowning of laboratory personnel in cleanroom garments. Laboratory equipment, especially pipettors and shared equipment, and laboratory personnel are the most common sources of PCR product contamination. Clinical laboratories or other analytical laboratories now employ physical separation of work areas and equipment in addition to a biochemical method using the enzyme uracil DNA glycosylase (UNG) to prevent amplicon contamination. If PCR reactions are performed with dUTP rather than dTTP, amplicons will be susceptible to digestion by UNG. This method has been called PCR sterilization. The need to avoid contamination with amplified DNA constitutes an incentive to use real-time PCR for qualitative as well as quantitative assays. With real-time PCR, tubes containing amplified DNA need never be opened. No matter what precautions are taken, negative controls containing no DNA template must always be included with each batch of PCR reactions. If low-copy detection is the assay goal, then multiple negative controls are required to adequately test for amplicon contamination.

## Reverse Transcription PCR

PCR can be used to detect and quantify RNA if RNA is first reverse-transcribed to complementary DNA (cDNA). Reverse transcriptases from Moloney murine leukemia virus (MMLV-RT) or avian myeloblastoma virus (AMV-RT) are commonly used. Either the buffer optimal for the reverse transcription (RT) or the PCR buffer is suitable in most cases, and reverse transcription is typically performed at 37–42°C for 15–60 min. Because these enzymes are heat-labile, they are inactivated during the first denaturation segment of PCR. Heat-stable reverse transcriptases enable reverse transcription at higher temperatures to denature RNA secondary structure. In the presence of  $Mn^{2+}$ , some DNA-dependent DNA polymerases such as the enzyme from *Thermus thermophilus* (*Tth* DNA polymerase) have activity on RNA templates. Using  $MnCl_2$  for reverse transcription can enable cDNA synthesis with the same enzyme used for PCR. Unless a balance between  $MnCl_2$  and  $MgCl_2$  is carefully identified,  $Mn^{2+}$  must be removed or chelated, and  $Mg^{2+}$  must be added prior to PCR. RT-PCR can be performed within intact cells (*in situ* PCR) to identify cells expressing particular genes or to assess the presence of disease-related genes. Thermal cyclers are available to automate the process on microscope slides.

## Applications

### Clinical Diagnostics

The main clinical application of PCR technology is in the diagnosis of disease, and in many cases the speed and simplicity of PCRs have revolutionized clinical diagnosis. Bacteria that are difficult or impossible to culture on artificial media can now be detected by PCR. The causative organism of syphilis, *Treponema pallidum*, can be detected by a PCR assay in which a gene encoding a specific membrane protein is amplified. PCR is being employed to investigate the pathogenicity of *Mycoplasma genitalium*. As few as four organisms can be detected by this method. PCR-based detection assays have also been developed for *Mycobacterium tuberculosis*, an organism that takes up to several weeks to identify by conventional means owing to its slow growth on artificial media. *M. tuberculosis*, the bacterium that causes tuberculosis, infects about one-third of the world's population and is most prevalent in developing countries. Recently, this organism has re-emerged as a significant pathogen in the developed world, especially among the poor and homeless, and antibiotic-resistant strains are developing. PCR tests have become essential for rapid diagnosis and epidemiology of diseases such as tuberculosis.

Amplification of viral genes forms the basis of PCR-based detection systems for viral pathogens. The human immunodeficiency virus (HIV), which is responsible for AIDS, is currently detected primarily by immunological means. However, as with most viral infections, it takes some time following infection for antibodies to develop, and there is therefore a period when a patient infected with the virus would appear negative by conventional tests. A PCR assay can be used to detect the presence of the virus in this case. Proviral DNA that has integrated into the genomes of leucocytes can be detected, or the viral RNA genome can be detected in plasma by RT-PCR. Identification of HIV by PCR methods has played an important role in AIDS diagnostics and research, as well as improved safety testing in blood banks. PCR assays have also been used in epidemiological studies and in the identification of other retroviruses such as human T-lymphotropic virus types 1 and 2 (HTLV-1 and HTLV-2). PCR procedures have also been developed to detect human papillomaviruses. This diverse group of viruses can cause a number of diseases. A PCR assay can distinguish the relatively harmless strains from those that cause serious diseases such as cervical cancer. Current laboratory methods for the diagnosis of human enteroviruses, which cause infection in children, are slow and lack sensitivity. Detection is hindered by the low level of viruses present in clinical specimens. PCR technology, however, provides a rapid and accurate diagnostic test. PCR assays have also been used to detect coronaviruses, cytomegalovirus, human herpesvirus type 6, adenovirus, Epstein-Barr virus, rotaviruses, human parvovirus, and herpes simplex viruses, among many others. In addition to detection, PCR assays can be designed to differentiate between different strains or serotypes of a particular virus.

Parasitic infections cause important human diseases such as leishmaniasis and malaria. About 40% of the world's population is at risk from malaria, and drug-resistant strains have emerged and are spreading rapidly. As well as being helpful in diagnosis of disease, a PCR assay can distinguish between drug-resistant and drug-sensitive strains, thus enabling effective treatment to be prescribed.

Many PCR diagnostic kits are now commercially available. Hoffman-La Roche has introduced kits for *Chlamydia*, *M. tuberculosis*, HIV, hepatitis C virus, and others. It is likely that an even wider range of PCR diagnostic kits will become available, and PCR testing will continue to have a huge impact in the clinical diagnostic laboratory. Despite the simplicity and rapidity of PCR from the perspective of the research laboratory, it is viewed as a complex test in the clinical laboratory. The expenses of highly

trained personnel and commercial PCR kits coupled with concerns about false positive test results due to contamination with previously amplified DNA have so far confined this promising technology to clinical research and specialty testing labs. Eventually, improved technology and assay automation will enable PCR testing to be as common in diagnostic laboratories as immunoassays.

### Molecular Genetics

In addition to being an invaluable tool for the detection and diagnosis of infectious diseases, PCR tests have also proved useful in the diagnosis and analysis of genetic diseases. Many genetic disorders are caused by specific mutations, and the fragment of DNA involved can be amplified by PCR. DNA sequencing, high-resolution size analysis on sequencing gels, single-strand conformation polymorphism analysis, Southern blotting and probe hybridization, allele-specific oligonucleotide probes, and other techniques are used to test for the presence of mutations. This approach has been used to detect mutations causing diseases such as sickle cell anemia, beta-thalassemia, cystic fibrosis, and many others. Genetic diseases caused by trinucleotide repeat expansion can be tested using PCR followed by high-resolution electrophoretic analysis on sequencing gels. A commercial test for fragile X syndrome employs novel PCR reagents and thermal cycling conditions to enable amplification of (CGG) repeats up to 800 repeat units in length in the same reaction containing a normal allele (usually 29–31 repeat units). In general, PCR should be considered a versatile template onto which can be superimposed known DNA sequences for any disorder, whether genetic or infectious in nature, to rapidly configure a high-performance test.

PCR is routinely used in the analysis of human leucocyte antigen (HLA) polymorphisms (HLA typing) to assess donor compatibility prior to bone marrow or organ transplantation. The use of PCR in transfusion medicine is also being investigated, and the basis of a system for ABO blood typing using PCR has been described.

### Prenatal Diagnostics

Prenatal diagnosis of genetic disease is normally carried out on a sample of amniotic fluid obtained by amniocentesis or by chorionic villus sampling (CVS). Conventional diagnostic tests for diseases such as cystic fibrosis can take as long as 2 weeks. Results of a PCR test can be available in 1 day. If the option to terminate a pregnancy is taken, this time savings can

be of enormous benefit. The availability of PCR technology creates an incentive for the development of noninvasive prenatal sampling techniques. Because amniocentesis and CVS carry a certain amount of risk, less invasive sampling procedures are desirable. A small number of cells of fetal origin circulate in the maternal blood. Methods to purify these cells have been described. Mutations and polymorphisms on the Y chromosome already can be tested using maternal blood samples without the need to purify fetal cells. It seems likely that in the future PCR tests will be used for prenatal diagnosis of a whole range of genetic diseases from a sample of maternal blood.

PCR testing is currently available for preimplantation genetic diagnosis for *in vitro* fertilization. One cell is removed from an eight-cell zygote for PCR testing.

### Basic Research

PCR has become an indispensable tool in the molecular biology laboratory. Many traditional methods and protocols have now been replaced by PCR-based techniques. One of the most common techniques employed by molecular biologists involves the cloning of DNA fragments into plasmid vectors. The plasmids are maintained in bacterial strains. To determine which bacterial colony contains the plasmid of interest, transformant colonies can be analyzed by PCR using primers designed from the vector (plasmid) DNA, bracketing the insert. In this way, colonies can be screened to find one containing a plasmid with an insert of a particular size, obviating the need to screen transformants by plasmid DNA preparation followed by restriction enzyme digestion. PCR can also be used to screen libraries for clones containing a particular DNA fragment or gene. This strategy dispenses with the filter hybridization step and can be both quicker and more sensitive than traditional methods. PCR is also invaluable in mutagenesis, radioactive, and nonradioactive labeling of DNA fragments, and many other applications. In addition, the technology developed for PCR amplification has been applied to the extremely important techniques of DNA sequencing and gene expression analysis.

PCR is being used in phylogenetic and evolutionary studies. PCR amplification and screening of ribosomal RNA (rRNA) genes and mitochondrial DNA have been used to estimate relationships between species and subspecies. Besides studies on the relationships of contemporary species, PCR allows studies on extinct species because it can be performed on material in museum collections. Amplification and sequencing of mitochondrial DNA are also used in evolutionary studies.

## Forensics

PCR has become an essential tool in forensic science. Many methods that exploit differences in DNA to identify individuals or to distinguish between individuals require large amounts of DNA (>100 ng). However, PCR permits the amplification of minute amounts of DNA from biological evidence found at the scene of a crime and has greatly facilitated the identification of individuals. Now, sufficient DNA for analysis can be obtained from very small amounts of biological material, such as a single hair or a small blood stain.

## Other Applications

Monitoring of foods for pathogens is increasingly performed using PCR analysis. Direct detection of microbial DNA or RNA can enable detection of contamination before visible signs appear. Assays for enterotoxigenic *E. coli* (*E. coli* 0157:H7), *Listeria monocytogenes*, *Salmonella*, and other pathogens are increasingly a component of food safety testing. Food products and ingredients can be tested for the presence of genetically modified organisms using PCR primers specific for the inserted gene sequence. Environmental samples can be tested for coliform and other bacteria or for biowarfare agents such as smallpox or anthrax. Because of its exquisite sensitivity, PCR amplification seems assured to find even more applications across research and analytical disciplines wherever there is a need to detect small amounts of genetic material.

*See also:* **Amplification Reactions. DNA Sequencing. Electrophoresis:** Nucleic Acids. **Forensic Sciences:** DNA Profiling. **Nucleic Acids:** Chromatographic and Electrophoretic Methods; Immunoassays.

## Further Reading

- Arnheim N and Erlich H (1992) Polymerase Chain Reaction Strategy. *Annual Reviews in Biochemistry* 61: 131–156.
- Bloch W (1991) A biochemical perspective of the polymerase chain reaction. *Biochemistry* 30: 2735–2747.
- Goodman MF (1995) DNA polymerase fidelity: Misinsertions and mismatched extensions. In: Innis MA, Gelfand DH, and Sninsky JJ (eds.) *PCR Strategies*, pp. 130–139. San Diego: Academic Press.
- Griffin HG and Griffin AM (1994) *PCR Technology, Current Innovations*. Boca Raton, FL: CRC Press.
- Innis MA, Gelfand DH, and Sninsky JJ (1999) *PCR Applications: Protocols for Functional Genomics*. San Diego: Academic Press.
- Innis MA, Gelfand DH, Sninsky JJ, and White TJ (1990) *PCR Protocols: A Guide to Methods and Applications*. San Diego: Academic Press.
- Markham AF (1993) The polymerase chain reaction: a tool for molecular medicine. *British Medical Journal* 306: 441–446.
- McPherson MJ, Hames BD, and Taylor GR (1995) *PCR 2: A Practical Approach*. Oxford: Oxford University Press.
- Mullis KB and Faloona FA (1987) Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods in Enzymology* 155: 335–350.
- Mullis KB, Ferré F, and Gibbs RA (1994) *The Polymerase Chain Reaction*. Boston: Birkhäuser.

# POLYMERS

Contents

**Natural Rubber**

**Synthetic**

**Polyurethanes**

## Natural Rubber

**A Krishen and M A Schafer**, The Goodyear Tire & Rubber Company, Akron, OH, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

During 2001, the world's natural rubber consumption was 7.07 million metric tons and 75% of it was

used in tires and related products. Natural rubber, which contains polyisoprene as the main constituent, is present in over 1000 plant species. However, the vast majority of the world's natural rubber production is derived from the tree *Hevea brasiliensis*. Written references to natural rubber date back to the late fifteenth century following the voyages of Columbus, although certain articles made of the rubber were used earlier in South America.

Serious investigations of rubber were first made in the mid-1700s by Charles de la Condamine and Francois Fresneau, who refer to it as caoutchouc,



## Forensics

PCR has become an essential tool in forensic science. Many methods that exploit differences in DNA to identify individuals or to distinguish between individuals require large amounts of DNA (>100 ng). However, PCR permits the amplification of minute amounts of DNA from biological evidence found at the scene of a crime and has greatly facilitated the identification of individuals. Now, sufficient DNA for analysis can be obtained from very small amounts of biological material, such as a single hair or a small blood stain.

## Other Applications

Monitoring of foods for pathogens is increasingly performed using PCR analysis. Direct detection of microbial DNA or RNA can enable detection of contamination before visible signs appear. Assays for enterotoxigenic *E. coli* (*E. coli* 0157:H7), *Listeria monocytogenes*, *Salmonella*, and other pathogens are increasingly a component of food safety testing. Food products and ingredients can be tested for the presence of genetically modified organisms using PCR primers specific for the inserted gene sequence. Environmental samples can be tested for coliform and other bacteria or for biowarfare agents such as smallpox or anthrax. Because of its exquisite sensitivity, PCR amplification seems assured to find even more applications across research and analytical disciplines wherever there is a need to detect small amounts of genetic material.

*See also:* **Amplification Reactions. DNA Sequencing. Electrophoresis:** Nucleic Acids. **Forensic Sciences:** DNA Profiling. **Nucleic Acids:** Chromatographic and Electrophoretic Methods; Immunoassays.

## Further Reading

- Arnheim N and Erlich H (1992) Polymerase Chain Reaction Strategy. *Annual Reviews in Biochemistry* 61: 131–156.
- Bloch W (1991) A biochemical perspective of the polymerase chain reaction. *Biochemistry* 30: 2735–2747.
- Goodman MF (1995) DNA polymerase fidelity: Misinsertions and mismatched extensions. In: Innis MA, Gelfand DH, and Sninsky JJ (eds.) *PCR Strategies*, pp. 130–139. San Diego: Academic Press.
- Griffin HG and Griffin AM (1994) *PCR Technology, Current Innovations*. Boca Raton, FL: CRC Press.
- Innis MA, Gelfand DH, and Sninsky JJ (1999) *PCR Applications: Protocols for Functional Genomics*. San Diego: Academic Press.
- Innis MA, Gelfand DH, Sninsky JJ, and White TJ (1990) *PCR Protocols: A Guide to Methods and Applications*. San Diego: Academic Press.
- Markham AF (1993) The polymerase chain reaction: a tool for molecular medicine. *British Medical Journal* 306: 441–446.
- McPherson MJ, Hames BD, and Taylor GR (1995) *PCR 2: A Practical Approach*. Oxford: Oxford University Press.
- Mullis KB and Faloona FA (1987) Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods in Enzymology* 155: 335–350.
- Mullis KB, Ferré F, and Gibbs RA (1994) *The Polymerase Chain Reaction*. Boston: Birkhäuser.

# POLYMERS

Contents

**Natural Rubber**

**Synthetic**

**Polyurethanes**

## Natural Rubber

**A Krishen and M A Schafer**, The Goodyear Tire & Rubber Company, Akron, OH, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

During 2001, the world's natural rubber consumption was 7.07 million metric tons and 75% of it was

used in tires and related products. Natural rubber, which contains polyisoprene as the main constituent, is present in over 1000 plant species. However, the vast majority of the world's natural rubber production is derived from the tree *Hevea brasiliensis*. Written references to natural rubber date back to the late fifteenth century following the voyages of Columbus, although certain articles made of the rubber were used earlier in South America.

Serious investigations of rubber were first made in the mid-1700s by Charles de la Condamine and Francois Fresneau, who refer to it as caoutchouc,

an American Indian word meaning weeping wood. Joseph Priestley is credited for coining the word rubber when he noticed that gum from the *Hevea* tree could be used to rub out pencil marks.

Rubber is composed of particles ranging from 0.01 to 15  $\mu\text{m}$  in size in the cytoplasm or vacuolar sap of plant cells and occurs as a milky cytoplasmic suspension that exudes from a wound or an incision made in the plant. *Hevea* contains an articulated anatomizing laticifer system, which can be drained by tapping into the phloem layer of the trunk. The rubber content in *Hevea* latex is 30–40%. *Hevea* clones in current usage have been selected for high rubber production and the replenishment of rubber following controlled tapping make this plant the most important commercial source of rubber.

Structural characterization of naturally occurring polyisoprenes sheds some light on the biosynthesis mechanism of rubber formation, the origin of outstanding properties of natural rubber, and the role of rubber in rubber trees. Nuclear magnetic resonance (NMR) analysis, based on terpenes and polyprenols as models, disclosed the structure of both terminal groups of the rubber chain. Structural evidence indicated the biosynthesis of natural rubber starts from *trans*, *trans*-farnesyl diphosphate or *trans*, *trans*, *trans*-geranylgeranyl diphosphate and terminates by dephosphorylation to form a hydroxyl terminal group. The biosynthesis of natural rubber was presumed to start from unidentified initiating species containing two *trans*-isoprene units and a peptide group and to terminate forming a phospholipid terminal group. The initiating group of natural rubber associated with proteins formed branch points, which can be decomposed by enzymatic deproteinization. The branch points formed by a phospholipid group were decomposed by transesterification with sodium methoxide. Rapid crystallization of natural rubber was explained by the presence of mixed fatty acids synergistically with linked fatty acids, which were included in phospholipid. Saturated fatty acids linked to the rubber chain-induced crystallization, while mixed unsaturated fatty acids acted as plasticizer and accelerated the crystallization rate. This was confirmed by the preparation of model *cis*-polyisoprene grafted with stearic acid. The green strength of natural rubber decreased to the same level as synthetic *cis*-polyisoprenes after transesterification, indicating the effect of branching formed by the phospholipid terminal group and fatty acids in natural rubber. The role of natural rubber in *Hevea* trees was analyzed using natural rubber from *Hevea* trees never tapped before. The formation of hard gel and oxidative degradation during the storage of natural

rubber in *Hevea* trees suggested that natural rubber acted as a radical scavenger to remove hydroperoxide.

Commercial unmilled rubber is not completely soluble in any of the common solvents – benzene, chloroform, or petroleum. The insoluble portion, called gel, contains most of the nitrogenous impurities while the solution contains molecules of rubber representing a wide range of relative molecular masses. The amount of gel, defined as residue on a filtration screen, depends on the solvent used and decreases on prolonged immersion. Fresh latex from trees that are regularly tapped has an initial low-gel content, which increases rapidly on storage. The gel phase consists of internally cross-linked insoluble particles with an average diameter of 120 nm. When dry rubber is extracted with a solvent, the uncrosslinked rubber diffuses out leaving the insoluble rubber particles held together by the residual nonrubber network.

Ribbed smoked sheets (RSS) are made from intentionally coagulated latex. In this process, strained latex is diluted with water to ~15% solids, strained again to remove particulate matter, and then coagulated by addition of an acid. The coagulum is compressed into 2–3 mm thick sheets and dried for 4–7 days in large smoke houses.

Technically specified rubber is produced from intentionally coagulated and crumbled field latex, which is converted to a blanket and dried in an air circulating drying oven at 110–125°C.

In addition to rubber, latex contains 0.3–0.7% ash, 1–2% nitrogenous compounds, ~2% resins, and 1–2% sugars and related compounds. The proteins represent most of the nitrogenous compounds in the latex and upon hydrolysis yield alanine, aspartic acid, dihydroxyphenylalanine, glutamic acid, histidine, leucine, ornithine, hydroxyproline, tyrosine, and valine. Lecithin, glycerophosphoric acid, fatty acids (palmitic, stearic, arachidic, oleic, and linoleic), sterols, resin acids, wax, and quebrachitol are some of the other nonrubber constituents. The mineral constituents present in the ash consist of potassium, magnesium with traces of copper, iron, and manganese.

## Identification and Characterization

Manufacture of rubber products requires the incorporation of fillers such as carbon black, silica, and clay, pigments, sulfur, accelerators, retarders, resins, antioxidants, antiozonants, extending oils, zinc oxide, and a variety of other elastomers. The complexity and the variety of compounding ingredients normally present in articles containing natural rubber necessitate the use of a multitude of analytical techniques depending on the information required.

## Microstructure

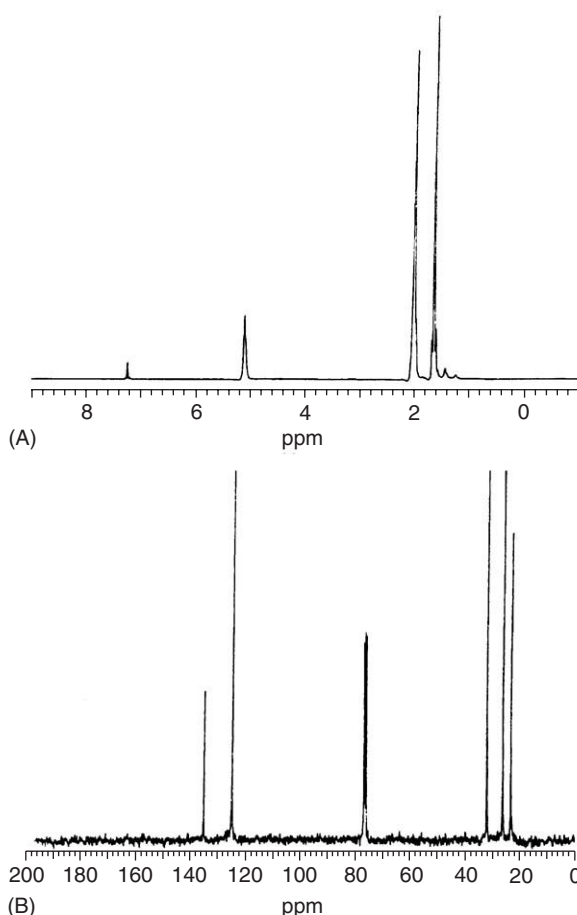
Microstructure is defined as the order and orientation in which monomer units combine together to form the polymer backbone chain. In *Hevea* rubber, the orientation of the isoprene units in the polymeric chain consists of a dimethylallyl terminal unit, three *trans*-1,4-isoprene units, a long block of *cis*-1,4-isoprene units, and a *cis*-terminal isoprenyl unit.

In comparison, four different orientations for the isomer units can be derived by synthetic polymerization of isoprene: *cis*-1,4-, *trans*-1,4-, *trans*-3,4-, and *trans*-1,2- additions. The last two addition orientations produce asymmetric carbon atoms. An identical configuration of asymmetric carbon atoms is referred to as isotactic, regular alternating is syndiotactic, and random is atactic. Another consideration is the successive addition of monomer units in head-to-head, head-to-tail, or tail-to-tail configurations, though head-to-tail is preferred due to steric hindrances. The microstructure of *Hevea* rubber is predominantly *cis*-1,4- units, eliminating the tacticity concern. Because of its biochemical origin, the addition of isoprene units is almost entirely in a head-to-tail orientation.

## NMR Spectroscopy

NMR spectroscopy and Fourier transform infrared (FTIR) spectroscopy are the main techniques used to provide microstructure information that is especially important for differentiating *Hevea* rubber from other types of naturally occurring and synthetic polyisoprene. Both proton ( $^1\text{H}$ ) and carbon ( $^{13}\text{C}$ ) NMR spectroscopy are used to obtain spectra of natural rubber in solution, and are shown in Figure 1. In the  $^1\text{H}$  NMR spectrum, the olefinic proton gives rise to a peak  $\sim 5.0$  ppm, the methylene protons  $\sim 2.0$  ppm, and the methyl protons  $\sim 1.6$  ppm.

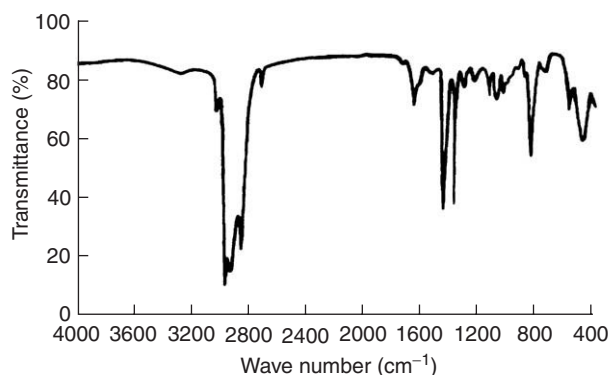
In the  $^{13}\text{C}$  NMR spectrum, the tetra-substituted olefinic carbon gives rise to the peaks from 130 to 140 ppm. The tri-substituted olefinic carbon accounts for the peaks from 118 to 128 ppm, and the aliphatic methyl carbon results in peaks from 20 to 25 ppm. This latter set of peaks can be used to distinguish *cis*-1,4-polyisoprene from *trans*-1,4- and 3,4-polyisoprenes. Either set of olefinic bands can differentiate *cis*-1,4- from 1,2-polyisoprene. Because of this,  $^{13}\text{C}$  NMR can readily be used for distinguishing other isomeric types of polyisoprene from *cis*-1,4-polyisoprenes, whether they are synthetic or natural in origin. Distinguishing a synthetic high *cis*-1,4-/low *trans*-1,4-polyisoprene from *Hevea* can present problems using  $^1\text{H}$  NMR. This problem can be overcome by using  $^{13}\text{C}$  NMR.



**Figure 1** Nuclear magnetic resonance spectra of natural rubber. (A) 300 MHz  $^1\text{H}$  NMR spectrum. Pulse delay 8.0 s; acquisition time 1.8 s; 32 transients co-added;  $50^\circ\text{C}$ ; 0.5% wt/vol in  $\text{CDCl}_3$ . (B) 75 MHz  $^{13}\text{C}$  NMR spectrum. Pulse delay 2.6 s; acquisition time 0.4 s; 2400 transients co-added; 5% wt/vol in  $\text{CDCl}_3$ ; ambient temperature. (Provided by courtesy of The Goodyear Tire & Rubber Company, Akron, OH.)

## FTIR Spectroscopy

Infrared spectroscopy is also used to provide microstructure information, though quantitative data must be obtained by comparison to standards of known composition. Typically, spectra are obtained from cast or pressed films of gum polymers. Carbon black and other fillers in compounded stock or vulcanizates can cause problems since they either absorb too much of the light energy or their characteristic spectra may be superimposed on the polymer spectra. Procedures using devulcanization, dissolution of lightly cross-linked fractions of the polymers, microtomed sections, and microdiamond anvil sampling cell with FTIR microscopy are commonly employed to overcome these problems. However, oxidation during some of these sampling procedures may affect the quantitative structural information.



**Figure 2** Infrared spectrum of natural rubber. Film cast from 1% wt/vol solution of rubber in carbon disulfide onto a KCl disk; 100 scans co-added; resolution  $2\text{ cm}^{-1}$ . (Provided by courtesy of The Goodyear Tire & Rubber Company, Akron, OH.)

An FTIR spectrum of natural rubber is shown in **Figure 2**. A number of peaks corresponding to unique molecular vibrations around the carbon-carbon double bond are observed. The CH stretch from 1,4-addition and the carbon-carbon double bond vibrational band are observed at  $3000$  and  $1665\text{ cm}^{-1}$ , respectively. In five other wavenumber regions,  $742$ ,  $764$ ,  $840$ ,  $1130$ , and  $1315\text{ cm}^{-1}$ , *cis*-1,4-polyisoprene has bands corresponding to vibrations of the  $-\text{C}(\text{CH}_3)=\text{CH}-$  group. In each case, *trans*-1,4-polyisoprene units have similar bands ranging from 3 to 20 wavenumbers higher, shifted toward shorter wavelengths. Resolution of the *cis* and *trans* bands is difficult, even with FTIR, in situations where low levels of *trans*-1,4-polyisoprene are blended with *Hevea*, or with synthetic 1,4-polyisoprenes containing a small amount of *trans*.

Infrared spectra of products obtained by heating the polymer in a test tube or in special heating equipment are also used to identify polymers in vulcanizates. The pyrolysis temperature is a critical parameter since the same polymer can give different spectra when pyrolyzed at different temperatures. Comparison of pyrolysis products obtained by heating carbon black containing vulcanized compounds can be used to distinguish *Hevea* from synthetic polyisoprene.

## Micromorphology

Performance of rubber compounds is affected by micromorphology, such as polymer compatibility and carbon black distribution. Transmission electron microscopy has been commonly used to examine micromorphology. More recently, atomic force microscopy, also referred to as scanning probe microscopy, has been used for this purpose. Tapping mode atomic force microscopy can be used to distinguish two different polymer domains in unfilled rubber

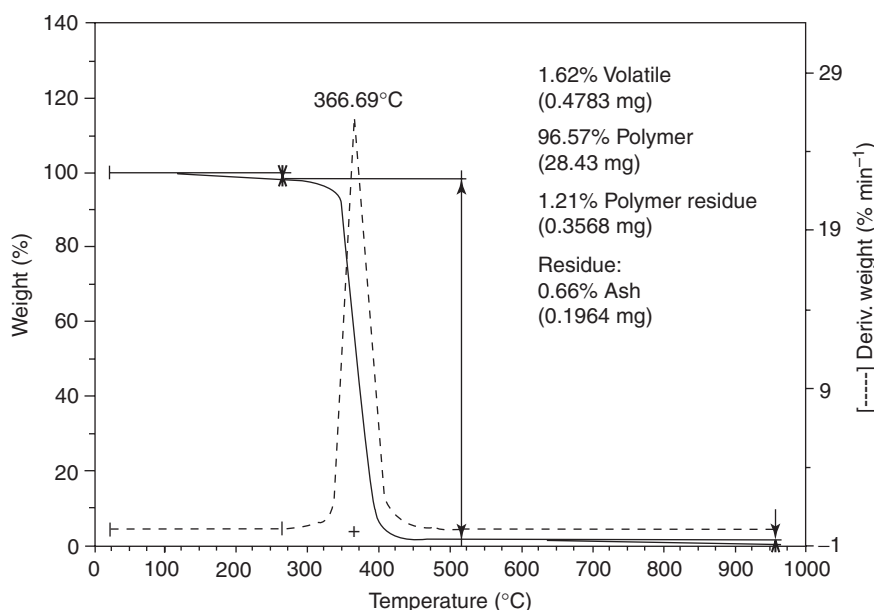
blends. Filler morphology was also examined in the filled natural rubber and filled rubber blend compounds. Silica and carbon black showed different behavior in the rubber blend. Carbon black lies predominantly in the polybutadiene rubber domain whereas silica exists in the natural rubber domain.

## Analysis of Nonrubber Constituents

Several nonrubber constituents such as  $\beta$ -sitosterol, fatty acids, and proteins, all naturally occurring in *Hevea*, are of importance in analysis. Natural rubber can be differentiated from a synthetic high *cis*-1,4-polyisoprene, as long as it exceeds 10% by weight in a mixture, through detection of  $\beta$ -sitosterol in the extract of either the rubber or vulcanized products. Typically,  $\beta$ -sitosterol is separated from other extracted components using thin-layer chromatography (TLC) followed by cupric acetate spray reagent for visualization. Gas chromatographic analysis of the silyl derivative can also be employed. Quantification of natural rubber based on the  $\beta$ -sitosterol content is not possible as  $\beta$ -sitosterol content is not consistent from batch to batch.

Fatty acids are determined after extraction using either gas chromatography (GC) or high-pressure liquid chromatography (HPLC). With either technique, the fatty acids can be measured either directly or after derivatization with methylating, silanizing, or ultraviolet (UV) chromophore-generating reagents. This approach can also be used to determine fatty and rosin acids added as compounding ingredients in cured samples.

The protein content of products intended for medical and health-related markets have been of special concern in recent years due to the observation of human allergic and anaphylactic reactions. Three colorimetric methods, modified Lowry, Bradford, and Pierce BCA reagent, are currently used to determine aqueous extractable proteins, though all three lack specificity to some degree. Electrophoresis, size-exclusion chromatography, amino acid analysis, and enzyme-linked immunosorbent assays (ELISAs) have been used to improve specificity, but are time consuming, expensive, and require an experienced analyst. The American Society for Testing and Materials (ASTM) has established two methods for determination of proteins, D5712 based on the modified Lowry colorimetric test and D6499 based on ELISA. Determination of the nitrogen content of *Hevea* can serve as another measure of its protein content, although this is not accepted by regulatory agencies. Standard test methods for nitrogen, ASTM D3533 and International Organization for Standardization (ISO 1656), use a factor to convert from nitrogen to protein.



**Figure 3** Thermogravimetric analysis of natural rubber (RSS): 6 mg heated in vacuum at  $10^{\circ}\text{C min}^{-1}$  to  $485^{\circ}\text{C}$ ; in nitrogen to  $570^{\circ}\text{C}$ , and in air at  $22^{\circ}\text{C min}^{-1}$  to  $990^{\circ}\text{C}$ . (Provided by courtesy of The Goodyear Tire & Rubber Company, Akron, OH.)

## Quantitative Compositional Analysis

Thermogravimetric analysis (TGA) provides a rapid scanning technique to determine the ratio of the major components in raw or cured stocks. Using a 10–50 mg sample with appropriate heating parameters, four fractions are observed (Figure 3):

1. Volatile matter representing moisture, oils, plasticizers, polymer-diluents, curatives, antioxidants, antiozonants, and other components with boiling points below  $300^{\circ}\text{C}$ .
2. Polymeric components representing polymers, fibers, processing oils, and resins that degrade between  $300^{\circ}\text{C}$  and  $750^{\circ}\text{C}$ .
3. Oxidizable materials representing carbon black, graphite, and carbonaceous residues from some polymers.
4. Inorganic ash representing metallic oxides, fillers, zinc oxide, and residues from elastomers.

The first two fractions are produced either in a nitrogen atmosphere or under vacuum at 70 Pa (0.5 mm Hg). Accurate quantification requires appropriate corrections to allow for the overlap of the process oil and resins with the polymer decomposition region along with the serious complication observed for elastomers with a heteroatom in the monomer unit like acrylonitrile-butadiene, chloroprene rubber, chlorobutyl rubber, polyvinyl chloride (PVC), chlorosulfonated polyethylene, polyacrylates, fluoroelastomers, epichlorohydrin rubber, etc. Most of these polymers produce volatile components below  $300^{\circ}\text{C}$  and leave a carbonaceous residue after degradation, which oxidizes with the carbon black

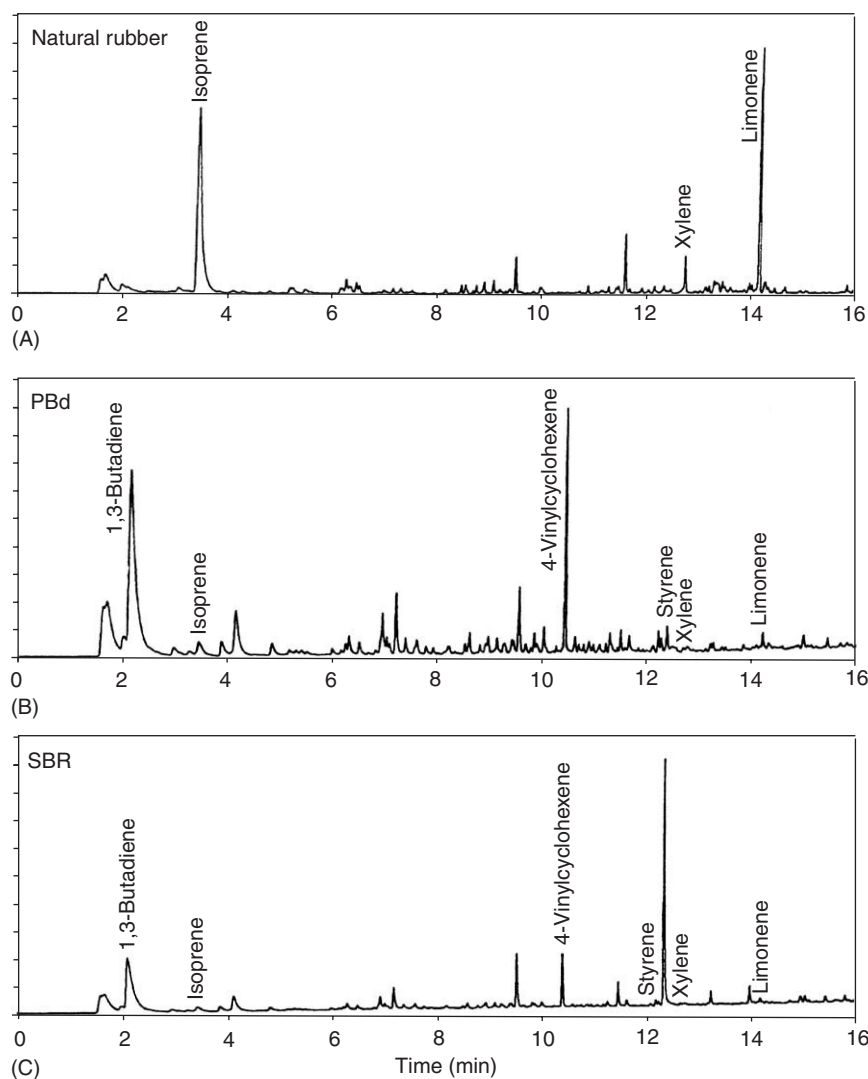
fraction. With the increasing use of short fibers to reinforce vulcanizates, their degradation also complicates interpretation and quantification.

Polymer ratio calculations based on the peak areas even for a mixture of polymers, which are well resolved, should be based on a calibration curve obtained from compounds of known composition. Modified heating programs and automated high-resolution instrumentation, which adjusts the heating rate when the start of a peak is sensed, are helpful in providing better discrimination in some cases.

Pyrolysis in an inert atmosphere under precisely controlled conditions (Figure 4) generates duplicable amounts of products, which are separated by capillary GC and provide an estimate of the ratio of polymeric constituents. Natural rubber produces isoprene and limonene as two of the characteristic products, which distinguish it from polybutadiene (BR), styrene-butadiene co-polymer (SBR), butyl rubber (IIR), and some of the other polymers. Quantification involving a mixture of polymers requires calibration curves derived from similar combinations of polymers (Figure 5). Cured and uncured formulations require separate calibrations and the differences in the microstructure of a polymer affect the products obtained on pyrolysis.

## Analysis for Additives in Vulcanized Rubber

As previously described, TGA can provide a wealth of information about the composition of rubber



**Figure 4** Pyrolysis/capillary gas chromatograms of (A) natural rubber, (B) BR, and (C) SBR; 80  $\mu\text{g}$  pyrolyzed at 770°C for 10 s and chromatographed on 30 m  $\times$  0.32 mm DB-1701 (J & W Scientific Company) capillary column temperature programmed from  $-20$  to 250°C at 10°C min $^{-1}$ . (Provided by courtesy of The Goodyear Tire & Rubber Company, Akron, OH.)

samples. Additional analyses of a cured product are occasionally required in cases addressing product integrity or to detect errors in compounding.

### Extraction

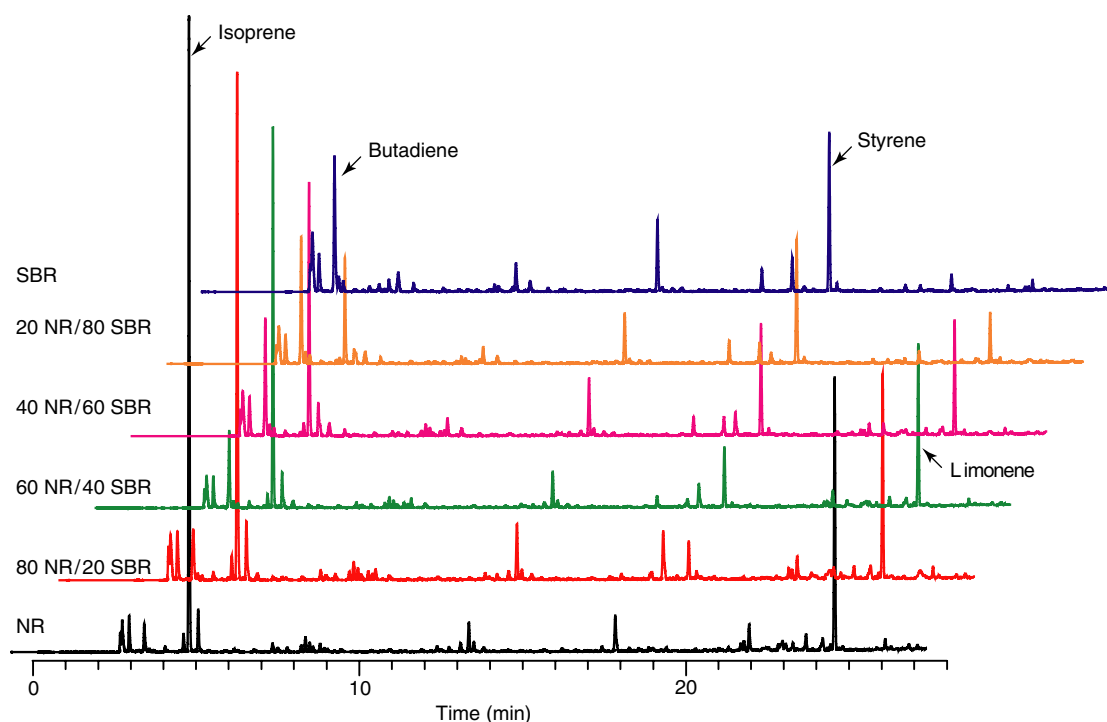
The first step involves sample preparation and extraction of additives. Preparation of samples for chemical analysis is listed in ASTM D3183 and ISO 4661/1.2. Acetone and methanol are common solvents used for Soxhlet extraction following procedures such as ASTM D297 and ISO 1407. Soxhlet extractions can often require up to 16 h for completion. Thus, rapid solvent extraction techniques using ultrasonic baths or homogenizers, microwave heating, or pressurized extractions such as supercritical fluids are finding ever-increasing use. Extractions solely utilizing heat have also been reported. The extract can contain

rubber additives such as antioxidants, antiozonants, accelerators (uncured samples), or accelerator fragments (cured), resins, fatty and rosin acids, and oils.

### Antioxidants and Antiozonants

Antioxidant and antiozonant types most commonly used are aromatic amines or phenolics, though others are also employed, and can be determined using a variety of techniques such as UV-visible spectrophotometry, FTIR, near-infrared spectroscopy, TLC, GC (if the material can be volatilized), supercritical fluid chromatography, and HPLC. Identification of unknown antioxidants requires a separation technique like chromatography followed by mass spectrometry, NMR, FTIR, X-ray crystallography, etc. Standardized TLC methods are given in ASTM D3156 and





**Figure 5** Pyrolysis/capillary gas chromatograms of natural rubber/SBR blends. (Provided by courtesy of The Goodyear Tire & Rubber Company, Akron, OH.)

ISO 4645/2, whereas a standardized HPLC method is listed in ISO 11089.

### Accelerators

As the name implies, accelerators are used to increase the rate of sulfur vulcanization reaction in polymers containing unsaturated backbones. Many accelerators contain sulfur that chemically reacts during vulcanization, resulting in decomposition of the original accelerator. Accelerator types such as sulfenamides, thiurams, and mercaptobenzothiazyl disulfide (MBTS) are typically not detectable as the intact component in cured rubber articles, but can usually be measured unchanged in the raw compounded stock. Upon vulcanization, mercaptobenzothiazole (MBT) is produced from MBTS and sulfenamides, the latter also giving an amine, while thiurams react to produce the corresponding dithiocarbamates, or in some cases carbon disulfide and the corresponding amine. Guanidine and dithiocarbamate type accelerators can be detected as the intact component, though the corresponding aromatic amine and secondary aliphatic amine, respectively, are also produced in lesser quantities. Occasionally, more than one accelerator is used in compounds and this may lead to difficulties in identification of the exact combination of accelerators added.

Unlike vulcanized samples, uncured compounds must be extracted with a cold solvent to prevent

onset of the vulcanization reaction. Methanol and isopropanol are the preferred extraction solvents for uncured compounds, which might contain thiurams since they can decompose even upon extraction with acetone. Techniques that have been employed for detection of accelerators from vulcanized samples include colorimetry, TLC, GC, and HPLC. GC is primarily used to determine MBT and amine fragments, whereas HPLC is useful for determining guanidines and MBT. The same techniques, as well as FTIR, have proven useful for analysis of uncured compounds. Here, HPLC is advantageous for the determination of sulfenamide, MBTS, and guanidine accelerators. ISO 11389 outlines the determination of accelerators using GC and TLC.

N-nitrosamines generated at low levels from secondary amines liberated from some of the accelerators can cause health concerns and have been subject to regulation. Volatile N-nitrosamines are quantified using GC with a thermal energy analyzer.

### Sulfur Cross-linking

Although relative amounts of crosslinking are estimated by physical properties, a rough estimate of the degree of curing is obtained by quantitating free sulfur (ISO 7269), extractable low relative mass polymeric species, or by determining the curing isotherm using differential scanning calorimetry.

The type and distribution of sulfidic links – mono, di, or poly – affects the properties of the vulcanized rubber and these are determined by scission with specific chemical reagents followed by determining the degree of swelling in a suitable solvent. NMR methods have also been developed to determine cross-linking.

### Extending Oils

Hydrocarbon oils are used as plasticizers or extenders in unsaturated polymers like *Hevea*. Oils often represent a significant amount of the rubber compound, up to 40 phr (parts per hundred of rubber) in natural/synthetic blends, but at lower levels when natural rubber is the only polymer. Oils are grouped into three types, paraffinic, naphthenic, and aromatic, but each type varies only in the proportions of these three different types. Paraffinic and naphthenic oils are most commonly used in compounds containing natural rubber. TLC provides a convenient means of identifying oil type by visualizing the separated spots under short (254 nm) and long (366 nm) wavelength UV lamps.

Aromatic oils have the lowest mobility and fluoresce at the long wavelength, while at the short wavelength produce a dark spot. Naphthenic oils have an intermediate mobility and are only observed as a dark spot under the short wavelength light, whereas paraffinic oils have the highest mobility but can only be observed as fleeting spots as the TLC solvent dries. GC, HPLC, and size-exclusion chromatography with photodiode-array and differential refractive index detectors have been applied to extender oil determination. Infrared spectroscopy can also be used to identify and quantify oil type based on absorption at multiple bands.

### Metals

The determination of several metals is important to assuring proper compounding and minimal metal contamination known to be detrimental to vulcanization. Procedures for ashing samples are given in ASTM D297 and ISO 247. Manganese, known to inhibit rubber curing, is determined by photometric methods, ASTM D1278 and ISO 7780. Zinc, added as zinc oxide

to the compound, is determined using a titrimetric method involving chelation with ethylenediaminetetraacetic acid as in ASTM D297 and ISO 2454. Methods using atomic absorption spectrometry (AAS) have also been standardized for manganese and zinc as ASTM D4004 and ISO 6101. Other metals such as silicon, iron, and copper are also quantified using AAS.

Energy dispersive (ED) and wavelength dispersive (WDX) X-ray fluorescence analysis are both capable of identifying a large variety of elements. Quantification can suffer from matrix effects and from the fact that these are near-surface techniques. The quantity of residual ash generated during TGA is sufficient for identification of metals using a scanning electron microscope equipped with EDS.

**See also:** **Gas Chromatography:** Overview; **Pyrolysis.** **Infrared Spectroscopy:** Overview; **Industrial Applications.** **Lipids:** Fatty Acids. **Nuclear Magnetic Resonance Spectroscopy-Applicable Elements:** Hydrogen Isotopes; Carbon-13. **Polymers:** Synthetic.

### Further Reading

- American Society for Testing and Materials (2003) *Annual Book of ASTM Standards 09.01*. Philadelphia, PA: ASTM.
- Brown R (1999) *Handbook of Polymer Testing*. London: Dekker.
- Forrest MJ (2002) Rubber analysis – polymers, compounds and products. RAPRA Review Reports.
- International Organization for Standardization (1992) *International Standards*. Geneva: ISO.
- Loadman MJR (1998) *Analysis of Rubber and Rubber-like Polymers*. Dordrecht: Kluwer Academic.
- St. Cyr DR (1988) *Encyclopedia of Polymer Science and Engineering*. New York: Wiley.
- Tanaka Y (2001) Structural characterization of natural polyisoprenes: Solve the mystery of natural rubber based on structural study. *Rubber Chemistry and Technology* 74: 355–375.
- Turi E (1997) *Thermal Characterization of Polymeric Materials*. New York: Academic Press.

## Synthetic

**S Podzimek**, SYNPO, Pardubice, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

A polymer is a substance composed of macromolecules, i.e., molecules built of a large number of small

molecules linked together by covalent bonds. Man-made polymers (synthetic polymers) are relatively new materials that did not exist until about hundred years ago. The first synthetic polymer, phenol-formaldehyde resin (Bakelite), appeared shortly before World War I. Nowadays, synthetic polymers are used in a variety of applications covering almost all areas

The type and distribution of sulfidic links – mono, di, or poly – affects the properties of the vulcanized rubber and these are determined by scission with specific chemical reagents followed by determining the degree of swelling in a suitable solvent. NMR methods have also been developed to determine cross-linking.

### Extending Oils

Hydrocarbon oils are used as plasticizers or extenders in unsaturated polymers like *Hevea*. Oils often represent a significant amount of the rubber compound, up to 40 phr (parts per hundred of rubber) in natural/synthetic blends, but at lower levels when natural rubber is the only polymer. Oils are grouped into three types, paraffinic, naphthenic, and aromatic, but each type varies only in the proportions of these three different types. Paraffinic and naphthenic oils are most commonly used in compounds containing natural rubber. TLC provides a convenient means of identifying oil type by visualizing the separated spots under short (254 nm) and long (366 nm) wavelength UV lamps.

Aromatic oils have the lowest mobility and fluoresce at the long wavelength, while at the short wavelength produce a dark spot. Naphthenic oils have an intermediate mobility and are only observed as a dark spot under the short wavelength light, whereas paraffinic oils have the highest mobility but can only be observed as fleeting spots as the TLC solvent dries. GC, HPLC, and size-exclusion chromatography with photodiode-array and differential refractive index detectors have been applied to extender oil determination. Infrared spectroscopy can also be used to identify and quantify oil type based on absorption at multiple bands.

### Metals

The determination of several metals is important to assuring proper compounding and minimal metal contamination known to be detrimental to vulcanization. Procedures for ashing samples are given in ASTM D297 and ISO 247. Manganese, known to inhibit rubber curing, is determined by photometric methods, ASTM D1278 and ISO 7780. Zinc, added as zinc oxide

to the compound, is determined using a titrimetric method involving chelation with ethylenediaminetetraacetic acid as in ASTM D297 and ISO 2454. Methods using atomic absorption spectrometry (AAS) have also been standardized for manganese and zinc as ASTM D4004 and ISO 6101. Other metals such as silicon, iron, and copper are also quantified using AAS.

Energy dispersive (ED) and wavelength dispersive (WDX) X-ray fluorescence analysis are both capable of identifying a large variety of elements. Quantification can suffer from matrix effects and from the fact that these are near-surface techniques. The quantity of residual ash generated during TGA is sufficient for identification of metals using a scanning electron microscope equipped with EDS.

**See also:** **Gas Chromatography:** Overview; **Pyrolysis.** **Infrared Spectroscopy:** Overview; **Industrial Applications.** **Lipids:** Fatty Acids. **Nuclear Magnetic Resonance Spectroscopy-Applicable Elements:** Hydrogen Isotopes; Carbon-13. **Polymers:** Synthetic.

### Further Reading

- American Society for Testing and Materials (2003) *Annual Book of ASTM Standards 09.01*. Philadelphia, PA: ASTM.
- Brown R (1999) *Handbook of Polymer Testing*. London: Dekker.
- Forrest MJ (2002) Rubber analysis – polymers, compounds and products. RAPRA Review Reports.
- International Organization for Standardization (1992) *International Standards*. Geneva: ISO.
- Loadman MJR (1998) *Analysis of Rubber and Rubber-like Polymers*. Dordrecht: Kluwer Academic.
- St. Cyr DR (1988) *Encyclopedia of Polymer Science and Engineering*. New York: Wiley.
- Tanaka Y (2001) Structural characterization of natural polyisoprenes: Solve the mystery of natural rubber based on structural study. *Rubber Chemistry and Technology* 74: 355–375.
- Turi E (1997) *Thermal Characterization of Polymeric Materials*. New York: Academic Press.

## Synthetic

**S Podzimek**, SYNPO, Pardubice, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

A polymer is a substance composed of macromolecules, i.e., molecules built of a large number of small

molecules linked together by covalent bonds. Man-made polymers (synthetic polymers) are relatively new materials that did not exist until about hundred years ago. The first synthetic polymer, phenol-formaldehyde resin (Bakelite), appeared shortly before World War I. Nowadays, synthetic polymers are used in a variety of applications covering almost all areas

of human life. They can replace conventional materials with properties that are better in many aspects when compared to classical materials. Low density, possibility to adjust properties to required applications, resistance to moisture and chemicals, and easy workability are the main advantages of synthetic polymers.

Synthetic polymers can be classified as thermoplasts, which soften under heat and can be reversibly melted, and thermosets, which by the action of heat or chemical substance undergo chemical reaction and form insoluble materials that cannot be melted. Mixtures of molecules of relatively low molar mass (hundreds to thousands) that are able to react mutually or with other compounds and form cross-linked materials are often called synthetic resins. That is, synthetic resins such as epoxy or polyester or phenol-formaldehyde resins are thermosets. The term oligomer refers to a polymer molecule with relatively low molar mass (roughly below  $10\,000\text{ g mol}^{-1}$ ) whose properties vary significantly with the removal of one or a few of the units.

## Structure of Synthetic Polymers

The part of a macromolecule from which the macromolecule is built is called monomer unit, while the smallest part of a macromolecule that repeats periodically is called structural repeating unit. Polymers can consist of one or more kinds of monomer unit. The former are called homopolymers and the latter copolymers. Synthetic polymers are usually varied mixtures of molecules of different molar mass ( $M$ ) and possibly chemical composition and/or molecular architecture; that is, they are nonuniform

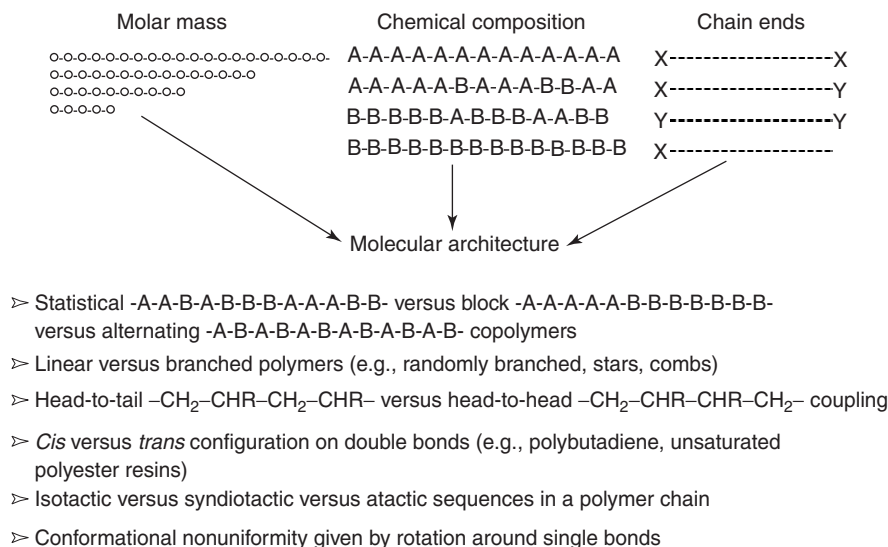
(polydisperse) materials. Examples of various non-uniformities are outlined in **Figure 1**. Polymers can be nonuniform in one or more properties. It is worth mentioning that monodisperse polymers (i.e., uniform with respect to all properties) are exceptional in the area of synthetic polymers.

It is essential for the polymer chemist and analyst to be aware of all possible nonuniformities of polymers in order to interpret the experimental data and understand the properties and behavior of polymers. Two polymer samples may be identical in one or more properties but may differ in others. Although the polymer properties are generally distributed, analysis can provide solely average values. Two polymer samples can be identical in an average property but the property distributions can be different. However, average properties are often used instead of distributions in order to simplify the description of a polymer sample or because the distribution cannot be determined due to the time or instrumental limitations.

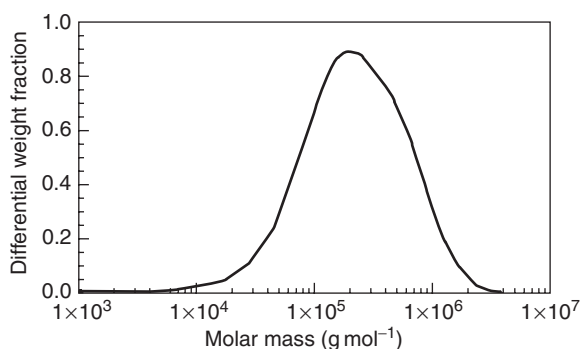
In addition to molecular nonuniformity, many commercially important polymer-based materials are polymer blends, i.e., mixtures of two or more polymeric components.

## Molar Mass Distribution

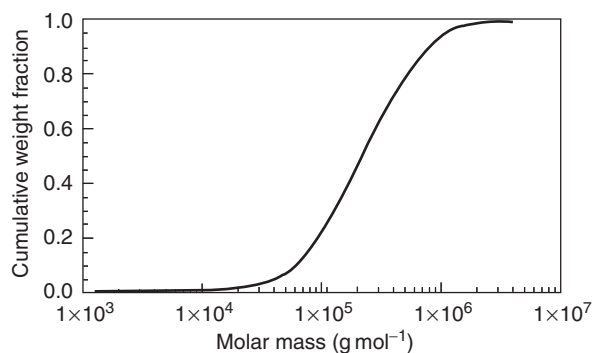
Several notations are used in polymer science to express molar mass, namely molar mass or molecular mass (mass of polymer divided by the amount of polymer, dimension  $\text{g mol}^{-1}$ ), and relative molecular mass or molecular weight (mass of the polymer relative to 1/12 of the mass of the  $^{12}\text{C}$  atom, dimensionless). The number of monomeric units in



**Figure 1** Schematic description of possible nonuniformities in synthetic polymers.



**Figure 2** Differential molar mass distribution curve of polydisperse polystyrene. Averages by SEC-MALS:  $M_n = 100\,000\text{ g mol}^{-1}$ ;  $M_w = 334\,000\text{ g mol}^{-1}$ ;  $M_z = 704\,000\text{ g mol}^{-1}$ ;  $M_v$  (by viscometry) =  $307\,000\text{ g mol}^{-1}$ . Zimm plot for this sample is shown in **Figure 4**.



**Figure 3** Cumulative distribution curve corresponding to the differential distribution shown in **Figure 2**.

a macromolecule or oligomer molecule is called degree of polymerization ( $P$ ).

In contrast to low molar mass compounds a polymer does not have a single molar mass and almost all synthetic polymers are polydisperse in molar mass. Frequencies of particular molar masses in a given polymer sample are described by the molar mass distribution (MMD). The differential  $f(M)$  and cumulative  $I(M)$  MMD curves are used for the description of MMD. An example of the differential distribution curve is shown in **Figure 2**. Note: It is usual practice to express the molar mass axis in logarithmic scale. The area  $f(M) dM$  in the graph represents the weight fraction  $w$  of material having the molar mass between  $M$  and  $M + dM$ . According to the graph in **Figure 2**, the value  $f(M)$  of  $\sim 0.84$  corresponds to the molar mass of  $300\,000\text{ g mol}^{-1}$ . Taking a molar mass of  $310\,000\text{ g mol}^{-1}$  as  $M + dM$  the differential  $d(\log M) = 0.014$ . Then,  $0.84 \times 0.014 = 0.012$  is the weight fraction of a polymer having the molar mass in the range of  $300\,000\text{--}310\,000\text{ g mol}^{-1}$ . Additional information obtainable from **Figure 2** is the entire molar mass range ( $\sim 1000\text{--}3\,000\,000\text{ g mol}^{-1}$ ) and the molar mass of the most abundant fractions ( $\sim 200\,000\text{ g mol}^{-1}$ ).

Another way of graphical representation of MMD is the cumulative MMD as shown in **Figure 3**. The cumulative MMD is defined as:

$$I(M) = \int_0^M f(M) dM \quad [1]$$

The cumulative MMD allows easy determination of the weight fraction of a polymer below or above a particular molar mass or the weight fraction of polymer in a specific molar mass range. Using the curve in **Figure 3**, one can easily see that  $\sim 23\%$  molecules have

molar mass below  $100\,000\text{ g mol}^{-1}$  (77% above this limit) or 95% molecules have molar mass above  $1\,000\,000\text{ g mol}^{-1}$  or 72% molecules have molar mass in the range of  $100\,000\text{--}1\,000\,000\text{ g mol}^{-1}$ .

Graphical overlay of distribution curves can serve for comparison of a set of samples and finding even subtle differences among them. Note: Instead using the weight fraction it is possible to use the mole fraction. The frequency functions are then called the number distribution functions and subscripts 'w' and 'n' can be used in order to differentiate between the weight distribution and number distribution, respectively.

The important properties of the MMD curve are the number of peaks and whether a peak is symmetrical or skewed. They can provide information about the polymer composition or polymerization process. Two or more peaks indicate a blend of two or more polymers with different MMDs. Or it may indicate the presence of a single polymer prepared under more than one set of polymerization conditions (temperature, concentration of monomers, or initiator). Shoulders at the high molar mass or low molar mass regions of the MMD are signs of appreciable amounts of species with very high or low molar masses. It is important to note that these species can significantly affect the properties of the polymer. A high molar mass tail on the distribution curve often indicates the presence of branched molecules.

Several average molar masses are used in polymer practice. Different instrumental techniques provide different kinds of averages. They are defined as follows:

Number average:

$$\begin{aligned} M_n &= \int_0^\infty M f_n(M) dM = \sum_i x_i M_i \\ &= \frac{\sum_i n_i M_i}{\sum_i n_i} = \frac{\sum_i m_i}{\sum_i m_i / M_i} = \frac{1}{\sum_i w_i / M_i} \end{aligned} \quad [2]$$

Weight average:

$$M_w = \int_0^\infty M f_w(M) dM = \sum_i w_i M_i$$

$$= \frac{\sum_i m_i M_i}{\sum_i m_i} = \frac{\sum_i n_i M_i^2}{\sum_i n_i M_i} \quad [3]$$

Z-average:

$$M_z = \frac{\int_0^\infty M^2 f_w(M) dM}{\int_0^\infty M f_w(M) dM} = \frac{\sum_i m_i M_i^3}{\sum_i m_i M_i} = \frac{\sum_i n_i M_i^4}{\sum_i n_i M_i^2} \quad [4]$$

Viscosity average:

$$M_v = \left[ \int_0^\infty M^a f_w(M) dM \right]^{1/a} = \left[ \sum_i w_i M_i^a \right]^{1/a} \quad [5]$$

In the above equations,  $n_i$  is the number of moles,  $m_i$  is the mass,  $x_i$  is the mole fraction, and  $w_i$  is the weight fraction of molecules with molar mass  $M_i$ . The exponent  $a$  is the exponent of the Mark-Houwink equation. In thermodynamically good solvents, the values of  $a$  mostly range from  $\sim 0.7$  to  $0.8$ . If the exponent  $a$  becomes 1, the viscosity average becomes identical to the weight average. The  $M_v$  value lies between  $M_n$  and  $M_w$ , closer to  $M_w$ .

The different averages are differently sensitive to different molar masses. Namely,  $M_n$  is sensitive mainly to the fractions with low molar masses while  $M_w$  and particularly  $M_z$  are sensitive to high molar mass fractions. For monodisperse polymers, all molar mass averages are identical. The order of molar mass averages for polydisperse polymers is  $M_n < M_v < M_w < M_z$ . The ratio of  $M_w/M_n$  is a measure of the broadness of the MMD and is often called polydispersity or polydispersity index. The ratio  $M_z/M_w$  can be used as an additional parameter or can be alternatively applied instead of  $M_w/M_n$  if the  $M_n$  value cannot be reliably determined.

Molar mass averages and polydispersity can be calculated theoretically for various polymerization reactions. In the case of polycondensation of monomer type a-R-b (e.g., hydroxy acid) or the equimolar mixture of monomers a-R<sub>1</sub>-a and b-R<sub>2</sub>-b the average polymerization degrees are as follow:

$$P_n = \frac{1}{1-q} \quad [6]$$

$$P_w = \frac{1+q}{1-q} \quad [7]$$

$$P_w/P_n = 1+q \quad [8]$$

where  $q$  is the conversion. In the case of polycondensation of monomers with average functionality larger than 2, the average polymerization degrees

are:

$$P_n = \frac{1}{1-(qf/2)} \quad [9]$$

$$P_w = \frac{1+q}{1-q(f-1)} \quad [10]$$

where  $f$  is an average monomer functionality (e.g., equimolar mixture of adipic acid and glycerol has  $f=2.5$ ).

The polydispersities of polymers prepared by free radical polymerization are 2 or 1.5 for the termination by disproportionation or combination, respectively. These values apply to products formed during a short time period when the concentrations of monomer and initiator are constant. In the course of free radical polymerization the monomer and initiator concentrations usually vary with conversion and consequently the polydispersity of the final product is broader.

The discrepancy between the experimental and theoretical values of molar mass averages and polydispersity can reveal side reactions. For instance, branching, which can arise as a result of chain transfer to polymer or due to the presence of polyfunctional impurities, markedly increases polydispersity.

Many important polymer properties depend on molar mass, e.g., melt and solution viscosity, tensile strength, toughness, impact strength, adhesive strength, elasticity, brittleness, abrasion resistance, flex life, softening temperature, solubility, chemical resistance, cure time, diffusion coefficient, film and fiber forming ability, ability to be fabricated, and processing temperature. Different molar mass averages can influence different polymer properties since either high molar mass or low molar mass species can primarily influence specific properties. The tensile strength is particularly related to  $M_w$  since it is most influenced by the large molecules in the material. The flex life is related to  $M_z$  because extremely large molecules are most important for this property. The glass transition temperature of oligomers increases with increasing  $M_n$  since the decreasing number of chain ends decreases the chain mobility and free volume. There is an empirical relationship of polymer melt viscosity having a 3.4 power dependency on  $M_w$ . Polymer rheology is strongly related to  $M_z$  reflecting the high molar mass tail of MMD.

It is important to note that there are no commonly good molar mass averages or MMD for a polymer sample. The optimum values depend on the nature of polymer and especially on the required end-use properties. A polymer sample known as good can serve as a reference to which other samples are compared.



## Analysis and Characterization of Synthetic Polymers

There are numerous techniques employed for the analysis and characterization of synthetic polymers. A number of them are used in other types of chemistry (e.g., nuclear magnetic resonance, infrared spectroscopy, liquid chromatography, gas chromatography, mass spectrometry) while some are applied mostly in the area of polymers (e.g., osmometry, light scattering, size-exclusion chromatography (SEC), field flow fractionation (FFF)). The focus of this article is in the methods used for the determination of average molar mass and MMD, because molar mass is a fundamental difference between the polymers and other organic compounds. In addition, it is essentially related to polymer properties and bears information about polymerization process and polymer degradation.

## Methods for the Determination of Molar Mass and Molar Mass Distribution

Molar mass of polymers can be determined by various methods (as summarized in Table 1). The following provides principal equations and points out advantages and limitations of particular methods.

### Method of End Groups

The results obtained by the method of end groups may be influenced by the presence of chain defects (overestimation of  $M_n$ ) or branching (underestimation of  $M_n$ ). The method is applicable solely to polymers with easily determinable end groups.

### Osmometry

Vapor pressure osmometry (VPO) calculates  $M_n$  using the equation:

$$\frac{\Delta T}{c} = B \left( \frac{1}{M_n} + A_2 c + \dots \right) \quad [11]$$

where  $\Delta T$  is the temperature difference of the solution and solvent,  $B$  is a calibration constant of the instrument obtained by measuring a compound of known molar mass,  $c$  is the concentration, and  $A_2$  is the second virial coefficient. The measurements are made at multiple concentrations and  $M_n$  is obtained by the extrapolation to zero. Serious disadvantage of VPO is excessive sensitivity to the presence of low molar mass compounds, such as residual monomers, solvents, or moisture, which can result in serious underestimation of  $M_n$ .

Membrane osmometry (MO) is another absolute method for the determination of  $M_n$ . The osmotic pressure  $\pi$  is related to the molar mass by the equation:

$$\frac{\pi}{c} = RT \left( \frac{1}{M_n} + A_2 c + \dots \right) \quad [12]$$

The extrapolation of the concentration dependency to zero provides  $M_n$ . If the sample contains oligomeric species that can permeate through the membrane the osmotic pressure is too low and the obtained  $M_n$  is overestimated. Consequently, polydisperse polymers with low molar mass tails are not suitable for the method. The advantage of VPO and MO is independency of chemical nonuniformity.

The second virial coefficient is an additional result that provides information about the polymer-solvent interactions and thus about thermodynamic quality of solvent that increases with increasing value of  $A_2$ . The  $A_2$  decreases with increasing molar mass, which explains decreasing solubility of polymers with increasing molar mass.

### Viscometry

Viscometric measurements are performed using a capillary viscometer (Ubbelohde, Ostwald). The measured quantity is the specific viscosity ( $\eta_{sp}$ ):

$$\eta_{sp} = \frac{\eta - \eta_0}{\eta_0} = \frac{t - t_0}{t_0} \quad [13]$$

where  $\eta$  and  $\eta_0$  are the viscosities of dilute polymer solution and pure solvent, respectively. In practice, the measurement of viscosity is replaced with the measurement of time ( $t$ ) needed for the solvent or solution to flow from one mark to the other. The concentration dependence of the specific viscosity can be described by Huggins equation:

$$\frac{\eta_{sp}}{c} = [\eta] + k_H [\eta]^2 c + \dots \quad [14]$$

where  $c$  is the concentration,  $k_H$  is the Huggins coefficient for given polymer, solvent, and temperature, and  $[\eta]$  is the intrinsic viscosity that is determined from the concentration dependency of the specific viscosity. The relation between the intrinsic viscosity and molar mass is given by the Mark-Houwink equation:

$$[\eta] = KM_v^a \quad [15]$$

where  $K$  and  $a$  are the constants for given polymer, solvent, and temperature. For many polymers these values can be found in the literature. The experimental determination of  $K$  and  $a$  requires a series of

**Table 1** Methods for the determination of molar mass of polymers

Method	Principle	Average type/range ( $\text{g mol}^{-1}$ )
Method of end groups	Determination of the concentration of the end groups by titration or spectral methods	$M_n$ Hundreds to 20 000
Vapor pressure osmometry (vapor phase osmometry)	Measurement of the temperature difference between droplets of polymer solution and pure solvent maintained in an atmosphere of solvent. Lower vapor pressure of solution compared to pure solvent results in the condensation of vapor into solution and temperature rise due to the released heat. Increase of vapor pressure due to the increased temperature compensates decrease of vapor pressure due to the presence of dissolved polymer	$M_n$ Hundreds to 20 000
Membrane osmometry	Measurement of osmotic pressure created by the flow of solvent into polymer solution through semipermeable membrane	$M_n$ 5000–1 000 000
Viscometry	Determination of the increase of viscosity of dilute polymer solution resulting from the presence of polymer molecules	$M_v$ Thousands to millions
Light scattering	Measurement of the intensity of light scattered by dilute polymer solution. The intensity is measured as a function of concentration and angle of observation	$M_w$ Hundreds to tens of millions
Ultracentrifugation		
(a) Sedimentation velocity	(a) Moving boundary is formed on application of a strong centrifugal field. The rate of boundary movement is measured	Various averages Thousands to tens of millions
(b) Sedimentation equilibrium Note: the method is rarely used for synthetic polymers	(b) The experiment is performed at lower rotation speed and measures the equilibrium concentration distribution of macromolecules that eventually form when sedimentation is balanced by diffusion	
Size-exclusion chromatography (gel permeation chromatography)	Separation of molecules according to their hydrodynamic volume during the flow through the column packed with porous packing. Determination of molar mass from the elution volume after column calibration, i.e., establishing the relation between molar mass and elution volume	Various averages and MMD Hundreds to millions
Matrix assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry	High molar mass ions are generated by irradiating a solid mixture of a polymer dissolved in a matrix compound with a pulsed laser beam. The matrix molecules absorb most of the laser's energy and thus the ionization process is 'soft' and produces molecular ions that are introduced into the mass spectrometer and analyzed. MALDI mass spectra are most often acquired using the time-of-flight technique that uses the differences in transit time through a drift region to separate ions of different masses	Various averages and MMD Hundreds to tens of thousands

samples with narrow MMD and known values of the intrinsic viscosity and molar mass. However, in practice the requirement of narrow MMD can be rarely fulfilled. The influence of polydispersity is relatively small when  $[\eta]$  is correlated with  $M_w$ . The constants can be considered as trustworthy for  $M_w/M_n < 1.5$ , correlation  $[\eta]$  versus  $M_w$  and molar mass range covering at least two orders of magnitude. The recommendations for reliable results are  $\eta_{sp}$  in the range of 0.2–0.8 and  $t_0 \sim 100$  s. An important feature of the intrinsic viscosity is that it increases with increasing temperature due to the expansion of polymer molecules.

### Light Scattering

Scattering of light is a natural phenomenon arising from the interaction of light with the matter. There are several equations describing the concentration and angular dependence of the intensity of light scattered by a dilute polymer solution. One of them is called the Zimm equation:

$$\frac{K^*c}{R_\theta} = \frac{1}{MP(\theta)} + 2A_2c + \dots \quad [16]$$

where  $c$  is the concentration of polymer in solution ( $\text{g ml}^{-1}$ ), for polydisperse polymer  $M$  is  $M_w$ ,  $A_2$  is

the second virial coefficient,  $K^*$  is an optical constant that for vertically polarized light is defined as:

$$K^* = \frac{4\pi^2 n_0^2}{\lambda_0^4 N_A} \left( \frac{dn}{dc} \right)^2 \quad [17]$$

where  $n_0$  is the refractive index of the solvent at the incident radiation wavelength,  $\lambda_0$  is the incident radiation wavelength at vacuum,  $N_A$  is the Avogadro number,  $dn/dc$  is the specific refractive index increment,  $R_\theta$  is the excess Rayleigh ratio defined as:

$$R_\theta = f \frac{(I_\theta - I_{\theta,\text{solvent}})}{I_0} \quad [18]$$

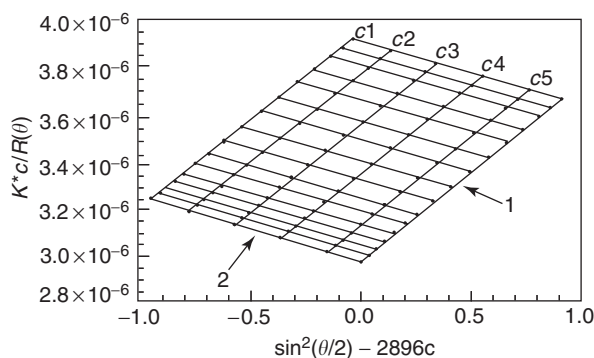
where  $I_\theta$  is the scattered light intensity of the polymer solution,  $I_{\theta,\text{solvent}}$  is the scattered light intensity of the solvent,  $I_0$  is the intensity of the incident radiation,  $f$  is an instrumental constant related to the geometry of the apparatus. The subscript  $\theta$  is the angle between the scattering direction and the incident light beam. In practice, the constant  $f$  is determined experimentally using a compound of well-known Rayleigh ratio (e.g., toluene).

$P(\theta)$  is the particle scattering function (the ratio of the intensity of radiation scattered at an angle of observation  $\theta$  to the intensity of radiation scattered at an angle zero,  $R_\theta/R_0$ ) that describes the decrease of the scattered light intensity with increasing angle of observation:

$$P(\theta) = 1 - \frac{16\pi^2}{3\lambda^2} R^2 \sin^2\left(\frac{\theta}{2}\right) + \dots \quad [19]$$

where  $\lambda$  is the wavelength of the incident light in a given solvent,  $R^2$  is the mean square radius (z-average in the case of a polydisperse polymer). The root mean square (RMS) radius ( $R$ , also called radius of gyration) describes the distribution of mass around the center of gravity and thus expresses the size of a particle regardless of its shape. It can be determined from the initial slope of the angular variation of the scattered light intensity. It is worth mentioning that the angular variation of scattered light intensity is negligible in the case of small polymers having  $R$  below  $\sim 10$  nm ( $M$  roughly  $100\,000 \text{ g mol}^{-1}$ ).

Light scattering experiments can be carried out either in batch mode using the light scattering photometer as a standalone instrument or in continuous flow mode when the light scattering instrument is used as a detector in SEC or FFF. In the batch mode, the sample is measured at several concentrations and the obtained data are processed using a Zimm plot, which allows simultaneous extrapolation to zero angle and zero concentration and thus the



**Figure 4** Zimm plot obtained by the measurement of five different concentrations ( $c1 = 3.28 \times 10^{-4} \text{ g ml}^{-1}$ ,  $c5 = 5.28 \times 10^{-5} \text{ g ml}^{-1}$ ) of polydisperse polystyrene. The intensity of light measured at 15 angles ( $23^\circ$ – $147^\circ$ ). Line 1 is the angular variation of the scattered light intensity at zero concentration; line 2 is the concentration dependence at zero angle. The slopes of lines 1 and 2 are used to compute  $R_z$  and  $A_2$ , respectively. The intersection of lines 1 and 2 is the reciprocal  $M_w$ . The factor 2896 was chosen automatically by software to spread the data points in the plot.  $M_w = 337\,000 \text{ g mol}^{-1}$ ;  $R_z = 34 \text{ nm}$ ;  $A_2 = 4.1 \times 10^{-4} \text{ mol ml g}^{-2}$ .

determination of  $M_w$ ,  $R_z$ , and  $A_2$ . An example of a Zimm plot is presented in Figure 4.

### Size-Exclusion Chromatography

SEC is the most important and widely used method for the determination of MMD. The most serious limitation of SEC is that it is a relative method requiring careful calibration of SEC columns, i.e., establishing the relation between molar mass and elution volume for the analyzed polymer. However, it has become a usual practice to calibrate the columns by polystyrene standards and to employ the polystyrene calibration to other polymers as well. This approach provides the polystyrene equivalent molar mass values, i.e., molar masses of polystyrene molecules of the same hydrodynamic volume as that of the analyzed polymer. It must be emphasized that the polystyrene equivalent molar masses can be up to a 100% or even more below or above the true values.

Application of a molar mass sensitive detector eliminates the column calibration. In principle, there are two kinds of the molar mass sensitive detectors. They are the light scattering photometers and viscometers. Actually the viscometers are not truly molar mass detectors since the measured quantity is the intrinsic viscosity and not the molar mass. The molar mass is determined from the experimental intrinsic viscosity and the so-called universal calibration, i.e., the relation  $\log(M[\eta])$  versus elution volume that is independent of polymer composition and structure. Likewise, in conventional SEC the obtained results are affected by the flow rate and temperature

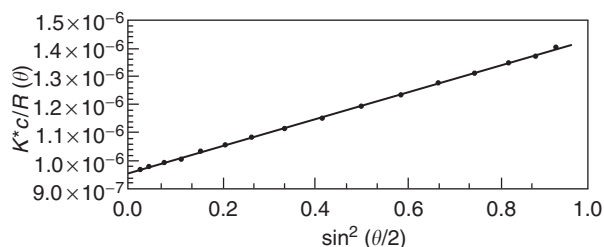
fluctuations, injected mass, injected volume, column performance, and non-size exclusion separation mechanisms.

Low-angle light scattering (LALS) and multiangle light scattering (MALS) detectors differ in the number of angles at which the intensity of light is measured. A LALS detector operates at a single angle low enough (e.g.,  $7^\circ$ ) that can be considered zero. The disadvantage of a LALS detector is a strong tendency to signal noise caused by particles (dust, column shedding) in the mobile phase. This is given by the fact that large particles scatter light with the highest intensity at very low angles. In the case of a MALS detector, the intensity of light is measured at multiple angles simultaneously and then extrapolated to zero angle, which avoids the measurements at very low angles and reduces signal noise. In SEC-MALS the data for each elution volume slice are extrapolated to zero angle and molar mass and RMS radius are obtained from the intercept and slope of the plot, respectively (Figure 5). A refractive index detector determines the concentrations at particular elution volume slices. Because of very low concentrations in SEC columns the term  $2A_2c \ll 1$  and thus the concentration dependence can be neglected.

### Field-Flow Fractionation

FFF is an alternative separation technique to SEC that crosses the borders from soluble region into particle region. The separation takes place in a thin channel between two parallel plates where the sample flows in the axial direction. The flow profile in the channel is parabolic with the maximum velocity in the middle. External field (e.g., flow, thermal) applied in the perpendicular direction forces molecules to different distances from the channel wall, i.e., to different velocity layers. Thus, the elution time increases with decreasing distance of particles from the channel wall.

FFF can be used for separation of particles in the range of a few nanometers to tens of micrometers. In



**Figure 5** Extrapolation of light scattering intensities to zero angle for a single SEC elution volume slice; 16 angles in the range of  $17^\circ$ – $147^\circ$ . The intercept and initial slope provide molar mass and RMS radius, respectively.  $M = 1\,043\,000\text{ g mol}^{-1}$ ;  $R = 47.1\text{ nm}$ ;  $c = 3.38 \times 10^{-5}\text{ g ml}^{-1}$ .

combination with a MALS detector, the FFF technique becomes efficient for the determination of the molar mass and size distributions. An important advantage of FFF is that it is a very gentle method. Compared to SEC the method eliminates significantly the possibility of shearing degradation. The fact that the channel does not contain any packing eliminates the possibility of enthalpic interactions that may strongly affect SEC results. Nonexistence of packing also eliminates entanglement that can occur during SEC separation of large highly branched molecules.

### Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry

In addition to MMD, MALDI-TOF mass spectrometry also provides compositional information (end groups and structure). However, in the case of polydisperse polymers MALDI commonly fails due to the mass dependent desorption ionization process or mass dependent detection efficiency. The inability of MALDI to provide reliable MMD for polydisperse polymers can be overcome by combination with SEC either by collecting SEC fractions and performing MALDI analysis offline or using a direct coupling. In the SEC-MALDI combination, the refractive index detector of the SEC instrument determines the polymer concentration and the corresponding molar mass is measured by MALDI. The determination of average molar masses of each fraction allows establishing the SEC calibration that can then be used to compute MMD by conventional SEC. However, the speed of the analysis and expensive instrumentation are the limitations of the SEC-MALDI methodology.

### Characterization of Long-Chain Branching

Long-chain branching (i.e., polymeric offshoots from a macromolecular chain) is another parameter strongly affecting polymer properties. Branching can arise from various side-reactions and is often overlooked in polymer characterization. The principle of the identification and characterization of long-chain branching is based on the fact that at a constant molar mass the molecular size decreases with increasing degree of branching. That means in order to characterize branching one has to obtain information about both the molar mass and the molecular size. Light scattering has been presented as the method providing simultaneously molar mass and RMS radius. However, the classical light scattering measurement of unfractionated sample provides molar mass and RMS radius of different moments ( $M_w$  versus  $R_z$ ) that cannot be directly compared.

**Table 2** Average branching ratio calculations for two randomly branched polystyrenes

Sample	Quantity	Experimental	Calculated for linear polymer	Equation used for calculation
#1	$[\eta]$ (ml g <sup>-1</sup> )	58.0	80.5	$[\eta] = 0.0117 \times M^{0.717}$
	$M_w$ (10 <sup>3</sup> g mol <sup>-1</sup> )	225	—	$R = 0.014 \times M^{0.585}$
	$R_z$ (nm)	26.8	37.4	
	$M_z$ (10 <sup>3</sup> g mol <sup>-1</sup> )	719	—	
	$g'$	—	0.72	
	$g$	—	0.51	
#2	$[\eta]$ (ml g <sup>-1</sup> )	82.5	146.5	$[\eta] = 0.0117 \times M^{0.717}$
	$M_w$ (10 <sup>3</sup> g mol <sup>-1</sup> )	519	—	$R = 0.014 \times M^{0.585}$
	$R_z$ (nm)	50.3	76.5	
	$M_z$ (10 <sup>3</sup> g mol <sup>-1</sup> )	2447	—	
	$g'$	—	0.56	
	$g$	—	0.43	

Branched samples can be characterized by the branching ratio (branching index) defined as the ratio of the mean square radii of the branched (br) and linear (lin) molecules at the same molar mass:

$$g = \left( \frac{R_{br}^2}{R_{lin}^2} \right)_M \quad [20]$$

An alternative branching ratio is based on the intrinsic viscosity:

$$g' = \left( \frac{[\eta]_{br}}{[\eta]_{lin}} \right)_M \quad [21]$$

The mutual relationship of the two branching ratios can be expressed as follows:

$$g^e = g' \quad [22]$$

where the draining parameter  $e$  varies in the range of 0.5–1.5.

The average  $g'$  for the unfractionated sample can be determined by measuring  $M_w$  with light scattering either in batch or SEC mode and the intrinsic viscosity with a viscometer. Then the intrinsic viscosity of the linear sample having the same  $M_w$  as the sample under investigation is calculated using the Mark–Houwink equation for the linear polymer. The branching ratio  $g'$  is calculated from the experimental and calculated intrinsic viscosities.

The average  $g$  can be calculated alike using the values of  $R_z$  and  $M_z$  determined by SEC–MALS. The RMS radius of a linear sample having the same  $M_z$  as the sample under investigation is calculated using the relation  $R = kM^b$  for the linear polymer. The average branching ratio  $g$  at the same  $M_z$  is calculated using the experimental and calculated RMS radii. Note: The two methods described do not provide identical results since the calculation is based on different size

parameters and different molar mass averages. Examples of calculation of average  $g$  and  $g'$  for two randomly branched polystyrene samples are given in **Table 2**.

The combined SEC–MALS technique overcomes the limitation of classical light scattering experiment since the molar masses and RMS radii are obtained for narrow fractions. Consequently, the branching ratio  $g$  can be determined as a function of molar mass. In addition, branching can be identified from the slope of the RMS radius versus molar mass plot even if the data for corresponding linear polymers are unavailable. However, the RMS radius versus molar mass plots may be biased by the abnormal elution behavior of some highly branched large molecules that can be entangled in the column packing and elute later together with smaller molecules and thus strongly increase the polydispersity within elution volume slices.

**See also:** **Centrifugation:** Analytical Ultracentrifugation. **Field-Flow Fractionation.** **Infrared Spectroscopy:** Overview; Industrial Applications. **Liquid Chromatography:** Size-Exclusion. **Mass Spectrometry:** Matrix-Assisted Laser Desorption/Ionization. **Particle Size Analysis.** **Polymers:** Natural Rubber; Polyurethanes.

## Further Reading

- Barth HG and Mays JW (eds.) (1991) *Modern Methods of Polymer Characterization*. New York: Wiley.
- Braun D, Cherdron H, and Ritter H (2001) *Polymer Synthesis: Theory and Practice*, 3rd edn. Berlin: Springer.
- Gedde UW (1995) *Polymer Physics*. London: Chapman and Hall.
- Kratohvil P (1987) *Classical Light Scattering from Polymer Solutions*. Amsterdam: Elsevier.
- Mori S and Barth HG (1999) *Size Exclusion Chromatography*. Berlin: Springer.

- Pasch H and Schrepp W (2003) *MALDI-TOF Mass Spectrometry of Synthetic Polymers*. Berlin: Springer.
- Pasch H and Trathnigg B (1999) *HPLC of Polymers*. Berlin: Springer.
- Podzimek S, Vlcek T, and Johann C (2001) Characterization of branched polymers by size exclusion chromatography coupled with multiangle light scattering detector. I. Size exclusion chromatography elution behavior of branched polymers. *Journal of Applied Polymer Science* 81: 1588–1594.
- Podzimek S and Vlcek T (2001) Characterization of branched polymers by SEC coupled with a multiangle light scattering detector. II. Data processing and interpretation. *Journal of Applied Polymer Science* 82: 454–460.
- Schimpf M, Caldwell K, and Giddings JC (eds.) (2000) *Field-Flow Fractionation Handbook*. New York: Wiley.
- Wyatt PJ (1993) Light scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta* 272: 1–40.

## Polyurethanes

**T M Schmitt**, BASF Corporation, Wyandotte, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Polyurethanes are addition polymers containing the urethane linkage,  $\text{—RNHCOOR'—}$ . They have been of great commercial importance for 50 years, with many applications as foam (furniture and automotive seating, insulation, carpet underlay), coatings (automotive and architectural paint), adhesives and sealants, and as elastomers for a variety of special purposes.

Most synthetic polymers are made in chemical plants, with the completed products shipped to fabricators for final molding, spinning, blowing, etc. Polyurethanes are different in that the polymer itself is made on the spot in the facilities of appliance and furniture manufacturers, at automotive suppliers, and on construction sites. For example, polyurethane insulation is formed within the shell of a refrigerator during appliance assembly. Usually, there is no need to drive off reaction byproducts to force the reaction to completion and the polyurethane is formed without addition of heat.

### Analysis of Polyurethane Raw Materials

Since most industrial customers purchase raw materials and form the polyurethane at point of use, analysis of the raw materials is commercially more important and much more common than analysis of the polyurethane product. Raw materials are analyzed to confirm their identity and to assess their suitability for use. The formation of a polyurethane, like the formation of other polymers, is quite sensitive to the purity of the reagents. Since it is

impossible to purify the polymer once it is formed, any faults contributed by the raw materials remain part of the polymer for the life of the product.

In simplest terms, the raw materials for urethanes consist of compounds having two or more isocyanate groups and compounds having two or more hydroxyl groups. These hydroxy compounds, commonly called polyols, are most often polyethers and polyesters, although some other materials are also encountered.

### Polyols

Long, linear, difunctional polyols give flexibility to polyurethanes while short, branched, polyols with three or more hydroxyl groups per molecule impart rigidity. It is critical that the polyol has the molecular weight and functionality specified, and that it has the reactivity appropriate for the particular application. The reactivity is controlled by the presence of catalysts (whether they be impurities or deliberate additives) and by chemical functionality, most importantly whether the hydroxyl groups are primary or secondary, and whether the reactants contain amine groups.

The most important compounds are polyetherols, polyesterols, and graft polyols (**Figure 1**). Polyetherols are the most common polyols and most commercial products are made chiefly from ethylene oxide and propylene oxide. Polyesterols are widely used in elastomeric polyurethanes. Typical compounds are aliphatic polyesters, such as poly(ethylene adipate), made so that there are only hydroxyl end groups, with no residual acid groups. Graft polyols (also called polymer polyols) consist of a matrix of a conventional polyol, which also contains rubbery (such as styrene/acrylonitrile copolymer) particles chemically bound to polyol molecules. These 'grafted' particles reinforce the final polyurethane, giving improved physical properties.



- Pasch H and Schrepp W (2003) *MALDI-TOF Mass Spectrometry of Synthetic Polymers*. Berlin: Springer.
- Pasch H and Trathnigg B (1999) *HPLC of Polymers*. Berlin: Springer.
- Podzimek S, Vlcek T, and Johann C (2001) Characterization of branched polymers by size exclusion chromatography coupled with multiangle light scattering detector. I. Size exclusion chromatography elution behavior of branched polymers. *Journal of Applied Polymer Science* 81: 1588–1594.
- Podzimek S and Vlcek T (2001) Characterization of branched polymers by SEC coupled with a multiangle light scattering detector. II. Data processing and interpretation. *Journal of Applied Polymer Science* 82: 454–460.
- Schimpf M, Caldwell K, and Giddings JC (eds.) (2000) *Field-Flow Fractionation Handbook*. New York: Wiley.
- Wyatt PJ (1993) Light scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta* 272: 1–40.

## Polyurethanes

**T M Schmitt**, BASF Corporation, Wyandotte, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Polyurethanes are addition polymers containing the urethane linkage,  $\text{—RNHCOOR'—}$ . They have been of great commercial importance for 50 years, with many applications as foam (furniture and automotive seating, insulation, carpet underlay), coatings (automotive and architectural paint), adhesives and sealants, and as elastomers for a variety of special purposes.

Most synthetic polymers are made in chemical plants, with the completed products shipped to fabricators for final molding, spinning, blowing, etc. Polyurethanes are different in that the polymer itself is made on the spot in the facilities of appliance and furniture manufacturers, at automotive suppliers, and on construction sites. For example, polyurethane insulation is formed within the shell of a refrigerator during appliance assembly. Usually, there is no need to drive off reaction byproducts to force the reaction to completion and the polyurethane is formed without addition of heat.

### Analysis of Polyurethane Raw Materials

Since most industrial customers purchase raw materials and form the polyurethane at point of use, analysis of the raw materials is commercially more important and much more common than analysis of the polyurethane product. Raw materials are analyzed to confirm their identity and to assess their suitability for use. The formation of a polyurethane, like the formation of other polymers, is quite sensitive to the purity of the reagents. Since it is

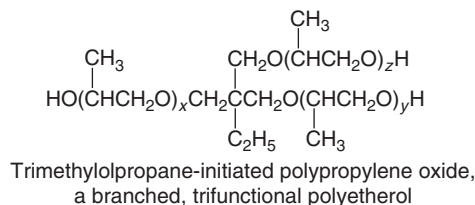
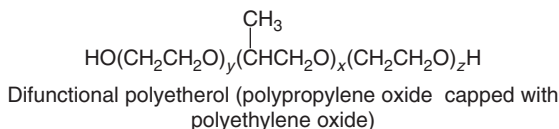
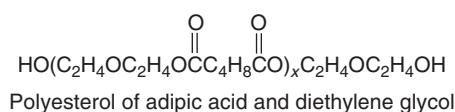
impossible to purify the polymer once it is formed, any faults contributed by the raw materials remain part of the polymer for the life of the product.

In simplest terms, the raw materials for urethanes consist of compounds having two or more isocyanate groups and compounds having two or more hydroxyl groups. These hydroxy compounds, commonly called polyols, are most often polyethers and polyesters, although some other materials are also encountered.

### Polyols

Long, linear, difunctional polyols give flexibility to polyurethanes while short, branched, polyols with three or more hydroxyl groups per molecule impart rigidity. It is critical that the polyol has the molecular weight and functionality specified, and that it has the reactivity appropriate for the particular application. The reactivity is controlled by the presence of catalysts (whether they be impurities or deliberate additives) and by chemical functionality, most importantly whether the hydroxyl groups are primary or secondary, and whether the reactants contain amine groups.

The most important compounds are polyetherols, polyesterols, and graft polyols (**Figure 1**). Polyetherols are the most common polyols and most commercial products are made chiefly from ethylene oxide and propylene oxide. Polyesterols are widely used in elastomeric polyurethanes. Typical compounds are aliphatic polyesters, such as poly(ethylene adipate), made so that there are only hydroxyl end groups, with no residual acid groups. Graft polyols (also called polymer polyols) consist of a matrix of a conventional polyol, which also contains rubbery (such as styrene/acrylonitrile copolymer) particles chemically bound to polyol molecules. These 'grafted' particles reinforce the final polyurethane, giving improved physical properties.



**Figure 1** Common types of polyols used in polyurethane products.

For two generations, wet chemical methods have been standard for quality control of polyols. Most of these methods are quite precise and give results directly related to suitability for use.

1. Hydroxyl number is a critically important parameter since it is the hydroxyl group which reacts with isocyanate to form the urethane linkage. The classical determination is by esterification of the hydroxyl groups with excess acetic anhydride or phthalic anhydride, followed by titration of the surplus reagent. Esterification remains the referee method, although it has been largely supplanted by near infrared (IR) spectrophotometry for routine quality control. Reaction with phenyl isocyanate is also sometimes used for hydroxyl determination, as is NMR spectrometry. Hydroxyl number by any of these techniques is indirectly proportional to number average molecular weight.
2. Acid number is of concern with both polyether and polyester polyols, since acidity affects the kinetics of the urethane reaction. Determination is by potentiometric titration.
3. Water content above a threshold value is critical because water reacts with isocyanate to form urea in a reaction competitive to urethane group formation. The determination is a straightforward application of the Karl Fischer titration, but precautions are necessary when weighing and transferring the sample because of the ease with which the polyol absorbs moisture from laboratory air. Usually, the sample is weighed out by difference from a disposable syringe.
4. Alkalinity is important because of its effect on reactivity. There is a potential for high alkalinity in polyether polyols, most of which are synthesized

using potassium hydroxide catalysis. Usually, the residual catalyst is removed by liquid–liquid extraction or by filtration. As an alternative or supplement to direct titration of alkalinity, residual sodium and potassium may be determined by atomic absorption or induction-coupled argon plasma emission spectroscopy. The latter approach is more convenient, since the polyol can be analyzed immediately after dilution in a suitably pure solvent, with no sample digestion required.

5. It is common in the industry to measure apparent pH of polyols after dissolution in an alcohol. Polyols with apparent pH below 4 or above 10 may give trouble in the urethane reaction. Because of poor reproducibility of measurements in the 5–9 range, there is less correlation with performance at intermediate apparent pH.

6. Nitrogen content (primary, secondary, tertiary) is important for reactivity considerations and is determined by classical nonaqueous potentiometric titration.

7. Reactivity is determined by mixing the polyol with an isocyanate compound, catalyst, and the other components of a standard foam recipe, then measuring the time required for maximum foam rise, maximum exotherm, or some related phenomenon.

8. Saponification number is characteristic of polyesters and is determined by titration.

9. Unsaturation is important for special cases, depending on the nature of the polyol. This is because, during polyetherol synthesis, an undesirable side reaction can lead to formation of a terminal double bond and no terminal hydroxyl group. The resulting polyol therefore has less functionality than intended.

10. Ash, color, refractive index, specific gravity, suspended matter, and viscosity are determined by traditional methods.

Most methods for quality control of polyols have been used internationally for decades and have official status as ASTM, or, in some cases, ISO standards. While there has been evolutionary change, most procedures have been altered little over the past generation.

Instrumental analysis is used more often for troubleshooting and research than for routine analysis.

1. Infrared and Raman spectroscopy are sufficient to identify the gross composition of polyols, but will not give the precise characterization of NMR spectroscopy. Near IR analysis is used in quality control laboratories for the determination of hydroxyl number and water, after calibration with a closely related set of polyols.

2. NMR spectrometry is used for determination of primary and secondary OH content and ethylene

oxide content, as well as determination of heteric (random) or block polyether structure and capping.

3. Gas chromatography (GC) is used mainly for determination of residual monomer and perhaps initiator fragments from the graft component.

4. Liquid chromatography (LC). While normal-phase or reversed-phase LC are sometimes applied for special purposes, the usual mechanism of polyol characterization is size-exclusion chromatography (SEC) using differential refractive index or evaporative light scattering detection. Many polyester polyols, for example, those based upon terephthalate esters, can also be detected by ultraviolet (UV) absorbance. SEC yields both number average and weight average molecular weight values, as well as information about the polydispersity and whether the polyol is a single polymer or is perhaps a blend of two or more components.

### Isocyanate Compounds

Toluene diisocyanate (TDI) and the various forms of methylene 4,4'-bis(phenylisocyanate) (MDI) are the most common isocyanate compounds in commercial polyurethanes. Aliphatic isocyanates (hexamethylene diisocyanate) are not generally used in foams, but

rather in elastomers, coatings, and other applications (Figure 2).

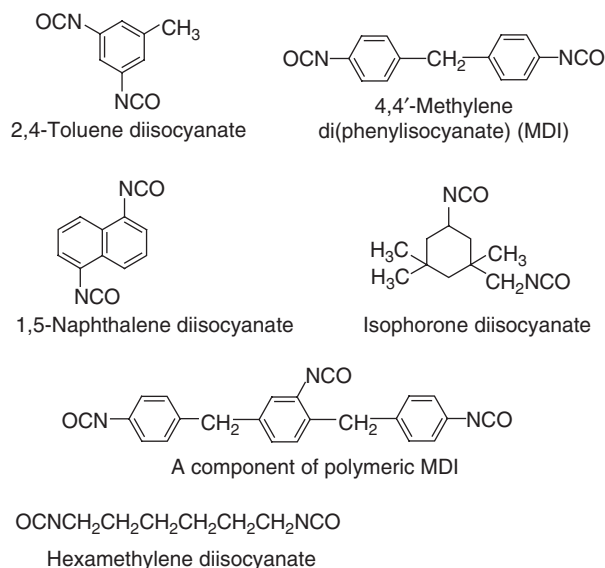
Because volatile isocyanate monomers present an industrial hygiene challenge, they are often supplied to the industrial user in a modified form. For example, an isocyanate-terminated 'prepolymer' may be formed by reacting an excess of isocyanate with the polyol to form a liquid material, which has little of the starting isocyanate. However, unless special measures are taken (e.g., reacting a large excess of diisocyanate monomer with the hydroxy compound, then stripping away the excess monomer), the prepolymer will contain free diisocyanate monomer and also oligomers of, for example, three molecules of diisocyanate reacted with two polyol molecules (Figures 3–5).

Because of the toxicity of isocyanates, laboratory operations are designed to minimize personal exposure and are performed in a fume hood. Exposure to high levels of isocyanates, particularly toluene diisocyanate, can cause 'sensitization' of a worker, so that afterward he/she cannot tolerate exposure even to levels of isocyanate considered safe for others.

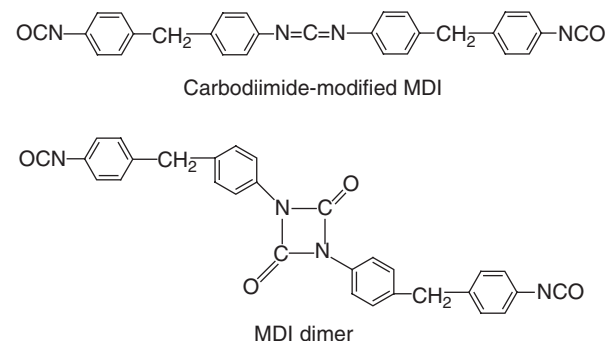
Isocyanates react readily with the moisture in laboratory air, forming ureas, so stringent precautions are taken when sampling to avoid contamination by moisture. This is usually accomplished by purging the headspace of sample containers with a stream of dry gas, most conveniently nitrogen.

Wet chemical analysis of isocyanates is well established.

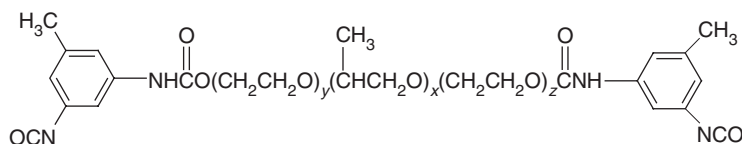
1. Isocyanate (NCO) content is critical for calculation of reaction stoichiometry. Total NCO content is



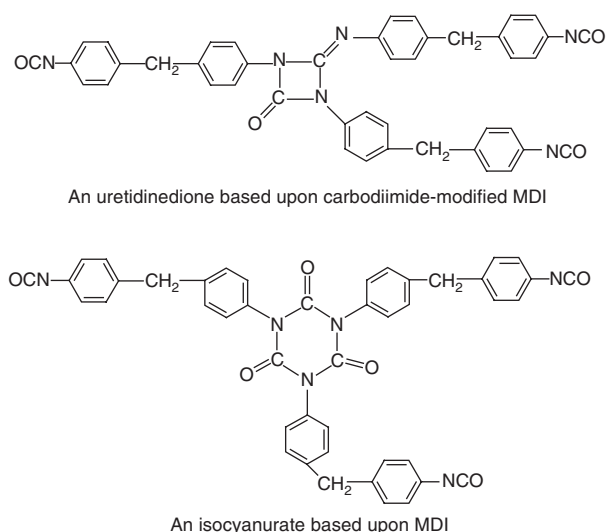
**Figure 2** Some commercially important isocyanate compounds.



**Figure 4** Two types of modified MDI.



**Figure 3** Prepolymer based on 2,4-TDI and a difunctional polyetherol.



**Figure 5** Two more modified MDI compounds.

determined by reaction with excess dibutylamine to form a substituted urea, followed by titration of the unreacted dibutylamine with acid.

2. Hydrolyzable chlorine content is important because of the effect of acidity on reaction speed. Most isocyanates are made by phosgenation of the corresponding amines, and side reactions lead to the presence of what can be thought of as organic addition compounds of hydrochloric acid. These are hydrolyzed under more or less rigorous conditions to yield chloride ion, which is titrated in the usual way with silver nitrate solution. 'Total chlorine' content is related to overall purity of the isocyanate. It is determined by combustion of the organic matter in a Schöniger oxygen flask or Paar bomb, followed by measurement of liberated chloride ion by titration or by ion chromatography.

3. In the absence of additives, acidity is stoichiometrically equal to easily hydrolyzable chlorine, determined above, and is useful for predicting reactivity.

4. Properties such as viscosity, color, and specific gravity are determined by the usual industrial methods.

Instrumental analysis may include:

1. X-ray fluorescence is the most convenient technique to determine total chlorine and bromine content. Bromine may be introduced as an impurity in the phosgene used in isocyanate synthesis.
2. GC is suitable for the determination of many simple isocyanates, including the most common, namely TDI isomers and monomeric MDI isomers. Both assay and isomer content are determined, and GC is a useful method for determining the

residual monomeric diisocyanate content of prepolymers. Impurities such as solvents are also apparent by GC.

3. LC is used in production facilities to characterize MDI and its precursor, methylene dianiline (MDA). Either normal-phase or reversed-phase conditions can be used. The isocyanates must first be derivatized by reaction with an alcohol. The isomer composition is determined by LC, as well as the distribution of monomeric and 2- and 3-ring oligomeric components. LC complements SEC for MDI/MDA characterization.

4. SEC is an excellent method of characterizing higher molecular weight isocyanates such as polymeric MDI and prepolymers. The isocyanate end groups are derivatized with methanol to prevent reaction and the sample is subjected to conventional SEC using THF as solvent. Polymeric MDI oligomers are separated up to about the 5-ring compounds. In the case of prepolymers, clear separation is obtained of oligomers, together with a peak for the unreacted isocyanate compound. This latter is most easily quantified using an internal standard such as toluene. SEC is especially useful for studying the 'pot life' of prepolymers. Samples can be pulled from the prepolymer over the course of several hours, quenching the samples promptly in methanol. The whole series is then analyzed in batch mode at a convenient time to observe the changes in molecular weight distribution and free isocyanate as the batch ages. A warning: allophanate cross-linking of the prepolymer will not be seen by this technique, since the cross-linked material is insoluble in the SEC system.

5. Infrared spectroscopy is effective for detecting carbodiimide groups in isocyanates, as well as following dimer formation and other modification. In times past, IR was used for determining TDI isomer content, but now this is almost always performed by GC.

### Other Polyurethane Raw Materials

Besides the polyol and isocyanate, many other materials are used in polyurethanes, albeit in much lower volumes. These can be grouped into materials, which become part of the polyurethane matrix itself (such as chain extenders), and materials, which are additives physically but not chemically mixed with the polymer (such as antioxidants and blowing agents).

Chain extenders and cross-linkers (encountered especially in thermoplastic urethanes) are usually short diamines or diols, such as 1,4-butanediol or *tetrakis*(2-hydroxypropyl)ethylenediamine. Alkanolamines and triols like glycerol are also used. Most of these can be easily determined by GC.

Blowing agents are gases released during formation of the urethane which cause the final product to be a rigid or flexible foam. The main blowing agent in flexible foams is water, which forms  $\text{CO}_2$  by reaction with isocyanate groups. Carbon dioxide has disadvantages in rigid foam, especially in foam used for insulation. Chlorofluorocarbons, which boil at the temperature of the polyurethane reaction were used in the past, but are no longer found due to ecological considerations. Hydrochlorofluorocarbons and hydrofluorocarbons are still found, although they have been replaced for many purposes by simple hydrocarbons such as pentane.

Blocking agents are common in coatings. They prevent the cross-linking reaction from occurring until heat is applied. Phenols, alcohols, and other compounds, which react reversibly with isocyanates are used for this purpose. Most are readily determined by GC.

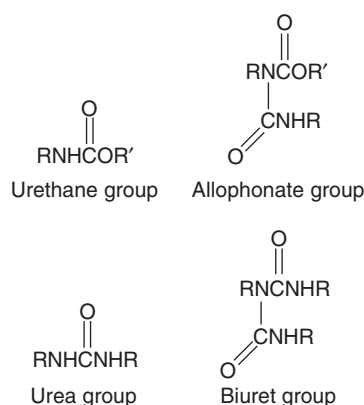
Oxidative stabilizers are added to prevent yellowing or 'scorching', especially in large foam blocks which develop high internal temperature during manufacture. Hindered phenols such as butylated hydroxytoluene are often used. Stabilizers are usually determined by HPLC or GC analysis, although UV absorbance measurement of a foam extract can often be used for quality control.

**Catalysts.** To make the final polyurethane, common catalysts are tertiary amines and organotin compounds, often used in mixtures. Amine catalysts favor the reaction of isocyanate with water, producing urea linkages and  $\text{CO}_2$ , which acts as a blowing agent. On the other hand, organotin catalysts favor the isocyanate/hydroxyl reaction. Additional catalysts and catalyst decomposition products may be present if graft polyols are part of the mixture. The most common of these catalysts are free radical initiators. Phosphorus compounds may be present in the case of carbodiimide-modified MDI.

Other materials may include pigments and fillers, surfactants, mold release/lubricants, and flame retardants. Coatings will have additional components, including UV stabilizers and hindered amine light stabilizers.

## Analysis of the Polymer

Most testing of polyurethane products is performed by specialists seeking information for well-defined purposes, whether it be to troubleshoot problems such as loss of physical properties, or perhaps to assess the ecological impact of the product in terms of what gases are released at equilibrium and under



**Figure 6** Linkages found in polyurethanes.

fire conditions, or what is the cause of a specific color or odor. Sometimes though, it is necessary simply to analyze the finished polyurethane in order to learn what are the chief components, namely, what were the isocyanate and polyol compounds from which it was made, and the catalysts, additives, blowing agents, and so on. This is a challenging task, because most polyurethanes cannot be dissolved without breaking chemical bonds and fundamentally changing the nature of the polymer.

Besides urethane bonds, other chemical functionalities are present due to reaction of isocyanate groups with water and other compounds. In fact, urea and isocyanurate bonds are more abundant than the urethane linkages in some commercial products that are designated as polyurethanes. Allophanate, biuret, urea, carbodiimide, and isocyanurate linkages may be found. These bonds certainly affect the properties of the final polymer, but unfortunately it is almost impossible to determine them quantitatively (Figure 6).

Polyurethane elastomers are made up of predominantly linear molecules which contain rigid and flexible segments. This is true whether the elastomer composes an elastic fiber for clothing (spandex) or a shock-absorbing component in an automobile. The rigid and flexible (also called hard and soft) segments are balanced to give the physical properties needed. The flexible segments are formed from longer diols, while the rigid segments are composed of short cross-linking units and the diisocyanate components.

## Direct Analysis of Polyurethanes

1. Infrared spectroscopy is effective for qualitative analysis of polyurethanes, confirming that the material is a polyurethane and, depending upon the

experience of the spectroscopist, even telling something about the raw materials. For example, IR can indicate whether a polyether or polyester polyol was used, whether the isocyanate component is aliphatic or aromatic, and if aromatic, whether a polymeric MDI was used. Because of overlapping absorbance bands, it is much more difficult to characterize a polymer in terms of urea, allophanate, carbodiimide, biuret, and urethane structures, and small differences between urethane polymers are undetectable by IR.

2. Solid-state NMR analysis is useful for special purposes, especially for understanding chemical changes caused by oxidation or other degradation reactions.

3. SEC is of limited applicability with polyurethanes. Some thermoplastic urethanes, including spandex, can be characterized by simply dissolving and performing conventional SEC analysis, although this is most valuable when used for comparative purposes rather than for developing absolute information. The more common cross-linked polyurethane foams used in furniture and insulation cannot be dissolved and analyzed, and in any case the concept of molecular weight is meaningless in such systems.

4. Quite a number of physical test methods are routinely applied to polyurethanes, both during development of the product and for quality control purposes. Their discussion is beyond the scope of this article, but it should be noted that most of the testing performed on the final product is physical property testing, with chemical analyses being comparatively rare.

5. Identification of cell gas. This analysis is almost always performed by GC-mass spectrometry (GC-MS), with the gas removed from the closed cells using a syringe or by crushing the foam in a closed container.

6. Filler content is usually determined by quantitatively ashing the polymer, then performing qualitative or quantitative analysis of the inorganic residue.

7. Flame retardants must be determined by methods specific for the individual compound. For melamine, this can be as simple as extracting a foam with hot water and measuring the material removed. Brominated flame retardants are most easily measured by X-ray fluorescence determination of total bromine.

8. Mold release and lubricants are usually extracted with a suitable solvent and identified qualitatively by IR or MS. Quantitative determination is rarely of interest, but is most easily performed by determination of the cation, if present. For example, zinc stearate is most easily determined by measuring the total zinc content.

## Analysis after Hydrolysis/Solvolysis

There is a substantial literature describing the decomposition of urethane compounds under various conditions of hydrolysis and solvolysis to give fragments chemically related to the materials that reacted to form the original polyurethane. Usually, the polyether polyols and chain extenders can be recovered as simple derivatives, while the isocyanate compounds are recovered as the amine analogs. The reaction products are recovered by liquid-liquid extraction at controlled pH, or perhaps by preparative liquid chromatography. In the more recent work, this degradation has been conducted online as part of pyrolysis GC-MS analysis. This can be a very rapid technique for qualitative analysis of a polymer, identifying the main components in only an hour or two.

Allophanate and biuret bonds in polyurethane foam can be determined by degrading the foam with amines and subsequently performing quantitative titrations. This methodology is not of interest for routine analysis but is sometimes useful for fundamental understanding of the polymerization reaction.

## Isocyanates and Amines in Prepolymers and Finished Products

Many useful methods have been published for the determination of trace isocyanates in air and in polyurethane products by derivatization and GC, LC, or SEC quantification. For example, industrial hygiene determination of TDI in air by OSHA method 42 is performed by drawing air through a filter impregnated with 1-(2-pyridyl)piperazine. Derivatives are formed during sampling, which are then determined by LC. Also of concern are the amine hydrolysis products of the isocyanates. For example, toluenediamine (TDA) may be monitored in furniture foam. This is generally accomplished by liquid-solid extraction or vacuum distillation, followed by GC or LC quantification.

*See also:* **Gas Chromatography:** Overview. **Infrared Spectroscopy:** Overview. **Liquid Chromatography:** Size-Exclusion. **Mass Spectrometry:** Overview. **Nuclear Magnetic Resonance Spectroscopy Techniques:** Solid-State. **Polymers:** Natural Rubber; Synthetic.

## Further Reading

The American Society for Testing and Materials, West Conshohocken, PA, maintains most of the methods for testing polyurethane raw materials. These are also listed on the web site of the Alliance for the Polyurethanes Industry, <http://www.polyurethane.org/>

Bonart R and Demmer P (1982) GPC determination of monomeric diisocyanate in prepolymer for segmented polyurethanes (in German). *Colloid and Polymer Science* 260: 518–523.



- David DJ and Staley HB (1969) *Analytical Chemistry of the Polyurethanes*. New York: Wiley.
- Frisch KC (1997) High-performance polyurethanes. In: Hatada K, Kitayama T, and Vogl O (eds.) *Macromolecular Design of Polymeric Materials*. New York: Dekker.
- Mehl JT, Murgasova R, Dong X, Hercules DM, and Nefzger H (2000) Characterization of polyether and polyester polyurethane soft blocks using MALDI mass spectrometry. *Analytical Chemistry* 72: 2490–2498.
- Oertel G (ed.) (1994) *Polyurethane Handbook: Chemistry–Raw Materials–Processing–Application–Properties*, 2nd edn. Cincinnati: Hanser Gardner.
- Servay T, Voelkel R, Schmiedberger H, and Lehmann S (2000) Thermal oxidation of the methylene diphenylene unit in MDI-TPU. *Polymer* 41: 5247–5256.
- Szycher M (1999) *Szycher's Handbook of Polyurethanes*. Boca Raton: CRC Press.
- Thomas P (ed.) (1998) *Waterborne and Solvent Based Surface Coating Resins and their Applications*, vol. III, *Polyurethanes*. Chichester: Wiley.
- Ulrich H (1997) Urethane polymers. In: Howe-Grant M (ed.) *Kirk-Othmer Encyclopedia of Chemical Technology*, 4th edn., vol. 24, pp. 695–726. New York: Wiley.

## POTENTIOMETRIC STRIPPING ANALYSIS

**M Honeychurch**, University of Queensland, Brisbane, QLD, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

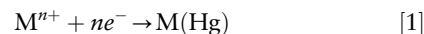
Potentiometric stripping analysis (PSA) has become more widely used over the last 20 years in part because of the advent of commercial instrumentation during that period. In many ways the method is analogous to voltammetric stripping methods insofar as it comprises an initial preconcentration step in which the analyte is accumulated onto or into the working electrode followed by a stripping step in which the analyte is 'stripped' back into solution. The electrochemical cell is under potentiostatic control during the accumulation period,  $t_{\text{acc}}$ ; however, during the stripping step the potentiostatic control is discontinued and the change in the working electrode potential versus time is measured. Therefore, the method is more appropriately termed chronopotentiometric stripping analysis. The output is a sigmoidal-shaped chronopotentiogram (**Figure 1**) with each plateau (parallel to the potential axis) corresponding to the transition or stripping time,  $\tau$ . This is analogous to a potentiometric titration with each plateau corresponding to an 'endpoint' equating to the complete reaction of the accumulated analyte with the difference being that the change in electrode potential is plotted versus time rather than versus the volume of titrant added.

Extremely low detection limits (below  $\mu\text{g l}^{-1}$ ) are achieved due to the preconcentration step. Additionally, the method allows multielement detection and the apparatus required for the analysis can be purchased for a relatively small cost.

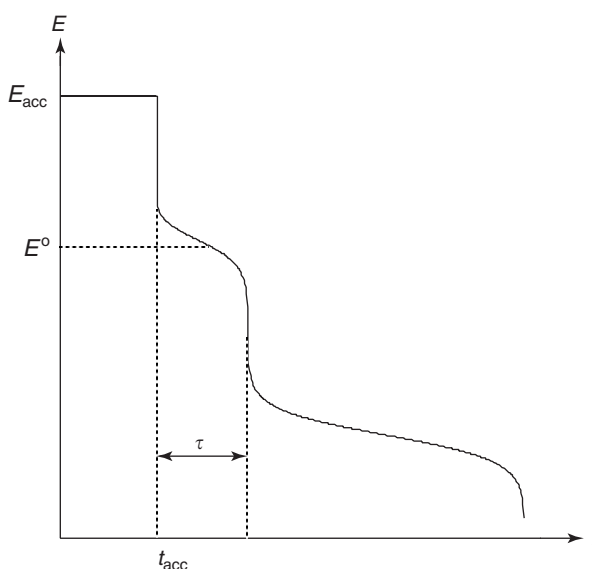
### Principles

#### The Preconcentration Step

By far the most common mode of preconcentration is the accumulation of a metal ion by the formation of an amalgam in a mercury drop or thin mercury film by electrolytic reduction at a fixed potential:



Amalgam-forming elements such as cadmium, copper, lead, zinc can be determined simultaneously in this way. Other elements such as arsenic, mercury, and silver can be determined by electrolytic reduction



**Figure 1** Typical stripping chronopotentiogram obtained from PSA. After an initial accumulation period,  $t_{\text{acc}}$ , at potential,  $E_{\text{acc}}$ , potentiostatic control of the electrochemical cell is discontinued and change in potential is plotted versus time.

- David DJ and Staley HB (1969) *Analytical Chemistry of the Polyurethanes*. New York: Wiley.
- Frisch KC (1997) High-performance polyurethanes. In: Hatada K, Kitayama T, and Vogl O (eds.) *Macromolecular Design of Polymeric Materials*. New York: Dekker.
- Mehl JT, Murgasova R, Dong X, Hercules DM, and Nefzger H (2000) Characterization of polyether and polyester polyurethane soft blocks using MALDI mass spectrometry. *Analytical Chemistry* 72: 2490–2498.
- Oertel G (ed.) (1994) *Polyurethane Handbook: Chemistry–Raw Materials–Processing–Application–Properties*, 2nd edn. Cincinnati: Hanser Gardner.
- Servay T, Voelkel R, Schmiedberger H, and Lehmann S (2000) Thermal oxidation of the methylene diphenylene unit in MDI-TPU. *Polymer* 41: 5247–5256.
- Szycher M (1999) *Szycher's Handbook of Polyurethanes*. Boca Raton: CRC Press.
- Thomas P (ed.) (1998) *Waterborne and Solvent Based Surface Coating Resins and their Applications*, vol. III, *Polyurethanes*. Chichester: Wiley.
- Ulrich H (1997) Urethane polymers. In: Howe-Grant M (ed.) *Kirk-Othmer Encyclopedia of Chemical Technology*, 4th edn., vol. 24, pp. 695–726. New York: Wiley.

## POTENTIOMETRIC STRIPPING ANALYSIS

**M Honeychurch**, University of Queensland, Brisbane, QLD, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

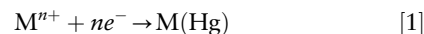
Potentiometric stripping analysis (PSA) has become more widely used over the last 20 years in part because of the advent of commercial instrumentation during that period. In many ways the method is analogous to voltammetric stripping methods insofar as it comprises an initial preconcentration step in which the analyte is accumulated onto or into the working electrode followed by a stripping step in which the analyte is 'stripped' back into solution. The electrochemical cell is under potentiostatic control during the accumulation period,  $t_{\text{acc}}$ ; however, during the stripping step the potentiostatic control is discontinued and the change in the working electrode potential versus time is measured. Therefore, the method is more appropriately termed chronopotentiometric stripping analysis. The output is a sigmoidal-shaped chronopotentiogram (**Figure 1**) with each plateau (parallel to the potential axis) corresponding to the transition or stripping time,  $\tau$ . This is analogous to a potentiometric titration with each plateau corresponding to an 'endpoint' equating to the complete reaction of the accumulated analyte with the difference being that the change in electrode potential is plotted versus time rather than versus the volume of titrant added.

Extremely low detection limits (below  $\mu\text{g l}^{-1}$ ) are achieved due to the preconcentration step. Additionally, the method allows multielement detection and the apparatus required for the analysis can be purchased for a relatively small cost.

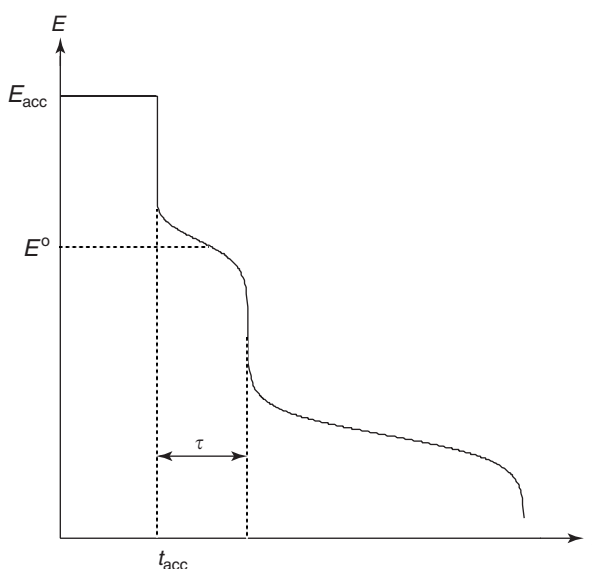
### Principles

#### The Preconcentration Step

By far the most common mode of preconcentration is the accumulation of a metal ion by the formation of an amalgam in a mercury drop or thin mercury film by electrolytic reduction at a fixed potential:

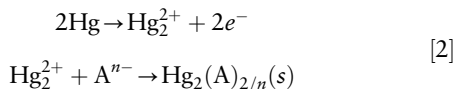


Amalgam-forming elements such as cadmium, copper, lead, zinc can be determined simultaneously in this way. Other elements such as arsenic, mercury, and silver can be determined by electrolytic reduction



**Figure 1** Typical stripping chronopotentiogram obtained from PSA. After an initial accumulation period,  $t_{\text{acc}}$ , at potential,  $E_{\text{acc}}$ , potentiostatic control of the electrochemical cell is discontinued and change in potential is plotted versus time.

at solid electrodes such as gold or bismuth. An alternative method of electrolytic preconcentration occurs by electrolytic oxidation during which the mercury working electrode is oxidized and an insoluble salt is formed between the electrode and (generally) inorganic anionic analytes:



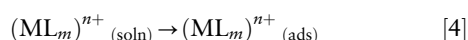
During the preconcentration step, only a small amount of analyte is being removed from the solution; therefore, the bulk concentration of the analyte in solution is assumed to be constant. It is also assumed that the kinetics of amalgamation is sufficiently rapid for the process to be mass transport controlled. The amount of the analyte that is accumulated during the preconcentration step is proportional to the bulk concentration. For example, the concentration of an amalgamated metal ion is

$$[\text{M}(\text{Hg})] \propto \frac{D_{\text{M}^{n+}} [\text{M}^{n+}] t_{\text{acc}} A}{\delta_{\text{acc}}} \quad [3]$$

where  $D_{\text{M}^{n+}}$  is the diffusion coefficient,  $A$  the electrode area,  $t_{\text{acc}}$  the accumulation time, and  $\delta_{\text{acc}}$  the diffusion layer thickness during the accumulation process. Increasing the  $t_{\text{acc}}$  increases the amount of analyte accumulated, thereby increasing the analytical signal and lowering the detection limit (Figure 2). Some trade-off is inevitably made between electrolysis

time and detection limit. Stirring the solution during the preconcentration reduces the diffusion layer thickness  $\delta_{\text{acc}}$ , thereby increasing the amount of analyte that is accumulated and reducing the accumulation period that is required.

Since the early 1980s, a number of methods have been developed to analyze various compounds by stripping voltammetry following the adsorptive accumulation of the species on the electrode. The adsorptive accumulation procedure has enabled analysts to measure extremely low concentrations of species that cannot be accumulated by amalgam formation. Species measured following adsorptive accumulation include metal complexes, organic compounds, and many biological macromolecules:



Adsorptive accumulation can also be used in chronopotentiometric stripping as the preconcentration step allowing a large number of organic compounds to be determined. It is assumed that the adsorption kinetics is sufficiently rapid for the process to be mass transport controlled. During the preconcentration step, the solution is normally stirred in order to enhance the mass transport of the analyte to the electrode surface, thereby reducing the time required for preconcentration. For nonelectrolytic or adsorptive accumulation,  $t_{\text{acc}}$  has an upper limit governed by the time required for the electrode to be saturated with a monolayer of adsorbate. This is determined experimentally and will depend on the mass transport conditions that exist in the electrochemical cell. For example, the amount of adsorbed metal ion complex would be

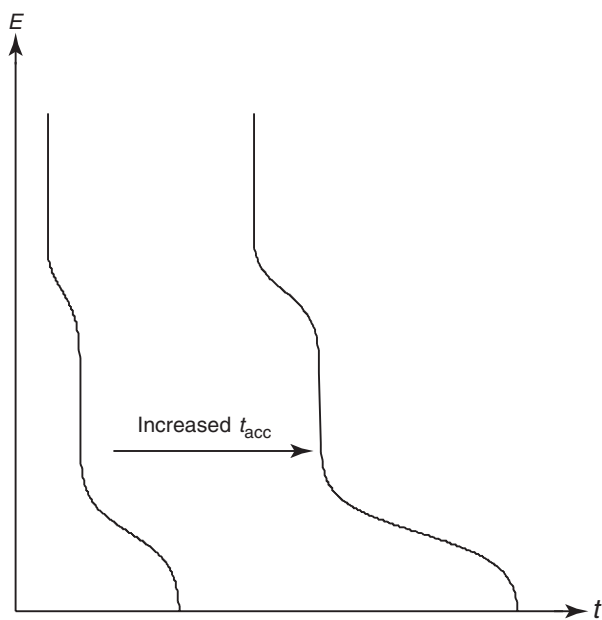
$$\Gamma_{\text{ML}_m^{n+}} \propto \frac{D_{\text{ML}_m^{n+}} [(\text{ML}_m)^{n+}] t_{\text{acc}} A}{\delta_{\text{acc}}} \quad [5]$$

where  $\Gamma_{\text{ML}_m^{n+}}$  is the surface excess of  $(\text{ML}_m)^{n+}$ , for  $t_{\text{acc}} \leq t_{\text{mon}}$ , with  $t_{\text{mon}}$  being the time required to form a monolayer and constant after that.

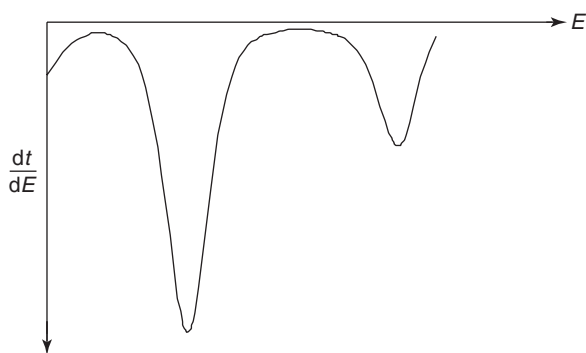
### The Stripping Step

In stripping voltammetry, the stripping step comprises the application of a potential sweep and the measurement of current versus potential. In chronopotentiometric stripping, the analyte is stripped by the application of a constant oxidative or reductive flux while the change in the potential of the working electrode versus time is measured. This is normally performed in an unstirred solution to minimize mass transport and consequently maximize the stripping time.

In the literature, the term 'potentiometric stripping analysis' is generally reserved for cases where the



**Figure 2** Increasing the accumulation time increases the stripping time.



**Figure 3** Commercial instruments display plots of  $dt/dE$  versus  $E$ , which are similar in appearance to stripping voltammograms. The area under the peak is equal to the stripping time,  $\tau$ .

constant flux is a constant chemical flux to the electrode surface and thus was originally referred to as chemical stripping. The other form of chronopotentiometric stripping that is frequently reported in the literature is referred to as constant current stripping to indicate that the constant flux is a flux of electrons.

Commercial instruments generally take the derivative of the chronopotentiogram and present the output of stripping measurements as a plot of the form  $dt/dE$  versus  $E$ . This is esthetically preferable since it produces peaks analogous to what one observes from stripping voltammograms (Figure 3). The area under each peak is the stripping time for that element.

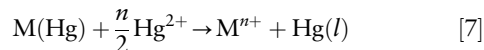
## Potentiometric Stripping Analysis (Chemical Stripping)

Chemical stripping has its origins in the 1960s but was popularized by Jagner in the 1970s and 1980s. During a chemical stripping step the electrochemical cell is switched to open circuit at the completion of the accumulation period. Chemical stripping relies on a redox reaction between the analyte and another redox active compound in solution. The stripping of amalgamated metals ions, the most common application of PSA, provides a good example. Here, the stripping compound is usually Hg(II), which is added to the cell to form the thin mercury film working electrode by electrolytic reduction, although dissolved oxygen will often suffice as the stripping oxidant. The potential of the working electrode under open circuit, in this case a thin mercury film, is given by the Nernst equation:

$$E = E^{\circ} + \frac{RT}{nF} \ln \frac{[M^{n+}]}{[M(Hg)]} \quad [6]$$

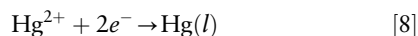
where  $[M^{n+}]$  is the concentration of the analyte in solution at the electrode surface and  $M(Hg)$  is the

concentration of the amalgamated metal. Under open circuit conditions, the amalgamated metal is oxidized by the chemical reaction with Hg(II) (or  $O_2$ ):



Therefore, the ratio  $[M^{n+}]/[M(Hg)]$  changes with time under open circuit as Hg(II) diffuses to the electrode surface and reacts chemically with the amalgamated metal and consequently the electrode potential becomes increasingly positive over time. The oxidant Hg(II) will oxidize those amalgamated metal ions that have lower (more negative) reduction potentials. For multielement analysis after the first element is oxidized, the potential of the working electrode moves rapidly positive until the oxidation of the next species begins.

It follows that the potential of the working electrode will be sufficiently negative during the stripping step for the direct reduction of Hg(II) to occur at the electrode:



A corresponding oxidation reaction will also occur at the counter electrode. This reaction contributes to the background that is measured during the stripping step.

Several oxidants could be present in solution. For example, in non-deaerated solutions dissolved oxygen can act as an oxidant along with any Hg(II) that may have been added to the solution. The total number of moles of oxidant,  $Ox$ , reaching the electrode is

$$Ox_j \propto \frac{A\tau}{\delta_{strip}} \sum D_{Ox_j} [Ox_j] \quad [9]$$

where  $[Ox_j]$  is the concentration of oxidant  $j$  in solution,  $\delta_{strip}$  is the thickness of the diffusion layer formed during the stripping process, and  $\tau$  is the transition or stripping time. The measured stripping time will correspond to the time taken for the complete re-oxidation of the amalgamated metal. This is given by combining eqns [3] and [9] and rearranging:

$$\tau \propto \frac{t_{acc} \delta_{strip}}{\delta_{acc}} \frac{D_{M^{n+}} [M^{n+}]}{\sum D_{Ox_j} [Ox_j]} \quad [10]$$

Since the absolute amount of analyte accumulated during the preconcentration process is small the concentration of oxidant(s) in solution is assumed to be constant during the stripping process. The ratio  $\delta_{strip}/\delta_{acc}$ , which is assumed to be between 5 and 10 under normal experimental conditions, can be kept constant by keeping the stirring rate during each accumulation period constant between experiments.

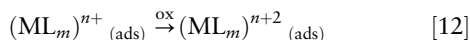
The diffusion layer during the stripping process should be constant since this is performed in unstirred solution. Therefore, the stripping time becomes dependant on the concentration of the analyte in solution and the accumulation period:

$$\tau \propto [M^{n+}] t_{acc} \quad [11]$$

Since time is more easily measured than current PSA has advantages over stripping voltammetry; however, in practice very short stripping times are often measured requiring computerized equipment with high sampling rates. Nevertheless, it is usually possible to implement PSA with reduced cost relative to stripping voltammetry.

Since dissolved oxygen can be used as an oxidant no deaeration is generally required for PSA unless one is endeavoring to minimize the total oxidant concentration in order to lower the detection limit. Dispensing with the deaeration procedure results in a large time saving per analysis relative to stripping voltammetry. On the downside, PSA stripping times are slightly dependent on the concentration of elements that are stripped before it. Also the oxidant concentration decreases during electrolysis, making it difficult to maintain a constant oxidant concentration between experiments. This can be offset by using dissolved oxygen as an oxidant and using an open cell to maintain a saturated oxygen concentration during the stirred electrolysis procedure. Other negatives such as the formation of intermetallic compounds, the influence of the mercury film history, and the overlap of the analytical signal with the background signal are common to both PSA and stripping voltammetry.

Adsorbed molecules can be chemically stripped using both oxidants and reductants depending on the nature of the analyte in question. In most systems the analyte remains strongly adsorbed in both its oxidized and reduced forms. For example, the two-electron oxidation of an adsorbed metal complex would be given by the equation



The stripping time is normally orders of magnitude less than the accumulation time when adsorptive accumulation is employed. During the accumulation period, which is normally carried out at a constant potential, the oxidant is being reduced, thereby creating a diffusion layer  $\delta_{ox}$ . When the oxidation of the adsorbate is controlled by the rate of mass transport of the oxidant the stripping time is given by

$$\tau \propto \frac{\Gamma_{\text{ML}_m^{n+}} \delta_{ox}}{[\text{Ox}] D_{ox}} \quad [13]$$

Combining eqns [5] and [13] gives a similar expression for the stripping time to that for the stripping of amalgamated metal ions:

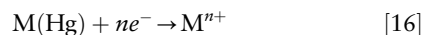
$$\tau \propto \frac{[(\text{ML}_m)^{n+}]}{[\text{Ox}]} \frac{D_{\text{ML}_m^{n+}} \delta_{ox}}{D_{ox} \delta_{acc}} t_{acc} A \quad [14]$$

If the experimental conditions are reproducible then this reduces to

$$\tau \propto [(\text{ML}_m)^{n+}] t_{acc} \quad [15]$$

## Constant Current Stripping

Constant current chronopotentiometric stripping of metal ions from amalgams has its origins in the 1950s with the theory for the stripping from a planar mercury electrode first described by Delahay. During a constant current stripping step, a constant current is applied to the electrochemical cell at the completion of the preconcentration step. For an electrolytic preconcentration, the constant current is intended to reverse the electrolytic reaction in an analogous way to the application of a potential sweep during the stripping step in stripping voltammetry. For example, an amalgamated metal would be stripped back into solution by applying a constant reducing current



This changes the ratio  $[\text{M}^{n+}]/[\text{M}(\text{Hg})]$  with time and consequently changes the electrode potential over time. The stripping time is

$$\tau \propto \frac{t_{acc}}{\delta_{acc}} \frac{D_{\text{M}^{n+}} [\text{M}^{n+}]}{i} \quad [17]$$

where  $i$  is the current applied to the cell. The stripping time can, therefore, be increased by reducing the magnitude of the current applied to the cell during the stripping step. The equation for the chronopotentiogram for both PSA and constant current stripping analysis (CCSA) stripping of amalgamated metals from a mercury film of thickness  $l$  is

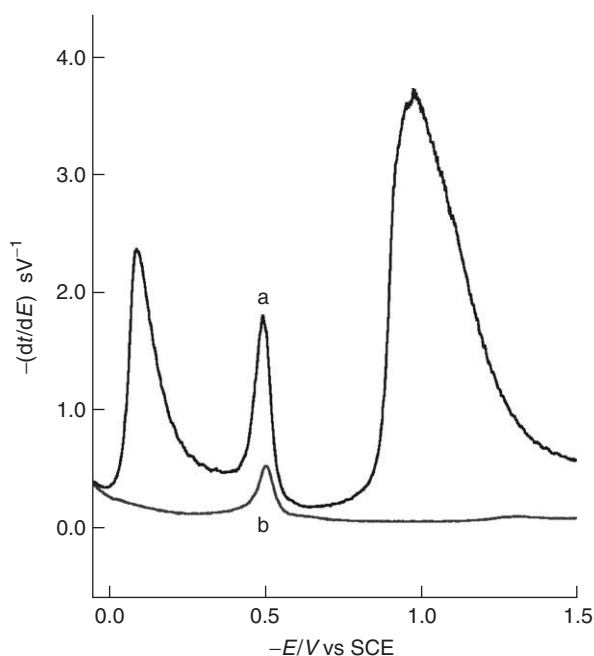
$$E = E^0 + \frac{RT}{nF} \ln \frac{2l}{(\pi D_{\text{M}^{n+}})^{1/2}} + \frac{RT}{nF} \ln \left( \frac{t^{1/2}}{\tau - t} \right) \quad [18]$$

For CCSA of adsorbed molecules the situation is relatively straightforward. The stripping time is inversely proportional to the current applied to the cell

$$\tau \propto \frac{\Gamma_{\text{ML}_m^{n+}}}{i} \quad [19]$$

Therefore,

$$\tau \propto \frac{D_{\text{ML}_m^{n+}} [(\text{ML}_m)^{n+}] t_{acc} A}{i \delta_{acc}} \quad [20]$$



**Figure 4** Experimental derivative chronopotentiograms of the reduction of mercurous cysteine thiolate in  $0.1 \text{ mol l}^{-1}$  phosphate buffer, pH 7.4: (A) before and (B) after deaeration. The presence of dissolved oxygen, which is simultaneously being reduced at the working electrode, enhances the response. The dissolved oxygen is reduced in two steps (peaks on either side of a). The peak on the left is the reduction to hydrogen peroxide and the peak on the right is the reduction of hydrogen peroxide to water. (Reproduced with permission from Honeychurch MJ and Ridd MJ (1995) The effect of a non-adsorbing electroactive species on the transition time in derivative adsorptive chronopotentiometric stripping analysis. *Electroanalysis* 7: 1047–1054; © Wiley-VCH.)

The equation for the chronopotentiogram for both PSA and CCSA of adsorbed molecules is

$$E = E^{\circ} + \frac{RT}{nF} \ln \frac{\tau - t}{t} \quad [21]$$

An interesting effect is obtained when stripping an adsorbed analyte if an electroactive species in solution is simultaneously being oxidized or reduced. For example, if an adsorbed molecule was being reduced with a constant current in the presence of dissolved oxygen, a part of the current applied to the cell goes into reducing the dissolved oxygen. Effectively, this means that less current is available to reduce the adsorbate and consequently, from eqn [19], an increased stripping time is observed. An example is shown in Figure 4. This enhancement of the stripping time can be quantified.

## Applications

The applications of chronopotentiometric stripping are identical to those of stripping voltammetry. For

example, elements such as Bi, Cd, Cu, Ga, In, Pb, Mn, Tl, Zn, Co, Ni, Fe, and Se have been determined by PSA or CCSA at mercury film electrodes. Additional elements such as Na and K can be determined from aprotic solvents. Other elements such as Hg, Ag, As, Sb, and Sn can be determined at gold electrodes. Detection limits below the microgram per liter level are generally obtained following an appropriate preconcentration period. Many stripping voltammetric methods using adsorptive preconcentration to quantify metal ion complexes or organic molecules have been repeated using PSA or CCSA.

While the applications of chronopotentiometric stripping are identical to stripping voltammetry, chronopotentiometric methods may be preferable for certain sample types. The current widespread use of PSA and derivative PSA is mainly due to the efforts of Jagner and co-workers, which centered on the PSA determination of metal ions following their potentiostatic deposition at a mercury thin film electrode followed by a chemical oxidation with either dissolved oxygen or mercury (II). Much of this early work focused on analysis of metal ions in what would normally be considered difficult sample matrix such as wine and fruit juice. PSA is suitable for low ionic strength and low conductivity solutions and, when employing electrolytic preconcentration, is less affected by the presence of organic compounds in the sample than in stripping voltammetry or CCSA. Thus, PSA is preferable in situations such as metal analysis in the field where sample pretreatment such as deoxygenation or removal of organic matter is necessary to improve accuracy and lower detection limits or for metal analysis in samples containing large amounts of organic material such as wine or blood.

**See also:** Voltammetry: Anodic Stripping; Cathodic Stripping.

## Further Reading

- Chau TC, Li YL, and Wu YL (1982) Studies on potentiometric stripping analysis. *Talanta* 29: 1083–1087.
- Estela JM, Tomas C, Cladera A, and Cerda V (1995) Potentiometric stripping analysis: a review. *Critical Reviews in Analytical Chemistry* 25: 91–141.
- Honeychurch MJ and Ridd MJ (1996) The potentiometric stripping analysis of proteins. *Electroanalysis* 8: 654–660.
- Hoyer B and Kryger L (1985) A theoretical and experimental study of the waveform in potentiometric stripping analysis with a rotating mercury film electrode – the reversible case. *Talanta* 32: 839–844.
- Hussam A and Coetzee JF (1985) Potentiometric stripping analysis: theory, experimental verification, and generation of stripping polarograms. *Analytical Chemistry* 57: 581–585.



- Jagner D (1982) Potentiometric stripping analysis. *The Analyst* 107: 593–599.
- Jagner D and Aren K (1982) Application of potentiometric stripping analysis to complexation titrations. *Analytica Chimica Acta* 134: 201–209.
- Kolev SD, Chow CWK, Davey DE, and Mulcahy DE (1998) Mathematical modelling of potentiometric stripping analysis. Chemical stripping in quiet solutions. *Analytica Chimica Acta* 377: 13–19.
- Labar C and Lamberts L (1981) Potentiometric stripping analysis at a stationary electrode. *Analytica Chimica Acta* 132: 23–33.
- Ostapczuk P (1993) Present potentials and limitations in the determination of trace elements by potentiometric stripping analysis. *Analytica Chimica Acta* 273: 35–40.
- Ruan X and Chang H (1988) Reciprocal derivative constant current stripping analysis. *Talanta* 35: 861–867.

## PRECIOUS METALS

**Y Hayashibe**, Mitsubishi Materials Corp, Saitama, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The platinum group metals (Ru, Rh, Pd, Os, Ir, and Pt), Ag, and Au are called ‘precious’ or ‘noble’ metals. Nobility and catalytic activity are unique properties of precious metals, that result in a wide range of applications, such as catalysts in various industrial fields, in electronic industries, and in jewelry. The chemical and physical properties of each precious metal are shown in Table 1. The determination of precious metals attracted the interest of analysts and developed rapidly because these metals are valuable and rare, and also very important for many products. Their concentration levels are very low in many natural sources, metallurgical intermediates, and environmental samples. Furthermore, precious metals are collectively handled in the analytical chemistry field, because of the close resemblance of their chemical properties and behavior. Precious metals are the subproducts in copper, zinc, or lead smelting and refining, which is the most important source of precious metals. Whereas many analytical methods for the ultratrace determination of precious metals in environmental or biological samples were recently published with the development of high-sensitivity analytical instruments, the classical fire-assay has been widely applied for the accurate determination of expensive precious metals.

### International Standard Method

The international standard method for the determination of precious metals has been mainly discussed in ISO/TC183 (copper, lead, and zinc ores and concentrates) and ISO/TC174 (Jewelry). Table 2 shows

the ISO international standard methods published or discussed. Accurate sampling methods, for example, fire assay, titrimetric method, and/or gravimetric method are regulated in almost all international standards for precise and true determination, because of international trading of expensive precious metals. Another method for the determination of precious metals content is by subtraction of the total content of impurities in the sample from 100%, which is also regulated. In these cases, atomic absorption spectrometry (AAS) and/or inductively coupled plasma-atomic emission spectrometry (ICP-AES) is utilized for the final detection.

### Sample Decomposition (Sample Preparation)

Sample decomposition for the determination of precious metals is very difficult, because of their high chemical stability and low content in samples.

Fire assay is the classical sample decomposition method. This method is also used due to its merit of accuracy, when collecting the precious metals from various sulfide ores, or when collecting the precious metals from a complicated matrix. Although lead fire assay is used for the collection of Au, Ag, Pt, Pd, Rh, etc., Os is volatilized by the lead fire assay method. Moreover, the recovery of Ru and Ir is poor. The nickel sulfide fire assay is used for a large number of precious metals, but the recovery of Au is quite low.

Chlorination technique has been developed to make the precious metals water soluble; that is, effective for a low concentration sample, alloy, or sulfide mineral. Chlorination consists of both dry and wet methods. Chlorination gives a low reagent blank, and is used in combination with high-sensitivity detection. In the wet chlorination method, the samples are pressurized in a closed vessel with an oxidizing agent and hydrochloric acid, and dissolved

- Jagner D (1982) Potentiometric stripping analysis. *The Analyst* 107: 593–599.
- Jagner D and Aren K (1982) Application of potentiometric stripping analysis to complexation titrations. *Analytica Chimica Acta* 134: 201–209.
- Kolev SD, Chow CWK, Davey DE, and Mulcahy DE (1998) Mathematical modelling of potentiometric stripping analysis. Chemical stripping in quiet solutions. *Analytica Chimica Acta* 377: 13–19.
- Labar C and Lamberts L (1981) Potentiometric stripping analysis at a stationary electrode. *Analytica Chimica Acta* 132: 23–33.
- Ostapczuk P (1993) Present potentials and limitations in the determination of trace elements by potentiometric stripping analysis. *Analytica Chimica Acta* 273: 35–40.
- Ruan X and Chang H (1988) Reciprocal derivative constant current stripping analysis. *Talanta* 35: 861–867.

## PRECIOUS METALS

**Y Hayashibe**, Mitsubishi Materials Corp, Saitama, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The platinum group metals (Ru, Rh, Pd, Os, Ir, and Pt), Ag, and Au are called ‘precious’ or ‘noble’ metals. Nobility and catalytic activity are unique properties of precious metals, that result in a wide range of applications, such as catalysts in various industrial fields, in electronic industries, and in jewelry. The chemical and physical properties of each precious metal are shown in Table 1. The determination of precious metals attracted the interest of analysts and developed rapidly because these metals are valuable and rare, and also very important for many products. Their concentration levels are very low in many natural sources, metallurgical intermediates, and environmental samples. Furthermore, precious metals are collectively handled in the analytical chemistry field, because of the close resemblance of their chemical properties and behavior. Precious metals are the subproducts in copper, zinc, or lead smelting and refining, which is the most important source of precious metals. Whereas many analytical methods for the ultratrace determination of precious metals in environmental or biological samples were recently published with the development of high-sensitivity analytical instruments, the classical fire-assay has been widely applied for the accurate determination of expensive precious metals.

### International Standard Method

The international standard method for the determination of precious metals has been mainly discussed in ISO/TC183 (copper, lead, and zinc ores and concentrates) and ISO/TC174 (Jewelry). Table 2 shows

the ISO international standard methods published or discussed. Accurate sampling methods, for example, fire assay, titrimetric method, and/or gravimetric method are regulated in almost all international standards for precise and true determination, because of international trading of expensive precious metals. Another method for the determination of precious metals content is by subtraction of the total content of impurities in the sample from 100%, which is also regulated. In these cases, atomic absorption spectrometry (AAS) and/or inductively coupled plasma-atomic emission spectrometry (ICP-AES) is utilized for the final detection.

### Sample Decomposition (Sample Preparation)

Sample decomposition for the determination of precious metals is very difficult, because of their high chemical stability and low content in samples.

Fire assay is the classical sample decomposition method. This method is also used due to its merit of accuracy, when collecting the precious metals from various sulfide ores, or when collecting the precious metals from a complicated matrix. Although lead fire assay is used for the collection of Au, Ag, Pt, Pd, Rh, etc., Os is volatilized by the lead fire assay method. Moreover, the recovery of Ru and Ir is poor. The nickel sulfide fire assay is used for a large number of precious metals, but the recovery of Au is quite low.

Chlorination technique has been developed to make the precious metals water soluble; that is, effective for a low concentration sample, alloy, or sulfide mineral. Chlorination consists of both dry and wet methods. Chlorination gives a low reagent blank, and is used in combination with high-sensitivity detection. In the wet chlorination method, the samples are pressurized in a closed vessel with an oxidizing agent and hydrochloric acid, and dissolved

**Table 1** Physical and chemical properties of precious metals

<i>Properties</i>	<i>Au</i>	<i>Ag</i>	<i>Pt</i>	<i>Pd</i>	<i>Rh</i>	<i>Ru</i>	<i>Ir</i>	<i>Os</i>
Atomic number	79	47	78	46	45	44	77	76
Relative atomic mass	196.96	107.90	195.09	106.42	102.90	101.07	192.22	190.23
Isotopic percentage	<sup>197</sup> Au (100%)	<sup>107</sup> Ag (51.8%) <sup>109</sup> Ag (48.2%)	<sup>195</sup> Pt (33.8%) <sup>194</sup> Pt (32.9%) <sup>196</sup> Pt (25.2%) <sup>198</sup> Pt (7.19%) <sup>192</sup> Pt (0.78%) <sup>190</sup> Pt (0.0127%)	<sup>106</sup> Pd (27.3%), <sup>108</sup> Pd (26.5%) <sup>105</sup> Pd (22.3%) <sup>110</sup> Pd (11.7%) <sup>104</sup> Pd (11.2%) <sup>102</sup> Pd (1.0%)	<sup>103</sup> Rh (100%)	<sup>102</sup> Ru (31.6%) <sup>104</sup> Ru (18.7%) <sup>101</sup> Ru (17.1%) <sup>99</sup> Ru (12.7%) <sup>100</sup> Ru (12.6%) <sup>96</sup> Ru (5.5%) <sup>98</sup> Ru (1.8%)	<sup>191</sup> Ir (37.3%) <sup>193</sup> Ir (62.7%)	<sup>192</sup> Os (41.0%) <sup>190</sup> Os (26.4%) <sup>189</sup> Os (16.1%) <sup>188</sup> Os (13.3%) <sup>187</sup> Os (1.6%) <sup>186</sup> Os (1.6%)
Color	Yellow	Bright white	White gray	White shiny	Silvery white	Dark gray	Silver white	Bluish-white
Luster	(Strong)							
Malleable				Easily forged				
Ductile								
Electronic configuration	(Xe)5f <sup>10</sup> 6s <sup>1</sup>	(Kr)4d <sup>10</sup> 5s	(Xe)5d <sup>9</sup> 6s <sup>1</sup>	(Kr)4d <sup>10</sup>	(Kr)4d <sup>8</sup> 5s <sup>1</sup>	(Kr)4d <sup>7</sup> 5s <sup>1</sup>	(Xe)5d <sup>9</sup>	(Xe)4f <sup>14</sup> 5d <sup>9</sup> 6s <sup>2</sup>
Melting point (°C)	1063	961.9	1769	1552	1970	2310	2447	3327
Boiling point (°C)	2970	2212	4530	2940	4500	4900	4428	5300
Heat of fusion (kJ mol <sup>-1</sup> )	12.8	11.1	19.7	17.6	21.6	25.5	26.4	31.7
Heat of vaporization (kJ mol <sup>-1</sup> )	343	258	469	362	494		612	738
Enthalpy of formation (kJ mol <sup>-1</sup> )	379	284	545	377	556	640	669	791
Ionization energy (eV)	9.22	7.574	8.62	8.33	7.46	7.364	9	8.7
Electron affinity (eV)	2.31	1.30	2.13	0.6	1.2	1.1	1.6	1.1
Pauling's electronegativity	2.4	1.9	2.2	2.2	2.2	2.2	2.2	2.2
Density (20°C) (g cm <sup>-3</sup> )	19.3	10.49	21.4	11.99	12.4	12.41	22.5	22.57
Crystal structure	Face central cubic	Face central cubic	Face central cubic	Face central cubic	Face central cubic	Hexagonal	Face central cubic	Closed packed hexagonal structure

Electric resistance (20°C) ( $\mu\Omega\text{ cm}$ )	2.20	1.61	9.85	9.93		7.4	4.71	9i
Thermal conductivity ( $\text{W m}^{-1}\text{ K}^{-1}$ ) (27°C)	315	427	71.4	75.5	150	117	147	87.6
Young's modulus (GPa)	79.5	73.2	168	126	380	420	514	550
Compressibility ( $\text{GPa}^{-1}$ ) (30°C)	0.00588	0.0101	0.00367	0.00538	0.00368	0.00349	0.00273	
Resistivity ( $\mu\Omega\text{ cm}$ ) (20°C)	2.35	1.59	9.85	9.93	4.33	6.71	4.71	8.12
Magnetic susceptibility ( $\text{mm}^3\text{ kg}^{-1}$ )	− 1.8	2.4	+ 12		+ 13.6	+ 5.4	1.6	+ 0.6
Mohs hardness number	2.5		4.3		6	6.5		7.5
Oxidation state (bold: stable)	0, + <b>1</b> , + <b>3</b>	0, + <b>1</b> , + 2, + 3	0, + 1, + <b>2</b> , + <b>4</b> , + 5, + 6	0, + <b>2</b> , + 4	− 1, 0, + 1, + 2, + <b>3</b> , + 4, + 5, + 6	0, + 1, + 2, + <b>3</b> , + 4, + 5, + 6, + 7, + 8 + 0.455 or + 0.249	+ 1, + 2, + <b>3</b> , + 4, + 5, + 6	0, + 1, + 2, + 3, + <b>4</b> , + 5, + <b>6</b> , + 7, + <b>8</b> + 0.687
Oxidation–reduction potential (V)	+ 1.691 ( $\text{Au}^+ + \text{e}^- = \text{Au}$ ) + 1.50 ( $\text{Au}^{3+} + 2\text{e}^- = \text{Au}^-$ )	+ 0.80 ( $\text{Ag}^+ + \text{e}^- = \text{Ag}$ ) + 1.98 ( $\text{Ag}^{2+} + \text{e}^- = \text{Ag}^+$ )	+ 0.73	0.99	0.758		+ 1.156	
Concentration of earth's crust (mg per kg)	0.004	0.08	0.01	0.0015	0.0001	0.0001	0.001	0.005
Bond energy ( $E_b$ , eV)	84.0	368.2	71.1	335.0	307.2	280.0	60.8	50.7

**Table 2** International standard methods for the determination of precious metals

<i>ISO no.</i>	<i>Title</i>
ISO 9202:1991	Fineness of precious metal alloys
ISO 10378:1994	Copper sulfide concentrates – determination of gold and silver contents – fire assay gravimetric and flame atomic absorption spectrometric method
ISO 11210:1995	Determination of platinum in platinum jewelry alloys – gravimetric method after precipitation of diammonium hexachloroplatinate
ISO 11426:1997	Determination of gold in gold jewelry alloys – Cupellation method (fire assay)
ISO 11427:1993	Determination of silver jewelry alloys – volumetric (potentiometric) method using potassium bromide
ISO 11489:1995	Determination of platinum in platinum jewelry alloys – gravimetric determination by reduction with mercury(I) chloride
ISO 11490:1995	Determination of palladium in palladium jewelry alloys – gravimetric determination with dimethylglyoxime
ISO 12740:1998	Lead sulfide concentrates – determination of silver and gold contents Fire assay and flame atomic absorption spectrometric method using scorification or cupellation
ISO 12743:1998	Copper, lead, and zinc sulfide concentrates – sampling procedures for determination of metal and moisture content
ISO 13756:1997	Determination of silver in silver jewelry alloys – volumetric (potentiometric) method using sodium chloride or potassium chloride
ISO 15247:1999	Zinc sulfide concentrates – determination of silver content – acid dissolution and flame atomic absorption spectrometric method
ISO 15248:1998	Zinc sulfide concentrates – determination of silver and gold contents fire assay and flame atomic absorption spectrometric method using scorification or cupellation
ISO 15249:1998	Zinc sulfide concentrates – determination of gold content acid dissolution/solvent extraction/flame atomic absorption spectrometric method

in hydrochloric acid. The advantage of this method is that the recovery of precious metals is quantitative without losing Ru and Os.

The acid decomposing method is widely used in many sample treatments. Generally, according to the matrices of a sample, a mineral acid or its mixture is used in a vessel that is either an open or a closed system. A hot plate system and/or a microwave oven are commonly utilized for sample digestion. In the case of an open vessel, the recovery of Os is poor because of volatilization.

Alkali fusion is used as a digestion technique for putting a persistent sample into solution. Sodium peroxide is mainly used as the flux for analyzing the precious metals. However, in the case of direct fusion of a metal or a sulfide sample, preliminary oxidization is necessary to perform fusion completely and safely.

## Separation and Concentration

Solvent extraction, solid–liquid sorption, ion exchange, and precipitation methods are commonly used for the separation or preconcentration of precious metals from various matrices.

### Fire Assay

The fire-assay method described above has some advantages as the separation and preconcentration method of precious metals. It is useful where large samples are required owing to their extremely low

concentrations in some materials and the heterogeneity of ores containing these metals. Lead, copper, copper–iron–nickel alloy and nickel sulfide have been used as collectors. Lead fire assay is an effective and widely used technique for the preconcentration of Au, Ag, Pt, and Pd when the concentrations of the last two elements are low. Nickel sulfide fire assay is very effective for the collection of precious metals. By reducing the button size to below 1 g, the precious metals are successfully determined directly in the button with inductively couple plasma-mass spectrometry (ICP-MS). For understanding the effects of fusion charge composition on the collection of precious metals into a minimized NiS button, the extraction ratio  $E$ , is expressed as  $E = (m_{\text{sample}} + m_{\text{flux}}) / (m_{\text{Ni}} + m_{\text{S}})$ , where  $m$  is the mass of the component indicated by the subscript.  $E > 30$  only due to the effect of the sample composition, particularly its sulfur content. Antimony fire assay was derived prior to concentrate precious metals from samples with simpler compositions. The separation of precious metals, including Os, from antimony can be achieved by cupellation.

### Gold

Au can be readily precipitated from extremely dilute, acid solutions by such reducing agents as tin(II) chloride, sulfurous acid plus hydrazine hydrochloride, Zn, and Mg if a suitable collector is used. Tellurium is very useful as a collector. The precipitate of

elemental tellurium, produced by reduction with tin(II) chloride, contains Au as telluride. If tellurium must be removed, the precipitate can be ignited strongly to volatilize tellurium oxide. Pd, Pt, Hg, and As will be co-precipitated with Au. Au can also be precipitated with various organic reducing agents such as Metol, *p*-phenylene diamine, and hydroquinone in hot 1 mol l<sup>-1</sup> HCl. Pd, Pt, Te, and Se are not precipitated under these conditions, and Pd and Pt at least show less tendency to contaminate in the precipitation of Au. One microgram of Au can be precipitated quantitatively with *p*-dimethylaminobenzylidenerhodanine or 2-mercaptobenzothiazole from 50 ml of dilute aqua regia or dilute HNO<sub>3</sub> solution. Ag, Hg, and Pd are also precipitated nearly quantitatively. Au(III) as chloroaurate is strongly adsorbed on a strongly basic anion-exchange resin from HCl solutions at any acid concentration. This behavior has been utilized for the separation of Au in seawater and of traces of Au in copper. Because a suitable eluant for Au has not been found, the resin is ignited in an alumina crucible at as low a temperature as possible. Au(III) has been separated from Ni, Cu, and Fe(III) by adsorption of Au on Dowex 1-X8 from 0.5 mol l<sup>-1</sup> HCl and elution with 0.3 mol l<sup>-1</sup> thiourea in 0.1 mol l<sup>-1</sup> HCl at 50°C. Au and Pt metals can be separated from base metal (Fe(III), Cu, Ni) by adsorption of the base metals on strongly acidic cation-exchange resin, Dowex 50W-X8, from diluted HCl solution of pH 1.0–1.5. A medium of 0.1 mol l<sup>-1</sup> HCl in 60% (v/v) acetone serves the same purpose. Au can be separated from Pt, Ir, Pd, and Rh by adsorption of Au on Dowex 50W-X8 from 6 mol l<sup>-1</sup> HBr – 0.0035 mol l<sup>-1</sup> bromine. Solvent extraction is one of the important techniques for the separation and preconcentration of precious metals. Many common extractants such as 4-methyl-2-pentanone (IBMK), tributyl phosphate (TBP), and trialkylphosphine oxide (TOPO) are still widely used. The extractability of chloroauric acid from hydrochloric acid medium, and of bromoauric acid from hydrobromic acid, by ether and ethyl acetate provides not only a method for isolating small amounts of Au but also a method for separating it from most of the precious group metals as well as other elements. The extraction coefficient (ether/aqueous phase) has a value of ~50 for 3–6 mol l<sup>-1</sup> HCl with macro concentrations of Au. Bromoauric acid is a better extraction form for Au than chloroauric acid. Isopropyl ether is a more satisfactory solvent than ethyl ether, ethyl acetate, or IBMK. Clark *et al.* reported an IBMK–amine synergistic iodide complex (MAGIC) extraction system for use in geological exploration samples. The method involves extraction of halides, mostly iodide,

complexes of metals into an organic phase that contains Aliquant-336, tricaprylyl tertiary amine, and IBMK. The extraction was carried out in HCl between 2.25 and 4.2 mol l<sup>-1</sup> in the presence of MAGIC salt solution composed of potassium iodide, bromide, chloride, and L-ascorbic acid. Precious metals containing Au, Pd, and Pt can be quantitatively extracted into organic solution. The organic extract can be analyzed directly by AAS or ICP-AES. For samples with high Fe content, extraction of iodide or bromide complexes into IBMK as ion associates with long-chain aliphatic polyamines is a promising method for multielement concentration.

### Silver

Precipitation of Ag as chloride, or better as bromide, with a slight excess of precipitant may be a useful method of separation. Traces of Ag have been separated by thallium iodide as the collector. The precipitation can be dissolved in diluted H<sub>2</sub>SO<sub>4</sub>–H<sub>2</sub>O<sub>2</sub> solution. A widely applicable method for the isolation of traces of Ag consists in co-precipitation with elemental tellurium formed by adding tin(II) chloride to 2 mol l<sup>-1</sup> HCl solution of the sample containing Te(IV) (see above). The extraction constant of primary Ag dithizonate is such that Ag can be extracted quantitatively from 6 mol l<sup>-1</sup> H<sup>+</sup> solution by the use of 0.001% (w/v) dithizone solution in carbon tetrachloride. Anions such as halides and thiocyanates that form slightly soluble or complex Ag salts must be absent when extraction is made in mineral acid medium. Other metals extracted at high acidity are Pd, Au, and Hg. Copper will also be extracted.

### Platinum

Minute amounts of Pt can be separated from almost all of the non-noble elements by precipitation with tin(II) chloride from HCl solution containing a small amount (< 1 mg) of tellurium(IV). Au, Ag, Pd, Rh accompany Pt, the first three usually quantitatively, and Rh incompletely but in large amount. Ir appears to be incompletely precipitated. Milligram amounts of Pt, Pd, Rh, and Ir can be separated from base metals by complexing the Pt metals with NaNO<sub>2</sub> and precipitating the latter at pH 7. The separation of Au from Pt can be effected by precipitating Au with a reagent such as oxalic acid, sulfur dioxide, nitrous acid, or hydroquinone, which do not reduce Pt to the metals, but a reprecipitation is required. Ag may be separated by precipitation as chloride with slight excess of HCl if the precipitate is dissolved in ammonia and again thrown down by acidification to release co-precipitated Pt. Pt will be present as ammine complexes in the filtrate from the reprecipitated



AgCl; in order to convert it to chloroplatinic acid, it is necessary to evaporate the solution, destroy  $\text{NH}_4\text{NO}_3$  with HCl and  $\text{HNO}_3$ , evaporate to fumes with  $\text{H}_2\text{SO}_4$ , and after eliminating most of the latter, treating with dilute aqua regia.

Anion-exchange chromatography of chloro-complexes of the Pt metals is especially useful. After adsorption on a strongly basic anion-exchange resin Amberlite IRA-400, Rh is eluted with  $2 \text{ mol l}^{-1}$  HCl, Pd with  $9 \text{ mol l}^{-1}$  HCl, and Pt with  $2.4 \text{ mol l}^{-1}$   $\text{HClO}_4$ . The Pt metals plus Au can be separated from Fe, Co, Ni, and Cu by extracting the noble metals with 4-octylaniline in diisobutyl ketone or the other organic solvents from dilute ( $3 \text{ mol l}^{-1}$ ) HCl medium. The Pt metals are back-extracted with  $7 \text{ mol l}^{-1}$   $\text{HClO}_4$ , but Au is not back-extracted. Pt and Pd also can be separated from Rh and Ir by extracting the former pair with sodium diethyldithiocarbamate and chloroform or IBMK from HCl–KI solution. Ethyl ether extraction of Fe(III), Au(III), and other metals from HCl solution may be of value at times for removal of large amounts of these from Pt(IV).

### Palladium

Minute amounts of Pd can be co-precipitated with elemental tellurium obtained by reduction of a tellurite with tin(II) chloride or similar strong reducing agents. The slight solubility of Pd(II) dimethylglyoximate and other dioximates in dilute HCl solution enables fairly small amounts of Pd to be separated from most other metals except Au (Pt tends to contaminate Pd dimethylglyoximate). Such precipitation fails at the low Pd concentrations frequently encountered, but  $<1 \text{ mg l}^{-1}$  of Pd can be co-precipitated with Ni dimethylglyoximate from a solution (pH 6.5) containing tartrate,  $(\text{NH}_3\text{OH})\text{Cl}$ , and acetate. Pd can be separated from large amounts of Fe, Ni, and Cu on a strongly basic anion-exchange resin at pH  $\sim 1.5$ ; the chloro complexes of Pd, Pt, Rh, and Ir pass through the column. Chloroform extraction of Pd(II) dimethylglyoximate provides the most selective separation of Pd from other elements. This extraction has not been systematically studied, but it is known that Pd can be quantitatively extracted from  $0.2\text{--}0.3 \text{ mol l}^{-1}$  HCl or  $\sim 0.5 \text{ mol l}^{-1}$   $\text{H}_2\text{SO}_4$  solutions. Fe(III) is reported to hinder the extraction of Pd. Extraction with IBMK from  $\sim 7.5 \text{ mol l}^{-1}$  HCl removes Fe(III), Mo(VI), etc.

### Rhodium

In the systematic macroanalysis of the platinum group metals, Os and Ru are first separated by volatilization of the tetraoxides, and Pd, Rh, and Ir are then precipitated as hydrous oxides in the

presence of bromate by adding  $\text{NaHCO}_3$ . The hydrolytic precipitation of Rh is complete at pH 6. A reprecipitation of the hydrous rhodium oxide is required to free it from Pt. Palladium is removed as the dimethylglyoximate, and Rh is then precipitated as metal, and thus separated from Ir, by reduction with  $\text{TiCl}_3$  in hot  $\text{H}_2\text{SO}_4$  solution. The Rh precipitate must be dissolved and reprecipitated to eliminate the small amount of Ir that has been carried down. The microgram quantities of Rh can be separated from  $2 \text{ mol l}^{-1}$  HCl by reduction with  $\text{TiCl}_3$  in the presence of Pd as collector. Elements precipitated with Rh include Pt, Ir (at least partly), Au, Ag, Te, Se, and Hg. Rh can be separated from Ir in HCl or  $\text{H}_2\text{SO}_4$  solution by reducing the former to the metal by boiling with Sb or Cu powder. Pt, Pd, and Rh (plus Ru and Ir) have been separated from each other by using organic co-precipitants including  $\text{SCN}^-$ -diantipyrinylmethane. A useful separation of Rh and other Pt metals from base metals (Cu, Fe, Ni) is based on the adsorption of the base metals on the cation-exchange resin Dowex 50W-X8 from HCl solution of pH 1.0–1.5 and nonadsorption of the Pt metals. Rh(III) can be extracted into 1-butyl-3-methylpyrazole in toluene from  $0.01 \text{ mol l}^{-1}$  HCl medium and separated from Pt(IV), Ir(III,IV), and many base metals including Fe(III), Cu, and Ni. Pd(II) is also extracted and Au(III) is partially extracted. Pt and Pd also can be separated from Rh and Ir by extracting the former pair with Na diethyldithiocarbamate and chloroform from HCl–KI solutions, with IBMK from HCl–KI solution, or with trioctylammonium chloride in toluene from  $\text{NaNO}_2$  solution. Pt and Pd are also extracted with trioctylamine in benzene from  $4\text{--}6 \text{ mol l}^{-1}$  HCl; Rh remains in the aqueous phase.

### Ruthenium

Ru in  $0.2 \text{ mol l}^{-1}$  HCl has been co-precipitated with  $\text{Re}_2\text{S}_7$  by  $\text{H}_2\text{S}$  under pressure. Ag has been separated from Ru, Rh, Pd by precipitation as AgCl (purified by dissolution in aqueous ammonia and reprecipitation). Ru can be separated from base metals (Cu, Fe, and Ni) by cation-exchange resin using pH 1.0 HCl. Ru(III) can be separated from Rh(III) by adsorption of Ru thiocyanate on polyurethane foam. The isolation and separation of Ru can be simplified by extracting  $\text{RuO}_4$  with an immiscible organic solvent instead of volatilizing it.  $\text{RuO}_4$  is soluble in inert solvents such as chloroform and carbon tetrachloride, which will extract very few inorganic substances under any conditions and still fewer from an  $\text{H}_2\text{SO}_4$  or  $\text{HNO}_3$  solution. The main problems arising in the analytical application of this extraction concern the oxidation of Ru to the +8 state and return of Ru

from carbon tetrachloride to an aqueous phase. AgO is very suitable for oxidation of Ru(III or IV) to Ru(VIII) in  $\text{H}_2\text{SO}_4$  or  $\text{HNO}_3$  medium. Ru can be removed quantitatively from the carbon tetrachloride phase by shaking with an aqueous solution of  $\text{SO}_2$ . The back-extraction is slow and 2 h of shaking is required for submicrogram quantities of Ru. Os can be removed by carbon tetrachloride extraction of  $\text{OsO}_4$  before oxidation of Ru to  $\text{RuO}_4$ . Iron(III) can be removed by extraction with ethyl ether from  $6 \text{ mol l}^{-1}$  HCl and with isopropyl ether from  $8 \text{ mol l}^{-1}$  HCl. Ru can be isolated by oxidizing to  $\text{RuO}_4$  in acid solution and volatilizing (distilling) this compound (b.p.:  $130^\circ\text{C}$ ). The distillation effects a separation from almost all other metals except Os. A strong oxidizing agent is needed to convert Ru in the +3 or +4 state to the +8 state. Hot perchloric acid is one of the good oxidizing agents. Os can be separated from Ru by boiling a  $5 \text{ mol l}^{-1}$   $\text{HNO}_3$  solution after preliminary oxidation with  $\text{KMnO}_4$ . Only Os is volatilized as  $\text{OsO}_4$ . Nitric acid is then eliminated by evaporation of the solution to dryness followed by evaporation several times with HCl. Finally, chloride is removed by fuming with  $\text{H}_2\text{SO}_4$ , after which distillation of Ru as  $\text{RuO}_4$  can be carried out.

### Iridium

In a standard method for separating the Pt metals from each other, Ir is precipitated as the hydrous dioxide by adding  $\text{NaHCO}_3$  to a boiling acid solution containing bromate until a pH of 6 is reached. Pd and Rh are also precipitated, but not Pt(IV); Os and Ru must be removed from the solution by volatilization of tetraoxides before the hydrolytic precipitation is made. Palladium is then precipitated from the (acid) solution of the hydrous oxide precipitate as dimethylglyoximate followed by precipitation of Rh with  $\text{TiCl}_3$  in the filtrate. Some Ir is carried down by metallic Rh, and the precipitate must be dissolved in hot concentrated  $\text{H}_2\text{SO}_4$  and the precipitation repeated. Ir is thus left with Ti, which is precipitated with cupferron (reprecipitation is necessary). A fairly satisfactory separation of microgram quantities of Ir from Rh can be obtained by boiling  $\text{H}_2\text{SO}_4$  solution with Sb powder; Rh is precipitated, whereas Ir remains completely in solution. Freshly reduced Cu powder also can be used as a precipitant (reductant) for Rh in  $1 \text{ mol l}^{-1}$  HCl solution; Ir remains quantitatively in solution. Ir, Pt, Pd, and Rh can be precipitated by thiourea and they can be separated from Cu, Fe, Ni, Pd, Pt(IV), Ir(IV), and Rh(III) as chloro complexes can be separated from base metals (Cu, Fe(III), Ni, etc.) by use of the cation-exchange resin Dowex 50W-X8. The base metals are adsorbed from

HCl solutions of pH 1.3–1.5, whereas the Pt metals are not. Ir and Rh can be separable from each other in thiourea solution by a cation exchanger. An anion-exchange method for separating Rh from Ir is based on the strong affinity of chloroiridate(IV) for the anion exchanger in contrast to the relatively weak affinity of chlororhodate(III). The resin is previously washed with water and Ce(IV) solution in  $0.8 \text{ mol l}^{-1}$  HCl. After adsorption from  $0.8 \text{ mol l}^{-1}$  HCl solution, Rh is eluted with  $0.8 \text{ mol l}^{-1}$  HCl containing Ce(IV) to retain Ir(IV). Ir is removed from the resin in a Soxhlet extractor with  $6 \text{ mol l}^{-1}$  HCl. Ir can be separated from Rh by TBP extraction of Ir(IV) from  $\sim 6 \text{ mol l}^{-1}$  HCl.

### Osmium

Losses of Os can occur during evaporation of hydrochloric acid solution of chloroosmate(IV) and during fusion of geomaterials. Os(IV) can be isolated by hydrous precipitation of the hydrous oxide. A 10%  $\text{NaHCO}_3$  solution is added to the boiling solution until a precipitate appears and suddenly coagulates. Os(IV) can be separated from base metals by adsorption of the latter on a strong-acid cation-exchange resin (Dowex 50W-X8) from an HCl solution of pH 0.8–1.5. Os(VIII) oxide can be extracted from aqueous medium by carbon tetrachloride and chloroform and also distilled (see Ru section). The distillate is collected in  $6 \text{ mol l}^{-1}$  HCl saturated with  $\text{SO}_2$ .

## Detection Method for Each Precious Metal

### Gold

Au content in sulfide ores is determined by fire assay using scorification or cupellation. Crucible fusion of the samples mixed with a litharge-based flux that, under reducing conditions, collects the precious metals in a metallic lead button. Oxidizing fusion helps the base metals present in the lead button to be substantially separated from the precious metals. Cupellation produces a bead largely comprising an Ag–Au alloy with small quantities of other metals. Au is separated from the primary bead by treatment with nitric acid. The Au prill is weighed. Au prills with a mass less than  $50 \mu\text{g}$  are dissolved in aqua regia and the Au is determined by AAS. Ag is determined in the parting solution by AAS. All residues are retreated to maximize the recovery of Au and Ag. The addition of collectors for either Au or Ag is not required, as both metals are present in sufficient amounts to be readily visible after the cupellation stage. The second bead is dissolved in acids followed by analysis of both metals by AAS. Contamination

by Au and Ag impurities in the reagents is corrected for by fusing the reagents without the test portion. Au in aqueous solutions can be determined as Au metal by using an appropriate reduction agent and igniting at 800°C. Oxalic acid is used for separating Au and the other platinum group metals. Tin(II) chloride, hydrobromic acid, and Rhodamine B are utilized as the chromogenic reagent for spectrophotometric determination of Au. Radiations of wavelengths 242.8 and 267.6 nm are used for the analytical line for atomic absorption spectrometric detection of Au. Each metal's sensitivity (1% absorption) is 0.2 and 0.4 mg l<sup>-1</sup> with Air-C<sub>2</sub>H<sub>2</sub> flame. There is no interference from other platinum group metals, but iron and copper interfere. The characteristic mass in picogram/0.0044 absorbance is 13 in GF-AAS. The most sensitive spectral line for ICP-AES is 242.795 nm. The detection limit is ~3 ng l<sup>-1</sup>. Mass number = 197 is used for ICP-MS detection. The detection limit is ~0.0005 ng l<sup>-1</sup>. TaO<sup>+</sup> (= 197) interferes with Au determination. The current popular stripping voltammetric for the determination of Au are based on the chloro-complex with gold on the carbon paste electrode. Au can be determined in the range from 5 to 250 nmol l<sup>-1</sup>. Neutron activation analysis (NAA) combined with radiochemical separation is a very sensitive technique. The Au in the sample is subject to the nuclear reaction: <sup>197</sup>Au(n,γ)<sup>198</sup>Au. The intensity from <sup>198</sup>Au radionuclide (*t*<sub>1/2</sub> = 2.7 days) that corresponds to the concentration of Au in the sample is measured. The sensitivity is 10<sup>-6</sup>–10<sup>-3</sup> μg. The interferences include Ag, Pt, and Br, which might be critical for the determination of Au at the nanogram per gram level. X-ray fluorescence spectrometry (XRF) is able to give a rapid estimation of the sample composition and often without damaging the sample. The accuracy, however, is not good enough to satisfy the official requirements in the analysis of jewelry, etc. The rapid analysis by XRF is utilized in routine quality control.

### Silver

Gravimetric method such as AgCl or fire assay is used. Gay-Lussac method involves the titration of Ag with NaCl and/or KBr in acid solution. The endpoint is taken as the point at which no further precipitation of silver halide is observed upon the addition of more precipitant. Fajans method is based on the precipitation of Ag with chloride or bromide ions, using an adsorption indicator to detect the endpoint. Rhodamine 6G or dichlorofluorescein is used as the adsorption indicators. Volhard method is based on the titration of Ag with potassium or sodium thiocyanate to form insoluble silver thiocyanate. This titration is

carried out in dilute acid medium in the presence of a small quantity of ferric ion as the indicator. For accurate work, an indicator blank must be determined and subtracted. The most generally suitable reagents at the present time for the photometric determination of small amounts of Ag are dithizone and *p*-dimethyl- or *p*-diethyl-amminobenzylidenerhodamine. Neither is specific for Ag. Au and Pd react with both. Small amount of Hg can be tolerated in the direct rhodamine method but not in dithizone. In practical analysis, separations will usually be required before dithizone or rhodamine can be applied. The nephelometric determination of Ag as chloride is also used as a conventional method. Radiations of wavelengths 328.1 and 338.3 nm are used for the analytical line for atomic absorption spectrometric detection of Ag. Each metal's sensitivity (1% absorption) is 0.05 and 0.15 mg l<sup>-1</sup> with Air-C<sub>2</sub>H<sub>2</sub> flame. The characteristic mass in picogram/0.0044 absorbance is 1.4 in GF-AAS. The detection limit is ~1.5 ng l<sup>-1</sup>. The most sensitive spectral line for ICP-AES is 328.068 nm. The detection limit is ~1 ng l<sup>-1</sup>. Mass number = 107 is used for ICP-MS detection. The detection limit is ~0.0009 ng l<sup>-1</sup>. ZrO<sup>+</sup> (= 107) interferes in Ag determination. The popular stripping voltammetric method for the determination of Ag is based on nitric acid media on a carbon paste electrode. Au can be determined in the range from 4 to 20 nmol l<sup>-1</sup>. The Ag in the sample is subject to the nuclear reaction: <sup>107</sup>Ag(n,γ)<sup>108</sup>Ag or <sup>109</sup>Ag(n,γ)<sup>110</sup>Ag. The sensitivity is 10<sup>-4</sup>–10<sup>-2</sup> μg. XRF is able to give a rapid estimation of the sample composition and often without damaging the sample.

### Platinum

There are no selective gravimetric reagents for the determination of Pt. The element must first be isolated from interfering metals. Ammonium chloride or sodium formate is the acceptable method for gravimetric procedure. The two important reagents for the spectrophotometric determination of Pt are tin(II) chloride and *p*-nitrosodimethylaniline. The advantage of the latter is its sensitivity. It is far less selective than the tin(II) chloride method and requires close control of conditions of color development. Radiation of wavelength 266.0 nm is used for the analytical line for atomic absorption spectrometric detection of Pt. The sensitivity (1% absorption) is ~0.05 mg l<sup>-1</sup> with Air-C<sub>2</sub>H<sub>2</sub> flame. The characteristic mass in picogram/0.0044 absorbance is 100 in GF-AAS. The detection limit is ~10 ng l<sup>-1</sup>. The most sensitive spectral line for ICP-AES is 265.95 nm. The detection limit is ~2 ng l<sup>-1</sup>. Mass number = 195 is used for ICP-MS detection. The

detection limit is  $\sim 0.0007 \text{ ng l}^{-1}$ .  $\text{HfO}^+$  ( $=195$ ) interfere Pt determination. The current popular stripping voltammetric methods for the determination of Pt are based on the catalytic effect of the formazone (a condensation product of formaldehyde and hydrazine) complex with platinum on the developed hydrogen at a mercury electrode. Pt can be determined polarographically. More sensitive approaches involve adsorption preconcentration of the platinum complex on a hanging mercury drop electrode, which is followed by cathodic stripping. Earlier stripping voltammetric methods for determining Pt have been based on the masking of the sulfide or thiourea peak by platinum or on the deposition of platinum on a graphite or glassy-carbon electrode from solutions containing  $\text{PtCl}_4^{2-}$  and  $\text{PtCl}_6^{2-}$  ions, followed by anodic stripping voltammetry. The sensitivity of these methods is not high (detection limit  $\sim 10^{-8} \text{ mol l}^{-1}$ ). On irradiation in a thermal neutron flux, the platinum in the sample is subject to the nuclear reaction:  $^{198}\text{Pt}(n,\gamma)^{199}\text{Pt} \rightarrow ^{199}\text{Au}$ . The intensity of a  $\gamma$ -peak ( $E_\gamma = 158.9 \text{ keV}$ ) from  $^{199}\text{Au}$  radionuclide ( $t_{1/2} = 3.15 \text{ days}$ ) that corresponds to the concentration of Pt in the sample is measured. The sensitivity is  $10^{-2} \sim 1 \mu\text{g}$ . The irradiated sample is decomposed and mixed with the carrier (Au), which is separated using the methods described in the section on gold. The rapid analysis by XRF is utilized in routine quality control.

### Palladium

Dimethylglyoxime in an alcohol solution is generally used as the gravimetric reagent. When a large amount of palladium is to be precipitated, the water-soluble sodium salt is more convenient. The gravimetric determination of palladium as the diiodide is not selective as with dimethylglyoxime, but quicker and at least as accurate. The precipitate is very stable, filters readily, and can be easily reduced to the metal. Palladium may be readily reduced to the metal by formate ion. This is best done in a 20% sulfuric acid medium since hydrochloric acid tends to redissolve the metal. The iodide titration as palladium iodide is one of the most acceptable methods. The isolation of palladium before titration is essential. Compounds containing the *p*-nitrosophenylamino group form strongly colored complexes with Pd(II). Both *p*-nitrosodimethylaniline and *p*-nitrosodiphenylamine may be used for the photometric determination of Pd. Radiation of wavelength 283.3 nm is used for the analytical line for atomic absorption spectrometric detection of palladium. The sensitivity (1% absorption) is  $\sim 0.01 \text{ mg l}^{-1}$  with Air-C<sub>2</sub>H<sub>2</sub> flame. The characteristic mass in picogram/0.0044

absorbance is 24 in GF-AAS. The detection limit is  $\sim 4 \text{ ng l}^{-1}$ . The most sensitive spectral line for ICP-AES is 248.89 nm. The detection limit is  $\sim 6 \text{ ng l}^{-1}$ . Mass number = 106 is used for ICP-MS detection. The detection limit is  $\sim 0.002 \text{ ng l}^{-1}$ .  $\text{Cd}^+$  ( $=106$ ) and  $\text{ZrO}^+$  ( $=106$ ) interfere in Pd determination. The Pd in the sample is subject to the nuclear reaction  $^{108}\text{Pd}(n,\gamma)^{109}\text{Pd}$ . The sensitivity is  $10^{-4} - 10^{-2} \mu\text{g}$ . Because of low content in natural samples, radiochemical neutron activation analysis (RNAA) is utilized instead of instrumental neutron activation analysis (INAA).

### Rhodium

There are no specific precipitants for Rh. Rh is precipitated as metal by hydrazine in alkaline solution. The hydrolytic procedure is an acceptable gravimetric procedure. Few satisfactory photometric methods are available for Rh. The tin(II) chloride method is generally preferred. The sensitivity can be increased by carrying out the reaction in the presence of bromide or iodide. Radiation of wavelength 343.5 nm is used for the analytical line for atomic absorption spectrometric detection of Rh. The sensitivity (1% absorption) is  $\sim 0.02 \text{ mg l}^{-1}$  with Air-C<sub>2</sub>H<sub>2</sub> flame. The characteristic mass in picogram/0.0044 absorbance is 12 in GF-AAS. The detection limit is  $\sim 100 \text{ ng l}^{-1}$ . The most sensitive spectral line for ICP-AES is 369.24 nm. The detection limit is  $\sim 3 \text{ ng l}^{-1}$ . Mass number = 103 is used for ICP-MS detection. The detection limit is  $\sim 0.009 \text{ ng l}^{-1}$ .  $\text{SrOH}^+$  ( $=103$ ) and  $\text{SrO}^+$  ( $=103$ ) interfere in Rh determination. The complexity of interferences negates the possibility of using electrochemical methods in routine quantitative work with Rh. The Rh in the sample is subject to the nuclear reaction  $^{103}\text{Rh}(n,\gamma)^{104\text{m}}\text{Rh}$ . The sensitivity is  $10^{-5} - 10^{-3} \mu\text{g}$ . The sensitivities obtained in NAA are several orders of magnitude lower than by other approaches, but it is hard to achieve them on real samples due to the complexity of interferences. XRF is often favored, when applicable, because of the speed of analysis. However, the sensitivity of XRF is relatively poor and limited to samples containing milligram or higher amounts. Lead interferes in Rh determination with XRF. A number of procedures have been developed for XRF analysis of Rh after separation and preconcentration by fire assay or by resin-impregnated filter paper.

### Ruthenium

Ru may be quantitatively precipitated as a hydrous oxide. It is necessary to reduce the distilled Ru(IV) oxide to the four-valent state. Sulfite and alcohol are

often used as reducing agents. There is no acceptable volumetric procedure for the determination of Ru. In some early studies of the oxidation states of dissolved ruthenium halides, Ru(IV) was determined directly by titration with tin(II) chloride as indicated by the color change from red-brown to light rose, and more accurately by addition of excess tin(II) chloride and back-titration with iodine in the presence of starch. Strongly colored procedures are obtained when an acid solution of a ruthenium salt is heated with thiourea. 1,4-Diphenylthiosemicarbazide forms a red-violet chloroform-extractable complex with Ru. In the presence of hydroxylammonium chloride, Ru(III or IV) reacts with phenanthroline to form yellow tris(1,10-phenanthroline) Ru(II) ion. The reaction is slow even at 100°C and after 2 h at this temperature the reaction is ~97% complete. Radiation of wavelength 372.8 nm is used for the analytical line for atomic absorption spectrometric detection of Ru. The sensitivity (1% absorption) is  $\sim 0.06 \text{ mg l}^{-1}$  with Air-C<sub>2</sub>H<sub>2</sub> flame. The characteristic mass in picogram/0.0044 absorbance is 32 in GF-AAS. The most sensitive spectral line for ICP-AES is 378.89 nm. The detection limit is  $\sim 60 \text{ ng l}^{-1}$ . Mass number = 102 is used for ICP-MS detection. The detection limit is  $\sim 0.001 \text{ ng l}^{-1}$ . Pd<sup>+</sup> (=102) and SrO<sup>+</sup> (=102) interfere in Ru determination. The Ru in the sample is subject to the nuclear reaction  $^{102}\text{Ru}(n,\gamma)^{103}\text{Ru}$ . The sensitivity is  $10^{-3}$ – $10^{-1} \mu\text{g}$ . Because of low content in natural samples, RNAA is utilized instead of INAA.

### Iridium

The hydrolytic procedure in combination with purification by chlorination is an acceptable gravimetric procedure. The Ir solution must be free from large concentrations of base metals and other Pt metals, with the exception of Pt metals. The acidic solution containing Ir is treated with sodium bromate while boiling and the solution is neutralized with addition of sodium bicarbonate solution. The precipitate is filtered, dried, and ignited. Tin(II) chloride-hydrobromic acid method will usually be preferred to other photometric methods for Ir. Radiation of wavelength 264.0 nm is used for the analytical line for atomic absorption spectrometric detection of Ir. The detection limit is  $\sim 1.0 \text{ mg l}^{-1}$  with N<sub>2</sub>O-C<sub>2</sub>H<sub>2</sub> flame. The characteristic mass in picogram/0.0044 absorbance is 275 in GF-AAS. The detection limit is  $\sim 50 \text{ ng l}^{-1}$ . The most sensitive spectral line for ICP-AES is 322.08 nm. The detection limit is  $\sim 70 \text{ ng l}^{-1}$ . Mass number = 193 is used for ICP-MS detection. The detection limit is  $\sim 0.0006 \text{ ng l}^{-1}$ . HfO<sup>+</sup> (=193) interferes in Ir determination. INAA is a highly sensitive

method of analysis with which low and sub-nanogram per gram limits of detection can be achieved. Ir in the sample is subject to the nuclear reaction  $^{191}\text{Ir}(n,\gamma)^{192}\text{Ir}$  or  $^{193}\text{Ir}(n,\gamma)^{194}\text{Ir}$ . The sensitivity is  $10^{-5}$ – $10^{-2} \mu\text{g}$ .

### Osmium

Os may be determined gravimetrically after precipitation as the hydrous oxide. The procedure is similar to that used for Ru, with difference that all ignitions must be done in hydrogen or an inert atmosphere to avoid loss as Os(VIII) oxide. The hydrous oxide is precipitated within the pH range 1.5–6.3 by adjusting the boiling solutions with sodium bicarbonate additions. A number of volumetric methods have been used for the determination of Os with accuracy, in some cases, as good as a few tenths of 1%. In most cases sealed-tube reactions are used to force the reactions. Hydrazine sulfate, tin(II) chloride, and iodide have been used to titrate Os(VIII). These methods are only applied to the analysis of pure salts. Thiourea is an old and well-established photometric reagent for Os, but it is not very sensitive. When really small quantities of Os must be determined, the catalytic method based on the reaction between As(III) and Ce(IV) will be chosen. Radiation of wavelength 305.9 nm is used for the analytical line for atomic absorption spectrometric detection of Os. The detection limit is  $\sim 0.4 \text{ mg l}^{-1}$  with N<sub>2</sub>O-C<sub>2</sub>H<sub>2</sub> flame. The characteristic mass in picogram/0.0044 absorbance is 1400 in GF-AAS. The detection limit is  $\sim 100 \text{ ng l}^{-1}$ . The most sensitive spectral line for ICP-AES is 290.91 nm. The detection limit is  $\sim 6 \text{ ng l}^{-1}$ . Mass number = 192 is used for ICP-MS detection. The detection limit is  $\sim 0.0005 \text{ ng l}^{-1}$ . Pt<sup>+</sup> (=192), YbO<sup>+</sup> (=192), HfO<sup>+</sup> (=192), and LuO<sup>+</sup> (=192) interfere in Os determination. The detection limit for Os achieved by XRF is of the order of milligram per kilogram for solid samples. NAA is a sensitive technique for Os determination. However, as a common major component in natural samples is iron (Os is a siderophilic element), which would cause a high background in the  $\gamma$ -ray spectrum by the Compton scattering effect, it is very difficult to determine microgram amounts of Os in metallic samples by NAA without separation of Os from iron either by INAA or RNAA. Although RNAA is time consuming compared with INAA, it provides a much higher sensitivity for Os determination below nanogram per gram levels.

See also: **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Mass Spectrometry:**

Inductively Coupled Plasma. **Fire Assay.** **Gravimetry.** **Ion Exchange:** Overview. **Sampling:** Theory; Practice. **Titrimetry:** Overview.

## Further Reading

- Alfassi ZB and Wai CM (1992) *Preconcentration Techniques for Trace Elements*. London: CRC Press.
- Balcerzak M (2002) Sample digestion methods for the determination of traces of precious metals by spectrometric techniques. *Analytical Science* 18: 737–750.
- Beamish FE and Van Loon JC (1977) *Analysis of Noble Metals, Overview and Selected Methods*. New York: Academic Press.
- Bugbee EE (1984) *A Textbook of Fire Assay*, 3rd edn. Colorado: Colorado School of Mines Press.
- Kolthoff IM and Elving PJ (eds.) (1963) *Treatise on Analytical Chemistry*, Part II, vols 4 and 8. New York: Wiley.
- Onishi H (1986) *Photometric Determination of Traces of Metals, Part IIA: Individual Metals, Aluminum to Lithium*. New York: Wiley.
- Onishi H (1986) *Photometric Determination of Traces of Metals, Part IIB: Individual Metals, Magnesium to Zirconium*. New York: Wiley.
- Qu YB (1996) Recent developments in the determination of precious metals: A review. *Analyst* 121: 139–161.
- Smith EA (1987) *The Sampling and Assay of the Precious Metals*, 2nd edn. Colorado: Met-Chem Research Inc., originally published by Charles Griffin & Company Ltd.
- Van Loon JC and Barefoot RR (1991) *Determination of the Precious Metals. Selected Instrumental Methods*. New York: Wiley.

# PRESSURIZED FLUID EXTRACTION

See **EXTRACTION: Pressurized Fluid Extraction**

# PRIONS

**B Brenig**, Universität Göttingen, Göttingen, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

## Transmissible Spongiform Encephalopathies

The term transmissible spongiform encephalopathies (TSE) is used broadly to cover a group of neurodegenerative diseases with strikingly similar histopathologies. The diseases of this group known to date in humans are Gerstmann–Sträussler–Scheinker disease, fatal familial insomnia, fatal sporadic insomnia, Kuru, classical Creutzfeld–Jakob disease (CJD), and a new variant of Creutzfeld–Jakob disease (vCJD). Also belonging to the group of TSE is scrapie in sheep and goats as well as bovine spongiform encephalopathy (BSE) in cattle. Transmissible Mink Encephalopathy first appeared in 1947 in mink farms in the USA, the last recorded case of this very rare disease worldwide was reported in the USA in 1985. Furthermore, TSE have been confirmed in eight different species of wild ruminants kept in captivity. An additional animal species affected is the family of cats, not just exotic large cats such as cheetahs, pumas, tiger, and ocelot, but also

domesticated cats, who are susceptible to feline spongiform encephalopathy.

The typical pathological disease profile of TSE is sponge-like (spongiform) changes in brain tissue caused by the formation of microscopically small holes. These particular features as well as additional fundamental characteristics of the demonstrable infectivity of this disease led to the name ‘transmissible spongiform encephalopathy’.

Transmission of a TSE was first achieved by Cuille and Chelle, who managed to infect healthy sheep and goats by intracerebral inoculation of spinal marrow homogenates from sheep suffering from scrapie. Conclusive evidence for the infectious nature of TSE came from long-term inoculation experiment, which succeeded in transmitting CJD and Kuru to chimpanzees.

## Nature and Pathological Action of the TSE Agent

The search for the causative agent of TSE unearthed numerous unusual features that distinguish it significantly from other previously known pathogens. The agent was resistant to formaldehyde and heat



Inductively Coupled Plasma. **Fire Assay.** **Gravimetry.** **Ion Exchange:** Overview. **Sampling:** Theory; Practice. **Titrimetry:** Overview.

## Further Reading

- Alfassi ZB and Wai CM (1992) *Preconcentration Techniques for Trace Elements*. London: CRC Press.
- Balcerzak M (2002) Sample digestion methods for the determination of traces of precious metals by spectrometric techniques. *Analytical Science* 18: 737–750.
- Beamish FE and Van Loon JC (1977) *Analysis of Noble Metals, Overview and Selected Methods*. New York: Academic Press.
- Bugbee EE (1984) *A Textbook of Fire Assay*, 3rd edn. Colorado: Colorado School of Mines Press.
- Kolthoff IM and Elving PJ (eds.) (1963) *Treatise on Analytical Chemistry*, Part II, vols 4 and 8. New York: Wiley.
- Onishi H (1986) *Photometric Determination of Traces of Metals, Part IIA: Individual Metals, Aluminum to Lithium*. New York: Wiley.
- Onishi H (1986) *Photometric Determination of Traces of Metals, Part IIB: Individual Metals, Magnesium to Zirconium*. New York: Wiley.
- Qu YB (1996) Recent developments in the determination of precious metals: A review. *Analyst* 121: 139–161.
- Smith EA (1987) *The Sampling and Assay of the Precious Metals*, 2nd edn. Colorado: Met-Chem Research Inc., originally published by Charles Griffin & Company Ltd.
- Van Loon JC and Barefoot RR (1991) *Determination of the Precious Metals. Selected Instrumental Methods*. New York: Wiley.

# PRESSURIZED FLUID EXTRACTION

See **EXTRACTION: Pressurized Fluid Extraction**

# PRIONS

**B Brenig**, Universität Göttingen, Göttingen, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

## Transmissible Spongiform Encephalopathies

The term transmissible spongiform encephalopathies (TSE) is used broadly to cover a group of neurodegenerative diseases with strikingly similar histopathologies. The diseases of this group known to date in humans are Gerstmann–Sträussler–Scheinker disease, fatal familial insomnia, fatal sporadic insomnia, Kuru, classical Creutzfeld–Jakob disease (CJD), and a new variant of Creutzfeld–Jakob disease (vCJD). Also belonging to the group of TSE is scrapie in sheep and goats as well as bovine spongiform encephalopathy (BSE) in cattle. Transmissible Mink Encephalopathy first appeared in 1947 in mink farms in the USA, the last recorded case of this very rare disease worldwide was reported in the USA in 1985. Furthermore, TSE have been confirmed in eight different species of wild ruminants kept in captivity. An additional animal species affected is the family of cats, not just exotic large cats such as cheetahs, pumas, tiger, and ocelot, but also

domesticated cats, who are susceptible to feline spongiform encephalopathy.

The typical pathological disease profile of TSE is sponge-like (spongiform) changes in brain tissue caused by the formation of microscopically small holes. These particular features as well as additional fundamental characteristics of the demonstrable infectivity of this disease led to the name ‘transmissible spongiform encephalopathy’.

Transmission of a TSE was first achieved by Cuille and Chelle, who managed to infect healthy sheep and goats by intracerebral inoculation of spinal marrow homogenates from sheep suffering from scrapie. Conclusive evidence for the infectious nature of TSE came from long-term inoculation experiment, which succeeded in transmitting CJD and Kuru to chimpanzees.

## Nature and Pathological Action of the TSE Agent

The search for the causative agent of TSE unearthed numerous unusual features that distinguish it significantly from other previously known pathogens. The agent was resistant to formaldehyde and heat

treatment when attempting to produce a vaccine against scrapie. Investigations showed that the agent is extremely resistant to ultraviolet (UV) and ionizing radiation.

According to the prion hypothesis, the typical TSE damage to nerve tissue is caused by an infectious agent that consists exclusively of a membrane protein altered in its secondary structure. The protein in question is a highly conserved glycoprotein composed of 254 amino acids with a molecular weight of 33–35 kDa. It is found in mono-, di-, and non-glycosylated forms. In its original, physiological form this protein is possibly involved in copper metabolism in the brain and according to recent research also plays a role in the capture of stress-induced oxygen radicals.

One knows that this protein is encoded by a single exon, in humans called the PRNP gene. It is primarily found on the cell surface of neurons, glial cells, and cells of the lymphoreticular system, and it is anchored in these cell membranes via a glucosyl-phosphatidyl-inositol group. The protein is also detected at low concentrations in muscle cells and cells of the immune system.

The trigger for TSE is a structural change in this membrane protein, where certain regions of the protein normally adopting an alpha-helix conformation refold into a beta-sheet structure. This is considered the initiating step for the development of TSE. Its effect on the physical and chemical properties of the protein lead to the extreme resistance to heat, formaldehyde, and UV light referred to above, amongst other properties.

The decisive new property responsible for the damage to nerve tissue is, however, the acquisition of partial resistance to proteases, since the refolded protein cannot be completely degraded. This leads to protein deposits in the form of amyloid plaques in the nerve cells and finally to cell death through apoptosis and other processes. The amyloid plaques are also called scrapie-associated fibrils (SAF). As a result, the brains of damaged individuals show focal and diffuse degeneration and vacuole formation, which are responsible for the typical spongiform appearance of the TSE brain.

A further essential feature of the altered protein is its apparent activity as a crystallization seed: in the proximity of altered proteins, normal proteins are similarly stimulated to refold.

Therefore, the altered protein possesses the essential properties of a disease-causing agent: it has a damaging effect and is able to propagate itself. The unique feature of this pathogenic protein is that in contrast to all previously known infectious agents it is apparently able to propagate itself without the help of nucleic acids. Prusiner named his discovery prion

derived from 'proteinaceous and infectious'. Since then the term prion has come to be used generally for the infectious agents of TSE, even if the agent in question is not one specifically defined.

Subsequently, the physiological, unaltered form of the prion was called PrP<sup>C</sup> for the cellular prion protein and the pathologic, altered form was called PrP<sup>Sc</sup>, derived from the scrapie prion protein.

Prusiner's hypothesis provided a plausible explanation for why TSE does not elicit an immune reaction and why no TSE-associated coding nucleic acids could be found as yet in infectious material. However, it is still not convincingly clarified as to how the conversion from PrP<sup>C</sup> to infectious PrP<sup>Sc</sup> happens without regulating nucleic acids. Although it is possible in cell culture to convert PrP<sup>C</sup> into PrP<sup>Sc</sup> by the addition of PrP<sup>Sc</sup> from infected animals, PrP<sup>Sc</sup> stemming from transgenic organisms are not infectious. Transgenic mice that produce elevated amounts of PrP<sup>Sc</sup> do develop several signs of TSE, but passing on the disease using PrP<sup>Sc</sup> derived in this way has as yet been unsuccessful. This means the conclusive evidence to support the prion hypothesis, in other words the *de novo* synthesis of infectious PrP<sup>Sc</sup>, still remains elusive.

Another hypothesis for the pathological action of the TSE agent is based on the idea that apart from the prion protein another virus-like structure is involved in the development of a TSE disease. This structure comprises a protein coat, a single-stranded DNA layer, and a core of SAFs, and has been called 'scrapie termed nemavirus' (NVP).

As mentioned above, it has not yet been possible to induce an infection by transmission from transgenically produced pure prions. Transmission of TSE is, however, possible with a mixture of single-stranded DNA derived from NVP structures from the brain of scrapie-infected hamsters, and the so-called binding proteins. From this he concluded that the structural changes in PrP<sup>C</sup> essentially require an accessory protein, possibly encoded by the single-stranded DNA of the NVP structure.

The participation of a so-called 'protein X' in the refolding of PrP<sup>C</sup> is also suspected by other authors. Taken together, it seems that although various theories about the pathological action of PrP<sup>Sc</sup> exist, the involvement of this protein in the development of a TSE is, however, generally recognized. How the various forms or phenotypes of TSE are generated can apparently not only be traced back to genetic differences between the affected hosts, but also results from interplay between genetic factors and the action of different pathogen strains.

Individual strains differ in respect to their incubation times as well as with reference to the findings of

histopathological brain examinations. The disruption of nerve cells caused by the accumulation of amyloid plaques leaves different pictures with the individual forms of TSE with respect to the distribution of the vacuoles and degree of tissue damage, producing agent-typical lesion profiles. Individual strains can be identified in experimentally infected laboratory mice by comparing the incubation time and the lesion profile typical for the pathogen. The typical strain characteristics are retained over several generations in experimentally infected mice and are independent of the host. Possibly, the strain-specific information responsible for the differences regarding lesion profile and incubation time is contained within the tertiary structure of the PrP<sup>Sc</sup>.

According to more recent investigations there are also differences with respect to the molecular weight and relative proportion of individual glyco-forms within the PrP<sup>Sc</sup> fraction between individual strains. In addition, it is possible to distinguish between them by using various antibodies.

Differences relating to the parameters mentioned above can be detected between vCJD, a variant of CJD that will be discussed in more detail later, and classical forms of CJD. Various findings concerning these criteria also result from comparisons between scrapie PrP<sup>Sc</sup> and BSE PrP<sup>Sc</sup>. In contrast, the vCJD PrP<sup>Sc</sup> of humans and the BSE PrP<sup>Sc</sup> of cattle, as well as sheep that have been infected with material from BSE cows under experimental conditions, show striking similarities.

## Possible Causes for the Formation of TSE Agents

The formation of PrP<sup>Sc</sup> through alterations in the structure of a normal host protein can have different causes. These are partly determined by differences in the disease progression and in the histopathology of the various TSE forms. In this connection both germ line mutations and somatic mutations are relevant. In addition, an infection can be initiated by implantation of infected tissue. A further source of infection can be oral consumption of the infective agent.

Examples for the different causes of a TSE disease are provided by the various subforms of CJD in humans. CJD was first observed in 1920, and remains one of the best analyzed prion diseases. Classic CJD is divided into three subforms according to the pathogenic mechanism: with 'iatrogenic CJD' patients are infected, for example, through administration of contaminated human growth hormone or transplantation of contaminated dura mater. In addition, the disease can be passed on by contaminated EEG needles in individual cases. In total, more than

267 cases of 'iatrogenic CJD' have been recorded in the last 20 years.

The cause of 'familial CJD' is a germ line mutation in the PRNP gene. To date, 12 mutations are known that seem to be unambiguously associated with familial CJD. The assumption that there is a direct connection between 'familial CJD' and mutations in the PRNP gene was supported experimentally by studies in mice: transgenic mice carrying the mutations mentioned above show several essential aspects of a TSE.

What triggers the appearance of the third form of CJD, 'sporadic CJD', which with one or two cases per million people, per year, worldwide constitutes the major proportion of all TSE in humans, remains unclear. In these cases no connection with mutations in the PRNP gene can be found. Somatic mutations in the PRNP gene have been suggested, although such speculation could not be confirmed experimentally as yet.

A variant of the classical CJD (vCJD) that first appeared in 1993 in the UK differs from the subforms described above in that it preferentially affects younger patients and shows somewhat slower clinical progression. One of the most frequently represented hypotheses for the cause of this disease is infection through oral consumption of products from BSE-affected cattle. Support for this theory comes from the already mentioned similarity between the prions isolated from the brainstem of vCJD patients and BSE animals, which clearly differ from the infective agent and histopathological findings with sporadic and inherited CJD. Lesion profiles that were comparable to the lesion profiles in brains from vCJD patients were found in histopathological analyses of the brain from macaques experimentally infected with BSE.

Additional support for this theory comes from an experiment where transgenic mice that exclusively produce human PrP<sup>C</sup> could be infected with BSE.

The causes for the appearance and spread of BSE are as yet not satisfactorily explained, and extensive research is still required. Experimentally, an infection can be achieved through intracerebral inoculation or oral administration of brain material from infected animals.

The possibility that cows transmit BSE to each other, for example, by infected placenta, amniotic fluid, or contaminated pastureland, could not be confirmed as yet.

Specific mutations that appear in sheep in association with scrapie and the knowledge derived from sheep breeding allowing the selection of scrapie, and BSE-resistant animals, could not be confirmed in cattle as yet.

It is generally assumed that the infective agent reached the cattle population through feeding them animal-meal produced from scrapie-affected sheep where heat treatment was of insufficient duration. Possibly, PrP<sup>Sc</sup> managed to reach the CNS, via peripheral nerves, directly from the digestive tract, or the pathogen first accumulated and proliferated in the lymph organs and later invaded the CNS via the nerve cell network.

In support of transmission via contaminated fodder is the fact that in England at the beginning of the 1980s, in connection with a rationalization measure, the production process for meat and bone-meal was simplified, which subsequently meant dispensing with one procedural step that included steam heat treatment. This step has been shown to reduce the infectivity of PrP<sup>Sc</sup>. The etiology of the BSE epidemic in Great Britain, therefore, indicates that feeding of contaminated animal-meal can be considered the major source of infection. Since the ban on feeding cattle animal-meal, which came into force in the UK in 1989, the number of BSE cases has dropped noticeably.

## Diagnosis of Prion Protein and Transmissible Spongiform Encephalopathies

Basically, as applies to all forms of TSE, a confirmed diagnosis is not possible in a living animal. Although the combination of a range of clinical investigations can justify a suspicion and corroborate it, absolute certainty is only obtained after a histopathological analysis of the CNS, which will be described in more detail later.

The following descriptions concentrate on the field of BSE diagnosis. In order to offer as comprehensive an overview as possible, also concerning theoretically useful methods for BSE diagnosis, additional methods are included that until now have only been applied in human medicine.

### Postmortem Diagnosis

All BSE tests approved so far are either based on histopathological brain analysis or direct detection of PrP<sup>Sc</sup>, which in cattle is still only possible in nerve tissue. Since in the course of the disease prion accumulation appears first and in the highest concentration in the Obex area of the brainstem at the transition between the spinal cord, tissue samples for analysis should strictly be taken from this region.

All practiced methods where diagnosis relies on detection of PrP<sup>Sc</sup> are based on the partial protease resistance of PrP<sup>Sc</sup>. When applying the following test

methods, there is a choice between rapid tests, recommended by the EU Commission for the analysis of larger numbers of samples, and the OIE test methods, which are approved by the International Animal Breeding Authority, and in contrast to the rapid tests used as a screening method, are suited for making a conclusive diagnosis.

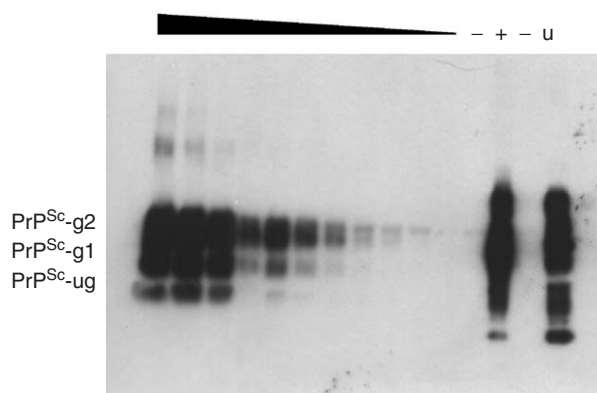
**ELISA rapid tests** As a screening method for analyzing larger sample numbers, two out of three procedures evaluated by the EU Commission as sufficiently sensitive are approved. The method preferred by most laboratories is an enzyme-linked immunosorbent assay (ELISA) procedure where homogenized tissue is subjected to proteinase K hydrolysis. PrP<sup>C</sup> is completely digested, while with PrP<sup>Sc</sup> only the first 62–80 amino acids are cleaved off. The remaining PrP<sup>Sc</sup> is fixed on an ELISA plate using antibody already bound to the plate. Binding of a further labeled antibody to the fixed PrP<sup>Sc</sup> leads to a color change that allows detection of PrP<sup>Sc</sup>.

A further ELISA works not through using a color change, but by photometric measurement of a chemiluminescence signal. This achieves higher sensitivity.

**Western blot rapid test** In contrast to the two ELISA rapid tests, a third rapid test procedure based on the Western blot principle offers the advantage that it not only detects the presence of PrP<sup>Sc</sup> but at the same time determines the molecular weight of the protein. This provides a control parameter for assessing the completeness of the proteinase K digestion. The disadvantage of this test is the lack of defined endpoint. The isolation of PrP<sup>Sc</sup> is achieved by separating the protein fraction of the homogenized brainstem preparation using SDS polyacrylamide gel electrophoresis (Figure 1).

**OIE Western blot** With the approved OIE Western blot test the isolation of PrP<sup>Sc</sup> is achieved in a different way from the Western blot rapid test, by using several successive ultracentrifugation steps during which the PrP<sup>Sc</sup>, which tends to form fibrillar structures, accumulates in a pellet and is thus separated from normal PrP<sup>C</sup>. This step, although laborious, considerably increases the test sensitivity, and is followed by a proteinase K digest as described above in the Western blot rapid test and subsequent detection by Western blotting. Due to the time consuming preparation of the material to be analyzed, this test takes several days and is not suited for use as a screening test.

**Immunohistochemical analysis** A further procedure approved by the International Epidemic Authority is



**Figure 1** Autoradiogramme of Western blot rapid test. Different dilutions (1:1 to 1:1024) of a brain homogenate from the Obex region of a BSE-infected cow were digested with proteinase K together with positive (+) and negative (-) controls. After digestion, the samples were loaded on a SDS-PAGE and separated. After blotting onto a nylon membrane, the filter was hybridized with a Prp-specific antibody and PrP<sup>Sc</sup> was detected by a chemiluminescence detection method after exposure to an X-ray film. PrP<sup>Sc</sup>-ug, unglycosylated PrP<sup>Sc</sup>; PrP<sup>Sc</sup>-g1, single glycosylated PrP<sup>Sc</sup>; PrP<sup>Sc</sup>-g2, double glycosylated PrP<sup>Sc</sup>.

the immunohistochemical analysis of brain tissue preparations. Tissue blocks 2 µm thick are hydrolytically autoclaved. Following this preparation accumulation of PrP<sup>Sc</sup> in nerve cells can be visualized using a PrP<sup>Sc</sup>-specific antibody and subsequent staining. This method is highly sensitive and also provides rough, ordered anatomical information.

**Histoblot** A further test procedure is the histoblot. Here, blocks of tissue 2 mm thick are transferred to nitrocellulose membranes, lysed, and subjected to proteinase K digestion. Following proteinase K digestion over 8 h at 55°C, protein is subsequently fixed to the membrane. After blocking the membrane with a protein solution the membrane is treated first with a PrP<sup>Sc</sup>-specific antibody and then a second antibody directed against the PrP<sup>Sc</sup>-specific antibody. This second antibody is stained with Formazan, meaning the immunoreaction can be visualized under the microscope. The histoblot permits sensitive protein detection and provides detailed anatomical information. The disadvantage is that only freshly frozen material can be used. Samples stored in formaldehyde solution that are perfectly suitable for carrying out other procedures cannot be used, since such material is too soft to dissect into sufficiently small tissue blocks for transferring to the nitrocellulose membrane.

**PET-blot** A very sensitive and new procedure that overcomes such problems is the so-called paraffin embedded tissue blot (PET-blot). By covering



**Figure 2** PET-blot analysis of a BSE-infected brain. Deposition of PrpSc in the brain was detected by the PET-blot technique. PET-blot provided with courtesy of Dr W. Schulz-Schaeffer.

samples stored in formaldehyde solution with paraffin, tissue blocks of 5–7 µm in size are cut that are small enough to be fixed on a nitrocellulose membrane. After the paraffin is released from the membrane-bound proteins, proteinase K digestion and subsequent immunodetection is carried out as with the histoblot. Even in cases where the immunohistochemical analysis gave ambiguous results, the PET-blot can clearly detect PrP<sup>Sc</sup> and is more sensitive than conventional Western blot and histoblot procedures (Figure 2).

**Histopathological analysis** Similarly, histopathological analysis is suitable merely to confirm and not to refute a suspected case of BSE, and detects vacuolization in certain areas of the brain. The majority of TSE cases are accompanied by focal or diffuse degeneration of the gray matter in the entire CNS, above all in the cerebellum, as well as the formation of vacuoles. As a result, damage to the nerve cells in addition to proliferation and enlargement of astrocytes is observed. During damage to the nerve tissue microscopically small holes are formed that are responsible for the sponge-like appearance of the brain in TSE.

**Electron microscope analysis** Electron microscope visualization of prion accumulation, also known as scrapie-associated fibrils or amyloid plaques, which are isolated by ultracentrifugation as described above, is also approved as a diagnosis procedure. This procedure was frequently used at the beginning of the BSE epidemic in England, although since then it has proved to be less reliable than the methods described above.

**Approaches to a Premortal Diagnosis**

In this section, clinical BSE diagnostic techniques will be discussed in detail. Subsequently, to complete the picture further methods that play more of a secondary role in BSE diagnosis will be covered briefly, since in part they have applications exclusively in human medicine or are still under development.

**Clinical analysis** BSE-infected cows normally show a distinct pattern of clinical signs, including three complexes, i.e., ‘behavior’, ‘sensitivity’, and ‘movement’.

As suspect behavioral disturbances changes in anxiety and character are typical. Most frequent are disturbed defensive reactions such as being easily startled (in 66% of animals), anxiety (in 66% of animals), and nervousness or restlessness (in 64% of animals). Concerning aggressive behavior, increased readiness to repulse approaches was apparent in 36% of the animals, maliciousness and lashing out in 20% of animals. Fear of narrow passages and small obstructions was more seldom.

To assess sensitivity, reactions to manipulation of the neck and head as well as to noises and light were tested and the so-called broom test was carried out where the severity of lashing out upon touching the shackles with a broom was observed. Typical observations of arousal during the sensitivity test were defensive movements or more sedate arousal

reactions such as salivation, snorting, nervous ear twitching, licking the rim of the mouth, and quivering muscles. Furthermore, repeated or even a progressively more severe shock reaction upon repeating the same light or noise stimulation was interpreted as an indication of BSE. Over-sensitivity to tactile stimulation was apparent in 98% of animals, 84% reacted hypersensitively to noise, but light sensitivity was observed in only 45% of animals.

As BSE-typical features for the animals’ movement, hesitation and anxiety are observed when crossing the septic tank or passing through the stall door, disturbed balance and ataxia of fore and hind legs and hypermetria. About 92% of the animals showed ataxia, whereby a generalized ataxia appeared twice as frequently as one restricted to the hind legs. Less frequent was problems with standing up, lying down, and falling over. Often the disturbances in movement were only weakly manifest.

Animals showing at least two of the three signs of complexes of behavior, sensitivity, and movement disturbances, in most cases in fact all three, are BSE positive. Therefore, precise examination of all three complexes is indispensable, since some signs may only be weakly manifest particularly in the early stages.

A clinical calibration of BSE-investigated animals into four diagnostic categories is based on these observations (**Table 1**).

Often milk production and underfeeding can be observed. The cows’ appetite is graded as normal, similarly the frequency of rumen contractions. From observations of reduced strength of contractions and the rumen content, it can be concluded that the poor nutritional state of the animals, despite the normal desire to feed, could be traced back to lower ingestion of food.

A further clinical finding considered as being of possible diagnostic interest is a bradycardia. Here,

**Table 1** Clinical BSE diagnosis

Group	Clinical diagnosis	Clinical findings (various possibilities)
I	BSE	All typical findings present (disturbed behavior, movement and sensitivity), or Two major signs clearly manifest (e.g., easily startled and atactic or nervous and sensitive to noise)
II	No BSE	No signs typical for BSE, or One single finding mildly positive (e.g., mildly startled by noise or mild sensitivity to manipulation of the neck), or Disturbed movement without disturbed behavior or sensitivity
III	Suspected BSE	Strongly pronounced behavior disturbances (nervous, anxious, easily startled), or Strongly pronounced disturbed sensitivity (over sensitive to manipulation, noise, light)
IV	BSE not excluded	Two main signs mildly pronounced (e.g., relatively easily startled and mildly sensitive to light), or One individual finding clearly positive (e.g., startled by noise), other major signs clearly negative



this refers to a significant slowing down of the pulse rate. This observation seems in contradiction to the nervous and anxious state of the animals. A possible explanation can be attributed to the pathological-anatomical changes in the vagus nerve, which could lead to increased vagal tone. Increased intracranial pressure is one of the possible causes for a bradycardia. It cannot be excluded that ECG findings after further detailed investigations could have a diagnostic significance.

The normal clinical blood chemistry and hematological values routinely determined in the laboratory, as well as the urine results, gave no indication of BSE. In contrast, the analysis of cerebrospinal (CS) fluid could be useful, since this can be a means to exclude other CNS diseases such as encephalitis, which in contrast to BSE are accompanied by inflammatory changes in the CS fluid. However, more detailed use of CS fluid data determined within the framework of routine clinical analyses for the diagnosis of BSE can be excluded. This is also confirmed in human medicine investigations.

**Blood analysis** To directly detect prion proteins in blood one could take advantage of the specific binding of PrP<sup>Sc</sup> to plasminogen, as soon as the specificity of this binding is clearly defined.

A further approach to diagnosing TSE is based on the detection of cellular nucleic acids in serum (CNAPS). Production of such nucleic acids is apparently very specific, for example, as a result of malignant diseases, autoimmune diseases, and infectious diseases and opens up the possibility of diagnosing such diseases through the detection of these nucleic acids. Possibly, this reaction to diseases is part of an older immune system. Investigations with diverse serum material from cattle provided optimistic results. The CNAPS are present in the liquid fraction of blood samples from both healthy and diseased cattle. The CNAPS were found in the blood fraction that contains primarily microvesicles. CNAPS profiles have been shown to differ in humans with multiple myeloma and mysterious neurologic syndromes such as Gulf War syndrome and chronic fatigue syndrome.

**Electroencephalogram (EEG)** In human medicine electroencephalographic findings are brought in as a further diagnostic aid. The so-called 'periodic sharp wave complexes' appear with ~60–70% of patients with sporadic CJD.

**Nuclear magnetic resonance spectroscopy** Also, nuclear magnetic resonance spectroscopy or nuclear spin tomography (MRI, magnetic resonance imaging),

which depicts body structures through imaging the distribution of free-moving hydrogen atom, has some importance as diagnostic tool. Amongst other things, this can visualize atrophy in the brain, although atrophy in several brain regions is not one of the major signs for confirming a suspected CJD. In various investigations, abnormal symmetrical elevated signals in the caudate nucleus and putamen were found in up to 67% of the suspected patients analyzed. Above all, MRI findings played a role in distinguishing between vCJD and sporadic CJD. Up to 80% of vCJD cases showed hyperintensities in the posterior thalamus that clearly exceeded the elevated signals in the caudate nucleus and putamen. This provides a possibility to clinically differentiate between vCJD and sporadic CJD.

**Positron emission tomography** Furthermore, positron emission tomography delivered useful results in a first analysis with a small number of patients. This method involves injecting patients with labeled ions, for example, oxygen, carbon, fluoride, or glucose. A reaction takes place, releasing radiation, when these positively charged particles meet a negatively charged counterpart in the body. This emission can be measured and localized, thus making it possible to build up a picture of the distribution of individual molecules in particular regions of the body. All the patients investigated showed reduced glucose metabolism in the brain. The distribution of the area of reduced glucose metabolism differed significantly from the typical pictures obtained with other brain diseases included as comparisons, such as Alzheimer's, Parkinson's, frontal temporal dementia, Lewy body dementia, Wilson's disease, Huntington's disease, and dementia due to blood vessel damage.

**Bioassay systems** Particularly in animal studies, an increasingly applied option for detecting the presence of a TSE infection is the inoculation of mice or hamsters with infectious material taken by biopsy from the brain of the animal or human under investigation. Against its use in the clinic is the highly variable sensitivity of this bioassay system, the not insignificant risk of damage following the brain biopsy for the patient under examination, and the disadvantage that due to the incubation time the results are only available much later.

## Cerebrospinal Fluid Analyses

### Direct Detection of Prions in CS Fluids

**Detecting infectivity in CS fluids** Standard diagnostic analyses of cerebrospinal fluids from cattle or even

from people have resulted in no notable findings. Evidence was provided for the infectivity of cerebrospinal fluids when infecting primates by inoculation with undiluted cerebrospinal fluid from CJD patients. Considering the high receptivity of primates for CJD, this transmission rate of only 15% is perhaps low, but can be rated as indirect evidence that CS fluids contain PrP<sup>Sc</sup>. In an Internet publication from the Paul Ehrlich Institute, where bovine tissue, secretions, and excretions are divided into four categories according to their prion content, CS fluids are allocated to the category – high infective agent content – and thus lie between the categories ‘very high infective agent content’ and ‘moderate infective agent content’.

**The SIFT method** Direct detection of PrP<sup>Sc</sup> in CS fluids is now possible. Using the method SIFT (scanning for intensely fluorescent targets), the lowest molecular concentration that could be detected was less than a femtomole. This method represents a further development of classical fluorescence spectroscopy. It includes an immunoreaction of the molecule to be detected with specific antibodies that are labeled with a fluorescent compound. A light beam causes these marker compounds to fluoresce at a characteristic wavelength. Light of the expected wavelength is selectively conducted by an emission filter or monochromator to a detector, where it is converted into electrical energy and transferred via an amplifier to a monitoring device.

The intensity of the fluorescent signal from the marker compound is weak, but increases 50-fold when it is bound to an antibody. This means fluorescent signals from possible free macromolecules in the sample under investigation can be distinguished from those arising from the substance being sought.

In classical fluorescence spectroscopy, in contrast to the SIFT method, the entire sample is exposed to the light beam and the fluorescence intensity, which results from the sum of the signals received from the preparation, is measured within a particular time-frame. Here, scattered and reflected light photons create the so-called background noise, which tends to drown out fluorescence signals from the small amounts of molecules to be detected. Therefore, with this method only amounts above the nanomolar level can be detected.

In order to cut down interfering effects such as background noise, only a small proportion of the volume is light-activated in the SIFT method, which then emits correspondingly little scattered light. If the labeled substance to be detected enters this focus, it emits several thousand fluorescence light photons within a few milliseconds in the form of sequential

rapid-fire flashes of fluorescence. These are detected using autocorrelation. Moreover, the fluorescence intensities measured in two directly sequential time intervals are multiplied together. If the emitting particle in the focus is the sought molecule, then the expected signal in the two sequential time intervals is measurable and the product of both measured intensities does not equal zero. A possible interference signal of background noise, in contrast, is expected to be in the form of a few flashes that are only visible for the duration of the measuring interval. The product of the two fluorescence intensities in the two sequential time intervals is in this case equal to zero. In this way, the principle of autocorrelation permits separation of specific signals from unspecific signals. Unspecific signals mostly arise from free, labeled antibodies or those bound to PrP<sup>C</sup> arriving in the measuring focus, which in contrast to those bound to PrP<sup>Sc</sup> are single and not present in the form of agglomerates.

In order to increase the specificity of the measurements, the developers of the SIFT method target the molecule to be detected with two antibodies instead of one. The antibodies, one of which is ‘stamped’ with a green fluorescence marker, the other with a red fluorescence marker, each bind specifically to two different sites on the molecule. The measuring device is equipped with two light-sensitive detectors, one for green and one for red. A specific reaction means a simultaneous measurement of both green and red fluorescence signals.

**Application of the SIFT method in clinical diagnosis** Concerning its application in clinical diagnosis the described procedure still shows major deficits. Investigations with a total of 24 CS fluid samples from confirmed CJD patients and 13 control samples from definite non-CJD patients resulted in no positive signal in the control samples. Therefore, diagnostic specificity was 100%. However, with a sensitivity of ~20%, where PrP<sup>Sc</sup> could only be detected in five out of 24 fluid samples, this method is far from being suited for use in clinical diagnosis. It remains to be established whether all CJD patients really have measurable amounts of PrP<sup>Sc</sup> in their CS fluid. There is still justifiable hope that this method, through appropriate optimization, can be developed in the future into a specific diagnostic tool.

## Surrogate Markers in CS Fluid

Due to the lack of available procedures for reliable, direct detection of prions in CS fluids, suitable surrogate markers for TSE were sought; in other

words, molecules that are as specifically as possible found in association with TSE fluids.

What was found in CS fluids of CJD patients was increased levels of the brain proteins: neuron-specific endolase (NSE), Tau protein, S-100b, G<sub>0</sub>-protein, and brain-specific creatine kinase, amongst others. All these proteins are observed to increase in concentration in CS fluids as the CJD condition progresses. In the final stage of the disease their concentrations decrease again. In addition, in CJD patients, as with Alzheimer's patients, a reduced concentration of A $\beta$ -amyloid in CS fluids was found.

Of particular significance as surrogate marker proteins are a group of brain proteins known as 14-3-3, which currently play an important role in the diagnosis of familial and sporadic CJD. Among other things, these proteins are involved in diverse metabolic processes in nerve cells. 14-3-3 proteins appear in the cerebrospinal fluid when acute nerve cell loss or acute brain disruption occurs, as with TSE, but also with strokes and several other diseases (herpes simplex encephalitis, multi-infarct dementia, acute infarct, subarachnoid hemorrhage (bleeding in cerebrospinal space), viral encephalitis).

In the literature, the parameters of specificity and sensitivity are used in particular to compare various surrogate marker tests and diagnostic procedures, as well as to provide a reference point for evaluating the predictive accuracy of individual tests with respect to a CJD diagnosis.

Specificity represents the percentage of marker-negative patients in the group of definitely not CJD-affected patients. This means it is a measure of the certainty with which the marker test recognizes a CJD-negative patient as being such.

Sensitivity indicates the proportion of marker-positive patients in the group of definitely or probably CJD-affected patients. Thus, sensitivity provides a reference point for the certainty that definite or probable CJD-positive patients can be identified within a suspect group using the marker test. Patients that are graded as probably CJD-positive demonstrate advanced dementia with a duration of less than 2 years in connection with at least two of the clinical characteristics of myoclonus, visual or cerebellar signs, pyramidal or extrapyramidal signs, as well as akinetic changes, and show in EEG 'periodic sharp wave complexes'.

Table 2 shows the sensitivity and specificity of selected surrogate marker test procedures in comparison with other diagnostic techniques in sporadic CJD. Sensitivity and specificity refer to a given selection of patients whose neurological and psychiatric condition fits in the differential diagnosis context.

**Table 2** Sensitivity and specificity of diagnostic techniques in sporadic CJD

Marker	n	Sensitivity (%)	Specificity <sup>a</sup> (%)
<i>Cerebrospinal fluid</i>			
14-3-3	1136	95	93
NSE > 35 ng ml <sup>-1</sup>	1276	81	92
S100 > 4.2 ng ml <sup>-1</sup>	135	84	91
Tau > 1400 pg ml <sup>-1</sup>	290	93	91
PrP <sup>Sc</sup>	34	20	100
MRI	208	63	92
EEG	805	66	74

<sup>a</sup>In neurological/psychiatric conditions, which are relevant in the differential diagnosis.

From Table 2 it is clear that the 14-3-3 test, in comparison to other diagnostic procedures, shows a particularly high specificity and sensitivity. Also, relatively high sensitivity and high specificity are reported for detecting Tau protein.

In a comparison of the Tau protein test with the 14-3-3 test with a group of 297 patients the Tau protein test performed just as well in the parameters of sensitivity and specificity as the 14-3-3 test. The authors concede, however, that the 14-3-3 test in these investigations gave poorer values than with earlier analyses carried out by them. They attribute the cause to the fact that in the earlier investigations the group of definite and probable CJD cases served as the basis for evaluating specificity and sensitivity, whereas the evaluation parameters in the investigation described here were calculated based merely on definite cases.

### Various Detection Procedures for 14-3-3 Proteins

**14-3-3 detection using two-dimensional gel electrophoresis** Using two-dimensional gel electrophoresis the proteins p130 and p131 were detected, which later turned out to belong to the group of 14-3-3 proteins, in CS fluids of 21 CJD patients. In an investigation with 100 healthy volunteers, 420 patients with other nervous disorders and 21 CJD patients, a sensitivity of 100% and a specificity of 91% was obtained.

The diagnostic usefulness of a combination of Harrington's p130/131 detection system with measurement of the NSE concentration in CS fluids was tested. Harrington's p130/131 detection gave a sensitivity of 84% for sporadic CJD and 81% for familial CJD. The specificity in these investigations was 100%. The p130/131 proteins were detectable in 47 out of 58 definite CJD cases, in 37 out of 46 probable cases, and in 23 out of 34 possible cases. In

44 definite cases of other diseases p130/131 could not be detected. Therefore, the p130/131 test is recommended as a backup diagnosis in CJD cases diagnosed clinically as possible or probable, and which show an NSE concentration of over 20 ng ml<sup>-1</sup> in CS fluids.

**14-3-3 Western blot immunoassay (immunostaining)** The p130/131 proteins were also detected in normal brain tissue and identified with the help of partial sequence analysis as the already known 14-3-3 brain proteins. This information made it possible to develop a new method for detecting the proteins using immunoassays.

This method involves electrophoretically separating the protein fraction of CS fluid and subsequently transferring these proteins to a nitrocellulose membrane according to Western blot procedures. After saturating the free binding sites on the membrane with a Tris-buffered salt solution, the membrane is treated with the 14-3-3-specific antibody ‘anti-14-3-3 $\beta$  polyclonal rabbit antibody’. A second antibody binds this antibody, the ‘alkaline phosphatase-conjugated anti-rabbit IgG antibody’. Detection is achieved by a colorimetric reaction.

With this new method a sensitivity of 96% was attained: 14-3-3 proteins were detected in CS fluids from 68 out of 71 definite CJD cases. Out of 137 samples from patients with other nervous disorders, 22 were 14-3-3 positive. This gives a relatively low specificity of 83.9%. If within the group of patients with other diseases one only takes into account the proportion that show dementia and suffer no stroke within a month of lumbar puncture, specificity increases to 99%: of the 91 patients who fulfilled these criteria, 14-3-3 proteins were detected in only one. This patient suffered from Alzheimer’s. Due to these false-positive results the authors recommend caution

in interpreting the 14-3-3 test results and stress that importance should only be attached to the results of these tests in connection with clinical investigations.

The new method was compared with the detection method of Harrington using two-dimensional gel electrophoresis (Table 3). In this experiment based on 50 samples the results of both tests matched except for the number of positive samples within the category ‘clinically certain’ CJD cases. Within this category 14-3-3 proteins could be detected using an immunoassay in all patients, whereas using two-dimensional electrophoresis detection was successful in only 13 patients. Therefore, the authors attribute a slightly higher sensitivity to the immunoassay than to the two-dimensional electrophoresis.

An immunoassay investigation detected 14-3-3 proteins in CS fluids from every one of the 20 patients investigated, who were definitely suffering from sporadic CJD, whereas the results of CS fluid analysis of 20 patients with dementia due to other diseases were negative for 14-3-3. The authors, therefore, rate the immunoassay as highly sensitive and specific for the diagnosis of CJD.

The high sensitivity and specificity of the 14-3-3 immunoassay was also confirmed by other investigations. In a study with 484 patients, the positive predictive value of the 14-3-3 immunoassay reached 94.7% and the negative predictive value 92.4%. The sensitivity within the group of definite CJD patients was 95.4%, within the group of probable CJD patients, 92.8%. Altogether, the sensitivity was 94% and the specificity 93.3%.

The appearance of 14-3-3 proteins in CS fluids was also investigated in 80 patients with paraneoplastic nerve defects (PND), in order to test whether a false CJD diagnosis might possibly be reached due to a positive 14-3-3 test result with these patients. As a positive control the team included 53 CS fluid

**Table 3** Comparison of two-dimensional electrophoresis for protein 130 and 131 with the 14-3-3 immunoassay

Diagnosis	Total number of samples	Positive samples	
		Two-dimensional electrophoresis	14-3-3 immunoassay
CJD			
Pathologically confirmed	5	5	5
Clinically certain <sup>a</sup>	10	8	10
Total	15	13	15
Other diseases			
Dementia	18	0	0
Other neurological diseases	13	1	1
Total	31	1	1
Healthy control samples	4	0	0

<sup>a</sup>Clinically certain is defined as rapid advancing dementia with myoclonus and characteristic encephalographic results.

samples from CJD patients; as a negative control, CS fluid samples from CJD-negative patients with various nervous defects were similarly subjected to the 14-3-3 test. Among the samples from the 80 PND patients, 12.5% showed a positive reaction for 14-3-3, while 14-3-3 proteins were detected in CS fluids from 48 out of the 53 CJD-positive patients (i.e., 90.6%). In the group of 55 CJD-negative patients, six were 14-3-3 positive.

An investigation with 805 patients resulted in a sensitivity for the 14-3-3 immunoassay of 94% and a specificity of 84%.

In general, the 14-3-3 immunoassay can be considered as a sensitive and specific parameter in the diagnosis of sporadic CJD. The 14-3-3 test is not suited for diagnosis of vCJD, since the sensitivity with these patients lies at just over 50%.

**14-3-3 Western blot immunoassay (chemiluminescence)** Bernheimer published the results of their 14-3-3 test on 20 suspected CJD patients, of which three proved to be actually CJD positive and five were assessed as probable cases, as well as 18 comparable cases with other diseases. The detection method was the only way the test differed from the procedure of Hsich described above: detection of the 14-3-3 proteins was not achieved using immunostaining, but according to the chemiluminescence procedure.

The 14-3-3 test results were distributed similarly to the investigations described above, even if the authors did reject direct comparison with previous studies of sensitivity and specificity on the grounds of low sample numbers. The three CJD-positive patients showed a positive 14-3-3 reaction. Amongst the five probable CJD cases, one was positive for 14-3-3. Two patients assessed as possible CJD cases reacted negatively for 14-3-3, and from the remaining 10 not diagnosed, suspected CJD patients, eight were 14-3-3 positive. Among the 18 control patients there were six with positive 14-3-3 test results. In investigations of CS fluids from vCJD patients, a combination of the 14-3-3 test with Tau protein detection improved the negative and positive predictive values of both tests. The positive predictive value for the 14-3-3 test was 86%, the negative predictive value 63%. The Tau protein test attained a positive predictive value of 93% and a negative predictive value of 81%. Combination of the results from both tests, however, resulted in a positive predictive value of 91% and a negative predictive value of 84%.

**14-3-3 detection using ELISA** On the basis of the immunoassay, an ELISA was developed that, in

contrast to the immunoassays described above, allowed not only qualitative but also quantitative detection of the molecule under investigation.

Here, the monoclonal 'capture anti-14-3-3 antibody' added to the wells of a so-called microtiter plate made up of a number of regularly assembled wells per plate, and binds tightly to the bottom of these wells. The samples to be investigated and 14-3-3 standards with known concentrations of 14-3-3 proteins are pipetted in the wells and after a reaction time during which the 14-3-3 proteins bind to the 'capture antibody' on the bottom of the microplate, they are rinsed with washing solution. A protein-containing solution added subsequently to the wells serves to block remaining unbound 'capture antibody'. Following more washing steps, two detection antibodies are added in sequence to the wells. These are the rabbit anti-14-3-3 $\beta$  polyclonal antibody, which binds specifically to the 14-3-3 protein, and the peroxidase-labeled goat anti-rabbit IgG antibody that binds to the first antibody. After the immunoreaction TMB is added as substrate for the colorimetric reaction. The enzymatic reaction that takes place leads to a certain degree of staining that depends on the amount of 14-3-3 protein bound in the well. Quantification is achieved by comparing the absorption values of the different samples with the absorption values of the 14-3-3 standards.

The lowest amount of 14-3-3 protein detected using the ELISA was a concentration of 2.1 ng ml<sup>-1</sup>. Comparison of the new ELISA method with the Western blot immunoassay using samples from 63 definite or probable CJD patients and 84 samples from patients with other neurological defects resulted in a sensitivity of 92.7% for the ELISA relating to the definite CJD patients; with reference to the definite and probable CJD patients, sensitivity was 88.9%. In contrast, the sensitivity of the Western blot immunoassay within the group of definite CJD patients was 95.1%, and in the group of definite and probable CJD patients it was 93.7%. The specificity for both tests was 97.6%.

At the same time a 14-3-3 protein ELISA was developed, which achieved a relative quantification of the 14-3-3-protein concentration in CS fluid from CJD and non-CJD patients. To determine the sensitivity of the ELISA, the 14-3-3 content in extracts from cattle was also measured in parallel using ELISA and immunoblots. For the ELISA the lowest amount of 14-3-3 protein detectable was a concentration of 0.5 ng ml<sup>-1</sup> and for the Western blot immunoassay a concentration of 4 ng ml<sup>-1</sup>. This test should make it possible, through its high sensitivity and quantification of the 14-3-3 concentration in investigated material, to distinguish between various

forms of CJD, which have different levels of 14-3-3 concentrations in CS fluids.

Comparison of the ELISA with a Western blot immunoassay revealed a higher sensitivity of 95% for the ELISA compared to 93% for the immunoassay. With reference to specificity, however, the Western blot immunoassay came off better, with 100%, than the ELISA with merely 92%.

*See also:* **Electrophoresis:** Blotting Techniques; Proteins.

## Further Reading

- Awerbuch G, Peterson P, and Sandyk R (1988) Elevated cerebrospinal fluid lactic acid levels in Creutzfeldt–Jakob disease. *International Journal of Neuroscience* 42: 1–5.
- Basler K, Oesch B, Scott M, *et al.* (1986) Scrapie and cellular PrP isoforms are encoded by the same chromosomal gene. *Cell* 46: 417–428.
- Bieschke J, Giese A, Schulz-Schaeffer W, *et al.* (2000) Ultrasensitive detection of pathological prion protein aggregates by dual-color scanning for intensely fluorescent targets. *Proceedings of the National Academy of Science USA* 97: 5468–5473.
- Braun U, Schicker E, and Hörnlimann B (1998) Diagnostic reliability of clinical signs in cows with suspected bovine spongiform encephalopathy. *The Veterinary Record* 143: 101–105.
- Brenig B, Schütz E, and Urnovitz H (2002) Cellular nucleic acids in serum and plasma as new diagnostic tools. *Berliner und Münchener Tierärztliche Wochenschrift* 115: 122–124.
- Bruce ME, Will RG, Ironside JW, *et al.* (1997) Transmissions to mice indicate that ‘new variant’ CJD is caused by the BSE agent. *Nature* 389: 498–501.
- Collinge J, Sidle KC, Meads J, Ironside J, and Hill AF (1996) Molecular analysis of prion strain variation and the aetiology of ‘new variant’ CJD. *Nature* 383: 685–690.
- Creutzfeldt HG (1920) Über eine eigenartige herdförmige Erkrankung des Zentralnervensystems. *Zeitschrift für die gesamte Neurologie und Psychiatrie* 57: 1–18.
- Gajdusek DC, Gibbs CJ Jr., Asher DM, and David E (1968) Transmission of experimental kuru to the spider monkey (*Ateles geoffreyi*). *Science* 162: 693–694.
- Hill A, Desbruslais M, Joiner S, *et al.* (1997) The same prion strain causes vCJD and BSE. *Nature* 389: 448–450, 526.
- Jakob A (1921) Über eigenartige Erkrankungen des Zentralnervensystems mit bemerkenswertem anatomischen Befunde (spastische Pseudosklerose-Encephalomyelopathie mit disseminierten Degenerationsherden). *Zeitschrift für die gesamte Neurologie Psychiatrie* 64: 147–228.
- Lasmezas CI, Deslys JP, Demaimay R, *et al.* (1996) BSE transmission to macaques. *Nature* 381: 743–744.
- Schulz-Schaeffer WJ, Tschoke S, Kranefuss N, *et al.* (2000) The paraffin-embedded tissue blot detects PrP(Sc) early in the incubation time in prion diseases. *American Journal of Pathology* 156: 51–56.
- Zerr I, Bodemer M, Otto M, *et al.* (1996) Diagnosis of Creutzfeldt–Jakob disease by two-dimensional gel electrophoresis of cerebrospinal fluid. *Lancet* 348: 846–849.

# PROCESS ANALYSIS

Contents

**Overview**

**Chromatography**

**Electroanalytical Techniques**

**Sensors**

**Acoustic Emission**

**Maintenance, Reliability, and Training**

**Bioprocess Analysis**

## Overview

**C McGill**, Avecia Ltd., Grangemouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Process analysis is the application of analytical science to the monitoring and control of industrial processes. The data obtained from process analysis allow the control and optimization of industrial



forms of CJD, which have different levels of 14-3-3 concentrations in CS fluids.

Comparison of the ELISA with a Western blot immunoassay revealed a higher sensitivity of 95% for the ELISA compared to 93% for the immunoassay. With reference to specificity, however, the Western blot immunoassay came off better, with 100%, than the ELISA with merely 92%.

See also: **Electrophoresis:** Blotting Techniques; Proteins.

## Further Reading

- Awerbuch G, Peterson P, and Sandyk R (1988) Elevated cerebrospinal fluid lactic acid levels in Creutzfeldt–Jakob disease. *International Journal of Neuroscience* 42: 1–5.
- Basler K, Oesch B, Scott M, *et al.* (1986) Scrapie and cellular PrP isoforms are encoded by the same chromosomal gene. *Cell* 46: 417–428.
- Bieschke J, Giese A, Schulz-Schaeffer W, *et al.* (2000) Ultrasensitive detection of pathological prion protein aggregates by dual-color scanning for intensely fluorescent targets. *Proceedings of the National Academy of Science USA* 97: 5468–5473.
- Braun U, Schicker E, and Hörnlimann B (1998) Diagnostic reliability of clinical signs in cows with suspected bovine spongiform encephalopathy. *The Veterinary Record* 143: 101–105.
- Brenig B, Schütz E, and Urnovitz H (2002) Cellular nucleic acids in serum and plasma as new diagnostic tools. *Berliner und Münchener Tierärztliche Wochenschrift* 115: 122–124.
- Bruce ME, Will RG, Ironside JW, *et al.* (1997) Transmissions to mice indicate that ‘new variant’ CJD is caused by the BSE agent. *Nature* 389: 498–501.
- Collinge J, Sidle KC, Meads J, Ironside J, and Hill AF (1996) Molecular analysis of prion strain variation and the aetiology of ‘new variant’ CJD. *Nature* 383: 685–690.
- Creutzfeldt HG (1920) Über eine eigenartige herdförmige Erkrankung des Zentralnervensystems. *Zeitschrift für die gesamte Neurologie und Psychiatrie* 57: 1–18.
- Gajdusek DC, Gibbs CJ Jr., Asher DM, and David E (1968) Transmission of experimental kuru to the spider monkey (*Ateles geoffreyi*). *Science* 162: 693–694.
- Hill A, Desbruslais M, Joiner S, *et al.* (1997) The same prion strain causes vCJD and BSE. *Nature* 389: 448–450, 526.
- Jakob A (1921) Über eigenartige Erkrankungen des Zentralnervensystems mit bemerkenswertem anatomischen Befunde (spastische Pseudosklerose-Encephalomyelopathie mit disseminierten Degenerationsherden). *Zeitschrift für die gesamte Neurologie Psychiatrie* 64: 147–228.
- Lasmezas CI, Deslys JP, Demaimay R, *et al.* (1996) BSE transmission to macaques. *Nature* 381: 743–744.
- Schulz-Schaeffer WJ, Tschoke S, Kranefuss N, *et al.* (2000) The paraffin-embedded tissue blot detects PrP(Sc) early in the incubation time in prion diseases. *American Journal of Pathology* 156: 51–56.
- Zerr I, Bodemer M, Otto M, *et al.* (1996) Diagnosis of Creutzfeldt–Jakob disease by two-dimensional gel electrophoresis of cerebrospinal fluid. *Lancet* 348: 846–849.

# PROCESS ANALYSIS

Contents

**Overview**

**Chromatography**

**Electroanalytical Techniques**

**Sensors**

**Acoustic Emission**

**Maintenance, Reliability, and Training**

**Bioprocess Analysis**

## Overview

**C McGill**, Avecia Ltd., Grangemouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Process analysis is the application of analytical science to the monitoring and control of industrial processes. The data obtained from process analysis allow the control and optimization of industrial

processes by providing information from which the chemical and/or physical composition of the process stream can be obtained.

Process analytical chemistry (PAC) techniques have been grouped into the following categories.

- **Offline:** A sample is removed from a process stream or reactor and transported to a centralized laboratory where it is analyzed by skilled analysts on sophisticated equipment. This has the advantage that expensive instrumentation can be shared for the analysis of different types of samples. However, these samples can take hours or even days to be analyzed and this can cause a delay to production. There can also be confusion caused by differing opinions on the priority of a sample and the ownership of the data produced by the analysis.
- **At-line:** A dedicated analytical instrument is placed next to or close to the process stream. This allows faster sample analysis with a simpler, less expensive instrument and gives process operators closer control of the analysis. An at-line instrument must be robust enough to cope with conditions on the chemical plant, e.g., temperature, solvent atmosphere, and vibrations.
- **Online:** A sample is automatically removed from the process, conditioned, and then passed to the process analyzer for measurement. Online analysis can be continuous, where the sample flows continually through the analyzer, or intermittent, where aliquots of the sample are injected into the instrument at various time intervals. A dedicated instrument is utilized for the analysis of the process stream. Included under the umbrella of online techniques are inline and noninvasive techniques.
- **Inline:** Analysis is carried out by a probe that is placed directly inside the process stream. This removes the need for a separate side arm to transport the sample for analysis to the instrument.
- **Noninvasive:** This is seen as being the main aim of any process analysis technique. With noninvasive techniques, there is no physical contact between the analyzer and the sample. This removes any problems with sampling that may be encountered, e.g., probe fouling.

The advantages and disadvantages of the different categories of PAC are summarized in **Table 1**.

The instrumentation used in at-line and online analysis differs from laboratory instrumentation in that it is rugged and dedicated to a particular task, i.e., continually monitoring the same process stream, whereas laboratory-based analytical instruments are usually used to determine many analytes. Process analyzers are usually installed in safe analyzer houses that are kept at a constant temperature and pressure and have air continually passed through them to

**Table 1** Advantages and disadvantages of the different categories of PAC techniques

	<i>Advantages</i>	<i>Disadvantages</i>
Offline	Expert analysts available Flexible operation Controlled environment Sophisticated instrumentation Low unit cost per test	Can be slow Lack of ownership of data Conflicts of priorities
At-line	Dedicated instrument Faster sampling process  Simpler instrumentation Ownership of data by production personnel Control of priorities	Low equipment utilization Equipment needs to be robust to cope with production environment Data can be of poor quality
Online	Fast  Automatic feedback possible Dedicated analyzer  Can potentially be noninvasive	Minimum downtime required Long/expensive method development 24 h troubleshooting/maintenance resource required Electrical classification required Data can be of poor quality

remove any harmful vapors from the process that may be leaking into the safe house.

Traditionally, process analysis is carried out by manually removing a sample from a process, followed by offline analysis in a centralized laboratory. The samples are analyzed by qualified technicians using state-of-the-art instrumentation. Typically, it can take a few hours or even a few days for the results to be obtained and reported back to production personnel. The results from the analysis are then used retrospectively to determine the yield obtained, whether the material needs to be reworked or discarded, or to assess material going on to the next stage of a multistage batch process. While the results are awaited from the analytical laboratory, the process may be sitting idle when reactors could possibly be getting prepared for the next process. In a multistage batch synthesis, the next stage of the process cannot be started until the analytical results are obtained from the laboratory. In both cases the cycle time of the process is increased. The advantages obtained from using at-line or online process analyses are increased efficiency of the plant, increased safety, and the reduction of pollution. The efficiency of the plant is increased as there is no down time while results are awaited from a centralized analytical laboratory. Real-time results from an online process analyzer can be used in a feedback loop to maintain the conditions of a reaction. This helps to lower the 'quality cost' of production – the loss experienced

when a batch of material fails to meet the required specification and so has to be reworked, sold as lower standard material, or discarded altogether.

Figure 1 shows the steps involved in offline and online process analysis and it can be seen that online analysis takes less time to produce the data required for process control.

Reports have shown that when online analysis was used to control various parameters of a gas liquefaction plant, large savings in energy usage and therefore cost can be made. An example of process analysis increasing safety is in the production of ethylene. The presence of oxygen in the process can lead to the reaction becoming out of control which jeopardizes lives and plant equipment, so it is prudent to have the oxygen content of the reaction mixtures continuously monitored online and in real-time. Online analysis helps to minimize the waste produced by a production plant because as process efficiency is improved, fewer resources are lost to effluent streams. The concentration of any pollutants can also be monitored directly with a process analyzer to ensure that the least possible amount of unwanted material is produced. Using online process analysis is a much preferred method of pollution control as it involves the 'prevention' of pollution rather than the 'cure' of treating waste streams before they are released into the environment. Online analysis requires long and expensive method development. Expertise from many individuals is required if a technique is to be implemented successfully, e.g., analytical chemists, process engineers, process research and development chemists, chemometricians, and the

production manager. Round the clock resources for maintenance and troubleshooting are required, as the analyzer would be expected to work continually with minimum down time. If an analyzer is to be placed directly into the plant, it may be necessary for electrical classification tests to be carried out to permit the analyzer to be situated in the plant atmosphere, especially if it contains a significant concentration of solvent. This may require the analyzer to be kept in a specially designed analyzer house, where the atmosphere around the instrument can be controlled.

The move toward online analysis has been caused by the increased international competitiveness between companies in the chemical industry. In an increasingly consumer driven market, chemical manufacturers must supply their customers with products of a consistently high quality, and be able to supply these products when required by the customer. The increased control obtained by using online analysis allows the quality of materials to be maintained, and reduces delays in the delivery of products by reducing the number of batches that require to be reworked or discarded. The adoption of online analysis by the chemical industry has been aided by developments in microcomputer technology and analytical instrumentation, allowing data and hence information to be generated almost instantaneously.

## Techniques Currently Used in Process Analytical Control

Surveying the literature, it can be observed that most of the major techniques used in analytical chemistry are already in use as online process analyzers.

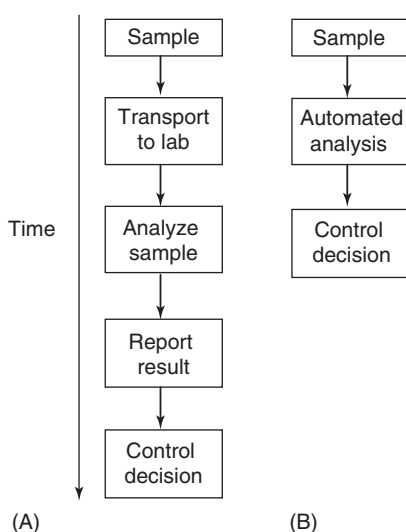
### Chromatography

Gas chromatography has become the one of the most widely used techniques in PAC for the analysis of volatiles. The applications in which it has been used include the determination of octane number in fuels using *n*-alkanes as reference peaks and to control the hydrogen/nitrogen ratio in ammonia production. A disadvantage of process gas chromatography is that there is a delay as the sample passes through the column before the results are known.

At the moment, process liquid chromatography is still under development to improve the reliability and ruggedness of the instrument although there have been some uses reported in the scientific literature.

### Optical Spectroscopic Techniques

Optical spectroscopic techniques are attractive for online analysis as remote sampling by fiber optics can



**Figure 1** (A) Steps involved in offline analysis. Time between sampling and a control decision being made can be days. (B) Steps involved in online analysis. Results obtained in seconds or minutes.

be employed. Fiber optic technology allows a probe to be placed directly into the stream or reactor and measurements made. This removes the need for a sample conditioning system (SCS) that many other online techniques require to regulate the flow rate, temperature, and pressure of the sample and perhaps also to remove particulates. Disadvantages of having an SCS include delays between sampling and analysis, and flushing between samples. It has also been reported that 90% of problems experienced by utilizing online analysis are traced back to the SCS. Care must be taken to ensure that the calibration procedure used takes into account probe fouling, temperature/pressure effects, presence of particulate matter, and any other interferences as there is no SCS present. By using fiber optics, the instrument can be kept out of hazardous areas and can be stored in a safe analyzer house. Different probe designs can be used to make either absorbance or diffuse reflectance measurements. These probes can be inserted directly into the process stream at the point of interest or be placed in a side stream. Contingency plans are necessary in the event of probe fouling. A quick and safe procedure is required to remove and clean the probe without having to close down the process or the stream affected.

Multiplexing, where multiple sampling points can be analyzed with a single instrument, is possible with spectroscopic techniques. Fiber optic cables going to different sampling points can share the energy from the optical source either by splitting the light between the different fibers (space multiplexing) or by passing all the energy down each fiber in turn (time multiplexing).

Ultraviolet (UV)–visible spectrometry, near-infrared spectrometry, mid-infrared spectrometry, and Raman spectrometry have all been used for online process analysis (including inline and noninvasive applications).

**UV–visible spectrometry** UV–visible spectrometry has been used to monitor *in situ* the dissolution of pharmaceutical products, the monitoring of batch reactions, and to monitor antioxidants in a polymer melt. Online UV–visible spectroscopy has also found use in the determination of metals in electroplating bath waste waters.

**Mid-infrared spectrometry** Mid-infrared spectrometry (MIR) has disadvantages that have hindered the widespread acceptance of the technique for online analysis. The wavelength of light used in mid-infrared spectrometry is highly attenuated by silica optical fibers. This limits the length of optical fiber that can be used for distancing the instrument from

the process. Special fibers have been produced, e.g., chalcogenide, which have a lower attenuation for mid-infrared wavelengths, but these fibers are expensive and they still have a very limited throughput which means that they can only be used over short distances. The fundamental absorptions in the mid-infrared region are very strong. If mid-infrared spectrometry was being used for the analysis of major constituents in a process stream, very short pathlengths would have to be employed to ensure the absorptions were not strong enough to deviate from the Beer–Lambert law. By using short pathlengths, there is a danger of small particulates in the process stream blocking the space between the windows of the cell. The problem of obtaining a short pathlength can be rectified by using an attenuated total reflectance (ATR) probe. ATR probes have some disadvantages though. For instance, they are susceptible to fouling and scratching of the surface, and the calibration has to be carried out carefully, as the ATR response is not always linear. ATR is affected by changes to the temperature and refractive index of the liquid being measured which also complicates the calibration procedure.

**Near-infrared spectrometry** Near-infrared spectrometry provides a vibrational spectrum of an analyte over the wavelength range of 1000–2500 nm ( $10000\text{--}4000\text{ cm}^{-1}$ ) produced by overtones and combinations of the fundamental vibrations of the chemical bonds in the analyte. Compared to the mid-infrared region, the near-infrared region is limited in its ability to provide detailed qualitative information of chemical structures due to the wide, overlapping bands that are observed. However, near-infrared spectrometry is very useful for providing quantitative information of molecules containing OH–, NH–, and CH– bonds.

Near-infrared spectrometry is preferred to mid-infrared spectrometry for making online measurements of a chemical process for a number of reasons. As it is desirable to keep an analytical instrument remote from the process, optical fibers are used to transport light to and from the flow cell/probe used for making the measurement. Unlike online MIR, silica fibers of a few hundred meters in length can be used to connect a near-infrared spectrometer instrument to the measurement probe. Absorptions in the near-infrared region are 10–100 times less intense than those in the mid-infrared region and because of this longer pathlengths ( $\sim 1\text{--}10\text{ mm}$ ) can be employed. Although the spectra obtained in near-infrared spectrometry are highly convoluted, precise quantitative information can be obtained by employing chemometric techniques when building a calibration model.

**Raman spectrometry** Raman spectrometry has the potential of being a very powerful technique in process analysis and has many advantages. Noncontact measurements can be made of solids and liquids, and fiber optics can be used to transmit the incident and scattered light, which means that *in situ* measurements can be made in harsh process environments. Raman spectrometers require only electricity to operate; cooling water, gases, and other services are not required. The technique can be used to measure analytes in aqueous solution, unlike infrared techniques, which give large absorptions for water. Another major advantage of Raman spectrometry is that glass can be used for cell windows. This means that laser light could be directed onto a sample through a glass window on a pipe or in the side of reactor, and the scattered light collected through the same window. This removes the problems of probe fouling that can be encountered with inline Raman, near-infrared, and mid-infrared spectrometries.

### **Mass Spectrometry**

Online mass spectrometry is very useful for obtaining process information concerning the molecular composition and structure of unknown compounds and for monitoring an analyte in a complex matrix, and has been used in many applications in the analysis of gaseous streams. Very little research has been carried out into the introduction of, and subsequent analysis of, liquid samples into a process mass spectrometer.

### **Nuclear Magnetic Resonance Spectrometry**

Nuclear magnetic resonance (NMR) spectrometry is a widely used analytical technique for the analysis of many different materials including organic chemicals, inorganic complexes, and large biological molecules. At the moment it is not commonly used in process analytical chemistry. This is partly due to NMR techniques being rather complicated and the signals obtained can be difficult for non-NMR specialists to understand. NMR instruments can be expensive and the technique is much less sensitive than others, e.g., MIR spectrometry.

However, NMR has the potential of being very useful in process analysis due to it being a noncontact and nondestructive technique. Another problem that was faced when NMR was initially tried as an online application was that laboratory instruments were simply moved into the plant. Process operators found problems in the calibration and maintenance of these instruments. Recently, small, dedicated NMR systems have been developed, which in turn has led to more reports of online NMR. Time domain measurements (low field, low resolution) are far more

common than frequency domain measurements (low field, high resolution).

### **Electroanalytical Techniques**

Electroanalytical techniques are not as widely used in process analysis as spectroscopic or chromatographic techniques, but are used in a number of applications. Conductivity measurements are often made to determine the concentration of ionic species in a solution, e.g., in a water demineralization unit, or to monitor the salt content of brine. Voltammetric and amperometric measurements have been used to determine the concentrations of metals in solutions, e.g., waste effluents. Ion-selective electrodes can be used to make potentiometric measurements to determine the concentrations of specific ions, and coulometric measurements have been used in process analysis to measure sulfur-containing compounds.

### **Choosing the Right Technique**

Choosing the best analytical technique to use for a particular application is rarely a straightforward task with many factors needing to be taken into account. The nature of the process stream often rules out a number of techniques, e.g., mass spectrometry and gas chromatography would not be considered for monitoring a blending process in the pharmaceutical industry. Often the nature of the process stream may determine the kind of sampling that is required, e.g., a highly corrosive stream may be more suited to a noninvasive technique rather than a technique that requires a probe to be inserted into the stream or reaction vessel. A full feasibility study should be carried out to determine the best technique to use. Often, the simplest option is by far the best, e.g., if the UV absorbance at a particular wavelength is shown to be selective and sensitive enough to give the required information, then a filter photometer will be much simpler and cheaper to install than online/inline UV-visible spectrometer.

Once a technique has been decided, a decision then has to be taken on which of the available instruments from different suppliers is to be purchased. Parameters that would have to be taken into consideration for this decision include cost, accuracy, precision, limit of detection, dynamic range, weather proofing, ease of calibration, materials of construction of wetted parts, power requirements, internal performance checking, running costs, and vendor support. Often the reputation of the potential instrument vendors can sway the decision on what instrument to purchase.

The technologies and philosophies used in online analysis, i.e., real-time *in situ* analysis, have also found use in environmental analysis and in medicine.

**See also:** **Process Analysis:** Chromatography; Electro-analytical Techniques; Sensors; Maintenance, Reliability, and Training; Bioprocess Analysis.

## Further Reading

Carr-Brion KG, Clarke JRP, and Harding EF (1994) *Sampling Systems for Process Analysers*. Oxford: Butterworth Heinemann.

Chalmers JM and Dent G (1997) *Industrial Analysis with Vibrational Spectroscopy*. Cambridge: The Royal Society of Chemistry.

Chalmers JM (ed.) (2000) *Spectroscopy in Process Analysis*. Sheffield: Sheffield Academic Press.

Maciel GE (1994) NMR in industrial process control and quality control. In: Maciel GE (ed.) *NMR in Modern Technology*. Dordrecht: Kluwer Publishing.

McLennan F and Kowalski BR (eds.) (1995) *Process Analytical Chemistry*. Glasgow: Blackie Academic and Professional.

Workman Jr. J, Veltkamp DJ, Doherty S *et al.* (1999) *Analytical Chemistry* 71: 121R.

## Chromatography

**P M Wiegand**, The Dow Chemical Company, South Charleston, WV, USA

**C P Sears**, The Dow Chemical Company, Freeport, TX, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A process chromatograph is an environmentally suitable automated device that measures one or more chemical components in one or more process streams and presents the results in a usable format. A process chromatographic system includes the associated sample system, utilities, waste handling, and data communication as well as the instrument itself.

In most cases, process chromatography is done in the gas phase. Many improvements have been made to liquid chromatographic systems that make them suitable for process use, but this article will focus only on process gas chromatography (PGC).

Process analyzers, whether chromatographs or other analytical technologies, are often compared with their laboratory counterparts. It is instructive to look at some of the differences between a laboratory gas chromatograph (LGC) and a PGC.

Sample introduction for an LGC is usually done with a manual or automated syringe that pierces a septum. A PGC introduces sample with a valve. The process equivalent of an autosampler is a set of stream-switching valves, which are used to choose the sample point of interest. The PGC must also be able to extract an aliquot of sample from a moving process stream some distance away from the analyzer. This requires a means of preserving the state of the sample and delivering a precise amount to the chromatograph without human intervention. Methods for accomplishing this are discussed in more detail below.

An LGC typically uses temperature programming to enhance separation. The PGC is more likely to be run at an isothermal oven temperature, often using multiple columns, with column-switching valves being employed to enhance the separation.

LGCs use mainly capillary columns, while PGCs still use packed columns. PGC with capillary columns is becoming more popular, but criteria such as detector compatibility, sample capacity, robustness, and temperature effects must be considered.

Detectors for lab and process instruments are very similar, but the digitized data may be handled differently. Lab instruments are designed to produce a visually readable report. Process instruments can usually produce a visual report, but their main function is to provide the data to another computer, which often uses the information to control the chemical process.

To summarize, a LGC is designed to be interactive and flexible and is able to handle a variety of samples. A PGC is designed to be independent and robust and is optimized for a particular type of sample.

## Service and Maintenance of Process Chromatograph

When considering the total cost of a process chromatograph, the cost of ongoing maintenance and service is frequently neglected. However, all process chromatographs need periodic maintenance to run reliably, and so planned maintenance should be included in any cost analysis.

### Consumables and Preventative Maintenance

Replacement of consumables can generally be considered routine for gas chromatography as large



The technologies and philosophies used in online analysis, i.e., real-time *in situ* analysis, have also found use in environmental analysis and in medicine.

**See also:** **Process Analysis:** Chromatography; Electro-analytical Techniques; Sensors; Maintenance, Reliability, and Training; Bioprocess Analysis.

## Further Reading

Carr-Brion KG, Clarke JRP, and Harding EF (1994) *Sampling Systems for Process Analysers*. Oxford: Butterworth Heinemann.

Chalmers JM and Dent G (1997) *Industrial Analysis with Vibrational Spectroscopy*. Cambridge: The Royal Society of Chemistry.

Chalmers JM (ed.) (2000) *Spectroscopy in Process Analysis*. Sheffield: Sheffield Academic Press.

Maciel GE (1994) NMR in industrial process control and quality control. In: Maciel GE (ed.) *NMR in Modern Technology*. Dordrecht: Kluwer Publishing.

McLennan F and Kowalski BR (eds.) (1995) *Process Analytical Chemistry*. Glasgow: Blackie Academic and Professional.

Workman Jr. J, Veltkamp DJ, Doherty S *et al.* (1999) *Analytical Chemistry* 71: 121R.

## Chromatography

**P M Wiegand**, The Dow Chemical Company, South Charleston, WV, USA

**C P Sears**, The Dow Chemical Company, Freeport, TX, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A process chromatograph is an environmentally suitable automated device that measures one or more chemical components in one or more process streams and presents the results in a usable format. A process chromatographic system includes the associated sample system, utilities, waste handling, and data communication as well as the instrument itself.

In most cases, process chromatography is done in the gas phase. Many improvements have been made to liquid chromatographic systems that make them suitable for process use, but this article will focus only on process gas chromatography (PGC).

Process analyzers, whether chromatographs or other analytical technologies, are often compared with their laboratory counterparts. It is instructive to look at some of the differences between a laboratory gas chromatograph (LGC) and a PGC.

Sample introduction for an LGC is usually done with a manual or automated syringe that pierces a septum. A PGC introduces sample with a valve. The process equivalent of an autosampler is a set of stream-switching valves, which are used to choose the sample point of interest. The PGC must also be able to extract an aliquot of sample from a moving process stream some distance away from the analyzer. This requires a means of preserving the state of the sample and delivering a precise amount to the chromatograph without human intervention. Methods for accomplishing this are discussed in more detail below.

An LGC typically uses temperature programming to enhance separation. The PGC is more likely to be run at an isothermal oven temperature, often using multiple columns, with column-switching valves being employed to enhance the separation.

LGCs use mainly capillary columns, while PGCs still use packed columns. PGC with capillary columns is becoming more popular, but criteria such as detector compatibility, sample capacity, robustness, and temperature effects must be considered.

Detectors for lab and process instruments are very similar, but the digitized data may be handled differently. Lab instruments are designed to produce a visually readable report. Process instruments can usually produce a visual report, but their main function is to provide the data to another computer, which often uses the information to control the chemical process.

To summarize, a LGC is designed to be interactive and flexible and is able to handle a variety of samples. A PGC is designed to be independent and robust and is optimized for a particular type of sample.

## Service and Maintenance of Process Chromatograph

When considering the total cost of a process chromatograph, the cost of ongoing maintenance and service is frequently neglected. However, all process chromatographs need periodic maintenance to run reliably, and so planned maintenance should be included in any cost analysis.

### Consumables and Preventative Maintenance

Replacement of consumables can generally be considered routine for gas chromatography as large

banks of gas cylinders can be set up to minimize the frequency of cylinder changes. Some moving parts, such as valves, will require periodic replacement of seals, inspection for fouling, corrosion, and so forth. A preventative maintenance program can help avoid unexpected problems. Sample cleanup and elimination of oxygen from the carrier gas can help improve column lifetimes.

### **Electronics**

Most electronic repairs consist of replacing entire printed circuit boards, as opposed to individual electronic components and wiring, which was required in early instrumentation. Cam timers have disappeared, to be replaced by digital electronics. Analog electronics have been minimized in order to preserve the integrity of the signal by digitizing as close as possible to the detector.

### **Software**

As in most instrumentation, process analyzers have undergone a fundamental change from hard-wired controllers, switches, and rotary knobs to microprocessor-based software control. With microprocessor control, more advanced functions have become available. Recent years have seen a trend in the incorporation of self-diagnostic routines in the instrument.

PGCs have the capability to be controlled with menu-driven setup screens or a user program. While the maintenance technician will probably not be required to write any programs, enough knowledge to edit programs and understand their function is useful.

Additional software knowledge is required for understanding how to move data from the analyzer to the process computer. Many analyzers have network capabilities and communicate using specific hardware and software protocols. While much of the inner workings of such networking is shielded from the user, a working knowledge of how the network communicates is required for troubleshooting.

### **Control Charts**

An invaluable tool for troubleshooting and maintenance is the instrument control chart. Anything useful can be control charted, such as detector response factors, deviation from expected retention time, etc., but the analysis results from a check gas is one of the most popular items to chart. The most important information a control chart can show is when the instrument needs maintenance or recalibration. While it may seem like a good idea to recalibrate every day, the act of recalibration has error associated with it, and thus overcalibrating can be as much of a problem as undercalibrating. Statistical tools can

be applied to the control chart trend data to show that a significant change has taken place and thus needs to be addressed. In addition, the control chart shows visually when the change began, if it has occurred more than once in the past, cyclical phenomena due to day–night temperature variation, and so forth. This simple, low-tech chart is one of the most underutilized tools in analyzer maintenance.

### **Sample Systems**

It is often said that most process analyzer problems can be traced to the sample system. While some progress has been made in modularizing sample systems, generally the type of maintenance that must be performed has not changed. In many cases, steam-traced lines have been replaced with electrically heated lines, which eliminate maintenance related to steam traps and steam supply. However, filters must still be replaced, valves must still be maintained, and materials of construction that corroded in the past will still corrode.

### **Applications**

Rather than present a specific application, an attempt is made below to define the criteria that would make PGC a suitable technique. After reading the criteria it becomes apparent that certain process points will usually be good candidates for PGC.

#### **Physical Characteristics of the Sample**

One obvious criterion to consider is the physical nature of the sample. Gas-phase streams are the most compatible, but liquid mixtures can also be accommodated if the boiling points are not too high. Small amounts of high boiling materials present in a sample can build up on the column, which will alter or destroy the column's separation ability. This can sometimes be alleviated by including an appropriate backflush step in the method. If there is an extremely wide range of boiling points in the mixture, vaporization of the sample becomes a little tricky as the high boilers may be discriminated against in the volatilization step. Also, a wide boiling range may require a multicolumn approach for obtaining reasonable separation times.

#### **Chemical Characteristics of the Sample**

The chemical nature of the sample mixture should be considered from both a materials of construction point of view and a chromatographic point of view. Extremely corrosive or reactive materials may not be good candidates for gas chromatography. Column stationary phases can react with some chemicals, and so column compatibility is also an issue. Some samples

are not amenable to vaporization; thermal instability can cause components in the sample to degrade.

### Safety

Some sample streams are too toxic, flammable, corrosive, etc. to be removed from the process. Since PGC is necessarily an extractive analysis, the sample must be removable from the process safely for PGC to be suitable. Streams of this nature may be handled more easily with a type of analysis that can be done *in situ*, such as spectroscopy.

### Analyte concentration

PGC can be done successfully at sub-ppm levels, but installation, method development, and maintenance are more difficult at extremely low concentration levels. The sample system, for example, might adsorb low levels of an analyte on surfaces, which can result in memory effects between separations.

## Process Sampling and Sample Systems

The main goal of the sampling system is to extract a sample from the process and deliver it to the analyzer in a manner that is timely, robust, and with the composition unchanged. The manufacturing process from which the chromatograph sample is taken may be under different physical conditions from those for which the chromatograph is designed. Therefore, the sample probe and sample system must change the physical conditions of the sample from process conditions to conditions suitable for the analyzer without changing the analytical information in the sample. The pressure, temperature, flow, and phase must be controlled. The pressure can be controlled by means of pumps and regulators. The temperature can be controlled using heaters and coolers. The flow can be controlled using valves and regulators. The phase can be controlled using heaters, coolers, and vaporizing regulators.

The sample point location is dependent upon the purpose, i.e., control or specification analysis. If the purpose of the analysis is to control the operation of, for example, a distillation column immediately upstream, the concentrations in a downstream storage tank may not be related to controllable process parameters. Better control can be achieved by sampling at the location where the concentration is changing the most rapidly in response to process parameters. Even the overhead and bottoms of a distillation column change very slowly in response to a change in distillation column operation. In this case, the best place to sample the column may be near the critical trays, which have large concentration gradients.

## Basic PGC Hardware

The basic components of a PGC are the same in principle as those for a LGC, but there are differences due to the requirement for unattended and continuous operation in the field.

### Carrier Gas

The carrier gas is identical in both PGC and LGC. It is selected based on the component matrix being analyzed and the detectors being used. The common carrier gases are hydrogen, helium, nitrogen, and argon. Hydrogen is usually best suited for high-speed analyses because its usage leads to minimal degradation in efficiency as carrier linear velocities are increased. Both hydrogen and helium are normally used in the analysis of hydrocarbons using thermal conductivity detectors due to the large difference in thermal conductivity between the carrier gas and the hydrocarbon components of interest. Nitrogen is sometimes used for analyzing air samples or for determining components with high thermal conductivities like hydrogen and helium.

PGC carrier gases are usually supplied by banks of high-pressure gas cylinders, often using an automatic manifold that switches from the 'in-use' bank to the 'reserve' bank when the pressure in the former drops below a preset level. The manifold may also be equipped with an alarm that indicates the reserve bank is in use and that the first bank needs to be replaced.

### Injection and Column-Switching Valves

Since PGCs are required to operate automatically and unattended for extended periods, the sample injection valve in these systems is of critical importance. A PGC sample valve must be more robust than the injection apparatus for a LGC.

Column-switching valves are also used routinely in PGC for backflushing, heartcutting, and so forth for increasing the speed of the analysis. In some applications, the same individual valve can be used for both injection of the sample and column switching if ported correctly.

The two types of sample injection valve are the external sample loop and the internal sample loop. External sample loop valves are typically used for injecting vapor samples. The volume of sample injected can be as large as desired and as small as the physical distance between ports and tubing diameter permit. External loop sizes typically range from ~2  $\mu$ l up to 10 ml.

Internal sample loop valves are characterized by a subtended sample volume inside the moving parts of the valve. Volumes of 0.05 to 20  $\mu$ l are typical and

are limited to the size of the groove or indentation that can be made into the internal parts of the valve. The volume can only be modified by changing the internal parts of the valve. Internal sample injection valves are typically used for introducing much smaller quantities of vapor samples or small aliquots of volatile liquid samples.

There is a variety of valves available for PGC. These are classified according to the mechanism by which the ports are switched and include rotary, piston, slider, and diaphragm valves. The different types of valve all have strengths and weaknesses in terms of operating lifetime, ease of maintenance, and the ability to handle various types of sample reliably. Valveless techniques, based on pressure balancing, are also seeing some application in PGC for column switching.

A rotary valve has partial circumferential grooves on a rotor that turns inside the valve body. The valve body is equipped with ports that emerge radially from the body. When the rotor is moved circumferentially from one position to the other, different combinations of radial ports are connected via the grooves.

Vaporizing liquid injection or piston valves function in a manner similar to syringe injection in LGC. They are typically used to inject moderate to high boiling liquids. Instead of a syringe, the valve contains a grooved piston or valve stem. In the sampling position, the flowing liquid stream passes through the valve and across the valve stem. During injection, the valve stem is inserted through a set of isolating seals and into a heated chamber. The sample aliquot, contained in the groove of the valve stem, is vaporized, and the flowing carrier stream then sweeps it into the column. Sample sizes are typically limited to between a few tenths of a microliter and a few microliters. The large thermal mass of these valves leads to relatively inefficient vaporization compared with laboratory syringe injection systems. This often leads to degradation in separation compared with separations done using laboratory instruments under similar conditions. With capillary columns, splitting of the sample is usually required to prevent column overload.

A slider valve has a flat plate with grooves that connect the valve ports in different combinations, depending upon the plate position. The plate is sandwiched between the body pieces, and the ports project from the body in a perpendicular fashion.

Finally, the diaphragm valve consists of an annular circular groove in the body, alternative ports of which are opened and closed by applying pressure on a diaphragm using either mechanical plungers or simple gas pressure. The diaphragm is sandwiched between the body and the plungers or pressure ports.

By dividing the plungers or pressure ports into two sets and alternatively activating them, two different flow paths are created across the diaphragm, thereby connecting different combinations of ports via the circular groove.

All of the valves are available with air actuators, and some are available with electric actuators. Electric actuators are not usually used in process analysis applications because of both safety and reliability considerations.

### **Packed and Open-Tubular Capillary Columns**

Much process chromatography is still performed using packed columns. Packed columns offer some advantages in terms of ruggedness, stability/reliability, sample capacity, ease of maintenance, range of selectivity, and compatibility with column-switching techniques. However, as process chromatographs decrease in size and as more capillary columns are offered in metal tubing with bonded stationary phases, an increase in the use of capillary columns should be expected.

A packed column typically uses 1–2 mm (called micropacked) or 2–4 mm internal diameter tubing, with some larger columns being used in older gas chromatographs. As its name implies, the packed column is filled with stationary-phase particles. These particles can be coated with a liquid phase, or they can be a stand-alone solid adsorbent.

Stationary-phase particles typically range in size somewhere between 40 and 120 mesh. When coated with liquid, the amount of loading changes the degree of retention on the column. Excessive pressure drops prevent long lengths of packed columns from being utilized. Therefore, separations using packed columns rely heavily on column selectivity rather than column efficiency.

Open-tubular capillary columns typically have an internal diameter ranging from 0.1 up to 0.53 mm. Capillary columns have both advantages and disadvantages over packed columns. Because the carrier gas takes a single flow path through the column, a reduction in the randomness of the mobile-phase path increases the efficiency. In addition, the pressure drop across an open-tubular column is relatively small. Therefore, a long column length can be used. Separations using capillary columns rely mainly on the column efficiency rather than column selectivity. The markedly smaller dimensions of the capillary columns in comparison with packed columns mean that the capillary columns require less carrier gas. Therefore, capillary columns are much more sensitive to temperature, pressure, and flow fluctuations than are packed columns. In addition, the sample

capacity of capillary columns is orders of magnitude smaller than that of packed columns. This can make it more difficult to perform trace analyses on capillary columns, despite their advantage in efficiency and the corresponding sharpness of the peaks.

Short capillary columns can also be used with hydrogen carrier gas to perform high-speed separations. This is due to the fact that the efficiency for short columns does not change dramatically as hydrogen carrier gas flow velocities are greatly increased from the optimal value. Some work has been done with high-speed separations in PGC, but this has been mainly limited to isothermal separations since fast temperature programming (both heating and cooling) is needed to analyze quickly matrices with a wide boiling-point range.

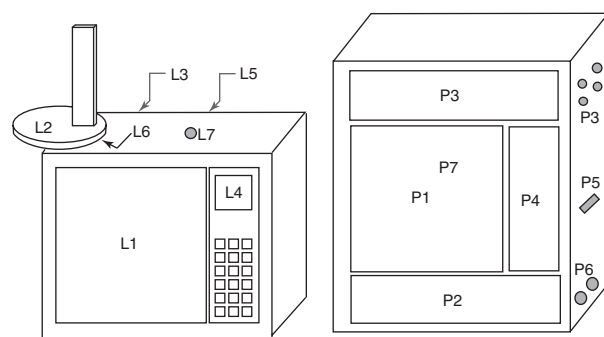
### Chromatographic Column-Switching Systems

In LGC, a single long capillary column is typically used in a temperature-programmed mode to perform the separation. Although this same separation scheme can be used in PGC, the major drawback is the speed of analysis. The solution is to use multiple column trains and column-switching techniques. This allows the chromatograph to be operated in an isothermal mode.

There are predominately four column-switching methods used in PGC. These are (1) backflush, (2) heartcut, sometimes referred to as a 'diverter', (3) trap-bypass, and (4) column selection. Backflush involves reversing carrier flow through the column. It is done in order to remove late-eluting, heavy sample components. This prevents interference in subsequent analysis cycles, prevents damage to the downstream columns, and helps ensure that sample impurities do not accumulate. Late-eluting components can be backflushed to another column that will separate them more efficiently, they can be backflushed directly to another detector to measure them as an aggregate, or they can be discarded to a vent.

Heartcut involves the removal of a chromatographic peak of interest from the background matrix in order to improve the analysis. Heartcut is typically used to transfer a portion of the sample to a second column so that a minor component can be measured without interference. It can also be used in general to reduce tailing from upstream columns and to remove other sample components that might interfere with the analysis.

Trap-bypass involves momentarily trapping part of the sample in a 'lights' column and allowing heavier components to bypass this column. It is typically used to analyze a component matrix with a wide boiling-point range using a single column train. This operation is of less importance today with the advent



**Figure 1** Chromatograph diagrams (L, lab; P, process). L1, column compartment; P2, column and valving compartment; L2, autosampler; P2, sample handling valve(s); L3, pneumatic connections (rear); P3, pneumatic connections (side) and regulator compartment; L4, user keypad and electronics; P4, electronics compartment; L5, data and remote control connections (rear); P5, process computer and local or remote control connection (side); L6, split/splitless injector with septum; P6, process sample input to inject valve and return outlet; L7, detector and vent; P7, detectors (interior).

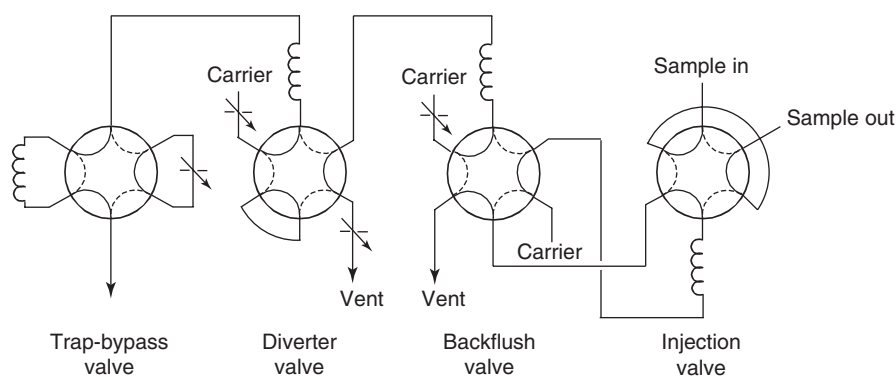
of parallel chromatography using multiple column trains and detectors.

Finally, column selection allows multiple column trains to be merged alternately together while maintaining a constant carrier gas flow. The side that is selected at any moment in time is transferred downstream for detection, while the other stream is sent to vent. This configuration prevents the two column trains from interfering with each other. This is also of less importance due to the use of parallel chromatography.

Valve configurations for performing the column-switching operations noted above using six-port valves are shown in **Figure 1**. Both the deactuated and actuated states are shown for clarity. It should be noted that these valve-switching schemes could also be designed using 10-port valves. In the case of backflushing, a single 10-port valve can be configured to both inject and backflush the sample if desired.

### Detectors

Among the variety of detectors, only the thermal conductivity detector (TCD) and the flame ionization detector (FID) are in broad use in PGC. The flame photometric detector, typically used for measuring trace sulfur containing species, and the photoionization detector, predominately used in environmental monitoring, also see some usage. The variety of detector types available for PGC tends to be limited because of the requirements for robustness and sensitivity to a variety of stream components. In addition, many PGC detectors are not optimized for use with capillary columns.



**Figure 2** Schematic of various valve configurations.

Thermal conductivity detectors in PGC are universal detectors in that they can be used to measure any component (Figure 2). Typically, TCDs are used to measure percent level components, but because it is a concentration-sensitive detector, the TCD can be designed for lower level analyses by optimizing the detector volume and sensing elements. TCDs used in PGC utilize either filaments or thermistor beads as their sensing elements. Filaments have the advantage that they can be operated at higher temperatures, but they are more prone to failure than are thermistors. The sensitivity is mainly dependent on the detector temperature and the thermal conductivity difference between the carrier gas and the component of interest.

FIDs are most often used in PGC to measure trace-level components because of their high sensitivity. The FID responds mainly to hydrocarbon-containing components. Therefore, it cannot be used for measuring some components such as fixed gases. The FID is a mass-sensitivity detector.

Most modern PGCs can handle simultaneous data acquisition from multiple detectors. Acquisition rates for most PGCs are high enough for integrating relatively sharp peaks properly. Only in the case of very high speed analyses are higher acquisition rates needed.

Peak integration in PGC is accomplished using fixed integration (often called fixed gating) or slope-based integration, similar to that used in LGC. Fixed integration uses start/stop gates that are fixed in time. It has the advantage of being very stable but can be difficult to implement with closely spaced peaks. Large errors can also occur if the peak elutes outside the start/stop gate times. On the other hand, slope-based integration is very good at handling closely spaced peaks and minor retention time drifting, but if the slope detection parameters (typically threshold and peak width) have not been carefully adjusted, it has a tendency to stop integrating a component suddenly when the concentration drops below a certain level.

## Ovens

Temperature control in gas chromatography is critical to achieving stable chromatographic retention times. Although most PGCs are installed in climate-controlled shelters, they must still be designed for installation in an outside shelter, where they are only protected from wind, rain, and dust but not from ambient temperature extremes.

Most PGC separations use isothermal ovens. Although temperature programming can also be used in PGC, it is relatively slow and usually requires installation of a special auxiliary oven. In addition, injectors and detectors are often neither closely coupled to the temperature-programmed oven nor independently heated. This can lead to a reduction in efficiency compared with a similar laboratory measurement.

PGC ovens today are heated electrically. Because the instrumentation must be able to be operated in electrically classified areas, care is taken by the manufacturers to ensure that the skin temperature of the heaters does not exceed specified limits. This can limit the speed at which the oven heats up and the maximum operating temperature that can be achieved. PGC ovens also tend to be fairly large in order to accommodate a number of valves, columns, and detectors.

Two types of ovens are used, the thermal mass or airless oven and the air-bath oven. The thermal mass oven consists of a thick metal enclosure that is heated with cartridge heaters. It is slow to heat up from ambient conditions to moderate temperatures but is very well suited to higher-temperature applications. Air-bath ovens rely on forced circulation of heated air throughout them for maintaining a uniform temperature. Although air-bath ovens heat up to moderate temperatures more quickly than do thermal mass ovens, the skin temperature of the heater element often imposes severe restrictions on the upper temperature limit. Fortunately, most PGC



methods operate at oven temperatures that are below 100°C.

### Basic Utilities and Connections

In addition to the sample inlet and sample return lines, PGCs require electrical power, utility air for valve operation and temperature control, carrier gas, fuel gas and air for FIDs, and signal cables. Sometimes, both digital and analog signal cables are used. The carrier gas, fuel gas, and air for the FIDs must be of high purity. In some installations, utility air is also used for air purging for maintaining electrical code conditions. Utility needs must be considered and accommodated early in the design phase of the process chromatographic system because there are often long distances between the nearest utility source and the chromatographic site.

See also: **Chemometrics and Statistics:** Multivariate Calibration Techniques. **Gas Chromatography:** Column

Technology. **Process Analysis:** Maintenance, Reliability, and Training.

### Further Reading

- Annino R (1993) Chromatographs – Gas. In: Lipták BG (ed.) *Analytical Instrumentation*, pp. 87–108. Radnor, PA: Chilton Book Co.
- Annino R and Villalobos R (1992) *Process gas chromatography – Fundamentals and Applications*. Research Triangle Park, NC: Instrument Society of America.
- Cooley RE (1996) Liquid chromatographic (LC) analyzers. In: Sherman RE and Rhodes L (eds.) *Analytical Instrumentation*, pp. 645–672. Research Triangle Park, NC: Instrument Society of America.
- Converse JG (1993) Process chromatography. In: Considine DM (ed.) *Process/Industrial Instruments & Controls Handbook*, 4th edn, pp. 648–657. New York: McGraw-Hill, Inc.
- Grob RL (ed.) (1985) *Modern Practice of Gas Chromatography*, 2nd edn, pp. 707–716. New York: Wiley.

## Electroanalytical Techniques

**A Ivaska and J Bobacka**, Abo Akademi University, Turku-Abo, Finland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Electrochemical methods are not as widely used in process analysis as spectroscopic methods and methods based on measurement of physical rather than chemical properties of the material to be analyzed or characterized. Calibration of the analyzer is always a problem and no general procedure can be recommended or applied because it depends on the particular application and analytical method. Poisoning and deactivation of the active surface of the electrode shortens the lifetime of the sensors. Accumulation of static electricity on the body of the sensor creates extra noise.

The advantages of electrochemical methods are that the signal is electrical and no transformation is required for transferring the information to control units. Some of the methods are rather specific and therefore can be used to detect only the species of interest. Other methods are rather nonspecific but can be used to advantage to detect only certain types of compounds such as ionic species in conductimetric measurements.

All electrochemical methods are based on the interaction of electrical energy and matter. The measurements are done in an electrochemical cell where the sample and at least two electrodes are placed. The electrochemical cell possesses a large variety of concentration-dependent physical characteristics that may be exploited for chemical analysis. The methods are mainly used in analysis of aqueous samples but are also applicable to nonaqueous solutions and gases. In most of the methods one concentration-dependent electrical parameter, like voltage, current, resistance, or charge, is measured while the others are kept constant or manipulated to receive the desired response that correlates to the sample composition.

The analytical methods can be divided into different groups depending on the parameter measured. A schematic presentation of the methods is shown in **Figure 1**. The presentation is based on the recommendation of the International Union of Pure and Applied Chemistry. The methods are first divided between those where either the response depending on a reaction at the electrode or a bulk property of the sample is measured. Measurement of electrical capacity of the cell and conductivity of the sample belong to the latter category. In the conductivity measurements the bulk ohmic impedance of the cell is measured. It is customary to report the conductance of the cell as the function of concentration of ions in the sample. All ions, anions and cations,

methods operate at oven temperatures that are below 100°C.

### Basic Utilities and Connections

In addition to the sample inlet and sample return lines, PGCs require electrical power, utility air for valve operation and temperature control, carrier gas, fuel gas and air for FIDs, and signal cables. Sometimes, both digital and analog signal cables are used. The carrier gas, fuel gas, and air for the FIDs must be of high purity. In some installations, utility air is also used for air purging for maintaining electrical code conditions. Utility needs must be considered and accommodated early in the design phase of the process chromatographic system because there are often long distances between the nearest utility source and the chromatographic site.

See also: **Chemometrics and Statistics:** Multivariate Calibration Techniques. **Gas Chromatography:** Column

Technology. **Process Analysis:** Maintenance, Reliability, and Training.

### Further Reading

- Annino R (1993) Chromatographs – Gas. In: Lipták BG (ed.) *Analytical Instrumentation*, pp. 87–108. Radnor, PA: Chilton Book Co.
- Annino R and Villalobos R (1992) *Process gas chromatography – Fundamentals and Applications*. Research Triangle Park, NC: Instrument Society of America.
- Cooley RE (1996) Liquid chromatographic (LC) analyzers. In: Sherman RE and Rhodes L (eds.) *Analytical Instrumentation*, pp. 645–672. Research Triangle Park, NC: Instrument Society of America.
- Converse JG (1993) Process chromatography. In: Considine DM (ed.) *Process/Industrial Instruments & Controls Handbook*, 4th edn, pp. 648–657. New York: McGraw-Hill, Inc.
- Grob RL (ed.) (1985) *Modern Practice of Gas Chromatography*, 2nd edn, pp. 707–716. New York: Wiley.

## Electroanalytical Techniques

**A Ivaska and J Bobacka**, Abo Akademi University, Turku-Abo, Finland

© 2005, Elsevier Ltd. All Rights Reserved.

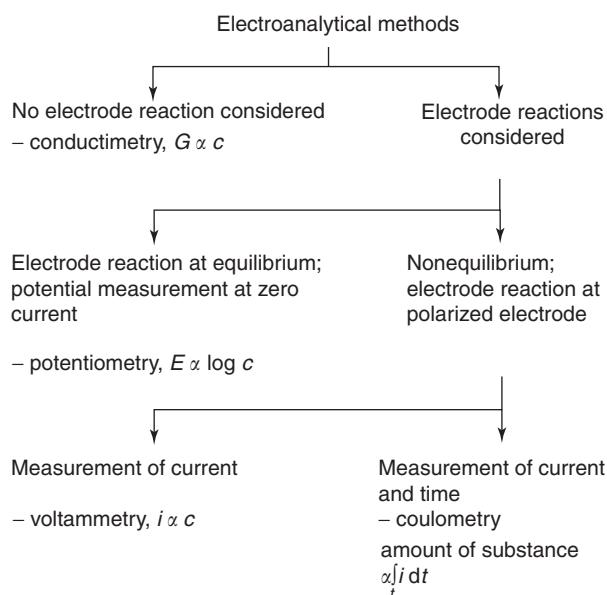
### Introduction

Electrochemical methods are not as widely used in process analysis as spectroscopic methods and methods based on measurement of physical rather than chemical properties of the material to be analyzed or characterized. Calibration of the analyzer is always a problem and no general procedure can be recommended or applied because it depends on the particular application and analytical method. Poisoning and deactivation of the active surface of the electrode shortens the lifetime of the sensors. Accumulation of static electricity on the body of the sensor creates extra noise.

The advantages of electrochemical methods are that the signal is electrical and no transformation is required for transferring the information to control units. Some of the methods are rather specific and therefore can be used to detect only the species of interest. Other methods are rather nonspecific but can be used to advantage to detect only certain types of compounds such as ionic species in conductimetric measurements.

All electrochemical methods are based on the interaction of electrical energy and matter. The measurements are done in an electrochemical cell where the sample and at least two electrodes are placed. The electrochemical cell possesses a large variety of concentration-dependent physical characteristics that may be exploited for chemical analysis. The methods are mainly used in analysis of aqueous samples but are also applicable to nonaqueous solutions and gases. In most of the methods one concentration-dependent electrical parameter, like voltage, current, resistance, or charge, is measured while the others are kept constant or manipulated to receive the desired response that correlates to the sample composition.

The analytical methods can be divided into different groups depending on the parameter measured. A schematic presentation of the methods is shown in **Figure 1**. The presentation is based on the recommendation of the International Union of Pure and Applied Chemistry. The methods are first divided between those where either the response depending on a reaction at the electrode or a bulk property of the sample is measured. Measurement of electrical capacity of the cell and conductivity of the sample belong to the latter category. In the conductivity measurements the bulk ohmic impedance of the cell is measured. It is customary to report the conductance of the cell as the function of concentration of ions in the sample. All ions, anions and cations,



**Figure 1** Classification of electroanalytical techniques.

contribute to the conductance and therefore the conductivity method is not specific. Measurement of the electrical capacitance of the cell is also an impedance measurement and the response that is not specific but a bulk characteristic property is related to the dielectric constant of the sample.

The methods where electrode reactions are considered can be divided into two categories: either the reaction is at the equilibrium or the electrode is polarized and a transient signal is measured. In the former case a thermodynamic property of the cell is used for analytical measurements, i.e., the equilibrium potential of the cell. The Nernst equation can be used to relate the measured potential to the activity of the analyte. The method can be regarded as rather specific because by the design of the electrochemical cell a rather high specificity can be achieved.

The methods where the electrode reaction is not at equilibrium are numerous. In **Figure 1**, only the two most common methods are considered: voltammetry and coulometry. In voltammetry, the electrical current is measured at different potentials and is linearly dependent on the concentration of the analyte. In coulometric methods, the current is integrated over a period of time giving the charge as the parameter measured. The amount of substance of the analyte is, according to Faraday's law, directly related to the charge consumed in the electrode reaction. Coulometric measurements may be performed either in the constant current or constant potential mode. Voltammetric and coulometric methods are not as specific as the potentiometric methods because

all substances that can undergo an electrochemical reaction at the potential of the experiment and that are present in the sample will contribute to the signal measured.

In the conductivity, potentiometric, and voltammetric measurements the response is correlated to concentration or activity of the analyte usually by using calibration curves. In coulometry, however, the charge measured gives directly the amount of substance and therefore no calibration is needed. However, in coulometry the sample is consumed in the measurements and the problem is that the method requires 100% current efficiency to be reliable. Conductimetry and potentiometry are sample nonconsuming methods. In voltammetry, only an insignificant amount of the sample is consumed and therefore the measurement can be repeated. Only in voltammetric stripping methods of very low concentrations of the analyte the amount consumed at the electrode reaction has to be considered if repeated measurements are to be done.

## Conductivity Measurements

The flow of electricity through the solution is accomplished by the movement of ions. The electrical conductivity of a solution is reciprocal of the bulk ohmic resistance and is associated with the various factors retarding migration of ions through the solution. Concentration of all ions present and their mobilities determine the electrical conductivity of a solution. Mobility depends on the charge and size of the ions, temperature of the solution, and the properties of the solvent like the dielectric constant and viscosity.

Measurement of the electrical conductivity of a solution is normally done in two different ways depending on the type of the sample and the purpose of the measurement by using either contact electrodes or electrodeless (inductive) measurement.

In contact electrode measurement two similar electrodes, normally of platinum and coated with platinum black, are immersed in the sample solution. The conductivity cell is a branch of a Wheatstone bridge. In order to avoid polarization of the electrodes an alternating potential of several volts and frequency  $< 1$  kHz is applied to the cell. The solution resistance is measured with the bridge. In the presence of ions in the sample the equilibrium cell potential and the interfacial impedances are unimportant. The measured conductance,  $G$ , can be expressed by the following equation:

$$G = \kappa \frac{A}{l} \quad [1]$$

where  $\kappa$  is the conductivity,  $A$  is the area of the electrode, and  $l$  is the distance between the electrodes. The unit of conductance is siemens, S, which is the reciprocal of ohm,  $\Omega$ . The conductivity is dependent on the concentration of the solute and is given in **Figure 2** for some common electrolytes in water. The common unit of  $\kappa$  is  $\text{S cm}^{-1}$ .

Conductivity measurement with contact electrodes is used in many different process analytical applications. In all of them the main interest is to detect the presence of ionic species in the sample. The effectiveness of distillation and demineralization units is normally controlled by continuous measurement of the conductivity of the outgoing stream. Any defect in the operation of the unit can immediately be seen in the increase in conductivity. The feeding water to steam turbines has to be as clean as possible in order to avoid corrosion in the system that operates under high pressure and at elevated temperatures. The conductivities of feeding water as well as the condensed steam are also continuously monitored by conductivity measurement. Increased conductivity is an immediate indication of corrosion in the system or leakage in the cooling system. This is of extreme importance in nuclear power plants where among other risks corrosion products may also become activated.

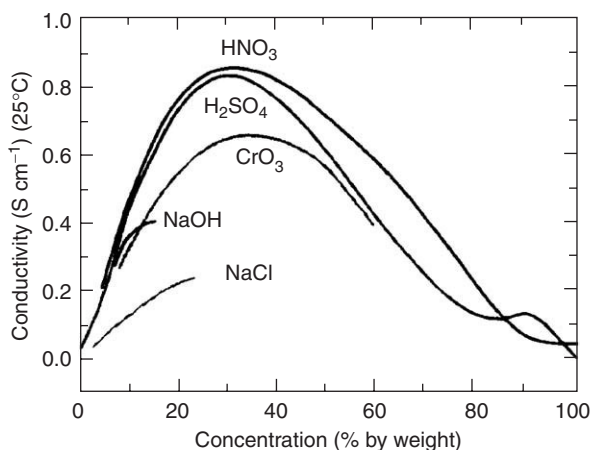
Conductivity measurements are also used in monitoring neutralization and precipitation reactions where the conductivity has its minimum at the equivalence point. Processes where soluble electrolytes are washed from insoluble materials are conveniently monitored by measuring the conductivity of the washing solution. Detection of corrosion and leaks in process heat exchangers are widely monitored by measuring the conductivity of

the streams. Any break in the tubes will result in an increase in the conductivity. Salinity in fresh water and salt water is conveniently determined by conductivity measurement. Pollution sources in fresh-water systems can be located by conductivity measurement because pollution generally increases the content of dissolved material in streams and lakes. The concentration of some gases in air is detected by purging the air through an absorbing solution. Sulfur dioxide is measured by absorbing it in a solution of hydrogen peroxide where sulfur dioxide is oxidized and sulfate and hydrogen ions are produced. Increase in the conductivity can be related to the sulfur dioxide content in the gas flow.

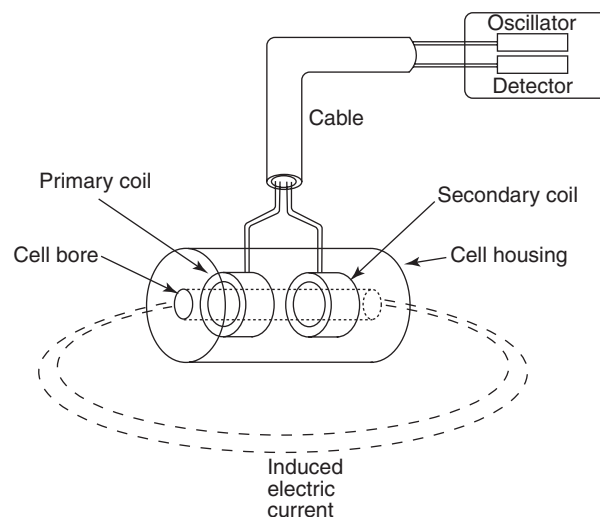
Conductivity or impedance measurements are also used to monitor the curing of an epoxy or a thermoset in the plastics industry. Here, the changes in the electrical properties of the materials give information about the chemical changes that take place in the curing process.

Measurement with the contact electrode cell can be done in samples where the specific conductivity is in the range  $0.1 \mu\text{S cm}^{-1}$ – $10 \text{ kS cm}^{-1}$ .

The electrodeless or inductive conductivity measurement in its essence measures the resistance of a closed loop of solution by the extent to which the loop couples two transformer coils. A schematic drawing of the system is given in **Figure 3**. The primary coil is connected to a transmitter that supplies an alternating voltage. Through the solution loop an alternating voltage is induced in the secondary transformer coil. With constant input excitation signal the induced response will be dependent on the ionic composition of the sample solution. The coils are normally enclosed in the walls of a hollow cylinder



**Figure 2** Conductivity as a function of concentration of some common electrolytes.



**Figure 3** Electrodeless or inductive conductivity measurement.

of inert material such as polyethylene and polytetrafluoroethylene. The cell is immersed in the sample solution that constitutes the loop linking of the coils together.

The electrodeless measurement is used with advantage in applications where the sample solution contains abrasive or fibrous solids and where the conductivity is measured in hot and highly corrosive solutions, or when the sample is an electrolyte of high concentration such as mineral acids, strong bases, and different salts. In all these applications a contact electrode cell cannot be used due to problems of serving the cell and maintaining the electrodes in good working condition. The sensitivity of the electrodeless measurement is low and is restricted to samples of high ionic conductivity. Conductivities in the range of  $1 \text{ kS cm}^{-1}$ – $1 \text{ MS cm}^{-1}$  are normally measured with the electrodeless cell.

## Potentiometric Measurements

In potentiometric measurements the potential of an electrode is measured when no current is drawn through the electrode, i.e., the electrode reaction is at equilibrium. Let us consider the electrode reaction [I]:



where Ox and Red denote the oxidized and reduced forms of the compound, respectively. The number of electrons involved in the electrode reaction is  $n$ . The potential of the electrode,  $E$ , follows the Nernst equation:

$$E = E_0 + \frac{RT}{nF} \ln \frac{a_{\text{Ox}}}{a_{\text{Red}}} \quad [2]$$

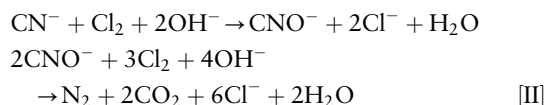
where  $E_0$  is the standard electrode potential for the electrode reaction,  $R$  is the gas constant,  $T$  is the absolute temperature, and  $F$  is the Faraday constant. Activities of Ox and Red are denoted by  $a_{\text{Ox}}$  and  $a_{\text{Red}}$ , respectively.

The potential of a single electrode cannot be measured but potential differences between two electrodes can be measured. Therefore, in potentiometric measurements a reference electrode has to be used. The reference electrode is constructed in such a way that it has a constant potential. The most common reference electrodes are the silver–silver chloride and the calomel electrodes.

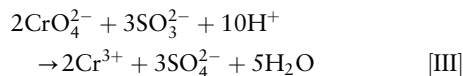
There are several types of indicator electrodes. The simplest is the electrode of some inert material such as platinum, gold, or glassy carbon. These electrodes are used to measure the oxidation–reduction (redox) potential of the solution. The electrode itself does not participate in the electrode reaction but functions only as an inert electron transferring material.

Measurement of the redox potential is nonspecific and all the redox couples in the sample solution contribute to the total potential.

Redox electrodes are mainly used as sensors in processes involving redox couples. An important application is in the metal plating industries where cyanide has to be removed from the plant effluents. Cyanide is eliminated in alkaline solutions by oxidation with chlorine gas in the first step to cyanate ions and then continuing the oxidation process to produce nitrogen and carbon dioxide as the final products:



Both reactions require close pH control and the completeness of the processes is determined by measuring the redox potential of the  $\text{Cl}_2/\text{Cl}^-$  couple at a platinum electrode. When all cyanide has been entirely oxidized the redox potential of the solution is determined by the  $\text{Cl}_2/\text{Cl}^-$  couple. Chromate ions in the process effluents can also be eliminated by reducing them to chromium(III) by sulfite ions:



The completeness of this reaction is also followed by measuring the redox potential of the  $\text{Cl}_2/\text{Cl}^-$  couple. The mineral ilmenite is used as the raw material in the titanium dioxide process. Ilmenite contains also iron(III), which has to be removed before the precipitation of titanium hydroxide in order to avoid co-precipitation of iron(II) hydroxide that would discolor the final product. Iron(III) is reduced with scrap iron to iron(II). The reduction reaction is allowed to go slightly further to reduce also some Ti(IV) to Ti(III) to be sure that the iron(III) has completely been reduced. This reduction step is followed by measuring the redox potential of the solution.

Another type of indicator electrodes are the ion-selective electrodes (ISEs). The potential of an ISE follows the Eisenman–Nicolson equation:

$$E = E_0 \pm \frac{RT}{n_A F} \ln \left( a_A + \sum_i K_{Ai} (a_i)^{n_A/n_i} + D \right) \quad [3]$$

where  $a_A$  and  $a_i$  are activities of the primary and the interfering ion, respectively.  $K_{Ai}$  is the selectivity coefficient for the interfering ion  $i$ . Charges of the ions are  $n_A$  and  $n_i$ .  $D$  is the detection limit. The positive sign is for cations and the negative sign for anions. It should be pointed out that the ISE responds only to

the activity of the free ion, not to the ion in complex form.

ISEs can be classified into several groups depending on the nature of the ion-sensing membrane. The most common and oldest ISE is the glass electrode for measurement of the hydrogen ion activity in the solution. The ion-sensing membrane is made of special glass that is sensitive toward hydrogen ions. The potential of the electrode is a measure of the pH of the sample solution:

$$E = E_0 - 0.059V \cdot \text{pH} \quad [4]$$

The slope of the electrode response, 0.059 V per pH unit, is valid at 25°C. The internal reference electrode in the glass electrode is a silver–silver chloride electrode. The pH electrode is used in most variable applications in process analysis. Most of them are in pH control of processes, effluents, and wastewater treatment plants. The sodium selective electrode is also a glass membrane electrode. It is primarily used in power plants to detect the presence of dissolved solids because sodium is usually a major portion of such contamination.

There are a number of ISEs where the membrane consists of an inorganic salt of the ion to be measured. Most common of such electrodes are fluoride, copper, chloride, and sulfide selective electrodes. In fluoride electrodes the membrane is made of  $\text{LaF}_3$  and the measurement has to be made within the pH range 5.5–8.5 in a special buffer in order to uncomplex the fluoride ion. The main process analytical applications of the fluoride electrode are monitoring the fluoride concentration in fluorinated drinking water systems and measuring the fluorine in process gases from an electrolysis aluminum plant. The gases are absorbed in citrate buffer and fluoride is determined in the buffer by a fluoride ISE. The membrane of a copper ISE is made of  $\text{Cu}_x\text{S}$ , consisting of both  $\text{CuS}$  and  $\text{Cu}_2\text{S}$ . In process analytical applications the electrode is used, for example, in determining copper in metal plating baths. The membrane of a chloride ISE is made of either  $\text{AgCl}$  or  $\text{Hg}_2\text{Cl}_2$  and mixed with some  $\text{Ag}_2\text{S}$  or  $\text{HgS}$ . The chloride ISE is mainly used in the determination of chloride in plant water, natural and drinking water, plant effluents, etc. The membrane of a sulfide selective electrode is made of  $\text{Ag}_2\text{S}$ . Sulfide ISEs have some applications in pulp-making processes where the process solution contains sulfide ions.

Calcium and potassium selective electrodes have a sensing membrane made of poly(vinyl chloride) (PVC) containing a ligand that forms a complex with the cation. The membrane also contains some additional components. Calcium ISEs are mainly

used in the determination of the hardness of different waters. By changing the composition of the membrane it is possible to measure the total water hardness, i.e., the total concentration of calcium and magnesium. Potassium ISEs have their main applications in clinical chemistry.

When the process waters from wood pulping and bleaching are recycled and reused it becomes important to continuously measure the ion content of the pulp suspension. ISEs based on PVC membranes have been optimized for online monitoring of sodium, potassium, and calcium ions in wood pulp suspensions. Potentiometric measurements with ISEs allow rapid measurements with low-cost instrumentation.

One of the main problems with ISEs in process analytical chemistry is the maintenance of the electrode and especially the calibration of the electrode. The short lifetime and poor selectivity are severe limitations of some ISEs.

Oxygen in stack gases is frequently determined by potentiometric measurement (Figure 4). The core of the sensor is a ceramic-like membrane of zirconium dioxide. Two nets of platinum are placed on both sides of the membrane and the potential difference over the membrane is measured. This sensor obeys

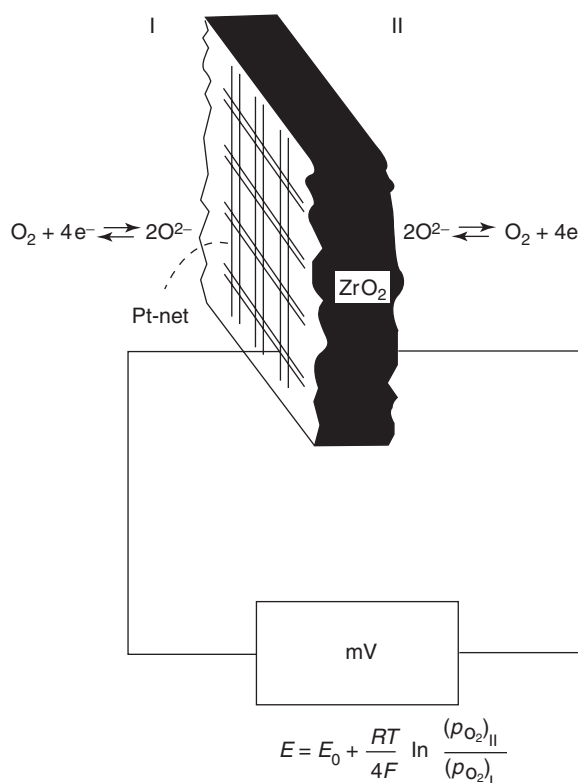


Figure 4 Potentiometric measurement of oxygen gas.



the Nernst equation for the electrode reaction:



where  $\text{O}^{2-}$  is the oxonium ion in the zirconium dioxide crystal structure:

$$E = E_0 + \frac{RT}{4F} \ln \frac{(p_{\text{O}_2})_{\text{II}}}{(p_{\text{O}_2})_{\text{I}}} \quad [5]$$

The partial pressures of oxygen,  $p_{\text{O}_2}$ , on both sides of the membrane are used instead of activities as in electrolyte solutions. When air is used as the reference gas on one side of the membrane, the potential of the sensor will be determined by the partial pressure of oxygen in the sample gas. In some of the latest modifications air as the reference gas has been replaced by solid nickelous oxide in contact with the zirconium dioxide providing a constant concentration of  $\text{O}^{2-}$  ions at the platinum contact electrode.

The zirconium dioxide electrode allows direct (in-line) determination of the partial pressure of oxygen in a steel melt (molten metal bath) within some seconds. The oxygen content is used to calculate the carbon content of the steel.

## **Voltammetric and Amperometric Measurements**

In voltammetric measurements the current at a polarized electrode is measured. The current is produced in the electrode reaction [I] that is forced to take place in one direction by applying an overpotential to the electrode from an external potential source. The current is normally proportional to concentration of the electroactive compound in the sample solution. The electrochemical cell consists of three electrodes: a working electrode at which the reaction takes place, a reference electrode, and a counter electrode. The current in the cell flows between the working electrode and the counter electrode. There are also applications where only two electrodes are used. In polarography, the working electrode is a dropping-mercury electrode (DME). It is not a practical electrode and is only used in some special applications. The main advantages of the DME are the renewable surface and the large overpotential to hydrogen evolution. The most common working electrodes in process analysis are platinum, gold, and glassy carbon.

Voltammetric measurements are done by applying a potential scan to the working electrode. When the potential reaches the range where the analyte will be either reduced or oxidized, a definite electrical current starts to flow. When the current is measured at a constant potential the technique is called

amperometry. All compounds that are reduced (oxidized) at less negative (positive) potentials will also undergo electrode reaction simultaneously with the analyte. This is a severe restriction of the voltammetric methods.

Amperometric methods are used to monitor concentrations of chlorine in drinking water and wastewater treatment plants. The methods are to determine the residual chlorine in water. Chlorine is allowed to react with iodide to produce chloride and iodine. The concentration of the produced iodine is measured amperometrically and is a measure of the chlorine concentration in the sample. Chlorine has also been widely used as a bleaching chemical in pulp-bleaching processes. Chlorine dioxide has replaced chlorine in most of the processes and is readily monitored by amperometry. The best material for the working electrode in these applications is glassy carbon. Measurement of the current at +0.5 V gives the concentration of unreacted chlorine dioxide. At +1.0 V the concentration of chlorite ion as the product of the bleaching reaction can be measured. When concentrations of the residual chlorine dioxide and chlorite ion are measured at a certain stage of the process it is possible to predict the brightness of the final product. In many pulp-bleaching plants hydrogen peroxide has replaced both chlorine and chlorine dioxide as the bleaching chemical. Hydrogen peroxide can also readily be determined by amperometry.

Dissolved oxygen in natural waters and wastewater treatment plants is mainly determined by an amperometric method. The sensing device consists of a gold cathode and silver-silver chloride anode immersed in an electrolyte solution of potassium hydroxide and chloride. The cell is enclosed in a plastic container and is in contact with the sample solution through a thin gas permeable membrane covering the cathode and leaving a thin layer of electrolyte between the cathode and the membrane. The cathode is polarized to a potential where reduction of oxygen takes place. Oxygen in the sample solution penetrates the membrane and is then reduced at the cathode producing electrical current. The current is a measure of the concentration of dissolved oxygen in the sample solution.

Voltammetric methods are also used to monitor concentrations of heavy metals in effluents of metal plants. This is done by anodic stripping voltammetry where the ions of the heavy metals are reduced and plated simultaneously with mercury ions, added to the sample, on a working electrode. The metals will form amalgam in the thin layer of mercury that is formed on the surface of the working electrode. This sampling step is performed at a potential where all

the metal ions of interest are reduced. After this step a positive potential scan is applied and the metals will be reoxidized each at their specific potential. The current produced by each metal is a measure of the concentration of the individual metal ions in the sample as long as the accumulation time and other parameters are kept constant.

The most severe problem with voltammetric methods is the poisoning of the working electrode surface. It can be cleaned by applying potential pulses that are capable of removing the adsorbed reaction products from the surface. The surface can also be cleaned mechanically by bombardment with abrasive grit or by rapid rotation in the presence of plastic abrasion pellets. Strong ultrasound has also been used to clean the electrode surface.

## Coulometric Measurements

In coulometric methods the amount of charge consumed in a reaction is measured. The reaction must be brought to completion. The charge is a measure of the amount of substance consumed in the reaction or reagent produced for complete reaction with the analyte. Coulometric methods rely on Faraday's law:

$$N = \frac{Q}{nF} \quad [6]$$

where  $N$  is the amount of substance consumed or produced by the charge  $Q$ . The fundamental requirement of coulometric analysis is that only a single reaction takes place and that the reaction proceeds with 100% current efficiency. In a coulometric experiment either the potential or the current is held constant or in general the current is integrated over the time of the experiment:

$$Q = \int_t i \, dt \quad [7]$$

One of the most frequently used coulometric process analyzers is the sulfur dioxide measuring instrument. A gaseous sample containing sulfur dioxide gas is first filtered to remove solid particles. It is then led through a solution containing sulfuric acid, potassium bromide, and bromine. Sulfur dioxide reacts with bromine producing sulfate and bromide. The concentration ratio between bromine and bromide is measured potentiometrically with a redox electrode. When the ratio is changed due to sulfur dioxide in the sample gas reacting with bromine, more bromine will be produced in the cell coulometrically from bromide to gain the original concentration ratio between bromine and bromide. The coulometric generation current is the measure of the sulfur dioxide

content in the sample gas. All gases reacting with bromine will interfere. Similar instruments for the determination of hydrogen sulfide and some organic sulfur compounds are used in process analysis. The total sulfur compounds and the different components can be measured by using appropriate scrubbers to remove certain components. The individual sulfur compounds can then be calculated from the differences of the signals.

## Electrochemical Sensors

Chemical sensors are small devices that convert chemical information (concentration, activity, partial pressure) into a measurable signal. A chemical sensor is composed of a recognition part and a transducer part. The recognition part interacts with the target molecules or ions in the sample and the transducer converts the chemical interactions into a measurable signal. If the signal is an electrical signal we talk about electrochemical sensors. Depending on the nature of the electrical signal, electrochemical sensors are further divided into potentiometric, amperometric, and conductimetric sensors. Consequently, the ISEs and the zirconium dioxide oxygen sensor described above may be called potentiometric sensors. Analogously, the amperometric method for determination of dissolved oxygen represents an amperometric sensor. Electrochemical sensors include also a large variety of biosensors, such as enzyme-modified electrodes for determination of glucose, immunosensors, and DNA sensors. Electrochemical sensors are continuously developed and may become important analytical tools for process analysis in the future.

**See also:** **Amperometry.** **Conductimetry and Oscillometry.** **Coulometry.** **Electrogravimetry.** **Ion-Selective Electrodes:** Overview; Glass; Solid-State; Liquid Membrane; Gas Sensing Probes; Water Applications. **pH. Polarography:** Overview. **Process Analysis:** Sensors. **Sensors:** Overview; Amperometric Oxygen Sensors. **Sulfur.** **Voltammetry:** Overview; Anodic Stripping. **Water Analysis:** Industrial Effluents.

## Further Reading

- Bakker E and Telting-Diaz M (2002) Electrochemical sensors. *Analytical Chemistry* 74: 2781–2800.
- Clevett KJ (ed.) (1986) Oxygen measurement: Electrochemical methods of measurement. In: *Process Analyzer Technology*, pp. 220–227. New York: Wiley.
- Clevett KJ (ed.) (1986) Water quality monitoring. In: *Process Analyzer Technology*, pp. 506–557 and pp. 579–587. New York: Wiley.
- Hitchman ML and Berlouis LEA (1995) Electrochemical methods. In: McLennan F and Kowalski BR (eds.)

- Process Analytical Chemistry*, pp. 217–258. Glasgow: Blackie Academic & Professional.
- Ivaska A (1979) Potentiometric sensors in on-line applications. *Proceedings of the Analytical Division of the Chemical Society* 16: 283–288.
- Koch KH (ed.) (1999) Process analytics of gases: Electro-metric methods. In: *Process Analytical Chemistry*, pp. 33–36. Berlin: Springer.
- Koch KH (ed.) (1999) Process analytics of liquid phases: In-line analytics and sensor technology. In: *Process Analytical Chemistry*, pp. 55–57. Berlin: Springer.
- Light TS (1997) Industrial use and applications of ion selective electrodes. *Journal of Chemical Education* 74: 171–177.
- Meites L, Nürnberg HW, and Zuman P (1976) The classification and nomenclature of electroanalytical techniques. *Pure and Applied Chemistry* 45: 81–97.
- Nichols GD (ed.) (1988) Electrochemical process analyzers. In: *On-line Process Analyzers*, pp. 89–116. New York: Wiley.
- Workman J Jr., Koch M, and Veltkamp DJ (2003) Process analytical chemistry. *Analytical Chemistry* 75: 2859–2876.

## Sensors

S B Adeloju, Monash University, Churchill, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

In recent years, the push for improved process control, online quality assurance, higher safety standards, and minimization of environmental impact has extended the interest in online process analysis from the traditional industries, such as chemical manufacturing, mineral processing, and petroleum refining, to other areas, such as biotechnology, food manufacturing, metal smelting, paint production, pharmaceutical industry, water and wastewater treatment. The diverse range of process analysis needed in these industries can be performed either directly with suitable sensors or indirectly in samples taken from the process stream at a controlled rate, with or without sample preparation, such as filtration and addition of reagents. In other words, the required process analysis can be performed online, inline, or offline, and can range from simple physical measurements, such as pressure, temperature, flow rate, density, humidity, and moisture content, to chemical measurements, such as pH, dissolve oxygen, conductivity, and a range of specific chemical composition. Online process analysis is particularly useful for gaining a snapshot of the current state of industrial processes with respect to a selected number of relevant parameters. The resulting signal from the sensor provides an indication of the concentration or other measures of the desired sample constituent(s). This signal may also be used with automated process control systems to maintain optimum process operation. For example, such a controller may be able to detect the difference between a set-point signal and the actual signal obtained from the sensor, and can, in turn, act to initiate a corrective measure(s) to bring the process

operation conditions back to optimum range. Consequently, to gain the best advantage from the results of a process analysis, the sensor must also be able to perform the analysis as rapidly as may be necessary for the corrective measures to be effective. In effect, the chosen sensor is central to the reliable analysis of the process streams and, hence, essential for the effective maintenance of process operation and control. With the aid of advanced computer technology it is possible to accomplish the ultimate in flexible and precise control of processes through combination with online sensors.

In selecting sensors for the analysis of physical and/or chemical components in industrial process streams, several factors must be taken into consideration. **Table 1** lists some of these important factors. It is possible to make some compromise between these factors depending on their relevance to a particular process analysis task.

**Table 1** Important factors in selecting suitable sensors for process analysis

Accuracy	Availability	Capital cost
Corrosion resistance	Credibility	Dynamic/linear range
Fault detection	Fault diagnosis	Field maintainable
Fouling resistance	Hysteresis	Installation cost
Intrinsic safety	Life expectancy	Maintenance requirements
Mechanical robustness	Noninvasive	Power requirements
Precision	Reliability	Response time
Running cost	Safety	Sampling requirements
Shock proof	Selectivity	Self-checking
Size	Stability – long/short term	Throw away (disposable)
Versatility	Weather proof	

Adapted from Clarke JRP (1990) Process analysis. *Analytica Chimica Acta* 238: 3–9.

- Process Analytical Chemistry*, pp. 217–258. Glasgow: Blackie Academic & Professional.
- Ivaska A (1979) Potentiometric sensors in on-line applications. *Proceedings of the Analytical Division of the Chemical Society* 16: 283–288.
- Koch KH (ed.) (1999) Process analytics of gases: Electro-metric methods. In: *Process Analytical Chemistry*, pp. 33–36. Berlin: Springer.
- Koch KH (ed.) (1999) Process analytics of liquid phases: In-line analytics and sensor technology. In: *Process Analytical Chemistry*, pp. 55–57. Berlin: Springer.
- Light TS (1997) Industrial use and applications of ion selective electrodes. *Journal of Chemical Education* 74: 171–177.
- Meites L, Nürnberg HW, and Zuman P (1976) The classification and nomenclature of electroanalytical techniques. *Pure and Applied Chemistry* 45: 81–97.
- Nichols GD (ed.) (1988) Electrochemical process analyzers. In: *On-line Process Analyzers*, pp. 89–116. New York: Wiley.
- Workman J Jr., Koch M, and Veltkamp DJ (2003) Process analytical chemistry. *Analytical Chemistry* 75: 2859–2876.

## Sensors

S B Adeloju, Monash University, Churchill, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

In recent years, the push for improved process control, online quality assurance, higher safety standards, and minimization of environmental impact has extended the interest in online process analysis from the traditional industries, such as chemical manufacturing, mineral processing, and petroleum refining, to other areas, such as biotechnology, food manufacturing, metal smelting, paint production, pharmaceutical industry, water and wastewater treatment. The diverse range of process analysis needed in these industries can be performed either directly with suitable sensors or indirectly in samples taken from the process stream at a controlled rate, with or without sample preparation, such as filtration and addition of reagents. In other words, the required process analysis can be performed online, inline, or offline, and can range from simple physical measurements, such as pressure, temperature, flow rate, density, humidity, and moisture content, to chemical measurements, such as pH, dissolve oxygen, conductivity, and a range of specific chemical composition. Online process analysis is particularly useful for gaining a snapshot of the current state of industrial processes with respect to a selected number of relevant parameters. The resulting signal from the sensor provides an indication of the concentration or other measures of the desired sample constituent(s). This signal may also be used with automated process control systems to maintain optimum process operation. For example, such a controller may be able to detect the difference between a set-point signal and the actual signal obtained from the sensor, and can, in turn, act to initiate a corrective measure(s) to bring the process

operation conditions back to optimum range. Consequently, to gain the best advantage from the results of a process analysis, the sensor must also be able to perform the analysis as rapidly as may be necessary for the corrective measures to be effective. In effect, the chosen sensor is central to the reliable analysis of the process streams and, hence, essential for the effective maintenance of process operation and control. With the aid of advanced computer technology it is possible to accomplish the ultimate in flexible and precise control of processes through combination with online sensors.

In selecting sensors for the analysis of physical and/or chemical components in industrial process streams, several factors must be taken into consideration. **Table 1** lists some of these important factors. It is possible to make some compromise between these factors depending on their relevance to a particular process analysis task.

**Table 1** Important factors in selecting suitable sensors for process analysis

Accuracy	Availability	Capital cost
Corrosion resistance	Credibility	Dynamic/linear range
Fault detection	Fault diagnosis	Field maintainable
Fouling resistance	Hysteresis	Installation cost
Intrinsic safety	Life expectancy	Maintenance requirements
Mechanical robustness	Noninvasive	Power requirements
Precision	Reliability	Response time
Running cost	Safety	Sampling requirements
Shock proof	Selectivity	Self-checking
Size	Stability – long/short term	Throw away (disposable)
Versatility	Weather proof	

Adapted from Clarke JRP (1990) Process analysis. *Analytica Chimica Acta* 238: 3–9.

## Classes and Types of Process Sensors

The two broad classes of sensors that are used for the analysis of industrial process streams are physical and chemical sensors. A selection of each of these classes of process sensors are discussed below.

### Physical Sensors

Despite the considerable development of new sensors and transducers for various chemical substances in recent years, the measurement of physical parameters such as temperature, pressure, density, flow rate, humidity, and refractive index are still very important aspects of industrial process analysis. Some of the

recent developments in the measurement of physical parameters, such as flow rate, temperature, pressure, humidity, turbidity, and moisture content, are discussed below.

**Flow and pressure sensors** A very important parameter that is often measured in most industrial process streams is flow rate. The range of flow sensors that are available for measuring liquid and gas flow rate in industrial process streams is summarized in **Table 2**. Of these, the accuracy, resolution, and cost of ultrasonic flow sensors are attractive and comparable to other flow sensors. Furthermore, it is now possible to use this flow sensor to monitor process

**Table 2** Available industrial process flow sensors

<i>Flow sensor</i>	<i>Measurement approach</i>	<i>Type</i>	<i>Applications</i>
Pressure head flow	Relates a change in pressure drop to a change in flow. Volumetric flow rate is proportional to the square root of the pressure drop	Rotating-vane sealed drum, lobed impeller meters	Liquids and gases for mixture ratio control, distribution of multiple lines, injection of secondary fluid
Positive displacement flow	Uses volumetric displacement principle to determine flow rate		Liquids
Open channel flow	Based on the principles of dilution ratio, or head flow rate relationship or velocity area product		
Ultrasonic flow	Uses piezoelectric transducers mounted either integrally within a flow type section or on an existing pipe together with associated electronics to process the ultrasonic signals. Transmits ultrasonic pulses diagonally through a pipe from the upstream and downstream sides. Measures flow rate from the detected time difference. Measures flow velocity, flow rate, and total flow output. Can also provide flow alarms	Pulse and Doppler	Liquids; can be used with process density analyzers to provide mass flow data
Vortex flow	Uses a circular cylinder across a pipeline and detects frequency of the vortex shedding signal generated. Vortex shedding is detected by sensing the resulting small velocity or pressure pulses	Thermal, ultrasonic, pressure	Liquids and gases
Turbine flow	Uses the proportionality of the fluid flow rate to the speed of revolution of a turbine wheel in a flow stream	Alternating current, direct current	Liquids and gases
Electromagnetic flow	Based on passage of a conductive liquid through a magnetic field, resulting in the generation of a voltage proportional to the average velocity of the liquid slurry		Liquids. Especially for corrosive acids, sewage, slurries, paper pulp. Liquid should have a conductivity of at least $1 \mu\Omega \text{ cm}$

Adapted from Liackman F (1990) *Process Control Engineering* May: 78–84.

flow rate without requiring any special installation at a point in the pipe, as was previously the case. There are many varieties of the clamp-on ultrasonic sensors that can now be used in this way.

A low-cost, low-power-consuming micromachined flow and pressure sensing device has also been reported. This sensor can be used for monitoring pressure and flow rate in clean fluids without particles and without the tendency to coat the channel. The pressure is measured with capacitive or piezoresistive pressure sensors and the flow rate is computed from the pressure drop over a well-defined, hydraulic resistance. Although this device is yet to be used for process analysis in industrial stream, it will gain more use in this area in the future.

**Temperature sensors** Another important parameter that is commonly measured in industrial process streams is temperature. Most chemical and biological reactions in industrial processes are known to be temperature dependent. The reliable and accurate determination of temperature in these cases is of paramount significance. Some of the industrial processes for which temperature measurement is frequently required include water treatment, wastewater treatment, conditioning of drinking water, sewage sludge treatment, and a wide range of chemical and biochemical processes.

The most commonly used temperature sensors for these purposes are thermocouples and resistive temperature detectors (RTDs). The thermocouples are more widely used because they are rugged and relatively inexpensive. These sensors are capable of measuring temperatures between  $-250$  and  $2300^{\circ}\text{C}$ . However, they are not as accurate and corrosion resistant as RTDs.

The temperature of automotive engine exhaust gases is usually monitored with RTDs. A thin platinum-based film is usually used as the sensing element and the resistance of this material increases with increase in temperature. Conduction is minimized by thermally decoupling the sensing element from the sensor body. The exposure of the sensing element to the exhaust gas is via an open housing that acts as a radiation shield. As a result of this unique design, RTDs are very sensitive with voltage drop across the sensing element exceeding  $1\text{ mV }^{\circ}\text{C}^{-1}$  with an accuracy of  $\sim 1.5\%$  and a response time of  $\sim 5\text{ s}$  for a temperature change of  $300\text{--}1000^{\circ}\text{C}$ . These temperature sensors are rugged and easy to set up. They are now commonly used in automotive engine and catalyst diagnostic and control.

In general, most of the available temperature sensors can be used in aqueous and gaseous media, but there are also Teflon-coated temperature sensors

that can be used for measurement in aggressive solutions.

**Turbidity sensor** The online measurement of turbidity and solid content are common practices in wastewater treatment plants. This is particularly significant in biological wastewater treatment, sludge recycling, and in the final effluent where the level of suspended matter provides a useful and immediate indication of the effectiveness of the overall treatment process. The common turbidity sensor used for this purpose is based on the use of an optical infrared scattered light sensor. Unfortunately, the reliance of this sensor on optical measurement is a major disadvantage, as the buildup of particulates on the surface can give incorrect measurement for turbidity and the solid content. The buildup of microorganisms on the surface of this device is a common problem in wastewaters. To remedy this problem, manual cleaning is often required or the use of wiper systems may be employed. However, the sensors equipped with wipers and jutting corners do not always give satisfactory results for the measurement of turbidity and solid content in wastewater treatment plants. Recent developments have led to the introduction of new sensors that incorporate ultrasound cleaning system to minimize or eliminate the buildup of particulate matter and microorganisms on the surface of the sensor. This, in turn, ensures that the sensors last longer and require less maintenance due to the prevention of accumulation of particulate matter from the onset. The ultrasound vibrations also prevent the collection of air bubbles on the surface of the sensor and, hence, ensure reliable measurement of turbidity and solid content. The sensors can measure turbidity within 0 and 4000 FNU and solid content in the range of  $0\text{--}300\text{ g l}^{-1}\text{ SiO}_2$ . These sensors can be used in various applications over several weeks without requiring maintenance.

**Humidity sensors** The measurement of humidity is essential for various industrial process operations, such as in the control of drying plant, ovens, and effluent gases from metal refining furnaces, mineral processing kilns, power plants, chemical plants, and incinerators. Adequate expression of the measured gas pollutant concentrations (on a dry basis) relies on a reliable measurement of the humidity in the process environment.

The range of available ceramic sensors for measuring humidity in industrial process environments can be classified into four categories: (1) ionic, (2) electronic, (3) solid electrolyte, and (4) rectifying-junction types. A solid electrolyte humidity sensor that uses doped strontium cerate has been reported



for high-temperature humidity monitoring. The sensor has been successfully used for reliable measurement in moist air streams and in a combustion rig under conditions where  $>10\%$   $O_2$  is present. Below this oxygen concentration the sensor's response becomes very sensitive to oxygen levels.

**Online moisture sensors** Moisture determination is an important aspect of many industrial processes. The range of available methods for moisture measurement in granular materials and the principles of these methods have been reported in a recent review. Of these, near-infrared (NIR) spectroscopy has gained most use for online moisture measurement in a wide range of granular materials. Also commonly used for moisture measurement in diverse range of granular materials are microwave phase shift (MPS) and microwave attenuation (MA) methods, radio frequency transmission, neutron moderation activation, and direct physical measurement (based on drag force principle) methods. Some of the specific applications of the methods are summarized in Table 3.

Evidently, MPS and MA are the best methods for online measurement of moisture in most materials. However, for materials where most methods are not applicable or have not been attempted, NIR spectroscopy is the method of choice.

### Chemical Sensors

This class of sensors consists of devices that can determine specific chemical composition in industrial process streams by measuring some physical properties, such as voltage, current, electrical or thermal conductivity, absorbance, density, refractive index.

Most of these sensors have gone through considerable developments in recent years and newer versions have been introduced in many cases. The distinct features of these newer or modern chemical sensors that are available for process analysis include improved sensitivity, better precision, better selectivity, faster response time, extended analytical range (component and concentration), capability for simultaneous or sequential multicomponent detection and quantification, improved portability (compactness), and robustness. Some of the newer or modern chemical sensors that are of direct and immediate significance for process analysis in various industrial process streams are described below.

**pH sensors** The acidity or alkalinity of most industrial process streams plays a significant role in the chemical, biological, and physical reactions that take place in the various processes. For this reason, pH is the most commonly measured parameter in most process streams. The solution pH is particularly significant in water and wastewater treatment processes. For example, the continuous online pH monitoring and control are essential in biological treatment of wastewaters where the pH of the water can have significant influence on the activity of the microorganisms that are used in this process. Also, extreme pH conditions in water and wastewater treatment can result in the destruction of metallic components (at  $pH \leq 6.5$ ), solubility of metallic and other components or precipitation of soluble substances. Reliable pH monitoring and control are also essential for various reactions, such as precipitation, neutralization, and detoxification, which are often

**Table 3** Some industrial applications of online moisture sensors

Industrial applications	Online moisture sensors/methods									Best method
	SGM	MPS	MA	NIR	CC	RFT	NMA	LR	DPM	
Carpet		✓	✓							MPS, MA
Cement							✓	✓	✓	DPM
Ceramics		✓	✓	✓					✓	MPS, MA
Coal and minerals	✓	✓	✓				✓	✓	✓	MPS, MA
Food products		✓	✓	✓	✓	✓				MPS, MA
Gases				✓						NIR
Grains		✓	✓		✓	✓				MPS, MA
Paper fibers				✓						NIR
Pharmaceuticals		✓		✓						MPS
Pigments				✓						NIR
Textile						✓				RFT
Tobacco				✓						NIR
Wood products		✓	✓	✓	✓	✓				MPS, MA

Moisture sensors/methods: standard gravimetric method (SGM); microwave phase shift (MPS); microwave attenuation (MA); near-infrared (NIR); capacitance and conductivity (CC); radiofrequency transmission (RFT); neutron moderation activation (NMA); low resolution (LR); direct physical measurement (DPM) based on the drag force principle.

required in many industrial processes. For example, industrial effluents from breweries and dairies often have to be pretreated in a neutralization plant in order to meet the desired pH range of 6.5–8.5, required for discharge into municipal sewer systems.

For the above reasons, most efforts in the design and production of pH sensors have focused on the development of high-precision sensors for industrial process streams owing to the diverse and often difficult conditions that are experienced in these media. These pH sensors are designed to have long-term stability and extended life, based on the use of industrial-grade electrodes with silver/silver chloride reference electrodes incorporated into gel-polymer electrolytes. Some of the industrial-grade pH sensors use modified gel-solid electrolyte, which is free of silver chloride and, hence, can be used in presence of sulfides.

The design of these industrial pH sensors is superior to those of conventional pH electrodes in terms of failures and durability. The online pH sensors are rugged and can withstand harsh industrial conditions. Various types of these pH sensors are available for use in contaminated sewage, drinking water, emulsions and suspensions, municipal and industrial sewage, polluted wastewater, protein- and sulfide-containing media. These sensors can measure pH in the various process streams between 0 and 14. Most of these also have integrated temperature sensors for automatic temperature compensation and a protection guard for the electrode. Some of the pH sensors have additional features, such as glass breakage detection and lightning protection system.

A special type of these pH sensors are designed as valve assemblies and are designed for installation in pipes or vessels. These assemblies allow the pH sensors to be inserted manually and retracted without interrupting the process stream. They are particularly useful for reliable and safe operation under tough process conditions (pressure vessels). The pH sensors can be recalibrated or cleaned when in retracted position without stopping the process.

**Conductivity sensors** A very significant parameter that gives a measure of the concentration of ions in industrial process streams is conductivity. The continuous measurement of conductivity in process streams is a common practice in the monitoring of boiler feed water, brackish water, demineralization, industrial process media, municipal and industrial wastewater, surface waters, and water treatment. Specific areas of application for conductivity sensors include control of drinking water and ultrapure water quality, determination of nonspecific contaminants, monitoring of salt load in wastewater

treatment plants, and process control in food and pharmaceutical industries.

Most of the development in conductivity sensors in recent years has been in the improvement of the design to enable attainment of better performance. In addition to the conventional two-electrode cells, more advanced design has recently resulted in the introduction of a four-electrode cell that gives a considerably better performance, particularly at the higher conductivity ranges. This new sensor design employs two separate electrode pairs, where the current-free voltage electrodes produce a stable and constant reference potential and a potentiostatic circuit is used to regulate the voltage drop at the current electrodes. The two distinct advantages of this new conductivity sensor design are (1) the elimination of measurement errors caused usually by polarization effect, and (2) the avoidance of contact resistance caused normally by contaminated electrodes. More significantly, these new conductivity sensors are capable of measuring conductivity in process streams in the range of  $0.1 \mu\text{S cm}^{-1}$  to  $2000 \text{ mS cm}^{-1}$ . The analytical conductivity ranges of the new conductivity sensors are higher than those of the conventional two-electrode cell with an upper limit of  $\sim 300 \mu\text{S cm}^{-1}$ , but the lower limit for the latter can be as low as  $0.1 \text{ nS cm}^{-1}$ .

**Dissolved oxygen sensors** In addition to pH and conductivity, the online monitoring and control of dissolved oxygen in process streams is of vital importance for efficient operation of many industrial processes, particularly in aquaculture, limnology, water and wastewater treatment, and water pollution control. For example, the efficient operation of biological nutrient removal in wastewater treatment plants requires precise and continuous measurement of dissolved oxygen level in the process stream. The oxygen level in such a process has significant influence on the activity of the microorganisms that are used and, hence, on the efficiency of the nutrient removal process.

The most significant development in this area has been in the design of low maintenance, mechanically more robust dissolved oxygen sensors with rugged foul-resistant membrane. These new oxygen sensors enable more reliable measurement with minimum sample flow and simple automatic calibration in air. The design of this high-performance dissolved oxygen sensor has resulted in an increasing shift from the conventional two-electrode arrangement. Typically, the new three-electrode oxygen sensor consists of a gold working electrode, a silver reference electrode, and a silver auxiliary (current-carrying) electrode. The two distinct advantages of the new oxygen

sensor are improved accuracy and stability. These advantages are due, in part, to the improved stability of the potential of the reference electrode (with no current flow) in the new sensor arrangement.

Some of the reported successful uses of the new dissolved oxygen sensor include monitoring of boiler feed water, drinking water, denitrification of biological sewage treatment, and heavily polluted industrial wastewater. These oxygen sensors can measure dissolved oxygen in process streams at high ( $0\text{--}60\text{ mg l}^{-1}$ ) and low ( $0\text{--}2000\text{ }\mu\text{g l}^{-1}$ ) ranges with response times between 30 and 180 s.

**Phosphorus and nitrogen sensors** A very important area where the use of sensors for process analysis has gained much interest is in wastewater monitoring. In particular, the use of analyzers for monitoring of nutrients (nitrate, nitrite, ammonium, and phosphate) and organic matters is of considerable interest.

Several online analyzers/sensors based on classical colorimetric methods have been reported for the monitoring of phosphate, ammonia, and nitrate concentrations in wastewater treatment plants. These sensing devices have been used to achieve significant improvements in the efficiency of wastewater treatment plants based on the implementation of a rule-based control strategy.

Table 4 provides a summary of some of the sensors/analyzers for online monitoring of ammonium, nitrate, and phosphate. Most of these involve common colorimetric reactions in automated analyzers.

**Chlorine sensors** Online potentiometric and amperometric chlorine sensors are commonly used to monitor chlorine concentration in process streams. The use of a potentiometric chlorine sensor involves adjustment of solution pH and subsequent reaction with lead iodide ( $\text{PbI}_2$ ) to produce iodine. The resulting iodine is detected potentiometrically by an electrode and is related to chlorine concentration. This sensor has been successfully used for continuous monitoring of chlorine levels in water treatment plants.

The use of an amperometric variation of the chlorine sensor has been reported for monitoring chlorine

dioxide in the bleaching stage of a chemical pulp bleaching. This amperometric sensor can be more widely used for monitoring and controlling the bleaching process in the pulp industry by measuring residual chlorine dioxide and chlorite.

**Online mercury sensor** A computer-controlled online sensor for monitoring mercury in streams and aqueous discharges has been reported. This device is based on the use of a dual-wavelength spectrophotometric analyzer. The monitor operates continuously unattended and can measure  $0.5\text{--}10\text{ }\mu\text{g l}^{-1}$  of mercury. It has facility for sounding alarm from a remote location to the central monitoring station. Figure 1 shows a 20-h segment of the continuous monitoring of mercury at upstream and downstream of East Fork Poplar Creek within the Oak Ridge Y-12 plant in Tennessee, USA.

## Process Analyzers

In addition to the physical and chemical sensors, there are a wide range of analyzers that are commonly used for analyzing chemical substances and/or other chemical characteristics in process streams. Specific examples of these analyzers are described below.

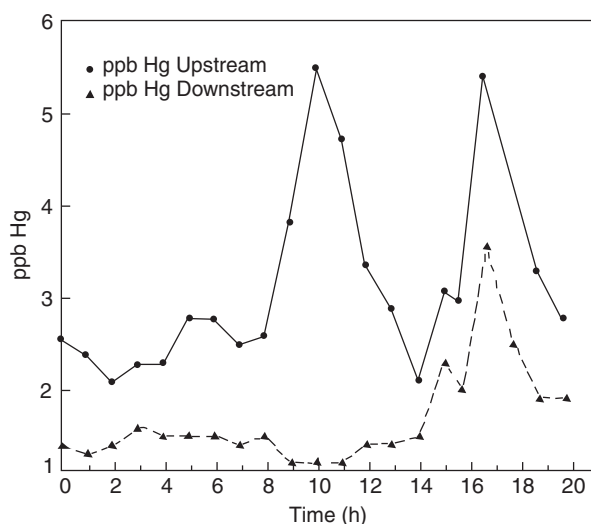
### Process Fourier Transform Infrared Spectroscopy

Fourier transform infrared (FTIR) spectroscopy is particularly useful for online monitoring of polymer composition. Its ability to fingerprint chemical components allows FTIR to determine the constituents of a chemical process. It has been applied to the analysis of solid, liquid, and gaseous process streams. These analyzers can perform complete measurements in as rapidly as one per second. The range of parameters that can be monitored by these analyzers include film thickness, impurities, chemical composition, and polymer curing rate. FTIR can be used for online monitoring of polymer reactions, analyze polymer melt processes, and monitor the gaseous by-products from polymer processing. With the aid of a computer it is possible to determine the concentration of the

**Table 4** Available nutrient analyzers/sensors for monitoring wastewater

Analyte	Chemical method	Alternative
Ammonium	Indophenol blue	Increase pH and measure $\text{NH}_3$ with a gas-sensitive electrode
Nitrate	Reduction of nitrite followed by formation of a purple diazo complex	Absorbance at 205 nm or ion-selective electrode
Phosphate	Molybdenum blue	None

Reproduced from Lynggaard-Jensen (1999) Trends in monitoring of waste water systems. *Talanta* 50: 707–716; © Elsevier.



**Figure 1** A 20-h segment of the continuous monitoring of mercury. (Reprinted with permission from Hinton ER, Jr., Rawlins LK, and Flanagan EB (1987) Development of an online mercury stream monitor. *Environmental Science and Technology* 21(2): 198–202; © American Chemical Society.)

constituents or determine their origins in the chemical process by FTIR spectroscopy. The use of fiber optics with FTIR means that the process stream can be monitored without being directly in contact.

Online process analysis of polymer composition has been accomplished by Rheo-Optical FTIR spectroscopy, which involves the measurement of the stress and strain of an elastomer. This approach utilizes the rapid scanning capabilities of FTIR to obtain vibrational spectroscopy data. Strain-induced crystallization that occurs in polymer gives important structural information that can be obtained by FTIR. The stress/strain process was observed by measuring the C–CH<sub>3</sub> in-plane deformation vibration (1126 cm<sup>-1</sup>) and the ν(C=C) absorption band (1662 cm<sup>-1</sup>), which represents the average polymer thickness reference band. This is particularly useful in the adhesive industry for monitoring the characteristics and composition of the resulting polymer when elastomers are blended in a milling process.

FTIR has also been used for online compositional analysis of polymer blends and copolymers from a plant extruder. Varying concentrations of styrene or acrylic polymer were identified from the resulting FTIR spectra. The online monitoring was accurate when temperature fluctuations were ±5°C. The off-line measurements were found to be less accurate than the online measurements.

Hyphenated FTIR systems have also been used to improve the detection of the by-products of the outgassing of polymer processes. A good example of such hyphenated systems is the use of thermal-desorption

gas chromatography/FTIR for the analysis of the outgassing of polymers. This hyphenated system was successfully used to examine numerous thermoplastics, such as polystyrene, polystyrene copolymer, polyacrylate, polycarbonate, polyamide, and polyacetal. An extended version of the hyphenated system incorporates a mass spectrometer and permitted online analysis of the outgassing by-products of thermal degradation of polymers, as well as the elucidation of evolved species.

### Process Near-Infrared Reflectance Spectroscopy

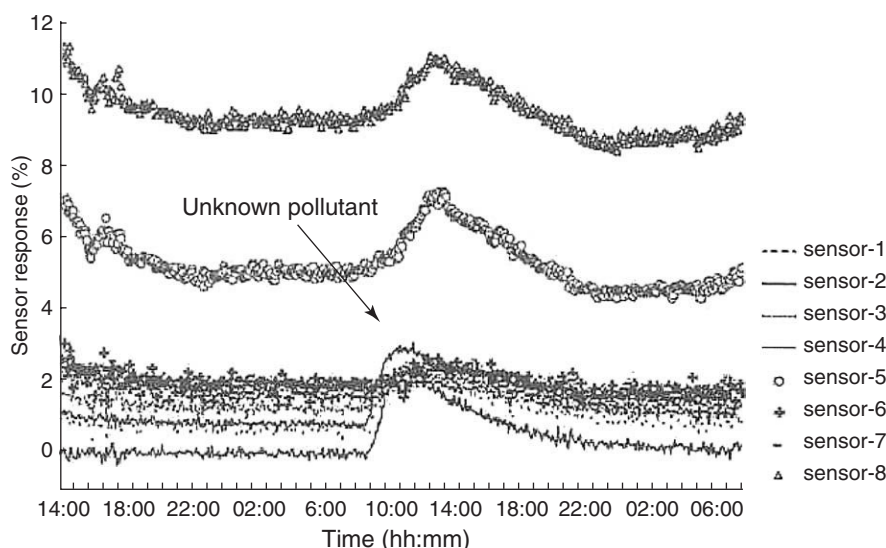
The use of NIR reflectance has gained considerable interest for process analysis in the food, pharmaceutical, petrochemical, and other chemical industries. This method requires little or no sample preparation and involves measurement of the reflectance of a liquid or solid sample relative to that of a standard in the same wavelength series. As illustrated in Table 3, process NIR has gained considerable use for moisture determination in coal, cosmetics, and detergent powders, as well as for protein contents of cereal and grain, and hydrogenation of unsaturated fats and oils.

The use of NIR with fiber optics has also been reported for online monitoring of fuel quality parameters, such as research octane number, motor octane number, olefin, naphthene and aromatics content of gasoline. The online analysis of the fuel quality parameters was obtained without loss of precision and accuracy when compared with the traditional motor test.

### Dispersive and Nondispersive Infrared Process Analyzers

In addition to FTIR and NIR, three categories of infrared analyzers have gained considerable interest in the process analysis of liquid and gaseous streams. These infrared analyzers include (1) dispersive, (2) nondispersive, and (3) bandpass optical filter. The two commonly used categories are the dispersive and nondispersive infrared analyzers.

The dispersive infrared analyzer, which is similar to the double-beam spectrometer, is capable of analyzing components of liquid and gaseous process streams. In contrast, the nondispersive infrared analyzer, which is a filter photometer, is more suited to the selective analysis of components of gaseous process streams. In general, the design and operation of nondispersive infrared analyzers are simpler than those of dispersive infrared analyzers. Also, the use of filtering techniques with the nondispersive infrared analyzers increases the selectivity of the analysis, but this can also reduce sensitivity. These infrared analyzers are commonly used for the analysis of



**Figure 2** Detection of an unknown discharge in the wastewater influent with the conducting polymer sensor array. (Reproduced with permission from Bourgeois W, Gardey G, Servieres M, and Stuetz RM (2003) A chemical sensor array based system. *Sensors and Actuators B* 91: 109–116; © Elsevier.)

gaseous substances, such as carbon monoxide, carbon dioxide, nitrogen oxides, sulfur dioxide, cyclohexane, vinyl acetate, ammonia, dioxane, Freon, and carbon tetrachloride in industrial processes and in ambient air monitoring station.

### Other Process Analyzers

The range of other available process stream analyzers includes process gas chromatograph, process high-performance liquid chromatograph, and online potentiometric analyzers. Process gas chromatograph is different to those commonly used in laboratories, particularly in its design to provide analytical results at a response time that is comparable to changes taking place within an industrial process stream. Also, it operates continuously, analyzing one or more process stream components. This analyzer is commonly used in the petrochemical and refining industries.

Interest in process high-performance liquid chromatography has only developed in recent years, but is gaining use in food and pharmaceutical industry. However, the resolution of problems associated with sampling, maintenance of column condition, and reliability are more difficult to accomplish than for process gas chromatography.

Online analysis of inorganic ions and dissolved gases has also been accomplished by use of potentiometric analyzers that employ specific or selective ion electrodes for continuous analysis of the individual components. This type of analyzer is commonly used for continuous online monitoring of chloride, cyanide, and fluoride in water, effluent, and

other process streams. The unique advantage of these analyzers is that their measurements are often not affected by the color and presence of suspended materials in the process streams. Also, their accuracies are comparable to those obtained in laboratory environments.

### New and Emerging Sensor Technologies

A much wider range of physical and chemical sensors will be available for process analysis in the future. Currently, there are several new sensors that have been developed, such as biosensors, conducting polymer sensors, electronic noses, and electronic tongues, to name a few, that will gain more interest in process analysis in the future. A recent application of one of these new sensors involves the use of a conducting polymer sensor array coupled with a headspace generating flow cell for a successful continuous monitoring of sudden changes in wastewater quality over a 12-month period. **Figure 2** shows that this sensor array is capable of detecting a pollution episode in wastewater, as evident from the rapid increase in the responses of four of the eight sensors in the array. The observed increases reportedly correspond to a strong petroleum smell in the wastewater at the sewage works.

This is only one of the many new dimensions of process analysis that will be opened up with the introduction of the new and emerging sensor technologies in this area in the future.

See also: **Process Analysis:** Chromatography; Electro-analytical Techniques. **Sensors:** Overview.

## Further Reading

- Cancilla PA, Barrette P, and Rosenblum FF (2003) On-line moisture determination of ore concentrates – ‘a review of traditional methods’ and ‘introduction of a novel solution’. *Minerals Engineering* 16: 151–163.
- Carr-Brion KG (1990) Sampling systems for on-line analysis of materials that are difficult to handle. *Analytica Chimica Acta* 238: 11–16.
- Docquier N and Candel S (2002) Combustion control and sensors: A review. *Progress in Energy and Combustion Science* 28: 107–150.
- Guillemin CL (1990) *Analytica Chimica Acta* 238: 17–33.
- Hassen MA, Clarke AG, Swetnam MA, Kumar RV, and Fray DJ (2000) *Sensors and Actuators B* 69: 138–143.
- Hauptmann P, Lucklum R, Puttmer A, and Benning B (1998) Ultrasonic sensors for process monitoring and chemical analysis: State-of-the-art and trends. *Sensors and Actuators A* 67: 32–48.
- Ivaska A, Forsberg P, and Heikka R (1990) Application of an amperometric sensor to in-line monitoring of pulp bleaching with chlorine dioxide. *Analytica Chimica Acta* 238: 223–229.
- Jansen JAJ and Haas WE (1987) *Analytica Chimica Acta* 196: 69–74.
- Jansen JAJ, Haas WE, Neutkens HGM, and Leenen AJH (1988) *Thermochimica Acta* 134: 307–312.
- McPeters HL (1990) On-line analysis of polymer melt processes. *Analytica Chimica Acta* 238: 83–93.
- Oosterbroek RE, Lammerink TSJ, Berenschot JW, et al. (1999) *Sensors and Actuators* 77: 167–177.
- Parisi AF, Nogueiras L, and Prieto H (1990) *Analytica Chimica Acta* 238: 95–100.
- Pedersen KM, Kummel M, and Soeberg H (1990) *Analytica Chimica Acta* 238: 191–199.
- Siesler HW (1985) *Applied Spectroscopy* 39: 761–765.
- Traversa E (1995) Ceramic sensors for humidity detection: The state-of-the-art and future developments. *Sensors and Actuators B* 23: 135–156.
- WTW Tetracon® 4-Electrode (2004) Wissenschaftlich-Technische Werkstätten GmbH & Co. Weinheim, Germany.

## Acoustic Emission

**R M Belchamber**, Process Analysis and Automation Ltd., Farnborough, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A transient elastic wave spontaneously generated by abrupt localized changes of strain within a body is the definition of acoustic emission given by Eitzen and Wadley. Acoustic emission monitoring is the name of the associated technique that finds application in a range of areas from nondestructive testing (NDT), machinery condition monitoring to process analytical technology.

Serious work on acoustic emission monitoring started in the 1950s. This mainly featured NDT of composite materials, storage vessels, and aircraft structures.

Since the 1990s, there have been significant developments in the application of acoustic emission to monitoring processes including fluidized beds, high-shear granulation, crystallization, entrainment of solid particulates, and liquid droplets in gas streams. All of these are dynamic processes involving at least two phases.

Acoustic emissions are broadband signals whose frequencies range from a few hertz to several

megahertz. However, acoustic emission is usually monitored at ultrasonic frequencies (50 kHz–1 MHz). This leads to a signal-to-noise ratio benefit as high frequencies are attenuated more strongly than low frequencies, which results in low background signals.

The key features of acoustic emission that make it attractive for process monitoring are:

- *Noninvasive* – this makes installation very easy. The transducers are attached to the outside of the vessel. There is no need to breach the wall of a vessel (in the case of a pressure vessel this is a particularly costly procedure); there are no sensor contamination issues.
- *Real-time information* – acoustic emission responds to real-time events. This means that acoustic emission is very suited to control applications.
- *Intrinsically safe* – acoustic emission transducers are often rated intrinsically safe so that they can be installed in a hazardous area without the need for purged or explosion-proof enclosures. This makes deployment easy in locations such as petrochemical plants.

## Theory

### Acoustic Emission Sources

In NDT applications, the main source of acoustic emission is the rapid release of energy during crack



See also: **Process Analysis:** Chromatography; Electro-analytical Techniques. **Sensors:** Overview.

## Further Reading

- Cancilla PA, Barrette P, and Rosenblum FF (2003) On-line moisture determination of ore concentrates – ‘a review of traditional methods’ and ‘introduction of a novel solution’. *Minerals Engineering* 16: 151–163.
- Carr-Brion KG (1990) Sampling systems for on-line analysis of materials that are difficult to handle. *Analytica Chimica Acta* 238: 11–16.
- Docquier N and Candel S (2002) Combustion control and sensors: A review. *Progress in Energy and Combustion Science* 28: 107–150.
- Guillemin CL (1990) *Analytica Chimica Acta* 238: 17–33.
- Hassen MA, Clarke AG, Swetnam MA, Kumar RV, and Fray DJ (2000) *Sensors and Actuators B* 69: 138–143.
- Hauptmann P, Lucklum R, Puttmer A, and Benning B (1998) Ultrasonic sensors for process monitoring and chemical analysis: State-of-the-art and trends. *Sensors and Actuators A* 67: 32–48.

- Ivaska A, Forsberg P, and Heikka R (1990) Application of an amperometric sensor to in-line monitoring of pulp bleaching with chlorine dioxide. *Analytica Chimica Acta* 238: 223–229.
- Jansen JAJ and Haas WE (1987) *Analytica Chimica Acta* 196: 69–74.
- Jansen JAJ, Haas WE, Neutkens HGM, and Leenen AJH (1988) *Thermochimica Acta* 134: 307–312.
- McPeters HL (1990) On-line analysis of polymer melt processes. *Analytica Chimica Acta* 238: 83–93.
- Oosterbroek RE, Lammerink TSJ, Berenschot JW, *et al.* (1999) *Sensors and Actuators* 77: 167–177.
- Parisi AF, Nogueiras L, and Prieto H (1990) *Analytica Chimica Acta* 238: 95–100.
- Pedersen KM, Kummel M, and Soeberg H (1990) *Analytica Chimica Acta* 238: 191–199.
- Siesler HW (1985) *Applied Spectroscopy* 39: 761–765.
- Traversa E (1995) Ceramic sensors for humidity detection: The state-of-the-art and future developments. *Sensors and Actuators B* 23: 135–156.
- WTW Tetracon® 4-Electrode (2004) Wissenschaftlich-Technische Werkstätten GmbH & Co. Weinheim, Germany.

## Acoustic Emission

**R M Belchamber**, Process Analysis and Automation Ltd., Farnborough, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A transient elastic wave spontaneously generated by abrupt localized changes of strain within a body is the definition of acoustic emission given by Eitzen and Wadley. Acoustic emission monitoring is the name of the associated technique that finds application in a range of areas from nondestructive testing (NDT), machinery condition monitoring to process analytical technology.

Serious work on acoustic emission monitoring started in the 1950s. This mainly featured NDT of composite materials, storage vessels, and aircraft structures.

Since the 1990s, there have been significant developments in the application of acoustic emission to monitoring processes including fluidized beds, high-shear granulation, crystallization, entrainment of solid particulates, and liquid droplets in gas streams. All of these are dynamic processes involving at least two phases.

Acoustic emissions are broadband signals whose frequencies range from a few hertz to several

megahertz. However, acoustic emission is usually monitored at ultrasonic frequencies (50 kHz–1 MHz). This leads to a signal-to-noise ratio benefit as high frequencies are attenuated more strongly than low frequencies, which results in low background signals.

The key features of acoustic emission that make it attractive for process monitoring are:

- *Noninvasive* – this makes installation very easy. The transducers are attached to the outside of the vessel. There is no need to breach the wall of a vessel (in the case of a pressure vessel this is a particularly costly procedure); there are no sensor contamination issues.
- *Real-time information* – acoustic emission responds to real-time events. This means that acoustic emission is very suited to control applications.
- *Intrinsically safe* – acoustic emission transducers are often rated intrinsically safe so that they can be installed in a hazardous area without the need for purged or explosion-proof enclosures. This makes deployment easy in locations such as petrochemical plants.

## Theory

### Acoustic Emission Sources

In NDT applications, the main source of acoustic emission is the rapid release of energy during crack

propagation as solid materials are subjected to tensile stress. The amplitude of the acoustic wave (stress wave) is proportional to the area of the new fracture surface, the velocity of fracture, and the square of the yield stress.

In process analysis, the sources of acoustic emission are typically the impact of particles on the walls of a vessel (e.g., fluid bed, granulator, pipe), the deformation or fracture of particles, and the oscillation of gas bubbles.

A single particle impact produces a single burst of acoustic emission. This is referred to as burst emission. These are also the typical acoustic emission signals generated in NDT applications. Cracks propagate in a jerky manner, every time they move they produce a burst of acoustic emission. These bursts are normally well time-resolved, i.e., the length of the burst is short compared with the interval between bursts. This has led to the development of a number of measurement techniques based on event counting or timing. However, these techniques are usually unsuited to process analytical work. In these applications, such as monitoring a fluid bed reactor, the impacts are so frequent that it is impossible to resolve individual events. This type of acoustic emission is referred to as continuous emission (Figure 1).

The acoustic emission arising from a particle impact is related to its velocity, mass, and how the particle interacts with its target. When a particle impacts

with a surface the interaction may be elastic, plastic, or cause fracture of the particle.

Elastic impact is a nondamage process. The intensity of the acoustic emission is controlled by the kinetic energy of the impact.

Plastic impact is similar to elastic impact except that the surface or particle is modified in shape as a result of plastic deformation.

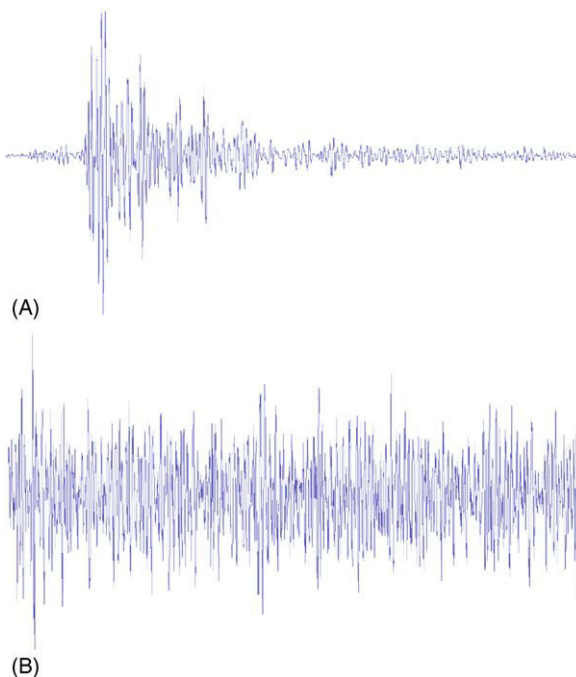
Acoustic emission caused by single particle impact is reasonably well understood. It is possible to quantitatively analyze acoustic emission signals to predict particle size. The analysis is based on a deconvolution of the acoustic emission source function from the measured acoustic emission signal. The measured acoustic emission signal  $V(t)$  is a convolution of the acoustic source  $S(t)$ , the propagation function through the metal plate  $G(t)$ , and the detector response function  $D(t)$ , where  $*$  represents convolution:

$$V(t) = S(t) * G(t) * D(t)$$

All the information of interest is contained in the acoustic source function. The theory works well for low-velocity impacts where the collisions are elastic but may be extended with some modifications to include plastic deformation. The method is nonempirical (requiring no independent calibration). However, the measurement requires a calibrated broadband sensor, the particles strike a well-characterized metal plate, at a known velocity, and the burst acoustic emission signals must be separated in time by  $\sim 1$  ms. In process analysis, it is unlikely that all (or any) of these requirements can be met.

Acoustic emission from large assemblies of particles, such as encountered in a fluidized bed or granulator is not well understood and there is virtually no published research. In contrast, there is a rich history of research on acoustic emission producing sands. These sands are known as booming and squeaking sands and occur naturally in deserts and beaches. They have been known of for thousands of years. Marco Polo in 1295 described evil desert spirits 'at times filled the air with the sounds of all types of musical instruments and also of drums and the clash of arms'. These sounds are produced when the sands are stressed or put into motion by an avalanche. Squeaking sands usually produce acoustic emission in the audible range 100 Hz–1 kHz. Booming is described as a 'beat frequency' in the 1–10 Hz band. There has been much speculation on the mechanisms responsible for the production of this acoustic emission.

Both observational and laboratory research have failed to come up with a complete or satisfactory theory explaining this phenomenon. The majority of evidence supports the theory that squeaking is caused



**Figure 1** Types of acoustic emission – (A) burst and (B) continuous.

by intergranular friction. The combination of narrow particle size distribution, high degree of sphericity, and resistance to shear are thought to be critical factors leading to 'squeaking'. Booming sands are characterized by unusually smooth surface textures of the granules.

Because of the lack of detailed theory and the difficulty of applying what is known to practical situations most process acoustic emission work has focused on the development of highly empirical models. The basis of these is the comparison of acoustic emission signals with those measured under known conditions.

### Acoustic Wave Propagation

The link between the source and the signal detected by the sensor is the acoustic wave. Much of the complexity seen in an acoustic wave is generated as the wave travels through the medium. Any understanding of the acoustic emission signal requires knowledge of the acoustic waves.

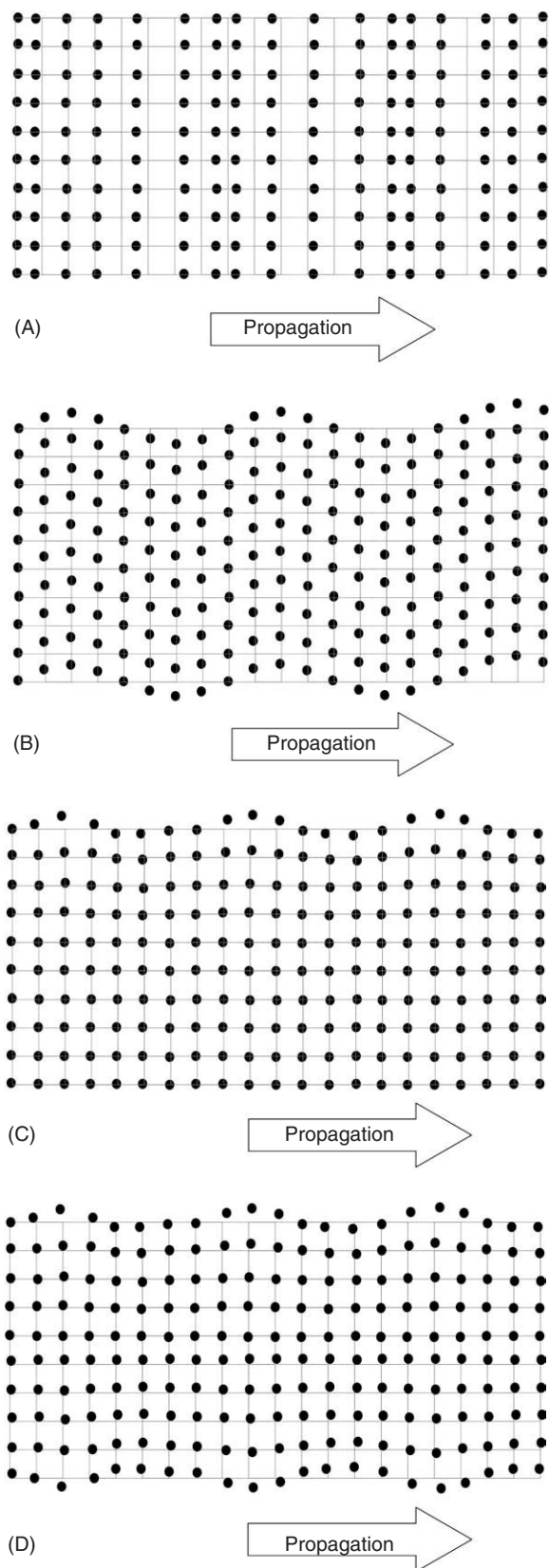
Acoustic waves are elastic waves that propagate through gases, liquids, and solids. In liquids and gases, longitudinal (compressional) waves are the only mode of propagation. The motion of the supporting medium is in the direction of propagation (Figure 2).

Bulk solids also support transverse (shear) waves. In these waves, the motion of the transmitting material is transverse to the direction of propagation.

In a bounded medium, such as thick metal plates, surface waves (Love waves and Rayleigh waves) occur. If the thickness of the plate is of the order of a few wavelengths, Lamb waves (plate waves) can propagate. These can be either symmetric or asymmetric. Generally, if the solid has moderate symmetry, unique waveforms can propagate, for example, tube waves.

The velocities of these propagation modes are different. Longitudinal waves are the fastest with about twice the velocity of transverse waves. The result is that the sensor detects a rather complex waveform. In process analytical applications this situation is somewhat relaxed due to these measurements being made in what is termed a 'diffuse field'. This arises for two reasons: (1) it is impossible to resolve individual acoustic events and (2) acoustic emission waves mix due to reflections from interfaces. This means that within a small area there is no real difference in the measured acoustic emission signal, no matter where the acoustic emission sensor is mounted or its orientation.

The main characteristics of an acoustic wave are its propagation mode ( $m$ ), velocity ( $v_m$ ), its amplitude ( $A$ ), and frequency (Table 1).



**Figure 2** Modes of acoustic propagation: (A) longitudinal or compressional wave; (B) transverse or shear wave; (C) surface or Rayleigh wave; and (D) plate or Lamb wave (symmetric mode).

**Table 1** Acoustic properties of some representative materials commonly found in the process environment

Material	Longitudinal velocity ( $m s^{-1}$ )	Transverse velocity ( $m s^{-1}$ )	Acoustic impedance ( $kg m^2 s^{-1}$ )
Aluminum (typical)	6320	3130	17.1
Steel (typical stainless)	5600	3120	46.5
Silica (fused)	5570	3520	14.5
PVC (hard)	2395	1060	3.35
Polyethylene (typical)	2667	No shear component	2.94
Motor oil (SAE20)	1740	—	1.51
Water (20°C)	1480	—	1.48
Ethanol	1180	—	0.93
Air	343	—	0.0004

The exact values of the quantities depend on the purity, method of production, and temperature.

The velocity of the wave is related to the density ( $\rho$ ) and the elastic constant ( $c_m$ ) of the medium through which it is propagating (in the equation shown below). The elastic constant is unique to the mode of propagation and to the material. For example, in liquids  $c_m$  is the adiabatic bulk modulus ( $B$ ):

$$v_m = \sqrt{\frac{c_m}{\rho}}$$

Acoustic impedance is an important parameter. When an acoustic wave passes from one medium to another, some of the energy is transmitted and some is reflected. The acoustic impedance controls the ratio of transmitted to reflected energy. In acoustic emission, monitoring the acoustic impedance explains the need for an acoustic coupling agent between the sensor and the surface on which it is mounted. Without this the detected signal will be very low, as most of the energy will be internally reflected at the wall.

The acoustic impedance of a material ( $Z_m$ ) is related, as is velocity, to density and elastic constant:

$$Z_m = \sqrt{\rho c_m}$$

The reflected intensity ( $I_r$ ) and the transmitted intensity ( $I_t$ ) (when the acoustic wave strikes an interface relates to acoustic impedance of both media ( $Z_1, Z_2$ )) are

$$I_r = \frac{(Z_1 - Z_2)^2}{(Z_1 + Z_2)^2}$$

$$I_t = \frac{4Z_1 Z_2}{(Z_1 + Z_2)^2}$$

## Instrumentation

The heart of an acoustic emission system is the sensor. The sensors are usually made of piezoelectric ceramic material such as PZT (lead zirconium titan-

ate). These materials are not naturally piezoelectric but are made piezoelectric by 'poling'. Poling consists of heating the material to a temperature in excess of its Curie temperature, applying a strong electrostatic field and then cooling it with the electrostatic field still applied. Above the Curie temperature the electrostatic dipoles become mobile, they align in the electrostatic field and on cooling lock into place. The Curie temperature for PZT ceramics is 328–365°C. After poling, when these materials are subjected to a physical stress the surfaces become charged. If they are set into vibration by an acoustic wave stress they produce an AC signal.

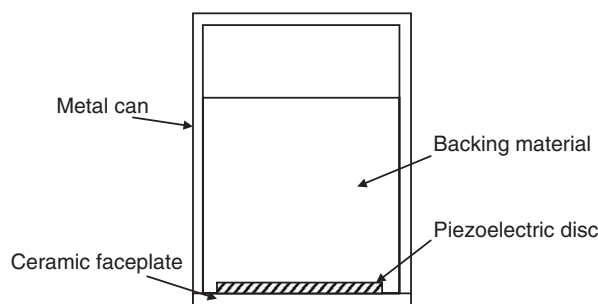
The Curie temperature represents the maximum temperature that a piezoelectric material can be exposed to. Above this temperature the electrostatic dipoles misalign and the piezoelectric properties are lost permanently.

The sensor consists of a small disc of piezoelectric coated on both sides with silver to make electrical contact. This disc is mounted in a metal can behind a ceramic faceplate for mechanical protection (Figure 3).

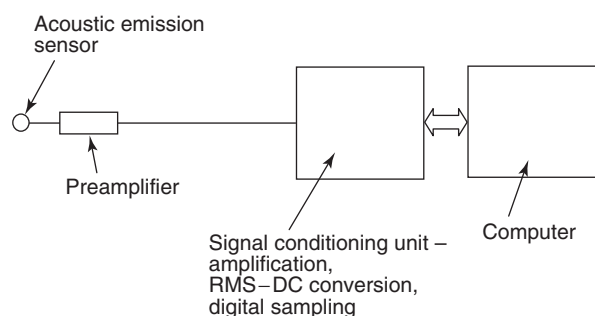
There are two main types of transducer. The most sensitive is the resonant transducer, which has its maximum sensitivity at the resonance frequency. Typical sensors have resonances in the range 70–600 kHz.

A second kind of transducer is the broadband sensor. In construction it is similar to a resonance sensor except that it has mechanical damping. These respond over range of several hundred kilohertz.

The sensor is attached to the outside wall of the vessel to be monitored. A small amount of acoustic coupling agent is placed between the faceplate of the sensor and the vessel surface. This provides impedance matching between the surface and the transducer, which greatly increases the sensitivity and reproducibility of the measurement. For temporary installations, grease or a proprietary acoustic coupling agent is used. For permanent installations an adhesive is used as it provides both acoustic coupling and mechanical strength.



**Figure 3** Acoustic emission sensor.



**Figure 4** Acoustic emission monitoring system.

A preamplifier is located close, usually within 2 m, to the acoustic sensor. This provides a low electrical impedance output so that the signal can be driven down a long coaxial cable to the signal-conditioning unit. Preamplifiers usually have a gain of 40 dB (Figure 4).

### RMS Conversion

The signal from a resonant transducer resembles an AM radio signal. It has a carrier wave, at the resonance frequency of the transducer, which is amplitude modulated by the process. The information about the process is in the modulation envelope. An RMS-to-DC (root mean square-to-direct current) converter is used to demodulate the signal. The output of this device is the amplitude of the envelope.

The envelope is digitally resampled at a frequency appropriate for the process. For instance, 50 Hz is a typical digital sampling rate for a fluid bed reactor.

In some simple applications it is sufficient to record the RMS signal. An example of this would be detecting fine material in a gas stream downstream of a cyclone.

### Direct Digitization

The acoustic emission signal may be directly digitized. This is useful if a broadband sensor is used or if

use is going to be made of digital signal processing (DSP) technology.

Sampling theory states that the signal must be sampled at twice the highest frequency present in the signal. If this criterion is not met then frequencies above the Nyquist frequency (half the digital-sampling rate) will be aliased; that is, they appear at spurious lower frequencies. To avoid this, very high sampling rates (10 MHz) must be employed.

DSP allows much of the signal processing to be performed in hardware. The DSP is able to calculate and output to its associated computer many parameters characterizing the signal such as RMS power, total energy in a given time, counts of pulses, and frequency analysis via hard coded FFT (fast Fourier transform).

### Counts/Ring Down Counts

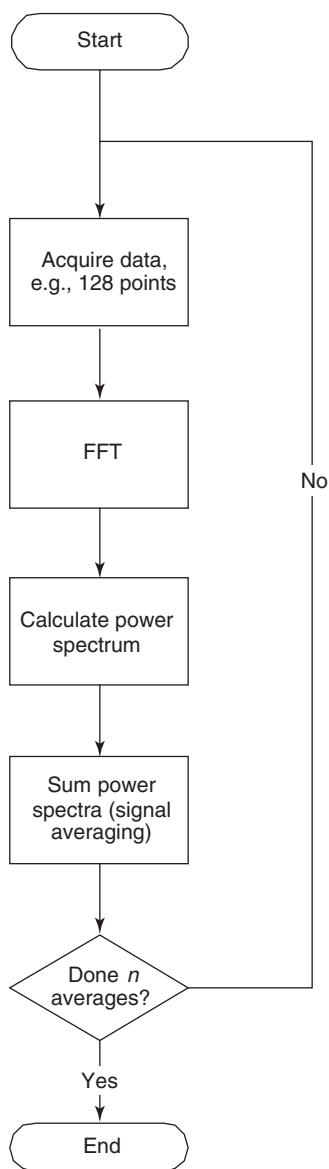
These measurements have been included for completeness. The origin of these parameters lies in the history of acoustic emission monitoring. These parameters are more significant in the realm of NDT than in process analysis. Counts are simply the number of acoustic emission bursts detected in a selected period. Ring down counts are an early attempt to measure the energy in a single burst. A voltage comparator is used to count how many detectable oscillations (threshold crossings) there are in burst. This provides a crude estimate of the energy.

### Data Analysis

Continuous acoustic emission is aperiodic, i.e., there is no start or stop to the signal. This means that it is not necessary to use signal-processing techniques that preserve phase. Power spectral analysis is therefore a useful technique. Because power spectra are coherent it is possible to gain a signal-to-noise improvement by averaging a number of power spectra. This spectrum may be of interest in its own right or used as signature for the state of the process. A scheme for calculating the power spectral estimate is given in (Figure 5). Other spectral analysis techniques, like the Gabor spectrogram, provide a good means of monitoring how the power frequency distribution varies with time (Figure 6).

Acoustic emission power spectra are similar in many respects to optical spectra and are amenable to chemometric processing (multivariate analysis). Principal component analysis, partial least squares (PLS), neural networks, and qualitative techniques such as SIMCA (soft independent modeling of class analogy; a pattern recognition technique) have been employed





**Figure 5** Power spectral analysis algorithm.

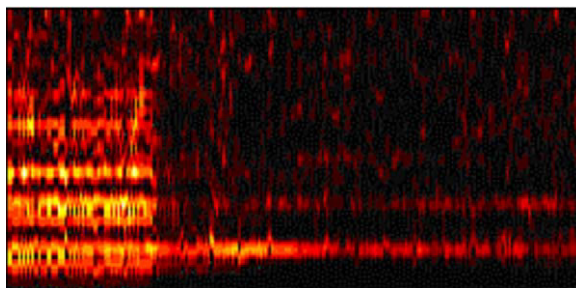
to analyze and empirically model acoustic emission producing processes.

## Applications

### Fluid Beds

Fluid bed technology is important in the oil, petrochemical, pharmaceutical, and food industries.

Catalytic cracking is an important process in the oil industry where petroleum vapor passes through a low-density bed of catalyst, which causes the heavier fractions to 'crack' producing lighter more valuable products.



**Figure 6** Gabor spectrogram representing a fingerprint of an entire high shear granulation process. The 'x-axis' represents time; the 'y-axis' represents frequency. The color represents intensity varying from black (low) through to yellow (high intensity).

In the petrochemicals industry they are used for producing polyolefins on a very large scale. The particles consist of growing polyolefin granules (polyethylene or polypropylene). The gas is the monomer ethylene or propylene.

Fluid beds are frequently used to produce granules or to dry particulate materials. These are of particular interest in the food and pharmaceutical industries.

In a fluidized bed, an upward current of gas suspends particles. The particles move upwards in the center of the bed and fall downwards, under gravity, at the walls. In dense beds a 'bubble phase' is often seen. These are pockets of gas moving upward through the bed. In the wake of these pockets is a stream of high-velocity particles. The bubbles are responsible for much of the mass transport of particles and are an important part of the mixing process. When bubbles grow too large they serve as conduits for gas by-pass thereby reducing gas-solid contact. In extreme cases when the bubble diameter approaches that of the vessel a phenomenon known as 'slugging' may develop. This lifts the material in the top sections of the bed causing loss of solid material and very poor efficiency.

Fluid beds produce large levels of acoustic emission. A resonant sensor, mounted on the outside wall of a fluid bed, senses the motion of the particles from their impact. Anything in the fluid bed, which changes the velocity, size, or hardness, brings about a change in the acoustic emission signals detected. The passing of bubbles through the bed causes a low-frequency modulation of the acoustic emission amplitude. As the bubble size changes the modulation frequency also changes. It is therefore easy to detect the onset of slugging behavior and to exercise a control measure.

Change in particle size and morphology influence the flow regimes within the bed and the acoustic



emission produced. Acoustic emission is a good indicator of the onset of bed disorders such as unwanted agglomeration of particles. In polyolefins production the product of the bed is a granule in the order of 200  $\mu\text{m}$  diameter. Under certain conditions these particles can agglomerate to form large lumps that disrupt the production process. Acoustic emission provides early indication that the nature of the bed is changing and that agglomerates are starting to form.

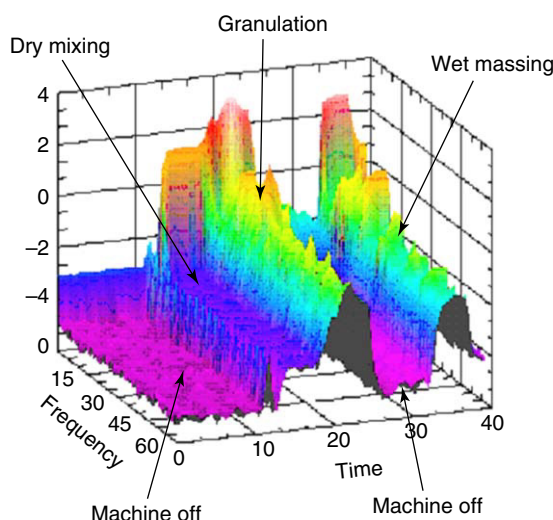
Pattern recognition (techniques such as SIMCA) performed on the acoustic emission spectrum has been used to detect deviations from normal operation and to predict endpoint conditions such as drying endpoint.

Powder transport lines can be considered as very lean fluidized beds. These are encountered in many industries; for instance, conveying powder-laden gases to cyclone separators. Here, acoustic emission proves to be a sensitive indicator for the presence of powders or liquid droplets, the detection of blockages, and as a semiquantitative indicator of the particulate mass flow.

### High-Shear Granulation

High-shear granulation is a process widely used in the pharmaceuticals industry. It is used to make granules ( $\sim 1000 \mu\text{m}$ ) from mixtures of much finer powders. The physical properties of the granules are very important as this affects their ability to be made into tablets. The powders need to be mixed to a very high degree of consistency, to ensure a uniform distribution of the active drug substance. The granules need to be freely flowing and have the right particle size distribution, density, and morphology for tablet making.

A high-shear granulator consists of a flat-bottomed mixing bowl and a multi-bladed impeller that rotates a very small distance above the bottom of the vessel. The first stage of the granulation process is dry mixing of the powder. After dry mixing, water and binder are sprayed into the vessel and the granulation process starts. Material is continuously mixed in the space above the impeller. The fine powder gradually agglomerates to form granules. At some point the addition of liquid is stopped and the material is left to complete the granulation. This process is known as 'wet massing'. The granulation process is stopped at a notional endpoint, i.e., when the particles have reached a satisfactory form for tablet making. Traditionally, the endpoint has been identified by visual inspection by an experienced operator. However, offline physical tests such as particle sizing, moisture content analysis, measurements of compressibility



**Figure 7** Change in power spectrum of RMS acoustic emission spectrum during high-shear granulation.

index (Carr's index), and flow characteristics are sometimes performed.

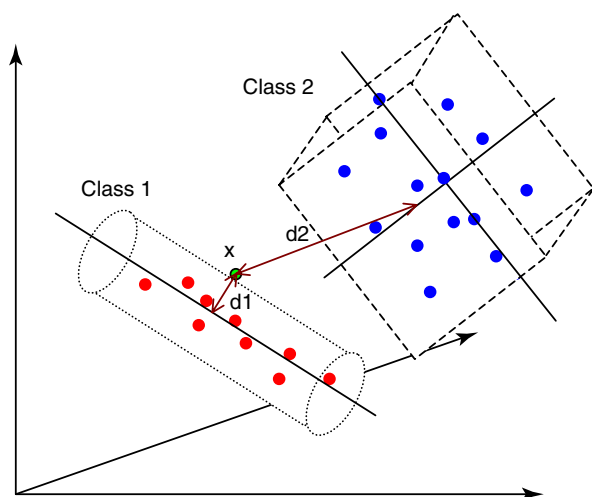
Under controlled conditions, these parameters may be quantitatively predicted from the acoustic emission spectrum using multivariate-modeling techniques such as PLS.

**Figure 7** shows the change in the RMS acoustic emission signal from a granulation process. Qualitatively it is easy to identify regions as static powder, dry blending, granulation, and wet massing. Dry mixing is characterized by a very flat spectrum (characteristic of truly random processes). As the liquid is added the powder starts to shear under the granulator blade. This produces intense broadband acoustic emission modulated by the passing of the impeller blades above the sensor. Similar levels of acoustic emission are seen during wet massing. As the material approaches the endpoint the acoustic emission signals tend to reduce in amplitude and stabilize.

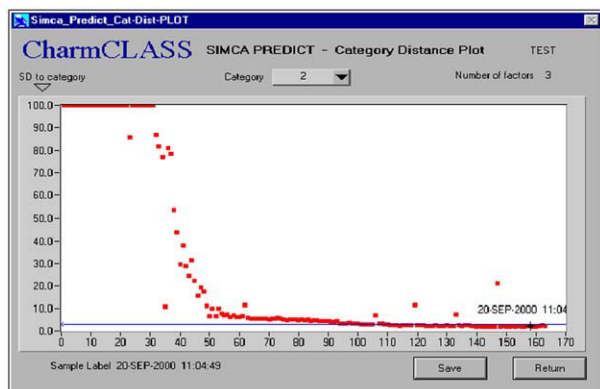
Pattern recognition techniques, like SIMCA, may be used to track the approach to endpoint by comparing the instantaneous acoustic emission spectrum with those obtained at the endpoint in previous production runs (**Figures 8 and 9**).

### Gas-Liquid Bubbles

Knowledge of bubble size distributions is important in gas-liquid systems, e.g., two-phase flow in pipelines, fermenters, and a variety of chemical reactors. In these cases, the acoustic emission extends from the audio region to  $\sim 50 \text{ kHz}$  and is usually detected by means of a submerged hydrophone.



**Figure 8** SIMCA modeling showing two multidimensional models (class 1 and class 2) made from acoustic emission signatures. In prediction mode the instantaneous acoustic emission signature (x) is compared to these models and a classification is made based on the distance the signature is from the centers of the models ( $d_1$ ,  $d_2$ ). The SIMCA distance is a measure of how well the instantaneous signal compares to the models. In this example the signature would be classified as belonging to class 1.



**Figure 9** SIMCA prediction output during high-shear granulation. The SIMCA distance reduces as the batch approaches its endpoint.

Each bubble has a number of possible modes of oscillation each producing a characteristic tone. For the air–water system the fundamental frequency of oscillation (Minnaert frequency)  $f_0$ , (where  $\gamma$  is the ratio of specific heats,  $s$  is the surface tension,  $P_\infty$  is the static pressure,  $R_0$  is the radius of the bubble,  $\rho$  is the density, and  $s$  is the surface tension) is given by:

$$f_0 = \left( \frac{3\gamma P_\infty}{\rho R_0^2} \right)^{1/2} / 2\pi - \left( \frac{2s}{\rho R_0^3} \right)^{1/2} / 2\pi$$

It is possible to model acoustic emission production from a population of bubbles of different sizes giving good correlation with physical measurements. The inverse problem has also been tackled; that is, the prediction of bubble size distributions from the acoustic emission spectrum. Good correlation (within 20% across the distribution) for bubble sizes from 0.2 to 10 mm has been obtained.

### Crystallization

Some chemical reactions produce acoustic emission directly. These require a phase change, i.e., liquid to gas, liquid to solid. Crystallization is a likely target for acoustic emission monitoring. There is some evidence that certain crystallization processes intrinsically produce acoustic emission. However, crystallization may be detected by the impact of the crystals on the walls of the reactor. As long as the fluid is well agitated acoustic emission gives a good estimate of the solid yield.

See also: **Chemometrics and Statistics:** Signal Processing. **Process Analysis:** Overview.

### Further Reading

- Betteridge D, Joslin MT, and Lilley T (1981) Acoustic emission from chemical reactions. *Analytical Chemistry* 53: 1046–1072.
- Bouchard JG, Beesley MJ, and Salkeld JA (1993) Acoustic diagnostics in process plant crystallisation monitoring. *Process Control and Quality* 4: 261–277.
- Boyd JWR and Varley J (1998) Sound measurement as a means of gas bubble sizing in aerated agitated tanks. *AIChE Journal* 44: 1731–1739.
- Buttle DJ and Scruby CB (1990) Characterization of particle impact by quantitative acoustic emission. *Wear* 137: 63–90.
- Eitzen DG and Wadley HNG (1984) Acoustic emission: establishing the fundamentals. *Journal of Research of the National Bureau of Standards* 89: 75–100.
- Minnaert M (1933) On musical bubbles and the sounds of running water. *Philosophical Magazine* 16: 235–248.
- Sholtz P, Bretz M, and Nori F (1997) Sound-producing sand avalanches. *Contemporary Physics* 38: 329–342.
- Tann CC, Thornhill NF, and Belchamber RM (2002) Principal component analysis of spectra with application to acoustic emissions from mechanical equipment. *Transactions of the Institute of Measurement and Control* 24(4): 333–353.
- Wade AP, Sibbald DB, Bailey MN, *et al.* (1991) An analytical perspective on acoustic emission. *Analytical Chemistry* 63: 497A–507A.
- Whitaker M, Baker GR, Westrup J, *et al.* (2000) Application of acoustic emission to the monitoring and end point determination of a high shear granulation process. *International Journal of Pharmaceutics* 205: 79–91.

## Maintenance, Reliability, and Training

**J Sung and A C Alves**, Rohm & Hass, TX, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The use of analytical instrumentation for continuous online measurement of industrial process streams for real-time information gained popularity in the late 1950s and early 1960s. This information has met various demands, the most notable of which include process control, safety, waste reduction, environmental emission monitoring and control, and product quality assurance. With the application of computers and electronics to analyzer technology in recent decades, online analyzers have become even more sophisticated and widely used. In many instances, the application of online analysis has become a necessity for survival in the face of the economic pressures of globalization in chemical manufacturing.

Since process analysis has become so critical to the computers that control optimal operating conditions, the information from the field measurements, such as flow, temperature, pressure, level, compositions, etc. must be reliable and accurate. These measurements must receive adequate maintenance in order for the control to be successful. Compositional information from online analyzers provides more detailed information for process control than conventional pressure, temperature, and level data; thus, online analyzers justifiably require more attention than normal.

The real-time process analyzer is significantly different from conventional instrumental measurements in combining analytical chemistry with instrumentation. Typically, there is a sample transportation and conditioning system associated with the analysis, as well as some form of data presentation for human or automatic interface. It is also different from the laboratory analytical instrument. While laboratory analysis occurs within environmentally controlled conditions, the process analyzer is typically installed in a harsh environment and the analyses are taking place around the clock. Because of these unusual characteristics, the technicians responsible for analyzer maintenance must be thoroughly familiar with the entire analyzer system (analyzer as well as sample conditioning system). Analyzer technicians are typically well trained and highly skilled, and dedicated solely to analyzer system maintenance. Occasionally, analyzer technicians work side by side with laboratory technicians.

### Prerequisites of a Successful Maintenance Program

The successful analyzer maintenance program should meet two prerequisites: a well-engineered and properly installed analyzer system, and a well-structured maintenance organization and system to provide service and support. The analyzer system installation and engineering includes the selection of an appropriate analyzer and properly designed and installed sample transportation and conditioning system. The maintenance organization should be staffed with a supervisor and properly trained personnel to execute a well-structured maintenance program.

An analyzer system with improper engineering or installation will require excessive maintenance time and effort. Similarly, if there is no maintenance system in place when the analyzer system is installed, even a properly installed analyzer system will fail before long.

### Analyzer System Engineering

1. Economics analysis: The cost and benefit balance of analyzer system installation merits thorough consideration. The cost should include the initial cost of engineering, installation, personnel training, and the estimated cost of lifetime operation. The benefits should include the potential process improvement (such as better yield, less waste, safer operation, less pollution, etc.) resulting from analyzer usage. The cost and benefit balance should fall within the typical guidelines of the particular corporation (e.g., pay-back within 3 years).
2. Analyzer selection: The selection of the analyzer and the vendor also bears consideration. Some useful questions to ask include: Is the analyzer suited to its intended purpose (sensitivity, accuracy, interference, speed of analysis, etc.)? Does the analyzer have self-diagnostic capability to show signs of malfunctioning? Are service and parts readily available from the vendor? Is the vendor accountable for support for the lifetime of the analyzer?
3. The respective locations of the sample system and analyzer: The sample system should be well suited to the intended analysis. The representative sample must be taken where it will give the most insight of the process, and sample transportation must be timely so that the analysis can be used for real-time process control. Sample return should be considered as well if the sample would create any environmental, health, or safety problems. The analyzer must be well

protected from the weather, and access should be convenient and reasonably comfortable for maintenance technicians. The sample system should be designed so that its function can be safely and easily monitored regularly. Critical operating parameters, such as temperature and flow, should be monitored at the control room computer. The system design must include a safe means for isolating the analyzer and sample system from the process for service so that not to interfere the production.

4. Tools, operating manuals, spare parts, and training: Any special tools, manuals, and training for each analyzer and the sample system must be provided. Any spare parts recommended by the manufacturer should be on hand. Training can be done several different ways: offsite, onsite, or on the job training. Also, the freedom of computer-based training allows training to be taken at any time and individual pace.

### Maintenance Organization

Analyzer system design, engineering, installation, and long-term operation involve many disciplines: analytical and physical chemistry, electrical, mechanical, chemical and project engineering, and management. Therefore, it is not always obvious where to place the maintenance organization. Sometimes, analyzer maintenance is an afterthought because of the immediate need to improve the process. After the analyzer has been installed, the maintenance of it is often assigned to whatever group happens to make the most sense under these particular circumstances.

Depending on the size of the plant, the number of analyzers, the culture of the location, and, to some extent, precedents in maintenance practice, the analyzer maintenance organization may be of any variety. Smaller plants may have a few analyzers that require only one person to maintain. A large plant may need to have a well-structured and staffed department to maintain the analyzers throughout the plant. Some plants may choose to have some analyzer technicians assigned to a certain production area, while others may assign technicians based on their skill levels. The analyzer maintenance organization may be localized or centralized.

The maintenance organization could be in one of many areas, including the instrument shop, engineering department, central laboratory, process control group, production area, or process research department.

Generally, a well-structured maintenance organization should consist of the following elements:

1. The supervisor: Functions as a central figure to coordinate and manage all the maintenance activities for the whole organization. This person should be intimately familiar with all the analyzers and the sample systems installed in the plant, and should have a general knowledge of analyzer technology. The supervisor should be able to answer any technical question from the technicians, and also be motivating to the technicians. The supervisor should be fully aware of each technician's skills and ability, and should be responsible to arrange proper technician training. The supervisor should be the single point of contact for customers regarding any analyzer related issue.
2. Technicians and training: The number of analyzers assigned to each technician depends on the complexity of the analyzer and, sometimes, on the importance of the analyzer and the skill level of the technicians. The technicians should have basic computer skills and familiarity with basic statistics. Technician training should be an on-going activity. At least 10–20% of working hours should be allocated to training.
3. Spare parts inventory: There should be a well-stocked and organized inventory of spare parts. Part of the maintenance organization should be dedicated to replenishing missing parts and refreshing the list periodically.
4. Tools and equipment: The analyzer group should be equipped with essential tools and basic equipment, such as computers, multimeters, scopes, etc.
5. Accountability: Accountability should be assigned to the supervisor and technicians for their assigned duties.
6. Vendors and contractors: The analyzer group should have both an updated vendor service contact list as well as the local analyzer service contractor list for external service and training support.
7. Close communication with customers/users: The maintenance organization should have periodic communication with the users to ensure that the analyzer data are being used correctly. The users should be notified when the analyzer is being serviced, and the maintenance group should be notified when the data are suspicious or erroneous.

### Analyzer System Maintenance

After the analyzer and the sample conditioning system are installed, commissioned, and turned over to the maintenance organization, it is time for the maintenance to ward off problems as long as possible and at minimal cost. The successful maintenance program can be broken down into these essential elements:

1. Technicians' knowledge: Analyzer technicians should have a comprehensive understanding of the

functions of the analyzer, sample system construction, and safe operation. They should also understand the purpose of the analyzer, both in production as well as its impact to the bottom line for the corporation. They need to understand and be aware of their contribution to successful analyzer operation.

2. Maintenance modes: The two modes of maintenance are reactive and proactive. A strictly reactive maintenance program has no maintenance plan, does not monitor the analyzer and the sample system regularly, and only performs maintenance when the system fails. Reactive maintenance is ineffective and inefficient, trading the cost of regular maintenance for the usually greater cost of repairing broken equipment. Proactive maintenance has a structured schedule and procedure on each analyzer system for the assigned technicians to perform. Such maintenance could range from a routine checklist to a complete overhaul of the analyzer or the sample system. This kind of maintenance program takes time to develop because each analyzer system installation and application is unique. A proven maintenance program for an analyzer system on one process may not work as well for the same analyzer system at another process, even if they are in the same facility. Statistical process control technique and a quality assurance program will help develop a proactive maintenance program.

3. Statistic process control: Dr. W. Edwards Deming revolutionized the manufacturing process by applying statistics to a simple control chart concept that can be used to tighten the manufacturing process and improve quality. This technique should be used to maintain the analyzer system. Data should be collected to construct control charts, and maintenance decisions regarding the analyzer system should be based on the control charts. For example, data collected from repeated analyses of a calibration standard are used to construct the control charts. The control charts contain the upper and lower control limits and the range. If subsequent periodic analysis of the same calibration produces a result that falls outside the control limits, this indicates a problem in the system. The maintenance plan should be able to identify the cause of the problem and resolve it.

4. Quality assurance (QA) program for analyzer and sample system: This is closely related to the control chart concept. The capability of the analyzer (e.g., accuracy, reproducibility, etc.) and the sample system (e.g., temperature, pressure, flow, etc.) should be established as soon as possible after installation. The control charts (mentioned above) of statistical process control technique should indicate the capabilities of the analyzer system. The analyzer should be checked against a calibration

standard and the operating parameters for the sample system (such as flow, temperature, etc.) should be monitored and documented periodically. These records are helpful to modify the maintenance schedule for the future. For example, if a flow check on the sample system remains the same for a month with daily checks, the maintenance schedule may be extended to once a month at the next revision. On the other hand, if pressure measurements of the sample system begin to decrease at the end of a week, the system may need to be cleaned more frequently. QA data can convert routine maintenance to preventive, and even to predictive, maintenance. The more QA data are collected, the better the system can become.

5. Documentation: Operating manuals, service procedures and schedule, cost and benefit analysis, analyzer history, service records (based on control charts), drawings, etc. should be well documented and organized.

6. Data network: With networking capability, analyzer data can often be viewed and archived on an internal data network. This is very convenient for the analyzer technicians and engineers to view the analyzer data at their offices and to analyze data along with other operating parameters for any correlation.

7. Maintain spare parts inventory: Part of the technicians' regular duty is to maintain adequate inventory of consumable parts, such as filter elements, calibration standards, etc. and critical parts for the analyzer and the sample systems, such as gas chromatography columns, special fittings and valves, etc. For a very critical operation, a redundant analyzer may be justifiable as a hot standby item.

8. Ongoing training: Technicians' skills and knowledge should be periodically reviewed and assessed. Refresher and new training for new installation or new personnel should be ongoing to maintain the adequate competency. Attending analyzer conferences should be encouraged, especially for the supervisor, in order to keep abreast with the latest analyzer technology.

9. Maintain regular communication with customers, lab, and analyzer vendors: The users (customers) should be regularly communicated for their feedback of the service provided. It is a good idea for the analyzer supervisor to attend the regular production meetings in order to keep updated with the level of support on the analyzer systems. The analyzer group should also maintain close communication with the vendors for new product introduction and training opportunity. The analyzer technicians should work closely with the lab personnel for mutual support of the analytical efforts.

10. Replace, upgrade, or remove analyzers: All the analyzer installations should be periodically reviewed for its operability and cost effectiveness. This should be performed with the production personnel to determine if the analyzer needs to be replaced, upgraded, or removed if it no longer serves the purpose.

## Analyzer System Training

As mentioned earlier, the analyzer system is typically composed of an analyzer, the sample transportation, and conditioning system. Maintenance activities should include all parts of the system in order to be successful. Thus, the training should be on all these aspects of the system. The technicians should be responsible and accountable for all the elements associated with the analyzer system. A well-calibrated analyzer without a representative process sample to analyze is useless. Wrong data is worse than no data. The technicians should be equipped with the proper knowledge and skills from the training programs.

The training can be achieved by several different ways: Classroom training, onsite or offsite training, computer-based training, hands-on training, on-the-job training and the latest, web-based training.

Training programs are usually categorized into the following areas:

1. Analyzer training: Subjects should include: how the analyzer works, includes some basic chemistry, physics, and electronics; some basic statistics knowledge, such as of standard deviation, is very helpful. The technicians should be familiar with the analyzer capability, such as the sensitivity, response time, analytical accuracy, and repeatability. The QA program and troubleshooting procedures should also be included.
2. Sample system training: The technicians should be familiar with the entire sample transportation and conditioning system. Several key questions should be asked: Where does the sample come from and return? Under what specific operating parameters, such as temperature, pressure, flow rate, should the sample system be operated? How to monitor the system performance and how to isolate the sample system from process for offline service? How to safely take a representative sample for offline analysis? With a QA program and use of the statistical process control, the technicians should be trained to determine when to service the sample system, such as water wash to clear the sample line; and when to replace some parts, such as a dirty filter.
3. Computer skills training: The analyzer technicians should have some basic computer skills in

word processing, spreadsheet, and file and data management. The technicians should be expected to write some simple reports on occasions for documentation purpose. Some commercially available or customized maintenance software can be helpful in keeping the analyzer and maintenance records. This maintenance software may also contain comprehensive manuals and troubleshooting guides for all the analyzers. There are companies that, for a fee, will provide the service to populate the data in the maintenance software for the customers and conduct user training. For a large facility with many installed analyzers, this may be a beneficial investment.

4. Conferences, research consortiums: To keep up-to-date with the new technology, the analyzer personnel should attend conferences on process analyzers, such as the annual meetings of International Forum on Process Analytical Chemistry (IFPAC) and Instrument Society of America Analysis Division (ISA A/D). Research consortiums (membership required), such as Center for Process Analytical Chemistry (CPAC) at University of Washington and Measurement and Control Engineering Center (MCEC) at University of Tennessee are the well-known centers in the USA devoted to the process analysis research and applications. Sponsoring members are usually exposed to cutting edge technologies in the process analysis areas. These conferences and research consortiums often sponsor useful training programs in the analyzer areas in conjunction with their biannual meetings.

## Analyzer System Reliability

### Definition of Reliability

Reliability is the probability that an equipment can perform continuously without failure for a specific period of time. Reliability can be also the capability of an equipment continuously to perform its intended functions. Normally, reliability is measured by mean time between repairs (MTBR). MTBR is equivalent to the total equipment or instrument downtime in a given period, divided by the number of failures in that period. Reliability engineering can be defined as a staff function. The prime responsibility is to ensure that maintenance techniques are applied effectively, that equipment is designed and modified to improve maintainability and operability, that ongoing maintenance technical problems are investigated, and appropriate corrective and improvement actions are taken. Analyzers and sample systems reliability depends on an effective design, good operation, and good maintenance practices.



### **Analyzer Reliability Vision**

In order to achieve operational consistency, process must be controllable and operated properly. Analyzers for a chemical process serve as a powerful tool to help process engineers and operators to effectively do their jobs. In the following examples, reliable analyzers are definitely required for these tasks:

1. For troubleshooting process upsets: Production personnel monitor and control the process in a consistent manner to produce quality products. When process upsets occur, there must be sufficient confidence in the analyzer data so that quick analysis of the 'out of control' situation is made and the appropriate corrective actions are undertaken.
2. For process improvement: Technical support will have the reliable process data to analyze production efficiency and formulate optimization strategies to improve process performance.
3. For process optimization research: Accurate and tight control will enable the execution of well-designed experiments in a controlled environment and provide data that can be used with confidence in process optimization.

### **Analyzer Reliability Strategy**

A reliability strategy is composed of the following programs:

1. Training program: A good maintenance analyzer facility will develop its own training program based upon the analyzers installed in the plant or facility. The analyzers training program will need clear, concise training objectives written that are based on verified data. A formalized evaluation process should be performed to determine the success of the training program. On-the-job training should be considered, with evaluation/certification of the trainee.
2. Develop 'Top Ten' bad actors list: The purpose of the bad actors list is to establish the analyzers within the facility that shall receive priority for resource allocation determination. These resources include preventive maintenance, spare parts, reliability study, and necessary training for technicians and operators. The 'Top Ten' bad actors list ranks the importance based on the following three categories:
  - frequency,
  - maintenance cost, and
  - safety/environment impact on business (productivity and quality). This list should be generated periodically for discussion with the production

personnel to take proper action to address the bad actors.

3. Elaborate or revise the maintenance procedures and preventive maintenance: We rely on maintenance procedures as a tool in achieving consistent performance. They set a right way based on experience and manufacturers advice and allow any qualified person to perform the maintenance work as expected. The maintenance procedures will define all requirements covering maintenance of the analyzers, including tracking performance, and alerting to overdue maintenance. It's a common assumption that certain individuals will always do a particular task and they will remember exactly the best way to do it every time, every day. We want reliable performance from our analyzers, and consistently doing the correct things leads to reliable performance.

4. Analyzers database implementation: Normally the data provided by the computerized maintenance management systems (CMMS) are not sufficient to provide the information needed to make the right decisions regarding improving analyzers reliability. The combination of the CMMS and an analyzer database provides a very powerful tool. Today, there are several commercial instrument database software available in the market. Alternatively, an analyzer database can be developed in-house using common database software such as Microsoft Access. An analyzer database should contain the following basic information:

- general information (manufacturer, model/serial number, applications, etc.);
- parts replaced over a period of time;
- calibration information (accuracy, deviation, range, as found/as left);
- name of the technician;
- date; and
- maintenance procedures.

5. Spare Parts evaluation: The analyzer maintenance expeditor should maintain close ties with the plant warehouse and the purchasing department to optimize parts availability and its location. Spare parts should be readily available when needed. Vendor information should be used as a guide to determine the parts to be included in stores. We have to be careful to avoid over-stocking of parts to control the cost. If the expensive parts can be readily obtained from vendors in a timely manner, then it need not be in the warehouse.

6. Metrics to evaluate the results: The metrics should be visible, understood, and used by the people involved in the process to improve analyzers reliability. The metrics will show how well the strategy to

improve analyzers reliability is working. These data should be reviewed periodically to improve the reliability and reduce the cost. Below are some metrics that can be used to measure analyzer performance:

- maintenance cost (labor and materials);
- MTBR: mean time between repairs;
- uptime: defined as the time that an item of equipment is in service and operating; and
- production loss caused by analyzer malfunction.

See also: **Process Analysis:** Overview; Chromatography; Electroanalytical Techniques; Sensors; Acoustic Emission; Bioprocess Analysis.

## Further Reading

Clevett KJ (1986) *Process Analyzer Technology*. New York: Wiley.

Clevett KJ (1991) Maintaining process analyzers for high reliability and performance. *Process Control and Quality* 1: 89–99.

Deming WE (1986) *Elementary Principles of the Statistical Control of Quality*. Tokyo: Nippon Kagaku Gijutsu Renmei.

Deming WE (1986) *Out of Crisis*. Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study.

Eustace DJ (1995) Process analysis: Maintenance and training. In: *Encyclopedia of Analytical Science*, pp. 4164–4172. Cambridge: Elsevier.

Jacob SM, Mehta SM (1987) Process quality measurement systems: An integrated approach to a successful program in process analytical instrumentation. *American Laboratory*.

Nichols GD (1988) *On-line Process Analyzers*. New York: Wiley.

Van Vuuren P (1993) Process analytics – an integrated approach to on-line and off-line functionality and organizational support. *Seventh International Forum on Process Analytical Chemistry*, January 1993.

## Bioprocess Analysis

**T Schweder**, Ernst-Moritz-Arndt-Universität, Greifswald, Germany

**P Neubauer**, University of Oulu, Oulu, Finland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Knowledge of the physiological state of cells and the physicochemical conditions of the bioreactor environment is required for an optimal bioprocess control. The analysis of bioprocesses is based on physical, chemical, and biological parameters. A selection of offline and online parameters, which are used for the analysis of bioprocesses, is shown in **Table 1**.

A comprehensive understanding of the physiology of industrially relevant prokaryotic or eukaryotic cells during bioprocesses is mandatory for a successful strain and process optimization. The physiological status of the cells during a bioprocess can be measured indirectly by analyses of external variables outside the cells or directly by measuring cellular components inside the cells. New molecular analytical methods allow the determination of detailed biomolecular parameters of the production cells. Physicochemical separation techniques like chromatography or gel electrophoresis allow an accurate separation of molecular components of cells. Furthermore, the analysis of key metabolites gives valuable information on the cellular physiology of the cells.

Gene expression activities of cells can be determined primarily by their mRNA and finally by their protein pattern. The proteomics approach, which is based either on electrophoretic separation of the cellular proteins (proteome) and their identification by mass spectrometry (MS) or on protein chips, allows a visualization of the final products of the gene expression activity of the cells, and thus reflects rather closely their *in vivo* physiological state. The proteome analysis technique, however, has some limitations and does not allow a determination of the expression of all process relevant genes. In this respect, the DNA array technology represents a further quantum jump in the molecular physiological state analysis. DNA arrays allow a faster analysis of the expression of process-relevant genes at a certain time during a bioprocess. In comparison to proteins, mRNA levels reflect changes in the bioprocess performance at a rather early stage. However, the combination of proteome and transcriptome analysis allows a nearly comprehensive view on the physiological state of the cells and may contribute to evaluation of process-related bottlenecks. The integration of different analytical data and the understanding of complex cellular regulatory networks offer new perspectives in the optimization and control of bioprocesses. In this respect, expert systems and models for control are very helpful tools to maintain the process at optimal conditions.

improve analyzers reliability is working. These data should be reviewed periodically to improve the reliability and reduce the cost. Below are some metrics that can be used to measure analyzer performance:

- maintenance cost (labor and materials);
- MTBR: mean time between repairs;
- uptime: defined as the time that an item of equipment is in service and operating; and
- production loss caused by analyzer malfunction.

See also: **Process Analysis:** Overview; Chromatography; Electroanalytical Techniques; Sensors; Acoustic Emission; Bioprocess Analysis.

## Further Reading

Clevett KJ (1986) *Process Analyzer Technology*. New York: Wiley.

Clevett KJ (1991) Maintaining process analyzers for high reliability and performance. *Process Control and Quality* 1: 89–99.

Deming WE (1986) *Elementary Principles of the Statistical Control of Quality*. Tokyo: Nippon Kagaku Gijutsu Renmei.

Deming WE (1986) *Out of Crisis*. Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study.

Eustace DJ (1995) Process analysis: Maintenance and training. In: *Encyclopedia of Analytical Science*, pp. 4164–4172. Cambridge: Elsevier.

Jacob SM, Mehta SM (1987) Process quality measurement systems: An integrated approach to a successful program in process analytical instrumentation. *American Laboratory*.

Nichols GD (1988) *On-line Process Analyzers*. New York: Wiley.

Van Vuuren P (1993) Process analytics – an integrated approach to on-line and off-line functionality and organizational support. *Seventh International Forum on Process Analytical Chemistry*, January 1993.

## Bioprocess Analysis

**T Schweder**, Ernst-Moritz-Arndt-Universität, Greifswald, Germany

**P Neubauer**, University of Oulu, Oulu, Finland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Knowledge of the physiological state of cells and the physicochemical conditions of the bioreactor environment is required for an optimal bioprocess control. The analysis of bioprocesses is based on physical, chemical, and biological parameters. A selection of offline and online parameters, which are used for the analysis of bioprocesses, is shown in **Table 1**.

A comprehensive understanding of the physiology of industrially relevant prokaryotic or eukaryotic cells during bioprocesses is mandatory for a successful strain and process optimization. The physiological status of the cells during a bioprocess can be measured indirectly by analyses of external variables outside the cells or directly by measuring cellular components inside the cells. New molecular analytical methods allow the determination of detailed biomolecular parameters of the production cells. Physicochemical separation techniques like chromatography or gel electrophoresis allow an accurate separation of molecular components of cells. Furthermore, the analysis of key metabolites gives valuable information on the cellular physiology of the cells.

Gene expression activities of cells can be determined primarily by their mRNA and finally by their protein pattern. The proteomics approach, which is based either on electrophoretic separation of the cellular proteins (proteome) and their identification by mass spectrometry (MS) or on protein chips, allows a visualization of the final products of the gene expression activity of the cells, and thus reflects rather closely their *in vivo* physiological state. The proteome analysis technique, however, has some limitations and does not allow a determination of the expression of all process relevant genes. In this respect, the DNA array technology represents a further quantum jump in the molecular physiological state analysis. DNA arrays allow a faster analysis of the expression of process-relevant genes at a certain time during a bioprocess. In comparison to proteins, mRNA levels reflect changes in the bioprocess performance at a rather early stage. However, the combination of proteome and transcriptome analysis allows a nearly comprehensive view on the physiological state of the cells and may contribute to evaluation of process-related bottlenecks. The integration of different analytical data and the understanding of complex cellular regulatory networks offer new perspectives in the optimization and control of bioprocesses. In this respect, expert systems and models for control are very helpful tools to maintain the process at optimal conditions.

**Table 1** Online and offline parameters for the monitoring of the physiology of cells during bioprocesses

Parameters	Analytes	Measuring device/methods
<i>Online</i>		
Growth, cell density	Cell population, broth	Spectrophotometry
pH	Broth	pH electrode
Metabolism	Metabolites	HPLC
Respiration	$pO_2$ level	$pO_2$ electrode
Exhaust gas composition	$CO_2$ , $O_2$	Gas analyzers, MS
<i>Offline</i>		
Cell viability/population analysis	Single cells	Flow cytometry, FISH, plating, microscopy
Energetic pool	Nucleotides	HPLC, enzymatic test
Gene expression	mRNAs	Real-time RT-PCR, DNA chips, Northern blot
	Proteins	2D-PAGE/MS, protein chips, enzymatic analyses, immunologic methods
Metabolism	Metabolites	NMR, TLC, HPLC, enzymatic analyses
Stress	Free amino acids	HPLC
	Stress specific mRNAs or proteins	Real-time RT-PCR DNA- or protein-chips, proteome analysis
Medium components	Carbohydrates, ammonia/nitrogen	Enzymatic, HPLC (Micro)-Kjeldahl, ion-sensitive electrode
	Phosphate	Chemical analysis
	Total protein	Chemical analysis

FISH, fluorescence *in situ* hybridization; HPLC, high-performance liquid chromatography; MS, mass spectrometry; NMR, nuclear magnetic resonance; PCR, polymerase chain reaction; RT, reverse transcription; TLC, thin-layer chromatography; 2D-PAGE, two-dimensional polyacrylamide gel electrophoresis.

## Analytical Methods for Bioprocess Monitoring

### Online Parameters

State-of-the-art bioreactors allow the online detection and control of a number of process variables. These online variables can be used to calculate culture-specific parameters, which characterize the cell's activity and the growth state. Generally, pH, dissolved oxygen tension ( $pO_2$ ), and temperature are measured continuously. In connection to other online data, such as level or fermenter weight, pressure, temperature of cooling water circuit, stirrer power input, and volume of additives, these measures can be used to evaluate the growth state. Much additional information can be obtained by outgas analysis (infrared for  $CO_2$ , paramagnetic for  $O_2$ , or other principles such as electroacoustic methods and MS). Volumetric rates, e.g., for oxygen consumption,  $CO_2$  formation, and heat production can be calculated. From these volumetric rates, specific rates (e.g., specific growth rate, specific substrate uptake rate) and yield parameters (e.g.,  $Y_O/S$ , RQ-value) can be derived that contain a high information content.

Methods for detection of biomass can be divided into direct and indirect methods (Table 2). Most online probes for direct monitoring of cell mass are based on optical principles, including light absorbance, light scattering, and fluorescence. Although such sensors may work well in specific cases, often

**Table 2** Methods for online analysis of biomass during fermentation processes

Direct	Electrical	Conductivity Capacitance
	Optical	Fluorescence Light absorbance Light scattering Real-time imaging
Indirect	Acoustic resonance densitometry Nuclear magnetic resonance spectroscopy	
	Oxygen uptake rate	
	Carbon dioxide evolution rate	
	ATP-production rate	
	NAD sensor	Fluorescence
	Heat production	Calorimetry
	Stirrer power input	
	Redox level	
	pH value/base addition	
	GFP	Fluorescence

GFP, green fluorescent protein; NAD, nicotinic acid adenine dinucleotide; ATP, adenosine tri-phosphate.

the influence of medium components disturbs the measurement. Furthermore, most of these sensors cannot be used in high-cell-density cultivations. Alternatively, sensors with electrical principles, such as capacitance and conductivity, have been developed. These sensors can be used at high cell densities and even in densely packed bioreactors like fixed-bed systems.

Indirect methods are based on the measurement of cellular activities and mostly detect viable cells. These methods often make use of standard online sensors, such as outgas analysis, pH, power input to stirrer, temperature input, or specific probes such as redox electrode or fluorescence sensors for the detection of NADH or green fluorescent protein (GFP), and subsequently apply calculations or simple models for the determination of biomass.

Generally, care must be taken in calibration of the online biomass sensors. The conditions of their use have to be standardized, as medium components, particles, gas bubbles, etc. may disturb the signal.

**Electronic noses** The so-called electronic noses consist of chemical gas sensors that are able to monitor changes in the offgas composition of fermentation processes. The different sensors of electronic noses are based on conductive polymers (CP), metal oxide semiconductors (MOS), metal oxide semiconductor field effect transistors (MOSFET), or quartz crystal microbalance (QCM). CP-based sensors use the electrochemical properties of polymers like polypyrrole or polyindole. The absorbance of selected molecules of the off-gas into the polymer film causes changes in the sensors conductivity. MOS sensors possess an electrochemically active surface of metal oxides like tin oxide or copper oxide. The sensitivity

of this type of gas sensors toward different inorganic (e.g., hydrogen, carbon monoxide, ammonia, hydrogen sulfide, nitrogen oxide, or sulfurous compounds) or organic compounds (e.g., alcohols and hydrocarbons) of the off-gas is modulated by the oxide composition, the amount of trace elements (e.g., palladium, gold or rhodium), and the temperature, which ranges from 100°C to 400°C. A MOSFET sensor is based on thin catalytic metal films covering a silicon oxide field effect transistor. The catalytic decomposition of organic and inorganic compounds on the surface causes shifts in the capacitance of the semiconductor device of this sensor. QCM sensors are able to measure the physical mass of a bound analyte. This is reached by determination of changes in the frequency of a quartz crystal by binding of the analyte to the sensor surface.

### Offline Parameters

**Medium components** The analysis of components in the cultivation medium during bioprocesses requires sampling procedures that consider the appropriate rate constants of their consumption or synthesis. A selection of different sampling procedures for the analysis of cellular components or medium compounds during fermentation processes is shown in Table 3.

**Table 3** Methods for sampling of cellular components or compounds in the fermentation medium

Example	Sampling	Quenching/inactivation		Extraction	Stabilization/storage
		Physical	Chemical		
Glycolysis intermediates	By overpressure (bioreactor) or underpressure (sampling vial)	Low temperature (–20 to –70°C)	Methanol	Perchloric acid	–70°C
mRNAs			Phenol/ethanol		RNase inhibitor, –70°C
Nucleotides			Perchloric acid	Repeated freeze/thaw cycles	–70°C after titration to pH 7
Nucleotides			Formaldehyde	Basic (KOH)	–70°C
Isotopomer distributions in intracellular proteinogenic amino acids		Low temperature, 0°C			–70°C
Proteins, 2D-PAGE			Chloramphenicol/ NaN <sub>3</sub>		–70°C
Medium components		Filtration through 0.2 or 0.4 µm filters			–0°C
Medium components		Centrifugation			–70°C
Medium components (glucose, acetate)		Centrifugation	Perchloric acid		–70°C after titration to pH 7

KOH, potassium hydroxide; 2D-PAGE, two-dimensional polyacrylamide gel electrophoresis.

Sampling techniques are challenging for medium components with a high rate constant and low concentration especially at high cell densities. Examples of medium components are carbon compounds (e.g., glucose, acetate), nitrogen compounds (e.g., ammonia,  $\text{NO}_2$ ), extracellular metabolites (e.g., amino acids), and signal compounds (e.g., extracellular cAMP). For medium sampling filtration techniques including dialysis through filters and bypass modules are normally used. Thereby filter clogging, binding of the analytes to the filter, and residual target-transforming activities have to be kept in mind and should be critically evaluated in each case.

Glucose is the most favored carbon and energy source of many industrial microorganisms and thus one of the most frequently used nutrients in biotechnological processes. Many processes apply the fed-batch technology, which is used to reach high cell densities as a basis for high product yield by the continuous growth-limiting addition of concentrated carbon substrate solutions. This technique combines the advantages of avoiding substrate inhibition and the production of unwanted metabolic side products, thereby allowing bioprocesses with high cell densities and without oxygen limitation. However, by definition the concentration of the limiting substrate is in the order of magnitude of the saturation constant  $K_s$ . For glucose, the saturation constant is in the  $\text{mg l}^{-1}$  range, but the glucose uptake rate of a culture at a higher cell density is in the order of  $10 \text{ mg l}^{-1} \text{ s}^{-1}$ . Therefore, the detection of actual glucose concentrations in glucose-limited fed-batch cultivations is a challenge and demands a fast inactivation of further cellular metabolic activities. Preferred methods are direct sampling of the cell broth in ice-cold quenching solutions, such as perchloric acid or methanol, by taking advantage of the overpressure in the fermenter, or sampling with a concentric double lumen catheter with inactivating liquid in the outer tube and inactivation at the point where the cell broth is sucked into the inner catheter. Although the latter method has been successfully used in line with a dialysis module as an online sampling tool in fermentations and for blood sugar analysis in the human body, the method is not widely used for bioprocess analyses due to the missing sterility barrier. Alternatively, syringes (monovettes) with underpressure, as they are also used for blood sampling, have been applied successfully for offline sampling in combination with cooling for bioprocess analyses. It must be taken into account that inactivation with quenching solutions can result in leakage of cellular compounds to the medium, which may be specific to the cultured organism.

The analysis of glucose and other sugars is generally performed by high-performance liquid

chromatography (HPLC) or by enzymatic assays. Enzyme electrodes have also been tested. They, however, cannot be used inside a bioreactor due to their limited heat stability. However, several flow injection systems (flow injection analysis, FIA) have been realized and are commercialized today. They allow a detection of the analyte with a short time delay of seconds to minutes with a sampling interval down to one sample per minute.

A specific problem for the use of such medium analyses in the control of a bioprocess can be caused by insufficient mixing of the bioreactor environment. Local zones of high or low nutrient concentrations or varying oxygen availability can produce gradients in the concentration of the analyte. It has been demonstrated that in large-scale fed-batch processes the concentration of the limiting substrate fluctuates due to slow mixing of the feed solution. Feed solutions are usually applied as highly concentrated and viscous solutions to minimize the dilution effects. The limiting substrate concentration in a fed-batch fermentation process depends on the sugar uptake kinetics, but is generally by the order of 10 000 smaller than the feed concentration. This creates a dynamic situation in large fed-batch bioprocesses. Cells repeatedly pass close to the feed point and are exposed, for a short time, to a much higher sugar concentration than the mean concentration in the reactor. A rapid utilization of the sugar and concomitant local oxygen limitation is expected at high cell densities. Furthermore, many microorganisms respond by overflow metabolism when exposed to glucose concentrations above a critical level. Furthermore, local oxygen limitations in the bioreactor environment create specific anaerobic conditions and provoke concomitant stress responses.

A nitrogen balance by monitoring inorganic and organic nitrogen may help to better understand the metabolic activities in a bioprocess. In dependence on the microorganism applied, microbial media may contain nitrogen in different oxidation states as nitrate, nitrite, ammonia, and organic nitrogen. Organic nitrogen occurs in proteins, peptides, amino sugars, nucleic acids, and urea. Ammonia and nitrate are alternatively used as inorganic nitrogen sources, whereas organic nitrogen is contained in complex media and in products of microbial activities including the biomass. Analytically, ammonia and organic nitrogen can be determined together using the Kjeldahl method or by ion-sensitive electrodes. Generally, for an accurate assessment of total nitrogen, all of the various nitrogen forms are separately measured and then the results are summarized. Consequently, analysis for total nitrogen is a time-consuming and difficult procedure. Most of the methods are wet



chemical methods including ultraviolet/visible spectroscopy requiring extensive sample preparation with several digestions and/or titration steps. New methods for the determination of total nitrogen, which simplify and speed up the analysis, are based on either gas chromatography with a flame ionization detector for gaseous nitrogen compounds or on chemiluminescence for liquid and solid samples with prior combustion to nitrogen monoxide and nitrogen dioxide and following reaction with ozone to form an excited state of nitrogen dioxide. Upon returning to ground state, light energy is emitted. Total nitrogen is measured using a chemiluminescence detector.

Phosphate is analyzed spectrophotometrically by a chemical reaction with ammonia molybdenum and iron sulfate or by ion chromatography.

**General cell parameters** Traditionally, a large number of methods has been applied to analyze general cell parameters in bioprocesses such as gravimetry (cell dry weight, cell wet weight), volumetry (sedimentation volume), optical methods (e.g., light absorbance in the range of 500–650 nm), visualization in microscopic counting chambers (cell count), or growth-based methods (replica plating or spreading out). Whereas all these methods have in common to consider the biomass as a unit, a number of modern methods aims for the analysis of subpopulations to get information about the variability in a bioprocess. In the analysis of subpopulations flow cytometry has been a breakthrough although the size of bacteria is still at the lower end of the technology. Combination with fluorescing labeling compounds makes flow cytometry a powerful analytical method. This method has the major advantage of being able to combine single cell analysis with a high sample throughput. Current flow cytometers allow the analysis of  $\sim 10^4$  cells per second and thus give statistically ensured data on the properties of the cells during different stages of a bioprocess. Cellular components or the integrity of cells can be analyzed by applying distinct staining procedures with fluorescing compounds that are available in a large number. The complexity of labeling protocols is the reason why this technique has not been performed online so far. The cells are analyzed cell by cell in a flowthrough chamber, where each cell is determined by a light source, usually a laser beam, and an appropriate detection unit, which measures the scattered and the fluorescence light emitted by the distinct fluorescence dyes. By this approach not only cell number, size, and shape can be explored but also the viability of cells, cell wall integrity, internal pH, inclusion bodies, storage material, and DNA, RNA, or protein content. The flow

cytometry analysis gives valuable information on the physiological state of the cells and thus on the performance of the whole bioprocess. Flow cytometry has a good potential to be combined with the methods described below for the analysis of specific RNAs and proteins.

**Cell components** The detailed analysis of cellular components may provide a deeper understanding of the behavior of cells in bioprocesses. Today, a number of methods can be applied to separate thousands of cellular compounds and even to quantify them. A few of the current methods, which were developed to handle the large complexity of biological samples, are briefly described below.

Generally, for all these methods the sampling and inactivation (quenching) of the cellular activities is a critical step taking into account the time frame of the chemical and enzymatic reactions as well as the synthesis and degradation processes. Each target compound or group of compounds needs an own protocol for sampling, inactivation, and stabilization, which has to be optimized for the organism and process of interest. **Table 3** summarizes a number of principles that have been described for the analysis of different target components. Generally, the time span from taking the sample out of the fermenter broth to complete quenching of the sample should be very short. Running bioreactors with a slight overpressure simplifies sampling. Additional underpressure in the sampling vial supports fast sampling. It is important to keep the sampling connection very small in comparison to the sample volume and/or to guarantee washing of the connection to avoid falsification of the analytical results by co-sampling of old material from the sampling pipe. The multiple sampling in very short intervals (milliseconds to seconds), has been a special challenge for which special devices have been constructed. The same problem of the inactivation interval also has to be considered in FIA. Therefore, different principles have been applied, such as concentric double lumen catheters, filtration modules (e.g., cross flow, hollow fibers), and ultrasonic waves in the megahertz range.

Quenching of the samples is necessary to freeze the target components at a state that does not deviate from their levels in the reactor. Generally, collection at a very low temperature is most favorable to inhibit enzymatic and chemical reactions, but freezing should be avoided if cells and media are separated in a later step, as thawing always leads to some cell lysis and release of cellular components into the medium. Quenching can be efficiently supported by chemicals such as sodium azide, formaldehyde, and methanol. Also, special cell inhibitors, such as

antibiotics, may be helpful, e.g., chloramphenicol for inhibition of protein synthesis, or protease and RNase inhibitors.

As in most cases at-time analysis is not practicable, the analysis protocol should contain an option for sample storage. Most common is the storage of aliquot samples at  $-70^{\circ}\text{C}$  or on liquid nitrogen.

Fast sampling has especially been of interest for metabolic flux analysis. This method is based on the analysis of a large number of cellular compounds. Mostly this method includes radioactive labeling, following fast sampling and analysis of the medium and cell fractions by nuclear magnetic resonance (NMR).

The offline analysis of selected metabolites of the cells during a bioprocess can be performed by MS, NMR spectroscopy, or chromatographic separation procedures. The NMR technique is based on the specific response of atomic nuclei that are exposed to a magnetic field. This technique is able to resolve the structures of small biomolecules and thus to determine selected metabolites during a bioprocess or for metabolic flux analyses. Advanced MS techniques allow a high throughput analysis of complex mixtures of metabolites by the determination of the specific masses of the metabolites.

Nucleotides are key metabolites playing a major role in energy metabolism as electron acceptors (NAD, NADP), in enzyme regulation (ATP, ADP, GTP), and in cell signaling (cAMP, ppGpp). Changes of their level and phosphorylation status belong to the earliest metabolic reaction of a cell to environmental signals. Therefore, their analysis has been interesting and challenging due to their high turnover rates. For fast quenching and extraction different procedures have been applied, such as direct sampling in formaldehyde, ice-cold perchloric acid, trichloroacetic acid, or concentrated KOH followed by neutralization. The selected method depends on the cell wall characteristics of the organism, the medium components, and may include additives such as ethylenediaminetetraacetic acid to complex divalent metal ions, which are activators of adenine nucleotide-converting enzymes. Different methods exist for the quantitative analysis including the detection of radiolabeled nucleotides, which is most sensitive, as well as luminometry and a variety of HPLC procedures. HPLC analysis is most suited for bioreactor analyses. Different HPLC techniques have been developed including ion exchange, reversed phase (RP), and ion-pair RP principles. Luminometric analysis is restricted to adenosine nucleotides. Radiolabeling of bioreactor samples is mostly not possible, although this is a highly sensitive method.

As mentioned above, chromosomal DNA and plasmid content of cells during a bioprocess can be monitored by flow cytometry. These cellular genomic constituents contain a decisive part of the blueprint of the bioprocess. For example, plasmids can carry recombinant genes, which code for the product formation of the cells. The stable distribution of these plasmids to the daughter cells is crucial for the yield of a bioprocess. Therefore, the analysis of the plasmid content of single cells is an important approach to monitor the productivity of the cells during recombinant bioprocesses. Another approach for the quantification of the plasmid number or DNA concentration is based on the real-time polymerase chain reaction (real-time PCR). Specific probes for selected sequences of the genes of interest, which are labeled with different fluorescence dyes, are used to quantify the plasmid molecules in a distinct number of cells throughout the PCR amplification. The quantitative detection of the DNA hybrids is realized by the detection of the bounded labeled specific hybridization probes or by double-strand DNA-specific fluorescence dyes like SYBR Green I.

The mRNA level of a bioprocess-relevant gene can be used as an early measure of its expression status. A simple but labor-intensive offline technique for the mRNA analysis of one or a limited number of genes is Northern blotting. Following the separation of the RNA sample in a polyacrylamide gel, the RNA is blotted on a membrane and detected by the hybridization of labeled sequence-specific probes. This technique does not only allow a quantitative analysis of the mRNA level but also gives information about the length and potential degradation products of the RNA.

Based on the knowledge of the base composition of a distinct bacterial genome, a global expression profiling of all genes of an organism can be performed by means of genomic DNA chips. The mRNAs are rewritten into cDNAs and labeled, usually with the fluorescence dyes Cy3 or Cy5. The labeled cDNA sample is incubated on the chip and binds at the chip surface by forming a duplex with its complement probe. The intensity of the hybridization signal allows an estimation of the level of distinct mRNAs in the investigated cell sample. The detection of the fluorescently labeled hybrids on the chip is usually based on confocal laser scanning.

The expression analysis by means of optical DNA chips is only useful for an offline analysis of selected critical bioprocess-relevant genes. The analysis time for the mRNA by these approaches from early cell disruption and RNA isolation to final data analysis lasts several hours. Alternative shorter and less labor-intensive techniques have been suggested and are

being developed. In comparison to common optical DNA chips, these techniques do not allow a parallel analysis of thousands of RNA species at one time, but are useful for process state analysis based on tens of key RNAs. One such approach is the real-time reverse transcription-polymerase chain reaction (real-time RT-PCR). This technique is also based on a rewriting of the mRNA with specific primers by reverse transcription into cDNA, which is then used as a template in the fast PCR reaction. Real-time RT-PCR allows the simultaneous analysis of ~10–40 mRNA species in ~1 h. The detection can be realized either by labeled specific hybridization probes or double-strand-specific fluorescence dyes.

Even faster analysis of mRNAs is possible with a sandwich hybridization system for which two oligonucleotide probes, acting either as capture or detection probe, are directly bound to the target RNA species. Although this method currently does not reach the sensitivity of real-time RT-PCR it has the advantage of being applicable for measurements directly in crude cell extracts. This method can be automated for large sample numbers in pipetting robots and has been combined with different read-out principles. An interesting recent development for bioprocess state analyses is an electrical chip-based read-out system for DNA/RNA hybrids by an electrochemical detection with interdigitated electrodes. This electrical chip technique is based on the coupling of the hybridization event to an enzymatic reaction, which produces a redox-active product. In this approach the mRNAs are efficiently captured by oligonucleotide probes located on magnetic beads or directly on the electrode surface. DNA/RNA hybrids are then detected with alkaline phosphatase-coupled DNA detection probes. The enzyme converts the redox-inactive substrate *para*-aminophenol phosphate into redox-active *para*-aminophenol, which by recycling between the electrodes acts as a second-stage signal amplifier.

The two-dimensional polyacrylamide gel electrophoresis technique (2D-PAGE) is used for the investigation of the protein pattern of a cell population, the proteome, at a distinct time point of a bioprocess. The proteome represents the sum of all proteins of one cell or cellular compartment based on the genome sequencing. The 2D-PAGE technique in combination with N-terminal protein sequencing or mass spectroscopic techniques, particularly matrix assisted laser desorption/ionization–time-of-flight MS, allows the identification of critical proteins that are relevant for the performance of biotechnological processes. Furthermore, the proteome analysis technique is used to characterize protein turnover or post-translational modifications, which are

important parameters for the quality control of recombinant protein production processes. The proteomics approach, however, has some limitations. Although 2D-PAGE analysis allows the identification of the majority of the proteins, it does not give a comprehensive overview of the gene expression activity of the cells. Critical proteins like hydrophobic, very small, or low abundant proteins cannot easily be visualized by the 2D-PAGE technique.

Selected proteins, which are of special interest for the monitoring of a bioprocess, can also be analyzed by protein chips. Protein chips are either based on antibodies, which capture the appropriate antigens out of a protein sample, or on different chromatographic surface properties, which attract only distinct proteins. The detection of antibody-based protein chips is done with labeled antibodies recognizing an additional suitable epitope of the target protein. These protein chips can be read out in a DNA chip laser scanner. The chromatographic protein chips possess, for example, different hydrophobic, hydrophilic, cation exchange, anion exchange, or immobilized-metal affinity surfaces. The proteins interacting with the chemistry of the chip surface can be analyzed in special protein-chip readers by a time-of-flight mass spectrometer.

The drawbacks of the proteomic approach can be partially circumvented by the DNA array technology. The combination of both techniques allows a comprehensive direct view on the physiological state of the cells during a bioprocess. The sequencing of the genomes in concert with proteome and transcriptome analyses, including the coverage of the most essential metabolites (metabolome approach) of different industrially relevant hosts, opens up completely new possibilities for solving problems connected with the control and optimization of biotechnological processes.

Metabolic changes in bioprocesses such as certain stress events are better reflected by the change of the synthesis rates of proteins than by the total composition of the cell. Therefore, in microbial physiology short-time labeling by radioactive precursors and following chasing by addition of the unlabeled form of the same precursor is a well-established principle. However, this method is difficult to apply in bioprocesses, where fermentation volumes are in the range of liters to m<sup>3</sup>. Such processes cannot easily be scaled down to the milliliter scale, which is commonly applied for these investigations. For offline labeling it must be taken into account that the conditions (especially aeration, nutrient supply, mixing) in the sampling vial are different from the fermenter and therefore very short labeling intervals (1 min) and reliable controls must be applied. Nevertheless,

the detection of synthesis rates has been applied for cell samples from fermentations. These analyses provide valuable information concerning the total cellular synthesis rates (e.g., for translation, transcription, and replication) of the production cells during a bioprocess.

See also: **Electrophoresis:** Two-Dimensional Gels; Nucleic Acids. **Enzymes:** Enzyme-Based Assays. **Flow Injection Analysis:** Principles. **Fluorescence:** Quantitative Analysis. **Lab-on-a-Chip Technologies.** **Mass Spectrometry:** Matrix-Assisted Laser Desorption/Ionization; Time-of-Flight. **Microelectrodes.** **Microscopy:** Overview. **pH.** **Process Analysis:** Overview; Chromatography; Electroanalytical Techniques; Sensors; Acoustic Emission; Maintenance, Reliability, and Training. **Proteins:** Overview. **Proteomics.** **Purines, Pyrimidines, and Nucleotides.** **Sensors:** Overview. **Spectrophotometry:** Overview.

## Further Reading

- Christensen B and Nielsen J (2000) Metabolic network analysis. A powerful tool in metabolic engineering. *Advances in Biochemical Engineering/Biotechnology* 66: 209–231.
- Enfors SO, Jahic M, Rozkov A, *et al.* (2001) Physiological responses to mixing in large scale bioreactors. *Journal of Biotechnology* 85: 175–185.
- Gabig-Ciminska M, Holmgren A, Andresen H, *et al.* (2004) Electric chips for rapid detection and quantification of nucleic acids. *Biosensors and Bioelectronics* 19: 537–546.
- Harms P, Kostov Y, and Rao G (2002) Bioprocess monitoring. *Current Opinion in Biotechnology* 13: 124–127.
- Hewitt CJ and Nebe-Von-Caron G (2001) An industrial application of multiparameter flow cytometry: Assessment of cell physiological state and its application to the study of microbial fermentations. *Cytometry* 44: 179–187.
- Komives C and Parker RS (2003) Bioreactor state estimation and control. *Current Opinion in Biotechnology* 14: 468–474.
- Larsson G and Törnkvist M (1996) Rapid sampling, cell inactivation and evaluation of low extracellular glucose concentrations during fed-batch cultivation. *Journal of Biotechnology* 49: 69–82.
- Mandenius CF (1999) Electronic noses for bioreactor monitoring. *Advances in Biochemical Engineering/Biotechnology* 66: 65–82.
- Meyer S, Noisommit-Rizzi N, Reuss M, and Neubauer P (1999) Optimized analysis of intracellular adenosine and guanosine phosphates in *Escherichia coli*. *Analytical Biochemistry* 271: 43–52.
- Neubauer P, Åhman M, Törnkvist M, Larsson G, and Enfors SO (1995) Response of guanosine tetraphosphate to glucose fluctuations in fed-batch cultivations of *Escherichia coli*. *Journal of Biotechnology* 43: 195–204.
- Rautio J, Barken KB, Lahdenperä J, *et al.* (2003) Sandwich hybridisation assay for quantitative detection of yeast rnas in crude cell lysates. *Microbial Cell Factories* 2: 1–9.
- Schaefer U, Boos W, Takors R, and Weuster-Botz D (1999) Automated sampling device for monitoring intracellular metabolite dynamics. *Analytical Biochemistry* 270: 88–96.
- Schuster KC (2000) Monitoring the physiological status in bioprocesses on the cellular level. *Advances in Biochemical Engineering/Biotechnology* 66: 185–208.
- Schweder T and Hecker M (2004) Monitoring of stress responses. *Advances in Biochemical Engineering/Biotechnology* 89: 47–71.

# PROTEINS

Contents

**Overview**

**Traditional Methods of Sequence Determination**

**Physiological Samples**

**Foods**

## Overview

**L Skipper**, Reading University, Reading, UK

## Introduction

Proteins are linear polymers of amino acids joined by peptide bonds. Proteins with  $M_r$  above 5000 Da are arbitrarily called polypeptides and those with smaller  $M_r$  are called peptides. They perform many roles: they carry oxygen in blood; they are part of the

the detection of synthesis rates has been applied for cell samples from fermentations. These analyses provide valuable information concerning the total cellular synthesis rates (e.g., for translation, transcription, and replication) of the production cells during a bioprocess.

See also: **Electrophoresis:** Two-Dimensional Gels; Nucleic Acids. **Enzymes:** Enzyme-Based Assays. **Flow Injection Analysis:** Principles. **Fluorescence:** Quantitative Analysis. **Lab-on-a-Chip Technologies.** **Mass Spectrometry:** Matrix-Assisted Laser Desorption/Ionization; Time-of-Flight. **Microelectrodes.** **Microscopy:** Overview. **pH.** **Process Analysis:** Overview; Chromatography; Electroanalytical Techniques; Sensors; Acoustic Emission; Maintenance, Reliability, and Training. **Proteins:** Overview. **Proteomics.** **Purines, Pyrimidines, and Nucleotides.** **Sensors:** Overview. **Spectrophotometry:** Overview.

## Further Reading

- Christensen B and Nielsen J (2000) Metabolic network analysis. A powerful tool in metabolic engineering. *Advances in Biochemical Engineering/Biotechnology* 66: 209–231.
- Enfors SO, Jahic M, Rozkov A, *et al.* (2001) Physiological responses to mixing in large scale bioreactors. *Journal of Biotechnology* 85: 175–185.
- Gabig-Ciminska M, Holmgren A, Andresen H, *et al.* (2004) Electric chips for rapid detection and quantification of nucleic acids. *Biosensors and Bioelectronics* 19: 537–546.
- Harms P, Kostov Y, and Rao G (2002) Bioprocess monitoring. *Current Opinion in Biotechnology* 13: 124–127.
- Hewitt CJ and Nebe-Von-Caron G (2001) An industrial application of multiparameter flow cytometry: Assessment of cell physiological state and its application to the study of microbial fermentations. *Cytometry* 44: 179–187.
- Komives C and Parker RS (2003) Bioreactor state estimation and control. *Current Opinion in Biotechnology* 14: 468–474.
- Larsson G and Törnkvist M (1996) Rapid sampling, cell inactivation and evaluation of low extracellular glucose concentrations during fed-batch cultivation. *Journal of Biotechnology* 49: 69–82.
- Mandenius CF (1999) Electronic noses for bioreactor monitoring. *Advances in Biochemical Engineering/Biotechnology* 66: 65–82.
- Meyer S, Noisommit-Rizzi N, Reuss M, and Neubauer P (1999) Optimized analysis of intracellular adenosine and guanosine phosphates in *Escherichia coli*. *Analytical Biochemistry* 271: 43–52.
- Neubauer P, Åhman M, Törnkvist M, Larsson G, and Enfors SO (1995) Response of guanosine tetraphosphate to glucose fluctuations in fed-batch cultivations of *Escherichia coli*. *Journal of Biotechnology* 43: 195–204.
- Rautio J, Barken KB, Lahdenperä J, *et al.* (2003) Sandwich hybridisation assay for quantitative detection of yeast rnas in crude cell lysates. *Microbial Cell Factories* 2: 1–9.
- Schaefer U, Boos W, Takors R, and Weuster-Botz D (1999) Automated sampling device for monitoring intracellular metabolite dynamics. *Analytical Biochemistry* 270: 88–96.
- Schuster KC (2000) Monitoring the physiological status in bioprocesses on the cellular level. *Advances in Biochemical Engineering/Biotechnology* 66: 185–208.
- Schweder T and Hecker M (2004) Monitoring of stress responses. *Advances in Biochemical Engineering/Biotechnology* 89: 47–71.

# PROTEINS

Contents

**Overview**

**Traditional Methods of Sequence Determination**

**Physiological Samples**

**Foods**

## Overview

**L Skipper**, Reading University, Reading, UK

## Introduction

Proteins are linear polymers of amino acids joined by peptide bonds. Proteins with  $M_r$  above 5000 Da are arbitrarily called polypeptides and those with smaller  $M_r$  are called peptides. They perform many roles: they carry oxygen in blood; they are part of the

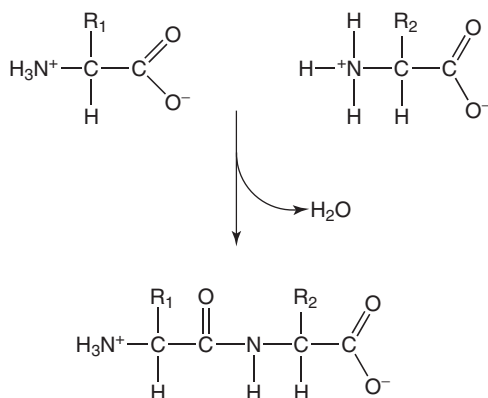
muscles that lever bones; they reinforce teeth and bone; they are the antibodies that fight infection; they are some of the hormones that provide signals to organs; they regulate gene function; and some of the proteins are the enzymes that control metabolism.

Protein function is related to protein structure, which is considered on four levels: the primary structure, which is the specific sequence of amino acids joined together by covalent peptide bonds; the secondary structure, which is based on how amino acids interact, promoting protein folding; the tertiary structure, which represents the overall three-dimensional shape of the protein; and the quaternary structure, where single proteins, or subunits, combine to make up a multisubunit protein.

## Primary Protein Structure

There are 20 common amino acids that in various numbers and combinations form all proteins, although there are a few exceptions. The essential ingredient of an amino acid is the three-atom group nitrogen-carbon-carbon. The nitrogen is present as part of an amine group and, under certain physiological conditions is positively charged. The central carbon, or the  $\alpha$ -carbon, has a side group or residue that contributes much of the chemical and physical properties to each of the 20 amino acids. The last carbon is part of a carboxylate group and under certain physiological conditions is either neutrally or negatively charged. Individual amino acids are joined together enzymatically by the elimination of water in a condensation reaction that yields a covalent peptide bond between the carboxyl carbon of one amino acid and the amine nitrogen of another. **Figure 1** shows peptide bond formation.

The first amide atom is represented as  $N_i$ , where  $i$  is the number of the amino acid in the chain,

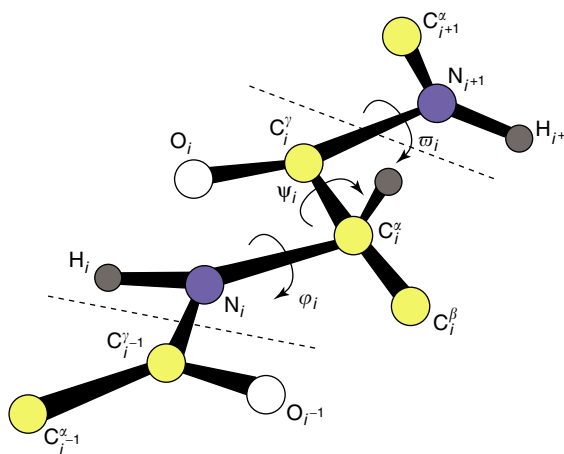


**Figure 1** Covalent peptide bond formation by condensation.

numbered from the amino end of the chain. The carbon to which the first carbon of the amino acid residue bonds is termed the  $\alpha$ -carbon,  $C_i^\alpha$ , and the carbon involved in the peptide bond is  $C_i'$ . The subsequent carbons in the side chain from the  $C_i^\alpha$  are termed  $C_i^\beta$ ,  $C_i^\gamma$ ,  $C_i^\delta$ ,  $C_i^\epsilon$ , and so on. The repeated atoms of  $N_i$ ,  $C_i^\alpha$ , and  $C_i'$  make up the backbone of the peptide. Three-dimensional crystal structures solved for small peptides have revealed the dimensions of the peptide group.

The average dimensions solved in many crystal structures show that almost all peptide bonds are in a *trans* configuration creating staggered residues due to steric hindrance between nonbonded atoms. One main exception occurs with the amino acid proline. Its cyclic side chain reduces the repulsion between atoms and so favors the *trans* and *cis* conformations equally (see **Figure 2**).

The expected length of a single C–N bond is 1.45 Å, as in the  $C^\alpha$ –N bond, and that of a C=N double bond is 1.25 Å. The actual length of the  $C'$ –N peptide bond is 1.33 Å, showing that it has partial double bond characteristics (40% double bond). Rotation can occur, in principle, around all three bonds  $\psi$ ,  $\phi$ , and  $\varpi$ , where  $\psi = \phi = \varpi = 180^\circ$ . This means that for a protein of 100 residues there are  $2 \times 10^{90}$  possible conformations, far more possible conformations than there would be protein molecules, even in a large sample. However, we know that a folded protein has a relatively stable conformation. This is due to many factors, one being the partial double bond characteristics of the  $C'$ –N peptide bond that limits it to a *trans* conformation (with the exception of proline), the atoms of the side chains restrict bond rotation due to excluded volume effects that dictate



**Figure 2** Diagram showing the typical *trans* conformation of the backbone atoms. Phi ( $\phi$ ) and psi ( $\psi$ ) bond angles determine the conformation. Dotted lines delimit a single amino acid.



that two atoms cannot exist in the same space as defined by the van der Waals radius and by repulsion effects of atoms in the side chains. The whole conformation is also stabilized by weak electrostatic interactions between atoms, and between atoms and the solvent and salts.

The chemical properties of amino acids make them weak buffers, so they can either accept or donate protons (see Figure 3). This means that their electrical properties can change with changes in pH. First, the amino and carboxyl groups of amino acids are ionizable. At a low pH, the amino and carboxyl groups will be protonated. The amino group will carry a positive charge and the carboxyl group will be neutral. The overall charge is therefore positive. At a high pH, both groups will be deprotonated and the overall charge will be negative. The pH at which the amino group is protonated and carboxyl group is deprotonated, and therefore has no overall charge, is said to be the amino acids' isoelectric point ( $pI$ ) (Table 1). This property of amino acids means that proteins are susceptible to changes in pH and salt concentration and that stability is dependent on interactions between protein, solvent, and intermolecular interactions. At a pH close to a protein's  $pI$ , these interactions cannot take place, which results in precipitation.

The properties of individual amino acids in the primary structure greatly affect the structure of a protein at the secondary, tertiary, and quaternary levels. In fact, a single mutation in the genetic code may alter only a single amino acid in a large protein but result in dramatic changes in structure. An example of this is the case of sickle-cell anemia, where a single amino acid change causes reduced function of hemoglobin. Hemoglobin comprises four subunits,  $2\alpha 2\beta$ . A single mutation in the gene that codes for the  $\beta$ -subunit of hemoglobin results in residue 6, Val, being replaced by Glu. Consequently, a charged

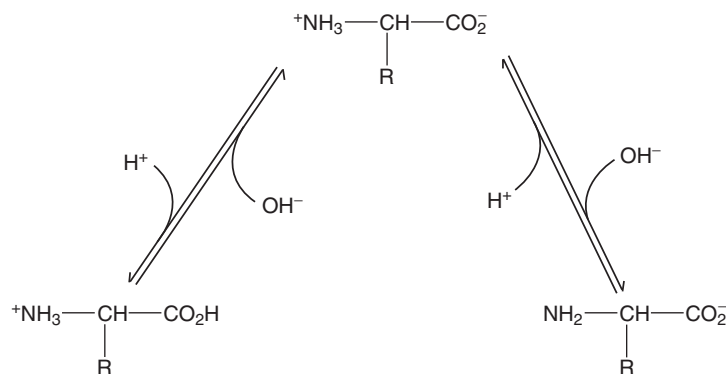
residue of a surface  $\alpha$ -helix is replaced with a hydrophobic residue. The hydrophobic region on the surface of the protein fits and binds into a hydrophobic pocket of a second hemoglobin molecule when in the deoxygenated form. The result is the polymerization of hemoglobin molecules in red blood cells. This causes the cells to deform into a sickle shape and their oxygen-carrying properties are lost. If the mutation only occurs in one gene copy, the effects are limited, but for homozygotes the disease is fatal without treatment.

The ability of proline to adopt a *cis* conformation in its peptide bond, for example, can create hairpin turns along the length of the protein. The thio groups of a cysteine side chain can be oxidized to form a covalent disulfide bridge between two adjacent cysteines. Disulfide bridges stabilize protein structure and can form large loops in proteins.

## Primary Structure Determination


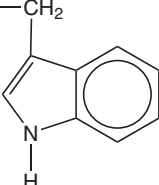
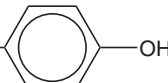
The amino acid sequence can be determined in a number of ways. The sequence can be established directly by combining a number of different techniques. The sequence of the protein can also be determined from the sequence of the gene that codes for it, and the primary, secondary, tertiary, and quaternary structures can be determined crystallographically, which gives a complete model of the protein.

The amino and carboxyl residues can be elucidated chemically by labeling them. Two common methods modify the N-terminal residue by the chemical addition of fluoro-2,4-dinitrobenzene or dansyl chloride. The protein is then hydrolyzed to its component amino acids. The amino acid labeled with fluoro-2,4-dinitrobenzene is colored yellow and that with dansyl chloride is fluorescent. This enables visualization of the labeled amino acid. Comparison of further



**Figure 3** Diagram showing the ionization of a single amino acid. Top species shows an amino acid's overall neutral charge when  $\text{pH} = pI$ . Bottom left shows protonation when  $\text{pH} < pI$ . Bottom right shows deprotonation when  $\text{pH} > pI$ .

**Table 1** Isoelectric points for the major protein amino acids

Side chain, R	Symbol	Name	pI	Type
—H	Gly	Glycine	5.97	Side chain is nonpolar
—CH <sub>3</sub>	Ala	Alanine	6.00	
—CH <sub>2</sub> (CH <sub>3</sub> ) <sub>2</sub>	Val	Valine	5.96	
—CH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	Leu	Leucine	5.98	
—CHCH <sub>2</sub> CH <sub>3</sub>   CH <sub>3</sub>	Ile	Isoleucine	6.02	
—CH <sub>2</sub> — 	Phe	Phenylalanine	5.48	
—CH <sub>2</sub> — 	Trp	Tryptophan	5.89	
—OC(=O)—CH—CH <sub>2</sub>   H <sub>2</sub> N—CH <sub>2</sub> —CH <sub>2</sub>	Pro	Proline	6.30	
—CH <sub>2</sub> OH	Ser	Serine	5.68	
—CH(OH)   CH <sub>3</sub>	Thr	Threonine	5.64	
—CH <sub>2</sub> — 	Tyr	Tyrosine	5.66	
—CH <sub>2</sub> CO <sub>2</sub> H	Asp	Aspartic acid	2.77	
—CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> H	Glu	Glutamic acid	3.22	
—CH <sub>2</sub> CONH <sub>2</sub>	Asn	Asparagine	5.41	
—CH <sub>2</sub> CH <sub>2</sub> CONH <sub>2</sub>	Gln	Glutamine	5.65	
—CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> NH <sub>2</sub>	Lys	Lysine	9.74	Side chain has a basic amino group
—CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> NH—C(=NH <sub>2</sub> )—NH <sub>2</sub>	Arg	Arginine	10.76	
—H <sub>2</sub> CC(=N)N=CH—   N—H	His	Histidine	7.59	
—CH <sub>2</sub> SH	Cys	Cysteine	5.07	
—CH <sub>2</sub> CH <sub>2</sub> SCH <sub>2</sub>	Met	Methionine	5.74	

chromatographic analysis of the labeled amino acids with standards identifies it. The Edman degradation is often used to determine the amino acid sequence. This involves the sequential degradation of the N-terminal amino acid. The protein is reacted with phenylisothiocyanate (PITC) to give phenylthiocarbamyl (PTC) derivative that makes the first N-terminal peptide bond unstable; mild hydrolyzation then liberates the first amino acid in the sequence but leaves the stable peptide bonds in place. The liberated amino acid can then be identified. The process is repeated and each amino acid in the chain is determined one at a time. This method has been automated to speed up the process and to make it more reliable, but it is not without its drawbacks.

The carboxyl group of the C-terminal amino acid is identified in a similar way, although the results are less reliable. One of the most common methods used involves converting all the C $_{\alpha}$  atoms into hydrazides. This liberates most, but not all, of the C-terminal residues in the sample and allows their identification. Further limitations occur as Asn and Gln cannot be liberated since their side chain amides are converted to hydrazides and Arg residues are converted to ornithine and are not liberated either. Another method uses acetic anhydride to form an oxalone specifically on the C-terminal residue. The converted residue can be labeled with deuterium or tritium and then liberated by hydrolysis and identified.

Mass spectroscopy is increasingly being used to resolve amino acid sequences due to recent developments in this technique. There are several different methods that can be used, but basically this method determines the mass-to-charge ratio of individual amino acids of a fragmented protein. The pattern of fragmentation can be extremely complex, but peptide bonds are mostly broken to cause the fragmentation. The patterns of fragmentation are known, and this allows the sequence to be determined.

The most commonly used, and most reliable, method for determining primary structure is the determination of the DNA sequence of the gene that codes the protein. It is necessary to know the sequence of the first and last 6–10 amino acids of the protein to be identified so that primers can be generated that correspond to the non-sense strand of the gene in question. Polymerase chain reaction (PCR) can then be used to ‘pull out’ a copy of the gene that can then be cloned, propagated, and sequenced. This method still has its challenges due to the redundancies of the DNA triplet code, where more than one triplet codes the same amino acid. If the wrong triplet is chosen in the primer then it may not hybridize to the corresponding sequence in the gene. Proteins also often undergo posttranslational

modification and only knowing the gene sequence will not give any information as to what these might be.

## Secondary Protein Structure

$\alpha$ -Helices,  $\beta$ -sheets, and triple helices are three types of secondary structures. All are formed and stabilized by noncovalent interactions, mainly hydrogen bonds. Proteins can contain only one or mixtures of secondary structure as well as portions of the protein that contain structures difficult to describe but not less ordered, such as reverse turns and  $\beta$ -bends.

$\alpha$ -Helices are regular right-hand turns of amino acids 3.6 residues long; 5.41 Å. Hydrogen bonding between the first backbone carbonyl oxygen atom and the fourth residue NH group stabilizes the structure; van der Waals interactions across the axis further stabilize the structure. There are some rare exceptions to this general scheme where hydrogen bonding can occur between three residues ( $3_{10}$ -helix) or between five residues ( $\pi$ -helix). These structures are much less stable than general  $\alpha$ -helices and are not normally favored. Any of the 20 amino acids can participate in an  $\alpha$ -helix but some are more favored than others. Ala, Glu, Leu, and Met are most often found in helices whereas, Gly, Tyr, Ser, and Pro are less likely to be seen. Proline, for instance, is rarely seen as its backbone nitrogen is bonded to its cyclic side group and cannot participate in hydrogen bonding. When prolines are found in  $\alpha$ -helices, they tend to cause the helix to bend due to steric hindrance caused by its side group. They can be found as the first or last residue in the helix where they do not cause bending. The side groups of the other amino acids point out and down relative to the helix. In globular proteins, those that are hydrophobic tend to be on one side of the helix and interact with other amino acids of the protein, and those on the other side are generally hydrophilic and interact with the solvent. For this reason,  $\alpha$ -helices of globular proteins are predominantly found on the protein surface and have polar, hydrophobic, and hydrophilic amino acids. On average,  $\alpha$ -helices in globular proteins have 11 residues,  $\sim 17$  Å long. Some  $\alpha$ -helices have mainly hydrophobic residues, which are found buried in the hydrophobic core of a globular protein, or are transmembrane proteins.

$\beta$ -Sheets are formed by the interactions between parallel regions of a protein chain. These either run in the same direction, parallel; or in the opposite direction, antiparallel. These structures are stabilized by hydrogen bonds between backbone carbonyl oxygen atom and the hydrogen of the amino group. In parallel  $\beta$ -sheets, the distances between the carbon

and nitrogen involved in binding on one strand and on the other differ. This means that the hydrogen bonds are at an angle in relation to the protein strand. This is thought to make parallel  $\beta$ -sheets less stable than antiparallel  $\beta$ -sheets. In antiparallel  $\beta$ -sheets, the atoms on opposite strands involved in hydrogen binding are the same distance so that hydrogen bonds are at  $90^\circ$  to the strand.  $\beta$ -Sheets are not flat but have a pleated appearance due to the  $C_i^\alpha$  atoms being successively above and below the plane of the sheet. The side groups are also successively above and below the plane of the sheet and they, therefore, cannot interact with each other. They do have significant interactions with neighboring side chains and with their backbone. Proteins can contain parallel  $\beta$ -sheets, antiparallel  $\beta$ -sheets, or a mixture of both, although mixed proteins account for only 20% of proteins with  $\beta$ -sheets. The strands are typically 5–10 amino acids long and  $\beta$ -sheets contain 2–15 strands. The strands in  $\beta$ -sheets always have a right-handed twist.

$\alpha$ -Helices and  $\beta$ -sheets make up  $\sim 50\%$  of a protein's secondary structure. These structures are connected by looped regions that change the direction of the protein strand.  $\beta$ -Bends connect subsequent strands in a  $\beta$ -sheet, so that they are aligned parallel. Each bend consists of two amino acids, usually Pro because of the flexibility of its C'-N peptide bond and Gly because its side chain gives the least crowding in the tight turn. The structure is stabilized by hydrogen bonding between the residues immediately before and after the bend.  $\Omega$ -loops are so called as they have the general shape of the letter omega. These are usually 6–16 residues long and connect  $\alpha$ -helices and  $\beta$ -sheets. They are predominantly found at the protein surface in globular proteins or connecting membrane  $\alpha$ -helices at the cytoplasmic or extracellular face. These structures are often involved in recognition processes. Some of the remaining structure is disordered but nonetheless important, as these regions often confer great flexibility to proteins essential to their function.

## Secondary Structure Determination

Secondary structure is difficult to determine directly, although models of a protein can be built using data gathered with nuclear magnetic resonance (NMR) or by X-ray diffraction of protein crystals. Because of these difficulties, unknown secondary structure is often predicted based on what is known about secondary structure of other proteins.

Many methods have been developed to try to predict secondary structure, based on the knowledge of the primary structure of a protein. Knowledge of the

position and properties of amino acid allows assignment of probable structure, although this can be unreliable and very limited. It is most successful in predicting the  $\alpha$ -helices involved in transmembrane stretches, and there are a number of computer programs that can do this because of the hydrophobic nature of the  $\alpha$ -helices. These programs also try to predict  $\alpha$ -helices and  $\beta$ -sheets of globular proteins but are much less reliable because they use information from models of known structure based on sequence similarities. Certain secondary structures, folds, or motifs, are conserved in families of proteins. Prediction programs and sequence comparisons can often spot these, but it is very difficult to predict how individual amino acids in the structure are influenced by their local environment, so these predictions often do not give an absolute determination of secondary structure.

Circular dichroism (CD) spectroscopy can reliably assign the proportions of  $\alpha$ -helices and  $\beta$ -sheets in a protein based on the differing spectral properties of their backbones. This does not give any more information about secondary structure but is useful because it does tell that a protein is folded, although not whether it is folded correctly. Comparison of CD spectra for a protein of unknown structure with that of known structure can provide a good interpretation.

## Tertiary Protein Structure

Tertiary structure is the folding of the secondary structure into distinct arrangements known as domains. This occurs spontaneously in cells in one to several seconds and is driven by the properties of the amino acid side chains. Hydrophobic side chains cannot break the hydrogen bonding of solvent water molecules and so the nonpolar residues Val, Leu, Ile, Met, and Phe tend to bury themselves in the center of the protein; this is one of the main forces that drives protein folding. The core of a globular protein is tightly packed, thus excluding most water molecules, and is highly ordered. Hydrophilic side chains favor the surface of globular proteins. Charged polar residues Arg, His, Lys, Asp, and Glu are mostly found on the surface and the uncharged polar residues Ser, Thr, Asn, Gln, Tyr, and Trp are found on both the interior and on the surface and are the hydrogen bond donors that help stabilize the tertiary structure. Ionic forces also contribute to structure stabilization at physiological pH. Asp, for instance, carries a negative charge and Glu carries a positive charge. These charges attract each other when in close proximity and form a salt bridge.

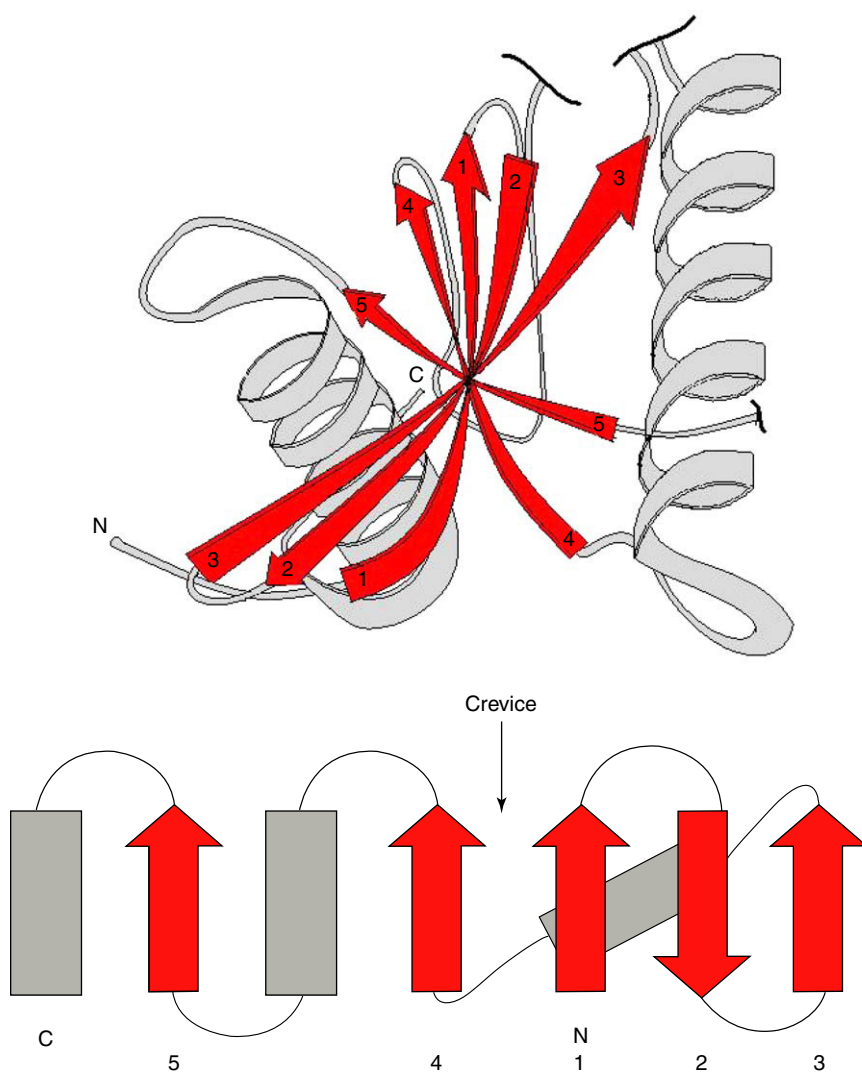
A protein can have one or several distinct domains linked by loop regions. These domains contain

100–200 residues typically and can give functional as well as structural properties to the protein. Protein domains can be divided into three main groups;  $\alpha$ -domains,  $\beta$ -domains, and  $\alpha/\beta$ -domains.  $\alpha$ -Domains consist mainly of  $\alpha$ -helices at the protein's core,  $\beta$ -domains consist of antiparallel  $\beta$ -sheets, and  $\alpha/\beta$ -domains consist of  $\beta$ -sheets surrounded by  $\alpha$ -helices in a  $\beta\alpha\beta$  conformation.

$\alpha/\beta$ -Domains are the most frequent structures in globular proteins. They can be divided into three main classes, depending on their conformation. The first is the  $\beta$ -barrel, characterized by a barrel-like structure where  $\alpha$ -helices connect parallel  $\beta$ -strands on the outside of the barrel. The second is the Rossman fold, after Michael Rossman, who first described the structure, which is an open, twisted

$\beta$ -sheet not unlike the rungs of a ladder surrounded on both sides by  $\alpha$ -helices. These folds are often associated with nucleotide binding (see Figure 4, the ATP binding domain of hexokinase for an example). The third is called the horseshoe fold as its shape resembles a horseshoe. The  $\beta$ -strands form a curved structure with  $\alpha$ -helices on the outside face of the curve. The horseshoe fold is also often called leucine-rich motif as it contains repetitive leucine residues.

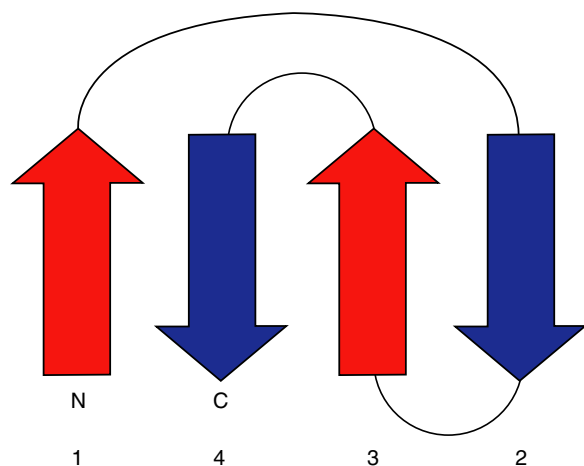
$\beta$ -Barrels often have active sites in very similar positions, usually at the carboxy end of the barrel. This is also true for the open folds but binding of ligands is very different. The open folds do not form a cylinder as in the barrel structures. Instead, binding occurs in a cleft formed by loops linking and going in and out of the sheet. The domain structures are very



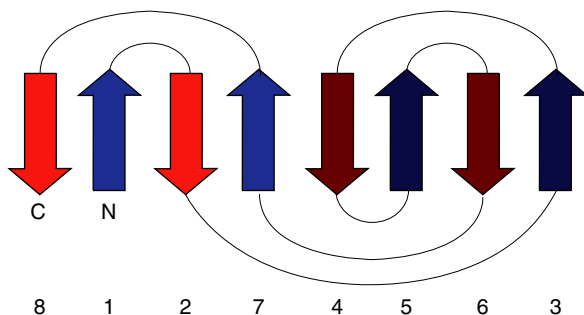
**Figure 4** Cartoon and topology diagram of the ATP binding domain of hexokinase. The crevice formed between  $\beta$ -sheets 1 and 4 is the site of ATP binding.

similar even though their amino acid sequences may differ somewhat. It is this similarity in structure and function that is used to assign proteins into families.

$\beta$ -Domains are the second most frequent structure in proteins and are the most diverse. They generally consist of two antiparallel sheets packed together to form a barrel and can be divided into three domain groups. Up-and-down  $\beta$ -barrels are adjacent antiparallel strands connected by hairpin loops. These structures are found in many proteins, which include transport proteins and viral coat proteins as well as many others. Greek key motifs are barrel structures with a basic unit made up of four-stranded antiparallel  $\beta$ -sheets. The first strand is connected to the fourth strand by a hairpin loop. Strands 2 and 3 are connected by a loop on the opposite end of the barrel (see Figure 5). These motifs are repeated to form the barrel. The Greek key motif is a component structure of the jelly-roll domain. A simple illustration is given in Figure 6.



**Figure 5** Topology diagram showing the organization of an antiparallel  $\beta$ -sheet into a Greek key motif.

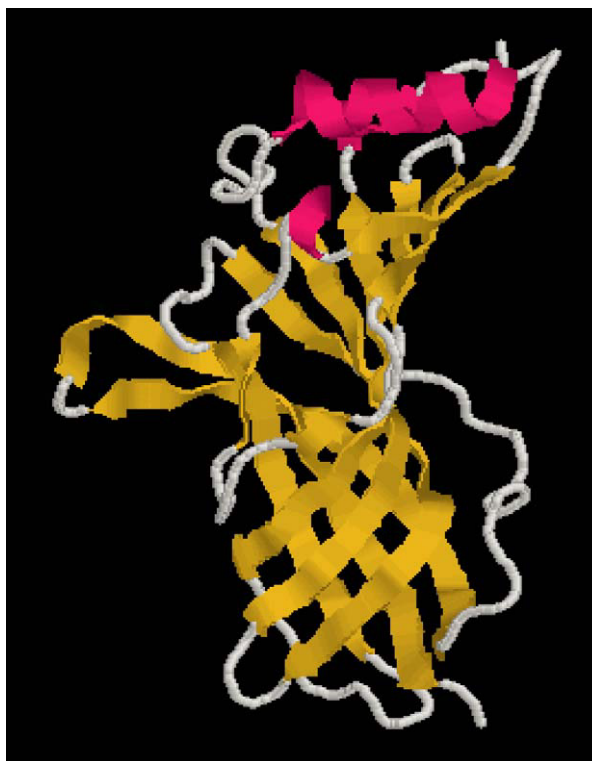


**Figure 6** Topology diagram showing the organization of two Greek key motifs into a jelly-roll fold.

## Tertiary Structure Determination

It is possible to predict the tertiary structure of a protein by comparing the amino acid sequence of the unknown protein with those whose structure is known. Structures can be assigned based on unique folds or domains that have conserved amino acids and, although the amino acid sequence of a specific fold may differ, amino acids with similar properties are conserved to maintain the structural and functional properties of the protein. It is possible, with relative accuracy, to assign proteins to a particular protein family in this way.

The complete protein structure can be derived by X-ray crystallography. The technique requires the target protein to be purified to homogeneity and concentrated, typically to  $5\text{--}10\text{ mg ml}^{-1}$ . The buffer that supports the protein is then slowly exchanged into a supersaturated solution containing a precipitant at a given pH. Salts and polyethylene glycol are commonly used precipitants. The type and concentration of precipitant and pH used are generally arbitrary as it is extremely difficult to predict under which conditions the protein will precipitate to form crystals. Crystal screens of  $\sim 100$  conditions are typical as a first step to derive satisfactory conditions for precipitation.



**Figure 7** X-ray crystallographically derived 3D model of snail acetylcholine binding protein.



A crystal is a regular repeating array of many protein molecules. The level of order in the crystal determines how well it will diffract X-rays when placed in an X-ray beam. A protein crystal scatters X-rays when placed in the beam emitting its own source of X-rays. The direction and intensity of these emissions are recorded on X-ray detection plate. The pattern of diffraction can provide information about the individual atoms of the crystal, but extracting this information requires significant computation. The result is an electron density map that can be used to interpret the structure of the target protein, although this in itself is complicated (see **Figure 7** for an X-ray crystal model of acetylcholine binding protein).

### Quaternary Protein Structure

Protein subunits, or the tertiary structures, can come together to produce a multisubunit protein. These can be two to several identical or different subunits. They may be functionally distinct or can act together cooperatively. Hydrogen and ionic bonds hold them together as well as disulfide bridges in some cases. Subunits may remain together for function or come together as part of their function.

### Quaternary Structure Determination

The quaternary structure is usually determined by X-ray crystallography, as described previously. However, when crystallographic data were difficult or impossible to gather, electron microscopy had provided some clues to quaternary structure.

It is relatively simple to determine the multisubunit composition of a protein experimentally. By disrupting the interactions that hold multisubunits together, it is possible to establish the molecular weight of each by using size-exclusion chromatography. If the method is repeated without bond disruption, the composition can be determined by using simple mathematics. Ultracentrifugation of a protein sample in a viscous glucose solution can ascertain molecular weights based on their sedimentation coefficients, and, in conjunction with light scattering, which measures the molecular dimensions of proteins in solution, some information about how the multisubunit protein is organized can be determined.

*See also:* **Proteins:** Traditional Methods of Sequence Determination; Physiological Samples; Foods.

## Traditional Methods of Sequence Determination

**D H Schlesinger**, New York University Medical Center, New York, NY, USA

© 2005, Published by Elsevier Ltd.

This article is reproduced from the previous edition, pp. 4215–4220, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

The need for automated methods to determine the primary structure of a protein was never more apparent than in Sanger's labor-intensive landmark and classic effort of the amino acid sequence determination of insulin in the early 1950s. Practically coincident in time with this successful *tour de force* was the development of a straightforward chemical method for the determination of the amino acid sequence of a peptide or protein by Edman, who not only fulfilled this need but provided the groundwork for automation of the method. Seventeen years later, Edman developed the first automated spinning-cup, or liquid-phase, sequencer. This technology became virtually universal and routine for the determination of the primary structure of any peptide, protein, or glycoprotein. Edman chemistry, also known as the phenylisothiocyanate degradation, remains in the

1990s as the most generally accepted procedure for the efficient sequencing of both small peptides and large proteins, although many variations on the basic theme have been developed to optimize the system.

Currently, Edman chemistry manifests itself in three modes: the automated liquid phase, solid phase, and gas phase, all three of which developed from the manual method, which today still has some appeal and usefulness in some laboratories. The most significant development in phenylisothiocyanate chemistry was the advent, due to Edman and Begg, of the automated spinning cup in 1967, which was successfully developed by Beckman Instruments. Just four years later, the first automated solid-phase sequencer, from Laursen, was described and developed by Sequemat, a Boston-based company founded by two graduate students of Laursen. Both of these machines overcame most of the limitations (lack of speed and labor-intensiveness) of the manual method, although both technologies had inherent problems of their own. For example, the spinning-cup sequencer encountered large peptide extractive losses when handling small peptides, and the solid-phase method

A crystal is a regular repeating array of many protein molecules. The level of order in the crystal determines how well it will diffract X-rays when placed in an X-ray beam. A protein crystal scatters X-rays when placed in the beam emitting its own source of X-rays. The direction and intensity of these emissions are recorded on X-ray detection plate. The pattern of diffraction can provide information about the individual atoms of the crystal, but extracting this information requires significant computation. The result is an electron density map that can be used to interpret the structure of the target protein, although this in itself is complicated (see **Figure 7** for an X-ray crystal model of acetylcholine binding protein).

### Quaternary Protein Structure

Protein subunits, or the tertiary structures, can come together to produce a multisubunit protein. These can be two to several identical or different subunits. They may be functionally distinct or can act together cooperatively. Hydrogen and ionic bonds hold them together as well as disulfide bridges in some cases. Subunits may remain together for function or come together as part of their function.

### Quaternary Structure Determination

The quaternary structure is usually determined by X-ray crystallography, as described previously. However, when crystallographic data were difficult or impossible to gather, electron microscopy had provided some clues to quaternary structure.

It is relatively simple to determine the multisubunit composition of a protein experimentally. By disrupting the interactions that hold multisubunits together, it is possible to establish the molecular weight of each by using size-exclusion chromatography. If the method is repeated without bond disruption, the composition can be determined by using simple mathematics. Ultracentrifugation of a protein sample in a viscous glucose solution can ascertain molecular weights based on their sedimentation coefficients, and, in conjunction with light scattering, which measures the molecular dimensions of proteins in solution, some information about how the multisubunit protein is organized can be determined.

*See also:* **Proteins:** Traditional Methods of Sequence Determination; Physiological Samples; Foods.

## Traditional Methods of Sequence Determination

**D H Schlesinger**, New York University Medical Center, New York, NY, USA

© 2005, Published by Elsevier Ltd.

This article is reproduced from the previous edition, pp. 4215–4220, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

The need for automated methods to determine the primary structure of a protein was never more apparent than in Sanger's labor-intensive landmark and classic effort of the amino acid sequence determination of insulin in the early 1950s. Practically coincident in time with this successful *tour de force* was the development of a straightforward chemical method for the determination of the amino acid sequence of a peptide or protein by Edman, who not only fulfilled this need but provided the groundwork for automation of the method. Seventeen years later, Edman developed the first automated spinning-cup, or liquid-phase, sequencer. This technology became virtually universal and routine for the determination of the primary structure of any peptide, protein, or glycoprotein. Edman chemistry, also known as the phenylisothiocyanate degradation, remains in the

1990s as the most generally accepted procedure for the efficient sequencing of both small peptides and large proteins, although many variations on the basic theme have been developed to optimize the system.

Currently, Edman chemistry manifests itself in three modes: the automated liquid phase, solid phase, and gas phase, all three of which developed from the manual method, which today still has some appeal and usefulness in some laboratories. The most significant development in phenylisothiocyanate chemistry was the advent, due to Edman and Begg, of the automated spinning cup in 1967, which was successfully developed by Beckman Instruments. Just four years later, the first automated solid-phase sequencer, from Laursen, was described and developed by Sequemat, a Boston-based company founded by two graduate students of Laursen. Both of these machines overcame most of the limitations (lack of speed and labor-intensiveness) of the manual method, although both technologies had inherent problems of their own. For example, the spinning-cup sequencer encountered large peptide extractive losses when handling small peptides, and the solid-phase method

never achieved the sequencing efficiencies of the liquid-phase machine with large proteins. To a large degree, these deficiencies were overcome with the development of the gas-phase sequencer of Hunkapiller and Hood in 1981. Applied Biosystems Inc., founded by Hood, commercially developed the first gas-phase sequencer, which is virtually ubiquitous in modern protein sequencing laboratories throughout the world.

## Edman Chemistry and the Manual Method

Edman degradation is a three-step procedure consisting of the coupling of phenylisothiocyanate (PITC) to the  $\alpha$ -amino group of a peptide or protein, cleaving the amino-terminal amino acid (via cyclization in strong per-fluorinated acid, typically trifluoroacetic acid (TFA), to a 2-anilino-5-thiazolinone), and converting the resulting thiazolinone in aqueous acid to its more stable isomer, the phenylthiohydantion (PTH) amino acid.

Coupling takes place most rapidly and completely under alkaline conditions at a pH above the pK of the amino group (pH = 7.8), at which it is uncharged, and at pH below 10, above which PITC is hydrolyzed. The coupling medium is a mixture of aqueous and organic solvents to enhance the solubility of both the peptide and the PITC reagent.

In the manual mode, coupling of the peptide is performed under nitrogen in a stoppered 5-ml conical centrifuge tube for 30 min at 54°C in 100  $\mu$ l of 0.4 mol l<sup>-1</sup> dimethylallylamine (DMAA) buffer in propanol–water (60:40 v/v), previously adjusted to pH 9.5 with TFA. Following coupling, a single extraction with 0.1 ml of benzene is performed and the organic phase is discarded. Cleavage is accomplished with 75  $\mu$ l of TFA under nitrogen at 54°C for 5 min. Cleavage products are dried under a stream of nitrogen for 30 s. The TFA-treated thiazolinone amino acid is extracted with 0.15 ml of benzene and, following drying, is converted to its PTH isomer in 0.2 ml of 1 mol l<sup>-1</sup> HCl at 80°C for 10 min. The PTH amino acids are then identified most commonly by liquid chromatography (LC). Alternative manual methods are available and are described in great detail by Edman (1950), and reviewed by Niall (1973), Allen (1981), and Schlesinger (1988) (see Further Reading).

One useful advantage of the method is that four to six peptides can be analyzed simultaneously, each cycle of degradation taking ~1 h. This multiple sample sequencing technique is especially useful for screening the chromatographic elution profile of a

mixture of peptides with respect to their quantitative yields and N-terminal sequence.

## Automated Liquid-Phase Sequence Analysis

The advent of the liquid-phase sequencer marked the beginning of the era of data acquisition of protein sequences in the early 1970s. The machine fully automated each of the three steps of Edman chemistry. In the coupling stage, buffer and PITC are delivered to a rapidly rotating cup to which the protein sample has been applied. Following coupling, extraction of excess of PITC and nonvolatile buffer is accomplished by delivery of a flow of organic solvents, which are removed by a scoop at the top of the spinning cup. Remaining buffer and derivatized protein are then dried *in situ*. Cleavage is accomplished by the automated delivery of TFA, followed again by a drying step. Extraction of the cleaved thiazolinone is effected by delivery of excess organic solvent. In the older machines, conversion to the PTH amino acid was performed manually but in the more modern devices the thiazolinones are automatically converted. A complete description of the liquid-phase sequencer has been given by Edman and Begg (1967), Niall (1973), Waterfield and Bridgen (1975), and Edman (1975).

Since the mid-1970s a number of refinements have been introduced to improve the spinning-cup sequencer's performance and efficiency. These include a cold trap to provide a better vacuum system and an automated conversion device to improve the recovery of the PTH amino acids, such as those of serine and threonine. In addition, combined delivery of benzene and ethyl acetate has been introduced to decrease peptide and protein extractive losses while still effecting removal of Quadrol, a quarternary amine buffer, and excess PITC. To achieve the same efficiency in sequencing small peptides, volatile coupling buffer systems such as dimethylallylamine and dimethylbenzylamine have been instituted in this automated methodology. Other alternative approaches have been used in conjunction with these peptide systems to decrease extractive loss in solvent washes. One involves modification of the peptide itself by reacting the  $\alpha$ - and  $\epsilon$ -amino groups of the peptide with naphthyl or 4-sulfonaphthyl isothiocyanate to render the peptide more polar, and hence less soluble, in organic solvent washes. A second method involves film-stabilizing reagents such as Polybrene. This reagent, although used initially in the liquid-phase sequencer, is currently used routinely and successfully in the gas-phase sequencer described below.

## Automated Solid-Phase Sequencing

In an attempt to decrease peptide and protein extractive loss, covalent attachment of peptides to insoluble supports was developed by Laursen in 1966. Numerous solid-phase supports are available. These include the original amino-polystyrene and triethyltetramine-polystyrene and, more recently, 3-amino-propyl glass, diisothiocyanate glass, and iodoacetyl glass. These porous glass supports have proved generally to be more effective for coupling to proteins than the polystyrene resins. These reagents couple to the  $\alpha$ - and  $\epsilon$ -amino groups of the protein, so that major disadvantages are that peptide covalent attachment is lost when sequence analysis proceeds through the last lysine residue and that the amino-coupled residue remains unidentified. For these reasons, the ideal site of attachment is the carboxy-terminal residue. However, no generally satisfactory method is available, although carbonyl diimidazole, carbodi-imide, and trifluoroacetic anhydride have been used with success in some cases. Attachment of the  $\alpha$ -amino groups of lysine, the side-chains of cysteine, aspartic and glutamic acids, and methionine to insoluble supports has been successfully accomplished, although the coupling yields, regardless of the procedure, remain variable. Nevertheless, the solid-phase method provides a wide variety of sequencing strategies to complete structure determination of relatively large proteins.

## Automated Gas-Phase Sequencing

The gas-phase sequencer was the first instrument designed to handle microquantities (0.5 pmol) of both peptides and proteins and avoid the extractive loss problems of the spinning-cup sequencer as well as the covalent attachment problems of the solid-phase approach. The gas-phase apparatus combined the miniaturization and efficient chemical flow circuitries of the solid-phase column system with the ease of sample application of the spinning-cup sequencer. In the gas-phase approach the sample is embedded in a film of Polybrene on the surface of a glass microfiber filter disk within the reaction cartridge. Because the sample is not covalently attached to the support, coupling base and cleavage acid must be delivered as vapors (carried by a stream of argon), which diffuse into the Polybrene film to effect Edman chemistry. Solvent washing of the film to remove excess reagent is very efficient, so that small solvent and reagent volumes contribute to minimal contamination.

The newest version of the gas-phase sequencer is the so-called pulsed liquid sequencer. Microvalving technology, designed in the original machine, is used

to deliver precisely pulses of liquid reagent to wet the sample disk without flowing past it. Once reactions are complete, volatile reagents are removed by evaporation in a stream of argon gas passing through the reaction chamber. The use of pulses of liquid reagents rather than the continuous flow of reagent gases (TFA and triethylamine) significantly increases reaction rates and decreases cycle times.

The newest gas-phase sequencers also incorporate an automated conversion chamber to facilitate the on-line analysis of the PTH amino acids by micro-bore LC. All of these features facilitate rapidity and high sensitivity in microsequencing of peptides and proteins and have drastically decreased the amounts of material needed for structure determination to the extent that direct N-terminal sequence analysis of over 20 residues from a 2-D gel is possible. As a result, this methodology has altered the strategy for protein isolation from the tissue in which it was initially detected.

With the realization that direct sequencing from 1-D and 2-D gels was becoming feasible in the mid-1980s, attention was focused on optimization of direct transfer of protein samples from gels to the sequencer cartridge. This process became known as electroblotting. The two major classes of blotting media used for this application have been either glass-fiber filter (GF/C) or inert polymer-based membranes. Included in the class of glass fiber type media are derivatized GF/C (i.e. aminopropyl, quaternary ammonium, siliconized-glass fiber (SGF)) and non-covalently modified GFC coated with quaternary ammonium polybases (polybase GFC). The most commonly used nonglass, polymer-based membranes are prepared from poly(vinylidene difluoride) (PVDF). Initial success in this direction was achieved using PVDF membranes commercially referred to as Immobilon-P.

For this technique to be effective, however, numerous problems had to be either circumvented or overcome. These included variability in electrotransfer for different proteins, long times for adsorption of proteins from solution onto PVDF membranes, variable recovery due to protein selectivity, and staining and destaining methods that decrease sequencing yields. One successful method circumvents these problems through the use of centrifugation of dissolved samples through a PVDF membrane and subsequent washings to remove contaminants such as trisglycine. Another alternative is the implementation of a modified reaction cartridge within a gas-phase sequencer for direct sequencing on polymeric membranes. This redesigned reaction vessel implements a vertical (rather than horizontal) cross-flow type reactor to accommodate the PVDF membrane.

## Identification of the PTH Amino Acids

Sequence analysis of a peptide or protein occurs in two parts. The first is based on the chemistry of Edman degradation reviewed above. The second is the identification of the resulting PTH amino acids following that degradation. The most popular method for the identification of the PTH amino acids is LC, which was initially developed for the purification of intact peptides and proteins or their fragments. Although thin-layer and gas-liquid chromatography (GLC) have been used for the identification of the PTH amino acids, LC is the technique of choice for many reasons. LC can rapidly (20–30 min) resolve and quantify all 20 PTH-derivatized naturally occurring amino acids; it is easily automated and more sensitive than thin-layer chromatography or GLC; and it is not as labor-intensive, expensive, or equivocal as back-hydrolysis to free amino acids.

A typical LC system consists of hydrophobic liquid-bonded phase column packings (such as a cyanopropyl, phenyl alkyl, C<sub>8</sub>, or C<sub>18</sub>) as the stationary phase. The mobile phase consists of a buffered aqueous solution, such as 0.01 mol l<sup>-1</sup> sodium acetate, pH 5.5, and a miscible organic phase, such as acetonitrile, at a temperature of 40–62°C. The mobile phase is run either isocratically or with an increasing percentage of organic phase (gradient). Various parameters, such as temperature, organic solvent gradient, and pH of the aqueous buffer, can be manipulated so that separation of all 20 PTH amino acids as well as side-products of Edman degradation (diphenylthiourea (DPTU) and dimethylphenylthiourea (DMPTU)) are baseline-resolved in a single analysis. In addition, some PTH amino acids such as PTH histidine and PTH arginine are very sensitive to pH and salt concentration and their elution peaks can easily be placed in gaps between peaks of other PTH amino acids in the elution profile. The first description of reversed-phase gradient elution of the PTH amino acids was that of Zimmerman in 1977, but some two dozen other systems have been described. With modern LC instrumentation employing microbore gradient pumps, low noise, variable-wavelength UV absorbance detectors and specially formulated microbore columns, separation and detection of subpicomole quantities of PTH amino acids has become routine. In addition to these recent improvements, the use of radioactivity has also been used to increase sensitivity in the identification of the PTH amino acids. This involves the use of a radioactive flow detector in series with the UV detector. The PTH amino acids can be labeled using [<sup>14</sup>C]PITC or [<sup>15</sup>S]PITC during the sequencing of unlabeled peptides or proteins, or the protein can be

intrinsically labeled prior to sequencing. Using both labeled and unlabeled PTH norleucine as marker, superposition of the radioactive and UV profiles aligning the norleucine peaks identifies the radioactive PTH amino acid unknown.

## Optimization of Sequencing Data

Ideally, in sequencing one protein, one would expect a single PTH amino acid after each cycle of the Edman degradation. In the real world, this is never the case: each chromatogram contains variable amounts of the PTH amino acids for several reasons. (1) Protein or peptide samples are unlikely to be absolutely pure. (2) Repeated exposure of the protein sample to cleavage acid during sequencing can cause low-level acid-catalyzed cleavage of the peptide chain giving rise to other N-termini that can participate in Edman chemistry. Thus, as the degradation proceeds, the PTH amino acid background increases. (3) While this is occurring, the recovery of the expected PTHs is slowly decreased during sequencing by side-reactions that block the amino terminus or by physical extractive losses of sample from the reaction chamber. (4) Removal of an amino-terminal amino acid at any given cycle of Edman chemistry is incomplete, appearing, therefore, to some variable degree, at the next cycle of degradation. This carry-over, or lag, is cumulative, resulting eventually in the identification of the PTH amino acid being out of phase with the sequence analysis or expected order of release of amino acids during sequencing. (5) The relative recoveries of the PTH amino acids from Edman chemistry vary greatly. Some, such as leucine, are recovered almost quantitatively, while others, such as serine and threonine, are largely destroyed before analysis. (6) LC sampling volumes also contribute to ambiguity in the PTH assignments. These factors, summed together, make correct amino acid assignments ever more difficult as sequence analysis proceeds further into the polypeptide chain.

Because of the large number of variables that are virtually impossible to control manually, LC peak-integration systems have been employed to translate these analog signals into a simpler set of digital numbers. This process, initially introduced in 1971, has evolved into a system of fully computerized data analysis and data reduction systems. Here, the computer first collects and stores digitized LC detector output. Detector baseline drift and high-frequency noise from the chromatograms are removed by application of digital filters, and peaks corresponding to the PTHs are located, quantified, and stored for all cycles in a sequencer run file for each PTH amino acid. A recursive least-squares fit defines the background

levels of each individual PTH amino acid as a function of the cycle. Comparison of signal and background levels on the basis of their standard deviations then permits automated objective assignments of the amino acid sequence. This assignment can be further refined by additional routines that correct for sequence lag from incomplete Edman chemistry and for variability in PTH sampling by the LC system. This combination of baseline, background, and lag correction routines decreases the ambiguity in amino acid identification and allows amino acid sequence assignments with a defined level of statistical confidence. This computerized capability permits significant extension of assignable sequence length over previously unadjusted recoveries.

### **Novel Applications of Automated Protein Sequencing**

Unfortunately for the protein chemist, automated protein sequencing in most cases is no longer the method of choice for determining the primary structure of large proteins. This is because DNA sequencing is more rapid and less laborious in terms of obtaining peptide fragments, each of which has to be sequenced, with sequence overlaps, and because an amino acid sequence of a protein can be inferred from its corresponding gene.

Nevertheless, protein sequencing frequently plays a crucial role in the DNA sequencing process. For example, proteins purified by following the activity associated with them can be sequenced to obtain limited sequence information. These data then can be used to construct oligonucleotide probes to assist in the identification of the corresponding gene from a cloned library. Once identified and isolated, the gene can be amplified and sequenced.

This information, however, defines only the primary translation product. The post-translational processing events crucial to the biological activity of many proteins can only be determined by direct structural analysis of the protein itself, such as determining the site of proteolytic processing and side-chain modification, including glycosylation and phosphorylation. In a similar fashion, site-specific sequencing also defines amino acid side-chains labeled by active site or other chemical modifications.

In addition, the quantitative nature of Edman degradation lends itself to determining the purity of a given preparation and the homogeneity of material in chromatographic column effluents, as described above under manual methods. Quantitative sequencing can also be used to define the stoichiometry of multi-subunit proteins that perform many of the complex enzymatic and structural functions.

Another important application of protein sequencing is in the field of solid-phase peptide synthesis. With the rapid development, success, and popularity of solid-phase peptide synthesis, automated peptide and protein sequencing has found a number of important uses and applications in this burgeoning field. By the very nature of the fact that protein sequencing represents the reverse process of solid-phase peptide synthesis, sequencing offers a unique method for evaluating the products of solid-phase synthesis. Firstly, sequencing can be used to verify the order of amino acids following assembly on to the insoluble support prior to its cleavage from the resin on which it was synthesized. In so doing, sequencing can determine sites of difficulty in synthesizing a given peptide, i.e. sites of deletion of amino acid addition (or the presence of failure sequences). This sequencing application is referred to as 'preview analysis' because the sequence C-terminal to the deleted amino acid appears one residue amino terminal to its expected position in the sequence. Peptide resins from the solid-phase peptide synthesizer can easily and directly be applied to the gas-phase sequencer cartridge so that sequencing of peptide resins is routine in most core-facilities laboratories.

Another very important way in which protein sequencing can be used to characterize products of solid-phase peptide synthesis is in the detection of the presence and removal of amino acid side-chain-protecting groups in synthesis. This is because most side-chain-protecting groups used in *t*-butyloxycarbonyl (but not in 9-fluorenylmethoxycarbonyl) peptide synthesis are stable to reagents in protein sequencing and the PTH side-chain-protected amino acids resolve easily from their unprotected PTH amino acid counterparts. Thus, the sequencing of peptide resins establishes the integrity of side-chain-protecting groups prior to cleavage of the peptide from the resin and side-chain removal during HF treatment. Similarly, direct sequencing of the crude synthetic product following HF cleavage provides an accurate indication of the degree of removal of the side-chain-protecting groups as well as other effects of HF on the crude product. Carrying the solid-phase synthetic process further, sequencing can be used to evaluate the degree of homogeneity of the final synthetic product following chromatographic purification. Thus, it is clear that protein sequencing is an invaluable tool for monitoring virtually each stage of the solid-phase synthetic process.

Recently, sequencing has found a very novel application in epitope selection by screening universal peptide libraries in a process called *in vitro* selection. These libraries are generated by the solid-phase peptide synthesis of peptides prepared in random fashion



of any desired length, but in a very particular way. The quintessence of the method is that the random peptides are prepared in such a manner that only one random peptide resides on one particular bead on which the peptide is constructed. This feat is accomplished by performing 20 separate coupling reactions with each of the 20 naturally occurring amino acids, then mixing or randomizing the amino acid or peptide resin to which the next amino acid has been coupled, dividing the amino acid resins into, again, 20 portions and coupling the next amino acid to the growing peptide chain. In such a way, for example, after just two couplings there is a total of 400 randomly prepared peptides, each bead containing only one peptide, but obviously many beads containing the same peptide. However, after each additional coupling followed by randomization of the peptide resins, the number of beads containing the same peptide rapidly diminishes. In fact, after eight or so couplings, over 25 billion ( $20^8$ ) different peptides exist. Once these libraries have been constructed, they can be screened using a particular antibody to which a fluorescent tag has been covalently bound. The antibody binds and recognizes theoretically only one peptide epitope or amino acid sequence that fits its binding site. Then, simply using a microscope, a fine pair of tweezers and a UV radiation source, this *in vitro* selection process can be completed. The now-fluorescent antibody-peptide-resin complex is removed with the tweezers, the antibody is dissociated from the peptide resin by washing, and the peptide resin can now be sequenced directly in the protein sequencer to identify the primary structure of the unknown synthetic epitope.

This article has summarized the most significant developments in the field of automated protein sequencing since the introduction of the Edman chemistry. It is not unreasonable to expect that the next major advance in protein sequencing by the end of the 1990s and into the twenty-first century will be based on the routine use of mass spectrometry, giving

orders of magnitude increases in both speed and sensitivity of analysis (see Proteins, sequencing by mass spectrometry).

See also: **Liquid Chromatography:** Instrumentation; Amino Acids. **Nucleic Acids:** Electrochemical Methods.

## Further Reading

- Allen G (1981) Sequencing of peptides and proteins. In: Work TS, Burden RM, and Allen G (eds.) *Laboratory Techniques in Biochemistry and Molecular Biology*, pp. 174–180. New York: Elsevier Science Publishers.
- BIOTOL (1992) Analysis of Amino Acids, Proteins and Nucleic Acids. Oxford: Butterworth-Heinemann.
- Edman P (1950) A method for the determination of the amino acid sequence of peptides. *Acta Chemica Scandinavica* 4: 283–293.
- Edman P (1975) Sequencing determination. In: Needleman SB (ed.) *Protein Sequence Determination*, pp. 211–279. Berlin: Springer.
- Edman P and Begg G (1967) A protein sequenator. *European Journal of Biochemistry* 1: 80–90.
- Hunkapiller MW, Hood LE, and Dreyer WJ (1981) A gas liquid solid phase peptide and protein sequenator. *Journal of Biological Chemistry* 256: 7990–7997.
- Laursen RA (1971) Solid phase Edman degradation. An automated peptide sequencer. *European Journal of Biochemistry* 20: 89–102.
- Niall HD (1973) Automated Edman degradation. The protein sequenator. In: Hirs CHW and Timashoff SN (eds.) *Methods in Enzymology*, pp. 179–238. New York: Academic Press.
- Wittmann-Liebold B, Sainkow J, and Erdmann VA (1986) Advanced Methods in Protein Microsequence Analysis. Berlin: Springer.
- Schlesinger DH (1988) High sensitivity microsequencing of peptides and proteins. In: Schlesinger DH (ed.) *Macromolecular Sequencing and Synthesis*, pp. 35–44. New York: Alan R. Liss.
- Waterfield M and Bridgen J (1975) Automated liquid phase sequence analysis. In: Perham RN (ed.) *Instrumentation in Amino Acid Sequence Analysis*, pp. 41–71. London: Academic Press.

## Physiological Samples

**A Bianchi-Bosisio**, Unita Socio Sanitaria, Milan, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 4183–4199, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

## Introduction

The differentiation and determination of proteins in body fluids is fundamental in clinical diagnosis, since qualitative and/or quantitative alteration can be linked to different disease conditions. The subject is

of any desired length, but in a very particular way. The quintessence of the method is that the random peptides are prepared in such a manner that only one random peptide resides on one particular bead on which the peptide is constructed. This feat is accomplished by performing 20 separate coupling reactions with each of the 20 naturally occurring amino acids, then mixing or randomizing the amino acid or peptide resin to which the next amino acid has been coupled, dividing the amino acid resins into, again, 20 portions and coupling the next amino acid to the growing peptide chain. In such a way, for example, after just two couplings there is a total of 400 randomly prepared peptides, each bead containing only one peptide, but obviously many beads containing the same peptide. However, after each additional coupling followed by randomization of the peptide resins, the number of beads containing the same peptide rapidly diminishes. In fact, after eight or so couplings, over 25 billion ( $20^8$ ) different peptides exist. Once these libraries have been constructed, they can be screened using a particular antibody to which a fluorescent tag has been covalently bound. The antibody binds and recognizes theoretically only one peptide epitope or amino acid sequence that fits its binding site. Then, simply using a microscope, a fine pair of tweezers and a UV radiation source, this *in vitro* selection process can be completed. The now-fluorescent antibody-peptide-resin complex is removed with the tweezers, the antibody is dissociated from the peptide resin by washing, and the peptide resin can now be sequenced directly in the protein sequencer to identify the primary structure of the unknown synthetic epitope.

This article has summarized the most significant developments in the field of automated protein sequencing since the introduction of the Edman chemistry. It is not unreasonable to expect that the next major advance in protein sequencing by the end of the 1990s and into the twenty-first century will be based on the routine use of mass spectrometry, giving

orders of magnitude increases in both speed and sensitivity of analysis (see Proteins, sequencing by mass spectrometry).

*See also:* **Liquid Chromatography:** Instrumentation; Amino Acids. **Nucleic Acids:** Electrochemical Methods.

## Further Reading

- Allen G (1981) Sequencing of peptides and proteins. In: Work TS, Burden RM, and Allen G (eds.) *Laboratory Techniques in Biochemistry and Molecular Biology*, pp. 174–180. New York: Elsevier Science Publishers.
- BIOTOL (1992) Analysis of Amino Acids, Proteins and Nucleic Acids. Oxford: Butterworth-Heinemann.
- Edman P (1950) A method for the determination of the amino acid sequence of peptides. *Acta Chemica Scandinavica* 4: 283–293.
- Edman P (1975) Sequencing determination. In: Needleman SB (ed.) *Protein Sequence Determination*, pp. 211–279. Berlin: Springer.
- Edman P and Begg G (1967) A protein sequenator. *European Journal of Biochemistry* 1: 80–90.
- Hunkapiller MW, Hood LE, and Dreyer WJ (1981) A gas liquid solid phase peptide and protein sequenator. *Journal of Biological Chemistry* 256: 7990–7997.
- Laursen RA (1971) Solid phase Edman degradation. An automated peptide sequencer. *European Journal of Biochemistry* 20: 89–102.
- Niall HD (1973) Automated Edman degradation. The protein sequenator. In: Hirs CHW and Timashoff SN (eds.) *Methods in Enzymology*, pp. 179–238. New York: Academic Press.
- Wittmann-Liebold B, Sainkow J, and Erdmann VA (1986) Advanced Methods in Protein Microsequence Analysis. Berlin: Springer.
- Schlesinger DH (1988) High sensitivity microsequencing of peptides and proteins. In: Schlesinger DH (ed.) *Macromolecular Sequencing and Synthesis*, pp. 35–44. New York: Alan R. Liss.
- Waterfield M and Bridgen J (1975) Automated liquid phase sequence analysis. In: Perham RN (ed.) *Instrumentation in Amino Acid Sequence Analysis*, pp. 41–71. London: Academic Press.

## Physiological Samples

**A Bianchi-Bosisio**, Unita Socio Sanitaria, Milan, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 4183–4199, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

## Introduction

The differentiation and determination of proteins in body fluids is fundamental in clinical diagnosis, since qualitative and/or quantitative alteration can be linked to different disease conditions. The subject is

so vast, however, that this article has been limited largely to the analysis of serum proteins, neglecting other body fluids such as urines, cerebrospinal fluids (CSFs), tears, saliva, and sweat. It describes the most common methods of analysis, followed by a table listing (according to electrophoretic migration in agarose) of most of the presently known serum proteins with details of their relative molecular mass, reference mobility, standard concentration intervals, functions and properties, and causes of elevated and decreased levels.

## Analysis for Proteins

The methods of analysis for proteins in body fluids can be roughly grouped as follows:

1. Quantitative measurements of total proteins.
2. Separations by electrophoresis that provide semi-quantitative estimations of the main classes of proteins present in a fairly high concentration.
3. Specific quantitative assay of particular proteins by immunochemical methods using specific antisera and measurement of the antigen-antibody (Ag-Ab) complexes by nephelometry, turbidimetry, radial immunodiffusion (RID) or radioimmunoassay (RIA), or enzyme immunoassay.
4. Detection and identification of abnormal proteins by RID, immunoelectrophoresis (IE), and immunofixation electrophoresis (IFE).

### Methods for Determination of Total Proteins in Serum and Plasma

The overall protein composition of a patient's plasma or serum should be studied first in terms of its total protein content. The total protein concentration of serum in healthy adults is  $60\text{--}80\text{ g l}^{-1}$ . There are two general causes of alterations of serum total proteins: change in the volume of plasma water and change in the concentration of one or more of the specific proteins in the plasma. A decrease in the volume of plasma water (hemoconcentration), as seen in inadequate water intake or excessive water loss, results in relative hyperproteinemia. Hemodilution (increase in plasma water volume), as seen in water intoxication or salt retention syndromes during massive intravenous infusion, is reflected as relative hypoproteinemia. In these cases the concentrations of all the individual plasma proteins are increased or decreased to the same degree. Individual serum proteins may also cause an increase or decrease of total plasma protein concentration: low levels of albumin cause hypoproteinemia (very commonly and with many causes), while a mild hyperproteinemia may be

caused by an increase in the concentration of specific proteins normally present in relatively low concentration (e.g., in acute-phase reactants, polyclonal immunoglobulins, in myeloma, or other malignant paraproteinemia).

When the total protein concentration in serum is measured, two assumptions are arbitrarily made: (1) that all protein molecules are pure polypeptide chains, containing on an average 16% by weight of nitrogen (an assumption that is not true at all), and (2) that each of the several hundred individual proteins present in serum reacts chemically like every other protein (not always true). These simplifying assumptions make the measurement of total protein a practical, though empirical, procedure.

**The Kjeldahl method** This consists of acid digestion to release ammonium ions from nitrogen-containing compounds. The ammonium can then be quantified by conversion to ammonia gas and titration as a base or by Nesslerization in which mercury(II) and potassium iodide are used to form a colored species with ammonia under alkaline conditions. A correction is made for nitrogen contributed by nonprotein compounds also present in serum, and the ammonia nitrogen value is multiplied by  $100/16$  (i.e., a factor 6.25) to express protein nitrogen as total protein. While determination of nitrogen content can be extremely precise, its use for calculation of protein concentration depends on the exact protein composition of the sample, since each protein has somewhat different nitrogen content according to its amino acid composition. The methods are time consuming and not simple, so they are not commonly used in clinical laboratories. However, they remain a means of defining reference standards for the biuret method (described next).

**The biuret method** The biuret method is a colorimetric technique specific for proteins and peptides. Copper salts in alkaline solution form a purple complex with substances containing two or more peptide bonds. The absorbance produced is proportional to the number of peptide bonds that are reacting and therefore to the number of protein molecules present in the reaction system. Thus, the biuret reaction with proteins is suitable for the determination of total protein by spectrophotometry (at  $540\text{--}560\text{ nm}$ ). The method is used extensively in clinical laboratories, particularly in automated analyzers in which protein concentration can be measured down to  $0.1\text{--}0.15\text{ g l}^{-1}$ . The use of bovine or human serum albumin to standardize the biuret method is well established. High-purity albumin contains only amino acids; its nitrogen content is a constant fraction

of its molecular mass and the number of peptide bonds per molecule is known. Since the peptide bond is the biuret-reacting unit in all proteins and the number of peptide bonds determines the absorbance of the colored product, albumin is a reasonable peptide bond standard for all proteins in the mixture.

**Refractometry** Refractometry is a quick and reasonably accurate alternative to chemical analysis for serum total protein when a rapid estimate is required. The refractive index of water at 20°C is 1.330; if solute is added to the water, the refractive index of a dilute solution increases linearly and proportionally to the solute concentration; at higher concentrations of dissolved solids (50–200 g l<sup>-1</sup>), the increase is nearly linear. Temperature affects appreciably the refractive index of a solution, so refractometers for clinical use compensate for temperature effects. Serum contains dissolved solids in concentrations of 80–100 g l<sup>-1</sup>, most of which are proteins. In the refractometry of serum, it is assumed that the concentration of inorganic electrolytes and nonprotein organic compounds does not vary appreciably from serum to serum and that the differences in the refractive index reflect primarily the differences in protein concentrations. The assumption has been shown to be reliable for clear, nonpigmented samples, but hemolysis, lipemia, icterus, and azotemia produce erroneously high results. The method cannot be used for urine protein measurement because of excess solutes in relation to the protein.

**Other methods** Although other methods have been used for total serum protein quantification, none is more convenient or practical than the biuret method. These methods are reviewed here predominantly as background information for the assay of total protein in urine or CSF. They are dependent on four general properties of proteins.

1. Reaction of tyrosine and tryptophan residues with the Folin–Ciocalteu or phenol reagent (tungstophosphoric or molybdophosphoric acid) to give a deep blue color. This property is more advantageous for assaying a pure protein whose composition and relative reactivity are known (e.g., fibrinogen) than it is for a mixture of individual proteins with different concentrations and reactivities. The Lowry method uses the biuret method followed by the phenol reagent, which greatly enhances color formation since the phenol reagent can react with biuret complexes involving all the peptide bonds. The absorbance of the colored species is measured at 650–750 nm. The Lowry method is sensitive to protein at concentrations of 10–60 µg ml<sup>-1</sup>, i.e., it is 100 times more

sensitive than the biuret reaction alone. The Lowry method is widely used in research to quantify tissue proteins and enzyme proteins in purified preparations. In urine and CSF, the method is not satisfactory because the reagents react nonspecifically with nonprotein compounds. Many drugs such as salicylate, chlorpromazine, tetracyclines, and some sulfa drugs also give a positive interference. The method is not used in clinical laboratories.

2. Measurement of the characteristic UV absorbance at 280 nm ( $A_{280}$ ), which depends largely on the tryptophan and also on the tyrosine and phenylalanine content of proteins. For accurate conversion of  $A_{280}$  readings to concentration, the molar absorptivity must be used, since each protein contains different amounts of these three amino acids. However, the  $A_{280}$  value of a mixture of proteins is not an accurate measure of protein content, since molar absorptivities vary greatly between different proteins. Because nucleic acids, which absorb strongly at 260 nm, may be present in protein preparations, total protein can be estimated by the equation

$$\text{Protein concentration (mg ml}^{-1}\text{)} = 1.55A_{280} - 0.76A_{260}$$

so that absorbance can be used for quantifying proteins in the range of 0.05 – 1.5 mg ml<sup>-1</sup>. The method has been used for CSF after removal of small interfering molecules by gel filtration. This approach is sensitive and simple but requires the use of an appropriate spectrophotometer and expensive (silica) cuvettes with high transmittance at 220–280 nm.

3. Precipitation of protein for turbidimetric or nephelometric assays, which can be achieved with sulfosalicylic acid alone, sulfosalicylic acid with sodium sulfate or trichloroacetic acid, or trichloroacetic acid alone. Precipitation methods for total protein assay require that both globulins and albumin are efficiently precipitated with the formation of small, well-dispersed particles of consistent size so that light scattering is a reproducible phenomenon. The particles must be homogeneously suspended in the medium at all times. Standard material must have as nearly as possible the precipitation behavior of normal and abnormal mixtures of plasma and proteins. Specific turbidimetric precipitation methods have been described for total protein determination in urine and CSF.

4. Ability to bind colored dyes, such as Amido Black 10 B, Ponceau S, and Coomassie brilliant blue (CBB). This property is not only utilized to stain protein bands after electrophoresis but can also be used in spectrophotometric methods for total protein determination. There are two major problems: the unequal affinities and binding capacities of individual

proteins for dyes, and the inability to define a consistent material for use as a standard. The CBB method was originally described for CSF and subsequently adapted to urine. The absorbance is measured at 495 and 595 nm (bichromatic mode). The method is simple, fast, and linear up to  $1.5 \text{ g l}^{-1}$  and can be automated (coefficient of variation from run to run  $<5\%$ ). Negative interferences are noted in urine specimens with a high content of sodium chloride or hydrochloric acid used as preservative, while positive interferences are seen with tolbutamide and high concentrations of urea. The color obtained with CBB and albumin is taken as 100%. Hemoglobin and transferrin give a similar color intensity, while globulin and  $\kappa$  and  $\gamma$  chains give only 60%. For the standardization of precipitation and dye-binding methods, a suitable dilution of a normal serum (or normal serum pool) with normal albumin/globulin ratio and quantified for total protein with a reference method (biuret or Kjeldahl method) is used. The intent is to have in the standard a mixture of proteins whose precipitation and dye-binding behavior are similar to that of proteins in the sample to be analyzed. Bovine or human serum albumin as standard prohibits the use of sulfosalicylic acid because these pure proteins give  $\sim 2.5$  times the turbidity given by serum globulins. With these pure proteins, trichloroacetic acid is the reagent of choice.

## Electrophoresis

Modern understanding of protein composition derives from the electrophoretic techniques introduced by Tiselius in 1937 whereby proteins were separated in an electrolyte solution contained within a U-shaped quartz tube through which an electric current was passed. At pH 7.6, four fractions designated albumin, alpha, beta, and gamma were identified and quantified by the change in the refractive index at the boundaries between these bands. Because separation was achieved in a homogeneous solution without a solid support medium, convective forces prevented resolution into more distinct zones. Hence, this technique was termed moving-boundary or frontal electrophoresis. Separation of the protein fractions into discrete bands or zones became possible with the introduction of filter paper as an anticonvection support medium. The term 'zone electrophoresis' was introduced to refer to the migration of charged macromolecules in a porous supporting medium such as cellulose paper, cellulose acetate sheets, agarose gel film, starch gel, and polyacrylamide gel. Zone electrophoresis differs from moving-boundary

electrophoresis in that it generates an electrophoretogram, a display of protein zones, each one sharply separated from neighboring zones on the electrophoretic support material. Solutes of interest in clinical chemistry are mainly macromolecular in size and colloidal in nature, and include proteins in serum, urine, CSF, and other physiological fluids as well as the proteins in erythrocytes and tissue.

The theory of electrophoresis is discussed in detail in another article, but it must be noted here that the rate of protein migration is dependent on such factors as net electric charge of the molecule, size and shape of the molecule, electric field strength, properties of the supporting medium, and the temperature of the operation.

Buffer ions have two purposes in electrophoresis: they carry the applied current and they fix the pH at which electrophoresis is carried out and thus determine the kind of electrical charge on the solute and the extent of ionization of the solute. Through these effects the buffer determines the direction of the electrophoretic migration to one of the two electrodes. The buffer ionic strength determines the thickness of the ionic cloud surrounding the charged molecules and thus also the rate of migration and the sharpness of the zones. Many buffer systems have been used in electrophoretic procedures: the barbital buffer (ionic strength  $0.025\text{--}0.075 \text{ mol l}^{-1}$ ) and the Tris-boric acid-EDTA buffer (ionic strength  $0.03\text{--}0.12 \text{ mol l}^{-1}$ ) are common.

### High-Resolution (High-Performance) Electrophoresis

This involves modifications to conventional electrophoresis that allow resolution of normal serum proteins into as many as 13 bands: a relatively high ionic strength buffer ( $0.075 \text{ mol l}^{-1}$  at pH 8.6) mixed with calcium lactate in an  $8 \text{ g l}^{-1}$  low-electroendosmosis agarose gel is used. When an electrophoretic support medium has a negative charge, the electromotive force tends to move such a charge toward the anode. However, since the solid support medium is fixed and stationary, buffer flows toward the cathode instead. This buffer flow is termed electroosmosis or endosmosis, which also carries the proteins with it to some extent. When the electroosmotic force is greater than the electrophoretic force acting on weakly anionic proteins (e.g.,  $\gamma$ -globulins), these proteins move from the application point toward the cathode although their charge is slightly negative. Endosmosis is minimal in electrophoretic media where surface charges are minimal (starch gel or polyacrylamide gel).

### Paper Electrophoresis

Paper electrophoresis using either Whatman no. 1 or 3 MM paper or their equivalents as support medium was very common in clinical laboratories in the past, but electroendosmosis was prominent and the separation time was long (14–16 h).

### Cellulose Acetate Electrophoresis

This is widely used today. The characteristics of the membrane will vary with the extent of acetylation, the prewashing procedure employed by the manufacturer, and the additives used, as well as the pore size and the thickness of the membrane. Cellulose acetate membranes may be made transparent (cleared) for densitometry by treatment with a solvent mixture that partially dissolves the cellulose acetate fibers and eliminates the original air spaces. Advantages are the speed of separation (20 min–1 h) and the ability to store the transparent membranes for long periods.

### Agarose Gel Electrophoresis

This has been used successfully for analysis of serum proteins, hemoglobin variants, lactate dehydrogenase isoenzymes, lipoprotein fractions, etc. Agarose, which is essentially free of ionizable groups, exhibits little endosmosis. Usually 5–10 g of agarose per liter of buffer provides a gel of desired strength and with good migration properties. The serum sample size used in this kind of electrophoresis is relatively small (0.6–3  $\mu$ l) and the electrophoresis time is short (30–90 min).

### Polyacrylamide Gel Electrophoresis

Polyacrylamide is an inert support whose porosity is easily adjusted by changing the composition of the acrylamide solution prior to polymerization. Although polyacrylamide gel electrophoresis (PAGE) is applicable to standard separations of native proteins, it can also be used for separating proteins according to molecular mass when they are denatured in the presence of sodium dodecyl sulfate (SDS); hence, the term SDS PAGE. This last technique is very powerful for resolving proteins and separating them into multitudinous subunits. It is used in the clinical laboratory for highly individualized investigations (e.g., studies of proteinuria).

### Starch Gel Electrophoresis

This also possesses the property of separating macromolecular ions on the basis of both surface charge and molecular size. The starch gel is partially hydrolyzed since native starch does not gel. Proper

preparation of the gel requires considerable skill and is relatively difficult. The concentration of starch gel is generally 100–160 g l<sup>-1</sup> and the pH of the buffer is between 8.6 and 9.0.

### Isoelectric Focusing

Isoelectric focusing (IEF) affords superior resolution of closely migrating proteins or various forms of a single protein that differ in charge owing to minor modifications. With this technique proteins migrate through a gel containing a pH gradient established with a mixture of carrier ampholytes (amphoteric oligoamino–oligocarboxylic acids). As each protein reaches the gel location where the pH is equal to its pI (pH at which the protein carries zero net charge), it comes to rest. Thus, the final pattern strictly accords with the pI value of the proteins.

### Immobilized pH Gradient Electrophoresis

This is a modification of IEF in which the pH gradient is not created by the electric field but is preconstituted by covalently grafting on to a polyacrylamide matrix a series of buffering and titrant acrylamide derivatives (known commercially as immobilines).

### Emerging Techniques

While not yet active in clinical analysis, two techniques appear to be of great interest: two-dimensional (2D) maps and capillary zone electrophoresis (CZE). In 2D maps, denatured polypeptide chains are separated in two dimensions, first by surface charge (IEF) and then orthogonally by size (SDS electrophoresis). In normal human sera, for example, 600 spots are separated and detected by silver staining.

In CZE, proteins are separated in 50–100  $\mu$ m i.d. capillaries with very high resolution, and direct on-line quantification is often achieved by exploiting the absorbance of the amide bond at 210 nm.

### Staining of Protein Fractions after Electrophoresis

After electrophoretic separation, protein fractions are visualized by staining with a suitable dye. Ideally, a dye should have equal affinity for all proteins and sufficient sensitivity to detect each protein. However, the most commonly used dye for cellulose acetate separation, Ponceau S, has 1.5 times greater affinity for albumin than for  $\gamma$ -globulin. Other widely used dyes that behave in a similar manner are Amido black 10 B (Amidoschwarz 10 B), Naphthol blue black or Buffalo black, a stain commonly employed for agarose separations, and CBB, a dye widely used for agarose and polyacrylamide gel separations. In



terms of sensitivity, Ponceau S on cellulose acetate or Amido black on agarose can detect proteins whose concentrations in the sample exceed 300–500 mg l<sup>-1</sup>. CBB is slightly more sensitive and is suitable for detection of an oligoclonal band in concentrated CSF. Silver staining binds to specific amino acid groups in proteins in a manner independent of the net protein charge and can routinely detect proteins at concentrations as low as 100 mg l<sup>-1</sup>, a sensitivity some 20–50 times greater than that of CBB. Silver staining is therefore an attractive choice for CSF and any other fluids when both a low protein concentration and a small volume are available. Its disadvantage is its greater complexity. In addition, the technique is less suited for serum and other body fluids. To quantify each electrophoretic protein band, the plate may be scanned with a densitometer.

### **Immunological Methods**

The strength or energy of interaction between an antibody and an antigen is described by two terms. 'Affinity' refers to the thermodynamic quantity defining the energy of interaction of a single antibody-combining site and its corresponding epitope on the antigen. 'Avidity' refers to the overall strength of binding of antibody and antigen and includes the sum of the binding affinities of all the individual sites on the antibody. For example, immunoglobulin G (IgG) has two affinity-binding sites, whereas immunoglobulin M (IgM) has 10 affinity-binding sites per antibody molecule. Thus, affinity is a property of the substance bound (antigen) and avidity is a property of the binder (antibody). For polyclonal antibodies, avidity is difficult to determine, primarily owing to the diversity of the antibody population.

Several forces act cooperatively to produce antigen–antibody binding. The three major contributing forces are van der Waals–London dipole–dipole interaction, hydrophobic interaction, and ionic Coulombic bonding. The binding of antigen to antibody is not static but is an equilibrium reaction that proceeds in three phases. The initial reaction (phase 1) of a multivalent antigen (Ag<sub>n</sub>) and a bivalent antibody (Ab) occurs very rapidly in comparison with the subsequent growth of the complex (phase 2). The third phase of the reaction involves the precipitation of the complex after a critical size is reached. The speed of these reactions depends on factors such as electrolyte concentration, pH, and temperature, as well as on antigen and antibody types, and the binding affinity of the antibody. The concentration of sodium chloride is important: in most cases 0.15 mol l<sup>-1</sup> is used. It is best to use dilute solutions for determining the influence of such factors as ionic

species, ionic strength, pH, and concentration of soluble linear polymers, and for optical analytical methods. The use of dilute solutions slows the growth of the antigen–antibody complexes and a more stable and more homogeneous population of complexes results. If the antigen or antibody of interest is bound to a solid phase such as a cell membrane or a synthetic particle (polystyrene or cellulose), the protein will exist in a microenvironment that is different from that of a protein free in solution. The water surrounding the protein is more highly ordered near the surface of the solid phase and a condition results that is more favorable for dispersion and dipole–dipole interactions and Coulombic bonding. This situation favors the formation of low-avidity as well as high-avidity antigen–antibody complexes and hence can provide better sensitivity for analytical applications. Some studies have shown radiometric solid-phase assays to be more sensitive than their counterpart solution assays. The difference in sensitivity is most likely a result of additional low-avidity antibody binding during the initial incubation in the solid-phase system, which does not occur as readily in the liquid phase.

### **Qualitative Immunological Methods for Detecting Proteins**

**Passive gel diffusion** Many qualitative and quantitative immunochemical methods are carried out in a semisolid medium such as agar or agarose. The primary advantage in using a gelatinous medium is stabilization of the diffusion process with regard to mixing caused by vibration or convection and to allow visualization of precipitin bands for qualitative and quantitative evaluation of the reaction. Two basic approaches to passive diffusion are in common use today. In 'single' immunodiffusion a concentration gradient is established for only a single reactant. It usually depends on diffusion of an antigen into agar impregnated with antibody. A quantitative technique based on this principle is RID. 'Double' immunodiffusion in two dimensions (Ouchterlony technique) is a widely used technique in immunology. In this technique a concentration gradient is established for both reactants (antigen and antibody); it allows direct comparison of two or more test materials and provides a simple and direct method for determining whether the antigens in the test specimens are identical, cross-reactive, or nonidentical. Note that a negative reaction does not necessarily imply absence of antibody or antigen. A negative reaction can result from using amounts of material below the detection limit of the method; alternatively, the antibody may be nonprecipitating.

**Immunoelectrophoresis (IE)** This couples electrophoretic separation with the 2D immunodiffusion reaction and is used for specific identification and semiquantitative estimation of a wide range of antigens. First, the electrophoresis of a protein mixture is carried out in agar gel on a plastic or glass support, and then a trough is prepared along one margin of the slide and a specific antiserum is applied in the trough. Diffusion is then allowed to proceed overnight. In the clinical laboratory this procedure has primarily been applied to the evaluation of human myeloma proteins. However, the method has slowly been replaced by IFE.

**Crossed immunoelectrophoresis** Also known as 2D IE, this is a variation of IE in which electrophoresis is also used in the second dimension to drive the antigen into a gel containing antibodies specific for the antigens of interest. This technique is more sensitive and produces higher resolution than is possible with IE. As with IE, the specimen is applied to a cylindrical well cut in an agar medium and is then subjected to electrophoresis in the first dimension, followed by electrophoresis at 90° to the first run through the gel, which contains antibodies. Precipitation occurs along the lateral margins as the antigen advances and results in precipitation 'peaks' for each antigen. Quantities of individual antigens can be estimated by measuring the peak area of the antigen under the precipitation arc and comparing it to the peak area of a standard antigen preparation. This technique has its greatest utility in evaluation of protein mixtures in the study of protein alterations subsequent to activation or interaction with other molecules. The method was developed by Laurell and has often been referred to as the 'rocket' technique. This method closely parallels that of RID. It offers the advantage of more rapid completion of the assay owing to the enhanced migration achieved by the use of electrophoresis. One important limitation of the 'rocket' technique is that the relative net charge of macromolecules at the pH used in the test must be estimated accurately, because it will determine the direction of migration of the molecules to either the anode or cathode.

**Counter immunoelectrophoresis (CIE)** Here two parallel lines of wells are punched in the agar. One row is filled with antigen solution and the opposing row is filled with antibody solution. If the solutions were allowed to diffuse passively over the next 18–24 h, a precipitation line would form between the opposing wells where antigen reacted with specific antibody. In CIE, this process is made to occur more rapidly by applying a voltage across the gel so that the antigen and antibody move in opposite directions

as a result of electroendosmosis and a precipitation line is formed where they meet. Qualitative information (i.e., identification of antigen) is provided within 1–2 h.

**Immunofixation electrophoresis** This is a method of immunochemical identification of proteins following electrophoretic separation. It is gradually replacing IE because of its rapidity and ease of interpretation. Several procedures and commercial kits are available for IFE and IE. While these procedures may differ in detail, their principles are the same. IFE is used mainly for estimation of paraproteins. The first step consists of an estimation of the paraprotein from the spike identified on the electrophoresis plate by densitometric scan. The second step is to identify heavy and light chains by IE or more simply by immunofixation. Specific antisera against  $\gamma$ ,  $\alpha$  and  $\mu$  heavy chains and against  $\kappa$  and  $\lambda$  light chains are used; antisera against immunoglobulins IgD and IgE are seldom required. Reaction of the paraprotein with only one heavy-chain and one light-chain antiserum confirms that the anomalous spike on protein electrophoresis is indeed a paraprotein and is not due to denaturation, fibrinogen, hemoglobin, or some other artifact. Samples of the patients' specimens are placed in six wells on an agarose gel and their major protein groups are separated by electrophoresis. One of these tracks is then treated with a chemical fixative solution to fix all proteins in the agarose and create an electrophoresis reference pattern for the specimen. The other five tracks are treated with specific heavy-chain and light-chain antisera that react with individual immunoglobulins in the specimens, causing them to become immunofixed in the agarose. All unreacted proteins in these five tracks are then washed out of the gel and all six tracks are stained to visualize the fixed protein bands. By comparing the locations of the stained immunofixed bands with a band at the same location in the reference pattern, a specific protein can be identified. In monoclonal gammopathies, the IFE patterns yield a distinct, sharply defined precipitation band with one heavy chain and one light-chain antiserum. The bands match the location of the particular immunoglobulin in the reference pattern. In polyclonal gammopathies, a diffuse precipitation band will occur with the specific antiserum in contrast to the sharp band observed in monoclonal gammopathies. The method can be used for other protein identification in a similar way.

### Quantitative Methods for Measuring Proteins

**Radial immunodiffusion and electroimmunoassay (EI)** Typically, the two most commonly encountered

gel-based methods for quantitative immunochemical studies are RID and EI (the 'rocket' technique): RID is a passive diffusion method in which a concentration gradient is established for a single reactant, usually the antigen. The antibody is uniformly dispersed in the gel matrix. Antigen is allowed to diffuse from a well into the gel until the antibody is in excess and immune precipitation occurs; the antigen-antibody interaction is manifested by a well-defined ring of precipitation around the antigen well. The ring diameter will continue to increase until equilibrium is reached. Standards are run at the same time as the sample, and a standard curve of ring area or diameter versus concentration is plotted. The original procedure described by Mancini requires the establishment of equilibrium before measurement of the precipitin ring diameter. Under equilibrium conditions, there is a linear relation between antigen concentration and the square of the ring diameter.

In EI, as in RID, a single concentration gradient is established for the antigen, but in this case an applied voltage is used to drive the antigen from the application well into a homogeneous suspension of antibody in the gel. This produces, as opposed to RID, a unidirectional migration of antigen and results in increased sensitivity. The height of the resulting rocket-shaped precipitation line is proportional to the antigen concentration. Quantification is effected by using standards on the same plate.

In general, gel-based methods for quantification of antigens require excellent technical skill, as well as tedious preparation and execution, to produce good results. In many clinical laboratories these methods are restricted to qualitative studies or are used as reference methods. Quantitative data are more commonly obtained by turbidimetric and nephelometric methods, RIAs, enzyme immunoassays, and fluorimetric immunoassays.

**Turbidimetric and nephelometric assay** Nephelometry and turbidimetry, because of their speed and ease of use, are most widely used. These techniques can be used either by measuring the amount of Ag-Ab complex formation (endpoint methods) or by measuring the rate of complex formation (kinetic methods). The kinetic methods are more rapid because measurements are accomplished within 20 s, and are more precise because sample blanks are not necessary. Kinetic assays are, however, somewhat less sensitive because low-affinity antibodies do not have time to react. The occurrence of Ag-Ab formation is related to the amount of light scattering and is used as the basis for antigen quantification. This approach has been accompanied by the development of

sophisticated instruments specifically designed rapidly to measure scattered light, i.e., nephelometers. Turbidimetry, in contrast, measures the decrease in intensity of transmitted light. Nephelometry has proved useful because of its sensitivity and specificity for rapidly detecting immune complex formation. Filtered light of a certain wavelength enters the analytical cell containing a suspension of reactant material and is scattered by the immune complexes. A photomultiplier tube located at an angle of 30–90° to incident light measures the intensity of the scattered light. The sensitivity can be enhanced by using an intense light source as used in a spectrofluorimeter. Also, use of enhancing reagents such as poly(ethylene glycol) has improved the assay by increasing the speed and sensitivity of Ag-Ab formation so that measurements can be made within seconds to minutes after antigen-antibody mixing. Within the last few years, improved instruments utilizing laser and other light sources along with microcomputerized data handling to determine reaction rates for precipitate formation have appeared. Such devices offer the promise of even more rapid and sensitive assays. One limitation of this technique is that the antisera and test specimens must have low levels of intrinsic light-scattering activity. The antiserum must have high affinity and monospecificity for the assay antigen. An antiserum exhibiting these properties is referred to as 'nephelometric grade'. Those nephelometric systems that employ 'steady-state' conditions require that optimum antigen-antibody ratios be established for use in the assays. This is necessary in order to determine the dilutions of antigen and antibody that are sufficient to produce light scattering by immune complex formation but avoid rapid development of large particles that cause an uneven distortion of the light-scattering response. Once the dilutions have been determined, the antiserum is mixed to the diluted test specimen either individually, in a continuous flow system, or under conditions of constant mixing as in a rotary chemical analyzer. Once 'steady-state' conditions have developed, light-scattering measurements are made and the peak height response from the photomultiplier tube is recorded. Standards that contain known concentrations of the antigen are measured in order to obtain a calibration graph.

The methods discussed so far rely on examining immune complex formation as an index of an antigen-antibody reaction. For measuring the initial phase, i.e., the primary reaction  $\text{Ag} + \text{Ab}$ , different physical and chemical techniques are required. Many of the methods have provided important information on the mechanism of the antigen-antibody reaction but are not applicable for use in a clinical laboratory.

The following section discusses selected methods appropriate for clinical laboratory use.

**Radioimmunoassay and immunoradiometric assay (IRMA)** RIA and IRMA methods are capable of measuring the primary reaction between hapten (hapten is an incomplete antigen that cannot elicit an immune response by itself but can react with antibody of block a specific antigen-antibody complex) or antigen and a single antibody. In RIA the hapten or antigen is labeled with a radioactive isotope, whereas in IRMA the antibody is the labeled species. RIA and IRMA are concerned only with the primary combination of antigen and antibody. The two standard procedures for RIA are termed competitive and sequential. In a competitive RIA, all reactants are mixed together simultaneously. Labeled antigen (Ag) and unlabeled antigen (Ag) compete for binding to the antibody. In such a system, the avidity of the antibody for both the labeled and unlabeled Ag compete for binding to the antibody. In such a system, the avidity of the antibody for both the labeled and unlabeled antigens must be the same. Under these conditions, the probability of the antibody binding the labeled antigen is inversely proportional to the concentration of unlabeled antigen; hence, bound counts are inversely proportional to unlabeled antigen concentration. For some antisera a competitive RIA is not as sensitive as a sequential RIA, especially in cases where the antiserum has different avidities for the labeled and unlabeled antigens. In the sequential approach, unlabeled antigen is first mixed with excess of antibody and binding is allowed to achieve equilibrium. Labeled antigen is then added and allowed to equilibrate, and, after separation, the bound and free counts are determined. With this approach, a higher fraction of the unlabeled antigen can be bound by the antibody than in a competitive assay, especially at a low antigen concentration. Sequential assays can provide a two- to fourfold increase in sensitivity compared to a competitive assay.

Labeled antibody assays (IRMA) have the advantage of not requiring a quantity of purified antigen because the antigen need not be labeled. This also obviates potential problems that may arise from iodination of labile antigens. Antibodies are more stable proteins and are less difficult to label without damaging the protein's function. An IRMA often requires additional steps for addition of reactants or washing. Typically, a 'sandwich' or two-site IRMA method is used. In these systems, antibody is first attached to a solid phase by passive absorption or by using reactions that result in covalent binding of the antibody to the solid phase. IRMA is a common

approach for protein antigen measurement, owing to its simplicity and ease of application in the routine laboratory. Indirect IRMAs have also been described.

**Nonisotopic immunoassays** Use of nonisotopic immunoassays has, in the last few years, become increasingly common in both research and clinical laboratories. Although these immunoassays do not employ radioactive labels, they may have sensitivities similar to those of classic RIA. The immunochemical mechanism and principles are the same as for RIA; only the label and the method of measurement are different. The most common labels are enzymes and fluorescent conjugates, which may be covalently bound to antibodies or haptens.

Enzyme multiplied immunoassay technique (EMIT) is one example of a nonisotopic immunoassay. The EMIT system is considered a homogeneous system. A heterogeneous enzyme immunoassay technique is that of enzyme-linked immunosorbent assay. One of the reaction components is nonspecifically adsorbed to a solid phase, such as a microtiter tray or a plastic bead. One of the major goals in the design of all nonisotopic immunoassays has been to provide a homogeneous system in which no separation of bound and free antibody or antigen is necessary and, therefore, to make these assays technically easier and more rapid. The choice of an enzyme label in any nonisotopic immunoassay depends primarily on the application. Alkaline phosphatase,  $\beta$ -glucuronidase, glucose-6-phosphate dehydrogenase, and horseradish peroxidase are commonly used in both commercial and research systems.  $\beta$ -Glucuronidase and alkaline phosphatase, among other enzymes, can be used with either a colorimetric or fluorimetric substrate. A biotin-avidin system uses a biotin-labeled first antibody. Biotin can be attached to the antibody in relatively high proportion without loss of immunoreactivity by the antibody. When an avidin-conjugated label is added, a complex of Ag-Ab-biotin:avidin-label is formed. Further amplification can be achieved by a biotin:avidin:biotin linkage.

Fluorescent labeled antibody methods are being used to quantify antigen and antibodies similarly to RIA methods with comparable levels of sensitivity. The procedures require extensive purification and characterization of the fluorescent reagents, as well as special instrumentation (time-resolved fluorimeter) to decrease nonspecific background emissions while increasing specificity and sensitivity. The fluorescent antibody methods have been used for some time and have the advantages of greater availability of well-characterized standardized reagents and procedures.

**Table 1** Principal plasma proteins and properties

<i>Protein</i>	<i>Relative molecular mass</i>	<i>Mobility<sup>a</sup> (agarose pH 8.6)</i>	<i>Reference intervals (g l<sup>-1</sup>) [IU ml<sup>-1</sup>]</i>	<i>Function, properties, genetic variants (GVs)</i>	<i>Elevated in</i>	<i>Decreased in</i>
Prealbumin (thyroxine binding prealbumin) PA (TBPA)	54 980 (tetramer)	Prealbumin	0.25–0.45	Binds and transports T3, T4, RBP, vitamin A GV: molecular variants associated with familial amyloidosis	Glucocorticoids Hodgkin lymphoma Alcohol	Negative acute-phase reactant Hepatocellular disease Malnutrition Nephrotic syndrome Childhood Estrogens, pregnancy
Retinol binding protein RBP	20 960	Prealbumin d <sub>2</sub> (free)	0.03–0.06	Binds and transports retinol (vitamin A) Forms 1:1 complex with PA	Glucocorticoids Oral contraception Renal insufficiencies	Negative acute-phase reactant Hepatocellular disease Nephrotic syndrome Hyperthyroidism Inherited deficiency of vitamin A
Albumin ALB	66 460	Albumin	37–53 [93–133]	Actions: oncotic pressure, amino acid reservoir, carries small molecules GV: 20 genetic variants (bisalbuminemia); analbuminemia (congenital absence of albumin)	Dehydration Intravenous infusion	Negative acute-phase reactant Chronic inflammatory or neoplastic diseases Protein-losing syndromes and malnutrition Altered distribution in an extravascular compartment (ascites) Congenital absence
$\alpha$ -Fetoprotein AFP	66 300	$\alpha$ 1	<10 $\mu$ g l <sup>-1</sup> [<7]	Early in embryonic life this protein functions as the major serum protein, as albumin later Produced by the fetal yolk sac and then by the fetal liver	Neonatal period, pregnancy Hepatocellular carcinoma, hepatoblastoma Testicular or ovarian chorion carcinoma Carcinoma of the pancreas, gut, and lung Cirrhosis of the liver Tyrosinemia Ataxia, telangiectasia Congenital hypothyroidism In maternal serum: neuronal tube defects, anencephaly of the fetus	In maternal serum: intrauterine death, trisomy 21

$\alpha$ -1 Acid glycoprotein (orosomicold)	41 000	$\alpha$ 1	M: 0.50–1.30 F: 0.40–1.20	Has low pI 2.7–3.5 and 45% is carbohydrate Inactivation of progesterone	Acute-phase reactant Glucocorticoids (endogen/esogen)	Estrogen/pregnancy Malnutrition Severe hepatic damage
AAG				Binds lidocaine, propranolol, drugs Role in immune system Modifies platelet aggregation Polymorphism with no clinical significance	Rheumatoid arthritis Systemic lupus erythematosus Malignant neoplasm	Severe protein-losing gastroenteropathies
$\alpha$ 1-Antitrypsin AAT $\alpha$ 1 protease inhibitor $\alpha$ 1 PI	54 000	$\alpha$ 1	1.4–3.2 [59–134]	Antiprotease activity Acts against trypsin, kallikrein, urokinase, plasmin, etc., and especially against neutrophilic elastase, and collagenase GV: Significant microheterogeneity: 75 allotypes described Genotypes PI MM (produces M protein), PI ZZ, PI MZ, PI SZ, and PI MS Ability to inhibit protease is directly related to the circulating level of M protein	Estrogens/pregnancy Acute-phase reactant Hepatocellular disease	AAT congenital deficiency (neonatal respiratory distress syndrome) Pulmonary emphysema Neonatal hepatitis Juvenile cirrhosis Nephrotic syndrome Acute pancreatitis Severe hepatitis
Serum amyloid P-protein  (9.5 S $\alpha$ <sub>1</sub> glycoprotein, $\alpha$ 1-macroglobulin) SAP	233 000–308 000	$\alpha$ 1	0.035–0.075	Noted in most amyloid deposits; similar to lecithin High-affinity for polyanions, cations, and fibronectin Has pentagonal ultrastructure similar to C-reactive protein (CRP)		Hepatocellular diseases
$\alpha$ 1-Microglobulin $\alpha$ 1-M	31 000	$\alpha$ 1	0.02–0.05	Immunosuppression Binds lipophilic molecules (it is similar to retinol-binding protein and $\alpha$ 1-acid glycoprotein)	Renal insufficiency In urine: tubular proteinuria	

*Continued*



**Table 1** Continued

<i>Protein</i>	<i>Relative molecular mass</i>	<i>Mobility<sup>a</sup> (agarose pH 8.6)</i>	<i>Reference intervals (g l<sup>-1</sup>) [IU ml<sup>-1</sup>]</i>	<i>Function, properties, genetic variants (GVs)</i>	<i>Elevated in</i>	<i>Decreased in</i>
$\alpha$ 1-Thiol proteinase inhibitor $\alpha$ -TPI	167 000–90 000	$\alpha$ 1 $\alpha$ 2	0.3–0.5	Inhibitor for thiol protease (cathepsins B1, C, H, L, and activated proteases by Ca <sup>2+</sup> )		Childhood
Transcobalamin I TCI	65 000	$\alpha$ 1	Traces	Binds vitamin B <sub>12</sub>	Hepatoma Polycythemia Myeloid leukemia	
$\alpha$ 1-Antichymotrypsin ( $\alpha$ 1-X-glycoprotein) $\alpha$ 1-X; AAC	68 000	$\alpha$ 1 slow Inter $\alpha$	0.3–0.6	Inhibits cathepsin G, mast cell kinase	Acute-phase reactant	Inherited deficit Childhood Nephrotic syndrome Asthma
Thyroxine-binding globulin TBG	54 000	$\alpha$ 1 Inter $\alpha$	0.01–0.03	Binds and transports T3 and T4 GV: qualitative and quantitative (X-chromosome linked)	Estrogens/pregnancy Inheritance Hepatitis B Hypothyroidism	Negative acute-phase reactant Glucocorticoids Anabolic steroids Hyperthyroidism Malnutrition Nephrotic syndrome
Transcortin Corticoid binding-globulin CBG	55 700	Inter $\alpha$	0.015–0.02	Binds and transports cortisol	Estrogens/pregnancy	Anabolic steroids Hypofunctional ovaries
Vitamin D-binding protein (GC-globulin) VDEP	50 800	Inter $\alpha$	0.2–0.55	Binds and transports vitamin D <sub>3</sub> Binds and removes actin GV: CG (more than 50 alleles)	Estrogens/pregnancy	Hepatopathies Protein-losing syndromes Childhood
Ceruloplasmin CP, CER	132 000	Inter $\alpha$	0.20–0.55 [64–176]	Copper-containing protein Oxidase activity toward many polyamine and polyphenol substrates, Fe(II) to Fe(III) Prevents and removes free-radical production in inflammatory diseases Genetic polymorphism	Acute-phase reactant Disease of reticuloendothelial system Hodgkin disease Infection or obstruction of the biliary tract Estrogens/pregnancy	Wilson disease Copper deficit Malnutrition Nephrosis Severe liver disease (particularly biliary cirrhosis) Childhood

Inter $\alpha$ trypsin inhibitor $\alpha_1$ -TI	160 000	Inter $\alpha$	0.2–0.7	Trypsin, acrosin, and secretory protease inhibitor Breaks elastase in two fragments: the smaller fragment is a high-affinity inhibitor for many serum proteases Precursor of LMW inhibitor in secretions and tissues	Inflammatory disorders Tumors Tumors	
Antithrombin III (heparinic cofactor) AT III	58 000	Inter $\alpha$	0.22–0.39 (plasma)	Thrombin, factor Xa, XII, XI, IX, kallikrein inhibitor Heparin accelerates the activity of AT III GV: qualitative and quantitative	Acute-phase reactant Anabolic steroids	Nephrotic syndrome Inherited deficit (hypercoagulability) Oral contraceptives Disseminated intravascular coagulation Venous thrombosis/pulmonary embolism
$\alpha_2$ -Macro-globulin $\alpha_2$ -M	72 500	$\alpha_2$	M:1.2–2.7 [50–113] F:1.4–3.2 [158–134]	One of the largest plasma proteins Inhibits proteases, coagulation factors, fibrinolysis, and complement Role in immunological and inflammatory processes Can bind and transport $\text{Zn}^{2+}$ Four subunits; dimeric structure GV: three polymorphic forms: system Xm	Nephrotic syndrome Hepatic synthesis Liver disease Diabetes mellitus Estrogens/pregnancy Childhood Neural tube defects Ataxia telangiectasia Down's syndrome	Fibrinolysis Disseminated intravascular coagulation Acute pancreatitis Stress Severe peptic ulcer Extracorporeal circulation
Haptoglobin Hp		$\alpha_2$	0.5–3.2	Binds free hemoglobin Prevents loss of iron Tetramer ( $2\alpha$ , $2\beta$ , chains); Hp2 forms polymers GV: qualitative and quantitative	Acute-phase reactant Glucocorticoids Hodgkin's lymphoma Bile duct obstruction Nephrotic syndrome (Hp 2-1 or 2-2) Ulcerative colitis	Intravascular hemolysis Ineffective erythropoiesis Estrogens/pregnancy Hepatocellular disease Nephrotic syndrome (Hp 1-1) Childhood
Hp 1-1	86 018		1.0–2.3			
Hp2-1	88 018 plus polymers		0.9–3.2			
Hp2-2	>20 0000 (polymers)		0.5–3.2			

*Continued*

**Table 1** Continued

<i>Protein</i>	<i>Relative molecular mass</i>	<i>Mobility<sup>a</sup> (agarose pH 8.6)</i>	<i>Reference intervals (g l<sup>-1</sup>) [IU ml<sup>-1</sup>]</i>	<i>Function, properties, genetic variants (GVs)</i>	<i>Elevated in</i>	<i>Decreased in</i>
2-HS glycoprotein 2-HS	50 000	$\alpha 2$	0.40–0.85	Component of bone and dentin matrix Affinity for Ca <sup>2+</sup> and Zn <sup>2+</sup> Opsonic property Inhibits cell-mediated immunity		Negative acute-phase reactant
C1 esterase-inhibitor C1 1NH ( $\alpha 1$ -neuro- amino- glycoprotein)	104 000	$\alpha 2$	0.15–0.35	Inhibits activated complement components C1r, C1s, coagulation factors and fibrinolysis, plasmin, thrombin, and kallikrein GV: quantitative and qualitative variants (some hypofunctional)	Acute-phase reactant	Inherited deficit (inherited angioedema)
C1 INA						
$\alpha 2$ -Antiplasmin $\alpha 2$ AP	65 000	$\alpha 2$	0.04–0.08	Inhibits plasmin (fibrinolysis) and kallikrein	Acute-phase reactant	Inherited deficit (hyperfibrinolysis)
Fibronectin (cold insoluble globulin) FN	440 000	$\alpha 2$ – $\beta 1$	0.25–0.4	Makes possible attachment to cell membranes Binds to fibrin (clots) and C1q (immunocomplexes) playing a role in opsonization and fibroblast adhesion	Chronic hepatitis	Negative acute-phase reactant Sepsis, shock Disseminated intravascular coagulation Acute leukemia Acute pancreatitis Inherited deficit
Transferrin (siderophilin) TRF	79 550	$\beta 1$	2.3–4.3 [80–150]	Principal plasma protein for transport of iron: 1 mg = 1.25 g Fe <sup>3+</sup> total iron binding capacity (TIBC) Unsaturated TRF is important during infections and parasitic diseases GV: more than 20 variants Polymorphism has no clinical significance except as the rare congenital defect attransferrinemia	Iron deficiency Estrogens/pregnancy	Negative acute-phase reactant Chronic inflammation Malnutrition Chronic liver diseases Protein-losing syndromes Iron overload Congenital defect attransferrinemia (severe hypochromic anemia)

Sex hormone-binding globulin ( $\beta$ steroid-binding globulin) SHBG	65 000	$\beta$ 1	M:0.001–0.012 F:0.003–0.015	Binds and transports testosterone, estriol, etc.	Estrogens/pregnancy	Anabolic steroids Thyroid hormones Growth hormones
Transcobalamin II TC II	53 900	$\beta$ 1	$> 25 \text{ ng l}^{-1}$	Binds and transports vitamin B <sub>12</sub> GV: some hypofunctional variants	Myeloproliferative disorders	Inherited deficit Childhood
Hemopexin HPX	60 000	$\beta$ 1	0.5–1.15	Binds free heme, prevents loss of iron	Estrogens/pregnancy	Severe <i>in vivo</i> hemolysis Nephrotic syndrome Childhood
$\beta$ 1 Pregnancy-associated glycoprotein SPI	90 000 (42 300)N	$\beta$ 1	$< 1 \times 10^{-6}$	Binds and transports steroid hormones Synthesized in the syncytiotrophoblast cells Concentration of SPI correlates placental weight	Pregnancy Trophoblastic diseases Malignant teratomas Twin pregnancy	Fetal distress Spontaneous abortion Placental insufficiency
C3 complement component ( $\beta$ 1-C-globulin) C3	180 000	$\beta$ 2	0.50–0.90 [0.64–1.15]	Glycoprotein Visible as a band on electrophoresis only if serum is fresh Polymer contains two subunits Classical and alternative pathways of complement activity Genetic polymorphism without clinical abnormalities	Subacute-phase reactant Biliary obstruction Immune complex diseases	Rheumatoid vasculitis Glomerulonephritis Gram-negative bacteremic shock Gram-positive bacteremia Disseminated cytomegalovirus infection Subacute bacterial endocarditis Inherited deficit (recurrent infections/childhood)
C4 complement-component ( $\beta$ 1-E globulin) C4	206 000	$\beta$ 2	0.13–0.37	Glycoprotein Not visible on cellulose acetate or agarose electrophoresis Polymer contains three subunits Classical pathway of complement activity Genetic polymorphism without clinical abnormalities	Subacute-phase reactant Estrogens/pregnancy Immune complex diseases	Systemic lupus erythematosus Subacute bacterial endocarditis Glomerulonephritis Gram-positive bacteremia Inherited angioedema Inherited deficit

Continued

**Table 1** Continued

<i>Protein</i>	<i>Relative molecular mass</i>	<i>Mobility<sup>a</sup> (agarose pH 8.6)</i>	<i>Reference intervals (g l<sup>-1</sup>) [IU ml<sup>-1</sup>]</i>	<i>Function, properties, genetic variants (GVs)</i>	<i>Elevated in</i>	<i>Decreased in</i>
Fibrinogen FI	340 000	$\beta 2$	2.0–4.0 (plasma)	Precursor of fibrin; it is the major constituent of blood clot GV: some qualitative and quantitative variants (hypofunctional)	Acute-phase reactant Estrogens/pregnancy Oral contraception	Disseminated intravascular coagulation Hepatocellular diseases Anabolic steroids Inherited deficit Childhood
$\beta 2$ -Microglobulin $\beta 2$ -M	11 818	$\beta 2$	0.001–0.003	Light or $\beta$ chain of human leukocyte antigen (HLA) on cell surfaces Single polypeptide chain, does not contain carbohydrate	Renal failure Inflammation Neoplasm, especially those associated with B-lymphocytes Fetal/neonatal period In urine: tubular damage, kidney transplant rejection	
C-reactive protein CRP	105 000–114 700	$\beta$ - $\gamma 2$ related to $\text{Ca}^{2+}$ concentration	<0.005	An activator of classical complement pathway ( $\text{Ca}^{2+}$ -dependent) Binds polysaccharides in many bacteria, fungi, and protozoal parasites Can initiate opsonization, phagocytosis, and lysis of invading cells In presence of $\text{Ca}^{2+}$ binds phosphorylcholine, lecithin, nucleic acid CRP is closely related to serum amyloid P No genetic abnormalities	Most sensitive acute-phase reactant (increases up to 2000 times) Increases are nonspecific	Childhood
Post $\gamma$ -globulin Cystatin C	11 500	Post $\gamma$	Trace	Inhibits cysteine, proteases	In cerebrospinal fluid (CSF): degenerative diseases In urine: tubular damage	

Lysozyme (muraminidase)	15 000	Post $\gamma$	0.036–0.078	In lysosomes intracellularly and in most extracellular fluids (exocrine secretions) Bactericidal for enzymatic activity against constituents in bacterial cell walls Produced by granulocytes and monocytes but not lymphocytes	Tuberculosis Sarcoidosis Granulomatous diseases Monocytic leukemia In urine: tubular damage	
Ferritin	450 000		M:30–220 $\mu\text{g l}^{-1}$ F:20–110 $\mu\text{g l}^{-1}$	Plasma concentration correlates iron storage (not in acute-phase reaction) Is present in many tissues: e.g., liver, spleen, bone marrow	Iron storage diseases (hemochromatosis, multiple transfusions) Acute and chronic inflammations Liver diseases	Iron deficiency Malnutrition
Carcinoembryonic antigen CEA	200 000		> 5 $\mu\text{g l}^{-1}$	Glycoprotein	Malignant tumor of colon, lung, pancreas, stomach Smoking Pulmonary infections Hepatitis Ulcerative colitis	
Immunoglobulin G (7S- $\gamma$ globulin) IgG	150 000	$\alpha 2-\gamma 2$	8.0–17.0 [92–196]	Antibodies, mainly, antiviruses, antibacteria, and toxins Crosses the placenta by an active transport Subclasses: IgG1, IgG2, IgG3, IgG4 GV: Allotype GM chain; allotype KM K chain	Chronic hepatitis, cirrhosis Chronic infections Parasitic diseases Autoimmune diseases Sarcoidosis	Acquired immunodeficiencies Congenital deficiencies (class or subclass) Protein-losing syndrome Pregnancy Macroglobulinemia of Waldenstrom
Immunoglobulin A IgA	160 000 and aggregates F:0.85–4.5 [51–268]	$\beta 2-\gamma 1$	M:1.0–4.9 [60–292]	Antibodies, subclasses: IgA1, IgA2 GV: Allotype AM of a chain; allotype KM of K chain	Chronic hepatitis, cirrhosis Chronic infections mainly of gastrointestinal and respiratory tract Intestinal inflammations and malignancies Autoimmune disorders,	Acquired immunodeficiencies Congenital deficiencies Ataxia, telangiectasia Protein-losing syndromes Childhood Non-IgA myeloma Macroglobulinemia of Waldenstrom

*Continued*



**Table 1** Continued

<i>Protein</i>	<i>Relative molecular mass</i>	<i>Mobility<sup>a</sup> (agarose pH 8.6)</i>	<i>Reference intervals (g l<sup>-1</sup>) [IU ml<sup>-1</sup>]</i>	<i>Function, properties, genetic variants (GVs)</i>	<i>Elevated in</i>	<i>Decreased in</i>
					mainly rheumatoid arthritis Wiskott–Aldrich syndrome IgA myeloma (often polymerized)	
Secretory IgA s IgA	385 000	$\gamma 1$		Antibodies, in secretory mucous membranes and in colostrum (2 IgA + secretory component (SC) + a J chain)		Deficit of IgA secretion
Immunoglobulin M(1s-globulin) IgM	971 000 pentamer	$\gamma 1-\gamma 2$	M:0.5–3.2 [58–368] F:0.6–3.7 [69–425]	First or primary response to an antigen Isohemoagglutinins (anti-A and anti-B) IgM can be present as monomer in childhood and in IgM monoclonal gammopathies	Acute and chronic infections, mainly viral Waldenstrom macroglobulinemia Nephrotic syndrome Autoimmune disorders Sarcoidosis Hepatocellular diseases IgM hyperdisgamma-globulinemia Congenital infections (newborn)	Acquired immunodeficiencies Congenital deficit Protein-losing syndromes Childhood Non-IgM myeloma
Immunoglobulin D IgD	175 000	$\gamma 1$	<0.15 [ $<100$ ]	Antibodies Surface receptors for antigen on B lymphocytes	Autoimmune infections Chronic infections IgD myeloma	Congenital deficit Childhood Acquired immunodeficiencies Non-IgD myeloma
Immunoglobulin E IgE	190 000		<0.0003 [ $<100$ ]	Antibodies (reagins)	Reaginic hypersensitivity reactions Parasitic infestations Non-IgE myeloma	Congenital deficit Acquired immunodeficiencies

<sup>a</sup> The proteins are listed in order of their electrophoretic mobilities at pH 8.6. The symbols  $\alpha 1$ ,  $\alpha 2$ ,  $\beta 2$ , and  $\gamma$  refer to zones of a given position in standard cellulose acetate electrophoresis and are correlated with their respective mobilities. Absolute values of intrinsic electrophoretic mobilities are not given here. M, male; F, female.

## Clinical Significance and Reference Intervals of the Best-Known Proteins

Table 1 lists the principal plasma proteins according to their electrophoretic mobilities in agarose with relative molecular mass, reference values, functions, and properties, and the causes of elevated and decreased levels. Other proteins are considered in detail in other articles in this encyclopedia (e.g., enzymes, lipoproteins, hemoglobin, and fibrinogen and other coagulation proteins).

See also: **Blood and Plasma. Capillary Electrophoresis:** Overview. **Cerebrospinal Fluid. Clinical Analysis:** Sample Handling. **Electrophoresis:** Principles; Isoelectric Focusing; Polyacrylamide Gels; Clinical Applications. **Immunoassays:** Overview. **Immunoassays,**

**Applications:** Clinical. **Immunoassays, Techniques:** Enzyme Immunoassays; Luminescence Immunoassays. **Liquid Chromatography:** Amino Acids.

## Further Reading

- Burtis CA and Ashwood ER (eds.) (2001) *Tietz Fundamentals of Clinical Chemistry*. Philadelphia: W.B. Saunders.
- Henry JB (ed.) (2001) *Clinical Diagnosis and Management by Laboratory Medicine*. Philadelphia: W.B. Saunders.
- Heusghem C, Alpert A, and Benson ES (1982) *Advanced Interpretation of Clinical Laboratory Data*. New York: Dekker.
- Tietz NW (ed.) (1995) *Clinical Guide to Laboratory Tests*. Philadelphia: W.B. Saunders.

## Foods

F Bonomi, Università degli Studi di Milano, Milano, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Determination of proteins in food still poses some problems to the chemist and the biochemist. The high complexity of most food matrices in terms of compositions and the many interactions between chemically different components in food contribute to the difficulties. Furthermore, all the physical, enzymatic, and microbiological treatments involved in the transformation of animal and plant materials into food call for adapting even the most standard analytical procedures to the particular properties and peculiar requirements of a given food.

Proteins are perhaps the most sensitive indicator of food quality. This not only applies to the relevant issue of preventing food adulteration and fraud, as detected by the presence of foreign proteins or of undesired degradation products, but to the appreciation of the freshness of ingredients, of proper processing and handling, and of adequate storage. Most of these indicators may be derived from the analysis of proteins in food and of their modifications.

Also relevant is the identification and quantification of proteins (and of their derivatives) that are known to be detrimental to the health of sensitive individuals, or that – on the contrary – are claimed to be beneficial to human health. In view of the increased consumer awareness with respect to these issues, analytical

procedures have been developed to detect the presence of proteins eliciting negative physiological responses (either true food allergies or food intolerances of various origin), and that of proteins or peptides having beneficial effects on the well-being of humans.

This article does not deal with analysis of food enzymes, which may rely on tests adapted from those used in clinical biochemistry, but rather deals with that ‘silent majority’ of proteins in food, which lack the signature features making them easily recognizable with dedicated analytical techniques or by suitable activity assays.

## Methods Based on Protein Nitrogen

The determination of the nitrogen content of food samples, based on its conversion into simple nitrogen derivatives, is still useful, and remains the official method for determining the overall protein content in many countries and for most foods. Separation of (specific) proteins from other components in the sample is not required, unless the food contains appreciable amounts of nonprotein nitrogen-containing material (e.g., nucleotides). Typically, the protein content of the food is given as a percentage calculated by multiplying the determined nitrogen content by a factor that depends upon the material being investigated. Typical coefficients are 5.70 for cereal-based foods, 6.38 for meats, milk, and dairy products, or 6.25, which represents the most widely used value for generic foods.

## Clinical Significance and Reference Intervals of the Best-Known Proteins

Table 1 lists the principal plasma proteins according to their electrophoretic mobilities in agarose with relative molecular mass, reference values, functions, and properties, and the causes of elevated and decreased levels. Other proteins are considered in detail in other articles in this encyclopedia (e.g., enzymes, lipoproteins, hemoglobin, and fibrinogen and other coagulation proteins).

See also: **Blood and Plasma. Capillary Electrophoresis:** Overview. **Cerebrospinal Fluid. Clinical Analysis:** Sample Handling. **Electrophoresis:** Principles; Isoelectric Focusing; Polyacrylamide Gels; Clinical Applications. **Immunoassays:** Overview. **Immunoassays,**

**Applications:** Clinical. **Immunoassays, Techniques:** Enzyme Immunoassays; Luminescence Immunoassays. **Liquid Chromatography:** Amino Acids.

## Further Reading

- Burtis CA and Ashwood ER (eds.) (2001) *Tietz Fundamentals of Clinical Chemistry*. Philadelphia: W.B. Saunders.
- Henry JB (ed.) (2001) *Clinical Diagnosis and Management by Laboratory Medicine*. Philadelphia: W.B. Saunders.
- Heusghem C, Alpert A, and Benson ES (1982) *Advanced Interpretation of Clinical Laboratory Data*. New York: Dekker.
- Tietz NW (ed.) (1995) *Clinical Guide to Laboratory Tests*. Philadelphia: W.B. Saunders.

## Foods

F Bonomi, Università degli Studi di Milano, Milano, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Determination of proteins in food still poses some problems to the chemist and the biochemist. The high complexity of most food matrices in terms of compositions and the many interactions between chemically different components in food contribute to the difficulties. Furthermore, all the physical, enzymatic, and microbiological treatments involved in the transformation of animal and plant materials into food call for adapting even the most standard analytical procedures to the particular properties and peculiar requirements of a given food.

Proteins are perhaps the most sensitive indicator of food quality. This not only applies to the relevant issue of preventing food adulteration and fraud, as detected by the presence of foreign proteins or of undesired degradation products, but to the appreciation of the freshness of ingredients, of proper processing and handling, and of adequate storage. Most of these indicators may be derived from the analysis of proteins in food and of their modifications.

Also relevant is the identification and quantification of proteins (and of their derivatives) that are known to be detrimental to the health of sensitive individuals, or that – on the contrary – are claimed to be beneficial to human health. In view of the increased consumer awareness with respect to these issues, analytical

procedures have been developed to detect the presence of proteins eliciting negative physiological responses (either true food allergies or food intolerances of various origin), and that of proteins or peptides having beneficial effects on the well-being of humans.

This article does not deal with analysis of food enzymes, which may rely on tests adapted from those used in clinical biochemistry, but rather deals with that ‘silent majority’ of proteins in food, which lack the signature features making them easily recognizable with dedicated analytical techniques or by suitable activity assays.

## Methods Based on Protein Nitrogen

The determination of the nitrogen content of food samples, based on its conversion into simple nitrogen derivatives, is still useful, and remains the official method for determining the overall protein content in many countries and for most foods. Separation of (specific) proteins from other components in the sample is not required, unless the food contains appreciable amounts of nonprotein nitrogen-containing material (e.g., nucleotides). Typically, the protein content of the food is given as a percentage calculated by multiplying the determined nitrogen content by a factor that depends upon the material being investigated. Typical coefficients are 5.70 for cereal-based foods, 6.38 for meats, milk, and dairy products, or 6.25, which represents the most widely used value for generic foods.

There are two procedures based on total nitrogen determination:

1. The Kjeldahl conversion of amino and amide groups into ammonia, which is achieved by treating the sample with strong acids at high temperatures in the presence of a catalyst. Following alkalization of the digest, the ammonia is steam-distilled into a nonvolatile acid and titrated.
2. The oxidative conversion of sample nitrogen into  $\text{NO}_x$ , which is then reduced to a chromatographically detectable form. Analysis is typically carried out in an automated fashion along the lines of the CHN analyzers.

Both methods require minimal if any sample preparation, and extensively automated systems are available. The highly corrosive chemicals and the harsh conditions used in the Kjeldahl digestion call for appropriate fume hoods and exhaust systems, and standardization of the digestion itself may sometimes be difficult. The relatively low sensitivity and the fairly large amount of sample required are usually no problem in the food industry. As for chromatographic methods, controlled oxidative pyrolysis of food releases a number of volatile compounds that may foul the separation columns. This requires careful maintenance of the equipment, and in particular of the precolumn that guards the separation apparatus.

## **Spectrometric Methods**

Application of infrared spectroscopy to food analysis has made possible the highly automated determination of protein content in foodstuffs. Available commercial instruments allow the simultaneous determination of water, lipid, fiber, and protein content in solid or liquid samples with little if any sample preparation and impressive throughput.

Although the peptide bonds of proteins absorb at several wavelengths, low-energy electronic transitions, overtones, and combinations of stretching and bending vibrations are best evaluated in the near-infrared region ( $0.8\text{--}3\mu\text{m}$ ). Mid-infrared instruments operating in the  $3\text{--}15\mu\text{m}$  range are set to read at  $6.46\mu\text{m}$ , where the secondary amide groups of peptide bonds absorb, for protein quantification. Near-infrared spectroscopy may be used for online monitoring of processes, and has been applied to issues such as crop selection and breeding studies.

Transmission instruments are used in milk analysis, and suitable provisional standards are available for fat, protein, and lactose determination in whole milk by use of a mid-infrared instrument. Near-infrared reflectance techniques are typically used for

assessing the protein content in solid samples, such as grains and flours, although some commercial instruments operate in the transmission mode. Near-infrared instruments are marketed for specific applications, and typically come with sets of selective filters that allow online determination of the required parameters.

The possibility of direct and online evaluation of nutritionally relevant compounds by mathematical treatment of spectral data, and the possibility of predicting properties of the sample that are relevant for further processing make extending the applications of these techniques to food protein analysis particularly appealing.

## **Methods Based on Amino Acid Analysis**

Amino acid analysis is a very useful tool in the analysis of food since it not only allows an accurate evaluation of the amount of protein in the sample but also provides valuable information on the quality of the food under investigation, in particular as for the content in nutritionally relevant amino acids. Further information on the quality of the investigated food may be derived from the determination of specific amino acids or amino acid derivatives that are recognized as markers of improper processing or storage.

Although the direct spectrometric determination of certain amino acids in food by derivative spectrometry in appropriate spectral regions or by specific reaction with suitable reagents after unfolding of the protein structure has been attempted, amino acid analysis typically requires hydrolysis of the proteins followed by separation of the liberated amino acids.

Liquid chromatography determination of amino acids relies on procedures based either on precolumn or postcolumn derivatization. A variety of reagents and detectors are available for postcolumn derivatization, whereas the choice is somewhat more limited as for precolumn derivatization. Sensitivity, repeatability, and accuracy are extremely good with either procedure. Available commercial instruments have detection limits in the picomole range.

In the particular case of foodstuffs, problems may arise from the presence of other food components, such as some lipids, reducing sugars, and phenolic compounds, which may form derivatives resulting in altered quantification. For problem samples, the use of internal standards added prior to hydrolysis is highly recommended.

Amino acid analysis is widely used for detecting adulteration, improper processing, or fraud. Some

amino acids are typical of certain proteins, in most cases as a consequence of protein-specific post-translational modifications. Well-known examples include 1- and 3-methylhistidine in muscle proteins, and the hydroxylated forms of lysine and proline in proteins from the connective tissue. Comparison with suitable standards can detect the addition of foreign proteins to meat-based foods or the use of lesser quality materials in high-quality formulae. Another possibility is the detection of amino acid derivatives that are formed as a consequence of extreme processing conditions or of the addition of stabilizing agents. High temperatures and extreme pH values in food processing or in food sanitization leads to a number of alterations in the amino acid structure. L-Amino acids can isomerize to the D-form; they can react with other amino acids or with their decomposition products to decrease the nutritional value of food (as exemplified by the formation of dehydroalanine from cysteine, and of lysinoalanine); they can react with other food components (sugars, in most cases) to form adducts that subsequently rearrange into stable products. Also of interest for food quality are the products of de-amination or de-carboxylation reactions, which modify amino acids yielding compounds that – besides their significance as markers of food spoilage – are also physiologically active. These products include biogenic amines (from decarboxylation) and some neurotransmitters (from de-amination).

In this frame, special procedures have been devised for the detection and quantification of specific products, including the development of fast and reliable liquid chromatography–mass spectrometry (LC–MS) methods that are also suitable for detecting multiple and simultaneous reactions at different amino acid side chains. These dedicated procedures are advantageous in that they may not involve the complete and time-consuming routine of hydrolysis and chromatography, but special attention must be paid to the standardization of the procedures, and to the presence of interfering materials.

## Methods Based on Mass Spectrometry

The development of dedicated methodologies, instrumental progresses, and the ever-improving quality/price ratio for mass spectrometers due to the increasing popularity of the technique is increasing the relevance of mass spectrometry in food protein analysis. Although application of the many variations of mass spectrometric techniques to food analysis is still under development and mass spectrometry labs dedicated to food analysis are not common, this could be the technique of choice when the analytical

requirements are too challenging for more established, common, and instrumentally cheaper procedures. In principle, the sensitivity of a mass spectrometer for protein molecules is in the  $10^{-15}$  mol range, by far superior to any other available detection method. A major methodological breakthrough has been the combination of high-resolution separation techniques such as two-dimensional electrophoresis or capillary reverse-phase high-performance liquid chromatography (HPLC) ionization with mass spectrometric analysis, which is termed as the ‘proteomic’ analysis.

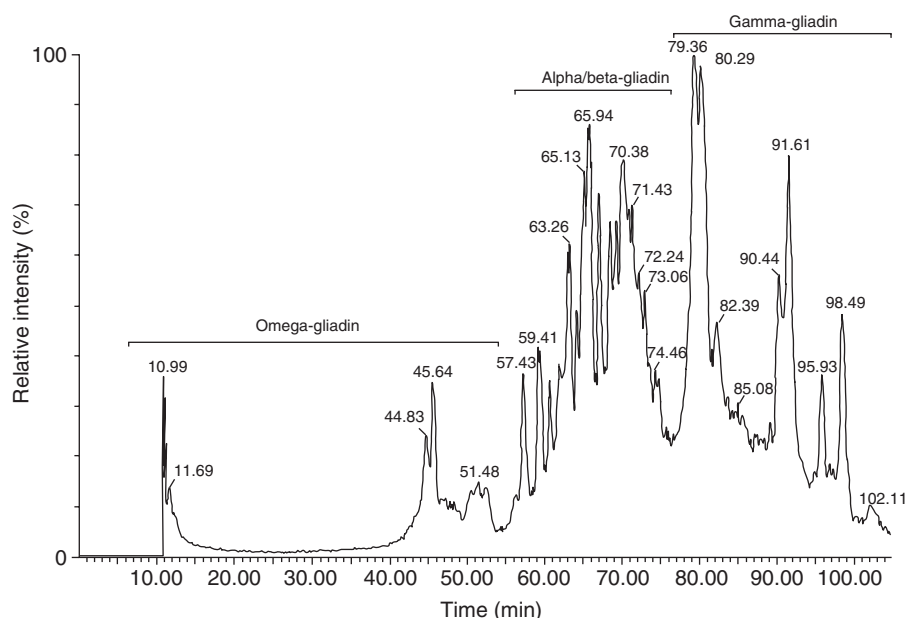
Electrospray mass spectrometry (ES–MS) or matrix-assisted laser desorption/ionization mass spectrometry (MALDI–MS) analysis can be used for determining the molecular weight of peptides and proteins. The accuracy of ES–MS for measuring molecular weight is impressive, as the standard deviation for routine measurements is represented by a mass shift falling within the range 2–3 Da for proteins up to 100 kDa. At the same time, MALDI-MS has complementary advantages over ES–MS, particularly when analyzing proteins with a mass greater than 100 000. The mass spectrometer can not only directly provide information on the mass of a particular peptide but can also be used to generate *de novo* amino acid sequence information, including post-translational modifications, from tandem mass spectra obtained either by postsource decay or by collision-induced dissociation (see **Figure 1**).

Also very interesting is the possibility of detecting process- or storage-induced modifications, and to monitor fundamental events that affect the structure and composition of proteins in the food industry, all in a single experiment and from minute amounts of sample. Estimating the extent and assessing the products of enzymatic or microbial proteolysis, and of process-induced modifications of food proteins (including the formation of isopeptide bonds by enzyme action or by harsh treatments) to whole food samples are some examples of how the straightforward applications of mass spectrometry may simplify food protein analysis.

## Applications of Electrophoresis and Chromatography to Food Proteins

Separative methods are advantageous in that they offer both qualitative and quantitative information, and therefore are the most powerful discriminating tool when it comes to identifying and quantifying the presence of particular proteins in food.

Application of standard HPLC procedures, based on various principles, to the separation, identification, and quantification of individual food proteins is



**Figure 1** LC-ES-MS chromatograms of the gliadin fraction from a mixture of wheat varieties. Gliadins were extracted from wheat in ethanol 70%, and directly analyzed by LC-ES-MS. (Courtesy of Prof. Pasquale Ferranti.)

often complicated by several factors, which may include: the limited solubility of many food proteins; the intrinsic heterogeneity of some proteins (as a consequence of post-transcriptional or post-translational modifications); and the occurrence of covalent and noncovalent modifications upon processing, and of hydrolytic processes during storage or ripening.

Solubility problems have been overcome by the use of denaturing agents, also in combination with reducing agents acting on disulfide bonds. This makes possible the quantitative chromatographic analysis of individual protein components also in complex systems, such as casein micelles in milk and the storage proteins in cereals and legumes. The development of specific separation protocols and the increasing use of LC-MS has changed the situation as for protein microheterogeneity, which represented a drawback for conventional methods, but is now exploited to detect subtle variations in the starting material or induced by processing/storage conditions.

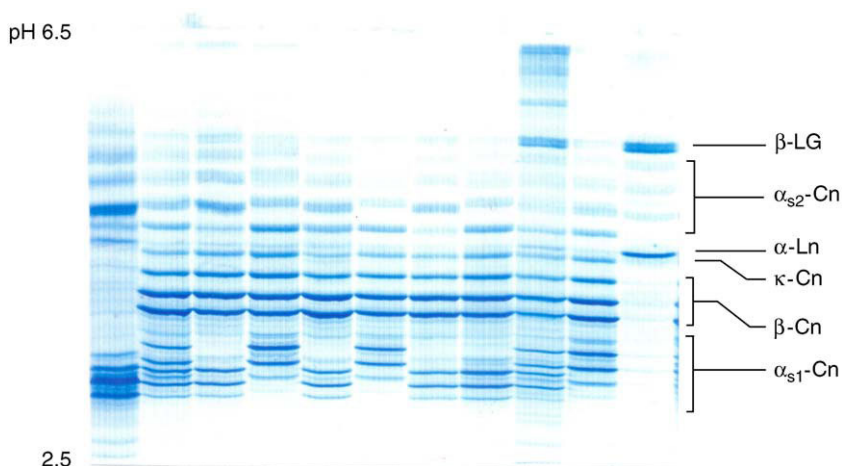
Electrophoretic techniques are also highly popular in protein food analysis. Polyacrylamide gel-based separation, either depending on the protein size in denaturing conditions or on its charge in specific solvent systems (as exemplified by those used for analysis of cereal proteins), is still highly popular because of their good sensitivity, ease of use, and capability to compare a relatively large number of samples on the same gel. This unique combination of features makes electrophoresis a popular choice also from an economic standpoint.

The high resolving power of isoelectric focusing (IEF) also has been thoroughly exploited in food analysis, in particular after immobilized pH gradients have become commonplace. Again, uncharged denaturants or nonionic detergents help in tackling most of the solubility issues, which are of particular relevance in IEF. Image analysis tools and the remarkable progresses in bioinformatics in relation to proteomics was advantageous to the application of gel-based electrophoretic techniques to food proteins. The introduction of the sensitive staining technique – such as the many types of silver and gold staining – has brought the detection limits of polyacrylamide-based electrophoresis to sub-nanogram quantities of protein (see **Figure 2**).

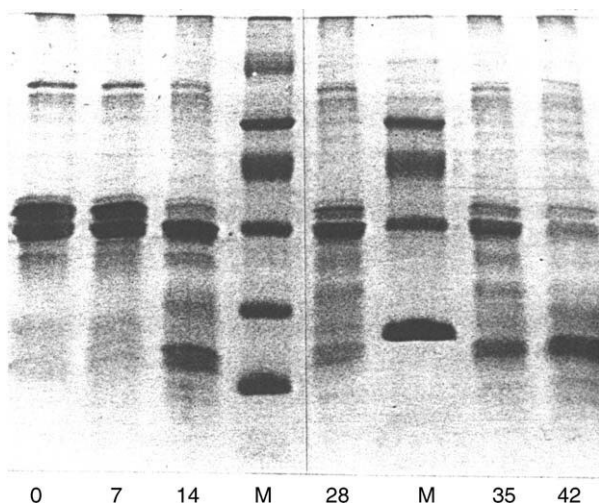
As of this date, in spite of its high resolving power and of extensive experimentation, the application of capillary electrophoresis to food protein analysis has not gained wide recognition, at least at the level of official or recommended methods, although this separation technique offers several advantages over conventional electrophoresis in terms of sensitivity (detection of  $10^{-10}$  g protein has been reported) and speed of operation (see **Figure 3**).

One of the major drawbacks of separation techniques when applied to food protein analysis is the availability of protein standards of suitable high quality. Commercial highly purified proteins are very often a mixture of microheterogeneous proteins, so that individual laboratories may have to rely on in-house prepared standards, and complications ensue at the validation level.





**Figure 2** IEF of milk from different goat breeds. It was performed in an immobilized pH gradient (pH 2.5–6.5) in the presence of  $8 \text{ mol l}^{-1}$  urea and  $1 \text{ mmol l}^{-1}$  dithiothreitol. Rightmost lane is goat milk whey proteins isolated by ultracentrifugation. Proteins are identified on the right side. (Courtesy of Dr Maria Feligini.)



**Figure 3** Time course of protein hydrolysis in the ripening of soft Italian cheese. SDS-PAGE was performed under reducing conditions on defatted, homogenized cheese samples. Age of cheese (days after brine) is given for each sample lane. M, molecular mass markers (two different sets).

Separation techniques allow further characterization of the isolated species. Peaks from chromatographic and capillary electrophoresis runs may be collected, and electrophoretic bands on gels may be excised or blotted onto suitable membranes. Then, more powerful and sensitive techniques may be used for absolute sample identification or for the detection of specific local modifications. Common examples of these postseparative techniques include: MS or LC-MS analysis and chemical or MS sequencing of the intact protein or of its enzymatic digests; immunochemical recognition of native or modified protein

regions; and identification of nonprotein constituents by suitable chemical or spectroscopic methodologies.

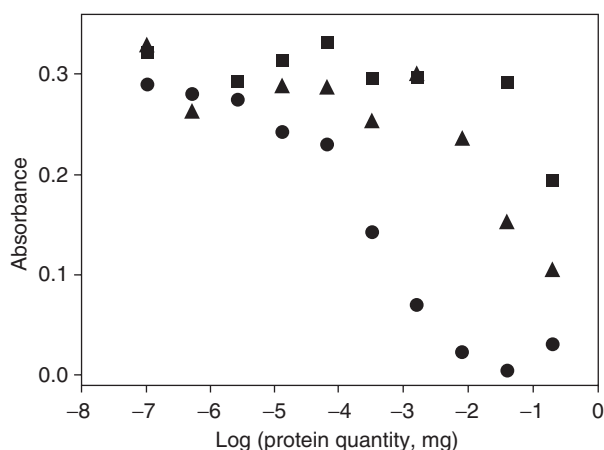
## Immunological Methods

Immunological methods in food analysis offer the intrinsic advantage of very high sensitivity and specificity, and of ease of automation. However, interferences from the complex food matrices and the requirement for sample preparation steps contribute to make their application to food analysis less straightforward than in other analytical fields.

Most food proteins are excellent antigens both in their native form and in the denatured form that is often isolated from separation procedures. This makes it relatively easy to raise suitable antibodies in any of the animal species commonly used for this purpose. A relatively large number of antibodies against food proteins are readily available commercially, as are several antibody conjugates useful for quantitative and semiquantitative analysis.

Techniques used for the detection and quantification of food proteins can be divided into purely immunological techniques and techniques that combine the specificity of detection by appropriate antibodies with separation method.

The most widely used immunological approaches are those based on enzyme-conjugated antibodies, offered as “kits” that allow immunochemical detection of allergens and toxic components in foods. These products are popular in that they offer ease of automation, the possibility of analyzing a large number of samples in a single approach, besides relying on safe chemicals and simple protocols. Either direct or competitive enzyme-linked immunosorbent assay



**Figure 4** Reduction of immunoreactivity upon digestion of casein with different proteases. Competitive ELISA was performed by using polyclonal anti-casein rabbit antibodies on sodium caseinate before (dots) and after controlled digestion with a protease of pancreatic (triangles) or bacterial origin (squares).

(ELISA) formats are commonly used, depending on the nature of the coating originally placed in the microplate wells and on the nature of the conjugate used for the final detection. These techniques offer high sensitivity (detection of sub-nanogram amounts is commonplace) and relatively fast responses, and in most cases do not require sample preparation steps (see Figure 4).

When sample preparation steps are required, caution must be exercised with respect to the fact that very often they have been designed for a specific food, and therefore are not directly transferable to another food matrix. Careful preliminary experimentation and use of internal standards are mandatory when adapting to a specific food (or to food processed in a different way) a procedure originally designed for some other food. Indeed, protein antigens in processed food may become hidden as a consequence of the structural and chemical modifications induced by the severe physical treatments. These modifications can affect both the protein antigen itself and, more commonly, the surrounding matrix, in such a way that accessing the protein of interest becomes difficult. Examples of these difficulties include foods with high starch or high fat content, where preliminary enzymatic hydrolysis of the nonprotein matrix may be required for proper quantification of some protein species. Modifications of the protein itself are particularly difficult to deal with, in particular when monoclonal antibodies are used. Epitope-specific monoclonal antibodies have a great deal to offer in terms of selectivity and specificity, but the regions they recognize sometimes

may be lost in the treatments used for food processing and preservation.

Combining the specificity of antibody recognition with protein separation techniques, mostly electrophoretic ones, offers obvious advantages, which are most often exploited at the level of research and development studies. Specialized procedures such as rocket electrophoresis and variations thereof are no longer used, mostly because of the large amounts of antibodies they require. Western Blotting has become the most popular methodology: proteins are separated electrophoretically (by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), IEF, etc.), and protein bands are then transferred onto a hydrophobic membrane (nitrocellulose, polyvinylidene fluoride (PVDF), etc.). Unspecific binding sites on the membrane are blocked with an inert protein, and the analyte on the membrane is reacted with a suitable primary antibody. The antigen-antibody complex remaining after washing is then detected by a conveniently labeled secondary antibody. These methods have been used extensively for the detection and quantification of foreign proteins in materials derived from a single major commodity, and to a lesser extent to assist with establishing the pattern of sensitivity of a given individual to the different proteins in a given food. In this latter case, the serum from the patient is used as the primary antibody. Limitations of these techniques in the analysis of food proteins are related to the difficulties encountered in the electrophoretic separation of small peptides (such as those originating from proteolytic ripening processes), and to the cross-reactivity toward different protein antigens that may occur in some antibody preparations (in particular when unfractionated sera are used).

## Protein Denaturation and Interaction with Other Food Components

Treatments used in food processing may modify the native structure of proteins independently of or in concomitance with chemical/enzymatic modification of the amino acids in their primary structure. In most cases, such structural changes are a highly desirable result of processing, and process conditions are chosen that result in extensive (but controlled) protein denaturation and in the exposure of reactive functional groups. In many other cases, however, protein denaturation represents a major drawback, in that it may alter the nutritional and functional properties of the processed material.

From a chemical standpoint, there are two major classes of modifications involving proteins: those leading to the disruption or formation of new covalent

bonds and those that involve only intrinsically weak intramolecular or intermolecular interactions.

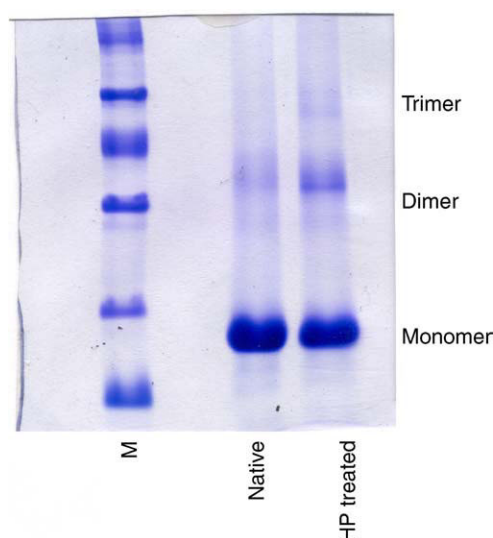
From a structural/functional standpoint, modification of a protein is reflected in changes in its aggregation state and in its surface properties. Changes in the aggregation state lead to changes in solubility and in network-forming capability, and in the sensitivity to proteolytic action as a consequence of modified exposure of the protein regions where proteases may act. Changes in the surface properties of proteins affect their emulsifying and foaming properties, and their ability to bind hydrophobic molecules or surfaces. These changes are often exploited by the food technologist, provided that they be controlled and used for specific goals. For these reasons, much effort has been dedicated to understanding and quantifying the denaturation of food proteins, and also in the general frame of all those modifications that affect the quality of a protein-based food and that are generally described as 'process damage'. Evaluating the extent of this damage, both in kinetic and in end-point studies, has provided some insights as for optimizing processes by minimizing nutritionally and technologically relevant modifications. These studies have been applied to edible fluids (milk and eggs, for instance), but are currently being extended to dry and semi-dry foods (see Figure 5).

Modifications ensuing from the formation of novel covalent bonds in food proteins include: (1) intra- and intermolecular formation of isopeptide bonds (also as a consequence of the use of transglutaminase

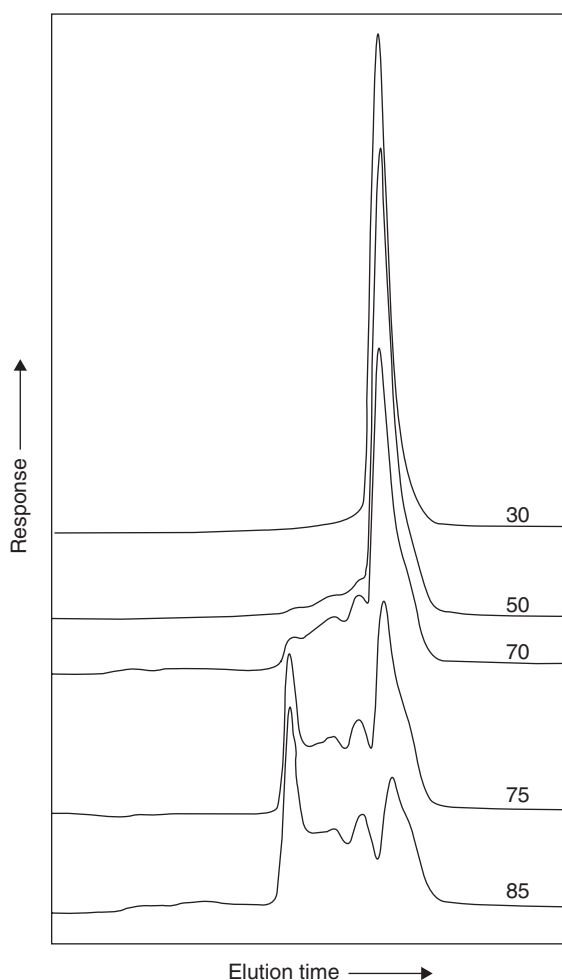
for improving the rheological properties of food); (2) formation of dehydroalanine from cysteine or serine residues, possibly followed by condensation of dehydroalanine with other amino acids (most notably, lysine) or with other food components; (3) Maillard-type reactions between amino acids and carbohydrates; and (4) intermolecular or intramolecular disulfide exchange. Several analytical techniques can detect these modifications. Formation of isopeptide bonds may be detected from the modified chromatographic and MS patterns in enzymatically hydrolyzed proteins, and the nature of the involved residues may be detected by MS. Chromatographic techniques, alone or in combination with MS, can monitor the formation of lysinoalanine, lanthionine, and of other condensation products between dehydroalanine and amino acids, as well as of condensation products between amino acids and sugars, such as furosine, which are recognized as markers of the intensity of thermal processing in milk and cereal-based foods. Maillard-type reactions can be monitored by the formation of their colored products, which can be detected by surface or immersion luminometry (such as tristimulus colorimetry) or by image analysis. Disulfide exchange leads to changes in the aggregation state of proteins, which can be detected by size-based separation under denaturing (but not reducing) conditions or, when extensive, by changes in protein solubility (see Figure 6).

Changes in the aggregation state may result in insolubilization of the protein, or its binding (not necessarily through covalent bonds) to other insoluble components. Thus, disappearance of a specific protein from the pool of soluble proteins in a given food may be detected by any of the techniques discussed above, and taken as a reliable index of process damage. A prime example of this simple approach to defining food quality at the molecular level may be the detection of individual whey proteins to define the quality of heat-treated milk.

However, protein insolubilization is often the final event in a sequence of modifications that involve all the levels of protein structural organization. For this reason, the mechanism of food protein denaturation is now the focus of research efforts aimed at defining the properties of the transient intermediates formed during processes and evolving into different final products as a consequence of their properties and of the kinetics of their formation. This approach opens new perspectives for modeling purposes and for directing the whole process toward the desired products. The approaches used in these studies are taken from structural biochemistry and physical chemistry, or combine protein chemistry approaches with others (such as calorimetric techniques) better suited for describing changes in



**Figure 5** Detection of disulfide-linked aggregates in pressure-treated  $\beta$ -lactoglobulin by SDS-PAGE separation under non-reducing conditions. Protein was incubated for 10 min at 37°C at 0.1 (native) or 600 MPa (pressure treated). M, molecular mass markers.



**Figure 6** Detection of thermal aggregation of  $\beta$ -lactoglobulin by gel-permeation chromatography. Solutions of  $\beta$ -lactoglobulin ( $3.8 \text{ mg ml}^{-1}$ ) in  $50 \text{ mmol l}^{-1}$  phosphate, pH 6.8, were heated for 15 min at the temperatures given in the figure ( $^{\circ}\text{C}$ ). Aliquots of the clear solutions of the heated protein were run on a Superdex column in  $50 \text{ mmol l}^{-1}$  phosphate,  $0.1 \text{ mol l}^{-1}$  NaCl, pH 7.0.

other food components that are interacting with the protein system under investigation.

## Examples of Determination of Specific Proteins in Processed Foods

### Immunochemical Detection of Gliadins and Gluten

Celiac disease represents an autoimmune disease associated with the consumption of gluten. Gluten represents the main water-insoluble protein fraction in several cereals, such as wheat, rye, barley, and triticale. Celiac individuals must absolutely follow a gluten-free diet, also in consideration of the fact that assumption of gluten has additive effects. However, successful implementation of a gluten-free diet is complicated by the frequent use of gluten-containing

grains, especially wheat, and of their derivatives in numerous food products, either alone or associated with other food ingredients. CODEX has established a 100 ppm limit for gluten content in gluten-free products, although the allowed daily intake of gluten in celiac patients is still much debated.

Analysis of gliadins ( $\sim 50\%$  of gluten proteins) is most conveniently carried out by immunochemical ELISA tests that all call for an initial step in which water-insoluble gliadins are extracted into aqueous ethanol (40–70% v/v). This step represents the main critical point of the procedure since the solubility of gliadins may be affected by any of the technological processes used for food production. In this frame, proper protocols for gliadin extraction are developed according to the particular food matrix.

ELISA gluten tests are also prone to interference from other food components. For example, a tannin-binding additive should be used for cocoa-based food to ensure that the detection of gluten is not inhibited. The detection of gluten in highly viscous samples of syrups obtained from starch hydrolysis is also difficult. Both the high amount of sugars in the sample and small residues of polysaccharides in the ethanolic gliadin extract result in heavy interference in the ELISA test. In this case, procedures other than sample dilution (such as sugar hydrolysis and removal) must be performed in order to reduce the glucose content to levels where it no longer interferes with the assay, while avoiding gliadin losses due to insolubilization or adhesion to the surfaces of the equipment.

### Detection of Modified Egg White Proteins

Ovalbumin is the major protein component of egg white, which represents 45% of the total proteins. As eggs age, or upon thermal treatment such as pasteurization of egg white, native ovalbumin is converted into a de-aminated form, known as S-ovalbumin. This event has obvious meanings, and is one of the many suggested markers of egg freshness, in particular when processed eggs are analyzed. However, since the denaturation temperature of S-ovalbumin is  $\sim 23^{\circ}\text{C}$  higher than that of the native protein, this conversion is also of practical relevance in that it may affect processing conditions for various egg-based products.

Detection of these ovalbumin isoforms is made difficult by the simultaneous occurrence of different glycosylated forms of the protein, which complicate the application of common and powerful separation techniques, such as reverse-phase HPLC or electrophoresis. Glycosylation not only affects the migration or elution patterns, it also makes difficult the

quantitation of the different isoforms by the common staining procedures used in electrophoresis, which are all affected to some extent by protein glycosylation.

For these reasons, methods for quantitative detection are based on ion-exchange chromatography, which exploits the charge difference stemming from de-amidation and allow easy resolution of the two ovalbumin isoforms on a suitable anion exchanger by applying a simple salt gradient in a total analysis time of 20 min. This compares favorably in terms of analysis time and amount of sample with other suggested procedures for detection of S-ovalbumin, such as those involving the use of thermal analysis techniques (differential scanning calorimetry, in this case) for discriminating protein isoforms with a different temperature sensitivity and for quantitating the enthalpy changes (and thus the protein amounts) associated with each protein species.

**See also:** **Mass Spectrometry:** Peptides and Proteins. **Peptides.**

## Further Reading

- Damodaran S and Paraf A (1997) *Food Proteins and their Applications*. New York: Marcel Dekker.
- Gagnaire LJ, Molle V, Pezennec D, and Bouhallab SS (2000) Application of chromatography and mass spectrometry to the characterization of food proteins and derived peptides. *Journal of Chromatography A* 881: 1–21.
- Kolakowski E (2001) Protein determination and analysis in food systems. In: Sikorski ZE (ed.) *Chemical and Functional Properties of Food Proteins*, pp. 57–112. Lancaster, USA: Technomic Publishing Company, Inc.
- Nakai S and Modler HW (2000) *Food Proteins: Processing Applications*. London: Wiley Europe.
- Nordlee JA and Taylor SL (1995) Immunological analysis of food allergens and other food proteins. *Food Technology* 49(2): 129–132.
- Recio I, Ramos M, and Lopez-Fandino R (2001) Capillary electrophoresis for the analysis of food proteins of animal origin. *Electrophoresis* 22: 1489–1502.
- Wetzel DLB and Charalambous G (1998) *Instrumental Methods in Food and Beverage Analysis*. London: Elsevier.

# PROTEOMICS

**R M Twyman**, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Proteomics is a rapidly growing area of molecular biology that is concerned with the systematic, large-scale analysis of proteins. It is based on the concept of the proteome as a complete set of proteins produced by a given cell or organism under a defined set of conditions. Proteins are directly involved in almost every biological process, so a comprehensive analysis of the proteins in the cell provides a unique global perspective on how these molecules interact and cooperate to create and maintain a working biological system. The cell responds to internal and external changes by regulating the level and activity of its proteins, so changes in the proteome, either qualitative or quantitative, provide a snapshot of the cell in action. The proteome is a complex and dynamic entity that can be defined in terms of the sequence, structure, abundance, localization, modification, interaction, and biochemical function of each of its components, providing a rich and varied source of data. The analysis of these various properties of the proteome requires an equally diverse range

of technologies. Proteomics provides a much more robust and representative picture of the functioning cell than other forms of large-scale biology, such as the sequencing of genomes or the global analysis of gene expression. The reasons for this are summarized in Table 1.

## Protein Separation

The analysis of proteins, whether on a small or large scale, requires methods for the separation of protein mixtures into their individual components. Protein separation methods can be placed on a sliding scale from fully selective to fully nonselective. Selective methods aim to isolate 'individual proteins' from a mixture usually by exploiting very specific properties such as their binding specificity or biochemical function. In contrast, nonselective separation methods aim to take a complex protein mixture and fractionate it in such a manner that all the individual proteins, or at least a substantial subfraction, are available for further analysis. Such methods lie at the heart of proteomics and exploit very general properties of proteins, such as their mass or net charge.

Many techniques can be used to separate complex protein mixtures in what at least approaches a



quantitation of the different isoforms by the common staining procedures used in electrophoresis, which are all affected to some extent by protein glycosylation.

For these reasons, methods for quantitative detection are based on ion-exchange chromatography, which exploits the charge difference stemming from de-amidation and allow easy resolution of the two ovalbumin isoforms on a suitable anion exchanger by applying a simple salt gradient in a total analysis time of 20 min. This compares favorably in terms of analysis time and amount of sample with other suggested procedures for detection of S-ovalbumin, such as those involving the use of thermal analysis techniques (differential scanning calorimetry, in this case) for discriminating protein isoforms with a different temperature sensitivity and for quantitating the enthalpy changes (and thus the protein amounts) associated with each protein species.

**See also:** **Mass Spectrometry:** Peptides and Proteins. **Peptides.**

## Further Reading

- Damodaran S and Paraf A (1997) *Food Proteins and their Applications*. New York: Marcel Dekker.
- Gagnaire LJ, Molle V, Pezennec D, and Bouhallab SS (2000) Application of chromatography and mass spectrometry to the characterization of food proteins and derived peptides. *Journal of Chromatography A* 881: 1–21.
- Kolakowski E (2001) Protein determination and analysis in food systems. In: Sikorski ZE (ed.) *Chemical and Functional Properties of Food Proteins*, pp. 57–112. Lancaster, USA: Technomic Publishing Company, Inc.
- Nakai S and Modler HW (2000) *Food Proteins: Processing Applications*. London: Wiley Europe.
- Nordlee JA and Taylor SL (1995) Immunological analysis of food allergens and other food proteins. *Food Technology* 49(2): 129–132.
- Recio I, Ramos M, and Lopez-Fandino R (2001) Capillary electrophoresis for the analysis of food proteins of animal origin. *Electrophoresis* 22: 1489–1502.
- Wetzel DLB and Charalambous G (1998) *Instrumental Methods in Food and Beverage Analysis*. London: Elsevier.

# PROTEOMICS

**R M Twyman**, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Proteomics is a rapidly growing area of molecular biology that is concerned with the systematic, large-scale analysis of proteins. It is based on the concept of the proteome as a complete set of proteins produced by a given cell or organism under a defined set of conditions. Proteins are directly involved in almost every biological process, so a comprehensive analysis of the proteins in the cell provides a unique global perspective on how these molecules interact and cooperate to create and maintain a working biological system. The cell responds to internal and external changes by regulating the level and activity of its proteins, so changes in the proteome, either qualitative or quantitative, provide a snapshot of the cell in action. The proteome is a complex and dynamic entity that can be defined in terms of the sequence, structure, abundance, localization, modification, interaction, and biochemical function of each of its components, providing a rich and varied source of data. The analysis of these various properties of the proteome requires an equally diverse range

of technologies. Proteomics provides a much more robust and representative picture of the functioning cell than other forms of large-scale biology, such as the sequencing of genomes or the global analysis of gene expression. The reasons for this are summarized in Table 1.

## Protein Separation

The analysis of proteins, whether on a small or large scale, requires methods for the separation of protein mixtures into their individual components. Protein separation methods can be placed on a sliding scale from fully selective to fully nonselective. Selective methods aim to isolate 'individual proteins' from a mixture usually by exploiting very specific properties such as their binding specificity or biochemical function. In contrast, nonselective separation methods aim to take a complex protein mixture and fractionate it in such a manner that all the individual proteins, or at least a substantial subfraction, are available for further analysis. Such methods lie at the heart of proteomics and exploit very general properties of proteins, such as their mass or net charge.

Many techniques can be used to separate complex protein mixtures in what at least approaches a



**Table 1** The unique position of proteomics in relation to other areas of large-scale biology

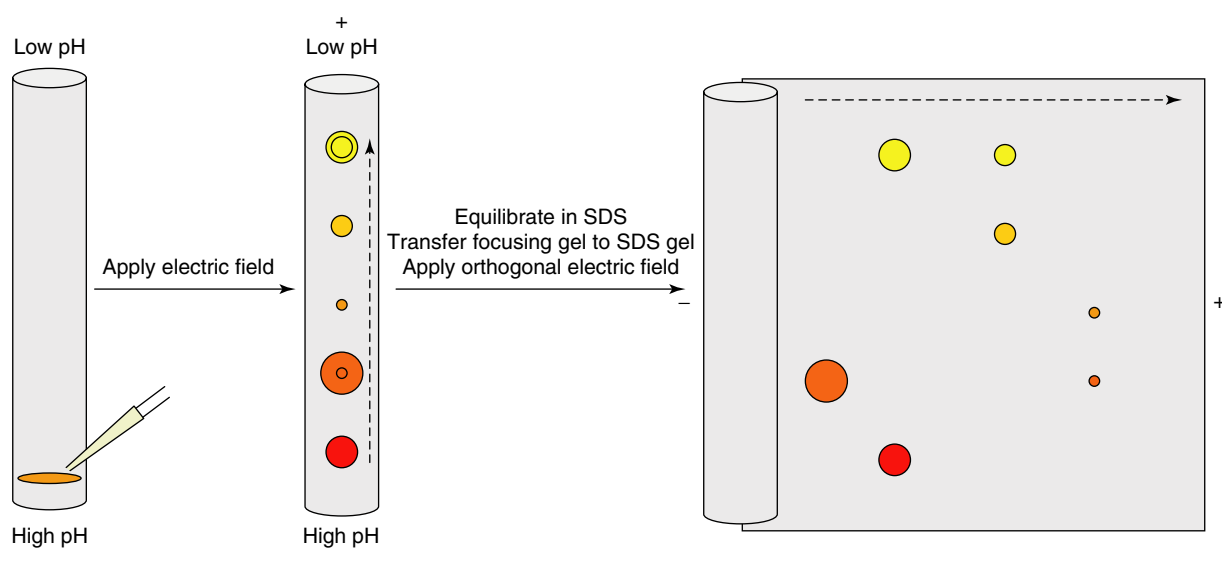
<i>Principle</i>	<i>Explanation</i>
The function of a protein depends on its structure and interactions, neither of which can be predicted accurately based on sequence information alone	Only by looking at the structure and interactions of the protein directly can definitive functional information be obtained
The abundance of a given transcript may not reflect the abundance of the corresponding protein	Messenger RNA analysis tells us the relative abundance of a transcript in the cell, and from this we infer the abundance of the corresponding protein. However, the two may not be related because of post-transcriptional gene regulation. Not all the mRNAs in the cell are translated, and rates of protein synthesis and protein turnover differ among transcripts
Protein diversity is generated post-transcriptionally	Many genes, particularly in eukaryote systems, give rise to multiple transcripts by alternative splicing. In some cases, the same transcript may give rise to multiple proteins whose individual functions cannot be studied other than at the protein level
Protein activity often depends on post-translational modifications, which are not predictable from the level of the corresponding transcript	Many proteins are present in the cell as inert molecules, which need to be activated by processes such as proteolytic cleavage or phosphorylation. In cases where variations in the abundance of a specific post-translational variant are significant, this means that only proteomics provides the information required in establishing a link between gene expression and function
The function of a protein often depends on its localization	While there are some examples of mRNA localization in the cell, particularly in early development, most trafficking of gene products occurs at the protein level. Many proteins are shuttled between compartments (e.g., the cytosol and the nucleus) as a form of regulation. In many cases, it is the distribution of a protein rather than its absolute abundance that is important
Some biological samples do not contain nucleic acids	One practical reason for studying the proteome is that many important samples do not contain nucleic acids. Most body fluids, including serum, cerebrospinal fluid, and urine, fall into this category
Proteins are the most therapeutically relevant molecules in the body	Although there has been recent development of drugs that target nucleic acids, most therapeutic targets are proteins and this is likely to remain so in the near future

nonselective fashion, but not all of these techniques are suitable for proteomics. One major requirement is high resolution. The separation technique should produce fractions that comprise very simple mixtures of proteins, and ideally each fraction should contain an individual protein. This essentially rules out one-dimensional techniques, i.e., those that exploit a single chemical or physical property as the basis for separation, since this simply does not provide enough resolving power. Proteomic techniques are therefore multidimensional, i.e., two or more different fractionation principles are employed one after another. The other major requirement in proteomics is high throughput. The separation technique should resolve all the proteins in one experiment and should ideally be easy to automate. The most suitable methods for automation are those that rely on 'differential rates of migration' to produce fractions that can be displayed or collected, a process generally described as separative transport. A final requirement is that the fractionation procedure should be compatible with downstream analysis by mass spectrometry (MS), as this is the major technology platform for high-throughput protein identification. The two groups of techniques that have come to dominate proteomics are two-dimensional gel electrophoresis (2DGE) and multidimensional liquid chromatography (LC).

## Two-Dimensional Gel Electrophoresis

In 2DGE, proteins are separated by two rounds of electrophoresis, in orthogonal directions, according to different separative principles (**Figure 1**). The first dimension separation in 2DGE is usually isoelectric focusing (IEF), in which proteins are separated based on their net charge irrespective of their mass. The underlying principle is that electrophoresis is carried out in a pH gradient, allowing each protein to migrate to its isoelectric point, i.e., the point at which its *pI* value is equivalent to the surrounding pH and its net charge is zero. Proteins with different *pI* values therefore focus at different positions in the pH gradient.

The pH gradient in an IEF gel can be established in two ways. The first is to use synthetic carrier ampholytes, which are collections of small amphoteric molecules with *pI* values corresponding to a given pH range. There are several problems with the use of ampholytes, and these lead to poor reproducibility in 2DGE experiments. One of the most serious limitations is cathodic drift, where the ampholytes themselves migrate to the cathode due to a phenomenon called electroosmotic flow (bulk solvent movement toward the cathode). This results in pH gradient instability, as basic ampholytes are progressively lost from the system. In practical terms, it is difficult to



**Figure 1** Two-dimensional electrophoresis using a tube gel for isoelectric focusing and a slab gel for SDS-PAGE. The proteins are separated in the first dimension on the basis of charge and in the second dimension on the basis of molecular mass. The circles represent proteins, with shading to indicate protein pI values and diameters representing molecular mass. The dotted line shows the direction of separation. (Reproduced with permission from Twyman RM (2004) *Principles of Proteomics*. Abingdon, UK: Bios/Garland Publishers.)

maintain a pH gradient that extends far beyond pH 7–8, resulting in the loss of many basic proteins.

These problems have been addressed by the development of immobilized pH gradient (IPG) gels, in which the buffering groups are attached to the polyacrylamide matrix of the gel. This is now the standard approach in proteomics, where reproducibility is a key issue. The IPG is established using immobilines, a collection of nonamphoteric molecules that contain a weak acid or base buffering group at one end, and an acrylic double bond to facilitate the immobilization reaction at the other. These chemicals are available from Amersham Pharmacia Biotech. The gel is run in the normal way but the pH gradient exists before the electric field is applied, and remains stable even when the gel is run for a long time. When the sample is loaded, the proteins migrate to their isoelectric points as in conventional IEF.

The second dimension separation in 2DGE is generally carried out by standard sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE), and separates the proteins according to molecular mass irrespective of charge. The basis of the technique is the exposure of denatured proteins to the detergent SDS, which binds stoichiometrically to the polypeptide backbone and carries a large negative charge. The presence of tens or hundreds of SDS molecules on each polypeptide dwarfs any intrinsic charge carried by the proteins themselves, and stoichiometric binding means that larger proteins bind

more SDS than smaller proteins. This has two important consequences that ensure separation on the basis of mass alone. First, all protein–SDS complexes have essentially the same charge density, and second, the relative differences in mass between proteins are maintained in the protein–SDS complexes. The gel enhances the size-dependent separation by sieving the proteins as they migrate.

Today's standard 2DGE systems, which are based on first dimension IEF using IPG strips followed by second dimension SDS-PAGE, are capable of resolving ~2500 protein spots on a routine basis. However, the proteome of a complex eukaryotic cell may be more than an order of magnitude greater than this. Even in the case of a simple eukaryotic system such as yeast, where alternative splicing and post-translational protein modifications are the exception rather than the rule, individual protein spots on standard 2D gels may comprise several different comigrating proteins, which can make downstream analysis very complex. The resolution of 2DGE depends on the separation length in both dimensions, and can thus be increased if very large format gels are used. For example, IEF tube gels and IPG strips > 30 cm in length have been used to achieve maximal separation in the first dimension, in combination with very large SDS gels that also provide a separation distance of > 30 cm. Although such gels can be difficult to handle, they do allow the separation of up to 10 000 protein spots. Another way to increase the

resolution of 2DGE is to use multiple IEF gels, each with a narrow pH range. These are known as zoom gels. Following second dimension SDS-PAGE and image analysis, the images of the separate zoom gels can be stitched together by computer to produce a composite of the entire proteome. Alternatively, to increase the resolution of proteins within a particular pH range, gels with nonlinear pH gradients can be produced. This is achieved simply by increasing the spacing between the appropriate immobilized reagents, and is often used to 'flatten' the pH gradient between pH 4 and 7, which accounts for the majority of proteins in the proteome. Finally, resolution can be increased by various forms of prefractionation prior to electrophoresis, to simplify the protein mixture that is being analyzed.

The data produced by 2DGE experiments are visual in nature, so downstream analysis involves capturing the images from stained 2D gels and then isolating particular spots for further processing and MS. This process is difficult to automate and therefore constitutes the most significant bottleneck in proteomic research. Until quite recently, manual analysis and spot picking from gels was very common. However, there are now various software packages available that produce high quality digitized gel images and incorporate methods to evaluate quantitative differences between spots on different gels. These can be integrated with spot excision robots that use plastic or steel picking tips to transfer gel slices to microtiter plates for automated digestion, cleanup, concentration, and transfer to the mass spectrometer. Several commercially available systems can fully automate the analysis and processing of 2D gels, and can handle 200–300 protein spots per hour.

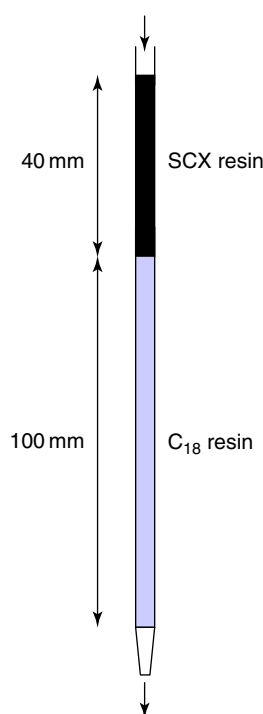
### Multidimensional Liquid Chromatography

In proteomics, LC has several advantages over 2DGE as a separation method, including its versatility, sensitivity, and the ease with which it can be automated and integrated with downstream analysis by MS. Unlike gel electrophoresis, LC is suitable for the separation of both proteins and peptides, and can therefore be applied either upstream of 2DGE to prefractionate the sample, downstream of 2DGE to separate the peptide mixtures from single excised spots, or instead of 2DGE as the separation technology of choice. Alternative LC methods can exploit different separation principles, such as size, charge, hydrophobicity, and affinity for particular ligands. As is the case for electrophoresis, the highest resolution separations are achieved when two or more separation principles are applied one after the other in orthogonal dimensions.

In LC methods used in proteomics, the stationary phase is a porous matrix, usually in the form of packed beads that are supported on some form of column. The mobile phase, a solvent containing dissolved proteins or peptides, flows through the column under gravity or is forced through under high pressure. The rate at which any particular protein or peptide flows through the column depends on its affinity for the matrix, and matrices with different chemical and physical properties can be used to separate proteins or peptides according to different selective principles, e.g., size exclusion, ion exchange, reversed phase, or affinity chromatography.

The sequential application of different chromatographic techniques exploiting different physical or chemical separative principles can provide sufficient resolution for the analysis of very complex protein or peptide mixtures. For example, the sequential use of ion-exchange chromatography (which separates proteins by charge) and reversed-phase high performance liquid chromatography (RP-HPLC; which separates proteins approximately in a mass-dependent fashion) can achieve the same resolution as 2DGE, with added advantages of automation, increased sensitivity and better representation of membrane proteins. The compatibility of RP-HPLC with the solvents used in matrix assisted laser desorption/ionization (MALDI)-MS and ESI-MS means that HPLC is almost universally used as the final separation method in multidimensional chromatography.

Initially, multidimensional chromatography was achieved by a discontinuous process in which fractions were collected from the ion exchange or gel filtration column and then manually injected into the HPLC column. However, the need for manual sample injection can be circumvented by equipping the first column with an automatic fraction collection system and a column-switching valve. Fractions are then collected from the first column across the elution range, and the switching valve can bring the RP-HPLC column in line to receive the fractions sequentially. Alternatively, some researchers have developed apparatus comprising a single ion-exchange column coupled, via an appropriate set of switching valves, to multiple HPLC columns arranged in parallel. In this scheme, fractions emerging from the first column are directed sequentially to the multiple HPLC columns, and the cycle is repeated when the first column has been regenerated. A third strategy for multidimensional chromatography separations is the use of biphasic columns, in which the distal part of the column is filled with reversed-phase resin and the proximal part with another type of matrix (Figure 2). As long as the elution solvents for each type of resin do not interfere with



**Figure 2** Continuous multidimensional chromatography using a biphasic column. In this example, simplified from the MudPIT method, a 140 mm  $\times$  0.1 mm fused silica capillary is packed at the distal end with 5  $\mu$ m C<sub>18</sub> (reversed-phase) particles and at the proximal end with 5  $\mu$ m strong cation exchange (SCX) particles. After introduction of the sample (top arrow) fractions are eluted from the SCX resin with a stepped salt gradient. After each salt step elution, the ion-exchange fraction flows into the reversed-phase material and is eluted using a gradient of acetonitrile (bottom arrow). Reversed-phase elution and reequilibration does not affect the SCX resin. This cycle is repeated until the SCX resin is exhausted.

each other, this allows the stepped elution of fractions from the first resin and the gradient elution of second dimension fractions from the second. This technique has been developed into a system called multidimensional protein identification technology (MudPIT).

## Protein Identification

The techniques described above allow protein mixtures to be separated into their components but do not allow those components to be identified. Indeed, the individual fractions produced by such methods are usually 'anonymous'. Each spot on a 2D gel and each fraction emerging from an HPLC column look very much like any other. In the case of 2DGE, even differences in spot size and distribution provide only vague clues about protein identity. The next stage in proteomic analysis is therefore to characterize the fractions and thus determine which proteins are actually present.

Proteins can often be characterized using probes – typically antibodies – that recognize unique structural features known as epitopes. This is a very powerful way to isolate and identify individual proteins but it is difficult to apply on a proteomic scale (see the section 'Protein chip technology' below). The gold standard method for identifying a protein is sequencing by Edman degradation, which involves the step-wise chemical removal of single amino acid residues, allowing short sequences to be determined that can be used to search sequence databases. While Edman degradation remains the most useful technique for N-terminal sequencing, it is slow and laborious, and many proteins are blocked to this technique because the N-terminal amino acid is chemically modified.

In the early 1990s, the identification of proteins was revolutionized by simultaneous development in two areas. First, in MS, techniques became available for the soft ionization of macromolecules, preventing the ions from fragmenting indiscriminately. The two techniques used most widely for ionization in proteomics today are:

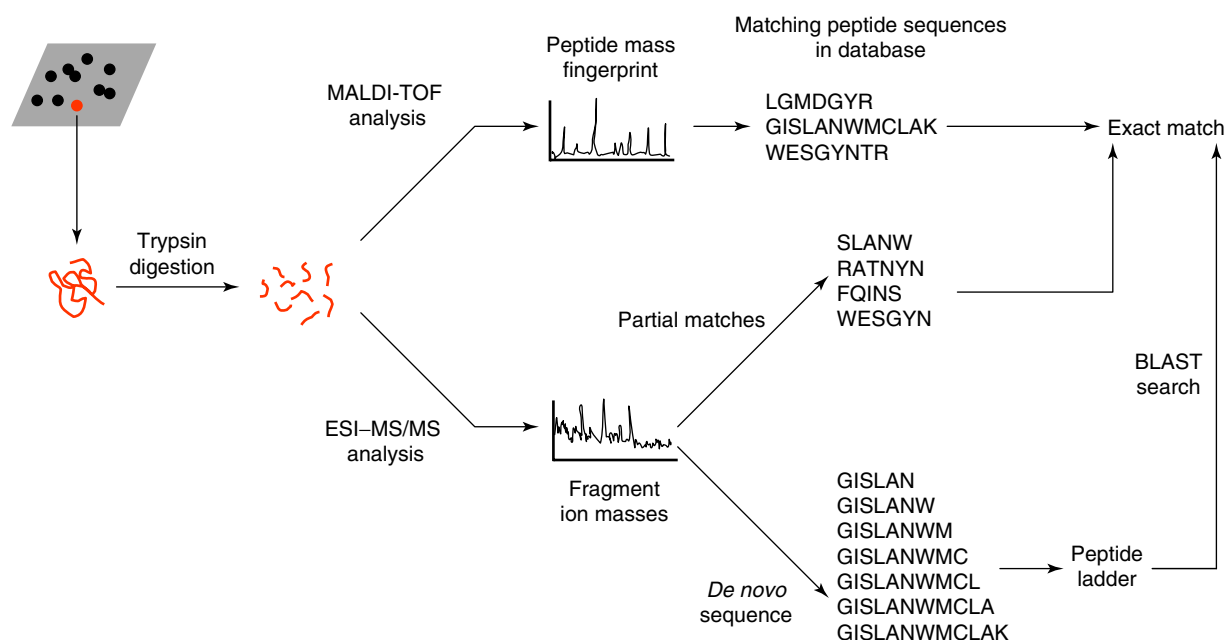
- **MALDI.** The analyte is mixed with a large excess of an aromatic matrix compound that can absorb energy from the laser used with the mass spectrometer. For example, the matrix compound  $\alpha$ -cyano-4-hydroxycinnamic acid can absorb the energy from a nitrogen UV laser (337 nm). The analyte and matrix are dissolved in an organic solvent and placed on a holder that can handle multiple samples. The solvent evaporates leaving matrix crystals in which the analyte is embedded. The holder is placed in the vacuum chamber of the mass spectrometer and a high voltage is applied. At the same time, the crystals are targeted with a short laser pulse. The laser energy is absorbed by the crystals and emitted (desorbed) as heat, resulting in rapid sublimation that converts the analyte into gas-phase ions. These accelerate away from the probe through the analyzer toward the detector. MALDI is used predominantly for the analysis of simple peptide mixtures, such as the peptides derived from a single spot from a 2D gel.
- **Electrospray ionization (ESI).** The analyte is dissolved and forced through a narrow needle held at a high voltage. A fine spray of charged droplets emerges from the needle and is directed into the vacuum chamber of the mass spectrometer through a small orifice. As they enter the mass spectrometer, the droplets are dried using a stream of inert gas, resulting in gas-phase ions that are accelerated through the analyzer toward the detector. Because ESI produces gas-phase ions from solution, it is readily integrated with upstream protein separation by liquid-phase methods, particularly capillary electrophoresis and

LC. Whereas MALDI-MS is used to analyze simple peptide mixtures, LC – ESI-MS is more suited to the analysis of complex samples.

These ionization techniques can be combined with a variety of instruments for the sensitive determination of peptide masses. Widely used instruments include triple quadrupole, time-of-flight (TOF), hybrid quadrupole-TOF, TOF-TOF, quadrupole-ion trap, and FT-ICR mass spectrometry.

The second development was the development of algorithms that could be used to search sequence databases with MS data. There are two general approaches, which are compared in Figure 3:

- The analysis of intact peptide ions. This allows the masses of intact peptides to be calculated, and these masses can be used to identify proteins in a sample by correlative database searching.
  - The analysis of fragmented ions. Intact peptide ions are fragmented randomly, generally by collision with a stream of inert gas (collision induced dissociation, CID). This allows the masses of peptide fragments to be determined, and the resulting CID spectrum can be used either for correlative database searching or to derive *de novo* sequences. In the latter case, the derived sequences can be used as standard queries in similarity search algorithms derived from BLAST and FASTA.
- The sample of interest should comprise a single protein or a simple mixture, e.g., an individual spot from a 2D gel or a single LC fraction. The sample is digested with a specific cleavage reagent, usually trypsin.
  - The masses of the peptides are determined, e.g., by MALDI-TOF-MS.
  - The experimenter chooses one or more protein sequence databases to be used for correlative searching.
  - The algorithm carries out a virtual digest of each protein in the sequence database using the same cleavage-specificity as trypsin (or whichever other reagent has been used experimentally) and then calculates theoretical peptide masses for each protein.
  - The algorithm attempts to correlate the theoretical peptide masses with the experimentally determined ones.
  - Proteins in the database are ranked in order of best correlation, usually with a significance threshold based on a minimum number of peptides matched.



**Figure 3** Protein identification by mass spectrometry. In a typical strategy, digested peptides are analyzed by MALDI-TOF-MS in order to determine the masses of intact peptides. These masses can be used in correlative database searches to identify exact matches. If this approach fails, ESI-MS/MS analysis can be used to generate peptide fragment ions. These can be used to search less robust data sources and to produce *de novo* peptide sequences. (Reproduced with permission from Twyman RM (2004) *Principles of Proteomics*. Abington, UK: Bios/Garland Publishers.)

There are several reasons why PMF may not work, including the absence of the sequence from the database, insufficient instrument sensitivity, nonspecific cleavage of the protein, the existence of several polymorphic variants of the protein with different masses, the presence of unanticipated post-translational modifications (either natural or artifactual) or the presence of contaminants. Where PMF fails to identify any proteins matching those present in a given sample, the CID spectrum of one or more individual peptides may provide important additional information. The data can be used in two ways. First, the 'uninterpreted' fragment ion masses can be used in correlative database searching to identify proteins whose peptides would likely yield similar CID spectra under the same fragmentation conditions. In probability based matching, virtual CID spectra are derived from the peptides of all protein sequences in the database and these are compared with the observed data to derive a list of potential matches. In cross-correlation, it is the degree of overlap between the observed and predicted peaks that determines the best potential match. Second, the peaks of the mass spectrum can be 'interpreted', either manually or automatically, to derive partial *de novo* peptide sequences that can be used as standard database search queries. The advantage of both these approaches is that correlative searching is not limited to databases of full protein sequences.

The interpretation of CID data is complex because fragmentation produces a diverse collection of thousands of different ions. The most informative fragments are those in which breakage has occurred along the polypeptide backbone, because these represent strings of contiguous and intact amino acids. If the charge remains on the N-terminal fragment, it is known as a b-series ion whereas if it remains on the C-terminal fragment, it is known as a y-series ion. Interpretation generally involves the arrangement of b- or y-series ions in order of increasing mass. The differences in mass between consecutive ions in either series should correspond to the masses of individual amino acids, and this can be used to derive a short sequence or peptide tag.

## Quantitative Proteomics and the Analysis of Post-Translational Variants

The objective in many proteomic experiments is to identify proteins whose abundance differs across two or more related samples. This may include variations in absolute protein levels or variations in the stoichiometry of different forms of modification, such as phosphorylation. At the current time, protein quantitation in proteomics relies primarily on the use of

general labeling or staining, or on the selective labeling or staining of particular classes of proteins. The chosen strategy depends largely on how the protein samples are prepared and fractionated, and can be divided into two broad categories: those based on the image analysis of 2D gels, and those based on differential labeling of samples for separation by LC followed by MS.

A range of stains and labels is available for use on 2D-gels and these differ in their sensitivity, specificity, and compatibility with MS. The most sensitive methods are silver staining, radioactive labeling, and fluorophore labeling, particularly the SYPRO reagents available from Molecular Probes Inc. (recently acquired by Invitrogen) which also have a broad linear dynamic range. Depending on the choice of label, the gel image is captured using a scanning device – a phosphorimager, a charge coupled device (CCD) camera, a densitometer or a fluorescence reader. Each spot is then converted into quantitative digital data, which requires sophisticated visual analysis software. Further software is required to recognize and compare corresponding spots on separate gels representing related samples (e.g., disease and healthy tissue), providing a spot-by-spot comparison across multiple gels. This approach is sometimes problematical because of the difficulties in reproducing running conditions accurately, and recently a number of multiplexed techniques have become available where several, differentially labeled samples can be run concurrently. Such techniques include difference gel electrophoresis, where individual samples are labeled with different fluorophores prior to electrophoresis, and the use of different stains to label different populations of proteins. The latter has been most useful for studying proteins with particular forms of post-translational modification (e.g., Pro-Q Diamond which selectively stains phosphoproteins, fluorescent labeled anti-phosphotyrosine antibodies, and Pro-Q Emerald which selectively stains glycoproteins) and proteins with different functional properties (e.g., the ability to bind certain drugs or substrates).

Protein quantitation by MS makes it possible to use in-line liquid-phase separation methods such as multidimensional chromatography and capillary electrophoresis. Quantitation can be carried out by comparing peptide ion currents but this is inherently inaccurate and is biased by instrument design. Instead, quantitation is often based on the use of stable isotopes. The general approach is to label alternative samples with equivalent reagents, one of which contains a heavy isotope and one of which contains a light isotope. The samples are mixed, separated into fractions, and analyzed by MS. The ratio of the two isotopic variants can be determined from the heights



of the peaks in the mass spectra and used to identify proteins with differential abundance.

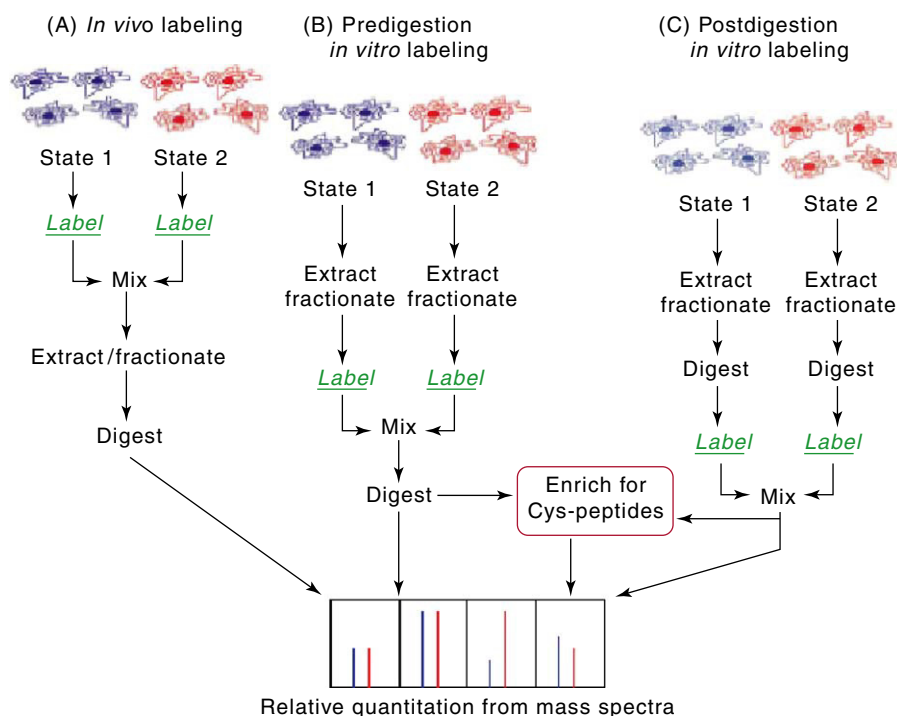
There are several different approaches to quantitative MS, including metabolic labeling *in vivo* (useful for cell assays), nonselective labeling of peptides with heavy oxygen during trypsin cleavage, and the use of isotope-coded affinity tags (Figure 4). These are biotinylated derivatives of iodoacetamide, a reagent that reacts with the cysteine side chains of denatured proteins. Two versions of the reagent are used, one normal or light form and one heavy or deuterated form in which a hydrogen atom is replaced by deuterium. The heavy and light forms are used to label different protein samples and then the proteins are combined and digested with trypsin. The biotin allows cysteine-containing peptides to be isolated from the complex peptide mixture through affinity to streptavidin, therefore considerably simplifying the number of different peptides entering the mass spectrometer.

Methods similar to ICAT can also be used for the LC-MS analysis of phosphoproteins and other variants. In this case, a reagent is chosen that selectively modifies the phosphate group. For example, ethanedithiol can be used to replace the phosphate group of phosphoserine and phosphothreonine residues by  $\beta$ -elimination under strongly alkaline conditions, leaving a thiol group that can be used to attach an isotope-coded biotin affinity tag. Phosphopeptides

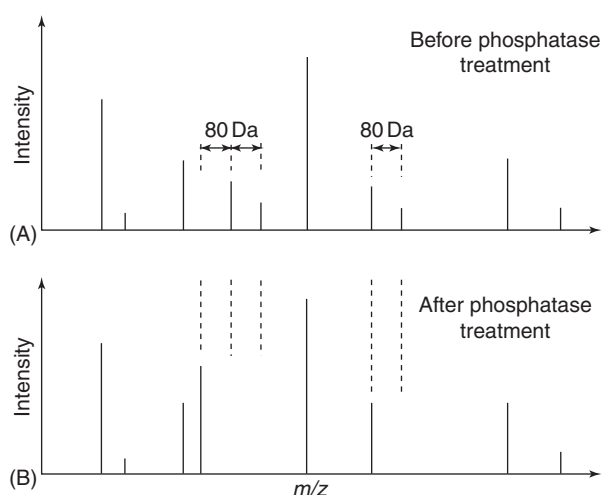
can be detected directly in tandem mass spectrometers using particular running modes. Standard MALDI-TOF analysis can identify ions with masses shifted by 79.983 compared to predicted masses, representing the additional phosphate group. Parallel analysis in which the sample has been treated with alkaline phosphatase, which removes phosphates, can also be helpful, since peaks corresponding to phosphopeptides in the untreated sample should be absent from the treated sample (Figure 5). In precursor ion scan mode, the first analyzer scans the ion stream and directs selected ions into the collision cell, while the second analyzer looks for diagnostic reporter ions, such as  $\text{H}_2\text{PO}_4^-$ ,  $\text{PO}_3^-$ , and  $\text{PO}_2^-$ , which have masses of  $\sim 97$ ,  $79$ , and  $63$ , respectively. In neutral loss scan mode, both analyzers are set to scan the ion stream. The first scans the full mass range, and the second scans a parallel range but at an  $m/z$  ratio that is  $98/z$  lower, with the intention of detecting the neutral loss of  $\text{H}_3\text{PO}_4$ .

## Other Proteomics Technologies

While 2DGE/LC and MS represent the major technologies for proteome analysis, proteomics is not limited to the analysis of protein abundance and expression. In the limited space available, a summary is provided of the other major technology platforms,



**Figure 4** Overview of MS-based strategies for quantitative proteomics. Depending on the point at which the label is introduced, most procedures are classified as: (A) *in vivo* labeling, (B) predigestion labeling *in vitro*, or (C) postdigestion labeling *in vitro*.



**Figure 5** Phosphopeptide identification by MALDI-TOF-MS mapping combined with alkaline phosphatase treatment. (A) The MALDI-TOF-MS spectrum of a proteolytic digest. Phosphopeptides are indicated by peaks shifted by multiples of 80 Da ( $\text{HPO}_3^- = 80 \text{ Da}$ ) relative to predicted unphosphorylated peptide masses. (B) The disappearance of such peaks upon treatment with a phosphatase confirms their identity as phosphopeptides.

which allow the investigation of protein sequences, structures, interactions, and biochemical functions.

### Sequence and Structural Proteomics

Although proteomics as we understand it today would not have been possible without advances in DNA sequencing, protein sequencing by Edman degradation for many years provided a crucial link between the activity of a protein and the genetic basis of a particular phenotype. It was not until the mid-1980s that it first became commonplace to predict protein sequences from genes rather than to use protein sequences for gene isolation.

The increasing numbers of stored protein and nucleic acid sequences, and the recognition that functionally related proteins often had similar sequences, catalyzed the development of statistical techniques for sequence comparison which underlie many of the core bioinformatic methods used in proteomics today. Nucleic acid sequences are stored in three primary sequence databases – GenBank, the EMBL nucleotide sequence database, and the DNA database of Japan (DDBJ) – which exchange data every day. These databases also contain protein sequences that have been translated from DNA sequences. A dedicated protein sequence database, SWISS-PROT, was founded in 1986 and contains highly curated data concerning over 70 000 proteins. A related database, TrEMBL, contains automatic translations of the nucleotide sequences in the EMBL database and is not manually curated.

Since similar sequences give rise to similar structures, it is clear that protein sequence, structure, and function are often intimately linked. The study of three-dimensional protein structure is underpinned by technologies such as X-ray crystallography and nuclear magnetic resonance spectroscopy, and has given rise to another branch of bioinformatics concerned with the storage, presentation, comparison, and prediction of structures. The Protein Data Bank was the first protein structure database and now contains more than 10 000 structures. Technological developments in structural proteomics have centered on increasing the throughput of structural determination and the initiation of systematic projects for proteome-wide structural analysis. A selection of proteomics databases is listed in Table 2.

### Interaction Proteomics

This branch of proteomics considers the genetic and physical interactions among proteins as well as interactions between proteins and nucleic acids or small molecules. The analysis of protein interactions can provide information not only about the function of individual proteins but also about how proteins function in pathways, networks, and complexes. It is a field that relies on many different technology platforms to provide diverse information, and is closely linked with functional proteomics and the large-scale analysis of protein localization. Conceptually the most ambitious aspect of interaction proteomics is the creation of proteome linkage maps based on binary interactions between individual proteins and higher order interactions determined by the systematic analysis of protein complexes. Key technologies in this area include the yeast two-hybrid system (a genetic assay for binary interactions) and MS for the analysis of protein complexes. Interactions between proteins and nucleic acids underlie many important processes including gene regulation, while protein interactions with small molecules are also of interest, e.g., enzymes interacting with their substrates and receptors with their ligands. These types of interactions are often investigated using biochemical assays and structural analysis methods such as X-ray crystallography. The characterization of protein interactions with small molecules can play an important role in the drug development process.

### Functional Proteomics

The most straightforward way to establish the function of a protein is to test that function directly. Functional proteomics is a relatively new development in which protein functions are tested directly but on a large scale. An example is the systematic

**Table 2** Important proteomics databases

Database	Contents	URL
<i>Primary sequence databases (raw and annotated sequence data)</i>		
GenBank <sup>a</sup>	Nucleotide and protein sequences	<a href="http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html">http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html</a>
EMBL <sup>a</sup>	Nucleotide and protein sequences	<a href="http://www.ebi.ac.uk/embl/">http://www.ebi.ac.uk/embl/</a>
DDBJ <sup>a</sup>	Nucleotide and protein sequences	<a href="http://www.ddbj.nig.ac.jp/">http://www.ddbj.nig.ac.jp/</a>
SWISS-PROT	Protein sequences and knowledgebase	<a href="http://us.expasy.org/sprot/">http://us.expasy.org/sprot/</a>
TrEMBL	Computer-annotated supplement to SWISS-PROT	<a href="http://us.expasy.org/sprot/">http://us.expasy.org/sprot/</a>
<i>Secondary sequence databases (information extracted and summarized from primary databases)</i>		
PROSITE	Sequence patterns associated with protein families and longer sequence profiles representing full protein domains	<a href="http://ca.expasy.org/prosite">http://ca.expasy.org/prosite</a>
PRINTS, BLOCKS	Highly conserved regions in multiple alignments of protein families. These are called motifs in PRINTS and blocks in BLOCKS	<a href="http://bioinf.man.ac.uk/dbbrowser/PRINTS">http://bioinf.man.ac.uk/dbbrowser/PRINTS</a> <a href="http://www.blocks.fhcrc.org">http://www.blocks.fhcrc.org</a>
Pfam, SMART, ProDom	Collections of protein domains	<a href="http://www.sanger.ac.uk/Software/Pfam">http://www.sanger.ac.uk/Software/Pfam</a> <a href="http://smart.embl-heidelberg.de/">http://smart.embl-heidelberg.de/</a> <a href="http://prodes.toulouse.inra.fr/prodom/current/html/home.php">http://prodes.toulouse.inra.fr/prodom/current/html/home.php</a>
Superfamily	HMM library and genome assignments	<a href="http://supfam.org/SUPERFAMILY/">http://supfam.org/SUPERFAMILY/</a>
PROT-FAM	Protein sequence homology database	<a href="http://www.mips.biochem.mpg.de/desc/protfam/">http://www.mips.biochem.mpg.de/desc/protfam/</a>
ProClass and iProclass	Protein classifications based on PROSITE patterns and PIR superfamilies	<a href="http://pir.georgetown.edu/iproclass/">http://pir.georgetown.edu/iproclass/</a> <a href="http://pir.georgetown.edu/gfserver/proclass.html">http://pir.georgetown.edu/gfserver/proclass.html</a>
ProtoMap	Automatic hierarchical classification of all SWISS-PROT and TrEMBL sequences	<a href="http://protomap.cornell.edu/">http://protomap.cornell.edu/</a>
SYSTEMS	Protein families database	<a href="http://systems.molgen.mpg.de/">http://systems.molgen.mpg.de/</a>
Interpro	A search facility that integrates the information from other secondary databases	<a href="http://www.ebi.ac.uk/interpro/">http://www.ebi.ac.uk/interpro/</a>
<i>Primary protein structural databases</i>		
PSD <sup>b</sup>	Gold standard database of three-dimensional macromolecular structures	<a href="http://www.rcsb.org/pdb/">http://www.rcsb.org/pdb/</a>
EBI <sup>b</sup>	Macromolecular structure database	<a href="http://www.ebi.ac.uk/msd/index.html">http://www.ebi.ac.uk/msd/index.html</a>
PDB <sup>b</sup>	Macromolecular structure database	<a href="http://www.pdbj.org/">http://www.pdbj.org/</a>
<i>Secondary protein structural databases (structural motifs and classifications)</i>		
CATH	Protein structure classification	<a href="http://www.biochem.ucl.ac.uk/bsm/cath_new/index.html">http://www.biochem.ucl.ac.uk/bsm/cath_new/index.html</a>
SCOP	Protein structure classification	<a href="http://scop.mrc-lmb.cam.ac.uk/scop/">http://scop.mrc-lmb.cam.ac.uk/scop/</a>
FSSP	Protein structure classification	<a href="http://www.bioinfo.biocenter.helsinki.fi:8080/dali/index.html">http://www.bioinfo.biocenter.helsinki.fi:8080/dali/index.html</a>
DALI	Dictionary of protein domains	<a href="http://www.ebi.ac.uk/dali/domain/">http://www.ebi.ac.uk/dali/domain/</a>

<sup>a</sup> The three primary sequence databases exchange and update data on a daily basis

<sup>b</sup> Now being integrated as the World-Wide Protein DataBank: <http://www.wwpdb.org/>

testing of expressed proteins for different enzymatic activities, as described in a landmark publication by Martzen and colleagues (see 'Further reading' section).

### Protein Chip Technology

Protein chips are miniature devices on which proteins, or reagents that capture proteins from solution, are applied in an array. There are many different types of protein chip, some used to analyze protein abundance and others used to study protein functions.

This emerging technology has the potential to considerably improve the throughput of protein analysis, particularly with the recent advent of a whole proteome chip for the yeast *Saccharomyces cerevisiae*. The various different types of protein chip that have been described are summarized below:

- Antibody chips. These consist of arrayed antibodies and are used to detect and quantify specific proteins in a complex mixture. They can be thought of as miniaturized high-throughput immunoassay devices.

- **Antigen chips.** The converse of antibody chips, these devices contain arrayed protein antigens and are used to detect and quantify antibodies in a complex mixture.
- **Universal protein chips (functional arrays).** These devices may contain any kind of protein arrayed on the surface and can be used to detect and characterize specific protein–protein and protein–ligand interactions. Various detection methods may be used, including labeling the proteins in solution or detecting changes in the surface properties of the chip, e.g., by surface plasmon resonance. Included within this category are lectin arrays, which are used to detect and characterize glycoproteins.
- **Protein capture chips.** These devices do not contain arrayed proteins, but other molecules that interact with proteins as broad or specific capture agents. Examples include oligonucleotide aptamers and chips containing molecular imprinted polymers as specific capture agents, or the proprietary protein chips produced by companies such as BIAcore Inc. and CIPHERGEN Biosystems Inc., which employ broad capture agents based on differing surface chemistries to simplify complex protein mixtures.
- **Solution arrays.** The latest generation of protein chips is being released from the 2D array format to increase their flexibility and handling capacity. Such devices may, for example, be based on coded microspheres or barcoded gold nanoparticles.

See also: **Electrophoresis:** Isoelectric Focusing; Polyacrylamide Gels; Two-Dimensional Gels; Proteins. **Liquid Chromatography:** Liquid Chromatography–Mass Spectrometry; Biotechnology Applications. **Mass Spectrometry:** Ionization Methods Overview; Electrospray; Matrix-Assisted Laser Desorption/Ionization; Multidimensional; Peptides and Proteins. **Proteins:** Traditional Methods of Sequence Determination.

## Further Reading

- Aebersold R and Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422: 198–207.
- Drewes G and Bouwmeester T (2003) Global approaches to protein–protein interactions. *Current Opinion in Cell Biology* 15: 1–7.
- Kalume DE, Molina H, and Pandey A (2003) Tackling the phosphoproteome: Tools and strategies. *Current Opinion in Chemical Biology* 7: 64–69.
- Mann M, Hendrickson RC, and Pandey A (2001) Analysis of proteins and proteomes by mass spectrometry. *Annual Review of Biochemistry* 70: 437–473.
- Mann M and Jensen ON (2003) Proteomic analysis of post-translational modifications. *Nature Biotechnology* 21: 255–261.
- Martzen MR, McCraith SM, Spinelli SL, *et al.* (1999) A biochemical genomics approach for identifying genes by the activity of their products. *Science* 286: 1153–1155.
- Patton WF and Beecham JM (2001) Rainbow's end: The quest for multiplexed fluorescence quantitative analysis in proteomics. *Current Opinion in Chemical Biology* 6: 63–69.
- Phizicky E, Bastiaens PIH, Zhu H, Snyder M, and Fields S (2003) Protein analysis on a proteomic scale. *Nature* 422: 208–215.
- Sali A, Glaeser R, Earnest T, and Baumeister W (2003) From words to literature in structural proteomics. *Nature* 422: 216–225.
- Sechi S and Oda Y (2003) Quantitative proteomics using mass spectrometry. *Current Opinion in Chemical Biology* 7: 70–77.
- Tyers M and Mann M (2003) From genomics to proteomics. *Nature* 422: 193–197.
- Wang H and Hanash S (2003) Multi-dimensional liquid phase based separations in proteomics. *Journal of Chromatography B* 787: 11–18.
- Zhu H, Bilgin M, and Snyder M (2003) Proteomics. *Annual Review of Biochemistry* 72: 783–812.
- Zhu H and Snyder M (2003) Protein chip technology. *Current Opinion in Chemical Biology* 7: 55–63.

## PSA

See **POTENTIOMETRIC STRIPPING ANALYSIS**

## PURGE AND TRAP

See **HEADSPACE ANALYSIS: Purge and Trap**

# PURINES, PYRIMIDINES, AND NUCLEOTIDES

**J-J Aaron**, Université Paris 7-Denis Diderot,  
Paris, France

**S Trajkovska**, University St Kiril i Metodij, Skopje,  
Macedonia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Purines, pyrimidines, nucleosides, and nucleotides belong to a biologically important class of compounds in which several of these derivatives constitute nucleic acid components (typical structures of these compounds are given in **Figure 1**). Purines and pyrimidines are azaheterocyclic bases. Purines are constituted of two nitrogen-containing rings (**Figure 1**, structures (1)–(3)), whereas pyrimidines contain only one nitrogen-containing ring (**Figure 1**, structures (4)–(6)). When a purine or a pyrimidine is attached to the C-1 position of a sugar, the corresponding structure is called a nucleoside (**Figure 1**, structures (7) and (8)), whereas a heterocyclic base-sugar-phosphoric acid unit is called a nucleotide (**Figure 1**, structures (9) and (10)). A dinucleotide consists of two nucleotide units connected by phosphate groups (**Figure 1**, structure (11)). These compounds are of major biological and biochemical interest since the backbone of nucleic acid molecules (DNA, RNA) is a polynucleotide chain, made up of a large number of various nucleotide units joined by phosphate groups.

The names of nucleosides are derived from the names of the corresponding heterocyclic bases combined with the name of the sugar (ribose or deoxyribose). In some cases, special names are used: for example, cytidine is cytosine- $\beta$ -D-ribose.

Nucleotides, which contain one, two, or three phosphate groups, are named as mono-, di-, or triphosphate derivatives of the nucleosides: for example, cytidine 5'-monophosphate (CMP), 5'-diphosphate, and 5'-triphosphate are the nucleotides derived from cytidine.

Many purine and pyrimidine derivatives, including mercapto-, fluoro-, and azasubstituents, are contained in pharmaceuticals, and therefore their metabolism has to be monitored in living organisms. For these reasons, their determination has been the object of much attention during the last 25 years, and a number of methods have been proposed. These procedures have been applied widely to the determination of purines, pyrimidines, and their derivatives in plant and animal tissues as well as in various biological fluids such as blood plasma, whole blood,

and urine. This article briefly reviews the various analytical methods used for determining purines, pyrimidines, and their derivatives, as well as the special methods for nucleotides. Performances and analytical figures of merit are compared. Practical applications for the determination of these compounds in biological fluids and tissues are also discussed.

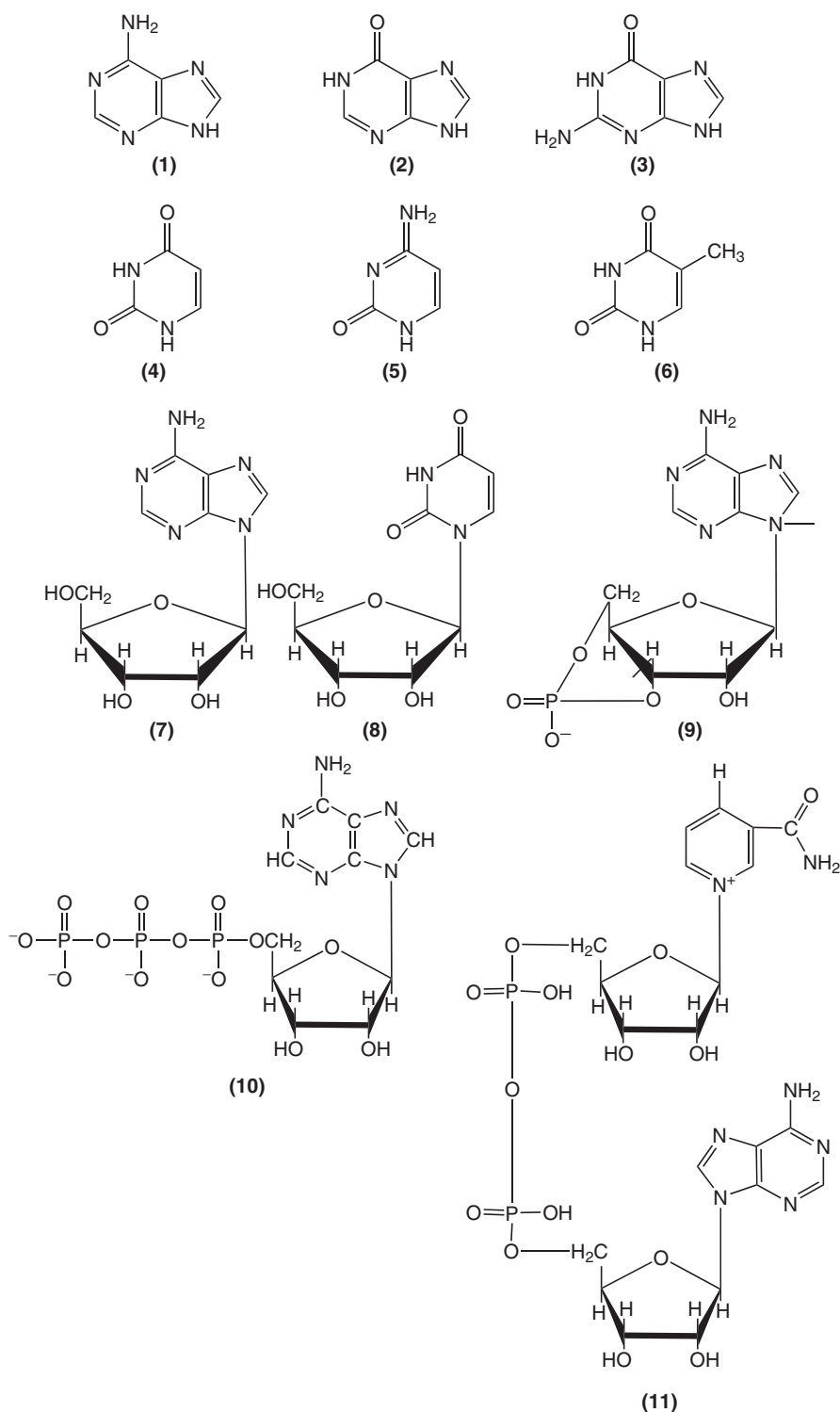
## Chromatographic Methods

Because biological samples contain generally a relatively large number of purines, pyrimidines, nucleotides, and/or nucleosides, it is necessary to use separation methods to simultaneously analyze multi-component mixtures of these derivatives. Several chromatographic methods have been reported such as paper chromatography (PC), thin-layer chromatography (TLC), gas chromatography, and liquid chromatography (LC). The performances of PC, TLC, and LC are described in detail below.

### Paper Chromatography

Although paper chromatography is often considered outmoded, it remains one of the most widely used techniques in analyzing mixtures of purine and pyrimidine derivatives, including nucleosides and nucleotides. These compounds generally absorb ultraviolet (UV) radiation in the 250–300 nm range. Therefore, they can be easily detected on chromatograms by irradiation with a UV lamp emitting in this wavelength region. Under these conditions, dark, UV-absorbing spots are observed against the natural fluorescence of the chromatogram paper, indicating the location of the derivatives. When elution is performed carefully, PC makes it possible to achieve quantitative recovery of purine and pyrimidine derivatives from chromatograms.

Several solvent systems have been proposed for PC such as butan-1-ol/acetic acid/water, butan-1-ol/ammonia/water, butanol-1-ol/water, ethanol/ammonium acetate, and methanol/formic acid/water. Generally, the separation of a complex mixture of bases, nucleosides, and nucleotides requires using successively two solvent systems for two-dimensional PC. Since  $R_F$  values may vary slightly with chromatographic conditions, it is recommended that identifications be checked by comparing the  $R_F$  values of the mixture components with those of authentic samples in several solvent systems.



**Figure 1** Examples of structures of purines: (1) adenine; (2) hypoxanthine; (3) guanine (G). Pyrimidines: (4) uracil; (5) cytosine (C); (6) thymine (T). Nucleosides: (7) adenosine (A); (8) uridine (U). Nucleotides: (9) 3',5'-cAMP; (10) adenosine 5'-triphosphate. Dinucleotide: (11) NAD<sup>+</sup>.

### Thin-Layer Chromatography

Because of its great versatility and simplicity, TLC constitutes an important analytical tool for the

separation and determination of purine and pyrimidine derivatives. The bases and nucleosides and nucleotides can be separated on thin layers of silica gel G, unmodified cellulose, chemically modified



cellulose, and aluminum oxide. The choice of solvent mixture depends on the specific compounds under study and on the type of thin layers used. Typically, 2-methylbutan-2-ol/formic acid/water (3:2:1, v/v) gives a good overall separation of nucleotides, while butan-1-ol/acetone/acetic acid/ammonia (5% v/v)/water (45:15:10:10:20) separates nucleoside mono-, di-, and triphosphates on unmodified cellulose. A high acetonitrile content seems to increase the rapidity of elution of nucleosides and nucleotides, as compared with alcohol-containing mixtures, when silica gel and microcrystalline cellulose are utilized.

Compounds separated by TLC on cellulose or silica gel can be detected by visualizing the spots in UV radiation from a suitable wavelength-emitting source (200–250 nm). An alternative, and generally more sensitive, detection method can be used on silica gel by spraying the developed chromatogram with a solution of a fluorescent compound, such as 2',7'-dichlorofluorescein (0.01% w/v in ethanol with a few drops of ammonia) or uranyl acetate (1% w/v in water). Limits of detection using TLC for purines range between ~0.1 µg per spot (UV detection) and 10 ng per spot (fluorescence detection).

### Liquid Chromatography

Liquid chromatography is a rapid, selective and sensitive technique which can be adapted to the analysis of extracts from a variety of biological sources. It constitutes an excellent method for the analysis of nucleotides, nucleosides, purines, and pyrimidines bases.

**Analysis of nucleotide mixtures in tissues** Nucleotide mixtures are found in human, animal, and plant tissues, and it is essential to select convenient extraction procedures in order to minimize chromatographic artifact formation. Also, preliminary purification is needed to remove interfering natural substances that can bind irreversibly to the chromatographic column packing, producing loss of resolution.

The extraction methods for free nucleotides from human, animal, and plant tissues are generally based on the use of cold acid extractants, which also have the role of suppressing enzymatic activity and precipitating proteins. Usually, biological samples are disrupted in the presence of an extractant, and the extracts are centrifuged or filtered. Dilute solutions of perchloric acid or trichloroacetic acid (TCA) are utilized as extractants in most cases. However, caution is recommended in the use of TCA preceding LC because it can produce chromatographic artifacts affecting the quantification of nucleotides. Formic

acid-based solutions have also been proposed for the extraction of nucleotides from microorganisms and plant tissues.

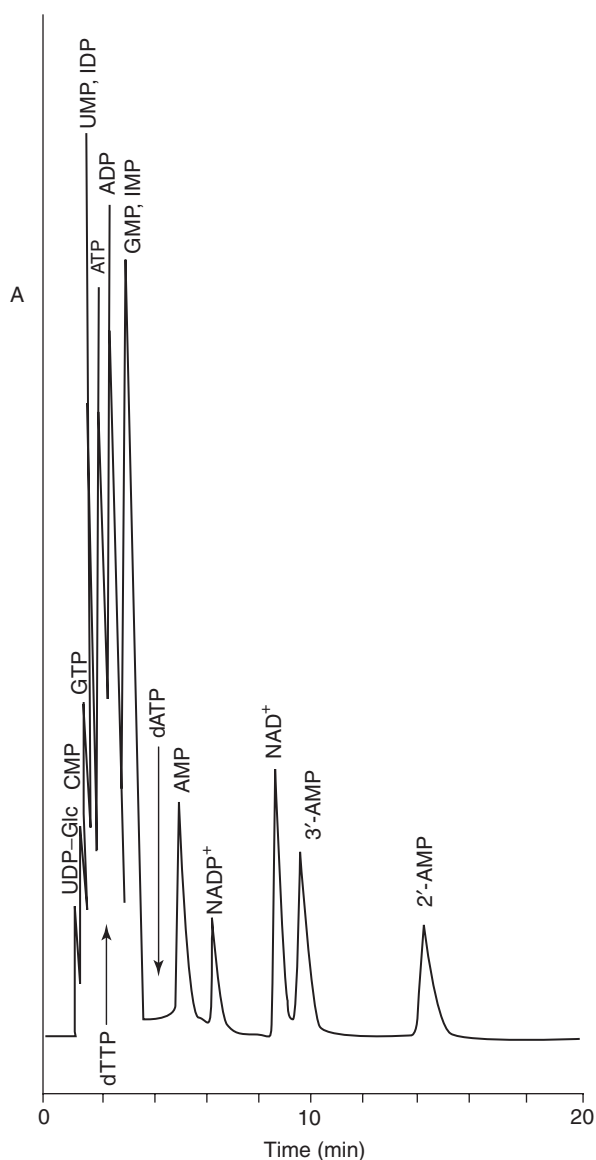
The purification procedure consists of several steps. First, the perchloric acid extract is centrifuged (2000–16 000 g, for 5–10 min), and the supernatant is neutralized with potassium hydroxide solution. Then the pH of the aqueous solution containing the nucleotides is adjusted to an appropriate value (between 5.0 and 7.0), and the precipitated potassium perchlorate, which has a low solubility in cold water, is removed by centrifugation. In the majority of human and animal tissues, and of microbial cells, it is convenient to chromatograph this extract. In some cases, however, especially where a large number of constituents have to be resolved in a single chromatographic step, it may be necessary to further purify the extract by a chemical reaction or by a preliminary chromatographic separation.

Liquid chromatographic analysis of nucleotide mixtures is based on several separation techniques such as ion exchange, reversed-phase, and reversed-phase ion-pair chromatography. The use of microparticulate, chemically bonded anion exchange allows the separation, on a single column, of a mixture of all the commonly occurring nucleotides, with a linear gradient elution system, as shown in **Figure 2**. Isocratic reversed-phase LC systems have also been proposed for the separation of a relatively large number of nucleotides, including ribonucleotides, deoxynucleotides, cyclic nucleotides, and deoxycyclic nucleotides with good speed and resolution.

**Determination of purines, pyrimidines, and nucleosides** Although several LC procedures are available for separating mixtures of purines, pyrimidines, and nucleosides, it is generally more convenient to use specific columns for differentiating between these different classes of compounds.

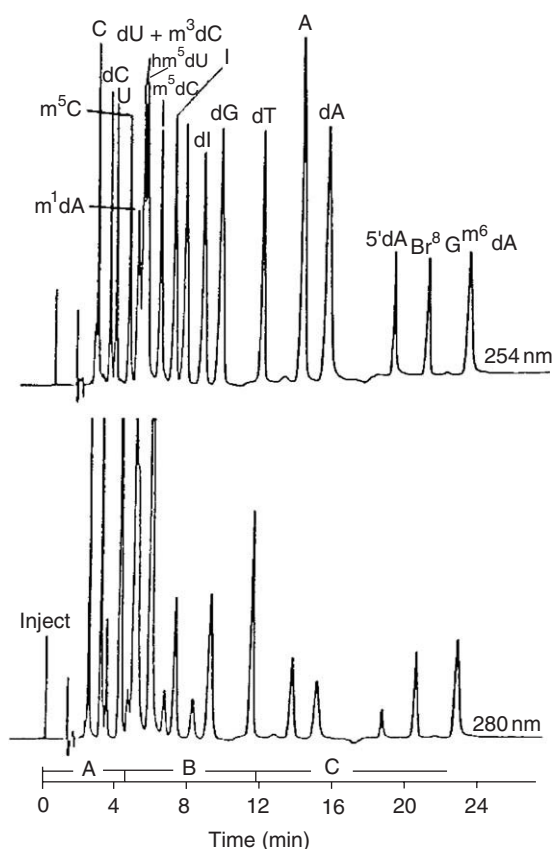
Liquid chromatographic analysis of free purine and pyrimidine bases can be carried out on either cation-exchange or reversed-phase columns. The separation of a mixture of 10 methylated xanthines has been accomplished using reversed-phase LC on a column of Apex Octadecyl 3 µm kept at 50°C, in less than 2 min. This mixture included caffeine, theophylline, and theobromine, which are very important compounds in the food and pharmaceutical industries.

The separation of nucleoside mixtures generally requires a reversed-phase system with a stepwise gradient elution, using phosphate buffers and increasing concentrations of methanol. Under these conditions, 20 nucleosides can be separated in less than 24 min (**Figure 3**).



**Figure 2** Reversed-phase LC separation on  $\mu$  Bondapak C<sub>18</sub> of a reference mixture of commonly occurring free-nucleotides. A linear gradient (30 min) was used with  $0.2 \text{ mol l}^{-1} \text{ KH}_2\text{PO}_4$  (pH 4.0) as the initial eluant, and  $0.2 \text{ mol l}^{-1} \text{ KH}_2\text{PO}_4$  containing 20% (v/v) methanol, also at pH 4. Flow rate  $2 \text{ ml min}^{-1}$ . The positions of 2'-deoxyadenosine 5'-triphosphate (dATP) and 2'-deoxythymidine-5'-triphosphate (dTTP), determined from other separations on the same column, are indicated by arrows (Reprinted with permission from Brown EG, Newton RP, and Shaw NM (1982) Analysis of the free nucleotide pools of mammalian tissues by high-pressure liquid chromatography. *Analytical Biochemistry* 123: 378–388; © Elsevier.)

Ultraviolet spectrophotometric, spectrofluorimetric, and electrochemical detectors are utilized for most compounds. Purines such as hypoxanthine and xanthine can be determined fluorimetrically by means of a reaction using xanthine oxidase and producing hydrogen peroxide. Generally, the LC



**Figure 3** Reversed-phase separation of a mixture of ribo- and deoxyribonucleosides on a Supelcosil LC 18-DB column using a ternary buffer stepped gradient as eluant. Column temperature  $35^\circ\text{C}$ . Flow-rate  $2.0 \text{ ml min}^{-1}$ . The buffers were all  $0.05 \text{ mol l}^{-1} \text{ KH}_2\text{PO}_4$  (pH 4.00) with increasing concentrations of methanol. Buffer A, 2.5% (v/v) methanol; buffer B, 5% (v/v) methanol; buffer C, 12.5% (v/v) methanol. Abbreviations: m'dA, 1-methyldeoxyadenosine; m<sup>5</sup>C, 5-methylcytidine; C, cytidine; dC, deoxycytidine; U, uridine; dU, deoxyuridine; m<sup>3</sup>dC, 3-methyldeoxycytidine; hm<sup>5</sup>dU, 5-hydroxymethyldeoxyuridine; m<sup>5</sup>dU, 5-methyldeoxyuridine; I, inosine; G, guanosine; dI, deoxyinosine; dG, deoxyguanosine; 5'dA, 5'-deoxyadenosine; Br<sup>8</sup>G, 8-bromoguanosine; m<sup>6</sup>dA, 6-methyldeoxyadenosine. (Reprinted with permission from Gehrke CW, McCune RA, Gama-Sosa MA, Emrich M, and Kuo KC (1984) Quantitative reversed-phase liquid chromatography of major and modified nucleosides in DNA. *Journal of Chromatography* 301: 199–219; © Elsevier.)

detection limits for purines, pyrimidines, and nucleosides are rather low, in the nanogram range.

## Electrophoresis

Electrophoresis has emerged recently as a promising method for the determination of purines, pyrimidines, nucleosides, and nucleotides. Two techniques are used: high-voltage paper electrophoresis (PE) and capillary electrophoresis (CE).

### High-Voltage Paper Electrophoresis

High-voltage PE is a separation technique based on the differences in the electrophoretic mobilities of charged compounds. Commercially available apparatus consist of a high-voltage (1–10 kV) supply that is connected across the opposite ends of an electrophoresis paper. Application of a convenient electric potential induces a current to flow along the sheet of paper, producing a migration of ions. Under certain conditions of pH and current, the extent of migration of the components of a mixture is related to the charges located on the individual chemical species. Therefore, the number of ionizable groups and their  $pK_a$  values are critical for determining the electrophoretic behavior. In the case of purines, pyrimidines, and nucleosides, the  $pK_a$  values of groups such as amino and keto groups are important for electrophoretic separation. Nucleotides can be separated from the other compounds on the basis of the  $pK_a$  values of their phosphoric acid groups. As would be expected from their difference in negative charge, nucleoside mono-, di-, and triphosphates migrate at different speeds toward the anode, and they can be separated easily one from another.

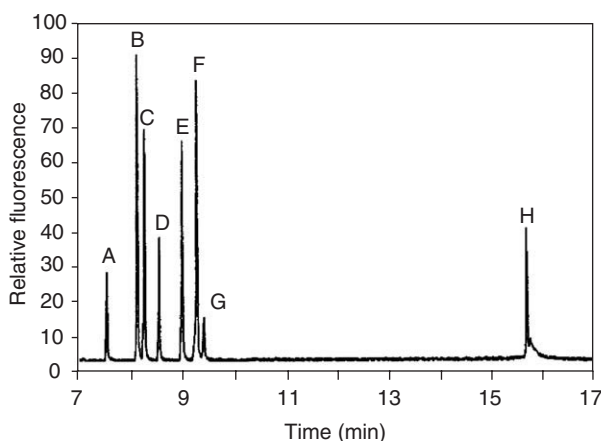
Volatile 'buffers' with a pH ranging from 2.0 to 9.0 are generally utilized in PE. Typically, a formic acid/acetic acid/water mixture at pH 2.0 is used for nucleotides. Noncyclic 5'-nucleotides can be separated from cyclic 3',5'- nucleotides, which are simultaneously present in animal, plant, and microbial tissues, by applying a nonvolatile sodium borate buffer at pH 9.3; indeed borate interacts specifically with the *cis-vicinal* hydroxyl groups of 5'-nucleotide ribose but not with cyclic nucleotides.

### Capillary Electrophoresis

In recent years, CE has developed into a very sensitive and selective analytical technique for the determination of purines, pyrimidines and their metabolites, and nucleotides. It allows charged compounds to be separated in a capillary filled with the electrophoresis buffer through application of a high voltage (typically, 10–30 kV) over the length of the fused silica capillary tube. The electrophoretic migration rate of charged compounds, which move toward the electrode with opposite charge, depends on the charge and the size of the molecules and is proportional to the electrical field strength. Generally, the efficiencies obtained in CE are much greater than those in other separation methods. However, a drawback of CE is that the mass flow is limited owing to the small dimensions of the capillary used, resulting in flow rates of only 5–15  $\mu\text{L min}^{-1}$ .

The separation of mixtures of purines, pyrimidines or nucleotides can take place in 15–25 min, depending on the molecular structure, charge, and experimental conditions. In **Figure 4**, an electropherogram is presented, showing that the separation of a mixture of eight nucleotide monophosphates (prepared from LC-purified samples) is complete in approximately 16 min. The method involves micellar electrokinetic capillary chromatography (MECC), in which a surfactant at concentrations greater than the critical micellar concentration is added to the buffer solutions.

Several detection systems are utilized in CE for the analysis of nucleoside and nucleotide mixtures. The performances of UV-visible absorption, conductance, electrochemical,  $\alpha$ - $^{32}\text{P}$  radiochemical and fluorescence detectors and mass spectral interfacing have been compared recently. Although UV-visible absorption is generally considered as not very sensitive, low limits of detection (LODs) of  $8 \times 10^{-8} \text{ mol L}^{-1}$  have been reported for purine metabolites using this method. The conductivity technique suffers from poor sensitivity. Electrochemical detection has a higher sensitivity, but its usefulness is limited by the fact that only electroactive species can be detected. Detection by mass spectrometry (MS) leads to poor sensitivity and implies expensive instrumentation. Radiochemical detection has been applied to  $\alpha$ - $^{32}\text{P}$ -labeled thymidine, cytidine, and adenosine



**Figure 4** MECC-laser induced fluorescence analysis of deprotonated dansyl hydroxide (A), dansyl-dTMP (B), dansyl-dCMP (C), dansyl-5 methyl dCMP (D), dansyl-dAMP (E), dansyl-dGMP (F), dansyl-8-OH dGMP (G), and dansyl 8-(*N*-2-acetylaminofluorene) dGMP (H). (Reprinted with permission from Lee T, Yeung ES, and Sharma M (1991) Micellar electrokinetic capillary chromatographic separation and laser-induced fluorescence detection of 2'-deoxynucleoside-5'-monophosphates of normal and modified bases. *Journal of Chromatography and Biomedical Applications* 565: 197–206; © Elsevier.)

**Table 1** Comparison of state-of-the-art LODs<sup>a</sup> with LC and CE in the analysis of various nucleosides and nucleotides

Detection system	LC		CE	
	Amount (mol)	Concentration (mol l <sup>-1</sup> )	Amount (mol)	Concentration (mol l <sup>-1</sup> )
Absorption	$2 \times 10^{-9b}$	$2 \times 10^{-4b}$	$7 \times 10^{-14b}$	$3 \times 10^{-5b}$
Electrochemical	$2 \times 10^{-13c}$	$2 \times 10^{-8c}$	—	—
Fluorescence	$2 \times 10^{-13d}$	$2 \times 10^{-8d}$	$6 \times 10^{-18d}$	$1 \times 10^{-9d}$
Radiochemical	—	—	$72 \times 10^{-18e}$	$1 \times 10^{-10e}$

<sup>a</sup>The LOD is defined as the amount (mol) or concentration (mol l<sup>-1</sup>) of analyte giving a signal-to-noise ratio of 3.

<sup>b</sup>For dansyl-2'-deoxyadenosine-5'-monophosphate (dAMP), dansyl-2'-deoxycytidine-5'-monophosphate (dCMP), dansyl-2'-deoxyguanine-5'-monophosphate (dGMP), and dansyl-2'-deoxythymidine-5'-monophosphate (dTMP).

<sup>c</sup>For 8-hydroxy-2-deoxyguanine.

<sup>d</sup>For dAMP, dCMP, dGMP, dTMP, and corresponding nucleosides-5',-mono-, di- and triphosphates.

<sup>e</sup>For adenosine-5'-triphosphate, cytidine-5'-triphosphate, and thymidine-5'-triphosphate.

(Reprinted with permission from Lee T, Yeung ES, and Sharma M (1991) Micellar electrokinetic capillary chromatographic separation and laser-induced fluorescence detection of 2'-deoxynucleoside-5'-monophosphates of normal and modified bases. *Journal of Chromatography and Biomedical Applications* 565: 197–206; © Elsevier.)

triphosphates, resulting in very low LODs of  $\sim 1 \times 10^{-10}$  mol l<sup>-1</sup>. Laser-induced fluorescence is also a very sensitive technique, with LODs in the  $10^{-9}$  mol l<sup>-1</sup> range for nucleoside monophosphates. Table 1 shows state-of-the-art LODs for several detection systems coupled to CE or LC in the analysis of derivatized or native nucleosides. It can be seen that the absolute LODs (expressed in moles) for CE are generally five orders of magnitude lower than those for LC, using the same detection system. In contrast, the concentration LODs (in mol l<sup>-1</sup>) have only a difference of one order of magnitude. This is mainly due to the smaller sample volumes used in CE as compared with LC.

## Spectrometric Methods

Spectrometric methods are extremely useful for structural investigation and identification of purines, pyrimidines, nucleosides, and nucleotides. They allow the provisional identification of these compounds, made on the basis of chromatographic and electrophoretic behavior, to be confirmed. The application of MS, UV absorption spectrophotometry, and phosphorescence spectrometry to the identification and determination of purines, pyrimidines, and their derivatives is reviewed in the following sections.

### Mass Spectrometry (MS)

Although the development of conventional mass spectrometric techniques has been relatively slow in the area of purines and pyrimidines, primarily because of the low volatility of these compounds and hence the difficulty of achieving ionization, recent advances such as fast atom bombardment (FAB) and collision-induced dissociation (CID) have opened up this field.

FAB is a method of relatively soft ionization that has been applied to many volatile and thermally labile, biologically important compounds over a large mass range. Although FAB–MS provides information on the molecular mass of cyclic nucleotides, it does not allow detailed structural identification. For instance, isomers often provide similar spectra. For example, the FAB mass spectra of the two isomers 3',5'- and 2',3'-cyclic adenosine monophosphate are practically identical. The employment of CID of ions results in mass-analyzed kinetic energy spectra (MIKES), which provide the means of differentiating these cyclic nucleotide isomers.

Therefore the use of FAB–MS combined with CID–MIKES can be applied to the qualitative analysis of cyclic nucleotides in mammalian and higher plant tissue extracts and to the identification of synthetic cyclic nucleotide derivatives. It provides unambiguous evidence of the identity of nucleotides, in contrast with chromatographic, enzymatic, and radioimmunoassay methods, which often lead to equivocal results.

Quantitative analysis of cyclic nucleotides in protein kinase activity can also be performed using the FAB–MIKES technique. It gives results comparable with those obtained by conventional radiometric assay for monitoring the kinetics of cyclic nucleotide phosphodiesterase activity.

The potential of laser microprobe mass spectrometry (LMMS) has been investigated for structural characterization of nucleosides and nucleotides. This technique is based on the measurement of ions formed promptly by direct desorption and ionization (DI) of solid microscopic samples. The DI process is very fast for nucleosides, which makes it possible to apply a relatively high laser energy to the sample. In the case of nucleosides, LMMS gives

satisfactory results based on the mass spectral peaks of positive and negative ions. The molecular mass of these compounds can easily be determined from the relatively intense signal. In contrast, LMMS analysis of nucleotides is more difficult. The sensitivity is lower, and the DI process is slower compared with nucleosides. More work is needed for optimizing the experimental conditions of analysis of nucleotide sodium salts.

### UV Absorption Spectrophotometry

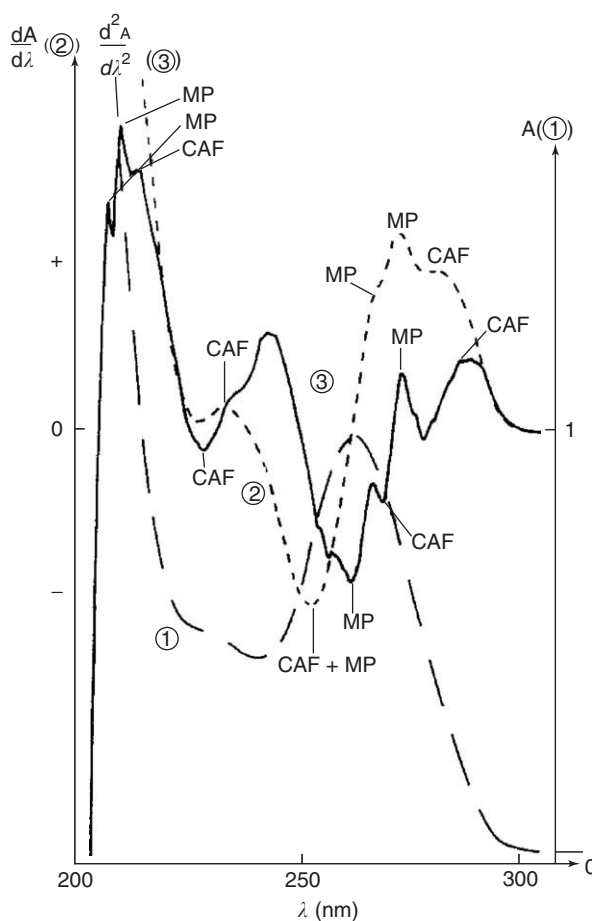
Ultraviolet absorption spectrophotometry constitutes a very rapid and simple method for the identification and determination of purines, pyrimidines, and their derivatives. Because many of these compounds can undergo keto–enol tautomerism – a pH-dependent phenomenon – and/or prototropic equilibria, caution should be exercised in selecting the pH of the aqueous solution used for spectrophotometric measurements. As a consequence, spectra obtained at arbitrary pH values may correspond to the summation of spectra belonging to various species. Knowledge of the  $pK_a$  values of purines and pyrimidines is necessary to choose convenient pH values for spectral measurements.

Generally, UV absorption spectra can be utilized to determine the family to which an unknown base, nucleoside, or nucleotide belongs. Indeed, the chromophore of the nucleosides or nucleotides is typical of the purine or pyrimidine structure, with different maximum absorption wavelengths in the adenosine, guanosine, cytosine, or thymidine series. For example, it is possible to differentiate readily the UV absorption spectra of guanosine (G), G-monophosphate (GMP), deoxy GMP (dGMP), G-diphosphate (GDP), and G-triphosphate (GTP), from those of adenosine (A), AMP, dAMP, ADP, and ATP.

For mixtures of several nucleotides and/or nucleosides, however, spectroscopic analyses become difficult because UV spectra contain only a small number of absorption bands and have a very similar shape. In these cases, it is necessary to evaluate precisely and in detail the minor differences in the spectra, such as the position of the shoulders and the relative intensities of the various bands, which are often sufficient for multicomponent analysis. Two approaches, the so-called principal component regression (PCR) and partial least-squares (PLS) regression analyses, with Fourier preprocessing of UV spectra, have been developed. These methods can monitor continuously entire spectral regions, and they are generally reliable and accurate. They have allowed the composition of 14 unknown mixtures of dAMP, dCMP, dGMP, and dTMP to be analyzed, with average percent errors of 0.60–1.85%. The presence of a protein

(up to 36% of albumin added to the mixtures) is found not to interfere significantly with the spectral results of the PCR and PLS approaches.

The analysis of mixtures of purines and pyrimidines can also be carried out by using first- and second-derivative UV spectra. Indeed, these derivative spectra contain many more features, such as peaks, shoulders, and/or trough wavelengths, useful for identification purposes, than the zero-order spectra. An example is provided in Figure 5, which shows zero-order first-, and second-derivative spectra of a caffeine-6-methylpurine (1:1) mixture. Second-derivative spectra can be utilized for the identification of binary mixtures of purines and pyrimidines, as well as for their determination. Moreover, several computer programs have been described for resolving



**Figure 5** Zero-order (1), first-derivative (2), and second-derivative (3) UV spectra of a mixture of caffeine (CAF) ( $5 \times 10^{-5} \text{ mol l}^{-1}$ ) and 6-methyl purine (MP) ( $5 \times 10^{-5} \text{ mol l}^{-1}$ ). (Reprinted with permission from Aaron JJ and Gaye MD (1988) Analysis of mixtures of purines and pyrimidines by first and second derivative ultraviolet spectrometry. *Talanta* 35: 513–518; © Elsevier.)

second-derivative UV spectra and elucidating the components of purine and pyrimidine binary mixtures. For example, mixtures of adenine–cytosine (1:1), adenine–purine (1:1), cytosine–thymine (1:1), theobromine–theophylline (1:1 and 1:9), guanine–6-iodopurine (1:1 and 1:9), and thymine–uracil (1:1) have been identified unequivocally using these computer-assisted methods.

### Phosphorescence Spectrometry

The luminescence properties of purines, pyrimidines, and nucleosides have been extensively investigated, mainly from a physicochemical standpoint. These compounds generally have low fluorescence quantum yields, between 0.01 and 0.06, according to the molecule, in glycol–water (1:1), pH 7, and at a low temperature (77 K). The fluorescence intensities of purines and pyrimidines are practically nil or very weak at room temperature, preventing the analytical use of this method.

In contrast, the phosphorescence quantum yields are significantly higher, and thus low-temperature (77 K) phosphorescence (LTP) and solid-surface room-temperature (298 K) phosphorescence (RTP) have been applied to the analysis of purines and pyrimidines. The preferred technique is RTP because it is rapid, simple, and sensitive and avoids the use of the expensive cryogenic equipment that is needed in the case of LTP. A variety of pH conditions, heavy ions ( $I^-$ ,  $Tl^+$ ,  $Pb^{2+}$ ,  $Sm^{3+}$ , etc.), and solid substrates (cellulose, silica gel, aluminum oxide) have been studied. Optimal conditions have been selected according to the relative enhancement effects of pH, heavy ions, and solid substrate on the RTP signal of purines and pyrimidines. In most cases, thallium ions, and filter paper as the solid substrate, are found to provide good analytical results, with the optimal pH depending on the  $pK_a$  value of the compound. Linear calibration graphs cover two to four orders of magnitude of concentration. The LODs for RTP are relatively low, ranging between  $\sim 40$  pg and 19 ng, according to the particular purine or pyrimidine. Binary mixtures of purines and pyrimidines have also been determined using synchronous RTP.

### Special Methods for Nucleotides

Nucleotides contain purines or pyrimidines in their structure (Figure 1). Nucleotide purines and pyrimidines have different and important functions in biological system. They act as the bases of genetic material (DNA and RNA). Nucleotide triphosphates (ATP, GTP, UTP, and others) are considered to be energy centers containing a linked chemical energy.

This energy can be transferred to many ‘inert’ compounds such as amino acids, lipids, and carbohydrates, which then can contribute to other anabolic or catabolic reactions. Many biochemical reactions are regulated by cyclic nucleotides, including cyclic AMP (cAMP) and cyclic GMP (cGMP), which have a role of second messenger within hormone functions (in animal world), and other substances similar to them, which act in plants, bacteria, and viruses. Some nucleotides are coenzymes and prosthetic groups of many enzymes – oxidoreductases ( $NAD^+$ ,  $NADP^+$ , FAD) – and can transfer groups. Also, nucleotides are contained in the structure of chlorophyll, vitamin B12, and coenzyme A SH. Other nucleotides are very important pharmacological compounds that are used to prevent the growing up and mitosis of cells, at different levels. As a result, the determination of nucleotide concentrations in various biological media constitutes an important aspect of many experimental biological projects and therapeutic applications.

Generally, the nucleotide concentrations are very low in biological systems, and those of cyclic nucleotides are more than 1000 times lower. The first step of nucleotide quantification is to extract them, by means of boiling buffers, strong acids (sulfuric acid, perchloric acid, TCA), *n*-butanol, or bromosuccinamide. The choice of the extractant depends on the biological material to be analyzed. Moreover, the extracted solution must be neutralized at a pH value optimum for maximum enzymatic activity. After their extraction, nucleotides can be determined either using bioluminescence methods (firefly and bacterial) or using radioassay methods (radioimmunoassay and binding-protein assay).

The firefly bioluminescence method enables us to determine ATP. Luciferin undergoes a reaction of oxidative decarboxylation at pH 7.75 in the presence of ATP, oxygen, and a firefly luciferase (EC 1.13.12.7), yielding oxyluciferin,  $CO_2$ , AMP, pyrophosphate, and light emission. At high levels of luciferin, the luminescence intensity is proportional to the ATP concentration. Besides ATP, firefly bioluminescence is also used to determine the concentrations of other nucleotides and of a large number of important biological compounds. In addition, the activities of many enzymes are able to produce or consume ATP in successive coupling reactions.

The bacterial bioluminescent reaction requires the contribution of the flavine mononucleotide (FMN) reduced form (FMNH), a long-chain aldehyde (tetradecanal), and oxygen. All these molecules produce an oxidative reaction in the presence of bacterial luciferase (EC 1.14.14.3), giving tetradecanoic acid, FMN, water, and light emission; under these conditions,



the luminescence intensity is found to be proportional to the FMNH<sub>2</sub> concentration. Other substances that produce or consume FMNH<sub>2</sub> and/or nicotinamide adenine dinucleotide reduced form (NADH) in successive coupled reactions can also be determined using the same bacterial bioluminescence method. Indeed, in the last case, hydrogen atoms are transferred from NADH<sub>2</sub> to FMN, forming FMNH<sub>2</sub> in a redox reaction.

The concentrations of cyclic nucleotides can be quantified by means of two different radioassays methods (radioimmunoassay and binding-protein assay). The radioimmunoassay method of cAMP and cGMP determination is based on competitive ligand binding. The tissue extract is first mixed with <sup>125</sup>I radioactive tracer nucleotide and then incubated with a specific cyclic nucleotide antibody at 4°C. The reactivity due to the antibody is measured and is found to be in inverse proportion to the concentration of the nonradioactive competing ligand. The binding-protein assay is based on the use of a binding protein, which is specifically associated with cAMP in competition with radioactive-labeled AMP. Radioimmunoassay has a greater sensitivity than the binding-protein assay.

In general, the bioluminescence methods are considered to be more sensitive, accurate, and precise than the radioisotopic methods. Moreover, the first methods are more appropriate than the second ones from an ecological standpoint.

*See also:* **Atomic Mass Spectrometry:** Laser Microprobe. **Bioluminescence.** **Capillary Electrophoresis:** Overview. **Chemometrics and Statistics:** Multivariate Calibration Techniques. **Extraction:** Solvent Extraction Principles. **Liquid Chromatography:** Overview; Reversed Phase. **Mass Spectrometry:** Principles. **Phosphorescence:** Principles and Instrumentation; Room-Temperature.

## Further Reading

Britz-McKibbin P, Nishioka T, and Terabe S (2003) Sensitive and high-throughput analyses of purine metabolites by dynamic pH junction multiplexed capillary electrophoresis: a new tool for metabolomic studies. *Analytical Sciences* 19: 99–104.

Brown EG (1991) Purines, pyrimidines, nucleosides and nucleotides. *Methods in Plant Biochemistry* 5: 53–90.

Brown PR, Robb CS, and Geldart SE (2002) Perspectives on analyses of nucleic acid constituents: the basis of genomics. *Journal of Chromatography A* 965: 163–173.

Delaye F, Gaye MD, and Aaron JJ (1989) Evaluation of computer-assisted second derivative ultraviolet spectrophotometry for analysis of binary mixtures. *Analytica Chimica Acta* 223: 395–402.

Donahue SM, Brown CW, and Scott MJ (1990) Analysis of deoxyribonucleotides with principal component and partial least-squares regression of UV spectra after Fourier preprocessing. *Applied Spectroscopy* 44: 407–413.

Gaye MD, Delamar M, and Aaron JJ (1990) Study of intermolecular interactions on several solid substrates using room temperature phosphorescence and XPS analysis. *Analysis* 18: 571–577.

Gaye-Seye MD and Aaron JJ (1994) Analysis of mixtures of purines and pyrimidines by synchronous room temperature phosphorescence spectroscopy. *Analytica Chimica Acta* 290: 166–171.

Geldart SE and Brown PR (1998) Analysis of nucleotides by capillary electrophoresis. *Journal of Chromatography* 828: 317–336.

Kricka LJ (2002) Stains, labels and detection strategy for nucleic acids assays. *Annals of Clinical Biochemistry* 39(2): 114–129.

Lim CK and Peters TJ (1989) Isocratic reversed-phase high-performance liquid chromatography of ribonucleotides, deoxynucleotides, cyclic nucleotides and deoxycyclic nucleotides. *Journal of Chromatography* 461: 259–266.

Newton RP, Brenton AG, Ghosh D, *et al.* (1991) Qualitative and quantitative mass spectrometric analysis of cyclic nucleotides and related enzymes. *Analytica Chimica Acta* 247: 161–175.

Nicolas JC, Boussieux AM, Joyeux A, *et al.* (1995) Applications of luminescence reaction in life science. *Compte Rendus des Sciences de la Société de Biologie et de ses filiales* 189(5): 740–753.

Trajkovska S, Dzhekova-Stojkova S, and Kostovska S (1994) Determination of validity for blood for transfusion by firefly bioluminescent assay for ATP. *Analytica Chimica Acta* 280: 246–248.

Tu SC, Lei B, Liu M, Tang CK, and Jeffers C (2000) Probing the mechanisms of the biological intermolecular transfer of reduced flavin. *The Journal of Nutrition* 130(2S): 331S–332S.

# PYRIMIDINES

*See* PURINES, PYRIMIDINES, AND NUCLEOTIDES

# PYROLYSIS

*See* GAS CHROMATOGRAPHY: Pyrolysis. MASS SPECTROMETRY: Pyrolysis



# QUALITATIVE ANALYSIS

**M Valcárcel and S Cárdenas**, Universidad de Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Chemical analysis can be divided into three main groups according to the purpose and type of information required:

- *Qualitative analysis*. This concerns the identification of the analytes present in a sample being subjected to the chemical measurement process.
- *Quantitative analysis*. This aims to determine the amount or proportions of an analyte, atoms, or chemical groups in the material studied and, therefore, it provides a numerical response suited to the particular question raised.
- *Structural analysis*. This involves the elucidation of (bio)chemical structure pertaining to an individual analyte or sample.

These three concepts are not independent as they can be related through a scope hierarchy. Thus, quantifying an analyte entails prior checking to find out whether it is present in the sample concerned. In addition, the identification of other interfering species that might be present in the sample will also be of interest. Likewise, structural analysis relies on the qualitative and quantitative information previously obtained. The boundaries between the three analyses are not very clearly defined as they share some connotations.

Two different terms can be found in the literature when dealing with qualitative analysis although they present slightly different connotations. The word 'detection' is normally used to refer to a chemical measurement process for qualitative purposes, opposed to determination, which is reserved for quantitative analysis. 'Identification' is usually employed for qualitative analysis aimed at recognizing the analyte (or its reaction product) from some chemical or physicochemical properties. As it entails the use of a standard for signal comparison, it is more appropriate as an alternative to qualitative analysis from a metrological point of view.

Qualitative analysis is an emerging area in pure and applied chemistry as its main objective is to obtain the maximum reliable information from the sample with minimum cost, effort, and time. The success or failure of such analysis is largely dependent on the skill and judgment of the analyst. Sometimes it turns to be a difficult task as many of the tests are not clear-cut and cannot be easily interpreted, being thus of little value in the hands of unskilled analysts.

It is interesting to note that although the classical slant of qualitative analysis had declined in the last decades with the advent of instrumental analytical techniques (the qualitative potential of which is under-exploited), simple qualitative tests are extremely relevant nowadays. It can be understood on the grounds of the increasing demand for achieving consistency between delivered and required analytical information that tends to be of a binary nature to a large extent: presence/absence of an analyte, positive sample/negative sample, or yes/no response according to a previously established threshold value.

## Qualitative Chemical Information

The principal objective of the analytical problem-solving process is to ensure the representativeness of the information delivered in such a way that the differences between the information demanded by the client and that delivered by the laboratory are minimal. Indeed, the amount of information that analytical laboratories provide usually exceeds the client's real needs. In this way, the analytical problem is not properly solved. The characteristics of the expected analytical information have constantly evolved from the classical approach to the present approach, which can be summarized as follows:

- greater simplicity;
- a substantial portion is of qualitative nature or based on the production of global indexes;
- relevance of productive properties (costs, safety, and expeditiousness); and
- information generated *in situ*.

Clients are usually interested in rapid, low-cost but reliable overall indicators (such as yes/no responses). For example, timely chemical information with 10% uncertainty would be better than very accurate and precise information but delivered too late to make a correct, profitable decision. An environmental body may be interested in knowing if an industrial soil is contaminated with hydrocarbons rather than in obtaining a detailed list of the analytes, including those at extremely low concentrations with their corresponding uncertainties.

Qualitative chemical information is characterized by the following features:

- it is usually of a binary nature (yes/no);
- it also presents quantitative connotations;
- it requires the definition of new analytical properties although the classical ones are still valid in this context; and
- it is scarcely supported by metrological standards and guides.

## The Binary Response

The expected response from a qualitative analysis is of a binary nature: yes/no. However, the level of information of such a response can be rather different

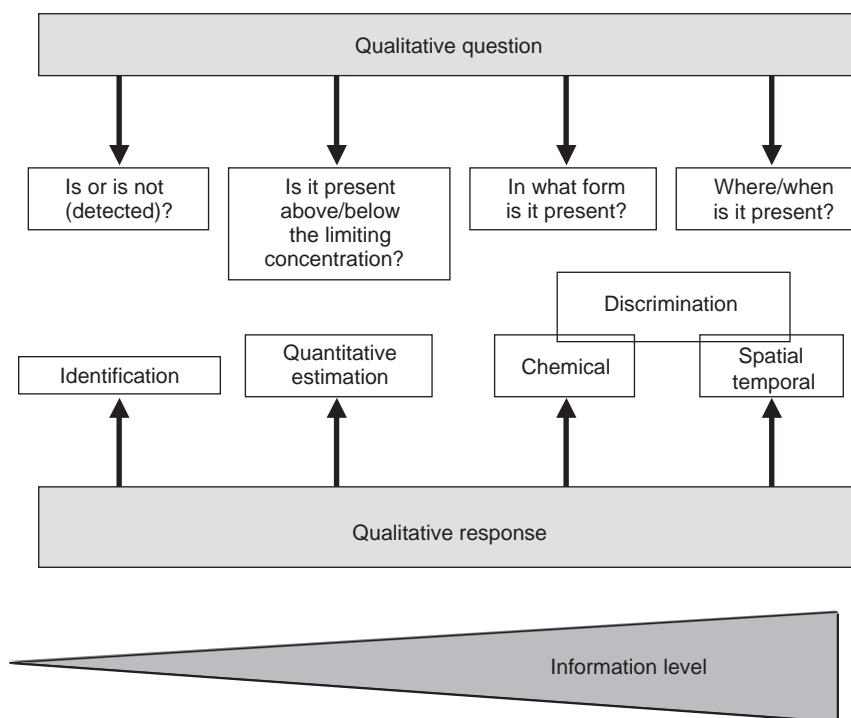
depending on the question raised on the sample; some examples are shown in **Figure 1**.

The lowest level of information corresponds to the identification of the analyte in the sample (is the analyte present/detected or not), thus being a prerequisite for the others. Sometimes, the laboratory may need to consider a concentration limit imposed by the client or legislation. The qualitative information given has more quantitative connotations than stated in the simplest question.

Even in its most simple possible form, the yes/no binary response can differ chemically, spatially, or temporally (third and fourth levels of information). Qualitative information can discriminate among the different forms in which an analyte may occur in a given sample (speciation). This chemical discrimination involves a multiple binary response (one per species potentially present). Finally, the highest level of information demands expands the binary response with temporal or spatial discrimination answering the question 'when' for dynamic objects or 'where' for heterogeneous ones.

## Quantitative Connotations

The quantitative connotations of qualitative information are derived from the data comparison carried out in response to an existing limiting



**Figure 1** Types of qualitative responses according to the information content they provide to properly answer the qualitative question posed by the client.

concentration/amount of analyte. It allows the laboratory to determine if a sample/object has the required quality. In this regard, three quantitative references can be clearly identified, namely: limit of detection, cut-off, and threshold/alarm limits. The limit of detection is an internal reference inherent in the chemical measurement process. The cut-off limit is established by the analyst/organization in setting a given probability level to ensure the obtainment of a correct binary response. The threshold concentration is the highest or lower level established by the client or legislation to be used in deciding whether or not the sample warrant assignment of a given attribute (e.g., contaminated, light, etc.). According to these definitions, the limit of detection must be lower than the other two references to allow detection/identification. The cut-off concentration must exceed the limit of detection as it involves a higher probability level to make the detection error-free. Finally, any externally imposed limit should be greater than the other two to ensure a reliable binary response. The higher the difference between the cut-off/threshold concentration and the detection limit, the more reliable will be the qualitative information given. The internal (cut-off) reference can always be established while the external one (threshold concentration) may or may not exist, coincide, or even differ with applicable legislation or client's specific needs.

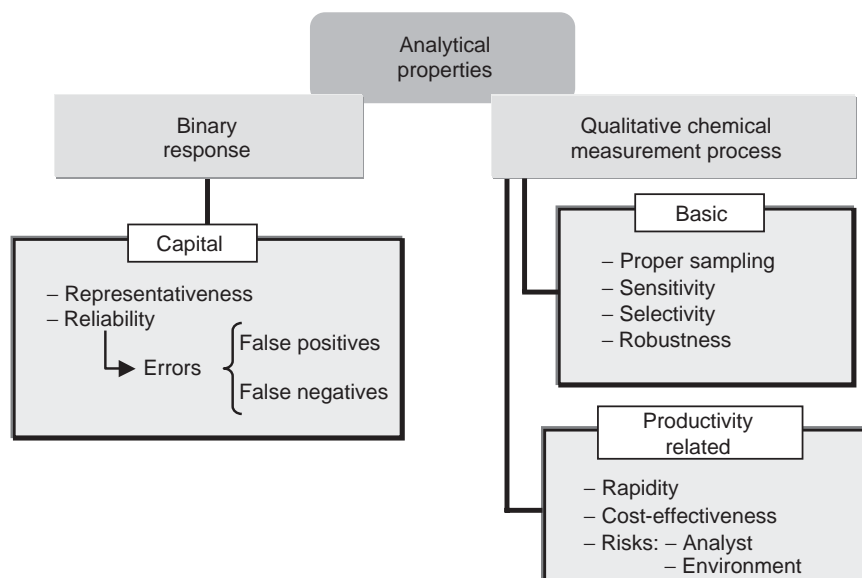
### Analytical Properties

The analytical properties that characterize a quantitative analytical process cannot be directly extrapolated to qualitative analysis owing to the peculiarities

of the binary yes/no responses. Some adaptations are needed as can be seen in Figure 2.

**Representativeness** The binary yes/no response should be representative of the test sample, the object, the analytical problem raised, and the information requested. This property is especially significant when tackling the problem with the imposed limits and threshold in mind. Proper sampling is the main requirement of this crucial property.

**Reliability** It is rather difficult to use the accuracy and precision concepts as these capital and basic properties are closely related in qualitative analysis. Their combination has produced a new property called reliability, which is defined as the proportion (percentage) of right yes or no answers provided by individual tests carried out on  $n$  aliquots of the same sample to identify an analyte or a family of the analytes. This definition represents the positive side of the errors in qualitative analysis: false positives and false negatives. The reliability of the binary response is not an independent property as it strongly depends on the basic properties of sensitivity and selectivity. Moreover, it is in contradiction with productivity-related properties. Reliability is equivalent to certainty and, in quantitative analysis, the uncertainty of a result is a parameter associated with reliability. Indeed, the term is included in the definition of traceability as every experimental datum is affected by specific variations or doubts. As it directly affects the quality of an analytical result, it is necessary to find out an equivalent method to express uncertainty in qualitative analysis. The term unreliability can be



**Figure 2** Analytical properties in qualitative analysis.

proposed to characterize the range of responses where the errors are produced. It is a property of binary yes/no response and it is also expressed by a measurand response range. Both reliability and unreliability define the same capital analytical property in the opposite sense.

**Sensitivity** It expresses the ability of a qualitative method to detect small amounts of the analyte in a sample to produce the binary yes/no response. It can also be defined as the ability to discriminate yes/no responses for samples containing similar amounts of analyte. From the different parameters that can be used in quantitative analysis, the only one relevant in qualitative analysis is the above-mentioned limit of detection.

**Selectivity** This property may refer to a single analyte (viz., the ability of the qualitative method to produce results that are exclusively dependent on the measurand for its identification) or to a family of the analyte and the absence of interferents from components other than the target compounds. Selectivity is a crucial support to reliability: the higher the selectivity, the better the certainty.

**Robustness** It is a basic analytical property of a qualitative method that describes its resistance to change the binary yes/no response when applied to individual amounts of the same sample under slightly different experimental conditions. The ultimate purpose is to find out the experimental weakness of the method, viz., the actual factors influencing qualitative responses. This property is directly related to certainty of identification (reliability).

**Productivity-related properties (rapidity, cost-effectiveness, and risks)** Although they are only occasionally considered in metrological approaches (validation), a systematic consideration is mandatory because of their practical impact. They can imply a serious limitation for routine application of the qualitative method even though it shows high levels of capital and basic characteristics. If the method is too expensive, slow, or involves risks, it cannot be appropriate to solve the analytical problem addressed (fitness for purpose).

The hierarchical foundation, and contradictory and complementary relationships among analytical properties are quite applicable in this context.

## Errors

The errors contained in qualitative information are specifically called 'false positives' and 'false negatives'. A false positive arises when the signal

from a sample containing the analyte at a level below the reference concentration yields a 'yes' response. A false positive also arises when the signal belonging to a compound different from the analyte is wrongly assigned to the target compound. On the other hand, a false negative results from the signal for a sample containing the analyte at a level above the reference concentration provides a 'no' response. Therefore, the assumption is made in rejecting a signal for a sample containing the analyte in the mistaken belief that its concentration is lower than the limit established. The relative concentration of the analyte in relation to the reference concentration affects the type of error made; if the concentration is slightly higher than the reference, then a false negative is expected while for concentrations lower but close to this reference value, false positives are to be expected.

The relevance of such errors depends on the particular analytical problem. In general, all positive results from a qualitative test will be systematically confirmed by a conventional chemical measurement process when any error made may have a significant social or economical impact and, therefore, their practical impact in the decision-making is low. However, quality assurance of negative responses is crucial, as their practical effects are more relevant. Indeed, they are especially serious when detecting or identifying a toxic chemical in environmental, food, or clinical samples as no confirmatory step is carried out.

## Classical versus Instrumental Qualitative Analysis

Qualitative analysis can be classified according to a variety of criteria. One of them considers the analytical technique used and, therefore, qualitative chemical measurement processes can be divided into two main blocks: classical and instrumental qualitative analyses. Their main characteristics will be briefly discussed below.

### Classical Qualitative Analysis

Classical qualitative analysis uses human senses (as sensors) and the brain (to process the signals) to detect the presence of an analyte that was previously subjected to a (bio)chemical or immune reaction with the view of yielding a product clearly identified from a well-defined change.

Identification is carried out by sensual comparison, possibly but not necessarily using a reference scale (e.g., test strips) from which the operator will obtain semiquantitative information. Such is the case with pH, active chlorine, and glucose measurements based on reagent strips.



The main limitations of classical qualitative analysis can be summarized as follows:

- low sensitivity due to limited capacity of the human operator to detect small changes and, hence, low discrimination among signals;
- the less variety of information that can be derived; and
- reliability relies on the chemical reaction(s) involved.

The chemical measurement process used in classical qualitative analysis can be either a direct or a systematic procedure using sequential separations to indirectly raise sensitivity and selectivity. The qualitative analysis also differs depending on whether a single analyte, a family of analytes, a small group, or a wide range of groups are to be identified.

There are three generic approaches to the identification of analytes in complex mixtures, namely:

- Direct individual identification of each analyte or family using highly selective and sensitive tests.
- Individual identification following systematic separations (precipitation, liquid–liquid extraction, ion exchange, etc.) to obtain spatially discriminate species (individually or in groups).
- Individual identification of each analyte in different sample aliquots, which may be subjected to a simple separation technique, included in a pre-established operational sequence.

Analytes can be identified using a variety of reagents and the many different combinations of analyte–reagent resulted in a wide range of situations of variable complexity. Such combinations dictate the type of identification reaction to be used in each case. Thus, inorganic reagent and analyte combination involves a precipitation, redox, or complex formation reaction; the organic reagent–inorganic analyte combination leads to a colored or fluorescent chelate formation; and in the organic or biochemical analyte combination, the identification reaction is of organic or biochemical type.

The reagents typically used in classical qualitative analysis can be grouped into three main categories:

- *Identification reagent.* Compounds of inorganic, organic, biochemical, or immune nature that give a reaction, the external effect of which can be readily detected by the human senses (precipitate, gas, color, change, etc.). The test associated should possess adequate sensitivity and selectivity, directly or after separations included in an analytical scheme.
- *Group reagents.* They affect the separation of the analyte(s) present in a sample in order to increase the reliability of individual identification.
- *Masking reagents.* They prevent interference from other species potentially present in the sample or isolated group of analytes. They should never significantly disturb the main analytical reaction.

### Instrumental Qualitative Analysis

Instrumental qualitative analysis converts a physico-chemical property of the analyte (or its reaction product) into an analytical signal that can be measured by an instrument with the view of identifying it. It is, therefore, more reliable and more widely applicable than the classical approach. Notwithstanding this, not all instruments are equally capable of providing reliable identification. In this context, reliability can be understood as a combination of the selectivity and the amount of information content of the response that they provide. For example, the use of a single signal (e.g., absorbance) obtained using a single instrumental parameter (e.g., wavelength) obviously provided less reliable qualitative identification than multiple signals recorded using many different instrumental parameter values (e.g., the bands in an infrared absorption spectrum).

Instrumental analytical techniques applied to qualitative analysis can be divided into three main categories:

- *Group 1.* They are useless in qualitative analysis as they provide general responses, usually one shared by many analytes (e.g., balance or piezoelectric sensor). This signal can be, however, altered in some way to increase its selectivity (e.g., thermogravimetry or piezoelectric sensor coated with selective sorbent material).
- *Group 2.* These instruments give scarcely selective responses (e.g., photometers), but can be improved by analyte derivatization to products with different spectral characteristics. Higher reliability can be achieved by using more selective techniques (e.g., fluorimetry). Moreover, online coupling of these instruments with continuous column chromatographic techniques results in greater selectivity as it permits the separation of the analytes prior to detection.
- *Group 3.* It includes highly selective instruments (atomic absorption and emission spectroscopies) or instruments providing information that contains many well-resolved signals obtained using multiple instrumental parameter values (infrared spectroscopy or mass spectrometry). Their use as detectors in chromatographic techniques has given

rise to the so-called hyphenated techniques, which provide significantly increased reliability in identification. This analytical property can be substantially improved by the joint use of the information provided by one instrument of the second group and another from the third group. The highest level of reliability results from the combination of a separation technique and two or more instruments of the second, or better, third group.

In this context, the fitness-for-purpose should be borne in mind, i.e., the selection of the most appropriate instrument on the grounds of the actual client's information needs rather than the selectivity provided by a given analytical technique.

## Modern Approach to Qualitative Analysis

Analytical information delivered by the laboratory should be coherent with that demanded by the client in order to properly solve the social problem raised. As was previously stated, qualitative information represents a substantial portion of the information required. Therefore, modern qualitative analysis needs to consider what the main objectives of the analysis are, externally imposed by the client or regulation needs. Under this consideration, the two types of qualitative analysis shown in **Figure 3** can be distinguished, namely: qualitative analysis for sample qualification/classification and qualitative analysis for analyte identification.

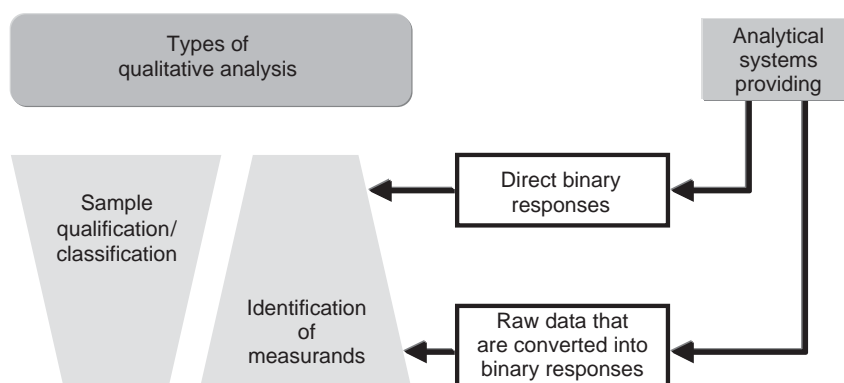
The difference is established according to the possible questions asked by the clients and not on the analytical technique used. The main objective of the former is to provide a rapid, reliable screening of the samples on the grounds of previously established criteria (cut-off or threshold values fixed by the

laboratory or the legislation, respectively). The second type permits the identification of an analyte or a family of the analytes as an answer to questions such as 'is this compound the analyte?' or 'is it detected in the sample?' using powerful spectroscopic and chromatographic tools. The two types of qualitative analyses are not independent, however. Identification/quantitation of the analyte is usually required prior to sample qualification/classification (viz., estimation of the concentration of a given analyte as regards the previously established cut-off or threshold).

As can be seen in the figure, qualitative analysis can be implemented using two general analytical tools. On the one hand, there are analytical systems, such as test kits, that provide the binary response in a direct way, which are more convenient for sample qualification/classification. On the other systems based on the use of analytical techniques providing raw data, which are further processed and converted into binary responses, are more appropriate for analyte identification.

### Qualitative Analytical Systems for Sample Qualification/Classification

Fast and reliable analytical processes designed for sample qualification/classification are generally known as sample screening systems. The term refers to analytical methodologies intended to identify and select a group of samples, from a starting set, in relation to a specific property of the sample, in order to provide a requested information (for example, if they contain one or more analytes above a preset concentration level). These straightforward, responsive analytical systems allow those samples to be filtered out, minimizing the preliminary operations of the chemical measurement process. Only those samples providing a positive response in the screening step will be subjected to the conventional analytical



**Figure 3** Types of qualitative analysis according to the objective of the analysis and their relationship with qualitative analytical systems.

process, both for confirmation and increasing the information level.

Test methods are also useful for sample qualification/classification. They can be defined as a combination of a test tool and procedures for detection and determination, which do not require complex sample treatment, sophisticated instruments, or complicated data treatment. In most cases they involve self-contained, single-use tools. The reagents and additives are supplied as ready-for-use solutions (in ampoules or droppers) or immobilized on a solid support such as paper, silica gel, or polyurethane foam. They are expected to be of reduced size and weight, easy to use, having speed of operation, and independent power supply. Either visual or colorimetric detection for comparing and assessing color can be normally used. Advances in other areas such as microelectronics, physics, and technology have become feasible to build portable, pocket size instruments such as photometers, reflectometers, luminometers, and turbidimeters. The most relevant applications of such devices are the determination of organic and inorganic species in water, monitoring of toxic gases, detection of vapors, alcohol, and narcotics, and the determination of glucose in whole blood.

Qualitative analysis for sample qualification/classification can be divided into two main groups depending on whether the information required by the client is related to one single chemical species or not. Indeed, sometimes the questions are related to a quality of the sample, which may not only be defined by single chemical species, for example, 'is this wine a Merlot?', is the hydrocarbon index above a threshold value?'. In these cases, the response expected from the analytical process is a global or total index. A global index can be defined as a measurand that describes a group (family) of chemical or (bio)chemical species from millions to few of them that have similar nature/structure (total fat), operation behavior (chemical oxygen demand), or both (total polychlorinated biphenyls). Calculation of global indices depends on the raw data provided by the qualitative chemical measurement process followed. One group provides a single primary signal at a fixed instrumental parameter with different contributions of the different measurands of the target group. The other group gives discriminate signals for each analyte that needs further conversion into a global index.

### Qualitative Analytical Systems for Identification of Measurands

Methodologies intended for both analyte(s) identification in complex samples and multideterminations can also be considered within the framework of

qualitative analysis. These systems use conventional sample treatments and/or powerful identification techniques (e.g., Fourier-transform infrared spectroscopy, inductively coupled plasma-mass spectrometry, nuclear magnetic resonance), with or without analyte separation using either chromatography or electrophoresis. The last alternative enhanced sensitivity and selectivity through (a) separation of the monitored compounds from matrix components and separation of the analytes themselves, which can be expanded using multidimensional or multimodal chromatography; and (b) increased identification power by using hyphenated techniques.

Qualitative procedures for identification of measurands have been successfully applied in different areas of interest. They are widely variable in the sample treatment proposed, which is directly related to the nature of the sample and analytes as well as the detector used.

**See also:** **Bioassays:** Overview; **Microbial Tests.** **Clinical Analysis:** Overview. **Food and Nutritional Analysis:** Overview. **Forensic Sciences:** Overview; **Drug Screening in Sport.** **History of Analytical Science.** **Microscopy:** Overview. **Spot Tests.** **Structural Elucidation.**

### Further Reading

- Hardcastle WA and Ellison E (2002) Expression of uncertainty in qualitative analysis. Approaches to the problem. *EURACHEM/CITAC Conference Discussion Paper*.
- Jungreis E (1997) Spot test analysis. In: Winefordner JD (ed.) *Chemical Analysis*, vol. 141. New York: John Wiley.
- Ríos A, Barceló D, Buydens L, Cárdenas S, Heydorn K, Karlberg B, Klemm K, Lendl B, Milman B, Neidhart B, Ríos A, Stephany R, Townshend A, Zschunke, and Valcárcel M (2002) Quality assurance of qualitative analysis in the framework of the European project 'MEQUALAN'. *Accreditation and Quality Assurance* 8: 68–77.
- Valcárcel M (2000) *Qualitative Aspects of Analytical Chemistry. Principles of Analytical Chemistry*, pp. 201–245. Heidelberg: Springer-Verlag.
- Valcárcel M, Cárdenas S, and Gallego M (1999) Sample screening systems in analytical chemistry. *Trends in Analytical Chemistry* 18: 685–694.
- Valcárcel M, Cárdenas S, and Gallego M (2000) Qualitative analysis revisited. *Critical Reviews in Analytical Chemistry* 30: 345–361.
- Valcárcel M, Cárdenas S, Barceló D, Buydens L, Heydorn K, Karlberg B, Klemm K, Lendl B, Milman B, Neidhart B, Ríos A, Stephany R, Townshend A, and Zschunke A (2002) *Metrology of Qualitative Chemical Analysis*. EUR 20605 EN, European Commission, Brussels.
- Zolotov YA, Ivanov VM, and Amelin VG (2002) Chemical test methods of analysis. In: Barceló D (ed.) *Comprehensive Analytical Chemistry*, vol. XXXVI. Amsterdam: Elsevier.

# QUALITY ASSURANCE

## Contents

### Quality Control

### Primary Standards

### Spectroscopic Standards

### Internal Standards

### Instrument Calibration

### Interlaboratory Studies

### Reference Materials

### Production of Reference Materials

### Method Validation

### Traceability

### Accreditation

### Laboratory Information Management Systems

### Clinical Applications

### Water Applications

## Quality Control

**D B Hibbert**, University of New South Wales, Sydney, NSW, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

From the earliest days of mass production, in which pikes of equal length and standard manufacture in the New Model Army of Oliver Cromwell is a well-documented example, there has been a clear desire to ensure that the output of a process is as it was intended. A manufacturer of a physical product has particular specifications and tolerances that allow testing of the product before it is unleashed on the unsuspecting public. The provider of an analytical service is under no less an obligation to provide a quality product, but it may be less obvious what the definition of quality in that product is. The problem is that each analytical result is a unique entity whose true value is usually unknown.

The analytical laboratory, in setting up a quality assurance scheme, must therefore employ strategies that allow it to infer quality in a result that is released to a client. This section is an overview of the types of approach that can be employed and some of the methodologies employed by laboratories to ensure, as far as possible, their products – analytical results – are of high quality.

## Definitions of Quality

The literature is not at a loss for definitions of quality:

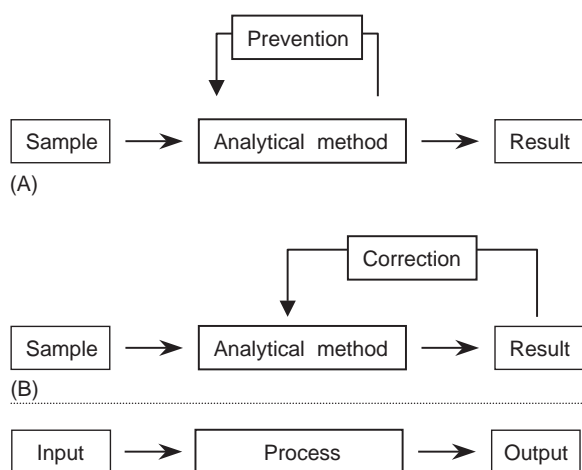
- Delivering to a customer a product or service that meets the specification agreed with the customer, and delivering it on time.
- Satisfying customer requirements.
- Fitness for purpose.
- Getting it right the first time.

While none of these seems to be quite right for analytical science, the intent is entirely clear. The analyst must deliver to his or her client a result that conforms to agreed specifications and can be used for the client's intended purposes. How the analyst does that is by implementing a regime in the laboratory that provides such assurances about the results being produced.

Different aspects of quality assurance are defined by the level in the organization that they are concerned with. The Association of Official Analytical Chemists has offered the following:

1. *Quality system*. Laboratory activities aimed at producing accurate work and a high-quality work product.
2. *Quality control*. Planned activities designed to provide a quality product.
3. *Quality assurance*. Planned activities designed to ensure that quality control activities are being properly implemented.

Thus, quality control (QC) is overseen by quality assurance (QA) and both are part of the laboratory's quality system. In these definitions, planned activities



**Figure 1** Analysis viewed in operational terms as a process operating on inputs to produce outputs. Quality control inspects (A) the process, initiating improvement, and preventative action, or (B) the output that triggers corrective action when nonconforming product is detected.

refer to a systems approach to management in which the operations of the enterprise may be viewed in terms of inputs, processes, and outputs, with the possibility of inspection of outputs and processes and hence feedback causing improvement in the quality of the product (Figure 1).

Hopefully, by inspection of the process (Figure 1A) potential concerns may be identified and dealt with before results are obtained. Traditional QC inspects the product (here the results) and nonconforming output initiates feedback to the process (Figure 1B). As discussed above, analytical science rarely has the luxury of knowing whether the output is of sufficient quality until the client complains.

## Measures of Quality in a Laboratory

The reputation for quality of a laboratory rests on a number of factors that may include: the scientific reputation of the laboratory; observed work practices; staff qualification and experience; a history of satisfactory achievement; good results in QA audits and reviews; and accreditation to international guides and standards.

In a survey of analysts the top rating indicators of quality were the use of validated methods and the use of reference materials followed by participation in proficiency testing and quality audits. The Valid Analytical Measurement program of the Laboratory of the Government Chemist in the UK lists six aspects of a quality laboratory that mirrors the list given above:

- work to an agreed customer requirement;
- use validated methods and equipment;

- use qualified and competent staff;
- participate in independent assessment of technical performance (proficiency testing);
- ensure comparability with measurements made in other laboratories (traceability and measurement uncertainty);
- use well-defined quality control and quality assurance practices.

The assurance of quality in the analytical laboratory differs from that in manufacturing industry in that although statistics are used extensively, and the application of the principles of statistical process control have grown in recent years, the scope of QA activities is much wider.

## The Basis for Quality Analytical Measurement

### Use of Validated Methods

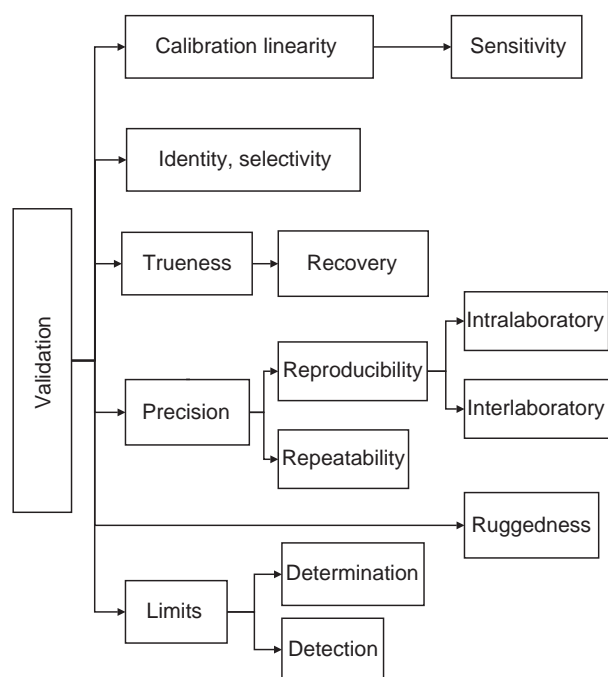
The use of any analytical method implies that the method, if employed properly, will yield an acceptable result. However, a client of an analytical service is entitled to have some documentary evidence that this assertion is true. Method validation (q.v.) is a lengthy process that is applied to a new method after development to provide evidence that the method will indeed produce results to satisfy the client's requirements. Figure 2 is a schematic of the different aspects of a method that may be considered during method validation.

Standard methods of analysis, for example, those published by American Society for Testing and Materials or International Organization for Standardization (ISO), have the imprimatur of the organization and have been developed and validated for use in field laboratories. They have usually been subjected to a method performance study, a kind of interlaboratory study (q.v.), which establishes repeatability and reproducibility precision and possibly method bias. The published methods also indicate the range of concentrations for which the method can be used, and the detection limit.

It is the responsibility of any laboratory to demonstrate the validity of the method used. Referring to some third party validation is sufficient if there is a statement that the laboratory can follow the method to ensure the maintenance of the validation. This requires some measure of verification to show, for example, that the analyst in the laboratory can achieve the repeatability precision indicated in the method.

Although there has been a move to requiring the demonstration of a target measurement uncertainty,





**Figure 2** Schematic of the aspects of an analytical method that may be assessed during a method validation.

there is much legislation in countries around the world that requires the use of a specified standard method.

### Assessment of Measurement Uncertainty

It is now understood that an analytical result comprises not only a number and its units but also the measurement uncertainty (q.v.). As part of accreditation requirements, and also as a proper approach to providing a quality result, the measurement uncertainty should be assessed and quoted. The fitness for purpose of the result may determine the rigor with which the measurement uncertainty is estimated, but however carefully it is determined the quoted expanded uncertainty (the range within which the true value may be found with given probability) should be correct. This means that the expanded uncertainty should be sufficiently wide such that the quoted probability is at least fulfilled. Perhaps a more detailed assessment of uncertainty would reduce the range, but it should never be underestimated. Particularly, the reproducibility from a standard method is not sufficient as a measurement uncertainty, although it may comprise the largest component.

### Traceability of Measurements

By giving a result in a particular unit implies that the result may be traced back to the definition of that unit. The use of the unit mol,  $\text{mol dm}^{-3}$ , or kg

implies that the measurement result is traceable to the definition of the unit of amount of substance, amount of substance concentration, or mass, respectively. The traceability chain is established by a series of calibrations using calibrators whose quantity values are themselves traceable. In a field measurement, the calibrator is usually an in-house standard that is traceable to a certified reference material. Without explicit traceability the result is only as good as the calibrator used, which may mean that the laboratory gives results that are traceable to values carried by material found only in that laboratory.

## Approaches to Quality Management

### Accreditation

Although accreditation *per se* does not guarantee the quality of a given result, it is clear that a laboratory that has been through a careful scrutiny of its operations and has been assessed against an international standard must be in a better position to claim quality of results. National or international bodies themselves accredited and members of the International Laboratory Accreditation Cooperation conduct assessments and confer accredited status to laboratories under their jurisdiction. Legislation frequently requires laboratories undertaking analysis to be accredited to International Standards Organization (ISO)/International Electrotechnical Commission (IEC) 17025. Here, two international standards are briefly discussed, ISO/IEC 17025 and Good Laboratory Practice (GLP) guidelines issued by the United States Food and Drug Administration (US FDA) and European Union.

**ISO/IEC 17025** This standard, entitled 'General requirements for the competence of testing and calibration laboratories,' has two parts, one relating to management structures that covers the ISO 9000 series of quality practices, and the second to practical aspects of testing and calibration. The standard specifies that the laboratory shall use validated methods, determine measurement uncertainty, and ensure the traceability of its results. The laboratory shall also report measurement uncertainty when the client requests it and when it is applicable; for example, when a result must be compared to a statutory limit. This standard also allows the analyst to offer interpretations and advice where appropriate.

**Good Laboratory Practice** In response to concerns in the mid-1970s about the veracity of reports on the safety of clinical medicines, the US FDA issued regulations in 1978 to prescribe principles of GLP



that should be applied to the nonclinical safety testing of test items contained in pharmaceutical products, pesticide products, cosmetic products, veterinary drugs, as well as food additives, feed additives, and industrial chemicals. These regulations cover much the same areas of management and methodology of laboratory testing as ISO/IEC 17025, but are aimed specifically at laboratories conducting nonclinical research, rather than general testing and calibration. There is an emphasis on the managerial structure (particularly with regard to QA), record keeping, and audit control.

### **Other Management Approaches to Ensuring Quality**

Before turning to specific measures of QC in the laboratory, a number of possible approaches that are available, if not widespread in analytical science, will be discussed.

**Peer review/visitors** Common in government and academic institutions, but not in the commercial sphere (for obvious reasons of confidentiality), the use of senior personnel from a similar organization to conduct reviews of a laboratory's operation is one way of receiving useful advice. The review rarely goes down to the day-to-day operation and often focuses on the more strategic aspects of the laboratory.

**Benchmarking** Akin to peer review, an organization can benchmark itself against what is seen as the best available practice. Internal personnel or an outside consultant prepare a report in which the procedures of the organization are compared against those of the target best-practice organization. As with peer review, benchmarking tends to concentrate on the bigger picture of the laboratory's operations.

**Total quality management** Total quality management is a popular approach to organizational management that can be adapted to a research-oriented analytical laboratory. It is considered a leadership style of management that promotes and cultivates a 'one team approach' by accepting risk taking and with an emphasis on continuous learning and improvement. One aspect that does promote a scientific approach to quality is the use of objective monitoring of progress using statistical tools. Brainstorming sessions of all the staff are encouraged to provide new ideas that can be tried and evaluated. It is clear that this approach is more suited to an R&D environment rather than the routine application of a standard analytical method.

**Project management** Project management is a structured approach to work that views a project in terms of defining, planning, implementing, and completing. There is use of project tools such as Gantt charts, with milestones and the attendant monitoring and reporting. Project management emphasizes a full specification of the task and communication between groups. Again this is suited to the management of large research efforts rather than day-to-day analysis.

**Six sigma** Six sigma is one of the more recent popular approaches to QA that is based on a tight statistical approach to the production of a product. The name arises from a desire to limit the tolerance of a product to plus or minus six standard deviations and thus have only 3.4 defects per million. (This is the fraction outside  $\pm 4.5$  standard deviations from the mean; the method allows for some measurement uncertainty.) In order for the statistics to hold, the system must be in statistical control and the defects must be random and normally distributed. There is a heavy reliance on control charts and the system is built around what to do if there is evidence for nonconformity. For a nonconforming product six sigma institutes an approach with the acronym DMAIC – define, measure, analyze, improve, control. This has been implemented in some organizations, such as pharmaceutical companies, which produce large volumes of chemicals. However, strict statistical control of chemical products is not always easy, and considerations of the measurement process also needs to be taken into account.

## **Tools for Quality Control**

So far this article has concentrated on management approaches to quality. This is appropriate, because no amount of QC at the laboratory bench can overcome shortcomings in the QA system or the culture of the organization. However, once management is committed to producing quality results what are the tools available to help the organization?

### **Cause and Effect Diagrams**

At the planning stage, cause and effect diagrams can be useful in analyzing the operation of a laboratory. Together with Pareto charts and other such indicators they can allow management to pinpoint areas that may be of concern. In particular, in the analytical laboratory the use of a cause and effect diagram (also known as an Ishikawa or fish bone diagram) can form the basis of the estimation of measurement uncertainty. It is applicable to analytical science because most measurement results require calculation

and are based on a measurement model that can be used as a basis of further dissection of the method. As a simple example, consider a titration in which the amount of substance concentration of the unknown ( $M_{\text{unknown}}$ ) is determined from the endpoint volume of the titration ( $V_{\text{titration}}$ ), and the volume ( $V_{\text{standard}}$ ) and amount of substance concentration ( $M_{\text{standard}}$ ) of a standard solution:

$$M_{\text{unknown}} = \frac{M_{\text{standard}} \times V_{\text{standard}}}{V_{\text{titration}}} \quad [1]$$

where the standard solution is made up by weighing an amount ( $m$ ) of a standard (molar mass  $MW$ ) of given purity ( $P$ ) and making up to a volume  $V$ :

$$M_{\text{standard}} = \frac{m \times P}{V \times MW} \quad [2]$$

A consideration of the method is now broken down to investigation of the influence factors of each of the three components of  $M_{\text{unknown}}$ . They are arranged on a cause and effect diagram as shown in Figure 3A.

Each branch of the diagram is then elaborated to include individual influence factors. It is possible to continue the ramification until all influence factors have been identified (Figure 3B). A number of components may be identified as contributing to the re-

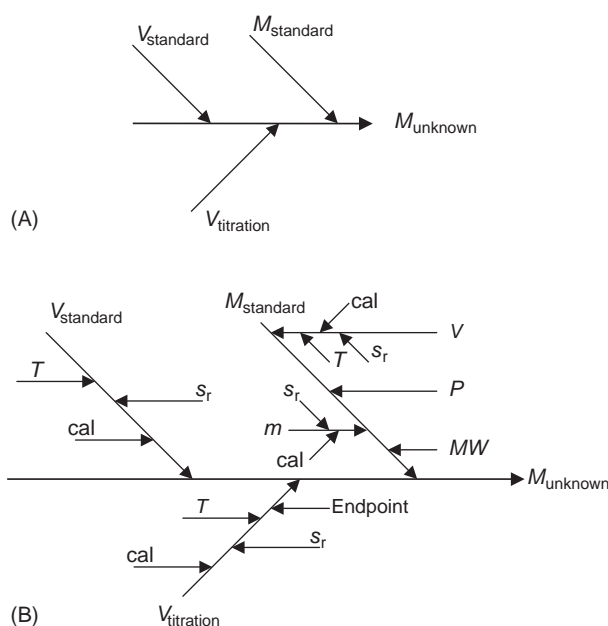
peatability or reproducibility precision. It may be possible to use validation data to provide much of the measurement uncertainty estimates.

The great advantage of the bottom-up approach to measurement uncertainty, even if precision data are to be used extensively, is that the analyst is made to consider the experiment in detail and he or she must understand the analysis and components of uncertainty. This may provide a useful insight into possible improvements that could be made to lower the combined uncertainty. Influence factors that are estimated to contribute less than, say, 20% to the measurement uncertainty can really be ignored when finding aspects for improvement. The contributions to measurement uncertainty can be displayed on a Pareto chart, which shows the individual components and cumulative effect. An example for a quantitative nuclear magnetic resonance (NMR) study is shown in Figure 4.

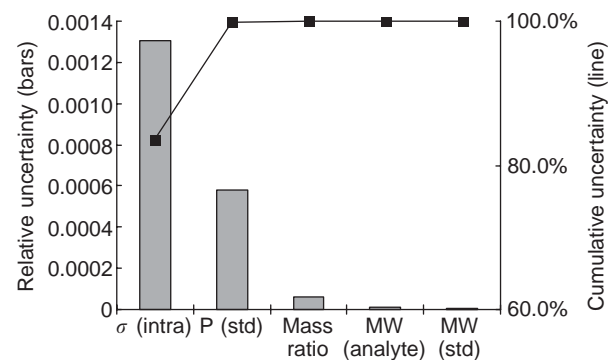
In this example, the intralaboratory precision is the greatest contributor to the uncertainty. There would be nothing gained, for example, in performing the mass measurements on a more precise balance.

### Control Charts

As part of the day-to-day operations of an analytical laboratory in which the same type of analysis is performed, it is usual to include check samples in the batches of unknowns. These are made from reference materials whose quantity values are known. If the analysis of this material correctly recovers the assigned quantity value within measurement uncertainty then the process is considered to be in control. Further information may be obtained by tracking these analyses over time, using control chart methodology developed during the early and mid-twentieth century for QC in the manufacturing industry.



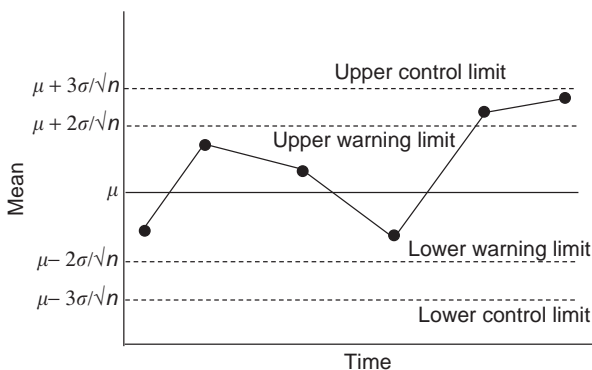
**Figure 3** Cause and effect diagram of a titration in which the standard solution is made up from pure solid. (A) Initial diagram with elements from the measurement model (see eqn [1]). (B) Diagram with expanded influence factors.  $T$ =temperature,  $s_r$ =repeatability,  $cal$ =manufacturer's calibration (see also eqn [2]). The mass measurements done by difference will contribute only linearity of calibration to the uncertainty labeled 'cal'.



**Figure 4** Pareto chart of contributions to the analysis of the purity of the agrochemical glyphosate by quantitative NMR. MW=molecular weight, P=purity,  $\sigma$  (intra)=intralaboratory standard deviation.

**Shewhart mean and range charts** A Shewhart means chart plots the average of determinations of a check sample against time. The time scale is chosen by the analyst and is whatever is most appropriate to the operation. If the interest is the long-term stability of the analysis, check samples analyzed each day may be averaged over a week, or if the batch-to-batch variation is of interest, the means of two check samples included in each run may be averaged. It is recommended that a sample size of at least two and no more than six should be chosen. The essence of the means chart is that the plot includes the expected mean and upper and lower lines that chart the expected limits of the results. If the system is in statistical control with mean  $\mu$  and standard deviation  $\sigma$  then the results should scatter about  $\mu$  with only five in every hundred results falling outside  $\pm 2\sigma/\sqrt{n}$  (where the sample size is  $n$ ), and three out of every thousand falling outside  $\pm 3\sigma/\sqrt{n}$ . Inspection of a means chart allows a check for any outliers or trends in the data. It is recommended that if two consecutive means fall outside  $\pm 2\sigma/\sqrt{n}$  (the so-called upper and lower warning limits) or if a single mean is outside  $\pm 3\sigma/\sqrt{n}$  (upper and lower control limits), then the analysis should be considered to be out of control and its use should be suspended followed by a careful investigation of possible causes of the problem. **Figure 5** is an illustrative means control chart in which the analysis has given two consecutive results over the upper warning limit. There are alternative formulae for the warning and control limits and more extensive rules for dealing with potentially out of control data, which may be found in the literature.

If the true mean ( $\mu$ ) and standard deviation ( $\sigma$ ) are not known the statistical parameters may be obtained from careful analyses of nonroutine samples to give results that may be accepted as within statistical control.



**Figure 5** A Shewhart means chart. The final two points are above the upper warning limit and would trigger suspension of the analysis and investigation of the process.

In addition to the mean being within statistical boundaries, the range of the data should also be monitored as a check of the process standard deviation. From the population standard deviation, or average range of a training set of data, it is possible to calculate expected ranges (i.e., maximum minus minimum values in the sample) and so construct upper and lower warning and control limits for the range. These are not symmetrical and must be obtained from tables.

**CuSum charts** In a CuSum chart, the cumulative sum of the difference between the mean result ( $\bar{x}_t$ ) and the target value ( $\mu_0$ ) is plotted against time. For a system in control, results are distributed about  $\mu$  and so the expectation of the difference is zero. A bias, however, will cause an accumulation of differences that can be detected. In the days of graph paper a V-shaped mask was placed over the graph with points outside the V indicating a system out of control. With calculations largely done in spreadsheets, the following calculations are performed. For a positive shift from the mean

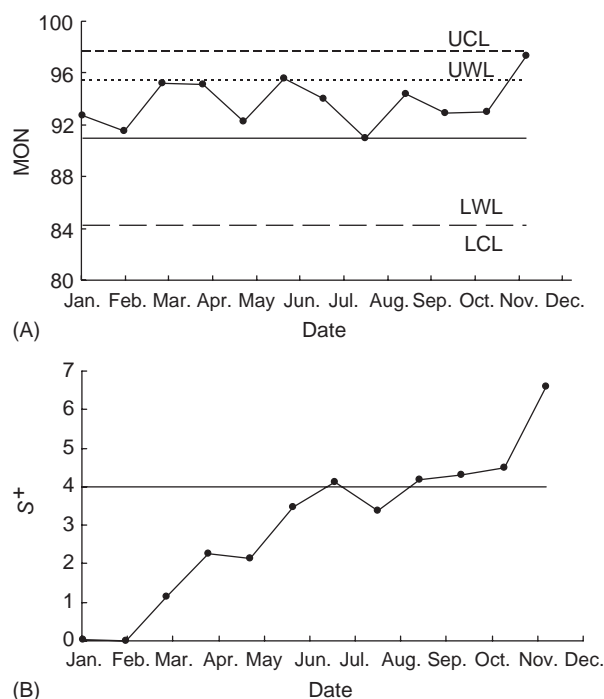
$$S_t^+ = \max \left\{ 0, S_{t-1}^+ + \frac{\bar{x}_t - (\mu_0 + k\sigma_{\bar{x}})}{\sigma_{\bar{x}}} \right\} \quad [3]$$

where  $\sigma_{\bar{x}}$  is the standard deviation of the mean,  $k$  the allowable change in one measurement period expressed in standard deviations of the mean (usually  $k=0.5$ ), and  $S_0=0$ . The equivalent equation for a negative shift is

$$S_t^- = \max \left\{ 0, S_{t-1}^- + \frac{\bar{x}_t - (\mu_0 + k\sigma_{\bar{x}})}{\sigma_{\bar{x}}} \right\} \quad [4]$$

$S_t^+$  and  $S_t^-$  are plotted against time, and when either goes over a threshold  $h=4$  the process is considered out of control. **Figure 6A** shows a Shewhart means chart for a measurement of motor octane number over a number of months. Although it appears that the results are consistently biased above the expected mean, apart from the last point, no result has triggered action based on the warning and control limits. A CuSum chart, however, does reveal a statistically significant trend (**Figure 6B**) when the value of  $S_t^+$  goes above  $h$  in July and then September and after.

**Average run length** A question that may arise in connection with the use of CuSum, mean, and range charts is ‘how soon will I discover that the mean has changed?’ An analysis that returns a higher value because of some change in the application of the method will have a greater probability of crossing the upper warning and control limits of a means chart



**Figure 6** Monthly data from analysis of a fuel for motor octane number (A) Shewhart means chart of duplicate results. The mean, warning and control limits are shown. (B) CuSum chart of  $S^+$  calculated from eqn [3], showing the limit  $h=4$  is exceeded in July.

than the expected 2.5/100 and 1.5/1000, respectively. The average run length (ARL) is how many points on average will follow a change before it is detected. A trivial example is that if the mean changes by  $3\sigma/\sqrt{n}$ , when it now sits on the upper control limit, on average it will take 1.5 points (i.e., 1 or 2 points) to go above the control limit. For a change of  $1\sigma/\sqrt{n}$  the ARL is 44. This highlights a characteristic of a means chart; it is much more sensitive to large changes than small changes in the process mean.

CuSum, as we have seen, is more responsive to a small bias in the results. The ARLs for different changes in the mean are tabulated in Table 1. When the shift of the mean reaches 2.5 standard deviations either method can detect the change quickly.

## Qualimetrics

In recent years, a branch of chemometrics devoted to quantitative methods of quality evaluation has been given the name qualimetrics. It covers the tools and approaches described in this article and takes an overall view of the quality issue, including customer perception as well as measurable quantities. Multivariate methods have also found their way into

**Table 1** Average run length rounded up to the next integer for a shift of the mean result by the given amount in numbers of standard deviations of the mean, for CuSum and Shewhart means charts

Shift in the mean ( $/\sigma_{\bar{x}}$ )	ARL (CuSum, $h=4$ )	ARL (Shewhart, $3\sigma_{\bar{x}}$ )
0	336	371
0.25	75	282
0.5	27	155
0.75	14	82
1.0	9	44
1.5	5	15
2.0	4	7
2.5	3	4
3.0	3	2
4.0	2	2

qualimetrics. Measurements that give large numbers of correlated data, for example, near-infrared, are ideal for multivariate methods such as principal components analysis. Soft modeling, as the use of these methods is sometimes called, can relate spectra to consumer parameters. In terms of the quality of analytical results themselves, rather than the quality of a product that is determined from analytical measurements, it is not clear that multivariate methods have such a great role to play.

**See also:** **Chemometrics and Statistics:** Statistical Techniques. **Quality Assurance:** Primary Standards; Spectroscopic Standards; Interlaboratory Studies; Reference Materials; Production of Reference Materials; Method Validation; Accreditation.

## Further Reading

- Garfield FM (1993) *Quality Assurance Principles for Analytical Laboratories*. Washington: AOAC International.
- Hawkins DM and Olwell DH (1998) *Cumulative Sum Charts and Charting for Quality Improvement*. New York: Springer.
- Kateman G and Buydens L (1993) *Quality Control in Analytical Chemistry*. New York: Wiley-Interscience.
- Martens H and Martens M (2001) *Multivariate Analysis of Quality – An Introduction*. New York: Wiley.
- Massart DL, Vandeginste BGM, Buydens LMC, de Jong S, Lewi PJ, and Smeyers-Verbeke J (1997) *Handbook of Chemometrics and Qualimetrics. Part A*. Amsterdam: Elsevier.
- Parkanay M (ed.) (1993) *Quality Assurance for Analytical Laboratories*. Cambridge: Royal Society of Chemistry.
- Parkanay M (ed.) (1995) *QA and TQM for Laboratories*. Cambridge: Royal Society of Chemistry.
- Shewhart SA (1931) *Economic Control of Quality Manufactured Product*. New York: Van Nostrand Reinhold.



## Primary Standards

**R Wielgosz**, Bureau International des Poids et Mesures, Sèvres, France

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

In science and technology, the word standard is used with two different meanings:

1. a measurement standard (etalon);
2. a widely adopted written technical standard, specification, technical recommendation, or similar document.

This article is concerned with the first meaning, and it should be noted that the qualifier 'measurement' is often omitted. The International Vocabulary of Basic and General Terms in Metrology (VIM) defines a measurement standard as:

material measure, measuring instrument, reference material (RM) or measuring system intended to define, realize, conserve or reproduce a unit or one or more values of a quantity to serve as a reference.

In order to ensure the metrological properties of measurement standards a hierarchy is established between them. A primary standard is defined as

a standard that is designated or widely acknowledged as having the highest metrological qualities and whose value is accepted without reference to other standards of the same quantity.

A secondary standard is defined as

a standard whose value is assigned by comparison with a primary standard of the same quantity.

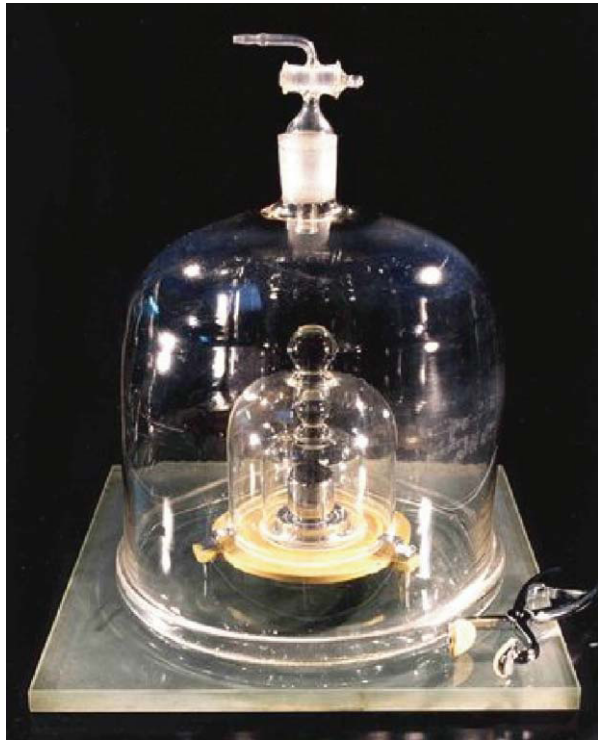
The property of standards or measurement results that allows them to be related to stated references, for example, a primary standard, through an unbroken chain of comparisons all having stated uncertainties is known as traceability. The unbroken chain of comparisons is referred to as a traceability chain (see Figure 1).

### Standards, Quantities, and Units

A measurable quantity is an attribute of a phenomenon, body, or substance that may be distinguished qualitatively and determined quantitatively. Quantities may be grouped together into categories of quantities that are mutually comparable. Lengths, diameters, distances, heights, wavelengths, and so on

would constitute such a category. Mutually comparable quantities are called 'quantities of the same kind'. If a particular example of a quantity from such a category is chosen as a reference quantity called the unit, then any other quantity from this category can be expressed in terms of this unit, as a product of this unit and a number. This number is called the numerical value of the quantity expressed in this unit.

A measurement standard is a physical realization of a quantity. The standard may represent a unit of measurement of the particular quantity, or a known value of the quantity.



**Figure 1** Primary standard of mass, the International prototype of the kilogram. The kilogram is the last remaining base unit of the SI that is still defined by a material artifact. The international prototype is kept with its six official copies in a vault at the BIPM. The international prototype was manufactured in the 1880s of an alloy of 90% platinum–10% iridium. Four of the six official copies date from the same period. In addition, copies of the international prototype have been manufactured by the Bureau International des Poids et Mesures (BIPM) for use as 1 kg national prototypes. The first of these were distributed in 1889. Since the 1880s the BIPM has produced more than eighty 1 kg prototypes in Pt/Ir. The BIPM also maintains a number of 1 kg copies for current use, one of which is illustrated in the figure. (Reproduced with the permission from International Organization for Standardization, ISO. This standard can be obtained from any ISO member and from the website, [www.iso.org](http://www.iso.org); © ISO.)

Measurement standards may be defined as primary or secondary depending on their metrological hierarchy. Additionally, they may be defined according to the role they fulfill. For example, an international measurement standard is recognized by an international agreement to serve internationally as the basis for assigning values to other standards of the quantity concerned. A national measurement standard is recognized by a national decision to serve, in a country, as the basis for assigning values to other standards of the quantity concerned. A reference standard is a measurement standard, generally having the highest metrological quality available at a given location or in a given organization, from which measurements made there are derived.

In the field of chemical measurement, certified reference materials (CRMs) and RMs are used as measurement standards. Depending on the metrological qualities of these standards, in certain fields CRMs provide a primary standard for a measurement system.

Pure substance CRMs provide the basis for many traceability chains in chemistry, and are often considered as primary standards for these measurement systems. The certification of substances for purity is an essential cornerstone of traceability in chemical measurements. These CRMs are used by laboratories to prepare calibration solutions or to certify or prepare other CRMs. The mass or amount fraction of the pure substance within the material is often determined by analyzing for possible impurities (often based on a consideration of the manufacturing process and degradation products) and the purity computed by difference.

## The International System of Units

A system of units (SI) (of measurement) is a set of base units, together with derived units, defined in accordance with given rules for a given system of quantities. The International system of units, SI, provides such a system, being the coherent system of units adopted and recommended by the General Conference on Weights and Measures (CGPM). A base unit is a unit of measurement of one of the quantities that, in a system of quantities, are conventionally accepted as functionally independent of one another. The seven base units of the SI are given in **Table 1**.

Formal definitions of all SI base units are approved by the CGPM. The first such definition was approved in 1889 and the most recent in 1983. These definitions are modified from time to time as techniques of measurement evolve and allow more accurate realizations of the base units. The definitions are

**Table 1** Base units of the SI

<i>Quantity</i>	<i>SI base unit</i>	
	<i>Name</i>	<i>Symbol</i>
Length	Meter	m
Mass	Kilogram	kg
Time	Second	s
Electric current	Ampere	A
Thermodynamic temperature	Kelvin	K
Amount of substance	Mole	mol
Luminous intensity	Candela	cd

Reproduced with permission from BIPM, [www.bipm.org](http://www.bipm.org).

published in the SI brochure by the Bureau International des Poids et Mesures (BIPM), which also lists CGPM decisions concerning the practical realization of the definitions of some key SI units. The framework within which standards laboratories must work if units they realize are to be in conformity with those defined by the SI is also presented there. The current definitions of the SI base units are summarized in **Table 2**.

## International Metrology

The Convention of the Metre is a diplomatic treaty that gives authority to the General Conference on Weights and Measures (Conférence Générale des Poids et Mesures, CGPM), the International Committee for Weights and Measures (Comité International des Poids et Mesures, CIPM), and the International Bureau of Weights and Measures (Bureau International des Poids et Mesures, BIPM) to act in matters of world metrology, particularly concerning the demand for measurement standards of ever increasing accuracy, range, and diversity, and the need to demonstrate equivalence between national measurement standards.

The Convention was signed in Paris in 1875 by representatives of 17 nations. As well as founding the BIPM, the Metre Convention established a permanent organizational structure for member governments to act in common accord on all matters relating to units of measurement. The task of the BIPM is to ensure worldwide uniformity of measurements and their traceability to the International System of Units (SI).

The Convention, modified slightly in 1921, remains the basis of international agreement on units of measurement. There are now 51 Member States, including all the major industrialized countries.

The CIPM is advised by Consultative Committees in various areas of metrology. The Consultative Committee for Amount of Substance; Metrology in



**Table 2** Current definitions of the SI base units

Meter	The meter is the length of the path traveled by light in vacuum during a time interval of $1/299\,792\,458$ of a second
Kilogram	The kilogram is the unit of mass; it is equal to the mass of the international prototype of the kilogram
Second	The second is the duration of $9\,192\,631\,770$ periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium 133 atom
Ampere	The ampere is that constant current that, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 m apart in vacuum, would produce between these conductors a force equal to $2 \times 10^{-7}$ newton per meter of length
Kelvin	The kelvin, unit of thermodynamic temperature, is the fraction $1/273.16$ of the thermodynamic temperature of the triple point of water
Mole	The mole is the amount of substance of a system that contains as many elementary entities as there are atoms in $0.012$ kilogram of carbon 12. When the mole is used, the elementary entities must be specified and may be atoms, molecules, ions, electrons, other particles, or specified groups of such particles
Candela	The candela is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency $540 \times 10^{12}$ hertz and that has a radiant intensity in that direction of $1/683$ watt per steradian

Reproduced with permission from BIPM, [www.bipm.org](http://www.bipm.org).

Chemistry (CCQM) is responsible for issues related to metrology in chemistry.

At the national level, National Metrology Institutes (NMIs) maintain measurement standards and capabilities and provide services in order to develop their national metrology infrastructures so as to underpin and facilitate the comparability of measurement results.

In October 1999, the directors of the NMIs of the industrialized states of the world signed a Mutual Recognition Arrangement (MRA) for national measurement standards and for calibration and measurement certificates issued by their institutes. This arrangement was drawn up by the International Committee of Weights and Measures (CIPM), under the authority given to it in the Metre Convention.

The objectives of the CIPM-MRA are to: establish the degree of equivalence of measurement standards maintained by NMIs; to provide for the mutual recognition of calibration and measurement certificates issued by NMIs; thereby to provide governments and other parties with a secure technical foundation for wider agreements related to international trade, commerce, and regulatory affairs. It is founded on the efforts of each individual NMI to base its measurements and measurement uncertainties on SI units.

The principal output of this MRA is a key comparison database (KCDB) at the Bureau International des Poids et Mesures ([www.bipm.org/kcdb](http://www.bipm.org/kcdb)) containing the results of international comparisons of national measurement standards and capabilities, known as key comparisons, which underpin and extensive lists of the calibration and measurement capabilities (CMCs) of NMIs. The CMCs of NMIs for chemical measurements are included in the KCDB, and listed under Amount of Substance. The chemical categories covered are: high-purity chemicals, inorganic

solutions, organic solutions, gases, water, metals and metal alloys, advanced materials, biological fluids and materials, food, fuels, sediments, soils, ores and particulates, pH and electrolytic conductivity, and other materials (see Figure 2).

### Degrees of Equivalence of Measurement Standards and Measurement Capabilities

The international comparisons of the CIPM-MRA result in a determination of the degree of equivalence of measurement standards and capabilities. The degree of equivalence of measurement standards is taken to mean the degree to which a standard is consistent with the reference value of a key comparison. The degree of equivalence of a national measurement standard is expressed quantitatively by two terms: its deviation from the reference value of a key comparison and the uncertainty of this deviation. The CCQM is active in coordinating international (key) measurement comparisons in the field of chemical measurements in the areas of organic analysis, inorganic analysis, gas analysis, electrochemical analysis, bioanalysis, and surface analysis. The results of these comparisons are publicly available and include the measurement results and their uncertainties, and a calculation of the degrees of equivalence between the measurements of the various laboratories that participate in the comparison. As an example, the graph of equivalence for a pH comparison ( $0.025 \text{ mol kg}^{-1} \text{ KH}_2\text{PO}_4 + 0.025 \text{ mol kg}^{-1} \text{ Na}_2\text{HPO}_4$  at  $15^\circ\text{C}$ ) is shown (Figure 3). The Physikalisch-Technische Bundesanstalt (PTB) piloted this comparison with the assistance of the National Institute of Standards and Technology (NIST) and the Slovak Institute for Metrology (SMU). The final

Gases, Environmental

**China, NRCCRM (National Research Center for Certified Reference Materials)**

Complete CMCs in Amount of Substance for Gases for China (.pdf file)

Matrix or material	Analyte or component	Range of certified values in reference materials	
		Amount-of-substance fraction in $\mu\text{mol/mol}$	Relative expanded uncertainty in %
nitrogen	carbon monoxide	5000 to 80000	1

Mechanism(s) for measurement service delivery: GBW08137

Internal NMI service identifier: NR CCRM/GAS 4

Gases, Environmental

**Finland, FMI (Finnish Meteorological Institute)**

Complete CMCs in Amount of Substance for Gases for Finland (.pdf file)

Matrix or material	Analyte or component	Dissemination range of measurement capability	
		Amount-of-substance fraction in $\mu\text{mol/mol}$	Absolute expanded uncertainty in $\mu\text{mol/mol}$
nitrogen	carbon monoxide	0.2 to 50	0.2 to 1.2

Mechanism(s) for measurement service delivery: Calibration

Uncertainty convention 2. No range of certified values in reference materials has been declared

Internal NMI service identifier: FMI/ PC-CO

Gases, Environmental

**United States, NIST (National Institute of Standards and Technology)**

Complete CMCs in Amount of Substance for Gases for United States (.pdf file)

Matrix or material	Analyte or component	Dissemination range of measurement capability		Range of certified values in reference materials	
		Amount-of-substance fraction in $\mu\text{mol/mol}$	Relative expanded uncertainty in %	Amount-of-substance fraction in $\mu\text{mol/mol}$	Relative expanded uncertainty in %
nitrogen	carbon monoxide	1 to 1.5E+05	0.5 to 1	10 to 130000	0.5 to 1

Mechanism(s) for measurement service delivery: NTRM, SRM 1677, SRM 1678, SRM 1679, SRM 1680, SRM 2636, SRM 2637, SRM 2638, SRM 2639, SRM 2640, SRM 2641, SRM 2642, SRM 2740, SRM 2741

Uncertainty convention 1

**Figure 2** An example of a calibration and measurement capability for gas standards. (Reproduced with permission from BIPM, www.bipm.org.)

report of the comparison contains full details of the measurements, and this together with statements on how the degree of equivalences is derived is found on the KCDB, allowing the standards or measurement capabilities to be compared.

## Primary Methods

The CCQM defined a primary method as ‘a method having the highest metrological qualities, whose operation can be completely described and understood, for which a complete uncertainty statement can be

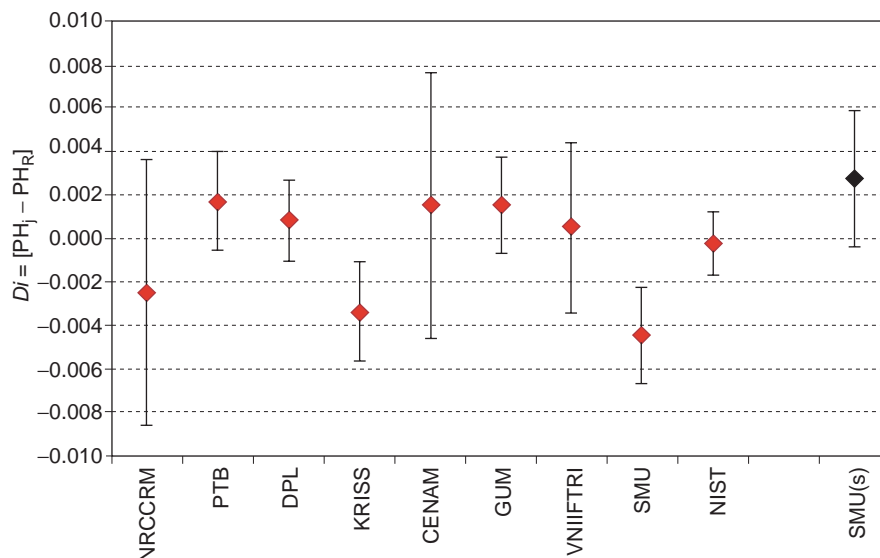
written down in terms of SI units’. A primary direct method: ‘measures the value of an unknown without reference to a standard of the same quantity’. A primary ratio method: ‘measures the value of a ratio of an unknown to a standard of the same quantity; its operation must be completely described by a measurement equation’.

A primary method provides the means to transform the abstract definition of an SI unit into practical measurements made in terms of that unit. This is sometimes referred to as a realization of that unit. The CCQM identified several methods with the potential of being primary methods of

MEASURAND: pH value of phosphate buffer  
 Sample 1:  $[0.025 \text{ mol kg}^{-1} \text{ KH}_2\text{PO}_4 + 0.025 \text{ mol kg}^{-1} \text{ Na}_2\text{HPO}_4]$   
 Measurements at 15°C

Nominal value: pH = 6.9 at 25°C

Degrees of equivalence  $D_i$  and expanded uncertainty  $U_i$  ( $k = 2$ )



◆ Indicates the degree of equivalence for SMU resulting from the subsequent bilateral comparison between SMU and PTB.

**Figure 3** Graph of equivalence for an international comparison of primary pH measurements. (Reproduced with permission from BIPM, [www.bipm.org](http://www.bipm.org).)

measurement: isotope dilution mass spectrometry, coulometry, gravimetry, titrimetry, and determination of freezing point depression.

The CCQM is not the only international organization that has established definitions for different types of measurement methods. Others include the International Union of Pure and Applied Chemistry (IUPAC), the International Organization for Standardization's Committee on Reference Materials (ISO-REMCO), and the International Federation of Clinical Chemistry (IFCC). The definitions of absolute and definitive measurements are similar to the definition of a primary method, but neither makes reference to the use of SI units or to the traceability of the results.

## Traceability and Uncertainty

Within a robust metrological system the values of measurement standards and measurement results are linked via comparisons or calibrations that take into account the measurement uncertainty of the linking

processes. Measurement uncertainty is the parameter associated with the results of a measurement, which characterizes the dispersion of the values that could reasonably be attributed to the measurand. The ISO Guide to the Expression of Uncertainty in Measurement (GUM) and the Eurachem guide on measurement uncertainty provide guidance of the evaluation of measurement uncertainty. The property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties, is termed traceability. Where these stated references are realizations of the SI units the term SI-traceable is used.

## Quantities Describing the Composition of Mixtures

Measurements in analytical chemistry are often performed on mixtures. The quantities used for expressing the composition of mixtures can be classified

into five groups: ratios, fractions, concentrations, molality, and contents. Since a number of these quantities are dimensionless the kind of quantity being measured should always be stated to avoid misinterpretation of measurement results.

## Certified Reference Materials

In the field of analytical chemistry the term certified reference material is more often used than measurement standard. A CRM is a RM, accompanied by a certificate, one or more of whose property values are certified by a procedure that establishes traceability to an accurate realization of the unit in which the property values are expressed, and for which each certified value is accompanied by an uncertainty at a stated level of confidence. All CRMs lie within the definition of 'measurement standard' previously given in this article.

The term reference material is used more generally to describe a material or substance one or more of whose property values are sufficiently homogeneous and well established to be used for the calibration of an apparatus, the assessment of a measurement method, or for the assigning of values to materials.

Some RMs and CRMs have properties that cannot be determined by exactly defined physical and chemical measurement methods. Biologicals provide an example where this is the case, where a measurement of biological activity is made relative to an internationally accepted artifact standard, to which an International unit (IU) has been assigned by the World Health Organization.

CRMs are required in many chemical analyses, in order to perform a calibration and/or validate the method being employed to ensure the accuracy, uncertainty, and traceability of the results obtained. Many modern analytical methods require the use of RMs the matrix of which is sufficiently close to the sample being analyzed to enable calibration and appropriate verification of the measurement method used.

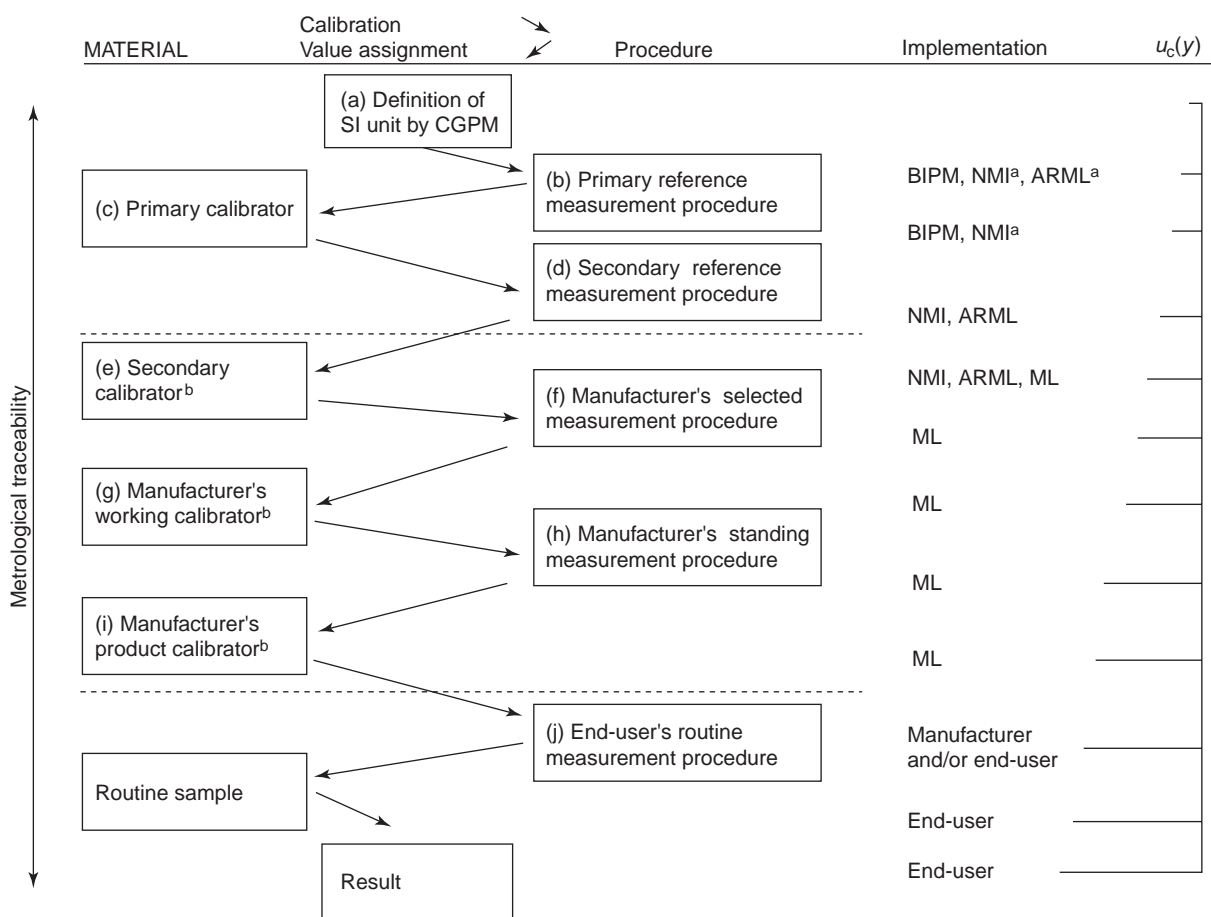
The COMAR (index Code of Reference Materials) database, initiated in the late 1970s, and now freely searchable via the internet ([www.comar.bam.de](http://www.comar.bam.de)), contains information on some 10 000 CRMs from more than 200 producers from throughout the world. The database was developed to aid chemists in finding the RMs they required. The database provides the following information on RMs: their basic area(s) of application, their certified values, their form, their country of origin, the producer, and relevant references.

The Reference Materials Committee (REMCO) of the International Organization for Standardization (ISO), formed in 1975, has been active in developing a series of guides to support best practice in the production and certification of RMs. The committee has developed two guides to assist in the establishment of facilities to produce and certify RMs and ensure their quality. ISO Guide 34 outlines the requirements to be met by a CRM producer to demonstrate competence, and ISO Guide 35 provides guidance on general and statistical principles related to the certification of reference materials, and cites models for the homogeneity testing, stability testing, and the characterization of a candidate CRM.

## Reference Measurement Systems

The concept of metrological traceability is particularly developed in a number of chemical measurement fields. In the field of laboratory medicine, ISO 17511 (*in vitro* diagnostic medical devices – Measurement of quantities in biological samples – Metrological traceability of values assigned to calibrators and control materials), describes reference measurement systems, which ensure traceability to the SI via reference measurement procedures and certified reference measurements. A calibration hierarchy with traceability to the SI is depicted in **Figure 4**. Depending on the possibility of metrological traceability to the SI, ISO 17511 envisages five typical upper ends of the metrological traceability chain:

1. Quantities for which results of measurement are metrologically traceable to SI.
2. Quantities for which results of measurements are not metrologically traceable to the SI, amongst which it is possible to differentiate:
  - a. quantities for which an international conventional reference measurement procedure and one or more conventional calibration materials with values assigned by these procedures are available;
  - b. quantities for which an international conventional reference measurement procedure is available but no conventional calibration materials are available;
  - c. quantities for which one or more conventional calibration materials with a protocol for value assignment are available, but for which there is no international conventional reference measurement procedure;
  - d. quantities for which neither reference measurement procedures nor reference materials are available.



**Figure 4** Calibration hierarchy and metrological traceability to SI (The terms and definitions taken from ISO 17511: 2003 *In vitro* diagnostic medical devices – Measurement of quantities in biological samples – Metrological traceability of values assigned to calibrators and control materials.) (Reproduced with permission from International Organization for standardization, ISO. This standard can be obtained from any ISO member and from the website, [www.iso.org](http://www.iso.org); © ISO.)

See also: **Quality Assurance:** Quality Control; Internal Standards; Instrument Calibration; Reference Materials; Traceability; Laboratory Information Management Systems.

## Further Reading

- BIPM (1998) *The International System of Units (SI)*, 7th edn. Sevres: Bureau International des Poids et Mesures (<http://www.bipm.org/en/si/>).
- Cvitas T (1996) Quantities describing compositions of mixtures. *Metrologia* 33: 35–39.
- ISO (1989) *Certification of Reference Materials – General and Statistical Principles*, Guide 35. Geneva: International Organization for Standardization (under revision 2003).
- ISO (1993) *ISO Standards Handbook (ISO 31)*, 3rd edn. Geneva: International Organization for Standardization.

ISO (1993) *The International Vocabulary of Basic and General Terms in Metrology*, 2nd edn. Geneva: International Organization for Standardization.

ISO (1995) *Guide to the Expression of Uncertainty in Measurement*. Geneva: International Organization for Standardization.

ISO (2000) *General Requirements for the Competence of Reference Material Producers*, Guide 34. Geneva: International Organization for Standardization.

ISO 17511 (2003) *In vitro Diagnostic Medical Devices – Measurement of Quantities in Biological Samples – Metrological Traceability of Values Assigned to Calibrators and Control Materials*. Geneva: International Organization for Standardization.

Mills I (ed.) (1988) *Quantities, Units, and Symbols in Physical Chemistry*, IUPAC. Oxford: Blackwell Scientific Publications.

Milton MJT and Quinn TJ (2001) Primary methods for the measurement of amount of substance. *Metrologia* 38: 289–296.



## Spectroscopic Standards

**J P Hammond**, Optiglass Ltd., Hainault, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Consider the following five observations?

Today, thanks to the genius of researchers, manufacturers, and commerce, there are more products on sale than ever before in the history of mankind. Increasing complexity is leading to a greater need for consumer protection. More and more products require testing to ensure compliance with specification and safety regulations before release to the market.

The growth in international trade, with its geographical variety of traditions, languages, standards, and habits, is adding to this need for consistency. Trade, even in simple commodities and products, demands supporting technical data, only available from analytical measurement. The achievement of the Single European Market and the GATT Agreement requires cross-frontier acceptance of test results.

The enforcement of regulation and often the administration of justice in the fields of health, safety, and environmental protection need to be based on a foundation of valid analytical measurement in which the public has confidence.

Competent calibration and traceability of measurements is essential for industrial manufacture and a major criterion in addressing product liability. It features prominently in both accreditation to ISO/IEC 17025 and certification to ISO 9001. It also underpins all testing.

The credibility of test results depends on effective method validation, which includes amongst its requirements accuracy, precision, repeatability, and reproducibility. In turn, these depend on the competence of the tester and the validity of the methods used. Those who have to accept goods must have confidence that laboratories conducting tests and calibrations are competent and that their results are valid.

In all the cases, effective quality assurance procedures can be seen as the tools that initially provide the confidence, and then maintain the compliance.

### General Requirements

Each area of science has its own specific vocabulary, and before proceeding further, it is worth considering how these definitions and principles are implemented when spectroscopic measurements are undertaken.

Often considered the primary responsibilities of the laboratory manager is the production of analytical results that are 'fit for purpose'; but again this phrase is open to interpretation.

Every result has a customer for whom the data are being generated, and addresses a specific need; e.g., to check for compliance with a specification or regulation, to monitor health, or to identify the cause of a plant malfunction. The data will be 'fit for purpose' provided it enables the customer to meet his or her needs.

Every analysis is undertaken in order to solve a problem of one type or another. The analyst needs to understand the problem and discuss it with the customer (i.e., the person requiring the data) so that an appropriate analytical procedure can be applied. It will also be necessary to take into consideration other factors that may be important to the customer. The analyst then uses all of this information to prepare a specification of the analytical requirement. The issues that need to be taken into account in preparing the specification may include:

- the nature of the samples and the way in which they were obtained;
- the information required to solve the problem;
- the criticality/acceptable risk for the analysis;
- time and cost constraints;
- bias (accuracy), precision, and degree of confidence required;
- identification or confirmation of the analyte;
- method development and validation requirements (where applicable);
- quality assurance (QA)/quality control (QC) requirements;
- regulatory or legal constraints; and
- third-party acceptance requirements (e.g., for documentation or traceability of data).

The specification should also include agreed acceptance criteria for the data. Unfortunately, disregard of the requirement to discuss and agree on the specification often results in the use of 'assumed' criteria by one or both parties. This may lead to the adoption of methodology, QA procedures, or documentation inappropriate to the purpose for which the data are required.

Increasingly, the chemical testing community is adopting QA principles, which, whilst not actually guaranteeing the quality of the data produced, increase the likelihood of it being soundly based. A quality system on its own will not, however, achieve



the desired aim unless the laboratory also adheres to 'good analytical practice'. This simple phrase embodies all the activities of the laboratory, whether technical, managerial, or quality assurance, which are needed to ensure valid results.

Good analytical practice is the basis of all the GxP ( $x = L, M$ , etc.) or cGxP standards long since mandatory in the pharmaceutical industry, and now finds an equally willing host in accreditation to ISO/IEC 17025. This new role is being driven in no small part by the conversion of ISO guide 25 in recent years to ISO/IEC standard 17025.

The International Standards Organization (ISO) clearly defines the difference between QA and QC, but these terms are sometimes mistakenly used interchangeably.

### Quality Assurance

"All those planned and systematic actions necessary to provide adequate confidence that a product or service will satisfy given requirements for quality." (ISO 8402 – 1994, 3.6)

QA describes the general quality procedures/requirements for a laboratory. These may include:

- educated, trained, and skilled staff;
- specified requirements for all chemical grades from reference material to reagent;
- suitable equipment calibrated and maintained;
- suitable laboratory environment;
- documented and validated methods;
- quality control procedures;
- check procedures for reporting and quality control;
- use of proficiency testing;
- quality system supported by a quality manager and documentation;
- a complaints procedure.

### Quality Control

"The operational techniques and activities that are used to fulfill requirements for quality." (ISO 8402 – 1994, 3.7)

Quality control procedures ensure the reliability of specific samples or batches of samples and may include routine analysis of:

- quality control samples (with results recorded on control charts);
- blanks;
- replicates (duplicate/triplicate);
- reference materials;
- 'spiked' samples;

- blind samples (i.e., samples where the laboratory staff are unaware that the result is already known).

### 'Standard' Definitions

The word standard and its use in the context of quality assurance of spectroscopic measurements can be broadly defined in three very different ways, each adding its own suffix to the term 'Standard'. These definitions can be broadly described as calibration, documentation, and reference – and such are the vagaries of the English language that these can often be subdivided further. However, each has its own vital contribution to the overall quality assurance of the spectroscopic measurements performed, and it is up to the laboratory manager to decide on the type or types required and their roles in ensuring that the data is indeed fit for purpose.

### Standard Calibration

Thanks to the Beer–Lambert law, which defines a linear relationship between absorbance ( $A$ ) and analyte concentration, the majority of analytical spectroscopic methods are quantitative in nature. These methods require the manufacture of a series of standards of increasing concentration from which a calibration function is derived, usually involving regression analysis for 'goodness of fit'.

From this method calibration, simple extrapolation of a measurement unknown will produce a concentration estimate. However, from a QA perspective, day-to-day comparisons of these simple calibration measurements, and the associated statistics, can often highlight aspects of reagent quality, apparatus (balance, volumetric glassware, etc.), and operator competence when compared. These data can be invaluable in assisting regulatory compliance, and ensuring that the complete measurement system is under control.

Whilst these standards are mostly prepared simply within the solvent matrix used, variations on this theme may result in inclusion of these standards actually within the sample matrix. These procedures are often referred to as internal standardization, standard addition, or standard recovery, depending on the actual implementation of how the known analyte addition is used within the spectrophotometric method. These procedures either provide spectral calibration or method quality assurance data. For example, spectroscopic methods will often require pretreatment to remove the interfering sample matrix. Addition of a known amount of the test analyte 'spike' (at the expected level) to a sample is

then subjected to the same pretreatment. This is then analyzed at the same time as the sample, the difference calculated, and the resultant calculated as a percentage of the original level. Clearly, anything significantly less than 100% indicates significant loss of analyte at the pretreatment stage.

### **Standard Documentation**

**Standard regulations** Briefly mentioned already, these standards all have the common requirement that, where possible, QA with traceable certified reference materials (CRMs) is required. Traceability is an essential requirement for compliance to these standards, and is fully covered elsewhere.

*Pharmaceutical 'good practices' – GxP ( $x = L, M$ , etc.) or cGxP standards* Formally introduced in the late 1970s by the FDA in the USA, these regulations were used as a basis to construct the internationally recognized Organization for Economic Cooperation and Development guidelines, issued in the early 1980s.

Broadly speaking, these standards require the submitting laboratory to provide 'evidence of control'.

*ISO 9001:2000* ISO develops and maintains the ISO 9000 series of International Standards for quality management and QA. First introduced in 1987, this series of standards has been employed by thousands of businesses worldwide and has been adopted by more than 70 countries as the national standard.

Fundamentally reviewed in 2000, this standard now places greater emphasis on customer communication and continuous improvement.

*ISO/IEC 17025:1999* Accreditation is the formal recognition of the competence of a body or an organization for a well-defined purpose. In contrast with the ISO 9001 processes, accreditation of a laboratory to ISO/IEC 17025 involves assessment of the technical competence and capability of the laboratory and its personnel. In practice, it is the procedure by which a laboratory is assessed to perform a specific range of tests or calibration measurements. Specific areas examined include infrastructure and staff qualifications, in addition to checks that an adequate quality management scheme is in place. The accreditation covers the range of materials tested or analyzed, the tests carried out, the method and equipment used, and the accuracy or precision expected, and is specific to the facility and the test.

Most national laboratory accreditation schemes have adopted this standard, which establishes the process not only as third-party audit, but an

assessment of the data generation process by a 'peer review'.

**Method** Any appropriate analytical method may be used providing it is suitable for the purpose intended, adequately validated, properly documented, and authorized by a competent person.

In many application areas, too numerous to mention, internationally recognized bodies such as ASTM International, AOAC, etc., have volumes of published standard methods. The reliability of a 'standard' method should not be taken for granted, no matter how impeccable the method's pedigree. The laboratory applying the method should satisfy itself that the degree of validation of the method is adequate for the specific purpose, and confirm that it is able to match any stated performance criteria. Validation of a method establishes, by systematic laboratory studies and/or appraisal of existing validation data, that the method is fit for purpose, i.e., its performance characteristics are appropriate to the requirements set out in the analytical specification.

### **Standard Reference**

Defined as a physical artifact, these standards may be further subdivided into the following groups.

**Reference material** ISO defines a reference material as 'a material or substance one of more of whose property values are sufficiently homogeneous and well established to be used for the calibration of an apparatus, the assessment of a measurement method, or for assigning values to materials'. In spectroscopic applications, these materials are usually substances with known, and proven spectral characteristics. They are invaluable in providing spectral data for either qualitative comparison by spectral matching or spectrophotometer qualification.

**Certified reference material** ISO defines a certified reference material as 'a reference material, accompanied by a certificate, one or more whose property values are certified by a procedure which establishes traceability to an accurate realization of the unit in which the property values are expressed, and for which each certified value is accompanied by an uncertainty at a stated level of confidence'.

These references are mandatory for regulatory compliance, and contain the essential certified value via a traceable route. Equally important from a QA perspective is a defined uncertainty budget. Given this budget figure, and the stated accuracy of the spectrophotometer (assuming the same level of confidence), simple addition of these two uncorrelated

functions will give the QA tolerance for use on the spectrophotometer of choice.

**Proficiency testing reference** Independent technical assessment of a laboratory's data provides an external benchmark, which enables the laboratory to monitor its own performance against other competent laboratories and to demonstrate its achievement of reliable results to customers. One of the best ways for an analytical laboratory to monitor its performance, both internally and externally, is to participate in proficiency testing (PT) schemes. These usually involve an organizing body that distributes an 'unknown' test sample to participating laboratories and then collates and assesses their results. Correctly used, these schemes help to highlight the success of a laboratory's quality procedures. Regular participation in PT schemes, and an associated improvement in performance, provides a means for laboratories to demonstrate their proficiency to third-party certification or accreditation organizations.

PT schemes or other types of interlaboratory comparison provides the analytical community with an alternative means to demonstrate consistency between a relatively small number of laboratories. Where the value of the test sample is independently determined these schemes also highlight systematic errors, i.e., bias. Including CRMs or results from recognized expert laboratories in a PT scheme or intercomparison is one way in which reference material producers establish certified values for new materials.

## Spectrophotometer Qualification and Calibration

The requirement that equipment should be tested and proven to be fit for purpose is effectively fulfilled by an approach known as equipment qualification (EQ), which is well established and proven in the highly regulated pharmaceutical industry. EQ is complementary to method validation for the spectrophotometer and is built on effective instrumental calibration validation (see below). EQ is a systematic procedure, which ensures that suitable equipment is purchased and that it remains fit for its chosen purpose throughout its operating life. The same requirement is present in ISO/IEC 17025, where section 5.5.5 'Records of Equipment' requires checks to ensure compliance with the specification, maintenance plans, etc.

### Calibration

Install and use any equipment in the environment of choice, and immediately the surroundings will begin

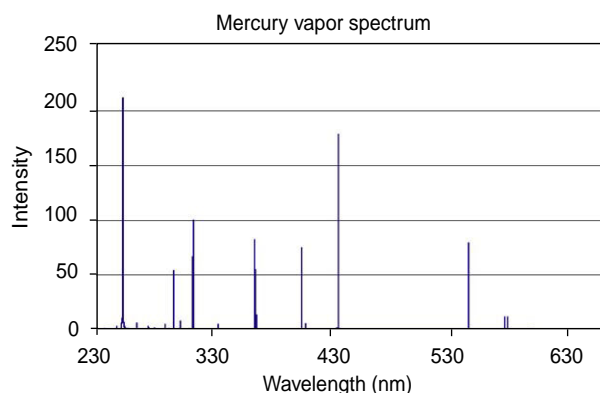
to act on that equipment causing change. Ultimately this process results in an unacceptable level of degradation in performance. This is commonly called drift, and when this happens the results become unreliable and no longer 'fit for purpose'. Whilst it may not be possible to eliminate drift, it can be detected and 'contained' using calibration. Calibration is the process of establishing how the response of a measurement process varies with respect to the parameter being measured.

CRMs are widely used to ensure both proper calibration and satisfactory use of the analytical methodology. When many laboratories use specific CRMs in a systematic way as part of QC procedures, consistency of data between those laboratories will be achieved. The main limitation in this respect is the frequent difficulty of obtaining CRMs appropriate to each type of analysis. This lack of supply is being addressed by an increasing number of secondary material producers, some of whom have already achieved ISO/IEC 17025 accreditation for production and certification of these references.

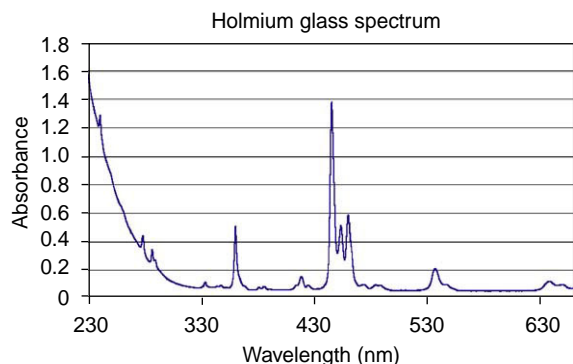
In the case of any spectrophotometric measurement, amongst the many parameters to keep in control, the three most usually measured are wavelength accuracy, linearity or absorbance accuracy, and stray light.

**Wavelength** There are several methods for checking wavelength accuracy. By far the best is to use an elemental line source. For analytical purposes, these sources can be deemed to be primary references because the line wavelengths are available as physical constants, with the wavelength determined to several orders of magnitude higher than required. For example, the principal line values in air for mercury (Hg) are established as: 253.6506, 296.7283, 312.5670, 365.0157, 404.6572, 435.8337, 546.0753, 576.9598, and 579.0663 nm (**Figure 1**). By definition, where accuracy to one or two decimal places is all that is required, these essentially become absolute wavelength calibration values. Not all instruments have the facility for fitting and using such a lamp, but recent years have seen this facility made available in top-of-the-range reference spectrophotometers. The deuterium lamp, fitted in most ultraviolet (UV) instruments, can be used in a similar manner, but it only provides two lines, at 486.00 and 656.10 nm. However, the 656.10 nm line is so intense that the second- and third-order harmonics at 1312.20 and 1968.30 nm, respectively, can be used for calibration in the near-infrared (NIR) region.

If line sources are not available, the spectra of the rare earth ions, in either liquid (as solution) or solid (as glass), can be used as useful references. In either



**Figure 1** Line spectrum produced from a mercury 'pen' lamp fitted as an alternative source in a high-performance reference commercial spectrophotometer.

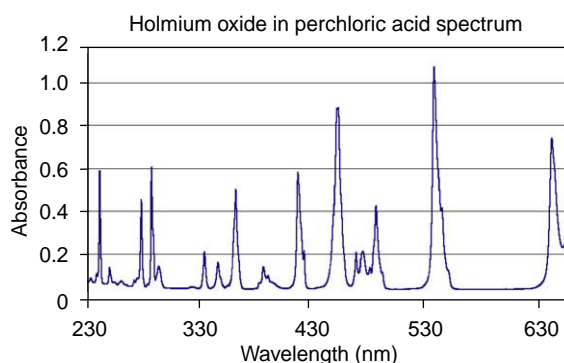


**Figure 2** Absorption spectrum produced by a holmium glass certified reference material measured at the sample position in a high-performance reference commercial spectrophotometer. The singlet peak at  $\sim 360$  nm and the triplet centered at  $\sim 450$  nm are particularly characteristic of this material.

form, these elements provide characteristic spectral information right from the far-UV region (200 nm), through visible, and into the NIR ( $3\text{ }\mu\text{m}$ ) region, with the glasses exhibiting the not unexpected peak broadening associated with the matrix (Figure 2). Holmium oxide in perchloric acid solution is a favorite for the 230–660 nm region, and is available from many commercial vendors, and as a primary reference from National Institute of Standards and Technology USA (NIST) as SRM 2034 (Figure 3).

**Absorbance** There are several methods for checking either the linearity or absorbance accuracy of an instrument.

The double aperture method is a physical, not a chemical, technique that at first seems to avoid many of the disadvantages of other methods. It relies on a plate with two independently operated shutters



**Figure 3** Absorption spectrum produced by holmium oxide in perchloric acid solution, provided as a 'sealed cell' certified reference material. Measured at the sample position, in a high performance reference commercial spectrophotometer, with air as the reference.

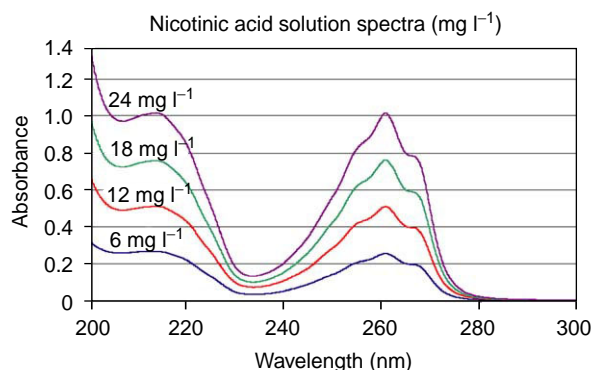
placed at a focal portion of the light beam and so is totally wavelength neutral. The United Kingdom's National Physical Laboratory has described it as 'the most accurate, sound, simple and fool-proof technique we know for testing linearity in a manual single beam dc mode instrument'.

However, the system described requires precision optical control and alignment, only found in high-performance commercial reference spectrometers such as those employed or constructed by National Metrology Institutes (NMIs). The double aperture method, therefore, is not ideal for the routine checking of normal spectrometers.

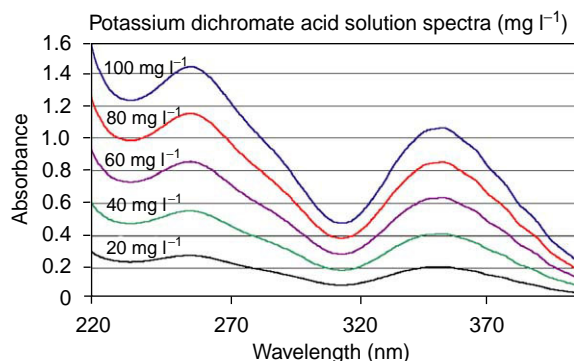
At the far-UV end of the spectrum (200–300 nm), nicotinic acid solution in the  $0\text{--}24\text{ mg l}^{-1}$  range provides a suitable stable reference (Figure 4).

In many application areas, the use of solutions of potassium dichromate is specified for checking absorbance accuracy. The NIST supply potassium dichromate (SRM 935a) as a solid material, from which the user prepares solutions in accordance with NIST Special Publication 260-54 (Figure 5). This process, as does any similar preparative procedure, therefore, relies on the laboratory chemist making up the solutions correctly, and is as much a test of his or her competence as it is of the instrument's performance. 'Preparative' used in this context can simply refer to the act of decanting a supplied bottled or ampouled solution into the measurement cuvette. This method can be extremely useful if one wishes to establish the total competence of a laboratory, but not so if the aim is to establish the calibration of the spectrophotometer only. If results are obtained that appear to question the accuracy of the spectrophotometer, is the instrument or the accuracy of the solution in question? To offer a remedy to this conundrum a few reference material producers now





**Figure 4** Absorption spectra produced by a series of acidic nicotinic acid solutions, provided as 'sealed cell' certified reference materials. Measured relative to a solvent blank, at the sample position, in a high-performance reference commercial spectrophotometer.



**Figure 5** Absorption spectra produced by a series of acidic potassium dichromate solutions, provided as 'sealed cell' certified reference materials. Measured relative to a solvent blank, at the sample position, in a high-performance reference commercial spectrophotometer.

offer sealed cells of potassium dichromate solution, prepared in perchloric acid to improve the stability, and under clean room conditions to minimize contamination.

Neutral density filters made of gray glass have been in use for over 30 years as absorbance reference materials and have shown themselves to be extremely stable and ideal for use in the visible portion of the spectrum, where they are now employed as the definitive reference. However, interest is now being expressed in extending the use of these glasses into both the NIR and mid-IR regions. They are available from NIST (as SRM 930, 1930, and 2030) or as secondary CRMs from a number of manufacturers. Care must be taken in mounting the filters, in manufacturing them to exacting standards, particularly with respect to parallelism of the optical faces, and in certifying

their absorbance values. The main disadvantage of these filters is that they cannot be used in the UV.

Metal on quartz filters have been proposed in recent years as an alternative method for checking absorbance accuracy. They are usable over a wide wavelength range, and have good thermal stability. However, their optical homogeneity is poor (1 in 100 instead of the required 1 in 1000), and their major disadvantage is that they work by reflecting light rather than by absorbing it. Reflected light can hit other optical surfaces and in turn be reflected back by them. Significant interreflection errors can occur. In fact, a major use of these filters is to check for such errors.

**Stray light** Stray light is one of the most difficult of the parameters to check. It is extremely wavelength and sample dependent, so any checks must be qualified. In a modern instrument employing master holographic gratings and coated optics, stray light is usually not an issue when the instrument is first manufactured. However, as the spectrophotometer optics and/or sources age and degrade, the instrumental stray light will increase with time. It is for this reason that from a QA perspective the often-neglected stray light check will give an overall 'health check' on the performance of the spectrophotometer.

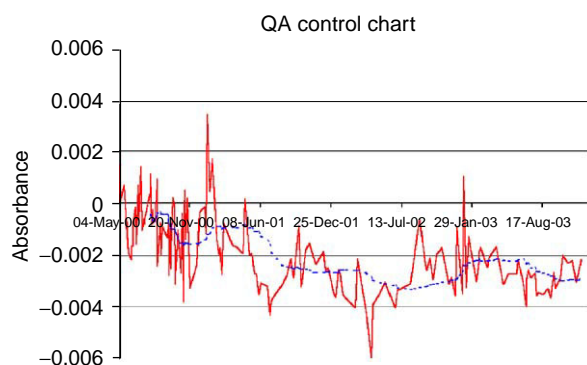
Most instrument manufacturers rely on the tests specified by ASTM International (standard E387), involving solutions with sharp cutoffs to measure stray light.

If the UV stray light has deteriorated, a useful check is to look at the number of hours the deuterium lamp has been in use. Today's better instruments include a 'lamp hours' feature as part of their software. The output of the deuterium lamp decays with age, and although the actual amount of stray light will be fairly constant with time, the relative amount increases as the lamp energy falls.

## Customer Requirements

In a modern laboratory, even simple QA procedures are an integral part of providing spectroscopic data that are 'fit for purpose'.

A considered assessment of just what is required to ensure both the quality of the data produced and more importantly the acceptance of the 'customer' is paramount to the whole QA concept. Procedures must be matched to requirements to ensure that whilst sufficient checks and controls are put in place, these are not perceived as excessive or overbearing. For example, in a manufacturing environment, a simple spectroscopic comparison reference measurement



**Figure 6** Quality assurance control chart of an absorbance scale certified reference material, measured regularly on a high-performance reference commercial spectrophotometer. The dotted trace indicates a moving average calculation on the data, showing a possible seasonal trend in the performance of the spectrophotometer.

of an unknown, against a known reference, may be all that a client will require as proof of compliance. However, if the same product is to be sold into a regulated environment, use of a CRM with full traceability is an essential requirement.

Therefore, effective combination and use of one or more of these 'standards', as appropriate, will provide the data for effective QA of modern spectroscopic methods.

One QA tool, which is yet to be mentioned, is the use of control charts. Having generated the spectroscopic data, their benefit as an immediate visual indication as to the control state of a process or procedure cannot be overstated.

Apart from displaying the obvious out-of-control conditions, trend analysis of the data will often reveal subtle changes in the instrument, not revealed on a day-to-day basis. For example, long-term drift

or seasonal variation may be observed. The example shown is one such case (Figure 6).

Increasing use of fluorescence, micro-plate readers, and fiber-optic based systems all pose their own individual QA requirements and problems, and these are only just now beginning to be addressed.

**See also:** **Quality Assurance:** Quality Control; Instrument Calibration; Interlaboratory Studies; Reference Materials; Production of Reference Materials. **Spectrophotometry:** Overview.

## Further Reading

ASTM standard E-387 (2003) *Annual Book of ASTM Standards*.

Burgess C and Frost T (eds.) (1999) *Standards and Best Practice in Absorption Spectrometry*. London: Blackwell Science.

Burke RW and Mavrodineanu R (1997) *Standard Reference Materials: Certification and Use of Acidic Potassium Dichromate Solutions as an Ultraviolet Absorbance Standard – SRM 935*. NBS Spec. Publ. 260-54 (August).

Mavrodineanu R and Baldwin JR (1980) *Standard Reference Materials: Metal-On-Quartz Filters as a Standard Reference Material for Spectrophotometry – SRM 2031*. NBS Spec. Publ. 260-68 (April).

Mavrodineanu R, Burke RW, Baldwin JR *et al.* (1994) *Standard Reference Materials: Glass Filters as a Standard Reference Material for Spectrophotometry – Selection, Preparation, Certification and Use of SRM 930 and SRM 1930*. NIST Spec. Publ. 260-116 (March).

Weidner VR, Mavrodineanu R, Mielenz KD *et al.* (1986) *Standard Reference Materials: Holmium Oxide Solution Wavelength Standard from 240–650 nm – SRM 2034*. NBS Spec. Publ. 260-102 (July).

Wenclawiak BW, Koch M, and Hadjicostas E (eds.) (2003) *Quality Assurance in Analytical Chemistry*. Berlin: Springer.

## Internal Standards

**R Whelpton**, University of London, London, UK

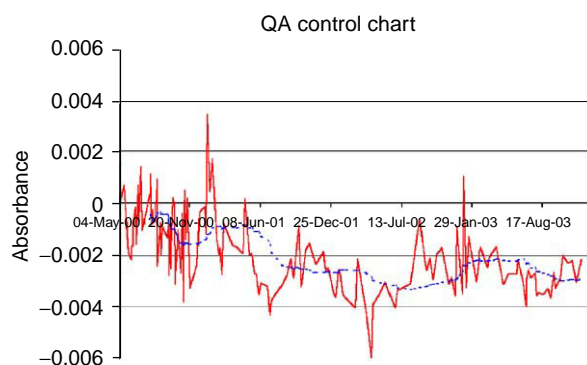
© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

This article considers the role of internal standards in the quality control of quantitative assays of organic compounds. An internal standard can be defined as a second compound, not the analyte, which is added at some stage in an assay with the intention of

correcting for systematic errors. The use of internal standards in analytical procedures such as chromatographic analyses was once considered mandatory by some. Indeed, assays utilizing internal standards were once assumed to be superior to those without, even when there was little or no supporting evidence. In the 1970s, these views were beginning to be questioned, and articles began to appear that demonstrated that the indiscriminate use of internal standards was much more likely to reduce the quality of an assay than to improve it. Moreover, with the





**Figure 6** Quality assurance control chart of an absorbance scale certified reference material, measured regularly on a high-performance reference commercial spectrophotometer. The dotted trace indicates a moving average calculation on the data, showing a possible seasonal trend in the performance of the spectrophotometer.

of an unknown, against a known reference, may be all that a client will require as proof of compliance. However, if the same product is to be sold into a regulated environment, use of a CRM with full traceability is an essential requirement.

Therefore, effective combination and use of one or more of these 'standards', as appropriate, will provide the data for effective QA of modern spectroscopic methods.

One QA tool, which is yet to be mentioned, is the use of control charts. Having generated the spectroscopic data, their benefit as an immediate visual indication as to the control state of a process or procedure cannot be overstated.

Apart from displaying the obvious out-of-control conditions, trend analysis of the data will often reveal subtle changes in the instrument, not revealed on a day-to-day basis. For example, long-term drift

or seasonal variation may be observed. The example shown is one such case (Figure 6).

Increasing use of fluorescence, micro-plate readers, and fiber-optic based systems all pose their own individual QA requirements and problems, and these are only just now beginning to be addressed.

**See also:** **Quality Assurance:** Quality Control; Instrument Calibration; Interlaboratory Studies; Reference Materials; Production of Reference Materials. **Spectrophotometry:** Overview.

## Further Reading

ASTM standard E-387 (2003) *Annual Book of ASTM Standards*.

Burgess C and Frost T (eds.) (1999) *Standards and Best Practice in Absorption Spectrometry*. London: Blackwell Science.

Burke RW and Mavrodineanu R (1997) *Standard Reference Materials: Certification and Use of Acidic Potassium Dichromate Solutions as an Ultraviolet Absorbance Standard – SRM 935*. NBS Spec. Publ. 260-54 (August).

Mavrodineanu R and Baldwin JR (1980) *Standard Reference Materials: Metal-On-Quartz Filters as a Standard Reference Material for Spectrophotometry – SRM 2031*. NBS Spec. Publ. 260-68 (April).

Mavrodineanu R, Burke RW, Baldwin JR *et al.* (1994) *Standard Reference Materials: Glass Filters as a Standard Reference Material for Spectrophotometry – Selection, Preparation, Certification and Use of SRM 930 and SRM 1930*. NIST Spec. Publ. 260-116 (March).

Weidner VR, Mavrodineanu R, Mielenz KD *et al.* (1986) *Standard Reference Materials: Holmium Oxide Solution Wavelength Standard from 240–650 nm – SRM 2034*. NBS Spec. Publ. 260-102 (July).

Wenclawiak BW, Koch M, and Hadjicostas E (eds.) (2003) *Quality Assurance in Analytical Chemistry*. Berlin: Springer.

## Internal Standards

**R Whelpton**, University of London, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

This article considers the role of internal standards in the quality control of quantitative assays of organic compounds. An internal standard can be defined as a second compound, not the analyte, which is added at some stage in an assay with the intention of

correcting for systematic errors. The use of internal standards in analytical procedures such as chromatographic analyses was once considered mandatory by some. Indeed, assays utilizing internal standards were once assumed to be superior to those without, even when there was little or no supporting evidence. In the 1970s, these views were beginning to be questioned, and articles began to appear that demonstrated that the indiscriminate use of internal standards was much more likely to reduce the quality of an assay than to improve it. Moreover, with the

increased use of reproducible injection systems, stable detection systems and extraction methods that gave reproducibly high recoveries, such as solid-phase extraction, the days when internal standards were routinely necessary appeared to be numbered. That was until the introduction of liquid chromatography (LC)–tandem mass spectrometry (MS–MS) when the need for suitable, and carefully chosen, internal standards became apparent. This article considers the arguments for the use of internal standards and illustrates some of the problems that can arise and how they may be minimized.

A constant amount of internal standard must be added with a high degree of precision to each sample. Accuracy is generally less important, and within limits, an internal standard does not need to be as pure as the analyte. Depending on the reasons for its inclusion, the internal standard, may be added at the beginning or at later stages in the assay procedure.

Before using an internal standard, the following should be considered:

1. Internal standards are useful for correcting systematic errors that occur in some assays, but the need for one should be identified before spending time and effort on finding one.
2. One should be prepared to spend at least as much time and effort on validating the use of the internal standard as that given for the analyte. Since using an internal standard frequently increases the analysis time, it is worth ensuring that the choice is the right one.
3. It should be demonstrated that the internal standard is performing as required and that its inclusion confers an overall benefit.
4. It may be necessary to develop the method without the internal standard if it fails to meet the requirements.

### Requirements of an Internal Standard

In 1981, Haefelinger listed his criteria for an internal standard, which, summarized, were as follows.

It must

1. be completely resolved from the known and unknown substances in the chromatogram;
2. be eluted near the peak of interest;
3. have a similar peak height (or area) to the substance to be determined;
4. be chemically similar to the substance of interest;
5. be chemically stable.

For the analysis of drugs in biological materials:

1. the internal standard should be added to biological samples in solution (preferably aqueous) and the samples thoroughly mixed to ensure uniform distribution of the internal standard;
2. where there are extraction steps, 'the internal standard should show similar behavior to the analyte; namely, the partition coefficients should be equal';
3. the internal standard should not be a metabolite of the drug of interest (nor a decomposition product or other drug that may be present);
4. it should not interfere with the metabolites of the drug or endogenous compounds.

To evaluate the validity of this list it is helpful to consider the rationale for the use of internal standards and to consider the types available.

### Reasons for Use

**Reproducibility of injection** Manual injection of small volumes of analyte solution on to gas chromatographic columns is not very precise. Provided that the detector response to the internal standard is linear, the signal given by the internal standard should be directly proportional to the volume of sample injected. Thus, if all the samples contain the same concentration of internal standard a calibration curve of:

$$\frac{\text{analyte response}}{\text{internal standard response}}$$

versus analyte concentration can be constructed. If unknown samples are treated in the same way, response ratios can be used to derive the analyte concentration – compensation for differences in injection volume having been made. The same reasoning applies to situations where a volatile injection solvent is used. In this case, the drug and internal standard concentrations will rise as the solvent evaporates, the response given by the internal standard being inversely proportional to the volume of solution remaining.

**Instability of the detection system** Some detection systems are prone to instability. Contamination of electron capture, nitrogen–phosphorus (NPD), and electrochemical detectors may lead to gradual or even sudden loss of sensitivity. A suitably chosen internal standard may compensate for such changes. However, the analyte–internal standard signal ratio must remain constant for a given mixture. If detector contamination results in nonlinear changes and different compounds are affected to differing degrees, the use of an internal standard is fallacious.

**Pipetting errors and evaporation of extraction solvent** Once an internal standard has been added, losses on sample transfer, e.g., taking less than the total amount of extraction solvent, should be reflected in its recovery. It can be argued, therefore, that inclusion of an internal standard obviates the need for quantitative transfer and so cheaper glassware, e.g., Pasteur pipettes, or quicker methods, such as pouring, may be used to separate the phases. If the transfer volumes are measured, good precision can be obtained without an internal standard. Moreover, measuring the volumes of the phases transferred is to be recommended as the internal standard response should be relatively constant and any deviation should alert the analyst to a problem. Evaporation of solvent (e.g., diethyl ether) during handling and changes in phase volumes will give errors (usually an overestimate of the analyte concentration) that can be reduced by inclusion of an internal standard.

**Extraction efficiency** For an internal standard to compensate for extraction efficiency it must be as chemically similar to the analyte as possible. Thus, isotopically labeled internal standards should be the most appropriate or if these are unsuitable, or unavailable, then the closest suitable homolog should be selected. A problem arises when several compounds, e.g., a drug and its metabolites, are to be quantified, as the extraction characteristics of each will be different. Some studies may require more than one internal standard. If the extraction efficiency is high, <95%, then inclusion of an internal standard may give only marginal improvement. The same is true of situations where the extraction efficiency is lower but reproducible.

**Derivatization and nonstoichiometric reactions** Ideally, only isotopic internal standards should be used, as it is naive to expect two chemically distinct entities to react identically. The causes of the imprecision should be investigated and the reaction conditions modified to make them suitably robust rather than rely on the inclusion of a second compound. However, potential problems with isotope internal standards need to be considered, particularly if the bond(s) linking the isotope to the rest of the molecule are involved in the reaction. It may be necessary use alternative sites of labeling.

**Protein binding, precipitation, and adsorption** The inclusion of internal standards to correct for differences in protein binding, losses on protein precipitation, and adsorptive losses is far from ideal unless the compound is as chemically similar as possible to the analyte, preferably an isotopically labeled internal

standard. When recovery is very variable, radio-labeled analyte, added as a tracer, can prove effective. Conversely, when the analyte is radioactive, the use of 'cold' analyte, as an internal standard and carrier, is valid and to be recommended. Normally, it should be possible to add sufficient carrier to use a relatively simple assay – e.g., liquid extraction and ultraviolet detection prior to radioactive counting.

### **Types of Internal Standard**

Internal standards can be considered under four major groups: isotope, homolog, generic, and chemically unrelated. To some extent the type of internal standard dictates its applicability.

**Isotope** Because isotopically labeled materials are as chemically close to the analyte as possible without being identical to it, these should be the most satisfactory internal standards, fulfilling the majority of requirements. However, difficulties arise. First, suitably labeled materials have to be available or synthesized and, second, they have to be quantified. When using radiolabeled internal standards, the sample normally has to be divided for the radioactivity to be counted. Molecules containing stable isotopes are usually quantified by MS. Furthermore, there is increasing evidence that isotopic effects are more common than was once thought and assumptions of identical chemical properties cannot be made.

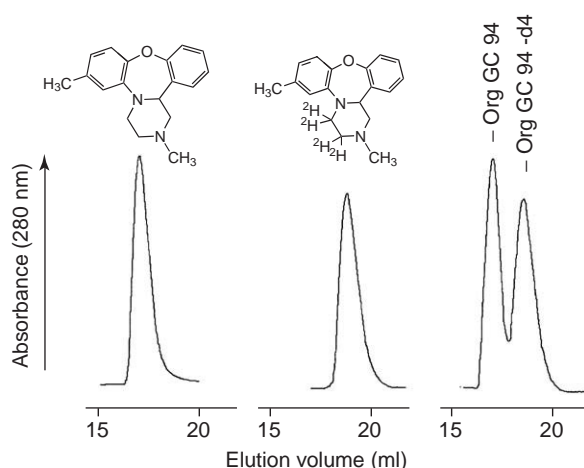
The nuclides suitable for incorporating in drugs and related biological materials include the stable isotopes deuterium ( $^2\text{H}$ ),  $^{13}\text{C}$ ,  $^{15}\text{N}$ , and  $^{18}\text{O}$ , and the unstable isotopes  $^3\text{H}$ ,  $^{14}\text{C}$ ,  $^{32}\text{P}$ , and  $^{35}\text{S}$ . Generally, synthesis of high specific activity radiolabeled drugs is beyond the scope of most analysts, so if the material is not commercially available it will have to be custom-synthesized. High-purity  $^2\text{H}$ -labeled reagents, e.g., acetic anhydride, iodomethane, and lithium aluminum anhydride, are available at reasonable cost so that preparation of a stable internal standard is possible provided that a suitable precursor is available. Sometimes the precursor can be derived from the drug either chemically or metabolically, e.g., a tertiary amine may be desalkylated and then reacted with the appropriate labeled iodoalkane to give the internal standard. Alternatively, an already labeled compound may be modified, e.g., [ $^2\text{H}$ -N-methyl]-7-hydroxychlorpromazine has been isolated from the feces of rats dosed with [ $^2\text{H}$ -N-methyl]-chlorpromazine.

Trace amounts of radioactive internal standards are used so that their contribution to the 'cold' assay

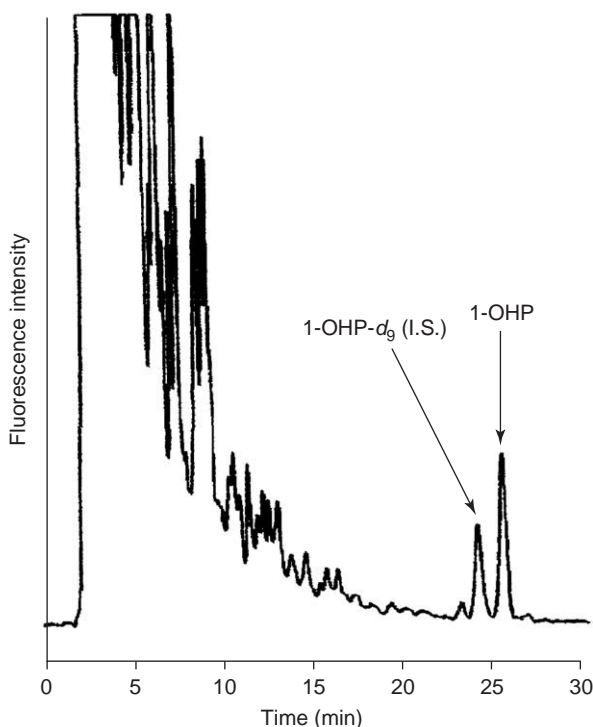
is negligible. Thus, for low-level assays high specific activity material will be required and there is potential for decomposition. For mass spectral analysis the internal standard is usually added at a higher concentration than the analyte, e.g., 100-fold. This results in the assay being conducted over a narrower 'total' concentration range, thus minimizing concentration-related effects such as adsorption. The isotopic purity of the material will limit the amount that can be added for a particular limit of quantification. As the amount of nonlabeled material in the internal standard is constant, a calibration curve with a significant positive intercept is permissible providing the blank and lowest calibration standard can be distinguished.

It cannot be assumed that labeled and nonlabeled compounds are chemically identical. They are not. Bonds linking heavier isotopes are more stable and if such a bond is broken, in a chemical reaction or due to decomposition during the assay, then the heavier isotopic material will react at a slower rate and the ratio of labeled to nonlabeled compound will change. These effects are most noticeable with the isotopes of hydrogen as the mass differences are two- and three-fold. Labeled material may show chromatographic differences and integration parameters may need to be adjusted to compensate for differences in retention times. Deuterium labeled internal standards elute slightly earlier from gas chromatography (GC) columns. The partial resolution of a deuterated internal standard by liquid chromatography (LC) was reported as early 1976 (Figure 1). The larger the number of deuterated atoms the higher the degree of resolution. Baseline resolution of analyte and internal standard allows alternative forms of detection, such as fluorescence (Figure 2). Such observations of differences in chromatographic behavior must cast doubt on assumptions about the chemical similarity of labeled and nonlabeled materials, particularly with regard to solvent partitioning and solid-phase extraction.

**Homolog (analog) internal standards** Isotopic internal standards are not routinely used as the increased costs, both for the materials and the instrumentation, do not justify it and homolog internal standards can provide a reasonable alternative. Often the chemical similarities between a compound and its closest homolog are such that this approach gives satisfactory results. If needs be, higher homologs can be used but this is may not be so successful. It should be noted the addition of even one methylene group can have a marked effect on the extraction efficiency and this should be



**Figure 1** Partial resolution of the mianserin analog, Org GC 94, and its tetradeuterated internal standard (Org GC 94- $d_4$ ) by LC. (Redrawn from de Ridder JJ and van Hal HJM (1976) Unexpected high-performance liquid chromatographic separation of Org GC 94 and [3,3,4,4- $^2H_4$ ]Org GC 94. *Journal of Chromatography* 121: 96–99; © Elsevier.)



**Figure 2** LC resolution of 1-hydroxypyrene (1-OHP) and its deuterated internal standard (1-OHP- $d_9$ ). Fluorescence detection:  $\lambda_{ex}$  240 nm,  $\lambda_{em}$  387 nm. (Reproduced with permission from Chetiyakornkul T, Toriba A, Kizu R, Makino T, Nakazawa H, and Hayakawa K (2002) Determination of 1-hydroxypyrene in human urine by high-performance liquid chromatography with fluorescence detection using a deuterated internal standard. *Journal of Chromatography A* 961: 107–112; © Elsevier.)

invested to ensure that the extraction conditions are optimum for both the analyte and the internal standard.



**Generic internal standards** This term is used to describe compounds, which are not homologs, but are chemically similar to the analyte, for example, compounds substituted with different halogen atoms, or containing an additional group. The chemical properties of compounds in this group would not be as similar as those expected of homologs but within certain constraints these compounds may prove to be useful internal standards. Care should be taken if they are used to compensate for extraction efficiency or derivatization reactions and, of course, the internal standard must not be a putative metabolite or decomposition product. These types of internal standards, however, are suitable for compensating for differences in injection volumes and changes in detector sensitivities.

**Unrelated compounds** These should be reserved for compensating for differences in injection volumes and transfer volumes only. They are sometimes chosen because alternatives are not available: their position on the chromatograph being the primary criterion for the choice.

## Choice and Supply

Suitable internal standards may be readily available from chemical suppliers, colleagues in organic chemistry departments, pharmaceutical companies, etc. If they are not, the possibility of having them synthesized or preparing them in the analytical laboratory should be considered. This is generally preferable 'to making do' with a less suitable compound. Preparation of internal standards should not be daunting. An internal standard need not be 100% pure and relatively small quantities will often suffice. At least, to prove the validity of using the chosen compound as an internal standard, prior to having a larger batch synthesized, should that proof be necessary. The reaction can be monitored using the analytical tools that are to be employed in the proposed analysis and, for sensitive assays, it is often possible to isolate sufficient product by thin-layer or liquid chromatography.

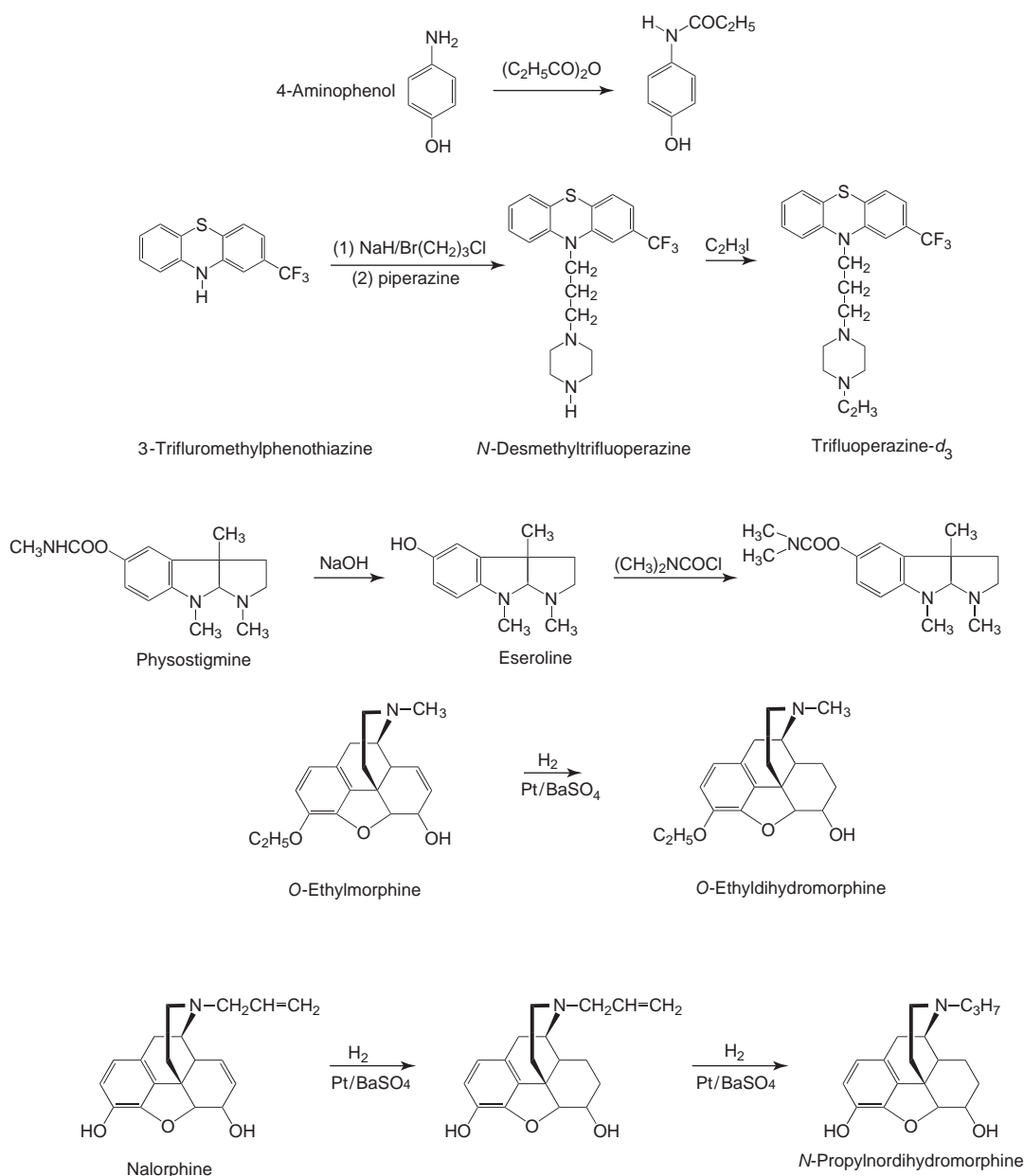
Often the chemistry required to synthesize an internal standard is not dissimilar to the derivatization reactions with which the analysis will be familiar. **Figure 3** shows examples of the types of reactions that have been used to prepare internal standards in an analytical laboratory. Whenever possible the closest homolog (or analog) should be prepared. The paracetamol homolog was produced by shaking 4-aminophenol hydrochloride with propionic acid anhydride in water until the product

precipitated (~30 min). The filtered solid was pure enough for use without further treatment. Crystallization from ethyl acetate conferred no additional benefit.

Tertiary and secondary amines are often metabolized by desalkylation and these metabolites may be available as reference compounds. This was the approach adopted for morphine internal standards. *N*-Desmethylnormorphine (normorphine) was used as the starting point for preparation of the *N*-ethyl and *N*-propyl homologs. The secondary amine was treated with the appropriate iodoalkane and the products separated by thin-layer chromatography. The example in **Figure 3** shows the synthesis of [<sup>2</sup>H-methyl]-trifluoperazine, the secondary amine precursor having been synthesized from readily available starting materials. Quaternary ammonium by-products present little problem as they do not extract into organic solvents and tend to remain at the origin of thin-layer plates. If the desalkyl material is not available and is difficult to synthesize then it may be possible to dealkylate the analyte. This approach was used for the alkaloid hyoscyne (scopolamine), as the cost of total synthesis would have been prohibitive. The desmethyl compound was obtained by permanganate oxidation and used to give a number of potential internal standards including: the *N*-ethyl, *N*-propyl homologs and deuterium and tritium labeled [*N*-methyl]-hyoscyne.

For more complex structures, a metabolite or decomposition product may provide the best approach. Esters may be hydrolyzed and the product reesterified to give the required product. Physostigmine (**Figure 3**) was hydrolyzed with sodium hydroxide under nitrogen to give the phenol, eseroline, which on treatment with *N,N*-dimethylcarbamoyl chloride gave the *N,N*-dimethylcarbamate analog. Ethyl and propyl homologs were prepared by treating the eseroline with ethyl or propyl isocyanate in the presence of a trace of sodium, but, after evaluation, these were rejected as internal standards.

Reference materials and internal standards may be obtained by simple oxidation or reduction reactions. Reduction of *O*-ethylmorphine gave the *O*-ethyl homolog of dihydrocodeine. The 7,8-hydrogenation was accomplished by bubbling hydrogen into an acidic solution of the compound in the presence of a little Pt catalyst on barium sulfate (**Figure 3**). The reaction was monitored by LC. When a similar approach was adopted to produce an internal standard for dihydromorphine, the hydrogenation of nalorphine proceeded in two stages (**Figure 3**), which were easily monitored by LC (**Figure 4**). The final product (checked by nuclear magnetic resonance) was *N*-propyl-7,8-dihydronormorphine.



**Figure 3** Examples of internal standards prepared in an analytical laboratory.

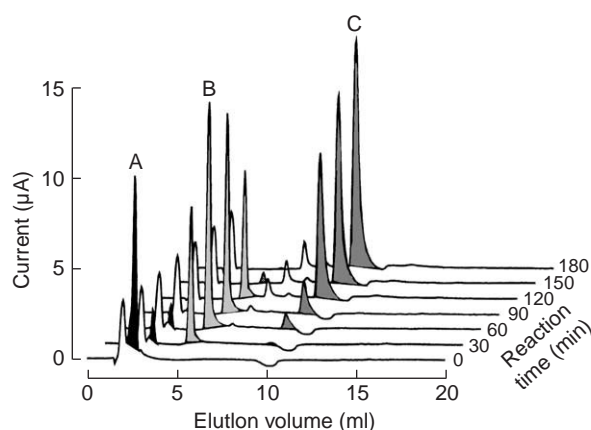
If simple chemical reactions are inappropriate then metabolism or decomposition may be alternative approaches. Isotope-labeled materials lend themselves to this approach. The labeled precursor can be incubated with enzymes or extracted from excretion products after administration to animals. The recovery of material is easier when the material is radioactive, but the technique can be applied equally well to nonradioactive material. Small quantities of otherwise expensive isotopic material can be obtained this way; for example, deuterium labeled 7-hydroxychlorpromazine (for GC-MS analysis) has been

extracted from the feces of rats dosed with labeled chlorpromazine.

#### Will an Internal Standard Improve the Assay?

That an internal standard is 'just something else to measure and something else to go wrong' may be an extreme view, but also it should not be assumed that its inclusion will inevitably improve the precision and accuracy of a chromatographic method. The use of an internal standard should not be expected to reduce the imprecision of an assay caused by random





**Figure 4** Use of HPLC to monitor the preparation of an internal standard. The chromatograms show the two stage reduction of nalorphine (A) to 7,8-dihydronalorphine (B) and subsequent reduction of the allyl group to give the *N*-propyl-7,8-dihydro product (C).

errors. Indeed, the random errors associated with the quantification of the internal standard are likely to 'increase' the overall imprecision rather than reduce it. From replicate assays of analyte and internal standard, mean responses (e.g., peak height or area) and their associated relative standard deviations  $RSD_a$  and  $RSD_{is}$  can be obtained. The relationship between the deviations in the measurement of the responses of the individual compounds and the ratio as used in a typical assay is given by:

$$(RSD_r)^2 \approx (RSD_a)^2 + (RSD_{is})^2 - 2r(RSD_a) \cdot (RSD_{is}) \quad [1]$$

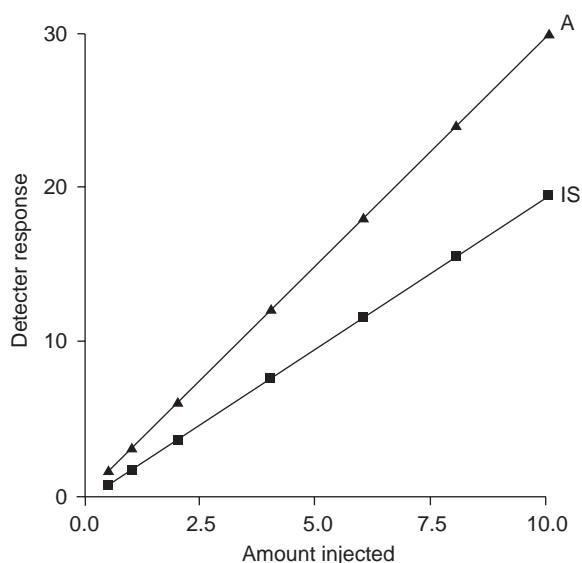
where  $r$  is the correlation coefficient between the responses to the analyte and internal standard. If the precision of the assay has been improved by the introduction of the internal standard then:

$$RSD_r < RSD_a \quad [2]$$

Substituting eqn [1] in eqn [2] and rearranging gives:

$$RSD_{is} < 2r \cdot RSD_a \quad [3]$$

Thus, depending on the correlation between the responses to analyte and internal standard and imprecision associated with those responses, inclusion of an internal standard may reduce or increase the overall imprecision. Even when  $r = 1$ , the RSD for the internal standard must be less than half that of the analyte, for an overall improvement in the assay. If the compound under test can be determined with high precision in the absence of internal standard, then the above equation confirms what one would instinctively feel should be the case, namely that



**Figure 5** Amount injected-response curves for an imaginary analyte (A) and internal standard (IS).

addition of an internal standard will not improve the situation, but rather, if anything, worsen it. A large RSD for the internal standard would indicate that the compound was a poor choice. To test the effect of the internal standard on assay accuracy and precision, calculate them using the analyte response data.

A crucial assumption that is generally made is that the response to the internal standard is linear. What is not so often considered is the effect that an intercept can have. **Figure 5** shows two 'perfect' amount injected-response curves. The line representing the response to the internal standard:

$$\text{response} = -0.5 + 2 \times (\text{amount injected}) \quad [4]$$

has a small negative intercept. A reasonable choice for the working amount of the internal standard would be 5 – this is in the middle of the linear range. If the internal standard is to perform as it should, then for a given ratio of amount of analyte to internal standard the response ratio should be constant, no matter what volume of solution is injected. Assume there is 1 mass unit per microliter of solution, thus injecting 5  $\mu\text{l}$  will give an internal standard response of 9.5. If half the volume were injected, the response would not be  $9.5/2 = 4.75$ , but 4.5 (from eqn [4]). This would result in the analyte amount being overestimated by 5.5%. If only 1  $\mu\text{l}$  were injected the corresponding overestimate would be 26.7%. This is a rather extreme example to illustrate the potential problem that can arise if the internal standard concentration-response curve does not pass through a 0,0-origin. One would expect such a problem to be identified during method development and validation.

However, when methods are modified, e.g., when seeking more sensitivity, the amount of internal standard may be reduced and the fact that there is an intercept may become significant. The concentration–response curve for the analyte need not be linear but clearly, if the internal standard curve is not linear then the results using:

$$\text{analyte response/internal standard response}$$

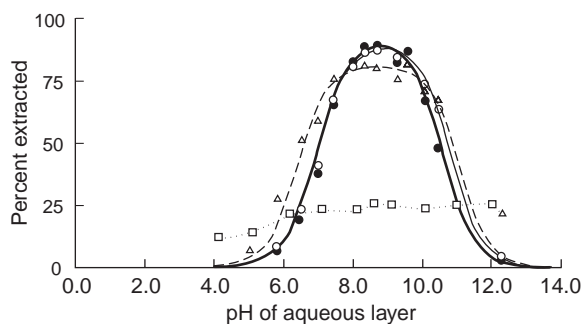
will be bizarre.

**Practical examples of use** An internal standard should be introduced when it is thought that the effects of a systematic error can be reduced by its inclusion. The potential usefulness of an internal standard can be tested by deliberately perturbing the system at the point where the error is thought to occur. For example, if accuracy of injection volume is suspect, then differing volumes of a test mixture of analyte(s) and internal standard should be injected to demonstrate that the analyte–internal standard ratio remains constant. The use of 1-propanol as an internal standard for ethanol determined by gas–liquid chromatography is typical of their use to correct for variability in injection volume (Table 1). Using peak height ratio, rather than peak height, generally gave smaller intraassay coefficients of variation and the correlation coefficient of the calibration graph increased from 0.9899 to 0.9993 when peak height ratios were used. If the pH of the extraction is thought to be critical, this should be tested by performing the extraction at pH values either side of the optimum. Figure 6 shows the pH–extraction curves for morphine and its *N*-ethyl and *N*-propyl homologs. The extraction characteristics of the ethyl homolog were very similar to those of morphine, but those of the propyl homolog were not so close. Furthermore, the ethyl homolog chromatographed with a suitable retention time whereas the propyl compound chromatographed close to the void volume of the column. Consequently, the ethyl compound was chosen. The extraction characteristics of dextromethorphan (Figure 5), which has been used as the internal standard in previous morphine assays, illustrate that this compound is far from being an ideal candidate. Comparison of the intraassay RSDs showed that inclusion of *N*-ethylnormorphine improved the precision of the assay at the three concentrations (100, 50, 10 ng ml<sup>-1</sup>) tested. The RSDs were 3.6, 5.1, and 6.0%, respectively, without internal standard and 0.6, 1.2, and 3.5% with the internal standard.

If detector stability is the problem it may be possible to study the effect of changing the conditions.

**Table 1** Comparison of intraassay precision of ethanol assay with and without 1-propanol as internal standard

Concentration (mg dl <sup>-1</sup> )	RSD (%)	
	Peak height evaluation	Peak ratio evaluation
25	4.40	3.79
50	10.24	8.75
100	5.47	5.42
150	9.48	3.44
200	4.45	4.95



**Figure 6** Comparison of the extraction characteristics of an analyte (morphine, ●) and its first homologs (*N*-ethyl, ○, and *N*-propyl, △). The extraction of dextromethorphan (□), used as internal standard in a published method, is shown for comparison. The extraction solution was 10 volumes butanol + toluene (1 + 9) to one volume of aqueous buffer.

In the example above, changes in electrochemical detector responses to morphine as function of applied voltage were paralleled by the responses to the ethyl homolog. The peak height ratio remained between 0.56 and 0.57 for a test mixture injected at applied potentials ranging from 0.6 to 1.0 V versus Ag/AgCl. Similarly, this approach was used successfully in the GC–NPD determination of fluphenazine. Repeated injections of trimethylsilyl fluphenazine resulted in a progressive decrease in detector response; however, addition of thioridazine to the injection solution compensated for this loss of sensitivity. The peak height ratio of fluphenazine–thioridazine for test solutions was shown to vary by <2% even when the peak height response to fluphenazine had declined by as much as 80%. The detector sensitivity was restored by periodically increasing the temperature of the rubidium silicate bead for a few minutes. This is an example of a ‘generic’ internal standard chosen on the basis of (1) the way detector responses to it parallel changes in analyte response when the detector conditions (hydrogen flow and bead temperature) were varied and (2) its position in the chromatogram.

It should not be assumed that because a compound was a suitable internal standard for a particular assay

**Table 2** Physostigmine–internal standard ratio in 0.1 ml fractions collected from five solid-phase extraction columns

Fraction (ml)	Diol	C-18	C-8	C-2	CN
0–0.1	0.8	1.8	1.2	1.3	2.3
0.1–0.2	1.0	0.8	0.8	0.8	0.8
0.2–0.3	1.5	0.3	0.8	0.6	1.0

that it will still be suitable if the method is changed. This is exemplified by the observations with physostigmine and its *N,N*-dimethyl analog, which had been shown to be a useful internal standard for liquid–liquid extraction. When a solid-phase method was developed it became apparent that the internal standard was no longer making a useful contribution. The ratio of drug to internal standard eluting from solid-phase columns is shown in **Table 2**. The peak height response ratios for three fractions from five types of column show that the analyte and internal standard are partially resolved. On the C-18, C-8, and C-2 columns the drug eluted first, whereas on the diol column the more lipophilic internal standard eluted first. With the nitrile column the internal standard eluted as a sharper peak. In no case did the ratio remain constant for the fractions. Fortunately, the precision of the solid-phase extraction was shown to be acceptable without an internal standard and so the method was developed without one.

**Internal standards for GC–MS and LC–MS–MS** Because of variations in the degree of ionization, the use of internal standards is vital when using MS for quantitative detection. Tandem MS – selecting the parent ion and a subsequent daughter ion – held the promise of specific and sensitive analyses with little or no sample preparation. However, because of interferences from the sample matrix, some form of chromatographic ‘cleanup’ is required. LC–MS–MS methods often use short chromatographic columns, relying on the mass spectrographic resolution of specific ions rather than chromatographic separations to impart the required selectivity; the advantage of this approach being short cycle times and high sample throughput. However, one ignores the coeluting but ‘unseen’ materials from the sample matrix at one’s peril. The most common phenomenon is a reduction of signal due to endogenous materials in the sample competing with the analyte during the ionization process. Ion suppression is frequently observed with electrospray interfaces and is largely unpredictable and variable. Clearly, this is a case for the use of isotopically labeled internal standards and it has been shown that this approach can be successful – although ion suppression may reduce the signal to the

extent that sensitivity is compromised. In some instances, for example, during drug development, isotopically labeled internal standard may not be available and a second compound may have to be used. Generally such a compound will be fully or partially resolved chromatographically and there is every possibility that the degree of ion suppression will be different. Increased, and uncompensated, ion suppression of the internal standard will lead to an overestimate of the analyte. That labeled and unlabeled compounds may partially resolve during chromatography must be borne in mind not only with regard to choosing the correct integration parameters, but also that the degrees of ion suppression may be different.

Any isotopically labeled material must be chosen such that bonds linking them are not broken during fragmentation as bonds bearing heavier isotopes are more stable and the resulting fragmentation pattern will differ from that of the analyte.

#### Working without an Internal Standard

When a suitable internal standard cannot be identified, it will be necessary to work without one. This does not mean that the method will inevitably suffer from poor precision and accuracy; however, several points will have to be observed:

1. Keep the number of steps to a minimum; for example, consider injecting the sample directly or after dilution, increasing the detector sensitivity or selectivity if necessary to accommodate this.
2. If ‘cleanup’ and concentration steps are required, consider solid-phase extraction as this often gives high recoveries with low standard deviations.
3. When liquid–liquid extraction is used the transfer volumes must be measured to ensure consistency across all samples. Volatile extraction solvents are best avoided as it is difficult to assess the volume change during handling. When feasible, concentration of the sample by back-extraction into a small volume of liquid is preferable to evaporation.
4. Completely (over) fill the sample loop when injecting on to LC columns to ensure minimum error. Similarly inject as large a volume as possible on to GC columns to reduce the errors associated with small volumes. Use of the ‘solvent-flush’ technique may be beneficial. A small volume of solvent is drawn into the syringe, then air until the meniscus can be seen; the sample is loaded using the movement of the meniscus to measure the volume and finally more air is withdrawn until the ‘slug’ of sample can be seen. The volume of sample, seen between the menisci, can be checked before injection. This approach had the added advantage that it reduces carryover between samples.

See also: **Extraction:** Solid-Phase Extraction. **Gas Chromatography:** Overview. **Liquid Chromatography:** Overview. **Quality Assurance:** Spectroscopic Standards.

## Further Reading

- Chang WT, Smith J, and Liu RH (2002) Isotopic analogs as internal standards for quantitative GC/MS analysis – molecular abundance and retention time difference as interference factors. *Journal of Forensic Science* 47: 873–881.
- Curry SH and Whelpton R (1978) Statistics of drug analysis and the role of internal standards. In: Reid E (ed.) *Blood Drugs and Other Analytical Challenges*, pp. 29–41. Chichester: Ellis Harwood.
- Dudley KH (1981) Policies for internal standards. In: Reid E (ed.) *Trace-Organic Sample Handling*, pp. 336–340. Chichester: Ellis Horwood.
- Haefelfinger P (1981) Limits of the internal standard technique in chromatography. *Journal of Chromatography* 218: 73–81.
- Sojo LE, Lum G, and Chee P (2003) Internal standard signal suppression by co-eluting analyte in isotope dilution LC–ESI-MS. *Analyst* 128: 51–54.
- Vessman J (1981) The usefulness of internal standards in quantitative bioanalytical methods. In: Reid E (ed.) *Trace-Organic Sample Handling*, pp. 341–347. Chichester: Ellis Horwood.
- Whelpton R (1984) Isotope-labelled materials as internal standards. In: Reid E and Wilson ID (eds.) *Drug Determination in Therapeutic and Forensic Contexts*, pp. 39–44. New York: Plenum Press.
- Wilson ID (1990) Observations on the usefulness of internal standards in the analysis of drugs in biological fluids. In: Reid E and Wilson ID (eds.) *Analysis for Drugs and Metabolites Including Anti-infective Agents*, pp. 79–82. Cambridge: Royal Society of Chemistry.

## Instrument Calibration

**J G K Ševčík**, Charles University in Prague, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The analytical chemical picture of a studied object is provided by a measuring instrument. The results of analytical measurements are required by a wide range of users, and thus standardization approaches must be employed to assure their accuracy, precision, and low uncertainty in their interpretation.

Any measuring equipment fulfills two tasks:

- It converts the analyte amount into an electrically measurable quantity, which is termed the signal
- It treats the signal electronically and using computing techniques.

The creation of the signal is connected with a physical, chemical, or biological interaction of the analyte within the part of the measuring instrument called the sensor, followed by conversion into a measurable electrical quantity. This section of the measuring instrument is called the transducer. The subsequent algorithmic treatment of the signal provides rough data that are further refined. The measuring procedure can be depicted as shown in **Figure 1**, from which the analyte amount/signal

conversion factor is obtained, which involves all the principal properties of the measuring system.

The analyte amount/signal conversion factor is termed the sensitivity. It follows from the scheme in **Figure 1** that the conversion factor involves many unknown parameters that are specific to the particular measuring instrument and that the sensitivity has no general significance but that it is specific for a given analytical application. The set of properties of the conversion factor is termed the characteristics of the measuring system. Two models have been formulated for determination of the measuring systems characteristics, namely, the signal model and the response model.

## Signal Model

The signal model expresses the relationship between the analyte signal ( $S_i$ ) and the analyte amount ( $m_i$ ). The sensitivity of the measuring instrument is expressed in terms of the product of the construction properties of the measuring instrument ( $k$ ) and a quantified physical, chemical, or biological property of the analyte ( $a_i$ ) that can be used for the measurement. It holds that

$$S_i = (k * a_i) * m_i^l \quad [1]$$

where  $l$  is the linearity coefficient.

See also: **Extraction:** Solid-Phase Extraction. **Gas Chromatography:** Overview. **Liquid Chromatography:** Overview. **Quality Assurance:** Spectroscopic Standards.

## Further Reading

- Chang WT, Smith J, and Liu RH (2002) Isotopic analogs as internal standards for quantitative GC/MS analysis – molecular abundance and retention time difference as interference factors. *Journal of Forensic Science* 47: 873–881.
- Curry SH and Whelpton R (1978) Statistics of drug analysis and the role of internal standards. In: Reid E (ed.) *Blood Drugs and Other Analytical Challenges*, pp. 29–41. Chichester: Ellis Harwood.
- Dudley KH (1981) Policies for internal standards. In: Reid E (ed.) *Trace-Organic Sample Handling*, pp. 336–340. Chichester: Ellis Horwood.
- Haefelfinger P (1981) Limits of the internal standard technique in chromatography. *Journal of Chromatography* 218: 73–81.
- Sojo LE, Lum G, and Chee P (2003) Internal standard signal suppression by co-eluting analyte in isotope dilution LC–ESI-MS. *Analyst* 128: 51–54.
- Vessman J (1981) The usefulness of internal standards in quantitative bioanalytical methods. In: Reid E (ed.) *Trace-Organic Sample Handling*, pp. 341–347. Chichester: Ellis Horwood.
- Whelpton R (1984) Isotope-labelled materials as internal standards. In: Reid E and Wilson ID (eds.) *Drug Determination in Therapeutic and Forensic Contexts*, pp. 39–44. New York: Plenum Press.
- Wilson ID (1990) Observations on the usefulness of internal standards in the analysis of drugs in biological fluids. In: Reid E and Wilson ID (eds.) *Analysis for Drugs and Metabolites Including Anti-infective Agents*, pp. 79–82. Cambridge: Royal Society of Chemistry.

## Instrument Calibration

**J G K Ševčík**, Charles University in Prague, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The analytical chemical picture of a studied object is provided by a measuring instrument. The results of analytical measurements are required by a wide range of users, and thus standardization approaches must be employed to assure their accuracy, precision, and low uncertainty in their interpretation.

Any measuring equipment fulfills two tasks:

- It converts the analyte amount into an electrically measurable quantity, which is termed the signal
- It treats the signal electronically and using computing techniques.

The creation of the signal is connected with a physical, chemical, or biological interaction of the analyte within the part of the measuring instrument called the sensor, followed by conversion into a measurable electrical quantity. This section of the measuring instrument is called the transducer. The subsequent algorithmic treatment of the signal provides rough data that are further refined. The measuring procedure can be depicted as shown in **Figure 1**, from which the analyte amount/signal

conversion factor is obtained, which involves all the principal properties of the measuring system.

The analyte amount/signal conversion factor is termed the sensitivity. It follows from the scheme in **Figure 1** that the conversion factor involves many unknown parameters that are specific to the particular measuring instrument and that the sensitivity has no general significance but that it is specific for a given analytical application. The set of properties of the conversion factor is termed the characteristics of the measuring system. Two models have been formulated for determination of the measuring systems characteristics, namely, the signal model and the response model.

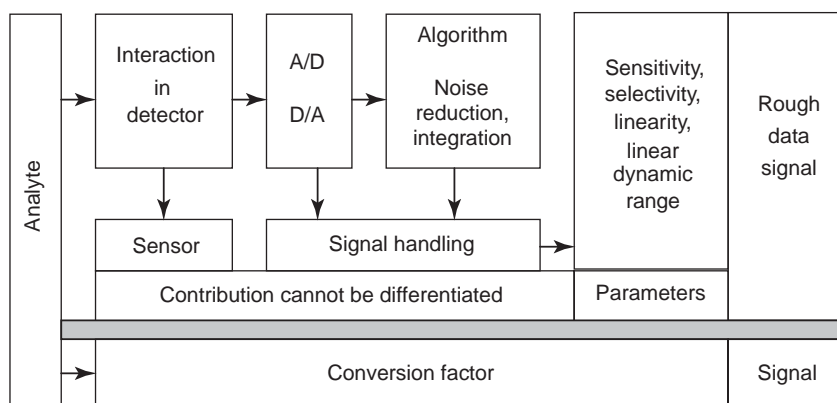
## Signal Model

The signal model expresses the relationship between the analyte signal ( $S_i$ ) and the analyte amount ( $m_i$ ). The sensitivity of the measuring instrument is expressed in terms of the product of the construction properties of the measuring instrument ( $k$ ) and a quantified physical, chemical, or biological property of the analyte ( $a_i$ ) that can be used for the measurement. It holds that

$$S_i = (k * a_i) * m_i^l \quad [1]$$

where  $l$  is the linearity coefficient.





**Figure 1** Measuring system components for an analytical method observing the traceability concept.

## Response Model

The response model, described by eqn [2], states that the real instantaneous signal of the measuring system ( $S$ ) consists of the contributions from the analyte ( $S_i$ ), from the sample matrix ( $S_{\text{matrix}}$ ), and from the instrument itself ( $S_{\text{instr}}$ ). Therefore,

$$S = S_i + S_{\text{matrix}} + S_{\text{instr}} \quad [2]$$

The analyte signal ( $S$ ) integrated from the beginning of the observation ( $t_{\text{start}}$ ) to its end ( $t_{\text{end}}$ ) is termed the response,  $R$ ,

$$R_i = \int_{t_{\text{start}}}^{t_{\text{end}}} S_i dt \quad [3]$$

Equations [1] and [2] demonstrate the variability of the measuring instruments from the point of view of the measuring principle (various physicochemical properties of analyte  $a_i$ , e.g., thermal conductivity, ionization efficiency, molar absorption coefficient) and various construction types (construction properties  $k$  and  $S_{\text{instr}}$  differ among the instrumentation manufacturers). This variability leads to differences in the measuring selectivity and linearity and to different values of the limit of detection and limit of quantitation.

## Signal Handling

The signal is the instantaneous output of the measuring instrument, usually in the form of an electrical quantity. It follows from eqn [2] that the measured signal is not a simple function of the analyte amount but that it corresponds to the sum of contributions from the analyte, the sample matrix, and the instrument itself. The contribution of the instrument,  $S_{\text{instr}}$ , is due to the construction properties

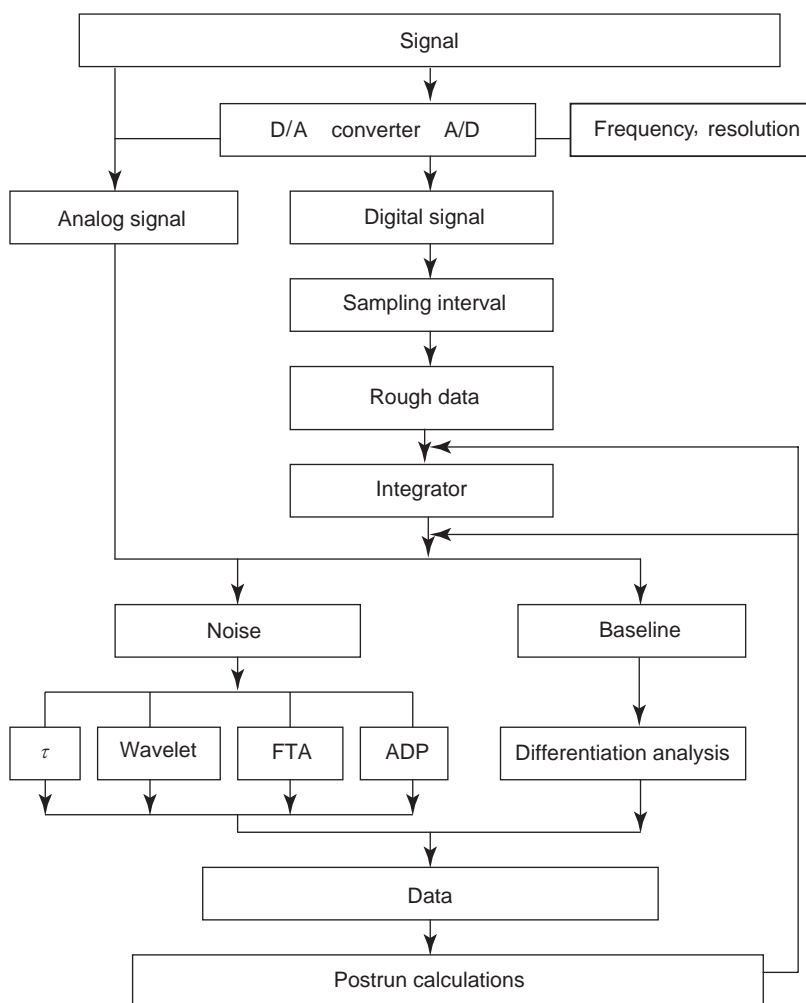
of the instrument and is called the baseline or the zero line of the system. The detection mechanism, ambient parameters, and construction materials are subject to noise, which is manifested by random increasing and decreasing of the signal. The measured signal of a detector is continuously handled and usually stored in the memory of the measuring instrument as a set of rough data that contain both the signal and the noise and the parameters of the baseline (the signal in the absence of an analyte) and are treated appropriately. A scheme of the signal treatment is shown in **Figure 2**.

The signals can be evaluated from the point of view of whether they are continuous in time or not and whether their amplitude is continuous or not. Various signal types can be distinguished (**Figure 3**), analog and digitally treated signals being most common.

The signal directly from the detector is almost always analog. It is processed in the detector, creating either an analog or digital signal that is further processed. Signals with discrete time (numerical or digital signals) are not continuous functions and are termed time series. Signals are treated in order to find their mean values, to sum or multiply them, and to reduce the noise.

The noise oscillating about the baseline is characterized by its frequency and amplitude, and as follows from statistical concepts, the sum of the positive and negative noise deviations equals zero, provided the time interval of summing is sufficiently long. A noise that has a sum of zero over the time interval of observation is called white noise. A noise that does not have a sum of zero is termed random noise. A noise that exhibits a time dependence over subsequent observation intervals is called drift (**Figure 4**). The noise is only significant within the observation interval (e.g., the elution peak width), and thus it is obvious that the significance of the various noise





**Figure 2** A scheme of treatment of the measuring instrument signal.  $\tau$  is the time constant.

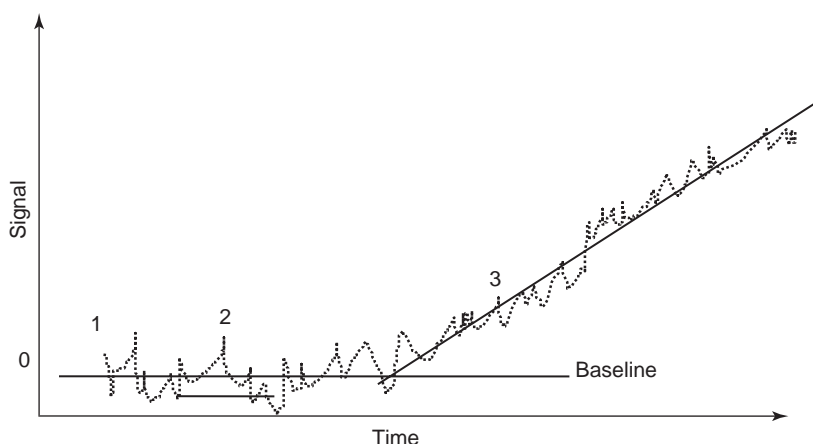
	Continuous time	Discrete time
Continuous amplitude	Analog signal 	Digital signal 
Discrete amplitude	Impulse signal 	Digital signal 

**Figure 3** Classification of signals according to whether they are continuous in time and amplitude or not.

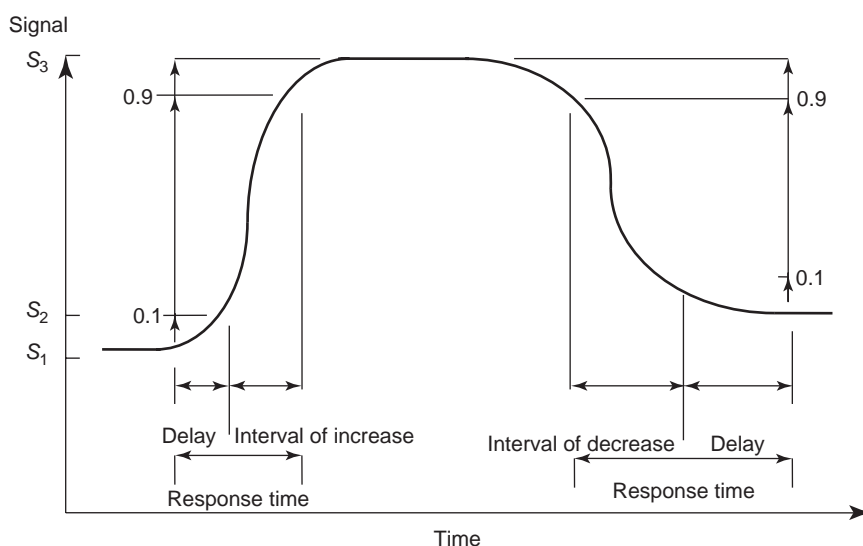
types changes with a change in the observation interval.

Various forms of the detector signal can be inter-converted by electrical components called A/D and

D/A converters (A stands for analog and D for digital) that are characterized by the conversion frequency and by the resolution format (expressed in terms of the number of bits).



**Figure 4** Noise types related to the observation interval. 1, white noise (the baseline); 2, random noise; 3, drift.



**Figure 5** Response time of a measuring instrument according to ISO 9169.

Analog signals are distorted by a number of time constants,  $\tau$ , that are proportional to the detector volume and depend on the way the signal is treated electronically. Signal distortion increases with increasing time constant and may lead to substantial suppression of the signal originally produced by rapid processes and thus to erroneous interpretation. On the other hand, analog output signals are rarely subject to white noise, and random noise is also strongly suppressed. Analog signals are visualized by recorder traces. As the time constants of recorders are rarely smaller than 0.3 s, the recorders are unsuitable for observation of processes shorter than 1 s.

In relation to the response rate of measuring instruments, the response time is defined, corresponding to the time required to attain 90% of the signal final value on a unit change in the analyte amount. The response time consists of the delay time interval

(attainment of 10% of the signal final value) and the increase or decrease interval (the dynamic behavior between 10% and 90% of the final signal). It should be emphasized that the response of nonphysical detectors (i.e., the chemical and biochemical ones) is usually nonsymmetrical, which is an important drawback of these detectors (**Figure 5**).

A digitally treated signal is produced by sampling an analog signal. The sampling consists of repetitive reading of the input (with a sampling frequency  $f$ ) over a certain time interval ( $w$ ) that is characterized by a mean value. The digital signal is not affected by the time constant and is suitable for observation of even very rapid processes because the sampling frequencies are high ( $f > 25$  kHz). However, the interval for the formation of the signal mean value is very narrow, usually  $\sim 40 \mu\text{s}$ , and thus the digital signal significantly 'copies' the noise of the analog signal.

The large number of instantaneous signal values (e.g., 25 000 values per second with a 25 kHz converter) cannot be directly used, both because the number is too large and because the values are subject to noise. Therefore, a certain time sequence of instantaneous signals is averaged over a longer time interval and the white noise is suppressed. The signal value thus obtained is characterized by an average value over a time interval that depends on the instrument construction and on the signal handling algorithm and is called a 'slice', whose values are stored in the instrument memory.

This signal handling provides rough data that represent a recording of the process observed. These rough data cannot be deconvoluted into the original signals, but their time shift compared to the original signal is negligible and does not affect the measurement result. A substantial advantage of the signal handling lies in reduction of the noise and creation of a set of data that can be evaluated repeatably.

Reduction of noise is a parallel step in the data treatment and is carried out by a number of procedures, e.g., in order of increasing sophistication,

- inclusion of the time constant ( $\tau$ );
- use of wavelets, i.e., of a digital, frequency-selective filter or band gate;
- use of Fourier transformation analysis (FTA); and
- use of adaptive digital filtering (ADF).

Appropriate software is available for all the noise reduction procedures, and the choice of approach depends on the aim of the particular analysis. The time constant approach (e.g., lengthening of the time

interval,  $t_{\text{int}}$ ) leads to suppression of all the noise types, the 'wavelet' method removes noise within a certain frequency range (digital, frequency-selective filters of the band gate (low-pass, high-pass)), FTA carries out frequency analyses over a wide range of frequencies, and ADF is a learning system capable of applying the experience obtained to new data sets.

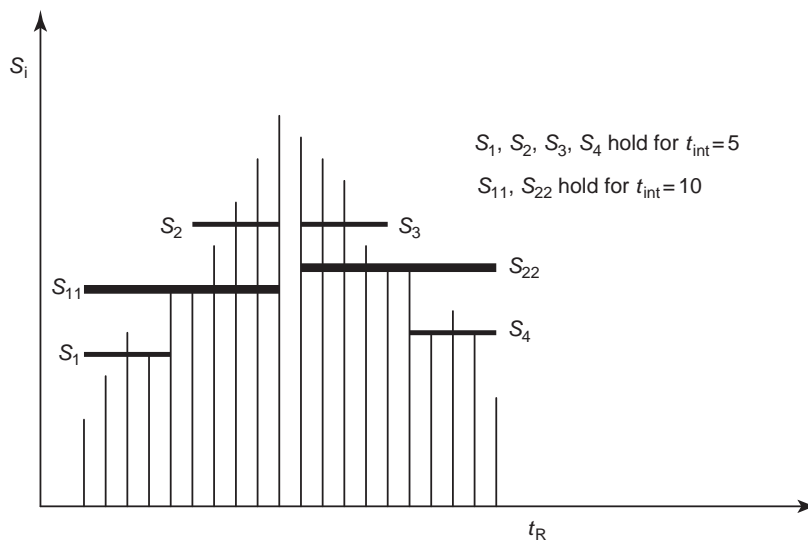
The disturbances to measurements also include sudden and intense changes in the measured signal, called spikes, originating from erroneous functioning of the measuring system. The spikes are not part of the noise because they do not exhibit a frequency structure and can only be removed during follow-up data treatment (postrun calculations). If spikes occur often, the hardware should be checked up.

The creation of a set of rough data ends the period of the 'operator passive role' in the measurement, during which the operator cannot influence the parameters of the electronic sampling, the A/D converter, the time constant of the measuring system, the slice integration interval, the parameters of the transfer of the rough data into the memory, etc.

The next phase of the signal treatment that affects the instrument calibration is an accumulation of instrument-dependent slices into an interval called a bunch, whose size is selected by the operator. During the treatment of the rough signal, all the partial signals are integrated over a certain time interval to which the mean signal value corresponds (Figure 6).

## Zero Line

In contrast to a set of rough data giving the magnitude of the overall signal, these newly created data



**Figure 6** A scheme for obtaining the mean signal value in dependence on the bunch length; this example shows the formation of four signal mean values at  $t_{\text{int}}=5$ , two values at  $t_{\text{int}}=10$ , and a single mean value at  $t_{\text{int}}=20$ .

represent signals above a predetermined zero line. The above treatment uses a suitable algorithm based on the finding that a time interval corresponding to a change in the observed process can be described by a limited number of equidistant time sections or points that are representative of the observed process. The length of the time section (bunch) depends on the character of the observed process and of the measuring method and is subject to optimization because a too-long section leads to suppression of sudden changes in the observed process, while a too-short section permits the data to be affected by noise.

If the time of the observed process changes during the measurement (e.g., the width of chromatographic peaks increases with increasing elution time), then it is necessary to maintain the number of points (bunch) describing the change constant in order to keep comparable conditions for earlier and later peaks. Under this condition, the number of rough data (slice) accumulated into a bunch increases with increasing observation time, but the observation conditions, the relative change in the signal between the bunches, the noise contribution to the signal, etc., remain independent of the observation time. The data corresponding to a bunch are characterized by the signal mean value, calculated as the mean of the slices contained in the bunch. Therefore,

$$S_{\text{bunch}} = \frac{\sum_n S_i}{n} \quad [4]$$

when the bunch length is given by

$$t_{\text{bunch}} = n * t_{\text{slice}} \quad [5]$$

where  $n$  is the number of rough data intervals and  $t_{\text{slice}}$  is the sampling interval given by the converter frequency.

This modification of the rough data suppresses random noise and permits determination of the signal zero line.

The determination of the zero signal is based on accumulation of the data into bunches, followed by differentiation analysis. It is assumed that the signal change caused by the analyte is larger than a possible random change in the zero line. It then follows from differentiation analysis of the data that the zero line corresponds to time intervals within which the first derivative of the signal lies below a certain predetermined value (the threshold value). The analytical signal is then any value whose first derivative exceeds the threshold value.

The length of  $t_{\text{bunch}}$  determines the shape of the signal zero line. If this value or the value of the signal derivative (threshold, in some data treatment software called the sensitivity) is too low, then the zero

line is subject to noise and the analyte signal can be erroneously attributed to noise. On the other hand, if the  $t_{\text{bunch}}$  or the derivative value is too large, then small changes in the analyte signal can be included erroneously in the zero signal (the baseline).

Figure 7 depicts the first derivative of a signal and demonstrates the importance of the  $t_{\text{bunch}}$  interval on the shape of the signal zero line. For a high frequency  $f_1$ , the zero line is 'not constant' because the threshold value is often exceeded; on the other hand, for a sufficiently large bunch length corresponding to  $f_3$ , the zero line is constant even at a low threshold value.

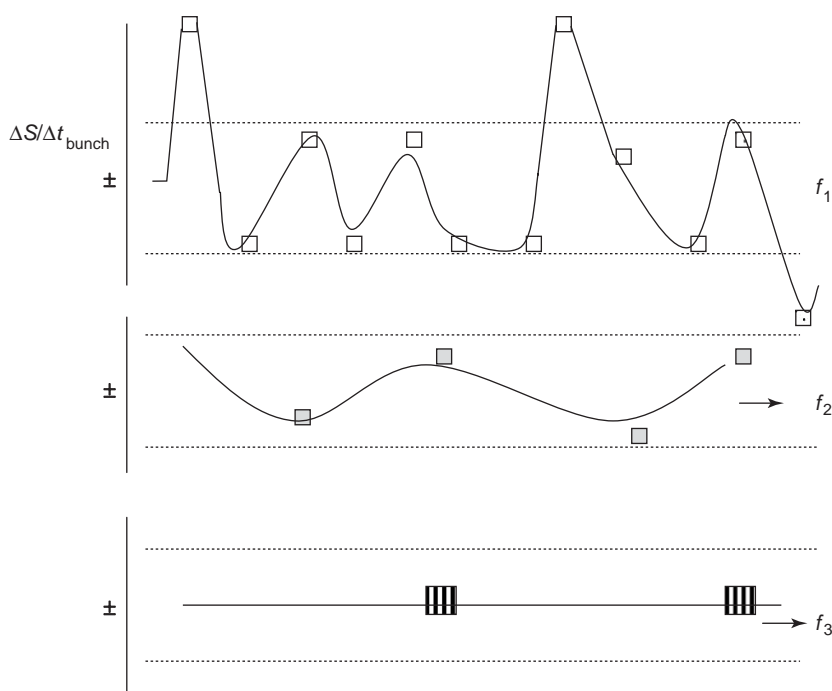
The area of an integrated signal is determined by the beginning of integration (the instant exceeding the predetermined threshold), the end of integration (the decrease in the signal first derivative below the threshold), and the zero line value between these two points. Because the analyte amount is determined from this area (sometimes from the signal height), the establishment of the zero line is decisive for the error of determination. The error increases with increasing tailing of the signal (increasing  $\tau/\sigma$  ratio) and with an increase in the ratio of the magnitudes of overlapping signals.

## Characteristics of Measuring Systems

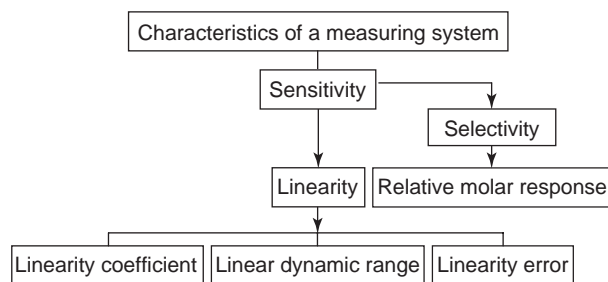
Observation in analytical chemistry is only based on interpretation of eqns [1]–[3]. It is thus understandable that evaluation of the variables in these equations has been standardized, especially in EN ISO/IEC 17025 and ISO 11843. The rules given below are in agreement with these standards and are part of the rules of good laboratory practice.

The necessity to know the properties of the conversion factor between the analyte amount and the analytical signal magnitude follows not only from the requirement of correct interpretation of the results of the measuring system but also from traceability requirements, i.e., demonstration of the chain of conversion of rough data into the analyte amount: 'traceability – property of the result of a measurement or the value of a standard whereby it can be related with a stated uncertainty, to stated references, usually national or international standards (i.e., through an unbroken chain of comparisons)' (IUPAC, 1998).

The hierarchic interconnection of the characteristics of the measuring system is shown in Figure 8 and involves the sensitivity from which the selectivity and relative molar response are derived and the linearity of the measuring system, composed of three parallel criteria, namely, the linearity coefficient, the linear dynamic range, and the linearity error.



**Figure 7** The effect of the  $t_{\text{bunch}}$  interval and of the threshold value on the determination of the signal zero value.



**Figure 8** Characteristics of a measuring system.

The sensitivity of a measuring system expresses the output/input transformation function and is the product of the construction properties ( $k$ ) and those specific to the analyte ( $a_i$ ). A rearrangement of eqn [1] leads to

$$k * a_i = \frac{S_i}{m_i^I} = \frac{1}{m_i^{(I-1)}} * \frac{S_i}{m_i} \quad [6]$$

from which it follows that the sensitivity of the measuring system will be constant and independent of the analyte amount,  $m_i$ , only when the linearity coefficient is equal to unity,  $I = 1$ . If this condition is not satisfied, then the sensitivity is a function of the analyte amount  $\frac{1}{m_i^{(I-1)}}$  and measurement over a wide analyte amount range is erroneous. Because a constant sensitivity value, independent of the analyte amount, is a prerequisite for accurate measurement, the output/input transformation function must be

linearized according to eqn [7]:

$$\sqrt[l]{k * a_i} = \frac{\sqrt[l]{S_i}}{m_i} \quad [7]$$

The dimension of the sensitivity corresponds to this definition and depends on the dimensions of the output (electrical) signal and of the input signal (analyte amount),  $\text{Vg}^{-1}$ ,  $\text{Appm}^{-1}$ ,  $\text{Hz mol}^{-1}$ ,  $\text{Vs g}^{-1}$ , etc. The dimension of the analyte amount follows from that of the sensitivity of the analytical determination.

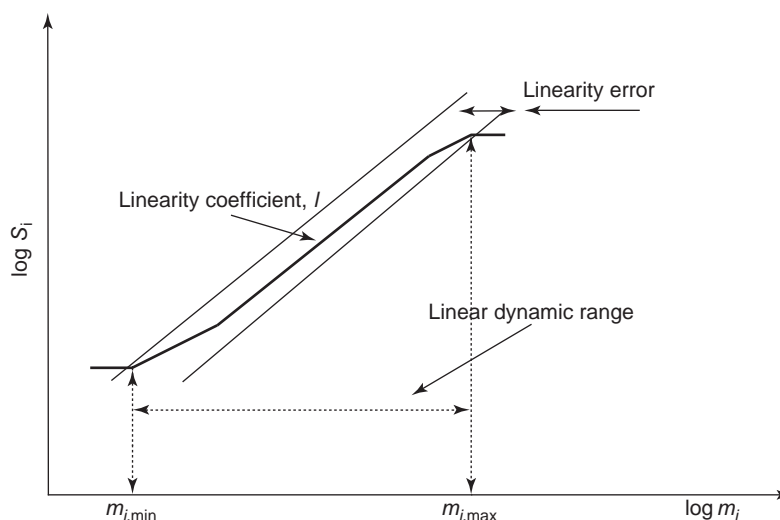
The selectivity,  $\Gamma_i$ , expresses the ratio of the sensitivities for two analytes in a given measuring system

$$\Gamma_i = \frac{k_i}{k_{st}} * \frac{a_i}{a_{st}} \quad [8]$$

It follows from eqn [8] that the selectivity is not a general property of the measuring system and that it is always relative and dependent on the selection of the standard analyte ( $a_{st}$ ).

The ratio of the magnitudes of the analytical properties of the analyte and standard ( $a_i/a_{st}$ ) in eqn [8], e.g., the thermal conductivity or ionization potential, is called the relative molar response ( $\text{RMR}_i$ ) and is used for normalization of the response magnitudes to a unit mass amount.

The linearity of the output of a measuring system is expressed by three parameters that describe the



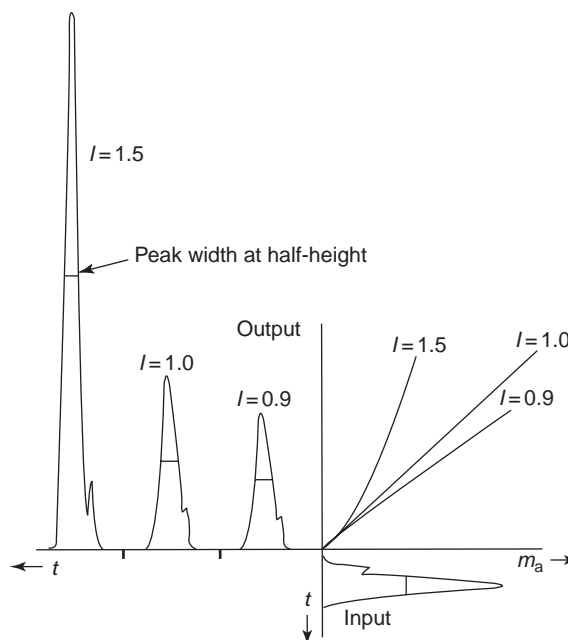
**Figure 9** Principal concepts of the measuring system linearity.

constant value of exponent  $l$  of the analyte amount causing signal  $S_i$  within the range of the analyte amount,  $\langle m_{i,\min}, m_{i,\max} \rangle$ , within a predetermined imprecision band (**Figure 9**). These parameters are called the linearity coefficient,  $l$ , the linear dynamic range, and the linearity error.

The linearity coefficient is defined as the slope of the  $\log(S) = -f(\log(m_i))$  dependence. It holds that

$$l = [\log S_i - \log(k * a_i)] / \log m_i \quad [9]$$

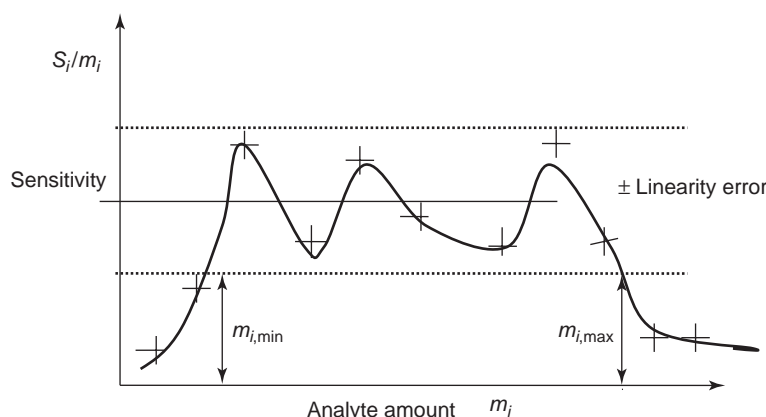
A measuring system is called linear when its linearity coefficient equals unity,  $l = 1$ . Only under this condition is the sensitivity constant and independent of the analyte amount. The linearity coefficient has the most pronounced effect on the detector signal, and it should be known for the used measuring system and the range of analyte amounts (for determination of the linearity coefficient, see below). Because the linearity coefficient is an exponential term, its effect on the signal values is large. A linearity coefficient greater than unity (e.g., with a flame-photometric detector) strongly alters the upper part of the peak and changes the value of its sharpness, whereas a linearity coefficient smaller than unity (e.g., with an electron capture detector) substantially lowers the upper part of the peak. The value of the linearity coefficient especially strongly affects large signals and thus can change the shape of an elution peak (**Figure 10**) and its characteristic quantities, e.g., efficiency, resolution, quantification. For nonunity linearity coefficients the sensitivity depends on the analyte amount (eqn [6]), and thus the determinations are subject to large errors.



**Figure 10** Distortion of the measuring system output caused by a varying linearity coefficient.

The linearity error represents a predetermined acceptable deviation from a constant sensitivity value, with a constant linearity coefficient. The linearity error is selected, and it is evident that selection of a larger error causes widening of the specified linear dynamic range but that it does not lead to improved linearization of the detector response. The linearity error should not be chosen greater than  $\pm 5\%$  to permit correct interpretation of the measured response.





**Figure 11** An example of evaluation of the measuring system parameters according to eqn [6]; the crosses denote the experimental points, the connecting line expresses the experimental changes in the calculated sensitivity.

The linear dynamic range is the range of the analyte amount within which it holds that the detector sensitivity is constant within the predetermined linearity error. Therefore, several linear dynamic ranges may exist of a measuring system, differing in the linearity coefficient. Practical measurements have a problem with distorted shapes of the response curve, as depicted in **Figure 10**. With dynamic measurements, e.g., elution profiles, the analyte passes through several linear dynamic ranges, and the detector response is a function of several linearity coefficients. Therefore, an evaluation of the instantaneous response (peak (signal) height) of the detector then leads to a more accurate representation of the function  $S = \log(m_i)$  than does an evaluation of the detector integral response.

The linear dynamic range of a measuring system can be obtained from the linearized form of the detector response, i.e., from the logarithmic form of eqn [1] (**Figure 9**) or from eqn [6], as demonstrated in **Figure 11**. The latter procedure does not permit determination of the linearity coefficient and can only indicate its deviations from unity.

Determination of the characteristics of a measuring system requires measurement of the detector response over a wide range of the analyte amount. In order to shorten the time required and to suppress errors caused by repeated analyte injection, continuous feeding of the analyte to the detector is advantageous, provided that a defined concentration profile is insured, e.g., by logarithmic dilution or through the concentration steps obtained in a step dilutor. Methods employing repeated injections of various analyte amounts are time consuming, are subject to a great dispersion, and should be used only exceptionally.

*See also: Quality Assurance: Quality Control; Spectroscopic Standards; Laboratory Information Management Systems.*

## Further Reading

IUPAC (1998) *Compendium of Analytical Nomenclature*, definitive rules 1997 ('Orange Book'), 3rd edn. Oxford: Blackwell.

## Interlaboratory Studies

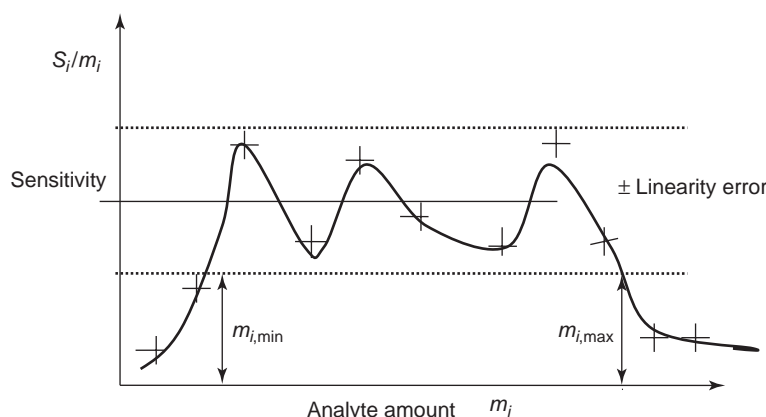
**D B Hibbert**, University of New South Wales, Sydney, NSW, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

One of the corner stones of quality control procedures is the analysis of test material by a number of laboratories, the so-called interlaboratory studies. The

results of the analyses allow comparisons to be made between, and information to be obtained about, the laboratories, methods, or the test materials. The laboratories may come from within an organization, or may encompass national metrology institutes across the world. The quantity value of the measurement may be known, or the object of the study may be to arrive at a consensus value. Common to all interlaboratory trials is one organization that takes responsibility for sourcing the material, sending it out



**Figure 11** An example of evaluation of the measuring system parameters according to eqn [6]; the crosses denote the experimental points, the connecting line expresses the experimental changes in the calculated sensitivity.

The linear dynamic range is the range of the analyte amount within which it holds that the detector sensitivity is constant within the predetermined linearity error. Therefore, several linear dynamic ranges may exist of a measuring system, differing in the linearity coefficient. Practical measurements have a problem with distorted shapes of the response curve, as depicted in **Figure 10**. With dynamic measurements, e.g., elution profiles, the analyte passes through several linear dynamic ranges, and the detector response is a function of several linearity coefficients. Therefore, an evaluation of the instantaneous response (peak (signal) height) of the detector then leads to a more accurate representation of the function  $S = \log(m_i)$  than does an evaluation of the detector integral response.

The linear dynamic range of a measuring system can be obtained from the linearized form of the detector response, i.e., from the logarithmic form of eqn [1] (**Figure 9**) or from eqn [6], as demonstrated in **Figure 11**. The latter procedure does not permit determination of the linearity coefficient and can only indicate its deviations from unity.

Determination of the characteristics of a measuring system requires measurement of the detector response over a wide range of the analyte amount. In order to shorten the time required and to suppress errors caused by repeated analyte injection, continuous feeding of the analyte to the detector is advantageous, provided that a defined concentration profile is insured, e.g., by logarithmic dilution or through the concentration steps obtained in a step dilutor. Methods employing repeated injections of various analyte amounts are time consuming, are subject to a great dispersion, and should be used only exceptionally.

**See also:** **Quality Assurance:** Quality Control; Spectroscopic Standards; Laboratory Information Management Systems.

## Further Reading

IUPAC (1998) *Compendium of Analytical Nomenclature*, definitive rules 1997 ('Orange Book'), 3rd edn. Oxford: Blackwell.

## Interlaboratory Studies

**D B Hibbert**, University of New South Wales, Sydney, NSW, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

One of the corner stones of quality control procedures is the analysis of test material by a number of laboratories, the so-called interlaboratory studies. The

results of the analyses allow comparisons to be made between, and information to be obtained about, the laboratories, methods, or the test materials. The laboratories may come from within an organization, or may encompass national metrology institutes across the world. The quantity value of the measurement may be known, or the object of the study may be to arrive at a consensus value. Common to all interlaboratory trials is one organization that takes responsibility for sourcing the material, sending it out

to the participating laboratories, receiving and statistically analyzing the results, and publishing a report. Exactly what is done in an interlaboratory study depends on the purpose of the study, that is what is being studied and why? In this article the different ways of setting up interlaboratory studies and the statistics used to analyze results will be described.

### Types of Interlaboratory Studies

Table 1 gives the major types of interlaboratory studies.

In each case a referee laboratory is responsible for providing the sample and organizing its distribution and the collation and analysis of the results. For the majority of studies the purpose of the study is to test the laboratory (proficiency testing), the test material (material certification), or the method (method validation study). The purpose determines some of the study parameters. For example, a method validation study requires all the laboratories to use the same method, but for other studies the method may not be prescribed. For materials certification the laboratories are chosen for their demonstrated proficiency in analysis of the material. It would not be appropriate if there was any dubiety concerning the assigned quantity value because of a suspect result.

### General Methodology

All interlaboratory studies have a common overall procedure:

1. A referee laboratory is appointed to oversee the study.

2. The participating laboratories are chosen. The criteria for choice of the number and nature of the laboratories depend on the kind of trial and will be discussed below.

3. The test material is sourced. For any interlaboratory study to be at all useful the material distributed to the laboratories must be homogeneous, i.e., each laboratory must receive a representative portion, and must be stable during transport and storage. For studies such as the Comité Consultatif de la Quantité de la Matière (CCQM) Key Comparisons in which material, often in complex matrices, is sent across the world, stability is a major concern. Homogeneity is assessed by analyzing, under repeatability conditions by a suitably precise method, at least 10 test samples. The within subsample homogeneity is established by analyzing about eight subportions of one subsample. These tests should be done for all materials that are not capable of being completely homogeneous (i.e., other than stable solutions).

4. If this is a requirement, the test material must be prepared or analyzed so that the quantity value is assigned. In interlaboratory comparisons of elements in water organized by IRMM under the International Measurement Evaluation Program (IMEP), the test materials were prepared by gravimetry or analyzed by a primary method such as isotope dilution mass spectrometry by an independent laboratory.

5. Test portions are sent to the participating laboratories with instructions and timetable for reporting measurement results.

6. On receipt of the measurement results, the organizing laboratory collates the data, and undertakes the

**Table 1** Types of interlaboratory study

<i>Name</i>	<i>Purpose</i>	<i>Comments</i>
Proficiency testing or laboratory performance study or round-robin study	To test the ability of a laboratory to obtain similar results to peer laboratories	Often employed as part of an accreditation scheme (e.g., to ISO/IEC 17025)
Cooperative trial	One-off comparison of laboratory performance	May be for contractual purposes
Material certification	To assign a consensus value to a test material	Used as part of a certification scheme, although this does not ensure traceability
Collaborative trial or method performance study or method precision study	To provide data for a validation study of a method	Determines the reproducibility precision of a method and if a CRM is used, the method or laboratory bias
Interlaboratory bias study	To determine method bias or laboratory bias of a standard method	Similar to a collaborative trial but with a specific aim
Improvement schemes	Validation of new or improved methods by comparison with fully validated method	Less costly exercise than full validation
Key comparisons	Assessment of national capability in analysis of materials	Organized by the CCQM and covers important areas related to trade, the environment and human health
IMEP		Organized by IRMM for the European Union

statistical analysis of the results. Of importance in any study is the treatment of outliers.

7. A report is distributed to the participating laboratories and may be made public as part of a process of deciding if a method is to be adopted as a standard, or if the study is a Key Comparison. Usually, the performance of individual laboratories is not disclosed by the use of codes to anonymously label the participants.

## Method Performance Study

National and international bodies that provide methods of analysis that have undergone a high level of testing and scrutiny will organize interlaboratory studies to determine the repeatability and reproducibility of a method. If the test material has a certified value then the method bias may be determined.

### Choice of Material and Laboratories

The participating laboratories should be chosen for their known ability to perform the method, or similar methods. As the method will be released for use by any competent laboratory, the choice of participants should reflect this. It may be that in a preliminary round the laboratories are sent practice samples. Indeed, a pilot trial may be done to provide an initial estimate of the precision of the method, which may give an opportunity to discontinue the study if it transpires the method does not have a minimum standard of performance. A pilot trial may also allow the method to be completely specified and any anomalies in the written procedure to be clarified before the full study.

To obtain reasonable statistics at least eight laboratories should participate in the study, although for adequate estimation of laboratory systematic errors 15 laboratories should be used.

### Measures of Precision

Repeatability precision is defined as a standard deviation of a mean result determined under conditions in which the same analyst performs a series of independent measurements of a test material with the same measurement system in a short period of time. The repeatability precision represents the smallest variability identified in an experiment.

Reproducibility precision is a standard deviation of a mean under conditions in which one or more aspects of the measurement system are changed. Interlaboratory reproducibility perforce changes the analyst, instrument, reagents, and so on. Intralaboratory reproducibility (sometimes known as intermediate

reproducibility) addresses changes within a laboratory, for example, by changing analyst, instrument, or by doing the analysis over an extended period of time.

For a method precision study, the method, of course, is fixed. The mean result  $\bar{y}$ , follows a model:

$$y_i = \bar{y} - B - \varepsilon \quad [1]$$

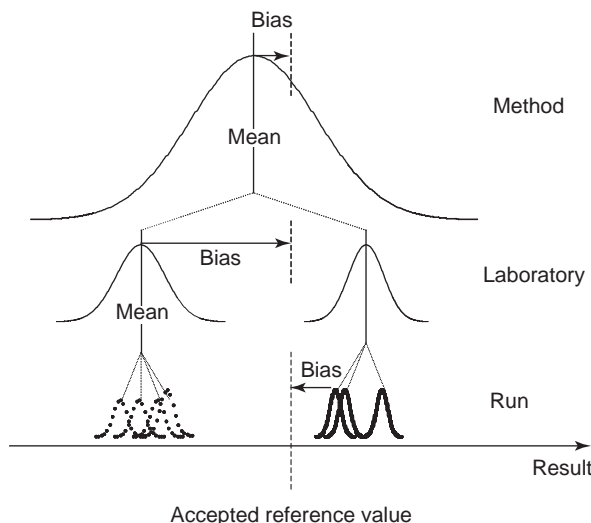
where  $y_i$  is the value obtained by laboratory  $i$ ,  $B$  the laboratory bias, and  $\varepsilon$  the within laboratory random error (which has an expectation of zero and standard deviation of  $s_r$ , the repeatability). Across a number of laboratories  $B$  will average to the method bias with standard deviation the between laboratory standard deviation ( $s_L$ ). Together these combine to give the reproducibility standard deviation ( $s_R$ ):

$$s_R = \sqrt{s_r^2 + s_L^2} \quad [2]$$

Thus, from the point of view of a laboratory  $B$  is a bias, while considering the interlaboratory study the  $B$ s contribute to the measurement uncertainty. This is illustrated in Figure 1.

### Method Performance Experiments

As precision is often dependent on concentration a sufficient range of samples should be chosen to obtain information across the method range. At least five (International Union of Pure and Applied Chemistry) or six (International Organization for



**Figure 1** Replicate results in a number of runs showing run bias and repeatability variance are combined to give the laboratory mean (hence laboratory bias) and intermediate reproducibility and further to the grand mean for the method (hence method bias) and reproducibility variance.

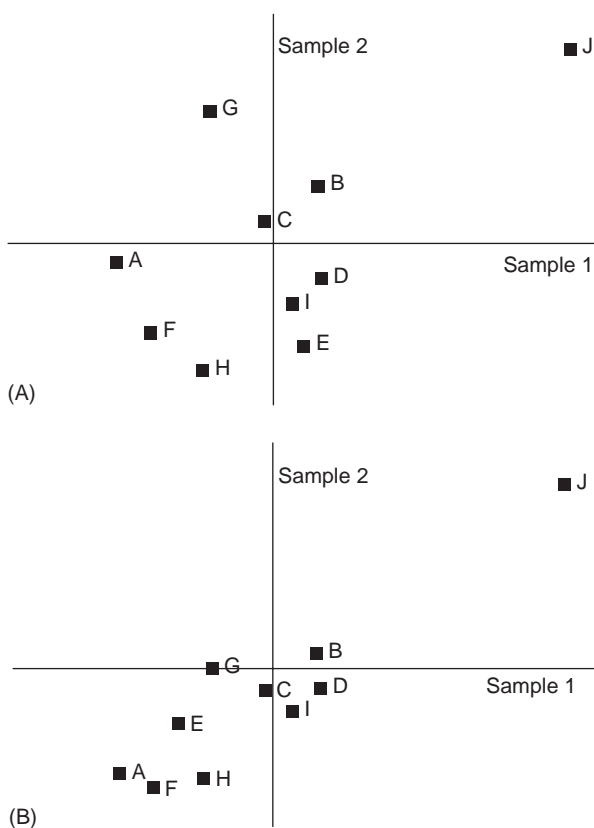
Standardization) concentration levels are recommended to be analyzed in duplicate. As the range of materials defines the limits of the study, it is up to the organizers of the study to choose an appropriate set of materials. A pharmaceutical company analyzing a product with a likely narrow range of concentration of analyte may decide to perform fewer analyses than those recommended for other materials. To avoid possible unconscious effects from a laboratory knowingly performing duplicate measurements, when results are found to have a smaller standard deviation than would have been the case in a completely blind trial, split-level experiments may be done, in which test materials having slightly different concentrations are submitted for a single analysis each.

Because the aims of method performance studies are to obtain estimates of the repeatability and reproducibility of a method, it is important that the data are shown to conform to the statistical assumptions made about them, namely that the laboratory means are distributed normally, and that there is an equality of variance among laboratories for each concentration level. A Cochran test is recommended to test the homogeneity of variances, followed by a Grubbs test for single and double outliers. Laboratory results that fail these tests should be removed from consideration as long as no more than two out of nine of the participating laboratories are so removed. If too many outliers are discovered the program is considered to be flawed and should be repeated after investigating causes of the outlying results.

If split-level samples have been sent to laboratories a Youden plot may be used to identify trends among laboratories and determine outliers. In such a plot the mean for the sample is subtracted from each result and the two results from a laboratory are plotted as a scatter plot against each other (Figure 2). Outliers are seen as isolated points away from a group, and the greater the tendency for laboratory bias, the more the points will lie in the first and third quadrants.

### Statistical Treatment of Results

If some data are missing, for example, if not all results are returned from a particular laboratory, the organizers must decide whether to include what data are reported or to exclude the laboratory altogether. Exclusion of the laboratory is generally favored unless this action would result in a lower than acceptable number of laboratories remaining for the statistical analysis.



**Figure 2** (A) Two-sample (Youden) plot of data from 10 laboratories showing a potential outlier (J) and mostly random scatter of results. (B) Two-sample (Youden) plot of data from 10 laboratories showing a potential outlier (J) and a large component of laboratory bias.

The organizing laboratory will collect the results, test for normality of variances and outliers, and then determine repeatability precision and reproducibility precision, either directly or from an analysis of variance (ANOVA) on the data. A one way ANOVA, with the laboratories being the factor studied, will give the repeatability variance as the within groups mean square, and the between groups mean square is the repeatability plus the number of replicates times the laboratory variance. For duplicate determinations these quantities may be obtained directly from the means of the two results and their differences, which will subtract out the laboratory bias.

For the  $i$ th laboratory out of  $k$  returning two measurement results,  $x_{1,i}$  and  $x_{2,i}$ , the repeatability variance is the pooled variance of  $k$  pairs of differences:

$$s_r^2 = \frac{\sum_{i=1}^{k-1} (x_{1,i} - x_{2,i})^2}{2k} = \frac{\sum_{i=1}^{k-1} d_i^2}{2k} \quad [3]$$

The sample variance, with  $k - 1$  degrees of freedom, of the means of the two measurement results ( $\bar{x}_i$ )



across the laboratories ( $s_{\bar{x}}$ ) is related to the interlaboratory variance

$$s_{\bar{x}}^2 = \frac{\sum_{i=1}^{k-1} (\bar{x}_i - \bar{\bar{x}})^2}{k-1} \quad [4]$$

where  $\bar{\bar{x}}$  is the mean of all results

$$s_L^2 = s_{\bar{x}}^2 - s_r^2/2 \quad [5]$$

from which the reproducibility variance follows from eqn [2].

The estimate of reproducibility may be compared with the expected value from the Horwitz curve. An analysis of 7500 method performance studies led to the proposal that the reproducibility expressed as a percentage of the concentration of the analyte ( $R\%$ ) was related to the concentration ( $c$ ) by

$$\log_2(R\%) = 1 - 0.5 \log_{10}(c) \quad [6]$$

where  $c$  is expressed as a fraction. For example,  $1 \mu\text{g per g} = 1 \text{ ppm} = 1:1 \times 10^{-6}$ , and hence  $R\%$  is estimated to be  $2^{1-0.5 \times (-6)} = 16\%$ . It is normally expected that a reproducibility determined in an interlaboratory study should fall within a factor of 2 of the Horwitz  $R\%$ .

Other statistical treatments that may be employed are Mandel's  $b$ - and  $k$ -statistics, box plots, and one way ANOVA when replicate data are available. When samples are analyzed at different concentrations, the relationship between precision estimates (repeatability and reproducibility) and concentration should be established.

## Reporting

Because the aim of the study is to provide information to laboratories using analytical methods, reports of method validation studies are usually made public. They comprise the specification of the analyte, a description of all materials, the number of laboratories taking part, the number of outliers, the grand mean, the assigned value (if it has been established), and the repeatability and reproducibility. The organizing panel must finally decide if the results of the study are acceptable.

## Interlaboratory Bias Study

A particular and limited form of a method performance study is one to determine the bias of a standard method, or the bias introduced by laboratories that perform the standard method. The organization of the study is much the same as for general method performance studies, but no pilot study is needed, and the number of laboratories and replicates is chosen on a statistical basis in order to

be able to determine either that the method does not have significant bias, or if the method has more than a stated *a priori* bias.

As the repeatability and reproducibility standard deviations should already be known, the values determined by the study may be tested against them by a  $\chi^2$  test. If the test reveals a significant difference, an investigation must be undertaken to decide why this is so, with possible repeats of the analyses. The method bias is calculated as the difference between the grand mean of the results of the laboratories and the assigned value of the measurand of the test material.

## Laboratory Performance (Proficiency) Tests

Proficiency tests are conducted by accrediting authorities or similar bodies as a method of comparing the performance of laboratories and maintaining their accreditation. Usually, these studies are ongoing with an appropriate frequency. Single proficiency tests, called cooperative trials, as a one-off assessment of a laboratory's ability to perform a particular test, may be undertaken among laboratories being considered for a contract for outsourced chemical analysis.

For continuing proficiency testing schemes laboratories are able to compare their measurement result (1) with their own past performance, (2) with other laboratory's present results and, if the material has an assigned value of the measurand (3) the true value.

## Organization

Unlike a collaborative trial, the organizing committee of a proficiency test must be strictly independent of the participating laboratories. As the status of the accreditation of a laboratory is often at stake the process must be scrupulously fair and must be seen to be so. In many cases the organization of a proficiency test is by a government laboratory or other statutory body. These bodies themselves can now be accredited as a statement of their competence to oversee proficiency testing schemes.

The general role of the organizer is as laid out above. It is important that all participating laboratories are aware of the rules and conditions of the test, and that they have indicated their agreement to a detailed plan. If results are to be used as part of compliance assessment to an accreditation program, then this is made clear beforehand.

## Choice of Materials, Methods, and Laboratories

The test material is chosen to be as similar to a routine sample as possible. If part of long-term trials,



the material used for each round must be different from those used in earlier rounds, and participants should be aware of this. Laboratories should not be tempted to 'second guess' the expected result from extrapolations of earlier rounds.

There is no requirement to use a particular method; only the result is of importance, but each laboratory should state the method used and the repeatability and reproducibility precision should be known. This will allow assessment of the overall performance of the laboratories against the expected precision.

Laboratories may be invited to participate as a condition of their accreditation, or through general advertisements or Internet announcements. It is recommended that all participants remain anonymous to each other and to other parties. This is not always possible or desirable. Certainly for any report of results of a round of the proficiency test the laboratories should be coded with any laboratory knowing only its own code. However, there may be some benefit from having occasional open meetings to discuss issues among laboratories. There is no requirement for any particular number of laboratories, but with fewer than five it is not possible to derive meaningful statistics. In practice, there is often a large number of laboratories (18–30 are typical) because of the economies of performing the trials. The benefit to potential customers is increased with the participation of a larger number of laboratories.

### Proficiency Test Procedures

The number of test materials will be determined by the requirements of the trial, the range of concentrations, and the matrices that must be considered. Some authorities are content with sending a single material and the harmonized protocol indicates that a maximum of six test materials should be sent out. Duplicates or split-level samples may be included, but as the laboratory must decide its own procedures these are not *a priori* necessary.

As outliers as such are laboratories that must be scrutinized and reasons for their poor performance understood, it is not good practice to remove them from consideration as happens in a method performance study. Unless there is a target mean and target measurement uncertainty (standard deviation) against which the results of all laboratories are assessed, the organizers must make a decision as to the actions to be taken with results that appear to be outlying. There are two approaches. First, outlier tests can be applied such as Cochran's test for laboratory variances (if

replicate results have been reported) or the Grubbs' test for single or two outliers of laboratory means. If one or two laboratories are rejected, the consensus mean and standard deviation of the remaining laboratories are calculated. Alternatively, a robust method is used to estimate the mean and standard deviation that can tolerate outlying results while providing useful information about the main body of laboratories.

Results are usually presented as  $z$ -scores, which are the results expressed as deviations from an accepted value (mean,  $\mu$ ) as a number of (given or target) standard deviations ( $\sigma$ ):

$$z = \frac{x_i - \mu}{\sigma} \quad [7]$$

or from sample means and standard deviations

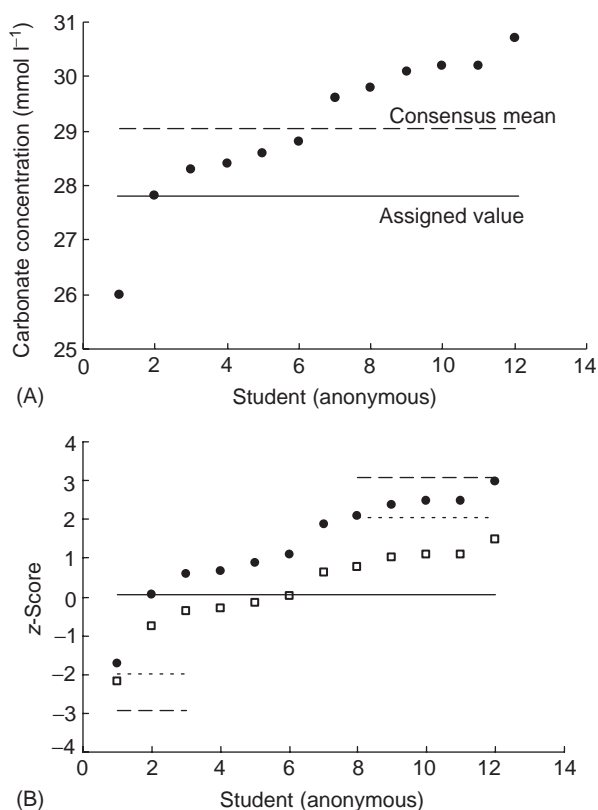
$$z = \frac{x_i - \bar{x}}{s} \quad [8]$$

Because the use of a sample mean, and particularly a sample standard deviation, relies on the assumption of normality of distribution of the data, outliers in the data must be suitably dealt with. Alternatively, robust  $z$ -scores may be calculated from the median and the normalized interquartile range (norm IQR).

$$z_{\text{robust}} = \frac{x_i - \text{median}(x)}{\text{norm IQR}} \quad [9]$$

Laboratories for which concern must be shown will be indicated as outliers or questionable results ( $2 < |z| \leq 3$ ) or extreme outliers ( $3 < |z|$ ). Where a number of samples, perhaps at different concentrations, have been analyzed a ranking method may be applied. For each sample (concentration level) the laboratories are ranked by the closeness of their results to the consensus or accepted mean, and then the ranks are summed over the samples. Laboratories with consistently bad performance will appear with the highest scores. It is also possible, but not necessarily recommended, to combine a number of  $z$ -scores from one laboratory. The sum of squared  $z$ -scores is distributed as  $\chi^2$  and as it ignores the sign of  $z$  more reflects the aim of the test, namely to highlight laboratories that have outlying results, whether too high or too low. The relative laboratory performance is the mean of the summed  $z$ -scores for a laboratory.

Figure 3 shows some data from a study conducted in the author's laboratory. Twelve students were given a test solution of a mixture of sodium carbonate and sodium bicarbonate. The results shown here are for the amount of substance concentration of carbonate. The data illustrate that



**Figure 3** (A) Ordered results from a proficiency test of 12 analysts. The amount of substance concentration of carbonate in the distributed test solution was assigned as  $27.8 \pm 0.2 \text{ mmol l}^{-1}$  with a target standard deviation for the group of  $1.0 \text{ mmol l}^{-1}$ . The consensus mean was  $28.9 \text{ mmol l}^{-1}$  with a sample standard deviation of  $1.3 \text{ mmol l}^{-1}$ . No points were outliers by a Grubbs test. (B) Data of (A) transformed to a z-score. Filled circles:  $z_i = ((x_i - \text{Assigned value}) / \text{Target standard deviation})$ , Open squares:  $z_i = ((x_i - \text{Consensus mean}) / \text{Sample standard deviation})$ . Lines are drawn at  $z = \pm 2$  and  $\pm 3$ .

significant differences in the interpretation of results can be made if the consensus mean and sample standard deviation are used to calculate z-scores. As may be seen, within the group there were no outliers, and only one student received a z-score greater than 2 ( $-2.3$ ). However, if the assigned value and a reasonable target uncertainty were applied, a number of results become suspect. Because the data contained no outliers and are symmetrically distributed about the mean, there was no great difference if the median and normalized IQR were used.

### Reporting of Proficiency Tests

The report should contain a clear statement of procedures for dealing with the data and the determination of test statistics (e.g., treatment of missing data, outliers, and method of calculation of z-scores). Usually quoted are the valid measurement results,

coded laboratory means, grand mean, and repeatability and reproducibility. Graphs of the z-scores against laboratory code, or ordered z-scores or laboratory means are often reported. Where split-level samples have been distributed, a Youden plot is usually given.

### Material Certification

Reference materials (RMs) and certified reference materials (CRMs) are used widely in quality control of analytical laboratories. Traceability of a measurement result is established through calibrators with traceable quantity values. In field laboratories making routine measurements the calibrators are usually CRMs, or working standards derived from CRMs, for which the certificate is the documentation of the traceability chain, usually to the unit in which the quantity value is expressed. At other stages in the analytical procedure RMs may be introduced to determine recovery and to establish run bias. To be a link in a traceability chain, a CRM must have its quantity value measured by comparison with a calibrator, whose value is established ultimately by the definition of the unit. A traceable quantity value is not *per se* established by the consensus of a number of laboratories, but if each of those laboratories can show they make a traceable measurement then the weight of the group of these high-level measurements is compelling evidence that the combined measurement result is as good an estimate as possible of the true value.

The organizer of an interlaboratory study to assign a quantity value to a reference material is usually a national or international certifying authority, such as the International Atomic Energy Agency (United Nations), the National Institute of Standards and Technology (USA), the Laboratory of the Government Chemist (UK), or the Community Bureau of Reference (EU).

### Choice of Materials, Methods, and Laboratories

Homogeneity and stability must be assured for the study to have any use at all. Usually, one laboratory has the responsibility for performing tests. Subsamples will be stored at different temperatures ranging from, say  $-20^\circ\text{C}$  to  $+50^\circ\text{C}$  and then analyzed once a month.

Use of a single method by a small number of laboratories runs the risk of introducing a method bias into the result. It is recommended that the laboratories are chosen for their high standard of analytical capability, and their ability to apply different methods, where this is appropriate. If necessary the laboratories will also be asked to employ different pretreatment methods. A large

number of laboratories should be chosen, with 20 in the initial phase of the certification.

### Procedure for Materials Certification

It is recommended that there be one or two rounds of preliminary tests to establish homogeneity and a likely consensus. A second round may be needed to allow understanding of any sources of inhomogeneity between laboratory results. Only those laboratories that achieve reasonable results in the initial rounds participate in the final certification round. It is important, however, to make sure that the one laboratory that does not agree with the results of the majority is not, after all, the only one with the right answer! At least eight determinations should be done in each laboratory to allow the estimation of the within laboratory repeatability. For example, four subsamples might be analyzed in duplicate. If a blank sample is also distributed, recovery experiments on the spiked blank may be performed. Samples may be distributed on different days to give information on another influence factor, time.

Normality of the distribution of results is checked by the Kolmogorov–Smirnov test and outlier tests are performed on variances and means. ANOVA is used to calculate the different contributions to the variance, between laboratory, between method, and within laboratory.

If the statistics indicate nothing untoward, the certified value is the grand mean of the results with a

95% confidence interval calculated from the standard deviation of the mean. A better estimate of the range containing the true value with 95% probability will be obtained from a full measurement uncertainty calculation, which should reflect the statistics of the trial.

### The IMEP®

In recent years, there has been an emphasis on the demonstration of national capability by the analysis of material of known composition. The IMEP® is carried out under the auspices of the European Union by the Institute for Reference Materials and Measurement (Geel, Belgium). Starting in 1989 there have been 19 rounds of IMEP®, which are detailed in Table 2.

Participating laboratories receive well-characterized samples with undisclosed values for analysis. These samples have been made, or characterized to have, certified reference values, with established traceability, and small stated measurement uncertainty. These values are established independently of any of the participating laboratories and are made known with the results of the study. This allows a direct comparison of a laboratory's result with the certified value, and so there is no requirement for statistical evaluation of means, or standard deviations that are characteristic of other studies. IMEP calls this result-oriented evaluation of measurement capacity.

**Table 2** The IMEP® rounds

<i>IMEP comparison</i>	<i>Material and matrix</i>	<i>Elements</i>	<i>Years of study</i>
19	Rice	Cu, Cd, Zn, and Pb	2002–03
17	Serum	Trace and minor constituents	2000–03
16	Wine	Pb	1999–2001
15	Water	Trace elements	2001–02
14	Sediments	Trace elements	1999–2000
13	Polyethylene	Cd, Cr, Hg, Pb, As, Cl, Br, S	1999–2000
12	Water	(Ag, B, Ba, Ca, Cd, Cu, Fe, K, Li, Mg, Mo, Ni, Pb, Rb, Sr, Tl, Zn)	2000–01
11	Car Exhaust Catalysts	Pt, Zr, Ce, Hf	1998–99
10	Polyethylene	Cd, Cr, Hg, Pb, As, Cl, Br, S	1997–98
9	Water	B, Ca, Cd, Cr, Cu, Fe, K, Li, Mg, Ni, Pb, Rb, Sr, U, Zn	1998–99
8	CO <sub>2</sub>	$n(13C)/n(12C)$ and $n(18O)/n(16O)$	1997–99
7	Human serum	Ca, Cl, Cu, Fe, K, Mg, Na, Se, Zn	1997–98
6	Water	Ag, B, Ba, Cd, Cu, Fe, Li, Mo, Ni, Pb, Rb, Sr, Tl, Zn	1994–95
5	Human serum	Fe	1991–94
4	Bovine serum	Li, Cu, Zn	1991–95
3	Water	B, Ca, Cd, Cu, Fe, K, Li, Pb, Rb, Zn	1991–93
2	Polyethylene	Cd	1990–91
1	Human serum	Li	1989

The matrices and elements that are studied target international trade, health, environmental issues, or politics that are associated with between country issues, and where possible are drawn from real-life samples. Laboratories are given the freedom to choose the method, measurands, or components that are relevant to their interests, and are encouraged to use their routine measurement methods. Usually, each laboratory is asked to report, as well as the value of the measurand, an expanded measurement uncertainty within which the value of the quantity being measured is claimed to be with a stated probability. The laboratories are also asked to indicate the method used, whether or not they consider themselves to be expert in the analysis, and to what guides or standards they may be accredited. As with any interlaboratory proficiency study full anonymity of the laboratories is guaranteed.

The results, without identifying the participating laboratories, are available on the website of IMEP<sup>®</sup> together with a breakdown by method, accreditation status, and self-certified experience. The results have been extremely illuminating, with many rounds showing little correlation between the quality of a laboratory's result and any of the collected parameters. The very first round, of lithium in serum, gave an excellent example of the problem of studies in which a consensus mean is used to compare performance. In this study, the one apparently outlying laboratory was the only one to achieve a result that was reasonable, with all other participating laboratories reporting results about twice the certified value.

## Key Comparisons of the BIPM

The Bureau International des Poids et des Mesures (BIPM) is the organization charged with overseeing the International System of Units (the SI system), and as such needs to be sure that standards maintained by nations carry the values they purport to carry. In 1999, the BIPM started the Key Comparisons program in which countries sign mutual recognition agreements (MRAs) to accept standards and if desired calibration and measurement certificates issued by other signatories' national measurement institutes (NMIs). To maintain membership of the MRA, an NMI must take part in rounds of the Key Comparisons. The Key Comparisons organized by the CCQM have consisted of a wide variety of matrices and measurands. Presently, 52 have been completed and eventually 80 are planned to cover all areas of chemical measurement. As with IMEP, the attempt is to cover all important areas that are involved in international commerce. Examples include health (cholesterol in fish), food (arsenic in fish),

environment (permanent gases in air), advanced materials (semiconductors), commodities (sulfur in fossil fuels), forensics (ethanol in air – breathalyzer), biotechnology (DNA profiling), and general analysis (pH).

Unlike the IMEP the laboratories are completely identified, as the point of the program is to demonstrate capability.

**See also:** **Quality Assurance:** Quality Control; Primary Standards; Reference Materials; Production of Reference Materials; Method Validation; Accreditation.

## Further Reading

- Association of Official Analytical Chemists (1989). Guidelines for collaborative study procedures. *Journal of the Association of Official Analytical Chemists* 72: 694–704.
- Association of Official Analytical Chemists (1995) AOAC/IUPAC revised protocol for the design, conduct and interpretation of method-performance studies. *Journal of the Association of Official Analytical Chemistry International* 78: 143A–160A.
- Horwitz W (1988) Harmonized protocol for the design and interpretation of collaborative studies. *Trends in Analytical Chemistry* 7: 118–120.
- International Organization for Standardization (1997) *ISO/IEC Guide 43-1:1997 (E) Proficiency Testing by Interlaboratory Comparisons Part 1: Development and Operation of Proficiency Testing Schemes*. Geneva: ISO.
- International Union of Pure and Applied Chemistry; Analytical, Applied and Clinical Chemistry Divisions (1988) Protocol for the design, conduct and interpretation of collaborative studies. *Pure and Applied Chemistry* 60: 855–964.
- Lawn RE, Thompson M, and Walker RF (1997) *Proficiency Testing in Analytical Chemistry*. London: Royal Society of Chemistry.
- Massart DL, Vandeginste BGM, Buydens LMC, *et al.* (1997) *Handbook of Chemometrics and Qualimetrics. Part A*, ch 14. Amsterdam: Elsevier.
- The website for the International Measurement Evaluation Program – <http://www.imep.ws>.
- The website for the Key Comparison program of the BIPM – [http://www.bipm.fr/enus/8.Key\\_Comparisons](http://www.bipm.fr/enus/8.Key_Comparisons).

## Glossary

<b>Interquartile range (IQR)</b>	The range of values that contains the middle 50% of data
<b>Median</b>	The middle value of data arranged in order (odd number of data), or the mean of the middle two values (even number of data)
<b>Normalized interquartile range (norm IQR)</b>	norm IQR = IQR/0.735. This is an estimate of the range that encompasses 68% of the data about the median, i.e., is an estimate of the standard deviation

## Reference Materials

P Quevauviller, Drève de Nivelles, Brussels, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Each week, a huge number of laboratories worldwide each produce perhaps a thousand or more analytical data and report on the quality of goods, food, the quality of the environment, or on the health of patients. These results form the basis for decisions on the use of goods, or on measures to be taken for the protection of the environment or the treatment of disease. In this respect, the accuracy of analytical measurements is a prerequisite for sound decision-making. Various means to achieve accuracy are available and are applied in good laboratories in the framework of quality systems. They include the existence of a quality assurance manual, training of personnel, a good managerial structure within the laboratory, the use of validated methods, the application of statistical control principles (e.g., control charts), external quality control measures (e.g., participation in interlaboratory studies leading to confrontation with results of other laboratories), and the use of certified reference materials (CRMs). In the limited scope of this article, attention will be mainly focused on the role of reference materials (RMs) and CRMs in chemical analysis. These materials are necessary for various of the above-mentioned requirements – method validation (CRMs), internal quality control using control charts (RMs), participation in interlaboratory studies (RMs), etc. – and thus represent one of the most necessary tools for achieving good analytical quality control.

### Definitions

The following definitions of ISO (International Organization for Standardization) are relevant:

- **RM:** a material or substance, one or more properties of which are sufficiently well established to be used for the calibration of an apparatus, the assessment of the measurement method, or assigning values to materials.
- **CRM:** a reference material, one or more of whose property values are certified by a technically valid procedure, accompanied by or traceable to a certificate or other documentation that is issued by a certifying body.

CRMs respond to the definition of ‘calibrants’. However, it should be noted that ‘matrix CRMs’ are generally reserved to the validation of analytical methods and not to their calibration (with the exception of certain techniques as discussed below). Another category involves the primary reference materials (PRMs), which are certified on the basis of primary methods. According to the above definitions, four main categories of reference materials may be distinguished:

- *Pure substances or solutions* used for calibration purposes and/or the identification of substances, or aimed at testing part or the totality of an analytical procedure. They are characterized by the establishment, in mass fractions, of (i) the remaining impurities in the purified substance and/or (ii) its isotopic composition.
- *Materials with known composition* representing materials used for the calibration of certain types of instruments, e.g., certified calibrating solutions or mixtures of gases or solids prepared on a gravimetric basis.
- *Matrix RMs* representing, as much as possible, the matrix to be analyzed and being characterized or certified for given parameters (e.g., concentrations of chemical substances). They consist of samples with a natural matrix, unknown or partially known, which are primarily used for the verification of a measurement process (CRMs) and/or the evaluation of laboratory performance in the framework of interlaboratory testing schemes (RMs). These materials may be used for the calibration of certain types of measuring instruments (i.e., techniques that require calibration with a material similar to the matrix analyzed, e.g., X-ray fluorescence (XRF)).
- *Operationally defined RMs* in which the assigned or certified values are directly linked to the applied method. These materials involve parameters that are associated with a certain properties, e.g., leachable or mobilizable fractions of elements from soils and/or bioavailable fractions, or an activity (e.g., enzymatic activity). The values are defined by the applied method following a very strict analytical protocol (e.g., a standardized method).

### Role of RMs

#### Use in Interlaboratory Testing Schemes

One of the well-known uses of RMs concerns the assessment of the performance of analytical



methods or laboratories in the framework of interlaboratory studies for research (development of new methods), standardization (development of standard methods), or laboratory evaluation (proficiency testing) purposes. RMs representative of the problems to be studied are distributed to participating laboratories by a central organization. Participants are requested to perform a given number of replicate determinations of one or several specified analytes and to report their results to the organizer of the study. The results may be discussed in technical meetings, where the sources of error are identified, and may later be removed in the laboratory. By comparing different techniques with different analytical steps (e.g., different pretreatments, separation, detection, calibration modes), possible bias due to a particular technique or errors of manipulation by a laboratory may be detected. Similarly, a laboratory's performance with a given technique (e.g., a standard method) may be evaluated by comparing the results obtained by this laboratory with those of other laboratories using the same technique. It is crucial that the RMs used in such studies are homogeneous and stable.

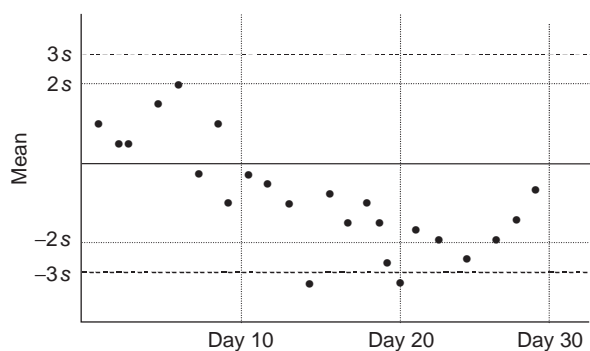
Robust statistics are increasingly used to evaluate the performance of laboratories and/or methods in the framework of interlaboratory proficiency testing schemes. One classical approach is based on the use of  $z$  scores, calculated according to the following formula:

$$\frac{\bar{X}_{\text{measured}} - \bar{X}_{\text{ref}}}{s/\sqrt{n}}$$

where  $\bar{X}_{\text{measured}}$  is the result of the analyses of the RM(s),  $\bar{X}_{\text{ref}}$  the assigned value of the RM (or the certified value in the case of a CRM),  $s$  the standard deviation of the measurements, and  $n$  the number of replicate analyses. The  $z$  scores that are below 2 are considered to be acceptable (method/laboratory of good performance), scores between 2 and 3 correspond to critical performance (e.g., insufficient precision/quality control), and scores above 3 are not acceptable.

### Use in Quality Control Schemes

RMs may be used for the verification of the long-term reproducibility of a method by setting up control charts. A control chart is a graphical representation of how results of RM analyses vary in time; it is used to detect possible systematic fluctuations (e.g., drift) in a method. The current practice is that one RM should be analyzed with 10–20 unknown samples and the results plotted on a



**Figure 1** Example of a Shewhart control chart. The warning limit corresponds to  $\pm 2s$ , and the action line is set at  $\pm 3s$ .

graph over time. In order to detect variations that are solely due to the method, the RM used should pose the same or similar problems to the analytical chemist as the unknown samples (analytical similarity), and their composition should be homogeneous and stable in time. One commonly used chart is the Shewhart control chart (Figure 1), which fixes limits beyond which a fluctuation has a 95% or 99% probability of being of systematic origin (designated as ‘warning’ and ‘action’, respectively). The results of a method are considered to be out of control if (1) the upper or lower control limit is exceeded (‘action’), (2) the same ‘warning’ line is exceeded twice in succession, and (3) more than ten successive measurements are on the same side of the target value line (‘mean’ line representing the reference value attributed to the RM).

Other charts that are also used are the cusum (cumulative sum) charts, in which the sum of the differences between the values found experimentally and the reference value (of the RM) are plotted in time. This latter type of chart allows for a more rapid detection of drifts.

## Role of Certified RMs

### Achievement of Accuracy

When developing a new analytical method or apparatus, and after having evaluated all its critical points, the analyst has to prove the accuracy of the results obtained. Usually the results will be compared with those obtained with a classical method, which implies that the latter is fully under control in the laboratory. However, some critical factors such as the influence of laboratory contamination cannot be resolved in the laboratory itself and have to be investigated externally (through participation in interlaboratory studies; as mentioned earlier). This



allows the comparison of laboratory results with those of other, preferably outstanding, laboratories. The use of CRMs allows a similar comparison and thus represents one of the ways to assess accuracy, but whenever necessary and instantaneously. CRMs may also be used to test a standardized method when it is applied for the first time in a laboratory.

In most chemical analyses, the measurement process contains steps where the sample is physically destroyed (e.g., acid digestion, fusion, dry-ashing) or the analyte to be determined is extracted from the matrix (e.g., using organic solvents). To ensure accuracy, it is necessary to demonstrate that no losses or contamination occurs during such sample treatment. The entire analytical procedure may be verified by using a CRM with a matrix similar to the unknown sample. Disagreement between the certified value and the value determined by the laboratory indicates an error, or errors, in the analytical procedure. The user may assess his/her laboratory bias from the difference between the certified value ( $\mu$ ) and the mean value of replicate measurements ( $\bar{X}$ ):  $\bar{X} - \mu$ . The criterion for acceptance is given in ISO Guide 33 (1988) as follows:

$$-a_2 - 2\sigma_L < \bar{X} - \mu < a_1 + 2\sigma_L$$

where  $a_1$  and  $a_2$  are adjustment values chosen by the user according to economic or technical limitations or stipulations, and  $\sigma_L$  is the long-term within-laboratory standard deviation of the user's method.

The use of CRMs for validation purposes is, however, not limited to the above intralaboratory verification of trueness (checking the absence of significant systematic errors). They also enable the user to estimate the precision of a method (repeatability and reproducibility), which should actually represent one of the first steps of the method validation. In this respect, the evaluation will have to take into account specific characteristics of the CRM, in particular, possible sources of uncertainties linked to the material heterogeneity which should in principle be considered for the calculation of the uncertainty of the certified values.

RMs that are certified according to the rules laid down in ISO Guide 35 (1985) allow the user's results to be traced back to those of the international scientific community. Additionally, they enable the user to assess his/her performance at any desired moment. In this respect, examples of the use of CRMs are described in the literature and are increasingly discussed among the analytical community, which highlight the need to clarify their advantages and limitations through the production of user-friendly guidelines.

## Calibration

In most analytical procedures, calibration is carried out by means of a calibration curve using compound(s) prepared with chemicals of an appropriate purity and verified stoichiometry. Matrix effects must often be taken into account and, consequently, the calibration solutions should be matrix-matched. CRMs of pure compounds may be used for calibration. However, matrix CRMs should in principle not be used for the purpose of calibration unless no other suitable calibrants are available, with the exception of those methods (e.g., spark source mass spectrometry, wavelength-dispersive XRF, etc.) that require calibration with CRMs of a similar, fully characterized matrix (e.g., metal alloys, cements). For such methods, accuracy can only be achieved when certified RMs are used for the calibration.

## Other Uses of CRMs

CRMs can also serve the purpose of demonstrating the equivalence of methods, enabling laboratories to follow the development of new analytical instrumentation, in particular when the certification has been based on the use of different analytical techniques. The analyst may compare the performance of his/her method with those from other laboratories and/or other techniques without the need for participating in interlaboratory studies. This sort of comparison is essential when harmonization of measurements is required at an international level.

CRMs are products of very high added value. Their production and certification are very costly and, therefore, these materials should not be used for routine quality control or external quality assurance (interlaboratory studies). In order to fulfil the needs related to these uses, noncertified RMs may be prepared, linking them to one or several CRMs as a means of evaluating the accuracy of the values assigned to the so-called 'secondary' materials. This approach is straightforward in the case of calibrants (pure compounds or calibrating solutions), but is more difficult in the case of matrix CRMs owing to the likely matrix differences.

## Frequency of Use

There are no absolute rules regarding the frequency at which CRMs should be used. The verification of the accuracy of a method based on CRM analyses will actually depend on the number of analyses carried out by the laboratory, method robustness, economic considerations, legal requirements, etc. Some laboratories prefer to analyze CRMs at regular intervals (e.g., once per month) rather than performing

daily verifications through control charts, which is certainly justified for analyses that are only performed on rare occasions (involving instruments and personnel that are usually engaged in other tasks). Others base their quality control on routine measurements of (noncertified) RMs, analyzing CRMs only once a year.

CRMs should, in principle, be systematically used if important modifications are brought to a routinely applied method (e.g., instrument change, new analyst, new batch of consumables, etc.). Finally, the use of CRMs may be required by customers, e.g., when analytical data have to be used as a decision-making tool in the framework of regulations.

## Other Types of Materials

Production requirements of matrix RMs correspond to the best compromise with respect to the representativity of the materials, but they of course cannot reflect the exact reality. In some cases, however, there is a need to prepare and store samples in such a way that they closely represent specific situations. This is the case for specimen banking purposes, e.g., for environmental or medical analyses, in which samples are collected at regular intervals and stored in liquid nitrogen to ensure long-term stability. These materials are not considered as RMs *sensu stricto* but they nevertheless act as 'references' for specific programs (e.g., Environmental Specimen Banking), with the aim of following contamination trends. Their preparation follows rules similar to those used for CRMs.

## Traceability of RMs

RMs represent an important tool for demonstrating the traceability of analytical measurements to given references. One should keep in mind that the traceability of chemical analyses is often more difficult to demonstrate than for physical measurements. This is due to major differences in the measurement processes (e.g., matrix influence on chemical analyses, various analytical problems linked to the analytes and the methods used, need for sample pretreatment, etc.). Contrary to physical measurements, the calibrants and CRMs used for chemical analyses are not only used for instrument calibration but also for a variety of other purposes (e.g., method validation). In terms of traceability, the theory implies that the certified values of a substance in a CRM should be traceable to the amount of the given substance expressed according to the relevant SI unit, i.e., the mol. Since there is no 'reference mol', this traceability can be established only in relation to the mass SI unit, i.e., the kg.

The degree at which CRMs will constitute themselves a way to achieve traceability of chemical measurements will actually depend upon the quality of the link with the SI unit. In this context, the demonstration of traceability of values of matrix materials to SI is often difficult to establish owing to the variety of matrices and parameters, and of the factors that influence their analysis. Traceability in chemical analysis links may be considered as follows: SI units (kg, mol) – International calibrants (kg) – atomic masses – pure compounds – primary methods – PRMs – validated methods – matrix RMs – routine methods – (not certified) RMs. In this respect, a hierarchy of RMs could be established, following a classification established by Pan Xiu Rong in 1997. In this classification, the highest metrological quality would be achieved by PRMs developed by a national metrological laboratory (traceable to SI units and verified through international intercomparisons) and certified by a primary method, a second level would correspond to CRMs developed by either a national metrological laboratory or a specialized organization (fulfilling the conditions of ISO Guide 30) and certified using reference methods through the comparison of different methods or a combination of the two approaches, and a third level would concern working (or laboratory) reference materials developed by an accredited organization (also fulfilling the conditions of ISO Guide 30) for which reference values are established by one or more validated methods and which are accompanied by a description of the traceability and an estimate of the uncertainty. Additional details on certification are described elsewhere in this encyclopedia.

## Examples of RMs

The variety of RMs is potentially very high, considering the unlimited combination of matrix types/parameters, and the purpose of this section is not to give specific examples (which may be obtained through information sources). Let us note that 'classical' solid materials used for environmental, food, and health-related analyses (e.g., sediment, soils, molluscs, leaves, plants, vegetables, biological tissues, dairy products, etc.) are often available in the form of dried powders (either oven- or freeze-dried) stored in brown glass bottles (units containing from one to hundreds of grams). Liquid materials may involve natural waters (basically all types of waters) stored in glass ampoules, glass or polyethylene bottles, extracts (e.g., fish extracts), natural oils, biological fluids, etc. Atmospheric CRMs also exist in the form of pure gases, ashes, dusts, etc. Finally, industrial CRMs may involve alloys, glass, ceramics, paints, etc. and also pure substances used for the quality control

of comparative methods (e.g., XRF spectrometry). Assigned or certified values to the various matrices represent a huge number of parameters, e.g., inorganic substances (trace elements, nutrients, etc.), organic compounds (PAH, PCB, pesticides, dioxins, etc.), chemical forms of elements (elements in different oxidation states, organometallic compounds), extractable forms of elements (determined on the basis of operationally defined extraction procedures), etc. and also microbiological parameters (e.g., *salmonella*, *listeria*, etc.). This list is far from exhaustive, and the reader is referred to the specialized literature.

**See also: Quality Assurance:** Quality Control; Instrument Calibration; Interlaboratory Studies; Production of Reference Materials; Accreditation; Clinical Applications; Water Applications.

### Further Reading

- Barwick V, Burke S, Lawn R, Roper P, and Walker R (2001) *Applications of Reference Materials in Analytical Chemistry*. Cambridge, UK: Royal Society of Chemistry.
- Hartley TH (1990) *Computerized Quality Control: Programs for the Analytical Laboratory*, 2nd edn. Chichester: Ellis Horwood.
- ISO (1985) *Certification of Reference Materials – General and Statistical Principles*, ISO/IEC Guide 35-1985. Geneva: International Organisation for Standardisation.
- ISO (1988) *Use of Certified Reference Materials*, ISO/IEC Guide 33-1988. Geneva: International Organisation for Standardisation.
- ISO (1991) *Terms and Definitions Used in Connection with Reference Materials*, ISO/IEC Guide 30-1991. Geneva: International Organisation for Standardisation.
- Pan Xiu Rong (1997) Hierarchy of reference materials certified for chemical composition. *Metrologia* 34: 35–39.
- Quevauviller Ph (2002) *Matériaux de référence pour l'environnement*. Tec&Doc, Paris.
- Quevauviller Ph and Maier EA (1999) *Interlaboratory Studies and Certified Reference Materials for Environmental Analyses*. Amsterdam: Elsevier.
- Stoeppeler M, Wolf WR, and Jenks PJ (eds.) (2001) *Reference Materials for Chemical Analysis – Certification, Availability and Proper Usage*. Weinheim: Wiley-VCH.
- Walker R and Lumley I (1999) *Trends in Analytical Chemistry* 18: 594–616.

## Production of Reference Materials

**P Quevauviller**, Drève de Nivelles, Brussels, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

### Production and Requirements

General aspects of production and requirements are common to all types of reference materials (RMs). The main differences relate to the procedure of certification or the establishment of assigned values. In the first case (certification), the material is associated to a guarantee given by the producer regarding the certified values (homogeneity, stability, traceability, etc.), the amount produced, and the existence of an infrastructure able to ensure a proper distribution, which is not often the case for non-certified RMs generally produced at a limited scale. The production infrastructure thus depends upon the type of RM and its use. It may vary from a simple laboratory to a semi-industrial structure. The quality of the produced RMs will directly rely on the resources that are made available for their preparation (personnel, time, equipment). In this context, quality is always rewarding but it also has a cost that the producer will have to consider according to various criteria, a main one being the final use of the certified reference material (CRM). Let us recall that the general quality of an analytical work strongly depends upon the available quality control tools.

Hence, it seems obvious that the evaluation of analytical results should not only consider results obtained with CRMs but also take into consideration the way they have been prepared and used.

Driving rules for RM production are detailed in the International Standards Organization (ISO) Guide 34, which summarizes requirements regarding the production, installation, collaborators and their competence, sample handling and storage, production planning, and aspects directly related to the material preparation (including homogeneity and stability testing). One should note that the production of CRMs represents a heavy responsibility since the materials are used by the laboratories, which will check their performance on the basis of their results compared with certified values and may change their methods or way of working depending on the results obtained. This high level of responsibility goes hand-in-hand with an increasing pressure for the accreditation of certification organizations. Indeed, CRM users are increasingly demanding information on the quality of the production process. In this respect, the ISO Guide 34 gives clear instructions for the interpretation of ISO 17025 and ISO 9000 standards as regards RM production. In addition, the European Commission has issued a guide addressed to laboratories that are willing to develop RM certification projects.

of comparative methods (e.g., XRF spectrometry). Assigned or certified values to the various matrices represent a huge number of parameters, e.g., inorganic substances (trace elements, nutrients, etc.), organic compounds (PAH, PCB, pesticides, dioxins, etc.), chemical forms of elements (elements in different oxidation states, organometallic compounds), extractable forms of elements (determined on the basis of operationally defined extraction procedures), etc. and also microbiological parameters (e.g., *salmonella*, *listeria*, etc.). This list is far from exhaustive, and the reader is referred to the specialized literature.

**See also: Quality Assurance:** Quality Control; Instrument Calibration; Interlaboratory Studies; Production of Reference Materials; Accreditation; Clinical Applications; Water Applications.

### Further Reading

- Barwick V, Burke S, Lawn R, Roper P, and Walker R (2001) *Applications of Reference Materials in Analytical Chemistry*. Cambridge, UK: Royal Society of Chemistry.
- Hartley TH (1990) *Computerized Quality Control: Programs for the Analytical Laboratory*, 2nd edn. Chichester: Ellis Horwood.
- ISO (1985) *Certification of Reference Materials – General and Statistical Principles*, ISO/IEC Guide 35-1985. Geneva: International Organisation for Standardisation.
- ISO (1988) *Use of Certified Reference Materials*, ISO/IEC Guide 33-1988. Geneva: International Organisation for Standardisation.
- ISO (1991) *Terms and Definitions Used in Connection with Reference Materials*, ISO/IEC Guide 30-1991. Geneva: International Organisation for Standardisation.
- Pan Xiu Rong (1997) Hierarchy of reference materials certified for chemical composition. *Metrologia* 34: 35–39.
- Quevauviller Ph (2002) *Matériaux de référence pour l'environnement*. Tec&Doc, Paris.
- Quevauviller Ph and Maier EA (1999) *Interlaboratory Studies and Certified Reference Materials for Environmental Analyses*. Amsterdam: Elsevier.
- Stoeppeler M, Wolf WR, and Jenks PJ (eds.) (2001) *Reference Materials for Chemical Analysis – Certification, Availability and Proper Usage*. Weinheim: Wiley-VCH.
- Walker R and Lumley I (1999) *Trends in Analytical Chemistry* 18: 594–616.

## Production of Reference Materials

**P Quevauviller**, Drève de Nivelles, Brussels, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

### Production and Requirements

General aspects of production and requirements are common to all types of reference materials (RMs). The main differences relate to the procedure of certification or the establishment of assigned values. In the first case (certification), the material is associated to a guarantee given by the producer regarding the certified values (homogeneity, stability, traceability, etc.), the amount produced, and the existence of an infrastructure able to ensure a proper distribution, which is not often the case for non-certified RMs generally produced at a limited scale. The production infrastructure thus depends upon the type of RM and its use. It may vary from a simple laboratory to a semi-industrial structure. The quality of the produced RMs will directly rely on the resources that are made available for their preparation (personnel, time, equipment). In this context, quality is always rewarding but it also has a cost that the producer will have to consider according to various criteria, a main one being the final use of the certified reference material (CRM). Let us recall that the general quality of an analytical work strongly depends upon the available quality control tools.

Hence, it seems obvious that the evaluation of analytical results should not only consider results obtained with CRMs but also take into consideration the way they have been prepared and used.

Driving rules for RM production are detailed in the International Standards Organization (ISO) Guide 34, which summarizes requirements regarding the production, installation, collaborators and their competence, sample handling and storage, production planning, and aspects directly related to the material preparation (including homogeneity and stability testing). One should note that the production of CRMs represents a heavy responsibility since the materials are used by the laboratories, which will check their performance on the basis of their results compared with certified values and may change their methods or way of working depending on the results obtained. This high level of responsibility goes hand-in-hand with an increasing pressure for the accreditation of certification organizations. Indeed, CRM users are increasingly demanding information on the quality of the production process. In this respect, the ISO Guide 34 gives clear instructions for the interpretation of ISO 17025 and ISO 9000 standards as regards RM production. In addition, the European Commission has issued a guide addressed to laboratories that are willing to develop RM certification projects.



## Selection

An RM or a CRM has to fulfill a specific task and so should be selected with great care on the basis of a series of requirements. The main prerequisite is certainly that the CRM should be as representative of the real sample as possible. The most critical aspect of a method validation is indeed that the material used should be at least as difficult to analyze than samples routinely analyzed. This requirement of representativeness means, in most cases, that there should be the best possible similarity between RM and sample in:

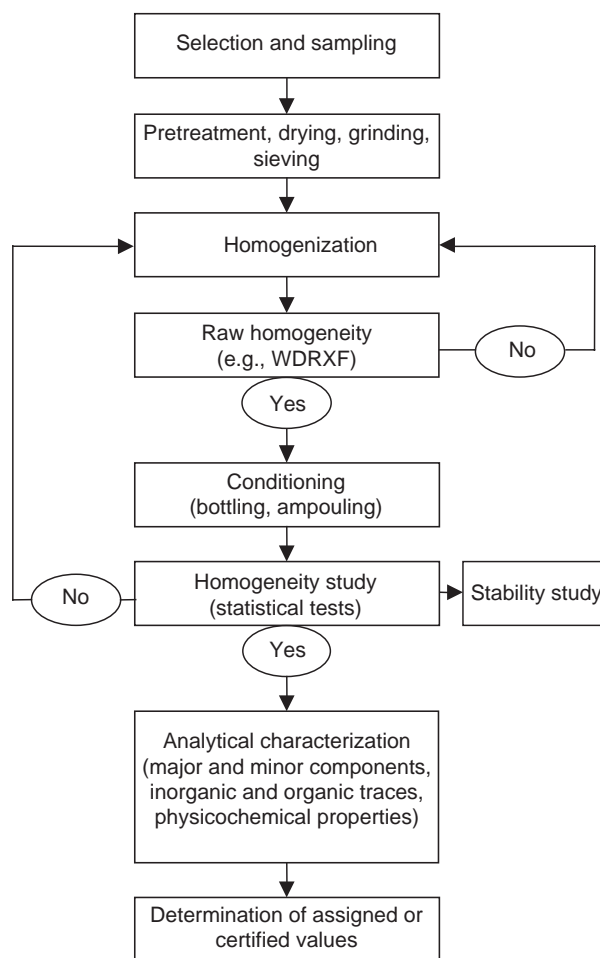
- matrix composition;
- measured substances and their level of concentrations;
- way of binding of the analytes;
- fingerprint pattern of possible interferences; and
- physical status of the material.

In selecting an RM, these factors should be carefully taken into consideration. For practical reasons, a close similarity cannot always be achieved. This actually depends on the availability of the samples and critical factors such as homogeneity and stability. For example, 'fresh' samples can hardly be stabilized and require specific treatment that will modify their physical status. Unstable chemical substances may also require specific stabilization procedures. Therefore, compromises have to be made by the producer and they have to be accepted by the user on the basis of clear justifications. In many instances, the RM preparation has to be adapted; this is discussed in the following sections.

## Preparation

An example of a planning flow chart for the preparation of an RM is shown in **Figure 1**.

The material to be used for the preparation of RMs or CRMs should be collected in a quantity sufficient to ensure an adequate and sufficiently lasting stock. The amount to be prepared depends on the analytical sample size, stability, shelf-life, and frequency of use. It may vary from 5 to 20 kg of solid material (e.g., for soils, sediments, biological tissues, etc.) or from 5 to 20 l (e.g., natural waters, calibrating solutions, etc.) for the preparation of RMs to be used in routine analysis, to up to 100 kg of solid material or several cubic meters of liquids when CRMs are to be prepared. In order to produce RMs and CRMs, laboratories need to be equipped to treat such large amounts of material without substantially changing the analytical methods. The treatment of 3–5 kg of raw material is already the limit for normal laboratory equipment and manual processing. For larger batches,



**Figure 1** Example of planning for an RM preparation. (Adapted from Muntau H (1079) *Proceedings of the First International Symposium on Production and Use of Reference Materials*, Berlin, pp. 185–218.)

and especially for larger volumes of material, it is necessary to scale up to half-industrial scale.

Typical operations required are crushing, grinding, sieving, filtering, and mixing and homogenization of the materials, which can only be performed in specialized laboratories or industries. Available tools for RM preparation are described in the literature.

## Stabilization

The usefulness of RMs and CRMs relies on the possibility offered to the laboratories to analyze them at any time, which require that their stability be guaranteed. The stabilization is, therefore, a crucial phase of the production. This required treatment should be such that it does not significantly affect the material representativeness. In this context, three types of phenomena may actually affect the

RM: physical effects, chemical reactions, and biological activity. Physical influences are generally due to effects of light or other radiations, temperature, and gravity. They can easily be avoided by adapting the storage and transport conditions. Chemical reactions may occur among substances present in the RM matrix, in particular in the case of 'fresh' materials. In the case of solid materials in a dried form, chemical reactions are either absent or are slowly processing. Storage at low temperature limits also the risk of chemical changes.

Microbial activities represent the instability factor that is the most difficult to control at the stage of RM preparation, in particular for food and environmental RMs. During treatment, it is, therefore, of paramount importance to eliminate or control the microbial activity either by killing the microorganisms or by stopping their activity for a given period. Various procedures are available to stop or limit the effects of microbial activity. RM may be heat-sterilized, which is the easiest and cheapest way, but this treatment can only be applied if the chemical substances to be analyzed/certified are not affected by very elevated temperatures (in this case, pasteurization may represent a compromise). Sterilization by gamma-irradiation is another possibility, but this process may also affect substances present in the matrix and hence the overall RM representativity. Finally, microbe destruction may be obtained by the addition of antiseptics, taking care that the added substance does not significantly affect the analytical procedure for the chemical substances to be analyzed/certified.

The stabilization of an RM has to be carefully planned, i.e., sufficiently well in advance in the overall production process to avoid further contamination or microbial growth during the subsequent steps. One should also note that the stabilization procedure may have limitations when it leads to changes that entail that the RM is no longer behaving like a 'natural' sample.

## **Treatment**

The material must be homogenized in such a way that subsamples taken in different units will be identical for the analyzed parameters (within a given confidence interval). For gases or liquids, homogenization is relatively straightforward, the only difficulty being when this operation is carried out at half-industrial scale (i.e., processing tons or cubic meters). Solid materials are more difficult to homogenize. They are generally composed of several phases, which are distributed according to various parameters such as particle size distribution and density. In principle, a narrow particle size

distribution is required to obtain similar sample intakes, which also depends on the size and density of the particles. At the planning stage of an RM preparation, it is important to define what should be the most appropriate grain sizes for a given material in order to select an adapted grinding (and if appropriate sieving) procedure. This may be particularly critical for heterogeneous materials, e.g., soils and sediments. One should also note that a very fine grinding process may lead to changes that may affect the analytical process, e.g., enhancing the extractability of some substances (because of a larger contact surface), generating static electricity for materials with very small particles (less than 1  $\mu\text{m}$ ), etc. Examples of grinding procedures and devices for RM homogenization are described in the literature.

## **Conditioning and Storage**

The vial (e.g., bottle or ampoule) in which the material will be conditioned, as well as storage conditions, should be chosen in such a way that no alteration will affect the material during and after its conditioning. The vial should be inert with respect to the RM matrix and the substances of interest so as to avoid any interactions that might generate either contamination or losses. The vial must also be such that the RM will be protected from external influences (e.g., light, temperature, humidity).

Great experience exists within RM-producing organizations, which may help in choosing the most appropriate vial and storage condition for a given matrix RM and type of substances. Optimal storage temperatures are generally determined on the basis of extensive stability experiments. As a final remark, let us note that conditioning for the transport of CRMs also has to be carefully planned in terms of both the protection of the materials and safety reasons (i.e., in the case of RM containing hazardous substances).

## **Homogeneity Study**

A subsample of an RM can be used in chemistry only once as it is generally destroyed during the analysis. Therefore, the amount of material in a bottle or ampoule should be sufficient to perform one or sometimes several determinations. The more subsamples taken, the lesser the chance that the bulk is still the same. A verification of the homogeneity must be performed to ensure that the contents within a vial and from one vial to another are the same (within- and between-vial homogeneity). The homogeneity of the material can be verified by performing, e.g., 10 replicate determinations of the element of concern within one vial (within-vial homogeneity) and by one



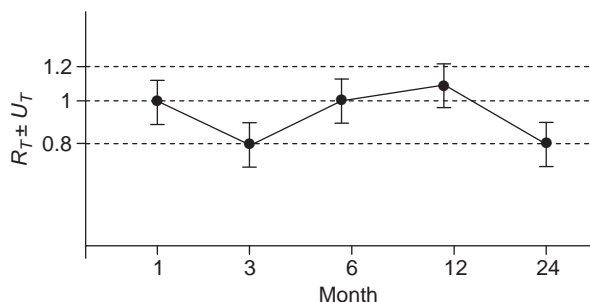
determination in vials set aside at regular intervals during the conditioning (2–5% of the vials can be used for this between-vial homogeneity). The coefficient of variation (CV) obtained may be compared with the CV of the final step of the method (assessed by, e.g., five replicate determinations of one digest or extract solution). Such comparison does not take into account that the extraction or digestion procedures contribute to the random uncertainty of a result; they, therefore, lead to conservative conclusions.

The minimum sample size for which the homogeneity is sufficient should be verified and provided by the producer. Below this sample size, the uncertainty caused by inhomogeneity contributes significantly to the uncertainty of the reference (or certified) values (i.e., a maximum of 30% of the total uncertainty). An additional problem is caused by segregation during transport and long-term storage, which requires that special care has to be taken for the rehomogenization of the material before taking a test portion.

## Stability Study

The material composition and the values of the RM parameters have to remain unchanged over its period of use, and this has to be demonstrated by the producer. The study of an RM's long-term stability is, of course, closely linked to the selected storage conditions. Such a study will be carried out following two basic principles: either following the possible losses of chemical substances in the sample, or detecting the apparition of known degradation products. In both cases, the verifications are based on analytical measurements, which should have a good reproducibility.

An approach to evaluate various storage conditions has been successfully used for RM certification within the last decade. It consists in studying the stability of the material stored at different temperatures (e.g., +20°C and +40°C) and to detect a possible evolution by comparing the results obtained over a given period of time (e.g., 12–24 months) with samples stored at –20°C, a temperature at which changes are assumed to be very small or negligible. The results obtained at +20°C may lead to an assessment of the sample stability at ambient laboratory temperature, whereas the results obtained at +40°C are used to assess the worst-case conditions (e.g., during transport) and allow the evaluation of the stability over longer periods of time. It is indeed assumed that a sample stable at +40°C over a period of, e.g., 1 year may be stable at +20°C for a longer period. This assumption does not hold in cases of spoilage by certain bacteria or molds having optimum temperatures for their metabolism



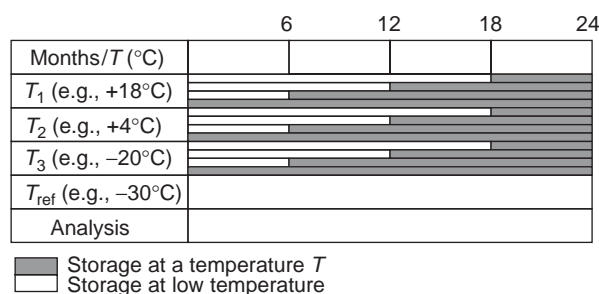
**Figure 2** Example of graph for the interpretation of results of a stability study. In this approach,  $R_T$  and  $U_T$  are the ratios (and their uncertainties) of mean results obtained at a given temperature (e.g., +20°C or +40°C) versus mean results obtained at a reference temperature (e.g., –20°C). In the ideal situation, the ratio should be 1. In practice, ratios may be comprised between  $[R_T - U_T]$  and  $[R_T + U_T]$ . In the example,  $U_T$  is equal to 0.2.

at +20–35°C, which is generally avoided thanks to a reduction of the water content of samples to less than 5% by mass and to the necessary care taken at the stabilization stage. In the framework of such experiments, results obtained at +20°C or +40°C are normalized to results obtained at –20°C, which enables to smoothen analytical variations over time. An example is shown in **Figure 2**.

The above approach is actually inspired by possible extrapolations using the Arrhenius law, which describes the activation rate in function of the temperature. Temperature is, indeed, an important factor for studying chemical reaction kinetics. The above assumption could then be backed up by such a law, e.g., materials tested at +40°C over 1 year could be assumed to be stable over 3–6 years at +20°C. In practice, however, these extrapolations are faced with difficulties linked to the complexity of RM matrices in which the kinetics of chemical reactions does not necessarily follow the Arrhenius law.

A more recent approach is based on isochronous studies in which samples are stored at different temperatures at different time intervals, performing all the analyses at the end of the study. This enables to plan the measurements well in advance and hence limits risks linked to an insufficient reproducibility of the method (which may be faced by the first approach). The drawback is that results are only available at the end of the study, which may generate a delay of the stability assessment and thus of the overall certification process. A flow chart illustrating this approach is shown in **Figure 3**.

A recent development in CRM production is the attempt to establish the maximum period over which the material may be used, which is, however, difficult to set out experimentally. Establishing such time windows is actually meant to fix the time frame



**Figure 3** Typical scheme of stability study based on isochronous measurements. (Adapted from Lamberty A, Schimmel H, and Pauwels J (1998) The study of the stability of reference materials by isochronous measurements. *Fresenius Journal of Analytical Chemistry* 360: 359–361.)

within which the producer's responsibility is legally engaged. It is not necessarily linked to the maximum period of stability of the materials.

Stability checks should be carried out at regular intervals (e.g., every 3 years) after certification, and if possible over the entire period of use of the CRM.

## Assigning or Certifying Values

Several technically valid approaches may be used to either assign reference values to or certify an RM. Table 1 describes eight modes that are followed by the National Institute of Standards and Technology (NIST), three of them leading to certified values and five used to establish 'indicative' (or assigned) values.

The certification of the purity of substances for calibration purposes represents a case apart in comparison to the approaches followed for certifying values in a matrix RM. This is usually carried out in specialized institutes (e.g., chemical companies). The basic principle is that the certified substances should be in the purest possible form, the major difficulty being that it is hardly possible to detect and quantify all potential impurities. The purity is controlled during the preparation and purification and at the end of the process. When a sufficient purity is attained, it is, therefore, possible to carry out certification analyses in order to quantify the purity and its uncertainty.

With regard to matrix RMs, the certification is less straightforward and the procedure has to follow strict rules described in the ISO Guide 35. Contrary to pure substances, matrix RMs cannot be certified on the basis of direct determinations of masses of constituents, and analytical methods have to be used that are often based on a transformation or a total destruction of the matrix. Three major approaches are described below.

**Table 1** Approaches for certifying/assigning values to a reference material

	C	A	I
1. Certification based on a primary method with confirmation by an independent method	X		
2. Certification based on two critically evaluated independent methods	X	X	
3. Certification/assignment of values based on a validated method and different methods used by external laboratories	X		
4. Assignment of values based on two or more laboratories using different methods		X	X
5. Assignment of values based on an operationally defined method		X	X
6. Assignment of values based on a single method (not responding to certification criteria)		X	X
7. Assignment of values based on external laboratories using a single method		X	X
8. Assignment of values based on data selected throughout interlaboratory studies		X	X

C means certified values, A refers to assigned values (not certified), and I corresponds to data given for information. (Adapted from Gills TE (1999) NIST standard reference materials for measurement assurance – practices, issues and perspectives. In: Fajgelj A and Parkany M (eds.) *The Use of Matrix Reference Materials in Environmental Analytical Processes*, Cambridge: The Royal Society of Chemistry, pp. 57–64.)

## Using a Primary Method in One Single Organization

The certification in this case is based on analyses carried out in one specialized laboratory, using a primary method operated by at least two independent analysts. Primary methods enable one to trace results back to SI units and hence they represent a high added value for certifying matrix RMs. In this respect, isotope dilution mass spectrometry is often used by major worldwide producers for the certification of trace elements. It should be noted that a RM certified on the basis of one single primary method may not be representative of results that will be used by routine methods. Furthermore, experience has shown that it is hardly possible to demonstrate the total absence of systematic errors for analyses carried out in a single laboratory, and producers following this approach are often checking their results through verifications with independent laboratories. It should also be noted that primary methods are not yet fully developed in analytical sectors such as the determination of trace organic or organometallic compounds, which require other approaches for assigning/certifying values (through interlaboratory studies).

### **Using Independent Reference Methods in One Single Organization**

The approach is based on the use of at least two reference methods with different chemical principles (possibly including a primary method), operated in the same laboratory by at least two independent analysts. 'Reference methods' refer to methods of which the accuracy has been demonstrated by an appropriate validation. The principle of this approach is that an agreement between two or more independent methods will limit the risk that systematic errors are left undetected. The basic question relates to the real independence of the analysts working in the same laboratory, and it is advisable to involve other laboratories, which brings us to the next case.

### **Independent Methods Applied by Several Independent Laboratories**

At least two methods (based on different chemical principles) are used by at least two independent high-quality laboratories. The most efficient approach is to develop a network of laboratories working according to clear instructions regarding certification analyses. All the methods have to be well characterized and validated. If possible, the inclusion of a primary method represents an added value with respect to the trueness of the final results. The selection of laboratories is based on their expertise in the analytical field concerned and their experience in participating in interlaboratory studies. Strict rules have to be followed for the organization of such trials, which are described in the literature. The wide variety of matrices in some fields (e.g., environmental, food analyses) generally justify such an approach, which involves several independent methods operated in expert laboratories, thus limiting the risks of systematic errors. Technical discussions among the laboratories represent an additional element that may help to improve the quality of the results. This concept relies on the assumption that there are a sufficient number of laboratories able to carry out the analytical work at a high analytical quality standard and that the differences between individual results (both within and among laboratories) are technically and statistically acceptable.

### **Voluntary Analyses through Interlaboratory Studies**

This approach corresponds to interlaboratory studies organized on a voluntary basis for specific purposes (e.g., proficiency testing). In this case, the selected laboratories and methods may be very diverse and not necessarily prepared of submitting results aimed at establishing assigned values to a RM. The

evaluation is often based on robust statistics and, while it is certainly useful for testing the performance of laboratories and/or methods, it is not recommended for the establishment of certified (or even assigned) values to a RM. The statistical evaluation of a population of results originating from various (uncontrolled) laboratories will, indeed, not replace a technical evaluation of results prior to the establishment of values to a RM. This approach is, however, followed by some organizations and has been considered in some instances as an alternative to the above described procedures.

## **Evaluation of Results**

The establishment of assigned or certified values relies on a thorough evaluation of the analytical results provided by the laboratories involved in the exercise. In this respect, one should distinguish between the technical and the statistical evaluations, which are discussed below.

### **Technical Evaluation**

The reliability of any assigned or certified values depends upon an appropriate technical evaluation, which has to be conducted prior to any statistics (which enable to study the distribution of a population of data but not to explain possible differences among them). In this respect, it is essential that the analytical method(s) used be examined in detail to verify that no systematic errors have been left undetected. In the case of certification based on an interlaboratory study, technical meetings with the participating laboratories represent a good means to scrutinize the results through discussions on the validity of the different analytical steps (e.g., extraction, separation, detection, calibration). Such technical evaluation is of particular importance for the purpose of interlaboratory-based certification, which is not solely dealing with an estimate of 'consensus values' arising from uncontrolled interlaboratory studies.

### **Statistical Evaluation**

Once a technical evaluation has been carried out, a statistical analysis is needed to confirm the absence of outlying means, examine the precision of the data, and evaluate the certified (or assigned) values and their uncertainties. A wide range of statistical tests exist in this respect; they are based on the calculation of the mean values obtained by different laboratories and/or methods and a study of their distribution, and an analysis of the variances obtained. Examples are the Dixon (Nalimov) test for the evaluation of the

distribution of means, the Cochran test to detect outlying variances, etc. As stressed above, major differences in the results should, in principle, be clarified at the stage of the technical evaluation. The basic rule is that no 'cosmetic' statistics should be used to make abstraction of data of questionable quality.

### Calculation of the Uncertainty

An important aspect of the statistical evaluation is the calculation of the uncertainty of the assigned/certified values. Recent developments through the guide for the uncertainty of measurements (GUM) are now being applied by major CRM producers to estimate the total uncertainty budget, taking account of all sources of uncertainties arising from the various production steps:

$$U_{\text{CRM}} = [u_{\text{char}}^2 + u_{\text{hom}}^2 + u_{\text{its}}^2 + u_{\text{sts}}^2]^{1/2}$$

where  $u_{\text{char}}$  is linked to the uncertainty of the material characterization (confidence interval of the certified values),  $u_{\text{hom}}$  corresponds to the variability linked to the (between-vial) homogeneity, and  $u_{\text{its}}$  and  $u_{\text{sts}}$  are related to the long-term (possible degradation during storage) and short-term (linked to possible degradation during transport) stabilities. A correct estimate of all uncertainties arising at the various production steps is not always possible, but it is recommended to examine all the sources of uncertainties, be they significant or not, to demonstrate to the users that they have been considered, whatever their order of magnitude.

The certified values will be calculated as the arithmetic means obtained by the various laboratories/methods. They will be provided with either a 95% confidence interval of the mean of laboratory means or the total uncertainty calculated according to the GUM as outlined above.

### Certificate

A certificate of analysis is mandatory for all CRMs (which is not the case for simple RMs). It contains information on the producer, a description of the material (characterization and preparation) and its use, storage recommendations, the certified values and their uncertainties, possible indicative values, identification of participating laboratories, and a legal notice with the signature of the certifying body. In some instances, a certificate may be complemented by a detailed technical report (also named certification report), which describes extensively all the production steps of the RM and provides all the individual data used for its certification along with a summary of their technical and statistical evaluations.

### Producers

The ISO Council on Reference Materials (REMCO) has identified more than 170 RM producers worldwide for the various fields of analyses. Two main bodies – the National Institute of Standards and Technology (NIST, USA) and the Institute for Reference Materials and Measurements (IRMM, European Commission, Belgium), successor of BCR (which is now a trade mark for materials produced by IRMM) – cover several fields and ensure long-term availability of the CRMs due to the large batches of materials produced. Other producers specialize in a particular field of interest, e.g., the National Research Council of Canada (NRCC, Canada), the National Institute for Environmental Studies (NIES, Japan), the IAEA Marine Environment Laboratory (Monaco), etc., specialize in marine analysis. The International Atomic Energy Agency (IAEA, Austria) mainly provides materials for nuclear measurements but also supplies RMs for non-nuclear analyses.

### Sources of Information

The REMCO publishes a directory for RMs and has gathered information on more than 1500 CRMs that were available or under preparation in 2002. In addition, the International Union of Pure and Applied Chemistry (IUPAC) issues a catalog of available CRMs. Some additional compilations of existing CRMs exist in more specialized fields, e.g., marine monitoring (publications of the National Oceanographic and Atmospheric Administration, USA).

An inventory established by the IAEA in collaboration with the UNEP program of the United Nations, and collecting information on 650 matrix RMs, may be consulted at: <http://www-naweb.iaea.org>

The major source of information on RMs, however, is the COMAR Data Bank, which is a joint enterprise between the Bundesanstalt für Materialforschung- und Prüfung (BAM, Germany), the Laboratoire National d'Essais (LNE, France), and the National Physical Laboratory (NPL, United Kingdom).

Finally, worth mentioning is the development of an 'European Virtual Institute on Reference Materials' (VI-RM), which is an initiative funded by the European Commission in which all the key players in RM production, certification, and distribution are collaborating for optimizing exchanges of knowledge and good practices.

**See also: Quality Assurance:** Quality Control; Primary Standards; Interlaboratory Studies; Reference Materials; Traceability.



## Further Reading

ISO (1985) *Certification of Reference Materials – General and Statistical Principles*. ISO/IEC Guide 35-1985. Geneva: International Organization for Standardization.

ISO (1996) *Quality System Guidelines for the Production of Reference Materials*. ISO/IEC Guide 34-1996. Geneva: International Organization for Standardization.

Quevauviller Ph and Maier EA (1999) *Interlaboratory Studies and Certified Reference Materials for Environmental Analyses*. Amsterdam: Elsevier, 558pp.

Stoeppler M, Wolf WR, and Jenks PJ (eds.) (2001) *Reference Materials for Chemical Analysis – Certification, Availability and Proper Usage*. Weinheim: Wiley-VCH, 297pp.

Walker RF (1997) Quality system requirements for the production of reference materials. *Trends in Analytical Chemistry* 16: 9–16.

## Method Validation

**D Brynn Hibbert**, University of New South Wales, Sydney, NSW, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

It is almost a truism that the use of an analytical method implies that it will, when implemented properly, deliver a useful result. However, someone must have first described the method but is the user of the method sure that, when it was developed, enough checks were made to demonstrate its validity? This leads to a further question of what appropriate checks may be, and how the validity of the method can be demonstrated.

This article catalogs and describes the different aspects of a method that must be studied to ensure that it is, indeed, a valid method. The relationship between systematic measurements made during development and validation, the so-called development–validation cycle will be discussed. When using a previously validated method, there is still a necessity to assert that the method performs as expected in the user's laboratory. This is known as verification.

### What is Method Validation?

Definitions found in the literature emphasize different aspects of validation. For example:

- The provision of documentary evidence that a system fulfills its predefined specification...involves the evaluation of the fitness of analytical methods for their purpose.
- The process of proving that an analytical method is acceptable for its intended purpose.
- Checks...to ensure that the performance characteristics of the method are understood and

demonstrate that the method is scientifically sound under the conditions under which it is to be applied.

- Analytical measurements should be made using methods and equipment that have been tested to ensure they are fit for purpose.
- Confirmation by examination and the provision of objective evidence that the particular requirements for a specific intended use are fulfilled.

Many of these definitions use the concept of fit for purpose. The importance of this is the implication that there is no absolute worth of a method. If the results are sufficient for the requirements of the client, and the client genuinely receives useful results, the method is fit for purpose. How well a client views a method is tied to its intended use and so the process of method validation must be ultimately guided by the requirements of intended use. This appears to imply that every laboratory will need to embark on extensive experimentation with each new job. Sensibly, it is recognized that a particular method is used in a broadly similar manner across the world, and so if an international agency such as American Society for Testing and Materials or Association of Official Analytical Chemists (AOAC) takes on the job of overseeing a validation, the work of the end laboratory will be to make sure that it carries out the method in such a way as to preserve the prior validation, and documents any aspects specific to that laboratory.

The second aspect of some definitions is that validation must be seen to be done through proper documentation, objective evidence as the last definition given above (taken from International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) 17025) says. If a laboratory is seeking accreditation to ISO/IEC 17025 (general requirements for the competence of

## Further Reading

ISO (1985) *Certification of Reference Materials – General and Statistical Principles*. ISO/IEC Guide 35-1985. Geneva: International Organization for Standardization.

ISO (1996) *Quality System Guidelines for the Production of Reference Materials*. ISO/IEC Guide 34-1996. Geneva: International Organization for Standardization.

Quevauviller Ph and Maier EA (1999) *Interlaboratory Studies and Certified Reference Materials for Environmental Analyses*. Amsterdam: Elsevier, 558pp.

Stoeppler M, Wolf WR, and Jenks PJ (eds.) (2001) *Reference Materials for Chemical Analysis – Certification, Availability and Proper Usage*. Weinheim: Wiley-VCH, 297pp.

Walker RF (1997) Quality system requirements for the production of reference materials. *Trends in Analytical Chemistry* 16: 9–16.

## Method Validation

**D Brynn Hibbert**, University of New South Wales, Sydney, NSW, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

It is almost a truism that the use of an analytical method implies that it will, when implemented properly, deliver a useful result. However, someone must have first described the method but is the user of the method sure that, when it was developed, enough checks were made to demonstrate its validity? This leads to a further question of what appropriate checks may be, and how the validity of the method can be demonstrated.

This article catalogs and describes the different aspects of a method that must be studied to ensure that it is, indeed, a valid method. The relationship between systematic measurements made during development and validation, the so-called development–validation cycle will be discussed. When using a previously validated method, there is still a necessity to assert that the method performs as expected in the user's laboratory. This is known as verification.

### What is Method Validation?

Definitions found in the literature emphasize different aspects of validation. For example:

- The provision of documentary evidence that a system fulfills its predefined specification...involves the evaluation of the fitness of analytical methods for their purpose.
- The process of proving that an analytical method is acceptable for its intended purpose.
- Checks...to ensure that the performance characteristics of the method are understood and

demonstrate that the method is scientifically sound under the conditions under which it is to be applied.

- Analytical measurements should be made using methods and equipment that have been tested to ensure they are fit for purpose.
- Confirmation by examination and the provision of objective evidence that the particular requirements for a specific intended use are fulfilled.

Many of these definitions use the concept of fit for purpose. The importance of this is the implication that there is no absolute worth of a method. If the results are sufficient for the requirements of the client, and the client genuinely receives useful results, the method is fit for purpose. How well a client views a method is tied to its intended use and so the process of method validation must be ultimately guided by the requirements of intended use. This appears to imply that every laboratory will need to embark on extensive experimentation with each new job. Sensibly, it is recognized that a particular method is used in a broadly similar manner across the world, and so if an international agency such as American Society for Testing and Materials or Association of Official Analytical Chemists (AOAC) takes on the job of overseeing a validation, the work of the end laboratory will be to make sure that it carries out the method in such a way as to preserve the prior validation, and documents any aspects specific to that laboratory.

The second aspect of some definitions is that validation must be seen to be done through proper documentation, objective evidence as the last definition given above (taken from International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) 17025) says. If a laboratory is seeking accreditation to ISO/IEC 17025 (general requirements for the competence of



calibration and testing laboratories), it must use methods for which the validation can be documented for scrutiny by the assessor. In the pharmaceutical industry, the requirements of government regulators, such as the Food and Drug Administration in the USA and the Therapeutic Goods Administration in Australia, are exact and specify complete documentation of the validity of in house methods that are used in quality control of the production of drugs.

## **When Should Methods be Validated?**

The extent and timing of validation is determined by the requirements of the problem. Of course, a completely new method must be validated to show that it is indeed capable of solving the analytical problem at hand. With an established method, as improvements are made, for example, a new column in chromatography, or improved detector in spectroscopy, the method should be sufficiently revalidated to determine the performance of the novelty. A method is validated for a particular analysis, and so if the method is used for a new problem it should be revalidated. The aspects of a method that are laboratory-specific must be validated whenever changes are made, a new analyst, different instrumentation, or a move to a new laboratory. The comparison between a new method and an existing method may also be seen as validation.

The person responsible for ensuring that a method being used is appropriately valid is the analyst himself or herself. However, much of the hard work is done by international organizations who publish standard methods. These are often prescribed in legislation and it is clear that such methods have to survive the greatest scrutiny. AOAC International is a strong supporter of interlaboratory trials as the preferred way of validating methods. AOAC has introduced its Peer Verified Method Program for the validation of methods by laboratories working with only one or two others, and more recently harmonized guidelines for single laboratory method validation have been published by International Union of Pure and Applied Chemistry (IUPAC). The suite of procedures adopted by a laboratory before applying a validated method is sometimes known as verification. Many laboratories rely on published validations, but it is important that some measure of verification is done for all methods.

## **The Validation Process**

There are two levels of validating a method. First, what parameters of a method need validating must

be determined by reference to the end use, preferably in consultation with the client, or the stakeholders group. At this stage the requirements are decided. For example, a target measurement uncertainty may be specified as a particular percentage of the measurement result, or the scope of validation is set to take in changes in instrument, operator, etc.

### **Development – Validation Lifecycle**

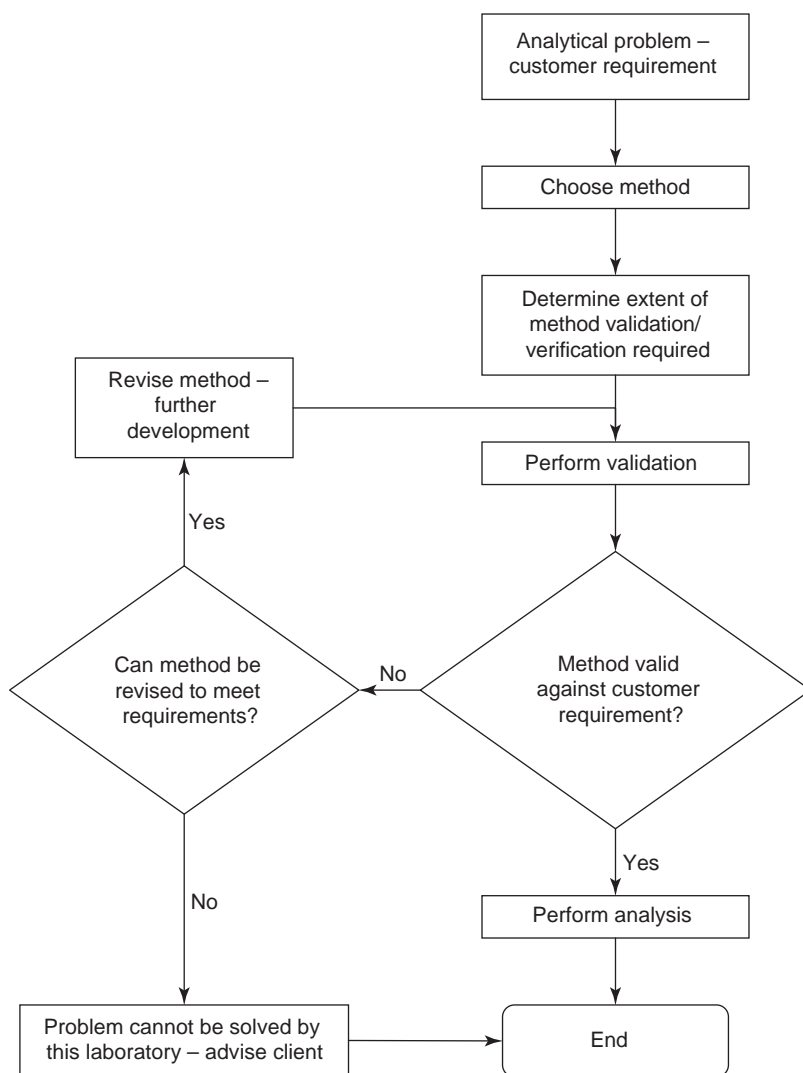
While validation is distinct from the development of method conditions it cannot be totally separated from it. Data from the development process can be used in validation studies, particularly precision data and determination of detection limits. It is unlikely that full robustness studies will have been performed during development, but if during validation it becomes evident that the method cannot meet the desired requirements then further development will be needed. Development/validation may therefore be seen as an iterated process.

The relationship between validation, development, and customer requirements is shown in **Figure 1**.

### **Validation Requirements for Different Purposes**

Methods are either valid or they are not, and it has been said that method validation is no more than the expression of the full measurement uncertainty of a method. However, historically, there are eight performance parameters that are considered to be capable of validation. These are given in **Table 1**. The hierarchy shows the order in which the validation parameters might be studied. For example, if the method does not actually analyze the required substance, then there is no point in worrying about precision or robustness.

Not all aspects of a full validation protocol are important for particular uses of some methods. For example, limits of detection or determination are not an issue for the analysis of the active in a 500 mg paracetamol tablet, but for trace impurities it is, of course, necessary to demonstrate these limits. However, in the documentation of a method validation it is useful to include each heading and at least give reasons for asserting that, as in the example above with limit of detection, one or more headings are manifestly satisfied. **Table 2** gives some examples of parameters that must be considered particularly carefully for different analytical requirements. It is stressed, however, that these are the parameters that address the particular requirements tabulated, and that other aspects might need to be validated to achieve a complete method validation.



**Figure 1** The relationship between validation, development, and customer requirements.

### Method Validation in the Pharmaceutical and Food Industries

Some highly regulated sectors, such as the pharmaceutical industry, have their own guides and regulations that must be followed. They largely concur with the more general approaches described here, but specific documents should be consulted to ensure compliance with national and international regulation. In particular, the ICH (International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use) guides cover requirements for the pharmaceutical industries in the USA, Europe, and Japan. The US has a number of regulations for the industry, published as Codes of Federal Regulations, which refer to US Pharmacopeia Compidual Methods. Use of these methods must be validated according to United States Pharmacopeial (USP) Convention

1225 – Validation of Compidual Methods. An example of regulations from other parts of the world is the Nordic Food Regulations, which govern method validation for the food industry in Scandinavia.

### Method Validation Parameters

#### Identity

It may not be trivial for an analyst to say that the method actually does measure what it purports to. When there are possibilities of speciation, or different isomers, then the analyst must be aware of the client's requirements and ensure that the method is specific for them. For example, analysis for 'total copper' is different from 'bioavailable copper', and with the latter the definition of 'bioavailable' must be carefully considered.

**Table 1** Performance parameters studied in a complete method validation

<i>Parameter</i>	<i>Description and comments</i>
Identity	
Selectivity	(Specificity). Determination of the extent of the effects of interfering substances
Limits	
of detection	Minimum result below which the presence of the analyte cannot be determined
of determination	(Quantitation). Minimum result attainable with given precision
Calibration parameters	(Linearity). Adequacy of the calibration model.
Sensitivity	Slope of a linear calibration plot. Values of coefficients of the calibration model
Trueness	(Accuracy). Demonstration of the absence of systematic error.
Recovery	The ability to determine the analyte in a given matrix
Precision	
Repeatability	Same analyst, same instrument, short time between replicates
Reproducibility	
Intralaboratory	Different analyst, instrument, longer time scale but within the laboratory
Interlaboratory	Precision between laboratories, measured in an interlaboratory study
Ruggedness	(Robustness). Ability of method to remain unaffected by small variations in method parameters Some authors make the distinction between the property robustness and a ruggedness test in which deliberate changes are made in a method in order to assess the robustness

**Table 2** Important method validation parameters for particular analytical requirements

<i>Analytical requirements</i>	<i>Important method validation parameters</i>	<i>Other validation parameters</i>
Qualitative analysis only	Selectivity, limit of detection	
Analysis of a particular form of a substance, e.g., oxidation state, isomer, conformer	Selectivity	Calibration, trueness, precision, ruggedness, (limits, recovery)
Measurement of an analyte in a matrix	Selectivity, recovery	Calibration, trueness, precision, ruggedness, (limits)
Trace analysis	Selectivity, limits of detection and determination, precision, recovery	Calibration, trueness, precision, ruggedness
Results that will be compared with those from other laboratories	Precision – interlaboratory reproducibility, trueness	Calibration, precision, ruggedness (limits, recovery)
Results that will be compared with limits or specifications	Trueness, precision, calibration parameters	Ruggedness (limits, recovery)
Method that is implemented by different analysts on different instruments	Precision – intralaboratory reproducibility, trueness, ruggedness	Calibration, precision, (limits, recovery)
Require to analyze to target measurement uncertainty	Precision	Calibration, trueness, ruggedness (limits, recovery)
A number of samples having a range of concentrations	Calibration parameters, trueness, precision	Ruggedness (limits, recovery)

A method is specific to a given analyte if it is 100% selective, i.e., there are no interferences of the analytical signal. Sometimes this may be obvious, but one of the main dangers in using a validated method for a new matrix is that novel interferences might affect the measurement result. For methods giving a signal peak, for example, gas or liquid chromatography or nuclear magnetic resonance (NMR), the peak purity must be established. This can be done in liquid chromatography by use of a multi-spectrum diode detector or by performing experiments with different columns; in mass spectrometry by increasing the mass resolution or by using a different ionization technique; and in NMR by the

use of higher field instruments or by observing different nuclei. When the likely interferences are known the response of the instrument to them can be assessed from spiking studies. Alternatively, the purity of a sample may be investigated by another more general method such as infrared spectroscopy, when the spectrum of the test sample can be compared with a library spectrum of the authentic analyte. This allows interferences that have an infrared spectrum above a certain detection limit to be deemed to be absent.

It is always better to have a method that only analyzes the desired measurand, by removing potential interfering substances in the sample preparation

or in a separation stage before detection (chromatography). If they cannot be removed, and the effect is small, then the presence of the interferent may be treated as a systematic error and either corrected for or allowed for in the measurement uncertainty. This limits the method to a particular matrix containing a particular interferent, or requires a series of experiments to quantify the effect of the interferent in all possible test samples.

Ultimately the analyst can never rule out a false positive, or the interference of some highly similar species that has not been considered. Professional judgment, the time and budget available will indicate to what extent experiments must be done to demonstrate specificity. As long as these are done and documented to the client's satisfaction, the method is considered properly validated. If a future user of the method has particular concerns about other interferences, then revalidation should be undertaken to study these substances.

## Limits

**Limit of detection** The most widely used definition of detection limit is derived from the signal of a blank plus three standard deviations of replicate blank determinations. Thus, for a linear calibration relation between concentration  $c$  and indication of the measuring system (sometimes called response)  $I$ ,

$$I = a + bc \quad [1]$$

$$c_{dl} = \frac{I_{blank} + 3s_{blank} - a}{b} \quad [2]$$

where  $s_{blank}$  is the standard deviation of 10 responses of a blank sample ( $I_{blank}$ ) performed under repeatability conditions. The use of eqn [2] presupposes that it is possible to measure a signal from a blank, or even create a sample that could be described as a blank. Not all techniques yield a sensible signal without sample, for example, ion-selective electrodes. This estimate also assumes that the blank signal is a normally distributed random variable, quite reasonable for a spectroscopic signal where the background is white noise, but not necessarily the case for all methods. Indeed, the blank signal is likely to be greatly influenced by the matrix.

If it is not possible to record a background or blank signal, the standard deviation of the blank can be estimated by the standard error of the regression of the calibration equation ( $s_{y/x}$ ) and  $I_{blank}$  by the intercept of the calibration equation ( $a$ ), when eqn [2] becomes

$$c_{dl} = \frac{3s_{y/x}}{b} \quad [3]$$

A more statistically valid approach that uses the calibration line has been recommended by IUPAC

$$c_{dl} = \frac{2t_{0.05', n-2} s_{y/x}}{b} \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{\bar{c}^2}{\sum_i (c_i - \bar{c})^2}} \quad [4]$$

where  $t_{0.05', n-2}$  is the one-tailed Student- $t$  value for  $\alpha=0.05$  and  $n-2$  degrees of freedom,  $m$  is the number of replicates of the blank measurement,  $n$  the number of points in the calibration line, and  $\bar{c}$  the mean value of the calibration concentrations ( $c_i$ ). It should be noted that eqns [3] and [4] are very sensitive to the linearity of the calibration line, and may overestimate the true detection limit if  $s_{y/x}$  is large.

Whichever method is used to estimate the detection limit, it should be clearly described in the method validation.

Having estimated the minimum detectable concentration, it is always a good idea to spike a matrix with that concentration and demonstrate the instrument response to it.

**Limit of determination** There are two approaches to the limit of determination (quantitation). First, a conventional way is to define the limit in terms of the standard deviation of 10 replicates of the blank ( $s_{blank}$ ) as  $4 s_{blank}$ ,  $5 s_{blank}$ , or  $10 s_{blank}$ .

Second, a series of replicate analyses at increasing concentrations (10 at each) are performed and the standard deviation calculated at each. Usually the relative standard deviation (RSD) increases as the concentration decreases, and the limit of determination is taken as the concentration at which the RSD reaches a predetermined value.

## Calibration Equation, Range, and Analytical Sensitivity

Most modern analytical methods require calibration, because the signal that is obtained from the instrument does not give the result directly, but must be compared with a signal from one or more samples that have a known quantity value. Thus, the relationship between the measurand and the instrumental signal is established. Historically, and also for good statistical reasons, a linear calibration relation is the most common one found and desired in analytical science. In common method validation parlance, working range is defined by the maximum and minimum concentrations that the method is to be used, and linearity refers to the demonstration of a linear calibration relation across the so-called linear range. Sensitivity is the slope of the linear plot, i.e., the increase in the instrumental signal with concentration of analyte. In more general terms, with the

possibility of nonlinear calibration, and multivariate calibration, the role of method validation is to assess the adequacy of the chosen calibration model and determine the range across which it applies.

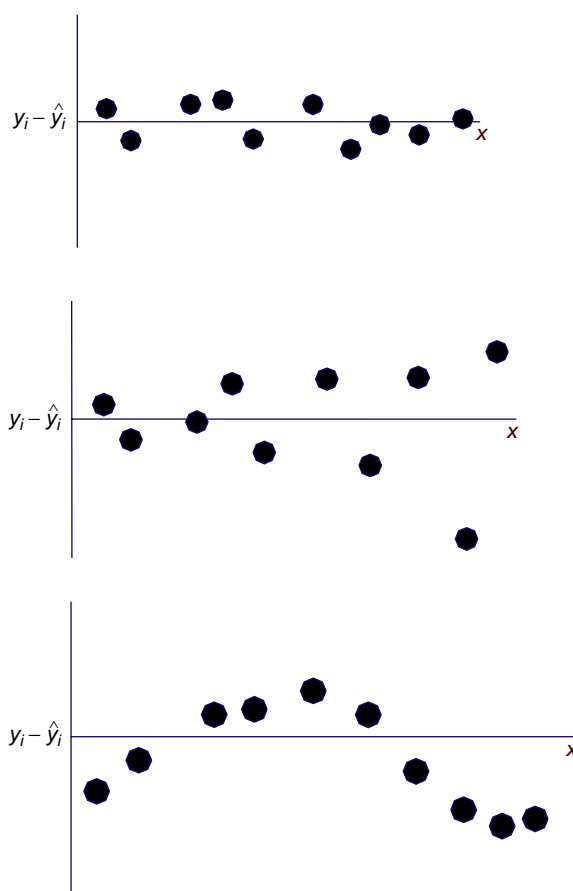
To obtain adequate statistics, the calibration range should be established by at least 10 independently prepared solutions. If a linear calibration model is assumed and the parameters of the model (slope and intercept) are determined by classical least-squares regression, the assumptions of the regression must hold. Namely:

- the model is a true description of the relationship between concentration and instrument response;
- the only errors are normally distributed in the response; and
- the errors are of constant variance across the linear range (homoscedasticity)

When constructing a linear calibration plot, we are not testing the linear relation. Hopefully this is established, for example, by Beer and Lambert in spectroscopy and Nernst in electrochemistry. It is important, however, to understand the limits of the range – the limits of detection and determination at the lower concentrations and instrumental saturation at the upper end. One procedure is to establish the lower limit as the prescribed limit of determination, then add points to the calibration until adding a further point at the high-concentration end violates some criterion. This may be when the error in the last point ( $y_n - \hat{y}_n$ ) becomes greater than an acceptable value, or when the estimation of the next point by extrapolation is not acceptable. If the measurement variance is known, the square of standard error of the regression can be compared against it by a one-sided *F*-test

$$F = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - 2)}{s^2} \quad [5]$$

Inspection of the calibration line itself, or using a high value of the correlation coefficient ( $r$ ) or coefficient of determination ( $r^2$ ), say better than 0.999, is not an adequate confirmation of linearity across the range, nor proof that when used to assign the concentration to a test solution it will not have unacceptable errors. A simple spreadsheet study has shown that data generated according to a nearly linear model, for example,  $y = x^{1.001}$ , and then fitted to a linear relation and used for calibration, errors of up to 70% were observed in the recovered values of  $x$ , despite  $r^2 = 0.999$ . The problem lies in the violation of the first assumption, that the data are inherently linear. Inspection of residuals is the best visual way of checking linearity. Nonlinearity,



**Figure 2** Examples of plots of residuals of calibration graphs. Top: normally distributed residuals with constant variance (homoscedastic data). Middle: data showing an increasing variance with increasing concentration (heteroscedastic data). Bottom: residual plot of a curved calibration relation.

outliers, and heteroscedacity are seen in distinctive shapes (Figure 2).

Having decided the range of the calibration, the parameters are determined from the data, making sure that at least 10 independent concentrations have been analyzed within the range. Modern spreadsheets give standard errors of the regression ( $s_{y/x}$ ) and of the slope and intercept, all of which should be quoted. The ultimate goal is to obtain the best-quality results from the calibration, and the analysis of an independent test solution, i.e., one that has not been used to construct the calibration line, is recommended. The measurement result can be compared with the assigned value, and the uncertainty calculated from

$$s_{\hat{x}_0} = \frac{s_{y/x}}{b} \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b^2 \sum_{i=1}^n (x_i - \bar{x})^2}} \quad [6]$$

where  $m$  is the number of replicates of the test solution (mean  $y_0$ ), and other terms refer to the

calibration. Equation [6] yields the 95% confidence on the estimate of  $x$  by multiplying by  $t_{0.05'', n-2}$ .

### Accuracy

Accuracy is a measure of the closeness of a result to the true value. It has two components, trueness and precision. In simple terms, lack of trueness indicates systematic error (bias in the case of instrumental results) and precision encompasses all aspects of random error.

**Trueness** Trueness is measured in terms of the systematic error (bias) in a measurement result and is the difference between a result obtained by the analyst and the true value of the measurand. For a measurement result, which is the mean of a series of analyses carried out within a single run, the bias so defined has three components, arising from the method, the laboratory, and the particular run. Assuming the method is in statistical control, the within run variation should be from a normal population with expectation zero. **Figure 3** shows these contributions to trueness. It would also be hoped that the laboratory could achieve a zero run bias. The systematic effect from the laboratory and method combined can be measured by analyzing a certified reference material (CRM) at least 10 times in several runs (assuming such material is available).

The difference between the mean result and the certified value is this bias. Collaborative trials between laboratories can establish the method bias and individual laboratory biases as shown in **Figure 3**. Collaborative trials also establish the reproducibility standard deviation.

Proficiency testing data may also be used to estimate laboratory bias.

Once bias has been estimated its magnitude is tested to decide if it is significant and must be corrected for. The bias and its uncertainty determined from the mean of repeated ( $n$ ) analyses of a CRM is

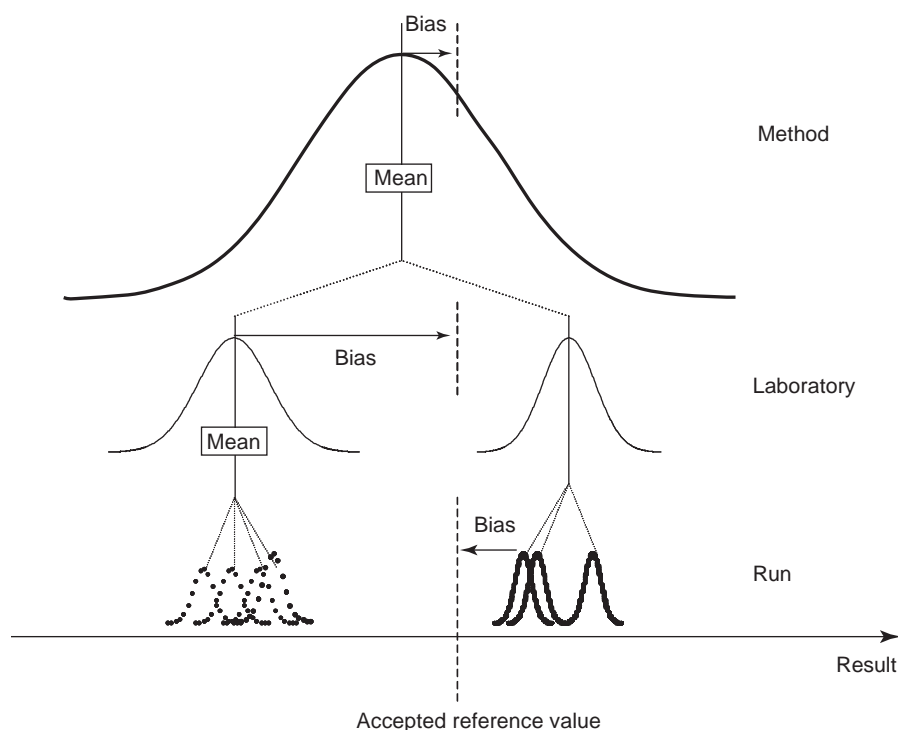
$$\delta = \bar{c} - c_{\text{CRM}} \quad [7]$$

$$u(\delta) = \sqrt{\frac{u_r^2}{n} + u^2(\text{CRM})} \quad [8]$$

where  $u_r$  is the repeatability standard deviation, and  $u(\text{CRM})$  is the uncertainty in the value of the CRM.  $\delta$  may be tested by a  $t$ -test with null hypothesis that there is no bias

$$t = \frac{|\delta|\sqrt{n}}{u(\delta)} \quad [9]$$

If the bias is deemed significant the result should be corrected. If the laboratory and method components



**Figure 3** Schematic showing the relationships among run bias, laboratory bias, and method bias. The standard deviation of the results of an interlaboratory method precision study includes the randomized components of run and laboratory bias.



are not independently known the use of the method is restricted to the laboratory in which it was validated. Bias may be a function of concentration and if this is suspected experiments should be done with a series of CRMs, to determine the proportional and fixed bias terms.

In the absence of a CRM, an uncertified reference material may be used if the analyst is satisfied that the material is well characterized enough to use for this purpose. Alternatively, a material may be analyzed by a reference method and the bias determined with respect to the reference method. Recovery studies (see below) may also give an estimate of bias.

**Precision** Precision is the closeness of agreement between results obtained under specified conditions, and is usually expressed as a standard deviation or RSD. Depending on the conditions specified, contributions may appear as bias or precision. For example, the temperature effect may be viewed as a random variable if sufficient experiments are done allowing the temperature to vary. On the other hand, if temperature is controlled and measured any difference from the target temperature can be corrected for as a bias. The measurement uncertainty is reduced, but still contains a contribution from the measurement of the effect (bias).

The repeatability is defined as the closeness of agreement between the results of successive measurements of the same measurand carried out subject to the following conditions:

- the same measurement procedure,
- the same observer,
- the same measuring instrument used under the same conditions,
- the same location, and
- repetition over a short period of time,

and the reproducibility as the closeness of agreement between the results of measurements of the same measurand, where the measurements are carried out under changed conditions such as:

- principle or method of measurement,
- observer,
- measuring instrument,
- location,
- conditions of use, and
- time.

In process control, measurements are often performed in duplicate. The difference between each measurement result may be tested against the

expected 95% probability interval of  $2 \times s_r \times \sqrt{2}$ , where  $s_r$  is the repeatability standard deviation at the specified concentration. This difference limit is known as the repeatability limit,  $r$ ,

$$r = 2 \times \sqrt{2} s_r = 2.8 s_r \quad [10]$$

When quoting reproducibility data it is important to include a statement of which aspects of the measurement system change. Standard methods that have been subjected to interlaboratory trials (interlaboratory method validation studies) quote the reproducibility as a 95% confidence interval: that is, the difference that two laboratories making measurements on portions of the same test material will exceed once in 20 times. The reproducibility precision will include the repeatability precision and components arising from the changes in the system. The relationship between this reproducibility limit,  $R$ , and the reproducibility standard deviation  $s_R$  is

$$R = 2 \times \sqrt{2} s_R = 2.8 s_R \quad [11]$$

For a laboratory performing an in-house method validation study, it is recommended that 10 test portions be analyzed at each concentration over the range of validation.

### Ruggedness (Robustness)

A method is rugged if it can remain sufficiently unaffected during normal use in which a number of changes in environmental conditions and the measurement system may occur. It is assessed by a robustness study in which deliberate changes are made and their effects determined. What is changed is chosen after consultation between analyst and client or by the professional judgment of the analyst. In the example of a robustness study of a chromatographic method parameters changed might include: column temperature, manufacturer of the column, composition of the mobile phase, and pH of the test portion. As a robustness study is a screen for unduly great changes in measurement result, it is only necessary to estimate the main (linear) effect of changing each parameter, not conduct a full determination including interaction or higher order effects. This is accomplished by a highly fractionated experimental design, such as a Plackett–Burman design in which  $(4 \times i - 1)$  parameters may be studied in  $4 \times i$  experiments ( $i = 1, 2, 3$ , etc.). Only two levels of each factor are studied, the method value and one changed by a challenging, but not ridiculous amount. A design for seven factors and eight experiments is shown in Table 3.

**Table 3** Contrast coefficients for a Plackett–Burman experimental design for seven influence factors

Experiment	Factor						
	1	2	3	4	5	6	7
1	+	+	+	–	–	+	–
2	–	+	+	+	–	–	+
3	+	–	+	+	+	–	–
4	–	+	–	+	+	+	–
5	–	–	+	–	+	+	+
6	+	–	–	+	–	+	+
7	+	+	–	–	+	–	+
8	–	–	–	–	–	–	–

The sign of the contrast coefficient indicates the level of the factor: –, normal experimental level; +, changed level.

The effect of a factor is obtained by summing each experimental response multiplied by the contrast coefficient divided by 4 (number of experiments/2).

Conventionally, the level labeled minus is the method value and the level labeled plus is the changed value. The calculation gives the average change in measurement result (which may be expressed as the change in the indication of the measuring system) when the parameter of interest changes from the method value to the changed value.

The specification of the method validation has to decide what an acceptable effect is for each factor studied.

### Recovery

Recovery is a bias usually associated with sample preparation or pretreatment. It is expressed as the measurement result as a percentage of the certified value. To obtain a proper estimate of recovery a matrix CRM should be analyzed. This value is sometimes known as the analytical yield, or apparent recovery, to distinguish it from the recovery of spiked samples described below.

If only pure analyte is available the surrogate recovery is obtained by analysis of a test material and then analyzed again after spiking with a known mass of pure analyte. The difference between the measurement results before and after spiking as a percentage

of the added material is known as the surrogate or marginal recovery. As with bias, recovery should be tested for significance and the result corrected if necessary.

**See also:** **Quality Assurance:** Quality Control; Interlaboratory Studies; Reference Materials; Production of Reference Materials; Accreditation.

### Further Reading

- Burgess C (2000) *Valid Analytical Methods and Procedures: A Best Practice Approach to Method Selection, Development and Evaluation*. Cambridge: Royal Society of Chemistry.
- Christensen JM, Kristiansen J, Hansen AaM, and Nielsen JL (1995) Method validation: An essential tool in total quality management. In: Parkany M (ed.) *Quality Assurance and TQM for Analytical Laboratories*, pp. 46–54. Cambridge: Royal Society of Chemistry.
- Eurachem Working Group (1998) *Eurachem Guide: The Fitness for Purpose of Analytical Methods*. Teddington: LGC (Teddington) Ltd.
- Fajgelj A and Ambrus A (eds.) (2000) *Principles and Practices of Method Validation*. Cambridge: Royal Society of Chemistry.
- Green JM (1996) A practical guide to analytical method validation. *Analytical Chemistry* 68(9): 305A–309A.
- ICH (1995) ICH Q2A Text on Validation of Analytical Procedures. *International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use*, Geneva, Switzerland.
- ICH (1996) ICH Q2B Validation of Analytical Procedures: Methodology. *International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use*, Geneva, Switzerland.
- Thompson M, Ellison S, and Wood R (2002) Harmonized guidelines for single laboratory validation of methods of analysis. *Pure and Applied Chemistry* 74: 835–855.
- United States Pharmacopeial Convention (1999) *The United States Pharmacopeia 24 – National Formulary 19, General Information, <1225> Validation of Compendial Methods*, United States Pharmacopeial Convention, Inc., Rockville, Maryland.

## Traceability

**M Sargent**, LGC Limited, Teddington, UK

### Introduction

Traceability is one of the principal requirements for achieving comparability of measurements across

**Table 3** Contrast coefficients for a Plackett–Burman experimental design for seven influence factors

Experiment	Factor						
	1	2	3	4	5	6	7
1	+	+	+	–	–	+	–
2	–	+	+	+	–	–	+
3	+	–	+	+	+	–	–
4	–	+	–	+	+	+	–
5	–	–	+	–	+	+	+
6	+	–	–	+	–	+	+
7	+	+	–	–	+	–	+
8	–	–	–	–	–	–	–

The sign of the contrast coefficient indicates the level of the factor: –, normal experimental level; +, changed level.

The effect of a factor is obtained by summing each experimental response multiplied by the contrast coefficient divided by 4 (number of experiments/2).

Conventionally, the level labeled minus is the method value and the level labeled plus is the changed value. The calculation gives the average change in measurement result (which may be expressed as the change in the indication of the measuring system) when the parameter of interest changes from the method value to the changed value.

The specification of the method validation has to decide what an acceptable effect is for each factor studied.

### Recovery

Recovery is a bias usually associated with sample preparation or pretreatment. It is expressed as the measurement result as a percentage of the certified value. To obtain a proper estimate of recovery a matrix CRM should be analyzed. This value is sometimes known as the analytical yield, or apparent recovery, to distinguish it from the recovery of spiked samples described below.

If only pure analyte is available the surrogate recovery is obtained by analysis of a test material and then analyzed again after spiking with a known mass of pure analyte. The difference between the measurement results before and after spiking as a percentage

of the added material is known as the surrogate or marginal recovery. As with bias, recovery should be tested for significance and the result corrected if necessary.

**See also:** **Quality Assurance:** Quality Control; Interlaboratory Studies; Reference Materials; Production of Reference Materials; Accreditation.

### Further Reading

- Burgess C (2000) *Valid Analytical Methods and Procedures: A Best Practice Approach to Method Selection, Development and Evaluation*. Cambridge: Royal Society of Chemistry.
- Christensen JM, Kristiansen J, Hansen AaM, and Nielsen JL (1995) Method validation: An essential tool in total quality management. In: Parkany M (ed.) *Quality Assurance and TQM for Analytical Laboratories*, pp. 46–54. Cambridge: Royal Society of Chemistry.
- Eurachem Working Group (1998) *Eurachem Guide: The Fitness for Purpose of Analytical Methods*. Teddington: LGC (Teddington) Ltd.
- Fajgelj A and Ambrus A (eds.) (2000) *Principles and Practices of Method Validation*. Cambridge: Royal Society of Chemistry.
- Green JM (1996) A practical guide to analytical method validation. *Analytical Chemistry* 68(9): 305A–309A.
- ICH (1995) ICH Q2A Text on Validation of Analytical Procedures. *International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use*, Geneva, Switzerland.
- ICH (1996) ICH Q2B Validation of Analytical Procedures: Methodology. *International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use*, Geneva, Switzerland.
- Thompson M, Ellison S, and Wood R (2002) Harmonized guidelines for single laboratory validation of methods of analysis. *Pure and Applied Chemistry* 74: 835–855.
- United States Pharmacopeial Convention (1999) *The United States Pharmacopeia 24 – National Formulary 19, General Information, <1225> Validation of Compendial Methods*, United States Pharmacopeial Convention, Inc., Rockville, Maryland.

## Traceability

**M Sargent**, LGC Limited, Teddington, UK

### Introduction

Traceability is one of the principal requirements for achieving comparability of measurements across

different places and different periods of time. If measurements are to be accepted everywhere, they must not only be reliable but also be made on a comparable basis to those obtained elsewhere or at other times. Hence, international quality standards such as ISO 17025 stress the need for traceability to appropriate measurement references, as well as the more familiar requirements of operating in accordance with a comprehensive quality system and using properly validated methods. Whilst it may be feasible for two results to be compared directly when necessary, a more general approach is needed to provide comparability between many results obtained at different times or places. This can be achieved by linking each measurement result to a common reference point or measurement standard. Results may then be compared through their relationship to that common reference. The concept of linking results to a stable common reference point is termed traceability.

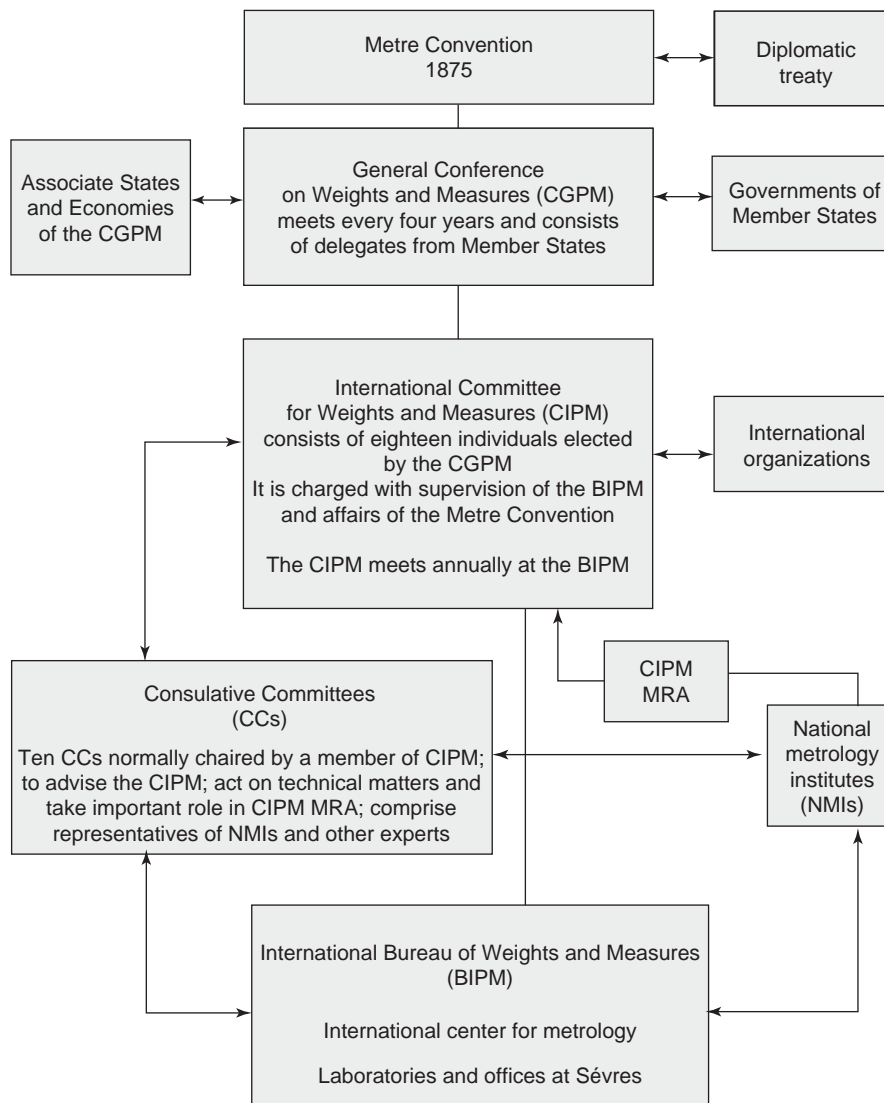
Traceability is formally defined as the property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties. This definition has achieved global acceptance in the metrology community, which is responsible for physical measurements such as time, mass, or length, and metrologists have worked for well over 100 years to achieve international comparability of their measurements in this way. The outcome has been an International Measurement System that is embraced by virtually every developed country in the world. This system is described in more detail later in this article. It is timely, however, to emphasize one aspect of traceability that is often misunderstood. Traceability is not an inherent property of the internationally recognized system of measurement units (the *Système International d'Unités*, SI), such as the kilogram, meter, or mole. The concept of traceability works because a complex infrastructure has been developed within the International Measurement System to underpin the realization of these units. This infrastructure allows measurement institutes around the world to regularly intercompare their measurement standards or procedures. This provides a global basis for the reference points that these institutes use for the measurement services they provide to industry and other users. These services include traceable standards or calibration facilities that are available to end users, either directly or through secondary suppliers. The International Measurement System also provides a framework within which these institutes can

collaborate to resolve measurement differences and improve measurement techniques.

The International Measurement System underpins accurate and comparable measurements by ensuring that each measurement result for a particular parameter is traceable to a reference that is accepted throughout the world. The original point of reference for each unit was a unique artifact kept at a laboratory near Paris (see below), for example, the international standard meter or kilogram. More recently, the emphasis has been on realization of the relevant base (SI) unit as a standard of measurement, usually achieved through development of an extremely accurate measurement procedure for that parameter. In either case, the concept of traceability depends on a chain of measurements linked back to the appropriate international primary standard through a series of calibrations (i.e., comparisons between two standards in the chain). Provided that the uncertainties of the comparisons are known, a measurement result obtained through calibration against one of these standards will itself be traceable to the agreed reference.

## **Development of the International Measurement System**

The basis for international metrology was established in Paris in 1875 when a diplomatic treaty entitled the 'Convention of the Metre' was signed by representatives of 17 nations. That Convention, which was modified slightly in 1921, now has 51 Member States, including all the major industrialized countries. It remains the basis of all international agreement on units of measurement and, as mentioned above, embraces not only the traditional physical quantities but also areas such as radiation, chemistry, and biology. In order to achieve its aims, the Convention established a permanent organizational structure to enable member governments to work together on matters relating to units of measurement. The key components of this structure, which are shown in **Figure 1**, were the *Conférence Générale des Poids et Mesures* (CGPM), the *Comité International des Poids et Mesures* (CIPM), and the *Bureau International des Poids et Mesures* (BIPM). Since 1875, the need to demonstrate equivalence between national measurement standards has been met by the BIPM, which is located at Sèvres near Paris, working in collaboration with the national metrology institutes (NMIs) of the member states. The same organizations have also worked together to meet the need for measurement standards of ever increasing accuracy, range, and diversity. These



**Figure 1** The structure of the international measurement system established by the Convention of the Metre in 1875. The chart shows the key international organizations and the links between them. (Reproduced with permission from the BIPM website.)

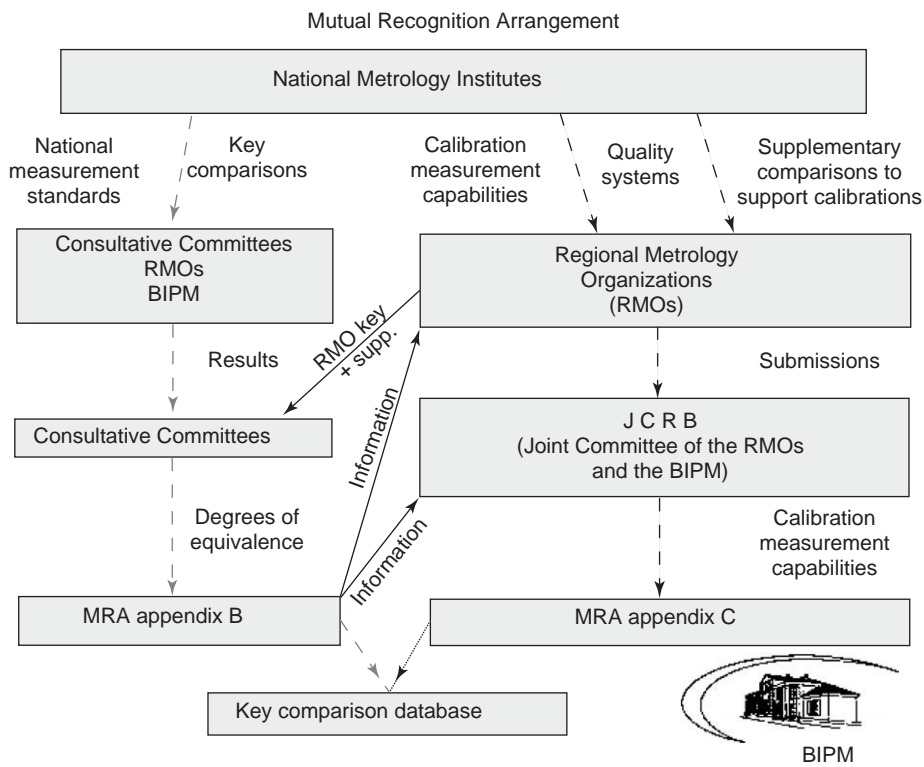
collaborative technical activities are achieved primarily through a number of Consultative Committees, each of which addresses a specific area of measurement and is normally chaired by a member of the CIPM.

A major step forward in the development of international metrology took place at a meeting held in Paris on October 14, 1999, with the establishment of a Mutual Recognition Arrangement (MRA) for national measurement standards and for calibration and measurement certificates issued by NMIs. The MRA was initially signed by the NMIs of 38 Member States and representatives of two international organizations. This Arrangement is a response to a growing need for a comprehensive scheme to give users reliable quantitative information on the

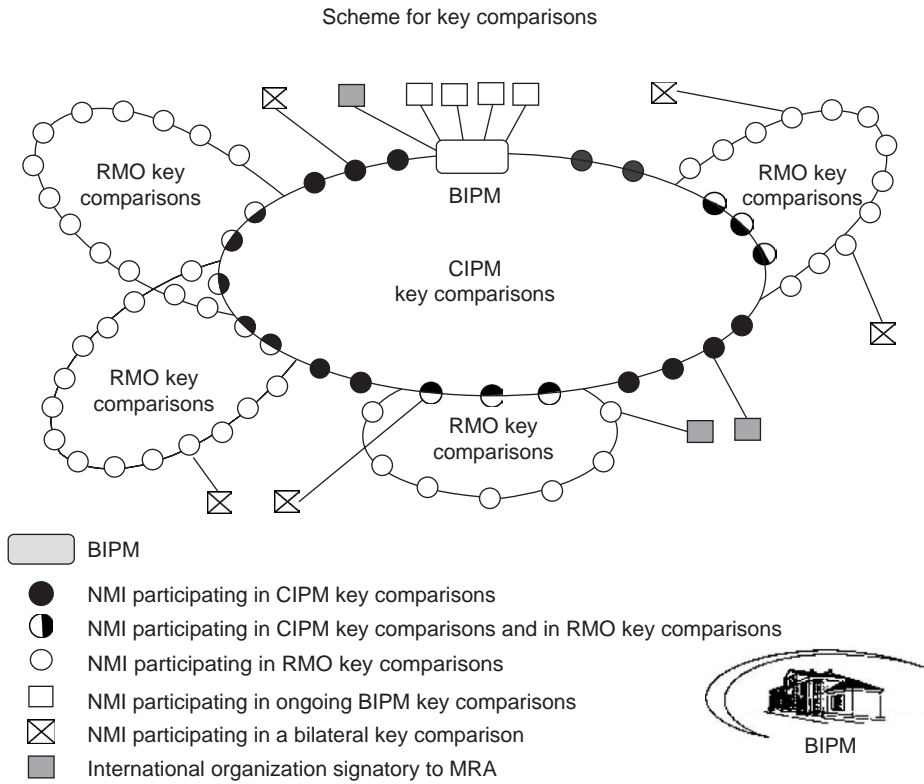
comparability of national metrology services and to provide the technical basis for wider agreements negotiated for international trade, commerce, and regulatory affairs. The organization and participation in the MRA is summarized in **Figure 2**.

A key feature of the MRA is the development of a BIPM key comparison database (KCDB), which includes the results of key and supplementary comparisons (KCs) and the calibration and measurement capabilities (CMCs) of the NMI signatories to the MRA. Key comparisons are high-level interlaboratory comparisons of the standards or measurement procedures of the participating institutes. The scheme for organizing such comparisons is shown in **Figure 3**. The comparisons are operated in the same way as proficiency testing (PT) schemes, i.e.,





**Figure 2** The organization of the BIPM Mutual Recognition Arrangement (MRA), agreed in 1999 between the directors of the national measurement institutes of 38 Member States and representatives of two international organizations. (Reproduced with permission from BIPM website.)



**Figure 3** The scheme of key comparisons established in order to support the aims of the Mutual Recognition Arrangement (MRA) and to derive the degree of equivalence between the standards or calibration services of participating national measurement institutes. (Reproduced with permission from BIPM website.)



standards or samples for intercomparison are circulated to laboratories and measured blind. Moreover, once an institute has agreed to participate in a KC it cannot withdraw its measurement results. The key comparison database is openly accessible through the BIPM website (<http://www.bipm.org>). Once agreed by the participants and approved by the relevant consultative committee, the results of each key comparison are added to the database.

The CMCs of the NMIs are also published in the key comparison database. These capabilities are submitted by each institute in accordance with the services that it offers. CMCs are subjected to two important safeguards to ensure that users of these services obtain traceable data of demonstrable equivalence with data provided by other member institutes. The safeguards are (1) the requirement that each institute should participate in relevant KCs, and demonstrate that its performance is commensurate with its claimed capabilities, and (2) a review procedure that involves experts from all MRA member institutes working at both the regional and global levels.

## The Need for Traceability in Analytical Chemistry

There are relatively few traceable chemical measurement standards in the sense used for physical measurement standards and the concept of traceable measurements is not widely understood by today's analysts. Nevertheless, the need for comparability of measurement results is just as important in analytical chemistry as in physical metrology. Classical analytical chemistry did, indeed, depend on traceable calibration of balances and volumetric glassware to achieve comparable data between laboratories. Prior to the extensive use of instrumentation, analytical laboratory procedures placed emphasis on the origin of the calibrated weights and glassware used, for example, in gravimetric or titrimetric analysis methods and the need to safeguard the integrity of these artifacts. In many analytical laboratories at that time, only a single set of balance weights was externally calibrated and traceable to a national reference; calibration of other balance weights or volumetric glassware was achieved using the set of reference weights and was an important, routine aspect of laboratory operations. In order to understand traceability in chemistry, it is important to appreciate what underlies the use of calibrated weights in this way. It will also become apparent why the concept of traceability seemed less relevant to chemistry as analytical laboratories moved away

from classical analysis and adopted instrumental techniques.

Regardless of the need for traceable calibration of balance weights and volumetric glassware, analytical chemistry is not concerned *per se* with the determination of mass and volume. Its purpose is to determine amount of substance, i.e., the amount of a specific chemical entity such as, for example, copper, potassium dichromate, or ethanol. This is reflected in the definition of the SI unit for chemistry, the mole (symbol: mol). The mole is the amount of substance of a system that contains as many elementary entities as there are atoms in 0.012 kilogram of carbon-12. When the mole is used, the elementary entities must be specified and may be atoms, molecules, ions, electrons, other particles, or specified groups of such particles. In some application areas such as, for example, the clinical sector, the mole is a widely used unit whereas in others, such as industrial or environmental analysis, it is more common to use units based on mass (e.g., milligram per kilogram) or parts per million (ppm). Nevertheless, it is important to remember that, regardless of the preferred unit, the chemist is measuring a chemical entity which must be correctly identified and for which a chemical standard is required. This is a fundamental requirement for traceability in chemistry.

In the classical analytical laboratory, it was common practice for the chemist to prepare chemical standards in-house using a knowledge of chemistry to ensure that they comprised the correct substance and were of sufficient purity for the purpose in hand. Clearly, any error in assessing the identity or purity of such a chemical standard will affect the reliability of measurements dependent on it, just as will errors in weight and volume when using an aliquot of the chemical standard to prepare a calibration solution. This situation was reflected in the growth of the chemical reagent industry, supplying initially the chemical materials and more recently certified calibration solutions. In order to ensure reliable reagents, each supplier adhered to an agreed specification, which might be its own, one agreed within an area of trade, or an international standard. These specifications were, however, largely local or sector-based; no attempt was made to adopt the concept of traceability to provide an international basis for all such materials or to implement a single, international infrastructure to underpin it.

This situation existed partly due to a widespread perception, which continues to this day, that errors in the preparation of calibration standards are not the most pressing problem facing analytical chemists. This arises because calibration using chemical standards is complicated by the dependence of the

chemical measurement process on the sample matrix. In the classical laboratory this problem was overcome by quantitative removal of the analyte from the matrix prior to measurement or by using appropriate chemistry to overcome matrix interferences. In the modern laboratory the analysis is almost invariably instrument based but the instrumental determination is often the final step of a complex analytical method involving extensive pretreatment of the sample. Hence, calibration of the instrument using a pure chemical standard, even a traceable one, is on its own insufficient to achieve reliable and comparable results.

The sample matrix problem has stimulated the development of two pragmatic solutions: matrix reference materials and interlaboratory comparisons. The matrix-matched, certified reference material (CRM) is a unique type of chemical standard commonly used to validate complete measurement methods and sometimes for instrumental calibration (e.g., in XRF). Such standards must be available for each required analyte/matrix combination. Similarly, interlaboratory comparisons are undertaken for each relevant analyte/matrix combination in order to establish comparability of data between laboratories. These comparisons range from round robin studies, which collaboratively test a new method, to formal PT schemes that assess agreement between laboratories on an ongoing basis.

CRMs and PT schemes have been used by analysts with reasonable success over many years but they both have a number of technical, practical, and economic limitations. The need for a wide variety of application-specific CRMs has led to fragmented production without any formal relationship between the certified values of CRMs produced for different applications or by different organizations. There are thousands of CRMs in use but many of those required for critical applications such as manufacturing, trade, health, or the environment are unavailable. In addition, production costs are high and it is difficult or impossible to manufacture sufficiently stable CRMs for some applications. Interlaboratory comparisons also have a number of limitations, particularly that they are time-consuming and expensive. Comparability usually extends only to the immediate participants in a single comparison because comparability between different comparisons is rarely established, even when they are coordinated by the same organization. It is impracticable to organize comparisons for every routine application or to organize a worldwide comparison involving all the laboratories requiring comparability for each measurement application.

These problems have long been recognized as a significant technical and economic limitation in

delivering sound chemical analysis data. The situation is steadily worsening with increasing demand from purchasers of data and by regulators for proven comparability of measurements. This is for several reasons. Global expansion of trade, which means more countries and more laboratories need to be brought into each interlaboratory comparison. In addition, increasing numbers of measurements are used in support of regulations, for which there is an expanding requirement for rigorously proven reliability and comparability. Finally, increasing use of subcontracted measurements, due to commercial pressures on laboratories, requires not only conformity of contractors to quality systems but also demonstration of the comparability of data from different contractors. To ensure reliable and comparable chemical measurements in the twenty-first century, it is desirable to have in place a unified international system based on the traceability of measurement results.

## **Achieving Traceability in Analytical Chemistry**

Traceable chemical measurement results require a measurement infrastructure analogous to the systems that underpin physical measurements, as described above. Chemical measurements have developed more or less on a sectorial basis and in a different culture, so that the systems developed for physical measurements cannot easily be directly applied to chemical or, indeed, biological measurements. In most countries, expertise in chemical metrology is also more widely dispersed than is the case for physical measurements, which are mainly focused on a single national measurement institute. Traceable measurements also require that the uncertainty of the entire chemical measurement procedure is fully understood. The uncertainty of the sample preparation and pretreatment is, however, largely an empirical estimate and the uncertainty associated with taking the initial sample or subsample is often overlooked. Developing reference methods that offer improved and rigorously determined levels of uncertainty for difficult sample matrices is a key factor in solving this problem.

In order to address both the technical and organizational problems, the CIPM (see **Figure 1**) decided in 1993 to establish an international, collaborative program of work in chemistry. This program is organized through the CIPM's Comité Consultatif pour la Quantité de Matière (CCQM) – the committee for metrology in chemistry. The CCQM's task is to resolve the practical difficulties of achieving comparable chemical measurements through traceability and to provide an international structure of

chemical laboratories. These laboratories are signatories to the BIPM MRA (see above) and are required to demonstrate the equivalence of their measurement data through measurement comparisons as well as implementing a quality management system for their calibration or measurement certificates. Once this has been done, claims may be submitted by each institute listing its measurement capabilities, i.e., standards or calibration services that it provides. These are vetted by peer review before inclusion in the BIPM Key Comparison Database. The CCQM has organized an ongoing series of key comparisons that reflect important applications relevant to industry, trade, health, and the environment, not just measurements on single-substance calibration standards. An example of some of the current CCQM entries in the BIPM database can be seen in Figure 4.

The results for one such Key Comparison, CCQM-K13, listed in Figure 4, are shown in Figure 5. This example illustrates a trace analysis application, the determination of lead in a sediment. The error bars

represent the expanded uncertainty reported by each participating national measurement institute. Bearing in mind that this is a difficult trace determination, these uncertainties and the degree of equivalence between the individual results (shown on the y-axis) illustrate how the participating institutes are working at the limits of current analytical methodology. This is necessary because, in order to underpin internationally traceable measurements, data from these laboratories form the ultimate reference points for calibration chains that end with routine measurements by field laboratories. Each calibration stage in such a chain introduces an increase in the uncertainty of the resulting measurement. Hence, unless the calibration chains can be kept very short, the national measurement institutes must achieve uncertainties substantially smaller than those needed by the field laboratories. As mentioned above, the BIPM Key Comparison Database provides access to the measurement capabilities of these institutes. Each such entry shows the dissemination range in terms of both the available concentration range and the

Key and supplementary comparisons

→ Search criteria: amount of substance, inorganics  
your request produced 11 result(s)

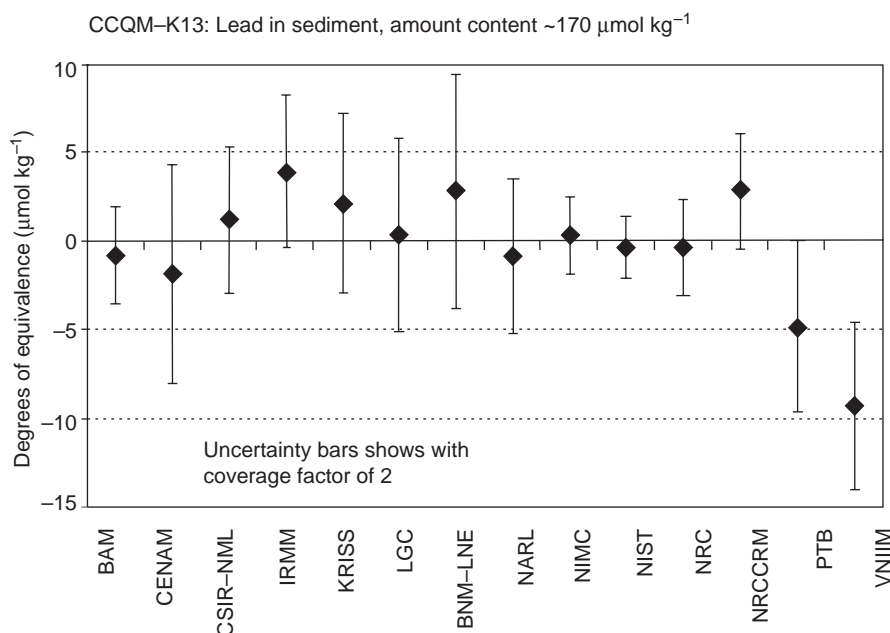
↙ List of comparisons

Click on a comparison identifier to view more

Page 1 2 3

<a href="#">CCQM-K2</a>	Cadmium and lead in natural water 1998
Comparison type, Field Status	Key comparison in amount of substance, inorganics Approved for equivalence, <a href="#">Results available</a>
<a href="#">CCQM-K8</a>	Monoelemental calibration solutions of Al, Cu, Fe, and Mg 1999–2000
Comparison type, Field Status	Key comparison in amount of substance, inorganics Approved for equivalence, <a href="#">Results available</a>
<a href="#">CCQM-K13</a>	Amount content of cadmium (Cd) and lead (Pb) in sediment 2000
Comparison type, Field Status	Key comparison in amount of substance, inorganics Approved for equivalence, <a href="#">Results available</a>
<a href="#">CCQM-K14</a>	Calcium in human serum 2003
Comparison type, Field Status	Key comparison in amount of substance, inorganics Measurements completed
<a href="#">CCQM-K24</a>	Cadmium (Cd) in rice 2001
Comparison type, Field Status	Key comparison in amount of substance, inorganics Approved for equivalence, <a href="#">Results available</a>

**Figure 4** Part of a table taken from the BIPM Key Comparison Database (KCDB) which may be viewed on the BIPM website ([www.bipm.org](http://www.bipm.org)). The table shows some of the chemistry key comparisons organized by the working groups of the CCQM. (Reproduced with permission from BIPM website.)



**Figure 5** Results of one of the key comparisons listed in the BIPM Key Comparison Database, CCQM-K13, which concerns the determination of lead in sediment samples at trace levels. This information may be viewed on the BIPM website ([www.bipm.org](http://www.bipm.org)). (Reproduced with permission from BIPM website.)

#### Calibration and measurement capabilities

Sediments, soils, ores, and particulates

United Kingdom, LGC (Laboratory of the Government Chemist)

Complete CMCs in Amount of substance for sediments, soils, ores, and particulates for United Kingdom (.pdf file)

Matrix or material	Analyte or component	Dissemination range of measurement capability	
		Mass fraction in $\mu\text{g per g}$	Relative expanded uncertainty in %
Sediment	Lead	30–90	2–3

**Figure 6** An example of a measurement capability (CMC) submitted for international peer review and accepted for inclusion in the BIPM Key Comparison Database. This example shows a relevant CMC for one of the institutes that participated in CCQM-K13 (illustrated in **Figure 5**). The database presently comprises several thousand CMCs, including a wide range of chemical measurements. (Reproduced with permission from BIPM website.)

corresponding range of measurement uncertainty. A typical CMC entry from one institute, for the capability tested by CCQM-K13, is shown in **Figure 6**.

## Traceability and Measurement Uncertainty

It should be clear from the above discussion that, regardless of the availability of standards, reference materials or calibration services having reference values exhibiting international traceability, the results dependent on them will not be traceable unless field laboratories are able to obtain reliable estimates of their own measurement uncertainty. This is, in fact, a requirement for accreditation to ISO 17025 but the practical difficulties of achieving it for

chemical measurements are considerable. It is well known that the uncertainties of analytical results are often large because many analytical methods require several quite complex operations (e.g., extraction, clean-up, or preconcentration) prior to the final measurement using a calibrated instrument. The uncertainty associated with these stages is frequently difficult, or sometimes impossible, to quantify and the analyst must resort to an estimate based on the available data and previous experience.

Nevertheless, it is essential that laboratories strive to achieve the best possible estimate of uncertainty at the method validation stage. A key factor in this regard is the availability of appropriate matrix reference materials. In order to achieve routine results that are both reliable and traceable, it is essential that

validation is based on materials that are both similar to the routine samples and have reference values with substantially smaller uncertainties than is required for routine data. Of equal importance is developing a base of knowledge and expertise that can help laboratories to resolve some of the problems encountered in evaluating the uncertainties of analytical results. Achieving these goals will not happen overnight but adopting the twin principles of measurement traceability and uncertainty offers analytical scientists a route map by which they may be achieved.

## Acknowledgments

The diagrams illustrating the international measurement system and the examples from the BIPM Key Comparison Database were provided by the Bureau International des Poids et Mesures, Sèvres, France, and are reproduced with their permission.

See also: **Quality Assurance:** Reference Materials; Method Validation.

## Further Reading

- EN-ISO/IEC 17025:2000 (2000) *Requirements for the Competence of Testing and Calibration Laboratories*. Geneva: International Standards Organisation.
- EURACHEM (1998) *The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and*

*Related Topics*. Available from the EURACHEM website (<http://www.eurachem.ul.pt>).

EURACHEM (2000) *Quantifying Uncertainty in Analytical Measurements*, 2nd edn. Available from the EURACHEM website (<http://www.eurachem.ul.pt>).

EURACHEM (2002) *The Selection and Use of Reference Materials; A Basic Guide for Laboratories and Accreditation Bodies*, EEE/RM/062. Available from the EURACHEM website (<http://www.eurachem.ul.pt>).

EURACHEM/CITAC (2002) *Guide to Quality in Analytical Chemistry*. Available from the EURACHEM (<http://www.eurachem.ul.pt>) or CITAC (<http://www.citac.cc>) websites.

EURACHEM/CITAC (2003) *Traceability in Chemical Measurement, A Guide to Achieving Comparable Results in Chemical Measurement*. Available from the Eurachem website (<http://www.eurachem.ul.pt>).

International Organization for Standardization (ISO) (1995) *Guide to the Expression of Uncertainty in Measurement*. Geneva: ISO.

ISO (1993) *International Vocabulary of Basic and General Terms in Metrology*. Geneva: International Standards Organization.

McNaught AD and Wilkinson A (1997) *IUPAC Compendium of Chemical Terminology*, 2nd edn. Cambridge, UK: Royal Society of Chemistry.

Stoeppler M, Wolf WR, and Jenks PJ (2000) *Reference Materials for Chemical Analysis*. New York: Wiley-Interscience.

Taylor JK (1993) *Handbook for SRM Users*, NBS Special Publication 260-100. Gaithersburg: National Institute of Standards and Technology.

## Accreditation

**E J Newman**, Newman, Bucknell Associates, Wimborne, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

One indication of the importance of analytical chemistry is that it has been estimated that ~3% of the gross domestic product of the more advanced industrial nations is spent on analytical testing. Other indications of importance come from major areas of application of analytical science, such as: food safety, drinking water quality, animal feeds, fertilizers and pesticides; air and coastal water quality monitoring, workplace health and safety monitoring, and applications to the environment generally. Also applications in the field of health care including the quality control (QC) of pharmaceuticals, and clinical analyses for both diagnosis and to monitor the effects of

therapy; the wide area of forensic science in which analytical chemistry is a core subject; and the many and varied important industries not included in the above, in which analysis provides vital QC, and which range from the long-established, such as steel-making or dyeing, to the modern manufactures of microprocessors, optical fibers, and other advanced materials. To meet all these and many other demands, analysts produce, as a conservative estimate, over a billion items of analytical data every year in the United Kingdom alone.

It is clear from this that analytical data provide the basis for thousands of important, even critical, decisions every day of the year. Furthermore, the importance of chemical analysis is increasing as legislators, considering scientific evidence, much of which was itself derived from analytical chemistry, continue to frame regulatory requirements, which can be enforced only by the endeavors of skilled analysts.



validation is based on materials that are both similar to the routine samples and have reference values with substantially smaller uncertainties than is required for routine data. Of equal importance is developing a base of knowledge and expertise that can help laboratories to resolve some of the problems encountered in evaluating the uncertainties of analytical results. Achieving these goals will not happen overnight but adopting the twin principles of measurement traceability and uncertainty offers analytical scientists a route map by which they may be achieved.

## Acknowledgments

The diagrams illustrating the international measurement system and the examples from the BIPM Key Comparison Database were provided by the Bureau International des Poids et Mesures, Sèvres, France, and are reproduced with their permission.

See also: **Quality Assurance:** Reference Materials; Method Validation.

## Further Reading

- EN-ISO/IEC 17025:2000 (2000) *Requirements for the Competence of Testing and Calibration Laboratories*. Geneva: International Standards Organisation.
- EURACHEM (1998) *The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and*

*Related Topics*. Available from the EURACHEM website (<http://www.eurachem.ul.pt>).

EURACHEM (2000) *Quantifying Uncertainty in Analytical Measurements*, 2nd edn. Available from the EURACHEM website (<http://www.eurachem.ul.pt>).

EURACHEM (2002) *The Selection and Use of Reference Materials; A Basic Guide for Laboratories and Accreditation Bodies*, EEE/RM/062. Available from the EURACHEM website (<http://www.eurachem.ul.pt>).

EURACHEM/CITAC (2002) *Guide to Quality in Analytical Chemistry*. Available from the EURACHEM (<http://www.eurachem.ul.pt>) or CITAC (<http://www.citac.cc>) websites.

EURACHEM/CITAC (2003) *Traceability in Chemical Measurement, A Guide to Achieving Comparable Results in Chemical Measurement*. Available from the Eurachem website (<http://www.eurachem.ul.pt>).

International Organization for Standardization (ISO) (1995) *Guide to the Expression of Uncertainty in Measurement*. Geneva: ISO.

ISO (1993) *International Vocabulary of Basic and General Terms in Metrology*. Geneva: International Standards Organization.

McNaught AD and Wilkinson A (1997) *IUPAC Compendium of Chemical Terminology*, 2nd edn. Cambridge, UK: Royal Society of Chemistry.

Stoeppler M, Wolf WR, and Jenks PJ (2000) *Reference Materials for Chemical Analysis*. New York: Wiley-Interscience.

Taylor JK (1993) *Handbook for SRM Users*, NBS Special Publication 260-100. Gaithersburg: National Institute of Standards and Technology.

## Accreditation

**E J Newman**, Newman, Bucknell Associates, Wimborne, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

One indication of the importance of analytical chemistry is that it has been estimated that ~3% of the gross domestic product of the more advanced industrial nations is spent on analytical testing. Other indications of importance come from major areas of application of analytical science, such as: food safety, drinking water quality, animal feeds, fertilizers and pesticides; air and coastal water quality monitoring, workplace health and safety monitoring, and applications to the environment generally. Also applications in the field of health care including the quality control (QC) of pharmaceuticals, and clinical analyses for both diagnosis and to monitor the effects of

therapy; the wide area of forensic science in which analytical chemistry is a core subject; and the many and varied important industries not included in the above, in which analysis provides vital QC, and which range from the long-established, such as steel-making or dyeing, to the modern manufactures of microprocessors, optical fibers, and other advanced materials. To meet all these and many other demands, analysts produce, as a conservative estimate, over a billion items of analytical data every year in the United Kingdom alone.

It is clear from this that analytical data provide the basis for thousands of important, even critical, decisions every day of the year. Furthermore, the importance of chemical analysis is increasing as legislators, considering scientific evidence, much of which was itself derived from analytical chemistry, continue to frame regulatory requirements, which can be enforced only by the endeavors of skilled analysts.



With matters of public health and safety, the reliability of manufactured products, the protection of the environment, and even the freedom of individuals depending on their results, analysts ought not to be surprised that those who pay for, use, or are affected by their data seek reassurance that the data are both reliable and fit for their intended purposes.

Analytical chemists recognize their responsibilities for the data they report and by their training and experience come to recognize many of the likeliest sources of significant errors in their methods. However, the point must be made that as analytical equipment becomes more automatic, instrumental, and computerized, the analyst must increasingly rely on indirect evidence that his analyses are proceeding as planned. Similarly, when quantities of sample and analyte are too small to be visible, the analyst has to depend on instrumental settings and readouts as being reliable indicators of experimental conditions. These difficulties have arisen during the same period that universities and training colleges have been obliged to provide an ever-broader range of practical experience at some sacrifice to the depth of understanding, whilst the problems that analysts are called upon to tackle have become increasingly complex.

The traditional approaches to assure the quality of analytical data involved careful standardization of reagents, accurate measurements of volumes, working to standard methods, replications of analyses, 'blind' analyses, and the use of reference materials. Other methods included participation in cooperative or 'round robin' analysis schemes with laboratories doing similar work, the witnessing of analysis, and verification and even certification of analysis by an independent laboratory. All these approaches are still in use along with other variants of QC, but it is now widely recognized that QC alone does not provide a high level of confidence without incurring very large costs, and that another rational approach is to ensure that all analytical and supporting staff are fully trained in and follow written instructions, and maintain appropriate records, of all the stages in the analytical process from the arrival of the sample in the laboratory to the reporting of the findings to the customer. A complete set of written procedures, training, practical activities, and records are known as the 'quality system'. Quality systems have been published as standards for companies, such as the widely known BS EN ISO 9000: 2000 series and Good Manufacturing Practice (GMP). Quality standards for laboratories include BS EN ISO 17025: 2000, 'General requirements for the competence of testing and calibration laboratories'; Good Laboratory Practice (GLP), which applies to laboratories involved in any testing of materials for toxicological

purposes; and Clinical Pathology Accreditation, specifically for the accreditation of pathology laboratories. The ISO 17025 standard is the most widely applied and is the main focus of this article.

One other development in support of analytical data quality must be mentioned, namely the use of 'round robin' analysis, which has developed into a number of laboratory performance assessment programs, and also known as proficiency-testing schemes. A proficiency-testing scheme does not eliminate the need for a quality system. In fact, the scheme can at best provide only a 'snapshot' of how a laboratory was performing on particular analyses on one occasion. Moreover, such schemes still exist only for a fairly limited range of sample and analyte combinations.

## **Analytical Quality Assurance**

The control and assurance of the quality of analytical data are important features of laboratory management. Since the great majority of analytical laboratories exist to provide services to customers in the same or other organizations, the appropriate definitions of quality are 'meeting customer requirements' and 'being fit for the intended purpose'. These do not wholly describe the situation of the analyst. Thus, it cannot be assumed that all customers really want unbiased analytical results when biased data might provide commercial advantage. Many analytical laboratories, especially those providing contract services to a wide range of industries, are not always told by their customers what their requirements are, and they might not know themselves. The best that the analyst can do in such situations is to exercise professional judgment based on experience and whatever knowledge they have of the industry concerned.

The term 'quality control' is used differently around the world. In the Far East, particularly Japan, it refers to what the Western industrial nations call quality assurance (QA), which is discussed shortly. QC in the West refers to the process of monitoring the product to establish whether or not it conforms to specified requirements, for example to demonstrate that a manufactured batch of a chemical used as a pharmaceutical raw material meets all the requirements of the specification published in the British Pharmacopoeia. In fact, this is not true QC, other than to prevent the release and use of substandard product. It only assesses quality, after the process is complete, and this is both costly and of limited operational use. In the analytical laboratory, QC refers to the analyses conducted along with a batch of samples, the results of which are used to judge whether the analyses were made correctly. For this

purpose, the QC sample may be made using previously or independently analyzed samples, reference substances, spiked samples, proficiency-testing samples, or simply replicates used to estimate precision. The criticism made above is valid; the QC results give data after the event. Moreover, unless the QC samples are 'blind' and introduced randomly, the information yielded by them could be biased.

The approach of QA is not, as sometimes stated, to replace QC, and neither is it an alternative name for QC, although some confusion exists about this. QA seeks to implement a system (the quality system) for managing all the operations and other features of a process that are known to influence quality, and QC forms part of that system. In the laboratory, this means not only the analytical methodology but all the other inputs, from the original customer instruction to sample storage and handling, the equipment and reagents used, and any calibrations made on them, all documentation, the environment in which the work was performed, the training and experience of the analyst, and finally, because the product is data, reporting the analyst's findings to the client in the most appropriate form.

Most laboratories have, in practice, developed their own approaches to handling these aspects of QA, but often they are not complete for various reasons. One purpose of the several standards mentioned in the section 'Introduction' is to provide complete systems. These and other successful private, official, and industry-based standards for analytical QA systems address a number of common features including:

- aspects of authorities and responsibilities within the laboratory;
- a description of how the quality system operates;
- staff qualifications, experience, and training;
- documentation and document control;
- equipment operation and maintenance;
- calibration and measurement traceability;
- quality control;
- sample storage and handling;
- review of customer requirements;
- accommodation and environment within the laboratory;
- record keeping;
- dealing with complaints and anomalies;
- internal audits and reviews of the system; and
- corrective action.

Both QC and QA have their advocates and, in fact, it is recognized that the delivery of analytical data quality requires first that the laboratory has the capability to perform satisfactorily on the type of analysis under consideration and second that it was

actually doing so for the analysis in question. Accreditation mainly addresses the first of these requirements, the control of capability: it is an independent confirmation that the laboratory has the ability to provide quality data. Analytical QC addresses the second requirement.

The standards for QA systems all emphasize the importance of well-trained staff and in order to be able to demonstrate the laboratory's commitment to training to assessors, for the purposes of accreditation, adequate training records must be maintained. The standards also emphasize the importance of using documentation to avoid ambiguities and communication problems, and to ensure that all staff are trained and working to the same procedures.

A third essential of any quality system is that it must be subject to regular internal audits and that all nonconformances to the standard use must be corrected as soon as possible. Audits ensure that all the components of the system itself are monitored and that not only is the standard maintained but that the system is gradually improved. It is a truism that no quality system can be said to exist unless it incorporates auditing. These three features: training, documented procedures, and audits are vital for quality systems for all situations. They are not independent, of course, but interact. For example, the documentation of training requirements for many laboratory operations will preclude the need to repeat instructions for these operations in standard operating procedures. Also, in auditing, the auditors must be specially trained staff who are normally analysts, and they must perform their duties according to documented procedures.

The responsibility for the achievement and maintenance of analytical quality rests on all the staff at all levels in the organization, and it needs to be stressed that the first line of management, the section leader or supervisor in charge of practical analysis has a key role to play. The section leader continually witnesses and monitors staff performance and identifies training needs, arranges for training to be provided when required, ensures procedures are followed, calibration and standardization routines are observed, reagents and equipment are maintained, that calculation checks are made, and records are properly kept. These are the basic disciplines of a quality system and they can only be consistently followed when there is a high standard of supervision at the bench.

## Accreditation in the United Kingdom

Accreditation of a laboratory means the official assessment and certification of its quality system as meeting the requirements of such a standard.

Assessment of an analytical laboratory quality system may be made by the laboratory itself. It will have little or no value in the eyes of many customers, however, since it will be seen as a subjective and favorable assessment. However, internal audits and management reviews of the system, made as a requirement of a quality system standard, are self-assessments and are valuable, as explained in the previous section.

Accreditation, also known as registration or certification by some assessment bodies, may be second- or third-party. Second-party accreditation is made by the customer's own QA department either to their own in-house standard or to an independent standard such as ISO 9001.

An independent body, having no other commercial interest in the organization it accredits, conducts third-party accreditation. In the United Kingdom, a testing (or calibration) laboratory would be accredited to ISO 17025 by the United Kingdom Accreditation Service (UKAS).

Assessment of a testing laboratory to the ISO 17025 standard is conducted against a schedule, which is a matrix of sample types tested, determinants measured, and the methods used. The laboratory may perform a much wider range of tests than those for which it seeks accreditation, but these will not be included in the assessment. The assessment team will include experienced analysts to ensure that the latest or most appropriate methods of test are used, that method validation studies have been made and recorded, and that appropriate QC is being conducted, including participation in proficiency-testing schemes where available. They also ensure that instruments are used and calibrated to meet the demands of the test methods. Another peer review aspect is the witnessing by assessors of a selection of the tests from the schedule. The assessors also ensure, of course, that all the other requirements of ISO 17025 are met (Table 1).

Once a laboratory has implemented a quality system in accordance with these requirements, it must make a formal application to UKAS to have its system assessed for the purpose of accreditation and at this stage submits its Quality Manual and its proposed schedule of accreditation for examination, along with any other documentation requested by UKAS. The manual describes the policies and practices through which the laboratory addresses the requirements of the standard. Provided UKAS is reasonably satisfied with the documentation received, and once the laboratory has been advised of any areas needing further attention before assessment, the arrangements for the assessment visit are made. It takes the form of a major assessment/audit

**Table 1** Main features of the ISO 17025 standard

<i>Management requirements</i>
Organization
Quality system
Document control
Review of requests, tenders, and contracts
Subcontracting of tests and calibrations
Purchasing services and supplies
Service to the client
Complaints
Control of nonconforming testing and/or calibration work
Corrective action
Preventive action
Control of records
Internal audits
Management reviews
<i>Technical requirements</i>
General (a consideration of factors affecting correctness and reliability of data)
Personnel
Accommodation and environmental conditions
Test and calibration methods and method validation
Equipment
Measurement uncertainty
Sampling
Handling of test and calibration items
Assuring the quality of test and calibration results
Reporting the results

The processes of assessment and registration of a quality system to ISO 9000 standards or of assessment and certification of compliance with the GLP Regulations are similar.

and review of the quality system by the assessment team who, at the conclusion, will report that they will recommend to the UKAS Executive that accreditation be offered or will describe what further action must be taken by the laboratory to achieve the standard. Further formalities include an agreement on the contents of the schedule, which is issued by UKAS, and the laboratory making a written undertaking to comply with regulations concerned with the use of the UKAS logo. The process does not end here: the laboratory will be subject to a surveillance assessment after 6 months, and then annually, and after 4 years the whole process of full assessment must be repeated for reaccreditation.

For completeness, it should be mentioned that laboratories providing QC and analytical support to pharmaceutical production, and also to the production of medical devices and diagnostic materials, are subject to assessment to ensure they comply with the relevant requirements of the Guide to GMP. These requirements are very similar to those found in the other standards. The laboratories are not accredited, but compliance with requirements is a condition of granting and maintaining a manufacturing license for the products concerned.

Assessments for both GMP and GLP are made by staff from the Medicines and Healthcare products Regulatory Agency (MHRA).

## Accreditation in Other Countries

Within the European Community (EC) the breaking of trade barriers and the need to overcome various attitudes between nations that were once rivals or worse, has led, in the analytical community, to a great amount of interest in the application and harmonization of standards. Some nations are well ahead in these processes, one of which is the United Kingdom. Others less advanced in the adoption of quality systems are now making progress particularly as EC legislation has obliged the adoption of GMP in several fields and is now beginning to specify the use of ISO 17025.

Much progress has been made toward the mutual recognition of accreditation and UKAS now has many agreements with its counterparts in other European countries and also with some organizations outside Europe. There is a forum for the discussion of accreditation harmonization known as the Western European Laboratory Accreditation Co-operation (WELAC). Another collaborative organization for the resolution of problems of mutual interest in analytical chemistry throughout the whole of Europe, known as EURACHEM, has focused its attention to date mainly on matters of QA, and together these two bodies have published a very useful booklet giving specific guidance on accreditation, now available as a UKAS publication.

However, there is no doubt that there are different perceptions about QA even within the Western European countries. Some chemists hold strongly to the view that QA is best managed through quality systems and accreditation. Others are equally committed to having long and careful professional training combined with participation in proficiency schemes.

Accreditation has long been established in Australia and New Zealand using standards similar to ISO 17025, and many other countries have now set up or are in the process of setting up organizations working to similar standards and principles.

Second-party accreditation is used widely in the United States, with many major companies, such as Ford, and official bodies, such as the US Environmental Protection Agency (EPA) acting as key players. The best-known public sector QA requirements are those of the US Food and Drug Administration (FDA) for registration and compliance with GLP and GMP, and the EPA Contract Laboratory Program (CLP). The FDA requires manufacturers

and importers of medical products to register with them and comply with a GMP regulation, which the FDA has a legal obligation to inspect and enforce. The FDA GMP regulation was developed in accordance with the basic principles of QA, which have as their goal the production of articles that are fit for their intended uses. The GMP requirements are similar to those of other quality standards and the inspections are conducted along the same lines as other assessments, but in great depth, usually applied to all aspects of QA of one or a few products.

The EPA's Contract Laboratory Program was designed to support the EPA's 'Superfund' program by providing analytical data of a known and documented quality. The major objective of CLP is to provide access to a range of analytical services on a high volume cost-effective basis. The CLP uses a highly structured, inflexible approach to analytical procedures to provide legally admissible evidence for use in EPA enforcement efforts. Before a laboratory is awarded a contract, it must be evaluated and meet a number of QA requirements that are largely similar to those already described under UKAS assessment, and which include a satisfactory performance on a performance evaluation sample. When the contract has been awarded, the laboratory must undergo a number of postaward evaluations including surveillance inspections and periodic performance evaluation testing.

Third-party accreditation is now increasing in popularity. Some laboratories now have quality systems registered to ISO 9000 series standards, and third-party accreditation for laboratories based on ISO 17025 has now become widely used through the National Voluntary Laboratory Accreditation Program (NVLAP) administered by the National Institute of Standards and Technology and the A2LA scheme of the American Association for Laboratory Accreditation.

## The Future of Accreditation

Laboratory quality systems and related quality activities such as participation in proficiency-testing programs are being adopted increasingly. They have existed in the United Kingdom, certain other Western European nations, Canada, Australia, New Zealand, and the United States for many years. They are coming into increasing use in India and the Far East, South Africa, and Eastern Europe.

Commercial pressures and legislation are both certain to increase the requirement for third-party assessment and registration of laboratory QA arrangements, whether to ISO 17025, ISO 9000, GMP, or GLP.



The need for harmonization of standards, where it is considered necessary, and mutual recognition of registration and accreditation schemes, is of increasing concern and importance to trade. This will inevitably result in progress toward a unified system of recognition of competent laboratories. However, difficulties between major groups of trading nations, and different perceptions of analytical quality and how it is achieved and maintained, will act to delay the process. It is therefore unlikely that true international agreement will be realized for some years yet, although it must always be an objective to pursue. It might be that such agreement could be possible only if analytical chemists in all countries are trained to a common core syllabus, which includes a significant content of analytical QA principles.

*See also:* **Quality Assurance:** Quality Control; Instrument Calibration; Interlaboratory Studies; Reference Materials; Production of Reference Materials; Laboratory Information Management Systems.

## Further Reading

*Accreditation for Chemical Laboratories* (UKAS Publication ref: LAB27). United Kingdom Accreditation Service, 2000.

Dempsey CH and Petty JD (1991) *Laboratory Accreditation and Data Certification: A System for Success*. Chelsea, MI: Lewis Publishers, Inc.

Dux JP (1986) *Handbook of Quality Assurance for the Analytical Chemistry Laboratory*. New York: Van Nostrand Reinhold.

Garfield FM (1984) *Quality Assurance Principles for Analytical Laboratories*. Arlington, VA: Association of Official Analytical Chemists, Inc.

*Guide to the UK GLP Regulations 1999* (2000). Department of Health; The United Kingdom Good Laboratory Practice Monitoring Authority.

ISO 9001 (2000) *Quality Management Systems – Requirements*. Geneva: ISO.

ISO/IEC 17025 (1999) *General Requirements for the Competence of Testing and Calibration Laboratories*. Geneva: ISO.

Laboratory of the Government Chemist (1993) *The VAM Bulletin*, issue 10, Accreditation Special, Autumn 1993.

Miller JM and Crowther SB (eds.) (2000) *Analytical Chemistry in a GMP Environment*. New York: Wiley.

*Rules and Guidance for Pharmaceutical Manufacturers and Distributors 1997* (2002) Medicines Control Agency (HMSO).

*The OECD Principles of Good Laboratory Practice* (1982) The Organisation for Economic Co-operation and Development.

# Laboratory Information Management Systems

**R D McDowall**, McDowall Consulting, Bromley, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## The Analytical Laboratory

### Role of an Analytical Laboratory

The role of an analytical laboratory within most organizations is to provide information upon which decisions can be made. A laboratory can be responsible for:

- process analysis of intermediates and bulk products;
- quality control of raw materials and/or finished products;
- research and development of new products.

The laboratory can be located within an industrial, medical, or academic environment, but regardless of the location, it still has the same overall responsibility. Therefore, it is crucial that as well as the need for information, the information must be produced in a timely manner with appropriate quality to allow effective decision-making.

While samples enter a laboratory where they are analyzed, it is rare for the information to be used there; the decision-makers usually reside outside of a laboratory. Therefore, a laboratory should integrate with its client departments to become an efficient part of an organization or enterprise. One element of this integration is to take the laboratory to the sample or process rather than the sample to the laboratory. When taking the laboratory to sample the analytical chemist can use conventional equipment or use some of the processes and techniques outlined in this article such as electronic noses and tongues, membranes, sensors of all descriptions, and continuous flow systems.

Regardless of the analytical system used, all it will do is to generate data. Other software applications will be needed to reduce the data to information and then to disseminate it to enable decisions to be taken.

### Need to Increase Productivity

Many laboratories are situated in industries that operate under regulatory guidelines, such as Good Manufacturing Practice (GMP) or Good Laboratory

The need for harmonization of standards, where it is considered necessary, and mutual recognition of registration and accreditation schemes, is of increasing concern and importance to trade. This will inevitably result in progress toward a unified system of recognition of competent laboratories. However, difficulties between major groups of trading nations, and different perceptions of analytical quality and how it is achieved and maintained, will act to delay the process. It is therefore unlikely that true international agreement will be realized for some years yet, although it must always be an objective to pursue. It might be that such agreement could be possible only if analytical chemists in all countries are trained to a common core syllabus, which includes a significant content of analytical QA principles.

*See also:* **Quality Assurance:** Quality Control; Instrument Calibration; Interlaboratory Studies; Reference Materials; Production of Reference Materials; Laboratory Information Management Systems.

## Further Reading

*Accreditation for Chemical Laboratories* (UKAS Publication ref: LAB27). United Kingdom Accreditation Service, 2000.

Dempsey CH and Petty JD (1991) *Laboratory Accreditation and Data Certification: A System for Success*. Chelsea, MI: Lewis Publishers, Inc.

Dux JP (1986) *Handbook of Quality Assurance for the Analytical Chemistry Laboratory*. New York: Van Nostrand Reinhold.

Garfield FM (1984) *Quality Assurance Principles for Analytical Laboratories*. Arlington, VA: Association of Official Analytical Chemists, Inc.

*Guide to the UK GLP Regulations 1999* (2000). Department of Health; The United Kingdom Good Laboratory Practice Monitoring Authority.

ISO 9001 (2000) *Quality Management Systems – Requirements*. Geneva: ISO.

ISO/IEC 17025 (1999) *General Requirements for the Competence of Testing and Calibration Laboratories*. Geneva: ISO.

Laboratory of the Government Chemist (1993) *The VAM Bulletin*, issue 10, Accreditation Special, Autumn 1993.

Miller JM and Crowther SB (eds.) (2000) *Analytical Chemistry in a GMP Environment*. New York: Wiley.

*Rules and Guidance for Pharmaceutical Manufacturers and Distributors 1997* (2002) Medicines Control Agency (HMSO).

*The OECD Principles of Good Laboratory Practice* (1982) The Organisation for Economic Co-operation and Development.

# Laboratory Information Management Systems

**R D McDowall**, McDowall Consulting, Bromley, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## The Analytical Laboratory

### Role of an Analytical Laboratory

The role of an analytical laboratory within most organizations is to provide information upon which decisions can be made. A laboratory can be responsible for:

- process analysis of intermediates and bulk products;
- quality control of raw materials and/or finished products;
- research and development of new products.

The laboratory can be located within an industrial, medical, or academic environment, but regardless of the location, it still has the same overall responsibility. Therefore, it is crucial that as well as the need for information, the information must be produced in a timely manner with appropriate quality to allow effective decision-making.

While samples enter a laboratory where they are analyzed, it is rare for the information to be used there; the decision-makers usually reside outside of a laboratory. Therefore, a laboratory should integrate with its client departments to become an efficient part of an organization or enterprise. One element of this integration is to take the laboratory to the sample or process rather than the sample to the laboratory. When taking the laboratory to sample the analytical chemist can use conventional equipment or use some of the processes and techniques outlined in this article such as electronic noses and tongues, membranes, sensors of all descriptions, and continuous flow systems.

Regardless of the analytical system used, all it will do is to generate data. Other software applications will be needed to reduce the data to information and then to disseminate it to enable decisions to be taken.

### Need to Increase Productivity

Many laboratories are situated in industries that operate under regulatory guidelines, such as Good Manufacturing Practice (GMP) or Good Laboratory



Practice (GLP) in the pharmaceutical industry, or voluntary quality schemes such as ISO 17025. This can result in a growing number of samples for analysis with stricter control of administrative procedures as a means of ensuring quality. Other laboratories operate in companies and industries where there is a need to contain costs whilst maintaining analytical information to make decisions. To overcome these issues, laboratories need to increase productivity as a means of curtailing rising staff numbers and/or labor costs.

If productivity increases are sought, then there are two possible approaches. The first approach is to change the process and the accompanying procedures so that some tasks are simplified or even eliminated; as such this may not involve the use of automation. However, the viability of this approach is subject to question unless labor costs are reduced dramatically. Therefore, the second approach is to use laboratory automation.

### The Aims of Laboratory Automation

The use of automation within an organization is an opportunity to create strategic advantage, which is manifested in one or more of the following factors:

- Better decisions based on valid data. This can lead to research that is more effective, faster time to market for new products, or to better products that due to their quality can increase customer satisfaction and market share.
- Improved organizational productivity.
- Reduced risk to employees.

Bringing these aims down to an analytical laboratory level, the automation of a task should produce one or more of the following advantages that will contribute to the aims above:

- Greater increases in productivity (either in numbers of samples assayed per unit time or speedier turnaround time); resulting in overall cost reduction and increased productivity.
- Automated systems that should be capable of the same or better precision and accuracy as existing manual methods with improved quality and reliability of measurement.
- The freeing of trained laboratory staff to do more creative or productive work than tedious tasks thus improving morale.
- Reduced human contact with biological or chemical hazards.
- Lower the consumption of sample and/or reagents used in an automated analysis.

### Problems with Laboratory Automation

How well have laboratories fared in implementing systems in practice? From a historical perspective not very well, Betteridge has stated: "the development of automation in analytical chemistry can be viewed as a series of steps in which the solution of today's bottleneck creates another problem for the future to solve."

This was shown succinctly by the development of automation in analytical chemistry:

- Before the 1960s, the barrier to automation was actually generating the data; the solution was the development of electronic control of apparatus. When apparatus was controlled, the new problem became the large amounts of chart paper produced.
- Therefore, in the 1960s the problem became data acquisition; this was solved by digitizers, which acquired the data but produced a new problem of manipulation.
- This was solved in the 1970s with the introduction of minicomputers to reduce the data; but many computers created the problem of data transfer.
- Data management was solved in the 1980s by the use of workstations and a laboratory information management systems (LIMS) but the problem moved to become the management of technical data.
- The interpretation of these data either by expert systems or artificial intelligence is the current problem for the 1990s to solve.
- The integration of the different data systems and LIMS to design the paperless laboratory is the task of the first decade of twenty-first century.

Part of the historical problem was the lack of technology or instrumentation at the various times discussed to allow an all-encompassing solution. The other part of the problem was, and still is, the lack of vision of many chemists to see a laboratory as an integral part of an organization.

## Laboratory Information Management Systems

### What is a LIMS?

The major function of most analytical laboratories is the creation and presentation of information quickly to make decisions. A LIMS is one of the major laboratory automation tools at the disposal of analytical chemists to help achieve this aim. Although a LIMS does not undertake analysis, a LIMS can be pivotal in integrating both the laboratory operations and the laboratory itself within an efficient organization. A LIMS can provide a laboratory with the means to

automate the processes of information creation and presentation, as well as being the platform for information dissemination to clients and senior management.

However, this is not always the case. A large number of systems failed to meet initial expectations that implied that LIMS are not fully understood. To help overcome this, a LIMS model was proposed that enabled the requirements of a system to be visualized conceptually. It has been adopted, modified, and used as the basis of the LIMS concept model in the ASTM LIMS guide. However, the LIMS model focuses on identifying and visualizing the user functions within the laboratory environment and less on the strategic positioning of a system.

To overcome this problem, a paper by McDowall proposed a matrix for the development of a LIMS with a strategic focus. It introduced the concept of an operational, logistic, and strategic LIMS. The matrix is formed by plotting these three types of LIMS versus the scope of laboratory and organizational tasks that can be undertaken by such a system.

### A LIMS has Two Targets

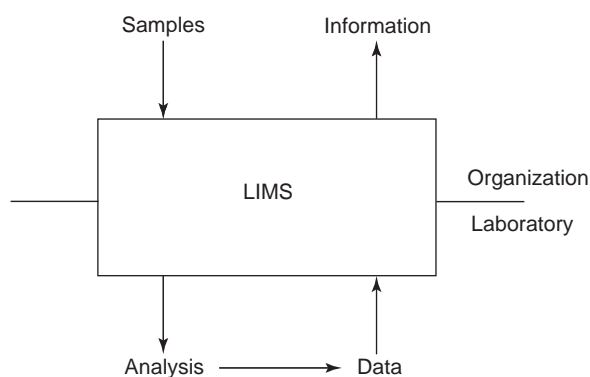
A LIMS is unlike any other piece of laboratory automation equipment available to the analytical chemist. It can provide benefits both within the laboratory and outside it. Thus, a LIMS has two targets:

- the laboratory: the information generator and
- the organization: the information user.

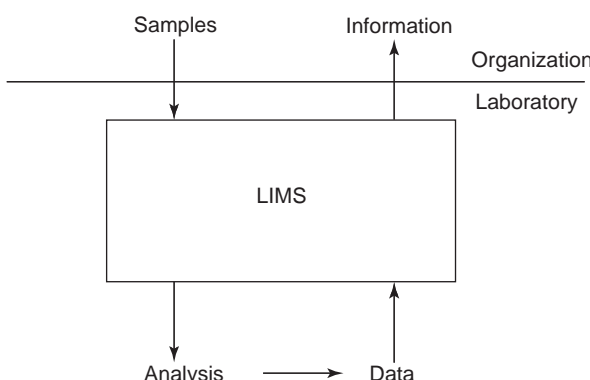
The problem is how to site and implement a system so that it hits both targets effectively.

Figure 1 shows an outline of the functions that a LIMS should undertake; the diagram shows a LIMS sited at the interface between a laboratory and an organization. Samples are generated in the organization and logged into the LIMS, the samples are analyzed within the laboratory, and data are produced and reduced within the LIMS environment to information that is transmitted back into the organization. Figure 1 represents the ideal setting of the LIMS: both the organization and the laboratory benefit. The line dividing the organization and the laboratory show the system is of equal benefit to both.

However, there are two other implementations that are possible with a LIMS, which result in different position of the interface between the laboratory and organization. Figure 2 shows the more common implementation, which is probably typical of the majority of early implementations of LIMS in the 1980s. The main functions carried out by the system are the same as that discussed previously but the emphasis of the implementation is different. The



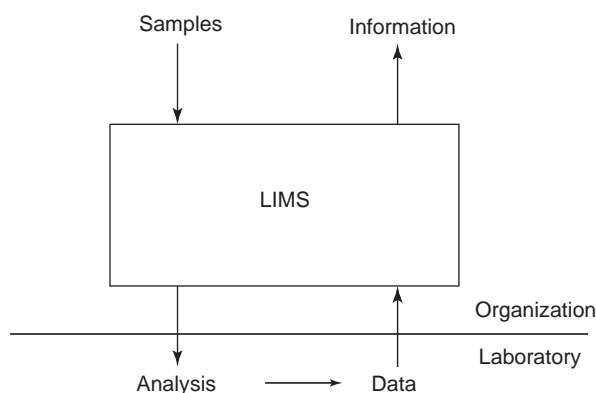
**Figure 1** Ideal implementation of a LIMS. The interface between the laboratory and the organization shows that the LIMS benefits both.



**Figure 2** A bottom-up implementation of a LIMS. The laboratory benefits but the organization does not.

boundary between the organization and the laboratory has been moved up and the benefit of the LIMS is almost exclusively that of the laboratory with little payback for the organization. Here, the LIMS is a toy for the laboratory that few others are allowed to play with. The system is built from the bottom up but with no consideration for anyone outside of the laboratory.

The rarest alternative implementation is presented in Figure 3. This is the top-down approach, where senior management or, worse, the information technology group has decided that a LIMS will be implemented. There has been little consideration for the laboratory, only the organization. The analysis and data gathering functions of a LIMS have been ignored, which allows the staff the latitude and the excuse to develop their own alternative local processing solutions. This system requires additional work by the staff to ensure its success in addition to the normal analytical function. The likelihood of failure with such a system is much higher than with the other two forms of implementation.



**Figure 3** A top-down implementation of a LIMS. The emphasis is on the overall system benefiting the organization but the implementation allows the laboratory to develop data analysis solutions independently of the LIMS.

As can be seen when comparing the three alternatives, there is a balance to be found between the needs of the organization and the laboratory. The interface between the two must be carefully defined; however, the initial implementation should be toward the analytical laboratory, the information generator. Automating the information generator is the key to success for the whole LIMS.

The importance of the LIMS matrix, described in the following section, allows that dividing line to be drawn accurately before starting the project and manage the implementation of a system.

### Construction of the LIMS Matrix

The matrix axes consist of LIMS type plotted against function in the laboratory or organization. Each will be outlined below.

**Three types of LIMS** There are three types of LIMS: operational LIMS, logistic LIMS, and strategic LIMS. The functions and aims of these systems are described in more detail.

**Operational LIMS** This is a basic system that automates analytical processes in the laboratory. This system increases the efficiency of a laboratory but the impact is only local. A computer system in this category would normally be concerned with the operation control of a laboratory and would automate functions such as sample entry, work list generation, and report preparation. Another way of looking at this system is usually the automation of the status quo. Such a system is probably typified in **Figure 2**.

**Logistic LIMS** In addition to the functions of an operational LIMS, a logistic LIMS provides users,

especially managers, with the information necessary to undertake their work. Such systems generate control information. Ultimately, this information improves the effectiveness of a laboratory.

**Strategic LIMS** The aims of a strategic LIMS are to integrate information and applications from different functional areas. From this information it may be possible to reshape operations. A strategic LIMS has the greatest impact upon the business by increasing the competitiveness of the laboratory.

Properly designed, a logistic or strategic LIMS will achieve the correct balance shown in **Figure 1**. However, too great an emphasis at the start of the project is likely to produce the system shown in **Figure 3**.

**The system scope** Where does a laboratory begin and end? Clear definition of these two points helps to determine the scope of a LIMS. The six areas of system scope that define the matrix are laboratory operations, monitor and control of operations, laboratory management, reporting and communications, analytical decision-making, and organizational integration. This is the maximum scope of a LIMS. These comprise the horizontal axis of the matrix. The function carried out by each of these areas is outlined in **Table 1**.

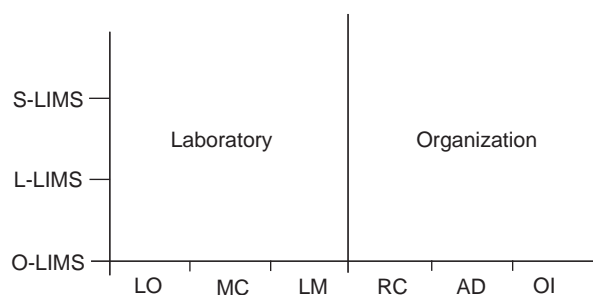
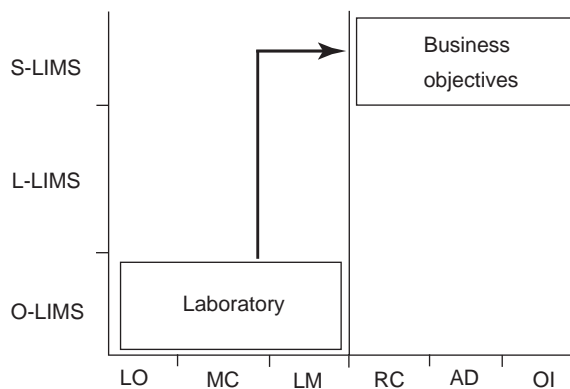
**Matrix Construction** The matrix can now be formed as a three by six block and is depicted in **Figure 4**. There is a division between the first three items of the system scope. Laboratory operation, monitor and control, and laboratory management are concerned with functions inside the laboratory. Organizational functions are represented as reporting and communications, analytical decision-making, and organizational integration.

The individual components of the matrix are shown in **Figure 5**; this is an outline of the functions of a LIMS within each cell.

Considering all of the functions in **Table 2**, one can see that the main functions of the laboratory are outlined in the lowest three cells of the matrix (operational LIMS: laboratory operations, monitor and control, and laboratory management). This represents the basic laboratory operation and is indicative of the data domain discussed previously. This is shown in **Figure 5** as the box labeled 'Laboratory'. Ideally, the major impact of any computer system, including a LIMS, should be strategic. The organization uses the information generated by the analytical laboratory and this can be considered as the information domain. The maximum impact of a LIMS is in the top right hand side of the matrix.

**Table 1** Functional scope of a LIMS

Functional area	Scope
Laboratory operations (LO)	<ul style="list-style-type: none"> <li>To automate and structure work</li> <li>The automation of the basic laboratory operations such as sample entry, work list generation, and results entry</li> <li>Work rationalization</li> </ul>
Monitor and control (MC)	<ul style="list-style-type: none"> <li>To evaluate performance</li> <li>The monitor and control of the laboratory operations by such processes as approving results, the use of quality control schemes, and the checking of transcription errors</li> <li>Provide standards, measures, and information for performance evaluation and feedback</li> </ul>
Laboratory management (LM)	<ul style="list-style-type: none"> <li>Support intellectual processes</li> <li>Organizing and managing the laboratory functions and operations</li> <li>Project and work planning</li> </ul>
Reporting and communications (RC)	<ul style="list-style-type: none"> <li>To augment human communication</li> <li>Here are the means to transmit results or reports and communicate with the clients of the laboratory</li> </ul>
Analytical decision-making (AD)	<ul style="list-style-type: none"> <li>To aid and speedup decision-making</li> <li>Providing quality information in a timely manner and the right format to make decisions</li> <li>Supporting processes in production, development, or research</li> </ul>
Organizational integration (OI)	<ul style="list-style-type: none"> <li>Facilitate intra and interorganization transactions</li> <li>Integrating with other functional groups in the corporation and between organizations</li> </ul>

**Figure 4** An outline of the LIMS matrix. S-LIMS, L-LIMS and O-LIMS are strategic, logistic, and operational versions of LIMS, respectively. The horizontal axis functions are laboratory operations (LO), monitor and control (MC), laboratory operations (LM), reporting and communications (RC), analytical decision-making (AD), and organizational integration (OI).**Figure 5** The LIMS matrix illustrating how some LIMS are implemented with the emphasis on laboratories and the overall business objectives where the best payback on the system investment would be obtained.

It consists of the strategic LIMS cells of results and communications, analytical decision-making, and organizational integration, and is marked 'Business Objectives' in **Figure 5**. It should be the aim of every LIMS to aim for the top right hand area of the matrix.

However, it should be stressed that not every LIMS will be developed to the full extent of the matrix. There will, inevitably, be some duplication with other computer applications that have been implemented or are being implemented within an organization. The aim, therefore, is to develop the matrix

around business objectives and a good financial justification to make the system pay for itself in a reasonable time period. One way to do this, and make the economical use of resources, is to link the implementation of the LIMS with existing applications to provide the functionality described in the matrix. Here, the LIMS was linked through a communications network to a document management software package that can be itself developed to produce computer assisted new drug applications (CANDAs) that will fulfill the requirements of the matrix.

**Table 2** The LIMS matrix

<i>Strategic LIMS</i>	Integrate with client operations	Compare requests with capacity	Resources based on workload	Integrate with document management software	Decision support system integration	EDI & CANDA Rapid commercialization
<i>Logistic LIMS</i>	Integrate laboratory operations	Automate regulatory compliance	Monitor test use	Electronic reports to clients	Highlight out-of-specification results	Remote online inquiry
<i>Operational LIMS</i>	Automate existing operations	Monitor and approve results	Sample status Billing	Paper reports to clients	Display results versus specification	Remote printing of results
	<i>Laboratory operations</i>	<i>Monitor and control</i>	<i>Laboratory management</i>	<i>Reporting and communications</i>	<i>Analytical decision-making</i>	<i>Client/lab e-mail</i> <i>Organizational integration</i>

### LIMS Matrix Views

Similar to the ability to use views to gain an insight into the contents of a database, the same concept can be used to gain an insight into the matrix and its impact. We will consider two main views in this section. The first is the vertical view through laboratory operations and a horizontal view across the logistic LIMS cells.

**Vertical view of laboratory operations** The vertical view through the laboratory operations cells of the matrix will see how this changes from the operational LIMS to the strategic LIMS. The lowest left hand cell of the matrix represents the core of the analytical laboratory; the inventory of assay methods and instrumental techniques. The operational LIMS would generally automate these processes and techniques. Business rules would be incorporated into the system such as who was allowed to create, modify, or delete methods within the LIMS, how many replicates were needed for each analysis. Samples would be logged in and work sheets would be produced for each analytical run. All functions are the usual LIMS functions within the analytical laboratory.

At the logistic LIMS level, the matrix is looking for integration of laboratory operations. This would be the linking of analytical instruments to the LIMS for electronic transfer of results, samples would be bar coded and would have their location monitored as they move through the laboratory. The various applications used for analyzing data and producing results would be integrated into the LIMS environment and would not be standalone applications requiring reentry of data.

At the strategic LIMS, the system should be integrating with client operations. Possible examples of this could be the clients logging onto the system to initiate sampling within their areas. The system would print bar coded labels for the samples from printers within the client's area. The samples would be submitted to the laboratory for analysis and logged into the system via the bar code. An alternative scenario, in a production environment, could be the analytical laboratory developing robust methods for the production staff to operate and use the LIMS as a means of collating and disseminating the results. The LIMS would be the medium for the laboratory staff to monitor the performance of the analytical methods and alert them if a method were getting out of control.

**Horizontal view of a logistic LIMS** Looking along the matrix cells at the logistic LIMS level would give us a view of a system that has integrated laboratory



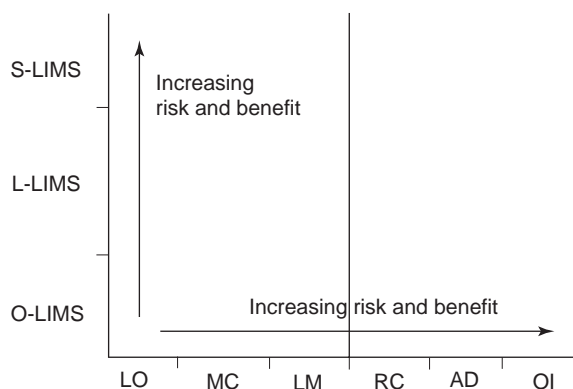
operations as described above. The monitor and control cell requires automated regulatory compliance, which would also include compliance with a quality scheme such as ISO 9000. There would be audit trail facilities as standard and the establishment of chain of custody from the time the sample arrived into the laboratory to when it was disposed or returned to the client. Automated regulatory compliance is a term that means that the documentation required to meet the quality standards that the organization is working to is generated by the system automatically. This would include receipt of samples and documentation of exceptions.

Laboratory management is concerned at this level with monitoring test use. Many laboratories accept requests for sample assays without questioning the rationale for the analysis or finding out if the information generated is ever used. The purpose of this cell is to question the analyses requested and monitor how the results are used. Is the laboratory generating the right information and offering the methods that the clients want? It is managing the client as much as managing the future direction of the laboratory with anticipating the services it will offer. The information from the LIMS, via a retrospective search of the database, will allow managers to follow up work that was done and how the results were used. Depending on the implementation of the whole system, this could be done electronically via feedback loops or manually.

The cells of reporting and communications, analytical decision-making, and organizational integration cover the functions whereby reports generated by the system are transmitted via the network, out-of-specification results are highlighted, and the clients can have remote access for online query of the database. The aim of these first two cells is to transmit the report to the client effectively and highlight aberrant results; this allows the client to focus immediately on problem areas. Some of the options available to implement this section are covered in more detail under 'Financial justification and risk assessment' in the next section.

These views of through the matrix are based on straight lines. However, not all systems will implement in straight lines and the views of individual systems will be based on functions required by an organization and a laboratory.

**Using the matrix for financial justification and risk assessment** The matrix can be used to justify the money spent on a LIMS on a cell by cell basis. The costs for each cell in the matrix can be calculated and the benefits quantified. This cost justification can be used to justify each cell of the matrix to senior



**Figure 6** The LIMS matrix showing increasing risk and benefits the further the system is implemented outside of the laboratory.

management. It will enhance the efficient use of resources and avoid duplication of functions within an organization. However, this idea can be taken further. Figure 6 shows as the LIMS is developed out of the operational level to the logistic and strategic levels and also into the organization there is increasing benefit but there is also a corresponding increase in the risk of the project. Therefore, a combined cost justification and risk assessment can be made for each matrix cell.

Let us look at two examples of this approach. In the logistic LIMS level of the matrix corresponding to the organization, communication is electronic, there is the ability to highlight out-of-specification results for the client for decision-making, and to help organizational integration, clients can access their results remotely. Ideally, the first two functions will help communicate results to the clients and enable them to make effective decisions. The third will enable the client to access the LIMS to see the progress of their samples and avoid the need to phone the laboratory to enquire about progress. All functions are desirable to integrate the laboratory and the organization. However, senior management is looking for a cost-benefit justification and an outline of the risks involved. Imagine two companies, one where information technology is well established and the second where use is not as successful as management would wish.

Scenario one would envisage an organization with an existing network linking all functional groups. This is a key component for organizational integration and provides the medium to move information. This organization is relatively sophisticated in its use of information technology with e-mail and online diaries available and used regularly. The main communications component, the network, is already installed. Therefore, the cost-benefit justification will only require expenditure for the customization of



reports and the linkage of the LIMS with the e-mail system to transmit a report to the client, a relatively minor cost component. To highlight out-of-specification results, more programming would be required to place an asterisk to highlight an out-of-specification result. The final cell in this level of the matrix requires external read only access via the network to the LIMS database. Security levels would be set up within the LIMS to prevent unauthorized users from changing results and roaming around the database; this may require a customized database view for each client or client group. With a network in place and working, the approach to implementation might be to authorize the communication cell and the organizational cell to proceed, with the ability to highlight out-of-specification results being implemented in the first place with the asterisk option. The risk assessment would be based upon a number of factors but would be low. The reasons are that there should be little impact on the ways of working as the staff already tend to work using the network facilities. The use of the matrix builds on the use of information technology and requires relatively little capital investment to be made.

In contrast, scenario two envisages an organization where information technology implementation is not always successful. A network is not in place throughout the organization and requires investment. Here the cost-benefit justification and risk assessment are changed dramatically to implement exactly the same functions. A higher capital investment is required to put in place the network. There may be the need to retrain or hire staff to run the network. Users would need to change their ways of working to use the system effectively, thus increasing the risk to the success of the whole project. Here the prudent approach would be to implement an operational level of the LIMS matrix: hard copy reporting first and concentrate on getting the LIMS working within the laboratory. As a future phase of the project, a network investment to the client groups could be justified on a piece by piece basis.

This illustrates the combined use of cost justification and risk assessment that should be applied to the use of the matrix. It also requires an assessment to be made of the use of information technology and the maturity of staff within an organization.

### Organizational Integration and LIMS

The use of information technology crosses functional boundaries: robots and other automation equipment are justified and used locally; however, the impact of a LIMS is far greater as shown by the matrix.

A LIMS has the ability to influence groups outside of the laboratory.

An operational LIMS has its impact mainly within the laboratory with little influence in the rest of the organization. As discussed earlier in this article, this is the typical implementation seen in the 1980s. The logistic LIMS has its main impact on the functional groups that work closely with the laboratory. They are linked via a corporate network that is fully realized with the strategic LIMS. The strategic LIMS has its greatest impact on the organization as well as providing benefits to the laboratory and the client groups served by the laboratory.

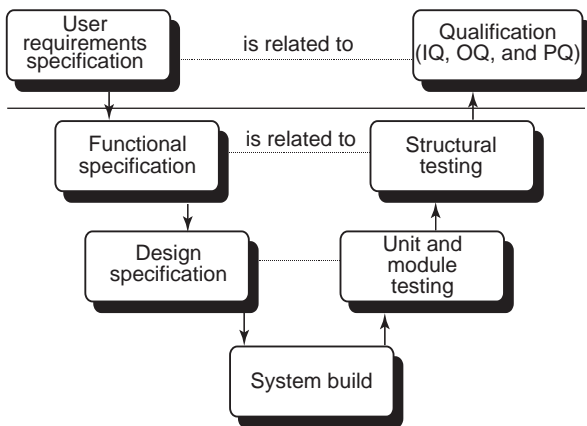
The implementation of a LIMS with increasing logistic and strategic elements means that there will be greater impact on the organizational structure. In instances where this occurs, a question arises: which processes should be automated?

Most LIMSs automate the current ways of working. These have evolved over time, but may not be efficient or cost-effective. Examples may be the way samples are received by the laboratory via multiple logbooks or several signatures are required to authorize a report. It may be prudent when considering such a system to consider reengineering or redesign of the laboratory operations prior to automating them. This is a higher risk approach, but it can improve the overall efficiency of the laboratory and its client groups and may be better than simply automating the status quo. We shall explore this approach in the next section of this module.

The other element in the equation is the impact of the LIMS on the existing organizational structures. It may be that organizational changes are required when implementing either a strategic or logistic LIMS. If this has not been anticipated or thought out, this can make this a high-risk factor. To avoid project failure due to this cause, an assessment of the impact of the system on the existing organizational structures is essential.

### LIMS and the System Development Life Cycle

The development and implementation of an information system such as a LIMS follows the traditional system development life cycle (SDLC) shown in Figure 7. Normally, equipment is purchased off-the-shelf and used with little or no modification; however, a different approach is necessary with complex laboratory automation such as a robotic system or a LIMS. In this instance, the users must state what tasks they want the system to undertake. Often this places the chemist in a difficult position as he or she



**Figure 7** Systems development life cycle.

is not certain what the system should or is capable of doing.

### Quality and the SDLC

Quality is designed into a computer application by defining the requirements and functions of the system. Quality can never be tested into a LIMS. Quality of a system is manifested in different ways such as user friendliness, undertaking the task it was designed to do, robustness of operation, and tolerance of error. During the development of the system by the supplier, quality is incorporated by ensuring that design and programming standards are adhered to. Furthermore, there are testing procedures that can be undertaken by the supplier to test the system prior to its release.

### Life Cycle Approach to LIMS Implementation

As can be seen, there are many stages to go through before the system is fully operational, thus the acquisition of a LIMS must not be undertaken lightly. The stages outlined in this module are guidelines, which based on the author's experience and can be recommended to any individual or organization contemplating an information system. These guidelines can be tailored or modified, as appropriate, to an individual situation.

As the life cycle unfolds the reasons for following this route will become apparent. The requirements and functions of a system must be fully defined, the responsibilities clearly stated, and good communications set up between all relevant parties. This is particularly important when the timescales of the SDLC are considered, the minimum overall time of a project is ~12–18 months and the maximum should be 24–36 months. However, timescales can be longer if the project is poorly managed. The timescales

obviously depend on the complexity of the project and how much configuration (changes of preset parameters in the software) or customization (programming) is involved as well how much time and resource can be committed by the organization. Configuration is preferable to customization for speed for implementation and long-term support issues.

The challenge for any organization is to implement a LIMS that supports the effective automation of 'information' production and distribution from the laboratory. Here the LIMS matrix can provide a high-level outline of what functions the LIMS should automate. Once the outline of the LIMS and the functions in the organization that it should automate has been discussed and agreed, the project proposal should be prepared.

### Project Proposal

The first stage in the life cycle is usually the preparation of a project proposal that will be reviewed by either senior management or a central computing resources committee. The information gathered during the laboratory review for the development of the laboratory strategy should be sufficient to state if computerization is a viable and cost-effective solution. Note the use of word 'computerization' and not LIMS, as the project proposal should be kept deliberately vague and not recommend definite hardware or software solutions.

The proposal document should contain an overview of the department and the problems that the system should overcome. It is important that the justification must be based on the business objectives of the department concerned and is not to be installed at the whim of management or an enthusiast. The outline justification must include the areas and estimates of cost savings that will be made when the new system is operational compared with an estimated cost of the package.

Outline approval by senior management to the often considerable financial outlay will be dependent on the ability of the laboratory to demonstrate, and where possible quantify, a return on the investment arising from benefits in:

- increased efficiency and productivity;
- reduced or limited increase in staff levels in relation to increased workload;
- increased scientific standards;
- production of more reliable data;
- improved speed of report production; and
- provision of management analysis functions.

The justification of a LIMS must be balanced between the 'tangible' benefits (i.e., those where a

return on investment can be calculated) and the 'intangible' benefits (i.e., where this cannot be done such as quality of data).

There is a balance between over-justifying the system to obtain project approval and being realistic about what the system will actually deliver in practice. This is very difficult to delineate; however, the review of the laboratory should give management the information to decide if the LIMS is worthwhile, if it is not, the project should not proceed.

A postimplementation review of the project, ~6–9 months after the LIMS is operational should temper overoptimistic justification of the system. The setting of criteria for success early in the project's life will help to make this review more objective than subjective.

Approval of the project proposal will normally allow the formation of a project team and allow the project to proceed to defining the requirements of the system, inviting tenders, and evaluating alternative systems.

### **The LIMS Project Team**

To solve the challenges posed by the design, implementation, and operation of LIMS a multidisciplinary approach is required. The formation of a LIMS project team to oversee all of the remaining stages of the SDLC up to the operation of the system is the normal solution. The minimum membership of the team will be users representatives (a vital component and often overlooked in some projects), laboratory management, and computing professionals.

There may be other members on the team such as personnel from client departments. When the system is to be operated under regulatory agency or government guidelines the involvement of a member of the quality assurance unit in all stages will be of direct benefit to the project. The project environment will entail computing and analytical experts working closely together; therefore, it is vital that each group has a full understanding of each other's areas of expertise and working practices. The computer scientist must know how the analyst works and vice versa, one area in particular must be the development of a common vocabulary for efficient and unambiguous communication between both parties.

A major problem in the implementation of any computer application, including LIMS, has been the crossing of disciplinary boundaries. The LIMS matrix can be used to help attain this goal: by giving both groups the common concepts and vision of a LIMS to help each group communicate with each other. Each discipline's own perspectives and strengths will be fused together in a project team

with the remit to implement a system successfully will enhance this communication.

The leader of the project team must be a member of the laboratory management, which clearly demonstrates to the users that it will be their system and not something that the computer department has decided may be suitable.

The project team must operate in a matrix management style, its members being responsible for what tasks should be undertaken and when they should be started and completed. The members liaise with their respective line managers who are responsible for providing the resources for the tasks and deciding how they should be carried out. This means that the project extends across department and discipline boundaries; therefore, communication is essential to ensure that members of the team and their respective line managers are aware of the progress and problems of the project.

### **User Requirements Specification and System Selection**

This is the most critical part of the system life cycle as it is the basis of all decisions and actions taken in the later phases. If a stage is wrong, then the rest of the life cycle can be jeopardized.

The assumption made in this section is that a commercial system will be selected rather than the organization changes their ways of working. A commercial system is a viable and cost-effective option as many systems are now available. Applications are networked and are scalable from small to large laboratories and can run from a single server across several sites and even continents.

The commercial approach has the advantage that development costs are shared over the whole user base and because of user feedback and competition there is continuous development of the software. A potential drawback may be that the vendor's conception of a laboratory may not be the way the package works and some configuration or customization of the package may be necessary.

**Understand the overall process** To obtain the best from the implementation of a LIMS, the project team will need to map and understand the whole process from sample generation to decision-making using the information generated. Inevitably, this will involve crossing functional boundaries and defining the hand-offs between the different departments involved in the process.

The team will need to identify where the process can be improved or redesigned to overcome bottlenecks in the existing process. As change is involved

with a LIMS project, the project team needs to give serious thought to changing the laboratory process to fit with the selected LIMS system. This will make the whole implementation faster than changing the LIMS to fit the existing laboratory processes.

**Write the user requirements specification** The new process map of the workflow gives the project team members the information needed to write the next document in the life cycle of a LIMS. This is the user requirements specification (URS). The URS is important as it gives the project team the outline tasks that the system should be capable of undertaking. In some respects, it can be regarded as a 'wish list' of functions that the system should be capable of performing.

The document should be written in a nontechnical style and follow a logical sequence starting from the entry of the sample into the laboratory, via analysis, through to the compilation of the analytical report and its distribution.

Each of the requirements needs to be uniquely referenced and be worded so that it can be tested. The URS is important as it defines what the laboratory wants from the system and stops the users being influenced by technology: the URS is the basis for the system selection.

**Write the invitation to tender** The URS is then converted into an invitation to tender or a request for proposal by appending to the document:

- A laboratory overview – describing the laboratory environment and the work undertaken. The analytical instrumentation and level of automation, the sources of work, and how results are reported. The present and past workload of the laboratory.
- Future of the laboratory – the future developments in automation, instrumentation, and techniques anticipated to be used by the laboratory. Estimated trends in workload and staffing. How is the laboratory changing?
- Integration with existing packages – data input from and output to other computers, software packages, and e-mail systems.
- Computing environment – describe the preferred hardware and software systems used at the local and corporate levels.

This information enables a prospective supplier to respond effectively to the invitation to tender. At this stage, for the project team to have as much choice as possible, the document should be sent to as many vendors as can be found.

To elicit similarly structured replies from all the vendors, either a section can be inserted in the tender document instructing the vendor how to respond or a checklist containing all the relevant questions that should be answered can be enclosed. It is important to know whether a function is available as standard in the package, needs to be configured within the standard package, customized, or is not available in the current released version of the system.

**Evaluation of prospective systems** The evaluation process is vital as it matches the laboratory requirements with those on offer from the suppliers. Unlike many other types of laboratory automation the full benefit from a system should not be expected until several months after the system has been introduced into the laboratory. This time lag means the impact of a wrong decision is much greater because much time, effort, and money have been expended. Therefore, the need to make the right decision the first time is very important.

The importance of requirements specification is now realized as it provides the written definition for the evaluation exercise. The key approach to evaluation is to develop an evaluation checklist for comparison of systems: the users should grade each requirement as either 'mandatory' or 'desirable'. All those features falling into the 'essential' category form the minimum requirements of the LIMS and the 'desirable' category features are those that are useful to have but the users do not require them in order to perform their normal role. To aid objectivity, essential items can be given a numeric value this will give a 'total score' for each system on offer.

The replies by suppliers to the tender document can be compared on paper to reject those obviously unsuitable, which should reduce the list to about four or five. These remaining should be visited and their replies discussed in depth and the systems evaluated by use. At the end of this phase, the systems may be reduced further to one or two for in-depth evaluation.

It should be noted that the sole use of a checklist should be discouraged as the subjective comments of the project team and users should be taken into consideration. This may take the form of comments on the interaction with the software or performance of the system.

Once a single system has been chosen, it should be tested in-house for up to a month (dependent on the complexity of the tasks it should undertake) to confirm that it is a suitable choice. Here the opportunity for all users to have hands-on access to the system must be encouraged. This in-house testing will usually entail the purchase of the license of a specified time from the LIMS vendor. If the evaluation is



successful, then the cost of the license should be deducted from the total cost of the system. If, however, the evaluation shows that the selected system is unsuitable then little time and money has been lost, which now means that the project team must reevaluate their requirements and consider other vendors and approaches and decide how to proceed with the project.

A final point for the evaluation: the project team must consider the laboratory environment whether it is changing or static. The reason is that software handles routine tasks well, but it is the unusual request that will prevent easy operation. Therefore, in a changing laboratory will the software be able to cope with changes in working practices?

This formal approach to the definition and evaluation of a LIMS will enable the project team to proceed to the purchase of a system with the confidence that it will undertake the tasks required of it. The in-house evaluation of the system will highlight areas that require tailoring to the way the laboratory wishes to operate as well as finding the deficiencies where custom software will be needed for full functionality.

**Vendor audit** In Figure 7, a line is drawn under the URS and the qualification stages of the V model. The reason is that for a commercial system, all the stages under the line are the responsibility of the vendor and one should see if the system offered is suitable in one's laboratory. In regulated laboratories such as in the pharmaceutical industry, there will be a requirement to see if the application software were developed in a quality manner. Hence, the need for a vendor audit.

**Purchase of the selected LIMS** The purchase of a LIMS will usually follow a company's procedures for capital equipment authorization, as the cost of a system will usually exceed the minimum limit for this process. The exception may be the instance when a company is purchasing the software alone to run on an existing processor and the cost may be considered as a revenue item.

Before placing an order, it is vital to check:

- Sufficient peripheral devices such as terminals, plotters, and printers are available to use the LIMS. This should ensure that there would be no complaints by users being unable to access the system.
- The small print of the contract should be examined to ensure that there are no escape clauses that can allow the vendor to avoid doing what they have promised to do.

- Ensure that commitments to enhance or modify the system are put in writing to protect both the purchaser and the vendor from any misunderstandings that might occur.

The purchase procedure will generally entail writing a further document, the basis for which is the project proposal. For most systems this will entail abstracting and updating the information already gathered. Include in the justification:

- business objectives of the laboratory;
- problems of the existing method of working;
- objectives of the system and how it should overcome the existing problems;
- design of the proposed system;
- detailed costing of the system including support, consultancy, and internal resources;
- competitive quotes from other suppliers, if required; and
- costs for laying information technology infrastructure and other equipment if required.

Once the purchase has been authorized and the order placed with the vendor, planning of the implementation should begin in conjunction with the vendor.

### Functional Specification

Once the final system has been chosen, the URS is used as a basis to input into the next document the functional specification that is the record of the configuration or customization of the application.

What is the difference between the requirements and functional specifications? An example of this could be a program to update records within the database of the LIMS. In the URS all that would be mentioned would be the need for a program to update records in the database. In contrast, the functional specification requires more detail:

- Functions that the program should undertake: create, copy, modify, view, and print records.
- Security considerations: who should have access to which functions, e.g., new or basic users may only use the view and print functions.
- Screen lay outs: how the program should appear to the user.
- Program logic: operation of the software if the module is to be customized or programmed.

As can be seen there is much more information required in a functional specification; it is for that reason that the configuration of a LIMS is faster and far more preferable than customization. It is also

easier to omit details, often essential, such as a 'delete record' function from the specification. This is the cause of many problems and one main reason for enhancement of the system functions later in the SDLC. An apocryphal definition of enhancement is 'what we forgot the first time around'. At first reading, it is funny; unfortunately it is also true.

Essentially, the functional specification is a document that can be read by the users and therefore is written with the minimum use of computer terminology. The functions to be undertaken by the system must be stated in a clear and concise manner. A major problem, if the wording is ambiguous or does not adequately describe the function, can be misinterpretation by programmers resulting in the program not working as required (see above). Close liaison between chemists and the supplier is essential, as the main problem is the crossing of disciplinary boundaries. The systems analyst who is drawing up the specification document must be aware of the intimate workings of the environment where the LIMS is targeted. To this end, spend one or more days walking him or her through the laboratory and show the operation. Great care is needed to include all working practices into the document to ensure that flexible software is written. Remember to define words to ensure common understanding.

Mistakes are expensive to correct and will cost time, effort, and credibility. The best advice is to consult the users, the laboratory management, and the client departments concerning the content of this document.

### **Qualification of the System**

There are three stages of this process:

- Installation qualification (IQ): is the system installed correctly.
- Operational qualification (OQ): does the system operate as the vendor says it should.
- Performance qualification (PQ): does the system work the way the users want it to.

The IQ and OQ can be performed in a relatively short time scale but there is usually a delay in undertaking the PQ as the system needs to be configured or customized before the testing can be started.

**Installation qualification** Prior to purchase of the LIMS, consideration should be given to the setting of the various pieces of hardware and to any special requirements for those areas. The LIMS computer could be sited in a central computer room that is air-conditioned room to control humidity and

temperature and with a false floor to allow for the large number of cables needed to connect the analytical instruments and peripheral devices to the processor. To allow controlled shutdown of the system in case of a power failure an uninterruptible power supply should be considered. The room housing the computer may need to be secure with access limited to authorized personnel only. The placement of the peripheral devices and the associated cables requires detailed discussion and planning with the computer department.

The system software and components will be delivered and installed, usually by the vendor's staff and any tests undertaken documented by them.

**Operational qualification** The OQ may be followed immediately from the IQ and can be performed by the vendor's staff or trained laboratory staff. This phase of the life cycle can be very short if there is extensive alteration of the system or relatively long if the system is used with little or no configuration.

**Performance qualification** PQ, also known as acceptance testing, has the objectives of confirming that the system operates as expected and to find errors in the LIMS.

It must be stated at the outset that the end user is responsible for PQ of the LIMS. However, it is impractical and virtually impossible to find all the errors in a system because of the multitude of different pathways through the software and the time it would take to test them all. Therefore, the PQ tends to be a compromise as software bugs will be found after the system becomes operational.

The size of most commercial LIMS is such that the PQ testing will be focused on the key system functions that acquire data, store it, transform it, and report it. If the system has been customized then the customized functions must also be tested as these are unique.

### **User Training and Roll-Out Strategies**

Once the LIMS has been validated and the documentation is available, the users can be exposed to the new system. There are three main ways this can be accomplished:

- total immersion,
- parallel operation, and
- selected use.

One approach to consider is a trial operation of the system, before the validation and the documentation of the system are finished, to find the working practices that will be used after the system becomes



operational. Regardless of which method is chosen, the learning curve is usually long, requiring much training and understanding. However, the benefits of a well-designed system should always outweigh the time spent on training.

**Total immersion** An implementation using total immersion uses the LIMS from a set date and abandons the old method of operation. This is a bold form of implementation and requires total confidence in the system that has been gained during both the acceptance testing and validation stages as well as total commitment from the users and management. This style of implementation is best suited when the LIMS is being used to implement change in the laboratory and its operation differs radically to the existing ways of working.

**Parallel operation** Parallel operation involves the running of both the old and new systems together to allow time to cross-check the data and information produced by the two methods and to build confidence in the LIMS. This approach will be applicable where tasks undertaken by both systems are similar; however, if the data produced by the two methods are not the same then the reason must be found and the problem resolved. If there are problems with the LIMS then the old system can still be used to produce results whilst a remedy is sought. This approach to implementation is resource intensive and a deadline should be given after which only the LIMS will be used.

**Selective use** Selective use, as the name implies, is the teaching of a small number of users how to use the LIMS. This is a cautious approach but has the advantage of allowing prompt attention to problems almost as they arise; moreover, in cases of slow learning then individual tuition can be given. This has an advantage over the other two styles that if the first groups of users are enthusiastic they can be used to 'sell' the benefits of the LIMS to other potential users by word of mouth. Care in the selection of the first group, therefore, is essential.

### **Project Close-Out**

Once the system is operational, the project team should formally close the project and hand the system over to the system manager. Normally the project team will disband; however, in the authors view, it should undergo a metamorphosis to become a 'departmental LIMS user group', which would be responsible for the future development of the system, monitoring its use, highlighting and solving problems, etc.

### **Postimplementation Review**

Usually after 6 months of operation the system should be reviewed. This should be a formal process where the actual performance of the LIMS in relation to the claims made at the project proposal or the purchase justification will be critically evaluated. Successes, failures, advantages, and problems should be documented: it is only now that the consequences of the decisions and actions that the project team took during the SDLC will be seen. It is essential for the organization that the successes and the failures be recorded and applied to other computing projects that either are ongoing or that will be initiated in the future.

### **Enhancement of the System and Controlling Change**

Once the initial configuration of the LIMS has been established, it will inevitably undergo change. This change comprises two areas.

The first area concerns the repair of hardware where a part is replaced due to malfunction, which should be logged in an instrument book to show the repair and maintenance records similar to an analytical instrument.

The second area of change involves the software: when an error is detected and remedied (by changing the code of the program) the software changes from one unique state to another. What is unknown is how many errors have been created by fixing the original fault. Hence, there is the need to revalidate either the software module or the whole system when a fault is remedied. Therefore, any changes to the system should also be authorized, logged, and validated to demonstrate that the system is under control.

If a LIMS is successful then the users will request change, which should be encouraged as it is a natural outgrowth of the system and a normal part of the SDLC. The changes requested by the users should be discussed at a meeting of the users, where they will be authorized, modified, or rejected as appropriate. The enhancement, if approved, can be communicated to a vendor to implement as part of a general improvement to the standard package or they can be contracted to undertake the work as a separate task or it can be carried out in-house with the appropriate staff.

*See also:* **Quality Assurance:** Quality Control.

### **Further Reading**

American Society for Testing and Materials (1993) *Standard Guide for Laboratory Information Management Systems (LIMS)*, Standard E5178. Philadelphia, PA: ASTM.

- Betteridge DJ (1987) Towards intelligent automation. *Analytical Proceedings* 24: 106–108.
- McDowall RD (1992) Strategic approaches to laboratory automation. *Chemometrics and Intelligent Laboratory Systems: Laboratory Information Management* 17: 265–282.
- McDowall RD (1993) A matrix for the development for a strategic laboratory information management system. *Analytical Chemistry* 65: 896A–901A.
- McDowall RD (1995) A matrix for a LIMS with a strategic focus. *Laboratory Information Management* 31: 57–64.
- McDowall RD (1998) What do you mean it doesn't work? *Scientific Data Management* March: 7–12.
- McDowall RD (1998) Who wrote this £\$%! program? Part 1. *Scientific Data Management* June: 8–14.
- McDowall RD (1998) Who wrote this £\$%! program? Part 2. *Scientific Data Management* September: 7–11.
- McDowall RD (2004) JALA, *Risk Assessment for LIMS and Laboratory Automation Projects* 9: 72–86.
- McDowall RD and Mattes DC (1990) An architecture for a comprehensive laboratory information management system. *Analytical Chemistry* 62: 1069–1075.
- Stein RR (1991) Evaluating the costs and benefits of laboratory information systems. *Chemometrics and Intelligent Laboratory Systems: Laboratory Information Management* 13: 15–36.

## Clinical Applications

**A Taylor**, University of Surrey, Guildford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Clinical work is concerned with the understanding of disease, its diagnosis, and treatment. Those involved with clinical research or those concerned with patient care have a range of disciplines that may be applied to a problem under investigation. Most of these activities require the exercise of skill and expertise whether, for example, it is for surgery in heart disease, radiotherapy to treat cancer, or molecular biology to reveal the genetic basis for disease. Disciplines such as clinical biochemistry, hematology, immunology, will, therefore, focus on disease, diagnosis, and treatment but their application is largely (although not exclusively) associated with analytical procedures. They involve both quantitative and qualitative analysis of specimens of body tissues and fluids – and the interpretation of these results in the light of observations of the patient. The quality of analytical data is crucial to clinical science.

### Clinical Analysis

Requirement for quality in laboratory medicine is best illustrated by one or two examples.

1. For healthy subjects, the concentrations of calcium in specimens of plasma will be within the narrow range of  $2.2\text{--}2.6\text{ mmol l}^{-1}$  ( $88\text{--}104\text{ mg l}^{-1}$ ). For any single individual the concentration will be maintained within an even narrower range, varying by less than  $0.04\text{ mmol l}^{-1}$ . A concentration

increase of no more than  $0.1\text{ mmol l}^{-1}$  ( $4\text{ mg l}^{-1}$ ) or  $\sim 4\%$  is sufficient to indicate the presence of pathology of the parathyroid gland with the possible consequence of surgery being necessary for that individual. Thus, calcium in plasma has to be measured accurately and precisely within these ranges and without interferences.

2. Children with the inherited disease phenylketonuria are unable to metabolize a component of their diet (phenylalanine) and will develop irreversible brain damage unless they are provided with appropriate foods. This condition can be diagnosed soon after birth by examination of a small blood sample to detect the presence of phenylalanine. Tragic consequences ensue should there be an error in the test procedure and the diagnosis is missed. Regular inclusion of positive and negative controls is essential to demonstrate the reliability of the procedure.

### Quality Management Systems

Clinical chemists were among the first to adopt the concept of analytical quality assurance in the late 1940s. Early studies were simply *ad hoc* distributions of specimens and comparisons of results among a few established laboratories, but they clearly demonstrated large differences in the values obtained. This work revealed for the first time the need for quality assurance, not just in clinical laboratories but also within all areas of analytical chemistry and various quality practices evolved over several years. To provide a common basis for nomenclature and application of quality assurance within hospital laboratories, the International Federation of Clinical Chemistry compiled a series of recommendations in

- Betteridge DJ (1987) Towards intelligent automation. *Analytical Proceedings* 24: 106–108.
- McDowall RD (1992) Strategic approaches to laboratory automation. *Chemometrics and Intelligent Laboratory Systems: Laboratory Information Management* 17: 265–282.
- McDowall RD (1993) A matrix for the development for a strategic laboratory information management system. *Analytical Chemistry* 65: 896A–901A.
- McDowall RD (1995) A matrix for a LIMS with a strategic focus. *Laboratory Information Management* 31: 57–64.
- McDowall RD (1998) What do you mean it doesn't work? *Scientific Data Management* March: 7–12.
- McDowall RD (1998) Who wrote this £\$%! program? Part 1. *Scientific Data Management* June: 8–14.
- McDowall RD (1998) Who wrote this £\$%! program? Part 2. *Scientific Data Management* September: 7–11.
- McDowall RD (2004) JALA, *Risk Assessment for LIMS and Laboratory Automation Projects* 9: 72–86.
- McDowall RD and Mattes DC (1990) An architecture for a comprehensive laboratory information management system. *Analytical Chemistry* 62: 1069–1075.
- Stein RR (1991) Evaluating the costs and benefits of laboratory information systems. *Chemometrics and Intelligent Laboratory Systems: Laboratory Information Management* 13: 15–36.

## Clinical Applications

**A Taylor**, University of Surrey, Guildford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Clinical work is concerned with the understanding of disease, its diagnosis, and treatment. Those involved with clinical research or those concerned with patient care have a range of disciplines that may be applied to a problem under investigation. Most of these activities require the exercise of skill and expertise whether, for example, it is for surgery in heart disease, radiotherapy to treat cancer, or molecular biology to reveal the genetic basis for disease. Disciplines such as clinical biochemistry, hematology, immunology, will, therefore, focus on disease, diagnosis, and treatment but their application is largely (although not exclusively) associated with analytical procedures. They involve both quantitative and qualitative analysis of specimens of body tissues and fluids – and the interpretation of these results in the light of observations of the patient. The quality of analytical data is crucial to clinical science.

### Clinical Analysis

Requirement for quality in laboratory medicine is best illustrated by one or two examples.

1. For healthy subjects, the concentrations of calcium in specimens of plasma will be within the narrow range of  $2.2\text{--}2.6\text{ mmol l}^{-1}$  ( $88\text{--}104\text{ mg l}^{-1}$ ). For any single individual the concentration will be maintained within an even narrower range, varying by less than  $0.04\text{ mmol l}^{-1}$ . A concentration

increase of no more than  $0.1\text{ mmol l}^{-1}$  ( $4\text{ mg l}^{-1}$ ) or  $\sim 4\%$  is sufficient to indicate the presence of pathology of the parathyroid gland with the possible consequence of surgery being necessary for that individual. Thus, calcium in plasma has to be measured accurately and precisely within these ranges and without interferences.

2. Children with the inherited disease phenylketonuria are unable to metabolize a component of their diet (phenylalanine) and will develop irreversible brain damage unless they are provided with appropriate foods. This condition can be diagnosed soon after birth by examination of a small blood sample to detect the presence of phenylalanine. Tragic consequences ensue should there be an error in the test procedure and the diagnosis is missed. Regular inclusion of positive and negative controls is essential to demonstrate the reliability of the procedure.

### Quality Management Systems

Clinical chemists were among the first to adopt the concept of analytical quality assurance in the late 1940s. Early studies were simply *ad hoc* distributions of specimens and comparisons of results among a few established laboratories, but they clearly demonstrated large differences in the values obtained. This work revealed for the first time the need for quality assurance, not just in clinical laboratories but also within all areas of analytical chemistry and various quality practices evolved over several years. To provide a common basis for nomenclature and application of quality assurance within hospital laboratories, the International Federation of Clinical Chemistry compiled a series of recommendations in

the 1970s while practical procedures were described in several textbooks of clinical chemistry. Meanwhile, important features of management, particularly in the context of preclinical trial work were codified as Good Laboratory Practice. These different themes were eventually picked up in a systematic way in international standards. Thus, ISO 9001:2000 addresses generic issues of 'quality management systems', while ISO/IEC 17025:1999 focuses on the 'General requirements for calibration and testing laboratories'. The requirements of ISO/IEC 17025 were amplified and applied to clinical analysis in a document, 'Essential Criteria for Quality Systems of Medical Laboratories', prepared by the European Communities Confederation of Clinical Chemistry (EC4, [www.uni-oldenburg.de/ec4/](http://www.uni-oldenburg.de/ec4/)). A very similar document was published in 2003 as ISO 15189:2003 'Medical laboratories – particular requirements for quality and competence'. The valuable experience over more than 60 years of quality assurance in clinical analysis is effectively captured in this last standard.

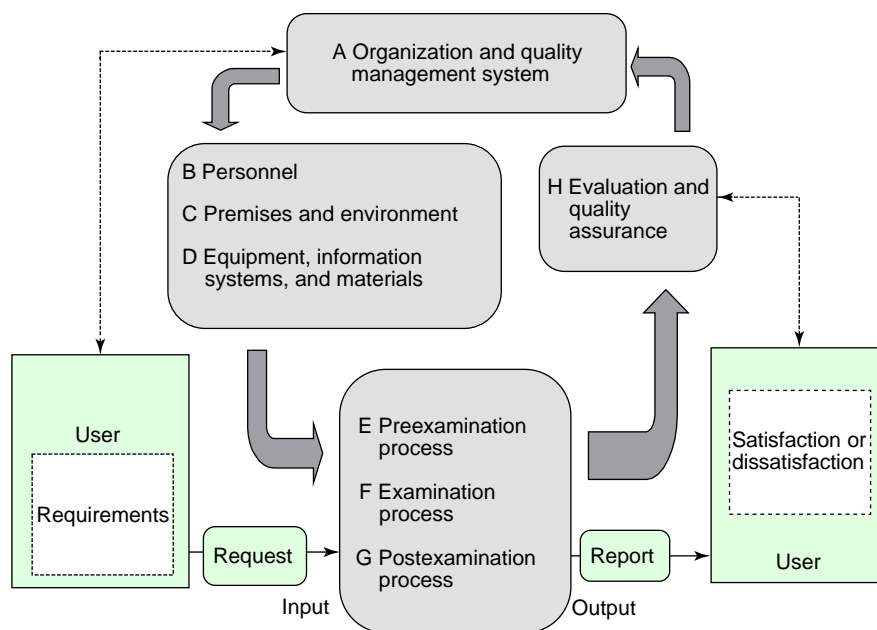
Alongside these developments has been the issue of accreditation, i.e., 'a procedure by which an authoritative body gives formal recognition that a body or person is competent to carry out specific tasks'. Those organizations responsible for accreditation of clinical laboratories draw extensively on the documents referred to above. ISO 9001 refers to a 'process based management system' in which elements of the quality management system (and,

therefore, of the work or process being undertaken) are subject to review and improvement to meet the requirements and comments of the user/customer. This may be viewed as a cyclical process where the work of the laboratory is periodically assessed in the light of user statements to modify the way in which the laboratory is organized and managed. One accreditation organization visualizes the process of laboratory organization and quality assurance in this way with its own accreditation standards (which are dependent on the ISO standards) forming essential components of the cycle (Figure 1). These concepts, together with extensive practical advice, are found in the guidebook by Burnett.

## Practical Approaches to Quality Assurance

### Introduction and Terminology

Analysts from different disciplines have adopted slightly different terms and also practical approaches to quality assurance. Recognizing this problem the international organizations for standardization have issued standards and guidelines addressing the definition of terms related to metrology, statistics, and analytical performances (International Vocabulary of Metrology, 1993; ISO Guide 30; ISO 3435, parts 1 and 2; ISO 5725, parts 1–6). In addition, a completely new concept has been introduced in 1993, with the ISO Guide on uncertainty of



**Figure 1** Relationship between a process-based quality management system and the CPA accreditation standards (from 'An approach to documentation in the medical laboratory'; [www.cpa-uk.co.uk](http://www.cpa-uk.co.uk)).

measurement, and, accordingly, the definition of terms related to analytical performances (namely within-run precision, between-run precision, and accuracy) have been revised to take into account this new development. Thus, the term 'accuracy' has been chosen to indicate the closeness of the agreement of an individual result with the 'true' value of the measurand. It is acknowledged that accuracy cannot be measured, but that the 'uncertainty' of measurement associated with a result, i.e., the interval of values that can be reasonably attributed to that measurand, may be determined. A new term, 'trueness', defines the closeness of the agreement of the mean of several results with an assigned value, and can be estimated as 'bias'. 'Precision' is a general term describing the closeness of the agreement between replicate measurements of the same sample by the same method. However, several different factors may affect estimates of precision and the terms 'repeatability' and 'reproducibility' are used to indicate two extreme conditions. Repeatability refers to measurements made under identical conditions, as far as possible (equivalent to within-run precision). Reproducibility is applied to estimates of precision that take account of all possible variables, including laboratories (more or less equivalent to between-laboratory precision). The term 'intermediate precision' refers to conditions lying between these extremes and it is recommended that the variables (time, operators, equipment, reagents, etc.) be indicated. As these terms are used in ISO 17025 and 15189 it is important to become acquainted with them. The complete definitions of the terms are given elsewhere in this Encyclopedia.

Although it is not universally adopted, most clinical laboratories now refer to 'quality assurance' with internal quality control (IQC) (intended to monitor precision and to detect significant changes in bias) and external quality assessment (EQC) (which can have a number of objectives including retrospective, independent supplementary checks on the effectiveness of the IQC procedures) as its major components.

### **Technical Components of Quality Assurance**

**Validation of a method** Method validation is the confirmation, by examination of objective evidence, that analytical procedures are appropriate for their intended use. Such evidence is generally already available for methods published as standards, or by recognized technical organizations. In other cases, such information may be included in the original published work or provided by the manufacturer of the equipment. Where an entirely novel method has

been developed, e.g., a procedure for the measurement by liquid chromatography of a new antibiotic agent in specimens of blood, or a mature method, e.g., the measurement of alkaline phosphates activity in serum with *p*-nitrophenylphosphate as substrate, is transferred to a different analyzer, performance characteristics need to be determined. Modifications to established methods may require only a limited study or a complete new validation, depending on the extent and importance of the changes.

The features to be determined to validate a procedure are repeatability (within-run) and intermediate (between-run) precision, limit of detection, limit of quantification, linearity of the response of the assay, specificity of the assay and whether there are any interferences, its trueness, the uncertainty associated with an individual result with a stated limit of confidence, and appropriate reference ranges. While examining these topics further information relevant to reagent stability, necessary frequency of recalibration, suitable IQC protocol, and overall assay weakness will be obtained.

**Precision** To measure repeatability one sample is repeatedly analyzed within a single assay batch. From the results given by this exercise the analyst can calculate the mean concentration, the standard deviation (SD), and the coefficient of variation (CV, sometimes referred to as the relative standard deviation, RSD). While there are no fixed number of replicates necessary for this measurement, it is usual to try and obtain at least 20 results, in order to ensure reliable statistics, and the number should always be stated in any description of the method's performance. The experiment should be repeated with further specimens having different concentrations of the analyte so that a clear indication is achieved of the precision over the range of concentrations that is measurable by the procedure (i.e., the precision profile). A similar series of measurements can be made to determine intermediate precision. Factors such as the preparation of reagents and calibration solutions, resetting of instrumental parameters, and the expertise of the analyst will contribute to the analytical variation. The influence of some of these variables can be critical and, unless great care is taken, the intermediate precision will be very poor. Inevitably, the CVs will be greater than the equivalent repeatability data.

Information derived from these experiments has two purposes. First, it can be decided if the methodological procedure is suitable for introduction into the repertoire of a laboratory. If the analytical variation is in excess of the resolution



needed for clinical applications it will not be possible to detect significant changes in patient condition with any certainty or confidence. The method will have to be improved or discarded. This issue of clinical need is separate from quality assurance but can only be understood in the context of good analytical performance. Attempts to quantify the requirements of analytical methods for clinical purposes (quality specifications) have been developed for individual analytes, on the basis of normal intraindividual day-to-day (or hour-to-hour) variations, or as a function of their reference ranges. While these calculations can be helpful, in practice, the fundamental goal is to achieve as good precision as is practicable. Second, the data provide the guidelines for IQC protocols (described later).

*Limit of detection (LOD) and limit of quantitation (LOQ)* If the analyte concentration is always at a level that is so high that sensitivity is not a problem, e.g., hemoglobin in blood, calculation of the LOD is unimportant. For a large number of other parameters the concentrations will always be very low. The LOD must be established so that a result can be quoted with an assurance that it has validity and does not represent an extreme of the background 'analytical noise'. With many investigations it is necessary to confirm whether or not a substance (e.g., HIV antibody) is present in the specimen. In these cases the minimum amount that can be determined must be known so that the sensitivity of the analysis can be considered when deciding if further action is necessary. The LOD is influenced by the inherent sensitivity of the instrumentation and by the methodological imprecision. To measure the LOD a specimen giving a reading close to that of the reagent blank should be analyzed with a large number (10–20) of replicates. The concentration equivalent to that of the reagent blank plus three times the SD of the series of results thus obtained represents the LOD. Although the LOD provides information on what can be distinguished from the blank, a high level of imprecision will affect the measurements of concentrations close to the blank. It is therefore sometimes useful to calculate the LOQ, i.e., the lowest concentration of an analyte that can be determined with acceptable precision (repeatability) and accuracy under the stated conditions of the test. If the accepted level of repeatability is taken as 10%, the LOQ can be calculated from the same series of data as the concentration of the reagent blank plus 10-fold the observed SD.

*Linearity of calibration graph* A simple experiment involving analysis of a series of standard solutions

can provide valuable practical information and show how many calibration standards are needed for an assay, i.e., whether or not the calibration graph is a straight line, and at what point does the graph flatten out and it becomes necessary to dilute specimens with high concentration of analyte. Regression analysis is not sufficient to establish linearity and visual inspection or the analysis of residues will give more information.

*Interferences* Most specimens of clinical importance have complicated and variable matrices. As a consequence there are many factors that can influence accuracy and have to be considered in the characterization of new or modified methods. Procedures to establish that a method is providing unbiased results will be considered in a later section but the features of biological systems and clinical specimens that can cause erroneous results include:

- Direct methodological interferences. Serum collected from patients with renal failure has increased concentrations of compounds such as urea while high levels of bilirubin are found in specimens from subjects with liver dysfunction. These are known to interfere (positively and negatively) with many analytical methods. Lipemic specimens cause light scattering that may produce aberrant results in spectrometric assays.
- Metabolites may be more or less active than the parent biomolecule, but a nonspecific methodology that is unable to differentiate between the various forms gives misleading results.
- A biological, protein-based matrix is more viscous and contains a high concentration of dissolved inorganic salts. The complex matrix can reduce the speed of pipetting and lead to differences in handling characteristics by tubing, syringes, etc., when compared with aqueous calibration solutions, and give erroneous results.
- Pathological conditions can give rise to sudden and unexpected large changes in the concentrations of a metabolite to be measured. Dilution of the specimen to allow quantification will adjust other components of the sample and may result in subtle changes to the measurement.
- Very many of the specimens collected for clinical investigation are from patients receiving drug treatment. Interferences associated with pharmaceutical agents or their metabolites are well recognized and documented.
- Measured concentrations can be influenced by characteristics of the methodology. It is usually unwise to rely entirely on data determined elsewhere



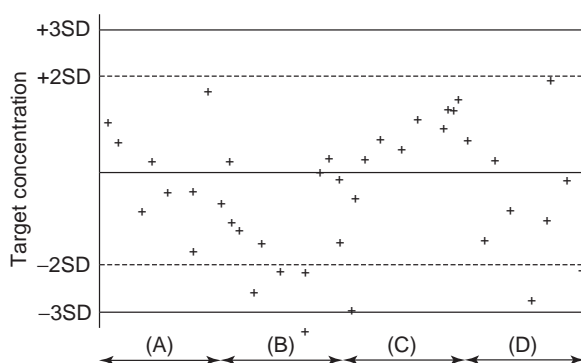
to define the concentrations within normal populations or associated with particular clinical states. A series of investigations to confirm or establish reference ranges should be undertaken.

At the conclusion of these experiments the analyst should have a complete set of information to characterize the method, with relevant figures of merit, review of interferences, speed and costs of the procedure, and data to indicate more qualitative features such as robustness and the degree of technical skill required for regular use.

**Internal quality control** The inherent imprecision of a method will have been determined as part of the validation, as it was implemented within the laboratory (intermediate precision). The information can then be applied to the IQC program that is designed to identify the intrusion of a bias and/or an alteration in the precision of the assay. Factors that can contribute to a bias or change in precision include:

- instrument stability or drift,
- deterioration of reagents,
- errors associated with the calibration material,
- problems with equipment, e.g., partial blockage within a dilutor or a shift in the monochromator calibration on a spectrometer,
- human errors, e.g., use of incorrect reagents, instrument settings, etc.

An IQC program can comprise several elements, but the most usual is to include a specimen(s) with known (or target) concentration/result within each batch, and to take that specimen(s) from a large stable pool so that it can be used in this way over a period of several months or longer. The precision of the method will show the analyst the allowable range of results to anticipate around the target concentration. An acceptance range of  $\pm 2SD$  is usually applied but these limits should not be so wide as to deprive the analysis of useful clinical value, a point that should be evident at the time when the method is developed and introduced into the laboratory repertoire. The practical approach is to draw a Levey–Jennings chart with the results of the IQC specimen(s) plotted (Figure 2). A series of decision criteria or ‘control rules’ have been developed by Westgard ([www.westgard.com](http://www.westgard.com)) for use with this type of chart. Westgard rules guide the analyst as to the number and concentration of IQC specimens to be included, the frequency of inclusion within a test run, and how QC results can be applied to decisions as to whether a run should be accepted or rejected. Results



**Figure 2** Results for an IQC specimen shown on a Levey–Jennings chart. This example presents four analytical scenarios: (A) good control where points are evenly distributed above and below the target concentration and  $\sim 95\%$  are within  $\pm 2SD$ ; (B) a sudden shift in accuracy – perhaps the calibration material has become contaminated; (C) a gradual shift in accuracy – perhaps the substrate solution has exceeded the expiry date; (D) very poor precision – perhaps a valve is sticking and causing erratic transfer of sample. Other types of display are proposed from time to time, which are claimed to offer increased sensitivity to detect errors. The Levey–Jennings plot, however, is very widely used.

determined for patients’ samples should be reported only if the concentration of the IQC specimens satisfies preset criteria, whether the Westgard rules or those set internally by the laboratory.

In addition to the systematic inclusion of quality control specimens as part of the analytical process there are other procedures that can be incorporated into the IQC program. These tests are usually applied to methods that are undertaken on a regular, almost daily, basis. They involve the carry-forward of specimens and the calculation of batch/day means. The sensitivity of daily mean plots is increased by the CUSUM technique (<http://www.westgard.com/lesson14.htm>).

Some analytical equipment include routines for calibration and quality control of analyses that are predetermined by the software developed by the manufacturers. In these circumstances, the principles and practices elaborated above may be unrealistic and analysts are forced to follow protocols as directed by the equipment.

No clinical laboratory should operate without an IQC program, but whatever approach is adopted the program will fail to have any impact unless it is accompanied by a well-defined policy for further action. The policy will form part of the quality management system and will describe the ways in which results of IQC specimens are to be recorded, the members of staff who should scrutinize these data, and the possible decisions that they can make. The outcomes of a properly implemented IQC program should be the early detection of errors or

analytical problems, the maintenance of a known level of performance even if there are changes among staff, departmental routine, or even relocation of the laboratory, and prolonged experience of a procedure that can act as a 'benchmark' against which alternative methods can be compared.

**Accuracy and external quality assessment** ISO 15189:2003 requires both the use of validated methods and the estimate of the uncertainty of measurement (i.e., the inaccuracy) associated with the individual results. To fulfill this requirement, determination of the method bias (trueness) is always necessary. In practice, trueness can be examined from two approaches. First, there is the estimate that should be part of the initial validation of the method (or a review if problems are found to have developed). An extension of this involves work with definitive and reference methods, primary standards, and reference materials and the traceability of results to SI units. A statement of the uncertainty of measurement should accompany each comparison and contribute to the estimate of the total uncertainty of measurement of the result. To this aim, manufacturers of *in vitro* medical diagnostic devices are now compelled to give information on the traceability of the results obtained using their products and further developments are to be expected in this field. The second approach to determination of trueness in the medical laboratory involves EQA. This has several synonyms, e.g., interlaboratory comparisons, round-robin exercises, or proficiency testing programs. In part, this work represents an ongoing surveillance of accuracy to supplement the laboratory's IQC program but EQA can fulfill a number of other functions including the provision of materials for which the concentration of analytes is known with reasonably good accuracy. These reference materials can subsequently be used for initial characterization or review of methods, or for the evaluation of instrument performance.

**Measurement of trueness** Wherever possible, investigations of the trueness of a method should include analysis of reference materials (RMs). The value of RMs for IQC has already been discussed but for studies of trueness it is those RMs with previously determined and well-defined concentrations of analytes that are required. The essential features of RMs, stability, homogeneity, composition of matrix, etc., are reviewed elsewhere in this Encyclopedia, but it is important to indicate here how the concentrations of clinical RMs may be derived and the degree of confidence that can be attributed to the stated levels. Certified reference materials (CRMs) are those with

concentrations defined with the smallest uncertainty and thoroughly documented. For clinical applications, however, the determinands that are present in CRMs are few in number and do not include most of those that make up the repertoire of clinical laboratories. Reliance has to be placed, therefore, on the 'assayed quality control' specimens (see below). Manufacturers of these specimens have different practices to determine the 'assayed values'. These include in-house measurements, results obtained by a limited number of expert laboratories, and distribution via an EQA scheme to a large number of laboratories to give consensus data. Almost certainly the assigned concentrations will be given together with an indication of the methodologies so that methodological biases can be taken into consideration. In addition to commercially available assayed RMs, the organizers of EQA schemes usually have specimens that are surplus to the requirements of the scheme. Organizers are generally willing to respond favorably to requests for one or more aliquots of the specimens and the relevant consensus data from the scheme.

A second experiment to investigate trueness is to calculate the recovery of a known amount of analyte added to a suitable specimen. Quantitative recovery from normal biological fluids must be followed by further experiments to determine whether there are interferences associated with pathological specimens. However, recovery experiments are not always possible in the clinical laboratory because many of the items that have to be measured, e.g., enzymes and immunoglobulins, are not available in a purified form.

Further studies on trueness are particularly valuable for newer work when the analyte under investigation is not included in any RM or EQA program. A series of specimens can be exchanged between laboratories and if the results obtained by the partners are not in good agreement the analysts must conclude that there are problems with trueness. This approach does not necessarily reveal interferences or biases associated with the technique and, therefore, it is desirable to take in addition a series of specimens for measurement by the method under investigation and by a method that involves an entirely different analytical technique.

**External quality assessment** Many countries have authoritative, regular, structured provision for EQA for at least general clinical laboratory work (biochemistry, hematology, microbiology). There are also large numbers of specialist EQA schemes, often international, for those analytes measured in fewer laboratories. In addition, schemes to monitor

laboratory activities that do not have a quantitative outcome, e.g., histological examination of tissue, the interpretation of analytical data, etc. are available.

As they play such an important role in the assessment of laboratory performance, a number of documents have been produced, suggesting minimal requirements for the managerial and technical aspects of EQA schemes. The issues addressed in these documents have been mainly integrated in the guide issued by ISO (ISO 43:1997, Parts 1 and 2) and the CPA standards for the accreditation of EQA schemes ([www.cpa-uk.co.uk](http://www.cpa-uk.co.uk)). The ISO standard for the accreditation of medical laboratories (ISO 15189:2003) also indicates that the laboratory should participate in EQA schemes that are substantially compliant with the requirements of the ISO guide.

Organization of EQA schemes consists of up to four components:

- management,
- data reduction,
- preparation of reports,
- assessment of laboratory performance.

Management of the EQA scheme is concerned with procurement and/or preparation and distribution of good quality specimens with analytes at concentrations appropriate for the needs of the participants, and at a volume consistent with usual laboratory routines. The number and frequency of distribution will be determined together with the time allowed for the analysis and return of results.

Data reduction involves the mathematical procedures applied to the results returned by the participants. Usually, there is a procedure to identify 'outliers' – results that are remote from the anticipated value, probably caused by a transcription error or analysis of a different sample. As inclusion of these results could distort subsequent calculations it is useful to apply a statistical routine (e.g., greater than  $\pm 3SD$ s from the mean) to eliminate these data. The 'trimmed' data are then taken through the chosen calculations that, with few exceptions, are determination of the number and range of results, the mean, SD, and CV. If it is suspected or has been definitely shown that the distribution of data is not normal, the median should be found. In schemes where there are hundreds of participants, subroutines are included to re-examine results and present these calculations for a specific method or other variable.

The operation of so many EQA schemes, with their own computer programs for the preparation of reports, makes it inevitable that reports will be very

different in appearance. Most organizers are so familiar with what they are doing that they are certain that participants will immediately comprehend the significance of all the calculations and information that is laid before them. This is a rash assumption and a clear display, logically laid out on a page with the salient features prominently positioned, is required. Bespoke reports that identify the result from the individual laboratory and its relationship to other participants are preferred. Where schemes include a large number of analytes it is helpful to provide a one-page summary of the participants' performance at the front of the report.

While a laboratory can review its own results against those of the other participants and decide if analytical performance is satisfactory, it has been shown that improvements are most effectively stimulated by a more formal independent assessment that includes some kind of measure or score. Scores allow laboratories to be compared and participants can look at their own performance relative to that of other colleagues. Changes in performance over time, within a single laboratory, are also revealed by regular calculation of a score.

As with report formats, performance assessments may be idiosyncratic to a scheme but some consistency in approach has been recommended in ISO 43:1997, which recommends use of the  $z$ -score, which is

$$z = \frac{x - X}{s}$$

where  $x$  is the laboratory result,  $X$  the target concentration, and  $s$  an appropriate estimate/measure of variability that is selected to meet the requirements of the scheme organizer. The  $s$ -value can be determined from data derived from the results of the participants of a particular scheme or in other ways, e.g., the SD achieved by reference laboratories. Scoring systems have undoubtedly prompted enormous improvements in performance and in established schemes that have run for several years; results reported by most participants are in close agreement and indicate good accuracy.

A further step in the application of EQA to laboratory performance is to establish quality specifications, i.e., performance scores that are indicative of good and poor performances and to a degree of quality to which all participants should aspire. Performance assessments are prepared for the benefit of participating laboratories and their users. In addition, performance scoring can be employed to assess the competence of laboratories where licensing is required.

The International Federation of Clinical Chemistry Working Group on Quality Control defined the functions of EQA schemes as:

- To supplement a laboratory's IQC procedures.
- To measure the performance of individual laboratories for comparison with other participants, or for changes with time (to which licensing and/or accreditation could be added).
- To obtain consensus concentrations for specimens.
- To investigate features within a laboratory that contribute to performance; e.g., method, size of laboratory, work load, test frequency, equipment, etc.
- To measure the 'state-of-the-art' for a test.
- To act as an educational stimulus.

The first three features in the list have been covered in the previous pages; the remaining aspects, especially that concerned with identification of reliable methods, etc., are relevant to laboratory management (see below).

It was emphasized at the beginning of this article that analytical accuracy is crucial to diagnosis and treatment of many clinical conditions. As with IQC, the laboratory must have a clearly defined policy to describe action taken on receipt of a report from an EQA scheme. Those involved with organization of such schemes can provide many examples of horror stories where reports have been ignored or deliberately suppressed on receipt, rather than admit that a problem exists. This type of bad and potentially

dangerous practice cannot occur if senior managers enforce the policy.

### Managerial Components of Quality Assurance

Laboratory management is typically concerned with general organization and policy arrangements, all of which contribute to quality. This will include selection and training of staff, evaluation, purchase and maintenance of equipment, selection of analytical methods, document control, protocols for clinical investigation, and audit. Other management topics such as preparation and issuing reports, corrective actions, resolution of complaints, risk assessments, and health and safety are equally relevant to quality assurance and all should form part of the departmental quality management system but are outside the scope of this review.

See also: **Clinical Analysis:** Overview.

### Further Reading

- Burnett D (2002) *A Practical Guide to Accreditation in Laboratory Medicine*. London: Venture Publications.
- International Federation of Clinical Chemistry. Expert Panel on Nomenclature and Principles of Quality Control in Clinical Chemistry. *Clinica Chimica Acta* (1975) 63: F25–F38; (1976) 69: F1–F17; (1977) 74: F1–F9; (1977) 75: F11–F20; (1978) 83: 189F–202F.
- ISO (1999) ISO/IEC 17025: General requirements for the competence of testing and calibration laboratories.
- ISO (2003) ISO 15189: Medical laboratories – Particular requirements for quality and competence.

## Water Applications

**J E Gaskin**, Environment Canada, Ottawa, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 4261–4280, © 1995, Elsevier Ltd., with revisions made by the Editor.

A fundamental consideration in the reliability of any analytical measurement is that of sample quality. Sample quality may be appraised on the basis of the quality of the resulting chemical, physical, and biological measurement data.

Assurance of acceptable data quality requires a system which ensures that all activities associated with the sampling program be well defined and supported by accepted and standardized practices.

### The Need for a Water Quality Field Sampling Quality Assurance/Quality Control Program

A field quality assurance/quality control (QA/QC) program must ensure the integrity and representativeness of the samples taken from aquatic sources such as surface waters, groundwater, and marine environments.

If nonrepresentative samples are collected, and sampling preservation techniques are inadequate, the quantitative results will be invalid. Emphasis must not be placed only on the collection of samples that are appropriately representative of the aquatic source from which they are taken, but also on the

The International Federation of Clinical Chemistry Working Group on Quality Control defined the functions of EQA schemes as:

- To supplement a laboratory's IQC procedures.
- To measure the performance of individual laboratories for comparison with other participants, or for changes with time (to which licensing and/or accreditation could be added).
- To obtain consensus concentrations for specimens.
- To investigate features within a laboratory that contribute to performance; e.g., method, size of laboratory, work load, test frequency, equipment, etc.
- To measure the 'state-of-the-art' for a test.
- To act as an educational stimulus.

The first three features in the list have been covered in the previous pages; the remaining aspects, especially that concerned with identification of reliable methods, etc., are relevant to laboratory management (see below).

It was emphasized at the beginning of this article that analytical accuracy is crucial to diagnosis and treatment of many clinical conditions. As with IQC, the laboratory must have a clearly defined policy to describe action taken on receipt of a report from an EQA scheme. Those involved with organization of such schemes can provide many examples of horror stories where reports have been ignored or deliberately suppressed on receipt, rather than admit that a problem exists. This type of bad and potentially

dangerous practice cannot occur if senior managers enforce the policy.

### Managerial Components of Quality Assurance

Laboratory management is typically concerned with general organization and policy arrangements, all of which contribute to quality. This will include selection and training of staff, evaluation, purchase and maintenance of equipment, selection of analytical methods, document control, protocols for clinical investigation, and audit. Other management topics such as preparation and issuing reports, corrective actions, resolution of complaints, risk assessments, and health and safety are equally relevant to quality assurance and all should form part of the departmental quality management system but are outside the scope of this review.

See also: **Clinical Analysis:** Overview.

### Further Reading

- Burnett D (2002) *A Practical Guide to Accreditation in Laboratory Medicine*. London: Venture Publications.
- International Federation of Clinical Chemistry. Expert Panel on Nomenclature and Principles of Quality Control in Clinical Chemistry. *Clinica Chimica Acta* (1975) 63: F25–F38; (1976) 69: F1–F17; (1977) 74: F1–F9; (1977) 75: F11–F20; (1978) 83: 189F–202F.
- ISO (1999) ISO/IEC 17025: General requirements for the competence of testing and calibration laboratories.
- ISO (2003) ISO 15189: Medical laboratories – Particular requirements for quality and competence.

## Water Applications

**J E Gaskin**, Environment Canada, Ottawa, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 4261–4280, © 1995, Elsevier Ltd., with revisions made by the Editor.

A fundamental consideration in the reliability of any analytical measurement is that of sample quality. Sample quality may be appraised on the basis of the quality of the resulting chemical, physical, and biological measurement data.

Assurance of acceptable data quality requires a system which ensures that all activities associated with the sampling program be well defined and supported by accepted and standardized practices.

### The Need for a Water Quality Field Sampling Quality Assurance/Quality Control Program

A field quality assurance/quality control (QA/QC) program must ensure the integrity and representativeness of the samples taken from aquatic sources such as surface waters, groundwater, and marine environments.

If nonrepresentative samples are collected, and sampling preservation techniques are inadequate, the quantitative results will be invalid. Emphasis must not be placed only on the collection of samples that are appropriately representative of the aquatic source from which they are taken, but also on the



appropriate preservation, transportation, and handling procedures that are used in the life of each sample in its journey between the collection site and the laboratory. In addition to the need for representativeness of samples, the data produced from the samples must be of known quality, sufficient to meet the needs of the sampling network and its data users. Some desired characteristics of data quality are accuracy, precision, confidence limits, data completeness, data compatibility, and data comparability.

### **QA/QC Documentation**

For a field sampling quality assurance/quality control (QA/QC) program, there are three documents that are required. The first is the 'Quality Assurance Program Plan', which commits management to a QA policy and sets out the requirements for data needed to support program objectives. The program plan describes the overall policies, organization, objectives, and functional responsibilities for achieving data quality goals.

The second document is the 'QA Project Plan', a technical document that specifies the QA and QC requirements for each project. The plan specifies any QA/QC activities required to achieve the data quality goals of the project. It describes how all data are assessed for precision, accuracy, completeness, comparability, and compatibility (components of data quality criteria and objectives). The QA Project Plan further requires that all data generated be thoroughly documented, and be in sufficient detail to permit unambiguous evaluation of project results.

The field sampling QA/QC manual is also a component of the QA Project Plan and must provide guidance on policy and procedures. This manual will contribute to the quality of the data generated.

The third document that is needed is a 'Program Implementation Plan'. A number of mechanisms must be put in place to ensure maximum coordination and integration of QA efforts within the overall program (covering sampling, laboratory analysis, and data handling). Management committees directing water quality programs must coordinate the monitoring and survey programs to ensure that data of known and acceptable quality are produced. Resource levels, schedules, turn-around times, responsibility centers, performance indicators, milestones, risk factors, implications, emerging issues, etc., must all be given attention, if the programs are to succeed.

## **Designing a Water Quality Monitoring Plan**

### **Sampling Programme**

The field QA/QC strategy includes a sampling plan which serves as the central element for directing the various operational activities. To produce a truly effective sampling plan, an interdisciplinary approach is necessary. The interdisciplinary contributions assure the development of a plan that is specific and has all the requirements for successful sampling operations.

The sampling plan must concern itself with the 'why', 'what', 'where', 'when', and 'how' questions. The 'why' question must address the broad objective(s) of the monitoring program, and the data quality objectives associated with the various specified projects. The rationale for each project should include an estimation of the project's utility and cost-effectiveness. The 'why' question can be influenced by considerations such as prevailing environmental concerns and problems; a combination of socioeconomic, sociopolitical, and environmental issues; the need for continuing and/or expanding existing projects or programs, and the need to provide baseline data at new specific sites.

The 'what to sample' should be a logical sequel from the answer to the 'why to sample' question. Specific water quality parameters may be issue-oriented, or may result from water quality objectives arising from negotiated water quality agreements between concerned and/or interjurisdictional water industry agencies.

The choice on 'where to sample' may be controlled by the decisions taken on the 'why to sample' and 'what to sample for'. However, once a macro-level choice (river basin, lake, or stream) has been agreed upon, selecting sampling stations (micro-level choices) is an exercise controlled by a number of exogenous and endogenous factors. These factors include hydrological data (flow rates, sediment discharges, etc.) for various sections of the aquatic body; topography of the macro-site; geographical location with regard to interjurisdictional boundaries; ease of access to sampling sites; runoffs and material infusions (sediment, water, sludge, sewage, etc.) from external sources; mixing currents; water depth; and anthropogenic activity.

The 'when to sample' choice is influenced by, for example, (1) objectives of the study; (2) available resources; (3) seasonality, including spring runoff, heavy snowfall, and heavy rainfall; (4) required estimated sampling frequency; (5) high or low flow rates within the sampling cross-section of the



aquatic body; (6) unusual environmental and ecological occurrences such as flooding and chemical spills; (7) intentional or unintentional toxic waste discharges; (8) anthropogenic activities which may be deleterious to the aquatic body; and (9) the density of airborne and nonpoint source pollutants.

The appropriateness of 'how to sample' can be addressed by consideration of nine principles:

1. Preliminary sampling, such as a pilot survey, should be conducted in order to provide basis for an evaluation of the sampling design and the statistical options.
2. Replicate samples should be taken with each combination of time, location, and any other controlled variable. Differences 'among' can only be demonstrated by comparison to differences 'within'.
3. An equal number of randomly allocated replicate samples should be collected for each combination of controlled variables. Collecting samples from 'representative' or 'typical' places is not random sampling.
4. Verification should be made that the sampling method is correct for the range of sampling conditions to be encountered. Note that variation in efficiency of sampling can bias 'among-area' comparisons.
5. If the area sampled has a large-scale environmental pattern, it should be broken up into relatively homogeneous subareas, and sample collections should be proportioned to the size of each subarea.
6. The sample size should be verified as appropriate to the quantity and spatial distribution of the parameter. An estimate should then be made of the number of replicate samples required to obtain the necessary precision.
7. To test whether a condition has an effect, samples should be collected both where the condition is present and absent, but all else is the same. An effect can only be demonstrated by comparison with a control.
8. On-going site evaluations for parameters such as pH, conductivity, turbidity, and dissolved oxygen should be incorporated into the sampling plan, with some consideration given to the possible need for more extensive studies at sites where specific problems have been identified.
9. The data from the pilot survey(s) should be tested and analyzed to determine whether the error variation is homogeneous, normally distributed, and independent of the mean. If this is not the case (a situation which is highly probable with biological parameters), then the data could be appropriately transformed, or distribution-free (nonparametric) procedures could be used.

## Site Selection

Site selection must be given detailed consideration during the development of the 'Sampling Programme Plan', since sampling techniques to be used in a given situation will depend not only on the data needs, but also on the nature of the flow and other conditions.

Appropriate siting is critical to the success of the monitoring network. Poor siting could result in low-quality, unrepresentative data. It is therefore important that comprehensive quality assurance/quality control procedures (QA/QCPs) be developed to address network siting. These procedures should be designed to meet a specific set of quality assurance objectives, namely:

- to assure that the geographical distribution of sampling locations meets network objectives;
- to ensure that the sampling sites (micro-locations) that are chosen are appropriately representative of the existing aquatic characteristics of the given water body; and
- to ensure that site documentation is sufficient for data users to assess the suitability of the sites for the required purposes.

Meeting these QA/QC objectives requires the implementation of the following procedures:

- Conducting a preliminary site assessment, followed by a report prepared within the first six months of operations, so that obvious problems are detected early, and the site can be changed before a large data set is obtained.
- Routine reporting of significant changes in data obtained from the site during period-to-period sampling.
- Proper documentation of sampling sites during program operations.
- Independent quality assurance audits of sampling sites. This should be conducted in order to (1) assess independently the accuracy and completeness of current site documentation, and (2) to assess the representativeness of each site with respect to its monitoring objectives. Site representativeness could be checked by looking at data for a number of physical parameters (turbidity, pH, conductivity, and dissolved oxygen).

Apart from using objectives to allocate sampling stations, other factors such as size of aquatic source and available resources must enter the selection scheme so that a reasonable degree of optimization can be achieved. Resources may be allocated on the

basis of the size of the aquatic source, sampling frequency, number of sampling sites within a macro-location, site accessibility, and distance of site(s) from the office which is coordinating or conducting the sampling operations. Different aquatic sources (rivers, lakes, streams, estuaries, etc.) require different or varying approaches in the selection of sampling stations.

**General considerations for station location** As a general rule, past or present monitored sites should be taken into account when locating a water quality sampling station, to ensure a long-term continuity in data sets or to identify seasonal cycles. Water quality monitoring stations should be located close to water quantity stations when these exist already, in order to facilitate the study of water quality–quantity relationships. The distance between the collection site and a gauge recorder should not be more than 60 to 80 m, unless it has been ascertained that there is no appreciable inflow between sites. If there is no gauging station at or near the sampling site, a discharge measurement should be made at the time of sampling.

### Sampling Frequency and Sample Size

Many water quality monitoring jurisdictions perform monthly sampling, but the day of the month and the time of the day may vary randomly. In the majority of cases, sampling is related to the estimated variance in the measurements of the water quality constituents of concern.

The simplest design case on a single station-single variable basis would be to select the sampling frequency which results in the desired confidence interval width about the annual mean (or geometric mean) for the specified water quality variable at the specified site. If the estimate of the variance (obtained from historical data or a pilot program) is assumed to be the population variance ( $\sigma^2$ ), then the number of samples  $N$  (per period) required to obtain a given degree of confidence in estimating a mean value for the parameter of interest, can be derived from the following equation:

$$N \geq Z_{\alpha/2}^2 \cdot \sigma^2 / E^2 \quad [1]$$

where  $Z_{\alpha/2}$  is the standard normal deviate corresponding to a probability of  $\alpha/2$ , and  $E$  is the error term, and is given by  $\mu - x$  for any specific or chosen value ( $x$ ).

If  $s$  is used instead of  $\sigma$ , and  $t_{\alpha/2}$  (Student's  $t$ ) is used in place of  $Z_{\alpha/2}$ ,  $N$  is given as

$$N \geq t_{\alpha/2}^2 \cdot s^2 / E^2 \quad [2]$$

A number of sources have provided a very comprehensive outline of methods for determining sampling frequencies for 'single station-single variable'; 'single station-multiple variables'; and 'multiple stations-multiple water quality variables'. The following is an example of the use of eqn [1].

'The samples to be measured for a given parameter are expected to have a mean annual concentration of  $\sim 0.1 \mu\text{g ml}^{-1}$  with a standard deviation of  $0.05 \mu\text{g ml}^{-1}$ , and the tolerable error in the stated value of the mean at the 95% confidence level ( $Z = 1.96$ ) is not to exceed 20% ( $0.02 \mu\text{g ml}^{-1}$ ). If it can be assumed that the measurement error is small in comparison with the measured values, find the number of samples required per year to estimate future annual means for the given parameter.'

With the above values, the number of samples required will be

$$N = (1.96)^2 \cdot (0.05)^2 / (0.02)^2 \\ = 24 \text{ samples per year, or two samples per month}$$

The appropriate size of samples to be collected for any given parameter will depend on the estimated natural concentration level of that parameter within the aquatic source. The smaller the existing analyte concentration, the larger the volume necessary for obtaining a reasonable measurement (a measurement equal to or greater than the limit of quantification after volume reduction).

For organic analytes, samples of 100 liters or more are collected for solvent extraction. New field *in situ* solvent extraction techniques are now providing alternatives to handling large liquid volumes. However, for a number of diverse parameters, there is still a need for estimating the sample volumes required to measure levels of interest. A number of empirical and semi-empirical methods have been developed for determining sample sizes so as not to exceed a predetermined level of sample uncertainty. One approach is based on the knowledge that the between-sample standard deviation decreases as the sample size increases.

A modified form of the Ingamells' equation can be adopted to determine the sample volume that is required for an estimated standard deviation and a predicted sampling uncertainty. The equation is of the form:

$$VR^2 = M_s \quad [3]$$

where  $V$  is the volume of sample analyzed,  $R$  is the relative standard deviation (in %) for repeat measurements on the sample analyte, and  $M_s$  is the sampling constant, corresponding to the volume of

sample required to limit the sampling uncertainty to a chosen percentage value (e.g., 1%) with a proposed level of confidence (e.g., 68%). The magnitude of  $M_s$  can be determined by estimating the between-sample standard deviation  $S_s$  from a series of measurements of samples of volume  $V$ . Once  $M_s$  is evaluated for a given sample, the minimum volume  $V$  required for a maximum relative standard deviation of  $R\%$  can be calculated.

An example of the use of  $M_s$  is as follows.

'Estimate the volume of sample to be collected for the analytical measurement of a given parameter, when the values of  $M_s$  and  $R$  are given as 100 ml and 20%, respectively'.

Inserting the values of  $M_s$  and  $R$  into eqn [3],  $V$  is calculated as

$$V = 100 \text{ ml}/(0.2)^2 = 2.5 \text{ l}$$

When a decrease of the original volume (via solvent extraction, distillation, or evaporation, etc.) is required before analysis, back calculations must be carried out to determine the original volume that must be collected to give the desired measurable level of the parameter. For example, if  $V_1$  is the volume calculated from the use of  $M_s$  and  $R$ , then the required original volume  $V_0$  is obtained from the product of  $V_1$  and the volume reduction factor. For a volume reduction of 100,  $V_0$  would be  $100V_1$ .

Another form of the Ingamells' equation:

$$WR^2 = K_s \quad [4]$$

can be applied to measurements involving bottom sediment samples.  $W$  represents the weight of the sample analysed,  $R$  has the same meaning as in eqn [3], while  $K_s$  replaces  $M_s$  in the same equation.

Whether volumes or weights are being considered, every effort should be made to assure homogeneity in the sample matrix before any subsampling is carried out.

### Field and Laboratory Analytical Practices

The sampling plan must consider the role of good field and laboratory practices. The basic QA/QC field testing and laboratory analytical protocols that will be adopted must be spelt out in the sampling plan. For the laboratory, the scope of both the intra- and interlaboratory QA/QC programs should be referenced, and should be consistent with the size of available resources.

The performance of an instrument must be checked and appraised by a qualified person before being transported to the field. This involves a visual inspection and verification of its operation including the zero and full-scale calibration. Documents on

standardization, calibration, maintenance, equipment safety, and spare parts should accompany each instrument.

The standardization of instruments should be performed using reference standards when these are available, or against certified standard instruments. This must be done before the instrument is transported to the field, or in some cases the reference instrument could be transported to the field and employed under actual field conditions. Laboratory standardized instruments should be rechecked upon arrival in the field.

Calibrations must be conducted under the same instrumental and chemical conditions as those that will exist during the measurement process. The frequency of calibration will depend on the accuracy requirements of the investigation and on the stability of the instruments. Daily calibration checks are recommended when the instrument is in daily use, or at other times, immediately prior to a series of measurements. For unstable instruments, the calibration should be checked before each series of measurements, between measurements, and after the last measurement.

The calibration process is vital to all measurements and should be governed by a calibration plan. Such a plan should provide for:

1. calibration procedures and record forms;
2. stated calibration frequencies;
3. appropriate sources for obtaining certified and high quality standards, or the best means of producing accurate in-house standards;
4. a list of all calibration standards (including nomenclature and assigned identification numbers);
5. specifications of environmental conditions; and
6. intended range of validity.

The sensitivity of each analytical method is an important characteristic, and must be known. The precision and accuracy of measurements must reflect the level of confidence placed on the measurements.

The accuracy required from the measuring system should be determined by the objectives of the sampling program, and by the use of the data. The measurement system should avoid superfluous and expensive accuracy and precision determinations if there is a failure in the monitoring program to meet the stipulated objectives. The analytical measurements must also be rugged and complete.

### Cost Considerations

Program managers involved with water quality monitoring program must be concerned with program cost-effectiveness. The operation of a quality cost

system would help to lessen the cost of operational activities which are directed toward controlling data quality. The basic concept of the quality cost system is to minimize total quality costs through proper allocation of planned expenditures for the quality assurance and quality control efforts, while maintaining an acceptable level of data quality.

The quality cost system structure provides a means of identification of quality-related activities and for organization of these activities into prevention, appraisal, and failure cost categories. These categories are defined as follows:

- Prevention costs are associated with planned activities whose purpose is to ensure the collection of data of acceptable quality and to prevent the generation of data of unacceptable quality.
- Appraisal costs are associated with measurement and evaluation of data quality. This includes the measurement and evaluation of materials, equipment, and processes used to obtain quality data.
- Failure costs are incurred directly by the monitoring agency or organization producing the failure (unacceptable data).

There is no preset formula for determining the optimum mode of operation. Rather, the cost-effectiveness of quality costs is optimized through an iterative process requiring a continuous analysis and evaluation effort. Maximum benefits are realized when the system is applied to a specific measurement method in a stable long-term monitoring program. For example, a monitoring program with a fixed number of monitoring sites, scheduled to operate for more than a year, would be a desirable candidate for a quality cost system.

The size and diversity of a water quality monitoring operation is an important determining factor in decisions related to formalization of the QA program. Operations concerned with defined programs of work may be more easily adaptable to formal QA activities, and the program costs can be more easily justified. For operations involving numerous repetitive measurements, a formal quality assurance program is a necessity, and the time and effort devoted to its development should be compensated by cost-effective reliable analytical data. The overall cost of the QA/QC activities should not exceed 15–25% of the total monitoring operational effort, and ‘repeats’ and lost data should be minimized.

### **Training of Field Personnel**

All personnel involved in sample collection, field analysis, data reduction, and QCPs must be adequately trained. They must be made aware of the

importance of obtaining data of known quality, must become familiar with the quality assurance program, and must be knowledgeable about the various ways of achieving and maintaining good quality data.

Training personnel for field operations will involve a broad range of disciplines, because of the diversified requirements (mechanical, physical, chemical, biological, hydrological, etc.). Personnel training must start before operations begin, and must continue during operations, with periodic reviews and retraining when indicated or required.

To maintain and foster consistency in the quality of collected data, all field personnel (particularly new employees) should become familiar with project objectives, sampling protocols, station location, equipment operation, and safety rules prior to the commencement of field work.

Suggestions for general on-the-job training for sampling and field-testing personnel, include:

1. studying operation manuals and the sampling program plan;
2. observing a trained and experienced operator perform the various tasks in the data collection process;
3. performing field operations under the direct supervision of a trained and experienced person, and then independently performing operations related to instruments, sample collection, sample analysis, and data handling;
4. attending and participating in in-house seminars which are directed towards problem solving and confidence building;
5. gaining experience and acquiring new skills and technical information through interchange programs between various water quality agencies, in cases where the interchange becomes a beneficial exercise; and
6. announced and unannounced periodic proficiency testing of operational personnel to determine the need for additional training.

### **Field Trip Preparation**

Before starting out on a water quality data collection trip, field personnel should determine the types of samples that are to be collected and the field measurements required to assemble the proper sampling and testing equipment. Field personnel should also ensure that the equipment being transported to the field is in good working order and that there are spares available for those pieces of equipment which are hard to replace once out in the field. Environment Canada’s manual on *Sampling for Water Quality*



(1983) gives a detailed outline of necessary preparation for field trips. The preparation entails:

1. obtaining specific instruments on sampling procedures;
2. preparation of an itinerary according to the sampling schedule;
3. preparation of lists of required equipment, supplies, and materials;
4. ensuring that all sample containers, covers, and attachments have been cleaned in accordance with standard procedures, and that they are safely packed;
5. ensuring that the laboratory has prepared the chemical reagents and standards needed for the trip; and
6. preparation of checklist (including road maps, station location descriptions, field sampling sheets, labels, equipment manuals, tool box, etc.).

### Selection of Sampling Equipment

The collection of water samples requires the use of a variety of sampling equipment depending on the

station, the medium to be sampled, and the water quality parameter list. The choice of sampler type must be closely related to the parameter list in order to avoid sample contamination. In addition to being parameter and station specific, sampling containers must be such that they provide suitable sample volumes, and are appropriate for use in a wide variety of environmental conditions.

**Sample bottles** Sample bottles are generally provided by the analytical laboratory. The selected bottles (including the liners and caps) must be free of contamination and suitable for the water quality parameters to be analyzed. **Table 1** gives a description of the recommended containers for specific water quality parameters, and also outlines the bottle washing and preparation procedures used for ensuring the cleanliness of the sample bottles. The volume of sample required depends on the type and number of parameters to be analyzed, the analytical method, and the expected concentrations of the parameters in the water.

**Table 1** Recommended containers and washing procedures for selected water quality parameters

<i>Parameters to be determined</i>	<i>Recommended container</i>	<i>Washing procedure</i>
Physical parameters	1 l polyethylene bottle	Rinse three times with tap water, once with chromic acid, three times with tap water, once with (1 + 1) nitric acid, and then three times with distilled water in the given order
Major ions		
Nitrogen: $\text{NO}_2/\text{NO}_3$		
Nitrogen: $\text{NH}_3$		
Nitrogen, total		
Dissolved organic carbon		
Dissolved inorganic carbon		
Suspended solids	2 l polyethylene bottle	Same as for physical parameters
Phosphorus, total	0.5 l glass (Soivrel) bottle	Rinse three times with tap water, once with chromic acid, three times with tap water, once with (1 + 1) nitric acid and then three times with ultrapure distilled water, in the given order
Metals, extractable	0.5 l polyethylene bottle	Rinse three times with tap water, once with chromic acid, three times with tap water, once with (1 + 1) nitric acid and then three times with ultrapure distilled water, in the given order
Chromium, total	0.25 l polyethylene bottle	Same as for metals but chromic acid is not used
Mercury	0.125 l glass (Soivrel) bottle	Rinse three times with tap water, once with chromic acid, three times with tap water, once with (1 + 1) nitric acid and then three times with ultrapure distilled water, in the given order
Organic toxic chemicals	1 l or 4 l glass (amber) bottles with cap lined with polytetrafluoroethylene (PTFE)	Rinse three times with tap water, once with chromic acid, three times with organic free water, twice with washing acetone, twice with pesticide grade hexane, and dry the uncapped bottle in a hot air oven at 360°C for at least 1 h
Urea	1 l glass (amber) bottle	Same as for organic toxic chemicals

Recommended sample containers, washing, and preparation procedures associated with the collection of suspended-sediment and bottom-sediment samples are outlined in *The Sampling for Water Quality Manual*.

Prior to sampling, every bottle (or sample container) should be rinsed twice with the water to be sampled. During the rinsing process, bottle caps should be put on loosely to ensure adequate rinsing of the cap, liner, and bottle threads. If replicate samples are taken, every set of bottles should be rinsed twice and then filled, one set after the other (i.e., sequentially).

Proper labeling of the sample bottle is necessary to avoid possible mistakes and confusion when the sample arrives at the laboratory. All sample and subsample containers must be affixed with a strip of vinyl labeling tape when the containers are warm and dry. Information such as the station number (station identification), sampling date and time, water quality parameter(s) to be analyzed, and the program submitter identification (Sub. ID) must be clearly printed on each container with a water-proof fine-tipped nylon marker. Bottles designated for blank samples should be similarly identified using a special code or number that has been reserved for the identification of field blank samples.

## **Water Quality Sampling**

### **Sample Representativeness**

For new sampling programs, pilot projects should be conducted to resolve logistic problems and implement procedures that would support the objectives for reaching the program goals. This should include an evaluation of each collection site by investigating temporal and spatial homogeneity. Also, the collection frequency for priority variables, sample size, preservation effectiveness, shipping efficiency, and minimization of contamination risks could be determined or checked at this stage. In addition, proper sampling techniques must be used to ensure that a sample is representative of the flow in the sampled cross-section.

The number of verticals to be sampled in a cross-section should relate primarily to the collection of a representative sample and secondarily to the volume of the sample required. If field measurements of specific conductance, temperature, pH, and dissolved oxygen show the water to be well mixed, then a sample obtained at a single vertical near the centroid of the flow may be assumed to be representative of the total flow.

*In situ* measurements of pH, temperature, specific conductance, and dissolved oxygen should be taken before actual water samples are collected, to minimize the effects of water column disturbance caused by the lowering and raising of the sampling apparatus. This is particularly important in the case of residence time waterbodies, and during periods of ice cover when the vertical distribution of many parameters may be highly variable. Once the *in situ* measurements are conducted, water samples may be collected, starting at the surface and proceeding down through the water column if applicable. Collection of sediment samples should be performed after all water column samples have been taken.

Special considerations are required for groundwater sampling. Groundwater vulnerability to contamination is affected by water depth, recharge rate, soil composition, topography (slope), as well as other factors such as the volatility and persistence of the analytes being determined. In planning groundwater sampling strategies, knowledge of the physical and chemical characteristics of the aquifer system is necessary (but almost never known). Groundwater presents special challenges for obtaining representative samples.

Samples that are to be collected for suspended sediment, or samples to be analyzed for total constituents should be collected by using approved and properly treated water-sediment depth-integrating samplers. Suggested instances where use of these samplers are not required are as follows:

1. Extreme low flow where use of the sediment sampler is impractical. Samples may be collected by immersing the bottle by hand.
2. Flood conditions when the velocity of flow is so great that the sampler (sampling apparatus) cannot be lowered.
3. Samples collected for dissolved chemical constituents that are well mixed within the sampled cross-section.
4. The necessity for aseptic collection techniques for bacteriological samples. Consequently, these samples can be collected at a single vertical near the centroid of flow, or by immersing the sample bottle by hand (when the water source is shallow).

In sampling for suspended sediment, cross-section samples should be obtained at every opportunity, using either the equal discharge increment (EDI) or equal width increment (EWI) method to obtain cross-section coefficients to apply to the point samples obtained by the sampler.



## Automatic Sampling

In many instances, particularly in accessible remote areas, automatic samplers are used effectively. In comparison to manual sampling, automatic sampling can be more cost effective at remote sites. The cost effectiveness is closely linked to the instrument's reliability, improved capabilities, and greater sampling frequency ability.

For the efficient and effective use of an automatic water quality sampler, flow data should be available at every sampling station where the sampler is used. Flow data can provide a sound basis for interpretation of water quality results. Unfortunately, automatic samplers are limited in their application. The sample handling procedures required in quantitative measurements for a wide range of water quality parameters (nutrients, major ions, total metals, toxic organics, etc.) make it virtually impossible to use automatic sampling for detecting and measuring parameters other than dissolved oxygen, pH, specific conductance, temperature, color, and turbidity.

Samples collected at remote sites by automatic samplers should be retrieved at the earliest possible time. Samples collected in this manner should be analyzed only for constituents that do not require on-site preparation and could be assumed to be representative of the associated flow event.

High concentrations of suspended solids and floating materials can limit the performance of automatic samplers and reduce the accuracy of the measurements made. Suspended sediment samples collected by automatic samplers should be analyzed individually, and the specific conductance should be measured for each sample. This helps to determine the sample representativeness of the individual samples.

Periodic manual sampling should be conducted to verify the accuracy and representativeness of results obtained from automatic samplers, and in selecting these instruments for limited water quality monitoring, attention must be paid to such factors as:

- the range of intended use;
- the level of accuracy desired;
- the skill level required for installing and servicing the automatic sampler;
- mechanical reliability of the sampler;
- adaptability of the sampler to varying weather patterns and flow regimes; and
- cost of the sampler.

## Sampling for Specific Parameters in Specific Matrices

**Toxic chemicals in water** Many pesticides and heavy metals are toxic to aquatic life even when

they are present in low concentrations. The toxicity of a particular substance is dependent to a large extent on other water quality characteristics such as temperature, pH, alkalinity, and dissolved oxygen. The complex nature of toxic chemicals (particularly trace organic toxic chemicals) demands the application of a number of QA/QC efforts which are additional to those generally used for less complex substances.

Although this section will focus on trace organic toxic chemical sampling, preservation, and analysis, the protocols and procedures that are outlined can be applied (with less complexity) to trace inorganic toxic chemicals.

In the toxic chemical monitoring programs, water, bottom sediments, fish (forage and predator), and other biota samples are collected and analyzed.

**Toxic organic chemicals** Sampling for toxic organic chemicals in water is conducted in a manner similar to any other type of water sample collection. Normally, 4 l amber glass bottles with PTFE-lined caps are used, and rinsing of these bottles in the field is not required. Sample preservation is carried out using the procedures listed in Table 2.

Recently, a number of improved field QA/QCPs have been instituted by Environment Canada with regard to toxic chemical survey projects. These procedures include the following:

- preparation of field blanks;
- preparation of spiked field blanks;
- preparation of spiked field samples (often in replicates);
- sequential replicate sampling; and

**Table 2** Procedures used for the preservation<sup>a</sup> of water samples collected for toxic chemical analyses

<i>Parameter</i>	<i>Preservation</i>
Pesticides	Add 200 ml of hexane per 4 l ample
Organochlorine	(in a 4.5 l amber glass bottle),
Organophosphorus	shake well and deliver to the
Polyaromatic	laboratory as soon as possible.
hydrocarbons	Place a piece of aluminum foil over
Polychlorinated	the mouth of the bottle before
biphenyls	screwing on the cover (except
	when there are PTFE liners in the
	cap).
Chlorophenols	Adjust to pH 4 with 25% H <sub>2</sub> SO <sub>4</sub>
Carbamates	Adjust to pH 3 with 25% H <sub>2</sub> SO <sub>4</sub>
Chlorobenzenes	Add 50 ml of hexane per 1 l and shake
Triazines	Adjust to pH 8 with 0.1 mol l <sup>-1</sup> NaOH,
	add 200 ml of methylene chloride
	(dichloromethane) per 4.5 l sample
	and shake

<sup>a</sup> All samples are kept as cool as possible in the field.

- preservation of spiked blanks and samples a few hours after spiking, to permit prolonged interaction between sample matrix and spiking solution.

In the preparation of a field blank, distilled deionized water is first added to a sample bottle prior to the field trip. The appropriate preservative is then added to the water blank in the field at the same time as the collected samples are being preserved. The resulting blank is called a field blank. The field blank is used to check on (1) the cleanliness of sample containers; (2) random field contamination; and (3) the purity of added preservatives.

In addition, solvent blanks are also run during the analytical determinations in the laboratory.

Prior to its use in the blanks' preparation, the distilled or deionized water should be checked for its purity in the laboratory. It is possible that the water may require some treatment (solvent extraction, filtering through XAD-type resin columns, etc.) before it is suitable for the required use.

Preparation of spiked field blanks is also required to check on the recovery of a spiked analyte of interest in a 'clean' matrix. Spiking solutions are prepared by laboratory personnel prior to field trips. Preparation of spiked field blanks also serves to monitor the stability of the spiking solution in a 'clean' matrix, and to check the accuracy of its preparation. The analytical results of the spiked field blanks and the spiked field samples (which are prepared at the same time in the field) are compared to see if the analytical measurement process is indicating analyte recovery at the same level in a natural sample (the presence of matrix effects) as in a blank (no matrix effects).

Spikes can be prepared by using an appropriate micro-syringe (which has been previously rinsed with solvent) to add a precise volume (generally 100 to 200  $\mu\text{l}$ ) of a spiking solution (of known concentration) to blanks and/or regular samples. All field personnel involved in the preparation of spikes should be trained in the correct use of a micro-syringe. Larger volumes of less concentrated spiking solutions may be appropriate for the preparation of spikes. This could minimize possible manipulation errors when a spiking solution is added to a sample.

Spiked samples are generally prepared in duplicate or triplicate in the field, to verify the reproducibility of spike recoveries, and to gain a better understanding of the spiking process (matrix effects, systematic errors, etc.). Replication of spikes on all occasions is recommended. Based on resource availability, duplication of spikes could be an acceptable minimum, although triplication would possibly provide more information.

The addition of preservative to spiked blanks and spiked samples should be carried out a few hours after spiking; this time delay permits physical interaction between the sample matrix and the spiking solution.

In reporting the analytical results on the spikes, both the absolute concentrations and the percent recoveries should be reported by the laboratory.

The frequency with which this set of field QA/QC samples, namely blanks, spiked blanks, and spiked samples (in replicate), should be prepared would be dictated by the water quality study or project and available resources.

Random 'sequential replicate sampling' (see below) should also be conducted regularly in addition to the above activities. Conducting sequential replicate sampling for 10–15% of all samples collected may be an adequate level of operation for a toxic chemicals QA/QC program. This sampling replication provides a means of monitoring the short-term variability of the parameters of interest. Additionally, the replication program together with information obtained from other field QA/QC samples may permit the discernment of contamination problems and/or contamination sources.

### **Extraction Techniques**

The use of liquid–liquid extraction techniques has permitted the detection of very low concentrations ( $\text{sub-ng ml}^{-1}$ ) of organic contaminants. In the extraction process, large quantities of solvent are needed for 40 l samples, and the purity of the solvent must be assured. When the 40 l sample is filtered by the extractor, the suspended solids retained on the filter can also be kept for analyses.

In the past, use has been made of Amberlite XAD-4 resin columns in the field for sample extraction of fenitrothion and aminocarb in rain water. The rain water was collected and extracted with the aid of a special sampling (collection) device equipped with Amberlite XAD-4 columns. It is possible that these types of extraction columns could have some field application in the extraction of large volumes of surface water for the determination of specific organic contaminants. However, one drawback associated with resin columns is the requirement of exhaustive cleaning with various solvents to remove all trace contaminants. Preparation of blanks, spiked blanks, spiked samples (in replicate), and sequential replicate sampling should be included as part of the specific QA/QCPs that are needed, if these Amberlite XAD-4 resin columns (or others) are used more extensively in the future.

The inclusion of QA/QCPs would assist in routinely checking the purity of the resin column(s), the cleanliness of the apparatus, the recoveries attained by the extraction and elution processes, the reproducibility of the extraction and elution efficiencies, and the representativeness of the water samples.

### Toxic Chemicals in Bottom Sediment and in Biota

The collection and interpretation of biological data present more difficulty than in the parallel cases with water and bottom sediment. Many of the organisms collected are mobile, and thus it is difficult to class them as representative of the site from which they are taken. Some water quality investigations involving biological parameters require monitoring or surveillance of an aquatic area over a long period of time. Such systematic resampling usually employs either a 'transverse' or a 'longitudinal transect' system, or a 'grid' or 'quadrant' system.

Regardless of the type or purpose of the biological sampling study, sample collection should be designed on a meaningful basis. A large number of samples collected at the wrong time or place have less value than a few samples carefully selected as to time and place of collection. The frequency of sample collection will depend on the variability of environmental and biological factors and on the study objectives. The greater the habitat variability, the more intensive the sampling program must be. Life history events of the organisms also must be considered in the design of a sampling program.

In spite of the temporal and spatial variability of aquatic populations, statistical techniques are available for the design of sampling program and for the evaluation of biological data.

In the quality assurance program for biota sampling, attempts must be made at maintaining statistically significant sample sizes for each site. In addition, where appropriate, factors such as population size, size of habitat, feeding habits, weight, sex, lipid content, and geographical disposition of habitat (relative to industrial sites, agricultural lands and developments, waste dumps and sewage inputs) should be determined to help in data interpretation.

Selection of discrete samples of the aquatic biota should be random, and the mean of five or more measurements on individual samples should be determined in order to give representative values of the parameters of interest during the laboratory analysis stage.

### Bacterial Sample Collection

Samples for bacteriological examination must be collected in bottles that have been carefully cleaned

and autoclaved for ~20 min at 121°C at 15 psi. Also glassware, except when in metal containers, may be sterilized in a hot-air oven for not less than 1 h at a temperature of 170°C. Glassware in metal containers may be sterilized at 170°C for not less than 2 h.

Bacteriological determinations should be commenced as soon as possible after sample collection, preferably within 1 h and not more than 6 h after sample collection. Samples must be chilled in ice during the time between collection and filtration. However, samples must not be subjected to freezing.

Residual chlorine in a water sample will destroy the biological population and may prevent an accurate determination of bacteria in the sample, unless the chlorine is destroyed at the time of sample collection. Therefore, if a sample from a halogenated water supply is taken, 1.0 ml of 10% solution of sodium thiosulfate should be added to the 1 l sample bottle, before sterilization in the field service unit. This reagent should neutralize ~15 mg l<sup>-1</sup> residual chlorine in the sample, and should show no effect upon variability or growth.

### Sequential Triplicate Sampling

The 'sequential triplicate sampling' program is designed to assess the efficiency and effectiveness of the overall water quality program. The program serves to evaluate whether the water sampling performed is representative; it aids in discerning contamination and/or analytical problems; and reveals problems associated with water quality data management. Overall, the sequential triplicate sampling process covers the entire quality assurance spectrum for the monitoring program by providing some information on the combination of field, laboratory, and data management components. When a problem is revealed through the use of sequential triplicate sampling, other quality control measures (field and/or laboratory) can be used to elucidate and correct that problem.

Sequential triplicate sampling may not define the exact source(s) of a problem, but can be used as an aid in determining the cause(s) of the problem.

Triplicate sampling is more effective than duplicate sampling; the use of three replicates delivers more statistical information than the use of two replicates. Higher replication (quadruplicates, quintuplicates, etc.) would be more statistically useful and significant, but the increased cost may outweigh the benefit of the potential additional information. In cases where there are very specific problems (in sampling, preservation, contamination, storage, laboratory analysis, etc.), increased repeat sampling may be deemed necessary in order to provide conclusive

experimental and statistical evidence relating to the source(s) of the problems.

In order to conduct a comprehensive evaluation of the triplicate results, it is essential first to define the acceptable degree of reproducibility between triplicate samples. One accepted rule-of-thumb is to set the maximum variation between triplicate results at a level of 10%. The arbitrary limit depends to some extent on the complexity of any or all of the stages within the monitoring program for any given parameter or suite of parameters. In the case of toxic organic chemicals, the maximum limit could possibly be extended from 10% to 20% for results at sub- $\mu\text{g ml}^{-1}$  and sub- $\text{ng ml}^{-1}$  levels or 50% or greater at or near the detection limit. However, upper parameter-based limits of variation should not be chosen in an *ad hoc* fashion, but should be consistent not only with valid and documented triplication data but also with parameter-associated operational complexities.

Table 3 presents a partial record of the results from a sequential triplicate sampling project conducted by Environment Canada in its Water Quality Data programme in the period 1984 to 1985.

#### Sample Handling, Preservation, Storage, and Transportation

Efforts must be made to minimize errors that can be introduced as a result of collecting and handling the sample. The objective is to provide the laboratory with a set of samples which closely represent the aquatic environment from which they are taken.

To ensure consistency and efficiency, sample handling (filtration, decantation, centrifugation, sample splitting, etc.), preservation, storage, and transportation procedures must be properly and accurately documented, and adhered to by field personnel.

Preserving agents should be prepared from Ultrex Grade or similar grade chemicals, and must be taken to ensure that the water sample is not contaminated by impurities residing in the added preserving agent.

In adding preserving agents to field blanks, the same level of caution exercised with actual samples must be extended to blanks. The practice of adding ultrapure distilled water to the field blank bottles in the laboratory prior to the field trip should be encouraged. The preservation of blanks can then be carried out in the field.

It is necessary to pass the water sample through a filter of specified porosity, type, and quality when a determination of the concentration of dissolved inorganic constituents has to be made. Some filters and filtration apparatus may require laboratory pretreatment and must also be rinsed with a portion of

**Table 3** Partial results for sequential triplicate sampling of selected stations in an LRTAP<sup>a</sup> program (Environment Canada, 1984–1985)

NAQUADAT <sup>b</sup> no.	Sample date	Manganese (Mn) extractable	
		Actual values ( $\text{mg l}^{-1}$ )	% from median
01NS01DA0004	02 May 1984	0.02	0
01NS01DA0004	02 May 1984	0.02 <sup>c</sup>	—
01NS01DA0004	02 May 1984	0.02	0
01NS01DA0006	02 May 1984	0.02	50
01NS01DA0006	02 May 1984	0.01 <sup>c</sup>	—
01NS01DA0006	02 May 1984	0.01	0
01NS01DA0007	02 May 1984	0.02	0
01NS01DA0007	02 May 1984	0.01	50
01NS01DA0007	02 May 1984	0.02 <sup>c</sup>	—
01NS01EA0010	02 May 1984	0.02	0
01NS01EA0010	02 May 1984	0.02 <sup>c</sup>	—
01NS01EA0010	02 May 1984	0.02	0
01NS01ED0013	02 May 1984	0.01	0
01NS01ED0013	02 May 1984	0.01 <sup>c</sup>	—
01NS01ED0013	02 May 1984	0.01	0
01NS01ED0021	02 May 1984	0.01	0
01NS01ED0021	02 May 1984	0.01 <sup>c</sup>	—
01NS01ED0021	02 May 1984	0.01	0
01NS01DA0003	16 Oct 1984	0.01	0
01NS01DA0003	16 Oct 1984	0.01 <sup>c</sup>	—
01NS01DA0003	16 Oct 1984	0.01	0

<sup>a</sup>LRTAP = Long Range Transport of Air Pollutants.

<sup>b</sup>NAQUADAT = National Water Quality DATAbase.

<sup>c</sup>Median value.

Over 80% of the variations from the median values were less than 10% for the total data set, indicating the possibility for achieving substantial success with the program.

the collected sample (subsequent to rinsing with ultrapure water) before the filtrate is collected. The filtrate should be preserved (if required) by adding the proper preserving agent. The glassware utilized should be specific to the parameter that is to be analyzed. The filtering apparatus should be placed in transportation cases that permit proper storage and provide protection against shock and dirt. The transportation case can also be used as a stable work platform for filtration.

Filtration should be performed at the end of the day, to ensure the same conditions for all samples. A filtration blank for each parameter or suite of parameters should be prepared from ultrapure distilled water using the same procedure as that used for the sample. At the end of a set of filtration



runs, the glassware must be rinsed with an appropriate solution (e.g., diluted acid solution for phosphorus and metal parameters) and afterwards with ultrapure distilled water. All glassware should be rinsed with ultrapure distilled water prior to filtration and between filtration of individual samples. The glassware and filter should be rinsed by filtering the excess water, and the sample bottle used for collecting the filtrate rinsed twice with filtrate.

To prevent chemical transformation due to photo-oxidative processes and/or biologically mediated reactions, bottles containing samples should be stored in coolers during transportation. Refrigeration at 4°C should be employed for the maintenance of the quality of both preserved and unpreserved water samples. In the absence of a refrigeration unit, 'koolatrons' should be used to keep the samples cooled during transportation.

Water samples must be well stoppered and packed, to prevent spillage and/or breakage. Labels bearing the sample identification, destination, and the word 'FRAGILE' must be attached to each container. The top of the carton must be clearly identified as 'THIS END UP', and the containers in a shipment must be numbered.

A careful check must be made to ensure that all samples recorded on the field sampling sheets have been placed in a given carton, before shipping is effected. The shipping date and mode of transport must be indicated on the field sampling sheet (Figure 1).

Samples from any one location should be kept together, except in cases where all bottles of one size must be shipped together because of container size. When samples from one station must be separated and placed in more than one carton, a copy of the field sampling sheet pertaining to the bottles must be enclosed in each box.

### Chain of Custody

Complete records must be kept of every transfer of data or samples to an individual, laboratory, or storage facility. Such records will permit an investigator to determine who had custody of the material and where it was at any given time. Custody documentation is part of the support data and should be available for review by an independent auditor.

Transfer documents must be updated and maintained at a central location such as in the office of the program or project manager. These records are of great importance in establishing the validity of any questionable data.

Chain of custody procedures must assure that:

1. Only authorized personnel handle the sample.
2. Only the field sampling techniques specified for the measurement program are used.
3. A record tag is attached to the sample immediately after collection, and the tag should include the following information: program or project identification; sample field number; location; depth; collection date; time; and collector.
4. Blank samples with and without added chemicals (preserving agents) are interspersed with the actual samples.
5. All record forms are completed.
6. The transfer of samples is documented.
7. Transfer procedures provide for proper protection and preservation. For example, if samples are mailed they should be sent by certified mail with a request for an acknowledgement of receipt.

### Field Safety

Safety procedures and guidelines affecting every phase of the sampling operation must not only be documented, but must be enforced if accidents and hazardous conditions are to be avoided.

All field personnel must have a copy of the field safety manual which gives the procedures and precautions to be observed during the collection of water samples, and personnel must be provided with the appropriate safety equipment (protection hat, boots, float-jackets, security belt, etc.)

Brochures, documents, and books relating to field safety, laboratory safety, occupational health and safety, hazardous chemicals, and transportation of dangerous goods should be made readily accessible and available to all field personnel. A copy of each document should be kept in the sampling vehicle or the mobile laboratory.

### Field Audit Program

Performance and systems audits must be instituted and should become stable and continuing features in the field QA/QC program.

The systems audit should consist of a review of the total data production process, which includes on-site reviews of the field's operational systems, the physical facilities for sampling, sample handling, storage, transportation, and the measurement protocols.

The audit system should be used as one of the channels for detecting, flagging, and correcting errors or defects at any point in the field QA/QC process.

Guidelines in the audit program should specify who will conduct the audit, what protocols and procedures will be used, and to whom the audit reports

will go. A number of procedures for conducting an audit program for laboratory operations have been proposed. These procedures and guidelines are easily adaptable to the field sampling and testing operations, since there is a similarity in the mechanics of

both types of program (e.g., constructing the audit team; planning prior to the audit; developing audit methods; subjective and objective measurement procedures; evaluating and assessing the audit findings; and using the audit report to effect change).



### Training on Quality Assurance in the Laboratory

Professional training on the subject of quality and particularly on accreditation, is fundamentally important. Background training centered on aspects such as the basic concepts, objectives, and advantages of QA and the QC system, as well as on the requirements laid down in the applied Standard(s), helps in acquiring and developing skills, attitudes, and performance, and is therefore highly relevant for the success of an action plan or work plan leading to the laboratory accreditation. In addition to focusing on quality aspects, it is essential to identify what type of background training is required by different staff members in all of the technical aspects which have to be implemented within the scope of the accreditation process, such as instrument calibration, method validation, uncertainty calculations, data validation, result trend analysis, etc.

### Laboratory QCPs

QCPs are the means by which a quality assurance program (QAP) is implemented. A QAP is usually divided into two classifications, namely the intralaboratory (in-house) and interlaboratory (between laboratory) quality control programs. The former program is a continuing and systematic regime carried out by each individual laboratory. Unless several appropriate standard reference materials are available for the specific parameter, substrate, sample matrix, and concentration levels under study, an in-house QAP can only reflect the performance of each laboratory in isolation. In order to assess and/or ensure the reliability and compatibility of sets of data generated by different laboratories, an interlaboratory QAP is necessary in addition to the intralaboratory QAP.

QCPs are intended to make available to analysts a standardized approach for minimizing analytical errors and to provide the assurance of the generation of good data with the best possible precision and accuracy. This is achieved by integrating into normal laboratory practices steps which would ensure freedom from sources of error such as contamination, matrix effects, human and instrumental bias and random errors, fluctuating instrumental sensitivity, and discrepancies in analytical standards.

#### Intralaboratory QCPs

The intralaboratory quality control program consists of a number of appropriately defined procedures. For every batch of analyses for water quality samples, the

following steps should be followed:

- Before deciding to use an analytical standard, a cross check should be made with standards of other manufacturers to determine purity and/or compatibility.
- Run an appropriate reagent blank to check for contamination.
- Spike reagent blanks at various levels of concentration to check the performance of the analytical instrument.
- Spike about every tenth sample with a concentration of analyte similar to that found in the natural sample. This step checks for matrix interference effects and determines recovery.
- If matrix interference is encountered, use the method of 'standard additions' to calculate the concentrations of analyte.
- In order to check the performance of the overall procedure, prepare synthetic samples at various concentration levels. Matching the sample matrix gives added information to the analyst.
- 'Close bracketing' of standards should always be exercised in order to check on changes in sensitivity. This involves running a standard about every tenth sample, as well as running the complete calibration graph before and after the samples.
- To determine the precision of  $n$  replicates of one sample, analyze 10 or more aliquots from a given sample, and use the formula given in eqn [5].
- Run a replicate of about every tenth sample to check the precision of the system. Use the formula for paired determinations to calculate the precision (eqn [6]).
- If available, run one or more certified standard reference materials (CRMs) to check the accuracy of the overall method (use eqn [7]).
- Run secondary in-house reference materials (control samples) in order to further check on recovery. Due to the limited supply and high cost of CRMs, in-house reference standards (control samples) should be used. They are also useful in the preparation of 'quality control charts'.
- For parameters determined routinely, quality control charts provide a powerful tool to establish control of the analytical system. This technique should be used whenever feasible.
- If difficulty is encountered in drawing the analytical calibration curve by the free-hand technique, use the 'method of least squares'.

**Precision** Precision or reproducibility describes the degree of closeness of agreement between the data generated from replicate or repetitive measurements by applying the same experimental procedure several

times under prescribed conditions. Statistically the concept is referred to as dispersion and it measures the variability of the analytical method resulting from random errors. Precision is generally reported as the standard deviation (SD) or relative standard deviation (RSD). The precision of an analytical method has two components which should be recorded for each method:

1. The within-run precision ( $SD_w$ ) measures the random error during processing of a single batch of samples analyzed at the same time.
2. The between-run precision ( $SD_b$ ) measures the variability between various batches of samples analyzed.

The precision of a method is ordinarily measured by the SD, where SD is defined as

$$SD = \left[ \sum (X_i - \bar{X})^2 / (n - 1) \right]^{1/2} \quad [5]$$

where  $n$  is the number of replicate results of the sample,  $\bar{X}$  is the mean of  $n$  determinations, and  $X_i$  is the value of the  $i$ th determination.

In the case where SD is independent of concentration (i.e., SD is constant for a given range of concentration), results of duplicate analyses of different samples may also be used to calculate the precision. In such a case:

$$SD = \left[ \sum (x_i - y_i)^2 / (n - 1) \right]^{1/2} \quad [6]$$

where  $x_i$  and  $y_i$  are  $n$  paired sets of duplicate analysis results of different samples (relating to the same parameter of interest).

Often precision is expressed relative to the concentration level at which it was determined. This term is called % relative standard deviation (%RSD) or coefficient of variation (cv), and is defined as  $100 SD/\bar{X}$ .

**Total precision** In the analysis of water samples, cyclic replication (e.g., triplication) at sampling sites is used in an attempt to identify contamination problems as well as sampling errors. To distinguish between the errors occurring in the field and those occurring in the laboratory, repetitive testing is conducted in the laboratory on replicate samples and single samples. It is then possible to extract the field 'component error' from the total error, using the following approximate equation:

Field precision

$$\approx \frac{\text{total precision}}{(\text{replicate samples})} - \frac{\text{Laboratory precision}}{(\text{repeats per sample})}$$

A sample subdivided in the field and preserved separately should be used whenever possible, to assess the variability of sample handling, preservation, and storage along with the variability of the analysis process. If the nature of the matrix, the sample acquisition procedure, or the analytical technique prevents the assessment of the entire measurement system, the replicate samples used to assess precision should be selected to incorporate as much of the measurement system as possible.

**Precision and concentration** Frequently the precision of a method is not constant over the concentration range of interest. When the precision varies significantly with concentration, a plot of precision versus mean concentrations should be made in order to describe or determine the concentration dependence. If the precision varies curvilinearly with the concentration level, a graphical plot should be presented with the equation for the graph included.

**Accuracy** Accuracy refers to the correctness of the data and defines the degree of agreement of the measurements with the true value of the magnitude of the quantity concerned. Inaccuracy results from imprecision (random error) and bias (systematic error) in the measurement process.

The accuracy of a determination is expressed as % Error as follows:

$$\% \text{ Error} = (\bar{X} - X_{SRM}) \times 100 / X_{SRM}$$

where  $\bar{X}$  is the mean value of replicate determinations of a standard reference material, and  $X_{SRM}$  is the certified value of the standard reference material.

In the absence of CRMs it is possible to estimate the accuracy by determining the 'spike recovery' of a given parameter. During method validation, spike recovery is established by spiking the analyte into an appropriate matrix at a minimum of three levels which span the range of interest. Since this procedure introduces the analyte in the soluble form, the recovery does not account for extraction efficiency from solid matrices such as sediment and biological tissue or from suspended and colloidal matter in liquid samples. Recovery is calculated on a percentage basis as follows:

$$\% \text{ Recovery} = (C_{\text{found}} - \bar{C}_{\text{blank}}) \times 100 / C_{\text{added}}$$

where  $C_{\text{found}}$  is the measured concentration in the sample,  $\bar{C}_{\text{blank}}$  is the average concentration of the blank, and  $C_{\text{added}}$  is the known concentration added to the sample.

Recovery can also be monitored by spiking at least 10% of the actual samples and calculated by the

equation:

$$\% \text{ Recovery} = (C_{(\text{sample}+\text{spike})} - \bar{C}_{\text{sample}}) \times 100 / C_{\text{added}}$$

where  $C_{(\text{sample}+\text{spike})}$  is the measured concentration in the spiked sample,  $\bar{C}_{\text{sample}}$  is the average concentration of the unspiked sample, and  $C_{\text{added}}$  is the known concentration added to the sample.

**Accuracy and concentration** As with precision, the accuracy of a method may not be constant over the concentration range of interest. When a significant concentration dependence is observed, it can be reported by either a linear slope and intercept, by an alternative functional relationship, or by a graph or table.

### Interlaboratory Quality Control

Interlaboratory quality control studies are designed to determine the degree of compatibility of the data generated by participating laboratories. Furthermore, such studies help to determine interlaboratory precision and accuracy and to standardize analytical methodology. They also provide valuable data for the certification of standard reference materials.

A well-designed multisample interlaboratory quality control program provides:

- The necessary documented information to data users on the overall competence and proficiency of a laboratory through time in providing data to costly technical, environmental and health programs.
- Confidence to correlate and present future data sets from a single or different laboratory since the interlaboratory levels of confidence through time can be established.
- A neutral evaluation of the effectiveness of in-house QCPs and early warning to the operational laboratories.
- A vital mechanism to spot bias in the laboratory measurement process that is undetected in intralaboratory QC.
- Quantitative measurement of systematic bias (through the use of a series of well-designed CRMs).
- A valuable database to compare different analytical systems for the same parameters.
- Realistic criteria for acceptability of data.

### Intralaboratory and Interlaboratory Precision

For analytical procedures, precision may be specified as either intralaboratory (within-laboratory) or interlaboratory (between-laboratory) precision. Estimates

of intralaboratory precision represent the agreement expected when the same laboratory (or operator) uses the same method to make repeated measurements on the same sample. Interlaboratory precision refers to the agreement expected when two or more laboratories (or operators) analyze the same or identical samples with the same or different methods. Intralaboratory precision is more commonly reported; however, where available, both intralaboratory and interlaboratory precision should be listed in the data compilation.

At present, there is no completely valid method available to directly calculate the interlaboratory precision from the intralaboratory precision or vice versa. This is unfortunate because frequently only one type of precision estimate is available for a method. However, in general, the following comments are applicable. First, the intralaboratory precision should be smaller than the interlaboratory precision because of additional variables in the latter. Second, if the interlaboratory precision is much larger than the intralaboratory precision, this indicates that the method is very technique-sensitive. Such information can be very important when considering a project that might involve several laboratories analyzing the same sample.

A semi-empirical formula that is useful in relating the intralaboratory precision with the interlaboratory precision is represented by the following relationship:

$$S_B = [S_{\bar{x}}^2 - S_w^2/N]^{1/2} \quad [7]$$

where  $S_w$  represents the within-laboratory precision,  $S_B$  represents the between-laboratory precision,  $S_{\bar{x}}$  denotes the overall standard deviation, and  $N$  is the number of replicate observations or measurements used for finding the average (mean). The following example illustrates the relationship between  $S_w$  and  $S_B$ .

The data for a water quality sample containing arsenic as tested at three laboratories are as shown in Table 4. Assuming that the three standard deviations are estimates of one and the same population standard deviation, it is quite proper to 'pool' the variances, and take the square root of the pooled variance. Using this procedure, the best estimate of the within-laboratory standard deviation ( $S_w$ ) is obtained as

$$S_w = \{\frac{1}{3}[(1.01)^2 + (1.22)^2 + (0.51)^2]\}^{1/2} = 0.961$$

If the standard error of the mean,  $S_{\bar{x}}$ , is found for the three averages 67.45, 66.75, and 65.95,

**Table 4** Arsenic concentrations in a water quality sample

Laboratory no.	Reported concentrations	Average concentration	Standard deviation
1	66.1 68.5 67.4 67.8	67.45	1.01
2	68.4 66.1 66.9 65.6	66.75	1.22
3	65.2 66.1 66.3 66.2	65.95	0.51

we obtain

$$S_x = \left\{ \frac{1}{2} [(67.45 - 66.72)^2 + (66.75 - 66.72)^2 + (65.95 - 66.72)^2] \right\}^{1/2} = 0.751$$

If the laboratories displayed no systematic differences, the standard deviation that has been calculated from the averages (of four replicates) should be equal to  $S_w/\sqrt{4}$ , and

$$S_w/\sqrt{4} = 0.961/\sqrt{4} = 0.486$$

The fact that the calculated value of  $S_x$  is larger than  $S_w/\sqrt{4}$  can be explained only through the presence of an additional component of variability. This component, which is the between-laboratory variability ( $S_B$ ), is calculated by subtracting the anticipated  $S_w^2/N$  from the observed variance  $S_x^2$ , and taking the square root. That is,

$$S_B = [(0.751)^2 - (0.486)^2]^{1/2} = 0.573$$

In general, if  $S_x$  and  $S_w$  are known, then  $S_B$  can be easily calculated using the formula of eqn [7].

### Detection Limits

The limit of detection (LOD) is one of the most important terms used for comparing various analytical procedures, techniques, or instruments. It is defined as being the lowest concentration of the analyte that can be distinguished with reasonable confidence from the blank or background. In water quality laboratories the confidence level of 95% is adopted as a standard for LOD for all analytes.

Different statistical approaches are reported in the literature to calculate limits of detection. The LOD for a determinant can easily vary through the use of

different statistical approaches. This problem is well documented in the literature.

The LODs can be represented in three different forms depending upon the requirements. The three representations are (1) the instrument detection limit (IDL), (2) the method detection limit (MDL), and (3) the practical detection limit (PDL).

**Instrument detection limit (IDL)** The IDL is the lowest concentration of analyte that an analytical instrument can detect and which is statistically different from the response obtained from the background instrumental noise. The IDL is established by adding the analyte in reagent (blank) water or appropriate organic solvent to give a final concentration within five times the estimated IDL and calculating the standard deviation by introducing the solution directly into the instrumental system to obtain seven or more replicate measurements. The IDL is then calculated using the 95% confidence level as follows:

$$IDL = t_{(n-1)} \times SD \quad \text{for } n \geq 7$$

where  $t_{(n-1)}$  is the value for a one-sided Student's  $t$ -distribution for  $n - 1$  degrees of freedom.

The IDL should be used to indicate the absolute LOD of the analytical technique and/or instrument.

**Method detection limit (MDL)** The MDL is the lowest concentration of analyte in distilled and/or deionized water that a method can detect reliably and is statistically different from the response obtained from a blank carried through the complete method including chemical extraction or pretreatment of the sample. When the MDL is experimentally evaluated for each matrix by the analysis of samples or spiked samples, then the PDL is determined (see below).

When repeated analyses of blanks show a positive response for the analyte, the MDL is defined as:

$$MDL = \bar{S}_b + t_{(n-1)} \times SD$$

where  $\bar{S}_b$  is the average signal (or level) for the blanks,  $SD$  is the standard deviation of the replicate determinations, and  $t_{(n-1)}$  is the one-sided Student's  $t$ -distribution for  $n - 1$  degrees of freedom at a confidence level of 95%.

**Practical detection limit (PDL)** The PDL is the lowest concentration of analyte in a real sample matrix that a method can detect reliably and is statistically different from the response obtained from a blank

carried through the complete method. It is calculated in the same manner as the MDL above.

For a specified method and analyte, the PDL will vary with different sample matrices since these may affect reproducibility, blanks, and interference levels. It is always  $\geq$  MDL.

**Limit of quantification (LOQ)** The LOQ is defined as follows:

$$\text{LOQ} = \bar{S}_b + 10 \text{ SD}$$

where  $\bar{S}_b$  is the average signal (or level) for the blank and SD is the standard deviation of the replicate determinations. This defines the level above which quantification is reliable and also a region between the MDL and the LOQ, where detection is reliable but quantification is not. The LOQ is the level above which quantitative results may be obtained with a specified degree of confidence.

### Reporting of Analytical Results

In no case should the term 0 (zero) be used to report an analytical result. In all cases for which no analyte was detected, the notations 'ND' (none detected) and 'L' (less than) should be used to report the result. The value following L is the MDL. For example, L 0.001 means that the analyte has been determined at a level less than the MDL of  $0.001 \mu\text{g ml}^{-1}$  (if the unit of measurement is in  $\mu\text{g ml}^{-1}$ ). If the result falls between the MDL and the PDL, the value may be reported in parentheses, with an explanation that the result is less than the PDL and the recovery and precision have not been evaluated at that level.

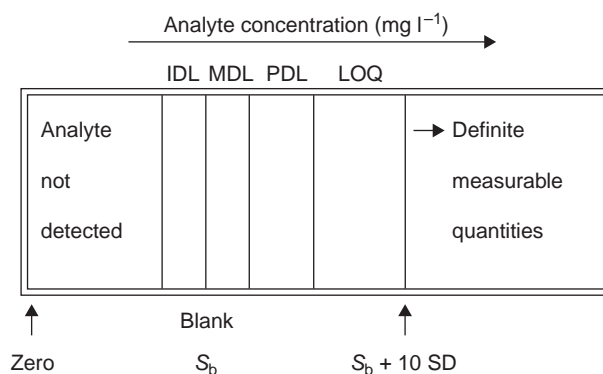
In some cases, data and analyte recovery or concentration may be reported down to instrumental detection limit values (left-censored data) provided that the necessary explanation accompanies the results, and indicates changes in sample size, final aliquots prior to analysis, or method quantification including the use of internal standards or surrogates.

Figure 2 shows a graphic summary in which the IDL, MDL, PDL, and LOQ are interconnected.

## Data Handling and Data Management

### Data Quality

The quality of data is to a large extent evaluated on the basis of their uncertainty when compared to end-use requirements. When data are accompanied by consistency, and their uncertainty is small when compared to the requirements, the data are considered to be of adequate quality. When excessively variable, or the level of uncertainty exceeds the



**Figure 2** Interconnection between IDL, MDL, PDL, and LOQ.

needs, the data may be said to be of low or inadequate quality. The assessment of data quality can thus be a relative determination. What is high quality for one situation could be unacceptable in another.

The virtue of water quality data may be assessed on the basis of two aspects: the accuracy of identification of the parameter or variable measured and the numerical accuracy. The qualitative identification must be made without reasonable doubt, and the quantitative measurements must be conducted precisely and accurately. Quantitative measurements must be made in such a manner that any error or uncertainty in the measurements can be tagged with a stated probability. To this end, measurements must be made in such a way as to provide statistical predictability.

Data quality assurance must deal with the aspects of data handling and data management shown in Table 5.

### Auxiliary Data

The practice of collecting supporting information during all phases of the measurement programs must be encouraged, and this auxiliary data must be easily available via electronic data processing facilities or field and laboratory records. The use of auxiliary data may become necessary during the data interpretation process.

The following information may be considered as auxiliary data:

- data charts and printouts;
- equipment performance records;
- calibration records;
- operation logs;
- environment conditions prior to and during sampling;
- measurement comparison records;
- quality control and system audit records; and
- records of corrective actions.



**Table 5** Data handling and data management

<i>Data recording and documentation</i>	
Data acquisition	Handwritten data
Transcription errors	Data storage
Data processing	Data retrieval
Computerized recorded data	
<i>Data transmission</i>	
Appropriate data format	Confidentiality
Clarity and comprehensiveness	Interlaboratory and interagency correspondence
Turn-around times adherence	
<i>Data validation and verification</i>	
Data inspection	Intermethod comparisons
Statistical testing	Interinstrument comparisons
Internal consistency checks (data plots; Dixon Ratio tests; Grubbs test; the Gap test; regression analysis; control charting; etc.)	Support data records
Control limits; quality control charts	Sample consistency (ion balances; Anion/cation ratio; mass balances; etc.)
Detection limits	Data comparability and compatibility
Data reduction	Calculation errors
Data corrections	Decimalization errors
Transmittal errors	Appropriate measurement units
Data acceptance/rejection criteria	Relationship between data results and programme or project
Data plots	Objectives
Test for suitability of data	Water quality objectives
Regression analysis	Sampling records
Test for goodness of fit and for outliers	Transportation and preservation records
	Analytical records
	Investigation of flagged questionable data
<i>Data analysis</i>	
Use of appropriate statistical techniques	Data representativeness
Data interpretation	Data comparison and compatibility with specified external
Data integrity	Data sets
Accuracy	Quality control charts
Precision	Data quality review and assessment
Completeness	Significant quality assurance problems
Method detection limit	Corrective action procedures
	Audit checks of data
<i>Data reporting</i>	
Performance evaluation reports	Responsibility centres
Data reports	Confidentiality
Interpretive reports	Provisional reports
Data presentation	Final reports
Turn-around times	Feedback from clients and data users, <i>re</i> data reports
Report control procedures	

Auxiliary data should be collected throughout the measurement program and reviewed periodically.

Auxiliary data are important in determining the validity of the measurement program data. For example, auxiliary or support data could be used in deciding whether or not an outlier is a valid value or an artifact. Ideally, unusual conditions occurring during the field sampling and laboratory analysis operations should be recorded on the field and laboratory data reports. Failure to report such occurrences may place the acceptance of important data in jeopardy, or may lead to the rejection of that data.

*See also:* **Extraction:** Solvent Extraction Principles. **Pesticides. Quality Assurance:** Quality Control; Reference Materials. **Sampling:** Theory. **Water Analysis:** Overview.

## Further Reading

- Agemian H (1988) *Quality Assurance in the National Water Quality Laboratory*. National Water Quality Laboratory. Burlington, ON, Canada: Water Quality Branch.
- Arseneault R and Howell G (1987) *The Water Quality Branch Atlantic Region Field Quality Assurance/Quality Control Program*, Version 2. Moncton, NB, Canada: Water Quality Branch, Inland Waters Directorate, Environment Canada.
- Desilets L (1986) *Water Quality Branch Strategy for Assessments of Aquatic Environmental Quality and Criteria for Basin Selection and Sampling Station Macro-Location*. Ottawa-Hull, Canada: Water Quality Branch.
- Environment Canada (1983) *Sampling for Water Quality*. Ottawa-Hull, Canada: Water Quality Branch, Environment Canada.



- Garfield FM (1984) *Quality Assurance Principles for Analytical Laboratories*. Arlington, VA: Association of Official Analytical Chemists.
- Gaskin JE (1988) *Quality Assurances Principles and Guidelines for Water Quality Sampling and Quality Assurance Principles and Guidelines for the Handling and Management of Water Quality Data*. Hull, QC, Canada: Water Quality Branch, Inland Waters Directorate, Conservation and Protection, Environment Canada.
- Gaskin JE (1993) *Quality Assurance in Water Quality Monitoring*. Ottawa: Canada Communications Group.
- Gaskin JE, Dafoe T, and Brooksbank P (1990) Estimation of analytical values from subdetection limit measurements for water quality parameters. *Analyst* 115: 507–510.
- Keith LH (1991) *Environmental Sampling and Analysis – Practical Guide*. Chelsea, MI: Lewis Publishers.
- Kent RT and Payne EK (1988) Sampling groundwater monitoring wells: special quality assurance and quality control procedures. In: Keith LH (ed.) *Principles of Environmental Sampling*, p. 231. Washington, DC: American Chemical Society.
- Kister LR and Garrett WB (1983) *Field Guidelines for Collection, Treatment and Analysis of Water Samples*. Arizona District, AZ, USA.
- Murdock A and MacKnight S (1991) *Handbook of Technique for Aquatic Sediment Sampling*. Boca Raton, FL: CRC Publisher.
- Quevauviller Ph (2002) *Quality Assurance for Water Analysis*. Chichester: Wiley.
- Rhodes RC and Hochheiser S (1971) *Quality Costs for Environmental Monitoring Systems*. American Society for Quality control, Technical Conference, vol. 77, p. 151.
- Sanders TG, Ward RC, Loftis JC, et al. (1983) *Design of Networks for Monitoring Water Quality*. Littleton, CO: Water Resources Publications.
- Stanley TW and Verner S (1985) The US Environmental Protection Agency's Quality Assurance Program. In: Taylor JK and Stanley TW (eds.) *Quality Assurance for Environmental Measurements*, ASTM STP 867. Philadelphia: American Society for Testing and Materials.
- Taylor JK (1987) *Quality Assurance of Chemical Measurements*. Chelsea, MI: Lewis Publishers.
- Taylor JK and Stanley TW (eds.) (1985) *Quality Assurance for Environmental Measurements*, ASTM STP 867. Philadelphia: American Society for Testing and Materials.

# R

## RADIOCHEMICAL METHODS

Contents

**Overview**

**Natural and Artificial Radioactivity**

**Radionuclide Monitoring**

**Technetium**

**Radon**

**Uranium**

**Radiotracers**

**Radio-Reagent Methods**

**Radioreceptor Assays**

**Gamma-Ray Spectrometry**

**Food and Environmental Applications**

**Pharmaceutical Applications**

### Overview

**K Buchtela**, Atominstitut der Österreichischen Universitäten, Vienna, Austria

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The application of radioactive materials in analytical chemistry is based on two outstanding features of radioactivity: the ease of measurement of the emitted radiation and the possibility of using radioactive isotopes as markers for labeling chemical elements or chemical compounds. Often, higher sensitivities in analytical determinations can be achieved using radiochemical methods or nuclear techniques than applying other analytical methods. The main disadvantages of the application of radiochemical methods are related to the problems connected with the handling of radioactive materials and the limited availability of suitable radioisotopes for some chemical elements.

Radiochemical methods in analytical chemistry are used for the qualitative and quantitative determination of most of the chemical elements or chemical species, or for the location or pathway of substances

in a system under investigation. Radioactive materials are also applied for quality assurance, improvement, and development of analytical procedures. To achieve these goals, radiochemical methods have been developed and applied. In fact, even the discovery of the radioactive elements polonium and radium (1898) by the pioneering work of Marie and Pierre Curie involved the application of radiochemical methods, the application of nonisotopic carrier materials. Another radiochemical method was reported a few years later by Hevesy and Paneth (1913) in a study of the solubility of lead sulfide; they used the natural lead isotope  $^{210}\text{Pb}$  as radioactive tracer. Isotope dilution was introduced by Hahn (1923) to determine the yield for the separation of isolated protactinium-234 using protactinium-231. The technique of neutron activation analysis had been applied by Hevesy and Levi after the discovery of the neutron and the possibility of nuclear reactions to produce radioactive specimens from stable elements. In 1938, the application of radiochemical methods saw the use of nonisotopic and isotopic carriers by Hahn and Strassmann resulting in the discovery of nuclear fission. Radiochemical methods also played an important role in the research work related to transuranium elements.

During the last years radiochemical methods for routine analysis seem to have lost ground to other,

nonradiochemical methods, primarily spectrometric, but because of their reliability they are indispensable and extremely useful for the certification of standard reference materials. Therefore, they play an important role in the field of quality control and quality assurance in analytic chemistry. There is also no doubt that the modern standard of the application of short-living radionuclides in diagnostic nuclear medicine depends on the application of radiochemical methods.

The applicability of radioactive substances in analytical chemistry depends on the properties of the radioactive material (type, energy, and half-life of the radioactive atoms), the availability of the radioactive material, and the availability of suitable radiation detection methods.

Some basic knowledge of the structure and rearrangement procedures of unstable atoms, properties of radiation, characteristics of radiation detectors, and production of artificial radionuclides is helpful for the understanding of radiochemical methods and their application in analytical chemistry. Also, it has to be pointed out that for the use of radioactive materials not only do the principles of radiation protection have to be observed but also one has to follow strictly those rules that depend on the legislation of the relevant country.

## Radioactive Atoms

Atoms, whether existing in nature or artificially produced, are made up of protons, neutrons, and electrons. The positively charged protons and the neutrons, which are particles without electrical charge, form the nucleus, and with the negatively charged electrons build up the shell of the atom. The overall number of particles and the ratio of protons to neutrons are of significance to the stability of an atomic nucleus. If the nucleus is too large or if there are either too few or too many neutrons for a given number of protons, the atomic nucleus turns out to be unstable. In an attempt to reorganize itself to a stable

configuration some rearrangement processes will take place in the atomic nucleus, which are accompanied by the release of particles or electromagnetic waves or both. Such nuclei are said to be unstable. The rearrangement process is usually called radioactive decay. Until now ~2000 such types of atoms, called radionuclides, have been identified.

## Types of Radioactive Decay

The various decay processes are listed in Table 1. Radioactive nuclides emit either nucleons (alpha particles, very rarely protons or neutrons) or electrons (negatrons, positrons). As an alternative to the emission of a positron, a proton may capture an electron of the K-shell (K-capture). By the emission of an alpha particle the mass number and the atomic number are reduced; by the emission of electrons either the number of neutrons ( $\beta^-$ -decay, negatron emission) or the number of protons ( $\beta^+$ -decay, positron emission) is reduced. By K-capture also the number of protons is reduced. Due to the missing electron in the K-shell, characteristic X-rays of the newly produced atomic species are emitted.

In most cases the emission of nucleons and electrons leads to an excited state of the resulting new nucleus. The excitation energy is given off in the form of one or several gamma-ray photons. In most cases this de-excitation takes place within an extremely short period of time (less than  $10^{-12}$  s) but in some cases this transition is delayed. Such a type is called an isomeric state of a nuclide.

Some very heavy nuclides have the tendency toward spontaneous fission, they decay to two (rarely to three) smaller nuclei.

## Rate of Radioactive Decay

Although it is known that the stabilizing rearrangement or decay of an unstable (radioactive) atom will occur sometimes, it is not known precisely when this will occur. For a statistically significant number of

**Table 1** Some decay processes

Decay	Radiation	Symbol	Example	Remarks
$\beta^-$ decay	Negatrons	$\beta^-$	$^{14}\text{C} \rightarrow ^{14}\text{N} + \beta^-$	Mainly for elements with high atomic number ( $Z > 83$ ) $\sim 10^{-12}$ s after $\alpha$ or $\beta$ decay excitation energy is given off
$\beta^+$ decay	Positrons	$\beta^+$	$^{11}\text{C} \rightarrow ^{11}\text{B} + \beta^+$	
Electron capture	Characteristic X-rays	$\epsilon$	$^{85}\text{Sr} + e^- \rightarrow ^{85}\text{Rb}$	
Alpha decay	Helium nuclei	$\alpha$	$^{226}\text{Ra} \rightarrow ^{222}\text{Rn} + \alpha$	
Gamma decay	Photons	$\gamma$	$^{60\text{m}}\text{Ni} \rightarrow ^{60}\text{Ni} + \gamma$	Delayed giving off of excitation energy
Isomeric transition	Photons	$\gamma$	$^{137\text{m}}\text{Ba} \rightarrow ^{137}\text{Ba} + \gamma$	Mainly for very high atomic numbers (atomic mass $> 240$ )
Spontaneous fission	Nuclei	$f$	$^{252}\text{Cf} \rightarrow \text{fission products}$	

unstable nuclei the rearrangement process will occur at a definite and predictable rate. This decay rate is characteristic for the radioactive nuclide under consideration.

Let  $N$  be the number of unstable (radioactive) atoms of a given radionuclide present at any time  $t$ . The change in the number of unstable atoms  $N$  per unit of time at any moment is proportional to the number of atoms present at that moment:

$$\frac{dN}{dt} = -\lambda N \quad [1]$$

where  $\lambda$  is the proportionality constant, termed the decay constant. The number  $N$  decreases with time; therefore, the negative sign is used. The decay constant can be understood as the fraction of the number of radioactive atoms decaying per unit of time at any moment:

$$\lambda = \frac{1}{N} \frac{dN}{dt} \quad [2]$$

Equation [1] may be integrated to give

$$N = N_0 e^{-\lambda t} \quad [3]$$

where  $N_0$  is the number of atoms present at any starting time ( $t=0$ ), and  $N$  is the number of unchanged atoms remaining after a period of time  $t$ .

The rate of decay,  $dN/dt$ , is termed the radioactivity, or simply activity ( $A$ ). From eqns [1] and [2], we can obtain

$$A = \lambda N, \quad A = A_0 e^{-\lambda t} \quad [4]$$

A convenient method for describing the radioactive decay is in terms of 'half-life', which is the time required for one-half of any starting amount of a radionuclide to undergo rearrangement. At the end of each half-life interval one-half of the starting material will be left unchanged. Also, the 'activity' (the rate of decay, the number of decaying atoms per unit of time) will be one-half of the initial activity. Radionuclides may have half-lives from fractions of seconds to billions of years.

## Units of Radioactivity

The unit of activity is the becquerel (Bq), named after Antoine Henry Becquerel, who discovered the phenomenon of radioactivity. This unit is defined as

$$1 \text{ Bq} = 1 \text{ disintegration per second}$$

In some countries an old special unit of activity is still used, the curie (abbreviated Ci). This was originally

defined as the radioactivity associated with 1 g of radium. According to a formal definition agreed upon in 1964:

$$1 \text{ Ci} = 3.7 \times 10^{10} \text{ disintegrations per second}$$

## Energy of Radiation

The energies of particles and waves emitted by radionuclides have characteristic values. The energies of alpha particles, characteristic gamma and X-rays are constant or discrete; the energies of beta particles ejected by a given radionuclide vary from zero up to a certain maximum value. This is because a variable part of the energy is carried by a neutrino that is emitted together with the beta particle. As a consequence, beta particles show a continuous energy spectrum from zero to a characteristic maximum energy.

As a unit of radiation energy the electron volt (eV) is used. Radiation energies are in the region of hundreds of keV to several MeV:

$$1 \text{ MeV} = 1.602 \times 10^{-13} \text{ J}$$

## Naturally Occurring and Artificially Produced Radionuclides

### Naturally Occurring Radionuclides

The elements with an atomic number higher than 83 do not have stable isotopes: they are called radioelements.

The naturally occurring radioelements are mainly found in thorium and uranium ores. Uranium-238, uranium-235, and thorium-232 have long half-life ( $^{238}\text{U}$ :  $4.5 \times 10^9$  years,  $^{235}\text{U}$ :  $7.04 \times 10^8$  years,  $^{232}\text{Th}$ :  $1.4 \times 10^{10}$  years). Otherwise, they would not exist in nature today because some  $10^9$  years have already passed since the origin of the elements. Some of the radioactive decay products of these three long-living radioelements have much shorter half-lives, they are always newly produced by the radioactive decay of the long-living 'parent radionuclides'. The stable end products of these naturally occurring radioactive decay series are  $^{206}\text{Pb}$  (for  $^{238}\text{U}$ ),  $^{207}\text{Pb}$  (for  $^{235}\text{U}$ ), and  $^{208}\text{Pb}$  (for  $^{232}\text{Th}$ ). There was a fourth decay series in nature, which is 'extinct' now because of the rather short half-life of the parent radionuclide ( $^{237}\text{Np}$ :  $2.14 \times 10^6$  years).

There are a number of other elements that also have naturally occurring radioactive isotopes besides the stable ones. They also have very long half-lives, similar to uranium and thorium; some of them are listed in Table 2.

**Table 2** Some long-living natural radionuclides

Radionuclide	Half-life (years)	Radiation	Isotopic abundance (%)
$^{40}\text{K}$	$1.3 \times 10^9$	$\beta^-$ , $\beta^+$ , $\varepsilon$ , $\gamma$	0.012
$^{87}\text{Rb}$	$4.7 \times 10^{10}$	$\beta^-$	27.8
$^{190}\text{Pt}$	$6.1 \times 10^{11}$	$\alpha$	0.0127
$^{187}\text{Re}$	$5 \times 10^{10}$	$\beta^-$	62.6
$^{176}\text{Lu}$	$3.6 \times 10^{10}$	$\beta^-$	2.6

These natural radionuclides are not daughter products of the uranium and thorium series.

**Table 3** Some naturally produced radionuclides

Radionuclide	Half-life	Radiation
$^{14}\text{C}$	5730 years	$\beta^-$
$^{10}\text{Be}$	$1.6 \times 10^6$ years	$\beta^-$
$^7\text{Be}$	53.4 days	$\varepsilon$ , $\gamma$
$^3\text{H}$	12.3 years	$\beta^-$

All these radionuclides are produced in the atmosphere by cosmic radiation.

Some radioactive nuclides are produced continuously by nuclear reactions caused by cosmic radiation. They are listed in Table 3. Some of them are of interest because they can be used for the determination of the 'age' of a sample material.

### Artificially Produced Radionuclides

Man has artificially produced radioelements (technetium, promethium, transuranium elements) and also many radioactive isotopes of the naturally occurring elements. Thus, the natural radioactivity on the earth is to some extent increased by man-made radioactive materials. Many analytical research works focus on the determination of man-made radionuclides, their migration, pathways, and accumulation in the environment. The aim of this analytical research work is either to know about fate, pathway, and metabolism of materials or to obtain information about possible environmental protection problems.

Radionuclides are produced by nuclear reactions. These nuclear reactions can take place between highly accelerated atomic nuclei or between nuclei and nucleons (e.g., protons). To overcome the Coulomb barrier the charged nuclei or nucleons have to be provided with high kinetic energies. This is achieved by the use of accelerator systems such as cyclotrons, synchrotrons, and linear accelerators. Frequently nuclear reactors are used for radionuclide production because the uncharged neutrons

present in large amount in nuclear reactors are interact easily with the target nuclei.

By nuclear reactions nonradioactive elements can be transformed to a radioactive species to enable their sensitive qualitative and quantitative determination (activation analysis).

The yield of such a nuclear reaction, expressed by the resulting activity of the product, is given by

$$A = N_T \Phi \sigma (1 - e^{-\lambda t}) \quad [5]$$

where  $A$  is the resulting activity,  $N_T$  the number of target atoms,  $\Phi$  the flux density of particles bombarding the target per unit of time and area,  $\sigma$  the cross-section of the target nuclei,  $\lambda$  the decay constant of the product, and  $t$  the reaction time.

After the nuclear reaction, the target material frequently has to be purified to remove by-products. Then, the material is used for the production and synthesis of 'radiochemicals'. The resulting materials are used for application in nuclear medicine, research, and industry.

Radiochemicals are available from a series of companies. Usually the purchase, handling, and use are subjected to licensing procedures. Details of such licensing procedures depend on the legislation of the country and have to be strictly observed.

## Radiation Detection Methods

The radiation from radionuclides interacts with matter and causes ionization, excitation, or chemical changes. These effects are utilized in the methods of radiation detection and measurement. Among them the most commonly used effects are the ionization in gases, the interaction of radiation with semiconducting materials, the orbital electron excitation in solids and liquids, and the specific chemical reactions in sensitive emulsions.

Radiation detectors are specially designed to provide information about the type (alpha, beta, gamma, etc.), energy (radiation spectrometry), and intensity (number of particles or quanta) of the radiation. Some detectors also provide information about the spatial distribution of the radiation (nuclear imaging detectors).

There is a wide and excellent selection of literature available related to radiation detection methods in the field of radioactivity analysis.

### Gas Ionization Detectors

Gas ionization detectors consist of a gas volume in an enclosure that is either sealed or constructed in such a way as to permit a continuous flow of the filling gas. The system is equipped with electrodes. The

outer wall frequently serves as the cathode, while a wire rod, a plate, or a grid in the middle of the gas volume serves as the anode. An electric field is applied across the electrodes. Ions and electrons formed by the interaction of the radiation with the gas molecules are collected at the electrodes producing an electric current that serves as the basis for the detection of the radiation. Special types of gas ionization counters depend on the field strength created by the applied voltage. In ionization chambers the field strength is just high enough to collect all the ions and electrons created by the interaction of the radiation with the gas; in proportional counters the ions are accelerated by the higher field strength and can create additional ions (ion multiplication). In Geiger-Müller tubes the field strength is very high and an ion avalanche of maximum size is formed which is independent of the primary ionizing effect produced by the radiation and is mainly governed by the counter tube construction. The electronic response of ion chambers and proportional counters depend on the energy deposited in the gas volume by the radiation. These counters provide some information about the type and energy of the radiation. They can also be constructed as imaging detectors. Geiger-Müller counters do not provide any information about the primary ionization events, they only register the presence of radiation without any additional information.

### Scintillation Detectors

In scintillation detectors the emission of photons due to the interaction of the radiation with the detector forms the basis of detection. The photons are registered and processed by a photosensitive device like a photomultiplier. By choice of suitable scintillating materials information about the type of radiation can be obtained. The number of photons created by the interaction of particles or quanta with the scintillation material depends on the energy of the radiation, by which radiation spectrometry can be performed.

### Semiconductor Detectors

Semiconductor radiation detectors are crystals whose electrical conduction is altered by the absorbed radiation. Their operation depends on their semiconducting properties. They can provide information about type, energy, and intensity of radiation and are the most important devices for radiation spectrometry.

### Photographic Emulsions

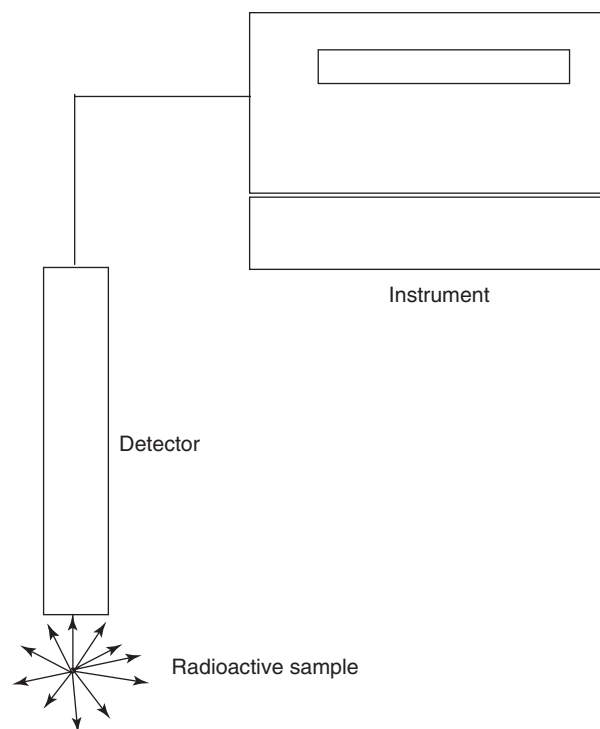
If radiation passes through a photographic emulsion, chemical changes in the material are caused by ionization. As in normal photography these changes

can be made visible by development procedures. In a microscopic scale tracks of those particles emitted by a radioactive material can be made visible. At a macroscopic scale radioactive material being incorporated in a sample material can be localized. This autoradiographic method is therefore used for localization of radioactive spots at a chromatogram and has found wide application in biological and biochemical research.

### Instrumentation

Besides a radiation detector a monitoring or measuring set-up includes several electronic units, for example, a power unit for the supply of the high voltage, an amplifier system for the amplification of the small electric current produced initially by the detector, a timing unit for running the counting system for a predetermined time, a pulse-height analyzer to sort the incoming electronic information for radiation spectrometry, etc.

A schematic diagram of a radiation measurement set-up is shown in **Figure 1**. The information displayed in the instrument shows the number of events registered by the detector (number of 'counts'),



**Figure 1** Schematic picture of a radioactivity measurement set-up. The indication at the instrument (number of 'counts') depends not only on the intensity of the radiation (given in Bq) but also on the 'counting geometry' (shape and size of radioactive sample, distance, absorption and scattering effects, etc.). In addition, type and energy of the radiation and suitability of the detector plays an important role in the 'counting efficiency'.



which is proportional but in most cases not identical with the number of emitted particles or rays of the radioactive material and depends on the counting efficiency of the system. This efficiency refers to the number of particles or photons emitted by a radiation source related to the number of interactions registered by the counting system. To measure the activity of a radiation source in Bq, the decay scheme of the radionuclide and the efficiency of the counting system must be known. The counting efficiency depends on the type and energy of the radiation, the size and shape of the sample, the distance between the detector and the radioactive source, the absorption and scattering of the radiation within the sample and by the materials between sample, and finally also on the detector properties. Calibration of counting systems is important if absolute activities have to be determined in Bq. In the application of radiochemical methods measurements are frequently performed to compare the activity of radioactive samples. In this case the knowledge of the efficiency is not necessary but it must be kept constant for all samples to be measured during the investigation.

### Background Activity

It has to be pointed out that radioactivity is a natural phenomenon, so there is always some low-level activity detected by the counting system. This activity is due to cosmic radiation and terrestrial radioactive materials. This activity is referred to as the background. This background counting rate of a measurement system must be subtracted from the gross counting level in order to obtain the net counting rate of the sample.

### Statistics

Because of the random nature of the radioactive decay, the laws of statistics also have to be applied for the interpretation of radioactivity measurements.

A minimum number of disintegrations (counts) has to be observed for a reliable quantitative determination. Therefore, either a minimum concentration of radioactive material is necessary for a reliable measurement or a minimum measurement time to collect a statistical reliable result is necessary. Usually, results are given including the standard deviation (68% probability that the 'true value' is within the given limits). Detailed information is available in the literature about radioassay techniques.

### Activity and Mass of Radionuclides

Because of the emitted radiation radioactive materials can be detected even when they are present in

**Table 4** Some examples for the mass of 10 Bq of carrier-free radionuclides

Radionuclide	Half-life	Mass (g)
$^{99m}\text{Tc}$	6.6 hours	$6 \times 10^{-17}$
$^{131}\text{I}$	8.06 days	$2 \times 10^{-15}$
$^3\text{H}$	12.3 years	$3 \times 10^{-14}$
$^{137}\text{Cs}$	30.4 years	$3 \times 10^{-12}$
$^{14}\text{C}$	5730 years	$6 \times 10^{-11}$

very low concentrations. Because of this phenomenon and because of the problems of radiation protection the actual amounts of radioactive materials used in normal laboratory experiments are usually extremely small. The mass  $m$  of a radioactive material is given by the following equation which can be derived from the mathematical expression for the radioactive decay (see eqn [2]):

$$m = \frac{AT_{1/2}M}{N_A \ln 2} \quad [6]$$

where  $m$  is the mass of the carrier-free radionuclide in g,  $A$  the activity of the radionuclide in Bq,  $T_{1/2}$  the half-life of the radionuclide in s,  $M$  the relative atomic mass of the radionuclide, and  $N_A$  is Avogadro's number.

Table 4 gives some figures for the mass of carrier-free radionuclides with an activity of 10 Bq.

### Isotopic and Nonisotopic Carrier Materials

Because of the small amounts of radioactive materials there are special problems in handling such materials. Adsorption phenomena on surfaces, such as sample containers, may result in losses of radioactive materials. Conventional separation procedures involving precipitation cannot be carried out with materials of such low concentrations.

Because the actual amounts of radioactive materials used in normal laboratory experiments are usually very small a chemical 'carrier' is added. This is a nonradioactive isotope of the radioactive material. This allows conventional chemical operations like precipitation. Care must be taken that the carrier compound is chemically identical to the compound of the radionuclide or equilibrates with the chemical species of the radioactive material.

If a compound of a nonradioactive isotope of the radionuclide is added, this type of carrier is called 'isotopic carrier'. Later on the radioactive species cannot be recovered by chemical separation from the nonradioactive carrier material because the

radioactive species and the nonradioactive carrier show exactly the same chemical behavior.

After addition of a 'nonisotopic carrier' material (the same chemical compound of an element with similar chemical behavior of the radionuclide), the radionuclide and the carrier show the same chemical behavior only in some chemical procedures. This can be used for separation processes. Later on the radionuclide can be separated from the nonisotopic carrier material by specific chemical separation procedures and recovered as a carrier-free radionuclide. Using this radiochemical method Marie Curie discovered radium by first using the nonisotopic carrier barium and finally isolating radium from barium by separately crystallizing barium and radium bromides.

## Specific Activity

The term specific activity (activity per unit mass) is important for radiochemical methods like isotope dilution. Samples of carrier-free radionuclides have the highest possible specific activity for this radionuclide. By the addition of carrier material the specific activity is lower, the total amount of material (stable and radioactive materials) is increased, and some handling procedures, like precipitation, might be easier. The change of specific activity is the basis of isotopic dilution techniques.

## Radioactive Tracers

Tracers are materials that are used as markers to show the location of a substance or to follow the pathway of a substance in a chemical reaction or physical process. Such tracers have to show the same physical and chemical behavior in the system under observation as the material that is actually observed. Radioactive isotopes are ideal tracers for a nonradioactive isotope which has to be traced because they show the same chemical behavior like the nonradioactive material. It is the activity of the radioisotope that is monitored to follow the process under investigation. In addition, the amount of radioactive material to be added as a marker can be kept extremely small. This is of great importance if biological processes are investigated because the amount of tracer material can be kept low enough not to interfere in physiological processes. The most important assumption made in the use of radiotracers is that the radioactive material will blend in perfectly with the system under study and that the emitted radiation does not affect any components of the system.

## Isotope Dilution Analysis

Isotope dilution analysis is a quantitative analytical technique. The basis of this radiochemical method is that by chemical processing the specific activity (activity per unit mass) of a mixture of stable and radioactive isotopes is not changed.

Direct isotope dilution analysis is applied if an amount of an analyte cannot be separated quantitatively for analytical determination. A known amount of a radioactive isotope of the element of interest is added to the sample containing the analyte. Then a portion of the analyte is isolated in high purity from the sample. This separation step need not be quantitative. The mass and activity of the isolated portion are measured and used to calculate the amount of analyte in the original sample. There are several varieties known of this radiochemical method, e.g., reverse isotopic dilution.

## Radio Reagent Method

In radio reagent methods a radioactive species is used in a quantitative reaction and the change in activity of that species in the course of the reaction is measured. The radioactive species may be a labeled reagent or the analyte.

After separation from excess reagent by any suitable chemical separation method like liquid-liquid distribution, chromatography, etc., the mass or concentration of this product is determined from activity measurement.

The advantage of radio reagent methods over classical analytical techniques arises from the high sensitivity of the activity measurements that are not subject to interference by other substances. The principle of the radio reagent method can be adapted to various procedures.

## Activation Analysis

In classical activation analysis a material to be analyzed is bombarded with neutrons, charged particles, or photons (gamma-rays). By this bombardment radionuclides are produced from elements of the target material. These radionuclides can be analyzed qualitatively and quantitatively by radioassay methods. From the results and the knowledge of the nuclear reactions during the bombardment an analytical determination of the target material can be achieved. The radioactive products of the bombardment can be measured after the irradiation and the emitted types, energies, and half-lives provide information used for qualitative analysis and the radiation intensity supplies data for the quantitative composition of the material to be analyzed.

In prompt gamma-ray activation analysis the measurement is done simultaneously with the bombardment, and excited intermediates of the nuclear reactions emitting the so-called prompt gamma-rays are determined. The characteristic wavelength (gamma-energy) of this emission can be used for qualitative analysis and their intensity for quantitative analysis of the target material.

Nuclear activation methods can provide high sensitivities for many but not all elements. These methods are capable of simultaneous multielement analysis and in some cases they may be essentially nondestructive analytical methods. One of the main disadvantages is the lack of information on chemical form (speciation) of the elements in the analyte. Neutron activation analysis is the most common form of activation analysis, but charged particle and photon activation analysis methods are also applied.

## Aspects of Safety Regulations

Work with radioactive materials is subjected to licensing procedures that may differ from country to country. Therefore, it is of utmost importance to become familiar with the regulations being issued by the relevant competent authority. Governmental regulations related to radioactive materials are also subjected to changes from time to time to maintain the state of the art of radiation protection. Scientists

planning experiments with radioactive materials are advised to check always on the regulations that apply to this particular establishment before commencing work with any radioactive materials.

**See also: Activation Analysis:** Neutron Activation. **Radiochemical Methods:** Natural and Artificial Radioactivity; Uranium; Radiotracers; Radio-Reagent Methods; Gamma-Ray Spectrometry.

## Further Reading

- L'Annunziata MF (1987) *Radionuclide Tracers: Their Detection and Measurement*. New York: Academic Press.
- L'Annunziata MF (ed.) (1998) *Handbook of Radioactivity Analysis*. New York: Academic Press (2nd edn., 2003).
- Brune D, Forkman B, and Persson B (1984) *Nuclear Analytical Chemistry*. Lund: Chartwell-Bratt Ltd.
- Choppin GR and Rydberg J (1980) *Nuclear Chemistry: Theory and Applications*. Oxford: Pergamon.
- Ehmann WD and Vance D (1991) *Radiochemistry and Nuclear Methods of Analysis*. New York: Wiley.
- Elving PJ (1986) *Nuclear Activation and Radioisotopic Methods of Analysis*. In: *Treatise on Analytical Chemistry*, vol. 14. New York: Wiley.
- Knoll GF (1989) *Radiation Detection and Measurement*. New York: Wiley.
- Parry JS (1991) *Activation Spectrometry in Chemical Analysis*. New York: Wiley.

## Natural and Artificial Radioactivity

**J Tölgyessy and M Harangozó**, Slovak Technical University, Bratislava, Slovak Republic

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Radiochemical methods are primarily concerned with the study of radioactivity in naturally occurring radioactive materials and in other materials in which radionuclides and their compounds are produced by irradiation. The foundation of such studies is the careful measurement of radioactivity in a variety of environmental samples, food samples, radiopharmaceuticals, etc. Such measurement can be divided into two major types:

1. Measurement of systems of naturally occurring radionuclides where objectives include the identification and determination of amounts of natural

radionuclides, and the use of this information to determine the presence and amounts of related elements in samples (e.g., uranium, thorium, and potassium).

2. Measurement of man-made (artificial) radionuclides, including identification and determination of amounts (e.g., fission products, fallout, radioactive spills, and leakage from waste disposal sites or other facilities).

In practice, radiochemical methods can be divided into two groups, one characterized by high sensitivity and the other offering high speed but less sensitivity.

## Sampling, Sample Preservation, and Sample Pretreatment

Air sampling is one of the most important and difficult steps in the surveillance of air pollution. Particulate air sampling techniques include filtration, electrostatic

In prompt gamma-ray activation analysis the measurement is done simultaneously with the bombardment, and excited intermediates of the nuclear reactions emitting the so-called prompt gamma-rays are determined. The characteristic wavelength (gamma-energy) of this emission can be used for qualitative analysis and their intensity for quantitative analysis of the target material.

Nuclear activation methods can provide high sensitivities for many but not all elements. These methods are capable of simultaneous multielement analysis and in some cases they may be essentially nondestructive analytical methods. One of the main disadvantages is the lack of information on chemical form (speciation) of the elements in the analyte. Neutron activation analysis is the most common form of activation analysis, but charged particle and photon activation analysis methods are also applied.

## Aspects of Safety Regulations

Work with radioactive materials is subjected to licensing procedures that may differ from country to country. Therefore, it is of utmost importance to become familiar with the regulations being issued by the relevant competent authority. Governmental regulations related to radioactive materials are also subjected to changes from time to time to maintain the state of the art of radiation protection. Scientists

planning experiments with radioactive materials are advised to check always on the regulations that apply to this particular establishment before commencing work with any radioactive materials.

**See also: Activation Analysis:** Neutron Activation. **Radiochemical Methods:** Natural and Artificial Radioactivity; Uranium; Radiotracers; Radio-Reagent Methods; Gamma-Ray Spectrometry.

## Further Reading

- L'Annunziata MF (1987) *Radionuclide Tracers: Their Detection and Measurement*. New York: Academic Press.
- L'Annunziata MF (ed.) (1998) *Handbook of Radioactivity Analysis*. New York: Academic Press (2nd edn., 2003).
- Brune D, Forkman B, and Persson B (1984) *Nuclear Analytical Chemistry*. Lund: Chartwell-Bratt Ltd.
- Choppin GR and Rydberg J (1980) *Nuclear Chemistry: Theory and Applications*. Oxford: Pergamon.
- Ehmann WD and Vance D (1991) *Radiochemistry and Nuclear Methods of Analysis*. New York: Wiley.
- Elving PJ (1986) *Nuclear Activation and Radioisotopic Methods of Analysis*. In: *Treatise on Analytical Chemistry*, vol. 14. New York: Wiley.
- Knoll GF (1989) *Radiation Detection and Measurement*. New York: Wiley.
- Parry JS (1991) *Activation Spectrometry in Chemical Analysis*. New York: Wiley.

## Natural and Artificial Radioactivity

**J Tölgyessy and M Harangozó**, Slovak Technical University, Bratislava, Slovak Republic

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Radiochemical methods are primarily concerned with the study of radioactivity in naturally occurring radioactive materials and in other materials in which radionuclides and their compounds are produced by irradiation. The foundation of such studies is the careful measurement of radioactivity in a variety of environmental samples, food samples, radiopharmaceuticals, etc. Such measurement can be divided into two major types:

1. Measurement of systems of naturally occurring radionuclides where objectives include the identification and determination of amounts of natural

radionuclides, and the use of this information to determine the presence and amounts of related elements in samples (e.g., uranium, thorium, and potassium).

2. Measurement of man-made (artificial) radionuclides, including identification and determination of amounts (e.g., fission products, fallout, radioactive spills, and leakage from waste disposal sites or other facilities).

In practice, radiochemical methods can be divided into two groups, one characterized by high sensitivity and the other offering high speed but less sensitivity.

## Sampling, Sample Preservation, and Sample Pretreatment

Air sampling is one of the most important and difficult steps in the surveillance of air pollution. Particulate air sampling techniques include filtration, electrostatic

precipitation, thermal precipitation, gravitational settling, centrifugal separation, and impingement. When the concentration of the radioactive component is high, the sampling procedure is simple. An enclosed metal or glass vessel equipped with inlet and outlet valves is filled simply by passing air through the vessel and subsequently closing the valves. When the radioactive component is present at a lower concentration, it is necessary to concentrate this radioactive component during the sampling procedure by using either sorption or compression. The absorption of carbon dioxide ( $^{14}\text{CO}_2$ ) in alkali and the adsorption of radon on charcoal are examples of such procedures.

In the case of air sampling by absorption in a collecting liquid solution, it is often advantageous to extract the absorbed analytes or concentrate them on ion-exchange materials, followed by elution. It is possible to dissolve the aerosols from the collecting filters and concentrate the analytes by ion exchange or extraction. Sometimes it is necessary to combine several filters to obtain sufficient analyte for measurement.

Wet- or dry-ashing technique may be needed to prepare particulates for dissolution. Aerosols and filter media are mineralized for these techniques.

Methods for sampling can be divided into two groups. In the first, the  $\alpha$ -,  $\beta$ -, or  $\gamma$ -activity (or  $\gamma$ -spectrum) is to be measured. In the second, a radiochemical analysis for certain radionuclides is to be performed.

For direct measurement, the sample is prepared by drawing a certain amount of air through a filter. To measure  $\alpha$ - and  $\beta$ -activities, the aerosols must be retained on the filter surface. Membrane filters, plastic filters, and glass fiber filters are used. For  $\gamma$ -activity measurement, the filters need not retain the sample at the surface.

When a radiochemical analysis has to be performed, the retention of aerosols on the filter surface is not necessary, and those filters that yield only small amounts of ash are preferred. This requirement is best satisfied with cellulose filters. In addition to filtration, radioactive aerosols can also be concentrated by electrostatic or thermal precipitation.

In water sampling, close attention must be given to the type of sample, the sampling equipment, sample container, holding times, and proper preconcentration techniques. In the case of radioactive material, it is very important that they are present generally in extremely low concentrations in water, so that sorption and volatilization can occur. It has been recommended that preservatives be added at the time of sampling unless suspended and dissolved fractions are to be separated. Concentrated hydrochloric or nitric acid may be used to adjust the pH (usually to  $<2$ ). Acidified samples should be held for at least

16 h before analysis. All preservatives and reagents should be tested for radioactivity, as well as all sampling equipment and containers. Because of the extremely low concentrations of many water constituents ( $\mu\text{g l}^{-1}$ ), it is frequently necessary to concentrate samples before analysis.

Sampling of aquatic and atmospheric systems is relatively straightforward because of the basic simplicity of the sample matrix. The situation is somewhat different in the lithosphere, where there are far greater variations. Rock and minerals of greatly differing accessibility, vegetation, animals, and various fluids (including water and sludges) may be present in different sampling situations.

In the preparation of samples from biological materials, the time between sampling and analysis should be kept as short as possible. For storage times of a few (2 or 3) days, refrigeration at  $4^\circ\text{C}$  may be adequate. In favorable cases, holding times can be prolonged by the addition of chemical preservatives (e.g., formalin (formaldehyde,  $\text{HCHO}$ ) and alcohol). Dehydration (by drying, combustion, and lyophilization) is also used.

Plant materials should be carefully washed free from sands, soils, and other physical artifacts. After draining away the wash water, leaves, stems, seeds, and fruits must be dried. Undried components are homogenized with a blender with glass walls and high-quality stainless-steel blades. Dry biological samples (seeds and grains) are generally ground in mills and sieved through fine-mesh screens to obtain uniformly sieved material.

Human and animal tissues are generally cut into thin sections or ground in mills. In some cases, samples are mineralized by dry- or wet-ashing technique.

## Analysis of Natural Radioactive Materials

Naturally occurring radionuclides consist of two basic groups: primordial and cosmogenic.

Primordial radionuclides are associated with the formation of the earth. A list of these is given in Table 1. The radioactive isotopes of the elements of higher atomic numbers are members of one of four possible decay series:

Thorium series,  $4n$ :  $^{232}\text{Th} \rightarrow ^{208}\text{Pb}$   
 Neptunium series,  $4n + 1$ :  $^{241}\text{Cm} \rightarrow ^{209}\text{Bi}$   
 Uranium series,  $4n + 2$ :  $^{238}\text{U} \rightarrow ^{206}\text{Pb}$   
 Actinium series,  $4n + 3$ :  $^{235}\text{U} \rightarrow ^{207}\text{Pb}$

The neptunium series, because of the relatively short life of its parent, is no longer present in nature. The

**Table 1** Primordial radionuclides

Nuclide	Half-life (years)	% Isotopic abundance	Emissions, decay <sup>a</sup>	Energy of primary emission (MeV)
<sup>40</sup> K	$1.3 \times 10^9$	0.0118	$\beta, \gamma, \varepsilon$	1.32
<sup>50</sup> V	$\sim 6 \times 10^{14}$	0.24	$\beta, \gamma, \varepsilon$	—
<sup>87</sup> Rb	$4.7 \times 10^{10}$	27.85	$\beta$	0.27
<sup>115</sup> In	$6 \times 10^{14}$	95.72	$\beta$	0.6
<sup>138</sup> La	$1.1 \times 10^{11}$	0.089	$\beta, \gamma, \varepsilon$	0.205
<sup>142</sup> Ce	$5 \times 10^{15}$	11.07	$\alpha$	1.5
<sup>144</sup> Nd	$\sim 5 \times 10^{15}$	23.85	$\alpha$	1.8
<sup>147</sup> Sm	$1.06 \times 10^{11}$	14.97	$\alpha$	2.24
<sup>148</sup> Sm	$1.2 \times 10^{13}$	11.24	$\alpha$	2.14
<sup>149</sup> Sm	$\sim 4 \times 10^{14}$	13.83	$\alpha$	1.84
<sup>152</sup> Gd	$1.1 \times 10^{14}$	0.200	$\alpha$	2.15
<sup>174</sup> Hf	$4.3 \times 10^{15}$	0.18	$\alpha$	2.5
<sup>176</sup> Lu	$3.6 \times 10^{10}$	2.59	$\beta, \gamma$	0.42
<sup>187</sup> Re	$7 \times 10^{10}$	62.93	$\beta$	0.008
<sup>190</sup> Pt	$7 \times 10^{11}$	0.0127	$\alpha$	3.11
<sup>192</sup> Pt	$\sim 10^{15}$	0.78	$\alpha$	2.6
<sup>204</sup> Pb	$1.4 \times 10^{17}$	1.48	$\alpha$	2.6
<sup>238</sup> U	(series) <sup>b</sup>			
<sup>235</sup> U	(series) <sup>b</sup>			
<sup>232</sup> Th	(series) <sup>b</sup>			

<sup>a</sup>  $\varepsilon$  = electron capture.<sup>b</sup> See **Figures 1–3**.

decay sequences of the uranium, actinium, and thorium are given in **Figures 1–3**.

Cosmogenic radionuclides are produced continuously as a result of nuclear reaction in the environment, such as interaction with cosmic radiation, spallation, and natural fission. A partial list is presented in **Table 2**.

Determination is usually relative; i.e., samples are compared with standards. For comparative measurements, it may be adequate to make relative measurements without standards, with an instrument calibrated for the purpose.

Direct measurements may be complicated by the phenomenon of self-absorption, especially in the case of low-energy  $\beta$ - or  $\alpha$ -radiation. Two approaches can be used to minimize self-absorption errors:

1. Measurement of saturation layers; that is, layers that are thicker than the penetration range of the measured radiation (e.g.,  $>0.1$  mm for most  $\alpha$ -particles).
2. Measurement of infinitesimally thin layers; layers that are thin enough for self-absorption to be negligible (layers 'density'  $<0.1$  mg cm<sup>-2</sup>).

### Determination of Natural Radioactive Elements

Naturally occurring radioactive elements can be determined by measuring the activity of the appropriate radioisotopes. These measurements are particularly suitable for radioactive elements with short half-lives.

The use of radioisotopes with long half-lives is less suitable, particularly as regards accuracy and sensitivity; even in this case, however, radiometric methods are usually simpler and more rapid than conventional procedures. The basic assumption is that a radioelement has at least one isotope that can be measured; hence the isotopic composition must be known. If an element is known to have a constant isotopic composition, measurements and calculations are relatively straightforward (e.g., potassium, rubidium, samarium, lutetium, rhenium, and francium).

When an element has more than one radioisotope, determinations and data analysis are generally more complex because the isotopes may differ in half-life, especially when a series is involved, e.g., radium, thorium, polonium, radon, actinium, protactinium, and uranium. One possibility is to make measurements after the decay of the short-lived radionuclides, but this may require long waiting times. In favorable cases, it is more convenient to measure the activity of decay products (e.g., radon, thoron (<sup>220</sup>Rn), actinon (<sup>219</sup>Rn)), or correct the measurements of the short-lived radioisotopes after determination of the isotopic composition.

Emanometric methods are radioanalytical methods that use measurement of radioactive isotopes of inert gases for the determination of appropriate elements. A good example is the use of the radon isotopes <sup>226</sup>Rn, <sup>222</sup>Rn, and <sup>219</sup>Rn to determine radon, thorium, radium, and actinium. Indirect determinations



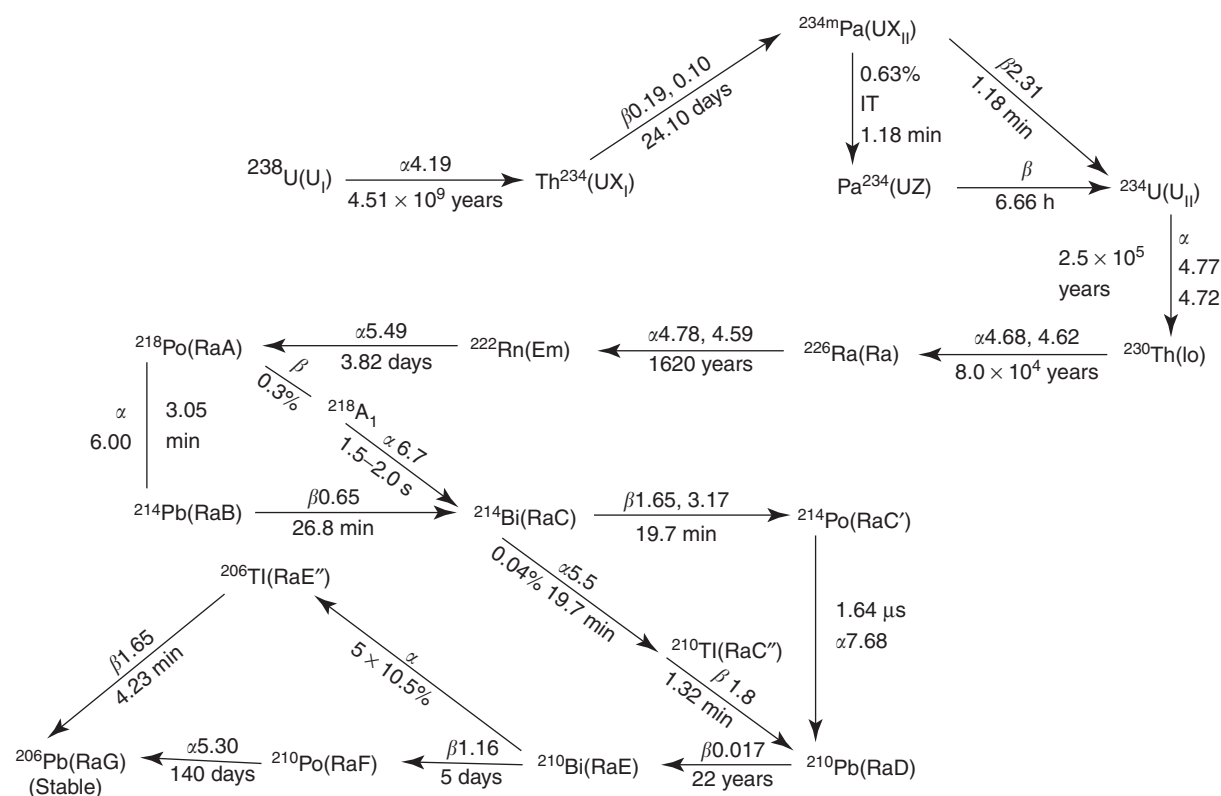


Figure 1 Uranium series decay scheme (energies in MeV).

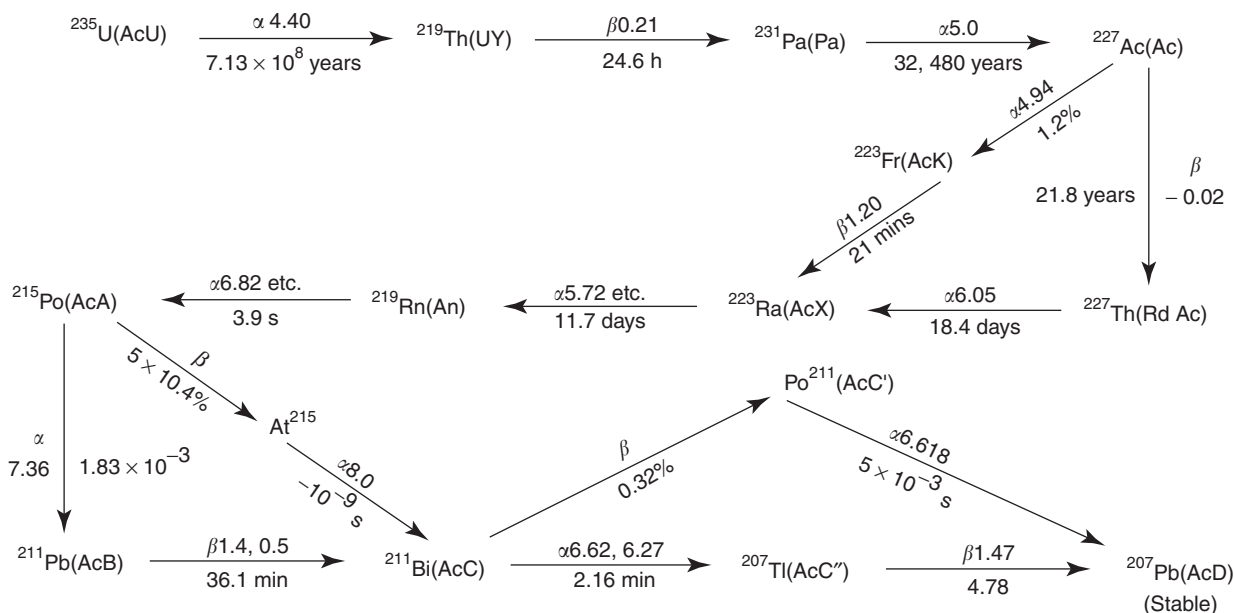
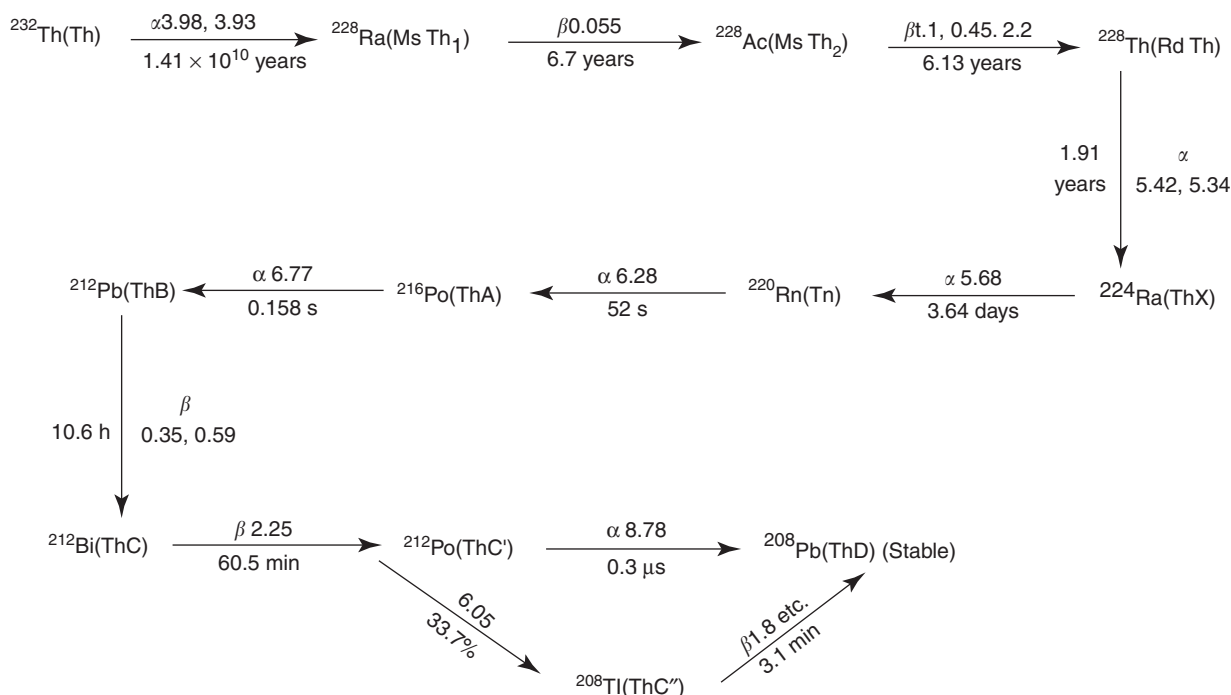


Figure 2 Actinium series decay scheme (energies in MeV).

(e.g., use of radon to determine radium) require both components to be in equilibrium, or present in a known ratio.

In general, these methods require the removal of the inert gas from the sample; the gas is then

transferred to a suitable detector. Solid samples must be converted into liquid form such as a solution or a melt (by fusion with sodium carbonate or borax). Once the sample is in liquid form, the gas may be swept out by bubbling air through the solution into

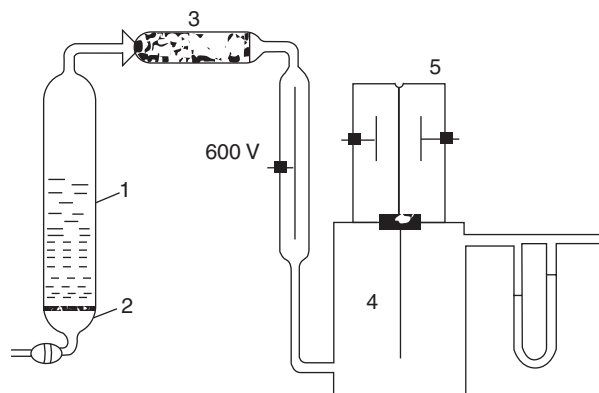
**Figure 3** Thorium series decay scheme.**Table 2** Major cosmic-ray activated radionuclides

Nuclide	Half-life	Emission, decay	Energy of primary emission (MeV)
$^3\text{H}$	12.26 years	$\beta$	0.0181
$^7\text{Be}$	53 days	$\varepsilon, \gamma$	—
$^{10}\text{Be}$	$2.7 \times 10^6$ years	$\beta$	0.56
$^{14}\text{C}$	5730 years	$\beta$	0.156
$^{22}\text{Na}$	2.58 years	$\beta^+, \varepsilon, \gamma$	0.54
$^{32}\text{Si}$	$\sim 650$ years	$\beta$	0.1 (to $^{32}\text{P}$ )
$^{32}\text{P}$	14.3 days	$\beta$	1.71
$^{35}\text{P}$	25 days	$\beta$	0.25
$^{35}\text{S}$	86.7 days	$\beta$	0.168
$^{36}\text{Cl}$	$3 \times 10^5$ years	$\beta, \varepsilon$	0.71
$^{39}\text{Cl}$	55 min	$\beta, \gamma$	1.91, 2.18, 3.43

an evacuated ionization chamber, by passing a continuous stream of air through the sample and detector, by spraying the liquid using a small air-screw, or by shaking the liquid sample with air. An example of emanation equipment for the determination of thoron is shown in **Figure 4**. Alternatively, an  $\alpha$ -scintillation cell coupled to a photomultiplier can be used.

### Air

The natural radioactivity in air is due mainly to radioactive emanations (radon, thoron, actinon) and to those of their deposits that produce radioactive

**Figure 4** Emanation equipment for the determination of thoron: 1, purge system; 2, fritted glass disk; 3, filter with glass wool; 4, ionization chamber; and 5, electrometer.

aerosols. For the analysis of air samples, a number of approaches are possible:

1. Direct emanometric determination with large-volume (20 l) differential ionization chambers that permit measurement at activity levels of  $\sim 10^{-7} \text{ Bq cm}^{-3}$ .
2. Preconcentration of radioactive emanation by condensation, adsorption on activated charcoal, cooling with liquid air, or by sorption in cooled organic solvents (e.g., toluene or carbon disulfide). After collection, the radioactive gases can be

purged from the collectors and transferred to an appropriate ionization chamber.

3. Capture (e.g., filtration) of aerosols and measurement of their activity with scintillation counting or by autoradiography.

### Water

The concentration of natural radioactive material in water varies considerably, depending on the nature, location, and history of the water. The principal natural radionuclides are  $^{238}\text{U}$ ,  $^{232}\text{Th}$ , and  $^{235}\text{U}$  and their decay products, such as  $^{226}\text{Rn}$  and  $^{222}\text{Rn}$  from  $^{238}\text{U}$ ,  $^{228}\text{Rn}$  and  $^{224}\text{Rn}$  from  $^{232}\text{Th}$ . If we discount the presence of uranium and thorium, the contribution of which to the total activity is usually relatively small, we can differentiate between temporary radioactivity caused by radon, and permanent activity caused by radium. For the determination of radon, the emanation can be flushed to an ionization chamber and measured. Radium can be determined by  $\alpha$ -counting of a barium-radium sulfate precipitate or by measurement of  $^{222}\text{Rn}$  produced from  $^{226}\text{Ra}$ .

### Analysis of Artificial Radioactive Materials

The analysis of natural radioactive materials is relatively straightforward, since it usually involves the determination of a small number of well-known radioisotopes of a few elements. The determination of artificial radionuclides in environmental samples, food samples, radiopharmaceuticals, etc., is more complicated, because several radionuclides of a number of elements may be involved.

The first aim of the qualitative analysis is identification of all radionuclides present in a sample (identification of all activities and the nuclides responsible). The procedure has two steps:

1. The characterization of all activities present, e.g., nature and energy of radiation, relative intensities, half-life values.
2. Identification of the radionuclides responsible for the various activities.

The nature and energy of radiation can be determined by

- $\gamma$ -spectrometry with NaI(Tl) scintillation crystals or solid-state detectors such as hyper-pure germanium, germanium/lithium, or silicon/lithium systems;
- $\alpha$ -spectrometry with surface barrier detectors;

- $\beta$ -spectrometry with silicon semiconductor detectors (surface barrier or silicon/lithium type); and
- absorption measurements.

Half-life values can be determined from decay curves showing the variation of activity with time; these values must be correlated with the types of radiation found. Once this correlation has been made, appropriate tables can be used for nuclide identification.

The aim of the quantitative analysis is to measure an amount of radioactivity in either a relative or absolute sense. Common units are Becquerels (Bq), Bq per unit mass or Bq per unit volume. In special cases, the activity is expressed as a mass of the radioactive nuclide. The determination of absolute activity presents special problems in addition to those encountered in relative activity measurements. Once a source of a radionuclide of known absolute activity is available, any detector may be calibrated in terms of this standard. This calibration will be valid for other samples of the same nuclide, provided they are measured under precisely the same conditions. The calibration may also be adequate for other radionuclides emitting radiation similar to those of the standard.

The main task in the analysis of a mixture of fission products is the identification of nuclides and the determination of yields. A combination of instrumental and chemical techniques is required and the work is laborious and difficult.

One of the most hazardous fission products is  $^{90}\text{Sr}$ , because of its long half-life (28 years) and because strontium is concentrated in bone. For the determination of total radioactive strontium and  $^{90}\text{Sr}$  in water, the following procedure is used.

A known amount of strontium (as the nitrate,  $\text{Sr}(\text{NO}_3)_2$ ) is added as carrier to a water sample. The carrier, alkaline-earth and rare-earth metals are precipitated as their carbonates to concentrate the radiostrontium. The strontium is then separated from other radioactive elements and inactive sample solids by precipitation as  $\text{Sr}(\text{NO}_3)_2$  from fuming nitric acid solution, and finally precipitated as strontium carbonate,  $\text{SrCO}_3$ , dried, weighed to determine the carrier recovery, and measured for radioactivity. A correction is applied to compensate for loss of carrier and activity during the purification procedure. A delay in counting will give an increased counting rate because of the production of  $^{90}\text{Y}$ . The total activity will be due to  $^{89}\text{Sr}$  and  $^{90}\text{Sr}$ . The amount of  $^{90}\text{Sr}$  can be determined by separating and measuring its daughter,  $^{90}\text{Y}$ , after equilibrium has been reached. Counting instruments used include internal gas-flow proportional counters and thin-window proportional or GM counters. The counting systems can be calibrated with  $^{137}\text{Cs}$  standards.

**Table 3** Radiochemical procedures for environmental and food samples for various radionuclides

Analyte	Sample	Preparation, Dissolution	Chemical separation	Method of measurement
$^{24}\text{Na}$	Precipitation	Add Na carrier	Na separated by ion exchange, hydroxides scavenged, Na measured in filtrate	$\beta$ , $\gamma$ Coincidence counter
$^{14}\text{C}$	Food, air	Combust to $\text{CO}_2$	$\text{CO}_2$ purified by expansion, condensation, drying agents	Internal gas counter, anticoincidence shielded
$^{14}\text{C}$	Water	Add oxalate Carrier	C oxidized to $\text{CO}_2$ in acidic solution, $\text{CaCO}_3$ collected	$\beta$ -Counter (gas flow or liquid scintillation)
$^{32}\text{P}$	Seawater	Add $\text{PO}_4^{3-}$ Carrier	P oxidized with $\text{HClO}_4$ , $\text{NH}_4$ molybdophosphate extracted into isoamyl alcohol; precipitated as $\text{NH}_4\text{MgPO}_4$	$\beta$ -Counter
$^{131}\text{I}$	Milk	Add HCHO and I carrier	$\text{I}^-$ separated on anion exchange resin, purified by extraction; $\text{Pdl}_2$ precipitated	$\beta$ -Counter
$^{137}\text{Cs}$	Milk, bone, vegetation, ash, soil, water	Fuse with $\text{Na}_2\text{CO}_3$ or leach with $\text{HNO}_3$	Ca precipitated; Cs separated on $\text{NH}_4$ molybdophosphate, purified by ion exchange, precipitated as $\text{Cs}_2\text{PtCl}_6$	$\beta$ -Counter
$^{210}\text{Pb}$	Bone	Ignite; dissolve in HBr	Pb extracted with quaternary amine; $\text{PbSO}_4$ precipitated and counted	$\beta$ -Counter
$^{232}\text{Th}$	Bone	Ignite at $600^\circ\text{C}$	Irradiate with neutrons, dissolve in HCl, separate $^{223}\text{Pa}$ by anion exchange	$\beta$ -Counter
$^{226}\text{Ra}$	Soil, vegetation	Plant – dissolve in acids; soil – exchange with ammonium acetate	$^{225}\text{Ra}$ , $^{133}\text{Ba}$ tracers and Ba carrier added; Ra precipitated with $\text{BaSO}_4$ , separated from Ba by ion exchange; electrodeposited for counting	$\alpha$ -Spectrometry
$^{90}\text{Sr}$	Bone	Ignite, dissolve with HCl	$^{90}\text{Y}$ extracted into ethylhexyl phosphoric acid; impurities removed by amine extraction	$\beta$ -Counter
$^{90}\text{Sr}$	Food, vegetation, milk	Evaporate, ignite, fuse with $\text{Na}_2\text{CO}_3$	Sr separated, purified by nitrate precipitations, $^{90}\text{Sr}$ precipitated and counted	$\beta$ -Counter
$^{110\text{m}}\text{Ag}$	Biological samples	Grind in blender with water	Slurry treated with NaOH-AgCN solution; Ag electroplated on Pt, dissolved, AgCl precipitated	$\gamma$ -Ray spectrometry ( $4\pi$ coincidence counter)
Th-Cf	Soil	Treat with acids, fuse – KF, $\text{Na}_2\text{S}_2\text{O}_7$ , dissolve HCl	Actinides co-precipitated with $\text{BaSO}_4$ , separated by solvent extraction after adjusting oxidation states, electrodeposited for counting	$\alpha$ -Spectrometry
U	Water		Extraction into methyl isobutyl ketone	$\alpha$ -Spectrometry

Continued

Table 3 Continued

Analyte	Sample	Preparation, Dissolution	Chemical separation	Method of measurement
Pu	Soil, air particulates	Soil – ash, extract with hot HCl; particulates – ash dissolve in mineral acids or by fusion	Pu separated and purified by cation and anion exchange; electrodeposited for counting	$\alpha$ -Spectrometry
Pu	Water, air, soil, food, vegetation	Treat with mineral acids	Pu purified by anion exchange; electrodeposited for counting	$\alpha$ -Spectrometry

Fission products may contain the iodine isotopes from  $^{129}\text{I}$  to  $^{135}\text{I}$ . Of these,  $^{131}\text{I}$  is considered the most significant hazard in drinking water. Three methods are available for the determination of radioactive iodine in water samples: precipitation, sorption on an anion-exchange resin, and distillation. The precipitation method is preferred because it is simple and requires the least time. In the precipitation method, iodate carrier is added to the sample and reduced to iodide with sodium sulfite. The iodide is precipitated as silver iodide. The precipitate is dissolved, and purified with zinc powder and sulfuric acid. The iodide is finally precipitated as palladium iodide,  $\text{PdI}_2$ , for counting in a low-background  $\beta$ -counter, or  $\beta/\gamma$  coincidence system.

In the determination of transuranium elements (or nuclides), the most important step is separation of the elements from the sample matrix. Differences in redox properties are used for the separation of the first four elements in the series (neptunium, plutonium, americium, curium). Since the higher members exist primarily in the same oxidation state (III), separation by ion-exchange chromatography is commonly used. The lighter transuranic elements can be determined by common chemical methods, and trace amounts are usually determined by radiometric methods such as  $\alpha$ -spectrometry.

Ion-exchange chromatography is generally used for the separation of the transplutonium elements (americium, curium, berkelium, californium, einsteinium, fermium, mendelevium, nobelium, lawrencium). Determinations are usually made directly by  $\alpha$ -spectrometry with solid-state detectors. Some elements (americium, curium, berkelium, californium) also have long-lived isotopes and can be determined by chemical methods such as ultraviolet–visible spectrophotometry.

Illustrative radiochemical procedures for environmental and food samples for various radionuclides are summarized in Table 3.

**See also:** **Air Analysis:** Sampling. **Ion Exchange:** Ion Chromatography Instrumentation. **Sample Dissolution for Elemental Analysis:** Dry Ashing; Oxygen Flask Combustion; Wet Digestion; Microwave Digestion. **Water Analysis:** Overview; Freshwater.

## Further Reading

- APHA (1986) *Standard Methods for the Examination of Water and Wastewater*, 16th edn. Washington: American Public Health Association.
- Environmental Instrumentation Group (1972) *Instrumentation for Environmental Monitoring*, LBL-1, University of California, Lawrence Berkley Laboratory, 1972 and updates.
- Johns FB (ed.) (1975) *Handbook of Radiochemical Methods*, EPA-68014-75-001, and updates. Las Vegas, NV: National Environmental Research Center, Office of Research and Development (USEPA).
- NCRP (1977) *Environmental Radiation Measurement*, NCRP Report No. 50. Washington, DC: National Council of Radiation Protection and Measurements.
- NCRP (1979) *National Council on Radiation Protection and Measurement*, Report No. 62. Washington, DC, 1979 and updates.
- Tölgyessy J and Bujdosó E (1991) *Handbook of Radio-analytical Chemistry*, vols. I and II. Boca Raton: CRC Press.
- Tölgyessy J and Klehr EH (1987) *Nuclear Environmental Chemical Analysis*. Chichester: Ellis Horwood.
- Tölgyessy J and Klehr EH (1991) *Jadernye metody chimicheskovo analiza okrushajushej sredy*. Moscow: Chimija.
- Tölgyessy J, Jakovlev JuV, and Bilimovitch GN (1985) *Diagnostika okrushajushej sredy radioanaliticheskimi metodami*. Moscow: Energoatomizdat.

## Radionuclide Monitoring

**J Tölgyessy and M Harangozó**, Slovak Technical University, Bratislava, Slovak Republic

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The development of applications of radioactive materials in research, industry, medicine, and agriculture, as well as the growth of nuclear power engineering programs, led to increasing amounts and varieties of radionuclides in the environment. This, together with the development of the nuclear sciences, has created the demand for proper analytical methods of monitoring radionuclides in the environment.

With the development of nuclear power engineering and nuclear technology, radionuclides, which formerly were not encountered, have entered the environment. In fact, artificial radionuclides, as well as natural radionuclides, are abundant and are introduced into the human organism in the form of food of plant and animal origin.

As a result, a new branch of science was formed to deal with monitoring of radionuclides in the environment. In a broader sense, this field involves the elaboration of relevant analytical procedures and the study of the ways radionuclides come into the environment and the health hazards they present.

From the practical standpoint, radionuclide monitoring techniques can be divided into two groups, one that is characterized by a high sensitivity and a second that offers high speed but relatively lower sensitivity. When selecting a proper analytical method, sometimes the faster method will be chosen and, other times, the more sensitive one.

### Radionuclides in the Environment

Radioactive contamination can result from either expected or accidental release of radionuclides during the treatment of uranium ores, the operation of nuclear reactors, the processing of burnt fuel elements from nuclear reactors, or the application of radionuclides in medicine, research, industry, and agriculture, as well as from radioactive fallout from the atmosphere.

Natural as well as artificial radionuclides can become sources of radioactive contamination in the environment. Primary natural radionuclides that should be considered are  $^{40}\text{K}$ ,  $^{87}\text{Rb}$ ,  $^{235}\text{U}$ , and  $^{238}\text{U}$ . Other natural radionuclides are members of

radioactive series and represent a majority of the total number of existing natural radionuclides. Radionuclides induced in nature can also be considered natural; some of these are  $\text{T}$ ,  $^7\text{Be}$ ,  $^{14}\text{C}$ ,  $^{22}\text{Na}$ ,  $^{32}\text{P}$ , and  $^{35}\text{S}$ . Artificial radionuclides that are significant from the standpoint of contamination are either direct or indirect products of nuclear fission reactions or eventually of thermonuclear processes. For example, on the average, during nuclear fission as many as 200 radioisotopes are formed of 35 different elements with medium atomic numbers. However, the abundance of many of them in the fission mixture decreases rapidly because of their short half-lives; this also leads to a rapid decrease in the total activity of the fission mixture. A regular operation of the nuclear reactor does not represent as serious a source of radioactive contamination of the environment as does the processing of burnt fuel elements. After removing fuel elements from the reactor core they are allowed to 'cool' for a certain time – 'the cooling time.' Then they are dissolved in acids. Radioisotopes of aerogens, largely of radioiodine and, in the case of an oxidation process, ruthenium (as  $\text{RuO}_4$ ), are simultaneously released. The most dangerous is  $^{131}\text{I}$ , which, after entering the atmosphere, is rapidly deposited on plants, thus contaminating the food and forage base. However, during the decomposition of burnt fuel elements, certain radioactive particles are also released, which, in spite of careful filtration, penetrate the surrounding atmosphere.

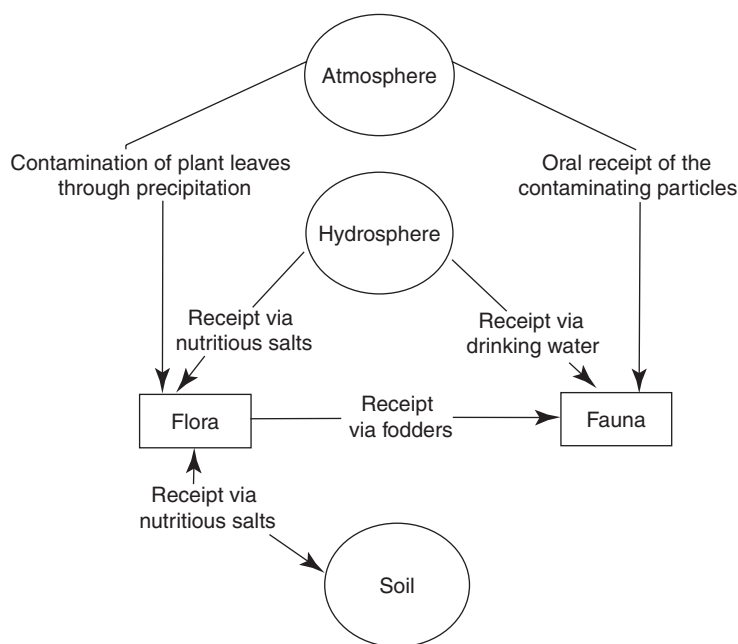
A significant amount of fission products passes into waste water. Storage of great quantities of high activity waste water is a relatively complicated problem. At the present time this is mostly solved by storage in underground tanks, concentration and conversion to solid forms, and purification by precipitation and ion-exchanging procedures.

A further source of radioactive contamination is the radionuclides induced with neutron radiation in nuclear reactors. Particularly large amounts of the following radionuclides are created:  $^{60}\text{Co}$ ,  $^{192}\text{Ir}$ ,  $^{198}\text{Au}$ ,  $^{131}\text{I}$ , and  $^{32}\text{P}$ .

From sea vessels powered by nuclear reactors,  $^{60}\text{Co}$ ,  $^{55}\text{Fe}$ ,  $^{59}\text{Fe}$ , and  $^{51}\text{Cr}$  pass into the sea with cooling water.

For the protection of man's health, studies of the contaminated biosphere, which represents a source of food, is of vital importance. Through a process that consists of several steps, mineral substances are converted to different forms of life by plants, the plants are eaten by animals, and the contaminated substances are incorporated into food. In this





**Figure 1** Radioactive contamination of plants and animals.

connection, all paths by which radionuclides pass from the atmosphere, hydrosphere, or soil to the biosphere are of interest. Some of the ways plants and animals are contaminated are sketched in **Figure 1**.

Plants can be contaminated by receiving radionuclides with nutrient solutions from water and soil. Also, soils can be contaminated by radioactive atmospheric precipitation and dust. In this way contamination is transferred to animals, since the contaminated plants are their forage. Animals can also be contaminated from the hydrosphere by drinking water.

The following essential factors should be considered when estimating radioactive contamination:

- the amount of the radionuclide;
- the character of the contamination (expected or accidental);
- the kind of radionuclide;
- the physical and chemical forms of the contaminating compound;
- the character of the contaminated area.

The nature of the contamination of the environment, as well as the degree of hazard, is given by properties of the radiation source and by the character of the contamination (whether it occurs during the usual operation of nuclear equipment or treatment of radioactive materials, or whether it results from an accident). The human organism can be exposed either directly (by inhalation, ingestion of contaminated water and/or food, or contamination of the skin) or indirectly (when contaminated materials enter into

the human organism with contaminated components of the biosphere).

The radioactivity of the environment can be measured regularly or from time to time to estimate the extent of the radiation hazard to man. Results of measuring radioactivity in samples taken from humans and from the biosphere are considered from the standpoint of maximum permissible levels based on the recommendations of national and/or international institutions. The maximum permissible levels for many radionuclides either in the whole body or in individual organs represent criteria for sensitivities for relevant radiochemical methods.

### Radioactive Contamination of the Environment

Knowledge of the composition of radioactive waste materials from nuclear equipment (regardless of their phases) is important in choosing a sampling protocol prior to any analytical procedures.

The recommendations of the International Commission on Radiological Protection (ICRP) for maximum permissible concentrations of radionuclides in air and water serve as basic criteria for determining radionuclide concentrations in air and water. For the first estimation, measurements of gross  $\alpha$ ,  $\beta$ , and  $\gamma$  activities are sufficient; however, in order to draw a responsible conclusion, individual radionuclides in the sample have to be determined.

Analysis of air holds a unique position, since air samples yield significant data about radionuclides that appear as either gases or finely dispersed particles transported by air.

Water samples come either from surface or underground water sources. Surface water can be contaminated with radionuclides either during atmospheric precipitations or in the course of the surface flow. Underground water can be contaminated with radionuclides from waste materials stored under the earth's surface; however, they may also contain natural radioactive compounds from their deposits. Some radioactive compounds are removed when surface water is conditioned for supply water; however, the most dangerous long-lived radionuclides, such as  $^{90}\text{Sr}$  and  $^{137}\text{Cs}$ , remain in the conditioned water.

Radionuclides in soil are a source of the contamination of forage and food with radionuclides, of which strontium, cesium, and radium isotopes are the most significant. Radionuclides penetrate into plants either from the atmosphere as deposits on soil surface or through roots from the soil. Plants radioactively contaminated are a significant hazard to man, either directly (food of plant origin) or indirectly (the milk of animals which receive contaminated forage).

$^{89}\text{Sr}$ ,  $^{90}\text{Sr}$ ,  $^{131}\text{I}$ ,  $^{137}\text{Cs}$ ,  $^{140}\text{Ba}$ , and  $^{226}\text{Ra}$  are found particularly often in plant materials. Of them,  $^{131}\text{I}$  and  $^{140}\text{Ba}$  represent a relatively short-term hazard, thanks to their half-lives; however, a long-term contamination can be expected of  $^{90}\text{Sr}$ ,  $^{137}\text{Cs}$ , and  $^{226}\text{Ra}$ .

A study of milk contamination is of particular interest, since strontium, cesium, and iodine enter into the human organism only with milk.

Contamination of cheese usually has its origin in the presence of  $^{89}\text{Sr}$  and/or  $^{90}\text{Sr}$  in milk. Contamination of meat can mostly be attributed to  $^{137}\text{Cs}$ . Contents of either  $^{131}\text{I}$  in the thyroid gland of animals or  $^{89}\text{Sr}$  and  $^{90}\text{Sr}$  in their bones may serve as indicators of the degree of environment contamination.

Water organisms (from plankton to fish) can concentrate certain radionuclides from their environment. Concentration effects were observed for the following elements: cobalt in mollusks; ruthenium in water plants; zinc, cadmium, and iron in fish bowels; calcium and strontium in bones and shells, etc.

Table 1 indicates the important radionuclides in evaluating contamination of the environment in the case of release from a nuclear facility and Table 2 lists those radionuclides in samples of human origin.

A somewhat different presentation of important nuclides is given in Table 3.

## Sampling, Sample Preservation, and Sample Pretreatment

In many cases, direct measurement cannot be made on environmental samples. Very often they must be

**Table 1** Samples for evaluating environmental contamination

Samples	Radionuclides
Air	All radionuclides dispersed with air
Water	All radionuclides dispersed with water
Soil	$^{90}\text{Sr}$ , $^{226}\text{Ra}$ , $^{137}\text{Cs}$
Plants	$^{131}\text{I}$ , $^{89}\text{Sr}$ , $^{90}\text{Sr}$ , $^{137}\text{Cs}$ , $^{226}\text{Ra}$ , $^{140}\text{Ba}$
Milk	$^{89}\text{Sr}$ , $^{90}\text{Sr}$ , $^{137}\text{Cs}$ , $^{131}\text{I}$ , $^{140}\text{Ba}$ , $^{226}\text{Ra}$
Milk products	$^{89}\text{Sr}$ , $^{90}\text{Sr}$ , $^{137}\text{Cs}$ , $^{226}\text{Ra}$
Meat	$^{137}\text{Cs}$
Thyroid gland	$^{131}\text{I}$
Bones	$^{89}\text{Sr}$ , $^{90}\text{Sr}$ , $^{226}\text{Ra}$
Water organisms	$^{89}\text{Sr}$ , $^{90}\text{Sr}$ , $^{137}\text{Cs}$ , $^{226}\text{Ra}$ , $^{141}\text{Ce}$ , $^{144}\text{Ce}$

**Table 2** Samples of human origin generally used in measuring radioactive contamination

Radionuclide	Basic sample	Other samples	Remarks
$^3\text{H}$	Urine	—	
$^{89}\text{Sr}$	Urine	Feces	
$^{90}\text{Sr}$	Urine	Feces, bones	
$^{131}\text{I}$	Thyroid gland	Urine	
$^{137}\text{Cs}$ – $^{137}\text{Ba}$	Urine	Blood	
$^{140}\text{Ba}$ – $^{140}\text{La}$	Urine	—	
$^{226}\text{Ra}$	Expired air, urine	Bones, feces	Rn is determined in expired air
$^{239}\text{Pu}$	Urine	Feces	
$^{210}\text{Po}$	Urine, blood	Feces	
Th (natural)	Urine	—	
U	Urine	—	

preserved and analytes must be separated, concentrated, or sometimes transferred into another phase. Generally, therefore, a number of steps may be required prior to final measurements.

Standard and reference materials are exceedingly important for the verification of each step of the total procedure and calibration of the final measurement technique. Careful attention must be given to their preparation and use.

Special consideration must be given to sampling for the monitoring of radionuclides in the vicinity of nuclear facilities (all phases of the nuclear fuel cycle, radionuclide production, accelerator, medical and research facilities), or in the event of accidents involving the release of radioactive materials during shipment or storage (e.g., spills during shipment). In such cases, a surveillance program is needed to provide information used for dose calculations, comparison with regulatory standards, assessment of environmental contamination, countermeasures, and public information.

The objectives of such a program are as follows:

- assessment of actual or potential exposure to humans;

**Table 3** Relative importance of environmental media as a function of effluent pathway and radionuclide(s)  
Effluent pathway – atmosphere

<i>Tritium</i>	<i>Noble gases</i>	<i>Iodine-131</i>	<i>MF&amp;AP<sup>a</sup></i>	<i>TRU<sup>b</sup></i>
Air	Direct radiation	Milk	Air	Air
Vegetables	Air	Vegetables	Vegetables	Deposition
		Air	Milk	Terrestrial biota
		Animal thyroids	Direct radiation	Soil
		Forage	Deposition	
			Forage	

*Effluent pathway – hydrosphere*

<i>Tritium</i>	<i>Iodine-131</i>	<i>MF&amp;AP<sup>a</sup></i>	<i>TRU<sup>b</sup></i>
Drinking water	Drinking water	Fish, shellfish	Sediment
Surface water	Milk	Waterfowl	Fish
Vegetables	Surface water	Surface water	Surface water
Groundwater	Vegetables	Milk	Vegetation
Honey	Forage	Sediment	
		Drinking and ground water	
		Terrestrial biota	

<sup>a</sup>Mixed fission and activation products.

<sup>b</sup>Transuranic nuclides, primarily <sup>239</sup>Pu.

Reproduced from CRC Press, Inc., Boca Raton, FL.

- evaluation of long-term trends in the environment;
- study of the fate of contaminants;
- baseline information and database maintenance;
- evaluation of offsite sources of radiation;
- compliance with regulations.

Factors that need to be considered include the following:

- potential hazards, including quantities and radio-toxicities;
- size and distribution of the exposed population;
- cost-effectiveness;
- availability of resources, including personnel, equipment, techniques, time, and money.

It must be possible to make adequate comparison with standards, reference materials, and background measurements.

The flow diagram given in **Figure 2** can be used as an aid in relating data requirements and critical path analysis to overall program planning.

In the case of nuclear facilities, both offsite and onsite sampling is required, in conjunction with pre-operational programs. A preoperational program should consist of:

- identification of critical exposure pathways;
- identification of population groups susceptible to exposure in the event of a release from the facility;
- selection of types of samples and sampling sites;

- accumulation of a background database;
- interpretation of information obtained.

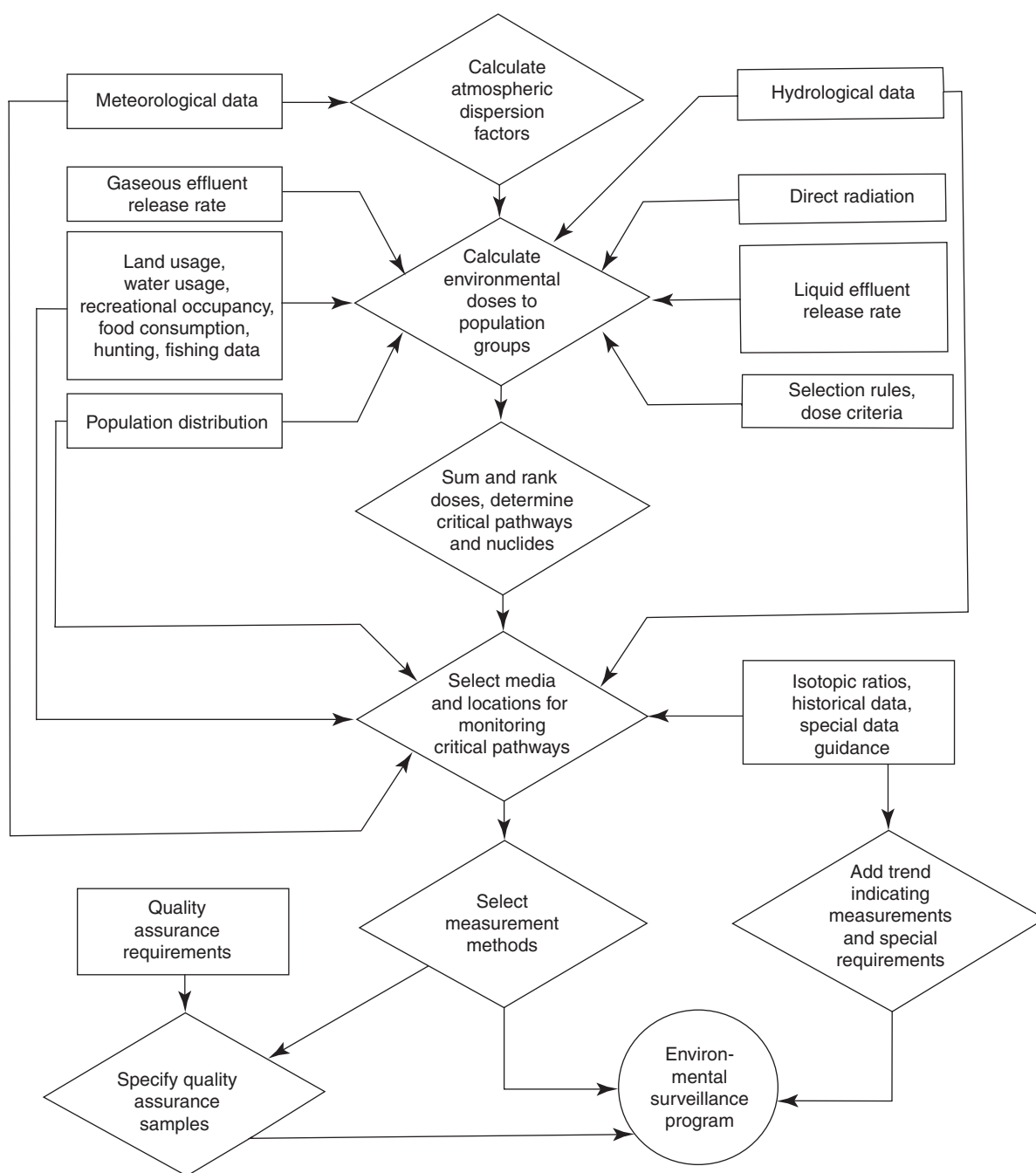
Once the facility is operational, continuous surveillance is needed to monitor effluents and the surrounding environment. Such programs rely heavily on the critical exposure pathway concept.

In general, sampling techniques for the monitoring of radionuclides in the event of a nuclear accident are based on standard procedures and are similar to those already discussed. It must be remembered that the sampling program is at least as important as the measurements made on the samples, and this is especially true for nuclear accidents.

### Analysis of Natural Radioactive Elements

Natural radioactive elements can be determined by measuring the activity of appropriate radioisotopes.

**Radon and radium** In nature, radon is most often associated with uranium deposits, since some of its isotopes are formed as part of the natural decay series. In isolated samples, radon will reach radioequilibrium in about 3 h, and the total rate of  $\alpha$  particle emission will be three times that for radon itself. This is because for each  $\alpha$  particle from radon decay, approximately one from <sup>218</sup>Po and one from <sup>214</sup>Po will also be emitted. Since the half-life of <sup>210</sup>Po (22 years) is long compared to analysis times, its decay can be neglected. However, since analysis times are still



**Figure 2** Environmental surveillance program design process. (Reproduced from CRC Press, Inc., Baco Raton, FL.)

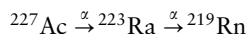
significantly long, the decay of the radon itself must be taken into account.

The determination of radium by the emanation method is based on the production of radon as a decay product of radium. From a liquid sample, the radon is first flushed out with a carrier gas. The end of the flushing is taken as the starting time for fresh radon accumulation in the sample. After a suitable time, the radon reaches equilibrium in the sample

and is then transferred to the counting device by a second flushing with a carrier gas (e.g., 15%  $\text{H}_2 + 85\% \text{N}_2$ ). After the radon isotopes have reached radioequilibrium with their daughters in the detector (3–4 h), a measurement is made. The measured radon activity is used to calculate the amount of radium in the sample. The method can detect as little as  $10^{-2}$  g of radium. Ionization chambers and  $\alpha$ -scintillation cells coupled to a

photomultiplier can be used for measurement of activity.

**Actinium** The nuclide  $^{219}\text{Rn}$  can be used for the determination of actinium:



The  $^{219}\text{Rn}$  activity of a sample is compared to that of a standard with a known content of  $^{223}\text{Ra}$ , which is prepared from a uranium ore containing a negligible amount of thorium. In such cases, it can be assumed that the ratio of  $^{235}\text{U}/^{238}\text{U}$  is 0.007. Since the decay products are short-lived, equilibrium is attained within a few minutes. Amounts of actinium as small as  $10^{-14}\text{ g}$  have been determined by this method.

**Thorium** The only gaseous product in the thorium decay series is thoron, which has a very short half-life (56 s). It can be used to determine  $^{220}\text{Rn}$ , and also its parent,  $^{232}\text{Th}$ . Thorium is determined by comparison with a standard, but because of the short half-life of  $^{220}\text{Rn}$ , decay correction must be made. Naturally occurring thorium cannot be used as a standard because of the presence of  $^{230}\text{Th}$ . If the sample is not in equilibrium, corrections must be made for the presence of long-lived isotopes such as  $^{228}\text{Ra}$  and  $^{228}\text{Th}$ . The  $^{220}\text{Rn}$  activity must be measured dynamically in the stream of carrier gas. A solution of known thorium content (1–10 mg) is used as a standard.

**Measurement of  $^{222}\text{Rn}$  and  $^{220}\text{Rn}$  in mixtures** The determination of  $^{222}\text{Rn}$  and  $^{220}\text{Rn}$  in mixtures is relatively simple because of the difference in half-life values (3.8 days and 56 s). One technique consists of measuring the combined activity of the nuclides as a function of time. The total activity first decreases by decay of  $^{220}\text{Rn}$ . After reaching a minimum value, it rises moderately with time because of the growth of daughter products of  $^{222}\text{Rn}$ . The ratio of the active components may be determined from an activity versus time curve.

**Protactinium** The significant isotope is  $^{231}\text{Pa}$ , a member of the  $^{235}\text{U}$  decay chain. Its radiometric determination is based on the measurement of  $\alpha$  activity, after a chemical separation of the element using a scintillation or semiconductor counter.

**Uranium** Determination of uranium by measuring its activity is of considerable practical significance in searching for uranium deposits, in its technical processing, and also when monitoring the environment. Direct methods, based on measurement of  $\alpha$ -particles

or photons emitted by uranium isotopes and indirect methods, based on measurement of other members of the decay chain, are used. The direct measurement of uranium in macroamounts is possible by use of the 0.184 MeV  $\gamma$ -ray of  $^{235}\text{U}$ . This technique also serves to determine the enrichment of uranium with  $^{235}\text{U}$ . Indirect methods are frequently used for uranium determination, even though difficulties can arise from the possible lack of equilibrium with decay products. It is therefore important to use methods that do not require equilibrium conditions, or to know the degree of equilibrium. Prospecting for uranium deposits with a GM or scintillation counter represents a field application of the indirect determination of uranium. Equilibrium is usually assumed, and the results are expressed as 'uranium equivalent' after comparison with a  $\text{U}_3\text{O}_8$  standard.

## Monitoring of Mixtures of Radionuclides

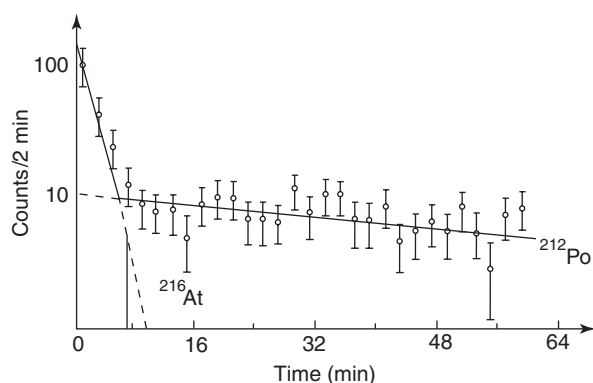
Monitoring of mixtures of radionuclides can include:

- identification of the number and nature of the radionuclides present in the mixture;
- determination of the relative radioactivity of each, e.g., the ratio of the count rates of one radionuclide in two complex mixtures of different composition;
- quantitative measurements, i.e., the determination of the quantity of each radionuclide (in becquerels).

The first of these tasks can be approached by considering the method of producing the mixture, the chemical behavior of the radioactive species compared to that of various inactive carriers, the half-life values, the type and energy of the radiation emitted, and the branching ratios of the radiation.

The second task is the one most frequently met in monitoring work. It can be solved in many different ways, but basically there are two typical approaches: (1) To add inactive carrier in known amounts, allow for isotope exchange, separate, and purify until further purification does not influence the characteristics of the radiation, and measure the radioactivity and chemical yield; the measurement of radioactivity in this case need not be discriminatory. (2) Without chemical pretreatment, to measure the intensity at one energy only (if no other radionuclides emitting at the same energy are present) or to use the difference in half-lives to differentiate between two (or more) radionuclides, as shown in Figure 3.

The third task, quantitative measurement, is equivalent to the determination of the number of



**Figure 3** Decay of  $^{214}\text{At}$  and  $^{212}\text{Po}$   $\alpha$  peak 8.78 MeV, 10 min irradiation of Th. Half-life of  $^{212}\text{Po}$  = 60.6 min, of  $^{214}\text{At}$  =  $1.28 \pm 0.19$  min. (Reproduced from Tölgyessy J and Kyrš M (1989) Radioanalytical Chemistry, vols. I and II. Chichester: Ellis Horwood.)

disintegrations of a given radionuclide per second (Bq). Absolute quantitative calibration here includes the determination of individual efficiency factors, which can be expressed as percentages and multiplied together to give the overall efficiency.

#### $\alpha$ -, $\beta$ -, and $\gamma$ -Spectrometry

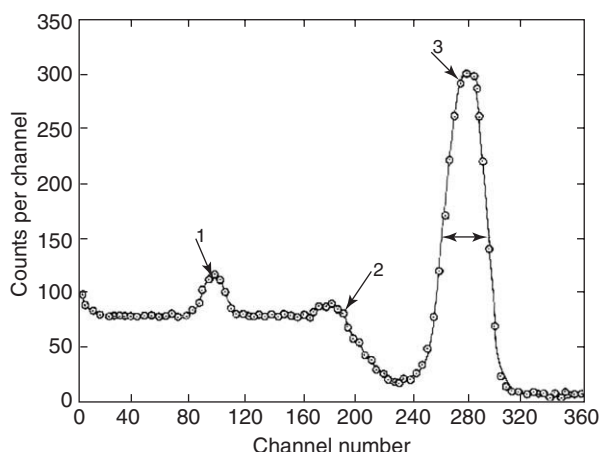
In  $\gamma$ -ray spectrometry, a mixture of  $\gamma$ -ray emitting radionuclides can be resolved quantitatively by pulse-height analysis. The analysis is based on the fact that the pulse heights (in volts) produced by a phototube are proportional to the amounts of  $\gamma$ -ray energy arriving at the scintillation detector. The amplification of the pulse is proportional to the voltage applied to the phototube, which can be adjusted so that the entire  $\gamma$ -spectrum can be examined.

In a single-channel counting procedure, only pulses of a certain height are counted; pulses of smaller or larger height are not registered ('seen') by the pulse-height analyzer. These pulse-height 'windows' are set by two pulse-height selectors.

In multichannel analysis the pulses are sorted according to height by a digital computer, and after a certain period of measurement, the distribution of the number of pulses according to height is obtained.

The number of counts registered in a certain channel (energy interval), plotted against channel number, is called a  $\gamma$  (or  $\alpha$ , or  $\beta$ ) spectrum. The  $\gamma$ -radiation of a simple radionuclide having a unique  $\gamma$ -energy does not produce a sharp spectral line, but a peak, owing to the statistical nature of the interaction of  $\gamma$ -rays with matter.

An NaI(Tl) scintillation detector spectrum for  $^{137}\text{Cs}$  is shown in **Figure 4**. The peak at channel No. 100 is due to backscatter and corresponds to 0.184 MeV; the edge at channel No. 180 is the Compton edge with energy 0.478 MeV, and the peak at



**Figure 4** NaI(Tl) spectrum of  $^{137}\text{Cs}$ . (Reproduced from Tölgyessy J and Kyrš M (1989) Radioanalytical Chemistry, vols. I and II. Chichester: Ellis Horwood.)

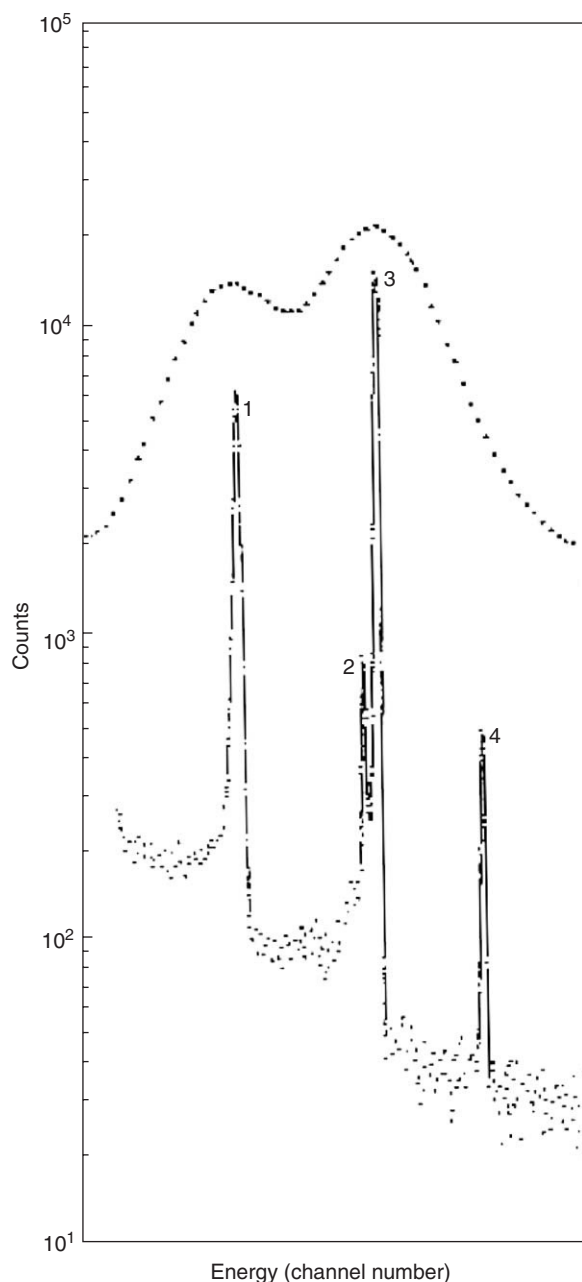
channel No. 280 is the photopeak (0.662 MeV), which is characteristic for  $^{137}\text{Cs}$ . The average width of the photopeak is 32 channels. The resolution (in per cent) characterizing the discriminating power of a detector is defined as  $100\delta E/E$ , where  $\delta E$  is the average width (full width of the peak at half of the maximum (FWHM) count level) of a photopeak, and  $E$  is its energy. For the purpose of finding the resolution of a detector (relative to a certain energy or source) the energy can be expressed in channel numbers. Therefore, in this case resolution with respect to  $^{137}\text{Cs}$  is  $100\delta E/E = 3200/280 = 11.5\%$ .

An energy calibration for a detector consists of finding the linear dependence between particle energy (MeV) and channel number.

It is evident that analyses of mixtures of radionuclides are not simple, for the multiplicity of types of interaction of  $\gamma$ -rays with the detector material generally leads to complicated  $\gamma$ -ray spectra. Besides the peak due to total absorption (photopeak at  $E$ ) peaks are caused by the following: pair production and escape of one or both annihilation photons ( $E$  0.51 or 1.02 MeV),  $180^\circ$  Compton scattering, single Compton scattering, multiple Compton scattering, external Compton scattering, and iodine K X-rays (from the NaI(Tl) crystal). For analysis and measurement of spectra of mixed  $\gamma$ -radiation, it is convenient and precise to measure only the contribution to the spectra resulting from total energy absorption.

If a semiconductor (such as hyper-pure germanium) is used instead of an NaI(Tl) scintillation detector, the resolution is improved by a factor of 30 and more. The tremendous increase in resolution compared to scintillation detectors is shown in **Figure 5**. Here, a spectrum is shown of lead bullet





**Figure 5**  $\gamma$ -Spectrum of an irradiated lead bullet. (Reproduced from Tölgyessy J and Kyrš M (1989) *Radioanalytical Chemistry*, vols. I and II. Chichester: Ellis Horwood.)

containing  $\sim 0.8\%$  Sb and 350 mg per kg As, which has been irradiated for 2 min at  $\sim 10^{12}$  neutrons  $\text{cm}^{-2}\text{s}$ . Peak 1 corresponds to 0.511 MeV (channel 226) from annihilation radiation, peak 2 to 0.559 MeV (channel 226) from  $^{122}\text{Sb}$ ; peak 4 to 0.603 MeV (channel 320) from  $^{124}\text{Sb}$ . The upper curve is an expanded portion of a spectrum taken with a  $3 \times 3$  in. NaI(Tl) crystal, with a resolution of 7% for 0.622 MeV. The lower curve is the spectrum of the same energy range taken with a Ge(Li)

low-energy photon detector, with an integral cooled preamplifier. On the other hand, the efficiency of Ge(Li) detectors is only 10–30% of that of a  $3 \times 3$  in. NaI(Tl) detector. The efficiency of one of the largest Ge(Li) detectors ever manufactured is 50% of the value that would be obtained from the standard  $3 \times 3$  in. NaI(Tl) detector.

The advent of high-resolution semiconductor  $\gamma$ -ray spectrometers opened a new era in the acquisition of accurate analytical data, which otherwise could not be obtained without great effort spent on chemical separation.

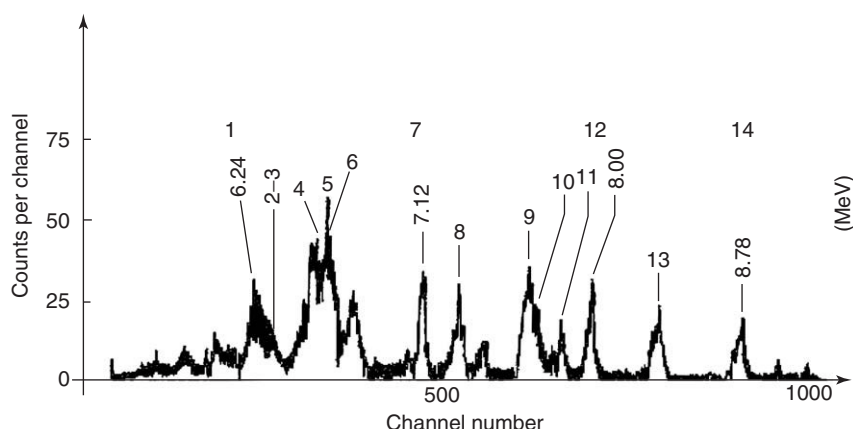
On the other hand, the interpretation of these spectra remains rather complicated, and analysis of the spectra is done by computer techniques. The major problem is the occurrence of interferences. Each peak used for quantitative analysis may have its own characteristic problems. The analysis begins with the integration of the counts in each peak of interest, and subtraction of the background counts, which are obtained from regions of the spectrum adjacent to the peaks.

$\beta$ -ray spectrometry, as a means of identifying and measuring mixtures of  $\beta$ -emitting nuclides by  $\beta$ -energy resolution, is very difficult because of the continuous energy spectra of  $\beta$ -radiation. On the other hand, monoenergetic conversion electrons yield spectra that can be relatively easily interpreted, and it is even possible to discriminate the peaks corresponding to the different binding energies of the electrons.

For  $\alpha$ -spectrometry, silicon charged-particle detectors (surface-barrier detectors) have been used. They can be used over an extensive range of energies (20 kV–200 MeV). The inherent resolution of these surface-barrier detectors is surpassed only by that of magnetic spectrometers.

Their resolution in a spectrum is calculated as  $\delta(\Delta E/\Delta C)$ , where  $\Delta E/\Delta C$  is the slope of the linear region of the calibration graph for  $\alpha$ -spectrometry (i.e., the slope of a plot of energy  $E$  (MeV) against channel number  $C$ ) and  $\delta$  is again the FWHM of the peak (in channel numbers). The resolution is then given in keV. In a typical  $^{210}\text{Po}$   $\alpha$ -spectrum, the energy peak at 5.305 MeV occurs in channels 512–528 with its maximum in channel 520; the resolution is  $(5.305/520) \times 8 = 0.082$  MeV.

Another advantage of these detectors is the fact that the detector output pulses rise rapidly, and hence they are well suited for fast ( $\sim 1$  ns) timing with coincidence circuitry in time-to-pulse-height converters. The efficiency of the active volume of these detectors is essentially 100%, and  $\Delta E/\Delta C$  is linear over a rather broad range. Compared to scintillation counters, gas proportional counters or ionization



**Figure 6**  $\alpha$ -Spectrum of spallation products after 1 min irradiation of Th with 600 MeV protons: 1 –  $^{226}\text{Th}$ ; 2 –  $^{226}\text{Th}/^{221}\text{Fr}$ ; 3 –  $^{222}\text{Ra}/^{223}\text{At}$ ; 4 –  $^{223}\text{Ac}/^{221}\text{Bi}$ ; 5 –  $^{222m}\text{Ac}/^{225}\text{Th}/^{221}\text{Ra}$ ; 6 –  $^{218}\text{Ra}$ ; 7 –  $^{219}\text{Fr}$ ; 8 –  $^{214}\text{Po}$ ; 9 –  $^{217}\text{Ra}$ ; 10 –  $^{218}\text{Fr}$ ; 11 –  $^{215}\text{At}$ ; 12 –  $^{213}\text{Po}$ ; 13 –  $^{214}\text{At}/^{212}\text{Po}$ . (Reproduced from Tölgyessy J and Kyrš M (1989) *Radioanalytical Chemistry*, vols. I and II. Chichester: Ellis Horwood.)

chambers, they also have good long-term pulse-height stability. Moreover, they are relatively inexpensive.

A typical detector may exhibit a resolution of 16 keV for  $^{241}\text{Am}$   $\alpha$  particles, have an active area of 50 mm<sup>2</sup>, and a depletion depth of 100  $\mu\text{m}$ .

An example of a complicated  $\alpha$  spectrum is given in Figure 6.

Prior chemical separation is used to simplify the subsequent determination based on physical principles. On the other hand, the introduction of this chemical operation renders the overall determination more complicated. Therefore, a decision to include chemical separation should be made carefully.

Prior separation can be either complete (i.e., individual substances are separated) or partial (with enrichment of some components in certain fractions). The latter is more frequently used when sophisticated measuring equipment is available. In many cases, prior chemical separation is done after the addition of isotopic carriers.

The main task in the monitoring of radionuclides in fission products (nuclides produced by primary and secondary processes associated with fission, generally of the uranium isotopes  $^{235}\text{U}$  and  $^{238}\text{U}$ ), transuranium nuclides, and transplutonium elements is the identification of nuclides and the determination of yield. A

combination of instrumental and chemical technique is required and the work is laborious and difficult.

**See also:** **Radiochemical Methods:** Overview; Natural and Artificial Radioactivity; Radiotracers; Gamma-Ray Spectrometry; Food and Environmental Applications. **Sampling:** Theory.

## Further Reading

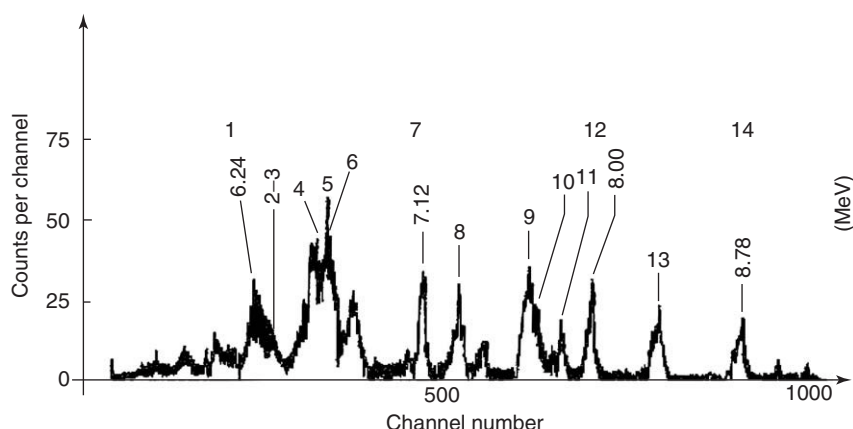
- Garten Rainer PH and Tölgyessy J (2001) *Radionuclides in Analytical Chemistry*, *Ullmann's Encyclopedia of Industrial Chemistry*, 6th edn. CD-Rom. Weinheim: Wiley-VCH.
- Tölgyessy J and Bujdosó E (1991) *Handbook of Radioanalytical Chemistry*, vols. I and II. Boca Raton: CRC Press.
- Tölgyessy J and Klehr EH (1987) *Nuclear Environmental Chemical Analysis*. Chichester: Ellis Horwood.
- Tölgyessy J and Varga Š (1974) *Nuclear Analytical Chemistry*, vol. III. Baltimore: University Park Press.
- Tölgyessy J, Dillinger P, and Harangozó M (2001) *Nuclear Chemistry*. Banská Bystrica: FPV UMB.
- Tölgyessy J and Harangozó M (2000) *Radioecology*. Banská Bystrica: FPV UMB.
- Tölgyessy J and Klehr EH (1991) *Jadernye metody chimicheskovo analiza okrushajushej sredy*. Moscow: Chimija.
- Tölgyessy J, Jakovlev JuV, and Bilimovitch GN (1985) *Diagnostika okrushajushej sredy radioanaliticheskimi metodami*. Moscow: Energoatomizdat.

## Technetium

**B Salbu**, Agricultural University of Norway, Ås, Norway  
**E Holm**, Lund University, Lund, Sweden

## Origin and Nature

The discovery of technetium is attributed to Perrier and Segre, who, in 1937, identified the element after



**Figure 6**  $\alpha$ -Spectrum of spallation products after 1 min irradiation of Th with 600 MeV protons: 1 –  $^{226}\text{Th}$ ; 2 –  $^{226}\text{Th}/^{221}\text{Fr}$ ; 3 –  $^{222}\text{Ra}/^{223}\text{At}$ ; 4 –  $^{223}\text{Ac}/^{221}\text{Bi}$ ; 5 –  $^{222m}\text{Ac}/^{225}\text{Th}/^{221}\text{Ra}$ ; 6 –  $^{218}\text{Ra}$ ; 7 –  $^{219}\text{Fr}$ ; 8 –  $^{214}\text{Po}$ ; 9 –  $^{217}\text{Ra}$ ; 10 –  $^{218}\text{Fr}$ ; 11 –  $^{215}\text{At}$ ; 12 –  $^{213}\text{Po}$ ; 13 –  $^{214}\text{At}/^{212}\text{Po}$ . (Reproduced from Tölgyessy J and Kyrš M (1989) *Radioanalytical Chemistry*, vols. I and II. Chichester: Ellis Horwood.)

chambers, they also have good long-term pulse-height stability. Moreover, they are relatively inexpensive.

A typical detector may exhibit a resolution of 16 keV for  $^{241}\text{Am}$   $\alpha$  particles, have an active area of 50 mm<sup>2</sup>, and a depletion depth of 100  $\mu\text{m}$ .

An example of a complicated  $\alpha$  spectrum is given in Figure 6.

Prior chemical separation is used to simplify the subsequent determination based on physical principles. On the other hand, the introduction of this chemical operation renders the overall determination more complicated. Therefore, a decision to include chemical separation should be made carefully.

Prior separation can be either complete (i.e., individual substances are separated) or partial (with enrichment of some components in certain fractions). The latter is more frequently used when sophisticated measuring equipment is available. In many cases, prior chemical separation is done after the addition of isotopic carriers.

The main task in the monitoring of radionuclides in fission products (nuclides produced by primary and secondary processes associated with fission, generally of the uranium isotopes  $^{235}\text{U}$  and  $^{238}\text{U}$ ), transuranium nuclides, and transplutonium elements is the identification of nuclides and the determination of yield. A

combination of instrumental and chemical technique is required and the work is laborious and difficult.

**See also:** **Radiochemical Methods:** Overview; Natural and Artificial Radioactivity; Radiotracers; Gamma-Ray Spectrometry; Food and Environmental Applications. **Sampling:** Theory.

## Further Reading

- Garten Rainer PH and Tölgyessy J (2001) *Radionuclides in Analytical Chemistry*, *Ullmann's Encyclopedia of Industrial Chemistry*, 6th edn. CD-Rom. Weinheim: Wiley-VCH.
- Tölgyessy J and Bujdosó E (1991) *Handbook of Radioanalytical Chemistry*, vols. I and II. Boca Raton: CRC Press.
- Tölgyessy J and Klehr EH (1987) *Nuclear Environmental Chemical Analysis*. Chichester: Ellis Horwood.
- Tölgyessy J and Varga Š (1974) *Nuclear Analytical Chemistry*, vol. III. Baltimore: University Park Press.
- Tölgyessy J, Dillinger P, and Harangozó M (2001) *Nuclear Chemistry*. Banská Bystrica: FPV UMB.
- Tölgyessy J and Harangozó M (2000) *Radioecology*. Banská Bystrica: FPV UMB.
- Tölgyessy J and Klehr EH (1991) *Jadernye metody chimicheskovo analiza okrushajushej sredy*. Moscow: Chimija.
- Tölgyessy J, Jakovlev JuV, and Bilimovitch GN (1985) *Diagnostika okrushajushej sredy radioanaliticheskimi metodami*. Moscow: Energoatomizdat.

## Technetium

**B Salbu**, Agricultural University of Norway, Ås, Norway  
**E Holm**, Lund University, Lund, Sweden

## Origin and Nature

The discovery of technetium is attributed to Perrier and Segre, who, in 1937, identified the element after

chemical isolation of products obtained from irradiation of molybdenum with deuterons. However, other claims to its discovery were made earlier, in 1908 by Ogawa, who named the element nipponium, and in 1925 by Noddack, Tacke, and Berg, who named it masurium.

The element, with atomic number 43 in the periodic table, was named technetium, a word derived from Greek meaning artificial. Thus, the name reflects the artificial origin of being produced by nuclear reactions, as no stable isotope of the element exists. Twenty-one radioactive isotopes of technetium and seven isomers are known from nuclear chemistry. Of these, only three isotopes have long physical half-lives:  $^{97}\text{Tc}$  ( $T_{1/2} = 2.6 \times 10^6$  years),  $^{98}\text{Tc}$  ( $T_{1/2} = 4.2 \times 10^6$  years), and  $^{99}\text{Tc}$  ( $T_{1/2} = 2.1 \times 10^5$  years). Being a long-lived fission product from neutron-induced fission of  $^{235}\text{U}$  and  $^{239}\text{Pu}$ ,  $^{99}\text{Tc}$  is the only technetium isotope of any radiological significance in the environment.

The major sources contributing to  $^{99}\text{Tc}$  in the environment are fallout from atmospheric nuclear weapons tests and releases from the nuclear fuel cycle, i.e., authorized or accidental releases from nuclear installations (e.g., reprocessing or enrichment plants, nuclear reactors), releases from waste disposal sites, and from dumping of nuclear materials. Contributions from natural processes, i.e., spontaneous fission of  $^{238}\text{U}$  in mineral ores such as pitchblende or nuclear reactions in molybdenum ores irradiated with cosmic-ray neutrons are negligible.

Technetium has been transferred to terrestrial and aquatic ecosystems due to the fallout from atmospheric nuclear weapons tests. It is assumed that 1 Mt fission energy corresponds to  $1.45 \times 10^{26}$  fissions and that  $^{137}\text{Cs}$  is produced with a representative fission yield of 5.57%. Assuming that the global stratospheric injection of nuclear debris was  $\sim 168.5$  Mt and a fission yield of  $^{99}\text{Tc}$  of  $\sim 6\%$ , similar to that of  $^{137}\text{Cs}$ , the global activity of  $^{99}\text{Tc}$  released into the stratosphere is 160 TBq. When local fallout is included, the total release of  $^{99}\text{Tc}$  into the environment from nuclear weapons tests between 1945 and 1963 should be within 180–200 TBq. Any releases from underground nuclear weapon tests are assumed to have a local impact only.

Releases from the nuclear fuel cycle may occur during reactor operations, nuclear fuel reprocessing,  $\text{UF}_6$  conversion, uranium enrichment, uranium fuel fabrication, high-level waste solidification, high-level and low-level waste disposals. The enrichment process, with a significant annual release through liquid discharges to the sea, is considered as the key contributor. So far, the total release of  $^{99}\text{Tc}$  to the

environment from the nuclear fuel cycle, e.g., reprocessing plants, is estimated to be  $\sim 1500$  TBq.

Reprocessing plants situated in coastal areas represent a major source of  $^{99}\text{Tc}$  to the marine environment. The Sellafield reprocessing plant on the Irish Sea has allowed the authorized release of  $\sim 770$  TBq to the marine system during 1952–94, with a maximum annual release of 178 TBq in 1978. Since 1994, the discharges from Sellafield into the Irish Sea increased significantly, due to processing of backlog liquid waste stored at the site since 1980s. The release of  $^{99}\text{Tc}$  in 1995 (190 TBq) was probably the highest annual release in the plant's history. In 2000, the permitted annual technetium discharges were reduced from 200 to 90 TBq, due to high activity concentrations in biota from the Irish Sea. Thus, during 1994–2001, additional 600 TBq was released to the marine environment. The release of  $^{99}\text{Tc}$  from the reprocessing plant Cap de la Hague, France, was estimated to be  $\sim 130$  TBq during 1976–2001. Information on authorized and accidental releases of  $^{99}\text{Tc}$  to terrestrial ecosystems or to freshwater drainage systems from large reprocessing facilities in the former USSR, e.g., Chelyabinsk (Mayak PA), Krasnoyarsk (KMCIC), or Tomsk (SCC) is not available. However, significant amount of  $^{99}\text{Tc}$  have been identified in ground water and artificial reservoirs associated with Techa River in the close vicinity of the Mayak PA.

Releases from nuclear power plants during normal operations are small, as  $\sim 1$  MBq (25% to the atmosphere) is released annually from a 1000 MW(e) boiling water reactor or pressurized water reactor. With a total capacity of 369 000 MW(e) in 1990 and an operational experience of 6000 reactor years with a mean capacity of 700 MW(e), the annual and integrated release of  $^{99}\text{Tc}$  corresponds to 350 MBq and 4 GBq, respectively. Uncontrolled releases during accidents are, however, more difficult to assess and are usually estimated from radionuclide ratios. In the fallout from the Chernobyl accident the radionuclide ratio  $^{99}\text{Tc}/^{137}\text{Cs}$  was reported to be  $1.5 \times 10^{-5}$ , i.e., significantly lower than the theoretical estimate for fission,  $1.43 \times 10^{-4}$ . Based on estimated total release of  $^{137}\text{Cs}$  and the reported radionuclide ratio, the total release of  $^{99}\text{Tc}$  from Chernobyl amounts to  $\sim 0.75$  TBq.

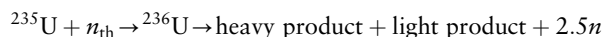
Information on releases of  $^{99}\text{Tc}$  from waste disposal sites or from dumping of wastes including reactors with fuel in the marine environment, e.g., Kara Sea including bays of Novaya Zemlya, is not available.

Radiopharmaceuticals containing  $^{99}\text{Tc}$  produced from molybdenum–technetium generators represent an additional source of  $^{99}\text{Tc}$  to the environment. The

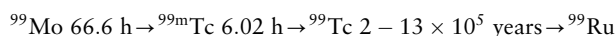
isotope ratio  $^{99}\text{Tc}/^{99\text{m}}\text{Tc}$  in a  $^{99}\text{Mo}$  generator used for medical purposes is  $\sim 0.1\text{--}2 \times 10^{-7}$  depending on the elution pattern. All the  $^{99}\text{TcO}_4^-$  produced from  $^{99}\text{Mo}$  generators will eventually reach the environment. With an annual discharge of  $\sim 200\text{ kBq}$  from an ordinary medical department, nuclear medicine will only contribute to minor releases of  $^{99}\text{Tc}$ .

## Chemical and Physical Properties

Technetium-99 is produced by fission, e.g.,



The products are neutron-rich  $\beta$ -emitters, and the precursors of  $^{99}\text{Tc}$  are the refractory elements zirconium, niobium, and molybdenum:



During release and initially after deposition of fission products, fractionation between volatile elements associated with condensed particles ( $^{137}\text{Cs}$ ,  $^{90}\text{Sr}$ ) and refractory elements associated with fuel particles ( $^{99}\text{Zr}$ ,  $^{99}\text{Nb}$ ) may take place. The behavior of  $^{99}\text{Tc}$  in the environment will, therefore, depend on whether technetium is released as such or formed from precursors after deposition. Technetium-99 is believed to be released from the nuclear fuel cycle as the volatile heptaoxide ( $\text{Tc}_2\text{O}_7$ ) in air emissions or as soluble and highly mobile pertechnetate ( $\text{TcO}_4^-$ ) in effluents.

Knowledge of the complex chemistry of technetium is primarily based on laboratory experiments using the isotopic tracers  $^{99\text{m}}\text{Tc}$  ( $T_{1/2} = 6.02\text{ h}$ ),  $^{97}\text{Tc}$  ( $T_{1/2} = 2.6 \times 10^6\text{ years}$ ),  $^{97\text{m}}\text{Tc}$  ( $T_{1/2} = 90\text{ days}$ ), and  $^{95\text{m}}\text{Tc}$  ( $T_{1/2} = 61\text{ days}$ ). In aqueous solution, the physicochemical forms of technetium will depend on pH and redox conditions (Eh). From Eh–pH diagrams it can be deduced that  $\text{TcO}_4^-$  is the only stable physicochemical form in aerobic natural waters. Under anaerobic conditions, however,  $\text{TcO}_4^-$  is reduced to  $\text{TcO}_2$ ,  $\text{TcO}(\text{OH})_2$ , or  $\text{TcCl}_6^{2-}$ . Technetium-species in oxidation state IV may associate with organic material such as humic substances in anoxic sediments, and thereby behave as organic colloids.

For  $^{99}\text{Tc}$ , the low-level energy  $\beta$ -particles emitted ( $E_{\text{max}} = 292\text{ keV}$ ) and the low concentrations in environmental samples represented an analytical challenge prior to the late 1970s. Based on a radiochemical procedure developed in 1969, the activity concentrations of  $^{99}\text{Tc}$  in waters ranged from  $0.019$  to  $1.8\text{ Bq l}^{-1}$ , and the detection limit was reached for 80% of the samples ( $n = 150$ ). For rainwater samples collected in 1967, the activity concentrations

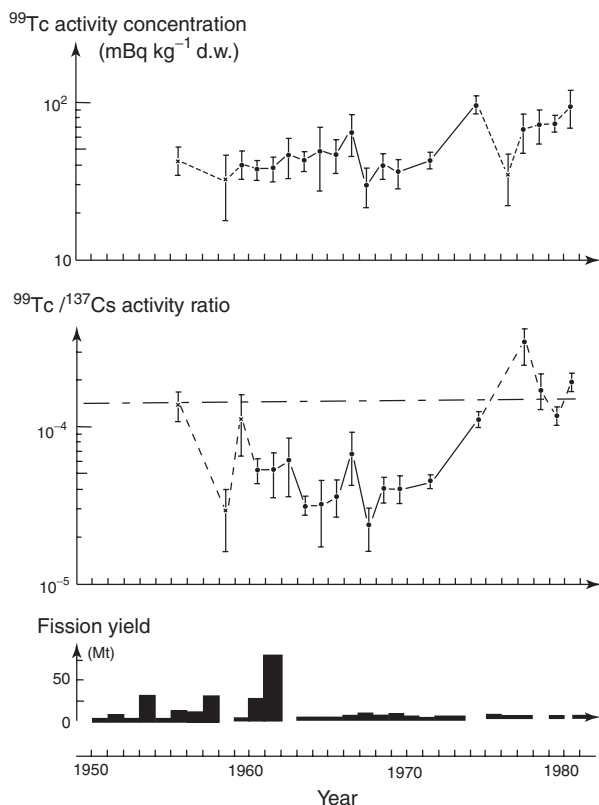
ranged from  $5.2 \times 10^{-5}$  to  $6.3 \times 10^{-4}\text{ Bq l}^{-1}$ , while  $7.5 \times 10^{-5}\text{ Bq l}^{-1}$  was reported for  $^{99}\text{Tc}$  in composite rain samples collected in 1975. Thus, analytical considerations should be taken when these results are compared with recent measurements.

During recent years, research has demonstrated that  $^{99}\text{Tc}$  behaves conservatively in natural water systems. Fractionation of discharge waters from the reprocessing plant in Sellafield showed that the colloidal fraction was significant for plutonium isotopes and refractory elements, while  $^{99}\text{Tc}$  was present as low molecular mass species, presumably  $^{99}\text{TcO}_4^-$ . Technetium-99 is therefore considered highly mobile and its retention in sediments and soils is most often negligible, especially when compared to transuranics. Since the 1980s,  $^{99}\text{Tc}$  and  $^{137}\text{Cs}$  released from European reprocessing facilities have successfully been used to trace water currents and identify water masses over temperate and Arctic waters of the North Atlantic. Analyses of  $^{99}\text{Tc}$  in groundwater from Nevada test sites also indicate the presence of mobile forms in solution, while  $^{99}\text{Tc}$  associated with colloids has been observed in ground water from Mayak. However, in reducing environments such as anoxic sediments,  $\text{Tc(VII)}$  can be reduced to  $\text{Tc(IV)}$ . Technetium in oxidation state IV can be retained in sediments or may associate with humic substances forming organic colloids. The relatively high retention of  $^{99}\text{Tc}$  in reservoir sediments at Mayak has also been attributed to the reduction of  $\text{Tc(VII)}$  in an anoxic environment.

## Common Matrices in Which the Element is Found

Deposition of  $^{99}\text{Tc}$  in terrestrial ecosystems is globally attributed to atmospheric fallout, while releases from installations associated with the nuclear fuel cycle may have a local or regional impact.

The deposition of  $^{99}\text{Tc}$  from weapon fallout has been estimated to be  $540 \pm 50\text{ mBq m}^{-2}$  at  $62^\circ\text{N}$ . Frequently, carpets of mosses and lichens are used as bioindicators for the estimation of deposited radionuclides. The concentration of  $^{99}\text{Tc}$  in lichens (mBq per kg dry weight), the fission yields (Mt) for nuclear weapons testing, and the observed and theoretical  $^{99}\text{Tc}/^{137}\text{Cs}$  radionuclide ratios are given in **Figure 1**. The present  $^{99}\text{Tc}/^{137}\text{Cs}$  ratio is expected to be  $3.2 \times 10^{-4}$  due to the physical decay of  $^{137}\text{Cs}$ . In general, the observed  $^{99}\text{Tc}/^{137}\text{Cs}$  ratio is lower than expected from fission probably due to a shorter residence time for  $^{99}\text{Tc}$  in the lichen carpet. The concentration of  $^{99}\text{Tc}$  and the  $^{99}\text{Tc}/^{137}\text{Cs}$  ratio in lichens has increased since 1975. Unusual mixtures of short-lived radioisotopes, e.g.,  $^{239}\text{Np}$  and  $^{99}\text{Mo}$ , in fallout



**Figure 1** Concentration of  $^{99}\text{Tc}$  (mBq kg<sup>-1</sup> dry weight) and  $^{99}\text{Tc}/^{137}\text{Cs}$  ratio in lichen. Individual results are marked with an x. The  $^{99}\text{Tc}/^{137}\text{Cs}$  ratio expected from fission is marked with a dotted line. In the lower part of the figure fission yields of atmospheric nuclear weapon tests are shown. (Reproduced with permission from Holm E and Riosco J (1987)  $^{99}\text{Tc}$  in the sub-arctic foodchain lichen-reindeer-man. *Journal of Environmental Radioactivity* 5: 343; © Elsevier.)

have also been reported on several occasions during 1975 and 1976, at a time when depositions from preceding nuclear tests were at their lowest.

Based on measurements of air filters from 1965 to 1967 and rainwater samples from 1967, the  $^{99}\text{Tc}/^{137}\text{Cs}$  ratio seems to be a factor of 10 higher than expected from the fission yield. The anomalous ratios of fission products observed in the atmosphere may partly be explained by fractionation of radionuclides during the detonation process. The precursors of  $^{137}\text{Cs}$  are gaseous or volatile elements, i.e., xenon and iodine, while the precursors of  $^{99}\text{Tc}$  are refractory elements, i.e., zirconium and niobium, which are usually incorporated in radioactive particles. Thus, the  $^{99}\text{Tc}/^{137}\text{Cs}$  ratio in the atmosphere may decrease with time after detonation due to the deposition of large radioactive particles. For deposited material releases of  $^{99}\text{Tc}$  with time should be expected due to weathering of particles. However, we cannot, at this stage, exclude additional sources contributing to releases of  $^{99}\text{Tc}$  to the atmosphere.

Such sources have not, however, significantly altered the integrated deposition.

Due to discharges from European reprocessing plants, radionuclide contamination of surface waters and marine biota are observed in the Irish Sea, North Sea, Norwegian Sea, Barents Sea, Kara Sea, North Atlantic, and the Polar Sea. The activity concentration in surface waters and biota such as *Fucus vesiculosus* decreases with distance to the source, e.g., reprocessing plants, or reflects global fallout. Using available data for  $^{90}\text{Sr}$ , which similar to technetium behaves conservatively in the water column, and known radionuclide ratios from fission, the concentration of  $^{99}\text{Tc}$  in surface waters is expected to be  $\sim 1 \text{ mBq m}^{-3}$ . However, from measurements of the bioindicator *F. vesiculosus*, having a concentration factor *F. vesiculosus* (dry weight)/water of  $\sim 10^5$ , the concentration in the surface waters should be  $\sim 10 \text{ mBq m}^{-3}$ . As seen in Table 1 enhanced values for  $^{99}\text{Tc}$  and for the  $^{99}\text{Tc}/^{137}\text{Cs}$  ratio are observed in areas exposed to releases from European reprocessing facilities. Radionuclide ratios in the range of 0.0018–0.049 are significantly much higher than those expected from fission.

Due to discharges from Mayak PA in the Urals, Russia, during 1949–51 into River Techa and during 1951 until present into Lake Karachay, groundwater, river water, and sediments (artificial reservoirs) were contaminated by a series of radionuclides, including  $^{99}\text{Tc}$ . In 1994, activity concentrations of  $^{99}\text{Tc}$  up to  $10 \text{ Bq l}^{-1}$  were observed in groundwater, reflecting vertical transport from the Lake Karachay. About 10–15% of  $^{99}\text{Tc}$  was associated with colloids, pseudocolloids, and particles. In 1996, the highest concentrations in river waters were observed in Reservoir 4, close to Mayak;  $\sim 2000 \text{ Bq l}^{-1}$  of  $^{137}\text{Cs}$ ,  $10 \text{ Bq l}^{-1}$  of  $^{99}\text{Tc}$ , and a  $^{99}\text{Tc}/^{137}\text{Cs}$  ratio of 0.005. The retention of  $^{99}\text{Tc}$  in sediments was relatively high ( $K_d \sim 10^4$ ) in the reservoirs and oxidizing reagents were needed for its extraction. Thus,  $^{99}\text{Tc}$  in the sediments was not present as pertechnetate, but was probably reduced and associated to organic sediment components. When sediments were exposed to artificial seawater, a few per cent of  $^{99}\text{Tc}$  could be remobilized. Thus,  $^{99}\text{Tc}$  could be released to the marine environment if contaminated sediments are transported by rivers to estuaries during flooding.

In areas contaminated mainly from global fallout, such as the Mediterranean, Davis Strait, and to some extent the Baltic proper before the Chernobyl accident, the concentrations and radionuclide ratios in surface waters are, however, also a factor of 5–20 higher than what should be expected from fission yields. This can partly be explained by the decay of



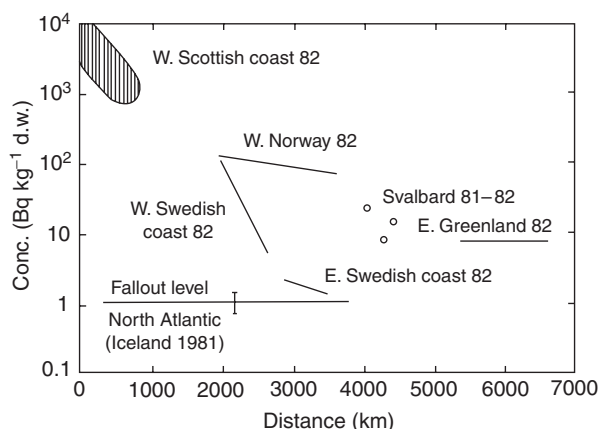
**Table 1** Activity concentrations of  $^{99}\text{Tc}$  and activity ratios to  $^{137}\text{Cs}$  in surface waters from different areas<sup>a</sup>

Year of collection	Sample area	$^{99}\text{Tc}$ ( $\text{mBq l}^{-1}$ )	$^{99}\text{Tc}/^{137}\text{Cs}$ activity ratio ( $\times 10^3$ )
1969	Irish Sea	$4.5 \pm 0.9$	8
1972	Irish Sea	$108 \pm 7$	36
1995	W. Irish Sea	$21 \pm 6$	—
1995	E. Irish Sea	100–500	1000–1600
1980	North Sea	$0.7 \pm 0.1$	—
1981	North Sea	$3.5 \pm 0.4$	20
1985–86	Mediterranean	$0.07 \pm 0.03$	20
1986–87	Mediterranean	$0.03 \pm 0.01$	0.8
1983	Baltic Sea	$0.07 \pm 0.01$	5
1986	Baltic Sea	$0.04 \pm 0.01$	0.2
1980	Norwegian Sea	$0.36 \pm 0.13$	7
1999	Norwegian Sea	$0.97 \pm 0.08$	17
1980	Barents Sea	$0.56 \pm 0.23$	7
1999	Barents Sea	0.3–1.1	—
1984	Davis Strait	$0.04 \pm 0.02$	11
1988	Greenland Sea	$0.08 \pm 0.01$	—
1999	Danish Straits	0.1–2	1–600
1994	Japan Sea	0.001–0.007	27
2001	Swedish W. Coast	$1.05 \pm 0.17$	18–85

<sup>a</sup>References: Brown JE, Iospje M, Kolstad KE, *et al.* (2002). Temporal trends for  $^{99}\text{Tc}$  in Norwegian coastal environments and spatial distribution in the Barents Sea. *Journal of Environmental Radioactivity* 60: 49; Lindahl *et al.* (2000); McCubbin D, Leonard KS, Brown J, *et al.* (2002) Further studies of the distribution of technetium-99 and Caesium-137 in UK and European coastal waters. *Continental Shelf Research* 22: 1417; Momoshima N, Sayad M, and Takashima Y (1995) Determination of Tc-99 in coastal seawater in Fukuoka, Japan. *Journal of Radioanalytical Nuclear Chemistry* 197: 245; Smith V, Fegan M, Pollard D, *et al.* (2001) Technetium-99 in the Irish marine environment. *Journal of Environmental Radioactivity* 56: 269.

$^{137}\text{Cs}$  and from a possible shorter residence time of  $^{137}\text{Cs}$  in the water column. The  $^{99}\text{Tc}/^{90}\text{Sr}$  ratio is also much higher than expected from fallout and the results are in agreement with more recent observations for atmospheric sample. We cannot, however, exclude systematically analytical errors in the earlier studies of  $^{99}\text{Tc}$  in seawater. After 1986, following the Chernobyl accident, a much lower  $^{99}\text{Tc}/^{137}\text{Cs}$  ratio is observed in contaminated areas since relatively small amounts of  $^{99}\text{Tc}$  were released. On the other hand, after 1994, much higher ratios ( $^{99}\text{Tc}/^{137}\text{Cs}$ ,  $^{99}\text{Tc}/^{90}\text{Sr}$ ) are found in areas contaminated from the discharges from Sellafield, due to the significant increase of  $^{99}\text{Tc}$  in effluents from that year.

To study the long-distance marine transport of  $^{99}\text{Tc}$  from European reprocessing facilities, samples of *Fucus* were collected in temperate and arctic regions of the North Atlantic and the Baltic Sea. **Figure 2** shows the activity concentration of  $^{99}\text{Tc}$  in *F. vesiculosus* ( $\text{Bq kg}^{-1}$  dry weight) as a function of distance from the Sellafield reprocessing plant. The

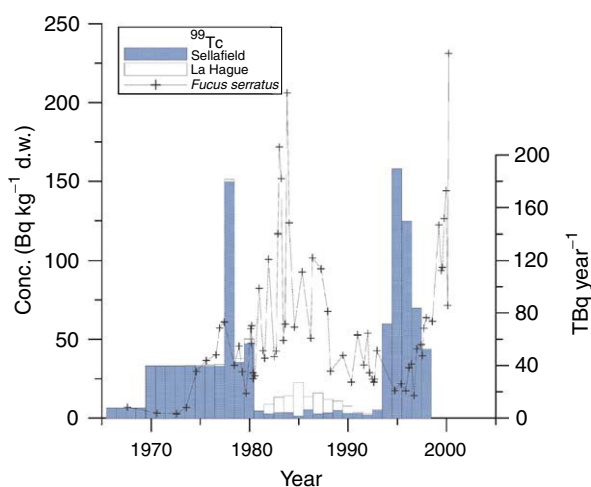


**Figure 2** Concentrations of  $^{99}\text{Tc}$  ( $\text{Bq kg}^{-1}$  dry weight) in *Fucus vesiculosus* from different areas as a function of transport distance (km) from European reprocessing facilities. The transit time (years) is indicated. (Reproduced with permission from Holm E and Rioseco J (1987).  $^{99}\text{Tc}$  in the subarctic foodchain lichen-reindeer-man. *Journal of Environmental Radioactivity* 5: 343; © Elsevier.)

estimated transit times to different regions is also indicated in the figure. The results show that  $^{99}\text{Tc}$  enters the North Sea and is effectively guided along the Norwegian coast by the Gulf current. The concentration decreases only by a factor of about three along the Norwegian coast (a distance of 2000 km), while the decrease along the Swedish coast into the Baltic Sea is more rapid due to the large yearly net outflow of water into the North Sea. The  $^{99}\text{Tc}$  activity concentrations in *F. serratus* at a site at the Swedish west coast (Särdal,  $56.76^\circ\text{N}$ ,  $12.63^\circ\text{E}$ ) collected from 1967 to 2001, with the annual releases from Sellafield and La Hague, are presented in **Figure 3**. The transit time of Sellafield-derived  $^{99}\text{Tc}$  to this site is shown to be 4–5 years. Utilizing the 1994/95  $^{99}\text{Tc}$  pulse from Sellafield, the transit time of  $^{99}\text{Tc}$  to the North Channel, North Sea, and Barents Sea seems to be more rapid than expected from previous estimates.

## Sampling, Handling, and Storage of Samples

The amount or volume of environmental samples needed for analysis for  $^{99}\text{Tc}$  depends on the degree of contamination (i.e., distance to the source) and the sensitivity of the measurement technique (i.e., detection limits). In the Irish Sea water, volumes of 10–50 L may be sufficient, while 50–100 l is needed in the Arctic Seas. For areas contaminated only from weapon fallout, volumes up to 1000 l may be needed. Especially in coastal areas or in high production areas filtration of waters through  $0.45 \mu\text{m}$  membrane is recommended, as algal materials, in particular,



**Figure 3** Technetium-99 activity concentration in *Fucus serratus* (primary axis) from Särda (56.76° N, 12.63° E), 1967–2000, and annual  $^{99}\text{Tc}$  activity discharge (secondary axis) from Sellafield (UK) and La Hague (France).

should be removed prior to the analysis. An effective method is to sorb technetium in the filtrate on activated carbon resins. Furthermore, the application of preconcentration techniques onboard the ship, e.g., anion-exchange chromatography or co-precipitation, are advantageous, as no storage effects will influence the analytical results. Also, the volumes transferred to laboratories are significantly reduced. Otherwise, waters should be stored cold (4°C) and in darkness in order to reduce algal or bacterial growth.

In freshwaters, *in situ* filtration and ultrafiltration techniques are useful to distinguish  $^{99}\text{Tc}$  associated with colloidal material such as humic substances.

In areas close to reprocessing plants the amount of marine bioindicators needed, i.e., *Fucus*, is ~1 g (dry weight), while in fallout areas 20–50 g (dry weight) should be sufficient. For lichens and mosses in fallout areas up to 200 g (dry weight) materials are usually needed. The bioindicators are most often air-dried prior to incineration and homogenization.

## Methods of Determination at Trace Levels

Even though neutron activation methods have been described for the determination of 0.1 pg amounts of technetium in environmental samples, radiochemical techniques have been most extensively used since the 1980s. In 1984, a general radioanalytical method was developed for the determination of  $^{99}\text{Tc}$  in environmental samples by radiochemical separation and low-level Geiger–Müller counting. A method suitable for large volumes of water was described in 1988. More sensitive analytical techniques using

mass spectrometry have been developed during the 1990s.

Determination of  $^{99}\text{Tc}$  in environmental samples represents an analytical challenge, as the low concentrations involved require the separation of a low-level  $\beta$ -emitter. The concentrations of  $^{137}\text{Cs}$  and  $^{210}\text{Pb}$  are frequently  $10^4$ – $10^6$  times higher than the concentrations of  $^{99}\text{Tc}$  and interferences from other major  $\beta$ -emitter contaminants such as  $^{103}\text{Ru}$ ,  $^{106}\text{Ru}$ , and  $^{110}\text{Ag}$  must be controlled. The separation of  $^{99}\text{Tc}$  is generally performed by anion-exchange chromatography followed by liquid–liquid extraction of technetium into cyclohexanone, extraction into 5% triisooctylamine, or extraction of  $^{99}\text{Tc}$  from a sulfuric acid solution into tributyl phosphate. Co-precipitation techniques in combination with anion-exchange chromatography have also been successfully applied.

When radiochemical separations are performed, the chemical yields reflecting possible losses must be determined. Among the tracers used (i.e.,  $^{99\text{m}}\text{Tc}$ ,  $^{97}\text{Tc}$ ,  $^{97\text{m}}\text{Tc}$ ,  $^{95\text{m}}\text{Tc}$ , and stable Re), the most suitable yield monitors, especially for preconcentration of large water volumes in the field, are  $^{99\text{m}}\text{Tc}$  and stable Re. When the short-lived  $^{99\text{m}}\text{Tc}$  obtained from  $^{99}\text{Mo}$  ( $T_{1/2} = 66.7$  h) is used, the Mo–Tc generator and a detector system must be brought in the field to perform yield determinations. When the stable chemical analog rhenium is used, the final source is a precipitate of the Re(Tc) tetraphenylarsonium salt,  $(\text{C}_6\text{H}_5)_4\text{AsReO}_4$ , and the recovery is obtained by gravimetric determination. When  $^{97\text{m}}\text{Tc}$  is used as a yield monitor, the quantities needed can frequently cause significant interference in the  $\beta$ -measurements. Although  $^{95\text{m}}\text{Tc}$  is more useful as a yield determinant than  $^{97\text{m}}\text{Tc}$ , it is not easily commercially available. The long-lived  $^{97}\text{Tc}$  is, however, useful in analysis of  $^{99}\text{Tc}$  using mass spectrometry techniques.

At very low levels of  $^{99}\text{Tc}$  in environmental samples and in the presence of much higher concentrations of interfering radionuclides, proper quality control may be a problem as  $\beta$ -spectrometry by solid-state detectors or liquid scintillation cannot be applied. So far, to our knowledge, two international analytical quality control exercises have been performed: one by the International Atomic Energy Agency, Monaco, using seaweed contaminated from a reprocessing plant and one performed by Environmental Measurement Laboratory, New York, using a grass sample contaminated from a gaseous diffusion plant. Both exercises were successful, but did not cover today's need for low-level intercalibration of  $^{99}\text{Tc}$  in environmental samples.

The recent development of mass spectrometry techniques such as inductively coupled plasma-mass

spectrometry (ICP-MS), high-resolution ICP-MS, accelerator mass spectrometry (AMS) improves the detection limits significantly (Table 2). For ICP-MS, the large range in the detection limits given reflects different instrument performances, depending on several factors such as systems (e.g., quadrupole) and type of sample introduction systems (e.g., nebulizers or ETV). Using time-of-flight instruments and AMS the detection limits can be lowered to the  $\mu\text{Bq}$  (fg) region. However, radiochemical separation is still needed prior to measurements and interferences from other nuclides, e.g.,  $^{99}\text{Ru}$  and  $^{99}\text{Mo}$ , should be removed or depressed. For analysis of  $^{99}\text{Tc}$  using ICP-MS and AMS, radiochemical separation based on ion chromatography (TEVA resins) and  $^{99\text{m}}\text{Tc}$  as yield monitor has proved successful.

## Areas of Special Interest and Safety Considerations

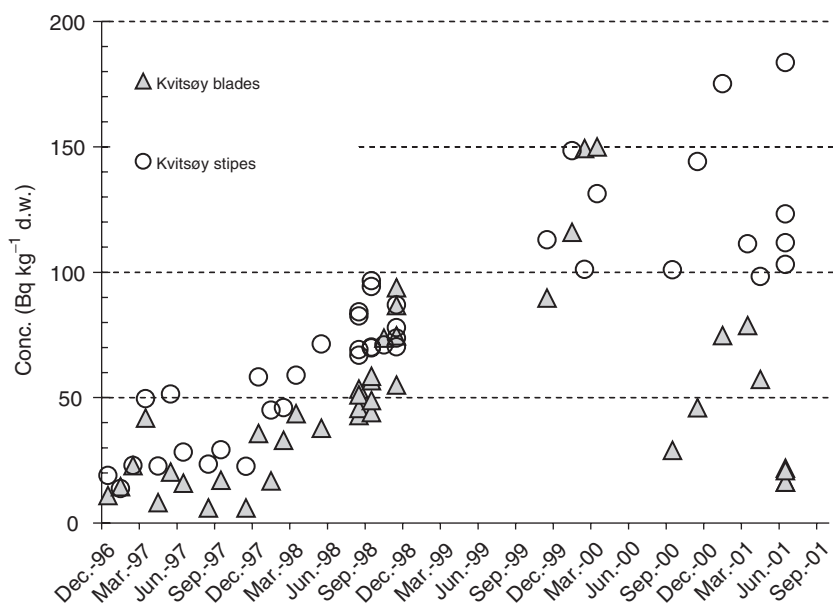
Radioecological interest in  $^{99}\text{Tc}$  started in the mid-1970s. In 1973, high concentration levels of  $^{99}\text{Tc}$

were reported in the seaweed *F. vesiculosus* as a result of effluents released from the nuclear fuel reprocessing plant on the Irish Sea coast. During the 1980s,  $^{99}\text{Tc}$  was successfully used together with  $^{137}\text{Cs}$  as a tracer to map water currents and to identify water masses over temperate and Arctic waters of the North Atlantic using known releases from European reprocessing facilities. As the disposal of high-level radioactive wastes in geological media could also result in releases, the potential migration of technetium in geological formations has also been discussed.

Information on biological uptake and accumulation of  $^{99}\text{Tc}$  in organisms is generally scarce, but has improved following the increased Sellafield discharges from 1994. In 1997, the  $^{99}\text{Tc}$  concentrations in crustacea, particularly in lobster, increased relatively rapidly, peaking at  $16\,000\text{ Bq kg}^{-1}$ , a factor of 10 higher than the European Council Food Intervention Level (CFIL) for  $^{99}\text{Tc}$  ( $1250\text{ Bq kg}^{-1}$  for adults) for postaccident situations. In 1998, the activity concentrations of  $^{99}\text{Tc}$  in *F. Vesiculosus* increased in the coastal area of Cumbria (up to  $200\,000\text{ Bq kg}^{-1}$ ), in the coastal areas of West Scotland (up to  $30\,000\text{ Bq kg}^{-1}$ ), in the coastal areas of Ireland (up to  $5000\text{ Bq kg}^{-1}$ ), and the North Sea ( $500\text{ Bq kg}^{-1}$ ) and the Norwegian Sea. As illustrated for *Laminaria hyperborea* (Figure 4), the accumulation of  $^{99}\text{Tc}$  in stipes exceeds that of the blades. The continuing accumulation of  $^{99}\text{Tc}$  in the algal stipes indicates that the peak effect of the discharges in *L. hyperborea* may not yet have occurred.

**Table 2** Limits of detection for the determination of  $^{99}\text{Tc}$

Analytical method	Detection limit (mBq)
Anti-coincidence shielded gas flow counting	1–5
Neutron activation ( $n, \gamma$ )	3–5
Liquid scintillation counting	1–25
ICP-MS	0.01–15
AMS	0.005–0.010



**Figure 4** Activity concentration of  $^{99}\text{Tc}$  in the stipes and blades of *Laminaria hyperborea* collected from the Western Coast of Norway, since 1997. (Mobbs H and Salbu B (2002) Temporal changes in technetium-99 activity in algae from the Norwegian Coast. In: Børretzen P, Jolle T, and Strand P (eds.) *Radioactivity in the Environment*, p. 610. Østerås: Norwegian Radiation Protection Authority.)

**Table 3** Concentration factors for Tc ( $\text{Bq kg}^{-1}$  biota/ $\text{Bq kg}^{-1}$  water)

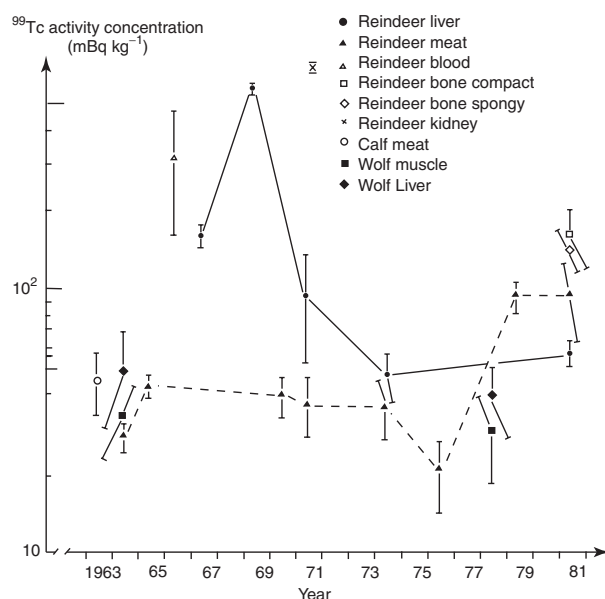
Specie	CF	Reference
Lobster (tail muscle)	6 850	Smith <i>et al.</i> (2001)
Lobster	8 000	Brown <i>et al.</i> (1999)
Lobster	1 000	IAEA (2004)
Prawns	2 800	Smith <i>et al.</i> (2001)
Mussels ( <i>Mytilus</i> )	970	Smith <i>et al.</i> (2001)
Mussels ( <i>Mytilus</i> )	486	Brown <i>et al.</i> (1999)
Oysters	280	Smith <i>et al.</i> (2001)
Molluscs	500	IAEA (2004)
Fish (flesh)	13	Smith <i>et al.</i> (2001)
Fish	80	IAEA (2004)
Zooplankton	100	IAEA (2004)
Phytoplankton	40	IAEA (2004)
Macroalgae	30 000	IAEA (2004)
<i>Fucus vesiculosus</i> <sup>a</sup>	132 000	Smith <i>et al.</i> (2001)
<i>Fucus vesiculosus</i> <sup>a</sup>	85 000	Holm <i>et al.</i> (1984)
<i>Fucus serratus</i> <sup>a</sup>	48 000	Holm <i>et al.</i> (1984)
<i>Fucus spiralis</i> <sup>a</sup>	46 000	Holm <i>et al.</i> (1984)
<i>Fucus disticus</i> <sup>a</sup>	102 000	Holm <i>et al.</i> (1984)
<i>Ascophyllum nodosum</i> <sup>a</sup>	178 000	Holm <i>et al.</i> (1984)
<i>Laminaria digitata</i> <sup>a</sup>	22 000	Holm <i>et al.</i> (1984)
<i>Halidrys siliquosa</i> <sup>a</sup>	63 000	Holm <i>et al.</i> (1984)
<i>Pelvetia canaliculata</i> <sup>a</sup>	102 000	Holm <i>et al.</i> (1984)

<sup>a</sup>Dry weight basis for algae.

For  $^{99}\text{Tc}$ , probably present as pertechnetate, the concentration factor (CF) varies according to biological species (Table 3). The CF is high for seaweed ( $\sim 10^5$ ) and low for fish fillet ( $\sim 10$ ). For lobster (tail muscle) recent results showed that the CF is  $\sim 10^3$ , which is a factor of 10 higher than previously measured in laboratory experiments. For edible parts of mussels and oysters, the CF is about 500 and 300, respectively. After the ingestion of contaminated lobster from the Irish Sea, 4–7% of  $^{99}\text{Tc}$  was retained in the body 1 year after intake. Thus, there seems to be a long-term compartment in the body retaining  $^{99}\text{Tc}$  and the dose conversion factor should be revised.

The possible uptake of technetium by plants and the risk of technetium transfer through the animal food-chain increased the environmental importance of this radionuclide. When studying  $^{99}\text{Tc}$  in carpets of lichen, the transfer to reindeer and to a limited degree to wolves and cattle was investigated. Figure 5 shows that the activity concentrations in liver and kidney of 2-year-old reindeer were significantly higher than those in their flesh.

A simple compartment model has been used to describe the transfer of technetium through the food-chain lichen-reindeer, as the uptake from food is much more important than from inhalation. The mean residence time for bone, liver, and flesh was 55, 11, and 5 days, respectively, giving a dose to the liver of  $0.02 \mu\text{Gy year}^{-1}$ , to the bone  $0.06 \mu\text{Gy year}^{-1}$ ,



**Figure 5** Activity concentrations of  $^{99}\text{Tc}$  ( $\text{mBq kg}^{-1}$  dry weight) in various reindeer tissues, wolves, and cattle during the period 1963–81. (Reproduced with permission from Holm E, Rioseco J, Ballestra S, and Walton A (1988) Radiochemical measurements of  $^{99}\text{Tc}$ : Sources and environmental levels. *Journal of Radioanalytical Nuclear Chemistry* 123: 167; © Elsevier.)

and to the flesh  $0.006 \mu\text{Gy year}^{-1}$ . Based on a literature survey, a  $^{99}\text{Tc}$  soil-to-plant concentration ratio of 50 ( $\text{Bq kg}^{-1}$  fresh weight tissue per  $\text{Bq kg}^{-1}$  dry weight soil) was applied, instead of the currently used value of 0.25, to evaluate the dose in man from a  $37 \text{GBq a}^{-1}$  release to the atmosphere from a hypothetical uranium enrichment facility. Later these dose evaluations were revised and maximum doses to the gastrointestinal tract and thyroid were estimated at 8 and  $13 \mu\text{Sv}$ , respectively, in an adult living 1600 m from the facility.

It should be emphasized that there are still sources and accumulating ecosystem compartments to be identified and our understanding of the biogeochemical behavior of technetium in the environment, especially in freshwater systems, is rather limited. Furthermore, information on the transfer to humans and the behavior of technetium in the human body is needed for assessing the long-term consequences of its release to the environment. However, major progress within the field of radioecology and dosimetry of technetium is expected in the years to come due to the development of highly sensitive analytical techniques such as high-resolution ICP-MS and AMS.

**See also:** Radiochemical Methods: Overview; Natural and Artificial Radioactivity; Radionuclide Monitoring; Food and Environmental Applications.

## Further Reading

- Aarkrog A, Carlson L, Chen QJ, *et al.* (1988) Origin of Technetium-99 and its use as a marine tracer. *Nature* 335: 338.
- Beasley TM, Gonor JJ, and Lorz HV (1982) A review of the biological and geochemical behaviour of technetium in the marine 10-environment. *Marine Environment Research* 7: 103.
- Chen Q, Dahlgaard H, Hansen HJM, and Aarkrog A (1990) Determination of  $^{99}\text{Tc}$  in environmental samples by anion exchange and liquid-liquid extraction at controlled valency. *Analytica Chimica Acta* 228: 163.
- Dahlgaard H, Hou X, and Nielsen S (2002) Radioactive tracers in Nordic waters,  $^{99}\text{Tc}$ ,  $^{137}\text{Cs}$  and  $^{129}\text{I}$ . In: Summaries of Studies Carried out in the NKS/BOK-2 project, p. 129.
- Geraedts K, Bruggeman C, Maes A, *et al.* (2002) Evidence for the existence of Tc(IV)-humic substance species by X-ray absorption near-edge spectroscopy. *Radiochimica Acta* 90(12): 879.
- Holm E, Gäfvert T, Lindahl P, and Roos P (2000) *In situ* sorption of technetium using activated carbon. *Applied Radiation and Isotopes* 53: 153.
- JRNEG (2004) Joint Russian-Norwegian Expert Group. Impacts on man and the environment in the northern areas from hypothetical accidents at Mayak PA, Urals, Russia. ISBN 82-996531-2-6.
- Keith-Roach M, Sturup S, Oughton DH, and Dahlgaard HH (2002) Comparison of two ICP-MS set-ups for measuring  $^{99}\text{Tc}$  in large volumes of water. *The Analyst* 127: 70.
- Lindahl P, Ellmark C, Gäfvert T, *et al.* (2003) Long-term study of  $^{99}\text{Tc}$  in the marine environment on the Swedish west coast. *Journal of Environmental Radioactivity* 67: 145.
- Morita S, Kim CK, Takaku Y, Seki R, and Ikeda N (1993) Determination of Technetium-99 in environmental samples by inductively coupled plasma mass spectrometry. *Radiochimica Acta* 63: 63.
- Salbu B, Lind OC, Børretzen P, and Oughton DH (2001) Advanced speciation techniques for radionuclides associated with colloids and particles. In: Brechignac F and Howard B (eds.) *Radioactive Pollutants – Impact on the Environment*, pp. 243–260. Les Ulis Cedex A: EDP Sciences.

## Radon

**L Tommasino**, National Agency for Environmental Protection and Technical Services, Rome, Italy

© 2005, Published by Elsevier Ltd.

## Introduction

Radon ( $^{222}\text{Rn}$ ) is the dominant source of human exposure to ionizing radiation in every country of the world. It is dominant in most circumstances: at home and at work, for individual persons and for whole populations. The worst characteristic of radon, apart from its carcinogenicity, is its ubiquity. Before radon became a matter of concern for human exposure, it was studied and measured for many purposes. It was an inert tracer for air masses, it was a geological indicator for radium and uranium, and it was a short-lived source of  $\gamma$ -radiation for cancer treatment. Radon plus beryllium was used as neutron source by Fermi for the discovery of neutron-induced fission reactions.

It would be difficult to design a radionuclide with characteristics more suitable for easy separation and determination than  $^{222}\text{Rn}$ . A noble gas with no stable compounds, it can be readily separated from the myriad of solid radionuclides and the other radon isotopes disappear rapidly by decay.

The radon can be absorbed on charcoal at room temperature or condensed at dry ice temperature.

Another useful radon characteristic is the decay to both short-lived and long-lived solid radionuclides. These can either be left with the radon to augment its radioactivity or separated on an air filter or electrostatic collector for separate measurement.

## Origin and Nature

Radon is a radioactive noble gas, discovered during the early studies of radioactivity at the turn of the century. The term radon was first introduced by Schmit in 1918 to designate the element with atomic mass 222 ( $^{222}\text{Rn}$ ), i.e., the gas associated with uranium-238 ( $^{238}\text{U}$ ). In precise usage, the term radon has come to denote the element with atomic number 86. This element has 26 isotopes ranging from  $^{199}\text{Rn}$  to  $^{226}\text{Rn}$ , three of which are naturally occurring. These are  $^{219}\text{Rn}$  (initially called actinon) from the  $^{235}\text{U}$  decay series discovered by Debierne and Giesel in 1903,  $^{220}\text{Rn}$  (thoron) from the thorium decay series discovered by Owens and Rutherford in 1889, and  $^{222}\text{Rn}$  (radon) from the  $^{238}\text{U}$  series discovered by Dorn in 1901. In the following the term radon will refer to  $^{222}\text{Rn}$  unless otherwise stated. It was William Ramsay who first separated radon and determined the quantity of radon in equilibrium with 1 g of radium. This quantity became the original definition of the radioactivity unit called the curie. Moreover,

## Further Reading

- Aarkrog A, Carlson L, Chen QJ, *et al.* (1988) Origin of Technetium-99 and its use as a marine tracer. *Nature* 335: 338.
- Beasley TM, Gonor JJ, and Lorz HV (1982) A review of the biological and geochemical behaviour of technetium in the marine 10-environment. *Marine Environment Research* 7: 103.
- Chen Q, Dahlgaard H, Hansen HJM, and Aarkrog A (1990) Determination of  $^{99}\text{Tc}$  in environmental samples by anion exchange and liquid-liquid extraction at controlled valency. *Analytica Chimica Acta* 228: 163.
- Dahlgaard H, Hou X, and Nielsen S (2002) Radioactive tracers in Nordic waters,  $^{99}\text{Tc}$ ,  $^{137}\text{Cs}$  and  $^{129}\text{I}$ . In: Summaries of Studies Carried out in the NKS/BOK-2 project, p. 129.
- Geraedts K, Bruggeman C, Maes A, *et al.* (2002) Evidence for the existence of Tc(IV)-humic substance species by X-ray absorption near-edge spectroscopy. *Radiochimica Acta* 90(12): 879.
- Holm E, Gáfvert T, Lindahl P, and Roos P (2000) *In situ* sorption of technetium using activated carbon. *Applied Radiation and Isotopes* 53: 153.
- JRNEG (2004) Joint Russian-Norwegian Expert Group. Impacts on man and the environment in the northern areas from hypothetical accidents at Mayak PA, Urals, Russia. ISBN 82-996531-2-6.
- Keith-Roach M, Sturup S, Oughton DH, and Dahlgaard HH (2002) Comparison of two ICP-MS set-ups for measuring  $^{99}\text{Tc}$  in large volumes of water. *The Analyst* 127: 70.
- Lindahl P, Ellmark C, Gáfvert T, *et al.* (2003) Long-term study of  $^{99}\text{Tc}$  in the marine environment on the Swedish west coast. *Journal of Environmental Radioactivity* 67: 145.
- Morita S, Kim CK, Takaku Y, Seki R, and Ikeda N (1993) Determination of Technetium-99 in environmental samples by inductively coupled plasma mass spectrometry. *Radiochimica Acta* 63: 63.
- Salbu B, Lind OC, Børretzen P, and Oughton DH (2001) Advanced speciation techniques for radionuclides associated with colloids and particles. In: Brechignac F and Howard B (eds.) *Radioactive Pollutants – Impact on the Environment*, pp. 243–260. Les Ulis Cedex A: EDP Sciences.

## Radon

**L Tommasino**, National Agency for Environmental Protection and Technical Services, Rome, Italy

© 2005, Published by Elsevier Ltd.

## Introduction

Radon ( $^{222}\text{Rn}$ ) is the dominant source of human exposure to ionizing radiation in every country of the world. It is dominant in most circumstances: at home and at work, for individual persons and for whole populations. The worst characteristic of radon, apart from its carcinogenicity, is its ubiquity. Before radon became a matter of concern for human exposure, it was studied and measured for many purposes. It was an inert tracer for air masses, it was a geological indicator for radium and uranium, and it was a short-lived source of  $\gamma$ -radiation for cancer treatment. Radon plus beryllium was used as neutron source by Fermi for the discovery of neutron-induced fission reactions.

It would be difficult to design a radionuclide with characteristics more suitable for easy separation and determination than  $^{222}\text{Rn}$ . A noble gas with no stable compounds, it can be readily separated from the myriad of solid radionuclides and the other radon isotopes disappear rapidly by decay.

The radon can be absorbed on charcoal at room temperature or condensed at dry ice temperature.

Another useful radon characteristic is the decay to both short-lived and long-lived solid radionuclides. These can either be left with the radon to augment its radioactivity or separated on an air filter or electrostatic collector for separate measurement.

## Origin and Nature

Radon is a radioactive noble gas, discovered during the early studies of radioactivity at the turn of the century. The term radon was first introduced by Schmit in 1918 to designate the element with atomic mass 222 ( $^{222}\text{Rn}$ ), i.e., the gas associated with uranium-238 ( $^{238}\text{U}$ ). In precise usage, the term radon has come to denote the element with atomic number 86. This element has 26 isotopes ranging from  $^{199}\text{Rn}$  to  $^{226}\text{Rn}$ , three of which are naturally occurring. These are  $^{219}\text{Rn}$  (initially called actinon) from the  $^{235}\text{U}$  decay series discovered by Debierne and Giesel in 1903,  $^{220}\text{Rn}$  (thoron) from the thorium decay series discovered by Owens and Rutherford in 1889, and  $^{222}\text{Rn}$  (radon) from the  $^{238}\text{U}$  series discovered by Dorn in 1901. In the following the term radon will refer to  $^{222}\text{Rn}$  unless otherwise stated. It was William Ramsay who first separated radon and determined the quantity of radon in equilibrium with 1 g of radium. This quantity became the original definition of the radioactivity unit called the curie. Moreover,



Ramsay was among the first to study the biological effects of the radioactive gas radon.

## Chemical and Physical Properties

Chemically, radon is a noble gas. As such, it is colorless, odorless, and almost chemically inert. Although radon is not chemically active, it is interesting to note that radon is not a totally 'inert' gas either. Studies on radon chemistry have been reported in which compounds such as clathrates and complex fluorides have been formed. Compared with the other noble gases, radon is the heaviest and has the highest melting point, boiling point, critical temperature, and critical pressure.

Some physical properties of radon are listed in Table 1. Radon is moderately soluble in water and, therefore, can be absorbed by water flowing through rock and sand containing radon. Its solubility depends on water temperature; the colder the water, the greater is radon's solubility. A measure of gas solubility in water is given by the solubility coefficient, which is defined as the ratio of the radon concentration in water to that in air. At 20°C, the solubility coefficient is  $\sim 0.25$ , which means that radon is distributed preferentially in air rather than in water (in

the ratio of 4:1 at 20°C). Another important property of radon is the fact that it is most soluble in organic liquids and it is readily absorbed on charcoal and silica gel. In a multiphase system at normal environmental temperature, radon concentrations are greatest, intermediate, and least in the organic liquid, gas, and the water phases, respectively. If the temperature is increased, the concentration in the gas phase increases at the expense of that in the liquid phase. These properties are successfully exploited for the determination of radon gas and for its extraction from other gases and/or liquids. For example, radon in water can be easily measured after its extraction by bubbling air through the water or by extracting radon from the water with organic liquid scintillators. Radon gas can be effectively removed from a sample of air by passing it through activated charcoal cooled to the temperature of solid carbon dioxide ( $-78.5^\circ\text{C}$ ). Radon desorption can be achieved by heating the charcoal to  $350^\circ\text{C}$ .

## Relative Abundance of the Radon Isotopes

The amounts of radon isotopes in the environment depend primarily upon the concentrations of  $^{238}\text{U}$ ,  $^{235}\text{U}$ , and  $^{232}\text{Th}$  (thorium-232) in the soil and rocks. Although these isotopes are radioactive, their half-lives are so long that the time since the formation of universe (i.e., the time since these radionuclides were formed) is not sufficient for them to have decayed to stable elements. In particular, the half-life of  $^{238}\text{U}$  is nearly equal to the estimated age of the earth. Only half of the earth's original endowment of uranium has decayed away. The abundance of  $^{232}\text{Th}$  in the earth's crust is somewhat higher than that of  $^{238}\text{U}$ , but because of the longer half-life of  $^{232}\text{Th}$  (see Table 2), the average rate of production of  $^{220}\text{Rn}$  in the ground is about the same as that of  $^{222}\text{Rn}$ .

Within the natural uranium isotopes, the relative abundance by weight is 99.28% for  $^{238}\text{U}$  and 0.72% for  $^{235}\text{U}$ . Hence,  $^{238}\text{U}$  predominates over  $^{235}\text{U}$  in the environment. These abundance ratios are necessarily the same for all the uranium isotopes on the earth, on all the planets, on meteorites, etc. A number of people have expressed surprise at the discovery of radon

**Table 1** Physical properties of  $^{222}\text{Rn}$

Property	Value
Atomic number	86
Boiling point at normal temperature and pressure	$-61.8^\circ\text{C}$
Density at normal temperature and pressure	$9.96\text{ kg m}^{-3}$
Coefficient of solubility at atmospheric pressure in water at the temperature of:	
0°C	0.57
20°C	0.250
37°C	0.167
100°C	0.106
Coefficient of solubility at atmospheric pressure and 18°C in:	
Hexane	16.56
Olive oil	29.00
Petroleum (liquid paraffin)	9.20
Toluene	13.24

Condensed data derived from NCRP (1988) *Measurements of Radon and Radon Daughters in Air*. National Council on Radiation Protection and Measurements, Report no. 97.

**Table 2** Properties of some members of the natural radioactive series

Series	Long-lived parent		Crustal abundance			Noble gas member	
	Isotope	Half-life (years)	$\mu\text{g kg}^{-1}$	$\text{Bq kg}^{-1}$	$\text{pCi g}^{-1}$	Isotope	Half-life
Uranium	$^{238}\text{U}$	$4.5 \times 10^9$	2.7	33	0.89	$^{222}\text{Rn}$	3.82 days
Thorium	$^{232}\text{Th}$	$14.1 \times 10^9$	8.5	34	0.92	$^{220}\text{Rn}$	55.6 s
Actinium	$^{235}\text{U}$	$0.7 \times 10^9$	0.02	1.5	0.04	$^{219}\text{Rn}$	3.96 s

on the moon. The real surprise should have been if no radon had been found, since such a finding would certainly require some new theories on the origin of the moon.

The slow decay of  $^{235}\text{U}$ ,  $^{238}\text{U}$ , and  $^{232}\text{Th}$  occurs through a long and complicated series of radioactive elements until they finally become stable lead isotopes. Most of the intermediate nuclides are isotopes of metals and are chemically reactive. Thus, they tend to stay within the material in which the uranium or thorium atom was originally deposited. The only exception is radon, which being a noble gas, does not interact chemically with other elements, and is free to move by diffusion and convection.

Of the three naturally occurring radon isotopes, only  $^{222}\text{Rn}$  has a sufficiently long half-life (3.825 days) to allow for release from the soil and rocks, where it was generated. This half-life is short enough to restrict transport by pure diffusion to short distances only, but once the radon has left the solid material and has become mixed with air, convection transport over longer distances (several meters) from the soil into both outdoor air and indoor environments is possible. In contrast, substantially less  $^{220}\text{Rn}$  reaches the atmosphere because its short half-life (55.6 s) limits the distance it can travel before decay. As for  $^{219}\text{Rn}$  (formed in the decay chain of  $^{235}\text{U}$ ), it

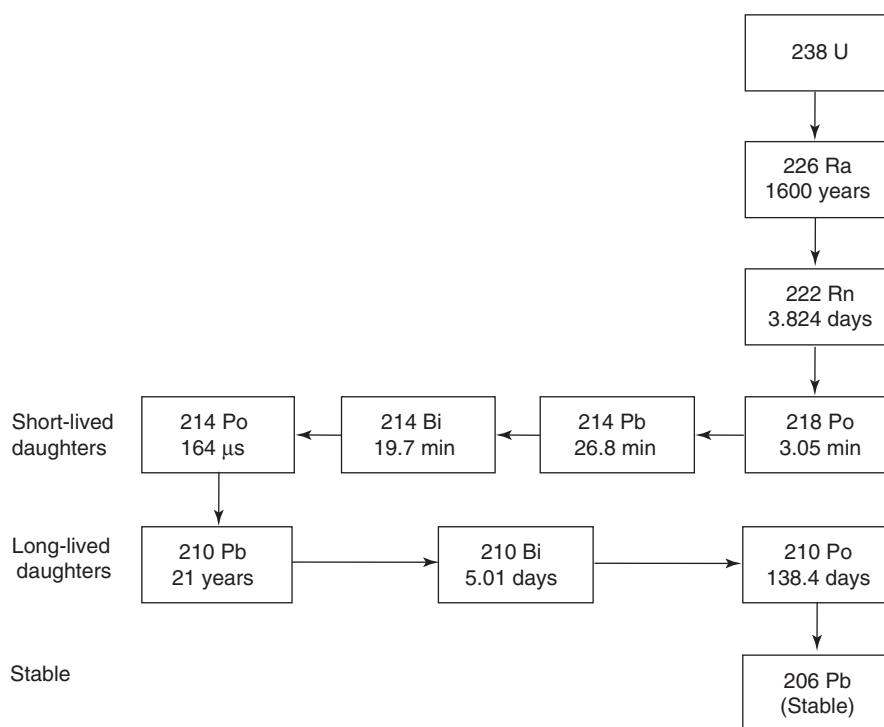
may be ignored entirely because of its short half-life (3.96 s) and the comparative scarcity of  $^{235}\text{U}$ .

## $^{222}\text{Rn}$ and its Decay Products

$^{222}\text{Rn}$  is preceded in the  $^{238}\text{U}$  decay series by radium ( $^{226}\text{Ra}$ ), which has a half-life of 1600 years.  $^{226}\text{Ra}$  decays to stable lead ( $^{206}\text{Pb}$ ), producing a chain of radioactive isotopes of solid elements. The decay chain of  $^{238}\text{U}$  is shown in Figure 1.

While the major removal of radon gas from air is by radioactive decay, the chemically radioactive decay products (formerly called radon daughters) may additionally be removed by processes such as wash-out and plating-out on surfaces. In air inside buildings, these radioactive decay products can attach themselves to walls, floors, people, or airborne particles that are inhaled into the lungs. Unattached radon decay products can also be inhaled and, subsequently, can become deposited on the lung tissue. As a consequence, radon decay products are seldom in radioactive equilibrium with radon in the lower atmosphere (near the earth's surface) or indoors.

An important characteristic from the radiological point of view is the fact that the four immediate radon decay products have half-lives of less than



**Figure 1** Uranium decay series showing radon and its decay products. (Reproduced with permission from Khan HA (1993) Passive dosimetry of radon and its daughters using solid state nuclear track detectors (SSNTDS). *Radiation Protection Dosimetry* 46: 149–170.)

30 min (Figure 1). Once inhaled, these products can undergo decay before they are removed by lung clearance mechanisms. In particular, the  $\alpha$ -particles emitted on decay are characterized by high linear energy transfer radiation, i.e., radiation with high biological effectiveness. Because of their short range in tissue (40–70  $\mu\text{m}$ ),  $\alpha$ -particles are stopped by the skin, but when emitted from inhaled decay products they irradiate epithelial cells, particularly in the bronchial region. The potential effect of the dose imparted to the lungs is an increase in the risk of lung cancer, as first observed with uranium miners. Because of its ubiquity, radon is the dominant source of human exposure to ionizing radiations.

### Radon in the Ground (Soil, Rocks, and Water)

The precursors of radon, namely radium and ultimately uranium, are present in all types of rocks and soil. Their amounts vary with the specific site and geological material. Uranium has a tendency to concentrate; it concentrates in magmas that form granites and in shales formed from marine mud high in organic matter. Because of uranium's chemical affinity for phosphates, phosphatic rocks often contain greatly elevated levels of uranium. Whenever rock is heated in the presence of a fluid, uranium has a tendency to move with the fluid until conditions change, when it stops and becomes concentrated. The end result is a wide variability in uranium concentration, even within the same rock formation or the same rock type. The concentration of both uranium and radium in the earth's crust as a whole is generally different from that of the soil. This is because soil concentrations are affected by the extent of leaching, porosity, precipitation from groundwater, etc. The effectiveness of radium in supplying radon to the soil pores for transport to the atmosphere depends both on the total concentration of radium atoms and on the fraction of those atoms in the soil (or rocks), which are located on the soil particle surfaces, so that the newly formed radon atoms can escape into the pores and capillaries.

For efficient release of radon into the air spaces of soil, the radon atom must be formed within 20–70 nm of the mineral surface for most common minerals. This distance is the recoil range of a radon atom at the instant of its formation from the decay of radium. The processes by which radon atoms escape from a given material are referred to as radon emanation. The emanation power or the coefficient of emanation is defined as the ratio of the number of radon atoms that escape from the solid to the number of radon atoms formed by radioactive decay of radium in the solid. The emanation power varies from about 0.02 to 0.7, depending upon the mineral structure and the water content.

A correlation between the amount of radon in the soil gas (or water) and the levels of  $^{238}\text{U}$  and  $^{226}\text{Ra}$  only exists for two extremes: the radon concentration in the soil gas is likely to be very low or very high if the radium content of the source materials (rock or sediment) is very low or very high, respectively. Outside these two extremes other factors dominate in controlling radon concentrations, with those factors responsible for the radon transport processes being especially important. These factors include soil porosity and permeability, density, moisture, barometric pressure, temperature, thickness of the soil over bedrock, and, in some cases, the underlying bedrock.

The most soluble of the noble gases, radon can be transported considerable distances by carrier gases or liquids such as carbon dioxide and water. The fracturing and faulting of a rock can alter its radon potential in several ways. Fracturing and faulting can create extensive migration pathways for radon, thus increasing the radon flow. Furthermore, fractures and faults are sometimes associated with elevated uranium concentrations because uranium-bearing fluids deposit the uranium within fracture or fault zones. Fractures and fault zones can be successfully identified by monitoring the radon concentration in the soil and they may be associated with houses that have the most severe indoor radon problems. Granitic rocks rich in uranium, especially when permeable and fractured, create the highest in-water concentration. Water penetrates into the voids present in the rocks and soil and dissolves radon that emanates into these spaces. Since seismic crustal activity should, in principle, affect the radon in the soil and groundwater, radon monitoring in the soil and well water has been used to study the complex precursor phenomena of earthquakes. When measuring radon in soil, the effects of changing weather on radon concentration are likely to be greater near the surface. A compromise between reducing the effects of weather and the practicability of sampling under field conditions is achieved by sampling soil radon from a depth of 0.5–0.7 m.

### Radon in the Atmosphere

Radon enters the atmosphere principally by crossing the soil–air interface. The transfer rate of radon across the interface between a solid phase and the atmosphere is referred to as a radon flux or exhalation rate. In SI units it is measured in  $\text{Bq m}^{-2} \text{s}^{-1}$ . The radon flux gives a measure of the source strength and varies strongly from soil to soil. A global average of  $\sim 17 \text{ mBq m}^{-2} \text{s}^{-1}$  from continental soil has been estimated.

Secondary sources of radon include the oceans and fluids that leave the earth and subsequently come

**Table 3** Sources of global atmospheric radon

Source	Input to atmosphere (million Ci per year) <sup>a</sup>
Emanation from soil	2000
Ground water	500
Emanation from oceans	30
Phosphate residues	3
Uranium mill tailings	2
Coal residues	0.02
Natural gas	0.01
Coal combustion	0.001

<sup>a</sup> 1 Ci =  $3.7 \times 10^{10}$  Bq.

From NCRP (1984) *Evaluation of Occupational and Environmental Exposures to Radon and Radon Daughters in the United States*. National Council on Radiation Protection and Measurements, Report no. 78.

into contact with the atmosphere (such as groundwater, natural gases, geothermal fluids, volcanic gases). Overall, as a global average, ~80% of the radon emitted into the atmosphere comes from the top layers of ground. **Table 3** lists the major sources of global atmospheric radon. The concentration of the radon in the atmosphere is governed by the source strength and dilution factors, both of which are strongly affected by meteorological conditions, particularly temperature, humidity, atmospheric pressure, and wind conditions at the surface. As a consequence, radon concentration exhibits both daily and seasonal variations that are often cyclic. Because of its modest half-life and origin at the ground surface, atmospheric radon exhibits a vertical concentration profile, which normally ranges from a maximum at the air–soil interface to an immeasurably low value in stratosphere. In particular, radon concentrations may drop by a factor of 10 in the first kilometer. For this reason, it is necessary to specify the altitude when talking of outdoor radon concentrations.

Atmospheric radon concentrations at 1 m from the ground are lower than those in soil by a factor of ~1000. This sharp drop is due to fast mixing in the air and the relatively short half-life of  $^{222}\text{Rn}$ . Isotopes of radon gas and their decay products provide a unique set of tracers for the study of transport and mixing processes in a wide variety of atmospheric phenomena. The  $^{222}\text{Rn}$  isotope has been more widely used because of its longer half-life and the greater relative abundance in the free atmosphere. For near ground level, tracer studies have been based on thoron mainly because of its short half-life (55.6 s). Since the ocean represents a poor radon source compared with the soil and rocks of continental areas, marine air masses contain only 1% or less radon per unit volume than air over large land areas. Collection

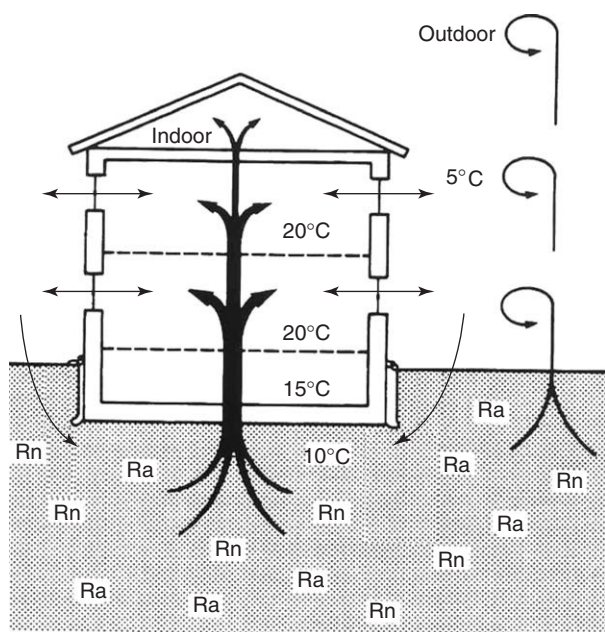
of air samples at sea can be useful in identifying radon-rich air masses that have originated from land areas.

### Indoor Radon

Most of the sources of outdoor atmospheric radon, such as soil, water, and natural gas, are also sources of indoor radon. The amount of air available for radon dilution in a structure is less than that available outdoors. As a consequence, levels of indoor radon are typically much greater than those found outdoors. As soil and groundwater are the major radon sources, large radon concentrations may exist in mines, caves, underground workplaces, and thermal spas. In the case of dwellings, emphasis was first placed on building materials as source of radon. It was not until the 1980s that soil was identified as an important source, especially for single family houses and ground-level workplaces. Subsequent detailed US studies led Nero to attribute virtually all of the radon excess over the outdoor level to the soil contribution. However, these conclusions may well apply to US homes, where the building material is essentially made of wood, but not in those countries where the building material could be even the soil itself.

It is believed that the basic mechanism that brings soil gas into the house is the pressure difference between the indoor and the outdoor environments. This pressure-driven flow (movement from a high- to a low-pressure area) is enhanced in the winter by a 'stack effect', similar to that encountered in the chimney. This chimney effect is created by the continuing rising of heated air as shown in **Figure 2**. In addition to the stack effect, contributions to the house depressurization arise from wind effects, the use of appliances that consume indoor air, etc.

The main characteristic of indoor radon levels, however, is their variability. Besides soil factors, building characteristics such as airtightness, permeability of cellar floors, doors between floors, building materials, microclimate (temperature, wind), water supply, number of occupants, and living habits have all been shown to influence indoor radon concentrations. Different comprehensive surveys of radon in houses have been carried out in several countries. One of the most striking results is the observed large variation in the radon levels from house to house. The value of the radon concentration covers a range from a few up to several thousands of  $\text{Bq m}^{-3}$  and in a few extreme cases, values up to ~200 000  $\text{Bq m}^{-3}$  have been measured. Just close to the house with extremely large radon concentration, it was possible to find other houses with very low radon levels.



**Figure 2** Radon infiltration from the subsoil into the indoor environment driven by the temperature differences (stack effect). (Reproduced with permission from Burkart W (1989) Radiation biology of the lung. *The Science of the Total Environment* 89: 1–230.)

A realistic quantitative prediction of the radon levels in existing houses seems impossible and measurements are finally the only reliable method for the detection of houses with high radon levels. Large-scale surveys of indoor radon have been carried out for different purposes. They can be classified respectively as:

- ‘Screening’ surveys for the investigation of geographical variations of the indoor radon distribution and for the identification of houses with high radon concentrations.
- Statistically representative surveys for the evaluation of the population doses.

Large-scale surveys for the assessment of the population exposure require a multitude of different detectors to be deployed in the primary living spaces (namely bedrooms) for year-long integrated measurements in several thousands dwellings. The most suitable detectors for these large-scale surveys are track-etch detectors. In contrast, short-term measurements are appropriate for screening surveys.

## Health Effects

The available epidemiological data of studies of underground miners exposed to radon decay products provide a direct assessment of lung cancer

mortality. An important aspect of this approach is that the levels of exposure are not a great deal higher than those found in the home, even for average home exposure. The analogy between working age males laboring in mines with families sleeping in their bedrooms may not be perfect. However, even though the risks of exposure to radon and its decay products in dwellings are not well characterized, their potential lung cancer risks need to be strongly addressed, both in homes and occupational environments. The risk of dying from a lifetime exposure to average indoor radon concentration is estimated to be in the range of 0.1–1%, depending on country and smoking habits. These risks are in the same order of magnitude as risk of dying in a fall at home or a home fire. Moreover, houses having several times the average radon concentration have radon risks that can equal or exceed the 2% risk of death in an automobile accident. Since it is not possible to predict where these high values will be found, screening surveys are needed to identify the houses with high indoor radon concentrations.

These data alone do not solve the radon problem. It is important to know that efficient and low-cost techniques for the reduction of the radon inflow from the soil, like subpressurization, are now being developed and tested. At the present stage of knowledge, the most effective remediation approach seems to be that of reversing the driving force behind radon inflow. The International Commission of Radiological Protection has recommended an intervention level of  $400 \text{ Bq m}^{-3}$  for existing dwellings as the concentration above which remedial action should be considered. The most reasonable approach is to apply a step-wise system of action levels for the concentration of radon. According to this approach, the higher the observed concentration, the shorter should be the time period for remedial actions.

## Monitoring Radon and Its Decay Products

There is no radionuclide with characteristics more suitable for easy separation and determination than  $^{222}\text{Rn}$ . A noble gas with no stable compounds, it can be readily separated from the myriad of solid radionuclides and the other radon isotopes disappear rapidly from decay. The radon can be absorbed on charcoal at room temperature or condensed at dry ice temperature. Another useful radon characteristic is the decay to both short-lived and long-lived solid radionuclides. These can either be left with the radon to augment its radioactivity or separated on an air filter or electrostatic collector for separate measurement.

Techniques for measuring the concentration of radon and its decay products have been developed in two large fields of investigations: earth sciences and radioprotection. In spite of such different fields of application, radon monitoring deals with the investigation of similar processes. For example, high indoor radon concentrations are related to gas migration processes that derive from the geological formation, from the soil and from underground water. The same processes are of interest for geological risk prediction (earthquake, vulcanism), mineral prospection (uranium and oil), hydrogeology, geothermy, etc.

The characteristics of the measuring techniques will depend on whether the quantity to be measured is the concentration of radon (be it in air, in water, or in soil) or of the radon decay products. Both these measurements are based on the detection of radiation emitted from radioactive decay in combination with a suitable sampling technique. The whole spectrum of radiation detectors could be used, but most methods rely on detection of  $\alpha$ -particles; some are based on detection of  $\gamma$ -emissions and only a few utilize  $\beta$ -decays.

## Airborne Radon Monitoring

In the case of radon gas measurement, the major problem is the low sensitivity of existing instruments for accurate routine analysis at radon concentrations near the indoor average, which has a worldwide arithmetic mean of  $40 \text{ Bq m}^{-3}$ . At this concentration, in each liter of air, there will be  $\sim 2.2$  decays per minute of  $^{222}\text{Rn}$ . Even with further decays of the radon progenies, this implies low counting rates. Each  $\alpha$ -particle decay of  $^{222}\text{Rn}$  is shortly followed by the  $\alpha$ -particle decays of  $^{218}\text{Po}$  and  $^{214}\text{Po}$ . Thus, there will be about seven  $\alpha$ -particles per minute for each typical value of radon concentration of  $40 \text{ Bq m}^{-3}$ . If these were counted with 100% efficiency, with no contribution from extraneous counts, a counting period of  $\sim 1 \text{ h}$  would be needed to yield 400 events. From simple statistical considerations, this means that the radon concentration is determined to within  $\sim 10\%$ , if the background rates are low compared to the radon counting rates. For most measurements of outdoor levels and for many measurements indoor, the radon concentration is less than  $40 \text{ Bq m}^{-3}$ , background counts cannot be neglected, and the problems of low counting rates are still greater.

Sufficient accuracy for these low-level measurements can be achieved by increasing the volume of air sampled, or by lengthening the sampling or analysis time. In particular, the time duration of the

measurement determines three broad classes of sampling procedures:

- grab sampling method for the measurement of the content of radon of a sample of air at a single instant in time,
- continuous sampling method for the study of the time dependence of the radon concentration, and
- integrating sampling methods for short- and long-term measurements, i.e., determination of radon concentration averaged over a period of a few days to a year.

Another important classification is based on whether the sampling method and/or technique is passive or active. Active instrument technique is used by forced sampling of air by pumps. In passive systems, the sampling of radon is based on natural diffusion without any power supply.

The terms active and passive are also often used to designate radiation detectors that operate with and without power supply, respectively. To avoid confusion, this classification should be made simply by denoting detectors with or without real-time response. In particular, a radon monitor may be formed by any possible combination of active or passive sampling systems and real-time or no real-time detectors. In spite of the many possible combinations of detectors and sampling procedures, some guidelines may be drawn from the experiences accumulated to date. In particular, total passive monitoring systems work well for the measurement of radon gas alone, being the most suitable for large-scale survey of radon concentrations in indoor air. By contrast, the measurements of radon daughters can successfully be carried out by total active monitoring systems, i.e., by using active sampling and real-time  $\alpha$ -detection and spectrometry. This type of monitoring is particularly useful for extensive investigations in a limited number of dwellings. Some of the currently available radon gas monitoring devices are described in the following and listed in Table 4.

### Grab Sampling Method

This method involves filling a container with an air sample of radon and transferring it to a laboratory for analysis. One of the most useful devices is the scintillation flask, known as the Lucas cell. This consists of a chamber whose walls are coated with a scintillating material such as silver-activated zinc sulfide. The cell is optically coupled to a photomultiplier tube to count the light pulses (scintillation) induced by the  $\alpha$ -particles interacting with the zinc sulfide coating. The scintillation flask can be filled either by evacuation or by airflow using a



**Table 4** Currently available radon measurement devices

<i>Instrument</i>	<i>Sampling type</i>	<i>Sensitivity<sup>a</sup> (Bq m<sup>-3</sup>)</i>
Scintillation cell	Grab or continuous	0.37–37
Diffusion electrostatic collection on scintillator	Continuous	<3.7
Ionization chamber	Grab or continuous	0.5–37
Two-filter method	Grab or continuous	0.74–3700
Activated charcoal	Integrating	7.4 (for 100 h exposure)
Track detector	Integrating	3.7 (for 100 days exposure)

<sup>a</sup>Sensitivity is defined as the smallest quantity that can be distinguished from background with a 95% confidence level.

Derived from Corthen CR and Smith JE Jr (1987) *Environmental Radon*. Environmental Science Research, no. 35. New York: Plenum Press.

pump. For accurate measurements, a delay of at least 3 h is used prior to counting to allow the development of a radioactive equilibrium of polonium-214. The efficiency of these cells is typically 70–80%. The lower level of detection (LLD) depends on several parameters such as the size of the flask, counting intervals; typical values are  $\sim 37 \text{ Bq m}^{-3}$  ( $1 \text{ pCi l}^{-1}$ ) or less.

### Continuous Sampling

For continuous sampling, there is no single system that can be favored over the others as in the case of grab sampling. In the following, four types of continuous radon monitoring will be described. In particular, it is important to note that these continuous monitoring methods can be also used for grab monitoring.

**Scintillation chamber monitor** This instrument is made of a scintillation flask in contact with a photomultiplier. Air is filtered to remove the radon daughters and is drawn continuously through the chamber. For continuous measurements of relatively high concentration of radon, such as in soil gas, an open scintillation cell may be used directly. The open end, usually connected to a tube, is inserted directly into the soil or water. The counting rate of the photomultiplier pulses is monitored continuously and changes proportionally to the radon concentration.

**Diffusion electrostatic method** Ambient air enters the sensitive volume of the instrument by molecular diffusion through the cover made of foam rubber. Radon progenies resulting from the decay of radon within the sensitive volume are attracted to a scintillator detector surface by an electric field (electrostatic collection) established by charging a surface near the scintillator. The rest of the electronics is identical to the scintillator cell method. One such instrument has a hemispherical detection chamber with a volume of several liters. The efficiency of the electrostatic collection depends on the humidity. This

problem can be solved by placing a desiccant in the detection chamber.

**Ionization chamber** The first radon measurements were made with various electrometers, which were primitive ion chambers. Most of the improvements in ion chambers over the years have been in the associated devices or electronics. Since the early applications of the ion chambers, the most important problem to be solved was the discrimination of the radon signal from that of the  $\gamma$ -radiation. The earlier system with paired chambers allowed automatic subtraction of background. One chamber, filled with air and sealed, responded to external  $\gamma$ -radiation. This response was subtracted from that of the chamber containing the sample. The present generation of ion chambers is based on fast pulse measurement rather than total ionization. In this case, pulse discrimination eliminates any  $\beta$ - or  $\gamma$ -response.

In this latter method air is introduced to an ionization chamber after being filtered to remove airborne radon progenies. The current in the chamber is proportional to the radon concentration. The lower detection limit depends on the volume of the ionization chamber and can be as low as  $0.5 \text{ Bq m}^{-3}$ . Ionization chambers are very complex laboratory instruments typically used as reference standards for their accuracy. In some laboratories, the ionization chambers are used as a standard for calibrating other instruments.

**Two-filter methods** This method consists of a two-filter tube with a scintillation phototube detector placed in front of the exit filter. The air is drawn continuously through the tube. The first filter removes all the radon decay products from the sampled air. This air is then passed through a long decay tube, where daughter products are allowed to develop and are collected on a second filter, the activity of which is proportional to the radon concentration in the sampled air.

The exit filter can be fixed or can be advanced automatically at preset time intervals. Sensitivity

depends mainly on the volume of the tube. For volumes of the tube between 1 and 1000 l the sensitivity ranges from thousands of  $\text{Bq m}^{-3}$  to  $0.7 \text{ Bq m}^{-3}$ .

### Integrated Sampling Methods

The most widely used integrating radon monitors are based on totally passive devices that present uniquely attractive characteristics for large-scale surveys. Totally passive devices can be successfully obtained for radon-only measurements in which radon diffuses in a detector housing and the radiation from the radon and its progenies is directly registered by passive types of detectors. These devices can be divided into two broad categories, consisting respectively of diffusion and permeation samplers (Figure 3). The upper part of Figure 3 shows the diffusion sampler, which consists of a tube with a detector located at one end of the diffusion zone formed by the tube. The other end of the diffusion tube is generally open to the atmosphere. By using a sufficiently long tube (with a length greater than 30 cm), the thoron decays before it reaches the sensitive volume of the detector.

The most popular geometry for the detection housing has been the cup-type container closed at one end with a porous filter such as fiberglass, a non-wetting cloth, or a microcore paper filter. These porous filters always ensure that aerosol radon daughters will not enter the sensitive volume of the detector, but they may not be efficient discriminators of thoron gas and water vapor.

A relatively more efficient way to eliminate the thoron and water vapor is to use a polymeric membrane (namely a polyethylene film a few tens of

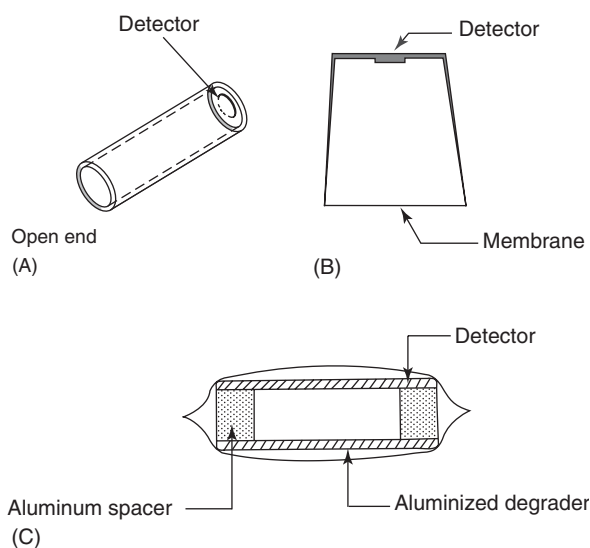
micrometers thick), in which the gas radon must first dissolve and then diffuse. These devices are referred to as permeation samplers (Figure 3). In particular, a permeation sampler can be simply formed by a heat-sealed plastic bag made of polyethylene, as shown in the lower part Figure 3. Enclosed in the bag are two track detectors held 2 cm apart by an aluminum or plastic spacer. The aluminized plastic films (referred to as degrader in Figure 3) are used to optimize the detector responses to the  $\alpha$ -particles from radon and its progeny.

In principle, different types of passive detectors can be used such as thermoluminescent materials, electret devices, and solid-state nuclear track detectors.

**Solid-state nuclear track detectors** For large-scale survey of long-term integrated radon measurements, the track detectors have the most favorable characteristics. For this reason they have been extensively applied in different national surveys for determining the annual average concentration of indoor radon. When compared to thermoluminescent materials and electrets, track detectors register  $\alpha$ -particles with high sensitivity and selectivity, with no response to lightly ionizing radiation ( $\gamma$  and  $\beta$ ). Several detector materials for the registration of  $\alpha$ -particles have been developed. The most suitable are, respectively, poly allyl diglycol-carbonate (known as CR-39, acronym of Columbia Resin 1939), cellulose nitrate (LR-115 from Kodak), bisphenol-A-polycarbonate (Makrofol from Bayer AG).

Alpha particles penetrating the material cause damage channels of near-atomic dimensions. Upon etching, the damage regions are enlarged to the point that they can be easily seen under the microscope. By counting the tracks in a given area, the exposure of the detector to  $\alpha$ -producing nuclides may be determined.

**Electret ionization chambers** A more recent application of the primitive total ionization chambers (such as the electroscopes used, for example, by Rutherford in the early 1900s), is based on the use of an electret, which holds a charge over a long period and is discharged by exposure to radiation. The loss of a charge is then measured by an electrostatic voltmeter and related to radon exposure through a calibration process. The chamber also responds to  $\gamma$ -radiation and the total signal must be corrected for this response. At average radon concentrations and  $\gamma$ -levels, the two signals are approximately equal. When used as a screening technique for short-term measurements, the assessment of the  $\gamma$ -exposure can be estimated with any available instruments, such as Geiger counter.



**Figure 3** Diffusion and permeation radon samplers: (A) diffusion sampler, (B) permeation sampler, and (C) plastic bag permeation sampler.

However, for the long-term measurements required for the assessment of human exposure, paired chambers must be used in order to ensure the accurate discrimination of the radon signal from that of the  $\gamma$ -background.

**Activated charcoals** Passive radon measurements can be achieved without a passive detector by absorbing radon with activated charcoal, which is analyzed by  $\gamma$ -ray spectrometry at the end of the exposure. These charcoal-based radon monitors are sufficiently sensitive for the assessment of short-term integrated (less than 1 week) radon exposures. They have been widely applied for screening surveys. Unfortunately, because of the 3.825 day half-life of  $^{222}\text{Rn}$ , the exposure period cannot be longer than 1 week and the detector must be analyzed in the laboratory soon after the exposure. For the same reason, this technique does not provide a true integration, but a response more closely related to the last day or two of exposure.

### Measurements of Airborne Radon Decay Products

For what concerns the techniques for the measurement of radon decay products, they are similar in principle to those discussed for airborne radon gas. As with the case of radon, the devices can be separated into the categories of those representing grab samples, continuous measurements, and integrations. Radon decay products are simpler to measure by instantaneous or grab sampling while radon gas measurement is easier using continuous or integrated sampling.

Knowledge of the average concentration of each of the radon decay products is important for the evaluation of the potential hazard. The unit 'working level' (WL) was first introduced to measure the concentration of short-lived radon daughters in the air of uranium mines and was also eventually applied for the assessment of the radon daughters in dwellings. One WL can be represented by any combination of short-lived radon progeny in 1 l of air that will result in the emission of  $1.3 \times 10^5$  MeV of potential  $\alpha$ -energy from the radioactive decay. The term potential refers to the fact that a given atom has a 'potential' to produce a certain amount of energy, if it decays through the rest of the decay chain. The WL avoids the complexity of specifying individual concentrations of each decay product and measures the exposure rate to the total  $\alpha$ -energy from  $\alpha$ -particles in 1 l of air. For this reason the measurement of WL has been appealing historically, since it only requires an air sample filtered for 5 min followed by a single  $\alpha$ -count.

The cumulative exposures for miners have been based on the unit 'working level month' (WLM),

which represents the exposure to one 'WL' for one 'WM' (based on 170 working hours). A better way to express the activity of the radon decay products is the potential alpha energy concentration (PAEC), which is the sum of all the potential  $\alpha$ -energies in a volume of air, divided by the volume of that air. A common unit for this quantity is  $\text{J m}^{-3}$ .

A typical evaluation of PAEC requires a filtered air sample and three subsequent  $\alpha$ -counts. The  $\alpha$ -spectrometry can replace the  $\alpha$ -counter in the detection process.

The PAEC can be readily calculated once the activities of the individual radionuclides have been determined from measurements. Direct measurements of the concentrations of all short-lived decay products of  $^{222}\text{Rn}$  are difficult and limited. They are estimated from considerations of equilibrium (or disequilibrium) between  $^{222}\text{Rn}$  and its decay products. An equilibrium factor  $F$  is defined that permits the exposure to be estimated in terms of the PAEC from the measurement of radon gas concentration. This equilibrium factor is defined as the ratio of the actual PAEC to the PAEC that would prevail if all the decay products in each series were in equilibrium with the parent radon. However, it is simpler to evaluate this factor in terms on an equilibrium equivalent radon concentration, EEC. This quantity, EEC, represents the activity concentration of the radon gas that would have to exist in complete equilibrium with the decay products if the short-lived decay products had the same PAEC as in the nonequilibrium mixture. The units of EC are  $\text{Bq m}^{-3}$ .

The equilibrium factor is then given by

$$F = \text{EEC}/C_m$$

where  $C_m$  is the measured concentration of the radon gas.

It is important to bear in mind that the EEC depends only on the concentrations of the decay products (as only these are considered to be radiobiologically important) and is given by

$$\text{EEC} = 0.105C_1 + 0.515C_2 + 0.380C_3$$

where  $C_1$ ,  $C_2$ , and  $C_3$ , are the activity concentrations of the decay products, namely  $^{218}\text{Po}$ ,  $^{214}\text{Po}$ , and  $^{214}\text{Bi}$ , respectively. The constants are the fractional contributions of each decay product to the total potential  $\alpha$ -energy from the decay of unit activity of the gas. In this way, a measured radon concentration can be converted to an EEC directly proportional to PAEC. This provides a measure of exposure in terms of the product of concentration and time. The EEC can be converted to the PAEC, when desired, by the

relationship:

$$1 \text{ Bq m}^{-3} = 5.56 \times 10^{-6} \text{ mJ m}^{-3} = 0.27 \text{ mWL}$$

Many measurements have been made of  $^{222}\text{Rn}$  and decay product concentrations, allowing estimates to be made of the magnitude of the equilibrium factor to be estimated in terms of both typical values and range. In general, it is unusual to find equilibrium factors less than 0.2, even in well-ventilated rooms or mines. For this reason, it is common practice to measure the concentration of radon gas, which is then used to infer the concentration of its decay products by using a known value for the equilibrium factor. This is highest for outdoor air, followed by those of indoor air and air in mines.

## Radon Measurement in Water, Soil, and Other Solid Matrices

Radon measurements in water, soil, and other solid materials are of interest for geoscientific or geoengineering purposes. Using  $\gamma$ -ray spectrometry it is possible to determine the radon concentration in samples of water, soil, and other solid materials by counting  $\gamma$ -rays from radon decay products, with particular regard to 1.76 MeV  $\gamma$ -rays from bismuth-214. This method, in addition to its low sensitivity, has the disadvantage of requiring two measurements 4 weeks apart. This is because the original radon concentration can be distinguished from the  $^{226}\text{Ra}$  concentration by repeating the analysis after at least 4 weeks. At this later time the original radon will have virtually all decayed and the only remaining radon is that in secular equilibrium with  $^{226}\text{Ra}$ . In other situations where the uranium mineral is resistant to weathering, the radium will tend to be in secular equilibrium with the uranium. In such minerals the radon loss is generally low and  $\gamma$ -ray spectrometry measurements give a good indication of both the radium and the uranium content.

The most common methods for measuring radon in soil and water are based on measurement of the radon gas after it has been separated from the host environmental sample by extraction and/or diffusion processes. Extraction of the radon sample is achieved by pumping gas from the soil using a rigid hollow tube. This is hammered into the ground to a convenient depth, causing minimum disturbance of the soil profile. The detection of radon is obtained by using the scintillation flask (Lucas cell) in a way similar to the grab sampling method. Radon measurements in soil can also be obtained by simply placing the detector in the ground at a depth of 0.5–0.7 m.

These measurements of soil radon by detectors in the ground are of interest when long-term

monitoring is employed to overcome the problem of temporal variation in the concentration of radon. Different detectors can be used, the choice of which depends on the particular application. Real-time detectors such as zinc sulfide scintillators or semiconductors are used when the time dependence of the radon concentration in soil is needed for a limited number of places. For a large-scale survey, the most popular radon monitoring devices are based on the track detector enclosed in the cup-type geometry, as described above.

The measurement of the radon content of water is based on extraction processes that exploit the high partition coefficient of radon either between gas and water or between organic liquid scintillators and water. In particular, by introducing fine gas bubbles to water, it is possible to extract radon very efficiently. Radon is thus bubbled out from water and collected in a Lucas cell. The detection limit for this method is very low,  $\sim 50 \text{ Bq m}^{-3}$ .

An alternative method requires the use of immiscible liquid scintillators. In this method, 5–10 ml of toluene scintillator is added to 10 ml of water in a glass vial and shaken vigorously. Because of the very large partition coefficient of radon between water and organic liquid scintillators, radon is effectively extracted from water and detected by a liquid scintillator counter. The detection limit is  $\sim 370 \text{ Bq m}^{-3}$  ( $10 \text{ pCi l}^{-1}$ ) for a 40 min counting time. This method makes it possible to obtain large-scale counting with automation as required for determining radon in a large variety of water samples. As for the measurement in soil, radon can be monitored under water simply by placing the detector directly in the water. This system has been applied to the monitoring of radon in the ground below the water table.

## Radon Flux Measurements

Radon flux measurements are important for the assessment of the transfer of radon across the interface between a solid phase and the atmosphere. The solid may be the soil, rocks, building materials, or other substances, while the atmosphere may be that of a closed laboratory vessel, a building, or the outdoor air. A variety of methods are available for measuring radon flux in the laboratory and in the field. These methods involve the use of various techniques described above for the assessment of the radon concentration.

### Laboratory Methods

Methods for laboratory measurements of soils and building materials are generally based on placing a sample of the material in a closed chamber and sampling the air of the chamber.

## Field Methods

Among the different methods available, the most applicable for the field measurements are the accumulation method and the adsorption method.

The accumulation method was one of the earliest procedures and consists, basically, of placing the open end of a vessel on the surface being measured. This method requires the use of a suitable sized container, ranging in volume from a few liters up to as much as 220 l. The radon concentration inside the vessel is then measured, either at some selected time or serially over a period of several hours. The requirements for the accumulation method are that the accumulation time is short compared with the 3.82 day half-life of  $^{222}\text{Rn}$ , that the concentration in the vessel is lower (perhaps 10% or less) than that in the solid material, and that the measuring device does not significantly affect the radon exhalation.

The adsorption method involves the use of an activated charcoal canister placed in close proximity to the solid surface. The canister procedure is simple, and the sampler can be readily sealed to vertical surfaces with modeling clay or flexible caulking materials. However, the small sampling area of the canister results in poor sensitivity. The LLD is  $\sim 2.5 \text{ Bq}$ , which amounts to a flux of  $\sim 4 \text{ mBq m}^{-1} \text{ s}^{-1}$  for a 1 day exposure, while the LLD of the accumulation method can easily be 10 times smaller.

**See also:** **Air Analysis:** Sampling. **Geochemistry:** Sediment; Soil, Major Inorganic Components; Soil, Minor Inorganic Components; Soil, Organic Components. **Radiochemical Methods:** Natural and Artificial Radioactivity; Radionuclide Monitoring; Uranium; Radiotracers; Gamma-Ray Spectrometry. **Water Analysis:** Freshwater.

## Further Reading

- Biancotto R, Marinaro M, and Tommasino L (1998) Natural and enhanced environmental radon. *Special issue. Radiation Protection Dosimetry* 78: 1–82.
- Campos-Venuti G, Janssens A, Olast M, *et al.* (1993) Indoor radon remedial action – the scientific basis and practical implications. Special issue. *Radiation Protection Dosimetry* 56: 1–387.
- Durrani SA and Ilic R (1997) *Radon Measurements by Etched Track Detectors – Applications in Radiation Protection, Earth Sciences, and the Environment*. Singapore: World Scientific.
- Furlan G and Tommasino L (1993) *Radon Monitoring in Radioprotection, Environmental and/or Earth Sciences*. Singapore: World Scientific.
- George AC (1996) State-of-the-art instruments for measuring radon/thoron and their progeny in dwellings – a review. *Health Physics* 70: 451–463.
- Griffith RV and Tommasino L (1991) Etch track detectors in radiation dosimetry. In: Kase K, Bjarnd BB, and Attix

FH (eds.) *Dosimetry of Ionising Radiation*, vol. III, pp. 326–426. New York: Academic Press.

Harley JH (1992) Measurements of  $^{222}\text{Rn}$ : A brief history. *Radiation Protection Dosimetry* 45: 13–18.

Nero A (1989) Earth, air, radon and home. *Physics Today* 42: 32–39.

Tommasino L (2001) Personal dosimetry and area monitoring for neutrons and radon in workplaces. *Radiation Measurements* 34: 449–456.

Tommasino L, Furlan G, Khan HA, and Monnin M (1990) *Radon Monitoring in Radioprotection, Environmental Radioactivity and Earth Science*. Singapore: World Scientific.

## Glossary

Activity	The rate of atomic disintegration (units: Becquerel or Curie).
Becquerel (Bq)	A unit used to describe the rate of radioactive decay. One Becquerel equals one disintegration per second.
Curie (Ci)	An old unit used to describe the rate of radioactive decay. One Curie equals $3.7 \times 10^{10}$ disintegrations per second.
Decay product	A nuclide resulting from the radioactive disintegration of a radionuclide, formed either directly or as the result of successive transformations in a radioactive series. A decay product may be either radioactive or stable.
Decay series	The consecutive members of a radioactive family of elements. A complete series commences with a long-lived parent such as $^{238}\text{U}$ and ends with a stable element such as $^{206}\text{Pb}$ .
Equilibrium-equivalent concentration (EEC)	The EEC of a nonequilibrium mixture of short-lived decay products in air is that activity concentration of the parent gas in radioactive equilibrium with the concentrations of its short-lived decay products that have the same PAEC as the nonequilibrium mixture. Units are $\text{Bq m}^{-3}$ .
Equilibrium factor	An adjustment used in converting from $\text{Bq m}^{-3}$ ( $\text{pCi l}^{-1}$ ) to working level concentration, which takes into account the possible absence of radioactive equilibrium between radon and its decay products.
Radioactive equilibrium	A state in which the rate of formation of atoms is equal to the rate of their disintegration by radioactive decay, so that the amount of the element or isotope is constant.
Flux measurements	Measurements made to determine how much radon is released into an enclosure

	from a source. They are used to determine the rate at which radon emanates from building materials.		
Half-life	The time it takes for one-half of any quantity of identical radioactive atoms to undergo decay.	Working level (WL)	A measure of decay product concentration that indicates the extent of $\alpha$ -particle release from short-lived products. Also a measure of the total $\alpha$ -activity in air.
Potential alpha energy (PAE)	The total energy emitted during $\alpha$ -decays if a quantity of radon decay		

## Uranium

**M Keith-Roach and E P Achterberg**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Uranium is a radioactive element that is ubiquitous in the environment, with typical crustal and seawater concentrations of  $2.7 \text{ mg kg}^{-1}$  and  $3.0 \mu\text{g l}^{-1}$ , respectively. Uranium-238 and  $^{235}\text{U}$  are primordial isotopes and their decay leads to the formation of series of daughter radioisotopes, amongst which is  $^{234}\text{U}$  (Figure 1). The natural isotopic ratio of  $^{238}\text{U}$ : $^{235}\text{U}$ : $^{234}\text{U}$  is 99.2745:0.7200:0.0055, and other isotopes (Table 1) are produced by nuclear processes.

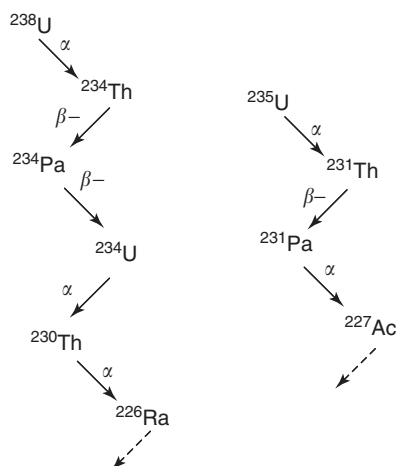
Uranium displays multiple oxidation states from +3 to +6, with +4 and +6 being the most common. Uranium(IV) is largely insoluble, as it reacts readily with particle surfaces. Uranium(VI), on the other hand, exists as a di-oxo cation,  $\text{UO}_2^{2+}$  in

solution, which forms stable complexes with oxygen-containing ligands, such as carbonate. This is seen in the high solubility of  $\text{UO}_2^{2+}$  as anionic carbonate complexes in seawater (e.g.,  $\text{UO}_2(\text{CO}_3)_2^{2-}$ ).

The main use of uranium is as nuclear fuel, with a lesser role in nuclear weapons, and its by-product, depleted uranium, is used in conventional weapon heads. Uranium oxides containing  $^{235}\text{U}$  enriched to  $\sim 4\%$  are the most common type of nuclear fuel, although metallic uranium with the natural isotopic ratio of uranium is also used in older technologies (e.g., Magnox). The fuel is bombarded with neutrons to induce fission but, as this process is not 100% efficient and high-energy neutrons result in neutron capture,  $^{236}\text{U}$  is also produced. The production and use of nuclear fuel results in significant changes to the natural isotope ratio, and the isotope ratio is therefore an important tool in identifying anthropogenic enhancement of uranium in the environment over variable natural concentrations. Isotope ratio measurements of natural uranium can also be used to examine environmental processes, for example, by using solubility differences between uranium and thorium, which is a precursor of  $^{234}\text{U}$  (Figure 1).

The study of uranium covers a wide range of interests, from natural systems with low anthropogenic inputs, to mining spoil heaps and contaminated environments, to quality control of fuel production and waste management. Analytical requirements are therefore highly varied with the need to measure uranium in different phases, from ultra-trace to high concentrations, from total uranium concentrations to high precision isotope ratios. The cost of analysis and availability of instrumentation must also be taken into consideration when selecting the appropriate analytical technique.

A range of sample types require radiochemical separation of uranium from the sample matrix and removal of elements that interfere with accurate



**Figure 1** Early stages of the uranium decay series.



	from a source. They are used to determine the rate at which radon emanates from building materials.		
Half-life	The time it takes for one-half of any quantity of identical radioactive atoms to undergo decay.	Working level (WL)	A measure of decay product concentration that indicates the extent of $\alpha$ -particle release from short-lived products. Also a measure of the total $\alpha$ -activity in air.
Potential alpha energy (PAE)	The total energy emitted during $\alpha$ -decays if a quantity of radon decay		

## Uranium

**M Keith-Roach and E P Achterberg**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Uranium is a radioactive element that is ubiquitous in the environment, with typical crustal and seawater concentrations of  $2.7 \text{ mg kg}^{-1}$  and  $3.0 \mu\text{g l}^{-1}$ , respectively. Uranium-238 and  $^{235}\text{U}$  are primordial isotopes and their decay leads to the formation of series of daughter radioisotopes, amongst which is  $^{234}\text{U}$  (Figure 1). The natural isotopic ratio of  $^{238}\text{U}$ : $^{235}\text{U}$ : $^{234}\text{U}$  is 99.2745:0.7200:0.0055, and other isotopes (Table 1) are produced by nuclear processes.

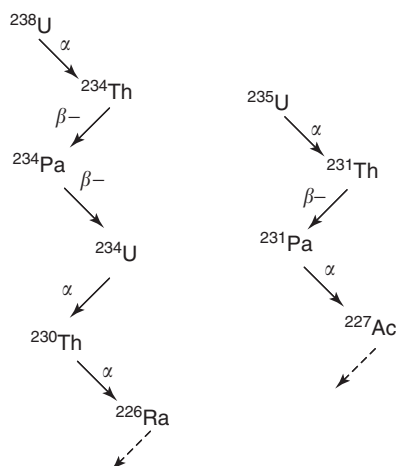
Uranium displays multiple oxidation states from +3 to +6, with +4 and +6 being the most common. Uranium(IV) is largely insoluble, as it reacts readily with particle surfaces. Uranium(VI), on the other hand, exists as a di-oxo cation,  $\text{UO}_2^{2+}$  in

solution, which forms stable complexes with oxygen-containing ligands, such as carbonate. This is seen in the high solubility of  $\text{UO}_2^{2+}$  as anionic carbonate complexes in seawater (e.g.,  $\text{UO}_2(\text{CO}_3)_2^{2-}$ ).

The main use of uranium is as nuclear fuel, with a lesser role in nuclear weapons, and its by-product, depleted uranium, is used in conventional weapon heads. Uranium oxides containing  $^{235}\text{U}$  enriched to  $\sim 4\%$  are the most common type of nuclear fuel, although metallic uranium with the natural isotopic ratio of uranium is also used in older technologies (e.g., Magnox). The fuel is bombarded with neutrons to induce fission but, as this process is not 100% efficient and high-energy neutrons result in neutron capture,  $^{236}\text{U}$  is also produced. The production and use of nuclear fuel results in significant changes to the natural isotope ratio, and the isotope ratio is therefore an important tool in identifying anthropogenic enhancement of uranium in the environment over variable natural concentrations. Isotope ratio measurements of natural uranium can also be used to examine environmental processes, for example, by using solubility differences between uranium and thorium, which is a precursor of  $^{234}\text{U}$  (Figure 1).

The study of uranium covers a wide range of interests, from natural systems with low anthropogenic inputs, to mining spoil heaps and contaminated environments, to quality control of fuel production and waste management. Analytical requirements are therefore highly varied with the need to measure uranium in different phases, from ultra-trace to high concentrations, from total uranium concentrations to high precision isotope ratios. The cost of analysis and availability of instrumentation must also be taken into consideration when selecting the appropriate analytical technique.

A range of sample types require radiochemical separation of uranium from the sample matrix and removal of elements that interfere with accurate



**Figure 1** Early stages of the uranium decay series.

measurement. A number of radiochemical separation techniques exist, notably co-precipitation, ion exchange, and solvent extraction. The application of measurement techniques to uranium will be described after a discussion of the sample preparation and radiochemical methods.

## Sample Digestion

Sample digestion is used to obtain a representative, homogeneous solution, to ensure that the uranium is present in solution as an inorganic complex or ion for the subsequent radiochemical separations, and to allow equilibration with the yield monitor (see below). Thus, liquids, as well as solid matrix samples, are often digested prior to analysis or radiochemical separation. The most common methods are acid digestion and fluxed decomposition. Acid digestion of dried soils, sediments, and foodstuffs often follows dry ashing of the organic matter at 450–550°C. The aggressiveness of the acid employed depends on the sample, and, for example, HF is required to digest silicate minerals while mixtures of HNO<sub>3</sub>, HCl, and H<sub>2</sub>SO<sub>4</sub> are suitable for many matrices. Use of fluoride can result in the formation of insoluble uranium fluorides and therefore fuming perchloric acid or boric acid should be added to prevent this.

In fluxed decomposition, or high-temperature fusion reactions, an accurate mass of sample is placed in a crucible with the flux material. Sodium hydroxide, sodium peroxide, sodium carbonate, and lithium tetraborate have been used for this purpose. The mixture is heated in a furnace or Meker burner until it forms a well-mixed, molten fusion. After cooling, this is dissolved in a dilute mineral acid. This method is arguably the best method for complete digestion, but other methods are used routinely and widely. A negative aspect of fusion is that it results in solutions with high total dissolved solids and potentially leads to contamination by the reagents. Solution-phase samples, such as natural waters, are digested using similar procedures, for example, nitric acid or aqua regia digestions may be used, but HF or fusion would give more quantitative dissolutions if small, recalcitrant uranium-rich particles were present.

The extent to which a digestion procedure brings the uranium into solution cannot be assessed through spiked experiments or addition of a yield monitor, as the spike is not representative of the actual sample. However, there are standard procedures published by the American Society for Testing and Materials (ASTM) and the Environmental Protection Agency (EPA) of the USA. Certified reference materials are also available from the International Atomic Energy Agency/NIST.

## Preconcentration and Separation

Analytical radiochemical procedures routinely use combinations of co-precipitation, solvent extraction, and ion exchange to preconcentrate and isolate uranium from bulk stable elements and potential interferences (e.g., for alpha and mass spectrometric techniques, see Table 1).

### Co-Precipitation

Iron(III) co-precipitation is an effective method for preconcentrating uranium from large-volume samples. An additional benefit is the removal of much of the bulk matrix, for example, salt in seawater. Iron(III) is added in milligram quantities and the pH is raised to ~9 using ammonia solution or sodium hydroxide. The hydrated iron(III) oxide precipitate is separated from the bulk solution by centrifugation or filtration and then dissolved in HCl.

**Table 1** Isotopes, half-lives, and main alpha emission energies of uranium and potential interferences in mass and alpha spectrometric techniques

<i>Element</i>	<i>Isotope</i>	<i>t<sub>1/2</sub> (years)</i>	<i>Main alpha energies/MeV<sup>a</sup></i>
Uranium	<sup>239</sup> U	$4.47 \times 10^{-5}$	Beta
	<sup>238</sup> U	$4.47 \times 10^9$	4.15 (0.23)
			4.20 (0.77)
	<sup>236</sup> U	$2.34 \times 10^7$	4.45 (0.22)
			4.50 (0.78)
	<sup>235</sup> U	$7.04 \times 10^8$	4.37 (0.16)
			4.40 (0.58)
	<sup>234</sup> U	$2.46 \times 10^5$	4.77 (0.71)
			4.72 (0.28)
	<sup>233</sup> U	$1.59 \times 10^5$	4.78 (0.15)
Plutonium			4.82 (0.83)
	<sup>232</sup> U	69.8	5.26 (0.31)
			5.32 (0.69)
	<sup>238</sup> Pu	87.7	5.46 (0.29)
			5.50 (0.71)
Neptunium	<sup>239</sup> Pu	$2.41 \times 10^4$	5.14 (0.15)
			5.16 (0.73)
	<sup>240</sup> Pu	$6.56 \times 10^3$	5.12 (0.27)
			5.17 (0.73)
Thorium	<sup>237</sup> Np	$2.14 \times 10^6$	4.77 (0.25)
			4.78 (0.48)
Radium	<sup>232</sup> Th	$1.41 \times 10^{10}$	3.95 (0.23)
			4.01 (0.77)
	<sup>230</sup> Th	$7.54 \times 10^4$	4.62 (0.23)
			4.69 (0.76)
	<sup>229</sup> Th	$7.34 \times 10^3$	4.85 (0.56)
Polonium			4.90 (0.10)
	<sup>228</sup> Th	1.9	5.34 (0.28)
			5.42 (0.72)
Radium	<sup>226</sup> Ra	$1.6 \times 10^3$	4.60 (0.06)
			4.78 (0.94)
Polonium	<sup>210</sup> Po	0.38	5.30 (1.0)

<sup>a</sup>Values in parentheses correspond to emission probabilities.

Other types of co-precipitation are used in the preparation of samples for analysis. For example, cerium(IV) hydroxide, cerium(III) fluoride, and neodymium fluoride can be used to prepare sources for alpha spectrometry (see below). LaF<sub>3</sub> co-precipitation is used in the preparation of samples for uranium measurement by neutron activation analysis (see below).

### Solvent Extraction

Solvent extraction is used on both analytical and industrial scales and utilizes the high solubility of uranium in many organic solvents, in contrast to the low solubility of elements such as iron, thorium, radium, and rare earth elements. Uranium must be present in an ionic form in an aqueous solvent (see sample digestion) as solvent extraction will not extract colloids or complexes. Separation funnels or counter-flow systems can be used, with a higher efficiency and throughput possible using the counter-flow approach.

Tertiary amines can be used to extract uranium from chloride or sulfate media, e.g.,



This process coextracts molybdenum, vanadium, and zirconium, but uranium can be separated from these elements by using a selective back-extraction process into the aqueous phase. For example, sodium chloride will strip uranium, while carbonate will remove uranium and molybdenum. This is relevant if, for example, the uranium will be quantified by titration (see below).

Tributyl phosphate (TBP) in paraffinic diluent can be used to isolate uranium and plutonium from a nitric acid ( $\sim 6 \text{ mol l}^{-1}$ ) medium, with uranium extracted as a neutral uranium nitrate complex:



Back-extraction can be achieved by washing with dilute nitric acid.

### Ion Exchange

Ion exchange is a method of separating complexes or ions in solution as a function of their differential attraction to a charged stationary phase. The stationary phase is held in a column and the mobile phase is changed to manipulate the speciation of the elements of interest. Anion exchange is a convenient method for removing bulk and trace interferences from the actinides, partly because of their complex solution chemistry.

Ion exchange resins are named according to the manufacturer, the purity, the resin type, and the cross-linkage of the resin beads. So, for example,

BioRad AG1-X8, refers to Bio-Rad's version of analytical grade type '1' resin with a cross-linkage of 8%. Type 1 resin has the functional group  $\text{R-CH}_2\text{N}^+(\text{CH}_3)_3$ , and is commonly used for uranium separation, as it has a high specificity for anionic chloride, sulfate, and oxalate complexes of uranyl over those of other metal cations. Cation exchange resins, on the other hand, are not used for uranium separations because of their low specificity. Using type 1 resin as an example, a high chloride concentration is required for the formation of negative uranyl chloro complexes, e.g.,  $[\text{UO}_2\text{Cl}_4]^{2-}$ . Therefore, uranium will not adhere to a type 1 resin in  $1 \text{ mol l}^{-1}$  HCl, but has a high affinity for the resin in  $9 \text{ mol l}^{-1}$  HCl. Neptunium(IV) also adheres to the resin in  $9 \text{ mol l}^{-1}$  HCl, but americium, plutonium, radium, and thorium pass through. Washing the column with  $8 \text{ mol l}^{-1}$  HNO<sub>3</sub> elutes uranium as  $[\text{UO}_2(\text{NO}_3)_2]$ , while retaining neptunium on the column. Other possible complexants include oxalate ( $0.5 \text{ mol l}^{-1}$ ), which causes the formation of anionic uranyl complexes that are retained on the resin, and a mixture of  $6 \text{ mol l}^{-1}$  HCl and  $1 \text{ mol l}^{-1}$  HClO<sub>4</sub>, which elutes uranium. The potential analytical interferences can therefore be removed by the differences in speciation in different solutions, and their affinity for the anion exchange resin.

### Extraction Chromatographic Resins

Specific extraction chromatographic resins have been developed for the isolation of actinides. They are used in a manner similar to ion exchange resins, but function as immobilized solvent phases, and thus work in the same way as solvent extraction. These resins have high specificities for the analytes of interest and are directed at solving problems associated with radio-nuclide separations. For example, U/TEVA (Eichrom Industries) is diamylphosphonate sorbed on an inert polymeric substrate, and has been developed for the separation of uranium from other actinides. If an oxidized solution is applied to a U/TEVA column in  $3 \text{ mol l}^{-1}$  HNO<sub>3</sub>, U, Np, and Th adhere to the column, while Pu and Am pass through. Th, Np, and Fe(III) can be completely removed by converting the column to a chloride form with  $9 \text{ mol l}^{-1}$  HCl and then washing with  $5 \text{ mol l}^{-1}$  HCl– $0.05 \text{ mol l}^{-1}$  oxalic acid. A clean solution of uranium can then be obtained by eluting with  $0.01 \text{ mol l}^{-1}$  HCl. Other specialist resins can also be used, for example, TRU.Spec resin (Eichrom Industries).

### Yield Monitors

A yield monitor is an isotope that is added to a sample prior to radiochemical separation and analyzed

postseparation to quantify the yield of the process. Most radiochemical separation techniques have variable recoveries that are influenced by the sample matrix, oxidation state control, and exact chemical and physical conditions. Therefore, a yield monitor is required for quantitation. The most reliable type of yield monitor is an isotope of the analyte that is not present in the sample of interest, because it will behave in exactly the same way as the analyte throughout the procedure. The addition of a yield monitor also compensates for experimental errors such as spills, boiling over, or losses to the beaker when heating to dryness. Yield monitors can be added either prior to or after digestion, since they cannot monitor the effectiveness of a digestion procedure. Dilution of a sample leachate may also be required prior to addition of a yield monitor.

The yield monitor must be measurable without spectral overlap with the uranium in the sample, and so artificial isotopes are used. These are prepared with minimal contamination of natural or environmentally relevant uranium isotopes. Uranium-232 is a suitable candidate for radiochemical preparation for analysis by alpha spectrometry as it has a suitable half-life and an alpha energy that does not interfere with the peaks from the other isotopes. Uranium-233 is preferred for mass spectrometric techniques because it has a long half-life and has few spectral interferences, whereas  $^{232}\text{U}$  is isobaric with  $^{232}\text{Th}$ . Both these yield monitors also function as internal standards for the measurement technique, which lowers the overall uncertainty associated with yield determination and measurement.

## Measurement Techniques

There are a wide range of techniques available for the measurement of uranium. The detection limits of a selection of widely used techniques are summarized in Table 2.

**Table 2** Summary of sample form and approximate detection limits of measurement techniques

<i>Method</i>	<i>Sample form</i>	<i>Approximate limit of detection</i>
Titration	Solution	$0.5\text{ g l}^{-1}$
Spectrophotometry	Solution	$10^{-6}\text{--}10^{-3}\text{ g l}^{-1}$
XRF	Solid	$10^{-4}\text{ g}$
Gamma analysis	Solid or solution	$10^{-5}\text{ g g}^{-1}$
LSC	Solution	From $10^{-8}\text{ g l}^{-1}$
NAA	Solid	From $10^{-12}\text{ g g}^{-1}$
Alpha spectrometry	Deposited on a planchet	$10^{-8}\text{--}10^{-7}\text{ g}$
ICP-MS	Solution	$10^{-15}\text{--}10^{-10}\text{ g}$
AMS	Prepared in matrix	$10^{-17}\text{ g}$

## Total Uranium Analysis

Total uranium can be analyzed by redox titration against acidified potassium permanganate, cerium(IV), or potassium dichromate. Uranium is reduced prior to the titration with zinc amalgam. Dichromate titrations are used to determine uranium concentrations in dissolved uranium metal, uranium ore, or yellow cake ( $\text{UO}_3$ ) samples. There are potential interferences from bismuth, manganese, platinum, silver, vanadium, and zirconium, which must therefore be removed prior to the titration.

Spectrophotometry also provides rapid, cost-effective measurement of total uranium concentrations, at levels down to tens of  $\mu\text{g l}^{-1}$ . In this method, a chelating agent is added to the digested solution to form colored uranium complexes. The most common chelator is 2-(5-bromo-2-pyridylazo)-5-diethylamino-phenol (bromo-PADAP), which may require removal of chromium, vanadium, or phosphate interferences by solvent extraction, prior to analysis.

Various forms of atomic absorption and atomic emission spectrometry are also able to perform rapid analyses of solutions. A high temperature is required to atomize uranium efficiently; therefore, Graphite furnace AAS and ICP-AES have low detection limits of  $20\text{--}30\text{ }\mu\text{g l}^{-1}$ , while Flame AAS has a detection limit of  $\sim 40\text{ mg l}^{-1}$ .

Uranium fluoresces and therefore can be analyzed by XRF. This allows direct analysis of uranium in solid matrices and overcomes the challenges of digestion and separation.

## Radiometric Techniques

Trace analysis of uranium can be carried out using a variety of radiometric techniques. The least sensitive method is gamma analysis of short-lived daughter products, for example, the  $^{234}\text{Th}$  daughter ( $t_{1/2} = 24.1$  days) of  $^{238}\text{U}$ . Short-lived daughter isotopes are in secular equilibrium with their parent isotope, unless they have been chemically separated, and the activity of  $^{234}\text{Th}$  is therefore exactly the same as that of the long-lived parent.

Lower detection limits are achieved using liquid scintillation counting (LSC) with  $\alpha$ - $\beta$  discrimination. The most sensitive system is the photon electron rejecting alpha liquid scintillation spectrometry, which can measure concentrations as low as  $1\text{ mBq l}^{-1}$ . However, other LSC systems have detection limits an order of magnitude or so higher. LSC requires thorough radiochemical separation to remove interfering isotopes, and the interferences are the same as for alpha spectrometry.

Neutron activation analysis is a sensitive technique for uranium analysis, but its use is limited by access

to a research reactor. During neutron bombardment,  $^{238}\text{U}$  is activated to  $^{239}\text{U}$ , which rapidly decays to  $^{239}\text{Np}$  via beta decay. Neptunium-239 has a half-life of 2.36 days, and can be measured by gamma analysis after removal from the reactor. Ultrasensitive delayed neutron counting is also possible with uranium, and this uses a pneumatic transfer system to carry the sample from the neutron flux to a neutron detector for counting after a short decay period (e.g., 20 s). Neutron activation analysis often requires prior radiochemical separation and the techniques described earlier, such as solvent extraction, ion exchange, and co-precipitation, are used typically for this purpose.

Alpha spectrometry has been used widely to measure uranium and is still used to a significant extent today. Direct counting of alpha emissions is achieved using surface-barrier semiconductor detectors. In effect, the detector is able to record the pulse of energy deposited as the alpha particle collides with it, and convert the pulse amplitude to an energy. Alpha particles are emitted with characteristic energies, allowing isotopes to be differentiated. Given that alpha particles have a high linear energy transfer, the uranium should optimally be prepared as an atom thin layer on a planchet, to prevent self-absorption of radiation. The source is then positioned close to the detector under high vacuum, to prevent energy spread of a given type of alpha emission.

With some planchet preparation techniques, peak resolution of 0.004–0.005 MeV can be achieved, and the common method of electrodeposition can allow resolution of  $\sim 0.02$  MeV. Resolution decreases as the peaks increase in size, and this has implications when the specific activity of one isotope is much lower than another. The activity of an isotope is a function of its half-life and percentage abundance. Therefore, the natural activity ratio of  $^{238}\text{U}$ : $^{235}\text{U}$ : $^{234}\text{U}$  is 0.483:0.0225:0.495. The alpha energies of  $^{238}\text{U}$ ,  $^{235}\text{U}$ , and  $^{234}\text{U}$  (Table 1) are resolvable, but it should be noted that when the count times are long enough to obtain a reasonable peak for  $^{235}\text{U}$ , the  $^{234}\text{U}$  and  $^{238}\text{U}$  peaks will be large and will typically have a base width of  $\sim 0.1$ – $0.2$  MeV. This hampers  $^{236}\text{U}$  analysis because of the relatively low activity concentrations of  $^{236}\text{U}$ , and the energies of its principle alpha emissions.

Typical efficiencies of surface barrier semiconductor detectors are 25–40%, and the backgrounds of uncontaminated detectors are close to zero, with a count recorded every few days. This makes it a sensitive technique for most alpha-emitting radionuclides. However, the low specific activity of uranium means that count times can be long (many days or weeks) for low-level samples.

The need to prepare very thin sample sources for counting necessitates a lengthy radiochemical separation procedure to remove bulk elements efficiently, as well as alpha emitters that will interfere with the spectra (Table 1). Radiochemical procedures and sample mounting onto a planchet are variable in efficiency and can be affected by sample-related factors. It is therefore imperative to use a yield monitor for the procedure if quantitation is required. After radiochemical separation the solution is prepared for alpha counting using electrodeposition, co-precipitation and filtration as a thin source, direct evaporation, electrospraying, or vacuum sublimation. If the planchet is contaminated with polonium, the planchet is heated to remove the volatile polonium.

### Mass Spectrometric Techniques

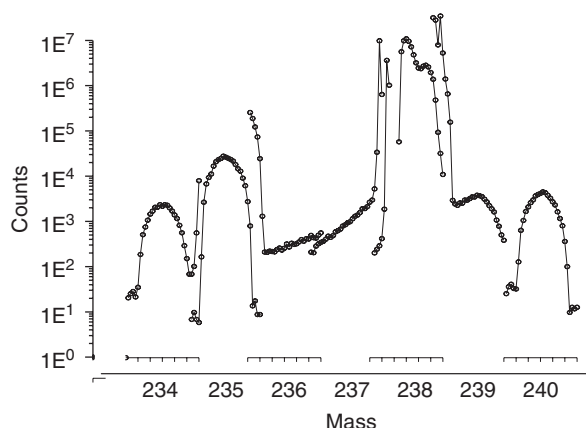
Atom counting techniques are highly sensitive methods for measuring the long-lived uranium isotopes  $^{238}\text{U}$ ,  $^{236}\text{U}$ ,  $^{235}\text{U}$ , and  $^{234}\text{U}$ . It is generally accepted that mass spectrometry is competitive with alpha counting for isotopes with half-lives greater than  $10^3$ – $10^4$  years, given the relationship

$$A = N\lambda$$

where  $A$  is the activity (Bq),  $N$  the number of atoms, and  $\lambda = \ln 2/t_{1/2}$ , with the half-life given in seconds.

However, there is a wide range of mass spectrometric instrumentation and these instruments have detection limits that range over many orders of magnitude. Mass analysis also allows the measurement of isotope ratios. Additionally, isobaric interferences associated with mass spectrometry are typically reduced at the high mass range, making uranium determination possible with limited sample preparation, at least for samples that do not require preconcentration. These factors alongside the rapid sample processing, and high sensitivity of some of these techniques make mass spectrometry an attractive measurement technique for uranium.

Inductively coupled plasma mass spectrometry (ICP-MS), laser ablation ICP-MS (LA ICP-MS), thermal ionization mass spectrometry (TIMS), secondary ion mass spectrometry (SIMS), glow discharge mass spectrometry (GDMS), resonance ionization mass spectrometry (RIMS), and accelerator mass spectrometry (AMS) have been used successfully to measure uranium concentrations and isotope ratios in a wide range of sample matrices. The specific details of the methods are described fully in the relevant sections of this encyclopedia. There are specific advantages associated with each method, which depend on the sample of interest and the information required.



**Figure 2** Mass spectrum of uranium extracted from a soil sample. Each mass was counted for a different time. Each peak was counted over a mass range greater than the peak; therefore, the edge of the neighboring peaks can be seen in the individual mass scans.

Briefly, LA ICP-MS, SIMS, and GDMS are suitable for the direct analysis of trace uranium in solid samples. RIMS offers near complete suppression of interferences. Quadrupole ICP-MS is relatively low cost and the most available. In sector field ICP-MS, the mass resolution can be increased up to 0.02 of a mass unit, which provides an inherent separation of many interferences. All types of ICP-MS can be coupled with a range of sample introduction systems to improve the delivery of sample to the plasma, and thereby increase the sensitivity of the measurement. TIMS provides the best isotope ratio measurements, with typical precisions of 0.25%, and AMS offers the lowest detection limits, in the region of  $10^4$  atoms, and effective reduction of interferences.

While  $^{238}\text{U}$  and, to a lesser extent  $^{235}\text{U}$ , are readily determined by all forms of mass spectrometry, there are challenges related to their relative concentrations. For example,  $^{238}\text{U}$  is orders of magnitude higher than any other isotope of uranium. When other isotopes are measurable, the  $^{238}\text{U}$  peak is very large and tailing from this peak can interfere with measurements at masses close to 238. **Figure 2** shows an ICP-MS spectrum in which the  $^{238}\text{U}$  count rate exceeds the capabilities of the electron multiplier detector. The peak tails into  $m/z$  236 and obscures any  $^{236}\text{U}$  present; however, it does not obscure the  $^{235}\text{U}$  peak. It should be noted that the count times for each peak were not equal.

Instruments have been developed to improve isotope ratio measurements, increase sensitivity, and reduce the effects of tailing, by reducing the energy

spread of the ion beam. Sector field ICP-MS uses a double-focusing system to achieve the lower energy spread, resulting in high sensitivity and highly accurate isotope ratio measurements, especially when fitted with a multicollector. Newer developments in ICP-MS such as collision cell ICP-MS improve sensitivity and isotope ratio measurements over quadrupole instruments by colliding the atoms prior to deflection; thus  $^{236}\text{U}$  can be measured in the presence of  $^{238}\text{U}$  at ratios  $3 \times 10^{-7}$ , and with a precision of 0.33%. The expensive and uncommon technique of heavy isotope accelerator mass spectrometry can deliver isotope ratio measurement of, for example,  $^{236}\text{U}$  and  $^{235}\text{U}$ , at a ratio of  $10^{-6}$ .

**See also:** **Activation Analysis:** Neutron Activation. **Extraction:** Solvent Extraction Principles. **Geochemistry:** Sediment. **Ion Exchange:** Overview. **Mass Spectrometry:** Overview. **Radiochemical Methods:** Gamma-Ray Spectrometry. **Sample Dissolution for Elemental Analysis:** Dry Ashing. **Spectrophotometry:** Overview. **X-Ray Fluorescence and Emission:** X-Ray Fluorescence Theory.

## Further Reading

- ASTM (2002) *D1971-02 Standard Practices for Digestion of Water Samples for Determination of Metals by Flame Atomic Absorption, Graphite Furnace Atomic Absorption, Plasma Emission Spectroscopy, or Plasma Mass Spectrometry*. ASTM International. For referenced ASTM standards, visit the ASTM website [www.astm.org](http://www.astm.org).
- ASTM (2001) *D4698-92 Standard Practice for Total Digestion of Sediment Samples for Chemical Analysis of Various Metals*. ASTM International. For referenced ASTM standards, visit the ASTM website [www.astm.org](http://www.astm.org).
- Becker JS (2003) Mass spectrometry of long-lived radionuclides. *Spectrochimica Acta* 58: 1757–1784.
- Burns PC and Finch R (1997) *Uranium: Mineralogy, Geochemistry and the Environment*. Washington: Mineralogical Society of America.
- Katz JJ, Seaborg GT, and Morss LR (1986) *The Chemistry of the Actinides*, 2nd edn. London: Chapman & Hall.
- L'Annunziata M (2003) *Handbook of Radioactivity Analysis*, 2nd edn. New York: Academic Press.
- Nancy AC (ed.) (1997) *HASL-300: The Procedures Manual of the Environmental Measurements Laboratory*, vol. I, 28th edn. New York: US DoE.
- Saito N (1984) Selected data on ion exchange separations in radioanalytical chemistry. *Pure and Applied Chemistry* 56: 523–539.
- Wilson PD (1996) *The Nuclear Fuel Cycle: from Ore to Wastes*. Oxford: Oxford University Press.



## Radiotracers

**P Viitanen**, VTT Chemical Technology, Espoo, Finland

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 4309–4317, © 1995, Elsevier Ltd.

### Introduction

Radiotracers have a variety of applications, the main uses being in engineering, medicine, and environmental sciences. Despite the variety of applications the same basic principles are used. This article gives an introduction to the principles of tracer techniques and measuring devices. Applications in industry, chemical, and physical research and clinical work as well as in food and environmental analysis are presented.

### Basic Principles of Radiotracers and Their Uses

A tracer defined in the simplest way is a substance that can be traced. It has some property that distinguishes it from its environment and makes it possible to detect it above the background.

Tracers are usually divided into inactive tracers (e.g., dyes) and radiotracers, i.e., radioactive substances. One advantage of radioactive tracers is their relatively easy and accurate detection compared to inactive tracers. Disadvantages include the radiation hazard involved and the expensive instrumentation demanded. Because of the radiation hazard the use of radiotracers is controlled in most countries. In order to minimize risks, an activity as low as possible should be used and short-lived isotopes are preferred.

The simplest form of tracer technique involves tracking the tracer or just measuring the tracer content in a sample. In these applications the tracer substance usually represents a larger amount of material, so by observing the tracer the bulk behavior can be studied. A homogeneity measurement can be achieved by measuring the tracer content of several samples. In medical work the presence or absence of the tracer in some location may be significant.

Measurement of the residence time distribution as the tracer passes through a system is another basic principle of tracer techniques. Typical applications are measurement of the transit time of a tracer pulse, a method applied in flow measurement, as well as measurement of a system response to a transient, a method that gives valuable information on a system, mainly industrial and environmental processes.

Isotope dilution is a method that can be utilized in mass and volume determination as well as in flow measurement. The method takes advantage of the decreased activity per volume or mass as a known amount of tracer is diluted to a larger volume or mass of inactive substance.

The main applications of radiotracers are in the fields of industry and chemical and physical research as well as in medicine.

The main applications of radiotracers in industrial work are:

- distinguishing and tracking of materials;
- flow rate measurement;
- study of process behavior;
- detecting and searching for leaks and cracks;
- determination of mass and volume;
- wear studies; and
- measuring mixture homogeneity.

Methods used in physical and chemical research make use of the same basic principles as the industrial applications. However, the work is of a slightly different character; some applications are mentioned below.

In clinical work, radiotracers are used in diagnostics and therapy as well as in pharmacokinetic and metabolic studies.

Applications in food and environmental analyses use the same principles as industrial and clinical applications. Food analysis has much in common with pharmacokinetic studies and environmental analysis is often related to industrial studies.

Another use of radiotracers, not clearly classified in the applications mentioned above, is age determination based on radioisotopes. Age determination is mainly applied in geological surveys, although some applications are also found in industry and food analysis.

### Counting Techniques

#### Detector Types

There are three different main types of radiation detectors. These are detectors based on gas ionization, scintillation detectors, and semiconductor detectors.

Detectors based on gas ionization are the ionization chamber, proportional counter, and Geiger–Müller counter. These devices measure the electric current pulses induced by radiation between electrodes in a gas-filled chamber. The ionization chamber has a low

bias voltage and can be used to measure only high-energy radiation. The current induced is proportional to the intensity of radiation. The ionization chamber can be connected to an electrometer in order to produce a time-averaged output proportional to the radiation dose rate. The proportional counter has a higher bias voltage and is energy-selective. The Geiger-Müller counter has a high bias voltage and one radiation quantum induces a high, constant-amplitude electric pulse.

The semiconductor detector is also based on ionization. The ionization takes place in the p-n junction of two semiconductor materials. At room temperature the signal-to-noise (S/N) ratio of semiconductor detectors is poor but increases when the detector is cooled, e.g., by liquid nitrogen. The semiconductor detector has a very good energy resolution.

Scintillation detectors are based on luminescence. Some materials have the special property that as they absorb a radiation quantum, a light flash is produced. The intensity of the flash is proportional to the radiation energy. The flash is amplified with a photoelectron multiplier. The energy resolution, though poor when compared with that of semiconductor detectors, is sufficient for many purposes.

### Counting Equipment

A basic radiation measuring system consists of the detector, a high-voltage power supply for bias voltage, a preamplifier, and a power supply for the preamplifier and signal amplifier. A pulse height discriminator is used to decrease the low-energy noise from the detector and the amplifier.

A rate meter generates a time-averaged number of pulses obtained from the signal amplifier. Rate meters are normally used in portable radiation dose rate monitors with a Geiger-Müller counter or a scintillation detector. A sealer is a device counting the number of pulses in a selected time.

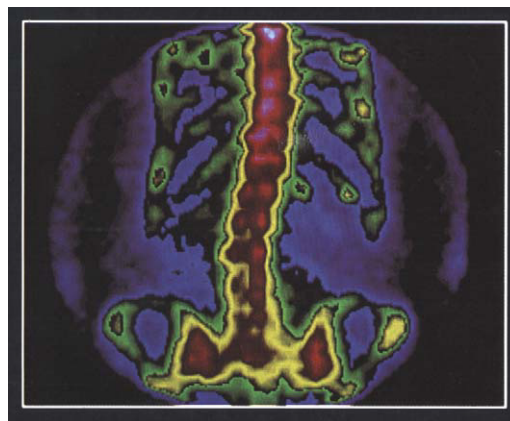
Pulse height analyzers are used more extensively for analysis purposes. The pulses from the amplifier are proportional to the radiation energy and can be analyzed either with a single-channel or a multichannel analyzer. A single-channel analyzer is used to count pulses in a certain energy range. A multichannel analyzer produces the energy spectrum of the measured radiation. Multichannel analyzer devices are usually expensive and their portability is restricted by the demand for cooling of the semiconductor detectors. Their use is necessary in some tracer applications, e.g., if a small amount of tracer has to be detected in the presence of strong background

radiation or if tracers with different energies have to be discriminated.

In many tracer applications it is not necessary to measure the absolute concentration of the tracer in a sample or in a material flow. In such cases a scintillation detector accompanied by a single-channel pulse height analyzer is a sufficient device.

A  $\gamma$ -camera, also referred to as a scintillation camera or Anger camera, is a device used mainly in clinical work. It consists of a large scintillation crystal, to which radiation is directed through a large pinhole collimator, and a matrix of photoelectronic multipliers. When a radiation-induced flash occurs in the scintillation material, the point of flash is calculated based on the intensity distribution registered by the photoelectron multipliers. Energy discrimination is used to decide whether the pulse was induced by the tracer or background. Modern computer techniques provide a means of monitoring the movements and accumulation of the tracer (**Figure 1**).

Positron emission tomography (PET) uses a device, in which several detectors are positioned circularly around the object. When a positron is emitted it almost immediately annihilates with an electron. As a result of the annihilation two  $\gamma$ -quanta are ejected in opposite directions. As two detectors detect simultaneously these 511 keV  $\gamma$ -quanta the occurrence is with a certain probability due to a positron emission at the line connecting the two detectors. As several occurrences at different angles are recorded an image of the object can be constructed.



**Figure 1** False-color scintigram of the human spine and ribs, revealing secondary cancers (metastases) in the vertebrae arising from a primary cancer of the prostate gland. A scintigram ( $\gamma$ -camera scan) is a record of radioactive emissions from an isotope (in this case,  $^{99m}\text{Tc}$ ) that is selectively absorbed by bone when injected into the body.  $\gamma$ -ray scintigraphy is frequently used to screen cancer patients for signs of secondary disease, often after their primary cancer has been treated. Here, the metastases appear as the pink and white 'hot spots' in the thoracic spine. (Reproduced from CNRI/Science Photo Library.)

## Counting Statistics

Radioactive decay is a statistical process. The number of decays per unit time obeys the Poisson distribution. The number of counts measured by the detector and the counting equipment is proportional to the activity of the sample. Factors affecting the proportionality are self-absorption by the source, absorption and scattering between the source and the detector, the efficiency of the detector, and the measuring geometry. The proportionality between the absolute sample activity and measured count rate is not necessary linear, but the assumption of linearity is sufficient in many cases.

An important characteristic of the Poisson distribution is that the mean and variance are equal. Thus, if  $N$  counts are measured in an arbitrary time  $T$ , then the mean deviation for the number of counts can be estimated as  $\sqrt{N}$ . If a background measurement shows  $N_b$  counts in time  $T$ , then  $N - N_b$  counts are due to the sample. The mean deviation for this is  $\sqrt{(N + N_b)}$ . The relative error of radiation measurement can be decreased by increasing the number of counts. Two alternatives are obvious: increasing the amount of radioactivity and increasing the counting time.

## Radiotracers in Industry and Research

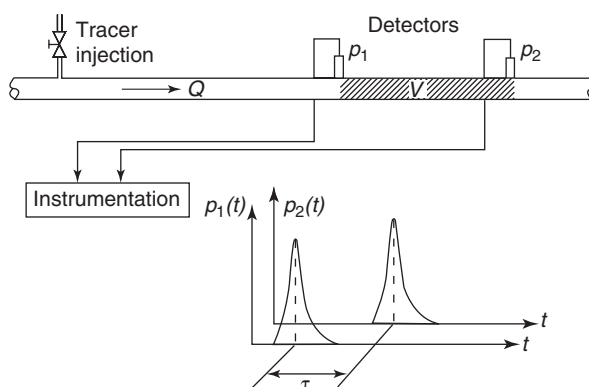
### Distinguishing and Tracking of Materials

Many tracer studies are intended to give only qualitative information such as information about the material of which an object is made or the location or route which the material or object is following. An example of this kind of application is the so-called pipe-pig used to clean long pipelines. A radiation source can be installed in the pig in order to localize it in case of blockage. In the early days of the industrial applications of radioisotopes the source in a pipe-pig was typically a 10 GBq cobalt-60 source. Nowadays more sensitive radiation detectors are used and short-lived radiotracers like manganese-56 or sodium-24 with activities of  $\sim 1$  GBq are used. These are installed in the pig each time just before the use.

Tracers have also been used to mark some special-purpose or special-quality materials, like stainless-steel alloy tubes for the heat exchangers of nuclear reactors.

### Flow Rate Measurement

Tracers can be used in flow rate measurement in order to calibrate existing flow meters or to measure flows where flow meter installation is difficult and the flow rate has to be known only temporarily.



**Figure 2** Flow measurement by the transit time method; flow rate is determined using the transit time ( $\tau$ ) of a tracer pulse between the two detectors.

Flow rate measurement with radiotracers is applicable to liquid and gaseous flows as well as to flows of solids.

The 'transit time method' is based on measuring the residence time of a tracer pulse between two measuring points. The flow rate  $Q$  is

$$Q = V\tau^{-1} \quad [1]$$

where  $V$  is the volume between the measuring points and  $\tau$  is the residence time. The transit time method is illustrated in Figure 2. The method is especially suitable for pipe flows. In order to compensate for the effects of radial velocity profile on the residence time, the flow must be turbulent and the tracer must be injected at least  $100 \times D$  ( $D$  is the pipe diameter) before the measuring interval in order to guarantee mixing across the pipe cross-section.

In the 'dilution method' tracer solution is fed into the main flow at a known flow rate  $q$ . The tracer concentration is measured from the original tracer solution and samples taken from the main flow. When the initial concentration is denoted by  $C_0$  and the sample concentration by  $C_1$ , it can be stated that

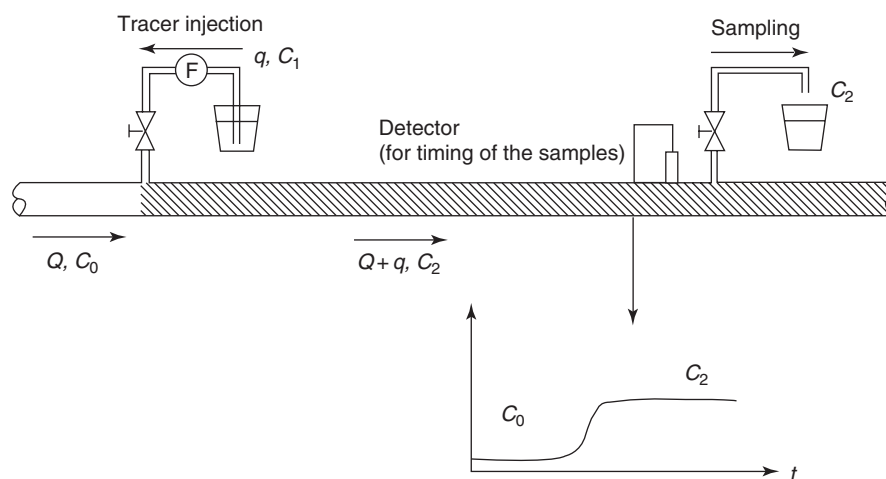
$$C_0 q = C_1 (Q + q) \quad [2]$$

Thus, the main flow rate  $Q$  is

$$Q = q \left( \frac{C_0}{C_1} - 1 \right) \approx q \frac{C_0}{C_1} \quad [3]$$

where the last identity is due to the fact that usually  $Q \gg q$ , and, thus, also  $C_0 \gg C_1$ . The method is illustrated in Figure 3.

'Pulse dilution methods' are based on a known amount of fed tracer and sampling from the main flow. At each point of the main flow downstream,



**Figure 3** Flow measurement by the dilution method; flow rate is determined by the dilution of the tracer solution in the main flow.

there is the relationship

$$\begin{aligned} A &= Q \int_0^{\infty} C(t) dt \\ &= C_0 V \end{aligned} \quad [4]$$

where  $A$  is the total amount of tracer,  $V$  is the volume of fed tracer,  $C_0$  is the concentration of the initial solution, and  $C(t)$  is the tracer concentration in the main flow at time  $t$ . The integral term can be determined, e.g., by continuous sampling during a time  $T$ , assuming that all tracer has passed the sampling location during this time. The tracer concentration of the sample is the average concentration:

$$\bar{C} = \frac{1}{T} \int_0^T C(t) dt \quad [5]$$

and the main flow rate can be determined as

$$Q = \frac{V C_0}{T \bar{C}} \quad [6]$$

In dilution methods the measurement of absolute concentrations is avoided by measuring count rates from the samples using a constant geometry.

### Study of Process Behavior

The determination of the residence time distribution of a process is probably the most used industrial application of radioactive tracers. The residence time distribution is the probability distribution of the time a piece of material is likely to spend in the process. The knowledge of this distribution is important when designing the process, as well as when the process is operated, in order to optimize the process operation conditions. The residence time distribution

can also be utilized in modeling many single- or multiphase flow processes.

The most straightforward way to determine the residence time distribution is to measure the system response for a short tracer pulse. If the system is closed, i.e., no back-mixing occurs at the process entrance and exit, the measured tracer concentration curve, when scaled to unity, is the residence time distribution.

### Detecting and Searching for Leaks and Cracks

There are a variety of leakage-detecting methods based on radiotracers, and choosing the method depends on the character of the suspected leak. The most frequently used methods with their applicability are listed below:

- In the case of a very large leak through a valve the leakage flow rate can be determined using standard methods of flow measurement with radiotracers.
- An arrangement where the flow entering a reactor is preheated in a heat exchanger by the flow leaving the reactor is quite common, for example, in the oil refinery industry. A leak in the heat exchanger can be detected by residence-time measurement. A tracer pulse is injected into the flow and the tracer concentration of the output flow is monitored. Any leakage is detected as a subsidiary peak preceding the main peak in the residence-time distribution. The amount of leaked material and the location of the leak can be estimated from this measurement.
- The direct tracer technique for leakage detection includes injection of tracer into a system and monitoring the system environment where the leak is suspected. The monitoring can be done either by using detectors *in situ* or by sampling. Sampling is preferable in the case of small leaks since the measuring of

samples can be done very accurately by increasing the analysis times. Moreover, the background can be eliminated.

- Leakage detection for pipes and long pipelines can be done by filling the pipe with a tracer and then leaving it under pressure in order for some of the tracer to leak out of the pipe. After the pipe is flushed out, the leak can be sought using a radiation monitor. In the case of underground pipes of depth more than  $\sim 0.5\text{--}1\text{ m}$ , a special pig can be used to localize the leaked tracer or the radiation monitoring can be done from strategically placed boreholes.

### Determination of Mass and Volume

Tracer methods are of practical interest when very large volumes of arbitrary shape are determined or a very small fraction of a large amount of material is mixed with a known amount of another material. If a small volume of tracer  $V_t$  is mixed evenly with a large volume  $V$ , and the count rates of the samples are  $c_t$  and  $c_0$  in constant geometry for the tracer and bulk material, respectively, and the count rate of the solution is  $c_s$ , then

$$c_0 V + c_t V_t = c_s (V + V_t) \quad [7]$$

This can be solved either for  $V$  or  $V_t$ ,

$$V = \frac{c_t - c_s}{c_s - c_0} V_t \quad [8]$$

and

$$V_t = \frac{c_s - c_0}{c_t - c_s} V \quad [9]$$

The first equation is for the isotope dilution method and the second for the reverse isotope dilution method.

### Wear Studies

The principle of isotopic wear studies is to label the material to be worn and monitor the material loss by radiation measurement. A typical example is studying car engine piston rings labeled, e.g., by neutron activation and measuring the accumulation of radioactivity in the lubricant oil. The measured count rates can be converted to mass by the appropriate calibration. Another example from the vehicle industry is to study the wearing of car tyres.

### Measuring Mixture Homogeneity

The determination of mixture homogeneity is based on labeling one component and taking samples from the mixture in order to determine the variations in tracer concentration.

As a measure for the mixture homogeneity the relative standard deviation of the measured count rates can be used. It is important to subtract the variations due to the statistical nature of radioactive decay, which would otherwise contribute to the measured variance. It has to be noted that the determination of homogeneity is not unambiguous. The variations in tracer concentration of samples depend on the sample size and on the microstructure of the material.

It is possible to measure the homogeneity of a ready-mixed composition by taking several samples from different positions. Another possibility is to monitor the variations in the tracer concentration during mixing. The maximum degree of mixing is achieved when the variations cease decreasing.

An online measurement of a continuous mixer can be made by injecting tracer continuously into the material to be mixed with the bulk material and measuring count rates after mixing. The sample size is now determined by the measurement geometry and the measuring time. The axial homogeneity can be determined as a variation of the measured count rates. If two opposed detectors are used, it is possible to obtain information about radial mixing as the difference in the response between the two detectors.

### Use in Physical and Chemical Research

The basic principles used in research work are similar to those used in industrial tracer applications.

Physical applications are mainly concerned with mass transfer. Typical applications are the study of separation processes, e.g., determining the efficiency of a filter.

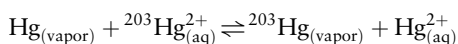
Very slow diffusion and self-diffusion processes in solids can be determined using radiotracers. Two methods are used. In one, a thin surface containing tracer is applied to the sample. After some time has elapsed, the sample is sliced parallel to the surface and the radioactivity of each slice is measured. Another method is to use an  $\alpha$ - or  $\beta$ -active tracer. As the tracer diffuses into the sample, a direct measurement of the surface radioactivity gives information about the diffusion process.

Adsorption of a gas on a solid surface may be of interest, e.g., when studying the efficiency of a catalyst. The adsorption measurement can be made using a labeled gas and measuring the surface activity. The background radiation due to the gas between the detector and the studied surface can be eliminated with a measurement in similar geometry without the adsorbent.

Chemical analysis can use tracer techniques to study reaction mechanisms and kinetics. Reaction pathways can be determined by labeling the starting

substance. When this is allowed to react in a solution containing an assumed intermediate, the intermediate character is verified if the intermediate contains tracer after the reaction. Reaction pathway studies can be more accurate if chemical methods are used to determine the position of the label. This involves the use of sophisticated processes, in which the labeled molecule is graded in successive reactions and the radioactivity of the intermediates are monitored.

Isotopic exchange is the best example of the power of radiotracer techniques in chemical research. Isotopic exchange reactions have no inactive analogies. A system with two states of an element in equilibrium apart from isotopic distribution can be studied:



Information on the character and strength of chemical bonds as well as the mechanism of the reactions accompanying the exchange can be obtained.

There are numerous isotopes that are used in physical and chemical research. To mention but a few elements that have radioactive isotopes with both relatively short and long half-lives, such as Na, Zn, Kr, Ag, Sn, Sb, I, and Hg. The most used isotopes to study reaction mechanisms and reaction kinetics are probably  $^{14}\text{C}$  and tritium,  $^3\text{H}$ , because they are both used in organic chemistry; tritium also has inorganic applications.

### Age Determination

Radioisotopic age determination is used for industrial purposes if a material is susceptible to aging. A method called double isotopic labeling can be used. The product is labeled when manufactured with a pair of isotopes having different half-lives and radiation energies. Examples of suitable pairs are chlorine-36 and manganese-54, chlorine-36 and niobium-95, and technetium-99 and niobium-95. As time passes, the initial ratio of the isotopes is changed.

Geological research can make use of radioisotopes in determining ages of minerals. This has been described in detail in another article.

## Radiotracers in Clinical Work

Tracers in clinical work are used in diagnostics and therapy as well as in pharmacokinetic and metabolic studies. The first clinical tracer tests were made in the 1920s in order to measure blood flow rate. Radon gas dissolved in salt solution was used as the tracer. Nowadays the clinical use of radiotracers covers diagnostic use, radiotracers are used as a tool in drug

design, and some forms of radiation therapy take advantage of radiotracers.

A variety of radiotracers are used in clinical work, the most used isotopes being technetium-99m, iodine-131, tantalum-201, xenon-133, and indium-113m. The use of technetium,  $^{99\text{m}}\text{Tc}$ , dominates, since it can be made to react with many substances having specific biological behavior.  $^{99\text{m}}\text{Tc}$  is obtained from an isotope generator, which is based on the radioactive decay of radioactive molybdenum,  $^{99}\text{Mo}$ . Pharmaceuticals containing  $^{99\text{m}}\text{Tc}$  are usually introduced by intravenous injection. Some radiopharmaceuticals may also be introduced orally, e.g., for those containing iodine this is the common procedure.

### Diagnostic Use of Tracers

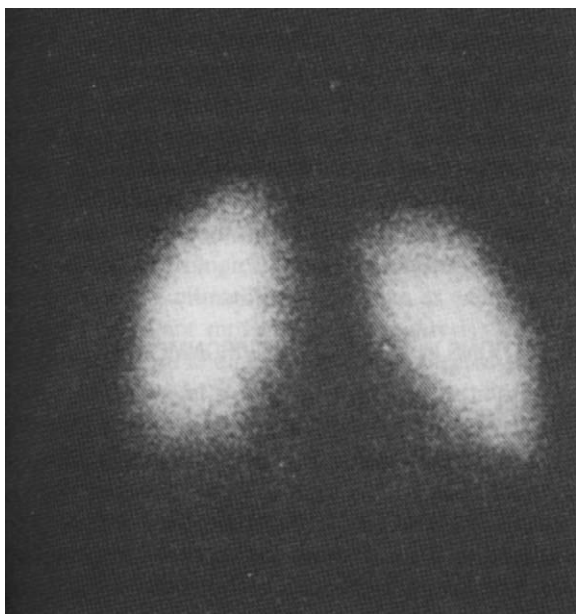
**Radioimmunoassay (RIA)** This is widely used to measure hundreds of substances of biological interest. The principle of classical RIA is to obtain the unknown concentration of the unlabeled antigen by comparing its inhibitory effect on the binding of radiolabeled antigen to a specific antibody with the inhibitory effect of known standards. The principle of RIA closely resembles the isotope dilution method.

In classical RIA, there is less binding protein, the antibody, present than there is antigen. At the end of the reaction period there are present a bound fraction and free fraction, both containing labeled antigen. The amount of labeled antigen in the bound fraction, however, decreases as the overall amount of unlabeled antigen increases and, thus, the ratio of radioactivity between the bound and free fractions can be used to determine the unknown amount of the unlabeled antigen. The technique for separation of antibody-bound antigen from free antigen is critical for the method. Nowadays the concept of RIA also covers the concept of labeled-antibody techniques.

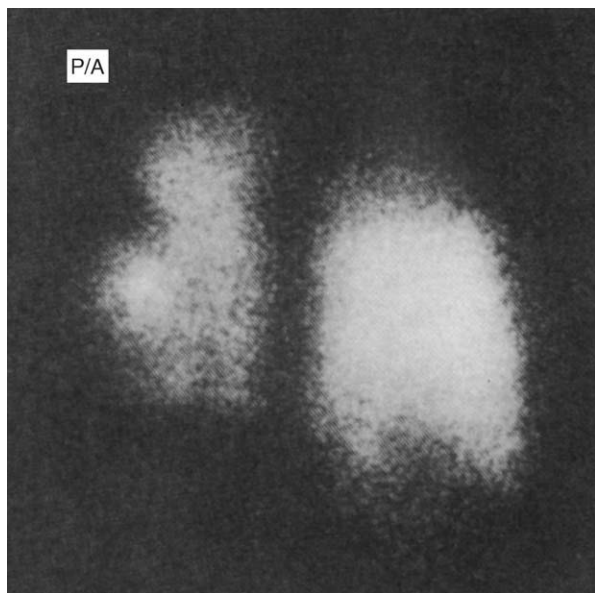
**Nuclear medians imaging techniques** These techniques have an important role in diagnostic work as a tool for detecting and localizing damage, tumors, etc., in tissue and organs. A  $\gamma$ -camera integrated with modern computer technology is a basic tool for data acquisition and processing in nuclear imaging. Technetium ( $^{99\text{m}}\text{Tc}$ ) labeled tracers are used in most imaging applications.  $^{99\text{m}}\text{Tc}$  is easily obtained from an isotope generator and has a favorable half-life and a suitable  $\gamma$ -energy, 140 keV, for detection.

Figures 4 and 5 give examples of gamma camera images. Figure 4 is the image obtained from a normal lung and Figure 5 shows a pulmonary embolism. Intravenously administered particles ( $\approx 20\text{ }\mu\text{m}$ ),  $^{99\text{m}}\text{Tc}$ -macroaggregates, were used for the scintigraphy of the lungs. The particles remain temporarily





**Figure 4** Gamma-camera image of normal lungs. (Reproduced from *Technetium Atlas* (1973) Petten, Holland: Philips-Duphar BV.)



**Figure 5** Pulmonary embolism shown by a gamma-camera. (Reproduced from *Technetium Atlas* (1973) Petten, Holland: Philips-Duphar BV.)

fixed in the arteria pulmonalis until they are cleared away mechanically and by phagocytosis.

Bone imaging is based on increased blood flow and the metabolic activity of new bone formation when bone is damaged by disease or injury. Many technetium-labeled bone seeking agents have been introduced, e.g.,  $^{99m}\text{Tc}$ -methylene diphosphonate ( $^{99m}\text{Tc}$ -MDP), which has a low cost, easy preparation,

high bone extraction rate, and rapid soft tissue clearance.

Radionuclide brain scintigraphy utilizes the accumulation of some tracers relative to healthy brain tissue when the integrity of the so-called blood-brain barrier is disrupted. The most commonly employed tracers are  $^{99m}\text{Tc}$ -pertechnetate ( $^{99m}\text{TcO}_4$ ),  $^{99m}\text{Tc}$ -diethylenetriaminepentaacetic acid ( $^{99m}\text{Tc}$ -DTPA), and  $^{99m}\text{Tc}$ -glucoheptonate ( $^{99m}\text{Tc}$ -GH).

The thyroid gland is an organ often studied with the help of radiotracers. Suitable tracers are the radioactive isotopes of iodine,  $^{123}\text{I}$  and  $^{131}\text{I}$ , as well as  $^{99m}\text{TcO}_4$ . Hyperthyroidism is indicated by enlargement of the thyroid, and hypothyroidism can be detected as a decreased uptake of the tracer. A heterogeneous tracer distribution also indicates an abnormality of the thyroid.

Cardiac scintigraphy can be used to obtain parameters of cardiac perfusion and function.  $^{99m}\text{Tc}$ -pyrophosphate can be used to assess acute myocardial infarction. In the case of myocardial infarction the agent localizes in the area of necrotic tissue, forming an intense spot of radioactivity. Global uptake of activity in the ventricular myocardium usually represents some other cardiac disease.

A cardiac perfusion study is done using tantalum-201 ( $^{201}\text{Ta}$ ) chloride at rest or during both exercise and rest. A defect is noted as the absence of or greatly reduced activity in the area of the myocardium. If the defect is noted during both exercise and rest, the area is considered to be irreversibly scarred and previous myocardial infarction is assumed. If the defect is noticeable only in the rest study, a reversible and often surgically correctable perfusion abnormality is assumed.

The ventricular ejection fraction can be determined using radiotracers. The activity in the ventricle is proportional to the amount of blood. Using a tracer that remains in the blood circulation and a  $\gamma$ -camera with a high-sensitivity collimator, the ejection fraction can be determined.  $^{99m}\text{Tc}$ -labeled red cells provide a suitable tracer for this purpose. The ejection fraction of the left ventricle is normally  $>50\%$  and this is increased during exercise. A decreased ejection fraction of the left ventricle during exercise is evidence of coronary disease. An abnormal ejection fraction of the right ventricle indicates a pulmonary disorder.

The liver and spleen can be studied with  $^{99m}\text{Tc}$ -labeled sulfur colloid. Space-occupying lesions such as cysts, abscesses, hematomas, or tumors can be detected in static images. For hepatobiliary evaluation derivatives of iminodiacetic acid labeled with  $^{99m}\text{Tc}$  are used.

Renal scintigraphy provides a means to assess aberrant renal anatomy and pathophysiology. The renal

cortices can be visualized with  $^{99m}\text{Tc}$ -dimercapto-succinic acid or  $^{99m}\text{Tc}$ -glucoheptanate. Determination of the functional status of a transplanted kidney is an important area of renal scintigraphy.  $^{99m}\text{Tc}$ -DTPA can be used as the tracer to evaluate semiquantitatively flow and function of a transplanted kidney.

PET is used mainly in the study of brain metabolism. The demand of the tracer to be  $\beta^+$ -active restricts the variety of the usable tracers. Isotopes mainly used in PET are carbon-11, nitrogen-13, oxygen-15, and fluorine-18. These are produced by cyclotron and have quite short half-lives (20.4, 10.0, 2.1, and 110 min, respectively).

### Use in Drug Design

Tracer methods provide a tool in different stages of drug development. In the design stage a few atoms in a drug can be replaced by radioactive ones, and the labeled drug allowed to react with cells or animal tissues. The autoradiograph reveals the binding sites as drug accumulation in a certain active site, a receptor. Thus, clues about the reaction mechanism between the drug and the biochemical target are obtained. New drugs developed are submitted to very strict safety tests. Pharmacokinetic and metabolic studies are used in this stage of drug development.

Thoroughly planned preparation and chemical purity of the radiolabeled drug are of essential importance for the success of pharmacokinetic and metabolic studies. It is desirable that the label site in the drug molecule is not lost through any probable metabolic mechanism. In cases, in which the drug is divided into two major entities, it may become necessary to label both portions.

Prior to human studies, tests of the radioactive drug are made with laboratory animals. Observations with animals include monitoring radioactivity in blood samples and monitoring accumulation of radioactivity in urine and fecal samples. Metabolic pathways are determined and possible organs or sites of drug retention are determined.

Human studies with a radiolabeled drug must be clearly justified on the basis of animal studies and analytical results obtained with the unlabeled drug. If there is any tendency for the drug to accumulate in some organs, it places such experiments in doubt. The main purpose of the human test made with a radiolabeled drug is to evaluate the complete material balance for the drug intake and output to ensure that no unknown metabolic pathway exists.

### Use in Radiation Therapy

Radiation therapy is based on the fact that infected cells or cancer cells usually have a poor resistance to

the harmful effects of radiation, while ordinary healthy cells have comparatively good resistance to radiation and are renewable.

One form of radiation therapy is isotope therapy, in which a labeled substance that is known to accumulate in the defective organ is introduced into the body. The radiopharmaceutical may be a compound labeled with radioactivity or a radioactive isotope that has the tendency to accumulate. The most-used application of isotope therapy is the use of radioactive iodine  $^{131}\text{I}$  to treat hyperthyroidism and cancer of the thyroid. The rest of the treatments are covered by treatment of blood illnesses with compounds containing radioactive phosphorus  $^{32}\text{P}$  or the treatment of joints with the yttrium isotope  $^{90}\text{Y}$ .

A recent innovation under study in the area of radiation therapy is boron-neutron capture therapy for brain tumor. This form of therapy uses an inactive tracer containing boron, an element with a high absorption rate of neutron radiation. When the radiopharmaceutical has accumulated in the tumor, the patient is exposed to neutron radiation, which is absorbed by boron. The energy of secondary low-range  $\alpha$ -radiation, which is formed in the boron-neutron reactions, is then absorbed by the tumor cells that are to be destroyed.

## Applications in Food and Environmental Analysis

### Food Analysis

Application in nutritional technology often resembles pharmacokinetic studies in clinical work. The aims of such studies can also be considered to be quite similar, i.e., to investigate the fate of food substances in the digestive system and to ensure that no unacceptable effects are present. For example carbon-14 ( $^{14}\text{C}$ ) can be used in studying metabolic effects related to food carbohydrates. Another application of the tracer technique is determining the origin of the food substance, e.g., the species of meat can be identified by RIA.

The principle of isotopic age determination is used in food analysis. Radiocarbon analysis can be used to detect whether a product is of natural or synthetic origin. The absence of  $^{14}\text{C}$  in the sample would indicate a synthetic origin, while the presence of  $^{14}\text{C}$  indicates a natural origin. The determination of tritium contents can be used, e.g., to determine the age of wine.

### Environmental Analysis

Environmental analysis with radiotracers is often related to monitoring industrial activities with

potential environmental impact. Wastewater management is an area where radiotracers are often used. The functioning of large aeration pools and sedimentation tanks is often difficult to predict. However, tracers provide a technique involving the measurement of the residence time distribution for this purpose. Flow measurement techniques using radiotracers can also be used in flow meter calibration and validating in arbitrary-shaped flow channels and for the large flow rates often met in wastewater management.

In agricultural research, radiotracers have been used to study the intake of fertilizers by plants as well as for studying the absorption of fertilizers by the soil. Isotopically labeled compounds have also been used to study the movement and degradation of pesticides.

The final disposal of nuclear waste is a major question in the environmental debate. One scenario for waste disposal is burying waste in bedrock. Radiotracers have been used to determine bedrock properties by the residence time distributions between boreholes. Another source giving information on how nuclear waste would behave in the bedrock is to study natural analogies, e.g., uranium deposits.

Global environmental research includes such items as studies of the hydrological cycle, deforestation, erosion, sedimentation, and climatological studies, as well as investigation of pollutant movements and analysis. Radiotracers can be utilized in all of these interconnected subjects. Global environmental studies are focusing on understanding the mechanism of the world climate and possible man-made changes to it, the 'greenhouse effect' being a major question.

Tritium ( $^3\text{H}$ ) has proved to be useful in studying surface water movements as well as movements in soil and bedrock. Age determination of groundwater and ice can be performed using the ratios of several natural isotopes.

Sedimentation in rivers, a major consequence of deforestation, can be studied by determining the age

of different sediment layers. The isotopes that have been utilized are lead-210, cesium-137, and polonium-210.

Erosion and desertification can be studied, e.g., by determining changes in cesium-137 levels in surface soil, the origin of cesium-137 being the fallout of early nuclear experiments.

Studies of pollutants include the investigation of pollutant movements in water, soil, and the atmosphere. The tracer may be of natural or artificial origin. A widely studied large-scale case is the Chernobyl accident of 1986.

**See also:** **Archaeometry and Antique Analysis:** Dating of Artifacts. **Bioassays:** Overview. **Drug Metabolism:** Metabolite Isolation and Identification; Isotope Studies. **Fertilizers.** **Food and Nutritional Analysis:** Meat and Meat Products. **Immunoassays, Techniques:** Radio-immunoassays. **Isotope Dilution Analysis.** **Pesticides.** **Pharmaceutical Analysis:** Drug Purity Determination. **Process Analysis:** Overview. **Radiochemical Methods:** Pharmaceutical Applications. **Water Analysis:** Industrial Effluents.

## Further Reading

- Bowen G, Rozanski K, and Vose P (1990) Isotopes in environmental research. *IAEA Bulletin* 32(4): 5–8.
- Charlton JS (1986) *Radioisotope Techniques for Problem-Solving in Industrial Process Plants*. Glasgow: Leonard Hill.
- Földiák G (1986) *Industrial Application of Isotopes*. Amsterdam: Elsevier.
- Hamner CE (1990) *Drug Development*, 2nd edn. Boca Raton, FL: CRC Press.
- Knoll GF (1988) *Radiation Detection and Measurement*, 2nd edn. New York: Wiley.
- Morris BA and Clifford MN (1985) *Immunoassays in Food Analysis*. London: Elsevier.
- Waggenger RG, Kereiakes JG, and Shalek RJ (1990) *Handbook of Medical Physics*, vol. III. Boca Raton, FL: CRC Press.

## Radio-Reagent Methods

**J Tölgyessy and M Harangozó**, Slovak Technical University, Bratislava, Slovak Republic

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A meaningful definition should reflect the differences between the radio-reagent method and indicator

analysis in addition to setting the boundary with isotope dilution analysis. In indicator analysis (sometimes called the method of labeled atoms; radiotracer methods), an appropriate isotope of the analyte is added before the chemical operations are started. After mixing, it is practically impossible to separate radioactive and inactive ions, molecules and atoms of the same chemical composition, so the amount of substance in every part of the system can be

potential environmental impact. Wastewater management is an area where radiotracers are often used. The functioning of large aeration pools and sedimentation tanks is often difficult to predict. However, tracers provide a technique involving the measurement of the residence time distribution for this purpose. Flow measurement techniques using radiotracers can also be used in flow meter calibration and validating in arbitrary-shaped flow channels and for the large flow rates often met in wastewater management.

In agricultural research, radiotracers have been used to study the intake of fertilizers by plants as well as for studying the absorption of fertilizers by the soil. Isotopically labeled compounds have also been used to study the movement and degradation of pesticides.

The final disposal of nuclear waste is a major question in the environmental debate. One scenario for waste disposal is burying waste in bedrock. Radiotracers have been used to determine bedrock properties by the residence time distributions between boreholes. Another source giving information on how nuclear waste would behave in the bedrock is to study natural analogies, e.g., uranium deposits.

Global environmental research includes such items as studies of the hydrological cycle, deforestation, erosion, sedimentation, and climatological studies, as well as investigation of pollutant movements and analysis. Radiotracers can be utilized in all of these interconnected subjects. Global environmental studies are focusing on understanding the mechanism of the world climate and possible man-made changes to it, the 'greenhouse effect' being a major question.

Tritium ( $^3\text{H}$ ) has proved to be useful in studying surface water movements as well as movements in soil and bedrock. Age determination of groundwater and ice can be performed using the ratios of several natural isotopes.

Sedimentation in rivers, a major consequence of deforestation, can be studied by determining the age

of different sediment layers. The isotopes that have been utilized are lead-210, cesium-137, and polonium-210.

Erosion and desertification can be studied, e.g., by determining changes in cesium-137 levels in surface soil, the origin of cesium-137 being the fallout of early nuclear experiments.

Studies of pollutants include the investigation of pollutant movements in water, soil, and the atmosphere. The tracer may be of natural or artificial origin. A widely studied large-scale case is the Chernobyl accident of 1986.

**See also:** **Archaeometry and Antique Analysis:** Dating of Artifacts. **Bioassays:** Overview. **Drug Metabolism:** Metabolite Isolation and Identification; Isotope Studies. **Fertilizers. Food and Nutritional Analysis:** Meat and Meat Products. **Immunoassays, Techniques:** Radio-immunoassays. **Isotope Dilution Analysis. Pesticides. Pharmaceutical Analysis:** Drug Purity Determination. **Process Analysis:** Overview. **Radiochemical Methods:** Pharmaceutical Applications. **Water Analysis:** Industrial Effluents.

## Further Reading

- Bowen G, Rozanski K, and Vose P (1990) Isotopes in environmental research. *IAEA Bulletin* 32(4): 5–8.
- Charlton JS (1986) *Radioisotope Techniques for Problem-Solving in Industrial Process Plants*. Glasgow: Leonard Hill.
- Földiák G (1986) *Industrial Application of Isotopes*. Amsterdam: Elsevier.
- Hamner CE (1990) *Drug Development*, 2nd edn. Boca Raton, FL: CRC Press.
- Knoll GF (1988) *Radiation Detection and Measurement*, 2nd edn. New York: Wiley.
- Morris BA and Clifford MN (1985) *Immunoassays in Food Analysis*. London: Elsevier.
- Waggner RG, Kereiakes JG, and Shalek RJ (1990) *Handbook of Medical Physics*, vol. III. Boca Raton, FL: CRC Press.

## Radio-Reagent Methods

**J Tölgyessy and M Harangozó**, Slovak Technical University, Bratislava, Slovak Republic

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A meaningful definition should reflect the differences between the radio-reagent method and indicator

analysis in addition to setting the boundary with isotope dilution analysis. In indicator analysis (sometimes called the method of labeled atoms; radiotracer methods), an appropriate isotope of the analyte is added before the chemical operations are started. After mixing, it is practically impossible to separate radioactive and inactive ions, molecules and atoms of the same chemical composition, so the amount of substance in every part of the system can be

determined at any time by measurement of the radioactivity. The conditions for the application of this method are that:

1. all the analytes are in reactive form at the start of the analysis; and
2. the proportionality constant does not change during the chemical operations.

The main feature of the 'isotope dilution method' is that complete isolation of the analyte is not necessary. Isotope dilution can be called the method of corrections. Another characteristic of isotope dilution analysis is its use of a change in specific activity.

The 'radio-reagent method' is based on the use of a radioactive species in a quantitative reaction, and measurement of the change in activity of that species in the course of the reaction. The radioactive species may be a labeled reagent, the analyte, or a substance able to give an exchange reaction with some compound of the analyte.

Consequently, characteristic features of this method are as follows:

1. the determination is based on a well-known chemical reaction;
2. the fraction of the labeled substance consumed in the reaction passes into a second phase;
3. this fraction is determined by the decrease, increase, or appearance of radioactivity in the selected phase; and
4. calculation of the concentration of the analyte is based on a knowledge of the corresponding factor relating radioactivity to concentration or mass (grams, grams per liter, moles, moles per liter, etc.).

After separation from excess reagent (by liquid-liquid distribution, chromatography, precipitation, etc.), the mass or concentration of this product is determined from activity measurement. The determination is based on a radioactive substance chemically different from the analyte substance (in contrast to isotope dilution analysis); therefore, the chemical reaction is of prime importance. By variation of this key reaction, the principle can be adapted to various procedures. The superiority of radio-reagent methods over classical separation techniques arises from the use of an inactive carrier and the high sensitivity of the activity measurements, which are not subject to interference by the carrier or other substances.

## Classification

The classification of radio-reagent methods is difficult. Depending on the type of chemical interaction

and relationship of the analyte to the radioactive substance measured, radio-reagent methods can be divided into three basic groups:

1. The radioactive substance is a typical reagent that is able to react with the compound to be determined.
2. The method is based on 'nonequivalent' competition; i.e., the radioactive substance is able to exchange with the analyte.
3. The method is based on 'equivalent' competition; i.e., the labeled substance is chemically identical to the substance to be determined.

The difference between these groups depends on which of the substances involved is radioactive. For example, if the precipitation of barium or radium sulfate is used in a determination, the following variants may occur: (1) the inactive barium is determined by precipitation with an excess of radioactive sulfate (or inactive sulfate is determined by a surplus of radioactive barium); (2) barium is determined by precipitation with an insufficient amount of sulfate in the presence of radium, which acts as an indicator; thus the higher the barium content, the lower the fraction of barium precipitated and, correspondingly, of radium co-precipitated; (3) a constant but insufficient amount of sulfate can be used for the determination of barium, the indicator for the completion of precipitation being radioactive barium.

From a practical viewpoint, it is convenient to divide radio-reagent methods into five groups:

1. Simple radio-reagent methods in which the reactions take place quantitatively, i.e., either the analyte or reagent is completely consumed and compounds of definite composition are formed.
2. Methods of concentration-dependent distribution (CDD) utilizing reactions where products of unstable composition are formed, but in which the extent of reaction is determined by the corresponding equilibrium constant or reaction time.
3. Isotope-exchange method based on the exchange of isotopes between two different compounds of one element, with a radioactive isotope in one of the compounds and a nonradioactive isotope in the other, usually the analyte.
4. Radio-release methods comprising procedures in which the analyte reacts with a radioactive reagent, thus releasing an aliquot of the reagent activity, in most cases into the gas phase.
5. Radiometric titration usually involves use of radioactive reagents to determine the equivalence point (endpoint) but these are variants that do not use a radio-reagent; instead they are based on a

**Table 1** Useful concentration regime of the radio-reagent method

Analyte	Radionuclide	Labeled reagent	Separation	Concentration regime
H <sup>+</sup>	<sup>131</sup> I	KI	Extr. I <sub>2</sub> into pyridine/chloroform	> 0.1 µg ml <sup>-1</sup>
H <sub>3</sub> BO <sub>3</sub>	<sup>18</sup> F	HF	Extr. HBF <sub>4</sub> into 1,2-dichloroethane <sup>a</sup>	≈ 0.1 µg B
NaLs	<sup>59</sup> Fe	[Fe(II)(phen) <sub>3</sub> ]	Extr. Fe(phen) <sub>3</sub> (LS) <sub>2</sub> into CHCl <sub>3</sub>	5–300 ng ml <sup>-1</sup> Na
Cl <sup>-</sup> , Br <sup>-</sup> , I <sup>-</sup>	<sup>203</sup> Hg	C <sub>6</sub> H <sub>5</sub> Hg <sup>+</sup>	Extr. C <sub>6</sub> H <sub>5</sub> HgCl into benzene	0.5–15 mg ml <sup>-1</sup>
Cl <sup>-</sup> , I <sup>-</sup>	<sup>110m</sup> Ag	AgNO <sub>3</sub>	Pptn. AgCl; AgI	0.4–250 µg ml <sup>-1</sup>
Cl <sup>-</sup>	<sup>203</sup> Hg	HgNO <sub>3</sub>	Pptn. Hg <sub>2</sub> Cl <sub>2</sub>	0.8–13 µg ml <sup>-1</sup>
Bi	<sup>131</sup> I	KI <sup>b</sup>	Extr. HBil <sub>4</sub> into n-butyl acetate	0.04–4 µg
Cationic detergents	<sup>131</sup> I	Rose Bengal	Extr. Ion-associate into CHCl <sub>3</sub>	

<sup>a</sup>In presence of methylene blue.<sup>b</sup>In presence of ascorbic acid + Na<sub>2</sub>SO<sub>3</sub> in 0.6 mol l<sup>-1</sup> H<sub>2</sub>SO<sub>4</sub>.

Extr. = solvent extraction; pptn. = precipitation; LS = lauryl sulfate; phen = 1,10-phenanthroline.

change in the intensity of radiation caused by its absorption or scattering in a medium containing the analyte.

## Simple Radio-Reagent Methods

These methods utilize a reaction with a suitable reagent, and sometimes with a third substance; one of the reagents is radioactive and the reaction is quantitative.

Simple radio-reagent methods can be classified into the following three groups:

1. determination with labeled reagents;
2. determination with labeled analyte; and
3. determination with labeled competing substances.

### Determination with Labeled Reagents

In this most widely used type of simple radio-reagent method, an excess of radioactive reagent solution of known analytical concentration is usually used. However, it is necessary to separate the active product from the excess of radioactive reagent. For this reason, this procedure makes use of various separation techniques, such as precipitation, formation of extractable chelates, or sorption. The majority of determinations are based on classical precipitation reactions.

In these determinations, the substance to be determined ( $n_R$  mol) forms a precipitate, extractable compound, or other separable substance. After the separation of the product, the radioactivity of the product ( $A_P$ ), or the excess of unreacted reagent ( $A_E$ ), or both, are measured. The following relationship is valid:

$$\frac{A_R}{n_R} = \frac{A_P}{zn_x} = \frac{A_E}{n_R - zn_x}$$

from which

$$n_x = \frac{n_RA_P}{zA_R} = \frac{n_R}{z} \left( 1 - \frac{A_E}{A_R} \right)$$

where  $A_R$  is the radioactivity of the reagent and  $z$  denotes the stoichiometric ratio in the compound formed (the number of moles of the reagent interacting with 1 mol of the test substance). In cases where the exact composition of the compound formed is not known, a calibration graph can be used. Quantitative isolation of trace amounts of the reaction products poses another important problem (Table 1).

A portable apparatus has been developed for the determination of chloride in water (0.001%) by precipitation with silver labeled with <sup>110m</sup>Ag. A rapid and routine method for the determination of sulfate in soils is based on its precipitation with barium salt labeled with <sup>133</sup>Ba.

In some cases filtration, sedimentation, or centrifugation can be omitted, for example, when the radioactive precipitant is a gas. <sup>14</sup>CO has been used in this way to determine hemoglobin in blood; and <sup>131</sup>I for the determination of bismuth in the presence of an excess of iodide ions using extraction separation.

### Determination with Labeled Analyte

This type of radio-reagent method is the opposite of determinations using a labeled reagent. Both groups of determinations have a common feature in that the radioactively labeled substance must be present in excess. Should the nonradioactive component be in excess, then the entire added radioactivity, irrespective of the quantity of the components, would be found in the phase of the reaction products and a determination would be impossible.

The radioactive indicator is added to the unknown sample and isotope exchange is allowed to take place; then the reagent is added in a substoichiometric but known quantity. After the reaction is complete, the phases are separated and activities measured. In those cases where the activity of the equilibrium solution is measured, the unknown concentration can be



calculated from the equation:

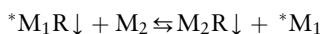
$$x + x_0 = \frac{A_{\text{init}}B}{A_{\text{init}} - A_{\text{equilib}}}$$

where  $x$  is the quantity of analyte (equivalents),  $x_0$  the amount of added radioactive indicator (equivalents),  $B$  the amount of unlabeled reagent (equivalents), and  $A_{\text{init}}$  and  $A_{\text{equilib}}$  represent the total activity of the initial and equilibrium solutions, respectively.

In this variant of the radio-reagent method, solvent extraction has found greater application.

### Determination with Labeled Competing Substances

Analyte  $M_2$  competes with the radioactively labeled substance  $^*M_1$  in the formation of a precipitated compound with a reagent  $R$



$^*M_1R$  and  $M_2R$  occur in a different phase than either substance  $M_1$  or  $M_2$ . The equilibrium constant of the reaction, as a rule, is  $> 1$  so that a significant exchange takes place. This method is used when the labeling of reagent  $R$  cannot be used or when the reagent and its compound with the analyte are transferred into the same phase. Generally, a relatively soluble precipitate with one radioactively labeled component (e.g.,  $^{45}\text{CaCO}_3$ ) is brought into contact with a solution containing the analyte ( $^{45}\text{Ca}^{2+}$ ) from the precipitate into the solution, and a less soluble precipitate ( $\text{PbCO}_3$ ) is formed. The resulting radioactivity of the solution is proportional to the initial amount of lead.

### Method of Concentration-Dependent Distribution

The method of CDD is based on the utilization of a calibration graph that shows the dependence of the distribution ratio in a two-phase system on the total concentration of the substance to be determined.

Characteristic features of the method of CDD are as follows:

1. The analysis is based on the distribution of a radioactive substance between two phases or between several parts of the systems used (paper chromatography, electrophoresis).
2. The ratio of the activities in both phases and parts of the system depends strongly on the initial concentration of the unknown substance.
3. The distribution is not given by the stoichiometric ratio of the reacting substances; that is, not by total

saturation of one phase or a nearly complete consumption of the reagent, but is given exclusively by the corresponding equilibrium or kinetic constants.

The determination of chloride in water samples by using a radioactive silver chloride ( $\text{AgCl}$ ) precipitate can be taken as an example. A water sample with an unknown concentration of chloride is shaken with an excess of labeled  $\text{AgCl}$  precipitate and the radioactivity of silver is measured in the equilibrium solution. The higher the chloride concentration the lower is the concentration of silver in solution (owing to the constancy of the  $\text{AgCl}$  solubility product). The  $\text{Cl}^-$  ions determined do not react with the  $\text{AgCl}$  precipitate, they only suppress its dissolution. The amount of chloride in the precipitate has no direct relationship to the concentration of  $\text{Cl}^-$  ions in the sample solution.

The second example is the determination of barium based on sorption on a hydrated iron(III) oxide precipitate. It has been found that the distribution coefficient of barium between the precipitate and the solution (under constant experimental conditions such as phase ratio, pH, concentration of other substances, temperature) depends on the total concentration of barium in the system. If this dependence is determined experimentally, it can be used for the construction of a calibration graph and the determination of barium. In this case, a known amount of a suitable barium radionuclide is added to the unknown amount of barium salt, and the adsorption is allowed to take place under standard conditions. From the activities of both phases, and use of the calibration graph, the unknown concentration of barium, corresponding to the measured ratio  $\text{Ba}(\text{precipitate})/\text{Ba}(\text{solution})$ , can be calculated.

In classifying the methods of CDD, it is useful to define two extreme types of determination: 'saturation analysis' and 'nonsaturation analysis'.

In saturation analysis, the cause of the change in the distribution of the radioactive substance is an increase in the saturation of the reagent by the given substance. In the example mentioned above, the hydrated iron(III) oxide adsorbed is successively saturated with barium ions. The term saturation analysis originally appeared in the literature in connection with the determination of biochemically important substances (steroids, hormones, vitamins) and is equivalent to the term radioimmunoassay.

For the determination of metals, reagents forming extractable chelates or ion associates or reagents forming neutral or negatively charged complexes are most often used. The ratio of complexed metal to uncomplexed metal is greater at lower metal concentration at constant reagent concentration under

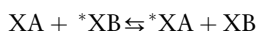
identical experimental conditions. The degree of complexation can be measured experimentally using liquid–liquid distribution, ion exchange, sorption, or by other means.

In nonsaturation analysis, the cause of the change in the distribution coefficient is the shift of the chemical equilibrium due to an increase in the concentration of the substance to be determined, irrespective of the extent of saturation. The concentration of the analyte may differ from that of the reagent by several orders of magnitude. The determination of some extractants can serve as a typical example. The extraction of microgram quantities of metals depends on the concentration of extractants, and the calibration graph may be used in the form  $\log D$  versus  $\log c_{\text{HA}}$ , where  $D$  is the distribution ratio of the metal in question and  $c_{\text{HA}}$  is the concentration of the reagent.

The number of procedures described that use nonsaturation analysis is limited. These procedures are generally characterized by extremely high theoretical sensitivity but low precision.

## Isotope Exchange Methods

This type of determination is based on the exchange of isotopes between two different compounds of the element X, one of which (XA) is nonradioactive, the other (XB) being labeled with a radioactive isotope. After isotopic equilibrium is reached,



The specific activities of the element X in both compounds are equal:

$$A_1/m = A_2/m_x$$

where  $A_1$  and  $A_2$  are the equilibrium radioactivities of XB and XA, respectively, and  $m_x$  and  $m$  are the amounts of X in XA and XB, respectively. The magnitude of  $m_x$  can be computed from

$$m_x = mA_2/A_1$$

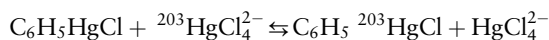
or from the calibration graph of  $m_x = f(A_2/A_1)$ ,  $m_x = f(A_2)$  or  $m_x = f(A_1)$ .

Isotopic exchange can be carried out in either a heterogeneous or a homogeneous system.

'Homogeneous isotope exchange' is usually carried out in organic solvents by extracting the element into the organic phase and adding to the separated extract another complex of the element, which is labeled. After equilibrium is attained, one of the chemical forms is reextracted into aqueous phase, provided that the difference between the magnitudes

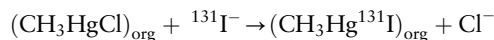
of the extraction constants of both complexes is high enough.

The determination of phenylmercury compounds is based on the isotope exchange between phenylmercury chloride in the sample and a standard solution of mercury-203 in  $3 \text{ mol l}^{-1}$  HCl. After reaching isotopic equilibrium for the reaction,



The specific activity of mercury is the same in both forms, and for  $m > 100m_x$  it is practically equal to the original specific activity of the added radiomercury. For this reason, the radioactivity of the separated organomercuric compound is directly proportional to its amount. For the separation of  $\text{C}_6\text{H}_5\text{HgCl}$  from  $\text{HgCl}_4^{2-}$ , liquid–liquid extraction can be used.

Methylmercury and phenylmercury are determined down to  $10 \text{ pg ml}^{-1}$  by liquid–liquid extraction from  $3 \text{ mol l}^{-1}$  aqueous HCl into benzene, and addition of  $\text{K}^{131}\text{I}$  to the separated organic phase. Inactive chloride is completely displaced by  $^{131}\text{I}$ :



The method has been used for the determination of methylmercury species in fish and in drinking water.

In heterogeneous isotope exchange, the aqueous solution containing an unknown amount of the elements to be determined is mixed with a solution of a known amount of a suitable complex of the same element in an organic diluent that does not contain an excess of the free complexing agent. One of the chemical forms of the element is labeled. This technique has been used to assay traces of silver in water by shaking the aqueous solution with a solution of labeled (with  $^{110\text{m}}\text{Ag}$ ) silver dithizonate in carbon tetrachloride.

## Radio-Release Methods

In radio-release methods, the nonradioactive test substance is brought into contact with a radioactive reagent. As a result of an appropriate chemical reaction, a portion of the activity of the radioactive agent is released, i.e., it passes from the solid to the liquid or gas phase, or from the liquid to the gas phase. A characteristic feature of radio-release methods is that the radioactive substance leaves the solid or the liquid phase without any nonradioactive substance replacing it.

The determination is carried out most often using a calibration graph, constructed by plotting the quantity of released radioactivity, or the rate of

release, against the concentration of the analyte. In some cases it is possible to calculate the amount of analyte from the amount of radioactivity released, if the specific activity of the reagent and the stoichiometry of the reaction are known. Radio-release methods in which the radioactive reagent is released into the gas phase require rather more complicated instrumentation than that used in simple radio-reagent methods.

Radio-release methods may be classified according to the type of the radioactive reagent employed, i.e.,

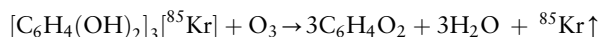
1. radioactive kryptonates;
2. radioactive metals; and
3. radioactive salts and other substances.

### Radioactive Kryptonates

The term radioactive kryptonates is used for substances into which atoms or ions of krypton-85 ( $^{85}\text{Kr}$ ) are incorporated (by diffusion of  $^{85}\text{Kr}$ , by bombardment with accelerated krypton ions, by introduction of  $^{85}\text{Kr}$  during sample preparation, by phase transition, etc.). The choice of technique for  $^{85}\text{Kr}$  introduction depends mainly on the character of the solid to be labeled and the reason for which the labeled solid is prepared.

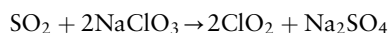
The basis of the use of radioactive kryptonates in chemical analysis is that during the chemical reaction the crystal lattice of the kryptonated carrier is destroyed, the carrier consumed, and the radioactive krypton released. The determination can be done with a calibration graph or by comparison with a standard.

In ozone determination with the radioactive clathrate hydroquinone kryptonate the following reaction is used:

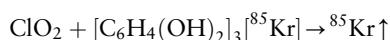


The logarithm of the activity of the released  $^{85}\text{Kr}$  is linearly proportional to the logarithm of the ozone concentration over a concentration range of  $10^{-10}$ – $10^{-6}$  g of ozone per liter of air.

Sulfur dioxide has also been determined by a method based on the mechanism of double release. In the first stage, sulfur dioxide reacts with sodium chlorate to release chlorine dioxide, which is a strong oxidizing agent. The chlorine dioxide then oxidizes radioactive hydroquinone kryptonate and gaseous  $^{85}\text{Kr}$  is released. The following reactions are involved:



and



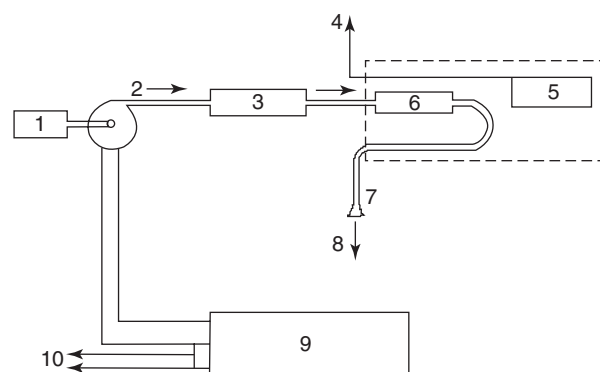
(The first equation is an oversimplification, since one molecule of sulfur dioxide releases between four and eight molecules of chlorine dioxide.) Much of the earlier work refers to use of sodium chlorite ( $\text{NaClO}_2$ ) but as the chlorine in this oxidation state is  $\text{Cl(III)}$ , it is difficult to see how it can be oxidized to  $\text{Cl(IV)}$  by  $\text{SO}_2$ .

### Radioactive Metals

Radioactive metals can be used in radio-release methods for the determination of oxidizing agents in aqueous solutions. The decisive factors in the choice of the metal are the following: the metal should not react with water, but should react with oxidizing agents to yield ions that do not form precipitates in aqueous media; the metal should have a radionuclide with suitable nuclear properties. These conditions are met by thallium ( $^{204}\text{Tl}$ ) and silver ( $^{110\text{m}}\text{Ag}$ ).

Dissolved oxygen in seawater, drinking water, and wastewater can be determined using metallic thallium labeled with  $^{204}\text{Tl}$ , which is quantitatively oxidized by oxygen. The laboratory equipment for oxygen determination consists of a glass column filled with labeled metallic thallium, which is connected to a flow-type Geiger-Müller (GM) counter. A block diagram of an analyzer used for oceanographic studies is shown in Figure 1.

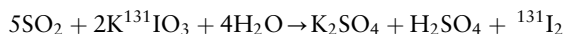
The vanadate ion is assayed by acidifying the sample ( $\text{pH} \sim 3$ ) and pouring it through a column containing radioactive metallic  $^{110\text{m}}\text{Ag}$ . The silver ions are released into the solution, where the concentration may be measured by a GM or a scintillation counter. The measurement of dichromate ion concentration in natural waters can be carried out similarly.



**Figure 1** Schematic diagram of an analyzer for oceanographic studies. 1 – input filter, 2 – pump, 3 – thallium column, 4 – coaxial cable to surface, 5 – electronics, 6 – GM counter, 7 – filter, 8 – discharge into sea, 9 – battery and transformer, 10 – control panel.

### Radioactive Salts and Other Radioactive Substances

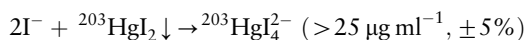
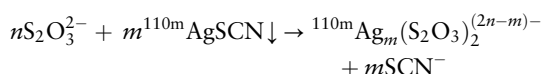
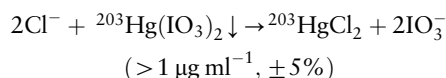
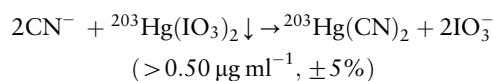
Sulfur dioxide is determined on the basis of the reaction:



This reaction takes place in an alkaline solution through which air containing sulfur dioxide is bubbled. After completion of the reaction, the solution is acidified and the iodine released is extracted.

Active hydrogen in organic substances may be determined by reaction with lithium aluminum hydride labeled with tritium ( ${}^3\text{H}$ ). The activity of released tritium is measured using a proportional counter.

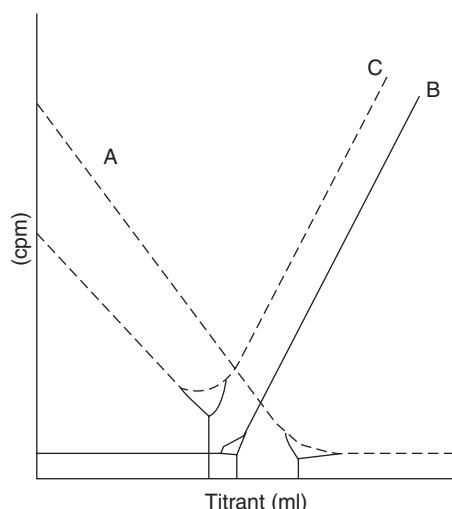
A number of determination are based on the reaction of a solution of an analyte that can form a soluble complex with a radioactively labeled precipitate. In this way it is possible to determine anions forming soluble complexes (e.g.,  $\text{CN}^-$ ,  $\text{S}_2\text{O}_3^{2-}$ ,  $\text{I}^-$ ,  $\text{F}^-$ ). The principle of the determination is given in the following reactions:



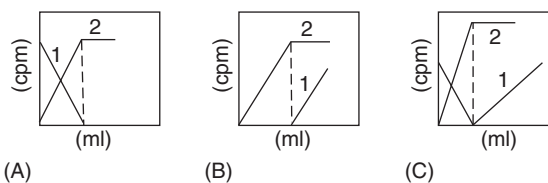
### Radiometric Titration

Radiometric titrations follow the relation between the radioactivity of one component or phase of the solution under analysis and the volume of added titrant. The compound formed during the titration must be easily separable from the excess of unreacted ions. This separation is directly ensured only in the case of precipitation reactions (Figure 2). In other types of reactions, the separation can be accomplished using an additional procedure. The endpoint is determined from the change in the activity of the residual solution or of the other phase.

According to the type of chemical reaction used, methods based on the formation of precipitates and methods based on complex formation can be distinguished (Figure 3). Because of the necessity for handling precipitates, precipitation radiometric titrations are difficult to apply to less than milligram amounts and, therefore, have no special advantages



**Figure 2** Radiometric precipitation titration curves. A – test solution active; B – test solution inactive, titrant active; C – both test solution and titrant active.



**Figure 3** Extraction radiometric direct titration curves. 1 – aqueous phase, 2 – organic phase, A – titration of radioactive test solution with inactive titrant, B – titration of inactive test solution with active titrant, C – titration of active solution with active titrant.

over other volumetric methods. The sensitivity of complexometric titrations is limited by the sensitivity of the determination of the endpoint. However, the use of radiometric detection can substantially increase the sensitivity of this type of determination. For the separation of the product from the initial component, liquid–liquid distribution, ion exchange, electrophoresis, or paper chromatography are most often used.

The application of radiometric titrations has declined over the past three decades. Their main advantage is where classical methods for detection of the endpoint are either impossible or subject to interference from the titration medium.

**See also:** **Immunoassays, Techniques:** Radioimmunoassays. **Isotope Dilution Analysis.** **Radiochemical Methods:** Radiotracers. **Titrimetry:** Overview. **Water Analysis:** Overview.

### Further Reading

Braun T and Tölgyessy J (1967) *Radiometric Titrations*. Oxford: Pergamon.

- Carden JE (1969) Radio-release. In: *Review with Special Emphasis on  $^{85}\text{Kr}$  Clathrates and Kryptonates*. ORNL-IIC-18, contract no.W-7405-eng.26.
- Garten Rainer PH and Tölgyessy J (2001) Radionuclides in analytical chemistry. In: *Ullmann's Encyclopedia of Industrial Chemistry*, 6th edn. Weinheim: Wiley-VCH.
- Kyrš M, Tölgyessy J, Bilimovich GN, *et al.* (1982) *Novye metody radioanaliticheskoi khimii*. Moscow: Energoizdat.
- Starý J (1986) In: Kolthoff IM Elving JM (eds.) *Treatise on Analytical Chemistry*, vol. 4, part 1, p. 241. New York: Wiley.
- Tölgyessy J and Bujdosó E (1991) *Handbook of Radioanalytical Chemistry*. vols I and II. Boca Raton: CRC Press.
- Tölgyessy J and Klehr EH (1987) *Nuclear Environmental Chemical Analysis*. Chichester: Ellis Horwood.

## Radioreceptor Assays

J W Ferkany, NovaScreen, Hanover, MD, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 4336–4342, © 1995, Elsevier Ltd.

### Introduction

Radioreceptor assays (RRAs) were introduced as an analytical tool in 1970. With the resurgence of interest in natural products, the development of techniques to generate random novel chemicals and advances in molecular biology, RRAs remain an essential tool for identifying novel substances, demonstrating the specificity of substances for target systems, tracking activities following fractionation of complex mixtures, and predicting the potential side-effects of new pharmaceuticals.

RRAs are (1) cost-effective, (2) offer a high degree of precision, (3) have a sensitivity comparable to other analytical techniques, (4) afford rapid turnaround times, (5) are amenable to evaluating large numbers of samples, and (6) do not require prior information regarding the structure of active compounds. RRAs offer the frequently unique advantage of providing information on the biological activity of new chemicals. The number of RRAs is now in the hundreds and is limited only by the availability of an appropriate ligand and a source of receptor.

This article focuses on the principles that underlie RRAs, the practical considerations of the method, and examples of specific applications.

### Methods and Techniques

#### General

RRAs are based on the specific and saturable binding of a radioligand (L) to a biological receptor or enzyme (R) at equilibrium. By 'specific' it is meant that L interacts with a single R. Saturability implies that at some concentration of L there is no further

increase in the amount of L bound to R since all R are occupied by a molecule of L. At equilibrium, the rate at which any single molecule of L dissociates from a molecule of R is equal to the rate that a molecule of L occupies another R.

These relationships are expressed in general terms by the mass action equation:



where [L] is the concentration of free ligand, [R] is the concentration of receptor, [LR] is the concentration of ligand–receptor complex, and  $k_1$  and  $k_2$  are, respectively, the association and dissociation rate constants of the ligand to the receptor.

Radioligands represent the radioactive forms of endogenous neurotransmitters, neuromodulators, hormones or enzyme substrates, or synthetic compounds that interact with the binding sites on receptors or enzymes. Ligands are made radioactive using a number of isotopes (e.g.,  $^3\text{H}$ ,  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{32}\text{P}$ , or  $^{14}\text{C}$ ) but in practice the most common method involves tritiation or iodination. To be useful a radioligand must have high affinity for the receptor or enzyme, be specific for the receptor, be easy to synthesize, and, most importantly, have similar biological, pharmacological, and functional properties as the parent compound. With few exceptions, ligands are currently obtained from commercial operations that specialize in the production of these materials.

For economic and environmental reasons, there is an emphasis on the development of nonradioactive ligands for use in RRAs. Emerging technologies rely on the substitution of a fluorescent moiety incorporated into the parent compound to track the binding of L to R. It is likely and desirable that these technologies will replace existing methods.

Historically, the source of receptors for use in RRAs has been animal tissues but the techniques of molecular biology are proving an increasingly important source of receptors or enzymes that are

- Carden JE (1969) Radio-release. In: *Review with Special Emphasis on  $^{85}\text{Kr}$  Clathrates and Kryptonates*. ORNL-IIC-18, contract no.W-7405-eng.26.
- Garten Rainer PH and Tölgyessy J (2001) Radionuclides in analytical chemistry. In: *Ullmann's Encyclopedia of Industrial Chemistry*, 6th edn. Weinheim: Wiley-VCH.
- Kyrš M, Tölgyessy J, Bilimovich GN, *et al.* (1982) *Novye metody radioanaliticheskoi khimii*. Moscow: Energoizdat.
- Starý J (1986) In: Kolthoff IM Elving JM (eds.) *Treatise on Analytical Chemistry*, vol. 4, part 1, p. 241. New York: Wiley.
- Tölgyessy J and Bujdosó E (1991) *Handbook of Radioanalytical Chemistry*. vols I and II. Boca Raton: CRC Press.
- Tölgyessy J and Klehr EH (1987) *Nuclear Environmental Chemical Analysis*. Chichester: Ellis Horwood.

## Radioreceptor Assays

J W Ferkany, NovaScreen, Hanover, MD, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 4336–4342, © 1995, Elsevier Ltd.

### Introduction

Radioreceptor assays (RRAs) were introduced as an analytical tool in 1970. With the resurgence of interest in natural products, the development of techniques to generate random novel chemicals and advances in molecular biology, RRAs remain an essential tool for identifying novel substances, demonstrating the specificity of substances for target systems, tracking activities following fractionation of complex mixtures, and predicting the potential side-effects of new pharmaceuticals.

RRAs are (1) cost-effective, (2) offer a high degree of precision, (3) have a sensitivity comparable to other analytical techniques, (4) afford rapid turnaround times, (5) are amenable to evaluating large numbers of samples, and (6) do not require prior information regarding the structure of active compounds. RRAs offer the frequently unique advantage of providing information on the biological activity of new chemicals. The number of RRAs is now in the hundreds and is limited only by the availability of an appropriate ligand and a source of receptor.

This article focuses on the principles that underlie RRAs, the practical considerations of the method, and examples of specific applications.

### Methods and Techniques

#### General

RRAs are based on the specific and saturable binding of a radioligand (L) to a biological receptor or enzyme (R) at equilibrium. By 'specific' it is meant that L interacts with a single R. Saturability implies that at some concentration of L there is no further

increase in the amount of L bound to R since all R are occupied by a molecule of L. At equilibrium, the rate at which any single molecule of L dissociates from a molecule of R is equal to the rate that a molecule of L occupies another R.

These relationships are expressed in general terms by the mass action equation:



where [L] is the concentration of free ligand, [R] is the concentration of receptor, [LR] is the concentration of ligand–receptor complex, and  $k_1$  and  $k_2$  are, respectively, the association and dissociation rate constants of the ligand to the receptor.

Radioligands represent the radioactive forms of endogenous neurotransmitters, neuromodulators, hormones or enzyme substrates, or synthetic compounds that interact with the binding sites on receptors or enzymes. Ligands are made radioactive using a number of isotopes (e.g.,  $^3\text{H}$ ,  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{32}\text{P}$ , or  $^{14}\text{C}$ ) but in practice the most common method involves tritiation or iodination. To be useful a radioligand must have high affinity for the receptor or enzyme, be specific for the receptor, be easy to synthesize, and, most importantly, have similar biological, pharmacological, and functional properties as the parent compound. With few exceptions, ligands are currently obtained from commercial operations that specialize in the production of these materials.

For economic and environmental reasons, there is an emphasis on the development of nonradioactive ligands for use in RRAs. Emerging technologies rely on the substitution of a fluorescent moiety incorporated into the parent compound to track the binding of L to R. It is likely and desirable that these technologies will replace existing methods.

Historically, the source of receptors for use in RRAs has been animal tissues but the techniques of molecular biology are proving an increasingly important source of receptors or enzymes that are



transiently or stably expressed. Cloned receptors offer the advantages of decreased reliance on animals, a high degree of receptor or enzyme expression, and homogeneity of receptor source. Disadvantages include a lack of knowledge of receptors (e.g., glycosylation patterns) cloned in prokaryote systems, leaving open the possibility that the clone has pharmacological characteristics different from the native receptor. Where receptors are composed of multiple, heterogeneous subunits (e.g., amino acid receptors), it is unwise to use cloned material until a better understanding of native receptor subunit composition is available.

## Binding Reactions: Theoretical and Practical Considerations

### Mathematics

The mathematics of ligand–receptor interactions are beyond the scope of the current article. Therefore, only important summary equations describing the bimolecular interaction of L with a single class of receptor are given below. These equations are analogous to the fundamental description of enzyme–substrate interactions first proposed by Michaelis and Menten, with the exception that L replaces the term for substrate, R replaces the enzyme, and LR substitutes for the enzyme–substrate complex. In binding reactions there is no term for product, since the binding of L to R does not result in the alteration of L.

Rearrangement and substitution of eqn [1] produces two summary statements important to the understanding of ligand–receptor interactions. These are respectively,

$$K_d = ([L][R]/[LR]) = k_2/k_1 \quad [2]$$

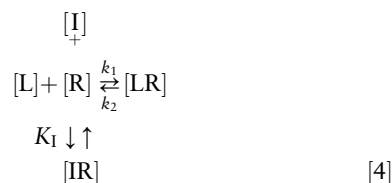
where  $K_d$  is the equilibrium dissociation constant for the reaction, and

$$[LR] = [L][R + LR]/K_d + [L] \quad [3]$$

Equation [3] is the general form of equation for rectangular hyperbola and a plot of [LR] versus [L] yields a curve with a horizontal asymptote of  $[R_{\text{total}}]$ . By definition, the  $K_d$  of the reaction is  $[R_{\text{total}}]/2$  and is the concentration of L necessary to half-maximally occupy the total receptor population.

Important to the interpretation of RRA data are the events occurring when fixed L and R are incubated in the presence of increasing concentrations of an unlabeled material (I) that prevents the binding of L to R. Experiments are generally referred to as ‘competition’ or ‘displacement’ protocols and

assume the general form shown in eqn [4]:



where [L], [R], and [LR] retain their former meaning, [I] represents the concentration of an unlabeled compound that binds to R, and [IR] is the concentration of the inhibitor–receptor complex. The term  $K_I$  is a constant representing the potency of I to bind to R.

Incubation of a fixed amount of L in the presence of increasing [I] leads to a progressive diminution of the amount of L bound to R. In practical terms, a decreasing amount of radioactivity is measured in the presence of successively higher amounts of I. Competition curves yield sigmoidal plots (Figure 1A) and data of this type are most often expressed by plotting the percentage of L specifically bound to R at each concentration of I. The two important equations derived from this interaction are the Cheng–Prusoff equation and a derivation of the pseudo-Hill plot. The Cheng–Prusoff equation is given by eqn [5]:

$$K_I = IC_{50}/(1 + [L]/K_d) \quad [5]$$

where  $IC_{50}$  is the concentration of I required to inhibit the binding of L to R by 50%.

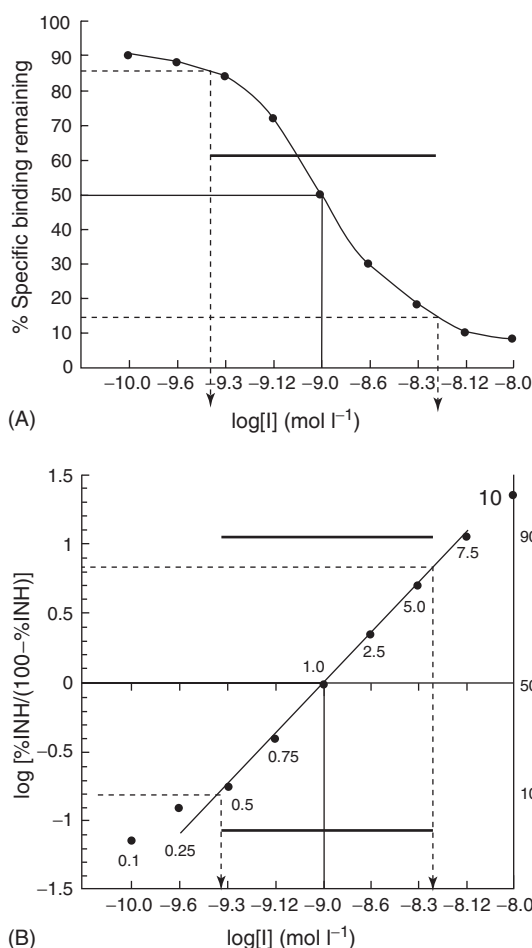
Equation [6] is a derivation of the pseudo-Hill plot:

$$\log \%INH/100 - \%INH = n \log \%INH + n \log IC_{50} \quad [6]$$

where %INH is the amount of inhibition of specifically bound L observed in the presence of a single concentration of I and  $n$  is a slope factor. Using this transformation, the data in Figure 1A have been replotted in Figure 1B and form the basis for quantitative measures in RRAs. These data are conveniently generated manually using log–logit paper or through the use of appropriate software.

### Selection of Ligand

There are a variety of ligands available to label biologically important receptors or enzymes (see Table 1). Many of these are highly specific for receptor subtypes, whereas others label a heterogeneous family of receptors. The choice of ligand will depend on the goals of the experiment. For example, it would be appropriate to select the generic ligand, [ $^3H$ ]naloxone, which labels all biologically relevant opiate receptors, if one were interested only in determining if a natural product extract contained



**Figure 1** Graphical representation of receptor binding data. (A) The curve is the standard representation of data resulting from experiments described by eqn [4]. Typically, fixed concentrations of L and R are incubated in the presence of increasing concentrations of an inhibitor, I, which competes for with the binding of L to R. As the concentration of I is increased, there is a progressive decrease in the amount of L bound to R. Dashed lines (and the bar at top) show the useful range of the assay; the bold line shows the concentration of I required to inhibit binding by 50%. (B) The data in (A) have been transformed using log-logit methods to yield a linear format useful for RRAs. Values associated with the line represent the absolute amount of I present in the assay assuming the concentrations shown and a final assay volume of 1 ml. Other notations are identical to those in (A).

opiates. On the other hand, if the object of the study were to identify selective  $\mu$  opiate antagonists, [<sup>3</sup>H]Tyr-D-Ala-Gly-NMe-Phe-Gly-ol (DAGO) would be more appropriate since this interacts with the desired receptor subtype. In every instance, however, ligands should label only a single family of receptors or enzymes.

### Linearity

Linearity is a correlate of receptor saturability. Intuitively, one would predict that as the concentration of

R added to incubations is increased, the amount of L specifically bound to the receptor would correspondingly increase as a linear function of [R]. Over a certain range of [R] linearity appears to be the case; beyond these limits, specific binding of L to R becomes nonlinear. In a manner analogous to enzyme reactions this occurs as a result of ligand availability becoming a driving force of the reaction. The interaction of L with R becomes 'substrate-dependent' and curves deviate from linearity.

For practical reasons it is important to determine the range of [R], which may be added to the RRA and still maintain linearity. Increasing [R] increases the amount of radioactivity measured in each incubation but can be a mistake since assays no longer adhere to the fundamental assumptions of ligand-receptor interactions.

### Equilibrium

Because binding assays follow mass action kinetics, a finite period of time is required for the reaction to reach equilibrium, the point at which the rate of association of L with R equals the rate of dissociation of the LR complex. To be a useful tool, RRA must always be performed under equilibrium conditions.

In an ideal world, it would be expected to be possible to predict the time required to reach equilibrium based on a knowledge of the affinity of R for L and the temperature at which the reactions are performed. Unfortunately, this is not the case and the time required to achieve equilibrium conditions must be determined in each laboratory for each assay. This is easily done by performing association experiments that determine the amount of L specifically bound to R as a function of time. In practice, most assays reach equilibrium within 1–2 h.

### Pharmacology

Beyond the characteristics described above (linearity, equilibrium conditions, saturability) an RRA must be specific. That is, the method must measure only those compounds or endogenous substances that are known to inhibit the receptor-ligand reaction in a competitive manner. By competitive it is implied that substances inhibit the binding of L to R in a concentration-dependent manner by binding to the same site on R that would otherwise be occupied by L.

In developing an RRA, the specificity of the system is shown by demonstrating that reference agents, known to interact with R, inhibit the binding of L. With the increasing reliance on cloned and expressed receptor systems, this latter step is crucial since expression of receptors in bacterial or mammalian

**Table 1** Some binding assays and ligands useful for RRAs

<i>RRA target site</i>	<i>Useful ligand(s)</i>
<i>Neurotransmitter receptors</i>	
Adenosine (nonselective)	<sup>3</sup> H-NECA
Adenosine A <sub>1</sub>	<sup>3</sup> H-CPX
Adenosine A <sub>2</sub>	<sup>3</sup> H-CGS 21680
Purinergic P <sub>2Y</sub>	<sup>35</sup> S-ADPS
Adrenergic <sub>1</sub> (nonselective)	<sup>3</sup> H-Prazosin
Adrenergic <sub>1A</sub>	<sup>3</sup> H-Prazosin
Adrenergic <sub>1B</sub>	<sup>3</sup> H-Prazosin
Adrenergic <sub>2</sub> (nonselective)	<sup>3</sup> H-RX 781094
Adrenergic <sub>2A</sub>	<sup>3</sup> H-RX 781094
Adrenergic <sub>2B</sub>	<sup>3</sup> H-RX 781094
Adrenergic (nonselective)	<sup>125</sup> I-Iodo-Pindolol
Adrenergic <sub>1</sub>	<sup>125</sup> I-Iodo-Pindolol (+ ICI 89,406)
Adrenergic <sub>2</sub>	<sup>125</sup> I-Iodo-Pindolol (+ ICI 118,551)
Dopamine (nonselective)	<sup>3</sup> H-Spiperone
Dopamine D <sub>1</sub>	<sup>3</sup> H-SCH 23390
Dopamine D <sub>2</sub>	<sup>3</sup> H-Sulpiride
Dopamine D <sub>4</sub>	<sup>3</sup> H-Clozapine
GABA <sub>A</sub>	<sup>3</sup> H-GABA
GABA <sub>B</sub>	<sup>3</sup> H-GABA (+ isoguvacine)
Histamine H <sub>1</sub>	<sup>3</sup> H-Pyramilamine
Histamine H <sub>2</sub>	<sup>3</sup> H-Tiotidine
Histamine H <sub>3</sub>	<sup>3</sup> H-Methylhistamine
Serotonin (nonselective)	<sup>3</sup> H-LSD
Serotonin 5-HT <sub>1</sub>	<sup>3</sup> H-5-HT
Serotonin 5-HT <sub>1A</sub>	<sup>3</sup> H-8-OH-DPAT
Serotonin 5-HT <sub>1B</sub>	<sup>125</sup> I-ICYP
Serotonin 5-HT <sub>1C</sub>	<sup>3</sup> H-Mesulergine
Serotonin 5-HT <sub>2</sub>	<sup>3</sup> H-Ketanserin
Serotonin 5-HT <sub>3</sub>	<sup>3</sup> H-GR 65630
Muscarinic (central, nonselective)	<sup>3</sup> H-QNB
Muscarinic (peripheral, nonselective)	<sup>3</sup> H-QNB
Muscarinic M <sub>1</sub>	<sup>3</sup> H-Pirenzepine
Muscarinic M <sub>2</sub>	<sup>3</sup> H-AFDX
Muscarinic M <sub>3</sub>	<sup>3</sup> H-N-CH <sub>3</sub> -scopolamine
Nicotinic	<sup>3</sup> H-NMCI
NMDA	<sup>3</sup> H-CGS 19755
Kainate	<sup>3</sup> H-Kainate
AMPA	<sup>3</sup> H-AMPA
Glutamate	<sup>3</sup> H-Glutamate
<i>Growth factors and peptides</i>	
Atrial natriuretic factor (ANF <sub>1</sub> and ANF <sub>2</sub> )	<sup>125</sup> I-ANF
Basic fibroblast growth factor	<sup>125</sup> I-BFGF
Corticotropin-releasing factor	<sup>125</sup> I-CRF
Epidermal growth factor	<sup>125</sup> I-EGF
Insulin	<sup>125</sup> I-Insulin
Nerve growth factor	<sup>125</sup> I-NGF
Oxytocin	<sup>3</sup> H-Oxytocin
Platelet-activating factor	<sup>3</sup> H-PAF
Platelet-derived growth factor	<sup>125</sup> I-PDGF
Thyrotropin-releasing hormone	<sup>3</sup> H-TRH
Tumor necrosis factor	<sup>125</sup> I-TNF
<i>Immunological factors</i>	
Complement C5a	<sup>125</sup> I-C5a
gp120/CD <sub>4</sub> capture	<sup>125</sup> I-gp120
Interleukin IL <sub>1</sub> , α	<sup>125</sup> I-IL <sub>1</sub> α
<i>Opioids</i>	
Opiate (nonselective)	<sup>3</sup> H-Naloxone
Opiate (μ)	<sup>3</sup> H-DAGO

**Table 1** Continued

<i>RRA target site</i>	<i>Useful ligand(s)</i>
Opiate (δ)	<sup>3</sup> H-DPDPE
Opiate (κ)	<sup>3</sup> H-U 69593
<i>Ion channels</i>	
Calcium channel (type T and type L)	<sup>3</sup> H-Nitrendipine
Calcium channel (type N)	<sup>125</sup> I-ω-Conotoxin
Chloride	<sup>3</sup> H-TBOB
Potassium (ATP-modulated)	<sup>3</sup> H-Glibenclamide
Potassium (low conductance Ca <sup>2+</sup> -activated)	<sup>125</sup> I-Apamin
Potassium (voltage dependent)	<sup>125</sup> I-Charybdotoxin
Sodium (site 1)	<sup>3</sup> H-Saxitoxin
Sodium (site 2)	<sup>3</sup> H-Batrachotoxin A
<i>Uptake sites</i>	
Adenosine	<sup>3</sup> H-NBTI
Choline	<sup>3</sup> H-Hemicholinium-3
Dopamine (cocaine site)	<sup>3</sup> H-WIN 35,428
GABA	<sup>3</sup> H-GABA
Norepinephrine	<sup>3</sup> H-DMI
Serotonin	<sup>3</sup> H-Citalopram
<i>Brain/gut peptides</i>	
Angiotensin II (type 1 and type 2)	<sup>125</sup> I-Angiotensin
Arg-Vasopressin (AVP <sub>1</sub> )	<sup>3</sup> H-AVP <sub>1</sub>
Arg-Vasopressin (AVP <sub>2</sub> )	<sup>3</sup> H-Phenylalanyl-3,4,5-AVP <sub>2</sub>
Bombesin	<sup>125</sup> I-GRF
Bradykinin	<sup>3</sup> H-Bradykinin
CGRP (type 1 and type 2)	<sup>125</sup> I-CGRP
Cholecystokinin, CCKA, and CCKB	<sup>125</sup> I-CCK
Endothelin (ET-A and ET-B)	<sup>125</sup> I-Endothelin
Neurokinin-NK <sub>1</sub> (substance P)	<sup>3</sup> H-Substance P
Neurokinin-NK <sub>2</sub> (NKA)	<sup>125</sup> I-NKA
Neurokinin-NK <sub>3</sub> (NKB)	<sup>125</sup> I-Eledoisin
Neuropeptide Y	<sup>125</sup> I-NPY
Neurotensin	<sup>3</sup> H-Neurotensin
Somatostatin	<sup>125</sup> I-Somatostatin
Vasoactive intestinal peptide	<sup>125</sup> I-VIP
<i>Prostaglandins</i>	
Leukotriene B <sub>4</sub>	<sup>3</sup> H-LTB <sub>4</sub>
Leukotriene D <sub>4</sub>	<sup>3</sup> H-LTD <sub>4</sub>
Thromboxane A <sub>2</sub>	<sup>3</sup> H-U46619
<i>Second messengers</i>	
Adenylate cyclase	<sup>3</sup> H-Forskolin
Protein kinase C	<sup>3</sup> H-PDBU
Inositol triphosphate	<sup>3</sup> H-IP <sub>3</sub>
<i>Steroids receptors</i>	
Oestradiol	<sup>125</sup> I-Oestradiol
Progesterone	<sup>3</sup> H-Promegestone
Testosterone	<sup>3</sup> H-R1881
<i>Regulatory sites</i>	
Benzodiazepine (central)	<sup>3</sup> H-Flunitrazepam
Benzodiazepine (peripheral)	<sup>3</sup> H-PK 11195
Glycine (strychnine insensitive)	<sup>3</sup> H-5,7-DCKA
Glycine (strychnine sensitive)	<sup>3</sup> H-Strychnine
Phencyclidine	<sup>3</sup> H-TCP
MK 801	<sup>3</sup> H-MK 801
Sigma	<sup>3</sup> H-DTG

cell-based systems may lead to differences in the pharmacological specificity of the clone.

Final validation requires the demonstration that RRAs provide similar or identical quantitative data to those generated using other analytical methods. The latter is particularly important when using RRAs to measure substances endogenous to biological samples.

## Practical Aspects of RRA

### Equipment

The equipment necessary to perform RRAs is generally available to any laboratory engaged in biomedical research. Common items include the obvious (balances, pH meters, tissue homogenizers, refrigerators/freezers, etc.) and most assays require a high-speed centrifuge capable of generating 50 000g. Because RRAs remain based on the use of radiolabeled ligands this necessitates access to  $\beta$ - and  $\gamma$ -counters. Access to animals and animal housing facilities are needed, although if cloned and expressed receptors represent the tissue source, tissue culture facilities must be available. Due to the resurgence of the popularity of RRAs, several commercial operations now offer stable freeze-dried receptor preparations for use.

Due to the availability of high-affinity ligands for many receptors, reactions are often terminated using vacuum filtration. In this procedure, the ligand that is not bound to the receptor is separated from the receptor–ligand complex by trapping the complex on a filter support. Commercial units are available for this purpose with the most popular capable of terminating 48 individual reactions simultaneously and in less than 1 min.

### Tissue Preparation

The identification of a tissue containing the receptor of interest or the transient or stable cloning and expression in cells is the first step in developing a reliable RRA. While the brain–spinal cord axis represents a major source of receptors, assays based on receptors specific to tissues including gut, platelets, lymphocytes, skeletal muscle, and endocrine organs are well known. Transformed cells are also useful as the point of departure to gather receptors; these may be obtained from public sources such as the American Type Culture Collection (Rockville, MD, USA). Cloned and expressed receptors may be obtained on an individual basis as needed.

Realization of a final receptor preparation depends on the starting material and the specific RRA. In general, preparative procedures are designed to concentrate receptors in a crude manner and remove endogenous substances interfering with the

sensitivity of the assay. Frequently, preparation involves homogenization of tissue(s) followed by low-speed differential centrifugation to remove cellular debris. Higher-speed centrifugation is performed to concentrate membrane fractions rich in the relevant receptor. The final material is ‘washed’ on several occasions using sequential resuspension and centrifugation in large amounts of buffers or water in order to remove endogenous substances that may interfere with the assay.

As a cautionary note, very few assays require tissue(s) obtained from human or potentially pathogenic animal sources (e.g., sheep brain). In these instances, equipment should be dedicated to tissue preparation, appropriate precautions should be taken during the preparation procedure, and equipment should be disinfected after use.

Cloned receptors offer the advantages of high expression, homogeneity, control of pathogenic factors, and the potential to eliminate endogenous substances that may interfere with assay sensitivity. When possible, the use of cloned receptors is desirable.

In some instances it is possible to prepare tissue in batches and store the preparation frozen for future use. This will vary for each assay and some receptors are unstable to freezing mandating the preparation of tissue on the day of assay.

### Sample Preparation

Preparation of samples for use in RRAs varies. If the substance of interest is in biological fluids (e.g., urine, spinal fluid, plasma) it may be possible to add fluid directly to the RRA. In many instances, however, components (salts, proteins) interfere with assay integrity. In this case, it is necessary to devise a preparation scheme suitable for use with the RRA and which does not alter the quantities of the substance of interest, or to incorporate an appropriate internal standard in the method (e.g., radioactive tracer amounts of the compound of interest preferably labeled with an isotope different from that incorporated in the ligand). In the case of natural product analysis, standard aqueous or organic extraction procedures are followed and the sample is freeze-dried or dried to provide the final preparation. Aqueous fractions can be redissolved in assay buffers whereas dimethylsulfoxide (DMSO) is a useful solvent for organic extracts since many RRAs tolerate  $\leq 1\%$  DMSO in the final assay.

### Assay Procedures

The elegance of the RRA is its simplicity, sensitivity, precision, high volume, and rapid turnaround time. In a typical day, as many as 150 samples may be

evaluated in any single procedure. Data are of the type shown in **Figures 1A** and **1B** are usually available the following day.

For assay, tissue preparations containing the receptor of interest are resuspended in buffer and an aliquot is added to tubes containing the radioligand and a sample containing the substance to be measured. In some tubes, solutions containing various concentrations of a reference agent are substituted for the sample. If the substance being measured is known (e.g., a drug or neurotransmitter), it is best to use the same compound to generate the reference curve in order to provide absolute values in the final analysis. When the identity of the substance of interest is not known (e.g., natural products) any appropriate reference substance may be used to generate standard curves and the final data are expressed in arbitrary units relative to the curve, or in terms of the amount of sample (e.g.,  $\mu\text{l}$  or  $\mu\text{g ml}^{-1}$ ) added to the assay.

Final assay volumes are 0.5–1.0 ml. Incubations are generally performed using triplicate tubes, each of which contains identical amounts of ligand, tissue, sample, or reference agent. In the case of sample tubes, it is useful to include at least two, and better yet, three different amounts of sample in triplicate in order to ensure the final result falls in the range of the reference curve.

Following an incubation at the time and temperature appropriate to each assay, reactions are terminated in order to separate L not bound to the R. This is done using one of two methods, centrifugation or filtration. In the former, the incubation mixtures are centrifuged at high speed (e.g., 50 000g; 10 min) to pellet the tissue. The incubation solution is decanted and the pellet is washed rapidly with 5–10 ml of ice-cold assay buffer. The pellet is dissolved in tissue solubilizers, scintillant is added, and the radioactivity contained in the pellet is counted using scintillation spectrophotometry.

In filtration assays, L is separated from the receptor–ligand complex by rapidly filtering the incubation mixtures through an appropriate support. Most assays call for the use of glass fiber filters. The filters are rapidly rinsed with 5–10 ml of ice-cold assay buffer, placed in counting tubes and radioactivity is quantified as described above.

From a practical perspective, filtration assays are generally preferred over centrifugation methods. Equipment is commercially available to filter simultaneously up to 48 samples in under a minute, making it possible to process 300–400 assay tubes in a single day. Assays can be performed in small volumes, reducing costs while substantially increasing assay sensitivity. Finally, emerging technologies

include the development of filter supports presoaked with scintillant, which allow direct assay of filters in the absence of scintillation fluor. When centrifugation assays must be used, commercial equipment and supplies make it possible to perform the entire assay in centrifugation tubes that double as scintillation vials.

### Cautionary Notes

As with any technique, RRAs are subject to limitations and cautions. Assays involving samples obtained from human tissues must be presumed to contain pathogens unless these have been specifically excluded by appropriate testing. In the absence of the latter, equipment used in the assay must be properly disinfected after each use and only individuals trained in correct safety methods should perform the tests.

At the practical level, it should be borne in mind that RRAs can be subject to a number of confounding factors. Chief among these is the possibility that substances contained in samples may interfere with assay specificity. Samples processed with acids or bases, if not properly neutralized, may yield false-positive data since RRAs are pH dependent. It would be ill advised to add organic extracts directly to an RRA since assays are performed in aqueous buffers that are not miscible with all organic solvents.

Receptor preparations often contain enzymes (proteases, esterases, etc.) that degrade compounds of interest. Assays designed to detect peptides or esters should include peptidase or esterase inhibitors and bovine serum albumin to minimize enzyme activity. Whenever enzyme inhibitors are added to assays, the effect of these additions on the characteristics of the assay must be determined to ensure that these additions do not inhibit (or enhance) ligand binding. Where interference is encountered, this is controlled by the addition of the interfering substance to all control incubations as well as in generating the reference curve.

At the practical level, there is always an urge to maximize the signal obtained in the final data output. For RRAs, this means maximizing the amount of radioactivity detected. As intuitively expected radioactivity can be increased by increasing [R], or by increasing [L] (or both) added to assay tubes. The detrimental effects of increasing [R] beyond certain limits have been described above. Increasing [L], while perhaps increasing assay precision, results in a decrease in the sensitivity of the method.

Consideration of the Cheng–Prusoff equation (eqn [5]) provides the rationale. Since  $K_d$  and  $K_1$

are constants,  $[L]$  determines the amount of  $I$  that must be present to achieve an  $IC_{50}$ . When  $[L]$  is increased the ratio  $[L]/[k_d]$  increases and the amount of  $I$  required to inhibit binding is increased. It is always desirable to use the least amount of  $L$  that provides reliable data and an acceptable error limit. A general rule of thumb is to maintain  $L$  at a concentration  $\ll$  (10-fold) the  $K_d$  of  $L$  for  $R$ .

Data analysis begins with the determination of the amount of  $L$  specifically bound to  $R$ . This is the difference between the total and nonspecific binding. The nonspecific binding is then subtracted and for each concentration of  $I$  or sample, the specifically bound decompositions per minute (dpm) are divided by the value for the total specific binding of  $L$ . This result is multiplied by 100 to yield the per cent inhibition of total binding at each concentration of  $I$  or sample.

As shown in **Figure 1A**, these data are plotted on semilog paper to yield the sigmoidal curve characteristic of binding reactions. To simplify data analysis, a linear transformation as shown in **Figure 1B** is useful. Alternatively, the log concentration of  $I$  can be directly plotted against the per cent inhibition using log-logit paper (or appropriate software packages).

As a final step, it is necessary to determine the amount of inhibitor present in experimental samples. This is accomplished by comparing the per cent inhibition of the specific binding of  $L$  caused by each sample to the reference curve and extrapolation to the  $x$ -axis of **Figure 1B**.

If the identity of the substance of interest is known, extrapolation to the  $x$ -axis yields the precise concentration of substance in the sample. As shown in **Figure 1B**, the amount of reference agent added to each tube of the standard curve can be expressed as either a concentration or an absolute amount. If the identity of the substance is unknown, as may be the case when screening for novel chemicals in natural product extracts, an arbitrary value is assigned to describe compound activity.

Like other analytical techniques, RRAs are associated with a degree of variability and it is not uncommon for results to differ by 10–15% among assays. Although sound opinion exists to the contrary, it is recommended that samples producing  $<15\%$  inhibition or  $>85\%$  inhibition of the specific binding be excluded from data analysis, or re-examined on a second occasion. In log-logit analysis, an infinite error is associated with samples producing  $>90\%$  or  $<10\%$  inhibition of binding if the assay has a day-to-day variance of 10%.

To be useful, an RRA must be sensitive and precise. That is, the assay must be capable of accurately detecting minute amounts of important substances in samples. The affinity ( $K_d$ ) of  $L$  for  $R$  is important since high-affinity ligands, in general, produce more robust assays. In practice, however, ultrahigh-affinity ligands present the investigator with technical difficulties. The mathematics of binding assume that the concentration of  $L$  added to the reaction equals the concentration of labeled  $L$  at equilibrium ( $>95\%$  of added  $L$  remains unbound to  $R$ ), and that the concentration of  $R$  is much less than the  $K_d$  of  $L$  for  $R$ . As the affinity of  $L$  decreases more  $L$  is bound to  $R$  at the same time that  $R$  is decreased. This combination of factors results in a progressive diminution of radioactivity measured in the final output. As measured radioactivity decreases, the error increases, resulting in a decrease in the precision of the RRA. For these reasons, the choice of ligand for an RRA must balance the goals of the experiment, the sensitivity which is required, and the practical challenges of assay precision.

The sensitivity of RRAs is optimized by decreasing assay volume or the selection of an  $L$  for which the  $IC_{50}$  of  $I$  of interest competing with the binding of  $L$  to  $R$  is decreased (made smaller). The effects of the latter are evident from consideration of eqn [4].

The sensitivity of RRAs can also be increased by minimizing the final assay volume. When  $[L]$  and  $[R]$  are decreased in a linear fashion relative to assay volume, the final amount (not concentration) of experimental samples  $f$  required to achieve an  $IC_{50}$  is decreased in parallel.

**See also:** **Bioassays:** Overview. **Clinical Analysis:** Sample Handling. **Quality Assurance:** Production of Reference Materials.

## Further Reading

- Enna SJ (1980) *Physio-chemical Methodologies in Psychiatric Research*, pp. 83–101. New York: Raven Press.
- Ferkany JW (1985) *Neuromethods*, vol. 3, pp. 117–153. Clifton, NJ: Humana Press.
- Ferkany JW and Enna SJ (1982) *Analysis of Biogenic Amines*, pp. 267–284. Amsterdam: Elsevier.
- Lefkowitz F, Roth J, and Pastan I (1970) Radioreceptor assay of adrenocorticotrophic hormone: New approach to assay of polypeptide hormones in plasma. *Science* 170: 633–635.
- Limbird LE (1986) *Cell Surface Receptors: A Short Course on Theory and Methods*. Boston: Martinus Nijhoff.
- Yamamura HI, Enna SJ, and Kuhar MJ (1990) *Neurotransmitter Receptor Binding*. New York: Raven Press.



## Gamma-Ray Spectrometry

**K Buchtela**, Atominstitut der Oesterreichischen Universitaeten, Vienna, Austria

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Presently, gamma-ray spectrometry is the most powerful tool in the field of radionuclide analysis. It is used for the qualitative and quantitative determination of radionuclides that emit gamma radiation. Most of the radionuclides send out gamma radiation during their transformations to stable decay products.

Most nuclear reactions are accompanied by the emission of gamma radiation and therefore this tool is used to observe such reactions or to analyze reaction products.

A gamma-ray spectrometer system consists of a detector, an electronic circuit for pulse handling, and a facility for data storage, data processing, and data display.

The detector is a tool that provides electric signals (pulses). These signals depend on the energy (wavelength) and the number of the gamma photons to be detected. The detector is connected to a system that handles these pulses. The pulse-handling system is an electronic circuit that shapes the pulses, counts them, and classifies them according to their pulse height. Thus, the electronic signals are converted to a pulse spectrum. This pulse spectrum represents the number and the distribution of the energy of the gamma photons. The spectrum is stored in a memory system and can be processed further by the use of appropriate computer programs. The results are shown as a graphical or numerical display.

The advantages of gamma-ray spectrometry are:

- Sample preparation for the analysis is extremely simple and in most cases a radiochemical separation procedure for analysis is not necessary.
- It is very useful for environmental radioassay and for the analysis of low amounts of radioactive materials.
- The data obtained by the detection system are extremely suitable for data processing by a computer and therefore appropriate for the design of automatic analysis systems.
- Gamma-ray spectroscopic devices, although being rather expensive, are available in the market with very high standards of performance and quality assurance.

The interpretation of gamma spectra is sometimes difficult due to the interaction of gamma radiation

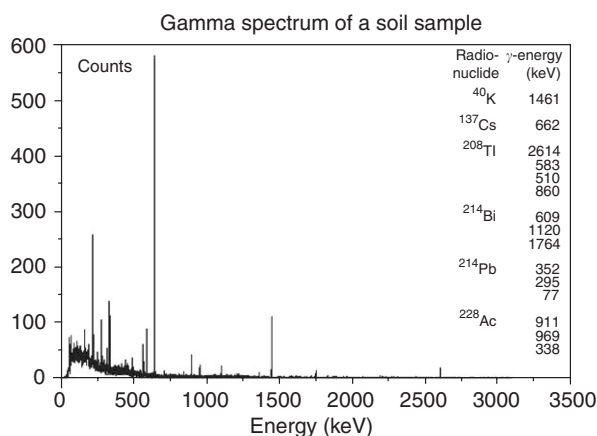
with the detector material, wherein additional gamma peaks appear in the spectrum. Therefore, some knowledge of the interactions of gamma radiation with matter is essential to understand the complex structure of the gamma spectra.

### Emission of Gamma Radiation

After an alpha or beta decay, the atomic nucleus is frequently in the so-called excited state, which means that the atomic nucleus still contains lots of excess energy. The transition from a higher to a lower energy state results in the dissipation of electromagnetic radiation of a very short wavelength of gamma radiation. The excited atomic nucleus either falls directly in one step to the ground state or descends in several steps to lower energy states and finally to the ground state. The energy dissipation is characteristic for the radionuclide under consideration and its spectrometric determination can be used for qualitative and quantitative analyses of radionuclides (see **Figure 1**). Gamma-ray spectrometry, therefore, is a specific and powerful tool for the analysis of gamma-emitting radionuclides.

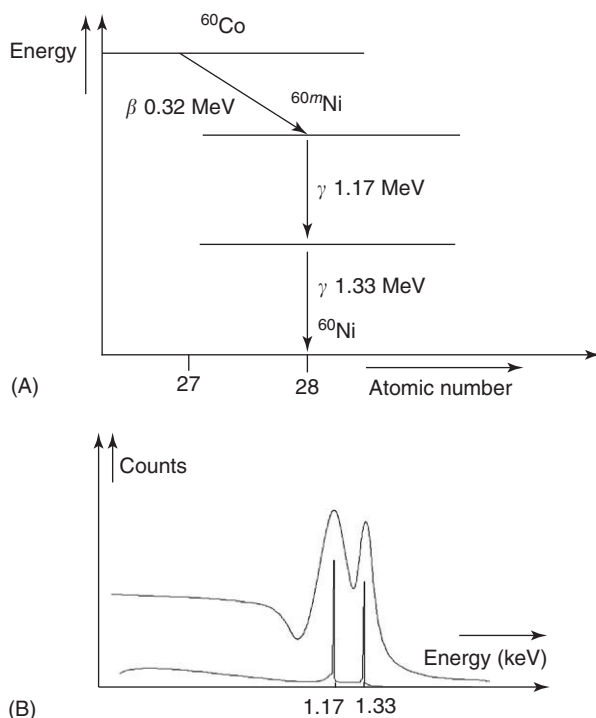
### Decay Schemes

The decay scheme of a radionuclide provides very useful information for its gamma-ray spectrometric determination.



**Figure 1** Gamma spectrum of a soil sample. Natural radionuclides from the uranium and thorium series are detected. The characteristic gamma line of cesium-137 at 662 keV is also displayed. The cesium-137 is due to the nuclear weapons tests and the Chernobyl reactor accident. The sample was taken at Vienna (Austria).

**Figure 2A** shows the decay scheme of cobalt-60, a radionuclide that is frequently used as a radiation source in industrial nuclear applications. Because of the unstable configuration of the atomic nucleus of cobalt-60, a neutron is changed to a proton and nickel-60 is obtained. This nuclear transformation, the change of a neutron to a proton, is accompanied by the emission of a particle with negative electric charge, namely an electron (a beta particle). The new nucleus of nickel shows already a stable configuration considering the amount of protons and neutrons but it contains lots of excess energy that will be dissipated by the emission of electromagnetic radiation due to gamma radiation. The excited nickel nucleus descends in two steps to the ground state. In the first step a gamma radiation of 1.17 MeV ( $1 \text{ MeV} = 1000 \text{ keV} = 1.602 \times 10^{-13} \text{ J}$ ) and in the next step a radiation of 1.33 MeV is released. Therefore, a gamma spectrum of cobalt-60 will show two characteristic lines, the so-called gamma lines as shown in



**Figure 2** (A) Decay scheme of cobalt-60. The decay scheme shows the transition from a high to a lower energy state and the alteration of the composition of the atomic nucleus. The atomic number (number of protons in the atomic nucleus) for cobalt changes to the atomic number of nickel due to the emission of a negatively charged beta particle and neutron is transformed to a proton. (B) Gamma spectra of cobalt-60. The spectra show the characteristic gamma lines of cobalt-60 (1.17 and 1.33 MeV). The two spectra are produced by two different radiation detector types. Broad peaks (poor resolution) are due to a scintillation detector (sodium iodide) and narrow peaks (good resolution) are due to a semiconductor detector (germanium of high purity).

**Figure 2B.** Such a gamma spectrum is a clear evidence for the presence of cobalt-60. A quantitative determination is done by measuring the intensity of the gamma lines, the area under the peaks.

In a gamma spectrum some other lines can also be found that are not directly related to the decay process but are due to the interaction of gamma radiation with matter. This interaction is a rather complex process. Some knowledge of the interaction will help to understand the structure of a gamma spectrum.

## Interactions of Gamma-Rays with Matter

Gamma spectra are more complex than is expected just from the energies of a gamma transformation of a radionuclide. Therefore, the correct interpretation of the features of a gamma-ray spectrum as recorded with a gamma-ray spectrometer is an important skill. To understand this, some knowledge regarding the interactions of gamma-rays with matter is necessary. There are three major modes of interaction of gamma-rays with matter: photoelectric effect, Compton scattering, and pair production.

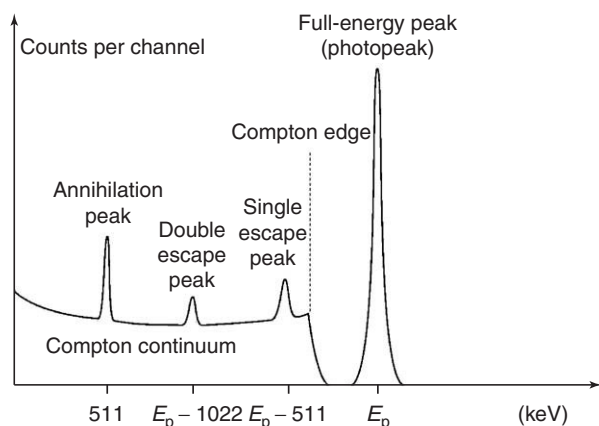
### Photoelectric Effect

In the photoelectric effect, the gamma-ray interacts with an electron of the inner shell of an atom of a radiation-absorbing material by transferring its energy totally to that electron. This results in the complete disappearance of the gamma photon and the ejection of an electron with energy equal to the energy of the initial gamma photon minus the binding energy of the electron. This binding energy is much smaller compared to the energy of the interacting gamma photon. The ejected electron can then induce events in a detector material. The electron gap in the inner shell is filled up by an electron of another shell and thus an X-ray photon is emitted.

The probability of the occurrence of photoelectric effect is related to the atomic number of the radiation-absorbing material and inversely related to the energy of the gamma-ray. Most frequently photoelectric absorption occurs with absorbing material of high atomic number and with low energies of the gamma-rays. The peak in the gamma-ray spectrum produced by detecting photoelectrons is called a full-energy peak because it provides immediate information about the energy of the gamma radiation (see Figure 3).

### Compton Scattering

A gamma photon may also interact with an electron of an atomic shell in such a way that it is scattered



**Figure 3** Schematic picture of a gamma spectrum of a radionuclide emitting gamma photons of a high single energy. The spectrum shows the photo-peak (or full-energy peak), corresponding to the energy of the characteristic gamma line of the radionuclide. By Compton scattering electrons with a wide energy distribution are produced (Compton continuum). By pair production and subsequent annihilation of the positron–electron pair a peak of 0.511 MeV (annihilation peak) is produced. If the gamma photons of 0.511 MeV escape from the detector without interaction a single and double escape peak appear in the spectrum.

and loses only part of its energy. The gamma-ray, now with less energy, is deflected from its initial path and the electron is ejected and may induce events in a detector material like a photoelectron, but these effects will be smaller compared with the total energy of the initial gamma-ray. If the scattered gamma-ray will produce a photoelectron within the detector material again a full-energy peak will be shown in the spectrum because the full energy of the gamma photon is dissipated in the detector material. But it is also possible that the scattered gamma-ray will escape from the detector without being detected. Then, only the dissipated energy of the scattered electron (Compton electron) contributes to the gamma spectrum. Because of the various scattering angles the Compton electrons show a distribution of energies over a wide range. This is called the Compton continuum. There is some gap between the full-energy peak and the Compton continuum (Compton edge or shoulder) corresponding to the minimum energy that is taken away by a back-scattered gamma photon (see Figure 3).

Compton scatter is most likely to occur in materials with high atomic number and for gamma-rays with energies from 0.6 to 4.0 MeV.

### Pair Production

Pair production occurs only for gamma-rays of high energy. By that effect the gamma-ray is transformed to matter in the form of a pair of negatively and

positively charged electrons (negatron and positron). Because an electron has a rest mass equivalent to 0.511 MeV of energy, a minimum gamma-energy of 1.02 MeV is required for this pair production. Any excess energy of the pair-producing gamma-ray is given to the electron–positron pair as kinetic energy. Most probably the positron will undergo annihilation by reaction with an electron in the detector material and by that two gamma photons of 0.511 MeV each will be created.

## Gamma Spectrum

Several additional peaks will appear in a gamma spectrum created by a high-energy gamma radiation. For a high-energy gamma-ray having produced an electron–positron pair, a full-energy peak will be displayed if all the energies of the electron–positron pair including the annihilation photons are dissipated in the detector material. But one of the annihilation photons or both of them may escape from the detector taking away 0.511 or 1.02 MeV of energy. As a consequence peaks related to lower energies than the full-energy peak will appear in the gamma spectrum: a single (or first) escape peak (full-energy peak minus 0.511 MeV) and a double (or second) escape peak (full-energy peak minus 1.02 MeV) are displayed in the spectrum.

Figure 3 is a schematic picture of a gamma spectrum related to a high-energy gamma radiation. Additional low-energy peaks caused by X-rays that are produced by fluorescence interactions with the shielding material of the detector may also appear. Lead X-rays occur in the regions from 74 to 88 keV.

It is possible that two gamma photons from the radioactive material are entering the detector simultaneously and are recorded as a gamma radiation with an energy equal to the sum of the two gamma energies. The resulting spectral feature would be a peak at energy equal to the sum of the individual incident gamma photons (sum peak). This phenomenon is observed at high radiation intensity and with small distances between radiation source and detector. A sum peak may be distinguished from a real full-energy peak by increasing the distance between source and detector. A sum peak will disappear but not a real full-energy peak.

## Detectors for Gamma Radiation

### Scintillation Detectors

Scintillation detectors emit photons of ultraviolet and visible light due to interaction of radiation emitted by radioactive materials.

Solid scintillation had been used earlier in nuclear science for the detection of radiation. The phenomenon of solid scintillation was discovered by Sir William Crookes in 1903. He exposed a layer of zinc sulfide to a sample emitting alpha particles and observed flashes of light. This effect later on was used for the assay of alpha radiation by visible counting of the light flashes. Many other scintillating crystalline substances useful for the detection of alpha particles, beta particles, gamma-rays, and even neutrons were discovered and applied. The application of solid scintillation for radiation detection and measurement can be found in most fields of nuclear science, and, of course, also in nuclear analytical techniques. For the measurement of gamma radiation, solid scintillation crystals of substances with high atomic number are used because of their so-called high stopping power. Among all the substances that are described in the literature, sodium iodide crystals are presently the most popular and most commonly used scintillation detector material for the measurement of gamma radiation. The sodium iodide crystal lattice contains a small amount of thallium that is introduced into the crystal lattice to serve as centers of activation to enhance luminescence.

The employment of solid scintillation detector materials like thallium-activated sodium iodide detectors for the measurement of gamma radiation requires the conversion of the visible photon emission of the scintillator to voltage pulses.

A photomultiplier tube with a photosensitive cathode and a series of dynodes is nowadays most frequently used for the detection and measurement of any photon-emitting process. Therefore, this type of equipment is also applied for counting the light flashes emitted by scintillation-detecting materials. The interaction of gamma radiation with the scintillator material produces light flashes and this visible light is converted to an electronic signal through a voltage pulse. The voltage pulses are processed further by electronic circuits to provide information about the spectral distribution of the energy of the gamma radiation.

### Semiconductor Detectors

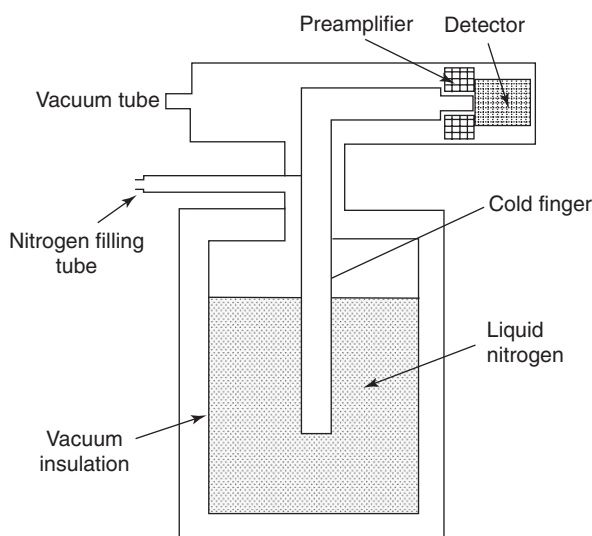
The principle of operation of this type of detector can be explained by the band structure of the electrons in the semiconductor crystal. The band gap of some semiconductor materials is very small, for germanium  $\sim 0.66$  eV. A passage of radiation may inject sufficient energy into the structure to raise an electron from the valence band to the conduction band. By this an electron-hole pair is created. The expression 'hole' has to be understood as the absence

of an electron. This electron-hole pair can be compared with an ion pair in a gas ionization counting tube. Both the electron and the hole can migrate through the semiconductor crystal in response to an electric field and can produce an electrical signal that indicates the passage of radiation.

Lithium drifted germanium detectors were in use during the first years of semiconductor detectors' application for gamma-ray spectrometry. They showed good resolution and efficiency but had to be kept at the temperature of liquid nitrogen all the time.

Nowadays, it is possible to produce semiconductor detectors made of germanium crystal materials with very high purity. This type of detector is called a high-purity germanium (HPGe) or intrinsic germanium detector. Without being connected to voltage supply they can be stored at room temperature all the time. They still have to be cooled to liquid nitrogen temperature for operation. The detector crystal is enclosed in a Dewar vacuum housing and connected by a copper rod (cold finger) to a reservoir of liquid nitrogen. **Figure 4** shows a schematic diagram of a semiconductor detector system cooled with liquid nitrogen. Portable germanium detectors with a small cryostat are even used for field work. The HPGe detectors nowadays have largely replaced the lithium drifted germanium Ge(Li) detectors. The high price of semiconductor detector materials is considered as one of the main disadvantages.

Due to the development of solid-state semiconductor detectors great progress with regards to spectral resolution had been achieved in gamma-ray spectrometry. By the use of these detectors spectra with excellent resolution are obtained.



**Figure 4** Schematic cross-sectional view of a semiconductor detector for gamma-ray spectrometry with liquid nitrogen Dewar.

Semiconductor detectors are usually made of germanium for the measurement of high-energy gamma radiation or of silicon for low-energy gamma- and X-rays.

## **Electronic Circuits for Pulse Handling**

An electronic preamplifier circuit is coupled with a semiconductor detector or a photomultiplier of the scintillation detector as close as possible to minimize any pulse degradation due to cable capacitance. The preamplifier amplifies the electronic pulses by increasing the signal-to-noise ratios before the pulses enter a cable leading to an amplifier. In addition, the preamplifier output pulses may be linearly amplified and shaped by a pulse-shaping circuit. By this unit the tailing of the preamplifier output pulse is removed and uniformly shaped output pulses of different magnitudes are produced. The magnitude of a pulse depends on the number of photons of a light flash created within the scintillation detector or the number of electron-hole pairs created within the semiconductor detector. Both the number of photons and the number of electron-hole pairs depend on the energy dissipated within the detector by a gamma-ray. Therefore, the magnitude of a pulse is a function of the gamma-ray energy. The number of pulses per unit of time is a function of the number of gamma-rays emitted by the radiation source. The number of gamma-rays emitted per unit of time by a radiation source is related to its activity in becquerel. Thus, the output signals of the amplifier provide information about the activity of the radiation source and the energy of the emitted radiation.

## **Pulse Height Analyzer**

### **Single Channel Analyzer**

There is a linear correlation between the gamma-ray energy dissipated in the detector and the amplitude of the output pulse of the amplifier pulse. To make use of the available information a pulse discriminator is connected with the counting system. A pulse discriminator is an electronic gate that permits pulses of a certain magnitude to continue on in the circuit to be counted. To select pulses of a well-defined magnitude two discriminators are used: a lower level discriminator and an upper level discriminator. Only pulses can pass the system that are high enough to overcome the lower level discriminator but also low enough to pass the barrier set by the upper level discriminator. The heights of the lower and upper level discriminators are set in such a way that a so-called window or channel is created that can only be passed

by pulses of a selected pulse height interval. In this way, only pulses that are related to a selected radiation energy interval are registered and counted. This can be used to measure the intensity of a specific gamma-ray energy region emitted by a radioactive sample, but this method is too time-consuming for spectrometric investigations. It is also too complicated to apply many such windows simultaneously for a spectrometric measurement and therefore another approach is used for multichannel analysis.

### **Multichannel Analyzer**

To make efficient use of the high-energy resolution of modern semiconductor detector systems a large number of analyzer channels is necessary. This can be achieved by an analog-to-digital converter, e.g., by conversion of the pulse height into a pulse length. The pulse length can then be measured digitally by means of a frequency standard. The digital result is used for storage at a corresponding address register. Nowadays, multichannel analyzer systems with at least 4000 or 8000 addresses are state of the art. This type of gamma-ray spectrometry is used for qualitative and quantitative analysis of gamma-emitting radioactive materials.

Gamma-ray spectrometers are either constructed as stand alone or hardware gamma spectrometers, or are connected with a computer or even simple boards in a computer. Excellent computer software is available for data processing to perform qualitative and quantitative analysis of gamma-emitting radionuclides. An example for a multichannel analysis of an environmental sample is shown in **Figure 1**.

## **Comparison of Detector Properties**

Detectors for gamma-ray spectrometry are usually characterized by their resolution, efficiency, and peak-to-Compton ratios.

Resolution is the ability to distinguish between closely spaced peaks in a spectrum. The resolution is given as the value of the full width of a peak at half of its maximum height (FWHM). This value is given in keV for a selected full-energy peak, usually for the 1332 keV peak of cobalt-60. Sometimes resolution is defined as a percentage of the absolute peak position. The shape of a peak and also the resolution is influenced by the count rate.

The efficiency is the quotient of the number of events (counts) registered by the detector and the number of gamma photons of given energy emitted by the radiation source.

Sodium iodide detector crystals give good efficiency for gamma-ray detection due to the high atomic

number of iodine. This is because of a large fraction of interactions by photoelectric effect. Sodium iodide crystals are available in many configurations from small crystals up to large detectors coupled with multiple photomultiplier tubes. Well-type configurations permit good counting geometry for small samples.

Sodium iodide crystals show good efficiency but rather limited spectral resolution. The peaks of a gamma spectrum from a sodium iodide crystal are rather broad. Peaks generated from gamma-rays with energies close together cannot be separated. This is a big problem if radioactive materials with many different gamma-ray energies or mixtures of gamma-emitting nuclides have to be analyzed. It may also be difficult to distinguish small peaks from a Compton or background continuum that superimposes the weak peaks.

Bismuth germanate detectors have even better efficiency than sodium iodide detectors but their energy resolution is poor. Therefore, this type of detector material is used for special purposes but not for spectrometry.

For analytical purposes particularly in the analysis of complex gamma spectra having many closely spaced peaks semiconductor detectors are used. They have excellent resolution but lower efficiency because of their lower atomic number. In addition, the volume of available semiconductor detectors is usually much smaller compared to sodium iodide detectors. Both of these limits result in much lower detection efficiency.

The efficiency of a semiconductor detector is described as a relative efficiency compared with a standard detector, which is a NaI(Tl) crystal, with a diameter of 3 in. and a length of 3 in. being exposed to a point source of radioactive material at a distance of 25 cm.

The peak-to-Compton ratio refers to the ratio of the number of counts in the 1332 keV peak of cobalt-60 compared to the number of counts in the Compton region. Germanium detectors usually have a resolution better than 2 keV at Cobalt-60 1332 keV, relative efficiencies are mostly smaller but can be ~100% or better, and peak-to-Compton ratios show values up to 50.

## Calibration of Gamma-Ray Spectrometry Equipment

### Energy and Efficiency Calibration

In gamma-ray spectrometry the pulse height scale has to be calibrated in terms of absolute energy in keV to identify unknown gamma-ray emitters. For this purpose the channel position of the maximum of each

**Table 1** Gamma-ray sources used for calibration of gamma-ray spectrometer systems

<i>Radionuclide</i>	<i>Gamma energy (keV)</i>	<i>Gamma photons per decay</i>	<i>Half-life</i>
<sup>109</sup> Cd	88.0	0.05	453 days
<sup>57</sup> Co	122.1	0.86	271 days
<sup>22</sup> Na	511.0	1.8	2.6 years
<sup>137</sup> Cs	661.6	0.85	30.17 years
<sup>54</sup> Mn	835.3	1.00	312.2 days
<sup>60</sup> Co	1173.2	1.00	5272 years
<sup>22</sup> Na	1274.6	1.00	2602 years
<sup>60</sup> Co	1332.5	1.00	5272 years

photo peak must be plotted versus the gamma-ray energy. A best fit through these data gives the energy corresponding to each channel in the multichannel analyzer. With good detectors a straight line is obtained; However, due to inherent nonlinearities in the detector or recording system this calibration may not be perfectly linear or pass exactly through the origin of the diagram.

For calibration of a gamma-ray spectrometry equipment a set of calibrated gamma-ray sources is necessary to determine various parameters like energy scale, efficiency, resolution, geometric effects, etc. Usually the activity of a calibrated source is given by the number of disintegrations per second (becquerel) or curie ( $1 \text{ Ci} = 3.7 \times 10^{10}$  disintegrations per second). The number of emitted gamma-rays per second is frequently less than the number of disintegrations depending on the branching ratio of the decay.

In Table 1, a list of radionuclides, their gamma-ray energies, and the fraction of gamma-rays per decay is listed. Such sets of gamma calibration sources are commercially available. The radionuclides are available either as point sources or as solutions. With point sources, the sample geometry cannot be changed whereas solutions of radionuclides can be filled in various containers like those that are used for the measurement of, e.g., environmental samples.

Catalogs of experimental measured gamma-ray spectra with both types of detectors, sodium iodide and semiconductor, have been published in the literature. This type of compilation is very useful for the prediction of the suitability of the detector types for a special application.

## Gamma-Ray Spectrometry in Quantitative Radionuclide Analysis

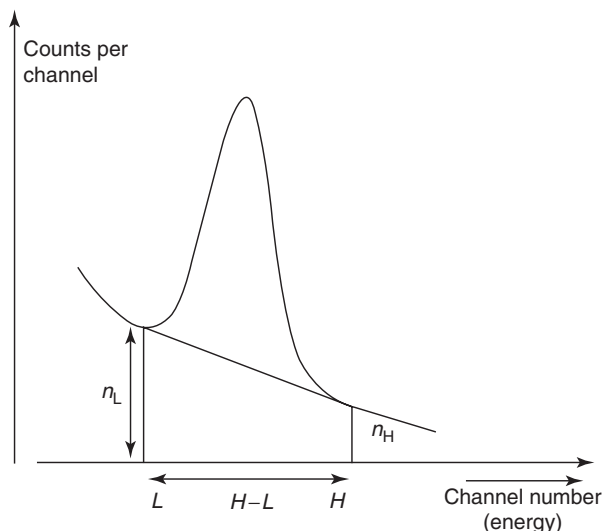
Gamma spectra are of great help for qualitative and quantitative analysis of radioactive samples of unknown composition.



The following method gives a simple example of how to perform a quantitative spectrum analysis. In any method, the number of integral (summation) counts of a photo-peak is proportional to the gamma-rays incident upon the detector. The total number of counts is displayed by the area of a photo-peak and this area has to be calculated. Nonoverlapping single peaks show a clear Gaussian shape. But the total area of a peak depends on the contributions of both the gamma photons from the sample and the gamma photons due to the background radiation and the Compton continuum of higher energetic peaks. Therefore, to obtain the contribution of the gamma-rays of the sample the contribution of the background radiation has to be subtracted. This background contribution to the peak area has the shape of a trapezoid. This trapezoid area can be easily calculated using the address (channel) numbers of the lower ( $L$ ) and higher ( $H$ ) end of the peak and the amount of counts being registered in the lower ( $n_L$ ) and in the higher ( $n_H$ ) address positions (channels). This is demonstrated by **Figure 5**

$$B = (H - L + 1) \cdot \frac{(n_L + n_H)}{2}$$

The area corresponding to the background contribution is subtracted from the total area, represented by the total number of counts being stored in all the channels from  $n_L$  to  $n_H$ . Very frequently this



**Figure 5** Determination of the net area of a peak in a gamma spectrum.  $T$  is the total peak area,  $L$  the lower end of the peak,  $H$  the higher end of the peak,  $B$  the contribution of the background to the total peak area,  $N$  the net peak area due to sample activity,  $n_L$  the counts being registered in the channel of the low end of the peak, and  $n_H$  the counts being registered in the channel of the high end of the peak:  $B = (H - L + 1) \cdot (n_L + n_H)/2$  and  $N = T - B$ .

calculation is much more complicated if there are fluctuations of channel contents in the neighborhood of channel  $n_L$  and  $n_H$ . In this case average values considering the relevant channel regions have to be calculated. To resolve overlapping peaks and construct their individual Gaussian shapes sophisticated and powerful computer programs have been developed. The interpretation of gamma spectra for quantitative analysis using computer programs nowadays has reached a very high standard of reliability and precision.

Finally, the radioactivity of a sample has to be calculated considering the counting efficiency of the detector for the gamma energy in question.

The simplest and most reliable method to determine the radioactivity will be the relative method in which the peak counting rate of a sample is compared quantitatively with that of the radioactive standard consisting of the same nuclide and geometry. Usually, the peak efficiency for each detector is prepared by means of experimental procedures using various kinds of radioactive standards. But theoretical methods based on Monte Carlo theory have also been designed.

## Special Types of Gamma-Ray Spectrometric Equipment

### Anti-Compton Spectrometers

One of the main problems in the interpretation of gamma spectra is the correct detection of low-level radioactivity in the presence of a high background. Such a high background can be due to a Compton continuum that is associated with an intense high-energy gamma peak. The detection of a small peak that may be situated in the region of such a Compton continuum is difficult. Such a problem can be avoided by the application of Compton suppression techniques. The best Compton suppression can be achieved by surrounding the detector for sample measurement by a so-called guard detector system. Most frequently sodium iodide or other scintillation detectors are used as guard detectors. Compton electrons are registered simultaneously by the guard system and by the detector facing the radioactive sample. This coincident information is subtracted from the information obtained by the detector for sample measurement. Thus, the background in the Compton region is reduced. By this method the Compton effect contribution is minimized and a peak-to-Compton ratio of 450:1 can be obtained.

Such a system is very helpful for environmental radioassay related to radionuclides with very low concentration in nature.

### Coincidence Spectrometers

Coincidence spectrometers are used for selectively measuring nuclear processes that occur with practically simultaneously emitting two gamma-rays of different energy or beta and gamma radiation within a very short time interval. By this selective measurements technique a radiation of low intensity in presence of an accompanying high-intensity radiation can be detected with reasonable sensitivity.

### Applications of Gamma-Ray Spectrometry

Nowadays gamma-ray spectrometry is the most important tool for radio assay. Quality control in radionuclide production, control of reactor operation, surveillance of the environment, and radio activation analysis are only a few examples for the application of gamma-ray spectrometry.

It has to be mentioned that the international control of nonproliferation of fissile materials and the

verification of the nuclear test ban treaty are mainly based on the application of gamma-ray spectrometric methods.

**See also: Activation Analysis:** Neutron Activation; Charged-Particle Activation; Photon Activation. **Radiochemical Methods:** Radionuclide Monitoring.

### Further Reading

- Brune D, Forkman B, and Persson B (1984) *Nuclear Analytical Chemistry*, 2nd edn. Lund: Chartwell-Bratt Ltd.
- Ehmann WD and Vance D (1991) *Radiochemistry and Nuclear Methods of Analysis*. New York: Wiley.
- Elving P J (ed.) (1986) *Treatise on Analytical Chemistry*, vol. 14. New York: Wiley.
- IAEA (1989) *Measurement of Radionuclides in Food and the Environment. A Guidebook*. Technical Report Series No 295. Vienna: IAEA.
- Knoll GF (1989) *Radiation Detection and Measurement*. New York: Wiley.
- L'Annunziata MF (ed.) (2003) *Handbook of Radioactivity Analysis*. London: Academic Press.

## Food and Environmental Applications

**J Tölgyessy and M Harangozó**, Slovak Technical University, Bratislava, Slovak Republic

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Radiochemical analysis is based on two outstanding features of radioactivity:

1. the high sensitivity and ease of measurement of radioactive radiation; and
2. the possibility of labeling chemical compounds with radioactive tracers.

Radionuclide techniques often have higher sensitivity than other analytical methods. The amounts of nuclides, correlated to an activity of 1000 Bq (see Table 1), can be derived from the law of radioactive decay. The amounts vary considerably, corresponding to the wide range of half-lives. For 90% of the commonly used nuclides, half-lives range from several minutes to several years, so the corresponding masses are extremely low.

Radionuclides are often diluted with inactive isotopes, but specific activities (i.e., activity per total mass of element) are still very high. Since the

background in nuclear spectrometry is very low, and sensitivity is high, activities as low as 0.2 Bq can be readily detected ( $^3\text{H}$ ,  $^{125}\text{I}$ ,  $^{132}\text{I}$ ), and even 0.01 Bq (several  $\gamma$ -emitters, e.g.,  $^{24}\text{Na}$ ,  $^{38}\text{Cl}$ ,  $^{42}\text{K}$ ,  $^{46}\text{Sc}$ ,  $^{59}\text{Fe}$ ,  $^{60}\text{Co}$ ,  $^{65}\text{Zn}$ ,  $^{110\text{m}}\text{Ag}$ ,  $^{182}\text{Ta}$ ,  $^{187}\text{W}$ ,  $^{198}\text{Au}$ ). Total activities  $\sim 0.1$ – $20$  kBq are often sufficient in

**Table 1** Mass corresponding to 1 kBq for carrier-free radionuclides

Nuclide	Mass <sup>a</sup> (ng)	Half-life <sup>b</sup>
$^{36}\text{Cl}$	820	$3 \times 10^5$ a
$^{14}\text{C}$	6.1	$5.8 \times 10^3$ a
$^{63}\text{Ni}$	0.57	120 a
$^{85}\text{Kr}$	$6.8 \times 10^{-2}$	10.6 a
$^{125}\text{I}$	$1.6 \times 10^{-3}$	60 d
$^3\text{H}$	$2.8 \times 10^{-3}$	12.4 a
$^{131}\text{I}$	$2.2 \times 10^{-4}$	8.04 d
$^{35}\text{S}$	$6.3 \times 10^{-5}$	87.2 d
$^{32}\text{P}$	$9.5 \times 10^{-5}$	14.3 d
$^{99\text{m}}\text{Te}$	$5.1 \times 10^{-6}$	6.02 h
$^{18}\text{F}$	$2.8 \times 10^{-7}$	1.83 h

<sup>a</sup>Commercially available with maximum specific activity of  $2 \times 10^4$  Bq g<sup>-1</sup>.

<sup>b</sup>a = year (annum).

Reproduced with permission from Garten Rainer PH and Tölgyessy J (2001) *Radionuclides in Analytical Chemistry, Ullmann's Encyclopedia of Industrial Chemistry*, 6th edn. CD-Rom. Weinheim: Wiley-VCH.

### Coincidence Spectrometers

Coincidence spectrometers are used for selectively measuring nuclear processes that occur with practically simultaneously emitting two gamma-rays of different energy or beta and gamma radiation within a very short time interval. By this selective measurements technique a radiation of low intensity in presence of an accompanying high-intensity radiation can be detected with reasonable sensitivity.

### Applications of Gamma-Ray Spectrometry

Nowadays gamma-ray spectrometry is the most important tool for radio assay. Quality control in radionuclide production, control of reactor operation, surveillance of the environment, and radio activation analysis are only a few examples for the application of gamma-ray spectrometry.

It has to be mentioned that the international control of nonproliferation of fissile materials and the

verification of the nuclear test ban treaty are mainly based on the application of gamma-ray spectrometric methods.

**See also: Activation Analysis:** Neutron Activation; Charged-Particle Activation; Photon Activation. **Radiochemical Methods:** Radionuclide Monitoring.

### Further Reading

- Brune D, Forkman B, and Persson B (1984) *Nuclear Analytical Chemistry*, 2nd edn. Lund: Chartwell-Bratt Ltd.
- Ehmann WD and Vance D (1991) *Radiochemistry and Nuclear Methods of Analysis*. New York: Wiley.
- Elving P J (ed.) (1986) *Treatise on Analytical Chemistry*, vol. 14. New York: Wiley.
- IAEA (1989) *Measurement of Radionuclides in Food and the Environment. A Guidebook*. Technical Report Series No 295. Vienna: IAEA.
- Knoll GF (1989) *Radiation Detection and Measurement*. New York: Wiley.
- L'Annunziata MF (ed.) (2003) *Handbook of Radioactivity Analysis*. London: Academic Press.

## Food and Environmental Applications

**J Tölgyessy and M Harangozó**, Slovak Technical University, Bratislava, Slovak Republic

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Radiochemical analysis is based on two outstanding features of radioactivity:

1. the high sensitivity and ease of measurement of radioactive radiation; and
2. the possibility of labeling chemical compounds with radioactive tracers.

Radionuclide techniques often have higher sensitivity than other analytical methods. The amounts of nuclides, correlated to an activity of 1000 Bq (see Table 1), can be derived from the law of radioactive decay. The amounts vary considerably, corresponding to the wide range of half-lives. For 90% of the commonly used nuclides, half-lives range from several minutes to several years, so the corresponding masses are extremely low.

Radionuclides are often diluted with inactive isotopes, but specific activities (i.e., activity per total mass of element) are still very high. Since the

background in nuclear spectrometry is very low, and sensitivity is high, activities as low as 0.2 Bq can be readily detected ( $^3\text{H}$ ,  $^{125}\text{I}$ ,  $^{132}\text{I}$ ), and even 0.01 Bq (several  $\gamma$ -emitters, e.g.,  $^{24}\text{Na}$ ,  $^{38}\text{Cl}$ ,  $^{42}\text{K}$ ,  $^{46}\text{Sc}$ ,  $^{59}\text{Fe}$ ,  $^{60}\text{Co}$ ,  $^{65}\text{Zn}$ ,  $^{110\text{m}}\text{Ag}$ ,  $^{182}\text{Ta}$ ,  $^{187}\text{W}$ ,  $^{198}\text{Au}$ ). Total activities  $\sim 0.1\text{--}20\text{ kBq}$  are often sufficient in

**Table 1** Mass corresponding to 1 kBq for carrier-free radionuclides

Nuclide	Mass <sup>a</sup> (ng)	Half-life <sup>b</sup>
$^{36}\text{Cl}$	820	$3 \times 10^5\text{ a}$
$^{14}\text{C}$	6.1	$5.8 \times 10^3\text{ a}$
$^{63}\text{Ni}$	0.57	120 a
$^{85}\text{Kr}$	$6.8 \times 10^{-2}$	10.6 a
$^{125}\text{I}$	$1.6 \times 10^{-3}$	60 d
$^3\text{H}$	$2.8 \times 10^{-3}$	12.4 a
$^{131}\text{I}$	$2.2 \times 10^{-4}$	8.04 d
$^{35}\text{S}$	$6.3 \times 10^{-5}$	87.2 d
$^{32}\text{P}$	$9.5 \times 10^{-5}$	14.3 d
$^{99\text{m}}\text{Te}$	$5.1 \times 10^{-6}$	6.02 h
$^{18}\text{F}$	$2.8 \times 10^{-7}$	1.83 h

<sup>a</sup>Commercially available with maximum specific activity of  $2 \times 10^4\text{ Bq g}^{-1}$ .

<sup>b</sup>a = year (annum).

Reproduced with permission from Garten Rainer PH and Tölgyessy J (2001) *Radionuclides in Analytical Chemistry, Ullmann's Encyclopedia of Industrial Chemistry*, 6th edn. CD-Rom. Weinheim: Wiley-VCH.

analytical applications. The precision of the final result can be better than 2%, if the counting error is kept below 1.5%. This can be obtained from counting times between 100 s and 5 h.

Radiochemical analysis relies on the assumption that different isotopes of the same element exhibit the same properties in any macroscopic physical or chemical process, and that radioactive labeling does not influence the other properties of a chemical species. This is generally the case, with deviations below 1% (with exception of hydrogen isotopes) owing to isotopic fractionation or radiation effects. For analytical purposes, the radiotracer and the analyte must be present in the same chemical form. This is usually easy to achieve, but specialized preparative techniques may be necessary for radioactive labeling of more complex organic compounds.

Measurements of ionizing radiation and radionuclides in foods and environmental samples are required for the assessment of exposure to both natural and artificial radiation sources, determination of compliance with government regulations, and studies of the movement and retention of artificial radionuclides in food and environmental media and of the composition of the natural radiation environment.

A wide assortment of equipment and methods are available with which to:

1. monitor the presence and ecological behavior of radionuclides that are present in food and environmental samples;
2. identify the pathways by which human exposure results; and
3. estimate the dose to humans.

The selection of foodstuffs depends on the behavior of analytes in the food chain and the eating habits of local populations. Materials frequently analyzed for radioactive contamination include food crops (fruits and vegetables), milk (cow and goat), domestic animals (deer, rabbits, and squirrels), and game birds (pheasant).

Laboratory measuring equipment used in radiochemical analytical procedures should show a high radiation sensitivity, high detection efficiency, low background, and high stability.

Counting rates measured with Geiger-Müller, scintillation, or semiconductor detectors are usually compared with suitable standards to obtain quantitative data. Pulse analyzers (either single or multichannel), when coupled to a suitable radiation detector, can yield qualitative as well as quantitative data about individual components of the sample.

## Preparation of Labeled Compounds

Aspects of the labeled substance involve the following requirements:

1. Radionuclide and radiochemical purity are essential. The labeled substance must react identically to the analyte.
2. Complete homogenization and isotope exchange must be achieved, requiring successive transformation of the active traces as well as of the inactive analyte to all chemical species that are considered to occur in the complete system.
3. Isotope effects must be negligible.
4. For most elements, suitable radionuclides for analytical use are commercially available. There are only seven elements of interest (B, He, Li, N, Ne, O, Mg), for which this is not the case.
5. The preparation of a specific isotope, of an isotopically modified (i.e., mixture of labeled and unlabeled), isotopically substituted, or even site-specifically labeled compound is often mandatory in analytical applications.
6. Synthesis (including biosynthesis) of these labeled compounds often requires much more effort than the actual radioanalytical experiment. A variety of labeled compounds and pharmaceuticals are available commercially from stock, and a further range of compounds will be prepared on request by a number of suppliers.
7. Cost of radionuclides for analytical applications are typically in the range US\$150–400 per 40 MBq, and higher where preparation is difficult or time consuming.

Synthesis of labeled compounds involve all of the classical biochemical and synthetic chemical reactions used in the preparation of nonradioactive chemicals.

Synthetic chemical methods for the preparation of carbon-14-labeled chemicals utilize a number of basic building components that are prepared from barium carbonate ( $^{14}\text{C}$ ). These are  $^{14}\text{C}$  versions of carbon dioxide, acetylene, benzene, sodium acetate (1- and 2- $^{14}\text{C}$ ), methyl iodide, methanol, sodium cyanide, and urea. Many complicated labeled compounds may be synthesized from these materials.

Isotope exchange can be used when the substance to be labeled is readily obtainable in the inactive form and conditions favorable for the exchange, e.g., temperature and exchange rate, can be found. This is so, for example, for Bengal rose and thyroxine, both of which may be labeled by exchange with  $^{131}\text{I}_2$ . The starting material must be free from iodinated or already iodinated compounds whose exchange rates are roughly equal to or greater than the rate of the

desired reaction. For preparative purposes, the isotope exchange method is largely confined to the elements bromine and iodine.

Biosynthetic methods are ideally suited for the synthesis of many radiolabeled compounds. Plants (e.g., potato and tobacco), when grown in the atmosphere of radioactive carbon dioxide utilize the  $^{14}\text{CO}_2$  as their sole source of carbon. After a suitable period of growth, almost every carbon atom in the plant is radioactive. In this way plants can serve as a source of labeled carbohydrates. Algae grown under similar conditions provide labeled amino acids, lipids, nucleotides, etc. Vitamin  $\text{B}_{12}$  is obtained from the metabolic products of microorganisms whose culture media contain cobalt-57 or cobalt-58.

Labeling with tritium ( $^3\text{H}$ ) is most commonly achieved by reductive methods, including catalytic reduction of olefin by deuterium gas ( $^3\text{H}_2$ ), catalytic reductive replacement of halogen by  $^3\text{H}_2$ , and metal hydride ( $^3\text{H}$ ) reduction of carbonyl compounds to tritium-labeled alcohols.

## Nonchromatographic Methods of Analysis

In general, there can be considered to be three working stages in radiochemical analysis:

1. sample preparation,
2. chemical separation, and
3. radioactivity measurement.

Following sample preparation, the element whose radioisotope (or labeled compounds) is to be determined is separated from the sample. The final measurement of the activity of the element (or labeled compounds) is made on the isolate.

One of the main differences between radiochemical analytical procedures and classical analytical methods is that the element (and particularly its radioisotope) to be determined is present in the sample in minor to trace amounts. Separation of radionuclides is performed with the aid of a suitable carrier. Generally, the carrier is a stable isotope (or a suitable compound) that is added to the radioactive compound in a small but detectable amount and has identical chemical properties. An isotopic carrier, i.e., a stable isotope of the element in question, is most frequently used. Both the radioactive isotope and the carrier must be in the same chemical form. The isotopic carrier is irreversibly mixed with the radioactive compound and cannot be separated from it again by chemical means. Such a carrier can therefore be used only when a lower specific activity is sufficient for the subsequent operations. For example, barium or lead can serve as carriers when

separating radium; iron or yttrium may be applied in separations of rare-earth elements, etc. In certain cases, when no stable isotope is available, e.g., as in plutonium separation, a separation of carrier-free radionuclide is also possible.

Essentially all the separation methods known from classical analytical chemistry can be applied to chemical separations of radionuclides and labeled compounds from samples to be analyzed: precipitation, electrolytic deposition, extraction, ion exchange, distillation, chromatography, etc.

As only small absolute amounts of radionuclides are to be determined, and thus only small amounts of elements are to be separated, separation cannot usually be properly performed by methods whose success depends on the amount of the component to be isolated (e.g., as in precipitation). Therefore, methods that are independent of amount (such as liquid-liquid extraction and ion-exchange methods) are more advantageous. Extraction procedures very often take advantage of additions of chelating components. For separating volatile radionuclides (such as iodine or ruthenium) from the sample matrix, distillation methods can be used advantageously. Electrolytic deposition has been shown to be applicable in the separation of polonium.

In contrast with precipitation procedures, these specific methods are time saving and allow the separation of carrier-free radionuclides. Radiochemical methods of determining radionuclides show the highest sensitivity. Radiochemical methods for determining fission products were originally devised for treating high-activity solutions; however, they are not always convenient for the determination of very small concentrations in biological materials. Radiochemical analytical procedures have been developed for those cases where the radionuclides should be separated from large amounts of stable salts after ashing the sample. Among fission products, strontium-90 is of vital importance; the determination of the other products is of minor significance.

The methods used for the detection and quantification of radionuclides and labeled compounds are determined by the type of emission ( $\beta$  or  $\gamma$ ), the energy of the radiation, and the efficiency of the system by which it is measured. Commercially available instrumentation is generally used, with exception of some research work and highly specialized routine analytical application. Detection of radioactivity can be achieved in all cases with the Geiger-Müller counter. However, for the weaker-emitting radionuclides, i.e.,  $^3\text{H}$ ,  $^{14}\text{C}$ , and  $^{35}\text{S}$ , large amounts of radionuclides are required for detection of a signal. In most cases this is both undesirable and impractical.

Liquid scintillation counting is by far the most common method of detecting and quantifying low-energy  $\beta$ -radiation. A less common technique is planchette counting. A film of the sample on a planchette, which is a flat metal pan, is brought within a fixed distance of a Geiger–Müller counter.

The detection and quantification of  $\gamma$ -radiation is accomplished by well counting in a scintillator, the sample being contained in a well or drilled hole.  $\gamma$ -Radiation is often measured by  $\gamma$ -ray scintillation or semiconductor spectrometry. Gas flow counting is a method for detecting and quantifying radionuclides on paper chromatographic strips and thin-layer plates.

The detection limit (DL) of the concentration of the analyte is given by

$$DL = \frac{f}{s} \sqrt{B}$$

Here  $f=3$  is appropriate for detection at the 99.9% confidence level,  $B$  is the blank level, including analytical blank and spectral background of the detection device, and  $s$  is the analytical calibration function, i.e., under idealized conditions:

$$s = \varepsilon \eta t a m_{\text{sample}}$$

where  $\varepsilon$  is the detection efficiency,  $\eta$  is the emission probability,  $t$  is the counting time,  $a$  is the specific activity of the analyte, and  $m_{\text{sample}}$  is the mass.

Specimen preparation for radioactivity measurements aims, in general, to obtain the analyte:

1. As completely as possible and with the highest possible degree of purity from extraneous radioactivity and from excess matrix matter.
2. Concentrated on the smallest area or volume of a solid, liquid, or gaseous (for H, C, S, Ar, Kr, Xe, Rn) target.
3. On a specimen support that is compatible with the chemical nature of the species analyzed, with the specific radiation to be measured, and with the detector system, e.g., liquid scintillation cocktails for weak  $\beta$ -emitters.

These preparation requirements are compatible with the more general rules for specimen preparation in trace and ultratrace analysis by any technique.

## Chromatographic Methods

Chromatographic methods are frequently used to separate radioactive nuclides and labeled compounds and can be used to analyze complex systems by using radioactive indicators. The principle of separation is the same as in standard chromatographic methods

but radiometric detection is used. It is possible to separate and detect radioactive compounds, compounds converted to radioactive derivatives, or compounds labeled with radioactive nuclides.

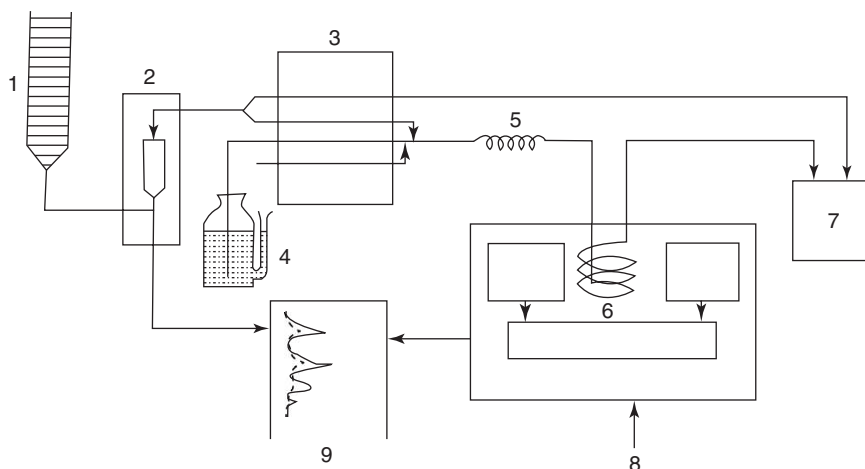
The detection of radioactivity is very sensitive, but there may be problems in manipulating the radioactive compounds because of contamination of various parts of the chromatographic system. Complete decontamination of those parts that have been in contact with radioactive compounds must be ensured. It is sometimes better to exchange inexpensive parts for new ones. Even small amounts of radioactive impurities can seriously interfere because of the high sensitivity of detection. The possibility of forming radioactive reaction products, such as tritiated water in reactions with compounds labeled with  $^3\text{H}$ , or  $^{14}\text{CO}_2$  in reactions with compounds labeled with  $^{14}\text{C}$ , must also be considered.

Considerable losses of samples with low concentrations of radioactive nuclides may occur because of adsorption on various parts of the chromatographic system. These difficulties can be overcome by the addition of a suitable carrier.

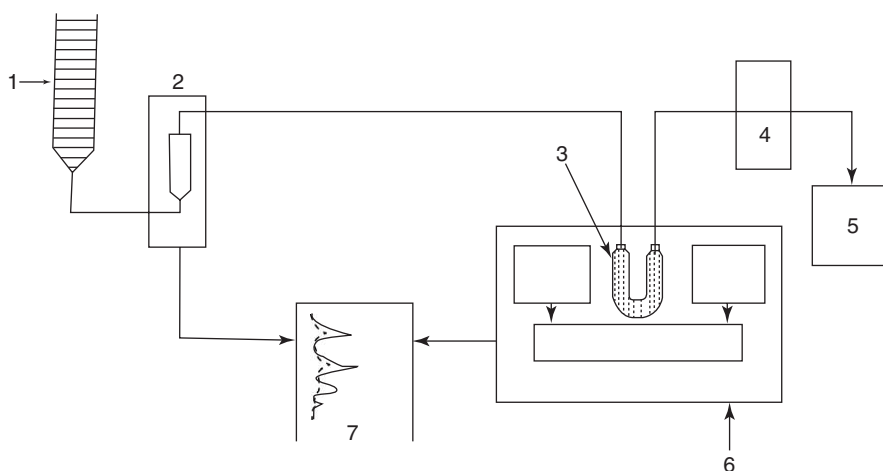
All chromatographic methods (liquid, liquid–liquid, liquid–solid, gas, gas–liquid, and gas–solid chromatography) are used in radiochemical analysis of food and environmental samples. Equipment for the evaluation of paper and thin-layer radiochromatograms is commercially available. The radiation detectors used vary according to the type and energy of the radiation emitted by the measured radionuclides. Use is made of ionization chambers, proportional counters, and Geiger–Müller, scintillation, and semiconductor detectors.

The measurement of radioactivity in chromatographic methods can be either discontinuous or continuous. Discontinuous analysis is a very simple method that depends upon measuring the radioactivity of the individual fractions of components that were separated in the column, chromatographic paper, adsorbent, etc. This method is used whenever the specific activities are very low and there is no time limit for the analysis. In this technique, however, results are obtained only after sometime and consequently parts of the fractions may be lost. In column chromatography fraction collectors are exchanged at regular time intervals or as soon as a detector (e.g., a standard ultraviolet (UV) absorbance detector) detects the passage of an analyte. The fraction is detected as soon as it leaves the column. A Geiger–Müller or a well-scintillator detector can be used for  $\beta$ -emitters (e.g.,  $^{32}\text{P}$ ) or  $\gamma$ -emitters. Semiconductor detectors can also be used to detect  $\beta$ -radiation. Cherenkov-type detectors are sometimes used for radiation of higher energy.





**Figure 1** Homogeneous measuring system (1 – column, 2 – UV detector, 3 – pump, 4 – scintillator, 5 – mixing chamber, 6 – flow cell, 7 – fraction collector, 8 – scintillation detector, 9 – recording device).



**Figure 2** Heterogeneous measuring system (1 – column, 2 – UV detector, 3 – flow cell, 4 – pump, 5 – fraction detector, 6 – scintillation detector, 7 – recording device).

In continuous analysis, radioactivity detectors are coupled serially with standard chromatographic detectors. The eluate passes through a suitable flow cell and the radioactivity is measured continuously. There are two types of measurements: heterogeneous and homogeneous. These two types of measuring systems for liquid chromatography are shown in **Figures 1** and **2**. In the heterogeneous mode, the measuring cell can be made of a scintillating material. Frequently used configurations are U-shaped cells packed with anthracene, scintillating lithium glass activated with cerium oxide, or other scintillators. For homogeneous measurements, the eluate is mixed with a liquid scintillator before entering the detector. The cell for homogeneous measurements is usually a spiral. It is placed between two photomultiplier tubes that are connected in a coincidence circuit.

Classical liquid radiochromatography has been used mainly for preparative work, and analytical applications have been rare. The method has been important in the separation and isolation of new radioactive elements and labeled compounds. Technical progress has made this method applicable to the direct analytical determination of biochemically important compounds, labeled preparations, components of food and environmental samples.

### Illustrative Examples from Environmental Science

Radioactive contamination can result from either expected or accidental release of radioactive materials during the treatment of uranium and thorium ores, the operation of nuclear reactors, or the

application of radionuclides and labeled compounds in medicine, research, industry, and agriculture, as well as from radioactive fallout from the atmosphere.

Man obtains food from three regions of the biosphere: seawater, the sweet water hydrosphere, and soil. Thus, radioactive contamination of these three regions represents the chief field of interest for radiochemical analysis.

Tritium in water samples is determined with a liquid scintillation spectrometer because of its low maximum  $\beta$ -particle energy (18 keV). Water samples are distilled to remove nonvolatile components whose presence could lead to extinction during the scintillation detection and could also be radioactive. The sample is distilled to dryness in order that the tritium present can be transferred quantitatively into the distillate. A portion of the distillate is mixed with the liquid scintillator, and the radioactivity is measured by a liquid scintillation spectrometer. A standard tritium sample and a sample for measuring the background are prepared in a similar way.

The radionuclide zirconium-95 ( $^{95}\text{Zr}$ ) can be found among direct products of nuclear fission. Its radioactive decay leads to the daughter niobium-95 ( $^{95}\text{Nb}$ ). In the determination of  $^{95}\text{Zr}$  and  $^{95}\text{Nb}$  in samples of seawater and sea plants, the samples are mixed with oxalic acid in order that zirconium and niobium complexes can be formed in the presence of nitric acid. The oxalic acid is destroyed with potassium chlorate, and zirconium and niobium are precipitated as zirconium phosphate and niobic acid, respectively. Activities of rare-earth elements are removed, and zirconium is separated as barium hexafluorozirconate and ashed to zirconium pentoxide. The niobium fraction is ashed to niobium pentoxide. Both radionuclides are finally determined by  $\gamma$ -ray spectrometry.

Radioruthenium enters plants because of direct contamination; however, it is usually concentrated in certain water plants, so that its determination is significant, particularly in regions where radioactive waste materials are stored in the sea. Ruthenium determination can be carried out by a simple and direct  $\gamma$ -ray spectrometric method. When it satisfies the other conditions, this procedure is rapid and economical. It is difficult to distinguish  $^{103}\text{Ru}$  and  $^{106}\text{Ru}$  because of their closely spaced  $\gamma$ -lines.

In many environmental samples, particularly those that are not contaminated with fission products, the radioiodine content can be determined by a direct  $\gamma$ -ray spectrometric method. The 364 keV  $\gamma$ -line is the most suitable for the spectrometric determination of radioiodine,  $^{131}\text{I}$ . Certain difficulties can be encountered when the radionuclides barium-140 and lanthanum-140 are present in the sample.

$\gamma$ -Ray spectrometry can also be employed for radium determination in seawater in combination with the separation with ammonium molybdophosphate.

A majority of the radiochemical methods for radium determination in environmental samples takes advantage of its isomorphous co-precipitation with barium as its sulfate. Very small amounts of the barium carrier are typically used to minimize the self-absorption of  $\alpha$ -particles from the radium.

Neptunium can easily be separated quantitatively from plutonium with the aid of anion exchangers. If both ions are in a concentrated hydrochloric acid solution, iodide ions are added to reduce plutonium to Pu(III) and neptunium to Np(IV). The solution then passes through a column filled with Dowex-1 anion exchanger. Plutonium is thus quantitatively eluted with concentrated hydrochloric acid, and neptunium can be desorbed with  $0.5 \text{ mol l}^{-1} \text{ HCl}$ .

## Illustrative Examples from Food Science

It was originally believed that radionuclides were introduced into the human organism by inhalation and/or through drinking water. Later it was shown that many radionuclides (e.g.,  $^{90}\text{Sr}$  and  $^{137}\text{Cs}$ ) came into the human organism only through food. Thus, attention should be paid not only to a study of the occurrence of radionuclides in air and water, but also to their determination in food and even in forage.

A study of milk contamination is of particular interest, since strontium, cesium, and iodine enter into the human organism only from milk.

Contamination of cheese usually has its origin in the presence of  $^{89}\text{Sr}$  and/or  $^{90}\text{Sr}$  in milk. Contamination of meat can mostly be attributed to  $^{137}\text{Cs}$ . Contents of either  $^{131}\text{I}$  in the thyroid gland of animals or  $^{89}\text{Sr}$  and  $^{90}\text{Sr}$  in their bones may serve as indicators of the degree of environmental contamination.

Labeled compounds offer many advantages for food science. The extreme sensitivity of radioactivity measurement enables one to determine and characterize biologically important trace elements and compounds (e.g., pesticides and pesticide residues) in food samples.

Many analytical methods have been developed for simultaneous radiochemical determination of several radionuclides, including  $^{89}\text{Sr}$ ,  $^{90}\text{Sr}$ ,  $^{140}\text{Ba}$ ,  $^{137}\text{Cs}$ ,  $^{131}\text{I}$ , and  $^{239}\text{Pu}$ , in a great variety of food samples, such as milk, corn, vegetables, fruits, fats, eggs, meat, etc.

Strontium-90 is commonly determined by measuring its daughter yttrium-90 ( $^{90}\text{Y}$ ) at the state of the radioactive equilibrium after the chemical separation. As radionuclides  $^{89}\text{Sr}$  and  $^{90}\text{Sr}$  emit no

$\gamma$ -radiation, they have to be separated from  $^{90}\text{Y}$  prior to their determination.

Strontium radioisotopes in milk can be determined by making use of an ion exchange procedure. The milk sample is allowed to stand for 2 weeks to achieve a sufficient  $^{90}\text{Y}$  accumulation. Citrate and both yttrium and strontium carriers are added. The solution is applied onto a column of cation then anion exchangers. Strontium and yttrium are retained on the cation and anion exchangers, respectively. Yttrium is washed out from the anion exchanger with hydrochloric acid and precipitated as yttrium oxalate. After weighing the precipitate and determining the chemical yield the  $\beta$  activity of the  $^{90}\text{Y}$  is measured. In the presence of  $^{140}\text{La}$  the yttrium oxalate is dissolved in concentrated nitric acid, and yttrium is extracted with tributyl phosphate. Lanthanum-140 remains in the aqueous phase. Yttrium is back-washed with nitric acid and precipitated as its oxalate. In an unfavorable case, i.e., when the amount of  $^{140}\text{Ba}$  exceeds by several orders that of  $^{90}\text{Sr}$ , the above separation procedure has to be repeated.

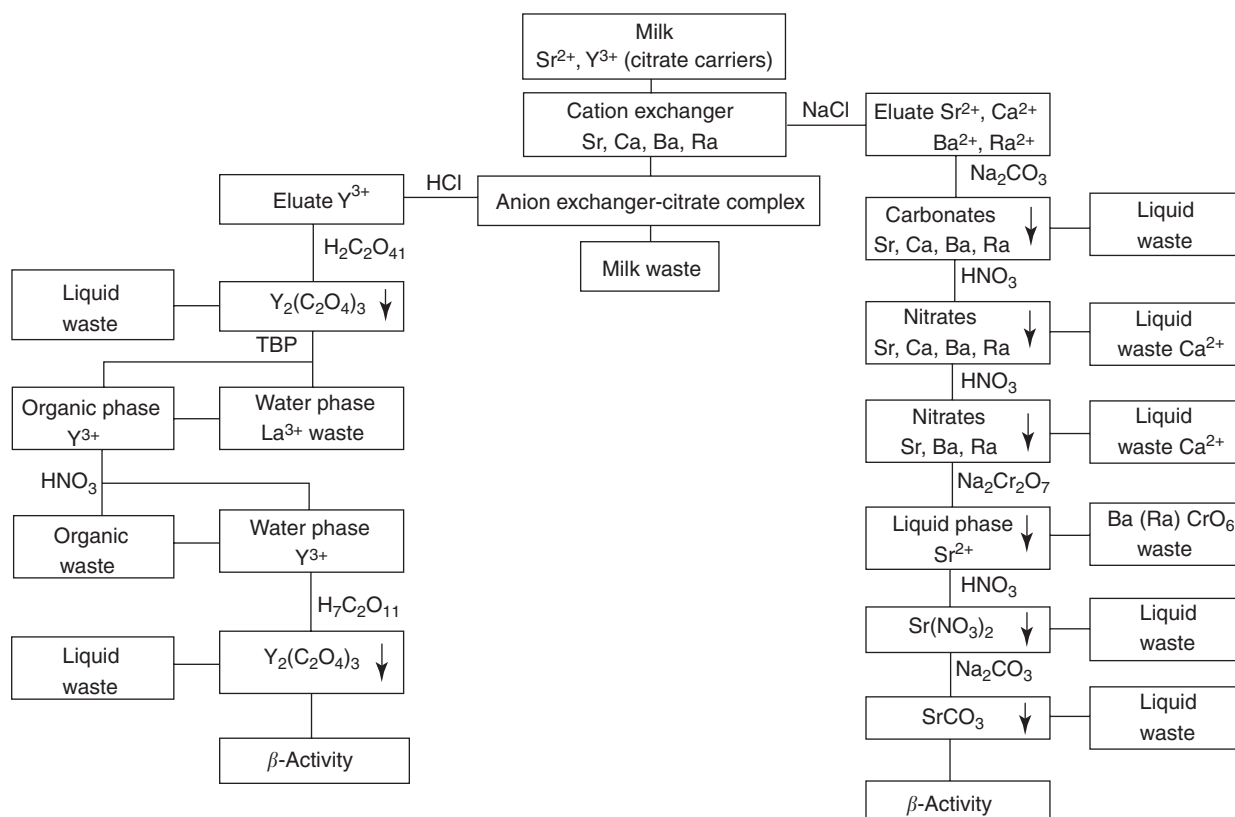
Strontium is washed out from the cation exchanger with a sodium chloride solution, concentrated in the eluate by precipitation as its carbonate, and purified through nitrate precipitation. Barium-140 and

radium are removed by chromate precipitation. When the strontium is precipitated as its carbonate, its chemical yield is determined, and the activity of all the radiostrontium is measured. The procedure for separating  $^{90}\text{Sr}$  and  $^{90}\text{Sr}$  from  $^{90}\text{Y}$  by ion exchange is shown in Figure 3.

The radioiodine ( $^{131}\text{I}$ ) content of milk can be determined on the basis of iodine extraction from milk with tetrachloromethane (carbon tetrachloride). In this way, only inorganic iodine in milk is determined, and organically bound iodine is obtained after alkali metal fusion and extraction. The activity is measured after the precipitation of silver iodide.

The  $^{137}\text{Cs}$  concentration in foods, plants, corn samples, etc., can be determined by direct measurement of radioactivity with a  $\gamma$ -ray spectrometer. This method requires a higher  $^{137}\text{Cs}$  concentration than in the radiochemical procedure, and radioactive impurities whose  $\gamma$ -rays have energies similar to that of cesium  $\gamma$ -rays (such as  $^{95}\text{Zr}$ ) can be present only in negligible amounts. The spectrometric method can be used either directly on the original sample or after concentrating  $^{137}\text{Cs}$ , depending on the detection efficiency and on the  $^{137}\text{Cs}$  concentration in the sample.

In the determination of uranium the food sample is decomposed with concentrated nitric acid, and



**Figure 3** Separation scheme for determining radiostrontium in milk by using ion exchange.

uranium is extracted with tributyl phosphate in paraffin. The solution of ammonium nitrate and iron(III) nitrate is employed as a salting-out agent. The organic phase is washed with nitric acid, and uranium is back-extracted into water, concentrated by evaporation, and its  $\alpha$ -activity is measured.

In aqueous food samples, plutonium can be separated by precipitation of bismuth phosphate and extraction of the cupferronate complex in chloroform. Finally, plutonium is precipitated by using an iron(III) carrier and measured by means of an  $\alpha$ -ray spectrometer.

Paper radiochromatography has been employed to determine vitamin B<sub>12</sub> labeled with cobalt-60 in foods, to follow its metabolism, and for various other studies. Thin-layer chromatography has been used for the determination of lipids labeled with <sup>3</sup>H, <sup>14</sup>C, or <sup>131</sup>I in foods.

Gas chromatography has been applied successfully to the separation and analysis of labeled compounds (chiefly those labeled with <sup>3</sup>H and <sup>14</sup>C) in food samples. Studies have been made of the biosynthesis of unsaturated fatty acids, the effect of insulin on the biosynthesis of individual fatty acids, the biosynthesis of cholesterol, etc. Numerous studies deal with the incorporation of labeled compounds through foods into living organisms.

**See also:** **Chromatography:** Principles. **Derivatization of Analytes.** **Food and Nutritional Analysis:** Meat and

Meat Products. **Pharmaceutical Analysis:** Overview. **Radiochemical Methods:** Overview; Natural and Artificial Radioactivity; Radionuclide Monitoring; Radon; Radiotracers; Gamma-Ray Spectrometry. **Water Analysis:** Freshwater; Seawater – Organic Compounds; Seawater – Inorganic Compounds.

## Further Reading

- Drenham DH (1982) In: Klement AW (ed.) *CRC Handbook of Environmental Radiation*. Boca Raton, FL: CRC Press.
- Environmental Radioactivity Surveillance Guide* (1972) ORP/SID-72-2, Office of Radiation Programs. Washington, DC: US Environmental Protection Agency.
- Kirk RE and Othmer DF (1982) *Encyclopedia of Chemical Technology*. 3rd edn., vol. 12. New York: Wiley.
- NCRP (1977) *Environmental Radiation Measurement*, Report No. 50. Washington, DC: National Council on Radiation Protection and Measurements.
- Šaršúnová M, Hanč O, Kakáč B, et al. (1990) *HPLC in Pharmacy and Biochemistry*. Heidelberg: Hüthig Buch Verlag.
- Tölgyessy J and Bujdosó E (1991) *Handbook of Radioanalytical Chemistry*, vol 1. Boca Raton, FL: CRC press.
- Tölgyessy J and Klehr EH (1987) *Nuclear Environmental Chemical Analysis*. Chichester: E. Horwood.
- Tölgyessy J and Koprda V (1999) In: Sohns T and Voicu VA (eds.) *NBC Risks*, pp. 523–529. Dordrecht: Kluwer Academic.
- Tölgyessy J and Varga Š (1974) *Nuclear Analytical Chemistry*. Baltimore: University Park Press.

## Pharmaceutical Applications

**R E Stringer**, Royal Liverpool University Hospital, Liverpool, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Radioactive compounds can be used in medicine for the diagnosis or treatment of disease. A radiopharmaceutical is an unsealed source of radioactivity in liquid, gaseous, or particulate form, which has to be removed from its container before administration. This article will discuss the analytical methods used for quality control of radiopharmaceuticals.

Most, but not all, radioisotopes require a carrier molecule (ligand) to be transported around the body and localize into the tissue or organ of interest. The applicability of a particular radiopharmaceutical depends on both the radioisotope and the ligand. There

should be selective concentration in the target organ with fast clearance from surrounding tissues (e.g., blood) giving a high target-to-background ratio. **Table 1** shows some of the many radiopharmaceuticals currently in clinical use.

The ideal isotope for use in diagnostic imaging is a gamma or positron emitter. These emissions are penetrating and can be detected outside of the body by the imaging equipment. The isotope should have a short half-life to minimize the radiation dose to the patient. Other requirements are that it should be nontoxic, readily available, inexpensive to produce, and chemically reactive. Technetium (<sup>99</sup>Tc<sup>m</sup>) fulfills all of these requirements having a gamma energy emission of 140 keV, 6 h half-life, and ready availability via a generator system. Technetium (<sup>99</sup>Tc<sup>m</sup>) labeled products are usually produced in the radiopharmacy from cold kits (from a licensed manufacturer) to which the isotope is added after elution

uranium is extracted with tributyl phosphate in paraffin. The solution of ammonium nitrate and iron(III) nitrate is employed as a salting-out agent. The organic phase is washed with nitric acid, and uranium is back-extracted into water, concentrated by evaporation, and its  $\alpha$ -activity is measured.

In aqueous food samples, plutonium can be separated by precipitation of bismuth phosphate and extraction of the cupferronate complex in chloroform. Finally, plutonium is precipitated by using an iron(III) carrier and measured by means of an  $\alpha$ -ray spectrometer.

Paper radiochromatography has been employed to determine vitamin B<sub>12</sub> labeled with cobalt-60 in foods, to follow its metabolism, and for various other studies. Thin-layer chromatography has been used for the determination of lipids labeled with <sup>3</sup>H, <sup>14</sup>C, or <sup>131</sup>I in foods.

Gas chromatography has been applied successfully to the separation and analysis of labeled compounds (chiefly those labeled with <sup>3</sup>H and <sup>14</sup>C) in food samples. Studies have been made of the biosynthesis of unsaturated fatty acids, the effect of insulin on the biosynthesis of individual fatty acids, the biosynthesis of cholesterol, etc. Numerous studies deal with the incorporation of labeled compounds through foods into living organisms.

**See also:** **Chromatography:** Principles. **Derivatization of Analytes.** **Food and Nutritional Analysis:** Meat and

Meat Products. **Pharmaceutical Analysis:** Overview. **Radiochemical Methods:** Overview; Natural and Artificial Radioactivity; Radionuclide Monitoring; Radon; Radiotracers; Gamma-Ray Spectrometry. **Water Analysis:** Freshwater; Seawater – Organic Compounds; Seawater – Inorganic Compounds.

## Further Reading

- Drenham DH (1982) In: Klement AW (ed.) *CRC Handbook of Environmental Radiation*. Boca Raton, FL: CRC Press.
- Environmental Radioactivity Surveillance Guide* (1972) ORP/SID-72-2, Office of Radiation Programs. Washington, DC: US Environmental Protection Agency.
- Kirk RE and Othmer DF (1982) *Encyclopedia of Chemical Technology*. 3rd edn., vol. 12. New York: Wiley.
- NCRP (1977) *Environmental Radiation Measurement*, Report No. 50. Washington, DC: National Council on Radiation Protection and Measurements.
- Šaršunová M, Hanč O, Kakáč B, et al. (1990) *HPLC in Pharmacy and Biochemistry*. Heidelberg: Hüthig Buch Verlag.
- Tölgyessy J and Bujdosó E (1991) *Handbook of Radioanalytical Chemistry*, vol 1. Boca Raton, FL: CRC press.
- Tölgyessy J and Klehr EH (1987) *Nuclear Environmental Chemical Analysis*. Chichester: E. Horwood.
- Tölgyessy J and Koprda V (1999) In: Sohns T and Voicu VA (eds.) *NBC Risks*, pp. 523–529. Dordrecht: Kluwer Academic.
- Tölgyessy J and Varga Š (1974) *Nuclear Analytical Chemistry*. Baltimore: University Park Press.

## Pharmaceutical Applications

**R E Stringer**, Royal Liverpool University Hospital, Liverpool, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Radioactive compounds can be used in medicine for the diagnosis or treatment of disease. A radiopharmaceutical is an unsealed source of radioactivity in liquid, gaseous, or particulate form, which has to be removed from its container before administration. This article will discuss the analytical methods used for quality control of radiopharmaceuticals.

Most, but not all, radioisotopes require a carrier molecule (ligand) to be transported around the body and localize into the tissue or organ of interest. The applicability of a particular radiopharmaceutical depends on both the radioisotope and the ligand. There

should be selective concentration in the target organ with fast clearance from surrounding tissues (e.g., blood) giving a high target-to-background ratio. **Table 1** shows some of the many radiopharmaceuticals currently in clinical use.

The ideal isotope for use in diagnostic imaging is a gamma or positron emitter. These emissions are penetrating and can be detected outside of the body by the imaging equipment. The isotope should have a short half-life to minimize the radiation dose to the patient. Other requirements are that it should be nontoxic, readily available, inexpensive to produce, and chemically reactive. Technetium (<sup>99</sup>Tc<sup>m</sup>) fulfills all of these requirements having a gamma energy emission of 140 keV, 6 h half-life, and ready availability via a generator system. Technetium (<sup>99</sup>Tc<sup>m</sup>) labeled products are usually produced in the radiopharmacy from cold kits (from a licensed manufacturer) to which the isotope is added after elution

from a generator. Following an incubation stage the product is ready to use without further purification. The kit is usually a lyophilized preparation containing ligand, reductant, and buffer. More recently, positron emitters such as fluorine ( $^{18}\text{F}$ ) have been introduced for use in positron emission tomography (PET). Radiopharmaceuticals for use in PET are

**Table 1** Examples of radiopharmaceuticals used in clinical practice

<i>Radiopharmaceutical</i>	<i>Example of clinical use</i>
Methionine ( $^{11}\text{C}$ ) injection	Imaging primary tumors and metastases in the brain
Fludeoxyglucose ( $^{18}\text{F}$ ) injection (FDG)	Imaging primary tumors and metastases
Chromium ( $^{51}\text{Cr}$ ) edetate injection	Determination of glomerular filtration rate
Yttrium ( $^{90}\text{Y}$ ) silicate injection	Radiation synovectomy agent
Technetium ( $^{99}\text{Tc}^{\text{m}}$ ) medronate injection	Bone imaging agent
Technetium ( $^{99}\text{Tc}^{\text{m}}$ ) mertiatide injection	Diagnostic renal imaging agent
Indium ( $^{111}\text{In}$ ) pentetereotide (OctreoScan)	Detection of primary sites and metastases of Gastro-entero-pancreatic endocrine tumors
Indium ( $^{111}\text{In}$ ) tropolonate white blood cells	Imaging infection and inflammation
Iodine ( $^{123}\text{I}$ ) Iobenguane (mIBG)	Diagnosis of neuroendocrine tumors
Iodinated ( $^{125}\text{I}$ ) albumin injection	Plasma volume studies
Sodium iodide ( $^{131}\text{I}$ ) injection and solution	Diagnosis and treatment of thyroid carcinoma
Thallous ( $^{201}\text{Tl}$ ) chloride injection	Myocardial visualization

generally produced in a 'hot cell' with a dedicated, automated synthesis unit.

Nonimaging diagnostic procedures are also used, for example, in kidney and hematological investigations. Longer-lived isotopes with half-lives of many days are used in these studies (e.g., chromium ( $^{51}\text{Cr}$ )).

Radiopharmaceuticals for use in therapy should have a long retention time and be highly selective. Low-energy beta emitting isotopes are used (e.g., iodine ( $^{131}\text{I}$ )). Some high-energy beta emitters (yttrium ( $^{90}\text{Y}$ ), phosphorus ( $^{32}\text{P}$ )) and alpha emitters (bismuth ( $^{213}\text{Bi}$ )) are also used.

The other main area of use for radiopharmaceuticals is as a research tool in drug formulation and drug delivery systems. For instance, molecular imaging using PET tracers can help in 'proof of concept' in drug development studies.

The radionuclides used in radiopharmaceuticals (Table 2) are produced by one of three methods. In cyclotron production a target material is bombarded with an ion beam and various isotopes are formed depending on the target material. Radionuclides can also be produced in nuclear reactors where nuclear fission leads to the production of a great number of radionuclides, some of which can be used *in vivo* after complex purification processes. Some radionuclides are produced from the decay of a radioactive parent. The parent generally has a longer half-life than the daughter radionuclide. In a generator system the parent is immobilized on an inert substrate and the daughter can be washed off ('eluted'). Technetium ( $^{99}\text{Tc}^{\text{m}}$ ) is produced by this method from the decay of molybdenum ( $^{99}\text{Mo}$ ).

**Table 2** Examples of radionuclides used in clinical practice

<i>Use</i>	<i>Radionuclide</i>	<i>Production method</i>	<i>Radiation emitted</i>	<i>Half-life</i>
Diagnostic (a) Imaging	$^{11}\text{C}$	Cyclotron	$\beta +$	20 min
	$^{13}\text{N}$	Cyclotron	$\beta +$	10 min
	$^{15}\text{O}$	Cyclotron	$\beta +$	2 min
	$^{18}\text{F}$	Cyclotron	$\beta +$	110 min
	$^{67}\text{Ga}$	Cyclotron	$\gamma$	78 h
	$^{81}\text{Kr}^{\text{m}}$	Generator	$\gamma$	13 s
	$^{99}\text{Tc}^{\text{m}}$	Generator	$\gamma$	6 h
	$^{111}\text{In}$	Cyclotron	$\gamma$	67 h
	$^{123}\text{I}$	Cyclotron	$\gamma$	13 h
	$^{133}\text{Xe}$	Nuclear reactor	$\beta - \gamma$	5.3 days
	$^{201}\text{Tl}$	Cyclotron	X-rays	73 h
	$^{51}\text{Cr}$	Nuclear reactor	$\gamma$	28 days
	$^{57}\text{Co}$	Nuclear reactor	$\gamma$	270 days
	$^{125}\text{I}$	Nuclear reactor	X-rays	60 days
Therapy	$^{32}\text{P}$	Nuclear reactor	$\beta -$	14 days
	$^{89}\text{Sr}$	Nuclear reactor	$\beta -$	52.7 days
	$^{90}\text{Y}$	Generator	$\beta -$	64 h
	$^{131}\text{I}$	Nuclear reactor	$\beta - \gamma$	8 days
	$^{213}\text{Bi}$	Generator	$\alpha$	46 min



Radiopharmaceuticals are usually formulated as sterile, pyrogen-free injections. There are exceptions, such as oral preparations, gaseous preparations, and the isotope generators themselves. While the preparation of radiopharmaceuticals is subject to the same regulations as any other pharmaceutical, due to the short half-life of many of the isotopes, products are often released for use before many of the quality control aspects can be completed. It is important that good manufacturing practice and good laboratory practice should be strictly followed. In the European Community, products bearing a product license number indicate that they have been approved for clinical use by the regulatory bodies and are termed 'licensed'. Unlicensed products may be used but responsibility for their efficacy rests with the individual clinician. Other countries, e.g., the USA, have similar regulations. Monographs listing the quality parameters for many radiopharmaceuticals are given in the Pharmacopoeiae and products supplied to patients must adhere strictly to these. The remainder of this article will outline the tests that should be carried out.

## Activity Check

There are very strict limits on the amount of radioactivity that can be administered for each type of investigation to be performed. These limits are set by the legislative bodies in individual countries. Therefore, all radioactive preparations must be checked for activity before administration. Most radionuclide calibrators used in radiopharmacy are ionization chambers. Commercial calibrators have built in scaling factors for individual radionuclides that take into account the ionizing ability of the isotope and give a readout in the appropriate units (kBq or MBq, or mCi). However, this type of calibrator is not ideal for all radionuclides. Low-energy radiation, such as that produced by iodine ( $^{125}\text{I}$ ), may be attenuated before reaching the gas and the measurement may be inaccurate. Also, high-energy beta particles interact with the chamber wall and the measurement of activity is based on the Bremsstrahlung radiation produced. The activities dispensed in a radiopharmacy need to be measured to a precision of 2–5%. It is very important that the radionuclide calibrator used is working correctly and a strict quality assurance regime should be followed to ensure this. The calibrator is checked daily for accuracy against a long-lived reference sealed source (e.g., cobalt ( $^{57}\text{Co}$ ) or americium ( $^{241}\text{Am}$ )), there should also be an annual linearity check. The accurate determination of low activities of X-ray and gamma-emitting radionuclides can be carried out using a well scintillation counter, in which a scintillant crystal produces a flash of light

when struck by an ionizing particle or photon. The light produced is converted to an electrical signal by a photomultiplier, which can then be amplified and counted. Gamma and X-rays are best detected with crystals of thallium-activated sodium iodide ( $\text{NaI(Tl)}$ ). In a well counter the source is placed inside a well drilled in the center of a cylindrical crystal. Liquid scintillation counting is used in the measurement of beta emitters, such as tritium ( $^3\text{H}$ ) and carbon ( $^{14}\text{C}$ ).

## Radionuclide Identity

A radionuclide can be identified by measuring the half-life of the radionuclide and/or determining the energies of the emitted radiation.

The half-life is measured with a suitable detector (see above). There must be sufficient activity in the sample so that it can be measured for several half-lives, but limited to minimize count rate defects and effects such as dead time losses. Multiple measurements are made in the same geometrical conditions, for a time at least equal to three expected half-lives. A graph is drawn with the logarithm of the instrument response against time. The half-life is calculated from the graph and should not differ by more than 5% from the half-life stated in the Pharmacopoeiae.

The nature and energy of the radiation emitted may be determined by several procedures depending on the type of radiation emitted. Radionuclides that emit gamma rays or detectable X-rays can be analyzed using gamma spectrometry. The spectrum can be used to identify which nuclides are present in a source and in what quantities. Nuclides that emit alpha and beta particles can be analyzed using liquid scintillation counting and spectrometry. Beta particle emitters can also be analyzed by the construction of an attenuation curve. The radiation source is placed in front of a Geiger–Mueller counter. The count rate is measured and a succession of aluminum sheets of increasing thicknesses are placed between the source and the counter and the measurement repeated. This is continued until there is no reduction in counts by the addition of more sheets. This procedure is repeated for a standard source. The mass attenuation ( $m_m$ ) can be calculated for the sample and standard using the following equation:

$$m_m = \frac{\ln A_1 - \ln A_2}{m_2 - m_1}$$

where  $m_1$  is the mass per unit area of the lightest screen,  $m_2$  the mass per unit area of the heaviest screen,  $A_1$  the count rate for mass per unit area  $m_1$ , and  $A_2$  the count rate for mass per unit area  $m_2$ .

The value obtained for the sample should not vary by more than 10% from the standard source.

## Radionuclidic Purity

This is the amount of radioactivity due to the radionuclide concerned compared to the measured activity of the radiopharmaceutical preparation. It is usually expressed as a percentage. The relevant monographs in the Pharmacopoeiae give limits to the radionuclidic impurities allowed in each preparation. If significant levels of other radionuclides are present then biological distribution may be altered. Radionuclide impurities can occur as a result of the manufacturing process, for example, for nuclides produced by cyclotron there can be contaminants due to impurities in the target or by the energy of the reaction. For example, yttrium ( $^{86}\text{Y}$ ), for use in PET, is produced by the reaction (p, n) on a target of strontium ( $^{86}\text{Sr}$ ) using protons with energy of 16 MeV. If energy of 30 MeV is used then yttrium ( $^{85}\text{Y}$ ) is also produced, which decays to strontium ( $^{85}\text{Sr}$ ). This radionuclide has a long half-life and targets bone, having serious implications for patients. Impurities can also arise due to the presence of the parent nuclide of the stated nuclide when the stated nuclide is obtained by a separation technique such as a generator elution, for example, the presence of molybdenum ( $^{99}\text{Mo}$ ) in a solution of technetium ( $^{99}\text{Tc}^{\text{m}}$ ). The presence of molybdenum ( $^{99}\text{Mo}$ ) in a technetium ( $^{99}\text{Tc}^{\text{m}}$ ) radiopharmaceutical would be detrimental for patients due to the beta emission of this radionuclide and the long half-life (66 h) giving an increased radiation dose.

The radionuclidic purity can be determined by gamma spectrometry, or by the determination of the half-life. The half-life method is only a qualitative assessment but is useful for very short lived isotopes for use in PET, such as oxygen ( $^{15}\text{O}$ ) (half-life 2 min).

A version of the attenuation method can be used to determine the amount of molybdenum ( $^{99}\text{Mo}$ ) in a solution of technetium ( $^{99}\text{Tc}^{\text{m}}$ ) as these two radionuclides emit gamma radiation of very different energies. The activity of the sample is determined in an ionization chamber on the technetium setting. The activity measured will be due to technetium ( $^{99}\text{Tc}^{\text{m}}$ ) and molybdenum ( $^{99}\text{Mo}$ ). The sample is then placed in a lead canister with wall thickness of 6 mm and measured again on the molybdenum setting. The radiation emitted by technetium ( $^{99}\text{Tc}^{\text{m}}$ ) has energy of 140 keV, which is absorbed by the lead. The radiation emitted by molybdenum ( $^{99}\text{Mo}$ ) has energy of 740 keV and is only attenuated by a third. Any measured activity is therefore due to the molybdenum ( $^{99}\text{Mo}$ ) present. The proportion of molybdenum in

the sample can then be calculated. The limit allowed is 0.1% contaminating molybdenum ( $^{99}\text{Mo}$ ). This test should be performed daily on the first elution from a generator.

Due to differences in the half-lives of the different radionuclides that can be present in a preparation, the radionuclidic purity changes with time. The requirement of the radionuclidic purity must be fulfilled throughout the period of validity.

## Radiochemical Purity Determination

It is important to know that the majority of the radioactive isotope is attached to the ligand and is not free or attached to another chemical entity as these forms may have a different biodistribution. This is termed the radiochemical purity (RCP). In diagnostic scanning the different biodistribution of contaminating radioactive chemicals could interfere with the clinical diagnosis by obscuring the region of interest and interfering with the interpretation of the scan. This could result in the patient returning to the department for a repeat scan with all the cost implications for the department and a repeat radiation dose for the patient, or, more seriously, to a misinterpretation of the images. Abnormal biodistribution can also be due to other causes such as patient medication; therefore, determining the RCP could help to clarify an abnormal scan by ruling out defective product. In radiopharmaceuticals used for therapy a low RCP could mean an unacceptable radiation dose to healthy organs and tissues.

For most radiopharmaceuticals the lower limit of RCP is 95%, that is, at least 95% of the radioactive isotope must be attached to the ligand. RCP determination can be carried out by a variety of chromatographic methods.

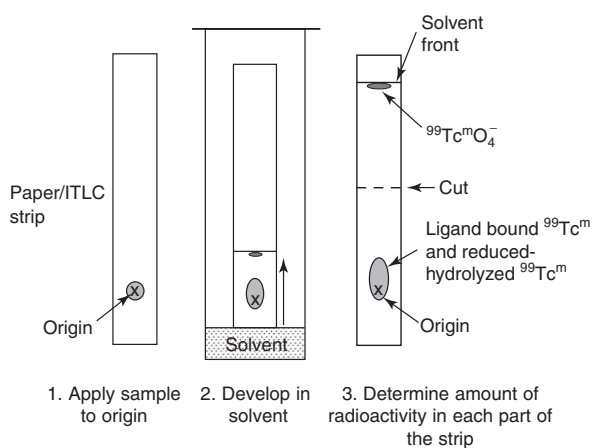
### Paper and Thin-Layer Chromatography

In these methods, a small drop of the radiopharmaceutical is put onto the bottom of a strip of support medium (e.g., paper, silica gel-coated sheets). The strip is put into a tank containing a small amount of solvent. The solvent migrates up the strip. The components of the radiopharmaceutical are separated according to the solubility in the solvent and adsorption to the support medium (Figure 1). Detection of the radioactivity in the strip can be carried out in a number of ways:

1. The simplest method is to cut the strip and count the activity in the sections in a radionuclide calibrator or a well scintillation counter. The percentage of activity in each section can then be determined. This method does have several limitations, however.

Radionuclide calibrators are inaccurate for samples of low activity due to the lower level of detectability and the accuracy of the calibrator at the lower range setting. Well scintillation counters should be avoided for samples of high activity, as these can exceed the count rate capabilities due to the resolving time of the detection system.

2. The strip can be imaged under a gamma camera. Regions of interest can be drawn around the areas of radioactivity and the percentage of counts in each region can be determined. This method has the advantage of imaging the whole chromatography strip enabling artifacts to be seen; however, it is not practicable for most hospital departments due to the cost of camera time.

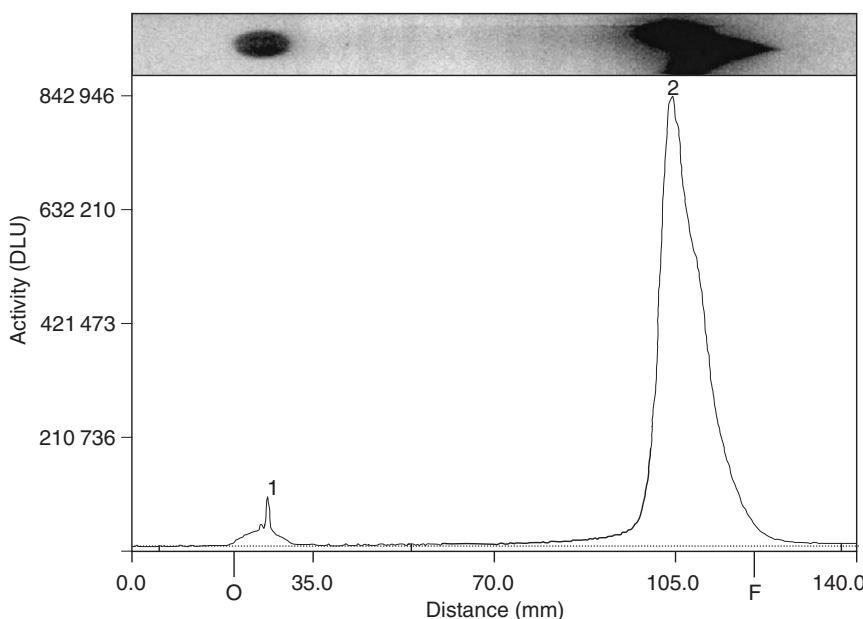


**Figure 1** Determination of radiochemical purity by paper and instant thin-layer chromatography (ITLC).

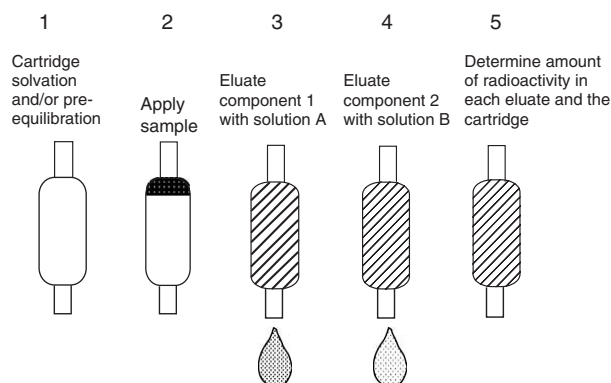
3. The strip can be imaged using a radiochromatogram scanner. A radiochromatogram scanner uses a sodium iodide detector to detect the radioactive emission. If the scanner is linked to an integrator then quantification of the peaks can be carried out. This equipment is not suitable for counting emissions from some isotopes (e.g., chromium ( $^{51}\text{Cr}$ ) or iodine ( $^{125}\text{I}$ )) due to statistics associated with counting low activities.

4. Analysis of the strip can be performed using storage phosphor imaging. A phosphor screen accumulates a latent image of the distribution of radioactivity on the strip. Scanning the screen with a laser allows the location and intensity of radioactivity to be analyzed and stored for future reference. Results are shown as an image of the strip. Regions of interest can be drawn and integration carried out to determine activity in each region. **Figure 2** shows an image of a chromatogram of sodium chromate ( $^{51}\text{Cr}$ ) solution obtained using a phosphor imager. By varying the time of exposure of the strip to the phosphor screen, radiopharmaceuticals containing many different isotopes, or of different activities, can be analyzed.

One of the biggest drawbacks to paper and thin-layer chromatography (TLC) methods of determining RCP is the resolving power of the methods. Most methods commonly used will only resolve one component and so two or three methods may be needed to identify all the major contaminants in a product. Time can also be a limiting factor with some methods taking 20–30 min to develop, or even longer.



**Figure 2** Image of a chromatogram of sodium chromate ( $^{51}\text{Cr}$ ) solution obtained using a phosphor imager.



**Figure 3** Determination of radiochemical purity using solid-phase extraction.

### Solid-Phase Extraction

A variety of bonded silica sorbents are available packed into disposable cartridges or columns. The sorbent will selectively retain specific types of chemical compounds from the sample loaded on the column. There are many different sorbents available and many suppliers and manufacturers of solid-phase extraction (SPE) columns and cartridges. **Figure 3** shows a typical extraction procedure using SPE for radiopharmaceuticals. Components can be selectively eluted by careful choice of solvents. Eluates are collected and have sufficient activity that they can be counted in a radionuclide calibrator. The procedure takes ~5 min ensuring that RCP determination can be carried out before administration of the patient dose.

### High-Performance Liquid Chromatography

TLC RCP methods may not be sufficient to identify all the compounds that are present in a product. High-performance liquid chromatography (HPLC) has a higher sensitivity and resolving power than simple TLC methods. Gamma emitters are detected using a well scintillation counter connected to a rate meter. Other detectors (ultraviolet or refractive index) can be connected in series allowing simultaneous identification of compounds.

It should not be necessary to perform HPLC on radiopharmaceuticals reconstituted from licensed cold kits. It is useful to have techniques available for the purpose of eliminating a cause of any abnormal patient scan. For radiopharmaceuticals prepared 'in-house' or novel compounds for research purposes, an HPLC method for estimating RCP is essential. It should be noted that HPLC does not detect colloidal contaminants and that this should be estimated using TLC methods.

### Electrophoresis

Electrophoresis separates components in a sample according to the charge and size of the molecules. The most frequently used electrophoresis method in radiopharmacy is in the RCP determination of iodinated ( $^{125}\text{I}$ ) albumin injection. A drop of iodinated ( $^{125}\text{I}$ ) albumin injection containing at least 0.5 mg albumin is placed on a strip of filter paper ( $5 \times 30$  cm). Electrophoresis is carried out using barbitone-acetate buffer at 500 V for 1 h.

### Chemical Purity

Chemical purity requires the identification and quantification of individual chemical constituents or impurities in a radiopharmaceutical preparation. A routine test for chemical purity is performed on technetium ( $^{99}\text{Tc}^{\text{m}}$ ) sodium pertechnetate solution eluted from a generator. Technetium ( $^{99}\text{Tc}^{\text{m}}$ ) sodium pertechnetate solution may be contaminated with aluminum, which originates from the alumina bed of the generator column. The presence of aluminum can interfere with the preparation of some technetium ( $^{99}\text{Tc}^{\text{m}}$ ) colloidal preparations and also with the labeling of red blood cells with technetium ( $^{99}\text{Tc}^{\text{m}}$ ), causing their agglutination. The aluminum in the eluate can be detected using a simple paper test that is commercially available. Filter paper is impregnated with a color complexing agent. A standard solution of aluminum ( $10 \mu\text{g ml}^{-1}$ ) is supplied. A spot of the standard solution causes a color change in the paper. A spot of generator eluate is compared to the standard spot. If the color is more intense in the eluate spot then the eluate contains more than  $10 \mu\text{g ml}^{-1}$  aluminum and implies a lack of stability in the column; consequently the eluate should be discarded.

It should not be necessary to perform other chemical tests on licensed radiopharmaceuticals. Unlicensed radiopharmaceuticals should be checked for chemical purity to ensure the quality of the product. This would include the quantification of the normal constituents of a labeling kit, i.e., the ligand and reductant. Synthesis precursors or catalysts used in the preparation should be tested for. Commonly, this can be carried out using HPLC. Gas chromatography methods are used to test PET tracers for residual solvents used in the synthesis of these agents.

### Particle Sizing and Counting

Particulate radiopharmaceuticals rely on a specific particle size to be effective. For example, technetium ( $^{99}\text{Tc}^{\text{m}}$ )-macroaggregates of human serum albumin (MAA) (for use in lung imaging) have a particle size

range of 10–100  $\mu\text{m}$ . Particles in this particular size range are trapped in the pulmonary capillary bed, which can then be imaged. Smaller particles will not be trapped and will remain in circulation giving a high background, or will be filtered out by the reticuloendothelial system giving high counts in the liver and spleen. Larger particles will not enter the smallest capillaries and so the whole lung will not be imaged. The proportion of small particles in the preparation can be determined by filtration. The preparation is filtered through a 3  $\mu\text{m}$  polycarbonate filter unit. The eluate is collected and the activity in the eluate and on the filter is counted. Activity in the eluate is due to free technetium ( $^{99}\text{Tc}^{\text{m}}$ ) or particles less than 3  $\mu\text{m}$  in size. There should be less than 5% of the activity in the eluate. This method can also be used to determine the RCP of the product.

Particle size and number can be determined by light microscopy using a hemocytometer.

## Sterility Testing

Radiopharmaceuticals for parenteral administration should be sterile. Preparation under aseptic conditions should ensure this; however, conditions and operator dispensing technique can be checked by carrying out sterility testing of products.

Sterility testing of short-lived radiopharmaceuticals can be carried out by inoculating broth culture media with a small volume of the preparation (0.3 ml) under aseptic conditions. The broth can then be incubated, for about 2 weeks, and examined for growth. If the facilities are available this can be performed in the laboratory; however, this is not

possible for most hospital radiopharmacies. In this case the inoculated broths are stored at room temperature under suitable shielding until the radioactivity has decreased enough for the samples to be transported to a quality control testing laboratory. An alternative to this method is filtration of the sample under aseptic conditions using a 0.2  $\mu\text{m}$  filter. The filter is then cultured in the broths mentioned above.

The results of sterility testing are retrospective and are chiefly a control of quality of production.

## Pyrogenicity Testing

Bacterial endotoxins (pyrogens) are polysaccharides from bacterial membranes. They are water soluble, heat stable, and filterable. If they are present in a preparation and administered to a patient they can cause fever and also leukopenia in immunosuppressed patients. To minimize the chances that pyrogens are present it is important that preparations are manufactured and dispensed under aseptic conditions and that all consumables and equipment used have been heat treated and known to be pyrogen free. Most licensed products are guaranteed pyrogen free. However, the Pharmacopoeiae state that certain radiopharmaceutical preparations are required to comply with the test for pyrogens and it may be necessary to test for pyrogens in unlicensed products or intermediates used in manufacture. This test can either be performed in rabbits or a test for bacterial endotoxin employing limulus amoebocyte lysate may be used. These tests can be carried out in the radiopharmacy or by a quality control testing laboratory.

**Table 3** Quality control tests required for radiopharmaceuticals manufactured by different methods

<i>Test</i>	<i><math>^{99}\text{Tc}^{\text{m}}</math> radiopharmaceuticals from licensed kits</i>	<i>Ready-to-use licensed products</i>	<i>PET radiopharmaceuticals</i>	<i>Research and unlicensed products</i>
Activity check	Each dose	Each dose	Each dose	Each dose
Radionuclidic identity	No	No	Every batch	No
Radionuclidic purity	Daily – first generator eluate	No	Every batch	Every batch
Radiochemical purity	Monthly or new batch of cold kit	No	Every batch	Every batch
Chemical purity	First and last generator eluate only	No	Every batch	Every batch
Particle sizing and counting	$^{99}\text{Tc}^{\text{m}}$ -MAA at regular intervals	No	No	No (unless particulate)
Sterility	Weekly	No	At regular intervals	Every batch
Pyrogenicity	No	No	At regular intervals	Every batch
pH	No	No	Every batch	Every batch
Visual examination	Each dose	Each dose	Each dose	Each dose
Biodistribution	No	No	No	Every batch

## pH

The pH of the preparation should be in the physiologically acceptable range (5.5–8) or at the optimum pH for the stability of the preparation. It is especially important to check the pH of PET radiopharmaceuticals as preparative operating procedures may include extreme pH conditions that need neutralization. pH can be checked by placing a drop of solution on pH paper. Reading the paper is done by referring to a colorimetric pH scale. Micro-pH meters are available that can read a drop of liquid (10 µl) placed on the electrodes.

## Visual Examination

As with all pharmaceutical preparations, radiopharmaceuticals should undergo a visual examination. The examination should take into account radiation protection issues for the operator and be conducted as quickly as possible with the preparation behind suitable shielding. Vials should be examined for insecure closures, cracks, glass particles in the liquid, and particulate contamination. Syringes should also be examined for particulate contamination.

## Biodistribution

Checking the biodistribution of radiopharmaceuticals is not necessary with licensed products and is not usually undertaken in most hospitals; however, the monographs for some agents give requirements for biodistribution studies. In development work with new agents, studies into the biodistribution are essential. In biomedical research radiopharmaceuticals are used to determine the biodistribution of the 'cold' drug under study. Biodistribution of radiopharmaceuticals can be checked by injecting animals and imaging under dedicated cameras especially designed for such work. The animals can be killed and dissected at suitable time points and organs, blood, urine collected and counted and the

biodistribution determined. Whole body slices can be made and the distribution of radioactivity can be visualized using autoradiography. It should be noted that the biodistribution in animals (even other primates) can be very different from that obtained in humans.

It is not necessary to perform all of the above tests on every product. **Table 3** summarizes the number and frequency of testing required for products manufactured by different methods.

*See also: Radiochemical Methods: Overview; Natural and Artificial Radioactivity; Radionuclide Monitoring; Technetium; Radon; Uranium; Radiotracers; Radio-Reagent Methods; Radioreceptor Assays; Gamma-Ray Spectrometry; Food and Environmental Applications.*

## Further Reading

- British Pharmacopoeia* (2002). London: The Stationery Office.
- European Pharmacopoeia* (2002), 4th edn. Strasbourg: Council of Europe.
- Fleming JS and Perkins AC (eds.) (2000) *Targeted Radiotherapy*. York: Institute of Physics and Engineering in Medicine.
- Perkins A and Frier M (eds.) (1999) *Nuclear Medicine in Pharmaceutical Research*. London: Taylor & Francis.
- Saha GB (1998) *Fundamentals of Nuclear Pharmacy*, 4th edn. New York: Springer.
- Sampson CB (ed.) (1999) *Textbook of Radiopharmacy: Theory and Practice*, 3rd edn. Amsterdam: Gordon and Breach.
- Sweetman SC (ed.) (2002) *Martindale: The Complete Drug Reference*, 33rd edn. Wallingford: Pharmaceutical Press.
- The United States Pharmacopoeia and the National Formulary* (2003). Rockville: The United States Pharmacopoeial Convention.
- Welch MJ and Redvanly CS (eds.) (2003) *Handbook of Radiopharmaceuticals: Radiochemistry and Applications*. Chichester: Wiley.

# RADON

*See RADIOCHEMICAL METHODS: Radon*



# RAMAN SPECTROSCOPY

Contents

**Instrumentation**

**Near-Infrared**

**Surface-Enhanced**

## Instrumentation

**M J Pelletier and C C Pelletier**, California Institute of Technology, Pasadena, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Raman spectroscopy is a powerful and versatile technique that is rapidly gaining popularity both in the laboratory and in the field. There are many variations of Raman spectroscopy including spontaneous Raman scattering, coherent anti-Stokes Raman scattering, Raman optical activity, hyper-Raman spectroscopy, and several types of Raman gain spectroscopy. Each of these has unique requirements for instrumentation. The simplest and most popular variation is spontaneous Raman scattering. The instruments used to measure spontaneous Raman scattering can be classified as filter instruments, dispersive spectrographs, or interferometers. This article describes filter instruments and dispersive spectrographs used for spontaneous Raman scattering.

Spontaneous Raman instruments illuminate a sample with monochromatic light and determine the intensities of new wavelengths of light created by the sample. Some important guiding principles for Raman instrument design can be drawn directly from the basic physics of Raman scattering. Raman scattering cross-sections are on the order of  $10^{-31} \text{ cm}^2$  molecule per steradian per, or  $\sim 15$  orders of magnitude smaller than those for fluorescence. A typical Raman sample will produce one Raman scattered photon from  $10^6$  to  $10^8$  excitation photons. Intense light sources and efficient collection of Raman photons are required in order for an adequate number of Raman photons to be detected. Even so, the number of Raman photons produced is usually small enough for measurements to be shot-noise limited, so the signal-to-noise ratio increases with the square root of the number of Raman photons detected.

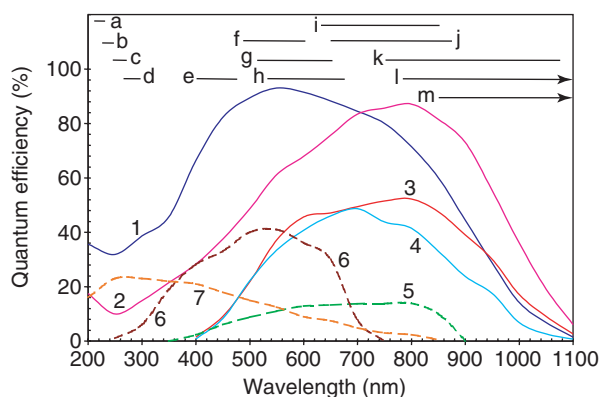
The weak Raman signal is easily overwhelmed if more intense light from other sources reaches the

detector. Excitation light elastically scattered (no wavelength change) from the sample is orders of magnitude more intense than the Raman signal and must be severely attenuated. Vanishingly small intensity at the wrong wavelength contaminating the excitation light, if present, will elastically scatter from the sample and can severely distort the Raman spectrum. If the sample or impurities in the sample fluoresce, the fluorescence intensity can easily overwhelm the Raman spectrum. Even room lighting elastically scattered from the sample can distort or obscure the Raman spectrum.

Raman scattering intensity is inversely proportional to the fourth power of the emission wavelength. Decreasing the excitation wavelength, therefore, rapidly increases the Raman signal. Unfortunately, decreasing the excitation wavelength also increases the likelihood that intense fluorescence from the sample will obscure the Raman spectrum. Raman measurements are often made in the near-infrared spectral region despite the loss in sensitivity in order to avoid fluorescence. Very low excitation wavelengths, below  $\sim 250 \text{ nm}$ , largely avoid fluorescence because most fluorescence that does occur is outside the spectral region of the Raman spectrum. Sample absorption and photolytic damage become more of a problem at these short wavelengths.

The spectral range, in nanometers, needed to include the entire Raman spectrum decreases with decreasing excitation wavelength, as illustrated in **Figure 1**. For example, with  $800 \text{ nm}$  excitation, a  $4000 \text{ cm}^{-1}$  Raman spectrum extends to  $1176 \text{ nm}$ , a  $376 \text{ nm}$  spectral range. At  $200 \text{ nm}$ , the same  $4000 \text{ cm}^{-1}$  Raman spectrum extends to  $217 \text{ nm}$ , only a  $17 \text{ nm}$  spectral range. To maintain the same Raman spectral resolution in  $\text{cm}^{-1}$ , shorter excitation wavelengths require greater spectral resolution (expressed in nanometers) from the Raman instrument. A Raman spectral resolution of  $5\text{--}10 \text{ cm}^{-1}$  is often adequate for condensed-phase samples, but shifts as small as  $0.1 \text{ cm}^{-1}$  can sometimes be used to extract important chemical information.

The remainder of this article examines the main subsystems of a Raman instrument: the laser source, the sample interface, the spectrometer, and the



**Figure 1** Quantum efficiencies of Raman detectors shown with the Raman spectral regions for various lasers. Solid curves represent CCD detectors: 1, back-illuminated; 2, back-illuminated, deep depletion; 3, front-illuminated, deep depletion; 4, front-illuminated. Dashed curves represent photomultiplier tubes: 5, GaAs photocathode; 6, GaAsP photocathode; 7, multialkali photocathode. The solid lines at the top indicate the spectral region of a 0–4000  $\text{cm}^{-1}$  Raman spectrum excited with various laser wavelengths: a, 213 nm Nd:YAG fifth harmonic; b, 224.3 nm HeAg; c, 244 nm argon-ion second harmonic; d, 266 nm Nd:YAG fourth harmonic; e, 400 nm GaN; f, 488.0 nm argon-ion; g, 514.5 nm argon-ion; h, 532 nm Nd:YAG second harmonic; i, 632.8 nm HeNe; j, 647.4 nm krypton-ion; k, 752 nm krypton-ion; l, 785 nm stabilized diode. 3650  $\text{cm}^{-1}$  and higher is beyond 1100 nm; m, 850 nm stabilized diode. 2675  $\text{cm}^{-1}$  and higher is beyond 1100 nm.

detector. Special emphasis is placed on the sample interface because it is the primary differentiator between different Raman instruments. It is also the part of the Raman instrument that the user interacts with the most. The article concludes by considering the integration of these subsystems into Raman instruments for the research laboratory, for routine analysis, for process control, and for field-portable applications.

## Laser Sources

Lasers with adequate power for Raman spectroscopy are now available at nearly any wavelength from the vacuum ultraviolet to well into the near-infrared spectral region. Several important laser wavelengths are included in Figure 1. The laser power used for a typical Raman measurement tends to increase with increasing wavelength, both because laser-induced damage thresholds tend to be greater and Raman sensitivities tend to be lower at longer excitation wavelengths. For example, most Raman measurements can be made with a few milliwatts of average ultraviolet laser power, a few tens of milliwatts of visible laser power, and a few hundreds of milliwatts of laser power at 1064 nm. It is not uncommon,

though, to use much greater laser power in order to enhance sensitivity when the sample can tolerate it. Raman measurements in aqueous solution often use several hundred to a few thousand milliwatts of laser power. Samples, and even optical elements, are often less tolerant to pulsed lasers, especially low-repetition-rate short-pulse lasers.

Most lasers have a preferred, stable, linear polarization. Those that do not are a potential source of measurement artifacts. Raman spectra can change with changing polarization of the excitation source. A randomly polarized laser does not generally provide equal intensity for all polarizations as is frequently assumed. To the extent that the polarization of the excitation light reaching the sample is not controlled, the resulting Raman measurements may not be reproducible.

Atomic gas lasers used for Raman spectroscopy, such as argon-ion lasers and helium–neon lasers, have a narrow linewidth and a fixed frequency. Their spectral linewidth and frequency uncertainty have a negligible effect on most Raman spectra. The frequency of a solid-state laser is more affected by temperature and optical configuration, so must be determined experimentally. For example, a frequency-doubled Nd:YAG laser changes wavelength with changing temperature at a rate of 0.003 nm per  $^{\circ}\text{C}$ . Tunable solid-state lasers can have a much larger operating frequency range of tens to hundreds of nanometers.

The water-cooled argon-ion laser has been a favorite laser for laboratory Raman spectroscopy. It provides several green, blue, and near-ultraviolet laser lines with hundreds to thousands of milliwatts of optical power. The resulting Raman wavelengths are well matched to the optimum sensitivity of photomultiplier tube detectors. Air-cooled argon-ion lasers provide hundreds of milliwatts at 514.5 or 488.0 nm in a smaller, portable package at a lower cost. The water-cooled krypton-ion laser complements the water-cooled argon-ion laser with laser wavelengths in the near-infrared, red, and violet regions of the spectrum.

The 632.8 nm helium–neon (HeNe) laser is also very popular for Raman spectroscopy because it is small, portable, mature, and very inexpensive. The resulting Raman wavelengths are well matched to the optimum sensitivity of charge-coupled device (CCD) detectors. HeNe lasers used for Raman spectroscopy typically deliver 5–30 mW of optical power. Sometimes they also sporadically deliver a weaker laser-like beam at 650 nm that can cause a huge line in the Raman spectrum at 418  $\text{cm}^{-1}$ . The 650 nm radiation is due to an intracavity pumped laser Raman transition.

Diode lasers in the near-infrared, red, and, more recently, violet spectral regions are popular excitation sources for Raman spectroscopy. An individual diode laser can lase over a range of several nanometers, but when operated at low power, will often lase at a single narrow-band frequency. Careful control of the diode laser temperature and drive current, along with elimination of back reflections into the laser diode, can stabilize the lasing frequency long enough to collect high-quality Raman spectra. Most diode lasers used for Raman spectroscopy rely on external cavity stabilization (optical feedback at the desired wavelength) to maintain a constant laser frequency indefinitely. Diode lasers are available that operate in a single spatial mode. Light from these diode lasers can be focused to a diffraction-limited spot. Diode lasers are also available with a larger emitting surface in order to provide higher optical power, but these lasers, called broad-area laser diodes, cannot be focused as tightly.

Continuous-wave diode-pumped Nd:YAG lasers provide 1064 nm laser light for Fourier transform Raman (FT-Raman) instruments. They are often frequency doubled to provide 10–100 mW of 532 nm light for other types of Raman instruments. The weaker 1074 nm laser line sometimes lases at much lower intensity than the 1064 nm line. The resulting 537 nm light present in a 532 nm laser creates a large artifact band near  $175\text{ cm}^{-1}$  in the Raman spectrum. Over 10 000 mW of 532 nm light is available from standard larger lasers, however. Q-switched (pulsed) Nd:YAG lasers provide enormous peak optical power at 1064 nm that can be converted into many shorter wavelengths using various nonlinear optical techniques. The high peak power can easily damage samples, however.

The titanium-sapphire laser is perhaps the ultimate near-infrared laboratory laser for Raman spectroscopy. It is a continuous wave laser that can deliver thousands of milliwatts of laser light and is continuously tunable from below 700 to above 1000 nm. It provides a narrow bandwidth with a high-quality spatial mode. The spontaneous emission from the titanium-sapphire crystal must be filtered from the laser beam.

Most of the original deep-ultraviolet Raman spectroscopy relied on laser light derived from a Q-switched Nd:YAG laser. Fourth and fifth harmonic generation provided light at 266 and 213 nm, respectively. The combined use of optical harmonic generation and dye-laser pumping provided a laser system continuously tunable down to 217 nm. The use of a Raman shifter, or Raman laser, with the 266 nm fourth harmonic of the Nd:YAG laser produced wavelengths of 240, 218, and 199 nm. All of

the wavelengths derived from a Q-switched Nd:YAG laser had low average power and high peak power pulses, so samples could easily be damaged. The spatial beam quality was often poor, making it difficult to effectively deliver the optical power to the sample. More recently intracavity-frequency-doubling of water-cooled argon and krypton lasers has provided continuous-wave deep-ultraviolet power of tens to hundreds of milliwatts with excellent spatial beam quality at several discrete wavelengths. The most popular wavelengths are 257.2, 244.0, 229.0, and 203.4 nm. Hollow cathode ion lasers are a type of pseudocontinuous wave deep ultraviolet source for Raman spectroscopy. They are small, inexpensive, and rugged. They are essentially HeNe lasers with the neon replaced by silver or copper formed dynamically by sputtering. The HeAg laser produces a few milliwatts of 224.3 nm laser light and the NeCu laser produces several milliwatts of 248.6 nm laser light. They are operated in a pulsed mode, but their duty cycle is much higher than that of a Q-switched Nd:YAG laser so sample damage is much less likely.

Microchip lasers also need to be mentioned because they are likely to have a big impact on inexpensive portable Raman instruments in the future. A microchip laser consists of a thin plate (typically 0.3 mm thick) of laser gain material (usually Nd:YAG or Nd:YVO<sub>4</sub>) with laser mirrors deposited on both sides. Axial pumping creates a positive thermal lens that stabilizes the otherwise flat-flat laser cavity. The laser cavity is so short that only one longitudinal mode can lase, making the laser inherently very narrow band. Additional plates can be added between the mirrors for passive Q-switching (Cr:YAG plate) or intracavity frequency doubling. The most popular frequency doubling crystal for the generation of 532 nm light is KTP (potassium titanyl phosphate, KTiOPO<sub>4</sub>). Potassium niobate, KNbO<sub>3</sub>, is commonly used to generate 473 nm light. Passively Q-switched microchip lasers can generate a wide range of wavelengths using harmonic generation and optical parametric oscillation. They can be mass-produced inexpensively using wafer bonding and cutting technology developed in the semiconductor industry. A Q-switched, frequency doubled microchip laser, complete with its pump laser, can occupy less than 1 cm<sup>3</sup> of volume.

## Spectrometers

The Raman spectrometer filters out the unwanted laser light and delivers the proper Raman wavelengths to the proper detector(s). Once most of the laser light has been removed, the Raman light can be

processed using either a monochromator or a spectrograph. A monochromator transmits a desired Raman wavelength to the detector and rejects all the rest of the Raman photons. Spectrographs direct different Raman wavelengths to different detector elements. Spectrographs make much better use of Raman photons from the sample, but tend to have poorer stray light performance than monochromators. Good laser line rejection filtering minimizes the importance of spectrograph stray light performance, making spectrographs preferable to monochromators in most cases.

### Laser Line Rejection Filters

The best filter commonly used with a Raman spectrograph for rejecting the laser light without rejecting low-frequency Raman photons is a zero dispersion double spectrograph. This filter separates the different wavelengths of light spatially, removes the laser frequency, and puts the remaining wavelengths back together at its exit slit, as shown in **Figure 2**. A zero dispersion double spectrograph is a tunable filter. The laser rejection wavelength can be changed by simply moving the diffraction gratings. The zero dispersion spectrograph is complex and typically has a relatively low transmission of  $\sim 10\%$ .

Single element laser line rejection filters reject more of the low-frequency Raman spectrum, but

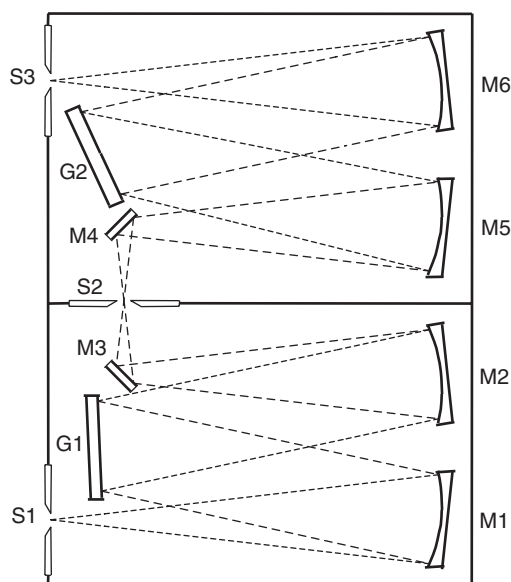
they tend to have a higher transmission, are simpler to set up and maintain, and are less expensive than a zero-dispersion spectrograph. The most common single element laser line rejection filters are thin-film interference filters, volume holographic filters, and rugate filters. Of the three, volume holographic filters provide the best laser line attenuation and least distortion of the Raman spectrum. Thin film interference filters are the least expensive. All three single element filters can be tuned over a limited wavelength range by tilting relative to the optic axis. Tilting causes the filter to have different transmission spectra for the two polarizations of light, which degrades the performance of the filter. Tilt angles greater than  $10^\circ$  are generally not recommended. As a result, a different single element laser line rejection filter is usually needed for each laser wavelength.

### Raman Monochromators

A Raman monochromator can be a diffraction grating instrument, two or three diffraction grating instruments in series (double or triple monochromator), or a single element bandpass filter. A triple monochromator is an extremely flexible instrument, but it is very complex and it uses Raman photons very inefficiently. All Raman photons outside the monochromator pass-band are lost. The first two monochromators serve as an excellent laser line rejection filter, though.

A thin film interference bandpass filter is often the monochromator of choice for simple gas-phase samples where low spectral resolution is acceptable. It is simple, inexpensive, and usually has high transmission. Multiple filters are often used together with multiple detectors in order to measure several Raman spectral regions simultaneously. Since the passband of a thin film interference filter shifts to shorter wavelengths with increasing angle of incidence, it is even possible to collect a full Raman spectrum by sequentially using a combination of tilt angles from different interference filters under computer control.

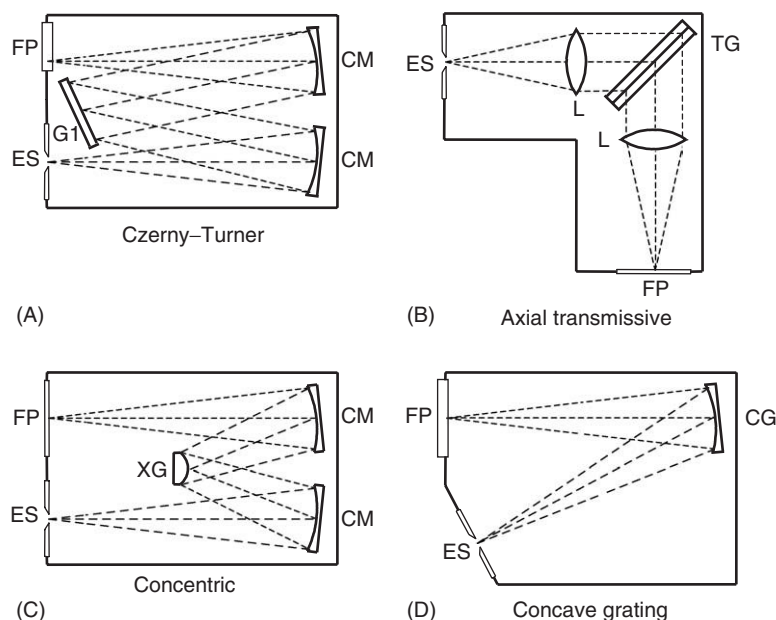
Tunable single element bandpass filters have also been used as Raman monochromators. The most popular are acousto-optic tunable filters and liquid crystal tunable filters. These filters are electrically tunable and are well suited to Raman imaging.



**Figure 2** Zero dispersion double spectrograph as a laser line wavelength rejection filter. Light passes through entrance slit S1 and forms a dispersed image at intermediate slit 2. Slit 2 is very wide and passes much of the spectrum, but not the laser wavelength. Grating G2 reverses the dispersion of grating G1 and focusing mirror M6 brings all the wavelengths back together at exit slit S3. M1, M2, M3, and M4 are concave mirrors. M3 and M4 are flat mirrors.

### Raman Spectrographs

Most modern laboratory Raman instruments use spectrographs with multielement detectors. **Figures 3** and **4** show optical diagrams of the more important spectrographs used for Raman spectroscopy. The Czerny–Turner spectrograph and its many variants are the most popular. Concave grating spectrographs



**Figure 3** One-dimensional spectrographs used for Raman spectroscopy. Several variations on these designs are available. FP, spectrograph exit focal plane; ES, entrance slit; CM, concave mirror; TG, transmission diffraction grating; L, multielement lens; XG, convex diffraction grating; CG, concave diffraction grating.

are advantageous for deep-ultraviolet work since there is only one optical element between the entrance slit and the detector. Off-axis imaging between the entrance slit and the detector for both the Czerny-Turner and the concave grating spectrograph degrades rapidly with decreasing  $f$ -number. A low  $f$ -number is often desirable, though, in order to maximize the Raman light intensity that can be delivered to the detector from a large sample area or optical fiber. Excellent imaging can be maintained at low  $f$ -number by replacing the concave mirrors of the Czerny-Turner spectrograph with multielement lenses. The stray light performance of the lenses is vastly inferior to that of mirrors, but is acceptable if a good laser line rejection filter precedes the spectrograph. Replacing the reflection grating with a transmission grating allows compact positioning of the optics with minimal vignetting. The result of these changes is the axial transmissive spectrograph shown in Figure 3. Concentric spectrographs are another approach to improve the imaging performance of a Czerny-Turner spectrograph at low  $f$ -number. They use a convex diffraction grating in a concentric configuration to correct optical aberrations.

Two-dimensional detectors provide the opportunity to divide a Raman spectrum into segments that can be stacked on top of each other at the detector, thereby increasing the spectral resolution for a given spectral range. Figure 4 shows two different ways to implement this concept. Raman instruments based

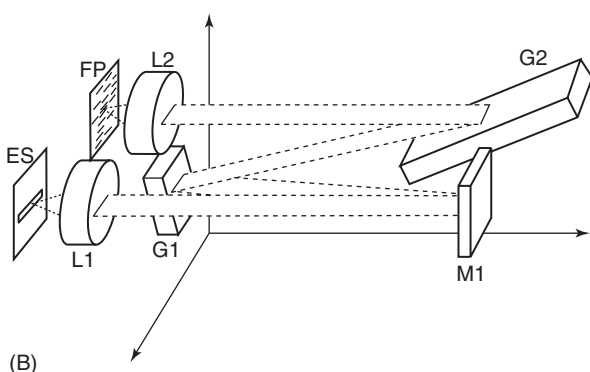
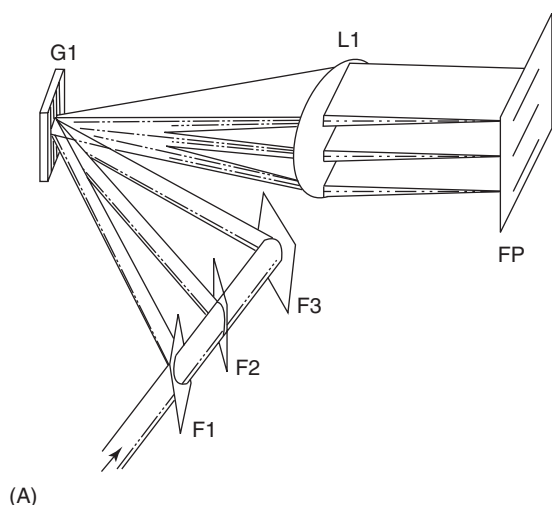
on echelle spectrographs and holoplex diffraction gratings are commercially available.

Optical elements not at normal incidence to both incoming and outgoing light generally have different properties for horizontally and vertically polarized light. As a result, spectrographs, monochromators, and even simple filters have a polarization-dependent transmission that can distort Raman spectra. Most Raman instruments provide a means to make the polarization of light entering the spectrograph independent of the polarization of the Raman photons from the sample in order to eliminate these distortions. One long-established way to do this is to use a stepped birefringent wedge (polarization scrambler) in the collimated space of the collection optics just prior to the spectrograph entrance slit.

## Sample Interface

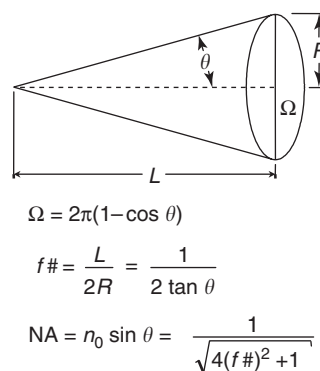
The sample interface brings together the laser illumination and the spectrometer field of view at the desired location on the sample. The three most common sample interfaces are a sample compartment, a fiber optic probe, and a Raman microscope. Common to all three interfaces are the needs to condition the laser beam that illuminates the sample and to collect light from the sample.

Laser illumination usually needs to be spectrally filtered and focused before it reaches the sample. Spectral filtering removes plasma lines, spontaneous



**Figure 4** Two-dimensional spectrographs used for Raman spectroscopy. (A) combination of a filter spectrograph and a diffraction spectrograph. F1, F2, and F3 are dichroic beam splitters; G1 is a diffraction grating; L1 is a camera lens; FP is the exit focal plane. (B) Echelle spectrograph for Raman spectrometry. ES is the entrance slit; L1, L2 are camera lenses; G1 is a diffraction grating; G2 is an echelle diffraction grating; FP is the exit focal plane. A two-dimensional spectrograph can also be made by using a holoplex transmission grating in the axial transmissive spectrograph shown in **Figure 3**.

fluorescence from the laser gain medium, or other unwanted light generated by the laser. Focusing the laser on the sample reduces the size of the Raman emitting object that must be imaged through the spectrometer. Raman emitting surfaces larger than 50 to 100  $\mu\text{m}$  place much greater demands on collection optics and spectrometers, and usually increase the size and cost of the spectrometer optics. Laser beams focused directly on the sample are usually 1–50  $\mu\text{m}$  in diameter. The resulting optical power density is high enough to sometimes cause excessive sample heating at that spot and even sample damage. Optical power density can be reduced without loss of sensitivity by illuminating the entire region of the sample that is imaged through the spectrometer entrance slit.



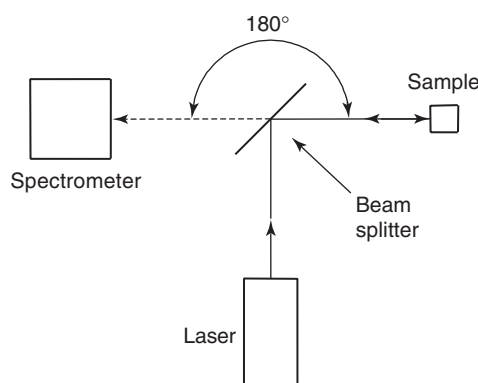
**Figure 5** Relationships between  $f$ -number, numerical aperture (NA), and solid angle. The variable  $n_0$  is the refractive index of the medium.

The collection optics image light from the sample onto the spectrometer entrance slit. The solid angle of light from a point on the sample to the collection optic surface is often expressed as an  $f$ -number ( $f\#$ ), a numerical aperture (NA), or a solid angle (steradians). The relationship between these is illustrated in **Figure 5**. Increasing the solid angle of collection delivers more Raman photons to the spectrometer entrance slit, but increases the magnification of the sample image at the slit as well. The magnification of the collection optics must be low enough for the image of the illuminated region on the sample to be transmitted by the spectrometer slit, and the solid angle of the light reaching the spectrometer slit must be equal to or less than that of the spectrometer collection optic. Otherwise, some of the light collected from the sample will be blocked from the detector by the spectrometer.

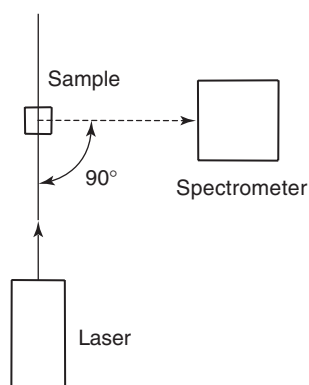
Two other important aspects of the collection optics are standoff and depth-of-field. Standoff is the distance between the sample and the collection optic. The standoff distance is often critical for noncontact analysis. Standoff distances can be increased without reducing collection efficiency by increasing the size of the collection optics. Depth-of-field is a measure of how far the sample can be from the focal plane of the collection optics without a substantial loss of Raman intensity at the detector. Increasing the solid angle of collection reduces the depth-of-field.

Raman collection geometry refers to the angular relationship between the exciting light and the Raman scattered light. The two most common collection geometries,  $90^\circ$  and  $180^\circ$  backscattering, are illustrated in **Figure 6**. The  $180^\circ$  backscattering collection geometry is the most popular largely because the alignment between the excitation and collection paths is built into the collection optics, greatly simplifying alignment of the sample to the Raman instrument.





(A) 180° Backscattering geometry

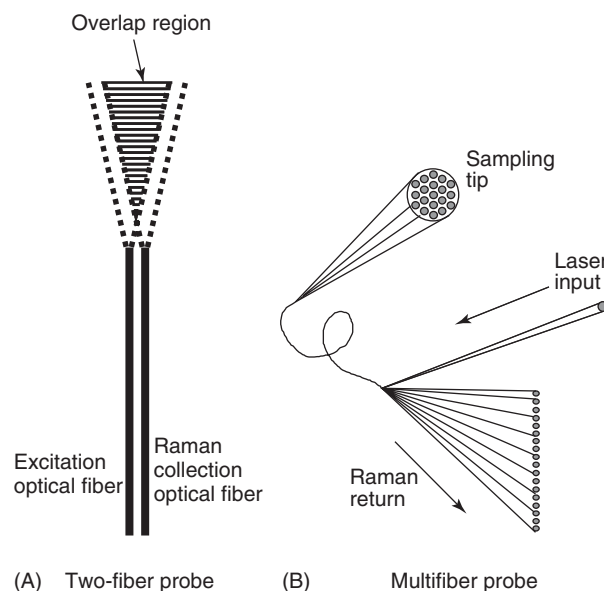


(B) 90° Scattering geometry

**Figure 6** Raman measurement geometries. The solid line represents laser light and the dashed line represents Raman scattered light from the sample. Laser light and Raman light overlap between the beam splitter and the sample in 180° backscattering.

### Sample Compartment

The sample compartment is the traditional and most flexible sample interface for a Raman instrument. The sample is usually mounted on a three-axis translation stage, sometime with additional rotational adjustments. The sample movements can be either manual or computer controlled. A computer-controlled translation stage is often used as an autosampler. Multiple samples are loaded into an array of containers, such as a 96-well plate, and each sample is moved to the observation region and analyzed without human intervention. Sometimes the sample is moved by spinning, translating, or flowing during the acquisition of the Raman spectrum in order to minimize sample damage or to average out sample heterogeneity. The sample compartment also provides a controlled environment for the sample. It excludes room light and sometimes controls temperature and humidity. The sample compartment is usually large enough to contain experiments to be



**Figure 7** Nonfocusing fiber optic Raman probe. The sample is placed in the region where the excitation and collection paths overlap: (A) illustrates the overlap for a nonfocusing probe having a single excitation fiber and a single collection fiber; (B) illustrates a common configuration using a single excitation fiber and multiple collection fibers.

monitored with Raman spectroscopy such as an electrochemical cell, a reo-optical modulator, or a microchemical reactor.

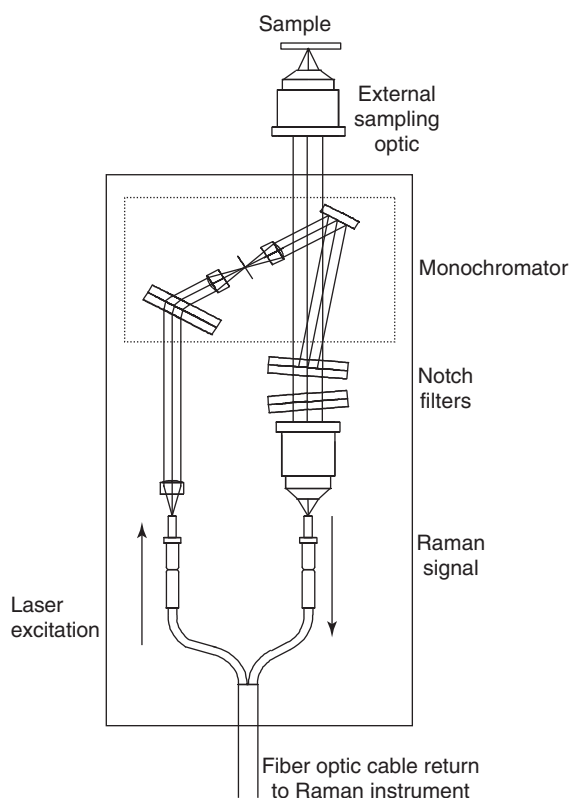
### Fiber Optic Probes

Fiber optic probes have expanded the utility of Raman spectroscopy by taking the Raman measurement capability to the sample rather than requiring the sample to come to the Raman instrument. An optical fiber is used to deliver laser light to the remote sample and a separate optical fiber or bundle of fibers sends Raman photons back to the Raman instrument. Laser light in the optical fiber generates Raman and fluorescence that, if not removed, can contaminate the Raman spectrum of the sample. Most Raman probes use optical filters to remove Raman and fluorescence generated in the optical fiber. These filters are not always necessary when short lengths of optical fiber are used due to the weakness of the fiber emission, or when samples are transparent because the unwanted light is not scattered into the collection fiber. There are two main classes of fiber optic Raman probes for examining single points on a sample: focusing probes and nonfocusing probes.

**Figure 7** illustrates a nonfocusing fiber optic Raman probe. Diverging laser light from the delivery fiber illuminates the sample. Adjacent optical fibers collect light scattered back from the sample and deliver the light to the spectrometer. The collection

fibers are stacked on top of each other at the spectrometer to optimize coupling to the entrance slit. No single collection fiber collects light very efficiently from any specific point on the sample, but the total amount of Raman light collected by all of the collection fibers is often competitive with more traditional focused collection. Optical filters can be deposited on the ends of the individual optical fibers, or spliced into the fibers a short distance from the sampling end to minimize the impact of laser-induced emission from the optical fibers themselves on the Raman spectrum from the sample. When the relative positions of the collection fibers are tracked or preserved, a Raman image of the sample can be obtained. Fiber optic analogs of line and global imaging (see Raman microscope below) can be produced this way.

Figure 8 illustrates a focusing fiber optic Raman probe. Light from the laser delivery fiber is sent through a miniature monochromator inside the probehead before being focused onto the sample. Light from the sample is filtered to remove laser light and then injected into the collection fiber for return



**Figure 8** Schematic drawing of a focusing fiber optic Raman probehead. An internal micro-monochromator eliminates unwanted light from the excitation and a notch filter removes laser light from the Raman signal. A second notch filter acts as a dichroic filter to overlap the excitation and collection paths.

to the Raman instrument. Only a small region on the sample is illuminated and light from that region is collected efficiently. The sampling lens on the probehead that both focuses the laser light on the sample and collects light from the sample can be changed to meet the needs of a particular application. The most commonly changed lens parameters are stand-off from the sample and  $f$ -number. The separation between the probehead and the sampling lens is typically small, but can be many meters since the light between the two is collimated. As a result, the sampling lens can be located in a location incompatible with the probehead without loss in sensitivity. The alignment tolerance between the probehead and the sampling lens is very forgiving.

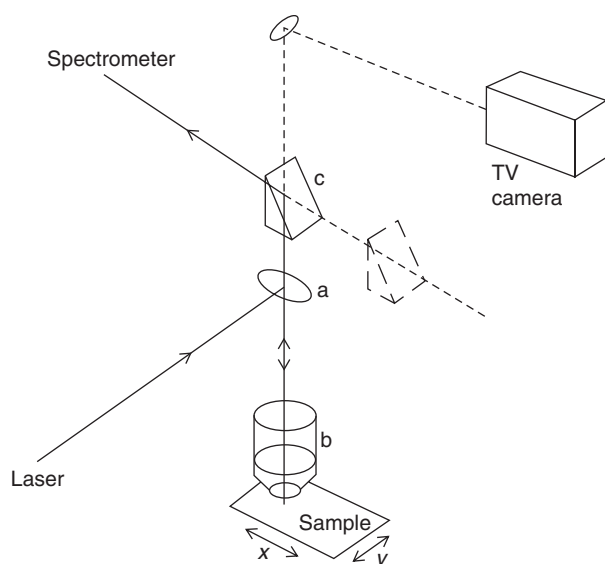
Table 1 compares the nonfocusing and the focusing fiber optic Raman probes. Both styles of probe have unique and important roles to play in fiber optic Raman spectroscopy.

### Raman Microscope

The first two Raman instruments that incorporated microscopes were reported in 1974. These were Raman microprobes, or Raman microspectrometers, because they acquired Raman spectra from microscopic sample volumes. One of these also provided for Raman imaging. The microscope enables these capabilities because its objective can focus a laser beam into a few-cubic-micrometer spot and collect scattered Raman photons through a wide angle. The percentage of Raman light collected is about three times higher for a high-efficiency nonimmersion objective than for a camera lens. In addition, a microscope's magnified optical image of the sample

**Table 1** Comparison of focusing and nonfocusing fiber optic Raman probes

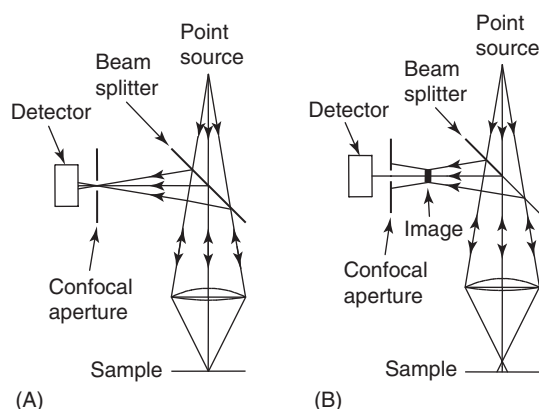
<i>Focusing Raman probe</i>	<i>Nonfocusing Raman probe</i>
Can collect polarized spectra	Cannot collect polarized spectra
Spot size on sample 0.001–0.1 mm	Spot size on sample 0.5–5 mm
Interchangeable optics	Light managed by tip shape
Strong laser rejection	Little, if any laser rejection
No silica Raman artifacts	Silica Raman artifacts possible
Uses two small diameter fibers	Uses several larger diameter fibers
Large probehead	Probehead can be < 1 mm diameter
Requires larger access port	OK for long, narrow openings
\$5000–10 000	Simplest probes <\$1000
Poor utilization of slit height	Good utilization of slit height
Many probes to one analyzer	One or two probes to one analyzer



**Figure 9** A basic Raman microprobe couples laser light onto the sample through a beam splitter (A) and a microscope objective (B). Raman photons are collected with the same objective, so retrace the path to the beam splitter (A), where they continue on to mirror (C), which directs the Raman light into a spectrometer and ultimately to a CCD detector. The mirror (C) can be moved out of the path to allow a magnified view of the sample using a TV camera.

allows the operator to position the laser beam on the sample with respect to features such as heterogeneity, inclusions, domain boundaries, layers, or contamination. **Figure 9** illustrates the basic design of a Raman microprobe.

Confocal scanning optical microscopy, developed in the 1960s, offered significantly improved spatial resolution and contrast over those available using conventional microscopes. Though the application of this technique to Raman microscopy was originally difficult, it has since become a central design element of many current Raman microprobes. Advances in wavelength-filtering technology, the sensitivity of modern CCD detectors, and holistically designed and optimized systems have overcome most of the early problems. In a confocal design, light from a diffraction-limited spot focused on the sample is collected and refocused to a tiny image. A pinhole is placed to allow light passing through this image to reach the detector. Most of the light rays that originate outside of the focused sample spot are blocked, as shown in **Figure 10**. A confocal Raman instrument can have 1- or 2- $\mu\text{m}$  spatial resolution in all three dimensions. This is a powerful technique for discriminating against background scatter or fluorescence and for probing below the surface of a nonopaque sample. Caution must be exercised, however, when interpreting the axial distances involved, due to the



**Figure 10** A confocal aperture combined with the strong divergence of tightly focused light yields fine axial (depth) resolution. Light from a point source (such as a focused laser beam) is focused by a lens (microscope objective) to a small spot on a thin sample in (A). Raman-scattered light from the spot is re-imaged by the lens into the confocal aperture. However, if the sample is positioned below the focal plane of the lens, as in (B), the sample spot becomes larger and no longer forms an image in the confocal aperture. Rather, a larger image forms in front of the aperture. Most of the rays from the image in (B) are blocked from reaching the detector since they no longer line up with the opening.

focal position distortion that occurs when light rays cross boundaries between materials that have very different refractive indices.

Raman imaging captures contrast, similar to that in a photograph, originating from the spatial distribution of Raman spectral features in the sample. Spectral features can express such properties as molecular composition, stress, and degree of crystallinity, among others. Three main approaches have been taken to Raman imaging: point mapping, line imaging, and global (widefield) imaging. In point mapping, a Raman spectrum is collected from each spatial point sequentially, so the number of spectral acquisitions is equal to the number of image pixels. Global imaging collects the whole image at once for a single spectral element, so the number of image acquisitions is equal to the number of Raman spectral elements in the spectrum. Line imaging is a hybrid of the two approaches: spectra from a spatial line on the sample are collected simultaneously to make a line image. Multiple line images can be collected and combined to form a two-dimensional image. So, for a square image, the number of line acquisitions is equal to the square root of the number of image pixels. Each of these approaches has strong and weak points, as summarized in **Table 2**. Point mapping is very slow, but is the most easily implemented with common Raman systems and can be adapted to solve many sample problems. Global

**Table 2** Comparison of Raman imaging strategies

	<i>Point mapping</i>	<i>Line imaging</i>	<i>Global imaging</i>
Spectral resolution ( $\text{cm}^{-1}$ )	1	1	10
Spectral acquisition speed	Seconds	Seconds to minutes	Minutes
Usually collect continuous spectra	Yes	Yes	No
Definition of image	Low	Moderate	High
Image acquisition speed	Hours to days	Seconds to minutes	Seconds
Raman photons used per acquisition	Most	Most	Fewer
Laser power required	Low	Moderate	High
Confocal axial resolution	Yes	Sometimes	No
Robust to different surface geometries	More	More	Less
Moving sample and/or optics required	Yes	Yes	No

imaging has somewhat lower spectral resolution, but can provide the fastest and highest definition images. Line imaging uses CCD area the most efficiently since the line is imaged vertically on the detector pixels, while the spectra are dispersed horizontally.

## Detectors

Photomultiplier tubes and CCD detectors are the dominant detectors for Raman spectroscopy. Less common detectors include avalanche photodiodes for specialized applications in the near-infrared, single element photovoltaic detectors such as germanium or InGaAs detectors for FT-Raman measurements, and single element silicon photovoltaic detectors for stimulated Raman measurements.

### Photomultiplier Tubes

A photomultiplier tube (PMT) consists of a photocathode followed by an electron multiplier. A single photon ejects an electron from the photocathode. Electric fields in the PMT accelerate the electron into another surface called a dinode. The collision of the electron with the dinode releases several new electrons, which are accelerated into another dinode.

This process is repeated several times producing a typical electron gain of  $\sim 10^6$ . The current from the PMT can be measured directly, but better signal-to-noise ratio can be obtained using a nonlinear process called photon counting.

At low enough optical intensities (less than  $10^7$  photons per second) each photon produces a discrete current pulse from the PMT. If the gain of the PMT is  $10^6$  then current pulses from the PMT consisting of significantly less than  $10^6$  electrons are noise and can be ignored. Current pulses of  $\sim 10^6$  electrons are counted. The sum of the counts for a given period of time is proportional to the detected optical intensity. The probability that more than one photoelectron is produced simultaneously is negligible at low count rates, but causes nonlinearity between the optical intensity and count rate at higher optical intensities. Multilevel discriminators can extend the linear range somewhat by classifying pulses as zero, one, or two photons.

PMTs have a relatively large detection area making them well suited to most monochromators. Their dark noise per unit detection area is the lowest of any detector (lower than even cryogenically cooled CCDs). Their quantum efficiencies are in the range of 1–40% in the ultraviolet and visible spectral regions, but fall off rapidly in the near-infrared. PMTs have nanosecond response times that are useful for time-resolved Raman measurements.

### CCD Detectors

CCD detectors are the most popular detectors for Raman spectroscopy. CCD detectors consist of a two-dimensional array of silicon photovoltaic detectors and the associated electronics to read the signal from each detector element. Photoelectrons are stored at each detector element until a read operation removes them for measurement. The detector elements in CCD detectors commonly used for Raman spectroscopy are  $13\text{--}27\mu\text{m}^2$  on edge and are arranged as arrays of  $200 \times 200$  detector elements up to  $2048 \times 1024$  detector elements. CCD detectors with more detector elements are available, but offer diminishing returns for much greater expense. The electrons from multiple detector elements are usually combined to give the signal at any particular Raman wavelength because the spectrograph slit image is almost always much larger than an individual detector element. Raman imaging instruments may use all the detector elements separately, however, to maximize spatial resolution.

Typical quantum efficiencies for CCD detectors range from 30% to  $\sim 100\%$ . Quantum efficiencies fall off with wavelength as the silicon band gap at

1100 nm is approached, but greater than 20% quantum efficiency is possible at 1050 nm with optimized detectors. Quantum efficiency versus wavelength curves for some popular CCD detectors are shown in **Figure 1**. CCD detectors are normally cooled to  $-40^{\circ}\text{C}$  to  $-120^{\circ}\text{C}$  in order to reduce dark noise. Dark noise is reduced by a factor of 2 for each  $8\text{--}10^{\circ}\text{C}$  decrease in temperature. Dark noise can be reduced to less than one photoelectron per hour per detector element (20 photoelectrons per second per square centimeter of detector area), but this level of dark noise reduction is usually not necessary. Shot noise is usually the dominant noise in Raman spectra measured with a CCD detector.

## Integrated Instrumentation

An integrated Raman instrument is a single package containing the laser source, sample interface, spectrometer, detector, and computer all working in harmony together. Different Raman instruments can be compared based on performance factors such as sensitivity, detector noise, spectral resolution, spectral range, spatial resolution, and acquisition speed. They can also be compared on the basis of operational parameters such as size, cost, power consumption, and adaptability, or on the basis of application-specific factors such as laser wavelength, laser power, sample interface, and user interface. Several different types of Raman instruments are in common use because an optimal compromise of these parameters differs widely among Raman applications and vendors.

Integrated Raman systems can be classified as instruments designed for the research laboratory, for routine analysis, for process control, and for portable, field-deployable applications. Research laboratory instruments offer new and state-of-the-art capabilities in exchange for compromised reliability and frequent need for support from a Raman expert. Research laboratory instruments are extremely adaptable to address unanticipated measurement needs. Routine analysis instruments provide limited flexibility with good reliability. They are operationally simple and contain enough Raman expertise built in for technicians to carry out repetitive assays efficiently and reliably. Process control instruments are typically fiber optic Raman systems that have been hardened to perform in the more challenging environmental conditions typical of a chemical production facility. A process control instrument usually runs continuously in a fully automated mode. There

are usually no Raman experts available at the plant to calibrate, reoptimize alignment, or restart the instrument. Any instrument downtime can be very expensive and even dangerous. Process Raman measurements are often made under adverse temperature, vibration, chemical exposure, and pressure conditions. Field-portable Raman instruments are still at a low level of maturity. Size, weight, and utility requirements currently take priority over Raman performance.

*See also:* **Raman Spectroscopy: Near-Infrared.**

## Further Reading

- Chalmers JM and Griffiths PR (2002) *Handbook of Vibrational Spectroscopy*, vol. 1. West Sussex: Wiley.
- Denton MB, Fields RE, and Hanley QS (1996) *Recent Developments in Scientific Optical Imaging*. London: Royal Society of Chemistry Information Services.
- Laserna JJ (1995) *Modern Techniques in Raman Spectroscopy*. London: Wiley.
- Lewis IR and Edwards HGM (2001) *Handbook of Raman Spectroscopy from the Research Laboratory to the Process Line*. New York: Dekker.
- McCreery RL (2000) *Raman Spectroscopy for Chemical Analysis*. New York: Wiley.
- Pelletier MJ (1999) *Analytical Applications of Raman Spectroscopy*. Oxford: Blackwell Science.
- Schrader B (1995) *Infrared and Raman Spectroscopy: Methods and Applications*. Weinheim: VCH.
- Turrell G and Corset J (1996) *Raman Microscopy: Developments and Applications*. New York: Academic Press.

## Glossary

Steradian	The SI unit of solid angle. There are $4\pi$ steradians in a sphere.
Shot noise	Fundamental noise due to the statistical nature of light. One standard deviation of the average intensity equals the square root of the total number of photons measured.
Monochromatic	Light having a spectrally narrow bandwidth.
Q-switched	Technique for generating short, intense laser pulses.
Polarization	Direction of the electric field vector in a beam of light.
Quantum efficiency	Ratio of photoelectrons to total number of incident photons absorbed by a detector.

## Near-Infrared

**C Ross**, Brock University, St Catharines, ON, Canada  
**K T Carron**, University of Wyoming, Laramie, WY, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

When radiation passes through a transparent medium, the species present scatter a fraction of the beam in all directions. C.V. Raman discovered that incident light is inelastically scattered from a sample and shifted in frequency by the energy of its characteristic molecular vibrations, a phenomenon known as Raman scattering. Today, laser photons over a wide range of frequencies from the near-ultraviolet (near-UV) to the near-infrared (NIR) region are used in Raman scattering studies. By choosing wavelengths which excite appropriate electronic transitions, resonance Raman studies of selected components of a sample or parts of a molecule can be performed.

Over the past few years, the range of excitation wavelengths has been extended to the NIR region in which background fluorescence is reduced and photoinduced degradation of the sample is diminished. The NIR region of the spectrum extends from the upper wavelength end of the visible region at  $\sim 770\text{--}2500\text{ nm}$  ( $13\,000\text{--}4000\text{ cm}^{-1}$ ). For many years, NIR excitation of Raman spectra was not considered viable. However, now high-intensity NIR diode lasers are easily available, making this region attractive for compact, low cost Raman instrumentation. Furthermore, the development of low noise, high quantum efficiency, multichannel detectors (charge coupled device (CCD) arrays) combined with high-throughput single-stage spectrographs used in combination with holographic laser rejection filters, have led to high-sensitivity Raman spectrometers.

Classical, spontaneous Raman scattering is a powerful analytical tool that allows for the investigation of the qualitative and quantitative composition of biological, pharmaceutical, and environmental samples. The following discussion of NIR-Raman spectroscopy will begin with a general review of Raman spectroscopy, followed by a description of NIR-Raman, with further discussion about instrumentation and applications of the NIR-Raman technique.

### General Aspects of Raman Spectroscopy

When a beam of monochromatic light is incident on a sample, some of the light is transmitted, some is

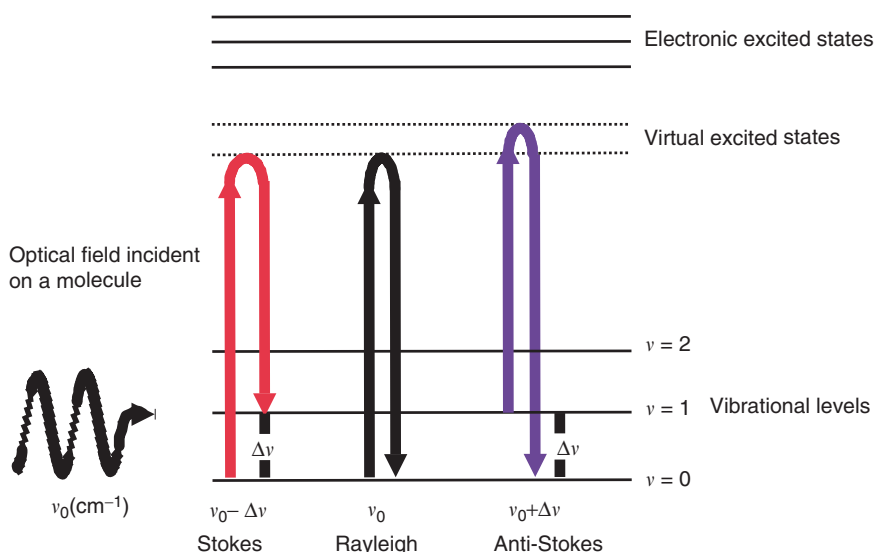
absorbed, and some is scattered. Most of the scattered light has the same wavelength as the incident light, an effect known as Rayleigh scattering. However, a small fraction of the scattered light ( $\sim 1$  out of  $10^8$  photons) is shifted in wavelength, resulting in inelastic scattering or Raman scattering. In the Raman scattering, the differences in energy between the incident and the scattered photons are quantized, corresponding to the energy differences within the vibrational or the rotational levels of the molecules.

In Raman spectroscopy, the wavelength shift is defined as the difference in wavenumber,  $\Delta\nu$ , ( $\text{cm}^{-1}$ ) between the observed radiation and that of the source. Two types of Raman scattering are present, Stokes and anti-Stokes lines. Stokes lines are found at wavenumbers that are lower than the Rayleigh peak while anti-Stokes peaks occur at wavenumbers greater than the wavenumber of the laser excitation source. Anti-Stokes lines are generally considerably less intense than the corresponding Stokes lines. Fluorescence may seriously interfere with the observation of Stokes shifts but not with anti-Stokes, thus with fluorescing samples, anti-Stokes signals may sometimes be more useful despite their lower intensities. With highly fluorescing samples, it is more common to use a different excitation source, circumventing the production of fluorescence. This strategy frequently involves moving from the use of a visible laser source as an excitation source to one lasing in the NIR region.

**Figure 1** depicts an energy-level diagram showing the sources of Raman and Rayleigh scattering. When a photon having energy  $\nu_0$  collides with a molecule, two different events can take place. If the scattered photon has the same energy as the exciting photon, the Rayleigh scattering can be observed. The middle arrow depicts the changes that produce Rayleigh scattering. No energy is lost in Rayleigh scattering, thus the collisions between the photons and the molecule are termed elastic.

The energy changes that produce Stokes and anti-Stokes emission are depicted on the left and right, respectively. The two differ from the Rayleigh radiation by frequencies corresponding to  $\pm\Delta\nu$ , the energy difference between the first vibrational level and the ground state. In Stokes scattering, the molecule loses energy  $-\Delta\nu$ , reaching the vibrational level  $\nu=1$  and not returning to its vibrational state  $\nu=0$ . In this case, the scattered photon has energy  $\nu_0 - \nu_1$ , where  $\nu_1$  is the frequency of the vibration. In anti-Stokes scattering, the molecule originally in the





**Figure 1** Energy level schematic of Raman spectroscopy.

$v = 1$  state, is first promoted to a virtual excited state, then successively returns to the ground vibrational state,  $v = 0$ . The energy of the scattered photon is  $\nu_0 + \nu_1$  and the spectral line is shifted to a lower wavelength and a higher frequency than the incident radiation. However, since the majority of the molecules are in the ground vibrational state at room temperature, the Stokes lines is stronger compared to the anti-Stokes lines.

## NIR-Raman Spectroscopy

### The Difference in Excitation

Many books on Raman spectroscopy were written before NIR lasers became a viable excitation source. Traditionally, the relative magnitude of Raman spectroscopy is explained using the  $\nu^4$  factor, which indicates that the Raman intensity for normal Raman scattering varies with the fourth power of the observed frequency, an event dependent on the excitation laser frequency. However, since Raman is too weak to be seen by the naked eye without the aid of filters, it is difficult to find good analogies for teaching. The  $\nu^4$  process provides the explanation for several common phenomena, the blue sky and red sunset. Rayleigh proved that elastic scattering of light (Rayleigh scattering) is responsible for the blue sky. When the sky is observed during the day, the vision is overwhelmed by light scattered by atmospheric molecules. Since high frequencies scatter more and blue light is the highest frequency of the visible spectrum, the sky appears blue. At sunset just the opposite occurs and all the high frequencies are scattered away to the west. Thus, only the low frequencies of red and

orange are visible. Raman may be considered in this context as the inelastic equivalent of Rayleigh scattering and follows the same rule.

This theory predicts that Raman spectroscopy should be optimally performed in the near-UV not the NIR region. For example, excitation using the blue line of an  $\text{Ar}^+$  laser at 454 nm will produce essentially nine times more intense Raman scattering than 785 nm excitation. Yet in recent years, there has been an overwhelming trend toward using NIR excitation. This trend stems from a single interference source and frustration in Raman spectroscopy: fluorescence. Fluorescence occurs when the excitation is able to promote an electron into an excited electronic state. When impure samples or pure samples with low energy transitions are placed in a Raman spectrometer, the spectrum recorded using visible excitation often shows a broad fluorescence background, frequently devoid of any Raman features. A good rule-of-thumb is that the number of electronic states decreases with energy. Thus NIR excitation is often preferred since it will produce high quality Raman spectra devoid of any fluorescence interference unattainable with a higher energy visible excitation source.

A simple property of any spectrum can provide a more quantitative picture to the problem of fluorescence in Raman spectroscopy. The signal-to-noise ratio (S/N) describes the quality of the spectrum. A high S/N is desirable for viewing a spectrum for both qualitative analysis and for concentration determination with quantitative analysis. Obviously, the S/N can be increased two ways: an increase in signal or a decrease in noise. For example, if the Raman spectrum of cyclohexane is taken from a sample and has been filtered to remove any particulates, the S/N ratio

will improve linearly with increasing laser power (for relatively low laser powers) or as the square root of the integration time. If one takes a spectrum of cyclohexane with a fluorescent impurity, the S/N will not increase linearly with laser power. Longer integration times will be limited by the dynamic range of the detector. This originates from the dependency of the noise on the signal in spectroscopy. When visible photons are detected, they produce a noise equivalent to the variance of the signal. If a sample has a background fluorescence that increases with the laser power or integration time, noise will also increase. This leads to the fundamental question of whether Raman spectroscopy is a true zero-baseline technique.

The answer is that Raman spectroscopy rarely fits that description of zero-baseline. Most samples possess a background due to either inherent or impurity fluorescence. If NIR excitation circumvents or strongly reduces the background, the reduction in intensity due to the  $\nu^4$  dependence of the Raman signal becomes irrelevant compared to the noise produced by a large fluorescence background if visible excitation is used.

A second problem arises from the shape of the fluorescent background. If the background were linear, it could be easily subtracted to produce a high-quality spectrum, albeit noisy due to the noise produced by the high background. However, in addition to adding noise to the spectrum, the fluorescence creates broad features in the background that are often very difficult to remove. This leads to serious difficulties in quantitative analysis because the Raman signal above the baseline is proportional to the concentration, not the signal above a zero-baseline. It also can lead to small frequency shifts in qualitative Raman spectroscopy. NIR-Raman spectroscopy often leads to higher quality spectra due to the minimization of background signals from the sample.

### The Difference in Detection

The S/N is a function of the excitation in Raman spectroscopy and as already discussed above, Raman spectroscopy can strongly benefit from using NIR excitation. The detection method also affects the S/N and is still perhaps a point of controversy among Raman spectroscopists. The controversy stems from the quasiphoton nature of NIR light. One may prefer to consider the light particle in nature and count the photons or one may prefer to consider the light as a wave. To explain this duality and controversy, a freshman chemistry experiment can be used as an illustration.

Consider the simple process of weighing coins on a balance. Assume the weight of three coins needs to be determined using a balance. The common laboratory balance has a standard error of  $e$  while a very good analytical balance has an  $e$  of 0.0001 g. Thus, if weighing  $0.001 \pm 0.0001$  g or even  $10 \pm 0.0001$  g, the error is not dependent on the mass being weighed. The three coins individually will weigh  $m_1 \pm 0.0001$  g,  $m_2 \pm 0.0001$  g, and  $m_3 \pm 0.0001$  g. The accuracy can be improved by weighing the coins in pairs. In this case,  $m_1 + m_2 = m_{12} \pm 0.0001$  g,  $m_1 + m_3 = m_{13} \pm 0.0001$  g, and  $m_2 + m_3 = m_{23} \pm 0.0001$  g.

The three weights of each pair can be subtracted to find the individual weights. For example,  $m_1 = \frac{1}{2}(m_{12} + m_{13} - m_{23}) \pm e_T$ . The total error,  $e_T$ , can be found from the propagation errors in the three measurements. It is  $\{[\frac{1}{2}(e_1 + e_2 - e_3)]^2\}^{1/2}$ . If the errors are random, the equation can be simplified to  $m_1 = [\frac{1}{4}(e_1^2 + e_2^2 + e_3^2)]^{1/2}$  and as stated above, the balance has an error of  $e$  regardless of the mass. This leads to  $m_1 = [\frac{1}{4}(3e^2)]^{1/2}$  or  $3^{1/2}/2e$  or 87% of  $e$ . In the actual measurement, the weight of  $m_1$  would be  $m_1 \pm 0.000087$ . In general, the more coins that are weighed, the better the result. If  $N$  coins were weighed, the error would be reduced by  $N^{1/2}/(N - 1)$ , or for a 1000 coins, the error would be reduced to 3% of its original value. This may be perceived of as the basis of Fourier Transform Raman spectroscopy (FT-Raman spectroscopy). In FT-Raman, the Raman scattered light can be considered as waves that are allowed to interfere with each other to produce an interference pattern (interferogram) that can be mathematically converted by an inverse FT into a spectrum. The benefit, called Fellgett's or Multiplex Advantage, arises from putting the whole spectrum at once onto the detector. This is exactly analogous to placing the many coins on the balance at once to find the weight of an individual coin.

The controversy arises from the fact that optical detectors are not always like an analytical balance. As the energy of a photon decreases, the prominent detector noise becomes the thermal excitation of electrons in the detector. When this noise source dominates, the detector is similar to the analytical balance. When the energy of the photon is large, the thermal noise becomes insignificant. This transition occurs in the NIR region. The noise from high-energy photons arises from their quantum nature. If individual photons were counted for 1-s intervals, in one interval, 100 may be counted, in the next 95, in the next 105, and so on. These differences occur because the photons are random events and may happen more often during one period than the next. Statistics dictate that the error or standard deviation of photons counted is the square root of the number

of photons being counted or as stated earlier, the signal is equal to the variance in the signal.

Using the weight of the coins, the error of the balance can be calculated. The same calculations can be made as described above, except the error over two weighings is  $2^{1/2}e$ . Calculations show that for  $m_1$ , the error is  $\pm(3/2)^{1/2}e$  or  $1.22e$ . The error has increased instead of decreasing. What was once an advantage for one detector is now a disadvantage for another. The result is two camps of Raman spectroscopists. When a sample is very fluorescent, even in the far-red, it is best to move further into the IR and use FT-Raman spectroscopy. If fluorescence can be eliminated or reduced, conventional NIR-excited dispersive Raman spectroscopy provides the best option.

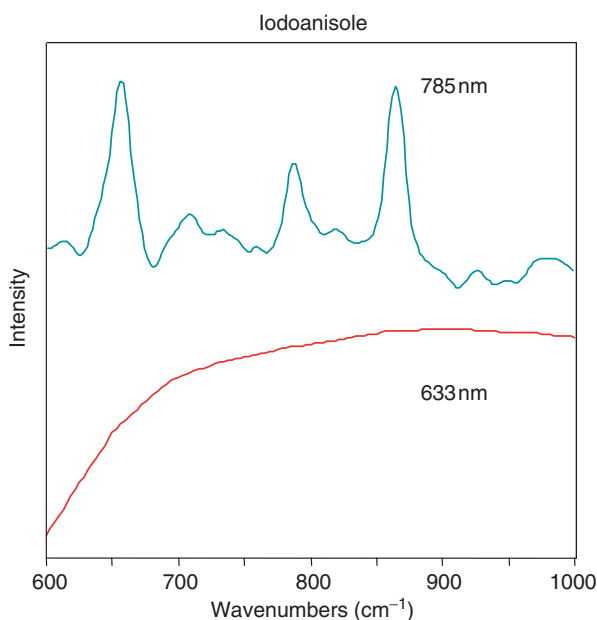
## Instrumentation

For many years, NIR excitation of Raman spectra was not considered viable; however, things changed with several key instrumentation developments. CCDs now allow parallel detection of Raman spectra with high quantum efficiency and low noise. Volume-holographic filters and gratings improve overall throughput and provide elastic light rejection for Raman spectrographs. In addition, economical, long-life laser diodes provide sufficient power for Raman applications and optical fiber probes allow remote chemical analysis by Raman scattering.

### Laser Source

The sources used in modern Raman spectroscopy are almost always lasers because their high intensity is necessary to produce Raman scattering of sufficient intensity to be measured with reasonable signal-to-noise ratio. Argon ion lasers operate at 488 or 514.5 nm, krypton lasers operate at 530.0 or 647.1 nm, and helium/neon (He/Ne) lasers operate at 632.8 nm. In NIR-Raman spectroscopy, the laser sources used are the diode laser (782 or 830 nm) and Nd/YAG (1064 nm). NIR sources have two major advantages over shorter wavelength lasers. The first is that they can be operated at much higher power (typically up to 50 W) without causing photodecomposition of the sample. The second is that they are not energetic enough to populate a significant number of fluorescence producing excited electronic energy states in most molecules. Consequently, fluorescence is generally much less intense or nonexistent with these lasers. The Nd/YAG line at 1064 nm is particularly effective in eliminating fluorescence. The two lines of the diode array laser at 782 and 830 nm also markedly reduce fluorescence in most cases.

Figure 2 provides an example where the diode laser source completely eliminates background



**Figure 2** Effect of different laser wavelengths, 633 nm (He/Ne laser) and 785 nm (diode laser), on the reduction of fluorescence in the analysis of iodoanisole.

fluorescence. The sample was iodoanisole and most of recorded signal using visible excitation arises from the fluorescence of that compound. The lower curve was obtained with conventional Raman equipment using the 633-nm line from a He/Ne laser for excitation. The upper curve is for the same sample recorded with a spectrometer equipped with a diode laser that emitted at 785 nm. Note the absence of fluorescence background signal.

Two basic cavity configurations are used to construct and stabilize NIR-Raman laser sources. The first design, the Littrow configuration, collimates light from the laser diode and sends it to the grating at a specific angle of incidence. The first-order diffracted light is sent directly back to the laser to provide optical feedback across a narrow wavelength range. The output beam is the light that is reflected from the grating, thus the output is at an angle that is approximately twice the angle of incidence from the optical axis. Because of its simple design, the external-cavity length can be quite short, resulting in wide spacing of the external-cavity mode. This spacing facilitates single-frequency operation where the goal is to have only one external-cavity mode fall within the bandwidth of the diffraction grating. The bandwidth of the grating is relatively large because the Littrow configuration uses a single-pass geometry.

In the second design, the Littman configuration, the collimated laser light strikes the grating near the grazing incidence so that the diffracted order does not return to the laser directly. Instead, the diffracted

light is reflected by a mirror, diffracted by the grating a second time and then returned as optical feedback to the laser diode. The advantage of the double-pass geometry is that the grating bandwidth is less than half of what it is in the Littrow case but the external cavity length tends to be longer. In practice, the Littman configuration is usually selected when narrower-frequency operation is desired.

### Spectrometer

A Raman spectrophotometer analyzes the radiation scattered by molecules when they are illuminated with monochromatic exciting radiation. Currently, most Raman spectrometers are either FT instruments equipped with cooled germanium transducers or multichannel instruments based upon CCDs. These transducers, in contrast to photomultiplier tubes, are sensitive to radiation at 785 nm produced by diode lasers, which provide Raman excitation of many compounds without significant fluorescence. CCDs are not sensitive to the 1064-nm radiation from an Nd/YAG laser.

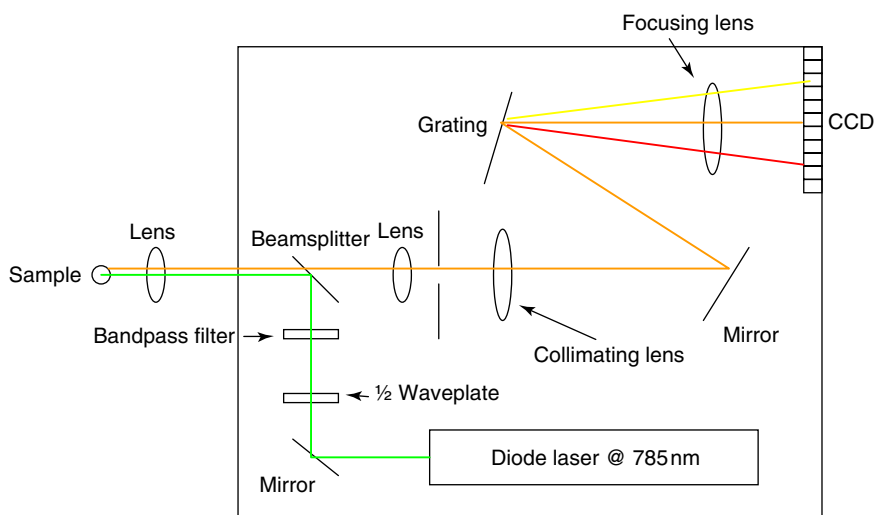
Figure 3 is a schematic representation of a typical NIR-Raman dispersing spectrometer with a CCD in the Littman configuration. The source is a diode laser and filter system that yields radiation at 785 nm. This beam is focused on the end of the excitation filter by means of a lens, and the Raman emission is transmitted through the collimating lens. A diffraction grating disperses the radiation and reflects it onto a CCD that is made up of a large number of pixels. These pixels store the charge produced by the light quanta. Since the spectral information seen by the individual pixels is accumulated simultaneously, an

array with  $n$  pixels is equivalent to  $n$  spectrometers accumulating separately or one spectrometer working  $n$  times.

### Applications

NIR-Raman spectroscopy has been used for a number of applications and is particularly useful for biological and biomedical uses. Fluorescence has been a limiting factor for much Raman analysis of biological samples, particularly whole-cell or whole-tissue samples. NIR excitation reduces interference from fluorescence and decreases photoinduced degradation of the sample, enabling researchers to obtain spectra for a variety of biomaterials and living cells.

Interest in both *in vivo* and *in vitro* use of NIR-Raman spectroscopy of tissues for diagnostics continues to grow. NIR-Raman has been evaluated as a diagnostic tool for cervical precancers. Although spectra could not be collected *in vivo* due to long integration times, NIR-Raman was found to have specificity that was superior to standard methodologies, such as colonoscopy and cytology, for differentiating squamous intraepithelial lesions (SIL) from non-SIL. Several research groups have described NIR-Raman instrumentation with CCD detection for biological tissue analysis, both *in vivo* and *in vitro*. In a specific study, an 810-nm excitation wavelength was used for *in vitro* laboratory analysis of aorta tissues. Subsequent work showed that fluorescence interference in tissue Raman spectra could be reduced even further using 830-nm excitation without compromising CCD sensitivity. The usefulness of NIR-Raman coupled with CCD detection for



**Figure 3** Typical Raman dispersing spectrometer with a CCD in the Littman configuration. (Courtesy of DeltaNu, Laramie, WY.)

studying cancerous changes in the colon, urinary bladder, breast, and soft tissue sarcomas has also been examined.

NIR-Raman has found applications in the pharmaceutical industry. McCreery and colleagues reported the use of the technique for identification of pharmaceuticals inside amber vials. Even with the signal attenuation through the glass, adequate spectra were obtained for determination of vial content with 1–60 s integration times. Using a library of spectra, identification of the pharmaceuticals in the vials was performed and identification was found to display accuracy between 88% and 96%. This work demonstrated the potential of NIR-Raman for online process monitoring.

The environmental community has also found uses for NIR-Raman analysis. A continuous method was developed by Weissenbacher and colleagues to detect trace organic pollutants using flow injection analysis and surface enhanced Raman spectroscopy (SERS). This method uses NIR excitation and FT-SERS detection to detect parts per million of pesticides in aqueous solutions. In this study, the authors describe the simultaneous detection of two pesticides, carbendazim and metazachlorine.

A variety of industrial processes benefit from the use of NIR-Raman spectroscopy. In a study examining unleaded petroleum gasoline, with excitation at 514 nm from an argon laser, no Raman features were observed above the strong fluorescence. With He/Ne laser excitation at 633 nm, several Raman features appeared above the background. With excitation at 785 or 852 nm, a good quality Raman spectrum was obtained. In the textile industry, NIR-Raman was used to examine ready-made textiles for the discrimination of the different raw materials.

NIR-Raman has been introduced for use into the food processing industry. In the production of oils and fats, the determination of the amount of unsaturation, such as *cis* and *trans* isomers, can be important for food processing and food labeling. NIR-Raman was reported to measure the total unsaturation and the *cis* and *trans* isomers online during the production process. Fluorescence interference with visible excitation of many fats was

drastically reduced at NIR wavelengths. In addition to food processing, NIR-Raman has been used for determination of food quality as the main components of food (carbohydrates, proteins, and lipids) all show characteristic Raman lines using NIR detection.

The pairing of SERS with NIR detection has proved to be a powerful DNA detection technique. NIR-SERS was reported to provide excellent discrimination against fluorescent interference and was nonresonant with most molecules. This allowed greater excitation intensities without photobleaching or destruction of the analyte. In one study, Kneipp and colleagues used this technique in rapid DNA sequencing to detect single-molecule DNA bases or nucleotides. NIR-SERS provided a method for detection and identification of the single DNA base, adenine, without any additional labeling. This study reported trace detection levels of adenine and AMP with well-resolved spectra.

*See also: Raman Spectroscopy: Instrumentation; Surface-Enhanced.*

## Further Reading

- Hanlon EB, Manoharan R, Koo TW, *et al.* (2000) Prospects for *in vivo* Raman spectroscopy. *Physical and Medical Biology* 45: R1–R59.
- Isola N, Stokes DL, and Vo-Dinh T (1998) *Analytical Chemistry* 70: 1352–1356.
- Keller S, Lochte T, Dippel B, and Schrader B (1993) Quality control of food with near-infrared excited Raman spectroscopy. *Fresenius Journal of Analytical Chemistry* 346: 863–867.
- Kneipp K, Kneipp H, Itzkan I, Dasari R, and Feld M (1999) Ultrasensitive chemical analysis by Raman spectroscopy. *Chemical Reviews* 99: 2957–2975.
- McCreery RL, Horn AJ, Spencer J, and Jefferson E (1998) *Journal of Pharmaceutical Science* 87: 1–8.
- Mahadevan-Jansen A, Mitchell MF, Ramanujam N, *et al.* (1998) *Photochemistry and Photobiology* 68: 123–132.
- Mulvaney SP and Keating CD (2000) Raman spectroscopy. *Analytical Chemistry* 72: 145R–157R.
- Weissenbacher N, Lendl B, Frank J, *et al.* (1997) *Journal of Molecular Structure* 410–411: 539–542.

## Surface-Enhanced

**R E Littleford, D Graham, and W E Smith**, University of Strathclyde, Glasgow, UK

**I Khan**, Imperial College, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Surface enhanced Raman scattering (SERS) is a sensitive spectroscopic technique for the detection and

studying cancerous changes in the colon, urinary bladder, breast, and soft tissue sarcomas has also been examined.

NIR-Raman has found applications in the pharmaceutical industry. McCreery and colleagues reported the use of the technique for identification of pharmaceuticals inside amber vials. Even with the signal attenuation through the glass, adequate spectra were obtained for determination of vial content with 1–60 s integration times. Using a library of spectra, identification of the pharmaceuticals in the vials was performed and identification was found to display accuracy between 88% and 96%. This work demonstrated the potential of NIR-Raman for online process monitoring.

The environmental community has also found uses for NIR-Raman analysis. A continuous method was developed by Weissenbacher and colleagues to detect trace organic pollutants using flow injection analysis and surface enhanced Raman spectroscopy (SERS). This method uses NIR excitation and FT-SERS detection to detect parts per million of pesticides in aqueous solutions. In this study, the authors describe the simultaneous detection of two pesticides, carbendazim and metazachlorine.

A variety of industrial processes benefit from the use of NIR-Raman spectroscopy. In a study examining unleaded petroleum gasoline, with excitation at 514 nm from an argon laser, no Raman features were observed above the strong fluorescence. With He/Ne laser excitation at 633 nm, several Raman features appeared above the background. With excitation at 785 or 852 nm, a good quality Raman spectrum was obtained. In the textile industry, NIR-Raman was used to examine ready-made textiles for the discrimination of the different raw materials.

NIR-Raman has been introduced for use into the food processing industry. In the production of oils and fats, the determination of the amount of unsaturation, such as *cis* and *trans* isomers, can be important for food processing and food labeling. NIR-Raman was reported to measure the total unsaturation and the *cis* and *trans* isomers online during the production process. Fluorescence interference with visible excitation of many fats was

drastically reduced at NIR wavelengths. In addition to food processing, NIR-Raman has been used for determination of food quality as the main components of food (carbohydrates, proteins, and lipids) all show characteristic Raman lines using NIR detection.

The pairing of SERS with NIR detection has proved to be a powerful DNA detection technique. NIR-SERS was reported to provide excellent discrimination against fluorescent interference and was nonresonant with most molecules. This allowed greater excitation intensities without photobleaching or destruction of the analyte. In one study, Kneipp and colleagues used this technique in rapid DNA sequencing to detect single-molecule DNA bases or nucleotides. NIR-SERS provided a method for detection and identification of the single DNA base, adenine, without any additional labeling. This study reported trace detection levels of adenine and AMP with well-resolved spectra.

*See also:* **Raman Spectroscopy:** Instrumentation; Surface-Enhanced.

## Further Reading

- Hanlon EB, Manoharan R, Koo TW, *et al.* (2000) Prospects for *in vivo* Raman spectroscopy. *Physical and Medical Biology* 45: R1–R59.
- Isola N, Stokes DL, and Vo-Dinh T (1998) *Analytical Chemistry* 70: 1352–1356.
- Keller S, Lochte T, Dippel B, and Schrader B (1993) Quality control of food with near-infrared excited Raman spectroscopy. *Fresenius Journal of Analytical Chemistry* 346: 863–867.
- Kneipp K, Kneipp H, Itzkan I, Dasari R, and Feld M (1999) Ultrasensitive chemical analysis by Raman spectroscopy. *Chemical Reviews* 99: 2957–2975.
- McCreery RL, Horn AJ, Spencer J, and Jefferson E (1998) *Journal of Pharmaceutical Science* 87: 1–8.
- Mahadevan-Jansen A, Mitchell MF, Ramanujam N, *et al.* (1998) *Photochemistry and Photobiology* 68: 123–132.
- Mulvaney SP and Keating CD (2000) Raman spectroscopy. *Analytical Chemistry* 72: 145R–157R.
- Weissenbacher N, Lendl B, Frank J, *et al.* (1997) *Journal of Molecular Structure* 410–411: 539–542.

## Surface-Enhanced

**R E Littleford, D Graham, and W E Smith**, University of Strathclyde, Glasgow, UK

**I Khan**, Imperial College, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Surface enhanced Raman scattering (SERS) is a sensitive spectroscopic technique for the detection and



characterization of analytes adsorbed on suitable metal surfaces. The effect was first discovered experimentally in 1974 by Fleischmann *et al.*, who reported very intense Raman bands from pyridine adsorbed onto an anodized silver surface. They attributed this strong signal to the presence of a large number of pyridine molecules present at the roughened electrode surface due to its large surface area. However, in 1977 Jeanmaire and Van Duyne, and Albrecht and Creighton, independently recognized that the increase in intensity could not be accounted for simply by the number of scatterers present. They observed that compared with the equivalent concentration of pyridine in solution, an enhancement of  $\sim 10^6$  in Raman scattering was obtained. They concluded that an intrinsic surface enhancement effect played a fundamental role in producing the enhanced Raman scattering.

In the basic process of SERS, the analyte is adsorbed onto a roughened metal surface of a suitable metal, usually silver or gold. On excitation of this surface with a laser beam, a change in polarizability of the analyte occurs in a direction perpendicular to the surface, leading to the enhanced scattering. The rough surface required for scattering can be provided in a large number of ways. Common methods are to use aggregated colloidal particles, roughened electrodes, or thin cold-deposited metal films.

## Mechanisms of Surface Enhancement

Since the discovery of SERS there has been much debate regarding the origins of the effect. However, it is generally accepted that there are two main contributions to the enhancement process, namely electromagnetic enhancement and charge transfer or chemical enhancement.

### Electromagnetic Enhancement

The collective excitation of an electron cloud on the surface of a metal is termed a surface plasmon. Surface roughness or curvature is required for the scattering of light by surface plasmons. The electromagnetic field at the surface is greatly enhanced upon surface plasmon excitation. If an analyte is present on the roughened metal surface, the molecule experiences a large electric field at the surface. The intensity of the Raman scattering is dependent upon the induced polarization caused by the electric field, and consequently Raman scattering is amplified by a factor of  $10^4$  or greater. Electromagnetic enhancement does not require a direct metal-analyte bond but becomes weaker the larger the separation between the analyte and the surface. It can occur

with a separation of up to  $\sim 20 \text{ \AA}$ . Surface selection rules have been developed based on the electromagnetic approach.

### Charge Transfer

Charge transfer assumes a bond is formed between the analyte and the metal surface. The energy levels of the molecule are shifted and broadened so that they overlap with the Fermi level of the metal and new electronic states are created. On absorption of the incident light by the metal these states serve as resonant intermediate states, allowing efficient charge transfer between the metal and the analyte and hence enhanced Raman scattering. The enhancement obtained in this manner is confined to being a first layer effect. In general, the enhancement predicted is  $\sim 10^2$ .

## Advantages and Disadvantages of SERS

SERS offers several advantages over normal Raman or resonance Raman scattering. The huge increase in signal intensity allows an extended concentration range to be studied, with detection limits considerably lower than those offered by resonance Raman scattering. Fluorescence from the analyte is quenched due to its proximity to the metal surface, providing an alternative, nonradiative route for energy loss. There is however a much broader low-intensity fluorescence from the SERS process. In addition, for studies of processes such as surface-ligand adhesion and corrosion, it allows the study of an adsorbate at less than monolayer coverage on a suitable metal surface *in situ*, in contact with water and other solvents.

SERS does, however, have some limitations. For the SERS effect to occur, the analyte needs to be adsorbed onto or in close proximity to the roughened metal surface, and only a few metals have so far been shown to be efficient at providing surface enhancement. SERS intensities are also dependent on the roughness of the metal surface, and there are significant problems associated with the preparation of reproducible substrates with uniform roughness features. The spectra obtained are also dependent on the orientation of the molecule on the metal surface, and vibrations with little to no intensity in normal Raman scattering can become relatively intense. This can in some cases make it difficult to identify the analyte positively. Further, contamination can also be a problem. Since SERS is very sensitive, it is possible that very small amounts of a contaminant can be enhanced, and if the analyte to be considered does

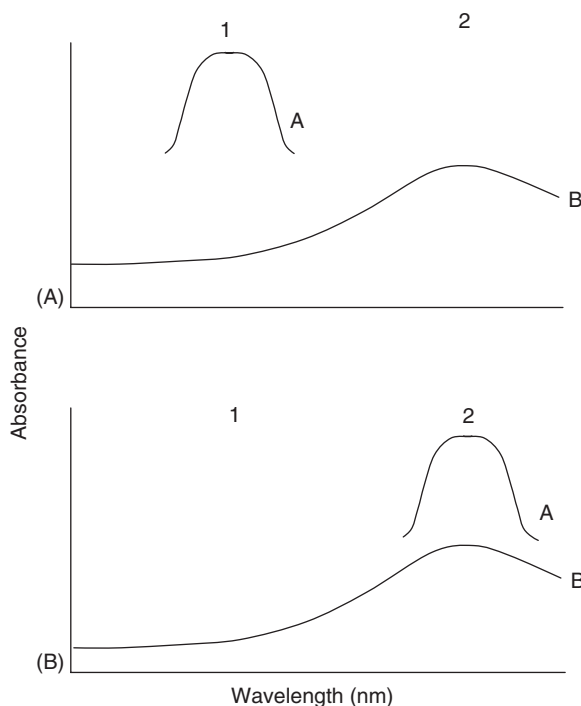
not adsorb efficiently or does not give good SERS, the contaminant can easily dominate the SERS. These problems limit the use of SERS as a quantitative analytical method. However, these problems are largely overcome if surface-enhanced resonance-Raman scattering (SERRS) is used.

## Surface Enhanced Resonance Raman Scattering

So far in this article, both colored molecules and colorless molecules have been treated simply as analytes, and the enhancement has been taken as a combination of electromagnetic and chemical enhancement. However, there are crucial differences where a colored adsorbate is adsorbed onto a metallic surface such that the plasmon resonance frequency and the frequency of an absorption band are close to the frequency of the incident light. This combination of surface enhancement and molecular resonance is called SERRS. Stacy and Van Duyne first reported SERRS in 1983. The increase in sensitivity over SERS can be quite large, with total enhancements of up to  $10^{14}$  reported.

When used practically, there are various combinations of laser frequency, plasmon resonance frequency, and molar absorptivity of the analyte that can be employed to obtain a condition that can be deemed to be SERRS rather than SERS. These are shown diagrammatically in Figure 1. If the excitation is matched to the molecular absorption maximum (Figure 1A), the surface selection rules expected for SERS are much less effective. This is due to the polarization of the laser excitation altering on interaction with the chromophore during the scattering process. This results in a degradation of the effective input polarization required for SERS selection rules, making SERRS less sensitive to the orientation of the molecule to the surface. This usually means that the spectrum resembles that of the resonance spectrum in solution, so that positive identification of the analyte is possible. In addition, the extra enhancement of the analyte discriminates against the detection of contamination. Thus, SERRS is more readily applicable for the development of methods for quantitative analysis.

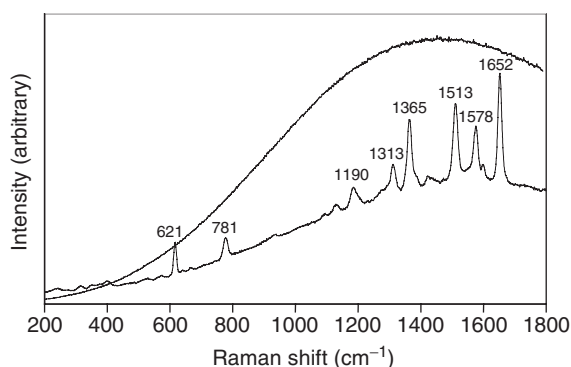
A second possible arrangement illustrated in Figure 1A is where the laser excitation is set away from the absorption maximum of the adsorbate and at the maximum of the plasmon resonance. This is described as preresonant scattering, and often SERRS taken in this way is written as SE(R)RS. The advantages of resonance still apply over quite a wide range of frequencies, gradually tailing off to SERS. Thus, SE(R)RS makes it simpler to pick out individual



**Figure 1** The different arrangements for SERRS: the curves represent (A) molecular absorbance and (B) plasmon resonance. In (A) the molecular absorbance maximum and the plasmon absorbance maximum do not coincide: position 1 represents excitation at the absorbance maximum, 2 that at the plasmon maximum. In (B) the molecular absorbance and the plasmon maximum coincide: position 1 represents excitation away from the absorbance and plasmon maximum where the spectrum has a preresonant component, 2 that at the absorbance maximum and plasmon maximum. (Rodger C, Smith WE, Dent G, and Edmondson M (1996) Surface-enhanced resonance-Raman scattering: An informative probe of surfaces. *Journal of Chemical Society, Dalton Transaction* 791–799; reproduced by permission of The Royal Society of Chemistry.)

resonant molecules in the presence of a matrix of interferences, but the further away the excitation is from molecular resonance, the more the effect will be dependent on the angle of the adsorbate to the surface. For surface studies this is a key point, and consequently this arrangement may be preferred for some surface analyses.

Figure 1B illustrates an alternative case in which the molecular chromophore coincides with the surface plasmon maximum. Similar considerations will apply to those discussed for the case represented in Figure 1A, but a further increase in sensitivity is likely with excitation at the molecular and plasmon resonance frequency. The orientation dependence will become more apparent the greater the difference in frequency between the excitation frequency used and the absorbance and plasmon maximum as the conditions become more appropriate for SERS (1 in



**Figure 2** Resonance Raman and SERRS spectra of rhodamine 6G with excitation at 514.5 nm. The high fluorescence background is effectively quenched in the SERRS spectrum due to adsorption of the rhodamine onto the surface of the silver nanoparticles. (Reproduced with permission from Rodger C and Smith WE (2002) SERS. In: *Handbook of Vibrational Spectroscopy*. New York: Wiley; © John Wiley & Sons Ltd.)

Figure 1B). As a consequence the sensitivity will decrease.

Where SERRS can be applied, it offers many advantages over Raman scattering, resonance Raman scattering, and SERS. It provides both vibrational and electronic information on the adsorbate. Very wide concentration ranges can be studied, and single molecule detection is easier. A further major advantage is the quenching of fluorescence by the surface, which allows fluorescing dyes, which are ineffective in visible resonance Raman studies, to give very good SERRS. For example, rhodamine 6G fluoresces strongly in the visible region, preventing the use of normal Raman scattering, but gives good SERRS (Figure 2). In addition, there is less dependence of the signal on orientation. The combination of sensitivity, selectivity, and robustness of signal makes SERRS a very sensitive and selective form of vibrational spectroscopy that can be used semiquantitatively or qualitatively.

## SERS Substrates

SERS activity requires that very specific substrates be prepared. The metal chosen requires a plasmon in the frequency region close to that of the excitation laser. In addition, the ratio of scattering to adsorption of the substrate is important. Since visible frequency lasers are widely used in this technique, silver, gold, and copper, which have plasmons resonant with visible light, are obvious effective substrates. Silver is particularly useful, having a better ratio of scattering to adsorption in the visible region than gold. Copper tends to be too chemically reactive, and although other metals including lithium, sodium, and some

transition metals are effective, silver is by far the most widely used. It is possible to form reasonably time-stable roughened silver surfaces that can be used under normal laboratory conditions.

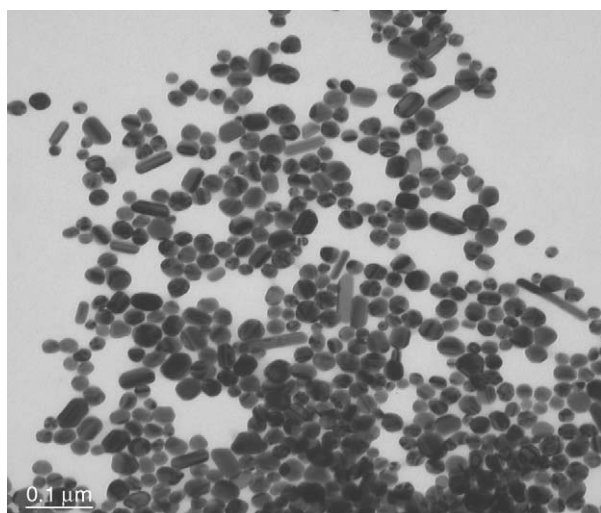
Many types of SERS-active substrate have been made, each containing many microscopic metal domains. These include the following: electrode surfaces roughened by oxidation–reduction cycles; island films consisting of small metal particles; cold-deposited films prepared by evaporation or sputtering in a vacuum; lithographically produced assemblies; metal gratings; metal colloids prepared by reducing a dissolved metal salt with an appropriate reducing agent; metal colloids encapsulated in sol–gel-derived xerogel layers; silver particle-doped cellulose gel films; and microbeads as a carrier of silver colloid. Metal nanoparticles, usually prepared as colloidal suspensions, are one of the most widely studied SERS substrates. They exhibit unique optical properties due to excitation of the surface plasmons. The frequency of the surface plasmon resonance depends strongly on the size, shape, and dielectric environment of the particles. Additionally, when particles are closely spaced, the surface plasmons of individual particles interact to create new resonances dependent upon the interparticle distance and the incident angle and polarization of the light. Thus, optical properties of individual particles and arrays of interacting particles can be conveniently tuned.

Significant progress has been made in controlling the size and shape of metal nanoparticles. Monodisperse single crystal nanocubes, nanorods, and nanoshells have all been prepared using simple methods. The size and shape of gold particles are more easily controlled than those of silver, and consequently silver-clad gold nanoparticles have been prepared and shown to possess distinct optical properties.

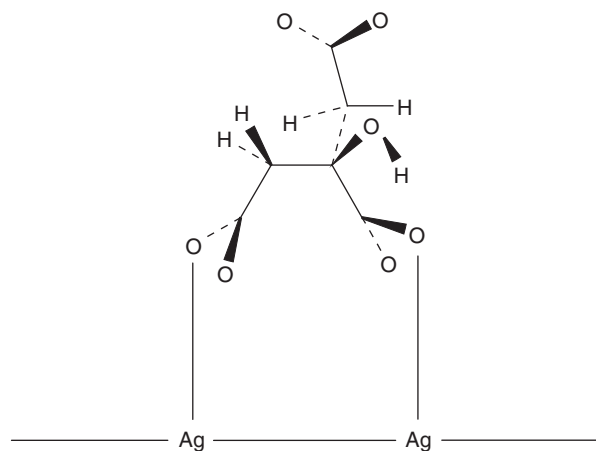
Silver colloid is popular due to the relative ease of manufacture, low cost, and stability. One widely used method of silver colloid preparation is citrate reduction of silver nitrate (Figure 3). This results in an overall net negative charge on the colloid, which is high compared with many preparations (Figure 4). As a result, the colloid is stable for months and even years, and in some cases, with careful attention to detail, quantitative SERRS can be obtained from this colloid with relative standard deviations of less than 5%. There is, however, a change in enhancement with time, and so a standard needs to be used.

## Studies of SERS/SERRS Enhancement

The reason for SERS/SERRS enhancement has been studied by a number of groups. It has been shown that aggregated silver colloids produce fractal



**Figure 3** A representative transmission electron microscopic image of citrate-reduced silver colloid. The particles are mainly hexagonal, with a largest face diameter of  $\sim 36$  nm. A few rods are present. (Reprinted with permission from Faulds K, Littleford RE, Graham D, Dent G, and Smith WE (2004) Comparison of surface-enhanced resonance Raman scattering from unaggregated and aggregated nanoparticles. *Analytical Chemistry* 76: 592–598; © American Chemical Society.)



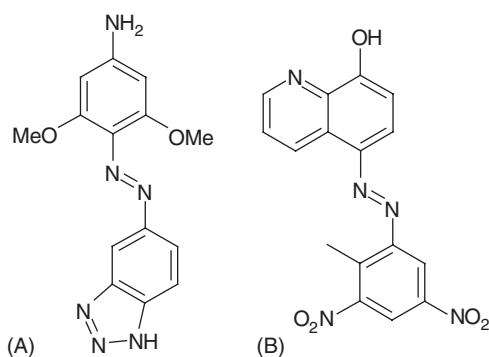
**Figure 4** Representation of the surface of a Lee and Meisel colloidal particle. The citrate is believed to be bonded to the silver, which is present as  $\text{Ag}^+$ , with negatively charged carboxylate groups. (Rodger C, Smith WE, Dent G, and Edmondson M (1996) Surface-enhanced resonance-Raman scattering: An informative probe of surfaces. *Journal of Chemical Society, Dalton Transaction* 791–799; reproduced by permission of The Royal Society of Chemistry.)

clusters. Upon excitation, the electromagnetic field over a fractal object is not distributed uniformly but is localized in hot spots much smaller than the wavelength of light (10–30 nm). A large volume of the fractal therefore remains inactive or active at a lower level, while the portions of the fractal within these hot spots carry the full activity. This

localization of optical excitation indicates that very high electromagnetic fields lead to the huge enhancement of Raman scattering. Other experiments have shown an anomalously high enhancement in the first layer, suggesting the need to consider the charge transfer or chemical mechanism of enhancement as well as the electrochemical enhancement.

Since only analytes attached to or close to metal surface are enhanced, the need for effective surface attachment is crucial if reproducible results are to be obtained. In many early experiments this was widely ignored. Commercial dyes were simply added to colloid or placed on surfaces and the results, where favourable, have been reported in many studies. The problem with this approach is that the surface chemistry of an element such as silver is complex. Little is understood about the surfaces on which the analyte is adsorbed, particularly in aqueous solution or, as is often done, with particles prepared in aqueous solutions and dried out on a surface for investigation. Further, where colloidal suspensions are used, it is common practice to aggregate the colloid in order to shift the frequency of the plasmon to a value that would place it in resonance with the laser. In fact, it is clear from ultraviolet–visible absorption that a range of clusters are made in most conditions and only a few are likely to be in resonance with the laser. However, this gives a bigger SERS enhancement than for single particles.

The aggregating agent usually breaks up a stable colloid by reducing the charge on the surface. This can often happen by chemical action. Under the aerobic conditions used in most experiments, and shown in some to be essential for good SERS, the silver surface in the colloidal suspension is likely to be coated with a silver oxide or related layer containing silver(I) ions. It may also contain other adsorbed molecules from the preparation of the colloid. When a reagent such as sodium chloride is added to this surface, it is likely that a layer of silver chloride will form. This is known to be the case for electrode surfaces. With solid substrates, similar approaches are sometimes taken. For example, using the most commonly used analyte for fundamental studies, rhodamine 6G, the normal process for obtaining the best signals is to activate the surface with the sodium chloride. This particular combination of rhodamine and sodium chloride does give a more effective spectrum than without the addition of sodium chloride. However, studies of other analytes with colloidal suspensions have shown that sodium chloride is not always the most effective aggregating agent and no special activation procedures have been required. Additionally, with higher concentrations of rhodamine, not all the dye adsorbs and large fluorescent



**Figure 5** Chemical structure of two dyes specifically designed to complex strongly to the silver surface for effective SERS. (A) 4(5'-azobenzotriazolyl) 3,5-dimethoxyphenylamine; (B) 5-(2-methyl-3,5-dinitro-phenylazo) quinolin-8-ol.

backgrounds are obtained (see Figure 2). This can limit the use of SERRS and is often a problem where weak signals are obtained, and the temptation is to use higher, not lower, concentrations.

Thus, for quantitative work and for obtaining a better understanding of the effect, effective surface adhesion is essential. In addition, the simpler the nature of the surface, the easier it is to understand the nature of the enhancement. For this reason, a set of special dyes has been developed for SERS/SERRS. These contain groups added to the dye specifically to obtain metal complexing with silver ions present on the surface. Two effective groups of this type are shown in Figure 5. One contains the benzotriazole group, known to be an effective method of preventing tarnishing in silver due to its strong attachment to the surface. The other contains an 8-hydroxy-quinoline group, which will create a complex with silver ions on the metal surface that is insoluble. Both these reagents have proved to be effective for obtaining SERS/SERRS with sodium chloride and other aggregating agents, including organic aggregating agents such as poly-L-lysine.

## Single Molecule Spectroscopy

In 1997 two groups independently claimed single molecule detection using SERS/SERRS using different approaches. In some of the single molecule studies, the difference between SERS and SERRS is ignored, but most, not all, of the studies use chromophores. Nie and Emory reported detection of single rhodamine 6G dye molecules adsorbed on immobilized single silver nanoparticles using SERRS, and Kneipp *et al.* reported detection of a single molecule of crystal violet adsorbed on aggregated clusters of silver particles in a colloidal suspension using near infrared excitation. The large SERS

enhancement was attributed to excitation of the coupled surface plasmons of the colloidal aggregate. The SERS enhancement quoted by both groups for the detection of single molecules is of the order of  $10^{14}$ . Such enhancements are many orders of magnitude greater than predicted ensemble averaged values. This can be attributed to the removal of population averaging effects, whereby all the molecules and particles are assumed to contribute equally to the observed signals. Instead, only the active particle or active site is considered. Other groups have subsequently reported single molecule detection using SERS.

Although detection of a single molecule adsorbed onto an immobilized single silver nanoparticle has been reported, studies on the protein hemoglobin adsorbed on immobilized colloidal particles revealed that the minimum state of aggregation necessary for SERS detection of a single hemoglobin molecule was the dimer. Detection of molecules adsorbed to single particles was not observed. Consequently, it is suggested that the surface enhancement is predominantly electromagnetic in nature and dominated by the increased local electric field between the two particles. Calculations suggest the maximum electromagnetic contribution to SERS enhancement for two interacting spheres to be of the order of  $10^{11}$ , and therefore an additional enhancement of  $10^3$  is required to explain the enhancement factors necessary for single molecule detection. This extra enhancement may be electromagnetic due to further surface roughness present on individual particles or may be chemical in nature. It is proposed that the additional electromagnetic enhancement is obtained only under special positions such as the interstitial site between two particles and outside sharp surface protrusions.

## Applications of SERS

### Sensors

Due to the molecular specificity of SERS/SERRS, combined with its inherent sensitivity, it has found utility in a wide range of applications, only a few of which can be described here. The development of SERS sensor technology has allowed the detection and identification of seven structurally similar monosaccharides in aqueous solution using a sample volume of only  $5\mu\text{l}$  with a concentration of  $1 \times 10^{-2} \text{ mol dm}^{-3}$ . This has led to the initial development of a glucose-based biosensor with true *in vivo*, real time, minimally invasive sensing.

The development of a volatile organic compound sensor using SERS detection has also been reported. The SERS substrate is chemically modified with a

thiol coating to prevent oxidation of the roughened silver surface and attracts the analyte of interest to the SERS surface. Detection of chlorinated solvents, aromatic compounds, and methyl *t*-butyl ether have all been demonstrated.

### Drugs

There is considerable interest in the development of novel platinum-based anticancer drugs that overcome the disadvantages associated with the widely used drug cisplatin, namely its inactivity against some types of tumors and toxic side effects. SERS has been shown to be suitable for the characterization of platinum complexes at physiological concentrations, allowing the determination of binding strengths of different ligands.

The large SERRS intensities from the anticancer drug mitoxantrone have been used to provide a quantitative method of detection. The method uses a flow cell, which causes a dilution of plasma and serum samples, thus overcoming the fluorescent background. Even with these dilution steps the detection limits are superior to the previously published high-performance liquid chromatography (HPLC) method and are sufficiently good to be used for clinical analysis. Since the SERRS technique takes minutes from sample introduction to detection, and the chromatography technique hours, the SERRS technique could be considered the technique of choice for analysis of this drug.

### Intracellular SERS

SERS from living cells has been reported using several techniques. Etched and silver-coated glass fiber tips have been used as the SERS substrate, allowing the recording of spectra of biological samples, such as plant tissue and microbiological cells, with high spatial resolution. The deposition of colloidal gold particles inside single living cells by fluid-phase uptake has been demonstrated, providing strong SERS signals from the native chemical constituents of the cells. The sensitivity of SERS allowed measurements to be made in relatively short collection times (1 s for one mapping point) using 3–5 mW of near infrared excitation. SERS mapping over a cell monolayer with a 1  $\mu$ m lateral resolution showed different Raman spectra at almost all places, reflecting the very inhomogeneous chemical constitution of the cells. This has opened up exciting possibilities for the study of cell biology and biomedical studies.

### Atomic Force Microscopy/Raman

Atomic force microscopy (AFM) uses a fine tip to probe surfaces. It has extremely high resolution. By

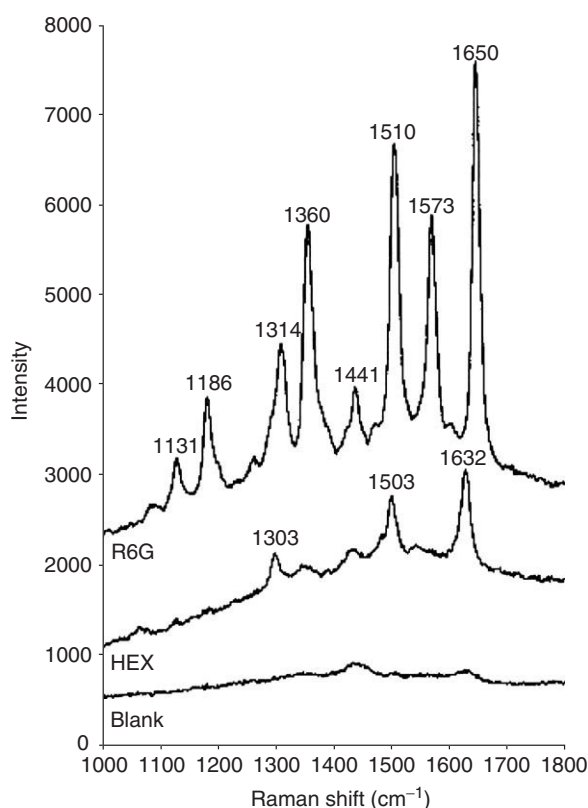
adding a roughened silver coating to AFM tips, a technique has been developed whereby when the AFM tip is touched on the surface it is irradiated with a laser, providing good SERS. The advantage of this technique is that the effective resolution of the technique is the dimension of the tip and consequently is much smaller than optical dimensions. The tip position is known from the AFM scan. This type of method has been used for analysis of thin films, which are undetectable with Raman microprobe systems but easily acquired with a suitably gold-coated AFM tip. More recently, a combined scanning electron microscope/Raman scattering detection system has been created. It is possible to detect the Raman scattering during the time of the SEM scan. This provides detailed structural information on the surface in a noncontact manner, which is faster and sometimes more effective than AFM. Since Raman scattering only occurs from the tip, and the tip position is known from the SEM, the dimensions of the tip are essentially the resolution of the instrument.

### DNA

The single molecule detection capacity of SERRS has been used to develop analytical techniques for DNA. These can use either a substrate or a colloid. In either case, this is most effective when the label is within the monolayer of the surface and hence the largest SERRS enhancement can be obtained. Methods have been developed that enable spatial separation of different events and their detection on a substrate. The use of SERS to monitor DNA hybridization of a fragment of the breast cancer susceptibility gene, BRCA1, on modified silver surfaces has been shown to be effective.

A number of SERRS active DNA probes with different dyes have been synthesized that are used with colloids. Detection limits that are better than those obtained with fluorescence can now be obtained with SERRS. In addition, whereas fluorescence gives a very broad signal, SERRS gives a sharp signal and consequently is much more effective for multiplexing. This technique has been improved by the addition of a flow cell and in particular a lab on a chip flow cell, where control of the process of aggregation through the laminar flow region of the chip has shown a great improvement in quantitation as well as an improvement in sensitivity. The sensitivity of this method is known to be down to single molecule level, and it is possible to obtain this using commercial equipment within a maximum of 10 s. This technique is showing considerable promise for more effective multiplexed forms of DNA analysis (Figure 6). Recently, multiplexing of six probes with





**Figure 6** SERRS from 2,5,1',3',9'-hexachloro-6-carboxyfluorescein (HEX) and rhodamine 6G labeled oligonucleotides at  $2 \times 10^{-9} \text{ mol l}^{-1}$  (1 s accumulation time). The unique SERRS spectral fingerprints from each dye allow multiplex analysis of SERRS/DNA complexes. (Reproduced with permission from Graham D, Mallinder BJ, and Smith WE (2000) Detection and identification of labeled DNA by surface enhanced resonance Raman scattering. *Biopolymers (Biospectroscopy)* 57: 85–91; © John Wiley & Sons, Inc.)

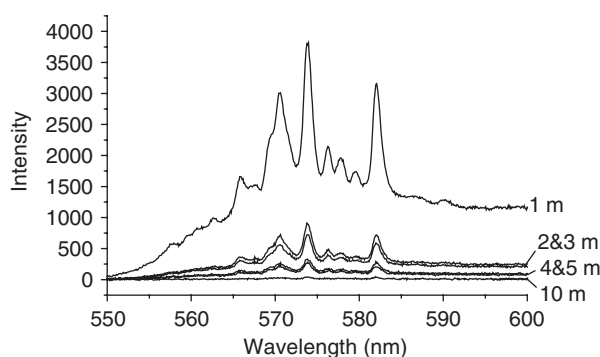
gold nanoparticles has been demonstrated, indicating the large multiple detection capability for SERRS/DNA assays.

### Distance Detection

The ability to obtain sharp molecularly specific spectra from SERRS prompted the concept that a number of codes could be written simply by changing the combination of analytes and by changing their concentration. In this way, it is believed, hundreds of thousands of codes can be written that can be read using a Raman spectrometer without any spatial separation. The initial demonstration for this for one dye has shown that Raman scattering from suitable polymer films can be detected at up to 10 m (Figure 7) and therefore there is considerable potential for further expansion in this direction.

### Proteins

One of the first important demonstrations of SERRS was to detect heme groups in proteins. It is a good



**Figure 7** Change in spectral intensity as the distance from the sample is increased. (Reproduced with permission from McCabe A, Smith WE, Thomson G, *et al.* (2002) Remote detection using surface enhanced resonance Raman scattering. *Appl Spectrosc* 56: 820–826.)

example of the selectivity of the method in that practically no signals are obtained from the protein other than from the heme group. The heme on the other hand can be detected down to below monolayer coverage, although a monolayer is often used since it helps prevent protein degradation. A silver colloid stabilized with citrate layers also helps. The advantage of SERRS is that marker bands are available for the oxidation state, spin state, and coplanarity of the vinyl groups with the heme group. In addition, compared with resonance Raman scattering, it is easier to obtain SERRS from weaker chromophore absorption bands such as the Q band.

### Bacteria

The treatment of bacteria with sodium borohydride provides a nucleating substrate for the reduction of silver ions, forming a rough silver metal coating around the microorganism. Intense SERS of the coated bacteria showed four different types of bacteria to all produce similar spectra, suggesting that the spectra are selective and sensitive to a specific molecular species that dominates the spectra and that is found in all the bacteria analyzed.

**See also:** Blood and Plasma. Clinical Analysis: Glucose. DNA Sequencing. Fluorescence: Overview. Forensic Sciences: Drug Screening in Sport. Microscopy Techniques: Electron Microscopy; Scanning Electron Microscopy; Atomic Force and Scanning Tunneling Microscopy. Nucleic Acids: Spectroscopic Methods. Raman Spectroscopy: Instrumentation. Sensors: Overview.

### Further Reading

Albrecht MG and Creighton JA (1977) Anomalously intense Raman spectra of pyridine at a silver electrode. *Journal of the American Chemical Society* 99: 5215–5217.

- Billmann J and Otto A (1982) Electronic surface state contribution to surface enhanced Raman scattering. *Solid State Communications* 44: 105–107.
- Creighton JA (1998) The selection rules for surface enhanced Raman spectroscopy. In: Clark RHJ and Hester RE (eds.) *Spectroscopy of Surfaces*, ch. 2, p. 37. New York: Wiley.
- Emory SR and Nie S (1997) Probing single molecules and single nanoparticles by surface enhanced Raman scattering. *Science* 275: 1102–1106.
- Fleischmann M, Hendra PJ, and McQuillan AJ (1974) Raman spectra of pyridine adsorbed at a silver electrode. *Chemical Physics Letters* 26: 163–166.
- Jeanmaire DL and Van Duyne RP (1977) Surface Raman spectroelectrochemistry. Part 1. Heterocyclic, aromatic, and aliphatic amines adsorbed on the anodized silver electrode. *Journal of Electroanalytical Chemistry* 84: 1–20.
- Kneipp K, Kneipp H, Irving I, Dasari RR, and Feld MS (2002) Surface enhanced Raman scattering and biophysics. *Journal of Physics: Condensed Matter* 14: R597–R624.
- Kneipp K, Wang Y, Kneipp H, *et al.* (1997) Single molecule detection using surface enhanced Raman scattering (SERS). *Physical Review Letters* 78: 1667–1773.
- Moskovitis M (1978) Surface roughness and the enhanced intensity of Raman scattering by molecules adsorbed on metals. *Journal of Chemical Physics* 69: 4159–4161.
- Otto A, Mrozek I, Grabhorn H, and Akemann W (1992) Surface enhanced Raman scattering. *Journal of Physics: Condensed Matter* 4: 1143–1212.
- Stacy AM and Van Duyne RP (1983) Surface enhanced Raman and resonance Raman spectroscopy in a non-aqueous electrochemical environment: Tris(2,2'-bipyridine) ruthenium (II) adsorbed on silver from acetonitrile. *Chemical Physics Letters*. 102: 365–370.
- Xu HX, Aizpurua J, Kall M, and Apell P (2000) Electromagnetic contributions to a single-molecule sensitivity in surface enhanced Raman scattering. *Physical Review E* 62: 4318–4324.
- Xu HX, Bjerneld EJ, Kall M, and Borjesson L (1999) Spectroscopy of single haemoglobin molecules by surface enhanced Raman scattering. *Physical Review Letters* 83: 4357–4360.

## REAGENTS

See **ANALYTICAL REAGENTS: Specification; Purification. RADIOCHEMICAL METHODS: Radio-Reagent Methods**

## REDOX INDICATORS

See **INDICATORS: Redox**

## REDOX TITRATION

See **TITRIMETRY: Overview; Potentiometric; Photometric**

## REFERENCE MATERIALS

See **QUALITY ASSURANCE: Reference Materials; Production of Reference Materials**

## REFLECTOMETRY

See **OPTICAL SPECTROSCOPY: Refractometry and Reflectometry**

# REFRACTOMETRY

See OPTICAL SPECTROSCOPY: Refractometry and Reflectometry

## REMOTE GAS SENSING

Contents

### Overview

#### Integrated-Path Remote Sensing

#### Range-Resolved Remote Sensing

### Overview

**M J T Milton**, National Physical Laboratory,  
Teddington, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Remote gas sensing is the process of measuring, or detecting, a gas at some point that is distant from the location where the analysis is carried out.

Various methods have been developed for remote gas sensing. These include: differential optical-absorption spectroscopy (DOAS), differential absorption lidar (DIAL), and a number of methods that use spectroscopic methods with an atmospheric path in place of a laboratory long-path cell, for example: tunable diode laser absorption spectroscopy (TDLAS) and Fourier transform infrared (FTIR) spectroscopy.

Generally, remote gas-sensing methods are sensitive to the presence of the target species at extended ranges, either intrinsically or as a result of their operation over different paths. Consequently, remote gas-sensing methods have applications to the measurement of poorly mixed atmospheres or when an averaged value for the concentration along a path is required. They also have advantages over point-sensing methods when measuring gases in locations that are inaccessible (e.g., at altitudes significantly above ground level) or whose location within a specified area is not known exactly.

Remote gas-sensing methods have often been considered to be competitors to point gas-sensing methods, which are only able to measure gases at the location of the apparatus. The examples given here

show that the two approaches are complementary and that remote gas-sensing methods are most likely to be applied effectively in applications that cannot be addressed by point sensing methods.

There are various methods that meet the definition of remote gas sensing in part but are not discussed here. These include sensors linked by optical fibers, or operated along the length of an optical fiber. None of these methods depend on atmospheric propagation, which is the common feature of the methods discussed here.

### Measurement Principle

In this overview, the basic principle of all optical remote gas-sensing methods is introduced and is used to explain how the accuracy and sensitivity with which a target gas can be measured are defined by the spectroscopic properties of the gas itself. These properties define the region of the spectrum in which it has significant absorption features together with their strength and structure.

The relationship between the intensity of a beam of electromagnetic radiation before and after it has traversed a path from range  $r' = 0$  to  $r' = r$  is given by the Beer–Lambert law:

$$I(r) = I(0) \exp \left( - \int_0^r [A(r') + \alpha C(r')] dr' \right)$$

where  $I(0)$  is the intensity of the light launched along the path and  $I(r)$  is its intensity at range  $r$ . The exponent includes all of the terms relating to the extinction and scattering of the beam. Since the atmosphere itself determines them, they are functions of the range. They can be conveniently divided into two groups. First, the absorption by the target species is given by the product  $\alpha C(r)$ , where  $\alpha$  is the

absorption coefficient and  $C(r)$  is the concentration at range  $r$ . Second,  $A(r)$  is the total optical attenuation along the path arising from all other mechanisms. These might include:

- Mie or Rayleigh scattering by aerosols,
- absorption by aerosols or other species, and
- absorption losses in the transmission or receiving optics of the remote gas-sensing system itself.

Both  $\alpha$  and  $A(r)$  are functions of the wavelength of the radiation as well as the range. A major challenge in practical applications of some remote gas-sensing methods is to distinguish changes in  $\alpha C$  from those in  $A$ . This is equivalent to establishing a 'zero level' for the measurement and is discussed under the section Calibration.

It can be seen from this simple analysis that the fundamental performance of a remote gas-sensing method in detecting a specific target gas is determined by  $\alpha$ . The spectral regions in which it might be possible to detect the target species are determined by the strength and structure of  $\alpha$  at wavelengths at which there is small or negligible absorption by other species present in the atmosphere.

The performance of any specific remote gas-sensing system is determined by its capability to detect changes in  $I(r)$  with respect to  $I(0)$ . This is usually expressed in terms of the signal-to-noise ratio.

## Absorption of Target Gases and of the Atmosphere

The monitoring of target gases in the atmosphere has to be undertaken in the presence of absorption by the gases normally present in the atmosphere. This imposes some important practical limitations on remote gas sensing.

In the ultraviolet spectral region, the short-wavelength monitoring limit in the atmosphere is set by the absorption of atmospheric oxygen. It absorbs weakly in the 260–230 nm range, and then increasingly strongly until at a wavelength of  $\sim 210$  nm the atmosphere is almost completely opaque. A 'clean' atmosphere is essentially transparent from this limit through the ultraviolet and visible spectral regions. A number of common atmospheric pollutants absorb in this region, including nitrogen monoxide, sulfur dioxide, nitrogen dioxide, ozone, and the single aromatic ring hydrocarbons. Hence, all of these species are readily detected by remote gas sensing in the ultraviolet spectral region.

In the infrared spectral region, some wavelength regions are inaccessible to remote gas sensing because of absorption by atmospheric carbon dioxide

**Table 1** Gases with characteristic absorption spectra that are potentially detectable by remote gas-sensing methods

220–900 nm spectral region	2–5 $\mu\text{m}$ spectral region	8–13 $\mu\text{m}$ spectral region
SO <sub>2</sub>	CO <sub>2</sub>	NH <sub>3</sub>
NO	CO	HNO <sub>3</sub>
SO <sub>2</sub>	N <sub>2</sub> O	H <sub>2</sub> S
O <sub>3</sub>	NO <sub>2</sub>	C <sub>2</sub> H <sub>2</sub>
Hg	HCl	C <sub>2</sub> H <sub>4</sub>
Benzene	HBr	C <sub>6</sub> H <sub>6</sub>
Toluene	HCN	O <sub>3</sub>
Xylenes	CH <sub>4</sub>	SO <sub>2</sub>
H <sub>2</sub> O	C <sub>2</sub> H <sub>2</sub>	SF <sub>6</sub>
O <sub>2</sub>	C <sub>2</sub> H <sub>4</sub>	CH <sub>3</sub> Cl
	C <sub>2</sub> H <sub>6</sub>	C <sub>2</sub> H <sub>3</sub> Cl
	C <sub>3</sub> H <sub>8</sub>	COCl <sub>2</sub>
	Higher hydrocarbons	CH <sub>3</sub> CHO
	COS	
	H <sub>2</sub> S	
	NH <sub>3</sub>	

(particularly the ranges 3–4  $\mu\text{m}$  and greater than 14  $\mu\text{m}$ ) and water (particularly in the ranges 2.5–3  $\mu\text{m}$  and 5–8  $\mu\text{m}$ ). Between these regions of absorption, there are substantial 'window' regions, which are free from interference by carbon dioxide and water, and which can be used for remote gas sensing. Since the amount of water vapor in the atmosphere varies, it is necessary to take great care with measurements made close to the edges of these window regions.

Table 1 lists some of the gases that can be observed in the 'window' regions of the atmosphere. It shows that a large number of target gases can be detected by remote gas-sensing systems that are capable of operation at the correct wavelengths.

Gas absorptions in the ultraviolet are due to electronic transitions in the target species and are relatively strong. They lead to well-defined line spectra that allow gases to be detected with considerable sensitivity. When combined with the use of detectors with near-quantum limited performance this leads to detection sensitivities of a few nanomoles per mole over typical path lengths.

The strongest infrared absorptions are due to the fundamental vibrations of molecules, and are mainly at wavelengths longer than 2.5  $\mu\text{m}$ . These are usually in bands of lines at approximately equal spacing. Weaker 'overtone' bands occur at shorter wavelengths. It is also possible to carry out remote gas sensing in the microwave region. The performance of detectors in these regions is poorer than in the ultraviolet. Consequently, remote gas sensing in the infrared is generally less sensitive than the ultraviolet or visible spectral regions and typical detection sensitivities are of the order of a few micromoles per mole over typical path lengths.

## Sensitivity and Specificity

The key characteristics of all remote gas-sensing systems are defined by a combination of the specifications of the apparatus (e.g., transmitter and receiver performance), the target species (e.g., absorption strength and bandwidth), and the atmosphere (e.g., interfering species and atmospheric attenuation). Knowledge of all of these leads to the definition of a number of parameters that characterize the method.

The first of these is the sensitivity. This expresses the minimum detectable concentration (often as a function of range) and is usually quoted in the absence of interference. It is defined mathematically by equating the product  $\alpha C$  with the noise level of the system. It is usually stated in units of path-integrated concentration, most correctly in units of  $\mu\text{g m}^{-2}$ , or alternatively in ppm m (where ppm is used to express the partial pressure of the target species as millionths of the total atmospheric pressure).

A specification of range-resolved remote gas-sensing systems that is closely related to the sensitivity is the maximum range. This is simply the range at which the sensitivity passes some defined threshold, such as the path-integrated concentration being equal to a stated value.

The second parameter used to specify the performance of a remote gas-sensing method is the specificity (or selectivity). This expresses the ability of the method to measure the target in the presence of other species. It is defined mathematically by equating  $\alpha C$  to  $A$ . The specificity will usually be expressed in the same units as the sensitivity.

This short discussion shows that the key parameters used to characterize the performance of a remote gas-sensing method are closely related to those used to characterize any spectroscopic instrument together with the added feature that the performance of a remote gas-sensing system is also related to the range over which the measurement is made.

## Types of Remote Gas-Sensing System

The emphasis of this overview has been on characterizing a remote gas-sensing system in terms of its sensitivity and specificity. It is also possible to group systems according to their mode of operation.

The capability of a system to measure a range of different species is determined by its ability to operate at different wavelengths. Single-wavelength systems (e.g., DIAL) are usually capable of measuring a single species, although they can sometimes operate with a wavelength tunable source to measure additional species. Multi-wavelength systems (e.g., DOAS and FTIR) operate over a wide wavelength

range and are intrinsically able to measure more than one species simultaneously.

Remote gas-sensing systems can also be grouped according to the type of equipment used at the 'remote' end. For example, double-ended systems, such as DOAS, have a remote source, while TDLAS is often operated with a remote reflector or detector. The principal example of a single-ended system is DIAL, which depends on atmospheric backscatter and therefore has a single apparatus that includes both the transmitter and the receiver.

## Advantages and Disadvantages of Remote Gas Sensing

Calculated values for the sensitivity and specificity of remote gas-sensing systems will generally appear to be poorer than many point sensing methods that are also less complex and available at lower cost. However, they must be judged on the basis of their capability to carry out measurements that are not readily performed using point sensing methods, for example:

- Remote gas sensing results represent a spatial average over an extended path. This may be extremely useful when concentrations vary spatially or temporally, or when the point concentration is too low to be detected directly, but the integrated-path concentration is above the detection threshold.
- The results of remote gas-sensing measurements can be used to estimate a flux of the target gas across the measurement path.
- Measurements can be made along a path that cannot be accessed directly. This could include passing through regions that are hostile because of high temperatures or pressures or have high concentrations of corrosive or toxic gases, for example, in measuring across emission sources such as stacks.
- Appropriate movements of the measurement path can lead to the development of a two- or three-dimensional map of the target species. In this way, a remote gas-sensing system may be able to replace several monitors carrying out point sensing.
- In an extension of the principle of mapping, a well-chosen path for the beam may have a greatly increased probability of locating emissions from a source whose location is not known.

There are some general limitations of remote gas sensing; the cost and complexity of remote gas-sensing monitors is often much greater than point monitors, which do not have the advantages listed above. This is not always the case for remote gas-sensing monitors designed for very specific measurements of a single gas, but is usually the case for monitors with

the capability to measure more than one species. Finally, there are some complex issues relating to the calibration and validation of remote gas-sensing methods that are discussed below.

In general, remote gas-sensing methods cannot match the sensitivity of similar techniques deployed with long pathlength cells under laboratory conditions. Similarly, they cannot match the specificity of the best nonoptical techniques (e.g., gas chromatography or mass spectroscopy).

## Calibration

The acceptance of any monitoring method into the widest range of applications is not dependent solely on its principle of operation, but also on the demonstration and widespread recognition that the data it produces are valid. This is usually achieved by operating within a recognized quality assurance regime together with the use of an appropriate calibration procedure.

In the early applications of remote gas-sensing systems, great emphasis was placed on the fact that they operated according to a well-validated 'law' of physics and consequently it was claimed that they were 'self-calibrating' when used with valid spectroscopy for the target species. Although, there is some truth behind such statements, they do not provide sufficient transparency to nonexpert users of the data to gain the acceptance of standards-making bodies. In particular, issues such as the bandwidth of the source, spectral interference, nonlinearity of the detector system, and all other wavelength-dependent effects in the optical system itself undermine this simplistic claim.

Direct calibration of the performance of a remote gas-sensing system can be achieved by placing a cell containing a known concentration of the target species in the path. This approach has been demonstrated in practice for all of the methods discussed here, but is limited by the practical difficulties of making a cell with windows that are completely transparent at the measurement wavelengths.

Although the principle of calibration of remote gas sensing monitors is straightforward – if inconvenient, a further difficulty lies with establishing the 'zero' level for the measurement. This is readily achieved in point-sensing instruments by flushing them with

'zero gas', but this is not possible for any instrument operating over an extended atmospheric path since it can never be assumed that the target species or an interfering species are not present. Alternative approaches to resolving this problem are by switching to a shorter or clearer path. It should be noted that DIAL, which intrinsically measures the concentration gradient and those methods that use spectral background fitting, such as DOAS and FTIR, do not have this drawback.

## Selected Applications

Remote gas-sensing methods have been demonstrated to be highly effective in a large number of applications. These include:

- direct measurements of gas fluxes from a wide area,
- the generation of two- or three-dimensional concentration maps of target gases,
- scanning along a line or across an area to provide an 'alarm'-type warning of the buildup of toxic or flammable gases,
- monitoring of gas concentrations at the boundary of an industrial plant in order to identify significant direct or fugitive emissions from the area,
- measurements along elevated paths, including measurements at heights substantially above ground level, and
- measurements of the lower and middle atmospheres from satellites.

*See also:* **Remote Gas Sensing:** Integrated-Path Remote Sensing; Range-Resolved Remote Sensing. **Sensors:** Overview.

## Further Reading

- Byer RL (1975) Remote air pollution measurement. *Optical and Quantum Electronics* 7: 147–177.
- Vaughan G (2002) In: Holton JR, Pyle J, and Currie JA (eds.) *Encyclopaedia of Atmospheric Science*. Amsterdam: Elsevier.
- Vaughan JM (2004) Remote optical sensing and measurement. In: Dakin J and Brown R (eds.) *Handbook of Optoelectronics*. Bristol: IOP Publishing.

# Integrated-Path Remote Sensing

**M J T Milton**, National Physical Laboratory,  
Teddington, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Integrated-path remote gas sensing is the process of measuring, or detecting, a target gas along an open



the capability to measure more than one species. Finally, there are some complex issues relating to the calibration and validation of remote gas-sensing methods that are discussed below.

In general, remote gas-sensing methods cannot match the sensitivity of similar techniques deployed with long pathlength cells under laboratory conditions. Similarly, they cannot match the specificity of the best nonoptical techniques (e.g., gas chromatography or mass spectroscopy).

## Calibration

The acceptance of any monitoring method into the widest range of applications is not dependent solely on its principle of operation, but also on the demonstration and widespread recognition that the data it produces are valid. This is usually achieved by operating within a recognized quality assurance regime together with the use of an appropriate calibration procedure.

In the early applications of remote gas-sensing systems, great emphasis was placed on the fact that they operated according to a well-validated 'law' of physics and consequently it was claimed that they were 'self-calibrating' when used with valid spectroscopy for the target species. Although, there is some truth behind such statements, they do not provide sufficient transparency to nonexpert users of the data to gain the acceptance of standards-making bodies. In particular, issues such as the bandwidth of the source, spectral interference, nonlinearity of the detector system, and all other wavelength-dependent effects in the optical system itself undermine this simplistic claim.

Direct calibration of the performance of a remote gas-sensing system can be achieved by placing a cell containing a known concentration of the target species in the path. This approach has been demonstrated in practice for all of the methods discussed here, but is limited by the practical difficulties of making a cell with windows that are completely transparent at the measurement wavelengths.

Although the principle of calibration of remote gas sensing monitors is straightforward – if inconvenient, a further difficulty lies with establishing the 'zero' level for the measurement. This is readily achieved in point-sensing instruments by flushing them with

'zero gas', but this is not possible for any instrument operating over an extended atmospheric path since it can never be assumed that the target species or an interfering species are not present. Alternative approaches to resolving this problem are by switching to a shorter or clearer path. It should be noted that DIAL, which intrinsically measures the concentration gradient and those methods that use spectral background fitting, such as DOAS and FTIR, do not have this drawback.

## Selected Applications

Remote gas-sensing methods have been demonstrated to be highly effective in a large number of applications. These include:

- direct measurements of gas fluxes from a wide area,
- the generation of two- or three-dimensional concentration maps of target gases,
- scanning along a line or across an area to provide an 'alarm'-type warning of the buildup of toxic or flammable gases,
- monitoring of gas concentrations at the boundary of an industrial plant in order to identify significant direct or fugitive emissions from the area,
- measurements along elevated paths, including measurements at heights substantially above ground level, and
- measurements of the lower and middle atmospheres from satellites.

*See also:* **Remote Gas Sensing:** Integrated-Path Remote Sensing; Range-Resolved Remote Sensing. **Sensors:** Overview.

## Further Reading

- Byer RL (1975) Remote air pollution measurement. *Optical and Quantum Electronics* 7: 147–177.
- Vaughan G (2002) In: Holton JR, Pyle J, and Currie JA (eds.) *Encyclopaedia of Atmospheric Science*. Amsterdam: Elsevier.
- Vaughan JM (2004) Remote optical sensing and measurement. In: Dakin J and Brown R (eds.) *Handbook of Optoelectronics*. Bristol: IOP Publishing.

# Integrated-Path Remote Sensing

**M J T Milton**, National Physical Laboratory, Teddington, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Integrated-path remote gas sensing is the process of measuring, or detecting, a target gas along an open

path defined between a transmitter and a receiver. The quantity measured by integrated-path gas sensing is the path-integrated concentration defined as  $\int \alpha C(r') dr'$ , where  $C(r')$  is the concentration at range  $r'$  and  $\alpha$  is the absorption coefficient at the wavelengths of the radiation used.

The simplest methods for integrated-path remote gas sensing use a receiver and a remote source of radiation as a transmitter. The measurement path is defined between the transmitter and the receiver. A simple development is through the use of a remote reflector (either a plane or a retro-reflecting mirror) to return the radiation to the receiver. The transmitter and receiver can then be co-located. In this configuration, the measurement is made over the return path between the transmitter and receiver.

Four different methods used for integrated-path remote gas sensing are discussed here. One of these (tunable diode laser absorption spectroscopy, TDLAS) uses a narrow linewidth source of radiation (usually a laser diode) and the other three methods use broadband sources of radiation. These three analyze the spectrum of the radiation after it has traversed the atmospheric path in different ways: both differential optical absorption spectroscopy (DOAS) and Fourier transform infrared (FTIR) spectroscopy analyze the entire spectrum over the spectral region of interest, whilst absorption correlation methods record the spectrum after it has been filtered optically with either an optical filter or a sample of the target gas itself. These four methods use an 'active' source of radiation. It is also possible to carry out integrated-path remote gas sensing using a 'passive' source.

Integrated-path remote gas sensing methods are the most widely used of all methods for remote gas sensing. This is because they have greater sensitivity than range-resolved methods since the signal strength is substantially higher. Consequently, they can be used with lower-powered sources than range-resolved remote gas sensing systems and are therefore less costly and complex.

## Tunable Diode Laser Absorption Spectroscopy

Many of the first demonstrations of remote gas sensing used integrated-path absorption usually in conjunction with a laser source. Examples included helium–neon, argon-ion, and carbon dioxide lasers. Subsequently, some successful practical devices were also made that used incandescent sources. However, the small size and high efficiency of diode lasers have given them a major application in this field as is reflected in the description 'TDLAS'.

In recent years, the field has been dominated by the use of tunable diode lasers. Initially, tunable cryogenic lead-salt diodes were used because they were the only practical means of providing tunable continuous-wave radiation in the infrared spectral region where many of the most important target gases absorb. More recently, distributed-feedback and grating-tuned diodes have become widely available at wavelengths close to the telecommunications bands in the near-infrared. In the future, it is expected that quantum cascade lasers, which are capable of tunable operation at wavelengths longer than 5  $\mu\text{m}$ , will find applications in integrated-path remote gas sensing.

An important part of the operation of any integrated-path remote gas sensing method is that in order to achieve a quantitative measurement that is distinct from the background signal, it is necessary to impose some form of modulation on the source radiation. The simplest way to do this is to choose two wavelengths, one of which is strongly absorbed by the target species and the other is not. Since the absorption of the target gas changes with wavelength, the intensity of the beam traversing a path containing the target will also vary as the wavelength changes.

A more sophisticated method is to modulate the wavelength of the beam such that the intensity of the measured signal is modulated at the same frequency. Some type of data processing can then make use of the modulation signal to recover this part of the measured intensity. More sophisticated versions of this approach make use of detection at the second or higher harmonics of the modulation frequency.

Figure 1 shows a schematic diagram of an example of a TDLAS system. The source is shown together with a chopper to modulate the transmitted radiation. It is usually possible to do this directly, in which case the chopper is not required. A retro-reflector is used to return the beam to the monitor and a gas cell

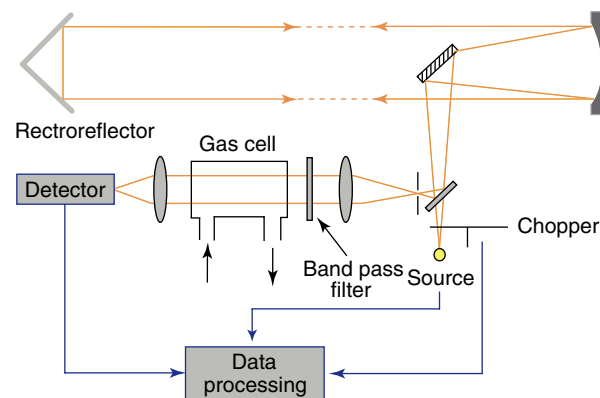


Figure 1 Principal components of a TDLAS monitor.

is shown that can be used as part of a calibration procedure.

The most common application of TDLAS is to the measurement of methane. This has been extremely successful, because methane has a well-defined line spectrum and is detected at high concentrations, usually close to its flammability limit.

## Differential Optical Absorption Spectroscopy

The TDLAS method described above uses differential absorption or wavelength modulation and operates over a narrow spectral region. In contrast, the DOAS method employs a broadband source and a spectrometer within the receiver to disperse the returned signal prior to detection. Consequently, a much larger spectral range including many spectral lines can be measured and a number of different target gases can be measured simultaneously. **Figure 2** shows a schematic diagram of a DOAS system. This example uses a polychromator together with a photodiode detector array.

DOAS systems are usually operated in the ultraviolet or visible spectral regions where high-intensity sources (such as xenon arcs) are available. In some applications to atmospheric measurements, very high sensitivity is achieved by using a measurement path through the whole atmosphere with the sun acting as the source. Radiation is dispersed by the spectrometer and detected in such a way that a complete spectrum of the source, as modified by absorption along the atmospheric path, is recorded. The spectrum is scanned repetitively, with each scan taking less than  $\sim 10$  ms in order to avoid significant intensity changes caused by atmospheric turbulence during the scan. Successive scanned spectra are then added over a predetermined time period and stored. The resulting atmospheric spectrum is fitted to the stored

spectra of the target gases, using a least-squares fitting method, in order to obtain the best estimate of the concentrations of any target gases present in the atmosphere. The spectrum of the source itself and the influence of other atmospheric effects are removed during the data processing using a similar spectrum taken over a very short optical path or by fitting a low-order polynomial curve to establish the background. The residual deviation between the fitted and recorded spectra gives a measure of the uncertainty in the determination of each target gas.

The DOAS technique has been used extensively in the ultraviolet and visible spectral regions for measurements of gases including sulfur dioxide, nitrogen dioxide, ozone, methanol, and aromatic compounds at  $\text{nmol mol}^{-1}$  levels and at distances of up to 1 km. It has even been used, in modified form, to measure concentrations of the OH radical, which is important in studies of atmospheric chemistry, but is only present at very low concentrations.

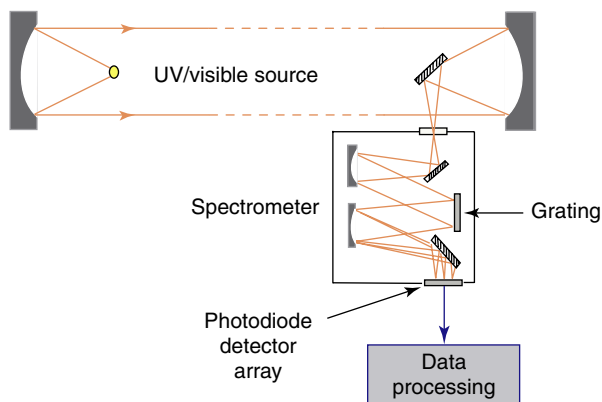
## Fourier Transform Infrared

The general principle of operation of the FTIR method is similar to the DOAS method, because they both measure and analyze over a broad region of the spectrum. The principal difference is that an FTIR spectrometer is used instead of a conventional dispersive monochromator. The benefits of using an FTIR in place of a monochromator can be summarized in terms of two types of advantage:

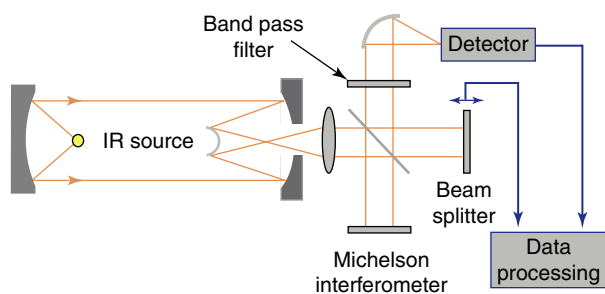
- Radiation from the entire spectrum is sampled throughout the measurement period; in general the duration of the measurement period limits the resolution of the measurement and not the spectral range.
- A much larger optical throughput is possible than with a dispersive monochromator with the same resolution, hence the signal-to-noise ratio is larger.

Long-path FTIR monitors are usually used in a 'bistatic' configuration with a source at the far end of the measurement path. Alternatively, the sun can be used as the source in order to allow the measurement of gases at any altitude along the atmospheric path between the sun and the observer. This approach is usually used to measure stratospheric gases from a site on the ground or from an aircraft, balloon, or satellite.

**Figure 3** shows the principal components of an integrated-path FTIR monitor. The absorption spectrum of the atmosphere within the band pass of the filter over the measurement path is recorded by the Michelson interferometer. This spectrum is then used



**Figure 2** Principal components of a DOAS monitor.



**Figure 3** Principal components of an FTIR monitor.

to identify and quantify any target gases by reference to their component spectra in the total spectrum. Early FTIR monitors used high spectroscopic resolution (typically  $0.05\text{ cm}^{-1}$ ) to record spectra with sufficient resolution to resolve the narrowest atmospheric lines completely. More recently, systems with lower resolution (typically  $1\text{ cm}^{-1}$ ) have come into use because they have sufficient spectroscopic resolution to achieve the accuracy required in most applications and allow the use of more compact and lower-cost FTIR spectrometers.

Multiple-species measurement in near real-time is possible with FTIR systems by comparing the atmospheric spectra obtained with stored spectra of a range of possible gases, using a 'least-squares' approach. Gas concentrations of tens of  $\text{nmol mol}^{-1}$  are typically measurable over paths of a few hundred meters. However, there can be problems of interference from other atmospheric species (especially water vapor) and due to the failure of Beer's Law because of the limited spectral resolution. These make it difficult to fit stored spectra accurately and automatically, and skilled operator intervention is generally required.

Integrated-path FTIR systems have been used for a variety of applications, including monitoring of volatile organic compounds from industrial plant, monitoring toxic and irritant gases in urban areas, and measuring gas release rates from coal mines and landfill sites.

## Absorption Correlation Spectroscopy

Absorption correlation spectroscopy works by making a real-time comparison between a portion of the atmospheric absorption spectrum and the absorption of a sample of the target gas itself. This contrasts with the DOAS and FTIR methods, which record the same type of spectrum over the atmospheric path, but recover the concentration of the target gas mathematically by reference to known spectra of the pure target gas. Two correlation methods are in use.

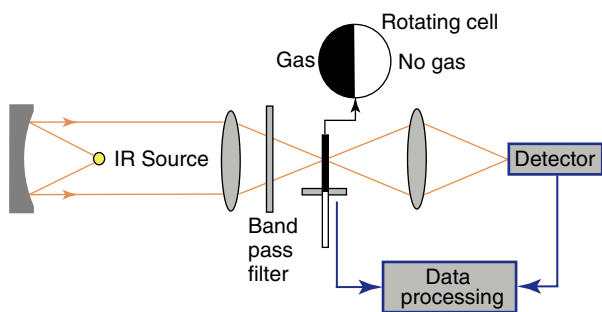
In 'dispersive correlation spectroscopy', radiation that has traversed an atmospheric path is dispersed by a monochromator. The spectrum is dispersed spatially onto a rotating disk with a series of slits. This disk contains in one-half slits that have been cut at points that correspond to wavelengths at which the target gas has absorption maxima, and in the other half slits at points where the target gas has absorption minima. A detector behind the disk collects the transmitted radiation. By evaluating the correlation between the signals from the two portions of the disc the analyzing electronics produce an output that is sensitive to the target gas but relatively insensitive to any other gases.

In 'gas filter correlation spectroscopy', the spectrometer and rotating disk are replaced by a cell containing a sample of the target gas at a known concentration (or by a pneumatic detector containing the target gas). Radiation traversing the atmospheric path then passes through either the cell of target gas or a similar empty cell, both followed by a detector. If any target gas appears in the atmospheric path then the detector signal for radiation that has passed through the empty cell will be correspondingly diminished. However, the signal from the radiation that has been passed through the target gas cell will not be reduced as much because the intensity at the principal wavelengths absorbed by the target gas along the atmospheric path will have been reduced. Subtraction of the target gas signal from the empty cell signal then gives an output that is sensitive to the target gas but insensitive to other gases.

Gas filter correlation is more attractive than dispersive correlation because it does not require a monochromator or a chopping disk with a specific slit pattern. Furthermore, it performs its correlation with the actual gas itself, giving a near ideal correlation between the target gas in the atmosphere and the cell. This technique is commonly known as non-dispersive infrared (NDIR) measurement, since there is no spatial dispersion of the wavelengths in the gas-analyzing beam. The principal components in an integrated-path gas correlation monitor are shown in Figure 4.

Both correlation methods require a target gas with a distinctive narrow-line spectrum, which restricts them largely to diatomic and triatomic species. The less specific the spectrum, in terms of the presence of well-defined spectral lines, the more interference will be caused by other gases.

The main advantages of gas correlation monitors are that they are relatively simple and can be constructed at low cost. They use a single detector for each gas and have good discrimination against other gases if the target gas has a distinctive spectrum.



**Figure 4** Principal components of a nondispersive gas correlation monitor.

Although they can be set up to monitor several gases by splitting the radiation into several different paths, they are generally only used to monitor one target species without modification of the correlation element.

NDIR are much used in the laboratory and in industry over short gas paths for gas concentration measurement. Recently, they have been applied to the longer paths required for 'stack gas' monitoring and the much longer paths (hundreds of meters) required for ambient air monitoring. Measurements of gases such as methane, ethane, hydrochloric acid, carbon monoxide, and carbon dioxide have also been demonstrated at industrial sites and roadsides.

## Passive Methods

All of the methods discussed above are described as 'active' because they make use of an 'active' source of radiation as part of the system. It is also possible

to measure gases remotely using passive methods that detect radiation emitted thermally by the target gas itself or by other objects within the field of view. Examples of this type of system are sensors used with a small number of spectral channels mounted on satellites and used to carry out qualitative identification of features on the earth's surface, such as vegetation or mineral species. Hyper-spectral imagers with a large number of spectral channels are capable of better qualitative measurements of the composition of objects in the field of view or along the measurement path. However, the derivation of true quantitative results from such instruments is problematic because of the difficulty of obtaining accurate information about the emissivity of the objects and their temperature. Nevertheless, this type of approach holds much promise for future development.

See also: **Remote Gas Sensing: Overview**; **Range-Resolved Remote Sensing**.

## Further Reading

VDI 4211 (2000) *Remote Sensing: Atmospheric Measurements near Ground with FTIR Spectroscopy*. Verein Deutscher Ingenieure.

Bosenberg J, Brassington D, and Simon PC (1997) Instrument development for atmospheric research and monitoring: Lidar profiling, DOAS and tunable diode laser spectroscopy. In: *Transport and Chemical Transformation of Pollutants in the Troposphere*, vol. 8. Berlin: Springer.

## Range-Resolved Remote Sensing

**M J T Milton**, National Physical Laboratory, Teddington, UK

© 2005, Elsevier Ltd. All Rights Reserved.

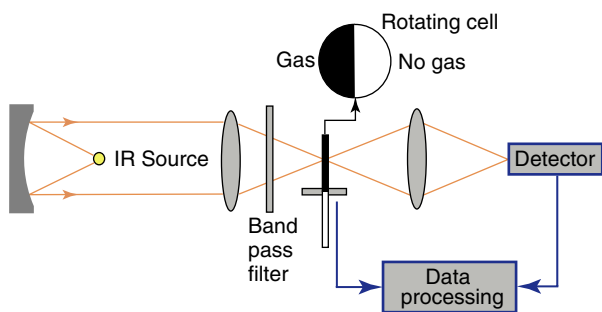
## Introduction

Range-resolved remote gas sensing is the process of simultaneously determining the concentration and distribution of a target species at locations remote from the measurement system using a single measurement. The range-resolved remote gas sensing methods that have been demonstrated successfully use pulses of optical radiation that are selectively absorbed or scattered by the target species. The range information is recovered by timing the propagation of the pulse to the absorbing, or scattering, event.

Another feature of these techniques, which is implicit in their ability to provide range information, is that they can operate from a location that is remote from the point being measured. They are distinguished from integrated-path remote gas sensing (q.v.) techniques, which are capable of performing remote measurements but are not capable of recovering information about the distribution of the target species along the measurement path.

Range-resolved gas sensing techniques have many advantages over the more widely used point-sensing methods. They are capable of scanning many locations and acquiring data that would require many individual point sensors. The addition of range information to concentration measurements enables maps of gas concentration to be compiled. The flexibility of range-resolved remote gas sensing has





**Figure 4** Principal components of a nondispersive gas correlation monitor.

Although they can be set up to monitor several gases by splitting the radiation into several different paths, they are generally only used to monitor one target species without modification of the correlation element.

NDIR are much used in the laboratory and in industry over short gas paths for gas concentration measurement. Recently, they have been applied to the longer paths required for 'stack gas' monitoring and the much longer paths (hundreds of meters) required for ambient air monitoring. Measurements of gases such as methane, ethane, hydrochloric acid, carbon monoxide, and carbon dioxide have also been demonstrated at industrial sites and roadsides.

## Passive Methods

All of the methods discussed above are described as 'active' because they make use of an 'active' source of radiation as part of the system. It is also possible

to measure gases remotely using passive methods that detect radiation emitted thermally by the target gas itself or by other objects within the field of view. Examples of this type of system are sensors used with a small number of spectral channels mounted on satellites and used to carry out qualitative identification of features on the earth's surface, such as vegetation or mineral species. Hyper-spectral imagers with a large number of spectral channels are capable of better qualitative measurements of the composition of objects in the field of view or along the measurement path. However, the derivation of true quantitative results from such instruments is problematic because of the difficulty of obtaining accurate information about the emissivity of the objects and their temperature. Nevertheless, this type of approach holds much promise for future development.

See also: **Remote Gas Sensing: Overview**; **Range-Resolved Remote Sensing**.

## Further Reading

VDI 4211 (2000) *Remote Sensing: Atmospheric Measurements near Ground with FTIR Spectroscopy*. Verein Deutscher Ingenieure.

Bosenberg J, Brassington D, and Simon PC (1997) Instrument development for atmospheric research and monitoring: Lidar profiling, DOAS and tunable diode laser spectroscopy. In: *Transport and Chemical Transformation of Pollutants in the Troposphere*, vol. 8. Berlin: Springer.

## Range-Resolved Remote Sensing

**M J T Milton**, National Physical Laboratory, Teddington, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Range-resolved remote gas sensing is the process of simultaneously determining the concentration and distribution of a target species at locations remote from the measurement system using a single measurement. The range-resolved remote gas sensing methods that have been demonstrated successfully use pulses of optical radiation that are selectively absorbed or scattered by the target species. The range information is recovered by timing the propagation of the pulse to the absorbing, or scattering, event.

Another feature of these techniques, which is implicit in their ability to provide range information, is that they can operate from a location that is remote from the point being measured. They are distinguished from integrated-path remote gas sensing (q.v.) techniques, which are capable of performing remote measurements but are not capable of recovering information about the distribution of the target species along the measurement path.

Range-resolved gas sensing techniques have many advantages over the more widely used point-sensing methods. They are capable of scanning many locations and acquiring data that would require many individual point sensors. The addition of range information to concentration measurements enables maps of gas concentration to be compiled. The flexibility of range-resolved remote gas sensing has



provided a way to study complex atmospheric transport processes and to locate and quantify industrial emissions.

## The LIDAR Principle

The methods that have been successfully used for range-resolved remote gas sensing all use the principle of light detection and ranging (LIDAR) to achieve range resolution. The LIDAR principle is the optical analog of radar (radio detection and ranging). It depends on the accurate timing of a pulse of radiation from a transmitter to a scattering event and back. The range is then calculated by converting the time of flight to a distance using a value for the speed of light. A schematic diagram of the major components in a LIDAR system is shown in Figure 1. The transmitted and received beams are usually directed into the atmosphere with a steering mirror in order to alter the direction in which the system measures.

The energy recorded at a detector collecting light from a telescope of area  $a$  from the transmission of a pulse of energy  $E$  is given by

$$P(r) = E\eta \frac{a}{r^2} B(r) \exp\left(-2 \int_0^r [A(r') + \alpha C(r')] dr'\right)$$

where  $B(r)$  and  $A(r)$  are the backscatter and absorption coefficients of the atmosphere, respectively, and  $C(r)$  is the concentration of the target species of absorption coefficient  $\alpha$ .

The difference between the various range-resolved remote gas sensing methods is in the nature of the processes used to return the light to the receiver. There are three processes that have been demonstrated

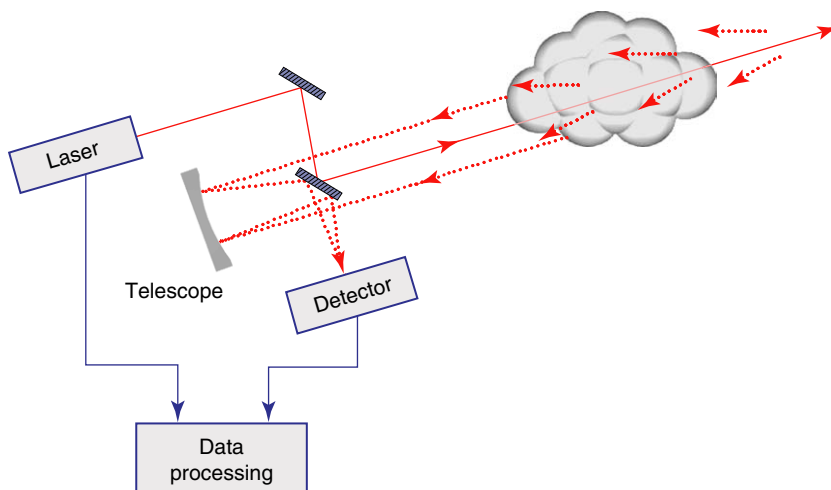
thoroughly: fluorescence, Raman scattering, and differential absorption combined with elastic (either Mie or Rayleigh) scattering.

Fluorescence has been used effectively in specialized applications (e.g., the measurement of the hydroxyl ion at short distances from airborne platforms) but it has not been applied widely to analytical measurement problems. Similarly, Raman scattering has been used in particular applications (e.g., the measurement of water vapor at high altitudes and the measurement of high concentrations of methane close to ground level). The cross-sections for fluorescence and Raman scattering are very low compared with elastic scattering, so they have only been used when there are specific reasons why the differential absorption and scattering method is not applicable. Additionally, they require either some form of spectral dispersion (in the case of Raman) or time gating (in the case of fluorescence) in order to distinguish them from the elastic scattered return.

The remainder of this article is devoted to the differential absorption LIDAR or DIAL method that is the most widely used technique for range-resolved remote gas sensing. It relies on elastic scattering from small particles and molecules that are always present in the atmosphere to return the signal. Speciation is achieved using the principle of differential absorption as described below.

## Differential Absorption LIDAR

The DIAL method makes use of two LIDAR measurements performed within a very short time interval at two different wavelengths. One is chosen to be at a wavelength that is absorbed by the target gas. The



**Figure 1** Schematic diagram of a differential absorption LIDAR (DIAL) system.

other is at a wavelength that is not absorbed by the gas of interest and forms a reference channel that measures the 'background' properties of the atmosphere.

The unknown concentration is then determined by taking the ratio of the returned intensity at the on-resonant wavelength ( $P_{\text{on}}$ ) to the off-resonant wavelength ( $P_{\text{off}}$ )

$$\log[P_{\text{on}}(r)/P_{\text{off}}(r)] = - \int_0^r (\alpha_{\text{on}} - \alpha_{\text{off}}) C(r') dr'$$

where  $\alpha_{\text{on}} - \alpha_{\text{off}}$  is the differential absorption coefficient between the on- and off-resonant wavelengths. If the absorption coefficient of the gas is known, then the concentration can be determined by calculating the ratio of the returned signals from the two measurements. Since both the measurements are time-, and hence range-resolved, it is possible to identify the distance to the point at which the concentration was measured. Performing this calculation at all ranges along the signal yields the concentration as a function of distance along the path of the measurement beams.

It is often stated that the principle of operation of LIDAR leads to a range resolution for the measured profile equal to half the length of the pulse. Although this is true for systems with long pulses and high-speed detectors, it is generally not true for systems using laser pulses with durations less than 50 ns. In typical examples of the DIAL technique, usable results can be achieved with range resolutions of several tens of meters achievable at ranges of many kilometers.

The species that can be measured by DIAL are most easily categorized according to the wavelength region of the radiation that is used to measure them. Examples of the species that can be measured by DIAL using presently available technology are shown in Table 1. Each of the wavelength regions listed in Table 1 requires a different type of laser transmitter and receiver system. These differences are sufficiently large that DIAL systems are usually designed to operate in just one wavelength region. Consequently,

each of these regions has become associated with particular choices of technology and applications.

## Applications of DIAL in the Ultraviolet and Visible Spectral Regions

High-quality components required for a DIAL system in the 220–900 nm spectral region are more readily available than those needed in other spectral regions. Excimer and solid-state lasers have sufficient pulse energies (in excess of 500 mJ), pulse lengths (between 5 and 50 ns), and repetition rates (between 5 and 100 Hz) for use as primary sources of laser energy. They can be used to pump tunable dye lasers or tunable solid-state lasers that can be used in conjunction with nonlinear optical techniques to generate the correct wavelengths in the ultraviolet and visible. Detectors such as photomultipliers and avalanche photodiodes which are capable of single-photon detection are available. The availability of suitable laser and detector technologies together with the strong backscattered signals from Rayleigh scattering led to this being the first region in which effective DIAL applications were demonstrated. The most important applications of DIAL in the ultraviolet and visible spectral regions include:

- studies of the movement of ozone within the stratosphere and troposphere;
- the monitoring of the dispersion of sulfur dioxide and nitrogen dioxide from industrial stacks;
- measurements of fugitive emissions of benzene, toluene, and xylene from chemical and petrochemical plants; and
- profiles of water vapor in the troposphere and lower stratosphere from ground-based and airborne platforms.

## Applications of DIAL in the Near-Infrared

The near-infrared between 2 and 5  $\mu\text{m}$  is perhaps the most technically challenging spectral region in which to perform measurements by DIAL. This is in part because of the weak and variable Mie scattering from particles at these wavelengths. Additionally, there are no simple primary laser sources available so the techniques of nonlinear optics must be used in order to generate suitable narrowband radiation. These include difference frequency mixing between fixed and tunable sources in the visible spectral region and the use of tunable optical parametric oscillators. These technical difficulties are well worth overcoming

**Table 1** Gases that can be measured using the DIAL technique

Wavelength range	Species measurable
220–900 nm	SO <sub>2</sub> , O <sub>3</sub> , NO, NO <sub>2</sub> , Hg, benzene, toluene, xylene, H <sub>2</sub> O
2–5 $\mu\text{m}$	Methane, alkanes, alkenes, alkynes, CO, HCl
9–11 $\mu\text{m}$	Chlorinated hydrocarbons, SF <sub>6</sub>

because of the wide range of gases that can be measured (see Table 1. Applications include:

- the identification of fugitive emissions of hydrocarbon vapors in chemical plants and
- detection of the build-up of flammable gases in industrial areas.

### Applications of DIAL in the Mid-Infrared

The use of the DIAL method in the mid-infrared spectral region has arisen from the availability of the high-powered carbon dioxide laser and the wide range of species with absorption lines in the range 9–11  $\mu\text{m}$ . However, the carbon dioxide laser cannot be tuned continuously across the spectrum so it is not possible to use a single carbon dioxide laser to measure all the species that have absorption lines in this region of the spectrum. Possible applications include:

- monitoring for the build-up of toxic gases generated as industrial by-products and
- ‘early warning’ of the presence of chemical agents used in weapons.

### Range, Range Resolution and Accuracy

When all the parameters of a DIAL system are known, it is possible to perform calculations of the range, range resolution, and sensitivity in the measurement of a specific target gas. However, such calculations are complex and the results are strongly

**Table 2** Performance of a DIAL system. The typical sensitivity applies at half the maximum range and the stated range resolution. The maximum ranges refer to analog detection. In some applications in the 220–900 nm region, they may be extended to 25 km using photon counting techniques

Wavelength range	Maximum range	Range resolution	Typical sensitivity
220–900 nm	1–3 km	10 m	1 ppb
2–5 $\mu\text{m}$	500 m to 1 km	15 m	100 ppb
9–11 $\mu\text{m}$	1–10 km	100 m	50 ppb

dependent on the exact measurement scenario and atmospheric conditions. A summary of typical performance in each spectral region is presented in Table 2.

See also: **Remote Gas Sensing:** Overview; Integrated-Path Remote Sensing.

### Further Reading

- Grant WB, Kagann RH, and McClenny WA (1992) Optical remote measurement of toxic gases. *Journal of Air and Waste Management Association* 42: 18–30.
- Measures RM (ed.) (1984) *Laser Remote Sensing*. New York: Wiley.
- Measures RM (ed.) (1988) *Laser Remote Chemical Analysis*. New York: Wiley.
- Milton MJT (1995) Tunable lasers for DIAL applications. *The Review of Laser Engineering* 23: 93–96.
- Vaughan G (2002) Remote sensing of atmospheric chemistry – lidar. In: Holton JR, Pyle J, and Currie JA (eds.) *Encyclopaedia of Atmospheric Science*. New York: Academic Press.

## ROBOTICS

See **SAMPLE HANDLING:** Robotics

## RUTHERFORD BACKSCATTERING SPECTROMETRY

See **SURFACE ANALYSIS:** Ion Scattering

# S

## SAMPLE DISSOLUTION FOR ELEMENTAL ANALYSIS

### Contents

#### Dry Ashing

#### Oxygen Flask Combustion

#### Wet Digestion

#### Microwave Digestion

### Dry Ashing

**M Hoenig**, Veterinary and Agrochemical Research Centre, Tervuren, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Nowadays, trace element determinations are performed using very sensitive analytical techniques. However, with a lowering of detection limits, the risk of errors suddenly appearing due to sample handling is increased. Prior to commercial introduction of ultrasensitive instrumentation as inductively coupled plasma-mass spectrometry (ICP-MS), these 'new' errors were practically imperceptible to the determination of relatively high analyte concentrations that were measured with less responsive techniques. The danger of contamination is now increasingly present: the choice of sample preparation procedure, the quality of its application, and the need for an adequate laboratory environment have therefore become most critical points defining successful trace element determinations.

In most cases, preparation of solid samples involves several stages: drying, homogenization, and/or grinding, followed by mineralization and dissolution of a subsample. The solution so obtained is ultimately diluted to volume. Ideally, the organic fraction of the sample has been decomposed and completely eliminated during these preparation steps and only dissolved inorganic compounds constitute the dissolved residue to be analyzed.

Before the analysis, samples of organic or of a mixed nature are subject to two distinct steps, which often take place simultaneously: mineralization and dissolution. Samples of purely inorganic composition are simply dissolved. The composition of biological and environmental samples varies from purely inorganic to purely organic, but, generally, they are an intermediate combination of these extremes. This implies that the total dissolution of samples usually cannot be achieved in a single step using a single reagent. In practice, the necessary number of steps and reagents is dictated by the matrix composition. Purely organic or mixed samples are usually brought into solution by some type of oxidation process combined with an acid dissolution of the resulting residue, as well as of the initial inorganic part of the matrix.

Already in 1844, Fresenius and Von Babo had developed a method for the destruction of animal tissues prior to elemental analysis. In the intervening years, many procedures were described for this purpose. However, despite numerous possible variations, almost all of the methods fall into one of two main classes, i.e., dry ashing and wet digestion. Dry ashing methods are especially appropriate for samples having high organic matter content. The first step of the method ensures the decomposition of organic matter by heating the sample to a relatively high temperature, with atmospheric oxygen serving as the oxidation agent. Chemical compounds (the so-called ashing aids) may sometimes be added to help this process. The second step of a dry ashing method is the subsequent solubilization of the resultant ash using an appropriate acid or a mixture of acids.

Depending on the sample type, the dissolution procedure generally involves several steps. Here, the terminology is precise: the term 'mineralization'

relates to samples having a totally or partly organic matrix only (animal and plant tissues, food samples, soils, etc.). Prior to the analysis, any organic compounds present must be decomposed and/or completely eliminated by the mineralization procedure. Using various reagents, the organic matter is decomposed into carbon dioxide, nitrogen oxides, and water, thus liberating into solution all elements initially associated with it. After the mineralization procedure, the resulting sample residue should be essentially inorganic: it will be subject to a final dissolution step similar to that used for a sample having an initially total inorganic composition (rocks, metals, etc.). For more complex samples (organic plus inorganic composition: soils, sludge, plant samples, etc.), chemical reagents and physical means are most often used to ensure these two roles (mineralization and dissolution) are simultaneously achieved.

The objective of the sample preparation stage is usually to bring all available means into play in order to determine as readily as possible the elements of interest. First, these means have to ensure the transformation and simplification of the matrix (mineralization: wet digestion, dry ashing). Second, they should convert the sample to a form compatible with the measurement technique utilized (generally a dissolution).

Dry ashing methods compete with the following techniques for sample dissolution of organic and biological species for elemental analysis:

- wet ashing in open or closed vessels,
- oxygen flask techniques,
- combustion tube techniques, and
- microwave digestion techniques in open or sealed vessels.

## General Principles

After the appearance, at the end of the 1970s, of commercial advertisements praising the universality and absolute necessity of wet digestion microwave heating devices for trace element analysis, several scientific papers have radically condemned dry ashing procedures, despite their long record of usefulness. In contrast, many respected institutions as well as numerous laboratories carry on the use of classical dry ashing in practical analyses of a number of materials of biological origin.

Our own extensive experience in the field of sample preparation has shown that, better than the other known mineralization procedures, dry ashing methods ensure the quantitative decomposition and elimination of organic matter. Usually, these procedures

are performed by calcination at atmospheric pressure in programmable muffle furnaces. The commonly utilized temperature for this step is  $\sim 450^{\circ}\text{C}$ . In addition to conventionally heated muffle furnaces generally employed for dry ashing purposes, the market now also provides microwave furnaces especially adapted to attain elevated temperatures. The unique advantage of the latter is the capacity to ensure application of very fast heating ramps. However, this property is not directly interesting for usual dry ashing procedures, where precisely slow heating ramps are needed. Additionally, a low-temperature ashing (LTA) procedure in electronically excited oxygen plasma exists, very desirable for sample preparation when volatile elements are to be determined. The instrumentation is, unfortunately, very expensive and not readily available at present. In addition, LTA is a particularly time-consuming procedure.

In the usual high-temperature ashing, fresh or dried (generally  $103\text{--}105^{\circ}\text{C}$ ) samples are weighed into suitable ashing vessels (vitreous silica, porcelain, platinum) and placed in the furnace. The temperature is then progressively elevated, following a convenient heating program, to attain  $450^{\circ}\text{C}$ , and then maintained for several hours. The resulting inorganic residue (ash) is dissolved using an appropriate acid. The solution is diluted to a known volume and analyzed. Depending on the initial sample condition, results are expressed based on a fresh- or dry-weight basis. The application of dry ashing methods is simple and large series of samples may be treated at the same time. This is not their unique advantage – compared with wet digestions, dry ashing procedures present several other interesting characteristics:

1. Possibility of treating large sample amounts and dissolving the resulting ash in a small volume of acid. This permits preconcentration of trace elements in the final solution, which is useful when very low analyte concentrations are to be determined. Such an advantage is not realizable with wet digestion methods. Additionally, heterogeneity is a typical property of many biological materials. The possibility of processing larger masses of sample, which, upon mineralization, provides a homogeneous solution, helps to minimize subsampling errors.
2. The resulting ash is completely free of organic matter. This is a prerequisite for ensuring accuracy with some analytical techniques (e.g., ICP-MS or electrochemical methods) wherein analyte response may be influenced by the presence of residual carbon or some undigested organic molecules. The resulting solutions are of very acceptable aspect

(clear, colorless, and odorless), rarely the case when wet digestion methods are used and residual carbon content often attains elevated values. This is systematically observed not only for both conventionally heated and open microwave digestions, but also for closed microwave-assisted digestions. With high-pressure bombs, the residual carbon content may be lower but it is never quantitatively eliminated.

3. Reagent volumes and their handling are reduced compared to wet digestion methods.

4. The acidity of final solutions can efficiently be controlled: the acid is added directly to the ash and only a small fraction is consummated during its dissolution. With wet digestion procedures, added acids must also ensure the destruction of the organic matter and their effective amounts utilized during these chemical reactions vary quite significantly. This results in unknown acid concentrations in the final solutions to be analyzed. This fact is at variance with well-known requirements for all methods based on atomic spectroscopy concerning the need for similarity of acidities between standards and samples. In some situations, this similarity is absolutely obligatory, e.g., for the determination of nickel by ICP-MS. The commonly used nickel cones in the ICP-MS interface usually produce relatively high nickel backgrounds due to their finite dissolution by the aerosol being introduced. With variable acid concentrations, the background can vary significantly from one sample to another, resulting in erratic nickel results. Such unfavorable conditions are avoided using dry ashing methods that ensure a practically constant acid concentration from sample to sample, allowing more consistent ICP-MS determinations of nickel to be made.

Despite these several advantages, one must also accept several drawbacks to dry ashing procedures:

1. The chemistry of charring process is very complex and the actual temperature in the sample remains unknown; in some cases, it may be several hundred degrees above that of the furnace. This may result in volatilization losses and then to poor recoveries of some elements.

2. Under particular experimental conditions and for some types of samples, retention phenomena can occur, resulting also in incomplete recoveries.

3. Dangers of contaminations by the ambient air are greater than in wet digestion procedures performed in closed systems.

### Chemistry of the Method

Results of studies concerning the dynamics and chemistry of organic matter degradation revealed

the strong exothermic character of decomposition during charring but also a possible variability, depending on the specific type of biological material. The former findings call for sufficient moderation of the charring step (ramp heating) in order to prevent local overheating of the sample and subsequent risk of loss of a fraction of the analyte due to its mechanical removal in the form of solid particles of aerosol (smoke). Only under such controlled conditions can classical dry ashing have the potential to yield accurate results.

The term 'dry oxidation' usually characterizes procedures wherein organic matter is oxidized by reaction with gaseous oxygen, generally with the supply of energy in some form. Included in this general term are methods in which the sample is heated to a relatively high temperature in an open vessel (conventional dry ashing), or in a stream of oxygen or air. In addition, related low temperature techniques employing excited oxygen, bomb methods using oxygen under pressure, and the classical oxygen flask technique in which the sample is ignited in a closed system must also be included. All these methods involve two main processes: they provoke evaporation of the moisture and of volatile materials and ensure the progressive oxidation of the nonvolatile residue until all organic matter is destroyed.

Although these processes occur in all dry oxidations, it is not always possible to distinguish them as separate events. They are probably most easily separated in the conventional ashing procedure in which the organic material is heated in an open vessel with free access to air or oxygen. In usual analytical practice, the first steps of such a procedure are usually conducted at a temperature much lower than that used to complete the oxidation. This is largely to prevent the ignition of the volatile and inflammable material produced by the process of destructive distillation and partial oxidation, as this would lead to an uncontrolled rise in the temperature resulting in an increased danger of analyte losses.

### Volatility Losses

The long experience has repeatedly shown that the most severe element losses are systematically observed during the heating ramp, not as a consequence of too high final ashing temperature. An inadequate heating ramp may provoke the autoignition of the sample and the resulting rapid temperature increase results in volatilization losses. The most often utilized means of avoiding ignition problems is direct insertion of the sample into the cold muffle furnace, followed by heating with an appropriately slow ramp.



Before the last stage of the process – progressive oxidation of the nonvolatile residue – the material remaining after the preliminary treatment is a more or less porous mass of charred organic matter containing variable amounts of inorganic material distributed throughout it. In reality, this picture is highly variable and it will depend not only on the type and the composition of samples analyzed, but also on the action of possible reagents added that can change the initial chemistry of the process. Consequently, the kinetics of oxidation of such material will be dependant on the nature of the material itself, the inorganic substances it contains, and its particle size and porosity. Tentative findings derived from such reactions with pure carbon or graphite can only be applied with caution to the complex chars existing in dry ashing of real samples. The temperatures generally recommended for dry ashing are apparently low compared with those reported for the oxidation of graphite, but the chars produced are probably far more reactive due to unknown catalytic effects of the inorganic constituents present.

In dry ashing procedures used for the analysis of environmental or biological samples (animal and plant tissues, food samples, blood, milk, etc.), the final temperature is maintained for several hours. If the oxidation is achieved under optimal conditions, it leads to white or light gray colored ashes, easily soluble in acids. Sometimes, depending on the sample type, the oxidation of organic matter is not completely achieved; in this case, the ash exhibits darker spots (dark gray to black) attributable to insufficiently oxidized carbon. Because this phenomenon is always responsible for a difficult subsequent dissolution (often resulting in incomplete recoveries for several elements), such a residue must be re-treated using a few drops of nitric acid and briefly recalcined at the usual ashing temperature. After this treatment, ashes generally become clear and easily soluble.

During the oxidation process, the analyte(s) will behave in one or more of a number of ways. Ideally, they will quantitatively remain in the residue (ash) arising from the oxidation, and in a form in which they can be readily recovered, generally by a simple dissolution of the ash in an appropriate acid. Fortunately, for the usefulness of the method, this is the case for most analytes and samples. In some cases, a part (or the total) of the analyte may be converted to a volatile form that may escape from the vessel (i.e., volatilization losses) or may be combined with the vessel surface or with some components of the inorganic residue remaining after oxidation (i.e., retention losses). In practical trace element analysis, the most often reported volatilization losses pertain mainly to mercury, arsenic, and selenium.

## **Retention Losses**

Retention losses result in poor recoveries of one or more analytes using the normal procedure for solubilization of the ash. They are generally observed for a particular quality of ashing vessel or in the presence of silicates or other insoluble compounds in the sample matrix. During intercomparison studies involving analyses of plant matrices, significant discrepancies amongst results are often observed between laboratories using simple mineralization procedures and those that apply procedures that include a hydrofluoric acid attack followed by evaporation to dryness. In the former, the values obtained are systematically lower because complete digestion is not achieved. In the often utilized wet digestion procedures, mixtures of various acids with hydrogen peroxide may also lead to poor recoveries due to the presence of silicate compounds in the sample or to (co)precipitation phenomena. As a consequence, Al, Fe, Cu, and Mn, in particular, are not completely recovered, depending upon the specific plant matrix, probably related to the binding of analytes with the insoluble residue. Several authors have noted similar problems with agricultural matrices such as composts, animal meats, or brewers yeast.

These statements, among others, are supported by studies concerning retention losses occurring during dry ashing of plant samples. For several plants, analyses of insoluble residues have revealed that a significant fraction of elements (major, minor, trace) is retained, depending on the type of sample and the element studied. The most affected element is Al (sometimes up to 95% retained). Consequently, it is useful to use the recovery of aluminum as a marker for the procedure quality; if the Al recovery is incomplete, it may be concluded that the dissolution step was not performed under good conditions and that many other elements may be affected in a similar fashion.

For an efficient dissolution of the ash of samples with silicate compounds in the matrix, such problems highlight the absolute necessity of utilizing an HF step followed by evaporation to dryness if the objective is the determination of total element content. However, this problem, typically associated with plant samples, is similar when applying a wet digestion procedure: if an insoluble residue remains, an additional HF step, followed by evaporation to dryness, must also be performed.

Another example highlighting this problem is associated to ICP-MS interferences from residual silicon resulted in up to 30% positive bias in intensities from  $^{63}\text{Cu}$ ,  $^{65}\text{Cu}$ , and  $^{55}\text{Mn}$  derived from soil and sewage sludge digests, due to spectral interferences

from  $^{28}\text{Si}^{27}\text{Al}^+$ ,  $^{28}\text{Si}^{35}\text{Cl}^+$ , and  $^{28}\text{Si}^{37}\text{Cl}^+$ . In such a case, the interfering Si will be easily removed by the above-mentioned HF step.

Such a procedure is not always easy to achieve with most commercially available microwave heating (closed) devices and is, in any case, much more difficult to apply than with a dry ashing procedure. Finally, plants are often considered as purely organic samples with some trace elements present. It is clear that the aforementioned problem is comparable to that encountered in soil, sediment, sludge, and rock samples where silicon is typically the primary matrix element. In this case, however, all analysts are aware of the absolute necessity of dissolving the entire sample if the total analyte content has to be determined.

An additional retention problem encountered with dry ashing procedures is that posed by the sequestering action of some materials produced during ignition. The binding of iron by condensed phosphates produced by the action of heat on simple phosphates, or retention of several elements on silicate compounds present in the sample are the best known examples. Nevertheless, many studies in this field often present contradictory observations, illustrating clearly the complexity of the retention problems and the need for an adequate dissolution step. On the other hand, with the exception of arsenic and selenium, these considerations indicate that several losses reported as being due to volatilization were, in reality, due to retention problems.

## Methodology

Dry ashing methods can be applied to mineralization of organic materials, biological tissues, and liquids, plant, and foodstuffs, sludge, etc. Well mastered, they ensure total destruction of the organic matter; the associated elements are generally transformed to carbonate or oxide forms. At present, they are generally performed using fully programmable (ramp and holding times/temperatures) muffle furnaces characterized by an efficient temperature control and reproducible thermal programs. Required intermediate evaporations to dryness are usually achieved on sand baths or on hot plates.

It is mandatory to select an ashing temperature that ensures the quantitative decomposition of organic matter without partial or total loss of analytes by volatilization or by their incorporation into a residue, which is insoluble in usual reagents. The latter may result from formation of refractory oxides, from combinations with other sample constituents present, as well as from reactions with the walls of the crucible. As noted earlier, one

of the causes of losses during dry ashing procedures is the reaction of the analyte with some of the solid matter present in the system. In order for a reaction of this nature to constitute a problem, it is first necessary that it occurs to a significant extent and, second, that the product of the reaction be insoluble in the reagents generally used for dissolving the resultant ash. The solid matter available for such a reaction is generally the material of the ashing vessel and the residue from the sample itself. It is obvious then that their nature will have a considerable effect on the extent of the losses.

The most common used ashing vessels are made of silica, porcelain, or platinum. Vitreous silica is a glass consisting almost entirely of  $\text{SiO}_2$ , with some Na, Al, Fe, Mg, and Ti oxide impurities, whereas the glaze on porcelain ware is a more complex material containing Al, K, Ca, and Na oxides (up to 30%) in addition to silica. For both vitreous silica and porcelain, the obvious reaction is between the oxide of the analyte and the ashing vessel to produce a complex silicate, resulting in a loss. Studies with radio-tracers have shown that retention of metals by reaction vessels made of vitreous silica may be very significant during dry ashing. This type of reaction clearly occurs but is dependent on many factors. Some oxides react much more readily than others and, even if silicates are formed, some will be stable to subsequent acid attack while others will readily be decomposed and so cannot be considered to cause losses. These reactions will, of course, be exacerbated if the ashing vessel is made of silica or porcelain, exhibits a marked weakening of the silicate structure, or a surface worn by extensive use.

Because the extent of such reactions remains unknown, the alternative practice of using essentially inert platinum crucibles is much more reliable. This metal is virtually unaffected by any of the usual acids, including hydrofluoric, but cannot be used for aqua regia digestion procedures. Of course, the initial cost for platinum is significantly higher than for other types of ashing vessel, but its lifetime is practically unlimited.

## Ashing Aids

In its initial form, a dry ashing procedure cannot be considered appropriate for preparation of samples to be used for the determination of As and Se. However, many of the reported ashing methods describe the addition of some extra inorganic compounds to the sample to improve the efficiency of the procedure. These added materials are generally called ashing aids, and they serve one or both of two purposes, i.e.,

to facilitate the decomposition of the organic matter, or to improve the recovery of the element to be determined.

The most common aid, used to purely hasten the oxidation of organic material, is nitric acid. It is generally added toward the end of the ashing process to decompose small amounts of remaining carbonaceous material. Since the ash from most biological materials contains up to several tens of percent carbonates, nitrates are formed *in situ* after the addition of nitric acid. Additional ashing is then, in fact, melting with nitrates and should help to remove the most resistant degradation products present in the organic matrix. This step, leading to the production of a clean ash, has to be performed with care because when appreciable amounts of organic material are still present, it can cause the ignition of the residue when it is returned to the furnace, resulting in a possible loss of material.

Some substances serve as auxiliary oxidants as well as serving other purposes. These are commonly the nitrates of light metals such as magnesium, calcium, or aluminum, which decompose on heating to yield oxides of nitrogen. These auxiliary oxidants also fulfill the important function of being inactive diluents in the process. As the organic matter in a sample is progressively decomposed, the analytes are brought into closer contact with the material of the vessel and other constituents of the residue. If a reaction with them is feasible, then the increased proximity will increase the chance of its occurrence. Under these circumstances, dilution of the ash with an inert material, such as magnesium oxide, should greatly reduce the possibility of undesirable solid-state reactions, resulting in improvement of recoveries. The well-known utilization of relatively unstable magnesium nitrate as an ashing aid likely offers both the advantage of more rapid oxidation and of decreased retention losses.

These oxidative-dilution agents improve recoveries without entering into any reaction with the sample itself. Another group of ashing aids achieves the same end by altering the chemical nature of some of the constituents. The best example of this is the use of sulfuric acid to convert volatile chlorides to nonvolatile sulfates; this may prevent losses of Cd, Pb, or Cu up to  $\sim 750^{\circ}\text{C}$ .

Arsenic and selenium determinations can, in some cases and under particular conditions, also benefit from the advantages offered by a dry ashing procedure. The addition of ashing aids – generally MgO and/or  $\text{MgNO}_3$  – can give rise to less volatile As or Se compounds during the ashing procedure. The successful use of ashing aids is, of

course, strongly dependent on the initial form of the analyte.

In any case, utilization of ashing aids is a particularly delicate step because some successful examples cannot lead to generalizations: for routine use, the procedure necessitates a serious and time-consuming validation for each type of sample analyzed. In addition, the utilization of ashing aids significantly increases the total dissolved solids content of solutions and enhances the dangers of contamination, limiting strongly the use of this approach for trace element analysis.

**See also:** Sample Dissolution for Elemental Analysis: Oxygen Flask Combustion; Wet Digestion; Microwave Digestion.

## Further Reading

- Anderson R (1987) *Sample Pretreatment and Separation*, ACOL series. Chichester: Wiley.
- Bock R (1979) *A Handbook of Decomposition Methods in Analytical Chemistry*. Glasgow: International Textbook Company.
- Dunneman L, Begerow J, and Bucholski A (1994) Sample preparation for trace analysis. In: *Ullmann's Encyclopedia of Industrial Chemistry*. Weinheim: Verlag Chemie.
- Gorsuch TT (1970) *Destruction of Organic Matter*. Oxford: Pergamon.
- Hoenig M, Baeten H, Vanhentenrijk S, Vassileva E, and Quevauviller Ph (1998) Critical discussion on the need for an efficient mineralization procedure for the analysis of plant material by atomic spectrometric methods. *Analytica Chimica Acta* 358: 85–94.
- Hoenig M (2001) Preparation steps in environmental trace element analysis – facts and traps. *Talanta* 54: 1021–1038.
- Hoenig M (2003) Dry ashing. In: Mester Z and Sturgeon R (eds.) *Sample Preparation for Trace Element Analysis*. Comprehensive Analytical Chemistry 41. Amsterdam: Elsevier.
- Kinoshita K (1988) *Carbon – Electrochemical and Physicochemical Properties*. New York: Wiley.
- Mader P, Száková J, and Curdová E (1998) Combination of classical dry ashing with stripping voltammetry in trace element analysis of biological materials: Review of literature published after 1978. *Talanta* 43: 521–534.
- Sulcek Z and Povondra P (1989) *Methods of Decomposition in Inorganic Analysis*. Boca Raton, FL: CRC Press.
- Ure AM, Butler LRP, Scott RO, and Jenkins R (1997) Preparation of materials for analytical atomic spectroscopy and other related techniques In: *Nomenclature, Symbol, Units and Their Usage in Spectrochemical Analysis, Part X. Spectrochimica Acta B* 52: 409–420.

## Oxygen Flask Combustion

**Y S Fung**, Hong Kong University, Hong Kong, People's Republic of China

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Oxidation of organic substances by means of combustion in an oxygen-rich atmosphere is one of the most popular techniques for sample decomposition for elemental analysis. During combustion, heteroatoms in the organic substances are converted into inorganic forms that can be readily analyzed subsequently using various techniques employed in elemental analysis. When a sample is oxidized in a closed system, loss of volatile compounds during combustion and contamination from the laboratory atmosphere can be avoided. Thus, a closed flask is normally used during combustion to reduce contamination.

The oxygen flask is the most simple closed combustion apparatus which is made of Pyrex (borosilicate) glass, it facilitates rapid elemental analysis in laboratories.

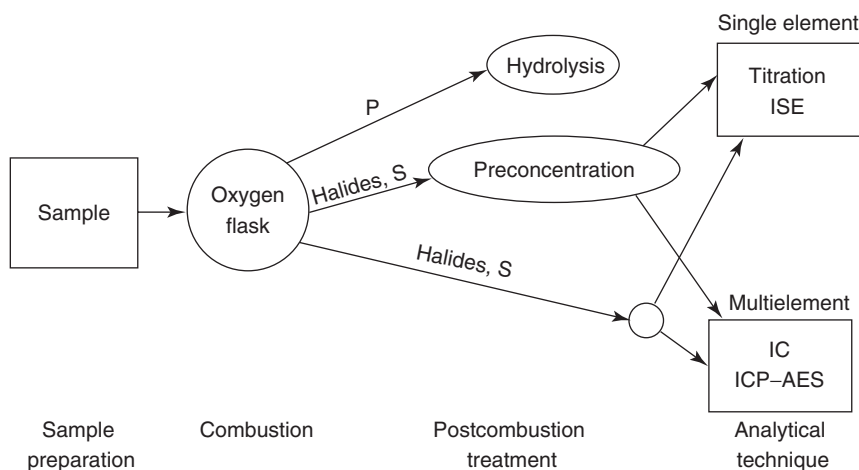
The oxygen flask method (also known as the Schöniger flask combustion method) has been designed for combustion of organic samples in the milligram region (normally less than 20 mg) prior to elemental analysis. The sample is wrapped in a piece of paper, placed in a platinum basket, and burned in a sealed flask filled with pure oxygen under 1–2 atm pressure. The flask is fitted with a ground glass

stopper and inverted. The absorption solution provides a liquid seal. Thereafter the sample is ignited either through an electrical discharge or infrared irradiation. This method is commonly used for elemental analysis of biological and organic materials. Compared with other methods for mineralization of organic matrices, the oxygen flask method is rapid, simple, and economic and uses glass apparatus readily available in most analytical laboratories.

The present article will cover the development of the methodology since 1995, with its focus on the direction taken by development, analytical techniques coupled with oxygen flask combustion, new areas of application, and future areas of development.

### Basic Procedures

The basic procedure using the oxygen flask method for elemental analysis is shown in the flowchart given in **Figure 1**. The procedures can be divided into three types. The first type includes wet chemical procedures used for sample preparation before combustion and after combustion treatments. The second type consists of hyphenated analytical techniques developed for elemental analysis using oxygen flask combustion. The third and the most important type is the oxidation of organic samples carried out inside a sealed oxygen flask under pure oxygen. A high recovery and a good absorption of oxidized products are needed in this step to ensure complete conversion of analytes into detectable inorganic forms. The basic oxygen flask combustion has not changed much since



**Figure 1** Flowchart showing the basic procedures using the oxygen flask method for elemental analysis. IC, ion chromatography; ICP-AES, inductively coupled plasma-atomic emission spectrometry; ISE, ion selective electrodes.

1995, and there has been a major change only in sample preparation before combustion and in post-combustion treatments; new hyphenated analytical techniques have also been developed for elemental analysis using oxygen flask combustion. These developments will be discussed in the following sections.

## **Sample Preparation Procedures**

The sample preparation procedure is mainly used for two purposes. The first purpose is distinguish different forms either for nutritional purposes, such as the determination of free or bounded iodine in food samples, or for differential toxicity so as to assess different fluorotensides in ecotoxicological experiments by their differential adsorption at an activated carbon surface.

The second purpose is to extend the scope of the radiochemical neutron activation analysis (RNAA) method. Preirradiation of inorganic samples to induce fission products prior to oxygen flask ignition and solvent extraction can be used to distinguish ionic and nonionic form. The use of double irradiation in RNAA allows the determination of nuclides with short, medium, and long lives in a single experiment. This enables simultaneous determination of  $^{127}\text{I}$ ,  $^{128}\text{I}$ ,  $^{235}\text{U}$ ,  $^{133}\text{I}$ , and  $^{203}\text{Hg}$ .

## **Postcombustion Treatment Procedures**

Procedures that have undergone significant development in the period are the various postcombustion treatment procedures developed with an aim to improving the absorption of gaseous analytes, to facilitate chemical conversion into a suitable form for detection, to concentrate the analytes to detectable levels, and enable the use of a selected analytical method for quantitation of analytes in total or free elemental form. Two major procedures have been developed, optimization of the composition of the absorption solutions and the development of techniques to concentrate analytes prior to their determination using an analytical technique for elemental analysis.

### **Optimization of Absorption Solutions**

The selection of a suitable absorption solution with high recovery of analytes is a critical area of elemental analysis using the oxygen flask method. The addition of  $\text{H}_2\text{O}_2$  to enhance the absorption of sulfur compounds is dealt with in several papers. To assist absorption of sulfur dioxide, two absorption solution systems have been developed. The first system uses 3% pure  $\text{H}_2\text{O}_2$  and the second absorption solution is

made up by mixing 100  $\mu\text{l}$  of 1%  $\text{H}_2\text{O}_2$  with 20 ml of 2  $\text{mmol l}^{-1}$  NaOH prior to making up to 100 ml by deionized (DI) water. To enhance bromine and iodine adsorption, 20 ml of 2  $\text{mmol l}^{-1}$  NaOH and 100  $\mu\text{l}$  of 5%  $\text{NH}_2\text{NH}_2$  are mixed prior to making up to 100 ml absorption solution with deionized water. An alkaline solution of 1,3-dibromo-5,5-dimethylhydantoin (DBH) has been used to improve iodine determination. In addition to enhancing analyte absorption, other reagents needed in the detection method are added to the absorption solution for formation of a colored complex or compounds required in subsequent titration and spectrophotometry. Various new reagents and complexes have been synthesized to enable the use of a specific analytical technique.

### **Preconcentration of Analyte**

There are two methods that can be used to concentrate analytes prior to analysis using the chosen method. The first method is conversion of different ionic products into a single form to increase the recovery. One example is the postcombustion hydrolysis used for phosphorus determination. The procedure consists of boiling the oxidation products in dilute  $\text{H}_2\text{SO}_4$  at 200°C for 20 min to ensure complete conversion from various different forms of phosphorus anions such as  $\text{P}_2\text{O}_7^{4-}$  and  $\text{P}_3\text{O}_{10}^{5-}$  into orthophosphate ( $\text{PO}_4^{3-}$ ) before ion chromatography (IC) analysis to increase the recovery.

The second method uses preconcentration to increase the analyte concentration in the absorption solution to detectable levels prior to analysis. One example is the use of solid-phase extraction with an anion exchange column as a pretreatment technique to remove interference from the sample matrix and to enhance the detection sensitivity prior to determination of the concentrated iodide anion through reversed phase high-performance liquid chromatography (HPLC) using an ultraviolet (UV) detector. The results show more than 90% recovery and good reproducibility for the determination of trace iodine in iodine-enriched eggs.

A summary of the new procedures reported since 1995 using the oxygen flask method for elemental analysis is given in Table 1 with a focus on the areas of development in the basic procedure. It is obvious from the table that the amount of work on multi-elemental analysis has increased, though techniques for single element determination are still the major thrust. Details of the development of the oxygen flask combustion method for single element and multi-element determination will be given in the next section.

**Table 1** Summary of new procedures developed using the oxygen flask method for elemental analysis since 1995

Analyte	Sample	Sample preparation and postcombustion treatment	Analytical methods	Remarks
I	Organic compounds	A new reagent, DBH, has been used to improve the determination of iodine due to the favorable removal of excess of DBH 5-sulfosalicyclic acid prior to determination	Volumetric titration	
	Food	Alkaline absorption solution	IC	
	Eggs	Use of solid-phase extraction and anion exchange column for Preconcentration of I	HPLC	Total I
	Food colors	Mix 20 ml of 2 mmol l <sup>-1</sup> NaOH with 100 µl of 5% NH <sub>2</sub> NH <sub>2</sub> and make up to 100 ml by DI water as absorption solution	Direct injection into IC for free I <sup>-</sup> and after oxygen flask combustion for total I	Use of ceramic carbon column for total I <sup>-</sup> and free I <sup>-</sup>
F	<i>p</i> -Fluorobenzoic acid and dexamethasone	New metallofluorescent reagent (CNDA) made to assist fluorophotometric detection	Fluorophotometric titration	Total F
	Organic compounds	Titration with AlCl <sub>3</sub> solution	Potentiometric titration	Total F
	Organic pharmaceutical compounds	Fluoride-ISE compared with volumetric titration, with excess Ce(III) back-titrated with ethylenediaminetetraacetic acid (EDTA). Addition of TISAB prior to fluoride-ISE determination	Fluoride-ISE	Total F with good agreement between two methods
	Hair	Addition of TISAB	Fluoride-ISE	Determine total F for fluorosis of workers in polluted environment
	Sediment and plants	Adsorption of fluorotensides at active carbon, filtration, dry and oxygen flask combustion. Addition of TISAB prior to fluoride-ISE determination.	Fluoride-ISE	Determine total F for ecotoxicology experiments
S	Plant materials	H <sub>2</sub> O <sub>2</sub> as absorption solution	IC and ICP–AES	Total S
	Organic materials	Improved recovery using sealed tube (580°C for 20 min) and 3% H <sub>2</sub> O <sub>2</sub> absorption solution for samples of 1 mg or below	Spectrophotometric titration	Total S
Cl <sup>-</sup> , Br <sup>-</sup> , SO <sub>4</sub> <sup>2-</sup>	Colour additives	20 ml of 2 mmol l <sup>-1</sup> NaOH and 100 µl of 5% NH <sub>2</sub> NH <sub>2</sub> to make up to 100 ml for Br absorption 20 ml of 2 mmol l <sup>-1</sup> NaOH and 100 µl of 1% NaOH/H <sub>2</sub> O <sub>2</sub> to make up to 100 ml for S absorption	IC developed by mobile phase of 2 mmol l <sup>-1</sup> Na <sub>2</sub> CO <sub>3</sub> , 1 mmol l <sup>-1</sup> TEA in 5% acetonitrile	Ceramic carbon column

Continued



**Table 1** Continued

Analyte	Sample	Sample preparation and postcombustion treatment	Analytical methods	Remarks
F, Cl, Br, S	Organic compounds	Suppress background of NO <sub>x</sub> and CO <sub>2</sub> using standard organic compounds	IC	Microdetermination
P, Cl, Br, S	Organic compounds	Dilute aqueous H <sub>2</sub> O <sub>2</sub> as absorption solution use of postcombustion hydrolysis to increase recovery within 0.3% of theoretical values	Boil in dilute H <sub>2</sub> SO <sub>4</sub> at 200°C for 20 min to ensure conversion from P <sub>2</sub> O <sub>7</sub> <sup>4-</sup> and P <sub>3</sub> O <sub>10</sub> <sup>5-</sup> to PO <sub>4</sub> <sup>3-</sup> before IC separation	Microdetermination
F <sup>-</sup> , Cl <sup>-</sup> , NO <sub>2</sub> <sup>-</sup> , Br <sup>-</sup> , NO <sub>3</sub> <sup>-</sup> , SO <sub>4</sub> <sup>2-</sup> , HPO <sub>4</sub> <sup>2-</sup> , I <sup>-</sup>	Organic compounds	Alkaline absorption solution	IC developed by mobile phase of 4 mmol l <sup>-1</sup> Na <sub>2</sub> CO <sub>3</sub> , 1 mmol l <sup>-1</sup> TEA in 7.5% acetonitrile	Ceramic carbon column coated with DBS
F <sup>-</sup> , Cl <sup>-</sup> , NO <sub>2</sub> <sup>-</sup> , Br <sup>-</sup> , NO <sub>3</sub> <sup>-</sup> , HPO <sub>4</sub> <sup>2-</sup> , SeO <sub>3</sub> <sup>2-</sup> , SeO <sub>4</sub> <sup>2-</sup> , SO <sub>4</sub> <sup>2-</sup>	Organic compounds	Water as absorption solution	Suppressed IC with anion micromembrane suppressor and conductivity detector	Shodex IC IF-424 anion exchange column
Se	Blood plasma and hair	Formation of a Se ternary inclusion compound at β-CD cavity with anionic surfactant SDS leads to a 10-fold enhancement in fluorescence	Ternary inclusion compound–fluorescent spectrophotometry	Low correlation between hair Se and Plasma Se levels
W	Organic compounds	Titration with Pb(NO <sub>3</sub> ) <sub>2</sub> by Pb ISE after removal of CO <sub>2</sub> and adjusting buffer pH	Potentiometric titration	Recoveries of W from 99.74–100%
Ce(III), Th(IV), and U(VI)	Chelates	Ce(III) determined as yellow formaldoxime complex at 400 nm; Th(IV) determined as arsenate(III) complex at 660 nm; U(VI) determined as complex at 555 nm	Volumetric titration and spectrophotometric titration	Microdetermination
<sup>127</sup> I, <sup>128</sup> I, <sup>235</sup> U <sup>133</sup> I and <sup>203</sup> Hg	Sediments and soil	Preirradiation to induce fission products prior to ignition of oxygen flask and solvent extraction	RNAA	Double irradiation enables analysis of short- medium-, and long-life nuclides

CNDA, 3-carboxy-2-naophthylamine-*N,N*-diacetic acid; DBS, dodecylbenzenesulfonic acid; IC, ion chromatography; ICP-AES, inductively coupled plasma-atomic emission spectrometry; ISE, ion selective electrodes; TEA, tetrabutylammonium hydroxide.

## Analytical Techniques for Elemental Analysis

The various analytical techniques used for elemental analysis after oxygen flask combustion can be divided into two different types, those targeting a single element and those focusing on multielement analysis. In general, analytical techniques based on titration, those using ion-selective electrodes (ISEs), spectrophotometry, and spectrofluorimetry are used in single element determination, whereas separation methods based on IC, and HPLC are used for multielement determination after oxygen flask combustion.

### Single Element Determination

The analytical techniques used for single element analysis can be divided into three groups. The first group consists of titration techniques using various means to detect the end point of titration, such as volumetric titration, fluorophotometric titration, potentiometric titration, and spectrophotometric titration. The second group includes direct detection techniques such as direct ternary inclusion compound fluorescent spectrophotometry and the use of ISEs. The third group is that of separation methods such as IC and HPLC, which are used in complicated sample matrices to reduce the sample matrix interference.

### Titration and Direct Detection

New procedures for titration have been developed based on the synthesis or availability of new reagents or complexes that enable sensitive and selective detection of a given analyte. For example, a new metallofluorescent reagent, 3-carboxy-2-naphthoylamine-*N,N*-diacetic acid (CNDA), has been synthesized for fluorophotometry and as a metallofluorescent indicator for chelatometric titration, and a new reagent, DBH, has been developed to improve the determination of iodine as the removal of excess DBH 5-sulfosalicylic acid prior to determination is more suitable than that of the formic acid currently in use. For potentiometric microdetermination of tungsten, a hydrolysis step is taken in boiling the solution for 6–7 min after combustion to ensure complete dissolution of the  $WO_3$  formed during combustion. Procedures have been developed for the removal of  $CO_2$  and the adjustment of buffer pH prior to the use of a Pb(II) ISE as indicator electrode for potentiometric titration of W(VI) using  $10 \text{ mmol l}^{-1}$   $Pb(NO_3)_2$  solution to reduce sample matrix interference and to produce results consistent to standard methods.

For direct detection, a fluorometry based on a new ternary inclusion compound has been developed to enhance the detection sensitivity for Se

determination. The conventional fluorometric method for Se determination uses a fluorescent compound (2,3-diaminonaphthalene (DAN)–Se) formed between Se and DAN. The formation of a ternary inclusion compound in the cavity of  $\beta$ -cyclodextrin ( $\beta$ -CD) by reacting anionic surfactant sodium dodecylsulfate (SDS) with Se has been shown to produce a 10-fold enhancement in fluorescence and a significant decrease in the detection limit of the method employed. Various procedures using the fluoride ISE (fluoride-ISE) have been developed for direct and indirect determination of fluoride after oxygen flask combustion. Total ionic strength adjusting buffers (TISABs) are added to the absorption solution after combustion prior to the use of fluoride-ISE to produce a uniform ionic activity and reduce sample matrix interference.

### Separation Method for Complex Samples

Procedures based on separation techniques such as HPLC and IC have been developed for single element analysis for the following two reasons. The first reason is to remove interferents in complicated sample matrices that can give rise to incorrect results, in particular for trace analysis in samples with a high organic content, such as the determination of total iodine in egg products. The second reason is to differentiate the total and free forms of a specific element, such as the determination of the free iodide ion and bounded iodine in food additives. The free iodide ion is determined by direct sample injection into the IC column, whereas the total iodine content is determined after oxygen flask combustion. Thus, both the free and bounded forms of iodine in food samples can be determined.

### Multielement Determination

The following three analytical techniques have been developed for multielement determination after oxygen flask combustion: inductively coupled plasma (ICP)–atomic emission spectrometry (AES), IC, and RNAA. ICP–AES is mainly used for trace metal analysis and RNAA for multiple isotope determination in inorganic materials. From the summary of new procedures developed since 1995 using the oxygen flask method for elemental analysis listed in Table 1, it is clear that IC is the major analytical technique behind the recent development of multielement determination for oxygen flask combustion, in particular for organic samples.

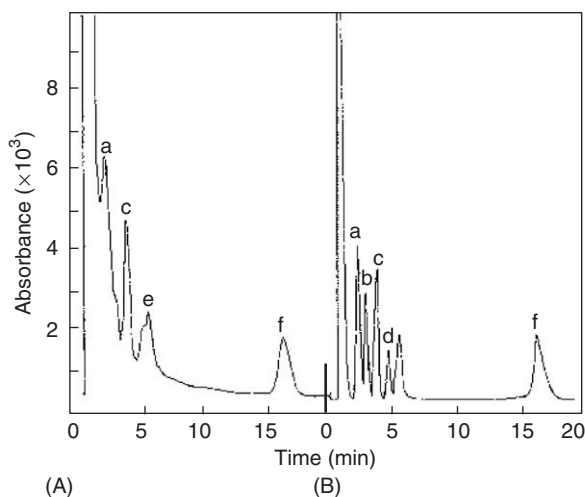
### Ion Chromatography

Suppressed IC was initially developed using an anion micromembrane suppressor and a conductivity

detector for multiple trace anion determination that could be used for multielement determination after oxygen flask combustion. However, the high cost of the equipment has restricted its use even in well-funded laboratories and the need for a second suppressor column has dictated the type of eluent used and limited the separation ability of the method. A strong alkaline mobile phase is used for separation, and this leads to the requirement of a special column such as the Shodex IC IF-424 anion exchange column for separation of the group of anions of interest in elemental analysis.

The recent advancement of nonsuppressed IC provides an attractive alternative method as ordinary HPLC equipment with UV detectors and reverse phase columns can be used for separation of the group of anions of interest in elemental analysis. Due to the strong competition among HPLC manufacturers in using their well established technology for an increasing market, the cost of HPLC equipment with UV detectors has come down considerably and within the budget for most laboratories. In addition, a weak acid mobile phase (typically 2 mmol l<sup>-1</sup> potassium hydrogen phthalate with eluent pH adjusted to 5.0) is used for separation using an ordinary reverse phase HPLC column (for example, a Hamilton PRP-X100 column). Thus, subsequent analysis of the anionic products generated from oxygen flask combustion can be performed in many laboratories equipped with HPLC apparatus. This has led to the popularity of using IC for multielement analysis as reported in the literature.

However, there are two major problems in the development of IC procedures for multielement determination. The first problem is the sloping of the baseline for samples with high organic contents such as wax, transformer oil, and lubricating oil. The second problem is the mutual interference of adjacent pairs of anions in an ion chromatogram with a large difference in concentration. The baseline stability problem is attributed to the high level of carbonates in solution as the result of absorption of carbon dioxide generated during combustion of samples with a high organic content. The effect of carbonate on the separation of common anions is shown in **Figure 2**. Carbonate is eluted just behind the injection peak, and the two peaks are so close that they are coeluted as a single peak in the ion chromatogram. Poor retention of CO<sub>3</sub><sup>2-</sup> is due to the formation of HCO<sub>3</sub><sup>-</sup>/H<sub>2</sub>CO<sub>3</sub> at an eluent pH of 5.0, a weak acid with little interaction with the analytical column. For CO<sub>3</sub><sup>2-</sup> concentrations greater than 50 µg ml<sup>-1</sup>, broadening and tailing of the injection peak are observed. Thus, early eluted monovalent anions such as F<sup>-</sup>, H<sub>2</sub>PO<sub>4</sub><sup>-</sup>, and Cl<sup>-</sup> could be

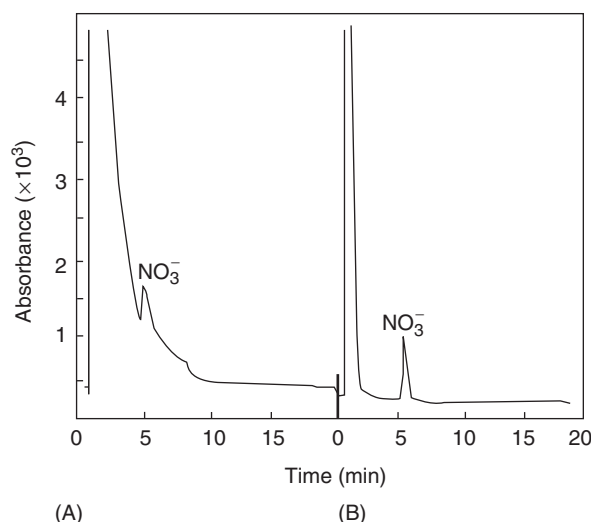


**Figure 2** The effect of carbonate on the separation of common anions. (A) anion standards with 60 µg ml<sup>-1</sup> CO<sub>3</sub><sup>2-</sup>; (B) without CO<sub>3</sub><sup>2-</sup>. Conditions: eluent = 2 mol l<sup>-1</sup> KHP, pH 5.0; detection at 272 nm; flow rate = 1.5 ml min<sup>-1</sup>; sample volume = 100 µl. Peaks: (a) 3 µg ml<sup>-1</sup> F<sup>-</sup>; (b) 5 µg ml<sup>-1</sup> H<sub>2</sub>PO<sub>4</sub><sup>-</sup>; (c) 4 µg ml<sup>-1</sup> Cl<sup>-</sup>; (d) 2 µg ml<sup>-1</sup> Br<sup>-</sup>; (e) 5 µg ml<sup>-1</sup> NO<sub>3</sub><sup>-</sup>; (f) 10 µg ml<sup>-1</sup> SO<sub>4</sub><sup>2-</sup>. (Reprinted with permission from Fung YS and Dao KL (1996) Elemental analysis of chemical wastes by oxygen bomb combustion-ion chromatography. *Analytica Chimica Acta* 334: 51–56; © Elsevier.)

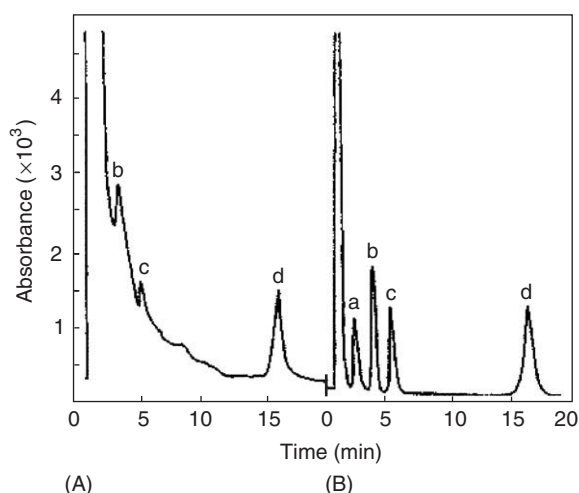
masked by the injection peak. Although diluting the CO<sub>3</sub><sup>2-</sup> solution could reduce the interference, it also lowers the sensitivity of the method to detecting other low-level heteroatoms.

The stability of the baseline can be improved as shown in **Figure 3** by passing the absorption solution through a cation exchange resin column to convert it into H<sub>2</sub>CO<sub>3</sub> prior to removing CO<sub>2</sub> by heating the solution. Thus, it enables the detection of low levels of F<sup>-</sup> and Cl<sup>-</sup> in real samples, as shown by the chromatogram of a motor oil sample as given in **Figure 4**. By spiking anion standard solutions after the combustion of kerosene prior to ion exchange treatment, over 96% recoveries are obtained for F<sup>-</sup>, H<sub>2</sub>PO<sub>4</sub><sup>-</sup>, Cl<sup>-</sup>, Br<sup>-</sup>, and SO<sub>4</sub><sup>2-</sup> with a relative standard deviation (RSD) less than 6%, linear ranges up to three orders of magnitudes, repeatability of less than 2%, and detection limits at sub µg ml<sup>-1</sup> levels for all analyte anions.

The mutual interference between closely eluted analyte peaks in an ion chromatogram are shown in **Table 2**. Three critical adjacent ion pairs have been identified as follows: H<sub>2</sub>PO<sub>4</sub><sup>-</sup> on F<sup>-</sup>, F<sup>-</sup> on H<sub>2</sub>PO<sub>4</sub><sup>-</sup>, and NO<sub>3</sub><sup>-</sup> on Br<sup>-</sup>. As the retention times of F<sup>-</sup> and H<sub>2</sub>PO<sub>4</sub><sup>-</sup> are 2.4 and 2.8 min, respectively, the peak resolution between F<sup>-</sup> and H<sub>2</sub>PO<sub>4</sub><sup>-</sup> decreases gradually as their concentration are increased. For the determination of 1 µg ml<sup>-1</sup> F<sup>-</sup>, the error is less than 5% for a H<sub>2</sub>PO<sub>4</sub><sup>-</sup> concentration up to 25 µg ml<sup>-1</sup>.



**Figure 3** Analysis of blank sample: (A) without removal of  $\text{CO}_3^{2-}$ ; (B) with ion-exchange treatment. Conditions same as those given in **Figure 2** and using 0.59 kerosene as sample. (Reprinted with permission from Fung YS and Dao KL (1996) Elemental analysis of chemical wastes by oxygen bomb combustion-ion chromatography. *Analytica Chimica Acta* 334: 51–56; © Elsevier.)



**Figure 4** The effect of carbonate on the analysis of 1 g motor oil: (A) without removal of  $\text{CO}_3^{2-}$ ; (B) with ion-exchange treatment. Conditions same as those given in **Figure 2**. Peaks: (a)  $1.2 \mu\text{g ml}^{-1} \text{F}^-$ ; (b)  $2.5 \mu\text{g ml}^{-1} \text{Cl}^-$ ; (c)  $4 \mu\text{g ml}^{-1} \text{NO}_3^-$ ; (d)  $7 \mu\text{g ml}^{-1} \text{SO}_4^{2-}$ . (Reprinted with permission from Fung YS and Dao KL (1996) Elemental analysis of chemical wastes by oxygen bomb combustion-ion chromatography. *Analytica Chimica Acta* 334: 51–56; © Elsevier.)

**Table 2** The interference between closely eluted pairs of analyte peaks on ion chromatograms

Anion 1	Concentration ( $\mu\text{g ml}^{-1}$ )	Anion 2	Concentration ( $\mu\text{g ml}^{-1}$ )	Error in measuring anion 1 in presence of anion 2 (%)
Fluoride	1	Phosphate	5	0
	1		10	–1
	1		15	–1.8
	1		20	–2.8
	1		25	–4.8
	1		30	–6.3
Phosphate	5	Fluoride	0.5	0.1
	5		1	0.2
	5		2	–1.1
	5		3	–2.1
	5		4	–5
	5		5	–11
Bromine	2	Nitrate	5	–0.2
	2		10	–0.8
	2		15	–1.3
	2		20	–4.1
	2		25	–5.5
	2		30	–8.6

However, due to the higher peak height response of  $\text{F}^-$ , the determination of  $5 \mu\text{g ml}^{-1} \text{H}_2\text{PO}_4^-$  leads to a more than 10% error for an  $\text{F}^-$  concentration greater than  $5 \mu\text{g ml}^{-1}$ . For  $\text{Br}^-$  and  $\text{NO}_3^-$ , only the interference of  $\text{NO}_3^-$  on  $\text{Br}^-$  is considered because elemental analysis of nitrogen is not possible due to the variable oxidation of atmospheric nitrogen during combustion. A less than 5.5% error has been obtained for measuring  $2 \mu\text{g ml}^{-1} \text{Br}^-$  in the presence of  $\text{NO}_3^-$  up to  $25 \mu\text{g ml}^{-1}$ . A suitable dilution

can be carried out to reduce the nitrate interference for samples with a low  $\text{Br}^-$  content.

To compensate for the interfering effect of  $\text{CO}_3^{2-}$  and  $\text{NO}_3^-$  on the determination of  $\text{F}^-$ ,  $\text{Cl}^-$ ,  $\text{Br}^-$ , and  $\text{S}^-$ , calibration curves based on different ratios of atomic compositions using organic standards with specified ratios of heteroatoms have been constructed. An automatic selecting program has been devised to select the most suitable calibration curve for a given sample. Highly accurate results have been

obtained, and the overall performance of the procedure is highly efficient and practical for the determination of a large number of samples with multiple elements.

### Metals, Isotopes, and Others

The following analytical techniques have been developed for multielement determination: ICP-AES, RNAA, and the chelating titration method. ICP-AES has been used for total sulfur determination. For multiple trace metal analysis, the trend is toward direct determination without combustion, in particular for organic compounds. For RNAA, a procedure has been developed to combine the determination of short-life nuclides with medium- and long-life nuclides by making a medium or long irradiation prior to short irradiation and radiochemical processing. Thus, an RNAA procedure has been developed using oxygen flask ignition of the irradiated samples for the determination of  $^{127}\text{I}$ ,  $^{128}\text{I}$ ,  $^{235}\text{U}$ ,  $^{133}\text{I}$ , and  $^{203}\text{Hg}$  in biological and environmental samples. Spectrophotometric procedures have been developed for determination of the chelates of Ce(III) at 400 nm, Th(IV) at 660 nm, and U(VI) at 555 nm.

### Applications

Applications of the oxygen flask method range from traditional elemental analysis of organic compounds to samples with more complicated matrixes such as food products, plant material (citrus leaf, canola seed, corn leaf, pine needle, red clover shoot, wheat flour, and orchard grass), organic pharmaceutical compounds, hair, sediment, and soil samples. For practical sample analysis, method validation is important. Two approaches have been adopted to check the validation of the method developed. First, the recovery of spiked standards is assessed normally, with results showing greater than 90% recoveries for all analytes investigated. Second, a parallel method comparison is performed to check the consistency of results, in particular on real samples with no primary standards available. Results of comparisons of various analytical methods employed for method validation are given in Table 3.

The results of the recently developed commercial automated oxidation equipment have been compared with those obtained from the traditional oxygen flask method for various plant materials. Out of the six methods compared (dry ashing/ICP-AES, oxygen flask/ICP-AES, oxygen flask/IC, automated combustion, microwave ( $\text{HNO}_3/\text{HClO}_4$ )/ICP-AES, microwave ( $\text{HNO}_3/\text{H}_2\text{O}_2/\text{HCl}$ )/ICP-AES), the oxygen flask/IC and oxygen flask/ICP-AES are found to give similar total sulfur content for all six plant tissues types

studied, indicating complete oxidation of the plant sulfur to sulfate. The low results obtained by other methods are attributed to incomplete destruction and sulfur volatilization during combustion.

Parallel method comparisons are used to establish the validity of a new method developed for five organic pharmaceutical compounds, food colors, and color additives. The standard methods such as the Japanese Standard Food Additives and Japanese Standard of Cosmetic Ingredients method, based on volumetric and gravimetric titration, have been used to establish new methods developed for the determination of I, Cl, Br, and  $\text{SO}_4^{2-}$  in food colors. The results obtained indicate good agreement in both accuracy and precision for procedures based on the oxygen flask method as compared with the standard methods. In addition to anion elemental analysis, method validation has also been carried out for metal analysis such as that of Ce(III), Th(IV), and U(VI), with the results showing acceptable limits of variation.

Analyzing elements with and without the use of the oxygen flask, the free and bounded forms of a given element can be determined. This scheme has been adopted in assessing free and bounded iodine for nutritional assessment. Different forms can also be differentiated by their differential adsorption during sample preparation prior to oxygen flask combustion. This scheme has found an application in assessing the risk for fluorosis of workers in polluted environments and in ecotoxicology studies.

Despite the recent development of commercial automated oxygen combustion equipment, the oxygen flask combustion method remains a method for choice for elemental analysis in many laboratories due to its simple, rapid, and economic procedure using readily available apparatus. From the record of the past 5 years, three directions of development of the oxygen flask method can be seen. First, it shows a trend toward microdetermination, in particular for samples of less than 1 mg. Second, new procedures have been developed for multielement determination after oxygen flask combustion, mostly based on IC separation of the anions produced. Third, extensive validation of the procedure developed has been carried out for real sample analysis with a parallel method comparison with two to six different procedures.

The three directions will certainly be continued in future. For microdetermination and multielement analysis, new procedures based on the recently developed capillary electrophoresis are expected to come, in particular for cases with limited sample sizes. The applicability to real sample analysis is expected to grow, together with work on validation of the methodology developed. A survey has been

**Table 3** Comparison of analytical methods for elemental analysis after combustion

Analyte	Sample	Methods compared	Summary of results	Remarks
S	Plant materials (citrus leaf, canola seed, corn leaf, pine needle, red clover shoot, wheat flour)	A (dry ashing/ICP–AES) B (oxygen flask/ICP–AES) C (oxygen flask/IC) D (automated combustion <sup>b</sup> ) E (microwave (HNO <sub>3</sub> /HClO <sub>4</sub> )/ICP–AES) F (microwave (HNO <sub>3</sub> /H <sub>2</sub> O <sub>2</sub> /HCl)/ICP–AES)	Average results for two daily batches of three subsamples from each of six tissues ( $n=36$ ) indicate $D = F > B = C = A > E$ B and C gave similar S values for all tissue types, indicating complete oxidation of plant S to sulfate	Total S
	Plant material (orchard grass)	A (oxygen flask) B (automated combustion <sup>b</sup> ) C (alkaline digestion) D (automated dry combustion <sup>c</sup> ) E and F (perchloric acids)	$A = B > C = D > E = F$ Incomplete destruction and S volatilization led to low S results	Total S
F	Five organic pharmaceutical compounds	A (EDTA back-titration) B (ISE potentiometric titration)	Good agreement for A and B for both accuracy and precision	
I	Food colors	A (oxygen flask/IC) B (gravimetric method of JSFA <sup>a</sup> )	Total I <sup>–</sup> by A in good agreement with B	Total I <sup>–</sup> and free I <sup>–</sup>
Ce(III), Th(IV), and U(VI)	Metal chelates	A (volumetric titration) B (spectrophotometry of metal complexes)	Results from A and B are within acceptable limits	
Cl <sup>–</sup> , Br <sup>–</sup> , and SO <sub>4</sub> <sup>2–</sup>	Color additives	A (oxygen flask/IC) B (titration method of JSCI) C (gravimetric method of JSCI)	Total Br by A in good agreement with B. Total S by A in good agreement with C	

<sup>a</sup>JSFA, Japanese Standard Food Additives; JSCI, Japanese Standard of Cosmetic Ingredients.

<sup>b</sup>Automated combustion, LECO CNS-2000.

<sup>c</sup>Automated dry combustion, Fisher automated dry combustion machine.

ISE, Ion selective electrode; ICP–AES, Inductively coupled plasma–atomic emission spectrometry; IC, Ion chromatography.

conducted using the oxygen flask method for measuring Se in blood plasma and in hair for toxic metal risk assessment. Given the extensive work done on method validation for elemental analysis of various plant materials using the oxygen flask method, this method is expected to provide a routine method for screening food products and plant material in the coming years as a validated method using a cheap and rapid procedure with readily available glass apparatus is the necessity for large-scale screening work.

See also: **Activation Analysis:** Neutron Activation. **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Blood and Plasma. Elemental Speciation:** Practicalities and Instrumentation. **Food and Nutritional Analysis:** Contaminants; Vegetables and Legumes. **Fuels:** Oil-Based. **Geochemistry:** Inorganic; Sediment; Soil, Major Inorganic Components; Soil, Minor Inorganic Components; Soil, Organic Components. **Ion-Selective Electrodes:** Food Applications; Water Applications.

**Radiochemical Methods:** Radionuclide Monitoring. **Spectrophotometry:** Inorganic Compounds. **Titrimetry:** Potentiometric; Photometric.

## Further Reading

- Campiglio A (1997) Potentiometric microdetermination of tungsten in organic compounds after combustion in a modified oxygen flask. *Microchemistry Journal* 55: 375–381.
- Dermelj M and Byrne AR (1997) Simultaneous radiochemical neutron activation analysis of iodine, uranium and mercury in biological and environmental samples. *Journal of Radioanalytical and Nuclear Chemistry* 216: 13–18.
- Hassouna M and Issa YM (1995) Volumetric and spectrophotometric micro-determination of cerium(III), thorium(V) and uranium(VI) in some chelates of 3-acetyl-1,5-diarylformazans and 3-cyano-1,5-diarylformazans. *Journal of the Indian Chemical Society* 72: 709–712.
- Hilp M (1998) Optimization of iodine determination according to Schöniger – analytical chemistry with



- 1,3-dibromo-5,5-dimethylhydantoin (DBH). Part 1, oxygen flask combustion, Part 7. *Fresenius Journal of Analytical Chemistry* 360: 184–191.
- Kowalenko CG (2000) Influence of sulfur and nitrogen fertilizer applications and within-season sampling time on measurement of total sulfur in orchardgrass by six different methods. *Communications in Soil Science and Plant Analysis* 31: 345–354.
- Kuboyama K, Sasaki N, Nakagome Y, Kataoka M, and Tamura C (1996) Simultaneous microdetermination of fluorine, chlorine, bromine and sulfur in organic compounds by ion chromatography. *Bunseki Kagaku* 45: 71–76.
- Nagashima H and Kuboyama K (1996) Microdetermination of phosphorus in organic compounds by ion chromatography. *Bunseki Kagaku* 45: 941–946.
- Okamoto T, Isozaki A, and Nagashima H (1996) Determination of chloride, bromide and sulfate in color additives by ion chromatography on ceramic carbon column. *Bunseki Kagaku* 45: 865–871.
- Soon YK, Kalra YP, and Abboud SA (1996) Comparison of some methods for the determination of total sulfur in plant tissues. *Communications in Soil Science and Plant Analysis* 27: 809–818.
- Sun Y and Li H (2000) Determination of trace selenium in human plasma and hair with ternary inclusion compound-fluorescent spectrometry. *Analyst* 125: 2326–2329.
- Woodget B (1995) Oxygen flask combustion. In: Townshend A, Worsfold PJ, Haswell SJ, *et al.* (eds.) *Encyclopedia of Analytical Science*, 1st edn., pp. 4495–4502.

## Wet Digestion

R M Twyman, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by A D Sawant, pp. 4503–4510, © 1995, Elsevier Ltd.

### Introduction

Wet digestion methods for elemental analysis involve the chemical degradation of sample matrices in solution, usually with a combination of acids to increase solubility. The various acid and flux treatments are carried out at high temperatures in specially designed vessels that help to minimize contamination of the sample with substances in the air, the local environment, and from the vessel walls. Losses from the sample may occur due to adsorption onto the vessel walls, volatilization, and coextraction, but these can be reduced by procedural modifications. The use of closed systems, where the digestion reaction is completely isolated from the surroundings, may help to reduce both contamination and sample loss.

### The Nature of the Sample Matrix

Many techniques employed for elemental analysis require the conversion of the sample matrix into a solution form. The selection of an appropriate treatment for sample dissolution depends on the nature of the sample, and different approaches are required for predominantly inorganic and predominantly organic matrices. Geological, geochemical, and soil samples generally contain silicate, metal oxides, carbonates, and, in many cases, organic matter. Such samples must be dried and ground to a fine powder to

facilitate dissolution. Minerals and coal often have a nonuniform distribution of elements, while fly ash is very fine and is composed of metal silicates and oxides. Both these types of sample are difficult to solubilize. Similarly, alloys can be difficult to dissolve because of the strong bonds between metal atoms and their brittle nature. Solid and crystalline samples may possess interstitial water and water of crystallization, so thorough drying of samples is necessary before and after grinding.

Biological samples must be processed with great care, since the dissolution and total decomposition of all organic matter is required for the release of trace elements. However, the use of oxidizing acids to decompose organic matter can produce violent reactions and the alternative procedure of dry ashing may be more suitable in some cases. Environmental and water samples often contain mixtures of organic and inorganic substances, so dissolution techniques need to be modified to take this composition into account. In particular, water samples may contain dissolved and suspended solids, colloids, and microorganisms. Elements embedded in such samples may be present both in dissolved and solid forms. The nature of the sample matrix must be given special attention during wet digestion.

### Extraction of the Analyte

Solid samples generally contain some adsorbed and/or absorbed water. In the case of inorganic materials, drying is carried out in an oven at 105–110°C for a few hours, although lower temperatures need to be used if the sample contains volatile components. On the other hand, higher temperatures may be required to remove water trapped within crystalline matrices.

- 1,3-dibromo-5,5-dimethylhydantoin (DBH). Part 1, oxygen flask combustion, Part 7. *Fresenius Journal of Analytical Chemistry* 360: 184–191.
- Kowalenko CG (2000) Influence of sulfur and nitrogen fertilizer applications and within-season sampling time on measurement of total sulfur in orchardgrass by six different methods. *Communications in Soil Science and Plant Analysis* 31: 345–354.
- Kuboyama K, Sasaki N, Nakagome Y, Kataoka M, and Tamura C (1996) Simultaneous microdetermination of fluorine, chlorine, bromine and sulfur in organic compounds by ion chromatography. *Bunseki Kagaku* 45: 71–76.
- Nagashima H and Kuboyama K (1996) Microdetermination of phosphorus in organic compounds by ion chromatography. *Bunseki Kagaku* 45: 941–946.
- Okamoto T, Isozaki A, and Nagashima H (1996) Determination of chloride, bromide and sulfate in color additives by ion chromatography on ceramic carbon column. *Bunseki Kagaku* 45: 865–871.
- Soon YK, Kalra YP, and Abboud SA (1996) Comparison of some methods for the determination of total sulfur in plant tissues. *Communications in Soil Science and Plant Analysis* 27: 809–818.
- Sun Y and Li H (2000) Determination of trace selenium in human plasma and hair with ternary inclusion compound-fluorescent spectrometry. *Analyst* 125: 2326–2329.
- Woodget B (1995) Oxygen flask combustion. In: Townshend A, Worsfold PJ, Haswell SJ, *et al.* (eds.) *Encyclopedia of Analytical Science*, 1st edn., pp. 4495–4502.

## Wet Digestion

**R M Twyman**, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by A D Sawant, pp. 4503–4510, © 1995, Elsevier Ltd.

### Introduction

Wet digestion methods for elemental analysis involve the chemical degradation of sample matrices in solution, usually with a combination of acids to increase solubility. The various acid and flux treatments are carried out at high temperatures in specially designed vessels that help to minimize contamination of the sample with substances in the air, the local environment, and from the vessel walls. Losses from the sample may occur due to adsorption onto the vessel walls, volatilization, and coextraction, but these can be reduced by procedural modifications. The use of closed systems, where the digestion reaction is completely isolated from the surroundings, may help to reduce both contamination and sample loss.

### The Nature of the Sample Matrix

Many techniques employed for elemental analysis require the conversion of the sample matrix into a solution form. The selection of an appropriate treatment for sample dissolution depends on the nature of the sample, and different approaches are required for predominantly inorganic and predominantly organic matrices. Geological, geochemical, and soil samples generally contain silicate, metal oxides, carbonates, and, in many cases, organic matter. Such samples must be dried and ground to a fine powder to

facilitate dissolution. Minerals and coal often have a nonuniform distribution of elements, while fly ash is very fine and is composed of metal silicates and oxides. Both these types of sample are difficult to solubilize. Similarly, alloys can be difficult to dissolve because of the strong bonds between metal atoms and their brittle nature. Solid and crystalline samples may possess interstitial water and water of crystallization, so thorough drying of samples is necessary before and after grinding.

Biological samples must be processed with great care, since the dissolution and total decomposition of all organic matter is required for the release of trace elements. However, the use of oxidizing acids to decompose organic matter can produce violent reactions and the alternative procedure of dry ashing may be more suitable in some cases. Environmental and water samples often contain mixtures of organic and inorganic substances, so dissolution techniques need to be modified to take this composition into account. In particular, water samples may contain dissolved and suspended solids, colloids, and microorganisms. Elements embedded in such samples may be present both in dissolved and solid forms. The nature of the sample matrix must be given special attention during wet digestion.

### Extraction of the Analyte

Solid samples generally contain some adsorbed and/or absorbed water. In the case of inorganic materials, drying is carried out in an oven at 105–110°C for a few hours, although lower temperatures need to be used if the sample contains volatile components. On the other hand, higher temperatures may be required to remove water trapped within crystalline matrices.

Frequently, the sample is not soluble in water and must be treated with acids or mixtures of acids to facilitate solubilization. The type of acid treatment must be given careful consideration, since particular acids may or may not oxidize the sample, and may be incompatible with certain elements. For example, sulfuric acid cannot be used to dissolve samples containing barium, while hydrochloric acid cannot be used to dissolve silver or samples containing lead and lead compounds. The choice of acid is also restricted by sample volatilization, e.g., hydrochloric acid should be avoided in samples containing arsenic since this is more volatile as a trichloride.

Naturally occurring inorganic materials, such as ores, must be given special treatments to facilitate solubilization. The two most common methods employed in dissolving samples are treatments with hydrochloric, hydrofluoric, nitric, sulfuric, or perfluoric acids (or various combinations thereof) and fusion with an acidic or basic flux followed by treatment with water or an acid. Organic materials are usually decomposed by wet digestion with a boiling oxidizing acid or acid mixture, ultimately producing carbon dioxide, water, and other volatile compounds that are driven off to leave behind salts or acids of the inorganic constituents of the sample. Wet digestions may be performed in open beakers on hot plates, but Kjeldahl flasks or specially designed containment vessels give results that are more satisfactory.

### Wet Digestion with Single Acids

The solvent action of an acid depends on several factors:

1. The reduction of hydrogen ions by metals that are more active than hydrogen, for example:  

$$\text{Zn(s)} + 2\text{H}^+ \rightarrow \text{Zn}^{2+} + \text{H}_2\text{(g)}.$$
2. The combination of hydrogen ions with anions of a weak acid, for example:  $\text{CaCO}_3\text{(s)} + 2\text{H}^+ \rightarrow \text{Ca}^{2+} + \text{H}_2\text{O} + \text{CO}_2\text{(g)}.$
3. The oxidizing properties of the acid anion, for example:  $3\text{Cu(s)} + 2\text{NO}_3^- + 8\text{H}^+ \rightarrow 3\text{NO(g)} + 4\text{H}_2\text{O}.$
4. The tendency of the acid anion to form soluble complexes with the sample cation, for example:  

$$\text{Fe}^{3+} + \text{Cl}^- \rightarrow \text{FeCl}^{2+}.$$

Ideally, the chosen reagent should cause the complete dissolution of the sample. As a general guide it is useful to classify the more common acid treatments according to whether they oxidize the sample or not. The nonoxidizing acids include dilute hydrochloric, hydrofluoric, sulfuric, and perchloric acids, whereas the oxidizing acids include hot, concentrated nitric, sulfuric, and perchloric acids. Dissolution of metals

by nonoxidizing acids is a process of hydrogen replacement.

Hydrochloric acid will dissolve metals above the standard reduction potential of hydrogen, salts of weak acids, and many oxides. Dilute sulfuric and perchloric acids are useful for metals above the standard reduction potential of hydrogen. Hot, concentrated sulfuric acid will often dissolve metals below the standard reduction potential of hydrogen. The most potent oxidizing conditions are obtained using hot concentrated perchloric acid, which will dissolve all common metals. Concentrated hydrochloric acid is an excellent solvent for many metal oxides as well as those metals that are more easily oxidized than hydrogen. In addition, it is often a better solvent for oxides than the oxidizing acids.

Hot, concentrated nitric acid will dissolve all common metals with the exception of aluminum and chromium, which are passive to the reagent as a result of surface oxide formation. Hot nitric acid also readily oxidizes many organic substances. Hot, concentrated sulfuric acid can be used to decompose and dissolve many substances in part because of its high boiling point (340°C), and it is particularly useful for the dehydration and oxidation of organic samples. Most metals and alloys are also attacked by this hot acid.

Perchloric acid is a potent oxidizing agent that leads to the formation of highly soluble perchlorate salts. As with sulfuric acid, perchloric acid dehydrates and oxidizes organic samples very efficiently. It also attacks iron alloys and stainless steel, which are resistant to other mineral acids. Care is required when using perchloric acid because it is explosive in contact with certain organic compounds and easily oxidized inorganic materials. Special chemical hoods are recommended. Perchloric acid, as a 72–74% solution, boils at 203°C.

Hydrofluoric acid is a weak, nonoxidizing acid that is particularly useful for dissolving silicate samples since it removes the silicon quantitatively as volatile  $\text{SiF}_4$ . In many cases, hydrofluoric acid dissolution can be achieved by adding sodium fluoride to samples treated with hydrochloric acid.

### Wet Digestion with Acid Mixtures

Acids in combination are preferred for certain inorganic matrices and are generally more advantageous for the decomposition of organic compounds. Wet digestion procedures using acid mixtures can be divided into four types:

1. Total decomposition, usually with hydrofluoric acid and another mineral acid.
2. Strong attacks, for routine analysis but leaving a residue of certain minerals, particularly

silicates. Carried out with various mixtures of sulfuric, nitric, and perchloric acids.

3. Moderate attacks, using weaker acid mixtures.
4. Partial digestions (acid leaching).

Both (3) and (4) are typically employed for environmental analysis where complete dissolution is either not required or is undesirable and the goal is to determine the presence of certain trace elements.

For geochemical samples containing silicates, the matrix is decomposed by heating with hydrofluoric acid in combination with either nitric or perchloric acid, each of which has a higher boiling point than hydrofluoric acid. The presence of the second acid with a higher boiling point ensures that, once the hydrofluoric acid has been boiled off and the dry sample redissolved, sparingly soluble metal fluorides are converted to salts that are more soluble. As stated above, however, caution should be exercised with the use of perchloric acid if the sample has a significant organic component. Perchloric acid is also more expensive than nitric acid, and can introduce chloride ions as contaminants.

For organic samples, a widely used mixture is aqua regia (1:3 nitric acid–hydrochloric acid). The nitric acid acts as the oxidizing agent, while the hydrochloric acid provides the complexing properties. The addition of bromine or hydrogen peroxide can sometimes increase the solubilizing power of mineral acids. Wet digestion is generally carried out in open flasks, covered loosely to avoid atmospheric contamination. However, it is becoming increasingly common to use closed vessels, such as polytetrafluoroethylene (PTFE)-lined containers or ultrapure quartz vessels, especially for small samples.

A 1:4 mixture of sulfuric and nitric acids is also widely employed for organic samples. The nitric acid decomposes the bulk of the organic matter but does not reach a temperature sufficient to destroy the last traces. However, as the nitric acid boils off, the sulfuric acid is left behind. Dense  $\text{SO}_3$  fumes evolve and begin to reflux in the flask, making the solution very hot and allowing the hot sulfuric acid to decompose the remaining organic matter. Because of the fumes produced in this method, it must be carried out under a fume hood. More nitric acid may be added to prolong the digestion and eliminate any stubborn organic material.

A very efficient acid mixture is nitric, sulfuric, and perchloric acid in a volume ratio of  $\sim 3:1:1$ . For a typical 10 g sample of tissue or blood, 10 ml of this solution is sufficient for complete dissolution. The samples are heated until the nitric acid boils off and perchloric acid fumes begin to appear. Heating continues until the perchloric acid boils off and  $\text{SO}_3$

fumes appear. There is little danger of perchloric acid explosions as long as sufficient nitric acid is present to decompose the bulk of the organic matter, and as long as sulfuric acid remains after the perchloric acid has evaporated to prevent the sample becoming dry. Perchloric acid should never be added directly to an organic sample. A mixture of nitric and perchloric acid may also be used.

The availability of strong hydrogen peroxide solutions allows a combination of sulfuric acid and hydrogen peroxide to be used for the decomposition of organic matter. Hydrogen peroxide is a vigorous oxidizing agent and is particularly useful for the degradation of resistant plastics. There is little danger of explosion if sulfuric acid is present in excess. Most elements can be recovered quantitatively in this procedure, with the exceptions of ruthenium, osmium, germanium, arsenic, and selenium. In the case of germanium and arsenic, loss is attributable to volatilization of chlorides. Additionally, precipitated calcium sulfate may retain lead and silver if not solubilized. After decomposition, the sulfuric acid solution should be diluted and boiled gently for 10 min to destroy any remaining hydrogen peroxide.

The recovery of trace elements after wet digestion using the methods described above is compared in Tables 1 and 2.

**Table 1** Recovery of trace elements after wet digestion of cocoa

Element	Recovery <sup>a</sup> (%)			
	$\text{HNO}_3$ & $\text{HClO}_4$	$\text{HNO}_3$ , $\text{HClO}_4$ & $\text{H}_2\text{SO}_4$	$\text{HNO}_3$ & $\text{H}_2\text{SO}_4$	$\text{HNO}_3$
Ag	91	99	100	95
As	99	99	98	91, 98
Cd	101	102	102	99
Co	98	99	101	100
Cr	100	101	100	101
Cu	100	99	100	100
Fe	99	99	102	97
Hg	79	89	93	—
Mo	97	98	101	98
Pb	100	99, 93	90, 93	100
Sb	99, 94	100	100	97
Se	101	101	79	1
Sr	100	97	98	95
Zn	99	101, 94	100	99

<sup>a</sup>Oxidation mixtures respectively as follows:

15 ml concentrated  $\text{HNO}_3$ , 10 ml 60%  $\text{HClO}_4$ .

15 ml concentrated  $\text{HNO}_3$ , 5 ml conc.  $\text{H}_2\text{SO}_4$ , 10 ml 60%  $\text{HClO}_4$ .

15 ml concentrated  $\text{HNO}_3$ , 10 ml conc.  $\text{H}_2\text{SO}_4$ .

Concentrated and fuming  $\text{HNO}_3$ .

500 ml flask used for 2 g cocoa sample.

Radioactive nuclides were used to determine recovery of 1–10 mg  $\text{kg}^{-1}$  level.

### Wet Digestion with Fluxes

Some materials are particularly resistant to acid digestion, e.g., certain rocks, mineral oxides, phosphates, and some iron alloys. For these samples, high-temperature fusion with an acidic or basic flux such as lithium metaborate ( $\text{LiBO}_3$ ) in the molten state can be used to render such materials soluble in water or dilute acid. Fusion decompositions are the most rigorous methods available and all silicate materials, including refractory substances like zircon and cassiterite, can be dissolved completely when fused with an appropriate flux. However, there are several disadvantages to this method including the introduction of additional salts into the final solution

**Table 2** Recovery of metals from various biomaterials after wet digestion

Element	Recovery (%)				
	Blood	Urine	Leaves	Animal tissue	Reflux condenser
As	93	94	97	92	101
Au	77	100	77	65	100
Fe	98	92	95	85	100
Hg	24	84	45	30	100
Sb	99	95	94	94	100

Ag, Co, Cr, Cu, Mn, Mo, Pb, V, Zn: 98–102% for all samples. Samples: 5 ml horse blood; 250 mg sugar beat leaves; 10 ml urine; 1.5 g animal tissue.

Digested for 1.5 h with 5 ml of a mixture of 98%  $\text{H}_2\text{SO}_4$ , 70%  $\text{HClO}_4$ , and 66%  $\text{HNO}_3$  in the volume ratio 1:1:3 in a micro-Kjeldahl flask.

Amount of metal present was 0.05–0.3 mg.

Urine was digested with concentrated  $\text{HNO}_3$  only.

(e.g., lithium metaborate, typically used at a 3:1 flux-sample ratio, introduces lithium and boron), loss of materials by volatilization, and the encouragement of sample contamination from the vessel walls, which also become solubilized. The properties of some useful fluxes are summarized in Table 3.

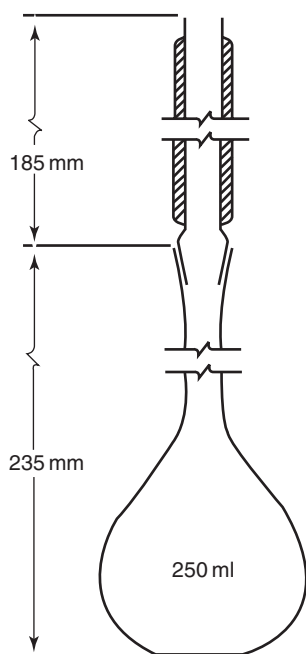
### Dissolution Techniques

Depending on the nature, origin, sample quantity, elements of interest, and concentration range, a variety of different techniques for sample dissolution can be used. Dissolution assemblies such as the Kjeldahl flask, Parr bomb, crucible, quartz and platinum vessels, PTFE containers, and specially designed apparatus can be used for wet digestion (Figures 1 and 2). Many suitable small vessels and devices for sample heating and filtering are available. In the case of classical milliliter, microliter, and nanoliter methods, solution concentrations are generally  $>0.01 \text{ mol l}^{-1}$ . The quantities of the decomposition reagents and the decomposition vessels are adjusted to the smaller sample weights.

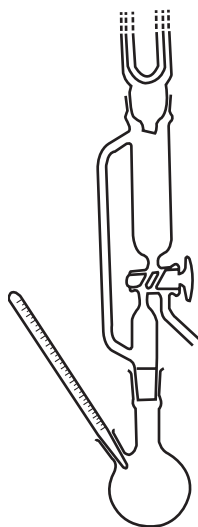
Inorganic materials subjected for trace metal analysis are generally decomposed and solubilized by treatment with mineral acids or by fusion with fluxes. Some inorganic samples, e.g., soils, may also contain considerable amounts of organic material, which can also be solubilized by acid treatment. Faster and more efficient decompositions of both organic and inorganic materials is possible using closed reaction vessels (bombs) that allow higher temperatures and pressures to be applied. For example, bombs have been designed for the decomposition of silicates with

**Table 3** Common fluxes

Flux	Melting point ( $^{\circ}\text{C}$ )	Type of crucible/dish for fusion	Type of substance decomposed
$\text{Na}_2\text{CO}_3$	851	Pt	For silicates and silica containing samples; alumina-containing samples; sparingly soluble phosphates and sulfates
$\text{Na}_2\text{CO}_3$ + an oxidizing agent such as $\text{KNO}_3$ , $\text{KClO}_3$ , or $\text{Na}_2\text{O}_2$	318	Pt (not with $\text{Na}_2\text{O}_2$ ), Ni	For samples requiring an oxidizing environment; that is, samples containing S, As, Sb, Cr, etc.
$\text{NaOH}$ or $\text{KOH}$	380	Au, Ag, Ni	Powerful basic fluxes for silicates, silicon carbide, and certain minerals; main limitation: purity of reagents
$\text{Na}_2\text{O}_2$	Decomposes	Fe, Ni	Powerful basic oxidizing flux for sulphides; acid insoluble alloys of Fe, Ni, Cr, Mo, W and Li; platinum alloys, Cr, Sn, Zr minerals
$\text{K}_2\text{S}_2\text{O}_7$	300	Pt, porcelain	Acid flux for slightly soluble oxides and oxide-containing samples
$\text{B}_2\text{O}_3$	577	Pt	Acid flux for decomposition of silicates and oxides where alkali metals are to be determined
$\text{CaCO}_3 + \text{NH}_4\text{Cl}$		Ni	Upon heating the flux, a mixture of $\text{CaO}$ and $\text{CaCl}_2$ produced; used to decompose silicates for the determination of the alkali metals



**Figure 1** Reaction flask with asbestos-jacketed air condenser. The flask is a 250 ml borosilicate volumetric flask. A condenser is made from ST 19/38 ground glass joint, with a jacket of woven asbestos tubing held in place by wrapping with PTFE sealing tape. (Reproduced with permission from Feldman C (1974) *Analytical Chemistry* 46: 1607; © American Chemical Society.)



**Figure 2** Apparatus for wet oxidation of organic material. Stopcock, having three positions, allows (1) refluxing, (2) distillation, (3) removal of distillate. (Taken from Bethge PO (1954) *Analytica Chimica Acta* 10: 317; and Gorsuch TT (1959) *Analyst* 84: 147. Copyright by the Chemical Society.)

hydrofluoric acid at temperatures up to 150°C and pressures in excess of 8000 kPa. In the milligram range, the Parr bomb using metallic sodium or the Wurzschnitt bomb using sodium peroxide can be used to decompose refractory substances.

## Contamination and Loss of Analyte

### Contamination

The introduction of foreign substances during wet digestion and sample processing may have serious consequences in elemental analysis, and sources of contamination and their likely impact should be anticipated in order to minimize the resulting errors.

Airborne contaminants, in the form of dust, corrosion products, and residues from paints, etc., can introduce many different elements into a sample, e.g., sodium, potassium, calcium, magnesium, and aluminum. This is a significant threat where open reaction vessels are used, and can be avoided by simple precautionary measures such as covering the dissolution vessel with a dish or beaker equipped with a side arm through which a stream of filtered air or gas is passed. More stringent measures include the use of a clean air room kept under positive pressure, with the reaction itself carried out under a laminar flow hood. Gaseous contaminants can be eliminated from the reaction vessel's air supply using appropriate adsorption filters.

Another significant source of contamination is the vessels used for reagent storage and decomposition. There is no such thing as a completely inert vessel, although the choice of an appropriate vessel that is the least reactive under the chosen storage or decomposition conditions is recommended. **Table 4** shows the elemental composition of four materials commonly used in decomposition and storage vessels, therefore revealing likely contaminants and the proportions in which they would be found. Dissolution vessels of Pyrex and quartz can introduce contamination from many elements, but ultrapure quartz can reduce the level of contamination from trace elements by up to 1000-fold as long as the surfaces are scrupulously cleaned before use. Polypropylene bottles, which are often used for the storage of acids, become brown after some time indicating the dissolution of the vessel's organic material. Even platinum vessels, which are required for fusions, contain small amounts of contaminating iron that can dissolve in contact with an acidic solution.

As well as the reaction and storage vessels, contamination may also arise from other stages of sample processing. Filter paper can contaminate samples with aluminum, barium, calcium, chromium, copper, iron, germanium, potassium, manganese, magnesium, sodium, antimony, gold, titanium, vanadium, zinc, zirconium, and rare earth metals. It is best to use filter media other than paper and carry out extensive preliminary washing with the operating solutions. Ultrapure water should be used as distilled water contains trace amounts of several



**Table 4** Average content (in  $\mu\text{g kg}^{-1}$ ) of current elements in the materials most commonly used for decomposition and storage vessels

Element	PTFE <sup>a</sup>	Ultrapure quartz	Quartz	Pyrex glass
Al		100	30 000	Main constituent
As		0.1	50	500–20 000
B		10	100	Main constituent
Ca		100	500–3000	1 000 000
Cd		0.1	10	1000
Co	1	1	1	100
Cr	20	2	5	3000
Cu	10	10	50	1000
Fe	10	100	1000	100 000
Hg	10 <sup>b</sup>	1	1	10–100
Mg		10	10	500 000
Mn		10	10	5000
Na	25 000	10	1000	Main constituent
Ni				2000
Sb	0.5	1	2	10 000
Si		Main constituent	Main constituent	Main constituent
Ti		100	1000	3000
Zn		50	50	3000

<sup>a</sup>Isostatic molded polytetrafluoroethylene.<sup>b</sup>Strongly dependent on cleaning conditions.From Buldini PL, Ricci L, and Sharma JL (2002) Recent applications of sample preparation techniques in food analysis. *Journal of Chromatography A* 975: 47–70.**Table 5** Illustrative contents of some trace elements in reagent grade and extra purity acids

Element	Content ( $\mu\text{g l}^{-1}$ )									
	HCl		HClO <sub>4</sub>		HF		HNO <sub>3</sub>		H <sub>2</sub> SO <sub>4</sub>	
	A <sub>max</sub>	B	A <sub>max</sub>	B	A <sub>max</sub>	B	A <sub>max</sub>	B	A <sub>max</sub>	B
Al		8		5	700		400	5	10	8
As	10	<1	50	5	50		10	1	500	<1
Ca		20		1000				7		10
Cd	10									
Co		<1		5				<1		<1
Cr		2								
Cu	500, 50	2	200	5	100	10	100	3, 10	500	3
Fe	100	7	500	100	1000	500	200	5	200	10, 3
Hg		<10						<10	<10	<10
Mg		4								
Mn		1	500	5		1		1, 5	1, 5	1
Ni	500	<1		2		20	100	<1	500	1, 7
Pb	50	<1	100	5		20		2		5
Zn	100	<1		40				<1		<1

A = analytical reagent, B = extra purity reagent ('electronic grade'), maximum content indicated by max; otherwise actual analysis.

inorganic compounds, resulting in contamination with copper, zinc, iron, nickel, aluminum, lead, and others.

Contaminants originating in the mineral acids used for digestion are also a significant source of error. **Table 5** shows the contaminants present in reagent grade and extra-pure mineral acids available from standard laboratory suppliers. Extra-pure acids have low levels of contaminants but are very expensive. It may be possible to eliminate many of the trace contaminants from reagent grade acids, or at

least reduce them to undetectable levels, by several rounds of distillation.

### Losses

During wet digestion, certain components of the sample can be lost, leading to the underestimation of particular elemental contents. Trace elements can be lost by adsorption to the vessel walls, volatilization, coprecipitation, and coextraction. Although the exact nature of adsorption losses is unclear, it may

be due to molecular and ionic interactions. Metal cations exchange with alkali or alkaline earth ions at the glass-solution interface and are replaced by  $\text{H}^+$  ions. Anion exchange may also occur, e.g.,  $\text{AuCl}_4^-$  for  $\equiv\text{SiOH}$  groups on the glass surface. Vigorous shaking of the solution may bring some of the absorbed species back into solution. Polyethylene and polypropylene also adsorb inorganic and molecular species but to a lesser extent and a mature acid strength of 10% is recommended in stored solutions. Solutions of  $0.1 \text{ mol l}^{-1}$  or stronger in mineral acid stored in polyethylene, Pyrex, or similar borosilicate glass bottles remain stable for weeks. Plastic bottles may be preferred because of reduced leaching of trace elements. Sample solutions of water should be acidified to prevent or minimize the adsorption of metals. Mercury concentrations in  $1 \text{ mol l}^{-1}$  nitric acid in borosilicate glass do not show any loss over a period of months. Loss of analyte also occurs when filtered through paper, sintered glass, or glass wool due to adsorption. Centrifugation is therefore preferable to filtration.

Volatile metals are less likely to be lost from basic solutions than from acidic ones. Metals with higher oxidation states tend to form volatile oxides and halides, e.g.,  $\text{OsO}_4$ ,  $\text{RuO}_4$ ,  $\text{Re}_2\text{O}_7$ ,  $\text{GeCl}_4$ ,  $\text{AsCl}_3$ ,  $\text{SbCl}_3$ ,  $\text{CrOCl}_2$ ,  $\text{SnCl}_4$ , and  $\text{TiF}_4$ . A few metals can be lost from solutions by volatilization at room temperature, particularly  $\text{OsO}_4$  and  $\text{RuO}_4$ . Losses associated with the reaction of samples with silica glass, porcelain, and platinum at higher temperatures may also take place.

### Matrix Effects and Modifiers

Many samples with inorganic or organic matrices pose serious problems in elemental analysis due to their composition and nature. It may therefore be necessary to add certain modifiers to the mixture to facilitate dissolution. In addition, dissolution assemblies may need to be modified. The use of hydrogen peroxide, a nonacidic modifier, has already been described. This can be used in conjunction with suitable acids for effective oxidation, and decomposition of organic matter usually proceeds without the danger of an explosion in a mixture of nitric and sulfuric acids at atmospheric pressure. One common modification of dissolution assemblies is the fitting of a condenser, to prevent the loss of volatile mercury, osmium, and ruthenium species. If excessive frothing occurs during wet digestion, a small amount of 2-octanol may be added or the preliminary treatment may be carried out in a beaker with glass beads to prevent bumping. With some organic materials, e.g., rubbers, coated fibers, and polymers, wet digestion in

sulfuric acid should not be carried out if the aim is to determine levels of arsenic. The ammonium salt of ethylenediaminetetraacetic acid is often used as a matrix modifier for the dissolution of sediments and other geological samples with hydrofluoric, sulfuric, and perchloric acids.

Digestion with a mixture of nitric, sulfuric, and perchloric acids is more efficient when a small amount of molybdenum(IV) catalyst is added. As soon as water and nitric acid have evaporated, oxidation proceeds vigorously with foaming and is completed in a few seconds. To avoid the loss of mercury from a digestion, volatile mercury compounds can be reduced to the metal with copper(I) and hydroxylammonium chloride at room temperature, while organic matter is degraded by potassium permanganate. The mercury can then be dissolved and the analysis completed. Nitrogen-containing compounds are commonly digested with sulfuric acid and potassium sulfate to increase the boiling point of the acid and thus the efficiency of the reaction. Proteins can be removed from organic samples by precipitation using trichloroacetic acid, tungstic acid with barium hydroxide, or zinc sulfate.

For the decomposition of inorganic samples by soda ash in platinum crucibles, the resistance of platinum is increased in an atmosphere of carbon dioxide. In a nitrogen atmosphere, platinum is corroded less by alkali metals. Peroxide bombs are good for wet ashing to destroy the organic matter in samples like coal. The PTFE acid digestion bomb prevents contamination and insures no losses by volatilization. Samples like coal can be analyzed in quantities of 50 mg or less.

In the fusion methods, cesium iodide is often added to the fusion mixture as a nonwetting agent to prevent the molten flux from adhering to the walls of the vessel, as well as to prevent incomplete transfer of the bead to the acid solution. In wet digestion of coal and fly ash using the Parr bomb, boric acid is added after digestion and the sample is heated for a further time on a water bath allowing the removal of unburned carbon.

No method of wet digestion is ideal, and both contamination and losses will occur to a greater or lesser degree whichever method is used. The task of the analytical chemist is to choose a procedure that will minimize interference from contaminants, reduce losses as much as possible, and therefore bring errors within acceptable limits. At all times, blank reagents and spiked recovery experiments are required to establish the degree of analyte contamination and loss from the dissolution process.

See also: **Sample Dissolution for Elemental Analysis:** Dry Ashing; Microwave Digestion.

## Further Reading

Hoenig M and de Kersabiec A-M (1996) Sample preparation steps for analysis by atomic spectroscopy methods: present status. *Spectrochimica Acta B* 51: 1297–1307.

Mader P and Curdova E (1997) Decomposition methods of biological materials for the determination of trace elements. *Chemické Listy* 91: 227–236.

Skoog DA and West DM (1976) *Fundamentals of Analytical Chemistry*, 3rd edn. New York: Holt, Rinehart & Winston.

Stoeppler M (ed.) (1997) *Sampling and Sample Preparation*. Berlin: Springer.

## Microwave Digestion

**C B Ojeda and F S Rojas**, Universidad de Malaga, Malaga, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The widespread use of microwaves for dissolution of samples is gradually replacing the conventional pressure reactors. This is particularly true in the determination of inorganic and organic analytes. Sample dissolution can be conducted in both closed and open systems.

A wet mineralization procedure using high pressure (and also high temperature), based on microwave decomposition with acid in hermetic Teflon containers, has recently been proposed. Sample heating is carried out in special microwave ovens that currently exist in the market. The speed of wet mineralization and the decomposition efficiency increase spectacularly when this type of equipment is used. For example, it has been observed that decompositions that require several hours at atmospheric pressure can be completed in a few minutes with the microwave technique. The other advantage is that it requires small quantities of the reactive substance, thus significantly reducing the magnitude of blanks.

Traditional methods of samples preparation do not give good reproducibility; however, the use of microwave irradiation rather than convection and/or conduction heating reduces errors in the sample preparation process. The method also addresses other parameters such as time required for analysis and sample contamination.

In 1975, microwaves were first used as rapid-heating sources for wet, open-vessel digestions. They were used to heat acids rapidly, in Erlenmeyer flasks, to digest biological matrices, and the conventional sample digestion times were reduced from 1–2 h to 5–15 min.

In the beginning, microwave sample preparation made use of common laboratory glassware and open Teflon vessels to digest matrices at the boiling point of the acids in commercial microwave ovens. In the 1980s, researchers began using specially designed

closed vessels for microwave digestions to achieve reaction temperatures above the atmospheric boiling points of the acids in order to increase the reaction rates and decrease reaction times.

## Theoretical Aspects

Microwaves are electromagnetic waves with frequencies in the range of 300–300 000 MHz. Microwave radiation neither possesses ionizing characteristics nor produces changes in the molecular structure; it can be used to measure the movement of molecules due to the migration of ions and the rotation of dipoles. Of the four frequencies most often used in industrial and scientific applications ( $915 \pm 25$ ,  $2450 \pm 13$ ,  $5800 \pm 75$ , and  $22\,125 \pm 125$  MHz), the one with a power range 600–700 W (domestic ovens) is 2450 MHz, which can provide, to the microwave cavity,  $\sim 43\,500$  cal in 5 min.

The warming of a sample through microwave action depends on the dissipation factor ( $\delta$ ), which is defined as the ratio of the dielectric loss factor of the sample ( $\epsilon''$ , which defines the efficiency of the dielectric medium to convert microwave energy into heat) and its dielectric constant ( $\epsilon'$ , a measure of the polarizability of the molecules in an electric field). Materials can reflect microwaves (e.g., metals), transmit them without being absorbed (e.g., transparent materials), or absorb them (e.g., dielectrics). All materials display these characteristics to varying degrees. The dissipation factor ( $\tau$ ) of the material is a good measure of its ability to heat up under microwaves: larger the dissipation factor, greater is the heat absorbed by the sample.

The dissipation of energy of the microwaves in a sample affects its transmission, which also depends on the frequency of the radiation (it is inversely proportional to the square root of the frequency).

Energy dissipation in a sample is caused by its warming and is basically produced by two mechanisms that generally operate simultaneously: ionic conduction and dipole rotation.

See also: **Sample Dissolution for Elemental Analysis:** Dry Ashing; Microwave Digestion.

## Further Reading

Hoenig M and de Kersabiec A-M (1996) Sample preparation steps for analysis by atomic spectroscopy methods: present status. *Spectrochimica Acta B* 51: 1297–1307.

Mader P and Curdova E (1997) Decomposition methods of biological materials for the determination of trace elements. *Chemické Listy* 91: 227–236.

Skoog DA and West DM (1976) *Fundamentals of Analytical Chemistry*, 3rd edn. New York: Holt, Rinehart & Winston.

Stoeppler M (ed.) (1997) *Sampling and Sample Preparation*. Berlin: Springer.

## Microwave Digestion

**C B Ojeda and F S Rojas**, Universidad de Malaga, Malaga, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The widespread use of microwaves for dissolution of samples is gradually replacing the conventional pressure reactors. This is particularly true in the determination of inorganic and organic analytes. Sample dissolution can be conducted in both closed and open systems.

A wet mineralization procedure using high pressure (and also high temperature), based on microwave decomposition with acid in hermetic Teflon containers, has recently been proposed. Sample heating is carried out in special microwave ovens that currently exist in the market. The speed of wet mineralization and the decomposition efficiency increase spectacularly when this type of equipment is used. For example, it has been observed that decompositions that require several hours at atmospheric pressure can be completed in a few minutes with the microwave technique. The other advantage is that it requires small quantities of the reactive substance, thus significantly reducing the magnitude of blanks.

Traditional methods of samples preparation do not give good reproducibility; however, the use of microwave irradiation rather than convection and/or conduction heating reduces errors in the sample preparation process. The method also addresses other parameters such as time required for analysis and sample contamination.

In 1975, microwaves were first used as rapid-heating sources for wet, open-vessel digestions. They were used to heat acids rapidly, in Erlenmeyer flasks, to digest biological matrices, and the conventional sample digestion times were reduced from 1–2 h to 5–15 min.

In the beginning, microwave sample preparation made use of common laboratory glassware and open Teflon vessels to digest matrices at the boiling point of the acids in commercial microwave ovens. In the 1980s, researchers began using specially designed

closed vessels for microwave digestions to achieve reaction temperatures above the atmospheric boiling points of the acids in order to increase the reaction rates and decrease reaction times.

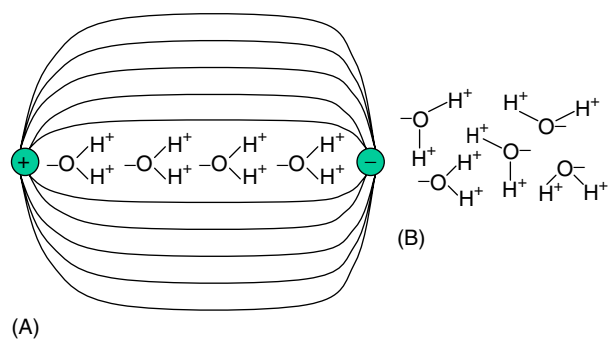
## Theoretical Aspects

Microwaves are electromagnetic waves with frequencies in the range of 300–300 000 MHz. Microwave radiation neither possesses ionizing characteristics nor produces changes in the molecular structure; it can be used to measure the movement of molecules due to the migration of ions and the rotation of dipoles. Of the four frequencies most often used in industrial and scientific applications ( $915 \pm 25$ ,  $2450 \pm 13$ ,  $5800 \pm 75$ , and  $22\,125 \pm 125$  MHz), the one with a power range 600–700 W (domestic ovens) is 2450 MHz, which can provide, to the microwave cavity,  $\sim 43\,500$  cal in 5 min.

The warming of a sample through microwave action depends on the dissipation factor ( $\delta$ ), which is defined as the ratio of the dielectric loss factor of the sample ( $\epsilon''$ , which defines the efficiency of the dielectric medium to convert microwave energy into heat) and its dielectric constant ( $\epsilon'$ , a measure of the polarizability of the molecules in an electric field). Materials can reflect microwaves (e.g., metals), transmit them without being absorbed (e.g., transparent materials), or absorb them (e.g., dielectrics). All materials display these characteristics to varying degrees. The dissipation factor ( $\tau$ ) of the material is a good measure of its ability to heat up under microwaves: larger the dissipation factor, greater is the heat absorbed by the sample.

The dissipation of energy of the microwaves in a sample affects its transmission, which also depends on the frequency of the radiation (it is inversely proportional to the square root of the frequency).

Energy dissipation in a sample is caused by its warming and is basically produced by two mechanisms that generally operate simultaneously: ionic conduction and dipole rotation.



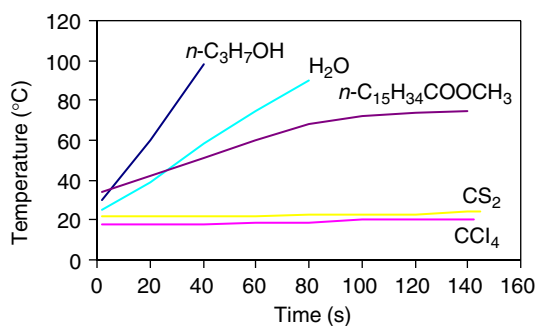
**Figure 1** (A) Aligned polarized molecules in the electromagnetic field. (B) Thermally induced disorder when the electromagnetic field is removed. (Adapted with permission from Kingston HM and Jassie LB (1998) *Introduction to Microwave Sample Preparation, Theory and Practice*, ACS Professional Reference Book Series, p. 11, Washington DC: American Chemical Society.)

*Ionic conduction* refers to the migration of ions in the applied electromagnetic field. Assume that the energy dissipation, in the form of heat, corresponds to  $iR^2$ , due to the resistance to the ionic flow. The ionic conduction depends on the nature and mobility on ions and on their concentration in the dissolution. In this mechanism, the dissipation of microwave radiation increases with increasing temperature, which controls the digestion processes in the final warming stage of sample dissolution.

*Dipole rotation* is used for molecules that possess a permanent or induced dipolar moment in an electric field. Upon applying microwave radiation to a sample that contains dipolar molecules, the electromagnetic field induces the organization of the system, which, upon increasing the temperature, is disordered and returns to the initial state, and thus thermal energy is released (Figure 1).

When a radiation of frequency 2450 MHz is employed, the orientation of the molecules is changed  $4.9 \times 10^9$  times per second. The result of this alignment-disorder process is a rapid self-warming of the sample.

The efficiency of the dipole rotation mechanism, with respect to warming of the samples, depends on the characteristics of the sample molecules, especially the dielectric easing time, which in turn depends on the temperature and the viscosity of the sample. The contribution of the dipole rotation to the warming of the samples is quantified through the dielectric easing time ( $\tau$ ): the time required for the molecules of the sample to return to its disordered state (up to 63%). This implies that the maximum energy conversion occurs when the frequency of the microwaves coincides with the inverse of the easing time; when the sample presents a value of  $1/\tau$  close to the microwave frequency, high dissipation of energy is produced, which implies increased yield of the warming processes. On the contrary, if  $1/\tau$  is very different



**Figure 2** Heating in the microwave oven. Adapted with permission from Empleo de los Hornos de Microondas en Química, Valencia 1990.

from the microwave frequency, it produces a low dissipation of energy.

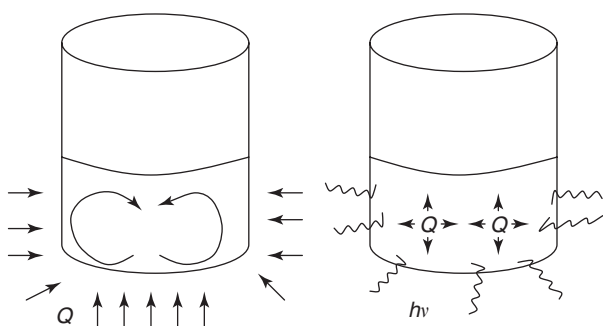
The value of  $1/\tau$  for water is greater than the corresponding value for the frequency of radiation (2450 MHz); it increases upon increasing the temperature: at temperatures of the order of 100°C, the ionic conduction prevails as compared to the dipolar rotation. Furthermore, for a given frequency, the warming of the samples is proportional to the applied power.

The customary procedure to establish the power of a microwave oven involves determining the temperature increase of a water pint (470 g) heated for 1 min; since  $\Delta E = cm\Delta T$ , where  $\Delta E$  is the absorbed energy,  $c$  the calorific capacity of water,  $m$  the water mass,  $\Delta T$  the temperature increase (in °F), so that  $P = \Delta T/18.2$ . Measuring  $\Delta T$  can be controlled by the power of the applied microwave radiation.

The clearly proportional (linear) behavior between warming and applied power is characteristic of very polar molecules; thus, for example, *n*-propanol behaves analogously to water, while carbon disulfide and carbon tetrachloride are practically not heated, and methyl palmitate presents a curve as a function of the time of exposure to the radiation that tends to be saturated, as can be seen in Figure 2.

From the above discussion, we can deduce that: as opposed to what occurs in a thermal stove, where the warming of the samples is produced by convection (transmission of thermal energy from the oven up to the walls of the container and from there to the molecules of the samples), in a microwave oven heat is generated by the polar molecules of the sample; the radiation results in an ordered state of the dipoles and is dissipated thermally in the sample itself, where heat is generated throughout, and there is no dissipation in the container on account of the slow convection processes that govern warming from an external source (Figure 3).

The practical conclusions of the processes induced in the dispersion of energy of the microwaves in an



**Figure 3** Comparison of heating in a thermal stove and a microwave oven. (Adapted with permission from Kingston HM and Jassie LB (1998) *Introduction to Microwave Sample Preparation, Theory and Practice*, ACS Professional Reference Book Series, pp. 16–17, Washington DC: American Chemical Society.)

oven are: very rapid warming and low energy consumption. Due to their rapidity and economy, the microwave systems can totally replace the traditional methods.

## Classification of Materials based on their Interaction with Microwaves

As mentioned earlier, materials are classified into three types, depending on their ability to absorb the microwave, allow the microwave to pass through, or reflect the microwaves.

Metals belong to the last group and therefore the use of metallic containers is not recommended. However, this characteristic is very important since it permits the design of a system of wave guides and protects users from radiation.

Transparent materials allow the microwaves to pass through without appreciable absorption. This property permits their use as insulating materials, as the heat loss of the samples is avoided and makes them particularly adapted for the construction of the vessels. Glass containers, ceramics, and some plastics are transparent. However, glass containers tend to be heated when they are exposed to high power for a long time.

Teflon is the most adequate material to build vessels: it is totally transparent to the microwave radiation, is resistant to all acids, has good thermal stability, and can sustain temperatures of up to 200°C, as opposed to polystyrene, which is only stable below 70°C. The only one application where it cannot be used is when concentrated phosphoric acid or sulfuric acid is used, since these acids have boiling points far above the fusion point of Teflon. In this case, quartz is the material that should be employed. Another material used often is polycarbonate, which resists the acids very well and has a fusion point of 135°C.

From the data for dissipation factors of different materials (at 3000 MHz, 25°C), it is deduced that compared to a value of  $\tan \delta = 157$  for water, quartz ( $\tan \delta = 0.06$ ) is the most transparent material. Teflon ( $\tan \delta = 0.15$ ), polyethylene ( $\tan \delta = 0.31$ ), polystyrene ( $\tan \delta = 0.33$ ), and some ceramics ( $\tan \delta \sim 0.55$ ) can be considered to be very transparent. Borosilicate glass ( $\tan \delta = 1.06$ ), phosphate glass ( $\tan \delta = 4.6$ ), Plexiglas ( $\tan \delta = 5.7$ ), poly(vinyl chloride) ( $\tan \delta = 5.5$ ), and nylon ( $\tan \delta = 12.8$ ) can also be considered as transparent, although they dissipate energy of the microwaves to a greater extent.

The materials comprising polar molecules, and especially liquids, absorb microwave radiation. Therefore, they can be heated easily in the ovens and hence are employed in reaction acceleration for dissolution and digestion of samples.

The different behaviors of the materials with respect to the microwaves must be taken into account during the manufacture of different components of the oven and for selection of vessels and acids for the dissolution of the samples.

## Components of Microwave Oven

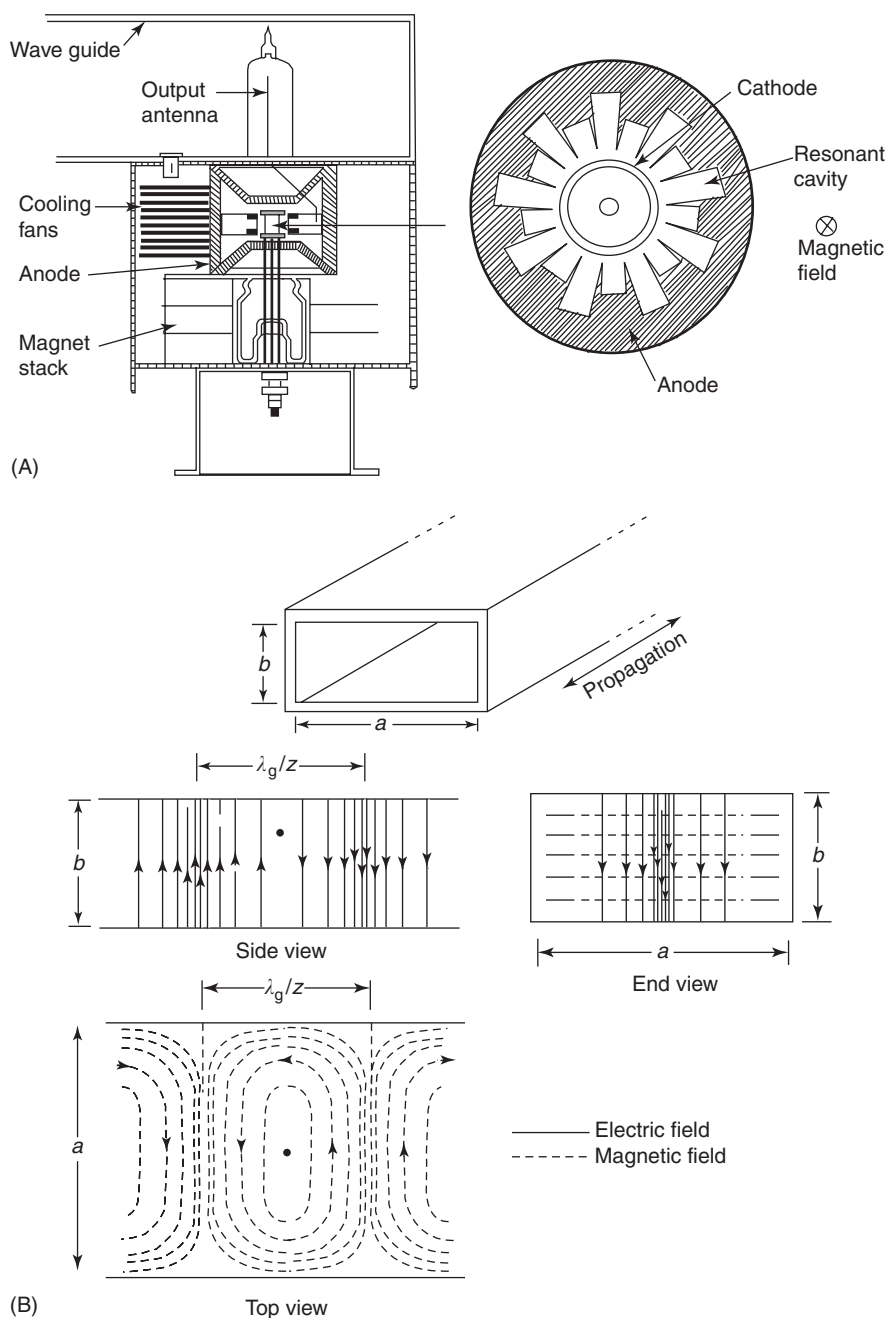
### Magnetron

A magnetron generates microwaves (Figure 4). It comprises a cylindrical diode working as a cathode, an anode, and a magnet that generates a magnetic field aligned with the cathode. Between the cathode and the anode are found a series of coupled resonating cavities that ensure that when a power of a few thousands of volts is reached in the diode, the electrons are freed, resounded by the influence of the magnetic field, and oscillate the magnetron, thus initiating the emission of an antenna placed under a pipe in which vacuum is created. The magnetrons used in microwave ovens tend to have an exit frequency of 2450 MHz and provide an electromagnetic energy of 600 W for an electrical entry power of 1200 W.

### Microwave Guide

The radiation is focused on the microwave cavity through a wave guide. Waveguides are rectangular, metallic conductors that permit the reflection of radiation and guide it toward an objective. In a guide of dimensions  $a$  and  $b$  (Figure 4), when the ratio  $a/b$  is between 2 and 2.5, there is an electromagnetic wave of wavelength  $< 2a$  can propagate. Figure 4B indicates the relative direction of the electric and magnetic fields, which constitute the radiation, in the wave guide. This spread pattern is unique for waves with  $2a > \lambda > a$ ; more complex spread patterns exist for other radiations.





**Figure 4** (A) Microwave oven components. (B) Microwave guide. (Adapted with permission from Kingston HM and Jassie LB (1998) *Introduction to Microwave Sample Preparation, Theory and Practice*, ACS Professional Reference Book Series, pp. 18–19, Washington DC: American Chemical Society.)

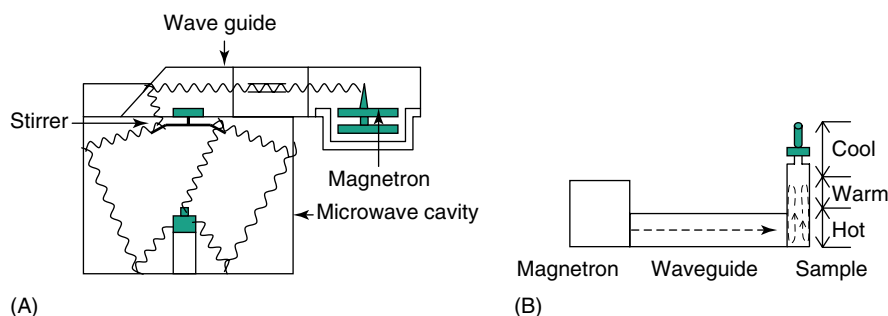
In practice, the optimum radiation for a guide is one whose wavelength corresponds to a value of  $1.3a$  and, as a rule, a good yield is obtained for radiation for  $\lambda$  between  $1.1a$  and  $1.7a$ .

In domestic microwaves, the radiation of the magnetron is focused toward the cavity, which results in preferential heating locations in the sample.

In microwave ovens manufactured for sample analytical studies, two strategies have been used

(Figure 5): (1) one similar to that of the domestic ovens; and (2) where radiation is focused on the sample (which is the basis for microdigestion systems).

The function of microwave guides is to lead the radiation from the magnetron up to the cavity. However, when it has worked for long periods of time, at high power, and with a small quantity of absorbent material, a large part of radiation



**Figure 5** (A) Schematic microwave oven of type 1. (B) Schematic microwave oven of type 2. (Adapted with permission from Kingston HM and Jassie LB (1998) *Introduction to Microwave Sample Preparation, Theory and Practice*, ACS Professional Reference Book Series, p. 18, Washington DC: American Chemical Society.)

that arrives at the cavity is reflected and it can be at this point that the magnetron initiates its destruction.

To avoid the concentration of reflected radiation, the microwave ovens designed for the sample digestion tend to carry a terminal circulator that employs ferrites and static magnetic fields to guide the radiation reflected toward a chamber where it is dissipated.

### Microwave Cavity

A microwave cavity is where the containers and the samples are introduced. It is here that the warming and attack processes develop.

A cavity is in the form of rectangular base prism with a volume of 10–20 l, except in the systems of type 2, where smaller cavities are used, designed to couple the digestors in their interior.

The distribution of radiation in the cavity is always uniform, creating greater density points of irradiation. It has been proved, for some ovens, that the positions next to the fund are more irradiated than the zones nearer to the center and the door; therefore, revolving or stirring plates are employed at the exit of the wave guide, which permit averaging of the exposure of the samples to the irradiation. Modifying the conditions of the samples or altering the direction of the microwaves is done in a uniform manner.

The walls of the oven are metallic and facilitate the reflection of radiation and, when interfaced with an absorbent material, dissipation results, and the provoking sample heats up.

In laboratory work, and as a rule, when samples are digested with acids, precautions have to be taken to avoid the attack of the walls. They have been used in systems for smoke extraction, closed containers, and special protection of the walls with a polymeric recovering.

There are two types of adapted microwave ovens, especially for the work in laboratory:

1. Equipment of the type 1: maintain the basic structure of the kitchen ovens; incorporate a system

of evacuation of smokes and a circulator to avoid the reflected microwaves; use digestors with a meticulous control of pressure and temperature during the digestion.

2. Equipment of type 2: very different from the conventional ovens; the cavity is designed especially to house the reactors; the reactors receive all the power from the magnetron through the wave guide; work at atmospheric pressure; possible to program all the stages, from the addition of acids and reagents to time and work power.

### Applications

The higher efficiency of microwave ovens to permit the warming of the samples allows drying, favors the digestion with acid, and accelerates chemical reactions that are controlled by a slow thermal stage.

It can be deduced that the chemical systems where a microwave oven can provide important advantages have the following characteristics: polar nature of the samples or the 'middle of reaction', processes that require warming, and those where reaction products are not decomposed with the increase in temperature, which leads to their degradation due to the lack of a strict control of the reaction or digestion conditions.

The last point above does not affect components that can be separated in gaseous form at temperatures between 100°C and 200°C, since they could be retained in the reaction system using properly sealed systems as is the case of the recovery of Hg and As in samples digested with acid.

### Acid Digestion of Solid Samples in Microwave Ovens

It is best to attack the samples with an acid or a mixture of acids. Dielectric liquids, when heated in contact with dielectric particles, induce the warming

of the superficial molecules of the sample. This creates a convection current that agitates and destroys the superficial caps of the samples. The destruction of this cap exposes a new part of the sample, which is now attacked, and in this way the digestion of the sample progresses quickly.

The “warming and attack” of the sample is not produced by the effect of an external source of heat but is generated by interactions of the reaction mixture with the microwave radiation.

The mixture of the acids and the samples acts as an excellent absorbent of radiation. Kingston and Jassie accomplished a systematic study of the behavior of the mineral acids, employing the type-1 system and closed digestors provided with pressure and temperature sensors. The data from this work are very interesting for studying the digestion conditions of numerous types of samples and they can be a guide to finding the best acid or mixture for use.

Normally, it is desirable that the acid form soluble salts with the metallic ion of interest, and because of this  $\text{HNO}_3$ ,  $\text{HCl}$ , and their mixtures are the most used acids.

### **$\text{HNO}_3$**

It is used essentially due to its characteristic as a strong oxidant acid and the fact that nitrates are very soluble salts, useful for the subsequent dissolution of the residues. It has a low boiling temperature ( $120^\circ\text{C}$ ), which limits its employment in open thermal systems. In closed vessels, it behaves ideally:

- 5 ml to 144 W:  $176^\circ\text{C}$  and 5 atm in 8 min (vessel of 100 ml); and
- 3 ml to 258 W:  $175^\circ\text{C}$  and 4.5 atm in 2.5 min.

This temperature ( $175^\circ\text{C}$ ),  $\sim 50^\circ\text{C}$  more than the boiling temperature, increases its oxidative power and favors the digestion of the samples.

### **$\text{HCl}$**

It is a strong acid and, although not an oxidant, has a great complexing power; therefore, it is very useful in the digestion of metals, giving easily soluble chlorides (7.3 g to 430 W:  $175^\circ\text{C}$  and 7 atm in 2 min). In these conditions  $\text{Cl}_2$  is produced, which increased the oxidant power of the acid.

These high pressures, which are used to attack with  $\text{HCl}$ , can be a drawback for use in closed containers; therefore, they should be used extremely cautiously to avoid losses or breaking of digestors.

### **Mixtures of $\text{HCl}$ and $\text{HNO}_3$**

Aqua regia is the most known and used mixture, and is a good oxidant due to the formation of nitrosyl chloride ( $\text{NOCl}$ ). Upon heating, dissociation of  $\text{NOCl}$  into  $\text{Cl}_2$  occurs, which continues attacking the samples and produces an effective digestion in closed containers. This effect is the cause of the oxidation of many materials, which occurs more effectively than with  $\text{HNO}_3$  or  $\text{HCl}$  alone (8 ml to 316 W:  $180^\circ\text{C}$  and 7 atm in 16 min).

Use of  $\text{HF}$ ,  $\text{H}_2\text{SO}_4$ ,  $\text{H}_3\text{PO}_4$ , and  $\text{HClO}_4$  has also been described and their use is less recommended.

### **$\text{HF}$**

$\text{HF}$  attacks the skin, and its steam can damage vision; therefore, one must be extremely cautious in its use. The use of gloves, plastic syringes, and Teflon containers, as well as safety eyewear and an extraction chamber with good shot is recommended.

### **$\text{H}_3\text{PO}_4$ and $\text{H}_2\text{SO}_4$**

These acids have high boiling points that can fuse the Teflon.

### **$\text{HClO}_4$**

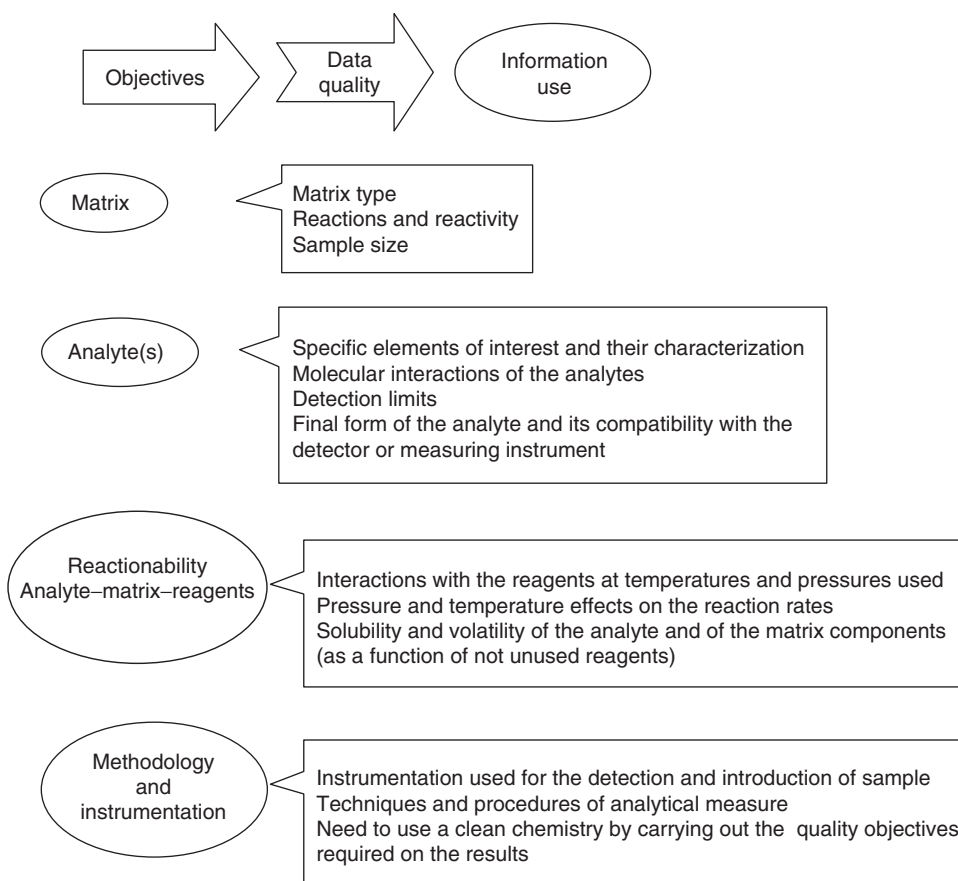
In the presence of organic samples,  $\text{HClO}_4$  could produce explosions.

## **Selection of a Microwave Sample Dilution Method**

Figure 6 gives a summary of factors that should be considered when selecting a microwave digestion method.

Acceptable errors in the results of analysis depend on the quality of data and so should in turn limit the acceptable errors of the process sampling, the blanks, etc. Often, for the same sample, not all the analytes possess the same priority and thus they do not require the same accuracy of the results.

The type of decomposition of the sample (reactive to use) depends on the nature of the matrix in reactors. Many matrices require  $\text{HF}$  to break the  $\text{Si-O}$  bonds of the silicates and  $\text{HNO}_3$  to oxidize the C. Organic samples (biological, botanical, polymeric pharmaceutical, etc.) and geological samples (soils, sediments, rocks, clays, etc.) are of interest for this type of treatment, since the C and the Si are transformed, by the acids mentioned, to volatile compounds ( $\text{CO}_2$  and  $\text{SiF}_4$ , respectively) and the

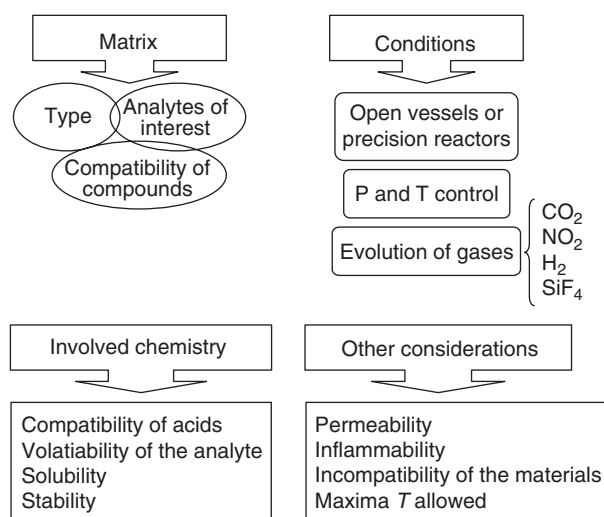


**Figure 6** Consideration of a selected method by microwave digestion. (Adapted with permission from Síntesis SA, *Toma y tratamiento de muestras*.)

transformation efficiency depends on the reaction temperatures.

With respect to the analyte, it must be taken into account that the matrix is soluble and stable in the digestion method selected. The concentration of the analyte in the sample should also be considered, since lesser the concentration greater is the need to obtain low quantification limits. The limits of detection and elemental interferences of the matrix can depend on the matrix-analyte-acid combination. The analysis technique should take extreme care to avoid interferences in the final detection. Thus, when inductively coupled plasma mass spectrometry is employed, the use of  $\text{H}_2\text{SO}_4$  and  $\text{HCl}$  can cause important spectral interferences that can completely annul the results.

Techniques that require the introduction of samples by nebulization, where physical interferences are avoided, for example, by variation of the viscosity of the sample are also interesting. On many occasions, to obtain a compatible dissolution, a subsequent treatment to the microwave digestion is required.



**Figure 7** Special considerations for the microwave oven sample preparation. (Adapted with permission from Síntesis SA, *Toma y tratamiento de muestras*.)

In summary, when microwave energy is employed in the sample digestion processes, the factors mentioned in **Figure 7** should be taken into account.

## Microwave Digestion of Various Sample Matrices

Methods based on microwaves have provided important improvements over the classical methods in terms of precision and efficiency.

### Organic Matrices

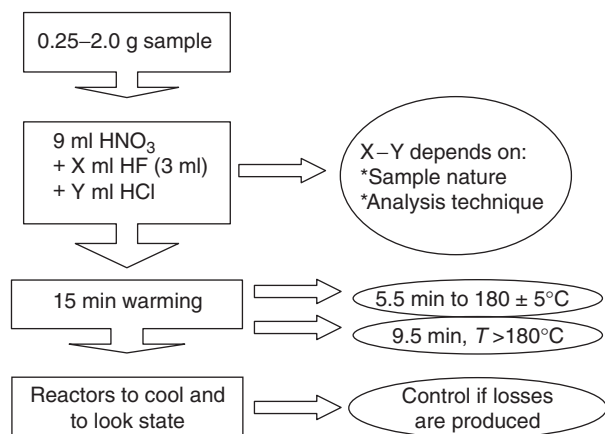
The decomposition of organic matrices in reactors to pressure by microwave is controlled temperature, that is, by adjusting the oxidation potential of  $\text{HNO}_3$ . The difficulty of mineralization of biological tissues depends on their composition.

In the case of botanical matrices, the addition of small quantities of HF is recommended to decompose the silicates present in small concentrations. The oxidation potential of  $\text{HNO}_3$  at  $180^\circ\text{C}$  is not sufficient for the decomposition of aromatic structures,

since these compounds or their degradation products cannot always be decomposed with  $\text{HNO}_3$  at temperatures  $\sim 300^\circ\text{C}$ . An incomplete decomposition of organic matter can give rise to serious interferences; for example, if electrochemical detection techniques are employed.

When biological samples are decomposed in open microwave systems, the use of  $\text{H}_2\text{SO}_4$  or  $\text{H}_2\text{SO}_4\text{--H}_2\text{O}_2$  is recommended; in the latter case, after burning of the organic matter by the addition of  $\text{H}_2\text{SO}_4$ , oxidation takes place completely due to  $\text{H}_2\text{O}_2$ , with the formation of  $\text{H}_2\text{SO}_5$ , which is decomposed at temperatures lower than the boiling point of  $\text{H}_2\text{SO}_4$ .

The dissolution of polymeric matrices is generally similar to that of other organic matrices ( $\text{HNO}_3$  is used, or in open systems  $\text{H}_2\text{SO}_4$  is added as an oxidizing agent). It must be taken into account that the hydrophobic nature of these samples causes a float on the surface of the reactors, which requires longer time for dissolution.



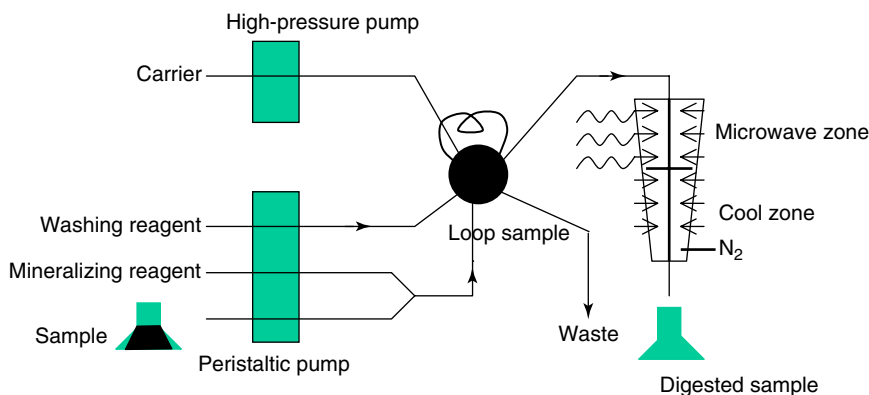
**Figure 8** Microwave assisted acid digestion of sediments, sludges, soils, and oils. (Adapted with permission from Síntesis SA, *Toma y tratamiento de muestras*.)

### Inorganic Matrices

Geological and metallurgical samples require very specific conditions for dissolution. Thus, if they contain silicates, HF should always be employed. A very general case of treatment would be the EPA-3052 method (Figure 8). This treatment results in appropriate dissolution of soils, rocks, sediments, etc.

### Environmental Matrices

One of the classic methods is 3051 developed by the EPA, and recently modified as method 3052. It is one of the most versatile methods and can be modified appropriately, depending on the characteristics of the matrices.



**Figure 9** Online schematic diagram of microwave digestion. (Adapted with permission from Síntesis SA, *Toma y tratamiento de muestras*.)

## Special Characteristics of Microwave Digestions for the Determination of Trace Analytes

The digestion methods in closed containers in reactors heated by microwaves provide an important reduction of the error risks introduced in the analysis when the samples are dissolved by traditional dry processes. One of the factors that contributes much to such reduction is the control of the blanks due to the following factors: the digestion takes place in

closed containers and pollution risks are reduced to a great extent; reagents of high degree of purity are used; the material of the reactors is very controlled and devoid of impurities; the time during which the sample is in touch with the reactor is very less compared to that in the classical methods; and risks are eliminated or reduced by the evaporation of volatile compounds.

Sample preparation using microwaves allows most reproducibility for treatment at ultratrace levels, in a given laboratory and between laboratories.

**Table 1** Official methods

<i>Analytes</i>	<i>Matrix</i>	<i>Type</i>	<i>Detection</i>	<i>Official method</i>
Water	Cheese	Microwave drying	Balance	AOAC 977.11: moisture in cheese
Water	Meat, poultry	Microwave drying	Balance	AOAC 985.14: moisture in meat and poultry products
Water	Tomato juice Tomato puree Tomato paste	Microwave drying	Balance	AOAC 985.26: solids (total) in processed tomato products
Water	Soil	Microwave drying	Balance	ASTM D4643-93: water (moisture) content of soil
Water	Wood	Microwave drying	Balance	ASTM E1358-90: moisture content of particulate wood fuels
Fat	Meat, poultry	Microwave drying	Balance	AOAC 985.15: fat (crude) in meat and poultry products
Ash	Carbon black	Microwave ashing	Balance	ASTM D1506-94b: carbon black – ash content
Al, Cd, Cr, Cu, Fe, Mn, Ni, Pb, Zn	Water	Microwave acid leach	ET-AAS, DCP-AES, FAAS, ICP-AES	ASTM D4309-91: total recoverable metals in water
As, Cd, Cu, Mg, Mn, Ni, Pb, Zn	Sediment, soil	Microwave acid leach	Not specified	ASTM D5258-92: acid extraction of elements from sediments
Ag, As, Ba, Be, Cd, Cr, Hg, Pb, Sb, Tl	Coal, coke, cement raw, feed materials, waste derived fuels	Microwave acid digestion	Not specified	ASTM D5513-94: industrial furnace feedstreams for trace element analysis
Pb	Paint	Microwave acid leach	Not specified	ASTM E1645-94: preparation of dried paint samples
Al, Ag, As, Ba, Be, Ca, Cd, Co, Cu, Cr, Fe, K, Mg, Mn, Mo, Na, Ni, Pb, Sb, Se, Tl, V, Zn	Water	Microwave acid leach	FAAS, ET-AAS, ICP-AES, ICP-MS	US-EPA 3015: aqueous samples and extracts
Ag, As, Ba, Be, Cd, Co, Cr, Cu, Mo, Ni, Pb, Sb, Se, Tl, V, Zn	Oil	Microwave digestion	FAAS, ICP-AES	US-EPA 3031: acid digestion of oils for metals analysis
Al, As, Ba, Be, Ca, Cd, Co, Cr, Cu, Fe, K, Mg, Mn, Mo, Na, Ni, Pb, Se, Tl, V, Zn	Sediment, sludge, soil	Microwave acid leach	FAAS, ET-AAS, ICP-AES	US-EPA 3050B: acid digestion of sediments sludges and soils

*Continued*



**Table 1** Continued

<i>Analytes</i>	<i>Matrix</i>	<i>Type</i>	<i>Detection</i>	<i>Official method</i>
Al, Ag, As, B, Ba, Be, Ca, Cd, Co, Cu, Cr, Fe, Hg, K, Mg, Mn, Mo, Na, Ni, Pb, Sb, Se, Sr, Ti, V, Zn	Oil, sediment, sludge, soil	Microwave acid leach	FAAS, ET-AAS, ICP-AES, ICP-MS	US-EPA 3051: microwave assisted acid digestion of sediments, sludges, soils and oils
Al, Ag, As, B, Ba, Be, Ca, Cd, Co, Cu, Cr, Fe, Hg, K, Mg, Mn, Mo, Na, Ni, Pb, Sb, Se, Sr, Ti, V, Zn	Fly ash, oil, sediment, sludge, soil	Microwave total digestion	FAAS, ET-AAS, ICP-AES, ICP-MS	US-EPA 3052: microwave assisted acid digestion of siliceous and organically based matrices
Al, Ag, As, B, Ba, Be, Ca, Cd, Co, Cu, Cr, Fe, Hg, K, Mg, Mn, Mo, Na, Ni, Pb, Sb, Se, Sr, Ti, V, Zn	Oil, sediment, sludge, soil	Microwave acid leach	FAAS, ET-AAS, ICP-AES, ICP-MS	US-EPA EMMC: microwave assisted acid extraction and dissolution of soils, sediments, sludges and oils
Al, As, Ba, Cd, Cr, Cu, Fe, Mn, Ni, Pb, Sb, Se, Zn	Domestic and industrial wastewater	Microwave acid leach	FAAS, ICP-AES, DCP-AES	US-EPA NPDES: closed vessel microwave digestion of wastewater samples for metals determination
Ag, Al, As, Au, Ba, Be, Bi, Ca, Cd, Ce, Co, Cr, Cu, Hg, Ir, K, Li, Mg, Mn, Mo, Na, Ni, Os, Pb, Pd, Pt, Rh, Sb, Se, Si, Sn, Sr, Th, Ti, Tl, V, Zn	Water	Microwave acid leach	FAAS, ET-AAS, CV-FAAS	Standard Methods 3030K: microwave assisted digestion
Al, As, Ba, Be, Ca, Cd, Co, Cr, Cu, Fe, K, Mg, Mn, Mo, Na, Ni, Os, Pb, Se, Th, V, Zn	Fish, shellfish	Microwave digestion	FAAS, ET-AAS, ICP-AES	Republic of China NIEA C303.01T: acid digestion of fish and shellfish
N	Milk, meat products, animal food, starch and starchy foods	Focused microwave digestion	Distillation or calorimetric	French Standard V 03-100: Kjeldahl nitrogen

## Safety of Microwave Digestion Processes

Digestion by microwaves is more secure than other types of treatments as long as it takes into account a series of factors such as avoiding the use of explosive mixtures.

The most important aspects that should be considered are: type and size of sample, types of acids, relationship of sample/solvent, primary and secondary chemical reactions, content or production of inflammable products, microwave power, ramp and time, instrument and conditions of the reactors, and interactions of the microwaves.

## Trends of Microwave Employment

In spite of the fact that the microwave instrumentation of pressure reactors is relatively recent, it is used

in numerous laboratories and at present its applications are impressive.

It can be said that microwaves constitute the standard treatment method for the analysis of ultratracés together with technical detection such as inductively coupled plasma–atomic emission spectrometry and inductively coupled plasma–mass spectrometry.

## Online Processes

The automation of systems for sample preparation is still far from being implemented in laboratories for routine analysis. Up until now, automation has been done only for preliminary chemistry when only relatively simple processes are required, such as filtration, dissolution in water, extraction systems. The methods of flow injection analysis have contributed to the automation processes.

Automation in the quantitative mineralization of complex samples is very difficult; furthermore, systems that require digestion in pressure reactors are practically impossible to automate. Automation has not been introduced in sample analyses that need complex treatments.

The employment of microwave open digestors is the easiest route to carry out online digestion processes by the wet route.

Different designs have been introduced, in both continuous flow and stop flow, which are being applied for the automation of the process.

Figure 9 shows a plan of the last proposed designs: with the aid of peristaltic pumps, the sample and the digestion reactive is impelled to the microwave unit. The digestion region is formed by a Teflon pipe that passes through the warming and cooling zones of the microwave oven.

There exist numerous bibliographic references that reveal the current interest in the development of on-line methods of sample treatment.

## General Information for the Treatment of Different Types of Samples

As microwave sample preparation has evolved, standard microwave procedures have been developed and approved by numerous standard methods organizations. Table 1 summarizes the different methods approved for either microwave drying or microwave acid dissolution by the Association of Official Analytical Chemistry (AOAC), American Society for Testing and Materials (ASTM), the United States Environmental Protection Agency (US-EPA), Standard Method, and French and Chinese national methods.

**See also:** **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Environmental Analysis. Sample Handling:** Comminution of Samples; Sample Preservation; Automated Sample Preparation; Robotics.

## Further Reading

- Agazzi A and Pirola C (2000) Fundamentals, methods and future trends of environmental microwave sample preparation. *Microchemical Journal* 67: 337–341.
- Cámara C, Fernández P, Martín-Esteban A, Pérez-Conde C, and Vidal M (2002) Preparación de la muestra para la determinación de analitos inorgánicos. In: Cámara C (ed.) *Toma y tratamiento de muestras*, pp. 202–218. Madrid: Síntesis.
- Chakraborty R, Das AK, Cervera ML, and de la Guardia M (1996) Literature study of microwave-assisted digestion using electrothermal atomic absorption spectrometry. *Fresenius' Journal of Analytical Chemistry* 355: 99–111.
- Cresser MS, Armstrong J, Cook J, *et al.* (1993) Atomic spectrometry update – environmental analysis. *Journal of Analytical Atomic Spectrometry* 8: 1R–77R.
- Kingston HM and Haswell SJ (1997) *Microwave-enhanced chemistry: Fundamentals, Sample Preparation and Applications*. ACS Professional Reference Book Series. Washington DC: American Chemical Society.
- Lamble KJ and Hill SJ (1998) Microwave digestion procedures for environmental matrices. *Analyst* 123: 103R–133R.
- Sengutta U (2003) Sample preparation with microwave technology. *Labor Praxis* 27: 34–37.
- Smith FE and Arsenault EA (1996) Microwave-assisted sample preparation in analytical chemistry. *Talanta* 43: 1207–1268.
- Yue Z, Yong Y, Huiqi H, and Changlin L (1996) Microwave digestion method in environmental analysis. *Journal of Environmental Sciences* 8: 407–413.

# SAMPLE HANDLING

Contents

**Comminution of Samples**

**Sample Preservation**

**Automated Sample Preparation**

**Robotics**

## Comminution of Samples

**B Markert**, International Graduate School (IHI), Zittau, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The analytical steps of sample preparation have varying significance relative to their effect on the

Automation in the quantitative mineralization of complex samples is very difficult; furthermore, systems that require digestion in pressure reactors are practically impossible to automate. Automation has not been introduced in sample analyses that need complex treatments.

The employment of microwave open digestors is the easiest route to carry out online digestion processes by the wet route.

Different designs have been introduced, in both continuous flow and stop flow, which are being applied for the automation of the process.

Figure 9 shows a plan of the last proposed designs: with the aid of peristaltic pumps, the sample and the digestion reactive is impelled to the microwave unit. The digestion region is formed by a Teflon pipe that passes through the warming and cooling zones of the microwave oven.

There exist numerous bibliographic references that reveal the current interest in the development of on-line methods of sample treatment.

## General Information for the Treatment of Different Types of Samples

As microwave sample preparation has evolved, standard microwave procedures have been developed and approved by numerous standard methods organizations. Table 1 summarizes the different methods approved for either microwave drying or microwave acid dissolution by the Association of Official Analytical Chemistry (AOAC), American Society for Testing and Materials (ASTM), the United States Environmental Protection Agency (US-EPA), Standard Method, and French and Chinese national methods.

**See also:** **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Environmental Analysis. Sample Handling:** Comminution of Samples; Sample Preservation; Automated Sample Preparation; Robotics.

## Further Reading

- Agazzi A and Pirola C (2000) Fundamentals, methods and future trends of environmental microwave sample preparation. *Microchemical Journal* 67: 337–341.
- Cámara C, Fernández P, Martín-Esteban A, Pérez-Conde C, and Vidal M (2002) Preparación de la muestra para la determinación de analitos inorgánicos. In: Cámara C (ed.) *Toma y tratamiento de muestras*, pp. 202–218. Madrid: Síntesis.
- Chakraborty R, Das AK, Cervera ML, and de la Guardia M (1996) Literature study of microwave-assisted digestion using electrothermal atomic absorption spectrometry. *Fresenius' Journal of Analytical Chemistry* 355: 99–111.
- Cresser MS, Armstrong J, Cook J, *et al.* (1993) Atomic spectrometry update – environmental analysis. *Journal of Analytical Atomic Spectrometry* 8: 1R–77R.
- Kingston HM and Haswell SJ (1997) *Microwave-enhanced chemistry: Fundamentals, Sample Preparation and Applications*. ACS Professional Reference Book Series. Washington DC: American Chemical Society.
- Lamble KJ and Hill SJ (1998) Microwave digestion procedures for environmental matrices. *Analyst* 123: 103R–133R.
- Sengutta U (2003) Sample preparation with microwave technology. *Labor Praxis* 27: 34–37.
- Smith FE and Arsenault EA (1996) Microwave-assisted sample preparation in analytical chemistry. *Talanta* 43: 1207–1268.
- Yue Z, Yong Y, Huiqi H, and Changlin L (1996) Microwave digestion method in environmental analysis. *Journal of Environmental Sciences* 8: 407–413.

# SAMPLE HANDLING

Contents

**Comminution of Samples**

**Sample Preservation**

**Automated Sample Preparation**

**Robotics**

## Comminution of Samples

**B Markert**, International Graduate School (IHI), Zittau, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The analytical steps of sample preparation have varying significance relative to their effect on the

precision and accuracy of the results of the analysis. Comminution of sample materials is one of the most essential, yet the most error-prone steps in the course of all chemical laboratory analysis tasks. Comminution usually refers to the conversion of solid samples with large particles mostly into powdery constituent substances of smaller particle size. The term is also infrequently used in conjunction with operations designed to convert liquids into droplets or gases into bubbles, the details of which will not be discussed further.

The aggregate state of the material to be homogenized does not undergo any change during these processes, whereas the physical and chemical characteristics of the sample may well change.

When comminuting crystalline substances, triboluminescence – a special form of luminescence – may be observed, so that the process can be made evident by way of adding triboluminescent substances.

The reasons for homogenizing sample material prior to performing the actual chemical/physical analysis are based on a great variety of requirements. The idea is usually to obtain a representative, i.e., evenly powdered sample for subsequent chemical/physical analysis. One of the problems encountered in sampling, especially in the field of ecological analysis, is the huge parent mass – anything between kilograms and tons – in the form of integral parts of a refuse dump, rock, debris of a specific stratum, etc. From this, a sample of only a few hundred milligrams may go into the measuring equipment for the actual analyses. In other words, the original sample is subjected to considerable decrease in bulk on its way to representing a proper analytical sample. It is important, however, to ensure that the make-up of this analytical sample represents the average chemical composition of the original sample.

Another major problem frequently encountered, specifically in the trace and ultratrace domains of ecological analysis, is the risk of contamination and/or volatilization of individual components. The risk of contamination is always present if the grinding elements themselves contain a high percentage of the substance to be determined (for example, iron and chromium in steel mills), or when certain readily volatilized combinations escape from the sample as a result of overheating.

The comminution process to be used normally depends on the following parameters:

- Total quantity and number of samples of the material to be homogenized.
- Initial particle size of the original sample.

- Ultimate fineness of the sample after comminution.
- Chemical/physical properties of the original sample and of the grinding elements (contamination/volatization).
- Hardness of the material to be homogenized.

There are three general comminution regimes:

- Coarse: grinding down to a particle size of  $\sim 5$  mm.
- Fine: grinding down to a particle size of  $63\ \mu\text{m}$ .
- Ultrafine: grinding down to a particle size of  $< 63\ \mu\text{m}$ .

## Theoretical Aspects

The basic operation of dividing solids into fractions by application of mechanical forces results in an increase of surface area, the magnitude of which depends primarily on the particle size, as is demonstrated in Table 1.

The term particle (or grain) size refers to the structural make-up of such substances as granulates, powders, dusts, granular mixes, and suspensions. Knowledge of the particle size, in conjunction with the comminution process, determines such details as grinding efficiency and ultimate product fineness. To establish particle sizes and their distribution within powdered systems, the user can have recourse to a number of different measuring processes designed to indicate, with appropriate particle definition, details of the probable equivalent diameter of a particle.

It follows that individual measuring methods depend upon particle definition. Particle definitions might include such details as number, length, surface area, mass, and volume, as a result of which one has the choice of using various counting methods – microscopy, image analyzers, laser diffraction, as well as sieving and air classification methods.

The details of a particle-size analysis are presented in tabular and diagrammatic form covering quantitative

**Table 1** Total increase of surface area upon disintegration of a cube of  $1\text{ cm}^3$  initial volume

Edge length	Number of cubes	Total surface area
1 cm	1	$6\text{ cm}^2$
1 mm	$10^3$	$60\text{ cm}^2$
0.1 mm	$10^6$	$600\text{ cm}^2$
0.01 mm	$10^9$	$6000\text{ cm}^2$
$1\ \mu\text{m}$	$10^{12}$	$6\text{ m}^2$
$0.1\ \mu\text{m}$	$10^{15}$	$60\text{ m}^2$
$0.01\ \mu\text{m}$	$10^{18}$	$600\text{ m}^2$
1 nm	$10^{21}$	$6000\text{ m}^2$

mass fractions, and cumulative and/or frequency density as a function of the equivalent diameter.

The following classification into millimeter ranges can be applied as a quantitative definition of particle/grain sizes:

Giganto grained > 300 mm  
 Giant grained 30–300 mm  
 Large grained 10–30 mm  
 Coarse grained 3–10 mm  
 Medium grained 1–3 mm  
 Small grained 0.3–1 mm  
 Fine grained 0.1–0.3 mm  
 Dense grained 0.03–0.1 mm  
 Microcrystalline 1–30  $\mu\text{m}$   
 Cryptocrystalline 0.1–1  $\mu\text{m}$   
 X-ray crystalline < 0.1  $\mu\text{m}$

An important factor governing the efficiency of any comminution process is the fracturing behavior of a sample. **Figure 1** demonstrates varying behaviors upon application of different stress-strain mechanisms:

1. *Fracturing behavior between two surfaces (Figure 1A).* The sample is wedged between two solid surfaces, which could be the surfaces of grinding assembly components or those of adjacent particles, and exposed to pressure load. Comminution here is, in particular, the result of pressure and frictional forces.

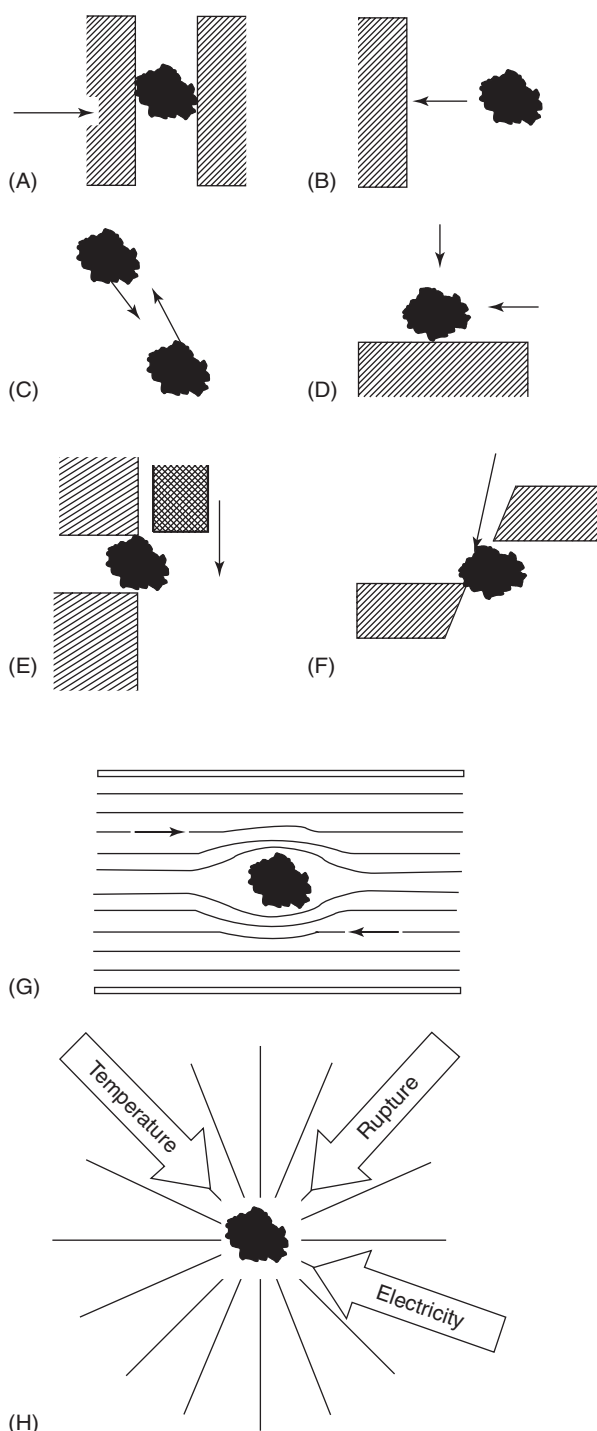
2. *Fracturing behavior upon contact with a single surface (Figures 1B and 1C).* The pressure effect here is produced upon contact with a single surface, which could be the solid surface of a grinding assembly component or of another particle within the sample.

3. *Fracturing behavior upon exposure to pressure and friction (Figure 1D).* Here the loading effect between two solid surfaces is in the form of a vertical force in combination with a horizontal, centric, or eccentric rotation, to bring about comminution.

4. *Fracturing behavior upon exposure to a shearing force (Figure 1E).* The loading effect is produced between two solid surfaces in the form of a shearing effect, i.e., comminution is brought about by way of impact and/or rebound effects generated between two counteracting surfaces or one moving and one static surface.

5. *Fracturing behavior upon exposure to a cutting effect (Figure 1F).* The loading effect is produced by two sharp-edged surfaces with their edges counteracting diametrically.

6. *Fracturing behavior upon exposure to the ambient medium (Figure 1G).* A loading effect due to the



**Figure 1** Fracturing behavior of sample constituents upon application of different stress-strain mechanisms. For details see text. (Modified from Pitsch H (1993) *Comminution*. Retsch Company, Rheinische Strasse 36, PO Box 1554, Haan W-5657, Germany; with permission.)

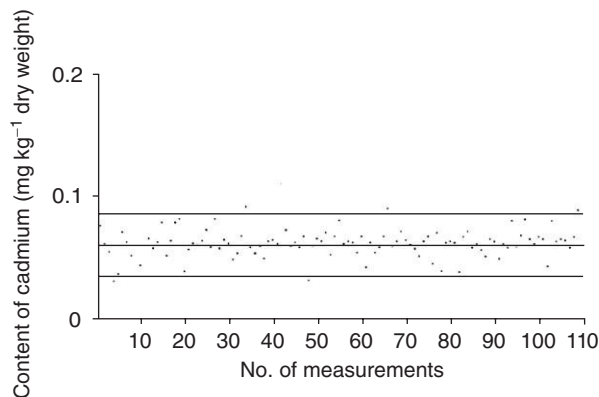
ambient or surrounding medium at a high shearing gradient can only be accomplished with low-strength materials, e.g., agglomerations or those with low degrees of hardness.

7. *Fracturing behavior upon exposure to nonmechanical energy* (**Figure 1H**). This loading effect on solid particles is caused by application of thermal loads without additional mechanical energy, i.e., by overheating, embrittling, or electric discharge.

Another important criterion for the quality of a comminution process is the subsequent homogeneity of the sample material, i.e., the distribution of the substance to be analyzed within the sample. A high degree of homogeneity may be obtained by blending the sample material, after comminution, in a suitable machine. This method counteracts the segregation tendencies of the sample material that result from variations in the distribution of particle sizes and is of particular importance whenever homogenized samples are bottled to serve as, for example, reference material (i.e., substances with a defined chemical composition) for quality control in chemical analysis. It must be ensured that each bottle of the total charge has exactly the same content ('between-bottle homogeneity') and that the contents of the single bottles represent the same average chemical composition ('within-bottle homogeneity'). To this end the material is analyzed for the component, with an appropriate number of repeat analyses. Once it has been established that the standard deviation is within tolerable limits, the sample material can be looked upon as being homogeneous. **Figure 2** shows an example of cadmium determination in white cabbage for the preparation of reference material for vegetable matrices.

## Contamination

Every comminution process involving solid surfaces in combination with kinetic energy input is likely to



**Figure 2** Quantitative determination of cadmium by solid sampling atomic spectrometry in a new candidate reference material (BCR-white cabbage; intake: 0.2–5 mg per measurement). Horizontal lines indicate tolerance limits. (From Markert, 1996.)

be accompanied by corresponding abrasion of the grinding assembly components. The resulting abrasive effect tends to increase grinding tool wear and to decrease tool life and grinding efficiency. More critically, the abrasion fines contaminate the sample stock during the grinding process. To minimize this risk of contamination, it is recommended practice to ensure that the abrasion resistance of the grinding assembly components is always higher than the hardness of the sample stock.

Hardness refers to the resistance offered by a solid body, i.e., the sample (glass, ceramics, minerals, etc.), when exposed to mechanical stress loads. High hardness ratings are typical for minerals, some of which are therefore well suited for use in lubricants and grinding and polishing compounds. Friedrich Mohs, an Austrian mineralogist, compiled an empirical scale of scratch hardness ratings in 1812 that is still applicable and is used to determine the hardness of minerals (see **Table 2**). Mohs hardness ratings are established by using the next higher mineral on the Mohs Hardness Scale to scratch-test a lower mineral on the scale: e.g., quartz (7) leaves a scratch on feldspar (6) but not on topaz (8).

The risk of contamination by individual process abrasives is particularly high in trace and ultratrace element analysis and is, therefore, primarily responsible for falsification of analytical results. Remembering that abrasion fines originating from grinding assembly components are always likely to contaminate sample material, the risk is best avoided by making sure that the component to be analyzed is not present in the grinding tools. The major components of individual grinding elements are listed in **Table 3**; the values necessarily represent only approximations and values are likely to vary with source and/or manufacturing process. Plastic material has recently been introduced for use in comminution equipment. The German government has launched an ecological sample base project in the course of which a mill with polytetrafluoroethylene rods (to serve as grinding elements) was conceived which produced

**Table 2** Mohs hardness scale

Hardness	Mineral
1	Talc
2	Gypsum
3	Calcite
4	Fluorite
5	Apatite
6	Feldspar
7	Quartz
8	Topaz
9	Corundum
10	Diamond



**Table 3** Approximate standard values of a selection of materials used in grinding assembly components<sup>a</sup>

Material	Composition (%)	Material	Composition (%)
<i>Agate (origin: Brazil or Uruguay)</i>		<i>Boron carbide</i>	
SiO <sub>2</sub>	99.91	B	77–79
Al <sub>2</sub> O <sub>3</sub>	0.02	C	21–23
Na <sub>2</sub> O	0.02	Si	0.1
Fe <sub>2</sub> O <sub>3</sub>	0.01	Fe	0.1
K <sub>2</sub> O	0.01	<i>Tungsten carbide</i>	
MnO	0.01	WC	94
CaO	0.01	Co	6
MgO	0.01	<i>Titanium</i>	
<i>Hard porcelain</i>		Ti	99.437
SiO <sub>2</sub>	61	Fe	0.1
Al <sub>2</sub> O <sub>3</sub>	34	O <sub>2</sub>	0.3
K <sub>2</sub> O	3	N <sub>2</sub>	0.07
Residual oxides	2	C	0.08
<i>Sintered corundum I</i>		H <sub>2</sub>	0.013
Al <sub>2</sub> O <sub>3</sub>	99.7	<i>Chrome steel</i>	
<i>Sintered corundum II</i>		Fe	84
Al <sub>2</sub> O <sub>3</sub>	83	Cr	12
SiO <sub>2</sub>	16.5	C	1.65
Residual oxides	0.5	Si	0.3
<i>Carbon steel</i>		Mn	0.3
Fe	99.11	P	0.03
C	0.15	S	0.03
Si	0.25	Mo	0.6
Mn	0.40	V	0.3
P	0.045	W	0.5
S	0.045	<i>Specialty steel</i>	
<i>Zirconium oxide</i>		Fe	84
ZrO <sub>2</sub>	97	Cr	13
MgO	~ 1.5	Mn	1
CaO	~ 1.5	Si	1
Fe <sub>2</sub> O <sub>3</sub>	~ 0.03	C	0.45
SiO <sub>2</sub>	~ 0.1		

<sup>a</sup> Allowance must be made for variation due to origin of material and manufacturing process.

Reproduced with permission from Pitsch H (1993) Comminution. Retsch Company, Rheinische Strasse 36, PO Box 1554, Haan W-5657, Germany.

noncontaminated homogenized ecological samples under liquid-nitrogen conditions.

## Methods of Sample Processing

Grinding equipment used for comminution operations in chemical analysis varies in design and layout with application requirements, i.e., sample make-up, desired degree of particle size decrease, etc. The selection of the grinding kit to be used depends on

the problem and the requirements with respect to the desired sample fineness and freedom from contamination. In view of the extremely wide range of grinding kits available on the market, it is best to concentrate on their major objectives.

In general, the complexity of the enormous variety of comminution problems and requirements rules out any hope for a single, ideal comminution process. It happens quite frequently that processing of one sample involves the combined use of different grinding processes; for example, coarse grinding followed by fine grinding.

Given the multitude of existing comminution processes, we shall discuss two coarse, two fine, and three ultrafine grinding methods.

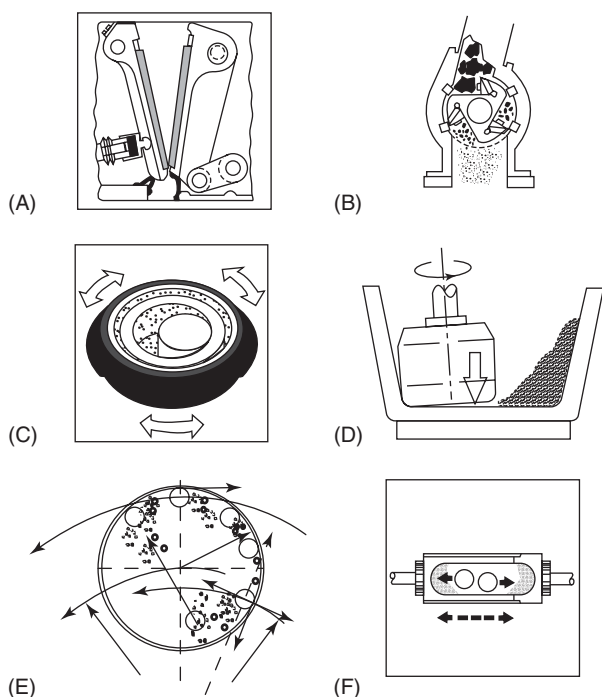
### Coarse Grinding (Down to 5 mm Particle Size)

There are many cases where the initial particle size of the sample requires preliminary reduction. Classical machines for this purpose are jaw crushers, normally used with samples of medium to extreme hardness as well as brittle to hard-tenacious samples, and cutting mills, used with samples of soft and tough make-up.

**Jaw crushers** These are used to break up substances with medium to extreme hardness ratings as well as brittle to tough materials, such as bakelite, bauxite, concrete, dolomite, ores, feldspar, ferroalloys, granite, glass, limestone, gravel, coal, coke, corundum, synthetic resins, quartz, salts, slag, silicates, sintered products, and many other substances.

Jaw crushers are normally upright floor models used for preliminary purposes in laboratories and at production plant level. They are designed for continuous operation and are therefore suitable for inclusion in fully automated and semiautomatic processing systems in production plants and laboratories. The efficiency and performance data (throughput rate, feed particle sizes, and ultimate fineness) of jaw crushers depend upon the disintegrating behavior and hardness rating of the material to be handled and on the size of the machine. **Figure 3A** shows that the grinding chamber is funnel shaped and tapers toward the adjustable outlet gap. These crushers incorporate a static and an eccentrically driven crushing jaw between which the breaking process takes place. The ellipsoidal movement of the driven crushing jaw provides for a continuous material flow toward the crusher outlet, from where it passes into a receptacle once broken down to the desired size.

**Cutting mills** These are designed to disintegrate bulky, soft, medium-hard, as well as fibrous and



**Figure 3** Operating principles of different comminution processes. See text for details. (Modified from Pitsch H (1993) Comminution. Retsch Company, Rheinische Strasse 36, PO Box 1554, Haan W-5657, Germany; with permission.)

cellulosic materials, and are used in agricultural research institutes and the plastics and pulp industries. Cutting mills are suitable for use with foliage, fibers, floor covering material, feedstuff, spices and herbs, rubber, grasses, hay, hops, wood, cardboard, plastics, leather, malt, Indian corn, paper, straw, tobacco, peat, paste products, roots, and twigs. **Figure 3B** shows that comminution of the material is brought about by a combined cutting and shearing effect. The material remains in the grinding chamber for no longer than is required to reach the desired fineness. The inlet zone is provided with a wooden tamper to facilitate handling of bulky or extremely light material. The attainable ultimate fineness depends on the size of interchangeable sieves located in the discharge zone.

#### Fine Grinding (Particle Size 5 mm to 63 $\mu\text{m}$ )

**Vibratory disk mills** These are designed for rapid, efficient, and lossless fine-grinding of medium-hard, hard-brittle, as well as tough materials in dry and wet conditions. They are specifically suitable for use with such substances as soil samples, ores, coal, coke, corundum, metal oxides, minerals, dry plant samples, slag, silicates, cement, cement clinker. **Figure 3C** shows that the sample material is ground down by

combined impact and friction effect generated between the grinding media and the jar wall surface. The grinding jar is fitted with the grinding tools (disk and/or rings) and loaded with the sample, after which the jar is closed and locked in the clamping fixture. When the machine is switched on, the mounting pad begins to oscillate, generating high dynamic mass forces during the grinding process.

**Mortar mills** These are designed for work with soft, hard, and brittle materials with hardness ratings of up to 8.5 Mohs. Mortar mills are particularly suitable for wet or dry grinding and homogenizing of such substances as ashes, chemicals, drugs, enamels, bauxite, dolomite, limestone, dry clay, kaoline, ores, coloring substances, rocks and debris, glass, synthetic resins, pharmaceutical products, salts, sand, chamotte, fireclay, slags, silicates, cement clinker, and similar materials. The ultimate fineness of samples produced in mortar mills depends on both the grinding time and the fracturing behavior of the substance to be handled. **Figure 3D** shows that the sample is loaded into the mortar located on a mounting pad. When the cover is closed and the machine switched on, the pestle follows the rotational movement of the mortar body, thus grinding the sample between the lower edge of the pestle and the rounded mortar wall.

#### Ultrafine Grinding (Particle Sizes <63 $\mu\text{m}$ )

**Planetary ball mills** These are designed to comminute soft and fibrous, medium to extremely hard, and brittle materials such as ashes, bauxite, soil samples, chemicals, chamotte, fireclay, diamonds, drugs, fertilizers, enamels, earth, ores, coloring substances, rock and debris, glass, domestic and industrial waste, wood chippings, cocoa beans, sewage sludge, coal, coke, plant constituents, pigments, quartz, salts, sands, slag, silicates, cement, cement clinker, and sugar. Planetary ball mills incorporate a high-speed transmission system generating extreme centrifugal acceleration rates, thus providing for ultimate finenesses well into the submicrometer particle size range. A multitude of grinding tools made of a great variety of materials provides for the processing of trace analysis samples, often without contamination. **Figure 3E** shows four grinding stations located on a horizontally rotating platform and intercoupled to move both around their own central axes and in counter-rotation around the axis of their common platform. The resultant effect is a continuously alternating superimposition of centrifugal forces sending the grinding balls along the jar wall to traverse the jar center line and back again from the opposite

side. The dynamics involved produce a violent rebound effect in combination with high friction and impact forces, providing for excellent comminution and homogenization.

**Vibratory mixer mills** These mills comminute and homogenize soft, fibrous, hard, and brittle substances in both dry and wet conditions. They are particularly suitable for processing basalt, building material, dolomite, earth, cereals, organic tissue samples, hair, cocoa, rubber, ceramics, bones, synthetic resins, alloys, minerals, plant constituents, salts, tablets, pills, peat, and wool. **Figure 3F** shows how grinding balls comminute the sample material. The grinding cylinder, holding the sample and loaded with one or two balls, is locked in the mounting bracket and energized into horizontal, radial oscillations, thus bouncing the balls against the rounded cylinder ends so as to disintegrate the sample material. The degree of reduction and the ultimate fineness attainable are governed by the mass of the grinding balls.

**Cold grinding** The terms cold grinding, freeze grinding, or cryogenic grinding refer to comminution methods employing mechanical and thermal energies to break down solids. This method is often the only means of obtaining sufficient comminution results with acceptable homogenizing efficiency, which is particularly important when handling materials with low melting points, high grease and/or moisture contents, as well as resilient or tough consistencies.

The disintegrating behavior of substances with any such characteristics can be substantially improved by addition of carbon dioxide 'ice' or by embrittlement with liquid nitrogen. The use of direct or indirect cooling, in combination with mechanical comminution techniques, facilitates adequate processing of many types and grades of plastic, rubber, organic tissue, pharmaceutical products, spice, herbs, and food samples.

The greatest care must be taken when handling coolants, especially liquid nitrogen, and the use of protective gloves, safety goggles, etc., in keeping with the safety instructions of grinding equipment manufacturers is most important.

### General Safety Aspects

Laboratory staff are always exposed to risk when undertaking their duties, comminution procedures included. Manufacturers have made great efforts to make their equipment foolproof against handling errors: mills, for example, cannot be unloaded until the machine has come to a dead stop.

Comminution of explosive materials represents a particularly high hazard and must, therefore, be left to qualified laboratory staff with special training in this area. Included in explosion hazard is the risk of dust explosions, always likely to be present in grinding operations involving high degrees of sample comminution.

*See also:* **Particle Size Analysis. Quality Assurance:** Reference Materials; Production of Reference Materials.

### Further Reading

- German Federal Ministry of Research and Technology (1988) *Umweltprobenbank*. Berlin: Springer (in German).
- Markert B (1996) *Instrumental Element and Multielement Analysis of Plant Samples/Methods and Applications*. Chichester: Wiley.
- Schladot JD, Backhaus F, and Reuter U (1985) Beiträge zur Umweltprobenbank, I. Studie zur Probenhomogenisierung bei tiefen Temperaturen unter Berücksichtigung der für die Umweltprobenbank notwendigen Parameter. Forschungszentrum Jülich, 'Jul. Spez. 330', 25pp.

## Sample Preservation

**R D McDowall**, McDowall Consulting, Bromley, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Know the Stability of an Analyte

The purpose of this article is to discuss the importance of knowing:

- The stability of the analyte in the sample matrix.

- Any measures that can be taken to preserve the sample if the analyte is unstable.
- The importance of avoiding contamination before or during analysis.

The first stage of an analysis is sampling, where a representative sample for analysis is taken and submitted for assay; this is followed by transport to the laboratory, storage, sample preparation, and analysis. Analytical methods can be highly specific and sensitive. However, the results of an analysis are only

side. The dynamics involved produce a violent rebound effect in combination with high friction and impact forces, providing for excellent comminution and homogenization.

**Vibratory mixer mills** These mills comminute and homogenize soft, fibrous, hard, and brittle substances in both dry and wet conditions. They are particularly suitable for processing basalt, building material, dolomite, earth, cereals, organic tissue samples, hair, cocoa, rubber, ceramics, bones, synthetic resins, alloys, minerals, plant constituents, salts, tablets, pills, peat, and wool. **Figure 3F** shows how grinding balls comminute the sample material. The grinding cylinder, holding the sample and loaded with one or two balls, is locked in the mounting bracket and energized into horizontal, radial oscillations, thus bouncing the balls against the rounded cylinder ends so as to disintegrate the sample material. The degree of reduction and the ultimate fineness attainable are governed by the mass of the grinding balls.

**Cold grinding** The terms cold grinding, freeze grinding, or cryogenic grinding refer to comminution methods employing mechanical and thermal energies to break down solids. This method is often the only means of obtaining sufficient comminution results with acceptable homogenizing efficiency, which is particularly important when handling materials with low melting points, high grease and/or moisture contents, as well as resilient or tough consistencies.

The disintegrating behavior of substances with any such characteristics can be substantially improved by addition of carbon dioxide 'ice' or by embrittlement with liquid nitrogen. The use of direct or indirect cooling, in combination with mechanical comminution techniques, facilitates adequate processing of many types and grades of plastic, rubber, organic tissue, pharmaceutical products, spice, herbs, and food samples.

The greatest care must be taken when handling coolants, especially liquid nitrogen, and the use of protective gloves, safety goggles, etc., in keeping with the safety instructions of grinding equipment manufacturers is most important.

### General Safety Aspects

Laboratory staff are always exposed to risk when undertaking their duties, comminution procedures included. Manufacturers have made great efforts to make their equipment foolproof against handling errors: mills, for example, cannot be unloaded until the machine has come to a dead stop.

Comminution of explosive materials represents a particularly high hazard and must, therefore, be left to qualified laboratory staff with special training in this area. Included in explosion hazard is the risk of dust explosions, always likely to be present in grinding operations involving high degrees of sample comminution.

*See also:* **Particle Size Analysis. Quality Assurance:** Reference Materials; Production of Reference Materials.

### Further Reading

- German Federal Ministry of Research and Technology (1988) *Umweltprobenbank*. Berlin: Springer (in German).
- Markert B (1996) *Instrumental Element and Multielement Analysis of Plant Samples/Methods and Applications*. Chichester: Wiley.
- Schlodt JD, Backhaus F, and Reuter U (1985) Beiträge zur Umweltprobenbank, I. Studie zur Probenhomogenisierung bei tiefen Temperaturen unter Berücksichtigung der für die Umweltprobenbank notwendigen Parameter. Forschungszentrum Jülich, 'Jul. Spez. 330', 25pp.

## Sample Preservation

**R D McDowall**, McDowall Consulting, Bromley, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Know the Stability of an Analyte

The purpose of this article is to discuss the importance of knowing:

- The stability of the analyte in the sample matrix.

- Any measures that can be taken to preserve the sample if the analyte is unstable.
- The importance of avoiding contamination before or during analysis.

The first stage of an analysis is sampling, where a representative sample for analysis is taken and submitted for assay; this is followed by transport to the laboratory, storage, sample preparation, and analysis. Analytical methods can be highly specific and sensitive. However, the results of an analysis are only

as good as the initial sampling and the sample preservation methods used to deliver the sample to a laboratory for analysis as well as ensuring the stability of an analyte after preparation and pending analysis.

Knowing that an analyte is stable is essential to enable an analyst to draw conclusions based upon the experimental observations. If an analyte is unstable, then steps such as the addition of preservatives or storing under known conditions can be used to stabilize an analyte. The stability of the analyte must be known from the time of sampling until analysis and calculation of final results; this overall time should include a contingency for any repeat analysis.

There are many approaches to sample preservation. For example, if the analyte shares structural similarities with another compound for which there is a proven preservation protocol, then all that may be required is a duplication of the procedure. For a novel compound, the instability problem may require a novel approach. It is important to realize that preservatives may only slow the rate of analyte degradation – they are not always a universal panacea.

Equally important during the sampling, storage, and sample preparation stages of an analysis is the need to avoid the introduction of contaminants from the environment, storage conditions, or sample container. The adsorption of the analyte onto the surface of containers can be confused with instability and complicates the sample stability and preservation problem.

Owing to the diverse nature of analytes and sample matrices, it is impossible to give specific advice about the storage and preservation under all circumstances. This must be determined on a case-by-case basis. It should be stressed that the analyst is responsible for the quality of information generated by his or her methods of analysis. Knowing the stability of the analyte from sampling through the analysis is a prime responsibility that helps ensure the quality of data generated by the analytical method.

### **General Aspects of Sampling**

There are a number of general points that set the scene for sample preservation. It begins with the assumption that the sample is homogeneous and representative of the whole from which it was derived. This is a key assumption that is covered in more detail elsewhere in this encyclopedia.

If this is not true and the sample and the bulk differ, the wrong conclusions could be drawn about the sample from the analytical results and extrapolated

to the original batch or lot of material. If the sampling site is remote from the laboratory or site of analysis, then the sample must be transported. Unless the sample results are required urgently, the sample will usually be stored at the laboratory until the analysis. In both cases, the samples will be stored prior to analysis. The storage conditions to preserve the analytes must be known and followed; otherwise the analytical results will be flawed. Sample preservation may also be required during the sample preparation stages to ensure the integrity of the analyte.

### **Sample Information**

The sampling information required by the analyst, with respect to sample preservation, are that a sampling protocol that ensures analyte preservation was followed correctly together with the history of the sample's transport and storage before receipt in the laboratory.

### **Sample Labeling**

All laboratories, including those operating under a quality scheme such as ISO 17025, Good Laboratory Practice (GLP) or Good Manufacturing Practice (GMP), must ensure that the sample is labeled correctly. It is the analysts' responsibility to ensure that this is done properly, either by themselves or by a third party who is performing the sampling. All relevant information should be gathered about the sample to aid the analyst. If the sample is from a system that changes with time, e.g., river water, then the date and time of sampling must be noted to ensure that the results have relevance to the sample, e.g., if a sample is connected with litigation over an illegal discharge into a river.

Labels must be appropriate to the conditions under which the samples will be taken and stored. It is not advisable, for example, to use labels marked with water-soluble ink if sampling river water. Equally, labels that could disintegrate with successive freeze/thaw cycles for the analysis of biological samples would be inviting disaster.

### **Sample Containers**

Containers are necessary to keep each sample discrete and are essential when the sampling site is remote from the laboratory. They should be of sufficient size and volume to ensure that they hold sufficient sample for the analysis and any repeat assays. This may not always be the case but the analyst should discuss with the submitter the minimum requirements for a particular assay.

Sample containers must be fit for the purpose to which they will be put. This has many aspects:

- They must be able to hold sufficient sample for the analysis and have sufficient sample for a repeat if required.
- They must be made of inert material that does not have any interaction with the sample such as absorption or adsorption. The containers used for sample storage must be evaluated for inertness and purity. Some containers or the stoppers can introduce phthalate plasticizer or other interfering compounds into an assay; care must be taken with the use of sample containers.
- They must be robust enough to cope with sampling, transport, and storage conditions that are likely to be encountered. The lids of any container must fit tightly to prevent any loss of sample during transportation and storage if tipped and to prevent any entry of air especially if the sample is susceptible to oxidation. Alternatively the cap must fit securely to avoid sample loss.
- Reuse of containers is a judgment based upon the analytical requirements, the information generated by the analyst, and economic factors of recycling versus purchase of disposable containers. If recycled, great care must be taken to avoid contamination by the cleaning process. This will be discussed later under contamination.

Although not normally taken into account, containers may influence the results of an assay, especially when quantifying trace amounts. To prevent the introduction of any artifacts they must be clean. The best way to determine the suitability of containers and the cleaning conditions is by experiment during method development and validation before starting work. However, there are many situations where the work starts, problems are seen and then investigated, thus wasting time and effort and potentially presenting the wrong information. Often in this context the use of disposable containers, especially for biological samples, is one way to ensure that the container is clean and avoids the need to cleanse the sample containers before reuse. Washing the containers for recycling is a process that can introduce contamination from the chemicals used; this again must be evaluated carefully before its introduction. For example, detergents containing phosphates cannot be used if phosphates or surface-active agents are being measured. In the same way, sulfuric acid cannot be used as a wash if traces of sulfate are being measured.

### Sample Transport

When transporting samples from the sampling site to the laboratory (this can encompass from laboratory

to laboratory and continent to continent) it is essential to ensure that the process does not prejudice the integrity of the sample or affect any results in any way. The analyst should investigate the transport of samples especially if movement of the sample is possible. Ideally, all samples should be transported to a laboratory as soon as possible after collection. The samples should be sealed and packaged in such a way as to avoid breakage, deterioration, or losing the contents.

Biological samples are usually frozen after sampling and transported in the same state, thus they will be solid and the major problem here will be thawing and possible analyte decomposition at ambient temperatures. Other samples may be liquid or a powder under the shipping conditions; care must be taken to ensure that movement of the sample does not break the container and careful packaging is essential to ensure safe arrival of the shipment at the laboratory.

### Sources of Analyte Loss

There are a number of processes that can cause analyte loss. Some, such as diffusion, are limited to fluid samples. Others, such as oxidation, may be applicable to many types of sample. The main processes contributing to analyte loss are discussed below. If a problem of sample stability arises, one or more of the following processes may be responsible for the loss. Therefore, the problem may require more than one solution to solve it.

- **Light** – Photodecomposition of material is a common cause of instability. Therefore, in the first instance, samples should be stored out of direct sunlight and in brown bottles. If photodecomposition is a particularly bad problem, then the samples should be kept in aluminum foil and all working solutions of the sample and the standards should be treated similarly. When measuring nifedipine in plasma, the analysis is best undertaken in yellow light to avoid photodecomposition; yellow light is suitable for the analysis of some vitamins.
- **Oxidation** – Compounds that are easily oxidized should be kept in containers that have tightly fitting lids to exclude air. If necessary, the sample container should be flushed with nitrogen or an alternative inert gas to displace all oxygen present or the container should be filled completely with the sample.
- **Temperature** – It is important to know the stability of samples awaiting analysis. This can take many forms depending on the type of sample. For instance, biological samples containing analytes must be investigated to find out the stability of analytes in



the particular matrix at ambient, refrigerated, and subzero temperatures. This situation may be most relevant to environmental samples where the compound or elemental species may undergo conversion to another form under the influence of microbial activity. A study on storage temperature will give an indication of the time that the sample can be stored before the analyte suffers degradation that renders the results useless. Samples taken at high temperatures from chemical processes may change composition upon cooling; care must be taken to ensure that the results relate to the actual material.

- **Biodegradation** – Enzymes present in the sample can hydrolyze conjugated material and release extra material for analysis or alter the analyte reducing the amount available for assay. Microbial contamination can have a similar effect and can metabolize components of the sample and produce others. This activity can affect dissolved oxygen and carbon dioxide through to nitrogenous compound and carbohydrates.

- **Preservatives** – Whatever chemical is used to preserve a sample, the effect on the overall assay must be assessed in order to prevent the introduction of another source of interference to the analytical procedure. This is especially important in trace element analysis where the preservative could introduce the very elements that are being quantified.

- **Evaporation or volatility** – Storing samples at low temperatures, in containers where there is no room for evaporation, or if the end of the container is firmly stoppered can prevent this from occurring.

- **Adsorption or absorption** – These can cause losses where the analyte is irreversibly attached on the surface of a sample container or solid particles contained in the sample. To avoid losses of this type either a container made from a different material should be tried or chemical modification of the interior surface of the container, e.g., by silanization, should be investigated. Absorption of carbon dioxide from the atmosphere can affect the pH value and conductivity of a sample.

- **Diffusion** – Some fluid samples can have this problem and it is the act of the analyte migrating into and eventually through the container material. It is most noticeable with low molecular weight gases, e.g., hydrogen and helium, with high molecular diffusivities. Liquids also exhibit diffusion, although to a lesser extent. To reduce diffusion, a material should be used to contain the sample that is cost effective, practical, and reduces or does not allow diffusion to take place. Also the sample should be assayed as soon as possible to avoid such losses with storage.

- **Chemical reaction due to chemical instability** – Molecules with labile bonds can decompose under

the storage conditions employed. This can be at ambient, subambient, or frozen temperatures. Reducing the temperature may sometimes stop the decomposition or it may only slow the rate of reaction. Specific tests must be carried out to determine the effects of the storage conditions and the effect on the analyte.

- **Analyte Precipitation** – Certain substances such as colloids that precipitate from solution affect the nature of the sample. Similarly, if the concentration of a compound is near the saturation point, it can precipitate if the storage conditions differ greatly from those at sampling. In some environmental samples, elements may precipitate due to changes in their oxidation state or sulfide formation that may lead to the scavenging or coprecipitating of other elements in a sample.

- **Polymerization and depolymerization** – Polymers can undergo depolymerization after sampling due to chemical and biological causes; conversely small molecules can polymerize. Storage and preservation methods should avoid this happening.

- **Apparent analyte gain** – In biological matrices this can occur when compounds are metabolized *in vivo* to either glucuronide conjugates or N-oxide derivatives. Some of these metabolites may be labile and can revert back to the parent molecule, thus increasing the concentration of the parent over time. This conversion will be seen with a pro rata fall in the metabolite concentration if there is a method available to measure it.

It follows from the areas outlined above that an analyst must be vigilant with respect to the problems that can occur with samples during collection and storage. As stated earlier, an analyst must know the stability of the compound to be measured under the conditions used in the method and the storage conditions under which the sample will be held pending analysis.

## **Storage Techniques**

A specimen must be kept under conditions that preserve the analyte. Some analytes remain stable for long periods, while others require storage at low temperatures, and some others require the addition of preservatives or stabilizers. In some instances, no practical method of preservation may be found.

The extent of these stability problems is a function of the chemical and nature of the sample matrix, its temperature, its exposure to light, the nature of the sample it is placed in, the time between sampling and analysis, and the conditions to which it has been submitted. It is important to minimize extremes of

these conditions and analyze the sample as quickly as possible. The longer a sample is stored, the greater the likelihood that stability problems will be observed.

The storage techniques that are possible for solid, liquid, and gaseous samples are presented below. Owing to the diverse nature of sample matrices, it is impossible to give specific guidance for all samples; therefore a general approach has been taken here with general advice. The analyst should then tailor this for a specific matrix and analyte combination.

### Storage Techniques for Solid Samples

The main forms of samples are usually powders or solids. The first consideration is settling of powders during transportation from the sampling site to the laboratory. The motion of transport should not have the effect of altering the composition of the sample; such an effect may be found with dietary mixtures of feed with a medicated product. The differences in particle size may cause zoning of the medicament. If zoning is suspected, this can be examined experimentally. Transport of solid material may cause breakage by the sample striking the side of the container. Again sample containers should be suitable for the sample to be collected and stored.

Solid samples may have problems with oxidation and, therefore, it may be necessary to remove air by flushing or purging with nitrogen or another inert gas such as argon. An alternative is to fill the container and exclude air, thus avoiding the problem.

Soil samples should be placed into a glass wide-mouthed bottle with a polytetrafluoroethene (PTFE) lined cap. The bottle should be half filled so that the sample can be homogenized before analysis. The samples should be stored in the dark to retard changes due to biological activity. Similarly, the sample containers should be darkened to minimize the effect of light during transport and storage.

Adsorption of the analyte onto the surface of the container is possible, more especially from powdered, than solid, samples. This effect should be examined by stability studies over time and at different concentrations.

### Storage Techniques for Liquid Samples

**Biological samples** The dangers of oxidation, hydrolysis, enzymatic or bacterial transformation, and photodecomposition are usually greater with organic analytes than inorganic ones. The effects are usually concentration dependent that can add to the difficulties of the analysis. With analytes in biological samples they are usually stored at low temperatures to slow down the kinetics of the alteration processes.

This is the most common way of preserving a biological sample. The temperatures used for storage can be subambient (4°C) or deep-frozen (−20°C, −40°C, −80°C, or even under liquid nitrogen). Usually, if a sample is deep-frozen, the temperature chosen will be the most convenient one available, usually −20°C. However, this is no guarantee of stability. Indeed, urine, depending on the amount of salts in a sample, is not completely frozen until −30°C and there are small cells of liquid at −15 to −20°C that may allow some decomposition processes to occur. A storage temperature of −40°C may be appropriate for some analytes in urine.

Often chemicals are added to biological samples to preserve and stabilize some analytes and enhance storage life. These preservatives can act as antioxidants, enzyme inhibitors, or inhibitors of microbial growth. A preservative can have a general action, e.g., ascorbic acid is a general antioxidant and is added in relatively high concentrations in a small volume to plasma to prevent the oxidation of catecholamines such as adrenaline and noradrenaline. At the sample time, a specific preservative, ethylenediaminetetraacetic acid, is added to the same sample to chelate the metal ions used as enzyme cofactors in the metabolism of catecholamines (catechol-O-methyl transferase; COMT). This deprives COMT of the cofactor necessary to function and stabilizes the catecholamines in the samples from *in vitro* degradation by this metabolic pathway. Thus catecholamines in plasma or serum provide an interesting example in the use of specific and general preservatives.

To inhibit microbial growth in blood samples taken for the determination of ethanol, sodium fluoride is added. As a volatile organic analyte is measured, the sample containers should prevent evaporation. Clinical chemistry samples have a variety of preservatives depending on the specific analysis, e.g., heparin or citrate to prevent clotting of a whole blood sample. In this instance the tube already has the preservative in place before the sample is added; all that is required is that the container is gently mixed to disperse the preservative throughout the sample.

Lyophilization (freeze drying) can be used to preserve nonvolatile analytes in liquid samples. The homogenized sample is frozen and the water content is removed over 12–24 h. This is very good for stabilizing some analytes that would be susceptible to enzymic degradation in biological samples and the specimen can then be stored at room temperature without the need for expensive freezer facilities. Analytes that are heat labile can be freeze-dried and then stored as a powder in a freezer. Lyophilization is

useful as the bulk of a sample can be reduced drastically but the process can be time consuming and labor-intensive.

**Water samples** Water can be analyzed for a number of analytes, therefore, the storage conditions and preservatives required can vary. As some changes in samples can occur, certain physical and chemical measurements (temperature, pH value, alkalinity, conductivity, dissolved gases, etc.) are best carried out at the site of sampling to ensure accurate values.

Water samples are usually refrigerated at 4°C for transport to the laboratory and in storage pending analysis. In addition, the sample preservation required depends on the stability of individual analytes. When sample preservation is required there are a number of possible approaches. The pH value of the nonchlorinated water samples can be adjusted by acidification to prevent bacterial growth, biological degradation, or retard acid or base catalyzed decomposition.

Some preservatives, however, present problems in their use. Mercury(II) chloride is used to inhibit bacterial growth. It is effective but has several disadvantages; it is toxic, environmentally unfriendly, and it can interfere with some assays involving activated charcoal where it lowers the recovery of some analytes.

To prevent oxidation, either ascorbic acid or sodium thiosulfate is often added to samples to remove free chlorine that could otherwise form trichloromethane. Polyaromatic hydrocarbons (PAHs), in particular benzo[*a*]-pyrene, are unstable in chlorinated water and sodium thiosulfate is added to preserve the analytes from degradation by chlorine. To ensure stability of the PAH analytes, the water sample is also acidified, again showing that often a single preservative is not always effective for some analytes.

### **Storage Techniques for Gaseous Samples**

The assumption made in this section is that a compound or substance is considered to be a gas if it normally exists in this state at normal pressure and temperature. Excluded from this section are substances that normally exist as a solid or liquid and have sublimed or volatilized due to their vapor pressure.

The most common use for gaseous samples is for environmental monitoring; monitoring the workplace for noxious substances, monitoring a location after a chemical accident, or assessing the level of air pollutants over time. Owing to the physical

nature of gases, there are special sampling procedures used that are related to the technique used for analysis.

There are three main approaches to sampling gaseous samples:

- Direct measurements on site using a portable instrument or a chemical reaction. As this method of monitoring analytes in gaseous samples does not involve sample storage, it will not be considered further here.
- Whole air sample, where a sample is taken and returned unchanged to the laboratory for analysis. This approach is very useful for maintaining sample integrity and where the analyte concentration is sufficiently high to avoid a sample preparation stage before analysis.
- Concentration of the analytes from the air by absorption, adsorption, condensation, or chemical reaction at the site of sampling after which they are returned to the laboratory.

**Whole air samples** Sampling can use glass, metal, or rigid PTFE containers with the neck sealed by a wax-filled cap or a stopcock. Alternatively, plastic bags, made of a variety of inert materials, have also been used. Many parameters can affect the storage of samples. Therefore, the analyst should investigate the storage characteristics of the sampling bags for the analytes being collected. This is achieved by filling the bags with air and injecting known amounts of the analyte. The bag can be sampled over time and the analyte determined to see if there is any adsorption to the bag or diffusion occurs. The results should show no significant drop in analyte over time. If sample bags are reused, they should be checked to see if there are any memory effects from the previous sample. This can be done by flushing the bag with clean air, filling with air, and analyzing after time in storage.

**Concentration sampling** Samples that are concentration sampled have a known volume of gas drawn through or left for a known time for passive diffusion to occur as described in the preceding article. The precautions needed to preserve the sample before storage depend upon the method of sampling. Concentration sampling methods should avoid high temperatures and extremes of relative humidity that may affect the adsorption equilibrium and may promote hydrolysis of some analytes. Where the collection medium is close to saturation, breakthrough of the analyte may occur. In these instances an analyst should know under what conditions this occurs and avoid these extremes; this should be an experiment undertaken during validation of the method.

Where more than one compound is to be collected, there may be the phenomenon of displacement of one analyte by the other. This can be a problem for silica gel as water vapor will displace nonpolar compounds and cause early breakthrough.

**Solid sorbent with thermal desorption** Each cartridge should be thoroughly conditioned by heating and purging with nitrogen for up to 24 h. Most sorbents used are gas chromatography packings, such as Tenax and Porapak, and are stable at the temperatures used for desorption. The thermal stability of the sorbent is very important as decomposition and oxidation of it after repeated reuse and conditioning may cause interference with the analysis. To reduce the background, pretreatment of the sorbent by Soxhlet extraction before packing into the sampling tube can be undertaken. Samples should be analyzed as soon as possible after collection. Storage for up to several weeks may be possible but only after determining the effect of such storage for each analyte.

## Evaluation of Contamination and Analyte Loss

Contamination of the sample must be prevented and degradation of the analytes must be minimized. Even if analyte stability has been fully evaluated and is known, the work of the analyst can be annulled if contamination of the sample is allowed to happen. Again it is the analyst's responsibility to ensure that the samples are taken and stored to minimize any extraneous artifact production or contamination.

### Evaluation of Stability

The stability of the analyte should be known for the normal conditions under which it could be exposed; this should include potential problems that could occur. The analyst should be proactive in this work rather than reactive. The following stability experiments can be undertaken, depending on the nature of the analysis, the analytical method, the matrix, and the analyte.

- The stability of the analyte from the time the sample is taken to receipt in the laboratory must be examined. Two experimental designs can be used, either a sample with a known amount or concentration is prepared or an actual sample with an unknown concentration. In either case, aliquots are taken for analysis at various times after sampling; a plot of the results should show a level display if the analyte is stable or a decreasing concentration if the analyte is unstable. This should identify whether there are any problems with analyte stability; if this

is so, work can begin on conditions that must be used to stabilize the analyte and develop an efficient sampling and preservation protocol.

- Stability during transportation must be tested. Here there are number of factors such as zoning of powders, thawing of frozen samples, which are dependent on the matrix. Investigations into the effect of these factors on analyte stability should be undertaken before large-scale work starts to prevent the wrong conclusions being drawn from incorrect data.
- The problems of long-term storage must be examined. Regardless of the analyte, the length of time that the analyte can be stored without degradation must be known. Long-term storage is a relative term: it can be hours for some analytes or months for others. A long-term stability test should be established under the storage conditions for the samples. If a range of concentrations are measured normally, then two concentrations at the top and bottom should be included. Enough samples should be prepared for a number of replicate (ideally six at each concentration) determinations at intervals over the anticipated storage time. Some additional samples should be prepared in case of a series being lost due to analytical problems or more intervals need to be included after the experiment has started. Again the results should indicate how long the samples could be stored until unacceptable degradation has occurred (say, >10–15%). This time should be the maximum time that the samples could be stored.

### Evaluation of Contamination

Evaluation of contamination is usually through the use of blank samples, i.e., samples that do not contain any analyte. As analytical techniques become more sensitive, and lower and lower concentrations are measured, blank samples are very important. The blank samples aid an interpretation of the result and add confidence. This is important where comparative methods of analysis are used such as chromatography or ultraviolet spectrophotometry. In these instances, the absence of an analyte signal in the blank sample allows an analyst to infer that the signal in the samples and standards is, in fact, due to the analyte.

Blank samples consist of two main types:

- Reagent blanks: these are made up of the reagents used in the assay. The sample matrix is water instead of, e.g., urine and is taken through the method as a sample. The purpose of this type of blank is to identify any contamination that may be due to the reagents used in the analysis.
- Matrix blanks: blank samples should be generated at the same time as the samples themselves, stored

under the same conditions, and taken through the analytical method as if they were samples.

Ideally, reagent and matrix blanks should be included every time a method is used. If this is not feasible then they should be included in a run every time reagents are made up or new batches of chemicals are used, although this does not inform the analyst about the day-to-day operation of the method.

In addition to the blank samples (negative control samples), positive control samples should be considered. These are samples containing a known amount of analyte in the sample matrix. They are included when a batch of samples is taken and stored with them until analysis. In this way there is a marker with every batch of samples to monitor storage conditions and analyte loss.

### Sources of Contamination

There are a number of sources of contamination that an analyst should be aware of, such as the environment, solvent impurities, microbes, and reagents.

**Environment** The environment of the laboratory can introduce contamination in the form of air particles, which may be very important in the case of environmental analysis. The environment of the sample, the sample container, is also a source of contamination. As the container can have an effect on the amount of analyte, it can also introduce contamination into a sample. Typical examples can be plasticizers from plastic containers, including phthalates from stoppers, etc.

**Sample containers and glassware** As discussed earlier, when cleaning sample containers for reuse, much care should be taken. The procedure should not be likely to introduce either contamination or the analyte into the analysis. Therefore when cleaning containers for analysis of volatile organic compound, for example, it would be foolish to use organic solvents as this could introduce the same compound that the method would be measuring.

Contamination can arise from sample containers and apparatus that have been cleaned. The detergents and cleaning agents can introduce contamination that can interfere with an analysis. Cleaning agents that have been tested by examining blanks in the analytical methods can be used routinely. New containers that will be recycled many times should be cleaned thoroughly before use the first time, as there may be dirt and interfering compounds from the manufacturing process or picked up during storage or transport.

**Reagents** Reagents are a source of potential contamination. The use of analytical grade reagents and chemicals is no guarantee of avoiding contamination. Examples of problems with liquid–liquid extraction are typical of the problems facing analytical chemists measuring compounds at the limits of sensitivity. A typical problem can be seen with the stabilizing agents and additives used for solvents.

A problem with liquid–liquid extraction that is often overlooked is the addition of compounds by manufacturers to prevent oxidation or decomposition of their product. For example, to prevent phosgene formation in chloroform, the solvent is often stabilized with 2% ethanol. In itself, this is no problem but when a method is to be established from the literature it may not be immediately obvious to the reader if the solvent contains ethanol (this assumes that the originating author knew the composition of the extracting solvent). The presence of ethanol can change the polarity of the solvent and affect the specificity and recovery of a method.

Diethyl ether can have any one of several chemicals added to it to prevent peroxide formation. These include 2% ethanol,  $2\text{--}3\ \mu\text{g m}^{-1}$  pyrogallol and  $<0.1\ \mu\text{g m}^{-1}$  butylated hydroxytoluene (BHT). Often the addition of a compound is not mentioned in the specification of the solvent. These additive compounds can affect the outcome of an assay. Ethanol, as mentioned above, can affect the polarity of the solvent and hence its selectivity. BHT could be concentrated upon evaporation and form a discrete liquid chromatographic peak, while pyrogallol is electrochemically active and can interfere with a liquid chromatographic analysis when using an electrochemical detector.

**Solvent impurities** The extracting solvent may introduce impurities into an analysis. The term ‘impurity’ is a function of the detection system used in the analytical scheme and the limits of detection required. It may be necessary to redistill or purify the solvent prior to use, e.g., ethyl acetate. Each reagent needs to be assessed for its purity. Moreover, when changing batches, a purity reassessment should be made as it is better to be safe than sorry.

**Microbial contamination** The impact that microbial contamination makes can depend on the nature of the sample and the analysis. Biological samples can be affected profoundly by microbial contamination: ethanol analysis in urine can be affected by elevated levels by microbial contamination. Fluoride is used to inhibit growth and alcohol production. Water samples are similarly affected by the growth of microorganisms for various analytes.

See also: **Air Analysis:** Sampling. **Clinical Analysis:** Overview; Sample Handling; Glucose; Sarcosine, Creatine, and Creatinine; Inborn Errors of Metabolism. **Polycyclic Aromatic Hydrocarbons:** Environmental Applications. **Sample Handling:** Comminution of Samples; Automated Sample Preparation; Robotics. **Sampling:** Theory.

## Further Reading

Barcelo D (ed.) (2000) *Sample Handling and Trace Analysis of Pollutants*. Amsterdam: Elsevier  
 International Standard ISO 5667/3 (1985) *Water Quality: Sampling*, Part 3: Guidance on the Preservation and

Handling of Samples. Geneva: International Standards Organization.

Koester CJ and Clement RE (1993) Analysis of drinking water for trace organics. *Critical Reviews in Analytical Chemistry* 24(4): 263–316.

Melcher RG, Peters TL, and Emmel HW (1986) Sampling and sample preparation of environmental material. *Topics in Current Chemistry* 134: 59–123.

Mester Z and Sturgeson R (2004) Sample preparation for trace element analysis. Amsterdam: Elsevier.

Moldoveanu SC and David V (2002) Sample preparation in chromatography. Amsterdam: Elsevier.

Pawliszyn J (2002) Sample preparation in field and laboratory. Amsterdam: Elsevier.

## Automated Sample Preparation

**R D McDowall**, McDowall Consulting, Bromley, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Automated Sample Preparation

Automated sample preparation can offer a means of improving efficiency and throughput within a laboratory. However, before deciding to implement an automated system, each method to be automated should be evaluated against its objectives, sample numbers, resources, and the amount of time available. Consideration of the physicochemical properties of the analyte, the sample matrix, and the proposed sample preparation technique for isolation will aid the efficient development of robust methods. It is important that the method strategy is thought out and that sufficient time is allowed to think it through. In general, the method used should be the simplest one available that is consistent with the objectives of the assay; it may not involve automated sample preparation.

If automation is viable, the type of automation chosen (e.g., robot, sample processor, or autosampler) will depend upon the expertise and availability within a given laboratory as well as consideration of the items above.

### The Role of Automation in Sample Preparation

Many laboratories ignore a major part of the analytical process that can be rate limiting for many methods and is responsible for a large proportion of

the variation of a method, namely sample preparation. If productivity increases are sought within a laboratory, then there are two possible approaches. The first approach is to change procedures so that tasks are simplified or even eliminated – as such this may not involve the use of automation. However, the viability of this approach can be subject to question unless labor costs are reduced dramatically. The second approach is to use laboratory automation. At the laboratory level, the automation of a task should produce one or more of the following advantages:

- Greater increases in productivity (either in numbers of samples assayed per unit time or speedier turn around time) resulting in overall cost reduction and increased productivity.
- Systems are capable of the same or better precision and accuracy as existing manual methods.
- Trained laboratory staff are free to do more creative or productive work thus improving morale.
- Human contact with biological or chemical hazards is reduced.
- Lower consumption of sample may be possible.
- Lower consumption of reagents used in an automated analysis, resulting in cost reduction and/or reduction of environmental hazard from their disposal.
- Miniaturization of automated sample preparation using microtiter plates containing 96, 384, or higher well capacities.

Instrument automation has been available in various forms for many years. However, the main emphasis of this area of automation used to be toward



See also: **Air Analysis:** Sampling. **Clinical Analysis:** Overview; Sample Handling; Glucose; Sarcosine, Creatine, and Creatinine; Inborn Errors of Metabolism. **Polycyclic Aromatic Hydrocarbons:** Environmental Applications. **Sample Handling:** Comminution of Samples; Automated Sample Preparation; Robotics. **Sampling:** Theory.

## Further Reading

Barcelo D (ed.) (2000) *Sample Handling and Trace Analysis of Pollutants*. Amsterdam: Elsevier  
International Standard ISO 5667/3 (1985) *Water Quality: Sampling*, Part 3: Guidance on the Preservation and

Handling of Samples. Geneva: International Standards Organization.

Koester CJ and Clement RE (1993) Analysis of drinking water for trace organics. *Critical Reviews in Analytical Chemistry* 24(4): 263–316.

Melcher RG, Peters TL, and Emmel HW (1986) Sampling and sample preparation of environmental material. *Topics in Current Chemistry* 134: 59–123.

Mester Z and Sturgen R (2004) Sample preparation for trace element analysis. Amsterdam: Elsevier.

Moldoveanu SC and David V (2002) Sample preparation in chromatography. Amsterdam: Elsevier.

Pawliszyn J (2002) Sample preparation in field and laboratory. Amsterdam: Elsevier.

## Automated Sample Preparation

**R D McDowall**, McDowall Consulting, Bromley, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Automated Sample Preparation

Automated sample preparation can offer a means of improving efficiency and throughput within a laboratory. However, before deciding to implement an automated system, each method to be automated should be evaluated against its objectives, sample numbers, resources, and the amount of time available. Consideration of the physicochemical properties of the analyte, the sample matrix, and the proposed sample preparation technique for isolation will aid the efficient development of robust methods. It is important that the method strategy is thought out and that sufficient time is allowed to think it through. In general, the method used should be the simplest one available that is consistent with the objectives of the assay; it may not involve automated sample preparation.

If automation is viable, the type of automation chosen (e.g., robot, sample processor, or autosampler) will depend upon the expertise and availability within a given laboratory as well as consideration of the items above.

### The Role of Automation in Sample Preparation

Many laboratories ignore a major part of the analytical process that can be rate limiting for many methods and is responsible for a large proportion of

the variation of a method, namely sample preparation. If productivity increases are sought within a laboratory, then there are two possible approaches. The first approach is to change procedures so that tasks are simplified or even eliminated – as such this may not involve the use of automation. However, the viability of this approach can be subject to question unless labor costs are reduced dramatically. The second approach is to use laboratory automation. At the laboratory level, the automation of a task should produce one or more of the following advantages:

- Greater increases in productivity (either in numbers of samples assayed per unit time or speedier turn around time) resulting in overall cost reduction and increased productivity.
- Systems are capable of the same or better precision and accuracy as existing manual methods.
- Trained laboratory staff are free to do more creative or productive work thus improving morale.
- Human contact with biological or chemical hazards is reduced.
- Lower consumption of sample may be possible.
- Lower consumption of reagents used in an automated analysis, resulting in cost reduction and/or reduction of environmental hazard from their disposal.
- Miniaturization of automated sample preparation using microtiter plates containing 96, 384, or higher well capacities.

Instrument automation has been available in various forms for many years. However, the main emphasis of this area of automation used to be toward

instrumental analysis and data acquisition and not toward sample preparation. In fact, sample preparation can now be regarded as the weak element in automated analysis for the following reasons:

- Sample preparation was seen to be highly application specific: applicable to one method but not another.
- Variation in the viscosity of some liquid and semi-liquid matrices can lead to variable results, especially when low-pressure pumping systems are used for sample transfer.
- Dilution or concentration steps may be required for some but not all samples.
- The particulate or semisolid nature of some matrices may adversely affect an automated analysis.

However, it has now been realized that some of these points are not major obstructions to automating sample preparation. They have been overcome, and increasing emphasis from users and instrument companies has made automated sample preparation a reality and *de facto* standard in many areas from DNA sequencing to combinatorial chemistry, often combined with the instrumental analysis technique.

Sample preparation is often one of the rate-limiting factors of sample throughput for many methods. Furthermore, it is a critical factor for determining the accuracy and precision of analytical results. The final aim of any sample preparation scheme must be to isolate and purify the compound of interest and present it in a form that is compatible with the analytical instrument. The task facing analytical chemists is to select and use the many diverse systems available, such as robots, sample processors, and autosamplers, to achieve that goal in a timely and cost-effective manner.

### **Islands of Automation**

It is important that a laboratory does not concentrate on simple task automation, such as sample preparation, but develops an overall strategic plan for automating the whole laboratory environment. The reason for this is that analysis is not performed for its own sake, except for educational purposes, but to provide information upon which decisions are made. The purpose of an analytical laboratory is to provide reliable information in a cost-effective and timely manner. To concentrate solely on automating bench-level tasks is inappropriate. The data generated by automated systems have to be interpreted and the resultant information distributed to the organization where decisions will be made.

### **Method Development: When to Automate?**

When developing a method: should an assay be automated immediately or be developed manually first and automated later? While no definite rules can be stated, it may well be easier for methods to be developed manually, first and then automated later. The main exceptions to this statement are column-switching in liquid chromatography (LC) and online dialysis, which can only be developed as automated methods. Another exception is when an existing automated method can be modified easily for use with new analytes, typically those that are structurally similar.

Within an industrial context, time constraints are the main reason for this approach; an assay may be required quickly and a manual method is usually quicker to develop than an automated one. An assay developed first manually allows control to be exercised over the method, its analytical principles and operation can be understood, and trouble shooting problems become easier without the added complication of automated equipment. This approach has the added benefit that results can be generated more rapidly, allowing work to proceed while a procedure is refined and automated.

Attempts to develop automated methods from first principles require that sufficient time, human skills, and resource are available and that these are not critical elements in the method development equation. Therefore, before beginning any work, it is important that sufficient information be assembled to enable a risk and technical feasibility assessment to be carried out on the proposed automated sample preparation method. Thus, it is important that the method selected comprises sample preparation processes that are amenable to automation. The techniques that have the greatest potential for automation are solid-phase extraction and high-performance liquid chromatography. Other sample preparation techniques such as liquid-liquid extraction, protein precipitation, and ultrafiltration are either difficult to automate or may not be cost-efficient to do so compared with alternative approaches.

### **Considerations for Automated Sample Preparation**

To obtain the best from an automated analysis, it is important that only a robust manual method is considered for automation. To use automation effectively, the following factors should be considered before attempting any work on a method.

**Analytical objectives** The following questions should be answered:

- Why is the method being developed or used?
- What questions will be answered by the information generated by the technique?

- How many analytes are to be measured by the procedure now and will this change over time? A method for multiple analytes will involve compromises being made. The number of analytes and their polarity range are factors influencing the choice of both the method of extraction and the whole automated system.
- What is the concentration/amount range over which each analyte is to be measured? If several analytes are to be quantified, each with a different range, this can lead to issues with the design and automation of the sample preparation method.
- For how many different sample matrices will the method be applied and which will need to be validated?

The answers to these questions will give the scope and limitations of the assay procedure. The information provided should enable efficient planning of the approach to take, as modifying a method several times is efficient. This planned approach should also give a better understanding of the total resource required for the work.

**Combine automated sample preparation with analysis** When automating a sample preparation method, strong consideration should be given to combining the automated sample preparation with the subsequent analytical technique. This will ensure that the greatest benefit will be obtained from the effort input. This is discussed in more detail at the end of this article.

**Total sample numbers** The total number of samples to be analyzed can be defined either per year or over the lifetime of a project. There is little point developing an automated sample preparation method for less than 250–500 samples as manual methods can usually be developed and implemented more rapidly. However, when an already established method is available for a structurally related compound, a slight modification may enable an automated method to be developed and validated in minimal time.

**Rate of sample arrival** How will the samples arrive and what is the turn around time required? While the overall numbers per annum may indicate that automation is required, the actual numbers arriving in the laboratory and the turn around time may dictate otherwise, e.g., in a laboratory dealing with emergency poisoning cases, samples may arrive at a continuous rate of a few samples per day, but a rapid turn around is required. It may then be appropriate

to develop an assay that has a simple manual sample preparation and chromatographic separation appropriate to the objectives of the analysis. Conversely, if samples arrive in larger numbers and turn around time is not as vital; an assay can be set up and run when sufficient samples have accumulated to make an automated analysis cost-effective.

**Limit of quantification (LOQ)** What LOQs are required and from what sample volume? Note that the lower the LOQ required for an analysis, the longer the method usually takes to develop; this is true whether an automated or a manual method is being developed.

**Batch size** The number of samples per batch should be considered in relation to the analytical objectives. A large number of batches containing many samples would benefit most from automated sample preparation and analysis as it allows the widest choice for an automated system: robotics, sample processors, or column-switching systems.

**Analyte stability** This factor can have a dramatic effect on the type of automated system used, or in the most extreme example, whether automated analysis can be used at all. Hence, it must be determined early in the method development and validation cycle. If a compound is temperature labile, then the use of cooled sample racks or trays may minimize this effect; if degraded enzymatically an inhibitor may need to be added to prevent breakdown of the analyte pending analysis. Analyte stability can have a profound influence on the number of samples processed per unit time and hence the batch size.

**Matrix composition and effect** Each matrix can have a different chemical composition that can affect the performance of the sample preparation portion in an assay. These must be known and understood for both manual and automated sample preparation methods.

**Matrix consistency** The consistency of a matrix may vary from a liquid to a solid. However, most instrumental techniques require that a sample is liquid for analysis. Therefore, unless the automation includes a liquefaction or homogenization step, most automated systems are usually restricted to liquid samples only. Therefore, some offline manual or online robotic manipulation may be required for some solid matrices.

**Sample preparation technique** The techniques used for preparing the sample for analysis must be

amenable to automation; for example, liquid handling is probably the easiest technique to automate, as there is a wide variety of equipment and approaches possible.

**High-risk samples** All biological samples have the potential to contain infective viruses such as hepatitis or human immunodeficiency virus. Dust and powders can cause serious problems if inhaled. Automation has the advantage that human contact with these types of samples is reduced. However, some automated techniques, e.g., the use of syringes, can generate aerosols containing infectious droplets. Thus, it is important to evaluate the operation of automated systems and house them in appropriate environments.

**Time** The time available can often determine the approaches taken in method development; therefore, it is important to use time effectively and plan the work well. Such constraints may not always be compatible with the development of automated methods. Again this emphasizes the need for a robust manual method to be selected for automation.

**Proposed sample preparation scheme** Knowledge of the physicochemical properties of the analyte can determine many of the techniques that can be used to develop the method, e.g., extraction using functional groups and interactions on the molecule and applicable detection systems. The processes involved in an automated assay should be as simple as possible. Complex automated assays requiring sophisticated online sample preparation and postcolumn reaction detection will take a long time to develop into routine and robust methods.

**Human skills** Never forget the human resource that is available to develop an assay. Do you have the right people to develop the assays you require?

Once these questions have been answered, an analyst is now in a position to make an informed judgment on the probability of success and the risks that are being taken when developing an individual automated sample preparation method. In fact, now the analyst has the information to decide when and why an automated method should be developed and perhaps, more importantly, why not and when not to develop such methods.

## **Factors in Design of a General Automated Sample Preparation System**

After an automated sample preparation method has been justified (by the criteria above), three factors

must be considered in the following order for the design of a suitable automated sample preparation system:

1. the consistency of the sample matrix,
2. easily automated sample preparation techniques, and
3. the appropriate automated instruments or systems.

Again, thought is required to design an automated sample preparation scheme for an individual analysis.

### **Matrix Consistency**

Sample matrices are liquid, semiliquid, or solid. The matrices that are easiest to incorporate in automated sample preparation schemes are liquid or semiliquid ones. As noted earlier, most analytical instruments require a liquid sample. Moving liquids, even relatively large volumes, can be easily accomplished by using syringes, sample loops, and low-pressure pumps. There may be a need to remove particulate matter from a sample prior to cleanup and this may be automated or done offline.

Semiliquid samples, such as plasma, can use to an extent the same techniques as liquid samples. Some samples usually need dilution to ensure that differences in viscosity between samples are minimized. Again, this can be done offline or by the sample processor.

Solid samples, varying from rocks, through tablets to biological tissues, can present a major problem for automation. The physical form of the sample can limit both the technique and the instrument that can be applied. This is an area that needs radical thought to solve the problems posed by the sample. For example, the automated microwave digestion of minerals still requires the crushing of the rock to be a manual offline process. To automate this and eliminate cross-contamination may not be easy or cost-effective.

### **Sample Preparation Techniques Amenable to Automation**

The major sample preparation techniques that are amenable to automation are solid-phase extraction, LC, dialysis, microwave sample preparation, flow injection analysis, and segmented flow analysis. Other sample preparation techniques, such as liquid-liquid extraction or ultrafiltration, may be possible to automate but may not be cost effective. This shortlist of amenable techniques may constrain an analysis, but there is a large body of experience in the literature to help a novice to use these procedures.

## Automated Instruments

Some automatic sample processors are also analytical instruments such as LC, flow injection analyzers, or autoanalyzers. There is a trend, wherever possible, to merge the sample preparation technique with the instrumental analysis. This allows greater efficiency and productivity. In addition, there are robots and sample processors that can be used to automate sample preparation.

Automated instruments can be classified into two main types, flexible and dedicated:

- Flexible automation can be reprogrammed but also reengineered to change the task that can be undertaken by it. This type of automation is typified by robotic arms.
- Dedicated automation is usually limited to a specific task, which may be varied by programming or parameterization of the basic commands of the program controlling the instrument. This can take the form of autosamplers, autoanalyzers, or batch processors, and these can offer a cost-effective means of automating an instrument assay. These autosamplers can automate techniques such as solid-phase extraction or dialysis.

## Classification of Automated Sample Processing

The operations comprising an analytical process have to be performed in the same sequence for each sample and in an identical manner. However, when a number of samples are to be analyzed, they can be either batch or sequentially processed. Refinements of these approaches to give shorter process times are parallel batch and concurrent sequential processing.

### Batch Processing

In batch processing, each single operation in the analytical process is carried out on each sample before the next operation is performed. Most manual techniques employ batch processing, due to the convenience of performing a single operation many times. The total time taken to process a group of samples is equal to the sum of the times for each unit operation in the analysis multiplied by the number of samples in the batch.

There are a number of disadvantages of batch processing. When the time taken to perform each operation are not identical then the time interval between the operations performed on each sample will be different, i.e., each sample does not receive identical treatment with respect to time. When the analytical process fails, all the samples in the batch are

lost. A high-priority sample cannot be analyzed preferentially since all the samples are treated in one batch.

### Sequential Processing

Each sample individually receives treatment from all of the operations before the next sample in the sequence is analyzed. Priority is given to the completion of all operations on a single sample and at any time only one sample is ever present in the analytical process. The total time taken to process a group of samples is equal to the sum of the times for each unit operation in the analysis multiplied by the number of samples in the sequence, i.e., the same as for batch processing. The disadvantage of sequential processing is that the timing of the operations must be very precise to ensure identical treatment of each sample.

### Parallel Batch Processing

When employing batch processing it is usual, in some of the operations, for the batch of samples to be treated in parallel, e.g., centrifugation, incubation with reagents, mixing, or passing liquid through solid-phase extraction columns with the assistance of vacuum or compressed gas. Parallel processing consideration reduces the total process time. The time saving increases with the number of samples in the batch.

Any errors due to variation in the length of treatment are eliminated because by parallel processing all the samples in one batch are treated simultaneously. It is important to ensure that the treatment actually starts at the beginning of the parallel treatment and finishes at the end of it. A further problem with parallel batch processing is monitoring the function of the operation. If an operation fails to function, all the samples are affected.

### Concurrent Sequential Processing

With concurrent sequential processing, as a sample passes along the chain of operations, further samples enter the system so that multiple operations function simultaneously. Once the system is running, the number of samples being treated at any one time is equal to the number of operations running concurrently and it is only these that will be lost in the event of a failure. The reduction in total processing time is proportional to the number of operations running concurrently and is independent of the number of samples being analyzed.

Of the four different types of processing, concurrent sequential processing requires the most complex system controls to coordinate the concurrent operations. Additionally, if the duration of the concurrent

operations is different, then, with respect to time, the first sample receives different treatment than do the subsequent samples. To speed up automated analysis, scheduling algorithms can be used.

## **Applications of Automated Sample Preparation**

Two applications will be described to illustrate automated sample preparation. The first will outline the automation of microwave dissolution of minerals, a solid matrix. The second will describe the use of online LC for the determination of urapadil and its metabolites in human plasma.

### **Robotically Controlled Microwave Dissolution of Minerals**

There are many ore samples that require analysis to guide mining operations and control the metallurgical processes for metal recovery. Despite advances in instrumentation, the sample preparation used is a lengthy manual acid digestion with nitric and sulfuric acids. Labrecque described a technique in 1988 for using a microwave oven to speed up the acid digestion of minerals. Showing the principles above, a manual microwave method was developed and automated using a commercially available robot arm. The robotic system took previously crushed and powdered ore samples and weighed an aliquot, added the appropriate acid reagents, capped the digestion vial, and then put the vial into a microwave oven. When the digestion was completed, the robot took the vial from the oven and placed it in a rack to cool.

This method illustrates a number of points on automated sample preparation. First, a new technique was needed to automate the preparation of a solid matrix. Second, not all of the methods could be automated in a cost-effective manner as the samples were powdered under human control. Finally, the sample preparation was standalone: the extracts were manually transferred to the analytical instruments.

### **Analysis of Urapadil in Plasma by Online LC**

This automated method involves the direct injection of a measured aliquot of plasma by an autosampler into a liquid chromatograph. A precolumn is used to trap the urapadil and its metabolites and wash the majority of the interfering plasma components to waste. A valve then switches the precolumn inline with the analytical column and the analytes are washed off the precolumn and separated by the analytical column. During the analysis, the precolumn is reconditioned to accept the next sample.

The sample in this method is a semiliquid and therefore it is easier to automate and to combine both the sample preparation and analysis into one operation. Furthermore, the cost of this system is little more than the additional switching valves over the chromatograph needed for a manual method. If the data system is combined with a laboratory information managing system (LIMS) virtually the whole of the analytical method can be automated.

## **Integrated Automatic Methods**

The principle that sample preparation should be performed automatically is almost universal and is independent of the instrumental technique. However, most so-called automatic methods are not truly automated – they are just mechanized. Most automated sample preparation systems have three factors missing: integration, equipment control, and communication.

Integration of sample preparation with the rest of the analysis is essential for full productivity gains to be achieved. This is achieved with some automated sample preparation techniques, e.g., LC column-switching illustrated above. The trend in automated sample preparation is clear: the technique will be online and part of the instrumental analysis. To use this approach, the sample must be liquid before the analysis can commence. Therefore, more effort needs to be spent investigating how to automate the sample preparation of solid samples.

The second factor concerns control of automated systems: what happens if there is a malfunction? At present, few automated sample preparation systems offer overall control. The probability, in the event of a malfunction, is that the autosampler will complete its cycle and valuable samples will be lost. Thus, developers must consider feedback control mechanisms that are essential to monitor the operation of a complex instrument that would be sampling, preparing, and analyzing simultaneously. If an error in this operation were detected and if the fault could not be corrected, the unit could be shut down, thus preventing the further loss of any samples. The development of these workstations able to make decisions may involve chemometrics, where there will be feedback between the analytical measurement and the experimental design.

The final factor is communication. Bidirectional communication with an LIMS is essential to produce an integrated sample preparation workstation for the analytical laboratory. Communications must be incorporated so that the workstation should be able to receive instructions such as the method and the



analysis conditions and the sample identities. When achieved, the checking for transcription error and administration for an analysis would be reduced drastically. When the analysis is complete, the LIMS receives a file with the sample ID and the corresponding analysis results for incorporation in the laboratory database. Failure to incorporate these facilities will mean that advances in automated sample preparation will be slow.

**See also:** **Liquid Chromatography:** Overview. **Quality Assurance:** Laboratory Information Management Systems. **Sample Handling:** Robotics.

## Further Reading

- Corkan LA and Lindsey J (1992) Experiment manager software for an automated chemistry workstation, including a scheduler for parallel experimentation. *Chemometrics and Intelligent Laboratory Systems: Laboratory Information Management* 17: 47–74.
- Huber R and Zech K (1988) Determination of drugs and their metabolites in biological samples by fully automated HPLC with on-line solid-liquid extraction and pre-column switching. In: Frei RW and Zech K (eds.) *Selection Sample Handling and Detection in High Performance Liquid Chromatography, Part A*, pp. 81–143. Amsterdam: Elsevier.
- Labreque JM (1988) Manual and robotically controlled microwave pressure dissolution of minerals. In: Kingston HM and Jassie LB (eds.) *Introduction to Microwave Sample Preparation: Theory and Practice*, p. 207. Washington, DC: American Chemical Society.
- Linder M (1990) Laboratory automation and robotics – Quo vadis? In: Karjalainen EJ (ed.) *Scientific Computing and Automation (Europe) 1990*, pp. 273–284. Amsterdam: Elsevier.
- McDowall RD (1989) Sample preparation for biomedical analysis. *Journal of Chromatography* 492: 3–58.
- McDowall RD (1992) Strategic approaches to laboratory automation. *Chemometrics and Intelligent Laboratory Systems: Laboratory Information Management* 17: 265–284.
- Meister Z, Sturgeon R (2004) *Sample Preparation for Trace Element Analysis*, Amsterdam: Elsevier.
- Moldoveanu SC and David V (2002) Sample preparation in chromatography. Amsterdam: Elsevier.
- Turnell DC and Cooper JDH (1989) Automation of liquid chromatographic techniques for biomedical analysis. *Journal of Chromatography* 492: 59–83.
- Wells DE (2003) *High Throughput Bioanalytical Sample Preparation*. Amsterdam: Elsevier.

## Robotics

**M D Luque de Castro and F Priego-Capote**,  
University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Even though pick-and-place robots had been in use for some time previously, commercial laboratory robots were first introduced at the 1982 Pittsburgh Conference by the pioneering firm Zymark Corporation. By 1994, robotics had seemingly reached maturity, so a specific nomenclature for laboratory robotics and automation was issued by the International Union of Pure and Applied Chemistry (IUPAC). Some of IUPAC's recommended terms are general and require the word 'robot' or 'robotics' for specific use (e.g., in 'controlled-path robots', 'corrosion-resistant robots', 'feedback in robotics', 'accuracy in robotics'); others are characteristic of robotic technology (e.g., 'arm', 'articulate structure', 'flexible automation', 'manipulator').

A robot (the name comes from the Czech word 'robota', which means work or servant) can be

defined as 'an automatically controlled, re-programmable, multipurpose, manipulative machine with several degrees of freedom, which may be either fixed in place or mobile for use in automation applications' or, more loosely, as 'a multipurpose machine which, like a human, can perform a variety of different tasks under conditions that may be unknown *a priori*'. The different types of robots currently available can be classified according to physical features such as hardware construction, degrees of freedom, coordinate system, or level of sophistication and technology.

The initially envisaged uses of a robot in the laboratory – which led several renowned manufacturers to fail in the past – have changed with time. At present, robots are devoted to things that users can hardly figure out how to do easily in any other way. These unique needs range from massive robotic systems for handling thousands of samples produced in high-throughput drug discovery and biochemical study programs to standalone devices that drone away at repetitive tasks.

Depending on both the number of tasks to be performed and their complexity, users can currently

analysis conditions and the sample identities. When achieved, the checking for transcription error and administration for an analysis would be reduced drastically. When the analysis is complete, the LIMS receives a file with the sample ID and the corresponding analysis results for incorporation in the laboratory database. Failure to incorporate these facilities will mean that advances in automated sample preparation will be slow.

**See also:** **Liquid Chromatography:** Overview. **Quality Assurance:** Laboratory Information Management Systems. **Sample Handling:** Robotics.

## Further Reading

- Corkan LA and Lindsey J (1992) Experiment manager software for an automated chemistry workstation, including a scheduler for parallel experimentation. *Chemometrics and Intelligent Laboratory Systems: Laboratory Information Management* 17: 47–74.
- Huber R and Zech K (1988) Determination of drugs and their metabolites in biological samples by fully automated HPLC with on-line solid-liquid extraction and pre-column switching. In: Frei RW and Zech K (eds.) *Selection Sample Handling and Detection in High Performance Liquid Chromatography, Part A*, pp. 81–143. Amsterdam: Elsevier.
- Labreque JM (1988) Manual and robotically controlled microwave pressure dissolution of minerals. In: Kingston HM and Jassie LB (eds.) *Introduction to Microwave Sample Preparation: Theory and Practice*, p. 207. Washington, DC: American Chemical Society.
- Linder M (1990) Laboratory automation and robotics – Quo vadis? In: Karjalainen EJ (ed.) *Scientific Computing and Automation (Europe) 1990*, pp. 273–284. Amsterdam: Elsevier.
- McDowall RD (1989) Sample preparation for biomedical analysis. *Journal of Chromatography* 492: 3–58.
- McDowall RD (1992) Strategic approaches to laboratory automation. *Chemometrics and Intelligent Laboratory Systems: Laboratory Information Management* 17: 265–284.
- Meister Z, Sturgeon R (2004) *Sample Preparation for Trace Element Analysis*, Amsterdam: Elsevier.
- Moldoveanu SC and David V (2002) Sample preparation in chromatography. Amsterdam: Elsevier.
- Turnell DC and Cooper JDH (1989) Automation of liquid chromatographic techniques for biomedical analysis. *Journal of Chromatography* 492: 59–83.
- Wells DE (2003) *High Throughput Bioanalytical Sample Preparation*. Amsterdam: Elsevier.

## Robotics

**M D Luque de Castro and F Priego-Capote**,  
University of Córdoba, Córdoba, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Even though pick-and-place robots had been in use for some time previously, commercial laboratory robots were first introduced at the 1982 Pittsburgh Conference by the pioneering firm Zymark Corporation. By 1994, robotics had seemingly reached maturity, so a specific nomenclature for laboratory robotics and automation was issued by the International Union of Pure and Applied Chemistry (IUPAC). Some of IUPAC's recommended terms are general and require the word 'robot' or 'robotics' for specific use (e.g., in 'controlled-path robots', 'corrosion-resistant robots', 'feedback in robotics', 'accuracy in robotics'); others are characteristic of robotic technology (e.g., 'arm', 'articulate structure', 'flexible automation', 'manipulator').

A robot (the name comes from the Czech word 'robota', which means work or servant) can be

defined as 'an automatically controlled, re-programmable, multipurpose, manipulative machine with several degrees of freedom, which may be either fixed in place or mobile for use in automation applications' or, more loosely, as 'a multipurpose machine which, like a human, can perform a variety of different tasks under conditions that may be unknown *a priori*'. The different types of robots currently available can be classified according to physical features such as hardware construction, degrees of freedom, coordinate system, or level of sophistication and technology.

The initially envisaged uses of a robot in the laboratory – which led several renowned manufacturers to fail in the past – have changed with time. At present, robots are devoted to things that users can hardly figure out how to do easily in any other way. These unique needs range from massive robotic systems for handling thousands of samples produced in high-throughput drug discovery and biochemical study programs to standalone devices that drone away at repetitive tasks.

Depending on both the number of tasks to be performed and their complexity, users can currently

choose between robotic stations and workstations. The former automate the entire processes; they often include facilities for sample insertion into analytical instruments and use the typical arm and an array of modules and peripherals to make tasks easier for the robot. Today, manufacturers design arms to function as ‘movers of stuff’ and operations such as transferring or filtering solutions are handled in modules around them. Workstations represent a newer vision of robotics. They are custom automation tools that assist workers in performing their job better and more efficiently. These stand-alone automated units are used for specific purposes such as handling liquids or preparing solid samples; they are growing in popularity as they can perform one to three tasks with reduced costs and sophistication. Some workstations can be interconnected or operated with robotic arms, while others use their own, dedicated arms (particularly to handle solid samples). One of the main differences between robotic arm systems and workstations is cost, the workstations being more economical, although this fact depends on its level of sophistication.

Because robotic technology continues to have some magical connotations in relation to laboratory automation, a number of manufacturers and users still use the words ‘robot’ and ‘robotic’ interchangeably to refer to both robotic stations and workstations. In addition, any instance of automation is also indiscriminately associated with robotics by many.

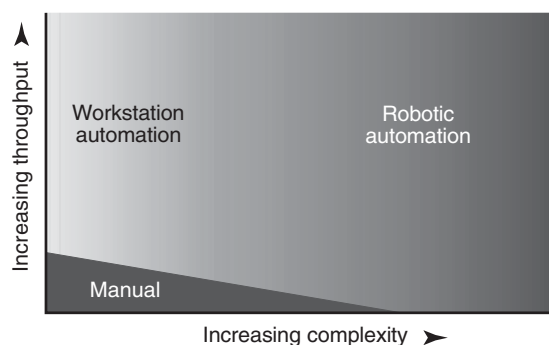
## Workstations, Robots, Modules, and Peripherals

The most salient difference between robotic stations and workstations is that, whereas a workstation can only be used for the tasks (all or some) for which it was constructed, robotic stations can be modified by changing their software, modules, or peripherals as required to undertake one or more specific tasks, or even a whole analytical process. As a result, describing a workstation is as simple as listing its intended functions, whereas characterizing a robotic station includes stating the type of arm it uses and the equipment that helps the arm perform its tasks.

Figure 1 depicts the scope of application of each option in terms of task complexity and throughput.

### Workstations

The most simple and common workstations are those for dilution and/or reagent addition to a number of samples in a simultaneous manner, either to all



**Figure 1** Fields of application of workstations and robotic stations according to task complexity and throughput. (Reproduced with permission of Elsevier Science.)

samples in a rack or to a line with a slide z-axis. Most workstations are designed to operate with liquid samples with specific equipment for liquid handling, solid-phase extraction, and preparation of liquid samples for insertion into chromatographs.

While considering solid samples, one alternative can be the use of workstations for automatic dissolution, which are very common in pharmaceutical dissolution testing. So, they can also be used to weigh and dissolve (or leach) some specific types of solids. However, there are workstations capable of automating solid sample treatments, including options as weighing, mixing, filtration, homogenization, and solid-phase extraction of samples for automatic insertion into chromatographs or transfer to other detection systems.

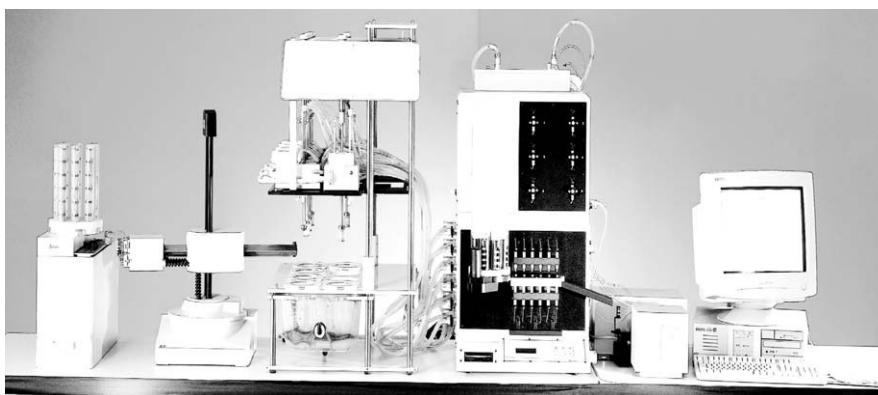
Figure 2 shows the tablet processing Zymark workstations, which was designed for treating solid samples.

### Robots

A laboratory robot constitutes the most flexible tool for automation as it can reproduce, with minimal adaptation, almost every task performed by an analyst.

A laboratory robot is essentially a machine consisting of various parts, namely:

1. The ‘manipulator’, which is a mechanism for grasping and/or moving objects, usually in several degrees of freedom. The essential parts of the manipulator are the body, the arm and a hand or end-effector.
2. The ‘controller’, an information processing device whose inputs are both desired and measured position, velocity, and other pertinent variables of the process. The robot can be controlled by the user, in a point-to-point manner and by a computer otherwise. The last are more common – ‘power



**Figure 2** Tablet processing II model Zymark workstations for solid sample treatment. (Reproduced with permission of Zymark Corporation.)

and event controller' (PEC) – and benefit from continuous improvement derived from the development of new software programming languages.

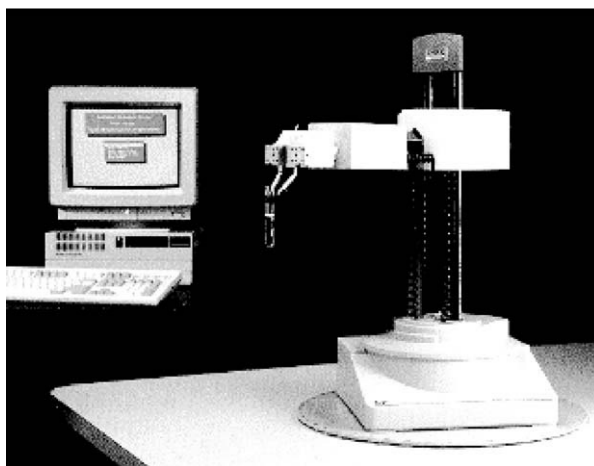
3. The 'power supply', which provides the energy – usually electrical or pneumatic energy sources – for locomotion of the different manipulator parts.
4. 'Sensors', which are used to boost performance in intelligent robots. They can be sorted according to type, physical and computational foundation, fixed or mobile and of the contact (force and tactile sensors) or noncontact type (vision, proximity, and range sensors).

One of the most important features of a robot, which determines its work envelope (i.e., the maximum extent and reach of the robot) is its position in the robotic station, which can be fixed or variable. The former is used in circular robotic stations, where the robotic arm is in the center of a circle and the different peripherals are included in a work envelope radius as removable pieces of a pie. An example can be seen in Figure 3.

A mobile arm in a robotic station is supported in a track that allows displacement of the arm in a length that varies, depending on the particular station, from 1 to 2 m. The capacity of the robotic station for locating peripherals in the work envelope of the arm can be expanded by using a more complex arm capable of operating on both sides of the track.

### Modules and Peripherals

The modules of a robotic station are the devices (apparatus, instruments, racks) used by the arm to perform its tasks. In circular configurations, the modules are referred to as 'peripherals'.



**Figure 3** Circular, PyTechnology, robotic station with fixed arm from Zymark. (Reproduced with permission of Zymark Corporation.)

The number of modules present in the work envelope of a robot arm varies with the number of tasks it is to perform in a given process. Some modules can be as complex as self-contained workstations. Although most of the modules required for the different steps of the process can be provided by either the arm's or an alternative manufacturer, some users design and construct their own modules, either because of the high specificity of the task or with a view to reducing costs.

The 'balance' is the most important module of a robotic station in dealing with solid samples. The only way of automating weighing is by using a robot or workstation. The balance must be connected to the PEC and be accessible to the robot arm, which should be furnished with a special end-effector or hand (a gripper hand) and a powder-pouring device (usually vibration based). Weight-based

measurements of liquids are frequently used in robotic stations as they provide more accurate data than do volume measurements. Zymark's new automated filter weighing system is an excellent example of the use of the balance with gaseous environmental samples. It provides automated and exact repetition of the analytical steps for compliance with EPA PM2.5 methods and uses no external air source in order to avoid introducing contaminants; also, filters are located before and after each weighing, and also ion-discharged before weighing.

A 'dissolve and dilute' module is a dual peripheral that includes a vortex mixer to facilitate dissolution and the obtainment of a homogeneous solution, respectively.

A 'master laboratory station' (MLS) is a module consisting of three syringes for dispensing liquids in conjunction with the dissolve and dilute module. However, the MLS can be used for additional purposes such as aspirating the phases involved in a liquid-liquid extraction (using a different syringe for each phase).

A 'centrifuge' is a key module when solid samples are to be leached or a precipitate forms during the process. The use of sensors to ensure correct positioning of the tubes with respect to the robot before and after the centrifugation step is mandatory here. Computer vision and neural networks can be used to detect errors in a robot system performing the automatic loading and unloading of a centrifuge.

'All-purpose hands' and 'syringe hands', available in a variety of designs, are also required elements of a robotic station. Differently sized objects (e.g., sample flasks, test tubes, probes, hold and press push-buttons) call for also different types of hand. A syringe hand facilitates the withdrawal of liquids from vessels.

'Liquid-liquid', 'solid-liquid', and 'liquid-solid' extraction modules are also devices required for implementation of a number of methods.

'Racks' can hold a variable number of test tubes, the positions of which are numbered so that the robot can distinguish them. In addition to holding the tubes, racks can accommodate devices such as the aspirator assay probe or even the sample probe if one is used.

Additional modules not always required in a robotic station include a 'capping-uncapping' module, used to remove and replace screw caps; a 'barcode reader', which is usually a laser barcode scanner combined with a turntable assembly capable of reading a label positioned anywhere around the circumference of a vial; and an 'ultrasonic bath', which is required not only for sonic mixing or cleaning,

but also, occasionally, to facilitate dissolution or leaching.

The performance of robots in some processes can be improved by using various approaches to expedite one or more steps thereof. One such approach involves adapting nonautomated devices for automated use in a robotic station; for example, the use of microwave digesters, which substantially accelerate solid sample pretreatment.

Robot control is a current research topic, which encompasses new hardware concepts for robot control, software environments for optimizing the productivity of robotic laboratories, and dedicated software for use in environmental laboratories. Finally, simulation and graphical animation of robots is highly useful with a view to both optimizing robot work and avoiding catastrophes on the workbench. Incorporating robotic stations into large-scale laboratories entails establishing appropriate links with a laboratory information management system (LIMS) to ensure precise control of the laboratory as a whole.

## **Detectors**

One of the essential units for completion of the overall analytical process by an automated system is the detector, which cannot be considered a module or peripheral as it preserves its identity as such and is only connected to the PEC for switching on, instrumental variable programming, and data collection. The robotic arm acts as a bridge between the other steps of the analytical process and detection, either by inserting the treated sample into the detector or by connecting the treated sample with the detector via fiber optics (occasionally furnished with a sensor) or electrical wire.

Photometric detectors are the most common choice for monitoring the end product of a robotic station. The easiest way to detect products here is by using an optical fiber to drive the beam from the light source to the solution under study, with passage through it and then back to the detector.

Luminescence molecular detectors have also been used for online monitoring of dissolution tests and the characterization of toxic residues using bioluminescence assays. Atomic (atomic absorption spectroscopy, inductively coupled plasma-atomic emission spectroscopy (ICP-AES)) detectors have been coupled to robotic stations either through a continuous system acting as interface or by direct aspiration into an instrument from a sample vial following treatment by the robot. Mass spectrometric and nuclear magnetic resonance (NMR) detectors

used in this context are also based on direct aspiration.

Electrochemical detectors (particularly probe sensors) are also widely used in robotic stations; for instance, the determination of soil pH by inserting a glass electrode into a test tube containing the soil leachate. A straightforward washing-rinsing station allows cleaning of the electrode, which is fouled in the measuring operation. A conductivity detector can also be used to determine soil salinity.

Finally, a balance can be used to develop a gravimetric method, for example, the determination of the oxidative stability of olive oil.

### Miscellaneous Considerations on Workstations and Robotic Stations

Two essential facts to be considered in purchasing a robotic station are that no single firm builds everything; there is a need to construct a robotic system tailored to one's needs, and compatibility between modules from different sources should thus be carefully checked prior to purchasing.

Procurement of supplies (pipette tips, autosampler vials, filters, etc.) and waste disposal are also two important issues in setting up and maintaining a robotic laboratory.

Because workstations can be designed and dedicated to a single, specific task, they are normally simpler mechanically and usually more reliable than are robotic stations. In addition, they can provide a data trail for regulatory compliance and are typically designed to be operated by non-experts. Commercial workstations vary in their level of sophistication, which allows the automation of even the simplest procedures.

Workstations operate best when there are one to three functions being automated; more operations or less defined analytical steps require the sophistication of robotic arm systems, which are more flexible and work best early on in research, when the methodology is not well defined.

Choosing between a robot plus peripherals and a workstation is a difficult task. From the beginning, automation held the promise of freeing analysts from cumbersome, time-consuming, repetitive tasks. In a quality control (QC) laboratory, workstations are typically the best solution as they are often more 'hardwired' and are better in QC laboratories, where the analytical steps are well understood and this equipment does save the laboratory's time and money. The best solution for implementing the complex treatments required by some solid samples is the sequential use of two workstations; when this is

impossible, a robotic station is the next-best choice in most instances.

## The Role of Robots in the Analytical Process

Automation of the analytical process by use of robotic equipment can extend from a single step to the whole analytical sequence. The number of steps that are robotized should be dictated by the user's experience and judgment, always as a function of the target process, costs, number of samples to be processed, etc. Straightforward single-task uses of robots, robotic sample preparation procedures, and fully robotized methods are discussed below, as are more rational uses in combination with other techniques intended to ensure optimum development of each step of the analytical process.

### Single-task Robots and Simple Uses of Robotics

One of the most important reasons of failure at the beginning of the laboratory robotics era, when no workstations were available, was the use of large robots to perform a single, repetitive task such as weighing, diluting or solid-phase extraction. One of the most common single tasks assigned to robots is weighing – the analytical step most difficult to implement using alternative automated approaches. Automating laboratory weighing may not increase the speed of weighing, but can free the valuable time of scientists and technicians.

Zymark has proposed simple and inexpensive workstations designed for specific tasks. Furthermore, some noncommercial systems for single tasks include a microwave digestion system for dissolution of Ti(IV) oxide and adaptations of workstations for special tasks such as the robotic–chromatographic method for the determination of glycosylated haemoglobin.

### Sample Preparation

Sample preparation is the most general application of both workstations and robotic stations as the tasks involved in this step of the analytical process are the most time-consuming, error prone, and difficult to develop by unskilled operators; in addition, safety restrictions apply when toxic materials are to be handled. The use of a specific approach depends on the number of steps involved and their complexity. **Table 1** summarizes the features of selected general and specific sample pretreatment procedures used in the environmental and clinical fields. Whereas most environmental samples subjected to a robotic treatment are solid, those dealt with by clinical



**Table 1** Use of robotic stations for sample preparation in environmental and clinical analysis

<i>Analyte(s)</i>	<i>Sample</i>	<i>Main steps</i>	<i>Subsequent steps</i>
Aflatoxin M1	Milk	SPE	HPLC, FI. det.
Aflatoxins	Peanut butter	Filt., Evap.	HPLC, FI. det.
Aldehydes	Surfactant	SLE, Filt.	HPLC
Atrazine, alachlor	Soil	SLE, Evap.	GC, N–P det.
Cations	Lichen	Leaching	CE
Characterization	Toxic residues	Mix., washing	pHmetric determination
Diene value	Fuels	LLE	Titration
Isocyanates	Adhesive	SLE, Filt.	SEC
Metals	Used oils		ICP-AES
Metals	Soil	SLE	FI manifold
Micropollutants	Water	SPE	GC-ECD
Nutrients	Environm. samples	Hot block digestion	Kjeldahl method
PCBs	Mineral oil	Mix., LLE	Pot. det.
Pesticides	Vegetables	SPE	GC-ECD
Phosphorus-31	Water	Mix.	NMR
Siloxanes	Hair	LLE	ICP-AES
Sulfur (mercaptans)	Fuels	LLE	Titration
TCP	Soil and fruit	SLE	ELISA
TCP	Urine	Hydrolysis	LLE, GC–MS
Trace element	Titanium dioxide	MD	–
BOD	Waste water	Filt.	DOM
Alosetron	Plasma and serum	SPE	HPLC, FI. det.
Amino acids	Pharmaceutical samples	SPE	SEC
Cocaine	Urine	LLE	GC–MS
Cocaine	Blood	SPE	GC–MS
Diclofenac	Human plasma	LLE	HPLC, Phot. det.
Drugs	Biological fluids	Centr., Evap.	HPLC
Drugs	Animal feed	SLE, Centr.	HPLC
Drugs	Plasma	Precipitation	LC–MS–MS
Felbamate	Human plasma	Mix., Centr., Evap.	HPLC, Phot. det.
Frenolicin B	Poultry feed	SLE	LC
Heroin	Heroin	Mix.	GC-FID
Hydrochlorothiazide	Human plasma	LLE, Evap.	HPLC
Mycophenolic acid	Human plasma	SPE	HPLC
Nefazodone	Human plasma	LLE	HPLC, Phot. det.
Porphyrins	Fecal samples	LLE	Reversed-phase HPLC
Ranitidine	Serum	SPE	HPLC, Phot. det.
Sulfadimidine	Serum	Imm.	Phot. det.
Terfenadine	Human plasma	SPE	HPLC, FI. det.
Theophylline	Human plasma	LLE, Centr.	HPLC, Phot. det.
Vitamin C	Foods	Centr. Filt. SPE	FI manifold
Vitamins A and B	Dairy products	LLE	HPLC

Filt., filtration; Evap., evaporation; FI. det., fluorimetric detection; SLE, solid–liquid extraction; SEC, size-exclusion chromatography; MD, microwave digestion; LLE, liquid–liquid extraction; Imm., immunoassay; Phot. det., photometric detection; FI. det., fluorimetric detection; BOD, biological oxygen demand; DOM, dissolved oxygen measurement; SPE, solid-phase extraction; Centr., centrifugation; Mix., vortex mixing; PCBs, polychlorinated biphenyls; Pot. det., potentiometric detection; TCP, 3,5,6-trichloro-2-pyridinol.

laboratories are typically liquid and highly complex in nature, as can be seen. The type of analyte to be isolated from a given sample and the main operations to be performed by either the workstation or the robot to prepare the sample for subsequent steps, as well as the main steps following the treatment, are also listed in **Table 1**. When the step following treatment of the sample is insertion into a highly discriminating instrument such as an NMR or MS/MS instrument, the robotic station is aimed at a 24-h working day of the high-price instrument in an

unattended manner, thus avoiding either the high maintenance personnel costs of continuous work or the purchase of another instrument. This task even justifies the use of a simple robotic station to insert samples with noncomplex matrices into the instrument.

Although workstations and even robotic stations commonly perform liquid–liquid and liquid–solid separations, these steps can be implemented in a continuous manner by using a more inexpensive setup and with shorter development times when the number of samples is not very high and the results

must be available within a short time. The unit operations that can be performed failure free by a robot include weighing, centrifugation, filtration, and special instances of solid and liquid transfer. On the other hand, reagent addition, dilution, heating, homogenization, derivatization, and insertion into a measuring instrument are generally more rapidly, economically, and efficiently performed by other automated alternatives, whether continuous or discrete.

### **Robotic Development of the Whole Analytical Process**

Most frequently, using a robotic station to develop an entire analytical process is unwarranted. However, well-established criteria exist to ensure correct use of the potential of robotic technology.

One well-justified use of a robotic station is in the gravimetric method for the determination of the oxidative stability of olive oil, where the station monitors the amount of oxygen absorbed by a sample heated under controlled conditions. The method was intended to relieve the typically heavy workload of olive oil analysis laboratories in winter and involves weighing the samples in test tubes that are heated for a few hours, allowed to cool and then weighed, the cycle being repeated as many times as required, depending on the stability of the oil, until the weight gain of all samples reaches a preset value.

Other example is the determination of fuel parameters including the diene value and mercaptan sulfur. First, the robot weighs the sample, refluxes it in the presence of maleic anhydride, and extracts the analytes into an aqueous phase, which is poured into the titration vessel, where the robot inserts the photometric or potentiometric probe. For the determination of mercaptan sulfur, the robot also weighs the sample and removes sulfide by precipitation with a  $\text{CdSO}_4$  solution and liquid-liquid extraction. Once the fuel is sulfide free, which is checked by using a photometric probe, it is poured into the titration vessel by the robot arm, which also plunges an Ag electrode prior to addition of the titrant (an  $\text{AgNO}_3$  solution). In both cases, the automated titrator acts as a module of the robotic station and is operated by the robotic arm.

### **Combining Robotic and Continuous Systems for More Reliable Development of the Whole Analytical Process**

The 'bridge' spanning the raw sample and reading of the analytical signal, which is a function of the analyte concentration, has traditionally been built around single batch, continuous, or robotic methodologies for driving the sample to the detector. Each

approach has its own advantages and disadvantages, and no single currently available choice for automation is the panacea. Combinations of a robotic station or a workstation with a liquid or gas chromatograph, or a capillary electrophoresis instrument, have been reported in which an autosampler was used as an interface, the robot arm either placing the vials in it or loading them (following positioning on the autosampler) with the treated samples.

Users experienced in flow injection (FI) methodology have proposed this continuous approach as the most easygoing friend of robotics in order to take advantage of both. Thus, the expeditiousness and simplicity of FI systems are offset by their inability to automate certain operations (some so essential as weighing); on the other hand, robots, which are the most powerful systems as regards the variety of processes they can develop, have an inherent relative slowness compared to FI systems in addition to a high purchase cost. However, the complementary features of the FI technique and robotics warrant their joint use with a view to exploiting their respective advantages while avoiding their shortcomings in isolation.

Combined FI-robotic systems can in principle be used in two ways, namely: (1) by having the two individual systems operate independently of each other, the user acting as an active interface between both, and (2) by setting up an integrated system where the two subsystems will interrelate to eliminate the need for human intervention. In the former approach, the robot is devoted to automating the complex preliminary operations the FI system cannot handle (e.g., placing partially treated samples in a rack for subsequent transfer to the FI system by the user); the FI system acts as an automated device for inserting samples into the instrument, with or without a prior separation and/or derivatization step. There are few reported instances of the combined use of an FI manifold and a robotic station operating independently. One application is the automation of a method for the determination of total vitamin C in foods. Here, the robotic station is used for homogenization of the sample, weighing, addition of an extractant, centrifugation, filtration, and cleanup through a  $\text{C}_{18}$  column. After this treatment, the sample is manually transferred to the FI autosampler. A derivatizing reaction is implemented along the FI manifold to obtain a fluorescent product prior to insertion into the spectrofluorimeter. Although not specifically stated, the information produced is also transferred manually between both systems. The previous example illustrates a way of transmitting information in offline coupled FI-robot systems,

namely: automatic processing by means of an online microcomputer or manual acquisition and treatment.

In the integrated approach, the FI manifold is a module of the robotic station that is used to insert the sample into the detector (also with or without a prior separation and/or reaction). This latter approach involves no human intervention, which is an obvious advantage with respect to the former.

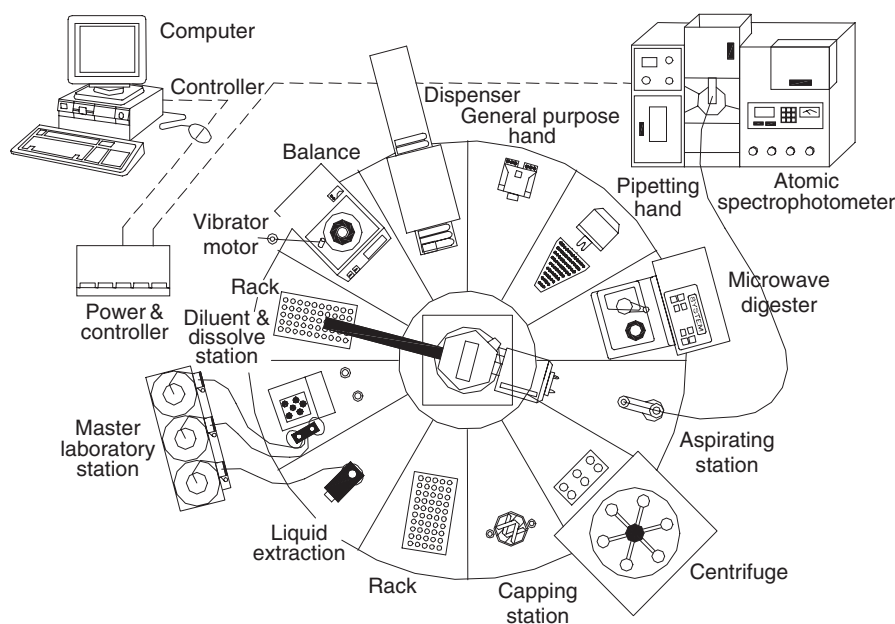
Even when the tasks to be conducted in the FI manifold are simple, coupling it to a robot provides some advantages. One example is the fully automated method for the determination of metals in soil. The use of a two-channel FI manifold to introduce robot-treated (weighed, microwave-digested, centrifuged) samples into the measuring instrument (e.g., an atomic absorption spectrophotometer) enables real and pseudodilutions with a dramatically broadened calibration range. In this way, the need for a prior estimation of the analyte concentration is avoided. In addition, such an expensive peripheral as a syringe hand is replaced with an inexpensive manifold, which reduces purchase costs. The overall operational setup for solid sample pretreatment plus determination is shown in Figure 4.

## Analytical Scope of Application of Robotics

Although robots can be used in virtually all analytical areas, its major applications encompass the clinical, pharmaceutical, and biotechnological fields.

The original robot arms gradually have replaced the well-known air-segmented continuous autoanalysers, mainly from Technicon in clinical laboratory and were eventually superseded by workstations for specific tasks such as pipetting, diluting, reagent addition (i.e., sample-handling tasks). Dedicated clinical laboratory workstations are used to receive samples in capped, barcoded vials in an indexing holding rack that the robot decaps in a decapper and introduces into the analyzer(s) as a function of the specific bar code. The use of robots in the healthcare sector is not limited to the clinical laboratory; in fact, it also encompasses tasks such as the assistance to surgical procedures, the elimination of mundane chores and the reduction of exposure of personnel to communicable diseases.

The pharmaceutical field has exploited robotic stations for the main purposes of dissolution testing and drug design. In fact, most of the procedures described in robotic manuals are concerned with the determination of drugs in pharmaceutical preparations. Robotic stations as such were initially also programmed for such a simple use as dissolution testing. Later, workstations for dissolution testing have flooded the market with specific designs in variable degrees of sophistication (e.g., for continuous monitoring through fiber optics, insertion into a chromatograph). In any case, robotic stations continue to be used to develop innovations in the pharmaceutical field. The impact of robotic high-throughput screening on drug discovery is a consequence of its ability to screen up to 100 000 samples



**Figure 4** Robotic station for fully automated determination of metals in soil. (---) Passive interface. (—) Tubing. Active interfaces of the computer to the robot and its peripherals are not shown. (Reproduced with permission of Elsevier Science.)

per day per instrument; this in turn results from the demand for new compounds, which is being met by combinatorial chemistry the most important pharmaceutical use of robotics at present.

Biotechnology is no doubt the most rapidly expanding field of application of robots. Most robotic biotechnological research work is conducted by workstations. Accelerated recombinant protein purification processes have been developed by using a workstation equipped with a fully integrated sample-holding robot that allows preparative cation-exchange chromatographic runs to be completed and the collected fractions analyzed by high-performance liquid chromatography (HPLC).

Various other fields benefit from the use of robotics. For example, the unstoppable advances in robotic technology have led to the development of a microrobotic arm constructed by using lithographic techniques that can be mobilized in an aqueous environment to pick up, lift, and reposition a 100- $\mu$ m glass bead. Agricultural laboratories are increasingly taking advantage of the use of robotics as seasonal overloads call for a 24-h working day without the involvement of temporary personnel. Thus, tasks such as weighing, leaching, filtering and measurement are frequently performed by robots in the determination of parameters such as pH, organic matter, conductivity, lime and phosphorus in soil, or bitterness in virgin olive oil.

## Present and Future Status of Robotics

Robotics as implemented in workstations and robotic stations has proved to be one of the most accurate and autonomous automated facilities, a fact that directly influences both the productivity and the quality of the analytical results it provides. However, despite the continuous advances, further expansion of robotic technology has been met with several serious hindrances, namely:

1. The scarcity or even lack of standards for checking and/or validating unit operations which led the official endorsement of a robotic method.
2. The high purchase cost of robotic stations, which makes them inaccessible to small laboratories – workstations are usually more affordable.
3. The need of professional assistance both from the manufacturer and dedicated specialists to keep robotic equipment operational.
4. The scarcity of literature on robotics, which is a consequence of the limited use of robotic equipment for research purposes outside companies while the results of private research teams are

rarely published in scientific journals. Moreover, propaganda publications magnifies advantages of robotics and hardly explains the underlying principle or working programs.

Current R&D efforts in robotics are being aimed at overcoming these weaknesses and provide clues for a number of strategies to be adopted in the future, namely:

1. Cooperative alliances between robotics firms and other manufacturing companies.
2. The current drivers for new robotic technology are the high-throughput screening and combinatorial chemistry markets.
3. Experts also expect automation to become more sophisticated through adoption of new technical achievements in both this area and computer science, which are being incorporated day by day to new equipment.
4. The simultaneous treatment of several samples in different analysis steps will be increasingly fostered in the future.
5. Robots will be driven to work in smaller dimensions (e.g., those of a 384-well plate) as they are still too large to deal with the increasingly smaller samples involved in clinical analyses. A saving of reagents will be one immediate result of this trend.
6. Internal and/or external sensors will be mandatory in order to construct robotic stations capable of performing new tasks (e.g., establishing when a solid sample has been completely dissolved, detecting the presence or appearance of a precipitate) and develop mobile robots. These can be either used in isolation or incorporated into robots or robotic stations in order to ensure traceability in the results by use of well-established barcoding systems.
7. The true revolution in analytical robotics will probably come from the development of artificial intelligence. Future robots will be managed by expert systems in such a way that they will be able to find solutions to analytical problems by using both the initial stored information and that obtained from the process it has developed previously. The use of artificial intelligence and neural networks to integrate laboratory automation will be one of the most exciting areas in future robotics research.

In this way, small laboratories will be able to programme their workstations to process different types of samples requiring also different treatments during 24-h working days.

See also: **Clinical Analysis:** Sample Handling. **Food and Nutritional Analysis:** Sample Preparation. **Gas Chromatography:** Overview. **Liquid Chromatography:** Overview. **Pharmaceutical Analysis:** Sample Preparation. **Quality Assurance:** Quality Control. **Sample Handling:** Automated Sample Preparation.

## Further Reading

Automated environmental lab combines robotics (1996) Software, *Chemical and Engineering News*, 14 Aug.  
García-Mesa JA, Luque de Castro MD, and Valcárcel M (1993) Coupled robot-flow injection analysis system for fully automated determination of total polyphenols in olive oil. *Analytical Chemistry* 65: 3540–3542.

Luque de Castro MD and Luque-García JL (2002) *Acceleration and Automation of Solid Sample Treatment*. Amsterdam: Elsevier.

Luque de Castro MD and Torres P (1995) Where is analytical laboratory robotics going? *Trends in Analytical Chemistry* 14: 492–496.

Nitzan D (1985) Development of intelligent robots: achievements and issues. *IEEE Journal of Robotics and Automation* RA-1: 3.

Velasco-Arjona A and Luque de Castro MD (1996) A robotic flow injection approach to the fully automated determination of starch in Fdo. *Analytica Chimica Acta* 333: 205–213.

Waldman H (1985) *Dictionary of Robotics*. New York: Macmillan.

# SAMPLING

Contents

**Theory**

**Practice**

## Theory

**R-Q Yu**, Hunan University, Changsha, People's Republic of China

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A chemical analysis is generally accomplished on only a small fraction of the material/population of interest. Only in some exceptionally rare cases are the specimens tested and the population of concern identical. Sampling is a necessary preliminary step practically in all analytical procedures. It is defined as the operation of acquiring an unbiased representative fraction (sample) from the entire population of material to be analyzed. The operation should provide a valid sample with minimum effort and cost commensurate with the precision required. The sampling step of the overall analytical scheme is of paramount importance and needs to be given considerable attention. The quality and utility of analytical data will not be better than the care exercised in sampling and depends critically on the validity of the sample and the adequacy of the sampling strategy. This article will briefly review the general statistical theory concerning various sampling

strategies as well as some refined theories modeling the heterogeneity pattern of the material to be analyzed and the sampling errors for bulk materials.

## General Considerations and Sampling Strategies

### Random Sampling

For random sampling, samples are simply taken at random from the whole population of the material. The only requirement of such a random sampling process is that samples are drawn in a way that all parts of the population have the same chance of being sampled. This can be done by means of a table of random numbers or by means of a computer program that produces pseudorandom numbers. If each of the  $n_s$  samples taken is analyzed  $n_a$  times, the overall variance of the whole procedure,  $\sigma_o^2$ , is

$$\sigma_o^2 = \frac{\sigma_s^2}{n_s} + \frac{\sigma_a^2}{n_s n_a} \quad [1]$$

where  $\sigma_s^2$  and  $\sigma_a^2$  are the variances of the material sampled and the analysis, respectively. Equation [1] can be used for the random sampling design. Assume that

$$\sigma_a^2 = \alpha \sigma_s^2 \quad [2]$$

See also: **Clinical Analysis:** Sample Handling. **Food and Nutritional Analysis:** Sample Preparation. **Gas Chromatography:** Overview. **Liquid Chromatography:** Overview. **Pharmaceutical Analysis:** Sample Preparation. **Quality Assurance:** Quality Control. **Sample Handling:** Automated Sample Preparation.

## Further Reading

Automated environmental lab combines robotics (1996) Software, *Chemical and Engineering News*, 14 Aug.  
García-Mesa JA, Luque de Castro MD, and Valcárcel M (1993) Coupled robot-flow injection analysis system for fully automated determination of total polyphenols in olive oil. *Analytical Chemistry* 65: 3540–3542.

Luque de Castro MD and Luque-García JL (2002) *Acceleration and Automation of Solid Sample Treatment*. Amsterdam: Elsevier.

Luque de Castro MD and Torres P (1995) Where is analytical laboratory robotics going? *Trends in Analytical Chemistry* 14: 492–496.

Nitzan D (1985) Development of intelligent robots: achievements and issues. *IEEE Journal of Robotics and Automation* RA-1: 3.

Velasco-Arjona A and Luque de Castro MD (1996) A robotic flow injection approach to the fully automated determination of starch in Fdo. *Analytica Chimica Acta* 333: 205–213.

Waldman H (1985) *Dictionary of Robotics*. New York: Macmillan.

# SAMPLING

Contents

**Theory**

**Practice**

## Theory

**R-Q Yu**, Hunan University, Changsha, People's Republic of China

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A chemical analysis is generally accomplished on only a small fraction of the material/population of interest. Only in some exceptionally rare cases are the specimens tested and the population of concern identical. Sampling is a necessary preliminary step practically in all analytical procedures. It is defined as the operation of acquiring an unbiased representative fraction (sample) from the entire population of material to be analyzed. The operation should provide a valid sample with minimum effort and cost commensurate with the precision required. The sampling step of the overall analytical scheme is of paramount importance and needs to be given considerable attention. The quality and utility of analytical data will not be better than the care exercised in sampling and depends critically on the validity of the sample and the adequacy of the sampling strategy. This article will briefly review the general statistical theory concerning various sampling

strategies as well as some refined theories modeling the heterogeneity pattern of the material to be analyzed and the sampling errors for bulk materials.

## General Considerations and Sampling Strategies

### Random Sampling

For random sampling, samples are simply taken at random from the whole population of the material. The only requirement of such a random sampling process is that samples are drawn in a way that all parts of the population have the same chance of being sampled. This can be done by means of a table of random numbers or by means of a computer program that produces pseudorandom numbers. If each of the  $n_s$  samples taken is analyzed  $n_a$  times, the overall variance of the whole procedure,  $\sigma_o^2$ , is

$$\sigma_o^2 = \frac{\sigma_s^2}{n_s} + \frac{\sigma_a^2}{n_s n_a} \quad [1]$$

where  $\sigma_s^2$  and  $\sigma_a^2$  are the variances of the material sampled and the analysis, respectively. Equation [1] can be used for the random sampling design. Assume that

$$\sigma_a^2 = \alpha \sigma_s^2 \quad [2]$$



Equation [1] can be rewritten as

$$\sigma_o^2 = \frac{\sigma_s^2}{n_s} + \frac{\alpha}{n_a} \left( \frac{\sigma_s^2}{n_s} \right) \quad [3]$$

From eqn [3], one can draw the following conclusions:

1. For given  $\alpha$ ,  $n_s$ , and  $n_a$ , the overall variance increases with increasing  $\sigma_s^2$ .
2. For given total number of analytical determinations,  $n_a n_s$ , and without cost-benefit considerations, the random sampling process should be planned with a reasonably large  $n_s$ . For example, six random samples with duplicate analyses would give better results than four samples each analyzed three times.
3. The overall variance is a linear function of  $\alpha$ . When  $\alpha$  is very small, that is, the variance of analytical determination is substantially smaller than the variance of the material sampled, the term  $\alpha/n_a(\sigma_s^2/n_s)$  would be negligible compared to the term  $\sigma_s^2/n_s$ . In such cases, there is little point in attempting to reduce the variance of analytical determination,  $\sigma_a^2$ . As a rule of thumb, when  $\sigma_a^2$  is one-third of or less than  $\sigma_s^2$ , further reduction of the analytical error by switching to expensive and sophisticated analytical procedures does not make much sense.

Practically, eqn [1] should be solved in combination with a cost-benefit analysis. Assume that the costs for sampling and analytical determinations are  $C_s$  and  $C_a$ , respectively, then the total cost,  $C$ , of the random sampling for  $n_s$  samples per  $n_a$  analyses should be:

$$C = n_s C_s + n_s n_a C_a \quad [4]$$

Combination of eqns [1] and [4] with subsequent differentiation would give the number of analytical determinations that should be carried out on each sample to minimize the total cost for fixed variances:

$$n_a = \frac{\sigma_a}{\sigma_s} \left( \frac{C_s}{C_a} \right)^{1/2} \quad [5]$$

The number of samples needed is

$$n_s = \frac{\sigma_s^2 + (\sigma_a^2/n_a)}{\sigma_o^2} \quad [6]$$

### Systematic Sampling

In systematic sampling, samples are collected on the basis of a given geometric or time pattern, for instance, at some regular intervals of spatial location or time. Sometimes the segregation of population

occurs and is known *a priori*, sampling should cover all parts of population more or less evenly, such as different depth in a tank of liquids. In these cases stratified sampling (*vide infra*) should really be used and systematic sampling would provide a quicker, more direct way as a compromise. The error and cost-benefit analysis of systematic sampling are similar to those of random sampling.

### Stratified Sampling

In stratified sampling the population to be examined is first divided into a number of portions called strata, followed by random sampling within each stratum. The overall variance,  $\sigma_o^2$ , is

$$\sigma_o^2 = \frac{\sigma_b^2}{n_b} + \frac{\sigma_s^2}{n_b n_s} + \frac{\sigma_a^2}{n_b n_s n_a} \quad [7]$$

Here  $n_b$  is the number of strata sampled,  $\sigma_b^2$  is the variance between strata,  $n_s$  is the number of samples per stratum, and  $\sigma_s^2$  is the variance within each stratum. The stratified sampling gives results as precise as random sampling when the whole population is homogeneous. When variance among strata is remarkable compared with variance within each stratum, sampling by stratified procedure is always preferred. Equation [7] has no unique solution for  $n_b$ ,  $n_s$ , and  $n_a$ . Compromises will be necessary, taking into consideration the relative cost of selecting a stratum  $C_b$ , sampling cost within each stratum  $C_s$ , and the cost of analytical determination  $C_a$ . The total cost of the process,  $C$ , is

$$C = n_b C_b + n_b n_s C_s + n_b n_s n_a C_a \quad [8]$$

When the variances are fixed, the optimum values of  $n_b$ ,  $n_s$ , and  $n_a$ , which minimize the total cost  $C$ , are

$$n_b = \sigma_b (\sigma_b C_b^{1/2} + \sigma_s C_s^{1/2} + \sigma_a C_a^{1/2}) / (\sigma_o C_b^{1/2}) \quad [9]$$

$$n_s = \frac{\sigma_s}{\sigma_b} \left( \frac{C_b}{C_s} \right)^{1/2} \quad [9a]$$

$$n_a = \frac{\sigma_a}{\sigma_s} \left( \frac{C_s}{C_a} \right)^{1/2} \quad [9b]$$

### Representative Sampling

The representative sampling as practiced in analytical chemistry is a stratified sampling that can provide an unbiased estimate of the population mean for the target component. For the case when the strata in stratified sampling are different in size and

in variance, in order to obtain an unbiased estimated population mean with minimum variance, the number of samples drawn from stratum  $k$ ,  $n_{sk}$ , should be taken in relation to its size,  $w_k$ , and standard deviation,  $(\sigma_s)_k$ :

$$\frac{n_{sk}}{n} = \frac{w_k(\sigma_s)_k}{\sum_k^{n_b} w_k(\sigma_s)_k} \quad [10]$$

For the case when all strata have the same standard deviations, eqn [10] simplifies to

$$\frac{n_{sk}}{n} = \frac{w_k}{\sum_k^{n_b} w_k} \quad [11]$$

Equation [11] states that the number of samples taken from different strata should be properly related to the size of the strata. Obviously, such samples cannot be taken by random sampling. Many analytical technical protocols prescribe the procedures for obtaining samples for specific purposes. These samples are considered by authoritative organizations as representative of specific objects. Such representative sampling procedures are usually designed with consideration of eqns [10] and [11]. Obviously, the estimate of population mean,  $\bar{x}$ , calculated from the stratum means,  $\bar{x}_k$ , weighted in relation to the size of strata will be an unbiased one 'representing' the given population:

$$\bar{x} = \frac{\sum_k^{n_b} w_k \bar{x}_k}{\sum_k^{n_b} w_k(\sigma_s)_k} \quad [12]$$

Sometimes the only information required from the analytical process is the average composition of the population. According to the principle of representative sampling, a composite sample can be prepared by some approved standardized protocols to provide the necessary information.

### Separation and Estimation of Variances Associated with Stratification, Sampling, and Analysis

Equation [7] contains variances  $\sigma_b^2$ ,  $\sigma_s^2$ , and  $\sigma_a^2$ , which are unknown. When analytical determinations are carried out for different strata, these variances will contribute to the overall variance  $\sigma_o^2$  simultaneously. Suppose again that for the  $n_b$  strata sampled,  $n_s$  samples are taken per stratum and each sample is analyzed  $n_a$  times. Let  $x_{kji}$  be the analytical result of the target component for  $i$ th determination of  $j$ th sample taken from the  $k$ th stratum. For each sample  $j$  from stratum  $k$  the mean of  $n_a$  determinations is  $\bar{x}_{kj}$ . One has  $n_b n_s n_a$  individual results. The estimate of  $\sigma_a^2$  or  $s_a^2$  can be obtained with  $n_b n_s (n_a - 1)$  degrees of

freedom:

$$V_a = \frac{\sum_k^{n_b} \sum_j^{n_s} \sum_i^{n_a} (x_{kji} - \bar{x}_{kj})^2}{n_b n_s (n_a - 1)} \quad [13]$$

$$V_a = S_a^2 \quad [14]$$

For  $k$ th stratum, one can write for  $n_s$  sample averages ( $\bar{x}_{kj}$ ,  $j = 1, \dots, n_s$ ):

$$V_s = \frac{\sum_k^{n_b} \sum_j^{n_s} (\bar{x}_{kj} - \bar{x}_k)^2}{n_b n_s (n_s - 1)} \quad [15]$$

where  $\bar{x}_k$  is the mean for stratum  $k$ .  $V_s$  is related with the estimates of  $\sigma_a^2$  and  $\sigma_s^2$ , which are  $s_a^2$  and  $s_s^2$ , respectively:

$$n_a V_s = n_a S_s^2 + S_a^2 \quad [16]$$

For  $n_b$  stratum averages,  $\bar{x}_k$ ,  $k = 1, \dots, n_b$ , and the grand average  $\bar{x}$ , one can write:

$$V_b = \frac{\sum_k^{n_b} (\bar{x}_k - \bar{x})^2}{n_b - 1} \quad [17]$$

$V_b$  is related with  $s_a^2$  and  $s_s^2$  together with the estimate of  $\sigma_b^2$  or  $s_b^2$ :

$$n_s n_a V_b = n_s n_a s_b^2 + n_a s_s^2 + s_a^2 \quad [18]$$

From eqns [14], [16], and [18], one can obtain the estimates of the variances associated with stratification, sampling, and analysis.

### Estimation of Minimum Number of Samples Using *t*-Statistics

So far the appropriate number of samples has been estimated in terms of true variances of the material sampled and the analysis. Only the estimates of the values of these true variances can be obtained from the limited number of samples and analyses. It is actually impractical to also obtain the true value of the population mean,  $\mu$ , from a limited number of analyses. The measured mean,  $\bar{x}$ , calculated from  $n$  times of analysis is close to the true value of  $\mu$  within certain confidence limits:

$$\mu = \bar{x} \pm \frac{t s_o}{\sqrt{n}} \quad [19]$$

where  $s_o$  is the estimate of overall standard deviation  $\sigma_o$ . Student's  $t$  is taken for a particular confidence level and  $(n - 1)$  degrees of freedom. The squared difference between  $\mu$  and  $\bar{x}$ ,  $(\mu - \bar{x})^2$ , is a measure of the overall error,  $e_o^2$ . With an analytical process including the sampling steps having an overall

standard deviation of  $s_o$ , the minimum number of analyses necessary to confine the overall error within  $\pm e_o$  is

$$n = \frac{t^2 s_o^2}{e_o^2} \quad [20]$$

In order to find out  $n$ , one needs the value of  $t$  for a given confidence level and  $(n-1)$  degrees of freedom. Since  $n$  is not yet known, the value of  $t$  for  $n = \infty$  and given confidence level can be used for calculating an approximate  $n$ .

By using the value of  $t$  for this approximate  $n$ , one can recalculate a more precise  $n$ . This iterative process is repeated until  $n$  converges to a constant value. In random sampling the number of analyses is  $n_a n_s$ . If the estimated standard deviation of the analytical measurement process,  $s_a$ , is negligible, and consequently the error of analytical measurement,  $e_a$ , is also negligible, one can write for the number of samples that must be taken in the sampling process,  $n_s$ , as

$$n_s = \frac{t^2 s_s^2}{e_s^2} \quad [21]$$

Equation [21] was derived for the case when the sample is very small in comparison with the whole population. For the case when the sample is a substantial part of the population, a so-called 'finite population correction' factor,  $(1 - n/N)^{1/2}$ , should be introduced. Here,  $N$  is the total number of samples contained in the whole population, the second term of the right-hand side of eqn [19] should be multiplied by this factor. Consequently, eqn [21] takes the following form:

$$n_s = \frac{t^2 s_o^2 N}{e_s^2 N + t^2 s_s^2} \quad [22]$$

Evidently, the finite population correction should be taken into consideration only for relatively small, finite population cases. It can be neglected whenever  $N$  is sufficiently large, as then  $(1 - n/N)^{1/2}$  approaches unity. Equations [21] and [22] can be used for estimating the number of samples necessary to confine the sampling error within  $\pm e_s$  with the confidence level chosen for the  $t$  value (95%, for example), provided the estimated variance within the sample unit is  $s_s^2$ .

Introduction of the  $t$  statistics and the concept of confidence limits makes the estimation of the number of samples needed for a specific purpose more realistic. Some expressions described in the previous sections should be modified when the true variances are substituted for their estimates and when the

confidence limits of the conclusions drawn from the corresponding calculations are considered. Equation [1] for random sampling, for example, can be rewritten as

$$e^2 = \frac{t^2 s_s^2}{n_s} + \frac{t^2 s_a^2}{n_s n_a} \quad [23]$$

This modified expression states that if one takes the  $t$  value for an appropriate degree of freedom and a confidence level of, say, 95%, one can be 95% confident that the overall error of the whole analytical process will not be greater than  $e^2$ . Similarly, expression [6] can be rewritten as

$$n_s = \frac{t^2 (s_s^2 + s_a^2/n_a)}{e^2} \quad [24]$$

## Sampling Constants

### Ingamells' Sampling Constant

Ingamells defined a sampling constant,  $K_s$ , for characterization of the homogeneity of a well-mixed laboratory sample. The relative standard deviation,  $R$ , expressed in per cent

$$R = 100 \times s_s / \bar{x} \quad [25]$$

is inversely dependent on the mass ( $w$ ) of the sample analyzed:

$$R^2 = \frac{K_s}{w} \quad [26]$$

The coefficient of proportion,  $K_s$ , which has the units of mass, is called the sampling constant. When  $w = K_s$ ,  $R = 1\%$ . As this expression contains  $R$ , which is related with the estimated value of standard deviation, one can imagine from the reasoning of the previous section that in eqn [26] there is an implicit  $t$  that equals unity. This is the  $t$  value for degree of freedom of infinity and confidence level  $\alpha$  of 0.3174. So the sampling constant is just the sample weight necessary to ensure a relative sampling error  $R$  of 1% with 68% ( $1 - 0.3174 = 0.6826$ ) confidence in a single determination:

$$K_s = R^2 w \quad [26a]$$

The square root of the  $K_s$  is numerically equal to the relative standard deviation of a set of results obtained from 1 g samples by a precise analytical method with negligible  $\sigma_a^2$ . This is one of the experimental approaches for determining  $K_s$ . Actually, one can estimate  $R$  for given  $w$  by repeated precise analysis using expression [25]. Once  $K_s$  is evaluated for the sample population, the minimum weight,  $w$ , needed for a maximum acceptable relative standard

deviation for this population can be estimated. This is an experimental method for determination of the minimum sample weight needed in the sampling process. It is possible to calculate this sample weight or the sampling constant itself by using some sampling theories (*vide infra*). If the population sampled is a well-mixed one, expression [26] holds fairly well. For extremely heterogeneous materials that are segregated or stratified, the value of  $K_s$  calculated for different batches would not be constant and might vary with  $w$ . This is one way for judging the degree of mixing of the population. For a segregated population, an additional segregation constant should be introduced.

### Visman's Sampling Constants

Visman developed the theory of sampling taking into consideration the effect of segregation. The experimentally estimated variance  $s_s^2$  is related to two sampling constants, a homogeneity constant  $A$ , which is similar to Ingamells' constant, and a segregation constant  $B$ :

$$s_s^2 = \frac{A}{wn_s} + \frac{B}{n_s} \quad [27]$$

where  $wn_s$  is the total mass of  $n_s$  samples. If the population is not segregated,  $B=0$ , one has the expression similar to eqn [26]. Considering eqn [25], one can find the relationship between Ingamells' sampling constant and Visman's homogeneity constant:

$$A = 10^{-4} \times \bar{x}^2 \times K_s \quad [28]$$

Constants  $A$  and  $B$  can be determined experimentally by collecting one series of small samples and one series of large samples. The sampling variances of these two series are estimated. The weights and variances of these two sets are used to set up two simultaneous equations of the form [27] and the values of  $A$  and  $B$  can be obtained by solving these equations. Equation [27] shows that the effect of the number of samples ( $n_s$ ) increases with the increasing value of  $B$ . A reasonably large  $n_s$  is preferred for reducing sampling variance. When  $B$  is zero, there is no segregation of the population and the number of samples does not affect the sampling variance as far as the total weight of the samples,  $wn_s$ , is fixed.

### Modeling the Heterogeneity Pattern and Sampling Errors for Bulk Materials

So far the sampling problem has been discussed from a general view by using statistical or experimental

methods. Actually, the sampling error for bulk materials is mainly caused by heterogeneities in the material under study. For a well-mixed gas or liquid, the homogeneity is of such a high degree that any small part of the whole will be really a representative sample of the population. A sample in the ordinary sense is too large and the precision of an analytical procedure is too low for the detection of the variation of composition of such homogeneous populations. The sampling problem arises when a small fraction is taken from a heterogeneous population. Naturally, modeling and statistical treatment of the heterogeneity pattern of the population to be sampled constitute the core of any sound theory of sampling from bulk materials.

### The Simplest Starting Model

Benedetti-Pichler drew an analogy between the sampling from bulk materials and counting the ratio of white and red beans in a silo containing hundreds of metric tons of a mixture of these two kinds of beans. As counting all the beans in the silo would involve years of work, one would switch to a more realistic, but less precise approach: take a sample from the silo and make the count for it. The sample size is determined by the precision desired and can be estimated by using the binomial distribution. A similar philosophy has been adopted in most modern theories for sampling from bulk materials.

Assume that the population is a mixture of density  $\rho$  of two kinds of cubic particles, A and B, of the same density. These particles are of the same size with a side of cube of  $u$ , the weight percentages of particles A and B in the population are  $w_A$  and  $w_B$ , respectively. As the simplest model, assume that only type A particles contain the sought-for component  $x$  of a weight percentage  $x_A$ . The weight percentage of the sought-for component in the whole population is  $x$ . One takes a sample of weight  $w$  from the population for chemical analysis. The sampling variance is  $\sigma_s^2$ . This model can be treated using the binomial distribution. For a binomial distributed variable, the variance  $\sigma^2$  is  $npq$ . In terms of the simplest sampling model under discussion, the number of particles ( $n$ ) in the sample is  $w/\rho u^3$ , the probability ( $p$ ) for taking at random a particle A is  $w_A\%$ , the probability ( $q$ ) for taking a particle B is  $w_B\%$  or  $(100 - w_A)\%$ . One can write for the  $\sigma_s^2$  expressed in numbers of particles:

$$\sigma_{s(\text{particle})}^2 = \frac{w \times w_A \times (100 - w_A)}{\rho u^3 \times 1000} \quad [29]$$

In terms of the percentage of the sought-for component,  $\sigma_{s(\text{particle})}$  should be divided by the number

of particles:

$$\sigma_s = \frac{\sigma_{s(\text{particle})}}{w/\rho u^3} \times 10^2 \quad [30]$$

The simplest case is that particle A is the pure sought-for component, e.g.,  $w_A = x$ , then

$$\sigma_s^2 = \frac{\rho u^3}{w} x(100 - x) \quad [31]$$

Equation [31] can be used to evaluate the sampling error for given sample weight or, conversely, estimate the minimum sample weight necessary for obtaining the desired sampling precision. This expression is based on an oversimplified model. For practical uses the model should be refined to make it closer to the analytical realities.

### Extension of the Model to More Realistic Cases

In the above model type A and B particles were assumed to have the same density. If type B particles have a density of  $\rho_B$  different from that of type A particles,  $\rho_A$ , and if not only type A particles contain  $x$ , but type B particles also contain the sought-for component and the weight percentage is  $x_B$ , then some corrections should be introduced to the simplest model:

$$\sigma_s^2 = \frac{\rho_A \rho_B}{\rho} \times \frac{u^3}{w} \times w_A(100 - w_A) \left( \frac{x_A - x_B}{100} \right)^2 \quad [32]$$

One notices that instead of the uniform density  $\rho$  of the whole population, the densities of particles A and B,  $\rho_A$  and  $\rho_B$ , respectively, are taken into consideration by replacing  $\rho$  with  $\rho_A \rho_B / \rho$ . The squared term  $(x_A - x_B/100)^2$  accounts for the difference in composition of two kinds of particles. It is obvious that the greater the difference between  $x_A$  and  $x_B$ , which is a measure of heterogeneity of the population, the greater the sampling variance.

Ingamells' sampling constant can be evaluated by comparing expressions [31] and [32] with eqns [30], [25]–[26a]. Ingamells himself gave several theoretical expressions for  $K_s$ , one was in the following form:

$$K_s = R^2 w \frac{(x_B \rho_B - x_A \rho_A)^2 (x_B - x)(x - x_A) u^3}{\rho_B (x_B - x)(x_B - x_A) + \rho_A (x - x_A)(x_B - x_A)} \times \frac{10^{18}}{x^2} \quad [33]$$

The expressions for binary mixtures can be extended to the case of 'more than two components'. The later case is complicated by the fact that since the sought-for component may be present in several types of particles, its variance will depend on the variances

and covariances of these various particles. The distribution of the component in a population has certain unusual statistical features in the application of bi- and multinomial statistics. For a mixture containing  $m$  types of particles of densities  $\rho_i$ , weight percentage in the population  $w_i$ , and weight percentage of  $x$  in the particle  $x_i$ , the expression for  $\sigma_s^2$  is given by Wilson's equation:

$$\sigma_s^2 = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \left( \frac{\Delta x_i \rho_i - \Delta x_j \rho_j}{100} \right)^2 \times \left( \frac{w_i w_j}{\rho_i \rho_j} \right) \times \left( \frac{u^3 \rho}{w} \right) \quad [34]$$

Here  $i$  and  $j$  refer to any two types of particles present out of  $m$  types,  $\Delta x_i$  is the difference between  $x_i$  and overall  $x$  content in the population. It is assumed that all the particles in the population are of uniform volume. When the particle densities differ greatly, eqn [34] is valid only if  $s_s$  is not too large.

### Introduction of Particle Property Factors: Gy's Theory

Gy's classical theory is a comprehensive theory for the sampling from heterogeneous materials developed on the basis of studying the sampling of granular materials. The simplest model expressed by eqn [31] can be modified according to Gy's sampling theory:

$$\sigma_s^2 = \rho u_3 x(100 - x) \left( \frac{1}{w} - \frac{1}{W} \right) f g l \quad [35]$$

In this expression, the term  $1/w$  of eqn [31] is replaced by  $(1/w - 1/W)$ . This replacement reflects the effect of the total mass of the bulk material,  $W$ , on the sampling variance. The sampling error decreases with increasing  $w$  and decreasing  $W$ . When  $w$  comes nearer to  $W$ , the sampling error approaches zero. When the size of population is much larger than the sample size,  $(1/w - 1/W)$  approaches  $1/w$ , the term used in eqn [31]. In expression [35], some additional parameters  $f$ ,  $g$ , and  $l$  are introduced. These parameters take into account the effects of the particle properties.

**Particle shape factor ( $f$ )** In derivation of eqn [31], all particles were assumed to be cubes of volume of  $u^3$ . In reality, particles may be different in shape. The factor  $f$  takes into account the deviation of the particle shape from cubicity. It is actually the ratio of the average volume of the real particles having a maximum linear dimension equal to the sieve size to the volume of the cube passing the same sieve. For the idealized case of eqn [31],  $f=1$ . If all particles are spheres of diameter  $u$ ,  $f=0.524$ . For most

populations  $f$  can be assumed to be 0.5, this is a better approximation than expression [31], which assumes  $f=1$ . For bulk materials with special shape such as particulate gold, the value of  $f$  might be as low as 0.2.

**Particle-size distribution factor ( $g$ )** For the simplest model of expression [31], all particles were assumed to have the same size. For a real mixture of particles the distribution of the size of particles should be considered and the factor  $g$ , the ratio of the upper size limit to the lower size limit, is introduced. For the idealized case of expression [31],  $g=1$ . When the size range is relatively large, such as the case when the bulk material is crushed and pulverized with no size classification, a value for  $g$  of 0.25 is recommended. For medium or small size ranges, the value for  $g$  may be taken as 0.50 or 0.75, respectively.

**Liberation factor ( $l$ )** For the simplest model it was assumed that the bulk material consisted of two kinds of particles, A and B. For eqn [31], it was even assumed that  $w_A=x$ . In real situations, it is possible that lumps of bulk material contain both kinds of particles. When the material is crushed, the particles that contain the sought-for component are liberated from the lump. This liberation process is taken into consideration by introducing the liberation factor,  $l$ , which is defined as

$$l = \left( \frac{u_1}{u_2} \right)^{1/2} \quad [36]$$

Here  $u_1$  is the size of the average grains of sought-for component in the lump and  $u_2$  is the size of the largest lump particle in the material. When  $u_2 \leq u_1$ ,  $l=1$ , no liberation process should be considered. Parameters for evaluation of the factor  $l$  may be estimated by microscopic examination and sieve tests.

**Composition factor ( $c$ )** Besides  $(1/w - 1/W)$ ,  $f$ ,  $g$ , and  $l$ , one can define a factor  $c$ , which is the contribution of the variation of the particle composition. Considering that the overall density  $\rho$  is a weighted average of densities of two kinds of particles,  $\rho_A$  and  $\rho_B$ , e.g.,

$$\rho = \frac{x}{100} \rho_A + \frac{(100-x)}{100} \rho_B \quad [37]$$

The factor  $c$  is defined as:

$$c = x(100-x) \left[ \frac{x}{100} \rho_A + \frac{(100-x)}{100} \rho_B \right] \quad [38]$$

Gy's sampling theory can be expressed by a very simple equation:

$$\sigma_s^2 = u^3 \left( \frac{1}{w} - \frac{1}{W} \right) c f g l \quad [39]$$

See also: **Air Analysis:** Sampling. **Clinical Analysis:** Sample Handling. **Food and Nutritional Analysis:** Sample Preparation. **Pharmaceutical Analysis:** Sample Preparation.

## Further Reading

- Gy PM (1979) *Sampling of Particulate Materials: Theory and Practice*. Amsterdam: Elsevier.
- Ingamells CO (1974) New approaches to geochemical analysis and sampling. *Talanta* 21: 141–155.
- Kratochvil B and Taylor JK (1981) Sampling for chemical analysis. *Analytical Chemistry* 53: 924A–938A.
- Laitinen HA and Harris WE (1975) Sampling. In: *Chemical Analysis*, 2nd edn, ch. 27, pp 565–582. New York: McGraw-Hill.
- Smith R and James GV (1981) *The Sampling of Bulk Materials*. London: Royal Society of Chemistry.
- Taylor JK (1987) *Quality Assurance of Chemical Measurements*, pp. 55–74. Michigan: Lewis.
- Yu RQ (1991) Analytical sampling theory. In: *Introduction to Chemometrics*, ch. 3, pp. 26–48. Changsha: Hunan Education Publishing House.

## Practice

**J G K Ševčík**, Charles University in Prague, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The purpose of sampling is to estimate the amount or properties of bulk material. Samples are taken

from a continuous stream, an individual lot, or a sequence of lots. A sampling standard is thus necessary because of the potential for numerous sources of variations within the bulk material, resulting from the sampling procedure employed, the preparation of composite samples, and measuring errors.

Sampling is a complex procedure for characterization of bulk material using individual



populations  $f$  can be assumed to be 0.5, this is a better approximation than expression [31], which assumes  $f=1$ . For bulk materials with special shape such as particulate gold, the value of  $f$  might be as low as 0.2.

**Particle-size distribution factor ( $g$ )** For the simplest model of expression [31], all particles were assumed to have the same size. For a real mixture of particles the distribution of the size of particles should be considered and the factor  $g$ , the ratio of the upper size limit to the lower size limit, is introduced. For the idealized case of expression [31],  $g=1$ . When the size range is relatively large, such as the case when the bulk material is crushed and pulverized with no size classification, a value for  $g$  of 0.25 is recommended. For medium or small size ranges, the value for  $g$  may be taken as 0.50 or 0.75, respectively.

**Liberation factor ( $l$ )** For the simplest model it was assumed that the bulk material consisted of two kinds of particles, A and B. For eqn [31], it was even assumed that  $w_A=x$ . In real situations, it is possible that lumps of bulk material contain both kinds of particles. When the material is crushed, the particles that contain the sought-for component are liberated from the lump. This liberation process is taken into consideration by introducing the liberation factor,  $l$ , which is defined as

$$l = \left( \frac{u_1}{u_2} \right)^{1/2} \quad [36]$$

Here  $u_1$  is the size of the average grains of sought-for component in the lump and  $u_2$  is the size of the largest lump particle in the material. When  $u_2 \leq u_1$ ,  $l=1$ , no liberation process should be considered. Parameters for evaluation of the factor  $l$  may be estimated by microscopic examination and sieve tests.

**Composition factor ( $c$ )** Besides  $(1/w - 1/W)$ ,  $f$ ,  $g$ , and  $l$ , one can define a factor  $c$ , which is the contribution of the variation of the particle composition. Considering that the overall density  $\rho$  is a weighted average of densities of two kinds of particles,  $\rho_A$  and  $\rho_B$ , e.g.,

$$\rho = \frac{x}{100} \rho_A + \frac{(100-x)}{100} \rho_B \quad [37]$$

The factor  $c$  is defined as:

$$c = x(100-x) \left[ \frac{x}{100} \rho_A + \frac{(100-x)}{100} \rho_B \right] \quad [38]$$

Gy's sampling theory can be expressed by a very simple equation:

$$\sigma_s^2 = u^3 \left( \frac{1}{w} - \frac{1}{W} \right) c f g l \quad [39]$$

See also: **Air Analysis:** Sampling. **Clinical Analysis:** Sample Handling. **Food and Nutritional Analysis:** Sample Preparation. **Pharmaceutical Analysis:** Sample Preparation.

## Further Reading

- Gy PM (1979) *Sampling of Particulate Materials: Theory and Practice*. Amsterdam: Elsevier.
- Ingamells CO (1974) New approaches to geochemical analysis and sampling. *Talanta* 21: 141–155.
- Kratochvil B and Taylor JK (1981) Sampling for chemical analysis. *Analytical Chemistry* 53: 924A–938A.
- Laitinen HA and Harris WE (1975) Sampling. In: *Chemical Analysis*, 2nd edn, ch. 27, pp 565–582. New York: McGraw-Hill.
- Smith R and James GV (1981) *The Sampling of Bulk Materials*. London: Royal Society of Chemistry.
- Taylor JK (1987) *Quality Assurance of Chemical Measurements*, pp. 55–74. Michigan: Lewis.
- Yu RQ (1991) Analytical sampling theory. In: *Introduction to Chemometrics*, ch. 3, pp. 26–48. Changsha: Hunan Education Publishing House.

## Practice

**J G K Ševčík**, Charles University in Prague, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The purpose of sampling is to estimate the amount or properties of bulk material. Samples are taken

from a continuous stream, an individual lot, or a sequence of lots. A sampling standard is thus necessary because of the potential for numerous sources of variations within the bulk material, resulting from the sampling procedure employed, the preparation of composite samples, and measuring errors.

Sampling is a complex procedure for characterization of bulk material using individual

samples. The individual steps in this process are as follows:

1. formulation of the terms of reference for the purpose of the sampling,
2. specification of the selection criteria (quality characteristics) for assessing the bulk material,
3. creation of a model of temporal and spatial variability of the properties of the assessed bulk material (quality variation), from which the sampling scheme follows,
4. selection of suitable instruments for taking individual samples and the operating requirements for sampling (the sampling procedure), and
5. requirements and instructions related to the methods and criteria for dividing the sample (sample preparation procedure).

The relationships and basic principles connected with sampling are given in **Figure 1**. The bulk material may consist of strata (mutually exclusive and exhaustive subpopulations considered to be more homogeneous with respect to the characteristics investigated than the total population) or of samples (subset of a specified population made up of one or more sampling units).

Sampling is undoubtedly the most important part of an analytical procedure; its importance is fully manifested in the concept of analytical uncertainty, and it fundamentally affects the magnitude of the expanded uncertainty of the method. Sampling

involves a well-defined method based on the experimentally observed properties of the evaluated bulk material. Standard processes offer various recommended sampling procedures, whose final form and scope form the subject of customer–contractor agreements and contracts. The results of sampling that is not a part of the quality system constitute only expensive information noise, which should be suppressed to the maximum extent possible.

Sampling is employed to ensure that a representative sample is obtained. The purpose of the sampling procedure is to ensure the expected probability of sampling, i.e., to fulfill the requirement that all the parts of the bulk material have the same probability that they will be sampled and that they appear in the sample for testing. Any deviation from this basic requirement may lead to unacceptable bias and the creation of a sample that is not representative.

There are various sampling procedures:

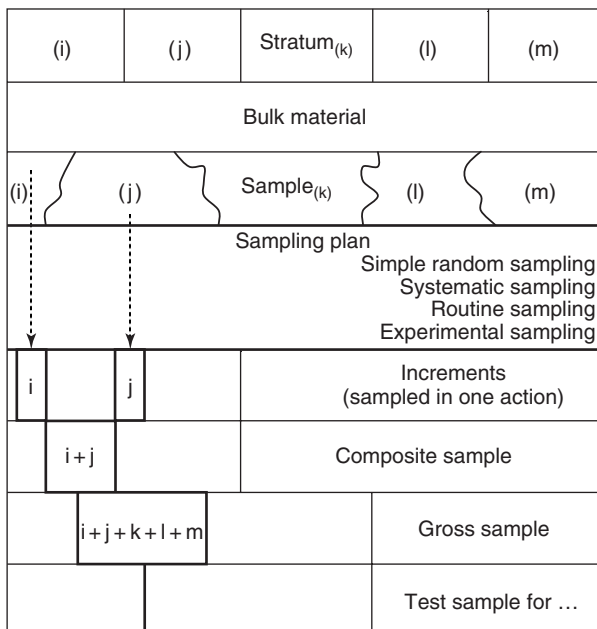
1. *Simple random sampling* – sampling where a sample of  $n$  sampling units is taken from a population in such a way that all combinations of  $n$  sampling units have the same probability of being taken.
2. *Systematic sampling* – sampling according to a methodical plan.
3. *Routine sampling* – sampling for commercial purposes carried out by the stipulated procedures in the specific international standard in order to determine the average quality of the lot.
4. *Experimental sampling* – nonroutine sampling where special-purpose experimental design is employed to investigate sources of variance and/or sampling bias.

In the framework of all the sampling procedures, an ‘increment’ sample is taken; the complex ‘composite’ and ‘gross’ samples are created from these increments, and from this, the ‘test’ sample is taken for a certain type of testing or analysis.

Technical standard regulations, including sampling standards, do not have legal force by law. Standard regulations are generally valid recommendations that may govern contract relations. The use of issued standards is voluntary, and their undoubted authority follows from their high technical level.

The scope and difficulties entailed in sampling, and the problem of preparing a general, unique sampling procedure for all potential cases have been discussed in the above. For a specific case of sampling, the sampling procedures are the subject of a contract agreement between the contractor and the customer.

National standards generally refer to international standards by stating that ‘This standard is the



**Figure 1** Basic terms in sampling of materials.

national version of International Standard ISO XXXXX-X:YYYY. International Standard ISO XXXXX-X:YYYY has the status of a 'National' Standard'. Under these conditions, mutual comparability of the recommended principles and procedures is achieved and thus contract regulation in exchange of goods and services is easier and less ambiguous.

A correct sampling scheme, i.e., a system of procedures and techniques specified by sampling plans, must facilitate determination of errors and individual steps in sampling and variations in the quality of the bulk material itself.

## Creation of a Sampling Scheme

The process of sampling is usually divided into three steps, where each step has its own variance:

1. the process of taking increments, with the variance caused during increment sampling,  $s_s^2$ ;
2. the process of sample preparation, with variance caused during test sample preparation,  $s_p^2$ ; and
3. the process of measurement, with variance characterizing the precision of the measuring method (analytical method) used,  $s_M^2$ .

The overall variability of the monitored characteristic for  $z$  individual samples and  $n_M$  repeated analyses is equal to  $s_T^2$  and it is given by

$$s_T^2 = \frac{s_s^2}{n} + s_p^2 + \frac{s_M^2}{n_M} \quad [1]$$

The whole of the bulk material is generally heterogeneous, and its mass is distributed in three-dimensional space. Consequently, in creating each sampling scheme, it is necessary, in the first stage itself, to determine the heterogeneity of the material in the framework of the total mass and its distribution in space.

The probability of exact determination of the quality characteristic, characterized by the variance  $s_T^2$ , increases with the amount of analyzed material (e.g., mass, volume, time, number of items). The ratio of the amount of analyzed material,  $m$ , to the total amount of material,  $m_T$ , is characterized as the sampling efficiency,  $\Phi$ :

$$\Phi = \frac{m}{m_T} \quad [2]$$

Here, the amount of analyzed material,  $m$ , can be expressed as the product of the number of samples,  $z$ , and their mass,  $m_i$ :

$$m = zm_i \quad [3]$$

$$s_T^2 = \left(1 - \frac{nm_i}{m_T}\right) \frac{s_s^2}{m} \quad [4]$$

where the heterogeneity of the material,  $s_s^2$ , is caused by two contributions:

$$s_s^2 = s_{s1}^2 + s_{s2}^2 \quad [5]$$

The first term of the right-hand side of eqn [5] can be designated as the short-term quality variation and is a consequence of the different compositions of the individual samples taken at successive, shortest possible intervals (20–40 individual samples). This short-term quality variation corresponds to the variance  $s_{s1}^2$ .

The second contribution is designated as the long-term quality variation and is a consequence of temporal and spatial changes during sampling. This long-term quality variation corresponds to the variance  $s_{s2}^2$ .

The results of repeated analysis of individual samples are given in the form of a variogram (the variogram is a plot of the variance as a function of the intervals between the original data values; the interval between consecutive data values is called lag one, that between every second data value is called lag two, etc.). It enables determination of the regression coefficients  $A$  (which is a function of  $s_{s1}^2$ ) and  $B$  (which is a function of  $s_{s2}^2$ ) and the variance of the evaluated material as a consequence of the segregation and aggregation,  $s_G^2$ , which is a function of the particle size,  $d$ . These quantities are further used in eqn [6] for the determination of the mass of the increment,  $m_i$ , and the number of increments,  $z$ , for sampling of bulk material with an overall mass  $m_T$ . It holds that

$$s_s^2 = \frac{Ad^3}{zm_i} + \frac{s_G^2}{z} + \frac{Bm_T}{6z^2} \quad [6]$$

These relationships indicate the broad variability of the sampling scheme. Economic aspects are the main criteria in decision making in selecting a scheme, both from the standpoint of the price of the sampled material (for expensive materials,  $m_i$  will be small and  $z$  large) or the price of the analysis (for expensive analyses,  $z$  will be small and the mass of the increments,  $m_i$ , will be large). It is also apparent that mechanical sampling leads to a large number of samples,  $z$ , and is especially suitable for bulk materials with a large overall amount,  $m_T$ .

## Experimental Sampling

Like the processes of sampling, the sample treatment and measurement are experimental procedures, and each of them is accompanied by an experimental error, which should have a random distribution, its magnitude should not significantly affect sampling. However, it may occur that the sampling scheme,

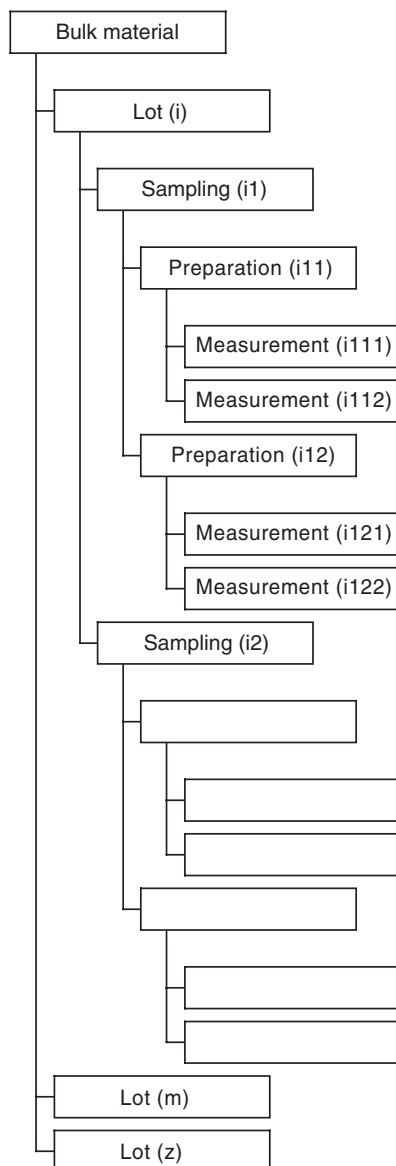
sampling procedure, or sample treatment employed will be accompanied by systematic errors. Such errors can be discovered and their magnitude can be determined only by comparative procedures, which form the basis of experimental sampling. The procedures of experimental sampling are generally based on the factorial design of the experiment.

Determination of the individual sources of errors in the sampling and heterogeneity of the material is made possible through creation of a suitable sampling scheme according to **Figure 2** and should precede systematic sampling.

The variability of the individual phases of sampling can be calculated on the basis of an experiment carried out according to **Figure 2**. **Table 1**

indicates the magnitude of variances in the individual phases. The values of variance determined in this procedure are then used for calculating the required sampling efficiency (eqn [4]) and for the required number of increments (eqn [6]).

The above principles are generally valid for materials in all states and for all applications. *A priori* simplification and the use of unverified values for the scattering variables for individual stages of sampling leads not only to erroneous determination of the magnitude of the quality characteristic but also to limited usefulness of the values determined and economic waste. As the creation of a sampling scheme is a key task, it is advisable to consult a professional for this step and ensure that the scheme is approved in the framework of contractor–customer relations.



**Figure 2** Full factorial design for experimental sampling.

## The Reason for Sampling

The creation of a sampling scheme always has a predefined purpose. For a given material, the sampling scheme will differ for the case of determination of the mean value of the quality characteristic of the material at the time of delivery (e.g., the water content in a delivery of coal) and for regulation of the quality characteristic of the material (e.g., the water content in coal used for coke production).

A large number of parallel standard regulations with similar internal structure exist for sampling of production processes and products. The politically highly topical area of the environment can serve as an example. Thus, one can find a standard for taking samples for measuring pollution levels in air, for stratified sampling for evaluating the quality of ambient air, for sampling for automatic determination of the mass concentration of gaseous components, and for the determination of the components of stationary emission sources, e.g., determination of the mass concentrations of PCDD/PCDF, HCl, sulfur dioxide, sulfur oxide, and sulfuric acid and the total content of sulfur oxides, ammonia, hydrogen sulfide, carbon disulfide, nitrogen oxide, fluorine,

**Table 1** Analysis of variances of sampling steps carried out in accordance with **Figure 2** for bulk material comprising  $z$  lots

Step	Source of variability	Degrees of freedom	Expected variance
1	Between lots	$z - 1$	$s_L^2 + s_S^2 + s_P^2 + s_M^2$
2	Sampling within lot	$z$	$s_S^2 + s_P^2 + s_M^2$
3	Sample preparation within sampling	$2z$	$s_P^2 + s_M^2$
4	Measurement within sample preparation	$4z$	$s_M^2$

etc. A similar situation exists for sampling of water and particulate matter.

It is obvious that the principles of the sampling scheme must be similar for all the above examples. The numerous one-purpose standards create the impression of separate, special sampling methods rather than a joint, global approach. In addition, a narrowly defined standardized approach can ignore the mutual interaction of the individual components of the bulk material and can thus lead to undesirable biased evaluation of the quality characteristic in question.

All sampling plans have a specified number of samples, sampling sites, type of sampling, and the necessary technical equipment and also describe the means of preserving the sample for subsequent analysis. However, specification of the site for taking individual increments requires a procedure that will differ from case to case. Knowledge of the properties of the material and its behavior during deposition is very useful in creating a sampling plan. When carrying out the sampling, a sampling protocol is filled out; this usually has prescribed information that must be entered by the person carrying out the sampling. Under these conditions, the sampling conditions are documented, and feedback to the description of the monitored quality characteristic of the bulk material is facilitated.

## Gas-Phase Sampling

The amount of component  $i$  in the gas phase is given by its partial pressure  $p_i$ , which is a function of the properties of the component monitored and the properties of other, simultaneously present substances and also of the ambient temperature and pressure. The composition of a gas phase is characterized by great complexity (all the chemical substances have the final vapor pressure value regardless of their polarity), instability (slow reactions take place between oxidizing and reducing substances that are simultaneously present), and low concentration levels (frequently below the detection limit in the *in situ* state). In addition to gases, the gas phase also contains suspended particulate matter and solid dust particles.

Ambient air pollution is understood to consist in human activity, during which one or more pollutants are introduced into the air. The pollution limit value is the highest permissible level of pollution of the air expressed in mass units per unit volume at normal temperature and pressure. The ambient air is most frequently sampled using methods of automatic continuous sampling, carried out by drawing the air directly into the analyzer or through a filter for analysis of dust and particulate matter (EN 12341:2000 *Air*

*Quality – Determination of the PM<sub>10</sub> Fraction of Suspended Particulate Matter*). In discontinuous sampling, sampling bottles and large-volume plastic sacks can be used.

The air quality in the working environment is generally regulated by law. The impact of dust, chemical substances, and biological factors is usually of concern. Several sampling techniques can be used here. These include analyzers for continuous sampling and measurement of the substances of interest, and also pumps and collectors utilizing adsorption, chemisorption, and dissolution to capture the substances to be determined (Table 2). These methods, in which the analytes are captured by a solid detector, are termed as solid-phase extraction methods. The method of personal passive sampling is used extensively in connection with the control of the working environment; here, the collector is located close to the nose and mouth of the worker.

When any type of collector is used, sampling leads to concentration of the individual substances and thus to an undesirable increase in the rate of reactions amongst them. In addition, when collectors are used, the homogeneous gas phase is changed into a heterogeneous environment with a certain level of chemical and surface activity.

Another problem connected with the use of collectors is their capacity to retain the sorbed substance in the collector during subsequent passage of analyzed air. There is a maximum volume of air passing through the collector for each substance in the given amount of sorbent at the given temperature and pressure, for which the sorbed substance will not be flushed out of the collector. This volume is termed  $V_{\max}$  and is specified for 1 g of sorbent and a temperature of 20°C; these values are mostly tabulated. This value can be determined experimentally using the gas chromatographic method. If the retention volume of the captured substance,  $V_R$ , in a chromatographic system with  $n$  theoretical plates is known, then

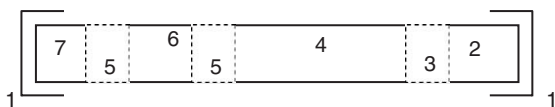
$$V_{\max} = V_R \left( 1 - \frac{2}{\sqrt{n}} \right) \quad [7]$$

These materials are usually placed in closed tubes, e.g., NIOSH (Figure 3). During sampling, the tubes are open and the sampled air is drawn through them. The volume of air drawn through the tube (of the order of liters) varies, depending on the analyte to be determined, as follows from eqn [7]. The material in the tube is divided into two parts. The first part of the detector is used to collect the sample, while the separate, connected, second part is used to control the sampling conditions (the determined analyte must not be present in the control part).

**Table 2** Sorption materials employed

Composition <sup>a</sup>	Name	Physical properties			
		Surface area (m <sup>2</sup> g <sup>-1</sup> )	Pore size (nm)	Limit (°C)	Polarity
Poly-S	Chromosorb 106	700–800	50	250	NP
Poly-S cross-linked	Chromosorb 103	15–25	3500	250	NP
S-DVB copolymer	Amberlit XAD-2	290–300	90	200	NP
	Amberlit XAD-4	750	50	200	NP
	Chromosorb 101	50	3500	275	NP
	Chromosorb 102	300–400	90	250	NP
	Porapak P	100–200		250	NP
EVD–DVB copolymer	Porapak Q	630–840	75	250	MP
CAN–DVB copolymer	Chromosorb 104	100–200	700	250	MP
	Separon CHN	650		220	MP
	Tenax GC (TA)	19–30	720	450	NP
Poly DPPO	Tenax GR	500	720	450	NP
DPPO + 23% graphite	Porapak S	450–600	76	250	SP
PVP	Porapak T	300–450	90	190	SP
EGDMA	Ambersorb XE-340	400	300		NP
	Carbosieve B	1000			NP
	Carbopack B	100	3000		NP
	Carbopack C	12	2000		NP
	Carbosphere	20	1200		NP

<sup>a</sup>S, styrene; DVB, divinyl benzene; EVD, ethyl vinyl benzene; CAN, acrylonitrile; DPPO, 2,6-diphenyl-*p*-phenyl oxide; PVP, polyvinyl pyridine; EGDMA, ethylene glycol dimethyl adipate; NP, nonpolar; MP, medium polar; SP, strongly polar.



**Figure 3** Schematic of air sampling tube. 1, sealing cap; 2, air inlet; 3, mechanical frit; 4, sampling compartment filled with adsorbent; 5, foam separator; 6, control compartment filled with adsorbent; 7, air outlet connected to pump.

A miniature variant of the above procedures consists in placing the sorption material on fibers. These collectors are called solid-phase micro-extraction devices and are suitable for sampling smaller volumes because of the smaller amount of sorbent.

## Liquid-Phase Sampling

Water sampling is similar to all special-purpose sampling and is divided into three large groups: sampling surface waters, sampling drinking water, and sampling waste waters. Standards of the EN 25667 and ISO 5667 series are the basic ones for water sampling.

Surface water is classified as:

1. water courses – flowing water (rivers, streams, and canals);
2. water reservoirs – still water (lakes, dams, and fishponds);
3. seas and oceans; and
4. glaciers and ice fields.

Sampling of surface waters consists in taking a representative part of the water body or other component of the environment (water, fluvial deposits, sediments, biological material) to determine various, exactly defined, quality indicators. This list indicates the great variability of sampling plans.

## Sampling Drinking Water

There are very detailed standards for sampling drinking water and the national systems are based on the above-cited international standards. Control of water is carried from water taps (intended for human consumption), at the outlets of wells and reservoirs, and at the sites of filling packages and containers. Drinking water is also sampled at the input and output of water tanks and water treatment facilities, and in water mains and swimming pools.

The manner of taking samples and treating them, including preservation, depends on the analyte to be determined, and, in general, on the quality characteristics. More than 40 quality characteristics are monitored in drinking water in relation to chemical compounds and physical properties, along with the biological characteristics and toxicity tests. This variability in the monitored quality characteristics corresponds to a large number of sampling plans, a significant part of which consists in sample treatment and preservation prior to subsequent analysis. The instructions thus encompass heating the tap outlet with a flame to disinfect it, rinsing the sample bottle



with the sampled water to suppress sorption from the sample on the walls of the vessel, and warnings about not shaking and not mixing the sample to prevent release of dissolved gases, etc.

The technical equipment for taking samples is determined by the sampling site and requirements on the number of analyses and analytical methods employed. The equipment includes bottles that are either closed and then opened at the sampling site (on the surface or at a certain depth) by opening the stopper (i.e., are not flow-through) or are open and rinsed by the surrounding water and then closed at the sampling site after a certain time. While the mechanical design of the sampling bottle is not decisive for the sampling, the chemical properties of the sampling bottle can lead to alteration of the sample. Consequently, clear and dark plastic (polyethylene, polytetrafluoroethylene, polyvinyl chloride, polyethylene terephthalate) or glass (hard glass, soft glass) bottles are used.

### **Sampling Waste Waters**

Waste waters are discharged into surface waters on the basis of a permit that specifies both their volume and their degree of pollution. The practice of sampling waste waters and sludges consists in monitoring indicators of the quality of treatment, management of treatment, and monitoring of the permitted limits. The sampling plans sometimes also prescribe the number and frequency of sampling operations, sample treatment, and statistical procedures for determining the monitored quality characteristic (for waste waters, the following are monitored: chemical oxygen demand determined by using the chromate method; undissolved substances; phosphorus; ammonia; nitrites; nitrates; absorbed organic halogenated substances; mercury; and cadmium). Because of the large content of solid substances in the sample, these are removed during preliminary treatment (homogenization and filtration). In contrast to surface and drinking waters, waste waters must be analyzed within 24 h of sampling. During sampling, potential infections must be taken into account, and thus gloves should be used during sampling, the sampling equipment should not be used for sampling other materials, etc.

### **Solid-Phase Sampling**

Materials in the solid phase are characterized by the highest degree of heterogeneity and thus they are most frequently sampled. Important groups of sampled materials include especially those that affect the

environment and human health. Consequently, great emphasis is placed on sampling wastes, ash from combustion of coal, welding sands, sludges from waste water treatment plants, soils and rubble, ash from waste incinerators, fishpond sediments, etc. from the standpoint of evaluating their hazardous properties, e.g., by determining their extractability or determining pollutants in dry matter. Sampling in agriculture, of agricultural products, plants, soils and fertilizers, genetically modified organisms, etc. is not only justifiable but is necessary for analyzing food-stuffs and the entire food chain. There is no doubt about the necessity for analyzing industrial materials in the form of individual items or free-flowing materials.

All these groups of materials are characterized by a high degree of heterogeneity of the monitored characteristics with respect to the spatial and temporal changes during sampling. The sampling plans for solid materials are most complex because of these properties of the bulk material.

Solid materials are frequently sampled using methods of continuous measurement on conveyor belts, in pipeline conveyors, etc. Under these conditions, the efficiency of sampling according to eqn [2] is high and the variance,  $s_{S2}^2$ , caused by spatial and temporal changes between subsequent sampling of increment samples is small; the monitored changes in quality correspond to the heterogeneity of the material,  $s_{S1}^2$  (eqn [5]). Manual, discontinuous sampling methods have very low sampling efficiencies and lead to greater variability in the monitored quality characteristic according to eqn [4].

*See also:* **Air Analysis:** Sampling. **Sampling:** Theory. **Water Analysis:** Overview.

### **Further Reading**

- Codex Alimentarius Commission Joint FAO/WHO Food Standards Programme, Codex Committee on Methods of Analysis and Sampling (2002) *General Guidelines on Sampling*, proposed draft, 24th Session Budapest, Hungary, 18–22 November 2002.
- Gy PM (1982) *Sampling of Particulate Materials – Theory and Practice*, 2nd edn. Amsterdam: Elsevier.
- Gy PM (1992) *Sampling of Heterogeneous and Dynamic Material Systems – Theories of Heterogeneity, Sampling and Homogenizing*. Amsterdam: Elsevier.
- ISO (2003) ISO 11648-1:2003 *Statistical Aspects of Sampling from Bulk Materials – Part 1: General Principles*.
- ISO (2001) ISO 11648-2:2001 *Statistical Aspects of Sampling from Bulk Materials – Part 2: Sampling of Particulate Materials*.

# SEDIMENT

See **GEOCHEMISTRY: Inorganic; Sediment**

## SEGMENTED FLOW ANALYSIS

**S Coverly**, Norderstedt, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Segmented flow analysis (SFA) was a pioneering development in laboratory automation. It was conceived in the 1950s by Leonard Skeggs, a researcher in a clinical laboratory, and within a decade of its development, became the dominant technique in automated clinical analysis.

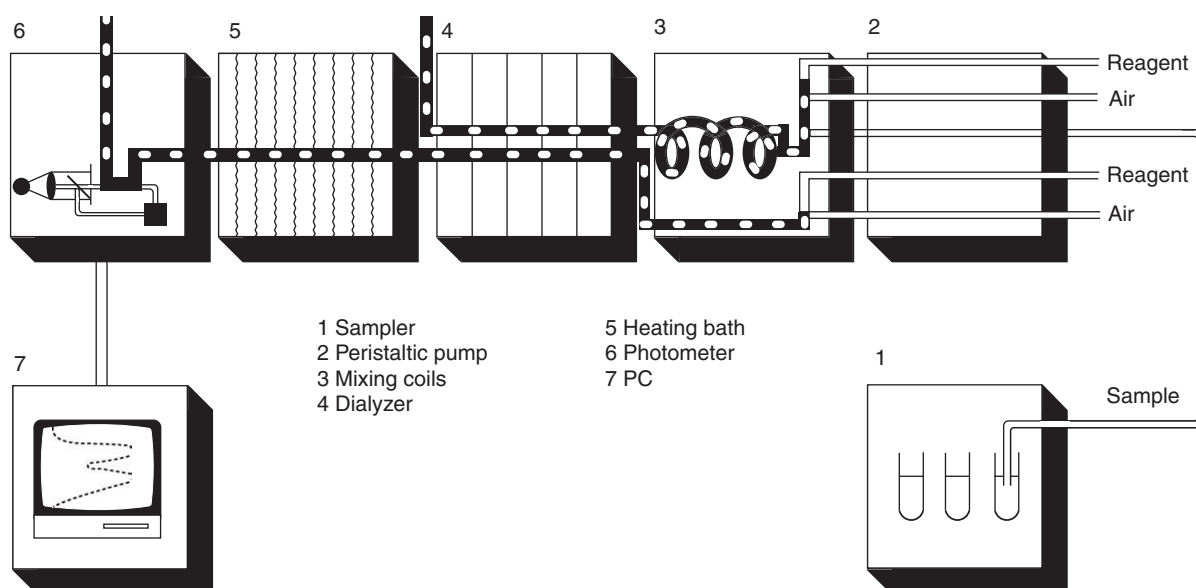
The rapid acceptance of SFA was due to its simplicity. Developed in the age before microprocessor control, when automating even simple procedures required complex mechanical timing and control devices, the low cost and high reliability of a flexible system employing only two moving parts – a peristaltic pump and an autosampler – was a major advance. The versatility of the system led to rapid development of methods and techniques by the manufacturer and by users in research institutions and in

industry, and by 1975 almost 8000 papers describing SFA and its applications had been published.

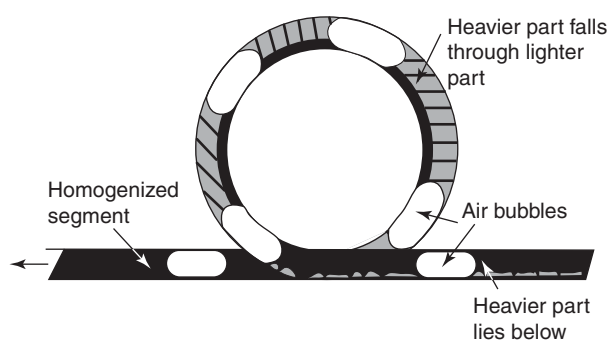
The use of SFA reached its peak in the mid- to late-1980s. Its use in clinical laboratories has now declined due to a move toward more highly automated systems requiring less operator involvement and with lower reagent consumption – an important factor for enzymatic reactions. In the industrial field, flow injection analysis (FIA), developed in 1975, is replacing SFA in small laboratories, but SFA is still dominant in larger laboratories, especially for determinations using long or complex procedures.

### Techniques

As shown in **Figure 1**, an SFA system consists of five basic components: sample changer, peristaltic pump, analytical manifold, detector, and data output system. The manifold contains the mixing coils, dialyzer, heater, and other components through which sample and reagents are pumped and where sample clean-up and reaction takes place. The data output



**Figure 1** Diagrammatic view of an SFA system using a dialyzer, heating bath, and colorimetric detection.



**Figure 2** Mixing coil, illustrating how liquids of differing densities are mixed by the vertical orientation of the coil.

system was originally a chart recorder, from which peak heights were measured and results were calculated by hand: recorders were later superseded by computers.

**Figure 2** illustrates the passage of liquid and air through the mixing coils in the analytical manifold. Sample and reagents are introduced through individual tubes in a continuously operating peristaltic pump. Samples follow one another from the auto-sampler; the only change in conditions within the system is thus the presence or absence of a sample. The air bubbles form barriers to longitudinal dispersion and separate the samples from each other and from the intersample wash solution. The bubbles are normally injected every 1–2 s, and the sampling time is typically 30–90 s. Each sample is thus divided into many segments, so that the small amount of carryover from one segment to the following one, which takes place via the thin film of liquid wetting the tubing surface, does not have any consequences in the segments toward the end of the sample. Indeed, the sampling time is usually selected so that at the end of the sample period there is a negligible change in concentration from one segment to the next. This results in a peak whose concentration at the end is not changing with time – a so-called ‘steady-state’ condition.

This low dispersion as samples pass through the system allows reactions lasting several minutes to take place in the flowing stream while maintaining a high sampling rate.

## Mixing and Reaction

Mixing takes place within each liquid segment due to both the internal motion of the liquid as it travels through the tubing and, where the liquids are of different densities, to gravity as segments are inverted during their passage through the mixing coils (see

**Figure 2**). Extending the length of the coil beyond that needed for complete mixing enables longer reactions to come to completion, a process that can be accelerated by heating the reaction mixture up to 95°C. Some special methods run at temperatures above 100°C, in a pressurized manifold.

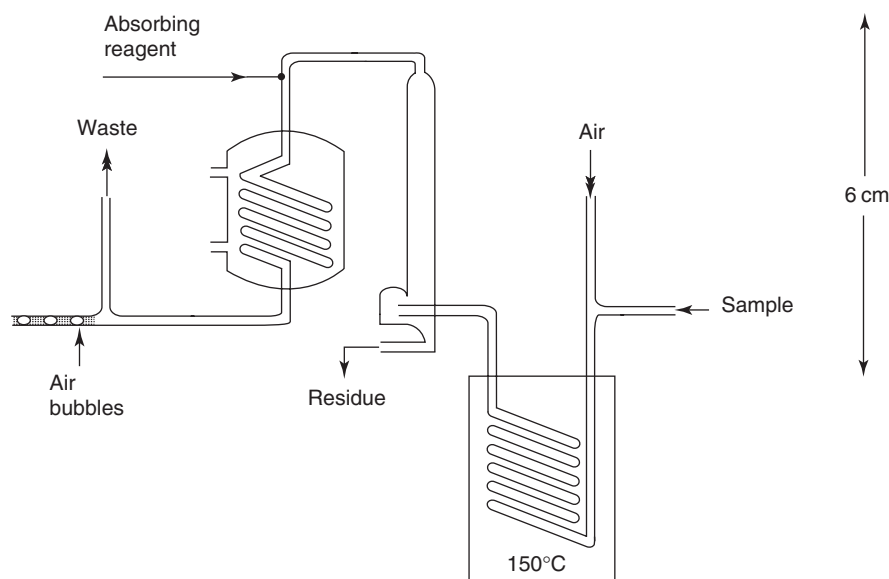
A further development of mixing and reaction is ultraviolet (UV) digestion. Here a quartz coil is wrapped around a UV lamp so that the sample stream is irradiated. Depending on the type of lamp and the spectral transmission characteristics of the coil, this can cause total breakdown of complex compounds, such as in the degradation of organic molecules in the presence of an oxidant for the subsequent measurement of total N, C, or, with milder conditions, partial decomposition as in the breakdown of complex cyanides to HCN.

## Dialysis

Dialysis was the first technique other than simple mixing to be developed for SFA. As shown in **Figure 3**, a donor stream consisting of the sample, normally diluted with a buffer or other reagent, passes over the surface of a permeable membrane made of cellulose acetate or similar material. A segmented stream of recipient reagent flows along the other side of the membrane (**Figure 1**, module 4). Ions and molecules with relative molecular mass less than ~1000, depending on the membrane type, diffuse through the membrane, while larger species such as fat and proteins remain in the donor stream. In a dialyzer of sufficiently long path length, equilibrium is eventually reached between donor and recipient streams. In practice, much shorter path lengths are usual, and the recovery of the wanted species is typically between 1% and 20%. Recovery depends on the path length, ionic strength, flow rates of the donor and recipient streams, temperature, and membrane type.

Dialysis was an important factor in the success of SFA in clinical applications, as it eliminated interference from blood lipids and protein, which formerly had to be removed by time-consuming manual procedures. In industrial applications, dialysis is commonly used to separate interfering materials such as those found in wastewater, and as a simple means of diluting high-concentration samples.

A variant of dialysis uses a membrane made of polytetrafluoroethene (PTFE), silicone rubber, or other hydrophobic material. These membranes allow the passage of gas while remaining impervious to liquids and dissolved solids. They are used in the determination of CO<sub>2</sub>, NH<sub>3</sub>, HCN, and other gases, either directly in the sample or following a reaction



**Figure 3** Online distillation in an SFA system.

to liberate them from other compounds present in the sample. An example is the enzymatic determination of lysine in food extracts. In this method, lysine decarboxylase degrades the amino acid to  $\text{CO}_2$ .  $\text{CO}_2$  in the donor stream diffuses through a gas-permeable membrane mounted between the plates of a conventional dialyzer into the recipient stream containing weakly buffered phenolphthalein, causing a color change that is measured photometrically.

## Distillation

This lengthy manual procedure is automated with SFA by passing the sample stream through a coil in a high-temperature heating bath, typically at 150°C (Figure 3). A nonvolatile liquid such as sulfuric acid or glycerol is added to the sample before distillation to maintain a liquid coating on the inner surface of the coil as the water evaporates. The vapor passes into a separation column and from there to a condenser, while the involatile residue is pumped to waste. After condensation, the sample stream is pumped into the analytical manifold for subsequent reaction.

Distillation is used to separate impurities, especially in the analysis of phenol and fluoride. If sulfuric or phosphoric acid is used as the distillation reagent, the combination of high temperature and the high acid concentration inside the distillation coil after most of the water has evaporated can also be used to decompose compounds such as polyphenols and complex cyanides.

## Solvent Extraction

When two immiscible liquids pass through a continuous-flow system they form alternating segments analogous to the gas-liquid segments in conventional SFA. The intrasegment motion as the segments pass along the manifold results in counter-current extraction at each phase interface, and solutes generally reach equilibrium between the two phases within 2 min. Phase separation normally relies on differences in density between the phases. Air segmentation is not usually necessary, although it is frequently used as a way of preventing small segments of each liquid phase from coalescing into larger segments that would hinder the extraction process and be more difficult to separate when their volume approaches that of the phase separator. When no air is present separation can take place through a hydrophobic membrane.

One application of solvent extraction in SFA is the determination of anionic surfactants in water, where methylene blue ion pairs are extracted into chloroform. Another is the determination of histamine in blood serum; here, interference is removed by extraction into butanol and back-extraction into an aqueous phase prior to the measurement of histamine by the formation of a fluorescent compound with *o*-phthalaldehyde.

Other techniques have been developed for SFA to solve specific problems. In continuous filtration the sample stream is pumped onto the upper surface of a moving strip of filter paper and the clarified filtrate is withdrawn by suction from the underside. Another

unit, now obsolete, was developed for continuous Kjeldahl digestion; the sample and digestion mixture were pumped into a rotating glass helix placed above heating elements. In practice, few compounds were completely decomposed in the short reaction time that could be achieved.

## Detectors and Peak Measurement

The first detector to be used for SFA was a photometer, and photometric determinations still form the vast majority of current methods. Other detectors in common use are UV spectrophotometers, used primarily for pharmaceutical compounds and for bitterness in beer; flame photometers, for potassium and sodium determination; fluorimeters, used primarily for measuring low levels of determinants in the presence of interferences, such as the determination of histamine in blood, and vitamins in food extracts; and ion-selective electrode and pH detectors. In principle, almost any detector with flow-through capability can be used with SFA systems, and determinations based on densitometry, thermometry, and luminescence have been published, among others.

Peak height is used almost exclusively as the basis for measurement in SFA. Commonly, the sample aspiration time is chosen so that the response in the detector reaches a constant or nearly constant value, giving a peak with a flat top. The signal from this steady-state section of the peak is averaged to give the final result. If the system has an irregular response, the sampling time can be extended to prolong the steady state, allowing irregularities due to hydraulic effects of unwanted solid particles or air bubbles in the stream to be eliminated. This mode of peak reading gives more accurate results than peak height measurements from sharp peaks such as those produced by chromatography or flow injection systems, and does not require the signal to return to near the baseline as does area-based calculation.

Evaluation of the peaks from a long run can follow up to four stages. First, the baseline at the beginning and end of a series of peaks is measured and the calculated baseline position for each peak, assuming linear drift, is subtracted from the overall peak height to give a baseline-corrected value. Second, sample interaction or carryover may be calculated from a set of known peaks and the contribution to each peak from the previous one subtracted. Third, if standards of known concentration have been repeated throughout the run, they can be used to calculate any change in sensitivity and this can be corrected. Finally, the corrected peak values are used to define the standard curve and the sample results are calculated.

## System Design and Development

### Dispersion

The purpose of Skeggs' innovation in flow analysis – the introduction of air bubbles into the flowing stream – was to reduce dispersion, so as to minimize the time taken for a steady-state condition to be reached in the detector. It was only in the 1970s that the definitive description of dispersion in segmented streams was derived by Snyder and Adler. This showed a complex relationship between internal diameter, liquid flow rate, segmentation frequency, residence time in the flow system, viscosity of the liquid, and surface tension. The most important conclusions from this model are, first, that for any given tubing diameter there are optimum values for flow rate and segmentation frequency to achieve minimum dispersion, and, second, that dispersion can be decreased without limit by reducing tubing diameter.

Other sources of dispersion are the transmission tubing carrying the sample from the autosampler to the pump, and the flow cell, where a debubbler removes the air segmentation before the liquid stream passes through the light path. The advent of electronic or software means for eliminating interference from air bubbles passing through the flow cell, sometimes known as bubble-gating, eliminated the need for flow cell debubblers and was a major contributor to reducing total system dispersion. Sample line dispersion can be reduced by introducing several air bubbles between samples and the intersample wash, using a so-called pecking-probe sampler. In practice, this is only useful for viscous samples such as blood serum.

The advances in system design roughly halved dispersion and doubled analysis rate with each new generation; a heated reaction stage would run at 30 samples per hour on a first-generation system, 60 per hour on a second-generation system, and 120 per hour on third-generation systems using 1 mm hydraulics (1 mm diameter tubing).

### Hardware

The first commercially available SFA system, the AutoAnalyzer, was produced in 1957 and remained in production for ~10 years. During this time all parts of the system underwent rapid design changes until, at the end of its life, it was very similar in principle to modern systems. The first systems, for example, aspirated air between samples, rather than using an intersample wash solution. This resulted in accumulations of analyte in the sample tubing and correspondingly high interaction between subsequent samples. The innovation of a water wash between

samples automatically introduced by the sampler first appeared in 1961.

The second generation of SFA hardware was introduced first in 1967 and subsequently in modular form in 1970; this system was being produced until 1997. These systems incorporated refinements whose effect was to increase the rate and precision of analysis. The sampler had a faster action so as to minimize the time the sample probe aspirated air between sample and wash, thus reducing the volume of the intersample air bubble and its consequential effect on the elasticity of the analytical stream; the pump incorporated an 'air bar' to introduce air bubbles of constant volume synchronized with the pulsation of the rollers, thus ensuring that every liquid segment contained the same volume of reagents and sample: the manifold tubing was reduced in internal diameter from 2.4 to 2.0 mm and was constructed almost entirely of high-tolerance glass, thus ensuring stable flow and reducing flow rates and reagent consumption. The debubbler needed to remove the air bubbles before the flow cell was integrated into the flow cell itself, greatly reducing the unsegmented volume with a corresponding reduction in dispersion and increase in sample rate. Flow rates were  $\sim 2 \text{ ml min}^{-1}$ , about half those of the earliest SFA systems.

Multichannel systems were produced for clinical analysis. Before microprocessors were available for calculating results, these systems used a combination of hydraulic phasing and mechanical timers to switch the steady-state part of each peak sequentially for each channel onto calibrated chart paper.

In 1975, a fully computer-controlled system was introduced that could run 20 channels in parallel. The hydraulics were designed for the first time with a proper understanding of the factors determining dispersion, and incorporated 1 mm glassware, liquid flow rates of  $\sim 0.5 \text{ ml min}^{-1}$ , and a segmentation frequency of 90 bubbles per minute, compared to  $30 \text{ min}^{-1}$  on the earlier systems. The bubbles passed directly through the flow cell, without debubbling, and their interference was eliminated electronically. These improvements resulted in a doubling of sampling rate and halving of flow rates, giving one-fourth the reagent consumption per sample.

This system was dedicated to clinical analysis, and marked a split in development between clinical and industrial systems. The wider variety of sample types and analytical requirements in the industrial market catalyzed the development of techniques such as distillation and UV digestion. Subsequently, 1 mm hydraulics were introduced to industrial systems and computer control enabled automatic photometer control and dilution of over-range samples.

## Further Developments

With the advent of systems using 1 mm hydraulics, with liquid and air deliveries close to the theoretical optimum, further reduction in dispersion could only be achieved by using smaller diameter tubing. This would place such high demands on pump design and so increase the risk of blockage that no system with smaller hydraulics has been produced commercially, and further reduction is unlikely. A system with theoretically zero carryover was introduced in 1987, designed for clinical laboratories. It used hydrophobic tubing that is coated internally with an inert fluorocarbon liquid. Aqueous segments were completely contained within the fluorocarbon, analogous to the air bubbles in a normal segmented system, and thus the thin film of liquid allowing adjacent segments to interact – the source of carryover – is absent.

A system introduced in 1992 can use an injection valve to introduce the sample into the flowing stream, which may be unsegmented, for FIA applications, or air-segmented to accommodate long reactions.

Other system improvements have centered on easier operation and higher automation. Computers, originally used only for calculating results, have taken over system control to automate functions such as photometer baseline and gain setting, and the control of random-access samplers. These samplers allow samples that fall off-scale to be diluted and reanalyzed in an extension to the original analysis run, so avoiding the need for manual dilution and re-run. An accessory to one system allows the reagents for several different tests on a multimethod manifold to be switched automatically between runs, allowing a single channel to be used for multiple analyses.

## Applications

The main fields of applications for SFA systems are, in order of current importance, water, soil and plant extracts, tobacco, food and beverages, chemical production, and clinical analysis. The reasons for the decline of the formerly dominant clinical market have been described; pharmaceutical applications, also previously important, have declined because standard methods developed for new products are generally based on separative techniques, especially liquid chromatography (LC). Table 1 shows the most common applications of SFA.

### Water Analysis

SFA is the dominant technique for the routine automated determination of ions in drinking, surface,



ground, waste, and sea water. Although FIA has made inroads into this area, its inability to accommodate reactions taking longer than  $\sim 1$  min while maintaining high analysis rates is sometimes disadvantageous, and complex determinations involving UV digestion or distillation, such as the measurement of total cyanide in wastewater, are more difficult to automate by FIA due to the high amount of dispersion generated. More recently, discrete analyzers, which use robotics to dispense small volumes of samples and reagents into a separate reaction vessel for each test, have started to enter the water analysis market. However, they are unable to perform the complex tests described above. Since 1995 several SFA methods for water and waste water have received ISO standard certification, including

nitrate, phosphate, total phosphorus, and ammonia. The ISO standard methods are documented for both SFA and FIA.

A simplified flow diagram for the SFA method for total cyanide is shown in **Figure 4**. The sample is acidified and passes through an online UV digester, where metallic cyanide complexes are decomposed into  $\text{CN}^-$  ions. The power and wavelength characteristics of the lamp are such that further breakdown into C and N radicals does not occur. Following digestion, the sample is acidified and distilled to separate the cyanide as HCN. The distillate is condensed and the cyanide ions react with pyridine and barbituric acid, producing a colored compound measured at 570 nm. Variations on the distillation stage of this analysis have been developed; in one method a stream of air is passed through the distillation system, resulting in a shorter residence time and lower dispersion; in another, the distillation stage is replaced by a gas dialyzer.

**Table 1** Applications of segmented flow analysis

<i>Water analysis</i>	<i>Food and drink</i>
Acidity	Alcohol
Alkalinity	Amylase
Ammonia	Bitterness (beer)
Chloride	Calcium
Cyanide	Carbohydrate
Dissolved organic carbon	Lactate
Fluoride	Malic acid
Hardness (Ca and Mg)	Nicotine
Iron	Phosphorus
Nitrate	Protein
Nitrite	Sulfur dioxide
Nitrogen, total	Sugars
Phenol	Vitamins
Phosphate, ortho- and total	
Silicate	<i>Chemicals</i>
Sulfate	Fertilizers (N, P, K, Ca, Mg)
Urea	Chlor-alkali production (NaOH, NaCl, NaOCl, NaClO <sub>3</sub> )
	Pharmaceuticals
<i>Soil analysis</i>	Penicillin and derivatives
Ammonia	
Calcium	
Nitrate	
Nitrogen, organic	
Phosphate	
Potassium	

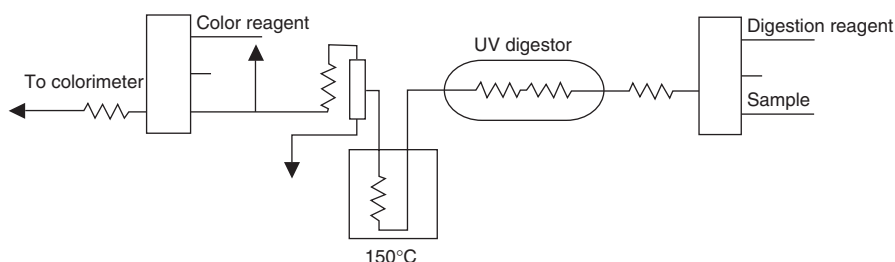
Well over 1000 SFA methods have been documented; this is a summary of the most common nonclinical applications in current use.

### Continuous Monitoring

SFA lends itself to continuous operation, and most methods can be run with a continuous flow of sample, rather than connected to sample changer. More than 100 papers specifically concerning online applications of SFA have been published. Modified laboratory systems have been less successful than those specially designed for continuous use, largely due to limited pump tube life and difficulties associated with calibration and sample introduction. In many online applications a response time of 10–15 min is adequate, and so the benefits of gas segmentation in limiting dispersion are questionable. As a result, many of the most successful commercial systems operate without air bubbles. The division between some of these systems and FIA systems is not sharply defined.

### Interface with Other Systems

SFA systems have been used with LC both for sample preparation before chromatography and for postcolumn derivatization. Sample purification by dialysis, filtration, or solvent extraction can be conveniently



**Figure 4** Simplified flow diagram for total cyanide determination with UV digestion and distillation.

performed with segmented flow. The low band broadening in segmented streams is advantageous both in precolumn sample preparation to allow lengthy cleanup procedures and postcolumn to maintain peak separation. In these hybrid systems, the output from the SFA manifold is pumped through the LC sample loop, which is actuated by a timer when sample concentration has reached its maximum.

Dialysis and solvent extraction are the most commonly used precolumn techniques. In an SFA unit specially developed for LC sample preparation to ensure solvent compatibility with the mobile phase, the evaporation to dryness module, sample dissolved in a volatile organic solvent is pumped onto a moving PTFE wire; the solvent is evaporated in a stream of hot gas and the residue redissolved in a second solvent compatible with the chromatography system.

Applications of SFA/LC include the determination of drugs in blood serum, pharmaceutical product analysis, and the determination of vitamin A in milk.

**See also:** **Blood and Plasma. Clinical Analysis:** Overview; Sample Handling. **Distillation. Extraction:** Solvent Extraction Principles. **Flow Injection Analysis:** Industrial Applications. **Food and Nutritional Analysis:** Overview. **Geochemistry:** Soil, Major Inorganic Components. **Membrane Techniques:** Dialysis and Reverse Osmosis. **Pharmaceutical Analysis:** Drug Purity Determination. **Water Analysis:** Freshwater; Seawater – Inorganic Compounds; Industrial Effluents.

## Further Reading

Furman WB (1976) *Continuous Flow Analysis: Theory and Practice*. New York: Dekker.

Patton CJ and Wade A (1990) Continuous flow analyzers. In: Ewing GW (ed.) *Analytical Instrumentation Handbook*, ch. 27, pp. 979–1051. New York: Dekker.

Snyder LR and Adler HJ (1976) Dispersion in segmented flow through glass tubing in continuous-flow analysis: the ideal and non-ideal models. *Analytical Chemistry* 48(7): 1017–1027.

# SELENIUM

**P C Uden**, University of Massachusetts at Amherst, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Selenium (Se) is essential for animals and humans, causing disease by deficiency, but being toxic on ingestion at an order of magnitude above essentiality. It is present in the environment in inorganic and organic chemical species and analytical speciation is vital as bioavailability and toxicity depend on the binding form.

## Selenium in the Environment

Selenium has six natural stable isotopes:  $^{74}\text{Se}$  (0.9%),  $^{76}\text{Se}$  (9.4%),  $^{77}\text{Se}$  (7.6%),  $^{78}\text{Se}$  (23.8%),  $^{80}\text{Se}$  (49.6%), and  $^{82}\text{Se}$  (8.7%). It usually occurs in the sulfide ores of the heavy metals but rarely as the free element. Volcanic soils are enriched in selenium as are some coal deposits. Selenium compounds are volatilized during the fuel combustion and during copper and other metal smelting and refining. There are numerous overviews of selenium with extensive information on geochemistry and occurrence, seleniferous

environments, solubility, speciation and transformation of selenium in soils, bioaccumulation, etc.

Selenate ( $\text{SeO}_4^{2-}$ ) is thermodynamically stable in alkaline and oxidized environments. Selenates are soluble in water and less strongly adsorbed than selenite ( $\text{SeO}_3^{2-}$ ), easily leached from soils, transported to ground waters, and readily taken up by plants. Selenite occurs in mildly oxidizing neutral pH environments, is less soluble than selenate, is strongly adsorbed by particles, and may be reduced to elemental  $\text{Se}^0$ . Selenides ( $\text{Se}^{2-}$ ) exist in reducing acidic environments but are very insoluble, resist oxidation, and are not taken up by plant and animals. Selenium pollution of waters, sediments, and soils results from industrial activities such as oil refining, aqueous discharges from the storage of coal, coal ash landfill, etc.

The toxicity of selenium depends on its biologically active form. In alkaline soils and oxidizing conditions, it may be oxidized enough to maintain its biological availability and be taken up by plants, but in acidic or neutral soils, it tends to remain relatively insoluble and biologically unavailable. It accumulates in living tissues and its level in human blood is  $\sim 200 \text{ ng ml}^{-1}$ . Selenium has been found in marine fishmeal at levels of  $\sim 2 \mu\text{g per g}$ ,  $\sim 50\,000$  times above that in seawater.

Selenium dioxide is an industrial exposure hazard since it forms selenite, an irritant, with water or

# SAPONINS

**K Hostettmann and A Marston**, Université de Lausanne, Lausanne-Dorigny, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 4540–4544, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

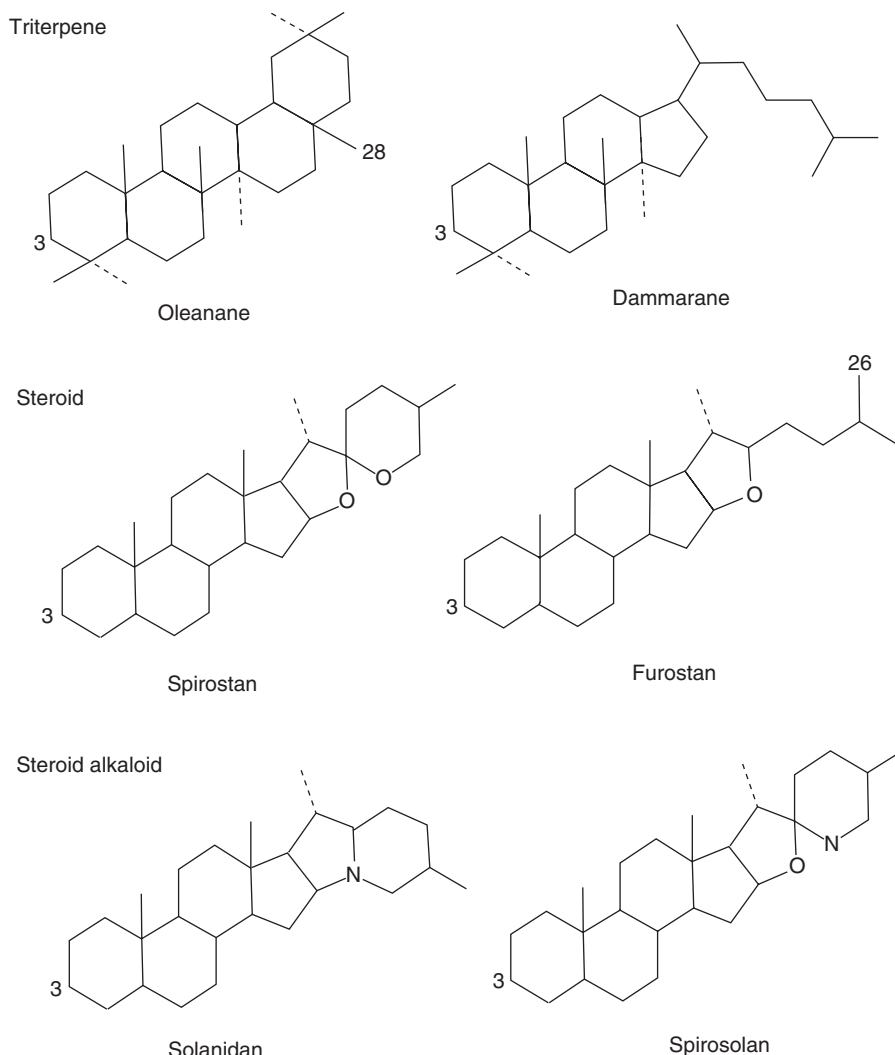
## Introduction

Saponins are  $C_{27}$  or  $C_{30}$  natural product glycosides. They are distributed widely in plants, sometimes in very large amounts. Saponins have surface-active properties and often foam strongly when shaken with water. Saponins consist of one or more linear or

branched sugar chains linked to an aglycone. There are three main classes of these glycosides:

- (1) triterpene glycosides;
- (2) steroid glycosides; and
- (3) steroid alkaloid glycosides.

Examples of the basic skeleton (aglycone) of these saponins are to be found in **Figure 1**. Hydroxylation at the carbon C-3 is almost always found, but functionalization at other positions is common. Monodesmosidic saponins have a single sugar chain attached normally at position C-3. Bidesmosidic saponins have two sugar chains, usually with one attached through an ether linkage at C-3 and the other at C-28 (triterpene saponins) or at C-26 (steroid saponins) or at C-26



**Figure 1** Aglycone skeletons of saponins.

(steroid saponins). Both pentacyclic (with  $\beta$ -amyrin,  $\alpha$ -amyrin, or lupeol skeletons) and tetracyclic (dammarane and lanostane) triterpene saponins are encountered frequently (Figure 1).

Saponins are distributed widely in the plant kingdom, with many examples to be found in plant foodstuffs such as sugar beet, potatoes, tomatoes, soya beans, peas, beans, and yams. The angiosperms are very rich sources of triterpene glycosides, and while these saponins have also been found in pteridophytes, they are virtually absent in the gymnosperms. Triterpene saponins are most often found in the dicotyledons (dammarane glycosides are frequent in the families Araliaceae, Cucurbitaceae, and Rhamnaceae), and steroid saponins are characteristic of the monocotyledons (especially the families Liliaceae and Dioscoreaceae). Steroid alkaloid glycosides are distributed widely in the family Solanaceae.

Certain marine organisms (sea cucumbers, sea urchins, and starfish) contain varying amounts of saponins, mainly for defense purposes.

## Sampling and Sample Pretreatment

The glycosidic nature of saponins means that they are polar compounds and often water-soluble. In extracts from plants, they are accompanied by sugars and phenolic material. Extraction is best achieved with methanol, ethanol, or aqueous mixtures of these alcohols. A defatting step needs to be included, either before or just after extraction with alcohol. In the past, precipitation with lead acetate, barium hydroxide, and tannic acid and salting-out with ammonium sulfate or magnesium sulfate was carried out. Precipitation with cholesterol is still sometimes performed. However, contemporary methods involve precipitation with diethyl ether, or dialysis to remove small water-soluble molecules. A final partition step between water and *n*-butanol is useful in concentrating saponins in the organic layer. Once obtained, these enriched preparations may contain complex mixtures of closely related compounds, differing in the nature of the aglycone or sugar moieties (nature, number, and position of attachment of the saccharides).

More recent sample preparation steps, especially for liquid chromatographic applications, have involved prepurification of extracts on Sep-Pak C<sub>18</sub> (or similar) cartridges. Glycoalkaloids may be purified preliminarily on Sep-Pak C<sub>18</sub> or NH<sub>2</sub> cartridges.

## Nonchromatographic Methods for Determination of Saponins

Several nonchromatographic methods have been employed for the identification and determination of

saponins – namely hemolysis, piscicidal activity, gravimetry, and ultraviolet (UV)–visible spectrophotometry. These are being replaced rapidly by chromatographic techniques, but hemolysis and UV–visible spectrophotometry are still of some importance.

### Hemolysis

One characteristic of many (but not all) saponins is their capacity to rupture erythrocytes (red blood corpuscles). By measuring the change in absorbance of the supernatant of an erythrocyte suspension after hemolysis, the saponin content can be calculated. Various amounts of the saponin-containing product or extract are mixed with a suspension of washed erythrocytes in isotonic buffer at pH 7.4. After 24 h, the mixture is centrifuged and hemolysis is indicated by the presence of hemoglobin (red) in the supernatant. In the European Pharmacopoeia, the quantity in milliliters of ox blood (diluted 1:50) that is totally hydrolysed by 1 g of test substance is measured. As a standard, the saponin mixture from the roots of *Gypsophila paniculata* (Caryophyllaceae) has by definition an activity of 30 000 units.

### UV–Visible Spectrophotometry

Unsaturated and hydroxylated triterpenes and steroids give colored products with aromatic aldehydes in strong mineral acids, with acetic anhydride in sulfuric acid, and with inorganic salts (cerium(IV) sulfate and antimony(III) chloride, for example) in an acidic solution. These reactions have been used as the basis for determination of saponins. The analysis of *Ginseng radix* (*Panax ginseng*, Araliaceae) in Pharmacopoeia Helvetica VII, for example, relies on reaction with glacial acetic acid/sulfuric acid and spectrophotometric determination at 520 nm of the red product. The  $\beta$ -aescine component of horse chestnut (*Aesculus hippocastanum*, Hippocastanaceae) saponin can be determined spectrophotometrically after treatment with a mixture of iron(III) chloride, acetic acid, and sulfuric acid.

## Chromatographic Methods

Spectrophotometric methods are sensitive but not really suitable for estimating saponins in crude plant extracts since the reactions are not specific and colored products may form with accompanying compounds such as phytosterols and flavonoids. Similarly, hemolytic methods also suffer from a lack of specificity. Consequently, alternative ways of analyzing for saponins using different chromatographic techniques have been developed.

### Thin-Layer Chromatography

Thin-layer chromatography (TLC) is of fundamental importance for both the identification and determination of saponins. Both pure products and crude extracts can be handled, the equipment is simple and inexpensive, and results can be obtained rapidly. A number of visualization reagents are available for spraying the TLC plates and give color reactions for the different classes of saponins (Table 1). For quantitative spectrophotometry, the required constituents are separated on the TLC plate, scraped off, and then extracted with a suitable solvent, such as ethanol. After treatment with a reagent, the colored solution is measured at a specified wavelength.

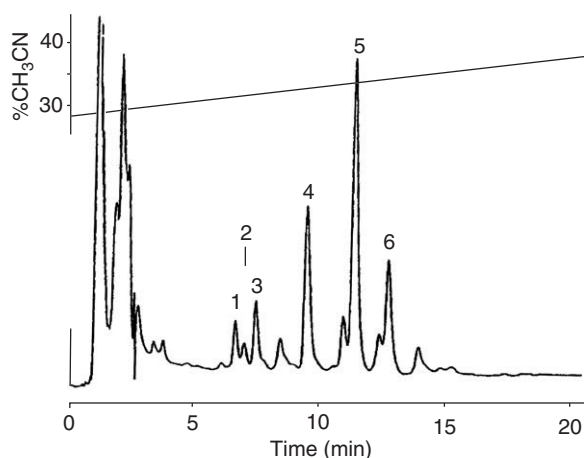
### Liquid Chromatography

Liquid chromatography (LC) is technically much more complicated than TLC. Its speed, sensitivity, and adaptability to nonvolatile polar compounds is ideal for the analysis of saponins and sapogenins. The vast majority of separations have been performed on octadecylsilyl columns, with acetonitrile–water mixtures as the eluent, although occasionally methanol–water mixtures and low concentrations of acid have been used. Acetonitrile is preferred to methanol at short wavelengths because of its smaller UV absorption. As long as gradient changes are small, UV detection (between 203 and 215 nm) is possible. An example of the separation of bidesmosidic saponins from berries of *Phytolacca dodecandra* (Phytolaccaceae) is shown in Figure 2. This used a shallow acetonitrile–water gradient and an RP-8 column. The more polar glycosides of bayogenin (1 and 2 in Figure 2) and hederagenin (3) were separated before the oleanolic acid derivatives (4–6). However, if the polarity difference of the saponin mixture is not too great, isocratic elution is sufficient for the separation.

The lack of a suitable strong chromophore is the reason for certain difficulties when chromatographing saponins. This problem can be overcome by

changing the detection method and employing refractive index and mass detectors, although these techniques have their own inherent complications. An alternative, assuming that the saponin has a suitable functional group, is to derivatize the sample, for example, the free carboxyl group often found in triterpene glycosides can be functionalized with 4-bromophenacyl bromide in the presence of potassium hydrogencarbonate and a crown ether (Figure 3). 4-Bromophenacyl derivatives absorb strongly at 254 nm, and UV detection can be performed at this wavelength without interference from the solvent. By this means, quantification of molluscicidal monodesmosidic saponins in aqueous extracts of *P. dodecandra* berries was possible (Figure 4). Analysis was performed on an RP-18 column with an acetonitrile–water gradient. Because derivatization of the extract allowed the introduction of a chromophore into the triterpene part of the molecule, the saponins were detected at 254 nm. Alfalfa (*Medicago sativa*, Fabaceae) root saponins have also been determined using this method. Alternative methods for analysis of poorly UV active saponins are used to prepare benzoate esters, fluorescent coumarin derivatives, or dienes.

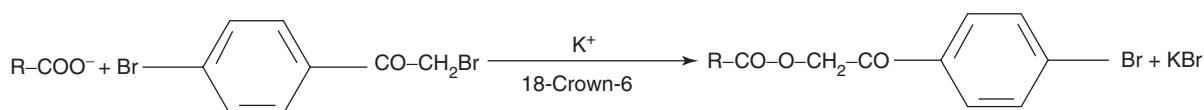
To avoid peak-broadening, some method of suppressing ion formation (of free carboxyl groups on the aglycone or saccharide moieties, for example) is required. This can be achieved by addition of a low UV-absorbing acid such as phosphoric acid or trifluoroacetic acid to the eluent. Another possibility is to use ion pair LC, with a counterion added to the mobile phase. The capacity factor of the ionic compounds is increased by forming ion complexes with the pairing reagent.



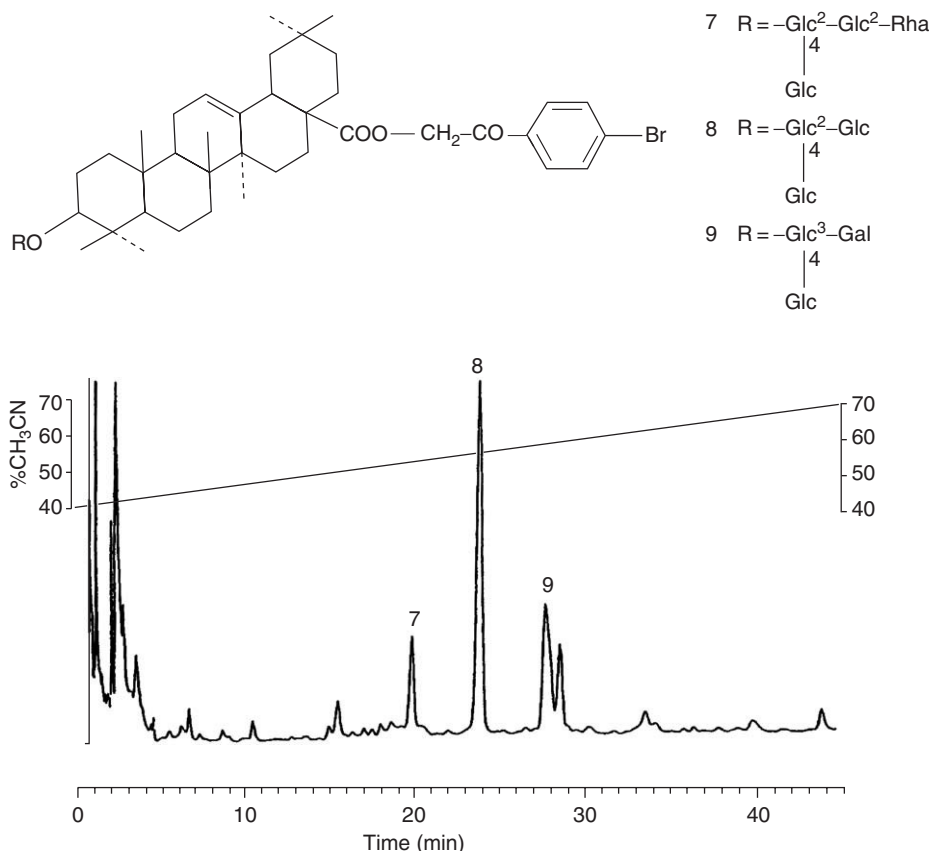
**Figure 2** LC of a methanolic extract of *P. dodecandra* berries. Column, RP-8; eluent, acetonitrile–water gradient from 28% to 38% acetonitrile in 20 min; flow rate, 1.5 ml min<sup>-1</sup>; detection, UV, 206 nm. 1 and 2, polar glycosides of bayogenin; 3, polar glycoside of hederagenin; 4–6, oleanolic acid derivatives.

**Table 1** Visualization reagents for the TLC of triterpene saponins

Vanillin–sulfuric acid
Vanillin–phosphoric acid
Liebermann–Burchard (acetic anhydride–sulfuric acid)
1% Cerium(IV) sulfate in 10% sulfuric acid
10% Sulfuric acid in ethanol
50% Sulfuric acid
<i>p</i> -Anisaldehyde–sulfuric acid
Komarowsky ( <i>p</i> -hydroxybenzaldehyde–sulfuric acid)
Antimony(III) chloride
Blood
Water



**Figure 3** Derivatization of saponins with a UV-active chromophore.



**Figure 4** LC analysis of a derivatized aqueous extract from *P. dodecandra* berries. Column, NovaPak C<sub>18</sub> 4  $\mu\text{m}$  (150 mm  $\times$  3.9 mm); eluent gradient from 40% to 70% acetonitrile over 45 min; flow rate, 1 ml min<sup>-1</sup>; detection, 254 nm. Peaks 7–9 correspond to structures 7–9 at the top of this figure.

If peak resolution on reversed-phase LC columns is insufficient, other approaches are to use hydroxyapatite or chemically modified porous glass columns or to carry out the LC of borate complexes.

Glycoalkaloids can be separated and quantified on RP-18 or amino columns. The solvent combination acetonitrile–water–potassium dihydrogenphosphate is indicated for the latter columns.

### Gas Chromatography

Rather than measuring saponins themselves, gas chromatography (GC) has been used to quantify aglycone moieties as their trimethylsilyl ethers after hydrolysis of parent saponins. This indirect method has been applied to the analysis of dammarane

glycosides from ginseng, analysis of soya bean saponins, and determination of 18 $\beta$ -glycyrrhetic acid after hydrolysis of *Glycyrrhiza* glycosides. Analysis of intact trimethylsilylated triterpene glycosides containing four sugar units or less from ginseng is also possible.

### Selected Applications

For determinations of saponins, LC is now the technique of choice, and better results are thereby obtained compared with UV–visible spectrophotometry, GC, and TLC. The amounts of the individual saponins in a mixture or extract can be determined, and adulterations are easier to discern.



For example, LC is now used routinely for quality control of commercial ginseng (*P. ginseng*) preparations and of the plant material itself. Dried vegetable material is extracted with methanol–water (80:20; v/v) at 60°C. After evaporation of the solvent, the extract is dissolved in water and introduced onto a Sep-Pak C<sub>18</sub> cartridge. Following washing with water and methanol–water (30:70; v/v), the saponins are eluted with methanol and then determined on an octadecylsilyl column (usually 5 µm). A typical elution profile with an acetonitrile–water solvent acidified with 1% phosphoric acid (15:85; v/v; A) and acetonitrile–water (80:20; v/v; B) is the following: 10 min with 10% B in A (isocratic), 15 min with a 10–20% linear gradient of B in A, 15 min with a 20–35% linear gradient of B in A, and 10 min with 35% B in A (isocratic). Elution is at 1 ml min<sup>-1</sup> and detection at 205 nm. By this means the glycosides of (20S)-protopanaxatriol, ginsenosides Rg<sub>1</sub>, Re, Rf, and Rg<sub>2</sub>, are eluted before the glycosides of (20S)-protopanaxadiol, ginsenosides Rb<sub>1</sub>, Re, Rb<sub>2</sub>, and Rd, in that order.

β-Aescine, the active constituent of *Aesculus hippocastanum* (Hippocastanaceae), has important antiinflammatory, antiedematous, and capillaro-protective properties. It is used in therapy of peripheral vascular disorders. β-Aescine is a mixture of saponins, but for standardization purposes, the two main glycosides of protoaescigenin can be determined by

LC. Samples of β-aescine are dissolved in water and analyzed on 5 µm RP-18 columns with the solvent system acetonitrile–water–20% phosphoric acid (33.5:66.5:0.1; v/vv), pH 3.2, at a flow rate of 1 ml min<sup>-1</sup> and UV monitoring at 205 nm.

Other determinations using LC employ similar procedures: these include analyses of licorice extracts, saikosaponins from *Bupleuri radix*, ivy (*Hedera helix*, Araliaceae) saponins, and saponins from *Primula veris* and *P. elatior* (Primulaceae).

**See also:** **Extraction:** Solvent Extraction Principles. **Liquid Chromatography:** Column Technology.

### Further Reading

- Hostettmann K, Hostettmann M, and Marston A (1991p) Saponins. In: Charlwood BV and Banthorpe DV (eds.) *Methods in Plant Biochemistry*, vol. 7, pp. 435–471. London: Academic Press.
- Price KR, Johnson IT, and Fenwick GR (1987) The chemistry and biological significance of saponins in foods and feedingstuffs. *CRC Critical Reviews in Food Science and Nutrition* 26: 27–135.
- Tschesche R and Wulff G (1972) Chemie und Biologie der Saponine. *Progress in the Chemistry of Organic Natural Products* 30: 462–606.
- Schwarz MW (2002) *Saponins In: Ullmann's Encyclopedia of Industrial Chemistry*. Heidelberg: Wiley-VCH.

## SCANNING TUNNELING MICROSCOPY

**See** MICROSCOPY TECHNIQUES: Atomic Force and Scanning Tunneling Microscopy

## SEALANTS

**See** ADHESIVES AND SEALANTS

## SEAWATER

**See** WATER ANALYSIS: Seawater – Organic Compounds; Seawater – Dissolved Organic Carbon; Seawater – Inorganic Compounds

## SECONDARY ION MASS SPECTROMETRY

**See** MASS SPECTROMETRY: Liquid Secondary Ion Mass Spectrometry. SURFACE ANALYSIS: Secondary Ion Mass Spectrometry of Polymers

performed with segmented flow. The low band broadening in segmented streams is advantageous both in precolumn sample preparation to allow lengthy cleanup procedures and postcolumn to maintain peak separation. In these hybrid systems, the output from the SFA manifold is pumped through the LC sample loop, which is actuated by a timer when sample concentration has reached its maximum.

Dialysis and solvent extraction are the most commonly used precolumn techniques. In an SFA unit specially developed for LC sample preparation to ensure solvent compatibility with the mobile phase, the evaporation to dryness module, sample dissolved in a volatile organic solvent is pumped onto a moving PTFE wire; the solvent is evaporated in a stream of hot gas and the residue redissolved in a second solvent compatible with the chromatography system.

Applications of SFA/LC include the determination of drugs in blood serum, pharmaceutical product analysis, and the determination of vitamin A in milk.

**See also:** **Blood and Plasma. Clinical Analysis:** Overview; Sample Handling. **Distillation. Extraction:** Solvent Extraction Principles. **Flow Injection Analysis:** Industrial Applications. **Food and Nutritional Analysis:** Overview. **Geochemistry:** Soil, Major Inorganic Components. **Membrane Techniques:** Dialysis and Reverse Osmosis. **Pharmaceutical Analysis:** Drug Purity Determination. **Water Analysis:** Freshwater; Seawater – Inorganic Compounds; Industrial Effluents.

## Further Reading

Furman WB (1976) *Continuous Flow Analysis: Theory and Practice*. New York: Dekker.

Patton CJ and Wade A (1990) Continuous flow analyzers. In: Ewing GW (ed.) *Analytical Instrumentation Handbook*, ch. 27, pp. 979–1051. New York: Dekker.

Snyder LR and Adler HJ (1976) Dispersion in segmented flow through glass tubing in continuous-flow analysis: the ideal and non-ideal models. *Analytical Chemistry* 48(7): 1017–1027.

# SELENIUM

**P C Uden**, University of Massachusetts at Amherst, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Selenium (Se) is essential for animals and humans, causing disease by deficiency, but being toxic on ingestion at an order of magnitude above essentiality. It is present in the environment in inorganic and organic chemical species and analytical speciation is vital as bioavailability and toxicity depend on the binding form.

## Selenium in the Environment

Selenium has six natural stable isotopes:  $^{74}\text{Se}$  (0.9%),  $^{76}\text{Se}$  (9.4%),  $^{77}\text{Se}$  (7.6%),  $^{78}\text{Se}$  (23.8%),  $^{80}\text{Se}$  (49.6%), and  $^{82}\text{Se}$  (8.7%). It usually occurs in the sulfide ores of the heavy metals but rarely as the free element. Volcanic soils are enriched in selenium as are some coal deposits. Selenium compounds are volatilized during the fuel combustion and during copper and other metal smelting and refining. There are numerous overviews of selenium with extensive information on geochemistry and occurrence, seleniferous

environments, solubility, speciation and transformation of selenium in soils, bioaccumulation, etc.

Selenate ( $\text{SeO}_4^{2-}$ ) is thermodynamically stable in alkaline and oxidized environments. Selenates are soluble in water and less strongly adsorbed than selenite ( $\text{SeO}_3^{2-}$ ), easily leached from soils, transported to ground waters, and readily taken up by plants. Selenite occurs in mildly oxidizing neutral pH environments, is less soluble than selenate, is strongly adsorbed by particles, and may be reduced to elemental  $\text{Se}^0$ . Selenides ( $\text{Se}^{2-}$ ) exist in reducing acidic environments but are very insoluble, resist oxidation, and are not taken up by plant and animals. Selenium pollution of waters, sediments, and soils results from industrial activities such as oil refining, aqueous discharges from the storage of coal, coal ash landfill, etc.

The toxicity of selenium depends on its biologically active form. In alkaline soils and oxidizing conditions, it may be oxidized enough to maintain its biological availability and be taken up by plants, but in acidic or neutral soils, it tends to remain relatively insoluble and biologically unavailable. It accumulates in living tissues and its level in human blood is  $\sim 200 \text{ ng ml}^{-1}$ . Selenium has been found in marine fishmeal at levels of  $\sim 2 \mu\text{g per g}$ ,  $\sim 50\,000$  times above that in seawater.

Selenium dioxide is an industrial exposure hazard since it forms selenite, an irritant, with water or

sweat. Other selenium compounds released during coal or petroleum combustion may be significant sources of exposure. Biogeochemical cycling involves geological erosion and anthropogenic influences to introduce selenium to sea water, mainly as selenite and selenate, at levels below  $1 \text{ ng ml}^{-1}$ . It is incorporated into organoselenium compounds in phytoplankton, zooplankton, lower and higher vertebrates, such pathways being considered to involve biomethylation. Natural selenium levels in soils vary widely from below  $100 \text{ ng per g}$  to  $10 \text{ } \mu\text{g per g}$  or above. Selenium volatilization contributes an estimated  $6 \times 10^6 \text{ kg}$  selenium per year to the atmosphere, dimethyl selenide and dimethyl diselenide being proposed as primary selenium carrier species. The concentrations in most plants and agricultural crops are below  $1 \text{ mg per kg}$  but some grown in selenium-rich soils may accumulate selenium to concentrations of several thousands of  $\text{mg per kg}$  dry mass. Hyperaccumulating plants such as *Brassica juncea* can also be used to remove selenium from contaminated soils by phytoremediation. Table 1 lists selenium species identified in the environment, inorganic forms and methylated species, selenoamino acids, selenoenzymes, and selenium nucleic acids. Traces of selenium ranging from  $0.001$  to  $10 \text{ ng l}^{-1}$  are found in drinking water. Concentrations of dimethyl selenide and dimethyl diselenide in air samples have been reported at  $\text{ng m}^{-3}$  levels.

Selenium's ecotoxicity is complex since selenium exists in many chemical forms, each of which has a different bioavailability and ecotoxic potential. Cytotoxic selenium species, such as selenocysteine, catalyze formation of free radicals and may be

harmful at very low concentration, but other species such as selenomethionine may present an ecotoxic threat without being directly cytotoxic. Organisms at higher trophic levels, such as birds, obtain most of their selenium from eating lower trophic level organisms, so the chemical forms of selenium to which they are primarily exposed may vary. The key to a mechanistic biochemical understanding of selenium ecotoxicity lies in a more complete knowledge of the processes of speciation, biotransformation, and accumulation in the food chain.

The many selenium species present in environmental and biological materials present a great analytical challenge. Biologically selenium is not bound by coordination but forms covalent C–Se bonds. Enzyme products arise from reactions such as reduction, methylation, and selenoamino acid synthesis. Selenium is incorporated into gene products, according to the UGA codon that encodes for the selenocysteinyl residue. To elucidate the complexities of selenium chemistry, the determination of 'total element', although essential to determine element mass balance, provides insufficient information and must be accompanied by speciation of selenium compounds, as related to biochemical cycles such as shown in Figure 1.

## Speciation of Nutritional Selenium

Differences in the selenium content of foodstuffs are due in part to geographical differences in the amounts of the element that are transferred from soil to the food chain. Selenium-deficiency diseases related to selenium-poor soil are recognized in parts of China and elsewhere, but soils rich in selenium (above  $5 \text{ ppm}$ ) are found in parts of the United States. The selenium in most plant-derived foods, present substantially as selenomethionine (SeMet) and selenocysteine (SeCys) derivatives, has reasonably good bioavailability. However, in animal-derived foods selenium has a wide range of bioavailability as assessed by its ability to increase liver concentrations and glutathione peroxidase activity.

Selenium was recognized as nutritionally significant in the late 1950s and in the 1970s: a discrete metabolic function was found as an essential component of the enzyme glutathione peroxidase (GPX) that contributes to antioxidant protection of cells by reducing hydroperoxides. Several selenium enzymes are now recognized, GPX isoforms, iodothyronine 5'-deiodinases (DIs), thioredoxin reductases (TRs), selenophosphate synthetase, and plasma selenoprotein P (SeP), all containing selenocysteine (SeCys). The nutritional essentiality of selenium may be due to the formation of the active selenol group ( $-\text{SeH}$ ) of

**Table 1** Some selenium compounds in environmental and biological systems

### Inorganic species

$\text{Se}^0$  (element), selenide –  $\text{Se}^{2-}$ , selenate –  $\text{SeO}_4^{2-}$ , selenite –  $\text{SeO}_3^{2-}$

### Simple organic and methylated species

Methylselenol ( $\text{MeSeH}$ ), dimethylselenide ( $\text{Me}_2\text{Se}$ ), dimethyldiselenide ( $\text{Me}_2\text{Se}_2$ )

Trimethylselenonium cation ( $\text{Me}_3\text{Se}^+$ ), dimethylselenone ( $\text{Me}_2\text{SeO}_2$ )

Dimethylselenoxide ( $\text{Me}_2\text{SeO}$ ), methylseleninic acid anion ( $\text{MeSe(O)O}^-$ )

Dimethylselenosulfide ( $\text{MeSSeMe}$ ), selenourea ( $\text{Se}=\text{C}(\text{NH}_2)_2$ )

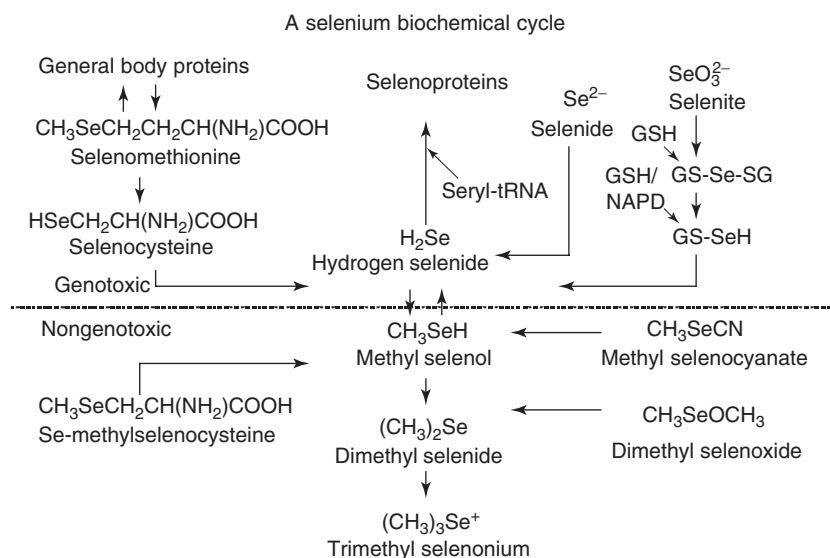
### Amino acids and low molecular mass species

Selenomethionine, selenocysteine, selenocystine, Se-methylselenocysteine, selenocysteic acid

Se-methylselenomethionine, selenomethionine selenoxide, selenocholine, selenobetaine

### Other compounds

Selenoproteins, selenoenzymes, Se-metal metallothionines



**Figure 1** Biochemical selenium pathways. (Reproduced from Uden C (2002) Modern trends in the speciation of selenium by hyphenated techniques. *Analytical and Bioanalytical Chemistry* 373: 423 (Figure 1); © Springer-Verlag.)

such SeCys proteins, antioxidant protection by the GPXs, energy metabolism affected by the DIs, and redox regulation of transcriptional factors and gene expression by the TRs.

Desirable selenium dietary levels span a relatively narrow range: food containing less than 0.1 mg per kg of food is deficient, but levels above 1 mg per kg lead to toxicity. Interest in selenium has focused on its role in the maintenance of low risk to cancer, relevant speciation studies being essential.

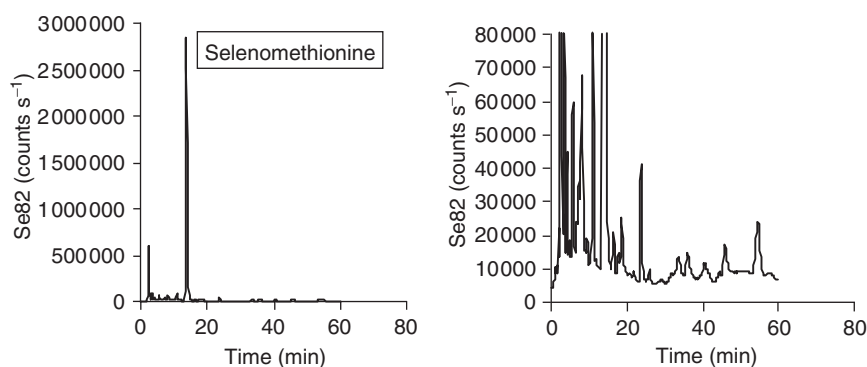
The *allium* plant family exhibits selenium chemistry and high-selenium garlic has found promise in mammary cancer prevention. At present, no selenium-enriched food or nutritional supplements have been fully characterized with respect to selenium compounds.

A common selenium supplement is 'selenized yeast', which has been evaluated in human nutritional clinical trials. The pattern of selenium compounds in different selenium-enriched products varies substantially. Only those made with active yeasts contain selenomethionine,  $\gamma$ -glutamyl Se-methylselenocysteine, and Se-adenosylselenohomocysteine, and perhaps other compounds with methyl-Se bonds such as Se-methylselenocysteine and methylselenides  $[\text{CH}_3]\text{SeR}$ , which may be the most directly anti-carcinogenic of the known selenium metabolites. Different selenized yeasts have been found to contain organically bound selenium in a range from 0% to 97% of total selenium with some showing only a few selenium species and others containing more than 20 different selenium compounds. **Figure 2** shows selenium-specific high-performance liquid

chromatography-inductively coupled plasma-mass spectrometry (HPLC-ICP-MS) chromatograms of enzymatic hydrolysis products of 1250 ppm selenium yeast using 0.1% HFBA ion pairing agent (right scale expanded).

## Selenium in Biology and Medicine

The essential metabolic need for selenium has as its basis formation of the active selenol group ( $-\text{SeH}$ ) center of glutathione peroxidase, thioredoxin reductase, and of other selenoenzymes. Cancer chemoprevention is associated with inorganic selenium salts, selenoamino acids, and other organoselenium compounds. Monomethylated forms of selenium such as methylselenol ( $\text{CH}_3\text{SeH}$ ) are thought to be important chemopreventive selenium metabolites. Many epidemiological studies have pointed to an inverse association of selenium status and risk to at least some cancers. Supplemental selenium has been found to promote cellular phenomena thought to be anti-carcinogenic, to support cellular redox regulation via GPXs and TRs, to enhance immune functions, to increase carcinogen metabolism, to promote apoptotic responses, and to inhibit angiogenesis (blood-vessel formation) in mammary cancer. In a multicenter, double-blind, randomized, placebo-controlled study a total of 1312 older Americans with histories of basal or squamous cell carcinomas of the skin were randomized to dietary supplement of selenium-enriched yeast (200  $\mu\text{g}$  Se per day) or placebo. After over 6 years of follow-up subjects taking supplemental selenium had lower incidences of total cancer, specifically cancers of the



**Figure 2** Selenium-specific HPLC–ICP-MS chromatograms of enzymatic hydrolysis products of 1250 ppm selenium-enriched yeast using 0.1% HFBA ion pairing agent (right scale expanded).

lung colon-rectum and prostate, and overall cancer mortality. Evidence suggests that anti-tumorigenic activities can be supported by metabolites of forms of the element that naturally occur in foods: the Se-amino acids, SeMet and SeCys, and methylated Se-compounds such as Se-methyl-SeCys. With varying efficiencies, these species can be converted to a number of selenium metabolites, including methyl-selenol ( $\text{CH}_3\text{SeH}$ ), which appears to be a key anti-tumorigenic metabolite. These findings make the consideration of ‘Se status’ important in the maintenance of overall health.

## Analytical Strategies

For practical speciation, an analytical method must be specific to a particular selenium species or that species must be separated in time or space before arriving at a selenium-specific detector. Radioimmunological assays for selenoproteins offer very low detection limits but require the proteins to be isolated in amounts sufficient for antibody production. Selenium has been speciated into selenate, selenite, elemental Se, and organoselenium using fractionation techniques and X-ray absorption spectroscopy, but little has been done in speciating organic forms.

The application of hyphenated methods, involves coupling of an electromigration or chromatographic technique with a mass spectral, atomic spectrometric, or other selenium-specific measurement. This may be offline, e.g., SDS-PAGE with instrumental neutron activation analysis, or online, as in HPLC–ICP-MS. Selenium chemistry may mirror sulfur chemistry, but as selenium species are typically present at three orders of magnitude below sulfur analogs, considerable analytical challenges occur. Sometimes, however, there are major differences between sulfur and selenium chemistries, notably in redox behavior.

Ionic, zwitterionic, and neutral selenium species are present in natural systems. **Table 2** shows selenium compounds of biological and clinical interest and species in living organisms are listed in **Table 3**. Selenoproteins contain selenocysteinyl residues but proteins that contain selenomethionyl residues are not formally classified as selenoproteins.

Analytical techniques for the determination of selenium species, including selenoproteins, have been reviewed. Many methods have only been applied to the commercially available standards and not to address specific biochemical problems.

The complexity of analyte matrixes and the low level of selenium compounds even in enriched samples make speciation difficult, but the combination of separation processes with selenium-specific detection is a powerful approach. High sensitivity is vital, and MS with an atmospheric pressure ionization source, such as the ICP, has proved successful for HPLC detection. Selenium presents problems due to moderate ionization efficiency and isobaric interferences, although these can be partly overcome with high-resolution mass spectrometers or dynamic reaction cell (DRC) technology. Significant isotopic overlap from  $^{40}\text{Ar}_2^+$  on the most abundant isotope  $^{80}\text{Se}$  (49.6%) may necessitate measurement of the less abundant isotopes  $^{82}\text{Se}$  (8.6%) or  $^{77}\text{Se}$  (7.6%) of total selenium.

## HPLC Modes for Selenium Speciation

Size exclusion, reversed-phase, paired ion reverse-phase, and ion exchange (cation and anion) HPLC have been used for selenium speciation. Size exclusion, also used in combination with affinity chromatography, has been used for the determination of selenoproteins and studies of interactions of selenium with proteins in the body. Ion-exchange and reversed-phase HPLC are often used for separation

**Table 2** Inorganic and organoselenium analytical target compounds

Selenous acid, selenite	$\text{SeO}_3^{2-}$
Selenic acid, selenate	$\text{SeO}_4^{2-}$
Selenocyanate	$\text{SeCN}^-$
Methylseleninic acid anion	$\text{MeSe(O)O}^-$
Methylselenenic acid anion	$\text{MeSeO}^-$
Dimethylselenide	$\text{Me}_2\text{Se}$
Dimethyldiselenide	$\text{Me}_2\text{Se}$
Methylselenol	$\text{MeSeH}$
Trimethylselenonium cation	$\text{Me}_3\text{Se}^+$
Selenocysteine	$\text{H}_3\text{N}^+-\text{CH}(\text{COO}^-)-\text{CH}_2-\text{SeH}$
Selenocystine	$\text{H}_3\text{N}^+-\text{CH}(\text{COO}^-)-\text{CH}_2-\text{Se}-\text{Se}-\text{CH}_2-\text{CH}(\text{COO}^-)-\text{NH}_3^+$
Selenomethionine	$\text{H}_3\text{N}^+-\text{CH}(\text{COO}^-)-\text{CH}_2-\text{CH}_2-\text{Se}-\text{Me}$
Se-methylselenocysteine	$\text{H}_3\text{N}^+-\text{CH}(\text{COO}^-)-\text{CH}_2-\text{Se}-\text{Me}$
Gamma-glutamyl-Se-methylselenocysteine	$\text{H}_3\text{N}^+-\text{CH}_2-\text{CH}_2-\text{CO}-\text{NH}-\text{CH}(\text{COO}^-)-\text{CH}_2-\text{Se}-\text{Me}$
Selenocystathionine	$\text{H}_3\text{N}^+-\text{CH}(\text{COO}^-)-\text{CH}_2-\text{CH}_2-\text{Se}-\text{CH}_2-\text{CH}(\text{COO}^-)-\text{NH}_3^+$
Selenohomocysteine	$\text{H}_3\text{N}^+-\text{CH}(\text{COO}^-)-\text{CH}_2-\text{CH}_2-\text{SeH}$
Se-adenosylselenohomocysteine	$\text{NH}_2\text{CH}(\text{COOH})\text{CH}_2\text{CH}_2\text{SeCH}_2\text{C}_4\text{H}_5\text{O}_3\text{C}_5\text{N}_4\text{NH}_2$

**Table 3** Selenium species in living organisms

<i>Selenium in proteins</i>	
Selenoproteins	Selenocysteiny residues
Se-containing proteins	Selenomethionyl residues
<i>Nonprotein selenium species</i>	
Inorganic selenium	Selenite ( $\text{SeO}_3^{2-}$ ), selenate ( $\text{SeO}_4^{2-}$ )
Methylated selenium	Monomethylselenol, dimethylselenide, trimethylselenonium ions
Selenoamino acids	Selenocystine, selenomethionine, Se-methylselenocysteine, selenogluthathione

of Se(IV), Se(VI), selenomethionine, and other selenoamino acids. A polymer-based reverse-phase column separated selenocystine, selenomethionine, and trimethylselenonium cation, with detection limits of  $\sim 1 \mu\text{g l}^{-1}$  for each species. Anion exchange was used for speciation of *in vitro* gastrointestinal extracts of cooked cod by ICP-MS but the degree of separation was strongly dependent on the pH of the mobile phase. A polymer-based strong anion-exchange column resolved selenite, selenate, selenomethionine, and selenocystine at pH 8.5 with 3% methanol in the mobile phase to enhance the ICP-MS signal.

If a 250 mg sample of selenium-enriched yeast with a selenium content of 1200 mg per kg is taken into 5 ml of solution, and target compounds are present at 10% of the total concentration, then prior to separation, the concentration of selenium in solution is  $6 \text{ mg l}^{-1}$ . If the separation procedure produces a dilution of 100, the concentration of material after separation is on the order of  $60 \mu\text{g l}^{-1}$ . The low concentrations of the target species present a

challenge and sample pretreatment must not change the chemical forms of the target species, or must change them in a known and controllable fashion.

Selenite was measured in human urine after selenomethionine supplementation, by anion-exchange chromatography with  $25 \text{ mmol l}^{-1}$  NaOH in 2% methanol mobile phase and ICP-MS detection. The concentration found ranged from  $0.4$  to  $7.1 \mu\text{g l}^{-1}$ , while the total selenium concentration was from  $12.4$  to  $97.6 \mu\text{g l}^{-1}$ . Assisted online species conversion hydride generation atomic absorption spectrometry (AAS) was used for quantitative selenium speciation analysis of selenate, selenite, selenomethionine, and selenocystine in CRM 402 reference material. Biological samples and selenium species were quantified in certified reference materials by anion exchange with interfaced direction injection nebulization (ICPAES) and graphite furnace AAS. An important advance in ICP-MS technique involved detection as  $^{80}\text{Se}$  by a DRC MS method. The most comprehensive ion exchange separation of selenium species yet reported separated cations on a cation-exchange column at pH 3 with a gradient of pyridinium formate, and anions on an anion-exchange column at pH 8.5 with a TRIS salicylate mobile phase. Selenomethionine-Se-oxide was observed at a high level in selenized yeast and dimethylselenonium propionate in an algal extract.

Many organoselenium compounds, especially selenoamino acids, are insufficiently hydrophobic to be retained and separated on  $\text{C}_8$  and  $\text{C}_{18}$  reversed-phase stationary phases with salt-free aqueous mobile phases, but retention and separation is increased by using ion-pairing reagent such as trifluoroacetic acid (TFA) or octane sulfonic acid. Perfluorinated carboxylic acids ion-pairing agents afford better resolution and are advantageous for characterizing samples



containing many different classes of organoselenium compounds. TFA is a good general-purpose system but compromises between resolution and retention time. However, heptafluorobutanoic acid (HFBA) provides considerable resolution enhancement and allows separation of many additional organoselenium species including the *cis-trans* isomers of Se-1-propenyl-dl-selenocysteine; eight inorganic and organoselenium species coeluting within an early eluted band with TFA were fully separated with HFBA. The method can also determine selenoxides and organoanions. Anionic species show some retention as protonated forms retained by reversed-phase partition. Semipreparative-scale reversed-phase columns were used for two-dimensional HPLC to isolate fractions of selenized yeast extracts for online and offline investigations by ICP-MS. Wild-type and genetically modified selenium-accumulating *Brassica juncea* were speciated with HPLC-ICP-MS and ES-MS detection. A mixed ion-pair technique in which butanesulfonate and tetramethylammonium hydroxide were used simultaneously enabled resolution of selenite, selenate, selenocystine, selenourea, selenomethionine, selenoethionine, selenocystamine, and trimethylselenonium.

#### **Speciation of Biomacromolecular Selenium Species by HPLC-ICP-MS**

Although separation of biomacromolecular selenium species by size exclusion chromatography (SEC) is mainly based on the molecular weights of the analytes, adsorption and ion exchange effects can play an important role. Resolution of small compounds of similar molecular mass may be achieved in addition to the possibility of the detection of selenoproteins. SEC-ICP-MS has enabled speciation of a number of metals bound to various macromolecular ligands but has had limited success for selenoproteins, the large dilution factor for selenium limiting sensitivity in attempts to speciate glutathione peroxidase. SEC-HPLC was the first procedure to measure selenoprotein-P in human plasma, but it lacks the resolution and sensitivity for separation of the major selenium proteins; a human serum sample yielded three signals, but none coeluted with the glutathione peroxidase activity. Speciation of human breast milk whey gave four selenium signals corresponding to apparent molecular weights of 15, 60, 1500, and >2000 kDa. A combination of affinity chromatography with SEC-ICP-MS separated three major selenium-containing proteins (albumin, glutathione peroxidase, and selenoprotein-P) in human plasma. Selenium incorporation into cyanobacterial metallothionein induced under heavy metal stress

was studied using SEC-ICP-MS, two pathways being indicated. Selenite and selenate metabolism in rats was investigated and exchange of endogenous and dietary selenium examined in brain, liver, and kidney.

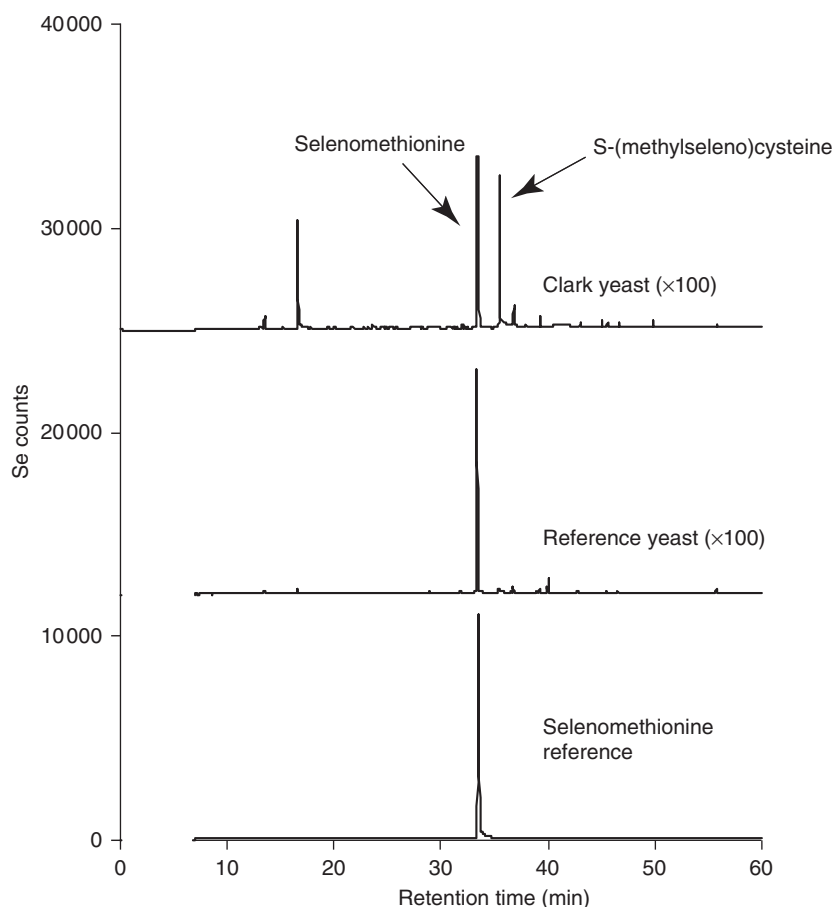
#### **Chiral Speciation of Selenium Species by HPLC-ICP-MS**

The presence of asymmetric carbons in selenomethionine, other  $\alpha$ -selenoamino acids, and related compounds produces different chiral enantiomers with different physiological activities. HPLC separation of enantiomers is possible with a range of chiral stationary phases. d- and l-Selenomethionine enantiomers have been resolved with an  $\alpha$ -cyclodextrin stationary phase and other species with a teicoplanin-based chiral phase. Hybrid chiral methodologies based on GC, HPLC, and capillary electrophoresis, coupled with ICP-MS are feasible. Enantiomers of d,l-selenocystine, d,l-selenomethionine, and d,l-selenoethionine were examined in a range of commercial dietary supplements using a chiral crown ether stationary phase and ICP-MS detection. Selenium-enriched onion, garlic, and yeast were analyzed and some of the selenoamino acid enantiomers were identified. 1-Fluoro-2, 4-dinitrophenyl-5-l-alanine amide was used to derivatize enantiomers of selenoamides for enhanced resolution.

#### **Microwave Plasma Atomic Emission GC Detection of Selenium**

Gas chromatography interfaced with atomic plasma emission spectroscopic detection (GC-AED) is an established tool for specific element detection of volatilizable species and has been used to detect and determine many volatile organoselenium compounds present in or produced by plants. Headspace-GC-AED is used to detect selenium compounds in members of the *Allium* family such as garlic, elephant garlic, onion, and broccoli (*Brassica*). A common structure is  $R-S_x(Se_y)-R'$ , where the R and R' are methyl or allyl groups and x or y may be 1–3. Natural abundance organoselenium compounds in human breath after ingestion of garlic can be identified using Tenax trap/cryogenic-GC-AED.

Free selenoamino acids, selenocysteine, Se-methyl selenocysteine, and selenomethionine in normal and selenium-enriched plants may be determined by GC-AED of ethylated derivatives. AED, flame photometry, and GC-MS detection have been compared for selective determination of selenomethionine in wheat samples.



**Figure 3** Selenium-specific GC-AED chromatograms (Se 196 nm) of ethylated enzymatic hydrolyzates of archived 200 µg Se tablets (top), reference 200 µg Se tablets (center), and selenomethionine (bottom). Peak at ~36 min contains sulfur also.

Selenomethionine is the predominant selenium moiety in plants whereas selenocysteine is probably formed from glycine and selenite in mammalian tissues. Analytical approaches are typically based on degradation of the original matrix to these and other amino acids followed by their determination. For this purpose, GC with element-selective or mass spectrometric detection provides a valuable alternative to HPLC due to improved sensitivity. Selenoamino acids are derivatized with isopropylchloroformate and bis(*p*-methoxyphenyl) selenoxide, and with ethyl chloroformate. **Figure 3** shows comparative selenium-specific GC-AED chromatograms for ethylated extracted enzymatic yeast hydrolyzates of archived selenized yeast employed in the ‘Clark trial’, a reference fresh selenized yeast and reference selenomethionine.

### GC with ICP-MS Detection

Although less widely adopted than HPLC–ICP-MS, GC–ICP-MS can fulfill some of the same speciation

functions as GC-AED for selenium. Selenomethionine enantiomers were resolved by capillary GC as trifluoroacetyl-*O*-isopropyl derivatives with a *l*-valine-*tert*-butylamide modified chiral stationary phase, detection limits below 250 pg being obtained for each isomer with ICP-MS. A glow discharge (rf-GD) mass spectral ion source was used as a sample introductory technique for selenoamino acids, with detection limits of 100 pg being seen for derivatized selenomethionine.

### Capillary Zone Electrophoresis with ICP-MS Detection

Capillary zone electrophoresis (CZE) is an effective technique for the definitive verification of the chromatographic purity of the target compound because of the large number of theoretical plates obtainable. The advantages of CZE, such as the possibility of analyzing for relatively labile species because of the absence of chromatographic packing, must be

considered given the need for ultrasensitive detection, such as high-resolution ICP-MS because of the small sample amount injected. A commercial interface for CZE-ICP-MS is available that optimizes electrophoretic and nebulizer flows and has minimal dilution and sample consumption.

The potential of CZE-ICP-MS has been particularly realized in the identification and determination of selenogluthathione and differentiation of methionine, selenomethionine, cystamine, and selenocystamine in milk. Se(IV), Se(VI), selenate-carrying glutathione (GSSeSG), selenomethionine, selenocystine, and selenocystamine can be speciated at the 10–50 µg per Se per liter level. A two-dimensional separation approach for selenized yeast speciation, based on size exclusion followed by CZE-ICP-MS, coupling to the ICP via a self-aspirating total consumption nebulizer affords limits of low molecular weight selenium species in the range 7–18 µg l<sup>-1</sup>.

## Mass Spectral Structural Identification

HPLC-ICP-MS, GC-AED, and CZE-ICP-MS offer sufficient sensitivity and selenium selectivity for the speciation of selenium compounds in many samples but do not give structural identification of known, unknown, or unpredicted compounds. This lack of identification becomes more important as more efficient separations are achieved. Chromatographic identification is based on migration time through a chromatographic support, but authentic standards of selenium compounds are often not available for retention-based identification, except for simple compounds. The major approaches to analyte authentication involve preparation of additional synthetic standards or isolation of purified selenium compounds from sample matrixes for further characterization by electrospray (ES-MS) or matrix-assisted laser desorption ionization.

The identification of selenium species by matching retention (migration) times with those of authentic standards demands the assurance that the resolution of the chromatographic technique produces at a given retention time a signal corresponding to this compound only, and that peak shapes are not distorted by sample overload or unacceptable adsorption. Thus, hyphenated molecular characterization by MS or NMR spectroscopy provides valuable confirmatory data for chromatographic speciation of standards and identifications of unknowns.

A combination of two-dimensional multinuclear NMR, electrospray MS, and GC-MS methods is valuable to characterize Se-methylselenomethionine, Se-methylselenocysteine, and dimethylselenonium

propionate in synthetic preparations. GC-MS is sensitive but restricted to volatilizable species; for the nonvolatilizable metabolites, there is often a discrepancy between the sensitivity of HPLC-MS techniques and that needed for speciation in real samples. For real-world samples, purification and preconcentration of selenium species are usually needed.

Electrospray MS that allows a precise ( $\pm 1$  Da) determination of molecular mass is an invaluable tool for the identification and a prerequisite for further characterization of compounds in speciation analysis. ESI-MS has been applied mostly to commercial standards or synthetic preparations. Pneumatically assisted ESI-MS is used to identify Se-adenosylhomocysteine in an extract of selenized yeast. Selenium in garlic has been speciated by parallel ICP-MS and electrospray tandem MS. A novel selenium metabolite in rat urine, the selenosugar diastereomer Se-methyl-N-acetylhexosamine, has been identified by this method. It should be emphasized that when authentic standards are not available the use of tandem MS is vital to confirm selenium speciation indicated by means of hyphenated selenium-specific detection.

**See also:** **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Capillary Electrophoresis:** Overview. **Liquid Chromatography:** Chiral.

## Further Reading

- Cai XJ, Block E, Uden PC, Sullivan JJ, and Quimby BD (1995) Allium chemistry: identification of natural abundance organoselenium compounds in human breath after ingestion of garlic using gas chromatography with atomic emission detection. *Journal of Agricultural and Food Chemistry* 43: 1751–1753.
- Clark LC, Combs GF, Turnbull BW, *et al.* (1996) Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skin. A randomized controlled trial. *Journal of the American Medical Association* 276: 1957–1963.
- Combs GF (2001) Selenium in global food systems. *British Journal of Nutrition* 85: 517–547.
- Combs GF and Clark LC (1999) Selenium and cancer. In: Heber D, Blackburn GL, and Go VLW (eds.) *Nutritional Oncology*, pp. 215–222. New York: Academic Press.
- D'Ulivo A (1997) Determination of selenium and tellurium in environmental samples. *Analyst* 122: 117R–144R.
- Ganther HE (1999) Selenium metabolism, selenoproteins and mechanisms of cancer prevention, complexities with thioredoxin reductase. *Carcinogenesis* 20: 1657–1666.
- Kotrebai M, Birringer M, Tyson JF, Block E, and Uden PC (2000) Selenium speciation in enriched and natural samples by HPLC-ICP-MS and HPLC-ESI-MS with perfluorinated carboxylic acid ion-pairing agents. *Analyst* 125: 71–78.

Lobinski R, Edmonds JS, Suzuki KT, and Uden PC (2000) Species-selective determination of selenium compounds in biological materials. *Pure and Applied Chemistry* 72: 447–461.

McSheehy S, Pannier F, Szpunar J, Potin-Gautier M, and Lobinski R (2002) Speciation of seleno compounds in yeast aqueous extracts by three-dimensional liquid chromatography with ICP-MS and ES-MS detection. *Analyst* 127: 223–229.

Michalke B and Schramel P (1998) Application of capillary zone electrophoresis-inductively coupled plasma mass

spectrometry and capillary isoelectric focusing-inductively coupled plasma mass spectrometry for selenium speciation. *Journal of Chromatography A* 807: 71–80.

Pryczynska K (2002) Determination of selenium species in environmental samples. *Microchimia Acta* 140: 55–62.

Sanz Medel A and Gonzalez EB (2001) Chiral speciation of trace elements: approaches to the speciation of seleno-aminoacid enantiomers in biological samples. *Journal of Analytical Atomic Spectroscopy* 16: 957–963.

## SEM

See **MICROSCOPY TECHNIQUES: Scanning Electron Microscopy**

## SENSORS

Contents

### Overview

### Amperometric Oxygen Sensors

### Calorimetric/Enthalpimetric

### Chemically Modified Electrodes

### Microorganism-Based

### Photometric

### Piezoelectric Resonators

### Tissue-Based

## Overview

**K Lau**, Dublin City University, Dublin, Republic of Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Sensor research has always been a very dynamic area. We have gone a long way since the primary form of sensor, litmus paper, was first made for crude pH measurements. Since the mid-1980s, there has been an explosion of research interests in sensor development as witnessed by the overwhelming amount of papers published during this period.

Sensor research has always been driven by commercial needs. It is also an area that crosses over several developing areas, and has benefited by the advances in material sciences, engineering, and

affordable computer technology. Consequently, in recent years, there is huge diversity of sensors available in the market place that covers industrial, environmental, and clinical applications. Yet, despite the large amounts of capital and human resources invested in the development of new devices, the best sensor commercially available is probably still the pH glass electrode in terms of its selectivity, sensitivity, dynamic range, response time, robustness, and ease of manufacture. The big successes in commercial terms since the late 1950s have been, in the chemical sector, the oxygen electrode based on the Clark cell, the fluoride electrode and ion-selective electrodes (ISEs) used in clinical analyzers, and the solid-state oxygen sensors based on zirconium oxide ( $\text{ZrO}_2$ ) and titanium oxide ( $\text{TiO}_2$ ) used in automobile air-to-fuel control; whereas more recently in the biosensor sector, the various glucose sensors based on glucose oxidase and the pregnancy test strips have been successful. There are clear advantages in using sensors in field (or *in vitro*) applications compared to

Lobinski R, Edmonds JS, Suzuki KT, and Uden PC (2000) Species-selective determination of selenium compounds in biological materials. *Pure and Applied Chemistry* 72: 447–461.

McSheehy S, Pannier F, Szpunar J, Potin-Gautier M, and Lobinski R (2002) Speciation of seleno compounds in yeast aqueous extracts by three-dimensional liquid chromatography with ICP-MS and ES-MS detection. *Analyst* 127: 223–229.

Michalke B and Schramel P (1998) Application of capillary zone electrophoresis-inductively coupled plasma mass

spectrometry and capillary isoelectric focusing-inductively coupled plasma mass spectrometry for selenium speciation. *Journal of Chromatography A* 807: 71–80.

Pryczynska K (2002) Determination of selenium species in environmental samples. *Microchimia Acta* 140: 55–62.

Sanz Medel A and Gonzalez EB (2001) Chiral speciation of trace elements: approaches to the speciation of seleno-aminoacid enantiomers in biological samples. *Journal of Analytical Atomic Spectroscopy* 16: 957–963.

## SEM

See **MICROSCOPY TECHNIQUES: Scanning Electron Microscopy**

## SENSORS

Contents

### Overview

### Amperometric Oxygen Sensors

### Calorimetric/Enthalpimetric

### Chemically Modified Electrodes

### Microorganism-Based

### Photometric

### Piezoelectric Resonators

### Tissue-Based

## Overview

**K Lau**, Dublin City University, Dublin, Republic of Ireland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Sensor research has always been a very dynamic area. We have gone a long way since the primary form of sensor, litmus paper, was first made for crude pH measurements. Since the mid-1980s, there has been an explosion of research interests in sensor development as witnessed by the overwhelming amount of papers published during this period.

Sensor research has always been driven by commercial needs. It is also an area that crosses over several developing areas, and has benefited by the advances in material sciences, engineering, and

affordable computer technology. Consequently, in recent years, there is huge diversity of sensors available in the market place that covers industrial, environmental, and clinical applications. Yet, despite the large amounts of capital and human resources invested in the development of new devices, the best sensor commercially available is probably still the pH glass electrode in terms of its selectivity, sensitivity, dynamic range, response time, robustness, and ease of manufacture. The big successes in commercial terms since the late 1950s have been, in the chemical sector, the oxygen electrode based on the Clark cell, the fluoride electrode and ion-selective electrodes (ISEs) used in clinical analyzers, and the solid-state oxygen sensors based on zirconium oxide ( $\text{ZrO}_2$ ) and titanium oxide ( $\text{TiO}_2$ ) used in automobile air-to-fuel control; whereas more recently in the biosensor sector, the various glucose sensors based on glucose oxidase and the pregnancy test strips have been successful. There are clear advantages in using sensors in field (or *in vitro*) applications compared to

conventional analytical instruments and therefore the search for better performing and cheaper sensors will always interest the researchers in both the industrial and academic worlds.

This article will concern itself only with devices that involve a chemical or biochemical transduction mechanism to generate the analytical information, with the processes occurring in a membrane or layer attached to the probe in such a manner that the analytical information can be accessed electronically from the outside world. This covers sensors that are for single use and for continuous monitoring because the basic chemistry and sensor configuration used are very similar for a particular application. Hence, the article does not cover techniques such as open-cell Fourier transform infrared or remote fiber spectroscopy, which can be used to 'sense' the chemical nature of the environment without involving the use of a bona fide sensor.

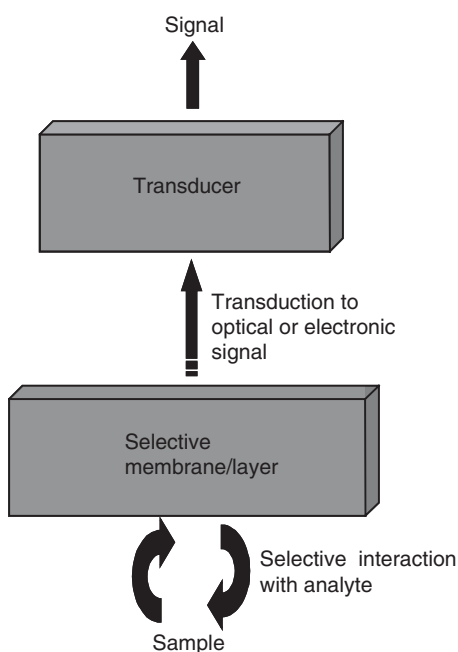
## Signal Generation

Figure 1 is a schematic representation of the main components of an idealized chemical sensor. The most important part of any chemical sensor is the selective membrane/layer, usually positioned on the sensor tip, which interacts with the sample in such a way as to provide analytically useful information about a particular component, the analyte, in the sample. In order to do this, the interaction of this

membrane/layer with the sample must be selective, i.e., it must be able to distinguish the analyte from the sample matrix. This is achieved by immobilizing receptor sites for the analyte species in the membrane/layer. Many different mechanisms have been used to achieve the required selectivity:

- ion-cavitand 'best-fit' interaction;
- electrostatic interactions;
- acid–base reaction;
- redox reaction;
- antibody–antigen interactions;
- enzyme–substrate reaction;
- ion-exchange reactions.

Generally, a membrane is configured in such a way that the selective chemical process or processes occurring at the membrane–sample boundary are coupled to a transducer using either electronic or optical means in order to make the signal accessible. This signal consists of an analytically useful component (i.e., related to the amount of analyte in the sample) in association with a component that does not convey useful information. The latter component can be subdivided into random noise, experimental errors, and errors arising from contributions from other species in the sample matrix to the signal (interferents). This component is minimized by maximizing the selectivity of the sensor, or by maintaining its contribution at a fixed level using techniques such as standard addition.



**Figure 1** Generalized scheme of the main elements of a sensor.

## Advantages of Sensors over Instrumental Methods

Although results obtained from sensor-based analytical methods are becoming more precise and accurate, instrumental methods are generally superior. However, sensors find their niche in applications that involve field measurements, measurements in remote or inaccessible places (e.g., *in vivo* monitoring), or where dictated by balancing cost against the degree of precision and accuracy required (sensor-based measurements are generally cheaper, and can be easily automated). Perhaps most importantly, sensors can often provide information as to how the target species concentration varies in real time, a function that is usually not available from analytical instruments.

## Sensor Characteristics

The ideal sensor should possess the following characteristics:

(1) 'Specificity for the target species'. This means that fluctuations in the signal obtained can be



unambiguously assigned to the target species. The function relating the response ( $S_i$ ) to the analyte concentration ( $C_i$ ) ideally should be linear, or easily linearized (although improvements in data processing have relaxed this constriction somewhat), i.e.,

$$S_i = M_i C_i + b_i \quad [1]$$

where  $M_i$  is the slope of the response function (selectivity) and  $b_i$  is the intercept on the signal axis.

In reality, sensors are selective rather than specific, and respond to greater or lesser degree to other chemical or biological species ( $j$ ) present in the sample:

$$S_i = M_i C_i + b_i + \sum_{j \neq i} K_j C_j \quad [2]$$

where the summation factor above incorporates the contribution of all interfering species ( $j$ ) along with a weighting coefficient  $K_j$ . In the ideal sensor, these coefficients are all zero, leading to a specific response to the analyte ( $i$ ) only.

A large amount of sensor research has been investigated in developing more selective sensors in order to minimize the error arising from the summation factor above, and to devise efficient methods for quantifying selectivity, so that the limitations of applicability of a particular device can be clearly identified. This is still a key area of research at present for most types of sensors. In addition to improving the selectivity of individual sensors, the use of arrays of sensors and advanced data-processing techniques has led to enhancements in selectivity through interpretation of the patterns obtained. With a multivariate approach, the sensor response characteristics are modeled, and the weighting factors for important interferences determined for each sensor in the array. This modeled error can then be subtracted from the analytical signal to give more accurate results.

(2) ‘Sensitivity to changes in the target species concentration’. The response slope ( $M_i$  in eqn [1]) of the sensor should be as large as possible to enable very small changes in the target species concentration to be detected. In some cases, it is possible to enhance the sensitivity by coupling the selective membrane interaction between the sensor and the analyte with further reactions (chemical amplification).

(3) ‘An extremely fast (instantaneous) response time’. This enables high-speed fluctuations in the analyte concentration to be monitored. For many applications, time constants of up to a few second can be tolerated.

(4) ‘An extended lifetime of at least several months’. This can be a major problem, particularly with biosensors based on enzyme or antibodies due to

gradual denaturation in the former and irreversible interactions in the latter. However, devices such as the glass electrode and other ISEs can have lifetimes extending into years when used carefully.

(5) ‘Small size and manufactured to high precision’. The integration of semiconductor manufacturing technologies with sensor fabrication is already an important aspect of the present sensor preparation techniques. These devices are generally small, solid state, well characterized, and identical in performance. These sensors have found their places in disposable sensor market and have important contributions to the progress into the use of sensor arrays rather than single devices. Advances in the areas of electronics and computing will enable powerful signal processing facilities to be made available in small palm-top instruments linked to the sensor arrays.

## Types of Sensors

It is possible to categorize sensors in many different ways including the mode of transduction (as in this case), the application (clinical sensors, gas sensors), or the size (nano, micro, mini, macro, industrial, bench-type sensors, etc.). The main differences in terms of operation and mode of transduction are summarized in Table 1.

### Electrochemical

#### Potentiometric

*Ion-selective electrodes* This is a classical field of chemical sensors, which bridges multidisciplinary research areas of host–guest chemistry, engineering, and material (membrane) science. In these devices, the selective membrane is symmetrically bathed by two electrolyte solutions, the internal reference solution and the external sample solution. The potential of an ISE is monitored with respect to an ideally invariant potential generated by a reference electrode under equilibrium conditions (i.e., zero current). Changes in the cell potential are related to the composition of the sample solution via the Nernst and Nicolsky–Eisenmann equations, which predicts a logarithmic relationship with the analyte or primary ion activity. While the logarithmic scale often leads to wide dynamic ranges for ISEs (typically four to five orders of magnitude, and up to 12 orders of magnitude with the glass electrode and the recently improved low detection limit ISEs), the fact that a rather vague thermodynamic quantity (activity) is monitored rather than concentration has led to difficulties in certain areas of application.

**Table 1** Sensor types classified by mode of transduction

<i>Sensor type</i>	<i>Chemical process involved</i>	<i>Parameter monitored</i>	<i>Conditions under which measurement is made</i>
Potentiometric	Spontaneous processes, e.g., movement of ions into membranes	Cell potential ( $E$ )	Zero current, i.e., equilibrium
Amperometric	Electrochemical reactions driven by an external potential	Limiting current $I_l$	Constant external potential ( $E$ )
Piezoelectric	Movement of material into or out of a surface or layer	Change in mass	External AC signal
Thermal	Any process involving a change in enthalpy ( $\Delta H$ )	Change in enthalpy	No heat transfer to surroundings
Optical	Any process resulting in a change in the optical properties of the immobilized selective membrane	Absorbance, fluorescence, refractive index, etc.	Detector must be shielded from extraneous radiation

The oldest, and still the best, ISE is the glass electrode, which was commercialized by the late 1920s. The introduction of poly(vinyl chloride)/neutral carrier membranes in the early 1970s led to a series of devices for cations ( $H^+$ ,  $Na^+$ ,  $K^+$ ,  $Ca^{2+}$ ,  $Mg^{2+}$ ,  $Ba^{2+}$ ,  $Cs^+$ ,  $NH_4^+$ ,  $Ag^+$ ) and anions ( $NO_3^-$ ,  $Cl^-$ ,  $HCO_3^-$ ). Solid-state ISEs (coated wire electrodes) have also been developed in which the sensitive membrane is coated directly onto a metal wire, usually a silver–silver halide. While these have the advantage of being small and easy to fabricate, they have been noted for their unpredictable properties and suffer from lifetime and stability problems. More sophisticated approaches involve the use of semiconductor planar fabrication technologies to deposit ion-sensitive layers onto semiconductor substrates to produce ion-selective field-effect transistors. These are conceptually very attractive but it has proven very difficult to produce devices as good as the equivalent ISE.

Solid-state (crystalline) devices have also proven successful, particularly with respect to the excellent fluoride electrode, and to a lesser extent with devices based on silver sulfide and related substances ( $Ag^+$ ,  $Cu^{2+}$ ,  $Cd^{2+}$ ,  $Pb^{2+}$ ). Potentiometric biosensors incorporating enzymes and antibodies have also been reported. In these devices, the biological reaction either consumes or produces a species that can be monitored by an ISE directly (commonly  $H^+$ ,  $NH_4^+$ ), or it is coupled with a second reaction. Examples include a sensor for glucose, which utilizes the coupling of the oxidation of glucose by glucose oxidase with the activity of the fluoride ions through the action of a second catalytic reaction on an organofluorine compound, and similar sensors for urea, penicillin, malate, asparagines, and sucrose. In a similar manner, a range of gas sensors have been produced that enable the gas to be monitored via the interaction of the gas with a buffer solution immobilized at the sensor tip by means of a gas permeable

membrane. Most involve the use of hydrogen ion-selective electrode and a pH buffer. Suitable gases include sulfur dioxide, carbon dioxide, ammonia, amines, and other acidic and basic gases. In addition, many devices have been reported that employ an ion-pair approach to enable ionizable drugs or organic ions such as histamine, nicotine, heroin, phenylephrine (cations) and carboxylates, amino acids, bile acids, and valproate (anions) to be monitored.

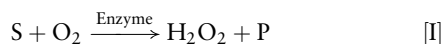
ISEs represent a mature sensor technology based on well-understood mathematical principles. Up until recently, the research areas have been mainly focused on applications rather than fundamental research. However, this research field has reinvented itself and a number of new and important research directions have emerged. The ongoing research to develop more efficient ionophores and also to expand the range of ions that can be successfully monitored are: the use of new membrane materials with low glass transition temperatures to eliminate the use of plasticizers and the use of redox membrane materials (e.g., polypyrrole) that act as ion-selective membranes and the solid contact at the same time; the studies of nonclassical response mechanisms; and, more significantly, the investigations and eventually breakthrough in obtaining low detection limits to picomolar levels, resulting in a class of very inexpensive sensors with detection limits comparable to inductively coupled plasma-mass spectrometry.

**Amperometric sensors** In contrast to potentiometric measurements where the monitored parameter, the cell potential, arises from spontaneous reactions, in amperometry the cell reaction is driven by an external fixed potential, and the current is monitored. With these sensors, in the idealized ‘specific sensor’ case, the signal resulting from a redox reaction is proportional to the concentration of the analyte species. Cells can be either two-electrode (working and reference) or three-electrode (working, counter, and

reference) systems, with the latter being introduced to circumvent problems arising from drawing current through the reference electrodes.

Electrochemical biosensors or electrochemical sensors for monitoring biologically important species are the fastest developing area of this sensor field. There are now commercially available devices for clinical, food, and environmental applications. Suitable target species for amperometric sensors are thus electroactive species (i.e., capable of being oxidized or reduced electrochemically), with the oxidation/reduction potential being as near zero volts as possible, as the number of possible interferents will increase with the magnitude of the external potential.

The best-known amperometric sensor is the Clark oxygen cell, developed by Clark in 1956. A wide range of biosensors have been reported in the literature in which the concentration of an enzyme substrate is measured indirectly through the consumption of oxygen by oxidase-catalyzed reactions or, in a similar manner, by the generation of hydrogen peroxide, i.e.,

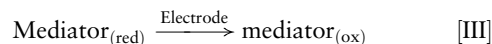
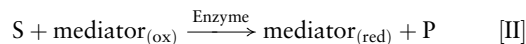


where S = substrate and P = products.

In most of these, the bioselective layer is deposited directly on the surface of a Clark-type working electrode using a variety of designs and immobilization techniques. While oxygen and hydrogen peroxide can be monitored successfully by polarizing the working electrode at  $-0.7\text{ V}$  or  $+0.7\text{ V}$ , respectively, the high potentials used leave the measurements open to interference from other redox-active species in the sample. In addition, the successful application of this approach depends on a sufficient oxygen supply, which can be problematic at high substrate concentration. To overcome these problems, there has been great interest in the development of efficient mediators that can shuttle electrons from the enzyme-catalyzed reaction through to the electrode surface at moderate external potentials. Molecules such as ferrocene and its derivatives, hexacyanoferrate(II)/(III) and benzoquinones, have been successfully used in this context, by coimmobilization in the bioselective layer. Attention has also been focused on the development of activated carbon paste electrodes, normally doped with small metal particles such as palladium, which functions as catalyst to reduce the overpotential required for hydrogen peroxide detection.

The most successful amperometric sensor in the 1990s is probably the various diabetes home care systems based on glucose sensor strips. These enzyme

electrodes are principally similar and are based on enzymatic reactions where a mediator is used for signal generation, i.e.,



where P = products.

Selectivity in these devices can be optimized through selection of moderate operating potentials, the use of enzymes that will selectively oxidize or reduce the analyte, or through selective control of mass transport via an additional outer permselective membrane. Mechanisms that have been used for this purpose include size (hydrolyzed cellulose acetate) and charge (Nafion<sup>TM</sup>) exclusion.

A major problem with all types of biosensor, and antibody-based devices in particular, is the short lifetime and irreversibility of the reactions involved. For monitoring purposes, it is just as important for the signal to return to baseline after a transient increase in the signal, as it is to detect the increase itself. This requires that the chemical processes in the sensor membrane that generate the observed signal should be fast and reversible, with stability constants ( $\log \beta$ ) for products typically lying in the 5–6 range. Most antibody-based reactions have stability constants many order of magnitude in excess of this, and regeneration of the original layer (i.e., removal of the substrate) usually requires extreme conditions that effectively destroy the biomembrane. Hence, reversibility is a major and continuing problem with many biosensors, and the subject of intensive research at present.

Various approaches have been investigated for immobilizing the bioreagents including direct covalent linkage, physical entrapment using porous membranes or hydrogels, incorporation into Langmuir–Blodgett (L–B) films, cross-linking with collagen and other cross-linking agents, and molecular wire approach where the enzyme is covalently attached to a redox polymer. Miniaturization has become an important research area especially for *in vivo* sensing. These small implantable electrodes modified with specific coatings are useful for online monitoring of very low concentration metabolites such as dopamine, nitric oxide, and insulin as well as important analytes such as glucose and ascorbate. Major problems still exist in the area of *in vivo* clinical monitoring including biocompatibility, stability, selectivity with respect to other matrix components such as drugs administered during therapy, and calibration.

## Optical Sensors

The development of inexpensive, efficient fiber-optic cables for applications in telecommunications and technological advances in the semiconductor sector has together stimulated much research into optical sensors. Typically, these devices incorporate a material at the tip or on the side of the cable (evanescent wave devices), which can generate an optical signal related to the concentration of the target species in the sample. Both fluorescence and absorbance have been used for this purpose, via both direct and indirect mechanisms. The fluorescence method has attracted much interest in recent years due to its superior sensitivity and low volume required. An important advantage of fluorescent measurements over absorbance/reflectance measurements is that it is possible to measure the fluorescence lifetime instead of intensity, therefore minimizing problems such as turbidity and ambient light associated with intensity-based measurements. The dual lifetime referencing scheme for converting fluorescence intensity into frequency domain or time domain has since been developed to include two fluorescent dyes, one active and the other inert to act as a reference. These systems provide a unique means to referencing the optical signal against background interferences including drifts resulting from optoelectronic components.

Various methods have been used for immobilizing the optically responsive material including physical entrapment with permeable layers and incorporation into L-B and sol-gel films. Sensors for hydrogen and other ions is a popular topic for research, and much activity continues in the design and synthesis of new materials and cocktails for these optodes. This can be achieved by functionalizing ionophores with ionizable chromophors or fluorophores or by arranging reactions that involve deprotonation of a chromophore to charge compensate the sequestering of the cation from the sample phase by the ionophore. Attachment of biological functionalities to molecular building blocks that incorporate fluorophores is an attractive proposition for the development of new types of immunoassays in particular. Reversibility is once again the key obstacle. Optical sensors for many gases including oxygen, ammonia, carbon monoxide, chlorine, anesthetics, and volatile organic compounds (VOCs) have also been described.

Another area of growing importance in optical sensing is the surface plasmon resonance (SPR). This technique makes use of the surface plasma wave (SPW) that may exist at the interface of two media with dielectric constants of opposite signs, e.g., a metal and a dielectric. The excitation of the SPW by

an optical wave results in resonant transfer of energy from the incident light into the SPW. This gives rise to resonant absorption of the energy of the optical wave to manifest SPR. Variations in the optical properties of the dielectric (the transducing medium) adjacent to the metal layer induce changes in the propagation constant of SPW. This change of optical parameters in the dielectric therefore can be detected by monitoring the interaction between the SPW and the optical wave.

The main feature of the SPR optical system comprises a radiation source and an optical structure where the changes in optical properties of the transducing medium, normally modified with specific (bio)chemical receptors, is interrogated. As the propagation constant of SPW is always higher than that of optical wave propagation in the dielectric, in order to excite SPW efficiently the momentum of the incident light is enhanced to match that of the SPW by using techniques such as diffraction at the surface of diffraction grating and attenuated total reflection in prism couplers and optical wave guides. The detection approaches commonly used are measurement of the intensity of the optical wave near resonance, and measurement of the resonant momentum of the optical wave including angular and wavelength interrogation of SPR. The applications of SPR are diverse and include the study of optical properties in metal films, film thickness, and refractive index measurements of organic layers on metal surfaces, protein adsorption to biomolecules, and the adsorption of gas molecules. Utilizing SPR as a biosensor has led to the first commercial SPR biosensor (BIAcore) launched in 1990. Since then increasing number of SPR systems employing different sensing approaches have been made available.

The main disadvantages of SPR-sensing devices are their lack of sensitivity when monitoring low molecular weight adsorbates and the limited use in kinetic analysis due to slow mass transport. These are still important directions for SPR sensor research and much effort is still being put in enhancing sensitivity and the study of nonspecific biomolecular interactions. On the instrument side, research into miniaturization and multichannel approach for high throughput analysis are emphasized.

## Mass Sensors

In these relatively new sensors, an acoustic wave is propagated by an externally applied alternative current between two electrodes or interdigitated electrodes deposited on a piezoelectric substrate such as quartz. There are several subclasses of piezoelectric transducers (bulk wave, surface acoustic wave, or

flexural plate wave) based on the way the acoustic wave is propagated between the electrodes. The fundamental frequency of the acoustic wave depends on the surface loading of the crystal, enabling these devices to function as very sensitive microbalances. Coating the crystal surface with a material, which will selectively interact with a sample phase, either liquid or gaseous, can induce a selective response. The increase in fundamental frequency obtained is simply related to the increase in mass in the surface layer, which in turn may be related through appropriate calibration to the bulk sample concentration.

Most applications for these devices have been for gas-phase monitoring, although recently, aqueous-phase applications are receiving more attention (glucose, immunosensors). In gas phase, hydrogen sulfide, carbon dioxide, oxygen, nitrogen dioxide, molecular hydrogen, mercury, formaldehyde, VOC sensors have been fabricated. A major problem has been the response of these transducers to humidity, which means that the humidity must be controlled or else compensated for with a reference transducer in different mode measurements. Selectivity has been another problem, although the use of arrays of piezoelectric sensors with different coatings and pattern recognition techniques can improve the situation markedly.

Specialized designs are available that enable acoustic sensors to be used as the working electrode in a voltammetric cell. This enables exciting fundamental experiments to be performed, where the mass changes during a voltammetric scan can give important insight into the mass transfer processes occurring at the electrode surface, although, as with any new technique, the results have to be interpreted with great care.

### **Calorimetric (Thermal) Sensors**

Almost every reaction that occurs is accompanied by a change in enthalpy, which can be measured as a rise or fall in temperature in adiabatic systems of known specific heat capacity. Furthermore, the reaction of interest can often be coupled with other reactions that can amplify the enthalpy change, a common example being enzyme-catalyzed reactions that involve a proton exchange being performed in tris(hydroxymethyl)aminomethane buffer.

The best known of these are the enzyme thermistors. Thermistors are devices whose resistance changes markedly with temperature and consequently are often employed as inexpensive, sensitive temperature sensors. In early designs, the matched thermistors were sited in a flow system downstream of packed enzyme and reference columns. Interaction

of enzyme with the substrate in the flowing stream produced a temperature change, which could be related to the concentration of substrate in the sample. More recently, the enzyme coating has been applied directly to the thermistor surface, with glucose once again being the most common substrate targeted. New sensors based on pyroelectric materials are revitalizing this rather dormant area of sensor research.

## **Sensor Developments**

### **Miniaturization and Arrays**

There is increasing interest in fabrication technologies that will enable miniaturized versions of conventional sensors to be produced with very reproducible characteristics. This will accelerate the trend toward the use of sensor arrays rather than the traditional standalone approach. Arrays can give much more information than the sum of individual components, and through self-diagnostics, the application of pattern recognition techniques, and rigorous calibration regimes, the information will be much more dependable and robust. Analysts have gradually adopted these methods as the instrumentation required for multichannel data collection is readily available; they are normally equipped with data collection/chemometric analysis software. The more robust and flexible methods offered by using arrays help to stimulate commercial activity that will in turn generate more finance for research. Sensor design is moving rapidly from the traditional, hand-assembled bench electrodes that have been the norm for the past 50 years or more, to planar technologies developed by the semiconductor industries. Deposition techniques such as vapor deposition, sputtering, screen printing, ink and bubble-jet printing have become standard technologies to place the correct amount of active components in micro- and miniaturized sensors in exactly the right place on the substrate.

### **Materials**

Materials research has been influencing the direction and diversity of sensor research. Novel semiconductor and superconductor materials have led to new and more efficient transducers. Advances in biotechnology have also resulted in a broader range of enzymes and monoclonal antibodies for use in biosensors. Examples for utilizing tissue and microbes in biosensors include the use of tomatoes, eggplant, grapes, horseradish, parsley, and Antarctic krill in electrochemical biosensors. The logic is that the active component (usually an enzyme) is protected by

the microenvironment offered by the tissue and can thus function for a longer time compared to the more expensive purified enzyme. Supramolecular chemistry is becoming a very lively area in synthetic research. Crown ethers, calixarenes, cavitands, and molecules with similar ion-complexing properties are now being used as the building blocks of more sophisticated structures containing chromophores, ionophores, and enzymes/antibodies. The possibility exists of producing molecules that can generate information by optical or electrochemical transduction of the same chemical process. Chemists are now calling some of these substances 'designer' molecules as the receptor-site morphology can be modeled using three-dimensional computer graphics in order to give a 'best fit' for the target species.

### Flow Injection Analysis

The coupling of sensors with flow injection analysis (FIA) is already a very popular option. The flow regime offers important advantages over discrete manual measurements that include (1) Sample preparation: processes such as reagent mixing, selectivity enhancement (e.g., removal of large molecular mass interferents such as protein by dialysis in clinical assays), and solvent extraction can all be carried out online. The improved sample preparation and more reproducible sample delivery result in improved measurement precision and accuracy. Drift is less of a problem as measurements are made of peak heights relative to a baseline. (2) Improved sensor lifetime: in flow analysis, the sensor may be exposed to the sample for only a short period of time, and maintained in a friendlier matrix between measurements that can help counteract or delay the deleterious effects of the sample. (3) Automation: the entire analysis can be

automated. Valves, samplers, and peristaltic pumps can be easily controlled from a personal computer. Flow speed and direction of flow can be controlled with very high precision and accuracy. Very high sample throughputs can be achieved, with more than 100 assays per hour being regularly reported. Furthermore, the combination of sensor arrays, sophisticated data processing, and FIA offers high levels of precision and accuracy in analysis coupled with almost instantaneous feedback (results are generated typically in 30–60 s).

**See also:** **Ion-Selective Electrodes:** Overview. **Process Analysis:** Sensors. **Sensors:** Amperometric Oxygen Sensors; Chemically Modified Electrodes; Piezoelectric Resonators.

### Further Reading

- Blum LJ and Coulet PR (eds.) (1991) *Biosensors Principles and Applications*. New York: Marcel Dekker.
- Catrrall RW (1997) *Chemical Sensors*. New York: Oxford University Press.
- Diamond D (1993) Progress in sensor array research. *Electroanalysis* 5: 795–802.
- Eggins BR (2002) *Chemical Sensors and Biosensors*. New York: Wiley.
- Fundamental Reviews published every even year in the journal *Analytical Chemistry*. The most recent is 2000 and 2002.
- Sensors Forecasts to 2006 & 2011. Published by The Freedonia Group, 2002/04.
- Surface Plasmon Resonance Sensors: review published in *Sensors and Actuators B* 54: 3–15 (1999) and review of surface plasmon resonance analysis of dynamic biological interactions with biomolecules in *Biomaterials* 21: 1823–1835 (2000).

## Amperometric Oxygen Sensors

**P C Hauser and S S Tan**, The University of Basel, Basel, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The determination of oxygen, mainly in dissolved form in aqueous samples, is by far the largest field of application for amperometric sensors. Because of their reliability, ease of use, and the fact that results are available almost instantaneously, these devices are widely used.

### Principles

Amperometric sensors are small electrochemical cells consisting of two or three electrodes that are usually combined in a single body. A constant potential is applied, i.e., the sensor operates as a Faradaic cell, and a dependence of the measured current on the analyte concentration in the sample is obtained. As in ordinary amperometry, this requires a diffusion layer on the surface of the working electrode. This diffusion layer, in which the analyte concentration is depleted, arises because the analyte is consumed in the electrode reaction. In order for this depletion to occur, the electrode kinetics has to be faster than the



the microenvironment offered by the tissue and can thus function for a longer time compared to the more expensive purified enzyme. Supramolecular chemistry is becoming a very lively area in synthetic research. Crown ethers, calixarenes, cavitands, and molecules with similar ion-complexing properties are now being used as the building blocks of more sophisticated structures containing chromophores, ionophores, and enzymes/antibodies. The possibility exists of producing molecules that can generate information by optical or electrochemical transduction of the same chemical process. Chemists are now calling some of these substances 'designer' molecules as the receptor-site morphology can be modeled using three-dimensional computer graphics in order to give a 'best fit' for the target species.

### Flow Injection Analysis

The coupling of sensors with flow injection analysis (FIA) is already a very popular option. The flow regime offers important advantages over discrete manual measurements that include (1) Sample preparation: processes such as reagent mixing, selectivity enhancement (e.g., removal of large molecular mass interferents such as protein by dialysis in clinical assays), and solvent extraction can all be carried out online. The improved sample preparation and more reproducible sample delivery result in improved measurement precision and accuracy. Drift is less of a problem as measurements are made of peak heights relative to a baseline. (2) Improved sensor lifetime: in flow analysis, the sensor may be exposed to the sample for only a short period of time, and maintained in a friendlier matrix between measurements that can help counteract or delay the deleterious effects of the sample. (3) Automation: the entire analysis can be

automated. Valves, samplers, and peristaltic pumps can be easily controlled from a personal computer. Flow speed and direction of flow can be controlled with very high precision and accuracy. Very high sample throughputs can be achieved, with more than 100 assays per hour being regularly reported. Furthermore, the combination of sensor arrays, sophisticated data processing, and FIA offers high levels of precision and accuracy in analysis coupled with almost instantaneous feedback (results are generated typically in 30–60 s).

**See also:** **Ion-Selective Electrodes:** Overview. **Process Analysis:** Sensors. **Sensors:** Amperometric Oxygen Sensors; Chemically Modified Electrodes; Piezoelectric Resonators.

### Further Reading

- Blum LJ and Coulet PR (eds.) (1991) *Biosensors Principles and Applications*. New York: Marcel Dekker.
- Cattrall RW (1997) *Chemical Sensors*. New York: Oxford University Press.
- Diamond D (1993) Progress in sensor array research. *Electroanalysis* 5: 795–802.
- Eggins BR (2002) *Chemical Sensors and Biosensors*. New York: Wiley.
- Fundamental Reviews published every even year in the journal *Analytical Chemistry*. The most recent is 2000 and 2002.
- Sensors Forecasts to 2006 & 2011. Published by The Freedonia Group, 2002/04.
- Surface Plasmon Resonance Sensors: review published in *Sensors and Actuators B* 54: 3–15 (1999) and review of surface plasmon resonance analysis of dynamic biological interactions with biomolecules in *Biomaterials* 21: 1823–1835 (2000).

## Amperometric Oxygen Sensors

**P C Hauser and S S Tan**, The University of Basel, Basel, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The determination of oxygen, mainly in dissolved form in aqueous samples, is by far the largest field of application for amperometric sensors. Because of their reliability, ease of use, and the fact that results are available almost instantaneously, these devices are widely used.

### Principles

Amperometric sensors are small electrochemical cells consisting of two or three electrodes that are usually combined in a single body. A constant potential is applied, i.e., the sensor operates as a Faradaic cell, and a dependence of the measured current on the analyte concentration in the sample is obtained. As in ordinary amperometry, this requires a diffusion layer on the surface of the working electrode. This diffusion layer, in which the analyte concentration is depleted, arises because the analyte is consumed in the electrode reaction. In order for this depletion to occur, the electrode kinetics has to be faster than the

rate of transport of the analyte to the electrode surface. For a linear dependence of current on concentration it is also essential that the transport through this depletion layer is by diffusion only and not by either of two other possible processes: electrostatic migration and convection.

The flux of the analyte to the electrode surface and therefore the current can be derived from the diffusion coefficient and the concentration gradient using Fick's law of diffusion. It is usually possible to employ the so-called Nernst approximation where the concentration ( $c$ ) is taken to linearly decrease from a certain distance ( $\delta$ ) where the concentration corresponds to the bulk of the sample ( $c_\infty$ ) to a concentration of zero on the electrode surface itself. This situation is illustrated in **Figure 1**.

The current is then given by the following combination of Fick's law of diffusion and the Faraday equation:

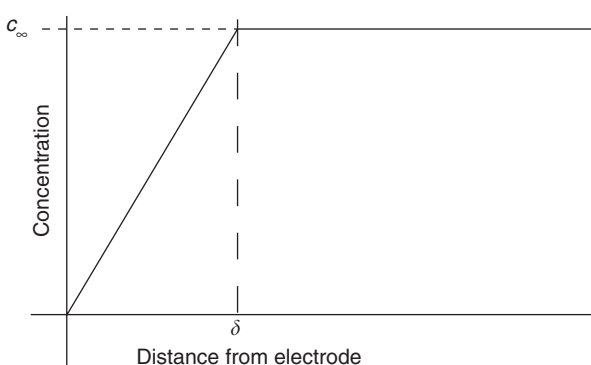
$$i = nFD \frac{c_\infty}{\delta} \quad [1]$$

where  $i$  is the measured current,  $n$  the number of electrons exchanged in the redox reaction,  $F$  the Faraday constant, and  $D$  the diffusion coefficient. From this it follows that the current is linearly dependent on the concentration in the bulk of the solution.

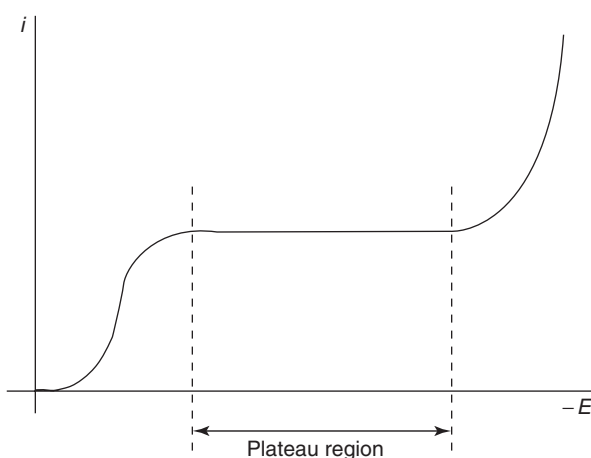
For ordinary amperometry in solution, the diffusion layer is formed as part or all of the stagnant or nonstirred layer of liquid that exists on solid surfaces. This layer does not take part in convection processes and therefore any flux of material through it is solely by diffusion. The thickness of the stagnant layer of the solution is determined by the flow velocity of the solution across the surface and therefore by the rate of stirring, the flow rate of the liquid, or the movement or rate of rotation of the electrode. If the diffusion layer extends over the entire thickness of the stagnant layer, then these flow parameters affect the performance of the cell. In amperometric sensors a membrane usually separates the working electrode from the sample solution. Often, the diffusion layer is largely confined to this membrane, so that the flow rate dependence is strongly reduced or even eliminated.

### Current/Voltage Behavior

A typical response of an amperometric cell is illustrated in **Figure 2**. On increasing the voltage, a rise in current is obtained when the redox potential of the species sensed is reached. A current plateau is reached for the diffusion limited case as discussed above, since this transport process is independent of the applied voltage. If the voltage is increased further,



**Figure 1** Concentration profile at an electrode according to the Nernst approximation.  $c_\infty$  is the concentration in the bulk of the sample and  $\delta$  the distance from the electrode to which the concentration is depleted.



**Figure 2** Current ( $i$ ) versus voltage ( $E$ ) plot obtained in amperometry for a cathodic reaction. The plot shows the rise of the current when the reduction potential of the analyte is reached, the plateau region, and the final increase in current resulting from the reduction of an interfering species or the solvent.

a different reaction such as the reduction or oxidation of a secondary species in the sample or eventually the decomposition of the solvent will set in. It is important that the plateau for the analyte is relatively wide so that the sensor is not sensitive to small changes in the applied voltage. Some discrimination against interferences can be obtained by choosing an appropriate potential. This, however, does not allow suppressing species that are reduced or oxidized at lower potentials than the analyte. The selectivity may be improved by the elimination of an interferent by chemical means in the sample solution. It is sometimes also possible to incorporate a membrane into the sensor, which shows selective transport or which contains a catalyst (such as an enzyme) enabling lower applied potentials and thus leading to improved selectivity.

The amperometric sensor has to be completed with counter and reference electrodes in order to carry the cell current and define the applied potential, respectively. In conventional voltammetry or amperometry, a three-electrode configuration is commonly used where two different electrodes fulfill these functions. In the design of amperometric sensors two-electrode systems are often used for simplicity. The counter electrode then serves as a reference as well. This requires a tolerance to a steady current over prolonged periods of time without a significant change in electrode potential. In either case, the current on the counter electrode causes the oxidation or reduction of a species that has to be considered when designing the sensor. The material consumed has to be incorporated into the probe on construction or provided as a refillable electrolyte solution. The amount available determines the lifetime of the sensor unless some other process (e.g., fouling of a membrane) has a more significant effect.

## Sensors for Dissolved Oxygen

The determination of dissolved oxygen in aqueous systems is of great importance in environmental and clinical applications. Prior to the introduction of oxygen sensors this analysis was mainly carried out with a chemical titration procedure according to Winkler. In this method, Mn(II) is oxidized to Mn(III) by analyte oxygen in a basic solution. This is followed by the oxidation of  $I^-$  to  $I_2$  by Mn(III) in acidic solution and the subsequent titration of the iodine produced with thiosulfate. The procedure is therefore very labor and time intensive and does not allow online or *in vivo* measurements. The use of a sensor for oxygen determination represents a significant improvement and has largely replaced the older chemical method.

The concentration of dissolved oxygen in solution that is in equilibrium with a gas mixture is determined by its partial pressure in the gas volume according to Henry's law:

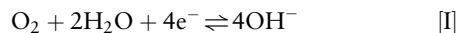
$$p = k'c \quad [2]$$

where  $p$  is the partial pressure,  $k'$  a constant that is specific for oxygen and the solvent used, and  $c$  the concentration in the solution. The partial pressure is dependent on the barometric pressure. For clean air the volume fraction of oxygen is a constant (20.095%) and may be used to derive the partial pressure from the barometric reading. However, air in contact with water will take up moisture and this will fractionally lower the partial pressures of all constituents in air. For precise oxygen determinations

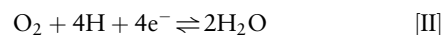
the partial pressure of water in air has to be considered and an appropriate correction for the partial pressure of oxygen also needs to be made.  $k'$  is dependent on temperature. The solubility of gases in water decreases with increasing temperature in a nonlinear fashion and is zero at the boiling point. A further dependence is found on the salt content of the solution. The solubility of oxygen decreases with increasing salt concentration. There is also a dependence on the pressure of the water body itself. The solubility of oxygen remains fairly constant, but the partial pressure increases with depth. This factor becomes important when determining dissolved oxygen in natural waters. Chemical methods like the Winkler titration method respond to the concentration, while amperometric oxygen sensors respond to partial pressure and therefore show this depth dependence. It is, however, possible to carry out a numerical correction based on thermodynamic principles for this pressure effect.

Oxygen can be reduced at a cathode according to the following reactions:

In basic solutions:

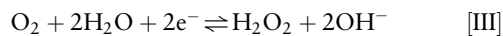


In acidic solutions:



The standard potentials,  $E^\circ$ , on platinum electrodes are +0.81 V at pH 7, +0.40 V at pH 14, and +1.23 V at pH 0. These standard potentials correspond to the potential where the rise in current is obtained in the current versus voltage plots. For basic solutions a plateau is obtained over several 100 mV before hydrogen evolution sets in at more negative potentials.

On a mercury electrode the reaction proceeds in two steps with distinct potentials:



A conventional polarograph with a dropping mercury electrode may indeed be employed for the determination of dissolved oxygen.

## Oxygen Sensing without Membranes

Currently, most amperometric sensors for dissolved oxygen contain a membrane to separate the working and counter electrodes from the sample solution. However, open cells had been used exclusively prior to the introduction of membrane based sensors. The electrolyte solution for these systems is the sample

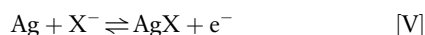
solution and the electrodes are therefore susceptible to interferences such as spurious electrode reactions by redox active substances in the sample and electrode poisoning. Another problem is the ill-defined counter electrode reaction that leads to a poor control of the cell voltage. Despite these drawbacks, open cells are still used when membrane contamination would be a problem (such as in sewage treatment). Frequent automated mechanical cleaning of the electrodes may be employed in such cases to eliminate electrode poisoning.

### Oxygen Sensing with Membranes

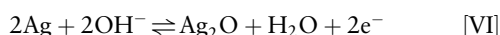
The disadvantages of the membraneless systems were largely overcome in 1956 by Clark who placed both working and counter electrodes behind a gas-permeable membrane that holds an internal electrolyte solution in place. This concept has since been widely accepted and this type of oxygen sensor is frequently called a Clark electrode. Note that the term electrode is used even though the sensor comprises a complete electrochemical cell.

A schematic drawing of a typical Clark-type cell is given in Figure 3. The working electrode is located behind the membrane separated by only a thin layer of electrolyte solution in order to achieve a fast response. Gold or platinum electrodes are almost universally used because of the scarcity of interfering reactions on these metals. In principle, it is possible to reduce oxygen at these electrodes over a wide pH range since oxygen reduction is obtained for acidic as well as basic solutions, albeit according to two different mechanisms as given in reactions [I] and [II].

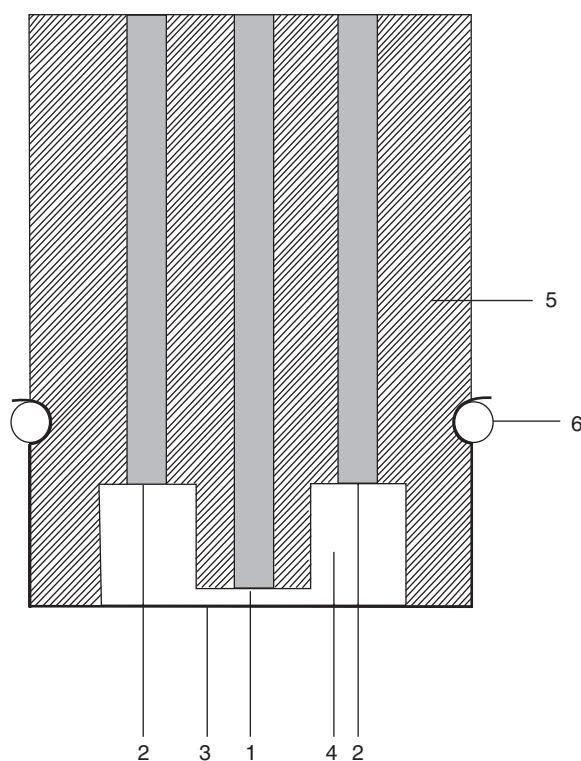
The counter electrode is commonly a silver anode, can be located at a distance from the membrane, and has to possess a relatively large surface area in order not to limit the current flow. It is necessary that the reaction products are obtained from a solid layer on the anode and do not enter the electrolyte solution as particles that may short-circuit the cell. The electrolyte solution is normally alkaline and the reference reaction is either:



or



A basic potassium chloride or bromide solution ( $\text{X}^- = \text{Cl}^-$  or  $\text{Br}^-$ ) is used as electrolyte solution. The depletion of the halide concentration in these cases over time will lead to the second reaction becoming the counter electrode process that will shift the cell potential to values  $\sim 120$  mV more positive.



**Figure 3** Schematic representation of the cross-section of a Clark electrode. 1, Au or Pt cathode; 2, Ag anode; 3, membrane; 4, electrolyte solution; 5, electrode body; 6, O-ring for fixing the membrane.

This may or may not shift the cell potential out of the plateau region for oxygen reduction, which would lead to a failure of the device because of hydrogen evolution. The basicity of the electrolyte solution tends to cause absorption of ambient carbon dioxide and the resulting lowering of the pH value may lead to hydrogen evolution and a failure of the device. This can be a problem for environments with high  $\text{CO}_2$  levels and may require frequent replacement of the electrolyte solution. Special Clark electrodes that use a  $\text{KHCO}_3$ -buffered electrolyte solution to prevent changes in pH value are available for critical cases. Silver is prone to sulfide poisoning, which may be a problem in environments containing high levels of sulfides.

### Membrane Materials

Different types of membrane materials such as polyethylene, silicon rubber, and poly(vinyl chloride) can be used. The sensitivity is higher and the response time of the sensor is shorter the thinner the membrane and the higher the permeability for oxygen. The permeability is a function of both solubility and diffusion rate of the gas in the polymer,

but is usually dominated by the latter. Silicone rubber has by far the highest permeability for oxygen, but in practice polyethylene or porous polytetrafluoroethylene (PTFE) of 10–20  $\mu\text{m}$  thickness is most often used, because these materials have high strength and can therefore be easily manufactured in thin sheets. For PTFE membranes the diffusion occurs through the pores that are not wetted due to the high hydrophobicity of the material. The membranes can usually be replaced by the user and are commonly held in place with an O-ring. It is important that on mounting the membranes are stretched flat across the electrode body without any creases as this would lead to irreproducible results.

### Interferences

The membrane material used imparts some degree of selectivity because the permeability varies for different gases. The diffusion rate and therefore the permeability is generally higher the smaller the molecule. On this principle, however, it is not possible to completely prevent gases other than oxygen from entering the electrolyte solution. Gases that have a reduction potential lower than that of oxygen on the precious metal cathode, such as  $\text{Cl}_2$ ,  $\text{NO}$ , and  $\text{CH}_2\text{O}$ , will be reduced and therefore will interfere. If these gases are only present at constant concentrations much lower than oxygen, this is not a serious problem, since it will result in a steady background current that can be electronically subtracted.  $\text{CO}_2$  interferes by lowering the pH value of the electrolyte solution as discussed above and  $\text{H}_2\text{S}$  and other sulfide gases tend to poison the silver anode. In severe cases it may be necessary to scrub interfering gases from the sample before measurements. Gases that cannot be reduced on the electrode, such as  $\text{NH}_3$ ,  $\text{CO}$ ,  $\text{CH}_4$ ,  $\text{N}_2$ ,  $\text{H}_2$ , and rare gases, do not interfere.

The cross-sensitivity to other gases has been employed to construct amperometric sensors for those species. Commonly used, for example, are sensors for dissolved chlorine (e.g., in swimming pools), which are very similar in construction to Clark electrodes for oxygen.

### Temperature Dependence

Amperometric oxygen sensors show a large dependence on temperature, an increase of current of  $\sim 2\% \text{ } ^\circ\text{C}^{-1}$  is observed. This is caused by the temperature dependence of the diffusion coefficient for oxygen in the membrane. A temperature sensitive element such as a thermistor is therefore often incorporated in the cell to allow electronic compensation for this source of error. It is obvious that

the temperature measuring component has to be located as close to the membrane as possible.

### Dynamic Range

Clark-type electrodes show a linear response over almost four orders of magnitude. The upper limit is determined by the saturation of the sample with dissolved oxygen ( $\sim 10 \text{ mg l}^{-1}$  in water at atmospheric pressure) while the lower limit is caused by the so-called residual current. This is due to impurities that can be reduced on the cathode, as well as trace oxygen present in the electrolyte solution or gassing out from the solid materials comprising the sensor body. A typical detection limit is  $0.05 \text{ mg l}^{-1}$  of dissolved oxygen.

### Sensitivity

The sensitivity of amperometric sensors can, in principle, be increased by making the working electrode surface larger. However, this may lead to an increased thickness of the diffusion layer that causes a reduction of the flux of the sensed species to the electrode and therefore a reduction in sensitivity. In practice, it is therefore often desirable to use microelectrodes. If the diffusion layer extends into the sample solution, increasing the stirring rate or the flow rate may lead to a reduction of the diffusion layer thickness and hence lead to an increased sensitivity as well as faster response times. Varying the thickness of a membrane may have a similar effect. The dependence of the sensitivity on flow rates may be a disadvantage and can again be overcome by using microsensors with diameters in the micrometer range. As an increase in temperature increases the diffusion coefficient this may be used to enhance sensitivity. However, the increase in current that comes with higher sensitivity also leads to a faster consumption of the counter electrode species and therefore to a reduced lifetime. A high sensitivity and therefore high rate of analyte consumption may also cause a depletion of the analyte species in the vicinity of the sensor leading to an erroneous low reading. This is, of course, particularly significant if small sample volumes are analyzed.

### Response Time

The response time of the sensor is given by the time required for establishment of equilibrium within the diffusion layer on changing the applied potential, or of more relevance, a change in the analyte concentration. This process is faster for thinner diffusion layers. In cases where the diffusion and the stagnant layers coincide, the response time can be controlled by stirring, flow rate, or rotation or other movement

of the sensor itself. If the diffusion layer is confined to a membrane it is the membrane thickness that determines the response time. The response time is then faster for thinner membranes. On the other hand, the diffusion layer for thin membranes might extend into the sample solution which again leads to a flow rate dependence.

The response time is usually expressed as the time required to reach 90% of the final change in current ( $t_{90}$ ). This is usually  $\sim 15\text{--}20$  s; however, it is necessary to wait for  $\sim 1$  min to obtain a stable reading. The response time of the electrode on switching on the cell voltage is, in principle, also determined by the time required for establishment of the diffusion layer. However, in practice, other factors, such as the reduction of an oxide layer on the electrode surface, the desorption of impurities from the electrode, and the increase of the pH value in the electrolyte solution layer immediately adhering to the electrode surface due to the production of hydroxide ions by the electrode reaction, further delay the attainment of the equilibrium current. The electrodes are therefore usually conditioned (polarized) by permanently applying the cell potential between measurements.

### Calibration

A calibration can be carried out by saturating water with air, pure oxygen, or known gas mixtures. The latter can be produced with flow meters or are commercially available in compressed form. It is also possible to use water saturated with air and mix it in known ratios with oxygen-free water. Alternatively, the oxygen content of different solutions may be established with a different procedure such as the chemical determination with the Winkler method. Once the linearity of the response has been established, it is, however, not always necessary to acquire a complete calibration curve. A two-point calibration check is usually adequate. The procedure is then similar to the calibration of pH electrodes. A solution of zero oxygen content and a solution with a known concentration are used to set the zero point and sensitivity of the electrode, respectively (intercept and slope of the calibration line). Solutions free of oxygen may be obtained by boiling, bubbling through an inert gas ( $\text{N}_2$ , Ar), or chemical decomposition with, e.g., 2% sodium sulfite. The solution used to adjust the sensitivity should show high oxygen content. The calibration for sensitivity may be carried out in air rather than in aqueous solution for simplicity. However, for precise measurements it has to be considered that the partial pressure of oxygen in air depends on the humidity of the air. To overcome this dependency it should be ensured that the air used for

calibration is saturated with moisture by holding the electrode in this process closely over the surface of clean water. The partial pressure of water for the given temperature and the resulting partial pressure of oxygen can be obtained from tables.

### Applications

The Clark electrode was originally conceived for use in online *in vivo* determination in blood streams during surgery. Noninvasive, even bedside, monitoring of blood oxygen content is now possible with the simple and elegant optical method of pulse oximetry (but note that it is oxygen saturation which is measured, not partial pressure as with the Clark electrode) but other applications remain highly important. Microelectrodes that allow high spatial resolution (e.g., measurements in single cells) are in use. The measurement of dissolved oxygen in blood samples is routinely carried out with Clark electrodes in clinical analysis as one of the standard parameters. The Clark electrode is also widely used in the determination of dissolved oxygen in natural waters (lakes, rivers, seawater) for environmental studies. In order to carry out measurements at large depths in natural waters at high pressure, special versions of the Clark electrode with pressure compensation are available, as are small sensors for use in spatial profiling of the sediment layer of water bodies. The determination of oxygen in fresh water supplies and boiler feed water is also often carried out. In the treatment of sewage, the continuous monitoring of oxygen content is highly important. The probes are widely used in the determination of biological and chemical oxygen demand of water. Clark electrodes have also been employed as the active sensing element in enzymatic sensors where oxygen is produced or consumed as part of the reaction with the analyte. Microscale Clark-type sensors have been created in silicon employing microlithographic manufacturing techniques.

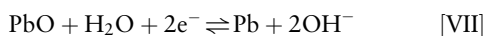
### Sensors for Atmospheric Oxygen

The Clark electrodes described above are not suitable for oxygen determination in dry gaseous samples such as air because the thin layer of electrolyte solution contained behind the membrane is prone to rapid drying. A different arrangement is therefore used for such applications. Amperometric gas sensors for oxygen (and sensors for other electroactive species in the vapor phase) usually consist of a porous PTFE membrane that bears a precious metal electrode deposited, also in porous form, directly on the backside. This keeps the diffusion length short while



allowing the inclusion of a larger volume of electrolyte solution (typically ~5–10 ml) and alleviating problems due to evaporation. An artificial additional diffusion barrier, such as a porous cover plate, is often needed in order to obtain a linear concentration dependence and to eliminate sensitivity to drafts in the sample gas.

If the reference electrode system is chosen such that the cell potential is sufficient to cause the sensing electrode reaction to proceed, the sensor acts as a galvanic cell and it is not necessary to apply an external voltage. It is possible to use, for example, Pb as anode with the following reference reaction:



The standard potential for these electrodes is  $-0.58\text{ V}$  and therefore the cell reaction occurs spontaneously. These cells are sometimes referred to as ‘Mackereth sensors’. Such galvanic sensors are also often termed ‘fuel cell sensors’ and it is possible to measure either the resulting current or the cell voltage. In the former case, the term amperometric sensor in its widest definition is still correct, although the cell is fundamentally different from the usual Faradaic amperometric systems. Note that for the determination of atmospheric oxygen an alternative has become available in recent years in the form of optical sensors based on fluorescence. These sensors are very robust as they do not contain electrodes or a liquid phase and show very fast response times.

One further important application of oxygen sensing is the direct analysis of automobile exhausts. For this purpose an amperometric sensor based on a solid

electrolyte phase can be used. Yttrium-stabilized zirconia ( $\text{ZrO}_2\text{--Y}_2\text{O}_3$ ) forms an electrolyte that is conductive for  $\text{O}^{2-}$  ions at temperatures above  $400^\circ\text{C}$ . Both the reduction and oxidation reactions can be catalyzed by porous platinum electrodes deposited on the zirconia:



If two such electrodes are separated by a thin layer of only zirconia, the application of a potential will lead to the pumping of oxygen from the cathode to the anode. This device can be used as an amperometric sensor for oxygen if a diffusion barrier restricts the flux of oxygen to the cathode. Note that similar devices are also often used as potentiometric sensors according to the Nernst equation (i.e., the lambda-probe in cars with catalytic converters). In this case one side of the cell has to act as a reference, e.g., by using ambient air.

See also: **Sensors:** Overview; Chemically Modified Electrodes.

## Further Reading

- Clark LC (1956) Monitor and control of blood and tissue oxygen tensions. *Transactions – American Society for Artificial Internal Organs* 2: 41–57.
- Göpel W, Hesse J, and Zemel JN (eds.) (1991) *Sensors, A Comprehensive Survey, Volume 2, Chemical and Biochemical Sensors*. Weinheim: VCH.
- Hitchman ML (1978) *Measurement of Dissolved Oxygen*. New York: Wiley.
- Janata J (1989) *Principles of Chemical Sensors*. New York: Plenum.

## Calorimetric/Enthalpimetric

**B Danielsson**, Lund University, Lund, Sweden

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 4597–4605, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor and O Wolfbeis.

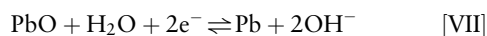
## Introduction

The most attractive feature of calorimetric sensors is the fact that they are based on an almost universal detection principle that is insensitive to the optical or electrochemical properties of the sample. The usefulness of calorimetry as an analytical tool in biology

has long been recognized, but it was not until immobilized enzymes had been introduced in bioanalysis that the development of simpler calorimetric sensors took place in the mid-1970s. The conventional (micro)calorimeters available before this time, although adequate for more precise analyses, were less suitable for routine analyses because of their high cost and time-consuming operation. Immobilized enzyme technology made it possible to construct simpler analytical devices by applying the enzymes in reactors or in layers close to the temperature transducer. Similar to the enzyme electrodes, calorimetric enzyme sensors found applications in many bioanalytical fields, such as clinical chemistry, environmental control, and bioprocess analysis. Enzymatic

allowing the inclusion of a larger volume of electrolyte solution (typically  $\sim 5\text{--}10\text{ ml}$ ) and alleviating problems due to evaporation. An artificial additional diffusion barrier, such as a porous cover plate, is often needed in order to obtain a linear concentration dependence and to eliminate sensitivity to drafts in the sample gas.

If the reference electrode system is chosen such that the cell potential is sufficient to cause the sensing electrode reaction to proceed, the sensor acts as a galvanic cell and it is not necessary to apply an external voltage. It is possible to use, for example, Pb as anode with the following reference reaction:



The standard potential for these electrodes is  $-0.58\text{ V}$  and therefore the cell reaction occurs spontaneously. These cells are sometimes referred to as 'Mackereth sensors'. Such galvanic sensors are also often termed 'fuel cell sensors' and it is possible to measure either the resulting current or the cell voltage. In the former case, the term amperometric sensor in its widest definition is still correct, although the cell is fundamentally different from the usual Faradaic amperometric systems. Note that for the determination of atmospheric oxygen an alternative has become available in recent years in the form of optical sensors based on fluorescence. These sensors are very robust as they do not contain electrodes or a liquid phase and show very fast response times.

One further important application of oxygen sensing is the direct analysis of automobile exhausts. For this purpose an amperometric sensor based on a solid

electrolyte phase can be used. Yttrium-stabilized zirconia ( $\text{ZrO}_2\text{--Y}_2\text{O}_3$ ) forms an electrolyte that is conductive for  $\text{O}^{2-}$  ions at temperatures above  $400^\circ\text{C}$ . Both the reduction and oxidation reactions can be catalyzed by porous platinum electrodes deposited on the zirconia:



If two such electrodes are separated by a thin layer of only zirconia, the application of a potential will lead to the pumping of oxygen from the cathode to the anode. This device can be used as an amperometric sensor for oxygen if a diffusion barrier restricts the flux of oxygen to the cathode. Note that similar devices are also often used as potentiometric sensors according to the Nernst equation (i.e., the lambda-probe in cars with catalytic converters). In this case one side of the cell has to act as a reference, e.g., by using ambient air.

See also: **Sensors:** Overview; Chemically Modified Electrodes.

## Further Reading

- Clark LC (1956) Monitor and control of blood and tissue oxygen tensions. *Transactions – American Society for Artificial Internal Organs* 2: 41–57.
- Göpel W, Hesse J, and Zemel JN (eds.) (1991) *Sensors, A Comprehensive Survey, Volume 2, Chemical and Biochemical Sensors*. Weinheim: VCH.
- Hitchman ML (1978) *Measurement of Dissolved Oxygen*. New York: Wiley.
- Janata J (1989) *Principles of Chemical Sensors*. New York: Plenum.

## Calorimetric/Enthalpimetric

**B Danielsson**, Lund University, Lund, Sweden

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 4597–4605, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor and O Wolfbeis.

## Introduction

The most attractive feature of calorimetric sensors is the fact that they are based on an almost universal detection principle that is insensitive to the optical or electrochemical properties of the sample. The usefulness of calorimetry as an analytical tool in biology

has long been recognized, but it was not until immobilized enzymes had been introduced in bioanalysis that the development of simpler calorimetric sensors took place in the mid-1970s. The conventional (micro)calorimeters available before this time, although adequate for more precise analyses, were less suitable for routine analyses because of their high cost and time-consuming operation. Immobilized enzyme technology made it possible to construct simpler analytical devices by applying the enzymes in reactors or in layers close to the temperature transducer. Similar to the enzyme electrodes, calorimetric enzyme sensors found applications in many bioanalytical fields, such as clinical chemistry, environmental control, and bioprocess analysis. Enzymatic

reactions are usually associated with rather high enthalpy changes, in the range of 20–100 kJ mol<sup>−1</sup>, which is often sufficient for designing assays based on only one enzymatic step. The combined heat production from sequential or cycling enzymatic reactions may also be utilized to increase the overall sensitivity. In addition, it is equally simple to combine calorimetric sensors with immobilized cells or enzymes or cells in flow streams.

Owing to their outstanding versatility in bioanalysis, this treatment of calorimetric/enthalpimetric sensors will focus on thermal bioanalyzers or biosensor systems. The most sensitive and practically useful calorimetric sensor arrangements involve continuous flow operation (viz., flow injection analysis), although several attempts have been made to create calorimetric enzyme probes simply by coating the temperature transducer with an enzyme layer. However, these measurements, which are carried out differentially using an inactive reference probe, suffer from poor sensitivity. Recent designs of miniaturized flow calorimetric sensors show that most of the sensitivity can be preserved concurrent with increased analysis speed and decreased equipment complexity. Such sensors may find use in portable monitoring, for instance, the home monitoring of glucose in diabetes care.

## Principles

The total heat evolved in a reaction is proportional to the molar enthalpy change:

$$Q = -n_p \Delta H \quad [1]$$

where  $Q$  is the total heat,  $n_p$  the moles product, and  $\Delta H$  the molar enthalpy change.

The total heat is also dependent on the heat capacity,  $C_s$ , of the system including the solvent:

$$Q = C_s \Delta T \quad [2]$$

Consequently, the measured temperature change,  $\Delta T$ , is dependent on the molar enthalpy change and the heat capacity:

$$\Delta T = -\frac{\Delta H n_p}{C_s} \quad [3]$$

Table 1 lists the molar enthalpy changes of some common enzyme-catalyzed reactions. Because enthalpimetric measurements are based on the sum of all enthalpy changes in the reaction mixture, it is common to coimmobilize oxidases with catalase, which approximately doubles the sensitivity at the same time as the oxygen consumption is reduced and

**Table 1** Molar enthalpy changes for some enzyme-catalyzed reactions

Enzyme	Substrate	$-\Delta H$ (kJ mol <sup>−1</sup> )
Catalase	Hydrogen peroxide	100
Cholesterol oxidase	Cholesterol	53
Glucose oxidase	Glucose	80
Hexokinase	Glucose	28 (75) <sup>a</sup>
Lactate dehydrogenase	Sodium pyruvate	62
NADH dehydrogenase	NADH	225
$\beta$ -Lactamase	Penicillin G	67 (115) <sup>a</sup>
Trypsin	Benzoyl-L-arginineamide	29
Urease	Urea (phosphate buffer, pH 7.5)	61
Uricase	Urate	49

<sup>a</sup>The  $\Delta H$  values in parentheses were obtained in TRIS buffer (protonation enthalpy  $-47.5$  kJ mol<sup>−1</sup>).

Reprinted from Blum LJ and Coulet PR (eds.) (1991) *Biosensor Principles and Applications*. New York: Dekker, by courtesy of Marcel Dekker, Inc.

deleterious effects of hydrogen peroxide are minimized. The high protonation enthalpy of buffer ions, such as tris(hydroxymethyl)aminomethane (TRIS), can be utilized to increase the total enthalpy change of proton-producing reactions (Table 1). Extreme increases in detection limit (to  $<10$  nmol l<sup>−1</sup>) can be obtained in substrate or coenzyme recycling enzyme systems, where the net enthalpy change of each turn of the cycle adds to the overall enthalpy change. An enthalpy change of 100 kJ mol<sup>−1</sup> is sufficient for substrate determinations with a detection limit better than 0.005 mmol l<sup>−1</sup>.

In enthalpimetric flow injection analysis, the sample pulses employed are too small to give a thermal steady state, but result in a temperature peak proportional to the enthalpy change and thereby to the substrate concentration. The peak height, the area under the peak, and the ascending slope of the peak are all linearly related to the substrate concentration. Sample introduction of a sufficient duration (several minutes) leads to a thermal steady state resulting in a constant temperature deviation, which is also proportional to the substrate concentration.

An inherent disadvantage of calorimetry is the lack of specificity: all enthalpy changes in the reaction mixture contribute to the measurement. It is therefore important to avoid nonspecific enthalpy changes from dilution or solvation effects. However, in the majority of cases this is not a serious problem. Differential measurement using a reference column containing an inactive filling is an efficient way of coping with nonspecific and matrix effects.

## Sensor Design

Thermistors, metal oxide resistors with a very high negative temperature coefficient of resistance, are normally used as temperature transducers in enthalpimetric sensors. They can be obtained in many different configurations, sizes (down to 0.1–0.3 mm beads or deposited as thin films), and resistance values. For narrow temperature ranges, the resistance–temperature relationship can be approximated by the following equation:

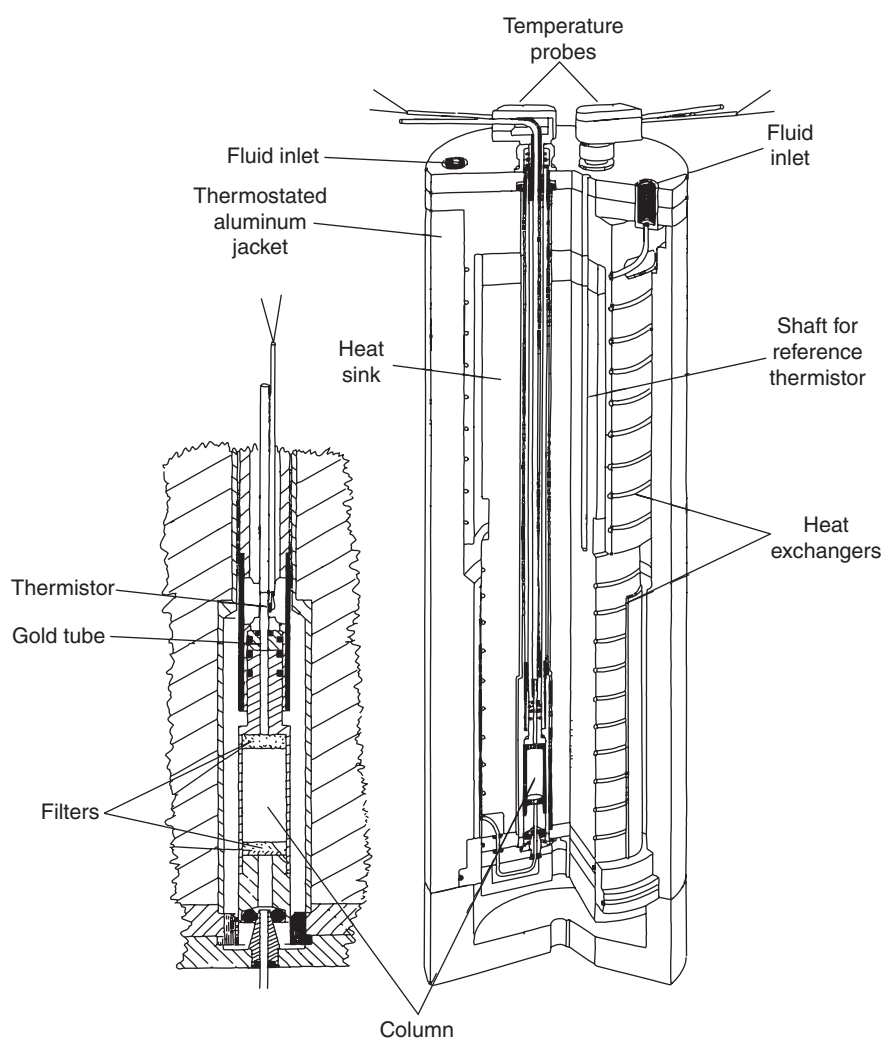
$$R_T = R_{T_0} e^{\beta(1/T - 1/T_0)} \quad [4]$$

where  $R_T$  and  $R_{T_0}$  are the zero-power resistances at the absolute temperatures  $T$  and  $T_0$ , respectively.  $\beta$  is a material constant that ranges between 4000 and

5000 K for most thermistor materials. This results in a temperature coefficient of resistance between  $-3\% ^\circ\text{C}^{-1}$  and  $-5.7\% ^\circ\text{C}^{-1}$ .

Other temperature transducers employed in calorimetric analyzers include Peltier elements, Darlington transistors, and thermopiles. Thermopiles can be made in very small dimensions using current semiconductor technology and are less flow-sensitive than thermistors, though the latter provide higher temperature resolution.

An example of the instrumentation used in the author's laboratory is shown in Figure 1. The calorimeter part of this two-channel instrument consists of an outer aluminum cylinder ( $80 \times 250$  mm) thermostatted to  $25^\circ\text{C}$ ,  $30^\circ\text{C}$ , or  $37^\circ\text{C}$  and a second inner aluminum cylinder with two column ports and a pocket for a reference thermistor. The solution to be



**Figure 1** Cross-section of the calorimeter of an enzyme thermistor with an aluminum constant temperature jacket. The two identical column ports (one temperature probe with a column is shown) can be used independently or one of the ports can be used for a reference channel. The column attachment is illustrated on the left. (Reprinted with permission from Danielsson B (1990) *Journal of Biotechnology* 15: 187–200; © Elsevier.)

analyzed is introduced via a thin-walled acid-proof steel tubing (typically 0.8 mm ID). Two-thirds of the tubing's length is in close contact with the thermostatted jacket, while the last third is in close contact with the inner cylinder, which acts as a heat sink. This arrangement results in exceedingly small temperature variations (short term) in the column. The column is inserted by a long Delrin tube containing the temperature sensor. The DC-type Wheatstone bridge is equipped with a chopper-stabilized low-drift operational amplifier and has maximum sensitivity of 100 mV per 0.001 °C. Temperatures in the range of 0.01–0.05 °C are commonly measured for enzymatic reactions.

## Performance

The equipment shown in **Figure 1** is operated with a continuous flow of a buffer (0.5–5 ml min<sup>−1</sup>) using sample volumes of 0.01–0.5 ml in flow injection mode. Small sample volumes (<25 µl) may be injected undiluted, while larger samples should be diluted at least 5- to 10-fold in buffer so as to avoid nonspecific thermal effects due to mixing.

Columns up to 1 ml (at 7 mm ID) can be used. The enzyme support should have good binding capacity and good mechanical, chemical, as well as microbial stability and involve relatively simple coupling procedures. These demands are satisfied by controlled pore glass (CPG), which has been the most common enzyme support in our investigations (80–120 mesh, pore size 500–2000 Å). Cells can be immobilized by entrapment in alginate or gelatin beads. Where possible a large excess of enzyme is applied; 100 units or

more. The major factor limiting column life is mechanical obstruction. If the solutions and samples are filtered through a 1–5 µm filter and assuming microbial growth is prevented, good operational stability and reproducibility for a large series of samples (thousands) can be achieved. The column lifetime can be several months of continuous use.

**Table 2** shows a number of substances for which enthalpimetric assays have been carried out. The linear ranges given may not in all cases represent what is ultimately achievable under optimal conditions. The data are from various studies from different laboratories and with different equipment. For an enthalpy change of ~200 kJ mol<sup>−1</sup>, such as for glucose oxidase combined with catalase, it is possible to come down to a limit of detection of 1 µmol l<sup>−1</sup> with the equipment described above. At the same time, the linear range may extend to several hundred millimoles per liter with a highly active column in reactions that are not limited by the availability of coreactants. The use of recycling techniques may result in several thousand times greater sensitivity. The precision of the measurement can be as good as 0.5% (relative standard deviation).

## Major Applications

### Metabolite Assays

A large number of enthalpimetric assays based on immobilized enzyme reactors have been proposed with potential interest to biotechnology, clinical chemistry, and food analysis. Several of these have been applied in bioprocess monitoring and recently bedside monitoring of glucose and urea have been

**Table 2** Linear concentration ranges of substances measured with enthalpimetric sensors using immobilized enzymes

Analyte	Enzyme(s) used	Concentration range (mmol l <sup>−1</sup> )
Ascorbic acid	Ascorbate oxidase	0.01–0.6
ATP (or ADP)	Pyruvate kinase + hexokinase	10 nmol l <sup>−1</sup> <sup>a</sup>
Cellobiose	β-Glucosidase + glucose oxidase/catalase	0.05–5
Cephalosporins	Cephalosporinase (β-lactamase)	0.005–10
Creatinine	Creatinine iminohydrolase	0.01–10
Ethanol	Alcohol oxidase	0.005–1
Glucose	Hexokinase	0.5–25
Glucose	Glucose oxidase/catalase	0.001–0.8 (75) <sup>b</sup>
L-Lactate	Lactate-2-mono-oxygenase	0.005–2
L-Lactate	Lactate oxidase/catalase	0.002–1
L-Lactate (or pyruvate)	Lactate oxidase/catalase + lactate dehydrogenase	10 nmol l <sup>−1</sup> <sup>a</sup>
Oxalate	Oxalate oxidase	0.005–0.5
Penicillin	β-Lactamase	0.005–200
Sucrose	Invertase	0.05–100
Urea	Urease	0.005–200

<sup>a</sup>With substrate recycling.

<sup>b</sup>With benzoquinone as electron acceptor.

demonstrated. To illustrate the possibilities and limitations of the technique a few assays will be discussed in some detail. The concentration ranges given have in general been obtained with 0.5 ml sample volume at a flow rate of  $1 \text{ ml min}^{-1}$ . The sensitivity can be adapted to higher concentrations by dilution and use of smaller sample volume.

**Alcohols** Lower alcohols have been measured with alcohol oxidase (EC 1.1.3.13) from *Candida boidinii* or *Pichia pastoris*. The latter is available with a higher specific activity and has a somewhat different substrate specificity. Coimmobilization with catalase increases the stability of the enzyme column to several months with an operating range of  $0.005\text{--}1 \text{ mmol l}^{-1}$  (0.5 ml samples) using  $0.1 \text{ mol l}^{-1}$  sodium phosphate, pH 7.0, as the buffer. This assay is useful for the determination of ethanol in samples from beverages, blood, and for monitoring fermentation.

**Cellobiose** Since the heat produced by the hydrolysis of cellobiose with  $\beta$ -glucosidase is too low to give acceptable sensitivity, a procedure has been designed in which the glucose formed in a precolumn containing  $\beta$ -glucosidase is measured with a glucose oxidase/catalase loaded enzyme thermistor. The operating range is  $\sim 0.05\text{--}5 \text{ mmol l}^{-1}$ .

**Cholesterol and cholesterol esters** Cholesterol has been determined in  $0.16 \text{ mol l}^{-1}$  phosphate buffer, pH 6.5, containing 12% (v/v) ethanol and 8% (v/v) Triton X-100 using cholesterol oxidase (EC 1.1.3.6) from *Nocardia erythropolis*. Cholesterol esters were measured after treating the samples with cholesterol esterase (EC 3.1.1.13). The measuring range is adequate for clinical use).

**Creatinine** The microbial enzyme creatinine deiminase (EC 3.5.4.21) works well in the enzyme thermistor giving a linear range from 0.01 to at least  $10 \text{ mmol l}^{-1}$ . This could be a very powerful assay for clinical use since it requires only one enzyme and is independent of endogenous ammonia. Unfortunately, the determination of creatinine in human serum requires a higher order of magnitude increase in sensitivity than is available with the present instrumentation.

**Glucose** Very useful assays for various bioanalytical applications have been developed based on calorimetry, usually with glucose oxidase coimmobilized with catalase. The addition of catalase more than doubles the total enthalpy change, eliminates the deleterious effects of the hydrogen peroxide formed in the glucose oxidase reaction, and restores half of

the oxygen consumed by this reaction. This procedure provides high sensitivity and specificity and it has no cofactor requirement. The enzyme columns are very stable. The most serious disadvantage is that the useful concentration range is limited to  $<0.7\text{--}1.0 \text{ mmol l}^{-1}$  by the supply of oxygen, which is soluble only to a small extent in water,  $\sim 0.25 \text{ mmol l}^{-1}$  at  $25^\circ\text{C}$ . A more soluble electron acceptor than oxygen, such as benzoquinone, increases the useful working range considerably to at least  $75 \text{ mmol l}^{-1}$ . Alternatively, the enzyme hexokinase can be used. The requirement for the cofactor adenosine 5'-triphosphate (ATP) is a drawback; however, a linear range of up to  $25 \text{ mmol l}^{-1}$  can be obtained.

**L-Lactate** L-Lactate can be determined with two different enzyme systems: the lactate-2-mono-oxygenase (EC 1.13.12.4) from *Mycobacterium smegmatis* providing a linear range of  $0.005\text{--}2 \text{ mmol l}^{-1}$  in  $0.2 \text{ mol l}^{-1}$  sodium phosphate buffer, pH 7.0, or the lactate oxidase from *Pediococcus pseudomonas* (EC 1.1.3.2) together with catalase giving a linear range of  $0.002\text{--}1 \text{ mmol l}^{-1}$  in  $0.1 \text{ mol l}^{-1}$  phosphate buffer, pH 7.0.

**Oxalate** Oxalate can be determined with oxalate oxidase (EC 1.2.3.4) from barley seedlings. The linear concentration range is  $0.005\text{--}0.5 \text{ mmol l}^{-1}$  in  $0.1 \text{ mol l}^{-1}$  sodium citrate buffer, pH 3.5, containing  $2 \text{ mmol l}^{-1}$  ethylenediaminetetraacetic acid (EDTA) and  $0.8 \text{ mmol l}^{-1}$  8-quinolinol. This assay has been found suitable for the determination of oxalate in urine, beverages, and food samples. Urine samples are preferably diluted 10-fold and passed through a  $\text{C}_{18}$ -cartridge to remove interfering substances.

**Penicillin G and V** Very useful procedures for the assay of  $\beta$ -lactams, such as penicillin G and V, have been designed using  $\beta$ -lactamases, such as penicillinase type I from *Bacillus cereus* (EC 3.5.2.6). The useful linear range is  $\sim 0.005\text{--}200 \text{ mmol l}^{-1}$ . Several industrial applications have been developed using both discrete samples and continuous monitoring on pilot-plant and production-scale fermentors. Alternatively, the penicillin amidase (EC 3.5.1.11) has been used which shows better specificity in fermentation broths. The sensitivity is, however, lower although sufficient for many purposes. In both cases, the enzyme columns are very stable and can be used for several months or for thousands of samples.

**Sucrose** In contrast to most other disaccharide splitting enzymes, invertase (EC 3.2.1.26) produces enough heat to allow direct determinations of sucrose in the range of  $0.05\text{--}100 \text{ mmol l}^{-1}$ , even in



the presence of glucose. Invertase columns are very stable and useful in food and bioprocess analysis.

**Triglycerides** Practical routine methods for the determination of the main blood lipid classes, cholesterol and cholesterol esters, phospholipids, and triglycerides have been proposed. Triglycerides have been determined with lipoprotein lipase (EC 3.1.1.34) immobilized on CPG with a pore size of 200 nm. The assay buffer was  $0.1 \text{ mol l}^{-1}$  TRIS buffer, pH 8.0, containing 0.5% Triton X-100. The linear response was  $0.05\text{--}10 \text{ mmol l}^{-1}$  for tributyrin and  $0.1\text{--}5 \text{ mmol l}^{-1}$  for triolein.

**Urea** Another common metabolite that has attracted a great deal of interest is biosensor analysis of urea. Urease, which gives a linear range of at least  $0.01\text{--}200 \text{ mmol l}^{-1}$ , offers a clinically useful assay that is independent of the ammonium concentration in the sample, the same as with creatinine. Urease is very sensitive to inhibition by heavy metals. This fact has even been exploited in the design of a reversible procedure for heavy metal determination. Addition of  $1 \text{ mmol l}^{-1}$  EDTA and  $1 \text{ mmol l}^{-1}$  reduced glutathione to the buffer protects the urease leading to a very stable enzyme column.

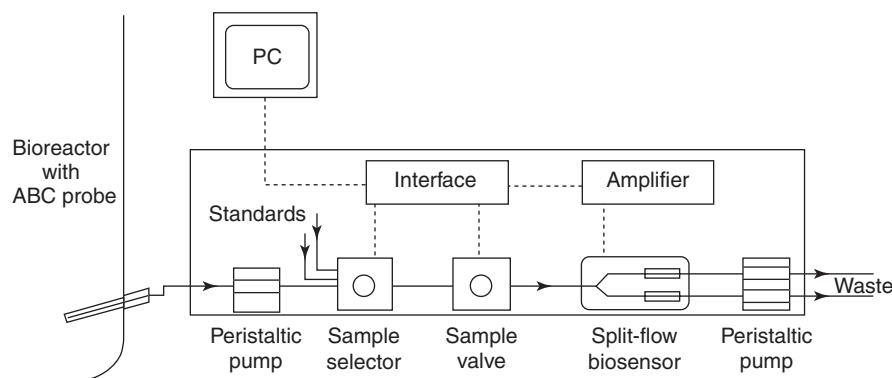
Other metabolites that have been measured with calorimetric sensors include ascorbic acid, ATP/ADP (adenosine 5'-diphosphate), cephalosporins, galactose, hydrogen peroxide, lactose, malate, phospholipids, uric acid, xanthine, and hypoxanthine.

### Process Control

Monitoring and control of (bio)processes was identified early on as a suitable application field for biosensors, but only recently have really practical

systems been introduced. Owing to their outstanding versatility, calorimetric sensors are particularly suitable for this purpose. Since a biosensor cannot usually be placed directly in the sample, for instance, inside a fermentor that needs to be sterilized before use, a sample stream has to be drawn from the process solution. This must be done under strictly sterile conditions to avoid microbial contamination of the bioprocess and to remove cells from the sample solution in order to prevent cells from attaching inside the apparatus and to stop changes in the composition of the sample during its transport. Preferably this is done by filtration or dialysis. There has been a lack of suitable devices for sample filtration, but recently various sterilizable, small cross-flow (tangential flow) filtration units have been introduced that work well, at least with larger fermentors where the spent sample stream can be discarded. A filtration probe that consists of a large (8 mm diameter)  $0.2 \mu\text{m}$  hollow-fiber filter in a holder that is inserted into the fermentor through a standard port and that can be steam sterilized *in situ* has been found to perform very well with larger as well as with smaller fermentors.

Such a filtration probe has been tested with very good results in a number of analyses with a special enzyme thermistor version designed for use in industrial environments. **Figure 2** is a schematic illustration of an enzyme thermistor arranged for process monitoring by repeated flow injection analysis. The interval between sample injections is chosen with respect to how fast the analyte concentration changes. A typical figure is four to five sample injections per hour. The injection valve and a sample selector valve for selection of different samples or calibration solutions and the pumps are controlled with a personal computer with a 386 processor, which also



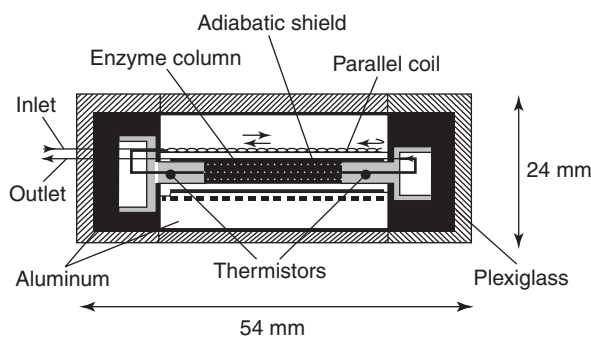
**Figure 2** Setup for bioprocess monitoring. A continuous sample stream ( $1\text{--}3 \text{ ml min}^{-1}$ ) is drawn from the bioreactor using a filtration probe (ABC probe) and a peristaltic pump. The signal from the split-flow enzyme thermistor is recorded with a computer via a 12-bit AD-converter and an interface that also connects with the valves and pumps. (Reprinted with permission from Rank M, Gram J, and Danielsson B (1993) Industrial on-line monitoring of penicillin V, glucose and ethanol using a split-flow modified thermal biosensor. *Analytica Chimica Acta* 281: 521–526; © Elsevier.)

calculates the analyte concentration from the enzyme thermistor signal and stores and displays the data. The equipment is placed close to the fermentor in a climatized cabinet flushed with cool, dry air. In order to minimize nonspecific heat signals, a split-flow enzyme thermistor equipped with an inactive reference column is generally used. Measurements of various metabolites, including glucose, lactate, ethanol, glycerol, acetaldehyde, and penicillin, have been very successful on fermentors ranging in size from 0.5 to 160 m<sup>3</sup>. Fermentations lasting for 2 weeks can usually be followed without change of enzyme column.

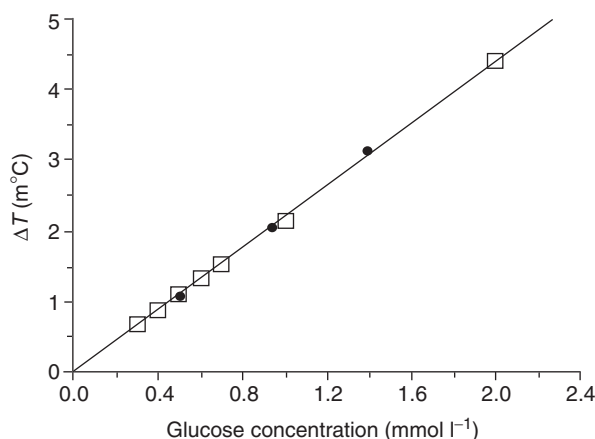
### Miniaturized Calorimetric Sensors

Recently, several laboratories have started investigations on miniaturized calorimetric sensors. One type of device involves scaled-down versions of the more conventional flow calorimeter designs, and another is based on constructions made by micromachining in materials such as silicon, quartz, and glass. These approaches should lead to rapid, versatile, and inexpensive enthalpimetric analyses requiring very small samples and minimal reagent/enzyme consumption. Several sensors have been constructed primarily for glucose measurements directed toward diabetes care. One thermal biosensor probe, for instance, was designed as an integrated-circuit biocalorimetric sensor with total dimensions of only 1 × 1 × 0.3 mm with Darlington transistors as temperature-sensitive elements. Another design is a small thermoelectric glucose sensor that employs a thin-film thermopile to measure the evolved heat. This device was reported to be sufficiently less affected by external thermal effects than calorimetric sensors based on thermistors that it could be operated without environmental temperature control. Work is in progress in the author's laboratory to construct miniaturized biothermal flow injection systems suitable for portable monitoring. Instruments with 0.1–0.2 mm (ID) flow channels and a flow rate of 25–100 µl min<sup>-1</sup> with sample volumes of 1–10 µl are presently being evaluated with emphasis on whole blood determination of glucose. **Figure 3** shows one version of a miniature calorimeter.

Operation with whole blood samples puts special demands on the enzyme column. The investigations have followed two main routes, the use of superporous agarose beads with pores large enough for cells to pass through and the use of spherical, uniform beads large enough to let the cells pass unobstructed outside the beads. Both approaches have led to acceptable results; for instance, columns that can be used with at least 100 blood samples with satisfying precision. An example of a calibration graph for a



**Figure 3** Cross-section of a miniaturized calorimetric immobilized enzyme sensor equipped with a 1.5 × 15 mm column.



**Figure 4** Calibration graph for a calorimetric device of the type shown in **Figure 3** with a glucose oxidase/catalase column. The sample volume was 20 µl. Open squares are for aqueous glucose standards and closed squares are for 10-fold diluted blood samples spiked with glucose.

superporous agarose column with glucose oxidase and catalase used in a device of the type depicted in **Figure 3** is shown in **Figure 4**.

### Other Applications

The outstanding versatility of calorimetric sensors makes them useful in many other modes of operation than those described above. Some of these are briefly discussed below.

#### Environmental Control

The calorimetric detection principle has been exploited mainly in two different ways in environmental analysis. The first measures the heat of conversion of a pollutant by a cell metabolic route or a specific enzyme directly.

The second alternative, which is much more sensitive, measures the inhibitory effect of a pollutant on

an enzyme or on the metabolism of suitable cells. To detect an environmental pollutant by quantifying its effect on a biological system is ideal. The highly sensitive determination of heavy metals ( $\text{Hg}^{2+}$ ,  $\text{Cu}^{2+}$ , and  $\text{Ag}^+$ ) based on the inhibition of urease has already been mentioned. The detection limit is comparable to that of atomic absorption spectrometry or stripping voltammetry ( $\text{ng g}^{-1}$  levels) and the measurement involves comparatively simple equipment and procedures. The assay is carried out on an aqueous solution and does not create any hazardous wastes.

Another example is the determination of cholinesterase inhibitors (pesticides) using acetylcholine or butyrylcholine as substrate. Since the inhibition of the enzyme in this case is irreversible, it is convenient to use a reversible enzyme immobilization technique that allows for simple replacement of spent enzyme. This can be accomplished by reversibly binding the cholinesterase to a concanavalin A-agarose column. Many enzymes require metal ions in their active sites for catalytic activity. The metal ion can be removed using strong chelating agents resulting in an inactive apoenzyme that can be reactivated upon exposure to a sample containing the appropriate metal ion.

Similarly, the use of immobilized cell columns in calorimetric studies of metabolic effectors has great potential due to its general applicability. The metabolic heat is very likely to be affected by the presence of a metabolic poison, such as an environmental pollutant, or substances with an activating effect, such as vitamins. A wide variety of cell immobilization techniques are available, including entrapment with alginate in  $\text{Ca}^{2+}$  solution and entrapment in gelatin or agarose beads. It is important to maintain a sterile environment and to make the beads sufficiently small, at least  $<1\text{--}2\text{ mm}$ , to provide a short response time. Immobilized whole cells may offer several advantages over pure enzyme systems in metabolite assays especially where the pure enzyme is unstable or difficult to purify. This may be the case with reactions involving sequential multienzyme systems. Furthermore, cells provide built-in systems for coenzyme regeneration.

#### **Thermometric Enzyme-linked Immunosorbent Assay**

Rapid, sensitive determination of larger molecules, such as hormones and antibodies present, for instance, in fermentation broth, can be accomplished by an automated, flow-through thermometric enzyme-linked immunosorbent assay (TELISA), which is a competitive assay. The unlabeled antigen in the sample (or standard) is mixed with a fixed amount of

enzyme-labeled antigen and the mixture is then applied to the immunosorbent column mounted in the enzyme thermistor apparatus. The more unlabeled antigen in the sample, the less enzyme-labeled antigen will be bound to the column. The amount of enzyme (e.g., peroxidase or catalase) bound to the column is determined by injecting a pulse of substrate in excess. Finally, bound antigen is removed from the immunosorbent by washing with  $0.2\text{ mol l}^{-1}$  glycine-hydrochloric acid, pH 2.2, thereby regenerating the column for the next assay. One assay cycle is completed in 10–15 min. This method is very suitable for monitoring fermentations or when rapid results are desired for a limited number of samples. The sensitivity can be increased with use of recycling enzyme systems.

#### **Determination of Soluble Enzymes**

The enzyme thermistor can be used for the determination of the activity of soluble enzymes, by modifying the flow system and using an empty or inactive column, preferably made of polytetrafluoroethene, as a reaction chamber. The sample solution and substrate solution are passed through heat exchangers and thoroughly mixed. The mixture is then passed through a short heat exchanger to eliminate mixing and solvation heats. Enzyme activities down to  $0.01\text{ U ml}^{-1}$  can thus be determined. The technique is general and can be used with both sample pulses and with continuous sample introduction. It should be of particular interest for monitoring enzyme purification processes.

#### **Measurements on Cell Suspensions**

In a similar manner, the total metabolic heat generated by cells can be registered by mixing a flow stream containing the cells with a substrate stream or simply by letting a flow with cells, for instance, from a fermentor, pass through an empty enzyme thermistor column. This technique should be useful in studies of metabolic, effectors, and in toxicological studies and as a general method of following the overall metabolic state of a fermentation. The heat or the temperature level recorded represents the total metabolism of the microorganisms, and changes in the metabolism are directly detectable as changes in the temperature level.

#### **Measurements in Organic Solvents**

Enthalpimetric measurements in organic solvents should be up to three times more sensitive than in water because organic solvents have a lower heat capacity than water. This and the fact that enzymes can be active in organic solvents, although often

with changed specificity – or even reversed direction of the reaction – make it interesting to directly measure reactants in processes which are carried out in organic solvents. Furthermore, it would be of interest to develop assays in organic solvents for substrates with poor solubility in aqueous solutions. As an example, triglycerides can be determined directly in cyclohexane. Since the calorimetric measuring principle is as useful with endothermic reactions as with exothermic ones, it has been demonstrated that the enzyme thermistor can follow synthetic as well as hydrolytic reactions catalyzed by  $\alpha$ -chymotrypsin.

See also: **Blood and Plasma. Clinical Analysis:** Glucose. **Enzymes:** Immobilized Enzymes; Enzyme-Based Electrodes; Enzymes in Physiological Samples; Industrial Products and Processes. **Ethanol. Flow Injection Analysis:** Principles; Detection Techniques. **Food and Nutritional Analysis:** Alcoholic Beverages. **Forensic Sciences:** Alcohol in Body Fluids. **Immunoassays, Techniques:** Enzyme Immunoassays. **Lipids:** Determination in Biological Fluids. **Pesticides. Process Analysis:** Bioprocess Analysis.

## Further Reading

- Blum LJ and Coulet PR (eds.) (1991) *Biosensor Principles and Applications*. New York: Dekker.
- Hagleitner C, Hierlemann A, Lange D, *et al.* (2001) Smart single-chip gas sensor microsystem. *Nature (London)* 414(6861): 293–296.
- Johannessen EA, Weaver JMR, Bourova L, *et al.* (2002) Micromachined nanocalorimetric sensor for ultra-low-volume cell-based assays. *Analytical Chemistry* 74(9): 2190–2197.
- Kohler JM, Zieren M, and Willnauer R (1999) Chip sensors for reaction, scanning, and ac-calorimetry in stationary and moving samples using thermoelectric transduction. *Proceedings of the Electrochemical Society* 99(23): 429–439.
- Ramanathan K and Danielsson B (2001) Thermal biosensors. *Biosensors and Bioelectronics* 16: 417–423.
- Schneider T, Politi MJ, and Baptista MS (2002) A smart-laser-based calorimeter. *Journal of Chemical Education* 79(4): 503–505.
- Xie B, Ramanathan K, and Danielsson B (2001) Mini/micro thermal biosensors and other related devices for biochemical/clinical monitoring. *Trends in Analytical Chemistry* 19: 340–349.

## Chemically Modified Electrodes

**S Dong and Z Peng**, Chinese Academy of Science, Changchun, Jilin, People's Republic of China

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The concept of controlling the molecular architecture at the electrode/electrolyte interface was not known until ~30 years ago. Electrodes that can be prepared by deliberately immobilizing selected chemicals to display the desired electrochemical and other properties of the immobilized species are called chemically modified electrodes (CMEs). Since the emergence of CMEs, the field of electrochemistry has moved from traditional studies confined to inert electrode materials (such as bare C, Au, Hg, and Pt) at the interface.

With respect to enhanced sensitivity and selectivity in electroanalysis, the study of CMEs has been the most popular research area with studies ongoing in many electrochemistry laboratories. There is now a large body of research articles published in this field. This article outlines the latest achievements in CMEs, emphasizing applications in analytical chemistry. Based on the progress of modification, this article is

divided into four parts: (1) self-assembled monolayers (SAMs) on gold electrode and three-dimensional (3D) assembly; (2) mono- and multilayer assemblies on carbon electrode; (3) polymer film electrode; and (4) nanoparticle modified electrode.

### SAMs on Gold Electrode and 3D Assembly

The earliest molecular monolayer was made by sorption and covalent bonding of chemicals to electrode surfaces; however, limitations exist in the ability to control the structural and dynamic aspects of the immediate microscopic environment of the immobilized chemicals. SAMs that form spontaneously at solid/liquid interfaces have attracted much attention lately as it provides a unique technique for controlling the properties of solid surfaces. SAMs on electrodes give the platform needed for probing the relationship between the microstructure on the electrode surface and macroscopic electrochemical responses, which reach the top level of the molecular monolayer CMEs.

The progress made with respect to sensing interfaces is mainly due to the discovery and development of SAMs of sulfuric compounds on gold and other

with changed specificity – or even reversed direction of the reaction – make it interesting to directly measure reactants in processes which are carried out in organic solvents. Furthermore, it would be of interest to develop assays in organic solvents for substrates with poor solubility in aqueous solutions. As an example, triglycerides can be determined directly in cyclohexane. Since the calorimetric measuring principle is as useful with endothermic reactions as with exothermic ones, it has been demonstrated that the enzyme thermistor can follow synthetic as well as hydrolytic reactions catalyzed by  $\alpha$ -chymotrypsin.

**See also:** **Blood and Plasma. Clinical Analysis:** Glucose. **Enzymes:** Immobilized Enzymes; Enzyme-Based Electrodes; Enzymes in Physiological Samples; Industrial Products and Processes. **Ethanol. Flow Injection Analysis:** Principles; Detection Techniques. **Food and Nutritional Analysis:** Alcoholic Beverages. **Forensic Sciences:** Alcohol in Body Fluids. **Immunoassays, Techniques:** Enzyme Immunoassays. **Lipids:** Determination in Biological Fluids. **Pesticides. Process Analysis:** Bioprocess Analysis.

## Further Reading

- Blum LJ and Coulet PR (eds.) (1991) *Biosensor Principles and Applications*. New York: Dekker.
- Hagleitner C, Hierlemann A, Lange D, *et al.* (2001) Smart single-chip gas sensor microsystem. *Nature (London)* 414(6861): 293–296.
- Johannessen EA, Weaver JMR, Bourova L, *et al.* (2002) Micromachined nanocalorimetric sensor for ultra-low-volume cell-based assays. *Analytical Chemistry* 74(9): 2190–2197.
- Kohler JM, Zieren M, and Willnauer R (1999) Chip sensors for reaction, scanning, and ac-calorimetry in stationary and moving samples using thermoelectric transduction. *Proceedings of the Electrochemical Society* 99(23): 429–439.
- Ramanathan K and Danielsson B (2001) Thermal biosensors. *Biosensors and Bioelectronics* 16: 417–423.
- Schneider T, Politi MJ, and Baptista MS (2002) A smart-laser-based calorimeter. *Journal of Chemical Education* 79(4): 503–505.
- Xie B, Ramanathan K, and Danielsson B (2001) Mini/micro thermal biosensors and other related devices for biochemical/clinical monitoring. *Trends in Analytical Chemistry* 19: 340–349.

## Chemically Modified Electrodes

**S Dong and Z Peng**, Chinese Academy of Science, Changchun, Jilin, People's Republic of China

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The concept of controlling the molecular architecture at the electrode/electrolyte interface was not known until ~30 years ago. Electrodes that can be prepared by deliberately immobilizing selected chemicals to display the desired electrochemical and other properties of the immobilized species are called chemically modified electrodes (CMEs). Since the emergence of CMEs, the field of electrochemistry has moved from traditional studies confined to inert electrode materials (such as bare C, Au, Hg, and Pt) at the interface.

With respect to enhanced sensitivity and selectivity in electroanalysis, the study of CMEs has been the most popular research area with studies ongoing in many electrochemistry laboratories. There is now a large body of research articles published in this field. This article outlines the latest achievements in CMEs, emphasizing applications in analytical chemistry. Based on the progress of modification, this article is

divided into four parts: (1) self-assembled monolayers (SAMs) on gold electrode and three-dimensional (3D) assembly; (2) mono- and multilayer assemblies on carbon electrode; (3) polymer film electrode; and (4) nanoparticle modified electrode.

### SAMs on Gold Electrode and 3D Assembly

The earliest molecular monolayer was made by sorption and covalent bonding of chemicals to electrode surfaces; however, limitations exist in the ability to control the structural and dynamic aspects of the immediate microscopic environment of the immobilized chemicals. SAMs that form spontaneously at solid/liquid interfaces have attracted much attention lately as it provides a unique technique for controlling the properties of solid surfaces. SAMs on electrodes give the platform needed for probing the relationship between the microstructure on the electrode surface and macroscopic electrochemical responses, which reach the top level of the molecular monolayer CMEs.

The progress made with respect to sensing interfaces is mainly due to the discovery and development of SAMs of sulfuric compounds on gold and other

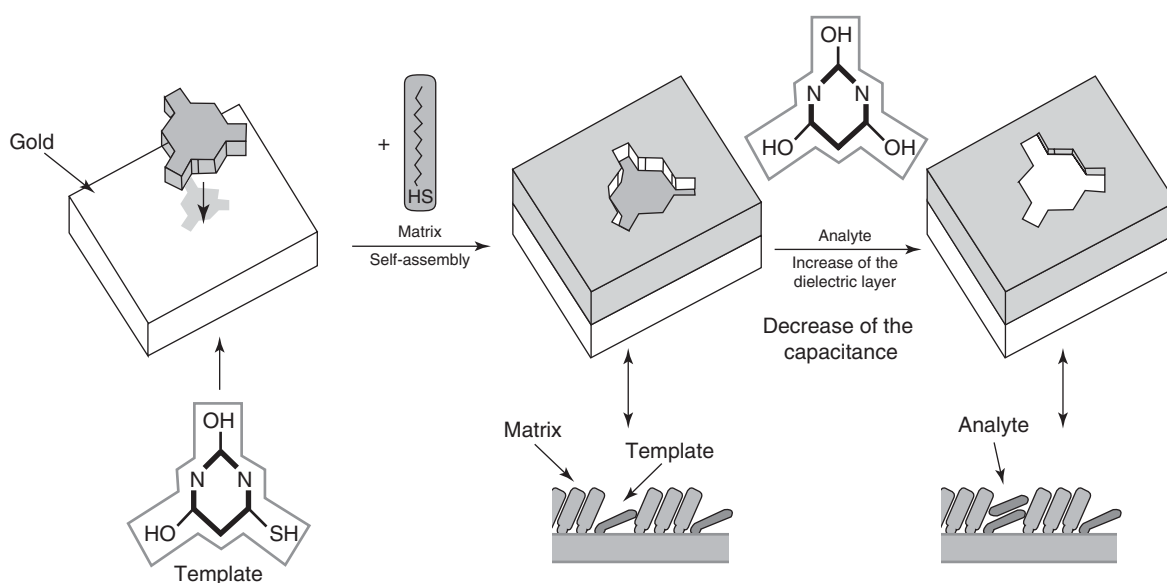
conductive solid surfaces. As a typical example, alkanethiols chemisorb on gold surface forming a gold–thiol bond to produce a densely packed, highly ordered monolayer into two dimensions (2D). The strength of the head group–substrate interactions, the lateral interactions, and the density of packing result in sufficient stability, and the monolayer is robust enough to endure prolonged electrochemical measurement. The preparation of SAMs is quite easy and simple.

The idea of molecular defects or ‘gates’ within a blocking SAM was introduced for sensor applications while preparing SAMs wherein the density and the size of the ‘gates’ can be controlled. The species present in the electrolyte solution can respond selectively by the mixed SAMs prepared by a long chain alkanethiol and a short thiol to form phase-separated domains. Here, the domains with short thiols perform the role of ‘gates’ for the selective response of the species studied. The dendrimer has been used as molecular ‘gate’ to control the electrochemical behavior of the redox species present in the electrolyte when co-assembled with long chain alkanethiols. The Langmuir Blodgett (LB) technique has also been used to prepare such molecular ‘gates’ for the species studied with the deposition of long chain alkanethiol. The density and the size of the species can be controlled rationally and thus signals appear that are related to the isolated species from molecular ‘gates’. The introduction of the molecular ‘gates’ into SAMs provide an avenue for the in-depth study of analytical applications.

Molecular imprinting suggests a general method to tailor specific recognition sites in polymers or

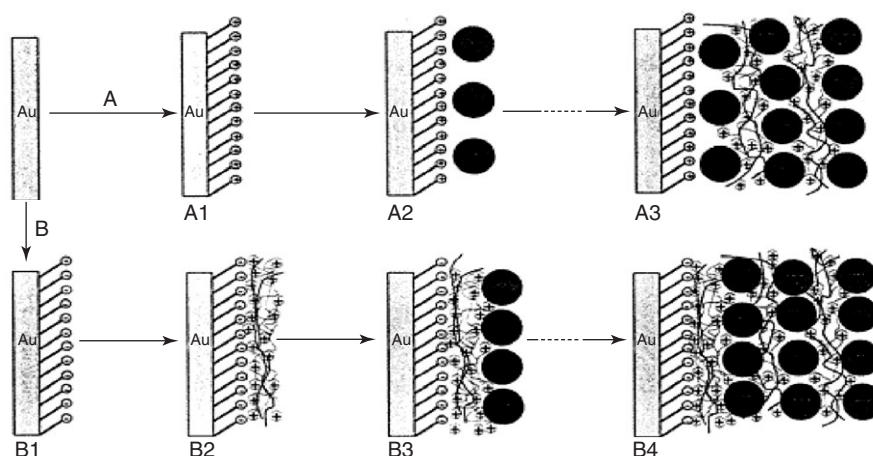
inorganic materials, which has been recently introduced in SAMs to form a sensing interface. The reduced dimensionality at 2D monolayers would facilitate the association of the guest molecules to the imprinted sites, resulting in a fast response of the interface. The imprinted molecular recognition sites are formed by adsorption of the sites in alkanethiols on gold. However, the electrochemical behavior in such a case is not stable. The influence of the stability of the imprinted sites caused by the lateral diffusion of alkanethiols can be overcome by the co-adsorption of two types of molecules: the ‘template’ molecules and the ‘matrix’ molecules (alkanethiols). The shape of the ‘template’ molecules should be similar to the species studied but with additional thiol groups to provide strong binding to the surface; however, the ‘matrix’ molecules should adsorb strongly to the surface to form a monolayer thicker than that of the template molecules (**Figure 1**).

The extension of a 2D monolayer to a 3D ordered assembly represents the recent advances to prepare well-defined multilayers. Among several different approaches, a method based on the electrostatic interaction between oppositely charged species layer-by-layer has been widely used to build up a variety of multilayer assemblies with precisely controlled thickness and layer sequences. A novel strategy for controlled fabrication of a well-defined monolayer and multilayer film containing polyoxometalates (POMs) on a gold electrode covered with alkanethiol SAMs has been proposed (**Figure 2**). During the assembly process, the formation of the multilayer by using electrochemical growth has been proven to have



**Figure 1** Preparation and operation of artificial two-dimensional imprinting sites for the binding of barbiturate. (Reproduced with permission from Mirsky *et al.* (1999) *Angewandte Chemie, International Edition* 38: 1108; © Wiley-VCH.)





**Figure 2** Schematic illustration of deposition procedures and conceptual models of monolayers and multilayers consisting of POMs (large circles) and QPVP-Os (thin curves) on two kinds of positively and negatively charged precursor alkanethiol SAMs. Cysteamine (A) and 3-mercaptopropionic acid (B) SAMs on Au electrodes are typically used. Depicted are various stages of the film formation: (A1) Au/Cyst; (A2) Au/Cyst/POMs; (A3) Au/Cyst/3POMs/2QPVP-Os; (B1) Au/MPA; (B2) Au/MPA/QPVP-Os; (B3) Au/MPA/QPVP-Os/POMs; and (B4) Au/MPA/3QPVP-Os/3POMs. (Reprinted with permission from Cheng L, Niu L, Gong J, and Dong SJ (1999) *Chemistry of Materials* 11: 1465; © American Chemical Society.)

more advantages over the commonly used immersion growth. The tunable multilayer films prepared may find applications in sensors, catalysis, electrochromism, and other thin film molecular devices.

### Mono- and Multilayer Assembly on a Carbon Electrode

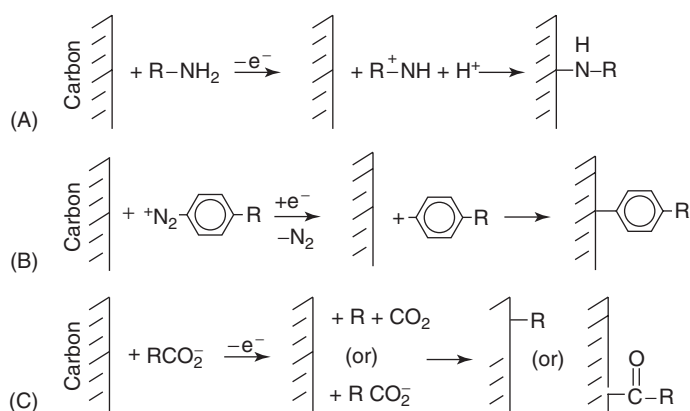
Beside properties of inertness and stability, carbon electrodes have a wide potential range when compared to gold electrodes, which are more popularly used in electrochemical laboratories and fuel cells. The modification of the carbon surface is of importance in electrocatalytic, analytical, and biotechnological applications.

In order to construct an ordered monolayer on a carbon surface, a premodified smooth and functionalized film is needed. Electrochemical reduction of diazonium salts or oxidation of amine-containing compounds and acrylate produces free radicals, resulting in covalent derivatization of the carbon surface in nonaqueous solutions (Figure 3). The premodification by free radical grafting methods is quite mild, simple, and rapid experimentally. The monolayers prepared as a precursor film for subsequent assembly are stable to harsh environments and can endure elevated temperatures, and especially are organized in order, compact, and charge-rich.

The premodification method has been proved to be a general one suitable for anchoring many kinds of organic and inorganic compounds, biomolecules,

and nanoparticles to construct ordered composite monolayers on carbon electrode surfaces. Starting from the obtained composite monolayer, a method based on electrostatic interaction between oppositely charged species, layer by layer, has been used to build up a variety of multilayer assemblies with precisely controlled thicknesses and layer sequences.

There are numerous research works on applications of catalysis and sensors by using the composite monolayer and multilayer modified carbon electrodes; for instance, NADH oxidation can be catalyzed at the dopamine grafted on glassy carbon (GC) electrode; determination of uric acid in the presence of 30-fold excess of ascorbic acid at GC electrode modified with cysteine layer by amine oxidation. The sensitivity ratio of chlorpromazine to uric acid is largely enhanced at the *p*-ethylphenyl monolayer grafted on the GC electrode; and a simple method for the fabrication of biotin-avidin based biosensors at the biotin immobilized on the GC electrode via amine oxidation is realized. Multilayer film systems based on 3D assembly on carbon electrode surfaces exhibit stronger activities, such as the multilayer films consisting of a cationic redox polymer QPVP-Os and POMs through layer-by-layer construction show high electrocatalytic activity on the reductions of  $\text{BrO}_3^-$ ,  $\text{IO}_3^-$ ,  $\text{HNO}_2$ , and  $\text{H}_2\text{O}_2$ ; organic-inorganic multilayer films composed of POMs ( $\text{P}_2\text{W}_{18}\text{O}_{62}^{6-}$ ) and metalloporphyrin ( $\text{CoTMP}_y\text{P}$ ) can catalyze  $\text{O}_2$  reduction mainly through  $4e^-$  to  $\text{H}_2\text{O}$ , whereas when using  $\text{SiW}_{12}\text{O}_{40}^{4-}$  instead of  $\text{P}_2\text{W}_{18}\text{O}_{62}^{6-}$ , the multilayer films exhibit high activity toward the hydrogen evolution reaction; a multilayer film constructed by vanadium-substituted POMs



**Figure 3** Some possible routes for the covalent modification of carbon surface using (A) amino, (B) diazonium, and (C) acrylate based compounds. (Reproduced with permission from Zen JM, Kumar AS, and Tsai DM (2003) Recent updates of chemically modified electrodes in analytical chemistry. *Electroanalysis* 15: 1073–1087; © Wiley-VCH.)

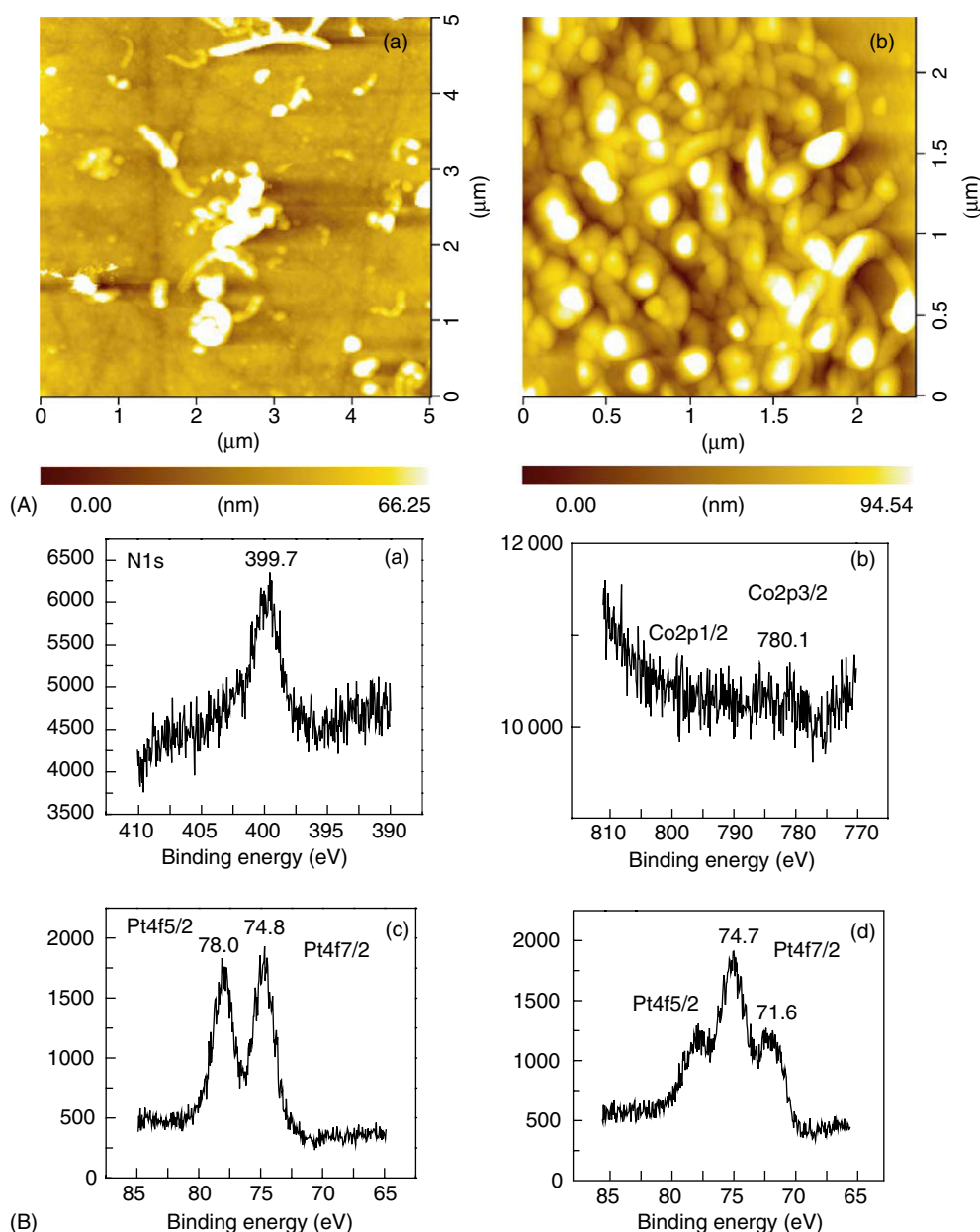
( $P_2W_{15}V_3O_{62}^{9-}$ ) with QPVP-Os can be used to detect ascorbic acid in the presence of dopamine; iron-substituted POMs ( $P_2W_{17}O_{61}Fe^{7-}$ ) with QPVP-Os formed multilayer films that give remarkable catalysis effects on  $H_2O_2$ ,  $BrO_3^-$ , and  $NO_2^-$  reactions.

Carbon nanotubes (CNTs) have attracted increasing attention from various fields recently. The high surface area and hollow geometry, combined with their high electronic conductivity and useful mechanical properties, show that CNTs have the ability to promote electron transfer when used as electrode material in electrochemical reactions. CNTs show a better electrochemical response when compared to ordinary carbon electrodes. One promising application of CNTs is their use in electrochemical sensors and nanoscale electronic devices. CNT electrodes have been utilized to detect: different types of analytes, some gas molecules such as NO,  $NO_2$ ,  $NH_3$ , small molecules such as dopamine, to catalyze biomolecules, and inorganic compounds. CNT electrodes were also successfully used in studying the electrochemistry of proteins. Chemically modified CNT electrodes with different functional groups would be more sensitive and selective in sensor application and catalysis. Recently, the CNTs/Nafion composite film immobilized on GC electrodes exhibited strong and stable electrocatalytic activity and electrogenerated chemiluminescence (ECL) of  $Ru(bpy)_3^{2+}$  responses to tripropylamine and oxalate. Based on the inhibition effect of dopamine on ECL of the above system, the determination of dopamine in the presence of ascorbic acid was realized. Interestingly, a hybrid thin film composed of CNTs/CoTMPyP/Pt on a GC electrode (Figure 4) shows remarkable electrocatalytic activity to oxygen reduction and high stability with promising application in fuel cells.

## Polymer Film Electrode

Polymer-based CMEs have been developed for analytical applications by either amperometric or potentiometric technique over the last decades. There are three kinds of polymers used extensively to prepare polymer film electrodes: (1) redox polymers; (2) ion-exchange and coordination polymers; and (3) electronically conductive polymers. Redox polymers contain electroactive functionalities either within the main chain or side groups pendent to the chain. Poly(vinylferrocene) is the typical example of redox polymers. Ion-exchange and coordination polymers can incorporate electroactive guest molecules although they are not electroactive themselves, for example, poly(vinylpyridine) film can incorporate electroactive counterions through an ion-exchange reaction. The perfluorosulfonate ionomer Nafion is the cation-exchange polymer used in extensive concentration and selective detection studies. Electrochemically conductive polymers contain electroactive chains. Their representatives like polyaniline, polypyrrole, and polythiophene are conjugative, and the cationic sites resulting from the oxidation would be delocalized along the polymer chain. This is in contrast to the case of a redox polymer like poly(vinylferrocene), where the positive charge produced during the oxidation would be localized within the ferrocene.

To prepare polymer film electrodes there are different techniques like dip coating, spin coating, electropolymerization from monomers, and molecularly imprinted polymers (MIPs) developed very recently. They are often used for both potentiometric and amperometric ion sensors. Some feature of different types of ion sensors based on polymer CMEs are shown in Table 1. Conducting polymer-based CMEs



**Figure 4** (A) AFM images: (a) MWNTs/CoTMPyP/PtCl<sub>6</sub><sup>2-</sup> film assembled on GC electrode (before electrochemical reduction); (b) MWNTs/CoTMPyP/Pt hybrid film (after electrochemical reduction at  $-0.7$  V in N<sub>2</sub>-saturated 0.1 mol l<sup>-1</sup> KCl solution). (B) XPS data of MWNTs/CoTMPyP film on GC surface (curves (a) and (b)) and MWNTs/CoTMPyP/PtCl<sub>6</sub><sup>2-</sup> film in the Pt (4f) region before (curve (c)) and after (curve (d)) electrochemical reduction. (Qu JY, Shen Y, Qu XH, and Dong SJ (2004) Preparation of hybrid thin film modified carbon nanotubes on glassy carbon electrode and its electrocatalysis for oxygen reduction. *Chemical Communications* 2004: 34–35; reproduced by permission of The Royal Society of Chemistry.)

can be well used in ion-selective electrodes (ISEs) for miniaturization.

The preparation of MIPs includes the formation of a complex between the functional monomer and guest molecules (template) in solution followed by freezing this complex through polymerization in the presence of high concentration of cross-linker. After removing the template by washing, the polymer

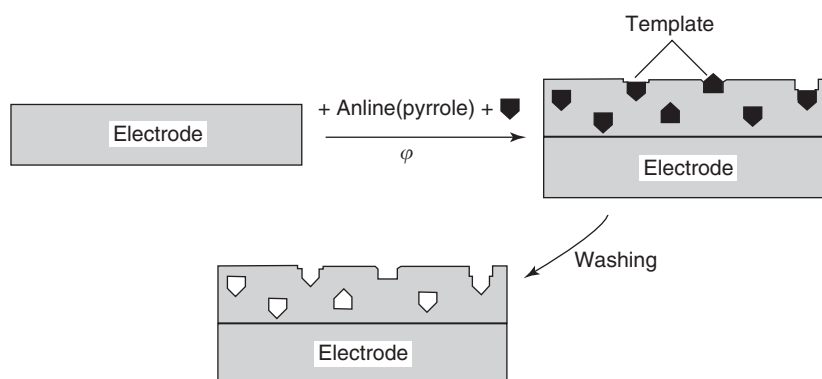
binding sites (imprints) are retained in the structure specifically for the template molecule. The preparation of MIPs can also be processed on the electrode surface by electropolymerization (Figure 5). This technique is convenient to deposit a sensitive layer with high precision on the electrode surface that can be used for various substances. This is also important for micro- and multisensor production.

**Table 1** Features of different types of ion sensors

Types of ISE	Structure	Fabrication methods	Sensitivity	Potential stability	Durability and miniaturizability
Conventional	L/ISM(I)	Solution casting	Ion	Excellent	Poor
CWE	S/ISM(I)	Solution casting	Ion	Poor	Good
SCISE	S/CP/ISM(I)	(1) Electropolymerization (2) Solution casting	Ion	Good	Good
SPISE	S/ISM(CP,I)	Solution casting	Ion/redox	Good	Good
CPISE	S/CP(I)	Electropolymerization or solution casting	Ion/redox	Good	Excellent

CWE, coated-wire electrode; SCISE, solid-contact ISE; SPISE, single-piece ISE; CPISE, conducting polymer-based ISE. L, aqueous electrolyte solution or hydrogel; ISM(I), ion-selective membrane containing ion-recognition sites (I); S, solid electronic conductor; CP, conducting polymer; ISM(CP, I), ion-selective membrane containing conducting polymer (CP) and ion-recognition sites(I); CP(I), conducting polymer containing ion-recognition sites(I).

Reprinted with permission from Bobacka J, Ivsaka A, and Lewenstam A (2003) Potentiometric ion sensors based on conducting polymers. *Electroanalysis* 15: 366–374; © Wiley-VCH.



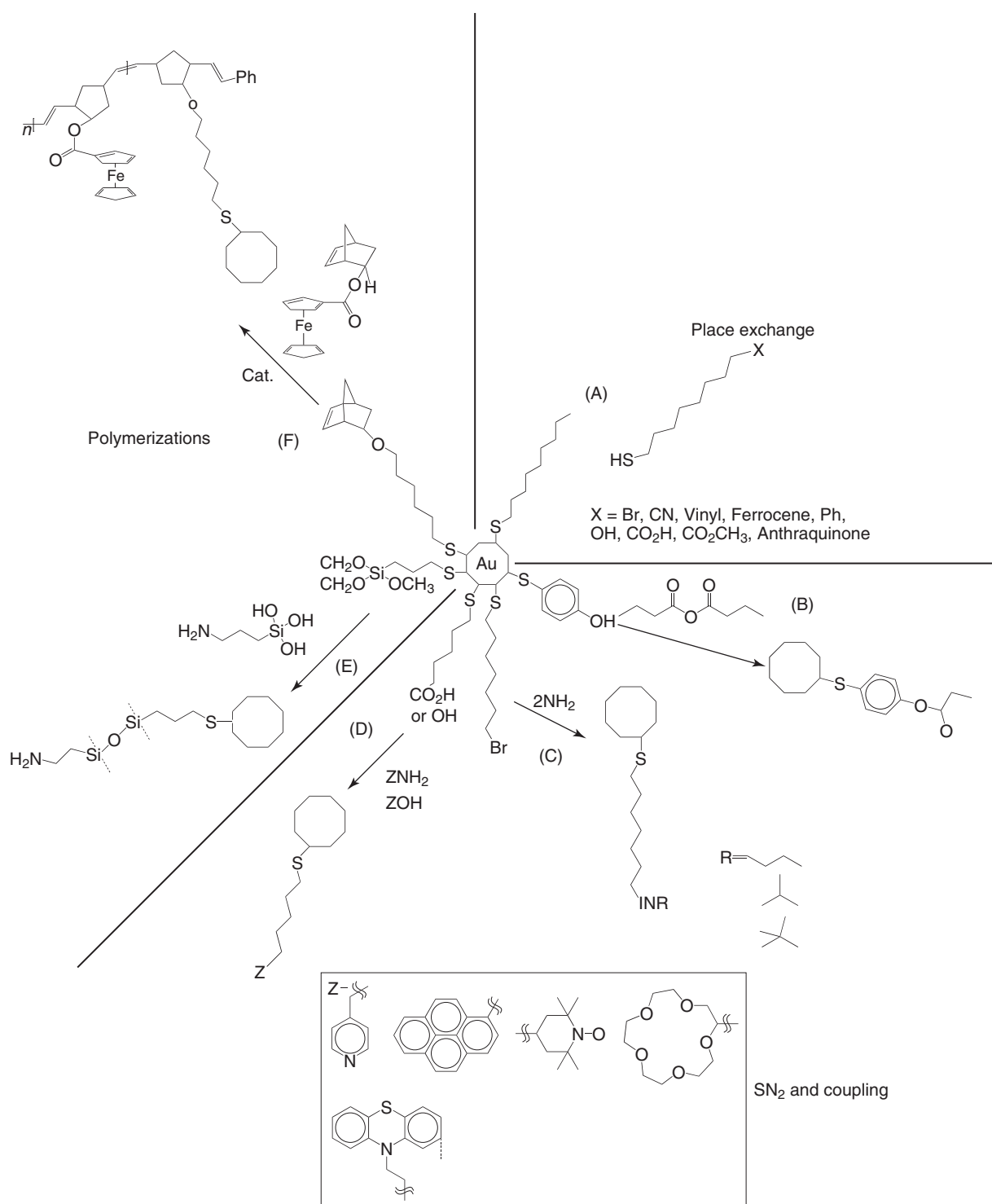
**Figure 5** Scheme for molecularly imprinted polymer film on electrode by electropolymerization. (Reproduced with permission from Piletsky SA and Turner APF (2002) Electrochemical sensors based on molecularly imprinted polymers. *Electroanalysis* 14: 317–323; © Wiley-VCH.)

Inorganic ions, drugs, nucleic acids, proteins, and even cells are successful examples of imprinting. In this way, affinity sensors, receptor sensors, and catalytic sensors based on MIPs have been explored. For affinity sensors, immunosensor-like devices were prepared by a 2D MIP technique with molecular imprinting on chemisorbed alkanethiol SAMs; then after necessary procedures, vitamins K<sub>1</sub>, K<sub>2</sub>, E, cholesterol, and adamantine could be detected by the strong electrochemical signals yielded. The sensors for nucleic acids, cholesterol, and catechol derivatives can be fabricated first by their adsorption as a template on the ITO surface and then by the treatment of the electrode with adsorbed template using trimethyl chlorosilane from the gas phase.

## Nanoparticle Modified Electrodes

Nanoparticles often display unusual physical and chemical properties mainly depending on their size

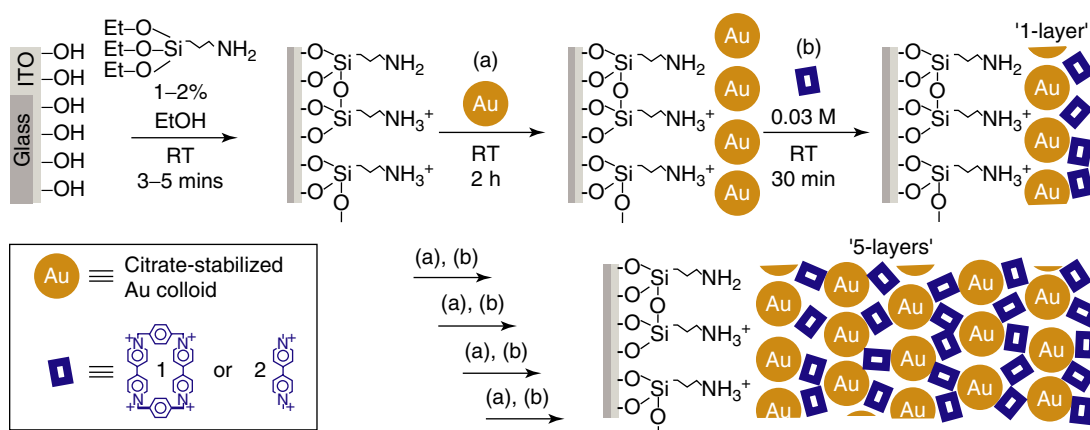
and shape, which have kindled much attention recently. Metal nanoparticles can act as catalyst, facilitate the electron transfer, and be modified with many biomolecules and ligands. All the properties permit obtaining a sensing interface that can be attached on the electrode surface. A nanoparticle modified electrode may be prepared by metal particles dispersed in a polymer film, in porous aluminum oxide film, with SAMs, or with films of porous glass based on sol–gel formation. Metal (especially Pt, Au) nanoparticles and their derivatives like monolayer protected clusters have extended applications in electroanalytical chemistry, for example, nonenzyme sensors and enzyme sensors. Gold nanoparticles with thiol derivatives as strong stabilizers can form a monolayer on the surface with strong coordination to modify metal nanoparticles with various organic ligands (**Figure 6**) for optical, chemical, and biosensing biomolecules (DNA) while retaining their activity. For nonenzyme sensors, metal nanoparticles or functionalized nanoparticles are used as sensing phase, while for enzyme



**Figure 6** Reactivity of monolayer-protected clusters: (A) place exchange, (B) reaction of *p*-mercaptophenol with propionic anhydride, (C)  $\text{S}_\text{N}^2$  reaction of bromoalkanethiolated MPCs with primary alkylamines, (D) amide and ester coupling reactions, (E) siloxane formation reactions, and (F) transition-metal-catalyzed ring-opening metathesis polymerization. (Reprinted with permission from Templeton AC, Wuelfing WP, and Murray RW (2001) Monolayer-protected cluster molecules. *Accounts in Chemical Research* 33: 27–36; © American Chemical Society.)

sensor, enzyme modified metal nanoparticles are used as sensing phase where nanoparticles act as mediators. For example, gold nanoparticles as the prepared

working electrode of the  $\text{SO}_2$  sensor catalyze the electrochemical oxidation of  $\text{SO}_2$ ; gold nanoparticles dispersed on the electrode surface can facilitate the



**Figure 7** Stepwise assembly of the three-dimensional array of organic oligocation-cross-linked Au nanoparticle superstructures on glass surfaces. (Reprinted with permission from Shipway AN *et al.* (2000) *Langmuir* 16: 8789; © American Chemical Society.)

electrooxidation of carbohydrates in alkaline media, which has been used as amperometric sensor for carbohydrates in flow injection and liquid chromatography; gold nanoparticles can also facilitate the electron transfer of redox proteins like horse heart cytochrome *c* and so forth. Another typical example in the selective electrochemical sensing of various substances like hydroquinone, ferrocene derivatives, adrenaline, dopamine etc., by the organization of layered gold nanoparticle architectures on ITO–glass support. It was processed by cross-linking the particles with oligocationic bis-bipyridinium cyclophene or molecular square transition-metal complex (Figure 7) where the organic cross-linker molecules act as receptors for  $\pi$ -donor substances, thus causing the concentration of such guests on the electrode.

Nanoparticle monolayers and multilayers can also be prepared directly by electrochemistry without performance of the nanoparticles. For example, Ag nanoparticles monolayer can be formed by  $\text{Ag}^+$  grafted on 4-aminophenyl monolayer on highly oriented pyrolytic graphite followed by pulse electroreduction, and then the uniformly dispersed Ag nanoparticles as ordered monolayers are prepared. By a similar design Pd and Pt ordered monolayers can also be fabricated. For Pt (or Pd) nanoparticle ordered multilayers, they can be assembled with  $\text{PtCl}_6^-$  (or  $\text{PdCl}_4^{2-}$ ) and CoTMPyP alternatively followed by electroreduction. The Pt or Pd nanoparticle ordered multilayers formed on the electrode surface demonstrate high catalytic activity to  $\text{O}_2$  and  $\text{H}_2\text{O}_2$  reduction.

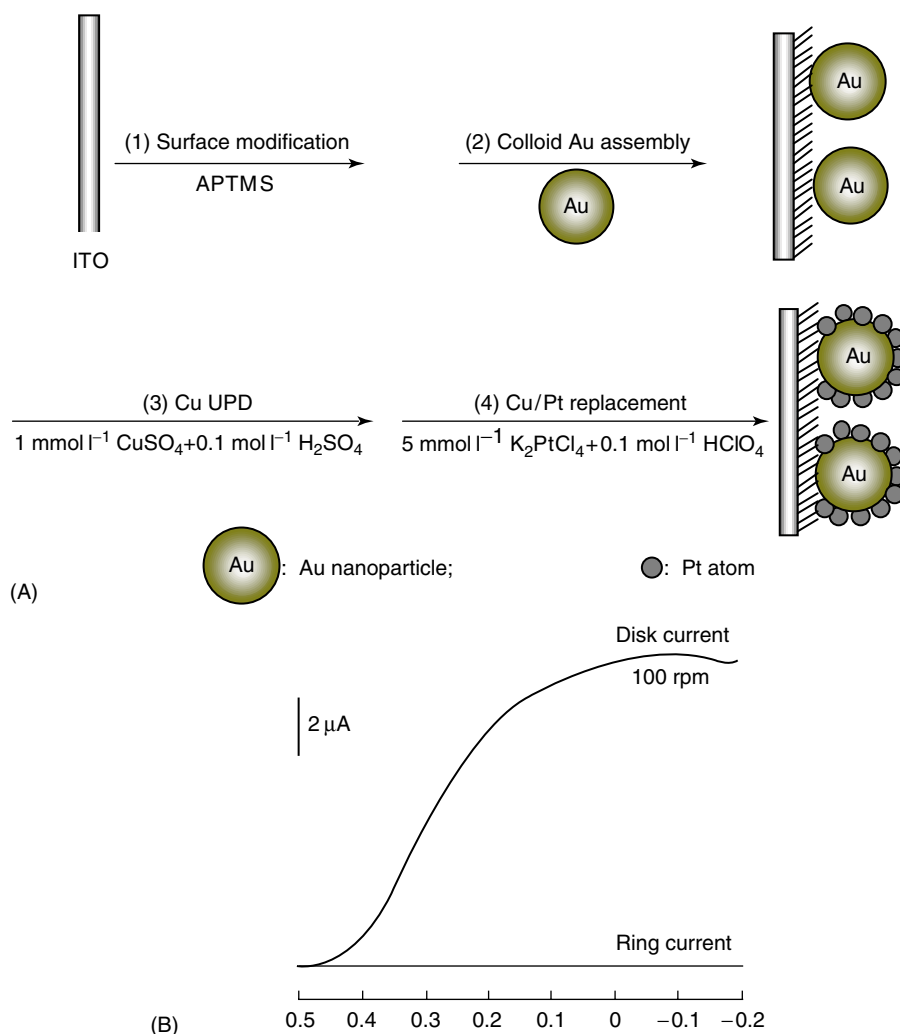
Another approach to prepare nanoparticle ordered multilayers by preceding the formation of the nanoparticle monolayer can be achieved by a bottom-up

way. A molecular cross-linker APTMS (a kind of amino-silane, (3-aminopropyl)-trimethoxysilane) layer is first modified on the solid substrate (e.g., ITO), and then the nanoparticles monolayer is assembled. The organization of a gold nanoparticle multilayer on ITO by using oligocationic bis-bipyridinium cyclophene or molecular square transition-metal complexes through cross-linking the nanoparticles has been achieved (Figure 7). Similarly, the nanoparticle ordered multilayer can be assembled alternatively by ‘molecular gels’ (dithiols, dipyrindinium, diamino, and thio-containing molecules) and nanoparticle sol. For example, on ITO electrode the thionine dye molecule bridged nanoparticle multilayers demonstrate porosity, permeability, and 3D conductivity.

Enzyme sensors are another important application of metal nanoparticles in CMEs besides nonenzyme sensors. Many enzymes can keep their activity when anchored onto gold nanoparticles. A novel method for fabrication of a biosensor based on the combination of sol–gel and self-assembled techniques has been introduced very recently. For example, the gold nanoparticles and enzyme horseradish peroxidase (HRP) can be successfully immobilized on gold electrode by the help of sol–gel with thiol groups, and the direct electrochemistry of HRP has been achieved and the biosensor thus prepared exhibits fast response, good reproducibility, and long-term stability.

Using underpotential deposition with the redox replacement technique a novel electrochemical approach for nanoparticle-based catalyst has been designed. Here, the as-prepared Pt-coated Au nanoparticle monolayer at atomic level on the electrode surface can reduce  $\text{O}_2$  predominately by 4e to  $\text{H}_2\text{O}$ , which was confirmed by rotating ring disk electrode technique (Figure 8).





**Figure 8** (A) Schematic illustrations of preparation procedures of ultrathin (nominally monolayer level) platinum-coated gold nanoparticle monolayer films on an ITO electrode surface. (B) Current–potential curves for the reduction of air-saturated  $\text{O}_2$  at a rotating platinum ring-graphite disk electrode with ultrathin Pt-coated Au nanoparticles monolayer adsorbed on the disk electrode. The potential of the ring electrode was maintained at 1.0 V. Rotation rate: 100 rpm. Scan rate:  $50 \text{ mV s}^{-1}$ ; supporting electrolyte:  $0.1 \text{ mol l}^{-1} \text{ H}_2\text{SO}_4$ . (Reprinted with permission from Jin YD, Shen Y, and Dong SJ (2004) Electrochemical design of monolayer-level platinum-coated gold core-shell nanoparticles monolayer films as novel nanostructured electrocatalysts for oxygen reduction. *Journal of Physical Chemistry B*, in press; © American Chemical Society.)

See also: **Molecularly Imprinted Polymers. Sensors:** Overview; Amperometric Oxygen Sensors; Calorimetric/Enthalpimetric.

## Further Reading

- Bakker E and Telting-Diaz M (2002) Electrochemical sensors. *Analytical Chemistry* 74: 2781–2800.
- Dong SJ, Che GL, and Xie YW (2003) *Chemically Modified Electrodes*. Beijing: Science Press.
- Downard AJ (2000) Electrochemically assisted covalent modification of carbon electrodes. *Electroanalysis* 12: 1085–1096.

- Fendler JH (2001) Chemical self-assembly for electronic applications. *Chemistry of Materials* 13: 3196–3210.
- Gooding JJ, Mearns F, Yang WR, and Liu JQ (2003) Self-assembled monolayers into the 21st century: recent advances and applications. *Electroanalysis* 15: 81–96.
- Hernandez-Santos D, Gonzalez-Garcia MB, and Garcia AC (2002) Metal-nanoparticles based electroanalysis. *Electroanalysis* 14: 1225–1235.
- Mirsky VM (2002) New electroanalytical applications of self-assembled monolayers. *Trends in Analytical Chemistry* 21: 439–450.
- Murray RW (ed.) (1992) *Molecular Design of Electrode Surfaces*. New York: Wiley.

Shipway AN, Katz E, and Willner I (2000) Nanoparticle arrays on surfaces for electronic, optical, and sensor applications. *Chemphyschem* 1: 18–52.

Zhao Q, Gan ZH, and Zhuang QK (2002) Electrochemical sensors based on carbon nanotubes. *Electroanalysis* 14: 1609–1613.

## Microorganism-Based

**D M Rawson**, Luton, UK

**B G D Haggett**, University of Luton, Luton, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The exploitation of microorganisms for sensing purposes covers a wide range of applications. In the most general case, a microorganism or consortium of microorganisms is exposed to a test environment and the effect(s) of exposure are compared to a control environment using any one of a number of measures. Such approaches might be divided into three broad areas:

- traditional microbial assays, using classical microbiological methods, e.g., cell counting;
- microbial assays using close coupling with physicochemical methods of transducing the signal, e.g., luminometer; and
- integration of microorganisms with physicochemical transduction methods, e.g., electrochemical biosensors.

The term ‘biosensor’ has been used loosely to describe all of these situations, but this article considers only those situations where the microorganisms are integrated with a transducer or the microorganisms are genetically modified and then used in close coupling with a transducer.

The exploitation of microorganisms as the biological component in biosensors is not as advanced or as widespread as the use of biomolecular biocatalysts in enzyme- and antibody-based biosensors. However, the stability of intact cells and the great diversity of microbial species enable the development of biosensors for use in a range of applications, including analyte determination in complex mixtures, process control, environmental monitoring, toxicity assessment, and mutagen detection. Electrochemical transducers dominated the earlier sensor types used to incorporate microbial cells, with amperometric oxygen electrodes and potentiometric ion-selective electrodes (ISEs) accounting for the majority of devices. Optical transducers, however, are becoming more

prevalent, with a wide variety of fluorescence- and luminescence-based devices having been described – particularly related to those sensors incorporating genetically modified organisms.

Monitoring of changes in (1) the level of microbial metabolic activity or (2) the concentration of products from microbial biotransformation, of specially selected biocatalysts, can provide rapid analysis or determination of sample quality. Whilst microbial-based sensors are able to demonstrate good sensitivity, they do not possess the inherent selectivity of enzyme- or antibody-based sensors. Therefore, the most promising areas for microbial biosensors will probably be in broad spectrum monitoring or those applications where specificity is not a problem. The recent scientific literature shows continuing interest in the use of microorganism-based sensors for environmental monitoring (particularly the development of rapid methods for estimation of biological oxygen demand, BOD). Recent global security issues also suggest new opportunities for such biosensors.

### Principles

#### Microbial Cells as Sensor Biocatalysts

A wide range of microorganisms have been used as biocatalysts. Bacteria dominate but cyanobacteria, microfungi, protozoa, and microalgae have all been successfully used in biosensors. Whilst living cells have many benefits as biocatalysts they also pose unique problems and these need to be successfully overcome before microbial-based sensors can become commercially successful. Six major benefits can be identified:

- high degree of stability, incorporating complete metabolic systems within a protected environment, and possessing damage repair capabilities allowing *in situ* recovery;
- wealth of known species covering many metabolic types, with a wide range of sensitivities and nutritional requirements;
- ability to perform complex reaction sequences;
- ease of maintenance in pure culture, low harvesting, and production costs;

Shipway AN, Katz E, and Willner I (2000) Nanoparticle arrays on surfaces for electronic, optical, and sensor applications. *Chemphyschem* 1: 18–52.

Zhao Q, Gan ZH, and Zhuang QK (2002) Electrochemical sensors based on carbon nanotubes. *Electroanalysis* 14: 1609–1613.

## Microorganism-Based

**D M Rawson**, Luton, UK

**B G D Haggett**, University of Luton, Luton, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The exploitation of microorganisms for sensing purposes covers a wide range of applications. In the most general case, a microorganism or consortium of microorganisms is exposed to a test environment and the effect(s) of exposure are compared to a control environment using any one of a number of measures. Such approaches might be divided into three broad areas:

- traditional microbial assays, using classical microbiological methods, e.g., cell counting;
- microbial assays using close coupling with physicochemical methods of transducing the signal, e.g., luminometer; and
- integration of microorganisms with physicochemical transduction methods, e.g., electrochemical biosensors.

The term ‘biosensor’ has been used loosely to describe all of these situations, but this article considers only those situations where the microorganisms are integrated with a transducer or the microorganisms are genetically modified and then used in close coupling with a transducer.

The exploitation of microorganisms as the biological component in biosensors is not as advanced or as widespread as the use of biomolecular biocatalysts in enzyme- and antibody-based biosensors. However, the stability of intact cells and the great diversity of microbial species enable the development of biosensors for use in a range of applications, including analyte determination in complex mixtures, process control, environmental monitoring, toxicity assessment, and mutagen detection. Electrochemical transducers dominated the earlier sensor types used to incorporate microbial cells, with amperometric oxygen electrodes and potentiometric ion-selective electrodes (ISEs) accounting for the majority of devices. Optical transducers, however, are becoming more

prevalent, with a wide variety of fluorescence- and luminescence-based devices having been described – particularly related to those sensors incorporating genetically modified organisms.

Monitoring of changes in (1) the level of microbial metabolic activity or (2) the concentration of products from microbial biotransformation, of specially selected biocatalysts, can provide rapid analysis or determination of sample quality. Whilst microbial-based sensors are able to demonstrate good sensitivity, they do not possess the inherent selectivity of enzyme- or antibody-based sensors. Therefore, the most promising areas for microbial biosensors will probably be in broad spectrum monitoring or those applications where specificity is not a problem. The recent scientific literature shows continuing interest in the use of microorganism-based sensors for environmental monitoring (particularly the development of rapid methods for estimation of biological oxygen demand, BOD). Recent global security issues also suggest new opportunities for such biosensors.

### Principles

#### Microbial Cells as Sensor Biocatalysts

A wide range of microorganisms have been used as biocatalysts. Bacteria dominate but cyanobacteria, microfungi, protozoa, and microalgae have all been successfully used in biosensors. Whilst living cells have many benefits as biocatalysts they also pose unique problems and these need to be successfully overcome before microbial-based sensors can become commercially successful. Six major benefits can be identified:

- high degree of stability, incorporating complete metabolic systems within a protected environment, and possessing damage repair capabilities allowing *in situ* recovery;
- wealth of known species covering many metabolic types, with a wide range of sensitivities and nutritional requirements;
- ability to perform complex reaction sequences;
- ease of maintenance in pure culture, low harvesting, and production costs;

- high levels of catalytic activity achieved with modest loading levels; and
- amenable to genetic manipulation, offering the possibility of specific trait incorporation and biocatalyst 'tailoring' for specific applications.

Whilst in the biosensor research laboratory these benefits can be exploited to the full, problems arise in developing sensors incorporating microorganisms that can be produced commercially. Central to this is the need to develop methods of handling microorganisms during fabrication and bulk production of biosensors. In particular, further development is required for:

- harvesting of cells in the appropriate metabolic state;
- appropriate immobilization techniques; and
- preservation methods to give good shelf life.

Optimization of these factors is required so as to enable batch consistency, stability, and extended operational life of the biosensor. These requirements are easier to satisfy with bacterial species, than with protozoa, algae, or fungi.

### Monitoring Biological Response

Unlike the biocatalytic or bioaffinity components of enzyme- and antibody-based biosensors, living microbial cells undertake a vast range of biochemical processes and respond to a multitude of physiological events. Monitoring of a microbial biocatalyst can therefore be directed at specific cellular events or at total metabolic activity. Present technology allows only a limited range of interrogation methods for use with living cells, and the full potential of microbial biosensors will only be achieved when a wider range of techniques are developed allowing nondestructive targeting of specific cellular events *in vivo*.

The majority of devices employ one of the following three electrochemical approaches to monitoring the biological component:

- monitoring oxygen or carbon dioxide to follow respiratory or photosynthetic activity;
- detection of metabolic products other than carbon dioxide or oxygen; and
- using redox mediators to monitor respiratory or photosynthetic activity.

Calorimetric methods can also be applied, exploiting the thermal changes that accompany enzymatic reactions within the microbial biocatalyst to monitor metabolic activity levels.

Transducer monitoring of microbial activity has many advantages over conventional microbial bioassay techniques, in particular:

- rapid response and assay/test times, often 15 min or less;
- ease of use by nonspecialists;
- lower assay cost; and
- potential for field-based testing and online use.

### Types of Biological Response

**Determination of specific analytes** The rapid response of bacterial cells to the appearance of respiratory substrates in their environment was the first feature exploited in microbial biosensors. Monitoring oxygen uptake of a heterotrophic species by means of the Clark oxygen electrode, it is possible to determine accurately the substrate concentration, over a limited range, from the respiratory response of the biocatalyst. Respiratory activity can also be monitored using carbon dioxide probes, mediated amperometry or calorimetry.

Unlike enzyme-based biosensors such responses are subject to interference from a large number of other compounds to which the microbial biocatalyst would respond. Determination of the concentration or presence of specific analytes in mixtures is therefore difficult, although several approaches to overcome this problem have been employed including: careful biocatalyst selection; biocatalyst conditioning or induction; and the use of inhibitors to suppress unwanted systems.

Alternatively, it is sometimes possible to gain some degree of selectivity by monitoring selected metabolic events within the cell, rather than total metabolic activity. The use of ISEs to detect product formation from a particular metabolic event can be used to determine the presence and concentration of substrates. Ammonia gas sensors have been used to detect deamination by bacterial cells for the determination of amino acids. However, the development of microbial biosensors capable of specific amino acid detection has to overcome the problem of interference from other amino acids and related compounds.

Developments in molecular biology have enabled the exploitation of genetically modified microorganisms to provide appropriate sensitivity to external analytes. Typically, there are two applications of genetically modified microorganisms in biosensor configuration – broad spectrum or analyte specific. In both cases the microorganism is genetically modified by the introduction of an appropriate reporter gene, whose expression is easily detected and monitored. In the majority of cases these are light

emitting or fluorescent reporter genes from either bacteria, for example, green fluorescent protein (*Aquorea victoria* gfp) or bioluminescence (*Vibrio fischeri* lux) genes, or the light emitting *luc* gene from the firefly. Microorganisms incorporating light-emitting reporting genes can be used in broad spectrum biosensors to monitor the metabolic status of the cell and respond to changes in substrate level or toxic challenge. Analyte specificity has been achieved in biosensors exploiting microorganisms that incorporate reporter genes linked to selected promoter genes triggered by the presence of specific analytes such as heavy metal ions (e.g., mercury and cadmium). Incorporation of additional genetic components has also been used to provide an internal reference signal so as to correct analytical responses due to nonspecific interferences.

The future use of analyte-specific microbial biosensors is likely to be in those situations where biomolecular-based sensors are impractical – for reasons of availability, stability, or cost.

**Broad spectrum analysis** The ability of microorganisms to respond to a wide range of materials can be a positive advantage, enabling their use in situations where analyte specificity would be a disadvantage. The monitoring of metabolic stimulation associated with substrate presence has been used in such applications as the monitoring of fermenter liquor substrate profiles. The recent scientific literature shows a continuing interest in the development of rapid methods for the estimation of BOD since the conventional 5-day BOD<sub>5</sub> is not suited to process control – for example, in waste water treatment plants for environmental protection.

In some situations it may be appropriate to use single-species biocatalysts where the response profile is known to match the expected analyte composition of the sample or environment to be monitored. In other situations it may be desirable to have as broad a response as possible in order to allow the detection of both the expected and unexpected. In these latter cases, a number of different microbial species may be used either in multisensor arrays or as a mixed microbial population incorporated into the sensor.

These different types of configuration are also appropriate for use in the detection of potentially toxic or mutagenic species. The use of microbial bioassays for mutagenicity and toxicity testing is now well established, and the coupling of such tests with appropriate transducers shows promise for rapid testing. In these applications, the metabolic activity of the microbial biocatalyst is monitored for signs of perturbation caused by the presence of cytotoxic or genotoxic chemicals. Recent developments in global

security issues suggest potential for development of microorganism-based sensing systems to be deployed at sensitive sites for the detection of chemical and biological terrorist attacks.

## Sensor Design

The choice of biocatalyst, transducer, and method of biocatalyst immobilization all influence sensor design, and many different combinations of microbial biocatalysts and transducers have been investigated.

### Configuration of Biocatalyst and Transducer

The preferred biosensor configuration for most applications is the electrode type, with intimate contact between the biocatalyst and the transducer element. Microorganism-based biosensors initially exploited existing ‘off-the-shelf’ electrochemical transducer electrodes and developed different strategies to incorporate the biological component. Whilst this approach was convenient for investigating biocatalyst response and exploring potential applications, conventional electrodes are large and expensive. For most applications the electrodes, incorporating the biocatalyst, need to be manufactured in bulk and at a cost appropriate for single-use disposable units.

In those cases where the transducer element does not lend itself to bulk manufacture and disposal after use, an alternative approach is to arrange for the biological component to be incorporated in a low-cost disposable unit, with a simple method for replacement when necessary. This is particularly suitable for flow-through-type measurement systems, where the biocatalyst can be upstream of the transducer – as in the flow-injection analysis (FIA) approach to monitoring.

Some microbial biomonitors have been described where the biological component is in the form of a cell suspension in a stirred reactor and the electrodes are immersed in the reactor liquid to monitor changes in parameters such as pH,  $pO_2$ , or  $pCO_2$ . Whilst this approach has much in common with microorganism-based sensors it is not appropriate to use the term biosensor for such devices.

### Biocatalysts

There is often a choice of microorganisms that can be considered for use in any particular application. This can be seen with commonly addressed applications for microbial biosensors such as the determination of single components, e.g., glucose or ethanol, or the detection of toxic pollutants, e.g., atrazine or pentachlorophenol (PCP). Many microorganisms could be selected for such applications, as a wide range of

microorganisms are able to utilize glucose or ethanol, while atrazine and PCP are toxic to most microorganisms. Even where unusual substrate detection is required, many bacterial and fungal species can be conditioned to utilize such compounds. Unlike enzyme- and antibody-based biosensors, microbial biosensors are not restricted to one or a very few biocatalysts for a particular application.

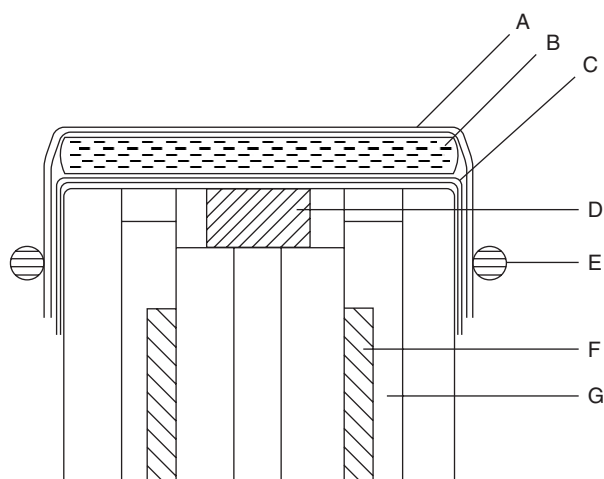
Ultimate selection of biocatalysts depends on several factors:

- screening tests for most responsive/sensitive species, using both conventional bioassay methods and biosensor monitoring;
- identification of the biocatalysts most conducive to monitoring using transducer technology;
- biocatalyst stability in the biosensor configuration;
- amenability to *in situ* preservation.

### Transducers

The majority of microbial biosensors employ one or other of the following transducer elements.

**Oxygen electrodes** The incorporation of an immobilized layer of microbial cells onto the outer surface of the polytetrafluoroethene (PTFE) membrane of a Clark amperometric oxygen electrode allows depletion of oxygen in the microbial layer due to changes in respiratory activity to be monitored continuously. Standard oxygen electrodes are most frequently used (Figure 1), but smaller and cheaper electrodes can be produced, typically using screen printing or microfabrication techniques.



**Figure 1** Biosensor assembly based on the Clark oxygen electrode. A, dialysis membrane; B, bacteria-loaded tortuous filter; C, PTFE membrane; D, platinum electrode; E, 'O' ring; F, silver/silver chloride electrode; G, potassium chloride solution.

**Potentiometric electrodes** Ammonia or carbon dioxide gas sensing probes can be employed in the same way as the oxygen electrode, with the immobilized microbial layer held in place on the surface of the gas-permeable membrane. The ammonia probe is commonly used to determine amino acids by detection of ammonia production by microbial deamination of amino acids. Carbon dioxide probes are alternatives to oxygen electrodes for monitoring respiratory activity.

**Mediated amperometric electrode** Redox mediators such as the ferricyanide ion, *p*-benzoquinone, or dimethylferrocene can be used as electron carriers allowing redox events within the cell to be monitored using suitably poised electrodes (Figure 2). Mediator reduced by enzymatic events within the cell is reoxidized at the working electrode surface resulting in a current flow. This technique has been successfully applied to the monitoring of bacterial and cyanobacterial respiratory and photosynthetic activity using nonmembrane-penetrating mediators to monitor dehydrogenase activity of the plasmalemma. Mediated amperometry can be applied to both probe and FIA-type configurations.

**Luminescence or fluorescence** A wide variety of microorganisms is naturally bioluminescent and/or has been genetically manipulated to incorporate genes encoding for proteins that bring about bioluminescence. In the simplest cases, the intensity of light emission is measured in cuvettes in a conventional luminometer. In other cases, the microorganisms may be attached to fiber optic cables.

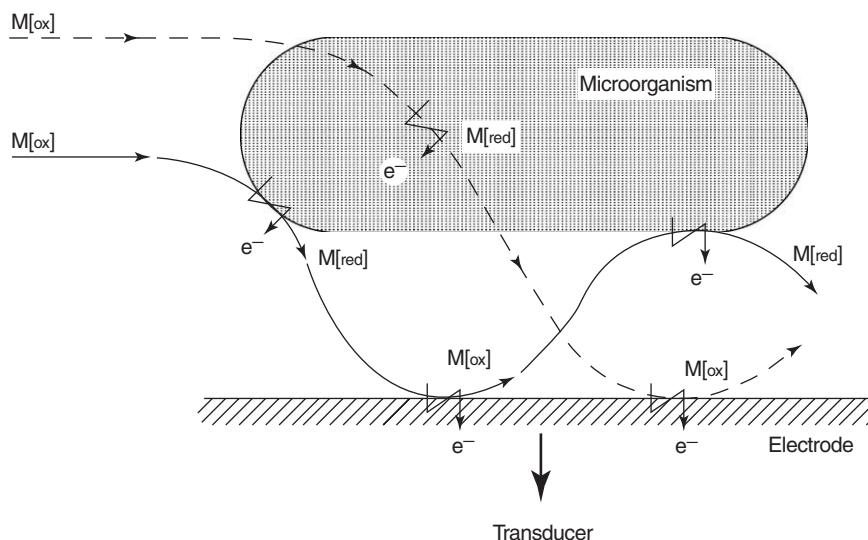
Fluorescence is similarly used for monitoring microbial activity. Chlorophyll fluorescence is widely used to monitor cyanobacteria or algae, while the incorporation of genes encoding for green fluorescent protein (gfp) or other fluorescent probes allows a wide variety of cell types to be monitored this way. Most conveniently, the cells are coupled to a fluorimeter using an optical fiber cable.

### Biocatalyst Immobilization

Physical entrapment is the most common method of immobilization, and this can take several forms:

- capture on tortuous or nontortuous filter discs held in place on the electrode/probe by dialysis or similar membranes;
- entrapment within a polymer matrix, such as calcium alginate, carrageenan, or collagen;





**Figure 2** Electron transfer events at the surface of a mediated amperometric bacterial biosensor, showing mediators assessing redox sites on the cytoplasmic membrane and within the cell.  $M[ox]$ , oxidized mediator;  $M[red]$ , reduced mediator.

- cell suspensions trapped between the PTFE membrane of an oxygen electrode or ISE and an outer semipermeable membrane.

Membrane immobilization gives the best accuracy of loading, whilst the use of cell suspension is the most difficult method for achieving reproducible loading. Where the biocatalyst is upstream of the transducer element there is scope for a wide range of immobilization methods. Most systems employ microreactors of the packed bed or filter support type. This approach also allows much biocatalyst to be loaded at much higher levels than does the probe type of approach to sensor systems.

## Performance

A considerable number of analytes have been detected by microbial biosensors (Table 1). The overall performance of these biosensors needs to be compared to other analytical methods before any judgment on their suitability for the task can be made. Microbial biosensors are of most interest when they offer unique capabilities or significant advantages in the areas of cost and/or ease of use. Some examples of typical performance profiles are given below.

### Single-Component and Analyte-Specific Determination

The simplest type of analysis is that where only one component is present and the sensor is being used to

determine concentration. Here, interference from other chemical species is not a problem, and microbial sensors perform well in this situation. However, when specificity is required to allow analyte detection in complex mixtures of unknown composition interfering chemical species can become a problem.

Bacterial biocatalysts incorporated on to oxygen electrodes and ISEs typically have response times to analytes of between 5 and 15 min and show good stability for 7–30 days with storage normally in buffered saline between assays. Limits of detection and detection ranges are often similar to, or overlapping with those of, enzyme-based systems, such as the examples given below.

Glucose determination by

glucose oxidase biosensor (oxygen electrode, low concentrations; pH electrode, high concentrations):

$$1 \times 10^{-5} - 7 \times 10^{-1} \text{ mol dm}^{-3}$$

*Ps. Fluorescens* oxygen electrode:

$$1 \times 10^{-5} - 1 \times 10^{-4} \text{ mol dm}^{-3}$$

*Saccharomyces cerevisia* CO<sub>2</sub> electrode:

$$1 \times 10^{-3} - 1 \times 10^{-2} \text{ mol dm}^{-3}$$

Glutamine determination by

glutaminase biosensor:

$$1.5 \times 10^{-4} - 3.3 \times 10^{-3} \text{ mol dm}^{-3}$$

*Sarcina flava* ammonia gas ISE:

$$1 \times 10^{-4} - 1 \times 10^{-2} \text{ mol dm}^{-3}$$

These two examples also show good specificity for the target analyte, but many microbial biosensors

**Table 1** Microbial biocatalysts with analytes detected

Analyte	Biocatalyst	Transducer
Acetic acid	<i>Trichosporon brassicae</i>	Oxygen electrode
Acrylonitrile	<i>Rhodococcus ruber</i>	Conductivity
Ammonia	<i>Nitrosomonas</i> sp./ <i>Nitrobacter</i> sp.	Oxygen electrode
$\alpha$ -Amylase	<i>Bacillus subtilis</i>	Oxygen electrode
Antibiotics	<i>Escherichia coli</i>	Potentiometric, fluorescence, luminescence
L-Arginine	<i>Streptococcus faecium</i>	Potentiometric
	<i>Streptococcus lactis</i>	Potentiometric
Aspartame	<i>Bacillus subtilis</i>	Oxygen electrode
Benzene	<i>Escherichia coli</i> (lac::luxCDABE)	Luminescence
BOD <sup>a</sup>	<i>Klebsiella</i> sp.	Oxygen electrode
	<i>Trichosporon cutaneum</i>	Oxygen electrode
	<i>Trichosporon cutaneum</i> / <i>Pseudomonas putida</i> / <i>Bacillus licheniformis</i>	Mediated amperometric
Carbon dioxide	<i>Pseudomonas</i> sp.	Oxygen electrode
Chemical warfare agents	Cyanobacteria/algae	Fluorescence
Cholesterol	<i>Nocardia erythropolis</i>	Oxygen electrode
Cyanide	<i>Pseudomonas fluorescens</i>	Oxygen electrode
	<i>Saccharomyces cerevisiae</i>	Oxygen electrode
Ethanol	<i>Trichosporon brassicae</i>	Oxygen electrode
	<i>Saccharomyces cerevisiae</i>	Mediated amperometric
	<i>Saccharomyces ellipsoideus</i>	Oxygen electrode
	<i>Gluconobacter oxydans</i>	Mediated amperometric
Glucose	<i>Saccharomyces cerevisiae</i>	Oxygen electrode, potentiometric
	<i>Pseudomonas fluorescens</i>	Oxygen electrode
	<i>Gluconobacter oxydans</i>	Mediated amperometric
Glutamine	<i>Sarcina flava</i>	Potentiometric
Heavy metals	<i>Escherichia coli</i> (lacZ, zntA, ECFP)	Fluorescence
	<i>Staphylococcus aureus</i> (pT0024)	Luminescence
	<i>Bacillus subtilis</i> (pT0024)	Luminescence
Herbicides	<i>Synechococcus</i> sp.	Mediated amperometric
	<i>Chlorella vulgaris</i>	Fluorescence
L-Lactate	<i>Hansenula anomala</i>	Mediated amperometric, dissolved oxygen
Methane	<i>Methylamonas flagellata</i>	Oxygen electrode
Mutagens	<i>Bacillus subtilis</i>	Oxygen electrode
NAD(P)H	<i>Escherichia coli</i>	Oxygen electrode
Naphthalene	<i>Pseudomonas putida</i>	Luminescence
Nitrate	<i>Azotobacter vinelandii</i>	Potentiometric
p-Nitrophenol	<i>Arthrobacter</i> sp.	Oxygen electrode
	<i>Moraxella</i> sp.	Oxygen electrode
Nystatin	<i>Saccharomyces cerevisiae</i>	Oxygen electrode, potentiometric
Phenol	<i>Pseudomonas putida</i>	Oxygen electrode
	<i>Pseudomonas</i>	Mediated amperometric
Tryptophan	<i>Escherichia coli</i> W2	Potentiometric
Urea	<i>Proteus mirabilis</i>	Potentiometric
Uric acid	<i>Pichia membranaefaciens</i>	Potentiometric
Volatile toxics	<i>Klebsormidium</i>	Fluorescence
Xylose	<i>Gluconobacter oxydans</i>	Oxygen electrode, potentiometric

<sup>a</sup> BOD, biological oxygen demand.

respond to several compounds limiting their application for specific analyte detection. An *Acetobacter pasteurianus* based oxygen electrode has been used for the determination of lactic acid, and shows good stability for 6 days, with a linear response range of  $1 \times 10^{-4}$ – $1.5 \times 10^{-3}$  mol dm<sup>-3</sup>. Whilst the sensor was insensitive to lactose, sucrose, and glucose, it responded with similar sensitivity to pyruvic acid,

acetaldehyde, ethanol, ammonium acetate, and sodium acetate as well as to L- and D-lactic acids.

Enhanced selectivity and sensitivity has been achieved in two ways: (1) induction of the desired metabolic or membrane transport systems, or (2) the inhibition or suppression of undesired systems. The former method has been successfully used for phenolics (*Trichosporon cutaneum*), tyrosine

(*Aeromonas phenologenes*), and lactate (*Hansenula anomala*). The inhibition of systems has proven successful in pyruvate monitoring (*Streptococcus faecium*) by inhibiting glycolysis with iodoacetamide and tyrosine decarboxylase with tyramine, and in microbial detection of glutamic acid (*Bacillus subtilis*) treated with chloromercuribenzoate and sodium fluoride to inhibit glucose uptake and metabolism.

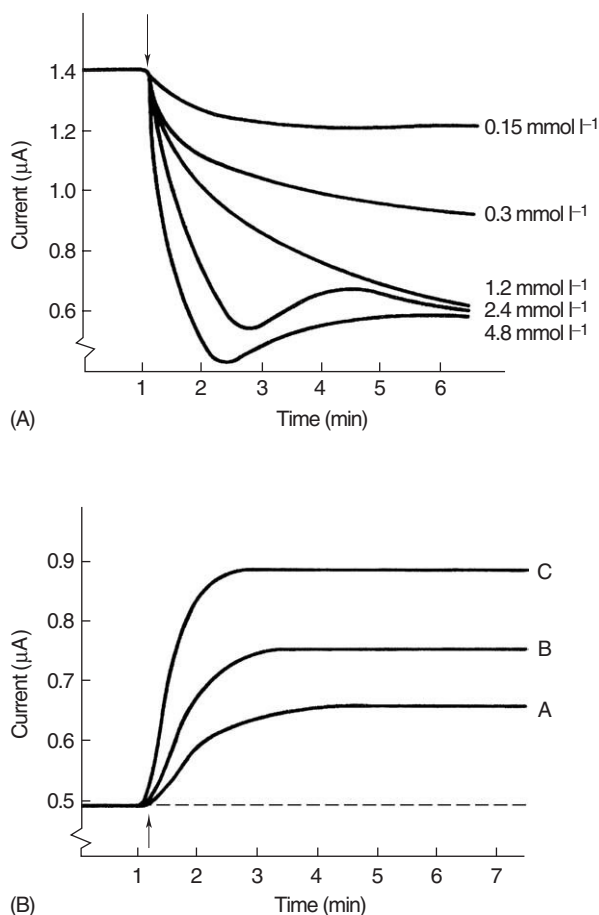
In comparison to enzyme-based biosensors, microbial biosensors show lower analyte selectivity, slightly slower response times, but often much better stability. Microbial biosensor determination of analytes such as amino acids, alcohol, and lactate show sensor stability over several days, whilst enzyme-based sensors may have operational lives of only 2–24 h.

### Broad Spectrum Monitoring

**Substrates** Monitoring total metabolic substances such as assimilable sugars or detecting the appearance/presence of potentially toxic compounds are applications ideally suited to microbial biosensors. In the case of substrates, *Brevibacterium lactofermentum* loaded oxygen electrodes have been operated for 10 days and up to a 1000 assays monitoring total glucose ( $<1 \text{ mmol l}^{-1}$ ), fructose ( $<1 \text{ mmol l}^{-1}$ ), and sucrose ( $<0.8 \text{ mmol l}^{-1}$ ) profiles in fermentation broth during glutamic acid production. In such situations the microbial biocatalyst mimics the fermenter organism and responds in much the same way to changes in substrate profiles. Typical responses obtained with bacteria loaded oxygen electrodes to a range of glucose concentrations are shown in Figure 3A.

**Toxicants** The use of microbial biosensors for detection of toxic or potentially toxic compounds is a recent development and one where the broad sensitivity of the microbial biocatalyst is vital. As toxicity is a biological phenomenon there is a need for intact organisms to be used in any such detection/assessment. In monitoring for toxicants there are positive advantages in following total metabolic activity or events that have an integrating effect on the metabolic events within the cell. Monitoring respiratory or photosynthetic activity for signs of perturbation has been used successfully to detect the presence of toxicants, using both mediated amperometric and oxygen electrode systems (Figure 3B).

Mediated amperometric biosensors incorporating bacterial and cyanobacterial biocatalysts have been shown to be sensitive to xenobiotics at environmentally relevant concentrations, enabling the biosensors to be used both for detecting toxic pollutants and



**Figure 3** (A) Typical current–time response curves of a bacterial oxygen electrode for a range of glucose concentrations. (↓) Glucose additions to the buffered minimal bathing medium. (B) Idealized current–time response curves for a bacterial oxygen electrode following the addition (↑) of toxicant solutions of different concentrations ( $A < B < C$ ) to the substrate supplemented bathing medium.

also for toxicity assessment. Sensitivity of detection increases with the toxicity of the analyte as shown by the inhibition of response of the mediated *Escherichia coli* biosensor to chlorinated phenols and biocides (Table 2). The magnitude of response to a single toxicant is determined by its concentration, as shown in the response to hypochlorite exhibited by biosensors incorporating a microbial consortium from activated sludge (Figure 4).

The hexacyanoferrate(III) mediator used in both examples accesses dehydrogenase enzymes on the bacterial outer membrane, and the biosensor signal reflects the concentration of reductant of the cells. Stimulation of the biocatalyst was by a cocktail of substrates. Both systems have operational lives of 3–5 days, being capable of continuous online applications, and response times of 15–30 min.

Toxicity screening by detection of respiration inhibition in *E. coli* has also been achieved by

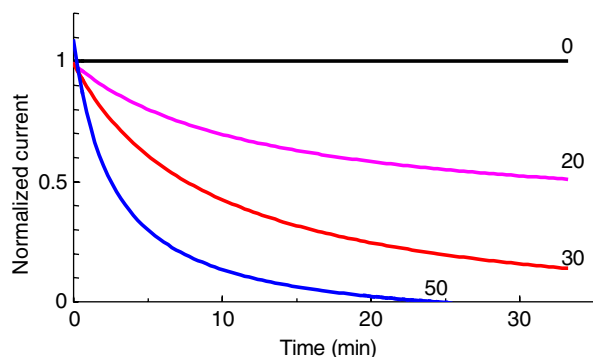
**Table 2** Sensitivity of *E. coli* to chlorophenols and biocides

Toxicant	EC <sub>50</sub> <sup>a</sup> (mg l <sup>-1</sup> )	Limit of detection (mg l <sup>-1</sup> )
Mercuric chloride	0.41	0.027
Phenol	> 50	> 50
2-Chlorophenol	> 50	> 50
2,5-Dichlorophenol	26	1.2
3,5-Dichlorophenol	14.5	1.0
2,4,5-Trichlorophenol	5.1	0.16
3,4,5-Trichlorophenol	6.5	0.25
2,3,5,6-Tetrachlorophenol	0.28	0.004
Pentachlorophenol	0.4	0.009
Dibutyltin dichloride	1.4	0.072
Tributyltin chloride	0.21	0.11
Ferbam	4.3	0.55
Thiram	—	0.48
Ziram	—	0.26
MCPA <sup>b</sup>	—	0.15
Formaldehyde	12.5 (EC <sub>25</sub> )	4.6
2,4-Dinitrophenol	1.65	0.95
Arochlor 1221	> 10	—

<sup>a</sup> Effective concentration resulting in a 50% decrease in biosensor response.

<sup>b</sup> 4-Chloro-2-methyl-phenoxyacetate.

—, Information not available.



**Figure 4** Normalized responses of mediated amperometric biosensors, incorporating a microbial consortium from activated sludge from a wastewater treatment plant, to increasing concentrations of hypochlorite. The figure shows the 'best curve fits' of three electrodes (one from six replicates at each hypochlorite concentration) exposed at time zero to 20, 30, and 50 ppm hypochlorite, compared to the normalized controls (0). The results from a full 32 electrode test gave effect concentrations for hypochlorite of EC<sub>10</sub> = 5.3 ppm; EC<sub>50</sub> = 25.5 ppm; and a 'no observed effect concentration' (NOEC) = 0.14 ppm.

incorporating the bacterium immobilized on filters on to a carbon dioxide electrode. Carbon dioxide production in glucose-enriched buffer in a stirred vessel quickly reaches a steady state, and the addition of a toxicant is easily detected by the resultant fall in carbon dioxide production within the bacterial layer. This electrode showed good stability, with constant response over 1 week, and the biosensor was still

functional after 6 weeks of storage in phosphate buffer at 5°C. Toxicant concentrations causing a 40% inhibition of respiration included hydrogen cyanide (3.5 mg l<sup>-1</sup>), cadmium (0.3 mg l<sup>-1</sup>), lead (0.13 mg l<sup>-1</sup>), and copper (1.1 mg l<sup>-1</sup>).

The use of microorganisms in mutagenicity tests suggests that microbial biosensors might also be capable of screening for mutagens. Revertant strains of *Salmonella typhimurium* and *E. coli*, and a Rec-strain of *Bacillus subtilis* have been used in combination with dissolved oxygen sensors to detect mutagens and carcinogens. Mutagen concentrations as low as 0.001 µg dm<sup>-3</sup> were achieved with an *S. typhimurium* sensor using a 10 h incubation period.

## Applications

Microorganism-based sensors bring the advantages of both classical bioassay techniques and sensor technology to analytical investigations, and in particular allow exploitation of the ability of microbial cells to respond to a wide range of metabolic substrates and inhibitors at low concentration in aqueous solution. Discrimination between different utilizable substrates in the same sample can be achieved but with difficulty. The use of microbial biosensors is better suited to applications where there are no such problems of interference or any need for high selectivity.

Where the determination of a single analyte or group of analytes is sought, it will be necessary to know the nature of possible interfering chemical species and the likelihood of any such material being present in the sample. The use of a microbial biosensor will be compromised if undesired components to which the biosensor will respond are expected in the sample, unless selectivity can be improved by pretreatment or conditioning of the biocatalyst.

## Process Control

Despite these potential problems, many microbial biocatalysts have been identified with sufficient selectivity and sensitivity to allow their use in single analyte determination. The low cost, stability, and ease of use makes such analyses attractive. The most commonly reported application for such sensors is process control, both in fermentation and downstream monitoring. A wide range of fermentation products have been monitored by microbial biosensors including alcohols, amino acids, antibiotics, organic acids, peptides, and vitamins.

Monitoring of fermentation substrate profiles is also possible with microbial biosensors, both for single analyte determination (e.g., glucose, acetic acid)

or for multianalyte determination such as total assimilable sugars. In the latter case, the broad spectrum response of the microbial biocatalyst is a distinct advantage, and bacterial biosensors are well suited to such profile monitoring. The high concentrations of substrates in fermenter liquor may require a dilution step before samples are presented to the biosensors. This dilution step has the advantage of facilitating modification of the small sample volumes drawn from the fermenter, allowing such steps as the adjustment of oxygen concentration, pH, salinity, and addition of mediator.

### Environmental Monitoring

The sensitivity of microorganisms to environmental pollutants makes them ideal for exploitation in biosensors for monitoring environmental quality. Monitoring both stimulation and inhibition of microbial metabolism allows their application in areas such as BOD determination, pollution incident detection in surface waters, measurement of effluent quality, and toxicity assessment. Biosensors incorporating living cells are unique amongst sensors in being able to detect analytes on the basis of their toxicity and assess toxicity levels; their use is a natural extension of microbial ecotoxicity testing. Similarly, biosensor determination of BOD is a logical development of the 5-day test using oxygen electrodes to monitor microbially seeded vessels. Along with the advantages of low cost and rapid test methods, biosensors also allow field testing and can be designed for online continuous monitoring. The most promising role for such biosensors in environmental monitoring is that of prescreening, with an alarm signal generated by the biosensors resulting in sample capture and offline analysis by more conventional technology.

See also: **Amino Acids. Bioassays:** Overview. **Chemical Warfare Agents. Ion-Selective Electrodes:** Overview.

**Process Analysis:** Bioprocess Analysis. **Sensors:** Amperometric Oxygen Sensors. **Water Analysis:** Biochemical Oxygen Demand.

### Further Reading

- Bentley A, Atkinson A, Jezek J, and Rawson DM (2001) Whole cell biosensors – electrochemical and optical approaches to ecotoxicity testing. *Toxicology in Vitro* 15: 469–475.
- Bousse L (1996) Whole cell biosensors. *Sensors and Actuators B* 34: 270–275.
- Chen J, Naglak T, and Wang HY (1992) An amperometric alcohol sensor on chemically permeabilized methylo-tropic microorganisms. *Biotechnology Progress* 8: 161–164.
- Hansen LH and Sorensen SJ (2001) The use of whole-cell biosensors to detect and quantify compounds or conditions affecting biological systems. *Microbial Ecology* 42: 483–494.
- Karube I (1987) Micro-organism based sensors. In: Turner APF, Karube I, and Wilson GS (eds.) *Biosensors: Fundamentals and Applications*, pp. 13–29. Oxford: Oxford Science Publications.
- Li Y and Chu J (1991) Study of BOD microbial sensors for waste water treatment control. *Applied Biochemistry and Biotechnology* 28/29: 855–863.
- Racek J (1995) *Cell-Based Biosensors*. Lancaster PA: Technomic.
- Riedel K, Renneberg R, Wollenberger U, Kaiser G, and Scheller FW (1989) Microbial sensors: fundamentals and applications for process control. *Journal of Chemical Technology and Biotechnology* 44: 85–106.
- Shah J and Wilkins E (2003) Electrochemical biosensors for detection of biological warfare agents. *Electroanalysis* 15: 157–167.
- Thevenot DR, Toth K, Durst RA, and Wilson GS (1999) Electrochemical biosensors: recommended definitions and classification. *Pure and Applied Chemistry* 71: 2333–2348.

## Photometric

**R Narayanaswamy**, UMIST, Manchester, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

A photometric (or optical) sensor can be described as a device which can be employed for the detection and determination of physical and chemical parameters through measurements of optical property. Optical methods have played a dominant role and continue

to do so in various fields of analytical sciences. Chemical and biochemical analytes have been qualitatively detected by spot tests using spectroscopic techniques including colorimetry and photometry. Such techniques, in conjunction with the use of dry reagent chemistries, have revolutionized the field of optical chemical sensing, particularly in the field of clinical analysis. A very familiar example of a photometric sensor is the pH indicator strip, which utilizes immobilized pH-sensitive reagents on cellulose for qualitative measurements of pH over a wide

or for multianalyte determination such as total assimilable sugars. In the latter case, the broad spectrum response of the microbial biocatalyst is a distinct advantage, and bacterial biosensors are well suited to such profile monitoring. The high concentrations of substrates in fermenter liquor may require a dilution step before samples are presented to the biosensors. This dilution step has the advantage of facilitating modification of the small sample volumes drawn from the fermenter, allowing such steps as the adjustment of oxygen concentration, pH, salinity, and addition of mediator.

### Environmental Monitoring

The sensitivity of microorganisms to environmental pollutants makes them ideal for exploitation in biosensors for monitoring environmental quality. Monitoring both stimulation and inhibition of microbial metabolism allows their application in areas such as BOD determination, pollution incident detection in surface waters, measurement of effluent quality, and toxicity assessment. Biosensors incorporating living cells are unique amongst sensors in being able to detect analytes on the basis of their toxicity and assess toxicity levels; their use is a natural extension of microbial ecotoxicity testing. Similarly, biosensor determination of BOD is a logical development of the 5-day test using oxygen electrodes to monitor microbially seeded vessels. Along with the advantages of low cost and rapid test methods, biosensors also allow field testing and can be designed for online continuous monitoring. The most promising role for such biosensors in environmental monitoring is that of prescreening, with an alarm signal generated by the biosensors resulting in sample capture and offline analysis by more conventional technology.

See also: **Amino Acids. Bioassays:** Overview. **Chemical Warfare Agents. Ion-Selective Electrodes:** Overview.

**Process Analysis:** Bioprocess Analysis. **Sensors:** Amperometric Oxygen Sensors. **Water Analysis:** Biochemical Oxygen Demand.

### Further Reading

- Bentley A, Atkinson A, Jezek J, and Rawson DM (2001) Whole cell biosensors – electrochemical and optical approaches to ecotoxicity testing. *Toxicology in Vitro* 15: 469–475.
- Bousse L (1996) Whole cell biosensors. *Sensors and Actuators B* 34: 270–275.
- Chen J, Naglak T, and Wang HY (1992) An amperometric alcohol sensor on chemically permeabilized methylo-tropic microorganisms. *Biotechnology Progress* 8: 161–164.
- Hansen LH and Sorensen SJ (2001) The use of whole-cell biosensors to detect and quantify compounds or conditions affecting biological systems. *Microbial Ecology* 42: 483–494.
- Karube I (1987) Micro-organism based sensors. In: Turner APF, Karube I, and Wilson GS (eds.) *Biosensors: Fundamentals and Applications*, pp. 13–29. Oxford: Oxford Science Publications.
- Li Y and Chu J (1991) Study of BOD microbial sensors for waste water treatment control. *Applied Biochemistry and Biotechnology* 28/29: 855–863.
- Racek J (1995) *Cell-Based Biosensors*. Lancaster PA: Technomic.
- Riedel K, Renneberg R, Wollenberger U, Kaiser G, and Scheller FW (1989) Microbial sensors: fundamentals and applications for process control. *Journal of Chemical Technology and Biotechnology* 44: 85–106.
- Shah J and Wilkins E (2003) Electrochemical biosensors for detection of biological warfare agents. *Electroanalysis* 15: 157–167.
- Thevenot DR, Toth K, Durst RA, and Wilson GS (1999) Electrochemical biosensors: recommended definitions and classification. *Pure and Applied Chemistry* 71: 2333–2348.

## Photometric

R Narayanaswamy, UMIST, Manchester, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

A photometric (or optical) sensor can be described as a device which can be employed for the detection and determination of physical and chemical parameters through measurements of optical property. Optical methods have played a dominant role and continue

to do so in various fields of analytical sciences. Chemical and biochemical analytes have been qualitatively detected by spot tests using spectroscopic techniques including colorimetry and photometry. Such techniques, in conjunction with the use of dry reagent chemistries, have revolutionized the field of optical chemical sensing, particularly in the field of clinical analysis. A very familiar example of a photometric sensor is the pH indicator strip, which utilizes immobilized pH-sensitive reagents on cellulose for qualitative measurements of pH over a wide



range. A major breakthrough was achieved when conventional spectroscopic analytical techniques were interfaced with optical fibers, which are offshoots of telecommunication industry. Optical fibers allow transmission of light over great distances and for chemical sensing typical monitoring distances required range from 1 to 100 m. In addition availability of low-cost optoelectronic components such as light-emitting diodes (LED), laser diodes, photodetectors, etc., and of rapid data acquisition and data processing techniques including chemometrics, has made considerable impact on the growth of and interest in optical chemical sensors over the past two decades.

Photometric chemical sensors based on optical fibers offer several advantages in various fields of application. These sensors are capable of observing a sample in its dynamic environment, no matter how distant, difficult to reach, or hostile this environment is. These devices are (1) intrinsically safe, involving a low optical power and are nonelectrical at the sensing point, (2) electrically passive and immune to electromagnetic disturbances, are geometrically flexible and corrosion resistant, (3) capable of being miniaturized, and (4) compatible with telemetry. Furthermore, the possibility of multiplexing several sensors to a single instrumentation unit can afford substantial economic advantages to this type of chemical sensor system. These features impart to the sensor immense potential importance in biomedical, process, and environmental monitoring applications.

Optical sensors possess more than a few limitations, which may need to be overcome if they are to replace existing sensor devices such as electrochemical sensors. The equilibrium involved in the chemical transduction restricts the concentration range over which the photometric sensors will be effectively useful. The dynamic range of optical sensors that have been studied to date extend over one or two orders of magnitude and, thus, a series of chemical transducers will be required in order to sense a single analyte at widely different levels of concentration. In such cases, a multisensor array system could be employed to widen the measurement dynamic range. In some chemical transduction systems, the equilibrium involved limits the reversibility characteristics of optical sensors. Ideally, these sensors should be capable of continuous use. However, in systems where a reaction is involved, the equilibrium state favors the formation of products which results in the consumption of the reagent by its conversion to some other form; in some cases the reagent can be regenerated using another chemical reaction. Another limitation is the response time of

spectroscopic sensors. Rates of chemical reactions and of diffusion of analytes and other materials determine the time taken for the attainment of a steady state in the chemical transduction system. Again the membrane, normally used as an envelope for the reagent phase, offers a barrier to mass transfer and contributes to response time. The lifetime of an optical sensor is determined by the stability of the reagent phases utilized in the chemical transducer particularly the photodecomposition characteristics of some of the reagents.

This article will review the practical methodology as applied to photometric sensors where measurements are confined to the ultraviolet and visible region of the electromagnetic spectrum. A previous article in this series on 'Sensors' has similarly discussed fluorescence measurement based sensors.

## Spectroscopic Techniques

In a photometric (optical) sensor, the optical property measured can be absorbance, reflectance, scattering, or emission. The first three of these methods of measurements are discussed in the following sections.

### Absorbance

The change in light intensity in an absorbance measurement is determined by the number of absorbing species in the optical path and is related to the concentration ( $C$ ) of the absorbing species via the Beer-Lambert relationship

$$A = \log_{10} \frac{I_0}{I} = \epsilon b C \quad [1]$$

where  $A$  is the absorbance,  $b$  the path length of light,  $\epsilon$  the molar absorptivity of the species, and  $I_0$  and  $I$  the intensities of the incident and transmitted light, respectively. When employing this technique with photometric sensors, the medium supporting the selective chemistry, and indeed the selective chemistry itself, should be optically transparent. The reagent and the supporting medium together are commonly referred to as the chemical transducer.

### Reflectance

When the chemical transduction medium used in a photometric sensor is opaque or when it transmits light only weakly, then measurements of the reflected light may be used. Reflection takes place when light infringes on a boundary surface, and two distinct types of reflection are possible. The first is the 'specular' (or mirror type) reflection, which occurs at the interface of a medium with no transmission

through it. The second type is the 'diffuse' reflection, where the light penetrates the medium and subsequently reappears at the surface after partial absorption and multiple scattering within the medium. Specular reflection can be minimized or eliminated through proper sample preparation and optical engineering.

The optical measurements of diffuse reflectance are dependent on the composition of the system. Several theoretical models have been proposed for diffuse reflectance, which are based on the radiative transfer theory, and all models consider that the incident light is scattered by particles within the medium. The most widely used theory in photometric sensors is the Kubelka–Munk theory, in which it is assumed that the scattering layer is infinitively thick, which may, in practice, be the case with the chemical transducers utilized in photometric sensors. The absolute value of the reflectance  $R$  is related to the absorption coefficient  $K$  and the scattering coefficient  $S$  by the equation

$$F(R) = \frac{(1 - R)^2}{2R} = \frac{K}{S} \quad [2]$$

where  $F(R)$  is known as the Kubelka–Munk function.  $K$  can be expressed in terms of the molar absorptivity  $\varepsilon$  and the concentration  $C$  of the absorbing species, as  $K = \varepsilon C$ . The above equation then becomes

$$F(R) = \frac{\varepsilon C}{S} = kC \quad [3]$$

where  $S$  is assumed to be independent of concentration. Equation [3] is analogous to the Beer–Lambert relationship (eqn [1]) and holds true within a range of concentrations for solid solutions in which the absorber is incorporated with the scattering particles and for systems in which the absorber is adsorbed on the surface of a scattering particle. The reflectance values ( $R$ ) are generally evaluated relative to the reflectance of standard reference materials such as barium sulfate.

Alternative theoretical models for reflectance have been formulated which are found to be valid for other types of systems. For example, the Pitts–Giovannelli theory is valid for scattering particles suspended in an absorbing aqueous medium, while the Rozenberg theory is applicable to mixtures of absorbing and nonabsorbing powders.

### Scattering

Unlike the processes of absorption and luminescence emission, scattering of light need not involve an electronic transition between quantized energy levels in atoms or molecules. Instead, a randomization in the

direction of light radiation occurs. Particles that are small when compared to the wavelength of optical radiation give rise to Rayleigh scattering, which is exhibited by all atoms and molecules. If the particles are larger than the wavelength of optical radiation, the scattering of light is called Mie scattering, in which the intensity of scattered radiation can be related to the concentration of the scattering particles.

In both Rayleigh and Mie scattering, the polarization of the particle remains constant. However, with molecules, the incident radiation can promote vibrational changes, which can alter the polarization of the molecule. The frequency of the light scattered by these molecules will be different from that of the incident light, and with much weaker intensity. Such a scattering phenomenon is known as Raman scattering, which can be observed if intense light sources (e.g., lasers) are used, and have also been used in photometric sensor development; discussion of this phenomenon is beyond the scope of this article.

### Evanescent Wave

When light is transmitted through an optical fiber or a waveguide by total internal reflection, a small portion of the light extends out of the central guide, which is referred to as the evanescent wave. In an optical fiber, the evanescent wave penetrates the cladding material. The intensity ( $I$ ) of the evanescent wave diminishes exponentially with increasing distance ( $z$ ) from the interface according to the equation

$$I = I_0 e^{-z/d_p} \quad [4]$$

where  $I_0$  is the electric field intensity at the interface and  $d_p$ , the depth of penetration, is the distance from the interface at which point the electric field intensity has diminished to  $1/e$  of its value at the interface. This characteristic depth of penetration is given by the formula

$$d_p = \frac{\lambda_0}{2\pi \sqrt{(n_1^2 \sin^2 \theta - n_2^2)}} \quad [5]$$

where  $\lambda_0$  is the wavelength of the propagating light,  $n_1$  and  $n_2$  are the refractive indices of the optically dense core and the surrounding medium, respectively, and  $\theta$  is the acceptance cone half-angle of the light entering the core. The distance  $d_p$  is typically of the order of a fraction of a wavelength but increases to infinity as  $\theta$  approaches the critical angle.

The existence of an evanescent wave is important for optical sensors, because devices in which the cladding material of an optical fiber has been removed and replaced by a thin layer of the reagent phase can be fabricated. Here, the optical properties

of the reagent phase can be monitored by means of the evanescent wave and the interaction between the analyte and reagent establishes a rapid equilibrium resulting in a fast response. It is also possible to couple light from one optical fiber to another using the evanescent wave by bringing the cores of the two fibers closer together than the penetration depth of the evanescent field. Consequently optical information can pass from one fiber to another. Surface plasmon resonance (SPR)-based sensors that utilize evanescent wave interaction of light with a metallic layer, usually silver, may also be considered as photometric sensors.

## Chemical Transduction

In photometric sensors, the chemical transducer usually consists of an immobilized reagent (or a reagent phase) which changes its optical property in the presence of an analyte. Reagents for a particular sensing application are often chosen from the analogous solution chemistry. A reagent phase in a photometric sensor may well be a solution of reagent itself whose optical property is monitored by conventional photometry. But for a reusable sensor, the reagent needs to be immobilized on to an inert and stable solid support, e.g., glass, silica gel, organic polymers, etc., by employing physical or chemical procedures. Physical methods of immobilization, which are simple and economical to carry out, include (1) entrapment, in which the reagent is immobilized within the interstitial spaces of a polymeric lattice; (2) surface adsorption, which involves a weak attraction (Van der Waals force or hydrogen bond) between the reagent molecules and the surface of the solid support; and (3) electrostatic attraction, which involves the formation of ion pairs between the charged reagent and the ionic centers of the solid support. Chemical immobilization method is based on the formation of a covalent bond between the reagent molecule and an activated or functionalized form of the polymeric solid support. This method produces the most firmly bound reagent, but several chemical reaction steps are usually required in the synthesis of a bondable reagent phase.

Like any other type of sensor, the ideal photometric chemical sensor should not perturb the sample. In practice, this requires that the amount of analyte present to react with the indicator reagent should be small compared to the total amount of analyte in the sample. If this requirement is not met, then the contact of analyte with the reagent in the chemical transducer will result in a change in the analyte concentration in the sample. This, in turn, will result in a

slow response of the sensor, in addition to the perturbation in the sample. The slow response arises due to the fact that the time required for sufficient analyte to diffuse and to come into contact with the indicator reagent, may be quite long. Such problems are likely to occur with (1) small sample concentration or mass, or (2) sensors with large indicator reagent concentration or mass.

## Transducer Chemistry

Some analyte chemical species are themselves capable of exhibiting optical changes related to their concentrations, which can be measured photometrically using optical fibers. Such sensors, in which separate chemical transduction systems will not be required, are commonly referred to as 'plain-fiber' sensors. Since many analytes do not themselves possess suitable optical characteristics, they may be sensed indirectly through their interaction with an appropriate chemical transduction system. The sensor response function in such chemical transducer-based optical sensors depends on the manner in which the analyte interacts with the reagent phase. For example, in a simple system where a reagent R reacts with an analyte species A forming a product AR can be represented by the reaction



where either R or AR exhibits measurable optical characteristics. The reagent R is usually selective for the particular analyte being sensed. The chemical transduction depends on the equilibrium of the above reaction, and can additionally be described by the following equation

$$K = \frac{[AR]}{[A][R]} \quad [6]$$

where  $K$  is the equilibrium constant of the transduction reaction; the square brackets indicate equilibrium concentration of the species involved.

During the chemical reaction, the reagent R can be consumed so that the measured optical signal due to it can be decreased, or the product AR can be formed when the optical signal would increase due to its formation. In either case, the change in the optical signal measured can be related to the concentration of the analyte A, which causes this change. If  $C_r$  is the total initial concentration of the reagent, then at any time during the reaction

$$C_r = [R] + [AR] \quad [7]$$

Thus, from eqns [6] and [7], the concentration of A can be related to that of R or AR by the following

equations:

$$[A] = \frac{1}{K} \left( \frac{C_r}{[R]} - 1 \right) \quad [8]$$

$$\frac{1}{[A]} = K \left( \frac{C_r}{[AR]} - 1 \right) \quad [9]$$

These equations illustrate that the reaction [I] is a reversible type, which provides sensing with a direct indicator and requires an appropriate equilibrium constant for the desired range of analyte concentration. The sensor response also depends on  $C_r$ . Furthermore, any uncontrolled variable that affects  $K$  will be a source of interference in the measurement, for example, ionic strength of the medium or temperature.

The most prominent example of the use of direct indicators can be found in optical pH sensors. Because of the multitude of pH indicators (also called acid–base indicators) available with different  $K$  values, it would not be difficult to find a pH indicator to cover virtually any desired pH range. Metal ion sensors based on direct indicators have been studied but they tend to be limited by the fact, in many cases, that the equilibrium constant depends on pH. Sensors for ammonia and carbon dioxide also involve the use of direct indicators. The chemical transducers in these sensor systems consist of suitable buffer and pH indicator enveloped by a gas permeable membrane.

If the measured optical parameter is dependent on the concentration ratio  $[AR]/[R]$ , then the sensor response is no longer dependent on  $C_r$ , but dependent on  $K$ . Thus

$$[A] = \frac{1}{K} \left( \frac{[AR]}{[R]} \right) \quad [10]$$

In general, in photometric sensors, reversible reactions are preferred in the chemical transduction systems because they can provide continuous and unperturbed measurements. The response time (i.e., the time to reach the equilibrium) is dependent on mass transfer. Reversible interactions of analytes with reagents, which do not involve chemical reactions, such as the dynamic fluorescence quenching by oxygen, can provide sensors with short response times. This is because of the fact that the fluorescence quenching effect takes place via the transfer of energy from the reagent to the analyte molecules through contact, and there is no consumption of analyte or reagent during this process.

Nonreversible chemical reactions may be employed in the photometric transduction systems, which can result in ‘one shot’ type sensor devices. Though there will only be a limited merit in utilizing

such reactions in sensors, they can provide high sensitivity for measurements. The reagent phases in these sensors can be formulated as flat strips or slides, so that high reproducibility in measurements may be achieved from sensor to sensor. In many cases, the reagents can be regenerated by the use of another chemical reaction and reused for further measurements. For example, the sensor for lead based on immobilized dithizone can be reversed using hydrochloric acid solutions.

Indirect chemical reactions involving two or more reagents and/or reactions can be used as optical chemical transducers. For example, consider the following reactions:



where  $X$  is the analyte,  $M$  the immobilized reagent, and  $S$  the species which competes with  $X$  for  $M$ . In the sensor, the optical property of free  $S$  is measured. In the absence of  $X$ , the reaction [II] will proceed to the right and the optical signal measured will be reduced due to increasing amount of  $S$  being bound to  $M$ . However, as the concentration of  $X$  in the sample increases, the reaction [III] will proceed to the right and that in reaction [II] to the left, thus resulting in free  $S$ . The optical signal measured due to free  $S$  will then be proportional to the concentration of  $X$ . This concept has been employed in the development of a glucose sensor. In this system,  $X$  was glucose,  $S$  was fluorescein-labeled dextran, and  $M$  was immobilized concanavalin A, a carbohydrate binding protein. The sensor response to glucose was detected by the measurement of the fluorescence of fluorescein-labeled dextran released in the optical path. In general, the use of indirect indicators in optical transducers offer the potential of varying the effective equilibrium constant which, in turn, determines the range of concentrations of analyte to which the sensor responds; this is not possible with direct indicators.

Additionally, phenomena such as catalysis, ion-pair interactions, chemi- or bioluminescence, redox changes, etc., may be employed in the chemical transduction process. In catalysis type of system, for example, the immobilized reagent catalyzes the conversion of analyte to form a product whose optical property is then measured. Several examples of such reactions include those which utilize immobilized enzymes acting as a catalyst. The response of the sensor in such a system, involves a steady state, in which the rate of product formation is equal to the rate at which it is removed from the transducer by processes such as diffusion and convection. The concentration of the steady-state product increases with

that of the analyte. Thus, this type of transduction system responds continuously but, with respect to mass transfer, controlled reaction conditions will be required.

### Transducer Design

A variety of transducer configurations that has been employed in photometric sensor devices fall into two sensor types: extrinsic sensors and intrinsic sensors. While in the former sensor type the optical fiber merely acts as a light guide, conveying the optical information between the optical source and the chemical transducer and between the chemical transducer and the detector, in the latter sensor type the optical fiber, probably in some modified form, would become a part of the transducer.

One of the critical elements in the design of a photometric sensor is the configuration of the sensing terminus. Its configuration often determines detectability, linear dynamic range, rate of response, the effects of interferences, site specificity, and useful lifetime of the sensor. Most important design considerations include (1) the method of immobilization and (2) the shape and size of the terminus and the amount of reagent phase.

According to the working principle, both extrinsic and intrinsic sensor types can be differentiated by their sensing schemes. As mentioned earlier, optical fibers have been employed in photometric sensors either as plain-fiber sensors or transducer-based sensors. In the former, the analyte species possesses some measurable optical property, which is measured and related to its concentration. The transducer-based optical sensor is the most common type of the two. The interaction between the light and the analyte/reagent system takes place at one region of the optical fiber and a variety of sensor configurations can be grouped into (1) transmissive configuration, in which the interaction occurs along the length of the fiber, and (2) nontransmissive configuration, where the interaction takes place at a distal end of the optical fiber system.

Some of the sensor configurations studied to date and commonly employed in both extrinsic and intrinsic types of photometric sensors are shown in **Figures 1A–1J**. Light is launched at one end of the optical fiber and is detected at the other end in the transmissive type of configuration with the liquid or gaseous analyte passing through a flow-cell (**Figure 1A**) or a gap (**Figure 1B**) interposed along the optical fiber, or with the analyte surrounding the cladding-removed fiber in a flow-cell (**Figure 1C**). On the other hand, the cladding of the optical fiber may be replaced with a chemical transducer (**Figure 1D**) whose

change in the optical property is measured as a function of the analyte concentration. In those cases where cladding-removed optical fibers are employed (**Figures 1C** and **1D**), optical signals are measured through the evanescent wave interaction of the light with the analyte or the reagent material.

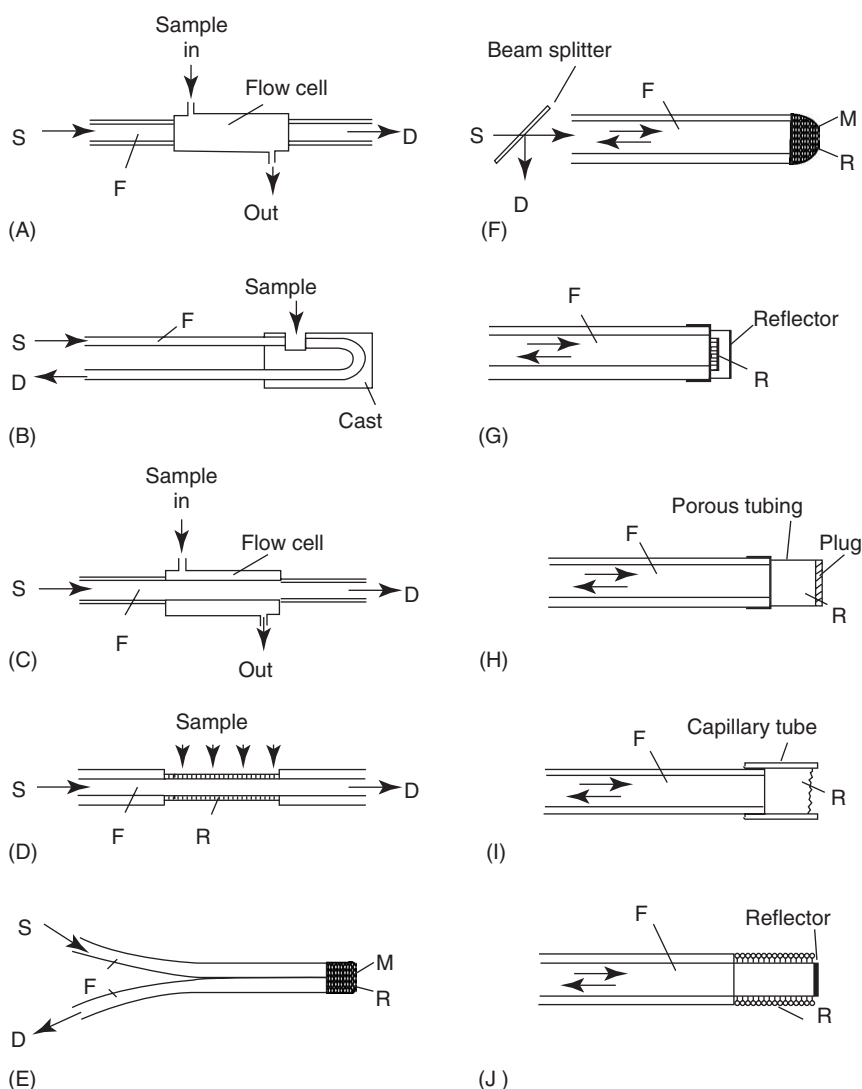
The nontransmissive type of configuration employed in photometric sensors are depicted in **Figures 1E–1J**, in which the incident light and the detected light are transmitted along the same or different optical fibers, and the sample system is encountered at a terminus of the optical fiber system. When single optical fibers are used in sensors (**Figures 1F–1J**), the light, supplied and collected by the same fiber, is discriminated either temporally or by wavelength with the aid of a beam splitter. However, in a bifurcated optical fiber system (**Figure 1E**), the incident and detected light travel along different fibers, and such a system can be fabricated using two optical fibers or an optical fiber bundle. In the latter, the light supply fibers and the detecting fibers are randomly distributed at the terminus of the fiber system. In general, sensors based on the use of single optical fibers are simpler to construct and the single fiber systems exhibit a greater field of view compared with the bifurcated optical fiber system, which also involves a blind volume in the field (**Figure 2**).

The reagent phases in chemical transducer-based sensors are normally fabricated by using physical or chemical methods, as described earlier in this article, and these are interfaced appropriately with optical fibers in order to produce an optimal sensor design with respect to a particular type of application. In these sensors aside from the chemical transduction process that recognizes the analyte interaction, the optical signals transmitted to the detector also need to be transduced further into a measurable electrical signal that can be displayed, usually, by a meter.

### Instrumentation

The basic instrumentation associated with photometric sensors is simple and requires both optical and electrical components. Apart from the optical fiber, the instrumentation system involves a light source, photodetector and associated display, optical components such as lenses, couplers and connectors, and monochromators or filters for wavelength selection. A typical instrumentation system employed in conjunction with photometric sensors is shown schematically in **Figure 3**.

The light source must be able to provide a stable and intense optical radiation. Several types of sources have been employed in photometric sensors, namely incandescent lamps, gas lasers, LEDs, and laser



**Figure 1** Transducer designs (see text for details). S, source; D, detector; F, optical fiber; R, reagent phase; M, membrane.

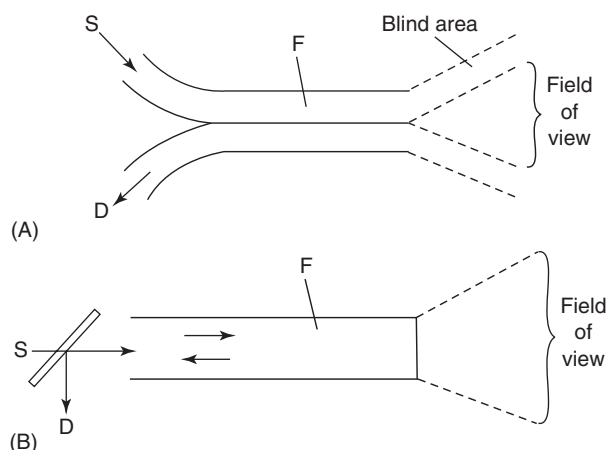
diodes. Incandescent sources, such as tungsten lamps and quartz-halogen lamps, emit a broad spectrum of optical radiation and are used as sources of ultraviolet and visible light in short-range optical sensors. Gas lasers are useful as general-purpose sources of highly intense and coherent radiation and are employed in long-distance or remote sensing systems. LEDs and laser diodes are miniature sources of high-intensity monochromatic radiation. LEDs produce incoherent light with a typical spectral bandwidth of 40–50 nm and are useful in short range sensing up to a distance of 1 km. Laser diodes, on the other hand, radiate a coherent beam of light with a narrower spectral bandwidth of the order of 5–10 nm or less and are excellent sources for remote sensing systems.

The detection utilized in the instrumentation is essentially a photon-counting device in which optical signals are converted into electrical signals which can

be easily amplified electronically. It is essential that the photodetector has peak sensitivity at the measurement wavelength, generate a minimal amount of noise with respect to the transmitted signal, and respond rapidly to variations in the intensity of light reaching it. Various types of photodetectors have been used in optical sensors including photomultiplier tubes, PIN photodiodes, photodiode arrays, and avalanche photodiodes.

In optical sensing system, it is necessary to employ optical couplers and lenses in order to focus the light beam to the optical fiber and direct radiation from the return fiber to the photodetector. Injecting light from a source into a fiber is a more tedious operation. Laser sources produce a coherent beam of light with a cross-section almost equal to the cross-sectional area of the optical fiber, and, therefore, can be coupled very efficiently to the fiber. On the



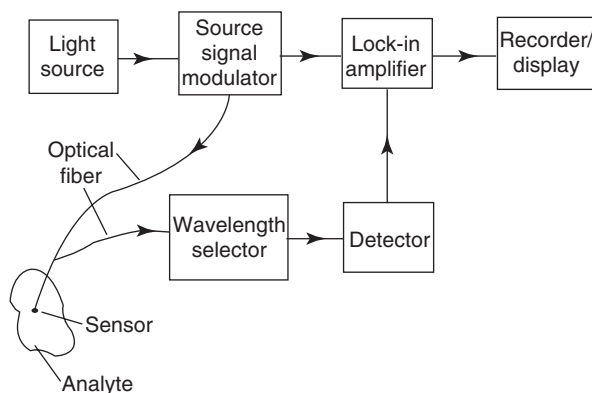


**Figure 2** Field of view with (A) bifurcated (or twin) optical fibers and (B) single optical fiber. S, source; D, detector; F, optical fiber.

other hand, incandescent lamps and LEDs radiate divergent light, which need to be focused using lenses in order to launch the light into the optical fiber.

The coupling of the photodetector to the optical fiber system is more easily accomplished due to the fact that the detectors have relatively large active surface areas and large acceptance angles. However, it will be necessary that wavelengths other than that of the analytical signal have to be excluded from the light that is reaching the detector. Such optical resolution needs to be carried out if broadband sources, e.g., incandescent lamps, are used in the optical system. Isolation of desired wavelengths is usually accomplished by the use of filters or monochromators. Simple optical filters attenuate significant amount of light and, thus, could lower the sensitivity of the system. Monochromators offer high efficiency of light throughput and can be adapted for different wavelengths.

Extraneous optical radiation can be excluded from reaching the detector by suitable modulation of the light source and by synchronizing the detector to this modulation frequency. Any instrumental drifts caused by the electronic components of the system, such as aging of the light source, fluctuations in the power supply, fouling, etc., may be eliminated by



**Figure 3** Schematic diagram of a typical instrumentation employed with photometric sensors.

employing internal referencing, which involves the monitoring of radiation that is not altered during the sensing process.

See also: **Sensors: Overview.** **Spectrophotometry: Overview.**

## Further Reading

- Harmer AL and Narayanaswamy R (1988) Spectroscopic and fibre-optic transducers. In: Edmunds TE (ed.) *Chemical Sensors*. London: Blackie.
- Narayanaswamy R (1991) Current developments in optical biochemical sensors. *Biosensors and Bioelectronics* 6: 467–475.
- Narayanaswamy R and Sevilla III FS (1988) Optical fibre sensors for chemical species. *Journal of Physics E: Scientific Instrumentation* 21: 10–17.
- Narayanaswamy R and Wolfbeis OS (2004) *Optical Sensors – Industrial, Environmental and Diagnostic Applications*. Berlin: Springer.
- Seitz WR (1988) Chemical sensors based on immobilised indicators and fiber optics. *CRC Critical Reviews in Analytical Chemistry* 19: 135–171.
- Wise DL and Wingard LB Jr. (eds.) (1991) *Biosensors with Fiberoptics*. Clifton, NJ: Humana Press.
- Wolfbeis OS (1991) *Fibre Optic Chemical Sensors and Biosensors*, vols. I and II. Boca Raton, FL: CRC Press.

## Piezoelectric Resonators

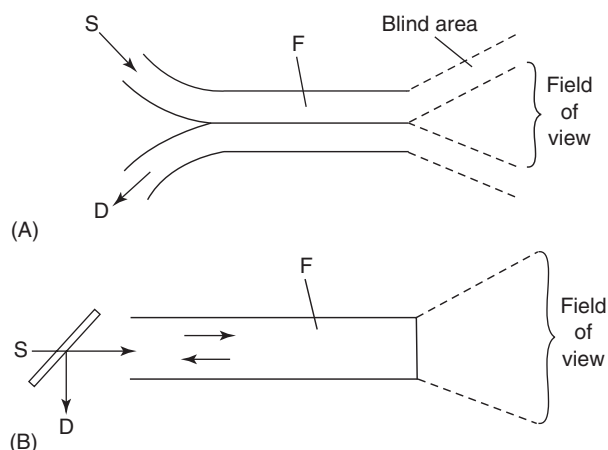
**C Steinem**, University of Regensburg, Regensburg, Germany

**A Janshoff**, Johannes-Gutenberg-Universitaet, Mainz, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Piezoelectric resonators are sensor devices that respond to changes in temperature, pressure, and, most importantly, to changes in the physical properties of

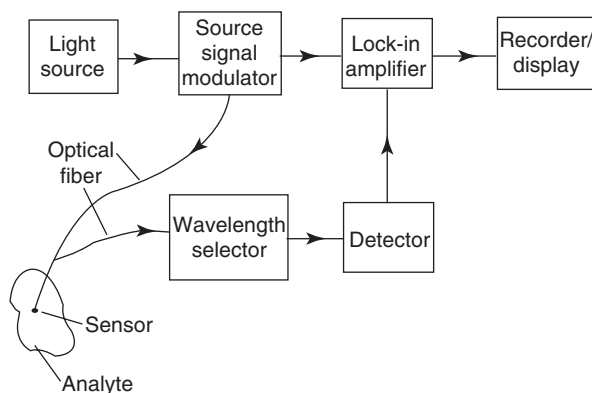


**Figure 2** Field of view with (A) bifurcated (or twin) optical fibers and (B) single optical fiber. S, source; D, detector; F, optical fiber.

other hand, incandescent lamps and LEDs radiate divergent light, which need to be focused using lenses in order to launch the light into the optical fiber.

The coupling of the photodetector to the optical fiber system is more easily accomplished due to the fact that the detectors have relatively large active surface areas and large acceptance angles. However, it will be necessary that wavelengths other than that of the analytical signal have to be excluded from the light that is reaching the detector. Such optical resolution needs to be carried out if broadband sources, e.g., incandescent lamps, are used in the optical system. Isolation of desired wavelengths is usually accomplished by the use of filters or monochromators. Simple optical filters attenuate significant amount of light and, thus, could lower the sensitivity of the system. Monochromators offer high efficiency of light throughput and can be adapted for different wavelengths.

Extraneous optical radiation can be excluded from reaching the detector by suitable modulation of the light source and by synchronizing the detector to this modulation frequency. Any instrumental drifts caused by the electronic components of the system, such as aging of the light source, fluctuations in the power supply, fouling, etc., may be eliminated by



**Figure 3** Schematic diagram of a typical instrumentation employed with photometric sensors.

employing internal referencing, which involves the monitoring of radiation that is not altered during the sensing process.

See also: **Sensors: Overview.** **Spectrophotometry: Overview.**

## Further Reading

- Harmer AL and Narayanaswamy R (1988) Spectroscopic and fibre-optic transducers. In: Edmunds TE (ed.) *Chemical Sensors*. London: Blackie.
- Narayanaswamy R (1991) Current developments in optical biochemical sensors. *Biosensors and Bioelectronics* 6: 467–475.
- Narayanaswamy R and Sevilla III FS (1988) Optical fibre sensors for chemical species. *Journal of Physics E: Scientific Instrumentation* 21: 10–17.
- Narayanaswamy R and Wolfbeis OS (2004) *Optical Sensors – Industrial, Environmental and Diagnostic Applications*. Berlin: Springer.
- Seitz WR (1988) Chemical sensors based on immobilised indicators and fiber optics. *CRC Critical Reviews in Analytical Chemistry* 19: 135–171.
- Wise DL and Wingard LB Jr. (eds.) (1991) *Biosensors with Fiberoptics*. Clifton, NJ: Humana Press.
- Wolfbeis OS (1991) *Fibre Optic Chemical Sensors and Biosensors*, vols. I and II. Boca Raton, FL: CRC Press.

## Piezoelectric Resonators

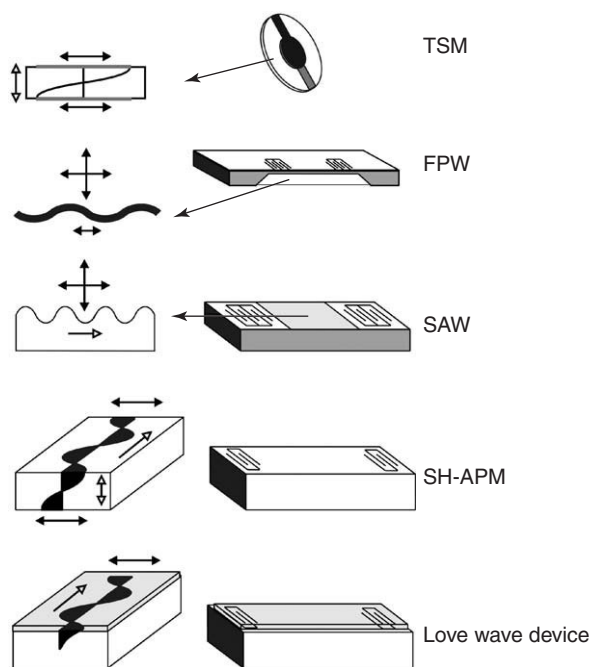
**C Steinem**, University of Regensburg, Regensburg, Germany

**A Janshoff**, Johannes-Gutenberg-Universitaet, Mainz, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Piezoelectric resonators are sensor devices that respond to changes in temperature, pressure, and, most importantly, to changes in the physical properties of



**Figure 1** Wave propagation modes of piezoelectric resonators. Closed arrows indicate particle displacement, open arrows direction of wave propagation. TSM, thickness shear mode; FPW, flexural plate wave; SAW, surface acoustic wave; SH-APM, shear horizontal acoustic plate mode. (Reprinted with permission from *Angewandte Chemie International Edition* (2000) 39: 4004–4032; © 2003 Wiley-VCH.)

the device surface's contacting materials. These physical properties include variations in interfacial mass density, layer thickness, viscosity, and viscoelasticity. The very sensitive and selective responses of the piezoelectric devices obtained by coating their surfaces with chemically or biologically selective films have been exploited by many researchers working in the field of analytical applications. This article will describe the basic principles behind piezoelectric resonators (Figure 1) and will review their major sensory application areas.

## Principles of Piezoelectric Sensors

### Piezoelectricity

Piezoelectricity (*piezin*, Greek, to press) was first described in 1880 by Pierre and Jacques Curie, who showed that upon mechanical deformation (torsion, pressure, bending, etc.) of a solid material along an appropriate direction, electrical charges occur on the material's opposing surfaces. Conversely, applying an external electric field to a material induces a mechanical deformation. This phenomenon is called the converse piezoelectric effect. Piezoelectricity can only occur in crystals with an inversion center and from a crystallographic viewpoint, 21 point groups fulfill

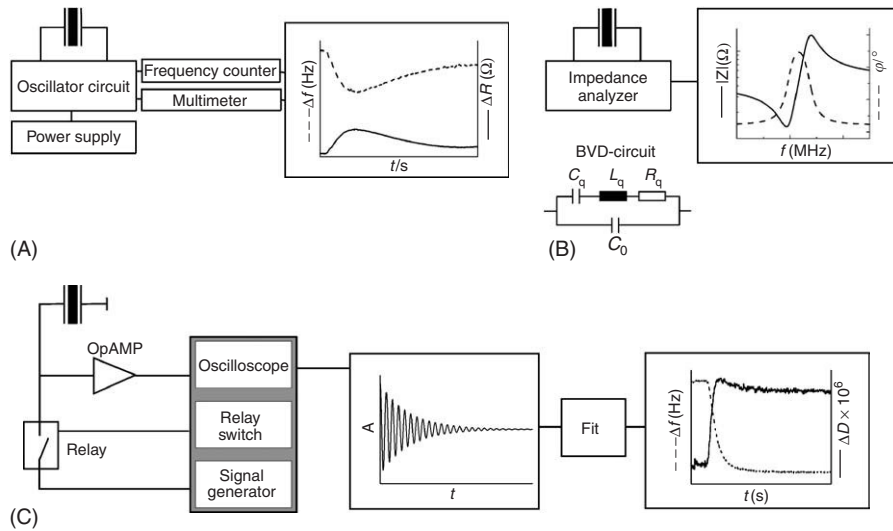
this requirement. However, only 20 point groups do have a nonzero piezoelectric constant. There are a large number of crystals commonly used as piezoelectric materials such as Rochelle salt, sodium chlorate, tourmaline, or quartz. Among them,  $\alpha$ -quartz ( $\text{SiO}_2$ ) is very unique, as it combines mechanical, electrical, chemical, and thermal properties, which has led to its commercial significance. The quartz crystal may provide a large variety of different resonator types depending on the cut-angle with respect to the crystal lattice. The cut-angle determines the mode of induced mechanical vibration. Resonators operating in thickness shear mode (TSM), face shear mode, or flexural mode can be obtained from the mother crystal with eigenfrequencies ranging from  $5 \times 10^2$  Hz to  $3 \times 10^8$  Hz.

### Thickness Shear Mode Sensors

For TSM resonators, which are also known as bulk acoustic wave sensors, AT-cut crystals are used that are prepared by slicing a quartz wafer at an angle of  $35.25^\circ$  to the optical  $z$ -axis. AT-cut quartz crystals are superior because they exhibit a tremendous frequency stability of  $\Delta f/f \approx 10^{-8}$  with a zero temperature coefficient between 0 and  $50^\circ\text{C}$ . Owing to the converse piezoelectric effect, the crystal plate can be excited to mechanical oscillation by applying an alternating potential difference via metal electrodes (commonly disk-like gold or silver electrodes) that are deposited on each side of the quartz plate. Shear waves of opposite polarity are generated at the electrodes on either side of the crystal with the shear displacement in plane with the crystal surface. Both waves traverse across the thickness of the quartz plate, are reflected at the opposing crystal face (phase shift  $\pi$ ), and then return to their origin. Constructive interference of incident and return waves occurs when the acoustic wavelength is an odd multiple of twice the crystal thickness ( $2d_q$ ) and standing wave conditions are then established. Particle displacement of the standing acoustic wave is maximal at the crystals' surfaces and a displacement node exists in the center of the resonator. Thus, the eigenfrequency of the TSM resonator is inversely proportional to its thickness  $d_q$  ( $f = nv_p/2d_q$ , where  $v_p$  is the phase velocity). For instance, a TSM resonator with a thickness of  $330 \mu\text{m}$  exhibits a resonance fundamental frequency  $f_0$  of 5 MHz with odd overtones  $n$  (15 MHz, 25 MHz, etc.). The lateral dimensions of the quartz plate depend on the mother crystal and are usually 10–25 mm in diameter.

### Detection principles

*Active oscillator mode* Stable oscillations of a quartz plate only occur at the resonance frequency



**Figure 2** Setups used for TSM resonators: (A) active mode, (B) impedance analysis of the oscillation (passive mode), and (C) ring down method (QCM-D<sup>TM</sup>).

of the crystal. If the crystal is incorporated into a feedback loop of an oscillator circuit, it becomes the frequency-determining element as its  $Q$ -factor is very large (20 000–50 000 for 10 MHz fundamental frequency in air). The frequency of the quartz plate can be read out by a frequency counter. Using an amplitude-controlled oscillator circuit that monitors the amplitude apart from the resonance frequency, damping of the quartz oscillation can also be recorded (Figure 2A). Self-sustaining oscillation of TSM resonators is, however, restricted to load situations with the phase maximum above zero.

**Passive oscillator mode** Impedance analysis of the forced oscillation of the quartz plate provides valuable information about the coating even if the active mode is not applicable anymore. For impedance analysis, a frequency generator is used to excite the crystal to a constraint vibration near resonance while monitoring the complex electrical impedance and admittance, respectively, dependent on the applied frequency (Figure 2B). For low load situations near resonance, an equivalent circuit with lumped elements – the so-called Butterworth–van-Dyke (BVD) circuit – can be applied to model the impedance data. The BVD circuit combines a parallel and series (motional branch) resonance circuit. The motional branch consists of an inductance  $L_q$ , a capacitance  $C_q$ , and a resistance  $R_q$ . An additional parallel capacitance  $C_0$  arises primarily from the presence of the dielectric quartz material between the two surface electrodes (parallel plate capacitor) also containing parasitic contributions of the wiring and the crystal holder (Figure 2B).

**QCM-D<sup>TM</sup> technique** Kasemo and co-workers have developed an interesting technique that measures the resonance frequency  $f$  and the dissipation factor  $D$ , which is the inverse of the  $Q$ -factor, of the oscillation simultaneously by a ring down method. The quartz plate is excited every second with a frequency generator followed by switching off the source and recording the free decay of the quartz oscillation. The dissipation factor and resonance frequency are obtained from each cycle by a curve fit of an exponentially damped harmonic oscillator function (Figure 2C).

**Mass sensitivity in the gas phase and in vacuum** The resonance frequency of an AT-cut quartz plate operating in shear mode is altered by adding mass to the surface of the crystal since this increases the thickness of the quartz plate. In 1959, Sauerbrey provided a mathematical treatment of the mass sensitivity of the quartz oscillation, which led to the term quartz crystal microbalance (QCM). He showed that a thin, rigid, and uniform film deposited on the electrode's surface results in a frequency decrease  $\Delta f$  that is proportional to the deposited mass  $\Delta m$  if the resonator is operated in the gas phase or in vacuum. Changes in the motional resistance ( $\Delta R$ ), i.e., damping, is zero. If the density of the mass layer is equal to that of the quartz crystal, the following relation applies:

$$\Delta f = -\frac{2nf_0^2}{A\sqrt{\mu_q\rho_q}}\Delta m = -S_f\Delta m \quad [1]$$

where  $A$  is the piezoelectric active electrode area (typically ranging between 0.2 and 0.4 cm<sup>2</sup>),

$\mu_q$  the piezoelectric stiffened shear modulus ( $2.96 \times 10^{10} \text{ N m}^{-2}$ ), and  $\rho_q$  the density of quartz ( $2.65 \text{ g cm}^{-3}$ ). The integral mass sensitivity  $S_i$ , also known as Sauerbrey constant, depends on the square of the fundamental frequency  $f_0$  and increases proportionally to the overtone number  $n$ . The sensitivity  $c_m^{\text{TSM}}$  of the TSM is defined as

$$c_m^{\text{TSM}} = \lim_{\Delta m/A \rightarrow 0} \frac{A}{\Delta m} \left( \frac{\Delta f}{f_0} \right) \approx - \frac{2f_0}{\sqrt{\mu_q \rho_q}} \quad [2]$$

A more detailed theoretical treatment of the propagating acoustic wave that solves the general wave equation of motion for the proper boundary conditions reveals that the shear amplitude along the crystal surface is not uniform but radially symmetric. This, in turn, means that the quartz resonator is not uniformly sensitive to the adsorption of a rigid foreign material. The amplitude is maximal in the center of a disk-like electrode and decreases monotonically with increasing distance from the center vanishing at the electrode edges. Empirically, the radial mass sensitivity can be described by a Gaussian function with its maximum sensitivity in the center of the electrode.

If the mass deposited on the electrodes is not rigid and measurements are performed in liquid phase, the TSM resonator does not behave like a simple mass sensor anymore, but provides valuable information about the viscous and viscoelastic properties of the deposited material. Many of the applications of the QCM, in which polymer films, biomolecules, or even whole cells are investigated, benefit from this additional information. In all these cases, the Sauerbrey equation does not apply strictly, and it might become impossible to extract corresponding mass binding quantities from merely recording changes in resonance frequency.

**Liquid loading** Most QCM applications require the sensor to work in a liquid environment. The coupling of the crystal surface to the liquid dramatically alters its resonance frequency as the shear motion of the surface generates motion in the liquid near the interface. Thus, a fraction of the kinetic energy of the oscillation is radiated into the bulk phase and is thereby dissipated. The frequency change upon contact of the resonator surface with a Newtonian liquid was first analytically described by Kanazawa and Gordon in 1985:

$$\Delta f = -f_0^{\frac{3}{2}} \sqrt{\frac{\eta_L \rho_L}{\pi \mu_q \rho_q}} \quad [3]$$

where  $\eta_L$  is the viscosity and  $\rho_L$  the density of the liquid. The characteristic decay length  $\delta$  of the shear wave within the bulk phase depends on the density

and viscosity of the liquid and is  $\sim 250 \text{ nm}$  (5 MHz TSM resonator) in water at  $20^\circ\text{C}$ .  $\Delta R$  can be expressed according to the acoustic load concept:

$$\Delta R = \frac{f_0^{3/2} L_q}{\sqrt{\mu_q \rho_q}} \sqrt{\eta_L \rho_L} \quad [4]$$

When using TSM resonators in liquids and in particular in aqueous solutions, other external parameters – besides the viscous load – affect the shear oscillation. These include the generation of longitudinal (compressional) waves, which are reflected at the liquid–air interface and alter the resonance frequency of the quartz plate. If working in an open crystal holder, periodical changes in the resonance frequency are recorded due to evaporation of the liquid. Other important parameters that influence the resonance frequency of the quartz crystal in solution are the ionic strength and dielectric constant of the electrolyte. If the QCM operator has to change the buffer conditions, an undesired parasitic frequency shift may occur, depending on the shape of the electrodes and conductance of the solution. Surface roughness of the quartz resonator also influences interpretation of adsorption phenomena. Alteration of hydrophilicity upon adsorption can lead to tremendous changes in resonance frequency since rough and hydrophilic surfaces entrap liquids in small cavities thus contributing to the overall mass detected by the device, while hydrophobic cavities are not wetted by the liquid resulting in the inclusion of air or vacuum. This implies that the resonance frequency shifts to smaller values when changing from a hydrophobic to a hydrophilic surface. When working in buffer solutions, the influence of the electrochemical double layer on the resonance frequency should also be taken into account. Considering the electrochemical double layer as a rigid mass with an extension of a few nanometers in the  $z$ -direction leads to an additional frequency change.

**Viscoelastic loading** The impact of viscoelastic materials on the electromechanical properties of a TSM resonator can be best described by the acoustic load concept. Adding a viscoelastic film with a sufficiently small shear modulus to the quartz crystal results in an additional frequency shift as expected for a thin rigid layer if the environment is air, or a decrease in frequency shift as expected from eqn [1] if the layer is adsorbed in liquid. The so-called ‘extra mass’ effect occurring in air can be expressed by a plus sign in eqn [5]:

$$\Delta f \propto 2\pi f_0 \rho_F h_F \left( 1 + \frac{1}{3} \frac{G'}{\rho_F |G|^2} (2\pi f_0 \rho_F h_F)^2 \right) \quad [5]$$

The increased shift in frequency by a thin viscoelastic film may increase the frequency response to added mass as compared to eqn [1], thus acting as an acoustic amplifier. The change in motional resistance upon addition of a film with a thickness  $h_F$ , a density  $\rho_F$ , and a complex shear modulus  $G = G' + iG''$  can be estimated as

$$\Delta R \propto 2\pi f_0 \rho_F h_F \left( \frac{1}{3} \frac{G''}{\rho_F |G|^2} (2\pi f_0 \rho_F h_F)^2 \right) \quad [6]$$

where  $G'$  is the storage modulus and  $G''$  the loss modulus. The change in resonance frequency is decreased if a soft film is deposited in a liquid environment referred to as the ‘missing mass’ effect as is obvious from the minus sign in eqn [7]:

$$\Delta f \propto 2\pi f_0 \rho_F h_F \left( 1 - \frac{2\pi f_0 \rho_L \eta_L}{\rho_F} \frac{G''}{|G|^2} \right) \quad [7]$$

while  $\Delta R$  reads:

$$\Delta R \propto 2\pi f_0 \rho_F h_F \left( \frac{2\pi f_0 \rho_L \eta_L}{\rho_F} \frac{G'}{|G|^2} \right) \quad [8]$$

### Surface Acoustic Wave Sensors

Surface acoustic wave (SAW) devices are characterized by SAWs electrically excited in a piezoelectric single crystal plate such as quartz or lithium niobate. Surface waves travel along the surface of a relatively thick solid material penetrating to a depth of one wavelength. The stress-free boundary of a solid gives rise to surface confined waves propagating as coupled longitudinal and transversal waves. The frequencies of SAW sensors are usually between 50 MHz and a few GHz. The displacement components decay exponentially within the solid. The surface acts as an acoustic waveguide. SAWs can be excited and detected by patterned interdigital transducers (IDT) on the surface of the piezoelectric crystals. Each ‘finger’ is the origin of a SAW. The transducer works most efficiently if the periodicity matches the wavelength of the surface wave occurring at  $f = v_p/d$ , in which  $v_p$  denotes the propagation velocity and  $d$  the distance between the interdigital fingers of one electrode, i.e., the periodicity. Generally, resonators can be distinguished as two-port delay lines and one-port resonators. Two-port delay lines work with one IDT as a transmitter and one as a receiver creating resonant traveling waves. One-port resonators consist of one IDT structure in between two reflectors thus producing a standing wave in both directions. The resonance frequencies of a

single-port resonator are given by  $f_n = nv_p/2l$ , in which  $l$  is the distance between the reflectors. The  $Q$ -factor of SAW devices is lower than that of the TSM and ranges between 6000 and 12 000.

**Mass loading** If the deposited mass is a thin rigid film, the kinetic energy of the synchronously vibrating system is increased without energy loss due to viscous damping. This leads to a decrease in propagation velocity, which is the strength of the SAW sensor since it reacts very sensitively to changes in the propagation velocity. The influence on the resonance frequency is given by eqn [9]:

$$\Delta f = f_0 \frac{\Delta v}{v_p} = c_m^{\text{SAW}} f_0 \frac{\Delta m}{A} \quad [9]$$

The mass sensitivity  $c_m^{\text{SAW}}$  (a negative value) increases with the fundamental frequency  $f_0$ , which is considerably higher ( $> 100$  MHz) than that of TSM resonators rendering the sensitivity of a 100 MHz SAW sensor 200 times higher than the mass sensitivity of a 5 MHz quartz crystal.

However, the use of SAW sensors in a liquid environment is not recommended, since the wave is damped by the liquid. Energy loss occurs due to the generation of compressional waves by displacement components parallel to the surface normal and the viscous coupling of displacement components parallel to the surface.

### Love-Wave, Shear-Horizontal SAW and Acoustic Plate Mode Sensors

A Love wave is a propagating shear mode wave supported on semiinfinite substrates with a waveguide layer that exhibits a shear acoustic speed lower than that of the substrate. Other acoustic wave sensors utilizing IDT to generate and detect propagating shear modes are shear horizontal polarized surface acoustic waves (SH-SAWs) and surface transverse waves. Shear horizontal acoustic plate modes (SH-APMs) are propagating modes involving the thickness of a thin piezoelectric plate and the detection and excitation by IDT. The SH-modes in SH-APM resonators can be considered as a superposition of plane waves with in-plane displacement reflected at a particular angle between the upper and lower face of the quartz resonator involving the full thickness of the resonator.

So far, numerical modeling of mass and liquid response was specifically directed to a particular device type. However, recent developments present a unified view of the different types of resonators by employing the dispersion equation approach. In brief, mass sensitivity depends on the point of operation on the



dispersion curve (phase velocity  $v$  as a function of normalized waveguide layer thickness  $z$ ), which is chosen to be the point of steepest slope corresponding to the transition where wave displacement is predominately in the substrate or the waveguide layer. The mass sensitivity of a Love-wave device can be expressed as:

$$c_m^{\text{Love}} = \lim_{\Delta m/A \rightarrow 0} \frac{A}{\Delta m} \left( \frac{\Delta v}{v_0} \right) \approx \frac{(1 - v_p^2/v_0^2)}{(1 - v_l^2/v_0^2)} \frac{f_0}{\sqrt{\mu_l \rho_l}} \left( \frac{d \ln(v)}{dz} \right)_{z=z_0} \quad [10]$$

$v$  is the phase velocity of the composed resonator as a function of  $z$ ,  $v_p$  that of the mass layer, and  $v_l$  the corresponding phase velocity of the waveguide layer.  $v_0$  is the phase velocity of the unperturbed system consisting of the waveguide and substrate at the point of operation  $z_0$  on the dispersion curve.  $\mu_l$  is the Lamé constant and  $\rho_l$  the density of the waveguide layer. The local slope of the  $v(z)$  dispersion curve might itself be frequency depending and as a consequence it would not be correct to claim that the sensitivity is proportional to the frequency at operation.

It was shown that eqn [10] can be employed to describe the sensor response of both SH-SAW and SH-APM resonators all with acoustic waveguides. In conclusion, the slope at the operating point  $z_0$  of the dispersion curve determines the sensitivity of the device.

### Flexural Plate Wave Sensors

Flexural plate wave (FPW) resonators are thin, rectangular membranes made of tension-free silicon nitride embedded in a frame of silicon manufactured photolithographically. Oscillations of these plates, which are only a few micrometers thick, can be excited piezoelectrically via IDT, electrostatically, or by using magnetic transducers. FPW sensors are characterized by a high  $Q$ -factor and low-energy loss in fluids at a low resonance frequency. Although the mechanical amplitude is rather high (100 nm), energy dissipation is low since the phase velocity of the acoustic wave is lower than the velocity of sound in most liquids (900–1500 m s<sup>-1</sup>). Low resonance frequencies (1–10 MHz) permit the use of low-cost electronics providing an attractive alternative to the less sensitive TSM resonators.

The simplest case of an oscillating isotropic plate includes an infinite set of waves known as Lamb waves. Two sets of waves can be distinguished: symmetric waves ( $S_0$ ,  $S_1$ , etc.) with particle displacements symmetric about the neutral plane and antisymmetric waves ( $A_0$ ,  $A_1$ , etc.), whose displacements have odd symmetry. For sufficiently thin plates only two waves occur,  $A_0$  and  $S_0$ . Thinner membranes exhibit

lower phase velocities for the  $A_0$  mode, in which the plate undergoes flexure as the wave propagates, while phase velocity reaches its maximum value for the  $S_0$  mode. In the case of the  $A_0$  mode the eigenfrequency decreases with decreasing thickness of the membrane at a given wavelength  $\lambda$ :

$$f = \frac{1}{\lambda} \frac{\sqrt{B}}{\rho d} \quad [11]$$

$B$  denotes the bending stiffness,  $\rho$  the density, and  $d$  the thickness of the membrane.

**Mass loading** Increasing the mass of the isotropic plate by a thin, rigid foreign mass layer results in a decreased phase velocity of the  $A_0$  Lamb wave:

$$\Delta f = c_m^{\text{FPW}} f_0 \frac{\Delta m}{A} \quad [12]$$

with  $c_m^{\text{FPW}} = -1/\rho d$ , the integral mass sensitivity, which can be increased by using thinner plates thus reducing the phase velocity and as a consequence the resonance frequency.

## Analytical Applications of Piezoelectric Sensors

### Gas-Phase Detection

The first gas-phase sensor based on TSM resonators was described by King in 1964. He used a range of standard chromatographic stationary phases to selectively adsorb vapors. The first analytical application of an SAW sensor was reported in 1979 by Wohltjen and Dessy, who used the device as a detector for gas chromatography, and as a thermo-mechanical polymer analyzer to determine glass transition temperatures and effects of ultraviolet curing on photoresists.

The key feature of all acoustic wave sensors for detecting vapors is that measurable characteristics of the acoustic wave is altered as a result of adsorption on the surface of a receptive layer or absorption into the bulk of a thin layer (Figure 3). After sorption of the vapor by a thin film on top of the acoustic resonator equilibrium conditions are established and as a consequence of the increased mass or more accurately the change in the phase velocity of the acoustic wave a signal is created. Surface coatings generally enhance the sorption of vapors with the key properties of selectivity and sensitivity while affording reversibility. Typically, rubbery polymers were used on SAW devices such as polyisobutylene or substituted polysiloxanes but also self-assembled

shift together with the assumption of a linear relation between frequency shift and mass load allow the determination of binding and rate constants of protein adsorption. Moreover, the QCM method has been evolved as a routine technique to control the quality of multilayers as prepared by the Langmuir–Blodgett technique and self-organization processes. In the early days of the QCM operated in liquids, it was postulated that the adsorbed amount of protein can be calculated from the frequency shift using eqn [1]. However, numerous publications yet established that protein adsorption performed in liquid often leads to larger frequency shifts than in air.

The high specificity of antigen–antibody reactions and the ability to generate antibodies against a variety of biological and nonbiological substances led to the development of a variety of piezoelectric immunosensors. The online detection of antibody–antigen reactions in aqueous solution was first reported by Roederer and Bastiaans on SAW resonators and by Thompson and co-workers on TSM resonators. Nowadays, the spectrum ranges from applications in clinical analysis through the determination of bacteria in food industry to environmental analysis for the detection of organic compounds by using an antigen–antibody reaction. Since the invention of phage libraries immunosensing systems based on the QCM has been successfully applied as a device for the screening of phage libraries and determination of antibody affinity.

A rather new and very exciting application of TSM resonators is the detection and characterization of pro- and eukaryotic cells, as the technique provides information of the viscous and viscoelastic load on the surface. Most piezosensors used for the detection of bacteria in solution are based on an antigen–antibody reaction, in which the bacterium binds to the corresponding surface confined antibody and thus can be monitored. In most cases, a linear relation between the bacterial cell number and frequency shift is found enabling one to calibrate the system for the bacterial cell numbers. However, whole cells do not always form complete layers on the surface; they have dimensions extending that of the acoustic

damping region in liquids and a viscoelasticity differing from simple Newtonian behavior. To obtain a set of data allowing one to extract cell-specific parameters impedance analysis in the range of the resonance frequencies of the quartz plate is hence required.

**See also:** **Sensors:** Overview; Amperometric Oxygen Sensors; Calorimetric/Enthalpimetric; Chemically Modified Electrodes; Microorganism-Based; Photometric; Tissue-Based.

## Further Reading

- Ballantine DS, White RM, Martin SJ, *et al.* (1997) *Acoustic Wave Sensors*. San Diego: Academic Press.
- Bandey HL, Martin SJ, Cernosek RW, and Hillmann AR (1999) Modeling the responses of thickness-shear mode resonators under various loading conditions. *Analytical Chemistry* 71: 2205–2214.
- Grate JW (2000) Acoustic wave microsensor arrays for vapor sensing. *Chemical Review* 100: 2627–2648.
- Janshoff A, Galla H-J, and Steinem C (2000) Piezoelectric mass-sensing devices as biosensors – an alternative to optical biosensors? *Angewandte Chemie International Edition* 39: 4004–4032.
- Johannsmann D (1999) Viscoelastic analysis of organic thin films on quartz resonators. *Macromolecular Chemistry and Physics* 200: 501–516.
- Lucklum R and Hauptmann P (2003) Transduction mechanism of acoustic-wave based chemical and biochemical sensors. *Measurement Science and Technology* 14: 1854–1864.
- Marx KA (2003) Quartz crystal microbalance: a useful tool for studying thin polymer films and complex biomolecular systems at the solution-surface interface. *Biomacromolecules* 4: 1099–1120.
- Rodahl M, Höök F, Krozer A, Brezezinski P, and Kasemo B (1995) Quartz crystal microbalance setup for frequency and *Q*-factor measurements in gaseous and liquid environments. *Review of Scientific Instruments* 66: 3924–3930.
- Wegener J, Janshoff A, and Steinem C (2001) The quartz crystal microbalance as a novel means to study cell-substrate interactions *in situ*. *Cell Biochemistry and Biophysics* 34: 121–151.

## Tissue-Based

**S B Adeloju**, Monash University, Gippsland, VIC, Australia

## Introduction

The use of plant and animal tissues, instead of isolated enzymes, for the construction of biosensors has attracted much interest in the last three decades.

shift together with the assumption of a linear relation between frequency shift and mass load allow the determination of binding and rate constants of protein adsorption. Moreover, the QCM method has been evolved as a routine technique to control the quality of multilayers as prepared by the Langmuir–Blodgett technique and self-organization processes. In the early days of the QCM operated in liquids, it was postulated that the adsorbed amount of protein can be calculated from the frequency shift using eqn [1]. However, numerous publications yet established that protein adsorption performed in liquid often leads to larger frequency shifts than in air.

The high specificity of antigen–antibody reactions and the ability to generate antibodies against a variety of biological and nonbiological substances led to the development of a variety of piezoelectric immunosensors. The online detection of antibody–antigen reactions in aqueous solution was first reported by Roederer and Bastiaans on SAW resonators and by Thompson and co-workers on TSM resonators. Nowadays, the spectrum ranges from applications in clinical analysis through the determination of bacteria in food industry to environmental analysis for the detection of organic compounds by using an antigen–antibody reaction. Since the invention of phage libraries immunosensing systems based on the QCM has been successfully applied as a device for the screening of phage libraries and determination of antibody affinity.

A rather new and very exciting application of TSM resonators is the detection and characterization of pro- and eukaryotic cells, as the technique provides information of the viscous and viscoelastic load on the surface. Most piezosensors used for the detection of bacteria in solution are based on an antigen–antibody reaction, in which the bacterium binds to the corresponding surface confined antibody and thus can be monitored. In most cases, a linear relation between the bacterial cell number and frequency shift is found enabling one to calibrate the system for the bacterial cell numbers. However, whole cells do not always form complete layers on the surface; they have dimensions extending that of the acoustic

damping region in liquids and a viscoelasticity differing from simple Newtonian behavior. To obtain a set of data allowing one to extract cell-specific parameters impedance analysis in the range of the resonance frequencies of the quartz plate is hence required.

**See also:** **Sensors:** Overview; Amperometric Oxygen Sensors; Calorimetric/Enthalpimetric; Chemically Modified Electrodes; Microorganism-Based; Photometric; Tissue-Based.

## Further Reading

- Ballantine DS, White RM, Martin SJ, *et al.* (1997) *Acoustic Wave Sensors*. San Diego: Academic Press.
- Bandey HL, Martin SJ, Cernosek RW, and Hillmann AR (1999) Modeling the responses of thickness-shear mode resonators under various loading conditions. *Analytical Chemistry* 71: 2205–2214.
- Grate JW (2000) Acoustic wave microsensor arrays for vapor sensing. *Chemical Review* 100: 2627–2648.
- Janshoff A, Galla H-J, and Steinem C (2000) Piezoelectric mass-sensing devices as biosensors – an alternative to optical biosensors? *Angewandte Chemie International Edition* 39: 4004–4032.
- Johannsmann D (1999) Viscoelastic analysis of organic thin films on quartz resonators. *Macromolecular Chemistry and Physics* 200: 501–516.
- Lucklum R and Hauptmann P (2003) Transduction mechanism of acoustic-wave based chemical and biochemical sensors. *Measurement Science and Technology* 14: 1854–1864.
- Marx KA (2003) Quartz crystal microbalance: a useful tool for studying thin polymer films and complex biomolecular systems at the solution-surface interface. *Biomacromolecules* 4: 1099–1120.
- Rodahl M, Höök F, Krozer A, Brezezinski P, and Kasemo B (1995) Quartz crystal microbalance setup for frequency and *Q*-factor measurements in gaseous and liquid environments. *Review of Scientific Instruments* 66: 3924–3930.
- Wegener J, Janshoff A, and Steinem C (2001) The quartz crystal microbalance as a novel means to study cell-substrate interactions *in situ*. *Cell Biochemistry and Biophysics* 34: 121–151.

## Tissue-Based

**S B Adeloju**, Monash University, Gippsland, VIC, Australia

## Introduction

The use of plant and animal tissues, instead of isolated enzymes, for the construction of biosensors has attracted much interest in the last three decades.

Several biological materials, such as animal tissues, bacterial cells, and plant tissues, have been used for this purpose. In earlier studies, animal materials, such as heart, liver, and kidney, were used exclusively for the construction of biosensors. However, since the introduction of the first plant-tissue-based electrochemical biosensor by Kuriyama and Rechnitz in 1981, the use of plant tissues has attracted the most interest for the development of tissue-based biosensors. The popularity of the use of plant tissues as natural enzyme sources for the construction of biosensors was further enhanced by the introduction of a banana tissue biosensor, known as the 'bananat-rode', in 1985. Since then several other plant tissues, such as fruits, leaves, roots, seeds, and vegetables, as listed in **Table 1**, have been used for the development of a wide range of tissue-based biosensors.

Plant leaves have attracted considerable interest in the development of tissue-based biosensors because of their unique natural structural arrangement. The

waxy eternal coating (cuticle) is hydrophobic in nature, but permits active gas exchange through the stomata openings. For this reason, leaves are excellent choice of material for development of tissue-based biosensors that employ potentiometric gas sensing electrodes, such as carbon dioxide electrode and ammonia gas sensing membrane electrode. One such example is the use of cucumber leaves attached to an ammonia gas sensing membrane electrode for tissue-based biosensing of cysteine. In this case, the substrate (cysteine) diffused into the spongy layer of the leaf where it gets broken down into products, which include ammonia that was readily detected by the ammonia electrode. There are several other examples of the use of plant leaves for the development of other biosensors, some of which are also listed in **Table 1** and others will be discussed later.

Other popular plant tissue materials of choice include several different fruits and vegetables. The considerable interest in these plant tissue materials is

**Table 1** Some of the reported plant-tissue-based biosensors

<i>Substrate/analyte</i>	<i>Plant tissue</i>	<i>Enzyme</i>	<i>Electrode/element</i>
Alcohol	Mushroom	Alcohol oxidase	Oxygen electrode
	Tomato seeds	Alcohol dehydrogenase	Carbon paste electrode
Amino acids	Parsley seeds	Amino acid oxidase	Ammonia gas sensor
Ascorbic acid	Cabbage, cucumber, yellow squash	Ascorbic acid oxidase (AAO)	Oxygen electrode/carbon paste electrode
Atrazine	Potato	Polyphenol oxidase (PPO)	Oxygen electrode
Catechol	Coconut, potato	PPO	Glassy carbon or O <sub>2</sub> electrode
Cysteine	Cucumber leaf	AAO	Ammonia gas sensor
Dopamine	Apple, avocado, banana, mushroom, potato	PPO	Oxygen electrode/carbon paste electrode
	Potato roots	PPO	Chemiluminescence
Ethyl paraoxon	Cucumber	AAO	O <sub>2</sub> electrode
Flavanols	Banana, potato, apple	PPO	Carbon paste electrode
Fluoride (and phosphate)	Asparagus	Asparagus peroxidase	Carbon paste electrode
	Potato	PPO	O <sub>2</sub> electrode
Glutamate	Yellow squash	Glutamate decarboxylase	CO <sub>2</sub> gas sensor
Glutathione	Cucumber	AAO	O <sub>2</sub> electrode
Glycolic acid	Spinach leaves	Glycolate oxidase	Carbon paste electrode/chemiluminescence
Hydroquinone	Sweet potato	Peroxidase	Graphite paste electrode
Mycotoxins	Mushroom	Tyrosinase	Carbon paste electrode
Oxalate	Spinach, banana peel	Oxalate oxidase	O <sub>2</sub> electrode
Paracetamol	Avocado	PPO	Graphite electrode
Pectin	Orange peel	Pectin esterase	Carbon paste electrode
Peroxide	Grape, asparagus, horseradish root	Catalase, horseradish peroxidase	O <sub>2</sub> electrode, carbon paste electrode
Phenol, <i>p</i> -cresol, <i>p</i> -chlorophenol	Mushroom	Tyrosinase	Carbon paste electrode
Pyruvate	Corn kernel	Pyruvate decarboxylase	CO <sub>2</sub> gas sensor, fiber-optic CO <sub>2</sub> sensor
Sulfoxide(s-methyl-L-cysteine sulfoxide)	Cabbage leaves	Allinase	NH <sub>3</sub> gas sensor
Tyrosine	Sugar beet	Tyrosinase	O <sub>2</sub> electrode
Urea	Soybean, jack bean, chrysanthemum	Urease	NH <sub>3</sub> gas sensor, chemiluminescence

centered on the fact that they contain many naturally occurring enzymes. The early attraction to the use of banana for the development of tissue-based biosensors is due to the fact that they are very rich in a copper-containing enzyme, known as ‘polyphenol oxidase (PPO)’, which can catalyze the conversion of dopamine to a quinone. PPO is well known for its inducement of browning in fruit and vegetables. This process, known as ‘phenolic oxidation’, has a positive role in disease resistance and possibly in photosynthetic regulation. The high level of PPO in banana has been readily exploited simply by placing a slice of the fruit on the surface of a dissolved oxygen electrode to produce a tissue-based biosensor for catecholamine. Common among the fruits and vegetables that have attracted interest for construction of tissue-based biosensors, as listed in **Table 1**, are avocado, banana, cabbage, cucumber, mushroom, potato, spinach, yellow squash, and zucchini.

Apart from some of the already highlighted benefits in using some tissue materials, the use of plant and animal tissues for the fabrication of tissue-based biosensors have the following specific advantages: (1) availability of tissues in abundance; (2) relatively high enzyme concentration; (3) availability of the additional components, such as co-factors, required for optimum catalytic activity; (4) access to enzymes that may not be available in isolated form; (5) high stability of enzymes in natural environment; (6) easier use for preparation of biosensor; and (7) very low cost compared to isolated enzymes. However, as can be expected, there are also some disadvantages in using animal and plant tissues for the construction of biosensors. A common problem experienced with this approach in earlier studies is mainly associated

with the difficulties in achieving adequate immobilization of the tissue materials. The use of thick layers of tissues, as a means of improving mechanical stability, resulted in very slow response times (10–30 min or more), as it act to create a long diffusion path between the solution and the electrode surface. For this reason, thinner layers of tissue materials are now commonly used and this has resulted in the reduction of the response time to less than 1 min in some cases, but typically not more than 10 min, as indicated in **Table 2**. In some cases, the use of homogenized tissue materials, for example, within a carbon paste matrix, has resulted in even more substantial reduction of the response time down to 30–40 s or less.

Another notable disadvantage of tissue-based biosensors is that their use is sometimes limited by interference from other enzymes or substances present in the tissues. However, several new approaches have reported for reducing these effects and in some cases taking advantage of the copresence of more than one enzyme for multicomponent analysis or the elimination of an interferant(s).

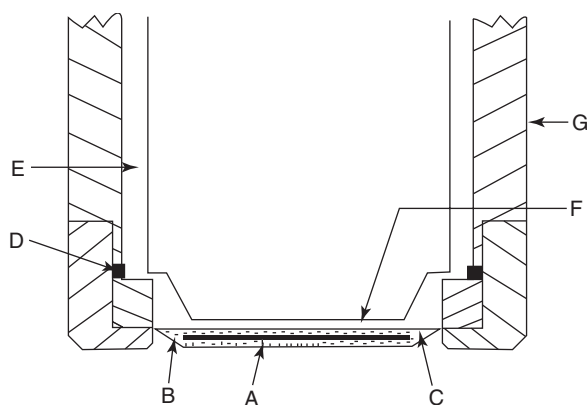
The principles and some of the important considerations in the development and use of tissue-based biosensors are described below. Also, some specific examples of the existing tissue-based biosensors that have employed a wide range of plant and animal tissues for the detection of a wide range of substances are discussed.

## Principles and Important Considerations

The working principles and operation of tissue-based biosensor are similar in many respects to those of the

**Table 2** Reported response times and lifetimes for some plant-tissue-based biosensors

<i>Substrate/analyte</i>	<i>Plant tissue</i>	<i>Response time (min)</i>	<i>Lifetime (days)</i>
Alcohol	Mushroom	2	
Atrazine	Potato		8–14
Catechol	Potato	3	1–2
Dopamine	Mushroom	4–9	120
Ethyl paraoxon	Cucumber	2	
Flavanols	Banana, potato and apple	40 s	68
Fluoride	Asparagus	1	12
Glutamate	Yellow squash	10	7
Glutathione	Cucumber		60
Glycolic acid	Spinach leaves	<1	12 (amperometry)
		1	> 30 (chemiluminescence)
Oxalate	Spinach	2–7	
Pectin	Orange peel		2
Peroxide	Grape	1	> 17 (O <sub>2</sub> electrode)
	Asparagus	2 s	> 30 (carbon paste)
Pyruvate	Corn kernel	6–12	7
Tyrosine	Sugar beet	5–10	8



**Figure 1** Schematic diagram of the arrangement of a tissue-based membrane biosensor: (A) slice of tissue; (B) bovine serum albumin conjugate layer; (C) carbon dioxide gas permeable membrane; (D) O-ring; (E) internal electrolyte solution; (F) pH-sensing glass membrane; (G) plastic electrode body. (Reproduced with permission from Kuriyama S and Rechnitz GA (1981) *Analytica Chimica Acta* 131: 91.)

conventional enzyme electrodes made from purified enzymes. However, a tissue-based biosensor is different in that it requires the placement of the tissue layer in an intimate contact with the chosen sensing electrode or other transducer that converts the biocatalytic reaction into a measurable analytical response, as illustrated in **Figure 1**. In most cases, the response to the associated biocatalytic reaction is usually measured by either amperometry or potentiometry. However, it is worth noting that amperometric detection is by far the most common detection mode used with tissue-based biosensors for the determination of various substances. The detection by amperometric and potentiometric modes are usually accomplished by combining the tissue with a dissolved oxygen electrode, a carbon paste electrode, or other gas electrodes, such as  $\text{CO}_2$  (**Figure 1**) and  $\text{NH}_3$  sensors. The optimization of the tissue and sensing electrode combination requires consideration of the influence of operational conditions, such as choice of tissue material (plant or animal), tissue thickness, choice of membrane, pH, buffer composition, and applied potential.

Another useful consideration is a knowledge of the physiology and metabolic processes of cellular material that can be beneficial for selecting appropriate tissue section(s) with high biocatalytic activity for the development of a reliable tissue-based biosensor. Such knowledge can also be helpful for characterizing and identifying interfering metabolic pathways. The transport of the substrate and of the product into, within, and out of the immobilized tissue forms the general basis for the response mechanism of tissue-based biosensors. Hence, with an understanding

of the tissue composition, it is also possible to exploit multiple enzymatic processes where several enzymes are present. The use of a combination of a tissue and a purified enzyme, as a hybrid, can also be considered for enabling the occurrence of multiple enzymatic processes in tissue–enzyme hybrid systems.

The potential for the coexistence of several enzymes in tissue materials can also be a major drawback as it can affect the selectivity of the device when used in complex sample media. Some of the strategies that have been employed for improving the selectivity of tissue-based biosensors include the use of activators to promote the primary reaction, inhibitors to suppress the undesirable reactions, or preincubation of the desired substrate. Under favorable conditions, the multienzyme activity of the tissue can be exploited for the detection of multicomponents in real samples. For example, with amperometric detection, such multicomponent detection may require the application of different potentials to achieve improved selectivity.

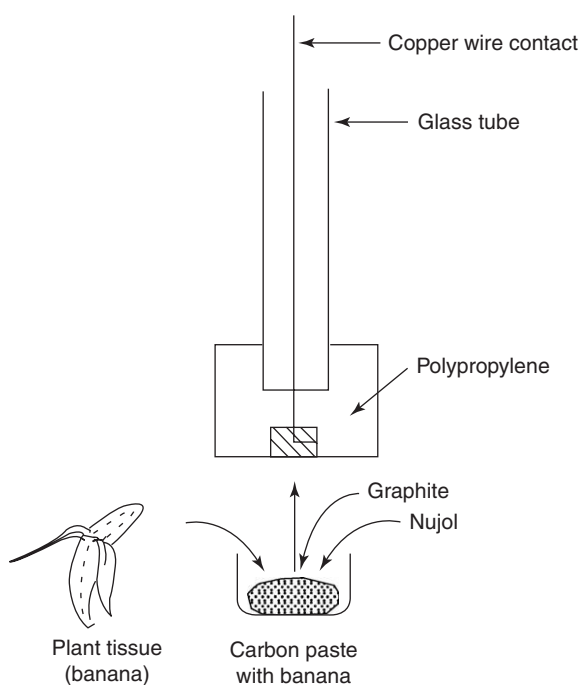
The problems of batch-to-batch reproducibility that were experienced in the early development of tissue-based biosensors were mainly associated with the difficulty in obtaining fresh and uniform sources of biocatalytic activities. Various attempts have been made to overcome these problems by use of a number of strategies, such as the use of plant tissue culture to produce fresh and consistent source of the biocatalytic activity, use of tissue homogenate, and pretreatment of tissue materials to remove possible interfering substances. Despite the batch-to-batch reproducibility problems, the recently reported analytical reproducibility (based on repeated analysis with the same tissue-based biosensor) has improved considerably, now commonly ranging from 0.3% to 4.0% relative standard deviation (RSD). In some cases, treatment of the tissue-based biosensors with a preservative, such as sodium azide, has also been found to improve performance by preventing bacterial growth and extending the lifetime of the biosensor.

Besides the availability of plant and animal tissues that contain the necessary enzymes and the associated issues with their use as described above, some of the other important considerations in the development and use of tissue-based biosensors include immobilization method, response time, and lifetimes. The significance of each of these factors is discussed below.

### Immobilization of Tissues

The physical retention of plant or animal tissues with either a membrane (**Figure 1**) or by incorporation





**Figure 2** A simplified illustration of a carbon paste banana-based electrode. (Reproduced with permission from Eggins BR, Hickey C, Toft SA, and Min Zhou D (1997) *Analytica Chimica Acta* 347: 281.)

into a carbon paste matrix (**Figure 2**) are the common methods of immobilization used for the construction of tissue-based biosensors. This is due to the fact that the matrices in which the tissues house the enzymes are already optimized and can, thus, be used directly in either of these ways. Consequently, in cases where the enzymatic activity is membrane bound, the use of tissues has a particular advantage over the use of purified enzymes. However, it is important to note that the choice of membrane or electrode components (in the case of carbon paste electrode) must enable the diffusion of the substrate into the biocatalytic enzyme layer of the tissue, while also preventing the diffusion of the biocatalyst away from the electrode surface. Suitable membrane materials for this purpose, in the case of the former, include cellophane dialysis membranes and nylon webbing of 150  $\mu\text{m}$  pore size. The diffusion of substrate and product to and from the tissue layer has been found to be unrestricted by both of these membranes. Other immobilization approaches that have attracted some interest in the fabrication of tissue-based biosensors include the use of alginate and hydrogels.

### Response Time

The ultimate test of the adequacy of the choice of immobilization of tissues by retention with a

membrane at a tip of an electrode or by incorporation into a carbon paste electrode is the achievable response times for substrate measurements. Depending on the electrode used in combination with the tissue, the response time can generally, as indicated by the data in **Table 2**, vary from less than 1 min to 10 min. Other factors that can influence the response time include tissue thickness, solution pH, membrane permeability, temperature, and stirring rate. In general, the response time increases with decreasing substrate concentration. It has been demonstrated that the two important factors that influence electrode response times are the tissue thickness and the effective solution substrate diffusion constant. To this end, the extent of external stirring influences the effective diffusion, while the bulk solution substrate diffusion coefficient influences the response time.

In terms of the choice of immobilization method, it is now widely accepted that the incorporation of the tissue into a carbon paste matrix is more effective for obtaining quick response and recovery times. A specific advantage of this immobilization approach is that the tissue becomes an integral part of the sensing electrode and, thus, enabling the achievement of not only a quick response time, but also very low detection limits. Another advantage is that, where necessary, it is possible to coimmobilize other reagents, such as redox mediators within the mixed tissue-carbon paste matrix. The nature of the mixed tissue-carbon paste matrix has also enabled the use of different designs, such as for thin-layer flow detectors and in miniaturized electrodes. **Figure 2** shows a simplified illustration of the process involved in the preparation of a carbon paste tissue-based electrode. Evidently, with the availability of an adequate plant tissue, this type of tissue-based biosensor can be prepared very quickly.

Other approaches that have been used for the immobilization of tissues for the construction of tissue-based biosensor include packing of tissues materials within the micropores of a porous carbon electrode, entrapment of tissue cells into electrochemically grown polymers, and incorporation into minireactors or column for flow analysis.

### Lifetimes

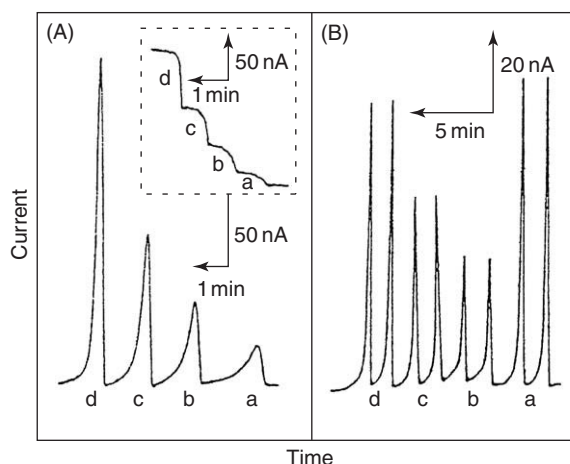
The usefulness and effectiveness of a tissue-based biosensor can be evaluated by its sensitivity, response time, detection limit, linear concentration range, and lifetime. The stability of the biocatalyst enzyme in the chosen tissue ultimately dictates the lifetime of a tissue-based biosensor. Factors that can affect the stability of the biocatalyst enzyme in the tissue include the chosen immobilization method, solution

pH, bacterial growth, storage conditions, and the presence of activators or inhibitors. Depending on the chosen tissue material, the lifetime of the biosensor can range generally from 1 to 30 days. However, as indicated by the data in Table 2, tissue-based biosensors with lifetimes longer than 3 months have been reported in recent years.

## Plant Tissue Biosensors

As discussed earlier, several fruits, plant leaves, and vegetables have been successfully used to develop tissue-based biosensors for a wide range of substances. Table 1 shows that a very diverse range of plant tissues have been used to develop various tissue-based biosensors with excellent detection limit and linear concentration range, commonly in the micromolar range. As indicated previously, the reproducibility of most of these biosensors is within 0.3–4.0% RSD and reported tissue-based biosensors for dopamine, glycolic acid, glutathione, and peroxide have lifetimes greater than 1 month and, in some cases, up to 4 months (Table 2). These various plant-tissue-based biosensors have been applied to the determination of various substances in a diverse range of sample materials, such as alcoholic beverages, river water, wastewater, urine, serum, whole blood, pharmaceutical preparations, cosmetic creams, vegetables, and fruits.

The information in Table 1 shows that plants that contain PPO such as apple, avocado, banana, coconut, mushroom and potato have attracted wide use for the construction of tissue-based biosensors for determination of atrazine, catechol, dopamine, and paracetamol. Also, plant materials that contain ascorbic acid oxidase (AAO), such as cabbage, cucumber, green zucchini squash, yellow crook neck squash, have attracted considerable interest in the development of tissue-based biosensors for the determination of ascorbic acid, glutathione, and organophosphorus pesticide. It is important to note here that while the determination of ascorbic acid was based on the direct biocatalytic effect of AAO, the measurement of glutathione and organophosphorus pesticide (ethyl paraoxon) was based on the inhibitory effect of these substances of the biocatalytic activity of AAO. Another common enzyme that is often used (Table 1) is tyrosinase, which was obtained from plant materials, such as mushroom and sugar beet, and used for the construction of tissue-based biosensor for the determination of mycotoxins, tyrosine, and phenolic compounds. It is also important to note that the use of some of the tissue-based biosensors for monitoring organic-phase biocatalytic

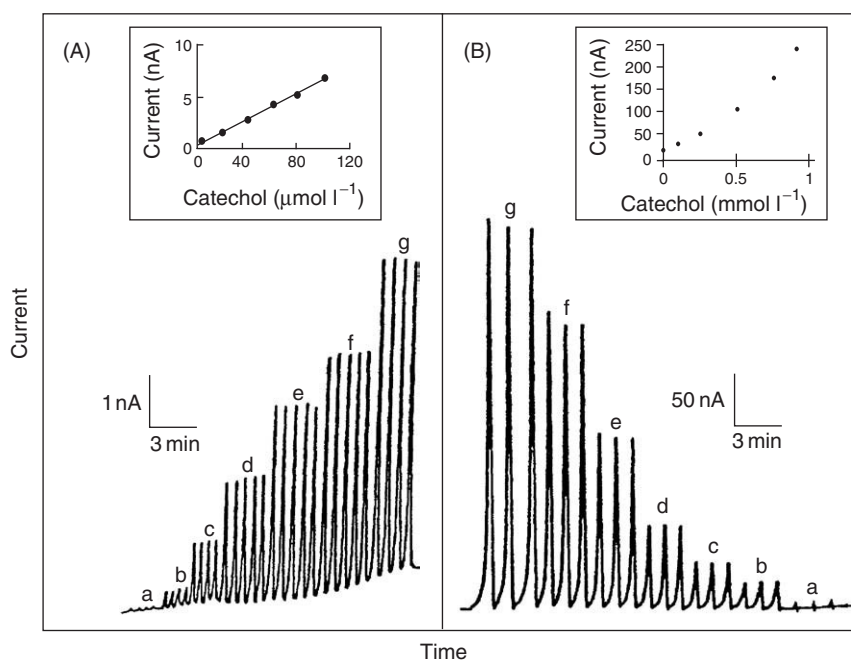


**Figure 3** Amperometric FIA detection of (A) different alcohols and of alcohols in (B) alcoholic beverages with a yeast-based biosensor. Alcohols in (A) are: (a)  $1 \text{ mmol l}^{-1}$  1-amyl alcohol, (b)  $1 \text{ mmol l}^{-1}$  1-butanol, (c)  $1 \text{ mmol l}^{-1}$  1-propanol, and (d)  $1 \text{ mmol l}^{-1}$  ethanol. Beverages in (B) are: (a) Polish vodka (Polanex), (b) red wine (Gallo), (c) white wine (Reunite), and (d) beer (Miller Lite); diluted 1:100 000, 1:100 000, 1:20 000, and 1:10 000, respectively. Inset is the corresponding batch current-time response for the addition of the different alcohols. (Reproduced with permission from Kubiak WW and Wang J (1989) *Analytica Chimica Acta* 221: 43.)

reactions has been reported. Also, an increasing use of these biosensors for flow injection analysis (FIA) and as detectors for high-performance liquid chromatography has been reported. Figure 3 shows the successful use of a yeast-based carbon paste bioelectrode for the FIA measurement of different alcohols and the determination of alcohol in alcoholic beverages. Also, a rapidly responding tissue- and microbe-based carbon paste electrode has been successfully used as an electrochemical detector for liquid chromatographic determination of *p*-cresol, dopamine, and ethanol.

The successful exploitation of the multienzyme composition of plant tissues in eliminating interferences and, hence, improving the selectivity and stability of tissue-based biosensors have also been reported. The presence of AAO in zucchini has been successfully used to eliminate ascorbic acid interference from the determination of dopamine or norepinephrine. Also, the incorporation of papaya tissue into a carbon paste matrix has been effectively used to destroy surface active proteins.

Another area of interest is the use of tissue-based biosensors in bioreactors. Figure 4 shows the successful use of a coconut-tissue-based bioreactor for the determination of catechol in river water and wastewater. Evidently, catechol concentrations can be determined reliably in these samples by this approach down to the micromolar levels.



**Figure 4** Quantification of catechol in (A) river water and (B) a paper plant waste water by FIA with a coconut tissue-based biosensor. (a) response for sample in absence of coconut bioreactor; (b) response for sample in presence of coconut bioreactor; (c)–(g) 20–100  $\mu\text{mol l}^{-1}$  spike for river water or 0.1–0.0  $\text{mmol l}^{-1}$  for waste water. (Reproduced with permission from Lima AWO, Nascimento VB, Pedrotti JJ, and Angnes L (1997) *Analytica Chimica Acta* 354: 325.)

## Animal Tissue Biosensors

In addition to plant tissues, several animal tissues have also been used to develop tissue-based biosensors. Some of the animal tissues that have been employed for this purpose and the associated enzymes and substrates are listed in **Table 3**. Evidently, two of the enzymes and substrates that have attracted considerable interest in the use of animal tissues for tissue-based biosensors are cytochrome *c* and lactate dehydrogenase for the determination of formic acid and lactate, respectively. Some of the other substances that have been determined with animal-tissue-based biosensors include adenine, alcohol, cholesterol, choline, guanine, and uric acid.

In one study, sliced or smashed (paste) porcine kidney was used, as a source of glutaminase, in conjunction with a conductance-surface acoustic wave resonator and a pair of parallel electrodes to develop a tissue-based biosensor for the determination of glutamine. The glutaminase in the porcine kidney catalyzed the hydrolysis of glutamine. The biosensor achieved a detection limit of 340  $\mu\text{mol l}^{-1}$  for glutamine and a wide linear concentration range of 680–6800  $\mu\text{mol l}^{-1}$ . The biosensor response was reproducible, achieving an RSD of 2.5%.

In another study, a slice of bovine liver attached to an oxygen electrode was used to construct a

tissue-based biosensor for hydrogen peroxide. The response of the biosensor to hydrogen peroxide was sensitive and stable. Other similar liver tissue biosensors for hydrogen peroxide have been reported and their lifetime is  $\sim 8$  days. Also, fluoride inhibition of the activity of liver esterase has been used to determine fluoride concentration down to 0.1  $\mu\text{mol l}^{-1}$ .

A recent interest in using animal tissue for construction of tissue-based biosensor involves the use of insect antenna. The olfactory abilities of up to 250 insects have been investigated to date. Of particular interest is the ability of these insects to detect odors. The two main groups of odor substances that these insects can detect are pheromones and host plant odors. These insects can detect  $\sim 250$  pheromones with very high sensitivity and selectivity, while  $\sim 400$  host plant odors can also be detected. The insect antenna is the key odor recognition component, consisting of several segments, equipped with sensilla that contain between one and three neurons that have special receptors for some odor substances. The use of the olfactory abilities of insects for the development of tissue-based biosensor has been reported. This involves the coupling of an insect (Colorado Potato Beetle) antenna to a field-effect transistor to detect the concentrations of the odor of *cis*-3-hexen-1-ol down to the parts per trillion

**Table 3** Some animal tissues that have been used for the development of tissue-based biosensors

<i>Animal tissue</i>	<i>Enzyme</i>	<i>Substrate/analyte</i>
Bovine erythrocytes	Acetylcholine	Choline
Bovine heart	Cytochrome <i>c</i>	Formic acid
	Lactate dehydrogenase	Lactate
Bovine pancreas	Cholesterol esterase	Cholesterol
Chicken heart	Cytochrome <i>c</i>	Formic acid
	Lactate dehydrogenase	Lactate
Equine liver	Alcohol dehydrogenase	Alcohol
Horse heart	Cytochrome <i>c</i>	Formic acid
Human erythrocytes	Lactate dehydrogenase	Lactate
Intestine	Adenosine deaminase	Adenosine
Pigeon breast	Cytochrome <i>c</i>	Formic acid
Porcine heart and muscle	Lactate dehydrogenase	Lactate
Porcine kidney	Glutaminase	Glutamine
Porcine liver	Uricase	Uric acid
Porcine pancreas	Cholesterol esterase	Cholesterol
Rabbit heart	Cytochrome <i>c</i>	Formic acid
	Lactate dehydrogenase	Lactate
Rabbit liver	Guanase	Guanine
Rat heart	Cytochrome <i>c</i>	Formic acid
Sheep heart	Cytochrome <i>c</i>	Formic acid
Trout muscle	Lactate dehydrogenase	Lactate
Yeast	Alcohol dehydrogenase	Alcohol

level. The electrical behavior of the antenna has also been studied by impedance spectroscopy and found that it is not just a passive electronic device, but that it reacts to voltage or odor concentration.

**See also:** **Flow Injection Analysis:** Principles. **Sensors:** Overview; Amperometric Oxygen Sensors; Calorimetric/Enthalpimetric; Chemically Modified Electrodes; Micro-organism-Based; Photometric.

## Further Reading

Bezerra VS, de Lima Filho JL, Montenegro MCBSM, Araujo AN, and da Silva VL (2003) *Journal of Pharmaceutical and Biomedical Analysis* 33: 1025.

Fatibello-Filho O, Lupetti KO, and Vieira IC (2001) *Talanta* 55: 685.

Li B, Zhang Z, and Jin Y (2002) *Biosensors and Bioelectronics* 17: 585.

Oungpipat W and Alexander PW (1994) *Analytica Chimica Acta* 295: 37.

Rekha K, Gouda MD, Thakur MS, and Karanth NG (2000) *Biosensors and Bioelectronics* 15: 499.

Schroth P, Luth H, Hummel HE, Schutz S, and Schoning MJ (2001) *Electrochimica Acta* 47: 293.

Sezginturk MK and Dinckaya E (2004) *Biosensors and Bioelectronics* 19: 835.

Sidwell JS and Rechnitz GA (1986) *Biosensors* 2: 221.

Timur S, Pazarlioğlu N, Pilloton R, and Telefoncu A (2004) *Sensors and Actuators B* 97: 132.

Wang J (1991) *Electroanalysis* 3: 255.

# SENSORY EVALUATION

**F Sinesio**, Istituto Nazionale di Ricerca per gli Alimenti e la Nutrizione (INRAN), Rome, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The use of human senses to measure and interpret flavor and sensory characteristics of foods, beverages,

or other materials is called sensory evaluation. All the human senses are involved in describing and measuring what can be detected by laboratory instruments (e.g., volatile chemicals by gas chromatography). The article consists of a general guide to the sensory and consumer testing methodologies and gives the basic principles of sensory evaluation. Sensory evaluation comprises a set of techniques that can provide analytical information of the sensory attribute of food or nonfood products to define the

**Table 3** Some animal tissues that have been used for the development of tissue-based biosensors

<i>Animal tissue</i>	<i>Enzyme</i>	<i>Substrate/analyte</i>
Bovine erythrocytes	Acetylcholine	Choline
Bovine heart	Cytochrome <i>c</i>	Formic acid
	Lactate dehydrogenase	Lactate
Bovine pancreas	Cholesterol esterase	Cholesterol
Chicken heart	Cytochrome <i>c</i>	Formic acid
	Lactate dehydrogenase	Lactate
Equine liver	Alcohol dehydrogenase	Alcohol
Horse heart	Cytochrome <i>c</i>	Formic acid
Human erythrocytes	Lactate dehydrogenase	Lactate
Intestine	Adenosine deaminase	Adenosine
Pigeon breast	Cytochrome <i>c</i>	Formic acid
Porcine heart and muscle	Lactate dehydrogenase	Lactate
Porcine kidney	Glutaminase	Glutamine
Porcine liver	Uricase	Uric acid
Porcine pancreas	Cholesterol esterase	Cholesterol
Rabbit heart	Cytochrome <i>c</i>	Formic acid
	Lactate dehydrogenase	Lactate
Rabbit liver	Guanase	Guanine
Rat heart	Cytochrome <i>c</i>	Formic acid
Sheep heart	Cytochrome <i>c</i>	Formic acid
Trout muscle	Lactate dehydrogenase	Lactate
Yeast	Alcohol dehydrogenase	Alcohol

level. The electrical behavior of the antenna has also been studied by impedance spectroscopy and found that it is not just a passive electronic device, but that it reacts to voltage or odor concentration.

**See also:** **Flow Injection Analysis:** Principles. **Sensors:** Overview; Amperometric Oxygen Sensors; Calorimetric/Enthalpimetric; Chemically Modified Electrodes; Micro-organism-Based; Photometric.

## Further Reading

Bezerra VS, de Lima Filho JL, Montenegro MCBSM, Araujo AN, and da Silva VL (2003) *Journal of Pharmaceutical and Biomedical Analysis* 33: 1025.

Fatibello-Filho O, Lupetti KO, and Vieira IC (2001) *Talanta* 55: 685.

Li B, Zhang Z, and Jin Y (2002) *Biosensors and Bioelectronics* 17: 585.

Oungpipat W and Alexander PW (1994) *Analytica Chimica Acta* 295: 37.

Rekha K, Gouda MD, Thakur MS, and Karanth NG (2000) *Biosensors and Bioelectronics* 15: 499.

Schroth P, Luth H, Hummel HE, Schutz S, and Schoning MJ (2001) *Electrochimica Acta* 47: 293.

Sezginturk MK and Dinckaya E (2004) *Biosensors and Bioelectronics* 19: 835.

Sidwell JS and Rechnitz GA (1986) *Biosensors* 2: 221.

Timur S, Pazarlioğlu N, Pilloton R, and Telefoncu A (2004) *Sensors and Actuators B* 97: 132.

Wang J (1991) *Electroanalysis* 3: 255.

# SENSORY EVALUATION

**F Sinesio**, Istituto Nazionale di Ricerca per gli Alimenti e la Nutrizione (INRAN), Rome, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The use of human senses to measure and interpret flavor and sensory characteristics of foods, beverages,

or other materials is called sensory evaluation. All the human senses are involved in describing and measuring what can be detected by laboratory instruments (e.g., volatile chemicals by gas chromatography). The article consists of a general guide to the sensory and consumer testing methodologies and gives the basic principles of sensory evaluation. Sensory evaluation comprises a set of techniques that can provide analytical information of the sensory attribute of food or nonfood products to define the

technical specification of the product and the hedonic perception of that product as defined by the user/consumer.

## Sensory Panels

Depending of their role, different types of panels with different sensory qualification can be distinguished.

### Expert Panel

It is composed of carefully trained and selected respondents with high degree of sensory sensitivity, discriminative ability, and experience of sensory methodology; the experts have developed a long-term memory and are able to make consistent and repeatable sensory assessments of products. Expert assessors can have additional experience as a specialist in the products and the process.

### Trained Panel

It comprises selected and trained assessors who have undergone sensory training and have experience of the methods under investigation. These people have developed the ability to perceive the differences and describe sample products' sensory characteristics and their intensity.

### Consumer Panel

In contrast to all other types of panels, consumers are not given any training, and therefore are only employed for preference and acceptance tests. Their answers are subjective and food acceptance is influenced by many factors like personal experience, sex, age, regular use of the product, eating habits, and social and nutritional aspects.

Consumers are recruited and screened for eligibility to participate in the test from a consumer database consisting of prerecruited consumers and are representative of the segment of population that regularly purchase or is expected to use the product or the product category.

## Panel Selection and Development

### Threshold and Sensitivity

The human sensory apparatus varies in sensitivity from person to person. Therefore, in the selection of panelists for sensory evaluation purposes the first step is to establish the limits of their sensory capacities by determining their perception thresholds for basic tastes, flavor, and odor recognition.

There are several definitions of threshold. The 'detection threshold' is defined as the lowest stimulus eliciting a sensorial response, even though no identification can be reached. The 'recognition threshold' or 'identification threshold' is the level of stimulus at which the specific stimulus can be identified. The 'differential threshold' or 'just noticeable' difference, is the minimum increase in a stimulus required to detect a difference in the perception.

There is also a 'terminal threshold', which is the magnitude of stimulus at which there is no increase in the perception of stimulus increasing.

A series of increasing concentrations of a chemically pure stimulus material in a neutral substrate (water or oil) may be used to determine sensory thresholds to those compounds. The assessors have to indicate the concentration at which stimulus is perceived for two successive concentrations. Increasing concentrations of sucrose, sodium chloride, citric acid, and caffeine are used for the basic tastes sweet, salt, acid, and bitter. Likewise, the olfactory threshold is determined by using some odorants in aqueous or alcoholic solutions (absolute ethanol is often used as a solvent), to measure the ability of the panelists to identify an odor impression and evaluate their odor memory.

### Background Information and Screening of Candidates

The most important problem in sensory testing is that although the sense receptors make an objective detection of food product, the personal psychology of consumers may lead them to have subjective judgments based on personal interpretations. An accurate sensory analytical methodology has been developed to obtain significant sensory results. This methodology requires trained sensitivity and a reliable and reproducible rating of small differences in sensory perception and excludes any subjective interpretation.

The most important criteria to be controlled when determining whether panel candidates are suited to the group dynamics are: health (normal sensitivity, infrequent colds, good dental condition, absence of specific allergies, diabetes, hypertension); motivation (willingness and interest in full participation); feeling of responsibility; sensory memory (ability to memorize many different odor and flavor impressions); ability to communicate and describe the sensory perceptions; ability to make judgments (honesty, self-criticism, objectivity, independence); avoidance of bias; ability to make decision (not easily discouraged, not timid); team



spirit; availability; attitude to foods (lack of aversions, or strong dislike for products, particularly those to assess, together with any cultural or ethnic reason for not consuming certain foods); and punctuality.

During this phase, panelists are screened and selected for their sensory acuity and potential for describing and communicating their sensory perception. Particularly, each candidate's ability to detect and describe, by screening tests, differences in flavor or texture characteristics and their intensities by scaling methods is checked, ability that will be better developed by suitable training. The choice of the screening test reflects the requirements of the selected assessors and the test are tailored to particular products to be evaluated.

### Training

The training program depends on the future role of the panel: to develop a descriptive panel that is able to identify and describe a wide range of sensory characteristics. The training is designed to:

- familiarize the assessors with test procedures;
- develop descriptive ability (ability to recognize and describe sensory stimuli);
- improve their sensitivity and sensory memory;
- learn to quantify their sensory perception.

In the training program, twice the number of assessors required in the final panel is recruited. Panelists are informed on how the perception of taste and odor occurs by lectures and demonstrations, and receive instructions on the technical procedure of tasting, such as size of mouthful to take, number of chews, and whether to be swallowed or not.

The next step of the training program is to develop a terminology system: a common list of terms for the whole panel and of which the panel share a common understanding of the meaning of each term. Sensory concepts are aligned and recognition of specific attributes can be re-enforced by the use of appropriate chemical or physical standards.

The subsequent step consists in training assessors to rate the perceived intensity of some descriptors of foods on an evaluation scale. Panelists learn to quantify their perception, initially by ranking series of single odor or taste or texture stimuli with respect to the intensity of a particular characteristic, then quantifying the perception on the evaluation scale. The test samples used are model systems: for instance, a single taste or flavor compound in water or other neutral media; or solid or semisolid materials differentiated in their texture properties; or samples obtained by spiking product samples with a flavor

compound, such as drinks added with an extra quantity of citric acid to increase the intensity of this acidic taste. The best approach is to start with simple products first, and then pass to more complex products when the assessors become more competent.

Panelists will define the direction of the scale (from minimum to maximum, or from maximum to minimum) and decide on the words needed to anchor the scales, such as 'none' to 'extreme' or 'slight' to 'strong'. Quantitative references can be used for defining anchors points at the end of each attribute. This is done, for example, using references or made-up samples to which a known concentration of a flavor substance has been added to remind the assessors of the exact meaning of a particular value on the sensory scale.

### The Test Room

Sensory evaluations are conducted in a distraction-free environment, to ensure panelists are not biased, or their sensitivity compromised. The test room should be at a comfortable temperature and ventilated with odor-free air. Lighting must be adapted to the testing using white light, which does not throw shadows. Colored lights can be used if appearance needs to be masked when only a flavor evaluation is required. Sound should also be restricted. The assessors must be comfortably seated in separate and not too narrow booths, in order to reduce distractions and panelists interaction during individual work. The booths should be provided with writing and spitting facilities and water for rinsing mouth. Plates and glasses for serving foods must be odorless and of neutral appearance so that no bias will be introduced.

## Sensory Evaluation Methods

### Discriminative Analysis

Discrimination analysis consists of difference and sensitivity tests. All of them are nonparametric; that is, they can determine whether two products differ, but they cannot determine how much. They are used when a simple answer is needed to a very simple question, i.e., whether or not there is a perceptible difference between two samples. It is possible for two samples to be chemically different in formulation but for human being not to perceive this difference.

#### Difference tests

*Paired comparison* There are two analytical sensory forms of this test: the difference paired comparison (also called as the simple difference test) and the

directional paired comparison (also known as the 2-alternative forced-choice or 2-AFC test).

The first technique is used when the experimenter wants to determine whether a simple sensory difference exists between two products (two-sided test) without specifying the dimension(s) (attributes) of the potential difference. Two odd or matched pair coded samples (A, B) are simultaneously evaluated in a randomized order of presentation between the assessors with equal number of the four possible serving sequences. (AB, BA, AA, BB). About 25 to 50 presentations are required to determine differences. The probability ( $P$ ) that an answer occurs by chance is  $1/2$ .

In the directional paired comparison method the experimenter wants to determine whether the two samples differ in a specified dimension, such as saltiness, hardness, crispness; samples are simultaneously presented to the assessors in two possible combinations (AB, BA), randomized across panelists, with an equal number of panelists receiving either sample A or sample B first. The assessors have to choose the sample within the pair that presents the higher intensity of a specified characteristic (one-sided test).

In both forms of the test the null hypothesis is that no distinction can be made between the products: in the two-sided test the alternative hypothesis is  $P_a \neq P_b$ ; in the one-sided test it is  $P_a > 1/2$ .

**Triangle** Three coded samples (two of which are the same and one is different) are distributed to the assessors in six possible sequences of two products A and B (AAB, BBA, BAB, ABA, BAA, ABB) that should be counterbalanced across all panelists. The assessors must indicate which the odd sample is. A forced choice is required also if the selection is based on a guess. These combinations of samples are distributed at random in groups of six among the assessors. In order to balance the results (each sequence uses an equal number of times), the number of assessors taking part in the test should be preferably in multiples of six.

The triangle test allows the sensory analyst to determine if two samples are perceptibly different, but not in which attributes the samples differed.

The probability ( $P$ ) of picking the right or the wrong sample by chance is  $1/3$  and therefore the null hypothesis will be rejected in favor of the alternative hypothesis at  $P > 1/3$ .

**Duo-trio** In this test, the panelists receive three samples simultaneously: one sample identified as a standard and two coded samples, one of which is the same formulation as the standard. A forced choice is required. Panelists have to recognize which of the

coded samples matches the standard. The probability of selecting the matching sample is  $0.5$ .

**Three-alternative forced choice (3-AFC)** The test allows to determine if two products differ in the specified dimension and which sample is higher in perceived intensity of the specified attribute. Three coded samples (two of which are the same and one is different) are distributed between the panelists in randomized or balanced order in three possible combinations (AAB, ABA, BAA, or BBA, BAB, ABB). As for the triangle the null hypothesis is  $P = 1/3$ .

**A, non-A** This test is essentially a sequential paired difference test. Panelists receive and evaluate a first sample that is then removed. Then the panelists receive a second sample and are asked to indicate whether the two samples are perceived to be the same or not. Since the panelists do not have the sample available simultaneously they must mentally compare the two samples and decide whether they are similar or different.

Another historical version of this test, rejected at present, consisted in presenting a single sample with a choice of response as A or non-A.

Like the difference paired comparison method, the A-non-A test has four possible serving sequences (AA, BB, AB, BA) that are randomized across panelists, with each sequence appearing an equal number of times. As in paired comparison, the null hypothesis is of no distinction between the samples and the alternative hypothesis is  $P_a > 0.5$ .

**Two out of five** This is a difference test involving five coded samples, two of one type and three of a different type. The tasters are asked to group the two sets of samples. They evaluate the samples in sequence from left to right. There are 20 possible combinations of the five samples (AAABB, AABAB, ABAAB, BAAAB, AABBA, ABABA, BAABA, AB-BAA, BABAA; BBAAA, BBBAA; BBABA, BABBA, ABBBA, BBAAB, BABAB, ABBAB, BAABB, ABABB, AABBB) that are randomly selected when the number of judges is less than 20. This method is statistically very efficient (the null hypothesis is  $P = 1/10$ ) but is affected by sensory fatigue and by the difficulty of memorizing the samples, and therefore is not usually used in flavor testing.

**Ranking** It is used for the classification of samples according to a specified difference between samples, e.g., a particular attribute, characteristic, impression, or preference.

A set of samples are presented to each assessor in random order and compared simultaneously for a

single sensory attribute. Five samples are a reasonable upper limit and at least seven selected assessors should be recruited, preferably with the same level of qualification (e.g., untrained assessors, selected assessors, experts). The assessors assign rank 1 to the sample with the strongest or weakest intensity of the attribute according to the given instruction. Data are merely ordinal, and no information on the degree of the difference is collected. A variety of statistical methods to analyze the data is available such as the Friedman rank test of the Page test.

**Sensitivity tests** Sensitivity tests are frequently used to define a threshold for taint and off-flavor in products and to select panelists for their sensitivity to key flavor compounds or any other component that can contribute to the perceived flavor.

**Threshold** This is for the recognition of taste, odor, and flavor components. A series of solutions in order of physical concentration of the stimulus is used to determine the absolute threshold (ascending forced choice).

**Dilution** The dilution indicates the smallest amount of test material that can be detected. By preparing a dilution of an aqueous extract (food substance or chemical mixture) the single flavor component can be analyzed.

## Descriptive Analysis

Descriptive analysis involves the description of qualitative and quantitative sensory aspects of a product by a trained panel relying upon objective reference standards.

**Scaling procedure** Quantification of sensory data being the perceived intensity or liking requires some form of scaling procedure. This can simply involve ordering the sample on the basis of intensity or grading by placing them into categories that can be predefined on one or more attributes.

Sensory data can be collected in different ways according to the purpose of the study. Four methods of measuring the data using four scales can be mentioned.

**Nominal scale** Data are simply descriptive: a series of categories, which are identified and labeled using a name or a number, is listed; samples are classified into groups on basis of the frequencies of each category. They contain very little information because no quantitative relationships are established.

**Ordinal scale** Samples are ranked for a single specific characteristic. Ordinal data are used for 'category scaling'. The tasters are asked to rate the intensity of a sensorial stimulus by assigning a score on a limited verbal or numerical scale. The number of points on the scale used for the determination of the different flavor attributes can differ by 5, 7, 9, and 10 points. Usually a description is attached to each value. In this scale the data are merely ordinal and intervals are not equidistant. The size of the difference between two values cannot be assumed to reflect the difference between the perceived intensities.

**Interval scale** Panelists score the samples in a structured scale, in which each value is separated by a constant interval. The exact intervals vary according to tradition (nine-point hedonic scale, 100-point scale). The psychological error of 'central tendency', i.e., avoidance of the endpoint scale, is frequently observed in scoring. Alternatively, a nonnumerical or graphical scale can be used, consisting of a horizontal line on which panelists mark a position on the scale; the scale is then converted to centimeters for numerical analysis. Anchors are presented along with the experimental samples to illustrate minimum and maximum intensities within the test. Line scales permit fineness of differentiation among responses and are examples of continuous scales.

The feature of interval data is that equal differences between numerical values correspond to equal differences between perceived intensities of the properties measured.

**Ratio data** This method is an estimation of magnitude by comparison with a reference stimulus. Data are collected on a magnitude estimation scale. Panelists receive a reference sample to which any number that seems appropriate (unspecified modulus) or a prescribed value (fixed modulus) is assigned. All the other samples are scaled in comparison with the reference sample. And this scale can be open or limited by the highest and lowest standards provided. A ratio scale has the same properties as an interval scale and, in addition, the ratio between the value allocated to two stimuli is equal to the ratio between the perceived intensities of these stimuli.

**Consensus profiling** Also named flavor profile analysis, this technique was developed to provide aroma and flavor impressions, and assess feeling factors and aftertaste. Judgments are made by four to six selected and trained panelists who undergo an advanced training period in order to orient them to a particular product to be analyzed. They work

as a team to reach a consensus opinion. This method employs intensity scaling for each sensory impression and rating to indicate the fullness of numerous aroma and flavor perceptions. With the help of the panel leader, panelists develop the appropriate terminology. Reference standard and definition for each descriptor are also created during the training phase to improve the precision of the consensus description.

This descriptive technique does not require statistical analysis of the data: confidence is derived from the reliability of skilled panelists' collective judgment. The data are reported in tabular or graphic form. One disadvantage of this method is that the panel's opinion may be affected by a dominant personality so that not all the panel members provide an equal input to the consensus.

**Conventional descriptive profiling** Descriptive profiling encompasses techniques based on descriptive quantitative analysis (QDA), a profiling method developed during the 1970s in response to dissatisfaction with the lack of statistical treatment of data obtained with the consensus profile method and the need to describe and measure the sensory parameters with mathematical precision. The basic approach to profiling of the QDA technique establish that the data are not generated through consensus discussion, that the panel leaders are not active participants, and that the assessors rate the intensity of the sensory attributes individually and the results are averaged over individual rating of assessors.

Although many different types of rating scales are in use, continuous unstructured line scales, 90 or 150 mm long, anchored at the end by terms that define the limits for each attribute are the most used. Six to 15 trained panelists are usually employed.

Descriptive profiling techniques provide for the intelligent use of human subjects as measuring instrument. When properly trained, human subjects can be effective analytical instruments able to accurately analyze flavor and aroma attributes present in a food system, and to generate statistical comparisons of samples products.

Panelists first take part in introductory meetings and discussions in order to choose and develop terms appropriate to the aim of the tests. One of the main points of the discussion is to reach an agreement on the understanding of the different sensory attributes. All member of the panel have to characterize each sensorial perception in the same terms. Round table discussions are useful to the tasters for discussing both aroma and flavor record attributes and their intensities, and avoiding any disagreements and doubts about the terminology used in the sensorial

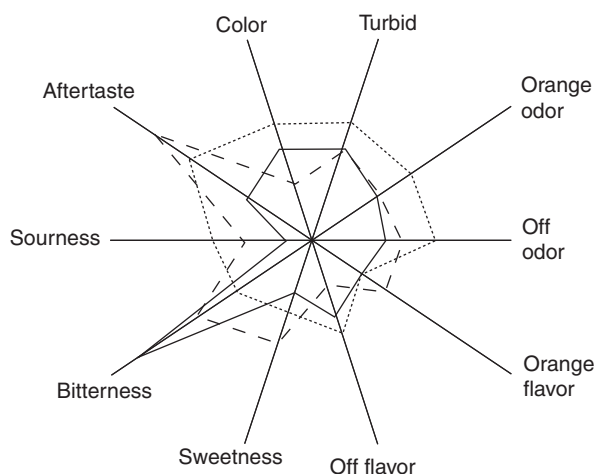
perception. In this phase, the panel prepares the score sheet containing the descriptive terms and provide an agreed definition for each term. The panelists also decide on the reference standards that should be used to anchor the descriptive terms.

Then the task of the assessors is to mark across the line the point that represents the intensity of their perceptions. Repeated judgments from each panelist for each sample are collected and arithmetic means of the score for each attribute are obtained. Panelists do not discuss the data after each taste session, to reach a consensus, but data are statistically analyzed to determine differences between the products and to control for panelists and panel performance.

Another feature of the method is the use of a graphic line scale that allows data to be presented clearly and easily. Each sensory attribute is presented by a graphic line radiating from the center as shown in Figure 1. The mean score of each parameter is located at the point on the scale corresponding to the intensity of the perception, and connected by lines in order to give a graphical profile of the descriptive results. The advantages of this kind of presentation are the following:

1. the different attributes in the comparison of the two different products are shown at a glance;
2. it provides a visual depiction of the appearance, flavor, and texture attributes of the products.

In any descriptive profiling program, before drawing any conclusion from the data, it is necessary to determine how the panelists are performing. When



**Figure 1** Quantitative descriptive analysis configuration of the aroma in orange juice samples. The average intensities for the various attributes are graphed on lines radiation outward from a value of 0 at the center point to a value of 9 at the outward perimeter.

panelists disagree, it is important to determine whether there is interaction due to a 'crossover effect' (not all the panelists rating the samples in the same order), or whether it is a matter of tasters seeing relatively small differences in intensity among the samples. Tasters may well be quite variable over time, very variable among themselves, and sometimes prone to bias.

**Free-choice profiling** The free-choice profiling technique was developed during the 1980s. According to this method each assessor develops his or her own list of descriptive terms instead of extensively training the panelists to generate a consensus vocabulary for the products or adopting common standards and agreed rating. Each panelist uses his or her unique list of term to evaluate the products. However, each individual assessor has to be able to use their developed vocabulary of terms consistently when evaluating the product. As for conventional descriptive profiling evaluations are replicated and assessors evaluated for their performance. A sophisticated multivariate statistical technique, known as the generalized procrustes analysis, it is used to analyze the data and to construct multidimensional maps that identify the descriptors that are commonly used to describe and discriminate between samples. With this technique each assessor's data are transformed into individual spatial configuration. These configurations from individual assessors are then matched by procrustes analysis to a consensus configuration that allows determining how different terms used by different assessors may be interrelated.

**Time-intensity profiling** The perceived intensity of sensory attributes can change as a function of time.

Time-intensity sensory evaluation consists in quantifying the continuous perceptual changes that occur in the specified attribute over time. Time-related methods have been used extensively to investigate the temporal behavior of tastants such as sweet and bitter stimuli and flavor release phenomena from food.

Continuous time-intensity scaling can be recorded using external devices such as strip-chart recorders, potentiometers, or computers. Computer data acquisition has facilitated the recording of time-intensity data: a scale is displayed on a computer screen and the panelists position the computer cursor on the scale to indicate the intensity of the attribute at each instant in time, using various data input methods such as moving a computer mouse or light pen, or joystick, or touching a touch-sensitive screen.

The following information can be obtained: the maximum intensity perceived during the time of

measurement, the time to reach the maximum intensity, the rate of intensity decrease after peak intensity (slope), the time to half-maximal intensity and to the estimation point, the duration of peak intensity, and the total duration of the sensation.

### Acceptance and Preference Testing

This section refers to the consumers' sensory tests used to evaluate product liking or to determine which of a series of products is the most acceptable or the most preferred.

Acceptance is an expression of higher degree of liking of a product, on a hedonic rating scale, which justifies the choice of one object over others. Acceptance or liking measurement can be done on single products and do not require comparison with other products.

Preference is an expression of positive attitude for an object and/or actual utilization. It is an expression of the level of liking. It includes the choice of one sample over another or more other samples.

**Preference test** There are different versions of preference tests.

*Paired preference test* In this test the assessors is presented with two coded products simultaneously and asked to identify the sample that it is preferred and offer a reason for preference. The test design ensures that the two possible sequences of samples (AB and BA) are balanced across panelists, so that each sample is assessed equally often in first or second position. Usually consumers are obliged to make a choice also when they have no preferences. However, there is a nonforced preference version of the test. A no-preference option is used in the score sheet and reported although usually excluded from the analysis.

*Multisample ranking for preference* This test is equivalent to the ranking test used in sensory difference testing except that attributes specified is preference or liking.

### Acceptance test

*Hedonic scaling* The method consists in recording the extent of liking of a product on a degree of liking scale that runs from 'extreme dislike' to 'extreme like'.

Hedonic scaling can be achieved using a five-, seven-, or nine-point scale. Such category scales are based on nonequal interval spacing, and for this are treated as ordinal scales. An alternative approach is to ask respondents to indicate their hedonic response

to the samples on an interval continuous scale or a line scale, sometimes anchored by like and dislike on each end, which assumes that intervals are equal and data can be summarized by averaging liking scores.

**Qualitative methods** Qualitative research methods are used to generate important terms for sensory attributes and to identify those attributes that are critical in consumer acceptance, i.e., positive attributes that have to be maximized in the product; and the characteristics that consumers do not like and think that must be minimized or eliminated from the product.

The purpose of the qualitative research analysis is to optimize product acceptance, to develop product prototypes, to establish criteria for data collection in a quantitative study, to gather information on how a product or product category is perceived, and the language that a consumer uses to talk about a product category. Group interviews are applied also in the early stage of descriptive analysis for terminology development for the use in the descriptive scorecard. This can help to ensure that everyone is speaking the same language.

Qualitative methods include one-on-one in depth interviews, group interviews, and focus group. The most commonly used qualitative research method is the focus group.

**Focus group** Focus groups vary in size and ideally consist of 8 to 12 participants. The consumer panel must be representative of the target population and it is drawn from the target population. The target population is defined as the segment of the population that regularly purchases or uses the product category.

**Quantitative methods** There are three types of quantitative affective tests using: (1) an in-house panel of consumers (laboratory test), (2) a field test

of consumers at home (home test), and (3) a field test in an institution (institutional test).

**Laboratory test** It is used when a controlled environment (odor, light, temperature, etc.) and product preparation (portion size, serving temperature) are needed.

**Home test** The aim of this test is to collect the opinion of an item by the whole family. The disadvantage of this method of data collection is the inability of the supervisor to control directly the validity of the testing procedure.

**Institutional test** Here data are collected by interview or questionnaire in institutions like markets, factories, hospitals, and schools. The advantage of this technique is that the testing procedure can be controlled. The disadvantage is the limited number of questions that can be asked to consumers.

## Further Reading

- Lawless HT and Heymann H (1998) *Sensory Evaluation of Food. Principle and Practice*. New York: Chapman & Hall.
- MacFie HJH and Thomson DMH (1994) *Measurement of Food Preference*. London: Chapman & Hall.
- Meilgaard M, Civille GV, and Carr BT (1999) *Sensory Evaluation Techniques*, 3rd edn. Boca Raton, FL: CRC Press.
- O'Mahony M (1986) *Sensory Evaluation of Food: Statistical Methods and Procedures*. New York: Dekker.
- Resurreccion AVA (1998) *Consumer Sensory Testing for Product Development*. Gaithersburg, MD: Aspen Publishers, Inc.
- Stone H and Sidel JL (1993) *Sensory Evaluation Practices*, 2nd edn. San Diego, Canada: Academic Press.
- Stone H, Sidel J, Oliver S, Woolsey A, and Singleton RC (1974) Sensory evaluation by quantitative descriptive analysis. *Food Technology* 28: 24–34.

# SEQUENTIAL INJECTION ANALYSIS

**C Lenehan**, Deakin University, Geelong, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Sequential injection analysis (SIA) was developed in 1989 at the laboratories of the Center for Process

Analytical Chemistry (CPAC), at the University of Washington. CPAC industry collaborators had been developing innovative instrumentation and methodologies using flow injection analysis (FIA), but had identified that the requirement for a different manifold for each analytical method and the relatively high maintenance of peristaltic pumps were major shortcomings when FIA was applied to process



to the samples on an interval continuous scale or a line scale, sometimes anchored by like and dislike on each end, which assumes that intervals are equal and data can be summarized by averaging liking scores.

**Qualitative methods** Qualitative research methods are used to generate important terms for sensory attributes and to identify those attributes that are critical in consumer acceptance, i.e., positive attributes that have to be maximized in the product; and the characteristics that consumers do not like and think that must be minimized or eliminated from the product.

The purpose of the qualitative research analysis is to optimize product acceptance, to develop product prototypes, to establish criteria for data collection in a quantitative study, to gather information on how a product or product category is perceived, and the language that a consumer uses to talk about a product category. Group interviews are applied also in the early stage of descriptive analysis for terminology development for the use in the descriptive scorecard. This can help to ensure that everyone is speaking the same language.

Qualitative methods include one-on-one in depth interviews, group interviews, and focus group. The most commonly used qualitative research method is the focus group.

**Focus group** Focus groups vary in size and ideally consist of 8 to 12 participants. The consumer panel must be representative of the target population and it is drawn from the target population. The target population is defined as the segment of the population that regularly purchases or uses the product category.

**Quantitative methods** There are three types of quantitative affective tests using: (1) an in-house panel of consumers (laboratory test), (2) a field test

of consumers at home (home test), and (3) a field test in an institution (institutional test).

**Laboratory test** It is used when a controlled environment (odor, light, temperature, etc.) and product preparation (portion size, serving temperature) are needed.

**Home test** The aim of this test is to collect the opinion of an item by the whole family. The disadvantage of this method of data collection is the inability of the supervisor to control directly the validity of the testing procedure.

**Institutional test** Here data are collected by interview or questionnaire in institutions like markets, factories, hospitals, and schools. The advantage of this technique is that the testing procedure can be controlled. The disadvantage is the limited number of questions that can be asked to consumers.

## Further Reading

- Lawless HT and Heymann H (1998) *Sensory Evaluation of Food. Principle and Practice*. New York: Chapman & Hall.
- MacFie HJH and Thomson DMH (1994) *Measurement of Food Preference*. London: Chapman & Hall.
- Meilgaard M, Civille GV, and Carr BT (1999) *Sensory Evaluation Techniques*, 3rd edn. Boca Raton, FL: CRC Press.
- O'Mahony M (1986) *Sensory Evaluation of Food: Statistical Methods and Procedures*. New York: Dekker.
- Resurreccion AVA (1998) *Consumer Sensory Testing for Product Development*. Gaithersburg, MD: Aspen Publishers, Inc.
- Stone H and Sidel JL (1993) *Sensory Evaluation Practices*, 2nd edn. San Diego, Canada: Academic Press.
- Stone H, Sidel J, Oliver S, Woolsey A, and Singleton RC (1974) Sensory evaluation by quantitative descriptive analysis. *Food Technology* 28: 24–34.

# SEQUENTIAL INJECTION ANALYSIS

**C Lenehan**, Deakin University, Geelong, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Sequential injection analysis (SIA) was developed in 1989 at the laboratories of the Center for Process

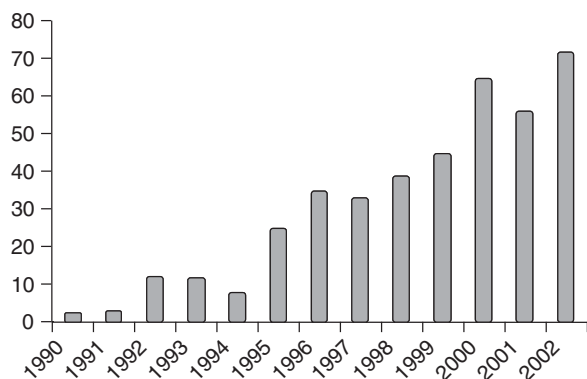
Analytical Chemistry (CPAC), at the University of Washington. CPAC industry collaborators had been developing innovative instrumentation and methodologies using flow injection analysis (FIA), but had identified that the requirement for a different manifold for each analytical method and the relatively high maintenance of peristaltic pumps were major shortcomings when FIA was applied to process

analysis. These inadequacies challenged researchers at CPAC to develop a simpler, more robust, automated liquid manipulation system.

The earliest experiments were performed with a disposable syringe, which acted as both pump and detection cell, and a multiposition valve. Upon reverse movement of the plunger, new reagents were drawn into the syringe, mixed by the resultant turbulence and the absorbance measured. Forward motion of the plunger expelled the reactants from the system in preparation for the next analysis. It was soon realized that the high degree of carryover and large reagent usage made this approach unsuitable for process analysis. Around this time Graham Marshall joined the project under the direction of Professor Jarda Ruzicka. Marshall noted that it was easier and more effective to flush the reactants from small internal diameter tubing (0.8 mm) than from the syringe barrel and consequently positioning of the detector was altered such that it was situated between the pump and multiposition valve.

In their first report of SIA in 1990, Ruzicka and Marshall operated a sinusoidal pump in reverse, sequentially aspirating reagents through a detection cell, giving rise to a primary response. This was followed by flow reversal (to expel the waste) and generated a secondary response that was broader and of lower intensity. Subsequent work by Gübeli, Ruzicka, and Christian moved the detection cell to the opposite side of the multiposition valve, resulting in a single peak; an approach that has been adopted by most SIA practitioners.

SIA has since rapidly established itself as a powerful and versatile flow-based sample-handling method as the instrumental setup facilitates the use of different chemistries without the need to reconfigure the manifold (see **Figure 1**). With the capacity to



**Figure 1** The number of SIA publications between January 1990 and December 2002, as reported by the chemical abstracts service SciFinder Scholar. Key words were sequential injection and SIA.

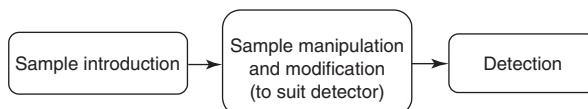
stack, in any order, precisely defined zones of sample and reagent that are mixed (by means of diffusion and flow reversal) to yield detectable products, SIA can be easily manipulated to incorporate a variety of detection methods and applied to a diverse range of analytical applications. This article provides a general overview of the principles of operation, instrumentation, detection techniques, and analytical applications of SIA.

## Principles of Operation

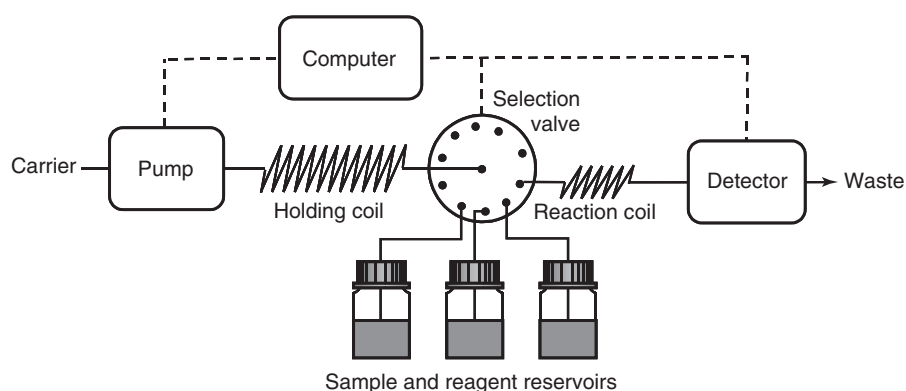
The functions used by SIA in order to perform an analytical method are similar to those used in FIA and are depicted in **Figure 2**. A zone of sample is aspirated into the instrument, the required chemistry performed, and the resultant product detected and propelled to waste. As with FIA, integral to the successful operation of SIA is the principle of controlled partial dispersion with reproducible sample handling. This is achieved using computer-controlled instrumentation comprising a selection valve, pump, and detector operating synchronously (see **Figure 3**). The instrumental configuration of SIA (**Figure 3**) allows the use of multireagent chemistries by sandwiching the sample between a variety of differing reagents and/or carriers.

SIA sequentially aspirates aliquots of sample and reagent into a holding coil by selecting the required reactant reservoir (with the selection valve) and operating the pump in reverse (see **Figure 4A** and **B**). Subsequent forward propulsion of this 'stack' of reactants results in mixing and formation of a zone of detectable product (see **Figure 4**). The order in which reactant zones are aspirated are dependent on the chemistry being used. Modifications of volumes and flow rates are achieved using appropriate computer control software.

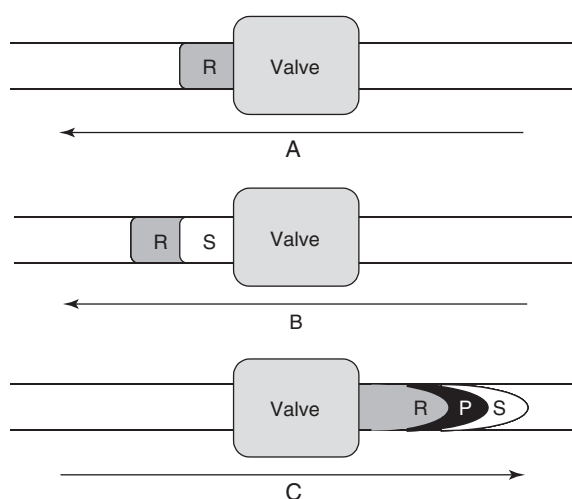
The power of SIA is evidenced in that some methodologies and chemistries are unique to SIA, and cannot be or are difficult to perform, with manual or other analytical techniques. The operational principles of SIA allow the potential for use with harsh and demanding matrices such as strongly acidic, alkaline, or organic solutions. This is due to the aspiration of reagent zones into a holding coil such that the reactants never make contact with the pump.



**Figure 2** The analytical functions performed by SIA.



**Figure 3** A schematic diagram showing a typical SIA manifold.



**Figure 4** Schematic diagram illustrating: (A and B) the pump operating in the reverse mode to aspirate a stack of sample and reagent and (C) flow reversal and the formation of a zone of detectable product, where the arrow indicates direction of flow.

## Dispersion

Not only does the flowing stream in SIA serve to propel the sample toward the detector; it also promotes mixing of the reactant zones to generate a detectable product. This process is commonly referred to as dispersion and can be defined as the dynamic merging of reactant zones as a result of forward and reversed flow through narrow bore tubing. As this merger is dynamic, the dispersion (and consequently the chemical reaction) never reaches equilibrium. Reproducible mixing is achieved by maintaining stringent control over the parameters affecting dispersion. These parameters include sample and reagent volumes, aspiration order, tubing internal diameter, tubing length, flow rate, number of flow reversals, type of reactor, and the geometry of valve,

**Table 1** Some of the terms that have been used to describe the dispersion in SIA

Term	Definition
$D$	Degree of dispersion
$S_{1/2}$	The volume required to reach a dispersion of 2 at the peak maximum
$P$	Degree of zone penetration

detectors, and connectors and are easily controlled. Two types of dispersion occur in SIA, axial and radial, which occur in the direction of and perpendicular to the direction of flow, respectively. Axial dispersion causes greater dilution and predominates in straight tubing whilst knitted/knotted reactors promote radial dispersion and minimize dilution. As might be expected, the use of more than one flow reversal results in increased dispersion and broadening of the signal. The use of tubing with internal diameters of either 0.8 or 1.5 mm has been reported to give improved precision over 0.5 mm without excessive decreases in dispersion. The terms  $D$ ,  $S_{1/2}$ , and  $P$  (see Table 1) have been used to describe the physical dispersion process and investigations have concluded that the most influential variables on dispersion were the length and internal diameter of the reaction coil. For single reagent chemistries the best results were achieved using at least twice as much reagent as sample, whilst keeping the total volume lower than half that required to give a dispersion of 2.

Whilst dispersion promotes chemistry between the reactants, and therefore sensitivity, it also causes dilution of the sample and subsequent loss of signal. Consequently it is important to optimize the reaction conditions and instrumental configuration to give the best analytical figures of merit for each application.

## Reaction Chemistry

One must not forget that the successful operation of SIA requires the judicious choice of detection chemistry. Although it may be possible to measure the analyte directly, typically the sample must be converted to a species compatible with the detection system. This is often a challenging, and sometimes daunting task, however, most common wet chemistry techniques such as colorimetric analysis and titration are easily adapted to SIA. As mentioned earlier, some methodologies and chemistries are unique to SIA, and are unable to be performed with other flow-based analytical techniques. Examples of such methods are the determination of factor XIII and the determination of morphine in nonaqueous process streams. The factor XIII assay requires the use of a complex series of reactions that are difficult to do manually, whilst the morphine sample is in a harsh organic matrix and is unable to be propelled through most pumping systems.

## Instrumentation

A typical SIA manifold is depicted schematically in Figure 2 and comprises a bidirectional pump, selection valve, reaction and holding coils, detector, and control software.

### Selection Valve

Central to the successful operation of an SIA instrument is a selection valve operated in synchronization with a pump, generally this valve is a multiposition valve, although auto samplers and magnetic snap valves have also been used for this purpose. The selection valve is connected to various sample and reagent reservoirs, typically via Teflon tubing.

### Propulsion Devices

It is imperative that the pumping device used in SIA is bidirectional and fully computer controllable. Initial reports of SIA utilized a sinusoidal pump, however, it was soon noted that 'any computer controllable piston pump capable of forward and reversed movement, would be suitable' and these were replaced with syringe pumps. Although the use of peristaltic pumps as propulsion devices for SIA was initially discounted due to the elasticity of the tubing, they can indeed be used and have the advantage of higher analytical throughput than syringe pumps. Recently the MilliGAT<sup>®</sup> pump has been designed specifically for use in FIA/SIA. These are self-priming positive displacement piston array pumps and do not contain syringes, check valves, or pulse dampeners. Alternate sources of fluid

propulsion, which have found limited utility are electroosmotic and gravity flow.

### Detectors

A large reason for the widespread utility of SIA is that any detection system that is capable of accepting a flowing stream may be utilized. It simply requires that the dimensions of the flow through cell be such that they allow solution propulsion with low-pressure pumps. Ideally detectors for SIA have a fast and stable response, high sensitivity, linearity of response over a wide range of analyte concentrations, low dead volume, low noise, do not fluctuate with temperature and flow rate, and are robust and relatively cheap. Unlike liquid chromatography and capillary electrophoresis, which can use nonspecific detection and rely on the resolving power of the separation, selectivity in SIA is achieved using prudent choice of reaction chemistries and detection parameters. Detection methods have included ultraviolet (UV)-visible, infrared, luminescence, and atomic spectroscopy along with electrochemical and turbidimetric measurements. Specific flow cell designs for sequential injection analysis include the 'fountain cell' and the 'jet ring cell'.

### Computer Control Software

In contrast to FIA where manual operation of pump and valve is easily achieved, the fundamental requirement for success in SIA is the accurate and precise synchronization of pumps, valves, and other hardware to obtain reproducible flow patterns. This, together with the benefits of electronic data acquisition and manipulation, necessitate the development of specialized software for overall instrument control. It has been noted that the lack of commercially available software was the most limiting factor in the development of SIA. In most cases specific 'in-house' programs have been developed to be compatible with the instrumentation available and software written in Turbo C++, Visual Basic, Basic, the Windows 95 environment, and LabVIEW<sup>®</sup> has been reported. Although recently commercially available software has become more widely available and the need to design purpose-built software has been lessened, some of this is instrument specific and many researchers choose to develop their own to suit the available components.

## Analytical Applications

SIA has been applied to the analysis of a wide variety of analytes in matrices as diverse as foods, beverages, bioprocesses, environmental, pharmaceutical, and

industrial processes. Published methods have incorporated sample treatment steps such as online dilution, preconcentration, separation on miniature columns, pH adjustment, phase transfer with membrane sampling devices, and immunoassay. Detection has been achieved using UV, visible and infrared spectrophotometry, turbidimetry, fluorescence, chemiluminescence, pH and ion selective electrodes, conductivity, and atomic spectroscopy, whilst some have employed chemometric manipulation of the resultant data. The following examples of analytical applications of SIA have been selected as they highlight the variety of sample treatment and detection options that can be incorporated into a method.

### Environmental Applications

The majority of analytical applications of SIA have involved the detection of analytes in environmental matrices. A considerable number of these methodologies have been concerned with the analysis of water samples, whilst some have been applied to soils, sediments, wastes, and plant materials. Applications have utilized molecular and atomic spectroscopy, fluorescence, radiochemistry, turbidimetry, and electrochemical detection techniques. Egorov and co-workers coupled chromatographic separation on a minicolumn with SIA for the determination of isotopes of americium, plutonium, and neptunium in vitrified nuclear wastes using inductively coupled plasma mass spectrometry (ICP-MS). Coupling of the ICP-MS detection system to the SIA was achieved using a four-port switching valve and ensured a continuous delivery of either the column eluate or aqueous nitric acid to the detector. The use of a confluence point (tee-piece) between the detector channel and a water reservoir has been reported to serve the same purpose when combining SIA with flame atomic absorption spectrophotometry.

### Food and Beverage Analysis

Spectroscopic, luminescence, turbidimetric, and electrochemical methods of detection have been combined with SIA for the successful determination of amino acids, sugars, and trace elements in matrices such as meats, vegetables, breads, wines, juices, and milks. Many of these methodologies required sample pretreatment and whilst most performed this in an offline manner there have been some reports of online sample cleanup. Microwave assisted digestion was performed in-line for the determination of phosphorous, calcium, magnesium, and iron in slurried foodstuffs, wine, milk, and soft drinks; whilst gaseous diffusion allowed interference removal for the determination of urea in milk.

### Industrial and Process Analysis

Industrial and processing solutions present some of the most challenging samples for analysis, as the matrices are typically complex and harsh. SIA has proved itself more than capable of such application and has been applied to the analysis of various analytes present in both industrial and fermentation processes. Determinations of a range analytes in industrial samples such as petroleum distillates, electroplating baths, mine process and effluent solutions, paper and pharmaceutical manufacturing have all been reported with SIA.

Wood pulp black liquors have been sampled directly from the processing vessel using SIA. The instrumentation automated the sample pretreatment and cleanup over an aluminum oxide minicolumn and then presented them to a high-performance liquid chromatography system for the determination of pyrocatechol, pyrogallol, gallic acid, and protocatechuic acid.

SIA instrumentation has also been modified to perform colorimetric titrations and one such apparatus has been applied to the determination of bromine number in petroleum distillates. Due to their organic nature, petroleum samples would not usually be able to be propelled through most pump tubings as they would be degraded. However, the use of SIA, which relies on reverse movement of the pump to aspirate the samples into a holding coil and via a selection valve, allows manipulation of such harsh matrices, as they never make contact with the propulsion device. The colorimetric titration system developed for bromine number assay utilized a reaction chamber fitted with an electrode for the *in situ* generation of bromine along with optical fibers for detection. Unfortunately interferences from sulfur-containing compounds resulted in errors when the method was applied to real samples.

### Bioprocess Monitoring

Several SIA applications have been reported for the determination of analytes such as ammonia, glucose, ethanol, lactic acid, and phosphate in a variety of fermentation media with spectrophotometric, turbidimetric, chemiluminescence, and electrochemical detection. This has included the indirect offline determination of glucose, lactic acid, and penicillin in *Penicillium chrysogenum* cultivars using chemiluminescence detection of enzymatically produced hydrogen peroxide. Whilst most SIA instrumental configurations consist of a single propulsion device, this system included a second pump for the injection of a reagent immediately prior to the detector.

The determination of D-lactic acid in *Lactobacillus delbrueckii* using SIA has been achieved with online sampling, filtration, dilution, and analysis based on the colorimetric detection of enzymatically formed NADH.

### Pharmaceutical Applications

The automation of certain pharmaceutical analyses through the utilization of SIA has received some interest within the research community. Applications to serial assays, drug dissolution testing, and drug screening have incorporated spectroscopic, electrochemical, and luminescence detection. Stopped flow SIA has been used to collect the chemiluminescence intensity versus time profiles of procaine, benzocaine, and tetracaine after their reaction with acidic potassium permanganate. The differing profiles were subsequently used to distinguish the analytes and resulted in reported detection limits as low as  $0.1 \mu\text{g ml}^{-1}$ .

SIA instrumentation has been applied to the automated dissolution studies of a sustained release formulation of ibuprofen. The instrument monitored the absorbance of seven replicate samples six times per hour. With the application of partial least squares regression analysis to the resultant absorbance data, this same instrumentation was capable of the simultaneous determination of aspirin, phenacetin, and caffeine in pharmaceutical formulations.

### Immunoassay

The manipulation of antigen–antibody interactions and subsequent detection of several analytes has been achieved using SIA. Whilst many of these applications have involved the use of SIA with bead injection (to be discussed later), some have used SIA for the direct analysis of cellular and bodily fluids.

Automation of a five-reagent procedure for the determination of factor 13, a chemistry that had only ever been performed manually, illustrates the potential of SIA for complex and difficult to perform chemistries. This methodology, reported by Guzman and Compton, uses a four-step sequence whereby factor 13 is activated by thrombin in the presence of calcium and UV radiation, the resultant activated product then mediates a condensation reaction between monodansylcadaverine and *N,N*-dimethylcasein to form a fluorescent product.

A specialized ‘fountain’ detection chamber equipped with an epifluorescence microscope was developed for live-cell studies of derivatized rat insulinoma cells. A coverslip coated with the cellular material was positioned in the chamber and SIA was

used to manipulate the type and rate at which fluid perfused the cells, whilst the resultant fluorescence was monitored. A comparison of this approach with corresponding FIA and fluid switching methodologies indicated that SIA was the most flexible and robust of the instrumentation tested.

### Renewable Microcolumns (Bead Injection)

Provided the instrumentation contains a suitable flow cell, small amounts of bead suspensions can be manipulated by SIA to form renewable microcolumns that act as disposable reaction surfaces, separation, and/or preconcentration devices. This involves the aspiration of a known volume of bead suspension followed by their trapping within a distinct geometry. The beads can then be perfused with carrier, sample, and/or reagents whereby the analyte is trapped on the surface of the beads and the appropriate chemistry is subsequently performed. Following detection, the beads are then discarded. Pollema and co-workers first reported this approach in 1992 when they trapped a solution of immunomagnetic beads within a magnetic field. A number of bead-trapping devices have been employed with the jet ring cell, porous frit, and rotating rod designs illustrated in **Figure 5**. This means of manipulating bead material for performing analysis has been applied to samples such as foods, beverages, urine, cellular material, and environmental and nuclear wastes.

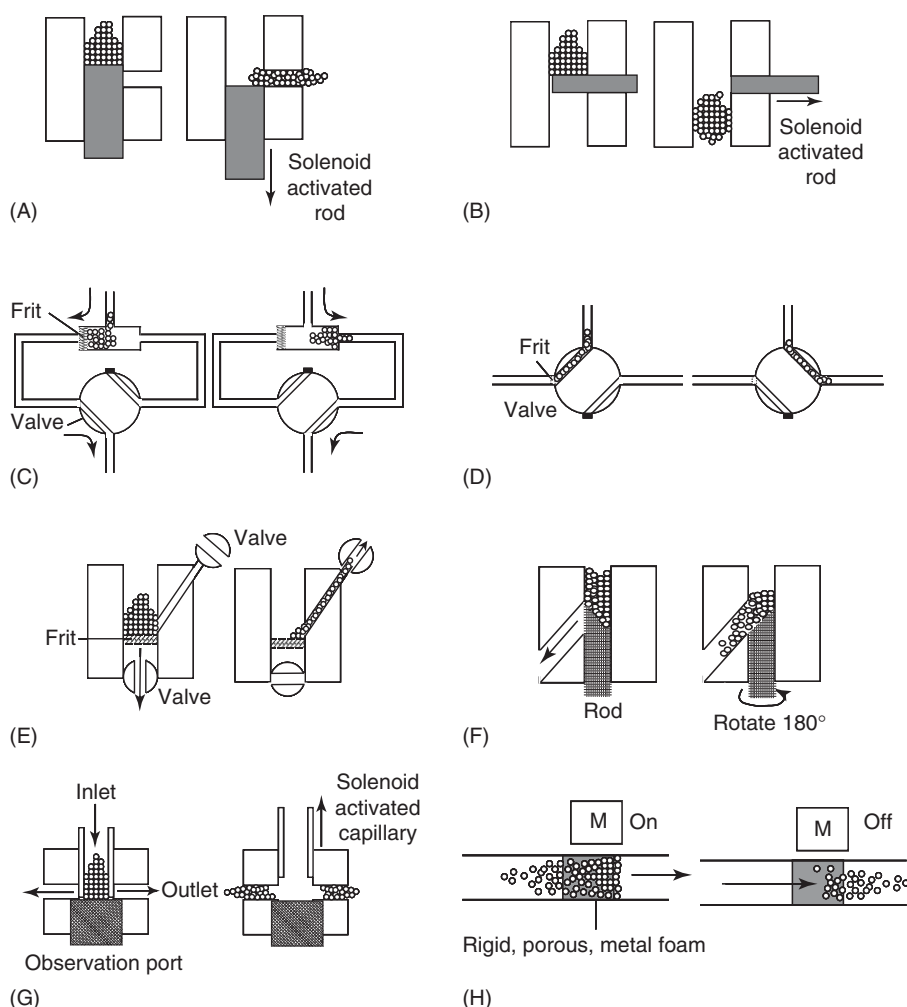
Electrically conductive glassy carbon beads have been manipulated with SIA to form renewable electrodes. The surface of the beads was coated with immobilized glucose oxidase and enabled the determination of glucose in beer and wine via the amperometric detection of hydrogen peroxide.

The utility of bead injection SIA for separation and preconcentration has been well demonstrated by the selective determination of strontium-90, americium-241, and technetium-99 along with various isotopes of plutonium and curium whose radionuclide activity was determined in nuclear waste samples using liquid scintillation counting.

### Sample Pretreatment and Preconcentration

Accurate determination of analytes in complex sample matrices is often problematic due to the presence of interferences, furthermore, analytes at low concentration may approach the detection limit of the method and consequently be unable to be directly determined. One of the great advantages of SIA is the ability to automatically preconcentrate samples from within various, and sometimes aggressive, solvents without risk of damage to the propulsion device.





**Figure 5** Schematic diagram representing some restriction devices that have been used for the formation of bead reactors. As published in *The Analyst*, 2002, 127, 997; reproduced by permission of the Royal Society of Chemistry.

After preconcentration, these samples can then be back-extracted into a similar solvent or into a matrix that is more compatible with the detection system. The use of SIA with bead injection was demonstrated for the preconcentration of nickel and bismuth on a cation-exchange resin prior to detection by electrothermal atomic absorption spectrometry (ETAAS) or ICP-MS.

Two liquid–liquid extraction steps were employed for the preconcentration and back-extraction of copper and lead prior to detection by ICP-MS. The analytes were complexed with ammonium pyrrolidine dithiocarbamate and extracted into isobutyl methyl ketone with the aid of a gravitational phase separator. As the ICP-MS is susceptible to organic solvents, a second liquid–liquid extraction whereby the analytes were displaced from the complex using Pd(II) and back-extracted into nitric acid was required.

**See also:** **Flow Analysis:** Overview. **Flow Injection Analysis:** Principles; Instrumentation; Detection Techniques. **Fluorescence:** Overview.

### Further Reading

- Chandler DP and Brockman FJ (2000) Renewable micro-columns for solid phase nucleic acid separations and analysis from environmental samples. *Trends in Analytical Chemistry* 19: 314–321.
- Lenahan CE, Barnett NW, and Lewis SW (2002) Sequential injection analysis. *Analyst* 127(8): 997–1020.
- Marshall G, Wolcott D, and Olson D (2003) Zone fluidics in flow analysis: Potentialities and applications. *Analytica Chimica Acta* 499: 29–40.
- Ruzicka J and Marshall GD (1990) Sequential injection: A new concept for chemical sensors, process analysis and laboratory assays. *Analytica Chimica Acta* 237(2): 329–343.
- Ruzicka J and Scampavia L (1999) From flow injection to bead injection. *Analytical Chemistry* 71: 257A–263A.

Taljaard RE and van Staden JF (1998) Applications of sequential injection analysis as process analyzers. *Laboratory Robotics and Automation* 10: 325–337.

Wang J, Hansen EH, and Miró M (2003) Sequential injection-bead injection-lab-on-valve schemes for on-line

solid phase extraction and preconcentration of ultra-trace levels of heavy metals with determination by electrothermal atomic absorption spectroscopy and inductively coupled plasma mass spectrometry. *Analytica Chimica Acta* 499: 139–147.

## SEWAGE

See **WATER ANALYSIS: Sewage**

## SIA

See **SEQUENTIAL INJECTION ANALYSIS**

## SIGNAL PROCESSING

See **CHEMOMETRICS AND STATISTICS: Signal Processing**

## SIZE-EXCLUSION CHROMATOGRAPHY

See **LIQUID CHROMATOGRAPHY: Size-Exclusion**

## SOFT DRINKS

See **FOOD AND NUTRITIONAL ANALYSIS: Soft Drinks**

## SOIL

See **GEOCHEMISTRY: Inorganic; Soil, Major Inorganic Components; Soil, Minor Inorganic Components; Soil, Organic Components**

## SOLID-STATE ELECTRODES

See **ION-SELECTIVE ELECTRODES: Solid-State**

## SOLID-PHASE EXTRACTION

See **EXTRACTION: Solid-Phase Extraction**

## SOLID-PHASE MICROEXTRACTION

See **EXTRACTION: Solid-Phase Microextraction**

## SOLVENTS

**B Buszewski**, Nicholas Copernicus University, Toruń, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

In modern chemical analysis, high detection sensitivity is achieved by using specific physicochemical methods optimized for the task of trace analysis where the reported detection levels are often measured in  $\mu\text{g ml}^{-1}$  (ppm),  $\text{ng ml}^{-1}$  (ppb), or  $\text{pg ml}^{-1}$  (ppt). The achievement of such low detection levels may be possible primarily due to the use of modern analytical instruments equipped with new types of detectors, and also due to improved sample preparation methods. Both these factors are closely related to the purity of solvents used as mobile phases in different variants of liquid chromatography (LC), electrochromatography (EC), capillary zone electrophoresis (CZE), or as extraction medium in liquid-liquid extraction (LLE), solid-phase extraction (SPE), membrane separation (MSP), filtration (FT), flotation (FL), etc. Moreover, these high-purity solvents are also used as media to dilute the analyzed samples for chromatographic, spectroscopic, and electrochemical analysis. Thus, solvents used in chemical analysis must fulfill many physicochemical requirements.

Many of the high-purity solvents now available for techniques including LC, electromigration and/or membrane separation, as well as for spectroscopy. However, purification and quality testing of even commercial products is often necessary before use, particularly in the above-mentioned techniques that are applicable to multicomponent systems. For

instance, the mobile phases in LC, CZE, SPE, and MSP are often modified by different organic and/or inorganic salts, and may interfere with the target analyte. Organic synthesis also requires high-purity solvents or solvent mixtures with different composition. Synthetic products of high purity are used as reference materials for analysis.

In this article, the most important physicochemical properties of commonly used solvents are briefly reviewed, including methods for their purification, recovery, and quality testing.

### Solvents: Properties and Their Usage

In many cases, the proper sequence of molecular processes occurring in multistep multicomponent analysis systems depends strongly on the purity of the solvents used. Often, the quality of measured data can be improved by using solvents of suitably high purity. From the chromatographic or spectroscopic point of view, a good solvent should be characterized by high purity and, consequently, by lower values of transparency at low wavelength (ultraviolet (UV) cut-off) and for transparency of the desired analytical wavelength and refraction (refractive index). Also, parameters such as reactivity and good miscibility are important criteria for solvent selection. In LC investigations, multicomponent solvents (e.g., binary/ternary mixtures) that have low boiling points (between 20°C and 60°C) and low viscosity ( $\eta < 50 \text{ cP}$ ) are preferred. A good solvent should be able to dissolve the sample, although this is seldom a problem in analytical separations.

The important physical parameters of the most commonly used solvents are listed in **Table 1**.

**Table 1** Physicochemical properties of solvents used in various analytical techniques

<i>Solvent</i>	<i>Usage area</i>	$\varepsilon^{\circ}$ (on silica)	UV cut-off (nm)	RI at 25°C	$d_{25^{\circ}\text{C}}$ (g ml <sup>-1</sup> )	BP (°C)	$\eta$ at 25°C (cP)	$\varepsilon$ at 25°C
Acetone	LC, E, F	0.43	330	1.356	0.791	56	0.30	20.70
Acetonitrile	LC, CZE, E, F	0.50	190	1.341	0.789	82	0.34	37.50
Benzene	LC, E, F	0.27	280	1.498	0.879	80	0.60	2.30
<i>n</i> -Butanol	LC, E, F	0.54	210	1.397	0.802	118	2.60	17.52
Carbon tetrachloride	LC, F	0.14	265	1.457	1.594	77	0.90	2.24
Chloroform	LC, F	0.26	245	1.443	1.483	61	0.53	4.80
Cyclohexane	LC, GC, E, F	0.03	200	1.423	0.774	81	0.90	2.02
Cyclopentane	LC, GC, E, F	0.04	200	1.404	0.809	78	0.42	1.97
Diethyl ether	LC, GC, E, F	0.29	218	1.350	0.708	35	0.24	4.30
Dimethylformamide	LC, F		268	1.428	0.950	153	0.80	36.73
Dimethylsulfoxide	LC, F	0.64	268	1.477	1.101	189	2.00	4.70
Dioxane	LC, E	0.56	215	1.420	1.034	101	1.22	2.21
Ethanol	LC, E, F	0.78	210	1.359	0.791	78	1.08	24.60
<i>n</i> -Heptane	LC, GC, E, F	0.00	195	1.385	0.684	98	0.40	1.92
<i>n</i> -Hexane	LC, GC, E, F	0.00	190	1.372	0.659	69	0.30	1.88
Isooctane	LC, GC, E	0.01	197	1.389	0.703	99	0.47	1.94
Methanol	LC, CZE, E, F	0.73	205	1.326	0.791	65	0.54	32.71
Methylene chloride	LC, E, F	0.32	233	1.421	1.327	40	0.41	8.93
<i>n</i> -Octanol	LC, E, F		205	1.427	0.827	195	7.31	10.30
<i>n</i> -Pentane	LC, GC, E	0.00	195	1.355	0.626	36	0.22	1.84
<i>i</i> -Propanol	LC, E, F	0.63	205	1.384	0.785	84	1.90	20.30
<i>n</i> -Propanol	LC, E, F	0.63	240	1.385	0.804	97	1.90	20.30
Tetrahydrofuran	LC, E, F	0.35	212	1.405	0.881	66	0.46	7.56
Toluene	LC, GC, E	0.22	285	1.494	0.867	110	0.55	2.40
Water	LC, CZE, E, F	> 0.73	200	1.333	1.000	100	0.89	80.00

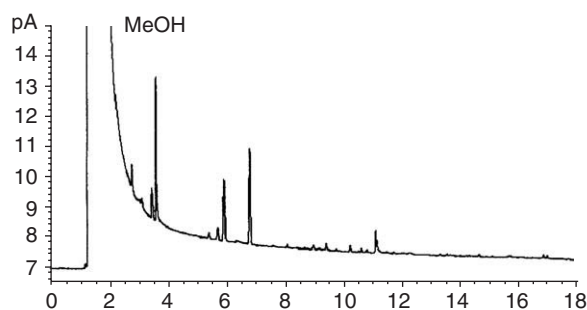
LC, liquid chromatography; GC, gas chromatography; E, extraction; F, filtration; CZE, capillary zone electrophoresis  
 $\varepsilon^{\circ}$ , eluotropic value; RI, refractive index;  $d_{25^{\circ}\text{C}}$ , density; BP, boiling point;  $\eta$ , viscosity;  $\varepsilon$ , dielectric constant.

In extraction techniques (LLC, SPE, etc.), normal-phase liquid chromatography (NPLC), and thin-layer chromatography, aliphatic hydrocarbons (e.g., *n*-hexane, *n*-heptane) are usually used. The elution strength of these solvents is often modified by addition of more polar solvents. The fundamental problem with the eluents in NPLC is dissolved water and trace amounts of olefins. These contaminations can induce a change in the wavelength cut-off values (UV detection, spectrophotometry), baseline perturbation, and poor reproducibility of retention data. Halogenated solvents such as dichloromethane can react with some organic solvents (e.g., acetonitrile) to form crystalline products.

In reversed-phase liquid chromatography (RPLC) and electromigration techniques (CEC, CZE), aqueous solutions of methanol, acetonitrile, tetrahydrofuran, and dioxane are used as eluents. In this mode of LC, a fundamental concern is the purity of water, which can contain phenols, hydrocarbons, etc. Tetrahydrofuran is used frequently as the solvent in gel permeation chromatography. It must be stabilized by butylated hydroxytoluene used as antioxidant. Similarly, when halogenated solvents (e.g., trichloroethylene) are used triethanolamine is added as a

stabilizer. However, in ion-suppression and ion-pair RP-HPLC, the mobile phases are often modified by the addition of compounds that can change the dissociation constants of analytes such as inorganic and organic acids or ionic substances (e.g., ammonium chloride, cetyl chloride, etc.). The same situation occurs when buffers are used. All these components may contain inconvenient impurities that can crystallize (precipitate) and interfere in detection during the elution process. As a result of these impurities, extra peaks can appear in chromatograms, confusing the identification of the analyzed substances. The impurities can interfere with the analyte's retention even if a UV transparent mobile phase (e.g., methanol–water, acetonitrile–water, and/or ethanol–water) is used. As an example, the GC-FID chromatogram of analytical grade methanol is shown in **Figure 1**. This chromatogram illustrates that together with the main MeOH peak, trace impurities are also detected. Thanks to chromatographic methods it may be often used for the purity testing and QA/QC (quality assurance/quality control) of solvent manufacture.

Another factor that influences the separation process and may result from solvent impurities during



**Figure 1** Example GC-FID chromatogram of methanol for chromatographic analysis. Chromatographic conditions: column – Poraplot Q (25 m  $\times$  530  $\mu$ m  $\times$  20  $\mu$ m), temperature program – 70°C held for 3 min to 200°C at a rate of 8°C, carrier gas – helium (1.4 bar).

chromatographic elution is peak tailing and/or peak splitting. In SPE, the irreversible sorption of solvent impurities on the active sites of packing materials (e.g., surface silanols or chemically bonded ligands with polar groups such as  $-\text{OH}$ ,  $-\text{NH}_2$ ,  $-\text{CN}$ ) can be manifested by differences in the reproducibility of recoveries.

Therefore, before chromatographic or spectroscopic measurements are undertaken suitable preparation (e.g., isolation, purification, and preconcentration) of the sample is recommended. For this purpose, methods such as LLE, SPE, FT, and MSP are used. It should be noted that high-purity solvents are required to prepare analytical samples.

## Methods of Solvent Purification

Even though there are many methods for the purification of solvents, the most popular methods are based on distillation: simple, fractional, and steam distillation. In recent times, solvent purification on solid-phase packing materials is becoming popular, which can be carried out by LC, SPE, and ultrafiltration by the utilization of membranes.

### Distillation

Distillation is the oldest, easiest, and cheapest procedure for purification of solvents and is based on the application of Raoult's law. The physical foundations of this separation technique depend on the distribution of the constituents between the liquid and vapor phases being at the equilibrium. In general, the composition of the vapor is different from the composition of the distilled mixture, but azeotropic mixtures distil without changing its composition. However, during distillation of normal mixtures, the

principal component with lowest boiling point distils first and then the compounds with higher boiling temperatures. The effectiveness of distillation depends on the physical properties of the components in the mixture, the equipment used, and the method chosen.

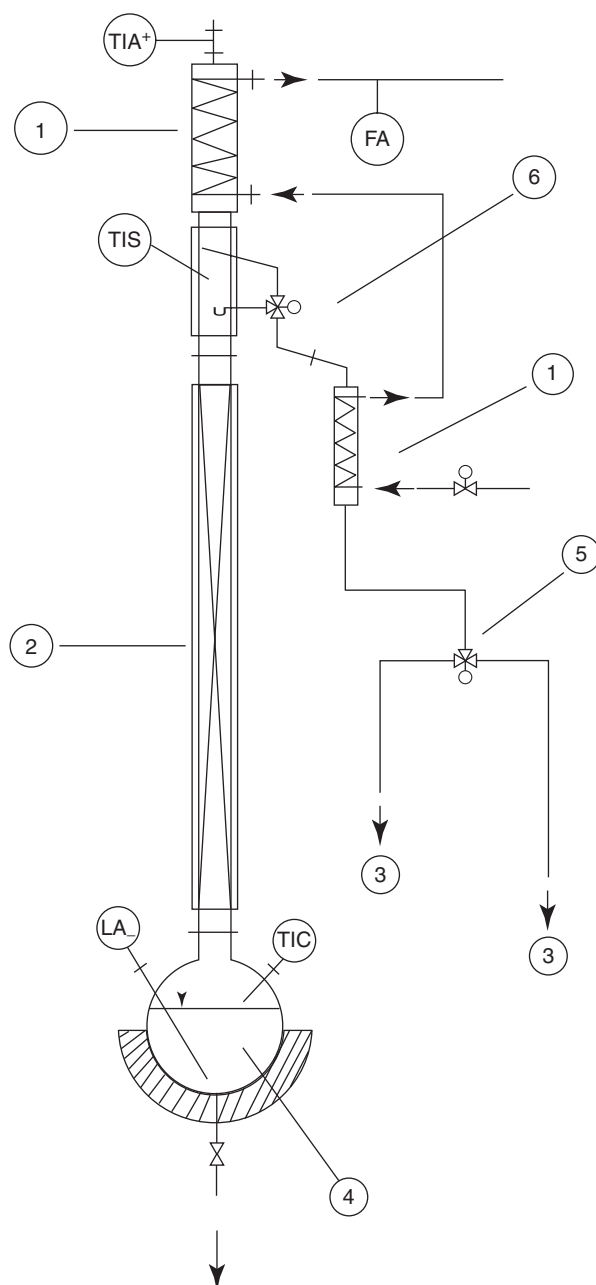
Simple distillation refers to the process in which molecules transferred from the liquid phase to the vapor phase are not subjected to partial condensation or contact with the condensed liquid prior to reaching the vapor condenser. The composition of the vapor near the liquid phase does not change as it moves along the condenser. In this technique, equipment requirements are minimal, and, usually, a flask fitted with a condenser and a product receiver is sufficient.

Fractional distillation is used when a more efficient separation process than simple distillation is required. This type of distillation is an equilibrium process, in which the composition of the distillate is constantly changing as the distillation proceeds and is changing along the distillation column toward the outlet. The main element of the apparatus is the distillation column consisting of a series of plates located one over the other in a suitable tube that is placed under the receiver. Liquid evaporating from one (lower) plate condenses on the other (higher) plate, where the evaporation process is repeated. In each plate equilibrium between the liquid and the vapor is established.

The vapor, enriched with the more volatile component, flows upwards, whereas the condensed fluid enriched with the less volatile component flows downwards. The efficiency of the column is increased by increasing number of plates. By fractional distillation (**Figure 2**), it is possible to recover solvents of chromatographic grade, such as acetonitrile or methanol obtained from an azeotropic mixture of methanol–water or acetonitrile–water.

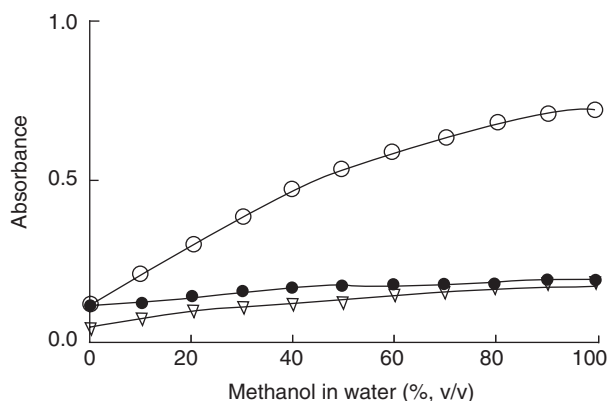
Steam distillation is a simple distillation procedure in which the vaporization of a mixture is achieved either by continuously blowing steam through a mixture or by boiling water and sample together. If sample contains both hydrophobic and hydrophilic components, two layers of distillate develop. In the typical steam distillation process, the two layers can be recovered separately. Aqueous distillation seems to provide the best compromise between time, cost and effort.

In many cases, satisfactory results are obtained when distillation is used for the purification of solvents. Normally, water used in HPLC investigations need not be purchased, but may be prepared by the user. A cheap and easy laboratory procedure for water purification is described below. Deionized



**Figure 2** Scheme of a fractional distillation system for organic solvent recovery: 1, cooler; 2, distillation column; 3, fraction flasks; 4, reboiler; 5, fraction split valve; 6, reflux valve; TIS, TIC, TIA – temperature sensors; FA – cooler sensor; LA – solvent level sensor. (Reproduced with permission from Stepnowski P (2003) Recovery of organic solvents after chromatographic analysis. *Analitique* 4: 39–41.)

water with an added alkaline solution of  $\text{KMnO}_4$  is left standing for a few days, and then the solution is distilled twice in an apparatus made of hard glass. The conductivity of the water purified in this fashion may be as low as  $10^{-6} \text{ S m}^{-1}$ .



**Figure 3** Plots of absorbance data for mixtures of water and methanol at 254 nm versus the composition of the solution obtained using 3 cm quartz cuvette. Key: ○, methanol (analytical grade and water (home purified); ●, methanol (HPLC grade, J.T. Baker, Deventer, The Netherlands); and ▽, water and methanol (HPLC grade, J.T. Baker). (Reproduced with permission from Buszewski B, Bleha T, and Berek D (1985) UV detection of solvent peaks in liquid chromatography with mixed eluents. *Journal of High Resolution Chromatography and Chromatography Communications* 8: 527–528. © Wiley-VCH.)

In Figure 3, plots of absorbance for three combinations of water and methanol of various purity are shown. The main source of UV absorption of the binary mobile phase was evidently the impurities in methanol. A comparison of both the bottom curves shown in Figure 3 also proves the effectiveness of water purification by distillation.

Other solvents, such as aliphatic and aromatic hydrocarbons, alcohols, ethers, including tetrahydrofuran, halogenated solvent, can also be purified through distillation methods. However, in many cases the solvent impurities, water in aliphatic hydrocarbons, or halogenated solvents can make the distillation process less efficient. Therefore, in these cases more effective methods for purification of solvents are necessary.

### Adsorption Methods

In adsorption chromatography (NPLC), control of the water content in solvents is important. In some cases, it is preferred to mix known amounts of dry and water saturated solvents together in order to know or control the percentage of water saturation. On the other hand, addition of activated molecular sieve beads ( $4 \text{ \AA}$  or  $5 \text{ \AA}$ ) to the solvent storage bottle clearly improves their purity and reduces the water content. Impurities, in addition to water, can often be removed by adsorption methods, particularly frontal analysis, which is utilized often in LC.



According to this method, a glass column packed with small adsorbent particles (0.2–0.15 mm diameter), usually comprising dried silica gel or alumina oxide, is used (Figure 4).

The column, before being used for solvent purification, is heated at 473 K for 8 h under vacuum ( $10^{-3}$  Torr) to remove physically adsorbed water. After this operation, the unpurified solvent is injected into the column using a siphon-type injector. The purified solvent is collected in a solvent receiver equipped with a moisture trap.

Frequently, purification and stabilization of the mobile phase, during HPLC investigation, is carried out on precolumns, located between the solvent reservoirs and the injection valve. A precolumn situated

before the analytical column can serve to remove precipitate material when the sample is introduced into the eluent stream at the injector. SPE, a technique employing the principle of frontal analysis, is also used for solvent clean-up. It is mainly used for purifying small amounts of solvent, particularly, for sample dilution.

## Filtration and Membrane Techniques

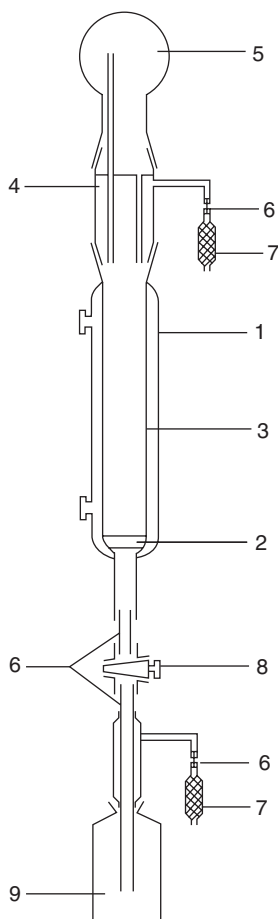
Where impurities are present as microparticulate material filtration affords a convenient technique for solvent purification. The mobile phase containing added buffers or reagents may be filtered through a 0.5  $\mu\text{m}$  or smaller filter to remove particulate matter that can damage the analytical system. The equipment for filtration is simple. Usually, it consists of an Erlenmeyer flask connected to vacuum and a reservoir in which a porous filter disk or membrane is placed. The porous disk is usually made from nonporous spherical glass beads (1–2  $\mu\text{m}$ ) and/or polytetrafluoroethylene (PTFE). Membrane materials are usually made from PTFE, cellulose, or nylon. To improve the efficiency of the separation process, the surface of the filter disks or membrane surface are often modified chemically, similar to that used for chemically bonded packing materials in RP-HPLC and/or SPE. In this case, the surface properties (hydrophobic or hydrophilic) of filters and/or membranes determine the extent of purification possible.

Water may be satisfactorily purified using a compact water purification system that combines filtration, deionization, and charcoal treatment in a convenient, high-volume unit.

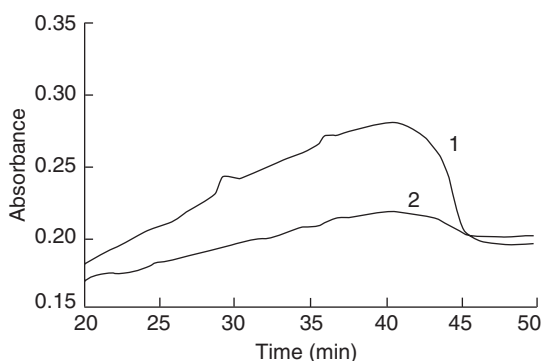
Ultrafiltration provides an alternative to filtration, where large molecules are separated from solution by employing membranes. Membranes are commercially available for separating molecules in the molecular weight range between  $10^3$  Da and  $10^6$  Da. Ultrafiltration is primarily used for isolation of low or high molecular weight substances from solvents. Figure 5 shows a comparison of the efficiency of acetonitrile purification using frontal analysis and the ultrafiltration technique.

Reverse osmosis is a technique similar to ultrafiltration, except that membranes of a much smaller pore diameter are employed, and the operating pressure is higher. The operating pressure must exceed the natural osmotic pressure. This technique has been applied to the purification of water and organic solvents.

In dialysis, solvents are purified by their ability to diffuse through membranes as a result of the



**Figure 4** Apparatus for solvent purification by frontal chromatography: 1, glass column with cooling jacket; 2, glass sinter; 3, silica gel or alumina adsorbent; 4, siphon-type injector; 5, container for unpurified solvent; 6, PTFE tube; 7, moisture trap; 8, valve; 9, container for purified solvent. (Reproduced with permission from Buszewski B, Lodkowski R, and Trociewicz J (1987) Purification of solvents for liquid chromatography. *Journal of High Resolution Chromatography and Chromatography Communications* 10: 527–528.)



**Figure 5** Liquid chromatograms with UV detection using acetonitrile as the mobile phase purified through (1) frontal analysis and (2) ultrafiltration.

**Table 2** Concentration of water in solvent and difference in absorbance for the  $\lambda$  of the cut-off ( $\Delta A$ ) before ( $C_b$ ) and after ( $C_a$ ) purification

Solvent	$\Delta A$ (% v/v)	$C_b$ (mg l <sup>-1</sup> )	$C_a$ (mg l <sup>-1</sup> )
Acetonitrile A	69.7	2216.0	167.0
Acetonitrile B	10.2	297.0	143.0
Benzene	24.6	308.0	27.5
Cyclohexane	16.8	49.0	4.5
<i>n</i> -Heptane	21.3	18.8	7.7
<i>n</i> -Hexane	17.2	19.6	8.6
Methanol A	77.05	650.0	134.0
Methanol B	3.6	120.0	114.0
Tetrahydrofuran	61.3	1081.0	32.6
Toluene	23.1	167.0	10.0

concentration gradient. The overall efficiency of this process is controlled by the ratio of the flow rates and the properties of membrane, fluid channel, and local fluid velocity.

## Recovery and Quality Control

After purification, quality control of solvent purity is necessary. For this purpose, many different analytical methods are utilized. Generally, chromatographic methods such as GC, GC-MS, and HPLC are used. Moreover, UV, infrared, and nuclear magnetic resonance spectroscopy can also be applied but they tend to be less sensitive toward trace impurities. Water in organic solvents is usually determined by Karl-Fisher titration. On the basis of experimental data obtained before and after purification, the efficiency of the clean-up procedure is determined. In general, the efficiency of purification, e.g., the recovery, is expressed by the coefficient  $R$ . This parameter is defined as the ratio of the amount of impurities removed to the amount of solvent before purification:

$$R \pm \sigma = (V_a \pm \sigma_a)(V_b \pm \sigma_b) \times 100\% \text{ [v/v]} \quad [1]$$

or

$$R \pm \sigma = (C_a \pm \sigma_a)(C_b \pm \sigma_b) \times 100\% \text{ [v/v]} \quad [2]$$

where  $V_a$ ,  $C_a$  and  $V_b$ ,  $C_b$  denote volume or concentration of the removed impurities and solvent, respectively, and  $\sigma$ ,  $\sigma_a$ ,  $\sigma_b$  are individual standard deviations.

Table 2 summarizes the results for solvent purification by frontal analysis. In each case, purification improves their absorbance at the  $\lambda$  cut-off point ( $\Delta A$ ). The small improvement in the case of HPLC-grade methanol B is expected for this high-purity

solvent. The high level of impurity (absorbance difference,  $\Delta A = 77.5\%$ ) in analytical grade methanol A precludes its use in HPLC. Similarly, toluene, benzene, tetrahydrofuran, and acetonitrile A tested in the study cannot be used in HPLC without purification.

In spite of the high contribution and frequent use of commercially prepared solvents in laboratory practice, there is a need to apply these solvents with special care. In many cases it is indispensable to apply high-purity 'fresh' portions of solvents prepared in laboratory conditions (called home-made). The possibilities of using different techniques for the purification of solvents depend on laboratory equipment and technical requirements, so as to obtain a product with a purity that meets or even exceeds the guarantees provided by manufactures. Recovery values for purification of water, acetonitrile, alcohols, ketones, and aliphatic and aromatic hydrocarbons obtained by distillation methods are usually in the 97% to 99.92% range. By utilizing membrane techniques for solvent clean-up it is possible to obtain recoveries in a similar range ( $R$  from 97% to 99.90% (v/v)). In the case of halogenated solvents this range is very narrow (i.e., 98–99.80%). In every case one should remember that the factor having much influence on solvent quality (purity) is the storage of the pure solvent (including packaging materials). For this purpose containers of colored glass and with small volume, or nonreactive metal containers, which allow storing of the solvents in an inert gas atmosphere are recommended.

**See also:** Analytical Reagents: Purification. **Distillation.** **Extraction:** Solvent Extraction: Multistage Countercurrent Distribution. **Liquid Chromatography:** Principles; Mobile Phase Selection; Normal Phase.

## Further Reading

- Brock TD (1983) *Membrane Filtration*. Madison: Science Technology.
- Hampel CA and Hawley GG (eds.) (1973/1974) *Handbook of Chemistry and Physics*, 54th edn. Cleveland: CRC Press.
- Higgins RS and Klinger SA (1990) *High Purity Solvent Guide*. Muskegon: Burdick & Jackson.

- Lide R (1995) *Handbook of Organic Solvents*. Boca Raton: CRC Press.
- Minear RA and Keith LH (eds.) (1984) *Water Analysis*, vol. III. Orlando: Academic Press.
- Poole CF and Poole SK (2003) *The Essence Chromatography*. Amsterdam: Elsevier.
- Riddick JA and Bunger WB (1970) *Organic Solvents*, 3rd edn. New York: Wiley.
- Sadek PC (1996) *The HPLC Solvent Guide*. New York: Wiley.

# SPECIATION

See **ELEMENTAL SPECIATION: Overview; Waters, Sediments, and Soils; Instrumentation**

# SPECTRAL DECONVOLUTION

See **CHEMOMETRICS AND STATISTICS: Spectral Deconvolution and Filtering**

# SPECTROELECTROCHEMISTRY

**J A Crayston**, University of St. Andrews, Fife, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Electrochemical techniques in themselves do not give any direct structural information on adsorbates, attached thin films, or electrogenerated compounds. However, the introduction of spectroscopic monitoring of redox reactions within an electrochemical cell has led to vastly improved understanding of electrochemical reactions, especially complex follow-up reactions in solution. It has also offered insights into surface processes such as adsorption and surface reconstruction as well as the double-layer structure. The spectroelectrochemical dimension also introduces specificity when several species are present, and distinguishes Faradaic from non-Faradaic processes. The term spectroelectrochemistry generally refers to an *in situ* method for monitoring electrochemical redox processes and subsequent follow-up reactions. Direct spectroscopic monitoring of an electrochemically generated intermediate can help to identify

products. The spectra of the products can also be used to identify the site of the redox orbital. For example, the site of reduction of a metal complex can be located at a metal or a ligand orbital with the help of techniques such as ultraviolet (UV)–visible and electron paramagnetic resonance (EPR) spectroscopy. With increasingly complex multicentered complexes and organic molecules, as well as the advent of supramolecular chemistry, the location of the redox orbital involved in the electron transfer can often be determined by spectroelectrochemistry.

Regarding surface spectroelectrochemistry, the intense interest in the electrochemistry of single crystal surfaces, adatoms, oxide layers, and monolayers in the last decade has seen not only the enhancement of traditional methods (e.g., infrared (IR) and Raman microspectroscopy) but also the development of relatively new surface specific methods such as second-harmonic generation, sum-frequency generation, surface plasmon resonance (SPR), and surface-enhanced resonance spectroscopy (SERS). The increased access to synchrotron radiation has led to X-ray absorption and X-ray diffraction methods becoming more feasible as *in situ* techniques for thin films or species generated close to the electrode. In this article, for

## Further Reading

- Brock TD (1983) *Membrane Filtration*. Madison: Science Technology.
- Hampel CA and Hawley GG (eds.) (1973/1974) *Handbook of Chemistry and Physics*, 54th edn. Cleveland: CRC Press.
- Higgins RS and Klinger SA (1990) *High Purity Solvent Guide*. Muskegon: Burdick & Jackson.

- Lide R (1995) *Handbook of Organic Solvents*. Boca Raton: CRC Press.
- Minear RA and Keith LH (eds.) (1984) *Water Analysis*, vol. III. Orlando: Academic Press.
- Poole CF and Poole SK (2003) *The Essence Chromatography*. Amsterdam: Elsevier.
- Riddick JA and Bunger WB (1970) *Organic Solvents*, 3rd edn. New York: Wiley.
- Sadek PC (1996) *The HPLC Solvent Guide*. New York: Wiley.

# SPECIATION

See **ELEMENTAL SPECIATION: Overview; Waters, Sediments, and Soils; Instrumentation**

# SPECTRAL DECONVOLUTION

See **CHEMOMETRICS AND STATISTICS: Spectral Deconvolution and Filtering**

# SPECTROELECTROCHEMISTRY

**J A Crayston**, University of St. Andrews, Fife, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Electrochemical techniques in themselves do not give any direct structural information on adsorbates, attached thin films, or electrogenerated compounds. However, the introduction of spectroscopic monitoring of redox reactions within an electrochemical cell has led to vastly improved understanding of electrochemical reactions, especially complex follow-up reactions in solution. It has also offered insights into surface processes such as adsorption and surface reconstruction as well as the double-layer structure. The spectroelectrochemical dimension also introduces specificity when several species are present, and distinguishes Faradaic from non-Faradaic processes. The term spectroelectrochemistry generally refers to an *in situ* method for monitoring electrochemical redox processes and subsequent follow-up reactions. Direct spectroscopic monitoring of an electrochemically generated intermediate can help to identify

products. The spectra of the products can also be used to identify the site of the redox orbital. For example, the site of reduction of a metal complex can be located at a metal or a ligand orbital with the help of techniques such as ultraviolet (UV)–visible and electron paramagnetic resonance (EPR) spectroscopy. With increasingly complex multicentered complexes and organic molecules, as well as the advent of supramolecular chemistry, the location of the redox orbital involved in the electron transfer can often be determined by spectroelectrochemistry.

Regarding surface spectroelectrochemistry, the intense interest in the electrochemistry of single crystal surfaces, adatoms, oxide layers, and monolayers in the last decade has seen not only the enhancement of traditional methods (e.g., infrared (IR) and Raman microspectroscopy) but also the development of relatively new surface specific methods such as second-harmonic generation, sum-frequency generation, surface plasmon resonance (SPR), and surface-enhanced resonance spectroscopy (SERS). The increased access to synchrotron radiation has led to X-ray absorption and X-ray diffraction methods becoming more feasible as *in situ* techniques for thin films or species generated close to the electrode. In this article, for

reasons of space, the emphasis has been placed largely on the *in situ* spectroscopy of solution species or electrogenerated species in a phase close to the electrode, such as in a thin film or polymer. These techniques are listed in Table 1. Omitted from this entry is a discussion of *ex situ* surface science measurements such as high-vacuum surface-specific spectroscopies: X-ray photoelectron spectroscopy (XPS), Auger electron spectroscopy, secondary ion mass spectrometry (SIMS), low-energy electron diffraction spectroscopy (LEEDS), high-resolution electron energy loss spectroscopy (HREELS). Other *in situ* measurements, such as gravimetry, impedance spectroscopy, and imaging techniques such as AFM and STM, have undergone spectacular growth over the last two decades, but do not yet involve radiation interacting with the electrode. But the coupling of many, if not all, spectroscopic probes to the scanning probe methods is certainly not far away. This will yield unprecedented imaging information. And, as with many other analytical techniques, spectroelectrochemistry is developing rapidly in response to better and less-expensive instrumentation and data handling, including the application of chemometric methods.

## UV-Visible Spectroscopy: Solution Species

The widespread use of the optically transparent thin-layer electrochemical (or OTTLE, Figure 1) cell led to the early growth in popularity of UV-vis spectroelectrochemistry. The OTTLE offers a way to measure both the redox potential ( $E^{0'}$ ) and the number of electrons transferred in the charge transfer process ( $n$ ) without requiring knowledge of the electron transfer kinetics, the diffusion coefficient, or, indeed, the concentration. A plot of the absorbance ratio versus potential ( $E$ ) should then be a straight line with a  $y$ -axis intercept giving  $E^{0'}$  and the slope giving  $n$ . OTTLEs have frequently been used in determining an accurate value for the redox potential,  $E^{0'}$ , of biological molecules with slow heterogeneous electron transfer rates. In these cases, a soluble redox mediator may be used to promote electron transfer in an indirect coulometric titration. OTTLEs are also used to study inorganic complexes with slow redox processes, for example, Cu(II) or Cu(I). Note that a great attraction of the technique is that neither the extinction coefficients nor the pathlength, or even the initial concentration are required as long as the limiting absorbances in each redox state (at sufficiently reducing and oxidizing potentials) can be measured.

The great advantage of the thin-layer design is that the small volume of analyte within the layer is

electrolyzed extremely rapidly, allowing the equilibrated spectrum to be recorded within seconds or minutes of applying the potential. The working electrode in these cells is usually a gauze, micromesh, or minigrid. Meshes of thickness as little as 5  $\mu\text{m}$  with up to 2000 lines per inch with 36% transmission are available. The thickness of the cell is determined by measuring the absorbance of a known compound and the volume by chronocoulometry. Typical values are 50  $\mu\text{m}$  and 0.09  $\text{cm}^3$ , with a characteristic time for complete electrolysis of 200 s for 2–4  $\text{mmol l}^{-1}$  solution. Such a cell would exhibit undistorted semiinfinite cyclic voltammograms (CVs) below  $\sim 20 \text{ mV s}^{-1}$  at concentrations less than 4  $\text{mmol l}^{-1}$ . Variable temperature cells generally have plenty of free volume to allow for expansion of the solvent and to prevent high pressures building up, leading to leaks.

The voltammetric response of the OTTLE cell should ideally be of the thin-film type with a peak separation close to zero and a symmetrical shape. The thin layer has restricted diffusion and exchange with the bulk solution, which leads to high uncompensated resistance and distorted responses at all but the lowest scan rates. OTTLE CV waves are often distorted with greater peak separations in the CVs than expected. Minimization of the  $iR$  (ohmic) drop is particularly applicable in the design of these thin-layer cells. Edge effects (i.e., diffusion to the edges that prevent the current in the cell reaching zero at long times – typically 1% residual current – is regarded as acceptable) can be minimized by limiting the optical area to a small central portion of the minigrid.

A thicker cell is possible if reticulated vitreous carbon (RVC) is used. For example, transmission of 13–45% for 100 p.p.i. (pores per inch) is possible for a 1.2–0.5 mm thick RVC plate. LIGAs are Lithographic-Galvanic structures that present some advantages over indium tin oxide (ITO) and metal meshes, namely improved stability, faster response times, and a greater range of electrode materials that can be used in the same cell. The electrodes are sandwiched between quartz rods coupled to the spectrometer using fiber optics.

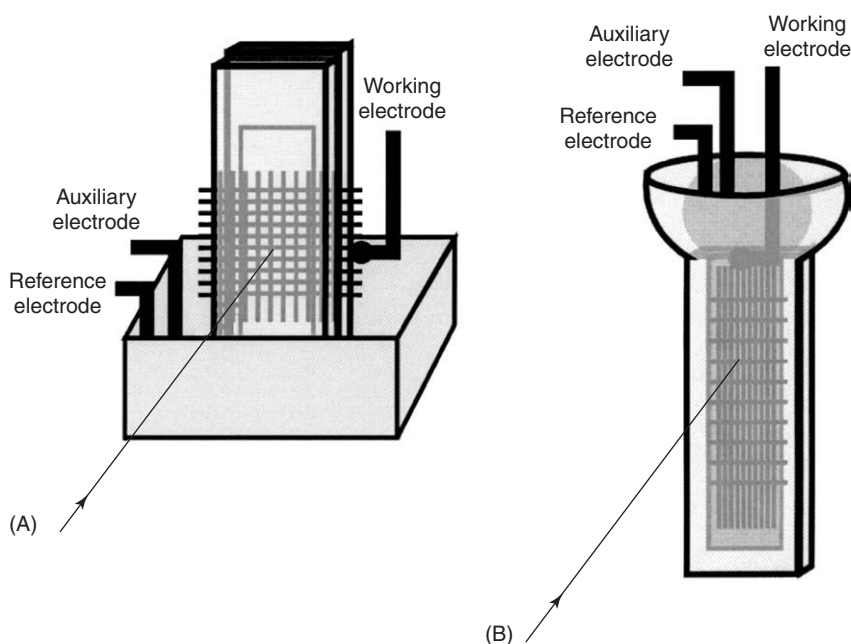
For a sequential reaction without any branching pathways (such as decomposition of a reactive intermediate), one would expect to see isosbestic points in the thin-layer electrolysis. It is advisable to carry out several oxidation–reduction cycles to test the stability of the spectral responses; in other words, to see if the starting material is always fully regenerated.

Like all thin-film cells, the OTTLE cell is useful for studying the slow follow-up reactions of the electrogenerated species ( $>20$  s) since diffusion and

**Table 1** Comparison of spectroelectrochemical methods

<i>Name</i>	<i>Sensitivity</i>	<i>Structural information</i>	<i>Surface selectivity</i>	<i>Time resolution</i>	<i>Applications/remarks</i>
UV–Vis OTTE/OTE	•••	•	•	•••	Redox potentials; <i>n</i> -values of solute redox species with or without mediator. Kinetics. Adsorption
UV–Vis reflection	•••	•	•	•••	Solution species at nontransparent electrodes. Faster response, e.g., DVCA
UV–Vis ATR	•••	•	•	•••••	Solution species at thin film coated prism. Fast response
UV–Vis LOPTLC	••••	•	•	••	Weak absorbers; adsorption. Can use opaque electrodes
UV–Vis photothermal	••••	•	••	•••••	Nonreflective; nontransmissive electrodes. Surface sensitive if high surface area electrode
Ellipsometry	•••••	••	•••	•••	Growth of multimonolayer; thick oxide or polymer films
ER/PMRS	•••••	•	•••••	•••	Sub-monolayer adsorption/deposition; double layer. Single crystals
SPR	•••••	•	•••••	•••	Submonolayer; adsorption. Limited to coinage metal films. Not single crystal electrodes
SHG	•••••	•	•••••	•••••	Submonolayer adsorption; double-layer. Electrode (incl. Single crystal)–liquid; liquid–liquid interfaces. Symmetry of adsorbed layer
FTIR OTTE	•••	••••	•	•	Greater information content than UV–Vis
EMIRS/PMIRRAS	•••••	••••	••••	••	Adsorption/deposition at single-crystal electrodes
Resonance Raman	••	••••	•	••	Solution species at greater sensitivity than IR. Aqueous media
SERRS	•••••	••••	•••••	••	Surface species. Adsorbates. Aqueous media
SFG	••••	••••	•••••	•••••	Submonolayer adsorption; double layer. Electrode (incl. Single crystal)–liquid; liquid–liquid interfaces
EPR	•••	••	•	•••	Solution species
NMR	•	•••	•	•	Difficult technique for liquids. Local structural info for solids. Amorphous powder electrodes
XAS (bulk)	•	•••	•	•	Local structural information. Solid-state electrochemistry, e.g., batteries
XAS (surface)	•	•••	•••	•	Local structure of surface atoms; single crystals
X-ray standing wave	•	•	•••••	•	Structure of double layer and polymer overlayers
XRD (bulk)	•	•••	•	•	Long-range structure. Solid-state electrochemistry, e.g., batteries
XRD (surface)	•	•••	•••	•	Detailed structure of surface. Single crystal electrodes
Mass spec (DEMS)	••	•••	•	••	Gaseous products, e.g., fuel cell electrochemistry
Mass spec (thermospray)	••	•••	•	••	Nonvolatile products of electrochemical reactions
Mössbauer	•	•••	•	•	Solid-state electrochemistry, e.g., batteries





**Figure 1** (A) Traditional OTTLE cell. The minigrid is sandwiched between two microscope slides and the spacing is defined by Teflon tape. The three exposed edges are sealed with epoxy adhesive. (B) More robust, modern version using a commercial 0.5 mm UV cuvette and gauze.

convection effects do not interfere. But the relatively long time to achieve equilibration of the stagnant thin layer of the OTTLE or reflection cell does not allow one to study the kinetics of faster reactions. To overcome this, the forced convection regimes of the rotating disc electrode (RDE) and the channel-flow cell allow steady-state currents to be produced, which provides access to shorter time regimes. Both approaches have been investigated with optical detection of electrogenerated species. The rotating optically transparent electrode has been used mainly for photogeneration of species rather than spectroscopy. On the other hand, a rotating disk surrounded by an optically transparent, insulating ring has been constructed for the spectroscopic determination of intermediates. Alternatively, a light beam can be introduced by fiber optics near-normal to the RDE and reflected from the electrode. In the case of the channel flow cell the steady-state current is proportional to  $V^{1/3}$ , where  $V$  is the flow rate, and hence the absorbance should be given by  $i/V$ , i.e., proportional to  $V^{-1/3}$ . The absorbance also follows the Nernstian response of the current. For faster reactions still, chronoabsorptometry at an optically transparent electrode is used (see below).

### Chronoabsorptometry and Derivative Cyclic Voltabsorptometry

The technique of chronoabsorptometry, also known as potential step chronoamperospectrometry and

chronocoulopectrometry, monitors the time dependence of the absorbance. Here, the optically transparent electrode (OTE) is used in a thick-layer cell so that semiinfinite diffusion conditions prevail. Such cells generally use a gold or ITO (indium-doped tin oxide)-coated electrode. Recently, boron-doped diamond has shown great promise as a new electrode material, but its transparency has not yet been exploited in UV-vis spectroelectrochemistry. The absorbance of an electrogenerated species,  $R$ , at a wavelength where only  $R$  absorbs, increases with time ( $t$ ) as  $t^{1/2}$ . Unlike chronoamperometry, there is no need to determine the electrode area, or  $n$  (the number of electrons passed). The slope of the absorbance-time plot allows the determination of the (usually) unknown extinction coefficient  $\epsilon_R$  for the electrogenerated species, providing that  $c$  and  $D_O$  are known. No knowledge of the electrogenerated species' diffusion coefficient  $D_R$  is necessary. Usually, the concentration is known and it is a simple matter to determine  $D_O$  from a chronoamperometry, chronocoulometry, or ultramicroelectrode experiment. The chronoabsorptometric response has been used to determine the heterogeneous rate constant for electron transfer to biological redox centers. The advantage of the optical technique over conventional electrochemical methods is that the measurements can be made successfully in the presence of other Faradaic or charging processes.

Derivative cyclic voltabsorptometry (DCVA) is the optical analog of CV. Usually a fixed wavelength is

monitored while the potential is cycled (typically at scan rates up to  $50 \text{ mV s}^{-1}$ ). If several species are being monitored by DCVA it is important to remember that peak heights are proportional not only to concentration but also to the molar extinction coefficients, which in general are unknown. DCVA provides a way to remove the effects of Faradaic processes that do not involve the colored species of interest.

### **Near-Parallel Configuration: Long Optical Pathlength Thin-Layer Cell and Spatially Resolved Spectroelectrochemistry**

The optical pathlength in an OTTLE cell is short, giving low optical sensitivity. With the long optical pathlength thin-layer cell (LOPTLC), the light beam of the spectrometer passes parallel to the electrode surface, thereby sampling a greater pathlength (typically 1 cm) of the region immediately in front of the electrode (up to 0.5 mm). Enhancements of  $\sim 100$ -fold are observed due to the increased surface-area to volume ratio. Another advantage is that, as in reflection spectroscopy, there is no requirement for the electrode to be transparent. The electrode may be mounted vertically or horizontally, the former more practical for modern spectrometer slits. An unusual design uses a block of graphite or glassy carbon with a  $500 \mu\text{m}$  hole drilled through it. The cylindrical geometry facilitated the simulation of the time response. The LOPTLC is largely used in the same manner as the OTTLE (i.e., bulk electrolysis), but it is also possible to compute the time-dependent absorbance either analytically or numerically.

If spectra of surface adsorption or deposition processes are not desired in a study it may be worth considering a method in which the light beam passes parallel to the electrode surface such as LOPTLC or diffusion-layer imaging techniques. Grazing-angle laser reflection with  $5 \mu\text{m}$  resolution is useful for diffusion-layer imaging. However, to some extent, diffusion-layer imaging such as this has been eclipsed by the advent of electrochemical STM techniques.

Probe-beam deflection is a technique in which a monochromatic source, typically a He-Ne laser beam, passing parallel to the electrode surface, is used to monitor refractive index changes in the diffusion layer. It is a simple and cost-effective way of profiling the diffusion layer, and is used to monitor the diffusion layer ingress and egress of ions, particularly protons. It is often used in conjunction with electrochemical quartz crystal microbalance (EQCM) measurements that monitor mass changes within or on the electrode layer itself.

### **Other UV-Visible Techniques**

When monochromatic light is absorbed by an electrode surface, it will heat up imperceptibly leading to small temperature changes near the electrode. Such temperature changes can be monitored by photothermal or photoacoustic spectroscopy. In the former, a thermistor is attached to the back of the electrode. Alternatively, the temperature changes are monitored indirectly using either a piezoelectric ceramic detector attached to the electrode surface in order to monitor the dimensional changes occurring as a result of absorption, or by monitoring refractive index changes in the electrolyte (thermal probe beam deflection; similar to probe beam deflection above). If the light is modulated (chopped) at acoustic frequencies, sound waves will be generated near to the electrode that are then detected by a microphone (photoacoustic spectroscopy). Though useful for nonreflective or nontransmissive surfaces such as powders, these techniques have not found widespread use in electrochemistry, though it seems that photothermal spectroscopy with optical detection of the refractive index change is set for a comeback due to its ultrafast response and potential in IR microspectroscopy imaging.

If the electrode is a semiconductor, or if it forms a semiconducting oxide layer (e.g., Fe or Pb), light absorption may be directly monitored by photocurrent spectroscopy. Such photocurrents arise from minority carriers and are much larger than photoemission effects at metallic electrodes.

UV-vis methods are often combined with other techniques. For example, combined UV-vis and EPR spectroelectrochemistry was used in the study of anthraquinone reduction. But the EPR spectroelectrochemical cell design is demanding to implement successfully and the presence of the metal electrodes within the cell reduces sensitivity considerably (see later).

### **UV-Visible Spectroscopy of Surface Species: Specular Reflectance, Ellipsometry, Internal Reflection, and Waveguides**

Sometimes the desired electrode cannot be rendered transparent. Then, the spectroelectrochemistry of solution species may be performed using a reflective electrode (Figure 2). The results are, of course, identical to those in the transmission mode, except that the path length and hence the absorbance of dissolved species is increased by the factor  $(2/\cos \theta)$  where  $\theta$  is the angle of incidence (angle between the

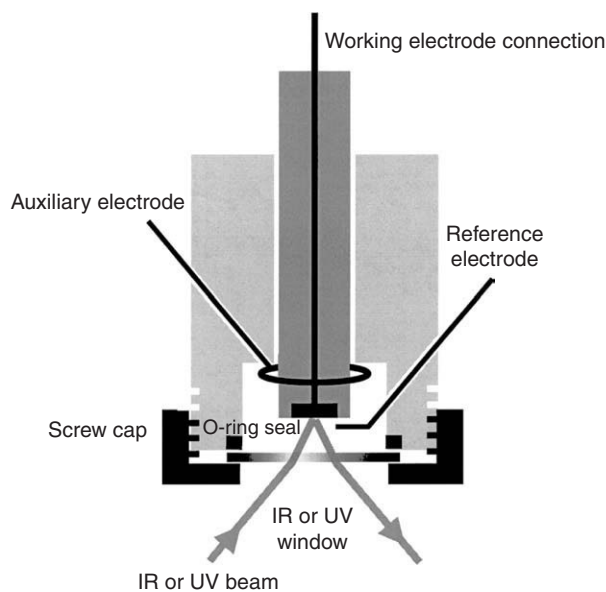
beam and the normal to the electrode surface). Typically,  $\theta$  is  $\sim 45^\circ$  so as to avoid oversampling of the diffusion layer.

If large surface selectivity is required, it is necessary to consider the physics of the reflection process. Consider only the electric vector of the electromagnetic radiation and ignore the perpendicular magnetic

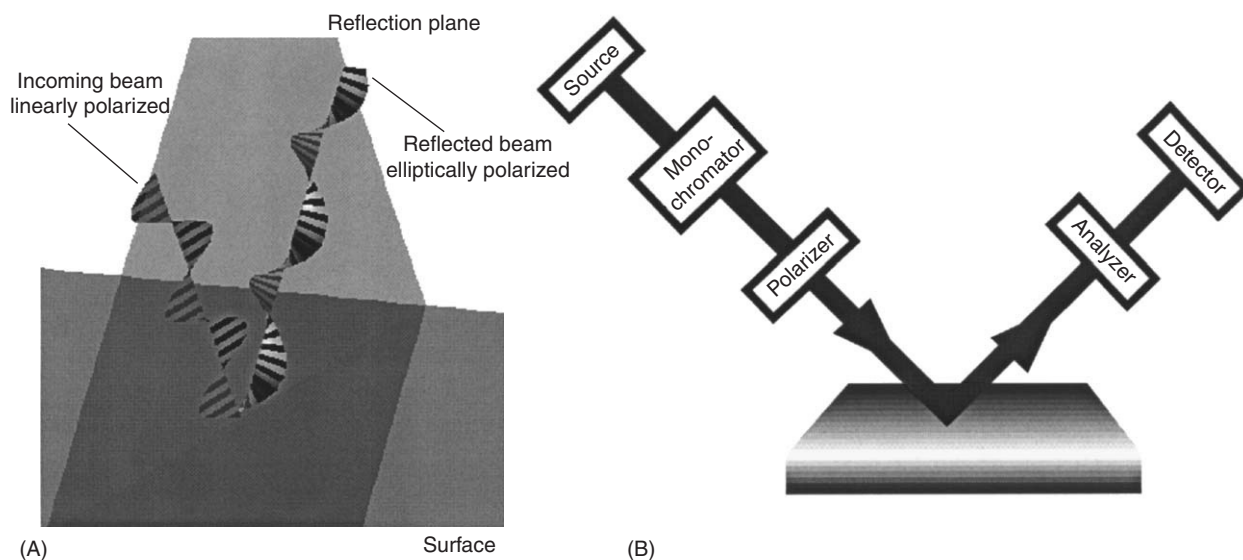
vector. For unpolarized light, the electric vector lies in all possible planes. Light with a component polarized perpendicular to the plane defined by the incident and reflected beams (the plane of reflection), i.e., parallel to the electrode, is termed the s-polarized component. Light polarized parallel to the plane of incidence is termed the p-polarized component. Both components will, in general, undergo different phase shifts and degrees of attenuation upon reflection, dependent on the angle of incidence and the refractive indices of the media.

The difference in phase, along with the ratio of the s- and p-polarized light intensities, is measured in the technique of ellipsometry, so called because the combination of two light vectors with a phase lag and unequal intensity, in general, leads to the electric vector tracing out an elliptical path in the direction of propagation (Figure 3A).

An ellipsometer (Figure 3B) measures the phase lag  $\Delta$  between the reflected s- and p-polarized light and the ratio of the extinction coefficients  $k_p$  and  $k_s$  ( $\tan \Psi = |k_p|/|k_s|$ ). The observed changes in intensity and phase as a result of the reflection at electrode surfaces depend on the refractive indices  $n_i$  of the various media,  $n = c/c_0 = (\epsilon\mu)^{1/2}$  where  $\epsilon$  is the dielectric constant and  $\mu$  the magnetic permeability, as well as on the absorbance of the media. The effect of the absorbance is often combined with the refractive index to yield a complex refractive index,  $\hat{n} = n - jk$  where  $k$  is the extinction coefficient. Similarly, a complex dielectric constant,  $\hat{\epsilon}$ , is given by  $\hat{n} = (\hat{\epsilon}\mu)^{1/2}$  with  $\hat{\epsilon} = \epsilon' - j\epsilon''$  where  $\epsilon' = (n^2 - k^2/\mu)$  and  $\epsilon'' = 2nk/\mu$ . Then, most optical materials are defined by three parameters  $\mu$ ,  $n$ , and  $k$ , or alternatively by  $\mu$ ,  $\epsilon'$ ,



**Figure 2** Reflection cell for solution/surface UV-vis or IR measurements. The body is fabricated from glass of Teflon. A similar cell is used in methods such as SHG. The cell may be vertically (avoids gas bubbles) or horizontally mounted. The light is delivered and collected by a reflection accessory or a fiber optic bundle. The working electrode may be mounted on a micrometer for greater control of the spacing between it and the window.



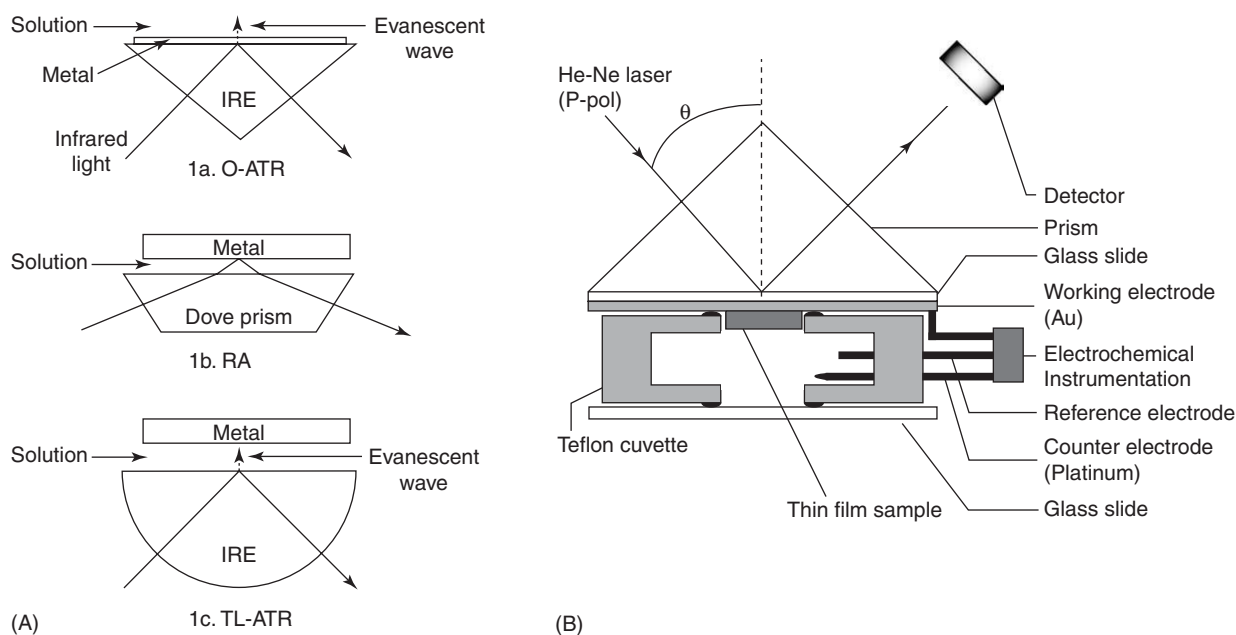
**Figure 3** (A) Diagram depicting the change from linearly to elliptically polarized light upon reflection at an electrode surface. (B) Schematic diagram of an ellipsometer setup.

and  $\epsilon''$ . Most reflectance and ellipsometry studies on growing surface coatings of thin films are modeled assuming the appearance of a third phase between the metal and the electrolyte due to the adsorbate, and the application of Fresnel's equations. The fitting of the data is carried out by computer. With two observed parameters the problem is underdetermined. In the past assumptions had to be made about the absorbance or the refractive index of the thin film, but nowadays the ratio of the intensity of the incident and reflected light is measured, or alternatively the angle of incidence and/or incident wavelength (spectroscopic ellipsometry) are varied, so removing any ambiguity about the refractive index of the new phase.

Due to its erstwhile greater speed in comparison to ellipsometry, specular reflectance using s- and p-polarized beams in the UV-vis region is a good method for studying surface adsorption and for rapidly determining the optical properties of substrates as a function of wavelength, with reflection changes of the order of  $10^{-5}$ – $10^{-6}$ . Unfortunately, in systems of electrochemical interest the reflectance changes are very small. To improve the situation, multiple reflections or glancing incidence are used together with chopping the light and phase-sensitive detection in order to measure absolute reflectivities. However, it is much easier to measure changes in reflectivity instead by using potential modulation (a small sinusoidal signal is superimposed) and phase-sensitive

detection in the technique known as electroreflectance, electrolyte electroreflectance, or potential-modulated reflection spectroscopy. Electrochemists were surprised to find very large changes in the apparent reflectance of noble metals using this method even in the absence of adsorption. This was traced to the modulating potential-induced motions of the free electrons in a layer 0.1 nm thick near the surface, giving it very different optical characteristics to the bulk. Under conditions in which this 'third phase' is forming the reflectance changes linearly with the amount of adsorbate or new phase, and it is found that the s-polarized beam intensity is half the p-intensity at  $\theta = 45^\circ$ . Electroreflectance can be very sensitive to surface layers; for example, in monitoring the production of adsorbed CO on single-crystal electrodes used to model fuel cell reactions.

An alternative reflection setup makes use of single (or multiple) internal reflection within an OTE (Figure 4). However, at each internal reflection, a small portion of the intensity 'leaks out' (it is correctly known as an evanescent wave) into the thin film electrode layer and beyond into the solution and can be used to detect any absorbing species. The method is also known as attenuated total reflection (ATR). The evanescent wave intensity decays exponentially as  $\exp(-\delta/x)$  with distance  $x$  from the interface. The penetration depth,  $\delta$ , depends on the wavelength and the optical properties of the substrate, electrode film, and solution:  $\delta = \lambda/(4\mu \text{ Im } \xi)$



**Figure 4** (A) (Top) Otto ATR configuration; (middle) reflection-absorption; (bottom) Kretschmann ATR configurations. (B) Electrochemical cell for thin film SPR spectroscopy using the Kretschmann configuration. (Reprinted with permission from Xia CJ, Advincula RC, Baba A, and Knoll W (2002) *Langmuir* 18: 3555–3560. © 2002, American Chemical Society.)

where  $\text{Im } \xi$  is the imaginary part of  $(\hat{n}_3 - n_1^2 \sin^2 \theta_1)^{1/2}$  where  $\hat{n}_3$  is the complex refractive index of the solution,  $n_1$  is the index of refraction of the substrate, and  $\theta_1$  is the angle of incidence within the substrate. Typically,  $\delta$  is the order of a quarter of a wavelength, 50–200 nm for UV–vis light. Usually, the absorption is small and so this is a very good technique for highly absorbing solutions. In other situations, the sensitivity is too low and multiple internal reflections are required. It is a useful technique for the determination of the rates of very fast, second-order follow-up reactions, since these will cause rapid loss of colored electrogenerated species before it has diffused very far from the electrode. Perhaps, however, the greatest advantage of ATR over other IR methods is its ability to sample only small path-lengths of strongly IR-absorbing solvent. Recent studies have looked at the potential of optical waveguides as internal reflection elements. Although the sensitivity is enormously increased due to the increased light intensity, by their very nature waveguides are limited to monochromatic probe wavelengths.

Reflection spectroscopy has also been carried out successfully by reflection off a water/immiscible liquid interface. Total internal reflection at a liquid–liquid interface can be used to monitor ion-transfer across the interface. For example, the kinetics of the reduction of TCNQ and the oxidation of 1,1'-dimethylferrocene by  $[\text{Fe}(\text{CN})_6]^{3-}$  in the aqueous phase have been considered. The kinetics of these reactions were studied by chronoabsorptometry, assuming diffusion control. Ion-transfer kinetics across interfaces have been treated theoretically and applied to the study of indicator transfer.

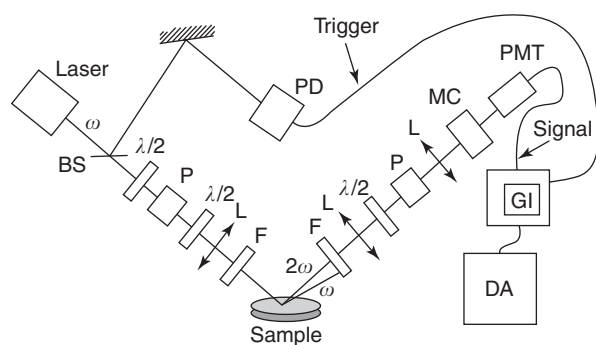
## Surface Plasmon Resonance

At a particular combination of refractive indices and incident angles an incident beam may come into resonance with the surface plasmons due to electron motion in the electrode. The minimum in the intensity (SPR minimum) is located precisely by varying the angle of incidence over a small range. Commercially available instruments generally employ the Kretschmann configuration as described above for ATR (Figure 4). The prism serves to fulfill the SPR condition in terms of refractive index and incident angle. The SPR minimum is observed over a very narrow range of incident angles (1–2°) and its shape and angular position are very sensitive to the optical constants and hence the presence of absorbates. Either the refractive index can be measured, or assuming a model for the monolayer, the extent of

coverage or thickness of the adsorbed layer can be determined. Other forms of spectroscopy such as luminescence or light-scattering are possible using SPR. Usually, a He–Ne laser is used and the angle is varied (though fixed angle, varying wavelength mode is also used). Alternatively, to eliminate the need for moving parts, a wedge of laser light is delivered to the surface and the reflected light is monitored using an area charge-coupled device (CCD) detector, as in some commercial instruments for bioassays. Unlike SERS, SPR uses a thin gold film as substrate or electrode because this offers far greater reproducibility than the roughened SERS electrode. It has been successfully used in monitoring the adsorption of biological molecules, self-assembled monolayers and double-layer structure. Greater versatility is obtained if gold films deposited on microscope slides are pressed against the prism. Waveguides prepared using multilayers on the substrate can be used to amplify the effect. The surface can be imaged by diverging the laser beam and detecting using a CCD detector. Very small refractive index changes can be measured, such as those caused by an electric field gradient within a thin film. SPR spectroscopy gives complementary information to EQCM measurements and the two techniques have been combined effectively. Despite its attractions, SPR places some restrictions on the electrode, but greater flexibility is possible using the Otto configuration (Figure 4).

## Second Harmonic Generation

When an intense light beam is directed to a noncentrosymmetric crystal then light may be reflected not only at the incident frequency,  $\omega$ , but also at the second harmonic frequency  $2\omega$ . Such second harmonic generation (SHG) is an example of nonlinear behavior, and is the basis for optical components such as frequency doublers. Since a surface or interface is inherently noncentrosymmetric, light from the electrode–liquid interface may show evidence for SHG (Figure 5). The intensity of the SHG light is dependent on the square of the incident intensity, and also on the square of the second-order nonlinear optical susceptibility, which in turn depends on the asymmetry of the interface. The effect is both angle and wavelength dependent. Although the effect is weak, typically  $10^{-12}\%$  of the incident intensity, requiring pulsed laser excitation and gated detection methods, the fact that the  $2\omega$  signal can only arise from the surface region lends the technique the advantage of surface specificity and blindness to the bulk solution. The  $2\omega$  signal intensity is sensitive to the presence of absorbed species and, in particular,



**Figure 5** Schematic SHG setup. Key: BS = beamsplitter, L/2 = half-wave plate, P = polarizer, L = lens, F = filter, MC = monochromator, PMT = photomultiplier tube, GI = gated integrator, DA = data acquisition, PD = photodiodes. (Reprinted with permission from Marrucci L, Paparo D, Cerrone G, *et al.* (2002) *Optics and Lasers in Engineering* 37: 601–610. © 2002, with permission from Elsevier.). The cell generally uses a reflection cell similar to that in **Figure 2** with a Dove prism (**Figure 4A** (middle)).

the rearrangement of adsorbed molecules (e.g., CO) or ions on the single-crystal electrode surfaces. It is also not particularly demanding of the nature of the metallic electrode (in contrast to SERS or SPR) and so metals such as platinum or mercury can be used. Furthermore, since the excitation source is visible light, less light is absorbed by the solvent and so longer path length cells can be used. In addition to simply measuring the intensity of the SHG beam as a function of potential or time, it is possible to derive molecular orientation information from the polarization dependence of the intensity, and symmetry information (if different from the surface) by rotating the sample or plane of polarization about the surface normal.

### Vibrational Spectroelectrochemistry: Fourier Transform Infrared Raman, and Sum Frequency Generation

Vibrational spectroelectrochemical techniques, particularly Fourier transform infrared (FTIR), with the advent of less-expensive and more sensitive spectrometers, have enjoyed a great increase in popularity over the last few years. The advantages over UV–vis spectroscopy include greater specificity and enhanced information content. An IR OTTE transmission cell with IR-transparent windows can be built by modifying a commercial liquid cell. Raman spectroscopy is also very easily carried out with an OTTE cell, but a flow cell is more useful for SERS studies of adsorption, since the solution can be replaced without disturbing the optical alignment.

Another popular configuration for FTIR is the reflectance geometry, which is clearly necessary when the electrode cannot be rendered transparent or the focus of interest is not the solution electrogenerated species but thin films growing on the surface of the electrode itself. A retractable electrode allows operation under both thin layer (15  $\mu\text{m}$ , 1 s electrolysis time) and semiinfinite conditions (**Figure 2**). An external reflection accessory is required to deliver the light to the electrode surface. This geometry is also used for Raman spectroelectrochemistry. Since solvent absorption is a usually major problem (though much less so for water solvent using Raman spectroscopy) the electrode is usually capable of close approach to the window surface in order to minimize the absorption due to the electrolyte. In early IR studies of surface processes the IR beam was monochromatic and the potential stepped between a background (where no Faradaic processes are occurring) and the working potential. This is called potential modulated or electrochemically modulated infrared spectroscopy (EMIRS). Alternatively, at fixed potential, the polarization of the beam can be modulated between s- and p-polarizations (polarization-modulated infrared reflection–absorption spectroscopy (PM-IRRAS) or simply (IRRAS). As mentioned above for the visible region, the s-polarized and p-polarized beams undergo different phase changes and attenuation on reflection. The signal is plotted as  $\Delta R/R = (I_p - I_s)/(I_p + I_s)$ . It turns out also that in the IR region the s-polarized light undergoes a phase lag on reflection  $\sim 180^\circ$ , leading to destructive interference between the incident and reflected beams. This means that there is a greater intensity of p-polarized light available for absorption by molecules and thus only vibrations with a dipole moment change perpendicular to the electrode are selected. This is the ‘surface-selection rule’, and operates in addition to the usual IR selection rule. Of course, since the solution is isotropic, no differential signal is observed from solution species.

PM-IRRAS can be implemented using modern commercially available FTIR spectrometers. Indeed, most modern surface IR studies simply use these instruments in conjunction with some form of subtraction. In subtractively normalized FTIR or potential difference IR, a single beam FTIR spectrum  $S_r$  is collected at a reference potential  $E_r$  (at which no Faradaic processes occur) and then at successively higher or lower potentials  $E_s$  scans  $S_s$  are obtained. Then plots (usually overlaid) of  $\Delta R/R = (S_s - S_r)/S_r$  are obtained for a range of working potentials. Unlike EMIRS or PM-IRRAS, there is no necessity for the electrode process under study to be reversible. The method is very popular, capable of investigating



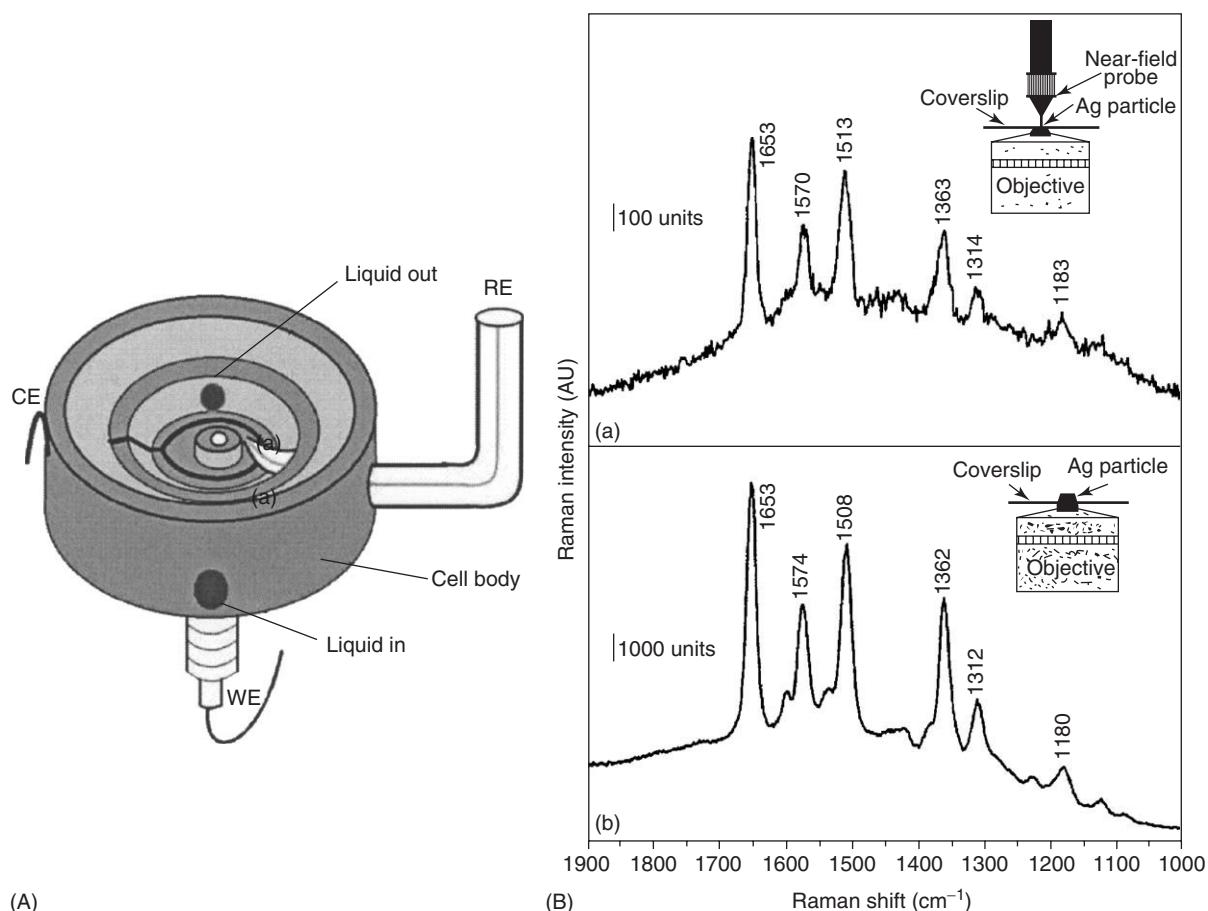
a wide range of electrode materials, but in contrast to the other methods it does not give absolute spectra and suffers from incomplete subtraction of strong absorbance bands due to solvent and electrolyte entering and leaving the thin-layer region. Often thin-layer cells must be used to reduce solvent absorption, and these introduce additional problems such as resistance effects and rapid electrolysis of solutes within the thin layer. If a thin film of gold or silver is evaporated onto the surface to provide islands for adsorption, it is possible to observe surface-enhanced IR absorption from adsorbates. This effect is closely related to the more familiar SERS effect in Raman spectroscopy.

In order to minimize the influence of strong solvent absorption, the IR beam can be internally reflected within an OTE, as for UV-vis internal reflection or ATR described above. The difference is the necessity for an IR-transparent OTE, preferably with a high refractive index, such as Si, Ge, or KRS-5 coated by a thin film electrode, typically gold or Pt. As explained above for the UV-vis version of this method, the penetration depth is related to the crystal refractive index, and is typically of the order of 1  $\mu\text{m}$  for IR radiation. This means that signals from the adsorbate, solute, and solvent are of comparable magnitude, and so the solvent is more easily subtracted. Also, there is no need for a thin layer configuration so mass transport is improved. Multiple internal reflections can be used to maximize the signal, and, as in external reflection, the potential may be modulated, or co-addition of several FTIR scans at working and background potentials may be used to improve the sensitivity. The technique obviously suffers from a lack of versatility in the choice of electrode substrate but the Otto configuration, in which the electrode is brought up to the uncoated crystal, may overcome this constraint.

The normal Raman effect is very weak: as little as  $10^{-12}$  of the incident intensity appears as Raman scattering. Although usually a scan is produced as in FTIR, Raman spectra may also be obtained at fixed wavenumber for time-dependent studies, with coaddition of large number of scans obtained by stepping between two potentials and allowing time for equilibration between cycles. The final signal is analyzed in a similar way to chronocoulometric techniques. This, of course, requires a reversible system. If more than one wavelength needs to be monitored then instruments are available with optical multichannel analyzer detection. The Raman signal can be very much greater if resonance enhancement or surface enhancement (SERS) is exploited. SERS, in particular, at a roughened Ag or Au electrode can lead to an enhancement of  $10^5$ – $10^6$ . The electrode is roughened

by a series of 'oxidation–reduction' cycles designed to produce surface features with nanometer dimensions. The enhancement arises from two separate effects: (1) the electromagnetic effect, whereby the nanometer structures concentrate the electric field through surface plasmon waves; (2) the chemical effect, i.e., chemisorption of the molecule on the surface gives rise to a new electronic state and a resonance Raman effect. The mechanisms can be distinguished somewhat by looking for evidence of chemisorption (band and intensity shifts on adsorption) and the excitation wavelength and potential dependences of the signal. Both mechanisms only operate over a short range from the surface, and in both cases SERS spectra give additional information on molecular orientation by application of the surface selection rules. Recently, it has been discovered that a SERS-active surface or colloid is far from homogeneous in response and that only particles of a certain size on the surface ('hot spots') deliver very high Raman intensities (enhanced by a factor of  $\sim 10^{14}$ ). This has led to several reports of single molecule spectra at colloids for molecules using surface-enhancement Resonance Raman spectroscopy. It remains to be seen whether this can be achieved routinely on electrode surfaces, perhaps with the help of scanning tip technology. SERS suffers from a time-dependence and lack of reproducibility, which makes quantitative applications difficult. There have been some successful attempts in recent years to prepare reproducible SERS surfaces by using, for example, monodisperse metal nanoparticles or by depositing metal films over latex nanospheres. Two other drawbacks are that SERS cannot be used on single-crystal electrodes, and that the effect is limited to the coinage metals, some of which make poor electrode materials. However, it has been shown that pinhole-free thin film of 3–30 atomic layers of other metals (e.g., Pt, Rh) do not reduce the SERS effect from the underlying silver layer by too much (factor of  $\sim 4$ – $5$ ). Two advantages of Raman over IR are that the laser light is hardly scattered at all by water as solvent and that low-frequency bands are readily observed.

Confocal Raman microspectroscopy is now yielding spectra at micrometer spatial (lateral) resolution, and can be carried out *in situ* (Figure 6). Diffusion profiles can also be mapped. The potential for FTIR microspectroscopy is also apparent, and results have been obtained that are superior to those from OTTE cells, with simpler cell design, thanks to the use of microelectrodes. But the discrimination against solvent absorption is poor for *in situ* work; it is hoped that higher intensity sources, such as synchrotron sources, in conjunction with PM-IR-RAS, can overcome this problem.



**Figure 6** (A) Cell for Raman or FTIR microscopy. (Reprinted with permission from Ren B, Li XQ, She CX, Wu DY, and Tian ZQ (2000) *Electrochimica Acta* 46: 193–205. © 2000, with permission from Elsevier.). (B) Near-field (a) and confocal (b) SERS spectra of Rhodamine 6G adsorbed on single silver nanoparticles. (Reproduced with permission from Emory SR and Nie SM (1997) 69: 2631–2635; © American Chemical Society.)

By analogy with SHG described above, if two light beams of fixed visible frequency  $\omega_{\text{vis}}$  and tunable IR frequency  $\omega_{\text{ir}}$  are focused onto a surface then due to nonlinear optical effects reflected light of sum-frequency  $\omega_{\text{sf}} = \omega_{\text{vis}} + \omega_{\text{ir}}$  may be detected, hence the name sum-frequency generation. The selection rules demand that only bands that are *both* IR and Raman active are observed. Unfortunately, sum frequency generation (SFG) suffers from the disadvantage of strong IR absorption by solvent, so irradiation from the back of the electrode is sometimes used.

## Electron Paramagnetic Resonance and Nuclear Magnetic Resonance

Electron paramagnetic resonance (EPR) has played a very important role in the study of electrogenerated intermediates, particularly of radical cations and anions of organic compounds. The *g*-values and the

hyperfine coupling constants help to identify the radical and the redox orbital. The obvious difficulty of *in situ* EPR, in which the radicals are generated at the electrode placed within the microwave cavity, is the presence of metal electrodes and the high dielectric constant solvents typically used in electrochemistry. These absorb microwaves, reducing the microwave intensity in the cavity. The electrodes also disrupt the resonant frequency of the cavity, leading to difficulties in tuning the cavity and requiring precise cell positioning. To avoid these difficulties, the cell is sometimes placed outside the cavity (external generation), particularly if very low temperatures are required to obtain meaningful EPR signals of some metal species. The external generation can be a conventional cell or an inline flow cell, for which detailed and reproducible mass transport characteristics are known. Nevertheless, many *in situ* EPR electrochemical cell designs have been published with the rectangular cell and mesh working electrode

flow cell proving most popular, due to its higher working electrode area and thus higher sensitivity. The flow cell allows ready replacement of the solution without having to retune the cavity, and also provides improved mass transport. Kinetic studies require modeling of the lifetime of the signal as a function of flow rate. Bubble formation is also suppressed if the solution enters from the bottom. The reference electrode is placed as close to the working electrode as possible, but outside the cavity, and the auxiliary electrode is placed downstream from the cavity. Even so, the mass transport characteristics of the cell are far from ideal and distorted CVs are inevitable. However, the signal can be correlated with the charge passed (recommended, since EPR is very sensitive to impurity signals, a notable example being  $\text{ClO}_2^\bullet$  from perchlorate oxidation). The line-widths of stable radicals can give information on their concentration and environment (freely tumbling dilute radicals give narrower lines). The line-width can be used to measure the self-exchange rate. Very unstable radicals can be trapped using a spin trap to give a more stable radical adduct whose spectrum can be observed at leisure.

The much lower sensitivity of nuclear magnetic resonance (NMR) and the need for higher fields and a spinning sample have thwarted many attempts at *in situ* electrolysis. However, metallized glass electrodes can be designed with low 'skin depth', i.e., the thickness of metal required to reduce the RF fields to  $1/e$  of its value. A recent practical cell uses a normal spinning 10 mm NMR tube into which dips a stationary electrode assembly suspended above the probe. If spinning can be avoided, then the stationary probes of solid-state NMR machines can be used. For example, this method has allowed the study of surface species on catalytic metal powder electrodes, e.g., the adsorption of CO from methanol on high-surface-area Pt particles. The use of NMR pulsed field gradient methods has been widely used for the study of ionic diffusion in electrolytes, such as polymer electrolytes for Li batteries, in the absence of electrode materials.

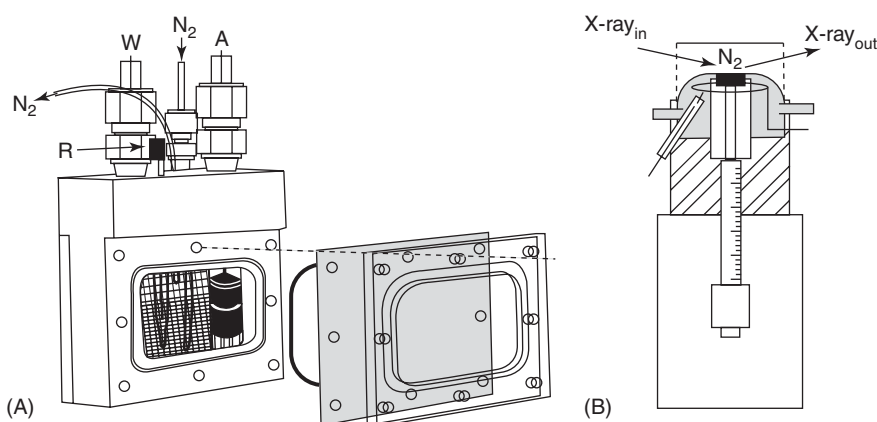
## X-Ray Absorption Spectroscopy (XAS)

X-rays are strongly scattered by matter with a cubic dependence on wavelength. Superimposed on the scattering spectrum, obtained in transmission mode, are small step-like features (absorption edges) due to absorption by X-rays at wavelengths sufficient to eject a core electron from the atom. The precise wavelength is related to the ionization energy and is thus element specific. As in light absorption, the

absorbance is defined from  $I = I_0 \exp(-\mu x)$  where  $\mu$  is the absorption coefficient at the particular wavelength and  $x$  is the distance into the material. In the spectrum  $\mu$  is plotted against the X-ray energy and the scattering curve is subtracted. Buried in the leading edge of the absorption ('near-edge') are absorptions due to excitation of electrons from the core to higher atomic orbitals. These low-intensity bands are called fine structure; hence, the terms X-ray absorption near edge structure or, alternatively, near-edge X-ray absorption fine structure. This can be analyzed to give information on the oxidation state and ligand environment. At higher energy than the absorption edge, more fine structure (extended X-ray absorption fine structure) is observed, associated with the interference patterns generated between ejected electron waves and recently other waves scattered back from surrounding atoms. This can be processed to give information on the local environment of the absorbing atom, and is very useful for amorphous materials and even solution species. Due to strong scattering from solvent molecules, and the low concentration of surface species or thin films, the best results have been obtained using thin-layer transmission cells ( $\sim 10 \mu\text{m}$  thick) and high-intensity sources, such as a monochromated synchrotron source, covering a wide range of wavelengths are required. Cells use polymer windows (such as Mylar or polythene) and thin-film electrodes on polymer substrates, metal foils, or gauzes (Figure 7A). The electrodes and electrolyte can even be placed in a plastic bag or pouch. As an alternative to transmission mode, the cell can be held at  $45^\circ$  to the beam and the detector picks up the X-ray fluorescence emitted at all angles from the recombination of atoms and electrons (and hence closely related to the absorbance). X-ray fluorescence detection is more sensitive. Using similar electrochemical cells, X-ray standing wave spectroscopy can be used to give information on electrode double-layer structure or polymer coatings.

## X-Ray Diffraction and Surface X-Ray Scattering Methods

Structural information on electrodes can also be obtained as a function of potential by X-ray using monochromatic synchrotron radiation and similar transmission cells to those used for XAS (Figure 7B). Transmission cells (Laue mode) are best suited to adsorbates, while reflection (Bragg) cells are used for thicker films. Laboratory rotating anode sources can also be used, and here the instrument usually demands a reflection cell geometry. Stable systems such as solid-state cells and batteries give better



**Figure 7** (A) Cell for transmission XAS (Reprinted with permission from Soderholm L, Antonio MR, Williams C, and Wasserman CR (1999) *Analytical Chemistry* 71: 4622–4628. © 1999, American Chemical Society.) (B) Cell for X-ray scattering (GIXD). (Reprinted with permission from Kondo T, Tamura K, Takahasi M, Mizuki J, and Uosaki K (2002) *Electrochimica Acta* 47: 3075–3080; © Elsevier.) Similar cells have been used for surface XAS.

stability over the long periods required for data collection, since these sources are  $10^3$ – $10^5$  times less intense than synchrotron sources. For surface thin films, i.e., surface X-ray scattering (SXS), greater selectivity over the bulk is achieved using grazing incidence X-ray scattering, also known as grazing-incidence diffraction. SXS, though technically very demanding, has yielded spectacular results on single crystal electrode surface reconstruction and underpotential deposition. Information from both the reflected beam (emerging at the angle of incidence in the plane of reflection) and the scattered beams (other directions) are used. The former gives information on the roughness of the surface while the latter gives information on the surface structure. This can be combined with XAS data using the same cell configuration.

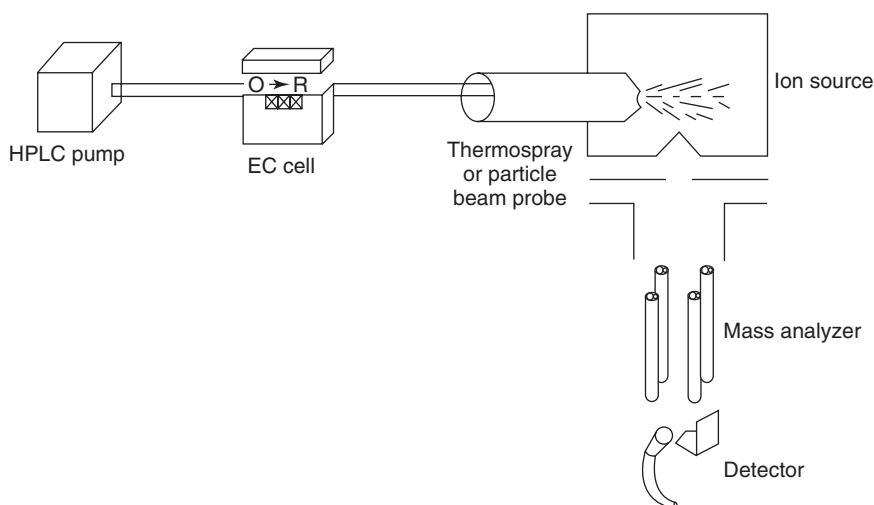
## Mass Spectrometry

Traditionally, products and adsorbates had to be volatile enough so that they could be carried from the cell into the mass spectrometer, either by headspace sampling, or, more commonly for near-simultaneous analysis (referred to as differential electrochemical mass spectrometry), across a nanoporous, gas-permeable membrane (e.g., Teflon) supported at the tip of a microcapillary placed close to the electrode. Alternatively, a Pt-coated membrane electrode can be used. But the advent of the so-called ‘soft’ atmospheric pressure desorption/ionization techniques associated with liquid chromatography–mass spectrometry has allowed the sampling of the solvent and involatile solutes directly. The spectra are more

interpretable as there is less fragmentation, and hence fewer lines not due to molecular ions. These techniques include atmospheric pressure chemical ionization, which uses a corona discharge, thermospray mass spectrometry (probe at  $290^\circ\text{C}$ ), particle beam (e.g., fast atom bombardment, FAB), and electrospray mass spectrometry (probe at high voltage). The sampling may be done using a membrane-coated capillary inlet in the electrochemical cell or by placing a flow cell immediately before the spectrometer (Figure 8). The time between electrogeneration and detection, the ‘dead time’ is 0.5–1 s, depending on the electrochemical cell’s conversion efficiency, flow rate, and ionization method. The best performance is usually seen for the thermospray systems, permitting simultaneous mass CVs to be recorded at scan rates of  $1$ – $10\text{ mV s}^{-1}$ . Electrospray ionization is of particular interest for the study of charged, electrogenerated inorganic species, but it is less useful for organic species that are not easily oxidized or protonated/deprotonated. It has long been suspected that the high voltage used in the electrospray process inadvertently causes electrochemical oxidation of neutral solutes at the tip. In order to enhance this oxidation for less readily oxidized organic analytes, a working electrode has been placed at the electrospray tip to apply an impressed potential.

## Other Methods

Cells have been designed to monitor insertion electrodes by Mössbauer spectroscopy. The technique is limited to relatively few nuclei. But these include



**Figure 8** Setup for electrochemical thermospray MS. (Reprinted with permission from Regino M, Weston C, and Brajter-Toth A (1998) *Analytica Chimica Acta* 369: 253–262; © Elsevier.)

$^{57}\text{Fe}$ , and hence Mössbauer spectroscopy finds uses in corrosion studies, although enrichment of the iron is sometimes required. The electrode must be very stable as a single spectrum can take several hours or days (depending on film thickness), so its use has been restricted largely to solid-state or battery systems. The sample must be thin enough to prevent the absorption or scattering of the  $\gamma$ -rays. In contrast to XRD methods Mössbauer spectroscopy offers local structural information on the iron or tin centers in mixed oxides. Finally, neutron reflectivity spectroelectrochemistry has proven useful for the investigation of electrode coatings. It could be regarded as the neutron equivalent of ellipsometry, with the advantage of much higher spatial resolution.

**See also:** **Infrared Spectroscopy:** Overview. **Mass Spectrometry:** Overview. **Nuclear Magnetic Resonance Spectroscopy:** Overview. **Raman Spectroscopy:** Surface-Enhanced. **X-Ray Absorption and Diffraction:** Overview.

## Further Reading

- Bard AJ and Faulkner LR (2001) *Electrochemical Methods*. New York: Wiley.
- Campion A and Kambhampati P (1998) *Chemical Society Reviews* 27: 241–250.
- Christensen P (2003) *In situ* Infrared Spectroelectrochemistry. In: Unwin PR (ed.) *Encyclopedia of Electrochemistry*. Weinheim: Wiley.
- Crayston JA (2003) *In situ* UV-visible Spectroscopy. In: Unwin PR (ed.) *Encyclopedia of Electrochemistry*. Weinheim: Wiley.
- Diehl G and Karst U (2002) On-line electrochemistry. *Analytical and Bioanalytical Chemistry* 373: 390–398.
- Frutos AG and Corn RM (1998) SPR of ultrathin organic films. *Analytical Chemistry* 70: 449A–455A.
- Harata A, Shen Q, and Sawada T (1999) Photo-thermal applications of lasers: Study of fast and ultrafast photothermal phenomena at metal–liquid interfaces. *Annual Review of Physical Chemistry* 50: 193–219.
- Lipkowski J and Ross PN (1992) *Adsorption of Molecules at Metal Electrodes*. Weinheim: VCH.
- Prenzler PD, Bramley R, Downing SR, and Heath GA (2000) High field NMR spectroelectrochemistry of spinning solutions: Simultaneous *in situ* detection of electrogenerated species in a standard probe under potentiostatic control. *Electrochemical Communications* 2: 516–521.
- Sharpe LR, Heineman WR, and Elder RC (1990) Exafs Spectroelectro-chemistry. *Chemical Review* 90: 705–722.
- Tadjeddine A and Peremans A (1998) In: Clark RJH and Hester RE (eds.) *Non-linear Optical Spectroscopy of the Electrochemical Interface*. *Spectroscopy for Surface Science*. London: Wiley.
- Tian Z Ren B (2003) Raman Spectroscopy of Electrode Surfaces. In: Unwin PR (ed.) *Encyclopedia of Electrochemistry*. Weinheim: Wiley.
- Toma HE and Araki K (2002) Spectroelectrochemical characterization of organic and metal-organic compounds. *Current Organic Chemistry* 6: 21–34.
- Webster RD, Bond AM, Coles BA, and Compton RG (1996) ESR-electrochemical cells: A Comparative Study. *Journal of Electroanalytical Chemistry* 404: 303–308.

# SPECTROPHOTOMETRY

Contents

**Overview****Diode Array****Derivative Techniques****Turbidimetry and Nephelometry****Inorganic Compounds****Organic Compounds****Biochemical Applications****Pharmaceutical Applications**

## Overview

P J Worsfold, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Ultraviolet-visible (UV-visible) spectrophotometry is primarily a quantitative analytical technique concerned with the absorption of near-UV (180–390 nm) or visible (390–780 nm) radiation by chemical species in solution. These regions of the electromagnetic spectrum (see **Table 1**) provide energy that gives rise to electronic transitions. The various colors of visible light and the complementary colors of solutions absorbing at particular wavelengths are shown in **Table 2**. Because of the superimposition of vibrational and rotational transitions the UV-visible spectrum of analytes in solution shows little fine structure. For this reason the technique is not commonly used for identification (although all molecules will have a wavelength of maximum absorption) but is one of the most widely used for quantitative analysis. Under

controlled experimental conditions the amount of radiation absorbed can be directly related to the concentration of the analyte in solution (Beer's law; see below). It can be used to quantify both organic (primarily in the near-UV) and inorganic (primarily in the visible) species.

## Spectra-Structure Relationships

In general, organic compounds absorb energy in the near-UV region due to  $\sigma \rightarrow \sigma^*$ ,  $n \rightarrow \sigma^*$ ,  $n \rightarrow \pi^*$ , and  $\pi \rightarrow \pi^*$  transitions. Most transitions of interest involve the

**Table 2** The visible spectrum

<i>Wavelength (nm)</i>	<i>Color absorbed</i>	<i>Color observed</i>
390–420	Violet	Green–yellow
420–440	Violet–blue	Yellow
440–470	Blue	Orange
470–500	Blue–green	Red
500–520	Green	Purple
520–550	Yellow–green	Violet
550–580	Yellow	Violet–blue
580–620	Orange	Blue
620–680	Red	Blue–green
680–780	Purple	Green

**Table 1** The electromagnetic spectrum

<i>Region</i>	<i>Wavelength range (m)</i>	<i>Wavelength range (commonly used units)</i>	<i>Type of transition</i>
$\gamma$ -Ray	$< 10^{-12}$	$< 1$ pm	Nuclear
X-ray	$10^{-12}$ – $10^{-8}$	0.001–10 nm	Inner shell electron
Far-ultraviolet	$10^{-8}$ – $1.8 \times 10^{-7}$	10–180 nm	Middle shell electron
Near-ultraviolet	$1.8 \times 10^{-7}$ – $3.9 \times 10^{-7}$	180–390 nm	Outer shell electron
Visible	$3.9 \times 10^{-7}$ – $7.8 \times 10^{-7}$	390–780 nm	Outer shell electron
Near-infrared	$7.8 \times 10^{-7}$ – $2.5 \times 10^{-6}$	0.78–2.5 $\mu$ m	Molecular vibration
Mid- and far-infrared	$2.5 \times 10^{-6}$ – $10^{-3}$	2.5–1000 $\mu$ m	Molecular vibration and rotation
Microwave	$10^{-3}$ –0.3	0.1–30 cm	Molecular rotation
Radio wave	$> 0.3$	$> 30$ cm	Electron and nuclear spin



promotion of  $n$  or  $\pi$  electrons to the  $\pi^*$  excited state and therefore involve molecules containing delocalized  $\pi$  electrons, i.e., aromatic and conjugated aliphatic species. The functional groups in these molecules that absorb energy are known as chromophores. The wavelength of maximum absorption of a molecule can be altered by the nature of the substituent(s) on the aromatic ring or the conjugated system. Substituents containing lone pairs of electrons, e.g.,  $-\text{OH}$ ,  $-\text{NH}_2$ , generally result in a longer wavelength of maximum absorption and a higher molar absorptivity of the molecule. Such groups are known as auxochromes.

Some inorganic compounds can be detected in the visible region by their inherent absorption of radiation. For example, the transition metal ions and their complexes are often colored and have a reasonable molar absorptivity due to transitions involving the 3d and 4d orbitals; lanthanoid and actinoid ions have characteristically sharp absorbance peaks due to effective screening of their 4f and 5f orbitals, respectively. Inherently nonabsorbing inorganic species can often be determined by a selective derivatization reaction wherein the non-absorbing analyte is reacted with an appropriate reagent to form an absorbing complex. Derivatizing reagents can be either universal (such as dithizone (diphenylcarbazone), PAR (4-(2-pyridylazo)resorcinol), and PAN (1-(2-pyridylazo)-2-naphthol) or selective for a particular species under appropriate reaction conditions (such as 1,10-phenanthroline for iron(II) and 1,5-diphenylcarbazide for chromium(VI)).

## Quantitative Aspects

The relationship between absorbance and concentration is known as Beer's law (also referred to by other names such as the Beer–Lambert law and the Bouguer–Lambert–Beer law) and is defined by the equation:

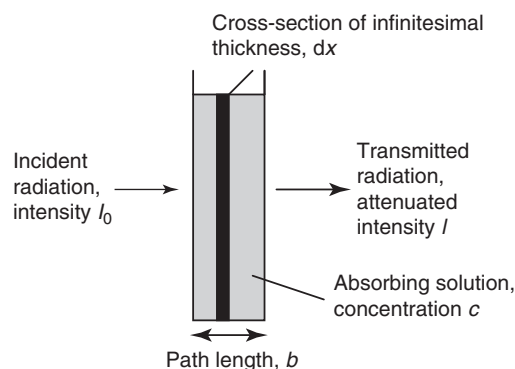
$$A = \epsilon bc$$

where  $A$  is the absorbance of the solution (no units),  $\epsilon$  is the molar absorptivity (units of  $\text{l mol}^{-1} \text{cm}^{-1}$ ),  $b$  is the path length of radiation through the absorbing medium (units of cm), and  $c$  is the concentration (units of  $\text{mol l}^{-1}$ ).

Beer's law can be derived as follows using the nomenclature shown in Figure 1. The transmittance ( $T$ ) is defined as the ratio ( $I/I_0$ ) and therefore

$$\%T = 100(I/I_0)$$

Absorbance ( $A$ ) is defined as  $\log_{10} I_0/I$ , which is the same as  $-\log T$ . Consider a thin layer of solution  $dx$ .



**Figure 1** Attenuation of radiation by an absorbing sample solution.

The decrease in power ( $dI$ ) across  $dx$  will be  $-\alpha I c dx$ , which can be arranged to give  $dI/I = \alpha c dx$ , where  $\alpha$  is a proportionality constant. Integrating over the path length of the cuvette gives

$$\int_{I_0}^I -dI/I = \alpha c \int_0^b dx$$

which gives

$$-\ln(I) - (-\ln(I_0)) = \alpha bc$$

which rearranges to give

$$\ln(I_0/I) = \alpha cb$$

Converting to  $\log_{10}$  gives

$$\log(I_0/I) = abc$$

which is the same as

$$A = abc$$

wherein  $a$  is another proportionality constant. Substituting concentration units of  $\text{mol l}^{-1}$  gives

$$A = \epsilon bc$$

Beer's law can also be applied to multicomponent systems by utilizing the additive effect of absorbances in solution. For example, two components (1 and 2) in a mixture can be quantified by measuring the absorbance at two wavelengths ( $x$  and  $y$ ) and knowing the molar absorptivities of 1 and 2 at the two wavelengths ( $\epsilon_x$  and  $\epsilon_y$ ) as follows:

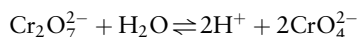
$$A_x = \epsilon_{1x}bc_1 + \epsilon_{2x}bc_2$$

$$A_y = \epsilon_{1y}bc_1 + \epsilon_{2y}bc_2$$

The unknown concentration  $c_1$  and  $c_2$  can be determined by solving the simultaneous equations above.

More complex mixtures can be analyzed in a similar way providing that there are no deviations from Beer's law. Derivative spectrophotometry can also be used to mathematically process the data after acquisition in order to improve spectral resolution in multi-component systems. In this approach the zero-order spectrum is derivatized to give first order ( $dA/d\lambda$ ) or higher plots of the rate of change of absorbance against absorbance.

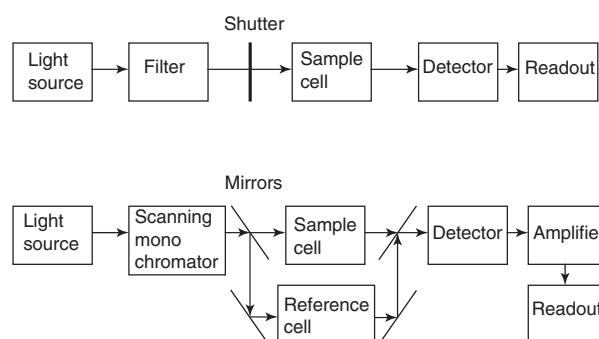
Beer's law is not valid at high concentrations ( $>0.01 \text{ mol l}^{-1}$ ) due to interactions between analyte molecules. There are also apparent limitations to the applicability of Beer's law due to instrumental and chemical deviations. Major instrumental deviations result from the fact that radiation impinging on the sample is polychromatic rather than monochromatic and the effect of stray radiation. Other, less important, instrumental factors are reflection losses, path length variability, light scattering and fluorescence (in the sample cell), and nonideal behavior in the signal processing and readout devices. Chemical deviations relate primarily to changes in absorbance due to changes in the chemical environment of the sample, e.g., pH. A classical example of this is the pH-dependent absorbance of Cr(VI) in solution due to the equilibrium between chromate (yellow) and dichromate (orange):



Other important equilibria that affect absorbance involve acid–base forms of a species (such as pH indicators) and metal–ion complexes. With all equilibria there will be one wavelength, known as the isosbestic point, where the absorbances of both species are equal because they have the same  $\epsilon$  value. The nature of the solvent (and the sample cell) and refractive index changes can also cause a positive deviation and this should be checked as part of the measurement protocol for the blank.

## Instrumentation

The basic instrumentation consists of a radiation source, a wavelength-selection device, a sample compartment, a detector, and an output device. For visible radiation a tungsten filament is the most common source and for near-UV radiation a deuterium lamp is usually used. Light emitting diodes can also be used as low cost sources with narrow bandwidths (typically 20–30 nm). For the best resolution a grating monochromator is used as the wavelength-selection device (this can also be used for scanning the range of interest); for lower-cost, lower-resolution systems a filter is used. The sample compartment is usually a cell with



**Figure 2** Block diagrams of single-beam (above) and double-beam (below) spectrophotometers.

a 1 cm square cross-section made of quartz (for the near-UV), glass, or plastic (for the visible). Consideration should also be given to the absorption characteristics of the solvent when developing a quantitative method. For detection a photomultiplier is the standard device but phototubes and, increasingly, photodiodes and diode arrays are also used.

The two most common optical configurations for UV–visible spectrophotometers are single-beam and double-beam arrangements (see **Figure 2**). Single-beam instruments are usually cheaper whereas double-beam instrument are more convenient for spectral scanning. An alternative optical arrangement is used in diode array spectrophotometry, wherein polychromatic light from the source passes through the sample cell and is then dispersed by a grating onto a linear diode array detector.

## Applications

UV–visible spectrophotometry is principally used for the quantitative determination of organic and inorganic constituents in a wide range of sample matrices, e.g., environmental, biochemical, pharmaceutical, clinical, and food. The technique can be used for the direct determination of absorbing species, such as the determination of nitrate in water at 208 nm, but care must be taken to avoid matrix interferences. Reaction rate measurements, used in methods based on catalytic reactions, can overcome such interferences and provide enhanced sensitivity. The technique can also be used for indirect determinations following derivatization of the analyte with a selective reagent to form an absorbing species, e.g., the determination of Fe(II) in water at 510 nm after reaction with the bidentate ligand 1,10-phenanthroline.

Systems based on diffuse reflectance from solid surfaces rather than absorbance in solution are also commercially available, particularly for clinical analysis, and fiber optic sensors based on spectrophotometric principles have been developed for

clinical and environmental applications. Spectrophotometry is also used extensively in combination with separation techniques such as liquid chromatography and capillary electrophoresis as a generic detector, as it is in nonchromatographic flow systems such as flow injection analysis.

See also: **Spectrophotometry:** Diode Array; Derivative Techniques; Turbidimetry and Nephelometry; Inorganic Compounds; Organic Compounds; Biochemical Applications; Pharmaceutical Applications.

## Further Reading

- Burgess C and Mielenz KD (1987) *Advances in Standards and Methodology in Spectrophotometry*. Amsterdam: Elsevier.
- Ingle JD and Crouch SR (1988) *Spectrochemical Analysis*. Englewood Cliffs: Prentice-Hall.
- Marczenko Z (1986) *Separation and Spectrophotometric Determination of Elements*. Chichester: Ellis Horwood.
- Perkampus HH (1992) *UV-VIS Spectroscopy and its Applications*. Berlin: Springer.

## Diode Array

**A V Medvedovici and V David**, University of Bucharest, Bucharest, Romania

© 2005, Elsevier Ltd. All Rights Reserved.

### Definition and General Considerations

Diode array spectrometry represents a particular approach for characterizing the result of the interaction between electromagnetic radiation and the sample, based on simultaneous measurement of light intensity over small spectral intervals having equal width. This is achieved through the following sequential stages, as schematically shown in **Figures 1** and **2D**:

1. interaction between the polychromatic beam and the sample;
2. spatial dispersion of the transmitted (eventually reflected or emitted) radiation according to wavelength by means of a fixed optical element;
3. imaging the dispersed radiation in a flat focal plane; and
4. simultaneous sampling of the dispersed radiation interval using photosensitive detectors, precisely positioned in the flat focal plane.

Each of the detectors measures the radiation intensity on a spectral width resulting from a division of the linearly dispersed wavelength interval by the number of photosensitive receivers disposed side to side. Alternatively, diode array spectrometry can be employed, which supposes a transformation of the continuum into discrete values, generally followed by interpolation and smoothing procedures.

This approach is essentially different from the other alternatives for characterizing polychromatic radiation. **Figure 2A** illustrates the basic principle of a conventional scanning spectrometer. The

polychromatic radiation produced by the source is incident on an optical device called monochromator through an entrance slit. Dispersion is achieved with a moving optical element (prism, plane, or concave diffraction grating) focusing at a time through the exit slit a specific wavelength related to the incident angle ( $\theta$ ). This wavelength passes through the sample, and the transmitted intensity is measured by a unique detector. Spectrum reconstitution is achieved sequentially by scanning the wavelength interval in question.

The Hadamard multiplexing technique (**Figure 2B**) is also based on the dispersion of the transmitted radiation by the sample with a fixed optical element (identical to the diode array approach), while detection is carried out by means of a unique large bandpass detector. Mask displacement allows simultaneous intensity measurement on multiple spectral channels at different times. Spectrum reconstitution is achieved by solving the system of mathematical equation generated by the mask movement in time.

The Fourier multiplexing technique (**Figure 2C**) is based on the interferometric process applied to the incident beam. For a given position of the moving mirror inducing different pathway lengths, constructive interference is obtained only for a series of specific wavelengths. The detector generates a periodic function  $I = f(t)$ , analyzed by means of a Fourier series. Then, the resulting sinusoidal or cosinusoidal pure functions are recombined to generate an intrinsic  $I = f(\lambda)$  dependence.

It is thus obvious that diode array spectrometry can be differentiated from the other spectrometric techniques by the following considerations:

1. the use of a polychromator placed after the passage of the incident beam through the sample;

clinical and environmental applications. Spectrophotometry is also used extensively in combination with separation techniques such as liquid chromatography and capillary electrophoresis as a generic detector, as it is in nonchromatographic flow systems such as flow injection analysis.

See also: **Spectrophotometry:** Diode Array; Derivative Techniques; Turbidimetry and Nephelometry; Inorganic Compounds; Organic Compounds; Biochemical Applications; Pharmaceutical Applications.

## Further Reading

- Burgess C and Mielenz KD (1987) *Advances in Standards and Methodology in Spectrophotometry*. Amsterdam: Elsevier.
- Ingle JD and Crouch SR (1988) *Spectrochemical Analysis*. Englewood Cliffs: Prentice-Hall.
- Marczenko Z (1986) *Separation and Spectrophotometric Determination of Elements*. Chichester: Ellis Horwood.
- Perkampus HH (1992) *UV-VIS Spectroscopy and its Applications*. Berlin: Springer.

## Diode Array

**A V Medvedovici and V David**, University of Bucharest, Bucharest, Romania

© 2005, Elsevier Ltd. All Rights Reserved.

### Definition and General Considerations

Diode array spectrometry represents a particular approach for characterizing the result of the interaction between electromagnetic radiation and the sample, based on simultaneous measurement of light intensity over small spectral intervals having equal width. This is achieved through the following sequential stages, as schematically shown in **Figures 1** and **2D**:

1. interaction between the polychromatic beam and the sample;
2. spatial dispersion of the transmitted (eventually reflected or emitted) radiation according to wavelength by means of a fixed optical element;
3. imaging the dispersed radiation in a flat focal plane; and
4. simultaneous sampling of the dispersed radiation interval using photosensitive detectors, precisely positioned in the flat focal plane.

Each of the detectors measures the radiation intensity on a spectral width resulting from a division of the linearly dispersed wavelength interval by the number of photosensitive receivers disposed side to side. Alternatively, diode array spectrometry can be employed, which supposes a transformation of the continuum into discrete values, generally followed by interpolation and smoothing procedures.

This approach is essentially different from the other alternatives for characterizing polychromatic radiation. **Figure 2A** illustrates the basic principle of a conventional scanning spectrometer. The

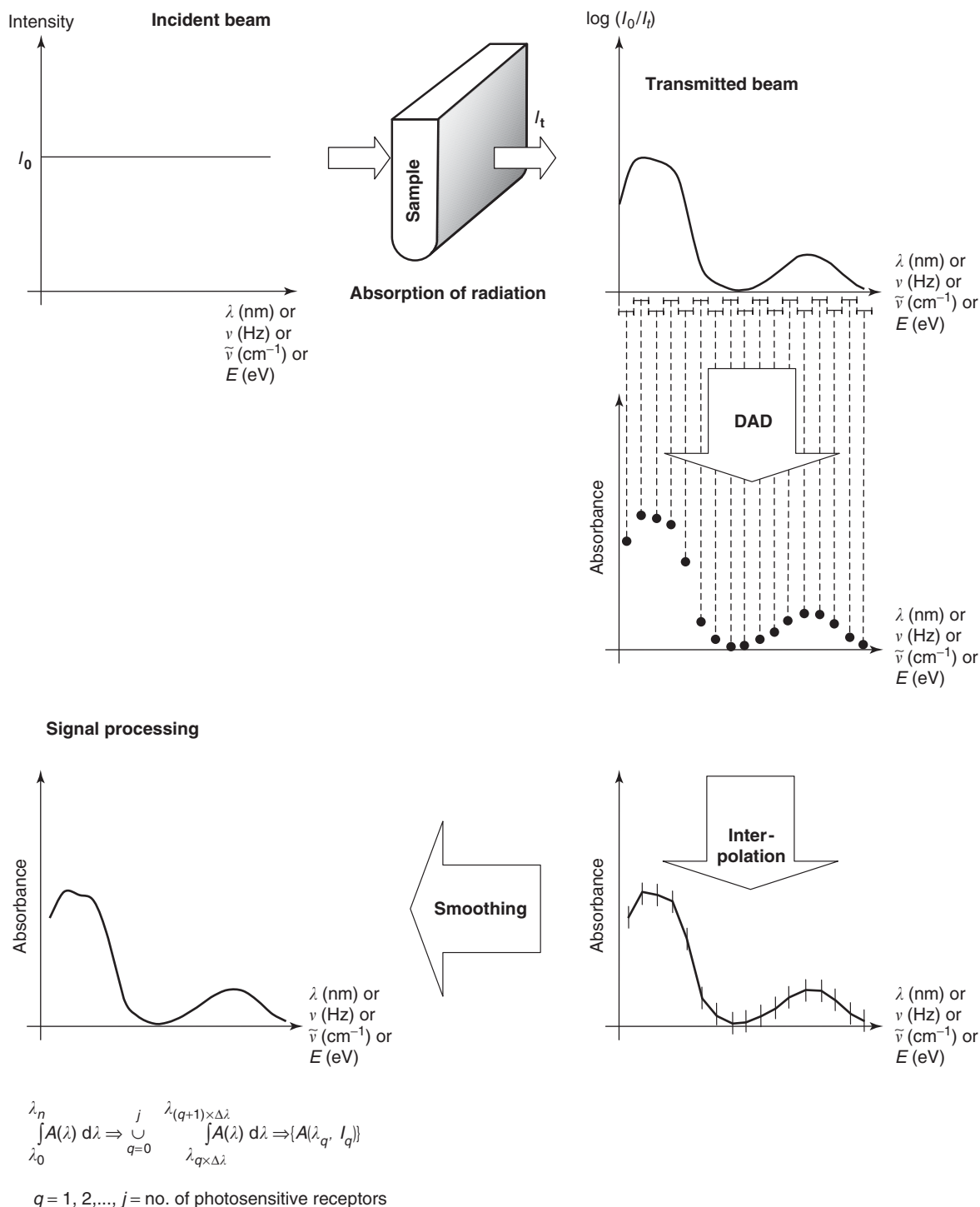
polychromatic radiation produced by the source is incident on an optical device called monochromator through an entrance slit. Dispersion is achieved with a moving optical element (prism, plane, or concave diffraction grating) focusing at a time through the exit slit a specific wavelength related to the incident angle ( $\theta$ ). This wavelength passes through the sample, and the transmitted intensity is measured by a unique detector. Spectrum reconstitution is achieved sequentially by scanning the wavelength interval in question.

The Hadamard multiplexing technique (**Figure 2B**) is also based on the dispersion of the transmitted radiation by the sample with a fixed optical element (identical to the diode array approach), while detection is carried out by means of a unique large bandpass detector. Mask displacement allows simultaneous intensity measurement on multiple spectral channels at different times. Spectrum reconstitution is achieved by solving the system of mathematical equation generated by the mask movement in time.

The Fourier multiplexing technique (**Figure 2C**) is based on the interferometric process applied to the incident beam. For a given position of the moving mirror inducing different pathway lengths, constructive interference is obtained only for a series of specific wavelengths. The detector generates a periodic function  $I = f(t)$ , analyzed by means of a Fourier series. Then, the resulting sinusoidal or cosinusoidal pure functions are recombined to generate an intrinsic  $I = f(\lambda)$  dependence.

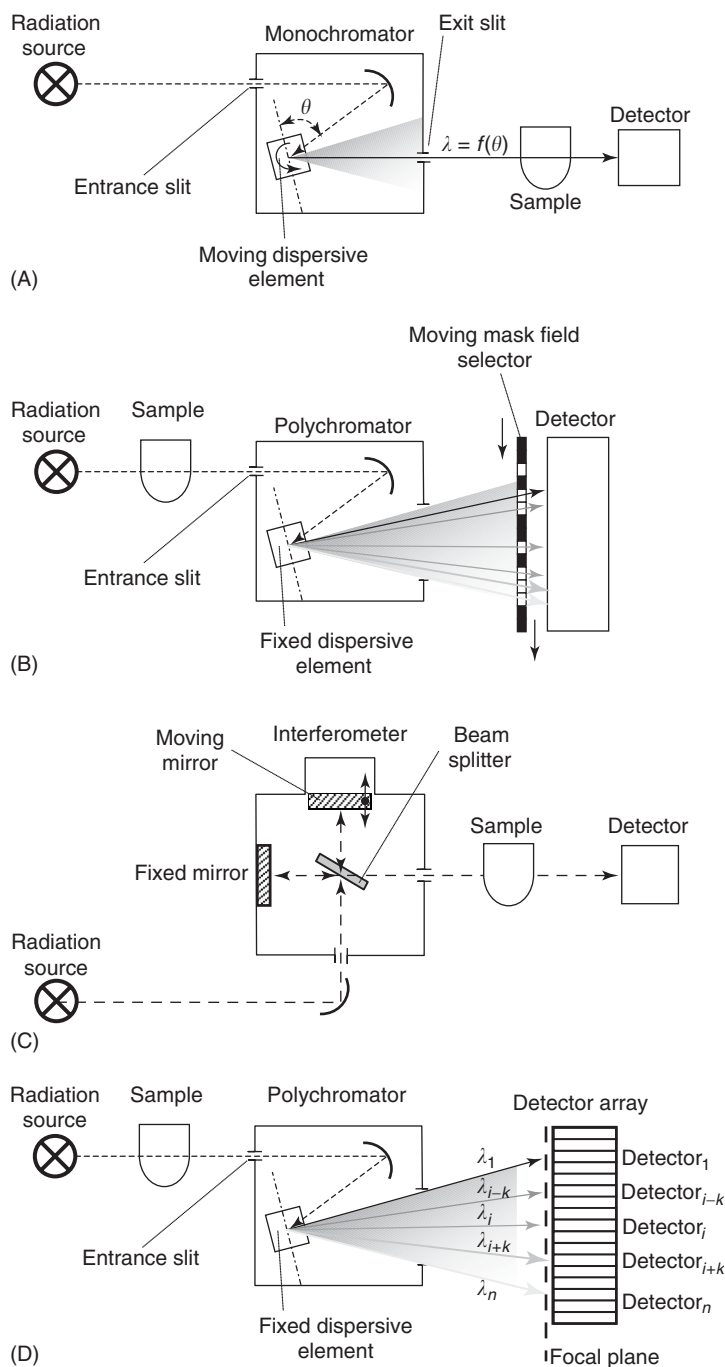
It is thus obvious that diode array spectrometry can be differentiated from the other spectrometric techniques by the following considerations:

1. the use of a polychromator placed after the passage of the incident beam through the sample;



**Figure 1** Basic principle of diode array spectrometry.

- the use of a polychromator that disperses light according to wavelength by means of a fixed optical element (holographic concave diffraction grating) and subsequent imaging of the dispersed light in a flat focal plane;
- the use of photosensitive detectors serially placed in the flat focal plane, generating simultaneous intensity readouts over a given spectral width; and
- spectrum reconstitution based on the correlation of the reading generated by each detector



**Figure 2** Comparative overview on different instrumentation principles used in molecular absorption spectrometry.

according to the position assigned to a dispersed wavelength.

## Concave Diffraction Gratings

The concave diffraction grating is an optical element combining two functions: (1) reflecting and focusing

light in a flat path (given its concave geometry) and (2) dispersing light according to wavelength (given its groove pattern). It consists of finely spaced lines produced on the concave face of a spherical surface.

Concave diffraction gratings were first introduced in spectrometric practice by Henry Rowland in 1883. Based on Fermat's principle, Rowland was able to



specify the shape and location of the grating lines, without having all technological facilities for producing them. According to Rowland, an aberration-free image can be produced for a given wavelength if the geometry of the grooves on the grating concave surface follows the intersection of a family of hyperbolae. The modern theory of the concave holographic diffraction gratings (the term holographic refers to the production technology of the grating) was proposed by Namioko *et al.*

As shown in Figure 3, the incident polychromatic light is focused on the concave diffracting grating by means of an entrance slit positioned on the Rowland circle, giving also the radius of curvature of the grating ( $R$ ). The length of the incident light beam is  $L_I$  and  $\alpha$  is the corresponding angle measured against the normal to the diffraction grating surface at the center.  $\beta_i$  are diffraction angles for wavelengths included in the specified spectral interval (from  $\lambda_0$  to  $\lambda_n$ ). The concave grating focuses the dispersed light in a flat focal (spectral) plane. All distances from the grating surface to the focal plane are denoted as  $L_i$  ( $i=0, \dots, n$ ) the index corresponding to the wavelength.  $L_H$  is the length of the diffracted wavelength ( $\lambda_H$ ) falling perpendicularly on the focal plane. The focal plane makes an angle  $\gamma$  with the normal to the last diffracted wavelength. Other

notations are as follows:  $N$  is the groove frequency at the center of the grating (number of grooves per millimeter);  $M$  the diffraction order;  $d$  the groove spacing at the center of the grating (nanometers); and  $\lambda_F$  the wavelength used during holographic grating fabrication.

The following equations describe the imaging in the plane of dispersion produced by a concave diffraction grating:

- the grating equation

$$\sin \alpha + \sin \beta_i = m \times N \times \lambda_i; \quad N = \frac{10^6}{d}$$

- the tangential (spectral plane) focus equation

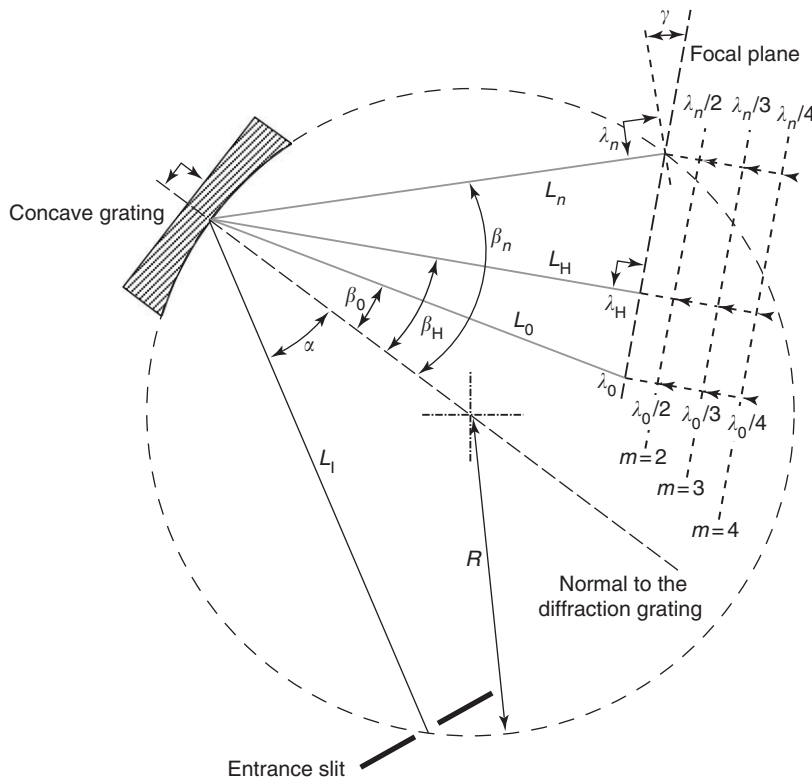
$$\frac{\cos^2 \alpha}{L_I} - \frac{\cos \alpha}{R} + \frac{\cos^2 \beta \times \lambda_i}{L_i} - \frac{\cos \beta}{R} - \frac{m \times \lambda_i}{\lambda_F} \times C_f = 0,$$

$$i = 0, \dots, n$$

- the sagittal (astigmatic) focus equation

$$\frac{1}{L_I} - \frac{\cos \alpha}{R} + \frac{1}{L_i} - \frac{\cos \beta_i}{R} - \frac{m \times \lambda_i}{\lambda_F} \times C_a = 0,$$

$$i = 0, \dots, n$$



**Figure 3** Basic principles of the concave diffraction grating.

- the coma equation

$$\frac{\sin \alpha}{L_i} \left( \frac{\cos^2 \alpha}{L_i} - \frac{\cos \alpha}{R} \right) + \frac{\sin \beta_i}{L_i} \left( \frac{\cos^2 \beta_i}{L_i} - \frac{\cos \beta_i}{R} \right) - \frac{m \times \lambda_i}{\lambda_F} \times C_c = 0, \quad i = 0, \dots, n$$

where  $C_f$ ,  $C_a$ , and  $C_c$  are constants depending on the experimental setup used during the production of the holographic concave diffraction grating.

The optical properties of a concave grating can be expressed as follows:

1. Angular dispersion – representing the angular separation between two wavelengths. The angular dispersion is calculated using the relationship

$$\frac{d\beta_i}{d\lambda_i} = (m \times N \times 10^{-6}) / \cos \beta_i$$

2. Linear dispersion – defined as the segment across the focal plane on which the spectral interval is spread out (generally expressed in nm/mm). The linear dispersion is given by the formula

$$\frac{d\lambda_i}{dx} = (10^6 \times \cos \beta \cos \gamma) / (m \times N \times L_i)$$

At the wavelength normal to the focal plane, the previous relation should be rewritten as

$$\frac{d\lambda_H}{dx} = (10^6 \times \cos \beta \cos^2 \gamma) / (m \times N \times L_H)$$

3. Wavelengths and diffraction orders – for a grating of given groove density and for given values of  $\alpha$  and  $\beta$ ,

$$m \times \lambda = \text{constant}$$

so that if the diffraction order  $m$  is doubled,  $\lambda$  is halved, and so on.

4. Resolving power – defined as the ability of the concave diffraction grating to separate adjacent spectral lines. Two consecutive wavelengths are resolved if their separation (considered as a distance) is such that the maximum of the first one falls at the starting minimum of the second (the Rayleigh criterion). It is demonstrated that the resolving power, RP, is given by the relation

$$RP = \frac{\lambda}{d\lambda} = m \times N \times W = m \times G$$

where  $W$  is the illuminated width of the grating and  $G$  represents the total number of grooves on the grating surface.

The resolving power of a grating depends on its width, the central wavelength to be resolved, and the geometry of the optical instrument.

## Manufacturing of Concave Diffraction Gratings

The mechanical ruling method for manufacturing concave diffraction gratings has been used since the late 1960s. Although A. Cotton in 1901 produced experimental interference gratings, even before the launch of the holography theory by Gabor, the availability of ion lasers and grainless photosensitive materials (photoresists) was lacking until the mid-1960s.

The interference produced by two coherent, equally polarized monochromatic optical waves of identical intensity, intersecting each other, generates a combined intensity distribution of fringes (bright and dark lines). Fringes are recorded on photosensitive plates with concave geometry. The light intensity varies sinusoidally with position, as the interference pattern is scanned along a direction. By modifying the intensity ratio between interfering sources, the exposure can be controlled. The separation between two adjacent maxima is  $d = \lambda_F / (2 \sin \varphi)$  ( $\lambda_F$  is the wavelength of the monochromatic sources and  $\varphi$  is the half-angle between the interfering beams).

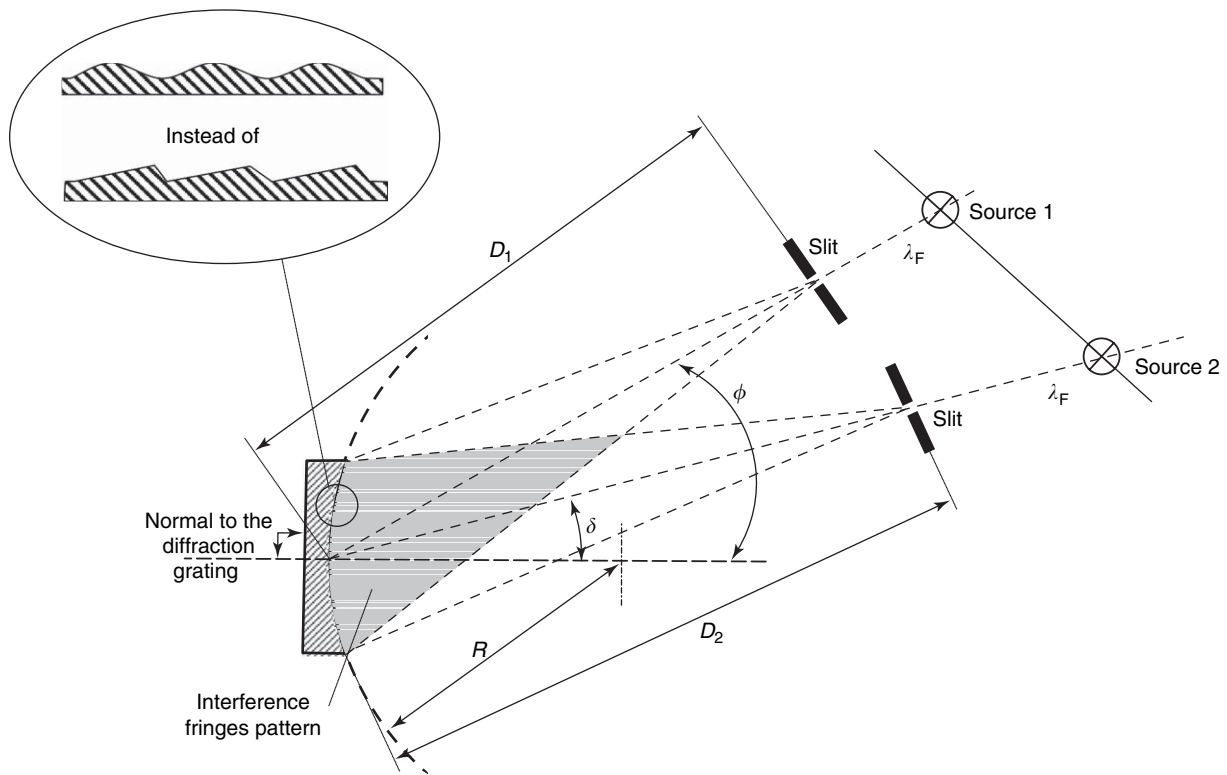
If the wavefronts have spherical section and are converging or diverging, the interference pattern produced will represent confocal hyperboloids (ellipsoids are obtained if one wavefront diverges while the other converges). Addition of auxiliary concave mirrors or lenses into the paths of the interfering beams can render the fringe pattern toroidal. The former situations lead to first and, respectively, second generation concave holographic gratings.

The recording process is schematically given in **Figure 4**. According to the manufacturing process parameters ( $\lambda_F$ ,  $\delta$ ,  $\phi$ ,  $D_1$ , and  $D_2$ ), one can control the constructive constants of the grating ( $C_c$ ,  $C_a$ , and  $C_f$ ).

It is worth noting that the groove profile resulting from holographic manufacture is sinusoidal, while mechanical ruling generates triangular profiles. Consequently, holographic gratings produce different efficiencies for a given wavelength and spectral order with respect to ruled gratings. Holographic gratings for the visible and ultraviolet (UV) spectral ranges exhibit less stray light and ‘ghost’ spectra than do classically ruled ones, because they have less random and systematic imperfections.

On recording the pattern of the fringes on concave substrates, the resulting groove image depends on the position as well as the local incidence angle. Diffracted rays are not parallel, and the gratings possess focal (imaging) properties superposed to dispersive ones.

After exposure, the pattern is etched into the substrate coated with the photoresist film. Blazing of



**Figure 4** Holographic manufacture of concave diffraction gratings.

the grating is achieved by means of a chemical etch, possibly followed by an ion etch to obtain the desired shape. Groove density ranges from 65 to more than 5000 grooves per millimeter.

According to the setup described in **Figure 4**, the following relations can be used for the calculation of the grating constants:

$$N \times \lambda_F = \sin \delta + \sin \phi$$

$$C_f = \frac{\cos^2 \phi}{D_1} - \frac{\cos \phi}{R} - \left( \frac{\cos^2 \delta}{D_2} - \frac{\cos \delta}{R} \right)$$

$$C_a = \frac{1}{D_1} - \frac{\cos \phi}{R} - \left( \frac{1}{D_2} - \frac{\cos \delta}{R} \right)$$

$$C_c = \frac{\sin \phi}{D_1} \left( \frac{\cos^2 \phi}{D_1} - \frac{\cos \phi}{R} \right) - \frac{\sin \delta}{D_2} \left( \frac{\cos^2 \delta}{D_2} - \frac{\cos \delta}{R} \right)$$

## The Photodiode Principle

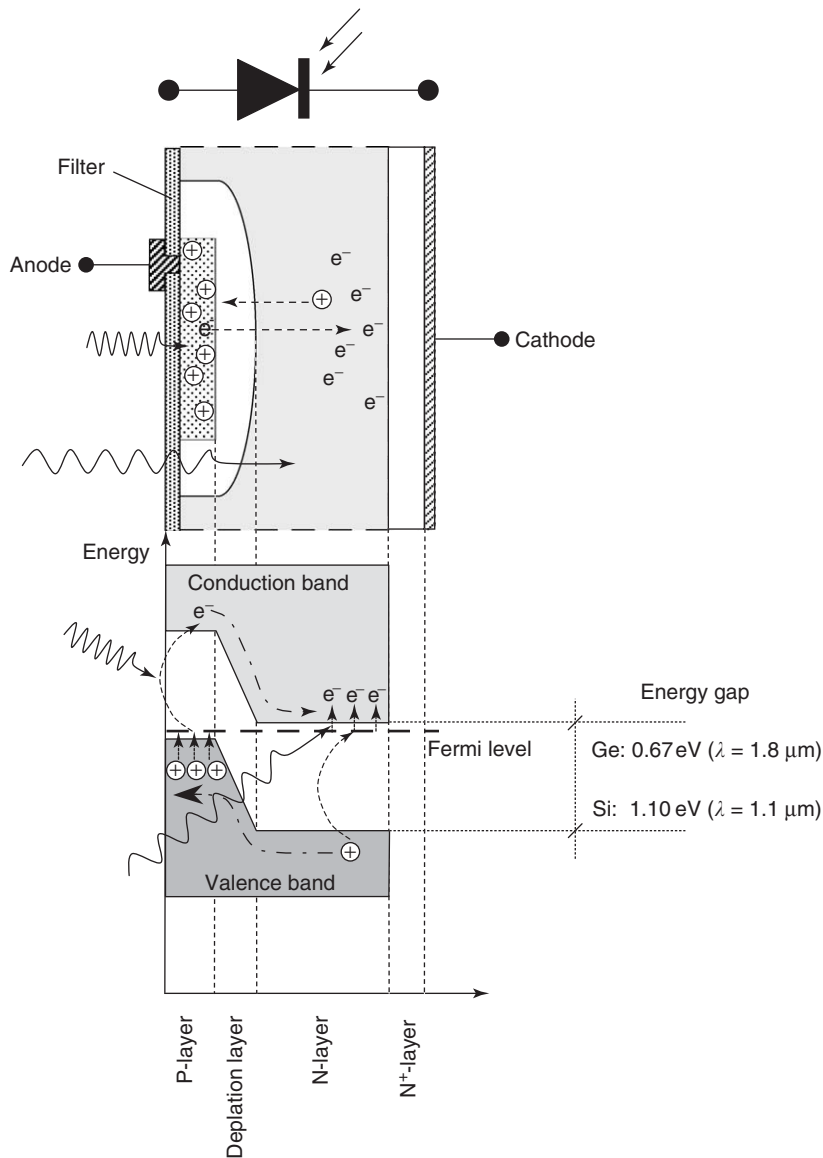
The photodiode is a semiconductor-based light-sensitive element, generating a current proportional to the illumination of the P–N junction (see **Figure 5**).

N-Type materials are intrinsic three-dimensional (3D) crystals of silicon or germanium subjected to

controlled impurification with chemical elements from the Va column of the periodic table (e.g., arsenic or phosphorous), behaving as a semiconductor because of the availability of unbound electrons belonging to the doping element atoms.

P-Type materials are intrinsic 3D crystals of silicon or germanium subjected to controlled impurification with chemical elements from the IIIa column of the periodic table (e.g., boron or aluminum), behaving as a semiconductor due to the availability of holes induced by the doping element atoms.

In semiconductors, the energy gap between the valence band and conduction band is relatively low. In N-type semiconductors, the Fermi level (defined as the highest energy level within a crystal populated by electrons at a temperature of absolute zero) is situated closer to the conduction band than in P-type semiconductors. Unbound electrons, at higher doping levels, are readily pushed into the conduction band with a small amount of energy transfer. In P-type semiconductors, the Fermi level is situated close to the valence band. Absorption of external energy results in an electron transfer at the Fermi level, leaving holes in the valence band. Further, energy transferred from external sources pushes electrons into a hole, generating other holes. The bound



**Figure 5** The photodiode functioning principle.

state of the electron is replicated in a new location, providing the appearance of a hole moving toward the opposite direction.

In a P–N junction generated in a single silicon or germanium crystal, the Fermi level is unique. Due to doping, the conduction and valence bands have different energies throughout the crystal, as can be seen from **Figure 5**. The neutral transition region within the P–N junction is called the depletion layer. When low-energy light strikes the photodiode, unbound electrons in the N-layer are promoted in the conduction band and the bound electrons in the P-layer are pushed onto the Fermi level. When higher-energy light strikes the photodiode, electrons from the Fermi level in the N-layer are promoted to the conduction

band, while valence electrons on the N-layer are forced over the Fermi level. Thus, the crystal becomes polarized due to electron concentration in the conduction band through the N-layer and hole concentration in the valence band through the P-layer. In the depletion layer, the electric field accelerates electrons toward the N-layer and holes toward the P-layer, resulting in a positive charge collected at the anode and a negative one placed on the cathode of the photodiode.

Controlling the thickness of the P, N, and depletion layers, as well as the dopant element concentration profile through the depletion layer, the photosensitivity ( $S$ ) and the quantum efficiency (QE) are controlled. The photosensitivity represents the

ratio between the energy of the incident beam striking the photodiode (expressed in watts) and the resulting photocurrent (expressed in amperes). QE is the photosensitivity of a photodiode measured at a specified wavelength. One can define also the spectral response of a photodiode as the photocurrent level generated for identical incident light intensity for specified wavelengths.

However, a small current is generated through a photodiode in dark conditions, when a reversed voltage is applied. This resulting current is known as a dark current.

If a charged condenser is connected in parallel to a photodiode, electrons will flow away from the N-layer into the condenser anode, while electrons from the condenser will compensate the hole excess in the P-layer. As a consequence, the condenser is discharged. The discharge of the capacitor through the photodiode will be proportional to the light intensity inducing the N-P layer junction polarization (Figure 6).

If the parallel circuit of a photodiode and a capacitor is electronically shifted between a direct current source and a current measuring device, with a given frequency, it is possible to monitor the light intensity falling on the anode. During the first stage, the direct current source charges capacitor. If anode is illuminated, the diode switches to the conduction mode and depletes the charge on the capacitor. The condenser discharge is proportional to the light intensity. When the solid-state electronic switch shifts in the measuring device loop, the residual current on the capacitor is measured.

Another working alternative is to charge, at regular intervals, the capacitor to a specific level. The amount of current needed to recharge the capacitor is thus measured, being proportional to the light intensity falling on the diode.

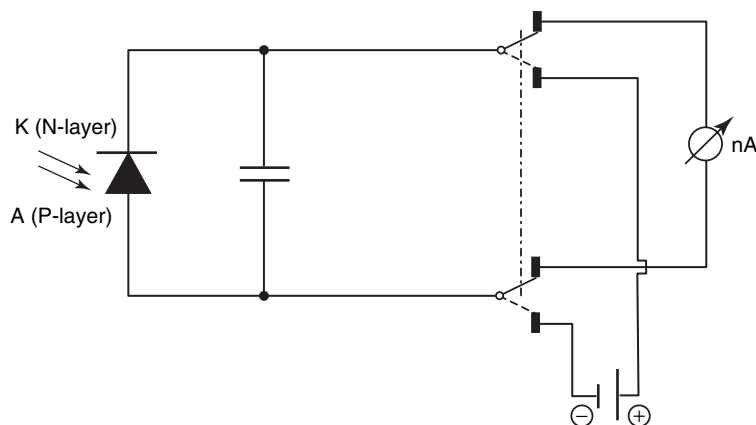
A diode array consists of a series of photodiode/capacitor parallel circuits positioned side by side on a silicon chip. Each such circuit is connected via its own solid-state switch controlled by a shift register, to a common output line. The shift register is controlled over a quartz oscillator (timer). The readout cycle, corresponding to the illumination time, falls in the 100 ms range. A schematic diagram of the diode array is given in Figure 7.

## Instrumentation

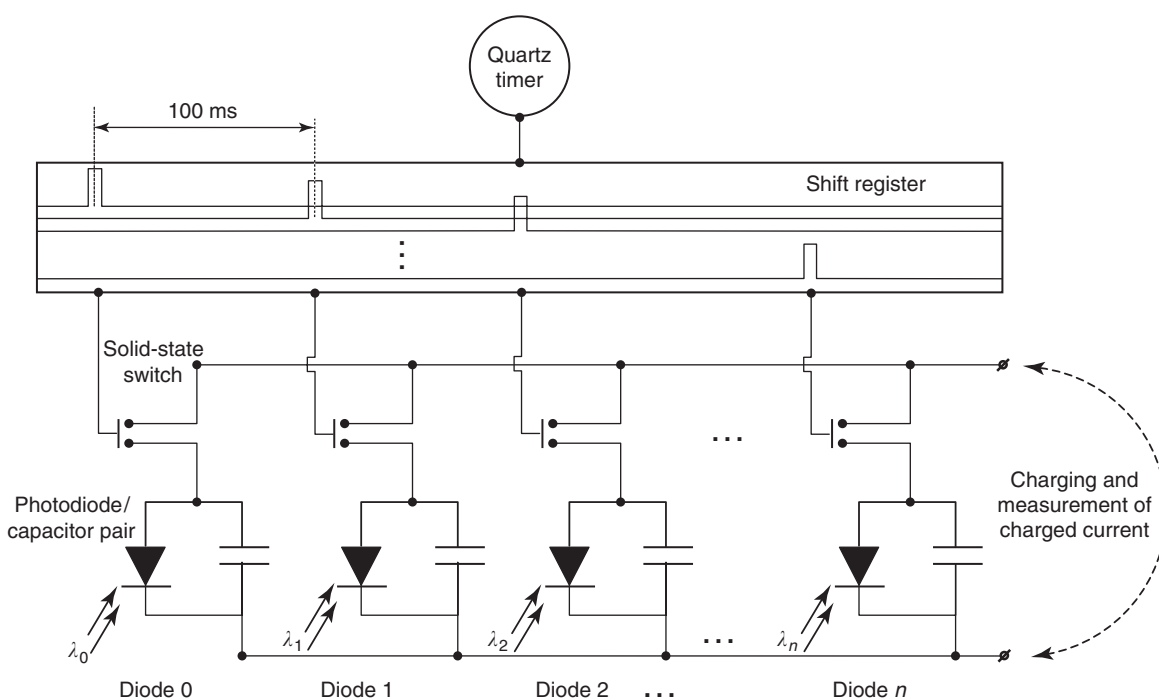
The basic design of a diode array spectrometer is given in Figure 8. Usually, spectral sources are aligned along a direction perpendicular to the sample cell. This requires a special construction of the deuterium lamp covering the UV domain (the shine through geometry) and avoids the use of a moving source selector mirror and corresponding focusing optical elements.

The shutter generally carries the holmium oxide filter (used for the calibration of the photodiode array) and the opaque surface used for the measurement of the dark current. Other optional filters can be included in order to afford fast built-in verification procedures for the spectrometer working parameters (e.g., neutral density glass filter for verification and calibration of absorbance scale in the visible domain).

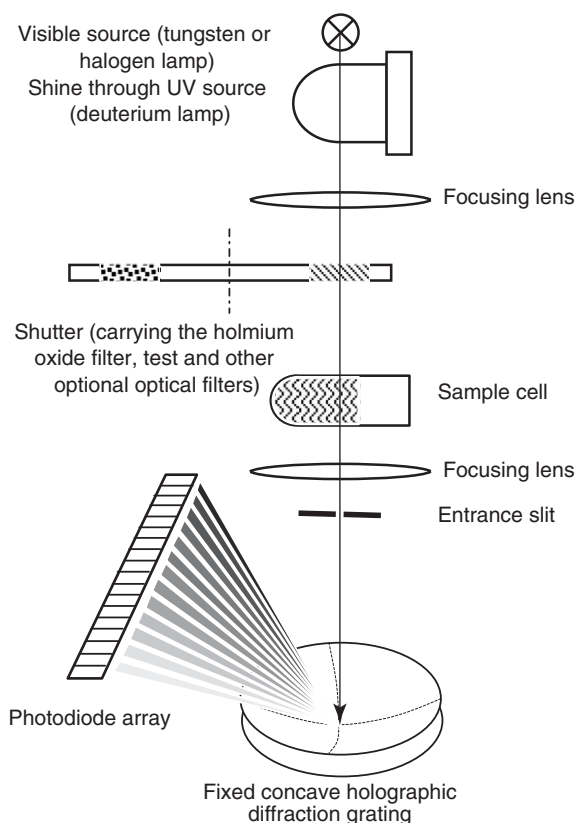
Incident light is focused on the sample cell. The transmitted radiation is refocused by means of lenses and collimated through the polychromator entrance slit, thus dispersed on the fixed holographic concave diffraction grating. Dispersed wavelengths are sampled on the photodiode array elements, positioned in the flat focal plane. The spectrum is obtained by electronic scanning through the sensitive receptor.



**Figure 6** Working principle of a photodiode array element.



**Figure 7** Schematic diagram of a photodiode array.



**Figure 8** Single-beam basic configuration of a diode array spectrometer.

Filters for reducing stray light due to the multiple high-order reflections are placed directly on the active receptor of each photodiode.

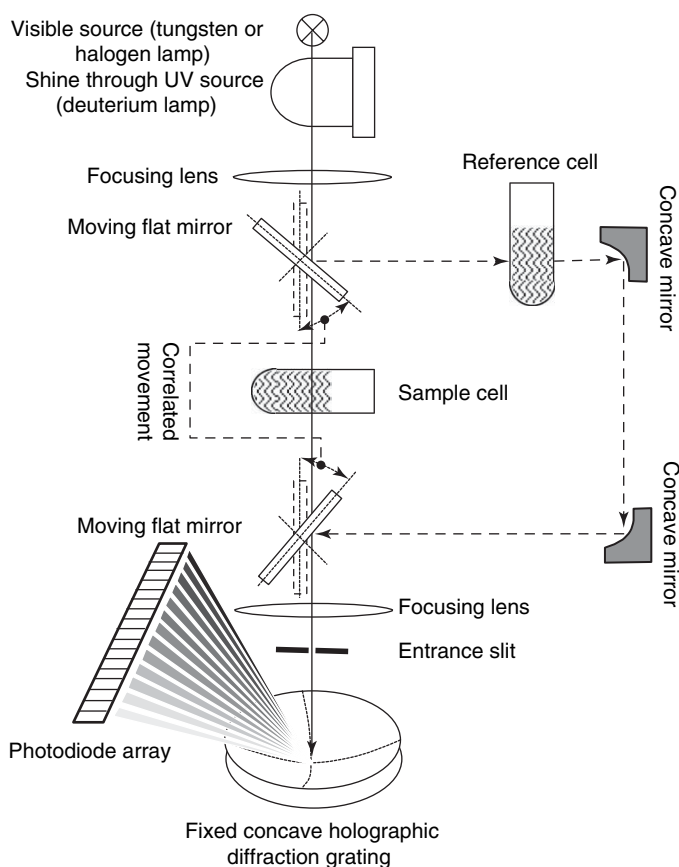
Some commercially available instruments are based on the double-beam design (see **Figure 9**). In such cases, moving beam directors (generally movement correlated flat mirrors) are used for alternatively orienting the incident beam on the reference cell and the sample cell, respectively. By controlling the movement of the flat mirrors and introducing auxiliary focusing optical reflective elements, multiple radiation beams can be obtained. Up to four sample paths have been obtained in a pseudo-double-beam design.

Some other spectrometers use different concave gratings positioned back to back. With one of the gratings, the low-resolution acquisition of the whole spectrum is achieved, while with the second grating, reduced spectral intervals are measured successively with a higher resolution.

In order to increase the resolution, dual-geometry polychromators with separate UV and visible entrance slits are also available. Each of the spectral intervals is measured sequentially on the same or on different photodiode arrays.

All these constructive alternatives suffer from increased complexity, the use of additional optically active components resulting in increased cost and





**Figure 9** Dual-beam basic configuration of a diode array spectrometer.

lower reliability together with the inherent reduction of wavelength reproducibility.

## Instrumental Parameters and Characteristics

### Resolution

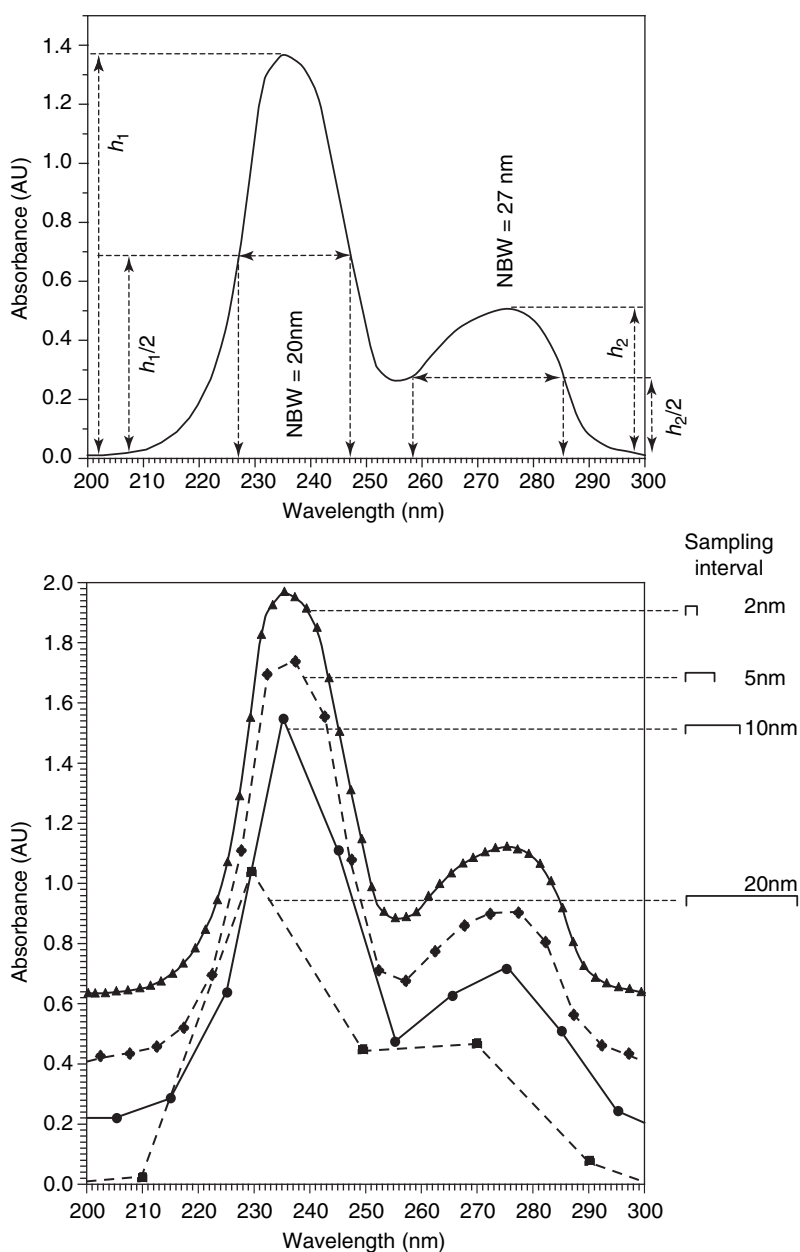
The spectral resolution for diode array spectrometers is dependent upon the number of photosensitive diodes distributed in the focal plane of the concave diffraction grating. The distance, in terms of wavelength, between two adjacent diode centers is called sampling interval. Higher resolution means not only more diodes composing the array but also smaller diodes. Spectrometers for the UV–Vis domain (180–820 nm) equipped with a 1024 photodiode array will consequently ensure a resolution of  $\sim 0.6$  nm. However, a 2 nm resolution can be considered as the usual value in diode array spectrometry, which means significantly lower capabilities compared to resolutions of 0.5, 0.2, and even 0.1 nm, which is quite usual for scanning instruments. The basic requirement in spectrometry is that the

sampling interval should be at least 10 times lower than the natural bandwidth (NBW) of the sample signal. For UV–Vis spectrometric applications on solutions, the NBW characterizing signals are generally higher than 20 nm, resulting in the acceptance of a 2 nm sampling interval. However, diode array spectrometry does not cover applications requiring high-resolution spectral acquisition. The influence of the sampling interval on resolution and the resulting spectra are illustrated in **Figure 10**.

### Dynamic Range

The dynamic range, defined as the ratio between maximum and minimum absorbance measured with 1% accuracy, is limited at high absorption values by the stray light and at low absorption levels by the noise.

Stray light generated by Fresnel reflection on lens surfaces, air bubbles in glass, and diffraction at aperture edges is less important in diode array instruments compared to conventional spectrometers because of less complex construction and lower optical surfaces. The stray light in diode array



**Figure 10** Resolution in diode array spectrometry.

spectrometry can be controlled by making measurements at absorption wavelengths in the neighborhood of the maximum, considering the inherent high wavelength reproducibility of diode array instrumentation.

Noise can be controlled in diode array spectrometry by spectral averaging. The Fellgett advantage (noise reduction equals the square root of acquired data points) leads to a significant improvement of the signal-to-noise ratio. Large bandwidth acquisition not only reduces noise but also lowers sensitivity.

### Fast Spectral Acquisition

The simultaneous sampling of all spectral channels by means of diode array spectrometry leads to real-time spectral data acquisition (usually 0.1 s for the whole spectrum). Data processing and storage may add a supplementary 0.5 s delay. Consequently, diode array spectrometry is a first choice for dynamic system measurements (e.g., flow injection analysis and liquid chromatography detection, process control, fast kinetic measurements). Spectral averaging capabilities represent a corollary of the increased acquisition speed.

Statistical information derived from spectral averaging represents an additional benefit. Calculation of standard deviation for each data point can be used to detect bad results or to select the optimum wavelength for a single-component analysis. Statistic information is also of particular importance in curve-fitting multicomponent analysis.

### Wavelength Resettability

Diode array instruments have no main moving optical elements; therefore, no mechanical errors or drift arise. Consequently, the widely accepted rule in scanning spectrometry emphasizing that an accurate quantitation requires the use of the absorption maximum as analytical wavelength is no longer critical for diode array instruments. Therefore, such a choice, especially in the case of multicomponent analysis, should be focused only on reasons related to selectivity.

### Multiwavelength Acquisition and Internal Referencing

Multiwavelength acquisition is an important tool for multicomponent analysis. The calculation of absorbance ratios at different wavelengths inside an absorption band or belonging to different absorption bands leads to confirmatory analysis and band purity check.

The internal referencing minimizes random, flow-induced effects and drifts. The principle is based on the subtraction, from the experimental values measured on a given spectral interval, of a baseline absorption value determined outside the analytical range, obtained with the same radiation source. This reduces many wavelength-dependent errors resulting from the spectral source instability or sample positioning. The effect is increased when a mean value of absorption, on wavelength interval, is used instead of a single reference value. The technique turns the single-beam instrumentation into a pseudo-dual-beam one.

### Sensitivity

The sensitivity should be considered as the ability of an instrument to detect and quantify low concentrations of the target compound. In absorption spectrometry, low concentration of the analyte means high transmitted radiation levels. This is the reason why the poorer sensitivity of silicon-based photodiodes compared to photomultipliers or coupled charge detectors is in fact not so important. The relatively lower sensitivity of photodiodes is also compensated

by low electronic noise levels and increased stability to intense radiations.

Photodiode array elements have also reduced optically sensitive area compared to a single detector. However, overall light intensity within a polychromator is higher than for the case of a scanning instrument, resulting in a compensatory effect.

Increasing the measurement time also induces sensitivity improvement due to noise reduction. Spectral averaging, however, should not be overestimated, as long as drift measurement errors will limit the acquisition duration.

### Reversed Optics

In scanning instruments, the sample is placed after the exit slit of the monochromator and in front of a wide acceptance angle detector (forward optics). In order to avoid ambient stray light to reach the detector, light-tight sample areas are required.

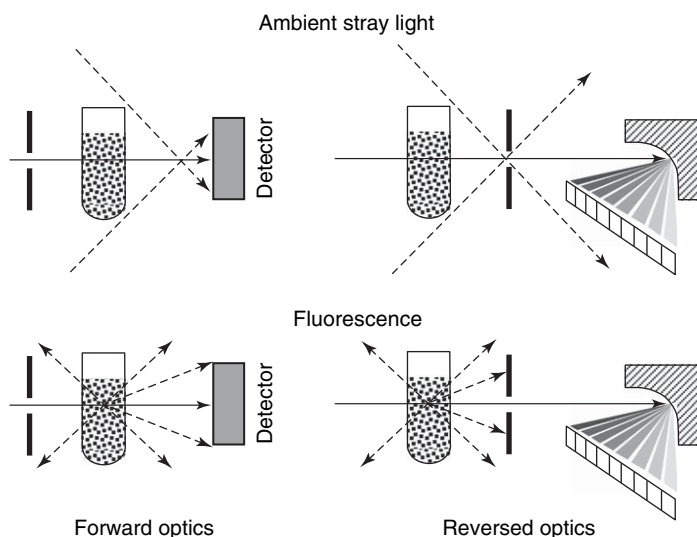
Diode array instruments have reversed optics (see Figure 11), which means that the sample is placed in front of the narrow-acceptance-angle entrance slit of the polychromator. Consequently, ambient stray light does not essentially affect measurements, even in the case of open sample area. Based on the same assumptions, fluorescence induced in samples affects diode array instruments to a lesser extent compared to the forward optics setup.

Open sample area improves productivity and allows easier accessory installation.

Forward and reversed optics are also discussed in terms of influence on sample stability. Reversed optics increases the probability of decomposition of photolabile compounds, because the whole wavelength range is passed through the sample. These are by no means conclusive results, as it is difficult to compare effects of a single full wavelength fast exposure of the sample (produced with a reversed optics instrument) with those of a cumulative long-term exposure made by a scanning instrument. Both examples of induced sample decomposition as well as successful analysis of well-known photosensitive compounds by means of diode array instrumentation have been reported in the literature.

### Built-in Calibration and Verification Features

By means of internal software routines, the emission lines of the deuterium source ( $\alpha$  line – 656.1 nm and  $\beta$  line – 486 nm, respectively) are checked for focusing on the corresponding preset  $i, j$  photodiodes from the array. This is made by sampling photodiode positions  $(i - a; i + a)$  and  $(j - a; j + a)$ , respectively. If deviations are higher than  $\pm 1$  nm (accepted displacement depends on the resolution of the instrument), the



**Figure 11** Comparative influence of ambient stray light and sample fluorescence on scanning and diode array instrumentation, respectively.

logic identification number of each diode in the array is redefined. For this purpose, the diode array contains additional sensitive elements placed on the right- and left-hand sides of its active length.

Wavelength accuracy is usually checked with the holmium oxide filter (absorption maxima at 361, 453.7, and 536.7 nm, respectively). Such a filter is brought in the optical path of the instrument by means of the shutter.

Alternatively, the shutter may introduce an opaque filter in the instrument optical path. Sampling each diode from the array for residual current with the highly absorbing filter placed in front of the entrance slit provides the dark current test result. Dark current values higher than a maximum accepted level indicate a system malfunction.

## Applications

Diode array spectrometry covers a broad range of wavelengths, from soft X-rays to the near-infrared (NIR) domain. A typical diode array spectrometer is designated to make spectral measurements from 200 to 1200 nm, with a resolution of 1 nm. A complete spectrum can be obtained in less than 100 ms. Due to these characteristics, this technique has been considered as a useful technique in different branches of analytical sciences (quantitative and qualitative determination) as well as in fundamental research.

Diode array spectrometers have been used extensively for pharmaceutical applications, mainly when the samples are not very complex and a

multicomponent procedure can be applied. For that purpose, absorption at a large number of wavelengths can be measured simultaneously, providing spectral information for solving the unknown concentration of multicomponent samples. Spectral acquisition speed represents a valuable tool for automated dissolution test applications. The high reproducibility and reliability are also cited as real advantages, demonstrating instrument performance stability over time.

NIR spectrometers based on InGaAs diode array have been realized and successfully used for the determination of nitrogen (protein origin) and water content in citrus leaves. Measurement in the 900–1800 nm spectral interval revealed a broad spectral band, with a maximum situated at ~1425 nm. This band for fresh whole leaf differs significantly from one for dried powdered leaf due to the abundant water masking the nitrogen features in the NIR spectrum.

Some other major applications of NIR diode array spectrometry are summarized in **Table 1**.

Recently, diode array systems have been used in fast transient absorption or chemiluminescence measurements due to their capability of providing extensive real-time spectral data. Enzyme kinetics as part of biochemistry relies on fast spectral multi-wavelength acquisition. At low costs, diode array instruments are ideal for portable microfluidic bioanalyzers and emerging large-scale integrated microfluidic technologies.

The use of a commercial diode array spectrometer to study the rapid reaction kinetics has great

**Table 1** Major application fields for NIR diode array spectrometry

<i>Chemical composition</i>	<i>Bulk properties</i>	<i>Physical properties</i>
Protein	Density	Particle size/fiber diameter
Humidity/water content	Digestibility	Temperature
Hydrocarbons	Viscosity	Mechanical properties
Carboxylic acids	Motor fuel octane number	Thermal and mechanical pretreatment
Amines	Reid vapor pressure of gasoline	Molar masses of polymers
Oil/fat	Seed germination	
Sucrose/glucose		
Additives in fuels	Distillation parameters	
	Fruit ripeness	
	Total dissolved solids	

potential due to the simultaneous acquisition of wavelength data. Addition of an optimized stopped-flow accessory makes such an approach affordable. For example, the reaction of Ellman's reagent with thioglycerol was used to illustrate how this combination can be used to acquire complex data, which are subsequently processed using multivariate global analysis methods. UV-Vis absorption cross-sections and atmospheric photolysis rates of some pollutants ( $\text{CH}_2\text{Br}_2$ ,  $\text{CH}_2\text{I}_2$ ,  $\text{CH}_2\text{BrI}$ ,  $\text{CH}_2\text{ICl}$ ;  $\text{CF}_3\text{I}$ ,  $\text{CH}_3\text{I}$  and  $\text{C}_2\text{H}_5\text{I}$ ) were studied by means of diode array instruments with spectral resolution of 0.6 nm. It has been proved that the reaction of these target compounds with HO radical is the dominant atmospheric loss process.

Diode array spectrometers are often used as multichannel analyzers designated for real-time spectral analysis. Spectral analysis provides useful information on the color characteristics of numerous materials and light sources. Spectral analysis techniques that use diode array technology are widely used in analyzing lamps and LED's emission and in quantifying the reflectance, transmittance, color, and haze of many materials. These instruments offer high sensitivity, low noise, and broad spectral response. Optionally, integral fiber cables, connecting the spectrometer to the light source, are used. Cables offer additional flexibility for remote positioning of the spectrometer. Spectrometers can be packed in self-contained compact enclosures that are easily connected to a computer via a serial port.

Another class of applications refers to environmental pollution monitoring. For instance, some atmospheric pollutants (ammonia, nitrogen oxides, and, in some cases, sulfur dioxide) in the flue gas stream can be analyzed simultaneously by means of UV spectrometry. Their determination at the nanogram per liter level can be achieved with a process

analyzer based on UV absorbance monitoring using diode array technology. Hydrocarbon monitoring in polluted water can be achieved by NIR diode array instruments after liquid-liquid extraction in chlorinated solvents.

Online applications are by far the most important utilization of diode array spectrometry. High-performance liquid chromatography, supercritical fluid chromatography, capillary electrophoresis, and flow-injection techniques produce enhanced sensitivity and structure-related information due to coupling with diode-array-based detectors. Emission of the microwave-induced plasma generated in atomic emission detectors for capillary gas chromatography is also analyzed by means of UV-Vis diode array instruments.

*See also:* **Spectrophotometry:** Overview; Derivative Techniques; Inorganic Compounds; Organic Compounds.

## Further Reading

- Brignell J and White N (1994) *Intelligent Sensor Systems*. Bristol: Institute of Physics Publishing.
- Harvey D (2000) *Modern Analytical Chemistry*. Boston: McGraw Hill Higher Education.
- Huber V and George SA (1993) *Diode Array Detection in HPLC*. New York: Dekker.
- Owen T (1996) *Fundamentals of Modern UV-Visible Spectroscopy*. Hewlett-Packard Co., publication no. 12-5965-5123E.
- Palmer C (2000) *Diffraction Gratings Handbook*, 5th edn. New York: Spectra Physics.
- Popov EK and Loewen EG (1997) *Diffraction Gratings and Applications*. New York: Dekker.
- Williams P and Norris K (eds.) (2001) *Near-Infrared Technology*, 2nd edn. St Paul: American Association of Cereal Chemistry Inc.

## Derivative Techniques

F Vogt, Arizona State University, Tempe, AZ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Derivative spectroscopy makes use of the first- and higher-order derivatives of absorbance or transmission spectra with respect to wavelength for improving qualitative and quantitative analysis. The two main advantages of derivative spectroscopy are (1) to enhance spectral features facilitating the analyses of samples containing substances with very similar absorption spectra and (2) to suppress large, broad background signals overlaying the absorption features of interest.

The concept of differentiating spectral data was first introduced in the 1950s, but it received little attention primarily because of experimental difficulties in generating derivative spectra with early ultraviolet (UV)/visible (Vis) spectrophotometers. Use of mathematical or optical methods to generate derivative spectra became feasible with the advent of microprocessors and microcomputers in the late 1970s. Nowadays, most instruments offer at least the first and second numerical derivatives as a standard feature. The combination of derivative spectroscopy and chemometrics for calibration and data evaluation has further increased the popularity of this technique. Derivative techniques are applied especially in UV spectroscopy for pharmaceutical, biomedical, and environmental measurement tasks and in food research.

In this article, the mathematical and experimental methods for generation and evaluation of derivative spectra are discussed.

### Generating Derivative Spectra

#### General

The absorbance spectrum  $A(\lambda)$  is linearly related to the absorption pathlength  $L$ , the wavelength-dependent molar absorption coefficient  $\varepsilon(\lambda)$ , and the concentration of an analyte  $c$ :

$$A(\lambda) = L \cdot \varepsilon(\lambda) \cdot c \quad [1]$$

There is a linear relationship between the  $n$ th-order derivative spectra and the concentration of a chemical compound in the analyzed samples:

$$\frac{d^n A(\lambda)}{d\lambda^n} = L \cdot \frac{d^n \varepsilon(\lambda)}{d\lambda^n} \cdot c \quad [2]$$

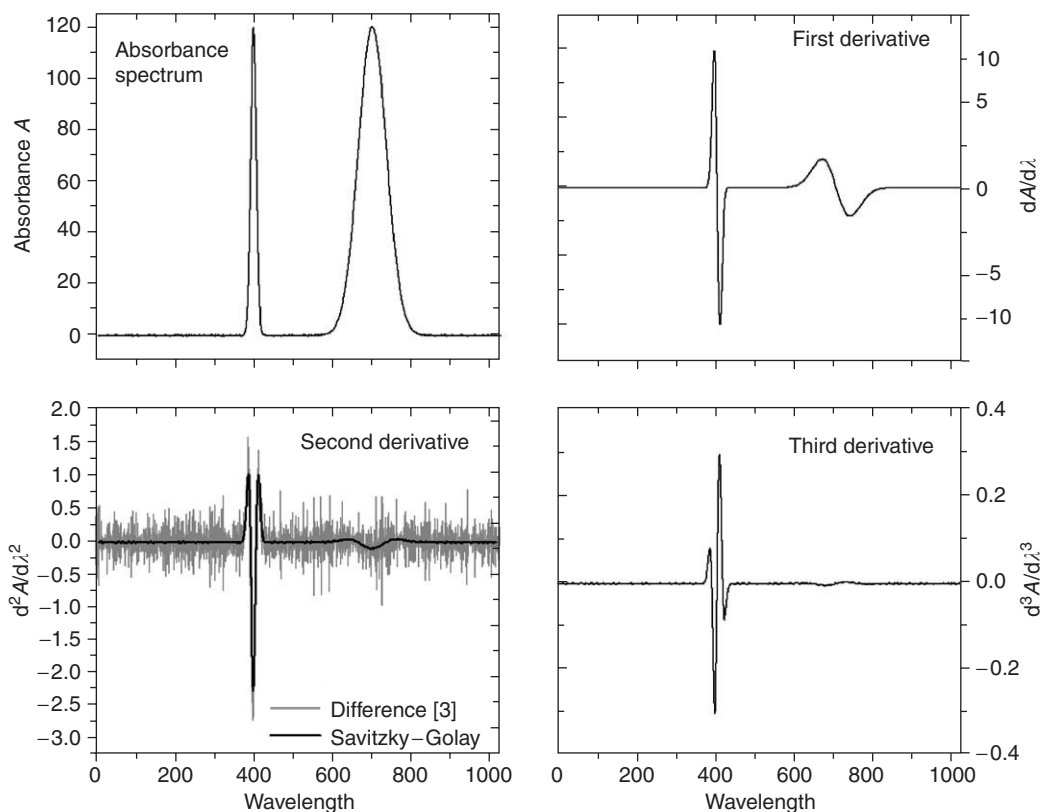
Usually, derivative orders  $n \leq 4$  are exploited in practical applications but second-order derivative spectroscopy is the most popular one.

The effect of differentiation is explained by means of Figure 1 independent from the technique used for determining derivative spectra: two synthetic Gaussian absorption bands of different width are shown in the top left graph. The first derivative (top right graph) describes how much the absorbance changes with wavelength. It is zero at wavelengths below and above the absorbance bands and it passes through zero for the maximum absorption. It is positive for wavelength below the maxima since the absorption increases. For wavelengths larger than the maxima's wavelengths the absorption decreases and hence the first derivative is negative. The most characteristic features of the second derivative (bottom left) are the negative band with minimum at  $\lambda_{\max}$  and the two positive satellite bands on either side of the main band. Note that the number of determined bands equals the order of the derivative plus one; for instance, four for the third derivative (bottom right graph). By comparing both absorption bands the tendency of derivatives to suppress broader features and pronounce narrower ones is evident.

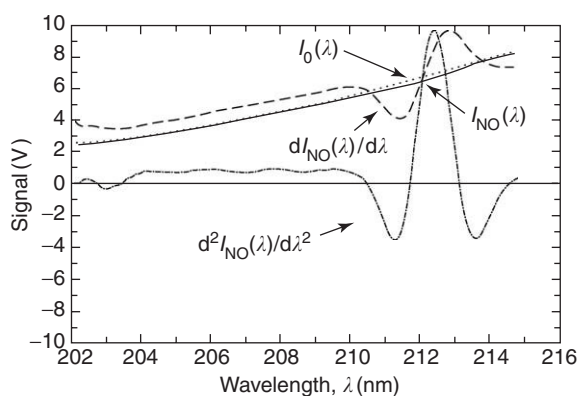
Another example involving experimental data acquired from a sample of 100 ppm gaseous nitric oxide (NO) diluted in nitrogen ( $N_2$ ) is given in Figure 2.  $I_0(\lambda)$  denotes the background spectrum obtained from a sample of pure  $N_2$  and  $I_{NO}(\lambda)$  the transmission spectrum of the NO sample. The low and featureless difference between  $I_0(\lambda)$  and  $I_{NO}(\lambda)$  makes NO quantification and the discrimination of other absorbers (not shown) difficult. The first derivative  $dI_{NO}(\lambda)/d\lambda$  and even more the second  $d^2 I_{NO}(\lambda)/d\lambda^2$  largely enhance the NO absorption features. Especially the second derivative suppresses the large background spectrum, which does not contain information but limits the resolution of the analog-to-digital conversion.

Derivative spectra can be obtained by mathematical, electronic, or optical methods. Mathematical approaches are the most common ones since they can be performed without additional hardware – a software package extension is sufficient. Electronic differentiation has been proposed but is not applied widely. Optical differentiation demands more elaborated setups; however, better results are obtained in comparison to the other two alternatives. This is discussed below.





**Figure 1** Synthetic example demonstrating numerical derivative spectra up to third order derived from a synthetic absorbance spectrum – the lower left graph compares the Savitzky–Golay method to the noise sensitive difference approach [3].



**Figure 2** Optical differentiation enhances low gaseous nitric oxide (NO) absorption features overlaid by a broad background spectrum.

### Numerical Differentiation Techniques

**Difference method** The most straightforward method for achieving first derivatives  $D_1$  of discretely sampled absorbance signals  $A$  is to calculate the difference of two consecutive measurement points number  $i$  and  $i + 1$  divided by the distance of the two points:

$$D_1(i) = \frac{A_{i+1} - A_i}{i + 1 - i} = A_{i+1} - A_i$$

The second derivative  $D_2$  is determined from the first derivative  $D_1$  in the same way:

$$D_2(i) = \frac{D_1(i+1) - D_1(i)}{i + 1 - i} = (A_{i+2} - A_{i+1}) - (A_{i+1} - A_i) = A_{i+2} - 2 \cdot A_{i+1} + A_i \quad [3]$$

Although this approach is easy to understand one should not use it for calculating derivatives since noise is amplified largely. The noise susceptibility is demonstrated in the lower left graph of **Figure 1** by comparing second-order derivatives obtained with this difference method [3] to the Savitzky–Golay approach discussed in the following paragraph, which is much less susceptible to noise.

**Savitzky–Golay method** This method determines a derivative spectrum by moving a spectral window comprising  $2 \cdot n + 1$  measurement points over an absorbance spectrum. Then a polynomial of order  $m$

$$P(\lambda) = a_0 + a_1 \cdot \lambda + a_2 \cdot \lambda^2 + \dots + a_m \cdot \lambda^m \quad [4]$$

is fitted to the measurement points inside the spectral window. This fit polynomial introduces smoothing,

which is dependent on the user selectable parameters  $n$  and  $m$ . From the resulting fit parameters  $a_0 \dots a_m$ , the derivatives at the window center  $\lambda_0$  can be derived easily:

$$\begin{aligned} \left. \frac{dP(\lambda)}{d\lambda} \right|_{\lambda_0=0} &= a_1 + a_2 \cdot 2 \cdot \lambda_0 \\ &\quad + \dots + a_m \cdot m \cdot \lambda_0^{m-1} = a_1 \\ \left. \frac{d^2P(\lambda)}{d\lambda^2} \right|_{\lambda_0=0} &= a_2 \cdot 2 + \dots + a_m \cdot m \cdot (m-1) \cdot \lambda_0^{m-2} \\ &= 2 \cdot a_2 \\ \left. \frac{d^3P(\lambda)}{d\lambda^3} \right|_{\lambda_0=0} &= a_3 \cdot 3 \cdot 2 + \dots + a_m \cdot m \\ &\quad \cdot (m-1) \cdot (m-2) \cdot \lambda_0^{m-3} = 6 \cdot a_3 \\ &\vdots \end{aligned} \quad [5]$$

Due to the design of this differentiation technique  $\lambda_0$  is located at the center of the window. Hence,  $\lambda_0 = 0$  can be assumed from a least-squares point of view – this has no physical meaning, though.

Once the derivatives are determined at  $\lambda_0$ , the window is moved one measurement point to the right followed by a polynomial fit inside this new window. Derivatives at the centers of these new windows are calculated again by means of [5]. This procedure is continued until the window reaches the end of the spectrum. A consequence of this mathematical method is that  $n$  data points are lost for the derivative spectra at both ends of the spectrum. The window width is a central parameter of the Savitzky–Golay differentiation method. A narrow window preserves more spectral features whereas a wider one introduces more smoothing. For example, **Figure 1** has been generated using a fourth-order polynomial and a window width of 21 points.tpb 1pc

### Electronic Differentiation Techniques

The output signal of a detector is proportional to the wavelength-dependent radiation intensity  $I(\lambda)$  illuminating the detector element. If a monochromator scans continuously over a certain wavelength interval, the measured radiation intensity is transformed into a time-dependent signal  $I(t)$ . An appropriate preamplifier modifies  $I(t)$  to an output voltage through an amplification factor  $g$ :

$$U_{\text{preamp}}(t) = g \cdot I(t) \quad [6]$$

$U_{\text{preamp}}(t)$  is then fed into an electronic differentiator. The basic electronic circuit for differentiating time-dependent signals utilizes an operational amplifier, a resistor  $R$ , and a capacitor  $C$ . This electronic

differentiator outputs:

$$U_{\text{out}}(t) = -R \cdot C \cdot \frac{dU_{\text{preamp}}(t)}{dt} \quad [7]$$

The capacitor is the main component of this circuit as it blocks DC parts of  $U_{\text{preamp}}(t)$  at the input of an operational amplifier and lets only AC parts pass, i.e., changes in  $U_{\text{preamp}}(t)$ . If the input voltage changes slowly, the output voltage is low; if there are fast changes on the input, the output voltage is high. Combining [6] and [7] leads to a signal, which is proportional to the time-dependent change of the intensity  $I(t)$  measured by the detector, i.e., the first derivative:

$$U_{\text{out}}(t) = -R \cdot C \cdot g \cdot \frac{dI(t)}{dt}$$

Given a known wavelength scan rate,  $dI(t)/dt$  can be transformed to  $dI(\lambda)/d\lambda$ . Cascaded circuits can be designed in order to determine higher derivatives.

### Optical Differentiation Techniques

Optical techniques for generating derivatives are based on wavelength-modulation strategies. A prominent method utilizes a laser diode, whose emission wavelength is tuned continuously over a certain wavelength range by adjusting the laser current. Another common wavelength-modulation technique is shown in **Figure 3**: a monochromator, usually a reflection grating, scans stepwise over the selected wavelength range. At every discrete wavelength position  $\lambda_0$ , the monochromator vibrates around an axis of rotation sinusoidally back and forth several times. In this way, the wavelength selected by the monochromator/slit configuration is a function of time. In the following equation,  $a$  denotes the amplitude of the modulation and  $\omega$  the vibration or modulation frequency:

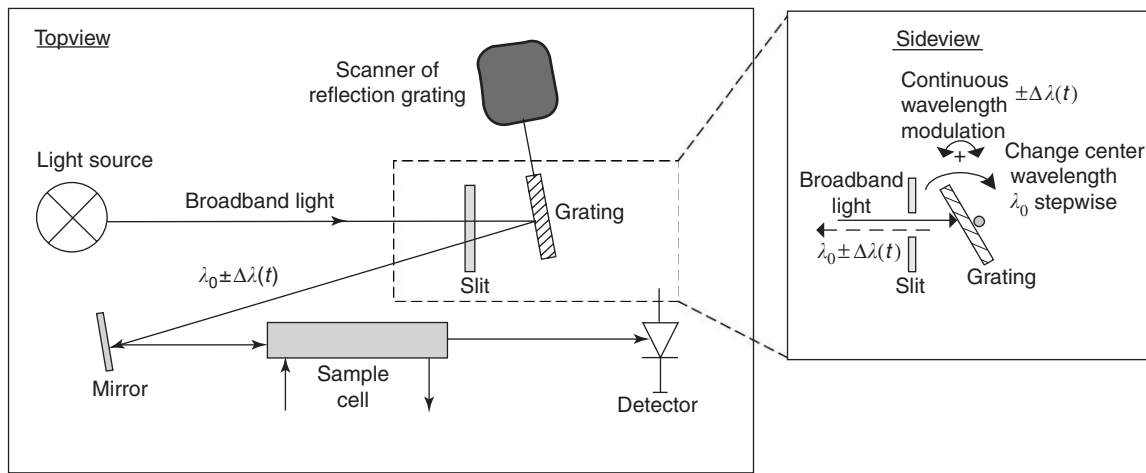
$$\lambda(t) = \lambda_0 \pm \Delta\lambda(t) = \lambda_0 + a \sin(\omega t) \Leftrightarrow \lambda - \lambda_0 = a \sin(\omega t) \quad [8]$$

The transmission spectrum (Beer's law)

$$I(\lambda) = I_0(\lambda) \exp[-L \cdot \varepsilon(\lambda) \cdot c] \quad [9]$$

with  $I_0(\lambda)$  as emission spectrum of the light source is expanded into Taylor's series around  $\lambda_0$ :

$$\begin{aligned} I(\lambda) &= I(\lambda_0, t) \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \cdot I^{(n)}(\lambda_0) \cdot (\lambda - \lambda_0)^n \end{aligned}$$



**Figure 3** Principle of wavelength modulation for generation of optical derivatives.

In [10],  $I^{(n)}(\lambda_0)$  denotes the  $n$ th derivative of the transmission  $I(\lambda)$  with respect to wavelength  $\lambda$  at the position  $\lambda_0$ . This expansion [10] is abbreviated to the order of required derivative; for instance, to second order. Using [8] in [10] derives an approximation of the transmission spectrum with  $I'(\lambda_0)$  and  $I''(\lambda_0)$  denoting the first and second derivatives at  $\lambda_0$ , respectively:

$$\begin{aligned} I(\lambda) &\approx I(\lambda_0) + I'(\lambda_0) \cdot a \sin(\omega t) \\ &\quad + \frac{1}{2} \cdot I''(\lambda_0) \cdot [a \sin(\omega t)]^2 \\ &= I(\lambda_0) + I''(\lambda_0) \cdot \frac{a^2}{4} + I'(\lambda_0) \cdot a \sin(\omega t) \\ &\quad - I''(\lambda_0) \cdot \frac{a^2}{4} \cos(2\omega t) \end{aligned} \quad [11]$$

In [11], the following formula was used:

$$\sin^2(\omega t) = \frac{1}{2} \cdot [1 - \cos(2\omega t)]$$

The wavelength modulation of the transmission intensity  $I(\lambda)$  induces in [6] a DC detector signal as well as AC signals with frequency  $\omega$  and higher harmonics. Due to the factorial factor  $1/n!$  [10], higher harmonic signals are decreasing quickly with  $n$ . In second-order derivative spectroscopy, the preamplifier output is directed to three different narrow band filters: a low pass filter determines the DC component and two lock-in amplifiers with reference frequencies  $\omega$  and  $2\omega$  measure the AC components of the signal at these frequencies. The resulting three signals are proportional to the transmission spectrum, the first and second derivatives:

$$\text{DC}(\lambda_0) \propto I(\lambda_0) + I''(\lambda_0) \cdot \frac{a^2}{4} \approx I(\lambda_0)$$

$$\text{AC}_\omega(\lambda_0) \propto I'(\lambda_0) \cdot a \quad [12]$$

$$\text{AC}_{2\omega}(\lambda_0) \propto -I''(\lambda_0) \cdot \frac{a^2}{4}$$

The approximation made in the first line of [12] is allowed since the contribution from the second derivative to the DC term is usually a few orders of magnitude lower than  $I(\lambda_0)$ . If the modulation amplitude  $a$  is constant during all measurements,  $a$  and the factor  $-1/4$  are just scaling factors. Once these three signals are measured at  $\lambda_0$  the monochromator is moved to the next wavelength position, and so forth.

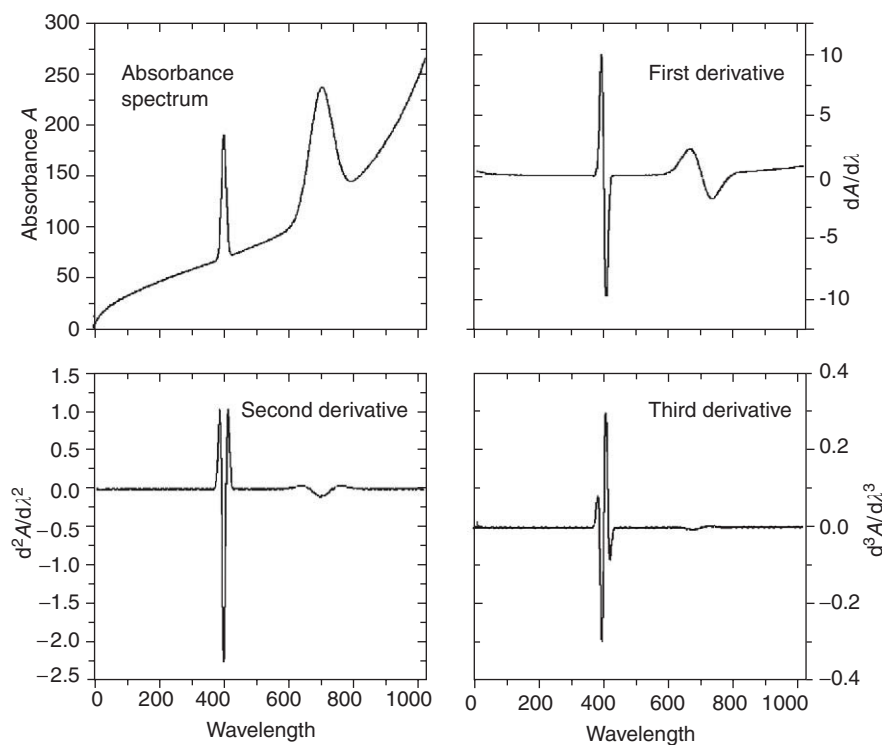
## Evaluation of Derivative Spectra

### Suppression of Background Spectra or Matrix Effects

Baseline drifts due to lamp and detector instabilities or due to optics getting dirty cause major concentration errors in spectrophotometry. Because background drifts are usually broad compared to absorption features suppression of such baseline shifts can be achieved by differentiating the spectra. **Figures 1 and 4** are based on the same spectroscopic features. However, in **Figure 4** (top left graph) a background drift unequal to a polynomial was simulated additionally. The first-derivative spectra (**Figures 1 and 4** right graphs) are slightly different, second and higher orders are basically equal. Hence, by means of second or higher derivatives such disturbing background drifts can be removed.

### Enhancing Low and Similar Absorbance Features

In spectrophotometry, the relevant absorption features are usually overlaid by a broad background



**Figure 4** Suppression of broad spectral features assigned to baseline shifts or matrix effects.

spectrum. Even if this background spectrum is stable in time, it limits the resolution of the analog-to-digital conversion. If most bits are needed for the background spectrum, which does not contain information, only a low number of bits are left for resolving the relevant absorption features. This limits the concentration resolution.

Experimental examples of enhancing minute spectral differences are given in **Figures 2** and **5**: in **Figure 2**, data obtained from a gaseous sample are shown; in **Figure 5**, aqueous samples of benzene, toluene, and *p*-xylene are considered. All three UV transmission spectra (left column of **Figure 5**) look very similar and are mainly determined by the emission spectrum of the light source; therefore, no compound-specific spectral features can be detected. Small differences are found in the first optical derivative spectra and even more in the second optical derivatives. These derivatives enable the discrimination of the analytes and the multicomponent quantification can be improved largely.

### Determination of Analytes

Derivatives of absorbance spectra

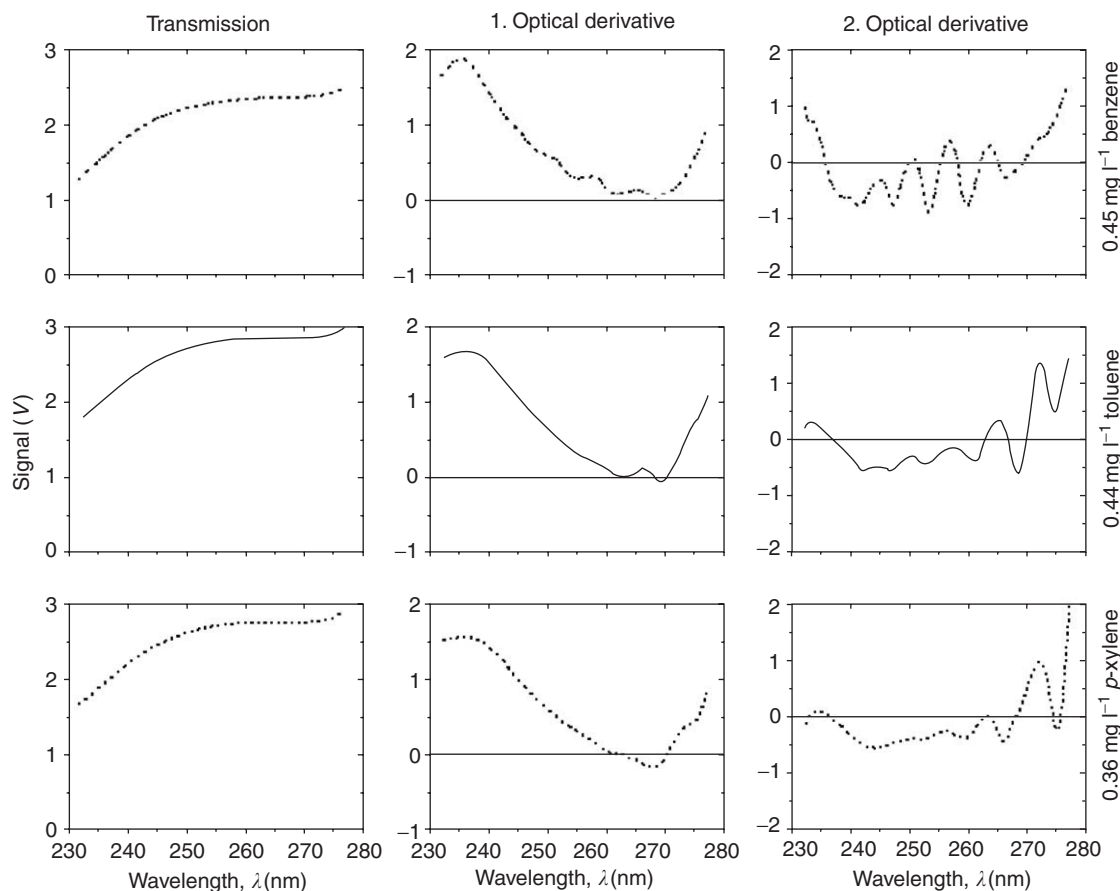
$$A(\lambda) = \ln \left[ \frac{I_0(\lambda)}{I(\lambda)} \right] = L \cdot \varepsilon(\lambda) \cdot c \quad [13]$$

are directly proportional to concentration of analytes [1]. In the following, it will be shown, how the first and second derivatives of the absorption spectra  $A'(\lambda)$  and  $A''(\lambda)$  are derived from the measured background and transmission spectrum as well as their first and second derivatives. All three types of derivatives, i.e., numerical, electronic, and optical, can be used in the same way for this purpose. Differentiating the transmission spectrum [9] using the product rule results in:

$$I'(\lambda) = I'_0(\lambda) \exp[-L \cdot \varepsilon(\lambda) \cdot c] - L \cdot \varepsilon'(\lambda) \cdot c \cdot I_0(\lambda) \exp[-L \cdot \varepsilon(\lambda) \cdot c] \quad [14]$$

In order to get a linear relation between measured data  $I'(\lambda)$ ,  $I'_0(\lambda)$ ,  $I_0(\lambda)$ , and concentration  $c$ , this equation is divided by [9]:

$$\begin{aligned} \frac{I'(\lambda)}{I(\lambda)} &= \frac{I'_0(\lambda) \exp[-L \cdot \varepsilon(\lambda) \cdot c]}{I_0(\lambda) \exp[-L \cdot \varepsilon(\lambda) \cdot c]} \\ &\quad - L \cdot \varepsilon'(\lambda) \cdot c \cdot \frac{I_0(\lambda) \exp[-L \cdot \varepsilon(\lambda) \cdot c]}{I_0(\lambda) \exp[-L \cdot \varepsilon(\lambda) \cdot c]} \\ &= \frac{I'_0(\lambda)}{I_0(\lambda)} - L \cdot \varepsilon'(\lambda) \cdot c \\ &\Downarrow \\ A'(\lambda) &= L \cdot \varepsilon'(\lambda) \cdot c = \frac{I'_0(\lambda)}{I_0(\lambda)} - \frac{I'(\lambda)}{I(\lambda)} \end{aligned} \quad [15]$$



**Figure 5** Enhancing the minute differences of analytes with very similar absorption spectra using optical derivative spectra. (Reprinted with permission from Vogt F, Tacke M, Jakusch M, and Mizaikoff B (2000) An ultraviolet spectroscopic method for monitoring aromatic hydrocarbons dissolved in water. *Analytica Chimica Acta* 422: 187–198; © Elsevier; cf. Erratum: (2001) *Analytica Chimica Acta* 431: 167.)

Differentiating [14] a second time and dividing the resulting equation by [9] relates the concentration to a measured transmission spectrum and its first and second derivatives:

$$\begin{aligned} \frac{I''(\lambda)}{I(\lambda)} &= \frac{I_0''(\lambda)}{I_0(\lambda)} - L \cdot \varepsilon''(\lambda) \cdot c \\ &\quad - 2 \cdot \frac{I_0'(\lambda)}{I_0(\lambda)} \cdot \underbrace{\varepsilon'(\lambda) \cdot L \cdot c}_{(*)} \\ &\quad + \left( \underbrace{(L \cdot \varepsilon'(\lambda) \cdot c)^2}_{(*)} \right)^2 \end{aligned}$$

The (\*)-marked terms in the latter equation are replaced by the right-hand side of [15]:

$$\begin{aligned} \frac{I''(\lambda)}{I(\lambda)} &= \frac{I_0''(\lambda)}{I_0(\lambda)} - L \cdot \varepsilon''(\lambda) \cdot c \\ &\quad - 2 \cdot \frac{I_0'(\lambda)}{I_0(\lambda)} \cdot \left( \frac{I_0'(\lambda)}{I_0(\lambda)} - \frac{I'(\lambda)}{I(\lambda)} \right) \\ &\quad + \left( \frac{I_0'(\lambda)}{I_0(\lambda)} - \frac{I'(\lambda)}{I(\lambda)} \right)^2 \end{aligned}$$

$$\begin{aligned} &= \frac{I_0''(\lambda)}{I_0(\lambda)} - L \cdot \varepsilon''(\lambda) \cdot c - 2 \cdot \left( \frac{I_0'(\lambda)}{I_0(\lambda)} \right)^2 \\ &\quad + 2 \cdot \frac{I_0'(\lambda)}{I_0(\lambda)} \cdot \frac{I'(\lambda)}{I(\lambda)} + \left( \frac{I_0'(\lambda)}{I_0(\lambda)} \right)^2 \\ &\quad - 2 \cdot \frac{I_0'(\lambda)}{I_0(\lambda)} \cdot \frac{I'(\lambda)}{I(\lambda)} + \left( \frac{I'(\lambda)}{I(\lambda)} \right)^2 \end{aligned}$$

Rearranging this equation results in the final equation relating the measured data linearly to the concentration  $c$ :

$$\begin{aligned} A''(\lambda) &= L \cdot \varepsilon''(\lambda) \cdot c = \left\{ \frac{I_0''(\lambda)}{I_0(\lambda)} - \left( \frac{I_0'(\lambda)}{I_0(\lambda)} \right)^2 \right\} \\ &\quad - \left\{ \frac{I''(\lambda)}{I(\lambda)} - \left( \frac{I'(\lambda)}{I(\lambda)} \right)^2 \right\} \\ &\approx \frac{I_0''(\lambda)}{I_0(\lambda)} - \frac{I''(\lambda)}{I(\lambda)} \end{aligned} \quad [16]$$

The approximation made in the third line of [16] is acceptable for several applications and simplifies experimental setups for measuring optical derivatives

because only one lock-in amplifier is needed for determining  $AC_{2\omega}$  [12].

This discussion can be easily extended to multi-component evaluation since Beer's law states that the absorbances  $A_{q=1\dots Q}$  of  $Q$  analytes superimpose linearly, i.e., [13] is replaced by:

$$\begin{aligned} A(\lambda) &= \ln \left[ \frac{I_0(\lambda)}{I(\lambda)} \right] = \sum_{q=1}^Q A_q(\lambda) \\ &= \sum_{q=1}^Q L \cdot \varepsilon_q(\lambda) \cdot c_q \end{aligned} \quad [17]$$

Equation [15] is replaced by:

$$A'(\lambda) = \frac{I'_0(\lambda)}{I_0(\lambda)} - \frac{I'(\lambda)}{I(\lambda)} = \sum_{q=1}^Q L \cdot \varepsilon'_q(\lambda) \cdot c_q \quad [18]$$

and equation [16] is replaced by:

$$\begin{aligned} A''(\lambda) &= \left\{ \frac{I''_0(\lambda)}{I_0(\lambda)} - \left( \frac{I'_0(\lambda)}{I_0(\lambda)} \right)^2 \right\} \\ &\quad - \left\{ \frac{I''(\lambda)}{I(\lambda)} - \left( \frac{I'(\lambda)}{I(\lambda)} \right)^2 \right\} = \sum_{q=1}^Q L \cdot \varepsilon''_q(\lambda) \cdot c_q \end{aligned} \quad [19]$$

Because the equations for the conventional transmission spectroscopy [17], the first [18], and the second derivative [19] are all of the same type, the following concepts can readily be adapted to all three cases. For example, in second-order derivative spectroscopy, the right-hand side of [19] is rewritten in vector notation as this facilitates the discussion, which paves the way to combine derivative spectrophotometry and chemometrics for sophisticated calibration and data evaluation tools:

$$A''(\lambda) = (L \cdot \varepsilon''_1(\lambda) \cdots L \cdot \varepsilon''_Q(\lambda)) \cdot \begin{pmatrix} c_1 \\ \vdots \\ c_Q \end{pmatrix} \quad [20]$$

If concentrations of  $Q > 1$  different analytes are to be determined,  $A''(\lambda)$  must be measured at  $N \geq Q$  different wavelength positions in order to set up an equation system:

$$\begin{pmatrix} A''(\lambda_1) \\ \vdots \\ A''(\lambda_N) \end{pmatrix} = \begin{pmatrix} L \cdot \varepsilon''_1(\lambda_1) & \cdots & L \cdot \varepsilon''_Q(\lambda_1) \\ \vdots & & \vdots \\ \vdots & \ddots & \vdots \\ L \cdot \varepsilon''_1(\lambda_N) & \cdots & L \cdot \varepsilon''_Q(\lambda_N) \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ \vdots \\ c_Q \end{pmatrix} \quad [21]$$

$$\mathbf{A}''_{(N \times 1)} = \mathbf{E}_{(N \times Q)} \cdot \mathbf{c}_{(Q \times 1)}$$

Calibration samples are used to determine the calibration matrix  $\mathbf{E}$ ; the concentrations  $\mathbf{c}$  of unknown

samples can then be determined from the measurement data vector  $\mathbf{A}''$  by means of solving the (over-determined) equation system [21]. Chemometrics provide the mathematical tools for these computations.

It is important to note that, by means of this approach ([21]), overlapping spectral features can be evaluated – there is no need for isolated absorption bands.

## Instrumental Assessments

### General

The use of numerical and electronic derivatives is not free from problems concerning the signal-to-noise ratio and versatility because not only baseline drifts are suppressed by differentiating spectra, but also broad, relevant absorption features (see **Figure 4**, absorption band  $\sim 700$  wavelength units).

### Signal-to-Noise Ratio

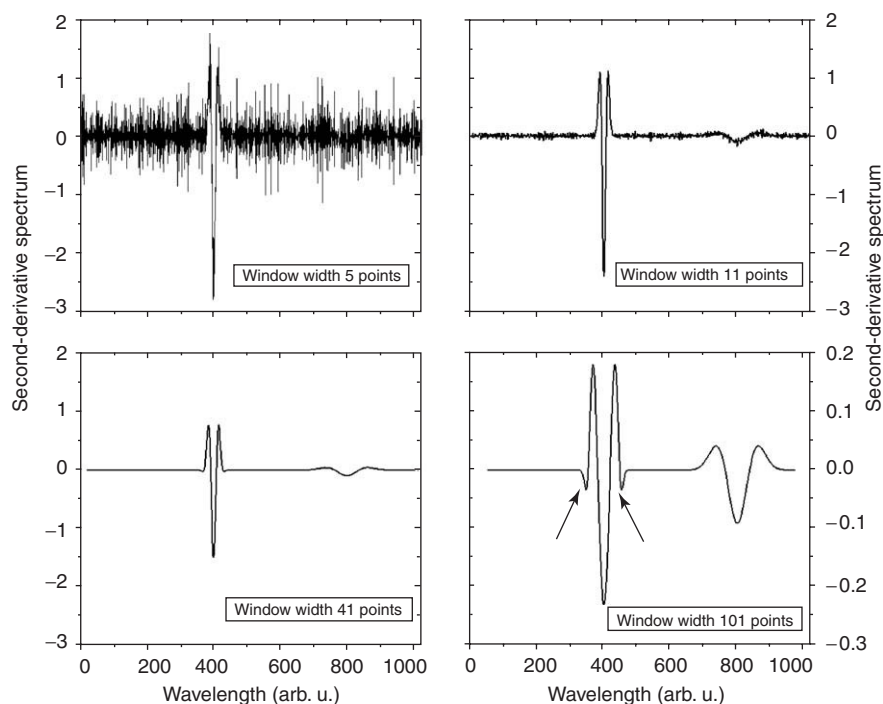
By comparing the different y-axis scales in **Figure 1**, it is obvious that numerical derivatives generally decrease the signal level because low-frequency signal components are suppressed. On the other hand, high-frequency signal components including noise are enhanced. The combination of suppressing low-frequency signals, which mostly represent relevant spectral information, and the enhancement of high-frequency noise decreases the signal-to-noise ratio.

Because numerical derivatives are computed from spectra, which are measured with a broadband amplifier, noise at all frequencies affects the measurement spectra. Optical derivatives, however, utilize a DC amplifier with a high time constant and lock-in amplifiers with narrow bandwidth  $\delta\omega$ . Only noise with frequencies  $\omega \pm \delta\omega/2$  and  $2\omega \pm \delta\omega/2$  influences the first and second derivatives. The DC measurement can only be affected by long time drifts – AC noise is averaged out. This improves the signal-to-noise ratio compared to the numerical approach.

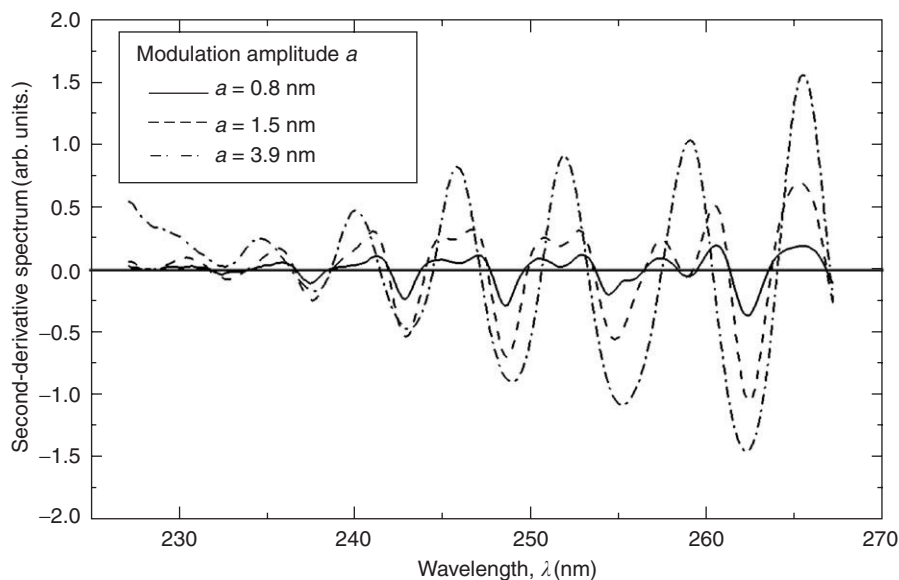
### Influence of the Window Width and the Modulation Amplitude on Derivatives

Of the two parameters the user has to select for the Savitzky–Golay method, the polynomial order  $m$  [4] is the less important one. Much more important is the window width  $2n+1$ . In **Figure 6**, second-derivative spectra of a synthetic absorption spectrum (**Figure 1**, upper left graph) are compared, which have been calculated using different window widths. A too narrow window does not incorporate sufficient smoothing (upper left graph) and suppresses relevant broad features – a too broad window causes artifacts





**Figure 6** Influence of the Savitzky-Golay window width on the shape of the derivatives – shown are second derivatives of the absorption spectrum plotted in the upper left graph of **Figure 1**.



**Figure 7** Influence of the modulation amplitude  $a$  [8] on the shape of a second-order derivative spectrum [19]. The example was determined by means of an aqueous sample containing approx.  $1 \text{ mg l}^{-1}$  of each benzene, toluene,  $p$ -xylene. (Reprinted with permission from Vogt F, Tacke M, Jakusch M, and Mizaikoff B (2000) An ultraviolet spectroscopic method for monitoring aromatic hydrocarbons dissolved in water. *Analytica Chimica Acta* 422: 187–198; © Elsevier; cf. Erratum: (2001) *Analytica Chimica Acta* 431: 167.)

(marked by arrows in the lower right graph). Fine tuning is necessary for a specific application in order to optimize smoothing and signal intensity (upper right and lower left graph).

Although electronic derivative spectroscopy has been proposed, such techniques never found broad acceptance or applications. One reason might be that optimizations for different applications are not

possible as the differentiation is defined unchangeable by the resistor and the capacitor of the differentiating circuit [7].

Analogous to the adjustment of the window width, different modulation amplitudes  $a$  [8] can be chosen in optical derivative spectroscopy. According to [12], larger derivative signals and hence better signal-to-noise ratio can be gained by increasing  $a$ . This was confirmed experimentally; however, a signal increase comes along with a washing out of derivative shapes and minute derivative features are lost (Figure 7). The values for the modulation amplitude given in Figure 7 are typical examples. Thus, the modulation amplitude can be optimized for different applications: if several analytes in mixtures have very similar and overlapping absorption spectra, a small modulation amplitude helps to pronounce minor differences for discrimination. If, however, clearly different absorption spectra are present, a large modulation amplitude is selected for improved signal-to-noise ratio.

See also: **Chemometrics and Statistics:** Multivariate Calibration Techniques. **Optical Spectroscopy:** Radiation Sources; Wavelength Selection Devices; Detection Devices. **Spectrophotometry:** Overview.

## Further Reading

- Bosch OC, Sanchez RF, and Cano PJ (1995) Recent developments in derivative ultraviolet/visible absorption spectrophotometry. *Talanta* 42: 1195–1214.
- Brown C, Vega-Montoto L, and Wentzell P (2000) Derivative preprocessing and optimal correction for

baseline drift in multivariate calibration. *Applied Spectroscopy* 54: 1055–1068.

- Dixit L and Ram S (1985) Quantitative analysis by derivative electronic spectroscopy. *Applied Spectroscopy Review* 21: 311–418.
- Fell A (1983) Biomedical applications of derivative spectroscopy. *Trends in Analytical Chemistry* 2(3): 63–66.
- Hager R and Anderson R (1970) Theory of the derivative spectrometer. *Journal of the Optical Society of America* 60: 1444–1449.
- Hawthorne A and Thorngate J (1978) Improving analysis from second-derivative UV-absorption spectrometry. *Applied Optics* 17: 724–729.
- Knowles A and Burgess C (1984) *Practical Absorption Spectrometry/Ultraviolet Spectrometry*. New York: Chapman & Hall.
- O'Haver TC and Begley T (1981) Signal-to-noise in higher order derivative spectrometry. *Analytical Chemistry* 53: 1876–1878.
- Sassenscheid K, Klocke U, Marb C, *et al.* (1998) Dynamic derivative UV-spectroscopy for combustion monitoring. *Proceedings of SPIE* 3535: 204–214.
- Sassenscheid K, Klocke U, Marb C, *et al.* (1998) Enhanced selectivity and sensitivity in UV-analysis of volatile organic compounds. *Proceedings of SPIE* 3533: 222–233.
- Savitzky A and Golay M (1964) Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* 36: 1627–1639.
- Steiner J, Termonia Y, and Deltour J (1972) Smoothing and differentiation of data by simplified least square procedure. *Analytical Chemistry* 44: 1906–1909.
- Vogt F, Klocke U, Rebstock K, *et al.* (1999) Optical UV derivative-spectroscopy for monitoring gaseous emissions. *Applied Spectroscopy* 53: 1352–1360.
- Williams D and Hager R (1970) The derivative spectrometer. *Applied Optics* 9: 1597–1605.

## Turbidimetry and Nephelometry

**D M Lawler**, University of Birmingham, Birmingham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Turbidity is an expression of the optical property of a medium, which causes light to be scattered and absorbed rather than transmitted in straight lines through the sample. The International Organization for Standardization (ISO) defines turbidity as the 'reduction of transparency of a liquid caused by the presence of undissolved matter'. It is, therefore, the opposite of clarity. The medium concerned is normally a fluid (but may be solid) in which light is

scattered by matter – usually small particles – suspended in the light path. Measurements of turbidity can be used in many analytical fields to determine the mass concentration of suspended particles in a sample and, for some simple contexts, particle size distributions. The field is hampered, however, by a lack of standardization in units, measurement devices and calibration techniques. Analytical determinations of concentrations tend to be empirical. Such methodological problems have recently driven a profusion of technical papers.

This article reviews turbidity theory, measurement principles, instrumentation systems, and applications, with particular reference to suspended sediment concentrations (SSCs) in natural waters (e.g., rivers, estuaries, and nearshore zones).

possible as the differentiation is defined unchangeable by the resistor and the capacitor of the differentiating circuit [7].

Analogous to the adjustment of the window width, different modulation amplitudes  $a$  [8] can be chosen in optical derivative spectroscopy. According to [12], larger derivative signals and hence better signal-to-noise ratio can be gained by increasing  $a$ . This was confirmed experimentally; however, a signal increase comes along with a washing out of derivative shapes and minute derivative features are lost (Figure 7). The values for the modulation amplitude given in Figure 7 are typical examples. Thus, the modulation amplitude can be optimized for different applications: if several analytes in mixtures have very similar and overlapping absorption spectra, a small modulation amplitude helps to pronounce minor differences for discrimination. If, however, clearly different absorption spectra are present, a large modulation amplitude is selected for improved signal-to-noise ratio.

See also: **Chemometrics and Statistics:** Multivariate Calibration Techniques. **Optical Spectroscopy:** Radiation Sources; Wavelength Selection Devices; Detection Devices. **Spectrophotometry:** Overview.

## Further Reading

- Bosch OC, Sanchez RF, and Cano PJ (1995) Recent developments in derivative ultraviolet/visible absorption spectrophotometry. *Talanta* 42: 1195–1214.
- Brown C, Vega-Montoto L, and Wentzell P (2000) Derivative preprocessing and optimal correction for

baseline drift in multivariate calibration. *Applied Spectroscopy* 54: 1055–1068.

- Dixit L and Ram S (1985) Quantitative analysis by derivative electronic spectroscopy. *Applied Spectroscopy Review* 21: 311–418.
- Fell A (1983) Biomedical applications of derivative spectroscopy. *Trends in Analytical Chemistry* 2(3): 63–66.
- Hager R and Anderson R (1970) Theory of the derivative spectrometer. *Journal of the Optical Society of America* 60: 1444–1449.
- Hawthorne A and Thorngate J (1978) Improving analysis from second-derivative UV-absorption spectrometry. *Applied Optics* 17: 724–729.
- Knowles A and Burgess C (1984) *Practical Absorption Spectrometry/Ultraviolet Spectrometry*. New York: Chapman & Hall.
- O'Haver TC and Begley T (1981) Signal-to-noise in higher order derivative spectrometry. *Analytical Chemistry* 53: 1876–1878.
- Sassenscheid K, Klocke U, Marb C, *et al.* (1998) Dynamic derivative UV-spectroscopy for combustion monitoring. *Proceedings of SPIE* 3535: 204–214.
- Sassenscheid K, Klocke U, Marb C, *et al.* (1998) Enhanced selectivity and sensitivity in UV-analysis of volatile organic compounds. *Proceedings of SPIE* 3533: 222–233.
- Savitzky A and Golay M (1964) Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* 36: 1627–1639.
- Steiner J, Termonia Y, and Deltour J (1972) Smoothing and differentiation of data by simplified least square procedure. *Analytical Chemistry* 44: 1906–1909.
- Vogt F, Klocke U, Rebstock K, *et al.* (1999) Optical UV derivative-spectroscopy for monitoring gaseous emissions. *Applied Spectroscopy* 53: 1352–1360.
- Williams D and Hager R (1970) The derivative spectrometer. *Applied Optics* 9: 1597–1605.

## Turbidimetry and Nephelometry

D M Lawler, University of Birmingham, Birmingham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Turbidity is an expression of the optical property of a medium, which causes light to be scattered and absorbed rather than transmitted in straight lines through the sample. The International Organization for Standardization (ISO) defines turbidity as the 'reduction of transparency of a liquid caused by the presence of undissolved matter'. It is, therefore, the opposite of clarity. The medium concerned is normally a fluid (but may be solid) in which light is

scattered by matter – usually small particles – suspended in the light path. Measurements of turbidity can be used in many analytical fields to determine the mass concentration of suspended particles in a sample and, for some simple contexts, particle size distributions. The field is hampered, however, by a lack of standardization in units, measurement devices and calibration techniques. Analytical determinations of concentrations tend to be empirical. Such methodological problems have recently driven a profusion of technical papers.

This article reviews turbidity theory, measurement principles, instrumentation systems, and applications, with particular reference to suspended sediment concentrations (SSCs) in natural waters (e.g., rivers, estuaries, and nearshore zones).

## Definitions and Applications

### Measurement Definitions

Turbidity can be measured using the techniques of turbidimetry or nephelometry (from *nephelo* = cloud (Greek)). Turbidimetry is the measurement of turbidity by quantifying the degree of 'attenuation' of a beam of light of known initial intensity. It is usually applied to media of fairly high turbidity in which the scattering particles are relatively large (e.g., natural waters), for reasons, which will be addressed below.

Nephelometry is the measurement of turbidity by the direct evaluation of the degree of light 'scattering' taking place in the medium. It is much more appropriate to media of lower turbidity in which the suspended particles are small. Turbidimetry and nephelometry can offer considerable time-saving advantages over gravimetric methods for the determination of particle concentrations, and are non-destructive techniques.

### Typical Applications

Turbidimetry and nephelometry have found many applications in scientific laboratories and in the chemical, pharmaceutical, foodstuffs, and beverage industries. In addition, turbidimetry and nephelometry are well-established procedures wherever filtration processes have to be effected, monitored, and controlled. Within the hydrological sciences, and water supply and wastewater management industries, turbidity values can act as simple and convenient surrogate measures of the concentration of suspended solids, sulfate ions (which are precipitated as  $\text{BaSO}_4$  in acidic media (HCl) with barium chloride), and other particulate material, and remain one of the most common applications of turbidimetry. Also, atmospheric and space physicists effect nephelometric analyses because of the importance of dust particles to radiation and other processes. In quantitative chemical and biological analysis, applications are common, especially the calculation of absolute molecular weights and dimensions of polymers in solution, as well as particle size determinations of suspended matter. Chemical profiles can also be obtained by observing turbidity changes deliberately induced by the addition of specific substances to the solution. Within microbiology, cell and bacteria growth can be monitored through the media turbidity changes such activity causes. In foodstuff manufacturing, turbidimetry is often used to monitor product quality and treatment process efficiency, especially in the dairy and brewing industries. Clarity (and hence turbidity) is also a key concern in the petrochemical industries. Determining turbidity

components – turbidity apportionment – has advanced with recent technological developments.

## Principles and Theory

Light passing through a liquid medium may be scattered and absorbed by inhomogeneities in the light path, especially suspended particles of silt, clay, algae and other plankton, microbes, organic matter, and other fine insoluble particulate substances. Bubbles and density discontinuities can also scatter light. Scattering occurs when 'a minute particle interacts with incident light by absorbing the light energy and then, as if a point light source itself, reradiating the light energy in all directions'. Absorption takes place when light is converted to other energy forms (e.g., heat) within the particle. Scattered light includes that reflected from the surface of the particle and that refracted within the particle, possibly after many internal reflections. Scattering is often accompanied by absorption.

The direct relationship, however, between turbidity data and suspended solids concentrations is weakened by the complex interactions of light energy with suspended particles. This interplay is heavily dependent on many factors, including:

- concentration of scattering particles suspended in the medium;
- size distribution of the scattering particles;
- shape, orientation, and surface condition of the scattering particles;
- refractive index of the scattering particles;
- refractive index of the suspension medium;
- wavelength of the light source employed.

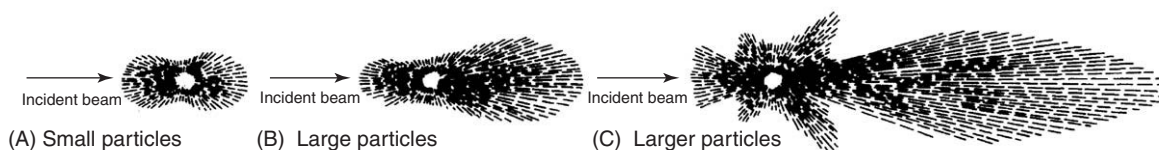
Consequently, separate bodies of theory have been developed to describe the many different processes that result. At its simplest level, light intensity is reduced during transmission through a collection of scattering particles in a sample according to an attenuation function of the form:

$$I = I_0 e^{-\tau l} \quad [1]$$

where  $I_0$  is the initial beam intensity,  $I$  the beam intensity after passing through a medium of length  $l$ , and  $\tau$  the turbidity coefficient of the medium. Equation [1] ignores losses of light through true absorption by suspended particles or reflection from the sides of the sample container.

### Light-Scattering Theory and the Influence of Particle Size

Appropriate light-scattering theory is governed by the diameter,  $D$ , of the scattering elements in relation



**Figure 1** Influence of particle size on the angular distribution of scattered light: (A) small particles ( $D < 0.1\lambda$ ); (B) large particles ( $D \sim 0.25\lambda$ ); and (C) larger particles ( $D > 1\lambda$ ). (From Vanous RD, Larson PE, and Hach CC (1982) *The theory and measurement of turbidity and residue*. In: Minear RA and Keith LH (eds.) *Water Analysis*, vol. 1, pp. 163–234. New York: Academic Press.)

to the wavelength,  $\lambda$ , of the light emitted by the measuring instrument. Indeed, theory is often specified in terms of the Mie size parameter ( $\alpha = 2\pi R/\lambda$ ), where  $R$  is particle radius. Particle size thus forms an appropriate basis for the subdivision of the theoretical discussion that follows.

### Small Particles

For particles where  $D < 0.05\lambda$ , Rayleigh scattering theory of 1871, originally developed for gases, is applicable to liquids with low concentrations of suspended particles which do not interact with each other. For such small particles, relatively symmetrical light-scattering distributions are obtained (Figure 1A). If a visible light source (i.e.,  $\lambda = 0.4\text{--}0.7\ \mu\text{m}$ ) is employed, then it follows that this theory is applicable for particles where  $D < \sim 0.03\ \mu\text{m}$ .

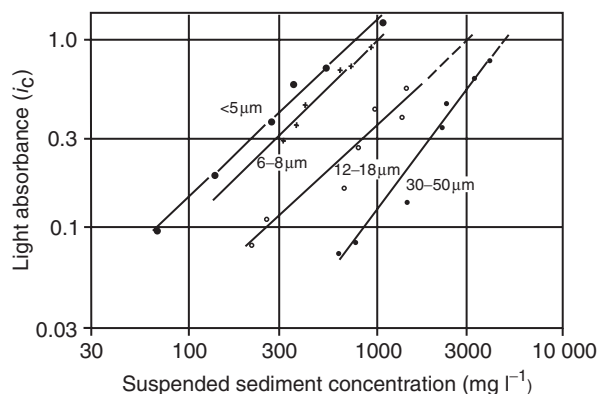
The Rayleigh equation describing the angular distribution of resultant scattering is:

$$i_\theta/I_0 = \{(n'/n) - 1\}^2 (NV^2/\lambda^4 r^2)(1 + \cos^2 \theta) \quad [2]$$

where  $i_\theta$  is the intensity of light scattered at angle  $\theta$ ,  $I_0$  is the initial light source intensity,  $n'$  is the refractive index of the particles,  $n$  is the refractive index of the suspension medium,  $r$  is the distance from the particles to the point of measurement, in terms of the number,  $N$ , of particles, each of volume  $V$ . Rayleigh's work thus shows that the intensity of scattered light varies: (1) with the square of the particle volume – and thus with the sixth power of the particle radius, assuming spherical shapes; and (2) inversely with the fourth power of the light wavelength used. Rayleigh theory has since been developed to allow relative molecular masses and sizes to be determined.

### Large Particles

For larger particles, however, where  $0.1\lambda < D < 0.8\lambda$ , the angular distribution of scattered light becomes asymmetrical. Destructive interference of light scattered in the backward direction leads to a bias in forward-scattered light (Figure 1B). In these contexts, Mie scattering theory of 1908 for larger spheres becomes more appropriate. For such larger



**Figure 2** Influence of sediment particle diameter on light absorbance by samples of different concentrations. Note how a given concentration effects much greater absorbance at the smaller particle diameters. (Reproduced with permission from Ward PRB and Chikwanha R (1980) *Laboratory measurement of sediment turbidity*. *Proceedings of the American Society of Civil Engineers, Journal of the Hydraulics Division* 106: 1041–1053.)

particles, scattering intensity is less dependent on wavelength.

### Very Large Particles

For larger particles still, where  $D > 0.8\lambda$ , the Mie equations are still workable, although for particles larger than  $\sim 0.4\ \mu\text{m}$  in diameter wide oscillations in scattering patterns emerge. For particles of  $D > 1\ \mu\text{m}$ , extreme concentration of scattering in the forward direction emerges (because of mutually destructive backscattering effects), along with secondary peaks in the angular distribution of scattered light (Figure 1C).

Theory and practice also demonstrate that the most efficient scattering elements are those of a diameter similar to the light wavelength used. Also, a given mass of small particles causes much greater light attenuation than the same mass of large particles (Figure 2).

One complexity is that much classical theory has been developed for identically sized spherical particles – conditions that may not be obtained in all laboratory or field situations. Indeed, many natural

waters, like the atmosphere, contain an ensemble of variably sized, irregularly-shaped, and randomly oriented particles, for which theory is still being developed. Furthermore, processes become highly complex when concentrations are so great that multiple scattering occurs (i.e., particles receive light previously scattered from other particles: this normally increases opportunities for light absorption).

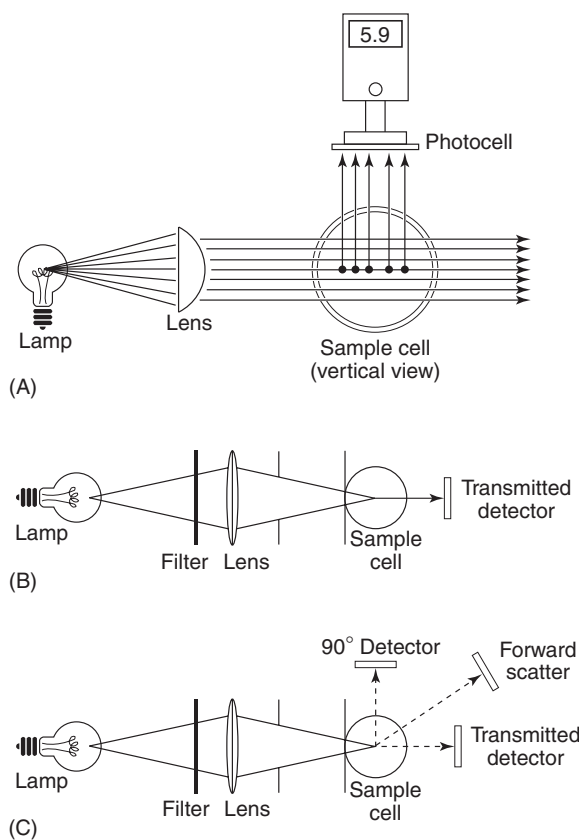
## Instrumentation

### Range of Turbidimetric and Nephelometric Systems

Early procedures were based on manual operation of analytical systems and visual turbidity assessment (e.g., the Secchi disk). Several instruments are now available, however, for quantitative turbidity determination in a variety of scientific, industrial, and process management applications. Choice will depend largely on the analytical aims (e.g., mass concentration, particle size distribution, molecular dimensions, or crystal/cell growth), the nature of the scattering elements and suspension medium, and whether field or laboratory measurement is needed. Turbidimeters also vary in optical geometry, mode of operation, sample handling capabilities, data recording options (e.g., automatic/manual or analogue/digital) and portability. Ultraviolet-visible spectrophotometry can also be used for turbidimetric measurements by measuring the absorption of light by particles at a fixed wavelength or full spectrum (e.g., for kinetic studies of the time decay of species). The discussion that follows focuses on the use of turbidimetric instrumentation to estimate the mass concentration of suspended matter in liquid samples, with particular reference to sediment in natural waters.

### Basic Elements of Measurement Systems

Modern measurement devices use photosensitive cells to quantify scattered and/or transmitted light. Figure 3 illustrates that most laboratory bench instruments usually have five basic components: a light source of known, constant intensity, and given wavelength characteristics; a lens to collimate the light beam; a sample cell; photosensor(s); and a meter or logger to record the output signals from the photosensor(s). Versions for continuous monitoring of turbidity values (e.g., for online industrial systems or process measurement in environmental sciences) include some kind of flow-through measurement chamber (instead of a 'static' sample cell) and outputs for a datalogger. The two basic measuring instruments are the nephelometer and the turbidimeter.



**Figure 3** Three basic designs of turbidity meter: (A) the nephelometer, which directly measures light scattered (usually at 90° to the beam direction) by suspended particles; (B) the turbidimeter, where the transmitted light is detected, in relation to initial beam intensity; (C) the ratio turbidimeter in which both transmitted and scattered light is detected. (Reproduced with permission from Hach CC, Vanous RD, and Heer JM (1982) Understanding turbidity measurement. *Technical Information Series*, Booklet No. 11, 1st edn., 11pp. Hach Chemical Co.)

A 'nephelometer' measures directly the intensity of light scattered by the sample, which is proportional to the amount of matter suspended in the light path, though the influence of size, shape, and refractive index of the scattering particles is also important. With nephelometers, the sensor is mounted at an angle to the traversing beam (often 90°) to record scattered light in one part of the angular distribution (Figure 3A). Some more sophisticated versions can monitor scattering intensity at many different angles: this allows angular summation values to be checked against initial and attenuated signals. Nephelometers usually provide better precision and sensitivity than turbidimeters and are normally used for samples of low turbidity containing small particles.

A 'turbidimeter', sometimes called a transmissometer, absorptiometer, or turbidity meter (the latter term is commonly used for field instruments in the earth and environmental sciences), measures the



intensity of the beam after it has passed through the sample, i.e., it quantifies the amount of transmitted light remaining (**Figure 3B**). Suspended matter in the light path causes scattering and absorption of some light energy, which reduces the incident illumination falling on the photocell. These instruments are more appropriate for relatively turbid samples in which the scattering particles are large in relation to the light wavelength used. This is because a significant reduction in the intensity of incident light is needed to yield precise results.

Some newer instruments, called ratio turbidimeters, incorporate measurement systems for light which is side-scattered (usually at  $90^\circ$ ), forward-scattered, and transmitted (**Figure 3C**). The turbidity value is obtained as the ratio of the  $90^\circ$  signal to the sum of forward-scattered and transmitted values. The ratio feature has a number of advantages: it increases the long-term stability of the sensor (by reducing effects of instrumental drift); it compensates for ageing of, and deposits on, the optics; it reduces the influence of temperature changes in the electronics; it minimizes the need for repeated recalibration; and it limits the effect of sample color on readings. This can be more appropriate for strongly and/or variably colored liquids, or for samples of high turbidity. A four-beam instrument version has emerged recently, which reduces error still further.

Recent developments include laser-based turbidimeters, reflectometers, or fiber-optic systems. The development of the optical backscatter sensor (OBS) has become popular for field deployment in the hydrological and oceanographic sciences: this instrument monitors water turbidity through the backscattering of pulsed infrared light emitted from the OBS instrument head. Also, remotely sensed turbidity measurement, using satellite or air borne instruments (e.g., the CASI (Compact Airborne Spectrographic Imager) system deployed by the UK Natural Environment Research Council), has recently eased the mapping of turbidity patterns over large spatial scales.

There is a strong dependency of scattering efficiency on light wavelength (see above). Consequently, for a given detector, light sources of short wavelength are more sensitive to, and therefore more useful for, the detection of small particles. Conversely, longer wavelengths are more appropriate for samples containing large particles (e.g., sediment in many earth or environmental science systems). The source-detector relationship can vary widely between instruments, and is cited as the key reason explaining the different readings obtained on the same sample by different devices.

## Units of Measurement and Instrument Calibration

The field is hampered by a nonstandard, ill-defined, and historically changing unit of measurement. The Nephelometric Turbidity Unit (NTU) is the most common unit employed. The precision with which turbidity data should be reported depends on how turbid the sample is, but should be to the nearest 1–10%, approximately, of the NTU value determined. For example, NTU values for distilled water, tap water, and raw water are 0.08, 0.54, and 3.52, respectively, but much higher values, well above 150 NTU, are common in many hydrological systems.

Formazin polymer, developed in 1926, can be used for turbidimeter calibration, and is straightforward to prepare, control, and reproduce. Standard procedures for the production of a stock formazin turbidity suspension of 400 NTU are given in American Public Health Association. Other calibration materials can be used (e.g., Fullers Earth or Hach Gelex 'fixed' standards – metal oxide particles permanently and statically suspended in silica gel) and may provide suitable alternatives, especially given the health concerns voiced in some quarters over formazin use.

## Field Calibration

In natural waters, suspended material may largely consist of particles in the size range of clay ( $D < 2 \mu\text{m}$ ), silt ( $2 < D < 63 \mu\text{m}$ ), or even sand ( $63 < D < 2000 \mu\text{m}$ ). It may also include organic matter and compounds and microscopic organisms. For field applications in hydrology or oceanography, analysts should preferably calibrate turbidity readings against known mass concentrations of the suspended sediment typical of that context, and declare the strength of the diagnostic statistics for derived relationships. Such correlations can be weak, reflecting temporal changes in suspended load composition (and hence its light-scattering efficiency), water color, or bubble presence. Predictive relationships can be strengthened by accounting for such changes (especially in sediment load particle size distribution), which can occur over various timescales (e.g., interannual, seasonal, subseasonal, flood event). It may even be necessary to produce multivariate or separate calibration equations to incorporate the effects of, for example, changing flow levels, sediment source areas, and season on sediment load constitution. For such field applications, the turbidimeter reading (often in arbitrary units) is converted to estimated SSC using a site-specific calibration curve.

### Turbidity Meters for Continuous Field Operation

For unattended automated field use a range of suitable turbidity meters with additional instrumental features are available. The measurement cell is replaced by a flow-through chamber, which must prevent stray light from reaching the photosensor. Instruments can be boom-mounted directly in the flow, or at the end of pump-lines connecting sampling point to measurement system.

Turbidity meters with narrow-band near infrared (NIR) light sources (peak output at  $0.86\ \mu\text{m}$ ; spectral bandwidth  $<0.06\ \mu\text{m}$ ) are recommended by ISO. Such instruments reduce problems of algal build-up on the optical surfaces, are less affected by color, and are more sensitive to the slightly larger particles typical of sediment transport systems. However, some relaxation of the infrared protocol is tolerable for field instruments operating in continuous monitoring mode.

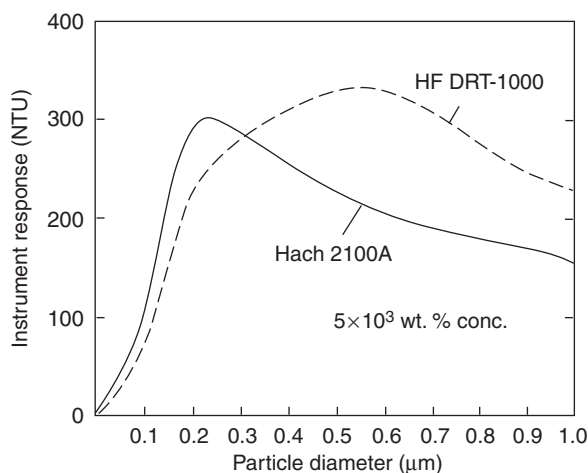
To limit further the impact of problematic algal growth on the optics, field turbidimetric systems can be equipped with a pulsed light source, antifouling chemicals or films, and/or 'wiper' blades for convenient (sometimes automated) cleaning of optical surfaces. Alternatively, a dual-beam (twin-gap) instrument to compensate for these effects can be deployed.

Power requirements for field instruments are important considerations. The use of low-consumption light-emitting diodes (LEDs) in sensors, and photovoltaic sensors which convert incident light directly into electrical energy, reduces power needs to a minimum. Solar panels are useful to trickle-charge instruments and dataloggers.

Some instruments are temperature sensitive, largely because LEDs can emit more strongly when warm, and photovoltaic detectors convert photons into electrons more efficiently at low temperatures. Given, for example, the annual range of river temperature in the UK is typically  $\sim 20\text{--}25\ \text{K}$ , some correction procedures may be necessary.

Figure 4 also demonstrates that many turbidity meters are relatively insensitive to very fine or very coarse particulate matter. Because most instruments work in the visible or NIR spectrum this means that the most readily detectable particles are those where,  $\sim 0.2 < D < 1.8\ \mu\text{m}$ . In standard sedimentological and engineering classifications, these are clay-sized particles, and hence the occasionally used term 'silt-meter' for turbidimetric instruments is not entirely appropriate.

Turbidity meters need to be interfaced with portable multimeters, dataloggers, or computers for recording and storage of turbidity data. Dataloggers



**Figure 4** Sensitivity of two turbidity meters to particle diameter (HF = H.F. instruments). Test material here is spherical latex particles of very narrow size distribution. (From Vanous RD, Larson CC (1982) The theory and measurement of turbidity and residue. In: Minear RA and Keith LH (eds.) *Water Analysis*, vol. 1, pp. 163–234. New York: Academic Press.)

equipped with an appropriate input channel and needing minimal power for operation are heavily used today. They provide quasi-continuous digital data on temporal variations in turbidity, which are ideal for computational analysis and the study of turbidity dynamics. Telemetry systems for real-time data acquisition, alarm facilities, and remote downloading capabilities are becoming increasingly common.

Many field scientists, however, despite the range of commercially available turbidity instrumentation, still recognize a need for low cost, rugged, and reliable systems. This is especially so when a network of instruments is required for permanent installation to define a turbidity field, for simultaneous manual turbidity measurements by a research team, or for specialist applications such as in subglacial environments. This has led many researchers to custom-build their own instruments.

### Hydrological Applications of Turbidimetry

The monitoring of turbidity and SSCs in rivers, estuaries, lakes, reservoirs, nearshore zones, etc. is attracting increasing attention from hydrologists, limnologists, geomorphologists, freshwater ecologists, engineers, oceanographers, glaciologists, water resource managers, and policy makers. Such measurement programs can allow inferences to be made about upstream hydrogeomorphological processes,

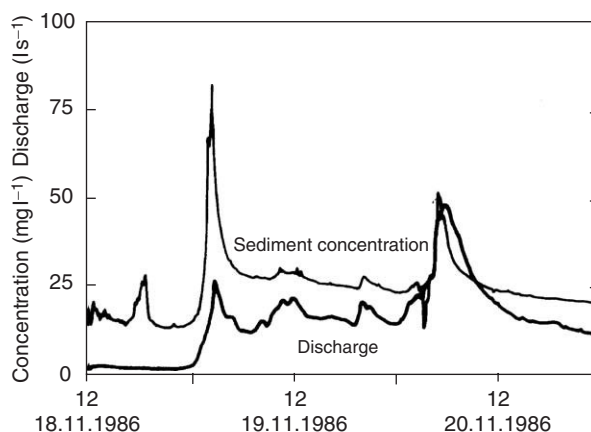
catchment erosion rates, downstream fluvial processes and sedimentation impacts, pollutant and contaminant transfer, and aquatic habitat quality. The recreational value of water bodies can be partly linked to their clarity, as demonstrated in the Lake Tahoe turbidity reports of 2001–02. Increasingly, there is a legal requirement for environmental impact assessments and water supply managers to consider the possibility of short- or long-term turbidity increases resulting from proposed development schemes.

Automated, in-stream, high-frequency turbidity monitoring has become increasingly popular, mainly because the alternative practice of sampling and subsequent laboratory processing of samples is laborious and resource-intensive. Sampling approaches thus inevitably constrain the level of temporal and spatial detail possible, making it difficult to reveal the patterns, dynamics, and processes present. Recent advances in the understanding of the hydrodynamics of fine-sediment transport in river, tidal, and nearshore environments, for example, would have been impossible without very high frequency (e.g., 5 Hz) monitoring of transient turbidity changes. In remote environments, and in the developing world, a likely paucity of suitable sample analysis facilities underlines the need for an automated and direct field-based method.

Three example applications of continuous monitoring are outlined below: river turbidity variation during rainstorms; very short-lived turbidity pulsing in glacial and coastal waters; and the definition of the estuarine turbidity maximum. These illustrate the many advantages of turbidity instrumentation over sampling-based approaches in quantifying and understanding complex temporal and spatial patterns of suspended sediment fluxes. They also demonstrate the substantial variations of turbidity and SSCs in natural systems in space and time.

### River Turbidity Variation through Individual Storm Events

SSC in streams can change appreciably over seasonal timescales and during high-flow events. Automated river turbidity monitoring is very useful in refining calculations of suspended sediment loads, because it detects the short-lived, but very important order-of-magnitude changes in SSC that occur in many rivers during complex storm events (e.g., **Figure 5**). The 1-min datalogging scan interval used in **Figure 5** defined all peaks and troughs in turbidity for most events. Following calibration, this permitted the full definition of exhaustion phenomena in SSCs, and



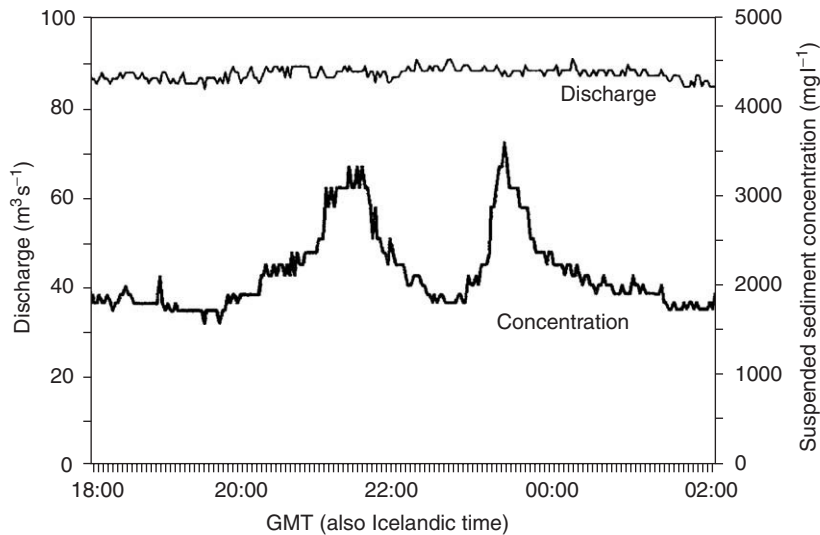
**Figure 5** Continuous record of stream suspended sediment concentration, in relation to river discharge changes during storm events, derived from calibration of a Partech turbidity record. (From Gippel CJ (1989) The use of turbidity instruments to measure stream water suspended sediment concentration, Monograph Series No. 4, Department of Geography and Oceanography, University College, University of New South Wales, Australian Defence Force Academy, Canberra, 204pp.)

hysteresis effects in relation to the discharge series (hitherto undetected for the system by earlier workers using a conventional sampling program). This led, in turn, to more securely based explanations of fine-sediment delivery processes.

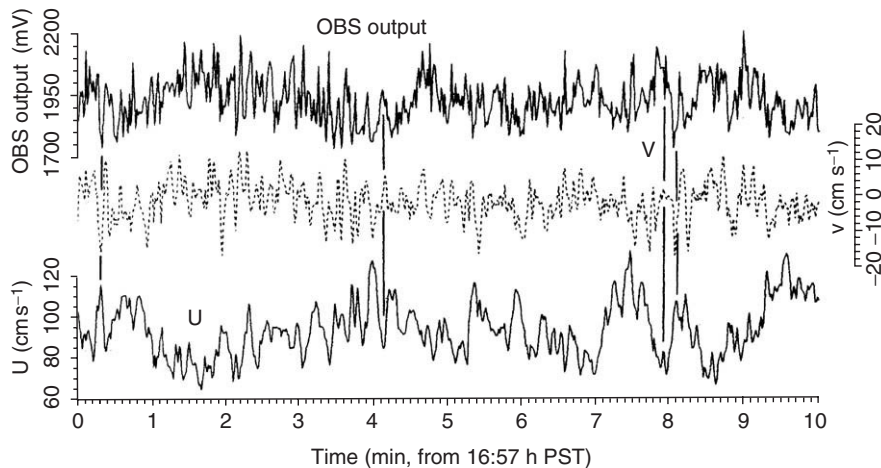
### Turbidity Pulsing in Rivers and Nearshore Zones

Turbidimetric instrumentation also facilitates the detection of very short-lived pulsing of suspended sediment, which characterizes many systems, especially proglacial meltwater environments. The example in **Figure 6**, for the Jökulsá á Sólheimasandi glacial river in southern Iceland, shows two substantial sediment-pulsing events detected by a 2-min turbidity scanning program. These were unrelated to river flow variations, and showed that other significant sediment mobilization processes were present in the system. The alternative approach of flow-triggered automated sampling is unsatisfactory in these situations where many sediment-flux perturbations are unrelated to water discharge.

One strength of automated turbidity monitoring is that the logging system can also be used to record, on the same time base, data on associated or explanatory variables. For glacial meltwater studies (**Figure 6**), these typically include energy budget components relevant to glacial ablation, rainfall intensity, river discharge, temperature, and electrical conductivity, and, with Photo-Electronic Erosion Pin (PEEP) sensors, even the erosion events themselves which generate sediment plumes. Such data can strengthen the



**Figure 6** Two-minute scanning of turbidity and river discharge (8–9 Aug 1988) showing a compound pulsing of suspended sediment concentration, unrelated to flow variations, in the Jökulsá á Sólheimasandi glacial river in southern Iceland. (Reproduced with permission from Lawler DM and Brown RM (1992) A simple and inexpensive turbidity meter for the estimation of suspended sediment concentrations. *Hydrological Processes* 6: 159–168; © John Wiley and Sons Ltd.)

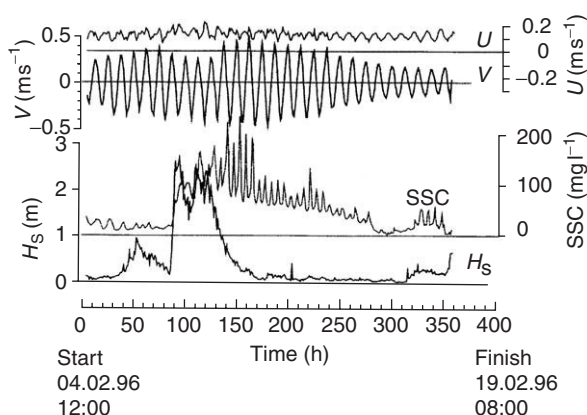


**Figure 7** A high-frequency, 5 Hz, 10-min time series of OBS turbidity alongside river flow components ( $u$ , streamwise;  $v$ , normal to the bed) for the Fraser River, near Mission, BC, Canada. (Reproduced with permission from Lapointe M (1992) Burst-like sediments suspension events in the sand bed river. *Earth Surface Processes and Landforms* 17: 253–270; © John Wiley and Sons Ltd.)

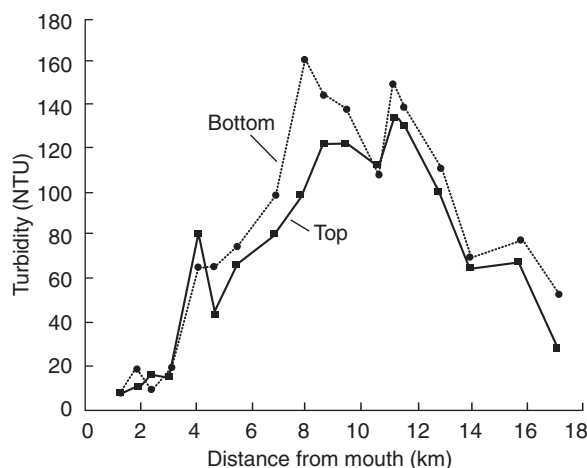
process-inference capabilities of the whole exercise. Thus, the correlation of high-frequency velocity and OBS turbidity series (Figure 7) allowed a clearer understanding of river sediment transport events to be gained. In coastal zones, very high frequency (5 Hz) monitoring of OBS turbidity, wave height, and currents (Figure 8), facilitated the definition of the critical flow velocities required to mobilize bed sediment. Knowledge of these threshold conditions is important for the stability, engineering, and protection of coastlines and their ecosystems.

### Estuarine Turbidity Maxima

The ‘estuarine turbidity maximum’ is the term given to the clear peak in mean SSC observable in many estuaries around the limit of saline intrusion. Improved explanations of turbidity fields in estuarine systems, including the nature, location, and migration of the turbidity maximum, and the tidal pumping processes responsible, have recently been obtained by supplementing water sampling approaches with detailed automatic turbidity monitoring. Furthermore,



**Figure 8** A high-frequency, 5 Hz record averaged to hourly time series of OBS turbidity (SSC), in relation to significant wave height,  $H_s$ ; cross-shore current,  $U$ ; and longshore current,  $V$ , for the North Sea nearshore zone at Holderness, UK. Data from Feb. 1996; water depth 16.8 m. The OBS here has been deployed within BLISS (Boundary Layer Intelligent Sensor System). (From Blewett J and Huntley D (1999) Measurement of suspended sediment transport processes in shallow water off the Holderness coast, UK. *Marine Pollution Bulletin* 37(3–7): 134–143.)



**Figure 9** Relationship between estuarine turbidity in NTU and distance from the mouth of St. Lucia Estuary, Natal, on a rising tide on 20 March 1981, showing a clear turbidity maximum ~8–12 km from the estuary mouth for both near-bed (bottom) and surface (top) waters. (From Cyrus DP (1988) Turbidity and other physical factors in Natal estuarine systems. Part 1: selected estuaries. *Journal of the Limnological Society of southern Africa* 14(2): 60–71.)

for any spatial survey, the fact that turbidity values are obtained at the field sites themselves, rather than in the laboratory subsequently, can allow instant decisions to be made regarding any further environmental sampling (including turbidity) that may be desirable. **Figure 9** illustrates the value of repeated estuarine turbidity measurement in revealing the spatial and temporal structure of the turbidity maximum.

**See also:** Color Measurement. **Environmental Analysis.** Geochemistry: Sediment. **Particle Size Analysis.** Sensors: Photometric. **Water Analysis:** Particle Characterization.

## Further Reading

- American Public Health Association (1998) *Turbidity. Standard Methods for the Examination of Water and Wastewater*. American Public Health Association, American Water Works Association, and Water Pollution Control Federation.
- Davies-Colley RJ and Smith DG (2001) Turbidity, suspended sediment, and water clarity: a review. *Journal of American Water Resources Association* 37(5): 1085–1101.
- ISO (International Organisation for Standardization) (1984). *Water Quality – Determination of Turbidity*, 1st edn., 5pp. ISO 7027 – 1984(E).
- Mitchell SB, Lawler DM, West JR, and Couperthwaite JS (2003) Use of continuous turbidity sensor in the prediction of fine sediment transport in the turbidity maximum of the Trent Estuary, UK. *Estuarine Coastal and Shelf Science* 58: 643–650.
- Schuerman DW (ed.) (1980) *Light Scattering by Irregularly Shaped Particles*, 334pp. New York: Plenum Press.
- Van de Hulst HC (1981) *Light Scattering by Small Particles*, 2nd edn., 470pp. New York: Dover Publications.
- Walling DE (1977) Limitations of the rating curve technique for estimating suspended sediment loads, with particular reference to British rivers. *Erosion and Solid Matter Transport in Inland Waters*. International Association of Hydrological Sciences Publ. No. 122, pp. 34–48.

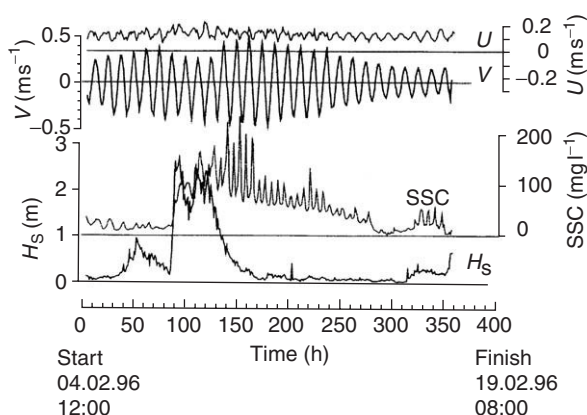
## Inorganic Compounds

**M A Zezzi-Arruda and R J Poppi**, University of Campinas, Campinas, Brazil

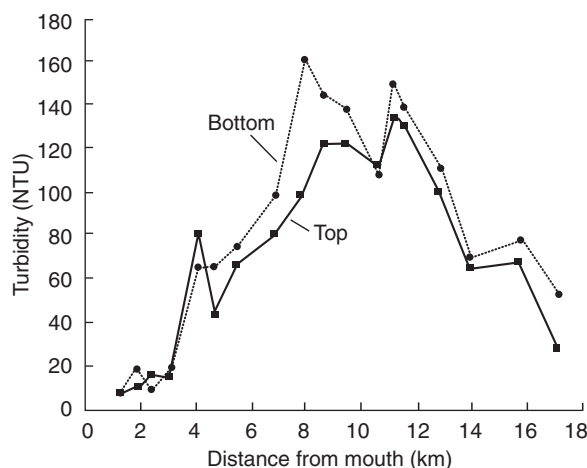
## Introduction

Spectrophotometry is an excellent alternative for the determination of inorganic compounds. It is characterized by a wide analytical working range,





**Figure 8** A high-frequency, 5 Hz record averaged to hourly time series of OBS turbidity (SSC), in relation to significant wave height,  $H_s$ ; cross-shore current,  $U$ ; and longshore current,  $V$ , for the North Sea nearshore zone at Holderness, UK. Data from Feb. 1996; water depth 16.8 m. The OBS here has been deployed within BLISS (Boundary Layer Intelligent Sensor System). (From Blewett J and Huntley D (1999) Measurement of suspended sediment transport processes in shallow water off the Holderness coast, UK. *Marine Pollution Bulletin* 37(3–7): 134–143.)



**Figure 9** Relationship between estuarine turbidity in NTU and distance from the mouth of St. Lucia Estuary, Natal, on a rising tide on 20 March 1981, showing a clear turbidity maximum ~8–12 km from the estuary mouth for both near-bed (bottom) and surface (top) waters. (From Cyrus DP (1988) Turbidity and other physical factors in Natal estuarine systems. Part 1: selected estuaries. *Journal of the Limnological Society of southern Africa* 14(2): 60–71.)

for any spatial survey, the fact that turbidity values are obtained at the field sites themselves, rather than in the laboratory subsequently, can allow instant decisions to be made regarding any further environmental sampling (including turbidity) that may be desirable. **Figure 9** illustrates the value of repeated estuarine turbidity measurement in revealing the spatial and temporal structure of the turbidity maximum.

**See also:** Color Measurement. **Environmental Analysis.** Geochemistry: Sediment. **Particle Size Analysis.** Sensors: Photometric. **Water Analysis:** Particle Characterization.

## Further Reading

- American Public Health Association (1998) *Turbidity. Standard Methods for the Examination of Water and Wastewater*. American Public Health Association, American Water Works Association, and Water Pollution Control Federation.
- Davies-Colley RJ and Smith DG (2001) Turbidity, suspended sediment, and water clarity: a review. *Journal of American Water Resources Association* 37(5): 1085–1101.
- ISO (International Organisation for Standardization) (1984). *Water Quality – Determination of Turbidity*, 1st edn., 5pp. ISO 7027 – 1984(E).
- Mitchell SB, Lawler DM, West JR, and Couperthwaite JS (2003) Use of continuous turbidity sensor in the prediction of fine sediment transport in the turbidity maximum of the Trent Estuary, UK. *Estuarine Coastal and Shelf Science* 58: 643–650.
- Schuerman DW (ed.) (1980) *Light Scattering by Irregularly Shaped Particles*, 334pp. New York: Plenum Press.
- Van de Hulst HC (1981) *Light Scattering by Small Particles*, 2nd edn., 470pp. New York: Dover Publications.
- Walling DE (1977) Limitations of the rating curve technique for estimating suspended sediment loads, with particular reference to British rivers. *Erosion and Solid Matter Transport in Inland Waters*. International Association of Hydrological Sciences Publ. No. 122, pp. 34–48.

## Inorganic Compounds

**M A Zezzi-Arruda and R J Poppi**, University of Campinas, Campinas, Brazil

## Introduction

Spectrophotometry is an excellent alternative for the determination of inorganic compounds. It is characterized by a wide analytical working range,



and therefore sample dilution is avoided. Versatility, low cost, easy implementation, etc. are in general inherent to this technique. Plug-and-play equipment is often used. These characteristics make spectrophotometry one of the most versatile of analytical techniques.

Further, the nature of the reagents used in spectrophotometric applications, used in catalytic procedures or to form complexes, allows chemical reactions involving mixed ligand and ion-association systems, organized media, etc. to be carried out to determine different analytes, thus increasing the applicability of spectrophotometry.

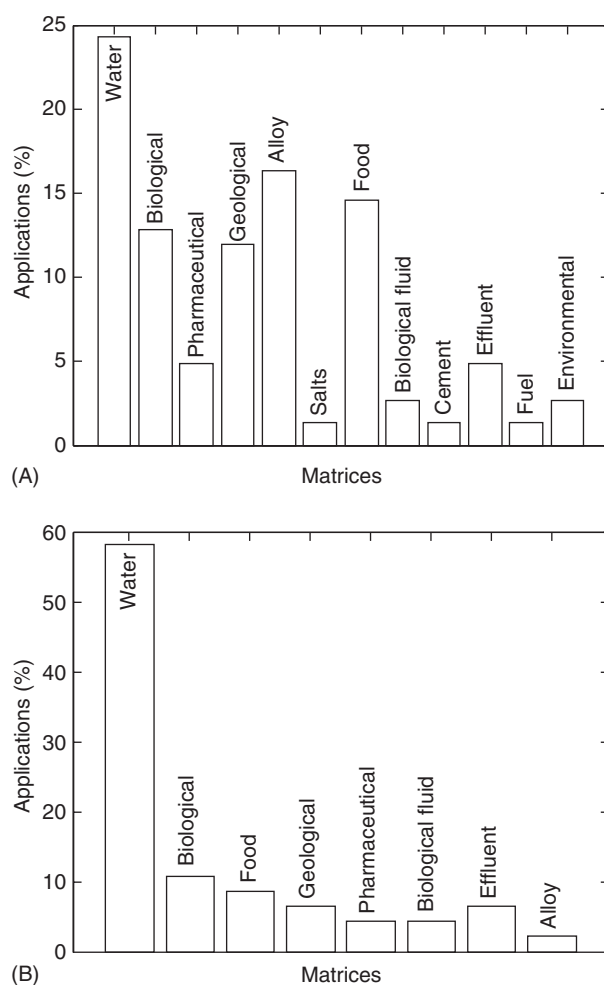
The majority of spectrophotometric applications are related to inorganic compounds. Hundreds of combinations can therefore be found, thus making it possible to determine one analyte, for example, using different spectrophotometric reagents.

In view of the wide spectrum of reagents, analytes, applications, systems, and their combinations involved, this article was written to inform readers about the reagents used in spectrophotometric analysis, with emphasis on inorganic species (cationic and anionic) because of their importance in analysis and to provide examples found in the literature, mainly from 1998 onwards. **Figure 1** provides examples of applications of spectrophotometry in different areas, such as environmental, biological, and pharmaceutical applications. From this figure, it can be concluded that the majority of applications relate to water, alloy, and food samples in which cationic species are determined. For anionic species, once again, water was the main matrix analyzed.

## Cationic Species

### Aluminum

Reagents such as quinizarin, alizarine violet N, semi-xylol orange, Eriochrome cyanine R, alizarin yellow, and chrome azurol S are used for aluminum determination. These reagents tend to form different complexes with Al, sometimes leading to good sensitivity of the method (nanogram per milliliter detection limits using alizarine violet N) or increasing its selectivity, thus allowing aluminum determination in complex matrices such as cement (quinizarin), plant digest (chrome azurol S), milk (alizarin), and botanical samples (Eriochrome cyanine R). However, when the selectivity is still not enough to guarantee accurate results, different strategies such as the use of extraction techniques, exchange resins, or surfactants are used.



**Figure 1** Applications (%) of (A) spectrophotometric determinations of cationic species and related matrices and (B) spectrophotometric determinations of anionic species and related matrices.

### Boron

Classical reagents such as azomethine-H and curcumin are frequently used in boron determination. Azomethine-H is used, for example, in boron determination in plants, geochemical materials, and soils. Curcumin is used in determination of boron as an impurity in silicon-doped gallium arsenide used in electronics and in high-purity molybdenum. The microgram per gram limit of detection (LOD) is attained when curcumin or 4-methoxy-azomethine-H (420 nm) is used. However, boron is also determined by the formation of the macrobicyclic complex Fe(Nx) (3) (BOH) (2) in the nioxime-boric acid-Fe(II) system (447 nm).

Boron has also been determined using the surfactant system formed with curcumin and cetylpyridinium bromide at 560 nm. A good Beer's law range has been attained ( $\leq 3 \mu\text{g}/25 \text{ ml}$ ).

Sometimes a preliminary extraction of boron with organic reagent is necessary for improving the selectivity as well as the use of a specific resin.

### Cadmium

Cadmium is commonly determined using 4-(2-pyridylazo)resorcinol; however, other ligands have been proposed, such as butylrhodamin B (607 nm), *p*-acetylbenzenediazoaminoazobenzene (475 nm), 2-pyridinediazoaminoazobenzene (530 nm), 5,10,15,20-tetrakis(4-carboxylphenyl)porphyrin (423 nm), and *meso*-tetra(4-trimethylammonium phenyl)porphyrin (434 nm).

A new azoamino reagent has been proposed recently for cadmium determination in wastewater, which at a pH value of 10.4 ( $\text{Na}_2\text{B}_4\text{O}_7/\text{NaOH}$  buffer) produces an orange-red chelate with cadmium (1:2, M:L) at 520 nm.

Some surfactants such as Triton X-100, OP emulsifier, and polyvinyl alcohol are commonly used to make some reagents soluble so that solutions can be prepared.

Cadmium is sometimes determined together with other metals such as zinc, lead, and mercury. Generally, these applications are based on the formation of metal-ligand complexes, and some kinds of statistical or chemometric tools (full factorial designs, sequential experimental Doehlert designs, artificial neural network, partial least square) are used for distinguishing the concentrations of each analyte. In addition, some applications are based on the kinetic differentiation of the reactions related to metals and the same ligands.

Problems with possible interferences are avoided by using either different masking agents such as picolinic acid, potassium cyanide, and potassium sodium tartrate or some kind of separation procedure.

Different samples are analyzed, independent of the chromogenic reagent used, such as beverages, powdered corn, powder milk, natural water and wastewater, nutritional supplements, fertilizers, lobster, and soya bean.

### Chromium

For chromium determination, reagents other than the classical 1,5-diphenylcarbazide, such as xylene cyanol FF (615 nm), *p*-amino-*N,N*-diethylaniline (550 nm), benzooxazol (538 nm), benzothiazol (546 nm), dimethylindocarboxyanine (560 nm) and dimethylindodicarboxyanine (640 nm), have been used.

Sometimes Cr(VI) is determined after sample preparation ( $\text{HNO}_3/\text{H}_2\text{SO}_4 + \text{KMnO}_4$  as oxidizing agent) by reacting chromium with 1,5-diphenylcarbazide. Determination of Cr(III) and Cr(VI) is also

emphasized nowadays. The general procedure for chromium speciation is carried out by determining Cr(VI) and total chromium. So the Cr(III) concentration is calculated as the difference between the Cr(VI) and total chromium values. In general, the reagent 1,5-diphenylcarbazide is commonly used for this task. However, other reagents have also been proposed: Cr(VI) reacts almost completely in a short period with *p*-amino-*N,N*-diethylaniline. Thus, Cr(III) is masked by cyclohexyldiaminetetraacetic acid and only Cr(VI) is determined. On the other hand, the sum of Cr(VI) and Cr(III) is determined after oxidation of Cr(III). From the difference between these two results it is possible to establish the Cr(III) concentration.

Complex samples are involved in chromium determination, such as industrial wastewater, natural water, pharmaceuticals, and soil, irrespective of the strategy adopted.

### Cobalt

Different chromogenic reagents have been proposed for cobalt determination, such as 2-(5-bromo-2-pyridylazo)-5-(*N*-propyl-*N*-sulfoethylamino) aniline (550 nm), 1-(2-thiazolylazo)-2-naphthol (572 nm), *N,N'*-bis(2-aminobenzoyl) ethylenediamine (470 nm), isonitroso-5-methyl-2-hexanone (400 nm), 1-(2-pyridylazo)-2-naphthol (650 nm), 4,5-dihydroxy-1,3-benzenedisulfonic acid, disodium salt (426 nm), and pyridixal-4-phenyl-3-thiosemicarbazone (450 nm). In addition, a new reagent has been proposed for this task, 5-(6-methoxy-2-benzothiazoleazo)-8-aminoquinoline (MBTAQ), which presents an apparent molar absorptivity of  $1.15 \times 10^5 \text{ l mol}^{-1} \text{ cm}^{-1}$  at 655 nm. In general, the molar absorptivity of the reagents used for cobalt determination ranges from  $10^4$  to  $10^6 \text{ l mol}^{-1} \text{ cm}^{-1}$ . However, some strategies are adopted for improving the LOD of these methods, such as the use of mechanized monosegmented flow systems (LOD nanogram per liter range), micellar media (LOD microgram per liter range), solid-phase spectrophotometry (LOD microgram per liter range), and preconcentration (LOD microgram per liter range).

A wide range of samples have been analyzed using complex formation with chromogenic reagents, micellar media, and catalytic reactions, among others, such as pharmaceutical, biological, and environmental samples, alloys, and industrial effluents. In addition, there are some chemometric applications, mainly those involving multivariate calibration.

### Copper

Some copper determination applications involve the use of organic solvents or surfactants such as

ethanol, Tween 80, sodium dodecyl sulfate (SDS), and Triton X-100. These are used to make the reagent more soluble or to improve reaction conditions. One example is the use of Triton X-100 in the DDTC-Cu<sup>2+</sup> reaction for spectrophotometric determination of copper in fungicides and fertilizers and a fourfold expansion in dynamic range has been attained.

Classical reagents such as 1-nitroso-2-naphthol, 1-(2-pyridylazo)-2-naphthol, 2-carboxy-2'-hydroxy-5'-sulfopharmazilbenzene are used for copper determination. In addition, they are also used for comparison with the results obtained with other reagents. Besides these reagents 4,5-dimercapto-1,3-dithiol-2-tionate (430 nm), 5,8-dihydroxy-1,4-naphthoquinone (330 nm), mercaptosuccinic acid (345 nm), di-2-pyridyl ketone benzoylhydrazone (370 nm), 5-mercapto-3-phenyl-1,3,4-thiadiazole-2-thione (395 nm), and others have also been proposed for this task. In addition, new reagents such as iminodibenzyl and 3-chloroiminodibenzyl (660 nm) or 3,3'-(1,3-propanediyl-diimine)bis-[3-methyl-2-butanone]dioxime (525 nm) have also been proposed for copper determination.

Different samples may be analyzed, jet fuels, soil and airborne dust particles, pharmaceuticals, alloys, dry tea, water, and industrial effluents, among others.

## Iron

Besides the classical chromogenic reagents commonly used for iron determination such as *o*-phenanthroline and Eriochrome cyanine R, others reagents have also been proposed for this task such as 4,5-dihydroxy-1,3-benzenedisulfonic acid, disodium salt (667 nm, Fe(III) determination), 1-(*p*-tolyl)-3-phenylpropane-1,3-dione (I) (520 nm, Fe(II) determination), 2-mercapto-3-pyridinol (640 nm, Fe(III) determination), *N,N*-dimethyl-*p*-phenylenediammonium dichloride (554 nm, Fe(III) determination).

Nowadays, the trend appears to be to determine simultaneously iron and other metals or Fe(II) and Fe(III) species. In the first case, some reagents such as 1,5-bis(di-2-pyridylmethylene)thiocarbonohydrazide, 2,4,6-tri-(2-pyridyl)-1,3,5-triazine are used. In the second case 4,5-dihydroxy-1,3-benzenedisulfonic acid, disodium salt (635 nm), 1-(5-bromo-2-pyridylazo)-2-naphthol-6-sulfonic acid (765 and 850 nm), azide complexes in tetrahydrofuran (396 nm), etc. are used

In simultaneous determination, some kind of resin or chemometric method (principal component regression) is used for improving the selectivity. For speciation, generally the same wavelength is used for determining different iron species. Sometimes

Fe(III) is determined first, and after using an oxidizing reagent, the total iron (Fe(II) + Fe(III)) is determined. So, by calculating the difference between the two results it is possible to establish the Fe(II) concentration.

Iron determination samples include pharmaceutical products, drinking and natural water, aluminum alloys, multivitamins, and infant milk products.

## Lead

For spectrophotometric lead determination the procedure normally used is the addition of sulfide to a solution, which results in the formation of a brown color due to the formation of colloidal lead sulfide. Ammonium, chloride, tartrates and citrates, as well as copper, bismuth, iron, and aluminum can cause interference.

Recently, highly sensitive and selective chromogenic reagents have been used for lead determination. Dibromo-*p*-methyl-methylsulfonazo, in phosphoric acid medium, reacts with lead to form a blue complex (642 nm). The limit of quantification (LOQ), LOD, and relative standard deviations (RSD) were found to be 7.30 and 2.21 ng ml<sup>-1</sup> and 1.1%, respectively. Dibromo-*p*-methyl-carboxysulfonazo, in a phosphoric acid medium, reacts with lead to form a blue complex (648 nm). The LOD and RSD were found to be 2.14 ng ml<sup>-1</sup> and 1.0%, respectively. *Meso*-tetra-(3,5-bibromo-4-hydroxyphenyl)porphyrin, in NaOH medium, reacts with lead to form a yellow complex (479 nm). The LOD, LOQ, and RSD are 0.70 and 0.21 ng ml<sup>-1</sup> and 1.0%, respectively. 2,5-Dimercapto-1,3,4-thiazole reacts with lead, in an acid medium, to produce a greenish-yellow chelate that absorbs at 375 nm.

A kinetic method has been developed to determine micromolar amounts of Pb(II) based on the catalytic effect of chromate in acidic media on the oxidation of mercaptosuccinic acid (420 nm and LOD of 20 µg l<sup>-1</sup>).

## Magnesium

There are two methods commonly used for magnesium determination. The first one is based on reaction, in sodium hydroxide media, with Titan Yellow (sodium salt of dehydrothio-*p*-toluidine sulfonic acid) to produce a colored red colloidal suspension. The other method is based on reaction with Solochrome Black to give a red soluble complex.

Magnesium(II) has recently been determined through complex formation with purpurin (PURP) at pH 9.5 (540 nm).

A method for simultaneous spectrophotometric determination of calcium and magnesium is based on

reaction with Eriochrome Black T at pH 10.1 and utilization of multivariate calibration. Also, this simultaneous determination has been accomplished by reaction of both analytes with methylthymol blue (MTB) in a strongly basic medium (610 nm). The magnesium was masked by 8-hydroxyquinoline for calcium determination alone.

A catalytic kinetic method is based on the inhibitory action of magnesium ions on the decoloration of the acid chrome blue K-Mn(II)-KIO<sub>4</sub> system at pH 11.9. The LOD is 7.6 ng l<sup>-1</sup>.

### Manganese

Usually, manganese is determined by oxidation to permanganate by periodate or persulfate (545 nm). New chromogenic systems have been recently proposed: 5,8-dihydroxy-1,4-naphthoquinone (naphthazarin) (695 nm), 9-phenyl-2,3,7-trihydroxy-6-fluorene with formation of a blue colored complex (591 nm), and *meso*-tetra(4-(carboxymethylenoxy)phenyl)porphyrin (TCMOPPH2) in the chromogenic system TCMOPPH2-Cd(II)-imidazole and *N,N'*-bis(2-hydroxy-5-bromobenzyl)-1,2-diaminopropane with formation of a brownish complex soluble in chloroform.

Another spectrophotometric method for determination of manganese oxide is based on the reaction of manganese with 1,2,4-trihydroxyanthraquinone (PURP) with an LOD of 68 ng l<sup>-1</sup>.

Kinetic spectrophotometric methods have also been used. The catalytic effect of manganese(II) on the oxidation of diphenylamine-4-azo-benzen-4'-sulfonic acid potassium salt with potassium periodate in the presence of 1,10-phenanthroline in weak media has been reported, with an LOD of 0.017 ng ml<sup>-1</sup>. A sensitive flow injection procedure is based on the catalytic effect of manganese on the oxidation of 2-2'-azinobis(3-thylbenzothiazoline-6-sulfonic acid) with periodate (415 nm).

### Molybdenum

Spectrophotometric determination of molybdenum can be accomplished by reaction with thiocyanate in an acid medium in the presence of tin(II) chloride (465 nm). Toluene-3,4-dithiol, usually referred to as dithiol, can also be used as a spectrophotometric reagent for molybdenum determination. In an acid medium, it forms a green complex that can be extracted using organic solvents.

A new selective and rapid method for the determination of molybdenum is based on reaction with Alizarin Red S in the presence of a water soluble polymer, poly(sulfonylpiperidinylmethylene hydroxide), which produces a red complex (500 nm). A new

sensitive and selective reagent, *p*-carboxyphenylfluorene, has been used for determination of molybdenum with measurements at 531 nm. The LOD of the method has been found to be 0.73 ng ml<sup>-1</sup>. Another new chromogenic reagent has been used, dimethoxyhydroxyphenyl-fluorone, which forms a red complex with molybdenum(VI) in the presence of Triton X-100 and a sulfuric-phosphoric acid medium.

Molybdenum can be determined using a kinetic spectrophotometric method based on the catalytic effect of molybdenum on the reduction of thionine by hydrazine monochloride in an acid medium. The reaction is monitored at 605 nm due to the decrease in absorbance of thionine.

### Nickel

The most usual method of spectrophotometric determination of nickel is using the reaction with dimethylglyoxime and an oxidizing agent (bromine, for example) in alkaline solution to form a red soluble complex (445 nm).

Novel sensitive and selective chromogenic reagents have been reported for nickel determination. A new thiazolylazo reagent, 2-(2-(5-methylbenzothiazolyl)-azo)-5-dimethylaminobenzoic acid in the presence of SDS, forms a blue-violet complex (640 nm). Azocalix(4)arene, 5,17-bis(quinolyl-8-azo)25,26,27,28-tetrahydro-xycalix(4)arene reacts with nickel in a basic medium (580 nm), with an LOD of 1.4 × 10<sup>-7</sup> mol l<sup>-1</sup>. MBTAQ, in the presence of cationic surfactants and a borax buffer solution, forms a blue colored complex with nickel. The LOD of this methodology is 45 ng l<sup>-1</sup>. Another reagent is *o*-carboxylbenzenediazoaminoazo benzene, which reacts with nickel to form a complex (540 nm) at a pH of 10, and the method has an LOD of 0.22 ng ml<sup>-1</sup>. Iminodibenzyl and 3-chloroiminodibenzyl are also used for nickel determination. The reaction is carried out in a neutral aqueous medium, and a blue complex (660 nm) is formed. Finally, another reagent used is di-2-pyridil ketone benzoylhydrazone, in which the yellow Ni(II) complex presents a maximum absorbance at 406 nm.

### Selenium

Trace amounts of selenium can be determined using its reaction with potassium iodine in an acid medium to liberate iodine. The liberated iodine can react with starch to form a blue colored species (570 nm) or can bleach the violet color of thionin (600 nm). This method can be used for the determination of selenium in water, soil, plant materials, human hair, and biological samples.

Several kinetic spectrophotometric methods have been developed for selenium determination. The catalytic effect of Se(IV) on the reduction of thionine by sulfide ions has been explored. The reaction is monitored spectrophotometrically by following the decrease in absorbance at 598 nm, and the method has an LOD of  $5 \text{ ng ml}^{-1}$ . Another method is based on the catalytic effect of selenium on the oxidation of methyl yellow by hydrogen peroxide in a nitric acid medium. It is monitored by the loss of the red color of methyl yellow. The method is extremely sensitive, with the possibility of measuring concentrations of  $0.2 \text{ ng ml}^{-1}$ . Selenium can be also determined based on its effect on the oxidation reaction of methyl orange with bromate in acidic media.

### Vanadium

There are two commonly used spectrophotometric methods for vanadium determination. The first one involves the reaction of peroxide with vanadium(V) in sulfuric acid media, producing a reddish-brown coloration due to the formation of vanadyl sulfate (450 nm). The second method is based on the reaction of phosphoric acid and sodium tungstate with an acid vanadate solution. The product of these reactions is a yellow, soluble phosphotungstovanadic acid.

Several sensitive and selective methods have been proposed recently for vanadium determination based on different chromogenic reagents, such as 1,5-diphenylcarbohydrazine (1,5-diphenylcarbazine) in a slightly acid medium (531 nm), 2-(5-nitro-2-pyridylazo)-5-(*N*-propyl-*N*-sulfopropylamino)phenol in a weakly acid medium (592 nm), 2-(2-thiazolylazo)-*p*-cresol in the presence of ascorbic acid (525 nm), and 6-chloro-3-hydroxy-7-methyl 1-2-(2-thienyl)-4H-chromen-4-one in a weakly acidic medium (425 nm). Also, the utilization of 4-(2-pyridylazo)-resorcinol (PAR) in the presence of hydrogen peroxide (542 nm, LOD  $0.0028 \text{ } \mu\text{g ml}^{-1}$ ) and in the presence of cetylpyridinium chloride and ethylenediaminetetraacetic acid (600 nm, LOD  $3.0 \text{ } \mu\text{g ml}^{-1}$ ) has been proposed.

Vanadium has also been determined at 560 nm using a mixed ligand system with PAR and *N*-phenylbenzohydroxamic acid in steels and Ti alloys. With this system the linear range was  $\leq 2 \text{ mg l}^{-1}$ . Kinetic spectrophotometric methods have also been used for vanadium determination. These methods are based on the catalytic effect of vanadium in the following reactions: diphenylamine and hydrogen peroxide in a concentrated solution of formic acid (583 nm); oxidation of indigo carmine by bromate (612 nm); reduction of thionine by ascorbic acid at pH 5 (598 nm); Victoria blue B and potassium

bromate in the presence of citric acid and in a diluted sulfuric acid medium (618 nm); and oxidation of Nile Blue by bromate in an acidic medium (585 nm).

Applications related to vanadium determination and micellar media can also be found. Different oxidation states of vanadium have been determined at pH 5 using gallic acid in a cationic micellar solution of cetyltrimethyl ammonium bromide (CTAB).

## Anionic Species

### Fluoride and Chloride

Fluoride is generally determined in water and toothpaste because of its use in protecting teeth. Nowadays, there are different strategies for determining fluoride, such as the extraction-spectrophotometric method. In such methods, solid phase extraction is used for extracting fluoride from the sample (generally an aqueous matrix), and an organic solvent is used for eluting the analyte.

Besides classical reagents as Eriochrome cyanine and acid zirconyl (SPADNS), reagents such as alizarin complexone, MTB, and 2,4-diaminophenol have been proposed for fluoride determination.

Fluoride is generally determined by forming a complex with the reagents; it sometimes acts as a catalyst or inhibits chemical reactions. Example, fluoride ions form stable complexes with Fe(III), thus reducing its catalytic effect in the oxidation of 2,4-diaminophenol by hydrogen peroxide. In the other case, ternary complexes can be used for fluoride determination, such as that based on La(III)-F<sup>-</sup>-alizarin complexone (ALC).

The fluoride concentration range that can be determined by spectrophotometry generally is in the milligrams per liter range.

Regarding chloride, it seems that there are only a few applications in the literature related to its spectrophotometric determination. Classical reagents such as  $\text{HgC}_6\text{Cl}_2\text{O}_4$  or  $\text{Hg}(\text{SCN})_2$  continue to be used; however, recent examples of chloride determination are based on exchange complexation reactions proceeding in the system with immobilized diphenylcarbazone-mercury(II). Another recent application uses a mechanized system for chloride determination, and the reaction is based on the sequential introduction of mercuric thiocyanate/iron(III)/nitric acid. Generally the analytical work range is milligrams per liter, and samples such as mineral water have been analyzed.

### Nitrate/Nitrite

Reagents such as naphthol green B, gallocyanine (530 nm), pyrogallol red (465 nm), safranin O

(520 nm), and phosphomolybdenum (820 nm) have been used for nitrate/nitrite determination.

Some applications for determining nitrite are based on the decrease in colour of a complex. For examples, nitrite is determined by the oxidation of a phosphomolybdenum blue complex. Nitrite is then determined from the decrease in absorbance of the blue complex. The same analyte is also determined using its catalytic effect on the oxidation of gall-oceanine by bromate in acidic media and the decrease in absorbance of the system. However, when safranin O or resorcinol is used, the complexes formed are determined by an increased absorbance.

In the case of nitrate, the majority of applications for its determination are based on copperized-cadmium or cadmium-coated zinc reductor columns, which are generally used for reducing nitrate to nitrite. After formation of color, specific reagents are added and nitrate is determined from the difference between the total nitrite and nitrate concentrations.

Generally nitrate and nitrite are determined in samples such as natural water, soil, vegetables, and fish.

### Phosphate

Phosphate is spectrometrically determined using the molybdenum blue reaction. In this reaction, ortho-phosphate and molybdate ions condense in an acidic medium to form molybdophosphoric acid (or phosphomolybdic acid), which is reduced to produce a blue colored compound (molybdenum blue), measured at 830 nm.

Traces of phosphate ions can be determined by the reaction between molybdate, phosphate, and 3,3',5,5'-tetramethylbenzidine in an acidic medium to form a color charge transfer complex ion (458 nm). Another new approach for *ortho*-phosphate determination is based on the reaction of phosphate with the by-products of the oxidation reaction between diphenylamine and molybdenum(VI) in a formic acid solution medium (340 nm).

Generally silicate is a interference in the methodologies described previously. Methods have been developed for simultaneous determination of phosphate and silicate, based on kinetic information or using multivariate calibration. A kinetic separation method is based on the reaction of molybdenum blue and involves monitoring at different times. A multivariate calibration method proposed is also based on the reaction of molybdenum blue. Ascorbic acid and oxalic acid are added to the sample (molybdate/antimony) for the formation of two molybdate

complexes, and the spectra are recorded in the range 410–820 nm.

### Silicate

Silicate is determined by reaction with a solution of molybdate in an acidic medium to produce a complex molybdosilicic acid with a yellow color (400 nm). More accurate determinations are obtained by reduction of the complex to molybdenum blue (815 nm). Phosphates, arsenates, and germanates cause interference because they react with molybdate to produce complexes of the same color.

Silicate has also been determined based on the ion-association system formed with molybdate and rhodamine B. The proposed method has been applied to iron alloys, and the analytical range was up to 3.5 µg/50 ml with a wavelength of 585 nm.

As stated in the phosphate section, there are methods for simultaneous determination of phosphate and silicate.

### Sulfate

Sulfate is usually determined by reaction with barium chloroanilate at pH 4 to form barium sulfate and the acid chloranilate ion. At this pH value, the acid chloranilate ion is purple (530 nm). Another method is based on the formation of the  $\text{FeSO}_4^+$  cation (355 nm). An indirect method has been proposed based on the reduction of sulfate to sulfite with metallic sodium. The product is then reacted with iron(III) phenantroline to reduce iron(III) to iron(II), forming a red colored complex (515 nm).

*See also:* **Chemometrics and Statistics:** Multivariate Calibration Techniques. **Color Measurement.** **Extraction:** Solvent Extraction Principles. **Flow Injection Analysis:** Detection Techniques. **Food and Nutritional Analysis:** Water and Minerals. **Kinetic Methods:** Principles and Instrumentation; Catalytic Techniques. **Optical Spectroscopy:** Detection Devices. **Spectrophotometry:** Overview; Derivative Techniques; Biochemical Applications; Pharmaceutical Applications. **Spot Tests.** **Water Analysis:** Overview.

### Further Reading

- Elving PJ, Meehan EJ, and Kolthoff IM (eds.) (1981) *Treatise on Analytical Chemistry*. New York: Wiley.
- Ingle JD Jr. and Crouch SR (1988) *Spectrochemical Analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Kellner R, Mermet J-M, Otto M, and Widner HM (eds.) (1998) *Analytical Chemistry*. Weinheim: Wiley-VCH.
- Kolthoff IM and Sandell EB (1967) *Textbook of Quantitative Inorganic Analysis*. New York: Macmillan.



McIntire GL (1990) Micelles in analytical chemistry. *Analytical Chemistry* 21: 257–278.

McKenzie HA and Smythe LE (eds.) (1988) *Quantitative Trace Analysis of Biological Materials: Principles and Methods for Determination of Trace Elements and Trace Amounts of Some Macro Elements*. Amsterdam: Elsevier.

Rao TP, Reddy MLP, and Pillai AR (1998) Application of ternary and multicomponent complexes to spectrophotometric and spectrofluorimetric analysis of inorganics. *Talanta* 46: 765–813.

Skoog DA, Holler FJ, and Nieman TA (1998) *Principles of Instrumental Analysis*. Philadelphia: Saunders College Publishing.

## Organic Compounds

**T Prasada Rao and V M Biju**, Regional Research Laboratory, Trivandrum, India

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Analysis of organic compounds not only encompasses functional group identification but also includes quantitative determination of trace amounts of hydrocarbons, organic solvents, heterocyclic materials, surface active agents, industrial chemicals like insecticides, fungicides, or drugs, and polymers and resins. This analysis assumes increasing importance in view of the widespread use of organic compounds in the pharmaceutical, clinical, industrial, and agricultural sectors. Further, the extreme toxicity of most of the organic compounds requires their critical monitoring in environmental, biological, geological, and industrial samples. This article presents ultraviolet (UV)–visible spectrophotometric approaches used for qualitative identification and quantitative analysis of organic compounds. Even though elemental analysis of physical properties including molecular weight determination, forms a part of organic compound analysis. These are not included in the present article. Rather, emphasis is given to quantitative trace analysis of organics either by direct determination or hyphenation techniques and chemometric approaches or by adopting suitable preconcentration procedures.

### UV–visible Absorption Spectra

The electromagnetic spectrum covers a wide range of energies (frequencies) and their wavelengths. Useful frequencies range from  $>10^{19}$  ( $\gamma$ -ray) to  $10^3$  Hz (radiowaves). The visible region, to which our eyes respond, is only a minute region of the entire spectrum. Such different types of radiation as gamma ( $\gamma$ )-rays or radiowaves differ from visible light only in the energy (frequency) of their photons. **Figure 1**

shows the regions of the electromagnetic spectrum that are used for spectroscopic analyses. The types of atomic and molecular transitions that result from interactions of the radiation with a sample are also shown. The energy level spacing of many molecules, especially large organic molecules, is such that they absorb light in the UV–visible wavelength range,  $\sim 190$ – $800$  nm. **Figure 2** shows visible spectra for 1,2,4,5-tetrazine that were obtained in vapor phase, hexane, and aqueous solutions. Many individual absorption peaks resulting out of transitions from the various vibrational and rotational states are clearly evident as the individual tetrazine molecules have separated from one another to vibrate and rotate freely. In the condensed state and in solution, however, the freedom to rotate is largely lost and rotational structure is rarely observed. Furthermore, frequent collisions and interactions with solvent molecules cause the vibrational levels to be modified energetically in an irregular way.

### Qualitative Organic Analysis

#### Direct Analysis

As stated above, the UV spectrum arises from electronic excitation ( $\sigma \rightarrow \sigma^*$ ,  $n \rightarrow \sigma^*$ ,  $n \rightarrow \pi^*$ ,  $\pi \rightarrow \pi^*$ ) of the molecule by the irradiating light. The energy required for the  $\sigma \rightarrow \sigma^*$  and  $n \rightarrow \sigma^*$  transitions is relatively high with absorption in the UV region, while the  $\pi \rightarrow \pi^*$  transition of electrons appears in the absorption in the UV–visible region. The important consequence is that the spectra in the UV region are diagnostic of unsaturation in the absorbing molecule. This comes about because with few exceptions only molecules containing multiple bonds have sufficiently stable excited states to give rise to absorption in the near-UV region. Thus, saturated hydrocarbons, alcohols, and ethers are transparent in this region. Further, monofunctional olefines, acetylenes, carboxylic acids, amides, and oximes have absorption maxima just outside the near-UV region ( $\sim 200$  nm) and in general show only end absorption.

McIntire GL (1990) Micelles in analytical chemistry. *Analytical Chemistry* 21: 257–278.

McKenzie HA and Smythe LE (eds.) (1988) *Quantitative Trace Analysis of Biological Materials: Principles and Methods for Determination of Trace Elements and Trace Amounts of Some Macro Elements*. Amsterdam: Elsevier.

Rao TP, Reddy MLP, and Pillai AR (1998) Application of ternary and multicomponent complexes to spectrophotometric and spectrofluorimetric analysis of inorganics. *Talanta* 46: 765–813.

Skoog DA, Holler FJ, and Nieman TA (1998) *Principles of Instrumental Analysis*. Philadelphia: Saunders College Publishing.

## Organic Compounds

**T Prasada Rao and V M Biju**, Regional Research Laboratory, Trivandrum, India

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Analysis of organic compounds not only encompasses functional group identification but also includes quantitative determination of trace amounts of hydrocarbons, organic solvents, heterocyclic materials, surface active agents, industrial chemicals like insecticides, fungicides, or drugs, and polymers and resins. This analysis assumes increasing importance in view of the widespread use of organic compounds in the pharmaceutical, clinical, industrial, and agricultural sectors. Further, the extreme toxicity of most of the organic compounds requires their critical monitoring in environmental, biological, geological, and industrial samples. This article presents ultraviolet (UV)–visible spectrophotometric approaches used for qualitative identification and quantitative analysis of organic compounds. Even though elemental analysis of physical properties including molecular weight determination, forms a part of organic compound analysis. These are not included in the present article. Rather, emphasis is given to quantitative trace analysis of organics either by direct determination or hyphenation techniques and chemometric approaches or by adopting suitable preconcentration procedures.

### UV–visible Absorption Spectra

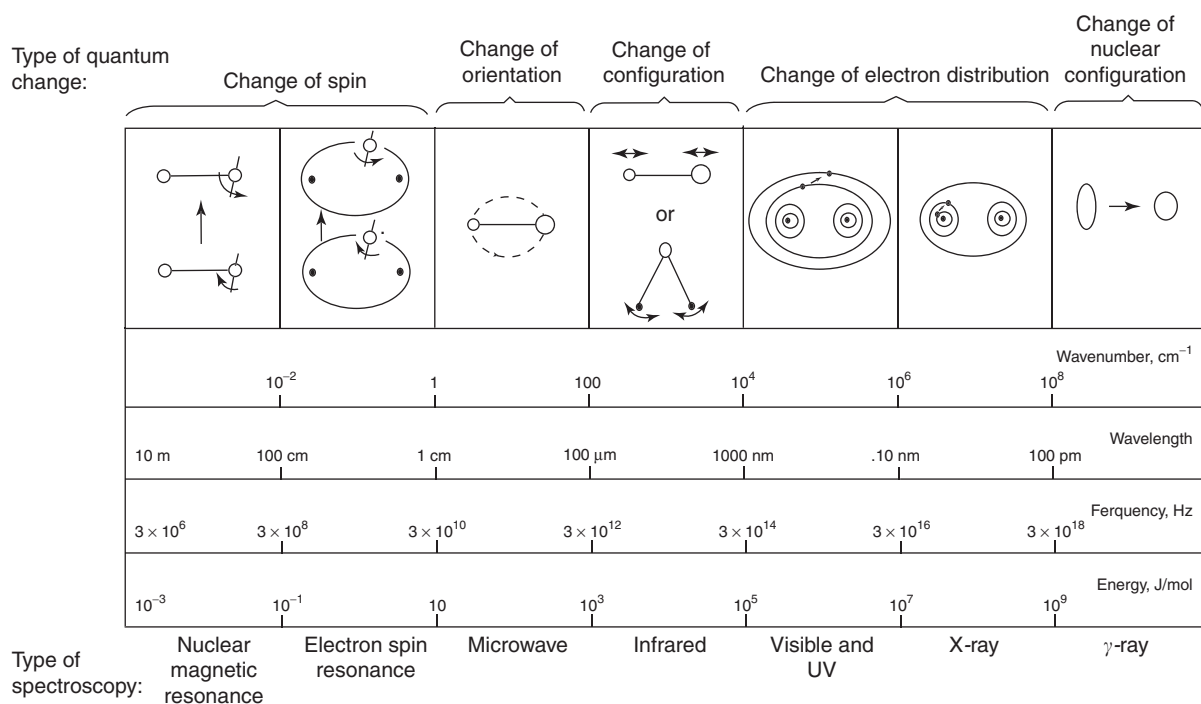
The electromagnetic spectrum covers a wide range of energies (frequencies) and their wavelengths. Useful frequencies range from  $>10^{19}$  ( $\gamma$ -ray) to  $10^3$  Hz (radiowaves). The visible region, to which our eyes respond, is only a minute region of the entire spectrum. Such different types of radiation as gamma ( $\gamma$ )-rays or radiowaves differ from visible light only in the energy (frequency) of their photons. **Figure 1**

shows the regions of the electromagnetic spectrum that are used for spectroscopic analyses. The types of atomic and molecular transitions that result from interactions of the radiation with a sample are also shown. The energy level spacing of many molecules, especially large organic molecules, is such that they absorb light in the UV–visible wavelength range,  $\sim 190$ – $800$  nm. **Figure 2** shows visible spectra for 1,2,4,5-tetrazine that were obtained in vapor phase, hexane, and aqueous solutions. Many individual absorption peaks resulting out of transitions from the various vibrational and rotational states are clearly evident as the individual tetrazine molecules have separated from one another to vibrate and rotate freely. In the condensed state and in solution, however, the freedom to rotate is largely lost and rotational structure is rarely observed. Furthermore, frequent collisions and interactions with solvent molecules cause the vibrational levels to be modified energetically in an irregular way.

### Qualitative Organic Analysis

#### Direct Analysis

As stated above, the UV spectrum arises from electronic excitation ( $\sigma \rightarrow \sigma^*$ ,  $n \rightarrow \sigma^*$ ,  $n \rightarrow \pi^*$ ,  $\pi \rightarrow \pi^*$ ) of the molecule by the irradiating light. The energy required for the  $\sigma \rightarrow \sigma^*$  and  $n \rightarrow \sigma^*$  transitions is relatively high with absorption in the UV region, while the  $\pi \rightarrow \pi^*$  transition of electrons appears in the absorption in the UV–visible region. The important consequence is that the spectra in the UV region are diagnostic of unsaturation in the absorbing molecule. This comes about because with few exceptions only molecules containing multiple bonds have sufficiently stable excited states to give rise to absorption in the near-UV region. Thus, saturated hydrocarbons, alcohols, and ethers are transparent in this region. Further, monofunctional olefines, acetylenes, carboxylic acids, amides, and oximes have absorption maxima just outside the near-UV region ( $\sim 200$  nm) and in general show only end absorption.



**Figure 1** The regions of the electromagnetic spectrum. The wavenumber, wavelength, frequency, and energy are the characteristics that describe electromagnetic radiation. (Reprinted with permission from Banwell C and McCash E (2003) *Fundamentals for Molecular Spectroscopy*. 4th edn. London: McGraw Hill; © The McGraw-Hill Companies, Inc.)

A third class of compounds comprising aldehydes, ketones, aliphatic nitro compounds, and esters is characterized by absorption maxima in the near-UV region. But they have intensities so low that they are useful only under special circumstances; i.e., acetone has a very low molecular extinction coefficient ( $\epsilon$ ) at its wavelength maximum of 270 nm. By way of comparison, a moderately strong bond would have an  $\epsilon$  of 1000–10 000  $\text{l mol}^{-1} \text{cm}^{-1}$ , and values as high as 100 000  $\text{l mol}^{-1} \text{cm}^{-1}$  are not uncommon. Table 1 lists some common chromophoric groups.

As seen from Table 1, UV spectroscopy would be limited severely if it is confined to monofunctional molecules. Fortunately, however, even functional groups such as the olefinic and carboxyl can give rise to strong absorption in the near-UV region when they are conjugated with one another. Thus, UV spectroscopy is a primary tool for the study of conjugated systems.

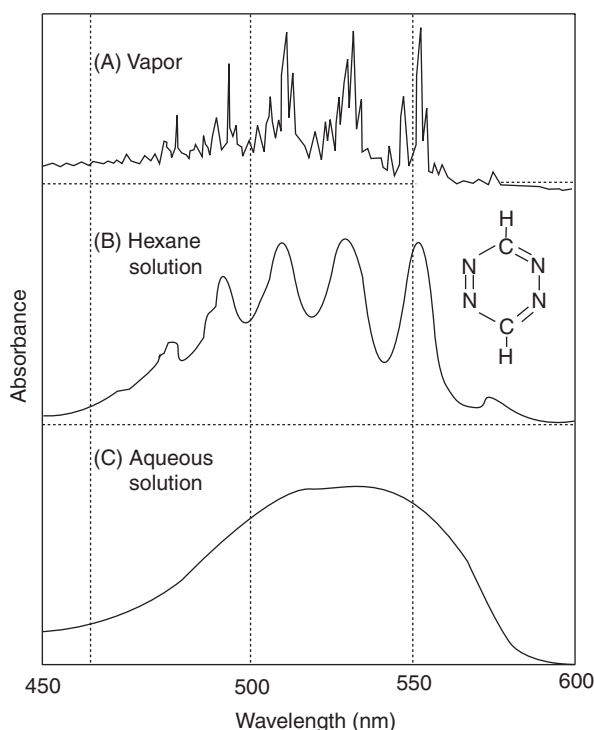
### Derivatization

Even though direct methods are often used for characterization of functional groups, in some instances, observation of the absorption spectra recorded after suitable derivatization leads to functional group identification. The types of reactions involved in such derivatization procedures include (1) formation

of azo compounds, (2) reaction with amines or phenols, (3) formation of nitro compounds, (4) formation of quinonoimino or polymethine compounds, (5) condensation of carbonyl compounds with various reagents, or (6) acidification of azomethines.

### Quantitative Trace Analysis of Organics Using UV-visible Spectrophotometry

Trace analysis refers to analysis for a minor component in a homogenous mixture. The definition of the term 'trace analysis' changes from time-to-time due to the rapid development of instrumentation or the use of sophisticated monitoring and measuring devices. By the current definition of the term 'trace component' proposed by the International Union for Pure and Applied Chemistry, the limit above which the term trace analysis can be used is 100 ppm ( $100 \mu\text{g g}^{-1}$ ). Microanalysis is a special case of trace analysis and is concerned with the analysis of a single small particle or a minor constituent in a heterogeneous mixture. Based on the analyte concentration in the sample to be examined, analytical methods and techniques are classified as below (see Table 2). The choice of analytical technique for a



**Figure 2** Typical ultraviolet absorption spectra of 1,2,4,5-tetrazine. (A) Gas phase spectrum where many lines due to electronic, vibrational, and rotational transitions are shown; (B) in a nonpolar solvent (hexane), where electronic transitions can be observed but the vibrational and rotational structures have been lost; and (C) in a polar solvent (water), where the strong intermolecular forces cause the blending together of the electronic peaks to give only a single, smooth absorption peak. (Reproduced with permission from Mason SF (1959) *Journal of Chemical Society*, p. 1265. The Royal Society of Chemistry.)

**Table 1** Examples of some common chromophoric groups

Functional group	$\lambda_{\text{max}}$ (nm)	$\epsilon$ ( $\text{l mol}^{-1} \text{cm}^{-1}$ )
R-CH=CH-R	185	8000
R <sub>2</sub> C=O	270	16
R-COOH	204	41
R-CO-Cl	235	53
R-NO <sub>2</sub>	275	12
	210	1200
R <sub>2</sub> -S	229	10900
R <sub>2</sub> C=N-NH-CO-NH <sub>2</sub>	350	140
RN=NR	300	100
RN=O	670	20

given analysis depends on the primary criteria, such as sensitivity, selectivity, precision, and accuracy and auxiliary criteria such as speed, cost of the hardware, and special requirements, if any. Spectrophotometry in the UV and visible regions of the spectrum is one of the oldest analytical techniques and is still the most commonly used. At one time, the light intensity in the visible region was measured by visual

comparison with standards, but now absorbance is measured using sophisticated instruments, which provide greatly increased sensitivity and selectivity, the keys to trace analysis.

Quantitative analysis of organic compounds using UV-visible absorption spectra involves one of the following: (1) direct determination; (2) diazotization and coupling reactions; (3) formation of charge transfer complexes; (4) oxidation or reduction reactions; (5) formation of colored binary or ternary (ion-association) complexes; (6) catalytic procedures; (7) enzymatic reactions; (8) genomic, proteomic, and phonemic analyses; and (9) online analysis. Examples of each are given below:

**Direct determination:** Tannic acid, phenazone derivatives, human hemoglobin, piperine, thermolabile drugs, and drugs containing benzene ring chromophores have been determined quantitatively using direct UV spectral measurements.

**Diazotization and coupling reactions:** Diazotization and coupling reactions are used for, for example, trace determination of primary or secondary aliphatic alcohols and aromatic amines by employing UV-visible spectroscopy.

**Formation of charge transfer complexes:** UV spectral measurements of tranquilizers and antidepressants like chlorpromazine, promethazine, promazine, thioridazine, progesterone, imprimane, and amitriptyline, hydrazines, pyrazolones, and corticosteroid drugs resulted in their determination at trace levels via formation of charge transfer complexes.

**Oxidation or reduction reactions:** Based on the oxidation reactions of (1) phenothiazines with H<sub>2</sub>O<sub>2</sub> in acetic acid or KBrO<sub>3</sub>, (2) phenols with periodic acid, (3) catechol amines, adrenaline, noradrenaline, and isopropyl noradrenaline with ceric sulfate, and (4) compounds containing *cis*- $\alpha$ -glycollic and sulfide functional groups with periodate results in trace determination of organics using UV-visible spectrophotometry. The reduction of copper(II)-phenanthroline chelate to a colored Cu(I) complex allows the determination of aliphatic thiols using UV-visible spectrophotometry. Trace amounts of pharmaceuticals like  $\beta$ -lactam antibiotics, drugs like terbutaline, isoxsuprine, and vitamin B<sub>1</sub>, phenothiazines, and dissolved proteins and amino acids like tryptophan and tryosine have been determined based on the reduction of heteropoly acid into molybdenum blue followed by UV-visible spectrophotometry.

**Formation of colored binary or ternary (ion-association) complexes:** Binary complexes. Amino acids such as glycine, L-alanine, L-serine, L-phenylalanine,

**Table 2** Classification of analytical methods and techniques based on analyte concentration in a sample

S. no.	Analyte concentration	General name of analyte	Common term for analytical procedure	Examples
1	< 1 ppt ( $< 10^{-10}\%$ )	Submicrotrace component	Ultratrace analysis	Determination of dioxins in samples of various matrices
2	< 1 ppb ( $< 10^{-7}\%$ )	Ultramicrotrace component	Ultratrace analysis	Determination of trihalomethanes in drinking water and human urine. Determination of volatile organic compounds in indoor air
3	< 1 ppm ( $< 10^{-4}\%$ )	Microtrace component	Trace analysis	Determination of CO in ambient air
4	< 100 ppm (0.01%)	Trace component	Trace analysis	Determination of methane in ambient air
5	< 1%	Secondary component (admixture)	Semimicroanalysis	Determination of CO <sub>2</sub> in ambient air
6	1–100%	Primary component	Macroanalysis	Determination of carbon in coal dust

L-glutamic acid, and L-lysine have been determined spectrophotometrically using *p*-benzoquinone as the color-forming reagent. Ethyl violet is used as a chromogenic reagent for determination of anionic surfactants in mineral water and soft drinks using UV–visible spectrophotometry. A new color-forming reagent,  $\alpha,\beta$ -dinitrostilbene, has been used for UV–visible spectrophotometric determination of primary and secondary amines when these are sufficiently basic and sterically unhindered. Primary and aromatic amines can be determined by reacting them with syringaldehyde in a HCl–methanol medium using UV–visible spectroscopy. Similarly, sulfonazo III has been used for spectrophotometric determination of ascorbic acid. Reserpine, amidopyrine, diazepam, and other benzodiazepine tranquilizers can be determined by extraction of their binary complexes with alizarin violet 3B or alizarin brilliant violet R. Penicillins and cephalosporins have been determined by formation of 1:1 and 2:1 complexes with chloranilic acid during UV–visible spectrophotometric determination. Coomassie Brilliant Blue G can be used for trace determination of urine proteins, plant proteins, acid soluble proteins, and thermally denatured and insoluble proteins using UV–visible spectroscopy.

Ternary complexes. Traces of proteins have been determined spectrophotometrically by formation of colored ternary ion-association complexes with pyrogallol red and molybdenum(VI) or pyrocatechol violet and molybdenum(VI). Compounds containing the sulfydryl group were determined spectrophotometrically based on the extraction of their ternary ion-association complexes of rhodamine 6G and as-traphloxine cations with hexachloroantimonate(III) from buffered aqueous solutions into toluene.

Hydroquinone has been determined spectrophotometrically by reducing iron(III) to iron(II), which was extracted into chloroform as the ferroin–bromophenol blue ion-association complex.

**Catalytic procedures:** Ultratrace amounts of catalysts such as formaldehyde and oxime iodides can be determined by catalytic oxidation of Brilliant cresyl blue bromate and Mn(III)–As(III) reactions, respectively. Furthermore, determination of organic compounds is possible by exploiting their influence on reaction rates based on the activation or inhibition effects. Thus, endosulfan is determined based on its activation of the catalytic oxidation of caraminic acid by H<sub>2</sub>O<sub>2</sub>. On the other hand, ascorbic acid, resorcinol, and oxalate can be determined based on the inhibition of Mn(II)-catalyzed oxidation of malachite green with KIO<sub>4</sub>, formaldehyde-catalyzed oxidation of neutral red by bromate, and tungsten-catalyzed oxidation of the iodide ion by H<sub>2</sub>O<sub>2</sub>. Organic compounds such as acetaminophen (paracetamol), phenol, and nitroprusside, have also been determined by an inhibition effect on the photolysis of a methyl viologen/EDTA/acridine yellow system that was irradiated with a halogen lamp and by measuring the time required for reaching a preset absorbance.

**Enzymatic reactions:** UV–visible absorption spectrophotometric determination of enzyme inhibitors is much more popular than that of enzyme activators. Drugs such as neostigmyne, chlorpromazine, and cisplatin can be determined by inhibition of soluble or bound enzymes, viz., acetylcholine sterase, Nicotinamide adenine dinucleotide hydrogenase (NADH), and peroxidase, respectively.

*Genomic, proteomic, and phenomic analyses:*

The genome is the total sum of all the genes in an organism. Normally, every cell in an organism contains an identical copy of the genome. During the past decade, an estimated 180 microbial genome sequencing projects churned out ~86 microbial genomes that have been sequenced completely and published. In molecular microarray fabrication, nanoliter sized volumes are arrayed instantaneously so that simultaneous characterization of genes and bacteria can be achieved in a very short time. Attention is now turning to the study of proteins as harvesting of knowledge from genome sequence data requires systematic and large-scale study of proteins. The ensemble of proteins related to a genome is known as a proteome. The understanding of proteomes is more useful than that of a genome; (1) as every cell in an organism contains an identical genome but proteins use the different genetic information to form different types of cell – liver, muscle, bone, or blood cells; (2) for selecting 20 or so genes out of 160 000; (3) because many human diseases are related to incorrect modification of normal proteins and thus allow diagnosis of the disease; and (4) because proteins are targets for nearly all drugs used for treatment. Proteomic analysis covers large-scale analysis of protein structures, protein expression, and protein interactions. Little research has been done on describing bacterial phenotypes in a more comprehensive manner, even though tremendous advances in DNA microarray and proteomic methodologies have been made. The advantage of phenomic research is the ability to conduct simultaneous testing of numerous bacterial phenotypes using an automated instrument by measuring the absorbance of a substrate color change of a tetrazolium indicator.

*Online analysis:* Online sample processing techniques such as flow injection provide advantages such as reliability, sample economy, ease of automation, measurement standardization, high speed, optional sample dilution, and the ability to derivatize the analyte so as to suit the analyzer/detector. These procedures facilitate the online monitoring of fermentation substrate materials, respiratory gases, and biomass. The modifications to flow injection analysis for accurate discontinuous flow operation include sequential injection analysis and bead injection spectroscopy. The most recent invention in online techniques is the introduction of the Lab-on-a-Valve, which opens the way to development of a novel type of microflow analytical system monitored by UV–visible spectrophotometry using fiber optics. This system is an ideal tool for fermentation monitoring.

## Sample Processing via Preconcentration/Enrichment

Typically, when measurements are made at the trace and ultratrace levels, the sample volumes are relatively large. Hence, direct determination of traces of organics using spectrophotometry is still difficult because of the insufficient sensitivity and selectivity of the methods used. Therefore, pretreatment of samples, such as preconcentration of the analyte before its determination, is frequently necessary. Preconcentration improves the analytical detection limit, increases the sensitivity by several orders of magnitude, enhances the accuracy of the results, and facilitates calibration. The preconcentration approaches that can be employed prior to spectrophotometric determination of traces of organics are summarized below.

### Liquid–Liquid Extraction

Liquid–liquid extraction (LLE), although used primarily as a separation technique, is an important enrichment technique. LLE is based on the distribution of an analyte between two essentially immiscible solvents. The distribution ratio of an analyte is defined as the ratio of its total concentration in the organic phase to that in the aqueous phase at equilibrium. When the desired organic compound is extracted, the larger the distribution ratio of the analyte and the smaller that of the matrix, the higher the recovery of the organic compound and enrichment factor. When the matrix is removed by extraction, the reverse is required for successful enrichment. After the extraction, back washing of matrix components selectively from the organic phase into the aqueous phase improves the enrichment factor.

### Supercritical Fluid Extraction

The term ‘supercritical’ denotes conditions above a critical temperature and pressure. However, interesting behavior occurs throughout the critical region. The advantages of supercritical fluid extraction (SFE) over conventional extraction methods include more rapid extraction rates, the possibility of more efficient extractions, increased selectivity, possible analyte fractionation during extraction, and high-enrichment factors and compatibility with online analysis methods. Offline SFE, ultrasonic SFE, and online SFE methodologies find widespread application for preconcentration of organics present in complex matrices. Extraction rates generally increase over an order of magnitude compared with conventional Soxhlet extraction methods. On the other hand, online SFE methods combine sample



preparation and analysis and provide the potential for rapid and highly sensitive analyses.

### Gel and Ultrafiltration

Gel filtration is based on the inclusion and subsequent elution for fractionating and enriching dissolved organic compounds according to their molecular size or molecular weight differences by passing through a porous polymeric gel such as a molecular sieve.

Ultrafiltration is employed to enrich organic compounds with appropriate cut-off molecular weights using a variety of ultrafiltration membranes.

### Dialysis

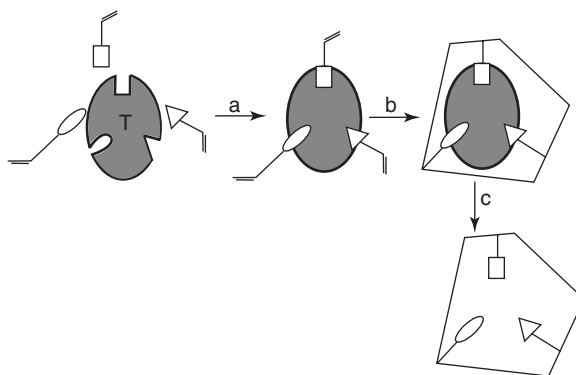
In an *in situ* dialysis technique, a dialysis bag filled with pure solvent is directly immersed into a natural aqueous sample *in situ* to enrich truly dissolved forms of organic compounds without substantial influence of adsorption on the walls of vessels and apparatus.

### Solid-Phase Extraction

Solid-phase extraction (SPE) has emerged as a powerful tool for separation/enrichment of organics. The basic principle of SPE is transfer of analytes from a liquid phase to active sites of an adjacent solid phase or sorbent. The sorbent material retains selectively the analyte, which is subsequently stripped by an appropriate solvent. Sorbents include activated carbon, silica and silica gel, glass beads, cellulose, polyurethane foam, and polymer supports such as macroreticular XAD resins, etc. The advantages of SPE over LLE include higher enrichment factors, absence of emulsion, safety with respect to hazardous samples, minimal costs due to low consumption of reagents, flexibility, ease of automation, and, more importantly, greater environmental friendliness as minute volumes of solvents and reagents are used.

### Molecular Imprinted Polymers for SPE (MIP-SPE)

Molecular imprinting is a way of making artificial locks for molecular keys. The selected molecule (target analyte) is first mixed with a variety of lock-building blocks (functional monomers). The building blocks and the key are allowed to bind each other strongly or loosely (Figure 3a). The so formed complexes consisting of building blocks and keys are glued together in order to fix the building block positioned around the key by polymerization (Figure 3b). Finally, by removing the molecular key, it leaves a construction that is selective for the original key and will not recognize any other key (Figure 3c).



**Figure 3** Schematic representation of molecular imprinting process: T, template; A, mixing and prearrangement; B, polymerization; C, template washing and Soxhlet extraction. (Reproduced with permission from Rao et al. (2004) *Trends in Analytical Chemistry* 23: 24.)

Thus, the molecular memory introduced into the polymer is now capable of enriching the analyte with high specificity. One of the many attractive features of the molecular imprinting method is that it can be applied to a diverse range of analytes. The imprinting of the organic molecules, viz., amino acids, peptides, pesticides, pharmaceuticals, nucleotide bases, steroids, and sugars is now well established. The importance of polymer- and template-related factors in the development of MIPs for use in SPE preconcentration of organics has been well brought out recently in the area of biopharmaceutical and environmental analyses. In addition, preparation of MIP-SPE materials is simple and convenient.

### Hyphenated Techniques

Hyphenation refers to the online combination of a separation technique and a spectroscopic detection method that provides structural information on the analytes concerned. Liquid chromatography (LC), mass spectrometry (MS), and gas chromatography (GC) are the most popular hyphenated techniques in use today. The choice of detection is important to the overall scheme of LC make up and is contingent upon criteria such as the noise, sensitivity, and linearity. Of the two basic categories of detectors, viz., solute and bulk property detectors, UV detection belongs to the former category.

### Hypernation and Extended-Hypernation Techniques

The real problem, as of now, is not how to improve the analytical performance but how to handle the plethora of data generated per run. Thus, a combination different hyphenated techniques is an obvious

choice for tackling challenging analytical problems and is usually called a hypernation – a term indicating that one is now ‘one higher than’ (or hyper) hyphenation. On the other hand, the obvious disadvantage of such hypernation is that, whilst it may be efficient in terms of run time, sample size, and data correlation, it is very inefficient in utilization of expensive spectrometers. Recent developments indicate the introduction of an extended hypernation technique that combines UV, MS, Inductively coupled plasma-MS (ICP-MS), nuclear magnetic resonance, Fourier transform infrared, etc. detectors for LC-based operations. However, the whole area of LC-based (extended) hypernation is still in an early stage of development, and major improvements can be expected. As the analytical problems that have to be solved tend to become increasingly complex, there is every reason to believe that in the longer term the benefits will certainly outweigh the costs – not for routine problem solving but for whatever is new, challenging, and complicated.

### Detectors in UV-visible Spectroscopy

Advances in detectors have been important for techniques like UV-Visible spectrophotometry and high-performance LC (HPLC)-UV, where rapid progress has been made over the recent past. Photomultiplier tubes continue to enjoy high popularity in forward optic instruments. However, in the past decade, detectors that permit simultaneous acquisition of spectral information over the whole wavelength range have dominated the market. Simultaneous detection may be achieved using a single detector that receives encoded information, which is later decoded mathematically (such as by Fourier transformation) or using multichannel techniques in which spectral information is either spatially or temporally dispersed (optical imagers). The rapid development of the later devices took place due to advances in semiconductor technology and proliferation of powerful and relatively inexpensive microprocessor systems. Thus, photodiode detector arrays (PDAs), despite their relatively high price, have been widely used as multichannel detectors for simultaneous monitoring at all wavelengths, in addition to detection of transient signals in HPLC and flow injection analysis. Thus, a complete spectrum can be recorded within a few milliseconds because the scan time is not affected by the movement of the diffractor. It is not exaggerating to say PDA detection in conjunction with LC will soon be the most frequently used technique for analysis of organic compounds that cannot be analyzed using GC.

### Chemometric Approaches in UV-visible Spectroscopy

Chemometrics is the science of relating measurements made on a chemical system (including dynamic chemical processes) to the state of the system via application of mathematical or statistical algorithms. It is clear from this definition that chemometrics is data based. The goal of many chemometric techniques is the production of an empirical model, derived from data, that allows one to estimate one or more properties of a system from measurements. The four important performance attributes that can be improved through the use of chemometric techniques are accuracy, precision, robustness, and reproducibility.

As in other analytical techniques, a natural trend in UV-visible spectrophotometry is an approach to multicomponent analysis. The quantitation of compounds with highly overlapping spectra in a mixture has always been a difficult analytical problem, especially at unequal analyte concentration levels. With the advent of the fast scanning PDAs mentioned above and low cost computers capable of processing complex data sets, new horizons have been opened for mathematic processing of the information acquired. Two powerful signal processing techniques, viz., multiwavelength and derivative spectrophotometry, have been developed rapidly in the recent past.

### Multiwavelength Simultaneous Analysis

The advent of low priced powerful microcomputers has enabled easy processing of spectrophotometric data. In addition, fast scanning detectors have allowed elimination of errors resulting from poor time stability when a dual wavelength method is used with a single wavelength spectrophotometer. Modern multiwavelength analysis utilizes the reversed matrix representation of Beer-Lambert's law. This can be combined with chemometric techniques such as principal component analysis (PCA) and partial least squares (PLS).

### Principal Component Analysis

The purpose of PCA is to give an overview of the dominant information patterns in the data. These are the relationships between the spectra and the wavelengths which can be explored mathematically or graphically. PCA sets out to express the main information of a calibration set of spectra in terms of a lower number of variables – these are the principal components of the original data matrix. This has the effect of maximizing the amount of information in a

greatly reduced number of variables and allows a much more efficient understanding and manipulation of the data. It is achieved by performing an eigen analysis of the covariance matrix of the spectral data matrix. One of the most important aspects of PCA is to choose how many of the components are significant for modeling the original data. The significance is of both a statistical and chemical nature. In essence, the information content of a large spectral dataset is distilled into a small number of principal components.

For quantitative analysis applications, these principal components can then be regressed with calibration values of the standards to produce the system model. This model is then used to predict the assay values of a test sample from its spectrum. Together with the assay figures, these models can indicate the errors associated with the prediction and validity of applying the model against the test sample. These are important diagnostics when the approach is a part of an automated analysis system and must be made to failsafe on the assay.

### Partial Least Squares

PLS is being used as a data reduction/modeling technique for spectral data. PLS is similar to PCA in extracting a series of principal components from the data but differs in that both the spectral data and the property or assay data are used together in an iterative fashion to build a model. The chemical data are used to find a pattern in the spectroscopic data that correlates with them. This ensures that the estimated regression factors have relevance toward the chemical values.

The advantage of PLS over PCA is to model a system with both fewer and more relevant dimensions than PCA. Hence, the model in PLS more robust and easier to be visualized by the user.

### Derivative UV-visible Spectrometry

The rapid evolution of microcomputers has led to the derivative transformation of spectral data, which offer a powerful tool for both qualitative and quantitative analysis of mixtures of organic compounds. The method has found increasing application in UV-visible spectrophotometric analysis of organics for background correction and for resolution enhancement. The ability to eliminate matrix interferences such as irrelevant absorption and light scattering has been of particular value.

It is clear that the second derivative of radiation intensity becomes nonlinear with concentration, except where the term  $(d\epsilon/d\lambda)$  is zero. It is common and convenient to measure absorbance,  $a(\log \phi_0/\phi)$ , at a defined wavelength,  $\lambda$ .

$$a = \epsilon bc$$

Thus

$$d^n a = d\lambda^n = d^n \epsilon / d\lambda^n bc$$

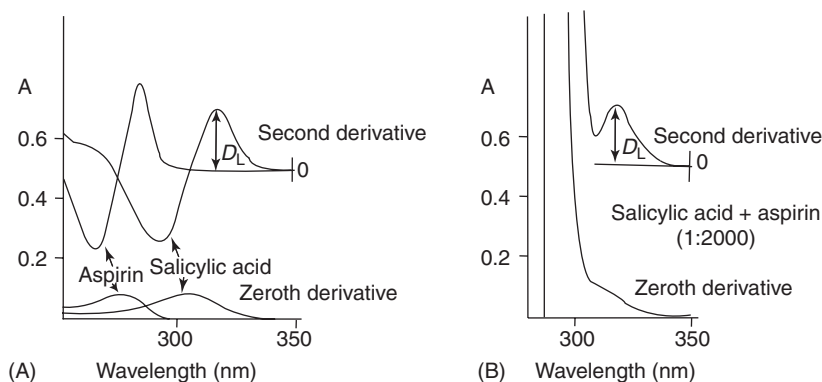
where  $n$  is the derivative order. When the matrix interference can be approximated as a linear function, the first derivative yields a function where the interference is reduced to a constant:

$$Y = ax + K$$

$$\frac{dy}{dx} = a$$

$$dy/dx = a$$

For total elimination of the matrix interference from the spectrum of interest, the derivative of order  $n + 1$  is therefore needed, where  $n$  is the highest power of the polynomial equation that represents the matrix. Matrix absorption can often be approximated by a linear function, so that the second derivative transformation can be used to eliminate



**Figure 4** Use of derivative spectra to determine salicylic acid in aspirin (A) the zeroth and second derivatives of both materials are shown. (B) The equivalent spectra of the mixture in a 1:2000 ratio are shown. (Reprinted with permission from Kitamura K and Majima R (1983) *Analytical Chemistry* 55: 54. © American Chemical Society.)

such interferences completely. An example of application of second derivative measurements is illustrated in **Figure 4**. In determination of salicylic acid in aspirin powder, the salicylic acid signal is obscured by the overwhelming concentration of acetyl salicylic acid. However, the second derivative spectrum yields a distinct peak for the salicylic acid, which can be used for the determination with good precision. In many situations, however, interferences can be attributed to the presence of other UV absorbing substances. Depending on the relative positions of the particular analytical bands concerned, polynomial equations of various degrees will be required in order to describe the interfering component effectively. Interference in such samples can often be described by a quadratic function requiring a third order derivative for complete deconvolution of overlapping UV–visible spectra.

*See also:* **Derivatization of Analytes.** **Extraction:** Solid-Phase Extraction. **Qualitative Analysis.** **Spectrophotometry:** Overview.

## Further Reading

- Bartsch RA and Maeda M (1998) *Molecular and Ionic Recognition with Imprinted Polymers*. Washington, DC: American chemical society.
- Christian GD and Callis JB (1986) *Trace analysis: Spectroscopic Methods for Molecules*. New York: Wiley.
- George WO and Willis HA (1990) *Computer Methods in UV–Visible and IR Spectroscopy*. Cambridge: Royal Society of Chemistry.
- Poole CF (2003) New Trends in Solid Phase Extraction. *Trends in Analytical Chemistry* 22: 362–373.
- Skoog DA, West DM, Holler FJ, and Crouch SR (1999) *Analytical Chemistry: An Introduction*, 7th edn. New York: Saunders College Publishing.
- Snell FD, Snell CT, and Snell CA (1967) *Colorimetric Methods of Analysis*, vols. IV, IV A, and IV AA. New York: van Nostrand.
- Weiss FT (1970) *Determination of Organic Compounds: Methods and Procedures*, vol. 32. New York: Wiley-Interscience.
- Wilson ID and Brinkmann UATh (2003) Hyphenation and hyphenation: the practice and procedures of multiple hyphenation. *Journal of Chromatography*, 1000: 325–356.

## Biochemical Applications

**F Sánchez Rojas and J M Cano Pavón**, Universidad de Malaga, Malaga, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Ultraviolet (UV)–visible spectrophotometry is widely used in biochemistry, both for the determination of species and for studying biochemical processes. This technique enables the determination of micromolar concentrations of substances and has a broad scope of application in this field since most biochemical compounds absorb in the UV–visible region or can be converted into some absorbing derivative. On the other hand, this often results in interferences during the determination of the compound of interest, interferences that are minimized by using either derivative spectrophotometry, particularly in the presence of strong background absorption or, to a lesser extent, differential spectrophotometry.

Spectrophotometric methods are usually less sensitive than fluorimetric ones, which accommodate concentrations of a few nanomoles per milliliter; however, their range of applications are much broader,

including inorganic species, organic compounds, proteins, etc., since UV–visible absorption is more universal a property than fluorescence. Spectrophotometry is especially useful for monitoring enzymatic reactions, either to determine the reaction products directly or to measure the reaction rate.

The most salient applications of UV–visible spectrophotometry to biological systems classified according to the type of substance studied are described herein.

### Inorganic Ions

Inorganic ions in biological systems can be quantified by spectrophotometric methods provided they are present at suitable concentrations (of a few micrograms per milliliter or higher) and other species in the medium exhibit little or no absorption. However, many elements including aluminum, cadmium, chromium, and nickel usually cannot be determined by UV–visible spectrophotometry, because they occur at very low concentrations in biological systems.

UV–visible spectrophotometric determinations of inorganic ions require the use of a reagent forming a

such interferences completely. An example of application of second derivative measurements is illustrated in **Figure 4**. In determination of salicylic acid in aspirin powder, the salicylic acid signal is obscured by the overwhelming concentration of acetyl salicylic acid. However, the second derivative spectrum yields a distinct peak for the salicylic acid, which can be used for the determination with good precision. In many situations, however, interferences can be attributed to the presence of other UV absorbing substances. Depending on the relative positions of the particular analytical bands concerned, polynomial equations of various degrees will be required in order to describe the interfering component effectively. Interference in such samples can often be described by a quadratic function requiring a third order derivative for complete deconvolution of overlapping UV–visible spectra.

*See also:* **Derivatization of Analytes.** **Extraction:** Solid-Phase Extraction. **Qualitative Analysis.** **Spectrophotometry:** Overview.

## Further Reading

- Bartsch RA and Maeda M (1998) *Molecular and Ionic Recognition with Imprinted Polymers*. Washington, DC: American chemical society.
- Christian GD and Callis JB (1986) *Trace analysis: Spectroscopic Methods for Molecules*. New York: Wiley.
- George WO and Willis HA (1990) *Computer Methods in UV–Visible and IR Spectroscopy*. Cambridge: Royal Society of Chemistry.
- Poole CF (2003) New Trends in Solid Phase Extraction. *Trends in Analytical Chemistry* 22: 362–373.
- Skoog DA, West DM, Holler FJ, and Crouch SR (1999) *Analytical Chemistry: An Introduction*, 7th edn. New York: Saunders College Publishing.
- Snell FD, Snell CT, and Snell CA (1967) *Colorimetric Methods of Analysis*, vols. IV, IV A, and IV AA. New York: van Nostrand.
- Weiss FT (1970) *Determination of Organic Compounds: Methods and Procedures*, vol. 32. New York: Wiley-Interscience.
- Wilson ID and Brinkmann UATh (2003) Hyphenation and hyphenation: the practice and procedures of multiple hyphenation. *Journal of Chromatography*, 1000: 325–356.

## Biochemical Applications

**F Sánchez Rojas and J M Cano Pavón**, Universidad de Malaga, Malaga, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Ultraviolet (UV)–visible spectrophotometry is widely used in biochemistry, both for the determination of species and for studying biochemical processes. This technique enables the determination of micromolar concentrations of substances and has a broad scope of application in this field since most biochemical compounds absorb in the UV–visible region or can be converted into some absorbing derivative. On the other hand, this often results in interferences during the determination of the compound of interest, interferences that are minimized by using either derivative spectrophotometry, particularly in the presence of strong background absorption or, to a lesser extent, differential spectrophotometry.

Spectrophotometric methods are usually less sensitive than fluorimetric ones, which accommodate concentrations of a few nanomoles per milliliter; however, their range of applications are much broader,

including inorganic species, organic compounds, proteins, etc., since UV–visible absorption is more universal a property than fluorescence. Spectrophotometry is especially useful for monitoring enzymatic reactions, either to determine the reaction products directly or to measure the reaction rate.

The most salient applications of UV–visible spectrophotometry to biological systems classified according to the type of substance studied are described herein.

### Inorganic Ions

Inorganic ions in biological systems can be quantified by spectrophotometric methods provided they are present at suitable concentrations (of a few micrograms per milliliter or higher) and other species in the medium exhibit little or no absorption. However, many elements including aluminum, cadmium, chromium, and nickel usually cannot be determined by UV–visible spectrophotometry, because they occur at very low concentrations in biological systems.

UV–visible spectrophotometric determinations of inorganic ions require the use of a reagent forming a

**Table 1** Characteristics of inorganic species by UV–visible spectrophotometric determinations

Inorganic species	Reagent	$\lambda_{\max}$ (nm)
Fe	1,10-Phenanthroline	510
	Bathophenanthroline	
Mo	Dithiol	680
Cu	Cuproine	546
	Neocuproine	
Mg	Thiazol yellow	540
Nitrate (reduction to nitrite)	Griess	540
Sulfide	<i>N-N</i> -Dimethyl- <i>p</i> -phenylenediamine	669
Ca	Sodium cloranalate	520

colored compound with the species to be determined. The compound can be a complex, an absorbing product (a lacquer), or an oxidized or reduced product, and should be stable under the experimental measurement conditions.

Ammonium ion is one of the inorganic species most frequently determined by UV–visible spectrophotometry. It is freed by the Kjeldahl method from organic matter and can be spectrophotometrically determined by reaction with phenol and hypochlorite in the presence of nitroprusside ion, with which it yields a blue compound (indophenol) with an absorbance maximum at 625 nm. Another alternative, especially avoiding the use of phenol, is by measuring the absorbance at 660 nm of an indophenol blue derivative generated by the reaction of ammonium ion with sodium salicylate in the presence of hypochlorite ion. These procedures are routinely used for the determination of ammonium ion with autoanalyzers.

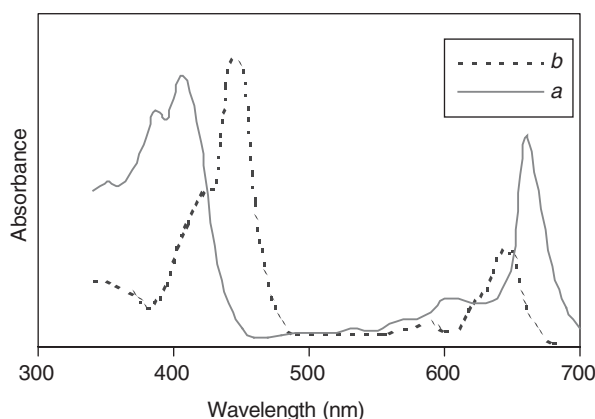
Phosphate is another inorganic species commonly encountered in biological systems that can be assayed spectrophotometrically. It reacts with ammonium molybdate to form 12-molybdophosphate heteropolyacid  $[\text{PMo}_{12}\text{O}_{40}]^{3-}$ , which is reduced by ascorbic acid to a blue product ( $\lambda_{\max} = 820 \text{ nm}$ ).

Other common UV–visible spectrophotometric determinations of inorganic species are listed in **Table 1**.

Kinetic catalytic methods make use of a number of inorganic oxidation indicator reactions, the most representative of which is probably the decomposition of hydrogen peroxide with different reagents catalyzed by copper, iron, manganese, etc., which allows these species to be determined by UV–visible spectrophotometry.

## Common Organic Compounds

Some organic compounds commonly occurring in biological systems absorb sufficiently in the UV–visible region, whereas others show little or no



**Figure 1** Absorbance spectra of chlorophylls *a* and *b* dissolved in 80% acetone. (Reproduced with permission from Pineda M and Cárdenas J (1988) *UV–Visible Spectrophotometry of Biological Compounds (Espectroscopía Ultravioleta–Visible de Compuestos Biológicos)*. Córdoba (Spain): Publications of Monte de Piedad and Caja de Ahorros of Córdoba.)

absorption so that they require some transformation into a colored, absorbing substance.

Chlorophylls are among the organic compounds whose intrinsic absorption is adequate for spectrophotometric quantification. These substances, responsible for the green color of plants, are insoluble in water but soluble in organic solvents (ether, acetone). The absorbance spectrum varies between chlorophylls (**Figure 1**) as a result of the presence of different substituents in the parent protoporphyrin–magnesium structure. Chlorophyll in plants is determined by grinding fresh vegetable material and extracting the pigment into an organic solvent.

However, organic biological compounds must be converted into colored species for their determination. This can be accomplished by means of a redox reaction such as that which takes place between reducing sugars (aldoses and ketoses) and ammonium molybdate (which yields molybdenum blue,  $\lambda_{\max} = 700 \text{ nm}$ ), picric acid (which is reduced to picramic acid,  $\lambda_{\max} = 460 \text{ nm}$ ), or uric acid (which reduces phosphomolybdate heteropolyacid). Reducing sugars can also be determined by reduction of hexacyanoferrate(III) to (II) and subsequent formation of Prussian blue (the Park and Johnson method). Other conversion reactions involve the formation of condensation products such as those of ketoses with diphenylamine or glyoxylic acid (resulting from the hydrolysis of allantoic acid, which in turn originates from hydrolysis of allantoin) with 2,4-dinitrophenylhydrazine. The converted products can also be of a different nature (e.g., a red salt obtained by reaction of creatinine and picric acid, the unknown



substance produced by glutathione and alloxan, and the red product yielded by fructosamine on reduction of nitroblue tetrazolium salts).

Unsaturated lipids in blood serum can be spectrophotometrically determined by using sulfuric acid and vanillin phosphate, which produces a red substance of unknown composition whose absorbance is measured at 520–530 nm.

## Amino Acids

Amino acids can easily be determined in a direct manner since they absorb significantly in the near-UV region. The main problem in their quantitation in biological systems is the strong interferences from other species present. Interferences can largely be overcome by using derivative spectrophotometry, which minimizes matrix effects. For example, phenylalanine and tryptophan can be determined using their second-derivative spectra; phenylalanine at 250–265 nm and tryptophan at 290–295 nm.

A similar procedure based on second- and fourth-derivative spectra can be used for the resolution of mixtures of phenylalanine, tryptophan, and tyrosine. Fourth-derivative spectra show strong bands for tyrosine and phenylalanine that are markedly affected by the polarity of the medium. Other amino acids have been determined in various biological samples, especially protein hydrolysates, with varying results.

Derivatization reactions are also commonplace in the determination of amino acids. One such reaction is that with ninhydrin-ascorbate. In the presence of amino acids, ninhydrin (triketohydrindane hydrate) is reduced to hydridantin by ascorbic acid while the amino acids undergo oxidative deamination with the formation of ammonium ion, which is condensed with hydridantin to yield a colored product with maximal absorbance at 405 and 575 nm. The procedure is often used for the determination of amino acids in protein hydrolysates.

Another interesting subject is the use of a reagent that allows the stoichiometry and reactivity of amino acid side chains to be monitored spectrophotometrically. For example, the classical reagent 5,5'-diethiobis-(2-nitrobenzoic acid) (known as Ellmans reagent) is very selective for the thiol side chain of cysteine residues. Others reagents used for this general purpose are: (1) trinitrobenzenesulfonic acids, which react well with the amine groups of reactive lysine residues, (2) tetranitromethane for modification of tyrosine side chains to give a colored product that is easily monitored by spectrophotometry, (3) diethylpyrocarbonate for histidine residues, (4) 2-nitrophenyl-sulphenyl chloride, which reacts with

tryptophan side chains to give products that can be measured spectrophotometrically.

## Proteins

Quantitative determination of proteins is often needed to determine their concentrations and changes in cell growth and differentiation processes or in enzyme purification procedures.

Occasionally, a peptide can be quantified directly provided it exhibits a significant, characteristic absorbance. Otherwise, a derivatizing reagent must be employed to obtain an adequate absorbance in the UV–visible region. Examples of these two types of application are detailed below.

### Direct Absorbance Measurements

Most proteins absorb strongly in the near-UV and visible regions. The polypeptide portion of the protein usually absorbs at short wavelengths; its absorbance, mainly due peptide bonds, appears as a weak band ( $\lambda_{\max} = 215$  nm,  $\epsilon = 100 \text{ l mol}^{-1} \text{ cm}^{-1}$ ) because of an  $n \rightarrow \pi^*$  transition and a strong band ( $\lambda_{\max} = 190$  nm,  $\epsilon = 7000 \text{ l mol}^{-1} \text{ cm}^{-1}$ ) produced by a  $\pi \rightarrow \pi^*$  transition. A number of other amino acids including tryptophan ( $\lambda_{\max} = 219$  nm) absorb in the same zone. Table 2 lists the characteristic absorbance parameters of some selected proteins.

The prosthetic groups responsible for the absorbance of many proteins at long wavelengths can be of very different nature. The principal compounds in this respect are flavins, tetrapyrroles, carotenoids, and flavonoids.

Flavins are riboflavin derivatives and includes a group of chromophoric substances that exhibit absorbance bands in the visible region (usually between

**Table 2** Absorbance parameters for selected proteins

Protein	Source	$\lambda$ (nm)	$\epsilon$ ( $\text{l mol}^{-1} \text{ cm}^{-1}$ )
Albumin	Bovine serum	280	44 700
Cytochrome <i>a</i>	Mitochondria	425	76 000
Cytochrome <i>b</i>	Mitochondria	563	21 000
Ferredoxin	Pumpkin	422	9800
Fibrinogen	Chicken	280	–
Lysozyme	White egg	280	36 800
Oxyhaemoglobin	Mammalian blood	275	35 500
Ribonuclease	Bovine pancreas	280	9100
RNA polymerase	<i>Escherichia coli</i>	280	7400
Trypsin	Bovine	280	35 500
Trypsinogen	Bovine	280	–

Reproduced with permission from Pineda M and Cárdenas J (1988) *Espectroscopía Ultravioleta–Visible de Compuestos Biológicos*. Córdoba (Spain): Publications of Monte de Piedad and Caja de Ahorros of Córdoba.

445 and 460 nm). They act as prosthetic groups in flavoproteins and flavoenzymes. Tetrapyrroles and their derivatives (porphyrins, haemes, chlorophylls, and bile pigments), all of which play major biological roles, absorb strongly in the near-UV and visible regions. Porphyrins absorb very strongly in the region from 450 to 650 nm. Their bands are highly sensitive to environmental factors and to metal ions and ligands; thus, the visible spectra of oxyhaemoglobin, deoxyhaemoglobin, and carboxyhaemoglobin are rather different from one another, which allows their selective quantification. The state of oxidation of iron in ferredoxin also influences its absorption. The spectra of the bile pigments bilirubin ( $\lambda_{\max} = 425$  nm) and biliverdin ( $\lambda_{\text{em}} = 378$  and 680 nm) are also rather different. On the other hand, carotenoids (terpenoids consisting of eight isoprene units with several conjugate bonds) can be of two types, xanthophylls and carotenes, which give absorbance bands in the 400–550 nm region.

The absorbance spectra of proteins are not the sum of those of its constituent amino acids and prosthetic groups, since the protein structure may undergo a number of interactions (hydrogen bonding, ionic interactions between charged or polarized groups) that prevent a detailed analysis of protein structure from UV–visible spectra alone. The problem is further complicated by the influence of the solvent used. All these factors diminish the qualitative value of the spectra. However, direct protein absorbance is very often used for quantitative purposes. Only if the absorbance is very weak or nonspecific does the protein need to be derivatized for quantification.

The most frequently used procedure for direct quantification of proteins is that of Warburg and Christian. It is based on the fact that tryptophan and tyrosine absorb UV radiation maximally at 280 nm and occur at fairly constant concentrations in proteins. Accordingly, the protein concentration of a solution can be determined by measuring its absorbance at 280 nm ( $A_{280}$ ). However, this procedure requires further corrections since nucleic acids and their nitrogen bases, either in free form or as nucleotides, also absorb in the measuring region ( $\lambda_{\max} = 260$  nm). This compels one to determine experimentally the  $A_{280}/A_{260}$  ratio for several samples containing different proportions of nucleic acid in order to calculate a correction factor that is multiplied by  $A_{280}$  to obtain the protein concentration.

### Derivatization

There are three prime methods for quantification of proteins based on derivatization reactions: the biuret, Lowry, and Bradford methods.

The biuret method is based on the fact that proteins (and, as a rule, all substances containing two or more peptidic bonds) react with copper to form a colored complex whose absorption ( $\lambda_{\max} = 454$  nm), in the presence of excess copper, is proportional to the amount of protein present. The reagent is obtained by dissolving 1–5 g of copper(II) sulfate and 6 g of sodium potassium tartrate tetrahydrate in 3% sodium hydroxide. Bovine serum albumin is used as standard. The most serious drawback of this method is its poor sensitivity.

The Lowry method, more sensitive than the biuret method, affords the determination of protein at the microgram per milliliter level. The procedure involves two reactions: that of the protein with alkaline copper solution (the biuret reaction) and reduction of the Folin–Ciocalteu reagent by tyrosine and tryptophan residues of the protein. The Folin–Ciocalteu reagent is prepared by boiling a solution containing sodium tungstate, sodium molybdate, and phosphoric and hydrochloric acids, which produces molybdophosphate and tungstophosphate ions. The absorbance is measured at 750 nm. The Lowry method has also been applied to proteins in whole cells. Like the biuret method, it is subject to some limitations and interferences (particularly those from ammonium salts, glycine, and mercaptans). In addition, variations in the tyrosine and tryptophan contents from protein may introduce some uncertainty in the determinations, so the method is more practical for determining concentration changes than absolute protein concentrations. The Lowry method has also been applied by exploiting solid-phase spectrophotometric techniques.

The Bradford method is based on the absorbance of Coomassie Brilliant Blue on the protein to be determined, which results in a spectrum shift of the dye from 465 to 595 nm. The absorbance at 595 nm is proportional to the protein concentration. Bovine serum albumin is used as standard. The reaction is susceptible to interferences from surfactants and alkalinity of the solution. The main advantage of the method is rapidity, which enables efficient application to many samples.

A micromethod has been developed using the Bradford method for studying the proteins present in cell membranes.

Proteins can also be determined by 1% bicinchoninic acid solution in alkaline medium, to produce a red compound. The absorbance is measured at 562 nm against a reagent blank. Proteins can be determined in the interval  $0.5\text{--}10\ \mu\text{g ml}^{-1}$ . The method offers a one-step alternative to the methods of Lowry and Bradford and is less subject to interferences.

In recent years, diverse chelating agents have been introduced for proteins quantification: Arsenazo III, fuchsin acid, methyl blue, bromophenol blue, methylene blue, etc.

Another common determination is that of haemoglobin, a major component of red blood cells. The method involves oxidation of haemoglobin with hexacyanoferrate(III) to form methaemoglobin, and further reaction with cyanide ion to form cyanmethaemoglobin that presents maximal absorptivity at 540 nm ( $\epsilon = 4.4 \times 10^4 \text{ l mol}^{-1} \text{ cm}^{-1}$ ). The procedure is carried out by using Drabkin's reagent, a solution containing potassium hexacyanoferrate(III), potassium cyanide, sodium phosphate, and a surfactant (e.g., Triton X-100). The method can also be applied to whole blood samples.

Haemoglobin can also be determined after complex formation with inositol hexaphosphate that is quantified by differential spectrometry to avoid the strong absorbance of the protein.

## Nucleic Acids

Nucleic acids absorb strongly in the near-UV region because of the purine and pyrimidine bases they contain. However, the energies of the strong transitions for the five most common bases are virtually the same, which results in very close absorbance maxima (260, 246, 259, 265, and 267 nm for adenine, guanine, uracil, thymine and cytosine, respectively), even though their exact position is pH dependent. This makes it rather difficult to assess the contribution of each band to the spectra of DNA or RNA, which exhibit a broad absorbance band between 240 and 280 nm due to the bases. The spectrophotometric quantification of nucleic acids is also hindered by marked variations in the spectra with pH as a result of the ionization of their constituent bases.

Interactions between nucleic acids and proteins are also hard to elucidate because of the difficulty involved in determining the individual contribution of each substance: protein bands are usually masked by the strong absorbance bands of nucleic acids.

Nucleic acids can also be quantified by derivatization. Thus, RNA is subjected to a multistep procedure in which the nucleic acids are decomposed to pentoses, which are subsequently converted into furfural by heating in an acidic medium. The furfural in turn reacts with orcinol to form a bluish-green condensation product with  $\lambda_{\text{max}} = 660 \text{ nm}$ . The deoxyribose of DNA is quantified with diphenylamine in an acid medium. The linear determination range is from 10 to 100  $\mu\text{g ml}^{-1}$  and the procedure is free from the interference of RNA.

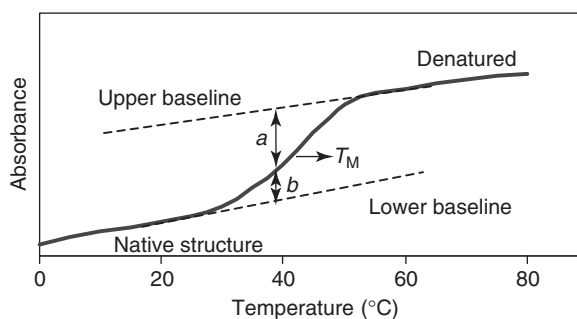
## Study of DNA Stability

Knowledge of the thermodynamics of DNA hybridization and secondary structure formation is necessary for understanding DNA replication fidelity, mismatch repair efficiency, and the mechanism of DNA triplet repeat diseases. In addition, RNA folding thermodynamics is an important aspect to understand ribozyme catalysis, as well as the regulation of protein expression, mRNA stability, and the mechanism of proteins synthesis by the ribosome. With the genome sequencing era upon us, to predict the folding and hybridization thermodynamics of DNA and RNA will be important, so that accurate diagnostics tests for genetic and infectious diseases can be developed.

For example, the temperature-induced transition between native and random coil states of a nucleic acid can be conveniently monitored by UV spectrophotometry. The reason for this is that stacked bases have a lower absorbance per base than unstacked bases; this is called hypochromicity, which is defined as:

$$\% \text{hypochromicity} = 100 \frac{A_{\text{denatured}} - A_{\text{native}}}{A_{\text{native}}}$$

where  $A_{\text{denatured}}$  and  $A_{\text{native}}$  are the absorbances at high and low temperatures, respectively. The absorbance versus temperature profile is commonly referred to as a UV absorbance melting curve (Figure 2). As the temperature increases, the ratio of molecules in the single stranded versus native states increases, resulting in an increase in the UV absorbance. The melting temperature,  $T_M$ , is defined as the temperature at which half of the strands are in the native state and half are in the 'random coil' state. Whereas many methods such as circular dichroism and nuclear magnetic resonance can be used to monitor



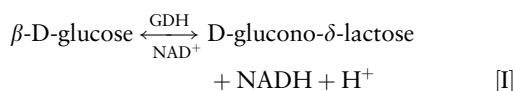
**Figure 2** Typical experiment UV melting profile. At a given temperature, the fraction of strands in the duplex state,  $\alpha$ , is given by the ratio  $a/(a+b)$ , where  $a$  and  $b$  are the respective vertical distances from the upper and lower baselines to the experimental melting curve. (Reproduced with permission from Gore MA (ed.) (2000) *Spectrophotometry and Spectrofluorimetry: A Practical approach*, p. 330. Oxford: Oxford University Press.)

thermal denaturation, UV absorbance is the most sensitive due to the high molar absorptivity of the bases.

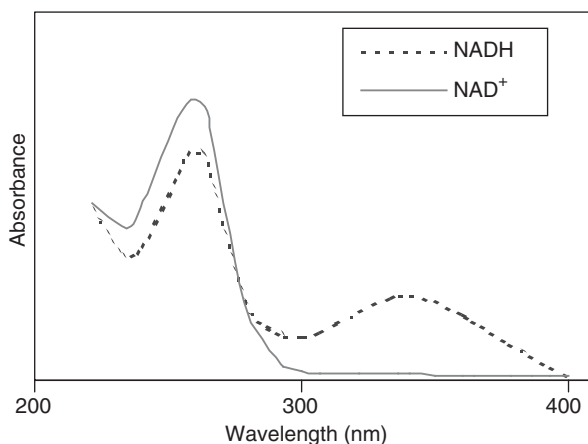
## Enzymatic Analysis

Enzyme analyses are common in clinical and biochemical laboratories, which exploit mostly UV-visible spectrophotometry for this purpose. Basically, the concentration of a substance that takes part in an generic enzymatic reaction  $A \rightarrow B$  can be determined in two ways: (1) by direct determination of the concentration of the reaction product (B) after the reaction is completed; and (2) by measuring the rate of the reaction, which must be related to the concentration of the studied substance (substrate) or the cofactor, activator, or inhibitor. In the former procedure, the reaction should be very fast, whereas the opposite holds true in the latter. In both cases, spectra of the reactant (A) and product (B) should be different. Experimentally, the procedure involves measuring the decrease in the absorbance of A or the increase in that of B. For example, the direct determination of uric acid ( $\lambda_{\text{max}} = 292 \text{ nm}$ ,  $\varepsilon = 1.2 \times 10^4 \text{ l mol}^{-1} \text{ cm}^{-1}$ ) with uricase requires the quantifications of absorbance lessening at 292 nm. In the determination of nicotinamide adenine dinucleotide ( $\text{NAD}^+$ ) with alcohol dehydrogenase, where  $\text{NAD}^+$  is reduced to NADH ( $\lambda_{\text{max}} = 340 \text{ nm}$ ,  $\varepsilon = 6.22 \times 10^3 \text{ l mol}^{-1} \text{ cm}^{-1}$ ), the increase in the absorbance is monitored at 340 nm. If A or B lacks a characteristic spectrum, then one must use coupled reactions in which B acts as substrate for a second enzymatic reaction (indicator reaction) whereby B is quantified. Thus, glucose can be oxidized to gluconic acid by glucose oxidase and the released hydrogen peroxide can be used to oxidize *o*-dianisidine in the presence of peroxidase to obtain a colored product to be spectrophotometrically measured.

Several enzymatic methods use the coenzyme nicotinamide adenine dinucleotide in its oxidized ( $\text{NAD}^+$ ) and reduced (NADH) forms, which makes a very useful reversible system as they are involved in several enzymatic reactions and have rather different absorption spectra (Figure 3). For example, glucose in blood serum can be determined with glucose dehydrogenase according to reaction [I] by monitoring the absorbance increase of NADH at 340 nm:

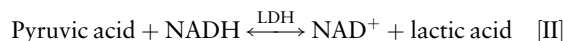


Lactate dehydrogenase (LDH) in serum can be measured by its catalytic action on the reaction

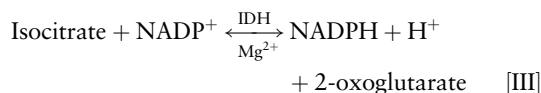


**Figure 3** Absorbance spectra of the coenzyme nicotinamide adenine dinucleotide in its oxidized ( $\text{NAD}^+$ ) and reduced (NADH) forms. (Reproduced with permission from Pineda M and Cárdenas J (1998) *UV-Visible Spectrophotometry of Biological Compounds (Espectroscopía Ultravioleta-Visible de Compuestos Biológicos)*. Córdoba (Spain): Publications of Monte de Piedad and Caja de Ahorros of Córdoba.)

between pyruvic acid and NADH (eqn [II]), which is also monitored by absorption decrease of NADH at 340 nm:



Nicotinamide adenine dinucleotide phosphate, which occurs in oxidized ( $\text{NADP}^+$ ) and reduced (NADPH) forms with markedly different absorption spectra and an absorptivity similar to  $\text{NAD}^+$  can also be used for enzymatic analysis. Thus,  $\text{Mg}^{2+}$  in serum can be measured by the oxidation reaction of isocitrate with  $\text{NADP}^+$ , catalyzed by isocitrate dehydrogenase (IDH) (eqn [III]):



If a fixed concentration of IDH and saturating amounts of isocitrate and  $\text{NADP}^+$  are used, the rate of the enzymatic reaction is proportional to the concentration of magnesium over the range from  $1.0 \times 10^{-6}$  to  $2.0 \times 10^{-4} \text{ mol l}^{-1}$ .

Spectrophotometric methods can be applied to both dissolved and immobilized enzymes. Immobilized enzyme columns are increasingly being used in flow injection analysis as a means of increasing the sample throughput.

## Other Aspects

In the recent past, a certain number of issues related to redox processes in biological systems and the

effect of free radicals have been widely studied. For example, reactive oxygen species (ROS) are produced during normal cellular function. ROS include hydroxyl radicals, superoxide anion, hydrogen peroxide, and nitric oxide. They are very transient species due to their high chemical reactivity that leads to lipid peroxidation and oxidation of DNA and proteins. Under normal conditions, antioxidant systems of the cell minimize the perturbations caused by ROS. When ROS generation is increased to an extent that overcomes the cellular antioxidants, the result is oxidative stress. It is now clear that several biological molecules, which are involved in cell signaling and gene regulation systems, are very sensitive to the redox state of the cell. On the other hand, antioxidants are substances that delay or prevent the oxidation of cellular oxidizable substrates. The prevention of the oxidation is an essential process in all aerobic organisms, as decreased antioxidant protection may lead to cytotoxicity, mutagenicity, and/or carcinogenicity.

In general, the determination of the redox processes and the evaluation of antioxidant capacity can be achieved by electrochemical techniques, such as cyclic voltammetry. On the other hand, as free radicals are very reactive they are usually analyzed by measurements of secondary products ( $\text{H}_2\text{O}_2$ , oxidized proteins, etc.) by fluorimetry in combination with fluorescent dyes, such as dichlorofluorescein. This procedure offers more advantages than the classical spectrophotometric determination based on the oxidation of *o*-dianisidine, due to which the sensitivity of this last procedure is limited to  $\sim 200 \text{ nmol l}^{-1} \text{ H}_2\text{O}_2$  under ideal conditions.

Diverse UV-visible spectrophotometric procedures have been described for the measurement of enzymatic activities in several diseases related to oxidative stress. Thus, catalase activity has been evaluated by measurement of  $\text{H}_2\text{O}_2$ , and glutathione peroxidase (GSPHx) was quantified by continuous photometric monitoring of oxidized glutathione; the conversion of NADPH to NADP was evaluated using UV absorbance at 340 nm. GSPHx activity was calculated after subtraction of the blank value as micromole of NADPH oxidized per minute per milligram protein (International unit per milligram protein) or micromole of NADPH oxidized per minute per gram Hb (International unit per gram Hb).

Flow techniques have been proposed as an alternative to conventional procedures. For example, due to the short lifetime of superoxide in aqueous solution, it still has been possible to adapt the useful and

versatile methods of stopped flow spectrophotometry to measure catalysis by superoxide dismutases. The method rapidly mixes superoxide into aqueous solutions containing the enzymes and follows the subsequent decay of superoxide by detecting its UV absorbance to 250 nm.

**See also:** **Derivatization of Analytes. Enzymes:** Immobilized Enzymes. **Flow Injection Analysis:** Principles. **Kinetic Methods:** Principles and Instrumentation. **Nucleic Acids:** Spectroscopic Methods. **Proteins:** Overview. **Spectrophotometry:** Inorganic Compounds; Organic Compounds.

## Further Reading

- Balestieri C, Colonna G, Giovane A, Irace G, and Servillo L (1978) Second derivative spectroscopy of proteins. A method for the quantitative determination in proteins. *European Journal of Biochemistry* 90: 433–440.
- Bergmeyer HU (ed.) (1983) *Methods of Enzymatic Analysis*, vols I and II. Basel: Verlag-Chemie.
- Boyer RF (1986) *Modern Experimental Biochemistry*. Reading MA: Addison-Wesley.
- Capitan Valvey LF, Duque O, Mirón García G, and Checa Moreno R (2001) Determination of protein content using a solid phase spectrophotometric procedure. *Analytica Chimica Acta* 433: 155–163.
- Fell AF (1979) The analysis of aromatic amino acids by second and fourth derivative UV-spectroscopy. *Journal of Pharmaceutics and Pharmacology* 31(suppl.): 1–23.
- Matés JM (2000) Effects of antioxidant enzymes in the molecular control of reactive oxygen species toxicology. *Toxicology* 153: 83–104.
- McKenzie HA and Smythe LE (eds.) (1988) *Quantification Trace Analysis of Biological Materials*. Amsterdam: Elsevier.
- Pineda M and Cárdenas J (1984) Enzymatic methods of analysis (Métodos enzimáticos de análisis). In: Pérez Bendito MD and Valcárcel M (eds.) *Métodos Cinéticos de Análisis*, pp. 113–140. Córdoba: Publications of Monte de Piedad and Caja de Ahorros of Córdoba.
- Sánchez Rojas F, Bosch Ojeda C, and Cano Pavón JM (1988) Derivative ultraviolet-visible region absorption spectrophotometry and its analytical applications. *Talanta* 35: 753–761.
- Silverman DN and Nick HS (2002) Catalytic pathway of manganese superoxide dismutase by direct observation of superoxide. *Methods in Enzymology* 349: 61–74.
- Smith PK, Krohn RI, Hermanson GT, et al. (1985) Measurement of protein using bicinchoninic acid. *Analytical Biochemistry* 150: 76–85.
- Zuo SS and Lundahl P (2000) A micro-Bradford membrane protein assay. *Analytical Biochemistry* 284: 162–164.

## Pharmaceutical Applications

**J Martinez Calatayud**, Universidad de Valencia, Burjassot, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Drug analysis encompasses tests on raw materials (purity criteria), pharmaceutical or veterinary formulations, and a number of other, more complex matrices such as foods of animal origin, drinks, and foodstuffs that are conducted for clinical, forensic, or veterinary purposes involving a variety of matrices including blood, urine, and tissues. Pharmaceutical analysis is restricted to drug analyses in raw materials and pharmaceutical formulations; in some instances, it also includes clinical analyses. Pharmaceutical analyses involve the determination of not only active components but also impurities, excipients, the stability of active components (and their degradation intermediates or end-products), and other parameters such as content uniformity, solubility, and dissolution rate.

Some spectrophotometric analyses focus on the properties of the target compound (its native spectrum), a chemical derivative, or the product of some separation or sample pretreatment. Direct analyses of target compounds are only possible if the wavelength used is not interfered with by other species or by background absorption; analyses of mixtures of components are also possible with recent developments in chemometrics and derivative spectrophotometry.

The first question to be answered is: what is the wavelength that provides the highest sensitivity and selectivity with the best reproducibility? Finding an answer entails deciding on optimal conditions of pH, solvent, and range of application of the Lambert-Beer law.

The situation is basically the same, albeit technically more complex, when some chemical reaction (e.g., derivatization of the analyte) is required prior to measurement.

Chemical pretreatments may be required due to lack of selectivity and/or sensitivity of direct spectrophotometry. The chemistry of spectrophotometric determinations in pharmaceutical analysis is by now well documented; common procedures are: redox reactions; metal ion chelation reactions for the determination of organic compound or metal ion components; charge transfer complex formation reactions; ion pairing; and diazotization.

If the objective is a quantitative determination by direct spectrophotometry, it suffices to read the absorbance at a given wavelength. By virtue of their simplicity, these methods are more precise than those involving chemical pretreatment. Therefore, if the absorption wavelength and intensity of the analyte are both adequate, pharmacopoeias and compilations of standard methods recommend using this approach. The choice between direct spectrophotometry and chemical derivatization is also influenced by the accuracy required in the result, the purity of the target analytes, and the interferences expected. If the sample allows for it, the solid, once powdered, can be extracted with water or another solvent (usually methanol or ethanol) at an appropriate pH, filtered, and diluted to the required concentration before measuring the absorbance.

### Determination of Pharmaceuticals without Chemical Derivatization of the Analyte

Most pharmaceuticals have a native spectrum without high molar absorptivities at any absorbing wavelength, which leads to mostly quantitation procedures using chemical derivatization of the analyte. Recent literature does not propose spectrophotometric methods based on the native spectrum of the drug, because they result in low sensitivity and can be only utilized when high accuracy is not required.

Another question to be considered is selectivity. The drug (even 'pure' bulk drug) may very often be accompanied by compounds with similar chromophores from different sources, namely, residual starting materials, products of synthesis, degradation products, the spectra of which may resemble closely to that of the analyte. Consequently, the selectivity of direct spectrophotometric methods is insufficient. The choice of solvent and the pH should be considered in connection with the sensitivity and selectivity of these procedures.

Examples include the spectrophotometric method developed to determine 1,4-benzodiazepines, diazepam, bromazepam, and clonazepam in pure forms, pharmaceutical preparations, and biological (urine) fluids measure the absorption spectra in methanolic potassium hydroxide solution; and the quality control of tablets containing mirtazapine is based on the measurement in methanol solution. The weighed and powdered sample is sonicated for



10 min and then made up to volume with methanol. To the aliquots of the filtered extract phosphoric acid is added. The same solvent has also been proposed for the estimation of mefloquine hydrochloride.

### Simultaneous Determinations (Derivative Spectrophotometry and Other Chemometric Approaches)

**Derivative spectrophotometry** The theoretical background is well known; however, its recent applicability is due to the availability of spectrophotometers that are able to scan the complete spectrum in a short time and provide data to the microcomputer to generate the corresponding derivative spectrum.

The analytical applications of derivative spectrophotometry focus on two aspects: background correction, and the resolution of overlapping bands for quantitative analysis and avoiding separation or other sample pretreatment steps. Other goals are to eliminate the effect of turbidity in the sample matrix and to improve spectral resolution.

Derivative spectrophotometry is highly effective for the analysis of mixtures, particularly with the quick recording when diode array spectrophotometers are used. The resolution of a binary mixture, when the spectra of the two compounds are different, requires a simple mathematical treatment of the data obtained at two wavelengths. However, drug mixtures having overlapped spectra require other mathematical treatment such as derivative spectrophotometry.

The resolution of binary mixtures of compounds with overlapping spectra by derivative spectrophotometry is frequently made on the basis of zero-crossing measurements, which is based on the measurement of the absolute value of the derivative spectrum of the mixture at a cutoff value (of wavelength) where the intensity of one of the components of the mixture is zero. At this wavelength intensity is directly proportional to the concentration of the other component.

Most studies deal with the simultaneous determination of binary or even ternary mixtures of drugs with overlapped spectra; or they determine one drug in a formulation by avoiding interferences from other compounds. Less frequent are stability studies in which derivative spectrophotometry is used as a tool for the determination of the drug in the presence of its degradation products. Examples of this include the determination of sumatriptan succinate by using first and second derivative spectra applied to bulk powder, laboratory-prepared mixtures, and a pharmaceutical dosage form; and determination of aceclofenac in the presence of its degradation product,

based on the determination of diclofenac by the third derivative spectrophotometry in pharmaceutical formulations. **Table 1** gives many examples of studied mixtures. Two stability-indicating methods have been developed for the determination of doxazosin mesylate and celecoxib in the presence of their degradation products. One method is based on the use of first derivative spectrophotometry in bulk powder, laboratory-prepared mixtures, and pharmaceutical dosage forms. The other, a derivative spectrophotometric method, is based on the simultaneous determination of amlodipine and its pyridine photodegradation product with spectral measurements on the third-order UV derivative spectrum. The method can be applied usefully to routine quality control of pharmaceutical formulations containing amlodipine and lisinopril in commercial dosage forms.

The purity of aztreonam can be maintained through its reaction with cerium(IV) in acidic medium. The spectrophotometric method involves the quantitation of the amount of ceric equivalent to aztreonam by measuring the absorbance at 317 nm and the corresponding first-derivative value at 284 nm for the blank solution against the reaction solution. It can be applied to the determination of aztreonam in pure form and in the presence of arginine both in laboratory mixtures and in commercial formulations. The determination of imipramine in the presence of iminodibenzyl as an impurity uses a first-derivative ratio spectrum and can be applied to pharmaceutical dosage forms.

Many of the procedures used in derivative spectrophotometry are also used in other chemometric procedures when dealing with the same analytical problem. Examples are (first and second mean, the first and the second derivatives, respectively): progesterone (first, by difference absorbance values at 227.2 and 253.6 nm); vasartan (second, at 205.6 nm); indapamide (first at 252.8 and second at 260.4 nm); ibesartan in the presence of hydrochlorothiazide (first at 263.0 nm); reboxetine (fourth after extraction); astemizole (second and zero crossings); cisapride (first at several wavelengths, e.g., 264 and 300 nm; and the second at 276 and 290 nm); glimepiride (second in dimethylformamide over the range 263.3–268.2 nm by peak to peak, the peak to zero at 268.2 nm, and tangent at 263.3–271.8 nm); mirtazapine (first and second, in methanol solutions over the range 225–360 nm); olanzapine (first at 298 nm); thonzylamine hydrochloride (first at 492 nm); nimesulide (second in different solvents and wavelengths); zinc in formulations (fourth after reaction with 1,2-thiazo-naphthol); lansoprazole (second at 200 and 400 nm); miconazole (second and

**Table 1** Simultaneous determination of drug mixtures by derivative spectrophotometry: tested mixtures

Cyproterone acetate, estradiol valerate	Cefotaxime sodium, cefadroxil monohydrate	Zidovudine, lamivudine	Piroxicam, tenoxicam	Theophylline, ephedrine
Amiloride, furosemide	Omeprazole, lansoprazole, pantoprazole	Hydrochlorothiazide, amiloride	Dorzolamide hydrochloride, timolol maleate Menadione, menadione sodium	Chlorphenoxamine hydrochloride, caffeine
Montelukast, loratadine	Trimethoprim, sulfamethoxazole	Mefenamic, paracetamol	Hydrocortisone, nystatin, oxytetracycline	
Terebinafine hydrochloride, triamcinolone acetonide	Dorzolamide hydrochloride, timolol maleate	Valsartan, hydrochlorothiazide	Dexamethasone, polymixin B, trimethoprim	Benazapril hydrochloride, hydrochlorothiazide
Fosinopril, hydrochlorothiazide	Perindopril, indapamide	Lisinopril, hydrochlorothiazide	Losartan potassium, hydrochlorothiazide	Paracetamol, methocarbamol
Benzocaine, cetylpyridinium	Chlordiazepoxide, clidinium bromide	Mephenoaloxone, paracetamol	Hydrochlorothiazide, amiloride hydrochloride	Epinephrine, norepinephrine
Chlorpheniramine maleate, phenylephrine hydrochloride	Epinephrine hydrochloride, theophylline	Thiamine hydrochloride, pyridoxine hydrochloride	Pseudoephedrine sulfate, dexbrompheniramine maleate, loratadine	Hydrochlorothiazide, benazepril hydrochloride, triamterene, cilazapril <sup>a</sup>
Mebeverine hydrochloride, sulfiride	Rimpanficin, isoniazid	Chlorpheniramine maleate, phenylephrine hydrochloride	Acrivastine, pseudoephedrine hydrochloride	Nifedipine, acebutolol hydrochloride
Benazapril hydrochloride, hydrochlorothiazide	Ampicillin sodium, sulbactam	Paracetamol, codeine	Cephalothin, cefoxitin	

<sup>a</sup> Binary mixtures of hydrochlorothiazide with each other.

after extraction); and, finally, haloperidol in the presence of parabens (first and the zero crossings).

**Chemometric procedures** Simultaneous spectrophotometric determinations on binary or ternary mixtures of drugs in a formulation, and avoiding separation methods and other sample pretreatments, can also be achieved by using chemometric-assisted spectrophotometric procedures other than derivative spectrophotometry. Analyzing synthetic mixtures containing title drugs by this approach has been validated using liquid chromatographic or capillary electrophoresis. The effort devoted to other chemometric methods is less than that devoted to derivative spectrophotometry. Tablets and sugar-coated pills are the formulations generally analyzed.

Studies on binary mixture samples frequently deal with classical least-squares, inverse least-squares, principal component regression and partial least-squares methods. These methods have been used for resolving mixtures of hydrochlorothiazide and spirinolactone in tablets; cyproterone acetate and estradiol valerate; amiloride and hydrochlorothiazide;

mefenamic acid and paracetamol; chlorphenoxamine hydrochloride and caffeine; benazepril hydrochloride and hydrochlorothiazide.

Mixtures of acetylsalicylic acid and ascorbic acid have been studied by using parallel factor analysis and partial least-squares. The former is used for spectral deconvolution, and  $pK_a$  estimation for both acids. The simultaneous determination of fosinopril and hydrochlorothiazide in pharmaceutical formulations consists of extracting both compounds in an aqueous solution, measuring by multiwavelength UV spectrophotometry; hydrochlorothiazide acts as an internal standard to verify the accuracy of the analysis.

Chemometrics has also been applied to solve other problems but not for the quantification of binary mixtures. Principal component analysis (PCA) has been used to plot dissolution curves and provide information about between- and within-batch variations. Differences in level or shape can be observed in the first two principal components (PCs). Irrelevant irregularities, which have a strong influence on the similarity factor, are resolved.

A kinetic spectrophotometric method has been developed for the simultaneous quantitative determination of acetaminophen and phenobarbital in pharmaceutical preparations. The basis of the method was the different kinetic rates of the analytes in a two-step chemical procedure: first the oxidative coupling reaction with 3-methylbenzothiazolin-2-one hydrazone in hydrochloric acid medium and then using Fe(III) as oxidant and continuous absorbance monitoring. An artificial neural network (ANN) coupled with PCA (PC-ANN) has been used to simulate this method.

Spectrophotometric monitoring with the aid of chemometrics has also been applied to more complex mixtures. To solve the mixtures of corticosteroid dexamethasone sodium phosphate and vitamins B<sub>6</sub> and B<sub>12</sub>, the method involves multivariate calibration with the aid of partial least-squares regression. The model is evaluated by cross-validation on a number of synthetic mixtures. The compensation method and orthogonal function and difference spectrophotometry are applied to the direct determination of omeprazole, lansoprazole, and pantoprazole in gastroresistant formulations. Inverse least squares and PCA techniques are proposed for the spectrophotometric analyses of metamizol, acetaminophen, and caffeine, without prior separation. Ternary and quaternary mixtures have also been solved using iterative algorithms.

## Determination Methods with a Prior Chemical Pretreatment

Chemical pretreatment is used to improve the analytical figures of merit with special emphasis on sensitivity and reproducibility. The spectrophotometric determination of pharmaceuticals relies on well-known reaction methodologies such as oxidation of the drug and the formation of metal complexes.

### Redox

Common strong oxidants are used to obtain colored products by acting directly on the pharmaceutical, e.g., Ce(IV), Cr(VI), metavanadate, BrO<sub>3</sub><sup>-</sup>, *N*-bromosuccinimide, and chloramine-T. Direct oxidation of imipramine and desipramine by ammonium metavanadate allows its determination in pharmaceutical dosage forms. The quantitative estimation of 11 phenothiazine drugs in dosage forms is also based on the interaction of these phenothiazine compounds with diphenylamine in presence of *N*-bromosuccinimide and sulfuric acid. Most of the studied phenothiazines yield bluish green products and the color is stable for at least 1 h.

The oxidation of nizatidine is used to develop a kinetic method for the determination of this drug in capsules. The reaction is with ammonium cerium(IV) sulfate in the presence of perchloric acid and the subsequent measurement of absorbance at 314 nm.

Sometimes the proposed method is an 'indirect' procedure to monitor the excess of oxidant or the color change of an auxiliary reagent such as the method described for the assay of azathioprine, either in pure form or in pharmaceutical formulations; the methods are based on the oxidation of the drug with excess *N*-bromosuccinimide or chloramine-T and determining the consumed reagent by the decrease in color intensity of celestine blue or gall-o-cyanine, respectively. Another example is the determination of ascorbic acid in fruit juice and pharmaceuticals on the basis of its inhibition effect on the reaction between hydrochloric acid and bromate. The decolorization of methyl orange due to the reaction products was used to monitor the reaction at 510 nm.

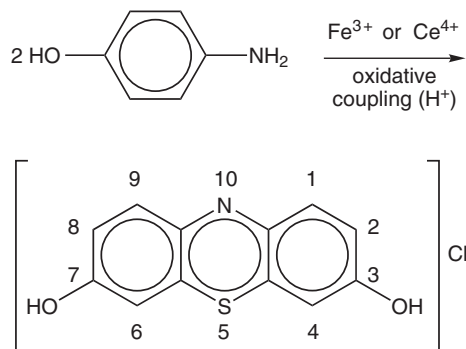
The spectrophotometric determination of the following antibiotics has been performed based on the reactivity with iodine: ampicillin, penicillin V, amoxycillin, cloxacillin, cefadroxil, ceftezoxime, griseofulvin, streptomycin, nicoumalone, and acebutolol hydrochloride. An excess of iodine of known concentration is added to the drug alkaline solution and the unreacted iodine monitored through the decrease in the absorbance of the dye wool fast blue at 540 nm. The determination of pentoxifylline is also based on adding a known excess of standard iodine solution under alkaline conditions and the excess of iodine determined at pH 3.0 with metol-isoniazid (620 nm) or also with wool fast blue (540 nm).

The oxidation of the phenothiazines by a known excess of chromium(VI) and subsequent determination of the unreacted oxidant by interacting with metol and sulfanilic acid resulted in colored species that exhibit maximum absorbance at 530 nm.

In some cases, the redox reaction is one step in the whole determination process. The well-known reductive behavior of ascorbic acid on Fe(III) results in Fe(II) ions that form a chelate with *o*-phenanthroline. This chelate is easily adsorbed on an anion-exchange gel of dextran type. The adsorbent is packed into the spectrophotometer cell and the absorbance of the resin monitored at 567 and 800 nm. The ascorbic acid is determined without any preconcentration in fruit juices, pharmaceutical formulations, human urine, and conservative liquids.

Different oxidizing agents can be tried, e.g., Ce(IV) and Fe(III), for the determination of *p*-aminophenol and acetaminophen in dosage forms and without prior separation. The method is based on the

reaction of *p*-aminophenol at room temperature with sodium sulfide in presence of the oxidant to produce a methylene blue-like dye. The color develops within 10 min and remains stable for at least 3 h.



Spectrophotometric methods for the determination of catecholamine derivatives pyrocatechol, dopamine, levodopa, and methyl dopa have been developed and applied to injections and tablets. One of the methods involves the oxidation of *o*-dihydroxybenzene derivatives by *N*-bromosuccinimide followed by oxidative coupling with isoniazide, leading to the formation of a red-colored product of maximum absorbance at 480–490 nm.

A spectrophotometric procedure for the determination of phenothiazines in pure form and in some pharmaceutical dosage formulations was based on the oxidation of the drug by a known excess of potassium dichromate. The unreacted potassium dichromate was estimated by reacting with excess of Fe(II) and complexing the released Fe(III) with thiocyanate.

### Paired Ionic Compounds

The formation of ion pairs between two ions of relatively large size and low charge – one of which will be the active principle – provides water-insoluble products. The experimental procedure is very simple, but is also slow as the insoluble compounds formed must be extracted with an organic solvent for their spectrophotometric determination. Ion-pairing reactions also provide substantially improved sensitivity and selectivity. Most reported methods of this type use organic counterions to avoid the extraction step. Table 2 gives a representative list of examples.

### Metallic Ion Chelates

These reactions are fast and provide simple experimental procedures, since, unlike ion pairing, they yield water-soluble products. In some cases (e.g., with ternary complexes), the products are made soluble with the aid of a surfactant. However, the

process is relatively simple and highly sensitive; molar absorptivities are adequate for the analysis of pharmaceutical formulations.

**Formation of metallic ion chelates for the determination of organic active principles** Some recent illustrative examples are as follows:

1. The determination in dosage forms of piperazine derivatives, ketoconazole, trimetazidine hydrochloride, and piribedil, by the formation of yellow orange complexes between iron(III) chloride and the drug. The method allows ranges of determination over 1–15, 1–12, and 1–12 mg l<sup>-1</sup>, respectively, and has been applied to different dosage forms.
2. The determination of amoxicilline, ciprofloxacin, and piroxicam in bulk and pharmaceutical preparations has been performed by combining the oxidation character of Fe(III) with the chelation property of Fe(II), affording highly colored complexes. The reagent solutions are: (1) a mixture of 1,10-phenanthroline and iron(III) ammonium sulfate in hydrochloride medium; and (2) 2,2'-bipyridyl and iron(III) ammonium sulfate in hydrochloride medium. The resulting Fe(II) concentration, proportional to the amount of drug present, is monitored by the measurement of absorbances of tris(*o*-phenanthroline) iron(II) or tris (bipyridyl) iron(II) complexes at 510 and 522 nm, respectively.
3. The formation of a chelate compound of strontium with bromopyrogallol red and cetylpyridinium chloride serves to develop a method for the spectrophotometric determination of cetylpyridinium chloride in pharmaceutical products. This indirect procedure is based on the influence of micellar media on the absorption spectrum of the strontium(II)–bromopyrogallol red complex; the formation of the ternary complex is accompanied by an observed increase in absorbance and a bathochromic shift in the maximal absorption of the complex from 555 to 627.5 nm. The optimum pH range for the reaction is 4.0–5.0 and Beer's law is obeyed over the concentration range 0.01–0.07 mg l<sup>-1</sup>.
4. The determination of two fluoroquinolone antibacterials (ciprofloxacin and norfloxacin), either in pure form or in tablets, is possible on the basis of the formation of a ternary complex palladium(II)–eosin–fluoroquinolone in the presence of methyl cellulose as surfactant. The ternary complexes show an absorption maximum at 545 nm and obey Beer's law in the concentration range 3–10 mg l<sup>-1</sup> for both quinolones.
5. The photometric reagent 2-(5-bromo-2-pyridylazo)-5-diethylaminophenol is used for the determination of zirconium; when fosfomycin is added, it

**Table 2** Formation of paired ionic compounds

<i>Drug</i>	<i>Reagent(s)</i>	<i>Maximum absorption (nm)</i>	<i>Stoichiometry, medium</i>
Nizatidine	Bromocresol purple	411	
	Picric acid	400	
Flucloxacillin	<i>p</i> -Nitrophenol	446	1:1; pH, 9.0
	2,4-Dinitrophenol	435	
	3,5-Dinitrosalicylic acid	442	
	Picramic acid	473	
	Picric acid	439	
Vitamin B <sub>1</sub>	Thymol blue, bromothymol blue, bromophenol blue, bromocresol green, phenol red, cresol red	Over the range 420–450	1:1; solubilization agents: polyvinyl alcohol, emulgent OP, Triton X-100, Tween-20
Piroxicam, tenoxicam	Alizarin		
	Alizarin red S		
	Alizarin yellow G		
	Quinalizarin		
Enoxacin	Bromophenol blue	412	Extract. <sup>a</sup> Chloroform
	Bromocresol purple	410	
Trazodone	Bromophenol blue	414	pH 3.4; extract. Chloroform
Ampicillin, dicloxacillin, flucloxacillin, amoxicillin	Mo(V)–thiocyanate		Extract. methylene chloride
Fluoxetine hydr. <sup>b</sup>	Methyl orange	433	pH 4.0
	Thymol blue	410	pH 8.0
Cisapride	Suprachen Violet 3B	595	Extract. Chloroform
	Erioglaucine A	640	
	Naphthalene Blue 12 BR	620	
	Tropaeolin 000	500	
Sparfloxacin	Bromothymol blue	385	
Cetirizine hydr.	Bromocresol purple	409	pH 2.64; extract. Chloroform
	Bromophenol blue	414	
Diltiazem hydr.	Bromothymol blue, bromophenol blue, bromocresol green	415	Acidic medium; extract. Chloroform
Penicillamine	2,6-Dichloroquinone-4-chlorimide	431	Extract. Dimethylsulfoxide
Ampicillin	Pyrocatechol violet	604	
Amoxycillin		641	
6-Minopenicillanic acid		645	
Cloxacillin		604	
Dicloxacillin		649	
Na-flucloxacillin		641	
Amineptine hydr., piribedil, trimebutine maleate	Bi(III)–iodide		pH 2.0–2.8; extract. 1,2-dichloroethane
Cysteine, cystine	Triiodide ion and hexadecylpyridinium	500	
Chloroquine, pyrimethamine	Molybdenum(V)–thiocyanate		Extract. methylene chloride

<sup>a</sup>Extract., extractive organic solvent.<sup>b</sup>Hydr., hydrochloride.

replaces diethylaminophenol by complexing with zirconium. This forms the basis for an indirect spectrophotometric method for fosfomicyn based on ligand exchange and is applied to dosage forms. The detection wavelength is at 605 nm, and the apparent molar absorption coefficient is found to be  $4.59 \times 10^4 \text{ l mol}^{-1} \text{ cm}^{-1}$ .

**Determination of metallic ions in pharmaceuticals** Spectrophotometric determination of a metal is present as an active component in dosage forms; it is most suitable for impurities, using atomic

absorption spectrometry for detection. However, the 'classical' formation of a metallic complex with an organic compound, frequently aided by liquid–liquid extraction or the presence of tensoactives, is also used for drugs in pharmaceutical formulations and biological samples. Some examples are given in Table 3.

**Diazotization** A procedure for amine compounds diazotization followed by a coupling reaction of the diazonium ions and phenols to produce strongly colored azo compounds ( $\text{ArN}=\text{NAr}'$ ) by electrophilic substitution. The azo link  $-\text{N}=\text{N}-$  gives a

**Table 3** Formation of metal chelates with the metal active ingredient

<i>Metal</i>	<i>Reagents</i>	<i>Method, stoichiometry</i>	<i>Samples</i>
Zn	Formazans (1,5-diphenyl-3-acetylformazan, 1-( <i>o</i> -carboxyphenyl)-3-acetyl-5-acetylformazan, 1-( <i>o</i> -carboxyphenyl)-3-acetyl-5-phenylformazan, and 1-( <i>o</i> -carboxyphenyl)-3-acetyl-5- <i>m</i> -tolylformazan); pH 8.5, 7.5, 5.5, and 6.5, respectively	Surfactant cetylpyridinium chloride; 1:1:2	Serum, human hair, phar. form. <sup>a</sup>
Zn	Azo-dye derivatives (2-(2',4'-dihydroxyphenylazo-1')benzimidazole, 3-mercapto-5-(2'-hydroxynaphthylazo-1')-1,2,4-triazole and 3-(5'-mercapto-1',2',4'-triazole-3'-azo)-2,4-dihydroxybenzoic acid)		Multimineral form. <sup>e</sup>
Zn	Xylenol orange; pH 5–6; 580 nm	Surfactant cetylpyridinium chloride; 1:2:4	Dermal ointments
Co(II) and Bi(III)	Isonitroso-5-methyl-2-hexanone	Extrac. <sup>b</sup> in chloroform	Synthetic mixtures, phar. form. biolog. <sup>c</sup> high speed steel
Bi(III)	Hydrogen tetraiodobismuthate(III); 482 nm	Extrac. propylene carbonate	Phar. form.
Bi(III)	Protriptylinium tetraiodobismuthate(III) into chloroform		Phar. form.
Bi(III)	Tetrabutylammonium tetraiodobismuthate(III); 485 nm	Adsorptive extrac. microcrystalline benzophenone and dissolution in ethyl acetate	Phar. form.
Fe(II) or Fe(III)	Iron(III)/azide complexes; Fe(II) oxidized; 396 nm	Tetrahydrofuran/water medium	Anemia form.
Fe(II)	Azo-dye derivatives of heteroazopyrocatechine (2-(3,4-dihydroxyphenylazo-1)-benzimidazole (BIAP); 3-mercapto-5-2-(3,4-dihydroxyphenylazo)-1,2,4-triazole and 2-carboxymethanotio-5-(3,4-dihydroxyphenylazo-1)-1,3,4-tiadiazole)		Multivitamin prep. <sup>d</sup>
Al(III)	1,2,4-Triazole series azo dyes (pyrocatechine: 3-(3',4'-dihydroxyphenylazo-1')-1,2,4-triazole and 3-(3',4'-dihydroxyphenylazo-1')-5-mercapto-1,2,4-triazole)	Aqueous-methanolic media; 2:1 and 3:1	Phar. <sup>f</sup> form.

<sup>a</sup>On different pharmaceutical formulations.<sup>b</sup>Extrac., extractive organic solvent.<sup>c</sup>Biolog., biological samples.<sup>d</sup>Prep., preparation.<sup>e</sup>Form., formulation.<sup>f</sup>Phar., pharmaceutical.

conjugation of aromatic rings resulting in a large system of  $\pi$  electrons, allowing light absorption into the visible region. The kinetic stability of resulting solutions is high.

The determination of flutamide in either pure form or in its pharmaceutical preparations can be performed with the aid of two different processes: (1) the diazotization of reduced flutamide, followed by coupling with alcoholic iminodibenzyl in acid medium to give a purple-colored product having a maximum at 570 nm; or (2) diazotization of the reduced flutamide and coupling it with 4-amino-5-hydroxy-2,7-naphthalenedisulfonic acid monosodium salt in a buffer medium of pH 12, which gives a red-colored

product presenting a maximum absorbance at 520 nm.

The estimation of catechol and its derivatives such as dopamine hydrochloride, levodopa, methyl dopa, and adrenaline hydrochloride in both pure form and in pharmaceutical formulation is based on the interaction of diazotized sulfanilamide with catechol derivatives in the presence of molybdate ions in acidic medium. Absorbance of the resulting red-colored product is measured at 490 nm for pyrocatechol and at 500 nm for other catechol derivatives. The color reaction is stable for 24–30 h.

The determination of nimesulide also requires the reduction of the group  $\text{NO}_2$  and then diazotization



followed by either coupling with alcoholic iminodibenzyl in acid medium to give a deep blue colored product (600 nm) or coupling with 3-aminophenol in acid medium to produce an orange red colored product with maximum absorbance at 470 nm. This method can be applied either to the pure form or pharmaceutical formulations.

The spectrophotometric determination of folic acid, either in pure form or in its pharmaceutical preparations, is based on the probable diazotization of *p*-aminobenzoylglutamic acid obtained after reductive cleavage of folic acid, followed by either coupling with iminodibenzyl (maximum absorbance at 580 nm) or with 3-aminophenol to produce an orange yellow colored product (460 nm).

The azathioprine is an immunosuppressive antimetabolite drug containing also a nitro group and its diazotization with excess nitrous acid also required the preliminary reduction of the drug. The spectrophotometric monitoring of either the consumed nitrous acid ( $\text{HNO}_2$ ) with cresyl fast violet acetate or by coupling reaction of the diazonium salt formed with *N*-1-naphthylethylenediamine dihydrochloride has been proposed. The reaction shown in Figure 1 depicts the proposed analytical procedure.

A spectrophotometric method for some sulfa drugs starts with the formation of an orange yellow colored azo product by the diazotization of sulfonamides and is followed by a coupling reaction with 3-aminophenol in aqueous medium. The absorbance of the resulting orange yellow azo product is measured at 460 nm and the product is stable for 6 days at 27°C. The method is successfully used for the determination in various pharmaceutical preparations of the sulfonamides dapsons, sulfathiazole, sulfadiazine,

sulfacetamide, sulfamethoxazole, sulfamerazine, sulfaguanidine, and sulfadimidine. The proposed reaction mechanism is depicted in Figure 2.

Another method for the determination of sulfa drugs is based on the formation of a violet-colored azo product by the diazotization of sulfonamides, followed by a coupling reaction with iminodibenzyl in alcohol medium. Absorbance of the resulting violet azo product is measured at 570–580 nm and is stable for 24 h at 27°C. Beer's law is obeyed in the concentration range of 0.05–6.0 ppm at the wavelength of maximum absorption. The method is successfully employed for the determination of sulfathiazole, sulfadiazine, sulfacetamide, sulfamethoxazole, sulfamerazine, sulfaguanidine, and sulfadimidine. The proposed reaction mechanism for the formation of the violet azo dye is shown in Figure 3.

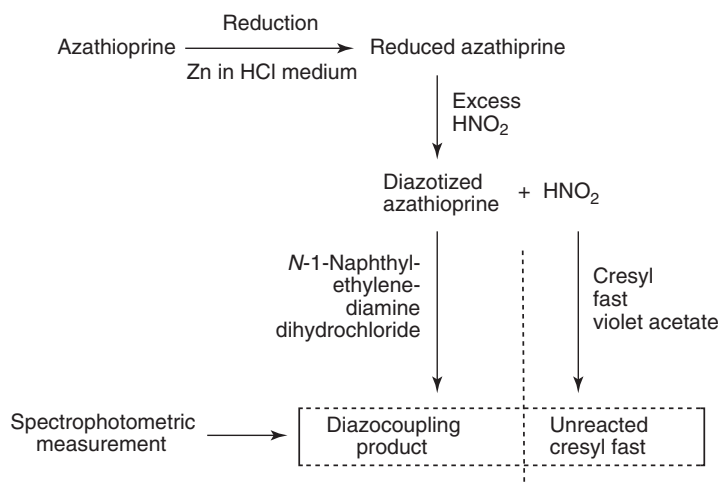
### Charge Transfer Complexes

Charge transfer complexes have high molar absorptivities and are produced when one reagent is an electron donor and the other is an electron acceptor. Light absorption is based on an electronic transition from the donor to an associated orbital of the acceptor. As a consequence, the excited electronic state is an internal redox process.

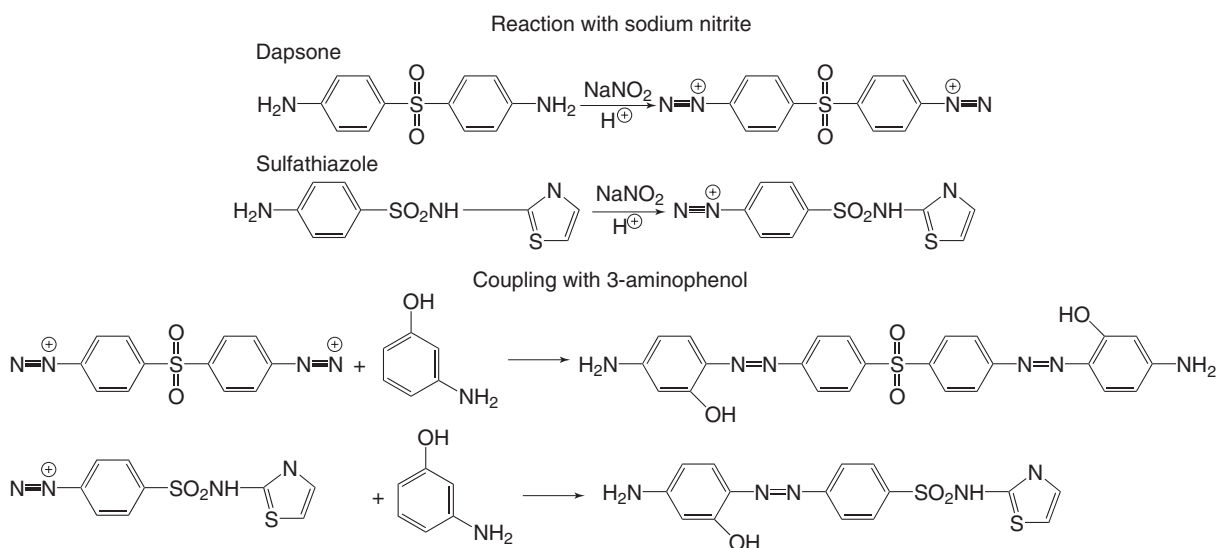
The active pharmaceutical ingredient should be the donor and the analytical challenge is to find a suitable acceptor. The resulting reactions is very quick and the compound finally formed is dependent on the polarity of the solvent (Table 4).

### Other Chemical Derivatization Procedures

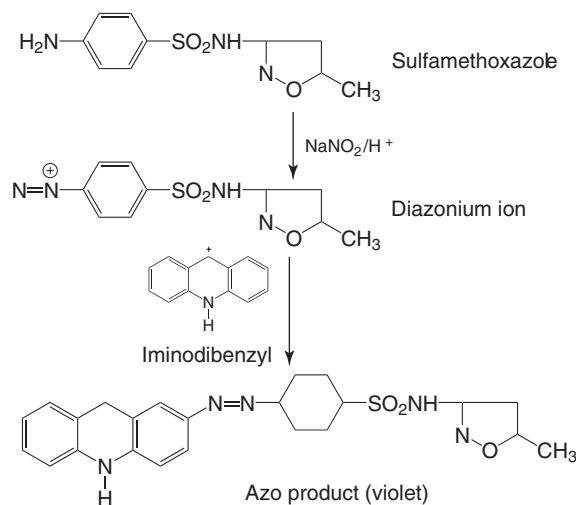
Procedures not included in the preceding sections and having a potential for derivatization are the use



**Figure 1** Analytical procedure for diazotization.



**Figure 2** Reactions in diazotization of sulfonamides.



**Figure 3** Reaction depicting the formation of violet azo dye.

of the light (more usual in dynamic flow methods) and classical titrimetric procedures.

The use of light as a 'reagent' is advantageous as regards the reproducibility of the process, the 'selectivity' of the reagent (irradiation wavelength), and its 'sensitivity' (lamp power). As a result, the photodegradation of organic compounds to photo-fragments gives improved spectrophotometric properties.

The irradiation of some sulfa compounds – such as sulfacetamide, sulfadiazine, sulfaguanidine, sulfamerazine, sulfamethoxazole, and sulfamethizole – gives an absorption band with the maximum at 332 nm.

With regards to spectrophotometric titrimetry, the chloride (hydrochloride) content in a drug can be determined by titration with mercuric nitrate, using a diphenylcarbazone–Bromothymol Blue mixture as indicator. For its spectrophotometric determination, a solution of the Hg(II)–diphenylcarbazone complex is titrated with the unknown sample (pharmaceutical formulation) and the absorbance of the resulting complex at 540 nm – which is decreased as a result of the diphenylcarbazone being displaced by the chloride present in the sample – is monitored.

### Structure–Spectra Correlation (Qualitative Analysis)

If the objective is identification (qualitative analysis), it suffices to compare the spectrum of the analyte with that of a standard, both recorded in the same solvent and at an identical pH. This is not the main application of UV–Vis spectrophotometry as the best results in this context are provided by spectroscopic methods considered more effective for the study of the molecular structure of organic compounds (infrared, nuclear magnetic resonance, mass spectrometry, and X-ray diffraction). However, UV–Vis spectrophotometry is a source of relevant supplementary information that helps in the elucidation of molecular structures of drugs, impurities, metabolites, intermediate compounds of degradation, etc.

UV–Vis absorption is essentially a classical quantitative tool for organic compounds, but a good level

**Table 4** Formation of charge transference complexes

<i>Drug</i>	<i>Reagent (acceptors)</i>
Piperazine derivatives; ketoconazole, piribedil and prazosin hydrochloride	2,3-Dichloro-5,6-dicyano- <i>p</i> -benzoquinone in acetonitrile
Amlodipine besylate	2,3-Dichloro 5,6-dicyano 1,4-benzoquinone
Cephapirin sodium, cefazoline sodium, cephalixin monohydrate, cefadroxil monohydrate, cefotaxime sodium, cefoperazone sodium, and ceftazidime pentahydrate	Iodine (in 1,2-dichloroethane-2,3-dichloro-5,6-dicyano- <i>p</i> -benzo-quinone (in methanol), 7,7,8,8-tetracyanoquinodimethane (in acetonitrile)
Lansoprazole(I) and pantoprazole sodium sesquihydrate	Acceptor 2,3-dichloro-5,6-dicyano-1,4-benzoquinone (DDQ) and with acceptor as iodine
Cinnarizine, analgin, norfloxacin as electron donors	7,7,8,8-Tetracyanoquinodimethane
Phenobarbital (1), thiopental (2), methohexital, phenytoin	Iodine; 7,7,8,8-tetracyanoquinodimethane; 2,3-dichloro-5,6-dicyano-1,4-benzoquinone; 2,3,5,6-tetrachloro-1,4-benzoquinone; 2,3,5,6-tetrafluoro-1,4-benzoquinone; 2,5-dichloro-3,6-dihydroxy-1,4-benzoquinone; tetracyanoethylene-2,4,7-trinitro-9-fluorenon (depending on the solvent polarity)
Haloperidol, droperidol	Iodine (in 1,2-dichloroethane); 2,4,7-trinitro-9-fluorenon (in 1,2-dichloroethane); 7,7,8,8-tetracyanoquinodimethane (in acetonitrile); 2,3-dichloro-5,6-dicyano- <i>p</i> -benzoquinone (in methanol); tetracyanoethylene (in acetonitrile); 2,3,5,6-tetrabromo-1,4-benzoquinone (in methanol)

of information about the correlation of molecular structure and spectra is required for empirical analysis. When the formula of the compound is under study, information obtained from the structure–spectra correlation can help predict several facts: the compound possesses a sufficiently strong absorption band; part of the spectrum in which the band (or bands) nesting occurs; and possible overlap with any interference present in the formulation. The analytical result is the information required to predict the possibility of spectrophotometric determination in a given sample matrix. The structure–spectra correlation is basically empirical and qualitative. If the objective is only the identification of a pharmaceutical active ingredient, all that is required is an agreement of the wavelength of the absorption maximum (minimum) with that of a standard sample. The determination of the specific absorbance at a given wavelength should only be given as much importance as the measurement of any other physical constant (melting point, density, specific optical rotation), as the absorbance value does not give conclusive evidence for assessing the quality of the bulk material.

Databases of spectral details include the complete spectrum (spectral atlases), wavelength of relative maximum (or minimum), and molar absorption coefficients. Basically, few factors should be required to be familiar with the relationship of the molecular structure and the spectrum of a given compound; factors such as the intensity of the absorption bands, the shape of the band, and the wavelength interval of the band. Additionally, the medium is

very important, which shows the critical influence of any solvent on the spectrum of the pharmaceutical.

**See also:** **Spectrophotometry:** Overview; Derivative Techniques; Inorganic Compounds; Organic Compounds; Biochemical Applications.

## Further Reading

- Basavaiah K and Swamy JM (2001) Application of potassium dichromate and iron–thiocyanate in the spectrophotometric investigations of phenothiazines. *Il Farmaco* 56(8): 579–585.
- Benamor M, Aguersif N, and Draa MT (2001) Spectrophotometric determination of cetylpyridinium chloride in pharmaceutical products. *Journal of Pharmaceutical and Biomedical Analysis* 26(1): 151–154.
- Blanco M, Coello J, Gené J, *et al.* (1996) Application of the Davidon–Fletcher–Powell algorithm to the resolution of multicomponent mixtures using UV–Vis spectrophotometry. *Analytica Chimica Acta* 327(2): 145–152.
- Görög S (1995) *Ultraviolet–Visible Spectrophotometry in Pharmaceutical Analysis*. New York: CRC Press.
- Jadranka Kuljanin IJ, Nedeljkovi J, Prstojevi D, and Marinkovi V (2002) Spectrophotometric determination of alendronate in pharmaceutical formulations via complex formation with Fe(III) ions. *Journal of Pharmaceutical and Biomedical Analysis* 28(6): 1215–1220.
- Lakshmi CSR and Reddy MN (1998) Spectrophotometric determination of azathioprine in pharmaceutical formulations. *Talanta* 47(5, 7): 1279–1286.
- Nagaraja P, Sunitha KR, Vasantha RA, and Yathirajan HS (2002) Iminodibenzyl as a novel coupling agent for the

spectrophotometric determination of sulfonamide derivatives. *European Journal of Pharmaceutics and Biopharmaceutics* 53(2): 187–192.

Nagaraja P, Yathirajan HS, Raju CR, *et al.* (2003) 3-Aminophenol as a novel coupling agent for the spectrophotometric determination of sulfonamide derivatives. *Il Farmaco* 58(12): 1295–1300.

Sabry SM, Abdel-Hay MH, Barary MH, and Belal TS (2000) Sensitive spectrofluorimetric and spectrophotometric methods for the determination of thonzylamine hydrochloride in pharmaceutical preparations based on coupling with dimethylbarbituric acid in presence of dicyclohexylcarbodiimide. *Journal of Pharmaceutical and Biomedical Analysis* 22(2): 257–264.

## SPOT TESTS

**E Jungreis**, The Hebrew University of Jerusalem, Jerusalem, Israel

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 4748–4760, © 1995, Elsevier Ltd., with revisions made by the Editor.

### Introduction

Sensitive and selective qualitative tests based on chemical reactions using small amounts of samples and reagents, often in single drops, are referred to as 'spot tests'. The procedure is the ultimate in simplicity. The target is to reach the utmost sensitivity and selectivity with a minimum of physical and chemical operations. Separations and conditioning reactions are generally integrated into the test procedure so that the final test becomes a single operation that can be applied directly for the identification of the substance in question.

The systematic development of spot test methods of analysis occupied Fritz Feigl in Vienna and Rio de Janeiro for half a century up to ~1970. Although in the past few decades chemical analysis has undergone a formidable process of sophistication with the development of advanced instrumental tools, there has been at the same time a contrary trend toward simplification in selected areas in the form of simple, rapid, and inexpensive spot and screening tests. Commercial companies are selling large numbers of compact spot test systems for the rapid establishment of the presence or absence of particular substances in clinical, food, water, soil, and forensic samples. The tests are essentially qualitative, but often can be semiquantitative if procedures as simple as visual comparison of color intensity are used.

### Techniques

The choice of techniques to be followed is generally dictated by the nature of the sample and the reagents

available. The tests are usually run by using one of the following techniques:

- (1) bringing together one drop each of the test solution and the reagent on porous or non-porous supporting surfaces such as paper, glass, or porcelain;
- (2) placing a drop of the test solution on a medium impregnated with appropriate reagents (filter paper, gelatin);
- (3) placing a drop of the reagent solution on a small quantity of the solid specimen (pulverized particles, evaporation or ignition residues);
- (4) subjecting a drop of reagent, or a strip of filter paper, to the action of gases liberated from a drop of test solution or from a minute quantity of the solid specimen.

### Equipment

The equipment required for simple spot test analysis is made of glass, porcelain, plastic, and metal. A classical spot test laboratory setup consists of assorted small sizes of beakers, volumetric flasks, suction flasks, Conway cells, crystallizing dishes, evaporating dishes, separating funnels, extracting pipettes, fritted glass crucibles, graduated cylinders, pipettes, burettes, weighing bottles, storage bottles, vials, test tubes, centrifuge tubes, microscope slides, white and black porcelain dishes, and spot plates. Tweezers and spatulas, microburners, are essential too. A balance, pH meter, ultraviolet (UV) lamp, and small centrifuge should also be available.

### Tests for Cations

It is often possible to carry out conclusive tests with as little as one drop of dilute sample solution, even though considerable concentrations of other substances are present. This possibility permits the microchemical accomplishment of the so-called 'ultimate' analysis, which demands only a decision as to the presence or absence of a particular substance. In

spectrophotometric determination of sulfonamide derivatives. *European Journal of Pharmaceutics and Biopharmaceutics* 53(2): 187–192.

Nagaraja P, Yathirajan HS, Raju CR, *et al.* (2003) 3-Aminophenol as a novel coupling agent for the spectrophotometric determination of sulfonamide derivatives. *Il Farmaco* 58(12): 1295–1300.

Sabry SM, Abdel-Hay MH, Barary MH, and Belal TS (2000) Sensitive spectrofluorimetric and spectrophotometric methods for the determination of thonzylamine hydrochloride in pharmaceutical preparations based on coupling with dimethylbarbituric acid in presence of dicyclohexylcarbodiimide. *Journal of Pharmaceutical and Biomedical Analysis* 22(2): 257–264.

## SPOT TESTS

**E Jungreis**, The Hebrew University of Jerusalem, Jerusalem, Israel

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 4748–4760, © 1995, Elsevier Ltd., with revisions made by the Editor.

### Introduction

Sensitive and selective qualitative tests based on chemical reactions using small amounts of samples and reagents, often in single drops, are referred to as 'spot tests'. The procedure is the ultimate in simplicity. The target is to reach the utmost sensitivity and selectivity with a minimum of physical and chemical operations. Separations and conditioning reactions are generally integrated into the test procedure so that the final test becomes a single operation that can be applied directly for the identification of the substance in question.

The systematic development of spot test methods of analysis occupied Fritz Feigl in Vienna and Rio de Janeiro for half a century up to ~1970. Although in the past few decades chemical analysis has undergone a formidable process of sophistication with the development of advanced instrumental tools, there has been at the same time a contrary trend toward simplification in selected areas in the form of simple, rapid, and inexpensive spot and screening tests. Commercial companies are selling large numbers of compact spot test systems for the rapid establishment of the presence or absence of particular substances in clinical, food, water, soil, and forensic samples. The tests are essentially qualitative, but often can be semiquantitative if procedures as simple as visual comparison of color intensity are used.

### Techniques

The choice of techniques to be followed is generally dictated by the nature of the sample and the reagents

available. The tests are usually run by using one of the following techniques:

- (1) bringing together one drop each of the test solution and the reagent on porous or non-porous supporting surfaces such as paper, glass, or porcelain;
- (2) placing a drop of the test solution on a medium impregnated with appropriate reagents (filter paper, gelatin);
- (3) placing a drop of the reagent solution on a small quantity of the solid specimen (pulverized particles, evaporation or ignition residues);
- (4) subjecting a drop of reagent, or a strip of filter paper, to the action of gases liberated from a drop of test solution or from a minute quantity of the solid specimen.

### Equipment

The equipment required for simple spot test analysis is made of glass, porcelain, plastic, and metal. A classical spot test laboratory setup consists of assorted small sizes of beakers, volumetric flasks, suction flasks, Conway cells, crystallizing dishes, evaporating dishes, separating funnels, extracting pipettes, fritted glass crucibles, graduated cylinders, pipettes, burettes, weighing bottles, storage bottles, vials, test tubes, centrifuge tubes, microscope slides, white and black porcelain dishes, and spot plates. Tweezers and spatulas, microburners, are essential too. A balance, pH meter, ultraviolet (UV) lamp, and small centrifuge should also be available.

### Tests for Cations

It is often possible to carry out conclusive tests with as little as one drop of dilute sample solution, even though considerable concentrations of other substances are present. This possibility permits the microchemical accomplishment of the so-called 'ultimate' analysis, which demands only a decision as to the presence or absence of a particular substance. In

**Table 1** Selected spot tests for cations

<i>Cation</i>	<i>Reagent</i>	<i>Test conditions</i>	<i>Reaction type</i>	<i>Visible change</i>	<i>Interference</i>	<i>Limit of identification (μg)</i>
Aluminum $\text{Al}^{3+}$	Morin, saturated soln in methanol	Neutral or slightly acidic test soln	Complex formation	Green fluorescence	$\text{Be}^{2+}$ ; $\text{Ga}^{3+}$ ; $\text{In}^{3+}$ ; $\text{Th}^{4+}$ ; $\text{Sr}^{2+}$	0.2
Ammonium $\text{NH}_4^+$	Red litmus paper	After liberation of ammonia with NaOH	Acid–base indicator	Blue color	Volatile bases, e.g., amines	0.01
Antimony $\text{Sb}^{3+}$ (i.e., $\text{SbO}_3^{3-}$ )	Filter paper impregnated with 5% molybdophosphoric acid	After addn of a test soln, hold paper over steam	Molybdenum blue formation	Blue color	$\text{Sn(II)}$	0.2
Arsenic $\text{As}^{3+}$ (i.e., $\text{AsO}_3^{3-}$ )	Silver nitrate, 1% aqueous soln	Test soln warmed with a few drops of $\text{NH}_3$ and a drop of 10% $\text{H}_2\text{O}_2$ in a test tube; acidified, and reagent added	Silver arsenate formation	Red–brown ppt	Chromates and $\text{Fe(CN)}_6^{3-}$	6
Barium $\text{Ba}^{2+}$	Sodium rhodizonate 0.2% aqueous soln	Neutral or slightly acidic test soln + reagent on filter paper	Complex formation	Red–brown ppt	$\text{M}^{2+}$	0.25
Beryllium $\text{Be}^{2+}$	Morin + EDTA	Ammoniacal	Complex formation	Yellow–green fluorescence		0.07
Bismuth $\text{Bi}^{3+}$	Cinchonine + iodide	Slightly acidic, on filter paper impregnated with reagent	Ion pair formation with $\text{BiI}_4^-$	Orange–red ppt	$\text{Cu}^{2+}$ ; $\text{Pb}^{2+}$ ; $\text{Hg}^{2+}$	0.14
Cadmium $\text{Cd}^{2+}$	Iron(II)–bipyridyl iodide	Weakly acidic, neutral, or ammoniacal test soln on filter paper	Ion pair formation with $\text{CdI}_4^{2-}$	Red ppt	$\text{Hg}^{2+}$ ; $\text{Sn(II)}$ ; $\text{Sb(III)}$ ; $\text{Bi}^{3+}$	0.05
Calcium $\text{Ca}^{2+}$	Glyoxal bis(2-hydroxyanil)	Neutral or acidic test soln. After reaction add NaOH and $\text{Na}_2\text{CO}_3$ ; extract with chloroform	Complex formation	Red color in the chloroform layer	$\text{Cd}^{2+}$	0.05
Chromium $\text{Cr}^{3+}$	Diphenylcarbazide	First add potassium peroxodisulfate and $\text{AgNO}_3$ . After 2–3 min add reagent	Complex formation	Violet color	$\text{Hg}^{2+}$ ; $\text{MoO}_4^{2-}$ ; $\text{VO}_3^-$	0.8
Cobalt $\text{Co}^{2+}$	Nitroso-R salt	Dowex-1 XI ion exchange resin treated on spot plate with reagent. $\text{CH}_3\text{COONa}$ and the sample. After 5 min add $\text{HNO}_3$ and heat on water bath	Complex formation	Red color	None	0.003
Copper $\text{Cu}^{2+}$	2,2'-Biquinolyl (cuproin)	Acidic test soln treated with hydroxylamine to reduce $\text{Cu(II)}$ to $\text{Cu(I)}$	Complex formation	Purple–red color	None	0.05
Gold $\text{Au}^{4+}$	<i>p</i> -Dimethylaminobenzylidene-rhodanine	Neutral or weakly acidic test soln placed on reagent paper	Formation of $\text{Au(I)}$ complex	Violet ppt	$\text{Ag}^+$ ; $\text{Hg}^{2+}$ ; $\text{Pd}^{2+}$	0.1
Iron $\text{Fe}^{2+}$	2,2'-Bipyridyl		Complex formation	Red color	None	0.03
Lead $\text{Pb}^{2+}$	Dithizone in $\text{CHCl}_3$	Test soln shaken with cyanide (CARE) and reagent soln	Complex formation	Red $\text{CHCl}_3$ layer	Specific	0.04
Magnesium $\text{Mg}^{2+}$	Quinalizarin	Mix test soln and reagent. Add NaOH soln until color changes to violet	Lake formation	Blue ppt (blank: blue–violet soln)	Metal ions precipitated at alkaline pH	0.25



Manganese Mercury $\text{Hg}^{2+}$	Periodate + tetrabase Diphenylcarbazone	Dilute acetic acid Test soln in $\text{HNO}_3$ placed on reagent paper	Oxidation to $\text{MnO}_4^-$ Complex formation	Deep blue color Violet–blue ppt	$\text{Cr}^{3+}$ Selective in the absence of chromate and molybdate	0.001 0.2
Molybdenum $\text{Mo(VI)}$ $(\text{MoO}_4^{2-})$	<i>o</i> -Hydroxyphenylfluorone	Acidic, test on filter paper. Add KF to mask interferences		Carmine red stain	None	1.7
Nickel $\text{Ni}^{2+}$	Dimethylglyoxime	Test soln and reagent placed on filter paper, held over $\text{NH}_3$	Complex formation	Red ppt	$\text{Pd}^{2+}$ ; $\text{Fe}^{2+}$ ; $\text{Co}^{2+}$	0.16
Palladium $\text{Pd}^{2+}$	Nickel-dimethylglyoxime paper	Treat with drop of test solution, wash with dil. acid	Protective layer formed over Ni complex	Red stain	No data	0.05
Platinum $\text{Pt(IV)}$	(1) Saturated aqueous $\text{TlNO}_3$ ; soln	Reagent (1) on filter paper. Add test soln. Wash with $\text{NH}_3$ and treat with reagent (2)	Reduction to Pt element	Orange–red stain	None	0.025
Potassium $\text{K}^+$	(2) $\text{SnCl}_2$ in strong HCl 0.05% $\text{AgNO}_3$ soln and a tiny crystal of $\text{Na}_3[\text{Co}(\text{NO}_2)_6]$	Neutral or acetic acid test soln placed on black spot plate with a drop of reagent and crystal	Precipitation	Yellow ppt	$\text{Li}^+$ ; $\text{Ti}^+$ ; $\text{NH}_4^+$	4
Silver $\text{Ag}^+$	Filter paper impregnated with saturated acetone solution of <i>p</i> - dimethylaminobenzylidene– rhodanine	Slightly acidic test soln	Complex formation	Red–violet spot against yellow–brown background	$\text{Hg}^{2+}$ ; $\text{Au(III)}$ ; $\text{Pt}^{2+}$ ; $\text{Pd}^{2+}$ ; $\text{Cu}^{2+}$	0.02
Sodium $\text{Na}^+$	Zinc uranyl acetate	Neutral or acetic acid test soln mixed on a black spot plate with reagent soln	Insoluble salt precipitated	Fluorescent, yellow ppt	High cone, of Li and K	2.5
Strontium $\text{Sr}^{2+}$	Sodium rhodizonate soln, filter paper impregnated with saturated soln of potassium chromate	Test soln is placed on paper. After 1 min a drop of reagent is added	Complex formation	Brown–red	Selective among the alkaline-earth metals	3.9
Tin $\text{Sn(II)}$	Ammonium molybdophosphate on filter paper	Test soln. placed on reagent paper	Reduction by $\text{Sn(II)}$	Blue	No data	0.03
Tungsten $\text{W(VI)}$ $(\text{WO}_4^{2-})$	2,4'-Diaminodiphenyl hydrochloride in HCl	Test soln mixed with reagent	Precipitation	White ppt	Molybdate in high conc.	6
Uranium $\text{U(VI)}$ $(\text{UO}_2^{2+})$	8-Quinolinol (oxide) soln in ethanol	Precipitate interfering ions $(\text{NH}_4)_2\text{CO}_3$ soln. Filtrate treated with reagent on spot plate	Complex formation	Red–brown ppt	None	10
Vanadium $\text{V(V)}$ $(\text{VO}^{3+}; \text{VO}_2^+)$	2-(Salicylidene-imino) benzoic acid on filter paper	Drop of acidic sample spotted on the paper	Complex formation	Violet stain	None	0.2
Zinc $\text{Zn}^{2+}$	Dithizone in $\text{CCl}_4$	Drop of test soln mixed on a watch glass with a drop of $2 \text{ mol l}^{-1}$ $\text{NaOH}$ and a few drops of reagent. Evaporate the $\text{CCl}_4$	Complex formation	Raspberry red soln (a colored ppt is disregarded)	None	5
Zirconium $\text{Zr}^{4+}$	Neothorin (Arsenazo I)	Drop of sample in conc. HCl treated with reagent	Complex formation	Blue–violet ppt	$\text{Hf}^{4+}$	3

ppt, precipitate; soln, solution.

many cases this can be achieved successfully without a preliminary separation by application of selective reactions.

It is also advisable to try out the reaction with pure solutions containing a series of concentrations of the substance to be detected.

Some selected examples of spot tests for cations are described in **Table 1**. Features such as the selectivity of the reaction, its sensitivity, the availability of the reagents, and simplicity of execution are the main parameters upon which the selection is based. The short descriptions given may be enough to allow the tests to be carried out, but Feigl's books should be consulted for more details and a deeper understanding of the reactions and to solve acute interference problems.

## Tests for Anions

Tests for anions are usually made in solutions that contain cations only of the alkali metals and thallium(I). This is achieved through preparation of sodium carbonate extracts by treating the test solutions with sodium carbonate and removing the resulting precipitate.

The metal cations in these procedures are transformed into their slightly soluble carbonates and hydroxides and are thus separated from the water-soluble alkali salts of the anions. Thallium(I), whose carbonate is soluble, is an exception. Spot reactions to detect the various anions can be successfully made on drops of the sodium carbonate extract.

Selected spot tests for anions are given in **Table 2**.

## Tests for Organic Compounds

Organic compounds may be detected by testing for a particular functional group (e.g., ketone, nitro compound) or for individual compounds. Tests for the more common functional groups are summarized in **Table 3**. Specific tests for some important individual compounds are given in **Table 4**. In the majority of cases, there is some previous knowledge about the origin of the sample, which may replace the necessity for complete specificity.

## Applications

Parallel to the advance toward instrumental sophistication, a trend toward simplification has been seen in some marginal, selected areas in the form of simple, rapid, and inexpensive spot and screening tests. A screening test is defined as a simple method that provides a sufficient answer the analytical question

with a minimum expenditure of time and money. It would be wasteful to devote resources to attaining precision or accuracy several magnitudes beyond that necessary in clinical urine analysis, for example, when pathological changes in the composition of urine are evidenced by a simple spot test.

The use of spot and screening tests is a marginal one; the margin, however, is quite significant. For the exact determination of most chemical substances, complex analytical procedures are unavoidable.

Commercial companies in the United States, Europe, and Japan are selling vast quantities of compact spot test systems to solve problems in clinical analysis, in control tests of air quality, in food, water, and soil analysis, and in forensic laboratories.

Simple, compact, and inexpensive analytical devices have been elaborated for semiquantitative evaluation of certain elements and compounds where such an approximated result has diagnostic value, in at least the first stage of the examination.

In the remainder of this article, selected examples of applications of spot tests are given.

## Clinical Applications

### Glucose

The first rapid tests in urine analysis were carried out in the 1940s by adding tablets to urine samples. The test for sugar was refined in the Clinitest<sup>®</sup> reagent tablets (Ames Co., Elkhart, IN). They are an ingenious adaptation of the alkaline copper reduction test in self-heating tablet form. Each tablet contains copper sulfate, sodium carbonate, and citric acid. Sodium carbonate and citric acid form an effervescent couple, which facilitates the rapid dissolution of the tablet and generates a little heat. Much more heat is liberated by the dissolution of sodium hydroxide and its partial neutralization by citric acid. In the alkaline medium, the sugar reduces the blue copper(II) sulfate solution to reddish insoluble copper(I) oxide. The carbon dioxide displaces the air above the reaction and prevents reoxidation of the copper(I) oxide during the test. The color of the mixture indicates the proportion of sugar in urine.

The Clinistix<sup>®</sup> (Ames Co.) is an enzymatic glucose test based on the activity of the enzyme glucose oxidase that uses dry-reagent chemical technology. A firm plastic strip, to which a stiff absorbent cellulose area is affixed, is impregnated with a buffered mixture of glucose oxidase, peroxidase, and *o*-toluidine. In the first stage of the reaction, glucose is oxidized by atmospheric oxygen in the presence of glucose oxidase to gluconic acid and hydrogen peroxide,

**Table 2** Selected spot tests for anions

Anion	Reagent	Test conditions	Reaction type	Visible change	Interference	Limit of identification ( $\mu\text{g}$ )
Azide $\text{N}_3^-$	Iron(III) chloride	Hydrazoic acid liberated in a test tube by addition of dilute HCl to test soln. Filter paper wetted with reagent soln covers the tube	Iron(III) azide formation	Red	$\text{SO}_4^{2-}$ , $\text{S}_2\text{O}_3^{2-}$	2.5
Borate $\text{BO}_2^{3-}$	Benzoin	Test soln evaporated to dryness with NaOH. After cooling add reagent	Fluorescent complex formed	Yellow–green	Chromate, periodate, peroxodisulfate, and hexacyanoferrate(III) must be reduced	0.04
Bromate $\text{BrO}_3^-$	Sulfanilic acid	Test soln and $\text{HNO}_3$ mixed with reagent	Selective oxidation to form azo compound	Violet turning yellow–brown	Selective in the presence of chlorate and iodate	5
Bromide $\text{Br}^-$	(1) Fluorescein in ethanol  (2) Acetic acid–hydrogen peroxide	Test soln placed on filter paper spotted with reagent (2). After drying the spot is treated with reagent (1)	Tetrabromofluorescein (eosin) formation	Red color produced		0.3
Carbonate $\text{CO}_3^{2-}$	Uranyl acetate soln treated with some drops of potassium hexacyanoferrate(II)	Sample mixed on a spot plate with reagent	Demasking of uranyl hexacyanoferrate(II)	Decoloration of brown solution	Phosphate	0.4
Chlorate $\text{ClO}_3^-$	Saturated manganese(II) sulfate and syrupy phosphoric acid (1:1 mixture, v/v)	Test soln and reagent mixture briefly warmed over a microburner and allowed to cool	Formation of a complex manganese (III) phosphate	Violet	Periodate	0.05
Chloride $\text{Cl}^-$	(1) Soln of 8-quinolinol in acetic acid (2) $\text{H}_2\text{O}_2$ + dilute acetic acid	Test soln warmed with both of the reagent solns and $\text{HNO}_3$ and $\text{AgNO}_3$ soln added	Reagent binds all halogens but chlorine	Colorless ppt		2
Chlorite $\text{ClO}_2^-$	$\text{NiSO}_4 \cdot 7\text{H}_2\text{O}$ soln and $0.1 \text{ mol}^{-1} \text{ NaOH}$	Reagent solns mixed with the test soln and heated in a water bath ( $40\text{--}50^\circ\text{C}$ ) for a few min	Formation of Ni(III) or Ni(IV) oxides	Black specks	None from chlorate, perchlorate, bromate, iodate or periodate	1
Cyanate $\text{OCN}^-$	Hydroxylammonium chloride and $\text{FeCl}_3$ soln	Test solution mixed with both reagents	Resulting hydroxyurea forms complex with $\text{Fe}^{3+}$	Transient violet color	$\text{SCN}^-$	30
Cyanide $\text{CN}^-$	<i>p</i> -Benzoquinone in dimethylsulfoxide	Sample mixed with the reagent	Unknown	Blue–green fluorescence	None	0.2
Fluoride $\text{F}^-$	Alizarin complexan buffered to pH 4.3 and cerium(III) nitrate soln	Sample mixed with the reagents	Ternary complex formation with alizarin complexan – Ce(III) chelate	Color changes from red to lilac blue	None	0.2

Continued

**Table 2** Continued

Anion	Reagent	Test conditions	Reaction type	Visible change	Interference	Limit of identification ( $\mu\text{g}$ )
Hexacyanoferrate(III) $[\text{Fe}(\text{CN})_6]^{3-}$	$\text{ZnCl}_2$ in ethanolic soln saturated with tetrabase	Weakly acidic test soln mixed on filter paper with the reagent	Oxidation to the quinoidal dyestuff	Blue	Chromate, dichromate, permanganate	1
Hexacyanoferrate(II) $[\text{Fe}(\text{CN})_6]^{4-}$	2,2'-Bipyridyl with several milligram of amidomercury(II) chloride	Neutral or sodium carbonate-containing test soln warmed with the reagent	Demasking of $[\text{Fe}(\text{CN})_6]^{4-}$	Red	Not hexacyanoferrate(III)	0.2
Hypohalite $\text{ClO}^-$ , $\text{BrO}^-$ , $\text{IO}^-$	Filter paper impregnated with Safranin O	Test soln dropped onto the paper and washed in water	Possibly an oxidative decomposition of the dye	Violet on the red paper	Hexacyanoferrate(III)	0.5 ( $\text{ClO}^-$ , $\text{BrO}^-$ ); 2.5 ( $\text{IO}^-$ )
Iodate $\text{IO}_3^-$	<i>p</i> -Aminophenol hydrochloride soln	Test soln mixed with the reagent	Selective oxidation to an indamine dye	Violet coloration	Periodate interferes; bromate and chlorate do not	0.5
Iodide $\text{I}^-$	Acetic acid and $\text{KNO}_2$ soln	Starch-containing filter paper is treated with acetic acid, test soln and $\text{KNO}_2$ soln in succession	Oxidation to $\text{I}_2$ , formation of colored compound with starch	Blue	Cyanide	0.025
Nitrite $\text{NO}_2^-$	Diphenylamine in conc. $\text{H}_2\text{SO}_4$	Test soln containing interfering oxidants taken to dryness, heated to 400–500°C and redissolved in water. Test soln is added to the reagent	Oxidation to a quinoida compound	Blue color	Chromate, vanadate, molybdate, antimony(V)	0.5
Nitrite $\text{NO}_2^-$	Sulfanilic acid in acetic acid and 1-naphthylamine oxalate impregnating a filter paper	Neutral or acetic acid test soln dropped on the reagent paper	Diazotization and azo dye formation	Blue–violet color	None	0.00005
Perchlorate $\text{ClO}_4^-$	Filter paper impregnated with thio-Michlers ketone	Take neutral or weakly acidic test soln to dryness with a drop of 2% $\text{CdCl}_2$ soln. Heat on a microburner, while covered with reagent paper	Chlorine formation	Yellow reagent paper turns blue	Halogenates and nitrates eliminated by fuming with HCl	0.1
Periodate $\text{IO}_4^-$	(1) Manganese(II) sulfate and syrupy phosphoric acid	Sample mixed with reagent (1) in a small porcelain dish	Formation of complex manganese(III) phosphate anions	Violet color which is intensified by addition of reagent (2)	Peroxodisulfate, chlorate	5
	(2) Diphenylcarbazine, ethanolic soln	Warmed briefly over a microburner and allowed to cool before adding reagent (2)				
Permanganate $\text{MnO}_4^-$	Cellulose in a thick fillet paper	Neutral or slightly acidic test soln soaked into the paper. Wait 3 min, then wash the paper with water	Reduction of $\text{MnO}_4^-$ to $\text{MnO}_2$ by cellulose	Brown circular stain		0.3 (in presence of 20 mg of $\text{K}_2\text{CrO}_4$ )

Peroxodisulfate $\text{S}_2\text{O}_3^{2-}$	(1) NaOH soln	Test soln and reagents (1) and (2) in succession are placed on a spot plate	Oxidation of $\text{Ni}(\text{OH})_2$ to $\text{NiO}(\text{OH})_2$	Black or grey ppt	$\text{H}_2\text{O}_2$	2.5
Phosphate $\text{PO}_4^{3-}$	(2) $\text{NiSO}_4 \cdot 6\text{H}_2\text{O}$ soln Ammonium molybdate soln and ascorbic acid soln	Acidic test soln placed on filter paper, followed by a drop of each of the reagent solns. The paper is held over $\text{NH}_3$	Formation of molybdenum blue	Blue stain	$\text{H}_2\text{O}_2$ , oxalate, fluoride, silicate	1.25
Pyrophosphate ( $\text{P}_2\text{O}_7^{4-}$ ) and polyphosphates	Iron(III) perchlorate soln and ammonium thiocyanate soln, freshly prepared	Sample on filter paper sprayed with the reagents	Demasking of the iron(III) thiocyanate complex	White spot on pinkish background	Phosphate and its esters do not react	30 (as P)
Silicate $\text{SiO}_4^{4-}$	(1) Ammonium molybdate soln in $\text{HNO}_3$ (2) Ascorbic acid soln	Test soln mixed with reagent  (1) Supernatant liquid transferred to a microcrucible and gently warmed. After cooling, 1% oxalic acid soln added, followed by reagent (2) and sodium acetate soln	Formation of molybdenum blue	Blue color	1.5 mg $\text{P}_2\text{O}_5$ does not interfere	12
Sulfate $\text{SO}_4^{2-}$	(1) $\text{Ba}(\text{NO}_3)_2$ soln	Weakly acidic test soln spotted on filter paper which has been treated successively with each of reagents (1), (2), and (3)	Demasking of Ba-rhodizonate and formation of silver rhodizonate	Red-brown ppt turning violet	Sulfite, sulfide, and thiosulfate do not interfere	0.3
Sulfide $\text{S}^{2-}$	(2) Sodium rhodizonate soln (3) Silver nitrate soln Sodium azide-iodine soln	Sample and reagent mixed on a watch glass	Catalytic acceleration of the reaction between the two components of the mixture	Brown color of iodine disappears gradually and bubbles of nitrogen appear	Thiosulfate, thiocyanate	0.3
Sulfite $\text{SO}_3^{2-}$	(1) Sodium azide soln	Reagent (3) placed on filter paper and spotted with reagent (1). A violet stain forms. Test soln (pH 5–6) and reagent (2) are added	Acceleration of the slow auto-oxidization of $\text{Co}(\text{II})$ -azide complex by sulfite	A light yellow color changes to blue		0.5
Thiocyanate $\text{SCN}^-$	(2) <i>o</i> -Tolidine in acetic acid soln (3) $\text{CoCl}_2$ soln Sodium azide-iodine solution (see sulfide)	Possible interferents (sulfide, thiosulfate) precipitated by addition of $\text{HgCl}_2$ . Reagent added to the clear supernatant liquid	Catalysts of the iodine-azide reaction	Brown color forms; bubbles of nitrogen develop	None	10

ppt, precipitate; soln, solution.

**Table 3** Detection of functional groups

Functional group	Reagent	Procedure	Visible change	Remarks	Identification limit ( $\mu\text{g}$ )
Aromatic compound	Freshly prepared, 0–2 ml formaldehyde in 10 ml conc. $\text{H}_2\text{SO}_4$	On a spot plate mix with a drop of the test solution	Red, green, brown–orange color development	Blank test with sulfuric acid alone is necessary	2–50
Alcohols (primary, secondary, and tertiary)	1 mg vanadium treated with 1 ml 2.5% 8-quinolinol in 6% acetic acid, shaken with 30 ml benzene	Mix reagent with the test solution, heat in a water bath at $60^\circ\text{C}$	Red color appears after 6–8 min	Reagent solution keeps 1 day	5–20
Secondary alcohols	2% solution of sulfur in carbon disulfide	Heat with sample in a microtest tube; cover mouth of tube with lead acetate paper	Black or brown stain on the paper	Hydrogen sulfide results	10–200
Phenols (volatile)	Filter paper impregnated with a saturated benzene solution of 2,6-dichloroquinone-4-chloroimide	Heat sample in microtest tube at $150^\circ\text{C}$ ; cover mouth of tube with reagent paper. After several min hold reagent paper over ammonia vapor	Blue stain	The phenol must have a free <i>para</i> position	0.3–1
Phenols	(1) Freshly mixed oxalic acid and hexamine  (2) Hydrazine sulfate boiled with sodium acetate in water	Mix reagent (1) and 1 drop of ethanol. Heat to $150\text{--}160^\circ\text{C}$ for 1–2 min. After cooling add a drop of reagent (2). Examine on filter paper	Blue–green fluorescence	No response is given by 2,4- and 2,6-dimethylphenol	0.25–5
Aliphatic polyhydroxides	(1) 5% potassium periodate  (2) 10% sulfuric acid (3) Saturated sulfurous acid (4) 0.1% fuchsin solution; $\text{SO}_2$ passed until the color is discharged	In a microcrucible mix test solution with reagents (1) and (2). Stand for 5 min. Excess periodic acid is reduced with reagent (3) and treated with a drop of reagent (4)	Red to blue color	Citric acid, inositol, pentaerythritol, pentacetylglucose, acetylcellulose do not react	2.5–100
Aldehydes (aliphatic and aromatic)	Pure indole freshly dissolved in pure conc. $\text{HCl}$	Test solution treated with reagent in boiling water	Pink to orange color		$10^{-4}$ –20
Aliphatic aldehydes and ketones	$\text{CuSO}_4$ and an alkaline solution of sodium potassium tartrate (Fehling's solution)	Sample added to reagent solution in a test tube and immersed in boiling water	Dark red crystalline ppt		5–25
Aromatic aldehydes	Solid <i>N,N</i> -dimethyl- <i>p</i> -phenylenediamine	Test solution mixed with reagent and warmed in a water bath	Red color		0.05–100
Ketones	Alcoholic <i>m</i> -dinitrobenzene solution; saturated methanolic potassium hydroxide	Reagents mixed, sample added, and warmed to $70\text{--}80^\circ\text{C}$	Violet–red color		5–200



Carboxylic acids	(1) Thionyl chloride	Solid analyte is treated with reagent (1), and is evaporated almost to dryness. Reagent (2) is added, the mixture made basic to litmus with alcoholic alkali, acidified with reagent (3) and treated with reagent (4)	Color varies from brown–red to dark violet	Positive responses also given by citric and thioacetic acids	10–33
Alkyl esters of carboxylic acids	(2) Saturated alcoholic hydroxylammonium chloride (3) 0.5 mol l <sup>-1</sup> hydrochloric acid (4) 1% aqueous iron(III) chloride (1) Very small piece of metallic sodium (CARE!)	Add to reagent (1) on a spot plate a benzene solution of the ester and reagent (2). After 1–2 min add a drop of water	Violet color		5–10
Ethers	(2) 2.5% benzene solution of dinitrobenzene (1) 5% aqueous copper ethylacetoacetate	Test solution heated in glycerol bath to 230°C. The mouth of the test tube is covered with a disk of filter paper, moistened by reagents (1) and (2)	Blue color	Peroxide formation	20
Amines (aliphatic, aromatic, and heterocyclic)	(2) 5% chloroform solution of tetrabase 5% dimethylsulfoxide solution of 1,3,5-trinitrobenzene	Mix test solution with reagent	Various colors	Complex formation. Glycine does not react	5–25
Primary and secondary amines (aliphatic and aromatic)	1% Dimethylaminocinnamaldehyde dissolved in methanol/trichloroacetic acid	Reagent solution placed on filter paper which is then spotted with ether solution of the amine. Heat at 100°C for 3 min	Primary amines, blue; secondary amines, purple	Schiff base formation	0.02–1.7
Primary and secondary amines (aliphatic)	(1) 1:1 mixture of ethanol:carbon disulfide	Alcoholic test solution mixed with reagent (1). After 5 min reagent (2) is added	Evolution of nitrogen gas	Dithiocarbamate formation and its catalytic action on reagent (2)	1–250
Aliphatic amines	(2) Iodine–sodium azide solution Filter paper impregnated with 2% diphenylcarbazide in formaldehyde	Test solution placed on reagent paper	Pink–violet color	No reaction in the absence of filter paper	0.3–1.6
Aliphatic secondary amines	(1) 5% aqueous copper sulfate solution	Acidic test solution added to reagent (1), made basic with ammonia and shaken with reagent (2)	Appearance of a brown benzene layer	Formation of copper dithiocarbamates	0.2–10
Aromatic amines	(2) 1:3 mixture carbon disulfide and benzene Furfural in glacial acetic acid	Test solution treated on a spot plate with reagent	Violet color	Schiff base formation	0.1–200

*Continued*

**Table 3** Continued

<i>Functional group</i>	<i>Reagent</i>	<i>Procedure</i>	<i>Visible change</i>	<i>Remarks</i>	<i>Identification limit (<math>\mu\text{g}</math>)</i>
Nitro compounds	5% Benzene solution of tetrabase	Ether or benzene solution of sample treated with reagent and dipped into boiling water	Appearance of a yellow melt	Formation of molecular complex	0.1–50
Primary aliphatic nitro compounds	(1) Freshly prepared alcohol solution of Fast Blue B	On filter paper spot reagent (1), the alcoholic test solution and reagent (2)	Orange–red spot	Condensation product formation	0.5
Nitro compounds (aliphatic and aromatic)	(2) $0.5 \text{ mol l}^{-1}$ sodium hydroxide (1) Zn dust and acetic acid	Sample reduced with reagent (1) to the hydroxylamine. Add reagents (2) and (3)	Pink or violet color	Color more stable at $0^\circ\text{C}$ ; hydroxylamine interferes	
Aromatic nitro compounds	(2) Benzenediazonium chloride (3) Saturated diphenylamine in acetic (1) 10% aqueous calcium chloride solution	Few milligrams of sample dissolved in 3 ml of hot ethanol treated with reagent (1). Reagent (2) is added and the tube is heated to strong boiling. Cooled filtrate treated with reagent (3)	Purple, blue, or green color	Chloromycetin can be detected likewise	0.3–15
<i>m</i> -Dinitro aromatic compounds	(2) Zinc dust (3) 1% aqueous $\text{Na}_3[\text{Fe}(\text{CN})_5\text{NH}_3]$ (1) Amberlite IRA-400 resin beads (Cl-form)	On a white spot plate sample added to a few beads of reagent (1) followed by reagents (2) and (3)	Blue–violet color	Specific for dinitro aromatic compounds	$10^{-4}$ – $1.6 \times 10^{-3}$
Amino acids	(2) Saturated aqueous sodium sulfite (3) Dimethyl sulfoxide (1) 0.01% Fluorescein isothiocyanate in $0.1 \text{ mol l}^{-1}$ sodium hydroxide	On a TLC plate dried sample sprayed with reagent (1), heated at $90^\circ\text{C}$ for 10 min, cooled and sprayed with reagent (2). Heated again at $90^\circ\text{C}$ , before being observed under visible and UV radiation	Different colors for each acid	UV produces fluorescence	0.5–1.0 (visible light)
	(2) 0.25% ninhydrin in acetone				0.3–1.0 (UV)

Nitriles	(1) Oxalic acid	Sample and reagent (1) heated in a glycerol bath at 200°C for 5 min. After cooling, reagent (2) is added and the mixture heated to 140–160°C in a glycerol bath	Orange product	The tests may not be applied in the presence of ammonium salts, amides of acids, amino acids, and ureides	2–20
Aliphatic nitriles	(2) Thiobarbituric acid (1) 1:1 mixture of CaCO <sub>3</sub> –CaO	Solid sample mixed with reagent (1), heated in a glycerol bath at 250°C, with the mouth of the tube covered with a filter paper disk moistened with reagent (2)	Blue stain after 3–4 min		5–50
Thioketones and thiols	(2) 5% solution of Cu ethylacetoacetate and 5% tetrabase in chloroform Sodium azide in iodine solution	Test solution mixed with a drop of reagent	Evolution of bubbles of nitrogen	Catalytic reaction. Inorganic compounds containing sulfur(II) must be absent	$3 \times 10^{-4}$ –0.6
Primary and secondary thiols	(1) Conc. NH <sub>3</sub>	Test solution treated with reagent (1) in test tube, covered with reagent (2), and heated in a boiling water bath	Black stain (PbS) with a metallic luster	Tertiary thiols and thioketones give a negative response	1–100
Sulfonic acids	(2) Lead acetate paper (1) Succinic or phthalic acid	Solid sample mixed with reagent (1), heated in a test tube in a glycerol bath at 200°C. A disk of filter paper moistened with reagent (2) is placed over the mouth of the tube and the temperature raised to 250°C	A blue stain appears	Sulfurous acid is formed during the pyrohydrolytic process	1–20
Organic cyanates	(2) Iron(III) hexacyanoferrate(III) solution Dimethylbarbituric acid in aqueous pyridine	Test solution mixed with a drop of the reagent	Blue spot appears within 5 min		36–180 (μmol)

TLC, thin-layer chromatography.

**Table 4** Detection of individual compounds

<i>Compound</i>	<i>Reagent</i>	<i>Procedure</i>	<i>Visible change</i>	<i>Remarks</i>	<i>Identification limit (μg)</i>
Acetone	(1) Ethanolic solution of salicylaldehyde	Mix reagents (1) and (2) with the sample. Heat for 2–3 min at 70–80°C	Deep red color	Specific test	0.2
Acetic acid	(2) 40% aqueous sodium hydroxide (1) 5% aqueous lanthanum nitrate solution (2) 1 mol l <sup>-1</sup> ammonia	On a spot plate mix the sample with reagents (1) and (2)	Blue to brown rings within a few minutes	Sulfate and phosphate interfere	50
Acetylene	(1) Ammoniacal silver chromate	In a small porcelain crucible mix reagents (1) and (2) and sample. Stir for 10 min	Red–brown residue	Protective layer formation by acetylene	1
Barbituric acid	(2) 6 mol l <sup>-1</sup> acetic acid (1) 2 mol l <sup>-1</sup> acetic acid	On a spot plate mix sample with reagents (1) and (2)	Red–violet color	Formation of a violuric acid	10
Carbon disulfide	(2) Saturated sodium nitrite solution Copper sulfate solution treated with conc. ammonia and piperidine	In test tube treat sample with reagent and extract with benzene	A yellow–brown extract	Formation of copper dithiocarbamate	3
Carbon tetrachloride	(1) Quartz dust	In a test tube mix test solution with reagent (1). Immerse in glycerol bath at 150°C; cover mouth of tube with a disk of reagent (2), moistened with reagent (3)	Formation of a blue stain	Chlorine release in the presence of quartz	5
Citric acid	(2) Congo-red paper (3) 5% hydrogen peroxide (1) Potassium permanganate solution	In a test tube mix test solution with reagents (1) and (2). Warm slightly and after cooling add solid reagent (3)	White ppt forms	Pentabromoacetone formation	6
Cysteine	(2) Saturated bromine water (3) Sulfosalicylic acid (1) Aqueous sodium nitroprusside (2) 1% aqueous potassium carbonate	On a spot plate a drop of the test solution is mixed with reagents (1) and (2)	Pink to deep red color		0.001
Ethanol	(1) Potassium permanganate solution in (1:1) sulfuric acid  (2) Aqueous morpholine and sodium nitroprusside	In a test tube treat test solution with reagent (1), with mouth of tube covered with paper moistened with reagent (2)	Blue stain on the paper	Acetaldehyde formation	3

Formaldehyde	0.1% of J-acid (6-amino-1-naphthol-3-sulfonic acid) in conc. sulfuric acid	On a glass fiber filter paper add aqueous test solution to the reagent	After 5–10 s a yellow fluorescence under UV. The yellow solution turns blue after addition of water	Xanthylum dyestuff production	0.01
2-Furaldehyde	(1) Ethanolic aniline	Sample treated with reagents (1) and (2)	Brilliant red color		20
Formic acid	(2) Aqueous acetic acid (1) <i>N</i> -methylquinaldinium toluene- <i>p</i> -sulfonate in dimethyl sulfoxide	In a test tube treat sample with reagent (1) followed by reagents (2) and (3). Heat in boiling water for 1 min	Blue to green color	Cyanine dyestuff formation	0.004
Glycerol	(2) Pyridine (3) Acetic anhydride (1) Aminophenol in ethanol	In a test tube evaporate reagent (1) at 110°C, add the sample followed by reagent (2). Heat at 140°C for ~15 min, cool and add reagents (3), (4), and (5); cool	Bluish-green fluorescence under UV radiation	Formation of 8-quinolinol and its fluorescent sail	0.5
Hydrazine derivatives	(2) 1% arsenic acid in conc. sulfuric acid (3) Conc. sodium hydroxide (4) Magnesium sulfate solution (5) Conc. ammonia (1) Dowex HCR resin (H <sup>+</sup> form)	In a test tube mix 5 beads of reagent (1) with test solution and reagent (2). Heat in a boiling water bath for 15 min. Place beads on a spot plate and mix with reagent (3)	Hydrazine, alkyl and arylhydrazines give orange–red bead surface		
Malonic acid	(2) 2% Dimethylaminobenzaldehyde in acidified ethanol (3) Aqueous 2.5% ammonia (1) Urea in methanol	In a test tube mix test solution with reagent (1). After evaporation heat tube for several min in a glycerol bath at 120°C. Add a drop of reagent (2) and heat for several minutes	Reddish-blue color which fluoresces red under UV radiation	Conversion into barbituric acid	50
Nitrobenzene	(2) Pyridylpyridinium dichloride in dimethylformamide (1) Zinc dust in slightly acidic medium  (2) Alkaline phenol solution	In a microcrucible treat test solution with reagent (1) to reduce analyte to phenylhydroxylamine, which rearranges with acid to 2- and 4-aminophenol. Add reagent (2) to form indophenol blue	Deep-blue color		0.01

Continued

**Table 4** Continued

<i>Compound</i>	<i>Reagent</i>	<i>Procedure</i>	<i>Visible change</i>	<i>Remarks</i>	<i>Identification limit (μg)</i>
Oxalic acid	Solid diphenylamine	In a test tube evaporate sample to dryness and melt with reagent over a flame. After cooling add ethanol	Blue color	Aniline blue formation	5
Phenylhydrazine	Cacotheline in sodium hydroxide solution	Treat test solution with reagent on a spot plate	Blue product		0.02
Piperidine	(1) Solid sodium carbonate	In a test tube mix acidic test solution with reagent (1). Immerse in water bath at 90°C and cover the mouth of the tube with a disk of filter paper moistened with reagent (2)	Red spot on yellow paper		20
Pyridine	(2) Aqueous sodium 1,2-naphthoquinone-4-sulfate Trisodium pentacyanoaminoferrate(III) in pH 6.5 buffer	Place sample on filter paper and spray with reagent	Yellow–orange color within 5 min	Replacement of the ammonia with pyridine in the coordination sphere of iron(III)	10–50
Salicylaldehyde	70% perchloric acid	Introduce the anhydrous test solution into a capillary tube, and dip into the reagent	A yellow color appears which turns red when standing in the meniscus	Protonation of the carbonyl oxygen	2
Urea	(1) Urease  (2) Nessler solution	On a spot plate treat the neutral or alkaline test solution with reagent (1) and after 2–3 min with reagent (2)	Brownish ppt	Enzymatic hydrolysis of urea	1



while in the second stage hydrogen peroxide/peroxidase oxidizes *o*-tolidine to a blue quinoidal product.

In another dry-reagent configuration in Diasti<sup>®</sup> (Ames Co.) the *o*-tolidine in the second stage of the reaction is replaced by potassium iodide, which is oxidized in the presence of peroxidase to form free iodine. Newer test strips contain hexokinase in place of glucose oxidase. A very sensitive spot test for Benedict-positive compounds is described, which uses 2,2'-bichinchoninate as a chromogen. This test is ~10 000 times more sensitive than the classical Benedict test.

Several companies are now offering plastic strips with eight or nine separate reagent areas affixed that fulfill in only 1 min the function of a routine urine analysis laboratory. These include N-Multistix<sup>®</sup> and Hema-Combistix<sup>®</sup> (Ames Co.), the Combur-9-Test, BMD Chemstrips, and BM 33071 glucose pad (Boehringer Mannheim), and the Rapignost (Hoechst AG).

To eliminate interpersonnel variations in interpretation of the dipstick colors, simple reflectance scanning instruments may be used. The Ames Clinitek Auto 200<sup>®</sup>, Glucometer II, and Seralyzer, the Urotron L R System, Reflotron, Rapignost Total Screen L, Reflolux II and Accu-Check II (Boehringer Mannheim), the Glucoscan 3000 (Lifescan, Inc., Mountain View, CA), the Super Action Analyzer SA 4220 (Kyoto Daiichi, Kogaku), the Cobas Bio Analyzer (BCL, Lewes, UK), the Y Si 23 A analyzer (Clandon Scientific, Aldershot, UK), and the Cheme-Trics Analyzer (Technicon, Tarrytown, NY) are all reflectance measuring instruments that irradiate the dipstick with polychromatic light and measure the reflected radiation.

### Protein in Urine

Next to glucose tests, protein tests are the most frequently performed tests in routine urine analysis.

The first spot test configuration for urinary protein was based on deproteinization and turbidimetric estimation of the coagulated protein. The BUM-INTFST<sup>®</sup> tablet (Ames Co.) contained a premeasured amount of sulfosalicylic acid; the amount of protein was visually estimated by the degree of turbidity of the solution in the MICROBUM-INTFST<sup>®</sup> (Ames Co.). There was no correlation what so ever with urinary pH.

The most important spot test for native albumin, widely applied today by manufacturers of diagnostic products, is based on 'the protein error'. Proteins and their decomposition products often interfere with colorimetric pH determination. The analytical use of

anomalies in color reactions carried out in the presence of protective colors was demonstrated by Feigl and Anger, but was only utilized by a commercial company 20 years later in a system applying dry-reagent technology (Albustix<sup>®</sup>, Ames Co.). This test exploits the fact that the blue aqueous solution of the potassium salt of tetrabromophenolphthalein ethyl ester becomes yellow on addition of dilute acetic acid. The blue can be restored by adding alkali. If, however, a solution or suspension of native albumin is added to a dilute blue solution, a great deal of acetic acid can be introduced without causing the blue to change to yellow.

### Bilirubin in Urine

There is a need for a simple, sensitive, and reliable bilirubin test that can be used routinely by the physician for the detection of bile in urine.

Barium chloride-impregnated dry paper strips are dipped in urine samples for 30–120 s and then spotted with two drops of Fouchet's reagent (10% FeCl<sub>3</sub>, 25% trichloroacetic acid in water). A positive test is indicated by the appearance of a green color.

### Nitrite in Urine

Testing of urine for nitrite is important because it can be used to detect urinary tract infection. The basis of the spot test used in the Nitur-Test<sup>®</sup> (Boehringer Mannheim) and in the Rapignost<sup>®</sup> strip system (Hoechst AG) is the classical Griess test.

In the strip formulation of the Nitur-Test<sup>®</sup> sulfanilamide reacts with nitrite in an acidic buffer medium to form a diazonium salt, which couples with 3-hydroxy-1,2,3,4-tetrahydrobenzoquinoline to give a red azo dye.

The nitrite test of the Merck Co. detects as little as 10 µg l<sup>-1</sup> nitrite. It uses filter paper impregnated with an aqueous solution containing methanol, 1–2 g of a diazotizable amine, 0.1–0.5 g of a coupling component, and 2–8 g of solid organic acid per 100 ml. The paper is dried and cut into strips and is specific for nitrite independent of urinary pH. The test reveals significant bacteriuria when the bacterial count reaches 1 × 10<sup>7</sup> ml<sup>-1</sup> of urine.

### Ethanol in Urine

The synergistic action of ethanol with other drugs makes the detection of alcohol important and urgent.

Alcoscan test strips contain alcohol oxidase and horseradish peroxidase with colorimetric reagents (3-methylbenzothiazolinone hydrazone hydrochloride and 3-dimethylaminobenzoic acid).

A 1–5 ml urine sample is placed in a 15 ml centrifuge tube and the test strip is suspended in the air

space above the sample. The stoppered tube is heated at 65°C for 5 min and the color development estimated from a color chart.

### Detection of Cystic Fibrosis

Cystic fibrosis (CF) is one of the most common of the genetic diseases that lead to early death. One of the classical clinical indications for the diagnosis of CF is an abnormally high concentration of sweat electrolytes.

An indicator system for CF consists of three filter paper strips impregnated with 0.1 mol l<sup>-1</sup> potassium dichromate. These are soaked in 0.05, 0.066, and 0.1 mol l<sup>-1</sup> silver nitrate solutions, respectively. The strips are covered with waxed paper for protection. The indicator system is attached to a suitable part of the body of the child for 15 min. The indicator strips are removed and read. A positive result is signaled by the decoloration of the originally red-brown papers. If none or only one indicator strip shows a positive reaction, the sweat chloride concentration is  $\leq 35$  mmol l<sup>-1</sup> and the child is considered normal. Children showing chloride concentrations of 35–55 mmol l<sup>-1</sup> (two of three strips reacting positively) need to be checked more thoroughly. Concentrations of  $> 55$  mmol l<sup>-1</sup> chloride (all three indicators reacting positively) indicate a high probability of CF.

## Forensic Applications

Of the objective physical evidence that a forensic scientist collects, a considerable amount is based on simple spot tests.

In homicides committed with a gun it is important to find out the shooting range. A great deal of corroborative evidence can be collected by detecting red blood cells and spermatozoa at the scene of a crime. The search for illicit drugs is greatly facilitated by selective spot tests in their preliminary identification.

### Firearm Discharge Residues

Reliable and sensitive methods for demonstrating the presence of residues of firearm discharge on the clothing of a victim or on the skin of a suspect are sometimes qualitative tests. The detection of nitrate and nitrite in such residues can be achieved by spot tests. The detection of traces of antimony and barium, which are present in many cartridge priming charges, on the web of the thumb and back of the hand, can supply data relevant to the question of whether a head wound, resulting from being shot with a revolver, was self-inflicted. The relatively recent discharge of a firearm is indicated by the detection of the most frequently encountered metallic components of firearm discharge residues, such as lead, copper, nickel, barium, and antimony.

The classical Griess reaction and its numerous modifications is a specific nitrite test, but the common obstacle faced by criminologists is the presence of blood. The color of the blood and that of the red azo dye produced by the Griess reaction are similar. Detecting nitrite by fluorescence under UV radiation solves this problem. Whatman No. 1 filter paper is soaked in a freshly prepared 5% solution of *o*-phenylenediamine dihydrochloride and partially dried. This filter paper is placed over the suspected area of the gunshot on the clothing, exerting slight pressure for better contact with all parts of the clothing. The paper is removed and placed in a Petri dish containing a 25% ammonia solution. The paper is illuminated with 254 nm UV radiation. The yellow fluorescence pattern may be photographed as permanent evidence.

### Simple Drug Tests

As the problem of drug abuse reaches epidemic proportions, the use of simple tests gains considerable importance. In recent years, various instrumental methods for identifying drugs by their physical properties have been greatly improved. These are enormously useful, but are outside the scope of this article; we are only concerned here with quick chemical methods feasible for field conditions. Some representative examples of such tests are presented below.

**Opiates** Several drops of Marquis reagent (8–10 drops of 40% formaldehyde in 10 ml of concentrated sulfuric acid) are added to 1–2 mg of drug. The presence of opiates is indicated by the appearance of an intense purple turning to blue-violet.

A simple and effective separation of opiates may be achieved by thin-layer chromatography (TLC). On silica gel 60 GF plates, CHCl<sub>3</sub>-hexane-triethylamine (9:9:4; v/v) is applied as the mobile phase.

Illicit heroin (diamorphine) may be screened by dissolving the sample in methanol and applying this to silica gel 60 GT 254 plates. The chromatograms are developed with toluene-acetone-ethanol-diethylamine (30:60:7:3; v/v) and the spots are visualized by UV irradiation at 366 nm.

**Detection of barbiturates in urine** A few drops of Zwikker A (1% cobalt(II) acetate solution) are added to a few milligrams of drug, followed by a few drops of Zwikker H solution (5% isopropylamine in methanol). The presence of barbiturates is indicated by the appearance of a blue-violet color. Blood and urine samples are analyzed by TLC on silica gel G with CHCl<sub>3</sub>-acetone (17:3; v/v) as mobile phase.

Visualization is effected by spraying with diphenylcarbazone– $\text{HgCl}_2$  reagent followed by methanolic 1% KOH.

Components in drug mixtures may be simply identified by dissolving the sample in 2 ml of ethanol, filtering, evaporating the filtrate to 0.5 ml, and applying 5  $\mu\text{l}$  portions to a silica gel–gypsum layer for development with ethyl acetate–ethanol–aq. 25%  $\text{NH}_3$  (18:2:1; v/v). Iodine vapor is used to detect phenobarbitone, bromvaletone, papaverine hydrochloride, and caffeine.

**Screening for LSD** Ehrlich's reagent produces a characteristic purple to deep-blue color with lysergic acid diethylamide (LSD).

**Screening for amphetamines** A highly sensitive screening strip for methyl amphetamine in urine may be prepared using the Simon reagent. A silica gel TLC plate is sprayed with equal volumes of 2% sodium nitroprusside and 5% aqueous sodium carbonate solution, and dried for 3 h at  $110^\circ\text{C}$ . A strip is cut from the plate to use in the test.

Urine (alkaline, 5 ml) is extracted with 2 ml of benzene and a 20  $\mu\text{l}$  portion of the benzene layer is spotted onto the strip, which is then exposed to acetaldehyde vapor; a blue color indicates the presence of  $>0.1\text{ }\mu\text{g}$  of methylamphetamine.

The same Simon reaction may be used for rapid screening of methylamphetamine in urine in a Sep-Pak  $\text{C}_{18}$  cartridge.

**Screening for marijuana** The most widely used illegal drug is marijuana, which consists of dried parts of the *Cannabis sativa* plant. Its concentrated, more potent, form is hashish, the resin form of drug which contains  $\sim 30$  derivatives.

In the optimized rapid method 5 mg of hashish or 50 mg of marijuana is extracted with light petroleum–diethyl ether (4:1; v/v). The extract is evaporated to dryness, the residue dissolved in 0.5 ml of 1% 4-dimethyl-aminobenzaldehyde in concentrated  $\text{H}_2\text{SO}_4$  and the solution is heated for 3 min in a water bath. A purple–red color is produced immediately, turning to indigo blue on pouring the mixture into cold water. This dual color reaction is specific for extracts of hashish and marijuana and can be used to distinguish them from other plant extracts.

**Screening for salicylates** Overdoses of aspirin and other mild analgesics are extremely common. Salicylate is commonly detected in urine by the classical violet iron(III)–salicylate chelate formation.

The same simple test is used for the detection of salicylate in hemolysed whole blood; 0.3 ml of blood

is mixed with 0.6 ml of methanol and the mixture is centrifuged at 3000 rpm for 5 min. The supernatant solution is filtered and 50  $\mu\text{l}$  of 0.5% iron(III) chloride solution is added to 0.2 ml of the filtrate. A violet color at the diffuse interface indicates the presence of salicylate.

**Screening for blood** Chemical testing for blood or traces of blood is often needed to provide physical evidence in cases in which direct visual inspection is not a decisive proof.

Blood is identified in most cases on the basis of the peroxidase-like activity of the heme group of hemoglobin. In the classical blood test, benzidine is oxidized to the quinoidal benzidine blue by hydrogen peroxide in the presence of hemoglobin. Since benzidine and its salts are now considered to be carcinogenic, 3,5,3',5'-tetranethylbenzidine is used instead.

**Screening for semen** The high incidence of sex crimes necessitates the rapid identification of semen. The male germ cells contain acriflavine, the presence of which has direct diagnostic value.

At high dilutions acriflavine fluoresces under UV radiation; this offers the basis of a very simple test for semen. Unfortunately, modern laundry brighteners also fluoresce very intensely.

Semen is generally detected by identifying acid phosphatase or choline. The suspected stain is cut and incubated for 30 min at  $37^\circ\text{C}$  with two drops of a 0.1% aqueous solution of disodium phenylphosphate; the phenol resulting from the enzyme-catalyzed hydrolysis can be detected with the Folin–Ciocalteu reagent. Phenol forms molybdenum blue on reaction with the molybdophosphate or tungstophosphate incorporated in this reagent.

Choline can be identified by a microscopic crystal test. The aqueous extract is treated with the Florence iodine reagent (6.6 g potassium iodide and 10.6 iodine dissolved in 120 ml distilled water). Characteristic rhombic crystals are observed.

## Water Quality Screening

In 1974, the Safe Drinking Water Act was passed by the US Congress and the Environmental Protection Agency (EPA) took responsibility for ensuring the safety of water supplies. The EPA proposed a list of materials that carry potential hazards when present in drinking water. Maximum allowable levels were set for more than a dozen inorganic species (As, Ba, B, Cd, Cr(VI), Cu,  $\text{CN}^-$ ,  $\text{F}^-$ , Pb,  $\text{NO}_3^-$ ,  $\text{NO}_2$ , Se, Ag) and for about the same number of organic pesticides. Turbidity and the amount of coli-form bacteria were also limited. Additionally, secondary

regulations were issued in 1977 dealing with the esthetic qualities of water. These are connected with the sulfate, iron, and manganese contents of drinking water.

Spot tests based on simple kits for quick and inexpensive water testing still find wide application and a number of commercial companies market variations of the test kit methods that are approved by the EPA for Safe Drinking Water Act reporting and recommended for onsite testing.

The reliabilities of spot test-based water analysis kits for ecological purposes have been evaluated. Statistical analysis of the results for a large series of samples showed that for routine purposes there was good agreement ( $r > 0.85$ ) with the results of standard methods. When high accuracy and precision were needed, however, the kits were unsuitable.

**See also: Amphetamines. Blood and Plasma. Clinical Analysis:** Glucose. **Fluorescence:** Quantitative Analysis. **Forensic Sciences:** Blood Analysis; Gunshot Residues; Systematic Drug Identification. **Heroin. Proteins:** Physiological Samples.

### Further Reading

Asselin WM and Caughlin JD (1990) Rapid and simple color test for detection of salicylate in whole hemolysed blood. *Journal of Analytical Toxicology* 14: 254–255.

Baudot P and Andre JC (1986) Identification and qualitative determination of LSD by fluorescence. *Analytical Abstracts* 2E44.

Burns DT, Townshend A, and Carter AH (1981) *Inorganic Reaction Chemistry*, vols 1 and 2. Chichester: Ellis Horwood.

Chen Q, Klemm N, Duncan G, and Jeng G (1990) Sensitive Benedict Test. *Analyst* 115(1): 109–110.

Feigl F and Anger V (1966) *Spot Tests in Organic Analysis*, 7th edn. Amsterdam: Elsevier.

Feigl F and Anger V (1972) *Spot Tests in Inorganic Analysis*, 6th edn. Amsterdam: Elsevier.

Jungreis E (1985) *Spot Test Analysis. Clinical, Environmental, Forensic and Geochemical Applications*. New York: John Wiley.

Jungreis E and Ben-Dor L (1980) Organic spot test analysis. In: Svehla G (ed.) *Comprehensive Analytical Chemistry*, vol. X. Amsterdam: Elsevier.

Nair NK, Navaratnam V, and Rajananda V (1986) Analysis of illicit heroin (diamorphine). *Journal of Chromatograph* 366–372.

Scheithauer R and Luta C (1988) Comparison of practical tests for detecting semen stains. *Archiv für Kriminologie* 181: 41–49.

Syoyama M, Oyama N, Sano T, and Maeda K (1987) Preparation of a highly sensitive screening chip for methylamphetamine in urine by use of Simon reagent. *Analytical Abstracts* 2D83.

Verebey K and Mule SJ (1986) Quick qualitative assay for alcohol in urine. *Journal of Analytical Toxicology* 10: 167–168.

## STABLE ISOTOPE RATIO MASS SPECTROMETRY

**See** MASS SPECTROMETRY: Stable Isotope Ratio

## STANDARDS

**See** QUALITY ASSURANCE: Primary Standards; Spectroscopic Standards; Internal Standards

## STATISTICS

**See** CHEMOMETRICS AND STATISTICS: Statistical Techniques

## STERILANTS

**See** BLEACHES AND STERILANTS

# STEROIDS

See **HORMONES: Steroids**

# STM

See **MICROSCOPY TECHNIQUES: Atomic Force and Scanning Tunneling Microscopy**

# STRUCTURAL ELUCIDATION

**M Novič and J Zupan**, National Institute of Chemistry, Ljubljana, Slovenia

© 2005, Elsevier Ltd. All Rights Reserved.

## Basic Principles

In analytical chemistry, many problems require completely different approaches or different type of automated computer methods in order to arrive at the desired solutions. One very important step is separation or isolation of the compound; however, we will begin from the stage in which a compound is available in a pure form. The problem we shall focus on in this article is structure elucidation based on different spectroscopic methods.

The problem of structure elucidation has a very broad range of goals: from the identification of only few constituent structural fragments of the unknown compound to the complete three-dimensional structure determination. The problem of establishing reliable correlations between different types of spectra (infrared (IR), ultraviolet, mass,  $^1\text{H}$  nuclear magnetic resonance (NMR),  $^{13}\text{C}$  NMR, etc.) and the chemical structure (or structural fragments) are diverse and have not yet been completely solved. In order to shed light on the prospects for solving this problem, some of the main difficulties and encouraging developments are summarized below.

The main difficulties are:

- The number of registered chemicals is over 20 million (Tuesday, March 4, 2003, at 05:34:17 there were 21 108 454 organic and inorganic substances registered; Monday, April 14, 2003, at 10:51:24 the number increased to 21 287 895). Approximately

4000 new substances are added each day to CAS Registry. Spectral data exist for only a minor portion of these compounds; the computer-readable formats, if available, are not standardized and thus difficult to exploit.

- The spectra held in collections may contain significant experimental errors, which are dependent upon the laboratory, spectrometer, source of material, sampling, experimental conditions, purpose of the measurement, and the operator.
- The number of theoretically possible structures even for a modest set of fragments may be prohibitively large. Extremely fast growth of the number of possible structures with an increasing number of fragments, called combinatorial explosion, virtually precludes their exhaustive generation and back-checking with spectral simulation.
- The logical rules for linking spectral features with the structural ones and vice versa are so diverse (dependent on very large structural or spectral parts) and so numerous that the decision between ‘encoding all, even minute rules’ and ‘encoding general rules and all exceptions from these rules’ has to be made. Since many rules or correlations are not even known it is practically impossible to collect them all.
- The level of cooperation between the human experts from different fields (chemists, spectroscopists, programmers, artificial intelligence researchers, psychologists, etc.) required to build a good chemistry-oriented expert system is in most cases far too low.

The list of problems could be even longer, but there are some encouraging developments that may aid the solution of spectra–structure correlation problems that should be mentioned as well. In some areas the

successes have been remarkable. Among the most encouraging factors are the following:

- Small changes in spectral (structural) features have been shown to be caused by small changes in the solutions sought, i.e., the structures (spectra). Such analogies can be widely used.
- Computer methods for the reduction of measurement spaces, such as fast Fourier and fast Hadamard transforms, principal component analysis, autocorrelation, genetic algorithm, and others, can be applied successfully.
- A number of software packages for accomplishing specific tasks (generators of all isomers, interactive structure editors, various spectra simulation programs, etc.) are available. Different input and output formats of spectral and structural data are often an optional choice. Besides, many databases are available online through the internet for browsing structural and spectral data.
- A number of heuristic rules linking structural fragments with parts of spectra and vice versa are known.
- In the laboratories performing routine analysis (analytical work, quality control, violation of regulations control, etc.) there is a considerable interest for automating the routine part of the work.
- A rapid increase in computational power accompanied by the lower price of hardware and more powerful software available on the market has aroused interest in using calculation-intensive methods for everyday applications.

The factors listed above are mostly unrelated and may not be pertinent to the solution of every problem of this kind. A thorough study of the suitable methods for learning, data reduction, feature extraction, or decision-making processes has to be made separately for each problem.

In general, methods of solving the spectra–structure correlation problems are based on three different approaches: on library searches, on expert systems, and on various modeling techniques.

Library searching techniques are of two kinds: identity searches and similarity searches. Identity searches (i.e., searches requiring an exact match between the query and the reference) are successful only if the sought spectrum or structures are ‘already’ in the reference library. If not, the search fails completely. In contrast, similarity searches depend on the metrics introduced into the representation of the spectra or structures. The metrics and consequently the distance or the similarity measure between the query object and the reference depend on the problem being studied; hence, the results obtained on the basis of similarity are not directly comparable when the problems are from different fields.

Another approach to solving structure elucidation problems is to use the ‘strategy’ of the expert system. This strategy consists of the logic rules-driven process that combines a broader knowledge base (including the logic rules) with the methods and procedures required to make partial or final conclusions. With this combination, systems capable of complex decisions in the structure elucidation process can be created. The best spectroscopic applications on expert systems are in the fields where the problems are well defined (large number of structure–spectra relationship logical statements) and the methods or procedures as to how to move from the description of the problem to the desired solution are established and formalized (hierarchy of rules and/or forward and backward chaining of questions and answers, etc.).

The third possible approach to the structure elucidation processes is to find procedures, the so-called modeling techniques, which are capable of making hypotheses on basis of which problem can be solved. For example, a set of procedures should be able to predict from spectral data all structural fragments, from which a structure generator could yield all possible structures. The resulting structures are then ranked and evaluated. In structure elucidation systems based on modeling techniques, the search procedures are only partially built on rules or other information provided in advance in a knowledge base. The solutions are learned from real examples rather than by using fixed rules. Such learning procedures are known as self-learning methods.

In practice, structure elucidation processes are very complex and for determining different specific structural features any of the three mentioned approaches can be used. It is possible to integrate the library searches and/or various modeling techniques into expert systems, or vice versa.

## **Use of Databases or Library Systems**

The success of a library search in analytical chemistry depends very much on the data representation. Different kinds of data (numerical, textual, spectral, structural) require appropriate treatments. Usually, spectral data collected from the instruments are not in a form suitable for direct input either into a collection or as a query for a library search. Once the spectrum is in the computer, a number of preprocessing steps (e.g., smoothing, baseline correction, normalization, peak and intensity detection, reduction of the measurement space, autocorrelation, deconvolution, shape tracing, etc.) should be applied in order to bring the data into a standardized format



and to extract the key features. The preprocessing steps must be applied selectively, bearing in mind that too extensive a use of preprocessing methods may cause a considerable loss of valuable information, thus jeopardizing the possibility of reaching a final successful solution. On the other hand, when starting to build a database of high quality, as many preprocessing methods as possible should be implemented to make the handling of the databases flexible and the system versatile.

The structure of the chemical compounds under investigation influences all aspect of its chemistry, particularly spectroscopy. Therefore, the handling of chemical structures is one of the basic features that have to accompany the library search. For the input of chemical structures into a computer, flexible on-line editing of chemical graphics is required.

All input spectra and structures must also be saved in direct access files for later use in the iterative process or for checking the results. Different data types (structures, spectra, sets of commands) must be coded uniformly in all system modules for input and output. The same is true for the files on which the final or intermediate results are stored. Because a resulting output file of one module may act as an input file to another one, or be used as an input for starting a new problem, all data exchange processes and files must be standardized. Good library search systems perform the reformatting of data transparently to the user.

Other types of possible input such as numbers and textual requirements that serve as supplemental information to the basic data are handled by special-purpose decoders that, unfortunately, are almost always custom designed.

Library systems designed to retrieve references identical to the query are called identity search systems. Since the identity of any object (spectrum, chromatogram, chemical structure, set of chemical features) is well defined, the identity search is reduced to the binary decision identical/not identical. The system has to be insensitive to varying instrumental conditions, but there is no need to discriminate between different degrees of similarity. The identity search system is useful only in a limited range of applications; for example, in the field of environmental control or drug control where the set of compounds that are expected or have to be identified are exactly specified. Any other, unspecified compound is irrelevant. In most other applications the unexpected compounds are of prime interest, so that the identity search is of limited use.

For the research and development purposes where the emphasis is on new compounds not yet contained in the libraries, similarity search systems able to

retrieve compounds similar to the unknown have been developed. The measure of similarity is related to the distance between two objects, and therefore depends on the object's representation, on the choice of metrics for the measurement space, and, consequently, on the choice of a mathematical expression for the distance calculation. If the similarity search is applied to spectra, it is not too difficult to calculate the similarity. The similarity search becomes more complex when the queries are chemical structures. At present, there is no generally accepted unique similarity measure for chemical structures. Different spectroscopic methods focus on different structural entities and so are concerned with different aspects of structural similarity.

Because identity is the upper limit of similarity, similarity search systems may always be used for identity searches. However, when used as an identity search, similarity search systems are not very efficient because the requirements for checking the similarity introduce lengthy algorithms to the system, which may use conflicting criteria and thus resulting in compromise. In order to distinguish identical compounds from not-identical ones, in a similarity search, the degree of similarity has to be defined in such a way as to give a significant difference between compounds with slightly different chemical structures, and a small difference between samples of identical chemical structures. On the other hand, the system should be tolerant of slight impurities in the sample of the unknown. It is very important to consider in a search system: (1) the selection of spectral features that are structure sensitive, and (2) how to normalize and scale the similarity measure. In order to make the correct choices, knowledge and experience in the particular spectroscopy are required.

## Expert Systems

### General

Expert systems are expected to solve problems always in close 'collaboration' with the user and not entirely by themselves. Dialog between the user and the system can lead to much better performance than if the systems were left by itself to work on the problem automatically. Expert systems help the user to solve problems; the more extensive and complex the help, the better the expert system.

There are many different opinions as to when a set of programs and data can be considered to be an expert system. A computer scientist working on the development of an expert system would favor the following definition:

Expert system = Knowledge base + Inference engine

where the inference engine interprets and exploits the knowledge base, which can be in many different forms (rules, frames, logic rules, tables, formulas, etc.). The 'inference engine' is usually separated from the knowledge base in such a way that the knowledge base can easily be changed without changing anything in the inference engine.

In the language of artificial intelligence, the expert system is able to recognize each possible situation (state) in the problem space and respond with an appropriate action, which will lead closer to the final solution. For example, on the basis of spectral features the system must be able to select possible structural fragments. One possibility is that the rules for the spectra-structure correlations are built into the inference engine by an expert. Another possibility is that the modules for clustering the spectra from available spectral collections provide automatic decisions about some spectra-structure correlations. Spectra in the same cluster have common spectral features. These features must be assigned to structural fragments common to all compounds having spectra from the cluster investigated.

It should be emphasized that expert systems are not expected to make fundamentally new discoveries; it is enough if they are able to identify situations, partial solutions, or ways toward a solution that were not known before, or at least that the system designers had not thought of in advance. It is very important that the expert system informs the user about the new situation when it is discovered, and demands from the user the explanation and instructions as to how to react or what to do in the particular case and in similar cases in the future.

The main components that must be known, accessible, or computable during the planning, developing, and testing an expert system are as follows:

- description (explicit or implicit) of possible states that can appear in the problem space;
- description (explicit or implicit) of possible solution states;
- means for the recognition of states;
- criteria for the evaluation of states;
- answers (reactions, ways) leading from the problem toward the solution;
- criteria for the evaluation of answers;
- means for the selection of the best possible answer (way); and
- means for the construction of new or intermediate states in the problem space.

Some of the above components might be omitted and others added. The description of states and different criteria for evaluations can be regarded

as a knowledge base, while the recognition of states, generation of answers, and the selection of the best answers can be regarded as parts of the inference engine. However, it is often hard to draw a clear line between both components.

The main guideline along which the expert systems work can already be recognized from the above components, which, of course, can be (and most often are) performed in an iterative way. From the initial state of the problem, the expert system must come to the final state (the solution) via a number of intermediate steps.

### **Acquiring the Knowledge**

Knowledge is the most important part of any expert system. The simplest way of implementing the knowledge is to transfer it to the expert system in one of the following forms:

- known facts (constants);
- mathematical formulas;
- tables or collections of data;
- description of representatives (frames);
- procedures (algorithms);
- logic rules (if 'condition' then 'action'); and
- heuristic rules.

For the designers of expert systems the logic and heuristic rules are regarded as the most interesting. Because branching is a fundamental property of logic rules, a collection of rules becomes very hard to follow, maintain, correct, and change when the number of rules increases. Additionally, if the solutions of the real-world problems are to be described with a tree of logic rules, the numbers of rules and exceptions to them (which are also described and resolved as logic rules) may become so immense as to become out of control. To avoid this problem, heuristic (experience based) rules and fuzzy rules are introduced. The difficulty initiated by the use of heuristic and fuzzy rules is not in their formulation, nor that they would not hold for certain, but in the fact that in most cases even the percentage of the confidence limits is not exactly known. In the fuzzy logic approach mathematical 'probability' is not appropriate and is replaced with an alternative theory called 'possibility' theory. The idea behind fuzzy systems is that values for 'truth' and 'false' are not only 1 and 0, respectively, but any value between 1 and 0. Fuzzy 'truth' values deal with the 'likelihood' or 'probability' that a fact or rule is true.

To handle large numbers of logic, heuristic, and fuzzy rules, a variety of supporting software (expert systems shells, languages specially designed to handle logic, etc.) has been developed and is now available

'off the shelf'. But in spite of all support offered, the basic scheme of rules must first be obtained from a human expert and then input by the designer into the expert system. Very often even the best human experts are not able to describe or clearly define the logic routes for solving a given problem. This obstacle in designing and programming of expert systems is called the 'Feigenbaum bottle-neck'.

Human experts can be aided and their work complemented by the use of self-adapting and self-learning procedures that can acquire knowledge by themselves. Two strategies have been developed for self-learning systems: supervised and unsupervised learning. If the system learns, i.e., acquires knowledge, with the aid of its tutor or by itself, the learning is called supervised or unsupervised, respectively. For both kinds of learning the selection of the database is very important. The designer of an expert system must use only collections of high quality to form the basis of self-learning procedures. It is also important to choose a qualified tutor, preferably a human expert in the field for which the computerized system is intended and not the designer of the expert system. The expert's task is to explain the meaning of the data which the computer system has selected from the partial solutions obtained at various stages (intermediate states) of the problem-solving process. The system first inspects the data, finds certain patterns, classes, clusters, or groups within them, and where it is not able to interpret the pattern, it requires the tutor's description of the meaning (or influence) of the patterns on the general solution. The tutor's answers are interpreted and connected to the already existing knowledge. In short, learning from examples rather than by following rules is favored in cases where the rules are difficult to write down explicitly.

### Neural Networks in Chemistry

The problem of establishing reliable correlations between different types of spectra (IR, mass,  $^1\text{H}$  NMR,  $^{13}\text{C}$  NMR, etc.) and the chemical structure or structural fragments of the corresponding compounds is so elaborate that a complex approach is needed. The neural network approach offers a valuable possibility to develop an automated system capable of dealing with such complex relations. The main effort of neural network research is training the network to obtain the model. After the model (neural network in our case) is obtained, it acts as a 'black-box' that provides reliable answers when a set of measured values (input variables) is fed into it. Although the training might be quite lengthy, the answer or prediction of a trained neural net is almost instantaneous,

especially if the net is implemented on a parallel computer. As a rule, the model is built on the available set of data such as a spectroscopic database, a set of historical recordings of process control, protein sequences, a set of multicomponent analyses, etc. Sets of data containing a number of interesting dependencies, feature correlations, and other information that cannot be deduced in a straightforward manner from first principles, by theoretical calculations, or even with numerical methods, are the best targets for neural network applications. Among all different types of neural networks, the feed-forward neural network based on the 'back-propagation of error' learning algorithm has been applied most frequently for structure-solving problems. The next very successful algorithm worth to be mentioned is self-organizing algorithm, originally used in Kohonen neural networks, which produces 2D self-organized maps (SOMs) of input objects. The information contained in 2D SOMs can be exploited for clustering, classification, variable selection, outlier detection, etc. The SOMs can be also employed into counter-propagation neural networks, which are based on initial self-organizing algorithm and extended by supervised learning algorithm yielding the response surface used for predictions.

### Communication between the User and the System

A vital part of the expert system that is too often neglected by the designers and by the researchers of the expert systems is the communication between the expert system and the user. Good communication provides the information about the point in the decision process from where a particular dialog starts; the user should understand the implications of his or her answers on the progression of the elucidation process and its final solution. To achieve this, the communication system must employ appropriate means such as menus, graphics, keyboard typing, pointing devices, help messages, links to other resources, and explanations. In a good expert system, the wording, the order, and the form of questions are adjusted to the level of a chemist, not to the level of a computer specialist. The online help must be available from all communication points of the system. Both general and context dependent help are better if supported with examples.

### Modeling Techniques for Special Methods and Procedures

#### General

The expert systems that are able to design and evaluate hypotheses can be regarded as models for solving

special problems. We shall focus here on spectroscopic systems with the limitation that they will be dealing with pure compounds only. The problem of mixtures (first qualitative identification of components and next quantitative determination of their amounts in the mixture) is far too complex to be solved for the general case. However, a well-designed spectroscopic expert system can offer a number of possibilities (structural features, combination of fragments, correlation of spectral parts with different types of compounds, etc.) if the spectra of a mixture are input.

A commonly used graphic presentation of such systems is a path of circular shape indicating a cyclic procedure of postulating and optimizing hypotheses. The process begins with the input of a spectrum of the unknown sample with all additional available data (known fragments, mass formula, constraints, chemical and physical constants) and finishes with the proposed structure and corresponding simulated spectrum (property). Between the input and output steps various intermediate procedures are necessary. The input data serve to predict one (or several) possible chemical structure(s) of the unknown sample. The simulation techniques are then applied to the proposed chemical structures and simulated spectra are compared with the original (input) spectrum. The predicted structure giving the simulated spectrum most similar to the original one is the most probable solution for the unknown structure. The whole procedure is shown in **Figure 1**.

This circularly designed system must allow the user to start his or her problem at different points of the solution-seeking process. The user is not always interested in the entire process. Often, the user has a partially solved problem and would like to know just a minor detail or alternative solution if it exists. Sometimes users have very simple problems such as searching a given spectrum, fragmentation of a structure, building a small database according to specific needs, or modifying the instrumental output (correcting baseline, removing noise, deleting part of the spectrum). Because the tools for solving minor problems are already available, each spectroscopic system should allow many different types of data as input and offer access to individual modules at user's will. In general, the solution-seeking process is two-directional, or, if looking from the algorithmic point of view, the process is cyclic and it allows the flow of data in both directions.

### **Input of Data**

The complex modularly built systems should be able to work with different types of data as well as with

the incomplete or corrupted data. The user may have only partial spectra or a few fragments. Handling different types of input must be designed carefully to enable the user to concentrate on the problem and not on data typing.

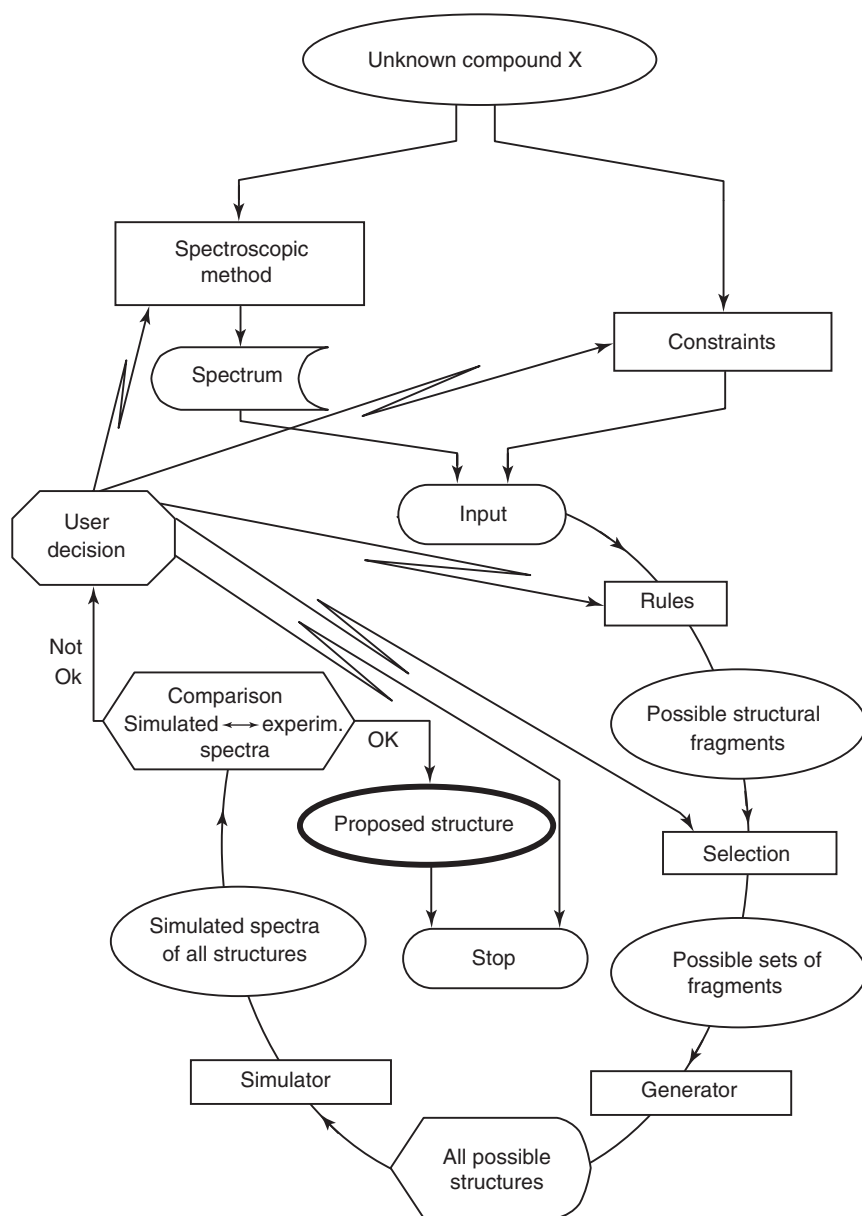
In both directions of the solution-seeking process (from the structure as input toward the spectrum, or vice versa) the system first transforms the input data into the representations it understands, i.e., the representations that enable the algorithms to be performed. The most standard spectral data are the peak positions, intensities, and shapes of the specific parts of the spectral curve, the area below the curves and the periodic patterns of peaks in the spectra. The main structural features are the spatial and topological relations of structural fragments with the aspect to the rest of the structure. Topological properties are deduced from connection tables, while spatial positions, orientations, and possible degrees of freedom of rotation or vibration of structural parts are usually obtained in an interactive online dialog with the user. However, the system must be able to detect the positions in a structure where certain spatial effects may occur and ask the user for detailed instructions if the questions cannot be resolved due to lack of data.

### **Building the Hypothesis**

Before the problem of structure elucidation can be completely solved, possible pieces of the unknown spectrum or structure must be obtained, from which a model for the solution can be constructed. Many different methods and processes are applied to obtain the building blocks of the solution. Among the most important ones are logical and heuristic rules of the type:

- If spectrum of type A contains feature B, then structural fragment C is present and D is absent.
- If structure A contains fragment B, then spectrum of type C will show feature D.

As was mentioned earlier, artificial intelligence methods such as pattern recognition, supervised and unsupervised learning, or neural networks can be successfully applied for building the hypothesis. The recognition of common patterns in spectra belonging to the same type of compounds and clustering of spectra in order to find common structural features are two main problems in the field. The key point is always the same: to correlate spectral features to the structure and then the structure to the sample's properties. At the hypothesis-building stage as much information as possible should be extracted from available data to achieve a good starting point. The



**Figure 1** Scheme of a spectroscopy oriented expert system. The data available on the unknown compound X (spectrum, constraints) form the input. All subsequent procedures are presented in rectangles, while intermediate results are shown in ellipses. At the point where the simulated spectra are obtained, they are compared with the original input spectra and the whole procedure stops if the result is satisfactory. If not, the user has the possibility of influencing certain steps (use another spectroscopy, add other constraints, change or add new rules, or make another selection of possible set of structural fragments), which is indicated by zigzagged arrows. (Reprinted with permission from Tušar L, Tušar M, Bohanec S, and Zupan J (1992)  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra simulation. *Journal of Chemical Information and Computer Sciences* 32: 299–303; © American Chemical Society.)

advantage of modeling systems designed for special purposes compared with classical expert systems is in a certain freedom in choosing possible solutions because of the feedback that rejects unreasonable results when the hypothesis is tested and evaluated.

The system must detect the cases when it is not able to produce a working hypothesis or when the hypothesis that has been created is causing the process to diverge ('explode'). If this happens, the system

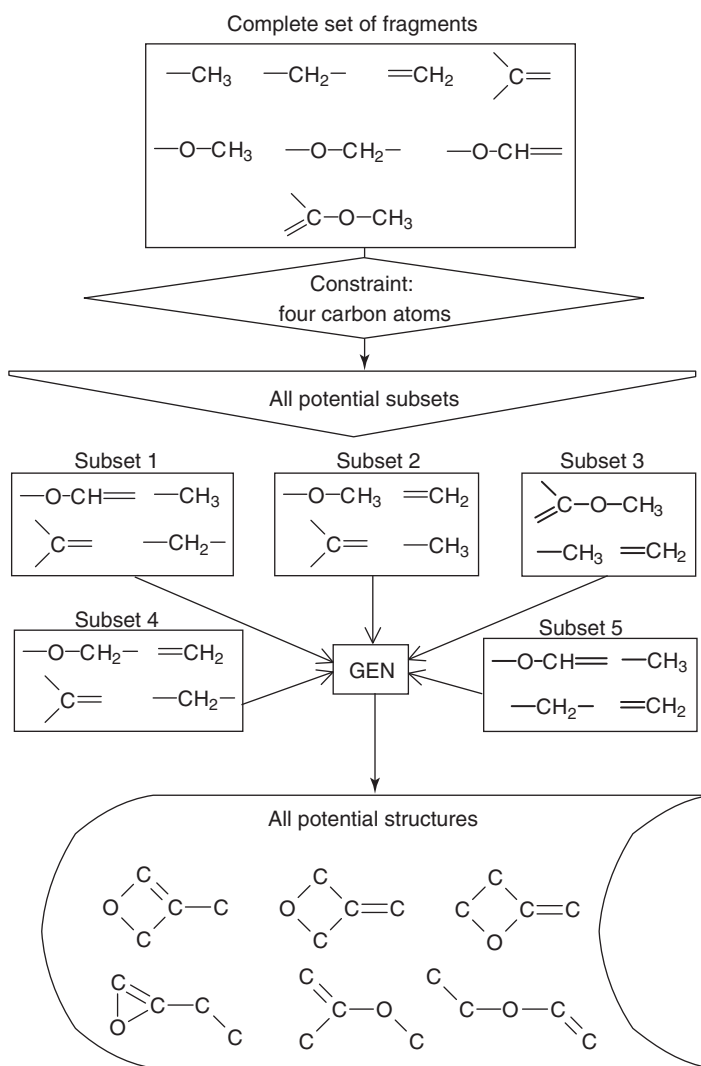
has to ask the user for more data constraints, for help in the selection of priorities among available hypotheses, and must accept suggestions from the user. Such an interaction from the user's side calls for an expert in the field and must be carefully recorded by the computer for possible feedback to the system in order to increase its knowledge.

The first complex procedure indispensable for all chemistry- and spectroscopy-oriented systems, as

well as for the procedures designed for special purposes, is a structure search. It can be implemented in three forms, i.e., as a substructure, as a superstructure, or as a maximal common structure search. If in the substructure search the input fragment is a complete structure, i.e., with no possible continuations given, the result can be either an exactly identical structure or no structure at all. A significant portion of the chemical and mathematical knowledge is built into the structure generator, or more precisely into the part of it that selects the constraints limiting the type of structures considered in the generation. The structure generator is a procedure that 'assembles' all topological possible structures from a given set of data: molecular formulas, set of fragments, etc. The problem may lead to an enormous number of

combinations (combinatorial explosion) that is computationally so intensive that, with the exception of the simplest molecular formulas, it is not possible to generate all potential structures. Therefore, the tree of structures has to be inspected only through the branches that represent the structures compatible with constraints given by the user (e.g., by considering only five- and six-membered rings, or no aromatic rings).

An indispensable procedure in such systems is the simulation of spectra from a known structure. The simulation of spectra can also be implemented as a stand-alone system. There are two different ways for solving the problem of simulations. The first is an *ab initio* calculation (from fundamental physical laws), while the second one is a heuristic or



**Figure 2** Selection of all possible subsets from the complete set of structural fragments and the resulting possible structures generated by the structure generator 'GEN'. (Reprinted with permission from Tušar M, Tušar L, Bohanec S, and Zupan J (1992)  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra simulation. *Journal of Chemical Information and Computer Science* 32(4): 299–303; © American Chemical Society.)



experience-based approach. Because of the excessive use of computer resources in the majority of cases, the *ab initio* approach is not suitable for implementation into the systems described in this section.

For the above reasons, the simulations are mainly performed in a heuristic and approximate way, using heuristic rules, statistical data, tables, semiempirical formulas, etc. For example, simulations can be made satisfactorily in various types of NMR spectroscopy, where long-range effects have much less influence compared to IR spectroscopy. IR spectra are very hard to simulate due to the 'spread' of vibrations through the entire structure. An interesting simulation technique is the determination of 'spectral contributions' of all fragments and atoms from large specialized collections of spectra with assigned structures. For the extraction of contributions a substructure search algorithm, which can search for a set of slightly different structural fragments, is required. The quality of the simulated spectrum obtained this

way depends on the size and quality of the collection for a particular spectroscopy. If the collection contains a structure very similar to the query, it is expected that the spectrum is similar as well and the available contributions are sufficient for the simulation of a good quality.

### Evaluation of the Hypothesis

Evaluation of the hypothesis, which may include simulated, retrieved, or calculated spectra, structures, and other features, is based mainly on the experience of human experts who set up the criteria and the coefficients according to which the evaluation is calculated. As discussed above, the distance between two spectra is relatively easy to evaluate, in contrast to the distance between two structures, which is very problem-dependent. In practice, we are looking for the relative similarity: which structure from the list of structures proposed in the hypothesis

**Table 1** Simulated  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of six possible structures

Structure no.	Structure	$^1\text{H}$ NMR spectrum (ppm)	$^{13}\text{C}$ NMR spectrum (ppm)
1		3.55, 3.55, 5.22, 5.22	74.2, 74.2, 106.5, 149.1
2		2.05, 3.55, 3.83, 3.90	39.4, 60.9, 78.4, 164.8
3		1.80, 4.35, 6.23	13.4, 74.1, 110.3, 135.6
4		1.00, 2.00, 6.24	7.5, 27.7, 105.7, 130.9
5		1.80, 3.50, 3.82, 3.90	23.7, 51.1, 76.5, 163.3
6	$\text{H}_5\text{C}_2\text{---O---CH=CH}_2$	1.30, 3.70, 4.04, 4.18, 6.47	14.8, 62.7, 83.5, 151.8
	Experimental	2.2 T, 3.6 S 3.8/4.1	20.8, 51.5, 76.7, 163.5

is more similar to the query? For this purpose some topological indices and distances are found to be suitable.

The most difficult of all evaluations is the final rating of how well the obtained structure or spectrum fits the input data. Probability measures should be available for each spectra–feature correlation. A simple mean value of all probabilities is a rough estimate of the accuracy of the result. The weights can be added to differentiate the influence of certain spectral regions or structural fragments on final evaluation of the hypothesis. In general, one spectroscopy is not enough to solve the structure completely; hence, data from different spectroscopies should be combined.

### Example

As an example of how the modeling techniques can be used, a real case where the  $^{13}\text{C}$  and  $^1\text{H}$  NMR simulations were employed to solve the unknown structure by the system CARBON is presented. The experimental data are  $^{13}\text{C}$  and  $^1\text{H}$  NMR spectra of the unknown.  $^{13}\text{C}$  NMR spectrum contains four peaks (at 20.8, 51.5, 76.7, and 163.5 ppm relative to the resonance line of tetramethylsilane – TMS) and  $^1\text{H}$  NMR spectrum consists of a triplet at 2.2 ppm, a singlet at 3.6 ppm, and a multiplet from 3.8 to 4.1 ppm, all relative to TMS. Using the ‘FRAGMENT’ option in the system CARBON, the two most common fragments at each  $^{13}\text{C}$  NMR peak were selected, yielding a set of eight fragments. From this complete set and the constraints of the target structure (four carbon atoms in the molecular formula), the module GENSTR has selected five possible subsets (Figure 2). From the five subsets, six different structures were generated.

For six possible structures  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra were simulated. From the obtained results (Table 1) it is clearly seen that the structure no. 5 has produced the best pair of spectra compared with the experimental ones. Indeed, the unknown compound was 2-methoxypropene as proposed by the CARBON system.

*See also: Chemometrics and Statistics: Expert Systems; Multicriteria Decision Making; Signal Processing.*

### Further Reading

- Brown S (1990) Chemometrics. *Analytical Chemistry* 62: 84R–101R.
- Hemmer MC (2002) Expert systems. In: Chalmers JM and Griffiths PR (eds.) *Handbook of Vibrational Spectroscopy*, vol. 3, pp. 1962–1981. Chichester: Wiley.
- Kamran P and Chignell M (1988) *Expert Systems for Experts*. New York: Wiley.
- Lavine BK and Workman J (2002) Chemometrics. *Analytical Chemistry* 74: 2763–2769.
- Munk ME (1998) Computer-based structure determination: then and now. *Journal of Chemical Information and Computer Sciences* 38: 997–1009.
- Pretsch E, Tóth G, Munk ME, and Badertscher M (2002) *Computer-Aided Structure Elucidation: Spectra Interpretation and Structure Generation*. Weinheim: Wiley-VCH.
- Varmuza K (1980) *Pattern Recognition in Chemistry. Lecture Notes in Chemistry*, vol. 21. Berlin: Springer.
- Warr WA (1993) Computer-assisted structure elucidation. 1. Library search and spectral data collections. *Analytical Chemistry* 65: A1045–A1050.
- Zupan J and Gasteiger J (1999) *Neural Networks in Chemistry and Drug Design*. Weinheim: Wiley-VCH.

## SUBLIMATION

J D Green, Beverley, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Sublimation is not a procedure that is generally regarded as an analytical technique. It is a process, however, by which compounds can be purified or mixtures separated and as such can be of value as a single step or as an integral part of a more complex analytical method. It is applicable to a range of solids

of inorganic or organic origin in a variety of different matrices and can be particularly useful when heat-labile materials are involved.

As a method of sample purification sublimation has been used to produce high-purity materials as analytical standards. A specific and common example of sublimation used as a means of purification is the removal of water from heat-labile materials in the process known as freeze-drying. The technique is described more fully below.

As a separation technique fractional sublimation has been used either to purify samples for analysis by

is more similar to the query? For this purpose some topological indices and distances are found to be suitable.

The most difficult of all evaluations is the final rating of how well the obtained structure or spectrum fits the input data. Probability measures should be available for each spectra–feature correlation. A simple mean value of all probabilities is a rough estimate of the accuracy of the result. The weights can be added to differentiate the influence of certain spectral regions or structural fragments on final evaluation of the hypothesis. In general, one spectroscopy is not enough to solve the structure completely; hence, data from different spectroscopies should be combined.

### Example

As an example of how the modeling techniques can be used, a real case where the  $^{13}\text{C}$  and  $^1\text{H}$  NMR simulations were employed to solve the unknown structure by the system CARBON is presented. The experimental data are  $^{13}\text{C}$  and  $^1\text{H}$  NMR spectra of the unknown.  $^{13}\text{C}$  NMR spectrum contains four peaks (at 20.8, 51.5, 76.7, and 163.5 ppm relative to the resonance line of tetramethylsilane – TMS) and  $^1\text{H}$  NMR spectrum consists of a triplet at 2.2 ppm, a singlet at 3.6 ppm, and a multiplet from 3.8 to 4.1 ppm, all relative to TMS. Using the ‘FRAGMENT’ option in the system CARBON, the two most common fragments at each  $^{13}\text{C}$  NMR peak were selected, yielding a set of eight fragments. From this complete set and the constraints of the target structure (four carbon atoms in the molecular formula), the module GENSTR has selected five possible subsets (Figure 2). From the five subsets, six different structures were generated.

For six possible structures  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra were simulated. From the obtained results (Table 1) it is clearly seen that the structure no. 5 has produced the best pair of spectra compared with the experimental ones. Indeed, the unknown compound was 2-methoxypropene as proposed by the CARBON system.

*See also: Chemometrics and Statistics: Expert Systems; Multicriteria Decision Making; Signal Processing.*

### Further Reading

- Brown S (1990) Chemometrics. *Analytical Chemistry* 62: 84R–101R.
- Hemmer MC (2002) Expert systems. In: Chalmers JM and Griffiths PR (eds.) *Handbook of Vibrational Spectroscopy*, vol. 3, pp. 1962–1981. Chichester: Wiley.
- Kamran P and Chignell M (1988) *Expert Systems for Experts*. New York: Wiley.
- Lavine BK and Workman J (2002) Chemometrics. *Analytical Chemistry* 74: 2763–2769.
- Munk ME (1998) Computer-based structure determination: then and now. *Journal of Chemical Information and Computer Sciences* 38: 997–1009.
- Pretsch E, Tóth G, Munk ME, and Badertscher M (2002) *Computer-Aided Structure Elucidation: Spectra Interpretation and Structure Generation*. Weinheim: Wiley-VCH.
- Varmuza K (1980) *Pattern Recognition in Chemistry. Lecture Notes in Chemistry*, vol. 21. Berlin: Springer.
- Warr WA (1993) Computer-assisted structure elucidation. 1. Library search and spectral data collections. *Analytical Chemistry* 65: A1045–A1050.
- Zupan J and Gasteiger J (1999) *Neural Networks in Chemistry and Drug Design*. Weinheim: Wiley-VCH.

## SUBLIMATION

J D Green, Beverley, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Sublimation is not a procedure that is generally regarded as an analytical technique. It is a process, however, by which compounds can be purified or mixtures separated and as such can be of value as a single step or as an integral part of a more complex analytical method. It is applicable to a range of solids

of inorganic or organic origin in a variety of different matrices and can be particularly useful when heat-labile materials are involved.

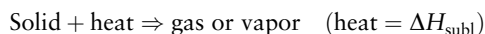
As a method of sample purification sublimation has been used to produce high-purity materials as analytical standards. A specific and common example of sublimation used as a means of purification is the removal of water from heat-labile materials in the process known as freeze-drying. The technique is described more fully below.

As a separation technique fractional sublimation has been used either to purify samples for analysis by

removing undesirable components of the matrix or to remove the analyte from the matrix for subsequent analysis.

## Principles

Sublimation is the direct conversion of a solid to a gas or vapor:



The heat supplied in this endothermic process is termed the heat of sublimation ( $\Delta H_{\text{subl}}$ ). The conditions under which sublimation occurs may be predicted for a given substance from its phase diagram, but in practice it is more common to use typical experimental parameters to determine the optimized procedure.

The heat of sublimation is a crucial parameter in deciding upon the applicability of sublimation to a particular substance, or indeed on the possibility of separating two components in a mixture.

An empirical approach to determine the appropriate temperature and pressure for sublimation can be used based upon previously determined data. The temperature ( $T$ ) and pressure ( $P$ ) of sublimation can be related by an expression of the form:

$$\log_{10} P \text{ (mm)} = A - B/T$$

in which the constants  $A$  and  $B$  for compounds of interest are available from published tables. The result of the sublimation process can be seen in a freezer where ice sublimates and resolidifies as crystals. Iodine is a common substance that sublimates at room temperature and pressure; the result of this can be observed in a reagent bottle of the element.

The theory and mechanism of sublimation is of less practical importance to analytical procedures than it is in some other specialized areas of chemical science. Knowledge of sublimation characteristics can aid improvements in the stability of materials used at high temperatures or low pressures. For analytical purposes it should be sufficient to recognize that rates of sublimation depend upon the topology of the vaporizing surface (dislocations, atomic steps, and ledge concentrations) and upon any atomic rearrangement that occurs during the sublimation process.

Experimentally, to effect sublimation, a number of criteria need to be satisfied. First, the sample in question must be maintained at a temperature that ensures a sufficiently high vapor pressure for sublimation to occur, whilst remaining below that point at

which the material either decomposes or melts. Second, a secondary surface must be available on which the sublimed vapor can recondense or resolidify. A number of experimental arrangements have been used that allow these criteria to be established; these are described in a later section of this article. It is also possible to enhance the sublimation process by changing the physical parameters under which the process is carried out:

- The sample may be heated in order to increase its vapor pressure.
- The application of a vacuum to the apparatus encourages vaporization and enhances the sublimation.
- Selectively cooling part of the apparatus increases the efficiency of the condensation process.
- Using an entraining gas can improve the mass transport in the system and thereby increase the overall efficiency of the sublimation process.

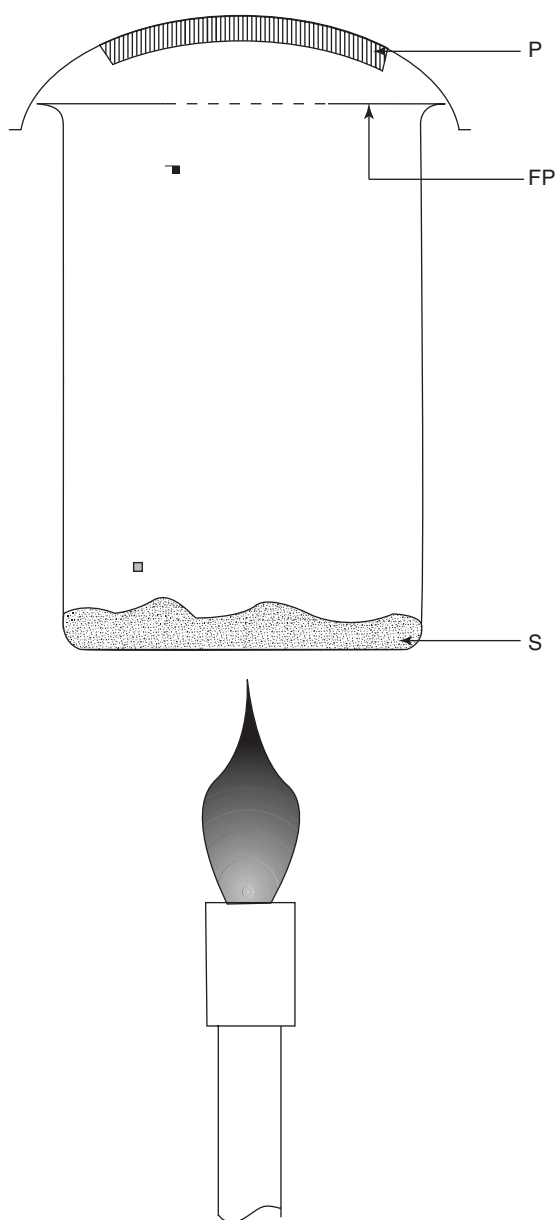
## Apparatus

The simplest form of sublimation apparatus consists of a beaker or porcelain dish on top of which is placed an upturned watch-glass. The beaker contains the solid to be sublimed and the underside of the watch-glass provides the surface upon which the sublimed components condense (**Figure 1**). A perforated filter paper is commonly placed between the beaker and the watch-glass to prevent sublimate falling back into the sample.

A variant upon the above system uses an upturned funnel instead of a watch-glass as the condensing surface and an appropriately placed sealing ring improves the performance (**Figure 2**). Coils through which coolant is circulated can promote the sublimation process.

An early form of sublimation apparatus of which the above arrangements are derivatives (**Figure 3**) included a means of cooling the surface on which the sublimate condenses. Cooling can be achieved in a number of ways, for example, using filter papers moistened with cool water, or in the case of the upturned funnel, a suitably shaped coil of circulating coolant liquid can be placed around the surface.

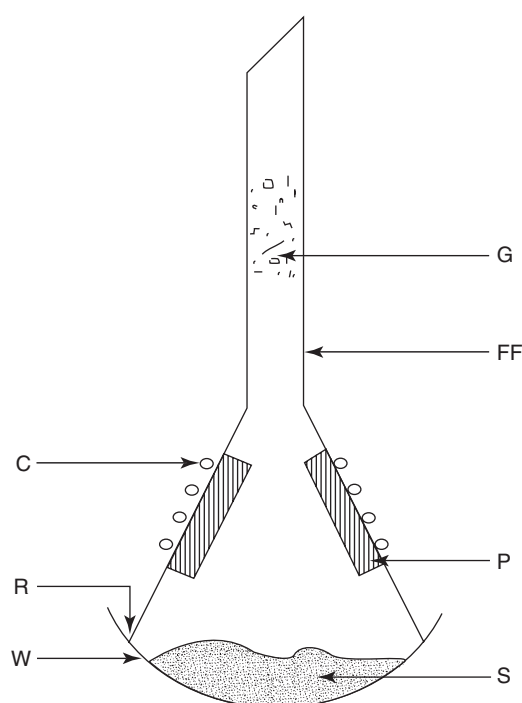
Sublimation under reduced pressure uses a modified form of apparatus in which a sealed enclosure allows a vacuum to be applied. The cooled surface is orientated with respect to the sample so as to maximize the condensation once sublimation has occurred (**Figure 4**). Reducing the distance that the sublimed substance(s) must travel is beneficial provided the necessary temperature gradient between



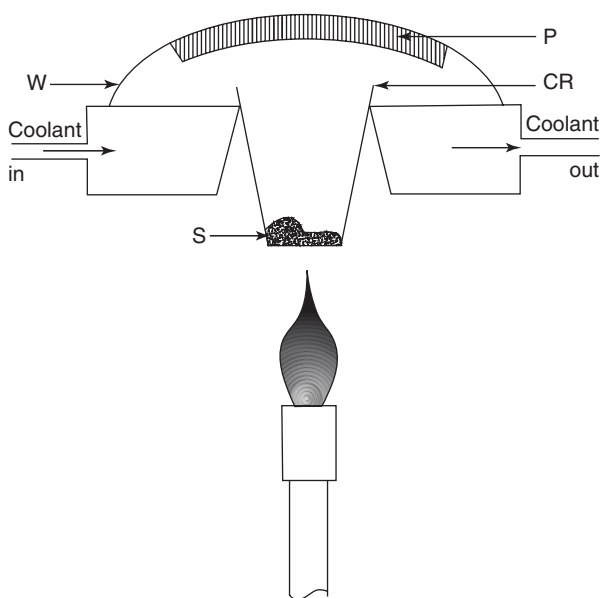
**Figure 1** Simple apparatus for demonstrating the principles of sublimation. S, sample; P, sublimate (product); FP, perforated filter paper.

sample and condensing surface can be maintained. An alternative arrangement for sublimation applications is shown in **Figure 5**.

Freeze-drying, a special application of the sublimation principle, uses apparatus of a different kind (**Figure 6**). The sample, usually a liquid at room temperature, is dispersed around the walls of a round-bottomed flask whilst it is frozen by immersion in a suitable freezing mixture, for example, dry ice/acetone. The flask is then attached to the evacuating system that usually comprises an oil vacuum pump protected from the ice sublimate by a train of

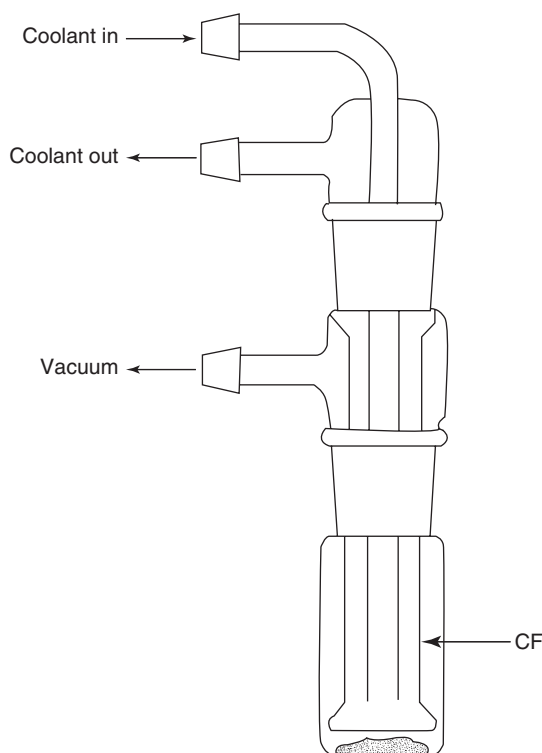


**Figure 2** Apparatus for simple sublimation at atmospheric pressures. Watch-glass, W, with sample, S, surmounted by filter funnel, FF, with cooling coils, C, glass wool, G, and collected sublimed product, P. A sealing ring, R, is included between the watch-glass and filter funnel.



**Figure 3** Early form of sublimation apparatus. The heated crucible, CR, rests in the cooling device. The sublimate product, P, collects on the underside of the watch-glass, W. S, sample; P, sublimed product.

condenser traps. Over a period of typically several hours the ice sublimates from the sample and condenses in the traps. Air is then admitted to the

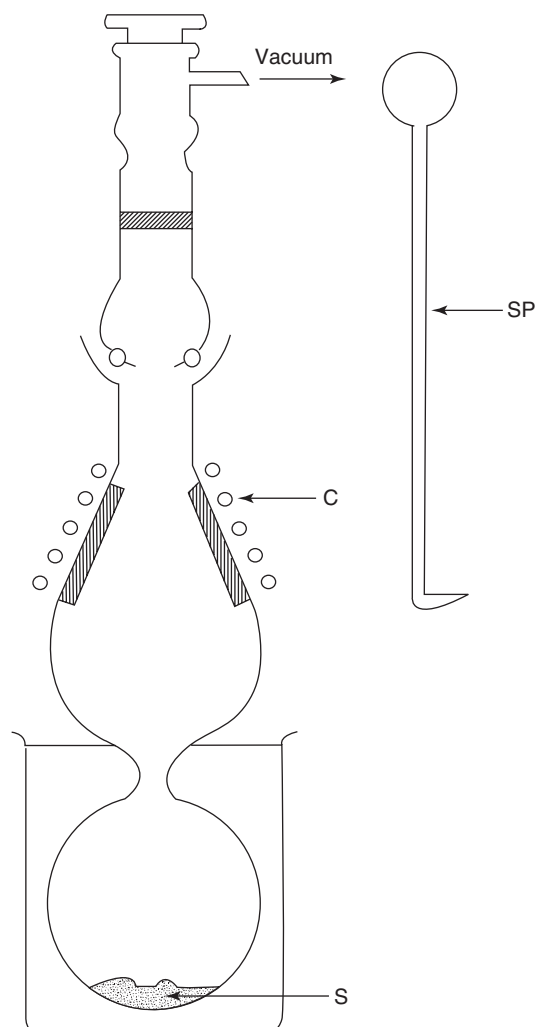


**Figure 4** Apparatus for sublimation at reduced pressure. Coolant is circulated through the cold finger, CF, whilst a vacuum is applied to the sample chamber.

apparatus and the dried sample can be removed whilst the sublimed ice is drained off as water through the drain tap. Frequently, this system is used to dry heat-sensitive materials such as enzymes and the process has been termed lyophilization.

Sublimation of metallic elements from rock or ore samples requires high temperatures. The equipment used is based upon silica furnace tubes in order to withstand the necessary conditions. The silica tube is heated in a furnace and the sublimate condenses either on a cool part of the tube or on a cooled surface immediately after leaving the tube.

The conditions of sublimation must be chosen according to the requirements of the application. For simple purification the sample temperature is raised slowly, under reduced pressure if necessary, until the sublimate is observed on the condensing surface. These established conditions should then be maintained until no further sublimation appears to be occurring, at which point the sample temperature can be raised again if other components of the sample can be further removed. At any point in this cycle the apparatus can be dismantled and the sublimate removed. This process allows selective separation or fractional sublimation to be carried out.



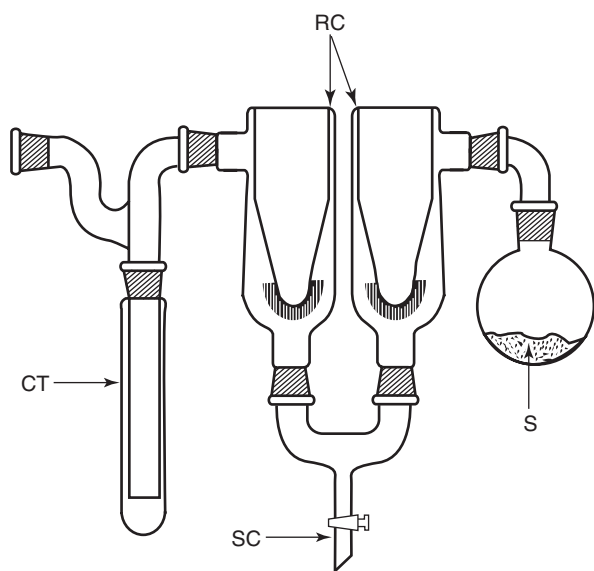
**Figure 5** Improved sublimation apparatus proposed by Eisenbraun *et al.* (1978). Sample, S, sublimates from the lower to the upper chamber where condensation takes place on the cooled surface, C, improves the condensation process. A specially formed spatula, SP, can be used to help remove the sublimate after the upper part of the apparatus is removed.

## Applications

Sublimation is applicable to a wide range of organic and inorganic compounds in an equally wide range of different matrices. Sublimable substances include ice, iodine, arsenic(III) oxide, cadmium sulfide, ammonium chloride, and a large number of organic compounds. Common matrices from which substances are sublimed include biological fluids, plant materials, carbonaceous materials, samples of crude organic solids, and samples of rocks and ores.

Sublimation as a method of applying substances to thin-layer chromatography (TLC) plates involves the sample being sublimed and the vapor produced being directed by means of a drawn capillary onto the





**Figure 6** Typical freeze-drying apparatus. The frozen sample, S, is attached to the condenser assembly and a vacuum is applied. A cold trap, CT, protects the pump as ice sublimates from the sample and subsequently condenses in the refrigerant condensers, RC. On completion of the drying the sample is removed and the collected ice melts and drains from the system through the stopcock, SC.

surface of a TLC plate that is slowly moved in one dimension. This results in the sublimed materials being deposited upon the TLC plate in a differential mode – the most easily sublimed compounds are deposited first whilst those requiring a higher temperature are deposited later. The TLC plate is then developed in the normal way to give what has been termed a ‘thermofractogram’ in which the substances are separated as a function of their heats of sublimation along one axis and as a function of their chromatographic characteristics along the other axis. This approach has been applied to a wide range of substances including pharmaceutical preparations, plant components, and foodstuffs.

Only a very limited number of standard methods are reported in which sublimation is an important aspect. These comprise an ASTM standard for measurement of sublimation from thermionic emitters, and two standards from Germany and Japan testing the stability of dyes and printing inks to sublimation. The first covers the determination of the quantity, rate, and identity of sublimed, evaporated, or sputtered materials, whilst the latter two are concerned with textile materials and semimanufactured products.

Sublimation of ninhydrin crystals has recently been reported as providing a significant improvement in detecting fingerprints on thermal paper. Sublimation allows fine detail of latent fingerprints to

become visible without the usual background staining experienced with the application of ninhydrin solutions. The method is applicable to the new Euro monetary notes.

Sublimation is often the mechanism by which pre-concentration of an analyte is effected, although this fact is frequently not appreciated. Dynamic headspace concentration from solid samples such as plant materials, foodstuffs, or polymeric materials occurs by sublimation of the volatile components. Indeed, given appropriate apparatus that can be operated at different temperatures for dynamic headspace concentration, the heat of sublimation can be determined for various compounds.

Derivatization procedures carried out on crude samples can produce materials with improved sublimation characteristics. This technique has been used to produce volatile compounds of lanthanides and actinides that have then been sublimed prior to analytical determinations. Derivatives have been made using  $\beta$ -diketones (hexafluoroacetylacetone or acetylacetone), benzoyltrifluoroacetone, and thenoyltrifluoroacetones.

Low-temperature sublimation, which in some circumstances is termed freeze-drying, has been used to separate water, as ice, from biological fluids such as serum, urine, or saliva. The technique has been particularly useful in pediatric cases where sample volumes have been extremely low. Determinations have then been accomplished using infrared spectroscopy or mass spectrometry. Preparation of physiological samples for determination of deuterium oxide has included sublimation techniques prior to spectrophotometric determinations.

Low-temperature sublimation has been used to prepare samples for cryo-scanning electron microscopy (SEM) analysis in order to examine herbicide particles in a water suspension. The sublimation of herbicide-containing frozen water droplets provides a suitable etching of the surface for the SEM technique.

High-temperature sublimations are often the methods of choice in sample preparations from mineral ores, particularly in the case of trace enrichment of noble metals and the actinides and lanthanoids prior to activation methods. Temperatures of 800–1200°C are typical. The procedure is carried out in silica tubes with entrainment gases, for example, air or argon, being used to increase the sublimation process.

Polycyclic aromatic compounds have been separated using sublimation techniques from a variety of samples including coal, solids derived from oil, coal, and petroleum processing, and residues (soots) resulting from the use of such fossil fuels.

A variety of miscellaneous applications have been developed for separation from difficult matrices and purification of specific materials. These include:

- Mercury separated from impurities by conversion to its iodide followed by sublimation.
- Isolation of proazulene and chamomile from the flower heads of plants.
- Isolation of aroma compounds from wheat and rye samples prior to determination using isotope dilution methods.
- Determination of tin in cassiterite.
- Selective sublimation of molybdenum and tungsten.

Sublimation is used in some procedures for the preparation of samples for SEM in which gold is sublimed in vacuum from a heated tungsten filament to the sample being examined.

See also: **Derivatization of Analytes.** **Geochemistry:** Soil, Major Inorganic Components; Soil, Minor Inorganic Components. **Polycyclic Aromatic Hydrocarbons:** Determination. **Quality Assurance:** Internal Standards. **Thin-Layer Chromatography:** Overview.

## Further Reading

- Chickos JS (1987) Heats of sublimation. In: Liebman JF and Greenberg A (eds.) *Molecular Structure and Energetics*, vol. 2. Weinheim: VCH Publishers.
- Davies M (compiler) (1991–92) Sublimation pressure for organic compounds. In: Lide DR (ed.) *CRC Handbook of Physics and Chemistry*, 72nd edn, pp. 5–91. Boca Raton: CRC Press.
- Eisenbraun EJ, Moyer CJ, and Vuppapalapaty P (1978) Sublimation apparatus: Design improvement. *Chemistry & Industry* 229–230.
- Holden CA and Bryant HS (1969) Purification by sublimation. *Separation Science* 4(1): 1–13.
- Somorjai GA (1968) Mechanism of sublimation. *Science* 162: 755–760.
- Mullin JW (2002) Sublimation. In: *Ullmann's Encyclopedia of Industrial Chemistry*. New York: Wiley.
- Schwartz L and Frerichs I (2002) Advanced solvent-free application of ninhydrin for detection of latent fingerprints on thermal paper and other surfaces. *Journal of Forensic Science* 47(6): 1274–1277.
- Stahl E (1976) Advances in the field of thermal procedures in direct combination with thin-layer chromatography. *Accounts of Chemical Research* 9: 75–80.

# SULFUR

J Räisänen, University of Helsinki, Helsinki, Finland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

In this article, the techniques most commonly employed for elemental sulfur determination are introduced. They are compared in the form of tables where relevant recent information such as detection limits and typical samples have been collected. Also, general information on the element sulfur is provided.

## Origin and Nature

Sulfur naturally consists of four stable isotopes  $^{32}\text{S}$  (95%),  $^{33}\text{S}$  (0.76%),  $^{34}\text{S}$  (4.22%), and  $^{36}\text{S}$  (0.014%). The atomic number of sulfur is 16 and its atomic weight is  $32.066\text{ g mol}^{-1}$ . It has the most allotropes of any element, the most stable of these being orthorhombic  $\text{S}_8$ . Commercial sulfur melts at  $\sim 119^\circ\text{C}$  and boils at  $444.6^\circ\text{C}$ . Its density varies from  $1.808\text{ g cm}^{-3}$  at  $115^\circ\text{C}$  to  $1.599\text{--}1.614\text{ g cm}^{-3}$  at the boiling point. Sulfur is an odorless, tasteless, and brittle solid. It is a poor conductor of heat and electricity. Pure solid sulfur is pale yellow. It is stable to air and water but burns if ignited. Almost all the

elements, apart from gold, platinum, and the inert gases, combine with sulfur.

Sulfur is found both in its native form and in metal sulfide ores. It occurs in its native form in the vicinity of volcanoes and hot springs. Sulfur is the 10th most abundant element, and it is found in meteorites, in the ocean, in the earth's crust, in the atmosphere, and in practically all plant and animal life. The abundance of sulfur in the earth's crust is 0.03–0.1%. Sediments (sulfur concentration  $\sim 4250\text{ }\mu\text{g g}^{-1}$ ) have accumulated a large proportion (43%) of the total sulfur of the earth's crust. Metamorphic and magma rocks, which are the other typical materials of the crust, contain less sulfur (concentration  $\sim 600\text{ }\mu\text{g g}^{-1}$ ). Pyrite is the most common mineral form of sulfur in sediments. The storage of sulfur in sediments and ocean crusts (the sulfur concentration of seawater is  $\sim 900\text{ }\mu\text{g g}^{-1}$ ) is balanced by the degassing of volcanic rocks. High-purity sulfur is commercially available in purities of 99.999% +.

## Common Matrices in which Sulfur is Found

Sulfur occurs in organic substances, e.g., eggs, hair, wool, albumin, garlic, mustard, horseradish,

A variety of miscellaneous applications have been developed for separation from difficult matrices and purification of specific materials. These include:

- Mercury separated from impurities by conversion to its iodide followed by sublimation.
- Isolation of proazulene and chamomile from the flower heads of plants.
- Isolation of aroma compounds from wheat and rye samples prior to determination using isotope dilution methods.
- Determination of tin in cassiterite.
- Selective sublimation of molybdenum and tungsten.

Sublimation is used in some procedures for the preparation of samples for SEM in which gold is sublimed in vacuum from a heated tungsten filament to the sample being examined.

See also: **Derivatization of Analytes.** **Geochemistry:** Soil, Major Inorganic Components; Soil, Minor Inorganic Components. **Polycyclic Aromatic Hydrocarbons:** Determination. **Quality Assurance:** Internal Standards. **Thin-Layer Chromatography:** Overview.

## Further Reading

- Chickos JS (1987) Heats of sublimation. In: Liebman JF and Greenberg A (eds.) *Molecular Structure and Energetics*, vol. 2. Weinheim: VCH Publishers.
- Davies M (compiler) (1991–92) Sublimation pressure for organic compounds. In: Lide DR (ed.) *CRC Handbook of Physics and Chemistry*, 72nd edn, pp. 5–91. Boca Raton: CRC Press.
- Eisenbraun EJ, Moyer CJ, and Vuppapalapaty P (1978) Sublimation apparatus: Design improvement. *Chemistry & Industry* 229–230.
- Holden CA and Bryant HS (1969) Purification by sublimation. *Separation Science* 4(1): 1–13.
- Somorjai GA (1968) Mechanism of sublimation. *Science* 162: 755–760.
- Mullin JW (2002) Sublimation. In: *Ullmann's Encyclopedia of Industrial Chemistry*. New York: Wiley.
- Schwartz L and Frerichs I (2002) Advanced solvent-free application of ninhydrin for detection of latent fingerprints on thermal paper and other surfaces. *Journal of Forensic Science* 47(6): 1274–1277.
- Stahl E (1976) Advances in the field of thermal procedures in direct combination with thin-layer chromatography. *Accounts of Chemical Research* 9: 75–80.

# SULFUR

J Räisänen, University of Helsinki, Helsinki, Finland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

In this article, the techniques most commonly employed for elemental sulfur determination are introduced. They are compared in the form of tables where relevant recent information such as detection limits and typical samples have been collected. Also, general information on the element sulfur is provided.

## Origin and Nature

Sulfur naturally consists of four stable isotopes  $^{32}\text{S}$  (95%),  $^{33}\text{S}$  (0.76%),  $^{34}\text{S}$  (4.22%), and  $^{36}\text{S}$  (0.014%). The atomic number of sulfur is 16 and its atomic weight is  $32.066\text{ g mol}^{-1}$ . It has the most allotropes of any element, the most stable of these being orthorhombic  $\text{S}_8$ . Commercial sulfur melts at  $\sim 119^\circ\text{C}$  and boils at  $444.6^\circ\text{C}$ . Its density varies from  $1.808\text{ g cm}^{-3}$  at  $115^\circ\text{C}$  to  $1.599\text{--}1.614\text{ g cm}^{-3}$  at the boiling point. Sulfur is an odorless, tasteless, and brittle solid. It is a poor conductor of heat and electricity. Pure solid sulfur is pale yellow. It is stable to air and water but burns if ignited. Almost all the

elements, apart from gold, platinum, and the inert gases, combine with sulfur.

Sulfur is found both in its native form and in metal sulfide ores. It occurs in its native form in the vicinity of volcanoes and hot springs. Sulfur is the 10th most abundant element, and it is found in meteorites, in the ocean, in the earth's crust, in the atmosphere, and in practically all plant and animal life. The abundance of sulfur in the earth's crust is 0.03–0.1%. Sediments (sulfur concentration  $\sim 4250\text{ }\mu\text{g g}^{-1}$ ) have accumulated a large proportion (43%) of the total sulfur of the earth's crust. Metamorphic and magma rocks, which are the other typical materials of the crust, contain less sulfur (concentration  $\sim 600\text{ }\mu\text{g g}^{-1}$ ). Pyrite is the most common mineral form of sulfur in sediments. The storage of sulfur in sediments and ocean crusts (the sulfur concentration of seawater is  $\sim 900\text{ }\mu\text{g g}^{-1}$ ) is balanced by the degassing of volcanic rocks. High-purity sulfur is commercially available in purities of 99.999% +.

## Common Matrices in which Sulfur is Found

Sulfur occurs in organic substances, e.g., eggs, hair, wool, albumin, garlic, mustard, horseradish,

cabbage, and many proteins. It also occurs in natural gas, petroleum, and coal. It is a minor constituent of fats, body fluids, and skeletal minerals. The element is commonly determined in environmental samples such as air particulates, water, and soil samples. Sulfur makes up a large proportion of the fine particulate matter found in atmospheric aerosol samples.

## Methods of Determination at Major and Trace Element Levels

Total sulfur can be determined by several techniques. In the more conventional chemical methods sulfur is converted to, e.g., oxides and then determined by infrared (IR) or gravimetric techniques or converted to hydrogen sulfide and then determined by gas-analysis techniques. Here, mainly methods used for the determination of elemental sulfur (direct methods) are considered; methods used for sulfur compound analysis are dealt with elsewhere in the Encyclopedia. Also, techniques used for radioisotope detection, for example, the multiphoton detection are excluded from the present survey. Readers interested in chromatographic and electrophoretic techniques employed for sulfur and sulfur-oxygen species analysis can consult also the review articles listed in the Further Reading section. The methods are divided into three main groups: general analytical methods, activation methods, and prompt ion beam methods. In order to find the most recent figures of merit and applications, computerized key word searches were performed using several databases. The collected important data concerning the methods and applications are summarized in Tables 1–4.

### General Analytical Techniques

In the following the various common analytical techniques are discussed briefly. The methods surveyed have been frequently used as routine methods for elemental analysis in laboratories, and easy-to-use apparatus have been developed for most of the commercially available methods. Details of the applicability of the methods to sulfur analysis are outlined in Table 1.

**Atomic spectrometry** Laser-induced atomic emission spectroscopy (AES) is a fast technique to determine directly elemental sulfur. Practically no matrix effects occur and the method is virtually nondestructive and easy to use. A disadvantage is the rather poor sensitivity, for example, a typical detection limit for sulfur in steel is  $\sim 70 \mu\text{g}$  per g. With indirect atomic absorption spectrometry clearly better

sensitivity is obtained, a detection limit of  $0.27 \mu\text{g}$  per g in soil has been achieved. The method needs, however, chemical pretreatment of the samples.

Different inductively coupled plasma (ICP) methods have been developed which can be used for sulfur analysis. Of these, ICP-AES or ICP-OES (optical emission spectroscopy) and ICP-MS (mass spectrometry) have been widely used. Both have a very broad analytical range.

In ICP-AES, intensity of the S emission spectrum is a direct function of sulfur concentration in a solution and does not depend on the type of sulfur-containing compound and the oxidation status of the analyzed sulfur atom. The method is subject to a large variety of spectral interference from various sources but these can be minimized by using high-resolution spectrometers and vacuum purge optics. The method is notable for its considerably low detection limits, wide dynamic range, and elemental coverage. It is used most in the field of environmental analysis but is of limited use without sample preconcentration. When the method is used to determine sulfur, its detection limit is substantially higher than the typical (nanogram per milliliter) values cited for most metals.

**Mass spectrometry** ICP-MS offers an accurate method for sulfur isotopic ratio measurements and a sensitive technique for elemental sulfur determination. The method may be applied as an absolute method, because the analyzed atoms themselves, and not the radiation they emit, produce the analytical signals. The technique is useful for the analysis of high-purity materials, such as metals and semiconductors. The detection limits obtainable in most cases by ICP-MS are considerably better than typical values reported for ICP-AES. However, the analytical precision and accuracy of the method are often poor. The utility of the method to determine sulfur is further complicated, as efficient ionization of sulfur is difficult to obtain because of the high ionization energy of the element.

There are also various sources of interference inherent in the method. The use of mass spectrometric methods for sulfur analysis is complicated by the isobaric interferences of  $^{16}\text{O}^{16}\text{O}^+$ ,  $^{16}\text{O}^{17}\text{O}^+$ , and  $^{16}\text{O}^{18}\text{O}^+$  for  $^{32}\text{S}^+$ ,  $^{33}\text{S}^+$ , and  $^{34}\text{S}^+$  analysis, respectively. Also, other possible polyatomic isobaric interferences, such as  $^{14}\text{N}^{18}\text{O}^+$  and  $^{15}\text{N}^{16}\text{O}^+$ , should be taken into account. Another problem has been the plural oxidation states of sulfur. Several procedures have been developed to minimize these problems, such as the use of the electrothermal vaporization technique, use of a hexapole filter, and a sector magnetic field to reduce molecular ions

**Table 1** Conventional analytical methods for sulfur determination

Method	Mass/volume needed	Interferences	Multielement method	Destructive	Typical samples	Typical detection limit	Remarks	
ICP-AES	ml	Yes	Yes	Yes	Oil	0.05 µg per g	Combined with electrolytic dissolution	(a)
					Wine	0.1 mg l <sup>-1</sup>		(b)
					Plant material	0.05 mg l <sup>-1</sup>		(c)
					Steel	0.6 µg per g		(d)
ICP-MS	ml	Yes	Yes	Yes	Liquid	0.01 ng per g	<sup>33</sup> S <sup>+</sup> isotope dilution	(e)
					Silicate	0.07 µg per g		(f)
					Silicate	0.3 µg per g	<sup>33</sup> S <sup>+</sup> isotope dilution	(f)
					Organic solutions	0.2 ng ml <sup>-1</sup>	Detected as SO <sup>+</sup>	(g)
SIMS	nl	Yes	Yes	Yes	Silicate glass	< 1 µg per g	In routine analysis	(h)
C		Yes	Yes	Yes	Sediment	420 µg per g	Py-GC + FPD GC + FPD	(i)
					Biological	40 ng		(j)
					Naphtha	20 ng per g		(k)
XRF	~ 1 g	Yes	Yes	No	Fly ash	8 µg per g	30 s measurement	(l)
					Aerosol filter	100 ng cm <sup>-2</sup>	In air measurement	(m)
					Oil	1.7 µg per g	Wavelength dispersive	(n)
TXRF	~ 10 µl ~ 10 µg	Yes	Yes	No	Polymer	11 µg l <sup>-1</sup>	1000 s measurement	(o)

(a) Wieberneit N and Heitland P (2001) Applications of ICP-OES with a new argon-filled CCD spectrometer using spectral lines in the vacuum ultraviolet spectral range. *Applied Spectroscopy* 55: 598–603.

(b) Sarudi I and Kelemen J (1998) Determination of sulfur and total sulfur dioxide in wines by an ICP-AES method. *Talanta* 45: 1281–1284.

(c) Wang J, Yi HQ, He CS, and Li HS (1999) Sample preparation method for the determination of total sulfur in plant materials. *Communications in Soil Science and Plant Analysis* 30: 599–603.

(d) Kondo H, Aimoto M, Ono A, and Chiba K (1999) Rapid determination of sulfur in steel by electrolytic dissolution-inductively coupled plasma atomic emission spectrometry. *Analytica Chimica Acta* 394: 293–297.

(e) Prohaska T, Latkoczy C, and Stingeder G (1999) Precise sulfur isotope ratio measurements in trace concentration of sulfur by inductively coupled plasma double focusing sector field mass spectrometry. *Journal of Analytical Atomic Spectrometry* 14: 1501–1504.

(f) Makishima A and Nakamura E (2001) Determination of total sulfur at microgram per gram levels in geological materials by oxidation of sulfur into sulfate with *in situ* generation of bromine using isotope dilution high-resolution ICPMS. *Analytical Chemistry* 73: 2547–2553.

(g) Bandura DR, Baranov VI, and Tanner SD (2002) Detection of ultratrace phosphorus and sulfur by quadrupole ICPMS with dynamic reaction cell. *Analytical Chemistry* 74: 1497–1502.

(h) Hauri E, Wang J, Dixon JE *et al.* (2002) SIMS analysis of volatiles in silicate glasses. 1. Calibration, matrix effects and comparisons with FTIR. *Chemical Geology* 183: 99–114.

(i) Colina M and Gardiner PHE (1999) Simultaneous determination of total nitrogen, phosphorus and sulfur by means of microwave digestion and ion chromatography. *Journal of Chromatography A* 847: 285–290.

(j) Choi SY, Kim MG, and Inoue H (1995) Determination of sulfur in biologically important substances by pyrolysis-gas chromatography. *Journal of Analytical and Applied Pyrolysis* 32: 127–136.

(k) Abdillahi MM, Alam K, Bari A, Siddiqui M, and Ali MA (1995) Determination of trace amounts of sulfur in hydrotreated naphthas: Comparative study using Raney nickel, Houston Atlas and gas chromatographic methods. *Analyst* 120: 1577–1582.

(l) Sprta V, Knob B, and Janos P (1999) X-ray fluorescence determination of total sulfur in fly ash. *Fresenius Journal of Analytical Chemistry* 364: 705–708.

(m) Samek L, Injuk J, Van Espen P, and Van Grieken R (2002) Performance of a new compact EDXRF spectrometer for aerosol analysis. *X-Ray Spectrometry* 31: 84–86.

(n) van Dalen G (1998) Determination of the phosphorus and sulfur content in edible oils and fats by wavelength-dispersive X-ray fluorescence spectrometry. *X-Ray Spectrometry* 27: 26–30.

(o) Vazquez C, Custo G, and Boeykens S (2001) Characterization of trace elements in high viscosity materials by total reflection X-ray spectrometry. *Spectrochimica Acta B* 56: 2253–2260.

**Table 2** Neutron activation analysis for sulfur

Reaction	Cross-section (mb)	Interferences	$T_{1/2}$ of product	$n$ flux ( $\text{n cm}^{-2} \text{s}^{-1}$ )	Detection limit	Remarks
<b>Fast neutron methods</b>						
$^{32}\text{S}(\text{n,p})^{32}\text{P}$	300	$^{31}\text{P}$ , $^{35}\text{Cl}$	14.3 d	$1 \times 10^{12}$ $1 \times 10^{13}$	$4 \times 10^{-3} \mu\text{g}$ 0.02 mg	Water, petroleum (a) Crude oil sample (b)
$^{32}\text{S}(\text{n,t})^{30}\text{P}$		—	2.5 min	$1 \times 10^{13}$	0.4 mg	(b)
$^{34}\text{S}(\text{n,p})^{34}\text{P}$	74	$^{34}\text{Cl}$	12.4 s	$3 \times 10^{10}$ $1 \times 10^{13}$	650 $\mu\text{g}$ 1 mg	(c) (b)
$^{34}\text{S}(\text{n},\alpha)^{31}\text{Si}$	150	$^{30}\text{Si}$ , $^{31}\text{P}$	2.62 h	$3 \times 10^{10}$ $1 \times 10^{13}$	250 mg 0.8 mg	(c) (b)
<b>Thermal neutron methods</b>						
$^{32}\text{S}(\text{n},\alpha)^{29}\text{Si}$	140		Stable	$3 \times 10^8$	$2.4 \times 10^{18} \text{ atoms cm}^{-2}$	Depth profiling (d)
$^{34}\text{S}(\text{n,p})^{34}\text{P}$		B, Cl, Co, Fe	12.4 s	$1 \times 10^{13}$	7 mg	(b)
$^{34}\text{S}(\text{n},\gamma)^{35}\text{S}$	240	$^{35}\text{Cl}$ , $^{38}\text{Ar}$	87.5 d	$1 \times 10^{12}$ $1 \times 10^{16}$	0.4 $\mu\text{g}$ 50 $\mu\text{g}$	(a) (b)
$^{36}\text{S}(\text{n},\gamma)^{37}\text{S}$	150	$^{37}\text{Cl}$ , $^{40}\text{Ar}$	5.1 min	$6 \times 10^{13}$	40 $\mu\text{g}$ per g	Fullerene sample (e)

(a) Pinta M (1978) *Modern Methods for Trace Element Analysis*. Ann Arbor: Ann Arbor Science.

(b) Parsons ML (1976) Nuclear methods. In: Winefordner JD (ed.) *Trace Analysis, Spectroscopic Methods for Elements*. New York: Wiley.

(c) Pepelnik R (1987) Capability of the 14 MeV neutron activation analysis at  $3 \times 10^{10} \text{ n cm}^{-2} \text{s}^{-1}$  with respect to sensitivities and interferences of all useful reactions. *Journal of Radioanalytical and Nuclear Chemistry* 112: 435–452.

(d) Downing RG, Fleming RF, Langland JK, and Vincent DH (1983) Neutron depth profiling at the national bureau of standards. *Nuclear Instruments and Methods in Physics Research* 218: 47–51.

(e) Rausch H, Braun T, Dodony I, and Lovas G (1999) Determination of sulfur as an impurity in commercial C60 fullerene soot and sublimed C60 polycrystals by INAA. *Analyst* 124: 417–419.

causing interference. Also, a dynamic reaction cell has been employed to oxidize sulfur to allow its detection as  $\text{SO}^+$ .

Sulfur isotope ratios  $^{34}\text{S}/^{32}\text{S}$  are known to vary significantly in nature and their determination has become of wide interest in, e.g., geochemical, biological, environmental, and industrial studies. ICP-MS provides a precise method for such isotope ratio determinations.

In secondary ion mass spectrometry (SIMS), secondary ions produced by primary low-energy ion bombardment are analyzed with a mass spectrometer. Sulfur has one of the highest relative secondary negative ion yields and is therefore very sensitive for SIMS detection using  $\text{Cs}^+$  as primary ions. SIMS is extensively used, especially in semiconductor technology, and has extreme detection limits and a wide dynamic range. Using the microvolume SIMS procedure nanoliter volumes of solution is sufficient for typical analyses. This technique is also commonly used for obtaining concentration–depth profiles. The problems encountered with SIMS in this case are related to calibration of both the concentration and depth scales. Persistent problem with SIMS is the dependence of the ion yield on the chemical composition of the matrix.

**Coupled chromatographic techniques** Several chromatography techniques (C) have been used for sulfur determination. These techniques allow generally very sensitive detection of sulfur-containing compounds (see Table 1). Usually, sulfur is first oxidized followed by ion chromatography. Typically, such analyses require sophisticated chemical procedures but this is not always the case; for example, the technique using pyrolysis-gas chromatography (Py-GC) equipped with a flame photometric detector (FPD) does not require sample pretreatment.

**Auger spectroscopy** Auger electron spectroscopy is a powerful surface analysis technique, as the analytical signals come from the top few monolayer of the specimen. In this technique, a primary beam of electrons is used to produce inner shell vacancies in atoms of the specimen. This method has been used for the determination of sulfur, e.g., on GaAs surfaces and to study sulfur surface segregation in Ni–Al solid solutions and the adsorption on a GaAs (001) surface by hydrogen sulfide exposure and heat treatment.

**X-ray fluorescence spectrometry** Conventional energy-dispersive X-ray fluorescence (XRF) has been widely used for sulfur determination by using a



**Table 3** Charged particle- and  $\gamma$ -activation procedures for sulfur

Ion/energy (MeV)	Reaction	Beam current	$T_{1/2}$ of product	Irradiation time	Interferences	Sample	Detection limit	Remarks
H <sup>+</sup> /22.5	$^{32}\text{S}(\text{p},\text{n})^{32}\text{Cl}$	5 $\mu\text{A}$	300 ms	1 s	$^{28}\text{Si}$ , $^{50}\text{Cr}$ , $^{40}\text{Ca}$ , $^{54}\text{Fe}$ , $^{46}\text{Ti}$	Petroleum on Al foil	50 $\mu\text{g cm}^{-2}$	(b)
H <sup>+</sup> /20	$^{32}\text{S}(\text{p},\text{n})^{32}\text{Cl}$	2 $\mu\text{A}$		10 irr.			30 ng per g	Interference free
H <sup>+</sup> /15	$^{33}\text{S}(\text{p},\gamma)^{34\text{m}}\text{Cl}$		32 min		Cl, Ar		30 $\mu\text{g per g}$	(c)
H <sup>+</sup> /13	$^{34}\text{S}(\text{p},\text{n})^{34\text{m}}\text{Cl}$	1–3 $\mu\text{A}$		30 min		Cu/Ni/Al alloys	$\leq 0.5 \mu\text{g per g}$	Chemical separation of $^{34\text{m}}\text{Cl}$ needed (d)
H <sup>+</sup> /12	$^{34}\text{S}(\text{p},\text{n})^{34\text{m}}\text{Cl}$	5 $\mu\text{A cm}^{-2}$		30 min		Fe, Al metal	$< 0.1 \mu\text{g per g}$	1 h measurement (e)
$^3\text{H}^+$ /12	$^{32}\text{S}(\text{t},\text{n})^{34\text{m}}\text{Cl}$	1 $\mu\text{A}$		1 h		Al	10 ng per g	(f)
$\gamma$ -Rays (a)	$^{32}\text{S}(\gamma,\text{n})^{31}\text{S}$		2.7 s	10 min/4 h		River sediment	0.2 $\mu\text{g}$	Large sample mass (~5 g) required (g)

(a) Bremsstrahlung from 30 MeV (100 mA) electron beam striking a W-target.

(b) Thomas J-P and Schweikert EA (1972) A rapid method for assaying sulfur using proton activation analysis. *Nuclear Instruments and Methods* 99: 461–467.

(c) McGinley JR and Schweikert EA (1975) Determination of lithium, boron and carbon by quasi-prompt charged particle activation analysis. *Analytical Chemistry* 47: 2403–2407.

(d) Pinta M (1978) *Modern Methods for Trace Element Analysis*. Ann Arbor: Ann Arbor Science.

(e) Vandecasteele C, Dewaele J, Esprit M, and Goethals P (1980) The determination of sulfur in copper, nickel and aluminum alloys by proton activation analysis. *Analytical Chimica Acta* 119: 121–127.

(f) Dabney SA, Swindle DL, Beck JN, Francis G, and Schweikert EA (1973) On the determination of sulfur by charged particle activation analysis. *Journal of Radioanalytical Chemistry* 16: 375–383.

(g) Bordes N, Blondiaux G, Maggiore CJ *et al.* (1987) Analytical possibilities of medium energy tritium beams ( $3 < E < 12$  MeV) and application to the analysis of oxygen in InP. *Nuclear Instruments and Methods in Physics Research B* 24: 722–724.

(h) Parsons ML (1976) Nuclear methods. In: Winefordner JD (ed.) *Trace Analysis, Spectroscopic Methods for Elements*. New York: Wiley.

chromium or copper anode in the X-ray tube. The determination of sulfur by X-ray detection poses several practical problems. As the energy of the characteristic X-rays is low (S-K $\alpha$  2.307 keV) absorption problems play an important role. Also the fluorescence yield of sulfur is rather low. The low-energy SK X-rays are usually measured in a vacuum, preferably by a windowless (or ultrathin window) X-ray detector. As a result, with techniques based on the detection of characteristic X-rays the achievable detection limits for sulfur are rather poor. For accurate sulfur determinations also large amounts of sample material and standards are required. An advantage of XRF is that chemical species analysis can be carried out by a high-resolution spectrometer.

The total reflection of primary X-rays (TXRF) has been employed to improve the XRF detection limits by decreasing the radiation-induced background. In a typical analysis, the sample material is prepared as a thin film, placed on a polished, optically flat carrier, and irradiated via a very small angle. TXRF is a convenient analysis technique because of the low sample masses required. The major drawback of this method is that only thin-film-type samples can be analyzed. It has been used for sulfur analysis from environmental, mineralogical, and biological samples.

Synchrotron radiation induced wavelength-dispersive XRF has gained increased attention as a sensitive technique for various analytical applications. It has been used, for example, in the analysis of ion-implanted sulfur from silicon wafers.

### Nuclear Activation Methods

Several methods are based on the measurement of radiation resulting from nuclear reactions induced either by neutrons, charged particle, or high-energy photon irradiation. These methods are generally multielemental and nondestructive.

**Neutron activation analysis** Neutron activation analysis (NAA) is not generally considered as a practical method for sulfur determination, but several reactions employing fast (14 MeV) or thermal neutrons may be used. The various applications of NAA have been collected in Table 2.

**Charged particle activation analysis** The method based on charged particle activation analysis utilizes delayed instrumental or radiochemical measurement of nuclear reaction-produced radionuclides in the same manner as in NAA. It differs from the other activation analysis techniques because of the limited penetration range of the charged particles used. This feature, however, is advantageous as it allows

**Table 4** Charged particle based prompt methods for sulfur

Method	Reaction	Ion/energy (MeV)	Measurement time/collected charge	Sample	Detection limit	Remarks	
PIXE		H <sup>+</sup> /3	3 $\mu$ C	Single animal cell	7.5 $\mu$ g per g	Microbeam, spatial resolution 1 $\mu$ m	(a)
		H <sup>+</sup> /2	50 $\mu$ C	Human brain tissue	20 $\mu$ g per g	In vacuum, beam diameter 2 mm	(b)
HIXE		Ar <sup>+</sup> /6	30 min	Atm. gas	1–10 ppm	Direct analysis for S in polluted air	(c)
HIXSE		Cl <sup>+</sup> /30	60 $\mu$ C	Insulator material	< 0.2 at%	Depth profiling, implanted samples	(d)
PIGE	<sup>32</sup> S(p,p' $\gamma$ )	H <sup>+</sup> /2.5–4	1 mC	Rock	5.2 $\mu$ g per g		(e)
		H <sup>+</sup> /3.4	1 mC	Geological	1.8 $\mu$ g	DL for thin target	(f)
					3.5 $\mu$ g per g	DL for thick target	(f)
		H <sup>+</sup> /3.43		GaSb	700 $\mu$ g per g	Interference from P, Si	(g)
	<sup>32</sup> S(t,n) <sup>34</sup> Cl	H <sup>+</sup> /4.5	30 min	Pyrite	1700 $\mu$ g per g	Macrobeam	(h)
					3200 $\mu$ g per g	Microbeam	(h)
		H <sup>+</sup> /5	5 min	Aerosol	100 ng cm <sup>-2</sup>		(i)
			5 min	Organic	100 $\mu$ g per g	Thick sample	(i)
	<sup>32</sup> S(t,d) <sup>33</sup> S	H <sup>+</sup> /7	1000 s	Aerosol	$\sim \mu$ g cm <sup>-2</sup>	External beam	(j)
		<sup>3</sup> H <sup>+</sup> /3	1 h	Nb	15 ppm	Interference from Ni	(k)
		<sup>3</sup> H <sup>+</sup> /3	1 h	Nb	20 $\mu$ g per g		(k)
ERDA		<sup>197</sup> Au/48		CuInSe <sub>2</sub> film	0.1 at%	TOF-ERDA	(l)

(a) Przybyłowicz NJ, Mesjasz-Przybyłowicz J, Pineda CA *et al.* (2001) Elemental mapping using proton-induced X-Rays. *X-Ray Spectrometry* 30: 156–163.

(b) Boruchowska M, Lankosz M, Adamek D, and Korman A (2001) PIXE analysis of human brain tissue. *X-Ray Spectrometry* 30: 174–179.

(c) Heitz Ch, Cailleret J, Iturbe J, Lagarde G, and Siffert P (1981) On the possibility of immediate analysis of chlorine and sulfur in air by argon ion induced X-ray emission. *Nuclear Instruments and Methods* 191: 558–564.

(d) Rosseel TM, Vane CR, Young JP, Zuhre RA, and Peterson RS (1989) High resolution HIXSE studies of the chemical environment of sulfur implanted in quartz glass. *Nuclear Instruments and Methods in Physics Research B* 43: 14–18.

(e) Borderie B (1980) Present possibilities for bulk analysis in prompt gamma-ray spectrometry with charged projectiles. *Nuclear Instruments and Methods* 175: 465–482.

(f) Clark PJ, Neal GF, and Allen RO (1975) Quantitative multielement analysis using high energy particle bombardment. *Analytical Chemistry* 47: 650–658.

(g) Chemin JF, Roturier J, Saboya B, and Petit GY (1972) Microanalysis of Si, S and Zn in GaSb by use of direct observation of (p,p' $\gamma$ ) reactions. *Journal of Radioanalytical Chemistry* 12: 221–232.

(h) Courel P, Trocellier P, Mosbah M *et al.* (1991) Nuclear reaction microanalysis and electron microanalysis of light elements in minerals and glass. *Nuclear Instruments and Methods in Physics Research B* 54: 429–432.

(i) Räsänen J and Lapatto R (1988) Analysis of sulfur with external beam proton induced gamma-ray emission analysis. *Nuclear Instruments and Methods in Physics Research B* 30: 90–93.

(j) Macias ES, Radcliffe CD, Lewis CW, and Sawicki CR (1978) Proton induced  $\gamma$ -ray analysis of atmospheric aerosols for carbon, nitrogen and sulfur composition. *Analytical Chemistry* 50: 1120–1124.

(k) Borderie B and Barrandon JN (1978) New analytical developments in prompt gamma-ray spectrometry with low-energy tritons and alpha particles. *Nuclear Instruments and Methods* 156: 483–492.

(l) Kemell M, Ritala M, Saloniemi H *et al.* (2000) One-step electrodeposition of Cu<sub>2-x</sub>Se and CuInSe<sub>2</sub> thin films by the induced co-deposition mechanism. *Journal of The Electrochemical Society* 147: 1080–1087.

near-surface analysis. The charged particles used in sulfur detection are usually protons (see Table 3 for details). Medium energy tritium beams (3–12 MeV) have also been used.

**Photonuclear activation analysis** In photonuclear activation analysis, high-energy photon irradiation of a sample is used to produce radionuclides that are subsequently measured by  $\gamma$ -ray spectroscopic methods.

These energetic photons, like neutrons, are capable of penetrating into thick solid samples. This allows bulk analysis without chemical pretreatment. A drawback of this method is the large sample mass required.

### Prompt Ion Beam Methods

The best known prompt accelerator based methods suitable for sulfur determination are particle-induced

X-ray emission (PIXE), heavy ion-induced X-ray emission (HIXE), particle-induced  $\gamma$ -ray emission (PIGE), Rutherford backscattering spectrometry (RBS), and elastic recoil detection analysis (ERDA). These methods are also multielemental and nondestructive. In general, when ion beam methods are used it should be kept in mind that sulfur-containing matter may be lost during the irradiation, and therefore sufficiently low beam currents should be employed. Also, sample homogeneity is vital since the volume probed by the ions is rather small.

**Particle-induced X-ray emission** In PIXE, mainly 1–3.5 MeV protons are used as bombarding particles. There are several studies in the literature concerned with investigations of the optimum bombarding proton energy for sulfur analysis. According to these findings, 1 MeV protons provide the best detection limits for sulfur analysis. As sulfur vaporizes easily when bombarded in a vacuum, an external beam mode where the particle beam is extracted into the ambient through a thin exit foil has been frequently used. The best results for sulfur analysis are obtained if helium is used as the ambient. Many types of samples have been analyzed for sulfur by PIXE; for example, aerosols collected on a filter, pigments, algae, artifacts, glass, and fingernails. The use of proton microprobes for elemental mapping (including sulfur) using PIXE has recently gained more interest in environmental and biological studies. A 2 MeV deuteron bombardment technique has also been tested simultaneously with nuclear reaction measurements for sulfur analysis. However, the sensitivity obtained is not as good as that with protons. Low-energy (below 10 MeV) heavy ion beams (HIXE) may be successfully used for sulfur analysis. Heavy ion induced X-ray satellite emission (HIXSE) spectroscopy together with efficient, high-resolution spectrometer has been used to obtain chemical information from sulfur-containing matrixes. For details, see Table 4.

**Particle-induced  $\gamma$ -ray emission** In PIGE, the prompt  $\gamma$ -rays from various reactions are detected during irradiation. Tritons ( $^3\text{H}$ ) as well as protons have been used as bombarding particles in this application. Alpha-particles do not produce  $\gamma$ -rays from sulfur bombardment at practical ion energies.

Of the several applicable reactions for this application,  $^{32}\text{S}(\text{p},\text{p}'\gamma)$  is the most frequently used. Information on the reactions used is collected in Table 4. An important advantage of PIGE with protons is the possibility of performing light element depth profiling. The method is based on the fact that in the megaelectronvolt energy range, the cross-sections for

reactions between protons and light nuclei exhibit sharp resonances. For sulfur depth profiling via the nuclear resonance broadening method, the resonance at 677 keV proton energy of the reaction  $^{34}\text{S}(\text{p},\gamma)$  has been employed successfully.

**Rutherford backscattering spectrometry and elastic recoil detection analysis** Conventional RBS analysis may be used for thin film sulfur analysis. Although the method in this case is not very sensitive, it has the advantage that depth profiles for sulfur and film thicknesses may be obtained using it. Ion channeling can be taken advantage of for studying near surfaces of crystalline materials. For example, microprobe channeling has been combined with PIXE analysis to determine the lattice location of sulfur in crystalline materials.

Heavy ion recoil spectroscopy can be used for analysis of surface layers and it has shown its usability as a characterization method especially to meet the needs of the semiconductor industry. The main reason for its success is that it provides depth profiling of all the sample atoms in one measurement. An example of a typical application of heavy ion ERDA is provided in Table 4.

### Miscellaneous Analytical Techniques

Laser-induced breakdown spectrometry enables simultaneous analysis of various elements in a short time. The obtained sensitivity for sulfur is not, however, as good as with the ICP techniques. For example, a detection limit of 45  $\mu\text{g}$  per g for sulfur in steel has been obtained using this technique.

Spectrochemical methods such as IR detection of  $\text{SO}_2$  formed after high-temperature combustion require large amounts of sample material and also standards for accurate sulfur determination. A detection limit of 0.2  $\mu\text{g}$  per g has been reported for sulfur in high-purity iron by this technique.

Flame photometric detection allows a fast and fully automated means for sulfur analysis. Typical limits of detection obtainable are 1.2 ng for solids and 0.2 ng for liquids.

Accurate determination of elemental sulfur in petroleum and its distillates (petroleum products) is of significant industrial importance. It is being determined routinely by several techniques, for example, by the differential pulse polarography technique. The detection limit is 0.1  $\mu\text{g}$  per g, but chemical treatment of the sample is needed. The wet chemical method of activated Raney nickel has been successfully employed with a detection limit of 0.1  $\mu\text{g}$  per g. Also in this case preconcentration of sulfur is needed. By using the Houston Atlas sulfur analyzer in which

total sulfur is reduced to  $\text{H}_2\text{S}$  a detection limit of  $<20 \text{ ng g}^{-1}$  can be achieved.

Electrode materials of electrochemical cells have been analyzed by a method in which free sulfur is first isolated from the sample by solvent extraction and then determined by molecular emission cavity analysis. The lowest limit of detection for free sulfur that can be analyzed is  $0.8 \mu\text{g per g}$ .

An objective evaluation of the methods on the basis of literature data is hardly possible, because of differences in experimental conditions, sample, amount of sample available, and measurement time. Therefore, the detection limit values stated in the literature given in **Tables 1–4** should be considered, at best, approximate. But it is clear that there is no single instrumental technique that meets all the analytical requirements. For example, some methods may be applicable over only a limited concentration range, may be subject to matrix effects or spectral interferences, or have a limited availability. Also, sometimes nondestructive fast methods are needed. The choice of an instrumental method depends on the material to be analyzed and the type of analysis required.

## Areas of Special Interest, Use, and Safety Considerations

The uses of sulfur are many and varied. Large quantities of elemental sulfur are used in the vulcanization of natural rubber, in lime–sulfur sprays to destroy plant parasites, in the manufacture of artificial fertilizers and certain types of cements and electric insulators, in certain ointments and medications, and in the manufacture of gunpowder and matches. Sulfur is a critical element in defining the physical properties of steel. It is a contaminant element originating from the blast furnace iron-making process and not desired at high levels. In controlled amounts it can, however, improve the machinability of steel. Sulfur is known also to have an important influence on the magnetic properties and embrittlement of some alloys. As an example of the new uses of sulfur is its key role in the formation of filled carbon nanotubes in nanotechnology.

Most of the sulfur currently used is burnt to form sulfur dioxide for the use in the pulp and paper industry, for bleaching, and for conversion to sulfuric acid. Sulfuric acid is generally considered to be the most important manufactured chemical. Compounds of sulfur have many commercial uses, for example, in the manufacture of chemicals, textiles, soaps, fertilizers, leather, plastics, refrigerants, bleaching agents, drugs, dyes, paints, paper, refined petroleum, and other products.

Due to the wide use of sulfur and its compounds, analytical applications for sulfur determination are needed in many fields of, e.g., technology, biology, environment, medicine, geology, fine arts, and archaeology. As sulfur is often a common industrial pollutant, testing environmental samples for the element has become a matter of significant interest.

Sulfur is a nontoxic element. However, carbon disulfide, hydrogen sulfide, and sulfur dioxide should be handled carefully. Small concentrations of hydrogen sulfide can be metabolized, but at higher concentrations it can quickly cause death by respiratory paralysis. It is insidious in that it quickly deadens the sense of smell. Sulfur dioxide is a dangerous component in atmospheric air pollution.

**See also:** **Activation Analysis:** Neutron Activation; Charged-Particle Activation; Photon Activation. **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Mass Spectrometry:** Overview. **Surface Analysis:** Particle-Induced X-Ray Emission; Auger Electron Spectroscopy; Ion Scattering; Nuclear Reaction Analysis and Elastic Recoil Detection. **X-Ray Fluorescence and Emission:** Wavelength Dispersive X-Ray Fluorescence; Energy Dispersive X-Ray Fluorescence.

## Further Reading

- Beauchemin D, Yves Le Blanc JC, Peters GR, and Craig JM (1992) Plasma emission spectrometry. *Analytical Chemistry* 64: 442R–467R.
- Erdtmann G and Petri H (1986) Nuclear activation analysis: Fundamentals and techniques. *Treatise on Analytical Chemistry* 14: 419–643.
- Godovac-Zimmermann J and Brown LR (2001) Perspectives for mass spectrometry and functional proteomics. *Mass Spectrometry Reviews* 20: 1–57.
- Jarvis KE, Gray AL, and Houk RS (1992) *Handbook of Inductively Coupled Plasma Mass Spectrometry*. Glasgow: Blackie.
- Johansson SAE, Campbell JL, and Malmqvist KG (1995) *Particle-Induced X-Ray Emission Spectrometry (PIXE)*. Chichester: Wiley.
- Knauth B and Otto M (2001) Nonmetals in the argon-inductively coupled plasma-optical emission spectrometry: I. Phosphorus, sulfur and carbon. *Fresenius Journal of Analytical Chemistry* 371: 1052–1056.
- Nutsch A, Erdmann V, Zielonka G, Pfitzner L, and Ryssel H (2001) Trace analysis for 300 mm wafers and processes with total reflection X-ray fluorescence. *Spectrochimica Acta B* 56: 2301–2306.
- O'Reilly JW, Dicoski GW, Shaw MJ, and Haddad PR (2001) Chromatographic and electrophoretic separation of inorganic sulfur and sulfur-oxygen species. *Analytica Chimica Acta* 432: 165–192.

Riciputi LR (1996) A comparison of extreme energy filtering and high mass resolution techniques for the measurement of  $^{34}\text{S}/^{32}\text{S}$  ratios by ion microprobe. *Rapid Communications in Mass Spectrometry* 10: 282–286.

Steinmeyer S and Kolbesen BO (2001) Capability and limitations of the determination of sulfur in inorganic and

biological matrices by total reflection X-ray fluorescence spectrometry. *Spectrochimica Acta B* 56: 2165–2173.

Tesmer JR and Nastasi M (eds.) (1995) *Handbook of modern ion beam materials analysis*. Pittsburgh: Materials Research Society.

# SUPERCRITICAL FLUID CHROMATOGRAPHY

Contents

**Overview**

**Applications**

## Overview

**T A Berger**, Berger Instruments, Newark, DE, USA

© 2005, Published by Elsevier Ltd.

## Introduction

Supercritical fluid chromatography (SFC) can be viewed as a subset of high-performance liquid chromatography (HPLC), sharing much of the same hardware and methodology. As in all chromatography, individual components of complex mixtures are separated from each other by differential interaction between a stationary phase and a mobile phase. In SFC, the mobile phase consists of a relatively dense, compressible fluid 'that acts as a solvent'. Instrumentally, the biggest difference between HPLC and SFC is the inclusion of a backpressure regulator downstream of most detectors that prevents the compressed mobile phase from expanding to a gas until after the separation is complete.

The name 'supercritical fluid' is obscure and confusing. Fluids have a characteristic 'critical point', discovered ~180 years ago, consisting of a critical temperature and pressure. Above (or 'super'-meaning above) this critical point, no increase in pressure, or temperature can cause two phases to form.

In a phase diagram, solid lines represent conditions where two phases are in contact. Crossing a line requires a phase transition and a massive energy transfer. A phase diagram of pure carbon dioxide is presented in **Figure 1** showing the regions of temperature and pressure defined as gas, liquid, and solid. The long diagonal line indicates the conditions where gas is in equilibrium with liquid. To the right, this line ends at the critical point. Some people

incorrectly draw horizontal and vertical lines (often dashed) extending from the critical point to the right and up to outline the supercritical region. There is no phase transition between a liquid and a supercritical fluid or between a gas and a supercritical fluid and such lines should not be drawn.

In SFC, the fluid must act as a solvent for the solutes of interest. Secondly, it should exhibit physical and chemical characteristics equal to or superior to normal solvents used as the mobile phase in HPLC. Two of the more intriguing characteristics of the fluids of interest are high diffusivity and low viscosity.

## Why SFC?

### Higher Diffusivity Equals Higher Speed Chromatography

In chromatography, the speed of separations is dictated in part by how fast the solute diffuses in the mobile phase. Generally, the more dense the mobile phase, the slower the diffusion and slower the separation. Typical solute binary diffusion coefficients ( $D_{1,2}$ ) in the gases used as the mobile phase in gas chromatography (GC) are in the order of  $0.1\text{--}1\text{ cm}^2\text{ s}^{-1}$ , while in liquids used as the mobile phase in HPLC are in the order of  $10^{-5}\text{ cm}^2\text{ s}^{-1}$ . In SFC using carbon dioxide as the mobile phase,  $D_{1,2}$  can be as high as  $10^{-3}\text{ cm}^2\text{ s}^{-1}$  when the fluid density is below  $0.1\text{ g cm}^{-3}$ , to as low as  $10^{-4}\text{ cm}^2\text{ s}^{-1}$  when the density approaches  $1\text{ g cm}^{-3}$ . Note that even at liquid-like densities, the diffusion coefficient remains substantially higher in the supercritical (and 'near-critical') fluid.

Higher diffusion coefficients translate directly to higher speed chromatography as indicated in the van Deemter equation. **Figure 2** is a van Deemter plot of plate height (HETP) versus linear velocity for both

Riciputi LR (1996) A comparison of extreme energy filtering and high mass resolution techniques for the measurement of  $^{34}\text{S}/^{32}\text{S}$  ratios by ion microprobe. *Rapid Communications in Mass Spectrometry* 10: 282–286.

Steinmeyer S and Kolbesen BO (2001) Capability and limitations of the determination of sulfur in inorganic and

biological matrices by total reflection X-ray fluorescence spectrometry. *Spectrochimica Acta B* 56: 2165–2173.

Tesmer JR and Nastasi M (eds.) (1995) *Handbook of modern ion beam materials analysis*. Pittsburgh: Materials Research Society.

# SUPERCRITICAL FLUID CHROMATOGRAPHY

Contents

**Overview**

**Applications**

## Overview

**T A Berger**, Berger Instruments, Newark, DE, USA

© 2005, Published by Elsevier Ltd.

## Introduction

Supercritical fluid chromatography (SFC) can be viewed as a subset of high-performance liquid chromatography (HPLC), sharing much of the same hardware and methodology. As in all chromatography, individual components of complex mixtures are separated from each other by differential interaction between a stationary phase and a mobile phase. In SFC, the mobile phase consists of a relatively dense, compressible fluid 'that acts as a solvent'. Instrumentally, the biggest difference between HPLC and SFC is the inclusion of a backpressure regulator downstream of most detectors that prevents the compressed mobile phase from expanding to a gas until after the separation is complete.

The name 'supercritical fluid' is obscure and confusing. Fluids have a characteristic 'critical point', discovered ~180 years ago, consisting of a critical temperature and pressure. Above (or 'super'-meaning above) this critical point, no increase in pressure, or temperature can cause two phases to form.

In a phase diagram, solid lines represent conditions where two phases are in contact. Crossing a line requires a phase transition and a massive energy transfer. A phase diagram of pure carbon dioxide is presented in **Figure 1** showing the regions of temperature and pressure defined as gas, liquid, and solid. The long diagonal line indicates the conditions where gas is in equilibrium with liquid. To the right, this line ends at the critical point. Some people

incorrectly draw horizontal and vertical lines (often dashed) extending from the critical point to the right and up to outline the supercritical region. There is no phase transition between a liquid and a supercritical fluid or between a gas and a supercritical fluid and such lines should not be drawn.

In SFC, the fluid must act as a solvent for the solutes of interest. Secondly, it should exhibit physical and chemical characteristics equal to or superior to normal solvents used as the mobile phase in HPLC. Two of the more intriguing characteristics of the fluids of interest are high diffusivity and low viscosity.

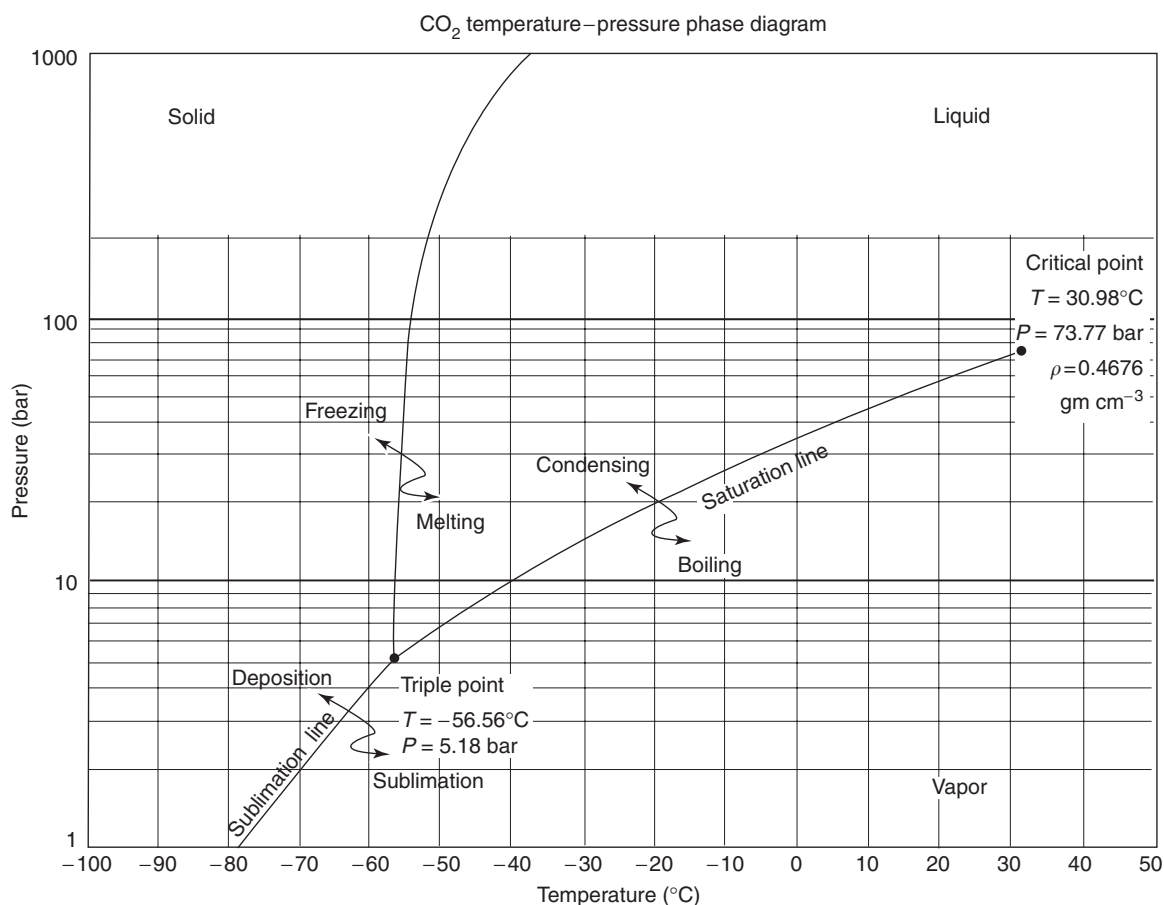
## Why SFC?

### Higher Diffusivity Equals Higher Speed Chromatography

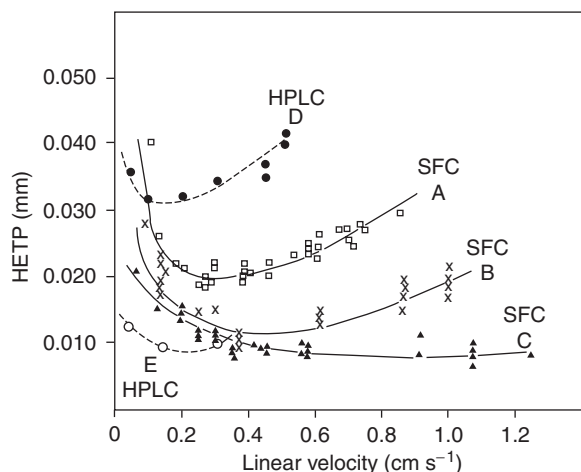
In chromatography, the speed of separations is dictated in part by how fast the solute diffuses in the mobile phase. Generally, the more dense the mobile phase, the slower the diffusion and slower the separation. Typical solute binary diffusion coefficients ( $D_{1,2}$ ) in the gases used as the mobile phase in gas chromatography (GC) are in the order of  $0.1\text{--}1\text{ cm}^2\text{ s}^{-1}$ , while in liquids used as the mobile phase in HPLC are in the order of  $10^{-5}\text{ cm}^2\text{ s}^{-1}$ . In SFC using carbon dioxide as the mobile phase,  $D_{1,2}$  can be as high as  $10^{-3}\text{ cm}^2\text{ s}^{-1}$  when the fluid density is below  $0.1\text{ g cm}^{-3}$ , to as low as  $10^{-4}\text{ cm}^2\text{ s}^{-1}$  when the density approaches  $1\text{ g cm}^{-3}$ . Note that even at liquid-like densities, the diffusion coefficient remains substantially higher in the supercritical (and 'near-critical') fluid.

Higher diffusion coefficients translate directly to higher speed chromatography as indicated in the van Deemter equation. **Figure 2** is a van Deemter plot of plate height (HETP) versus linear velocity for both





**Figure 1** Phase diagram of carbon dioxide. The supercritical region is to the right and above the critical point at the end of the line separating vapor from liquid (sometimes called the boiling line).



**Figure 2** A van Deemter plot of plate height (plate height = HETP = height equivalent of a theoretical plate) vs. linear velocity showing that SFC exhibits similar, low (good) plate heights but at much higher velocities (and flow rates) compared to HPLC. Lines A, B, and C represent efficiency vs. flow on 10, 5, and 3  $\mu\text{m}$  particles, respectively. Lines D and E represent HPLC on 10 and 3  $\mu\text{m}$  particles, respectively.

packed column HPLC and SFC using several different sized particles. The figure shows that the optimum linear velocity in SFC, yielding the lowest plate height, is substantially higher than the optimum velocity in HPLC, yielding the same plate height. Thus, the same quality of separation is achieved in less time in SFC compared to HPLC.

Similarly, low viscosity allows fluids to be pumped at higher linear velocities (flow rates) without the pressure drop across the column becoming excessive.

The desired high diffusivity and low viscosity results from the lack of strong intermolecular forces between solvent molecules. After all, if there were strong intermolecular forces, the solvent would condense to form a liquid. Without such strong interactions, the solute molecules are less impeded as they diffuse through the fluid. Despite weak intermolecular forces, when an external force is applied to push the molecules close together, they collectively act as a solvent.

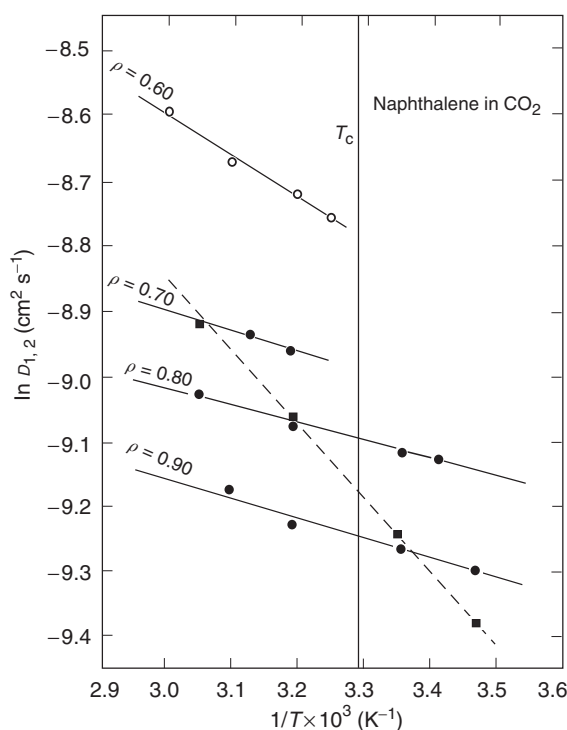
### Subcritical, Near-Critical, and Enhanced Fluidity Chromatography

For years, some users have differentiated between just subcritical and just supercritical conditions. This is an artificial distinction. For example, there has been an increase in the use of very hot water as a mobile phase. As the temperature is increased above 100°C, a backpressure regulator is required to keep the fluid from boiling (as in SFC). As temperature increases, diffusivity increases, while viscosity decreases, similar to SFC. The dielectric constant decreases, and water becomes less 'polar'. Temperature programming (usually from high to low temperatures) of pure water allows polarity tuning. The only negative aspect of water is the high temperatures required. Solute molecules and columns must be stable in water up to at least 300°C. This technique is sometimes called high-temperature HPLC or enhanced fluidity chromatography.

The general confusion about the physical and chemical meaning of the words 'super' and 'critical' has led to a plethora of terms such as 'subcritical' fluid chromatography, 'near-critical' fluid chromatography, and 'enhanced fluid' chromatography to deal with the fluids when they are nearly but not quite 'super' critical. All these names, including SFC, actually describe the same technique.

A subcritical fluid by definition is a liquid (or possibly a gas). Many workers have assumed that since some of the fluids used in SFC are defined as liquids they must have more typical liquid-like characteristics, such as much lower solute binary diffusion coefficients. Thus, crossing the border from just supercritical to just subcritical was assumed to cause a major change in physical properties (and was the tacit reason for the multiple names and the dashed lines in phase diagrams). To the contrary, some of the earliest work showed that diffusivity does not change significantly when the 'definition' of the fluid changes from just supercritical to just subcritical. There is no discontinuity, no change in slope, and nothing to support such assumptions. A plot of diffusion coefficients versus the reciprocal of temperature ( $1/T$ ), such as in Figure 3, produces a straight line even though the defined state changes from super- to subcritical in the middle of the temperature range.

SFC is actually misnamed because both just supercritical fluids and just subcritical fluids exhibit nearly identical chemically and physically desirable characteristics. Further, not all supercritical fluids act as solvents. Helium under some of the conditions found in a typical open tubular column in GC is supercritical but does not act as a solvent. In general, changing from supercritical to a gas or vice versa does not produce useful conditions. Thus, the term



**Figure 3** A plot of diffusion coefficient vs. the reciprocal of temperature at constant density for naphthalene in carbon dioxide. The line denoted  $T_c$  separates the supercritical region on the left from the subcritical (liquid) region on the right. Note that there is no discontinuity when the fluid changes from supercritical to subcritical. The solid lines represent constant densities, from top to bottom of 0.60, 0.70, 0.80, and 0.90 g cm<sup>-3</sup>, even though temperature is changing continuously from left to right. The dashed line indicates a constant pressure line.

supercritical fluid chromatography excludes many of the most desirable chromatographic conditions while including some completely useless conditions.

### Evolution of the Technique

Lovelock originally proposed SFC (he called it 'critical state chromatography') as an inorganic technique using compressed gases such as sulfur dioxide as the mobile phase, although apparently no experiments were performed.

Klesper, called the Father of modern SFC, used short-chain hydrocarbons such as pentane, sometimes modified with dioxane or other polar liquids heated to near 300°C. He and his co-workers reported eluting polymer homologs with molecular weights up to several million. The technique was usually called dense GC. The use of short-chain hydrocarbons at such high temperatures was widely considered to be both impractical and dangerous.

In 1968, Giddings proposed an elutrophic series (a rating of solvents for relative elution strength)

placing dense carbon dioxide beside isopropyl alcohol in solvent strength, based on estimated interaction parameters he used to calculate a Hildebrand solubility parameter. In retrospect, this estimation is clearly far too optimistic. Nevertheless, this series was used for many years, and was one basis for widespread disagreements among practitioners. Since Giddings, carbon dioxide has been the fluid of choice for SFC.

Significant effort was expended during the mid-1980s to extend SFC to the separation of smaller, more polar solutes. However, by the late 1980s there was widespread dissatisfaction with the apparent inability of SFC using pure carbon dioxide to separate more polar solutes. Many other fluids were tried including many chlorofluorocarbons, and fluorocarbons, nitrous oxide, and even xenon with similar poor results. The 'activity' of column packings was generally blamed for this lack of progress, not the (inadequately understood) inherent low polarity of the fluids used as the mobile phase.

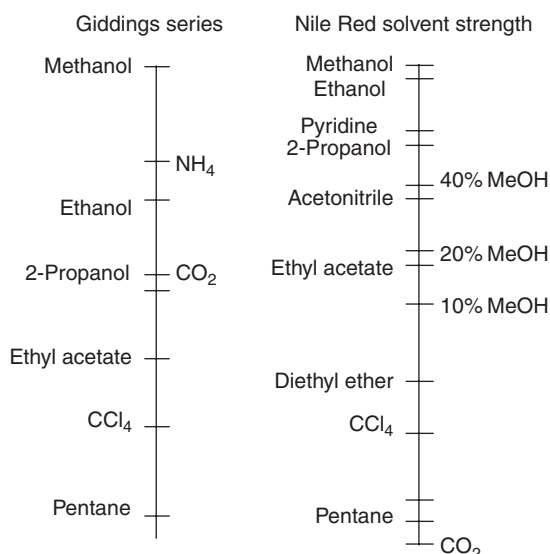
Practitioners broke into two camps. One pursued the use of pure carbon dioxide with apparently more inert open tubular columns, while the other experimented with more polar binary and ternary mobile phases and continued to use silica-based packed columns.

Today, packed column SFC is primarily practiced using binary or ternary mobile phases where a polar modifier or co-solvent is added to carbon dioxide. In some cases a third component such as trifluoroacetic acid or dimethylethylamine is also added to the modifier. The solvent strength of pure carbon dioxide is now accepted as somewhat similar to pentane, although the two are probably in different solvent families. In **Figure 4**, Giddings elutrophic series is compared to a modern polarity scale based on the solvatochromic dye (a dye that changes color depending on the 'polarity' of the solvent sheath surrounding it), Nile Red.

## Modifiers

When polar, liquid modifiers are added to the non-polar main fluid they tend to negate the speed advantage of the technique more or less in proportion to the concentration of modifier. If SFC with pure carbon dioxide at a density of  $0.85 \text{ g cm}^{-3}$  is 10 times faster than HPLC, then 20% methanol in carbon dioxide is only five times faster, while 50% methanol is only two times faster.

Fortunately, solvent strength is a highly nonlinear function of modifier concentration. The first small additions of modifier can create very large increases in solvent strength. Methanol is one of the most polar solvents completely miscible with carbon

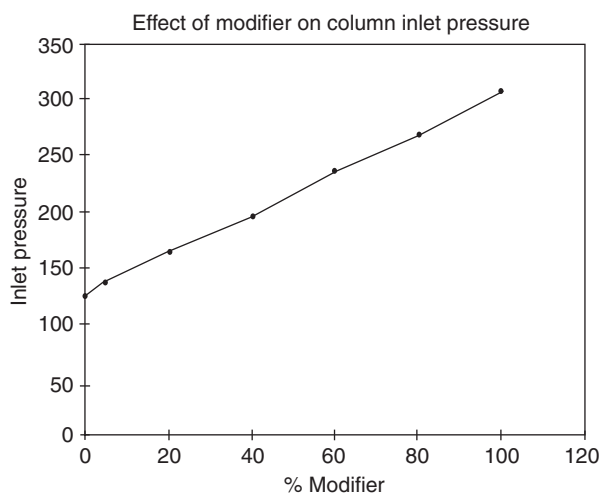


**Figure 4** Giddings elutrophic series originally presented in a *Science* article in 1968 is replotted on the left, while a solvent strength scale based on the solvatochromic dye, Nile Red, is presented on the right. Note that pure carbon dioxide is much less polar on the Nile Red scale than in Giddings series. Significant concentrations of polar modifier are required (like ~15% methanol) in the Nile Red scale to equal the estimated value in Giddings series. It is now nearly universally accepted that the Giddings estimate was much too optimistic.

dioxide. Using solvatochromic dyes, one can measure the solvent strength of mixtures. The addition of 1% methanol to carbon dioxide increases the apparent solvent strength to 10% of the difference between the two pure solvents, while 10% modifier increases the solvent strength to ~50% of the difference as indicated in **Figure 4** for the binary fluid methanol-carbon dioxide.

This nonlinear relationship means that small additions of modifier cause large increases in solvent strength while retaining much of the speed advantage of SFC over HPLC. The general rule of thumb is that SFC, with modifiers is three to five times faster than HPLC.

The viscosity of binary supercritical and near-critical fluids is also much lower than normal liquids, due, again to a lack of intermolecular forces. The addition of modifier increases viscosity more or less in direct proportion to modifier concentration as shown in **Figure 5**. This low viscosity is helpful in that the pressure drop across a column is as little as 1/10th of that in HPLC, at the same flow rate. Since SFC is three to five times 'faster', the flow rate of the mobile phase must be three to five times higher. Thus, even at the higher flow rate the pressure drop tends to be lower than in HPLC. The lower pressure drop can be exploited for using smaller particle packings or longer columns.



**Figure 5** Pressure drop as a function of modifier concentration. The pressure drop is directly proportional to viscosity. With pure carbon dioxide pressure drop is 9 bar. With pure methanol pressure drop is  $\sim 200$  bar. Flow rate ( $5 \text{ ml min}^{-1}$ ), outlet pressure (125 bar), and temperature ( $35^\circ\text{C}$ ) were all held constant. Column:  $4.6 \times 150 \text{ mm}$ ,  $6 \mu\text{mol l}^{-1}$  BI Cyano. Changing column inlet pressure has a significant impact on the compressibility of the carbon dioxide, which is a function of pressure. Without adequate dynamic compressibility compensation the delivery from the carbon dioxide pump will roll off as the pressure increases (making the total flow and actual composition unknowns).

As a 'rule of thumb', any compound soluble in methanol or a less polar molecule is an ideal candidate for separation by SFC. Any solutes requiring an aqueous or buffered aqueous environment is probably a poor candidate. Peptides are an example of an application where SFC has had little or no impact. On the other hand, lipophilic peptides up to 30 amino acids have been routinely separated. Most small drug-like compounds like those intended for oral administration are ideally suited to SFC. The 'limits' of SFC are poorly defined.

## Where does SFC Fit into Separation Science?

As indicated above, there are two forms of SFC although one is seldom performed today. Open tubular SFC is most often thought of as an extension of GC to heavier, less volatile solutes by replacing some thermal energy with solvation energy. Packed column SFC can be thought of as a subset of HPLC, employing binary and ternary mobile phases and gradient composition programming.

### Open Tubular SFC

Open tubular SFC was relatively popular during the 1980s employing a single pump, usually a large

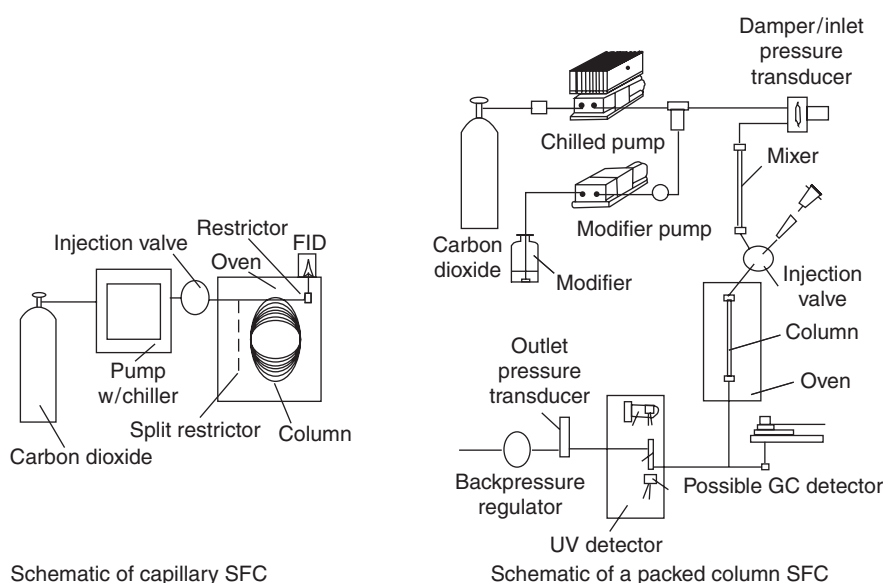
syringe pump, delivering pure fluids, used as a pressure source, and a column oven with a flame ionization detector (FID) (see Figure 6). Columns tended to be  $50 \mu\text{m}$  ID, by 10 m long with thin coatings of bonded stationary phases. Very small volume (i.e., 60 nl), high pressure injection valves were used, sometimes with timed splits to minimize the amount of sample put on-column. Pressures used ranged between 50 and 600 bar, while temperatures tended to be  $70\text{--}200^\circ\text{C}$ . Pressure programming allowed the user to progressively increase the solvating power of the fluid, somewhat analogous to temperature programming in GC. Since pure fluids are all nonpolar, the technique was limited to relatively nonpolar solutes.

Today, few new applications of open tubular SFC are developed and no hardware is manufactured specifically for this use. The general lack of interest in open tubular SFC stems from the inherent low polarity of carbon dioxide, and the lack of other more polar alternatives. There are also performance issues.

Even on the small diameter columns used (i.e.,  $50 \mu\text{m}$ ) diffusion coefficients dictated an optimum linear velocity ( $\mu_{\text{opt}}$ ) in the order of  $0.1 \text{ cm s}^{-1}$ . Since columns tended to be 10 m long, the optimum column hold-up time was in the order of 16.7 min. Since this is unrealistically long for routine analysis, most workers operated up to 10 times  $\mu_{\text{opt}}$ , making the column hold-up time less than 2 min with typical run times of 30 min to 2 h. Efficiency dropped from 200 000 plates near optimum to roughly 20 000 plates at the higher velocity. Very small column diameters made it difficult to make reproducible injections.

The biggest problem with open tubular SFC arose from the use of fixed restrictors to limit the flow of the mobile phase through the column. There are no dynamically adjustable devices capable of controlling the minute flows required in open tubular SFC. 'Integral' restrictors were initially made by heating the end of the fused silica column until it closed, then grinding until an opening of  $1\text{--}2 \mu\text{m}$  was achieved. This end was then inserted into the base of the FID. Obviously, such tiny holes make good filters for the ubiquitous dust particles found throughout chromatographs. As the holes plugged, their flow characteristics changed. 'Frit' restrictors consist of a plug of microporous 'water glass' deposited in the end of a  $50 \mu\text{m}$  tube. The length is trimmed to attain a specific flow. Since there are multiple flow paths, it is more difficult to plug.

All fixed restrictors share a common problem to a greater or lesser degree. As the pressure is programmed from low to high values, the density of the



**Figure 6** Schematic representations of capillary (open tubular) SFC and packed column SFC.

fluid in the column increases, causing the diffusion coefficients of the solutes to decrease. Since the density is typically increased from  $\sim 0.05\text{--}0.75\text{ g cm}^{-3}$ , the diffusion coefficients drop  $\sim 15$  times. To maintain constant efficiency, the linear velocity of the mobile phase should also decrease 15 times in proportion to the density increase. Unfortunately, the mass flow through a fixed restrictor actually increases with the applied pressure. The best fixed restrictors allow a modest increase in linear velocity while the worst cause a major increase. Thus, efficiency changes significantly from the beginning to the end of a pressure program.

Since the initial efficiency is usually degraded by starting at near 10 times optimum, fixed restrictors tend to cause a further ( $\sim 10\text{--}15\times$ ) degradation as pressure is increased. In the analysis of a homologous series, the greatest efficiency is needed near the end of the run where the differences between consecutive members of the series are the least. With fixed restrictors, efficiency is at its worse in this region of the run.

### Packed Column SFC

Modern SFC is much closer to HPLC than GC in that it is performed almost entirely on packed columns and the solvent strength of the mobile phase is varied by adding a polar organic modifier like methanol to the main, nonpolar fluid such as carbon dioxide. Since one fluid is under pressure and the modifier is usually a normal liquid, gradient valves cannot be used. Consequently, packed column SFC requires

more complex hardware (see **Figure 6**), including at least two high pressure pumps. Standard HPLC-like injection valves are used to make injections. A column oven is highly desirable although temperatures are most often controlled in the  $20\text{--}70^\circ\text{C}$  range. Columns are typically  $15\text{--}25\text{ cm}$  long,  $2\text{--}4.6\text{ mm}$  ID (for analytical) with  $3\text{--}6\text{ }\mu\text{m}$  silica-based packings. The most common detector remains the ultraviolet-visible (UV-Vis) light absorbance detector, used in HPLC, with a high pressure flow cell. After the detector, a backpressure regulator controls the column outlet pressure to insure that there is a single phase throughout the chromatographic flow path.

**SFC is normal phase chromatography** Perhaps  $80\text{--}90\%$  of HPLC is 'reversed phase', using aqueous based polar mobile phases and nonpolar stationary phases such as  $\text{C}_{18}$ . The fluid composition is programmed from polar (i.e., pure water) to less polar (more organic). Initially, HPLC was a 'normal phase' technique in which the fluid composition was programmed from less polar (hexane) to more polar (ethanol) during the course of a separation and the stationary phase (silica) was more polar than the mobile phase. Normal phase HPLC has been largely replaced by reversed phase because it is slow, has very slow reequilibration, and tends to have poor reproducibility.

SFC can be viewed as a rebirth of normal phase chromatography with vastly better physical and chemical characteristics. As indicated previously, the optimum flow rate is three to five times higher than even reversed-phase HPLC and reequilibration



is as much as 10 times faster. The last major use of normal phase HPLC is in chiral separations of enantiomers that are mirror images of each other. The separation of enantiomers is a very important aspect of pharmaceutical development. SFC is rapidly replacing normal phase HPLC in this application.

There are a wide range of polar stationary phases available for SFC. They include: bare silica, cyano, amino, diol, ethylpyridine, and many others. Virtually all chiral stationary phase types have been used in SFC.

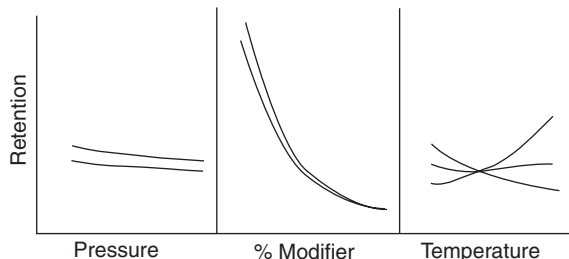
**Method development** There are three instrumental parameters that affect retention and selectivity: mobile phase composition, pressure, and temperature. Their relative effects are represented schematically in Figure 7. The composition of the mobile phase has the most effect on the retention of polar solutes but tends to have little effect on selectivity. Composition programming is used for gradient elution. Temperature tends to have only a moderate effect on retention but can have a relatively large effect on selectivity. Pressure tends to be a secondary control variable with a moderate effect on retention but little effect on selectivity. The stationary phase affects both retention and selectivity.

**Finding a mobile phase** There is no fundamental principle that makes SFC method development easier to understand than HPLC. On the contrary, there are more variations in SFC, which can be used to further optimize separations. The advantage of SFC lies in its speed and rapid reequilibration after perturbations. Many more experiments can be run per unit time to empirically find the best conditions.

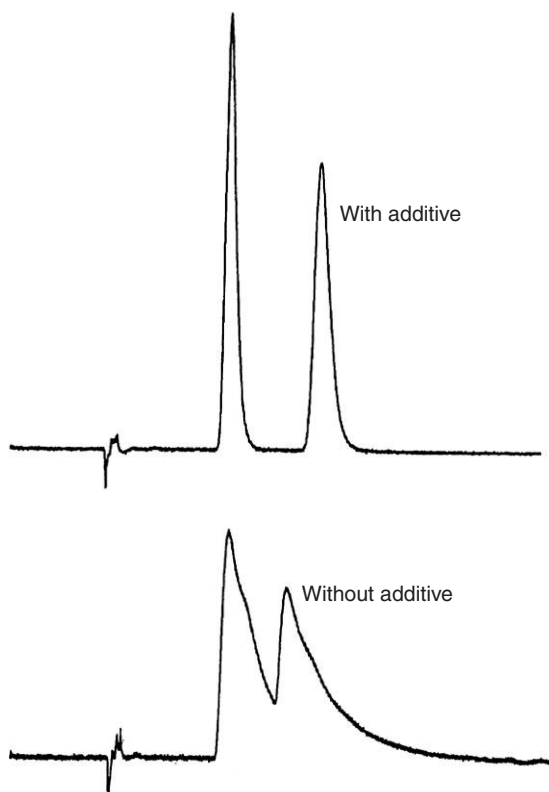
The most common mobile phase is methanol-carbon dioxide. Methanol is the most polar common

solvent completely miscible with carbon dioxide. However, methanol is sometimes too polar with 1% or 2% methanol yielding little or no solute retention. The methanol can be replaced with ethanol or isopropanol or an even longer chain alcohol. One can also change to a different solvent family, similar to HPLC. Acetonitrile is possibly the most common nonalcoholic modifier, but practically any organic solvent can be used.

**Additives** Many highly polar solutes, such as primary aliphatic amines either do not elute or elute with poor peak shapes when binary fluids such as methanol in carbon dioxide are used as the mobile phase. The addition of a third component, usually called an additive to the mobile phase usually dramatically improves peak shapes, as shown in Figure 8 for propranolol. The additive is placed in the modifier and the mixture is pumped as a single fluid. In general, basic solutes require a basic additive and acidic solutes require an acidic additive. Ion pairing (adding an acid to a base or a base to an acid) seldom appears to work. Common additives include



**Figure 7** Schematic representations of the effects of mobile phase composition, pressure, and temperature on retention in packed column SFC using binary fluids. Adjusting composition has the most effect on retention, but, since few of the lines cross, selectivity is only modestly affected. Temperature has little effect on retention but sometimes has a strong influence on selectivity. With binary fluids, pressure is a secondary control variable with modest influence on both retention and selectivity.



**Figure 8** The effect of additive on peak shapes. For solutes that are strong acids and bases, a simple binary mixture of methanol and carbon dioxide may not be polar enough, and peaks tail. Addition of a stronger acid or base as a minor component of the mobile phase often dramatically improves peak shapes as shown for propranolol.



trifluoroacetic acid (TFA) and dimethylethylamine (DMEA).

There appear to be several different possible mechanisms behind suppression of peak tailing but few systematic studies have been published. In general, the polarity limits of SFC remain poorly understood. It appears that the addition of small amounts of water (plus additives) to the methanol significantly increase the polarity of solutes that can be eluted.

The general approach to finding a mobile phase in method development is to quickly try a high concentration (i.e., 40%) of methanol in carbon dioxide. If the peaks emerge with little retention, try a lower concentration until the peaks all elute within a few minutes. The general 'rule of thumb' is that halving modifier concentration doubles retention. It is generally a good idea to not make very large changes in modifier concentration. Each experiment is very fast, so it is often worth the little extra time and effort to make multiple small changes (i.e., 40–20%, then 20–10%). If lowering the modifier concentration fails to yield significant retention, switch to a less polar modifier and repeat the process. If the peaks do not elute or elute with poor peak shapes at high modifier concentration, try an additive in the mobile phase.

**Finding a stationary phase** Finding an appropriate stationary phase is also a trial-and-error process. In general, however, phases similar to the solutes tend to give the best selectivity, giving some indication of where to start. Acids and alcohols tend to have the best selectivity on bare silica and diol phases. Bases tend to give the best separations on amino or ethylpyridine stationary phases. A cyano phase tends to give the least retention of all the polar stationary phases but also tends to yield the poorest selectivity. Similar compounds with different length hydrocarbon chains typically co-elute on all the polar phases. To separate such compounds a hydrocarbon phase, such as C<sub>4</sub>, C<sub>8</sub>, or C<sub>18</sub> may be appropriate.

Many modern packed column SFCs are equipped with a column selection valve and a solvent selection valve. Software 'Wizards' are available that allow simple programming where a large number of different stationary and mobile phases can be automatically tested to find the best conditions without operator intervention.

It is particularly difficult to predict the appropriate stationary phase for chiral separations (where the solutes are chemically identical mirror images of each other). The standard approach is to try a large number and see which of those work.

## Semipreparative SFC

Packed column SFC is scalable even to commercial production of semifinished products. Hardware similar to analytical scale instrumentation can be used with the addition of a device to recover the separated fractions. The recent widespread use of SFC for semipreparative separations has demonstrated several other practical advantages of SFC over HPLC. The fluids remain dense only as long as an external pressure is applied. If that pressure is relieved the fluid rapidly expands to a gas with little or no solvent power.

Separation of polar solutes generally requires polar modifiers like methanol. When the pressure is reduced at the end of the column, the mobile phase breaks down into two phases: one nearly pure carbon dioxide, the other nearly pure methanol. The carbon dioxide phase will be a low density, non-solvating gas, the methanol phase will be a solvating liquid. The polar organic solutes will remain dissolved in the small volume of the liquid phase. Thus, the task of removing the solvent from the collected fractions is greatly reduced.

Other aspects of SFC attracting considerable attention include the very low cost of the main fluid (carbon dioxide), the low disposal costs, and the relatively low toxicity of the fluids used. With large, bulk storage containers carbon dioxide can cost as little as 1/100th the cost of organic solvents.

The carbon dioxide used is recycled and considered 'green' compared to the solvents it replaces. It is generally vented into the atmosphere at the low volumes associated with analytical and semipreparative scale chromatography. At higher flow rates, it can be recycled. The modifier becomes the primary liquid waste stream. Depending on the concentration of modifier used, waste disposal tends to involve less volume and lower toxicity (methanol replaces the acetonitrile) than in HPLC.

## Collecting Fractions

Unlike in HPLC, one simply cannot divert the effluent from the detector into different test tubes or bottles. The fluid volume expands >500 times downstream of the backpressure regulator. The most typical result is the formation of large amounts of aerosol containing the solute molecules. Besides losing up to 70% of the solute, such aerosols can be extremely dangerous. Several devices exist which allow near 100% recovery without exposing anyone to an aerosol.

**Cyclone separator** In the past, the primary device used to collect solute was a cyclone separator. When

a peak was detected, a valve diverted the effluent into a specific cyclone separator. The pressure of the fluid emerging from the detector was reduced until two phases formed. The fluid was then impinged onto the inside wall of a relatively large closed cylinder at an oblique angle. Any liquid (modifier or solute) tends to run down the wall while any gas phase tends to swirl around cyclone-like and exit out the top of the cylinder. The cylinder could be replaced with a cone tapering down to a small hole at the bottom, or consist of a number of connected cylinders, each smaller in diameter than the previous one.

The pressure in a cyclone separator is only slightly lower than the pressure at the end of the column (i.e., 60 or 70 bar versus 100 bar) and the density of the gaseous phase can be relatively high, making it straightforward to recycle the fluid. The use of binary mobile phases make this somewhat more difficult but not impossible.

Cyclone separators work very well in situations where large amounts of the same material need to be purified. For situations where only small amounts of many different solutes need to be purified, they tend to be big and bulky, and require significant time and solvent to clean.

**Berger separator** An alternate phase separator is commercially available that allows small amounts (up to 100 g) to be purified in an automated fashion. It is actually often used to collect fractions when each injection involves a different sample. This approach only functions when binary fluids are used. The pressure is dropped in multiple stages, which initially creates two slow moving phases in intimate contact. This slow movement allows time to add heat. This heat is carefully added to decrease the solubility of carbon dioxide in the modifier. The result is two distinct fluids in equilibrium that are similar in composition to the final fluids at atmospheric pressure. The entire process takes place inside small ID tubing and is self-cleaning. After the fluids are separated, a valve

diverts the two fluids into a collection vessel where the gas phase exits out of the top and the liquid phase drips to the bottom. While it superficially resembles a cyclone separator, the mechanism of phase separation is more complex.

*See also:* **Supercritical Fluid Chromatography: Applications.**

## Further Reading

- Anton K and Berger C (1999) *Supercritical fluid chromatography with packed columns, techniques and applications*. New York: Dekker.
- Berger TA (1995) 'Packed Column SFC', *RSC Chromatography Monograph Series*. London: Royal Society of Chemistry.
- Caude M and Thiebaut D (eds.) (1999) *Practical Supercritical Fluid Chromatography and Extraction*. Amsterdam: Harwood Academic Publishers.
- King JW and List GR (eds.) (1996) *Supercritical Fluid Technology in Oil and Lipid Chemistry*. Champaign, IL: AOCS Press.
- Markides KE and Lee ML (eds.) (1989) *SFC Applications: 1989 Symposium/Workshop on Supercritical Fluid Chromatography*: Brigham Young University Press.
- Smith RM (ed.) (1988) *Supercritical Fluid Chromatography*. RSC Monograph Series. London: Royal Society of Chemistry.
- Berger TA, Fogelman K, Staats T, *et al.* 'Development of a Semi-Preparatory Scale SFC for High-Throughput Purification of 'Combi-Chem' Libraries', *J. Biochem. Biophys. Methods*, vol. 43, No.'s 1-3, E. Heftmann, Special Editor, Elsevier, New York 2000.
- Wilson WH and Berger TA, 'High-speed screening of combinatorial libraries by gradient packed-column supercritical fluid chromatography, *J. Biochem. and Biophys. Methods*', Jan. 2000. *J. Biochem. Biophys. Methods*, vol. 43, No.'s 1-3, E. Heftmann, Special Editor, Elsevier, New York 2000.
- York P, Kompella UB, and Shekunov BY (eds.) (2004) *Supercritical Fluid Technology for Drug Product Development*. Drugs and the Pharmaceutical Sciences Series. vol. 138, NY: Dekker.

## Applications

**T A Berger**, Berger Instruments, Newark, DE, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Supercritical fluid chromatography (SFC) has been practiced for more than 40 years and has produced a

remarkable array of applications. Two variants of SFC have emerged. Open tubular SFC almost invariably uses pure fluids, like carbon dioxide, pressure programmed at elevated temperatures and tends to act as an extension of gas chromatography (GC) to heavier, and less volatile materials. Solvation energy is used to partially replace thermal energy to allow such separations at lower temperatures than in

a peak was detected, a valve diverted the effluent into a specific cyclone separator. The pressure of the fluid emerging from the detector was reduced until two phases formed. The fluid was then impinged onto the inside wall of a relatively large closed cylinder at an oblique angle. Any liquid (modifier or solute) tends to run down the wall while any gas phase tends to swirl around cyclone-like and exit out the top of the cylinder. The cylinder could be replaced with a cone tapering down to a small hole at the bottom, or consist of a number of connected cylinders, each smaller in diameter than the previous one.

The pressure in a cyclone separator is only slightly lower than the pressure at the end of the column (i.e., 60 or 70 bar versus 100 bar) and the density of the gaseous phase can be relatively high, making it straightforward to recycle the fluid. The use of binary mobile phases make this somewhat more difficult but not impossible.

Cyclone separators work very well in situations where large amounts of the same material need to be purified. For situations where only small amounts of many different solutes need to be purified, they tend to be big and bulky, and require significant time and solvent to clean.

**Berger separator** An alternate phase separator is commercially available that allows small amounts (up to 100 g) to be purified in an automated fashion. It is actually often used to collect fractions when each injection involves a different sample. This approach only functions when binary fluids are used. The pressure is dropped in multiple stages, which initially creates two slow moving phases in intimate contact. This slow movement allows time to add heat. This heat is carefully added to decrease the solubility of carbon dioxide in the modifier. The result is two distinct fluids in equilibrium that are similar in composition to the final fluids at atmospheric pressure. The entire process takes place inside small ID tubing and is self-cleaning. After the fluids are separated, a valve

diverts the two fluids into a collection vessel where the gas phase exits out of the top and the liquid phase drips to the bottom. While it superficially resembles a cyclone separator, the mechanism of phase separation is more complex.

*See also:* **Supercritical Fluid Chromatography: Applications.**

## Further Reading

- Anton K and Berger C (1999) *Supercritical fluid chromatography with packed columns, techniques and applications*. New York: Dekker.
- Berger TA (1995) 'Packed Column SFC', *RSC Chromatography Monograph Series*. London: Royal Society of Chemistry.
- Caude M and Thiebaut D (eds.) (1999) *Practical Supercritical Fluid Chromatography and Extraction*. Amsterdam: Harwood Academic Publishers.
- King JW and List GR (eds.) (1996) *Supercritical Fluid Technology in Oil and Lipid Chemistry*. Champaign, IL: AOCS Press.
- Markides KE and Lee ML (eds.) (1989) *SFC Applications: 1989 Symposium/Workshop on Supercritical Fluid Chromatography*: Brigham Young University Press.
- Smith RM (ed.) (1988) *Supercritical Fluid Chromatography*. RSC Monograph Series. London: Royal Society of Chemistry.
- Berger TA, Fogelman K, Staats T, *et al.* 'Development of a Semi-Preparatory Scale SFC for High-Throughput Purification of 'Combi-Chem' Libraries', *J. Biochem. Biophys. Methods*, vol. 43, No.'s 1-3, E. Heftmann, Special Editor, Elsevier, New York 2000.
- Wilson WH and Berger TA, 'High-speed screening of combinatorial libraries by gradient packed-column supercritical fluid chromatography, *J. Biochem. and Biophys. Methods*', Jan. 2000. *J. Biochem. Biophys. Methods*, vol. 43, No.'s 1-3, E. Heftmann, Special Editor, Elsevier, New York 2000.
- York P, Kompella UB, and Shekunov BY (eds.) (2004) *Supercritical Fluid Technology for Drug Product Development*. Drugs and the Pharmaceutical Sciences Series. vol. 138, NY: Dekker.

## Applications

**T A Berger**, Berger Instruments, Newark, DE, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Supercritical fluid chromatography (SFC) has been practiced for more than 40 years and has produced a

remarkable array of applications. Two variants of SFC have emerged. Open tubular SFC almost invariably uses pure fluids, like carbon dioxide, pressure programmed at elevated temperatures and tends to act as an extension of gas chromatography (GC) to heavier, and less volatile materials. Solvation energy is used to partially replace thermal energy to allow such separations at lower temperatures than in

GC. Today, there are no commercial sources for instruments designed for open tubular columns, although packed column and micropacked column instruments can be adapted for this use.

Packed column SFC tends to use relatively short columns with polar stationary phases, and modified mobile phases at modest temperatures. While packed columns can be used with pure carbon dioxide for many of the same sample types as open tubular columns, the greatest advantages of the technique are in the analysis of small polar molecules using multi-component mobile phases, similar to high-performance liquid chromatography (HPLC). Packed column SFC is a replacement technique for HPLC. Small changes in temperature tend to modify selectivity. Pressure tends to be a secondary control variable, modestly changing retention but having little effect on selectivity.

The best reasons for using packed column SFC stem directly from a few physical and chemical characteristics. From the van Deemter equation it is obvious that the optimum linear velocity of the mobile phase in any chromatographic technique is related to the diffusion coefficient of the solutes in the mobile phase. The diffusion coefficients of solutes in carbon dioxide are up to 10 times higher than in normal liquids, making SFC up to 10 times faster than HPLC. When a modifier is added, this speed advantage is degraded more or less in proportion to the modifier content. Fortunately, solvent strength is a highly nonlinear function of modifier concentration. The first small additions of modifier cause significant increases in polarity.

The viscosity of these fluids is also as little as 1/10th normal liquids, making it possible to make longer columns, pump at higher velocities, or use smaller particles. The fluids are typically compressed gases. When the pressure is dropped, they turn back into a gas, very important in preparative chromatography, since fractions collected will be much smaller (only the modifier plus solute is collected). Finally, SFC using carbon dioxide is considered a 'green' technology since it is recycled and replaces far nastier solvents.

## Historical Development of Application Types

The solvent characteristics of supercritical fluids were recognized more than 100 years ago. Ernst Klesper, a German polymer chemist, published the first paper on SFC in 1961 separating organometallics, after Jim Lovelock suggested the possibility of SFC as an inorganic separation technique,

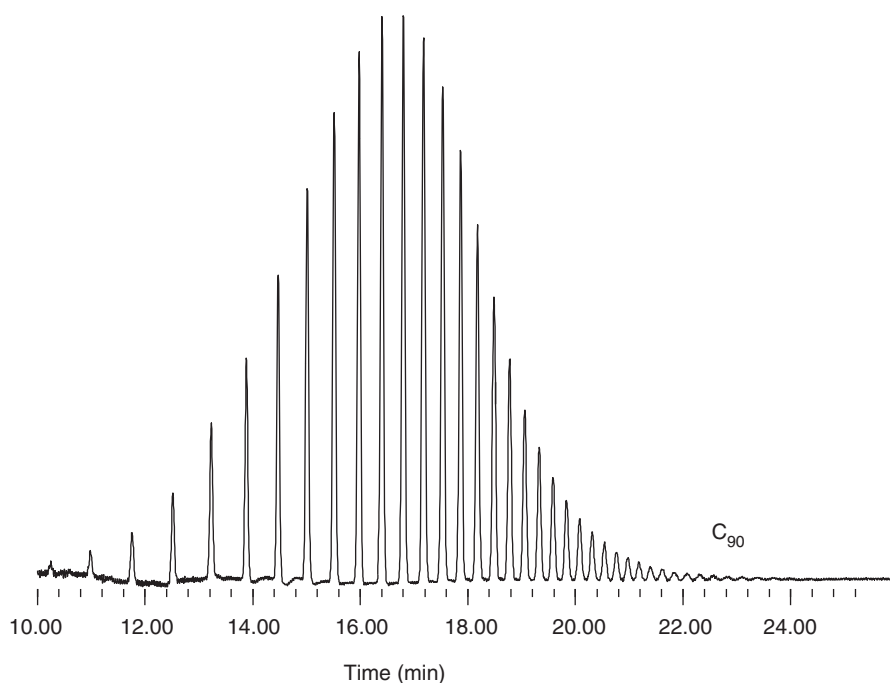
at an international symposium on GC. Klesper subsequently published a series of papers on polymer analysis using packed columns. The mobile phases he used were often supercritical pentane ( $T > 280^\circ\text{C}$ ) or pentane mixed with a more polar solvent like isopropanol. As the technique evolved, most workers switched away from using such highly flammable hydrocarbons requiring elevated temperatures, toward the use of carbon dioxide or, less commonly, a fluorocarbon refrigerant, or one of a small number of other viable fluids such as nitrous oxide.

Most of the work during the 1980s involved open tubular (capillary) column SFC using pressure programming of pure fluids for the separation of low to moderate polarity, often reactive, homologous series, including heavy hydrocarbons, waxes, natural oils, silicone oils, nonionic surfactants, mono-, di-, and triglycerides, phospholipids, derivatized polysaccharides, isocyanates, and light polymers such as polystyrenes and methacrylates, and many others. The separation of a hydrocarbon boiling point standard is shown in **Figure 1**. Typical run time was an hour.

During this time there was a controversy between open tubular and packed column SFC. There was a widespread belief that carbon dioxide was much more polar than it actually is. Carbon dioxide is roughly as polar as pentane but probably in a different solvent family. Initially, efforts to extend SFC to more polar solutes were limited by the (largely unrecognized) low polarity of carbon dioxide, since there was also a reluctance to abandon the inherent instrumental simplicity of pressure programming a pure fluid.

It was demonstrated that many homologous series could be eluted from an open tubular column, which apparently could not be eluted from the packed column with pure carbon dioxide. The inability to elute polar compounds from packed columns was attributed to excess column activity, not to the low polarity of the mobile phase. Today, we recognize that the open tubular column has perhaps 100 times higher void volume per unit surface area than a typical packed column. This difference in phase ratio makes inherent retention 1/100 as great on the open tubular column, compared to a totally porous column packing with small pores.

The development of binary/ternary mobile phases was foreshadowed by the first of many patents (in 1971) for decaffeination of tea (and later coffee). Caffeine is quite polar with an octanol/water partition coefficient ( $P$ ) more polar than 95% of commercial pharmaceuticals ( $\log P \sim -1$ ). Extraction with pure carbon dioxide was ineffective, but the



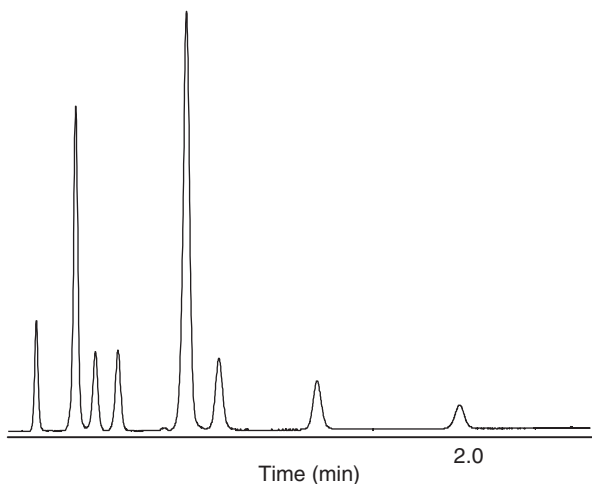
**Figure 1** Capillary SFC separation of a simulated distillation standard (Polywax 500) on a  $50\ \mu\text{m} \times 10\ \text{m}$  methylsilicone column; 80–400 bar at  $25\ \text{bar min}^{-1}$  after 5 min initial hold;  $150^\circ\text{C}$ .

extraction yield dramatically improved when water was added.

During the 1980s, work on packed columns was largely ignored but a steady progression of publications showed that mixing more polar solvents with carbon dioxide allowed the elution of much more polar solutes from packed columns than could be eluted from open tubular columns with pure carbon dioxide. Today, almost all packed column SFC applications use binary/ternary mobile phases. Methanol has become the modifier of choice, since it is among the most polar solvents completely miscible with carbon dioxide, is inexpensive, widely available, and of relatively low toxicity. Most organic solvents can be used as the modifier, but most yield less polar mixtures than carbon dioxide/methanol. The polarity range covered by carbon dioxide/methanol mixtures is wider than any binary pair available in HPLC.

For solutes that are strong acids and bases a third component is often required in the mobile phase. In general, acidic solutes require an acidic additive (like trifluoroacetic acid). Basic solutes require a basic additive (like dimethylethylamine). These additives are mixed with the modifier at 0.1–0.5%. Polar solutes are separated on polar stationary phases (silica, amino, diol, cyano, ethyl pyridine, etc.)

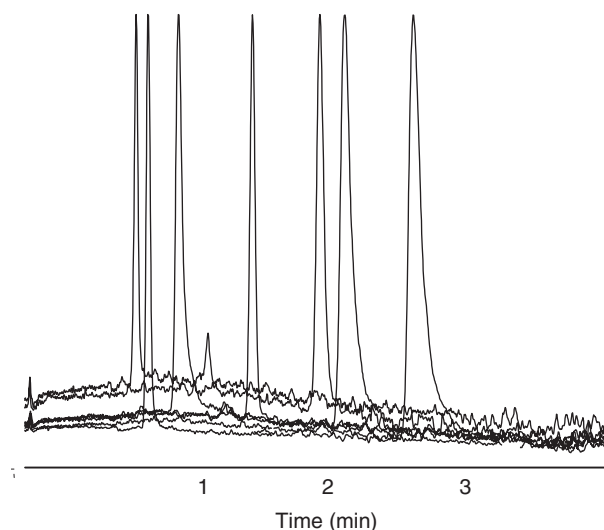
It was only in the mid- to late-1980s that it became clear that the speed advantages of SFC over HPLC were best realized using packed columns. Very rapid SFC separations were possible on standard HPLC



**Figure 2** Rapid SFC separation of steroids. Conditions:  $2.0 \times 250\ \text{mm}$ ,  $5\ \mu\text{m}$  Lichrospher silica,  $2.5\ \text{ml min}^{-1}$  of 20% methanol in carbon dioxide at  $70^\circ\text{C}$ , 200 bar. Solutes: progesterone, methyltestosterone, testosterone, estrone, estradiol, cortisone, hydrocortisone, estriol.

columns. For example, the 1989 SFC separation of hydroxysteroids was performed at 10 times the optimum flow rate in HPLC ( $2.5\ \text{ml min}^{-1}$  in SFC versus  $0.25\ \text{ml min}^{-1}$  in HPLC) without apparent degradation of column efficiency, as shown in Figure 2.

A rule of thumb states that any compound soluble in methanol or a less polar solvent is an ideal candidate for SFC. The corollary is that any solute requiring an aqueous medium or a buffered aqueous



**Figure 3** Relative elution of nucleic acids and nucleosides. Conditions: 5 ml min<sup>-1</sup> of 35–80% at 10% min<sup>-1</sup> of (methanol/10% water/0.4% ammonium acetate/0.4% formic acid) in carbon dioxide at 50°C, 200 bar, 4.6 × 200 mm, 5 μm Hypersil Amino. Solutes: uracil, adenine, adenosine cytidine, guanine, cytosine, guanosine dissolved in methanol/formic acid.

medium to dissolve is a poor candidate. Actually, the border between HPLC and SFC remains ill-defined. For example, nucleic acids, particularly guanine, are not soluble in methanol, and actually are not soluble in pure water. Nucleosides are even more problematic but both the acids and nucleosides can be readily separated by SFC with a carbon dioxide/methanol/water/ammonium acetate/formic acid mobile phase, as shown in Figure 3.

Today, most applications are based on packed columns at relatively low temperatures with binary and ternary mobile phases. It has become clear that the characteristics of interest are not limited to ‘supercritical’ conditions. Most packed column applications actually use conditions just below the critical temperature, but well above the critical pressure of the mixed fluids, i.e., subcritical (in other words, ‘liquid’). Unlike many observers’ expectations, important characteristics such as diffusion coefficients and viscosity are virtually identical whether the conditions are just supercritical or just subcritical.

There are two primary application areas of packed column SFC: pharmaceuticals and, to a lesser extent, petroleum. Several other areas such as pesticide analysis and chemicals are practiced but not on a widespread scale. A noncomprehensive list of applications is presented in Table 1.

## Pharmaceuticals

The biggest advances in SFC use have occurred in industrial pharmaceutical labs. Small drug-like

**Table 1** Selected applications of SFC

<i>Pharmaceuticals</i>	<i>Natural products</i>
Chiral separation	Alkaloids glycosides
Rapid method development	Amino acids/nucleic acids
Multiple CSP's in series	Bile acids
Chiral detectors	Sugars-polysaccharides
Enantiomeric excess	Cholesterol
Chiral purification	Steroid hormones
Combi-chem screening	Ecdysteroids
Universal gradients	Digitoxins
Speed vs. purity	Taxol
SFC-MS	Wood resin/rosins
High throughput purification	Fatty acids, prostaglandins
Traditional analytical QA/QC	Mono-, di-, and tri-glycerols/
Quantification	glycerides
Purity	Phospholipids
Excipients	Ubiquinones
Fermentation extracts	Terpenes
ADME	Lactones
Metabolites in biological	Favonoids
SFC-MS of metabolites	Carotenoids
Antidepressants	Fat and water soluble vitamins
Antipsychotics	Essential oils
Stimulants	Waxes
<i>Herbicides and pesticides</i>	<i>Chemicals</i>
Sulfonylurea pesticides	Surfactants
Phenylurea pesticides	Isocyanates
Carbamates	Silicone oils
Organochlorine	Polymers
Organophosphorus	Polymer additives
Triazine herbicides	Epoxy resins
Pyrethrins/pyrethroids	<i>Petroleum</i>
<i>Environmental</i>	Aromatics in diesel
Phenols	Olefins in gasoline
Phthalates	Simulated distillation
PAHs	
PCBs	

molecules are ideal candidates for SFC. Most commercial drugs are neither too ‘polar’ nor too ‘lipophylic’. The former would be eliminated from the body too rapidly through the urinary tract. The latter are trapped in body fat and can actually become toxic. Pharmaceuticals taken orally have an average log *P* of 2.5–3. Compounds with log *P* < 0 (such as most proteins) tend to be administered by injection, which is considered undesirable. Those with log *P* > 5 tend to be toxic. Compounds shown to have been eluted by SFC have log *P* values between –1 and 7. Thus, SFC appears to be the perfect match for small drug-like molecules intended for oral administration. There is an extensive literature, including thousands of papers, on the SFC separation of the most common orally administered pharmaceuticals.



## Chiral Separations

Fundamentally, packed column SFC is a replacement technique for normal phase HPLC (n-HPLC). The most obvious application demonstrating the superiority of SFC is in chiral analysis and purification. Chemically, chiral molecules are identical mirror images of each other. They can only be separated by differences in shape.

SFC uses the same stationary phases as HPLC. Selectivity is similar but not identical. One of the greatest differences from n-HPLC is in speed. SFC optimum flow is inherently three to five times faster, while peak shapes are often significantly better. Unlike n-HPLC, SFC reequilibrates after passage of only a few column volumes. Overall, SFC is often much more than 10 times faster than HPLC. A fast chiral SFC separation is shown in **Figure 4**. Note that the column is not 'high speed' but a standard  $4.6 \times 250$  mm column with  $10\text{ }\mu\text{m}$  particles. Some industrial pharmaceutical companies have dropped HPLC for chiral analysis and use SFC for less polar solutes, and high-performance capillary electrophoresis (HPCE) for water-soluble solutes. Unlike HPCE, SFC is scalable.

The mobile phase used in SFC costs as little as 1/20th HPLC solvents for the same work. In semipreparative applications, the absolute cost savings can be dramatic. In SFC, a 20–21.2 mm ID column can cost as little as  $\$0.46\text{ h}^{-1}$  (more typically  $\$3\text{ h}^{-1}$ ) while the solvent cost on the same column in HPLC can be as much as  $\$27\text{ h}^{-1}$ . Since the SFC performs three to 10 times more work in the same time, the cost per unit of work is dramatically better. Further, sample dry down is substantially easier since the carbon dioxide flashes away when the pressure drops, leaving behind the fractions in much smaller volumes of the modifier. Finally, waste generation is also reduced. Several organizations have claimed a single semipreparative SFC running 21.2 mm ID

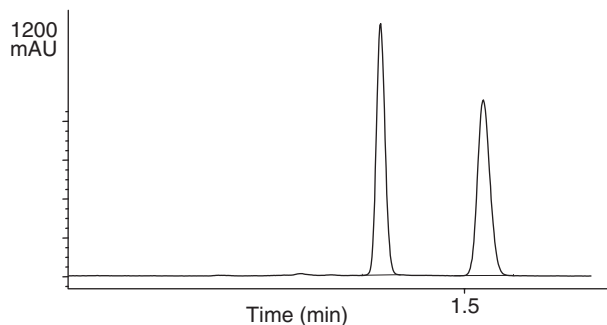
columns overall saves up to  $\$1\,000\,000\text{ year}^{-1}$  associated with solvent purchase, dry down, and disposal.

## Supporting Combi-Chem

For the last 5–10 years, combinatorial chemistry has been the method of choice for developing huge numbers of small drug-like compounds that can be tested for biological activity by high throughput screening (HTS). Small groups are often presented with the task of synthesizing, analyzing, purifying, and registering 40 000 compounds per year or more. The consensus opinion has shifted over time to indicate that relatively pure compounds should be tested by HTS to avoid false positives and negatives. Although synthesis can be optimized to create relatively pure compounds, it is more typical to produce larger numbers of relatively dirty compounds, requiring both high throughput analysis and purification.

Most practitioners perform a version of triage on the samples. The intent is to eliminate from consideration all the samples where the desired compound is not found, where there is not enough substance to purify, or where there are large numbers of potentially interfering contaminant peaks present. HPLC with mass spectrometric detection (HPLC–MS) is used for analysis, followed by semipreparative HPLC for purification. The mass spectrometer verifies that the compound of interest is present. Other detectors are used to attempt to quantitate how much is present. After the analytical step, the samples that have been approved are purified. In some cases, the peaks are collected based on triggering by another mass spectrometer on the preparative instrument. In other cases, the retention time of the peak of interest in the analytical run is translated or scaled to the preparative unit, and the peak collection occurs in time windows or by UV triggering. HPLC tends to use large amounts of toxic organic solvents like acetonitrile, which becomes toxic waste. Fractions are water/acetonitrile mixtures, which require extensive lyophilization to remove from the solutes.

Since SFC is three to five times faster, for the same work output, SFC is becoming an attractive alternative to HPLC for both analysis and purification. SFC is 'normal phase', and the elution order is generally opposite to reversed-phase HPLC (RPLC). Thus, SFC purification makes most sense following SFC analysis. The same general approach is followed, using SFC–MS, to identify which samples should be purified by SFC. The tendency has been to use SFC to achieve higher resolution in the same time compared to HPLC, rather than increase the number of samples per unit time. A single preparative



**Figure 4** Chiral separation of *t*-stilbene oxide, on a  $4.6 \times 250$  mm,  $10\text{ }\mu\text{m}$  Chirapack AD column using  $5\text{ ml min}^{-1}$  of 40% methanol in carbon dioxide. Outlet pressure was 100 bar, temperature was  $35^\circ\text{C}$ .

instrument is theoretically capable of more than 100 000 purifications of up to 100 mg each per year.

As in chiral purification, the mobile phase is dramatically cheaper than in HPLC. Less toxic waste is generated, and much less power is required to dry down samples. One of the biggest advantages of SFC is actually in the postpurification sample handling. The fractions tend to be much smaller and are 100% organic, allowing much faster, much easier, much cheaper dry-down.

### Natural Products

Packed column SFC has been used extensively to analyze natural products. Fat-soluble compounds like squalene, vitamins A, D, K and E, carotenoids, ubiquinones, cholesterol, mono-, di-, and triglycerides, and fatty acids have all been extensively studied. Hydroxy fatty acids, prostaglandins, and phospholipids have also been separated. Omega-3-fatty acids are separated out of fish oil on a commercial scale by packed column SFC.

Polar derivatives of cholesterol, such as keto and hydroxy steroids, bile acids, and ecdysteroids have all been extensively studied.

Nonvolatile terpenes, including alcohols and acids, wood resins, lactones, bittering acids from hops, etc., have been analyzed. Many alkaloids, and glycosides, toxins, and stimulants have been extracted and chromatographed. Compounds like caffeine, theophylline, and theobromine (see Figure 5), taxol, digitalis glycosides, reserpine, ginkgosides, garlic and onion extracts, and 'loco weed' extracts. Sugars,

polysaccharides, and glycosides have been eluted with an evaporative light scattering detector (ELSD), or a mass spectrometer.

Both native and derivatized amino acids are readily eluted. Most low polarity peptides containing up to perhaps 30 amino acids and small cyclic peptides are easy to elute. A carbon dioxide/methanol/water/formic acid mobile phase can also elute many 'water-soluble' vitamins.

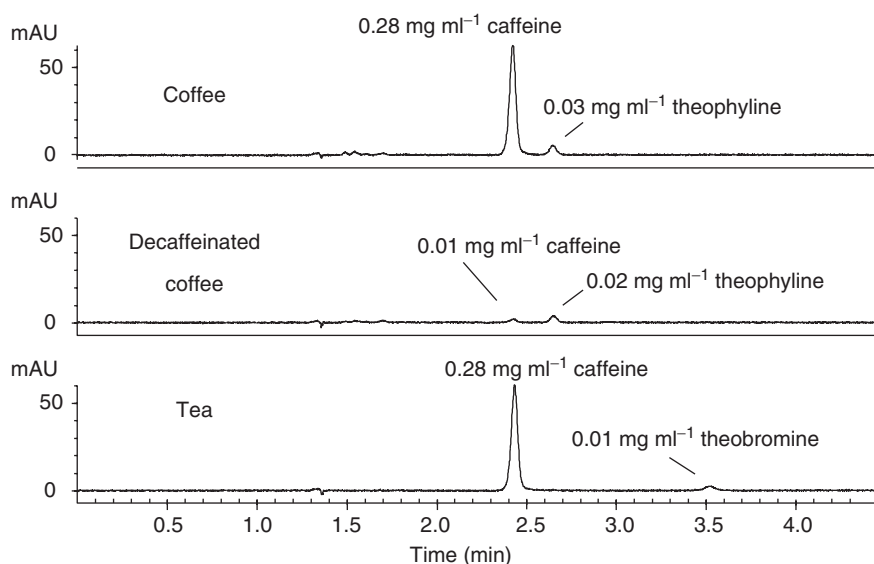
### Petroleum

#### Aromatics in Diesel

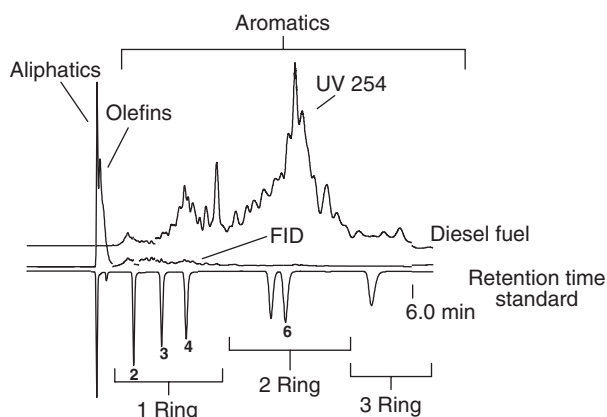
There are two American Society of Testing Materials (ASTM) methods based on SFC that have major implications in environmental health.

Most industrial countries have set as a goal a major reduction in the aromatic content in diesel fuel to cut down on particulates emissions. Traditional methods such as fluorescence indicator analysis are slow and require significant solvent but more importantly have difficulty quantitating low levels of aromatics.

The aromatic content in diesel fuel (ASTM Method #5186) is a group separation method where a bare silica column elutes all aliphatic hydrocarbons as a single peak, olefins as a second nearly resolved single peak, and a series of largely resolved peaks each representing a set of ring compounds, as shown in Figure 6. The first such peak contains compounds with a single ring, the next peak contains compounds with two rings, etc. All compounds respond in a



**Figure 5** SFC analysis of caffeine analogs: 0.5  $\mu$ l neat injection onto a 4.6  $\times$  250 mm, 5  $\mu$ m Lichrospher silica column with 2.5 ml min<sup>-1</sup>, of 8% methanol in carbon dioxide at 30°C, 250 bar.



**Figure 6** Aromatic content of diesel fuel with indication of ring number. Relative area under the FID trace is used to quantitate the different groups.  $4.6 \times 250$  mm,  $5 \mu\text{m}$  Lichrospher silica,  $2.5 \text{ ml min}^{-1}$  pure carbon dioxide,  $27^\circ\text{C}$ , 140 bar.

flame ionization detector (FID), which is used for quantitation. A UV detector is also sometimes used.

The method requires a resolution of four between  $\text{C}_{16}$  (or  $\text{C}_{22}$ ) and toluene. In reality it is relatively trivial to achieve resolutions up to 14 on some silicas. However, other silicas do not separate this standard at all.

### Olefins in Gasoline

Recently, the group separation method has been extended to further resolve the olefins from the aliphatic hydrocarbons, particularly in gasoline. A silver column is placed in series with the silica column and is switched in and out of line. Olefins are strongly retained on the silver column while the aliphatics are not. After the olefins are trapped on-column, the flow is switched, so the aromatics are diverted directly into the FID. Later, flow is switched again and the olefins are back-flushed off the column into the detector.

### High-Temperature Simulated Distillation

A still may take 15–24 h to adequately characterize a crude oil sample by manually or automatically recording the weight loss of the sample versus temperature over time. Stills have been ‘simulated’ using a temperature programmed GC method that reaches  $450^\circ\text{C}$  and only takes 70 min.

There are several problems with the GC method. A fraction ( $\sim 0$ –5%) of some crude oils is not volatile and is deposited on the head of the column, changing the column characteristics. Depending on the concentration of this nonvolatile material, the user may need to cut 10–20 cm off the front of the column after 1–10 injections. It is generally accepted that this

contamination and manual manipulation results in an average column lifetime of no more than 100 injections. In the extreme, a single injection can ‘kill’ a column. Further, there is some controversy over the consequences of using  $450^\circ\text{C}$  as the injection port and column temperatures. One camp suggests that some compounds will break down into smaller pieces when they hit these high temperatures, distorting the mass distribution (which affects the value of the oil). Others dismiss such concerns as overly conservative.

In SFC, a pressure program from 70 to 400 bar, at a temperature of  $150^\circ\text{C}$  provides an alternative to the GC method. The sample, injection valve, and transfer line must all be heated. A methyl packed column performs with selectivity similar to a methyl silicone GC column. Components with boiling points greater than  $\text{C}_{130}$  have been eluted.

The solvating power of the dense carbon dioxide replaces most of the thermal energy of the high temperature in the GC method. The sample (including the nonvolatile part) is soluble in carbon dioxide. There is a simple test that corrects for the nonvolatile component. Since the nonvolatile components do not accumulate, a single packed column has been used for as many as 4000 analyses without serious degradation, dramatically improving column life. The data have been shown to be equivalent to that obtained by the GC method.

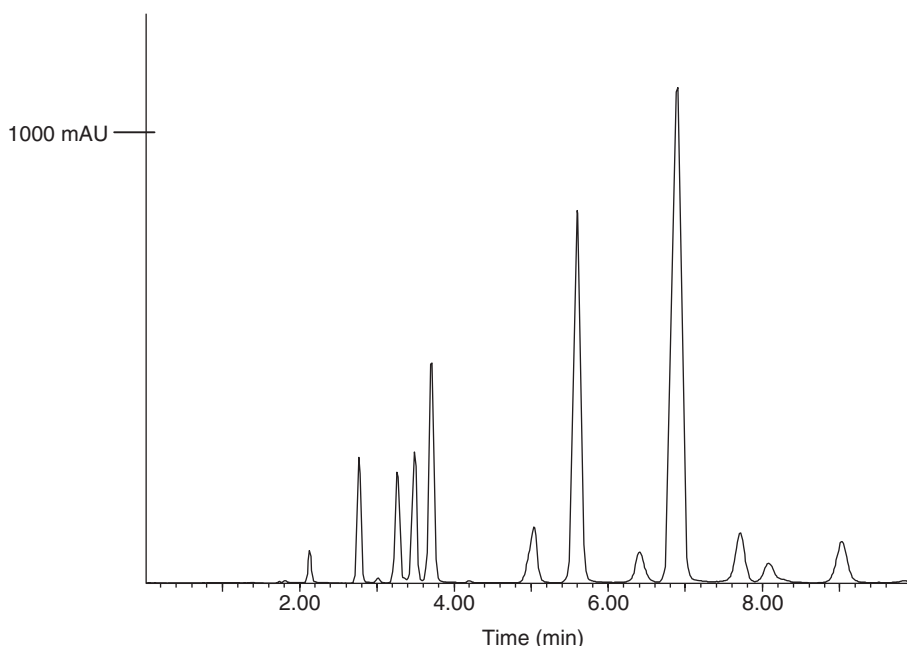
## Other Applications

### Agricultural Chemicals/Environmental Analysis

Packed column SFC has been shown to be widely applicable in the analysis of agricultural chemicals and in environmental analysis but has not achieved widespread use.

SFC can rapidly separate many of the thermally labile pesticide families such as sulfonylurea, phenylurea, and carbamate pesticides in addition to organochlorine, organophosphorus, and triazines. All these tolerate modified fluids and polar columns operated at modest temperatures. The carbamates are an example of thermally labile pesticides usually separated by HPLC. A rapid isocratic separation is presented in Figure 7.

Pesticides and herbicides have been trapped out of water and detected with various GC-like detectors such as the nitrogen phosphorus detector and the electron capture detector as well as UV photodiode array spectra. SFC is fully compatible with mass spectrometry and many other GC and HPLC detectors.



**Figure 7** Separation of carbamate pesticides on a  $4.6 \times 250$  mm,  $5 \mu\text{mol l}^{-1}$  Lichrospher Diol column with  $2.5 \text{ ml min}^{-1}$  of 5% methanol in carbon dioxide at  $30^\circ\text{C}$ , 200 bar;  $2 \mu\text{l}$  of  $2 \text{ mg ml}^{-1}$  each.

Polycyclic aromatic hydrocarbons, polychlorinated biphenyls, phenols, and phthalates have all been separated rapidly and effectively.

### Polymers and Homologous Series

**Silicone oils** Polysiloxanes (silicone oils) are widely used in industrial products. They consist of homologous series. The mass distribution in a sample dictates its physical properties, such as viscosity. The simplest member, dimethylpolysiloxane has a repeating unit of  $\text{H}_3\text{C}-\text{SiO}-\text{CH}_3$ , with a molecular weight of 74. There are now packed column separations (at  $150^\circ\text{C}$ ) that show individual maxima for more than 160 such units, suggesting a maximum molecular weight approaching  $\sim 15\,000$ . Heavier components can be eluted but the chemical differences between consecutive homologs are too small to result in individual maxima. There are often subseries of peaks offset from the main series that give some insight into the reaction mechanism or pathways.

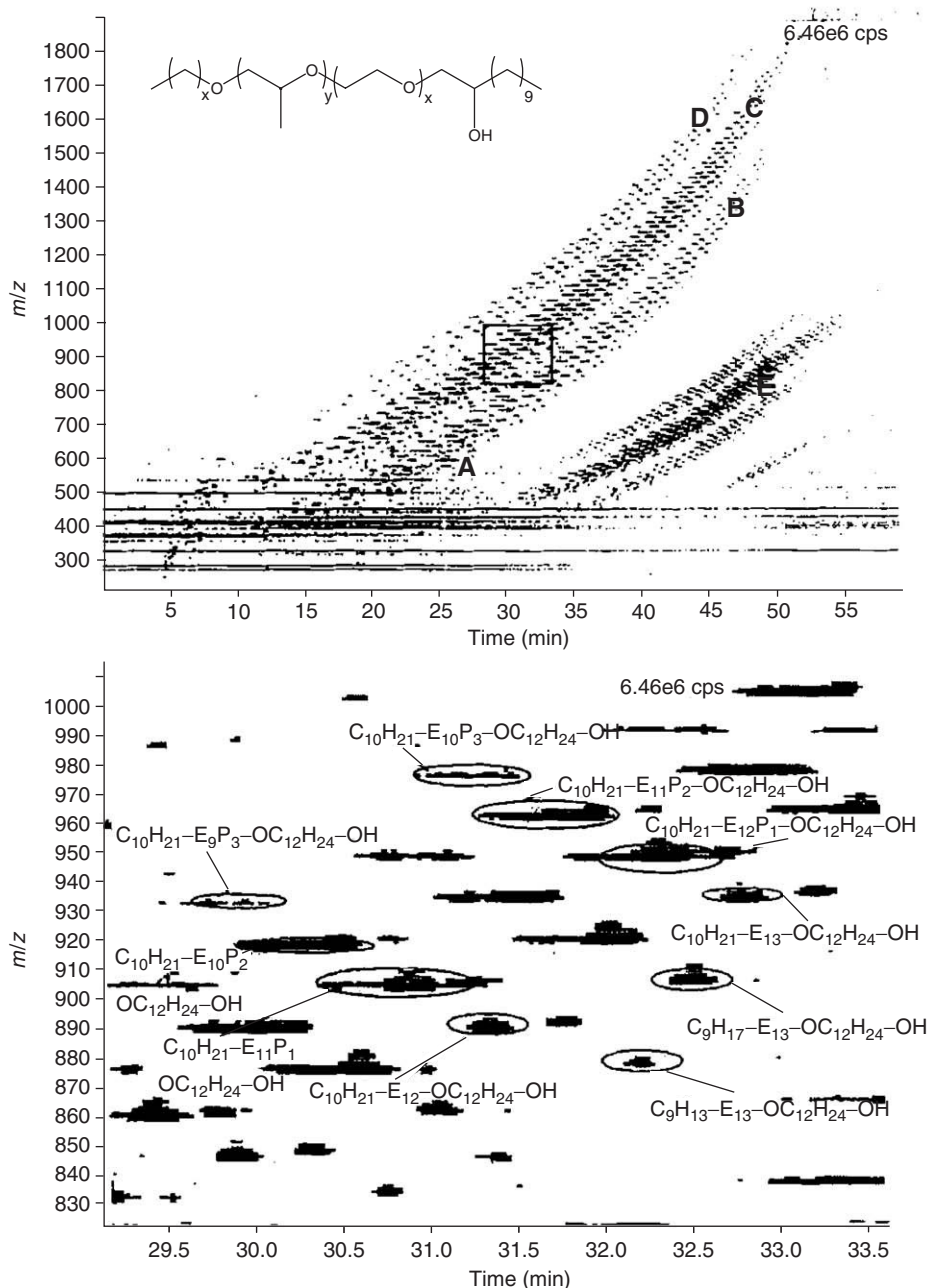
Although carbon dioxide is often described as similar in polarity to pentane or hexane, the addition of 10% hexane to carbon dioxide significantly increases the maximum molecular weight to between 25 000 and 40 000 that can be eluted. This should not be surprising. Some of Klesper's early work used pentane and modified pentane at  $260^\circ\text{C}$  to elute certain polymers with molecular weights up to several million.

All older chromatograms of silicone oil were collected using an open tubular column with an FID as detector. More recently, it has been shown that carbon dioxide is completely transparent to below 190 nm in the UV. All the siloxanes absorb light below 195 nm. Since packed columns allow larger, more reproducible injection volumes, and UV detectors do not need fixed restrictors, the method of choice appears to be packed columns with UV detection at 190 nm.

**Nonionic surfactants** Ethoxylates and propoxylates are widely used in liquid detergents. They contain a fatty acid tail and an ethyl-oxy or propyl-oxy repeating unit and are often terminated with a hydroxyl group. Each repeating unit produces a major set of peaks but for each repeating unit there is a subset of peaks representing the various chain length fatty acids.

These compounds are modestly polar but can be eluted with pure carbon dioxide from either an open tubular or packed column. Peak shapes are modestly improved by derivatizing the terminal hydroxyl group. As with the silicone oils, these compounds absorb UV at 190 nm.

Pinkston published several of the most incredible chromatograms ever (see Further reading). He made a long packed column to maximize chromatographic resolution then combined it with a mass spectrometer. The resulting two-dimensional contour plots reveal thousands of isomers in some of these samples. An example is shown in Figure 8. Such a



**Figure 8** Top: A contour plot of a low molecular weight alkoxyated polymer with mass on the y-axis and retention time on the x-axis. Each 'dot' represents a different component in the mixture. (A) Is a single component C8E10 internal standard. (B) Is the band of related components that are singly capped EO-PO chains, lacking a C12 chain. (C) Represents dicapped chains as shown in the structure at the top. (D) Dicapped EO-PO chains with a dimmer C12-OH cap. (E) Represents doubly charged species. Column: 4.6 mm  $\times$  1 m,  $d_p = 5 \mu\text{m}$  Deltabond Cyano. Mobile phase: 3% methanol for 3 min, then  $0.5\% \text{ min}^{-1}$  to 30%. Flow:  $2 \text{ ml min}^{-1}$ ;  $100^\circ\text{C}$ ,  $10 \mu\text{l}$  injected. Bottom: Blowup of square in top chromatogram with individual components identified.

chromatogram yields minute details about the reaction mechanism.

## Future Perspectives

In the year or so during which this summary of applications was written there has been a significant

increase in the number of SFC users although the specific applications remain the same. The general population of HPLC users is beginning to realize that SFC is actually a subset of HPLC with enhanced fluid characteristics and an environmentally more friendly fluid that offers dramatic decreases in solvent cost. While the primary uses remain chiral separations in

pharmaceuticals and group separations in petroleum, the inherent superiority of the fluids is being recognized more generally and the technique is being more generally applied, particularly with pharmaceuticals. The biggest driver is at the semipreparative scale where the higher throughput and dramatic decrease in costs seems to be rewriting the cost model for chromatography as an industrial process.

*See also:* **Gas Chromatography:** Detectors. **Supercritical Fluid Chromatography:** Overview.

## Further Reading

Anton K and Berger C (eds.) 1998 *Supercritical Fluid Chromatography with Packed Columns: Techniques and Applications*. Chromatographic Science Series, vol. 75. New York: Dekker.

Berger TA (1995) *Packed Column SFC*. RSC Chromatography Monograph Series. London: Royal Society of Chemistry.

Berger TA (1997) Separation of polar solutes by packed column supercritical fluid chromatography. *Journal of Chromatography A* 785: 3–33.

Caude M and Thiebaut D (eds.) (1999) *Practical Supercritical Fluid Chromatography and Extraction*. Amsterdam: Harwood Academic Publishers.

Heftmann E (ed.) *Journal of Biochemical and Biophysical Methods* 43(1–3), Special Issue.

King JW and List GR (eds.) (1996) *Supercritical Fluid Technology in Oil and Lipid Chemistry*. Champaign, IL: AOCs Press.

Markides KE and Lee ML (eds.) (1989) *SFC Applications: Symposium/Workshop on Supercritical Fluid Chromatography*. Brigham: Young University Press.

Parcher JF and Chester TL (eds.) (2000) *Unified Chromatography*. ACS Symposium Series 748. Washington DC: American Chemical Society.

# SUPERCritical FLUID EXTRACTION

*See* EXTRACTION: Supercritical Fluid Extraction

# SURFACE ANALYSIS

Contents

**Overview**

**X-Ray Photoelectron Spectroscopy**

**Particle-Induced X-Ray Emission**

**Auger Electron Spectroscopy**

**Appearance Potential Spectroscopy**

**Desorption Techniques**

**Ion Scattering**

**Nuclear Reaction Analysis and Elastic Recoil Detection**

**Secondary Ion Mass Spectrometry of Polymers**

**Secondary Neutral Mass Spectrometry**

**Laser Ionization**

**Low Energy Electron Diffraction**

**Infrared Spectroscopy**

## Overview

**R M Twyman**, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A surface can be defined as any interface between two components of a system with different compositions or properties. The simplest example is the interface between materials of two different phases, i.e., a solid and a gas, a liquid and a gas, or a



pharmaceuticals and group separations in petroleum, the inherent superiority of the fluids is being recognized more generally and the technique is being more generally applied, particularly with pharmaceuticals. The biggest driver is at the semipreparative scale where the higher throughput and dramatic decrease in costs seems to be rewriting the cost model for chromatography as an industrial process.

*See also:* **Gas Chromatography:** Detectors. **Supercritical Fluid Chromatography:** Overview.

## Further Reading

Anton K and Berger C (eds.) 1998 *Supercritical Fluid Chromatography with Packed Columns: Techniques and Applications*. Chromatographic Science Series, vol. 75. New York: Dekker.

Berger TA (1995) *Packed Column SFC*. RSC Chromatography Monograph Series. London: Royal Society of Chemistry.

Berger TA (1997) Separation of polar solutes by packed column supercritical fluid chromatography. *Journal of Chromatography A* 785: 3–33.

Caude M and Thiebaut D (eds.) (1999) *Practical Supercritical Fluid Chromatography and Extraction*. Amsterdam: Harwood Academic Publishers.

Heftmann E (ed.) *Journal of Biochemical and Biophysical Methods* 43(1–3), Special Issue.

King JW and List GR (eds.) (1996) *Supercritical Fluid Technology in Oil and Lipid Chemistry*. Champaign, IL: AOCs Press.

Markides KE and Lee ML (eds.) (1989) *SFC Applications: Symposium/Workshop on Supercritical Fluid Chromatography*. Brigham: Young University Press.

Parcher JF and Chester TL (eds.) (2000) *Unified Chromatography*. ACS Symposium Series 748. Washington DC: American Chemical Society.

# SUPERCritical FLUID EXTRACTION

*See* EXTRACTION: Supercritical Fluid Extraction

# SURFACE ANALYSIS

Contents

**Overview**

**X-Ray Photoelectron Spectroscopy**

**Particle-Induced X-Ray Emission**

**Auger Electron Spectroscopy**

**Appearance Potential Spectroscopy**

**Desorption Techniques**

**Ion Scattering**

**Nuclear Reaction Analysis and Elastic Recoil Detection**

**Secondary Ion Mass Spectrometry of Polymers**

**Secondary Neutral Mass Spectrometry**

**Laser Ionization**

**Low Energy Electron Diffraction**

**Infrared Spectroscopy**

## Overview

**R M Twyman**, University of York, York, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A surface can be defined as any interface between two components of a system with different compositions or properties. The simplest example is the interface between materials of two different phases, i.e., a solid and a gas, a liquid and a gas, or a

solid and a liquid. However, the interface between two different solids or two immiscible liquids is also a surface. In strict terms, such interfaces comprise two surfaces (e.g., at the interface between a solid and a liquid, one can refer to the surface of the solid and the surface of the liquid as separate entities). The interface between a solid or liquid and a vacuum is, in contrast, a single surface.

In purely geometric terms, a surface has no intrinsic thickness, but in practical terms a surface must comprise one or more layers of atoms or molecules. When dealing with the properties of solids and liquids, the surface and interior are usually considered separately. This is partly because, for any material of significant mass and volume, the number of atoms at the surface is negligible compared to the number of internal atoms. Mostly, however, it reflects the fact that the mathematical treatment of ordered substances is simplified if the structural periodicity is assumed to be infinite. The important point, and the reason that surface analysis is of interest at all, is that the surface differs in its properties from the bulk (internal) material even though it is made of exactly the same atoms. This is because atoms at the surface are relaxed, i.e., not all the bonds are constrained by interactions with neighboring atoms in the same material. Relaxed atoms have free bonds that are available to interact with atoms and molecules on the surface of the adjacent phase, such as the gas or liquid surrounding a crystalline solid.

The properties of surfaces have always been important to humans, from the development of the first stone tools through to the microprocessors and digital media of the present day. However, the analysis of surfaces did not become a mainstream scientific discipline until the 1960s, when it became possible to achieve ultrahigh vacuum on a routine basis. The ability to generate ultrahigh vacuum (generally defined as  $10^{-7}$ – $10^{-8}$  Pa) brought about a revolution in surface analysis for two reasons. First, it became possible to prepare atomically clean surfaces and maintain them in a contamination-free state over a typical experimental period. Under high vacuum or even very high vacuum conditions (down to  $10^{-6}$  Pa), the number of gas-phase molecules is sufficient to generate a monolayer of contaminants on the surface by molecular collision in a matter of seconds, particularly if the surface is reactive. In contrast, under ultrahigh vacuum, a full monolayer takes up to 1 h to form. Second, ultrahigh vacuum conditions allow low-energy electron and ion-based analytical methods to be used without interference from gas-phase scattering. These methods are discussed briefly below, and in more detail in the subsequent articles.

## Techniques for Surface Analysis

Almost all of the techniques currently used for surface analysis are based on interactions between the surface and incident photons, electrons, or particles (summarized in Table 1). At the simplest level, a surface can be examined visually, with either the naked eye or using a microscope. Where higher resolution is required, scanning electron microscopy can be used to show the detailed structure or morphology of a surface down to a resolution of  $\sim 2$  nm, with the added advantage of a large depth of field. This involves the use of a finely focused beam of high-energy, raster electrons (energy of tens of thousands of electron volts) that releases low-energy electrons from the surface, which can be detected and interpreted as an image. Other useful methods include scanning tunneling microscopy and atomic force microscopy, each of which can be used under normal atmospheric conditions although experiments are often carried out under vacuum so that the results can be compared with those obtained from spectroscopy. In each technique, a tip travels across the surface of the analyte and the deflection is measured generating a topological map of the surface at near atomic resolution. In the case of scanning tunneling microscopy, both the tip and the surface are conducting. The tip is not in direct contact with the surface and a tunneling current is measured between the surface and the tip. Atomic force microscopy can be used to analyze nonconducting surfaces. The tip can be constantly in contact with the surface, it can oscillate at high frequency, and tap the surface with a specific periodicity, or it can scan a few tenths of a nanometer above the surface. Additionally, the probe tip can be coated with a magnetic substance allowing the magnetic fields immediately above the surface to be measured.

The arrangement of atoms on a surface can be probed by various spectroscopy techniques. Some techniques make use of electron diffraction, where the electron energy can range from a few volts to many thousands. These methods have been the mainstay of surface analysis for many years, particularly for the detection of surface structures on single crystal faces. The area of analysis depends on the electron beam diameter and the interference length, and this usually means that the area from which the signal originates is  $\sim 1$  mm<sup>2</sup> with low-energy beams. Other methods use a beam of ions to strike the surface and displace atoms, which can be ionized and detected by mass spectrometry. Yet other methods involve bombarding the surface with electrons or X-rays, and then measuring the photons that are emitted. Depending on the technique, only the true surface

**Table 1** Overview of surface analysis techniques

<i>Technique</i>	<i>Principles and applications</i>	<i>Variations</i>
<i>Electron spectroscopy techniques</i>		
Appearance potential spectroscopy (APS)	A group of techniques in which the energy of the excitation beam (X-rays, electrons) gradually increases, allowing the onset of excitation at a core level to be detected	Auger electron appearance potential spectroscopy (AEAPS) Soft X-ray appearance potential spectroscopy (SXAPS) Disappearance potential spectroscopy (DAPS)
Auger electron spectroscopy (AES)	The sample is bombarded with an electron beam, leading to the ejection of core electrons from the analyte atoms. The resulting vacancies are filled by other electrons falling from higher levels, and their energy is transferred to so-called Auger electrons, which leave the atom with a specific kinetic energy. AES identifies elemental compositions of surfaces by measuring the kinetic energies of Auger electrons. Widely used for the analysis of surfaces and thin films	X-ray induced Auger electron spectroscopy (XAES) Auger photoelectron coincidence spectroscopy (APECS) Scanning Auger microscopy (SAM)
Electron energy loss spectroscopy (EELS)	The sample is bombarded with a monoenergetic beam of electrons. The inelastic scattering of low-energy electrons is used to measure vibrational spectra of surface species, allowing elemental characterization. Often used in concert with electron microscopy	High-resolution electron energy loss spectroscopy (HREELS) Spin polarized electron energy loss spectroscopy (SPEELS)
Electron momentum spectroscopy (EMS) (also known as (e,2e) spectroscopy)	The surface of the analyte is bombarded with an electron beam, causing incoming electrons to be scattered and resident electrons to be ejected. The scattered and ejected electrons are detected simultaneously	
Extended X-ray absorption fine structure (EXAFS)	The sample is targeted with a monochromatic X-ray beam and the photon energy is gradually increased until it exceeds the absorption edge of one of the components of the analyte. This results in a large increase in absorption and the resulting photoelectrons are backscattered by the atoms surrounding the emitting atom	Near edge X-ray absorption fine structure (NEXAFS) (also known as X-ray absorption near edge structure (XANES)) Surface extended X-ray absorption fine structure (SEXAFS)
Infrared spectroscopy	A group of techniques using infrared irradiation to study molecules vibrating on a surface. Vibrating molecules with a significant dipole moment will absorb infrared light at specific frequencies depending on the molecular species. Infrared spectra therefore show absorption peaks that are characteristic of particular molecules and the way they are bonded to the surface	Transmission infrared spectroscopy (TIRS) Diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS) Reflection–absorption infrared spectroscopy (RAIRS) (also known as infrared reflection absorption spectroscopy, IRAS or IRRAS) Multiple internal reflection spectroscopy (MIR) Attenuated total reflectance (ATR) Surface enhanced Raman scattering (SERS)
X-ray photoelectron spectroscopy (XPS) (also known as electron spectroscopy for chemical analysis, ESCA)	The sample is bombarded with X-rays, leading to the emission of photoelectrons. The kinetic energy of the excited photoelectrons depends on the energy of the incident X-ray (generally in the range 200–2000 eV) minus the binding energy of core level electrons, which are specific for each element. XPS is used for the quantitative compositional analysis of surfaces	Spin polarized X-ray photoelectron spectroscopy (SPXPS)

Table 1 Continued

<i>Technique</i>	<i>Principles and applications</i>	<i>Variations</i>
Ultraviolet photoelectron spectroscopy (UPS)	Similar in principle to XPS, but the sample is irradiated with vacuum ultraviolet radiation (10–45 eV) which is only capable of ionizing electrons from the outermost (valence) levels of atoms. Typically used to determine the electronic structure of solids and the adsorption of relatively simple molecules on metals	Angle resolved ultraviolet photoelectron spectroscopy (ARUPS) also known as angle resolved photoelectron spectroscopy (ARPES) Spin polarized ultraviolet photoelectron spectroscopy (SPUPUS)
Inverse photoemission spectroscopy (IPES) also known as Bremsstrahlung isochromat spectroscopy (BIS)	The sample is targeted with a beam of electrons and the light emission is measured, providing information on the empty density of states above the Fermi level	K-resolved inverse photoemission spectroscopy (KRIPES) Spin polarized inverse photoemission spectroscopy (SPIPES)
<i>Electron diffraction techniques</i>		
X-ray photoelectron diffraction (XPD)	Similar in principle to XPS, but involving the analysis of crystal surfaces. Under these circumstances, emitted photoelectrons may experience diffraction effects, which help to determine the crystal structure in the vicinity of the emitting atoms	
Auger electron diffraction (AED)	Similar in principle to AES, but involving the analysis of crystal surfaces. Under these circumstances, emitted Auger electrons may experience diffraction effects, which help to determine the crystal structure in the vicinity of the emitting atoms	
Low-energy electron diffraction (LEED)	Bombardment with electron beam in the 20–1000 eV range. Emerging diffracted beams are accelerated onto a phosphorous screen. Used to control the surface quality of a sample prepared under UHV conditions	Reflection high energy electron diffraction (RHEED)
<i>Ion emission techniques</i>		
Secondary ion mass spectrometry (SIMS)	Bombardment of a sample surface with a primary ion beam followed by mass spectrometry of the emitted secondary ions. Used for the analysis of trace elements in solid materials, continuous erosion by ion beam allows depth analysis	Secondary/sputtered neutral mass spectrometry (SNMS)
Ion scattering spectroscopy (ISS)	The sample surface is bombarded with a beam of ions. A certain number will undergo elastic reflection and the intensity and emission angle of these scattered ions provides structural information about the surface	Low-energy ion spectroscopy (LEIS) Medium-energy ion spectroscopy (MEIS) High-energy ion spectroscopy (HEIS) Rutherford backscattering (RBS)
Temperature-programmed techniques	The sample is heated in a vacuum and the residual gas is detected using a mass analyzer. As the temperature rises, certain absorbed species will have enough energy to desorb and will be detected	Temperature-programmed desorption (TPD) also known as thermal desorption spectroscopy (TDS) Temperature-programmed reaction spectroscopy (TPRS)
<i>Microscopy</i>		
Light microscopy (LM)	Low-resolution visual analysis of surfaces	
Electron microscopy (EM)	High-resolution analysis of surfaces using high-energy electrons that cause resident electrons to be ejected. When these are detected, they can be converted into an image	Scanning electron microscopy (SEM) Transmission electron microscopy (TEM) Scanning transmission electron microscopy (STEM) Scanning Auger microscopy (SAM)

Continued

Table 1 Continued

Technique	Principles and applications	Variations
Field ion microscopy (FIM)	The specimen is formed into a sharp tip to which a positive potential is applied. The tip is surrounded by inert gas atoms, which are drawn towards it. As they strike, electrons tunnel into the tip leaving the gas atoms ionized. The ions then accelerate away and can be detected as spots on a screen, corresponding to atoms at the tip surface	Atomic probe field ion microscopy (APFIM)
Scanning probe microscopy (SPM)	Techniques in which a surface is imaged at high (and in some cases atomic) resolution moving a probe across it, either in direct contact or just above the surface	Atomic force microscopy (AFM) Ballistic electron emission microscopy (BEEM) Chemical force microscopy (CFM) Lateral force microscopy (LFM) Magnetic force microscopy (MFM) Near-field scanning optical microscopy (NSOM) Scanning tunneling microscopy (STM)

may be analyzed (the uppermost layer of atoms or molecules) or the probe may penetrate up to several micrometers into the bulk material.

Many of the available surface analysis techniques are described in more detail in **Table 1**. The large number of techniques sometimes makes it difficult to determine which would be the most suitable for the analysis of particular materials. Often, the choice is made based on empirical knowledge, but in many cases individual researchers stick to the methods they understand the best and work with most of the time. Depending on the chemical composition and physical properties of a surface, many or only a few of the available analytical techniques might provide useful information. Several methods are often used in parallel to derive the maximum amount of information from a particular surface.

## Contamination and Decontamination

There are many reasons for studying and characterizing surfaces, and the applications of surface analysis cover a wide range of basic science disciplines, industrial applications, and manufacturing processes. Surface analysis is usually carried out to obtain information about the chemical composition of a surface, the concentration of trace impurities, the physical structure, or alterations that occur when surfaces are contaminated, stressed, or subjected to different types of modification. Such information is useful to researchers or manufacturers because they must understand the behavior of the materials they are working with in order to verify theories or make better products and procedures.

Contamination on the surface of the sample may represent the 'target' of the analytical procedure. One case in point is the use of surface analysis to verify cleaning procedures. For example, it would be helpful to know if a cleaning procedure used to remove environmental contaminants (e.g., organic depositions from vehicle exhaust fumes on buildings) also caused damage to the underlying material that should be preserved. It would also be helpful to know whether a cleaning procedure used to remove a harmful contaminant was in fact exchanging that contaminant for another surface species that could have similarly hazardous properties. Contamination often affects the surface of a material more than the bulk, even in materials with a stated purity level in excess of 99.99%. Polycrystalline metals, for example, often have impurities that segregate at grain boundaries.

In other cases, contamination on the surface of the sample is undesirable, and detracts from the analytical procedure. Almost all surfaces exposed to the atmosphere have an adventitious overlayer containing organic carbon and oxygen, even where precautions have been taken in specimen handling. While many of the analytical techniques used to investigate surfaces allow *in situ* cleaning after sample loading, it is still a good idea to clean the sample prior to analysis. Careless handling, including contact with fingers or storage in nonclean environments, often results in the formation of a surface layer that completely masks the underlying substrate. The contamination resulting from careless handling can also have deleterious effects on many surface analysis systems.

The modification of surfaces, even those normally considered to be inert, caused by atmospheric exposure can also result from reactions involving simple gases. Many metals, alloys and even polymers possess a very thin surface layer of oxide or hydroxide because of reactions with oxygen and/or water. Carbon monoxide also reacts with these substrates, so the consequences of such interactions must be taken into account when interpreting experimental data. Various cleanup strategies can be employed to overcome such problems, including sputtering (ion bombardment), heating, or chemical reduction, but this depends on the composition of the analyte. While heating will often desorb surface contaminants, many samples cannot withstand the high temperatures. Ion bombardment and chemical reduction can sometimes modify the surface chemistry of the sample in an undesirable way, i.e., the procedures are destructive. This also applies to routine analysis methods using ion bombardment, e.g., secondary ion mass spectrometry, although the destructive effects can be limited by maintaining a low ion current. Electron beam methods such as scanning electron microscopy and Auger electron spectroscopy can alter some samples as a result of interactions with the probe electron beam. Other approaches such as X-ray photoelectron spectroscopy and infrared reflection absorption spectroscopy usually do not alter the specimen.

## Information Derived from Surface Analysis

Although many of the techniques described in Table 1 can be used to obtain chemical and elemental information about surfaces, the data are in some cases restricted to particular chemical groups and moieties, while in other cases particular elements cannot be detected. For example, Auger electron spectroscopy and X-ray photoelectron spectroscopy cannot detect hydrogen directly, and only in a few instances can the presence of hydrogen be inferred from spectral features. Both techniques exploit the fact that electrons ejected from atoms in solids will not retain their characteristic quantized energy in collisions with bulk atoms. Thus, the observed signals come only from atoms in the first few layers of the solid. With techniques that depend on the emission of electrons, photons, or ejected material, the analyst must be vigilant of possible overlapping signals from unexpected species. Techniques such as ion scattering spectroscopy, which measures the energy reflected from a surface irradiated with an incident ion beam (usually  $\sim 1$  keV), are not able to discriminate between elements with similar properties, i.e., they

cannot distinguish small changes in atomic number or atomic weight.

Estimates of the relative amounts of the elemental and chemical surface species are often very important, so the detection level required for a given probe must be considered. Occasionally, surface sensitivity in the parts per million regime is needed and a technique such as secondary ion mass spectrometry would be the most appropriate. In this approach, a beam of energetic ions is used to eject material from the specimen surface for analysis by mass spectrometry. Where a lower sensitivity is acceptable, such as when the surface material needs to be identified but not necessarily quantified, then other techniques would be sufficient. The ability to quantify the surface composition of a sample accurately has a strong influence on the choice of analytical technique. The dynamic range, i.e., the variation in sensitivity to different elements or chemical species, must also be taken into account. For example, secondary ion mass spectrometry has sensitivity factors that differ over a million-fold range for different elements. This contrasts with Auger electron spectroscopy and X-ray photoelectron spectroscopy where the range of sensitivities differs by  $\sim 20$ -fold for different elements.

Spatial resolution is also important in surface analysis and the ability to resolve differences in surface composition on the nanometer scale is important in fracture analysis and many other areas. In this respect, secondary ion mass spectrometry has rather poor resolution, with a sample area generally in excess of  $7500 \text{ nm}^2$ . In contrast, techniques such as Auger electron spectroscopy, in which the probe diameter can be finely tuned, allow sample areas of  $70 \text{ nm}^2$  or less to be analyzed.

**See also:** **Surface Analysis:** X-Ray Photoelectron Spectroscopy; Particle-Induced X-Ray Emission; Auger Electron Spectroscopy; Appearance Potential Spectroscopy; Desorption Techniques; Ion Scattering; Low-Energy Electron Diffraction.

## Further Reading

- Briggs D and Seah MP (1990) *Practical Surface Analysis*, vol. 1. New York: Wiley.
- Briggs D and Seah MP (1991) *Practical Surface Analysis*, vol. 2. New York: Wiley.
- Desjonqueres MC and Spanjaard D (1993) *Concepts in Surface Physics*. Berlin: Springer Verlag.
- Riviere JC (1990) *Surface Analytical Techniques*. Oxford: Clarendon Press.
- Woodruff DP and Delchar TA (1994) *Modern Techniques of Surface Science*. Cambridge: Cambridge University Press.



## X-Ray Photoelectron Spectroscopy

**S Tougaard**, University of Southern Denmark,  
Odense M, Denmark

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

X-ray photoelectron spectroscopy (XPS) is the most broadly applicable surface analysis technique today. It relies on measurements of the energy distribution of photon-excited electrons from atoms in the surface region of the solid. It can detect all elements except hydrogen with a sensitivity that varies by only a factor of  $\sim 30$  across the periodic table. The absolute sensitivity is  $\sim 0.01\text{--}0.3\text{ at.}\%$  depending on the element. The electron-binding energies are sensitive to the chemical state of the atom and this is used to determine the chemical state of the atoms. Because of this feature, the XPS technique is also known under the acronym ESCA (electron spectroscopy for chemical analysis). It is a nondestructive technique (although damage by the X-ray beam can be significant in some organic materials) and in general it is less destructive than the other electron or ion impact techniques (Auger electron spectroscopy (AES) and secondary ion mass spectroscopy (SIMS)). The technique is highly surface sensitive with a probing depth of  $\sim 1\text{--}10\text{ nm}$ . The surface area detected can be as large as  $1\text{ cm} \times 1\text{ cm}$  or as small as  $3\text{ }\mu\text{m} \times 3\text{ }\mu\text{m}$ , and imaging XPS has become a common feature of the current generation of XPS spectrometers. In recent years, lateral resolutions in the  $100\text{ nm}$  range have been achieved by using the new generations of synchrotron light sources with high brightness. With special design of the spectrometer, XPS may be used to analyze gaseous and liquid samples, but it is mostly used for solids.

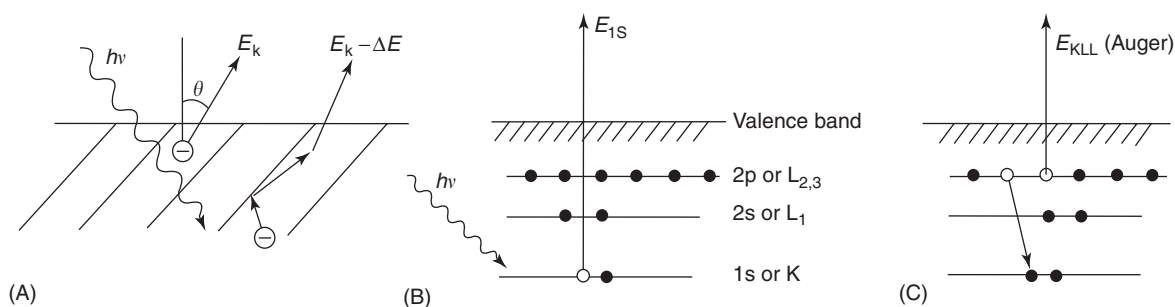
The photoelectron spectra can be generated using a primary beam of ultraviolet radiation (known as ultraviolet photoelectron spectroscopy (UPS)), or X-rays to yield an XPS spectrum. It can also be generated by synchrotron radiation (SR), in which case the electron spectrometer must be attached to the beam line at a synchrotron facility. UPS is essentially a molecular spectroscopy that provides high-resolution valence band spectra, while XPS is used predominantly to study core level transitions, and also to study valence band electrons.

### Principles of XPS

The physics of the XPS process is essentially straightforward. A primary beam of soft X-rays (usually Al  $K\alpha$  or Mg  $K\alpha$ ) interacts with the specimen. This leads to the excitation and subsequent ejection of low-energy electrons resulting from the photoionization of the atoms, molecules, or ions of the specimen. This is indicated in the schematic of **Figures 1A** and **1B**. The incoming photon of energy  $h\nu$  will, in this particular case, photoexcite a  $1s$  electron of binding energy  $E_B$  relative to the Fermi level; an additional term, the sample work function  $\Phi_{\text{sample}}$ , is needed to remove the electron from the solid. This energy is gained as the electron enters the spectrometer where it, in turn, has to overcome the work function  $\Phi$  of the spectrometer. Conservation of energy requires that

$$E_B = h\nu - E_K - \Phi \quad [1]$$

Thus, the determined electron-binding energy is independent of the work function of the sample. In the case of an insulator, the sample may become charged due to the emission of electrons during X-ray bombardment and the measured energy will be shifted. This is corrected for by setting a well-defined peak



**Figure 1** (A) Schematic of the XPS excitation process. (B) Schematic representation of the electron energy levels of an F atom and the photoionization of a F  $1s$  electron. (C) Auger emission relaxation process for the F  $1s$  empty-state produced in (B).

within the spectrum to its known binding energy value, e.g., carbon 1s to 285.0 eV.

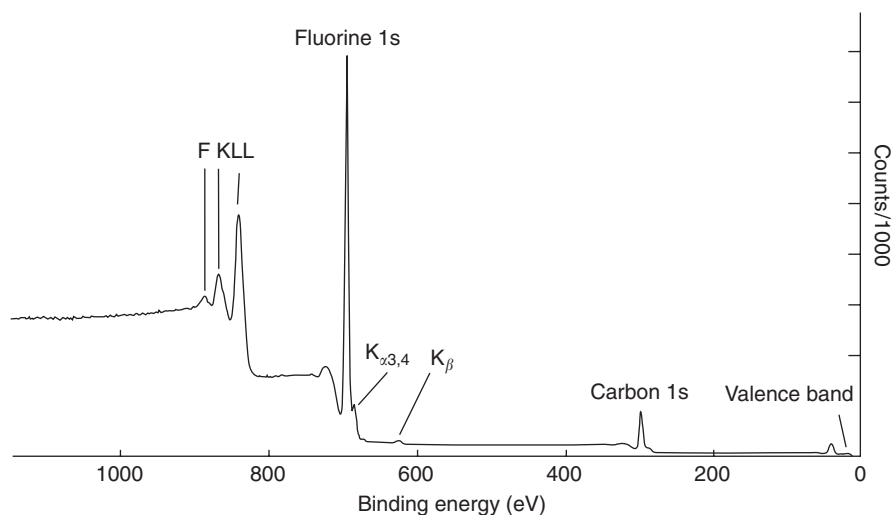
The X-ray photoelectron spectrum is recorded either as a function of electron-binding energy  $E_B$  or directly as the kinetic energy  $E_K$  of the measured photoexcited electrons. The two are equivalent and are linked by eqn [1]. **Figure 2** shows the XPS spectrum of polytetrafluoroethene (PTFE). The use of PTFE as an example also illustrates another strength of XPS – its ability to analyze insulating materials. In addition to the ejection of the photoelectron, the photoionized atoms may relax by the emission of an Auger electron (see **Figure 1C**). In the Auger process, the empty core hole left behind after the photoexcitation (**Figure 1B**) is filled with an electron from one of the outer orbitals and the energy released is given up to a third electron that is emitted. This Auger electron is characterized by the levels involved and for fluorine, it is seen as a range of peaks (F KLL) corresponding to the possible combinations of the different  $L_1$ ,  $L_{2,3}$  energy levels involved (see **Figure 2**). The peak has a lower kinetic energy than the corresponding photoelectron and consequently it occurs at an apparent higher ‘binding energy’. For Auger electrons the ‘binding energy’ has no real physical significance and is a mere mathematical consequence of representing the observed electron energies by eqn [1]. Electron-excited AES is an important separate analytical technique, but much useful information can also be obtained from the X-ray-excited Auger spectrum. The Auger process involves typically one or more weakly bound valence electrons. Since the energy distribution of valence electrons may be strongly affected by the chemical bonds, the Auger peaks often contain considerable

information on the chemical state of the atoms. Details of the structure of the F KLL peak structures are thus markedly different for solids of different covalent character (e.g., NaF, NaBF<sub>4</sub>) and this is used as a fingerprint technique to distinguish between different chemical forms.

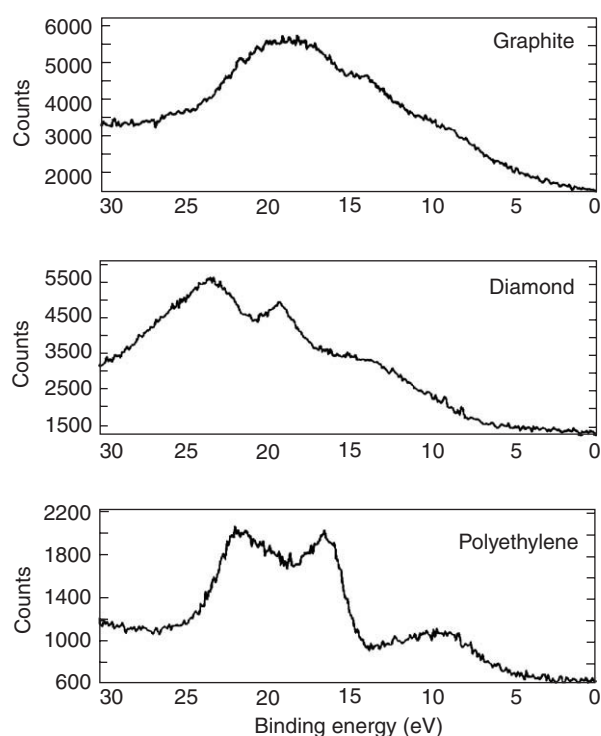
The electrons in the valence band, situated in the low-binding-energy region (0–35 eV), have peaks much less intense than the 1s (core level) peaks of carbon and fluorine, as a result of the very low photoelectric cross-section compared with the core levels. These features can be seen in **Figure 2**. The electron states in the valence band is studied extensively in UPS where use of low-energy photons make it possible to study transitions from occupied to unoccupied energy levels with high-energy resolution. At higher photon energies (typical of XPS), the excited electron is in a free electron state since it has an energy well above the Fermi level where the density of final states is featureless. The measured energy distribution in valence band XPS reflects therefore directly the density of occupied states. As a result, chemical bonding information may often be extracted from these spectra. **Figure 3** shows an example of spectra from carbon in three different forms (graphite, diamond, and ethylene) and it is clear that the valence spectral regions are distinctly different. This phenomenon is also used extensively to study variations in the electronic structure of alloys and semiconductor–metal interfaces.

## Qualitative Interpretation

The observed peaks in the XPS energy spectra of emitted electrons are essentially uniquely related to



**Figure 2** XPS spectrum of the polymer PTFE, showing the core level (carbon 1s and fluorine 1s, fluorine 2s at 31 eV and valence electron emission, X-ray induced Auger peaks (F KLL), and X-ray satellites ( $K_{\alpha 3,4}$  and  $K_{\beta}$ )). Note the step-like increase in intensity on the left-hand side of each major peak which is due to inelastically scattered electrons (see **Figure 7**).



**Figure 3** XPS valence band spectra of graphite, diamond, and polyethylene.

the binding energies of specific elements. Then by assigning the major peaks in a spectrum one readily determines which elements are present. Since the electron mean free path for the excited photoelectrons is only 0.5–3 nm the observed elements are situated in the outermost 5–20 atomic layers of the analyzed sample.

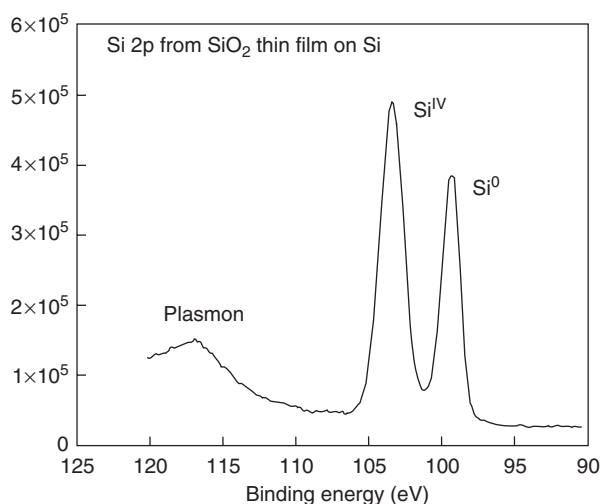
The assignment of the peaks, which, as described above, may be XPS transitions or their associated Auger peaks, is accomplished with the aid of a handbook or computer database of peak positions. All elements of the periodic table, with the exception of hydrogen, can be detected and XPS is able to analyze light elements with ease, a feature that makes it particularly attractive for the analysis of organic polymers. Minor X-ray satellites will be present if an achromatic X-ray source has been employed and arise from photoelectron transitions excited by the minor components ( $K_{\alpha 3,4}$  and  $K_{\beta}$ ) of such X-radiation. In **Figure 2**, the major photoelectron peaks (fluorine 1s and carbon 1s) are a result of Al  $K_{\alpha 1,2}$  induced photoemission. Features labeled  $K_{\alpha 3,4}$  and  $K_{\beta}$  arise from fluorine 1s electrons ejected by the Al  $K_{\alpha 3,4}$  and Al  $K_{\beta}$  components of the radiation, respectively. Instruments are often equipped with a monochromator that remove the  $K_{\alpha 3,4}$  and  $K_{\beta}$  satellite lines from the beam of X-rays.

## Chemical Analysis

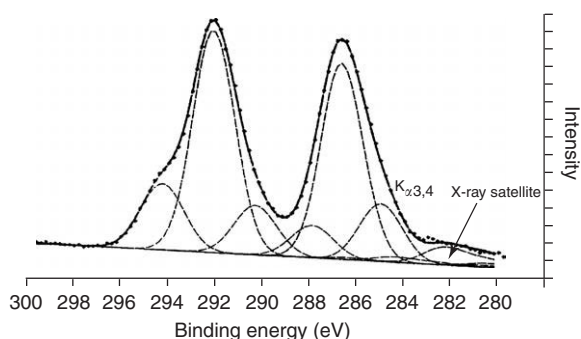
The chemical information implicit in the X-ray-induced photoelectron spectrum is a very important feature of the XPS technique and is, as mentioned above, the source of the alternative acronym ESCA. For almost all elements the binding energy of the photoelectron peak shows small perturbations that reflect the chemical environment of the element under consideration.

As mentioned above, both the valence region of the spectrum and the Auger peaks give information on the chemical bonding because the bonds will usually change the density of states in the valence band. However, the most important source of chemical state information in XPS comes actually directly from the energy position of the photoelectron core level peaks. Although the core levels are not directly involved in the chemical bonding process, their binding energy depends on the chemical environment. In a simplified picture, this effect, which is called the ‘chemical shift’, originates from the change in electrostatic potential inside the orbitals of the valence electrons (i.e., at the location of the core electrons). Thus, as the effective valence electron charge around a given atom changes by the amount  $\Delta q$ , the potential of all core electrons of this atom changes by  $\Delta q/(4\pi\epsilon_0 r)$ , where  $r$  is the radius of the valence electron orbital, and the core level binding energies change correspondingly. Large variations in chemical shifts are thus expected for different oxidation states but can also arise from difference in lattice site and molecular environment. In a more accurate model, the electron orbitals cannot be assumed to remain frozen during photoemission and the electron-density around the atom relaxes in response to the creation of the positively charged core hole. This reduces the kinetic energy of the outgoing photoelectron and thereby increases the derived electron-binding energy. These effects are known as shake up and may be seen as distinct features in the spectrum (see below).

An example of the chemical shift phenomenon is illustrated in **Figure 4**, which shows the Si 2p peak from a thin SiO<sub>2</sub> film on a Si substrate. This ability to distinguish between Si and SiO<sub>2</sub> is widely used by the semiconductor industry to monitor the morphology of nanometer thin SiO<sub>2</sub> structures. **Figure 5** shows the carbon 1s spectrum of a complex fluorocarbon surfactant. The experimental spectrum is indicated by the individual points and the computer convolution of the various components is represented by the solid line. Components resulting from carbon in the following chemical states are identified as C–H and C–C (lowest binding energy), C–O and C–N, C=O,



**Figure 4** Si 2p spectrum from a thin  $\text{SiO}_2$  film grown on a Si substrate. The  $\text{SiO}_2$  film is  $\sim 3.2$  nm thick.



**Figure 5** C 1s spectrum of a complex fluorocarbon surfactant. The assignment of the individual components is as follows: (left to right) C–H and C–C (285.0 eV), C–N and C–O (286.4 eV), C=O (287.8 eV), C–F (290.2 eV), C–F<sub>2</sub> (291.8 eV), C–F<sub>3</sub> (294.2 eV).

C–F, C–F<sub>2</sub>, and C–F<sub>3</sub> (highest binding energy). The low-intensity feature at 282 eV is the CF<sub>2</sub> component excited by the Al K $\alpha$  X-ray satellite. The binding energy of the carbon 1s electrons increases as they become associated with the more electronegative atoms; this shift also increases with the number of fluorine atoms attached to each carbon. Such a positive binding energy shift is not always observed; in the case of sulfur, a positive shift (relative to elemental sulfur) is seen for the sulfate ion, but a negative shift is observed in the case of a sulfide. The same phenomenon is observed in the carbon 1s spectrum of carbide compounds.

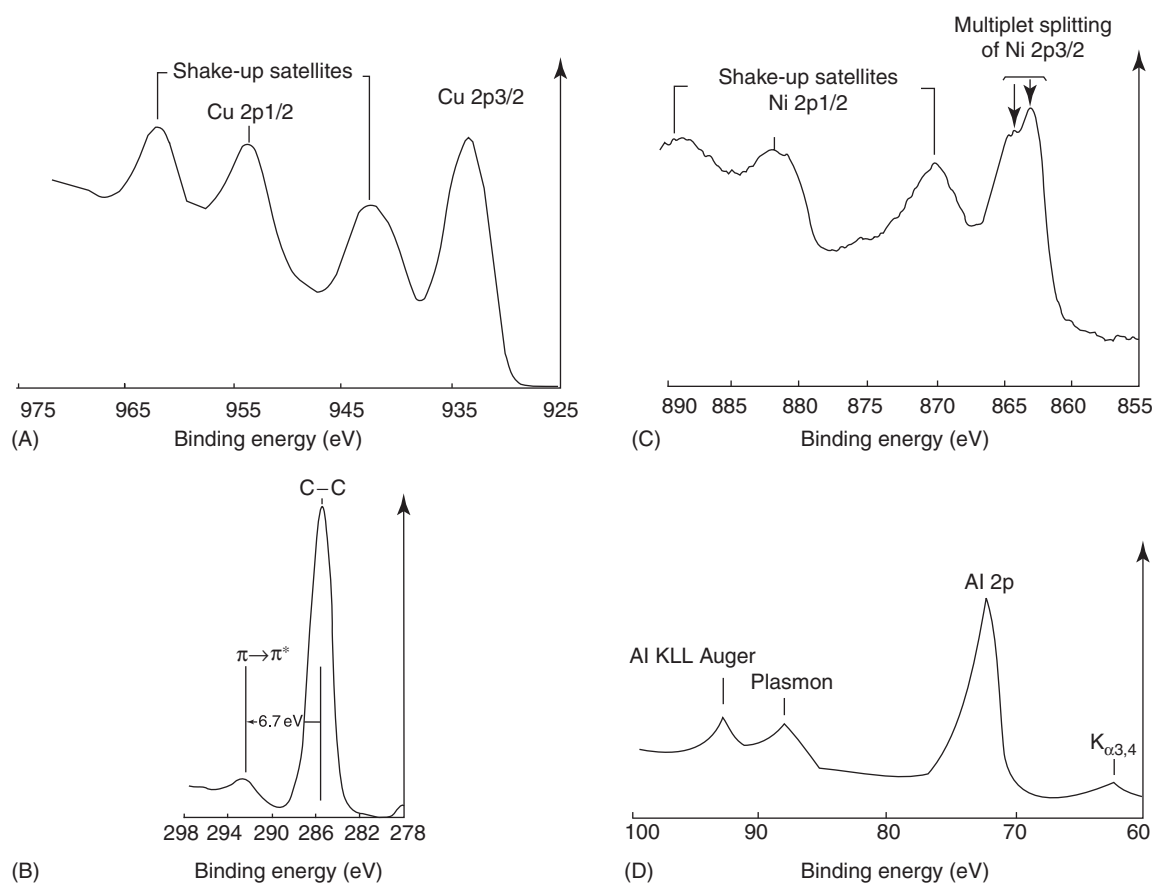
Such core level shifts are well documented in handbooks and databases and are widely used to get chemical state information. But in some cases, the change in potential is insignificant although the valence electrons are changed considerably due to

different chemical bonds. In such cases, it is of value to examine those transitions that are directly affected by the valence electron distribution (i.e., the valence band region and the Auger peaks). As an example, **Figure 3** illustrates the valence band spectra of three forms of carbon: polyethylene, graphite, and diamond. The carbon 1s core level spectra of these three materials are very similar, but the valence band spectra (carbon 2s and 2p electrons) are very characteristic and provide a ready means of differentiation between the three materials. When recorded at better energy resolution it is even possible to differentiate between individual polyolefines and their conformations. Although this region of the spectrum is not amenable to elemental quantification, it is possible to use such spectra in conjunction with molecular orbital calculations, or merely as a ‘fingerprint’ of a specific material. Another example is copper oxides. While the Cu 2p peak in CuO is shifted by  $\sim 1.5$  eV relative to Cu (metal), the Cu 2p peak in Cu<sub>2</sub>O is not shifted. Here, the three forms of Cu can, however, be easily distinguished by shifts in their Auger peaks. Thus, the L<sub>3</sub>VV peak is shifted by  $\sim 0.6$  eV for Cu<sub>2</sub>O and by  $\sim 1.9$  eV for CuO relative to Cu (metal).

## Energy Loss Features in the XPS Spectrum

The outgoing photoelectron may interact with electrons in the solid which gives rise to additional energy loss processes that will lower the measured kinetic energy of the electron and yield additional features in the XPS spectrum at a binding energy slightly greater than the core level peak.

The processes are traditionally divided into two types. The first type is known as ‘final-state effects’ or ‘intrinsic excitations’ that take place at the time immediately after the photoexcitation process before the photoelectron has moved away from the atom. Thus, the interaction with the suddenly created core hole gives rise to excitations of valence electrons to empty states in the valence band, known as ‘shake-up’ processes. This reduces the kinetic energy of the emitted core electron slightly and consequently it appears at a slightly higher binding energy within the photoelectron spectrum as a characteristic satellite structure. Shake-up satellites occur in several classes of spectra, the most notable being the 2p spectra of the d band metals as shown in the copper 2p spectrum of CuO in **Figure 6A**, and the bonding to antibonding transitions of the  $\pi$  molecular orbitals in aromatic organics, **Figure 6B**. A related process is that of ‘shake-off’, in which the valence electron is ejected to higher energy levels. Because the density of

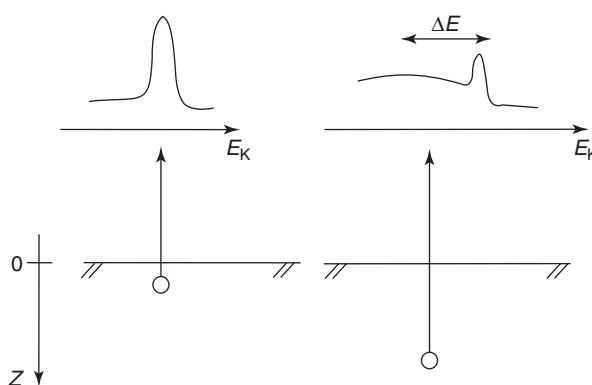


**Figure 6** Extrinsic and intrinsic loss features in XPS. (A) CuO: Cu 2p shake-up satellites, (B) C 1s and  $\pi \rightarrow \pi^*$  shake-up satellite, (C) NiO: Ni 2p shake-up and multiplet splitting, and (D) Al plasmon loss feature from pure aluminum.

these final states is broad and featureless, such losses generally occur as a very broad feature within the large background of inelastically scattered electrons.

Multiplet splitting of a photoelectron peak may occur in a compound that has unpaired electrons in the valence band. The effect arises because an unpaired core electron that remains after the photoemission will couple to unpaired spin electrons in the valence band and this leads to more than one final state peak which is seen as a peak splitting. Since the distribution of unpaired electron spins in the valence band changes with the chemical bonding so does the size of the multiplet splitting. Multiplet splitting effects are observed in cobalt and nickel (2p levels), manganese and chromium (3s levels), and the rare earths (4s levels). The spectrum of nickel(II) oxide in **Figure 6C** illustrates the diagnostic utility of such features; the multiplet splitting exhibited by this compound distinguishing it quite conclusively from nickel(II) hydroxide from which it is absent.

The second type of loss features, known as 'extrinsic' effects, occur during transport of the photoelectron from the emitting atom to the surface of the



**Figure 7** Schematic of the loss of peak intensity with increasing excitation depth and the corresponding rise in intensity of inelastically scattered electrons at lower kinetic energies  $E_K$ .

solid where the electron will excite valence electrons of the solid by plasmon and interband excitations. The distribution of energy loss depends characteristically on the distance traveled and thereby on the depth distribution of photon-emitting atoms. The situation is illustrated in **Figure 7**. Analysis of

the measured XPS peak shape in a 50–100 eV energy range is therefore used to determine the depth distribution of atoms (see below). In general, the extrinsic energy losses is rather featureless but for some solids, with distinct structure in the valence band density of states or with very narrow plasmon excitations, these inelastically scattered electrons may be seen in the spectrum. This is seen for SiO<sub>2</sub> in Figure 4 and for aluminum in Figure 6D.

## Depth of Analysis in XPS

The photoelectrons in XPS with typical energies in the 100–1500 eV range can only travel a short distance in the solid before they lose energy. This is the reason for the high surface sensitivity of XPS. If  $I_0$  is the flux of electrons originating at depth  $z$ , the flux detected at the peak energy, i.e., without energy loss is

$$I_d = I_0 \cdot e^{-z/(\lambda \cos \theta)} \quad [2]$$

where  $\theta$  is the angle of electron emission and  $z/\cos \theta$  the distance traveled through the solid. The quantity  $\lambda$  is the electron attenuation length that depends on several factors but the most important is the electron kinetic energy. As a general guide, this parameter varies as  $E_K^{0.75}$ . Values of  $\lambda$  are usually in the range of 0.5–2 nm for the electron energies of interest in XPS. From eqn [2] it follows that 65% of the photoelectron peak signal emanates from a depth of  $1\lambda$  (at normal electron emission), 85% from  $2\lambda$ , and 95% from  $3\lambda$ . The analysis depth in XPS is often taken as three times the electron attenuation length, although it must be appreciated that the analysis obtained within such a depth is heavily biased toward the outer region of the specimen for the reasons outlined above. This gives a characteristic analysis depth of 1.5–6 nm. Additional surface sensitivity in XPS may be obtained by tilting the sample within the spectrometer and thus changing  $\cos \theta$ . The reduction of the electron take-off angle in this manner, to reduce analysis depth, is known as angular-resolved XPS, which is also used to get information on the in-depth distribution of atoms.

As described above,  $\lambda$  is a function of electron kinetic energy and in the XPS experiment this can be varied for the electron orbital of interest by changing the photon source energy. In the laboratory this can be achieved by using an alternative conventional X-ray source (e.g., Al K $\alpha$  or Mg K $\alpha$ ,) or by increasing the analysis depth by using a high-energy source (e.g., Zr L $\alpha$ , Ag L $\alpha$ , or Ti K $\alpha$ ). SR can be extremely useful in this respect as it becomes possible to match the kinetic energy, and hence the analysis depth, for

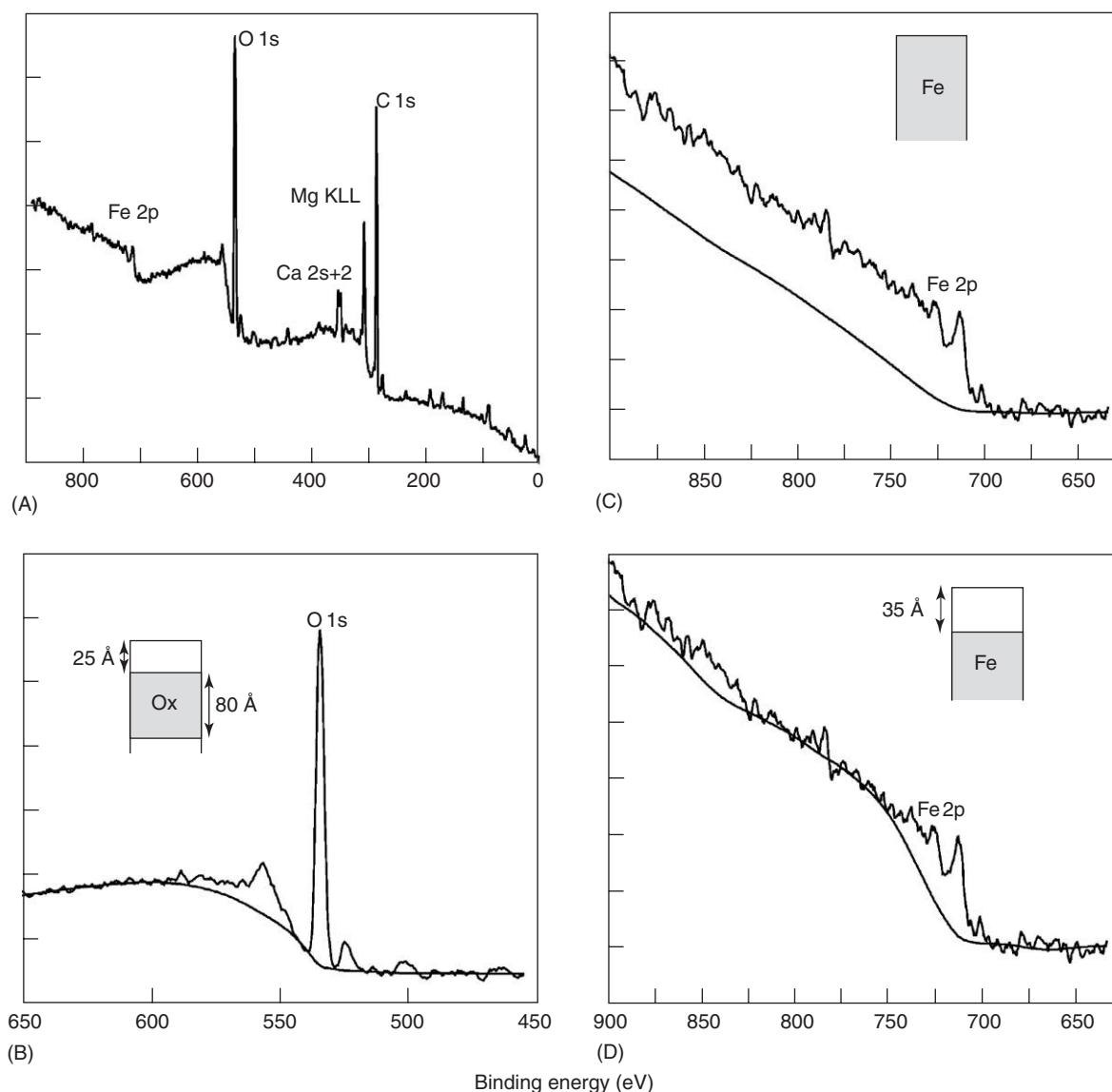
each electron orbital to be studied. This has been used to great effect in the study of semiconductor interfaces.

## Quantitative Interpretation of XPS

Quantification is simple if the atoms are homogeneously distributed with depth since the intensity of each XPS peak is then directly related to the abundance of that particular element at the specimen surface. The peak intensity will usually be reported as a peak area and this will be normalized using atomic sensitivity factors (the intensity of the photoelectron transition of interest,  $I$ , is related to the concentration of that element within the XPS analysis volume, and the sensitivity factor,  $S$ , in the following way:  $I = \text{concentration} \times S$ ). Such atomic sensitivity factors are a function of the basic physical parameters, such as the relative photoelectron cross-sections of the different elements, electron attenuation lengths, and instrumental parameters, such as analyzer transmission functions, of the XPS experiment. The ratio of normalized peak area to the sum of normalized peak areas for the major peaks of all elements detected in the spectrum provides an analysis as an atomic fraction (or when multiplied by 100, atomic %).

If the depth distribution is not homogeneous over the analysis depth, this quantification scheme is not valid because of the strong signal attenuation with depth (eqn [2]). It is simple to correct for this, if the in-depth distribution is known. But since this is rarely the case, it is customary to still use the above simple quantification scheme that assumes a homogeneous depth distribution; however, the resulting quantification is then quite inaccurate. The accuracy is improved considerably by analysis of the XPS peak shape in a 50–100 eV energy region around the peak. This is because photoelectrons that are ejected from atoms close to the surface have a low probability for energy loss and most of these will end up at the original excitation energy, i.e., at the peak energy. Photoelectrons that are excited from atoms at larger depths will have a much lower chance of escaping without energy loss and most of these will end up in the broad background at lower kinetic energies as shown schematically in Figure 7. The shape of the peak region including the continuous background of inelastically scattered electrons is therefore characteristic of the in-depth distribution of emitting atoms and the in-depth distribution of atoms can be determined by analysis of this energy distribution. A practical example of this is in Figure 8A, which shows the XPS spectrum of an iron sample that has been

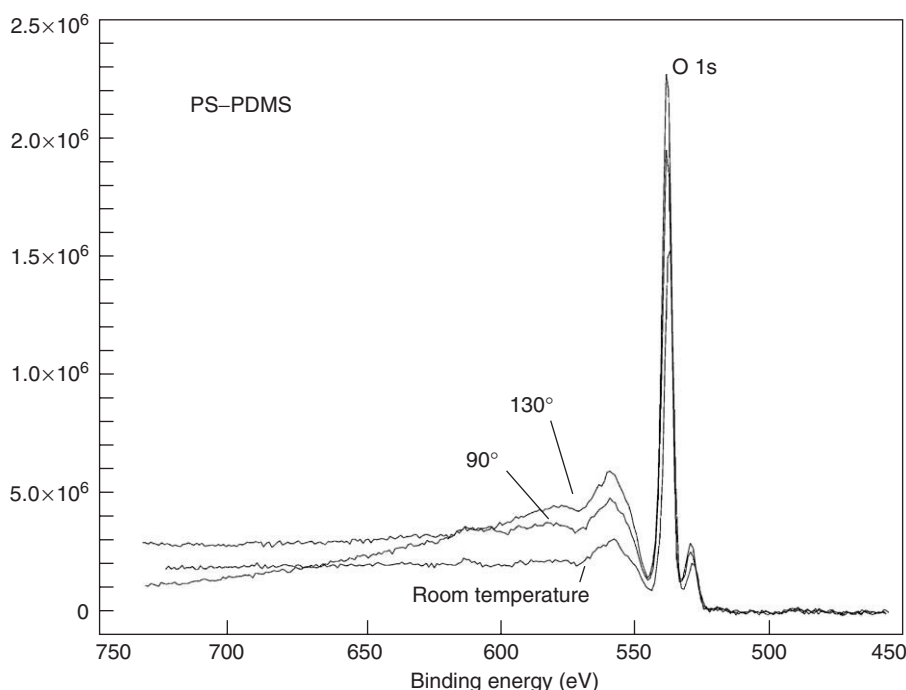




**Figure 8** (A) XPS of an iron sample that has been exposed to a corrosive marine environment. The presence of Fe, O, C, Mg, and Ca are seen. The C and Ca peaks have high intensity and rather low steps in background intensity and these atoms must therefore be near the surface (according to **Figure 7**). The Fe 2p peak has a low intensity and a steeply rising background of inelastically scattered electrons which readily shows that the iron concentration increases with depth, (B) Quantitative analysis of the O 1s peak shape shows that the oxide forms a 80 Å thin film covered by an overlayer of ~25 Å (of predominantly C and Ca). The analysis of the Fe 2p peak shape in (C) and (D) shows that the Fe concentration is not constant with depth but the Fe substrate is covered with a ~35 Å iron-free overlayer.

exposed to a corrosive marine environment. The presence of Fe, O, C, Mg, and Ca are seen. The C and Ca peaks have high intensity and rather low backgrounds and these atoms must therefore be near the surface. The Fe 2p peak has a low intensity and a steeply rising background of inelastically scattered electrons, which readily shows that the iron concentration increases with depth. In comparison, the O 1s peak has high-peak intensity and a relatively lower background that indicates that it is closer to the surface. Quantification of the peak shapes is done by

application of an integrated software package that simulates the inelastic background for assumed depth distributions which are varied until good account is found. The analysis assuming different distributions of Fe (**Figures 8C** and **8D**) shows that the sample consists of Fe which is covered by ~35 Å overlayer while a similar analysis of O 1s (**Figure 8B**) reveals the depth distribution of oxygen. The Mg KLL peak is mixed with the C and Ca peaks and therefore requires a more elaborate analysis of the background.



**Figure 9** XPS of a diblock copolymer that consists of poly(styrene)–poly(dimethylsiloxane) (PS–PDMS). PDMS contains oxygen atoms which are not in PS and the shape of the O 1s peak can be used to distinguish the in-depth distribution of the two polymer blocks.

Figure 9 shows the XPS spectrum of a diblock copolymer that consists of poly(styrene)–poly(dimethylsiloxane) (PS–PDMS). PDMS contains oxygen atoms that are not in PS and the O 1s peak can be used to distinguish the distribution of the two polymer blocks. The O 1s peak intensity increases gradually and the inelastic background decreases as the sample is annealed, which shows that PDMS segregates to the surface. The quantitative nanostructure of the polymers can be determined as above by computer simulation of the peak shape. Alternative possibilities to determine if a sample is homogeneous and to distinguish the composition of the outermost surface from the bulk is by varying  $E_K$  or by varying the emission angle as was discussed in the previous section.

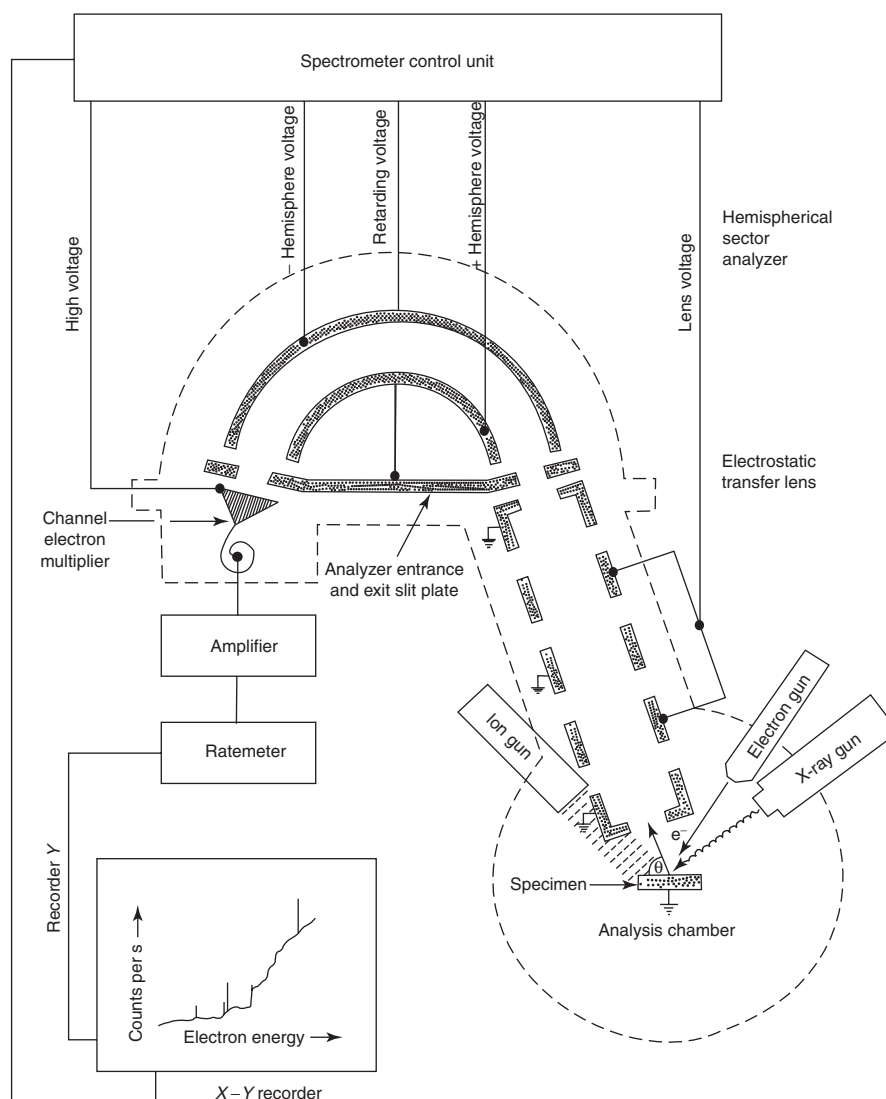
### Spatial Resolution in XPS

There has been a strong drive to see the unique advantages of XPS applied to smaller and smaller surface areas and imaging XPS has now become a common feature of the current generation of XPS spectrometers. Today, areas of  $\sim 100\mu\text{m} \times 100\mu\text{m}$  are the typical lower limit with conventional spectrometers and X-ray sources. Using a scanning XPS microprobe to produce a focused X-ray beam on the sample, lateral resolutions of  $\sim 3\mu\text{m}$  is obtained in

the best commercial instruments. With the new generations of synchrotrons, it has been possible to increase the lateral resolution tremendously due to their high brightness. Today lateral resolutions better than  $\sim 100\text{nm}$  is routinely achieved and in the near future it is expected that  $\sim 2\text{--}5\text{nm}$  resolution will be reached. This development will open up a new exciting area of applications. However, the time required to get sufficient signal-to-noise ratio at high lateral resolution increases considerably.

### Instrumentation for XPS

The illustration of Figure 10 is a schematic diagram of a commercial electron spectrometer. The essential features are an X-ray source, the specimen, an electron transfer lens, an electron energy analyzer, and an electron detector (channel electron multiplier or channel plate), all contained within an ultrahigh vacuum (UHV) chamber. A lens system is used in front of the entrance slit of the hemispherical analyzer to provide a comfortable free working range of typically 30–50 mm above the sample. The lens system is typically designed with a multielement zoom lens type that can easily be set to operate at different magnifications. This is used for an easy variation of the analyzed area of the sample and for imaging.



**Figure 10** Schematic of a typical XPS spectrometer.

The X-ray source is generally of a twin anode design providing a broad beam source of Al  $K\alpha$  ( $h\nu = 1486.6$  eV) or Mg  $K\alpha$  ( $h\nu = 1253.6$  eV) X-radiation. Other sources can be used to probe deeper core levels. Possible high-energy photon sources for XPS include Si  $K\alpha$  (1740 eV), Zr  $L\alpha$  (2042 eV), and Ag  $L\alpha$  (2984 eV). One factor that limits the spectral resolution attainable in XPS is the natural linewidth of the X-ray source. This can be reduced substantially, and the extraneous radiation such as X-ray satellites and Bremsstrahlung removed, by the use of a crystal monochromator by which the spectral resolution is improved substantially as the X-ray linewidth is reduced from 0.85 eV to  $\sim 0.35$  eV, but the intensity is also reduced.

The only restriction on the solid sample in XPS is that it is vacuum compatible, although even this may

be relaxed as a cryo-stage, or differential pumping in the region of the specimen, will often be available. The limit on sample size depends on the mode of sample handling but a  $1\text{ cm}^2$  sample is the optimum size for most spectrometers. During analysis the sample will be located on a holder within a high-precision manipulator. An inert gas ion gun is used to remove material by ion sputtering and when used sequentially with XPS it is a convenient way to make a compositional depth profile to a depth of  $\sim 1\text{ }\mu\text{m}$ .

The XPS spectrum is typically recorded with a hemispherical sector analyzer. The general configuration of the analyzer is seen in **Figure 10**. By applying appropriate potentials to the hemispheres and prere-tardation grids it is possible to select the energy of the electrons that will reach the electron multiplier. The spectrum is then acquired by scanning these

potentials. The electron detection stage of the analyzer is a channel plate or more usually a channel electron multiplier, which today are used in arrays of five, seven, or more to enhance the energy resolution and signal strength.

This hardware is all enclosed in a UHV chamber with pressure below  $\sim 10^{-9}$  mbar, constructed of or shielded by, a low magnetic permeability material such as mu-metal. This is necessary to avoid the field from the Earth and nearby magnetic objects to affect the electron trajectory. The UHV requirement is, in part, a result of the very surface sensitivity of XPS itself. At a vacuum level of  $10^{-6}$  mbar a monolayer of gas is adsorbed on a solid surface in a few seconds if the surface atoms are reactive to the gas. At a vacuum of  $10^{-9}$  mbar (UHV) the time to form an adsorbed monolayer is several hours.

The final component of an XPS spectrometer, and one that has been alluded to several times already, is the data system for control of the spectrometer, X-ray sources, other hardware items such as auto-carousel, ion gun, and analysis position in small-area XPS and imaging XPS. This will record the spectrum or image and retain it for further data processing such as quantification or curve fitting. Such a computer system is an integral part of the spectrometer and is provided by the manufacturer when the instrument is purchased.

## Applications

XPS is probably the most widely used of the 'popular' surface analysis methods that are generally available (the others being AES and SIMS). This is due to its unique advantages – that it readily gives chemical state information, is easily quantifiable, and applicable to insulators. The principal disadvantages are that it has lower spatial resolution although this has improved tremendously over the past decade and that it has a detection limit of  $\sim 0.1$  at.%. This latter value is rather better than AES but is well removed from that of SIMS which can be in the  $\text{ng g}^{-1}$  range in favorable cases.

In this section, a few typical applications of XPS will be presented to illustrate the type of information that is readily attainable. Although the examples quoted are invariably parts of larger research programs, they illustrate the type of information that can be obtained quite readily within a very short space of time, perhaps by buying a single day of instrument time from a consultancy organization.

### Corrosion

In the examination of corrosion processes, there are usually two considerations of interest: the chemical

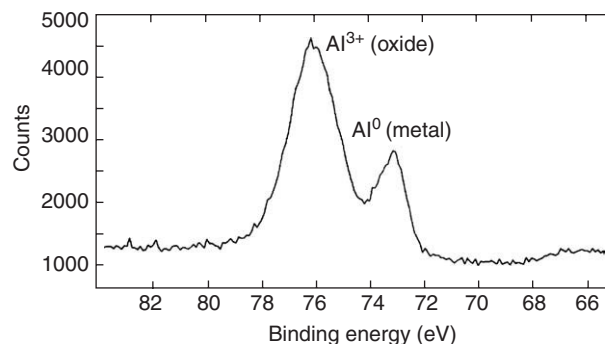
nature of the corrosion product (which may be a passive *film*) and the electrochemical history of the specimen. The subsequent breakdown of the passive film and other localized corrosion phenomena are generally studied by scanning Auger microscopy (SAM) but with the improvements that have taken place in imaging XPS and the inherent difficulties with SAM (possible beam damage and image instability resulting from electrostatic charging of thick corrosion products), XPS is now important in these investigations as well.

Figure 8 shows, as already described above, how nanothin oxide films can be easily quantified. Figure 11 shows the aluminum 2p spectrum of aluminum foil covered with a thin, oxide layer formed by exposure to air. The contributions from the metallic substrate, at a binding energy of 73 eV and the oxide over-layer aluminum 2p peak shifted to 76 eV, are easily distinguished in the experimental spectrum. By computer peak fitting of the spectrum it is possible to determine (as peak areas) the relative contribution from the two chemical states of aluminum. A chemical shift of this type is exhibited by almost all metals and provides a convenient way of separating the oxide component from its parent metal. In metals of variable valency it may be necessary to resort to an additional method of confirming the oxidation state, e.g., in copper compounds the X-ray-induced Auger peaks together with the CuO shake-up satellite of Figure 6A, when all taken together, provide unambiguous identification of Cu(metal),  $\text{Cu}_2\text{O}$ , and CuO.

If the overlayer is flat and covers the surface as a uniform layer, the oxide thickness,  $d$  is

$$d = \lambda \sin \theta \ln(I_{\text{oxide}}/I_{\text{metal}} + 1)$$

This does not apply if the surface is rough or if the layer is not uniform and then analysis of the peak shape may be used to determine the nanostructure of the film as discussed above (Figure 8). The electrochemical history of a specimen may be readily



**Figure 11** Al 2p spectrum from an oxidized aluminum foil. The oxide is  $\sim 3.8$  nm thick.

estimated by a surface-sensitive method such as XPS. This arises from the fact that if a metal electrode is polarized cathodically in an aqueous solution it will preferentially retain cations on removal, which are identified in the electron spectrum. The reverse is true of anodically polarized electrodes.

### Polymer Analysis

XPS provides a powerful method for the surface analysis of polymers and is employed routinely for this purpose in many laboratories around the world. All the constituent elements of organic polymers, with the exception of hydrogen, are detectable and have chemical shifts that indicate their chemical environment. An example of the segregation of a diblock copolymer during annealing is shown in **Figure 9** where the quantitative nanostructure of the PDMS segregation to the surface is determined by analysis of the peak shape. Combination of XPS with time-of-flight SIMS, which provides molecular specificity, improves the information attainable. The main advantage of XPS for many polymer applications is, however, obtained from the peak energy and peak fitting of the C 1s spectrum, as this provides chemical information that is superior to that provided by other characteristic spectra of polymer constituents. The type of information attainable in this manner is illustrated in the carbon 1s spectra of **Figure 5**. As indicated in an earlier section, the XPS valence band spectra of polymers can be extremely informative for the characterization of such materials.

### Adhesion Science

In adhesion science, the concern is often to identify the exact locus of failure between two dissimilar materials. XPS can easily identify and quantify very thin organic overlayers on metallic substrates and it is widely used in adhesion studies. In addition, the manner in which aggressive species (such as water or ions) that bring about adhesion failures aggregate at or near the interface can be monitored.

### Microelectronics and Other Areas of Application

In addition to the brief examples quoted above, XPS is widely employed in many other areas including, among others, microelectronics and thin film technology, catalysis, tribology, and oxidation.

*See also:* **Mass Spectrometry:** Liquid Secondary Ion Mass Spectrometry. **Polymers:** Natural Rubber; Synthetic. **Surface Analysis:** Auger Electron Spectroscopy; Secondary Ion Mass Spectrometry of Polymers.

### Further Reading

- Artyushkova K and Fulghum JE (2003) XPS imaging. In: Briggs D and Grant J (eds.) *Surface Analysis by Auger and X-Ray Photoelectron Spectroscopy*, ch. 24, pp. 677–704. Chichester: IM Publications.
- Beamson G and Briggs D (1992) *High Resolution XPS of Organic Polymers. The Scienta ESCA 300 Database*. Chichester: Wiley.
- Briggs D and Seah MP (eds.) (1991) *Practical Surface Analysis by Auger and X-Ray Photoelectron Spectroscopy*. Chichester: Wiley.
- Drummond IW (2003) XPS: instrumentation and performance. In: Briggs D and Grant J (eds.) *Surface Analysis by Auger and X-Ray Photoelectron Spectroscopy*, ch. 5, pp. 117–144. Chichester: Im Publications.
- Köver L (2003) Chemical effects in XPS. In: Briggs D and Grant J (eds.) *Surface Analysis by Auger and X-Ray Photoelectron Spectroscopy*, ch. 16, pp. 421–464. Chichester: IM Publications.
- Moulder JF, Stickle WF, Sobol SE, and Bomben KD (1992) *Handbook of X-Ray Photoelectron Spectroscopy*. Perkin-Elmer Corp., Physical Electronics Division, Eden Prairie, USA.
- Tougaard S (1998) Accuracy of the non-destructive surface nano-structure quantification technique based on analysis of the XPS or AES peakshape. *Surface Interface Analysis* 26: 249–269 and [www.quases.com](http://www.quases.com)
- Tougaard S (2003) Quantification of nano-structures by electron spectroscopy. In: Briggs D and Grant J (eds.) *Surface Analysis by Auger and X-Ray Photoelectron Spectroscopy*, ch. 12, pp. 295–344. Chichester: IM Publications.

## Particle-Induced X-Ray Emission

**E Koltay**, Institute of Nuclear Research, Debrecen, Hungary

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by R D Vis<sup>†</sup>, pp. 4965–4973, © 1995, Elsevier Ltd.

<sup>†</sup>Deceased.

### Introduction

Particle-induced X-ray emission (PIXE) is a technique for elemental analysis using a beam of high-energy particles as probe and characteristic X-rays of the elements as analytical signal. The value of the technique was first demonstrated at the Lund Institute of Technology (Sweden) in 1970. Since then, PIXE has grown to maturity, a growth that can be

estimated by a surface-sensitive method such as XPS. This arises from the fact that if a metal electrode is polarized cathodically in an aqueous solution it will preferentially retain cations on removal, which are identified in the electron spectrum. The reverse is true of anodically polarized electrodes.

### Polymer Analysis

XPS provides a powerful method for the surface analysis of polymers and is employed routinely for this purpose in many laboratories around the world. All the constituent elements of organic polymers, with the exception of hydrogen, are detectable and have chemical shifts that indicate their chemical environment. An example of the segregation of a diblock copolymer during annealing is shown in **Figure 9** where the quantitative nanostructure of the PDMS segregation to the surface is determined by analysis of the peak shape. Combination of XPS with time-of-flight SIMS, which provides molecular specificity, improves the information attainable. The main advantage of XPS for many polymer applications is, however, obtained from the peak energy and peak fitting of the C 1s spectrum, as this provides chemical information that is superior to that provided by other characteristic spectra of polymer constituents. The type of information attainable in this manner is illustrated in the carbon 1s spectra of **Figure 5**. As indicated in an earlier section, the XPS valence band spectra of polymers can be extremely informative for the characterization of such materials.

### Adhesion Science

In adhesion science, the concern is often to identify the exact locus of failure between two dissimilar materials. XPS can easily identify and quantify very thin organic overlayers on metallic substrates and it is widely used in adhesion studies. In addition, the manner in which aggressive species (such as water or ions) that bring about adhesion failures aggregate at or near the interface can be monitored.

### Microelectronics and Other Areas of Application

In addition to the brief examples quoted above, XPS is widely employed in many other areas including, among others, microelectronics and thin film technology, catalysis, tribology, and oxidation.

*See also:* **Mass Spectrometry:** Liquid Secondary Ion Mass Spectrometry. **Polymers:** Natural Rubber; Synthetic. **Surface Analysis:** Auger Electron Spectroscopy; Secondary Ion Mass Spectrometry of Polymers.

### Further Reading

- Artyushkova K and Fulghum JE (2003) XPS imaging. In: Briggs D and Grant J (eds.) *Surface Analysis by Auger and X-Ray Photoelectron Spectroscopy*, ch. 24, pp. 677–704. Chichester: IM Publications.
- Beamson G and Briggs D (1992) *High Resolution XPS of Organic Polymers. The Scienta ESCA 300 Database*. Chichester: Wiley.
- Briggs D and Seah MP (eds.) (1991) *Practical Surface Analysis by Auger and X-Ray Photoelectron Spectroscopy*. Chichester: Wiley.
- Drummond IW (2003) XPS: instrumentation and performance. In: Briggs D and Grant J (eds.) *Surface Analysis by Auger and X-Ray Photoelectron Spectroscopy*, ch. 5, pp. 117–144. Chichester: Im Publications.
- Köver L (2003) Chemical effects in XPS. In: Briggs D and Grant J (eds.) *Surface Analysis by Auger and X-Ray Photoelectron Spectroscopy*, ch. 16, pp. 421–464. Chichester: IM Publications.
- Moulder JF, Stickle WF, Sobol SE, and Bomben KD (1992) *Handbook of X-Ray Photoelectron Spectroscopy*. Perkin-Elmer Corp., Physical Electronics Division, Eden Prairie, USA.
- Tougaard S (1998) Accuracy of the non-destructive surface nano-structure quantification technique based on analysis of the XPS or AES peakshape. *Surface Interface Analysis* 26: 249–269 and [www.quases.com](http://www.quases.com)
- Tougaard S (2003) Quantification of nano-structures by electron spectroscopy. In: Briggs D and Grant J (eds.) *Surface Analysis by Auger and X-Ray Photoelectron Spectroscopy*, ch. 12, pp. 295–344. Chichester: IM Publications.

## Particle-Induced X-Ray Emission

**E Koltay**, Institute of Nuclear Research, Debrecen, Hungary

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by R D Vis<sup>†</sup>, pp. 4965–4973, © 1995, Elsevier Ltd.

<sup>†</sup>Deceased.

### Introduction

Particle-induced X-ray emission (PIXE) is a technique for elemental analysis using a beam of high-energy particles as probe and characteristic X-rays of the elements as analytical signal. The value of the technique was first demonstrated at the Lund Institute of Technology (Sweden) in 1970. Since then, PIXE has grown to maturity, a growth that can be



followed by surveying the proceedings of the triannually organized PIXE conferences as well as those of the regular International Conferences of Nuclear Microbeam Technology and Applications.

Although frequently used as a bulk technique, PIXE is surface sensitive owing to the limited range of the impinging ions in matter. The availability of dedicated accelerators producing ions in the MeV region and, even more important, the development of microprobes with ion beams focused down to below 1  $\mu\text{m}$ , give PIXE a secure place among the various microanalytical techniques. It is important to note that if MeV ions are incident on a specimen, multi-signal detection will lead to a more complete characterization of that specimen. PIXE is therefore very often used in combination with methods such as Rutherford backscattering and nuclear reaction analysis.

This article describes the potential of PIXE, its properties, and figures of merit, and a number of illustrative applications.

## Theory

PIXE relies on the Coulomb interaction between the incoming ions and inner-shell electrons in order to create vacancies in the K-shell (or for high-atomic number elements the L-shell), after which characteristic X-rays are emitted. The cross-section of the process of inner-shell ionization was initially described in the plane-wave Born approximation (PWBA) or a binary encounter model, but recently almost exclusively the so-called ECPSSR treatment is in use. This theory is a modification of the PWBA description and includes the deflection and velocity change of the projectile owing to the nuclear Coulomb field ( $C$ ), the perturbation of the atomic stationary states (PSS) by the projectile, relativistic effects ( $R$ ), and energy loss ( $E$ ) during the collision. These corrections are treated as a series of modifications to the projectile energy and the binding energy (using effective energies) of the inner-shell electron involved.

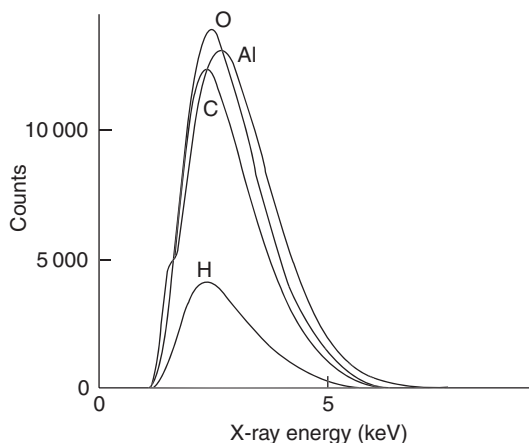
The projectile energy of interest for PIXE is commonly considered as being between 1 and 4 MeV for protons. Although this is not the energy where the cross-section is at its maximum (which is the case if the velocity of the projectile matches the velocity of the orbiting electron), these energies provide the highest peak-to-background ratio. This is because of the low values of cross-sections for nuclear events, which occur simultaneously with inner-shell ionization, and because of limitations on the maximum energy that can be transferred to electrons. This

energy transfer is important, as the major contribution to the background in PIXE spectra is the continuum bremsstrahlung X-radiation emitted during the deceleration of secondary electrons. To shift this continuum to lower energy values it is necessary to lower the incident particle energy that leads to reduced inner-shell ionization cross-sections. There is therefore a compromise between a reasonable cross-section for inner-shell ionization and an acceptable background at the position in the spectrum where the analytical signal is expected. The bremsstrahlung background is also dependent on the composition of the matrix; higher-atomic number elements with a higher electron density will give rise to both an increased background level and a shift toward higher energies owing to the collisions of protons with bound electrons. In **Figure 1**, typical backgrounds are given for a number of elements irradiated with 2.5 MeV protons. The low-energy cutoff is caused by the filtering of low-energy X-rays by the detector window. The proton energy for which the best detection limits (or the best signal-to-noise ratios) can be expected has been calculated by several authors and lies between 1 and 3 MeV for protons, the most frequently used projectile. ECPSSR cross-sections for particles of this energy are extensively tabulated for PIXE users.

The yield of characteristic X-rays for a given element  $Z$  is proportional to the amount of the element  $Z$  and can be written as

$$Y(Z) = \frac{N_p M(Z)}{S} \frac{\sigma_Z(E_0) \omega_Z b_Z \epsilon_Z N_0}{A_Z} \quad [1]$$

where  $M(Z)$  and  $A_Z$  are the mass present and the relative atomic mass of the element,  $\sigma_Z(E_0)$  is the



**Figure 1** Typical background spectra for a number of common matrix elements for 1  $\mu\text{C}$  accumulated charge of protons using energy-dispersive detection with a Si(Li) detector.

ionization cross-section (in  $\text{cm}^2$ ) for proton energy  $E_0$ ,  $\omega_Z$  is the K- or L-fluorescence yield,  $b_Z$  is the fraction of X-rays that appear in the  $K_\alpha$  or  $L_\alpha$  line,  $\varepsilon_Z$  is the absolute detection efficiency including the absorption in windows of the scattering chamber and the detector,  $N_p$  is the number of protons incident on the sample, and  $N_0$  is Avogadro's number.  $S$  is the area of the beam spot (in  $\text{cm}^2$ ). Equation [1] describes the situation of a homogeneous beam that envelops the sample. If the beam only covers a part of the specimen, it is customary to express the analyte in  $\text{g cm}^{-2}$  (area density).  $M(Z)$  in eqn [1] is then replaced by  $M_a(Z) \times S$ , where  $M_a$  is the area density. This has the advantage that  $S$  cancels out. The same formula is applicable whether the beam envelops the specimen or not.

Equation [1] only holds for thin specimens. Thin means that (1) the energy loss of the proton beam is so small that one can consider  $\sigma_Z(E)$ , which is energy dependent, as being constant; and (2) one can neglect absorption of X-rays in the specimen on their way to the detection system. As an example, for biological sections, often a few micrometers thick, proton energy loss is sufficiently small so that eqn [1] applies. The self-absorption depends on the analyte; for light analyte elements correction procedures are needed. For thin specimens it should be possible to perform absolute analysis with accuracy within 10% provided that the overall efficiency of the detector,  $\varepsilon_Z$ , is adequately calibrated and the accumulated charge,  $N_p$ , is monitored properly.

If the specimens are thicker, correction algorithms are necessary. In eqn [1],  $\sigma_Z(E_0)$  has to be replaced by the integral over this cross-section from the incoming ( $E_0$ ) to the outgoing proton energy (or, if specimens are thicker than the range of the projectiles, to 0). Moreover, for the outgoing X-rays, the attenuation as a function of depth of their origin has to be calculated. These calculations are invariably done numerically by dividing the specimen into a number of layers. For each layer, eqn [1] is applied with the cross-section for the proton energy in that layer and with an extra attenuation term to take into account self-absorption. Several computer programs are available to do these calculations. These calculations, however, can only give accurate results if one has a detailed knowledge of the matrix composition, because this composition determines both the energy loss of the beam and the attenuation of X-rays in the specimen. If this composition is not known, RBS analysis performed simultaneously can be very helpful. The accuracy of thick target analysis is less than for thin specimens because of uncertainties in the database needed for these correction procedures. Frequently, standards with comparable matrix

composition as the unknowns are used to improve accuracy.

The minimum detection limit (MDL) also depends on target thickness. If the MDL is defined in the usual way as being the concentration for which the corresponding X-ray line intensity (yield) exceeds three standard deviations of the background in that spectral region, it can be shown that for a thin specimen the MDL is proportional to

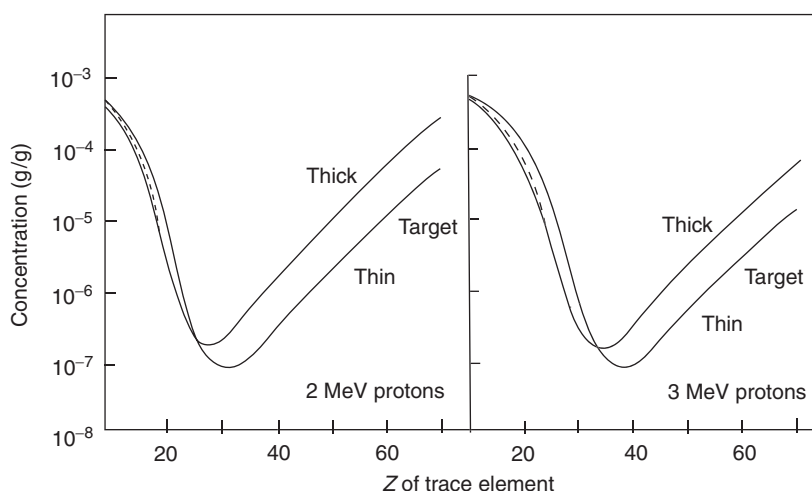
$$\frac{\sigma_B^{1/2}}{\sigma_Z} \left( \frac{\text{FWHM}}{N_p \varepsilon_Z} \right)^{1/2} \quad [2]$$

in which  $\sigma_B$  is the production cross-section of background in the spectral region of the characteristic peak of element Z, FWHM is the full-width at half-maximum of this peak, and  $N_p$  is the number of protons incident on the target.

Other definitions of MDL are sometimes used. For ease of calculation some authors define the requirement that the counting rate of the peak is equal to the counting rate of the background, which results in a peak-to-background ratio of 1. In Figure 2, the results of calculations of MDL for a carbon matrix are given for both thin and thick targets. These results hold for energy-dispersive detection with a Si(Li) detector. Figure 2 illustrates well the main features of the MDL, showing a minimum around  $Z=30$ , a minimum that is shifted upwards by increasing the proton energy. It illustrates the advantages of using a thin specimen, especially for heavier analyte elements. Also shown is that in optimal conditions, MDL values are  $0.1 \mu\text{g per g}$ . It should be noted here that absolute MDL values can be very low indeed since the irradiated sample mass, especially with well-focused proton beams, can be as low as a few nanograms, leading to absolute values of MDL in the femtogram region. It is also worth noting that research with other ion beams, especially beams of  $\alpha$ -particles, has demonstrated that at  $\sim Z=20$  lower relative MDL values ( $10^{-2} \mu\text{g per g}$ ) can be obtained.

## Instrumentation

PIXE instruments are normally divided into two groups, those instruments suitable for macro-PIXE with a beam with cross-section of several  $\text{mm}^2$ , and those suitable for micro-PIXE with a beam spot of  $\sim 1 \mu\text{m}^2$  cross-section. The obvious requirement for both PIXE facilities is the availability of an ion accelerator capable of producing beams of a few megaelectronvolts. In the early days of PIXE, parasitic use of accelerators purchased for research in nuclear physics was common. The majority of these



**Figure 2** Detection limits as a function of  $Z$  for two proton energies and for thin and thick specimens. The minimum detection limit is defined here as the concentration giving a peak-to-background ratio of 1. The calculations are done for a carbon matrix. (Reprinted with permission from Vis RD (1985) *The Proton Microprobe: Applications in the Biomedical field*. Boca Raton, FC: CRC Press; © CRC Press.)

machines were of the Van de Graaff type; at a few places cyclotrons were employed for PIXE. The mainstream research in nuclear physics shifted toward higher-energy machines and nowadays at several institutes accelerators are fully dedicated for ion beam analysis. Recently, a few laboratories have acquired modern electrostatic accelerators, especially produced for analytical work. These machines feature a high stability and a good energy resolution of the beam and are more user-friendly. They are often of the tandetron type, with a solid-state diode stack to provide the high potential, or of the pelletron type, in which a chain of pellets rather than a belt carries the charge toward the terminal. Both single-ended and tandem machines are in use.

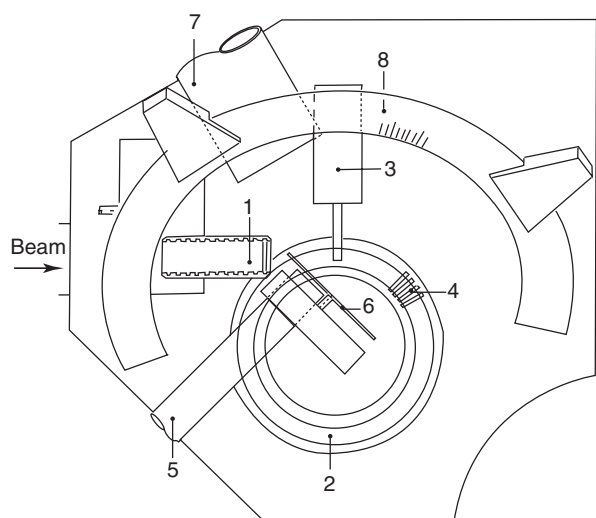
### Macro-PIXE

For macro-PIXE, the accelerator has to be equipped with a beam-guiding system to transport the beam to the sample chamber. This beam transport system encompasses one or more bending magnets, focusing quadrupole magnets, and beam steerers for minor corrections. The whole system should be evacuated to a pressure below 1 MPa. In **Figure 3A**, the various features of a PIXE chamber are illustrated. The PIXE chamber has (1) diaphragms to collimate the beam to the desired dimension; (2–4) a sample changer; (5) an X-ray detector; (6) absorbers to filter out low-energy X-rays in order to limit the total counting rate of the detector(s); and very often also (7), a  $\gamma$ -detector for the detection of  $\gamma$ -rays resulting from nuclear reactions on low- $Z$  elements and (8) an in-angle adjustable surface barrier detector (SBD) for RBS. The

chamber has a Faraday cup to monitor the beam. Electronics for data acquisition are standard; after amplification the signals are fed into a multichannel analyzer. **Figure 3B** shows a complete setup used for both macro- and micro-PIXE.

### Micro-PIXE

Since around 1980, nuclear microprobes, sometimes also referred to as ion microprobes (not to be confused with the low-energy ion microprobe for SIMS), have been developed. These are instruments in which the ion beam is focused to micrometer dimensions. All nuclear microprobes consist of an object slit for primary collimation, followed by one or two additional slit pairs to intercept scattered particles and to limit the divergence of the beam, and a lens system. The lens is normally a set (2–4) of high-precision magnetic quadrupoles. At the present stage of development beam spots can be produced of below  $0.5\ \mu\text{m}$  diameter with sufficient current (50–100 pA) to analyze at trace level. However, the quality of state-of-the-art microbeam optical channels can be characterized by the fact that spot sizes of  $35 \times 75\ \text{nm}$  have recently been measured in scanning transmission ion microscopy mode at a beam current as low as 10 000 protons per second. The specimen chamber for micro-PIXE is more complicated than that for macro-PIXE because in addition high-precision positioning of the specimen is needed and the sample has to be viewed by an optical microscope for tuning purposes and for proper definition of the scan area. Scanning can be performed either by moving the target or by sweeping the beam with deflection



(A)



(B)

**Figure 3** (A) Schematic view of a PIXE chamber. Key: 1, diaphragms to collimate the beam; 2, carousel for the samples; 3, lifting mechanism to position the sample in front of the beam; 4, sample holders; 5, X-ray detector; 6, wheel with different absorbers to filter out low-energy X-rays; 7,  $\gamma$ -detector; 8, protractor to position a surface barrier detector. (B) Photograph of a typical PIXE environment with a pelletron accelerator and micro- and macro-PIXE beam lines. (Reprinted with permission from Vis RD, Kramer JLAM, van Langevelde F, and Mars L (1993) The upgraded Amsterdam nuclear microprobe. *Physics Review B: Nuclear Instrument Methods* B77: 41–44; © Elsevier.)

coils. In both cases, fast electronics should correlate the detector events with the position along the scan in order to be able to construct trace element concentration maps. **Figure 4** is a schematic representation of a micro-PIXE chamber. Data storage is often done in list mode, which means that every detector event is labeled with the position on the sample from which it originated. This opens the possibility of replaying the data with a chosen lateral resolution or as a function of time to check the integrity of the sample during irradiation.

## Detectors

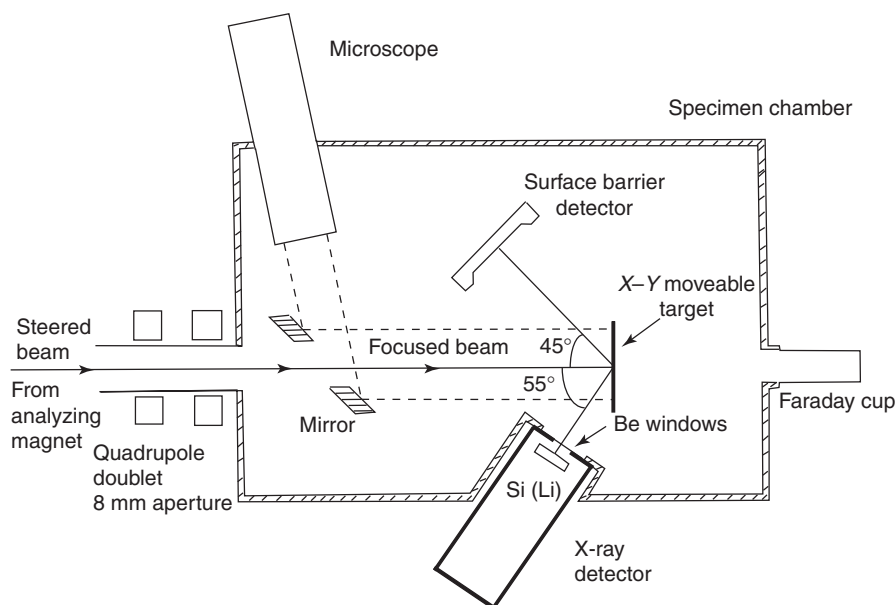
For the detection of X-rays lithium (Li)-drifted silicon (Si) detectors are generally used. The energy resolution of these detectors varies between 130 and 180 eV at 5.9 keV and is dependent on the size of the silicon crystal used. Large detectors have the advantage of a large solid angle at the price of resolution.

During PIXE analyses, the sample is normally surrounded by several detectors. **Figure 5** shows such a multidetector arrangement for full characterization of the specimen. Besides the Si(Li) detector for trace element determination by PIXE, an SBD placed backwards records the RBS spectrum. This is used to determine the matrix composition so that matrix correction can be calculated. A germanium (Ge) detector measures  $\gamma$ -radiation emitted by light sample nuclei, such as boron, fluorine, and sodium, thus providing additional information about elements not detected by the Si(Li) detector. In conventional PIXE setups, the X-rays of these light elements are absorbed in the detector window (usually 25  $\mu\text{m}$  beryllium); thinner windows are not used in such devices because protons scattered from the sample must be prevented from entering the Si(Li) detector. Still, PIXE is also a good tool for light element analysis due to its high cross-sections in the region of low atomic numbers. The technical difficulties caused at low  $Z$ -s by X-ray absorption in detector window and the presence of scattered bombarding particles entering the detector can be circumvented by the application of an X-ray detector of ultrathin ( $\sim 300$  nm) window (UTW) and a permanent magnetic circuit acting as a magnetic deflection trap, respectively. With the combined application of UTW and Be-windowed Si(Li) detectors in recent PIXE systems, complete analytical characterization of elemental constituents from carbon to uranium is possible in a single irradiation process.

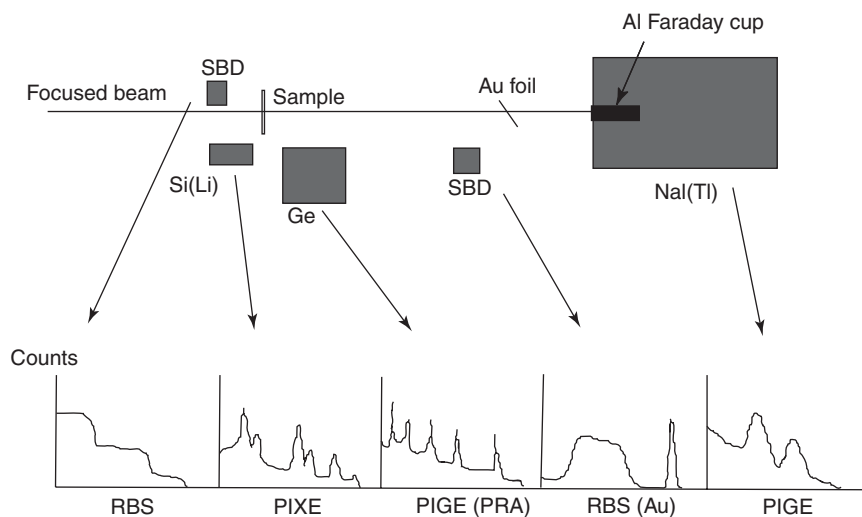
Several methods can be used to monitor the target thickness of specimens that are transparent to the beam. In **Figure 5**, protons transmitted by the sample are scattered off a thin gold foil and recorded with a second SBD. This is an elegant procedure since the accumulated charge is measured by the area of the gold peak and the target thickness by the energy of that peak. Finally, a thallium-doped sodium iodide well-type scintillator envelops an aluminum Faraday cup in which  $\gamma$ -radiation originating from inelastic scattering on aluminum is used for monitoring the beam intensity and stability.

## Applications

Macro- and micro-PIXE find applications in different types of analytical work.



**Figure 4** Schematic view of a specimen chamber used for micro-PIXE. Only two detectors are indicated – a Si(Li) for PIXE and a SBD for RBS.



**Figure 5** A multidetector setup used in combination with a nuclear microprobe. Types of spectra obtained are indicated. PIGE, particle-induced  $\gamma$ -ray emission; and PRA, prompt radiation analysis. See text for further explanation.

### Macro-PIXE

From the earliest days of PIXE, biological applications have attracted much attention. The role of essential trace elements and correlations between the concentrations of these elements and different types of diseases have been widely studied. The more recent interest in environmental problems led to an increase in the use of PIXE. Since PIXE is an almost ideal technique for the multielement analysis of aerosols filtered out of the atmosphere, many

laboratories embarked into programs on environmental pollution. The filters or other collection surfaces from various types of samplers can be placed in the beam without any pretreatment, and the subsequent PIXE analysis can be done very quickly and, if necessary, can be fully automated in order to obtain large datasets that are of statistical significance. To an increasing extent, geological and archaeological samples have been analyzed by PIXE, including objects of art. However, other analytical techniques have been refined to give better detection limits,

higher accuracy, etc., and therefore macro-PIXE now has a number of serious competitors. Contrary to the earlier broad scope of biological applications of PIXE, biologists will nowadays very often use in-house equipment based on competitive methods for their analytical needs. These have made the biological applications of PIXE less attractive than in the early days. Present-day biological applications of PIXE are concentrated on cases, where its advantage of being able to perform a full microelemental analysis on a very small amount of sample material (mg) is of primary importance. The majority of biological and biomedical applications is based on the micro-PIXE technique. As an illustration of biological work, Figure 6 shows a PIXE spectrum for the former NBS 1571 orchard leaves standard reference material, together with an indication of the elements and their concentrations. The nondestructive nature of PIXE means that there is no need for laborious destruction of the specimens. This is very important in a number of fields, among others in studying geological and archaeological samples.

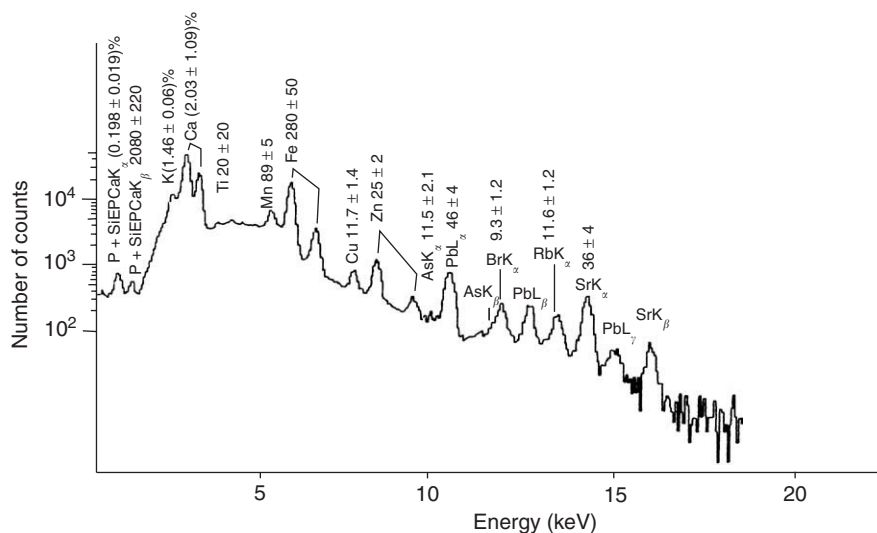
PIXE finds special fields of applications in geological investigations. Of special interest is the study of the contribution of arid regions to the global aerosol burden. PIXE analyses performed on volcanic samples from the neighborhood of a volcano and from its emission plume contribute to a more complete characterization of the processes governing the eruption in its different phases. The method finds applications in research in mineralogy and petrology, geogas analysis, as well in prospecting for raw material. As possible fields of PIXE applications measuring

sea-air exchange of materials, dating of ice cores in ice sheets can also be mentioned.

Aerosol work remains a major and convincing application of macro-PIXE. Many PIXE workers are nowadays specialists in aerosol physics and chemistry. The method plays a role in global and regional air quality monitoring through time dependent and sequential measurements of aerosol, in the separation of natural and anthropogenic components, and in source apportionment. Low trace element detection limits make the method applicable in studying aging and transformation of primary aerosols together with transport and deposition phenomena. Regional signatures deduced from PIXE datasets can be used as tracers in following long-range transport tracks over thousands of kilometers.

In environmental problems PIXE finds applications in acid-rain investigations, river water analysis, monitoring the deposition of marine and lake sediments, and the mixing of salt and fresh water in coastal regions. Location and qualification of sources contaminating soils and agricultural products can be performed with PIXE analysis. It is a versatile tool both in fundamental research on the processes in the respiratory tracts and in respiratory health survey aiming at guaranteed occupational safety in work environment.

Research in art and archeology uses PIXE as an analytical tool in the fields of early development of pottery, metallurgy, and glassmaking, in numismatics, in studying painting material and color additives, paper-like objects and manuscripts, and in provenance studies of artifacts.



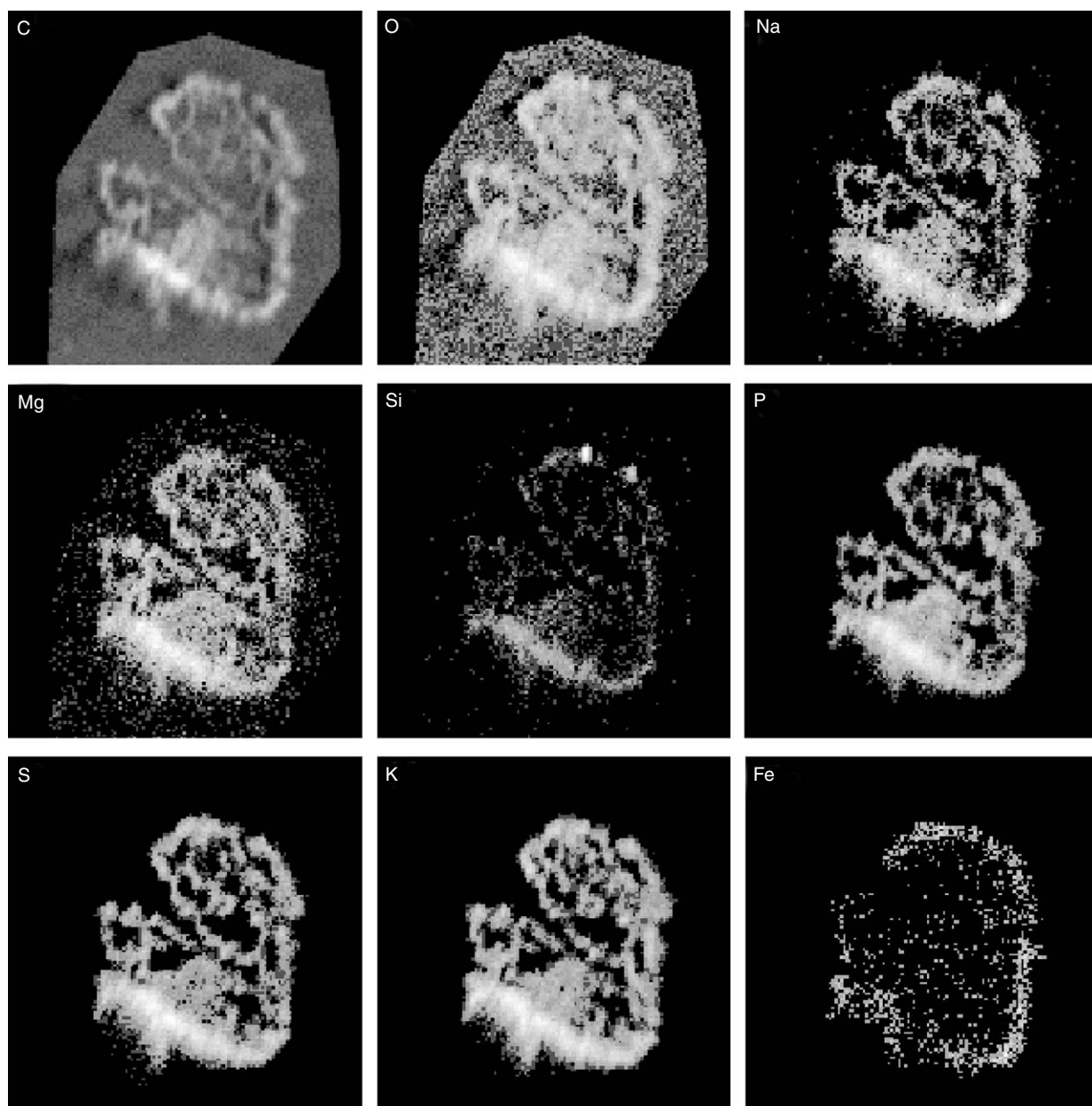
**Figure 6** The PIXE spectrum of orchard leaves (NBS 1571), showing the concentrations measured ( $\mu\text{g/g}$ , unless indicated otherwise).  $E_p = 3.0$  MeV;  $I = 42$  nA;  $t = 2350$  s; absorber =  $200\ \mu\text{m}$  PTFE. (Reprinted with permission from Vis RD (1985) *The Proton Microprobe: Applications in the Biomedical field*. Boca Raton, FC: CRC Press; © CRC Press.)



### Micro-PIXE

The development of nuclear microprobes brought about a strong revival of PIXE as an analytical method. The capability of PIXE to analyze on a micrometer scale at trace levels in a quantitative way has meant that there is little competition. Moreover, as mentioned before, the combination of methods possible with a megaelectronvolt ion beam focused down to  $<1\mu\text{m}$  offers unique analytical possibilities.

In biology, the ability to measure at the cellular level, as has been shown in an increasing number of papers, finds recent interest. The production of elemental maps for biological sections of both hard and soft tissues has provided new information about the role of essential trace elements and on biological processes like metal uptake by microorganisms, the appearance of elemental gradients across membranes, matter exchange between symbiotic partners, physiological functions and root structure of plants. As an example, in **Figure 7**, elemental maps



**Figure 7** Micro-PIXE elemental maps for the distribution of C, O, Na, Mg, Si, P, S, K, and Fe in root of plant rape (*Brassica napus*) (size of scanned area:  $500 \times 500\mu\text{m}$ ). (Reproduced with permission from Kertész Zs, Uzonyi I, and Kiss ÁZ (2003).)

taken in the Debrecen (Hungary) micro-PIXE beam-line on root sections of heavy metal accumulating rape plant (*Brassica napus*) are shown. The combined use of UTW and Be-windowed X-ray detectors made the mapping of elements of low atomic number such as C, O, Na, Mg together with higher Z elements Si, P, S, K, Fe shown here also possible. In the biomedical field hair has been studied extensively with micro-PIXE in order to find correlations between trace elements in hair and blood or tissue; radial scans across a single hair can distinguish between the different routes followed during the incorporation of trace elements in hair, including environmental exposure. Longitudinal scans, on the other hand, provide information in time. Light can be thrown on the role of trace elements in physiological and pathological processes on a cellular scale. Degeneration mechanisms in aging process, structural changes in calcified tissues, and wearing processes in oral and orthopedic implants can be followed by micro-PIXE measurements, as well.

Aerosol work has also profited from micro-PIXE, as single particle analysis has become feasible. Accurate correlation between the chemical composition of such particles and their size and shape provides information about the sources of the particles and about the transformations that take place during atmospheric transport. In other environmental studies, micro-PIXE has been used to measure the uptake of trace elements in plants, roots, and trees. In particular, tree rings, which also record time information, have been analyzed in several laboratories.

In materials science, micro-PIXE is used, among other applications, in microelectronics analysis, to study insulating materials, binary alloys, surface layers, and multilayers, in locating dopant atoms in crystals through PIXE channeling, to monitor the homogeneity of ceramic superconducting materials, to study segregation in metals, and to analyze the presence and distribution of trace metals in polymers. Catalysts are also studied with micro-PIXE, in particular the diffusion of the catalytically active metal species into the supporting material. The analysis of impurities deposited on the walls of nuclear fusion devices can be mentioned as a special field of applications, too.

In geology, there is such a wide variety of applications that a few laboratories with a microprobe use their accelerator exclusively for this type of research. Fundamental geological studies, not only work on fluid inclusions, zonation in several minerals, etc., but also the search for precious metals in ores, are issues amenable to micro-PIXE. Extraterrestrial material has also been investigated with the nuclear microprobe. The lateral resolution and the

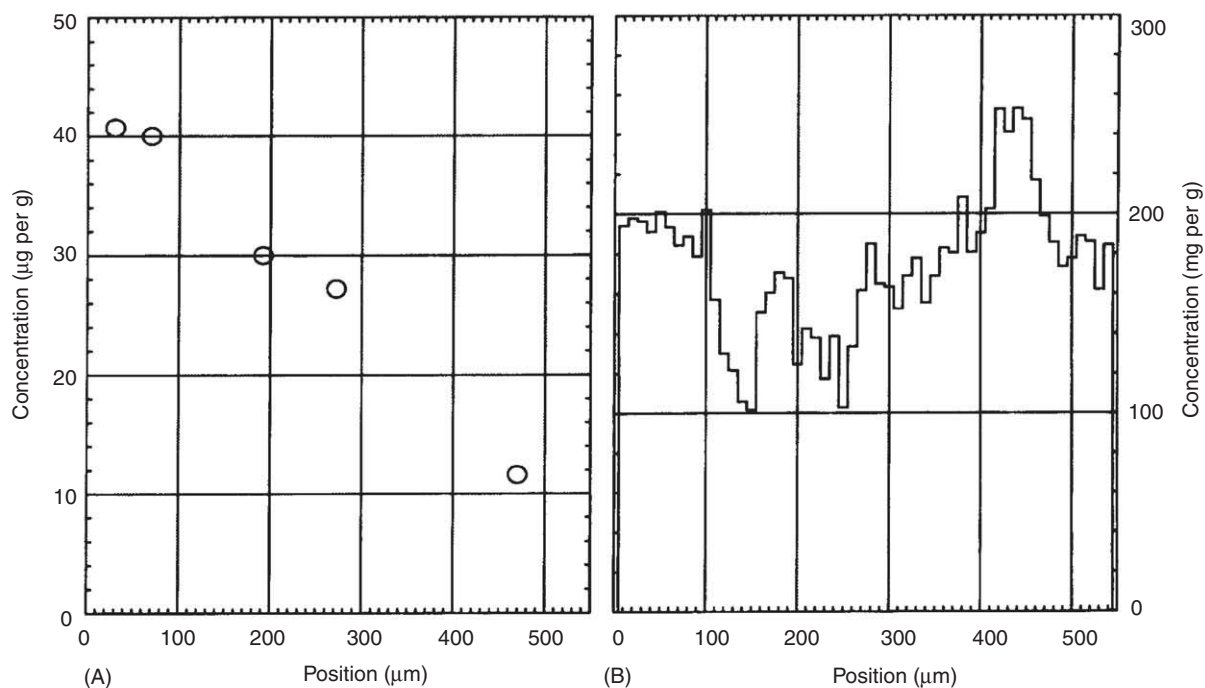
multielement character of micro-PIXE offer new information for cosmology and cosmochemistry on the composition and morphology of meteorites and cosmic dust, on origin of microspherules, and on trace element constituents in lunar rocks. Scans across typical inclusions, called chondrules, in ordinary chondrites, a class of stony meteorites, reveal the distribution of trace elements. A study of the distribution may shed light on the cooling history and sequence of solidification of minerals in such objects. As an example, **Figure 8** gives the results of a line scan over a chondrule. Zinc is the element of interest in this instance. Zinc is a volatile element and is present in such a low concentration that only spot measurements are done. The distribution of the major element silicon is also given in **Figure 8**. The micrograph shows the scan, made visible after the measurements by deposition of carbon. The chondrule can be seen on the right-hand side. The results clearly show a depletion of zinc in the chondrule as compared with the surrounding matrix.

Micro-PIXE offers a powerful nondestructive analytical method of micrometer-scale lateral resolution for research in art and archeology. Among other fields of application studies in corrosion processes, searching for characteristic features of painting and fabrication techniques, identification of art objects through trace element qualifications, and testing of restoration methods can be mentioned here.

In addition to the applications mentioned above, a wide variety of lesser activities in all branches of science are studied using PIXE. In all cases where it is beneficial to use a trace element analytical technique that combines a lateral resolution of  $\sim 1\text{ }\mu\text{m}$  with trace element detection limits ( $1\text{--}10\text{ }\mu\text{g g}^{-1} \equiv 10^{-15}\text{--}10^{-16}\text{ g}$ ), micro-PIXE is an option well worth considering.

## PIXE for Surface Analysis

As stated above, PIXE is surface sensitive in the sense that the limited range of the incoming ions and the decrease of the contribution to the analytical signal from deeper layers means that most information originates from the outermost  $10\text{ }\mu\text{m}$ . Nevertheless, the examples of applications given above are not real examples of the analysis of the very surface (outermost atomic layers) or of an interface. In contrast to the resonances in the cross-sections for nuclear reactions, which are therefore used for high-resolution depth profiling within the range of the bombarding particles, PIXE cross-sections have a smooth behavior with depth. Still, special deconvolution algorithms to be applied on a series of experiments with incident



**Figure 8** Results of a line scan over the matrix–chondrule interface shown in the micrograph (C). The chondrule is the whitish area on the right-hand side. The scan is visible as a black mark because of carbon deposition and was 0.55 mm long. Results are given for zinc (A) and silicon (B). Only spot measurements were feasible for zinc because of the low concentration present.

bombarding energies, varied over the appropriate energy interval, offer a technique for deducing depth distribution curves of the elements detected in PIXE spectra. PIXE's high sensitivity, however, enables

detection of very small contaminations or a very thin surface layer (a few atomic layers) on a suitable substrate, especially if the analyte has a higher  $Z$  value than the substrate. Heavier incident particles are even more

suitable here because of their much higher stopping power.

Even higher sensitivity is obtained if the ion beam impinges on the specimen at a glancing angle. This lengthens the path traveled through the surface layer of interest and also decreases the bremsstrahlung background owing to reflection of ions at the surface, resulting in lower production of secondary electrons. PIXE used in this mode is a true surface technique but a price is paid in terms of reduced lateral resolution because of the small angle between the beam and the target surface.

Other developments include improvements in detector technique to such an extent that, besides information about the elemental composition, chemical information also become available. The replacement of commonly used energy-dispersive Si(Li) X-ray spectrometer with high-resolution wavelength dispersive bent crystal spectrometer in combination with proportional counters opens a new field of applications for PIXE by resolving the pattern of satellite lines that is dependent on the chemical environment of the analyte atom. In such investigations beams of multiple charged light heavy ions are preferentially used. The underlying physical processes and the quantitative features of the fine structures in the spectra to be used in chemical state analysis are the subjects of recent investigations. Instrumental improvements aim towards sub-micrometer lateral resolution with appropriate beam intensity. The fact that, unlike electrons, incoming ions show virtually no scatter away from their path through the specimens helps in maintaining this extreme lateral resolution during real analyses.

See also: **Air Analysis:** Sampling; Outdoor Air; Workplace Air. **Archaeometry and Antique Analysis:** Dating

of Artifacts; Metallic and Ceramic Objects; Art and Conservation. **Forensic Sciences:** Hair. **Surface Analysis:** Nuclear Reaction Analysis and Elastic Recoil Detection. **X-Ray Absorption and Diffraction:** Overview. **X-Ray Fluorescence and Emission:** Particle-Induced X-Ray Emission.

## Further Reading

- Johansson SAE and Campbell JL (1988) *PIXE: A Novel Technique for Elemental Analysis*. Chichester: John Wiley.
- Johansson SAE, Campbell JL, and Malmqvist KG (1995) In: Winefordner JB (ed.) *Particle Induced X-Ray Emission Spectroscopy (PIXE)*, vol. 133 of Chemical Analysis: A Series of Monographs on Analytical Chemistry and its Applications. Chichester: Wiley.
- Proceedings of the International Conferences on PIXE and its Analytical Applications. *Nuclear Instrument Methods*: Johansson SAE (ed.) 142 (1977); Johansson SAE (ed.) 181 (1981); Martin B (ed.) B3 (1984); van Rinsvelt H, Bauman S, Nelson JW, and Winchester JW (eds.) B22 (1987); Vis RD (ed.) B49 (1990); Uda M (ed.) B75 (1993); Moschini G and Valkovic V (eds.) B109/110 (1996); Malmqvist KG (ed.) B150 (1999); Campbell JL (ed.) B189 (2002).
- Proceedings of the International Conferences on Nuclear Microbeam Technology and Applications. *Nuclear Instrument Methods*: Grime GW and Watt F (eds.) B30 (1987); Legge GJF and Jamieson DN (eds.) B54 (1991); Lindh U (ed.) B77 (1993); Yang F, Tang J, and Zhu J (eds.) B104 (1995); Doyle BL, Maggiore CJ, and Bench G (eds.) B130 (1997); Prozesky VM, Przybylowicz WJ, and Pineda A (eds.) 158B (1999); Moretto Ph and Bonnin-Mosbah (eds.) B181 (2001); Takai M and Kamiya T (eds.) B210 (2003).
- Watt F and Grime GW (eds.) (1987) *Principles and Applications of High-Energy Ion Microbeams*. Bristol: Adam Hilger.

## Auger Electron Spectroscopy

**W E S Unger**, Federal Institute for Materials Research and Testing (BAM), Berlin, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by G E McGuire, pp. 4907–4912, © 1995, Elsevier Ltd.

## Introduction

The Auger electron emission phenomenon was revealed by P. Auger in 1925 when he observed

nonradiative relaxation of excited ions in a cloud chamber. L.A. Harris observed Auger electron emission examining solid surfaces in 1968. Since then the spectroscopic method utilizing Auger electron emission has developed rapidly and is in widespread use today. In this review of Auger electron spectroscopy (AES), the basic principles of the technique will be described, a description of a modern instrument will be given, examples of modes and applications of AES will be provided, and trends in the use of AES in surface and thin film analysis will be reviewed.

suitable here because of their much higher stopping power.

Even higher sensitivity is obtained if the ion beam impinges on the specimen at a glancing angle. This lengthens the path traveled through the surface layer of interest and also decreases the bremsstrahlung background owing to reflection of ions at the surface, resulting in lower production of secondary electrons. PIXE used in this mode is a true surface technique but a price is paid in terms of reduced lateral resolution because of the small angle between the beam and the target surface.

Other developments include improvements in detector technique to such an extent that, besides information about the elemental composition, chemical information also become available. The replacement of commonly used energy-dispersive Si(Li) X-ray spectrometer with high-resolution wavelength dispersive bent crystal spectrometer in combination with proportional counters opens a new field of applications for PIXE by resolving the pattern of satellite lines that is dependent on the chemical environment of the analyte atom. In such investigations beams of multiple charged light heavy ions are preferentially used. The underlying physical processes and the quantitative features of the fine structures in the spectra to be used in chemical state analysis are the subjects of recent investigations. Instrumental improvements aim towards sub-micrometer lateral resolution with appropriate beam intensity. The fact that, unlike electrons, incoming ions show virtually no scatter away from their path through the specimens helps in maintaining this extreme lateral resolution during real analyses.

See also: **Air Analysis:** Sampling; Outdoor Air; Workplace Air. **Archaeometry and Antique Analysis:** Dating

of Artifacts; Metallic and Ceramic Objects; Art and Conservation. **Forensic Sciences:** Hair. **Surface Analysis:** Nuclear Reaction Analysis and Elastic Recoil Detection. **X-Ray Absorption and Diffraction:** Overview. **X-Ray Fluorescence and Emission:** Particle-Induced X-Ray Emission.

## Further Reading

- Johansson SAE and Campbell JL (1988) *PIXE: A Novel Technique for Elemental Analysis*. Chichester: John Wiley.
- Johansson SAE, Campbell JL, and Malmqvist KG (1995) In: Winefordner JB (ed.) *Particle Induced X-Ray Emission Spectroscopy (PIXE)*, vol. 133 of Chemical Analysis: A Series of Monographs on Analytical Chemistry and its Applications. Chichester: Wiley.
- Proceedings of the International Conferences on PIXE and its Analytical Applications. *Nuclear Instrument Methods*: Johansson SAE (ed.) 142 (1977); Johansson SAE (ed.) 181 (1981); Martin B (ed.) B3 (1984); van Rinsvelt H, Bauman S, Nelson JW, and Winchester JW (eds.) B22 (1987); Vis RD (ed.) B49 (1990); Uda M (ed.) B75 (1993); Moschini G and Valkovic V (eds.) B109/110 (1996); Malmqvist KG (ed.) B150 (1999); Campbell JL (ed.) B189 (2002).
- Proceedings of the International Conferences on Nuclear Microbeam Technology and Applications. *Nuclear Instrument Methods*: Grime GW and Watt F (eds.) B30 (1987); Legge GJF and Jamieson DN (eds.) B54 (1991); Lindh U (ed.) B77 (1993); Yang F, Tang J, and Zhu J (eds.) B104 (1995); Doyle BL, Maggiore CJ, and Bench G (eds.) B130 (1997); Prozesky VM, Przybylowicz WJ, and Pineda A (eds.) 158B (1999); Moretto Ph and Bonnin-Mosbah (eds.) B181 (2001); Takai M and Kamiya T (eds.) B210 (2003).
- Watt F and Grime GW (eds.) (1987) *Principles and Applications of High-Energy Ion Microbeams*. Bristol: Adam Hilger.

## Auger Electron Spectroscopy

**W E S Unger**, Federal Institute for Materials Research and Testing (BAM), Berlin, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by G E McGuire, pp. 4907–4912, © 1995, Elsevier Ltd.

## Introduction

The Auger electron emission phenomenon was revealed by P. Auger in 1925 when he observed

nonradiative relaxation of excited ions in a cloud chamber. L.A. Harris observed Auger electron emission examining solid surfaces in 1968. Since then the spectroscopic method utilizing Auger electron emission has developed rapidly and is in widespread use today. In this review of Auger electron spectroscopy (AES), the basic principles of the technique will be described, a description of a modern instrument will be given, examples of modes and applications of AES will be provided, and trends in the use of AES in surface and thin film analysis will be reviewed.

## Basic Principles and Qualitative Surface Analysis

Focused electron beams are the primary means of stimulating Auger electron emission, although, other energetic beams, e.g., X-ray photons or ions, may also be used. **Figure 1** is a schematic diagram of the Auger emission process in a metal. When an electron beam of energy,  $E_P$ , much greater than core electron binding energies of the solid,  $E_K$ , impinges on the surface, ionization occurs as depicted in the left panel of the figure due to hole formation in a core level, here the K shell. The atom is left in an excited state which decays radiatively or nonradiatively by emission of X-rays or Auger electrons, respectively. **Figure 1** shows an electron from the  $L_1$  level filling the hole created in the K shell while a second electron, the Auger electron with the kinetic energy  $E_A$ , is emitted from the  $L_{2,3}^*$  level. This results in a doubly ionized state shown in the right-hand panel of **Figure 1**. Various combinations of the discrete core levels and also valence band electrons may enter into this three-electron process resulting in series of Auger transitions with various energies characteristic of the emitter element. These are tabulated in databases, denominated, e.g., as KLL, LMM or MNN series, and often displayed in a table as in **Figure 2**. The competing X-ray emission (radiative core hole decay) influences the relative probability of relaxation by Auger electron emission.

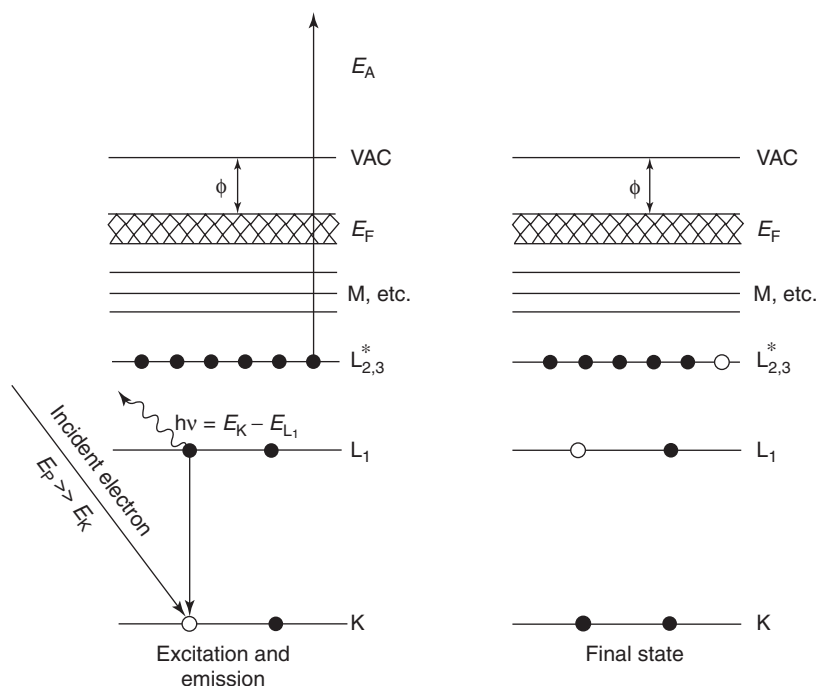
The kinetic energy of the ejected Auger electron illustrated in **Figure 1** is

$$E_{KL_1L_{2,3}} = E_K - E_{L_1} - E_{L_{2,3}}^* - \Phi_A \quad [1]$$

where  $E_K$  and  $E_{L_1}$  are the binding energies of the atomic energy levels K and  $L_1$ .  $E_{L_{2,3}}^*$  is the binding energy of the  $L_{2,3}$  level in the ion and, therefore, different from  $E_{L_{2,3}}$  for the ground state of the emitter atom.  $\Phi_A$  is the work function of the energy analyzer.

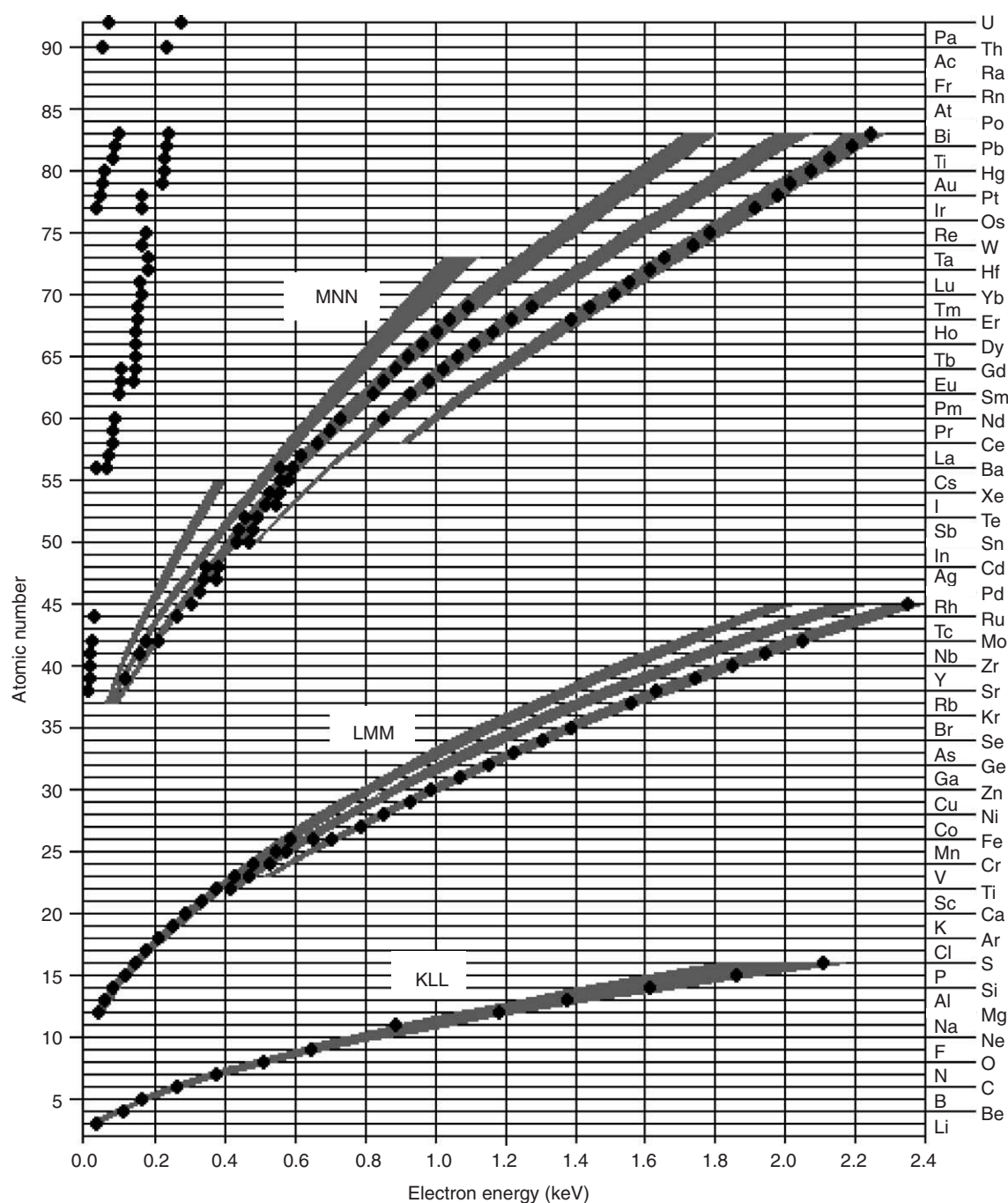
The kinetic energy of an Auger electron is characteristic for the atomic levels and, hence, for a given element, qualitative elemental analysis is enabled. Chemical shifts, similar to those observed in X-ray photoemission, may occur due to initial effects (chemical bonding) or final state effects (intra- and extra-atomic relaxation). Thus chemical bonding frequently influences the fine structure of a series of Auger transitions and the resulting chemical shifts of their kinetic energies can be utilized to derive chemical state information on the emitter atom. **Figure 3** shows highly resolved Si Auger KLL spectra of metallic Si, SiN, and SiO<sub>2</sub>, illustrating the dramatic changes of peak shapes and positions in the Auger spectrum of Si upon compound formation.

There are popular databases and handbooks that provide characteristic Auger electron spectra, beginning with lithium and ending with uranium. These spectra were obtained under controlled



**Figure 1** Schematic diagram of the Auger emission process using electron beam excitation.





**Figure 2** Survey of the most useful Auger transitions for all elements in the periodic table. Black diamonds indicate the strongest peaks usually used in practical AES applications. Gray bands indicate the kinetic energies of less intense peaks. (Reprinted from [www.eaglabs.com](http://www.eaglabs.com) with permission of the Evans Analytical Group.)

conditions. They can be used for chemical state identification by AES.

Because AES is a method of surface analysis its information depth should be considered. **Figure 4** contains the relevant information. The incident electron beam striking the specimen penetrates to a

depth of 1–3  $\mu\text{m}$ , depending on its initial kinetic energy and inelastic scattering in the solid. Within the primary excitation volume a variety of elastic and inelastic processes occur. Auger emission will occur within this volume until the energy of inelastically scattered electron falls below the threshold energy

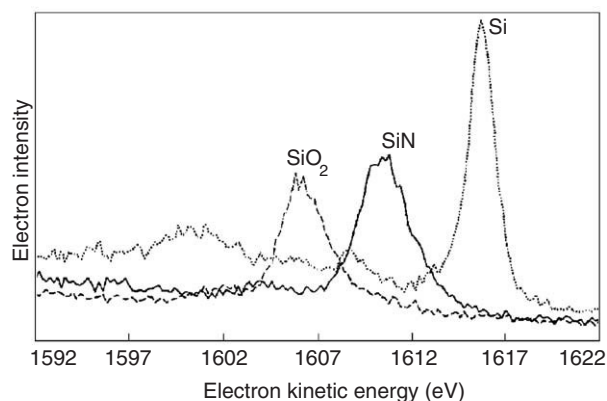
for ionization. However, most of the secondary electrons, including the Auger and photoelectrons, never reach the surface because of their very short inelastic mean free paths. Figure 5 shows the principal dependence of the electron inelastic mean free path,  $\lambda$ , on the electron energy. Therefore, in the energy range of interest in AES, from 0 to 2000 eV, the information depth is in the range of 0.5–7.5 nm, i.e., only a few monolayers, depending on the material analyzed and the individual Auger transition used.

## Quantitative Surface Analysis

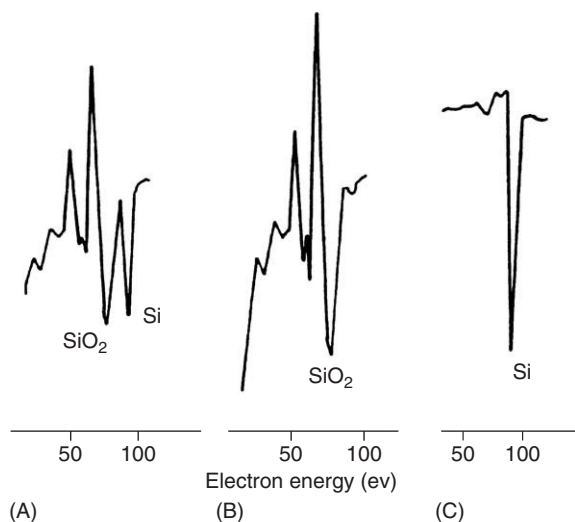
There is a well-founded understanding of the correlation of a measured Auger electron intensity with

the concentration of the respective emitter atom in a solid sample. Basic parameters involved are:

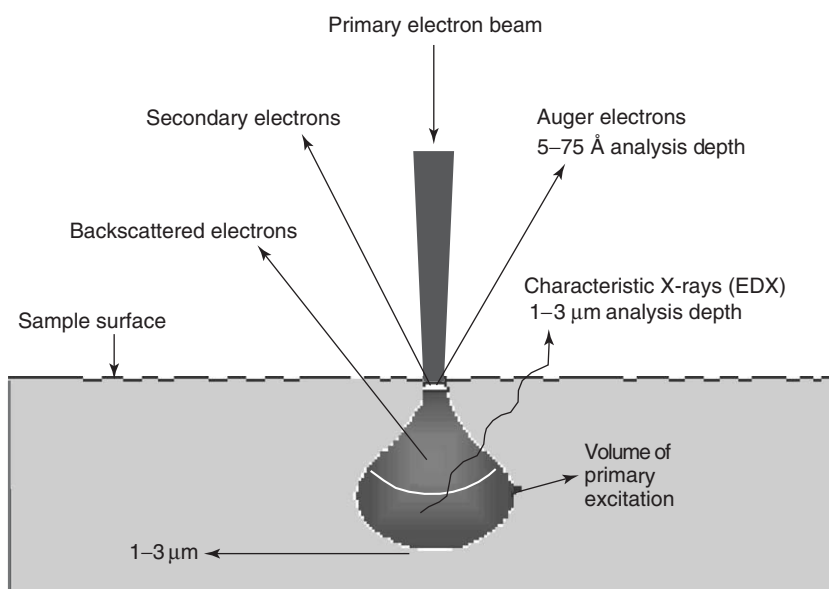
- primary electron beam current and energy;
- elemental core hole ionization cross-sections as a function of the primary electron beam energy;
- Auger transition probability;
- probability of core ionization by energetic electrons scattered in the solid (so-called backscatter effect) characteristic of the material;
- inelastic mean free path,  $\lambda$ , depending on the Auger electron energy and the material;



**Figure 3** Highly resolved Si KLL Auger transitions of silicon dioxide (dashed line), silicon nitride (solid line), and zero-valent silicon (dotted line). (Reproduced with permission from ULVAC-PHI.)



**Figure 5** Electron inelastic mean free path,  $\lambda$ , versus kinetic energy of (Auger) electrons. (Reproduced with permission from Seah MP and Dench WA (1979) *Surface and Interface Analysis*, vol. 1, 4p. Wiley, © Wiley.)



**Figure 4** Primary excitation volume and escape depth for Auger electrons. (Reproduced with permission from ULVAC-PHI.)

- spectrometer function (analyzer transmission, detector efficiency, geometry of the experiment, etc.).

Matrix effects, which are somewhat stronger in comparison to those in quantitative X-ray photoelectron spectroscopy, mostly originate from the backscatter effect. Many of the parameters mentioned above are difficult to determine. Therefore, in practical quantitative AES empirical relative sensitivity factors (RSFs) are used and the concentration  $C_i$  (at%) of element  $i$  in the thin AES analytical volume is derived for a certain matrix from measurements of the intensities measured for element  $i$ ,  $I_i$ , and all elemental constituents  $j$  by

$$C_i = 100 \text{RSF}_i I_i / [\sum_j (\text{RSF}_j I_j)] \quad [2]$$

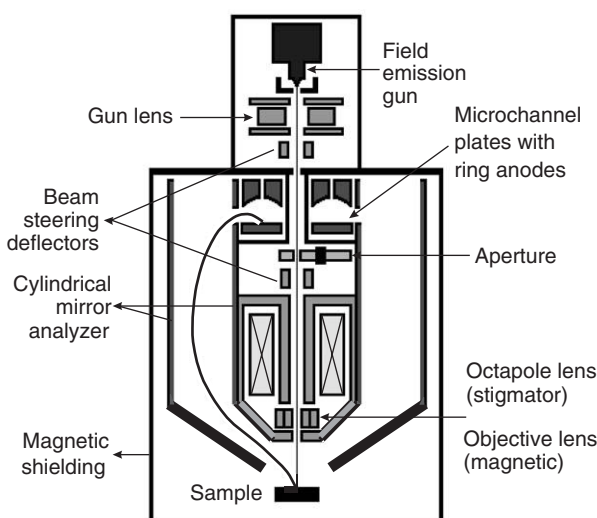
Matrix effects are thus 'embedded' in the RSFs. There are RSFs for elements and selected matrices (compounds) published in the literature. Certified reference materials are very useful to obtain reliable RSFs. However, sample homogeneity is required for application of eqn [2]. The primary electron beam diameter must be narrower than the scale of lateral heterogeneity. Heterogeneity of the analytical volume in depth is also crucial. Here, special quantitative models were developed by different authors.

The detection limits of quantitative AES are in the range of 1–0.01 at% depending on the element and the matrix under consideration.

## Instrumentation

A variety of electrostatic analyzers are used as electron spectrometers. The cylindrical mirror analyzer is the one most frequently used in AES because of its combined excellent energy resolution and high transmission efficiency. A schematic diagram of a recent Auger spectrometer design is shown in Figure 6. The cylinders of the analyzer are arranged coaxially; the inner cylinder is at ground potential while the outer cylinder is held at a negative potential. Electrons emitted from the analyzed sample enter the zone between the cylinders through an aperture and are deflected by the negative potential on the outer cylinder. By sweeping the potential on the outer cylinder, Auger electrons of varying energy can pass through the analyzer.

The primary electron gun, electron optical devices, deflection plates and the micro channel plate Auger electron detector are mounted coaxially with the cylinders of the analyzer. The focal point of the analyzer is coincident with the focal point of the primary electron beam. Scanning Auger microscopy (SAM) is



**Figure 6** Schematic diagram of a recent Auger electron spectrometer design. (Reprinted from [www.eaglabs.com](http://www.eaglabs.com) with permission of the Evans Analytical Group.)

enabled by using the deflection plates which direct the primary electron beam across the analysed surface. Modern instruments use high-brightness field emission primary electron guns operated at up to 25 keV. Beam diameters less than 10 nm are achievable. Usually, Auger spectrometers are equipped with a secondary electron detector, which additionally allows secondary electron imaging.

Another standard component is an ion gun providing ions which may impinge on a defined sample position seen by the Auger electron analyzer. The ions gradually erode the sample surface so that when combined with the surface analysis intrinsic to AES an elemental in-depth profile is obtained from a well-defined small area of the sample. Inert gases, such as argon, are typically used so that the ions do not react with the sample surface. The erosion rate is material dependent but can be 50–100 nm min<sup>-1</sup>.

It should be mentioned here that AES must be carried out under ultrahigh vacuum conditions. Samples to be measured must be compatible to this and stable under energetic electron irradiation.

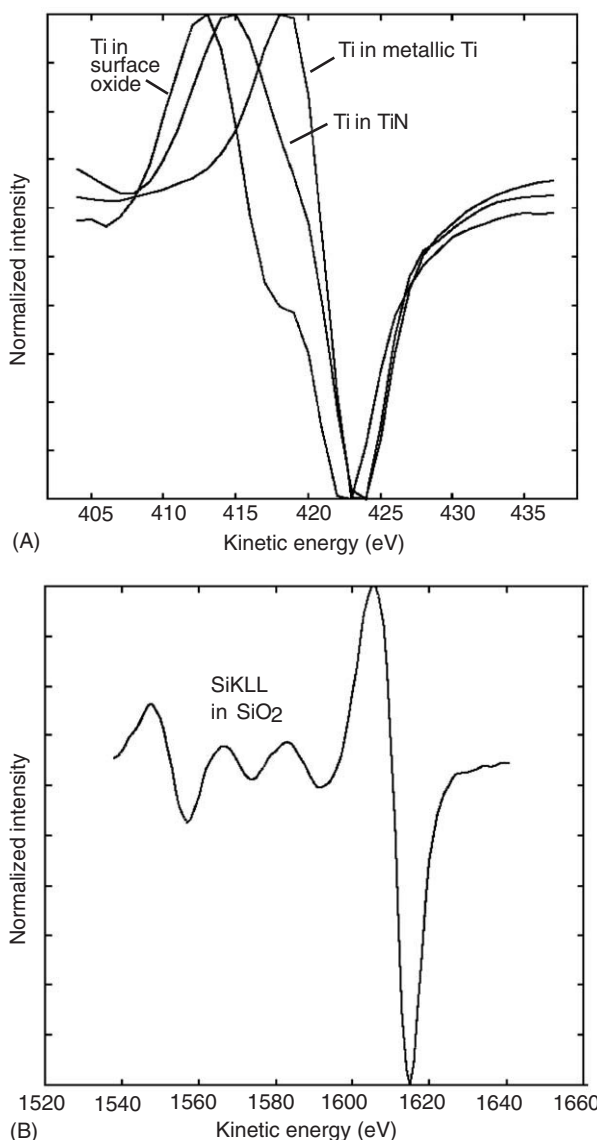
Sample charging is a frequent problem encountered in the analysis of insulators by AES. Any surface charging will distort the Auger signal intensity and shape as well shift the spectrum rendering the analysis meaningless. Charging may also occur for conductive material embedded in insulating surroundings. It has been shown that a positive low-energy argon ion beam can be successfully used to neutralize the surface charge. Ion current and energy as well as the irradiated area are the critical parameters here.

## Applications

AES is sufficiently mature and its application broad. Many general reviews of the technique deal with specific applications in general surface and thin film analysis. AES is often used to solve problems in metallurgy, plating, corrosion, and catalysis. Reviews covering these applications are listed in the Further Reading section. Because the primary electron beam can be focused down to a diameter of less than 10 nm, information about local compositions on a specimen's surface can be obtained. This special feature makes AES very attractive for applications in the semiconductor technology where submicrometer features are of interest. To satisfy semiconductor manufacturers, AES systems that are able to handle 300 mm silicon wafers are commercially available now.

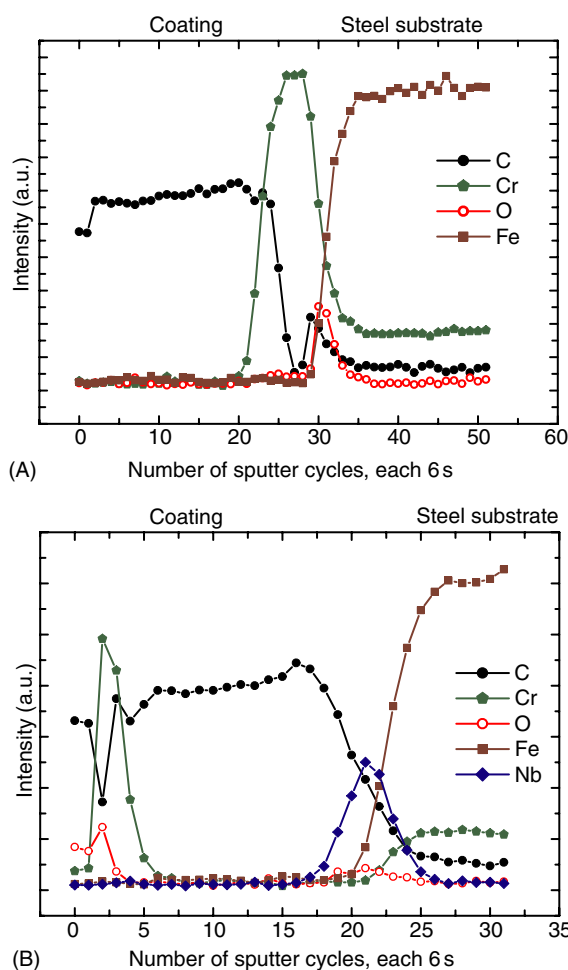
There are different modes of application of AES: (1) spectroscopy, also for rather small areas, (2) depth profiling, and (3) element mapping.

Considering the spectroscopic mode, a typical survey scan covers the range from 0 to 2000 eV of kinetic energy. Since there are characteristic kinetic energies of Auger transition series, i.e., KLL, LMM, and MNN, elemental identification for elements with  $Z \geq 3$  is excellent and quantification is possible. The Auger yield of the characteristic Auger transition series decreases, in competition with X-ray emission, as a function of increasing atomic number. However, there is always at least one intense Auger transition available for characterization. Since AES makes use of a focused electron beam for excitation, high spatial resolution is achieved. This permits analysis of, e.g., particles, residues, corrosion products, and adsorbed contaminants that may be nonuniformly dispersed over the surface. As mentioned earlier, Auger spectra contain a bounty of chemical information. To derive this, the spectroscopy mode has to be used at high-energy resolution. Historically, Auger data was taken in the derivative mode and plotted as  $d(N(E) \times E)/dE$  versus kinetic energy. In this mode, the peak minima were used to characterize the transition and the peak shape was frequently used as fingerprint of the chemical oxidation state. **Figure 7A** displays the changes of the Ti  $L_{3}M_{23}V$   $d(N(E) \times E)/dE$  spectrum depending on the chemical state of Ti. **Figure 7B** shows the Si KLL  $d(N(E) \times E)/dE$  spectra of  $\text{SiO}_2$  which can be compared to the respective  $N(E)$  spectrum given in **Figure 3**. Modern instruments collect the data in the  $N(E) \times E$  or  $N(E)$  mode (providing outputs as displayed in **Figure 3**) and electronically take the derivative when desirable. In any case  $N(E) \times E$  or  $N(E)$  spectra are better suited for quantification.



**Figure 7** (A) Ti  $L_{3}M_{23}V$   $d(N(E) \times E)/dE$  Auger spectra of a Ti surface oxide, TiN, and metallic Ti. (B) Si KLL  $d(N(E) \times E)/dE$  Auger spectra of silicon dioxide. (Reproduced with permission from ULVAC-PHI.)

Surface analysis by itself is rarely sufficient to characterize a sample. Sputter depth profiling in combination with surface analysis by AES is essential in order to characterize a sample thoroughly. **Figures 8A** and **8B** illustrates the insight that may be obtained for different tribological coatings on a steel substrate. By monitoring the C, Cr, Nb, O, and Fe Auger transitions as function of sputtering time, the in-depth composition of the coating was determined. After the initial surface contamination was removed by sputtering, a layer of  $\sim 50$  nm of carbon originating from a diamond-like carbon (DLC) layer was observed (**Figure 8A**). The DLC layer was



**Figure 8** (A) AES depth profile of steel substrate plated with a Cr interlayer and a diamond-like carbon layer using stepwise sputter erosion of the sample. (B) AES depth profile of steel substrate plated with an Nb interlayer, a diamond-like carbon layer, and a thin Cr top layer using stepwise sputter erosion of the sample. (Reproduced with permission from ULVAC-PHI.)

deposited on the steel substrate, which is represented by Fe and Cr Auger transitions, using a Cr interlayer of  $\sim 5$  nm to promote adhesion. The small C and O peaks between the Cr interlayer and the steel substrate point to some contamination of the steel substrate before coating. **Figure 8B** displays results obtained with a more complex system. Here a very thin Cr top coating on a thicker DLC layer was found and the adhesion promoting interlayer was made of Nb.

The primary result of an AES depth profile analysis is an Auger transition intensity versus sputter time presentation. Quantification leading to atomic concentrations is possible when RSFs or reference materials are available. A measurement of sputter rates which depend on the sputtered material enables

changing of the sputter time scale into a depth scale. However, because the effort required to determine reliable RSFs and sputter rates for an individual sample is rather high, often the Auger transition intensity versus sputter time representation of depth profiles is successfully used to compare related samples.

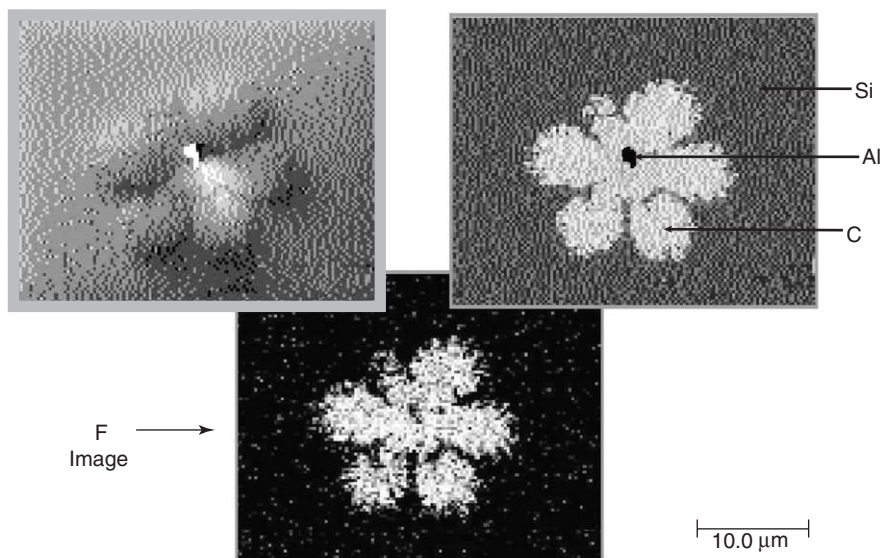
The third mode in which Auger data may be taken is the AES mode called scanning Auger microscopy, or SAM. Since the electron beam can be focused to a rather small spot, AES may be used to analyze localized features of a sample surface. This is accomplished in AES by rastering the electron beam over the area being analyzed while synchronously detecting the Auger signal. This is usually done by setting the analyzer transmission so that electrons of a specific energy reach the detector. Elemental distribution maps for different elements must be obtained sequentially. The distribution of compounds or elements in a certain oxidation state may be mapped when a chemically shifted peak is available for analysis. **Figure 9** shows the secondary electron micrograph of a 'flower defect' on a silicon wafer with low contrast because the petal-shaped residue is only 1–2 nm thick. SAM is the only method to detect such a thin contamination because of its unique spatial resolution and surface sensitivity. An accompanying overlay of Si, Al, and C AES maps as well as an individual F AES map is also given in **Figure 9**. These elemental maps reveal that the defect comprises a submicrometer scaled central particle with Al as a constituent and a fluorocarbon contaminant petal surrounding it.

## Quality Assurance in AES

One critical aspect of the reliable application of AES in research and technology by warranted quality of analytical results is the establishment of reference data, reference materials, reference procedures and, finally, written international standards. There are different bodies involved in this work.

At first the Versailles Project on Advanced Materials and Standards (VAMAS) to support international trade through projects aimed at providing the technical basis for drafting codes or practice and specifications in advanced materials should be mentioned here. The VAMAS Technical Working Area (TWA) 2 'Surface Chemical Analysis' has the production of reference procedures, data, and materials for surface chemical analysis as its objectives. Rather important interlaboratory comparisons were organized as part of VAMAS TWA2 projects. A survey of working and finalized TWA2 projects as well as





**Figure 9** Secondary electron image (upper left) and Auger elemental maps of C, Al, Si, and F of silicon wafer contamination comprising an Al central particle with fluorocarbon petals. (Reproduced with permission of ULVAC-PHI.)

related publications and reports is given at <http://www.vamas.org>.

The ISO Technical Committee 201 'Surface Chemical Analysis', the second body that is active in this area, is dedicated to the standardization in the field of surface chemical analysis in which beams of electrons, ions, neutral atoms or molecules, or photons are incident on the specimen material and scattered or emitted electrons, ions, neutral atoms or molecules, or photons are detected. Valid ISO standards on surface chemical analysis with relevance to AES are (January 2004, see ISO TC 201 at <http://www.iso.ch>):

- ISO 14606 Sputter depth profiling – optimization using layered systems as reference materials;
- ISO 14975 Information formats;
- ISO 14976 Data transfer format;
- ISO/TR 15969 Depth profiling – measurement of sputtered depth;
- ISO 17973/17974 – Medium-resolution/high-resolution Auger electron spectrometers – calibration of energy scales for elemental analysis;
- ISO 18115 Vocabulary;
- ISO/TR 19319 Auger electron spectroscopy – determination of lateral resolution, analysis area, and sample area viewed by the analyzer.

Moreover, there are a number of AES-related ISO standards under development. They are dedicated to the description of selected instrumental performance parameters (DIS 15471), specimen handling (CD

18116 and 18117), use of experimentally determined relative sensitivity factors for the quantitative analysis of homogeneous materials (CD 18118), derivation of chemical information (AWI 18394), determination of lateral resolution (WD 18516), peak intensity determination (WD 20903), linearity of intensity scale of AES spectrometers (PRF 21270), peak detection methods to perform qualitative or quantitative analysis (WI 22474), and repeatability and constancy of the intensity scale of AES spectrometers (D 24236).

The American Society for Testing and Materials Committee E-42 on Surface Analysis (<http://www.astm.org>), as a third body, continues to be active in developing standards. E-42 members have developed a Standard Terminology (E673-02b), as well as AES-related Standard practices and Standard Guides on the identification of elements (E827-02), minimizing unwanted electron beam effects (E983-94), identifying chemical effects and matrix effects (E984-95), background subtraction (E995-97), reporting data (E996-94), depth profiling (E1127-03), determination of current density of ion beams for sputter depth profiling (E684-95), and reporting ion beam parameters (E1577-95).

As a result, AES as an analytical method can be viewed to be sufficiently mature and the level of standardization reached now allows the preparation of standard operation procedures (SOPs) for its application. Thus, an integration of AES into practical quality management systems following ISO 17025 is possible and today there is a number



of accredited testing laboratories running SOPs using AES.

**See also: Quality Assurance:** Instrument Calibration; Interlaboratory Studies; Reference Materials; Method Validation; Traceability; Accreditation. **Surface Analysis:** Overview; X-Ray Photoelectron Spectroscopy.

## Further Reading

*Annual Book of ASTM Standards*, vol. 03.06.

*AVS Surface Science Spectra Database* published as a journal (Surface Science Spectra) and available online at <http://sss.avs.org>.

Briggs D and Seah MP (eds.) (1990) *Practical Surface Analysis, vol. 1, Auger and X-Ray Photoelectron Spectroscopy*. Oberentfelden: Sauerländer Verlage AG.

Briggs D and Grant JT (eds.) (2003) *Surface Analysis by Auger and X-Ray Photoelectron Spectroscopy*. Chichester, UK: IM Publications.

Childs KD, Carlson BA, LaVanier LA, et al. (1995) *Handbook of Auger Electron Spectroscopy*, 3rd edn. Eden Prairie, MN: Physical Electronics Inc.

*NIST Standard Reference Database 71: NIST Electron Inelastic-Mean-Free-Path Database*. Available online at: <http://www.nist.gov/srd/nist71.htm>.

Rivière JC and Myhra S (eds.) (1998) *Handbook of Surface and Interface Analysis*. New York: Dekker.

## Appearance Potential Spectroscopy

**A R Chourasia**, Texas A&M University-Commerce, TX, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The structural and electronic properties of the top few atomic layers of metals and semiconductors play a dominant role in the chemical (catalysis, corrosion) and electronic (semiconductor device interface) properties of a range of materials. For the study of electronic structure, surface analytical techniques typically employ particles such as photons, electrons, atoms, molecules, or ions. These particles must have just enough energy to probe only a few tenths of a nanometer into the solid, thereby investigating the surface atoms rather than the bulk atoms. Since low-energy electrons have inelastic mean free paths of only a few tenths of a nanometer in solids, they are made use of in many techniques, either as the incident projectile or as a particle emitted from the surface region.

The distribution of the energy states (density of states, DOS) in the vicinity of the Fermi level,  $E_F$  (the highest energy occupied level), forms the valence (occupied states) and conduction (unoccupied states) bands of solids. The technique employed for the study of the electronic structure of surfaces, therefore, fall broadly into two groups: techniques that investigate the occupied DOS (such as X-ray photoelectron spectroscopy (XPS) and Auger electron spectroscopy (AES)) and those that investigate the unoccupied density of states (e.g., X-ray absorption spectroscopy (XAS), inverse photoemission (IPE), Bremsstrahlung isochromat spectroscopy (BIS), and

appearance potential spectroscopy (APS)). This article discusses the different techniques that explore the conduction band states, concentrating on the various aspects of APS that apply to surface characterization of materials.

In XAS, the incident X-rays traverse through the sample. X-rays of appropriate energy are absorbed by the atoms and as a result a core electron is excited to unoccupied states above  $E_F$ . The decrease in the transmitted X-ray intensity is measured. In this spectroscopy, selection rules are rigorously obeyed and therefore the DOS of all symmetries must be considered to understand completely the electronic structure of the conduction band. The schematic of this X-ray excitation of a core level atom, in the one-electron approximation, is shown in **Figure 1**. In the one-electron approximation, the activity of one electron is considered at any instant, assuming all other electrons to be fixed. Because the incident X-rays traverse the whole sample, little surface information is obtained.

In BIS, the sample to be investigated serves as the anode of an X-ray tube. An X-ray spectrometer is tuned to detect X-rays of certain definite energy,  $\hbar\omega_0$ , which is arbitrarily selected from the Bremsstrahlung spectrum ( $\hbar$  is  $h/2\pi$ , where  $h$  is Planck's constant and  $\omega_0$  is the selected angular frequency). As the accelerating energy  $E$  across the X-ray tube is slowly varied from just below to just above the threshold energy  $\hbar\omega_0$ , the observed intensity begins to rise and then shows structure. The curve of change in intensity for one fixed energy  $\hbar\omega_0$  as a function of incident electron energy in such an experiment is called an isochromat. The electrons with initial energy  $E(E \geq \hbar\omega_0)$ , which undergo radiative transitions emitting photons of energy  $\hbar\omega_0$  will then occupy

of accredited testing laboratories running SOPs using AES.

**See also: Quality Assurance:** Instrument Calibration; Interlaboratory Studies; Reference Materials; Method Validation; Traceability; Accreditation. **Surface Analysis:** Overview; X-Ray Photoelectron Spectroscopy.

## Further Reading

*Annual Book of ASTM Standards*, vol. 03.06.

*AVS Surface Science Spectra Database* published as a journal (Surface Science Spectra) and available online at <http://sss.avs.org>.

Briggs D and Seah MP (eds.) (1990) *Practical Surface Analysis, vol. 1, Auger and X-Ray Photoelectron Spectroscopy*. Oberentfelden: Sauerländer Verlage AG.

Briggs D and Grant JT (eds.) (2003) *Surface Analysis by Auger and X-Ray Photoelectron Spectroscopy*. Chichester, UK: IM Publications.

Childs KD, Carlson BA, LaVanier LA, et al. (1995) *Handbook of Auger Electron Spectroscopy*, 3rd edn. Eden Prairie, MN: Physical Electronics Inc.

NIST Standard Reference Database 71: NIST Electron Inelastic-Mean-Free-Path Database. Available online at: <http://www.nist.gov/srd/nist71.htm>.

Rivière JC and Myhra S (eds.) (1998) *Handbook of Surface and Interface Analysis*. New York: Dekker.

## Appearance Potential Spectroscopy

**A R Chourasia**, Texas A&M University-Commerce, TX, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

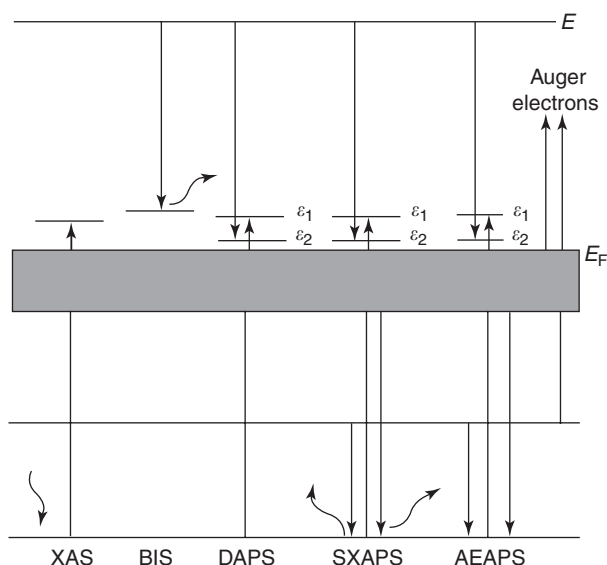
The structural and electronic properties of the top few atomic layers of metals and semiconductors play a dominant role in the chemical (catalysis, corrosion) and electronic (semiconductor device interface) properties of a range of materials. For the study of electronic structure, surface analytical techniques typically employ particles such as photons, electrons, atoms, molecules, or ions. These particles must have just enough energy to probe only a few tenths of a nanometer into the solid, thereby investigating the surface atoms rather than the bulk atoms. Since low-energy electrons have inelastic mean free paths of only a few tenths of a nanometer in solids, they are made use of in many techniques, either as the incident projectile or as a particle emitted from the surface region.

The distribution of the energy states (density of states, DOS) in the vicinity of the Fermi level,  $E_F$  (the highest energy occupied level), forms the valence (occupied states) and conduction (unoccupied states) bands of solids. The technique employed for the study of the electronic structure of surfaces, therefore, fall broadly into two groups: techniques that investigate the occupied DOS (such as X-ray photoelectron spectroscopy (XPS) and Auger electron spectroscopy (AES)) and those that investigate the unoccupied density of states (e.g., X-ray absorption spectroscopy (XAS), inverse photoemission (IPE), Bremsstrahlung isochromat spectroscopy (BIS), and

appearance potential spectroscopy (APS)). This article discusses the different techniques that explore the conduction band states, concentrating on the various aspects of APS that apply to surface characterization of materials.

In XAS, the incident X-rays traverse through the sample. X-rays of appropriate energy are absorbed by the atoms and as a result a core electron is excited to unoccupied states above  $E_F$ . The decrease in the transmitted X-ray intensity is measured. In this spectroscopy, selection rules are rigorously obeyed and therefore the DOS of all symmetries must be considered to understand completely the electronic structure of the conduction band. The schematic of this X-ray excitation of a core level atom, in the one-electron approximation, is shown in **Figure 1**. In the one-electron approximation, the activity of one electron is considered at any instant, assuming all other electrons to be fixed. Because the incident X-rays traverse the whole sample, little surface information is obtained.

In BIS, the sample to be investigated serves as the anode of an X-ray tube. An X-ray spectrometer is tuned to detect X-rays of certain definite energy,  $\hbar\omega_0$ , which is arbitrarily selected from the Bremsstrahlung spectrum ( $\hbar$  is  $h/2\pi$ , where  $h$  is Planck's constant and  $\omega_0$  is the selected angular frequency). As the accelerating energy  $E$  across the X-ray tube is slowly varied from just below to just above the threshold energy  $\hbar\omega_0$ , the observed intensity begins to rise and then shows structure. The curve of change in intensity for one fixed energy  $\hbar\omega_0$  as a function of incident electron energy in such an experiment is called an isochromat. The electrons with initial energy  $E(E \geq \hbar\omega_0)$ , which undergo radiative transitions emitting photons of energy  $\hbar\omega_0$  will then occupy



**Figure 1** A schematic diagram showing energy levels involved in techniques used to probe unoccupied density of states of a metal. XAS, X-ray absorption spectroscopy; BIS, Bremsstrahlung isochromat spectroscopy; IPE, inverse photoemission; DAPS, disappearance potential spectroscopy; SXAPS, soft X-ray appearance potential spectroscopy; and AEAPS, Auger electron appearance potential spectroscopy.

final states in the conduction band of the sample at an energy  $E - \hbar\omega_0$ . Assuming a constant radiative matrix element, the transition probability will be proportional to the density of unfilled states. The above-measured intensity will, therefore, provide the density of empty one-electron states. The electron transitions involved in the process are shown schematically in Figure 1.

IPE is an analog of BIS but with low-energy photon emission. It is an important and growing area in present-day surface analysis. IPE represents radiative capture of an electron by a solid surface, and is the inverse of photoemission. In IPE, the sample is bombarded by electrons with energy of only a few tens of electron volts. The technique is sensitive to the surface region because of the finite pathlength of low-energy electrons in solids. The photon energy spectrum emitted from a solid under electron bombardment contains information on the density of electronic states above  $E_F$ . The photon yield, which is extremely small, is measured dispersively by energy selection detection techniques. The curve of photon intensity as a function of  $E - \hbar\omega$  ( $\hbar\omega$  denotes the photon energy) measures the band structure above  $E_F$ . In actual isochromat mode, the energy of the detected photon  $\hbar\omega$  is held constant and the spectrum is obtained by sweeping  $E$ . The spectrum is considered to be a replica of the unoccupied DOS above  $E_F$ .

In APS, the surface of the specimen is bombarded by electrons in the 0–2000 eV range. When the energy of the incident electron equals that of a particular core level, the incident electron imparts its energy to the core level electron as a result of the inelastic collision. The system is then in an excited state and consists of a core hole and two electrons above  $E_F$  in the conduction band. The excitation is followed by radiative (X-ray fluorescence) and non-radiative (Auger electron) processes that compete for the de-excitation of the core holes. When the intensity of the emitted X-rays is measured the method is called soft X-ray APS or SXAPS. When the total current of the secondary electrons is measured the method is called Auger electron APS or AEAPS. In another mode, the current of the elastically reflected electrons is measured. The number of electrons that are scattered inelastically at the threshold energy of the core level excitation increases. They disappear from the measured current, and the method is, therefore, called disappearance potential spectroscopy or DAPS. The energy level diagram for APS is shown in Figure 1.

## Theoretical Background

APS measures the probability for electronic excitation of a core level as a function of incident electron energy. The energy of the incident electron is gradually increased, and the dependence of the total signal strength on this energy is measured. At certain energies a sudden change (increase or decrease) in the signal is observed, which corresponds to the excitation of a given energy level of the sample. The signal is extracted from the background with the help of an electronic differentiation technique, which enhances the signal-to-noise ratio. The intensity of the features in the APS spectrum at the threshold energy and above depends on the core hole excitation rate at that energy. The final state in the excitation process will consist of a core hole, an excited core electron, and the scattered incident electron. The fact that the shape of an APS feature is independent of the relaxation mechanism suggests the utilization of a non-dispersive scheme for this technique. Since neither an electron energy analyzer nor an X-ray monochromator is needed to obtain the experimental data, APS is undoubtedly the simplest method for studying the unoccupied DOS of solid surfaces. The surface sensitivity of APS is due to the short inelastic mean free path for the primary electrons in the energy range 0–2 kV. A primary electron with energy close to the threshold that experiences a characteristic energy loss upon penetrating the sample is no longer

able to excite the atoms. The depth of information is, therefore,  $\sim 10 \text{ \AA}$ .

Low-energy electrons interact with atoms by elastic collisions, by the emission of electromagnetic radiation, and by inelastic collisions. At the threshold energy of the core hole excitation a certain fraction of the incident electrons is involved in ionization. These electrons give up their energy and can, therefore, no longer produce Bremsstrahlung radiation in the sample. When the incident electron is captured by a state  $\varepsilon_1$  above  $E_F$ , the energy may be conserved by the excitation of a core electron into a state  $\varepsilon_2$  such that

$$\varepsilon_1 + \varepsilon_2 = E - E_b \quad [1]$$

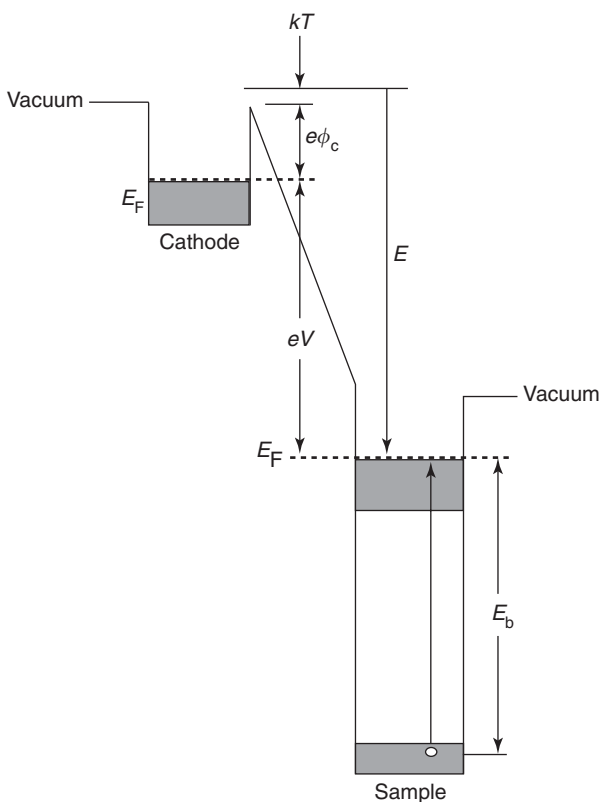
where  $E$  is the energy of the incident electron and  $E_b$  is the binding energy (BE) of a particular core level and  $E_F$  is taken as the zero of energy. The core level DOS is considered to have a negligible energy width. The incident energy is determined by  $eV + e\phi_c + kT$ , where  $V$  is the electron accelerating potential,  $e\phi_c$  is the emitter (cathode) work function, and  $kT$  is the thermal energy spread of the emitted electrons. Here,  $e$  is the charge of an electron,  $k$  is the Boltzmann constant, and  $T$  is the absolute temperature of the cathode. The final states must take into consideration all possible combinations of energies of both electrons (incident and core) that are consistent with the conservation of energy.

The characteristic emission has a distinct energy threshold (appearance potential), which occurs at  $E = E_b$  (i.e., for  $\varepsilon_1 = \varepsilon_2 = 0$ ). This results in the transfer of both the incident and core electrons to  $E_F$ , as shown in Figure 2. This event is signaled by the 'appearance' of a small bump in the total emission versus incident electron energy curve. The excitation process is the same for all APS spectra. It seems that differences in the decay step are responsible for the spectral differences between SXAPS, AEAPS, and DAPS. In terms of this model DAPS provides information directly about the excitation process and is free of the relaxation complications. SXAPS and AEAPS include additional information dependent on the different decay (relaxation) steps.

The excitation probability,  $P(E, E_b)$  of the APS process at a given incident energy  $E \geq E_b$  will be given by

$$P(E, E_b) \propto \int_0^{E-E_b} P_1(E \rightarrow \varepsilon_2) P_2(E_b \rightarrow \varepsilon_1) N(\varepsilon_1) N(\varepsilon_2) \times \delta(\varepsilon_1 + \varepsilon_2 - E + E_b) d\varepsilon_1 d\varepsilon_2 \quad [2]$$

The energy conservation is ensured by the  $\delta$  function. The factors  $P_1(E \rightarrow \varepsilon_2)$  and  $P_2(E_b \rightarrow \varepsilon_1)$ , which represent the transition probabilities, may depend on the



**Figure 2** Threshold electron excitation in APS that occurs when the incident electron energy  $E (= eV + e\phi_c + kT)$  is equal to the binding energy  $E_b$  of a core level, resulting in the transfer of both incident and core electrons to  $E_F$ . The core hole may subsequently decay by the emission of a characteristic X-ray or Auger electron.

selection rules and are not known precisely. If these factors are assumed to be constant over a small range of  $E$  values above the threshold  $E_b$ , then they can be taken as simply proportional to the DOS  $N(\varepsilon_1)$  and  $N(\varepsilon_2)$  and the equation simplifies to

$$P(E, E_b) \propto \int_0^{E-E_b} N(E - E_b - \varepsilon_2) N(\varepsilon_2) d\varepsilon_2 \quad [3]$$

which is simply the self-convolution of the density of the conduction band states.

In APS the signal is extracted from the background by modulation techniques. The resulting spectrum, therefore, represents the differential of the convoluted density of unoccupied states near  $E_F$ . The strength of the APS signal is given by

$$\text{APS}(E) \propto \int_0^{E-E_b} N(\varepsilon_2) \frac{d}{dE} N(E - E_b - \varepsilon_2) d\varepsilon_2 \quad [4]$$

Therefore, APS is particularly suitable for studying materials that have high DOS above  $E_F$ . There are

three main groups of elements that provide strong signals: 3d transition metals, alkaline earths, and rare earths. If the DOS is low or have disadvantageous character, APS cannot be utilized as a universal analytical technique. Nevertheless, APS has the potential for providing interesting results that cannot be obtained by other techniques.

The binding energies of the core level peaks are calculated from the corresponding appearance potential peaks. As core binding energies are characteristic of specific elements, chemical analysis of the surface is feasible. Since the APS yield is proportional to the self-convolution of the density of the unoccupied states broadened only by the finite lifetime of the core hole and by the finite experimental resolution, precise knowledge about the conduction band can be obtained by using deconvolution techniques. In SXAPS and AEAPS the intensity depends upon the fluorescence and Auger yields, respectively, which depend on the type of electron shell, on the binding energy, and on the local atomic environment.

As a tool for chemical analysis, APS has its merits when compared with other techniques. Electron-induced X-ray emission spectroscopy employs primary electrons of several kiloelectronvolts in most experiments, so that information originates mainly from the bulk. Since the dispersion of X-rays by monochromators presents considerable experimental problems, very little attempt has been made to study surface properties by analyzing X-ray spectra. APS, however, has the advantage that it utilizes primary electrons with energy less than  $\sim 2$  keV for the excitation of X-ray spectra; the shallow penetration depth of these electrons makes APS an attractive technique for surface analysis. AES utilizes the non-radiative process following the ionization of a core level of an atom as a tool for surface analysis. This technique employs energy analysis to determine the electronically differentiated energy distribution of the Auger electrons. AEAPS is the derivative of the total secondary emission electron current with respect to incident electron energy and gives information about the density of unoccupied states above  $E_F$ . However, the applicability of AEAPS as an analytical tool for determining the surface composition in a manner similar to AES is limited, since AEAPS is quite insensitive to a number of elements. In XPS, the principle of the experiment is to measure the kinetic energies  $E_{kin}$  of photoelectrons liberated from their bound states with energies  $E_b$  by interaction with monochromatic X-rays of energy  $h\omega$  and is given by

$$E_{kin} = h\omega - E_b \quad [5]$$

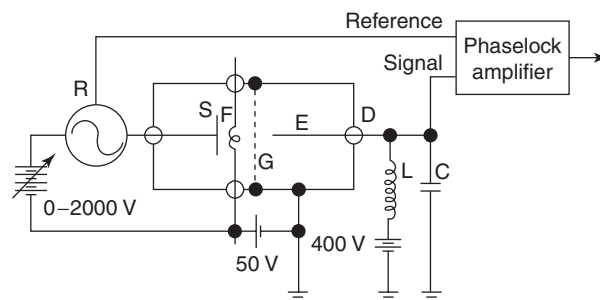
Soft X-ray sources with energies in the range of  $\sim 1$  keV are used, which makes the technique

sensitive to the surface properties of materials. XPS is suitable for studying the occupied DOS and the chemical shifts in the core electron-binding energies. In this respect, APS, AES, and XPS are complementary techniques.

## Experimental: Soft X-Ray Appearance Potential Spectroscopy

A schematic of the SXAPS spectrometer is shown in Figure 3. In SXAPS, the total soft X-ray intensity emitted by a sample under electron bombardment is measured as a function of incident electron energy. Electrons from a tungsten filament (F) impinge on the sample (S) to be studied. The collector electrode (E) and the stainless-steel chamber wall to the right of the grid (G) constitute an X-ray detector (D). The grid, which is negatively biased with respect to the filament, offers no effective obstruction to the X-rays emitted, while electrically screening the X-ray detector assembly from the filament-target assembly. The entire system is maintained under ultrahigh vacuum conditions. X-rays passing through the grid strike the walls of the chamber, which act as a photocathode. The resulting photoelectrons are collected by a positively biased collector electrode. The extraction of the signal is accomplished by a potential modulation technique. A reference signal (R) is superimposed on the target and the resulting variations in photocurrent are synchronously detected in the first differential mode with a phase-lock amplifier. The appearance potential spectrum represents the detection output versus the accelerating potential. It is generally advantageous to make AEAPS measurements in the second-derivative mode because the secondary electron emission does not exhibit a linear dependence on incident electron energy.

The resolution of the spectrometer depends on several factors, such as the voltage drop across the filament, the thermal energy spread of the incident



**Figure 3** The soft X-ray appearance potential spectrometer. S, sample; F, filament; E, collector electrode; D, X-ray detector; G, grid; R, reference signal; L, induction coil; C, condenser.

electrons, and the amplitude of the modulation voltage. Taking such factors into consideration, the resolution in APS can be kept below 0.5 eV.

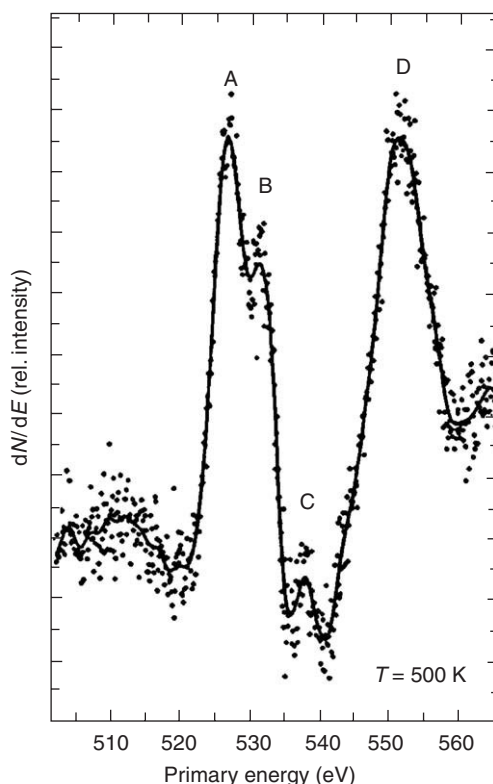
## Experimental Spectra

Unoccupied states of the insulator MnO (100) have been investigated by core electron energy loss (CEELS) and SXAPS at the oxygen 1s and manganese 2p thresholds. The SXAPS at the oxygen 1s threshold is shown in Figure 4. The peaks B and C have been identified as transitions into unoccupied metal 4s and 4p bands, which hybridize with oxygen 2p bands. The peak D is also observed in the CEELS spectrum and is due to multiple scattering processes of electrons excited into continuum states at neighboring oxygen ions. No differences were observed between SXAPS and CEELS spectra indicating delocalized character of the states at the ligand. Therefore, the electron correlation can be neglected in the SXAPS spectrum at the oxygen 1s threshold.

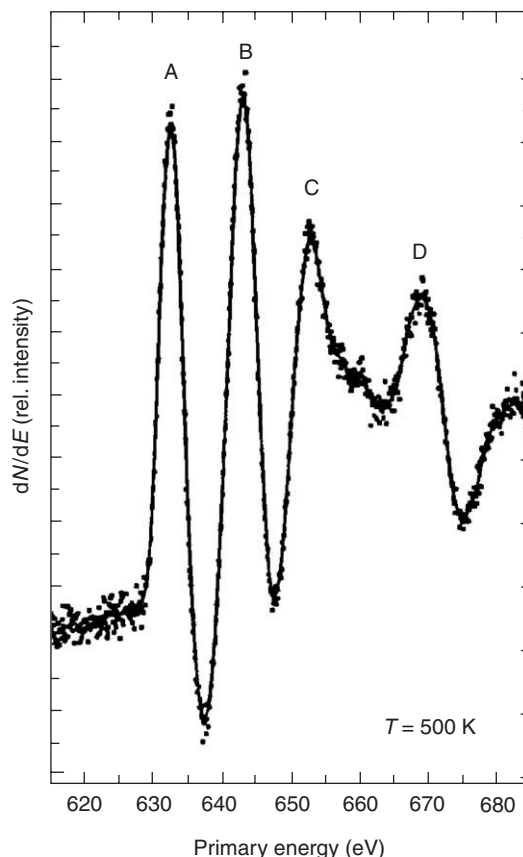
The SXAPS spectrum at the manganese 2p threshold is shown in Figure 5. The first two peaks (peaks A

and B) are interpreted as excitations of manganese 2p electrons into empty metal 3d levels. Significant differences were observed between this spectrum and the corresponding CEELS spectrum. The peak C is not a one-electron process but an excitation with two electrons involved, because this peak is absent in the CEELS spectrum. The differences have been attributed to the different number of localized electrons involved in the respective excitation channels and to the influence of electron correlation in the SXAPS spectrum at the Mn 2p threshold. The energetic difference between peaks B and C in Figure 5 was observed to be  $\sim 10$  eV, which is nearly the value of the electron correlation energy  $U_{dd}$  calculated on the basis of a configuration-interaction cluster model.

The DAPS spectra of La and Ce have been compared with ionization loss spectroscopy (ILS) spectra under various oxygen exposures. These two spectroscopies may respond in different ways to the changing chemical environment if Coulomb correlation effects are not negligibly small. The DAPS and ILS spectra for lanthanum are shown in

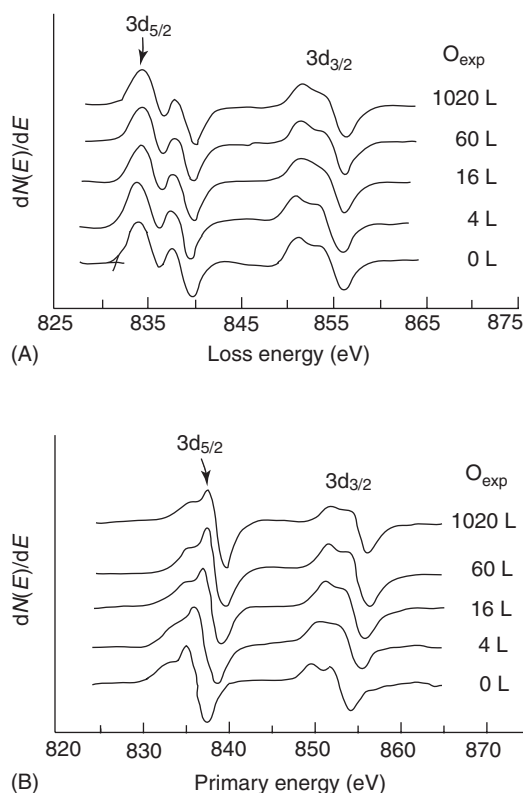


**Figure 4** SXAPS at the oxygen 1s threshold. (Reprinted with permission from Henig C, Untiet C, Merz H (1995). Electron correlation in MnO(100): A comparison of SXAPS and CEELS spectra. *Journal of Electron Spectroscopy and Related Phenomena* 76: 547–551; © Elsevier.)



**Figure 5** SXAPS at the manganese 2p threshold. (Reprinted with permission from Henig C, Untiet C, Merz H (1995). Electron correlation in MnO(100): A comparison of SXAPS and CEELS spectra. *Journal of Electron Spectroscopy and Related Phenomena* 76: 547–551; © Elsevier.)





**Figure 6** (A) 3d ILS of La at various oxygen exposures. Both line shapes and peak positions change slightly with increasing oxygen exposures. (B) 3d DAPS of La at various oxygen exposures. Compared with ILS, DAPS change more significantly in peak positions. (Reprinted with permission from Lu M, Quang-ji Z, and Hua Z (1995) Observations of Coulomb correlation effects during oxidation processes of La, Ce, Cr and Ti by ILS and DAPS. *Surface Science* 341: 182–189.)

**Figure 6.** In the case of lanthanides the 4f electrons are highly localized and are well correlated. The Fermi level,  $E_F$ , crosses the 5d6s valence band and the unoccupied 4f level is located above  $E_F$ . During the oxidation of La, the three hybridized (5d6s) valence electrons are transferred to the oxygen 2p derived level until the final oxide phase  $\text{La}_2\text{O}_3$  is formed. At this stage, the shapes of spectra of La and  $\text{La}_2\text{O}_3$  should be similar, a fact that has been experimentally confirmed. The DAPS peaks were observed to show greater changes in peak energies than the corresponding ILS peaks. Further analyses of the data showed that the difference between the peak positions in these two spectroscopies contained three items of Coulomb correlation energies:

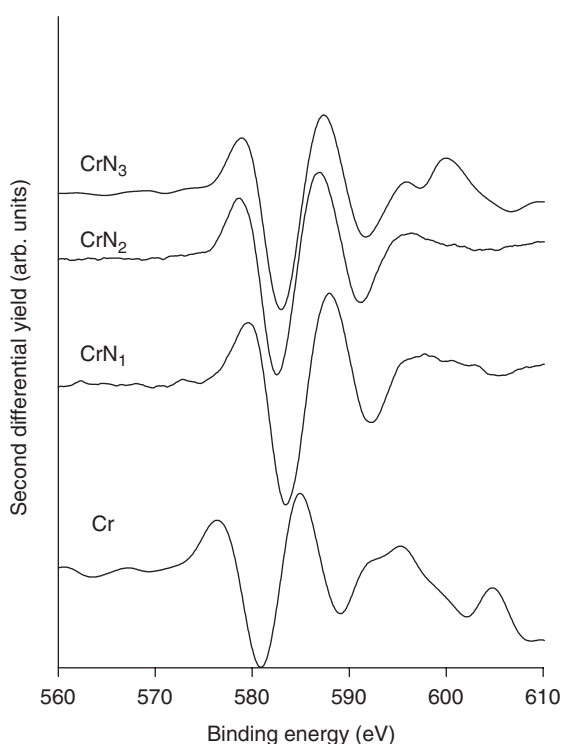
$$\delta D = \delta U_{ii} + \delta \varepsilon_i + \delta U_{ic} \quad [6]$$

where  $D$  is the difference between peak positions in DAPS and ILS,  $U_{ii}$  is the Coulomb repulsion energy between the excited electron and the impinging

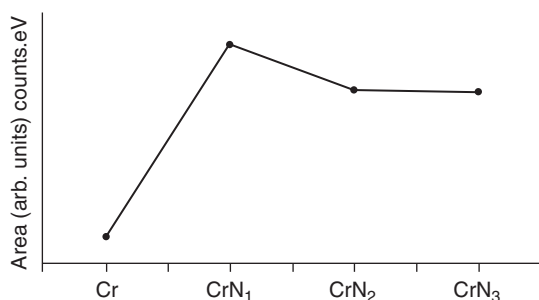
electron at level  $i$ ,  $\varepsilon_i$  is the unoccupied energy level  $i$  on which the excited electron resides, and  $U_{ic}$  is the Coulomb attraction between the excited electron at level  $i$  and the corresponding core hole created in the excitation process. The  $\delta$  indicates changes in the corresponding values. From the spectral analysis,  $\delta D$  was observed to be  $\sim 2.0$  eV between  $\text{La}_2\text{O}_3$  and La. With increasing oxidation of La, the increase in the  $\delta D$  values has been explained as due to the competition between the increase of the 4f energy level  $\varepsilon_f$  and the relatively less decrease of  $U_{fc}$  (the Coulomb attraction between the excited electron at  $E_F$  and the corresponding core hole created in the excitation process) until the  $\text{La}_2\text{O}_3$  layer is formed.

In case of cerium, the spectra showed indication of the 4f electrons taking part in bonding during the formation of  $\text{CeO}_2$ . After an exposure of 1020 Langmuir, a broadening of the spectra was observed and has been attributed to the overlapping of the  $\text{Ce}_2\text{O}_3$  and  $\text{CeO}_2$  spectra. The value of  $\delta D$  was found to increase with increasing oxygen exposure and was largely attributed to the increase in  $\varepsilon_f$ . Compared to  $\text{Ce}_2\text{O}_3$ , this  $\varepsilon_f$  for  $\text{CeO}_2$  showed a dramatic decrease and has been explained in terms of the localized f electron taking part in bonding (from  $\text{Ce}_2\text{O}_3$  to  $\text{CeO}_2$ ). As a result of this bonding the interaction between the excited electron and the electrons below  $E_F$  should be greatly reduced. This will then lead to a significant decrease in  $\varepsilon_f$ , as observed.

Chromium nitride thin films have received increasing attention in recent years due to their interesting physical properties, such as oxidation resistance, corrosion resistance, and wear behavior. The technique of AEAPS has been used to investigate the bonding characteristics in  $\text{CrN}_x$  thin films as a function of nitrogen content. The N/Cr atomic ratio as determined by AES is 0.2, 0.4, and 1.0. The structure has been determined by X-ray diffraction and was found to change from bcc + hexagonal to hexagonal to cubic with increasing nitrogen content. The  $\text{L}_{2,3}$  core level spectra of chromium in the films are shown in Figure 7. In this figure, sample 1 corresponds to the N/Cr atomic ratio of 0.2, sample 2 to 0.4, and sample 3 to 1.0. It is a plot of the second differential Auger electron yield as a function of the sample potential. The spectra have been normalized for equal intensity just below the threshold. In AEAPS, the inelastic event occurring leaves the incident and the core electrons in available states at  $E_F$ . Above this threshold, the electrons can lie in a range of vacant states that conserve energy. The strength of the signal at the threshold therefore represents the unoccupied DOS at  $E_F$ . From the Figure 8 it is seen that the signal strength increases in all of the CrN films compared with elemental chromium. This represents an



**Figure 7** Normalized chromium  $L_{2,3}$  level AEAPS spectra for the CrN films and elemental chromium. The emission current was 2 mA and the modulation voltage was 1 V<sub>p-p</sub>. The strength of the signal (i.e., the density of unoccupied states at the Fermi level) is more for the films. (Reproduced with permission from Chourasia AR and Hood SJ (2001) Auger electron appearance potential spectroscopy. *Surface and Interface Analysis* 31: 291–296; © John Wiley and Sons Ltd.)



**Figure 8** Area under the Cr  $L_{2,3}$  levels in CrN films and elemental chromium. The minima were aligned and the areas under the peaks were evaluated in a fixed energy window after the background subtraction. The plot represents variation in the total unoccupied density of states in the films compared with elemental chromium. (Reproduced from Chourasia AR and Hood SJ (2001) Auger electron appearance potential spectroscopy. *Surface and Interface Analysis* 31: 291–296.)

increase in unoccupied DOS at  $E_F$  as a result of chemical bonding between Cr and N. The increase in the unoccupied states of chromium, therefore, represents a charge transfer from Cr to N in the CrN

films. The established direction of charge transfer is consistent with Pauling's electronegativity criterion. In the films, however, the signal strength is found to decrease with an increase in the nitrogen content. The unoccupied DOS, therefore, decreases as the concentration of nitrogen increases in the films. This shows a decrease in charge transfer from Cr to N in the films. These observations are consistent with XPS and XAS investigation performed on the CrN films.

In the case of metal–nonmetal refractory compounds, the bonds are formed in two ways: a  $\sigma$ -bond is formed by an overlap of nonmetal 2p with metal 3d- $e_g$  orbitals, and a  $\pi$ -bond is formed by an overlap of non-metal 2p with metal 3d- $t_{2g}$  orbitals. In the CrN<sub>x</sub> films increasing the N/Cr atomic ratio will increase the number of antibonding nitrogen p-states. This would lead to raising of the unoccupied Cr d-states resulting in a decrease in both the  $\sigma$ - and the  $\pi$ -bonds as the concentration of nitrogen increases. This is consistent with the conclusions on charge transfer in these films.

In electron spectroscopies the investigation of excitation and transport processes poses problems owing to the many possible electron interactions. Because of the threshold energies used in APS the inelastic interaction excludes the incident electron from the number of electrons that can contribute to the signal, unlike in other spectroscopies. Also, the processes connected with the excitation of a particular core level can be studied separately. Since DAPS involves no relaxation process, it is possible to investigate the excitation probability for particular sublevels.

Knowledge of information depth obtained from a surface analysis technique determines its wide use in surface analysis. Only a few attempts have so far been made to gather information depth based on APS; this is one of the reasons why APS has not become a popular technique for surface analysis. The information depth obtained from APS technique correlates well with that determined by AES.

## Density of States

The technique of APS reveals a localized DOS because the matrix element governing the core hole production involves the very short range wave function of the initial core electron states. APS reveals information regarding the total DOS of all symmetries in the conduction band. Also, APS does not require a dispersive analyzer. This accounts for the extreme simplicity of the APS spectrometer. APS provides information regarding the elemental identification, chemical bonding, density of unoccupied

states, nearest-neighbor configuration in the surface layer, and the mechanism of excitation and transport processes occurring in the surface region of the sample. Wide scope still exists for theoretical and experimental work in this field. When the different aspects of APS are fully exploited, this spectroscopy will become a popular technique for materials characterization of surfaces.

See also: **Surface Analysis:** Auger Electron Spectroscopy; Ion Scattering; Overview; X-Ray Photoelectron Spectroscopy.

## Further Reading

- Barco JL and Ferron J (1990) Information depth of Auger electron and appearance potential spectroscopies: A Monte Carlo study. *Surface Science* 232: 393–398.
- Briggs D and Seah MP (eds.) (1990) *Practical Surface Analysis by Auger and X-Ray Photoelectron Spectroscopy*, vol. 1. New York: Wiley.
- Chourasia AR and Chopra DR (1997) Auger electron spectroscopy. In: Settle F (ed.) *Handbook of Instrumental Techniques for Analytical Chemistry*, pp. 793–808. NJ: Prentice-Hall.
- Dose V and Fauster T (1979) Deconvolution of appearance potential spectra II. *Applied Physics* 20: 299–303.
- Gerlach RL, Houston JE, and Park RL (1970) Ionization spectroscopy of surfaces. *Applied Physics Letters* 16: 179–181.
- Henig C and Merz H (1998) Inner shell excitations of the correlated insulator MnO(100). *Journal of Electron Spectroscopy and Related Phenomena* 93: 189–192.
- Hinkers H, Stiller R, and Merz H (1989) Near-threshold excitation in light rare-earth compounds: A new interpretation of 3d appearance-potential spectra. *Physical Review B* 40: 10594–10603.
- Houston JE and Park RL (1972) Difference in the Cr L<sub>3</sub>/L<sub>2</sub> intensity ratio measured by soft X-ray and Auger-electron appearance potential spectroscopy. *Physical Review B* 5: 3808–3809.
- Kirshchner J and Staib P (1975) Disappearance potential spectroscopy. *Applied Physics* 6: 99–109.
- Koningsberger DC and Prins R (eds.) (1988) *X-Ray Absorption: Principles, Applications, Techniques of EXAFS, SEXAFS, and XANES*. New York: Wiley.
- Park RL, Houston JE, and Schreiner DG (1970) A soft X-ray appearance potential spectrometer for the analysis of solid surfaces. *Review of Scientific Instrumentation* 41: 1810–1812.

## Desorption Techniques

**M Sasaki**, University of Tsukuba, Ibaraki, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Although many techniques have been used to analyze solid surfaces, most of them provide information on the static behavior of surface properties. By desorbing the surface materials and adsorbates into vacuum and by observing these species, we can obtain the dynamic features of the surface reaction and temporarily existing intermediates as well as the static behavior of the surface properties.

As in the case of other surface analytic methods, in order to examine well-defined surfaces, the sample surface must be prepared by cleaning and subsequent treatments such as adsorption and/or annealing processes in an ultrahigh vacuum (UHV). After that, the surface species must be desorbed from the surface by a specific excitation method and are analyzed by a specific detector. To induce desorption of the adsorbed species or the surface atoms, the surface is thermally excited by increasing the temperature (as in most cases) or electronically excited by irradiation

with electrons or photons. Although desorption can be induced by other methods such as energetic particle bombardment and application of electric fields, these are merely adapted to analyze surface properties. As a detector, a mass spectrometer is usually used. From the mass and the amount of desorbed species as a function of the experimental parameters, one obtains information on the surface properties and dynamic and temporal features of surface reactions. Optical methods are also used to selectively detect desorbed species.

In this article, the commonly used thermal desorption technique is first discussed in detail. Then other related methods are introduced briefly. Surface analyses based on electronic transition-induced desorption that provide unique information on the surface structure are presented.

## Thermal Desorption

By raising the surface temperature, adsorbates, trapped or bound on the surface, and surface atoms can be desorbed, even though they only exist on the surface during the surface reaction. Weakly bound surface species desorb early when the surface

states, nearest-neighbor configuration in the surface layer, and the mechanism of excitation and transport processes occurring in the surface region of the sample. Wide scope still exists for theoretical and experimental work in this field. When the different aspects of APS are fully exploited, this spectroscopy will become a popular technique for materials characterization of surfaces.

*See also:* **Surface Analysis:** Auger Electron Spectroscopy; Ion Scattering; Overview; X-Ray Photoelectron Spectroscopy.

## Further Reading

- Barco JL and Ferron J (1990) Information depth of Auger electron and appearance potential spectroscopies: A Monte Carlo study. *Surface Science* 232: 393–398.
- Briggs D and Seah MP (eds.) (1990) *Practical Surface Analysis by Auger and X-Ray Photoelectron Spectroscopy*, vol. 1. New York: Wiley.
- Chourasia AR and Chopra DR (1997) Auger electron spectroscopy. In: Settle F (ed.) *Handbook of Instrumental Techniques for Analytical Chemistry*, pp. 793–808. NJ: Prentice-Hall.
- Dose V and Fauster T (1979) Deconvolution of appearance potential spectra II. *Applied Physics* 20: 299–303.
- Gerlach RL, Houston JE, and Park RL (1970) Ionization spectroscopy of surfaces. *Applied Physics Letters* 16: 179–181.
- Henig C and Merz H (1998) Inner shell excitations of the correlated insulator MnO(100). *Journal of Electron Spectroscopy and Related Phenomena* 93: 189–192.
- Hinkers H, Stiller R, and Merz H (1989) Near-threshold excitation in light rare-earth compounds: A new interpretation of 3d appearance-potential spectra. *Physical Review B* 40: 10594–10603.
- Houston JE and Park RL (1972) Difference in the Cr L<sub>3</sub>/L<sub>2</sub> intensity ratio measured by soft X-ray and Auger-electron appearance potential spectroscopy. *Physical Review B* 5: 3808–3809.
- Kirshchner J and Staib P (1975) Disappearance potential spectroscopy. *Applied Physics* 6: 99–109.
- Koningsberger DC and Prins R (eds.) (1988) *X-Ray Absorption: Principles, Applications, Techniques of EXAFS, SEXAFS, and XANES*. New York: Wiley.
- Park RL, Houston JE, and Schreiner DG (1970) A soft X-ray appearance potential spectrometer for the analysis of solid surfaces. *Review of Scientific Instrumentation* 41: 1810–1812.

## Desorption Techniques

**M Sasaki**, University of Tsukuba, Ibaraki, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Although many techniques have been used to analyze solid surfaces, most of them provide information on the static behavior of surface properties. By desorbing the surface materials and adsorbates into vacuum and by observing these species, we can obtain the dynamic features of the surface reaction and temporarily existing intermediates as well as the static behavior of the surface properties.

As in the case of other surface analytic methods, in order to examine well-defined surfaces, the sample surface must be prepared by cleaning and subsequent treatments such as adsorption and/or annealing processes in an ultrahigh vacuum (UHV). After that, the surface species must be desorbed from the surface by a specific excitation method and are analyzed by a specific detector. To induce desorption of the adsorbed species or the surface atoms, the surface is thermally excited by increasing the temperature (as in most cases) or electronically excited by irradiation

with electrons or photons. Although desorption can be induced by other methods such as energetic particle bombardment and application of electric fields, these are merely adapted to analyze surface properties. As a detector, a mass spectrometer is usually used. From the mass and the amount of desorbed species as a function of the experimental parameters, one obtains information on the surface properties and dynamic and temporal features of surface reactions. Optical methods are also used to selectively detect desorbed species.

In this article, the commonly used thermal desorption technique is first discussed in detail. Then other related methods are introduced briefly. Surface analyses based on electronic transition-induced desorption that provide unique information on the surface structure are presented.

## Thermal Desorption

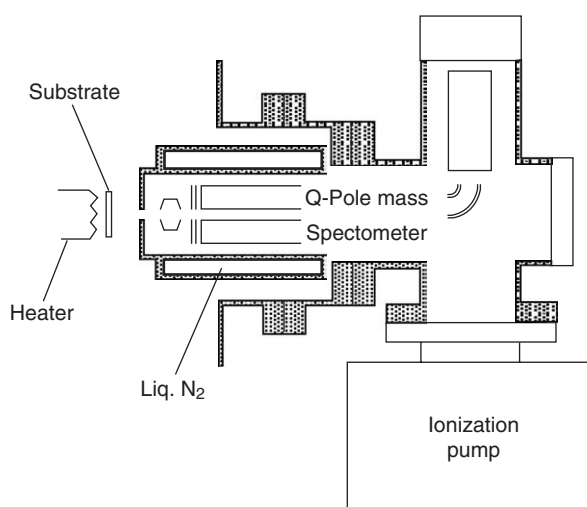
By raising the surface temperature, adsorbates, trapped or bound on the surface, and surface atoms can be desorbed, even though they only exist on the surface during the surface reaction. Weakly bound surface species desorb early when the surface

temperature is gradually increased. Thus, the amount of the desorbed species as a function of temperature shows the order of the weakness of the adsorption. This technique is called thermal desorption spectroscopy, temperature programmed desorption (TPD), or occasionally flash desorption. The temperature of desorption depends on the surface species, the adsorption site, and the surface properties as well as the speed of the temperature rise. Since desorption occasionally occurs through a surface reaction, the desorption temperature also depends on the detailed surface reaction. Thus, through the thermal desorption technique, one obtains a wide variety of information on the surface, such as the surface species, the number of them, the adsorption site, the surface structure, the binding energy of the adsorbates, and the interadsorbate interaction.

## Experimental Technique for Thermal Desorption

Thermal desorption can be performed with a rather simple UHV setup, consisting of a heater for raising the temperature of the sample surface, a detector for sensing the desorption species, and a gas doser for preparing the sample surface. A mass spectrometer is usually used as a detector and is helpful in identifying the desorbed species based on their mass spectrum. A multiplexer is often used to detect several species simultaneously (strictly speaking, sequentially with repeated quick scans) with a single temperature rise. It is necessary to minimize the temperature rise of the materials contained in the UHV system other than the specimen.

The concentration of adsorbates and surface atoms to be analyzed is limited to  $\sim 10^{15} \text{ cm}^{-2}$  for a single monolayer. Although the detection sensitivity of the desorbed species may be enhanced by using an integral method where the pumping speed is reduced and the desorbed species are accumulated, this method is no longer adopted at present because unknown factors due to the wall of the vacuum chamber hamper the interpretation of the results. Nowadays, one directly detects the species as they are ejected from the sample surface by using a system with a sufficiently high pumping speed. In this case, the detection sensitivity is crucially important. To detect species with high sensitivity, the ionizer of the mass spectrometer is placed just in front of the sample. In order to suppress the background signal, the ionizer is sometimes covered with a shield case with an aperture in the line of sight of the sample surface, which is optionally differentially pumped. The shield case is made of an inert material such as quartz, on which



**Figure 1** A typical setup for desorption measurements.

species are not adsorbed. To avoid detecting species desorbed from wall surfaces other than the sample, the shield case is sometimes cooled using liquid nitrogen as shown in **Figure 1**. Furthermore, a quick rise in the temperature improves the sensitivity. The heat capacity of the materials around the sample should be reduced. For a quantitative discussion, the temperature ramp rate needs to be carefully controlled, therefore a temperature controller with a temperature sensor and a feedback loop is occasionally adopted. Furthermore, a quick response is required for time-resolved measurements, which will be discussed later. In this case, the local pumping speed around the sample and the spectrometer should be made high enough by equipping them with a cryopanel or an open structure. In order to supply chemicals to the sample surface, a gas doser is used. A single capillary or multicapillary is sometimes used to restrict the gas-exposed area just around the sample surface and to align the direction and control the exposure distribution.

## Interpretation of Thermal Desorption Results

To extract quantitative information, mathematical considerations are necessary. Most simply, the peak position and the shape of the temperature programmed desorption spectrum are analyzed on the basis of the method proposed by Redhead. Here a sufficiently high pumping speed is assumed. Since the detected signal is proportional to the amount of the desorbed species under the experimental conditions, the desorption rate during a temperature rise is given

by the Polanyi–Wigner formula:

$$N(t) = -\frac{d\theta}{dt} = v_n \theta^n \exp\left(-\frac{E}{kT}\right) \quad [1]$$

where  $E$ ,  $n$ ,  $\theta$ , and  $v_n$  are the activation energy of desorption (binding energy of adsorption species), the order of desorption, the surface coverage, and the rate constant, respectively. For monomolecular and bimolecular desorptions,  $n$  is, respectively, 0 and 1.

Mostly a linear temperature rise,  $T = T_0 + \beta t$  is used, where  $\beta$  is the ramp rate of the temperature.  $E$ ,  $n$ , and  $v_n$  are assumed to be constant as a function of surface coverage, which may be valid only in the case of the low-coverage regime. The desorption rate being an exponential function of  $-E/kT$ , it increases with temperature. Since the desorption rate is proportional to the surface coverage of adsorbates as well, it decreases after reaching a low coverage. Thus, a desorption spectrum shows a peak. For the peak position ( $T_p$ ) of the TPD spectrum for  $n$  of 1 and 2, the following expressions hold:

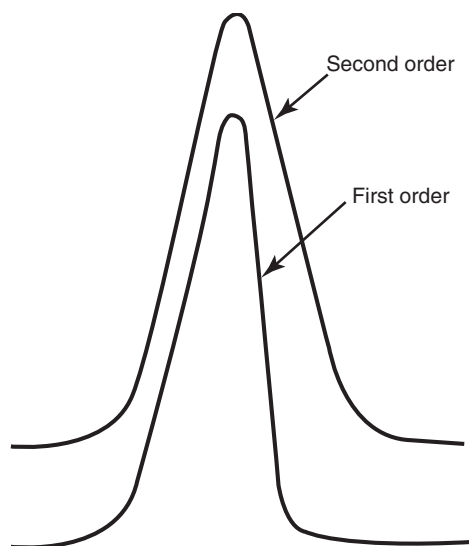
$$E/kT_p^2 = (v_1/\beta) \exp(-E/kT_p) \quad (n = 1) \quad [2]$$

$$E/kT_p^2 = (v_2\theta_0/\beta) \exp(-E/kT_p) \quad (n = 2) \quad [3]$$

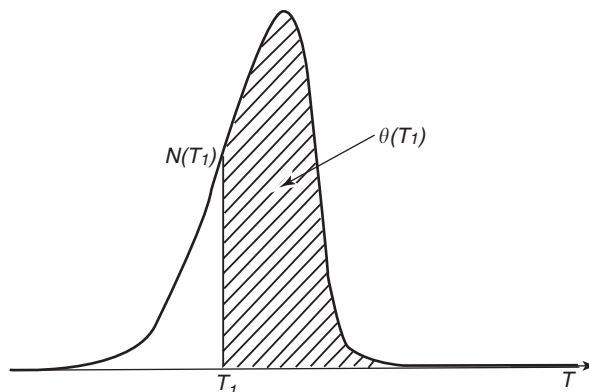
where  $\theta_0$  is the initial coverage of adsorbates. The activation energy is a key for discussing the reaction pathway. For analysis, it is derived from the peak position under an assumed rate constant of  $v = 10^{13} \text{ s}^{-1}$ . It should be noted that the peak temperature varies with the ramp rate of the temperature. Using a set of measurements from different ramping rates, the rate constant and the activation energy can be derived simultaneously.

For the first-order desorption, the peak position is independent of the initial coverage, and for desorption of the second or higher orders, it shifts toward a lower temperature with increasing coverage. Thus, from the peak positions for different coverages, we can obtain the reaction order. The peak shape of the TPD spectrum is also an interesting feature. While the peak is asymmetric for the first-order desorption, it is, in an ideal case, symmetric for second-order desorption as shown in Figure 2.

When the coverage is high, the activation energy and other parameters depend on the coverage, e.g., as a result of interadsorbate interactions and changes in atomic arrangement. Quantitative analyses for the desorption parameters that take into account coverage have been described on the basis of desorption rate isotherms, expressing the desorption rate as a function of the coverage for a constant temperature. The surface coverage during the TPD measurement



**Figure 2** Schematic shapes of TPD spectra for first- and second-order desorptions.



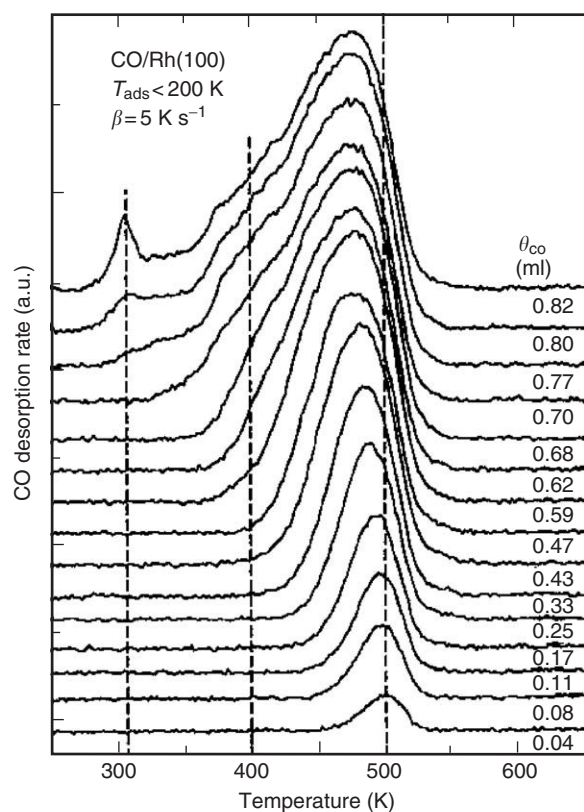
**Figure 3** Schematic illustration of the method of estimating the adsorbate coverage during TPD measurement from integration of the TPD spectrum.

(Figure 3) can be derived from the TPD spectrum obtained, using

$$\theta = A \int_T^\infty N dT \quad [4]$$

where  $A$  is the unit conversion constant. For such analysis, a set of TPD spectra obtained at different initial coverages are used. The example in Figure 4 shows a set of spectra. From these spectra, the desorption rate for a given coverage and temperature is derived and is called the desorption rate isotherm. For a given coverage, the logarithm of the desorption rate as a function of  $1/T$  can be plotted (Arrhenius plot). The slope of this plot gives the activation energy for a given coverage without assuming the desorption rate (the preexponential factor in the



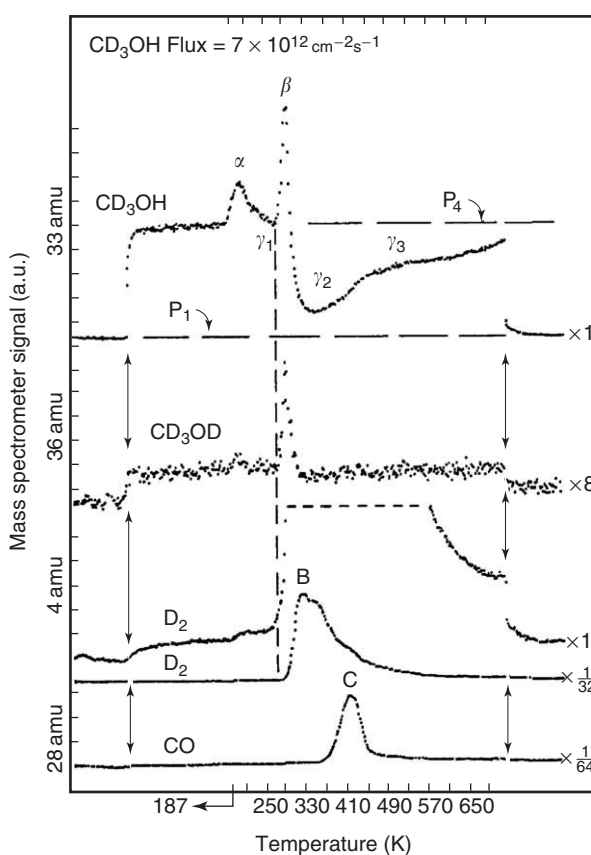


**Figure 4** Set of TPD spectra for various initial coverages; CO desorbed from a CO-adsorbed Rh(100) surface. The ramp rate is  $5 \text{ K s}^{-1}$ . (Reproduced with permission from van Bavel AP, Hops-taken MJP, Curulla D, *et al.* (2003) Quantification of lateral repulsion between co-adsorbed CO and N on Rh (100) using temperature-programmed desorption, low-energy electron diffraction, and Monte-Carlo simulations. *Journal of Chemical Physics* 119: 524–532; © American Institute of Physics.)

equation). From a set of TPD spectra for various temperature ramp rates one also obtains the desorption rate isotherm. Other methods such as threshold TPD have been proposed.

## Other Methods Related to Thermal Desorption

The surface reaction generally proceeds through multiple complex pathways. These pathways are unraveled by using the temperature-programmed technique. Under the experimental conditions of TPD, the reaction products and other surface species are desorbed from the sample surface by increasing the surface temperature after a sufficient amount of chemical species is adsorbed at a low temperature. This method is called the temperature programmed reaction method. Furthermore, we can observe the desorbed species during their exposure to the source



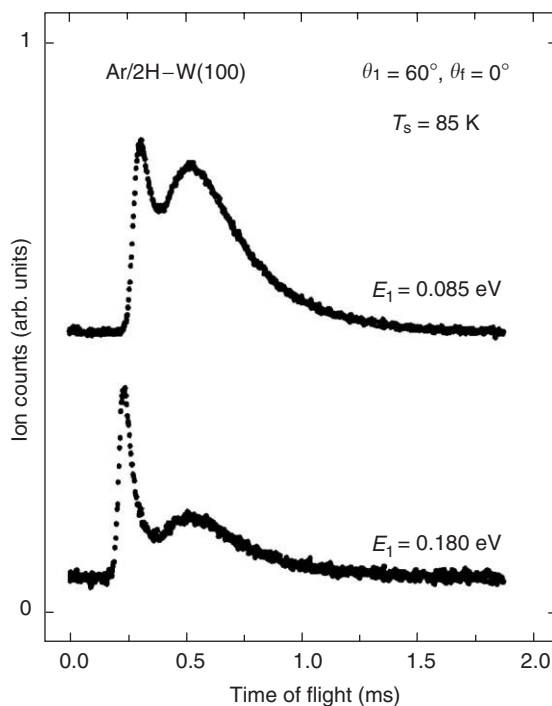
**Figure 5** Results of SKS measurements; the signal intensities of desorbed species from an Ni(111) surface under irradiation of  $\text{CD}_3\text{OH}$  as a function of the temperature. (Reproduced with permission from Gates SM, Russell JN Jr., and Yates JT Jr. (1985) Scanning kinetic spectroscopy (SKS): A new method for investigation of surface reaction processes. *Surface Science* 159: 233–255; © Elsevier.)

species. This method is called scanning kinetic spectroscopy (SKS). **Figure 5** shows examples of scanning kinetic spectra. In this method, the desorbed species are observed using a collimated quadrupole mass spectrometer when the temperature is raised at a constant ramp rate under molecular beam irradiation with a controlled flux.

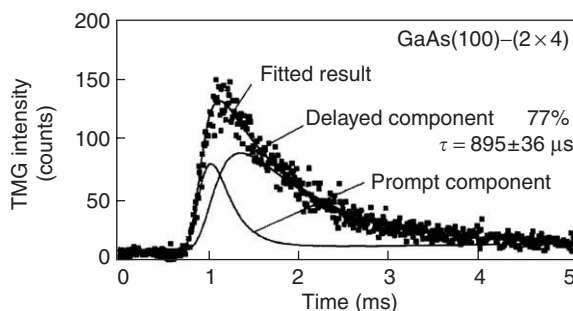
The desorption feature provides information on the reaction taking place at the surface. The velocity and angular distributions of desorbed species are measured by using a special setup with a rotating chopper with slits for a time-of-flight (TOF) measurement and a rotatable detector. A pseudorandom chopper is often used to perform time-resolved measurements efficiently that reflect the dynamical behavior of the surface reaction. The mechanism can be discussed on the basis of the detailed balance of desorption and adsorption. In the case of the absence of an activation barrier for adsorption, desorption shows the behavior for thermal equilibrium

conditions, i.e., an angular distribution under the cosine law and a Maxwellian velocity distribution at the surface temperature. On the other hand, when there is an activation barrier for adsorption, these distributions are different. Since the desorbed molecules are accelerated by the potential barrier for adsorption, the angular distribution is centered to the surface-normal and the velocity distribution is shifted to the higher-velocity side. The detailed behavior can be discussed in the framework of the multidimensional potential energy surface. As for the desorption resulting from surface reactions, the desorbed species are sometimes hyperthermally excited, showing specific velocity and angular distributions. It has been demonstrated that 'hot atoms' are generated during a surface reaction. In the case of misoriented surfaces, another feature is observed due to the anisotropy of the surface. In order to observe the vibrational and rotational states of the desorbed species, which are strongly related to the reaction mechanism, optical methods such as laser-induced fluorescence and resonance-enhanced multiphoton ionization are used to detect directly the electronically, vibrationally, and rotationally excited states of the desorbed species.

In order to obtain information on the dynamic behavior of the surface reaction, the modulated beam method is sometimes used. In this method, the desorbed species are detected using a time-resolved method, while the sample surface is irradiated with molecular beams that are temporarily modulated using a chopper or pulsed molecular beams. In the past, the distortion of the beam shape due to the scattering (adsorption/desorption) from the surface was mainly observed using, for example, a lock-in detection method where the dynamical feature of the reaction is described in terms of the phase shift. Recently the TOF technique has been used in combination with pulsed molecular beams and a multichannel scaler to obtain results that reflect the dynamical behavior of the inelastic surface scattering. If the molecules are trapped in a precursor state at the surface and energy equilibrium is attained, the velocity distribution should be the Maxwellian of the surface temperature; otherwise, it should be narrower than the Maxwellian, partially retaining the velocity distribution of the incident beams. It should be noted that the observed TOF spectrum is a convolution of the velocity distributions of the incident and scattered beams (Figure 6). From the TOF spectrum, one obtains the degree of energy exchange during scattering, which is important in discussion of the surface reaction in detail, especially for the direct reaction. In the case where the surface residence time is long compared with the time resolution of the



**Figure 6** TOF spectra of Ar scattered from the H-terminated W(100) surface for pulsed Ar beams with incident energies of 0.085 and 0.18 eV. The shape of the spectrum is well reproduced by summation of the direct inelastic and thermalized components. (Reproduced with permission from Rettner CT, Schweizer EK, and Mullins CB (1989) Desorption and trapping of argon at a 2H-W(100) surface and a test of the applicability of detailed balance to a nonequilibrium system. *Journal of Chemical Physics* 90: 3800–3813; © American Institute of Physics.)



**Figure 7** TOF spectrum of trimethylgallium (TMG) from the (2 × 4)-reconstructed GaAs(100) surface at 638 K when a pulsed TMG beam is supplied to the surface. The spectrum is well reproduced by convolution of the incident and scattered velocity distributions as well as a surface residence time of 0.9 ns. (Reproduced with permission from Sasaki M and Yoshida S (1996) Scattering of pulsed trimethylgallium beam from GaAs(100), –(110), and –(111) B surfaces. *Surface Science* 356: 233–246; © Elsevier.)

measurement, we can directly measure the residence time. In this case, the residence time is derived from deconvolution of the spectrum containing the surface residence time, as shown in Figure 7. Since the

precursor state does not have a barrier for adsorption, the Arrhenius plot of the inverse of the surface residence time provides the activation barrier as mentioned in the discussion of TPD.

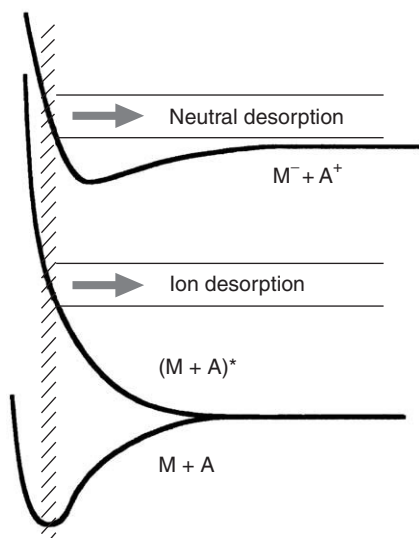
## Desorption Induced by Electronic Excitation

The surface species are also desorbed as ions or neutrals through electronic excitation induced by the irradiation of electrons or photons with energies from 10 eV to more than 1000 eV. This is called desorption induced by electronic transition (DIET). Photon-stimulated desorption shows a clear selection rule, and the results are easy to interpret. However, the number of photon energies available in the laboratory is limited. On the other hand, electron energies can be varied easily. Thus, electron-stimulated desorption is more widely used. By measuring the desorption rate through electronic excitations, one obtains information on the atomic geometry at the surface as well as the adsorbed species and the desorption mechanisms.

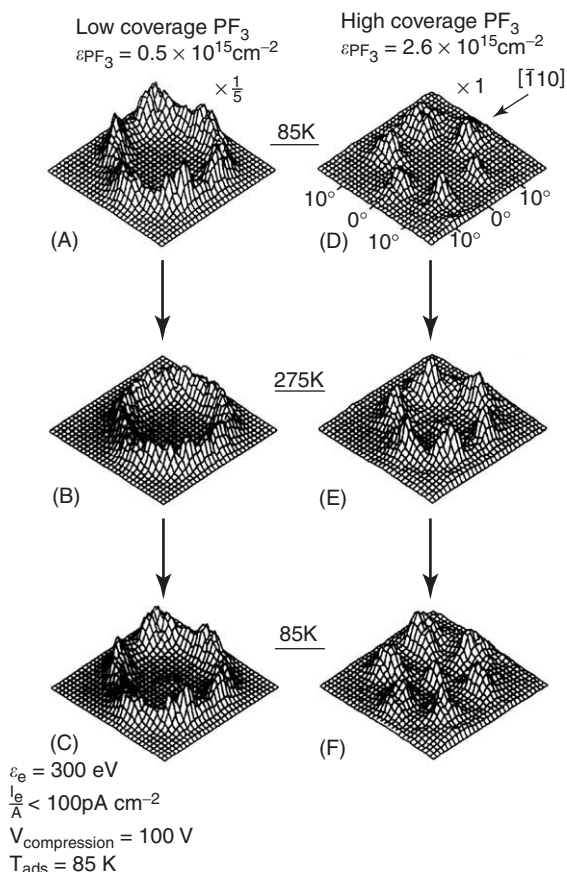
The mechanism of DIET was first discussed on the basis of the Franck–Condon principle. In this framework, the surface is electrically excited to a number of states, keeping the atomic location unchanged. In case of excitation to the repulsive part of the potential, ions or neutrals are released and desorbed when being accelerated by the repulsive potential as shown in Figure 8. This mechanism is called the Menzel–Gomer–Redhead mechanism. As for excitation by the impact of incident photons or electrons with

higher energies, inner shell electrons can be also excited. In this case, the electrons in a bonding state drop into the vacant inner shell, removing the same bonding state electron with a kinetic energy to conserve energy (Auger process). The surface species that are charged by losing a bonding electron are desorbed along the repulsive potential. This mechanism is called the Feibelman–Knotek model. In this mechanism, positive ions can be desorbed even from ionic crystals. The desorbing ions may be attracted from the surface by the image force. Occasionally, these species are neutralized and desorbed again if their lifetime is long enough.

The threshold energy or excitation function, as well as the velocity distribution of desorbed species, is a reflection of the desorption mechanism. From this information, the mechanism can be discussed. However, it should be taken into account that the



**Figure 8** Schematic illustration of the potential energy during DIETs.



**Figure 9** Typical result of ESDIAD for F ions desorbed from a  $\text{PF}_3$  preadsorbed Ni(111) surface with different initial coverages. The result at low coverage demonstrates that the  $\text{PF}_3$  rotates at a higher temperature. (Reproduced with permission from Yates JT, Alvey MD, Dresser MJ, Lanzillotto A-M, and Uram KJ (1988) Observation of molecular rotors on surfaces by ESDIAD: Studies of  $\text{PF}_3$  and  $\text{NH}_3$  chemisorption on Ni surfaces. In: *Desorption Induced by Electronic Transitions DIET III*, pp.100–108. Berlin: Springer-Verlag.)

spectra obtained are often broadened or shifted with respect to that for the gas-phase reaction due to the surface effect.

In the case of electronic excitation, the transition occurs in accord with the Franck–Condon principle. Here, the surface species are suddenly exposed to the repulsion potential, maintained at the same location as that just before excitation. In this case, the species are desorbed to the direction opposite to the broken bond. By using a setup with a spherical grid mesh generating a radial electric field and a fluorescent screen with a microchannel plate, one obtains 2D images corresponding to the atomic arrangement (Figure 9). This method is called electron-stimulated desorption ion angular distribution (ESDIAD). By using this method, it has been demonstrated that an adsorbate with a fixed orientation to the substrate in a low-temperature regime starts to rotate in a high-temperature regime.

The surface species can be desorbed also by a strong electric field of the order of  $10^{10} \text{ V m}^{-1}$ . It is well known that the electric field is then enhanced at the apex of the protrusion. This method is used in field ion microscopy.

**See also:** **Infrared Spectroscopy:** Overview; Photothermal. **Mass Spectrometry:** Overview; Time-of-Flight. **Microscopy Techniques:** Electron Microscopy. **Surface Analysis:** Overview.

## Further Reading

Asscher M and Somorjai GA (1992p) Reactive scattering. In: Scoles G (ed.) *Atomic and Molecular Beam Methods*,

vol. 2, pp. 488–514. New York: Oxford University Press.

Auerbach DH (1992p) Multiple-phonon inelastic scattering. In: Scoles G (ed.) *Atomic and Molecular Beam Methods*, vol. 2, pp. 444–462: Oxford University Press.

Falconer J and Madix RJ (1977) Desorption rate isotherm in flash desorption analysis. *Journal of Catalysis* 48: 262–268.

Gates SM, Russell JN Jr., and Yates JT Jr. (1985) Scanning kinetic spectroscopy (SKS): A new method for investigation of surface reaction processes. *Surface Science* 159: 233–255.

Head-Gordon M, Tully JC, Rettner CT, Mullins CB, and Auerbach DJ (1991) On the nature of trapping and desorption at high surface temperatures. Theory and experiments for the Ar–Pt(111) system. *Journal of Chemical Physics* 94: 1516–1527.

Madix RJ and Benziger J (1978) Kinetic processes on metal single-crystal surfaces. *Annual Review Physical Chemistry* 29: 285–306.

Matsushima T (2003) Angle-resolved measurements of product desorption and reaction dynamics on individual sites. *Surface Science Reports* 52: 1–62.

Miller JB, Siddiqui HR, Gates SM, *et al.* (1987) Extraction of kinetic parameters in temperature programmed desorption: A comparison of methods. *Journal of Chemical Physics* 87: 6725–6732.

Redhead PA (1962) Thermal desorption of gases. *Vacuum* 12: 203–211.

Yates JT Jr., Alvey MD, Dresser MJ, Lanzillotto A-M, and Uram KJ (1988) Observation of molecular rotors on surfaces by ESDIAD: Studies of  $\text{PF}_3$  and  $\text{NH}_3$  chemisorption on Ni surfaces. In: Stulen RH and Knotek ML (eds.) *Desorption Induced by Electronic Transitions DIET III*, pp. 100–108. Berlin: Springer.

## Ion Scattering

I Stensgaard, University of Aarhus, Aarhus C, Denmark

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Ion scattering is concerned with accelerating ions, causing them to impinge on a sample as a well-collimated beam, and analyzing the angle and energy distributions of the scattered ions. Ion scattering distinguishes itself by being a relatively simple technique, which can be applied for the analysis of both surface composition and surface structure. The technique was developed in the late 1960s and has since then matured to become an important research and analysis tool in many laboratories.

Traditionally, there has been a distinction between two different energy regimes for ion scattering as applied to surface analysis. At high ion energies (larger than  $\geq 100 \text{ keV}$ ), the analysis is normally carried out with protons or helium ions, the scattering process is determined by the Rutherford cross-section, and the scattered particles suffer little neutralization along the exit path. Depending upon the exact application, the techniques are termed Rutherford backscattering spectrometry (RBS), high-energy ion scattering (HEIS), or medium-energy ion scattering (MEIS). On the other hand, low-energy ion scattering (LEIS) or ion-scattering spectroscopy (ISS), as the technique is also known, is based on scattering of either inert-gas ions (He or Ne) or alkali ions (Li, Na, or K) with energies of the

spectra obtained are often broadened or shifted with respect to that for the gas-phase reaction due to the surface effect.

In the case of electronic excitation, the transition occurs in accord with the Franck–Condon principle. Here, the surface species are suddenly exposed to the repulsion potential, maintained at the same location as that just before excitation. In this case, the species are desorbed to the direction opposite to the broken bond. By using a setup with a spherical grid mesh generating a radial electric field and a fluorescent screen with a microchannel plate, one obtains 2D images corresponding to the atomic arrangement (Figure 9). This method is called electron-stimulated desorption ion angular distribution (ESDIAD). By using this method, it has been demonstrated that an adsorbate with a fixed orientation to the substrate in a low-temperature regime starts to rotate in a high-temperature regime.

The surface species can be desorbed also by a strong electric field of the order of  $10^{10} \text{ V m}^{-1}$ . It is well known that the electric field is then enhanced at the apex of the protrusion. This method is used in field ion microscopy.

**See also:** **Infrared Spectroscopy:** Overview; Photothermal. **Mass Spectrometry:** Overview; Time-of-Flight. **Microscopy Techniques:** Electron Microscopy. **Surface Analysis:** Overview.

## Further Reading

Asscher M and Somorjai GA (1992p) Reactive scattering. In: Scoles G (ed.) *Atomic and Molecular Beam Methods*,

vol. 2, pp. 488–514. New York: Oxford University Press.

Auerbach DH (1992p) Multiple-phonon inelastic scattering. In: Scoles G (ed.) *Atomic and Molecular Beam Methods*, vol. 2, pp. 444–462: Oxford University Press.

Falconer J and Madix RJ (1977) Desorption rate isotherm in flash desorption analysis. *Journal of Catalysis* 48: 262–268.

Gates SM, Russell JN Jr., and Yates JT Jr. (1985) Scanning kinetic spectroscopy (SKS): A new method for investigation of surface reaction processes. *Surface Science* 159: 233–255.

Head-Gordon M, Tully JC, Rettner CT, Mullins CB, and Auerbach DJ (1991) On the nature of trapping and desorption at high surface temperatures. Theory and experiments for the Ar–Pt(111) system. *Journal of Chemical Physics* 94: 1516–1527.

Madix RJ and Benziger J (1978) Kinetic processes on metal single-crystal surfaces. *Annual Review Physical Chemistry* 29: 285–306.

Matsushima T (2003) Angle-resolved measurements of product desorption and reaction dynamics on individual sites. *Surface Science Reports* 52: 1–62.

Miller JB, Siddiqui HR, Gates SM, *et al.* (1987) Extraction of kinetic parameters in temperature programmed desorption: A comparison of methods. *Journal of Chemical Physics* 87: 6725–6732.

Redhead PA (1962) Thermal desorption of gases. *Vacuum* 12: 203–211.

Yates JT Jr., Alvey MD, Dresser MJ, Lanzillotto A-M, and Uram KJ (1988) Observation of molecular rotors on surfaces by ESDIAD: Studies of  $\text{PF}_3$  and  $\text{NH}_3$  chemisorption on Ni surfaces. In: Stulen RH and Knotek ML (eds.) *Desorption Induced by Electronic Transitions DIET III*, pp. 100–108. Berlin: Springer.

## Ion Scattering

I Stensgaard, University of Aarhus, Aarhus C, Denmark

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Ion scattering is concerned with accelerating ions, causing them to impinge on a sample as a well-collimated beam, and analyzing the angle and energy distributions of the scattered ions. Ion scattering distinguishes itself by being a relatively simple technique, which can be applied for the analysis of both surface composition and surface structure. The technique was developed in the late 1960s and has since then matured to become an important research and analysis tool in many laboratories.

Traditionally, there has been a distinction between two different energy regimes for ion scattering as applied to surface analysis. At high ion energies (larger than  $\geq 100 \text{ keV}$ ), the analysis is normally carried out with protons or helium ions, the scattering process is determined by the Rutherford cross-section, and the scattered particles suffer little neutralization along the exit path. Depending upon the exact application, the techniques are termed Rutherford backscattering spectrometry (RBS), high-energy ion scattering (HEIS), or medium-energy ion scattering (MEIS). On the other hand, low-energy ion scattering (LEIS) or ion-scattering spectroscopy (ISS), as the technique is also known, is based on scattering of either inert-gas ions (He or Ne) or alkali ions (Li, Na, or K) with energies of the



order of 1 keV. At this energy, there are strong screening effects on the Coulomb potential, and a significant, frequently path-dependent neutralization of the scattered particles is observed.

As will be evident from the following, there are two major differences between the two energy regimes. Quantitative measurements of composition are easier in the high-energy regime than in the low-energy regime. Also, in structural analysis HEIS and MEIS determine the position of surface atoms relative to the bulk-atom positions, while LEIS (ISS) measurements reveal relative positions between surface atoms.

## Theory

### Surface-Composition Analysis

Ion-scattering techniques are well suited for a quantitative analysis of the surface composition because the energy spectrum of scattered particles shows mass dispersion, and the scattering process has a well-defined cross-section at high energies. The two main techniques applicable in the high-energy regime are RBS and nuclear reaction analysis (NRA). RBS is most commonly used in cases where the atomic mass of a surface impurity or an adsorbate is larger than that of the substrate. Concentrations of light elements may be determined with NRA.

**High-energy regime** RBS is the subject of a separate chapter of this encyclopedia. Here, a short description of RBS is presented, with the emphasis on applications for surface analysis in the monolayer regime.

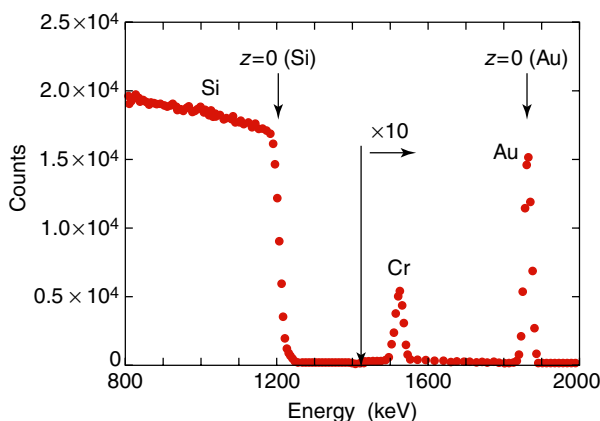
The energy  $E_1$  of the scattered particle is proportional to the incident energy  $E_0$ :

$$E_1 = KE_0 \quad [1]$$

where  $K$ , the kinematic factor, is given by

$$K = \left( \frac{(M_2^2 - M_1^2 \sin^2 \theta)^{1/2} + M_1 \cos \theta}{M_1 + M_2} \right)^2 \quad [2]$$

$M_1$  and  $M_2$  are the atomic masses of the incident ions and the target atoms, respectively, and  $\theta$  is the scattering angle in the laboratory system. This property is illustrated in **Figure 1**, which shows a  $^4\text{He}^+$  backscattering spectrum from a Si crystal covered with very thin layers of Cr and Au. Ions backscattered from the Si surface appear at  $\approx 1200$  keV. The continuum at lower energies is due to ions backscattered from Si after penetration into the crystal, leading to an energy loss. The two peaks corresponding to backscattering from Cr and from Au are clearly resolved and appear at energies determined by eqns [1] and [2]. The masses



**Figure 1** Rutherford backscattering spectrum ( $E_0 = 2$  MeV,  $\theta_{\text{LAB}} = 140^\circ$ ) from an Si crystal with thin layers of Cr ( $3 \times 10^{15}$  atoms  $\text{cm}^{-2}$ ) and Au ( $8 \times 10^{14}$  atoms  $\text{cm}^{-2}$ ) evaporated onto the surface. The continuum at low energies is due to ions backscattered from Si after penetration into the bulk. The energy of such ions can be related to the scattering depth.

of the surface constituents can therefore be identified with an energy-calibrated detection system. Unfortunately, the mass resolution is inherently poorer for heavy target atoms than for light ones.

For a quantitative analysis we consider the differential scattering cross-section in the laboratory reference frame, given by

$$\frac{d\sigma}{d\Omega} = \left( \frac{Z_1 Z_2 e^2}{2E} \right)^2 \frac{1}{\sin^4 \theta} \frac{\{[1 - ((M_1/M_2) \sin \theta)^2]^{1/2} + \cos \theta\}^2}{[1 - ((M_1/M_2) \sin \theta)^2]^{1/2}} \quad [3]$$

where  $Z_1$  and  $Z_2$  are the atomic numbers of the incident particle and target atom, respectively,  $E$  is the particle energy just before scattering, and  $e$  is the electron charge. Hence, the scattering yield  $Y$  from a given adsorbate of area density  $N_t$  (atoms  $\text{cm}^{-2}$ ) for a total number  $N_i$  of incident particles will be

$$Y = N_i N_t \left( \frac{d\sigma}{d\Omega} \right) \Delta\Omega \quad [4]$$

where  $\Delta\Omega$  is the solid angle of the detector system. Absolute densities of different elements present at the surface can therefore be determined.

Because of the small cross-sections, Rutherford backscattering is not well suited for an accurate determination of the abundance of light elements such as carbon and oxygen. Instead, NRA, sometimes called nuclear microanalysis, can be applied. The energy released in the reactions (the  $Q$  value) is quite high, resulting in high-energy reaction products that can be easily detected. In this case, the cross-sections



**Table 1** Nuclear reactions used for detection of light surface adsorbates

Element	Reaction
d( <sup>2</sup> H)	d( <sup>3</sup> He, p) $\alpha$
<sup>12</sup> C	<sup>12</sup> C(d, p) <sup>13</sup> C
<sup>14</sup> N	<sup>14</sup> N(d, p) <sup>15</sup> N
<sup>14</sup> N	<sup>14</sup> N(d, $\alpha$ ) <sup>12</sup> C
<sup>16</sup> O	<sup>16</sup> O(d, p) <sup>17</sup> O
<sup>16</sup> O	<sup>16</sup> O( <sup>3</sup> He, $\alpha$ ) <sup>15</sup> O
<sup>16</sup> O	<sup>16</sup> O( <sup>3</sup> He, p) <sup>18</sup> F
<sup>18</sup> O	<sup>18</sup> O(p, $\alpha$ ) <sup>15</sup> N

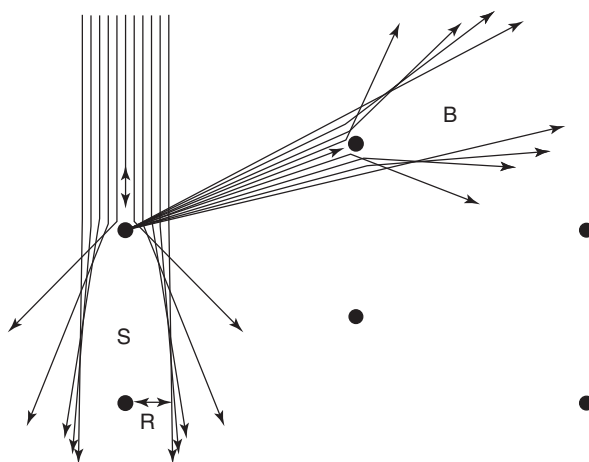
have no simple energy dependence. Table 1 lists some of the reactions used in surface analysis.

**Low-energy regime** Here, quantitative surface-composition analysis is less straightforward. The energy spectrum of scattered ions immediately gives a qualitative impression of the surface constituents, but it is difficult to convert the observed yields to absolute coverages for the following reasons: (1) the scattering cross-sections are less well known due to the strong screening of the Coulomb potential; (2) in most experimental systems only scattered ions are detected; and (3) shadowing effects may be important. The probability that an incident ion reaches the analyzer without neutralization depends on several parameters such as ion type, scattering depth, and the detailed trajectory. Some of the difficulties can be overcome by the choice of specific ions, application of time-of-flight (TOF) techniques, comparison to calibration samples, etc., but quantitative analysis is, in general, not straightforward in the low-energy regime.

However, LEIS (ISS) offers one advantage: an extreme surface sensitivity can be obtained with He ions because ions scattered at subsurface atoms will neutralize at the exit path. If only scattered ions are detected, the signal will therefore originate exclusively from the surface monolayer of atoms. This property is extremely useful in, for example, studies of surface segregation.

### Surface-Structure Analysis

Suppose that a beam of well-collimated, high-energy ions is incident on a crystalline substrate along a major crystallographic axis as illustrated schematically in Figure 2. Ions incident at the lowest impact parameters (the impact parameter is the distance at which the incident ion would pass the atom if no deflection took place) will undergo large deflections and scatter out of the crystal, while those at large impact parameters suffer small deflections and determine the flux of ions into the crystal. As a



**Figure 2** Formation of a shadow cone (S) behind a surface atom for beam incidence along a (vertical) row of atoms. The radius of the shadow cone at the second atom is denoted R. Additional deflections at the second atom have been neglected in the figure. As shown schematically, scattered ions are blocked in certain directions giving rise to a blocking cone (B).

consequence, a so-called shadow cone is formed behind the first atom, i.e., a region in space where, in a static representation, the incident flux is zero. The width of the shadow cone depends on the atomic numbers of the projectile and target atoms, and it increases with decreasing energy.

**High-energy regime** In the high-energy regime, a typical shadow-cone radius measured at the position of the second atom is  $\approx 0.1$  Å. For most materials, the one-dimensional, root-mean-square (rms) vibration amplitude of the atoms at room temperature is also  $\sim 0.1$  Å. The backscattering probabilities from the atoms behind the surface atom must therefore be taken into account. They may be calculated by a convolution of the flux distribution with the position distribution of the atoms. In general, such convolutions cannot be carried out on an analytical basis, and the calculation of backscattering probabilities is therefore based on Monte-Carlo-type computer simulations of a large number of ion trajectories.

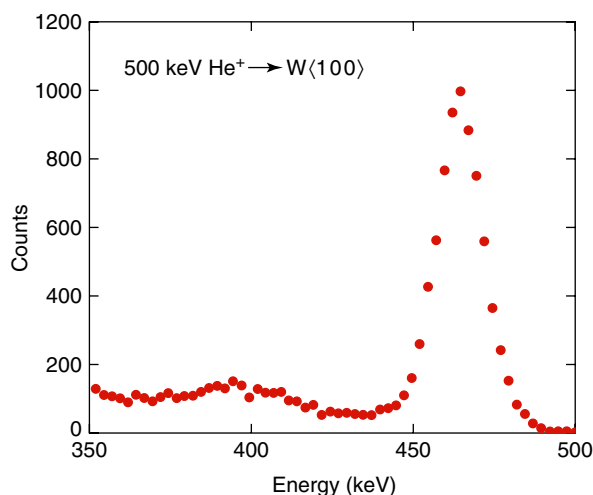
Instead of calculating only actual backscattering events (which are very rare because of the small cross-sections), the backscattering probabilities for the atoms along the row are determined. Assuming that the row of atoms coincides with the  $z$ -axis, the normalized probability can be shown to take the following form:

$$P = \frac{\cos \psi}{2\pi\sigma^2 NdN_i} \sum_{i=1}^{N_i} \exp\left(-\frac{x_i^2 + y_i^2}{2\sigma^2}\right) \quad [5]$$

where  $\psi$  is the angle of incidence with the atomic row,  $\sigma$  the one-dimensional rms vibration amplitude,  $d$  the interatomic distance along the row,  $N$  the atomic density in the crystal,  $N_i$  the number of ion trajectories, and  $(x_i, y_i)$  determines the impact parameter.

While the outermost atoms are fully exposed to the incoming beam, the backscattering probability becomes negligible a few atomic distances into the crystal due to the shadowing effect. This strong reduction in scattering yield is illustrated in Figure 3, which shows a backscattering spectrum obtained on a W crystal for ion-beam incidence along a  $\langle 100 \rangle$  direction. The high-energy end of the spectrum exhibits a peak that corresponds to backscattering from the surface atoms. As shown in the applications section, information about the surface structure can be derived from an analysis of the energy and/or angular dependence of this surface peak. This is the basis of HEIS for surface structure analysis.

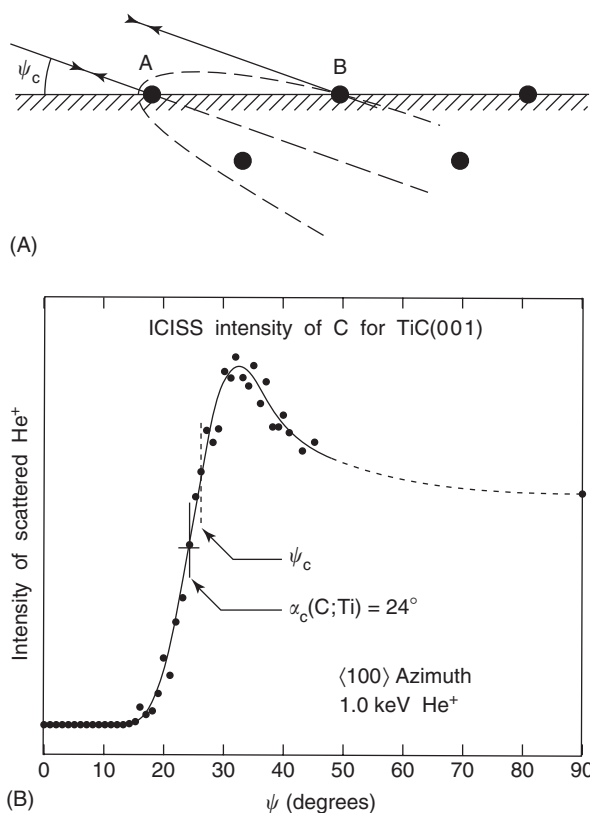
If backscattering takes place from a subsurface atom, this atom will act as a point source emitting particles, and the outgoing ions will be blocked in certain directions where a blocking cone is formed by a process similar to shadow-cone formation. This effect is also illustrated in Figure 2. To investigate the surface crystallography with MEIS, the angular distribution of the backscattered flux of particles is measured with the incident beam aligned to a major crystallographic axis. By a relatively simple extension, the calculation scheme outlined above may be used for the case of double alignment (shadowing and blocking) as well. Structural parameters are derived by optimizing the fit between the measured



**Figure 3** Backscattering spectrum for incidence of a 500 keV  $\text{He}^+$  ion beam along a  $\langle 100 \rangle$  direction of a W crystal. A surface peak, corresponding to backscattering from the surface layers, appears in the high-energy end of the spectrum.

angular distribution of scattered ions ('blocking spectrum') and calculated distributions for model structures.

**Low-energy regime** In the low-energy regime, quite similar directional effects are observed; however, due to the much lower energy, the shadow cones are considerably wider. This is exploited in a special version of LEIS, called impact collision ion scattering spectroscopy (ICISS) where a scattering angle close to  $180^\circ$  is used (in practice  $\geq 160^\circ$ ). In this mode, backscattered ions have experienced head-on collisions with the target atoms, and relative positions of surface atoms can be determined in the following manner (see Figure 4): If the angle of incidence  $\psi$  (measured relative to the surface plane) is below a critical value  $\psi_c$ , a surface atom B will be in the shadow cone from its nearest neighbor A, and consequently, no backscattering from the surface is



**Figure 4** Basic principle of ICISS. (A) For beam incidence at an angle  $\psi_c$  the edge of the shadow cone from atom A coincides with the position of atom B. (B) Experimental ICISS spectrum showing the scattering intensity from C in the TiC(001) surface as a function of angle of incidence  $\psi$ . The critical angle  $\psi_c$  is indicated. (Part B is reproduced with permission from Aono M and Souda R (1985) Quantitative surface atomic structure analysis. *Japanese Journal of Applied Physics* 24: 1249; © Institute of Pure and Applied Physics, Tokyo, Japan.)

possible. When  $\psi$  is increased, a sudden onset of backscattering intensity is observed around  $\psi_c$ , at which angle the edge of the shadow cone from atom A coincides with the position of atom B. If the detailed shape of the shadow cone is known (from theory or from measurements on known structures), the relative positions of atoms A and B can be determined, as will be discussed in the Applications section. Positions of subsurface atoms relative to surface atoms may be determined as well, although the analysis becomes less straightforward due to multiple deflections of the ions.

## Instrumentation

A detailed description of the experimental systems used for ion-beam crystallography cannot be included in the present article. Basically, an ultra-high-vacuum (UHV) system is coupled to a high- or low-energy ion accelerator via a differentially pumped beam line. There are two features that set the ion-scattering systems apart from most other UHV setups. One is a goniometer (high-precision manipulator), which allows the necessary, precise orientation of the crystalline sample relative to the incident ion beam over a large angular range. The other is some kind of detection system that provides an energy dispersion of the scattered ions.

In the high-energy regime, bakeable, implanted Si detectors are very convenient for this purpose and they have been used extensively in HEIS. They detect both neutral and charged scattered particles but have typical energy resolutions of 5 keV for protons and 15 keV for He particles, corresponding to a depth resolution  $\delta x > 10$  nm. Therefore, energy analysis of the backscattered particles with solid-state detectors does not provide a distinction between, for example, adsorption at surface and subsurface sites. Much better depth (and mass) resolution can be achieved with magnetic or electrostatic analyzers where energy resolutions of  $\Delta E/E \approx 3 \times 10^{-3}$  are typical. For MEIS toroidally shaped electrostatic analyzers are available, which, when equipped with a position sensitive detector, permit detection of ions in a large interval of scattering angles, thereby allowing a full angular distribution to be recorded at a fixed analyzer setting. To limit the size of the electrostatic analyzer, MEIS is typically performed at energies  $\sim 100$  keV. Energy resolutions below  $10^{-3}$  have been achieved, which in conjunction with special scattering geometries has led to monolayer depth resolution!

In the low-energy regime, electrostatic analyzers are most commonly used for energy analysis, but TOF

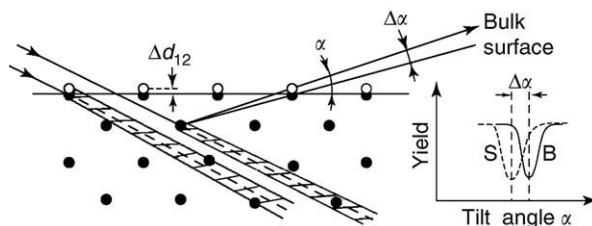
systems are also widespread. Such systems can detect both ions and neutrals, resulting in an increased sensitivity. However, for He scattering, where particles that are scattered beneath the surface layer are very effectively neutralized in the exit path, a loss of surface sensitivity is incurred in TOF systems.

## Applications

### Surface-Structure Analysis: Relaxation and Reconstruction

The termination of a solid by a surface modifies the atomic surroundings in terms of the atomic coordination and the distribution of conduction electrons. It is then to be expected that the equilibrium positions of the surface atoms will change in order to minimize the surface free energy. If the entire surface layer is shifted collectively inward or outward, the surface is said to be relaxed (contracted or expanded). If the atomic arrangement within the surface plane changes, the surface is said to be reconstructed. Both relaxation and reconstruction phenomena can be analyzed by ion scattering.

**Relaxation** Many structural problems can be addressed experimentally in both the high- and the low-energy regimes. As an example, consider surface relaxation. The application of the shadowing and blocking technique in MEIS to measure relaxations is illustrated in **Figure 5**, which shows (in a schematic way) the atomic configuration in a plane perpendicular to a face centered cubic (fcc) (001) surface that is assumed to be relaxed (contracted) by  $\Delta d_{12}$ . The ion beam is made incident along a crystal axis, and the scattered beam intensity is measured in an angular interval around an exit direction that coincides with a bulk axis. Any relaxation of the surface will result in an angular shift  $\Delta\alpha$  of the surface blocking cone away from the direction of the bulk axis, as observed in the recorded blocking spectrum.

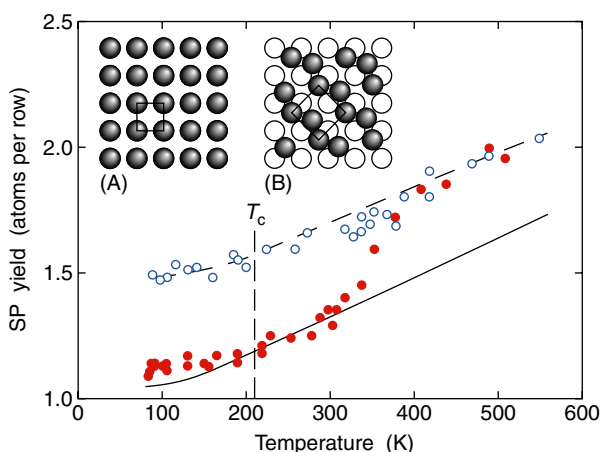


**Figure 5** Basic principles for analysis of surface relaxation based on double alignment (shadowing and blocking). A relaxation will give rise to an angular shift between the blocking directions for bulk and for surface scattering.

This technique relies to a high extent on simple geometry in the basic analysis in the sense that the angular shift of the surface blocking dip is relatively insensitive to the magnitude of the surface vibrations, which, however, influence the shape and depth of the dip. However, to derive the exact geometry a detailed comparison between the experimental blocking spectrum and data calculated for various model structures (including multilayer relaxations) must be carried out for the scattering data in the full angular interval.

A special type of relaxation is observed on some compound surfaces where the two kinds of constituent atoms have opposite (inward and outward) displacements, resulting in a so-called 'rumpling' relaxation. This type of relaxation can be analyzed with low-energy ions using ICISS, as illustrated in **Figure 4B**, which shows some results from a study of the TiC(001) surface using 1 keV He<sup>+</sup> ions. The intensity of He<sup>+</sup> scattered from C atoms is shown as a function of angle of incidence measured from the surface, and the onset of scattering intensity is well defined although somewhat smeared out because of the thermal vibrations. It can be shown that the angle corresponding to roughly 80% of the maximum intensity is a better estimate of  $\psi_c$  than the angle corresponding to half-maximum intensity. Based on this rule,  $\psi_c$  is estimated to be  $\sim 26^\circ$ . Using the fact that the shape of the shadow cone from Ti (atom A in **Figure 4A**) is known from other measurements, the height of the C atoms (atom B) can then be determined relative to Ti. The result indicates that the rumpling of the TiC(001) surface is negligible ( $\leq 0.01$  nm).

**Reconstruction** The simple observations made in the section 'Theory: Surface structure analysis' can be used to determine the nature of phase transitions involving surface reconstruction. The W(100) surface and its interaction with hydrogen will serve as an example. Below a critical temperature  $T_c$  of  $\approx 210$  K, the clean W(100) surface is known to exhibit a  $(\sqrt{2} \times \sqrt{2})R45$  LEED pattern that corresponds to a reconstructed surface in which the surface atoms have significant lateral displacements. By saturating the surface with hydrogen or by heating the crystal above  $T_c$ , the LEED pattern changes to a  $(1 \times 1)$  pattern. The question is then whether this  $(1 \times 1)$  pattern corresponds to a truncated bulk surface. Lateral displacements of surface atoms will expose underlying atoms to the ion beam at normal incidence and hence increase the effective number of atoms per row, visible to the ion beam, above that expected for a truncated bulk surface. Some results from an ion-scattering study of the different phases



**Figure 6** Experimental surface-peak yields for normal incidence on W(100) as a function of temperature, measured with 1 MeV He ions. Open circles are for the clean surface, solid circles for the hydrogen exposed surface. The solid line is a computer simulation for an unreconstructed surface. The dashed line corresponds to a case in which all surface atoms have a lateral displacement component of 0.15 Å.  $T_c$  is the critical temperature for the phase transition. (Reproduced with permission from Stensgaard I, Purcell KG, and King D (1989) Evidence for a temperature induced order-disorder phase transition on W(100). *Physical Review B* 39: 897; © American Physical Society.)

are shown in **Figure 6**. The surface-peak yield was measured at normal incidence for the surface in the clean and the hydrogen-exposed states. It is evident that the data points for the clean surface (open circles) lie well above the calculated yield for the truncated bulk surface (solid line) and that the data points do not drop to this level above  $T_c$ . In fact, there is no discontinuity in yield across the transition-temperature region. This indicates that the surface atoms have similar lateral displacements in the high-temperature and the low-temperature phases that are therefore both reconstructed. The detailed nature of the phase transition (for example, whether it is a genuine order-disorder transition) and the structure of the high-temperature phase are topics beyond the scope of this article, but clearly the very simple measurements described above represent some important constraints on possible models.

**Figure 6** also displays a series of surface peak measurements (solid circles) as a function of temperature for a hydrogen exposure, which, if the hydrogen atoms stick to the surface, is sufficiently large to saturate the surface. The adsorption of hydrogen below  $\approx 320$  K obviously causes the surface to become unreconstructed, i.e., that the  $(\sqrt{2} \times \sqrt{2})R45$  low-temperature reconstructed phase converts into a truncated bulk structure. Above  $\approx 350$  K, it is not possible to saturate the W(100) surface with hydrogen.

How large are the lateral displacements of the surface atoms in the reconstructed phases? Some information on this problem can be deduced from a measurement of the number of visible surface atoms (surface peak yield) as a function of energy. The shadow-cone radius scales roughly inversely with the square root of the ion energy. For a given transverse displacement, the shadow cone is sufficiently wide at low energy to shadow the underlying atom, which, in turn, becomes fully exposed at high energies. In practice, the thermal vibrations tend to smear out the transition from the fully shadowed to the fully exposed state but the lateral displacements can still be deduced by comparison to simulated surface yields.

### Chemisorption Positions

A detailed knowledge of the positions of chemisorbed atoms is a necessary prerequisite for an understanding of most surface properties. The chemisorption positions can be derived from ion scattering in both the high- and the low-energy regimes.

In general, pure shadowing is not very well suited for determination of adsorption positions with high-energy ions. The adsorbates may shadow the substrate-surface atoms in a certain direction of incidence (cf. Epitaxy, below), but unless this direction coincides with a major crystal axis, the shadowing effect is small. Also, the thermal vibrations will reduce the shadowing.

Shadowing and blocking using MEIS can be applied if the adsorbate gives rise to a significant blocking dip, i.e., if the atomic number is not very small and the thermal vibrations are not too large. A full surface-structure analysis must then be undertaken to incorporate a possible adsorption-induced relaxation.

In the low-energy case, the analysis may proceed much along the same lines as described above for an ICISS analysis of the TiC(001) surface (the C atoms could be thought of as being chemisorbed on a Ti surface) although a more thorough analysis, of course, involves a complete scan of the incident angle as well as angular scans along different azimuthal directions.

### Epitaxy

The epitaxial growth of one material onto another (heteroepitaxy) is extremely important, especially in the field of semiconductors. The quality of relatively thick films can be investigated by the ion-channeling process, but ion scattering has also provided several examples of very basic studies of the initial stages of heteroepitaxy.

A classical example is the study of the initial growth of Au on Ag(111). Gold and silver form a

lattice-matched system, both being face centered cubic materials and having lattice parameters differing by only 0.2%. **Figure 7** shows room-temperature backscattering spectra from the clean Ag surface and for increasing thicknesses of the deposited Au layer. Note how the Ag surface peak diminishes as an Au surface peak appears. This indicates directly that the Au atoms give rise to a perfect shadowing of the Ag surface atoms and hence that the Au overlayers are registered with respect to the substrate, i.e., that the growth is epitaxial.

Epitaxial growth systems may be divided into different classes, according to the detailed growth mode: monolayer-by-monolayer growth (Frank-van der Merwe), pure three-dimensional islanding (Volmer-Weber) or three-dimensional islanding after completion of one or a few monolayers (Stranski-Krastanov). In the Au on Ag(111) case, the growth mode was revealed by monitoring the Au-Au shadowing effect. The ratio of the Au signal in the aligned to the nonaligned direction exhibited a pronounced decrease only after a total coverage of one monolayer had been reached, and the detailed shape of the curve was found to be consistent with the Frank-van der Merwe growth mode (layer by layer). Note that the total Au coverage is obtained by RBS.

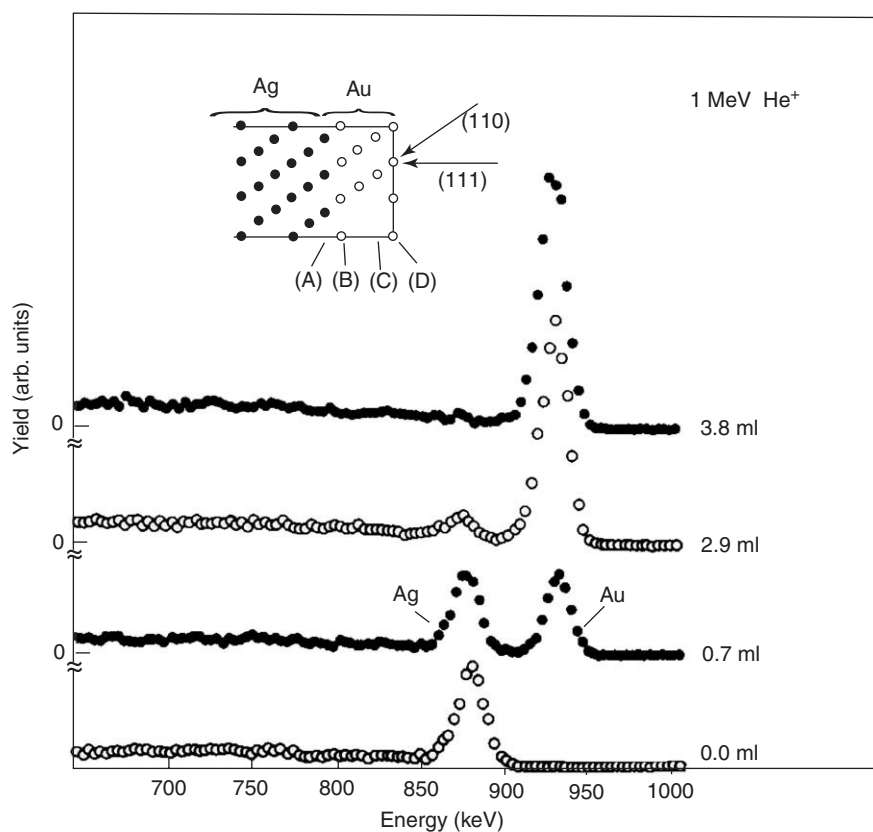
The almost perfect lattice match between Au and Ag is, of course, uncommon. For lattice mismatched systems the initial growth may again be pseudomorphic, but after completion of a critical number of layers, the strain energy built up in the thin film is relieved through the onset of misfit dislocations which destroy the shadowing.

Heteroepitaxial growth leads to buried interfaces with very important structural properties that determine, e.g., the Schottky barrier heights in the case of metal-semiconductor junctions. With the improved depth resolutions of the detection systems, structural studies by ion scattering of such interfaces are becoming increasingly feasible.

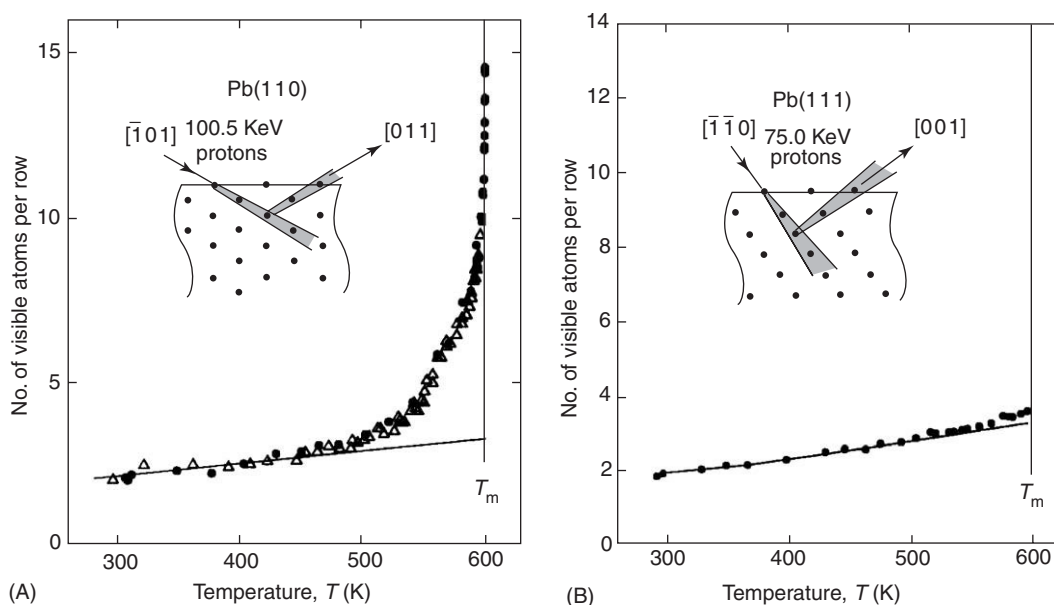
### Surface Dynamics and Melting

Based on numerous ion scattering studies of surfaces, a general picture has emerged (although exceptions have been reported): surface atoms have larger vibration amplitudes than bulk atoms, and the thermal displacement distributions for surface atoms are anisotropic, in general in such a way that rms amplitudes are larger for out-of-plane displacements than for in-plane displacements. Little is known about displacement correlations in the surface region.

As the temperature of a crystal is increased, the anharmonicity of the interatomic potentials gives rise to thermal expansion of the lattice and more



**Figure 7** Rutherford backscattering spectra from a Ag(111) crystal for ion incidence along a  $\langle 110 \rangle$  direction. The four spectra are for increasing coverages of Au (bottom to top). (Reproduced with permission from Culbertsen RJ, Feldman LC, Silverman PL, and Boehm H (1981) Epitaxy of Au on Ag (111) studied with high-energy ions scattering. *Physical Review Letters* 47: 657; © American Physical Society.)



**Figure 8** (A) Surface backscattering yields from Pb(110) as a function of temperature. The bulk melting temperature  $T_m$  is indicated. (B) As for (A) but for the Pb(111) surface. (Reprinted with permission from van der Veen JF and Frenken JWM (1991) *Surface Science* 251/252: 1; © Elsevier.)



pronounced anharmonic vibrational distributions. Some surfaces will exhibit roughening characterized by a proliferation of surface steps generated at a roughening temperature at which the free energy for creation of an atomic step becomes zero. Finally, when the thermal excursions of the atoms exceed a critical value, the crystal will begin to melt. As mentioned above, the atoms in the surface layer(s) show greater thermal excursions than the bulk atoms. This suggests that the surface layer may melt at a lower temperature than the bulk, and it may subsequently act as a nucleation center for further layer-by-layer melting, thus precluding superheating.

Direct observations of a reversible melting transition on crystalline surfaces can be observed with ion scattering. Using a shadowing/blocking geometry, the number of 'visible' surface layers of a Pb(110) crystal has been measured as a function of temperature. As shown in **Figure 8A**, the backscattering yield first increases slowly with temperature due to the expected increase in vibration amplitude. From  $\sim 470$  to  $\sim 560$  K, an additional increase is observed, which must reflect a positional disordering of the atoms in the surface region ('premelting'). As the temperature is further increased, the thickness of the quasiliquid layer diverges, following growth laws that can be predicted on the basis of thermodynamic theories of surface melting.

Not all surfaces exhibit surface melting; it appears to depend on the specific free energy  $\Delta\gamma$  that the dry solid surface has in excess of a solid surface wetted by its own liquid. For positive values of  $\Delta\gamma$ , the surface lowers its free energy by melting; for negative values, it stays dry. The condition for surface melting is more readily fulfilled for open surfaces such as fcc(110) than for close-packed ones. In accordance

with this, no surface melting was observed on Pb(111) (cf. **Figure 8B**).

So far, only a limited number of surfaces have been investigated for surface melting, but this is certainly a field where ion-scattering experiments have made a very substantial contribution.

*See also:* **Surface Analysis:** Overview.

## Further Reading

- Aono M and Souda R (1985) Quantitative surface atomic structure analysis by low-energy ion scattering spectroscopy (ISS). *Japanese Journal of Applied Physics* 24: 1249–1262.
- Czanderna AW and Hercules DM (eds.) (1991) *Ion Spectroscopies for Surface Analysis*. New York: Kluwer Academic/Plenum.
- Feldman LC and Picraux ST (1977) In: Mayer JW and Rimini E (eds.) *Ion Beam Handbook for Materials Analysis*, ch. 2 and 4. New York: Academic Press.
- Feldman LC and Mayer JW (1986) *Fundamentals of Surface and Thin Film Analysis*. New York: North Holland.
- Niehus H, Heiland W, and Taglauer E (1993) Low-energy ion scattering at surfaces. *Surface Science Reports* 17: 213–303.
- Rabelais W (2002) *Principles and Applications of Ion Scattering Spectrometry: Surface Chemical and Structural Analysis*. London: Wiley.
- Stensgaard I (1992) Surface studies with high energy ion beams. *Reports on Progress in Physics* 55: 989–1033.
- van der Veen JF (1985) Ion beam crystallography of surfaces and interfaces. *Surface Science Reports* 5: 199–288.
- van der Veen JF, Pluis B, and Denier van der Gon AW (1988) Surface melting. In: Vanselow R and Howe R (eds.) *Chemistry and Physics of Solid Surfaces VII, Springer Series on Surface Science*, ch. 16, p. 455. Heidelberg: Springer.

## Nuclear Reaction Analysis and Elastic Recoil Detection

**R W Smith**, Centre d'Etudes Nucléaires de Bordeaux Gradignan (CENBG), Gradignan, France

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Nuclear reaction analysis (NRA) and elastic recoil detection (ERD) are part of the suite of ion beam analysis (IBA) techniques. They are commonly used for the elemental depth profiling of materials in a wide range of fields, e.g., from biological and medical to the semiconductor industry.

IBA techniques, which also include Rutherford backscattering spectrometry (RBS) and particle induced X-ray emission (PIXE), require the use of a particle accelerator to produce a beam of mono-energetic MeV ions which is then incident on a target. The ions may interact with atomic electrons within the target to produce characteristic X-rays or they may collide with nuclei. If an ion collides with a nucleus it may scatter, cause the nucleus to be ejected (recoiled) or undergo a nuclear reaction resulting in the emission of particles and/or  $\gamma$ -rays. NRA involves the detection of particles or  $\gamma$ -rays caused by nuclear reactions in the target while ERD involves the

pronounced anharmonic vibrational distributions. Some surfaces will exhibit roughening characterized by a proliferation of surface steps generated at a roughening temperature at which the free energy for creation of an atomic step becomes zero. Finally, when the thermal excursions of the atoms exceed a critical value, the crystal will begin to melt. As mentioned above, the atoms in the surface layer(s) show greater thermal excursions than the bulk atoms. This suggests that the surface layer may melt at a lower temperature than the bulk, and it may subsequently act as a nucleation center for further layer-by-layer melting, thus precluding superheating.

Direct observations of a reversible melting transition on crystalline surfaces can be observed with ion scattering. Using a shadowing/blocking geometry, the number of 'visible' surface layers of a Pb(110) crystal has been measured as a function of temperature. As shown in **Figure 8A**, the backscattering yield first increases slowly with temperature due to the expected increase in vibration amplitude. From  $\sim 470$  to  $\sim 560$  K, an additional increase is observed, which must reflect a positional disordering of the atoms in the surface region ('premelting'). As the temperature is further increased, the thickness of the quasiliquid layer diverges, following growth laws that can be predicted on the basis of thermodynamic theories of surface melting.

Not all surfaces exhibit surface melting; it appears to depend on the specific free energy  $\Delta\gamma$  that the dry solid surface has in excess of a solid surface wetted by its own liquid. For positive values of  $\Delta\gamma$ , the surface lowers its free energy by melting; for negative values, it stays dry. The condition for surface melting is more readily fulfilled for open surfaces such as fcc(110) than for close-packed ones. In accordance

with this, no surface melting was observed on Pb(111) (cf. **Figure 8B**).

So far, only a limited number of surfaces have been investigated for surface melting, but this is certainly a field where ion-scattering experiments have made a very substantial contribution.

*See also:* **Surface Analysis:** Overview.

## Further Reading

- Aono M and Souda R (1985) Quantitative surface atomic structure analysis by low-energy ion scattering spectroscopy (ISS). *Japanese Journal of Applied Physics* 24: 1249–1262.
- Czanderna AW and Hercules DM (eds.) (1991) *Ion Spectroscopies for Surface Analysis*. New York: Kluwer Academic/Plenum.
- Feldman LC and Picraux ST (1977) In: Mayer JW and Rimini E (eds.) *Ion Beam Handbook for Materials Analysis*, ch. 2 and 4. New York: Academic Press.
- Feldman LC and Mayer JW (1986) *Fundamentals of Surface and Thin Film Analysis*. New York: North Holland.
- Niehus H, Heiland W, and Taglauer E (1993) Low-energy ion scattering at surfaces. *Surface Science Reports* 17: 213–303.
- Rabelais W (2002) *Principles and Applications of Ion Scattering Spectrometry: Surface Chemical and Structural Analysis*. London: Wiley.
- Stensgaard I (1992) Surface studies with high energy ion beams. *Reports on Progress in Physics* 55: 989–1033.
- van der Veen JF (1985) Ion beam crystallography of surfaces and interfaces. *Surface Science Reports* 5: 199–288.
- van der Veen JF, Pluis B, and Denier van der Gon AW (1988) Surface melting. In: Vanselow R and Howe R (eds.) *Chemistry and Physics of Solid Surfaces VII, Springer Series on Surface Science*, ch. 16, p. 455. Heidelberg: Springer.

## Nuclear Reaction Analysis and Elastic Recoil Detection

**R W Smith**, Centre d'Etudes Nucléaires de Bordeaux Gradignan (CENBG), Gradignan, France

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Nuclear reaction analysis (NRA) and elastic recoil detection (ERD) are part of the suite of ion beam analysis (IBA) techniques. They are commonly used for the elemental depth profiling of materials in a wide range of fields, e.g., from biological and medical to the semiconductor industry.

IBA techniques, which also include Rutherford backscattering spectrometry (RBS) and particle induced X-ray emission (PIXE), require the use of a particle accelerator to produce a beam of mono-energetic MeV ions which is then incident on a target. The ions may interact with atomic electrons within the target to produce characteristic X-rays or they may collide with nuclei. If an ion collides with a nucleus it may scatter, cause the nucleus to be ejected (recoiled) or undergo a nuclear reaction resulting in the emission of particles and/or  $\gamma$ -rays. NRA involves the detection of particles or  $\gamma$ -rays caused by nuclear reactions in the target while ERD involves the

detection of nuclei recoiled by the incident beam. RBS involves detecting the back scattered ions and PIXE involves detecting the characteristic X-rays.

One of the many advantages of IBA techniques is their nondestructive nature. All the techniques have their own strengths and weaknesses. To use NRA, knowledge of the sample's composition is required to determine which reactions to use. Although it generally has a better sensitivity and resolution than ERD, in most cases it can only profile one element at a time. This is also a benefit in the sense that unwanted signals from other components of the target do not interfere. ERD has the advantage of being very easy to use and unlike RBS can detect hydrogen. ERD is generally used for the easy profiling of light elements in a heavy matrix while RBS is more suited to depth profiling heavy elements in a light matrix. PIXE is a useful multielemental technique for determining a sample's composition but cannot be used to depth profile elements in the same way as NRA, ERD, and RBS.

The technique to use depends on a number of factors including the information required and the equipment available. In their simplest form, NRA (nonresonant), ERD, and RBS can all be carried out using the same equipment.

## Theory/Principles

### Nuclear Reaction Analysis

A typical nuclear reaction caused by an ion incident on a target may be written as

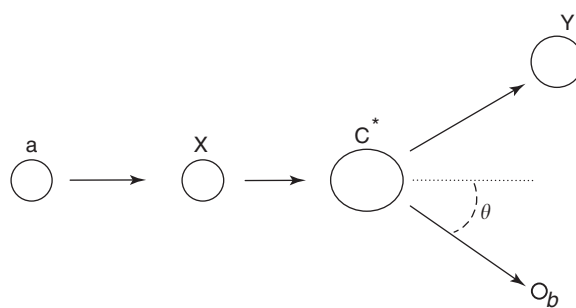
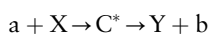


Figure 1 NRA geometry.

cross-section while resonant NRA utilizes reactions that have narrow resonances in the cross-section.

**Nonresonant NRA** In the basic reaction, conservation of total energy gives

$$m_X c^2 + E_X + m_a c^2 + E_a = m_Y c^2 + E_Y + m_b c^2 + E_b \quad [1]$$

where  $E_X$ ,  $E_a$ ,  $E_Y$ , and  $E_b$  are the kinetic energies and  $m_X$ ,  $m_a$ ,  $m_Y$ , and  $m_b$  are the rest masses of the particles. The  $Q$  value, the energy released during the reaction, is written as

$$Q = (m_X + m_a - m_Y - m_b)c^2 \quad [2]$$

If the  $Q$  value is positive the reaction is said to be exothermic and if the  $Q$  value is negative the reaction is said to be endothermic. By use of conservation of energy and linear momentum, an equation for  $E_b$  dependent on the incident ion energy  $E_a$  and detector angle  $\theta$  can be derived:

$$E_b^{1/2} = \frac{(m_a m_b E_a)^{1/2} \cos \theta + \{m_a m_b E_a \cos^2 \theta + (m_Y + m_b)[m_Y Q + (m_Y - m_a)E_a]\}^{1/2}}{m_Y + m_b} \quad [3]$$

where  $a$  is the energetic ion,  $X$  is the target, and  $C^*$  is the compound nucleus from which the reaction products  $Y$  and  $b$  (particle or  $\gamma$ -ray) are emitted. The geometry for this type of reaction is shown in Figure 1. Usually  $b$  is detected and its properties used for analysis. A more compact way of expressing the same reaction is  $X(a,b)Y$ ; the compound nucleus is not included because it is an intermediate state.

The probability of a reaction taking place is governed by its cross-section  $\sigma$  and is normally measured experimentally for a given incident particle energy and detector angle  $\theta$ . Nonresonant NRA uses reactions that have a smooth slowly changing

An equation for  $E_Y$  can be derived in the same way.

As the incident ions traverse the sample they lose energy primarily due to collisions with atomic electrons. This causes the energy of the products from nuclear reactions to change, hence producing a range of energies corresponding to different depths. By measuring the energies of these products (typically at backward angles) and taking account of the energy lost by both the incident ion on entering the target and the reaction product on leaving the target, the depth of interaction can be determined and hence an elemental depth profile produced. In some cases a particle filter is used to reduce the background in the detector and should be accounted for in the energy loss calculations.

The yield of a nuclear reaction from an element of thickness  $t$  at depth  $d$  can be calculated, providing that the cross-section changes slowly with energy and that the energy loss is small over the thickness, from the following:

$$\text{Yield}_b = \frac{Q_C \Omega (d\sigma/d\Omega)(\theta, E_a) N_{AV} t \rho}{Q_a \cos((\pi/2) - \phi) A} \quad [4]$$

where  $\text{Yield}_b$  is the number of counts (particles) resulting from a reaction with the element at depth  $d$  (cm) (normal to the surface) in the target,  $Q_C$  is the collected charge (C),  $Q_a$  is the charge on the ion a (C),  $\Omega$  is the solid angle (sr) subtended by the detector at the target,  $(d\sigma/d\Omega)(\theta, E_a)$  is the differential cross-section at angle  $\theta$  ( $\text{cm}^2 \text{sr}^{-1}$ ) for the incident ion's energy  $E_a$  at depth  $d$ ,  $N_{AV}$  is Avogadro's constant ( $6.023 \times 10^{23} \text{ mol}^{-1}$ ),  $t$  is the thickness at depth  $d$  in the target (cm),  $\rho$  is the density of the element at depth  $d$  in the target ( $\text{g cm}^{-3}$ ),  $\phi$  is the angle of incidence (rad) to the target and  $A$  is the atomic weight of the element (g).  $N_{AV} t \rho / A$  (areal density) is sometimes replaced with  $Nt$  where  $N$  is the number of atoms per cubic centimeter.

By measuring the number of counts originating from an element at depth  $d$  in the target for a set time, the concentration (areal density) of the element at that depth can be calculated using the above equation. If the energy loss is not small or if the cross-section changes quickly with energy, the single layer should be split into multiple layers for which the above equation can be used.

The depth resolution can be improved by reducing the incident beam energy (i.e., increasing the stopping power) or tilting of the sample. When using either method, changes in the cross-section should be taken into account. Both these methods have the disadvantage of decreasing the profiled depth.

**Resonant NRA** This uses a reaction that has one or more narrow resonances in the cross-section. Due to the energy loss of the ions traversing the sample, the depth at which resonance occurs can be changed by altering the incident beam energy. By scanning the energy of the beam and detecting the reaction products (normally  $\gamma$ -rays), the depth profile of a particular element can be produced. The concentration (areal density) and depth can be calculated in a similar way with modifications to that for nonresonant NRA. For the best depth resolution, the narrowest resonance should be chosen but the energy and cross-section of the resonance should also be taken into consideration.

## Elastic Recoil Detection

Ions incident on a target may collide with an atom of lower mass causing it to recoil in the forward direction. This recoiling atom, which is detected and used for analysis, has an energy  $E_2$  as follows:

$$E_2 = K \times E_0$$

where

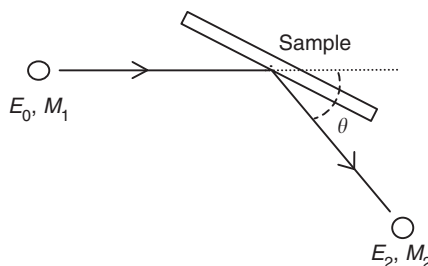
$$K = \frac{4M_1 M_2 \cos^2 \theta}{(M_1 + M_2)^2}$$

known as the kinematic factor,  $M_1$  and  $M_2$  are the mass of the incident and recoiled particles,  $\theta$  is the angle of scatter and  $E_0$  is the energy of the incident particle as shown in **Figure 2**. This technique is sometimes known as forward recoil spectrometry (FRoS).

Recoiled atoms of different mass from the surface therefore have different energies and measuring these energies at a particular scattering angle enables their identification. Also, as the incident ions traverse the sample they lose energy causing the energy of the resultant recoils to be less than those from the surface. The depth profile and concentration (areal density) can be calculated in a similar way to that for nonresonant NRA. For the concentration calculation the differential cross-section is replaced by the differential recoil cross-section – Rutherford or non-Rutherford – which can be calculated. A particle filter is normally placed in front of the detector to prevent scattered incident ions and unwanted recoils from entering it. When analyzing the data the energy lost by recoils in this filter must be taken into account.

The standard setup for many materials (e.g., polymer films) usually uses a 3 MeV  $^4\text{He}$  beam incident on the sample at a glancing angle (to improve the resolution) but there are a few variations.

**Low-energy ERD (LE-ERD)** The use of low-energy ions reduces the thickness of the particle filter required and increases the stopping efficiency of the beam. This results in an improved depth resolution but limits the depth that can be profiled.



**Figure 2** ERD geometry.

**Heavy-ion ERD (HI-ERD)** The use of heavy ions such as  $^{12}\text{C}$  increases the stopping efficiency of the beam, minimizing the thickness of particle filter required. This results in an improvement in depth resolution but limits the depth that can be profiled. The possibility of beam damage must also be considered. The heavy-ion accelerators required for this technique are not as widely available as those used for standard ERD.

**Transmission ERD** Transmission ERD uses a different geometry to standard ERD and requires the sample to be thinner than the range of the recoiled atoms. The beam is usually incident on the sample at or near normal incidence and recoils are detected using a detector placed directly behind it. It has the advantage of an improved sensitivity but the profiled depth is very limited.

**Time-of-flight ERD (TOF-ERD)** In ERD there is a built-in mass–depth ambiguity which can complicate the analysis of the results. This ambiguity is present because the energy of the recoiled atoms depends on both the mass of the target atom and the depth of interaction. This ambiguity can be avoided by using TOF-ERD. The geometry of the sample is the same as for standard ERD but before the particle is detected it passes through a TOF detector (two thin timing foils) that enables determination of its velocity. The velocity of ions with the same energy but different masses will be different. This results in measurements of both the energy and mass of the recoiled atoms. This setup does not require a particle filter because it can discriminate between recoils and

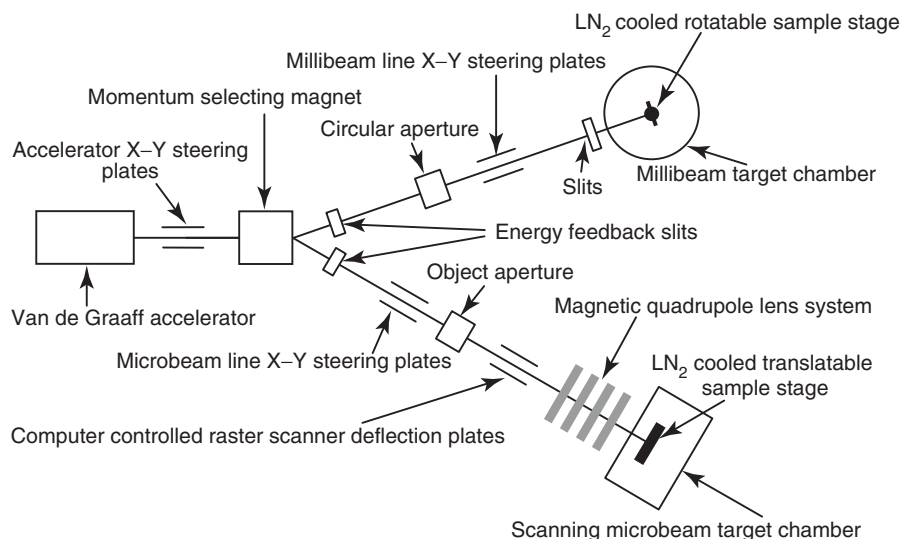
scattered ions. This improves the depth resolution but the equipment required is expensive and complicated.

## Instrumentation/Experimental Apparatus

The equipment required to perform NRA and ERD is very similar. A particle accelerator (e.g., Van de Graaff or Tandem Van de Graaff) is used to produce a beam of monoenergetic MeV ions, which is then directed down a beam line and into a target chamber, all of which are under vacuum. There are many different experimental setups in use. **Figure 3** shows the main components of both a simple millibeam line and a microbeam line. The millibeam line is typically used for nonresonant NRA and ERD and produces a beam of diameter of order millimeters. The microbeam line shown here has been used for a position scanned microbeam NRA technique and produces a beam of diameter of order micrometers. The resonant NRA technique would use a similar setup to those shown here but with the addition of equipment to scan the energy of the beam.

After the accelerator, a set of X–Y deflection plates direct the beam into a momentum selecting magnet which selects ions of a particular energy and species and directs them down the desired beam line. The energy feedback slits are linked to the accelerator and help to ensure that the beam energy is kept constant.

In the millibeam line, the ions may then pass through an optional circular aperture of a few millimeters after which they are guided into the target



**Figure 3** The basic components of a simple NRA and ERD setup.

chamber by a set of X–Y deflection plates. Before the chamber there is a pair of vertical slits which may be used to reduce the beam width if necessary. Inside the chamber the sample stage can be rotated.

In the microbeam line, X–Y deflection plates are used to guide the ions through a circular aperture of size ranging from 25  $\mu\text{m}$  to 1 mm that also acts as the object for a magnetic quadrupole lens system. They then pass through a set of computer controlled X–Y raster scanner deflection plates that enable them to be scanned over an area of the sample a few square millimeters. Finally the magnetic quadrupole lens focuses the collimated beam by a factor of five in both the  $x$  and  $y$  directions onto the surface of the sample. Inside the chamber the sample stage can be moved in the  $x$  direction to allow other samples to be positioned in the beam.

The chambers are isolated from each beam line using a gate valve. Particles are detected at backward or forward scattering angles using silicon surface barrier detectors inside the chambers.  $\gamma$ -Rays can be detected using a scintillation detector (such as sodium iodide activated with thallium – NaI(Tl) or bismuth germanate –  $\text{Bi}_4\text{Ge}_3\text{O}_{12}$  usually referred to as BGO) or an intrinsic germanium detector which is normally located outside the chamber. The detectors are connected to acquisition electronics and computer software, which for the microbeam line enable particles or  $\gamma$ -rays detected to be correlated with the position of the scanning beam. The samples can also be cooled in the chambers using an isolated liquid  $\text{N}_2$  cold finger, the use of which is discussed in the applications section. The charge incident on the sample can be measured by using a Faraday cup before the sample or by measuring the current through the sample using a current integrator.

An external beam can be produced by passing the ions (focused if necessary) through a small thin window into air. With one of these setups, instead of using deflection plates, the beam can be scanned over a large surface of the sample by moving the sample itself.

## Applications

Tables 1 and 2 show some of the recently used non-resonant and resonant nuclear reactions, respectively. In general these reactions are used to depth profile the element of interest but can also be used to determine the concentration of a particular element in a sample. The following sections describe some of the many applications of NRA and ERD.

### Hydrogen Analysis

Hydrogen is probably the most common elemental contaminant, the presence of which can seriously

**Table 1** Selected recently used nonresonant nuclear reactions

Target element	Reaction	Q value (MeV)
D	$\text{D}({}^3\text{He},\text{p}){}^4\text{He}$	18.352
${}^3\text{He}$	${}^3\text{He}(\text{D},\text{p}){}^4\text{He}$	18.352
${}^7\text{Li}$	${}^7\text{Li}(\text{p},\alpha){}^4\text{He}$	17.347
${}^{11}\text{B}$	${}^{11}\text{B}(\text{p},\alpha){}^8\text{Be}$	8.582
${}^{12}\text{C}$	${}^{12}\text{C}({}^3\text{He},\text{p}){}^{14}\text{N}$	4.779
${}^{12}\text{C}$	${}^{12}\text{C}(\text{d},\text{p}){}^{13}\text{C}$	2.719
${}^{14}\text{N}$	${}^{14}\text{N}(\text{d},\text{p}){}^{15}\text{N}$	1.305
${}^{14}\text{N}$	${}^{14}\text{N}(\text{d},\alpha){}^{12}\text{C}$	9.146
${}^{16}\text{O}$	${}^{16}\text{O}(\text{d},\text{p}){}^{17}\text{O}$	1.919
${}^{16}\text{O}$	${}^{16}\text{O}(\text{d},\text{p}_1){}^{17}\text{O}$	1.048
${}^{18}\text{O}$	${}^{18}\text{O}(\text{p},\alpha){}^{15}\text{N}$	3.980
${}^{31}\text{P}$	${}^{31}\text{P}(\alpha,\text{p}){}^{34}\text{S}$	0.632

Data from Tesmer JR and Nastasi M (eds.) (1995) *Handbook of Modern Ion Beam Materials Analysis*. Pittsburg: Materials Research Society.

**Table 2** Selected recently used resonant nuclear reactions

Target element	Reaction	Commonly used resonance energies (MeV)
H	${}^1\text{H}({}^{15}\text{N},\alpha\gamma){}^{12}\text{C}$	6.385, 13.350
H	${}^1\text{H}({}^{19}\text{F},\alpha\gamma){}^{16}\text{O}$	6.418, 16.440
H	${}^1\text{H}({}^7\text{Li},\gamma){}^8\text{Be}$	3.070
${}^{13}\text{C}$	${}^{13}\text{C}(\text{p},\gamma){}^{14}\text{N}$	1.748
${}^{19}\text{F}$	${}^{19}\text{F}(\text{p},\alpha\gamma){}^{16}\text{O}$	0.341
${}^{27}\text{Al}$	${}^{27}\text{Al}(\text{p},\gamma){}^{28}\text{Si}$	0.992
${}^{30}\text{Si}$	${}^{30}\text{Si}(\text{p},\gamma){}^{31}\text{P}$	0.620

Data from Tesmer JR and Nastasi M (eds.) (1995) *Handbook of Modern Ion Beam Materials Analysis*. Pittsburg: Materials Research Society.

affect the electrical, mechanical, and chemical properties of some materials. Other than resonant NRA and ERD, there are very few techniques available to measure hydrogen contamination. This has led to their common/widespread use making hydrogen analysis one of their more important applications. The depth profiling of hydrogen is generally carried out using the resonant nuclear reaction  ${}^1\text{H}({}^{15}\text{N},\alpha\gamma){}^{12}\text{C}$  with the 6.385 MeV resonance which for most applications offers the best depth resolution and sensitivity. Depth resolutions of  $\sim 5$  nm in thin polymer films with the beam at an angle of  $12^\circ$  to the sample have been achieved but the resolution quickly degrades with depth. Profiling depths of up to several micrometers are possible. The  ${}^1\text{H}({}^{19}\text{F},\alpha\gamma){}^{16}\text{O}$  and  ${}^1\text{H}({}^7\text{Li},\gamma){}^8\text{Be}$  reactions have also been used with the advantages that a  ${}^{19}\text{F}$  beam is easier to produce and the  ${}^7\text{Li}$  reaction allows profiling to greater depths.

ERD can be easily used to profile hydrogen but generally has a poorer depth resolution ( $\sim 80$  nm with a profiling depth of  $\sim 0.8$   $\mu\text{m}$  for thin polymer films) than that obtained using resonant NRA.



However, it does have the advantage of being a very simple technique that does not require the complicated beam energy scanning equipment that is required for resonant NRA.

These techniques have been used to profile the hydrogen content of many materials, e.g., electronic materials, metals, superconductors, and polymers. They have also found use in water diffusion studies by profiling of the hydrogen component as discussed later.

### Depth Profiling Deuterium-Labeled Molecules

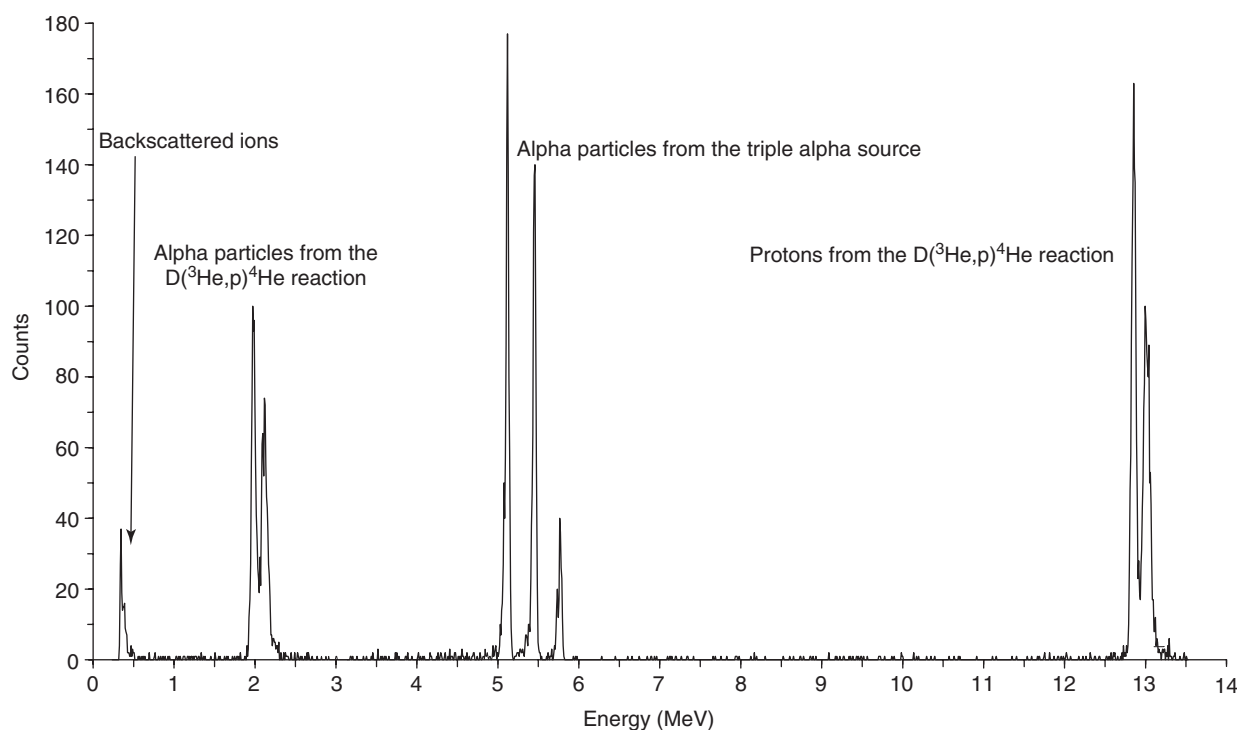
The  $D(^3\text{He},p)^4\text{He}$  nonresonant reaction has been used to study the diffusion of deuterium and deuterium-labeled molecules into a wide variety of materials.  $^3\text{He}$  ion beam energies of between 0.7 MeV, giving a resolution of 30 nm and a profiling depth of 0.1  $\mu\text{m}$  with the beam at a  $15^\circ$  angle to the sample, and 2 MeV giving a resolution of 400 nm and a profiling depth of 8  $\mu\text{m}$  with the beam normally incident, with the reaction products detected at an angle of  $165^\circ$  are usually used. These produce high-energy protons of  $\sim 12\text{--}14$  MeV which are well separated in the particle spectrum from the alpha particles of  $\sim 1\text{--}3$  MeV. Due to the kinematics of this reaction as the  $^3\text{He}$  ions lose energy by traversing the sample the

resultant protons and alpha particles emitted at backward angles increase in energy.

For several years this reaction was used to study interpolymer diffusion in thin films by deuterating the polymer of interest. **Figure 4** shows the particle spectrum resulting from a 0.7 MeV  $^3\text{He}$  ion beam incident on a test sample consisting of 32.3 nm DPS/63.3 nm PS/32.3 nm DPS (PS, polystyrene; DPS, deuterated polystyrene) at an angle of  $10^\circ$ . The high-energy proton and low-energy alpha particle peaks are from the DPS layers and the triple alpha peaks are from a radioactive source in the chamber used for calibration of the system. The second proton and alpha particle peaks are lower due to a decrease in the cross-section, i.e., the  $^3\text{He}$  beam energy has fallen below 0.7 MeV where the cross-section peaks.

ERD has also been used to depth profile deuterium in polymer thin films. It is able to profile both hydrogen and deuterium simultaneously but has a smaller profiled depth and poorer resolution when compared to the  $D(^3\text{He},p)^4\text{He}$  NRA technique.

To further increase the profiling depth of this technique the use of this reaction was combined with a position scanned  $^3\text{He}$  ion microbeam. The resulting technique has been used to study molecular ingress/diffusion into various materials by labeling the molecules of interest with deuterium. Water molecules are labeled by adding heavy water ( $\text{D}_2\text{O}$ ) and other



**Figure 4** Particle spectrum acquired from a DPS/PS/DPS sample with a 0.7 MeV  $^3\text{He}$  beam at an angle of  $10^\circ$ . A threshold level was applied to the acquisition electronics.

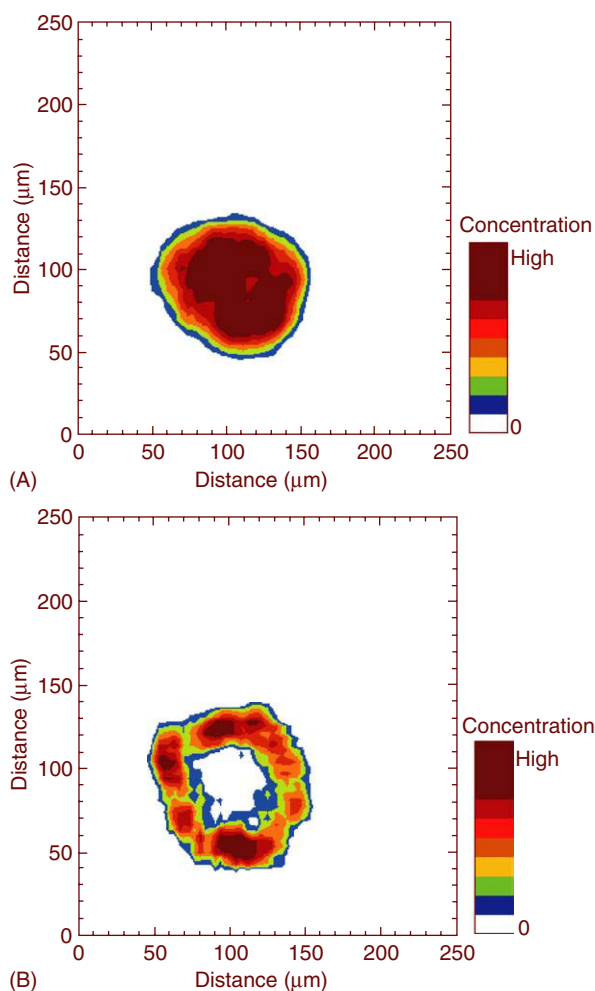
hydrogen containing molecules are labeled by deuteration. After the sample has been exposed to the deuterium-labeled molecules, it is cut to reveal a cross-section and then plunged into liquid nitrogen. This prevents the evaporation of the diffused molecules and fixes the diffusion profile. A  $^3\text{He}$  micro-beam is then raster scanned over the surface of the cross-section and the detected protons from the  $\text{D}(^3\text{He},\text{p})^4\text{He}$  reaction correlated with the position of the beam to produce a two-dimensional area map showing the location of the diffused molecules. During analysis, the sample is normally back-cooled with liquid nitrogen to prevent the molecules from being lost to the vacuum and also to reduce heating induced by the beam. If combined with PIXE, elements within the sample (e.g., sulfur present in hair or silicon present in glass) or the molecules themselves can also be located. The resolution depends on the size of the beam ( $\mu\text{m}$ ) and the distance between each dwell point, but the profiled depth is in principle only dependent on the ability to expose the sample to the beam. This technique has recently been used to study the diffusion of a deuterated hair dye into hair. For this hair dye to be effective and remain fast it was necessary for it to diffuse below the surface of the hair. **Figure 5A** shows a two-dimensional plot of detected sulfur X-rays showing the distribution of sulfur and hence the location of the hair. **Figure 5B** shows a two-dimensional plot of detected protons showing the distribution of deuterium and hence the location of the deuterated hair dye. It is clear from observation of these images that the hair dye has diffused well into the hair proving its effectiveness. Other applications have included the study of surfactants into hair, water diffusion into fiber optic pressure sensors, and water and drug diffusion in drug-release polymers.

### The Investigation of Nuclear Materials Using NRA

Minor actinides present in some nuclear waste produce large quantities of helium when they decay. When considering materials for storage of this waste, it is necessary to have some knowledge of helium diffusion in the material matrix. Helium diffusion into possible materials has been simulated by exposing them to a  $^3\text{He}$  ion beam (i.e., implantation). The implanted helium has then been profiled using the  $^3\text{He}(\text{D},\text{p})^4\text{He}$  nonresonant nuclear reaction enabling diffusion coefficients to be determined.

### Techniques to Study Water Diffusion

There are a few techniques that can be used to profile water ( $\text{H}_2\text{O}$ ) diffusion. The oxygen component can be profiled by labeling the water with  $^{18}\text{O}$  and using



**Figure 5** Two-dimensional plots showing the distribution of (A) sulfur and (B) deuterium in the hair. (Reprinted with permission from Smith RW and Clough AS (2002) Depth profiling of diffusion into cylindrical matrices using a scanning micro-beam. *Nuclear Instruments and Methods in Physics Research B* 188: 126–129; © Elsevier.)

the nonresonant  $^{18}\text{O}(\text{p},\alpha)^{15}\text{N}$  nuclear reaction to profile it. The hydrogen can be profiled using the  $^1\text{H}(^{15}\text{N},\gamma)^{12}\text{C}$  resonant nuclear reaction or ERD. If deuterium-labeled water ( $\text{D}_2\text{O}$ ) is used then its diffusion can be profiled using the  $\text{D}(^3\text{He},\text{p})^4\text{He}$  nuclear reaction techniques as described previously or ERD. The technique to use depends on the material being investigated, the depth to be profiled, and the ability to distinguish the water from the matrix. The hydrogen and oxygen nuclear reactions have both been used together to profile water diffusion into quartz. Using a combination of techniques in this way makes it possible to investigate the diffusion of the hydrogen and oxygen components separately and to study their effect on the diffusion mechanism.

### Analysis of Implanted Ions Using NRA

Ion implantation of materials, to change its properties, has many important applications in materials science and semiconductor technology. Several nonresonant and resonant nuclear reactions have been used for the profiling of such implantations. The 0.992 MeV resonance of the  $^{27}\text{Al}(p,\gamma)^{28}\text{Si}$  resonant nuclear reaction has, for example, been used to profile the thermal diffusion of aluminum in aluminum-implanted stainless steel when heated to temperatures of between 450°C and 650°C.

### NRA and ERD with External Beams (Including Simultaneous Analysis of Carbon, Nitrogen, and Oxygen)

The nondestructive nature of MeV IBA methods make them especially suitable in investigating fragile and precious objects. The use of PIXE with an external beam is a well-established technique that enables elemental analysis of large and/or delicate samples that could not be put in a vacuum chamber, e.g., art and archaeological objects. NRA (in conjunction with PIXE and RBS) has been used with an external deuteron beam for the simultaneous depth profiling of carbon, nitrogen, and oxygen in historical objects by using, among others, the  $^{12}\text{C}(d,p_0)^{13}\text{C}$ ,  $^{14}\text{N}(d,p_5)^{15}\text{N}$ , and  $^{16}\text{O}(d,p_1)^{17}\text{O}$  nuclear reactions. Preliminary experiments using ERD with an external  $^4\text{He}$  beam have been carried out to depth profile hydrogen and deuterium in precious stones. These have been relatively successful but there are some difficulties with the setup.

### Nonstandard ERD Techniques

LE-ERD and HI-ERD have both been used to profile deuterium in PS thin films with a better depth resolution than standard ERD but suffer a large decrease in profiling depth. LE-ERD has used a 1.3 MeV  $^4\text{He}$  beam with tilting of the sample to give a depth resolution of 14 nm and a profiling depth of 150 nm. HI-ERD has used a 2.4 MeV  $^{12}\text{C}$  beam with tilting of the sample to give a resolution of 8 nm and a profiling depth of 100 nm.  $^{20}\text{Ne}$  and  $^{40}\text{Ar}$  beams have also been used giving similar resolutions and profiling depths. HI-ERD can detect elements heavier than hydrogen and deuterium but radiation damage to the sample is a possibility.

### Developments in Data Analysis Programs

There have been several recent developments in the computer programs used to analyze NRA and ERD data.

The IBA DataFurnace utilizes the simulated annealing algorithm to analyze RBS, ERD, and

nonresonant NRA data in a fully automated way with minimum intervention. The user has to only input the data, the experimental conditions, and the elements present in the sample. No iterative procedure is involved although data can be simulated if required. Data from the same sample using different experimental conditions or techniques can be analyzed simultaneously and all the information taken into account to produce the final result.

SIMNRA is a program intended for the simulation of non-Rutherford backscattering, nonresonant NRA, and ERD data. The user is required to construct a layered target from which a simulation algorithm produces a simulated spectrum. The user then compares the simulation to the data and makes changes to the target description as necessary. In order to produce a target description the user must have some knowledge of the target composition. It does, however, have the option of fitting spectra using a least squares algorithm and parameters supplied by the user.

WinRNRA is a fitting program that converts resonant NRA yield curves into concentration depth profiles. It can be used to analyze data from a single element or isotope in an otherwise homogeneous target.

## Concluding Comments

NRA and ERD can be used to study elemental/molecular diffusion into a wide range of materials and can be used in conjunction with other complementary techniques such as PIXE and RBS. The sensitivity of the different techniques described here depends on a large number of factors including the experimental setup, detection efficiency, etc. but levels of about parts per thousand to parts per million have been reported. Excluding the position scanned microbeam NRA technique, depth resolutions ranging from several nanometers to several 100 nm for depths of up to several micrometers have been reported, with the resolution often degrading with depth.

There are a few potential future developments. The possibility of combining the position scanned  $^3\text{He}$  microbeam NRA technique and the  $\text{D}(^3\text{He},p)^4\text{He}$  nonresonant nuclear reaction technique to produce three-dimensional profiles has been discussed in the literature.

The size of focused ion beams has decreased in recent years so that it is now possible to produce beams of order 100 nm in size, i.e., nanobeams. The use of smaller beams with the position scanned microbeam NRA technique will improve the spatial resolution but will decrease the yield. For delicate samples where the current cannot be increased, this

may make it necessary to acquire data for longer and/or improve the particle detection efficiency.

When using an external beam there is also the possibility of producing large elemental maps of several centimeters over the sample surface by moving the sample itself over the beam.

Several new data analysis programs have been developed in recent years making the interpretation and analysis of data simpler and quicker. With modern computers it is also possible to analyze the data on-line, i.e., almost as soon as it has been acquired.

**See also:** **Archaeometry and Antique Analysis:** Dating of Artifacts; Metallic and Ceramic Objects; Art and Conservation. **Forensic Sciences:** Hair. **Polymers:** Natural Rubber; Synthetic. **Surface Analysis:** Particle-Induced X-Ray Emission; Ion Scattering.

## Further Reading

Calligaro T, Dran J-C, Moignard B, *et al.* (2002) Ion beam analysis with external beams: Recent set-up improvements. *Nuclear Instruments and Methods in Physics Research B* 188: 135–140.

Clough AS and Jenneson P (1998) Ion beam analysis of diffusion in heterogeneous materials. *Nuclear Instruments and Methods in Physics Research B* 139: 51–57.

Composto RJ, Walters RM, and Genzer J (2002) Applications of ion scattering techniques to characterize polymer surfaces and interfaces. *Materials Science and Engineering R* 38: 107–180.

Geoghegan M (1999) MeV ion beam profiling of polymer surfaces and interfaces. In: Richards RW and Peace SK (eds.) *Polymer Surfaces and Interfaces*, vol. III, pp. 43–73. New York: Wiley.

Hollands R, Rutt H, Clough AS, Peel R, and Smith R (2001) Preliminary results for the diffusion of water into fibre optic oil well sensors. *Nuclear Instruments and Methods in Physics Research B* 174: 519–525.

Jeynes C, Barradas NP, Marriot PK, *et al.* (2003) Elemental thin film depth profiles by ion beam analysis using simulated annealing – a new tool. *Journal of Physics D: Applied Physics* 36: R97–R126.

Landry F and Schaff P (2001) Simulation and deconvolution program WinRNRA for depth profiling of light elements via nuclear resonance reactions. *Nuclear Instruments and Methods in Physics Research B* 179: 262–266.

Mayer M (1997) SIMNRA User's Guide, Report IPP 9/113. Max-Planck-Institut für Plasmaphysik, Garching, Germany (<http://www.rzg.mpg.de/~mam/>).

# Secondary Ion Mass Spectrometry of Polymers

D Briggs, Malvern, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Secondary ion mass spectrometry (SIMS), when carried out in the 'static' (i.e., zero damage) mode, is a uniquely powerful tool for the molecular characterization of polymer surfaces. Its molecular specificity also allows the detection and characterization of additives and contaminants, whose presence at the surface can cause major changes in surface behavior, at the submonolayer level. Modern instruments based on time-of-flight (ToF) mass analyzers and liquid-metal ion sources are capable of chemical imaging with a spatial resolution of well below 1  $\mu\text{m}$ .

## Physical Basis

In SIMS, the sample, in ultrahigh vacuum (UHV) ( $<10^{-7}$  Pa), is bombarded with energetic ions (typically 8–25 keV depending on the source, see below). These penetrate below the surface to a distance of

several nanometers whilst transferring their energy to the material. In the immediate vicinity of the primary ion track, where the energy density is highest, many bonds are broken and atoms are set in motion. Secondary and higher generation collisions set up a collision cascade radiating from the impact point. Isomerization of the original translational energy also leads to electronic and vibrational excitation. As a result of the collision cascade, fragments of the material close to the surface can be desorbed (sputtered) into the vacuum. These fragments range in size from single atoms to large assemblies. Most of the sputtered material is neutral but  $\sim 1\%$  is charged (positively or negatively). These secondary ions, the vast majority being singly charged, can be extracted into a mass spectrometer for mass analysis.

As with gas-phase organic mass spectrometry, SIMS spectra are information-rich and provide fingerprints. However, the processes leading to the spectra (both positive and negative ion spectra are utilized) are not well understood. The charged fragments can be formed directly in the collision cascade, or by ion/electron attachment, electron impact, and ion–molecule reactions in the selvedge (the affected region near the surface which may have

may make it necessary to acquire data for longer and/or improve the particle detection efficiency.

When using an external beam there is also the possibility of producing large elemental maps of several centimeters over the sample surface by moving the sample itself over the beam.

Several new data analysis programs have been developed in recent years making the interpretation and analysis of data simpler and quicker. With modern computers it is also possible to analyze the data on-line, i.e., almost as soon as it has been acquired.

**See also:** **Archaeometry and Antique Analysis:** Dating of Artifacts; Metallic and Ceramic Objects; Art and Conservation. **Forensic Sciences:** Hair. **Polymers:** Natural Rubber; Synthetic. **Surface Analysis:** Particle-Induced X-Ray Emission; Ion Scattering.

## Further Reading

Calligaro T, Dran J-C, Moignard B, *et al.* (2002) Ion beam analysis with external beams: Recent set-up improvements. *Nuclear Instruments and Methods in Physics Research B* 188: 135–140.

Clough AS and Jenneson P (1998) Ion beam analysis of diffusion in heterogeneous materials. *Nuclear Instruments and Methods in Physics Research B* 139: 51–57.

Composto RJ, Walters RM, and Genzer J (2002) Applications of ion scattering techniques to characterize polymer surfaces and interfaces. *Materials Science and Engineering R* 38: 107–180.

Geoghegan M (1999) MeV ion beam profiling of polymer surfaces and interfaces. In: Richards RW and Peace SK (eds.) *Polymer Surfaces and Interfaces*, vol. III, pp. 43–73. New York: Wiley.

Hollands R, Rutt H, Clough AS, Peel R, and Smith R (2001) Preliminary results for the diffusion of water into fibre optic oil well sensors. *Nuclear Instruments and Methods in Physics Research B* 174: 519–525.

Jeynes C, Barradas NP, Marriot PK, *et al.* (2003) Elemental thin film depth profiles by ion beam analysis using simulated annealing – a new tool. *Journal of Physics D: Applied Physics* 36: R97–R126.

Landry F and Schaff P (2001) Simulation and deconvolution program WinRNRA for depth profiling of light elements via nuclear resonance reactions. *Nuclear Instruments and Methods in Physics Research B* 179: 262–266.

Mayer M (1997) SIMNRA User's Guide, Report IPP 9/113. Max-Planck-Institut für Plasmaphysik, Garching, Germany (<http://www.rzg.mpg.de/~mam/>).

# Secondary Ion Mass Spectrometry of Polymers

D Briggs, Malvern, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

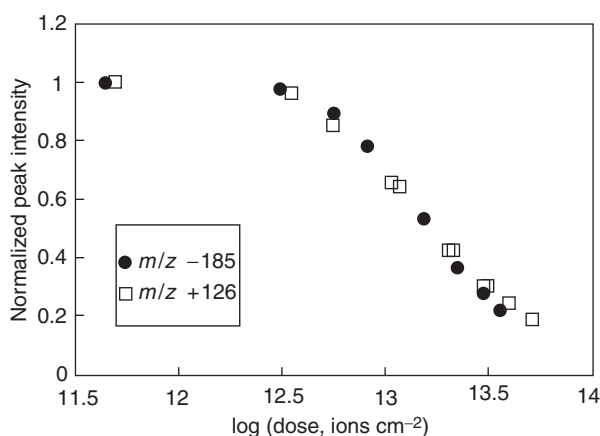
Secondary ion mass spectrometry (SIMS), when carried out in the 'static' (i.e., zero damage) mode, is a uniquely powerful tool for the molecular characterization of polymer surfaces. Its molecular specificity also allows the detection and characterization of additives and contaminants, whose presence at the surface can cause major changes in surface behavior, at the submonolayer level. Modern instruments based on time-of-flight (ToF) mass analyzers and liquid-metal ion sources are capable of chemical imaging with a spatial resolution of well below 1  $\mu\text{m}$ .

## Physical Basis

In SIMS, the sample, in ultrahigh vacuum (UHV) ( $<10^{-7}$  Pa), is bombarded with energetic ions (typically 8–25 keV depending on the source, see below). These penetrate below the surface to a distance of

several nanometers whilst transferring their energy to the material. In the immediate vicinity of the primary ion track, where the energy density is highest, many bonds are broken and atoms are set in motion. Secondary and higher generation collisions set up a collision cascade radiating from the impact point. Isomerization of the original translational energy also leads to electronic and vibrational excitation. As a result of the collision cascade, fragments of the material close to the surface can be desorbed (sputtered) into the vacuum. These fragments range in size from single atoms to large assemblies. Most of the sputtered material is neutral but  $\sim 1\%$  is charged (positively or negatively). These secondary ions, the vast majority being singly charged, can be extracted into a mass spectrometer for mass analysis.

As with gas-phase organic mass spectrometry, SIMS spectra are information-rich and provide fingerprints. However, the processes leading to the spectra (both positive and negative ion spectra are utilized) are not well understood. The charged fragments can be formed directly in the collision cascade, or by ion/electron attachment, electron impact, and ion–molecule reactions in the selvedge (the affected region near the surface which may have



**Figure 1** Relative intensities of diagnostic fragments from poly(methyl methacrylate) (+  $m/z$  126 and -  $m/z$  185) as a function of primary ion dose (using 8 keV  $\text{Cs}^+$ ). Individual spectra were acquired with a dose of  $<5 \times 10^{11}$  ions  $\text{cm}^{-2}$ . (Reproduced with permission from Briggs D and Fletcher IW (1997)  $\text{Cs}^+$  ion beam damage of poly(vinyl chloride) and poly(methyl methacrylate) studied by high mass resolution ToF-SIMS. *Surface and Interface Analysis* 25: 167–176; © John Wiley & Sons Ltd.)

some characteristics of a plasma), or by unimolecular decomposition of fragments with high internal energies in the vacuum.

For the study of organic materials, in particular, it is essential to limit the primary ion dose, so that the spectrum is representative of undamaged material (the 'static' SIMS regime). The aim is to ensure that no part of the surface receives more than one primary ion impact (affecting an area of order 10 nm diameter) during the analysis. By measuring spectra as a function of primary ion dose it is possible to determine the 'static SIMS limit', i.e., the dose above which the spectra are dose dependent. **Figure 1** illustrates such data for poly(methyl methacrylate), a representative polymer (the origins of the chosen fragments are described in section 'polymer spectra'). Under the conditions used the static SIMS limit is a primary ion dose of  $\sim 3 \times 10^{12}$  ions  $\text{cm}^{-2}$ .

There are few hard data that relate to the SIMS information depth for polymers, but the consensus view is that this is of the order of 1 nm. Atomic fragments may emerge from somewhat deeper below the surface but large molecular species arise only from the uppermost molecular layer or two.

## Instrumental Aspects

The early developments of static SIMS (largely driven by polymer surface analysis requirements in terms of practical application) were carried out with noble gas

ion sources and quadrupole mass analyzers (QMS). The latter had the benefits of compactness, ready availability, and relatively straightforward adaptability to the use of SIMS in UHV systems. Although all the essential features of polymer surface analysis were introduced using this technology, the QMS has major limitations for static SIMS. First, it is a serial (scanning) device, so that only one mass at a time can be detected. Second, it has a limited mass range ( $<1000$  Da) and the transmission (i.e., sensitivity) decreases with mass (by at least  $\text{m}^{-1}$ ). Third, it is only possible to achieve a uniform mass resolution ( $m/\Delta m$ ) over this range of about unity. These limitations led to the introduction of energy-compensating ToF mass analyzers.

In ToF-SIMS the primary ion beam is pulsed, producing primary ion 'packets' of  $\sim 1$  ns duration. The secondary ions are accelerated over a short distance to typically 3 keV energy and injected into the ToF analyzer. In a linear ToF this is simply a field-free tube in which the ions of different mass travel with different velocities and separate, to arrive at a detector sequentially. The flight time is related to the square root of the mass, so the 'time spectrum' can be converted to a mass spectrum by suitable calibration. The mass analysis, which takes  $<200 \mu\text{s}$ , is carried out between primary ion pulses. The spectrum is therefore acquired in parallel and the ion dose is minimized. The secondary ions leave the surface with an energy distribution that, whilst only a few percent of the energy after acceleration, is enough to degrade the mass resolution. Different ToF designs compensate for this initial energy distribution in different ways, but modern instruments achieve  $m/\Delta m$  of 10k or more. The mass range is theoretically infinite, but in practice the difficulty of registering the arrival of very large slow ions gives an upper limit of  $\sim 10$  kDa. For most polymer surface studies an upper mass of 2 kDa is more than sufficient and over this range the transmission is independent of mass. Overall, the ToF approach has a  $10^4$  increase in sensitivity compared to a QMS.

The pulsed nature of ToF-SIMS provides a further benefit for the analysis of insulating materials. These charge-up under positive ion bombardment and stabilizing the surface potential is essential, otherwise secondary ion detection is rapidly curtailed. In ToF-SIMS this is efficiently achieved by irradiating the sample with a broad flood of low-energy electrons (typically 20 eV) between each primary ion pulse. This is especially important during high spatial resolution imaging (which is virtually impossible with a QMS because the ion beam current density is much higher and continuous). Because of all these advantages, ToF-SIMS is the preferred method for polymer



surface analysis and during the last decade the QMS approach has been eclipsed.

A variety of ion sources are used in ToF-SIMS. Electron impact sources are used to produce  $\text{Ar}^+$ ,  $\text{Xe}^+$ , and  $\text{SF}_5^+$ . Surface ionization sources produce  $\text{Cs}^+$ . Both of these types of source are used mainly for spectroscopy where spatial resolution is not important (although beam diameters of a few tens of micrometers are used). Liquid metal ion sources produce  $\text{Ga}^+$ ,  $\text{In}^+$ , and, most recently,  $\text{Au}^+/\text{Au}_2^+/\text{Au}_3^+$ . These sources allow the production of highly focused beams with diameters down to 30 nm. There is an advantage to using high mass primary ions since the efficiency of production of large organic fragments (i.e., the secondary ion yield divided by the primary ion dose) increases significantly with primary ion mass at any given ion energy.

An important aspect of imaging by ToF-SIMS is the ability to store (in memory) the whole spectrum for each pixel in the scanned array. This large dataset, called a raw data file, can then be interrogated retrospectively. Two smaller data files are first created: one in which all the spectral intensity is summed for each pixel to generate a total intensity image and another in which the spectra from each pixel are summed to generate a total or average spectrum. The intensity variation in the total image is due to both topography effects (as in SEM) and major compositional differences. Regions of interest can be selected by bounding a number of pixels and the spectrum from them recreated – retrospective microanalysis. From the total spectrum, peaks of interest can be selected (singly or collectively) and the intensity in all pixels displayed – retrospective chemical imaging or mapping. Thus, two scans over the field of view to collect both positive and negative ion raw data files capture all the available information.

## Spectral Features and Information

### Detection of Elements

Modern, high mass resolution, ToF-SIMS instruments make this task relatively easy. With the standard isotopic mass defined for  $^{12}\text{C} = 12.0000$ , all the elemental isotopes from  $^{16}\text{O}$  (15.9949) upward have masses slightly below the nominal unit mass. Organic fragments, being largely composed of carbon and hydrogen ( $^1\text{H} = 1.0078$ ), have exact masses that are slightly above the nominal unit mass. Thus, for example, the single isotope of  $\text{Mn}^+$  ( $^{55}\text{Mn} = 54.9380$ ) would have a serious interference with a common hydrocarbon fragment, of the same nominal mass,  $\text{C}_4\text{H}_7^+$  (exact mass = 55.0548). However, with a mass resolution of only 1500 (defined as

$m/\Delta m$ , where  $\Delta m$  is the full-width at half-maximum peak intensity) these two peaks are completely resolved. Modern instruments can usually deliver a mass resolution well above this figure, even for insulating samples that pose some problems because of surface texture or shape, and  $m/\Delta m$  of  $> 3000$  would be expected. Hence, the identification of elemental species is usually straightforward.

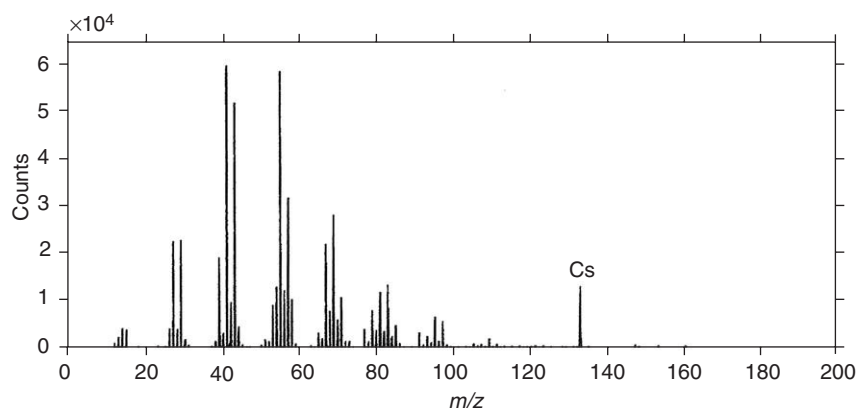
Hydrogen is detected with high yield with both polarities but most other elements of interest are detected in the negative ion spectrum. Carbon is detected as  $\text{C}^-$ , but the accompanying  $\text{CH}^-$  is always more intense. Similarly, oxygen is detected as  $\text{O}^-$  and  $\text{OH}^-$ . The relative intensity ratio  $\text{OH}^-/\text{O}^-$  is much more variable (structure dependent) than  $\text{CH}^-/\text{C}^-$ . The halogens give very high negative ion yields. When present in high concentration and with the appropriate structural proximity they may be observed as addition fragments  $\text{X}_2^-$  and  $\text{X}_3^-$ . Nitrogen is not usually detected as an elemental peak at  $m/z$  14 but rather as a combination ion,  $\text{CN}^-$  and additionally  $\text{CNO}^-$  if the structure is appropriate. Oxidized nitrogen functions produce  $\text{NO}_2^-$  and  $\text{NO}_3^-$ . Sulfur is detected as  $\text{S}^-$  (interfering  $\text{O}_2^-$  can usually be neglected) and  $\text{SH}^-$ . Oxidized sulfur appears as  $\text{SO}_x^-$  ( $x = 1-4$ ) and  $\text{SO}_4\text{H}^-$ . Phosphorus is usually present in the oxidized state and detected as  $\text{PO}_2^-$  and  $\text{PO}_3^-$ .

### Polymer Spectra

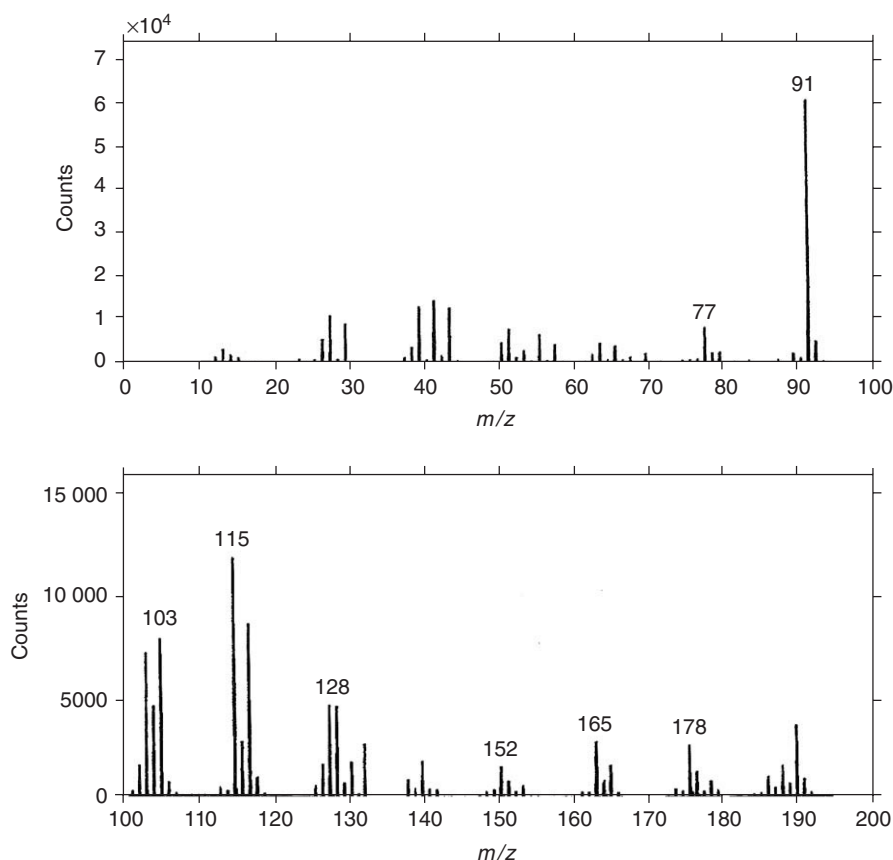
The interpretation of polymer spectra has been the subject of a significant body of literature. The *Static SIMS Library* contains ToF-SIMS spectra of some 200 reference polymers together with the structures of characteristic fragments. The following is, therefore, intended only to illustrate some general principles.

**Hydrocarbon polymers** Figure 2 shows the positive ion spectrum from high-density polyethylene,  $(\text{CH}_2\text{CH}_2)_n$ . Each cluster of peaks is due to  $\text{C}_n\text{H}_m^+$  where  $m = n, \dots, 2n + 1$ . For more complex aliphatic hydrocarbon polymers with pendant groups the smooth decrease in intensity of the clusters from  $\text{C}_2$  is interrupted, and the relative intensities of the ions within each cluster changes, as certain alkyl ions with methyl substituents are resonance-stabilized. Unsaturated aliphatic hydrocarbon polymer spectra are additionally characterized by a general weighting of intensity within each  $\text{C}_n\text{H}_m^+$  cluster toward lower values of  $m$  compared with saturated analogs.

The positive ion spectra of aromatic hydrocarbon polymers are exemplified by that of polystyrene in Figure 3. Compared to the aliphatic hydrocarbon spectra the  $\text{C}_n\text{H}_m^+$  clusters extend to much higher



**Figure 2** Positive ion ToF-SIMS spectrum from high-density polyethylene. (Reproduced with permission from Vanden Eynde X (2001) Quantitative analysis of polymer surfaces. In: Vickerman JC and Briggs D (eds.) *TOF-SIMS: Surface Analysis by Mass Spectrometry*, Ch.16. Manchester: SurfaceSpectra/IM Publications; © SurfaceSpectra/IM Publications.)



**Figure 3** Positive ion ToF-SIMS spectrum from polystyrene in the mass range  $m/z=0-200$ . (Reproduced with permission from Briggs D (1998) *Surface Analysis of Polymers by XPS and Static SIMS*. Cambridge: Cambridge University Press; © Cambridge University Press.)

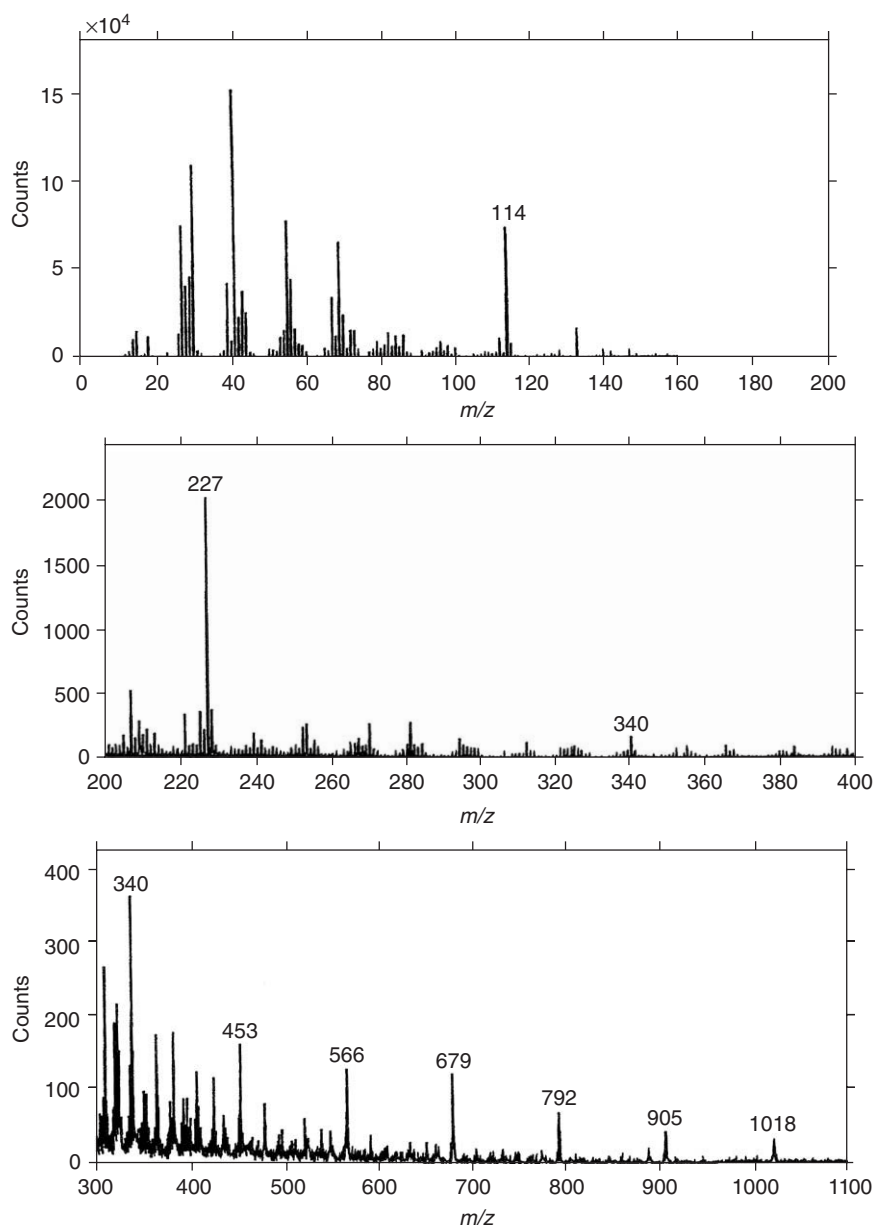
masses and have much lower  $m/n$  ratios as expected from the higher degree of unsaturation. Most of the prominent peaks are due to cyclic and polycyclic hydrocarbon ions, with the cyclic tropylium ion,  $C_7H_7^+$

( $m/z$  91), the most diagnostic. The negative ion spectra of all hydrocarbon polymers are very simple.  $C_1$  and  $C_2$  clusters dominate and above  $C_4$  (aliphatic) or  $C_6$  (aromatic) clusters the spectra are very weak. The

$C^-/CH_2^-$  ( $m/z$  12/14) intensity ratio is significantly higher for aromatic hydrocarbons compared with aliphatic hydrocarbons (saturated or unsaturated).

**In-chain functionality** In the spectra of hydrocarbon polymers there is no obvious manifestation of the repeat unit chemistry. Introduction of in-chain functionality changes the situation. **Figure 4** shows the positive ion spectrum of nylon-6 (polycaprolactam),  $[(CH_2)_5CONH]_n$ , where the amide function interrupts a hydrocarbon backbone. In addition to mainly hydrocarbon fragments at low

mass, a prominent series of peaks appears due to multiple repeat unit fragments  $[nM + H]^+$  where  $M$  is the repeat unit mass (113 Da). This behavior is much more typical of polymer spectra with multiple repeat unit fragments visible in either positive or negative ion spectra, or both, depending on the specific chemistry and opportunity for large stable fragment ion formation. (The relative intensity of these high mass ions is somewhat unusual in this case but they, in general, can be dramatically enhanced by the use of a polyatomic primary ion, such as  $SF_5^+$ .)



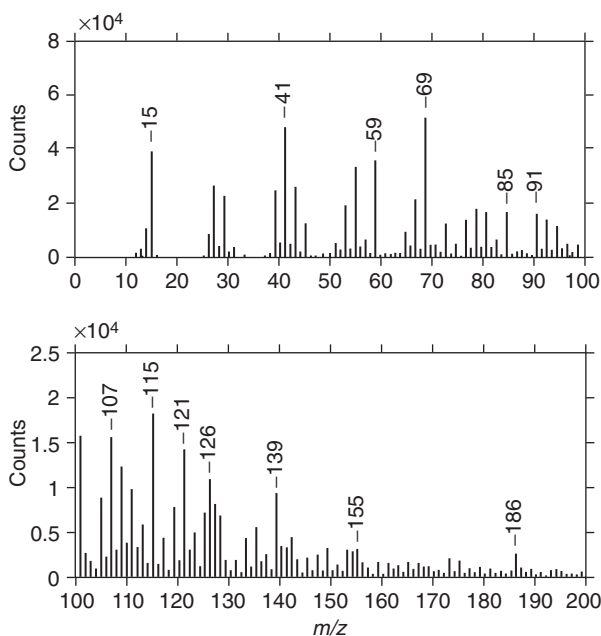
**Figure 4** Positive ion ToF-SIMS spectrum from nylon 6 (polycaprolactam). Highlighted peaks represent multiple repeat unit fragments  $[nM + H]^+$  from  $n = 1$  ( $m/z$  114) to  $n = 9$  ( $m/z$  1018). (Reproduced with permission from Briggs D (1998) *Surface Analysis of Polymers by XPS and Static SIMS*. Cambridge: Cambridge University Press; © Cambridge University Press.)

Main-chain oxygen functionalized polymers also behave in this way. Thus, linear polyethers give series such as  $[nM + H]^+$  and  $[nM + OH]^-$ . In-chain polyesters either of the poly(lactone)  $(R-COO)_n$ , or the poly(diacid-diol)  $(R_1COOR_2COO)_n$  types give series such as  $[nM + H]^+$  and  $[nM - OH]^-$ , with probable  $R-C\equiv O^+$  structures, and  $[nM + H]^-$  and  $[nM + OH]^-$ , with probable  $R-COO^-$  structures. Of course, there are many other ions formed by fragmentation at other points in the chain.

**Side-chain functionality** Several classes of polymers consist of a hydrocarbon backbone with an oxygen-containing side chain. Collectively, they can be grouped into vinyl and acrylic polymers. Vinyl polymers have the general structure  $(CH_2CHX)_n$ . In poly(vinyl ethers)  $X = -OR$ , in poly(vinyl ketones)  $X = -C(=O)R$ , and in poly(vinyl carboxylates)  $X = -OC(=O)R$ . These polymers give positive ion spectra with hydrocarbon contributions from the backbone and from the side chain R group if this has a significant number of carbon atoms (especially if linear). Repeat unit ions are observed for the ethers but the dominant ions for the other polymers tend to be side chain fragments, especially  $R-C\equiv O^+$ . In the negative ion spectra, repeat unit ions are seen for the ketones,  $RO^-$  dominates in the case of the ethers and  $RCOO^-$  dominates in the case of carboxylates.

Acrylic polymers are based on esters of acrylic acid and methacrylic acid and co-polymers are very common. Figure 5 shows the positive ion spectrum of poly(methyl methacrylate) (PMMA). Compared with the other oxygen-containing side-chain polymers just discussed the spectrum appears to be influenced to a greater extent by hydrocarbon fragments from the backbone. However, at high mass resolution, most peaks are seen to be multicomponent, often with a hydrocarbon component, but frequently due only to oxygen-containing fragments – some examples are illustrated in Figure 6. Interesting repeat unit fragments are even-mass radical cations (e.g.,  $m/z$  126, 154, 196, 200); these appear as essentially single-component peaks as illustrated by  $m/z$  126 in Figure 6. With higher methacrylates (R larger than  $CH_3$ ) the positive ion spectrum becomes more hydrocarbon-like. The common diagnostic ion is the methacryloyl ion at  $m/z$  69.

The negative ion spectrum of PMMA is shown in Figure 7. With the absence of 'hydrocarbon noise' this is a much simpler spectrum and the pattern of peaks clearly reflects the repeat unit mass of 100 amu. Negative ion peaks characteristic of all methacrylates, because the side-chain alkyl group (R) has been lost, are at  $m/z$  85, 109, 125, and 139.

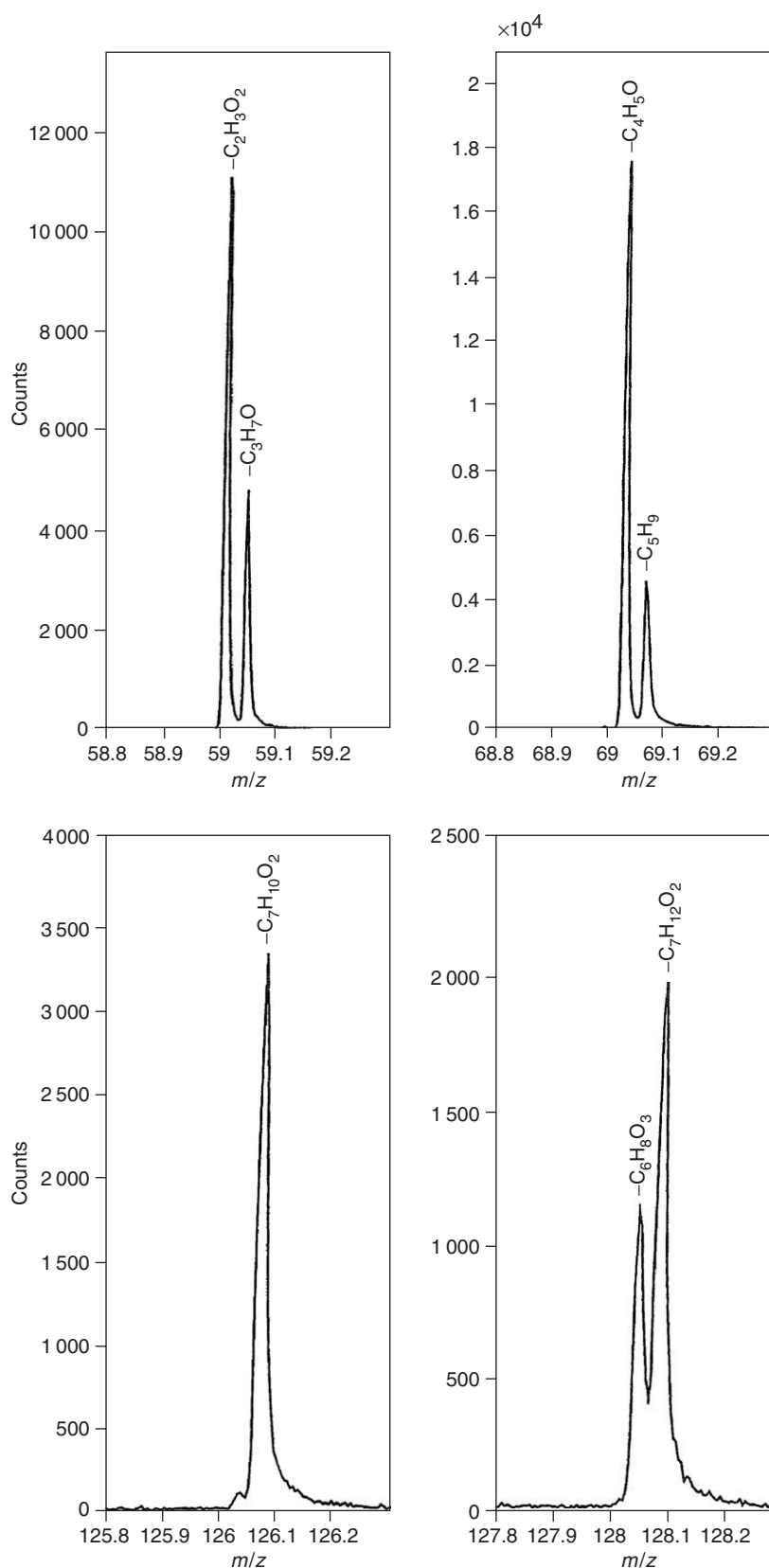


**Figure 5** Positive ion ToF-SIMS spectrum from poly(methyl methacrylate) in the mass range  $m/z = 0-200$ . (Reproduced with permission from Briggs D and Fletcher IW (1997)  $Cs^+$  ion beam damage of poly(vinyl chloride) and poly(methyl methacrylate) studied by high mass resolution ToF-SIMS. *Surface and Interface Analysis* 25: 167–176; © John Wiley & Sons Ltd.)

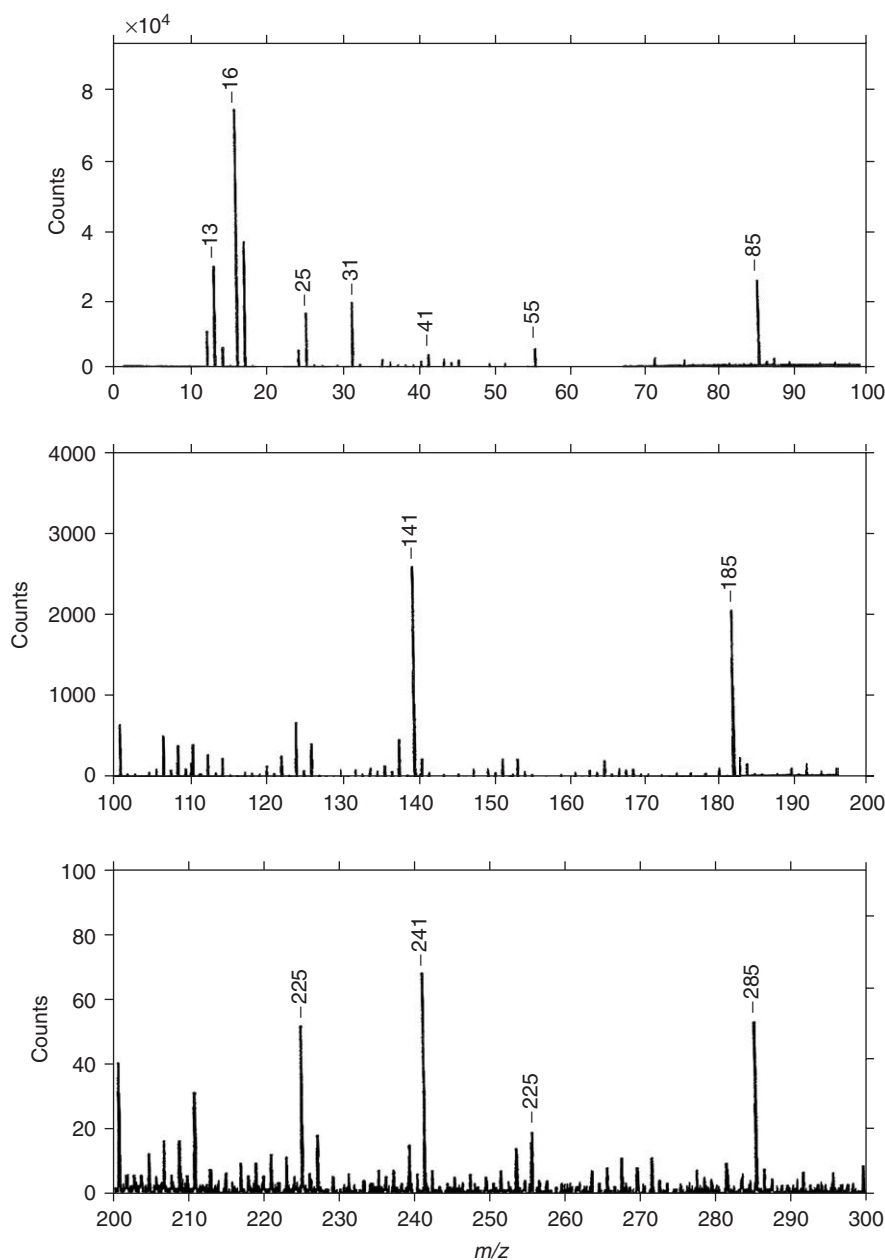
Specific ions having one or more intact side chains have the generic formulae:  $[M + 55]^-$ ,  $[2M - 15]^-$ ,  $[2M + 41]^-$ ,  $[2M + 55]^-$ , and  $[3M - 15]^-$  where  $M$  is the monomer mass. In the PMMA spectrum these are at  $m/z$  155, 185, 241, 255, and 285.

Acrylate homopolymers behave similarly, except that the repeat unit ions are much less stable. The characteristic common positive fragment is the acryloyl ion at  $m/z$  55. Common negative ions appear at  $m/z$  65, 67, 71, and 81, whilst specific ions are  $[M - H]^-$ ,  $[M + 13]^-$ ,  $[M + 27]^-$ ,  $[M + 39]^-$ , and  $[M + 53]^-$ .

The spectra of random acrylic co-polymers display the monomer-specific features of the components together with fragment ions that reflect the linking of the different monomers in the chain. These are the equivalent of the multiple repeat unit ions from a homopolymer, but now the monomers need not be the same. In all cases the negative ion spectra are the most informative. Thus, for example, the  $[2M - 15]$  ion from a methacrylate homopolymer could appear as  $[2M_1 - 15]^-$ ,  $[2M_2 - 15]^-$ , or  $[M_1M_2 - 15]^-$ , where  $M_1$  and  $M_2$  are the two monomers in a random co-polymer. The spectra of co-polymers in general are discussed in more detail below.



**Figure 6** High mass resolution details of some of the peaks in **Figure 5**. (Reproduced with permission from Briggs D and Fletcher IW (1997)  $Cs^+$  ion beam damage of poly(vinyl chloride) and poly(methyl methacrylate) studied by high mass resolution ToF-SIMS. *Surface and Interface Analysis* 25: 167–176; © John Wiley & Sons Ltd.)



**Figure 7** Negative ion ToF-SIMS spectrum from poly(methyl methacrylate) in the mass range  $m/z=0-300$ . (Reproduced with permission from Briggs D and Fletcher IW (1997)  $\text{Cs}^+$  ion beam damage of poly(vinyl chloride) and poly(methyl methacrylate) studied by high mass resolution ToF-SIMS. *Surface and Interface Analysis* 25: 167–176; © John Wiley & Sons Ltd.)

**Nitrogen-containing polymers** The polyamide nylon-6 has been discussed above. Other polylactams behave similarly and display a characteristic pair of peaks at  $[2M + \text{H}]^+$  and  $[2M - \text{OH}]^+$ . Diacid-diamine polyamides,  $(\text{R}_1\text{CONHR}_2)_n$ , show some similarity to in-chain polyesters having dominant fragments of the type  $\text{R}-\text{C}\equiv\text{O}^+$ . However, since there is no equivalent of the  $\text{RCOO}^-$  ion from nylons, the negative ion spectra are usually lacking in

high mass molecular fragments. The characteristic peaks are simply  $\text{CN}^-$  and  $\text{CNO}^-$ . Nitrogen content is generally obvious from the positive ion spectra through the higher than normal incidence of even mass peaks (due to fragments with one, or an odd number of, N atoms).

Another important class is the polyurethane, where a polyether,  $(\text{X})_n$ , is reacted with a di-isocyanate (in this case MDI) and chain extended with either



a diamine or a diol (Y). The positive ion spectrum is principally a superposition of the polyether spectrum (discussed above) and fragments of the di-isocyanate. In the case of 4,4'-methylenebis(phenylene isocyanate) (MDI), characteristic peaks are at  $m/z$  106, 132, 180 and 233. The polyether can be replaced by a polyester in which case its spectrum will be a significant component of the polyurethane spectrum.

**Halogen-containing polymers** These polymers are characterized by intense negative ions from the halogen. Aliphatic chlorine-containing polymers such as poly(vinyl chloride)  $(\text{CH}_2\text{CHCl})_n$  give positive ion spectra that at first sight resemble hydrocarbon spectra. However, at high mass resolution many peaks are seen to have additional components containing one or two Cl atoms.

Highly fluorinated polymers give an intense  $\text{CF}_3^+$  peak at  $m/z$  69, even if this group is not present in the molecule. Poly(tetrafluoroethylene)  $(\text{CF}_2\text{CF}_2)_n$  gives generic fragments  $\text{C}_x\text{F}_y^{+/-}$  and poly(vinylidene fluoride)  $(\text{CH}_2\text{CF}_2)_n$  gives generic fragments  $\text{C}_x\text{F}_y\text{H}_z^{+/-}$ . An important class of lubricants is the perfluoropolyethers. These are homo- or co-polymers involving the units  $(\text{CF}_2\text{O})$ ,  $(\text{CF}_2\text{CF}_2\text{O})$ , and  $(\text{CF}_2\text{CF}(\text{CF}_3)\text{O})$ . Complex repeat unit sequences are seen in both polarity spectra with characteristic low mass ions being  $\text{CFO}^+$  ( $m/z$  47) and  $\text{CF}_3\text{O}^-$  ( $m/z$  85). Low molecular weight materials with the same repeat units are often differentiated by their end groups, which give diagnostic peaks.

**Aromatic polymers** Apart from hydrocarbon polymers containing styrene, discussed above, this is a large class of polymers. Attachment of heteroatoms to the phenyl ring, either in-chain or in a pendant group, leads to characteristic ions. Thus, ring-substituted styrene polymers give substituted tropylium ions,  $\text{C}_7\text{H}_6\text{X}^+$  (e.g.,  $\text{X} = \text{OH}$ ,  $\text{F}$ ,  $\text{Cl}$ ,  $\text{Br}$ ), and aromatic polyethers give  $\text{C}_6\text{H}_5\text{O}^-$ .

**Silicones** Silicone rubbers and oils are usually based on polydimethylsiloxane  $(\text{Si}(\text{CH}_3)_2\text{O})_n$ . They all give spectra readily identified through the positive ions at  $m/z$  28, 47, 147, 107, 221, and 281. The relative intensities of these peaks vary as a function of molecular weight, end groups, and cross-link density in the rubbers.

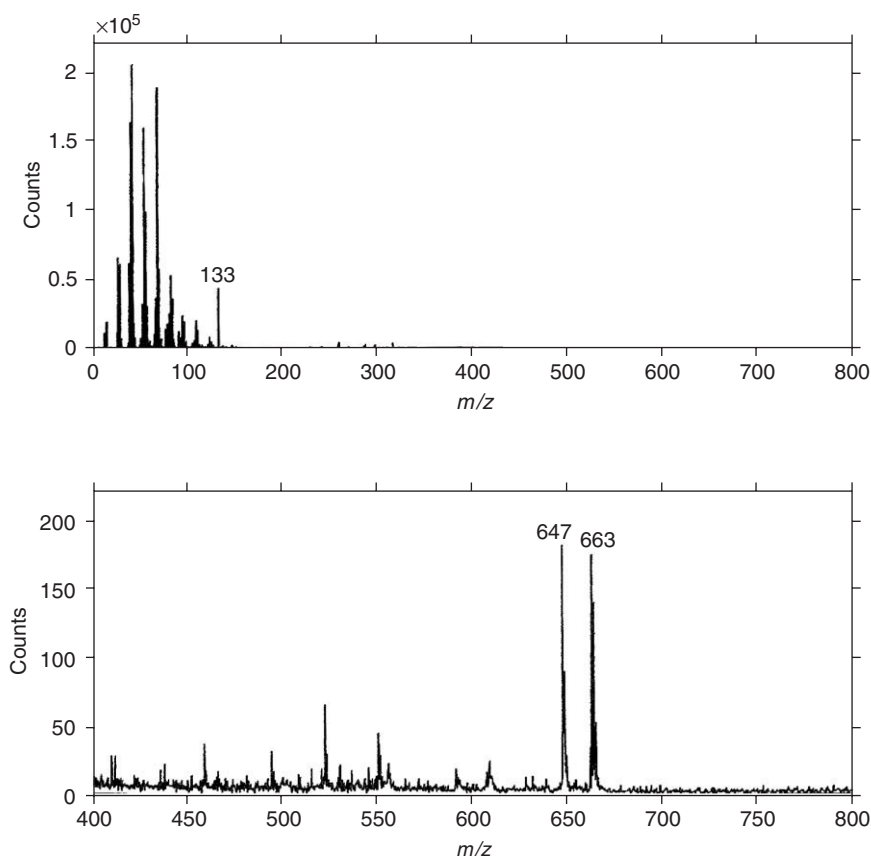
### Organic Molecules on Polymer Surfaces

Probably the most important aspect of ToF-SIMS analysis involves the use of the technique to understand the variation in surface properties, such as

wettability, adhesion, and friction, which are sensitive to submonolayer coverage of molecular species. ToF-SIMS is unique in its sensitivity and molecular speciation capability. However, unlike the situation with gas-phase organic molecule identification by electron impact mass spectrometry there are, as yet, no well-established 'rules' that allow molecular identification directly from the spectra. The situation is also complicated, in most cases, by the superposition of the spectrum of the substrate and the potential presence of more than one organic molecule.

Organic molecules usually give quasimolecular ions of the type  $[\text{M} + \text{H}]^+$  and  $[\text{M} - \text{H}]^-$ , where  $\text{M}$  is the molecular weight, in addition to characteristic fragments. In the case of polymer substrates, the intensity from the substrate spectrum is, typically, only high in the region  $m/z$  0–200. Quasimolecular ions are usually of higher mass, and even though their absolute intensity may be rather low, they are easily observed relative to the contribution from the substrate. An example is given in Figure 8, which shows the positive ion spectrum from the surface of pretreated (electrochemically oxidized) polypropylene. Below  $m/z$  120 the spectrum is essentially due to hydrocarbon polymer fragments (compare with Figure 1). Higher mass peaks are due to contamination and segregated additive molecules. The phosphite antioxidant Irgafos 168 gives rise to a distinctive set of peaks around  $m/z$  650. Although  $[\text{M} + \text{H}]^+$  for the phosphite is expected at  $m/z$  647 it is believed that the spectrum is mainly due to the oxidized phosphate product (the molecule is present to decompose hydroperoxides), which would give  $[\text{M} + \text{O} + \text{H}]^+$  at  $m/z$  663.

Pattern recognition is the key to organic molecule identification. Common 'surface contamination' molecules fall into a relatively small number of chemical families, depending on their function, e.g., surfactants, lubricants, and the various types of polymer additives (plasticizers, antioxidants, etc.). Thus, many surfactants are based on quaternary ammonium salts, many plasticizers are dialkyl phthalates, most antioxidants are di-*t*-butyl phenol derivatives, etc. The result is that each family member is likely to give a spectrum with some common fragment ions whilst having unique molecular ions. Alternatively, the whole spectrum may conform to a particular pattern but with components 'shifted' by fixed mass intervals, e.g., fatty acids that have a common hydrocarbon series in the positive ion spectrum plus molecular ions/fragments that increase in mass by 14 amu ( $\text{CH}_2$ ) as the homologous series is ascended. The easiest way to recognize these patterns is to study the collections of relevant spectra, collated by function, in *The Static SIMS Library*.



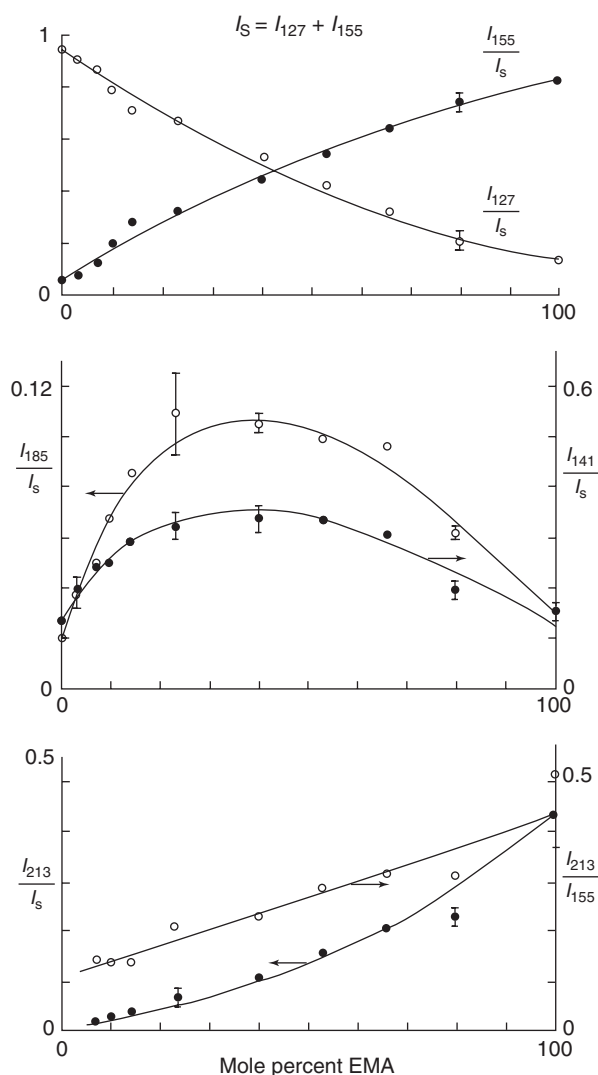
**Figure 8** Positive ion ToF-SIMS spectrum from the surface of pretreated (electrochemically oxidized) polypropylene, illustrating the typical signal-to-noise ratio of polymer fragments at low mass and additive molecular ions at high mass (see text). The peak at  $m/z$  133 is due to primary  $\text{Cs}^+$  ions. (Reproduced with permission from Vanden Eynde X (2001) Quantitative analysis of polymer surfaces. In: Vickerman JC and Briggs D (eds.), *ToF-SIMS: Surface Analysis by Mass Spectrometry*, Ch.16. Manchester: SurfaceSpectra/IM Publications; © SurfaceSpectra/IM Publications.)

## Quantitative Aspects

SIMS is an inherently nonquantitative technique – there is no direct relationship between peak intensity and the concentration of the species it represents. This is because secondary ion yields are strongly dependent on the chemical environment (or matrix), examples of which were noted in the previous section. Matrix effects in SIMS are most notable for yields of elemental species from inorganic matrices, but the situation is less severe for polymer matrices. Thus, relative intensities such as  $\text{O}^-/\text{CH}^-$  or  $\text{F}^-/\text{CH}^-$  correlate with X-ray photoelectron spectroscopy (XPS) atomic ratios (O/C, F/C) when trends in composition are restricted to similar materials; however, the  $\text{Cl}^-:\text{CH}^-$  ratio for a variety of chlorine-containing polymers does not correlate with the XPS Cl/C atomic ratio.

There have been many quantitative studies of co-polymers, in which the intensities of fragments

representing monomers or larger units have been correlated with composition. An early example, shown in **Figure 9**, concerns random hydroxyethylmethacrylate/ethylmethacrylate (HEMA/EMA) co-polymers and the use of characteristic negative ions (as discussed above). Fragments at  $m/z$  127 and 155 are essentially representative of HEMA and EMA monomers, respectively, and relative intensity plots of the form  $A/A + B$  give smooth trends (upper curves). The intensities of fragments at  $m/z$  141 and 185, believed to represent HEMA–EMA linked monomers (middle curves), maximize in the midrange as expected. The intensity of the fragment at  $m/z$  213, representing EMA–EMA links increases rapidly with EMA content (lower curve), also as expected. The ratio  $I(213)/I(155)$  represents the ratio of diads to singles. For a random co-polymer where diad (AA) signals come from sequences  $\text{B}-(\text{A})_n-\text{B}$  where  $n > 2$ , a statistical analysis predicts the ratio of diads to singles (isolated A units) will be proportional to the



**Figure 9** Relative intensities of characteristic negative ion fragments from HEMA-EMA copolymers (see text) plotted as function of bulk composition. Error bars in (A) represent the largest spread of values for any measurement; those in (B) and (C) represent the spread of values for all repeated measurements. (Reproduced with permission from Briggs D and Ratner BD (1988) A semi-quantitative SIMS analysis of the surfaces of random ethyl methacrylate: hydroxethyl methacrylate copolymer films. *Polymer Communications* 29: 6–8; © Butterworths.)

mole fraction of A, as seen in the lower plot. The choice of fragments for such plots is not always so obvious and in many cases better results are obtained by using multivariate statistical analysis techniques that utilize the whole spectrum.

The end groups of a polymer can have a marked effect on the SIMS spectra. If these are chemically distinct from the rest of the chain, because of the use of specific initiation or termination reactions, they will give rise to readily identifiable peaks. However, in all cases end groups can have a disproportionate effect on spectra because only one bond needs to be broken to produce a fragment whereas two bonds need to be broken to generate an in-chain fragment. Ratios of fragments due to end groups and in-chain fragments in homopolymer spectra depend on molecular weight, typically up to 20–30k. This provides a unique method for monitoring surface average molecular weight, which can be particularly important in this range. Low molecular weight fractions of the distribution may migrate to the surface and significantly alter mechanical properties relative to the bulk polymer as well as leading to adverse surface behavior.

See also: **Surface Analysis:** Overview; Nuclear Reaction Analysis and Elastic Recoil Detection.

## Further Reading

- Briggs D (2001) Surface analysis of polymers. In: Vickerman JC and Briggs D (eds.) *ToF-SIMS: Surface Analysis by Mass Spectrometry*. Manchester: SurfaceSpectra/IM Publications.
- Briggs D, Fletcher IW, and Goncalves NM (2000) The positive secondary ion mass spectrum of poly(methyl methacrylate): A high mass resolution ToF-SIMS study. *Surface and Interface Analysis* 29: 303–309.
- Briggs D, Fletcher IW, and Goncalves NM (2002) A high mass resolution study of chlorine-containing polymers. *Surface and Interface Analysis* 33: 178–184.
- Reichlmaier S, Bryan SR, and Briggs D (1995) Surface trimer crystallization on poly(ethylene terephthalate) studied by time-of-flight secondary ion mass spectrometry. *Journal of Vacuum Science and Technology A* 13: 1217–1223.
- Reichlmaier S, Hammond JS, Hearn MJ, and Briggs D (1994) Analysis of polymer surfaces by SIMS 17. An assessment of the accuracy of the mass assignment using a high mass resolution ToF-SIMS instrument. *Surface and Interface Analysis* 21: 739–746.
- The Static SIMS Library*, version 3 (2002). Manchester: SurfaceSpectra (CD-ROM, see [www.surfacespectra.com](http://www.surfacespectra.com)).

## Secondary Neutral Mass Spectrometry

**H Oechsner**, Universität Kaiserslautern, Kaiserslautern, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The removal of atoms and molecules from a solid surface with 'atomic spoons' and their subsequent mass spectrometric identification constitutes the most direct method of determining the surface composition of a solid. In practice, the ejection of surface particles is achieved by bombardment with primary atomic species, mostly ions, a process generally referred to as 'sputtering'. It is well established both by real and computer experiments that for moderate bombarding energies up to the lower keV regime particle ejection occurs mainly from the topmost atomic layer or, with rapidly decreasing probabilities, from a very shallow depth of a few atomic distances beyond the surface, thus providing information that is extremely surface sensitive.

It has been well established that the vast majority of the ejected surface particles consists of neutrals. Secondary ions contribute to the ejected particle flux at most to some  $10^{-2}$  for ionic targets as metal oxides, but remain by orders of magnitude lower in most other cases. Nevertheless, secondary ions have long been utilized for surface analysis (secondary ion mass spectrometry, SIMS), since they can immediately be identified in mass spectrometers employing electric or magnetic fields for particle separation. However, the ionization probability during the ejection process has also long been recognized to depend strongly on the chemical environment at the ejection site. Hence, SIMS signals are generally subjected to strong matrix influences that are difficult to take into account quantitatively, and, therefore, usually prevent quantitative information about the composition of an unknown sample surface from being derived. Deleterious matrix effects can be avoided to some extent by transforming the surface into a new distinct chemical state, e.g., by oxidizing it through oxygen flooding or a sufficiently dense bombardment with oxygen ions as primary particles. More recently, the signals of  $\text{MCs}^+$  ions produced under bombardment with  $\text{Cs}^+$  primary ions have also proved to be less matrix sensitive.

Such disadvantages clearly do not exist when the dominating neutral particle flux from an ion bombardment surface is evaluated. Even when the usually

very low secondary ion yields vary by orders of magnitude with changing surface concentrations, the flux of sputter ejected neutrals remains stable and reflects the true surface composition. In order to allow their mass spectrometric identification the ejected neutral atoms and molecules must, however, be 'postionized' in an appropriate way. Such experimental complications were the main reason why secondary (or 'sputtered') neutral mass spectrometry (SNMS) has been introduced as a surface analytical technique much later than SIMS. The main difference between these two techniques is clearly that particle ejection and ionization occur in the same event in SIMS, but are strictly separated in SNMS. This means that in SNMS, for appropriate performance, the particle-specific postionization probability becomes a constant of the analytical apparatus. This makes quantification of SNMS, i.e., the correlation between the mass spectrometric signals and the surface concentration of the respective particle, straightforward and simple.

Among the possibilities available to postionize sputter-ejected neutral surface particles, electron impact ionization has been employed in a variety of experimental approaches. More recently, photoionization by resonant or nonresonant multiphoton absorption processes has been established as another very effective technique in SNMS. Other processes as Penning ionization or charge exchange play only a minor role in postionization for SNMS.

Because of its attractive features SNMS is expected to replace SIMS when quantitative analysis of inhomogeneous samples and of layered structures is required. SNMS microprobes both based on electron impact postionization and photoionization enable microanalysis with the full quantifiability of this technique. While SIMS is quantitatively applicable only when the analyte is diluted in a homogeneous matrix at concentrations sufficiently below 1 at.%, i.e., when each secondary ion is emitted from an identical chemical environment, SNMS is an appropriate analytical technique across the entire concentration range varying from values close to unity down to well below the ppm level. For a comparison with electron spectroscopic techniques as Auger or photoelectron spectroscopy (AES or XPS), one has to consider that the original surface composition is in general changed under ion bombardment owing to preferential particle removal or atomic mixing effects. AES and XPS are monitoring such altered surface compositions. In contrast, SNMS collects

with a specific but constant detection efficiency the sputter-removed particles, thus providing quantitative information about the original sample composition.

The following sections will deal first with the physical effects involved in postionization, then with the respective instrumentation, and finally with the quantification schemes in surface and depth profile analysis with SNMS. A critical comparison of the different SNMS techniques will be addressed where appropriate.

## Postionization of Sputtered Neutral Particles

Historically, the first attempts to postionize sputter-released particles were performed with electron beam arrangements. Similar to residual gas analysis (RGA), an electron beam ionizer was placed between the sputtered surface and the entrance of a mass spectrometer. Compared with RGA very moderate postionizing efficiencies in the order of  $10^{-5}$ – $10^{-6}$  ionized particles per neutral sputtered particle entering the ionizer have been achieved, resulting in very low mass spectrometric signals and, hence, a low detection power. Such unattractive features, which are essentially due to the very short dwell time of sputtered particles in the ionizing electron beam, ensured an almost negligible interest in SNMS. It was found only after those first SNMS attempts that the energy distribution of sputtered particles is determined by the relatively violent atomic collisions in the sputtering cascades. The influence of the surface binding energy  $U_0$ , which in general is in the order of several eV, causes the energy distributions of sputtered atoms to peak at several eV (more precisely at  $U_0/2$  for a fully developed isotropic sputter cascade) and the average energy of sputtered atoms to be in the region of 10 eV, thus making the particles to be postionized very fast.

A considerably more effective way of electron impact postionization for mass spectrometric identification of sputtered neutral species was introduced by employing a spatially expanded hot Maxwellian electron gas (see below). Much higher postionization efficiencies of the order of  $10^{-2}$  were established by this method for which the acronym SNMS was first introduced.

Both for electron beam and electron gas postionizations the efficiency  $\alpha_x^0$  for postionizing a sputtered species  $X$  is described by

$$\alpha_x^0 \sim n_e l \iint Q_x(v_e) f(v_e) v_e N_x(v_x) v_x^{-1} dv_e dv_x \quad [1]$$

where  $f(v_e)$  is the electron velocity distribution,  $Q_x(v_e)$  the ionization function of  $X$ , and  $N_x(v_x)$  the

velocity distribution of the sputtered species to be ionized.  $\alpha_x^0$  is proportional to the electron density  $n_e$  and the individual dwelling times  $t_x = 1/v_x$ . For electron beam postionization the distance  $l$  along which postionization occurs is determined by the beam diameter and  $f(v_e)$  is a narrow distribution around the nominal beam velocity that can be matched to the maximum of  $Q_x$ , i.e., the maximum total cross-section for electron impact ionization of an individual species  $X$ . In the electron gas mode  $l$  is the traveling length of sputtered particles through the postionization electron gas (typically in the order of 5 cm), and  $f(v_e)$  is in general a Maxwellian distribution with a maximum corresponding to an electron energy of  $\sim 10$  eV. The ionization functions  $Q_x(v_e)$ , starting at the respective ionization threshold energy (a few eV), display a broad maximum at electron energies of 50–100 eV. Hence, both the much larger traveling distances  $l$  and the broader overlap between  $f(v_e)$  and  $Q_x(v_e)$  are the main reasons for the much higher ionization efficiencies  $\alpha_x^0$  for electron gas compared with electron beam postionization.

With the availability of high power lasers providing up to a few 100 mJ for nanosecond and to a few 10 mJ for picosecond pulses, photoionization based on resonant or nonresonant multiphoton absorption schemes can be employed as another effective postionization technique. Ionization efficiencies close to unity are achieved when a resonant step is included in the multiphoton absorption process (resonant-enhanced multiphoton ionization, REMPI). REMPI provides unrivalled detection sensitivity when the photon energy is tuned to a resonant step for the specific atomic energy structure of the species to be detected. REMPI has also been successfully employed to sputtered organic molecules. Nonresonant multiphoton ionization (NRMPI) has first been developed for the analysis of inorganic samples (surface analysis by laser ionization). Saturation of all photoion signals for the different species in the sputtered particle flux can be achieved in nonresonant schemes for laser intensities in the  $10^{12} \text{ W cm}^{-2}$  regime. Quasiresonant intermediate atomic states may become effective in such cases also. Below saturation, the dependence of the NRMPI signals of the laser intensity (or the photon density in the ionizing volume) was not always found to follow the expected power law, for instance, a quadratic dependence for two photon processes. The respective deviations may in part be referred to contributions from photon-induced dissociative ionization of the molecule fraction in the particle flux from the sample. Because of the reduced photon density in the edge regions of the laser beam such influences may

prevail, even when atomic photoionization in the center of the beam profile becomes saturated.

Postionization via a Penning process, i.e., by the interaction of a sputtered particle with a metastable atom as, e.g.,  $\text{Ar}(^3\text{P}_{2,0})$  with a stored energy of 11.5 or 11.7 eV, necessarily requires a high collision rate between heavy particles in an appropriate gas discharge. The corresponding analytical technique known as glow discharge mass spectrometry (GDMS) employs direct current (DC) or radio frequency (RF) discharges maintained in a noble gas at pressures of the order of  $10^{-2}$ – $10^{-1}$  mbar. The collisional interactions being necessary for the Penning process destroy the kinetic properties of the sputtered released surface particles, and lead to the formation of new molecular species (e.g., 'argides') by associative ionization. The latter effect may seriously hamper the quantification of the GDMS spectra from an unknown sample. Backscattering of previously sputter-removed particles to the sample via elastic collisions reduces depth resolution during thin film analysis. Since GDMS is in general more considered as a technique for bulk analysis and in general not compatible with ultrahigh vacuum (UHV) conditions, it will not be discussed further in the context of this article.

## Instrumentation for SNMS

### Electron Beam Postionization

After the advantages of SNMS for quantitative surface analysis had been recognized, a number of improved systems for electron beam postionization were developed. A corresponding setup is schematically depicted in Figure 1. Such an arrangement can be used either for SIMS or SNMS studies alternatively by switching off or on the postionizing electron beam from an electron-emitting filament. Typically, electron energy of 50–80 eV is used for postionization purposes in e-beam SNMS employing total electron currents up to a few mA. Typical cross-sections of the electron beam in such arrangements are at best in order of several  $\text{mm}^2$ . In order to suppress the secondary ions sufficiently, a retarding voltage in the order of 100 V is applied between the sample and the mass spectrometer entrance. Otherwise it is possible that higher secondary ion signals influenced by the already mentioned matrix effects would superimpose on the corresponding SNMS intensities. Since an arrangement such as sketched in Figure 1 acts simultaneously as a residual gas analyzer, the postionized sputtered particles have to be separated from the thermal residual gas ions that are ionized with much higher efficiency because of their

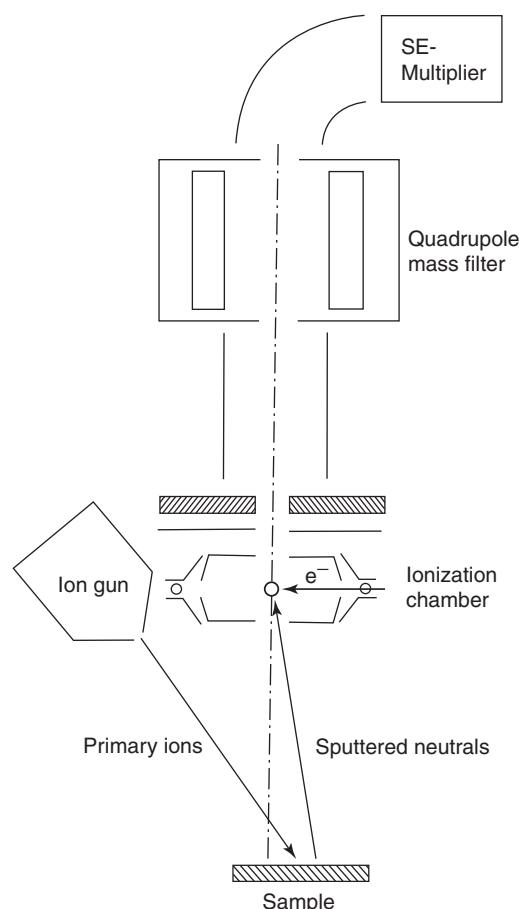


Figure 1 Schematic setup for electron beam SNMS.

low kinetic energy. This is achieved by another potential step between the ionization chamber and the mass spectrometer. The height of this potential step has to be adjusted such that the low-energy residual gas ions are retarded and only the faster fraction of postionized sputtered particle flux can enter the mass analyzer. Background signals can be reduced to a level of  $10^{-8}$  of the SNMS signals with an additional deflection capacitor between the residual gas suppression and the mass spectrometer and by choosing the bombarded spot at the sample surface slightly outside of the ion optical axis. In any case, one has to consider that low-energy sputtered particles from the sample surface being postionized in the electron beam can be trapped in the negative space charge well of the beam, and are thus lost for the analytical information.

### Electron Gas SNMS

Electron impact postionization with much higher efficiency is established when the sputtered neutral particles travel a longer distance through a volume

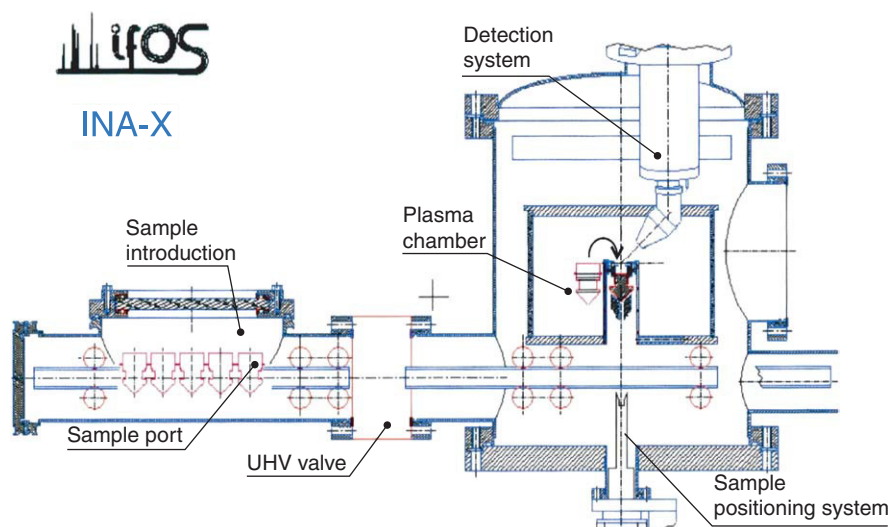


with a high density of electrons with sufficient energy. Such an 'electron gas' can be established in an effective way by the electron component of an electrodeless RF plasma maintained at gas pressures well below the usual glow discharge regime in a noble gas. Electron cyclotron wave resonance (ECWR) has proved to be a very suitable method for producing in Ar electron densities  $n_e$  of  $10^{10}$ – $10^{11}$  cm $^{-3}$  at pressures in the  $10^{-4}$ – $10^{-3}$  mbar range. Such high  $n_e$ -values are possible since the electron space charge is compensated by the background of positive plasma ions with energies only little above the thermal energies of the neutral gas particles. In contrast to the electron beam method, electron gas postionization is, therefore, not impeded by space charge effects in the ionizing volume. The electron component of an ECWR plasma in Ar was found to have a Maxwellian energy distribution with electron temperatures  $T_e$  equivalent to 10–15 eV, depending on the excitation conditions. Because of the low working pressure, heavy particle collisions can be neglected for traveling distances  $l$  of  $\sim 5$  cm that are usually employed for electron gas postionization. Plasma excitation by ECWR is well compatible with UHV conditions, and can be performed in an internal chamber inside a UHV vessel.

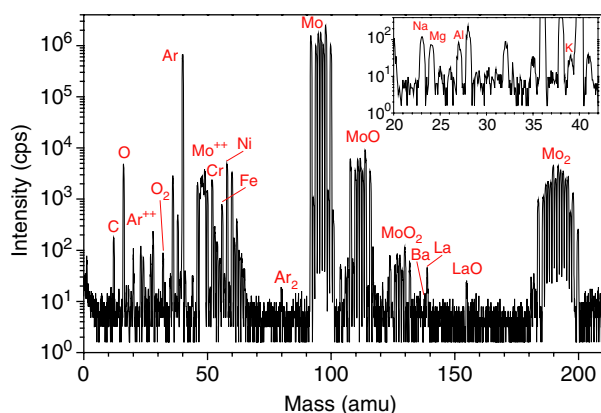
The layout of a commercial apparatus for electron gas SNMS is depicted in **Figure 2**. The sample head is introduced directly into the postionizing electron gas. Particle emission from the sample surface is excited by extracting ions from the ECWR plasma by a simple ion optical arrangement in front of the sample (direct bombardment mode, DBM). Since the angular distribution of sputtered particles changes from a torus-like configuration to an over-cosine behavior when the

bombarding ion energy is increased from the 100 eV regime to a few keV, the sputtered particle flux along the sample normal and, hence, the SNMS signals vary accordingly. In order to minimize the resulting element-specific dependence of the individual SNMS signals on the bombarding ion energy, a take-off angle of  $30^\circ$  against the normal to the sample is chosen. The neutral sputtered particles being postionized with an efficiency  $\alpha^0$  in the order of several  $10^{-2}$  are then guided to a quadrupole mass spectrometer via a  $45^\circ$  electrostatic deflector as an effective means to improve the signal-to-background ratio. Since the full energy distribution of the sputtered particles arrives at the entrance of the ion optical system, suppression of low-energy plasma ions can be readily established by a small positive potential step at the optics entrance. An SNMS spectrum obtained with the DBM from a sintered Mo sample is shown in **Figure 3**. The inset displays a mass interval with some low-concentration elements (or impurities) in the specimen. The retarding potential at the entrance lens of the ion optics was adjusted such that the peak of  $^{40}\text{Ar}$  ions from the ECWR plasma is reduced to the level of the main sample signals.

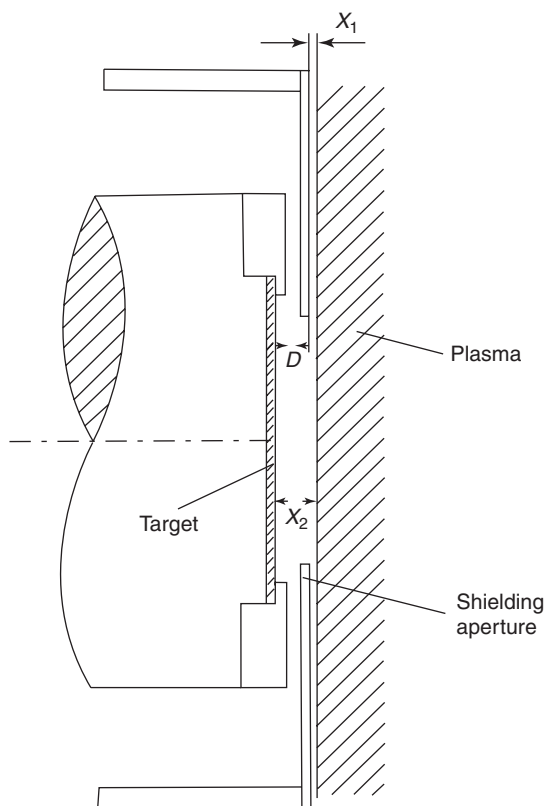
Whereas sample bombardment with an ion gun as employed in electron beam SNMS is usually performed at energies in the lower keV regime, much lower energies can be used in the DBM of electron gas SNMS. A schematic diagram of the corresponding ion extraction system is shown in **Figure 4**. By matching the distances  $D = X_2 - X_1$  properly for a given bombarding voltage applied between the ECWR plasma and the sample, an ideally plane plasma boundary can be established across a plane sample. Under such conditions the sample is



**Figure 2** Schematic of a commercial instrument for electron gas SNMS. (By IFOS GmbH, Kaiserslautern.)



**Figure 3** SNMS spectrum from a sintered Mo sample obtained with 1.8 keV  $\text{Ar}^+$  ions at  $2.1 \text{ mA cm}^{-2}$ .



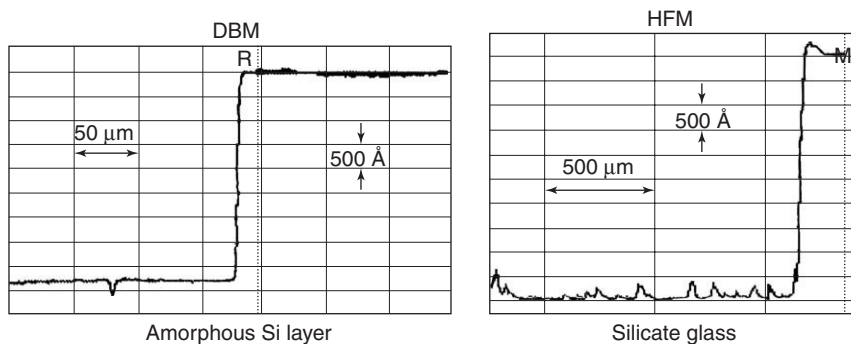
**Figure 4** Scheme of the ion extraction system for the direct bombardment mode of SNMS.

bombarded with extremely high lateral homogeneity at very low bombarding energies down to a few tens of eV at ion current densities in the order of  $1\text{--}2 \text{ mA cm}^{-2}$ . Thus, bombardment-induced atomic mixing effects at the sample become minimized and particle ejection is confined to the uppermost atomic layer. Both the high lateral homogeneity of sputter removal and the low bombarding energy make the

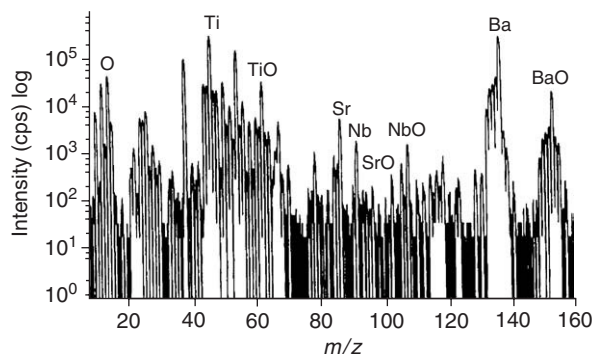
DBM extremely well suited for depth profile analysis. For amorphous samples the depth resolution obtained with this technique has been shown to be limited only by the atomic microroughness extending to  $\sim 4$  atomic distances or  $1\text{--}1.5 \text{ nm}$ , which develops as a consequence of the statistical sputter removal. A section of a bombarding crater obtained under such conditions is shown in **Figure 5A**.

The analysis of insulating samples at bombarding ion energies down to 100 eV or less becomes possible with the so-called high-frequency mode (HFM) of SNMS. For this operation mode a square wave high-frequency voltage is applied to the sample instead of a constant DC voltage as employed for DBM. Such an RF voltage can be basically looked at as a DC voltage, providing constant ion optical conditions, which is switched off periodically to admit short flux pulses of plasma electrons onto the insulating sample for compensating the positive charge from the preceding ion bombardment interval. The frequency of the applied square wave RF voltage and the ratio between the time intervals for the ion and electron fluxes within each period have to be chosen such that the variation of the surface potential due to charging up during the ion bombardment interval remains negligibly small. The bombarding ion energy is then controlled by the amplitude of the square wave voltage, and the same conditions for the analysis of dielectric samples are achieved with the HFM as for conducting samples with the conventional DBM. This is demonstrated in **Figure 5B**, which shows the edge region of a bombarding crater in an electrically nonconducting glass. An SNMS spectrum obtained with the HFM technique for a completely nonconducting ceramic ( $\text{BaTiO}_3$ ) is shown in **Figure 6**. Despite the periodically interrupted ion bombardment with  $\text{Ar}^+$  ions of only 800 eV the signal intensities for the main components approach  $10^6 \text{ cps}$ . The possibility to identify low-concentration admixtures that control the dielectric constant of the material is essential in this case.

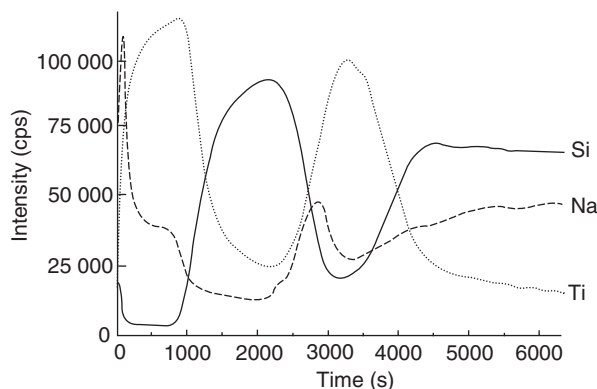
**Figure 7** depicts a schematic diagram of another operation mode of electron gas SNMS. In this 'external bombardment mode' (EBM) the sample is mounted outside the positionizing plasma chamber and bombarded with an external ion gun. Sputter-removed neutral surface particles enter the SNMS plasma through an electrical diaphragm that can be opened in any direction for charged particles of arbitrary sign. In EBM, the positive ion charge transferred to an insulating sample from the ion gun can be compensated by an electron current being extracted from the SNMS plasma through the diaphragm. The charge compensation becomes directly visible by the variation of the SNMS signals as a



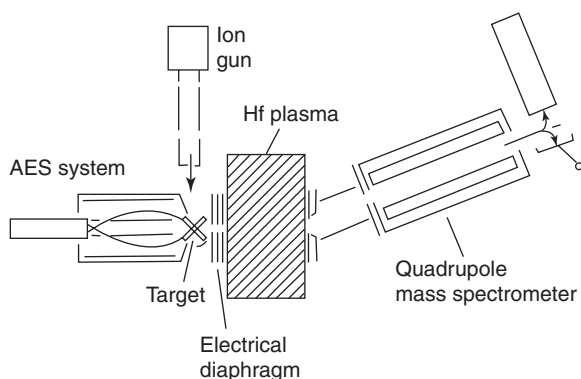
**Figure 5** Edge regions of ion bombarding craters produced in amorphous Si by the DBM technique with 200 eV  $\text{Ar}^+$ , and in a nonconducting silicate glass by the HFM version of SNMS with 480 eV  $\text{Ar}^+$ .



**Figure 6** SNMS spectrum acquired with HFM-SNMS from a ceramic  $\text{BaTiO}_3$  sample containing a number of minor admixtures. (Normal bombardment with 800 eV  $\text{Ar}^+$ ; ratio between ion-to-electron flux time intervals 1:1; operation frequency 500 kHz.)



**Figure 8** SNMS depth profile of an insulating oxidic layer structure (65 nm  $\text{TiO}_2$ /90 nm  $\text{SiO}_2$ /65 nm  $\text{TiO}_2$ /float glass). Analysis with charge compensated external bombardment mode with 5 keV  $\text{Kr}^+$ .



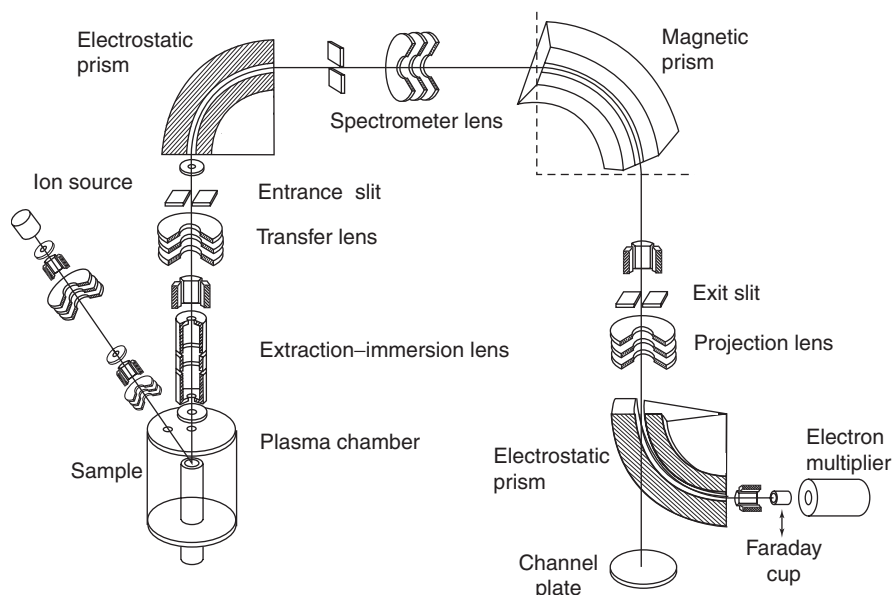
**Figure 7** Schematic diagram of the external bombardment mode of electron gas SNMS, including an Auger electron spectrometer for complementary analysis.

function of the potentials applied to the diaphragm electrodes: Without an electron flow to the insulating sample no SNMS signals appear, since sputtering is prevented by charging-up of the sample. For an appropriate choice of the diaphragm potentials the SNMS signals appear and become stationary at well-defined diaphragm voltages, because now a

low-energy electron current from the plasma penetrates to the specimen. Exact charge neutralization can be readily determined from the point where the SNMS signals just reach their saturation values.

An SNMS depth profile of an oxidic layer structure on a float glass substrate acquired with the EBM is presented in **Figure 8**. Complete charge compensation of the 5 keV  $\text{Kr}^+$  beam scanned across the sample surface was continuously maintained when profiling through the highly insulating sample structure with a 65 nm  $\text{TiO}_2$ /90 nm  $\text{SiO}_2$ /65 nm  $\text{TiO}_2$  layer stack on a glass substrate. Interesting features as the influence of the individual interfaces acting as diffusion barriers for Na, or the surface segregation of this component, become clearly visible.

When a highly focused ion beam, for example, of  $\text{Ga}^+$  ions from a liquid metal ion source with a beam diameter of typically a few 100 nm is employed in the EBM or directed to the sample through the ECWR plasma, a secondary neutral microprobe – a competitor to the secondary ion microprobe – is established. **Figure 9** shows a layout of a secondary neutral



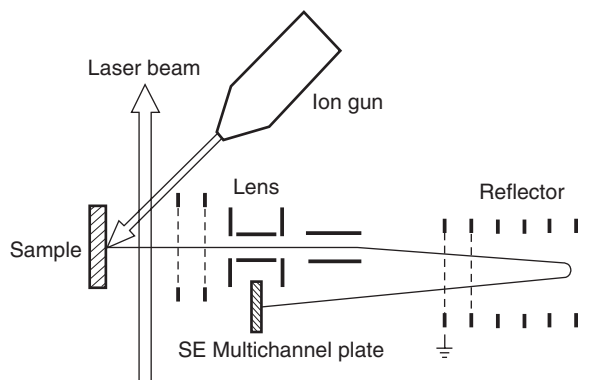
**Figure 9** Schematic layout of an electron gas secondary-neutral microprobe.

microprobe based on electron gas postionization. The plasma chamber is ion-optically coupled to a high transmission double-focusing mass spectrometer by a four-electrode electrostatic lens system and a dual quadrupole deflector. Such an instrument provides detection limits better than in the mg per g range when elemental surface distributions are imaged with a primary 20 keV  $\text{Ga}^+$  beam of a spot size of 1  $\mu\text{m}$ .

Although an SNMS microprobe is principally designed to obtain secondary neutral micrographs on inhomogeneous sample structures, it can also be operated in the DBM of electron gas SNMS. Because of the high transmission of the double focusing magnetic mass spectrometer, and since the mass independent background in the respective DBM spectra is below  $10^{-2}$  cps, detection limits in the nanogram-per-gram range are achieved.

### Laser SNMS

Postionization by resonant or nonresonant multiphoton absorption is mostly used in conjunction with a time-of-flight (TOF) mass spectrometer, since a postionizing laser pulse with a duration in the nanosecond, picosecond, or even subpicosecond regime provides an ideal start signal for the TOF measurements. A corresponding setup for laser SNMS is schematically shown in **Figure 10**. KrF excimer lasers operating at  $\lambda = 248.6$  nm corresponding to a photon energy of 4.99 eV are frequently employed for pulse widths in the nanosecond-range providing up to several 100 mJ per pulse. For shorter pulse times



**Figure 10** Scheme of a laser SNMS microprobe.

down to a few 100 fs often hybrid excimer – dye laser systems or frequency multiplexed Nd:YAG lasers are used. The latter extend into the VUV-regime with photon energies above 10 eV, which enable single photon ionization (SPI) for almost all elements. In contrast to REMPI, SPI provides another nonselective approach for laser postionization. It was originally elaborated for the detection of desorbed or sputtered organic molecules, but is also employed for inorganic surface analysis.

TOF mass spectrometers of the reflectron type are particularly suitable because they compensate, via different path lengths in an electrostatic reflector, for the time spreading caused by the energy distribution of the sputtered particles. In contrast to quadrupole or magnetic sector mass spectrometers, TOF instruments enable parallel detection of all species in a bunch of postionized sputtered particles. Repetition

frequencies of the postionizing lasers are often in the 10 Hz regime, but extend sometimes to the order of  $10^2$  Hz. Hence, sample bombardment can also be performed in a pulse mode, thus reducing the consumption of sample material in addition to the high particle economy achieved by the parallel detection in a TOF spectrometer. As a characteristic figure,  $\sim 10^{-10}$  g of sample material is consumed for ppm analysis.

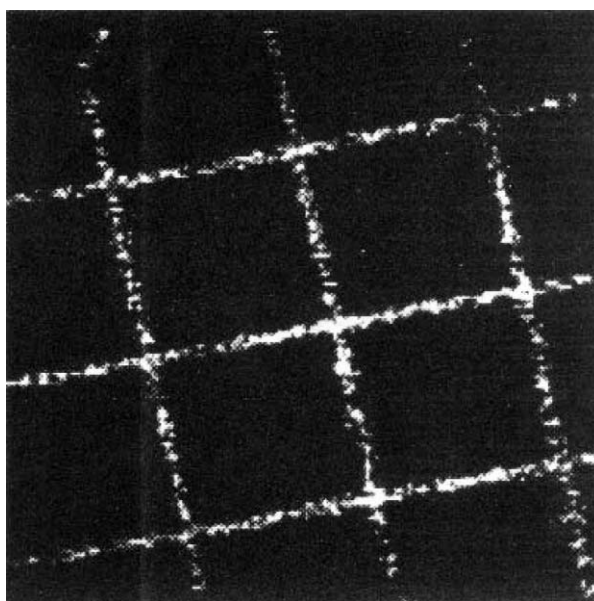
Laser SNMS requires the operation with properly selected duty cycles that control the delay times between the primary ion pulse, a pulsed extraction voltage for separating the secondary ions from postionized neutrals, and the firing of the postionizing laser pulse. Such duty cycles have, in addition, to be synchronized with the stepwise motion of the pulsed primary ion beam across the sample surface in the microprobe mode of laser SNMS. The selection of appropriate duration and decay times of the ion and laser pulses, of the laser intensity, and beam shape is important to make the photoion yields independent on the sputtered particle velocities. The detection volume must be matched to the entrance ion optics of the TOF such that it becomes independent of the individual ionization process. Usually, laser intensities in the range from  $10^6$  to  $10^{12}$  W cm $^{-2}$  are applied. While the particle density in the detection volume is monitored at small laser intensities, the particle flux is measured at high photon densities.

An example for surface imaging by laser SNMS is presented in **Figure 11**. The signal of the  $^{64}\text{Ni}$  isotope with an abundance of 1.1% was used to image a Ni grid pressed onto an Ag substrate. **Figure 11** demonstrates a detection power again below  $10^{-2}$  for surface microanalysis with laser SNMS. When an acquisition time of 1 s in the electron gas mode of SNMS is compared with signal collection over 100 laser pulses in the nonresonant photoionization mode, almost identical values of primary ion charge  $\sim 10^{-10}$  C per pixel are necessary for surface imaging with a lateral resolution  $\leq 1 \mu\text{m}$  and a detection power of  $10^{-3}$ – $10^{-2}$ .

## Surface Analysis and Depth Profiling with SNMS

### Quantitative Surface Analysis

The simple and straightforward quantifiability that is expected from the strict separation between particle ejection and ionization and has been demonstrated by many examples is the main justification for any postionization technique employed in the mass spectrometry of solids. The quantification procedure being outlined below has been developed primarily



**Figure 11** Laser SNMS micrograph of a Ni grid (20  $\mu\text{m}$  bars) displaying the  $^{64}\text{Ni}$  signal. Analysis conditions: 25 keV  $\text{Ga}^+$  of 6 nA with 1  $\mu\text{m}$  spot size; nonresonant multiphoton ionization at  $\lambda = 248 \text{ nm}$ . Viewed area  $1 \times 1 \text{ mm}^2$ , imaged with  $128 \times 128$  pixel during 164 s.

for the electron gas version, but applies also to the other SNMS techniques. The basic relation for the SNMS signal  $I(X^0)$  of a neutral species  $X$  being sputter released from the sample is given by

$$I(X^0) = I_p Y_x \alpha_x^0 \eta_x^v (1 - \alpha_x^+ - \alpha_x^-) \quad [2]$$

where  $I_p$  is the bombarding ion current onto the sample and  $\eta_x$  the detection efficiency for a postionized particle  $X^0$ . The quantity carrying the analytical information is the partial sputtering yield  $Y_x$ , which describes the formation probability of a particle  $X$  to be mass analyzed. The postionization probability  $\alpha_x^0$  for  $X$  must be corrected with regard to those particles contributing to  $Y_x$  but being ejected already as positive or negative secondary ions  $X^{+,-}$  with the respective ionization coefficients  $\alpha_x^{+,-}$ . As mentioned earlier, the values of  $\alpha_x^{+,-}$  in the vast majority of all cases are much smaller than unity. Hence, the term in parentheses in eqn [2] can generally be neglected without an appreciable loss in accuracy.

Since the postionization probability  $\alpha_x^0$  of an analyzed species  $X$  is determined by the postionization method and thus becomes a constant of the specific SNMS apparatus, the product  $\alpha_x^0 \cdot \eta_x$  can be combined to one apparatus constant  $D_x$ , which is the detection coefficient for  $X$  in the particular SNMS system. Equation [2] then reads:

$$I(X^0) = I_p Y_x D_x \quad [2a]$$



When the flux of sputtered particles contains only a negligible fraction of molecules, which has been shown to be the case for samples with metallic or covalent bonding, the partial sputtering yield  $Y_x$  under stationary condition is given by

$$Y_x = c_x \cdot Y_{\text{tot}} \quad [3]$$

where  $c_x$  is the atomic concentration of a sample component  $X$  and  $Y_{\text{tot}}$  the total sputtering yield under the respective bombarding conditions. When eqn [2a] is combined with eqn [3], the atomic concentrations of two elements  $i$  and  $j$  are related by

$$\frac{c_i}{c_j} = \frac{I(X_i^0) D_j}{I(X_j^0) D_i} \quad [4]$$

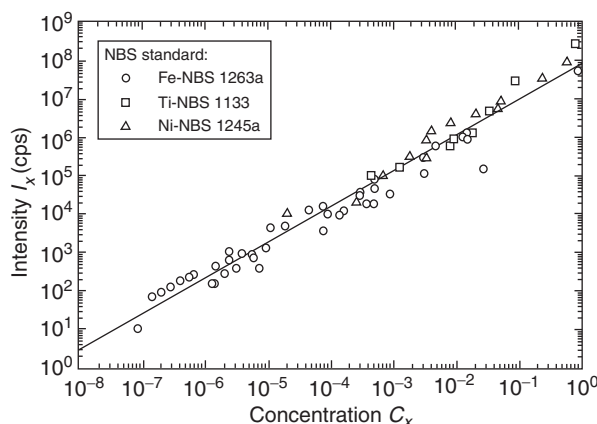
By introducing relative detection factors  $D_i^{\text{rel}}$ , all individual  $D_x$  are referred to the value of  $D_R$  for one identical reference element  $R$ , i.e.,  $D_x^{\text{rel}} = D_x/D_R$ . Since  $\sum c_x = 1$ , the relation

$$c_x = \frac{I(X^0)/D_x^{\text{rel}}}{\sum_i I(X_i^0)/D_i^{\text{rel}}} \quad [5]$$

is easily obtained for the concentration of the atomic constituent  $X$  in the sample.

For the quantification via eqn [5] all SNMS intensities have to be taken from the same SNMS spectrum or measured under identical experimental conditions. The values of  $D_i^{\text{rel}}$  can be easily determined from calibration measurements with standard samples of well-known composition. Since, particularly for low energies of the bombarding primary ions, the energy and angular distributions of the sputtered neutrals can be subjected to non-negligible changes, the absolute detection factors  $D_x$  may too become dependent from the bombarding energy. The relative detection factors  $D_i^{\text{rel}}$  will be less affected by such influences, especially for an oblique take-off angle as discussed earlier.

In Figure 12, the SNMS intensities taken directly from the mass spectra obtained with electron gas SNMS for different standard samples with specified compositions are plotted versus the concentrations  $c_x$  of the sample constituents. It can be seen that the relative detection factors for electron gas SNMS vary only little for a large variety of elements. Figure 12 demonstrates that the SNMS signals obtained with the electron gas method describe the sample composition within a factor of  $\sim 2$  without any further evaluation procedure being employed. Such a simple and very convenient behavior can be well understood from eqn [1] since the integral for  $\alpha_x^0$  is not much sensitive to the variations of the individual functions referring to an analyzed species  $X$ . This will be



**Figure 12** SNMS signals of various elements in three different standards vs. their specified concentrations.

different for electron beam SNMS where  $\alpha_x^0$  depends on the position of the electron beam energy along the ionization function  $Q_x$  of a species  $X$ . Large differences between individual detection factors occur for laser SNMS, particularly when near-resonance conditions for photon absorption are met for certain species  $X$ . For the same reason the detection coefficients in laser SNMS vary also with the photon energy. Different variations of the photoionization probability with the applied laser power (see Figure 13) as discussed earlier are still an obstacle for the quantification of laser SNMS, when saturation of the postionization process for all species  $X$  is not achieved.

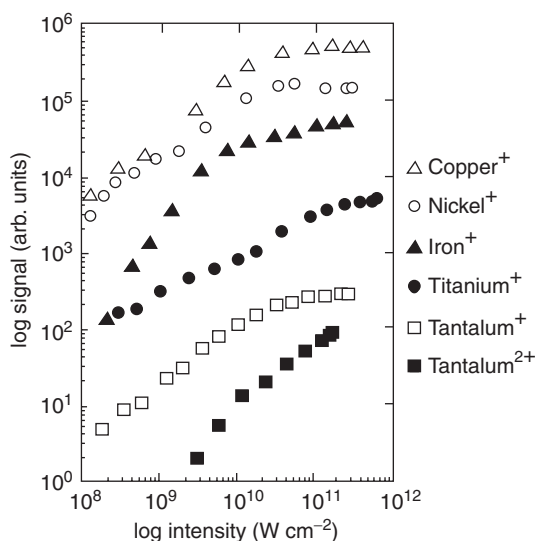
The small variations of the relative detection factors in electron gas SNMS for the different species  $X$  can be reduced significantly by putting

$$D_x \sim A_x^{1/2} / U_x^{\text{ions}} \quad [6]$$

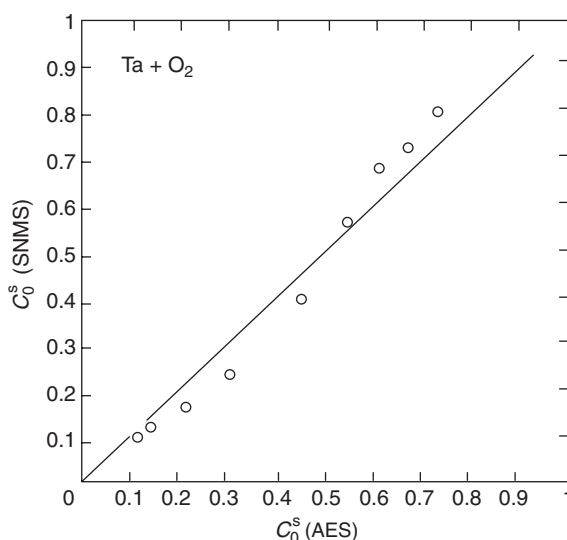
where  $A_x$  is the atomic mass number and  $U_x^{\text{ions}}$  the ionization potential of  $X$ . Equation [6] results readily from plausibility arguments on the individual dwelling times in the postionizing volume and the postionization probabilities for different species  $X$ . A much higher precision, however, is obtained with the evaluation procedure from eqn [5] when employing the exact relative detection factors for the respective SNMS instrument. An example is given in Figure 14, where for an unknown multielement sample the concentrations  $c_x$  obtained using eqn [5] from the corresponding SNMS spectrum are compared with the results from conventional analysis by wet chemistry.

The basic eqn [2] is also valid for neutral sputtered molecules. In that case the influence of electron impact dissociation can be included in the respective detection factors  $D_x$  for a molecular species  $X$ . As an

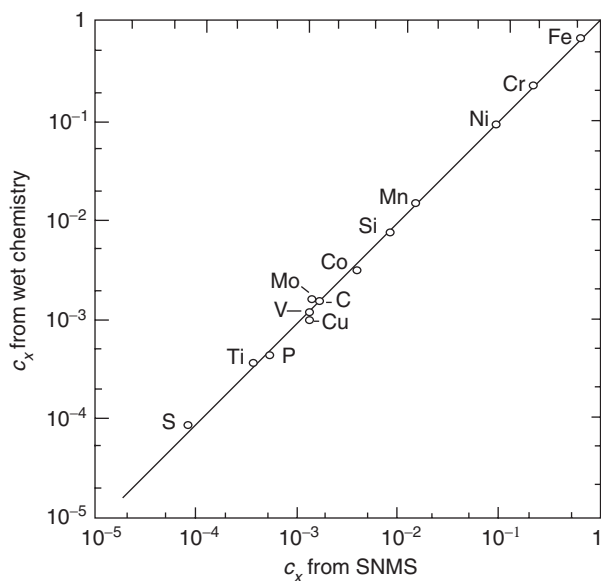




**Figure 13** Various photoion signals vs. laser intensity in non-resonant multiphoton ionization with a KrF excimer laser ( $\lambda = 248$  nm). (By S. Kaesdorf, H. Schröder and K.L. Kompa.)



**Figure 15** Correlation plot for atomic surface oxygen concentrations on polycrystalline Ta determined by SNMS and by *in situ* Auger electron spectroscopy AES.



**Figure 14** Correlation plot for the atomic concentrations in a metal sample determined by SNMS and conventional analysis by wet chemistry.

example for such species, diatomic oxide molecules like TaO with strong ionic bonds are sputter-generated via the so-called direct emission process. For a binary sample like an oxide or a hydride the formation probability for a sputtered molecule  $MB$  and, hence, the respective SNMS signal  $I(MB^0)$  are predicted to pass through a maximum for equal surface concentrations of  $c_B^S = c_M^S = 0.5$ .  $B$  refers to the

lighter and  $M$  to the heavier component, e.g., to oxygen and a metal or hydrogen and silicon. From the respective formation mechanism the relation

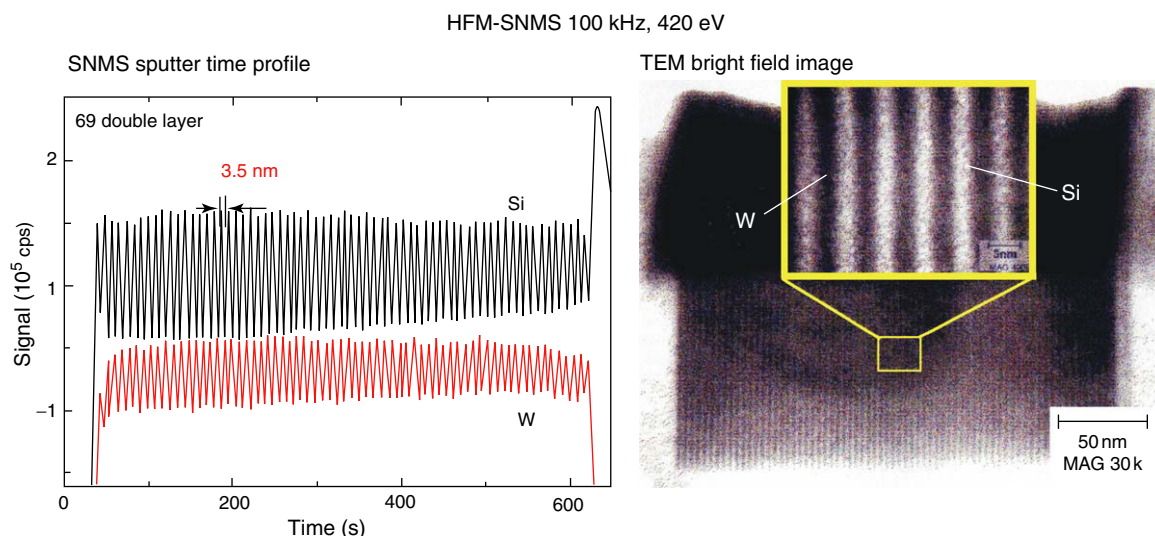
$$c_B^S = 0.5 \left[ \left\{ 1 - \frac{I(MB^0)}{I_{\max}(MB^0)} \right\}^{0.5} \right] \quad [7]$$

is readily derived for the actual surface concentration  $c_B^S$ .  $I_{\max}(MB^0)$  is the maximum SNMS signal for  $MB$  referring to  $c_B^S = 0.5$ .

A correlation plot for the  $c_0^S$  values derived from the variation of  $I(\text{TaO}^0)$  from an oxygen covered Ta surface via eqn [7] is shown in **Figure 15** where the SNMS results are compared with the respective oxygen surface concentration being determined *in situ* by AES. The convincing agreement between both sets of results for  $c_0^S$  proves simultaneously the direct emission model for molecular sputtering leading to eqn [7]. More important, eqn [7] makes SNMS to a self-calibrating surface analytical technique without the necessity of external standards when the conditions for the maximum formation probability of  $MB$  molecules, i.e., at  $c_B^S = c_M^S = 0.5$  for a binary system, are passed through during the analysis.

### High-Resolution Depth Profiling

The very favorable features of electron gas SNMS for high-resolution depth profiling being achieved with the DBM technique (see above) become exemplified by **Figure 16**. This figure shows the depth profile through a W-Si multilayer stack on a ceramic



**Figure 16** SNMS analysis of a W-Si multilayer structure. A cross-sectional TEM image of the layer stack is included (IFOS GmbH Kaiserslautern).

support. Profiling was performed with the high-frequency mode on SNMS (see above) using a rectangular high-frequency voltage of 100 kHz with a full amplitude of 420 V. For a double-layer thickness of only 3.6 nm the individual W- and Si-sublayers consisting of 7–8 atomic layers each become well resolved down to the substrate. The measured interface widths between the sublayers amount to 3–4 atomic distances and correspond well with the previously mentioned physical limitations for the depth resolution during sputter removal. For comparison a TEM image of a cross-section through the layer stack is added in **Figure 16**.

Depth profile analysis with SNMS also provides the possibility of determining the absolute depth scale quantitatively from the variation of the SNMS signals themselves. The respective evaluation procedure is based on the characteristic property of SNMS that all sputter-removed particles are collected with individual but constant factors when sputter profiling through a layer structure of varying composition.

The time-dependent sputter eroded depth  $z$  is given by

$$z(t) = \int_0^t j_p \frac{Y_{\text{tot}}(t')}{n(t')} dt' \quad [8]$$

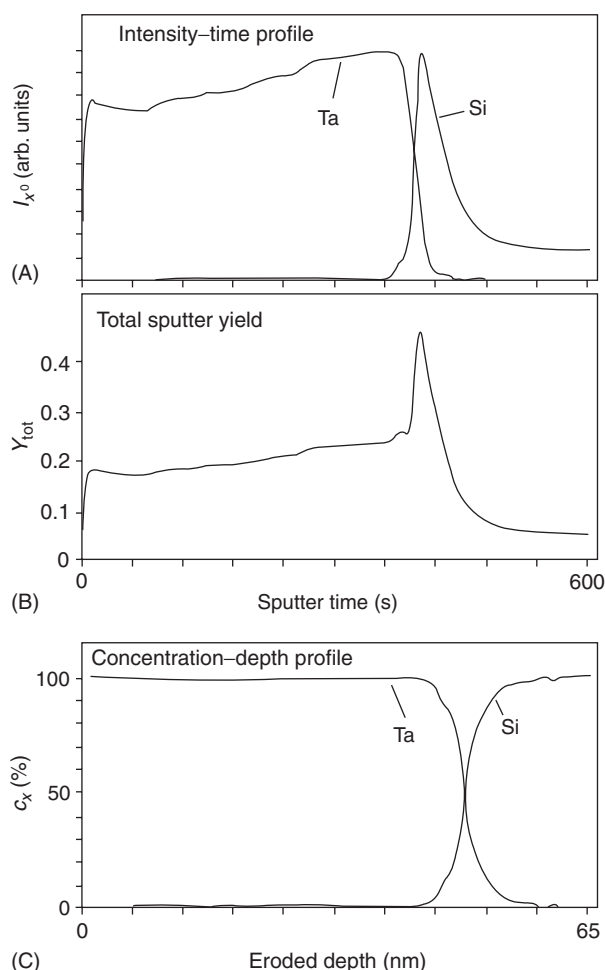
where  $j_p$  is the bombarding ion current density employed for the sputter removal and  $n$  is the time- (or depth-) dependent particle density in the sample. The time-dependent total sputtering yield  $Y_{\text{tot}}$  can be readily determined by adding all partial sputtering yields  $Y_x$ . These quantities are directly related with the time-dependent individual SNMS signals via eqn

[2a]. Finally, the time-dependent sample density can be approached by

$$n(t) = \sum_x c_x(t) n_x \quad [9]$$

where the  $c_x(t)$  are derived from the SNMS signals according to eqn [5]. As the only approximation entering the depth calibration procedure, the densities  $n_x$  are replaced by the respective bulk densities of the elemental bulk material X. By inserting  $Y_{\text{tot}}(t)$  and  $n(t)$  into eqn [8], the eroded depth  $z$  for a certain sputtering time is exclusively obtained from the time-dependent SNMS signals  $I(X^0, t)$ .

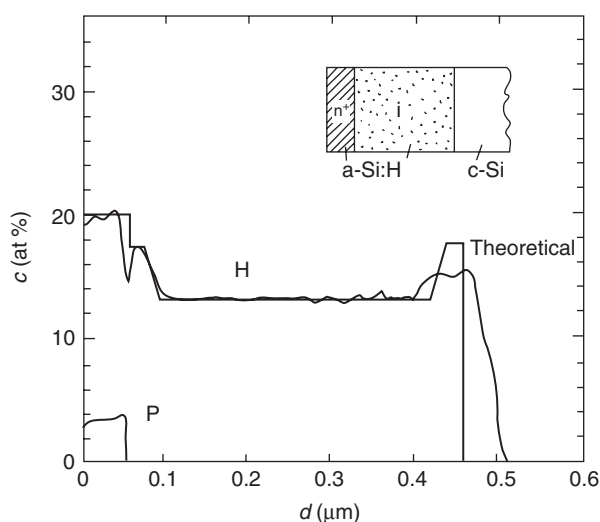
As an instructive example the results from sputter depth-profiling through a sputter-deposited Ta layer on a Si(100) substrate are shown in **Figure 17**. If literature values for the sputtering yields of Ta and Si are used to convert the sputter-time axis into a depth scale, a 84–16% interface width of 2.1 nm is obtained from the intensity-sputter time profile in **Figure 17**. However, the actual time- (or depth-) dependent total sputtering yield  $T_{\text{tot}}(t)$ , determined from the SNMS signals via eqn [2a] with  $D_{\text{Ta}} = D_{\text{Si}}$  as a good approximation, is found to vary with the sputter time as shown in **Figure 17B** with a peak at the interface region. The subsequent evaluation according to eqn [8] gives the true concentration-depth profile with a 84–16% interface width of 3.6 nm (**Figure 17C**). Hence, interface widths that are determined from sputter-time profiles employing surface analytical techniques become questionable in general, if time- or depth-dependent variations of the total sputtering yield (or the removal velocity) are not carefully taken into account. The obtained mean



**Figure 17** SNMS depth profiling of a sputter deposited Ta layer on Si(1 0 0) with 250 eV Ar<sup>+</sup> ions at 1 mA cm<sup>-2</sup>.

thickness of the Ta layer of 48 nm agrees with that expected from the deposition procedure and was independently confirmed through measurements with a step profilometer. The behavior of the SNMS signal, which is determined by the actual sputtering yield, points to structural inhomogeneities across the Ta film.

As with all mass spectrometric techniques, SNMS can be employed for isotope analysis, and is, in contrast to Auger electron spectroscopy, not restricted to elements with atomic numbers  $\geq 3$ . For example, quantitative analysis of the hydrogen content in a sample can – apart from the evaluation of the H-signal itself – be achieved from the ratio of the SNMS signals of the hydride molecule and the respective hydride forming element. According to the direct emission model for the generation of sputtered molecules as mentioned earlier, this ratio is directly proportional to  $c_H$ . An example is presented in **Figure 18** for the depth-dependent hydrogen concentration in a



**Figure 18** Hydrogen depth profile measured with SNMS in an amorphous hydrogenated Si layer with a P-doped n-layer on top. Curve 'theoretical' depicts the H-profile expected from the partial pressure variations during the sputter deposition process.

hydrogenated amorphous silicon structure that was deposited under varying partial hydrogen pressure  $p(\text{H}_2)$ . The solid line in **Figure 18** represents the hydrogen profile that is expected from the variation of  $p(\text{H}_2)$  during deposition. As a detail the P-concentration in a phosphorous-doped layer on top of this structure as determined from the respective electron gas SNMS measurements is included.

*See also: Surface Analysis: Overview; X-Ray Photoelectron Spectroscopy; Auger Electron Spectroscopy; Desorption Techniques; Ion Scattering; Secondary Ion Mass Spectrometry of Polymers; Laser Ionization.*

## Further Reading

- Becker CH and Gillen KT (1984) Surface analysis by non-resonant multiphoton ionization of desorbed or sputtered species. *Analytical Chemistry* 56: 1671–1677.
- Jede R, Ganschow O, and Kaiser U (1992) Sputtered neutral mass spectrometry (SNMS). In: Briggs D and Seah MP (eds.) *Practical Surface Analysis*, 2nd edn., vol. 2, pp. 425–506. New York: Wiley.
- Kaesdorf S, Hartmann M, Schröder H, and Kompa KL (1992) Influence of laser parameters on the detection efficiency of sputtered neutrals mass spectrometry based on non-resonant multiphoton ionization. *International Journal of Mass Spectrometry and Ion Processes* 116: 219–247.
- Oechsner H (1984) Secondary neutral mass spectrometry (SNMS) and its application to depth profile and interface analysis. In: Oechsner H (ed.) *Thin Film and Depth Profile Analysis*. Topics in Current Physics., vol. 37, pp. 63–85. Berlin: Springer.

Oechsner H (1993) Inorganic mass spectrometry for surface and thin film analysis. *Analytica Chimica Acta* 283: 131–138.

Pellin MJ, Young CE, Calaway WF, *et al.* (1987) Sensitive, low damage surface analysis using resonance ionization

of sputtered atoms. *Nuclear Instruments and Methods in Physics Research B* 18: 446–451.

Wucher A (1993) Microanalysis of solid surfaces by secondary neutral mass spectrometry. *Fresenius Journal of Analytical Chemistry* 346: 3–10.

## Laser Ionization

**L Van Vaeck**, University of Antwerp, Antwerp, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Laser ionization of atomic or molecular neutrals in the gas phase offers an unsurpassed potential with respect to analytical specificity and sensitivity. Unlike other ionization methods, photon absorption imparts to the analyte a well-defined amount of energy. Addressing the transitions between specific energy levels allows one species in a complex environment to be analyzed with isotopic selectivity. Because lasers deliver sufficiently high fluxes to saturate the population of an excited state, virtually each neutral in the irradiated volume is ionized. Hence, the ionization efficiency,  $\sim 1$ , exceeds that of current methods by a factor of  $10^2$ – $10^4$ .

Sputtered neutrals mass spectrometry (SNMS) involves the measurement of the postionized neutrals, generated either on a macroscopic scale (e.g., atomization in a graphite furnace) or from a microscopic volume of a solid (e.g., by focused primary ion or laser beam). In comparison to current ionization methods in SNMS, e.g., electron and low-pressure plasma ionization, lasers allow sensitivity and specificity of the postionization step to be ultimately optimized. Surface analysis requires the initial neutrals to be generated exclusively from the upper few monolayers of a solid, preferentially with a lateral resolution in the micrometer range. Essentially, two methods exist. Laser microbeam irradiation evaporates neutrals from the upper 10–100 nm of a solid with a diffraction-limited lateral resolution (spot diameter of  $0.5\ \mu\text{m}$  at a wavelength ( $\lambda$ ) of 250 nm). Alternatively, bombardment of the solid with a limited dose ( $< 10^{12}\ \text{ions cm}^{-2}$ ) of primary ions with energy in the kiloelectronvolt range generates neutrals from only the upper (few) monolayers over a surface area of up to  $250\ \mu\text{m} \times 250\ \mu\text{m}$  (size depends on sputter and ionization yield). The monolayer erosion makes primary ion bombardment the

method of choice for ‘surface analysis by laser ionization’ (SALI).

Lasers and primary ion beams bring sufficient energy into the sample to break even the strongest bond in a molecule. In spite of the resulting atomization, the ultrafast energy deposition and desorption processes still allow a significant fraction of (even thermolabile) analytes to be released in the form of intact molecules. Hence, SALI can address elemental as well as molecular (and organic) analysis.

From the 1980s onwards, dedicated setups (and acronyms) have been developed, e.g., SALI (Becker and Gillen, 1984), sputter initiated resonance ionization spectrometry (SIRIS) (Parks, 1990), surface analysis by resonant ionization of sputtered atoms (SARISA, Pellin *et al.*, 1989), and the Chicago–Argonne resonance ionization mass analysis (CHARISMA, Ma *et al.*, 1995). The recent proliferation of commercial time-of-flight (TOF) secondary ion mass spectrometry (SIMS) instruments stimulates their conversion into SALI setups.

### Fundamentals

#### Postionization SNMS versus SIMS

When an atomic primary ion of 10–25 keV impacts on a solid, it penetrates into the sample over a distance of typically 10–50 nm and dissipates its energy by inelastic collisions with the sample atoms along its trajectory. The resulting displacement of atoms and destruction of the molecular structure start a complex series of electronic, vibrational, and ultrafast thermal processes that finally cause electrons, atoms, and intact molecules, radicals, and secondary ions to be set free from the upper monolayer(s). Depending on the composition of the analyte and matrix, the ratio  $I_{\text{SIMS}}$  of directly emitted secondary ions over neutrals may range from  $10^{-2}$  to  $10^{-6}$ . These large differences in  $I_{\text{SIMS}}$  and detection sensitivity severely hamper the application of SIMS for survey analysis. In contrast, the relative abundance of sputtered neutrals remains relatively constant (i.e., 99 and 99.9999% for  $I_{\text{SIMS}}$  of  $10^{-2}$  and  $10^{-6}$ , respectively). On the condition that the gas-phase ionization step

Oechsner H (1993) Inorganic mass spectrometry for surface and thin film analysis. *Analytica Chimica Acta* 283: 131–138.

Pellin MJ, Young CE, Calaway WF, *et al.* (1987) Sensitive, low damage surface analysis using resonance ionization

of sputtered atoms. *Nuclear Instruments and Methods in Physics Research B* 18: 446–451.

Wucher A (1993) Microanalysis of solid surfaces by secondary neutral mass spectrometry. *Fresenius Journal of Analytical Chemistry* 346: 3–10.

## Laser Ionization

**L Van Vaeck**, University of Antwerp, Antwerp, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Laser ionization of atomic or molecular neutrals in the gas phase offers an unsurpassed potential with respect to analytical specificity and sensitivity. Unlike other ionization methods, photon absorption imparts to the analyte a well-defined amount of energy. Addressing the transitions between specific energy levels allows one species in a complex environment to be analyzed with isotopic selectivity. Because lasers deliver sufficiently high fluxes to saturate the population of an excited state, virtually each neutral in the irradiated volume is ionized. Hence, the ionization efficiency,  $\sim 1$ , exceeds that of current methods by a factor of  $10^2$ – $10^4$ .

Sputtered neutrals mass spectrometry (SNMS) involves the measurement of the postionized neutrals, generated either on a macroscopic scale (e.g., atomization in a graphite furnace) or from a microscopic volume of a solid (e.g., by focused primary ion or laser beam). In comparison to current ionization methods in SNMS, e.g., electron and low-pressure plasma ionization, lasers allow sensitivity and specificity of the postionization step to be ultimately optimized. Surface analysis requires the initial neutrals to be generated exclusively from the upper few monolayers of a solid, preferentially with a lateral resolution in the micrometer range. Essentially, two methods exist. Laser microbeam irradiation evaporates neutrals from the upper 10–100 nm of a solid with a diffraction-limited lateral resolution (spot diameter of  $0.5\ \mu\text{m}$  at a wavelength ( $\lambda$ ) of 250 nm). Alternatively, bombardment of the solid with a limited dose ( $< 10^{12}\ \text{ions cm}^{-2}$ ) of primary ions with energy in the kiloelectronvolt range generates neutrals from only the upper (few) monolayers over a surface area of up to  $250\ \mu\text{m} \times 250\ \mu\text{m}$  (size depends on sputter and ionization yield). The monolayer erosion makes primary ion bombardment the

method of choice for ‘surface analysis by laser ionization’ (SALI).

Lasers and primary ion beams bring sufficient energy into the sample to break even the strongest bond in a molecule. In spite of the resulting atomization, the ultrafast energy deposition and desorption processes still allow a significant fraction of (even thermolabile) analytes to be released in the form of intact molecules. Hence, SALI can address elemental as well as molecular (and organic) analysis.

From the 1980s onwards, dedicated setups (and acronyms) have been developed, e.g., SALI (Becker and Gillen, 1984), sputter initiated resonance ionization spectrometry (SIRIS) (Parks, 1990), surface analysis by resonant ionization of sputtered atoms (SARISA, Pellin *et al.*, 1989), and the Chicago–Argonne resonance ionization mass analysis (CHARISMA, Ma *et al.*, 1995). The recent proliferation of commercial time-of-flight (TOF) secondary ion mass spectrometry (SIMS) instruments stimulates their conversion into SALI setups.

### Fundamentals

#### Postionization SNMS versus SIMS

When an atomic primary ion of 10–25 keV impacts on a solid, it penetrates into the sample over a distance of typically 10–50 nm and dissipates its energy by inelastic collisions with the sample atoms along its trajectory. The resulting displacement of atoms and destruction of the molecular structure start a complex series of electronic, vibrational, and ultrafast thermal processes that finally cause electrons, atoms, and intact molecules, radicals, and secondary ions to be set free from the upper monolayer(s). Depending on the composition of the analyte and matrix, the ratio  $I_{\text{SIMS}}$  of directly emitted secondary ions over neutrals may range from  $10^{-2}$  to  $10^{-6}$ . These large differences in  $I_{\text{SIMS}}$  and detection sensitivity severely hamper the application of SIMS for survey analysis. In contrast, the relative abundance of sputtered neutrals remains relatively constant (i.e., 99 and 99.9999% for  $I_{\text{SIMS}}$  of  $10^{-2}$  and  $10^{-6}$ , respectively). On the condition that the gas-phase ionization step

features a uniform yield for all analytes, the ion intensities in SALI directly reflect the composition of the sputtered cloud, which in turn corresponds to the sample composition (unless selective sputtering occurs). Moreover, postionization under vacuum conditions avoids the matrix effects that result from interspecies interactions during the one-step sputtering ionization in SIMS.

The basic design of a typical postlaser ionization SNMS instrument in **Figure 1** resembles that of a TOF SIMS setup except for the laser beam, which is aligned parallel to the sample at a distance of 1–3 mm to the surface. The beam diameter is typically 1 mm. Pulsed biasing of the sample holder and ion extraction electrode potential prevents directly emitted secondary ions and electrons to enter the mass spectrometer. The laser is fired when the optimal density of neutrals in the irradiated volume is reached.

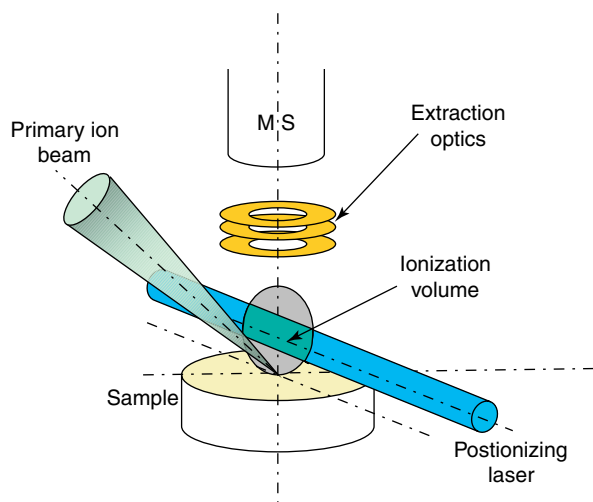
The number of parent ions ( $N_d$ ), detected from the upper monolayer when  $1\text{ mm}^2$  of sample is bombarded with  $N_{pi}$  primary projectiles, can be estimated by factorizing the following contributions:

$$N_d = N_s \times N_{pi} \times S \times G(\theta) \times D(E_{kin}, \Delta E_{kin}, \Delta t_p) \times I(N_\phi, \sigma) \times F(E_{exc}) \times T(MS, m/z, t, \dots) \quad [1]$$

where  $N_s$  = number of analyte molecules or atoms in  $1\text{ mm}^2$  of a monolayer.

$S$  = number of analyte neutrals sputtered by one primary ion.

$G(\theta)$  = geometry factor, i.e., fraction of  $S$  that is emitted with an angle within the solid angle of the ionization volume (seen from the sample surface).



**Figure 1** Schematic diagram of the basic SALI experiment (see text for details).

$D(E_{kin}, \Delta E_{kin}, \Delta t_p)$  = delay efficiency, i.e., fraction of  $S$  that is present in the ionization volume when the laser is 'fired'. The delay between the primary ion and laser pulse must account for the kinetic energy ( $E_{kin}$ ) of the neutrals (time needed to reach the ionization volume). The spread on  $E_{kin}$  and the duration of the primary ion pulse (or delayed release of neutrals) can reduce  $D$  significantly.

$I(N_\phi, \sigma)$  = yield of laser postionization, i.e., number of ions produced over number of neutrals present in the ionization volume. The yield can be expressed in terms of the cross-section for photon absorption  $\sigma$  (function of the analyte and photon energy  $E_\phi$ ) and the number of photons  $N_\phi$  in the ionization volume during the entire pulse.

$F(E_{int})$  = fragmentation factor ( $\leq 1$ ) taking into account that some parents are detected as daughter ions at a different  $m/z$ . Atomic ions are a favorable case with  $F = 1$  but in particular (polar) organic analytes readily undergo fragmentation depending on the internal stress  $E_{int}$ , i.e. (in a first approximation) the energy difference between the absorbed photons and ionization potential (IP).

$T(MS, m/z, t, \dots)$  = transmission and detection efficiency ( $\leq 1$ ), i.e., ratio of ions detected over those generated in the ionization volume. The factor depends on the instrumental optimization of ion extraction, transmission, and detector efficiency.

Introduction of realistic estimates for the different factors in eqn [1] evidences the potential advantage of laser postionization TOF-SNMS relative to TOF-SIMS. For instance, the upper monolayer of a copper target contains  $2.2 \times 10^{13}$  atoms  $\text{mm}^{-2}$ . Use of a 10 ns pulse from a primary ion beam ( $1\text{ }\mu\text{A mm}^{-2}$ ) gives  $N_{pi} = 6.24 \times 10^4$  projectiles per bombardment event. In a first approximation, each primary ion can be assumed to sputter constituents from a surface area with a diameter of 10 nm and a monolayer thickness. This volume contains  $1.7 \times 10^3$  Cu atoms. Postionization of Cu atoms with ArF or KrF excimer lasers can achieve complete ionization ( $I(N_\phi, \sigma_A) = 1$ ) while  $F = 1$  for atomic ions to stay on the safe side. The assumption of a laser beam with diameter of 1 mm at a distance of 3 mm to the surface allows a value of  $3.8 \times 10^{-2}$  to be calculated for  $G(\theta)$ .

The factor  $D$  is more difficult to derive. Sputtered Cu atoms ejected with a mean  $E_{kin}$  of  $\sim 3\text{ eV}$  have a velocity of  $3 \times 10^3\text{ ms}^{-1}$ . The time needed to reach and traverse the ionization volume is  $\sim 1$  and  $0.33\text{ }\mu\text{s}$ , respectively. Accounting for the emission angle and velocity distributions of the sputtered



neutrals,  $\sim 15\%$  of the Cu neutrals are within the ionization volume when a 10 ns laser pulse is fired at  $\sim 1.5 \mu\text{s}$  after the primary ion pulse. State-of-the-art instruments allow a combined efficiency  $T$  for ion extraction, transmission, and detection as high as 30% to be achieved.

Substituting the values in eqn [1] gives the number of ions, detected in laser TOF SNMS per primary ion pulse from  $1 \text{ mm}^2$  of copper:

$$\begin{aligned} N_d(\text{SNMS}) &= (1.7 \times 10^3) \times (6.24 \times 10^4) \times (6) \\ &\quad \times (3.8 \times 10^{-2}) \times (0.15) \times (1) \times (0.3) \\ &= 1.1 \times 10^6 \end{aligned} \quad [2]$$

The corresponding figure for the directly emitted secondary  $\text{Cu}^+$  ions in SIMS (without laser postionization) can be estimated using eqn [1] without the factors  $G$  and  $D$ . The ionization yield  $I_{\text{SIMS}}$  is  $10^{-4}$  under the specified conditions while the specific ion properties (emission angle and velocity distributions, mean  $E_{\text{kin}}$  and  $\Delta E_{\text{kin}}$ ) reduces  $T$  to 0.1. This gives:

$$N_d(\text{SIMS}) = N_s \times N_{\text{pi}} \times S \times I_{\text{SIMS}} \times F \times T \quad [3]$$

$$\begin{aligned} N_d(\text{SIMS}) &= (1.7 \times 10^3) \times (6.24 \times 10^4) \times (6) \\ &\quad \times (10^{-4}) \times (1) \times (0.1) = 6.4 \times 10^3 \end{aligned} \quad [4]$$

Otherwise stated, the detected number of  $\text{Cu}^+$  ions from a given primary ion pulse is  $\sim 200$  times higher in laser postionization SNMS than in SIMS. The situation becomes different for analytes with high  $I_{\text{SIMS}}$ . For instance, the  $I_{\text{SIMS}}$  of  $\text{Al}^+$  from aluminum targets is  $\sim 1.25 \times 10^{-2}$ , and thereby virtually annihilating the advantage of SNMS with respect to sensitivity. The reduction of matrix effects remains an advantage of SNMS.

### Laser Ionization Schemes

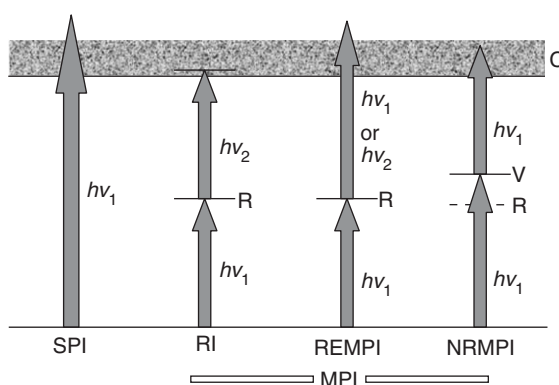
Photon absorption is by far the most elegant way to give neutrals a well-defined amount of energy, which is almost exclusively converted into electronic excitation.

Single photon ionization (SPI) requires  $E_\phi > \text{IP}$  (see Figure 2) with  $E_\phi$  given by

$$E_\phi = hc/\lambda \quad [5]$$

where  $h$  is Plank's constant and  $c$  the speed of light.

Modern lasers provide monochromatic beams in the UV and VUV range down to  $\sim 200 \text{ nm}$  ( $\text{IP} < 6 \text{ eV}$ ) while the excimer line of  $157 \text{ nm}$  allows transitions of  $7.9 \text{ eV}$  to be addressed. The overall ionization efficiency for elemental species primarily depends on the absorption probability, which is in turn a



**Figure 2** Survey of laser ionization schemes for gas-phase analytes: single photon ionization (SPI), multiphoton ionization (MPI) with distinction between resonant ionization (RI), resonance enhanced multiphoton ionization (REMPI), and nonresonant multiphoton ionization (NRMPI). Real and virtual states are denoted by R and S, respectively, while C refers to the ionization continuum.

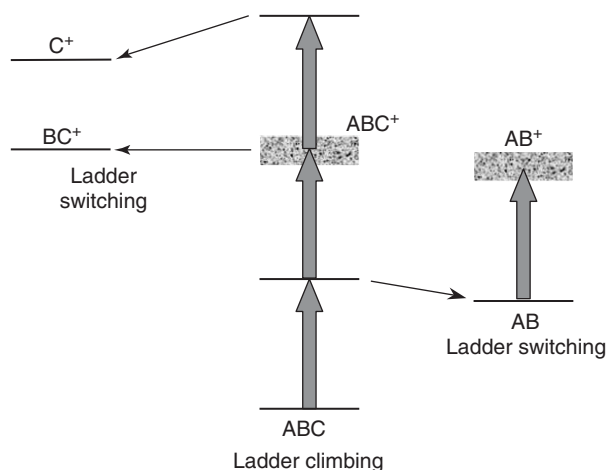
function of the density of the neutrals and the photon flux (intensity) in the ionization volume. Current lasers provide sufficient intensity to ionize nearly all atoms in the ionization volume. The situation is different for molecular species. As only little  $E_{\text{kin}}$  is given to the departing photoelectron, parent ions  $\text{M}^{+\bullet}$  obtain an amount of internal energy  $E_{\text{int}}$  corresponding to  $(E_\phi - \text{IP})$ . Unlike ionized atoms,  $\text{M}^{+\bullet}$  tend to relax their internal stress by fragmentation. Depending on the  $\lambda$ , the ratio of daughter ions over parent can be tuned for maximal yield of either  $\text{M}^{+\bullet}$  or fragment ions to obtain molecular weight or structural information, respectively. Unfortunately, each daughter costs the life of her parent and generation of fragments lowers the signal intensity of  $\text{M}^{+\bullet}$ .

Strictly speaking, distinction must be made between resonant and nonresonant SPI. In the resonant case, the  $E_\phi$  exactly matches the transition from the ground state into an autoionizing bond state above the IP. As a result, the high  $\sigma$  allows complete ionization to be achieved with lower laser intensities than when nonresonant SPI is used. In the latter case, the  $E_\phi$  drives the analyte 'somewhere' into the continuum above the IP.

Analytes with  $\text{IP} > 8 \text{ eV}$  require multiphoton ionization (MPI), i.e., consecutive absorption of two or more photons. A first photon transfers the analyte into a specific, electronically excited state and subsequent uptake of photon(s) is required to overcome the IP. The efficiency of this sequential process depends on the decay rate of the intermediate state relative to the photon flux density that promotes its population. The balance between the two competitive effects can be quantified by a quality factor as in

electric resonance circuits. The ultimate efficiency is obtained in resonant (R) MPI, often denoted as resonant ionization (RI). In this case, the  $E_\phi$  of the photons used exactly corresponds to the energy gaps between the ground level and the long-lived excited state on the one hand, and between that excited state and an autoionizing level on the other. Selectivity is optimized when molecules have low internal excitation. For instance, jet cooling of the gas-phase analytes reduces the population of vibrational states and the  $\sigma$  shows sharp maxima as a function of  $\lambda$ . Even isotope-labeled molecules can be discriminated from unlabelled ones. However, the situation is less ideal for the ion-beam sputtered neutrals because of their vibrational excitation. Therefore, resonance enhanced (RE) MPI is preferred in SALI.

Use of photons with a  $\lambda$  that does not match the transition to a long-lived intermediate state, allows nonresonant (NR) MPI to be performed by the 'quasisimultaneous' absorption of two or more photons. The sensitivity and selectivity of NRMPI is much lower than that of REMPI. The photon flux density (intensity) becomes a prime factor in NRMPI. Using laser pulses with a duration of  $<10$  ns and peak power densities above  $10^{10}$  and  $10^{13}$  W cm $^{-2}$  are required for NRMPI involving two and more photons, respectively. In the case of three-photon MPI of molecules, further increase of the flux density does not really improve the yield as a result of the so-called 'ladder mechanism'. Above a given threshold, the laser intensity primarily determines the extent of fragmentation. **Figure 3** illustrates the ladder concept



**Figure 3** The effects of fragmentation and ladder switching in the MPI process of a molecule ABC at high photon fluxes. (Reprinted from Vorsa V, Kono T, Willey KF, and Winograd N (1999) Femtosecond photoionisation of ion beam desorbed aliphatic and aromatic amino acids: fragmentation via  $\alpha$ -cleavage reactions. *Journal of Physical Chemistry B* 103: 7889–7895; © American Chemical Society.)

for a neutral ABC of which the ionization requires isoenergetic photons. The first photon brings ABC one step higher on its ladder, where competition occurs between uptake of another photon (ladder climbing, requires high flux density) or fragmentation into AB and C (ladder switching). In the latter case, the fragments AB and C can start their own process of ladder climbing. When a second photon brings the already excited ABC neutrals above the IP, fragmentation into, e.g.,  $BC^+$  competes with absorption of an additional photon yielding an excited state of  $ABC^{+\bullet}$ . The ladder concept explains the breakdown of stable molecules, such as benzene, into graphite-like  $C_n^+$  ions using photons of 4.8 eV.

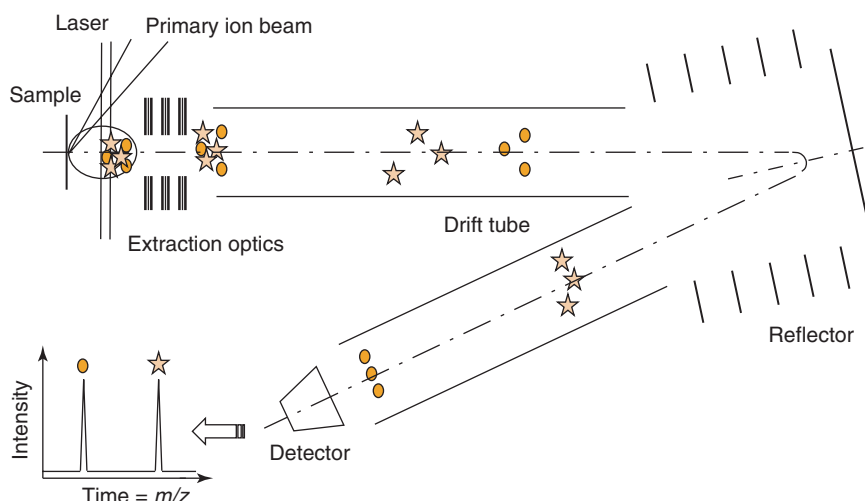
Ladder switching in organic molecules often involves processes of internal conversion (IC) and intersystem crossing (ISC), which bring molecules in vibrationally excited states and favor their fragmentation. As the rate constant of IC and ISC is typically  $<10^6$  and  $10^{11}$  s $^{-1}$ , respectively, use of lasers with a pulse duration in the femtosecond range makes excitation faster than the depopulation of the excited state and increases the yield of  $M^{+\bullet}$ . The growing availability of the tabletop ultrafast laser systems brings a tremendous potential for the ionization of large fragile organic molecules within the reach of the analytical chemist.

## Instrumentation

A TOF-mass analyzer is by far the instrument of choice for SALI applications. **Figure 4** illustrates the mass separation process. Electrostatic acceleration is the simplest way to distinguish between ions of different  $m/z$  in a given bunch. A potential difference  $V$  (typically 3 kV) between the sample and extraction electrode (on a distance of  $\sim 2$  mm) gives all ions the same  $E_{\text{kin}}$  and, hence, an  $m/z$ -dependent velocity  $v$ :

$$\frac{mv^2}{2} = zeV \Rightarrow v = \sqrt{2eV} \sqrt{\frac{z}{m}} \quad [6]$$

The obvious way to measure  $v$  consists in monitoring the time needed to travel a given distance in a field-free region. Hence, the accelerated ion bunch is sent through a tube of 1–2 m (kept on acceleration potential). After some time, the low  $m/z$  ions travel in front of the high  $m/z$  ions. The detector (channel plate with electron multiplier, pulse counting, and time-to-digital conversion) registers the arrival of ions as a function of time. Reference compounds are used to calibrate the time scale as a function of  $m/z$ . A basic requirement is that all ions start their travel in the field-free region at the same time, i.e., typically within 1–10 ns. Since the initial  $E_{\text{kin}}$  of the neutrals



**Figure 4** Schematic diagram of a typical SALI instrument with TOF mass analyzer.

before the ionization is kept during acceleration, the ions of the same  $m/z$  exhibit after acceleration a given  $\Delta E_{\text{kin}}$  and thereby  $\Delta v$ , which broadens the mass spectral peakwidth. Therefore, the ion reflector produces a retarding field aligned under a small angle with the first and second drift tubes. Ions of a given  $m/z$  with a higher  $E_{\text{kin}}$  penetrate deeper into the retarding field and spend more time on the longer path than the ions with lower  $E_{\text{kin}}$ . Adjustment of the retarding field allows, in principle, low and high  $E_{\text{kin}}$  ions of the same  $m/z$  to arrive at the same moment on the detector. The width of the mass spectral peak is reduced and the mass resolution improved without sacrificing sensitivity.

A TOF-MS provides several analytical advantages in comparison to magnetic sector, quadrupole, and ion trap analyzers. An ion beam is readily collimated to the diameter of the detector, allowing high transmission to be achieved. Because no fields have to be scanned, full mass spectra are recorded from each ion bunch. The efficiency, expressed as the ratio of number of ions used for mass analysis over the number of ions to be produced in the source, is high in the case of pulsed ionization. While mass separation requires ions to be introduced in the drift tube during only 1–10 ns, the mass analysis itself takes 100–500  $\mu\text{s}$ . Therefore, a pulsed primary ion gun is turned on during 1–10 ns and turned off during mass analysis. As a result, no sample is wasted. The  $m/z$  range in TOF-MS is only limited by the mass resolution of typically 1000–3000 (10% valley criterion) when ionization pulses of 1–10 ns are applied.

Currently used primary ions are  $\text{Ar}^+$ , produced by electron ionization (EI) of Ar, or  $\text{Ga}^+$ , generated by a liquid metal ion gun (LMIG). Typical beam spot diameters on the sample are 10–25  $\mu\text{m}$  and 50–100 nm

for the EI source and LMIG, respectively. The primary ion energy ranges from 5 to 25 keV. The beam is pulsed by electrostatic bunching. An incidence angle of  $\sim 45^\circ$  with the normal on the sample optimizes the sputtering yield. The release of many secondary electrons alters the surface potential of nonconducting samples and thereby the extraction field. Electron flooding, a conductive coating or mask on the sample, partly solves the problem. Pulsing the voltages on the sample holder and extraction optics keeps directly emitted secondary electrons and ions out of the analyzer.

Alignment of the postionizing laser as close as possible to the sample surface improves  $G(\theta)$ . Depending on the selected ionization scheme, irradiation is achieved with one or more lasers simultaneously. The delay between the primary ion and laser pulses must match the time needed for the neutrals to reach the ionization volume. Hence, adjustment of this parameter is critical to maximize the density of neutrals in the ionization volume. The laser pulse duration of typically 1–20 ns is selected to optimize both ionization probability and mass resolution. The spread on velocity and direction of the neutrals motivates application of relatively long pulses. As a result, the mass resolution in SALI experiments is lower than that for the SIMS mode with the same TOF instrument.

A vacuum of at least  $10^{-8}$  Torr is needed to avoid deposition of contaminants on the sample. Specifically, it takes only 1 s at a pressure of  $10^{-6}$  Torr for a residual vacuum component with sticking coefficient of 1 to form a complete monolayer on the surface.

Apart from the Perkin–Elmer SALI Model 7000, most instruments were developed in house. While the original SALI setup of Becker and Gillen (1984) used

an ion reflector, the SIRIS instrument of Atom Sciences (Oak Ridge, Tennessee) used a linear geometry. The Nd:YAG pumped tunable dye laser and/or mixing with the YAG fundamentals allowed photons with  $\lambda$  between 220 and 570 nm to be used. Isotope selectivity was not possible because the laser bandwidth exceeded the natural linewidth of the atoms. No ion reflector was needed because the RI avoided generation of isobars. The sensitivity for elements was in the low nanogram per gram range. The SARISA instrument built at the Argonne National Laboratories generated neutrals with a 3.5 keV  $\text{Ar}^+$  primary ion beam, focused to a spot with diameter of 250  $\mu\text{m}$ . In contrast to most setups, the primary ion beam impinged perpendicularly on the sample using ingenious ion optics that allowed photoionized neutrals to be extracted along the same axis. A so-called 'energy and angle refocusing' TOF system with two spherical analyzers optimized both mass resolution and acceptance of ions with different  $E_{\text{kin}}$  and emission angles. Elimination of the interference from  $^{56}\text{Fe}^+$  and  $^{28}\text{Si}_2^+$  by the RI scheme yielded a detection limit of 2 ng per g for Fe in Si, which is a demanding key analysis in semiconductor applications. The collaboration between the Chicago University and the Argonne National Laboratories yielded the beautiful acronym 'CHARISMA' for an instrument with elaborated laser optics permitting SALI and laser microprobe mass spectrometry (LMMS) with laser postionization to be compared directly. Primary ion bombardment was achieved either by 5 keV  $\text{Ar}^+$  or  $\text{Ga}^+$  from an LMIG with spot diameters of 500 and 0.1–1  $\mu\text{m}$ , respectively, while a Schwarzschild microscope was used to focus the laser on the sample in the LMMS mode. In both cases, the postionizing laser beam with a diameter of 100  $\mu\text{m}$  was aligned on a distance of only 0.2–0.5 mm above the sample. An in house developed laser system produced pulses as short as 100 fs at  $\lambda = 248$  nm for ultrafast MPI.

## Applications

Virtual absence of commercial instruments makes SALI a niche technique, particularly appreciated in areas where other methods fail. Specifically for elemental analysis, the use of SALI instead of the experimentally more simple and less expensive (S-)SIMS method is motivated by:

- The superior sensitivity for elements with low  $I_{\text{SIMS}}$ .
- The relatively uniform ionization yield for most elements (unlike SIMS) when the population of excited states is driven into saturation under NR

conditions. Unless preferential sputtering occurs, relative peak intensities of the atomic ions in SALI reflect directly the stoichiometry of the sample. This quick 'look and see' analysis is a valuable asset in any analytical strategy.

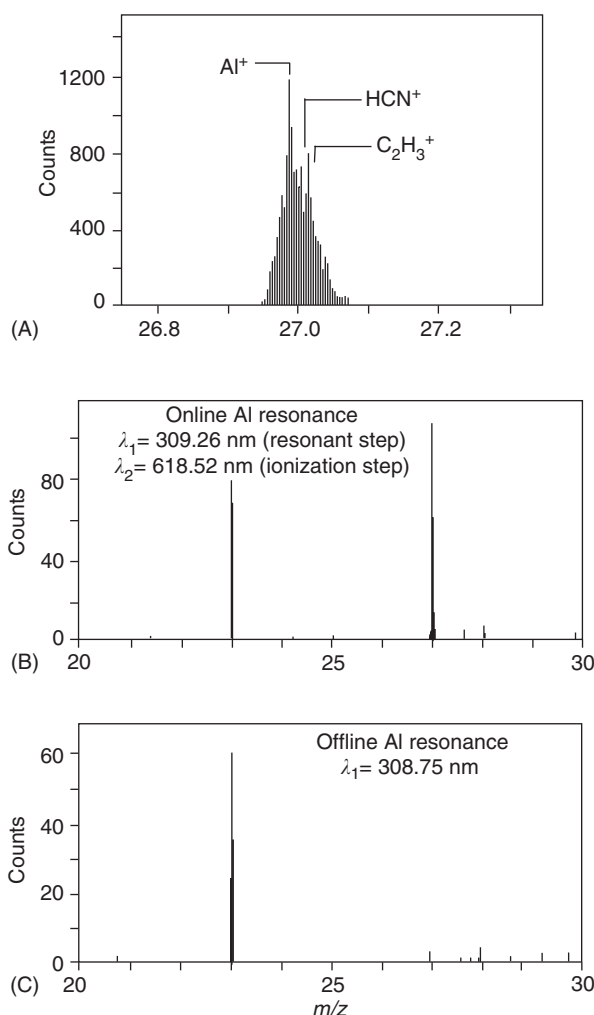
- Reduction of the matrix effects because ionization of neutrals in vacuum limits the interaction between constituents in the analyzed microvolume. Residual matrix effects are due to the sputtering step but remain orders of magnitude smaller than those for directly emitted secondary ions. Hence, quantification in SALI is by far less demanding with respect to calibration standards than SIMS.
- Elimination of isobaric interferences by the ionization itself, allowing lower mass resolution and higher transmission to be used in comparison with SIMS.

Figure 5 surveys the elements that can be addressed by laser postionization involving either SPI or MPI. The initial excitation of sputtered neutrals makes the investment of money and experimental efforts for RI often not paying in SNMS. In contrast, NR conditions are ideal for surveying the elemental composition. A typical example is the analysis of gallium arsenide, illustrated in Figure 6. The significant difference of  $I_{\text{SIMS}}$  for Ga in an As matrix and As in a Ga matrix hampers the quantitative analysis of major component analysis by means of the directly emitted secondary ions. In contrast, application of SALI yields peak intensities that correlate directly with the stoichiometric ratio of Ga and As over a wide range of compositions.

Figure 7 illustrates the advantage of SALI in comparison to SIMS for the detection of  $\text{Al}^+$  in a brain tissue sample. Use of the directly emitted secondary ions in the analysis of a sample containing 500  $\mu\text{g}$  per g Al requires the  $\text{Al}^+$  ions ( $m/z$  26.98) to be separated from the isobars due to  $^{26}\text{MgH}^+$  ( $m/z$  26.99),  $\text{C}^{15}\text{N}^+$  ( $m/z$  27.00),  $\text{HCN}^{+\bullet}$  ( $m/z$  27.01),  $\text{C}_2\text{H}_3^+$  (27.02). The presence of the first two interfering ions has been verified with high mass resolution mass spectrometry but cannot be traced back in the TOF-MS signal. The peak overlap prevents accurate quantification of the Al contribution. Application of an REMPI scheme in SALI eliminates the interfering ions and brings the detection limit to  $\sim 3$   $\mu\text{g}$  per g (resonant  $\lambda$  of 309.26 nm).

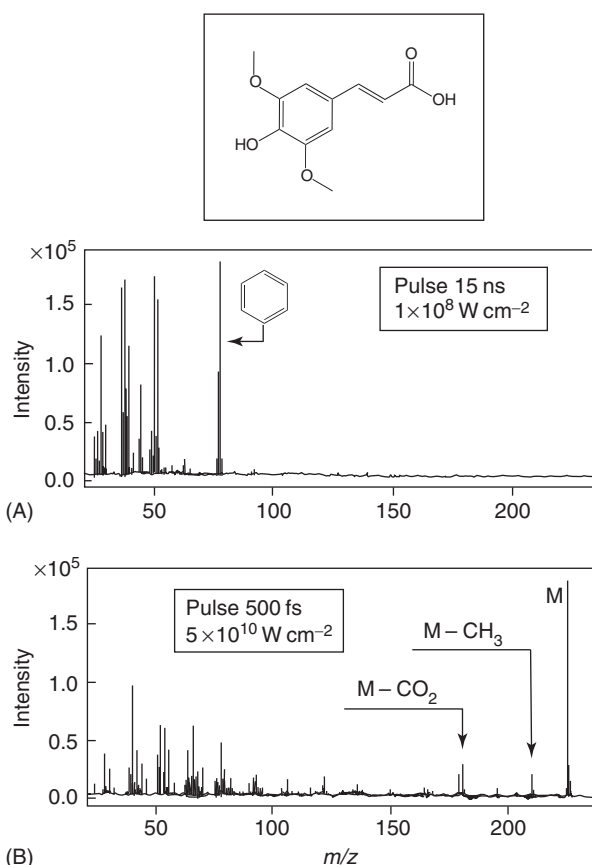
The relative insensitivity of SALI to matrix effects in comparison to SIMS makes the method particularly suited to the characterization of surface coatings, e.g., inorganic overlayers of nitrides and oxides, intentionally or accidentally deposited organic layers on an inorganic surface. Continuing bombardment of the sample with an ion dose above





**Figure 7** Positive ion mass spectra detected from a cortical tissue sample. (A) Isobars present detected around  $m/z$  27 in the SIMS mode without postionization. Mass spectra recorded using postionization under resonance (B) and nonresonant (C) conditions. (Reprinted from Jones OR, Perks RM, Abraham CJ, Telle HH, and Oakley AE (1997) A comparison of the techniques of secondary ion mass spectrometry and resonance ionisation mass spectrometry for the analysis of potentially toxic element accumulation in neural tissue. *Rapid Communication in Mass Spectrometry* 11: 179–183; © John Wiley.)

organic applications with SALI. Specifically, the latter team has pioneered the use of REMPI for post-ionization of the neutrals generated by primary ion bombardment of (sub-)monolayers of polyaromatic hydrocarbons. The sensitivity of SALI experiments has exceeded that of SIMS by a factor of 100. Detection limits in the femtomole range have been reached for fragile biomolecules, such as tryptophane, serotonin, and  $17\beta$ -oestradiol, deposited on silicon. **Figure 8** illustrates the significant gain in molecular information obtained even in the case of the simple molecule sinapinic acid by the application



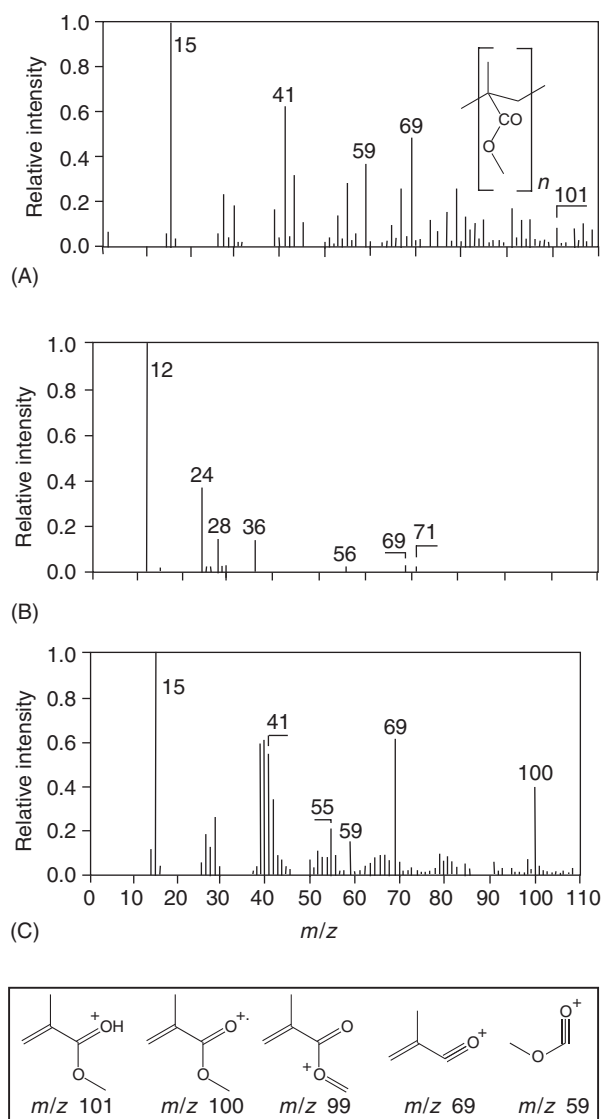
**Figure 8** Positive ion mass spectra taken from sinapinic acid using primary ion beam sputtering and laser postionization with 15 ns (A) and 500 fs (B) pulses of  $\lambda = 248$  nm. (Reprinted from Möllers R, Terhorst M, Niehuis E, and Benninghoven A (1992) Resonant photo-ionisation of sputtered organic molecules by femtosecond UV laser pulses. *Organic Mass Spectrometry* 27: 1393–1395.)

of 500 fs instead of 15 ns laser pulses ( $\lambda = 48$  nm). Ultrafast MPI with subnanoseconds laser pulses offers a particularly interesting potential for the analysis of labile substances such as organometallics and large biomolecules.

The use of an SPI scheme for the laser postionization of ion beam sputtered neutrals from polymers reduces the fragmentation in comparison to MPI-SALI. **Figure 9** compares the positive ion mass spectra recorded by S-SIMS, SALI, with SPI ( $\lambda = 118$  nm) and MPI (248 nm). Whereas an S-SIMS detects a small signal due to the protonated monomer, SALI with MPI exhibits extensive fragmentation causing the complete loss of the parent ion signal. In contrast, the limited fragmentation in SALI with SPI allows fragmentation the  $\text{M}^{+\bullet}$  of the monomer at  $m/z$  100 to be seen.

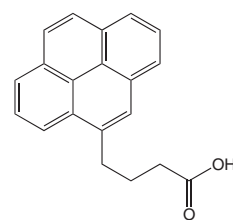
**Figure 10** illustrates the difference between the S-SIMS and SALI mass spectra of pyrene butyric acid,



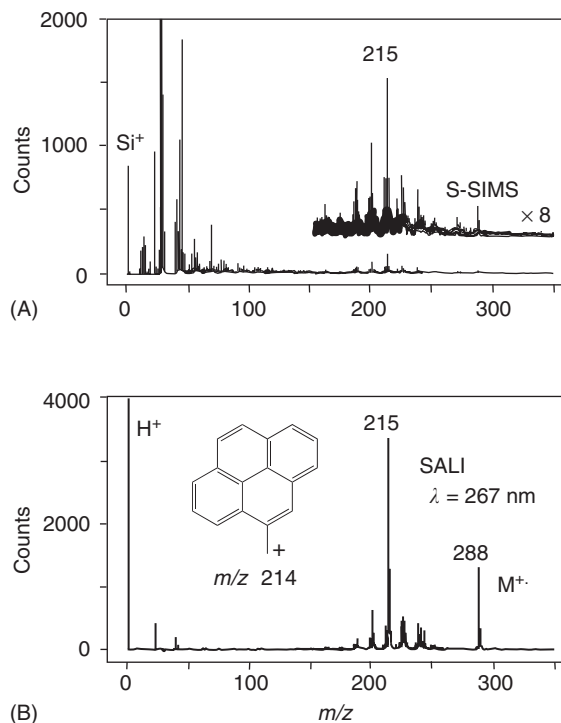


**Figure 9** A comparison of the positive ion mass spectra obtained from the surface of polymethylmethacrylate (PMMA) with  $\text{Ar}^+$  primary ions in the SIMS mode without postionization (A) and with laser postionization (B and C) using MPI ( $\lambda = 248 \text{ nm}$ , power density of  $10^7 \text{ W cm}^{-2}$ ) and SPI ( $\lambda = 118 \text{ nm}$ , power density =  $3 \times 10^3 \text{ W cm}^{-2}$ ), respectively. The structural assignment of major diagnostic ions is illustrated at the bottom. (Figure (A) is reproduced from Brown A and Vickerman JC (1986) A comparison of positive and negative static SIMS spectra of polymer surfaces. *Surface and Interface Analysis* 8: 75–81; and Figure (B) is reproduced from Pallix JB, Schuhle U, Becker CH, and Huestis DL (1989) Advantages of single photon ionisation over multiphoton ionisation for mass spectrometric surface analysis of bulk organic polymers. *Analytical Chemistry* 61: 805–811.)

bound to silver beads on a silicon substrate. The directly emitted secondary ions give virtually no signal in the molecular weight region and deprotonated methylpyrene fragments at  $m/z$  215 must be used to characterize the presence of the ad-layer. In contrast,



MW288



**Figure 10** Positive ion mass spectra of pyrene butyric acid desorbed from silver beads without (A) and with (B) laser postionization by means of 250 fs pulses at  $\lambda = 266 \text{ nm}$ . (Reprinted from Willey KF, Vorsa V, Braun RM, and Winograd N (1998) Postionization of molecules desorbed from surfaces by KeV ion bombardment with femtosecond lasers. *Rapid Communications in Mass Spectrometry* 12: 1253–1260.)

the intensity of the  $\text{M}^{+\bullet}$  signal in SALI is still 50% of the base peak due to the deprotonated methylpyrene. As a result, it has been possible to visualize the distribution of organic molecules at the surface of beads with a diameter of 40–100  $\mu\text{m}$  in the form of a ‘chemical’ image, in which the brightness and contrast directly reflects the abundance of a specific substance in the upper monolayer at the surface.

## Future Developments

The proliferation of adequate lasers and commercial TOF-SIMS instruments with advanced features for focused primary ion bombardment under static conditions, highly efficient mass analysis, and detection

as well as imaging is expected to make the virtues of SALI increasingly appreciated in material analysis. The recent use of polyatomic primary ions for increased sputtering yields with low ion-beam induced damage of the subsurface, which emerges as the most promising development in S-SIMS, has not been explored yet in SALI but ultimately offers increased possibilities for molecular detection and depth profiling of organic materials. At this moment, SALI is often considered as a niche technique to be exploited for the tasks that S-SIMS cannot handle because of isobaric interferences. The selectivity of RI and REMPI schemes reduces the need for high mass resolution. Simple TOF analyzers with high transmission and thereby favorable detection sensitivity are adequate. The laser postionization often yields a significant gain in ion yield in comparison to the direct emission of secondary ions in (S-)SIMS. The advent of ultrafast lasers with pulse durations in the subnanosecond range is foreseen to give additional impetus to the use of SALI in a variety of material science applications that require the characterization of the molecular composition and structure of the analytes in the ultimate surface layers.

**See also:** **Atomic Mass Spectrometry:** Laser Microprobe. **Mass Spectrometry:** Time-of-Flight. **Nitric Oxide. Surface Analysis:** Secondary Ion Mass Spectrometry of Polymers; Secondary Neutral Mass Spectrometry.

### Further Reading

- Lockyer NP (2001) Laser post-ionisation for molecular analysis. In: Vickerman JC and Briggs D (eds.) *TOF-SIMS: Surface Analysis by Mass Spectrometry*, pp. 417–445. Chichester: IM Publications and Surface Spectra Ltd.
- Pellin MJ, Young CE, Calaway WF, *et al.* (1987) Sensitive low-damage surface-analysis using resonance ionisation of sputtered atoms. *Nuclear Instruments and Methods in Physics Research B* 18: 446–451.
- Pellin MJ, Calaway WF, and Vervoykin IV (2001) Laser post-ionisation for quantitative elemental analysis. In: Vickerman JC and Briggs D (eds.) *TOF-SIMS: Surface Analysis by Mass Spectrometry*, pp. 375–415. Chichester: IM Publications and Surface Spectra Ltd.
- Wucher A (2001) Laser Post-ionisation: Fundamentals. In: Vickerman JC and Briggs D (eds.) *TOF-SIMS: Surface Analysis by Mass Spectrometry*, pp. 347–373. Chichester: IM Publications and Surface Spectra Ltd.

## Low-Energy Electron Diffraction

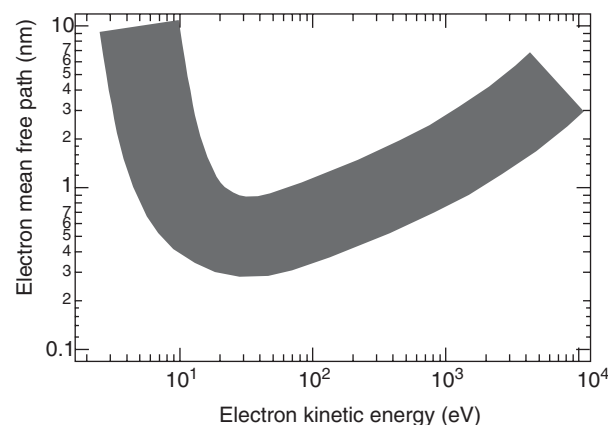
**K Sakamoto**, Tohoku University, Sendai, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

A diffraction experiment designed for crystal structure analysis requires a probe with de Broglie wavelength shorter than the interatomic spacing of the solid. The wavelength of an electron of energy  $E$  is given by  $E = (h/\lambda)^2/2m$ , and thus it is convenient to estimate the wavelength from the relation  $\lambda = 1.23/\sqrt{E}$  (nm), where the energy  $E$  is given in electronvolts. This indicates that wavelengths of electrons with energies of 20–300 eV, which are commonly called low-energy electrons, are 0.071–0.28 nm and thus meet the requirement for diffraction experiment. **Figure 1** shows the average mean free paths of electrons inside a solid (universal curve) in the energy range of 2–10 000 eV. As shown in **Figure 1**, the mean free path of low-energy electrons is in a range of 0.1–1 nm, a length that corresponds to a few atomic layers at low-index surfaces. The coincidence of the ideal wavelength and the surface sensitivity forms the basis for low-energy electron

diffraction (LEED) from solid surfaces. When low-energy electrons are reflected from an ordered crystal surface, the elastically backscattered electrons are distributed according to the two-dimensional (2D) grid of the surface and will form a Fraunhofer diffraction pattern. The Fraunhofer diffraction pattern is the Fourier transform of the surface atom arrangement, and thus it reflects the symmetry and the crystalline order of the surface. The qualitative



**Figure 1** Electron mean free path.

as well as imaging is expected to make the virtues of SALI increasingly appreciated in material analysis. The recent use of polyatomic primary ions for increased sputtering yields with low ion-beam induced damage of the subsurface, which emerges as the most promising development in S-SIMS, has not been explored yet in SALI but ultimately offers increased possibilities for molecular detection and depth profiling of organic materials. At this moment, SALI is often considered as a niche technique to be exploited for the tasks that S-SIMS cannot handle because of isobaric interferences. The selectivity of RI and REMPI schemes reduces the need for high mass resolution. Simple TOF analyzers with high transmission and thereby favorable detection sensitivity are adequate. The laser postionization often yields a significant gain in ion yield in comparison to the direct emission of secondary ions in (S-)SIMS. The advent of ultrafast lasers with pulse durations in the subnanosecond range is foreseen to give additional impetus to the use of SALI in a variety of material science applications that require the characterization of the molecular composition and structure of the analytes in the ultimate surface layers.

**See also:** **Atomic Mass Spectrometry:** Laser Microprobe. **Mass Spectrometry:** Time-of-Flight. **Nitric Oxide. Surface Analysis:** Secondary Ion Mass Spectrometry of Polymers; Secondary Neutral Mass Spectrometry.

### Further Reading

- Lockyer NP (2001) Laser post-ionisation for molecular analysis. In: Vickerman JC and Briggs D (eds.) *TOF-SIMS: Surface Analysis by Mass Spectrometry*, pp. 417–445. Chichester: IM Publications and Surface Spectra Ltd.
- Pellin MJ, Young CE, Calaway WF, *et al.* (1987) Sensitive low-damage surface-analysis using resonance ionisation of sputtered atoms. *Nuclear Instruments and Methods in Physics Research B* 18: 446–451.
- Pellin MJ, Calaway WF, and Vervovkin IV (2001) Laser post-ionisation for quantitative elemental analysis. In: Vickerman JC and Briggs D (eds.) *TOF-SIMS: Surface Analysis by Mass Spectrometry*, pp. 375–415. Chichester: IM Publications and Surface Spectra Ltd.
- Wucher A (2001) Laser Post-ionisation: Fundamentals. In: Vickerman JC and Briggs D (eds.) *TOF-SIMS: Surface Analysis by Mass Spectrometry*, pp. 347–373. Chichester: IM Publications and Surface Spectra Ltd.

## Low-Energy Electron Diffraction

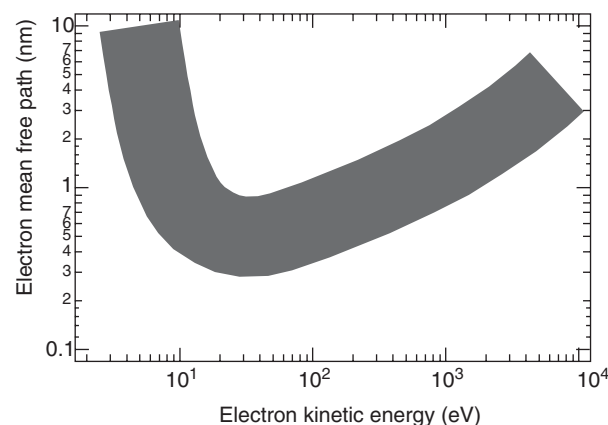
**K Sakamoto**, Tohoku University, Sendai, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

A diffraction experiment designed for crystal structure analysis requires a probe with de Broglie wavelength shorter than the interatomic spacing of the solid. The wavelength of an electron of energy  $E$  is given by  $E = (h/\lambda)^2/2m$ , and thus it is convenient to estimate the wavelength from the relation  $\lambda = 1.23/\sqrt{E}$  (nm), where the energy  $E$  is given in electronvolts. This indicates that wavelengths of electrons with energies of 20–300 eV, which are commonly called low-energy electrons, are 0.071–0.28 nm and thus meet the requirement for diffraction experiment. **Figure 1** shows the average mean free paths of electrons inside a solid (universal curve) in the energy range of 2–10 000 eV. As shown in **Figure 1**, the mean free path of low-energy electrons is in a range of 0.1–1 nm, a length that corresponds to a few atomic layers at low-index surfaces. The coincidence of the ideal wavelength and the surface sensitivity forms the basis for low-energy electron

diffraction (LEED) from solid surfaces. When low-energy electrons are reflected from an ordered crystal surface, the elastically backscattered electrons are distributed according to the two-dimensional (2D) grid of the surface and will form a Fraunhofer diffraction pattern. The Fraunhofer diffraction pattern is the Fourier transform of the surface atom arrangement, and thus it reflects the symmetry and the crystalline order of the surface. The qualitative



**Figure 1** Electron mean free path.

observation of the diffraction pattern is widely used to determine the periodicity and the symmetry of the surface structure and to identify ordered adsorbate phases.

The application of LEED is limited to ultrahigh vacuum (UHV) conditions because of the character of the probe and its high surface sensitivity. The main application is the study of single crystal metal and semiconductor surfaces, and adsorption on these surfaces. Study of surfaces of thin insulator films is also possible. Recently, the application extends to study the surface structure of quasicrystals. LEED is widely applied to determine the symmetry and periodicity of surface structures and thus to identify reconstructions and adsorbate phases by observing the diffraction pattern. It is also a technique for quantitative surface structure analysis. LEED is also applied as the quantitative determination of thermal vibrations at surfaces.

## Diffraction from Surfaces

A solid surface is intrinsically an imperfection of a crystalline solid by destroying the three-dimensional (3D) periodicity of the structure. That is, the unit cell of a crystal is usually chosen such that two vectors are parallel to the surface and the third vector is normal or oblique to the surface. Since there is no periodicity in the direction normal or oblique on the surface, a surface has a 2D periodicity that is parallel to the surface. By considering the symmetry properties of 2D lattices, one obtains the possible five 2D Bravais lattices shown in **Figure 2**. The combination of these five Bravais lattices with the 10 possible point groups leads to the possible 17 2D space groups. The symmetry of the surface is described by one of these 17 2D symmetry groups.

Surfaces frequently form superstructures by reconstruction and/or by adsorbate. The unit cells of these surface superstructures are usually describe by using the unit cell of the truncated bulk structure as  $(n \times m)$  or  $(n \times m)R\alpha^\circ$ , where  $n$  and  $m$  are integers or square

roots of integers that give the factors by which the translation vectors of the surface unit cell are enlarged compared to the bulk unit cell, and  $R\alpha^\circ$  denotes the rotation of the unit cell with respect to the substrate unit cell.

If an electron beam with wave vector  $\mathbf{k}_0$  is incident on a 2D system and the wave vector of the diffracted electron is  $\mathbf{k}_g$ , the conservation of energy gives

$$k_{0\parallel}^2 + k_{0\perp}^2 = k_{g\parallel}^2 + k_{g\perp}^2 \quad [1]$$

The suffixes  $\parallel$  and  $\perp$  denote the components parallel and perpendicular to the surface. The conservation of momentum gives

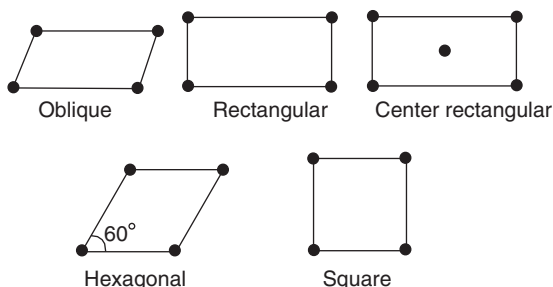
$$\mathbf{k}_{g\parallel} = \mathbf{k}_{0\parallel} + \mathbf{g}_{hk} \quad [2]$$

$\mathbf{g}_{hk}$  is a vector of the 2D reciprocal lattice

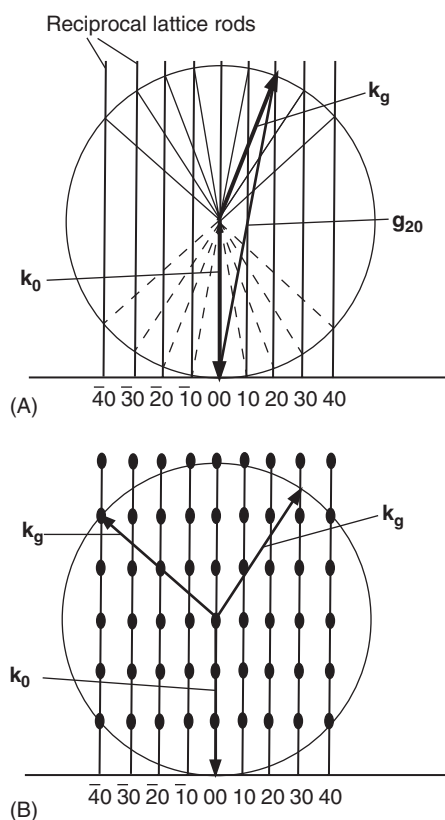
$$\mathbf{g}_{hk} = h\mathbf{a}^* + k\mathbf{b}^* \quad [3]$$

where  $\mathbf{a}^*$  and  $\mathbf{b}^*$  are the vectors of the reciprocal unit cell and  $h$  and  $k$  are integers. A convenient graphical representation of these Laue conditions in reciprocal space is the Ewald sphere construction. **Figure 3A** shows the Ewald construction for an electron beam incident normal to the surface. The wave vector of the incident electron beam,  $\mathbf{k}_0$ , is positioned with its end at the (00) reciprocal lattice point and a sphere with radius  $k$  is constructed around its starting point. Any line from the center of the sphere to the intersection point of the sphere and a reciprocal lattice rod represents a diffracted beam. One diffracted beam,  $\mathbf{k}_g$ , is shown in **Figure 3A** together with its associated reciprocal lattice vector  $\mathbf{g}_{20}$ . The diffracted beams are thus characterized by the index of the reciprocal lattice rods ( $hk$ ).

The presence of random steps or islands destroy the perfect 2D periodicity of the surface and thus the delta function character of the reciprocal lattice rods. Furthermore, the finite penetration depth of the low-energy electrons allows diffraction from not only the outermost atomic layer, but also from the 3D surface 'region'. This 3D scattering forms thicker regions in the reciprocal lattice rods according to the 3D Laue condition (**Figure 3B**). When the Ewald sphere crosses a thicker region, the diffracted beam has a strong intensity, and the intensity of the diffracted beam is weaker when the sphere crosses a thinner region. In the condition shown in **Figure 3B**, the intensities of the  $(\bar{4}0)$  spot and the  $(40)$  spot are stronger than the intensities of other spots. This leads to an important application of LEED. By changing the energy of the incident electron beam, one can change the radius of the Ewald sphere. The Ewald sphere passes through stronger and weaker regions of the rods as its radius



**Figure 2** The five 2D Bravais lattices.



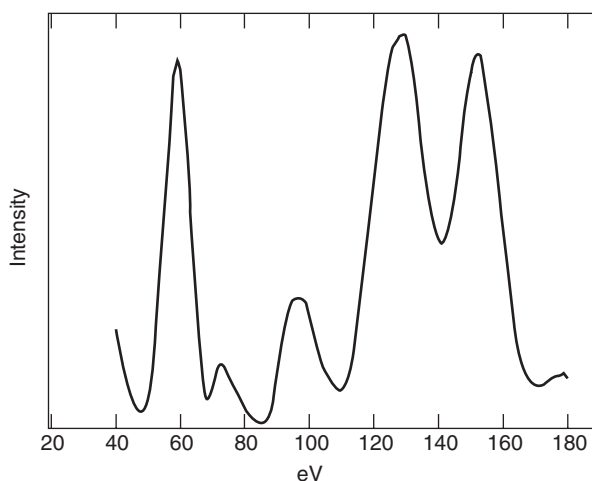
**Figure 3** Ewald sphere construction for an electron incident normal to the surface on (A) a perfect 2D system and (B) a quasi-2D system. In (A), the reciprocal lattice vector  $\mathbf{g}_{20}$  is shown associated with the relevant scattered wave. The dashed scattered wave vectors propagate into the bulk and thus are not observable.

varied. Measuring the intensity of a diffracted beam as a function of energy (the  $I/V$  curve ( $I$ , intensity;  $V$ , accelerating voltage of electrons)) and comparing the experimental result with model calculations is a usual method of structure analysis. A typical  $I/V$  curve is shown in Figure 4.

Although the most common  $I/V$  curve measurement is performed at normal incidence because the symmetry can be used to orient the sample, the intensity distribution in reciprocal space may also be measured by rotating the crystal about the surface normal or by changing the angle of incidence. However, these modes are rarely used, because of the experimental difficulty to measure the angles precisely.

## LEED Experiment

The first step in LEED experiments is the surface preparation. A surface with atomic-scale flatness is achieved by annealing a single crystal sample, which has a thickness of less than 1 mm and has at least one



**Figure 4** A typical LEED  $I/V$  curve.

face well oriented in the desired crystallographic orientation. Ion sputtering followed by additional annealing are used to remove impurities adsorbed on the surface and thus to obtain a clean ordered surface. Ordered adsorbate surfaces can be obtained by the adsorption of a monolayer or submonolayer of atoms or molecules onto the clean surface. The quality of the clean surface and/or the adsorbate surface is characterized using Auger electron spectroscopy (AES), X-ray photoelectron spectroscopy, and LEED.

## General Features

LEED experiments are usually performed in a UHV chamber that is maintained at pressures below  $10^{-10}$  mbar ( $\sim 10^{-8}$  Pa). The maintenance of a sufficiently low pressure is important to avoid residual gas adsorption and thus to keep the surface free of impurity during the measurement. In fact, a surface can be contaminated by a monolayer of gas in  $\sim 1$  s with a sticking coefficient of 1 with an ambient pressure of  $10^{-6}$  mbar. Thus, the timescale of the experiment may be estimated to be  $\sim 1000$  s before the surface is contaminated with one monolayer of the residual gas.

The quantitative measurement of diffracted intensities requires a planar and homogeneous surface of several  $\text{mm}^2$  in area and the sample must be orientated to within  $0.1^\circ$ . Furthermore, the earth magnetic fields must be balanced or shielded since it can deflect low-energy electrons. In addition, sample preparation requires heating to high temperatures to cleaning, whereas it is desirable to cool down the sample to liquid nitrogen temperature or below to reduce the thermal diffuse scattering in the case of intensity measurements.



### Experimental Setup

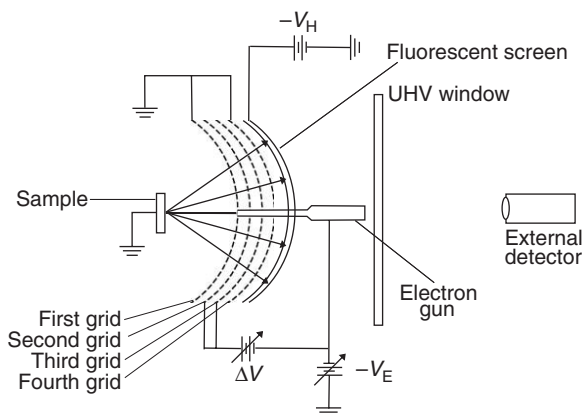
There are several experimental setups in use. The apparatus most commonly used to display the LEED pattern is a display setup where the diffracted electron beam is made visible on a fluorescent screen. Typically, the display setup consists of three or four hemispherical concentric grids that act as a high-pass filter for the inelastic scattered electrons. The four-grid systems can also be used for AES measurements to control the cleanliness of the surface. A schematic drawing of a back viewing arrangement of a four-grid system with an external detector is shown in **Figure 5**. This arrangement has the advantage that the grids and the sample holder do not hinder the observation of the diffraction pattern, and additional experimental tools near the crystal sample do not shadow the screen. The first grid is connected to earth ground to provide an essentially field-free region between the sample and the first grid, which minimizes the electrostatic deflection of diffracted electrons. A suitable potential ( $-V_E + \Delta V$ ) is applied to the second and third grids to retard inelastically scattered electrons. The fourth grid is grounded to reduce the field penetration of the second and third grids that can originate from the high potential ( $V_H$ , usually a few kilovolts) applied to the fluorescent screen. A typical diffraction pattern at normal incidence is shown in **Figure 6**.

The primary beam current is usually in the order of  $1\ \mu\text{A}$ , with an energy spread of  $\sim 0.5\ \text{eV}$ . This current density is rather high and materials that are sensitive to electron (for example, molecular overlayer structures) or insulator surfaces may suffer severe damage. For this purpose channel plates are used either in front of a fluorescent screen or in front of a position-sensitive detector, to intensify the diffracted beam. The use of two channel plates amplify the

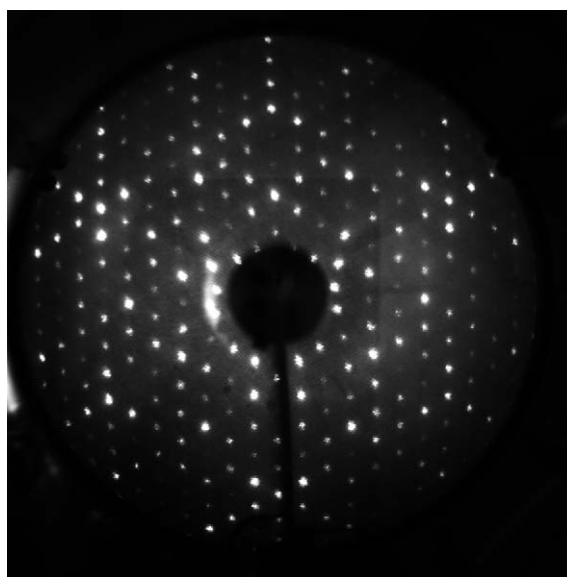
diffracted beam current by a factor of  $\sim 10^7$ , and thus allows the primary beam current to be in the order of  $0.1\text{--}1\ \text{nA}$ . The current of the diffracted beam is usually in the order of  $10^{-2}\text{--}10^{-4}$  of the primary beam and can be measured directly with a video camera from the fluorescent screen. By using the channel plate equipped system, the time required to measure the intensity of a diffracted beam at a particular energy is 20 ms, and the measurement of a complete set of  $I/V$  curves is less than 20 s.

### LEED Pattern

LEED pattern shows directly the size and orientation of the surface unit cell. However, the symmetry of the unit cell is not necessarily identical to the experimentally observed symmetry in LEED including spot intensities. The properties of symmetries such as rotation axes or mirror planes affect the intensities of the spots. The distinction between the structural symmetries of the surface and symmetries observed among spot intensities in LEED can be performed by changing the direction of the incident electron beam. At oblique incidence only mirror planes can be observed when the incident beam is parallel to the symmetry plane, otherwise the diffraction pattern exhibits no symmetry. In cases where the surface is not well ordered one may obtain additional symmetry information from the diffused LEED spots. The characteristic distribution of diffuse intensity in reciprocal space indicates the existence of short-range ordered antiphase domains or twin domains.



**Figure 5** Schematic diagram of a four-grid LEED display system.



**Figure 6** LEED pattern of the reconstructed  $\text{Si}(111)\text{--}(7\times 7)$  clean surface.



The analysis is analogous to that in diffuse X-ray diffraction.

### Domains

Rotational, translational, and mirrored domains are prevalent in the presence of adsorbate superstructure or adsorbate-induced reconstruction. A surface structure with symmetry lower than the symmetry of the substrate can produce such domains by applying all symmetry operations of the substrate. Since different domains with the same internal structure have the same structural energy, these domains should have an equal existence probability on a surface. If the areas of any domains presented on the surface are larger than the lateral coherence length of the incident electron beam, which is typically in the order of few tenths of a nanometer, the LEED pattern becomes the sum of the individual diffraction patterns of each domain. However, in cases where the symmetry of the surface is distorted by steps or by a strain, it is possible to eliminate certain domains selectively and one domain becomes dominant. **Figure 7A** shows an example of a well-oriented surface that contains three rotational domains, and **Figure 7B** displays a one-domain sample obtained using a stepped surface.

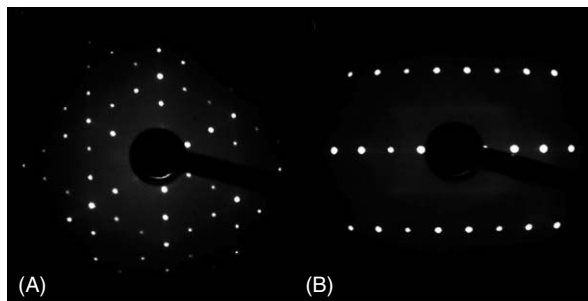
### Glide Plane Symmetry

Glide planes can be identified by the systematic absence of odd-order reflections in the glide plane. **Figure 8** shows the LEED pattern of a surface with a glide plane symmetry, in which spot extinctions are clearly observed along some lines. The absence of spots occurs when the direction of the incident electron beam is normal to the surface and when it is parallel to a glide plane of the surface. In some cases, there are invisible weak spots though there is no glide plane symmetry on the surface. However, this kind of missing spots and the extinctions due to glide lines can be distinguished by changing the direction of the incident electron beam because they have a different incident angle dependence, and by considering the

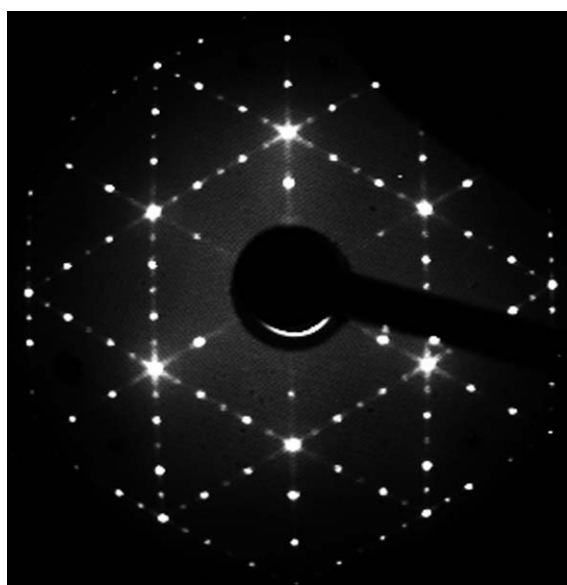
positions of the missing spots because they occur at different positions. In the case of a single domain surface, the direction of the lines of missing spots indicates directly the orientation of the glide lines. When different domains exist on the surface, even though one domain produces missing spots, other domains can give nonzero intensities at those positions. This problem can be solved by changing the sizes of domains that leads to an asymmetry in the spot intensities.

### Structural Analysis

Together with the determination of the symmetry and the periodicity of the surface structure, the atomic arrangement of the surface, the surface composition, and the thermal vibrations at surfaces can be determined by a quantitative analysis of the diffracted intensities. In most cases, the  $I/V$  curves are used for quantitative analysis. The small penetration depth of low-energy electrons restricts the applicability of the technique to  $\sim 1$  nm depth. Due to the large atomic scattering cross-section for low-energy electrons in a solid, the kinematic scattering theory does not describe the diffracted LEED intensities correctly, and thus a dynamical theory calculation is necessary for the quantitative description of diffraction intensities. In this context, the term dynamical is used synonymously for multiple scattering, in contrast to the kinematical theory that considers single scattering only. The dynamical theory of X-ray diffraction is, in principle, equivalent to the



**Figure 7** LEED patterns of (A) a three-domain  $\text{Si}(111)-(3 \times 1)$  surface and (B) a single-domain  $\text{Si}(111)-(3 \times 1)$  surface.



**Figure 8** LEED pattern of a three-domains  $\text{Si}(111)-(6 \times 1)$  surface.

dynamical theory of electron diffraction, but differs in the formulation of multiple scattering processes. This difference leads to the fact that it is not possible to invert LEED experimental data directly to obtain the atomic structure of the surface, like using the Patterson function in X-ray diffraction (the use of the Patterson function is based on the applicability of the kinematic theory). The structural analysis using LEED is conventionally done by comparing the measured intensities with the model calculations using a full dynamical theory. Tensor LEED is a perturbation method for calculation of the diffraction intensities in the vicinity of a reference structure, and has the advantage of a very fast calculation. The essential ingredients of the multiple scattering processes are the ion core scattering, the multiple scattering, the inelastic scattering, and the temperature effects.

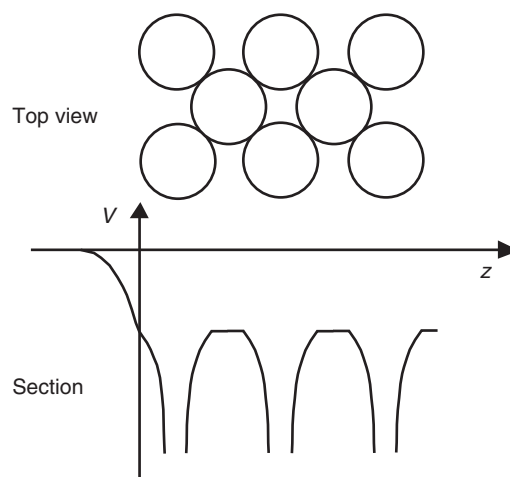
### Multiple Scattering

The calculation of the diffracted intensities usually proceeds in two steps. The first step is the construction of the crystal potential and the calculation of the scattering amplitudes from a single atom, and the second step is the calculation of scattering processes within a single atomic layer and the calculation of scattering between different atomic layers. In the second step the multiple scattering processes are based on the condition that the scattered wave from one atom is an incident wave on all other atoms. This leads to a set of linear equations that is solved by matrix inversion. The formulation of the theory is based on the KKR (Korringa-Kohn-Rostocker) method used for band structure calculations.

A muffin-tin model (Figure 9), in which spherically symmetric ion core scattering potentials are truncated at the radii at which they just touch and the interstitial region is assumed to be at a constant potential (muffin-tin constant), is often used for the first step. The crystal potential is obtained from a superposition of free atom electron densities that is spherically averaged inside the muffin-tin sphere. The muffin-tin constant cannot be calculated precisely enough and is treated as a variable parameter in the calculation. Its energy dependence is frequently included assuming that of the free electron gas, but an energy-independent value is in use as well. Common values are in the range of 5–15 eV.

The scattered intensity is conventionally written as

$$I(\mathbf{k}_0, \mathbf{k}_g) = \left| \sum_v F_v(\mathbf{k}_0, \mathbf{k}_g) \exp[2\pi i(\mathbf{k}_0, \mathbf{k}_g) \mathbf{r}_v] \right|^2 \quad [4]$$



**Figure 9** Top and sectional views of a muffin-tin potential,  $V$ , which is conventionally used in LEED multiple scattering calculations.

that is, using an equation that is formally equivalent to the sum over scattered amplitudes used in the kinematical theory. The sum over  $v$  runs over all atoms in the surface at position  $\mathbf{r}_v$ .  $F_v(\mathbf{k}_0, \mathbf{k}_g)$  is a generalized scattering amplitude that relates the scattered wave leaving atom  $v$  to the incoming wave without further scattering. The difference between the dynamical and kinematic calculations is that in the kinematic calculation the scattering amplitude is a function of the scattering vector  $\mathbf{q} = \mathbf{k}_0 - \mathbf{k}_g$  and depends on the kind of the atom, whereas the scattering amplitude is an explicit function of both vectors,  $\mathbf{k}_0$  and  $\mathbf{k}_g$ , in the dynamical calculation. The generalized scattering amplitude is a complex number and both the modulus and phases of the amplitudes  $F_v(\mathbf{k}_0, \mathbf{k}_g)$  vary strongly with the atomic arrangement in the neighborhood as a result of multiple scattering.  $F_v(\mathbf{k}_0, \mathbf{k}_g)$  can be calculated starting from the single scattering amplitude from a spherically symmetric atom. The scattering from a spherically symmetric potential is given by the solution of the Schrödinger equation that can be calculated numerically, and the scattering amplitudes are given by

$$f(\mathbf{k}, \mathbf{k}') = \frac{-2\pi i}{k} \sum_l (2l+1) [\exp(2i\delta_l) - 1] P_l(\cos \theta_{k,k'}) \quad [5]$$

where  $\theta_{k,k'}$  is the angle between the vectors  $\mathbf{k}$  and  $\mathbf{k}'$ . The quantities  $\delta_l$  which are used as input parameters, are the phase shifts of the spherical waves and describe the scattering properties of the atom. Usually, 6–10 phase shifts are included in the calculation. This number increases with the energy and atomic number of the atom.

### Inelastic Scattering

Inelastic scattering proceeds to the damping of an electron wave as

$$\Psi \sim \exp(-r/\lambda) \exp(ik_r r) \quad [6]$$

where  $k_r$  is the real part of a complex  $k$  that describes the damping. This equation indicates that the damping effect correlates to the mean free path  $\lambda$ . By regarding  $k$  as a complex energy, the relation between  $\lambda$  and the imaginary component of energy  $V_i$  is

$$\lambda \approx 0.39E^{1/2}/V_{0i} \quad [7]$$

where  $E$  and  $V_{0i}$  are expressed in electronvolts and  $\lambda$  in nanometers. In calculation, either  $\lambda$  or  $V_{0i}$  is usually kept constant for convenience.

### Thermal Effects

Thermal vibrations reduce scattering coherence and therefore reduce the intensities of diffracted electron beams. An atomic scattering factor is reduced by

$$\exp(-\frac{1}{2}\Delta k^2 \langle u^2 \rangle) = \exp(-M) \quad [8]$$

$\langle u^2 \rangle$  is the mean square vibrational amplitude measured in the direction of  $\Delta k$  and  $M$  is the Debye-Waller factor. The thermal effects affect the relative peak intensities, and contributes to reducing multiple scattering.

### Comparison with Experimental Data

A number of different  $R$ -factors are currently in use to quantify the agreement between experimental and theoretical  $I/V$  curves. There is no consensus as to which is the most appropriate. The most commonly used  $R$ -factor is  $R_p$  (the Pendry  $R$ -factor),  $R_1$  and  $R_2$  (the most common X-ray  $R$ -factors), and  $R_{ZJ}$  (the Zanazzi-Jona  $R$ -factor). The definition of Pendry  $R$ -factor is

$$R_p = \frac{\int (Y_e - Y_t)^2 dE}{\int (Y_e^2 + Y_t^2) dE} \quad [9]$$

with

$$Y = L/(1 + V_{0i}^2 L^2) \quad [10]$$

where  $L = l'/l$  is the logarithmic derivative of the  $I/V$  curve, and  $V_{0i}$  is the imaginary part of the electron self-energy. A value of  $R_p$  below 0.2 can be considered as good-to-excellent agreement, 0.3 indicates average agreement, and above 0.4 the agreement is marginal. With low-index beams and in the lower-energy range the level of agreement is in general better than for high-index beams and high energies.

$R_1$  and  $R_2$  are defined as

$$R_1 = \frac{\int (l_e - c_l) dE}{\int l_e dE} \quad [11]$$

$$R_2 = \frac{\int (l_e - c_l)^2 dE}{\int l_e^2 dE} \quad [12]$$

with

$$c = \frac{\int l_e dE}{\int l_t dE} \quad [13]$$

Both can be defined as integrals over continuous curves as well as sums over discrete points. In the latter case  $R_1$  has been called  $R_{DE}$ .

$R_{ZJ}$  is defined empirically by a more complicated expression using the second derivative of the  $I/V$  curves. The second derivative is used from the fact that the maxima are frequently not clearly separated. A disadvantage of using  $R_{ZJ}$  is that its absolute value depends on the energy range. It is therefore less frequently used. To avoid local minima in the  $R$ -factor curves it has been found that an average of different  $R$ -factors can be used to provide a more pronounced minimum.

### Accuracy of the Structure Determination

The error bars of the structure analysis are difficult to estimate because of the unknown influence of the approximations used in the LEED theory. Together with the main unknown factors, the nonspherical ion core potentials, anisotropy and energy dependence of inelastic scattering, and anisotropy and correlation effects of the thermal vibrations, there are further uncertainties due to systematic errors in the experimental data, misalignment of the surface orientation, and imperfections in the surface structure. The precision of the analysis is frequently estimated from the sensitivity of the  $R$ -factors to a variation of structural parameters assuming statistical errors. The standard deviation of the structural parameters is calculated from the number of beams and the energy range, assuming an average density of peaks in the  $I/V$  curves. For the back-scattering geometry, the highest precision is obtained for the outermost surface atoms and the precision becomes smaller for atoms in deeper layers. The estimated error is usually 0.001–0.003 nm for the layer distances and 0.005–0.01 nm for the parallel components of the atomic positions in the top two layers.

### Application of LEED

The relatively simple LEED experiment and the fact that LEED systems are commercially available have

led to a broad application in surface science. One of the main advantages of the LEED method is that the whole diffraction pattern is visible on a screen. A qualitative interpretation of the diffraction pattern is the most common use of LEED. It allows the surface unit cell to be identified, the state of order to be estimated, and surface phases in adsorption systems to be identified. The observations of 1D and 2D phase transitions are also performed using LEED. Quantitative structure analyses have mainly been performed for metal surface and adsorbate layers on metals. Surface defect distributions and structure of disordered adsorbate layers are performed by analyzing the angular beam profiles in the diffuse background. In the study of defect distributions the kinematic theory of diffraction can be applied as long as only the shapes of the profiles are evaluated and not the intensities. Multiple scattering effects are short-ranged, and the mean free path is of the order of several interatomic distances, so that for sufficiently widely spaced defects multiple scattering effects can be neglected. The main application of the study of defects has been the determination of step distributions in homo- and heteroepitaxial systems.

#### **Diffuse LEED (DLEED)**

The large cross-section of low-energy electrons makes it possible to measure the diffuse elastic background intensity arising from point defects. There are two methods to measure the background intensity. The first method is measurement of the spatial distribution of background intensity at different energies, and the second method is measurement of the energy dependence of the background intensity at certain points in the 2D reciprocal unit cell. The spatial distribution of background intensity can be measured using a fluorescent screen or a channel plate. After subtracting the diffuse scattering of the clean surface, the intensity distribution arising from point defects, which is formed by, for example, the adsorption of single atoms or molecules, can be analyzed quantitatively. To eliminate the effect of correlated defects that lead to maxima in the intensity distribution, the Y-function shown in eqn [10] is analyzed. The main uncertainty in this method is the subtraction of the thermal diffuse background. For the calculation of the energy dependence of the background intensity, conventional LEED programs can be used by modeling a defect site with a single atom or molecule in a large unit cell.

#### **Very-Low-Energy Electron Diffraction**

In the energy range between 0 and 10 eV, only the specularly reflected beam can be observed if the

surface unit cell is not too large. Very-low-energy electron diffraction (VLEED) is restricted to the measurement and analysis of the reflectivity. The advantages of the method are that (1) the intensity of the reflected beam in this energy range is high and (2) the experiment is relatively simple to perform. Because the reflectivity depends sensitively on the shape of the potential barrier across the surface, the method is used to study the potential barrier. A further advantage is the sensitivity to light elements such as hydrogen. A well-ordered superstructure is not required. The calculation of the reflectivity in this energy range is not very time-consuming because few beams and few phase shifts are required in the multiple scattering calculations. Special care has to be taken, however, to take account of the small optical potential at energies below the plasmon threshold. This makes conventional programs inapplicable for VLEED. Applications of the method include the study of surface resonances. VLEED has also been applied to locate hydrogen atoms at surfaces, to study disordered surfaces, and to eliminate models for which the full dynamical calculation at higher energies would be too tedious. The disadvantage of the method is the small database, which does not allow the study of complex structures.

### **Comparison with Other Related Surface Structure Analysis Techniques**

A number of techniques are used to determine structures of surfaces. These are (1) microscopic methods: scanning tunneling microscopy, atomic force microscopy; (b) spectroscopic methods: photoelectron diffraction, surface X-ray diffraction (XRD), extended X-ray absorption fine structure, and related techniques; and (3) ion scattering techniques: impact collision ion scattering spectroscopy, low-energy ion scattering, medium-energy ion scattering, and Rutherford back scattering. The microscopic methods have an advantage to obtain local surface structure in real space, but the information is limited to the arrangements of the outermost surface atoms. The technique that is most comparable to LEED is XRD. The main difference between LEED and XRD lies in the evaluation of the data. In the case of XRD, the kinematical diffraction theory is valid, which allows the application of structure analysis and refinement techniques developed for 3D structures, like the Patterson function, difference Fourier analysis, and the application of direct methods. XRD allows the determination of complicated surface structures that are not yet accessible with LEED. XRD is not limited to UHV conditions and in most cases

photon-induced damage is negligible. In practice, XRD requires intense sources. Mostly synchrotron radiation, but rotating anodes are also used for strongly scattering elements. The XRD experiment is more difficult to carry out as long measuring times are required and precision goniometers are necessary in UHV or precise movement of UHV chambers. With XRD, the full diffraction pattern cannot be observed. Most X-ray chambers as well as other surface structure instruments are therefore equipped with LEED screens to monitor the crystalline state of the surface. The resolution normal to the surface is not as high for XRD as for LEED as the momentum transfer normal to the surface at grazing incidence is limited to  $|\mathbf{k}|$ , while it is  $2|\mathbf{k}|$  in the case of LEED.

*See also:* **Microscopy Techniques:** Atomic Force and Scanning Tunneling Microscopy. **Surface Analysis:** X-Ray Photoelectron Spectroscopy; Auger Electron Spectroscopy; Ion Scattering. **X-Ray Absorption and Diffraction:** X-Ray Absorption. **X-Ray Fluorescence and Emission:** X-Ray Diffraction – Single Crystal.

## Further Reading

- Heinz K (1988) Structural analysis of surfaces by LEED. *Progress in Surface Science* 27: 239–326.
- Heinz K and Muller K (1982) *LEED Intensities—Experimental Progress and New Possibilities of Surface Structure Determination in Structural Studies of Surfaces*, Springer Tracts in Modern Physics, vol. 91. Berlin: Springer-Verlag.
- Marcus PM and Jona F (1984) *Determination of Surface Structure by LEED*. New York: Plenum Press.
- Pendry JB (1974) *Low Energy Electron Diffraction*. New York: Academic Press.
- Rous PF (1992) The tensor LEED approximation and surface crystallography by low energy electron diffraction. *Progress in Surface Science* 39: 3–63.
- Van Hove MA and Tong SY (1979) *Surface Crystallography by LEED*. Berlin: Springer-Verlag.
- Van Hove MA, Weinberg WH, and Chan CM (1988) *Low Energy Electron Diffraction*. Berlin: Springer-Verlag.
- Woodruff DP and Delchar TA (1994) *Modern Techniques of Surface Science*, 2nd edn. Cambridge: Cambridge University Press.

## Infrared Spectroscopy

**H Hoffmann and T Leitner**, Technical University of Vienna, Vienna, Austria

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

With the advent of sensitive Fourier transform infrared (FTIR) spectrometers, surface infrared spectroscopy has developed into one of the major surface analytical techniques. Highly specific information on molecular composition and structure, nondestructive sampling, and a fairly inexpensive instrumentation usable under ambient (nonvacuum) conditions are distinct advantages of IR spectroscopy over other surface science methods. Today, various surface-specific IR sampling techniques are available, which allow a wide variety of sample surfaces and adsorbate systems to be studied in the monolayer and sub-monolayer regime. They can be classified into reflection and nonreflection techniques according to the light path through the sample (**Figure 1**). The main reflection techniques are reflection absorption infrared spectroscopy (RAIRS), internal reflection or attenuated total reflectance (ATR), and diffuse reflectance infrared Fourier transform (DRIFT). Transmission measurements (TIR), emission

spectroscopy (EMS), and photoacoustic spectroscopy (PAS) are the most important nonreflective surface techniques. The optical properties of the sample (reflectivity, absorptivity) and its physical condition (powder, film, single crystal, etc.) will decide on which method to choose. RAIRS and ATR are typically used for the so-called low surface area samples (thin films or adsorbates on flat smooth surfaces) while powdered samples with a high surface area can be studied with DRIFT or transmission measurements of pressed pellets. EMS and PAS are the most versatile techniques regarding the sample properties and can be used for samples of any shape. Most of these methods are discussed in separate chapters of this Encyclopedia and this article will therefore focus on RAIRS. After a brief treatment of the theoretical background, instrumentation and signal processing techniques will be discussed. The main body of the text will then deal with some representative applications for surface analysis.

### Theory

A simplified model of three isotropic phases (**Figure 2**) shall be used to describe the light absorption process in a reflection-absorption experiment. The

photon-induced damage is negligible. In practice, XRD requires intense sources. Mostly synchrotron radiation, but rotating anodes are also used for strongly scattering elements. The XRD experiment is more difficult to carry out as long measuring times are required and precision goniometers are necessary in UHV or precise movement of UHV chambers. With XRD, the full diffraction pattern cannot be observed. Most X-ray chambers as well as other surface structure instruments are therefore equipped with LEED screens to monitor the crystalline state of the surface. The resolution normal to the surface is not as high for XRD as for LEED as the momentum transfer normal to the surface at grazing incidence is limited to  $|k|$ , while it is  $2|k|$  in the case of LEED.

*See also:* **Microscopy Techniques:** Atomic Force and Scanning Tunneling Microscopy. **Surface Analysis:** X-Ray Photoelectron Spectroscopy; Auger Electron Spectroscopy; Ion Scattering. **X-Ray Absorption and Diffraction:** X-Ray Absorption. **X-Ray Fluorescence and Emission:** X-Ray Diffraction – Single Crystal.

## Further Reading

- Heinz K (1988) Structural analysis of surfaces by LEED. *Progress in Surface Science* 27: 239–326.
- Heinz K and Muller K (1982) *LEED Intensities—Experimental Progress and New Possibilities of Surface Structure Determination in Structural Studies of Surfaces*, Springer Tracts in Modern Physics, vol. 91. Berlin: Springer-Verlag.
- Marcus PM and Jona F (1984) *Determination of Surface Structure by LEED*. New York: Plenum Press.
- Pendry JB (1974) *Low Energy Electron Diffraction*. New York: Academic Press.
- Rous PF (1992) The tensor LEED approximation and surface crystallography by low energy electron diffraction. *Progress in Surface Science* 39: 3–63.
- Van Hove MA and Tong SY (1979) *Surface Crystallography by LEED*. Berlin: Springer-Verlag.
- Van Hove MA, Weinberg WH, and Chan CM (1988) *Low Energy Electron Diffraction*. Berlin: Springer-Verlag.
- Woodruff DP and Delchar TA (1994) *Modern Techniques of Surface Science*, 2nd edn. Cambridge: Cambridge University Press.

## Infrared Spectroscopy

**H Hoffmann and T Leitner**, Technical University of Vienna, Vienna, Austria

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

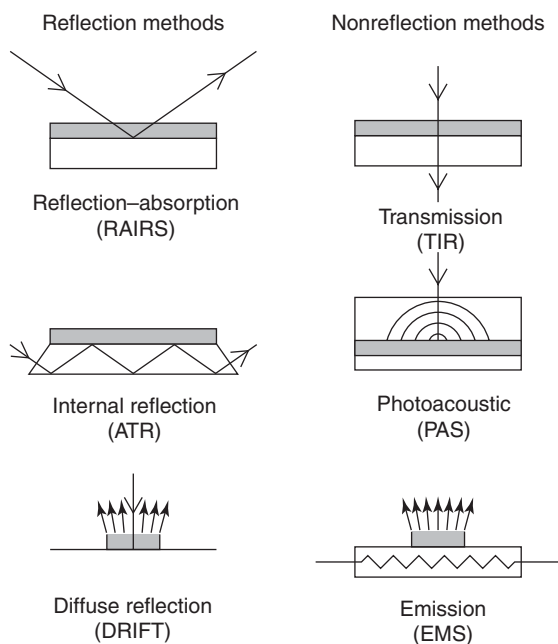
With the advent of sensitive Fourier transform infrared (FTIR) spectrometers, surface infrared spectroscopy has developed into one of the major surface analytical techniques. Highly specific information on molecular composition and structure, nondestructive sampling, and a fairly inexpensive instrumentation usable under ambient (nonvacuum) conditions are distinct advantages of IR spectroscopy over other surface science methods. Today, various surface-specific IR sampling techniques are available, which allow a wide variety of sample surfaces and adsorbate systems to be studied in the monolayer and sub-monolayer regime. They can be classified into reflection and nonreflection techniques according to the light path through the sample (**Figure 1**). The main reflection techniques are reflection absorption infrared spectroscopy (RAIRS), internal reflection or attenuated total reflectance (ATR), and diffuse reflectance infrared Fourier transform (DRIFT). Transmission measurements (TIR), emission

spectroscopy (EMS), and photoacoustic spectroscopy (PAS) are the most important nonreflective surface techniques. The optical properties of the sample (reflectivity, absorptivity) and its physical condition (powder, film, single crystal, etc.) will decide on which method to choose. RAIRS and ATR are typically used for the so-called low surface area samples (thin films or adsorbates on flat smooth surfaces) while powdered samples with a high surface area can be studied with DRIFT or transmission measurements of pressed pellets. EMS and PAS are the most versatile techniques regarding the sample properties and can be used for samples of any shape. Most of these methods are discussed in separate chapters of this Encyclopedia and this article will therefore focus on RAIRS. After a brief treatment of the theoretical background, instrumentation and signal processing techniques will be discussed. The main body of the text will then deal with some representative applications for surface analysis.

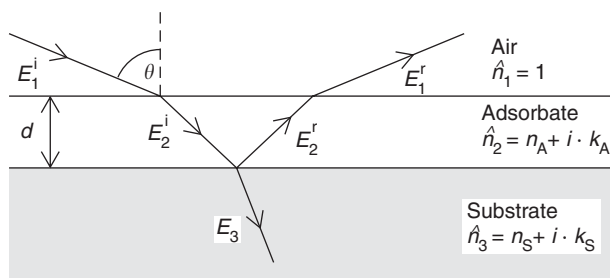
### Theory

A simplified model of three isotropic phases (**Figure 2**) shall be used to describe the light absorption process in a reflection-absorption experiment. The





**Figure 1** Sample illumination in different surface sensitive infrared techniques.



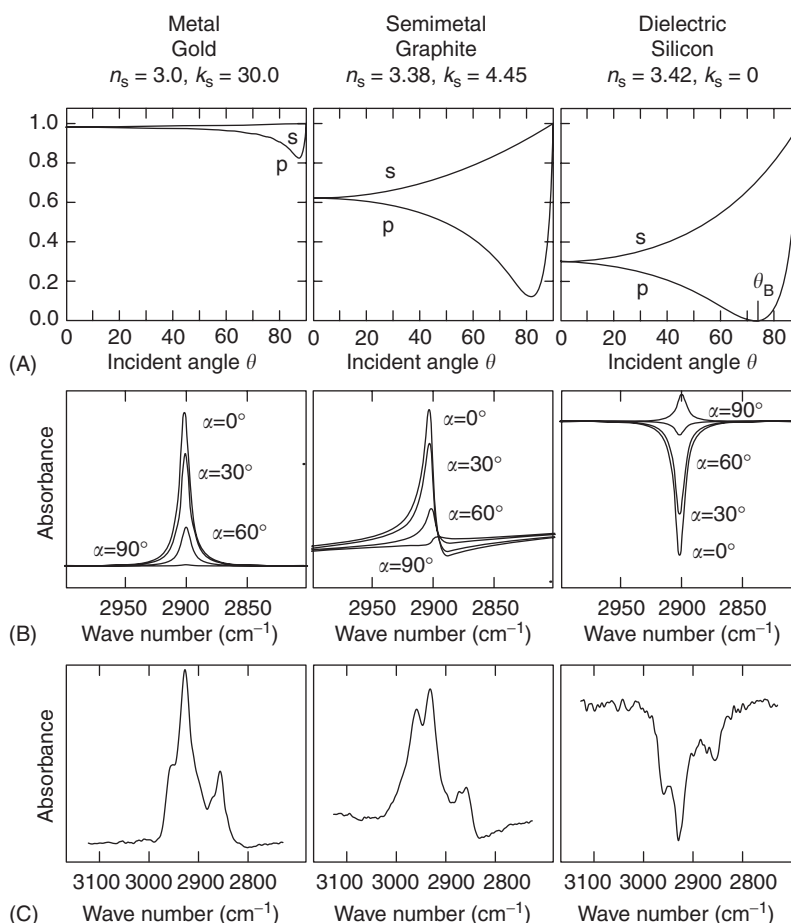
**Figure 2** Optical sample model for reflection absorption infrared spectroscopy.

optical properties of each phase are described by the complex refractive index  $\hat{n} = n + ik$  ( $n$  is the refractive index and  $k$  the absorption index). Phase 1 – the incident medium – is assumed to be air ( $n = 1, k = 0$ ), phase 2 is an adsorbate layer of thickness  $d$ , and phase 3 is the semi-infinite substrate phase. Infrared radiation (incident electric field  $E_1^i$ ) impinges on the surface at an angle  $\theta$  to the surface normal, bisects the adsorbate layer, and is reflected in the specular direction. The ratio  $\Delta R/R_0 = (R_0 - R_d)/R_0$  as a function of frequency, where  $R_d$  and  $R_0$  are the reflectivities of the adsorbate-covered and the bare substrate surface, respectively, and gives the reflection-absorption spectrum of the adsorbate. The theoretical treatment is based on calculations of the amplitude and phase changes of the electric field  $E$  of the incident radiation due to partial reflection and transmission at each interface and absorption by the

adsorbate. The calculations are carried out separately for light polarized perpendicular (s-polarized, where the s comes from the German word ‘senkrecht’, meaning perpendicular) or parallel (p-polarized) to the plane of incidence defined by the central incoming ray and the surface normal. The absorption intensities  $\Delta R = R_0 - R_d$  are determined by the so-called surface electric field, which is the vectorial sum of the incident and reflected electric field vectors  $E_2^i$  and  $E_2^r$  in the adsorbate layer, while the substrate reflectivity  $R_0$  is determined by the ratio  $E_1^r/E_1^i$  of the reflected to the incident electric field vectors for the bare substrate surface. Basic electromagnetic theory can be used to derive analytical expressions for the substrate reflectivity  $R_0$  and the absorption intensities  $\Delta R/R_0$  of an adsorbate as a function of the optical constants, the incidence angle, the layer thickness, and the frequency. In **Figure 3**, calculated and experimental RAIRS spectra of a hydrocarbon adsorbate film are compared on three different types of substrates. The main parameter, which determines the reflection properties of the substrate itself as well as the intensity and shape of an adsorbate vibrational absorption, is the absorption index  $k_s$  of the substrate, which can range from  $k_s = 0$  for nonabsorbing, dielectric substrates to  $k_s = 30$  for metal substrates. Thus, a strongly absorbing metal substrate ( $k_s > 10$ , e.g., gold) is compared in **Figure 3** with a weakly absorbing semimetal ( $1 < k_s < 10$ , e.g., graphite) and a nonabsorbing dielectric substrate ( $k_s = 0$ , e.g., silicon). **Figure 3A** first shows the reflectivities of the pure substrate as a function of the light incidence angle for s-polarized and p-polarized radiation, **Figure 3B** shows a calculated model absorption for p-polarized light at grazing incidence ( $80^\circ$ ) with different dipole moment tilt angles  $\alpha$  and **Figure 3C** shows experimental RAIRS spectra of thin, physisorbed hydrocarbon films on gold, graphite, and silicon. Without going into details, the following characteristics of RAIRS on different substrates are illustrated in **Figure 3**.

### Metal Substrate

The reflectivities for both s- and p-polarized light are  $\sim 1$  for all angles of incidence (**Figure 3A**). An adsorbate probed with p-polarized light yields symmetric, undistorted absorptions (**Figure 3B**), whose intensity is negligible for parallel dipole moment orientation ( $\alpha = 90^\circ$ ) and reaches a maximum for perpendicular orientation ( $\alpha = 0^\circ$ ). Thus, only the perpendicular component of a vibration, which is proportional to  $\cos \alpha$ , gives rise to absorption, because the electric field parallel to the surface is vanishingly small. The overall intensity is therefore



**Figure 3** Optical properties of three different substrate materials – a metal, a semimetal, and a dielectric – and their influence on IR reflection spectra of thin adsorbate films. (A) Substrate reflectivities of p-polarized (p) and s-polarized (s) radiation as a function of the light incidence angle  $\theta$ . (B) Simulated IR absorption of an adsorbate for different tilt angles  $\alpha$  of the vibrational transition dipole moment, probed with p-polarized light at 80° incidence. (C) Experimental CH stretching absorptions of a thin film of paraffin oil ( $d = 30$  nm), measured with p-polarized light at 80° incidence.  $n_s$  and  $k_s$  denote the refractive index and the absorption index of the substrate, respectively.

proportional to  $\cos^2 \alpha$ . This is the origin of the so-called ‘metal surface selection rule’, which can be used to determine surface orientations of molecules. A physisorbed adsorbate film with isotropic surface orientation yields a transmission-like, undistorted spectrum in a reflection absorption experiment (Figure 3C).

### Semimetal Substrate

With decreasing absorption index  $k_s$  of the substrate, the reflectivities for both s- and p-polarized light start to decrease and to diverge with  $R_p$  being always lower than  $R_s$  and going through a minimum  $\sim 80^\circ$  for the particular substrate (graphite) in Figure 3A. The simulated absorptions on graphite in Figure 3B are still quite similar to the metal substrate and yield a maximum peak intensity at  $\alpha = 0^\circ$  and only

minor absorption at  $\alpha = 90^\circ$ . The peakshape, however, shows a distinct asymmetry with a dip on the low-frequency side and a tail on the high-frequency side of the absorption, which is also visible in the experimental reflection spectra of Figure 3C. This asymmetry indicates the onset of the band inversion observed on dielectric substrates, which is caused by the adsorbate’s refractive index  $n$  dominating over the absorption index  $k$ .

### Dielectric Substrate

The reflectivity at normal incidence ( $\theta = 0^\circ$ ) is typically less than 30% (Figure 3A) and the p-polarized reflectivity goes to zero at the substrate’s Brewster angle  $\theta_B$  ( $\theta_B = \arctan n_s$ ). Perpendicular and parallel vibrational components give rise to absorptions of opposite ‘sign’, i.e., the corresponding peaks point in

opposite directions in the reflection spectrum. For incidence angles  $\theta > \theta_B$  as chosen in **Figure 3B** ( $\theta = 80^\circ$ ), the absorption points in the ‘negative’, downward direction for perpendicular dipole moment orientation ( $\alpha = 0^\circ$ ) and in the upward, regular direction for parallel orientation ( $\alpha = 90^\circ$ ). Consequently, the absorption vanishes for a certain tilt angle  $\alpha$ , where the perpendicular and parallel components cancel each other (e.g., at  $\alpha = 67^\circ$  for the example chosen in **Figure 3B**). A disordered, liquid-like film, where each dipole moment adopts an average orientational angle of  $\alpha = 54.7^\circ$  relative to the surface normal, therefore, yields an inverted absorption spectrum as shown in **Figure 3C**, where the band profiles are again undistorted as on metal substrates, but point in the opposite, downward direction.

## Instrumentation

### Optical Configuration

A typical setup for a reflection-absorption experiment is shown in **Figure 4**. IR radiation is passed from the spectrometer through a polarizer onto an off-axis paraboloidal mirror, which focuses the beam at an incident angle  $\theta$  onto the sample surface. A focal length  $> 15$  cm is advisable in order to keep the cone angle of the focused radiation at the sample surface small. From there the beam is reflected in the specular direction, is recollimated by a second off-axis paraboloidal mirror, and is focused onto the detector. Since a monolayer of an adsorbate on a smooth substrate is equal to only  $\sim 10^{14}$  mol cm $^{-2}$  surface area, the total amount of sample material probed by the light is very low (typically  $10^{-9}$  mol or less) and sensitivity is the major concern. The main

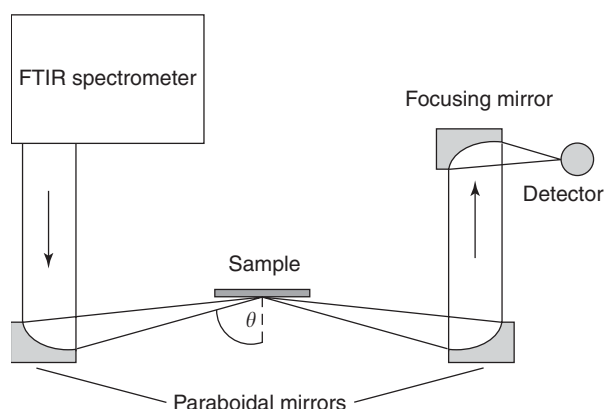
instrumental parameters and their influence on the signal-to-noise ratio (SNR) is briefly discussed below.

**Light source** Apart from the traditional blackbody sources for the mid-IR range ( $4000\text{--}400$  cm $^{-1}$ ) such as global, Nernst glower, ceramic sources, alternative brighter sources such as lasers, synchrotron radiation, or free electron lasers (FELs) are becoming available, which provide significant advantages for investigations of surfaces and interfaces. They provide highly focused, polarized radiation, which can be applied either in a continuous or in a pulsed mode with pulse durations from nanoseconds (synchrotron radiation) to femtoseconds (lasers, FELs). Despite some drawbacks such as the limited wave-number range of tuneable lasers, the high expense, and the usually complex instrumentation these novel sources open the access to a range of new experiments including time-resolved studies on surfaces (e.g., relaxation processes of excited vibrational and electronic states), surface imaging (e.g., near-field IR microscopy), and studies of low-energy vibrations (substrate-adsorbate binding vibrations, crystal lattice vibrations, etc.) in the far-IR region.

**Incidence angle** The SNR in a RAIRS spectrum is given by the absorption intensity  $\Delta R/R_0$  divided by the noise level, which in the mid-IR region ( $4000\text{--}400$  cm $^{-1}$ ) is inversely proportional to the reflectivity  $R_0$  of the substrate:

$$\text{SNR} \approx \frac{\Delta R/R_0}{1/R_0} \approx \Delta R$$

The optimum sensitivity is therefore achieved where the absolute reflectivity change  $\Delta R$  rather than the relative peak intensity  $\Delta R/R_0$  is largest. Since  $\Delta R$  depends strongly on the type of substrate and on the incidence angle, an optimum angle  $\theta_{\text{opt}}$  exists for each substrate, where the best SNR is achieved. **Table 1** lists calculated  $\theta_{\text{opt}}$  values and corresponding relative SNRs for a number of common substrate materials. Whereas the optimum angle for p-polarized light lies between  $70^\circ$  and  $87^\circ$  for most substrates, the predicted SNR values differ largely and are generally 1–2 orders of magnitude lower than the



**Figure 4** Instrumental setup for reflection absorption infrared measurements.

**Table 1** Optimum incidence angles and relative signal-to-noise ratios (SNR) for external reflection infrared spectroscopy with different substrate materials using p-polarized light

Substrate	$\theta_{\text{opt}} (^\circ)$	SNR
Gold	87	100
Graphite	72	6
Silicon	86	2
ZnSe	84	0.6

SNR on a metal substrate. However, the use of a highly sensitive IR detector can compensate for most of these inherent sensitivity differences, as will be discussed below, and usually allows high-quality spectra in the monolayer coverage regime to be obtained on almost any type of substrate material.

**Detector** RAIRS is a typical low-light application (high throughput losses because of limited sample size and/or low substrate reflectivity) and therefore the most sensitive detectors will generally give the best results. The specific detectivity  $D^*$  as a measure of detector sensitivity ( $D^* = A^{1/2}/\text{NEP}$ , where  $A$  is the detector area ( $\text{cm}^2$ ) and NEP the noise equivalent power ( $\text{W Hz}^{-1/2}$ )) and the desired wavenumber range will be the main criteria for choosing the optimum detector. The most commonly used detectors are liquid nitrogen cooled semiconductors of the photoconductive type such as narrow-band mercury-cadmium-telluride ( $D^* \approx 5 \times 10^{10} \text{ cm Hz}^{1/2} \text{ W}^{-1}$ ,  $\Delta\nu = 5000\text{--}800 \text{ cm}^{-1}$ ) or of the photovoltaic type such as indium antimonide ( $D^* \approx 2 \times 10^{11}$ ,  $\Delta\nu = 10\,000\text{--}2000 \text{ cm}^{-1}$ ). In the far-IR region liquid helium cooled doped germanium or silicon bolometers ( $D^* \approx 5 \times 10^{10}$ ,  $\Delta\nu = 5000\text{--}200 \text{ cm}^{-1}$ ) are often used.

### Signal Processing and Sensitivity Optimization

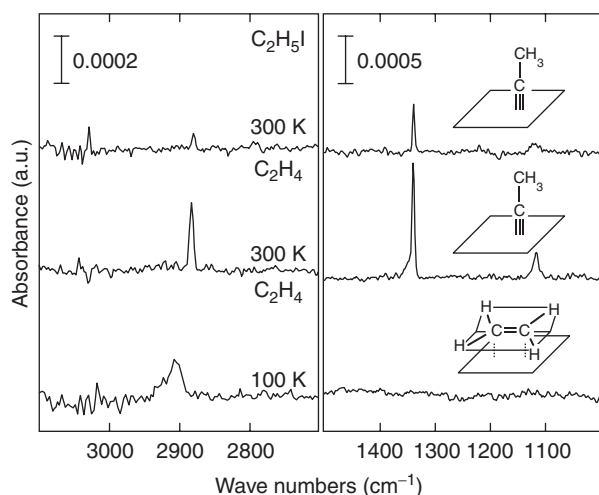
In the conventional single beam operation a spectrum of the adsorbate-covered substrate is ratioed against a spectrum of the clean substrate to yield the reflection absorption spectrum of the adsorbate. For weakly reflecting nonmetallic substrates and/or small sample sizes only a small fraction of the source intensity reaches the detector and the SNR will be limited by detector noise. The use of a more sensitive detector and longer signal averaging are the only ways to improve the sensitivity. With highly reflecting substrates such as metals, however, the signal intensity at the detector often exceeds the dynamic range of the analog-to-digital converter and the SNR of a single-beam experiment will be limited by digitization noise. Several ways of circumventing this limitation have been proposed, aiming at a suppression of the centerburst in the interferogram by measuring only a difference signal proportional to the amount of light absorbed by the sample. One approach classified as dual-beam spectrometry is based on optical subtraction of the transmitted and reflected beams in the interferometer, which are  $180^\circ$  out of phase and whose modulated components cancel upon recombination. This can be realized experimentally by using either two input beams from two light sources or two output beams that are recombined

after passing through the interferometer. If the sample is placed in one of the two beams, only a difference signal proportional to the light absorbed by the sample is detected. This small signal can then be amplified until detector noise appears in the interferogram. The amplification factor corresponds directly to the SNR improvement over a single-beam experiment. An alternative method of reducing the dynamic range of the signal is polarization-modulation (PM-RAIRS), which is based on the same principle of measuring a difference signal from a sample that discriminates between two different states of polarization. A conventional single-beam optical system can be used together with a photoelastic modulator, which changes the polarization of the incident radiation from parallel to perpendicular at frequencies  $\sim 100 \text{ kHz}$ . A lock-in amplifier connected to the detector output and tuned to the polarization modulation frequency is used to obtain the difference signal between s- and p-polarized radiations.

## Applications

### Single Crystal Surfaces

RAIRS is widely used in combination with other surface science techniques to study adsorption and surface reactions on single crystal surfaces in an ultrahigh vacuum (UHV) environment. The vast majority of these studies have been carried out with strongly absorbing model compounds such as CO or NO and detection limits below 1% of a monolayer coverage have been achieved on metal surfaces. Recently, the focus of these investigations has shifted to practically more relevant systems owing to the increased sensitivity of modern FTIR instruments. The surface reactivity of simple hydrocarbon species, for example, has been monitored IR-spectroscopically on various catalytically important metal substrates. Molecular orientations of the adsorbed species have been determined based on the surface selection rules and related to their surface reactivity. **Figure 5** shows an example of the thermal decomposition of ethylene and ethyl iodide on a Pt(111) surface. Adsorption of ethylene at 100 K results in molecular chemisorption with the C–C bond axis parallel to the surface. Only a weak CH stretching band at  $2900 \text{ cm}^{-1}$  can be detected in the IR spectrum. At 300 K ethylene decomposes to form ethynylidyne ( $\text{CH}_3\text{C}\equiv$ ). Three absorptions at 2883, 1339, and  $1118 \text{ cm}^{-1}$  are observed and assigned to  $\nu_s(\text{CH}_3)$ ,  $\delta_s(\text{CH}_3)$ , and  $\nu(\text{CC})$ , all of which have a transition dipole moment parallel to the C–C bond. Since no other absorptions with different dipole moment orientations are detected, the C–C axis of ethynylidyne must stand



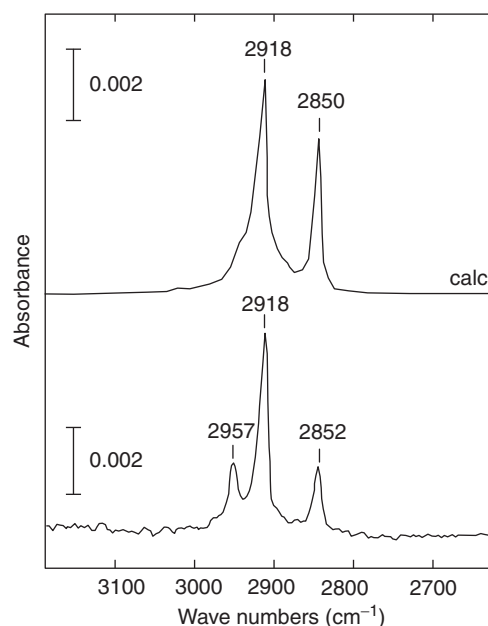
**Figure 5** RAIRS spectra of ethylene ( $\text{C}_2\text{H}_4$ ) chemisorbed at 100 K, ethylene chemisorbed at 300 K, and ethyl iodide ( $\text{C}_2\text{H}_5\text{I}$ ) chemisorbed at 300 K on a Pt(111) single crystal surface. (Reprinted with permission from Hoffmann H, Griffiths PR, and Zaera F (1992) A RAIRS study on the surface chemistry of ethyl iodide on Pt(111). *Surface Science* 262: 141–150; © Elsevier.)

perpendicular to the surface. Ethylidyne is also produced in the same adsorption geometry but with a lower surface concentration by adsorbing ethyl iodide at 300 K. This latter decomposition can be followed in the IR spectra at lower temperatures (100–300 K) and proceeds via the formation of ethyl radicals as stable intermediates on the surface. Alkyl radicals are known to be key intermediates in various industrial catalytic hydrocarbon reactions. They can be identified by RAIRS among other hydrocarbon fragments owing to the high resolution of IR spectroscopy.

A severe limitation for practical applications is the large pressure difference between single crystal experiments under UHV conditions and the real reaction conditions of heterogeneous catalysis and other surface processes. Most surface science techniques are restricted to pressures below  $10^{-6}$  Torr because they require hot filaments and high mean free paths for the probing particles. IR spectroscopy as an optical technique can be used in normal atmosphere as well as with liquid surroundings and is therefore ideally suited for *in situ* studies of surface processes under real reaction conditions.

### Thin Film Studies

Ordered monolayers of organic compounds, which can be prepared using the Langmuir-Blodgett or molecular self-assembly technique, have received growing attention both as model systems for synthetic organic interfaces and as technological products such



**Figure 6** Calculated (above) and experimental (below) RAIRS spectra for a methyl-16-mercaptohexadecanoate ( $\text{HS}(\text{CH}_2)_{15}\text{COOCH}_3$ ) monolayer on a gold surface. (Reprinted with permission from Nuzzo RG, Dubois LH, and Allara DL (1990) Fundamental studies of microscopic wetting on organic surfaces. *Journal of the American Chemical Society* 112: 558–569; © American Chemical Society.)

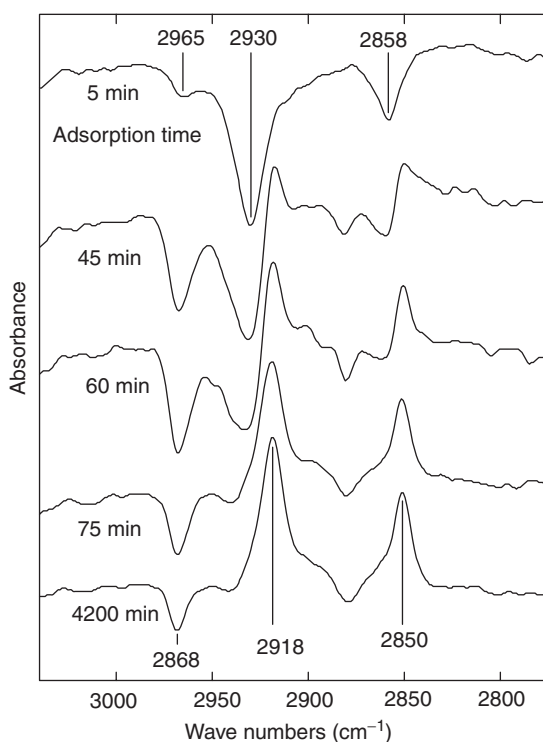
as nonlinear optical devices, chemically modified electrodes, lubrication layers, or synthetic membranes. Most of these applications are based on the densely packed, highly ordered structure of these monolayers, resulting from van der Waals interactions between the film molecule's hydrocarbon chains. Due to its nondestructive character RAIRS is one of the few surface analytical techniques that can provide quantitative information on the chemical composition and structure of these films. **Figure 6** shows an example of experimental and calculated spectra for a monolayer of  $\omega$ -mercapto-methylhexadecanoate ( $\text{HS}-(\text{CH}_2)_{15}-\text{COOCH}_3$ ) adsorbed on a gold substrate. The film molecules are bonded to the surface through their thiol end-groups and form a well-ordered monolayer film, whose interface to the ambient consists of the terminal ester groups. The sharp and intense bands in the IR spectrum reflect the densely packed, crystalline-like state of the adsorbate. The orientation of the film molecules on the surface can be obtained by comparing the experimental spectrum to a calculated spectrum of a monolayer film with isotropic molecule orientation. According to the surface selection rule the ratio  $I_{\text{obs}}/I_{\text{calc}}$  is proportional to  $\cos^2 \alpha$  for each vibrational mode, where  $I_{\text{obs}}$  and  $I_{\text{calc}}$  are the observed and calculated peak intensities, and  $\alpha$  is the angle

between the surface normal and the transition dipole moment of the vibration. For the example shown in **Figure 6**, a  $34^\circ$  tilt of the molecule axis toward the surface normal and a  $55^\circ$  twist of the C-atom plane from a configuration perpendicular to the surface has been derived from the  $\nu(\text{CH}_2)$  absorptions at  $2918$  and  $2850\text{ cm}^{-1}$ .

Recently, studies of monolayer film structure and composition have also been extended to nonmetal substrates, which have long been considered as unsuitable for RAIRS measurements due to their inherent low reflectivity and the often very complex band shapes of the film absorptions (see **Figure 3B**). In addition, the partial transparency of nonmetallic substrates also allows other surface techniques such as transmission or ATR spectroscopy to be used. The main advantages of RAIRS, on the other hand, are:

1. RAIRS requires no special sample geometry; it is based on a simple optical configuration and can therefore be used as an *in situ* technique for many surface processes.
2. RAIRS spectra are not affected by substrate absorptions that often block certain spectral regions in transmission and ATR spectra.
3. The complex band profiles of RAIRS spectra on nonmetal substrates contain highly specific information and allow—in combination with spectral simulations—a very detailed analysis of the film structure and composition.

An example of the unique sensitivity of RAIRS with nonmetals toward changes of the adsorbate structure is shown in **Figure 7**, where the formation of a monolayer film of octadecylsiloxane (ODS,  $\text{C}_{18}\text{H}_{37}\text{SiO}_x$ ) is monitored by RAIRS. Initially, the film molecules adsorb in a disordered, liquid-like configuration giving rise to 'negative' absorptions (pointing downward in an absorbance spectrum in accordance with an isotropic  $54.7^\circ$  dipole moment tilt angle, see **Figure 3C**) at  $2858\text{ cm}^{-1}$  ( $\nu_s(\text{CH}_2)$ ),  $2928\text{ cm}^{-1}$  ( $\nu_{as}(\text{CH}_2)$ ), and  $2960\text{ cm}^{-1}$  ( $\nu_{as}(\text{CH}_3)$ ). In the course of the monolayer formation process, the negative  $\nu(\text{CH}_2)$  absorptions are successively converted into positive peaks at lower wavenumbers ( $2850$  and  $2918\text{ cm}^{-1}$ ), indicating the formation of ordered film domains with vertically aligned hydrocarbon chains (dipole tilt angle  $0$ – $10^\circ$ ) to yield the final and highly characteristic spectrum of an ordered, densely packed ODS monolayer with sharp positive  $\nu(\text{CH}_2)$  peaks at  $2950$  and  $2918\text{ cm}^{-1}$  and negative  $\nu(\text{CH}_3)$  peaks at  $2878$  and  $2968\text{ cm}^{-1}$ . The same process on a metal substrate would lead to only minor band shifts and would be much harder to identify and to analyze than on silicon, where the



**Figure 7** RAIRS spectra in the CH stretching region monitoring the growth of an octadecylsiloxane monolayer on silicon.

disorder–order transition is associated with band inversions.

### Spectroelectrochemistry

IR-spectroscopic *in situ* studies of electrochemical processes at the electrode/electrolyte interface have recently gained great importance. The identity of reaction intermediates as well as their orientation and bonding on the electrode surface as a function of the electrode potential can be determined from RAIRS spectra. Besides the weak signal intensity of the adsorbed species, the strong absorption of the electrolyte imposes an additional problem that can be overcome by a special design of the electrochemical cell and the application of signal modulation techniques. Minimization of light absorption by the electrolyte requires the use of thin layer cells with a thickness of a few micrometers. The resulting drawbacks in the electrochemical behavior of these cells are a high ohmic resistance across the electrolyte film, a slow response to potential changes, and a hindered diffusion of reactants, which limit their application to low current processes. To eliminate the absorption of the bulk electrolyte solution several signal modulation techniques have been developed, among which PM-RAIRS and subtractively normalized interfacial FTIR spectroscopy (SNIFTIRS) are



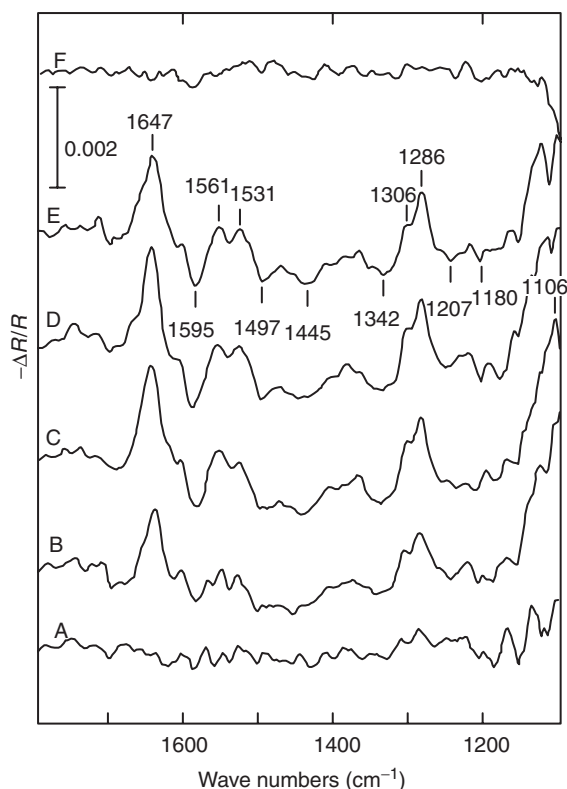
most commonly used in conjunction with FTIR spectrometers. PM-RAIRS has been described earlier as a centerburst suppression technique, yielding a difference spectrum  $R_p - R_s$  between the reflectances of s- and p-polarized radiation. In an electrochemical cell a surface layer of  $\sim 1 \mu\text{m}$  thickness is probed with this technique and absorptions of the bulk electrolyte are eliminated by dividing two spectra taken at different electrode potentials. The SNIFTIRS method is based on a modulation of the electrode potential and yields a difference spectrum  $R^S - R^0$  between the reflectances at the sample and the reference potential. A square wave potential modulation with frequencies  $\sim 100 \text{ mHz}$  is usually employed allowing repeated co-addition of interferograms at each potential step until an acceptable SNR is achieved.

Figure 8 shows an example of a SNIFTIRS study of the electrochemical oxidation of 2,5-dihydroxythiophenol (DHT) on a gold electrode surface according to Scheme 1.

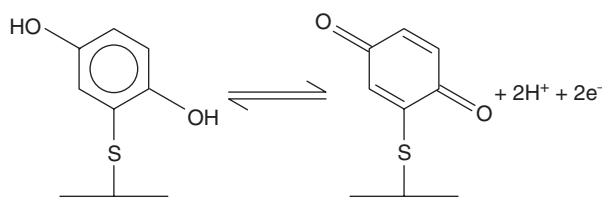
The spectra in Figure 8 were taken at different sample potentials between 0.1 and 0.6 V and were normalized to the reference potential of 0.1 V. Downward-pointing peaks correspond to a loss in starting material DHT and upward-pointing peaks to a gain in reaction products. The most prominent loss features are assigned to the C–O stretching ( $1207$ ,  $1180 \text{ cm}^{-1}$ ), the in-plane OH deformation ( $1342 \text{ cm}^{-1}$ ), and the aromatic CC stretching mode ( $1595 \text{ cm}^{-1}$ ) of DHT. The positive absorptions correspond to the C=O stretching ( $1647 \text{ cm}^{-1}$ ), the C=C stretching ( $1531$ ,  $1561 \text{ cm}^{-1}$ ), and the C–C stretching vibration ( $1286$ ,  $1306 \text{ cm}^{-1}$ ) of the oxidized benzochinonethiol. All these vibrations have a transition dipole moment parallel to the molecular plane. Their relatively large intensity provides evidence that both species are adsorbed with the molecular plane close to normal to the surface.

### Liquid Surfaces and Buried Interfaces

IR surface analysis is not restricted to solid/gas or solid/liquid interfaces, but can also be used to investigate liquid substrate surfaces (liquid/gas interfaces) and even solid/solid interfaces buried in the bulk of a solid sample. Even though the experimental setup and the measurements of such samples are fairly complex and far from a routine analytical technique, two examples shall be described here in order to show the potential and the diversity of IR spectroscopy in the field of surface analysis. In both cases, an experimental setup equal or similar to RAIRS is used with the light incidence angle being the only variable that must be adapted to the specific requirements of the sample system. In the first



**Figure 8** SNIFTIRS spectra of 2,5-dihydroxythiophenol adsorbed on a gold electrode at different electrode potentials and normalized to a 0.1 V reference potential: 0.2 V (A), 0.4 V (B), 0.5 V (C), 0.6 V (D), 0.4 V (E), and 0.1 V (F). (Reprinted with permission from Sasaki T, Bae IT, Scherson DA, Bravo BG, and Soriaga MP (1990) Oxidation-state changes of molecules irreversibly adsorbed on electrode surfaces as monitored by *in situ* Fourier transform infrared reflection absorption spectroscopy. *Langmuir* 6: 1234–1237; © American Chemical Society.)



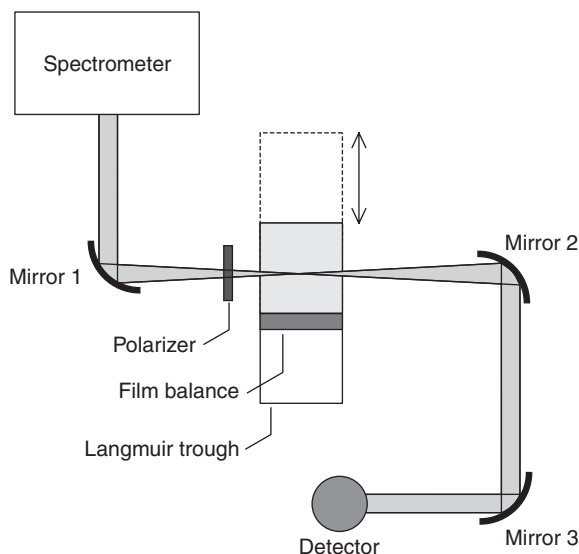
**Scheme 1**

example, the goal is to detect and analyze an adsorbate layer formed at the surface of water in contact with air. Ordered monolayer films of water-insoluble compounds can be formed on a water surface via the Langmuir–Blodgett technique, where the film molecules are first spread onto the water subphase in a so-called Langmuir trough and are then compressed with the film balance and thereby forced into a densely packed, uniform orientation. The IR beam is focused onto the water surface just like with

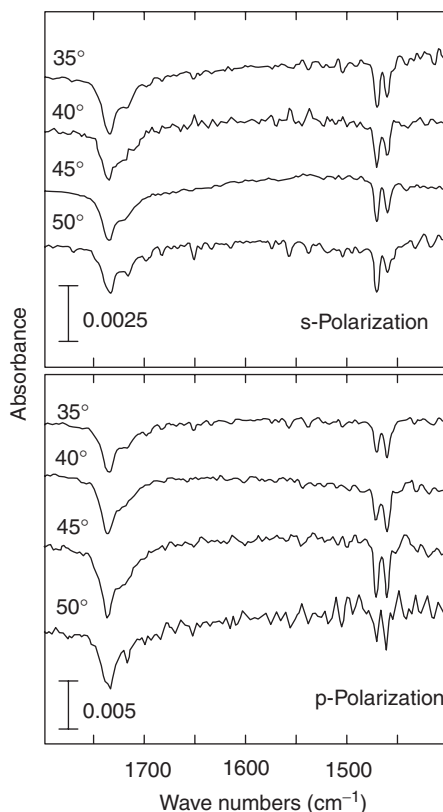
a solid substrate, whereby the highest sensitivity is generally achieved with p-polarized light and an incidence angle close to the water's Brewster angle ( $53^\circ$ ). A typical instrumental setup is shown schematically in **Figure 9**, where the Langmuir trough is moved via a computer-controlled shuttle system between the sample position (compressed monolayer film) and the reference position (bare water subphase). Note that the sample surface in **Figure 9** is oriented horizontally and the light incidence plane lies vertically, opposite to the conventional RAIRS setup shown in **Figure 4**. **Figure 10** shows a series of spectra of a behenic acid methyl ester ( $\text{C}_{21}\text{H}_{43}\text{COOCH}_3$ ) monolayer on  $\text{D}_2\text{O}$ , measured with s-polarized and p-polarized light at different angles of incidence. The ester carbonyl band is observed at  $1737\text{ cm}^{-1}$  with a shoulder at  $1720\text{ cm}^{-1}$ , indicating that most of the carbonyl groups are monomeric and uncoordinated (main peak at  $1737\text{ cm}^{-1}$ ) whereas only a small fraction is H-bonded to the water subphase. The  $\text{CH}_2$  scissor vibration near  $1465\text{ cm}^{-1}$  is split into two components, which is a highly characteristic indication for an orthorhombic subcell structure of the hydrocarbon chains with an all-*trans* chain orientation normal to the water surface.

An even more challenging task is the investigation of solid/solid interfaces, which are ubiquitous in any

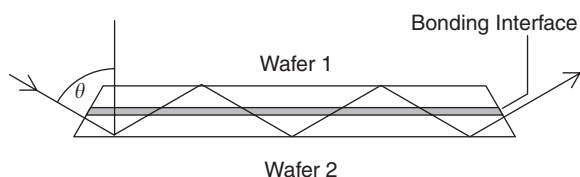
object containing welded, soldered, or glued parts. IR spectroscopy is one of the few analytical techniques that can not only penetrate to a solid/solid interface within a bulk solid sample, but can also provide a unique sensitivity enhancement up to monolayer sensitivity at such interfaces. A special sampling technique termed multiple internal transmission (MIT) has been developed originally to investigate a bonding method known as silicon wafer bonding, where two silicon surfaces are activated and brought into close contact, upon which a spontaneous bonding reaction takes place to form a covalent linkage between the two wafers. The optical lightpath through the sample in such an MIT experiment is shown in **Figure 11** and is reminiscent of the ATR technique, the difference being that MIT probes the interface layer not by the evanescent wave, but by the primary light beam that tunnels through the solid/solid interface and carries on the spectral information of the interface region. The best sensitivity is achieved for high incidence angles (large  $\theta$ ), where



**Figure 9** Optical setup for external reflection infrared measurements of Langmuir–Blodgett monolayers on liquid water surfaces. The arrow indicates the shuttling direction of the Langmuir trough in order to switch between the sample (monolayer on water) and the reference (pure water) position. (Reprinted with permission from Flach CR, Gericke A, and Mendelsohn R (1997) Quantitative determination of molecular chain tilt angles in monolayer films at the air/water interface: infrared reflection/absorption spectroscopy of behenic acid methyl ester. *Journal of Physical Chemistry B* 101: 58–65; © American Chemical Society.)



**Figure 10** External reflection infrared spectra of behenic acid methyl ester monolayers on water, measured with s-polarized and p-polarized light at different incidence angles. (Reprinted with permission from Flach CR, Gericke A, and Mendelsohn R (1997) Quantitative determination of molecular chain tilt angles in monolayer films at the air/water interface: infrared reflection/absorption spectroscopy of behenic acid methyl ester. *Journal of Physical Chemistry B* 101: 58–65; © American Chemical Society.)

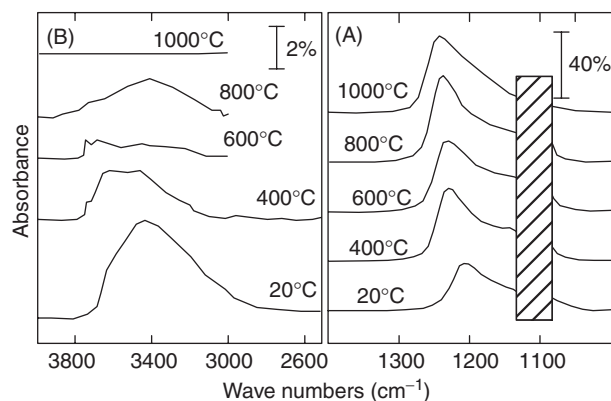


**Figure 11** Light path of the IR beam within a bonded silicon wafer pair used in the multiple internal transmission method. The IR radiation is totally reflected at the outer Si/air surfaces and tunnels repeatedly through the bonding interface.

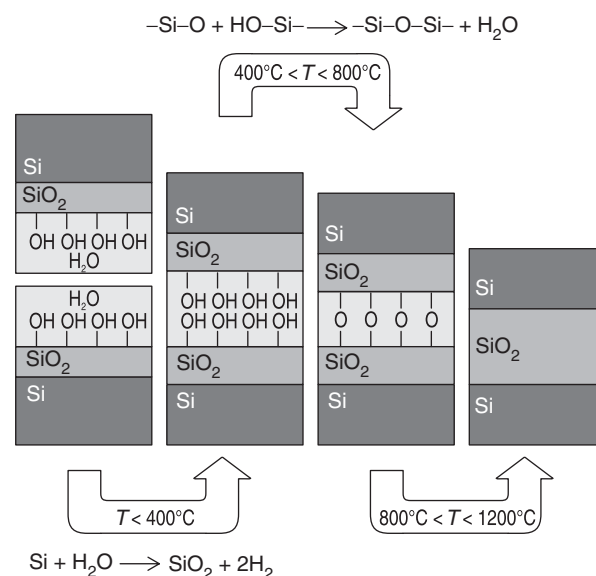
the perpendicular vibrational components at the interface are enhanced by several orders of magnitude compared to a standard transmission measurement. Thus, the optical setup and the surface selection rules are essentially identical to the RAIRS method with metal substrates. **Figure 12** shows the changes in the  $\nu(\text{SiO})$  and in the  $\nu(\text{OH})$  spectral regions of MIT spectra monitoring the bonding process as a function of the annealing temperature. The  $\nu(\text{SiO})$  absorption continuously increases and shifts to higher frequencies, indicating a thickness increase of the native surface oxide layers at the bonding interface. Concomitantly, the broad  $\nu(\text{OH})$  band at  $3500\text{ cm}^{-1}$  decreases, transforms into a sharp single absorption at  $3740\text{ cm}^{-1}$  at  $600^\circ\text{C}$ , reappears as a broad intense band at  $800^\circ\text{C}$ , and completely disappears at  $1000^\circ\text{C}$ . Based on these spectral observations, a bonding mechanism shown schematically in **Figure 13** has been proposed based on the following reaction sequence: at temperatures  $< 400^\circ\text{C}$ , the water molecules trapped at the interface diffuse through the oxide layers and form additional oxide at the Si/SiO<sub>2</sub> interface; the oxide thickness increases and the surface hydroxyl groups come into intimate contact. Between 400 and  $800^\circ\text{C}$ , a condensation reaction takes place between the interfacial hydroxyl groups resulting in Si–O–Si linkages between the two wafers and the formation of water, as indicated in the IR spectra by the reappearance of the  $\nu(\text{OH})$  water absorptions at  $800^\circ\text{C}$  and the growth of a characteristic Si–O–Si absorption at unusually high frequency due to the strained geometry of the bridging Si–O–Si bonds. Finally, between 800 and  $1200^\circ\text{C}$ , a combination of viscous flow in the oxide layer and elastic wafer deformations leads to complete closure of the interface under attainment of the ultimate bonding strength of a bulk quartz crystal.

### Surface Enhanced Infrared Absorption

The surface enhanced infrared absorption (SEIRA) effect, which is closely related to the better known surface enhanced Raman (SERS) effect, was



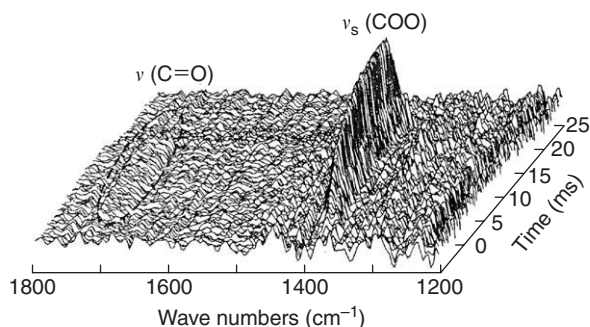
**Figure 12** MIT-IR spectra in the  $\nu(\text{SiO})$  region (A) and in the  $\nu(\text{OH})$  region (B) of the interface between two silicon wafers as a function of the annealing temperature. An optical setup as in **Figure 10** with an incidence angle of  $60^\circ$  and six internal passes was used. The shaded area indicates a region of intense substrate absorption. (Reprinted with permission from Weldon MK, Marsico VE, Chabal YJ, *et al.* (1996) Infrared spectroscopy as a probe of fundamental processes in microelectronics: silicon wafer cleaning and bonding. *Surface Science* 368: 163–178; © Elsevier.)



**Figure 13** Schematic bonding process of two native silicon wafers upon heating to  $1200^\circ\text{C}$ .

discovered more than 20 years ago but has been experimentally exploited primarily within the last decade. It is based on the phenomenon that an adsorbate deposited onto a microscopically rough metal substrate experiences an enhancement of its vibrational modes up to three orders of magnitude as compared to the same adsorbate on a flat substrate. The theoretical origin of this effect is still not completely

understood, yet it is widely used today as sensitivity amplifier in trace analysis, biosensing, catalysis, electrochemistry, time-resolved spectroscopy, etc. Properly prepared samples usually consist of an IR-transparent substrate ( $\text{CaF}_2$ ,  $\text{KBr}$ ,  $\text{ZnSe}$ ,  $\text{Si}$ ) onto which a SEIRA-active metal (usually  $\text{Ag}$ ,  $\text{Au}$ ,  $\text{Cu}$ ,  $\text{Pt}$ ) is deposited with strict control over thickness and morphology (shape, size, and density of the metal islands). Subsequently, the analyte is adsorbed onto this substrate and measured by either transmission, internal reflection (ATR), or external reflection (RAIRS). The obtained SEIRA spectra are usually not only more intense than the unenhanced reference spectra, but show very different relative band intensities and sometimes also bipolar or inverted absorptions, which initially increase very strongly with the adsorbed amount of analyte but level off at adsorbate thicknesses of a few nanometers. These results indicate that the enhancement is a truly local surface effect restricted to the first few layers of an adsorbate. If a surface process to be investigated is reversible, the sensitivity enhancement provided by SEIRA can be used to obtain time-resolved spectra with submicrosecond time resolution using step-scan FTIR. An example is shown in Figure 14, where the adsorption and desorption of fumaric acid ( $\text{HOOC}-\text{CH}=\text{CH}-\text{COOH}$ ) on a gold electrode, triggered by a square-wave potential modulation of the Au substrate, is monitored at  $100\text{ }\mu\text{s}$  intervals. Two major absorptions grow and disappear in Figure 14 synchronously, the carbonyl vibration  $\nu(\text{C}=\text{O})$  at  $1720\text{ cm}^{-1}$  and the symmetric carboxylate vibration  $\nu_s(\text{COO}^-)$  at  $1380\text{ cm}^{-1}$ . The concurrent presence of both bands indicates that fumaric acid bonds to the surface by adsorption and deprotonation of one carboxylic group, whereas the other  $\text{COOH}$  group is left intact and points away from the surface.



**Figure 14** Surface-enhanced time-resolved IR reflection spectra monitoring the adsorption and desorption of fumaric acid on a gold electrode. (Reproduced with permission from Osawa M (2002) Surface-enhanced infrared absorption spectroscopy. In: Chalmers JM and Griffiths PR (eds.) *Handbook of Vibrational Spectroscopy*, vol. 2, pp. 785–799. Chichester: Wiley; © John Wiley & Sons Ltd.)

## Trends

Despite being in use for over a 100 years, IR spectroscopy still experiences new and exciting developments, which steadily enlarge the range of applications and the types of samples to be studied. This applies, in particular, for IR surface spectroscopy, where several complete new branches of IR spectroscopy developed over the past 20 years for bulk samples have recently been adapted and applied to surface probes. One of them is time-resolved spectroscopy, which has been pushed down to the femtosecond range for conventional, transmission-type experiments with bulk samples utilizing pulsed light sources and ultrafast detectors. Obvious applications of time-resolved studies in surface science are the dynamics of surface reactions and surface diffusion, the monitoring of film growth processes such as chemical vapor deposition or mechanistic studies of surface-catalyzed reactions, characterization of transient species, etc., to name a few. Time resolutions of a few microseconds have recently been achieved with powdered, high surface area samples using step-scan FTIR. On single crystals and other low surface area supports with intrinsically lower SNRs, sampling times in the tenths of a second range are still required for covering a broader spectral range with FT-based techniques. Certain laser-based IR methods, on the other hand, such as sum frequency generation, allow time resolution down to femtoseconds, but their drawbacks are a limited spectral range (typically  $<100\text{ cm}^{-1}$ ) and a complex and expensive instrumentation. This gap between FT and laser-based techniques in terms of time resolution, spectral range, and equipment costs is getting narrower by steady advances on both sides. Another closely related goal in IR surface spectroscopy, imposed by new scientific branches such as nanoscience and nanotechnology, is the increase in spatial resolution, i.e., the ability to probe very small areas of a sample surface or to obtain an IR image of a surface with high spatial resolution, thereby combining spectral information with sample visualization. Based on the transmission IR microscope developed in the 1980s, their beam condensing and magnifying optics have been combined with various surface IR techniques such as ATR, RAIRS, or DRIFT and are now commercially available as surface IR microscopes for a wide range of solid, plane-surface, or powdered samples. As with any optical technique, the spatial resolution is determined by the diffraction limit and lies for surface probes in the mid-IR typically  $\sim 10\text{ }\mu\text{m}$ . Two major, recent developments in this area, which will be one of the main topics of surface analysis with IR in the next decade, can only be briefly mentioned here: near-field vibrational spectroscopy, which makes use of the fact that

an aperture or scattering center of subwavelength dimensions creates in its immediate neighborhood ('near field') an electric field with submicron wavelengths, which can be used to image objects with submicron resolution, and, secondly, IR imaging utilizing focal plane array (FPA) detectors, which will especially benefit in the upcoming years from their recent declassification by the US Department of Defense and the immense effort and funding currently channeled into IR detector technology. Even a combination of the latter two technologies, a near-field vibrational image of a surface obtained with an FPA detector in a few seconds with nanometer resolution is no longer unthinkable and might eventually provide instant chemical information of surfaces with a lateral resolution in the nanometer range.

See also: **Infrared Spectroscopy:** Overview; Photo-thermal.

## Further Reading

- Brunner H, Mayer U, and Hoffmann H (1997) External reflection infrared spectroscopy of anisotropic adsorbate layers on dielectric substrates. *Applied Spectroscopy* 51: 209–217.
- Chabal YJ (1988) Surface infrared spectroscopy. *Surface Science Reports* 8: 211–357.
- Chalmers JM and Griffiths PR (eds.) (2002) *Handbook of vibrational spectroscopy*. Chichester: John Wiley.
- Griffiths PR and deHaseth JA (1986) *Fourier transform infrared spectroscopy*. New York: John Wiley.
- Hayden BE (1987) Reflection absorption infrared spectroscopy. In: Yates JT and Madey TE (eds.) *Vibrational Spectroscopy of Molecules on Surfaces, Methods of Surface Characterization*, vol. 1, pp. 267–344. New York: John Wiley.
- Hirschmugl CJ (2002) Frontiers in infrared spectroscopy at surfaces and interfaces. *Surface Science* 500: 577–604.

# SURFACE-ENHANCED RAMAN SPECTROSCOPY

See **Raman Spectroscopy: Surface-Enhanced**

# SURFACTANTS AND DETERGENTS

**P López-Mahía, S Muniategui, and D Prada-Rodríguez**, University of A Coruña, A Coruña, Spain  
**M C Prieto-Blanco**, Arteixo Quimica, Arteixo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Some chemical compounds such as surfactants are linked to the technological development of human communities. The use of the first known surfactant, soap, has been documented in Sumerian civilization back to ~2500 BC. The first synthetic surfactant, sodium lauryl sulfate (SLS) was synthesized in the laboratory by Dumas and Peligot in 1830; however, until the next century there was lack of necessary technology for industrial-scale production. A wide range of surfactant and commercial applications emerged after the Second World War with the development of petrochemistry, and it caused a change of hygienic–sanitary customs.

Surfactants are a group of organic compounds achieved by chemical synthesis and characterized for specific behavior in solution that makes them especially suitable for many human activities. Surfactant is an abbreviation for a surface-active agent that refers to its ability to reduce the interfacial tension between two phases (liquid–liquid, liquid–solid, and liquid–gaseous). This behavior is caused by the molecular composition in the surfactant, which has a hydrophobic part, composed of alkyl chains, and another part that is an ionic or hydrophilic group.

Detergents are chemical formulations made up of principal (surfactant) and complementary components. The latter may act to enhance the performance (builders, alkalis, complexing agents, ion exchangers, enzymes, antiredeposition agents, foaming stabilizers, whitener or optical brightener, viscosity improver, conditioner or softener, corrosion inhibitor, solvents) and enhancing the appearance (perfumes, dyes, opacifier). The number and type of complementary components depend on further use of formulation: clean fabrics, hard surface, or body.

an aperture or scattering center of subwavelength dimensions creates in its immediate neighborhood ('near field') an electric field with submicron wavelengths, which can be used to image objects with submicron resolution, and, secondly, IR imaging utilizing focal plane array (FPA) detectors, which will especially benefit in the upcoming years from their recent declassification by the US Department of Defense and the immense effort and funding currently channeled into IR detector technology. Even a combination of the latter two technologies, a near-field vibrational image of a surface obtained with an FPA detector in a few seconds with nanometer resolution is no longer unthinkable and might eventually provide instant chemical information of surfaces with a lateral resolution in the nanometer range.

See also: **Infrared Spectroscopy:** Overview; Photo-thermal.

## Further Reading

- Brunner H, Mayer U, and Hoffmann H (1997) External reflection infrared spectroscopy of anisotropic adsorbate layers on dielectric substrates. *Applied Spectroscopy* 51: 209–217.
- Chabal YJ (1988) Surface infrared spectroscopy. *Surface Science Reports* 8: 211–357.
- Chalmers JM and Griffiths PR (eds.) (2002) *Handbook of vibrational spectroscopy*. Chichester: John Wiley.
- Griffiths PR and deHaseth JA (1986) *Fourier transform infrared spectroscopy*. New York: John Wiley.
- Hayden BE (1987) Reflection absorption infrared spectroscopy. In: Yates JT and Madey TE (eds.) *Vibrational Spectroscopy of Molecules on Surfaces, Methods of Surface Characterization*, vol. 1, pp. 267–344. New York: John Wiley.
- Hirschmugl CJ (2002) Frontiers in infrared spectroscopy at surfaces and interfaces. *Surface Science* 500: 577–604.

# SURFACE-ENHANCED RAMAN SPECTROSCOPY

See **Raman Spectroscopy: Surface-Enhanced**

# SURFACTANTS AND DETERGENTS

**P López-Mahía, S Muniategui, and D Prada-Rodríguez**, University of A Coruña, A Coruña, Spain  
**M C Prieto-Blanco**, Arteixo Quimica, Arteixo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Some chemical compounds such as surfactants are linked to the technological development of human communities. The use of the first known surfactant, soap, has been documented in Sumerian civilization back to ~2500 BC. The first synthetic surfactant, sodium lauryl sulfate (SLS) was synthesized in the laboratory by Dumas and Peligot in 1830; however, until the next century there was lack of necessary technology for industrial-scale production. A wide range of surfactant and commercial applications emerged after the Second World War with the development of petrochemistry, and it caused a change of hygienic–sanitary customs.

Surfactants are a group of organic compounds achieved by chemical synthesis and characterized for specific behavior in solution that makes them especially suitable for many human activities. Surfactant is an abbreviation for a surface-active agent that refers to its ability to reduce the interfacial tension between two phases (liquid–liquid, liquid–solid, and liquid–gaseous). This behavior is caused by the molecular composition in the surfactant, which has a hydrophobic part, composed of alkyl chains, and another part that is an ionic or hydrophilic group.

Detergents are chemical formulations made up of principal (surfactant) and complementary components. The latter may act to enhance the performance (builders, alkalis, complexing agents, ion exchangers, enzymes, antiredeposition agents, foaming stabilizers, whitener or optical brightener, viscosity improver, conditioner or softener, corrosion inhibitor, solvents) and enhancing the appearance (perfumes, dyes, opacifier). The number and type of complementary components depend on further use of formulation: clean fabrics, hard surface, or body.



Surfactants not only have detergency functions, but also other functions in formulation (softener, opacifier, conditioner, foam booster).

The most common classification of surfactants is based on the criterion of ionic group charge: there are anionic, cationic, nonionic, and amphoteric surfactants (Table 1). The anionic surfactants are primary surfactants that represent the principal detergent basis of formulations and a greater production of surfactants (millions of tons). The nonionic surfactants are the second group for amount of production and are used as emulsifiers and humectants. The property of cationic surfactants to adsorb on a surface makes possible their application as a softener, conditioner, and disinfectant. The amphoteric surfactants are compounds used frequently in cosmetics as detergents because they are less irritating and aggressive than anionic surfactants.

The analytical methodologies of surfactants and detergents are published by national organizations. To mention some of them, the British Standards Institute, BSI (London), Association Française of Normalisation Tour Europe, AFNOR (Paris), the Deutsches Institut für Normung, DIN (Berlin), The American Society for Testing Materials, ASTM (Philadelphia), and the Japanese Industrial Standards Committee, JISC (Tokyo). Information relating to standard methodologies accepted by national organizations might be obtained through the International Standards Organization (ISO).

The development of methods expanded with the appearance of new types and matrices of surfactant as well as new analytical techniques. Current trends include the study of environmental matrices, residual components, application of techniques such as capillary electrophoresis (CE) and combined techniques such as liquid chromatography and mass spectrometry (LC-MS).

This article discusses the state of the art of surfactant analyses taking into account classical methods and new approaches. Titrimetric and instrumental methods, analysis in two matrices (formulation and environmental), sample treatment, and qualitative and quantitative analysis will be examined.

## General Methods of Analysis

The analytical chemistry of surfactants presents some difficulty because most products contain a homologous chain and positional isomer in the hydrophobic part and may have an ethoxylated oligomer in the hydrophilic part. Also, they are compounds of low volatility with molecular weights above 200.

The titrimetric methods for determination of anionic and cationic surfactant concentrations are based on specific reactions between anionic and cationic surfactants in a two-phases system, chloroform-water, and with a transfer phase indicator for

**Table 1** Commonly used surfactants

	<i>Carboxylates</i>	<i>Sulfates</i>	<i>Sulfonates</i>
Anionics	Acyl sarcosinates Acylated peptides	Sulfated amides Alkyl sulfates Alkyl ether sulfates Sulfated alcohols Sulfated esters	Alkylbenzene sulfonates Dialkylsulfosuccinates Olefin and alkane sulfonates Sulfonates amides Sulfonated esters Sulfonates fatty esters Metal alkylphosphates Soaps
Cationics	Monoalkyl quaternary ammonium salts Dialkyl quaternary ammonium salts Amine oxides Imidazolinium derivatives Alkylamidoamines derivatives Alkylquinolinium and alkylisoquinolinium salts		Polymeric cationic Alkylpyridinium salts Alkylbenzylammonium salts Bis-quaternary ammonium derivatives Amines
Nonionics	Alcohols, ethoxylated Alkylphenols, ethoxylated Fatty alcohols Esters of polyhydric alcohols Esters of polyoxyalkylene glycols		Ethers of polyoxyalkylene glycols Dialkyl polyoxyalkylenephosphates Alkyl glycoside Polyoxyalkene oxide copolymers
Amphoterics	Carboxybetaines		Sulfobetaines

location of the endpoint. The most used indicator is a mixture of dimidium bromide and disulfine blue. Hyamine 1622 is used as titrant for anionic surfactant analysis and SLS for cationic surfactant analysis. This method is more sensitive than traditional titration and is quick and effective for quality control. A potentiometric titration using a surfactant-selective electrode is an alternative to two-phase titration that has advantages including a shorter analysis time and avoiding the use of a hazardous solvent such as chloroform.

Gas chromatography (GC) allows the distribution of the alkyl chain and ethoxylates to be ascertained but a sample treatment to convert surfactants into volatile compounds is necessary. Anionic surfactants with sulfonates and sulfates can be broken down by acid hydrolysis to yield alcohols. Derivatization using several agents is another way to achieve a volatile compound, for example, alkylaryl-, alpha olefin-, alkene-, and paraffinsulfonates can be converted to sulfonyl chlorides or methyl esters of alkyl sulfates (AS) to trimethylsilyl derivatives and linear alkylbenzenesulfonates (LAS) and AS to thiotrifluoroacetyl derivatives. Cationic surfactants of the alkyltrimethyl- and dialkyldimethylammonium type are thermally degraded to tertiary amines or by Hoffmann degradation to an olefin and alkylamine. Nonionic surfactants are treated by derivatization or cleavage, except for low molecular weight compounds that are analyzed directly. Alcohol ethoxylates (AE) and alkylphenol ethoxylates (APE) can be derivatized to form trimethylsilyl derivatives and acetate esters or fragmented in cleavage reactions with HBr, HI, or acetyl chloride.

LC provides information on the distribution of homologs and makes possible quantitative analysis of individual surfactants in mixtures without sample pretreatment. Three classes of aromatic surfactants (anionic as alkylarylsulfonates, cationic as alkylpyridinium and alkylbenzyltrimethylammonium, nonionic as APE) are analyzed using ultraviolet (UV) absorbance detection. A suitable alternative for nonaromatic surfactants such as alkenesulfonates, AS, or alkyltrimethylammonium is indirect detection with a mobile phase containing a UV-absorbing ion. Conductometric detection, combined with reversed-phase and ion chromatography, is used for analysis of all common anionic and nonaromatic cationic surfactants (alkyltrimethyl- and dialkyldimethylammonium). Simultaneous separation of an individual surfactant (alkylpyridinium, alkylbenzyltrimethyl-, and alkyltrimethylammonium) and its homologs may be achieved using a refractive index detector. Fluorescence detection is chosen for trace analysis of surfactants.

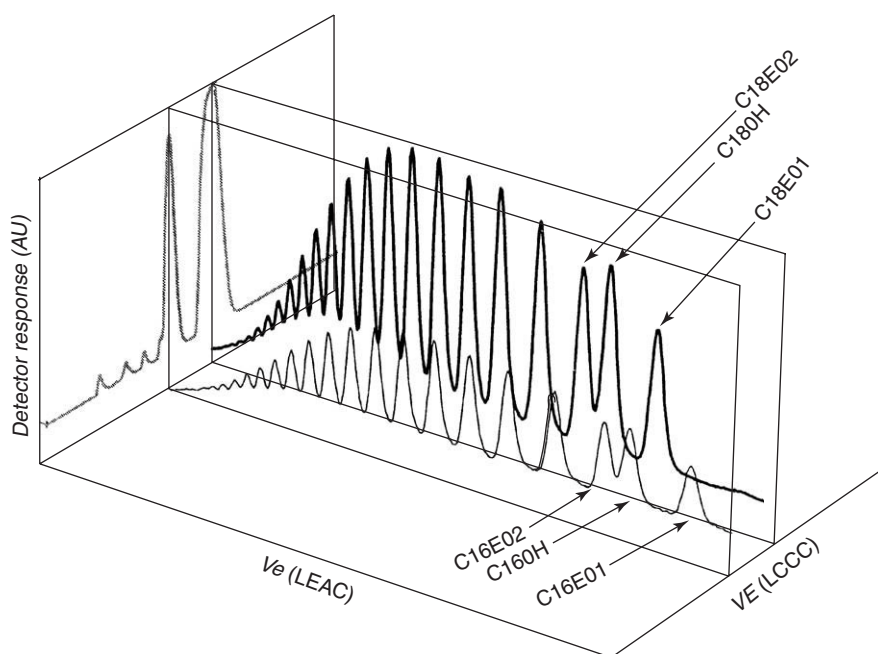
Reversed-phase LC, occasionally with paired ions, allows alkyl chain homologs to be separated from anionic and cationic surfactants, although for the latter, normal and reversed-phase LC on a cyano column is applicable. For nonionic surfactants, in general, reversed-phase LC using a C<sub>18</sub> column gives a separation on the basis of alkyl chain length and normal-phase chromatography according to the ethoxy chain distribution.

Automated chromatographic systems allow the determination of the ethoxy chain distribution and alkyl chain length by combining two separation methods for a nonionic surfactant. It is even possible to use two-dimensional chromatography under critical conditions (LCCC) and liquid exclusion-adsorption chromatography (LEAC). For example, separations are carried out for the alkyl chain in fatty AE by LCCC and the ethoxy chain by LEAC using refractive index detection (Figure 1).

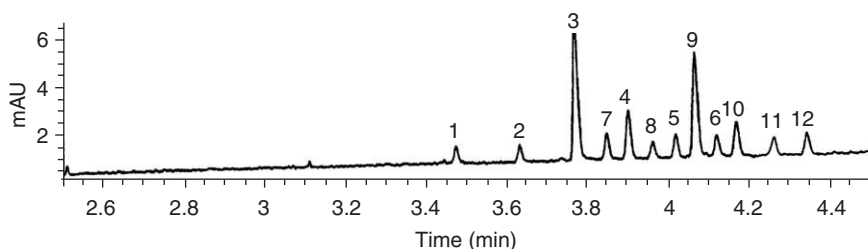
CE is a separation technique that may be an alternative to LC to determine alkyl chain distribution of anionic and cationic surfactants. Most of the studies refer to alkylbenzyltrimethylammonium with UV absorbance detection (Figure 2). Other uses of CE include the separation of homologs of anionic surfactants (as AS) and cationic (as alkyltrimethylammonium) by isotachopheresis with conductometric detection and homologs of non-UV absorbing surfactants (AS, alkylsulfonate, alkylsarcosinates and dialkyldimethylammonium) by capillary zone electrophoresis using indirect detection.

Infrared (IR) spectroscopy and nuclear magnetic resonance (NMR) spectroscopy are two techniques of great relevance for the characterization of surfactants and their mixtures which are applied for quality control. The degree of branching of the alkyl chain in LAS can be measured by either of them. <sup>13</sup>C NMR has been utilized for determination of individual LAS isomers. Other examples of NMR applications include the measurement of the ratio of alkyl to ethoxy chain length in APE, to quantify mixtures of dialkyldimethyl and monoalkyltrimethylammonium, and to characterize alpha olefinsulfonates. Regarding IR spectroscopy, some applications focused on quality control are the determination of ethylene oxide content in ethoxylated surfactants or number average molecular weight by measurement of the hydroxyl groups.

MS is utilized for the characterization of structures and determination of molecular weight. The production of gaseous ions from surfactants is carried out using soft ionization techniques. The chosen technique depends on the surfactant but those frequently used are field desorption and fast atom bombardment



**Figure 1** Two-dimensional separation of Dehydol T A5: C<sub>16</sub>–C<sub>18</sub> ethoxylated alcohol. (Reprinted with permission from (2002) *Journal of Chromatography* 952: 149–163; © Elsevier.)



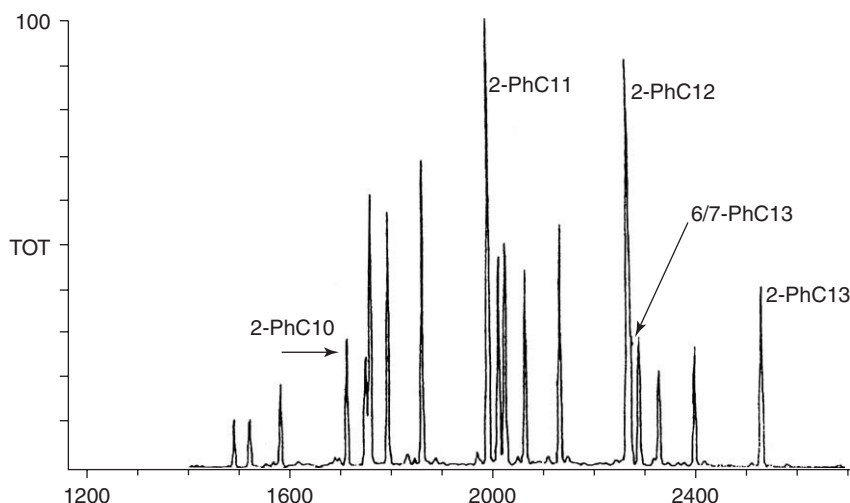
**Figure 2** Simultaneous separation of alkylamidobenzyltrimethylammonium (BAA) and alkylbenzyltrimethylammonium (BAK) by CE. Peaks: 1 = 8-BAK, 2 = 10-BAK, 3 = 12-BAK, 4 = 14-BAK, 5 = 16-BAK, 6 = 16-BAK, 7 = 8-BAA, 8 = 10-BAA, 9 = 12-BAA, 10 = 14-BAA, 11 = 16-BAA, 12 = 16-BAA. (Reprinted with permission from Turnes-Caron *et al.* (2002) *Chromatographia*, 56(9/10): 605–609.)

(FAB). The information of the molecular weight distribution of surfactant oligomers is provided by the parent ion spectrum. Other structural information include location of positions of unsaturation, isomer distribution, and degree of branching. The use of FAB-MS for anionic surfactant analysis makes possible the determination of chain length and branching as well as distinguishing between linear alkylbenzenesulfonates (ABS and LAS). In addition, it has been applied to ethoxylates and cationic surfactants analysis.

MS can be coupled with GC, LC, supercritical fluid chromatography, and MS (Figure 3). Most of the studies by MS have referred to commonly used surfactants such as LAS, APE, and alkylbenzyltrimethylammonium in environmental matrices.

## Analysis of Detergent Formulations and Other Formulated Products

The formulation class (tooth paste, shampoo, fabric softener, etc.) provides basic information about the components present. Standard shampoo formulations contain two types of surfactants (either two anionic or one anionic and one amphoteric surfactant); foam booster (nonionic surfactant); conditioner, viscosity improver, opacifier, dye, perfume, chelate, and preservatives. Normally, formulated products are available as liquid, solid, emulsion, dispersion, etc, with a great number of components (in some cosmetic products there may be 20 components) and some of them, such as the preservatives, are found at low concentrations.



**Figure 3** Separation of LAS alkyl chain isomers converted into thiotrifluoroacetates by GC-MS. (Reproduced (reprinted) with permission from *Rivista Italiana delle Sostanze Grasse* (2001), 78: 273–276.)

Initial examination generally starts with the determination of total solids, volatile matter, and water intended for mass balance calculations. The difference between the two later parameters is the content of organic solvent and perfumes. A general procedure consists of the following: from solids, the ethanol-soluble material (where surfactants are found) is separated from the insoluble material (containing inorganic compounds). Isolation of different classes of surfactants (anionic, cationic, amphoteric, and nonionic surfactants) from the organic fraction is realized by ion exchange, column and thin layer chromatography, and liquid–liquid extraction. Alternatives to classical methods, solid-phase extraction (SPE) and supercritical fluid extraction (SFE), reduce analysis time and solvent consumption. Solid-phase microextraction has demonstrated useful properties for sampling aqueous samples. The choice of separation mode depends, at least, on the type of required information, on the matrix, and available instrumentation.

IR is a nondestructive technique suitable for the analysis of formulated products, and gives a considerable amount of information about the compounds present. Near-infrared (from 13 000 to 4000  $\text{cm}^{-1}$ ) and Fourier transform infrared (FTIR) (from 4000 to 400  $\text{cm}^{-1}$ ) spectroscopies are used. Qualitative analysis of the ethanol soluble fraction allows the identification of functional group types such as hydrotropes (xylenesulfonate and toluenesulfonate). In addition, zeolite, alkalis, polymers, and builders may be identified in the insoluble ethanol fraction. For quantitative analysis, method development is slow because a great number of calibration standards

are necessary; however, sample analysis is achieved in a few minutes. Alkyl ether sulfate (AES) and fatty alkanolamide are quantified in shampoos by FTIR spectrometry with similar results obtained by two-phase titration. Also, powder soap, heavy-duty laundry detergent, and bar soap are analyzed by FTIR spectrometry using diffuse reflectance for solid samples.

Surfactants, hydrotropes, and sequestering agents are analyzed without sample pretreatment in liquid detergent by  $^{13}\text{C}$  NMR spectroscopy and phosphorus species present in detergents are quantified by  $^{31}\text{P}$  NMR spectrometry.

The analysis of solvent and perfumes in formulations is performed by GC. For solvent analysis in surfactants sample pretreatment with organic solvent is used for the separation of inorganic impurities. Regarding surfactants and complementary compounds, applications may be mentioned such as the separation of a hardness agent (sodium laurate) in liquid laundry detergent by GC and the Hofmann degradation and analyses of cationic surfactant of the alkyltrimethyl- and dialkyldimethylammonium type in fabric softener and hair rinse.

LC is a separation technique of great importance for formulations since it allows the quantification of a broad spectrum of compounds from surfactants to minor compounds in matrices of detergents, cosmetic products, and industrial products. LC using UV absorbance detection is used in the determination of anionic surfactants (alkylbenzene sulfonate) and nonionic surfactants (APE) in liquid pesticide formulations; cationic surfactants such as cetylpyridinium in pharmaceutical tablets; and benzalkonium

chloride in ophthalmic preparations. Nonchromophore surfactants are analyzed in cosmetic products (shampoos, foam bath, make-up remover lotion) by normal-phase and reversed-phase elution with short-wavelength UV detection. These compounds may be nonionic surfactants (alkanolamides), amphoteric surfactants (carboxybetaine and carboxyimidazoline), anionic surfactants (alkylsulfates, alkyl-ethoxysulfates, sarcosinate, and types of sulfonates such as taurate and sulfosuccinate). Examples of other detection modes are the determination of alkyltrimethylammonium (cationic surfactant) in skin moisture formulation using refractive index detection. Also, alkylsulfonates in shampoo and LAS in laundry detergent by LC with postcolumn addition of a solid-phase reagent followed by conductometric detection. The evaporative light-scattering detector is utilized for the determination of alkyl ether sulfates, synthetic and petroleum sulfonates, and AE in industrial formulated products.

Ion chromatography with inverse conductivity detection gives good results for the analysis of minor components such as buffering agents (triethanolamine, monoethanolamine, diethanolamine, and silicate) or an enzyme stabilizer (formate) in laundry detergent. Also, a sequestrant such as citrate may be determined in powdered and liquid laundry detergent by ion-exchange chromatography with UV absorbance detection. Other components, such as enzymes, can be quantified in detergent formulations by specific procedures involving measurement by plasma desorption, laser desorption, or ion-spray MS.

Inductively coupled plasma-atomic emission spectrometry allows the determination of anionic surfactants (LAS and AS) and inorganic compounds (phosphate, silicate, zeolite, sulfate). Other techniques, such as X-ray fluorescence spectroscopy and X-ray powder diffraction, have been used for the qualitative analysis of inorganic detergents. For surface analysis, optical light microscopy, scanning electron microscopy, and transmission electron microscopy characterize particles, deposition of surfactant, or other detergent ingredients on fabric.

Studies performed by MS on cosmetic products demonstrate the utility of this technique for analysis of hair-care surfactants. After sample pretreatment for the separation of different kinds of surfactant, the usual identification of anionic surfactant such as AE, AES, and less common surfactants (lauryl sulfosuccinate, *N*-acyl-*N*-methyl taurate, paraffin sulfonate) is possible. Amphoteric surfactants (cocoamidopropylbetaine) and those present less frequently (lauryl hydroxysultaine) have been identified by this technique as well. For other kinds of components

that are found in hair-care formulations, nonionic surfactants (AE and cocodiethanolamide, CDEA) and cationic surfactants (alkyltrimethylammonium and dialkyldimethylammonium), MS is a valuable technique.

The application of LC-MS achieves separation and identification of surfactants in formulated products. A representative example is the analysis of homologs of a cationic surfactant (ditallowdimethylammonium) and its impurities (tritallowmethylammonium) by normal-phase LC-MS FAB mode.

## Environmental Analysis

The analysis sequence for environmental matrices starts with sampling to provide a representative, sufficient, and well-preserved sample (refrigerator, formaldehyde, glass vessel, etc.). After a concentration step, isolation of other materials and cleanup with different techniques, depending on the matrix (liquid, liquid with particles, solid), is carried out. For aqueous matrices, liquid-liquid extraction with direct evaporation or solvent sublation when emulsions are formed and steam distillation (separation of nonylphenol from water) is used. Solid samples (sediment, sludge, etc.) are Soxhlet extracted. If class fractionation is required (for example, cationic or anionic surfactants), good results are obtained using SPE with ionic-exchange resins. SFE is used for specific problems when there is chemical breakdown in the extraction of solid materials (cationic compounds with ester group) and also for isolation of ionic and nonionic surfactants in sludge. SFE has advantages with respect to classical extraction techniques because it does not use hazardous solvents.

The purpose of analysis is to try and gain information on several aspects: total surfactant concentration, specific surfactant identity, or minor compound identity whose presence is linked to the use of detergent. For the first case, nonspecific methods such as colorimetric or indirect measurements such as degradation (CO<sub>2</sub> or total organic carbon) are enough. Colorimetric methods for cationic and anionic surfactants are based on two-phase extraction, of a complex formed between the surfactant and a colored reagent identified by spectrophotometric measurements. For anionic surfactants, the most common reagent is methylene blue; for cationic surfactants it is disulfine blue. The methods for nonionic surfactants are based on complex formation between the ethoxylated chain and a metal (Dragendorff reagent, ammonium thiocyanatecobaltate) and quantification by AAS or colorimetry. The

results of these methods are expressed as amount of disulfine blue active substance, methylene blue active substance, or cobalt active substance.

Instrumental techniques (HPLC, GC, CE, MS) are applied for the study of specific surfactants. GC, using desulfonation techniques, is used for LAS analysis in river water, sewage, superficial water, and sewage sludge. Also, biodegradation products of LAS are examined by this technique occasionally with MS detection. Homologs and isomers can be separated by LC, albeit with lower resolution. Normally, fluorescence detection is used when concentrations are at levels of submicrograms per liter in river water, treatment plants, and even seawater, sediment, and fish tissue. The first degradation products of LAS, sulfophenylcarboxylic acids, are analyzed in treatment plants as well. LC is available to separate alkyl chain LAS of nonionic surfactants as APE in reversed-phase using fluorescence detection.

Concerning nonionic surfactants, AEs are determined in surface and waste waters by GC after HBr cleavage. Homologs and oligomers of APE can be analyzed after derivatization using normal- and reversed-phase chromatography with fluorimetric detection at microgram per liter levels. Likewise, alkylphenoxycarboxylic acids have been determined by LC in studies of APE biodegradation. The latter may form halogenated by-products during disinfection treatment. LC-MS is employed for APE and halogenated APE analysis in sludge, river sediments, and surface, drinking, and waste waters (Figure 4).

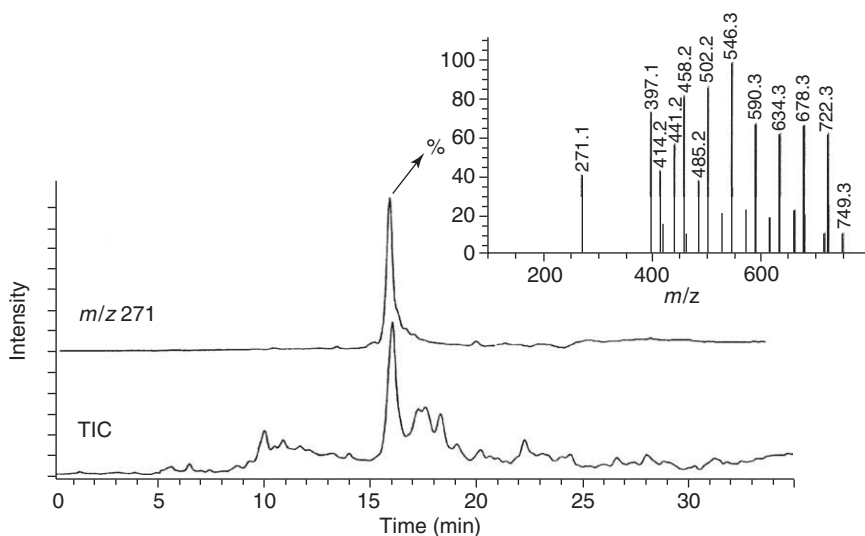
Dialkyldimethylammonium surfactants and specifically DHTDMAC are the cationic surfactants most used, as fabric conditioners. DHTDMAC has been

studied by LC in a nonaqueous medium using conductometric detection at concentrations of submicrogram per liter levels in river water, sewage treatment, sludge, sediment, and soil. Another method applied to DHTDMAC in river water uses fluorescence or UV absorbance detection with a post-column detector system. Cationic surfactants are detected because these form ion-pairs with methyl orange or dimethoxyanthracene sulfonate, providing good detection limits.

Homologs of alkylbenzyltrimethylammonium, particularly benzalkonium chloride (BAK), used as disinfectants and bactericides can be determined at microgram per milliliter levels in effluents from hospitals by LC using postcolumn derivatization and fluorescence detection. Also, BAK was determined after sample treatment using SPE with LC-MS and LC-MS/MS detection in waste and river waters.

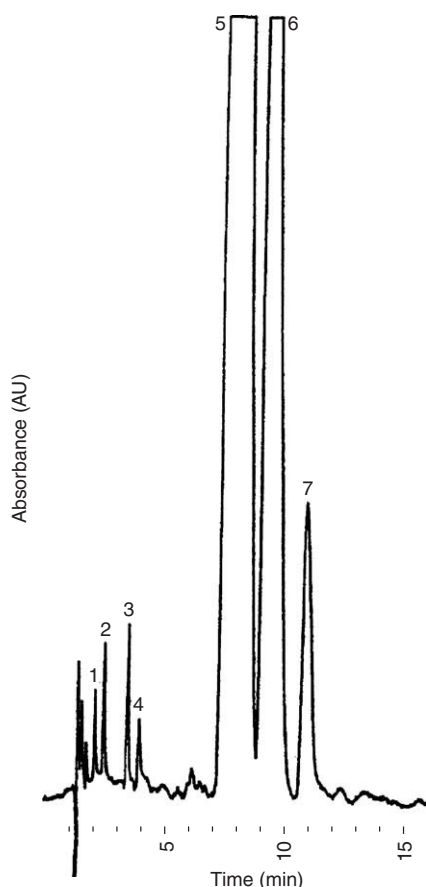
## Residual Components in Surfactants

Residual components in surfactants are undesirable because of their toxicity or inadequacy for the finished product, the result of side and unfinished reactions, mainly from raw materials, and the process. Metals at low concentrations may be encountered in surfactants. Some metals such as iron, nickel, and chromium originate mainly from corrosion of process plant materials; mercury comes from alkalis. Their detection is performed by three modes of AAS: flame AAS for iron determination in detergents; furnace AAS for chromium, nickel, and iron in powder



**Figure 4** Determination of an APE (nonylphenol ethoxylate) in sludge by LC-MS (TIC = total ion current). (Reprinted with permission from *Analytical Chemistry* (2000), 72: 4560–4567; © American Chemical Society.)





**Figure 5** LC separation of benzyl chloride during BAK synthesis. Peaks: 1 = benzyl alcohol, 2 = benzaldehyde, 3 = benzyl chloride, 4 = unidentified impurity, 5 = C<sub>12</sub> homolog, 6 = C<sub>14</sub> homolog, 7 = C<sub>16</sub> homolog. (Reproduced from Prieto-Blanco MC, López-Mahía P, and Prada-Rodríguez D (1999) *Journal of Chromatographic Science* 37: 295–299 by permission of Preston Publications, A Division of Preston Industries, Inc.)

soap with high concentrations of phosphate; and cold vapor generation AAS for mercury.

Organic residual components are the most worrying because of their toxicity. Some of these compounds are formed as by-products. Volatile organic compounds are determined by headspace GC, GC-MS. Intermediate products, such as sultones and sulfones, from sulfonation of olefin and alkylbenzene, respectively, can be detected by LC. Unreacted products, like ethylene oxide from the synthesis of ethoxylated nonionic and anionic surfactants, are studied by GC; benzyl chloride from the quaternization of tertiary amines and aliphatic amines from amidation reaction are determined by LC (Figure 5).

N-Nitrosamines deserve special attention owing to their toxic and carcinogenic properties. They can be formed by nitrosation of secondary amines such as diethanolamine present in triethanolamine and coconut diethanolamide. Their analysis methods are diverse: colorimetric, polarographic, and chromatographic (GC, GC-MS, and LC).

See also: **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Cosmetics and Toiletries. Derivatization of Analytes. Electrophoresis:** Isotachopheresis. **Environmental Analysis. Enzymes:** Overview. **Extraction:** Supercritical Fluid Extraction; Solid-Phase Extraction; Solid-Phase Microextraction. **Ion Exchange:** Ion Chromatography Applications. **Liquid Chromatography:** Reversed Phase; Liquid Chromatography-Mass Spectrometry. **Nuclear Magnetic Resonance Spectroscopy – Applicable Elements:** Carbon-13; Phosphorus-31. **Perfumes.**

## Further Reading

- Cross J (ed.) (1987) *Nonionic Surfactants. Chemical Analysis*. New York: Dekker.
- Cross J and Singer E (eds.) (1994) *Cationic Surfactants. Analytical and Biological Evaluation*. New York: Dekker.
- Di Corcia A (1998) Characterization of surfactants and their biointermediates by liquid chromatography-mass spectrometry. *Journal of Chromatography A* 794: 165–185.
- Karsa DR (ed.) (1992) *Industrial Applications of Surfactants III*. Cambridge: Royal Society of Chemistry.
- Kiewiet AT and De Voogt P (1996) Chromatographic tools for analysing and tracking non-ionic surfactants in the aquatic environment. *Journal of Chromatography A* 733: 185–192.
- Longman GF (1975) *The Analysis of Detergents and Detergent Products*, Ch. X, pp. 215–234. Chichester: Wiley.
- Marcomini A and Zanette M (1996) Chromatographic determination of non-ionic surfactants of the alcohol polyethoxylate type in the environment. *Journal of Chromatography A* 733: 193–206.
- Milwidsky BM and Gabriel DM (1982) *Detergents Analysis. A Handbook for Cost-Effective Quality Control*. New York: Wiley.
- Porter MR (1991) *Recent Developments in the Analysis of Surfactants*. Critical Reports on Applied Chemistry, vol. 32, chs. II, III, and VI. Barking: Elsevier Applied Science.
- Richmond JM (ed.) (1990) *Cationic Surfactants. Organic Chemistry*. New York: Dekker.
- Schmitt TM (1992) *Analysis of Surfactants*. New York: Dekker.

# SWEETENERS

**M C Yebra-Biurrun**, University of Santiago de Compostela, Santiago de Compostela, Spain  
© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Sweeteners are defined as food additives that are used or intended to be used either to impart a sweet taste to food or as a tabletop sweetener. Tabletop sweeteners are products that consist of, or include, any permitted sweeteners and are intended for sale to the ultimate consumer, normally for use as an alternative to sugar. Foods with sweetening properties, such as sugar and honey, are not additives and are excluded from the scope of official regulations. Sweeteners are classified as either high intensity or bulk (**Figure 1**). High-intensity sweeteners possess a sweet taste, but are noncaloric, provide essentially no bulk to food, have greater sweetness than sugar, and are therefore used at very low levels. On the other hand, bulk sweeteners are generally carbohydrates, providing energy (calories) and bulk to food. These have a similar sweetness to sugar and are used at comparable levels.

## High-Intensity Sweeteners

High-intensity sweeteners (also called nonnutritive sweeteners) can offer consumers a way to enjoy the taste of sweetness with little or no energy intake or glycemic response and they do not support growth of oral cavity microorganisms. Therefore, they are principally aimed at consumers in four areas of the food and beverage markets: treatment of obesity, maintenance of body weight, management of diabetes, and prevention and reduction of dental caries. There are several different high-intensity sweeteners. Some of the sweeteners are naturally occurring, while others are synthetic (artificial) or semisynthetic. Most of the more commonly available high-intensity sweeteners and/or their metabolites are rapidly absorbed in the gastrointestinal tract. For example, acesulfame-K and saccharin are not metabolized and are excreted unchanged by the kidney. Sucralose, stevioside, and cyclamate undergo degrees of metabolism, and their metabolites are readily excreted.

Acesulfame-K, aspartame, and saccharin are permitted as intense sweeteners for use in food, virtually worldwide. In order to decrease cost and improve taste quality, high-intensity sweeteners are often used as mixtures of different, synergistically compatible

Sweeteners	High-intensity (non-nutritive)	Synthetic artificial	Acesulfame Alitame Aspartame Cyclamate		Neotame Saccharin Sucralose
		Semisynthetic	Neohesperidine dihydrochalcone		
		Natural	Glycyrrhizin Stevioside Thaumatococin		
	Bulk (nutritive)	Caloric	Sucrose Molasses Honey and maple syrup		
			Starch-derived sweeteners		Glucose Fructose
		Low-caloric	Sugar alcohols	Monosaccharides	Erythritol Mannitol Sorbitol Xylitol
				Disaccharides	Isomalt Lactitol Maltitol
Hydrogenated starch hydrolyzates (HSH)					
		Tagatose			

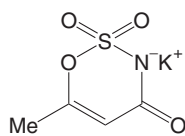
**Figure 1** Classification of sweeteners.

sweeteners. Sweetness characteristics of high-intensity sweeteners are shown in Table 1.

The more important properties of the high-intensity sweeteners that are permitted for use in food and drink applications are shown in Table 2.

### Acesulfame-K

Acesulfame-K [I] has an excellent stability under high temperatures, and good solubility which makes it suitable for numerous products. It is approved for use in food and beverage products in ~90 countries, including the USA, Switzerland, Norway, UK, Canada, Australia, and the European Union (EU).



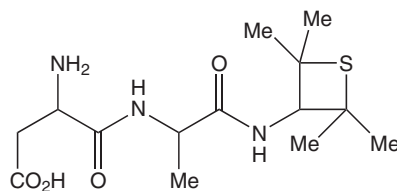
Acesulfame-K

Systematic name: 6-methyl-1,2,3-oxathiazin-4(3H)-one-2,2-dioxide, potassium salt

### Alitame

Alitame [II] is nutritive, but due to its intense sweetness, the amounts used are small enough for it to be considered and classified as a non-nutritive sweetener. Alitame is formed from the amino acids L-aspartic acid and D-alanine with a novel amide moiety (formed from 2,2,4,4-tetramethylthienanylamine). Alitame exhibits superior stability under a variety of conditions because of its unique amide group. Alitame has been approved

for use in some countries such as Australia, Mexico, New Zealand, and China, but not in the USA or EU.

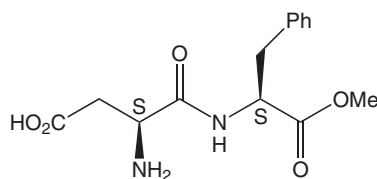


Alitame

Systematic name: L- $\alpha$ -aspartyl-N-(2,2,4,4-tetramethyl-3-thietanyl)-D-alaninamide

### Aspartame

Aspartame [III] is made from two amino acid components, L-aspartic acid and L-phenylalanine. Although nutritive, containing 4 kcal g<sup>-1</sup> like any other protein substance, due to its intense sweetness, the amounts used are small enough for aspartame to be considered and classified as a nonnutritive sweetener. In the dry form, the stability of aspartame is good and little decomposition is observed if the moisture content is kept below 8%. However, in solution, and under heat, it is not stable and undergoes hydrolysis to the free dipeptide and methanol, and cyclodehydration to its diketopiperazine derivative, in both cases with loss of sweetness. People suffering from the metabolic disorder phenylketonuria are unable to metabolize L-phenylalanine resulting from the hydrolysis of the dipeptide and are advised to avoid this sweetener. Hydrolysis of the ester functionality affords methanol but in insufficient quantities to be harmful. Aspartame is approved in more than 90 countries (the EU, the USA, Canada, South America, Australia, Japan, etc.) for use in numerous food-stuffs.



Aspartame

Systematic name: N-L- $\alpha$ -aspartyl-L-phenylalanine-1-methyl ester

**Table 1** Sweetness characteristics of high-intensity sweeteners (the scale uses sucrose as a sweetness of 1, and compares the sweetness of other sweeteners to sucrose)

Sweetener	Relative sweetness (sucrose = 1)	Aftertaste
Acesulfame-K	150–200	Very slight bitter
Alitame	2000–3000	Not unpleasant
Aspartame	160–220	Prolonged sweetness
Aspartame–acesulfame salt	350	–
Cyclamate	30–40	Prolonged sweetness. At high concentrations a distinct sweet–sour lingering
Glycyrrhizin	50–100	Prolonged sweetness (liquorice)
Neohesperidine dihydrochalcone	1000–2000	Lingering menthol–liquorice
Neotame	7000–13 000	Not unpleasant
Saccharin	300–600	Bitter metallic
Stevioside	250–300	Bitter and unpleasant
Sucralose	400–800	Not unpleasant
Thaumatococin	2000	Liquorice

### Aspartame–acesulfame salt

The salt is prepared by heating an ~2:1 ratio (w/w) of aspartame and acesulfame-K in solution at acidic pH. Aspartame–acesulfame dissolves completely in saliva and gastric juice. Although this salt it mainly consists of the two approved sweeteners, it is considered as a separate compound, which requires specific approval in certain countries. In the EU, it is part of the Proposed Amendment to the Sweeteners

**Table 2** Characteristics and selected physical properties of high-intensity sweeteners

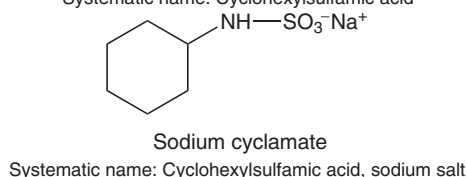
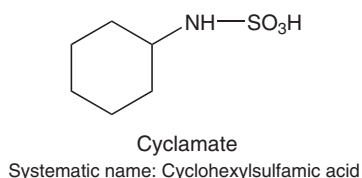
Sweetener	Molecular formula	Molecular weight	m.p. (°C)	Solubility in H <sub>2</sub> O at 20°C (%)	ADI (mg per kg bodyweight)	
					JECFA	SCF
Acesulfame-K	C <sub>4</sub> H <sub>4</sub> NO <sub>4</sub> SK	201.2	200	27	15	9
Alitame	C <sub>14</sub> H <sub>25</sub> N <sub>3</sub> O <sub>4</sub> S	331.43	136–147	14.3 (pH 7)	1	0.3
Aspartame	C <sub>14</sub> H <sub>18</sub> N <sub>2</sub> O <sub>5</sub>	294.31	246	1	40	40
A–A salt	C <sub>18</sub> H <sub>23</sub> O <sub>5</sub> N <sub>3</sub> S	457.46	–	–	Covered by the ADI values previously established for aspartame and acesulfame-K	
Cyclamate	C <sub>6</sub> H <sub>13</sub> NO <sub>3</sub> S	179.24	169–170	7.7 Na salt: 19.5	11	7
Glycyrrhizin	C <sub>42</sub> H <sub>62</sub> O <sub>16</sub>	822.93	–	–	Not specified	Not evaluated (100 provisional)
NHDC	C <sub>28</sub> H <sub>36</sub> O <sub>15</sub>	612.6	156–158	0.05	Not evaluated	5
Neotame	C <sub>20</sub> H <sub>30</sub> N <sub>2</sub> O <sub>5</sub>	378.46	80.9–83.4	1.3	2	Not evaluated
Saccharin	C <sub>7</sub> H <sub>5</sub> NO <sub>3</sub> S	183.18	228.8–229.7	0.3 Na salt: 83	5	5
Stevioside	C <sub>38</sub> H <sub>60</sub> O <sub>18</sub>	804.9	198	0.125	Not evaluated	Not acceptable
Sucralose	C <sub>12</sub> H <sub>19</sub> O <sub>8</sub> Cl <sub>3</sub>	397.63	125	25.7	15	15
Thaumatococin	–	~ 22 000	–	60	Not specified	Unlimited

A–A salt, aspartame–acesulfame; NHDC, neohesperidine dihydrochalcone; m.p., melting point; ADI, acceptable daily intake; JECFA, Joint Expert Committee on Food Additives of the Agriculture Organization/World Health Organization; SCF, Scientific Food Committee of the European Community.

Directive expected to be adopted during 2004. It is approved in the USA, Canada, UK, Mexico, Russia, and China.

### Cyclamate

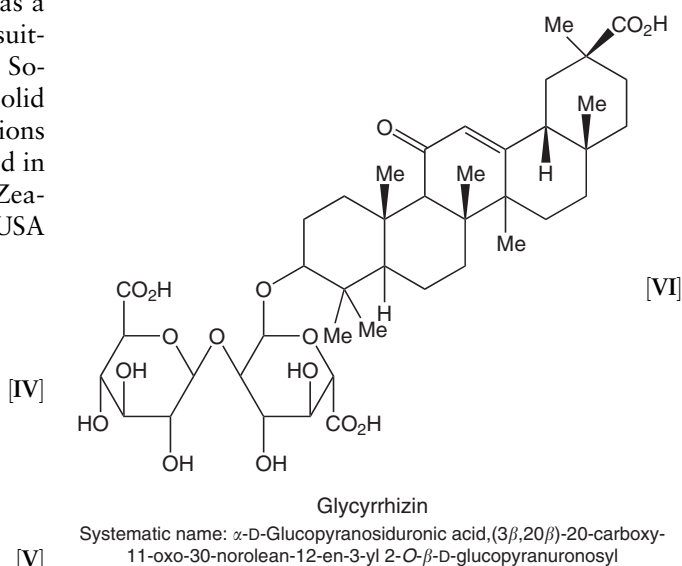
Cyclamate [IV] is generally used in the form of a sodium salt [V] because it is more soluble in water than the free acid. The calcium salt is also used as a sweetener, but, for some applications, it is not suitable as it can cause gelation and precipitation. Sodium cyclamate exhibits good stability in the solid form and is also stable in soft drink formulations within the pH range 2–10. Cyclamate is permitted in several countries (EU, Australia, Canada, New Zealand, etc.). However, it has been banned in the USA after controversial toxicity studies.



### Glycyrrhizin

Glycyrrhizin [VI] is a terpenoid glycoside and is isolated from the liquorice root plant *Glycyrrhiza*

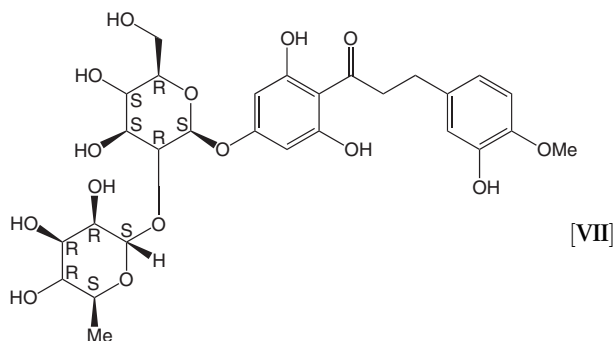
*glabra* L. Glycyrrhizin as the ammonium salt is soluble in both hot and cold water and is stable in its dry form. Glycyrrhizin is used in Japan and in other countries as sweetening agent. In the USA, it is approved for use as a flavor and flavor enhancer.



### Neohesperidine dihydrochalcone

Neohesperidine dihydrochalcone [VII] is a semisynthetic sweetener prepared from neohesperidin or naringin, two flavanones extracted from citrus peel. In aqueous solutions, neohesperidine dihydrochalcone is stable in the pH range 2.5–3.5. Neohesperidine

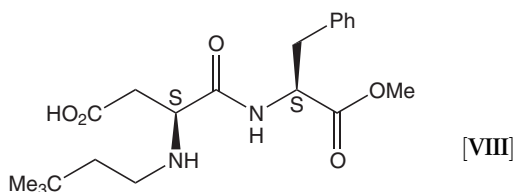
dihydrochalcone is currently allowed for many applications within the EU. In the USA, it is approved for flavoring food products.



Neohesperidine dihydrochalcone  
Systematic name: 3,5-dihydroxy-4-(3-hydroxy-4-methoxy-hydrocinnamoyl)phenyl-2-O-(6-deoxy- $\alpha$ -L-mannopyranosyl)- $\beta$ -D-glucopyranoside

### Neotame

Neotame [VIII] is a derivative of the dipeptide composed of the amino acids aspartic acid and phenylalanine. The optimum pH for maximum stability is  $\sim 4.5$ . Neotame has been approved in the USA, Australia, and New Zealand.

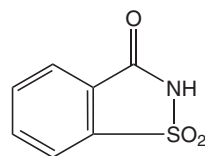


Neotame  
Systematic name: L-phenylalanine, N-(3,3-dimethylbutyl)-L- $\alpha$ -aspartyl-2-methyl ester

### Saccharin

This is the oldest high-intensity sweetener. It is commercially available in three forms: acid saccharin [IX], sodium saccharin [X], and calcium saccharin. Sodium saccharin is the most commonly used form because of its high solubility and stability. Saccharin and its salts in their solid form show good stability under conditions present in soft drinks. However, at low pH they can slowly hydrolyze to 2-sulfobenzoic acid and 2-sulfamoylbenzoic acid. Saccharin continues to be used in food and drink formulations in at least 90 countries despite controversy over its safety. Many studies have shown that there is no significant risk of cancer in humans associated with consumption of large quantities of saccharin. However, in the USA, an accompanying warning label was required until 2000. In 2000, after more than 20 years of scientific studies and further research, legislation was

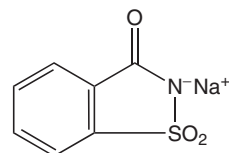
passed giving saccharin a clean bill of health and the warning label was allowed to be removed.



[IX]

Saccharin

Systematic name: 1,2-benzisothiazol-3(2H)-one-1,1-dioxide



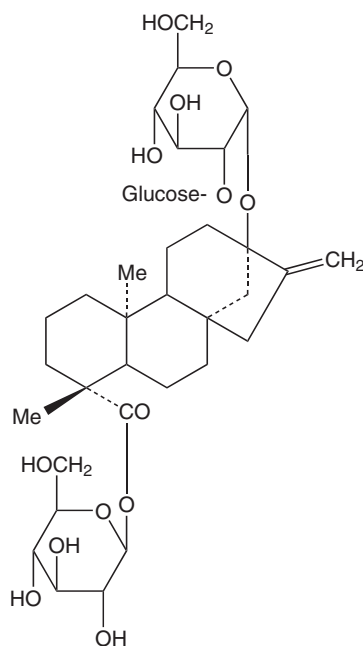
[X]

Sodium saccharin

Systematic name: 1,2-benzisothiazol-3(2H)-one-1,1-dioxide, sodium salt

### Stevioside

Stevioside or stevia [XI] is the name given to a group of sweet diterpene glycosides extracted from the leaves of *Stevia Rebaudiana* plant (native of South America). Steviosides show good stability in the solid form. They are also quite stable in acidic condition beverages at 22°C. Steviosides are approved for food use in several South American and Asian countries, but lack approval in Europe and North America.



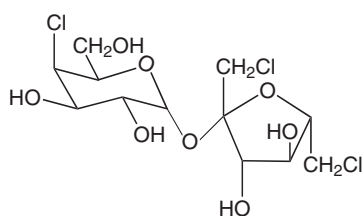
[XI]

Stevioside

Systematic name: 13-[(2-O- $\beta$ -D-glucopyranosyl- $\alpha$ -D-glucopyranosyl)oxy]-kaur-16-en-18-oic acid-4 $\alpha$ - $\beta$ -D-glucopyranosyl ester

## Sucralose

Sucralose [XII] is the common name for a sweetener derived from ordinary sugar through a multistep patented manufacturing process that selectively substitutes three atoms of chlorine for three hydroxyl groups on the sugar molecule. One advantage of sucralose for food and beverage manufacturers and consumers is its exceptional stability. Under forcing conditions in acidic solution it slowly hydrolyzes to its constituent chlorinated monosaccharides 4-chloro-4-deoxy-D-galactose and 1,6-dichloro-1,6-dideoxy-D-fructose. Following a lengthy safety evaluation it has been approved in the USA and in more than 35 countries around the world. In the EU, it is part of the Proposed Amendment to the Sweeteners Directive expected to be adopted during 2004.



[XII]

Sucralose

Systematic name: 1,6-dichloro-1,6-dideoxy- $\beta$ -D-fructofuranosyl  
4-chloro-4-deoxy- $\alpha$ -D-galactopyranoside

## Thaumatococcus

Thaumatococcus is a group of intensely sweet basic proteins isolated from the fruit of *Thaumatococcus danielli* (West African Katemfe fruit). It consists essentially of the proteins Thaumatococcus I and Thaumatococcus II. Thaumatococcus is a taste-modifying protein that functions as natural sweetener or flavor enhancer. Thaumatococcus is stable in aqueous solutions between pH 2.0 and 10 at room temperature. As occurs with aspartame it is nutritive, containing 4 kcal g<sup>-1</sup>, but due to its intense sweetness, the amounts used are small enough for thaumatococcus to be considered and classified as a nonnutritive sweetener. Thaumatococcus is approved for a number of uses in UK, Japan, Australia, the EU, and in many other countries. In the USA, it is approved as a flavor enhancer.

## Analytical Methodology

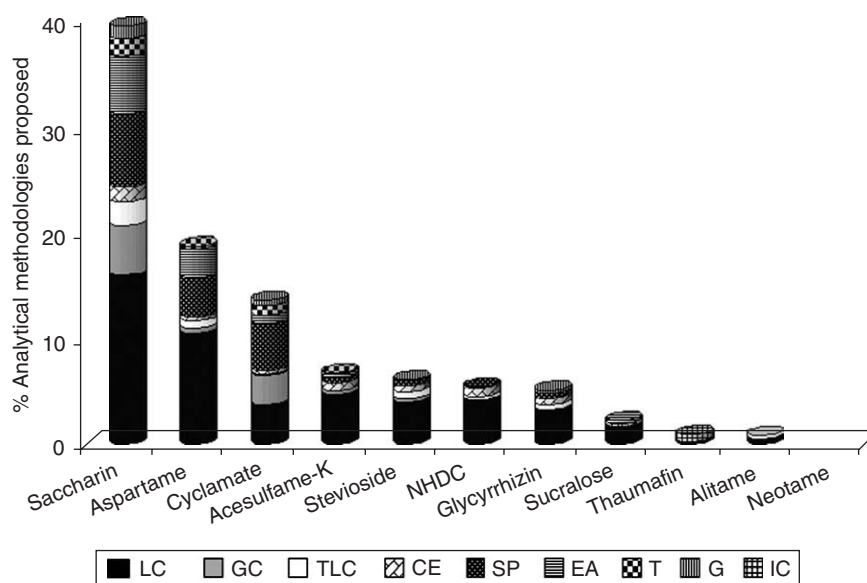
Nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) are useful techniques for structural elucidation of unknown compounds but the results obtained are difficult to quantify. Liquid chromatography (LC) methods have been used extensively for the determination of highly intense sweeteners because in many instances the sample

matrices from which they are to be determined may be complex. In addition, a sweetener may be used in combination with other sweetener(s). Nevertheless, in recent years, capillary electrophoresis methods were developed that compete successfully with LC methodologies. Figure 2 shows the most common techniques used for the determination of high-intensity sweeteners, and it can be seen that LC is the most widely used technique for all, except for cyclamate. There are no methodologies proposed for the determination of neotame and only a few methods for alitame and thaumatococcus. Most methods for the determination of thaumatococcus involve immunochemical assays (IC) and measurement in an enzyme-linked immunosorbent assay reader. On the other hand, the largest number of methodologies has been proposed for the determination of saccharin because it is the oldest known and used high-intensity sweetener. Automated continuous determinations by flow injection analysis (FIA) have been developed for acesulfame-K, aspartame, cyclamate, and saccharin; generally, these procedures involve spectrophotometric and electro-analytical detections.

**Sample preparation** Sample preparation for the determination of high-intensity sweeteners is relatively simple. Carbonated soft drinks are degassed prior to analysis. Liquid beverages and tabletop sweeteners are diluted or dissolved in water. Sweeteners in complex foods are extracted with water or an appropriate solvent. Then, the extract can be clarified, centrifuged, or cleaned by using solid-phase extraction techniques.

**Liquid chromatography** LC using amino, ion-exchange, or reversed-phase columns allows the separation of a sweetener from other sweeteners and additives present in foods in a single chromatographic run. The mobile phase can be methanol-water, methanol-acetic acid, methanol-phosphate buffer, methanol-ammonium citrate, or acetonitrile-phosphate buffer, when reversed-phase columns (RP-C18) are used. Methanol-phosphoric acid and Na<sub>2</sub>CO<sub>3</sub> or NaOH solutions were used with amino and ion-exchange columns, respectively. Acesulfame-K, alitame, aspartame, glycyrrhizin, neohesperidine dihydrochalcone, saccharin, and stevioside can be determined by ultraviolet (UV) absorbance (192–282 nm), and by amperometric or conductimetric detection in the case of ion-chromatographic procedures. Aspartame can be detected by fluorimetry ( $\lambda_{\text{EX}} = 205$  nm,  $\lambda_{\text{EM}} = 284$  nm). Cyclamate and sucralose are poorly detected by UV absorbance due to the absence of a chromophore. These sweeteners can be detected by UV absorbance if converted





**Figure 2** Analytical methodologies for the determination of high-intensity sweeteners. NHDC, neohesperidine dihydrochalcone; LC, liquid chromatography; GC, gas chromatography; TLC, thin-layer chromatography; CE, capillary electrophoresis; SP, spectrophotometric; EA, electroanalytical; T, titrimetry; G, gravimetry; IC, immunochemical. Source of information: *Chemical Abstracts* (until January 2004).

into a derivative possessing strong UV absorption: sucralose by treatment with *p*-nitrobenzoyl chloride, and cyclamate by conversion to dichlorohexylamine for UV detection or to a fluorescence derivative for fluorimetric detection. An alternative for the determination of these sweeteners is the postcolumn ion-pair extraction where the eluted sweetener is mixed with an appropriate dye (methyl violet or crystal violet) being detected by visible absorption. Furthermore, sucralose and cyclamate can be detected directly by refractive index.

**Gas chromatography** Acesulfame-K, aspartame, cyclamate, saccharin, and stevioside are determined by gas chromatography, but the main drawback of this technique is that a derivatization is required. Acesulfame-K is methylated with ethereal diazomethane, aspartame is converted into its *N*-(2-methylpropoxycarbonyl) methyl ester derivative, menthol and isobutyl chloroformate are used to convert aspartame to 3-[(isobutoxycarbonyl)amino]-4-[[ $\alpha$ -(methoxycarbonyl)phenethyl]amino]-4-oxobutyric acid, cyclamate is determined as cyclohexene resulting from the reaction with nitrite, saccharin is converted to *N*-methylsaccharin, and stevioside is hydrolyzed. Detection is carried out utilizing flame-ionization, flame-photometric electron-capture detectors or nitrogen-phosphorus detection.

**Thin-layer chromatography and paper chromatography** Qualitative, semiquantitative, and quantitative

methodologies have been described for alitame, aspartame, cyclamate, glycyrrhizin, neohesperidine dihydrochalcone, saccharin, stevioside, and sucralose. Quantification is achieved using scanning densitometry.

**Capillary electrophoresis** Capillary electrophoresis techniques, such as capillary zone electrophoresis and micellar electrokinetic capillary chromatography have been used to analyze high-intensity sweeteners in foods. These methods are rapidly gaining acceptance for the determination of sweeteners because they are comparable in resolution and precision to LC, but are faster and less expensive to operate. Aspartame, saccharin, acesulfame-K, alitame, and the other food additives are well separated in less than 12 min using an uncoated fused-silica capillary column with a buffer consisting of sodium deoxycholate, potassium dihydrogenorthophosphate, and sodium borate operating at 20 kV. In the micellar electrokinetic chromatographic mode, carbonate buffer at pH 9.5 is used as the aqueous phase and sodium dodecyl sulfate is used as the micellar phase. The determination can be performed by direct or indirect (cyclamate) UV detection, or by potentiometric detection with coated-wire ion-selective electrodes.

**Electroanalytical techniques** Polarographic procedures are described for acesulfame-K, cyclamate, and saccharin. Cyclamate was decomposed by heating with sodium nitrite and the sulfate liberated is

precipitated as lead sulfate, and the excess of  $\text{Pb}^{2+}$  is the specie measured. Sensors including biosensors such as enzyme electrodes or filter-supported bilayer lipid membranes are proposed for acesulfame-K, aspartame, cyclamate, saccharin, and sucralose. A fluoride-selective electrode is used for the kinetic potentiometric monitoring of the reaction between 2,4-dinitrofluorobenzene and aspartame. Ion-selective electrodes based on ion associates, liquid membrane electrodes based on crystal violet and brilliant green, and silver electrodes are used for the potentiometric titration of saccharin.

**Spectrophotometric techniques** Extractive spectrophotometric techniques using colorimetric reagents (oxazine dye, Sevron blue 5G for acesulfame-K; *p*-dimethylaminobenzaldehyde, 1,4-benzoquinone, ninhydrin for aspartame; picryl chloride, *p*-quinone for cyclamate; vanillin for glycyrrhizin; Nile Blue, Azure A, B, or C, Sevron blue 5G, Brilliant cresyl blue for saccharin; and anthrone for steviosides) are used for the direct determination of the high-intensity sweeteners. Indirect spectrophotometric methodologies are also proposed. UV absorbance detection is possible for aspartame, cyclamate, glycyrrhizin, and saccharin. Nevertheless, as cyclamates are not readily detected by spectroscopic techniques, a chemical derivatization is performed; generally to convert it to cyclohexylamine. Fluorimetric determinations are suggested for aspartame, where its reaction product with fluorecamine is detected, and for saccharin since this forms a fluorescent complex with sodium carbonate. Flame atomic absorption spectrometry can be used for the indirect determination of saccharin and cyclamate.

**Titrimetric methods** Titrimetric assays have been developed for acesulfame-K (titrated with sodium methoxide in benzene), aspartame, sodium cyclamate, and sodium saccharin (titrated with perchloric acid), and for saccharin (acid form) with potassium hydroxide as titrant. Precipitation, chelatometric, and redox titrations are proposed for the determination of cyclamate. The oldest methods for saccharin involve its determination by means of a Kjeldahl procedure.

**Gravimetric methods** Cyclamate, saccharin, glycyrrhizin, and stevioside can be determined by gravimetric procedures. Those for cyclamate and saccharin are included in the method book of the Association of Official Analytical Chemists (AOAC).

## Bulk Sweeteners

Bulk sweeteners, defined as those delivered, in solid or liquid form, for use in sweeteners per se or in foods in quantities greater than 22.5 kg, are disaccharides and monosaccharides of plant origin. Sucrose from sugarcane and sugar beet and starch-derived glucose and fructose from maize (corn), potato, wheat, and cassava are the major sweeteners sold in bulk to the food and beverage manufacturing industry, or packers of small containers for retail sale. Relative sweetness of bulk sweeteners and their caloric values are shown in Figure 3.

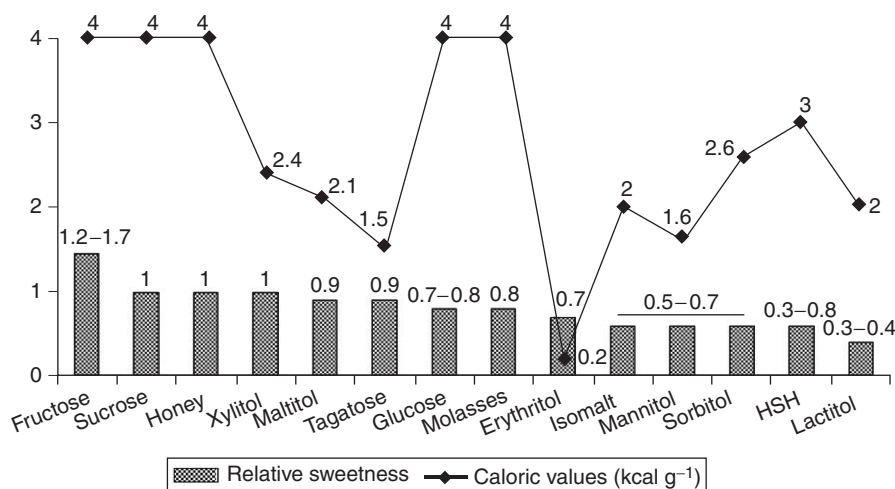
### Caloric Sweeteners

**Sucrose** Sucrose is composed of one molecule of glucose and one molecule of fructose,  $\alpha$ -D-glucopyranosyl-D-fructofuranoside. It is the traditional sweetener, table sugar. It is highly water soluble and is present in most fruits, some root vegetables, many trees, and grasses. It is more concentrated in sugarcane, *Saccharum officinarum*, and the sugar beet, *Beta vulgaris*, the latter in temperate zones. From either crop, juice is extracted, purified, and concentrated to syrup. Sucrose is crystallized from the syrup by serial crystallization; the residual viscous syrup is molasses. From both crops comes a range of white sugar products: solid granulated, powdered, cubes, liquid sucrose, and invert syrups (blends of sucrose, glucose, and fructose). Primarily from sugarcane come brown sugars, edible molasses, golden syrups, and dark cane syrups.

**Molasses** Molasses, the residual material from sugarcane and sugar beet processing, is traded in bulk, primarily for animal feed where it is mixed with fibrous residues, and serves to increase caloric value and palatability. It is not sold as a sweetener, although it is blended with other sweeteners to make special syrups.

**Honey and maple syrup** Honey, produced by honeybees (*Apis mellifera* and *Apis dorsata*) is a liquid product, ~80% solids, containing a mixture of simple carbohydrates: 25–45% fructose, 25–45% glucose, 2–12% maltose, and 0.5–3% sucrose, with traces of many other sugars depending on the bees' floral source.

Maple syrup and sugar, made from sap of the sugar maple tree (*Acer saccharum*) through concentration and crystallization, are, like honey, mixtures of simple sugars. Freshly made maple syrup, generally ~65% solids by weight, contains 50–63% sucrose and 0–8% combined glucose and fructose,



**Figure 3** Relative sweetness for bulk sweeteners and their caloric values. HSH, hydrogenated starch hydrolyzates.

but no maltose. As maple syrup is stored, the sucrose will invert to glucose and fructose.

### Starch-Derived Sweeteners

**Glucose** Glucose, commercial name dextrose, in the aldohexose form  $\alpha$ -D-glucose ( $C_6H_{12}O_6$ ), is the major product from starch hydrolyzed by acid and/or enzymes. The major starch source in the USA and Japan is corn (*Zea mays*) and in Europe, wheat and potato. There is some starch and starch hydrolyzate production from cassava in the tropics. Glucose is sold as anhydrous dextrose; more commonly as dextrose monohydrate, as glucose syrup or corn syrup.

**Fructose** Fructose, in the ketohexose form  $\beta$ -D-fructose ( $C_6H_{12}O_6$ ), is produced from glucose by an isomerase enzyme (glucose–fructose isomerase), which converts glucose to fructose, and subsequent enrichment of the fructose fraction (equilibrium conversion is  $\sim 50\%$ ), or isolation of fructose and crystallization. Products are high-fructose corn syrup, the most widely used monosaccharide sweetener, at 42, 55, and 90% fructose (with glucose, the other major component) and crystalline fructose.

### Low-Caloric Sweeteners

**Sugar alcohols, polyhydric alcohols, or hydrogenated sugars** Sugar alcohols or polyols differ from sugars in that the aldehyde or ketone function of the sugar molecule is reduced to an alcohol. They can also be categorized as sugar substitutes because they can replace sugar sweeteners. Most common as bulk sweeteners are sorbitol, mannitol, xylitol, and erythritol

(monosaccharides); isomalt, maltitol, and lactitol (disaccharides); and hydrogenated starch hydrolyzates (HSH, a mixture of sugar alcohols). Many sugar alcohols are found in nature, but it is not commercially feasible to isolate and concentrate them from their sources. Sorbitol, mannitol, maltitol, and the HSH are produced by enzymatic hydrolysis of a starch. Lactitol, xylitol, and isomalt are produced in a similar manner, except that they are not derived from starch. Erythritol has been commercially produced by a fermentation process. All retain sweetness through heating and their bulking properties are similar to those of sucrose. All polyols are absorbed slowly and incompletely from the intestine by passive diffusion. Therefore, these sweeteners provide low energy and offer potential health benefits (e.g., reduced glycemic response and reduced dental caries risk). All these sweeteners except mannitol do not represent a hazard to health and the Joint Expert Committee on Food Additives of the Agriculture Organization/World Health Organization (JECFA) deemed it not necessary to assign a numerical value for ADI, but instead assigned the most favorable term ‘not specified’. Furthermore, in the USA they are considered Generally Recognized as Safe. Nevertheless, an excess consumption of mannitol may have a laxative effect, and for this reason, JECFA has allocated a temporary ADI of 50 mg per kg.

**Tagatose** This is a new low-calorie sweetener. It is a bulk sweetener, as it is used where the bulk (mass) of sugar is important to the final product qualities. It is a monosaccharide sugar derived from lactose. Therefore, tagatose is a naturally occurring sugar derived from dairy whey.

### Analytical Methodology

Analysis of bulk sweeteners falls into two categories: (1) chromatographic methods, used in starch hydrolyzate production and for most sweetener-containing products, and (2) polarimetric and wet chemical methods used in sucrose production. There are some UV-visible spectrophotometric methods for sugar analysis; these are generally used in clinical or biological sugar analysis, and not for bulk sweeteners, and so will not be further considered here. FIA methodologies have been developed for a large number of methods for the determination of bulk sweeteners, above all for reducing sugars. The most recent methodologies involve automation of enzymatic determinations and biosensors. Sample preparation procedures are shown in Table 3.

**Polarimetric methods** These methods generally determine some overall feature of the bulk sweeteners such as total carbohydrates. Polarimetry, using a flow-through polarimeter at 589 nm, is the general method of sucrose analysis in bulk raw and white sugars. Traditional polarimeters are being replaced by polarimeters using light of longer wavelengths,  $\lambda = 880$  nm, which can be used for monitoring

a colored (though not a turbid) solution, and so decreases solid waste generated in the clarification step required by traditional polarization.

**Titrimetric and other methods** Titration with a copper(II) salt (Lane–Eynon titration) is the standard method for the determination of reducing sugars (glucose and fructose) in bulk raw and white sugars.

Moisture in solid sugars is determined generally by oven drying; in liquid products, by Karl Fischer titration. Inorganic content is determined by either conductivity in solution, or sulfated ash gravimetric procedures.

In syrups, the solids content (Brix, or refractometric dry solids (RDS)) is determined by refractive index measurement. Tables correlating refractive index with sucrose and invert content are published in the International Commission for Uniform Methods of Sugar Analysis and AOAC methods books.

**Gas chromatography** This was the first quantitative chromatographic system for sugars analysis, but it has been replaced by liquid chromatographic methods, which are simpler, less expensive, and require no derivatization, except in the fermentation industries, where the ease of analysis of alcohols and other fermentation products along with starting (glucose,

**Table 3** Sample preparation procedures for analysis of bulk sweeteners

<i>Technique</i>	<i>Sample preparation procedures</i>
Polarimetry (589 nm)	White sugar: do not require clarification Raw sugar, brown and yellow sugars: addition of lead subacetate, along with a diatomaceous earth filter aid, with filtration to obtain a clear solution of little color. Subacetate can be replaced by an aluminum salt or ultrafiltration, but for raw cane sugar these reagents do not provide satisfactory clarification
Polarimetry ( $\geq 880$ nm)	Addition of diatomaceous earth before filtration
Titrimetry	Samples other than molasses require no pretreatment
GC	Sucrose, glucose, fructose, syrups, and sugar alcohols can all be analyzed with pretreatment derivatization of the sugars as either aldonitrile or trimethylsilyl derivatives Simple sugars and sweetened beverages: generally require only filtration through at least a 0.5 $\mu$ m filter to remove suspended solids
LC	Fats and proteins present in food must be removed. Fats by extraction with an organic solvent (petroleum ether) and extraction of sugars with water or water/ethanol mixtures; proteins by precipitation with a solution of zinc acetate–potassium hexacyanoferrate(II)–water
Enzymatic methods	Extraction with cartridges of an appropriate adsorbent For methodologies supplied in kit form: filtration of the dissolved sample For biosensors containing immobilized enzymes: generally requires only filtration. Time may also be required for treatment with enzymes other than the detecting enzyme, e.g., mutarotase treatment before glucose oxidase addition

GC, gas chromatography; LC, liquid chromatography.

sucrose, molasses) and intermediate materials has helped in the continuation of the use of this technique.

**Liquid chromatography** Four LC systems are in general use for analysis of bulk sweeteners and sugar alcohols (all systems require a sample with sugars concentration from <1 to 10% and the detection of separated peaks is usually by differential refractometry):

1. Separation on a cation-exchange column, in the calcium or sodium form, with an aqueous mobile phase. Cation-exchange columns separate by a combination of liquid exchange and size exclusion; higher molecular weight sugars elute earlier than the smaller sugar molecules, but after ionic inorganic components, because the smaller molecules complex more strongly with cations on the column. Sugar alcohols may be separated on these columns, but only with pulsed amperometric detection and long separation times.
2. Separation on an amino-bonded silane column with an acetonitrile–water mobile phase. These were the first LC systems utilized for the separation of carbohydrates. The main drawback is the use of acetonitrile as solvent, but its expense and hazard can be decreased by recycling the mobile phase over a desugaring adsorbent. The column separates sugars because of their varying affinities in an aqueous/acetonitrile solvent for the bonded amino groups on the packing material. Monosaccharides (glucose, then fructose) elute first, followed by di- and trisaccharides. This system is operated at room temperature. Analysis for trisaccharides or sugar alcohols can add 40 min to the normal 10–15 min analysis time.
3. Separation on reversed-phase columns with an aqueous mobile phase. This rapid and inexpensive system is used where only separation of mono- and disaccharides is required without further separation within either group. The nonpolar column packing is silica particles coated with octadecyl (C-18) groups. Separation occurs in order of decreasing polarity and, therefore, for bulk sweeteners, in order of molecular weight.
4. Separation on anion-exchange columns with an alkaline aqueous mobile phase. This system for sugar analysis, originally called ion chromatography, is now, for carbohydrates, more selectively called high-performance anion-exchange-pulsed amperometric detection. It has gained use in the general analysis of bulk sweeteners because of its versatility and selectivity. It is the preferred system for analysis of low levels of bulk sweeteners.

The selectivity and retention time may vary by changing the pH and ionic strength of the mobile phase. The responses of the detector, a gold electrode, may be set at specific potentials to analyze for specific carbohydrates. Sugar alcohols can be analyzed along with sugars in a single analysis; isocratic elution is usually satisfactory for food and beverage products, but gradient elution is available for more complex mixtures.

**Enzymatic methods** Enzymatic methods are available for the analysis of bulk sweeteners in food and beverages; enzyme electrodes and detection kits are available for several sweeteners (e.g., sucrose, glucose, etc.). The usage of enzyme methods is determined by conditions under which the enzyme is viable, i.e., heat, substrate concentration, water availability, and interferences. This methodology is a popular alternative to chromatographic methods due to its speed, portability, and wide range of application.

**Capillary electrophoresis** Recently, this technique has revealed its highest capability in separation of bulk sweeteners. Since bulk sweeteners lack both a charge and a strong UV chromophore, several derivatization reactions have been proposed (e.g., interaction with oxoacid or metal ions). An alternative for bulk sweeteners is their separation in a fused-silica capillary using indirect fluorescence detection or indirect UV detection. Electrokinetic micellar chromatography has extended the applicability of electrophoretic techniques even to neutral molecules such as carbohydrates.

**Stable isotope ratio analysis** Adulteration of fruit juices or honey by addition of bulk sweeteners can be detected by stable isotope ratio mass spectrometric analyses, because the naturally occurring carbon isotope ratio of  $^{14}\text{C}$  to  $^{12}\text{C}$  in honey and fruit is different from that in corn or sugarcane although not from that in sugar beet. The oxygen isotope ratio may be used for beet sugar addition.

**Near-infrared spectrometry** This is an alternative for the process analysis of bulk sweeteners, e.g., raw sugar.

*See also:* **Capillary Electrophoresis:** Food Chemistry Applications. **Chiroptical Analysis.** **Enzymes:** Immobilized Enzymes; Enzyme-Based Electrodes. **Extraction:** Solvent Extraction Principles. **Flow Injection Analysis:** Principles. **Ion-Selective Electrodes:** Food Applications. **Liquid Chromatography:** Food Applications. **Spectrophotometry:** Organic Compounds.

## Further Reading

- Abreu Gloria MB (2000) Intense sweeteners and synthetic colorants. In: Nollet LML (ed.) *Food Analysis by HPLC*, 2nd edn, pp. 523–573. New York: Dekker.
- Association of Official Analytical Chemist (AOAC). (2003) *Official Methods of Analysis*, 17th edn. Arlington: AOAC.
- Grenby TH (1996) *Advances in Sweeteners*. London: Blackie Academic and Professional.
- Hollenbeck CB and Kretchmer (1991) *Sugars and Sweeteners*. Boca Raton: CRC Press.
- International Commission for Uniform Methods of Sugar Analysis (ICUMSA). (1994) *ICUMSA Methods Book*. Port Talbot, UK: ICUMSA Publications.
- International Commission for Uniform Methods of Sugar Analysis (ICUMSA). (1998) *ICUMSA Methods Book Supplement*. Port Talbot, UK: ICUMSA Publications.
- Marie S and Piggot JR (1991) *Handbook of Sweeteners*. Glasgow: Blackie and Son.
- O'Brien N (2001) *Alternative Sweeteners*, 3rd edn. New York: Dekker.
- Prodoliet J 1996. Intense sweeteners. In: Nollet LML (ed.) *Handbook of Food Analysis*, vol. 2, pp. 1835–1865. New York: Dekker.
- Salminen S and Hallikainen A (2002) Sweeteners. In: Branen AL, Davidson PM, Salminen S, and Thorngate JH (eds.) *Food Additives*, pp. 477–500. New York: Dekker.
- Von Rymon Lipinski G (2002) Sweeteners. In: Watson D (ed.) *Food Chemical Safety, vol. 2: Additives*, pp. 228–248. Boca Raton: CRC Press.
- Von Rymon Lipinski GW (2003) Intense sweeteners: status and new developments. *International Sugar Journal* 105: 308 (see also 310–312).
- Yebra-Biurrun MC (2000) Flow injection determination of artificial sweeteners: A review. *Food Additives and Contaminants* 17: 733–738.



# T

## TEA

See **FOOD AND NUTRITIONAL ANALYSIS: Coffee, Cocoa, and Tea**

## TECHNETIUM

See **RADIOCHEMICAL METHODS: Technetium**

## TEXTILES

Contents

**Natural**

**Synthetic**

### Preliminary Note

To distinguish between natural and synthetic fiber materials, the easiest method is to conduct a burning test. If a fiber based on natural material, with the exception of asbestos, is slowly moved to a flame, it starts to burn. Most synthetic fibers melt and burn only if the fiber is held directly onto the flame. Some synthetic fibers, e.g., Aramid, do not burn and can be distinguished under an optical microscope.

### Natural

**M Halbeisen**, EMPA St Gallen, St Gallen, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Ira Block, pp. 5116–5122, © 1995, Elsevier Ltd.

### Introduction

Fibers based on natural materials are divided into two categories: natural fibers found in the plant, animal, and mineral kingdoms; and regenerated cellulosic fiber, which are manmade fibers made out of natural polymers.

Plant fibers are cellulosic, animal fibers are proteinaceous materials, and asbestos is a crystalline silica compound. Almost every fibrous material has had some use in historic or prehistoric times. The most important natural fibers used in the textile industry are: cotton, linen, wool, and silk (moth silk) (see **Table 1**).

Regenerated cellulosic fibers are made out of cellulose from wood or cotton lint. The most used fibers are viscose, acetate, cupro (very small quantity), and lyocell. These fibers are produced in various titers as filament (endless fiber) or staple fiber.

### Fibers

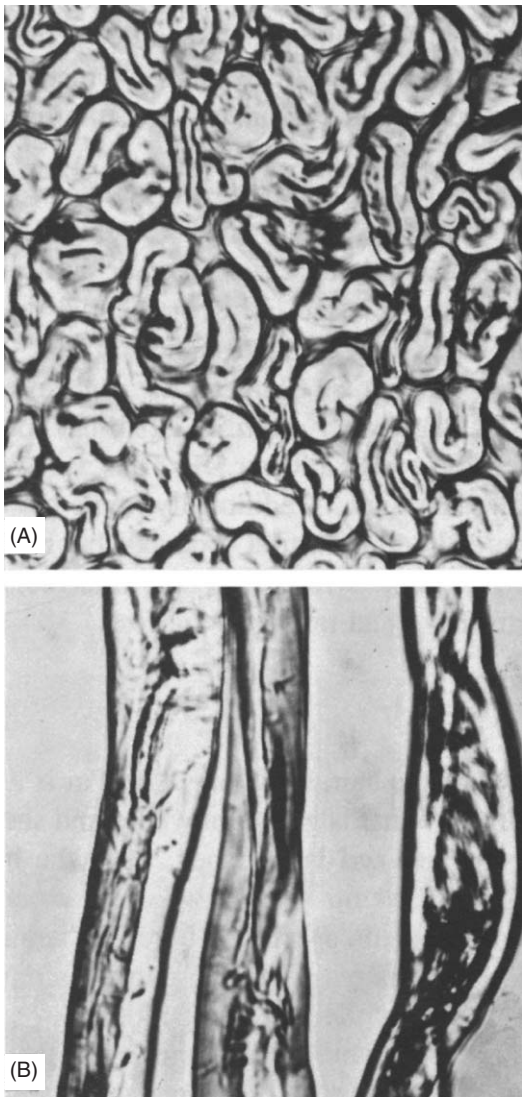
#### Cotton

This is the most abundant fiber, accounting for ~40% of the world's production of all fibrous materials. The fiber is the seed hair of plants of the genus *Gossypium* and exists in many commercial varieties, e.g., Egyptian, Sea Island, and Uplands. Its appearance is quite characteristic; it is a flattened, twisted hollow tube, as shown in **Figure 1**. Commercial fibers range in length from 0.3 to 5.5 cm, are ~10 µm in diameter, and generally have dull white color.

The structure of the fiber comprises a thin, dense outer shell (the primary wall) surrounding many

**Table 1** The most used natural fibers in the past and present

Plant fibers				Animal fibers				Mineral fibers
Seed	Bast		Fruit	Wool	Hair		Silk	
	Stems	Leaves			Fine	Coar		
Cotton	Linen	Sisal	Coco	Sheep	Alpaca	Cattle	Moth silk	
Kapok	Hemp	Manila hemp			Lama	Horse	Spider silk	
Akon	Jute	Yukka			Vicunja	Goat	Mussel silk	
	Ramie	Banana			Guanaco			
	Kenaf				Camel			
					Angora			
					Mohair			
					Cashmere			
					Yak			



**Figure 1** Cross-section (A) and longitudinal view (B) of typical cotton fibers. (Originally published in the AATCC Technical manual, as a part of AATCC Test Method 20; reprinted with permission from the American Association of Textile Chemists and Colorists.)

concentric cellulose layers (the secondary wall) that enclose a hollow core (the lumen). The primary wall is composed of cellulose containing small amounts of various vegetable oils and waxes. The secondary wall consists of numerous chains of cellulose deposited in layers within the tube formed by the primary wall. This deposition occurs as the glucose in solution polymerizes to cellobiose with a degree of polymerization (DP) of  $\sim 3000$ . That is, on average, each cellulose chain is made up of 3000 cellobiose units. The lumen is the hollow tube that is left when all the glucose solution from which the fiber was created has been used up or dried out. The fiber is made up of over 95% cellulose.

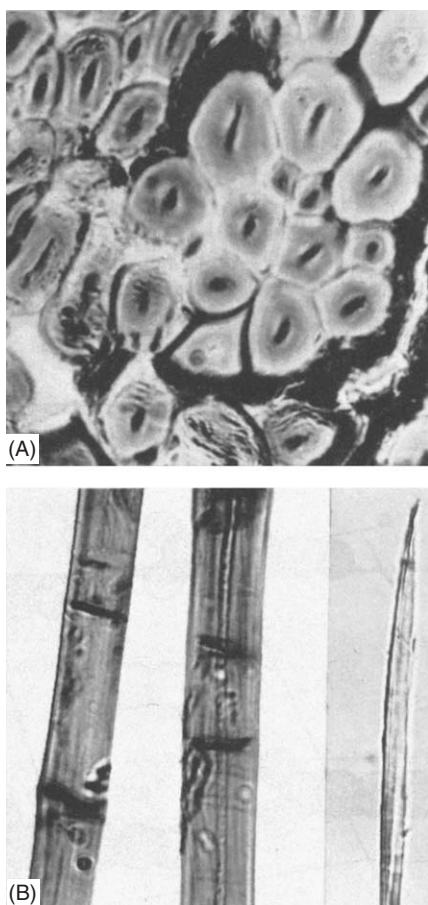
Morphologically cotton cellulose is highly crystalline. Estimations of the extent of crystallinity vary, depending upon the measurement method used, from  $\sim 75\%$  to  $95\%$ . This configuration is almost a rectangular parallelepiped containing a cellobiose unit in each corner, with a central cellobiose unit in the center, antiparallel to the other four.

Mercerized cotton has a different structure. With the use of a strong alkaline process, cotton fibers swell. The cross-section of mercerized cotton is round, like the premature cotton, and untwisted.

### Linen

Linen is the fiber taken from the stem of the flax plant, *Linum usitatissimum*. Photomicrographs of the cross-section and longitudinal view are shown in Figure 2. Commercial linen consists of bundles of individual fibers cemented together by lignin as well as vegetable waxes. The cement accounts for  $\sim 30\%$  of the weight of the fiber. The remainder is cellulose. Individual fibers range in length from  $\sim 4$  cm to 1 m and are  $\sim 20 \mu\text{m}$  in diameter. The color of linen is whitish blonde.

Individual fibers have a bamboo-like structure consisting of single cells, length ranging from  $\sim 0.3$  cm



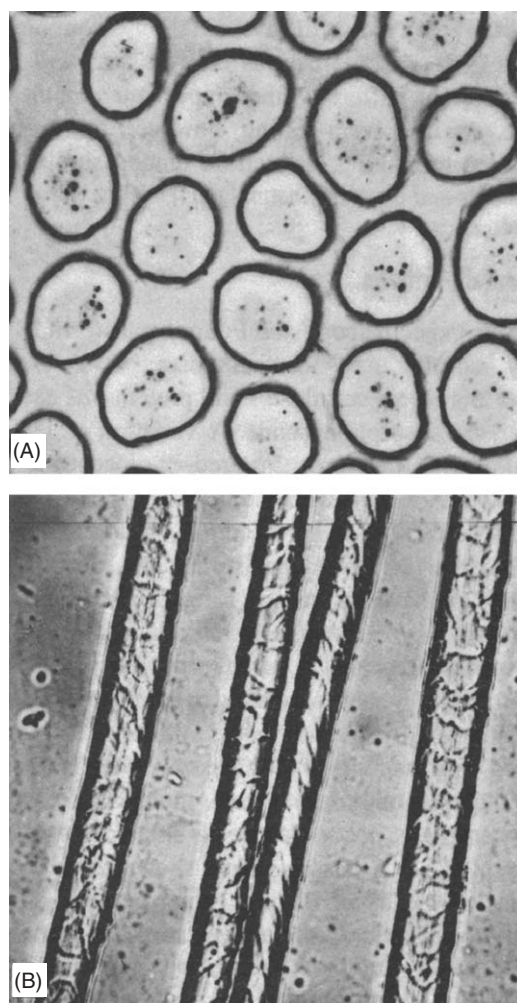
**Figure 2** Cross-section (A) and longitudinal view (B) of typical linen fibers. (Originally published in the AATCC Technical manual, as a part of AATCC Test Method 20; reprinted with permission from the American Association of Textile Chemists and Colorists.)

to as much as 5 cm, joined end to end. These cells are composed of smaller fibrils surrounding a thin lumen and give the longitudinal view of the striated appearance seen in **Figure 2**. The junctions of the cells produce a slight swelling, called nodes, that are characteristic of bast fibers. Lignin, a component of all woody plants including flax, is a complex, cross-linked, tarry substance, whose structure has not yet been clearly defined.

Morphologically, the cellulose in linen is in the Cellulose I form, with a DP of  $\sim 3000$ . The chains within the cells form an extended helix that is nearly straight and almost parallel to the fiber axis. They do not exhibit the reversals of direction found in cotton.

## Wool

Chemically and structurally, wool is typical of all animal hair and fur. Its major constituent is the



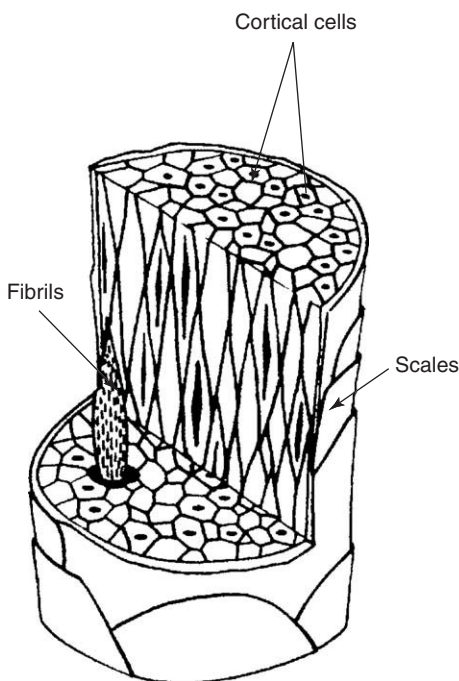
**Figure 3** Cross-section (A) and longitudinal view (B) of typical wool fibers. (Originally published in the AATCC Technical manual, as a part of AATCC Test Method 20; reprinted with permission from the American Association of Textile Chemists and Colorists.)

protein 'keratin'. Wool fibers range in length from 2 to 50 cm and from 10 to 20  $\mu\text{m}$  in diameter. The fibers of the highland sheep (e.g., merino) are highly crimped and shorter, compared to the lowland sheep (e.g., cheviot), which are low or uncrimped and longer. They may be white, brown, gray, or black. The finest wools are taken from young lambs. In its natural state, wool contains a high percentage of lanolin, a water-repellent oily material that is often removed in processing.

The cross-section and longitudinal views of wool are shown in **Figure 3**. **Figure 4** is a schematic diagram of the structure of the fiber. Enclosed within an outer scaly layer of keratin (the cuticle) are three sets of cortical cells: *para*, *meso*, and *ortho*. The cells may surround a hollow core, known as the medulla.

The *para* cells are characterized by close-packed microfibrils and a high cystine content. The *ortho*





**Figure 4** Cutaway view of the wool fiber.

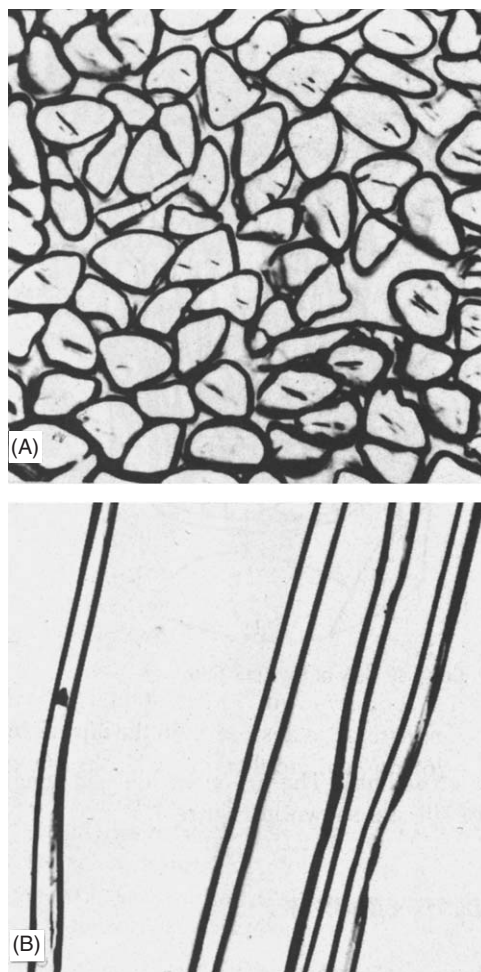
cells have their microfibrils more widely spaced, a lower cystine content, and a high concentration of glycine and tyrosine. The *meso* cells have properties in between those of the *ortho* and *para* cells. Highly crimped wools are those with a preponderance of *para* cells while those with little crimp contain mainly *meso* and *ortho* cells.

### Silk

Silk, obtained from the cocoon of the silk moth *Bombyx mori*, is the only natural filament fiber. It is composed of the protein 'fibroin'. Unlike wool, silk has an extended chain, pleated-sheet morphology, the  $\beta$ -helix. The  $\beta$ -form permits a much greater degree of intermolecular hydrogen bonding than does the  $\alpha$ -form, so that the crystallinity of silk is much higher than that of wool. Thus, silk is somewhat stronger than wool, even though it lacks cystine to provide cross-links. The cross-section and longitudinal views of silk are shown in Figure 5.

### Viscose

Viscose fibers – also known under the name rayon – are manmade cellulosic fibers. Cellulose – extracted from wood (wood pulp) – reacts with 18% caustic soda to give alkali cellulose, which is treated with carbon disulfide after the so-called preripening (breakdown of the cellulose molecules into smaller units), the white alkali cellulose changing into orange-yellow cellulose xanthogenate. Diluted



**Figure 5** Cross-section (A) and longitudinal view (B) of typical silk fibers. (Originally published in the AATCC Technical manual, as a part of AATCC Test Method 20; reprinted with permission from the American Association of Textile Chemists and Colorists.)

caustic soda dissolves the xanthogenate into a syrup mass, the spinning solution or viscose, after which this method of operation has been named. Pumped through the holes of the shower-like spinnerets (5000–250 000 holes) into a flowing spinning bath, which contains diluted sulfuric acid, sodium sulfate, and zinc sulfate, the viscose coagulates into fine white filaments of pure cellulose.

Viscose fibers have nearly the same physiological properties but poorer physical properties than cotton.

### Acetate

For the production of acetate and triacetate, high-grade cellulose – e.g., cotton lint – is converted into cellulose acetate with acetic anhydride, and dissolved in acetone. In contrast to triacetate, only 2.5 of the three hydroxyl groups of a basic cellulose unit are

esterified ( $2\frac{1}{2}$  acetate or diacetate). The viscous spinning solution is pressed through spinnerets, falling through 3–6 m high spinning ducts where the volatile solvent acetone is evaporated. The dissolved cellulose acetate is transformed into solid, filament yarns with a silk-like luster.

Acetate is thermoplastic and the softening rate has a range between 205°C and 215°C. The fibers are soluble in acetone.

### Lyocell

The production of lyocell is a new technology for producing cellulosic fibers by means of a direct solvent process. This means that a genuine solution of the cellulose in *N*-methylmorpholine-*N*-oxide (NMMO) monohydrate is produced and spun, and not a cellulose derivate as in the case of viscose. The cellulose can, however, be obtained from used textiles or by crushed wood pulp. After filtration of the honey-like spinning solution, it is spun through an air gap into diluted aqueous NMMO solution.

Lyocell has nearly the same physiological properties as cotton and the physical properties are much better than viscose.

## Fiber Identification Methods

### Ignition

Burning tests are the simplest means of determining the generic class to which an unknown fiber may belong. Protein fibers – all fibers from animals – burn slowly, with a smoky, sputtering flame and give off a characteristic odor of burning hair. Natural and regenerated cellulose fibers – with the exception of acetate – are readily ignited and burn rapidly. A glowing red cinder is obtained after burning, along with the characteristic odor of burning paper.

Acetate is not a pure cellulosic fiber, and so the ignition is completely different. They burn very quickly, first as a melted ball and later as gray ash. The odor is acidic. Mineral fibers do not ignite or burn.

### Microscopy

Having determined that a fiber is either of animal or vegetable origin, examination under a 10–20× microscope is often sufficient to identify it. If, for example, one has an animal fiber, comparison of the specimen with a known, previously mounted standard allows one to readily differentiate the various fur and hair fibers. Wool is easily distinguished from camel hair or rabbit fur by the types of scales and size of medulla. Further comparison with samples taken from different breeds of sheep will yield a more exact

identification. The use of a scanning electron microscope is also possible, especially if there are very fine details to show. Similarly, seed hair, leaf, and bast fibers also possess distinguishing characteristics. It should be recognized that this method will not permit one to identify the exact animal or plant from which a fiber may have been taken. More sophisticated instrumental methods are usually required and some are discussed below.

### Solvents

Various fibers may be identified, or separated from each other, through sequential dissolution in various solvents. As seen in Figure 6, silk is soluble in HCOOH/ZnCl<sub>2</sub>, while wool does not dissolve. For bast fibers it is possible to discern some types with phloroglucin/HCl under a microscope. Jute will stain red, while hemp is only slight pink, and ramie and flax are unaltered. With the use of Kuoxam, ramie is dissolved while flax remains undissolved. For these tests, it is absolutely necessary to use reference materials!

This technique is not capable, however, of distinguishing between different wools or between different silks.

### Stains

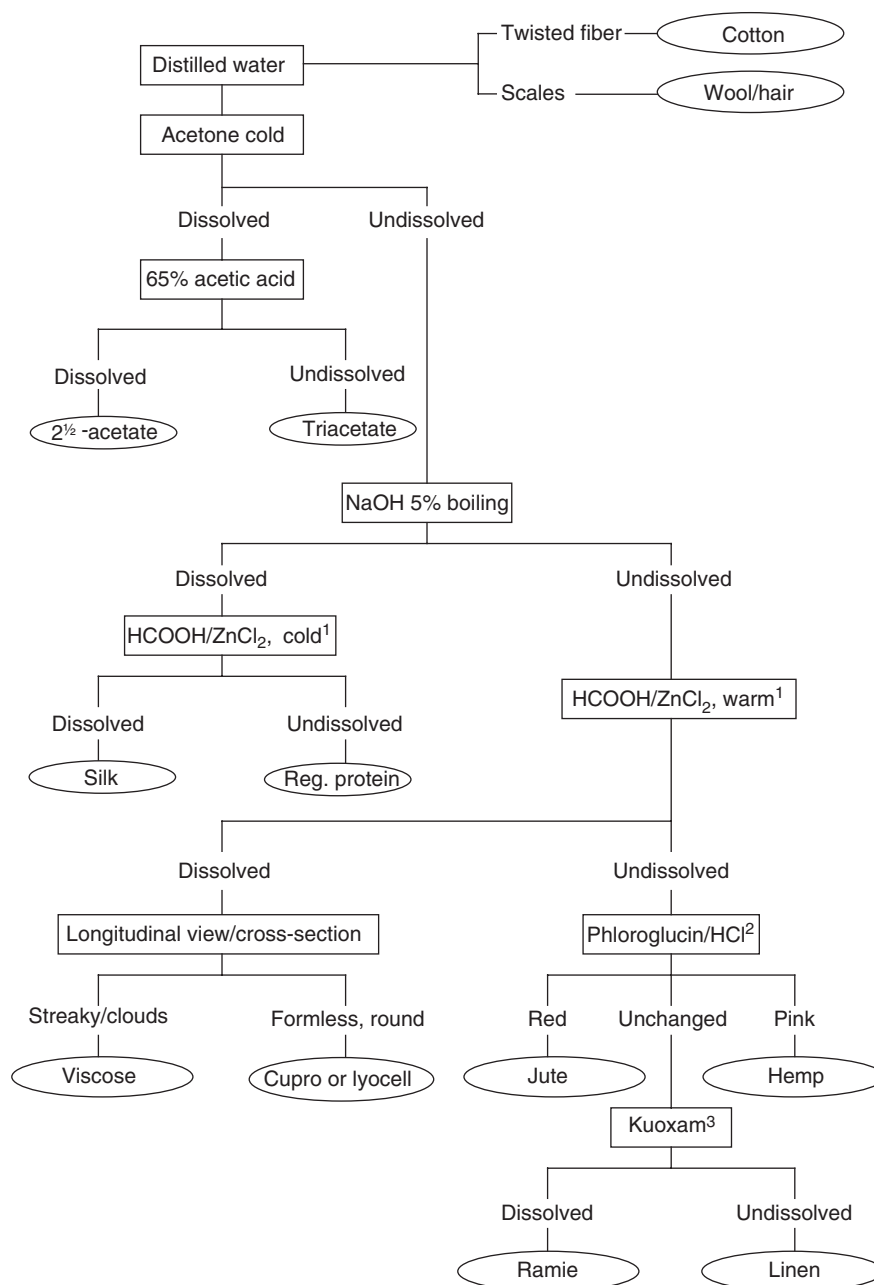
Fibers may be identified by their response to different dyes. Animal fibers readily take up acid dyes, while plant fibers are more responsive to direct dyes. Combinations of dyes that will permit identification of different fiber types are commercially available. The most common commercial solution is Neocarmin. See technical literature or the Internet for more information. As in solvent testing, however, this technique is not suitable for making distinctions among similar fibers.

### Instrumental Methods

Infrared (IR) spectroscopy and Fourier transform infrared (FTIR) spectroscopy can be used to distinguish between generic classes of fibers, i.e., cellulose fibers and proteins, as each have characteristic absorption bands. A variation of this technique, FTIR-photoacoustic spectroscopy can nowadays be used to determine the extent of cross-linking in cotton fabrics treated with polycarboxylic acids by measuring the absorption at 1725 cm<sup>-1</sup> of the ester groups formed.

Polyacrylamide-gel electrophoresis in sodium dodecyl sulfate combined with staining by Coomassie Blue could be used to distinguish among merino wool, mohair, and cashmere.

Analysis of DNA by polymerase chain reaction has also been shown to be a useful, but expensive, means of distinguishing between different hair and wool fibers.



<sup>1</sup>20 g water-free  $\text{ZnCl}_2$ , dissolve in 80 g formic acid 85%.

<sup>2</sup>10% Phloroglucin (alcoholic solution) mixed with the same quantity HCl 100%.

<sup>3</sup>8 g  $\text{CuSO}_4 \times 5\text{H}_2\text{O}$  dissolved in 40 ml distilled water. Add 10 g  $\text{NH}_4$  and fill up with distilled water to 100 ml. Then add 4 ml NaOH 20%. If cotton dissolves quickly, a lower quantity of NaOH should be used.

**Figure 6** Reaction diagram for solvent tests.

## Fiber Characterization

The properties and performance of the natural fibers are dependent upon both their macrostructure (large-scale physical features) and their microstructure (chemical composition). The major macrostructural

characteristics of importance in the use of these fibers are the length, crimp, and fineness.

### Macrostructure

Fiber length is measured by combing the fibers of a given sample, sorting these fibers according to length,



measuring the length of the individual fibers, and determining a histogram of the length distribution. Over the years, the procedure has been instrumented and image analysis is often used.

Fiber crimp (waviness) affects the ease with which fibers may be spun into yarn, as well as many of the comfort characteristics of cloth made from those yarns. Computerized image analysis techniques can be used to measure the degree of crimp of a fiber.

Fiber fineness is determined by compressing a standard weight of fiber into a known volume and measuring the rate of airflow through the compressed mass under standard conditions. The result is known as the Micronaire reading. Various instruments have been developed to measure the fineness of wool and cotton fibers.

### Microstructure

The size and arrangement of macromolecules within a fiber and the chemical structure of those molecules determine many of the physical and, of course, all its chemical properties. Molecular weight, crystallinity, substituent groups, and chemical moieties introduced as a result of processing or use are the major microstructural factors.

The physical and mechanical properties of fibers are a consequence of their macromolecules of very high molecular weight. Thus, both from commercial and scientific viewpoints, molecular weight measurements give important insights into the structure and behavior of these materials.

The molecular weight of the natural fibers is determined by standard procedures. End-group analysis is less useful for natural fibers, which do not yield unambiguous results, than for the synthetic polymers. Physical methods, such as viscometry, osmotic pressure, and light scattering, are all appropriate for the natural fibers. Gel chromatography is not generally used because it is more intricate than the other methods without yielding more significant information. The reader is referred to standard texts on polymer science for these procedures.

### Crystallinity

The natural fibers exhibit varying degrees of crystallinity. Although simple models that assume a given fraction of perfectly crystalline material existing within a matrix of perfectly amorphous material are suitable for estimating fiber mechanical properties, fiber morphology spans the full range from perfect crystal to perfect glass. The degree of crystallinity affects both mechanical and chemical properties, in that it is the crystalline regions that give a

fiber its strength, while the extent to which reactive moieties are accessible to chemical reaction is determined by the amount of amorphous material.

Crystallinity may be measured by chemical reaction, such as hydrolysis, deuteration, or saponification, or by physical methods, which include X-ray diffraction, density, IR spectroscopy, and birefringence. It should be noted that, since each of these methods measures a property that is affected by crystallinity in different ways, agreement among the different methods is rare.

### Chemical Reactivity

The chemistry of the cellulosic fibers is similar to that of the simple sugars, but more complex, as the stereochemistry of the alcohol and hemiacetal structures is affected by degree of crystallinity, e.g., the  $-OH$  groups on the carbon atoms 2, 3, and 6 have different reactivities. Reaction with mineral acids leads to cleavage of the glycosidic bond and formation of a reducing end ( $RO-CH-OH$ ).

The number of these ends, and thus the extent of degradation, is determined by reacting a fiber sample with Fehling's solution to produce copper(I) oxide. The result is known as the 'copper number'. Cellulosic fibers degraded in this manner are known as hydrocelluloses.

The major environmental effects on cellulose are oxidation of the hydroxyl groups and hydrolysis of the glycosidic bonds. Both atmospheric acidity and light are the major catalysts for weathering. Oxidation of cellulose leads to the formation of carbonyl ( $C=O$ ), carboxyl ( $COOH$ ), and shorter-lived peroxy ( $COO$ ) species. Carbonyl groups may be measured by titration with silver nitrate. Carboxyl species are determined by reaction with methylene blue dye. Peroxides, formed in greater quantity by light than by heat, are determined by measurement of the absorbance of a phenolphthalein-copper(II) sulfate solution that has reacted with a weighed amount of oxidized fiber.

Both carbonyl and carboxyl species are revealed, semiquantitatively, by the growth of a peak in the IR spectrum at  $\sim 1740\text{ cm}^{-1}$ . In addition, carbonyl groups produce an increase in absorption in the ultraviolet region at  $\sim 265\text{ nm}$  and are thought to be the chromophore that leads to the yellowing of celluloses.

In the absence of oxygen, cellulose is highly resistant to alkali. Oxidized cellulose, however, is far less resistant to alkaline hydrolysis as formation of  $C=O$  groups adjacent to the glycosidic bond greatly weakens it. Celluloses that have been oxidized, either with oxygen or more specific oxidizers such as periodate, are known as oxycelluloses leading to the destruction of the material.

The chemistry of the protein fibers is more complex. Amino acids exhibit amphoteric behavior in that they contain both a Lewis base (the amine) and a Lewis acid (the COOH). In acid media the active hydrogen exists in the form of an ammonium ion ( $-\text{NH}_3^+$ ), while in basic media the negative ion is the carboxylate ( $-\text{COO}^-$ ). In addition to the peptides of the backbone chain, the pendant groups in protein fibers provide additional reactive sites.

Acids cause hydrolysis of the main-chain peptides, while alkalis are less selective in their action. They may react with the carboxylic acids found in the aspartic acid, glutamic acid, and proline components, with the amino groups in arginine and lysine residues, or with the S–S bond in cystine, as well as with the main-chain links.

Cystine is important in the chemistry of the wool fiber, since the disulfide ( $-\text{S}-\text{S}-$ ) links form the intermolecular cross-links that give the fiber its coherence. Cystine is stable to acid, but readily attacked by dilute alkali. Mild oxidants such as hydrogen peroxide, the preferred bleach for protein fibers, will slowly convert cystine to cystic acid. Oxygen, in the presence of light also reacts with cystine, but its major effect is to attack the tryptophan residues and produce yellowing of the fibers.

**See also:** **Asbestos.** **Carbohydrates:** Dietary Fiber Measured as Nonstarch Polysaccharides in Plant Foods. **Electrophoresis:** Polyacrylamide Gels. **Forensic Sciences:** DNA Profiling; Fibers. **Fourier Transform Techniques.** **Functional Group Analysis.** **Infrared Spectroscopy:** Overview. **Liquid Chromatography:** Size-Exclusion. **Microscopy Techniques:** Light Microscopy; Scanning Electron Microscopy. **Proteins:** Overview. **Textiles:** Synthetic. **X-Ray Absorption and Diffraction:** Overview.

## Further Reading

- Agster A (1982) *Färberei- und textilchemische Untersuchungen*, 10th edn. Berlin: Springer.
- Kosowski HJ (1999) *Dictionary of Man-Made Fibers*, 1st edn. Frankfurt am Main: International Business Press.
- Latzke and Hesse (1988) *Textile Fasern – Rasterelektronenmikroskopie der Chemie- und Naturfasern*. Frankfurt am Main: Deutscher Fachverlag.
- Lewin M and Pearce EM (eds.) (1985) *Handbook of Fiber Science and Technology*, vol. IV, *Fiber Chemistry*, chs. 1–5. New York: Dekker.
- Lewin M and Pearce EM (eds.) (1985) *Identification of Textile Materials*, 7th edn. Manchester: The Textile Institute.
- Schenek A (2000) *Naturfaser Lexikon*. Frankfurt am Main: Deutscher Fachverlag.

## Synthetic

M Schmid, EMPA St Gallen, St Gallen, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Fibers, the basis for most textiles, consist of natural fibers, chemically modified natural fibers, or synthetic fibers (a description of the differences of these fiber types is given in the article).

Synthetic fibers, the subject of this article, are produced from linear synthetic polymers by different spinning processes. Depending on whether the process is based on polymer melt or polymer solution, the processes are divided into dry or wet spinning.

According to DIN 60001-T4, the synthetic fibers are divided into materials synthesized by different chain building processes: polymerization, polycondensation, and polyaddition. Polymerization is subjected to monomers containing a vinyl group (double bond) in the molecular structure. The chain reaction will be induced by radical reaction. Polycondensation

reaction needs bifunctional monomers, which are connected to a polymer chain by condensation reaction (secession of water or other small molecules). Polyaddition also needs bifunctional monomers but the formation of molecular bonding releases no side-products. The most common fiber forming polymers produced with polymerization reaction are: polyethylene, polypropylene, poly(vinyl chloride), poly(vinylidene chloride), poly(vinyl alcohol), polyacrylonitrile, polytetrafluoroethylene. Polycondensation and polyaddition leads to polyester, polyamide (PA), aramid, and polyurethane. **Figure 1** summarizes the principles of the chain forming reactions, chemical details of the concerned monomers regarding functional groups, and the resulting polymers. Besides the above-mentioned homopolymers (homopolymer = polymer derived from only one monomer type or a monomer couple), there exists a huge variety of the so-called co-polymers, which are polymers derived from two or more monomers. Additionally, two different polymer types may be extruded simultaneously in the same spinning

The chemistry of the protein fibers is more complex. Amino acids exhibit amphoteric behavior in that they contain both a Lewis base (the amine) and a Lewis acid (the COOH). In acid media the active hydrogen exists in the form of an ammonium ion ( $-\text{NH}_3^+$ ), while in basic media the negative ion is the carboxylate ( $-\text{COO}^-$ ). In addition to the peptides of the backbone chain, the pendant groups in protein fibers provide additional reactive sites.

Acids cause hydrolysis of the main-chain peptides, while alkalis are less selective in their action. They may react with the carboxylic acids found in the aspartic acid, glutamic acid, and proline components, with the amino groups in arginine and lysine residues, or with the S–S bond in cystine, as well as with the main-chain links.

Cystine is important in the chemistry of the wool fiber, since the disulfide ( $-\text{S}-\text{S}-$ ) links form the intermolecular cross-links that give the fiber its coherence. Cystine is stable to acid, but readily attacked by dilute alkali. Mild oxidants such as hydrogen peroxide, the preferred bleach for protein fibers, will slowly convert cystine to cystic acid. Oxygen, in the presence of light also reacts with cystine, but its major effect is to attack the tryptophan residues and produce yellowing of the fibers.

**See also:** **Asbestos.** **Carbohydrates:** Dietary Fiber Measured as Nonstarch Polysaccharides in Plant Foods. **Electrophoresis:** Polyacrylamide Gels. **Forensic Sciences:** DNA Profiling; Fibers. **Fourier Transform Techniques.** **Functional Group Analysis.** **Infrared Spectroscopy:** Overview. **Liquid Chromatography:** Size-Exclusion. **Microscopy Techniques:** Light Microscopy; Scanning Electron Microscopy. **Proteins:** Overview. **Textiles:** Synthetic. **X-Ray Absorption and Diffraction:** Overview.

## Further Reading

- Agster A (1982) *Färberei- und textilchemische Untersuchungen*, 10th edn. Berlin: Springer.
- Kosowski HJ (1999) *Dictionary of Man-Made Fibers*, 1st edn. Frankfurt am Main: International Business Press.
- Latzke and Hesse (1988) *Textile Fasern – Rasterelektronenmikroskopie der Chemie- und Naturfasern*. Frankfurt am Main: Deutscher Fachverlag.
- Lewin M and Pearce EM (eds.) (1985) *Handbook of Fiber Science and Technology*, vol. IV, *Fiber Chemistry*, chs. 1–5. New York: Dekker.
- Lewin M and Pearce EM (eds.) (1985) *Identification of Textile Materials*, 7th edn. Manchester: The Textile Institute.
- Schenek A (2000) *Naturfaser Lexikon*. Frankfurt am Main: Deutscher Fachverlag.

## Synthetic

M Schmid, EMPA St Gallen, St Gallen, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

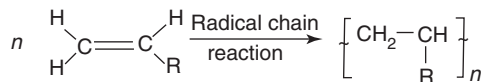
Fibers, the basis for most textiles, consist of natural fibers, chemically modified natural fibers, or synthetic fibers (a description of the differences of these fiber types is given in the article).

Synthetic fibers, the subject of this article, are produced from linear synthetic polymers by different spinning processes. Depending on whether the process is based on polymer melt or polymer solution, the processes are divided into dry or wet spinning.

According to DIN 60001-T4, the synthetic fibers are divided into materials synthesized by different chain building processes: polymerization, polycondensation, and polyaddition. Polymerization is subjected to monomers containing a vinyl group (double bond) in the molecular structure. The chain reaction will be induced by radical reaction. Polycondensation

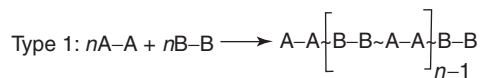
reaction needs bifunctional monomers, which are connected to a polymer chain by condensation reaction (secession of water or other small molecules). Polyaddition also needs bifunctional monomers but the formation of molecular bonding releases no side-products. The most common fiber forming polymers produced with polymerization reaction are: polyethylene, polypropylene, poly(vinyl chloride), poly(vinylidene chloride), poly(vinyl alcohol), polyacrylonitrile, polytetrafluoroethylene. Polycondensation and polyaddition leads to polyester, polyamide (PA), aramid, and polyurethane. **Figure 1** summarizes the principles of the chain forming reactions, chemical details of the concerned monomers regarding functional groups, and the resulting polymers. Besides the above-mentioned homopolymers (homopolymer = polymer derived from only one monomer type or a monomer couple), there exists a huge variety of the so-called co-polymers, which are polymers derived from two or more monomers. Additionally, two different polymer types may be extruded simultaneously in the same spinning

## (I) Polymerization



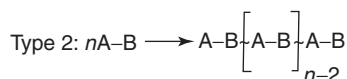
R = H: Poly(ethylene) terephthalate  
 R = CH<sub>3</sub>: Polypropylene  
 R = Cl: Poly(vinyl chloride)  
 R = OH: Poly(vinyl alcohol)  
 R = CN: Polyacrylonitrile

## (II) Polycondensation/polyaddition



A = COOH, B = OH:  
 $\longrightarrow$  Polyester

A = COOH, B = NH<sub>2</sub>:  
 $\longrightarrow$  Polyamide



A = NCO, B = OH:  
 $\longrightarrow$  Polyurethane

For example:

A-A: Terephthalic acid, B-B: ethyleneglycol  $\longrightarrow$  Poly(ethylene terephthalate) (PET)

A-A: Adipic acid, B-B: hexamethylenediamine  $\longrightarrow$  Polyamide 66 (PA 66)

A-B:  $\omega$ -Aminohexanoic acid  $\longrightarrow$  Polyamide 6 (PA 6)

**Figure 1** Principle of chain forming reactions and fiber forming synthetic polymers.

**Table 1** Properties of fibers derived from vinyl polymers

	<i>Polyolefins</i>	<i>Acrylics</i>	<i>Chlorofibers</i>	<i>Vinylls</i>
Specific gravity (g cm <sup>-3</sup> )	0.95 <sup>a</sup> 0.90 <sup>c</sup>	1.14–1.18 <sup>b</sup> 1.31–1.37 <sup>d</sup>	1.38–1.70	1.30
Molecular weight ( <i>M<sub>w</sub></i> )	250 000–500 000 <sup>a</sup> 100 000–400 000 <sup>c</sup>	65 000	–	65 000
Birefringence	0.052 <sup>a</sup> 0.034 <sup>c</sup>	– 0.003 <sup>b</sup> 0.003 <sup>d</sup>	0.005	–
Tenacity (N/Tex)	0.4–0.7 <sup>a</sup> 0.4–0.8 <sup>c</sup>	0.19–0.32	0.20–0.35 <sup>e</sup>	0.35–0.88 <sup>e</sup>
Extensibility (%)	10–20 <sup>a</sup> 20 <sup>c</sup>	33–64	15–90 <sup>e</sup>	9–26 <sup>e</sup>
Initial modulus (N/Tex)	2.2–4.4 <sup>a</sup> 4.0–9.0 <sup>c</sup>	1.8–4.0	1.3–4.4 <sup>e</sup>	2.2–11.5 <sup>e</sup>
Glass transition <sup>f</sup> <i>T<sub>g</sub></i> (°C)	– 100 <sup>a</sup> – 18 to 25 <sup>c</sup>	75	– 75 to – 10	70–85
Melting (°C)	133 <sup>a</sup> 160–165 <sup>c</sup>	Does not melt	185–265	215–265
Moisture regain (%) at 20°C, 65% RH	< 0.1	1.0–2.0 <sup>b</sup> 3.5–4.0 <sup>d</sup>	0	4.5–5.0

<sup>a</sup> Polyethylene.

<sup>b</sup> Acrylic fibers.

<sup>c</sup> Polypropylene.

<sup>d</sup> Modacrylic fibers.

<sup>e</sup> Low-to high-tenacity ranges.

<sup>f</sup> Quoted value dependent on experimental method used.

Sources: Lewin M and Pearce EM (eds.) (1985) *Identification of Textile Materials*, 7th edn. Manchester: The Textile Institute. Yang HH (1989) *Aromatic High Strength Fibers*, chs. 1 and 2. New York: Wiley.

process to produce bi-component fibers or blend fibers. Thus, an almost unmanageable number of different fibers with different properties exist.

The entire range of properties of a synthetic fiber depends especially on the molecular composition (monomers and tacticity; 'tacticity' means the geometric regularity of polymer side chains) and the

chain length of the polymer. Also, a marked influence on the mechanical properties is raised by the manufacturing conditions. Strength and tensile behavior (e-modulus), crystallinity and density, thermal properties (melting and glass transition points), moisture absorption, and solubility are some properties that can be mentioned in this connection. **Tables 1 and 2**

**Table 2** Properties of fibers derived from nonvinyl polymers

	Nylon	Aramid	Polyester <sup>a</sup>	Polyurethane
Specific gravity (g cm <sup>-3</sup> )	1.13 <sup>b</sup> 1.14 <sup>d</sup>	1.38 <sup>c</sup> 1.45 <sup>e</sup>	1.38	1.1
Molecular weight ( $M_n$ )	14 000–20 000	— <sup>f</sup>	18 000–30 000	—
Birefringence	0.049 <sup>a</sup> 0.056 <sup>d</sup>	Too high to measure	0.160	—
Tenacity (N/Tex)	0.4–0.6 <sup>g</sup> 0.6–0.9 <sup>h</sup>	0.4–0.6 <sup>c</sup> 1.3–2.6 <sup>e</sup>	0.35 <sup>g</sup> 0.65–0.85 <sup>h</sup>	0.05–0.14
Extensibility (%)	20–40 <sup>g</sup> 15–20 <sup>h</sup>	22 <sup>c</sup> 4 <sup>e</sup>	15–25 <sup>g</sup> 7–13 <sup>h</sup>	400–650
Initial modulus (N/Tex)	1.5–3.5 <sup>g</sup> 4.0–5.0 <sup>h</sup>	9–17 <sup>c</sup> 88 <sup>e</sup>	8–13	Very low
Glass transition <sup>i</sup> , $T_g$ (°C)	50–60	270 <sup>c</sup> > 300 <sup>e</sup>	69–90	Very low
Melting point, $T_m$ (°C)	210–216 <sup>a</sup> 252–260 <sup>d</sup>	370 <sup>c</sup> > 500 <sup>e</sup>	250–290	250
Moisture regain (%) at 20°C, 65% RH	4–4.5	5 <sup>c</sup> 0 <sup>e</sup>	0.4–0.6	< 0.1

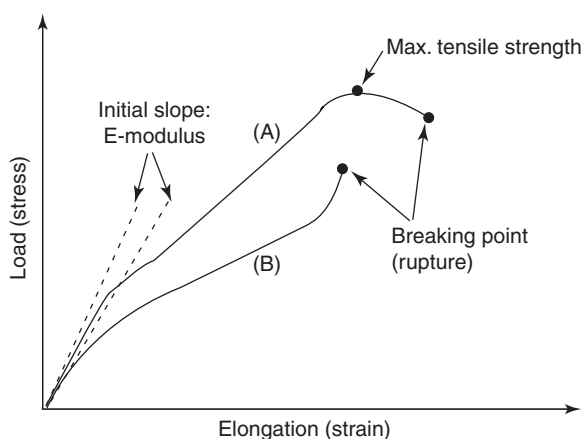
<sup>a</sup> Poly(ethylene terephthalate).<sup>b</sup> Nylon 6.<sup>c</sup> Nomex.<sup>d</sup> Nylon 6.6.<sup>e</sup> Kevlar 29.<sup>f</sup> Weight average molecular weight for unfractionated Kevlar is 1680–63 000. The average degree of polymerization is ~100.<sup>g</sup> Regular.<sup>h</sup> High tenacity.<sup>i</sup> Quoted value dependent on experimental method used.Sources: Lewin M and Pearce EM (eds.) (1985) *Identification of Textile Materials*, 7th edn. Manchester: The Textile Institute. Yang HH (1989) *Aromatic High Strength Fibers*, chs. 1 and 2. New York: Wiley.

summarize the basic properties of a few typical synthetic fibers made from basic polymers. In the following, the analytical methods to investigate the physical, mechanical, and chemical properties of synthetic fibers will be discussed. Methods to analyze the molecular weight are also introduced. Most methods are not specific to fiber materials but generally are suitable for analysis of polymeric materials.

## Analysis of Mechanical and Physical Properties

### Strength and Elasticity

Strength and elasticity properties play an important role in both the production and the utility of textiles. Apart from finished products, fibers and yarns are also investigated to determine their strength. In most cases, this is done by means of a tensile test on prior conditioned fibers (standardized conditions: 20°C/65% RH). The material tensioned between two clamps is increasingly loaded along its axis until it breaks. If the load and the elongation (change in length) are plotted during this process, the resultant graph shows a load–elongation or stress–strain diagram (see Figure 2).



**Figure 2** Load-elongation (stress–strain) diagram with typical curves of a tensile test: (A) with, and (B) without a maximum point (maximum tensile strength).

The maximum force plotted in the load–elongation diagram is described as the maximum tensile strength (also tearing strength). The associated elongation is the elongation at maximum tensile strength (tear elongation). Textiles can often be further elongated to ultimate rupture. This is termed the elongation at break, and the associated force is termed the breaking load.



In comparing the strengths of textile fibers and yarns, however, it is more convenient to use a measurement based on the linear density of the fiber: tenacity (or specific load), expressed as N/tex. The tensile elongation undergone by the fiber is defined as the percentage change in sample length, or percentage extension.

The initial slope of the load–elongation curve is a measure of the extent to which the fiber resists stretching for small extensions. This property is termed the ‘modulus of elasticity’ (E-module). The E-module of fibers is a fineness-related force. It is quoted in cN/tex.

Fibers that show a steep initial slope (high E-module) usually break soon after the yield point. If, however, the stress–strain curve flattens out and continues on, the fiber will yield under high loads. If the applied stress is removed before the breaking point, the fiber may partially recover from extension. This ‘elastic recovery’ is defined as elastic extension (or distance recovered) divided by the total extension (usually expressed as a percentage).

The mechanical properties of textile fibers, yarns, and fabrics may be more fully determined by subjecting the substrate to small forces in directions other than along the fiber axis. Tear, bending, and shear strengths, as well as recovery from bending and abrasion resistance, etc., also influence the wear properties of textiles. Finally, time-dependent extension and recovery, termed creep, or creep recovery, respectively, is of importance in determining the performance of fibers in industrial applications. A discussion on the measurement of these parameters, however, is beyond the scope of this article.

### Crystallinity and Density

Synthetic polymers usually consist of a crystalline part and an amorphous part (ordered and disordered regions on the molecular scale). Besides the molecular composition, the degree of the crystalline fraction largely determines all physical characteristics of textile fibers. Fibers with a high degree of crystallinity exhibit high densities and high tensile strengths. Crystallinity is induced by three main factors: high chain linearity, regular tacticity of the side chains, and/or strong interchain attractive forces. These forces can be exemplified in polyester and nylon. Whereas the cyclic units in polyester enforce a rigidly linear chain conformation, the polar amide linkages in nylon allow the establishment of strong hydrogen bonding forces between polymer chains. Thus, both polymers are able to form highly crystalline conformations.

The most common method to analyze the crystallinity in fibers is X-ray analysis. The degree of sharpness versus diffuse scattering of an X-ray pattern produced from a synthetic fiber is a reliable indication of the relative proportion of its crystalline and amorphous fraction. Additionally, the spacing between rows of ‘reflections’ provides information on the distance of polymer repeating units. Based on theoretical calculations, further information on the extent of chain folding, and the stereochemical structure (tacticity) of a vinyl polymer, can be indicated. An approximation of crystallite size may also be obtained from the degree of diffuseness of the discrete spots on the X-ray diagram. In all polymeric systems the molecular chain length is much greater than crystallite length; hence, a single polymer chain is considered to interconnect through a network of several crystalline and amorphous regions.

### Thermal Properties

As already mentioned above, polymeric materials always consist of a crystalline phase and more or less an amorphous phase (glassy like). Thus, when polymers are heated they usually show two different thermal transition incidence connected to these two phases. The first, the so-called glass transition ( $T_g$ ), is connected with the amorphous state. It means the reversible change of a (partially) amorphous material from a hard and relatively brittle to a viscous or rubbery one. From the thermodynamic point of view  $T_g$  is a so-called second-order transition not connected with a change of the state of aggregation, but a change in molecular mobility. The second thermal transition is the melting point ( $T_m$ ). The melting point is known as a first-order transition. This transition occurs over a relatively narrow range, and is associated with the melting of crystallites within the polymeric system. Generally,  $T_g$  increases with increase in  $T_m$  (as a rule of thumb:  $T_g$  is around two-thirds of  $T_m$  (K)). In addition to these phase transitions ( $T_g$  and  $T_m$ ), chemical reactions and eventual irreversible decomposition may also occur whilst heating.

The most extensively used methods for monitoring these heat induced changes in fibers are differential thermal analysis (DTA) and differential scanning calorimetry (DSC). In DTA, both sample and inert reference are heated at the same rate by the same heat source. When a thermally induced transition occurs in the sample, a temperature difference between the sample and reference results can be recognized. A plot of the difference in temperature between the two against increasing applied temperature exhibits deviations from the initial baseline, depending on



whether the transitions are endothermic or exothermic. These are characterized by a downward or upward deflection, respectively, in the thermogram. The temperature range of interest is usually 25–400°C (at 400°C most polymers decompose). DSC is similar in concept to DTA, but in the case of DSC the heat flow from the oven to the sample is recorded. Integrating the heat flow over time the enthalpy change of a sample can be directly measured with DSC. Thermal transitions are highly characteristic and reproducible, and hence may be used for fiber identification. Especially for moisture absorbing synthetic fibers it has to be pointed out that the glass transition decreases significantly with increasing moisture content. The  $T_g$  of PA 6 in dry condition is ~65°C, but for 'wet' PA 6 the  $T_g$  is close to room temperature!

Dilatometry and thermomechanical analysis (TMA) are also techniques used to monitor the thermal behavior of fibers. They both employ a sensitive probe in contact with the surface of the sample, and the thermal transitions are detected either by a change in volume or modulus of the sample, respectively. In the latter case, the probe necessarily penetrates the sample surface. A variable transformer records the voltage output that is directly proportional to the degree of displacement of the probe during a thermally induced transition. TMA is a more sensitive technique than either DTA or DSC for detecting thermal transitions.

### Moisture Absorbance

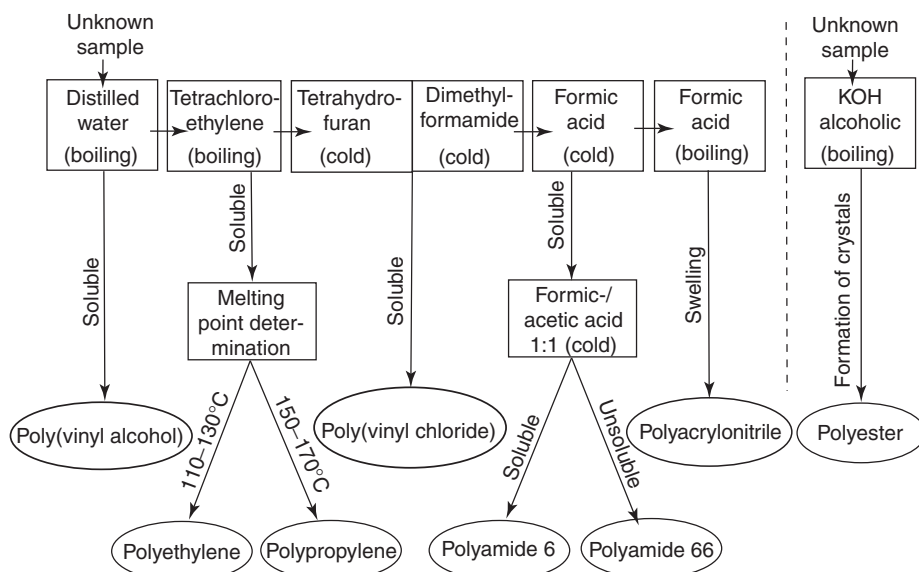
The tendency of a fiber to absorb moisture (water and/or water vapor) depends on the chemical

constitution of the fiber and is defined as the percentage of weight of moisture present in a fiber. The measurement of the moisture content can either be simply realized by the determination of the oven-dry weight (in case of high amounts) or due to more sophisticated methods for lower quantities of moisture. These methods are based upon vaporization of the moisture. The water content is then measured either by means of a coulometric Karl–Fischer titration or due to the measurement of the partial pressure of hydrogen. The hydrogen is released from a gas-phase reaction from water with calcium hydride (a corresponding machine named 'Aquatrac' has been commercialized by Brabender OHG, Duisburg, Germany).

A clear trend is recognized between fiber polarity and moisture uptake. The amount of polar amide (NH–CO) linkages and hydroxyl (OH) pendant groups into the methylene chain structures of polyamide and poly(vinyl alcohol) fibers, respectively, render them significantly more moisture absorbent (4–5%) than the polyolefins and elastomeric fibers (less than 0.1%). Densely packed polymer networks, as in the cases of polyester, poly(vinyl chloride), and particularly Kevlar fibers, exclude the entry of water molecules, again resulting in low moisture content.

### Solubility

The solubility of synthetic fibers in different inorganic or organic solvents depends, as most other properties, on the crystallinity and the chemical constitution of fibers. These differences in solubility can be used to distinguish between synthetic fibers in a simple way. The scheme in Figure 3 shows the way



**Figure 3** Schematic procedure for the determination of an unknown fiber sample due to differences in solubility.

for the characterization of an unknown synthetic fiber through application of different organic/inorganic media. A more effective and also easy characterization of unknown synthetic polymers can be realized with infrared (IR) spectrometry (see later).

### Microscopic Analysis

The characteristic features of a fiber are readily visible by light microscopy, within magnification levels of 200–400 $\times$ . The longitudinal sections of synthetic fibers, however, are rather smooth and structureless, and cross-sections of any desired shape may be produced; hence, microscopic examination of a fiber's appearance is not a useful tool for identification purposes. The cross-sections of synthetic fibers are mostly round or trilobal. Other existing forms are dog-bone or bean-shaped and many others (see Figure 4).

The dispersion of TiO<sub>2</sub> or the regularity of pigments is well visible under an optical microscope. With the use of polarization microscopy it is possible to control the linearity of the quenching process. Nonlinear quenched fibers are recognized by a color change.

In certain instances it may be of interest to examine the etched surface of a fiber or determine the proportion of a particular cross-section in a fiber mixture. This is done by scanning electron microscopy (SEM), where magnifications of up to 100 000 $\times$  can be achieved with greater depth of focus and resolution. Additionally, the cross-sectional appearance of bicomponent and polyblend fibers may be examined by SEM, revealing 'side-by-side' (Figure 4), 'sheath-core', or 'island-in-the-sea' composites.

## Analysis of Chemical Constitution

### Infrared Spectroscopy

This method of analysis is suitable for fiber identification, as well as for characterization of any chemical differences between two fibers of the same class. For instance, acrylic and modacrylic fibers containing copolymers of varying constitution and proportion may be identified. Identification depends on matching the additional bands with those in the IR spectrum of a fiber whose identity is already known. Both the wavelength and intensity differences of bands in each spectrum must be taken into account. Fiber samples may be prepared in the form of a pressed disk, where very finely divided particles of the fiber are uniformly distributed in powdered potassium bromide. The mixture is pressed into a small disk of about 1 mm thickness in a vacuum die under

pressure. Alternatively, a solvent-cast film of the fiber may be prepared. A suitable solvent for dissolving the fiber is necessary. A solution of the fiber is placed on a glass surface, and the solvent is carefully evaporated, preferably in a vacuum, so as to avoid bubble formation. The sample film is peeled from the glass plate and analyzed directly.

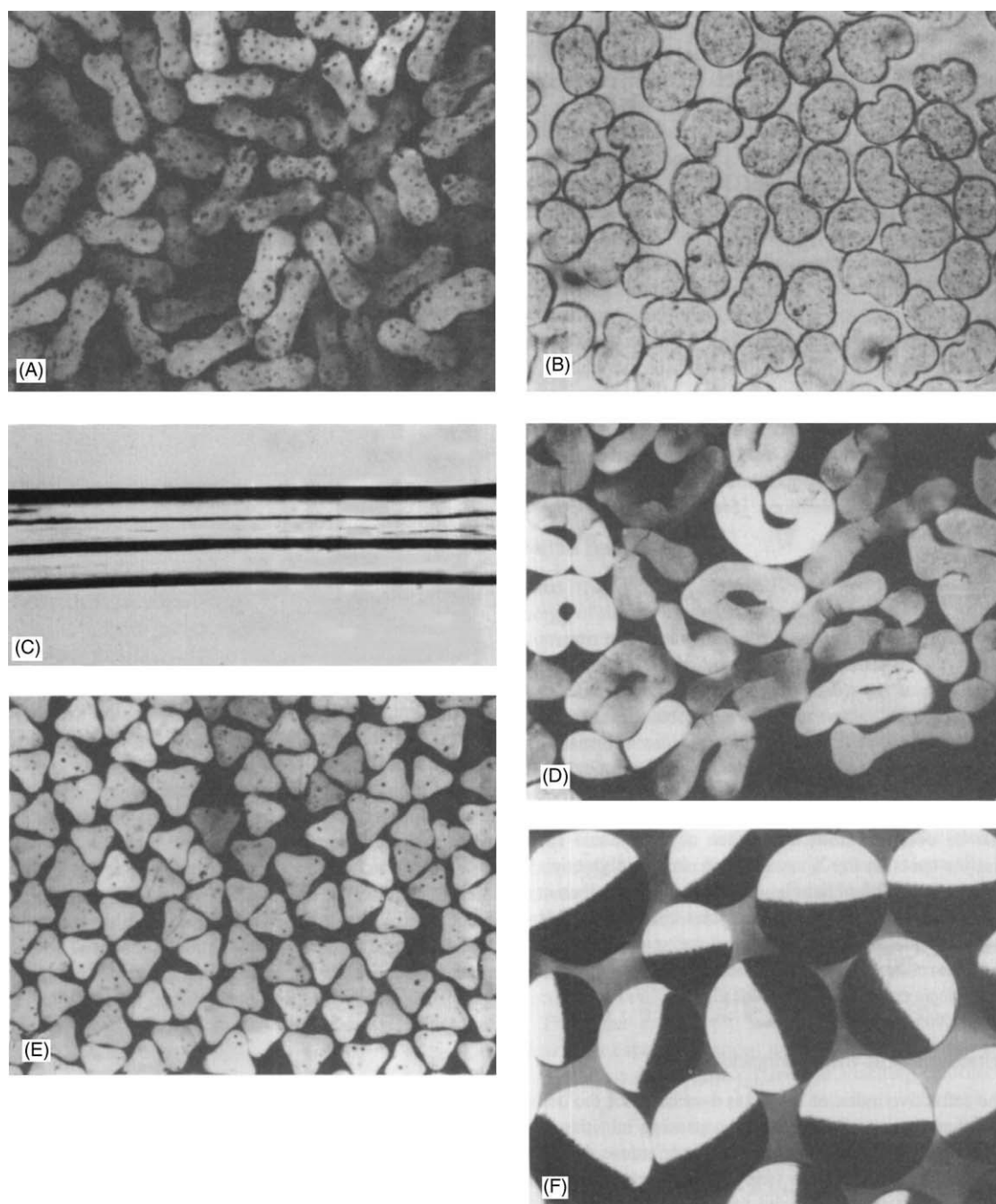
An extremely convenient instrumental technique that may be used to measure the IR spectrum of fibers or fabrics without prior preparation is attenuated total reflectance (ATR) spectroscopy. The technique is based on obtaining an IR absorption spectrum of a sample that is in contact with a reflecting surface. The incident IR beam is directed into the reflecting plate at an angle greater than the 'critical angle'; the beam is subsequently reflected after penetrating a few micrometers into the sample. A very simple to use ATR equipment suitable for almost all IR spectrometer types is the so-called 'Golden-Gate' accessory (commercialized by Grasby Specac Ltd., St. Mary Cray, Orpington, UK). All sorts of samples, liquids, powders, granules, fibers, etc., can be measured directly without any prepreparation of samples.

### NMR Analysis

Nuclear magnetic resonance (NMR) spectroscopy is a powerful but theoretically complex analytical tool. It can be used for the wide field of the structure analysis of organic materials. Also, synthetic polymers can be analyzed with NMR spectroscopy as long as the materials are soluble in suitable solvents (almost always chloroform). A specific application for NMR analysis in the field of synthetic fibers is the analysis of spin finish (processing aid). For this special application the fiber material is analyzed regarding the amount of protons in the solid and the liquid phase. The 'liquid protons' can be assigned to the amount of spin finish, if the technique is pre-calibrated well. Analysis time is typically 1 min per sample and no sample preparation is required. Thus, NMR is a rapid and accurate alternative to the solvent extraction techniques and can be used as an almost online production control (a corresponding machine for spin finish analysis by NMR is commercialized by Oxford Instruments Ltd, Abingdon, Oxon, UK).

## Analysis of the Molecular Weight Distribution

Due to the mechanism of polymerization processes (radical polymerization as well as polycondensation), all polymers exhibit a more or less pronounced distribution of polymer chain length, which means that



**Figure 4** (A) Acrylic fiber (Orlon 42). Dogbone-like cross-section shape;  $\times 750$ . (B) Acrylic fiber (Acrilan). Bean-shaped cross-section;  $\times 240$ . (C) Modacrylic fiber (Dynel). Deeply-fluted structure;  $\times 750$ . (D) Modacrylic fiber (Dynel). Folded shapes with deep re-entrants;  $\times 750$ . (E) Nylon 6,6, trilobal cross-section;  $\times 300$ . (F) Bicomponent polyester fiber, side-by-side type cross-section;  $\times 160$ . (Reproduced with permission from Lewin M and Pearce EM (eds.) (1985) *Identification of Textile Materials*, 7th edn. Manchester: The Textile Institute.)

no polymer consists of uniform molecules, where all chains have the same length with the same number of repeating units (monodisperse distribution). Thus, from polymers only average molecular weights, molecular weight distributions can be determined. But the knowledge of that is an important analysis

for every polymeric system. Especially for polymers subjected to a spin process the knowledge of these data is essential. If the molecular weight is too high, the melt viscosity may be too high for spinning process; if it is too low, adequate tensile properties may not be achieved.

The average molecular weight may be determined as a number average or weight average depending on the method of measurement. In the former case, the molecular weight is based on the mole fraction (or sample weight per mole), while in the latter case, the weight fraction determines the average molecular weight:

$$M_n = \frac{\sum M_i N_i}{\sum N_i} \quad M_w = \frac{\sum N_i M_i^2}{\sum N_i M_i}$$

where  $M_n$  is the number average molecular weight,  $M_w$  the weight average molecular weight,  $M_i$  the molecular weight of each species, and  $N_i$  the number of moles of each species.

$M_w$  is therefore greater than  $M_n$ , and the ratio  $M_w/M_n$  is indicative of the molecular weight distribution of a polymer sample (polydispersity  $P$ ). The more similar these values, the narrower the molecular weight range.

The most widely used method for determining molecular weight distribution for fiber forming polymers is gel permeation (size exclusion) chromatography (GPC or SEC). Other commonly used methods in this case are end group analysis, osmometry, and viscometry. These methods are also introduced below.

### Gel Permeation Chromatography

GPC (or SEC) is generally the technique of choice in determining the molecular weight distribution of synthetic fibers that are soluble in common organic solvents like tetrahydrofuran or chloroform. The technique is based on chromatographic separation of polymer molecules according to size by elution through a column of porous material (cross-linked polystyrene or silica gel). The smaller polymer molecules diffuse into the pores of the packing material and thus elute more slowly; the higher molecular weight fractions therefore elute first. A chromatogram is obtained relating detector response with elution volume, where the size of the peak is proportional to the amount of polymer eluted. Detection of the polymer eluent is accomplished by refractive index or ultraviolet. In order to correlate eluent volume with molecular weight it is necessary to calibrate the column with standard polystyrene fractions of known molecular weight. Due to the calibration of the GPC columns by means of polystyrene standards the obtained molecular weight data for all other polymers are only relative.

A more sophisticated method to determine molecular weights is the so-called universal calibration. This method, based on the principle that  $\log([\eta]M)$

versus elution volume is a constant for all polymers, for a given column, temperature, solvent, and elution volume. In other words,

$$[\eta]_1 M_1 = [\eta]_2 M_2$$

where  $[\eta]_1$  and  $[\eta]_2$  are intrinsic viscosities of the standard and sample, respectively, and  $M_1$  and  $M_2$  are the molecular weight fractions of the standard and sample, respectively. The intrinsic viscosity can be determined with a special viscosity detector (commercialized by Viscotek, Houston, TX, USA) for a standard and an unknown sample and the equation is solved for  $M_2$  (molecular weight of the unknown sample).

Absolute  $M_w$  data can be obtained by using a light scattering detector. A beam of monochromatic polarized light (laser) is passed through the sample, and the intensity of the scattered light is measured at one or multiple angles after passing through a slit system and a second polarizer. The intensity of the scattered light ( $R_0$ ) is a function of the weight average molecular weight of the polymer ( $M_w$ ), the concentration ( $C$ ), and an optical constant ( $K$ ):

$$R_0 = K \cdot C \cdot M_w$$

The optical constant  $K$  depends on the refractive index (RI) of the polymer in the given solvent and is known for a few common polymers or can be determined from the signal of the RI detector (see above). Due to some given, not further explained boundary conditions, this method is suitable for polymers and synthetic fibers with an average molecular weight ( $M_w$ ) greater than  $100\,000 \text{ g mol}^{-1}$ .

### End Group Analysis

End group analysis is particularly suitable for condensation polymers in the molecular weight range of  $10\,000$ – $20\,000$ , such as polyesters and polyamides. These polymers contain titratable end groups that are distinct in character from the main chain. The carboxylic acid end groups of a polyester may be titrated by dissolution of the polymer in benzyl alcohol and the addition of sodium benzyolate as the base. The carboxyl or amino end groups of polyamides may be determined in phenol or *m*-cresol by titration with alkali or acid. The endpoint is determined in the presence of an indicator, or conductometrically.

### Osmometry

There are two principal methods of osmometry that are suitable for determining average molecular weights of polymers: membrane and vapor



pressure osmometry. While the first one is suitable for molecular weights between 50 000 and 2 million ( $\text{g mol}^{-1}$ ), the second one is applicable for 'short' polymeric chains below 40 000  $\text{g mol}^{-1}$ . Both methods deliver the absolute value of the number average molecular weight ( $M_n$ ).

In the first case, a solution of a polymer and the pure solvent are placed in compartments separated by a semipermeable membrane. The membrane allows diffusion of small solvent molecules, but restricts the larger polymer chains to one compartment only. Hence, a net diffusion of solvent takes place from the solvent side to the solution side until sufficient hydrostatic pressure develops that prevents further diffusion. This hydrostatic pressure is the osmotic pressure, which is related to molecular weight by the van't Hoff equation extrapolated to zero concentration:

$$\left\{ \frac{\pi}{C} \right\}_{C=0} = \frac{RT}{M_n} + A_2 C$$

where  $\pi$  is the osmotic pressure,  $C$  the concentration of polymer ( $\text{g l}^{-1}$ ),  $T$  the temperature (K),  $R$  the gas constant, and  $A_2$  the second virial coefficient (solvent dependent).

A plot of  $\pi/C$  versus  $C$  is a straight line of slope  $A_2$ , and  $\gamma$ -intercept equal to  $RT/M_n$ . The permeability of the membrane to low molecular weight chains renders membrane osmometry useful for polymer molecular weights greater than 50 000, while inaccuracy in the measurement of very small osmotic pressures sets the upper limit at 2 million.

The second method is based on the vapor pressure difference of pure solvent and a polymer solution. A sample of the solution and pure solvent are introduced into a temperature-controlled measuring chamber, which is saturated with solvent vapor. Since the vapor pressure of the solution is lower than that of the solvent, solvent vapor condenses on the solution sample causing its temperature to rise. This temperature difference ( $\Delta T$ ) can be measured for different concentrations ( $C$ ) of the polymer solution and  $1/M_n$  can be calculated according to the following formula:

$$\left\{ \frac{\Delta T}{KC} \right\} = \frac{1}{M_n} + A_2 C$$

$K$  is a measuring constant determined for a given solvent and temperature with an organic substance (e.g., benzoin) with exactly known molecular weight. A plot of  $(\Delta T/KC)$  versus  $C$  delivers  $1/M_n$  as the  $\gamma$ -axis intercept (slope  $A_2$ ).

## Viscometry

A routinely used method of molecular weight determination is to measure the intrinsic viscosity  $[\eta]$  in a suitable solvent, and then use a correlation based on the Mark-Houwink-Sakurada relationship:

$$[\eta] = KM_v^\alpha$$

where  $M_v$  is the viscosity average molecular weight, which lies between the corresponding values of  $M_w$  and  $M_n$ . The constants  $K$  and  $\alpha$  are the intercept and slope, respectively, the values of which have already been published for a wide range of polymers, solvents, and temperatures. Although this method does not provide an absolute determination of molecular weight, it is an extremely simple procedure and hence frequently utilized in commercial practice.

Viscosities are measured in dilute solutions (0–5 g per 100 ml) by monitoring the flow time of a certain volume of solution through a capillary of fixed length. The device used is called a capillary viscometer. Polymer solutions must be filtered carefully using microfilters, as dust and dirt particles significantly affect the flow time.

**See also:** **Infrared Spectroscopy:** Overview; Sample Presentation; Industrial Applications. **Liquid Chromatography:** Size-Exclusion. **Microscopy:** Overview; **Microscopy Techniques:** Light Microscopy; Scanning Electron Microscopy. **Nuclear Magnetic Resonance Spectroscopy:** Overview. **Textiles:** Natural.

## Further Reading

- Koslowski HJ (1999) *Dictionary of Man-Made Fibers*, 1st edn. Frankfurt am Main: International Business Press.
- Lewin M and Pearce EM (eds.) (1998) *Handbook of Fiber Science and Technology*, vol. IV, *Fiber Chemistry*, chs. 1–5. New York: Dekker.
- Mitchell J Jr. (ed.) (ed.) (1987) *Applied Polymer Analysis and Characterization*. Munich: Hanser Verlag.
- Morton WE and Hearle JWS (1975) *Physical Properties of Textile Fibers*. Manchester: The Textile Institute.
- Robinson JS (ed.) (1980) *Fiber-Forming Polymers Recent Advances*. Park Ridge, NJ: Noyes Data Corporation.
- Robinson JS (ed.) (1980) *Manufacture of Yarns and Fabrics from Synthetic Fibers*. Park Ridge, NJ: Noyes Data Corporation.
- Stevens MP (1990) *Polymer Chemistry. An Introduction*, 2nd edn. Oxford: Oxford University Press.

# THERMAL ANALYSIS

Contents

## Overview

### Temperature-Modulated Techniques

### Coupled Techniques

### Sample-Controlled Techniques

### Nonbasic Techniques

## Overview

**J O Hill**, La Trobe University, Wodonga, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Thermal analysis (TA) can now be considered as a ‘mature science’, having been continuously developed over some 11 decades. In the last decade, there have been a variety of developments, not all associated with TA techniques. Those developments that have been associated with TA techniques relate to new methodologies, such as ‘temperature modulation’, ‘sample controlled regimes’, ‘online data analysis’, and ‘robotic operational regimes’. The major developments relate to the fields of application of TA techniques and the use of thermal methods in the kinetic analysis of chemical reactions. Hence, to write simply an ‘update’ presents a dilemma. This task has already been achieved via the recent publication of two books in the field, by Haines and by Brown. Further, any present coverage of the field needs to recognize that it has become increasingly difficult to differentiate between ‘TA’ and ‘calorimetry’ and many TA techniques can equally be classified as calorimetric techniques. Thermal methods of analysis have developed in accordance with the need to study the changes in the properties of a sample as a consequence of heating, whereas calorimetric methods measure such heat changes directly. TA and calorimetric techniques and the associated fields of application are uniquely complementary. The present review concentrates on developments associated with a wide range of TA techniques, with an emphasis on recent developments in applications of these techniques. In this context, it is the scope of the field overall that has developed very significantly, not only with respect to ‘techniques’ but also with respect to a range of associated features. These are highlighted here, in

addition to reviewing the developments associated with TA techniques.

## Revised Grouping of Techniques

For the purpose of the present review, TA techniques are grouped into four major categories: temperature modulated techniques, coupled techniques, sample controlled techniques, and less common techniques. ‘Modulation’ is probably the most significant development in terms of a ‘new’ TA technique. Modulation has had an impact on TA techniques similar to that of Fourier transform on infrared (IR) spectroscopy in that essentially, it has allowed a more complete analysis and a greater degree of refinement of a thermal event. Modulation has mainly been developed in conjunction with differential scanning calorimetry (DSC) to create a ‘new’ technique – temperature modulated (TM) DSC (TMDSC). Essentially, the use of TMDSC with appropriate frequencies and amplitudes of modulation allows separation of reversing processes, such as glass transitions, from nonreversing processes, such as relaxation endotherms or cure reactions. Further, baseline curvature on the cyclic signal is generally minimal – thus overcoming a major drawback of conventional DSC and making it easier to distinguish between baseline effects and real transitions. The overall resolution of TMDSC is higher than that of conventional DSC since with the former very low heating rates can be used. TMDSC has now become a major investigative technique in materials science and is used excessively in (synthetic) polymer science. The TM concept has also been applied to thermogravimetry (TMTG), thermomechanical analysis (TMTMA), dynamic mechanical analysis (TMDMA), and thermoelectrometry (TMTE) but to a lesser extent than to DSC. However, this family of ‘TM techniques’ deserves a separate review.

The second selected grouping of TA techniques is ‘coupled techniques’. These have also been traditionally grouped as ‘simultaneous TA techniques’,



although the range is now considerably greater than a combination of two TA techniques. With the coupled TA techniques, two or more types of measurement are made on the same sample at the same time. The benefits of such coupling are obvious. For a given material, it is unlikely that any single TA technique is able to provide a sufficiently comprehensive range of data for characterization purposes. Complementary data are usually required that may be obtained using another TA technique or via some other analytical process. With coupled TA techniques, the challenge is to correlate the individual sets of data obtained from the individual analysis systems that form the coupled technique. In simple terms, data that correlate precisely lead to a synergistic effect so that the total value of the information is greater than the sum of the parts, creating a 'multiplying effect'. However, although coupled systems are capable of excellent levels of precision and sensitivity, it is a fundamental principle that such systems do not perform the measurement in question as well as the dedicated separate units. Thus a compromise scenario prevails with the coupled TA systems. The earliest simultaneous TA system dates from 1955 and was developed by the Paulik brothers in Hungary. The instrument was called the 'derivatograph' and was essentially a combination of thermogravimetry (TG) and differential thermal analysis (DTA). The derivatograph has since been developed further to incorporate dilatometry and evolved gas analysis and has been used to pioneer the development of 'controlled rate methods' of TA. A range of commercial TG-DTA simultaneous TA systems is now available, and TG-DTA is the most widely used of the coupled TA techniques family. The 'coupling' can be further extended to analyzing the gases evolved during a thermal event – thereby creating an evolved gas analysis (EGA) system. This is achieved by coupling an STA system with a gas chromatograph and/or a mass spectrometer or, more recently, a Fourier transform IR (FTIR) spectrometer. Such highly sophisticated coupled systems are intrinsically very powerful 'complete analysis' systems.

It is also possible to couple TG with DSC to produce a simultaneous TG-DSC system. Apart from providing calorimetric data on a sample, subambient analysis is also possible with TG-DSC, which is especially useful for the study of systems containing moisture and other volatiles.

Hot-stage microscopy can be considered as a 'coupled TA technique' and has been widely applied. The technique involves observing a sample with a microscope as it is heated progressively on a 'hot-stage'. Such observations are often useful in the preliminary

interpretation of the thermal characteristics and behavior of materials by revealing the nature of the transition or reaction taking place. With particularly reactive materials, hot-stage microscopy can provide a preliminary 'thermal screening' that can be accommodated to prevent damage to a TG or DSC system in a subsequent detailed TA of such materials. Essentially, although a qualitative technique, hot-stage microscopy can reveal melting, bubbling, cracking, creeping, shrinking, swelling, decrepitation, and, significantly, interaction with crucible materials. In addition, the technique can reveal solid-solid transitions – particularly if associated with a color change and gas evolution with associated sample volume changes. It is also possible to combine TG, DTA, and DSC with microscopy – the latter systems have been particularly valuable for the investigation of coal samples, particularly in the determination of the pyrolysis and combustion characteristics of coal types.

Microscope hot-stages designed for transmitted light investigations and involving specially designed heating accessories have been used in conjunction with a FTIR spectrophotometer for studying physical and chemical changes in a heated sample. By using a hot-stage with a DSC cell incorporated, simultaneous DSC-FTIR is achieved. More flexibility can be achieved by using an IR microscope accessory to collect reflectance spectra from a sample contained in a DSC cell. Such a system is particularly useful for studying polymer curing phenomena and for studying structural changes occurring during glass transitions and melting.

X-ray diffraction (XRD) is particularly useful for the characterization of reaction intermediates and products in TA experiments. Coupling XRD to a TA technique potentially overcomes the difficulty of isolating the reaction intermediate or product from the TA system during a dynamic heating program – thereby avoiding 'time-lapse' phenomena and interactions with the prevailing atmosphere. Systems involving the coupling of TG with XRD and DSC with XRD have been devised and are particularly useful for a study of the thermal behavior of catalysts.

A relatively new TA technique, known as microthermal analysis ( $\mu$ -TA) can be considered as a coupled TA technique since usually two or more measurements are made simultaneously.  $\mu$ -TA combines the imaging capabilities of atomic force spectroscopy with a form of localized TA, and this technique offers nanotechnological TA applications.

The coupled TA techniques group consists not only of the commercial simultaneous TA systems but also a wide range of ingenious systems involving the coupling of TA systems with spectroscopic, optical, chromatographic, and a range of other analytical

techniques. Such coupled systems enhance very significantly the integrity of TA as a major analytical science, offering a diversity of application potential and a diversity of application range.

The third selected grouping of TA techniques is 'sample controlled techniques'. Sample controlled thermal analysis (SCTA) essentially overcomes the problems associated with linear heating rates in conventional TA. Linear heating rates cause significant temperature and pressure gradients in the sample. The judicious use of heating regimes that are modified in some manner by the reaction rate can greatly reduce these problems, giving enhanced resolution in analytical and characterization investigations. In conventional TA, the temperature of the sample follows some predetermined path as a function of time. In SCTA, it is a parameter of the sample that follows a predetermined path as a function of time, this being achieved by temperature variation of the applied heating program. Thus, SCTA represents a new paradigm in TA and was first introduced in 1989 by Rouquerol under the generic title 'Controlled Rate Thermal Analysis'. Essentially, in SCTA, it is the rate of mass loss of the sample material that is maintained constant during the applied temperature regime. An immediate advantage of the SCTA technique over conventional TA techniques is that the effect of material 'self-heating' is eliminated during the TA experiment. Over the last decade, a family of SCTA techniques has emerged. 'temperature-jump TG' and 'rate-jump TG', together with the very recent 'Hi-Res' TG<sup>TM</sup> and 'dynamic rate TG' – which are all members of this emerging SCTA family.

SCTA has many advantages over conventional TA techniques: SCTA data are less liable to contain artifacts due to inhomogeneous reactions within the sample matrix; SCTA generates superior kinetic data for the measured thermal event; SCTA offers defined preparative conditions for porous or finely divided solids such as coals – it is ideal for char preparation under defined conditions. Also, SCTA confines thermal events to occur within a narrow temperature range and clearly delineates the beginning and end of such events, thereby offering superior resolution of thermal events associated with materials. However, there is a major disadvantage of SCTA as compared with conventional TA techniques. The experimental time span required to complete an SCTA experiment is markedly increased over that for a conventional TA run due to the obvious fact that it is the thermal characteristics of the sample itself that control the time frame of the experiment rather than the operator.

Since SCTA usually provides improved control of the sample environment, it is frequently used as a method for preparing porous and finely divided

solids by thermal decomposition, specifically for use as catalysts. SCTA is particularly advantageous for the study of the reaction kinetics of materials. Temperature modulation is also being incorporated into SCTA to further enhance the unique capabilities and application range of these techniques.

The last group of TA techniques to be reviewed is that of 'nonbasic (or less common) techniques'. It is generally accepted that the basic TA techniques are TG, DTA, DSC, and the thermomechanical techniques, thermomechanical analysis (TMA) and DMTA. However, over the last decade, the range of TA techniques has widened considerably, and many of the so-called less common techniques have emerged into prominence. One such technique is thermomagnetometry. When a sample is placed in a magnetic field, it may experience either attractive or repulsive forces, the extent of which can be measured as an 'apparent' mass gain or a mass loss. Hence thermomagnetometry is particularly useful for analyzing the thermal behavior of paramagnetic and ferromagnetic materials. It has particular application to minerals and inorganic materials – particularly transition metal compounds. The apparatus for thermomagnetometry is the same as that for TG, with the addition of a magnet for producing a magnetic field around the sample. Temperature calibration of TG systems is effected by using 'magnetic standards' and is hence achieved by thermomagnetometry.

Emanation TA is another prominent less common TA technique. This technique involves the measurement of the release of (occluded) inert (usually) radioactive gas from a sample during the heating thereof. The rate of gas release reflects the changes that are occurring in the sample during the heating process and relate to the microstructure of the sample. The occluded gases are trapped at lattice imperfections, and these defects serve as both traps and diffusion pathways, and hence the migration of occluded gases in a solid effectively gives a three-dimensional defect map of the material under investigation.

Thermosonimetry (TS) and thermoacoustimetry have both attracted increased application in the last decade. In TS, sound waves emitted by a sample are measured as a function of temperature during heating of the sample. In thermoacoustimetry, the characteristics of imposed sound waves passing through a sample are measured as a function of temperature during heating of the sample. These techniques are particularly useful for detecting mechanical vibrations in a sample arising from dehydration, decomposition, melting, and solid–solid phase changes. TS data are usually used in combination with other thermoanalytical data (TG and DSC) to rationalize the thermal behavior of materials.

Thermoelectrometry embraces a small family of TA techniques that measure the electrical properties of a sample as a function of temperature. These techniques measure changes in conductance, capacitance, and dielectric properties as a function of temperature. The most prominent of these techniques is dielectric thermal analysis (DETA), which measures both the capacitance and the conductance of a sample as functions of time, temperature, and frequency. The capacitance is a measure of the ability of the sample to store charge, while the conductance is a measure of its ability to transfer charge. As for most TA techniques, it is the changes in these properties during thermal events in the sample that are of paramount interest rather than the absolute values of the properties themselves. DETA is particularly suited to the study of changes in molecular orientation during phase changes or chemical reactions or order-disorder crystal transitions. DETA is also useful for the resolution of multiple transitions in the solid state – particularly those associated with long chain and functional group motions in synthetic polymers.

Very recently, microwave thermal analysis (MWTa) has been enunciated. In MWTa, the sample is heated by microwave radiation, which has the advantage of providing direct molecular heating of the sample rather than via conduction or convection as in conventional heating regimes, and hence temperature gradients in the sample are minimized. Further, such uniform heating of the sample enhances the resolution of thermal events. Essentially, MWTa depends on changes in the dielectric properties of the sample, but because of its superior heating regime, it appears to have an illustrious future.

## Thermal Analysis Applications

Since TA techniques are conventionally applied to the study of the thermal behavior of materials in the solid state, it can be assumed that such techniques collectively have a major application in materials science. The thermal behavior of liquids can also be studied, but these require specialized variations of conventional TA techniques. Over the last decade, it is the development and extension of the application ranges of TA techniques that have been greater than the developments associated with the techniques *per se*. The range of application of TA techniques seems to be limited only by the sensitivity of the measurements and adaptation of the ‘problem’ to a ‘thermal solution’.

TA techniques can be used to study a wide variety of phenomena both physical and chemical in nature – such as melting and glass formation, solid-solid

phase changes, oxygen content, nonstoichiometry, structural defects, sintering and cracking, thermal expansion, dehydration, swelling, decomposition, heat capacity, sublimation, fusion, vaporization, linear and bulk deformations, reaction kinetics, surface morphology, and electrical properties of solids. As such, TA techniques inevitably have unlimited application.

With respect to the TA of individual compounds, as distinct from bulk materials, the thermal decomposition mechanisms of inorganic transition metal coordination compounds continues to be an expanding field of application – together with the application of thermal methods for deriving the kinetics of decomposition. In these contexts, such studies are becoming increasingly more creditable with the added characterization of reaction intermediates, evolved gaseous products, and final residues – usually achieved using EGA techniques for evolved gaseous products and IR and/or XRD for solid reaction intermediates and final residues. Such studies are usually correlated with the structures and thermal stabilities of metal complexes, particularly those of catalytic and biological activity.

With respect to organic compounds, those that are precursors to pharmaceuticals have attracted increasing interest from the viewpoint of their thermal properties, and TA techniques are used frequently in quality control procedures for such compounds. A wide variety of common and less common pharmaceuticals have also been investigated using a variety of thermal methods – mainly DSC and mainly in the context of purity determination. There is also potential for using TA techniques to identify pharmaceutical activity and for ‘shelf-life’ determinations.

In terms of (bulk) materials, the greatest use of TA techniques has been and continues to be focused on polymers. Techniques of particular prominence in this domain are DSC and the thermomechanical techniques. DSC is routinely used to study glass transitions in polymers together with curing phenomena of polymer blends. Thermomechanical methods are invaluable for the study of the mechanical properties of polymers in both the bulk form and in the form of fibers. New TA techniques such as  $\mu$ -TA will inevitably enhance and considerably refine these studies.

Minerals of great variety have been studied using TA techniques – particularly clay minerals and coal, lignite, and wood. Traditionally, the simultaneous methods have been applied in these cases, but increasingly, new methods such as sample controlled TA are becoming more prominent, particularly for studies of pyrolysis and combustion of coals. The rationale for using SCTA over conventional TA for the study of coal is simply related to overcoming the

problem of self-heating, which usually leads to premature ignition in coal combustion studies. Also, 'preparative SCTA' is available for 1 g samples – allowing the preparation of chars under precisely controlled conditions.

Progressive development of high-temperature DTA has allowed the thermal characterization of materials such as ceramics and studies of molten inorganic salts as stable nonaqueous solvents in synthetic chemistry and to some extent in industry. This latter field has attracted increased interest recently in terms of its potential 'green chemistry overtones'. Explosives, propellants, and pyrotechnics also form a major TA application area – particularly in terms of composition (phase diagrams) and thermochemical properties.

Other types of materials that lend themselves to TA investigation are electronic materials – with particular emphasis on the use of thermomechanical and thermoelectrical techniques. Heat capacity, conductivity, capacitance, dielectric phenomena, and structural morphology are all key features that can be studied using TA techniques.

There is much commercial interest in 'thin film' and 'nano' technologies, and these pose a new challenge for TA techniques since conventional TA deals with 'statistically averaged properties and phenomena'. However, with the advent of a range of 'modulated (temperature) techniques', micro sample sizes, and consequential rapid response times, the sensitivity of TA techniques generally is sufficiently enhanced to be able to address these immediate application challenges.

A wide range of biological materials ranging from hair and ancient papyri, to foodstuffs such as chocolate and edible oils and fats have been studied using TA techniques. The activity of biological materials in general is highly dependent upon temperature and frequently is confined to a narrow temperature range. The shelf-life of foodstuffs is obviously dependent on temperature, and the decay processes are delicately influenced by enzymatic action. It is these intriguing processes that can readily be studied using such techniques as DSC.

Finally, in terms of methodologies, TA techniques are routinely used for purity determination, heat capacity measurement, and kinetic analysis – the latter representing a major advance in the experimental reaction kinetics domain.

## Auxiliary Developments

It was emphasized (see Introduction) that the recent developments of TA techniques have been

comprehensively detailed in two recently published texts. A futuristic overview of further developments in the twenty-first century is given by a collection of papers in *Thermochimica Acta* (W. Hemminger, ed.). For completeness of the present overview of the field, some 'auxiliary (supplementary) developments' in TA are included here but are not further discussed in the subsequent four articles.

Thermal analysts operate under the acronym 'SCRAM' – 'sample', 'crucible', 'heating rate', 'atmosphere', and 'sample mass'. A thorough understanding of each of these components is critical if quality, reproducible results are to be obtained from a TA investigation. The art of TA is to relate the thermo-analytical data obtained directly to the sample under investigation. However, a myriad of experimental 'effects' affect such data, and hence, although thermal analytical techniques are generally regarded as 'easy to use', data correlation is most often not an easy task. Fortunately, over the last decade, sample, crucible, heating rate, atmosphere, and sample mass effects have been very effectively delineated, and a compendium of advice is now available as to how the majority of these effects can be overcome or avoided, so as to achieve thermoanalytical data of integrity and reliability.

Other 'technique related' developments include TA conducted under pressures greater than ambient. In high pressure DSC (HPDSC), the entire DSC cell can be pressurized to several megaPascals. HPDSC is useful in enhancing understanding of crystallization and melting of polymers and to determine the properties of materials produced under high pressure processing conditions such as injection molding. Another derivative of DSC is photo-DSC, in which samples can be irradiated with ultraviolet radiation to initiate reaction. This technique has particular application in polymer curing technology. Finally, in this context, robotic systems for handling multiple samples without direct operator intervention are available – particularly for use in DSC and quality control determinations. The sequence in which samples are handled and their individual treatments can be preselected and programmed, and the total data obtained are stored for subsequent analysis.

Any discussion of developments in TA would not be complete without reference to the concurrent development of software packages for TA data processing, and the availability of powerful spreadsheet packages, such as Microsoft Excel and Corel Quattro, has virtually removed any necessity for the writing of dedicated software. Databases for TA data and for kinetic analyses of TA results continue to develop. TA networks are proliferating.



Nomenclature in TA continues to be a controversial issue. The International Confederation of Thermal Analysis and Calorimetry (ICTAC) has essentially overseen developments in this domain over some four decades. However, practical nomenclature is determined by general acceptance, and a variety of factors affect this. The nomenclature recommendations released by Hemminger and Sarge in 2001, subsequent to widespread and intensive discussion with thermal analysts worldwide, have yet to be formally accepted by ICTAC. Likewise, modifications to the '*Recommendations for Reporting Thermal Analysis Data*' – as reported in the late 1960s and early 1970s – have yet to be approved by ICTAC. In short, developments on these issues have been severely curtailed by a lack of international agreement on the core principles.

Calibration of TA instrumentation and development of standards for calibration continue to be administered by ICTAC in conjunction with ASTM. The standardization Committee of ICTAC has certified a range of materials for temperature calibration of TA systems, and in addition, standards for calibration of 'mass' (known as 'Class M Standards') are available for this purpose. A range of certified reference materials are available for enthalpy calibration in DSC. Temperature calibration for TMA and dynamic mechanical analysis (DMA) is effected by using disks of pure metals (silver, aluminum, and tin) separated by alumina disks. Load or force calibration for DMA is a complex process involving the use of calibrated weights. Temperature calibration for DETA is effected by measuring the melting transition of benzoic acid and dielectric calibration is

simply achieved by measuring the dielectric constant of the empty cell. Developments in the area of calibration in TA involve increasing the range of calibration materials available, which involves exhaustive 'round-robin' assessment exercises of potential materials.

Developments in TA *per se* depend on the continuity of ICTAC and European Society for Thermal Analysis and Calorimetry conferences and the continuity of the publications on TA and calorimetry. Also, the future of TA (and calorimetry) depends on tertiary education institutions providing courses in these areas, so as to ensure a future supply of trained thermal scientists.

**See also: Thermal Analysis:** Temperature-Modulated Techniques; Coupled Techniques; Sample-Controlled Techniques; Nonbasic Techniques.

## Further Reading

- Brown ME (2001) *Introduction to Thermal Analysis – Techniques and Applications*. Amsterdam: Kluwer.
- Duna JG (2002) Recommendations for reporting thermal analysis data. *Thermochimica Acta* 390: 1–3. Amsterdam: Elsevier.
- Haines PJ (ed.) (2002) *Principles of Thermal Analysis and Calorimetry*. Cambridge, UK: Royal Society of Chemistry.
- Hemminger W (ed.) (2000). *Thermal Analysis & Calorimetry – Towards the New Century*. *Thermochimica Acta* 355. Amsterdam: Elsevier.
- Hemminger W and Sarge SM (1998) In: Brown ME (ed.) *Handbook of Thermal Analysis and Calorimetry*, vol. 1. Amsterdam: Elsevier.

## Temperature-Modulated Techniques

**J O Hill**, La Trobe University, Wodonga, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

It has always been recognized that there is both a subtle difference and a synergy between 'thermal analysis' and 'calorimetry' and the associated groups of experimental techniques. There has been generations of interest in 'thermal events' as a means of articulating changes in materials and the consequences of such changes in terms of the laws of thermodynamics. The analysis of thermal events may be

approached in two different ways, which intrinsically overlap. Either the analysis is designed to measure specific properties of a material, such as heat capacity, enthalpy, entropy, free energy, with high precision and accuracy at predefined temperatures and conditions, or thermal properties may be studied over a temperature range using a controlled temperature program. Thus, calorimetry is the measurement of heat changes that occur during a process, whereas thermal analysis measures a property of a sample as a function of temperature. It is evident that calorimetry and thermal analysis are not synonymous terms. However, even with these simple definitions of both types of techniques, it is difficult to classify a

Nomenclature in TA continues to be a controversial issue. The International Confederation of Thermal Analysis and Calorimetry (ICTAC) has essentially overseen developments in this domain over some four decades. However, practical nomenclature is determined by general acceptance, and a variety of factors affect this. The nomenclature recommendations released by Hemminger and Sarge in 2001, subsequent to widespread and intensive discussion with thermal analysts worldwide, have yet to be formally accepted by ICTAC. Likewise, modifications to the '*Recommendations for Reporting Thermal Analysis Data*' – as reported in the late 1960s and early 1970s – have yet to be approved by ICTAC. In short, developments on these issues have been severely curtailed by a lack of international agreement on the core principles.

Calibration of TA instrumentation and development of standards for calibration continue to be administered by ICTAC in conjunction with ASTM. The standardization Committee of ICTAC has certified a range of materials for temperature calibration of TA systems, and in addition, standards for calibration of 'mass' (known as 'Class M Standards') are available for this purpose. A range of certified reference materials are available for enthalpy calibration in DSC. Temperature calibration for TMA and dynamic mechanical analysis (DMA) is effected by using disks of pure metals (silver, aluminum, and tin) separated by alumina disks. Load or force calibration for DMA is a complex process involving the use of calibrated weights. Temperature calibration for DETA is effected by measuring the melting transition of benzoic acid and dielectric calibration is

simply achieved by measuring the dielectric constant of the empty cell. Developments in the area of calibration in TA involve increasing the range of calibration materials available, which involves exhaustive 'round-robin' assessment exercises of potential materials.

Developments in TA *per se* depend on the continuity of ICTAC and European Society for Thermal Analysis and Calorimetry conferences and the continuity of the publications on TA and calorimetry. Also, the future of TA (and calorimetry) depends on tertiary education institutions providing courses in these areas, so as to ensure a future supply of trained thermal scientists.

**See also: Thermal Analysis:** Temperature-Modulated Techniques; Coupled Techniques; Sample-Controlled Techniques; Nonbasic Techniques.

## Further Reading

- Brown ME (2001) *Introduction to Thermal Analysis – Techniques and Applications*. Amsterdam: Kluwer.
- Duna JG (2002) Recommendations for reporting thermal analysis data. *Thermochimica Acta* 390: 1–3. Amsterdam: Elsevier.
- Haines PJ (ed.) (2002) *Principles of Thermal Analysis and Calorimetry*. Cambridge, UK: Royal Society of Chemistry.
- Hemminger W (ed.) (2000). *Thermal Analysis & Calorimetry – Towards the New Century*. *Thermochimica Acta* 355. Amsterdam: Elsevier.
- Hemminger W and Sarge SM (1998) In: Brown ME (ed.) *Handbook of Thermal Analysis and Calorimetry*, vol. 1. Amsterdam: Elsevier.

## Temperature-Modulated Techniques

**J O Hill**, La Trobe University, Wodonga, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

It has always been recognized that there is both a subtle difference and a synergy between 'thermal analysis' and 'calorimetry' and the associated groups of experimental techniques. There has been generations of interest in 'thermal events' as a means of articulating changes in materials and the consequences of such changes in terms of the laws of thermodynamics. The analysis of thermal events may be

approached in two different ways, which intrinsically overlap. Either the analysis is designed to measure specific properties of a material, such as heat capacity, enthalpy, entropy, free energy, with high precision and accuracy at predefined temperatures and conditions, or thermal properties may be studied over a temperature range using a controlled temperature program. Thus, calorimetry is the measurement of heat changes that occur during a process, whereas thermal analysis measures a property of a sample as a function of temperature. It is evident that calorimetry and thermal analysis are not synonymous terms. However, even with these simple definitions of both types of techniques, it is difficult to classify a



technique such as differential scanning calorimetry (DSC) either as a calorimetric or a thermal analysis technique, since in DSC heat changes are measured as a function of temperature and hence DSC clearly epitomizes the synergy between calorimetry and thermal analysis. The present article concentrates on the historical development of DSC and, more particularly, on the development of modulated temperature DSC (MTDSC), together with selected applications and future possible enhancements. Also, other temperature modulated thermal analysis techniques are briefly addressed.

## Development of DSC

Historically, DSC is a development of differential thermal analysis (DTA) and both techniques have a common origin in the measurement of temperature. The fundamental concept of both techniques is simple—to measure thermal changes in a sample relative to a thermally inert reference as both are subjected to a controlled temperature program. In classical DTA, the temperature difference between sample and reference is measured as a function of temperature: in classical DSC, the energy difference between sample and reference is measured as a function of temperature. Hence, DSC is simply ‘quantitative DTA’, or more precisely, DSC is a combination of DTA and adiabatic calorimetry. DSC is the more recent technique and was developed for quantitative calorimetric measurements over a wide temperature range from subambient to 1500°C. DTA is not appropriate for such precision measurements and has been progressively replaced by DSC, even for high-temperature measurements, as the major thermal analysis/calorimetric technique. DSC is a ‘differential calorimeter’ that achieves a continuous ‘power compensation’ between sample and reference.

Three major types of DSC have emerged that are classified as ‘scanning’, ‘isoperibol’, and ‘twin calorimetric’. The scanning type effects a power compensation between two separately heated calorimeters, the latter two types rely on measurement of the heat flux of two calorimeters subjected to a single heating system but differing in terms of the positions of the controlling temperature sensors. The heat-loss problem is minimized in DSC by the relatively short experimental time frame and the differential nature of the measurements reduces this phenomenon further. Thus, DSC precision is similar to that of the classical Nernst adiabatic calorimeter over the temperature range 150–450 K. However, the DSC operational range has now been effectively extended to 1500°C with concomitant precision and accuracy. DTA remains the technique of choice for the qualitative

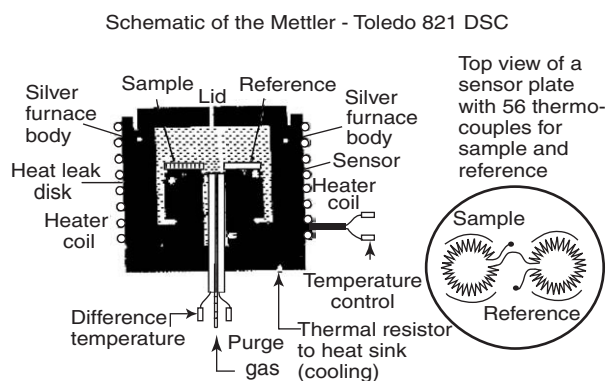
characterization of materials and for thermal measurements in excess of 1500°C – at which temperatures severe design constraints are imposed on the equipment. Both DTA and DSC are most commonly applied to study the thermal properties of solids, but with more sophisticated experimental techniques liquids can also be studied. Sample sizes in the milligram range are typical together with experimental timeframes in minutes and these features have been paramount in the popularization of DSC coupled with its unrivaled application range.

The practical distinction between DTA and DSC is simply related to the nature of the signal output from the equipment. For DTA, this is proportional to the temperature difference between the sample and a thermally inert reference when both are subjected to the same temperature program. For DSC, this is proportional to the difference in thermal power between the sample and the inert reference. For both DTA and DSC, the classical temperature program is a linear temperature change with respect to time. However, more complex programs can be invoked by different heating and cooling rates with isothermal periods. A common example is ‘stepwise heating’, which can be used to detect the onset of melting under quasi-isothermal conditions. A most important innovation in terms of temperature programming has been to overlay the linear heating change with a regular modulation – thereby creating MTDSC. In the original form of the technique described by Reading in 1993, the modulation was sinusoidal. Other forms of modulation have since been introduced, such as square wave and saw-tooth. MTDSC has been shown to have many advantages over conventional DSC, including enhanced sensitivity and resolution and the ability in some cases to resolve multiple thermal events.

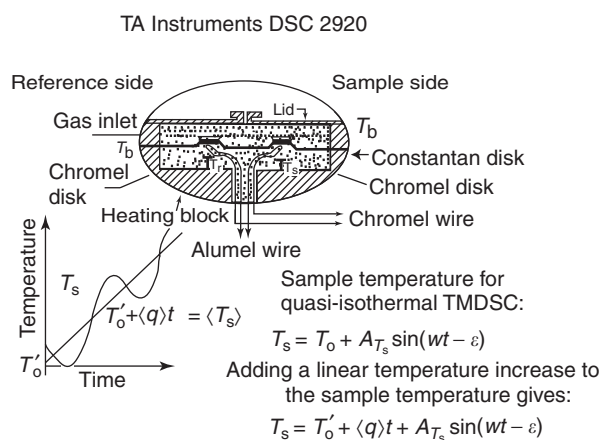
Three developments of calorimetry were thus combined in the last decade to dramatically enhance the capabilities of thermal analysis techniques and hence the study of the thermal properties of materials. These were the high precision of conventional adiabatic calorimetry, the speed of operation, and small sample size of DSC and the measurement of frequency dependence of thermal events and thus the MTDSC system evolved. ‘Modulation’ is perhaps the most significant development with respect to thermal analysis techniques paralleling in significance the ‘Fourier transform’ development of infrared spectroscopy.

## Development of MTDSC

The essential comparison between conventional DSC and MTDSC is shown schematically in **Figures 1** and **2**.

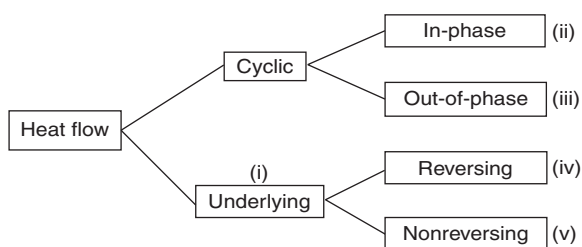


**Figure 1** Schematic of a classical heat flux DSC. (Reproduced with permission from Bernhard W (2000). Temperature modulated calorimetry in the 21st century. *Thermochimica Acta* 355: 43–57.)



**Figure 2** Schematic of a temperature modulated DSC. (Reproduced with permission from Bernhard W (2000) Temperature modulated calorimetry in the 21st century. *Thermochimica Acta* 355: 43–57; © Elsevier.)

In conventional DSC (**Figure 1**), the multiple temperature sensors average the calorimeter temperatures. In MTDSC (**Figure 2**), the linear heating program of conventional DSC is modulated by superimposing a periodic wave form of small amplitude on the linear temperature program. The most commonly applied wave form is a sine wave, as shown by the equations for sample temperature in **Figure 2**. Essentially, MTDSC provides periodic ‘heating’ and ‘cooling’ within each cycle but the overall effect is the same as in conventional DSC—a linear change in average temperature with time. The resultant heat flow signal is a composite of the response to a thermal event occurring in the sample and the response to the underlying heating program. Four-component signals thus derive from an MTDSC experiment, as shown schematically in **Figure 3** and



**Figure 3** The component output signals from TMDSC. (Reproduced with permission from Brown ME (2001) *Introduction to Thermal Analysis – Techniques & Applications*, p. 62. Amsterdam: Kluwer Academic.)

described as (i) the average or underlying signal, equivalent to DSC, (ii) the in-phase cyclic component, (iii) the out-of-phase component, (iv) the reversing component to the underlying heat flow, and (v) the nonreversing heat flow. Two of these components are interrelated – the reversing component (iv) is derived from multiplying the heat capacity of the sample by the heating rate and subtraction of (iv) from the underlying signal (i) gives (v) the nonreversing heat flow. Hence, in the simplest terms, MTDSC is a description of the heat flow into the sample resulting from the sinusoidal modulation of the temperature program. Two properties of the sample can be investigated by MTDSC – the heat capacity that is directly related to the ‘reversing component’ and a kinetically hindered thermal event that is related to the ‘nonreversing’ component. Conventional DSC provides only a measure of the total ‘heat flux’ into a sample as a function of temperature whereas MTDSC allows the heat capacity and kinetic components to be separated. However, the conditions of an MTDSC experiment are more critical than in conventional DSC since the selection of the period and amplitude of the modulation need to be selected in addition to the underlying heating rate. Sample size and the period of modulation influence the ability of the sample to follow the temperature modulation and hence the apparent value of the heat capacity. The underlying heating rate can be set to zero, in which case measurements are carried out under quasi-isothermal conditions.

Alternative modulation functions and data analysis procedures have been applied in MTDSC. Different instrument manufacturers have applied square wave or saw-tooth modulation, coupled with Fourier transform analysis methodologies. Alternative modulation functions provide some additional advantages – for example, a square wave function ensures that a steady state is achieved over an isothermal plateau since the signal during such a period is the nonreversing contribution. The amplitude provides a measure of the reversing signal. However, there are

some disadvantages of these alternative modulation functions, accuracy and resolution may be compromised since not all of the necessary data are made available.

There are many advantages of MTDSC over conventional DSC. Use of the former with appropriate frequencies and amplitudes allows separation of reversing processes, such as glass transitions, from nonreversing processes such as relaxation endotherms or cure processes. With MTDSC, baseline curvature of the cyclic signal is usually insignificant, thereby making it easier to distinguish between baseline effects and real transitions. The signal-to-noise ratio of the cyclic measurement of heat capacity is generally greater since all drift or noise at frequencies other than that of the modulation is ignored by the Fourier transform analysis. In general, in MTDSC, resolution of thermal events is improved because very low underlying heating rates are used.

## Applications of TMDSC

It is generally acknowledged that DSC is the preeminent thermal analysis technique and that it has progressively become the established technique for the study of the thermal behavior of polymeric materials. Conventional DSC links 'thermal power' to heat capacity and its integral to energy and entropy. Thus, DSC has primarily been applied to determine heat capacities of materials. Heat capacity is a primary thermodynamic parameter because of its intrinsic importance in materials characterization and its relationship to other parameters such as enthalpy, entropy, and Gibbs free energy. Conventional DSC is able to determine heat capacity to an uncertainty of 1–2%; MTDSC is able to measure heat capacity to an uncertainty of less than 1%, with reproducible reliability. Thus, fundamentally, MTDSC is able to measure thermodynamic quantities with uncertainties approaching those or equaling those of adiabatic calorimetry. It is the 'temperature modulation' feature of TMDSC that has confirmed this technique as the most versatile and most reliable of the thermal analysis techniques. Its versatility is further qualified by its ability to characterize the thermal behavior of materials without the need to have a detailed knowledge of the fundamental theoretical principles that underscore the basis of the technique. As for DSC, calibration of TMDSC is essential and this has presented a challenge. At present, the nematic to smectic-A transition in a cyanobiphenyl liquid crystal (80CB) has been suggested for calibration of TMDSC.

Five components of the MTDSC signal have been identified and these are conveniently shown by the separated MTDSC signal for a run on quenched

poly(ethylene terephthalate) (PET) in Figure 4. The difference in the Cp profiles corresponding to 'reversing Cp', 'kinetic Cp', and 'non-reversing Cp' are particularly striking and reveal the ability of MTDSC to provide 'complete' heat capacity characterization of a polymer sample.

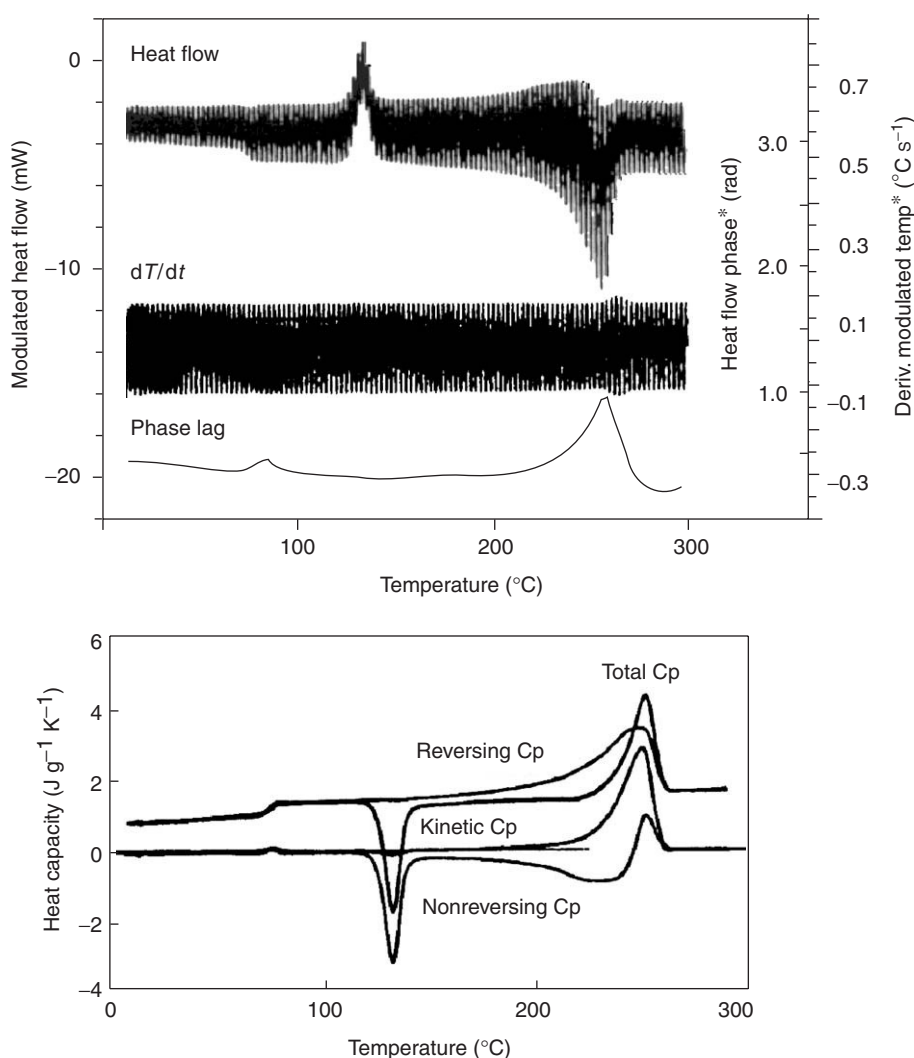
MTDSC has major advantages over conventional DSC in terms of differentiating overlapping thermal events. Figure 5 shows the MTDSC curves for a PET-acrylonitrile/butadiene/styrene (PET-ABS) blend. The total heat flow profile for the blend and its separation into the reversing and nonreversing components is shown. The glass transition temperature for PET is around 65°C and that for ABS is at 105°C. Crystallization of PET is revealed as a peak in the nonreversing curve and is also around 105°C, which would mask the glass transition of ABS in conventional DSC.

A more detailed theoretical treatment of the relationship between heat capacity and the reversing component of MTDSC shows that a 'reversing time correction factor' has to be included, which depends on mass and sample. By using multiple frequencies for modulation, the required correction can be made in a single MTDSC run, which further indicates the supremacy of TMDSC in heat capacity measurements.

Applications of TMDSC in general rely on the enhanced precision of the technique compared to conventional DSC as attained by the maintenance of a steady state and a negligible temperature gradient within the sample throughout the run and hence it is possible to measure heat capacity quasi-isothermally. In order to achieve the highest precision in TMDSC, the theory of the technique indicates that a high heating rate should be applied together with a large sample mass. These conditions lead to rapid analysis of large samples and give a further advantageous feature of DSC.

TMDSC is routinely applied to determine glass transitions of polymers. TMDSC offers quantitative investigation of polymer glass transitions as a function of 'thermal history', since glass transition kinetics vary with the enthalpy of the glass. Further, the ability of MTDSC to separate reversing and nonreversing processes during phase transitions offers quantitative investigation of cold crystallization and melting with superheating transitions. MTDSC therefore offers unparalleled opportunities for the quantitative analysis of the thermal properties and thermal behavior of materials, particularly synthetic polymeric materials.

A summary selection of recent specific applications of TMDSC is given here to illustrate the present and future potentials of the technique. Reactions involving partial diffusion can be studied by TMDSC such as



**Figure 4** Component TMDSC signals for a quenched PET sample. (Reproduced with permission from Brown ME (2001) *Introduction to Thermal Analysis – Techniques & Applications*, p. 63. Amsterdam: Kluwer Academic.)

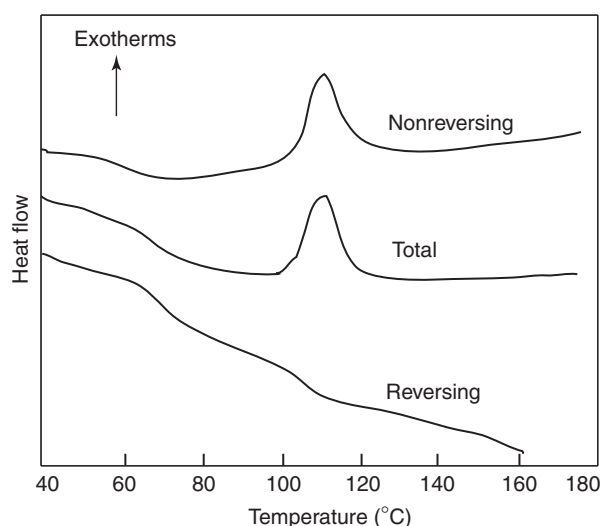
epoxy resin curing. The glass transition temperature is measured as a function of the conversion fraction. TMDSC is able to separate the glass transition and the melting transition in interpenetrating polymer networks prepared by in situ copolymerization. Additionally, an estimate of the degree of crosslinking in the copolymer is possible from the composite thermal data together with an estimate of the glass transition temperature of the copolymer phase.

The Wunderlich ATHAS laboratory has investigated other types of modulation, including 'saw-tooth' and 'square' and has shown that such modulations are likely to be incorporated into future generations of TMDSC systems for enhanced precision of measurement of heat capacity and glass transitions.

Numerous TMDSC applications were presented at the 8th European Symposium on Thermal Analysis & Calorimetry (ESTAC 8) in Barcelona, 2002. TMDSC

has been applied to study the morphology and recrystallization kinetics of nickel sulfide. Such information is of significance in understanding the failure of toughened glass panels when subjected to external stresses. TMDSC is shown to be ideal for rationalizing the melting behavior of polytetrahydrofurans and their blends. The PTHF oligomer shows two endothermic peaks – the lower peak shows temperature dependence on molar mass whereas the higher peak is independent of molecular mass of the polymer. The order–disorder transition behavior of triblock copolymer gels, swollen with paraffinic oils, has been studied by TMDSC. The order–disorder transition temperature of these gels is detected, reproducibly, as an abrupt change in heat capacity of the sample and is dependent on the polymer content of the gel and on the molecular weight of the oil. Thus, TMDSC assists in the understanding of the





**Figure 5** TMDSC curves for a PET-ABS blend. (Haines PJ (ed.) (2002) *Principles of Thermal Analysis & Calorimetry*. Cambridge: Royal Society of Chemistry; reproduced by permission of The Royal Society of Chemistry.)

mechanism of order-disorder transitions in triblock copolymers swollen with solvents. TMDSC has been applied to study 'intelligent' polymer systems, which show major changes in properties with small physical or chemical stimuli. The kinetics of 'demixing' (shrinking) and 'remixing' (swelling) are important considerations in possible applications of these systems. In this context, it is shown that TMDSC is ideal for the characterization of the kinetics of phase separation in partially miscible polymer blends and polymer solutions. 'Mixing/demixing' phenomena are manifested as large excess contributions to the apparent heat capacity signal of the system on the time scale of the temperature modulation that enable the real-time monitoring of demixing, remixing, and vitrification processes occurring. TMDSC thus allows investigation of nanoscale miscibility phenomena in water-soluble polymers and related hydrogels. TMDSC has been shown to be valuable for studying the deconvolution of 'nonreversing' and 'reversing' heat flow effects in reacting polymer systems. TMDSC has been further extended to study temperature-induced phase separation in polyethylene/polyethersulfone blends. TMDSC has also been applied to model the effect of additives on the cure kinetics of epoxy resins and for in situ detection of reaction-induced phase separation. The cure temperature is crucial in determining the final, phase-separated morphology and impact properties of such resins. TMDSC, in conjunction with TG, has been applied to characterize elastomers. From the component pyrolysis temperatures, TMDSC can be used to quantitatively determine the individual polymer

compositions. TMDSC has also been applied to study the vitrification, devitrification, and dielectric relaxations associated with the nonisothermal curing of an epoxy-amine system. Such systems are characterized by continuous heating transformation cure diagrams, which are readily obtained by TMDSC. The cold crystallization and melting of PET and poly(ethylene-2,6-naphthalene dicarboxylate) have been studied by TMDSC with a net zero heating rate. This application is typical of routine TMDSC analysis of synthetic polymer systems. TMDSC has also been applied to study high-temperature polymeric materials, such as bismaleimide resins, which find extensive use as light-weight fiber-reinforced structural composites and electrical insulators. Finally, maximum precision in TMDSC is obtained with very slow underlying heating rates that allow an adequate number of temperature modulations over a region of changing heat capacity such as a transition region. This condition has recently been achieved by the new DSC cell and 'Tzero' technology, developed by TA instruments (USA). In this new system, the thermal capacitance and resistance of the DSC measuring system are effectively eliminated and this feature, coupled with overall improved temperature control capabilities, has made much shorter modulation periods practical. Hence, 'fast TMDSC' is now a reality.

It is apparent from the rapid routine adoption of DSC for the characterization of materials, particularly synthetic polymers, that TMDSC with its additional advantages, is currently recognized as the most versatile analytical technique in materials science. It is inevitable that applications of TMDSC will dominate the thermal analysis and calorimetry literature in the current century.

## Other Temperature Modulated Thermal Analysis Techniques

In terms of 'modulated thermal analysis techniques', TMDSC clearly dominates the group. However, there is limited literature on temperature modulated thermogravimetric analysis (TMTGA) and temperature modulated thermomechanical analysis (TMTMA). Modulation principles have been applied to some less common thermal analysis techniques such as DMA and thermally stimulated current analysis and these developments will be briefly addressed here. It appears that the major development in thermal analysis in the next decade will be in the temperature modulated domain.

TMTGA has been developed by TA Instruments and essentially provides a kinetic analysis of solid-state

decomposition processes of inorganic salts, such as carbonates, sulfates, and oxalates. Thermogravimetry has long been applied to study the kinetics of a wide range of decomposition processes—it appears that TMTGA provides a refinement and an extension of this application. In TMTGA, a sinusoidal temperature modulation is superimposed on the linear heating rate program used in traditional TGA. This modulated temperature stimulus results in an oscillatory response in the rate of mass loss, deconvolution of which via a real-time discrete Fourier transform gives the relevant kinetic parameters of the reaction under study. While many materials decompose in a single step, others decompose in a stepwise manner, involving sequential mass losses. These stepwise mass losses may be widely separated in temperature, which leads to long experimental TMTGA times. To overcome the latter disadvantage, TA Instruments have also developed dynamic rate ‘High-Res’ TG<sup>TM</sup> TGA, which essentially involves rapid heating during thermal events and slow heating between them and is thus a sample controlled thermal analysis technique. This technique, in conjunction with TMTGA, provides high resolution of thermal events with short experimental times. A further advantage of TMTGA is that the reaction activation energy is determined as a function of the extent of reaction and hence no prior knowledge of the form of the relevant rate equation is required. The variation of activation energy is indicative of the reaction decomposition mechanism. Thus, for a single reaction mechanism, the associated activation energy shows no variation with the percentage conversion, whereas variation of activation energy with extent of reaction is indicative of ‘*n*th order’ processes. In addition to traditional thermal decomposition studies of materials, TMTGA has considerable promise for the study of metal oxidation and coal combustion processes, for which reliable kinetic data have considerable commercial value. It is also particularly applicable to study autocatalytic decomposition processes of commercial polymers such as polystyrene and polytetrafluoroethylene. Thus, ‘temperature modulation’ of TGA has produced a similar ‘revolution’ in thermogravimetry compatible with ‘temperature modulation’ of DSC and has certainly qualified thermogravimetry as the major technique for derivation of the reaction kinetics of materials decomposition.

Thermomagnetometry is directly related to thermogravimetry and involves heating the sample in a constant applied magnetic field. Magnetic transformations in the sample are manifested as ‘apparent’ mass changes. Temperature modulated thermomagnetometry has been applied to study the nanophase crystallization characteristics of amorphous

alloys. Nanocrystalline magnetic alloys are obtained by partial devitrification of the amorphous precursor and have ‘soft’ magnetic properties with high saturation magnetic flux density and high permeability. The nanocrystalline phase is obtained from the amorphous precursor under controlled annealing conditions. The ‘controlled conditions’ and the study of ‘soft magnetic properties’ are provided by TMTM and hence this technique has a very promising future in alloy nanotechnology.

In the same way that applying a temperature modulation to DSC can separate thermally reversing thermal events such as a glass transition from thermally nonreversing events such as crystallization and curing, these principles can be similarly applied to thermalmechanical analysis (TMA). TMTMA allows for the separation of reversible dimensional changes due to expansion from irreversible dimensional changes such as stress relaxation and creep. As for TMA, measurements can be made under tension (films and fibers) or under compression (rods and blocks). Similarly, temperature modulated DMA has been applied to study the reversible melting of polymers. Essentially, temperature modulation of thermomechanical techniques provides for enhanced resolution of polymeric transformations and hence has a promising future in this domain.

Finally, temperature modulated thermally stimulated current analysis has been applied to study dielectric materials via the separation of reversible pyroelectric effects from nonreversing thermally stimulated discharges.

**See also:** **Thermal Analysis:** Overview; Coupled Techniques; Sample-Controlled Techniques; Nonbasic Techniques.

## Further Reading

- 8th European Symposium on Thermal Analysis & Calorimetry (ESTAC 8), August 2002. Abstracts cited: 6-02, 6-03, 6-P1, 6-P2, 6-P3, 6-P14, 6-P16, 6-P23, 6-P26, P&F-06: Nomen R (Symposium Chair), Institut Quimic De Sarria, Universitat Ramon Llull, Barcelona, Spain.
- Boller A, Jin Y, and Wunderlich B (1994) Heat capacity measurement by modulated DSC at constant temperature. *Journal of Thermal Analysis* 42: 307–330.
- Buckman N, Blaine R, and Dallas G (1998) Modulated thermogravimetry. *Chemistry in Australia*: 22–23.
- Haines PJ, Reading M, and Wilburn FW (1998) Differential scanning calorimetry. In: Brown ME (ed.) *Handbook of Thermal Analysis and Calorimetry*, vol. 1, ch. 5. Amsterdam: Elsevier.
- Kamasa P and Myslinski P (1999) Study of the thermal effects during microstructural changes of metastable alloys by temperature modulated thermomagnetometry. *Thermochimica Acta* 337: 51–54.



- Kwon YK, Androsch R, Pyda M, and Wunderlich B (2001) Multi-frequency sawtooth modulation of a power compensation DSC. *Thermochimica Acta* 367–368: 203–215.
- Montsarrat S, Roman F, and Colomer P (2003) Study of crystallization and melting region of PET and PEN and their blends by TMDSC. *Journal of Thermal Analysis and Calorimetry* 72: 657–666.
- Price DM (2000) Modulated temperature thermomechanical analysis. *Thermochimica Acta* 357–358: 23–29.
- Pyda M, Kwon YK, and Wunderlich B (2001) Heat capacity measurement by sawtooth modulated standard heat-flux DSC with sample temperature control. *Thermochimica Acta* 367–368: 217–227.
- Reading M (1993) Modulated differential scanning calorimetry – a new way forward in materials characterisation. *Trends in Polymer Science* 8: 248–253.
- Reading M, Elliott D, and Hill VL (1993) A new approach to the calorimetric investigation of physical and chemical transitions. *Journal of Thermal Analysis* 40: 949–955.
- Sharp EJ and Garn LE (1982) Use of low frequency sinusoidal temperature waves to separate pyroelectric currents from non-pyroelectric currents. *Journal of Applied Physics* 53: 8980–8987.
- Wunderlich B (2000) Temperature modulated calorimetry in the 21st century. *Thermochimica Acta* 355: 43–57.
- Wunderlich B, Jin Y, and Boller A (1994) Mathematical description of DSC based on periodic temperature modulation. *Thermochimica Acta* 238: 277–293.
- Wurm A and Merzlykov Schick C (1999) Temperature modulated dynamic mechanical analysis. *Thermochimica Acta* 330: 121–130.

## Coupled Techniques

J O Hill, La Trobe University, Wodonga, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

As the number of thermal analysis (TA) techniques increased and the application field widened, it became apparent that for a detailed investigation of materials, it is improbable that any single TA technique is able to provide sufficient information to characterize its thermal behavior. Complementary information is usually necessary, which may be derived from another TA technique or from a supplementary analytical technique. The type of analytical technique differentiates between ‘simultaneous thermal analysis (STA) techniques’ and ‘coupled techniques’. By inference, the latter group includes a wider range of techniques than does the former group. The challenge for both groups is to be able to correlate the data from the individual analyses. Essentially, STA is a combination of (usually) two TA techniques such as thermogravimetric analysis (TGA)–differential thermal analysis (DTA), TGA–differential scanning calorimetry (DSC), and involves measurements of two thermal properties on the same sample subjected to the same temperature program. Each of the two thermal properties may be monitored continuously or they may be monitored in a repetitive sequence to allow for the requirements of data capture. Coupled techniques include the ‘evolved gas analysis’ (EGA) group of techniques

and involve the ‘coupling’ of two different types of analytical techniques, such as TGA–gas chromatography (GC) or TGA–mass spectrometry (MS). However, with the introduction of new thermoanalytical techniques and the refinement of traditional analytical techniques, the ‘simultaneous’ group and the ‘coupled’ group have both expanded considerably and these are now generally regarded as powerful ‘complete’ analytical systems for the investigation of a wide range of materials.

A simple differentiation between ‘simultaneous’ and ‘coupled’ techniques is that for the former both techniques are TA techniques, whereas for the latter only one of the two is a TA technique. However, the identification of two groups derived from combining two analytical techniques has significance beyond that of semantics. In STA, it is necessary to refine the term ‘simultaneous’ and to differentiate it from two other types of measurement – ‘parallel’, in which different portions of a sample are examined using different techniques and ‘concurrent’, in which different portions of the same sample are contained in different crucibles in a single furnace and are hence subjected to a common temperature program. There are advantages of STA over ‘parallel’ and ‘concurrent’ simultaneous measurements. Fundamentally, it is difficult to correlate results obtained on different portions of a sample with different techniques due to ‘sample’ and ‘instrument’ factors, whereas, in general, results obtained on the same sample by two TA techniques simultaneously are reproducible. In general, STA may be viewed as yielding data that are more significant than the sum of the two sets of

- Kwon YK, Androsch R, Pyda M, and Wunderlich B (2001) Multi-frequency sawtooth modulation of a power compensation DSC. *Thermochimica Acta* 367–368: 203–215.
- Montsarrat S, Roman F, and Colomer P (2003) Study of crystallization and melting region of PET and PEN and their blends by TMDSC. *Journal of Thermal Analysis and Calorimetry* 72: 657–666.
- Price DM (2000) Modulated temperature thermomechanical analysis. *Thermochimica Acta* 357–358: 23–29.
- Pyda M, Kwon YK, and Wunderlich B (2001) Heat capacity measurement by sawtooth modulated standard heat-flux DSC with sample temperature control. *Thermochimica Acta* 367–368: 217–227.
- Reading M (1993) Modulated differential scanning calorimetry – a new way forward in materials characterisation. *Trends in Polymer Science* 8: 248–253.
- Reading M, Elliott D, and Hill VL (1993) A new approach to the calorimetric investigation of physical and chemical transitions. *Journal of Thermal Analysis* 40: 949–955.
- Sharp EJ and Garn LE (1982) Use of low frequency sinusoidal temperature waves to separate pyroelectric currents from non-pyroelectric currents. *Journal of Applied Physics* 53: 8980–8987.
- Wunderlich B (2000) Temperature modulated calorimetry in the 21st century. *Thermochimica Acta* 355: 43–57.
- Wunderlich B, Jin Y, and Boller A (1994) Mathematical description of DSC based on periodic temperature modulation. *Thermochimica Acta* 238: 277–293.
- Wurm A and Merzlykov Schick C (1999) Temperature modulated dynamic mechanical analysis. *Thermochimica Acta* 330: 121–130.

## Coupled Techniques

J O Hill, La Trobe University, Wodonga, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

As the number of thermal analysis (TA) techniques increased and the application field widened, it became apparent that for a detailed investigation of materials, it is improbable that any single TA technique is able to provide sufficient information to characterize its thermal behavior. Complementary information is usually necessary, which may be derived from another TA technique or from a supplementary analytical technique. The type of analytical technique differentiates between ‘simultaneous thermal analysis (STA) techniques’ and ‘coupled techniques’. By inference, the latter group includes a wider range of techniques than does the former group. The challenge for both groups is to be able to correlate the data from the individual analyses. Essentially, STA is a combination of (usually) two TA techniques such as thermogravimetric analysis (TGA)–differential thermal analysis (DTA), TGA–differential scanning calorimetry (DSC), and involves measurements of two thermal properties on the same sample subjected to the same temperature program. Each of the two thermal properties may be monitored continuously or they may be monitored in a repetitive sequence to allow for the requirements of data capture. Coupled techniques include the ‘evolved gas analysis’ (EGA) group of techniques

and involve the ‘coupling’ of two different types of analytical techniques, such as TGA–gas chromatography (GC) or TGA–mass spectrometry (MS). However, with the introduction of new thermoanalytical techniques and the refinement of traditional analytical techniques, the ‘simultaneous’ group and the ‘coupled’ group have both expanded considerably and these are now generally regarded as powerful ‘complete’ analytical systems for the investigation of a wide range of materials.

A simple differentiation between ‘simultaneous’ and ‘coupled’ techniques is that for the former both techniques are TA techniques, whereas for the latter only one of the two is a TA technique. However, the identification of two groups derived from combining two analytical techniques has significance beyond that of semantics. In STA, it is necessary to refine the term ‘simultaneous’ and to differentiate it from two other types of measurement – ‘parallel’, in which different portions of a sample are examined using different techniques and ‘concurrent’, in which different portions of the same sample are contained in different crucibles in a single furnace and are hence subjected to a common temperature program. There are advantages of STA over ‘parallel’ and ‘concurrent’ simultaneous measurements. Fundamentally, it is difficult to correlate results obtained on different portions of a sample with different techniques due to ‘sample’ and ‘instrument’ factors, whereas, in general, results obtained on the same sample by two TA techniques simultaneously are reproducible. In general, STA may be viewed as yielding data that are more significant than the sum of the two sets of

data – a synergistic effect operates. This effect is readily apparent – TGA can only detect thermal events associated with a mass change – so a phase change such as ‘melting’ is not detected by this technique; however, melting can readily be detected by DTA. Thus, a combination of TGA and DTA shows no sample mass change associated with the endothermic (melting) DTA peak. However, DTA (or DSC) alone is unable to differentiate between types of sample enthalpy changes such as melting and decomposition. A combination of TGA and DTA (or DSC) is able to produce such differentiation, since ‘decomposition’ is associated with a mass change. Thus, STA is able to characterize a thermal event both qualitatively and quantitatively. The synergistic effect operating in STA is readily apparent from an examination of a material that melts with decomposition. The overall thermal event is readily characterized by STA, whereas TGA and DTA alone can only partially define such a transformation. Other obvious advantages of STA are ‘time saving’ (only need to do one experiment instead of two), ‘sample saving’ (only requires one portion of a sample instead of two), and ‘capital cost savings’ (only needs one TA instrument instead of two). However, generally, the sensitivity of individual TA techniques is decreased on combination due to essential compromises in instrumental design. Also, it is technically difficult to combine TG with DSC primarily because of the limited temperature range of DSC and its associated sophisticated temperature programming capabilities. Essentially, the major advantage of STA is associated directly with the unambiguous correlation of two sets of TA information on the same sample subjected to the same temperature regime. This feature promotes STA for the examination of complex materials such as coals, minerals, ceramics, pharmaceuticals, and commercial polymers.

Other STA techniques include the thermoptometry systems, such as the combination of TGA with an optical microscope to give a thermomicroscopy system – of which there are many variants. Thermomicroscopy systems allow visual observation of thermal events such as melting, bubbling, creeping, shrinking, swelling, decrepitation, and corrosion. DSC and DTA can also be combined with microscopy. X-ray techniques, such as X-ray diffraction (XRD), have also been combined with TA techniques, particularly to identify solid reaction intermediates and products. The most promising recent development with respect to simultaneous techniques is micro-thermal analysis ( $\mu$ -TA) and this has unrivalled potential for using TA to investigate materials at the nanosubmicroscopic level.

Coupled techniques have attained enhanced popularity as quantitative EGA systems. Essentially, the

coupled analytical technique characterizes the products evolved from a TA investigation. The initial coupled technique was evolved gas detection (EGD) and involved the coupling of a TGA to a gas chromatograph. However, it was recognized that essentially EGD was a qualitative technique and that the ultimate aim of coupled techniques is to provide complete information on a thermal event by not only identifying mass changes and temperatures of thermal transitions but also by characterizing the products of a thermal decomposition. With the coupling of TGA to mass spectrometry, this primary aim was achieved and EGA systems rapidly developed. Thus, EGA provides an opportunity to obtain ‘online’ chemical information associated with a thermal event. EGA has been accorded the title of ‘a complete analysis system’ by instrument manufacturers in recognition of its primary features. However, advances in EGA depend very significantly on the many technical problems associated with the ‘coupling’ of the two techniques in question and this feature is perhaps a limiting factor on the range of EGA applications. It is certainly a primary technique for the characterization of the thermal behavior of materials such as coals and commercial polymers. Other types of analytical techniques may be coupled to a TA technique, such as a Fourier transform infrared (FTIR) spectrophotometer to give the EGA technique TGA–FTIR. A gas chromatograph may also be interfaced with this system to give TGA–GC–FTIR.

The present article focuses primarily on the most significant coupled techniques that envelope and extend the traditional STA techniques. The basic principles of these techniques will be described together with a range of associated applications.

## Simultaneous Thermal Analysis Techniques

The pioneering STA system was the ‘derivatograph’, developed by J Paulik and F Paulik in 1955. It involved the direct combination of TG with DTA to give TG–DTA. Over subsequent decades, the derivatograph was developed to incorporate dilatometry and EGA. Over the last decade, controlled rate methods were also included, thus making the derivatograph the most versatile of the STA techniques. STA instruments involve the mounting of a DTA head onto the thermobalance suspension or rise rod, which is connected directly to a microbalance. Modern TG–DTA instruments are capable of a resolution of the order of 1  $\mu$ g with sample masses ranging from 5 to 100 mg. As for all TA techniques, temperature calibration is essential for STA, particularly when

the TG data are used for kinetic analysis of reactions. Temperature calibration of an STA is conveniently achieved by reference to the melting temperatures of pure metals. The TG mass loss scale is usually calibrated by using the ICTAC Curie Point (magnetic) standards. As previously noted, STA has many advantages over independent TG and DTA. Since sample mass is continuously monitored, DTA peak measurements can be directly correlated with sample mass after allowing for lower temperature mass losses such as dehydration. Also, peak area measurements can be correlated with single process thermal events in cases where the corresponding TG curves show that no decomposition or sublimation is taking place. This is particularly advantageous in the high-temperature studies of alloys and ceramics.

STA has been widely used in coal studies. Combustion of coal is highly sensitive to experimental conditions and TA experiments are usually associated with low reproducibility. STA markedly enhances coal combustion reproducibility and this feature accounts for its popularity as an analytical technique for studies of the thermal behavior of coals. In simple terms, the initial drying process of a coal can easily be separated from the subsequent oxidation, which in STA can be assigned directly to a 'dry weight'.

Simultaneous DTA can often provide an explanation of unusual features of a TG curve. For example, abrupt changes in the rate of mass loss can occur simultaneously with sample melting or, conversely, the TG curve may show that melting occurs with decomposition. The latter feature is particularly prevalent with organic compounds. The rate of reaction in a solid-state system or the rate of a solid-gas reaction may be associated with the intermediate fluid phase, the formation of which is revealed by an arrest of mass loss in the TG profile and the corresponding temperature range may be determined from the simultaneous DTA profile. Irregular sharp features on a TG curve due to 'bubbling' in a viscous melt are usually seen as sharp spikes on the corresponding DTA curve. Thus, STA has major advantages in the characterization of melting processes and this explains its value in widespread TA studies of commercial polymer systems. Phase diagrams for alloy systems are readily derived by high-temperature STA studies and 'molten solvent' applications have been widely studied by STA, using specially designed stainless steel instrumental accessories. A further major application of STA has been directed at catalyst composition and activity – particularly low-temperature catalytic activity and the associated economic consequences.

Over the last decade, DSC has progressively replaced DTA as the major TA technique and hence

STA has been extended to include TG–DSC. However, the complexity of power-compensated DSC sensors and their operation to a relatively restricted temperature range has precluded combination of power-compensated DSC with TG. Hence, TG–DSC involves the combination of TG with heat-flux DSC. The advantages of TG–DSC include all those associated with TG–DTA with the addition of quantitative calorimetric data on thermal events. Also, DSC is particularly suitable for the study of 'subambient' thermal events and this is particularly important for the study of systems containing moisture or other volatiles. Hence, subambient TG–DSC can be applied to study the multiple solid-state phase transitions in ammonium nitrate. It is known that traces of moisture can influence these transitions both in terms of 'extent' and 'temperature' – STA is able to determine the moisture content associated with each thermal event and hence an unambiguous TA of ammonium nitrate is achieved. A further application of TG–DSC is the determination of vaporization enthalpies of liquids. These are generally difficult to obtain by other methods due to 'partial' or 'incomplete' vaporization. In TG–DSC, the extent of vaporization is determined directly from the TG profile and the corresponding vaporization energy is obtained from the area of the associated DSC endotherm. For this type of investigation, the sample is sealed in a container with a small pin-hole such that a smooth effusion of vapor is achieved, corresponding to a constant rate of mass loss. Calibration of the STA system is necessary with a sample of well-defined vaporization enthalpy. TG–DSC can also be applied to determine sublimation enthalpies – using benzoic acid or ferrocene as calibration standard. A notable application in this context is the determination of the sublimation enthalpies of volatile coordination compounds, particularly those of the later transition metals. In terms of studying the volatility of (toxic) organic compounds, STA has a promising future as a valuable technique in promoting 'green chemistry'.

Thermooptometry is the generic term relating to the observation or measurement of an optical property of a material as a function of temperature. Samples can be viewed by reflectance or by transmission and the light intensity in each case can be measured by a photocell, thereby quantifying the technique. The earliest system was the well-known hot-stage microscope, which has had widespread application and is still used extensively. The modern equivalent is a thermomicroscopy system, which consists of a microscope with a hot and/or cold stage, a sample holder, gaseous atmosphere control (including vacuum), light sources, and a system for processing



and recording visual observations together with the sample temperature. Visual observation of a sample during a thermal event is often useful in characterizing the event in terms of its 'completeness', shrinkage, swelling, creeping, bubbling, cracking, and extent of surface activity. Further, many solid-solid transitions are detectable by a color change or volume change – which can be monitored by thermomicroscopy. Hot-stage microscopy has been used to determine the ignition temperature of pyrotechnic compositions. Modifications of the basic system were necessary, including mounting the sample between two microscope cover glasses that rested on a heating block within an atmosphere-controlled chamber. The block contained a sapphire window for light transmission and a platinum resistance thermometer was used for measurement of sample temperature and for the control of the heating/cooling program. The transmitted light intensity was measured by a silicon photodetector. A subambient facility further extends the operating temperature range of hot-stage microscopy systems from  $-180^{\circ}\text{C}$  to  $600^{\circ}\text{C}$ .

Simultaneous thermomicroscopy/thermal analysis techniques have been developed. A TG-thermomicroscopy system has been developed for a study of the pyrolysis and combustion of coals. The microscope is combined with a video camera to continuously monitor the thermal events involved and the sample holder for microscopy is attached to the microbalance by an aluminum oxide capillary. Events such as partial fusion, swelling, and internal structural changes associated with coal can be observed microscopically and correlated directly with noted associated mass losses. The system is particularly suitable for coal analysis involving rapid heating rates and atmosphere switching.

Conventional DSC has been modified to carry out simultaneous measurements of the intensity of light reflected from the sample. Changes in sample surface characteristics may not be associated with measurable enthalpy changes but may be manifested by changes in reflectance. A binocular microscope was employed so that one eyepiece photographs the actual appearance of the sample while the photodetector is the other eyepiece. Changes in the intensity of the reflected light are cumulative so that the net curves recorded are related to DSC curves as integral to derivative.

Other spectroscopic techniques have been combined with TA techniques. Thermomicroscopy has been combined with FTIR spectrophotometry and by using a hot stage with a DSC cell incorporated, simultaneous DSC-FTIR is achieved. However, with these later combinations, thin sample films are required that give poor DSC sensitivity. More flexibility

can be obtained by using an IR microscope accessory to collect reflectance spectra from a sample mounted in a DSC cell. DSC-FTIR is particularly suitable for studying curing processes in an amine-cured epoxy material and to study changes in poly(ethylene)terephthalate (PET) through its glass transition and melting regions.

TG, DTA, and DSC have been combined with XRD to give enhanced thermooptometry techniques. These simultaneous techniques are particularly advantageous for the investigation of the thermal behavior of materials in the solid state. A TG may be coupled with an XRD by mounting a top-loading thermobalance onto a diffractometer such that the sample is coincident with the X-ray beam. Simultaneous TG-XRD has been used to investigate the complex series of thermal events associated with the hydrogen reduction of tungstic oxide. The associated TG curve is featureless and difficult to interpret but the XRD spectra indicate that the reduction proceeds via the production of many intermediate tungsten-oxygen phases and through two tungsten modifications.

XRD has also been coupled to a DSC and the resulting simultaneous system is particularly suited to the study of catalyst regeneration by indicating the nature of the intermediate phases and the energy transformations involved. However, a difficulty with this technique is that XRD requires slow temperature scans to maximize sensitivity, which decrease DSC sensitivity. This problem appears solvable by using a high-intensity synchrotron X-ray source. The latter system appears particularly suitable for a TA study of liquid crystals.

Hemminger and Felder-Casagrande have comprehensively reviewed the wide range of applications of thermooptometry, which range from an investigation of the hydration/dehydration of gypsum to form the hemi-hydrate 'plaster of paris' to the nucleation of graphite on the surfaces of natural diamond. Thermomicroscopy is also particularly valuable for the investigation of the thermal behavior of solid solutions, liquid crystals, explosives, and pharmaceuticals. The singular advantage of the simultaneous technique involved in these investigations is its ability to characterize a thermal transition, not only in terms of a mass or energy change, but also spectroscopically. It is this correlation that unambiguously defines the thermal behavior of the material under investigation.

A recent addition to the family of STA techniques is micro-thermal analysis ( $\mu$ -TA). This technique combines the imaging capabilities of atomic force microscopy (AFM) with a form of 'localized' TA and has the capability of imaging thermal behavior of

solids over an area of a few microns.  $\mu$ -TA has major potential as an analytical technique in the nanotechnology era. The sensor is a very fine, pointed platinum wire mounted on the tip of an AFM probe. The sample (or sample region) is heated at a linear rate by passing a current through the wire and the resistance of the wire sensor is used simultaneously to measure the sample temperature. Since the sensor has a small thermal mass, heating rates in excess of  $20^{\circ}\text{C min}^{-1}$  can be used. The enhanced imaging capabilities of the AFM are used to map the surface area of interest as a function of temperature. The use of a controlled temperature (microscope) stage allows the entire sample to be heated or cooled and a thermal probe is then not necessary. Standard AFM contact or non-contact probes can then be used to obtain surface images of high resolution.

There are several 'forms' of  $\mu$ -TA. Micro-DTA involves comparison of the temperature difference of the sample with respect to a thermally inert reference material, as in normal DTA. If the probe is loaded with a small force, microthermomechanical analysis results. The  $\mu$ -TA configuration can be used for rapid pyrolysis of a small area of the sample and the evolved gases can be analyzed 'online' by a GC-MS system. This mode is particularly valuable for the compositional analysis of synthetic polymers. Sinusoidal temperature programs have also been applied in  $\mu$ -TA leading to temperature modulated  $\mu$ -TA.

$\mu$ -TA is more difficult to calibrate than macro-TA techniques. Temperature calibration of  $\mu$ -TA is achieved by using powdered organic compounds of well-defined melting point, such as biphenyl or benzoic acid. Pure metals cannot be used for this purpose because the calibrant comes into direct contact with the platinum wire sensor. With organic calibrant materials, the probe is easily cleaned by heating in air. For synthetic polymers, a suitable calibrant is PET.

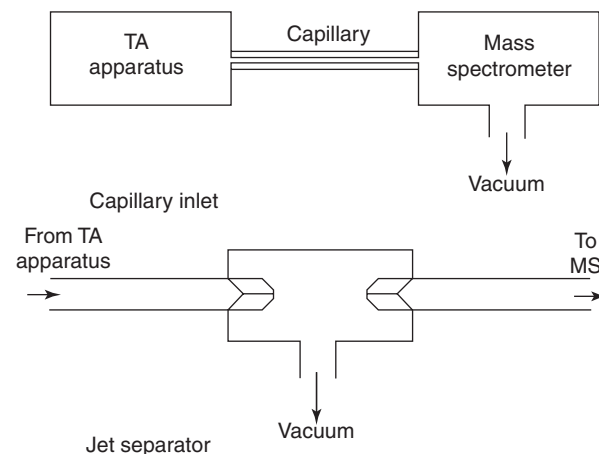
$\mu$ -TA offers a new analytical dimension in the materials and surface sciences. Typical applications relate to synthetic polymers, pharmaceuticals, microelectronics, cellular biology, and multilayered systems such as emulsion paints.

## Coupled Thermal Analysis Techniques

The coupled TA techniques specifically relate to EGA. Essentially, EGA characterizes the gases evolved during a TA investigation and hence augments the total information obtained. The most common EGA techniques are the coupling of TG and/or DTA with GC and/or MS. The coupling relies on a flowing purge gas stream through the system such that the evolved gases are swept into the

coupled analysis system. The most comprehensive EGA system is a TG-DTA-GC-MS, which combines the advantages of STA with the gas separation ability of GC and the identification ability of MS. TG-MS and TG-FTIR systems are also available commercially. For all EGA systems, the direct coupling of an analytical instrument to a TA system poses significant challenges and the interface has to recognize the different conditions associated with the individual instruments involved and the different time intervals required for sampling. For example, in TA-MS, a TA system usually operates at ambient pressure whereas an MS operates at high vacuum. The necessary coupling interface has to allow for the major pressure gradient between the TA and MS systems while simultaneously transferring a representative sample of the evolved gases from the TA system to the MS system. The most significant problem in this context is to overcome possible condensation of vapors in the interface. Heating of the interface may remedy this problem but may also cause decomposition of the evolved gaseous components (see Figure 1).

TA-FTIR is gaining popularity as a major EGA technique. In contrast to TA-MS, the bulk gas flow from the TA system passes through the FTIR gas cell so the interface does not change the flow rate and pressure of the purge gas – thereby overcoming the possible condensation problems associated with TA-MS. Full-range FTIR spectra can be recorded or spectral sections (windows) can be preselected. Output presentation is in the form of a plot of overall integrated intensity of the IR absorption as a function of temperature. However, the areas of the absorption peaks are dependent on the nature of the species contributing to the absorption; hence, unlike



**Figure 1** Schematic diagram of a TA-MS system showing the capillary inlet and jet separator components. (Reproduced with permission from Brown ME (2001) *Introduction to Thermal Analysis – Techniques and Applications*. Amsterdam: Kluwer.)



typical DTA or DSC peaks, the relative amounts of species cannot be determined. Full or partial IR spectra can be obtained at points on the TA profile to assist the identification of gaseous species evolved at that stage. In cases involving complex mixtures of evolved gases, interfacing a GC system between the TA and FTIR systems refines the overall analysis by effecting online separation of the gaseous components prior to identification.

Fundamentally, EGA systems require powerful software control and data analysis systems, which in addition to controlling the overall hardware and displaying and analyzing output data, may also include libraries of MS or FTIR spectra to assist identification of the gaseous products of a thermal decomposition. However, EGA represents the most comprehensive group of thermoanalytical systems available for the characterization of the thermal behavior of materials. Hence, EGA applications are widespread and diverse. Coordination chemistry has been considerably enhanced by TA studies of a wide range of metal complexes. The emphasis of these studies has been to derive thermal decomposition mechanisms for such metal complexes. Prior to the development of EGA, such mechanisms were suspect, since the volatile decomposition products and the solid intermediates were, in general, not identified directly but characterized indirectly on the basis of mass loss data. This procedure is flawed when multiple gaseous decomposition products and non-stoichiometric intermediates are involved. In this context, even 'simple' inorganic compounds such as copper sulfate pentahydrate are associated with complex thermal decomposition mechanisms. TG-MS is particularly well suited to studies of the thermal decomposition of metal complexes, particularly those of biochemical significance.

TG-DTA-MS has obvious synthetic polymer applications. TA-MS has been applied to study the thermal behavior of homopolymers, copolymers, polymeric blends, composites, residual monomers, solvents, additives, and toxic degradation products. In the latter context, HCl evolution from heated poly(vinyl chloride) materials is readily quantified by TA-MS and such data are of major significance in the design of fire-resistant polymeric materials. Pyrotechnic materials have been studied by TA-MS. A complex sequence of thermal events relates to the decomposition of these materials involving interactions between the nitrocellulose, perchlorate, and metal components with periodic release of carbon dioxide and oxygen. Only by EGA is it possible to rationalize the thermal behavior of such materials. TA-FTIR has also been applied extensively to study the thermal characteristics of synthetic polymers

such as polystyrene. A particular advantage of TA-FTIR in this context is that the output data can be represented as a three-dimensional surface with coordinates absorbance, wavenumber, and temperature. Thus, specific decomposition products, such as water and carbon dioxide, are continuously monitored quantitatively. Various TA-EGA systems have been used to study materials of environmental significance such as contaminated soils, waste (polymeric) products, and packaging materials. Such studies are important in the context of recycling strategies.

Finally, it should be noted that a major disadvantage of TA-EGA systems, particularly TG-MS and TG-FTIR systems, when applied to samples such as synthetic polymers that degrade to produce a variety of complex fragments simultaneously, is the inability to separate the decomposition products prior to analysis. The resulting analytical data are invariably difficult to interpret unambiguously. However, recently, Lever *et al.* have described the use of online sorbent tubes as an alternative to cold traps for condensing the emergent volatile decomposition products. Subsequent analysis of these products is affected by GC-MS by coupling the collection tube to the inlet of a GC-MS system and partially or completely thermally desorbing the decomposition products. A new EGA technique has hence been created – 'evolved gas collection'.

**See also:** Thermal Analysis: Overview; Temperature-Modulated Techniques; Sample-Controlled Techniques; Nonbasic Techniques.

## Further Reading

- Brown ME (2001) *Introduction to Thermal Analysis – Techniques and Applications*, chs. 7 and 8. Amsterdam: Kluwer.
- Haines PJ (ed.) (2002) *Principles of Thermal Analysis and Calorimetry*, ch. 6. Cambridge: Royal Society of Chemistry.
- Hemminger W and Felder-Casagrande S (1998) In: Brown ME (ed.) *Handbook of Thermal Analysis and Calorimetry*, vol. 1. Amsterdam: Elsevier.
- Paulik F and Paulik J (1981) Simultaneous thermoanalytical examinations by means of the derivatograph. In: Svehla G (ed.) *Comprehensive Analytical Chemistry*, vol. X11(A). Amsterdam: Elsevier.
- Raemaekers KGH and Bart JCJ (1997) Application of simultaneous thermogravimetry – mass spectrometry in polymer analysis. *Thermochimica Acta* 295: 1–58.
- Roedolf D, Lever TJ, Price DM, and Schaap K (2000) Proceedings of the 12th ICTAC, Paper P1-9.
- Warrington SB (1992) In: Charsley EL and Warrington SB (eds.) *Thermal Analysis – Techniques and Applications*, Special Publication 117. Cambridge: Royal Society of Chemistry.

## Sample-Controlled Techniques

J O Hill, La Trobe University, Wodonga, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Sample controlled thermal analysis (SCTA) is the generic title for a group of thermal analysis (TA) techniques in which the heating rate is varied so as to produce a constant rate of mass loss. This title has replaced the former title of 'controlled rate thermal analysis' (CRTA). The historical development of 'controlled rate' methodology has been comprehensively reviewed by Reading. SCTA essentially overcomes the problems associated with linear heating rates in conventional TA techniques that can cause significant temperature and pressure gradients in the sample. The judicious use of heating regimes, which are modified in some manner by the reaction rate, can greatly reduce these problems giving enhanced resolution in analytical and characterization investigations. Over the last decade, a family of 'controlled rate' techniques has emerged. The first of these, CRTA, is a method pioneered by Rouquerol and the Pauliks, which involves variation of the sample temperature to maintain a constant reaction rate. This condition essentially ensures pseudoequilibrium conditions are maintained throughout the sample. A constant rate of reaction is maintained by using the measured mass loss or evolved gas pressure as feedback to the furnace temperature controller. The resulting conditions for the analysis correspond to near isothermal and hence 'quasi-isothermal controlled rate thermal analysis' results. Similarly, if the pressure of the evolved gases is maintained constant during a TA experiment, 'quasi-isobaric controlled rate thermal analysis' results. The similarity between these techniques is immediately apparent. Both maintain the reaction rate constant and both control the pressure of the evolved gases in the system. Closely related techniques, which are particularly useful for deriving kinetic information, are those involving a 'temperature jump' or a 'rate jump' – leading to 'stepwise (iso)thermal analysis' (SIA). In the latter, upper and lower limits are set for the reaction rate and when the rate falls below the minimum preset value, the temperature is increased (at a generally fast linear rate) until the reaction rate exceeds the maximum set value. The temperature is then held constant until the reaction rate subsequently falls below the minimum. Further variations of SCTA have been developed by Parkes in which the

reaction rate for gas–solid reactions is controlled by programming the concentration of reactive gas whilst keeping the temperature constant. Also, further variants involve resolution in the temperature and time domains. Time-resolved SCTA techniques are useful in situations where it is necessary to isolate and collect the gaseous products and/or the residual solids of a TA experiment for subsequent identification by other analytical techniques. SCTA has also been developed for the investigation of 'large' samples (1 g) – thereby characterizing SCTA as a 'preparative' technique. A further variant within the SCTA family is 'Hi-Res' TG<sup>TM</sup> or 'Dynamic Rate' TG, in which the heating rate is reduced when the system crosses the upper threshold set for the reaction rate. The changes in heating and cooling rates are controlled by adjustment of the values of two numerical parameters – 'resolution' and 'sensitivity'. Temperature modulated SCTA is a further recent development. Since SCTA essentially eliminates sample 'self-heating' and inhomogeneous reaction within a sample, the resolution of SCTA is generally greater than that associated with conventional TA techniques.

### The Basic Principles of SCTA Techniques

It has already been stated that SCTA envelopes the original CRTA techniques. Essentially, there are two forms of CRTA – the quasi-isothermal/quasi-isobaric methods (Paulik) and the constant rate method (Rouquerol). The Paulik method is based on thermogravimetry (TG) and is known as 'Q-TG'. It depends on a control system that maintains a constant rate of mass loss and the measured parameter is temperature as a function of time. CRTA relies on using a transducer to monitor the pressure of evolved gas in a continuously evacuated chamber. The sample is heated in such a way as to maintain the monitored gas pressure constant. Since the pressure is maintained constant, the rate of gas pump-off is maintained constant and thus the rate of mass loss (when a single gas is evolved) is also maintained constant. Hence, CRTA is effectively vacuum thermogravimetry. The similarity between the Paulik and Rouquerol techniques is immediately apparent. Both maintain the reaction rate constant and both control the pressure of the evolved species in the reaction environment. Hence, the differences in the techniques are purely semantic. These techniques allow a precise

control of the uniformity of the reaction environment, which is effected by controlling the evolved gas pressure and the temperature and pressure gradients within the sample. Further, these techniques avoid artifacts that arise from a nonuniform sample environment such as artificial broadening of the temperature interval over which a reaction occurs and hence controlled rate techniques are ideal for kinetic studies. The ability of controlled rate techniques to resolve multiple thermal events is generally superior to conventional TA methods and is hence a further advantage.

Over the last decade, it has been internationally accepted that the generic title for TA techniques whereby the heating rate is controlled using a feedback loop as some function of the rate of reaction or the chemical or physical process under study should be sample controlled thermal analysis (SCTA) and this title now embraces Q-TG, CRTA, SIA, and dynamic rate thermal analysis.

Parkes *et al.* have described further variations of SCTA involving resolution in the time and temperature domains. In TA, resolution is related to the extent to which adjacent or partially overlapping thermal events are separated. In a conventional TA experiment, involving linear sample heating, the resolution in either the time or temperature domain is essentially the same and thus the TA profiles as related to either temperature or time are similar. However, this criterion is not the same for SCTA experiments in which the relationship between time and temperature is always nonlinear when a reaction is occurring. Consequently, SCTA curves displayed as a function of time are markedly different from those displayed as a function of temperature. Overall resolution is markedly enhanced with both types of 'resolved SCTA' techniques.

Conventionally, high resolution in TA is achieved by using low linear heating rates and small sample masses so that the sample reacts under near equilibrium conditions. However, these conditions lead to long experimental times and generally to low sensitivity. CRTA and SIA are basically similar, as essentially heating is slow through thermal events and fast between them. Such a heating regime provides all the benefits of a very low linear heating rate but with a reduced time penalty as compared to a TA experiment using the same linear heating rate, since the heating rate is faster during periods corresponding to no thermal events. Such heating strategies cause processes to occur over a very narrow temperature range (zero temperature range in the case of SIA) in both cases and leads to an increase in resolution in the temperature domain. However, one severe limitation of both SIA and CRTA is long experimental

times as compared to the time taken to complete a conventional TA experiment using a linear heating rate. Parkes *et al.* have described a temperature-resolved SCTA technique using 'proportional heating' (PH) in which the heating rate is proportional to a function of the reaction rate. Essentially, PH offers maximum temperature resolution within a minimum time frame. PH smoothly varies the heating rate between preset maximum and minimum values so as to provide the optimum balance between resolution (which is to be maximized) in the temperature domain and experimental time (which is to be minimized). The operation of the PH method depends on four main parameters: maximum and minimum heating rates, a 'target' reaction rate, and a function relating heating rate to reaction rate. When the reaction rate is zero, the heating rate is set to its maximum value and when the reaction rate is at, or above the preset target level, the heating rate is set to its minimum value. Between these limits, the heating rate is related to the reaction rate by one of several possible functions.

Temperature-resolved SCTA techniques operate by controlling the sample temperature as some function of the difference between the measured reaction rate and a preset 'target' rate. However, it is difficult to set a single level that is equally appropriate for complex systems that are associated with thermal events of different magnitudes. The basic principle of time-resolved SCTA is focused on maintaining a linear heating rate throughout a thermal event and reducing it to a very low level between two consecutive thermal events. However, a proviso applies – that the consecutive thermal events are well resolved. The initial linear heating rate imparts enhanced sensitivity since the process is forced to occur over a short time interval. Between the two thermal events, the very low heating rate increases the time between them and thus enhances resolution in the time domain. Thus, time-resolved SCTA is opposite in principle to temperature-resolved SCTA. Parkes *et al.* have described a time-resolved SCTA technique – termed peak-slope heating (PSH) in which the heating rate becomes zero when the rate of change of the reaction rate changes from positive to negative – immediately following a peak maximum. Conversely, the heating rate is increased as soon as the end of a thermal event is detected. PSH has a major advantage over temperature-resolved SCTA techniques in that the change in heating rate takes place at the same relative point on the peak, irrespective of its absolute magnitude. This eliminates the need to select a target reaction rate prior to the start of each experiment. For cases involving partial overlap of two thermal events, a more complex form of time-resolved SCTA

applies – known as peak-resolved heating (PRH), which relies on switching the heating rate at the point corresponding to the greatest acceleration of the reaction rate.

Summarizing, SCTA techniques are associated with six types of sample heating regimes: linear heating (LH) – the sample is subjected to a predetermined heating regime, including isothermal and cooling regimes; constant rate (CR) – the rate of a thermal event is made to proceed at a constant rate by variation of the sample temperature; stepwise isothermal – the sample is either heated or held isothermally so as to maintain the reaction rate between two predetermined thresholds; PH – the heating rate of the sample is directly proportional to its reaction rate (rapid heating of the sample occurs during thermal events and slow heating of the sample occurs between thermal events); PSH – the heating regime of the sample is switched from ‘heating’ to ‘isothermal’ at maximum reaction rate (slow heating of the sample occurs during thermal events and rapid heating of the sample occurs between thermal events). PSH is opposite to PH; and PRH – which is a more advanced form of PSH in which switching of sample heating corresponds to the point of maximum acceleration of reaction rate.

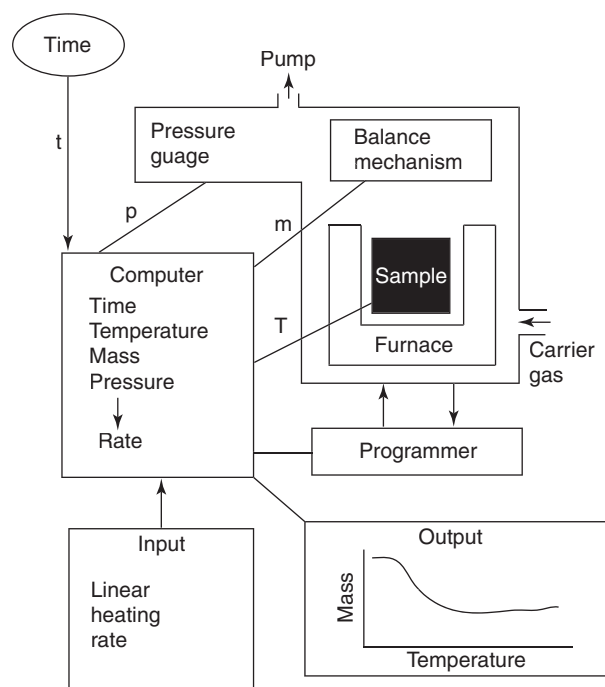
Temperature- and time-resolved SCTA techniques have major advantages over conventional TA techniques. Temperature-resolved SCTA can give information on reaction mechanisms that are not apparent from corresponding linear heating experiments. Temperature-resolved SCTA is also particularly advantageous for the *in situ* preparation of porous or finely divided solids by thermal decomposition and for the study of catalytic materials. Time-resolved SCTA is particularly useful in conjunction with an evolved gas analysis system whereby time resolution of evolved gases leads to their unambiguous identification.

The basic principles of SCTA have recently been incorporated into a commercial TG system developed by TA Instruments (USA) and known as ‘Hi-Res’ TG<sup>TM</sup> – an acronym for ‘High-Resolution TG’. This system allows for four heating regimes – a conventional constant heating rate; a stepwise isothermal mode, whereby the sample is heated until a reaction is detected and then held constant until the reaction is complete; a constant reaction rate, whereby heating of the sample is controlled to achieve a constant rate of mass change; and ‘dynamic rate TG’, which varies the heating rate smoothly and continuously in response to the rate of sample decomposition, so that the resolution of mass change is maximized. In addition, many ‘homemade’ SCTA systems have been developed, as described in the next section.

## SCTA Instrumentation

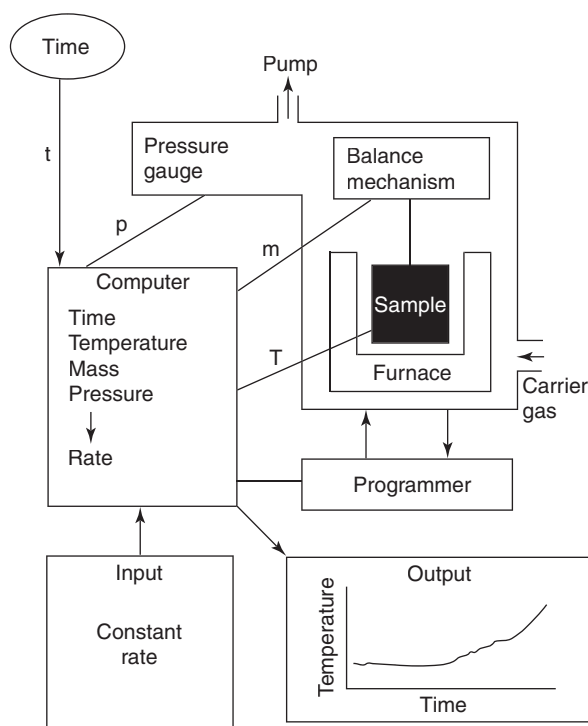
The essential difference between conventional TA and SCTA is the input to the control system. In conventional TA, a linear (sample) heating rate is imposed, whereas in SCTA the heating rate is related directly to the reaction rate. Figures 1 and 2 schematically show the key features of conventional TG and SCTG and define the ‘input’ condition for each system.

It is well-known that sensitivity and resolution in conventional TA are achieved by small sample sizes and rapid heating rates. SCTA offers unique possibilities for ‘preparative scale TA’ whilst maintaining high levels of sensitivity and resolution. The University of Huddersfield Centre for Thermal Studies has recently developed a ‘preparative scale TG’ system, which incorporates the principles of SCTA – specifically, the PH facility. The focus of the SCTG system is a five-decimal place top-pan balance (Stanton-Model BP211D) and a high-temperature water-cooled furnace (Stanton Redcroft), located above the balance and allowing operation to 1000°C (Figure 3). The sample is contained in a silica or platinum crucible and sits on top of a rise rod connected directly to the balance pan. Samples in excess of 1 g can be investigated, but routinely samples of 500 mg are usually involved. Grinding of powdered samples to mesh size 150–250 µm is advantageous. The sample temperature is monitored by a thermocouple located



**Figure 1** Schematic of a TG system. (Reproduced with permission from Brown ME (2001) *Introduction to Thermal Analysis – Techniques & Application*. Amsterdam: Kluwer. )

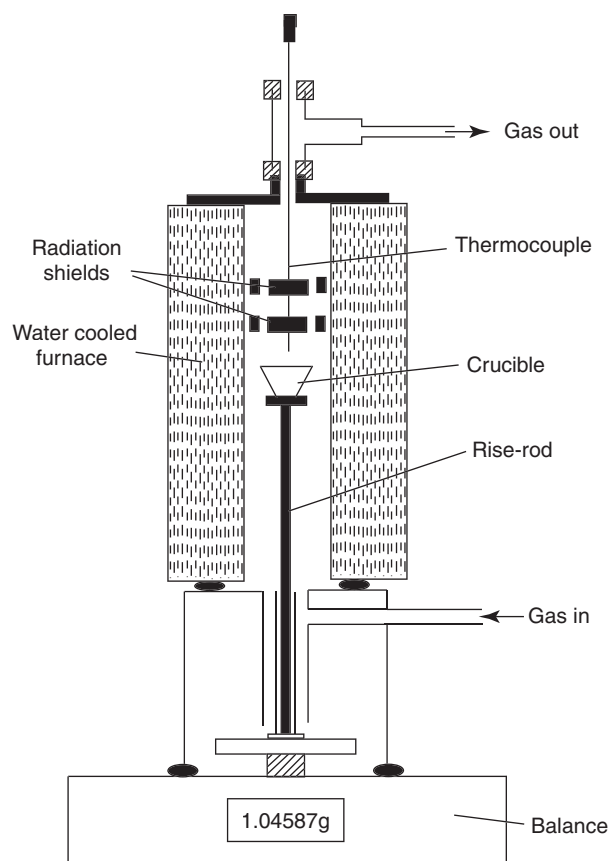




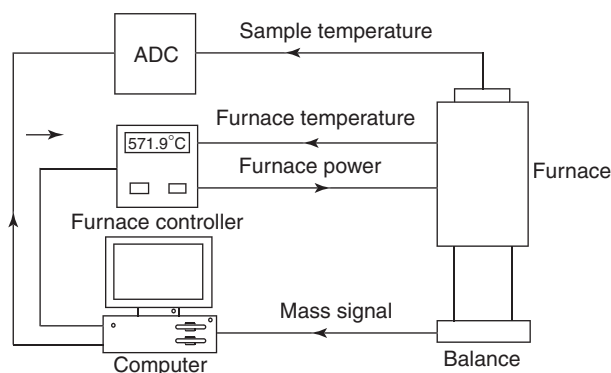
**Figure 2** Schematic of an SCTA system. (Reproduced with permission from Brown ME (2001) *Introduction to Thermal Analysis – Techniques & Application*. Amsterdam: Kluwer.)

centrally within the furnace tube. Radiation shields attached to the thermocouple reduce convection currents within the sample region. A gas diffusion chamber, situated between the balance and the furnace, permits experiments to be carried out in an upward flow of gas at a constant flow rate via continuously pumping the system. Although the overall system is not sealed, this net gas flow prevents air entering the system and results from the differential flow rates at the inlet and at the outlet, as measured by individual flow meters. Volatile decomposition products can be trapped using a combination of filters and activated charcoal and liquid nitrogen traps.

A block diagram of the data acquisition and control systems is shown in Figure 4, which are a development of those used for sample controlled studies based on evolved gas analysis techniques. The temperature of the furnace is continuously varied by a temperature controller and the output from the balance and from the sample temperature thermocouple are logged by a Pentium computer, fitted with a 16-bit dynamic range data acquisition board. The signals are processed via a program derived using a combination of Virtual Basic and C<sup>++</sup> and the derivative of the TG curve is used to control the heating program via the temperature controller. Several SCTA modes are available with this system – including



**Figure 3** Schematic of an SCTG system. (Reproduced with permission from Professor Edward L. Charsley, Centre for Thermal Studies, The University of Huddersfield, Huddersfield, UK.)



**Figure 4** SCTG control system. (Reproduced with permission from Professor Edward L. Charsley, Centre for Thermal Studies, The University of Huddersfield, Huddersfield, UK.)

linear heating TG and PH TG. In addition, the control software has a 'stop-mass' facility, by which 'products' can be produced at any preselected degree of reaction – thereby conforming to the 'preparative' aspect of the SCTG system. Modifications to the control system give rise to time-resolved SCTA

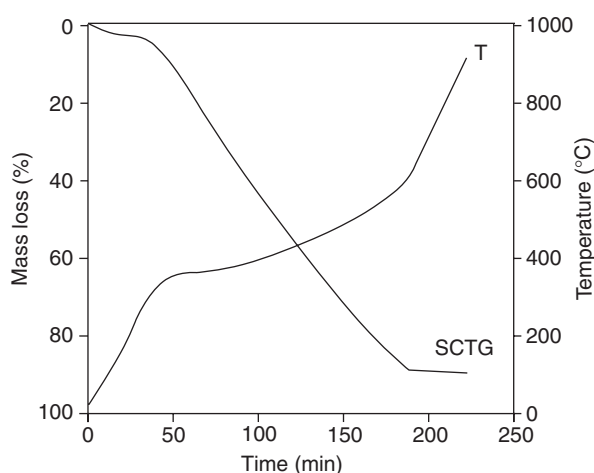
incorporating peak slope heating and dynamic rate modes. This 'in-house' construction is thus a very versatile SCTA system that incorporates the principal advantages of 'sample controlled' methodology.

## SCTA Applications

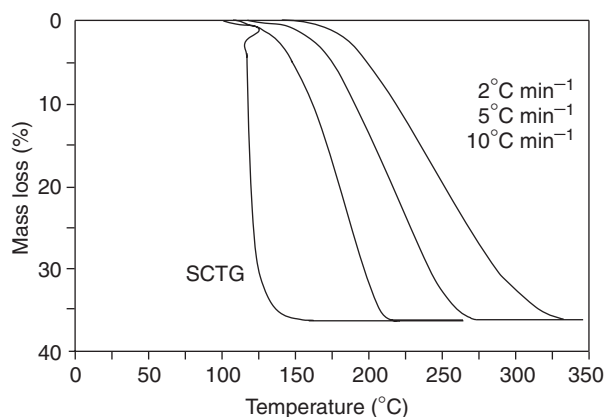
Major developments in SCTA together with diverse applications of these techniques have been made by the University of Huddersfield – Centre for Thermal Studies and the Centre for Applied Catalysis – and the following discussion represents a summary of the recent achievements of these research centers in SCTA. Specifically, these centers have been responsible for the development of temperature- and time-resolved SCTA and these techniques have been systematically applied to study the thermal behavior of coals and catalytic materials with consequential major advances in these fields.

The principal advantage of SCTA over conventional TA techniques is that in the former, thermal events can be studied under controlled conditions. A simple example of 'controlled decomposition' is the thermal decomposition of sodium bicarbonate and SCTG and TG curves for this process are shown in Figure 5. The most striking feature of the (proportional heating) SCTG curve is controlled decomposition over a narrow temperature range. The corresponding TG curves show a reduction of the decomposition temperature range with decreasing heating rate but the SCTG decomposition temperature range corresponds to zero heating rate of the sample, and hence via SCTA, the most realistic thermal decomposition temperature ranges are obtained. This feature is of particular significance

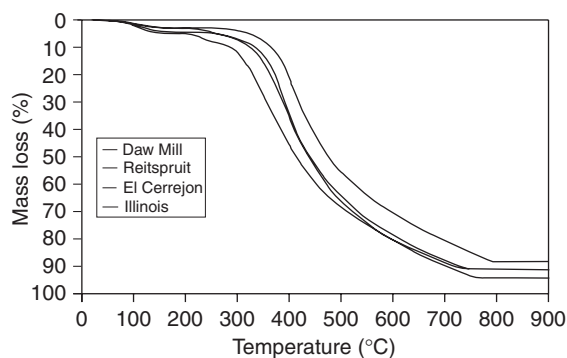
for samples that violently decompose or ignite such as pyrotechnics and coals. The primary advantage of the sample controlled approach in reducing both temperature and concentration gradients in a sample is illustrated by the oxidation of a Drayton coal sample (Figure 6). Even at a 500 mg sample level, a controlled oxidation is apparent and ignition is avoided. In conventional TA using the same conditions, self-heating of the sample leads to ignition and hence a loss of meaningful data after ignition, coupled with 'apparent' decomposition temperature ranges. Similarly, the controlled oxidation of four different coals is shown in Figure 7. Although the SCTG profiles are similar in character and reveal controlled oxidation, the subtle differences in oxidation behavior of these coals are revealed. Such differences are not readily deciphered from TG curves



**Figure 6** SCTG curve for the oxidation of Drayton coal. (Reproduced with permission from Professor Edward L. Charsley, Centre for Thermal Studies, The University of Huddersfield, Huddersfield, UK.)



**Figure 5** SCTG and TG curves of sodium bicarbonate. (Reproduced with permission from Professor Edward L. Charsley, Centre for Thermal Studies, The University of Huddersfield, Huddersfield, UK.)

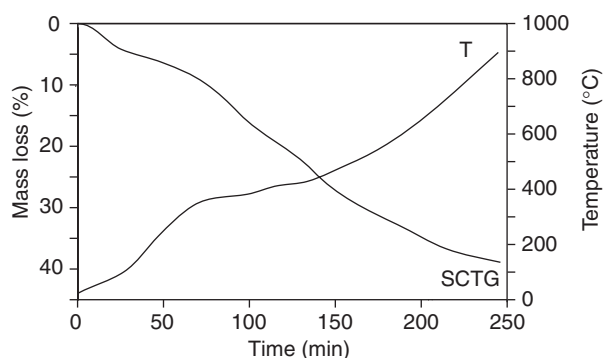


**Figure 7** SCTG curves for the oxidation of different coals. (Reproduced with permission from Professor Edward L. Charsley, Centre for Thermal Studies, The University of Huddersfield, Huddersfield, UK.)

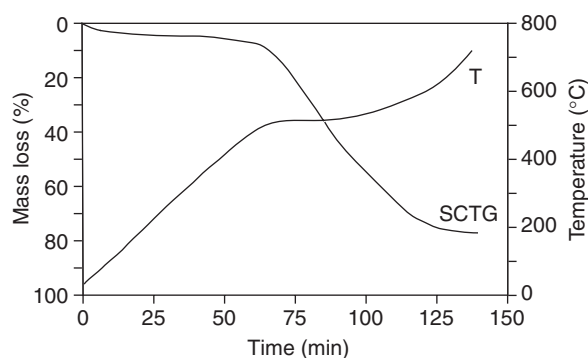


since the dominating feature is ignition that masks the intermediate thermal events. Coal pyrolysis can also be studied by SCTA in a nitrogen atmosphere. The PH SCTG curve for the pyrolysis of a sample of

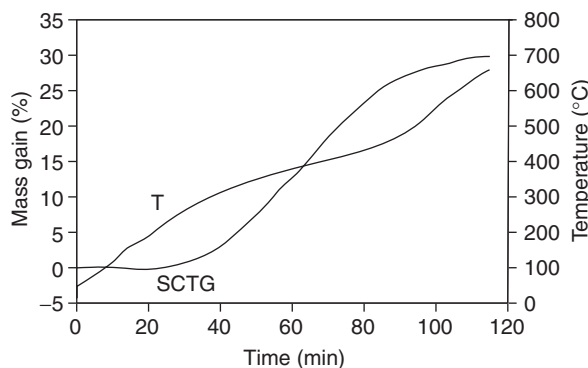
Illinois coal is shown in **Figure 8**. Initial pyrolysis occurs over a relatively narrow temperature range and the slow nature of the latter stages of pyrolysis is indicated by the temperature program returning to



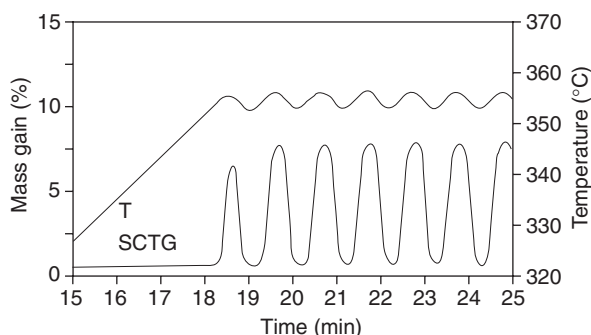
**Figure 8** SCTG curve for the pyrolysis of Illinois coal. (Reproduced with permission from Professor Edward L. Charsley, Centre for Thermal Studies, The University of Huddersfield, Huddersfield, UK.)



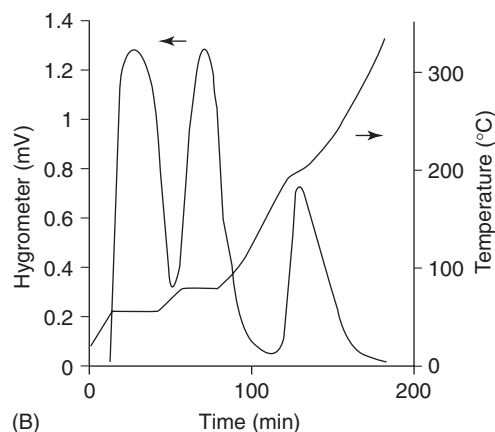
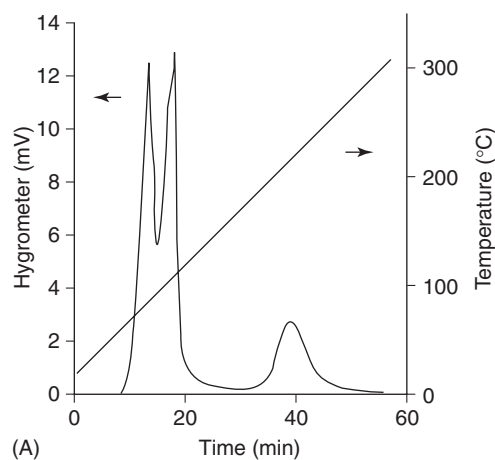
**Figure 9** SCTG curve for the oxidation of Drayton coal char. (Reproduced with permission from Professor Edward L. Charsley, Centre for Thermal Studies, The University of Huddersfield, Huddersfield, UK.)



**Figure 10** SCTG curve for the oxidation of zirconium powder. (Reproduced with permission from Professor Edward L. Charsley, Centre for Thermal Studies, The University of Huddersfield, Huddersfield, UK.)



**Figure 11** SC-thermomagnetometry curve for 'ICTAC Nickel'. (Reproduced with permission from Professor Edward L. Charsley, Centre for Thermal Studies, The University of Huddersfield, Huddersfield, UK.)



**Figure 12** LH (A) and PH (B) curves for the dehydration of copper sulfate pentahydrate. (Reproduced with permission from Parkes GMB, Barnes PA, and Charsley EL (1999) *Analytical Chemistry* 71: 2482–2487.)

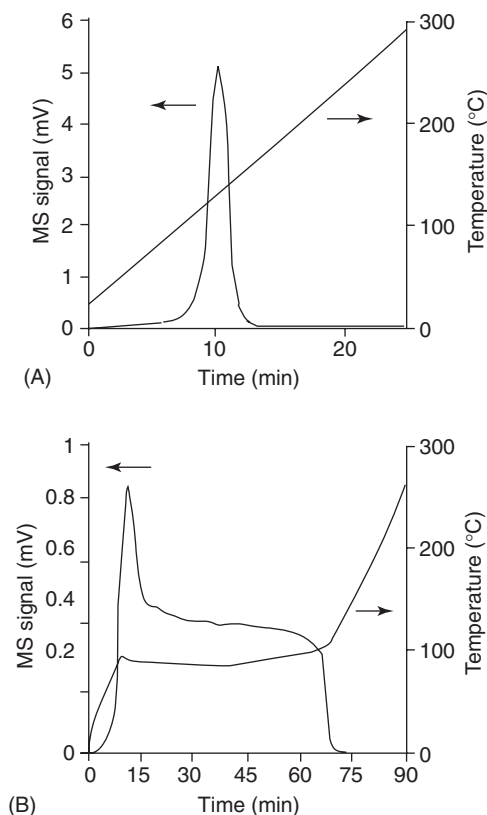
maximum heating rate. The 'stop-mass' facility of an SCTG system provides 'preparative opportunities' and in the context of coal pyrolysis, chars can be produced at any stage of the pyrolysis process that can subsequently be characterized by other analytical techniques. Further, the oxidation of such chars can be monitored by SCTG as shown in Figure 9 for a Drayton coal char. Prior to the development of SCTA, coal char TA was beset with controversy due to the indecisive nature of the derived data.

The high reactivity of finely divided zirconium powder toward aerial oxidation makes conventional TA studies difficult to perform under controlled conditions. Figure 10 shows an SCTG curve for a 10 mg sample of zirconium powder, which in a TG-DTA experiment under similar conditions but using a linear heating rate shows ignition at  $\sim 410^\circ\text{C}$ . However, it is apparent that under PH SCTG conditions, a smooth oxidation reaction process results, which is 50% complete below  $400^\circ\text{C}$ . These types of studies are most suitable for pyrotechnic mixtures, since reliable ignition data are generated.

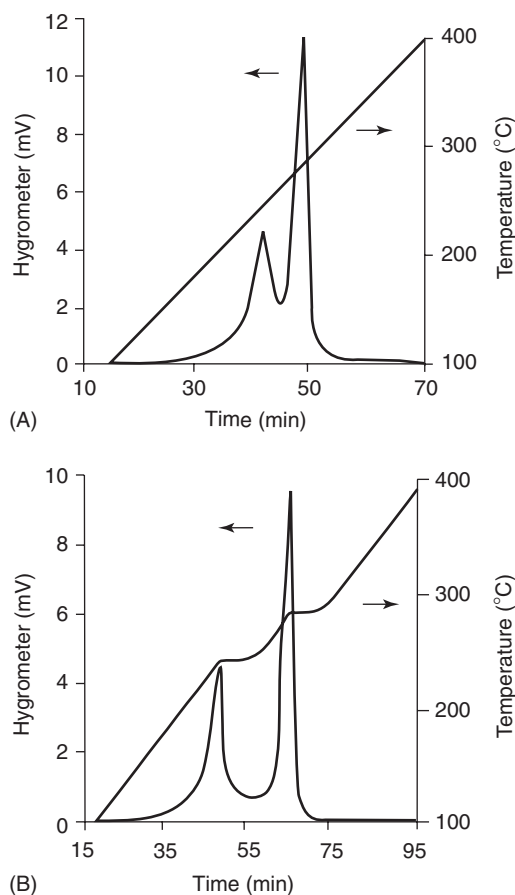
The possibility of performing sample controlled thermomagnetometry is illustrated by an SCTM

experiment on ICTAC nickel, with a magnet placed above the thermobalance furnace (Figure 11). During the heating program, the sample oscillates between the magnetized and unmagnetized states. If a zero cooling rate is applied, this transition can be made to take place over a narrow temperature range thereby characterizing the magnetic property of the metal.

The scope of temperature- and time-resolved SCTA has been revealed by Parkes *et al.* via studies of the decomposition of inorganic salts. Temperature-resolved SCTA is particularly useful for resolving the dehydration characteristics of copper sulfate pentahydrate. A typical TG profile for this salt is shown in Figure 12 and compared with a PH SCTG profile over the same temperature range. The typical 2:2:1 water loss is shown in both profiles but the resolution of these peaks is superior in the SCTG profile. Further, the latter shows how the heating rate is reduced through each dehydration peak and increased between each peak, which is the defining feature of temperature-resolved SCTA. Further,



**Figure 13** LH (A) and PH (B) curves for the decomposition of ammonium hydrogencarbonate. (Reproduced with permission from Parkes GMB, Barnes PA, and Charsley EL (1999) *Analytical Chemistry* 71: 2482–2487.)

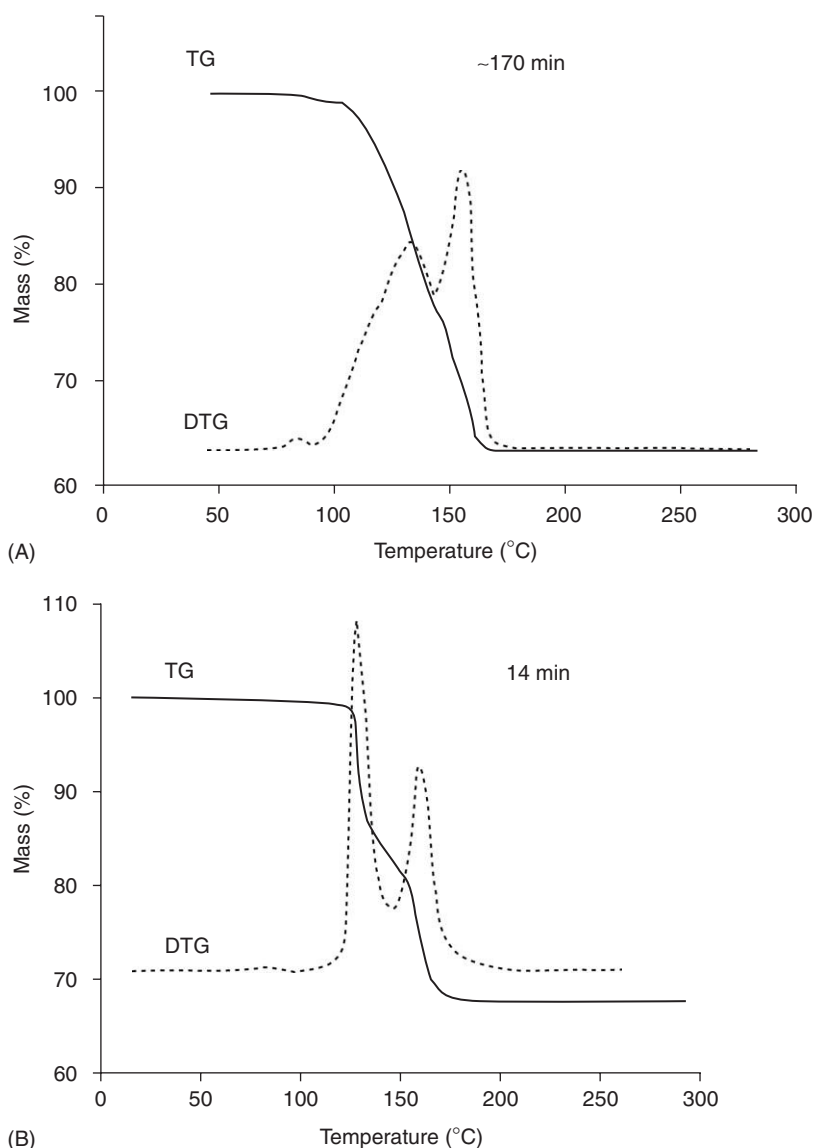


**Figure 14** LH (A) and PH (B) curves for a 1:2 mixture of copper and zinc hydroxycarbonates. (Reproduced with permission from Parkes GMB, Barnes PA, and Charsley EL (1999) *Analytical Chemistry* 71: 2482–2487.)

temperature-resolved SCTA can be used in mechanistic studies of complex reactions. The decomposition of ammonium hydrogencarbonate is associated with the co-evolution of ammonia, water, and carbon dioxide. PH SCTA in conjunction with mass spectrometry was applied to study this decomposition process and linear heating and PH TA curves are shown in Figure 13. A single decomposition peak, centered at 125°C is apparent from the linear heating profile, whereas the PH profile shows the presence of a nucleation stage prior to decomposition thereby providing mechanistic indications of the decomposition mode.

An application of the time-resolved SCTA technique is shown in Figure 14 relating to LH and PSH

of a 1:2 mixture of copper hydroxycarbonate and zinc hydroxycarbonate. Decomposition of this mixture yields both carbon dioxide and water and the latter was monitored using a hygrometer. The principal feature of time-resolved SCTA is shown by the PSH profile. It is apparent that the switch to zero heating rate on the falling edge of each peak is rapid, whereas the increase in heating rate as the first decomposition approaches completion is slower. Also, the switching point between heating and nonheating is at the same relative point on each peak despite their difference in size. The enhanced resolution of the decomposition peaks without distortion of shape is clearly apparent from the PSH profile and this feature is indicative of the ability of PSH



**Figure 15** TG (A) and 'Hi-Res' TG<sup>TM</sup> (dynamic rate) (B) curves for a sodium/potassium carbonate mixture. (Reproduced with permission from Haines PJ (ed.) *Principles of Thermal Analysis and Calorimetry* (2002). Cambridge: Royal Society of Chemistry.)

to time-resolve gaseous decomposition products for 'online' evolved gas analysis.

It should, however, be noted that although both PH SCTA and peak-resolved SCTA provide enhanced resolution in the temperature and time domains, respectively, both of these modes are typically associated with increased experimental times compared to those of conventional TA experiments. This feature at present eliminates SCTA as an analytical technique for the rapid characterization of materials and in quality control applications. However, these techniques offer many other advantages, most notably in providing mechanistic and kinetic information on thermal decomposition behavior of materials.

TA Instruments (USA) have recently produced a commercial SCTA system – known as 'Hi-Res' TG<sup>TM</sup> or 'Dynamic Rate' TG<sup>TM</sup>. The basic principle of this system is that there is a marked decrease in heating rate when the system crosses a preset upper reaction rate threshold. The changes in heating or cooling rates are controlled by adjustment of the values of two numerical parameters, referred to as 'resolution' and 'sensitivity'. 'Dynamic Rate' TG varies the heating rate smoothly and continuously in response to the rate of sample decomposition so that the resolution of the mass change is maximized. This mode allows rapid heating in regions where no transitions occur and proportionally slow heating during reactions. The latter form of 'Hi-Res' TG<sup>TM</sup> is illustrated by conventional TG (linear heating) and dynamic rate TG profiles for a sodium/potassium hydrogencarbonate mixture (**Figure 15**). It is apparent from the latter profile that enhanced resolution of the decomposition peaks is achieved in a time frame of less than 10% of the conventional TA experiment.

It is apparent that the future of SCTA is likely to be focused on maximizing the resolution/experiment time ratio inherent in these techniques. PH SCTA is a refinement of the Rouquerol CRTA technique but the associated long experimental times present an operational drawback. PSH SCTA, which heats quickly into thermal events but slowly between them, gives enhanced resolution in the time domain whilst maintaining sensitivity levels comparable to conventional TA techniques, but again at the expense of increased experimental times. The TA Instruments 'Dynamic Rate' TG<sup>TM</sup> appears to address both of these parameters and hence has much potential as a 'high resolution/rapid TA system'.

Finally, it is well recognized that TA can be applied in the complex domain of reaction kinetics. Since 'reaction rate' is the key concept on which SCTA techniques are based, these techniques collectively are becoming increasingly applied to study reaction mechanisms and to derive kinetic data for thermal decomposition processes of a wide range of materials. It is this type of application that will characterize SCTA as a dominant group of techniques for studying reaction energetics.

**See also:** **Thermal Analysis:** Overview; Temperature-Modulated Techniques; Coupled Techniques; Nonbasic Techniques.

## Further Reading

- 8th European Symposium on Thermal Analysis and Calorimetry (ESTAC 8), Barcelona, August 2002. Abstracts cited: IP-26: Nomen R (Symposium Chair), Institut Quimic De Sarria Universitat Ramon Llull, Barcelona, Spain.
- Barnes PA, Parkes GMB, and Charsley EL (1994) A high performance EGA system for catalyst characterisation. *Analytical Chemistry* 66: 2226–2231.
- Charsley EL, Rooney JJ, Hill JO, *et al.* (2003) Development and applications of a preparative scale sample controlled thermal analysis system. *Journal of Thermal Analysis and Calorimetry* 72: 1091.
- Gill PS, Sauerbrunn SR, and Crowe BS (1992) High resolution thermogravimetry. *Journal of Thermal Analysis* 38: 255–266.
- Parkes GMB, Barnes PA, and Charsley EL (1999) New concepts in sample controlled thermal analysis: Resolution in the time and temperature domains. *Analytical Chemistry* 71: 2482–2487.
- Paulik F and Paulik J (1971) Quasi-isothermal thermogravimetry. *Analytica Chimica Acta* 56: 328–331.
- Paulik F and Paulik J (1986) Thermoanalytical examination under quasi-isothermal/quasi-isobaric conditions. *Thermochimica Acta* 100: 23–59.
- Reading M (1988) The kinetics of heterogeneous solid state decomposition reactions: A new way forward. *Thermochimica Acta* 135: 37–57.
- Rouquerol J (1989) Controlled transformation rate thermal analysis – the hidden face of thermal analysis. *Thermochimica Acta* 144: 209–224.
- Sorenson OT and Rouquerol J (2003) *Sample Controlled Thermal Analysis*. Amsterdam: Kluwer.
- TA Instruments Hi-Res TM<sup>TM</sup>, TGA Application Brief (not dated). TA Instruments Inc., Wilmington DE, USA.
- Toft-Sorenson O (1981) Quasi-isothermal methods in thermal analysis. *Thermochimica Acta* 50: 163–175.

## Nonbasic Techniques

J O Hill, La Trobe University, Wodonga, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

'Less common thermal analysis (TA) techniques' encompass those TA techniques that monitor 'less obvious' properties of a sample as a function of temperature and hence these are specialized techniques that require specialized instrumentation. 'Less common techniques' are not synonymous with 'less used techniques' but relate to a field of TA that is less well developed as a result of being highly specialized. Some of the 'less common TA techniques' are 'standalone techniques' and are associated with International Confederation for Thermal Analysis and Calorimetry (ICTAC) approved definitions. Others in this group are variants of 'more common' TA techniques and do not have ICTAC approved definitions but, nonetheless, are viable TA techniques. In the former category, thermal techniques such as emanation thermal analysis (ETA), thermoelectrometry, and thermosonimetry (TS) are noteworthy. In the latter category, a range of techniques has emerged that focus on the method of heating of the sample. Techniques such as 'microwave thermal analysis' (MWTa) and 'programmed probe analysis' fit into this category. The former technique involves microwave heating of the sample and the latter technique is effectively an evolved gas analysis (EGA) technique and involves the rapid heating of the sample by the probe of a mass spectrometer. Techniques currently in the 'less common' category, such as micro-thermal analysis ( $\mu$ -TA), are becoming 'more common' as the major advantages of such techniques become increasingly self-evident. Also, with a progressive blurring of the edges of TA and a progressive inclusion of 'calorimetry' into the thermal studies domain, a wide range of 'special purpose calorimeters' has emerged which have 'less common' application, but which have very significant impact in terms of advancing the overall emphasis and application of TA techniques. The 'less common TA techniques' group provides a further insight into the diversity, complexity, authenticity, and superiority of this category of analytical techniques, which collectively have revolutionized studies of materials over several decades and promise analytical opportunities in support of the nanotechnological age.

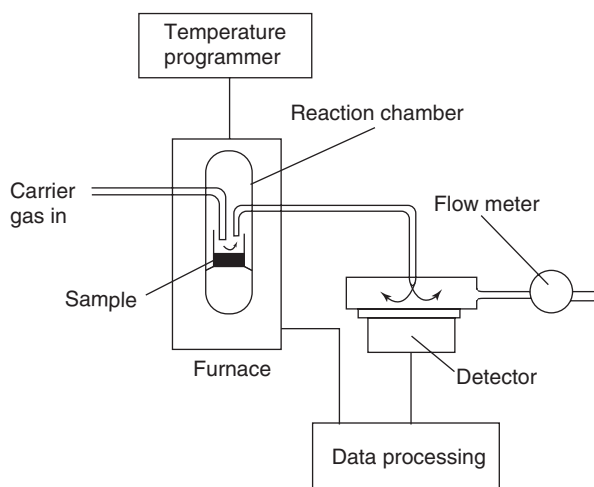
### Selected 'Less Common' TA Techniques

Emanation TA involves the measurement of the release of inert (usually radioactive) gas from a solid sample as a function of temperature. The rate of such gas release is essentially an indication of the changes taking place in the sample and a comparison of ETA data with those of other TA techniques, particularly thermogravimetry (TG) and EGA, provides information on the microstructure of the sample material. However, most of the solid samples studied by EGA are 'spiked' with 'inert gas', which is trapped at lattice imperfections and the latter also provide diffusion pathways for adsorbed gases. It is this pre-spiking procedure that characterizes ETA as a less-common TA technique. Several methodologies are available to effect the required 'spiking'. These are essentially divided into two groups: techniques for introducing the parent nuclide of an inert gas and techniques for introducing the inert gas itself. The former group of spiking techniques is preferable since labeling the sample with a parent nuclide gives an enhanced timeframe over which the inert gas can be desorbed from the sample, since it is formed continuously *in situ*. With the second group of techniques, the absorbed inert gas may be lost from the sample over the timeframe of a single run.

ETA is usually carried out in conjunction with other TA techniques, most notably differential thermal analysis (DTA) and EGA. In this context, ETA can be considered as a coupled TA technique. Carrier gas, at a constant flow rate, is used to carry released gas from the sample to appropriate detectors – usually radioactive counting devices. In the case of desorbed radon, a scintillation counter is used, whereas Geiger counters are used for krypton, xenon, and argon. A typical ETA system is shown schematically in **Figure 1**. An ETA curve is a plot of 'emanating power'  $E$  as a function of time and the  $E$ /time relationship has been developed and defined by Balek.

A major difficulty of ETA is that preparation and handling of samples requires sophisticated radiochemical facilities coupled with the associated precautions. However, the net amounts of radioactive gas incorporated into samples are so small that the evolved gas, after dilution with the carrier, does not pose a significant hazard.

One of the major applications of ETA is the characterization of powders. Emanating power is related directly to surface area and hence changes in grain



**Figure 1** Schematic of an ETA system. (Reproduced with permission from Brown ME (2001) *Introduction to Thermal Analysis – Techniques and Applications*. Amsterdam: Kluwer.)

size and the occurrence of sintering during heating can readily be detected. Kinetic parameters related to sintering can also be derived. Solid-state phase changes can also be manifested as changes in emanating power. In this context, ETA data can be correlated with DTA, differential scanning calorimetry (DSC), and thermo-mechanical analysis (TMA) data related to phase changes of solid-state materials. ETA has also been applied to study solid–gas reactions such as the oxidation of metals and the reduction of metal oxides. In this respect, ETA data are useful for studying surface phenomena and hence ETA has potential for future applications in the study of corrosion and catalytic processes. Solid–solid reactions, such as spinel formation between zinc oxide and iron(III) oxide can be studied by ETA by labeling ZnO with a radioactive thorium isotope. It is apparent that the two reactant oxides interact in a series of stages involving an initial surface diffusion process and a final volume (bulk) diffusion process. These processes are not readily detected by other TA techniques such as DTA or TMA curves but are clearly shown by increased emanation with clear definitions of the corresponding temperature ranges. The reactivities of oxides can readily be determined by ETA in terms of the mutual reactivities toward spinel formation.

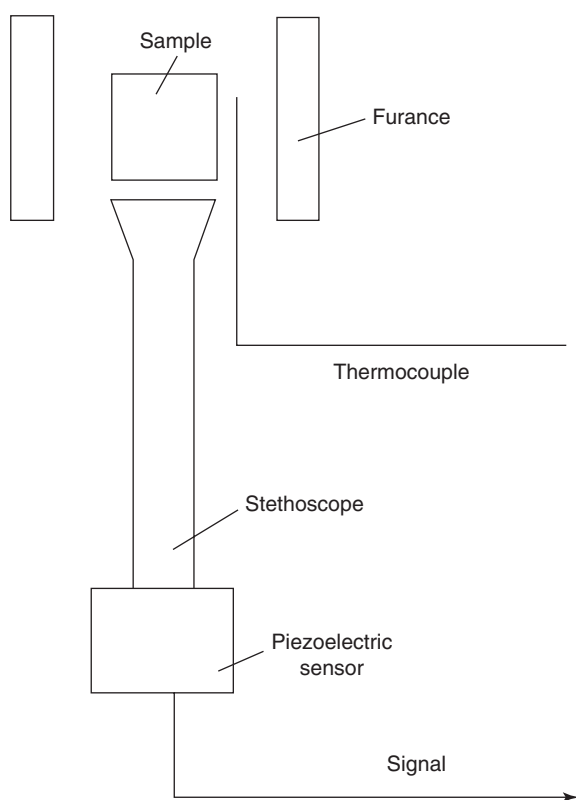
TS involves the measurement of the sound intensity emitted by a sample as a function of temperature. Sound emission originates from the release of thermal stresses in the sample, such as movement of dislocations, generation and propagation of cracks, nucleation of new phases, relaxation processes, and discontinuous changes in physical properties. For example, at a glass transition temperature, a discontinuous change in free volume generates elastic waves

that cause an acoustic effect. Such stress–relief processes are not usually detectable by conventional TA techniques, due to the low associated energy; hence, TS has many unique applications such as the detection of radiation damage, crystal defect content, and degree of annealing of polymeric samples. It is also complimentary to TMA for the detection of mechanical effects associated with melting, dehydration, decomposition, and for studying a range of solid-state phase changes.

TS is usually carried out in conjunction with DTA and sound energy is emitted as mechanical vibrations prior to and during thermal events associated with the sample over defined temperature ranges. Such sonic activity is detected by a specially adapted stethoscope. The mechanical waves are converted to electrical signals by conventional piezoelectric transducers. The stethoscope is constructed of fused silica for operation to 1000°C – for higher operating temperatures, ceramics or noble metals can be used. The sample is contained in the DTA sample head, which in turn acts as an acoustic transformer and is connected via a transmitting rod to a piezoelectric sensor, fixed on a recoil foundation and a seismic mount to prevent interference from external noise. A schematic diagram of a TS–DTA system is shown in **Figure 2**. Several special aspects of the instrumentation enhance sensitivity. The waveguide system transmits the acoustic emissions from the heated sample to the transducer at ambient temperature so as to nullify the effect of temperature on the transducer and the sample temperature thermocouple is located close to but not in contact with the sample since insertion of the sensor in the sample can cause mechanical damping effects that directly affect acoustic activity.

The output of a TS experiment consists of a rapid cascade of decaying signals that may be recorded in several ways; as the number of signals of peak amplitude greater than a set threshold value in a given time; the time for which the signal amplitude exceeds the threshold value; the time frequency that signals pass through a chosen voltage level in a positive direction; and the root-mean-square amplitude level (energy or as a set of frequencies). In general, interpretation of TS data is complex. Attempts have been made to relate frequency distributions to processes (thermal events) occurring in the sample. In the simplest interpretation, frequency distributions can be used as ‘fingerprints’ of sample origin. Superficial interpretation of TS data is also possible in that ‘bubble release’ is associated with low frequencies and crystal fracture or distortion is associated with high frequencies. In general, TS, with suitable refinements, is a powerful technique for rationalizing a wide variety of solid-state phenomena. For example, TS has been

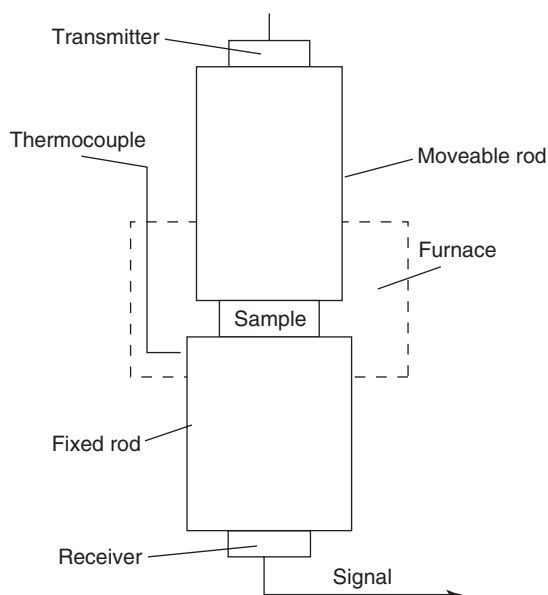




**Figure 2** Schematic of a TS-DTA system. (Reproduced with permission from Brown ME (2001) *Introduction to Thermal Analysis – Techniques and Applications*. Amsterdam: Kluwer.)

applied to study the dehydroxylation behavior of kaolins, which show two regions of acoustic activity. The low-temperature region corresponds to dehydroxylation and the high-temperature region corresponds to recrystallization to metakaolinite. TS is also useful for identification of 'pretransition' activity in a sample. The TS curve of potassium perchlorate shows acoustic activity prior to the transition temperature of 299°C, which corresponds to the release of included fluid. Similarly, the TS curve of potassium sulfate shows acoustic activity prior to the transition temperature of 582°C, which corresponds to the formation of microcracks in the sample structure. A DTA study of potassium perchlorate shows melting at 340°C and two successive exotherms due to decomposition to potassium chloride over the range 540–610°C. Simultaneous DTA-TS shows a total of four acoustic events for potassium chlorate. The first two TS peaks are associated with 'postmelting' activity directly related to the formation and evolution of gas bubbles in the melt.

A complementary technique – thermoacoustimetry – involves the measurement of the characteristics of imposed acoustic waves after passing through a sample as a function of temperature. The instrumentation for thermoacoustimetry is complex and is shown



**Figure 3** Schematic of a thermoacoustimetry system. (Reproduced with permission from Brown ME (2001) *Introduction to Thermal Analysis – Techniques and Applications*. Amsterdam: Kluwer.)

schematically in Figure 3. In one version of a thermoacoustimetry system, the sample is contained between a pair of lithium niobate transducers under an applied pressure of 300 kPa. One transducer induces the incident acoustic signal and the other detects the transmitted signal. Thermal expansion of the sample during the heating program is monitored continuously by a linear variable differential transducer. Atmosphere control of the sample environment is also possible. The incident signal is generated by a pulse generator. The transmitted signal received by the second transducer is inverted and amplified and an attenuated version of the driving pulse is added to this output signal. This summation procedure enables detection of the both compressional (P) and shear (S) waves. The velocities of the P and S waves and hence the corresponding elastic moduli are thence computed over temperature intervals and the final output is a plot of velocity or modulus versus temperature. The instrument is calibrated by an aluminum standard for which the P and S wave velocities are accurately known.

Thermoacoustimetry has been applied to differentiate between different grades of oil shales. Both the P and S wave velocities decrease with increasing temperature and with increasing organic content. Discontinuities and peaks in the P wave velocity/temperature plots relate directly to loss of included water and decomposition of some hydrocarbon fractions. Thermoacoustimetry, in conjunction with DTA, has been applied to examine the characteristics of

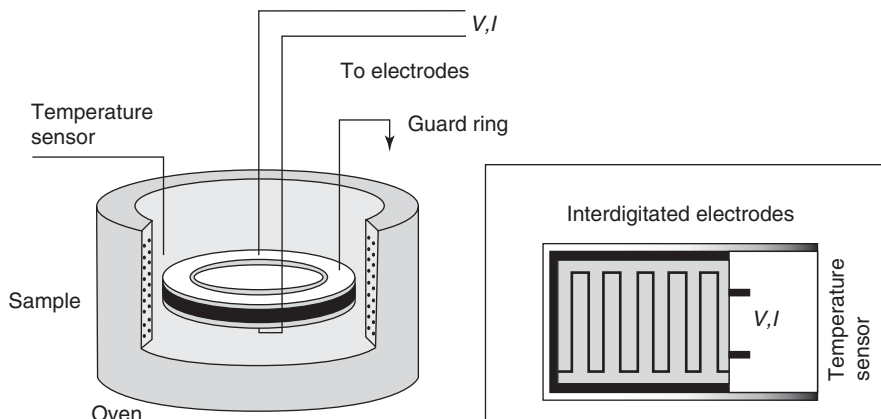
synthetic fibers. Generally, there are two significant increases in signal output, the first at the glass transition and the second prior to melting. The glass fiber form shows no thermal activity in this temperature range.

It is apparent that both these 'acoustic' TA techniques examine the 'fine-structure' associated with solid-state thermal events. Thus, these techniques are particularly suitable for a detailed TA of 'surface' and 'bulk' properties of solids, together with lattice imperfections. Microimpurities and the interactions of these with the host are also able to be evaluated by these specialized techniques. In particular, water inclusions in a host lattice can, in principle, be characterized in terms of the degree of bonding involved and the dehydration characteristics of the material under examination can be determined. These techniques have major analytical potential in materials science, mainly in terms of providing refinements to the traditional understanding of solid-state physical phenomena.

Thermoelectrometry or thermoelectrical analysis (TEA) is the generic title for a group of techniques involving the measurement of electrical properties of a sample as a function of temperature. The electrical properties commonly measured in TEA are conductance, capacitance, and dielectric properties. The most prominent technique in TEA is dielectric thermal analysis (DETA), which involves measurements of both the capacitance and the conductance of the sample as functions of time, temperature, and frequency. The former is the ability of the sample to store charge whilst the latter is a measure of its ability to transfer charge. Four parameters are associated with DETA: the permittivity, the loss factor, the dissipation factor, and the ionic conductivity. These parameters provide information related to molecular motion within the sample and as for most TA

techniques, it is the variation in these parameters during thermal events that is of primary interest rather than their absolute values. Hence, in DETA, the sample is subjected to an oscillating sinusoidal electric field and the applied voltage produces a polarization within the sample and causes a current to flow which in turn leads the electric field by a phase difference ( $\partial$ ). DETA is somewhat related to the thermomechanical techniques, in that the net current flow is similar in nature to the deformation brought about by applied mechanical stress and represents a measure of the freedom of charge carriers in the sample to respond to the applied field. Also, DETA is able to differentiate between different types of 'charge carriers' in the sample. Dipoles will attempt to orientate with the applied electric field, whereas ions, which are often present as impurities, will migrate toward the electrode of opposite polarity. DETA is therefore of significance as an analytical technique for 'quality control' applications.

A schematic diagram of a typical DETA system is shown in **Figure 4**. The sample is presented as a thin film, typically no more than 1–2 mm thick, sandwiched between two parallel plates so as to form a simple electrical capacitor. A grounded electrode surrounding one plate, known as a 'guard ring', is incorporated so as to nullify stray electric fields. The sample temperature sensor is a thermocouple or a platinum resistance thermometer, which is placed in contact with one of the plates. The sample is connected to the inverting input of an operational amplifier, configured as a current-to-voltage converter with capacitive feedback. The measured phase shift of this network can be related to the dielectric properties of the sample. There are two operational modes in DETA: using a continuous range of frequencies (50 Hz to 1 MHz) at preselected temperatures, or over a continuous range of temperatures at



**Figure 4** Schematic of a DETA system. (Reproduced with permission from Brown ME (2001) *Introduction to Thermal Analysis – Techniques and Applications*. Amsterdam: Kluwer.)

selected frequencies. The latter is the more common operating mode for the study of phase transitions and chemical reactions in a sample. Temperature calibration in DETA is effected by measuring the melting transition of a low-molar-mass organic powdered material, such as benzoic acid.

DETA is a highly specialized TA technique but is particularly useful for detecting polar impurities in nonpolar materials such as commercial polymers. Thus, DETA and the thermoelectrometric techniques in general occupy a specialized niche in the TA application field in the context of addressing a specific problem. For example, the TA of thin films can be problematic due to the very small amount of material involved. The glass transition behavior of adhesives affects the tack and bonding behavior and the analysis of adhesive films by DSC is difficult unless the film is removed from any backing support material. DSC analysis of such films is usually undertaken by swelling the adhesive in a suitable solvent and scraping it off the supporting substrate. However, the intrinsic properties of the adhesive may change in this process, even if the sample is subsequently dried to remove residual solvent. DETA, however, is ideally suited to the direct analysis of thin films, and so long as the support does not undergo any thermal transitions in the region of interest, DETA can measure the thermal properties of the adhesive *in situ* without any sample pretreatment.

DETA essentially involves monitoring the viscosity of a system via its ability to store and transport electrical charge. Changes in the degree of alignment of dipoles together with the ion mobility provide information relating to physical transitions in the sample and to changes in properties such as viscosity, rigidity, reaction rate, and cure state. DETA is particularly useful for characterizing the cure process of polymers. Polymers contain ionic impurities and the application of a voltage between two electrodes creates an electric field that forces such ions to migrate to the oppositely charged electrode. However, the conductivity of the material depends on the magnitude of the viscous drag experienced by these ionic impurities. Ions moving through polymerized (rubbery) materials have low mobility and conductivity corresponding to the high viscosity of the medium. Since the degree of cure in synthetic polymers is proportional to the degree of polymer cross-linking, the cure process can be monitored by monitoring the conductivity of the material throughout the entire curing process and DETA is able, more than any other TA technique, to determine the 'cure endpoint'.

DETA can also be used very effectively for the *in situ* monitoring of the ultraviolet (UV) degradation of adhesives such as tinted self-adhesive plastic films

used to screen out direct sunlight from penetrating buildings and vehicles. Attached to the window glass, the adhesive directly receives solar radiation and must be stabilized against photodegradation by a suitable blend of polymer and stabilizer. A modified form of DETA, in which the sample temperature sensor is an interdigitized single surface dielectric sensor, can be used to study such systems. The window film samples are mounted on glass plates for accelerated aging. A small hole is cut into the plastic film and the interdigitized sensor is applied to the exposed surface of the adhesive after freezing the plate to aid the removal of the backing. Measurements of the dielectric and permittivity factors are made at ambient temperature as a function of frequency. Such measurements correlate UV aging of the film with changes in dielectric behavior of the adhesive and, hence, optimized polymer/stabilizer formulations can be identified.

Thermally stimulated current analysis (TSCA) is another member of the TEA group of techniques. In TSCA, the sample is subjected to a constant electric field and the current that flows through the sample is measured as a function of temperature. The mode of operation usually involves heating the sample to a high temperature under the applied field followed by quenching to a low temperature. This process aligns dipoles within the sample. The polarization field is then switched off and the sample is reheated and the current flow resulting from the relaxation of the induced dipoles back to the disordered state is monitored.

TSCA finds applications in the pharmaceutical industry in which the stability of lyophilized products is a critical issue. Freeze-drying a drug formulation is commonly used to preserve the active component during storage. The active ingredient is diluted and embedded in an excipient formulation during the freeze-drying process. The physical properties of the excipient in the solid phase determine the overall stability of the formulation. Knowledge of the glass transition and melting temperatures of the formulation are essential to the prediction of product shelf-life. Also, the interaction of the product with water is essential to modeling the freeze-drying process. It is generally recognized that melting transitions are best determined by DSC. However, glass transitions are usually associated with very small energy changes and hence are often difficult to detect unambiguously by DSC. A feature of TSCA is that the current flow is directly proportional to the strength of the applied electric field. Thus, it is possible to magnify weak transitions simply by increasing the polarization voltage.

Dielectric analysis measures changes in properties of a sample as a function of temperature as it is

subjected to a periodic applied electric field. Changes in the dielectric constant of a material with temperature can arise from changes in molecular motion as a result of phase transitions or chemical transformations. For example, dielectric analysis of copper sulfate pentahydrate shows stepwise changes of dielectric constant in unison with stepwise dehydration – the latter being revealed by TG. Dielectric analysis is particularly useful for studying order–disorder crystal transitions and the diffusion dynamics of lattice-bound water molecules in inorganic salts.

## Miscellaneous ‘Less-Common’ TA Techniques

A form of EGA involves heating the sample at a constant rate in the probe of a mass spectrometer and the resulting TA technique is known as ‘programmed probe analysis’ or more simply as ‘analytical pyrolysis’. By such a process, samples are pyrolyzed in high vacuum since the sample environment is equivalent to the operating pressure of the mass spectrometer. This technique has been used by Bratspies *et al.* to determine thermal decomposition mechanisms of metal complexes, particularly tin dithiocarbamate complexes and this definitive study has been reviewed.

Mandelis has reviewed ‘photothermal’ TA techniques. Thermal waves may be optically induced in solid samples by modulated irradiation. These thermal waves then interact directly with the sample and such interaction is detected by suitable sensors. Acoustic waves may be simultaneously induced and detected. These techniques have specialized application to solid-state systems to determine thermal transport properties such as thermal conductivity, diffusivity/effusivity, and specific heat capacity. These techniques are of particular significance in the determination of mechanisms of solid-state phase transitions.

Parkes *et al.* have described a new technique – ‘microwave thermal analysis’, which, as the name implies, involves heating the sample by microwave radiation. This heating method has a major advantage over the conventional method using a programmed furnace, in that the microwave radiation heats by direct molecular interaction rather than by conduction or convection and thus temperature gradients in the sample are markedly decreased. The uniformity and rapidity of heating of samples enhances the resolution of thermal events. MWTA is related to the thermoelectrometry techniques in that it depends on changes of dielectric properties of the sample. With suitable refinements, MWTA has a promising future.

## TA Developments (1996–2005): A Summary Overview

Traditionally, the full range of ‘classical’ TA techniques has been associated with three major problems: the ‘relatively’ long experimental measurement times; the small (mass) size of samples; and the ‘averaged nature’ of the data output. A fourth drawback of this group of techniques is that, in general, no single technique is able to provide sufficient information to characterize the thermal behavior of a material. It is these drawbacks that have perhaps limited the development and application range of these techniques. However, these EAS reviews have revealed that these limitations have been essentially overcome, not only by the introduction of new TA techniques and further development of the less common ones, but also by refinement of the well-known traditional TA techniques. ‘Long experimental times’ have been very significantly reduced in modern DSC systems, such that DSC has become the definitive method for measurement of heat capacity in the solid state. New systems such as ‘Hi-Res’ TG<sup>TM</sup> have significantly reduced experimental time but not at the expense of resolution and sensitivity. The new SCTA techniques can be associated with long experimental times, but ‘time-resolved SCTA’ appears to have addressed this problem. ‘Preparative-scale’ TG systems are now in evidence, which provide sensitivity levels comparable to the wide range of small-scale TG systems. The problem relating to the ‘averaged nature of data output’ has been comprehensively overcome by the development of ‘modulated’ TA techniques and perhaps this has been the most significant recent advance in TA technology. Overcoming the principal disadvantages of TA techniques, coupled with the development of new, powerful, specialized techniques such as  $\mu$ -TA and DETA, have ensured a prominent future for TA in the twenty-first century – particularly in terms of TA being the preeminent analytical group of techniques in materials science.

*See also:* **Thermal Analysis:** Overview; Temperature-Modulated Techniques; Coupled Techniques; Sample-Controlled Techniques.

## Further Reading

- Balek V (1991) Emanation thermal analysis and its application potential. *Thermochimica Acta* 1921–1917.
- Bellchamber RM, Betteridge D, Collins MP, Lilley T, and Wade AP (1986) Quantitative study of acoustic emission from a model chemical process. *Analytical Chemistry* 58: 1873–1877.

- Betteridge D, Joslin MT, and Lilley T (1981) Acoustic emissions from chemical reactions. *Analytical Chemistry* 53: 1064–1073.
- Clark GM (1978) Instrumentation for thermosonimetry. *Thermochimica Acta* 27: 19–25.
- Hill JO and Chirawongaram S (1994) Thermal analysis studies of tin dithiocarbamate complexes: A short review. *Journal of Thermal Analysis* 41: 511–518.
- Lonvic (1993) Thermosonimetric (TS) investigation of the quartz-cristobalite conversion of quartzites. *Thermochimica Acta* 214: 51–57.
- Mandelis A (1991) Photothermal applications to the thermal analysis of solids. *Journal of Thermal Analysis* 37: 1065–1101.
- Mraz T, Rajeshwar K, and Dubar J (1980) An automated technique for thermoacoustimetry of solids. *Thermochimica Acta* 38: 211–223.
- Parkes GMB, Barnes PA, Bond G, and Charsley EL (2000) Qualitative and quantitative aspects of microwave thermal analysis. *Thermochimica Acta* 356: 85–96.
- Price DM (1997) Thermal analysis in industrial environments. *Journal of Thermal Analysis & Calorimetry* 49: 953–959.
- Rajeshwar K (1974) Dynamic dielectric analysis – a new thermal analysis technique. *Thermochimica Acta* 33: 157–168.
- Saffell JR, Matthiesen A, McIntyre R, and Ibar JP (1991) Comparing TSCA with other thermoanalytical methods to characterise the amorphous phase of polymers. *Thermochimica Acta* 192: 243–264.
- Shimada S (1990) A study of the thermal decomposition of sodium and potassium perchlorates by acoustic emission thermal analysis. *Thermochimica Acta* 163: 313–316.
- Shimada S (1993) Thermosonimetry and thermomicroscopy study of the decomposition of sodium and potassium perchlorate. *Journal of Thermal Analysis* 40: 1063–1068.

## THERMAL LENSING SPECTROMETRY

**T Imasaka**, Kyushu University, Fukuoka, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 5167–5173, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

### Introduction

In conventional absorption spectrometry, the light intensity absorbed by a sample is given by the Lambert–Beer law:

$$\log(I_0/I) = A \quad [1]$$

where

$$A = \epsilon bc \quad [2]$$

$I_0$  and  $I$  are the light intensities before and after passage through the sample, and  $A$  is the absorbance. The parameter  $\epsilon$  is the molar absorptivity,  $b$  is the pathlength, and  $c$  is the sample concentration. When the sample is diluted and the absorbance is sufficiently small, eqn [1] can be rewritten as

$$(I_0 - I)/I = 2.303A \quad [3]$$

In order to measure low light absorption, it is necessary to stabilize the intensity of the light introduced into the sample and to reduce fluctuations of the

transmitted light. Such fluctuations are caused by variation of the beam path due to refractive index changes that occur when there is turbulent flow in the sample. Thus, the sensitivity of conventional absorption spectrometry is independent of the light intensity and cannot be improved even when a laser with a large output power is used as a light source.

In 1964, Gordon and co-workers observed the build-up and decay of a transient optical phenomenon when a sample cell was placed in the resonator of a helium–neon laser. This phenomenon is produced by a thermal lens effect induced by low light absorption in the sample. Local heating and the resulting refractive index gradients generated in the sample cause it to act as a diverging lens. Since the magnitude of this effect is proportional to the intensity of the light, it is possible to measure low levels of light absorption by increasing the output power of the laser. This spectroscopic technique has been used for measuring low levels of light absorption caused by overtone vibrations of the molecule in the visible region and for measuring fluorescence quantum yields of dyes.

### Principles

A schematic diagram of a thermal lens spectrometer is shown in **Figure 1**. The laser beam is focused by a lens and is directed at the sample. When the sample is placed at the focal point of the lens, a strong thermal



- Betteridge D, Joslin MT, and Lilley T (1981) Acoustic emissions from chemical reactions. *Analytical Chemistry* 53: 1064–1073.
- Clark GM (1978) Instrumentation for thermosonimetry. *Thermochimica Acta* 27: 19–25.
- Hill JO and Chirawongaram S (1994) Thermal analysis studies of tin dithiocarbamate complexes: A short review. *Journal of Thermal Analysis* 41: 511–518.
- Lonvic (1993) Thermosonimetric (TS) investigation of the quartz-cristobalite conversion of quartzites. *Thermochimica Acta* 214: 51–57.
- Mandelis A (1991) Photothermal applications to the thermal analysis of solids. *Journal of Thermal Analysis* 37: 1065–1101.
- Mraz T, Rajeshwar K, and Dubar J (1980) An automated technique for thermoacoustimetry of solids. *Thermochimica Acta* 38: 211–223.
- Parkes GMB, Barnes PA, Bond G, and Charsley EL (2000) Qualitative and quantitative aspects of microwave thermal analysis. *Thermochimica Acta* 356: 85–96.
- Price DM (1997) Thermal analysis in industrial environments. *Journal of Thermal Analysis & Calorimetry* 49: 953–959.
- Rajeshwar K (1974) Dynamic dielectric analysis – a new thermal analysis technique. *Thermochimica Acta* 33: 157–168.
- Saffell JR, Matthiesen A, McIntyre R, and Ibar JP (1991) Comparing TSCA with other thermoanalytical methods to characterise the amorphous phase of polymers. *Thermochimica Acta* 192: 243–264.
- Shimada S (1990) A study of the thermal decomposition of sodium and potassium perchlorates by acoustic emission thermal analysis. *Thermochimica Acta* 163: 313–316.
- Shimada S (1993) Thermosonimetry and thermomicroscopy study of the decomposition of sodium and potassium perchlorate. *Journal of Thermal Analysis* 40: 1063–1068.

## THERMAL LENSING SPECTROMETRY

**T Imasaka**, Kyushu University, Fukuoka, Japan

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 5167–5173, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

### Introduction

In conventional absorption spectrometry, the light intensity absorbed by a sample is given by the Lambert–Beer law:

$$\log(I_0/I) = A \quad [1]$$

where

$$A = \epsilon bc \quad [2]$$

$I_0$  and  $I$  are the light intensities before and after passage through the sample, and  $A$  is the absorbance. The parameter  $\epsilon$  is the molar absorptivity,  $b$  is the pathlength, and  $c$  is the sample concentration. When the sample is diluted and the absorbance is sufficiently small, eqn [1] can be rewritten as

$$(I_0 - I)/I = 2.303A \quad [3]$$

In order to measure low light absorption, it is necessary to stabilize the intensity of the light introduced into the sample and to reduce fluctuations of the

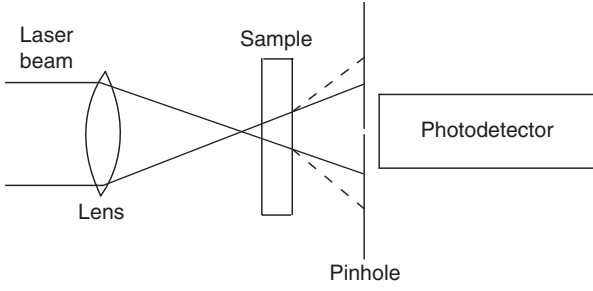
transmitted light. Such fluctuations are caused by variation of the beam path due to refractive index changes that occur when there is turbulent flow in the sample. Thus, the sensitivity of conventional absorption spectrometry is independent of the light intensity and cannot be improved even when a laser with a large output power is used as a light source.

In 1964, Gordon and co-workers observed the build-up and decay of a transient optical phenomenon when a sample cell was placed in the resonator of a helium–neon laser. This phenomenon is produced by a thermal lens effect induced by low light absorption in the sample. Local heating and the resulting refractive index gradients generated in the sample cause it to act as a diverging lens. Since the magnitude of this effect is proportional to the intensity of the light, it is possible to measure low levels of light absorption by increasing the output power of the laser. This spectroscopic technique has been used for measuring low levels of light absorption caused by overtone vibrations of the molecule in the visible region and for measuring fluorescence quantum yields of dyes.

### Principles

A schematic diagram of a thermal lens spectrometer is shown in **Figure 1**. The laser beam is focused by a lens and is directed at the sample. When the sample is placed at the focal point of the lens, a strong thermal





**Figure 1** Schematic diagram for a thermal lens spectrometer.

lens effect takes place. However, the lens effect cannot be measured using the laser beam since it is formed at the focal point of the laser. The sample is then shifted slightly from the focal point and the beam expansion caused by the thermal lens effect is measured. This is accomplished quantitatively by placing a pinhole in front of a photodetector, which then measures the change in light intensity at the beam center. Typical waveforms are shown in **Figure 2**.

A model has been proposed to describe the thermal lens effect in which the signal intensity is expressed by

$$[I_{be}(t=0) - I_{be}(t=\infty)]/I_{be}(t=\infty) = 1 - \theta + \theta^2/2 \quad [4]$$

$$\theta = P_{abs}(dn/dT)/\lambda k \quad [5]$$

where  $I_{be}(t=0)$  and  $I_{be}(t=\infty)$  are the light intensities at the beam center at  $t=0$  and  $r=\infty$ , respectively. The parameter  $\theta$  is a dimensionless quantity and is approximately the difference between the thermally induced phase shifts at  $r=0$  and  $r=\omega$ , where  $r$  is the distance from the beam center and  $\omega$  is the beam radius. The value of  $P_{abs}$  is the power absorbed by the sample,  $(dn/dT)$  is the variation of refractive index with temperature,  $\lambda$  is the wavelength of the laser, and  $k$  is the heat conductivity. The higher-order term,  $\theta^2/2$ , can be neglected when the thermal lens is optically thin and the signal intensity is sufficiently small (e.g.,  $<0.1$ ). In this case,  $P_{abs}$  is expressed by

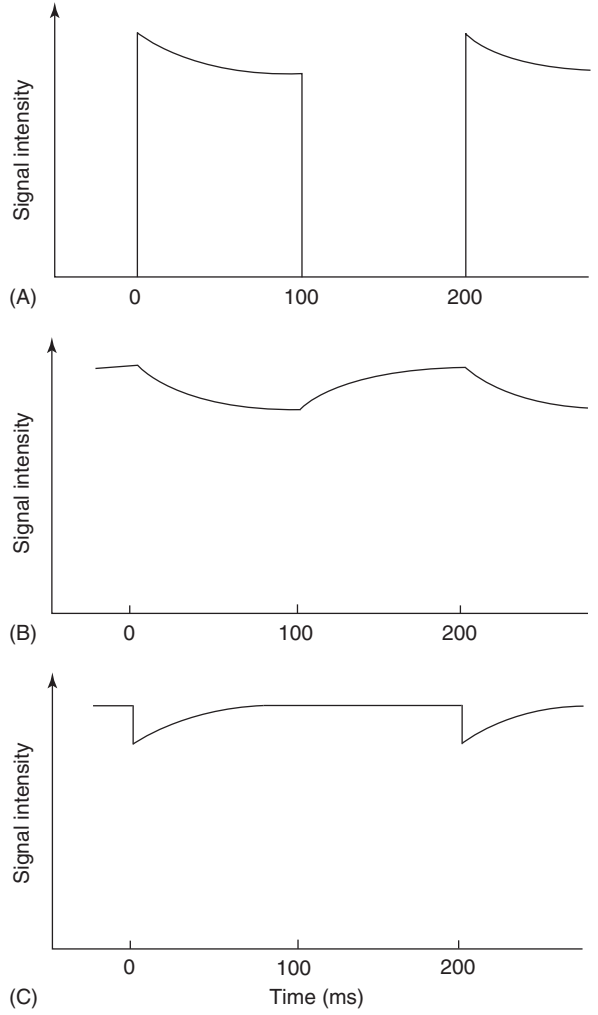
$$P_{abs} = 2.303PA \quad [6]$$

where  $P$  is the power of the laser beam introduced. Equation [4] then becomes modified to:

$$[I_{be}(t=0) - I_{be}(t=\infty)]/I_{be}(t=\infty) = 2.303E_cA \quad [7]$$

$$E_c = P(dn/dT)/\lambda k \quad [8]$$

where  $E_c$  is the enhancement factor, i.e., the sensitivity of thermal lens spectrometry relative to conventional absorption spectrometry, obtained under continuous-wave (CW) excitation.



**Figure 2** Transient curve of thermal lens signal: (A) single-beam system using CW laser; (B) dual-beam system using CW exciting laser; (C) dual-beam system using pulsed exciting laser.

When a pulsed laser is used as an exciting source, a different model is used. Under one-photon excitation, the signal intensity is expressed by

$$[I_{be}(t=\infty) - I_{be}(t=0)]/I_{be}(t=0) = 2.303E_1A \quad [9]$$

$$E_p = -3^{3/2}(E_l/\lambda_p\omega_{0p}^2)(1/\rho C_p)(dn/dT) \quad [10]$$

where  $E_p$  is the enhancement factor obtained under pulsed laser excitation,  $E_l$  is the pulse energy of the exciting laser,  $\lambda_p$  is the wavelength,  $\omega_{0p}$  is the beam waist radius,  $\rho$  is the density of the sample, and  $C_p$  is its specific heat capacity.

The enhancement factor is proportional to the output power of the exciting laser under CW laser excitation and to the pulse energy under pulsed laser excitation. Accordingly, the sensitivity of thermal

lens spectrometry is improved by using an intense laser source. In CW thermal lens spectrometry, the enhancement factor is not affected by the laser beam radius, so that the sensitivity is independent of the focusing condition. This is due to the more rapid heat dissipation across the beam radius that occurs when the laser beam is more tightly focused. In pulsed thermal lens spectrometry, the enhancement factor is inversely proportional to the square of the beam radius. Thus, the sensitivity can be substantially improved by concentrating the beam tightly, although this reduces its effective pathlength.

When the exciting beam is interrupted, the signal decays by thermal diffusion across the heating region. The transient decay curve is given by

$$I_{\text{be}}(t)/I_{\text{be}}(t=0) = [1 - \theta(1 + t_d/2t)^{-1}]^{-1} \quad [11]$$

(model of eqn [4])

or

$$I_{\text{be}}(t)/I_{\text{be}}(t=0) = (1 + 2t/f_d)^{-2} \quad (\text{model of eqn [9]}) \quad [12]$$

$$t_d = \omega^2/4D \quad [13]$$

$$D = k/\rho C_D \quad [14]$$

where  $t_d$  is the decay time constant,  $\omega$  is the beam radius at the sample, and  $D$  is the thermal diffusivity. Therefore, the decay time depends on the beam radius in CW thermal lens spectrometry as well as in pulsed thermal lens spectrometry.

In later work the observed signal was found to be lower than the theoretical value due to the aberrant nature of the thermal lens. In this case, the experimental value becomes consistent with the theoretical value when it is multiplied by a factor of 0.52.

## Enhancement Factor

The sensitivity of thermal lens spectrometry is estimated simply by calculating the enhancement factor. The values in **Table 1** are obtained by assuming that an argon ion laser producing 1 W or a dye laser (500 nm,  $\omega = 0.1$  mm) producing 1 mJ is used as a typical exciting source. The enhancement factor of the medium is not determined by the solute or sample gas but is primarily determined by the solvent or buffer gas.

These calculated values provide a guideline for the application of thermal lens spectrometry to trace analysis. An enhancement factor of 140 is obtained under CW laser excitation even in the worst case, that of water. Thermal lens spectrometry is therefore

**Table 1** Enhancement factors obtained under CW ( $E_c$ ) and pulsed ( $E_p$ ) laser excitation

Medium	$E_c$ (1 W)	$E_D$ (1 mJ)
<i>Liquid</i>		
Water	140	20
Ethanol	2400	210
Benzene	4700	430
Carbon tetrachloride	5900	440
<i>Gas</i>		
Helium	0.8	140
Air	35	770
Carbon dioxide	87	950
Benzene	535	1700

Values multiplied by 0.52 (see text).

useful for sensitive detection of the sample. Moreover, organic solvents give enhancement factors 20–40 times greater than that of water. This is due to organic solvents having heat conductivities about four times smaller than that of water, while the variation of refractive index with temperature is about six times larger for organic solvents than for water. This result indicates that extraction of the sample to be analyzed from an aqueous solution into an organic solvent is advantageous in terms of improved sensitivity. However, the enhancement factor is rather small for gas-phase samples in CW thermal lens spectrometry; this is due to the rapid dissipation of heat in gaseous samples. On the other hand, the enhancement factor is relatively large for gas-phase samples in pulsed thermal lens spectrometry, although the values for liquid-phase samples are much lower than those in CW thermal lens spectrometry.

These differences originate in the thermo-optical properties of the different media and in the nature of CW or pulsed laser excitation. For example, the decay time is generally longer for a liquid-phase sample, since it has a higher density. Heat generated locally within the sample builds up sufficiently in CW thermal lens spectrometry. In gas-phase samples heat is dissipated more rapidly, producing a weaker thermal lens. Furthermore, the background signal due to light absorption by impurities and overtone vibrations in the solvent can be effectively subtracted when a CW laser with a stable output power is used. Therefore, CW thermal lens spectrometry is the preferred technique for application to liquid-phase samples.

In contrast, pulsed laser excitation forms a thermal lens instantly and this lens effect is detected before heat dissipation in the sample medium occurs. The density of a gas-phase sample is low, which decreases the heat capacity per unit volume and increases the

temperature rise. Thus, a strong thermal lens occurs in gas-phase samples in pulsed thermal lens spectrometry. The enhancement factor can be further improved by tightly focusing the beam for a gas-phase sample. In liquid-phase samples laser breakdown occurs, limiting the maximum pulse energy that can be used. Pulsed thermal lens spectrometry is therefore preferable for gas-phase sample applications.

### Improvement of Enhancement Factor

In order to improve the enhancement factor, several approaches have been developed and their advantages have been verified experimentally. When the temperature of a binary liquid mixture of aniline and cyclohexane is adjusted to be slightly below the critical temperature ( $T_c = 30.585^\circ\text{C}$ ), the signal intensity is enhanced 44 times over that for the individual solvents. This anomalous phenomenon is considered to be due to a thermally induced concentration gradient. A similar effect is observed for the supercritical fluid carbon dioxide. The value of  $(dn/dT)$  is greatly increased at around the supercritical point ( $T_c = 31.0^\circ\text{C}$ ,  $P_c = 72.8\text{ atm}$ ), providing an enhancement factor 150 times larger than for carbon tetrachloride. The apparent enhancement factor increases when an exothermic (or endothermic) chemical reaction occurs during irradiation by the exciting laser. Local heating induces a strong temperature gradient, and therefore an anomalous thermal lens effect is (apparently) observed. The effects described above provide an extremely large enhancement factor, although an increasing fluctuation of the background signal also occurs. It is necessary, therefore, to carefully control experimental parameters such as temperature, pressure, and mechanical stability of the optical table.

### Exciting Laser

The enhancement factor calculated theoretically is obtained experimentally only when an ideal laser source is used. The enhancement factor is strongly affected by the beam quality of the laser. Thus, a laser with a single transverse mode must be used to obtain an enhancement factor close to the theoretical one. A CW laser, such as an argon ion laser, is usually operated in the single mode and is preferred in thermal lensing spectrometry. A pulsed laser, such as a dye laser pumped by an excimer laser, provides rather poor beam quality, unless the beam shape is specially controlled by using a pinhole followed by amplification. The enhancement factor obtained experimentally in pulsed thermal lens spectrometry is usually small in comparison with the theoretical value. Aberration of the focusing lens may also affect the

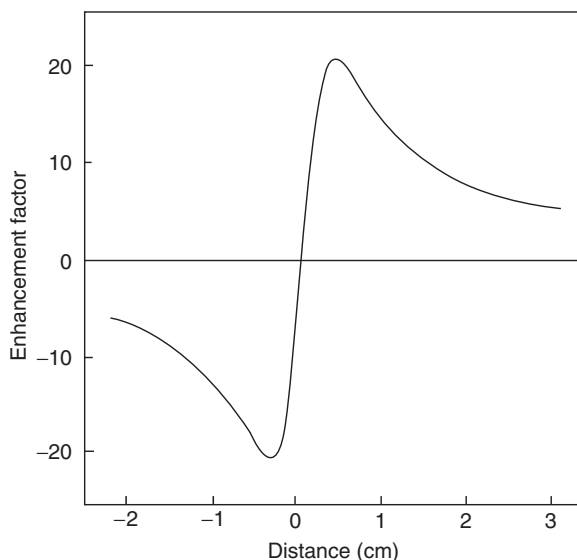
enhancement factor and should be minimized by careful selection of the lens.

### Sample

The enhancement factor is strongly affected by the sample position, as shown in Figure 3. Thermal focusing and defocusing effects are observed when the sample is positioned before and after the focal point, respectively. In theoretical calculations of the enhancement factor, the sample is assumed to be positioned at a point a distance of  $\sqrt{3}$  times the confocal distance after the focal point to provide a maximum signal. The confocal distance is defined as the distance between the focal point and the position at which the beam radius is  $\sqrt{2}$  times that at the beam waist. Thus, in order to achieve the optimum enhancement factor the sample should be carefully positioned and be sufficiently thin. Particles suspended in the sample introduce severe scatter into the laser light, since the laser beam is tightly focused in the sample. The sample must therefore be filtered carefully before making the measurement. Temperature variation of the sample may also affect the sensitivity, but most experiments are carried out without temperature control since it is only the temperature gradient that occurs in a short time period that is measured in thermal lens spectrometry.

### Measurement

In order to measure the thermal lens signal, the light intensities at  $t = 0$  and  $t = \infty$  must be measured. The



**Figure 3** Dependence on cell position of enhancement factor.

transient signal is recorded by a waveform digitizer and is analyzed by a microcomputer. The observed signal is fitted to a theoretical curve, and  $I(t=0)$  and  $I(t=\infty)$  are calculated. Using this technique, a minimum absorptivity of  $7.8 \times 10^{-8}$  is detected using a 160 mW argon ion laser. By placing reference and sample cells before and after the focal point, the background drift due to variation of the laser output power is cancelled out. In most experiments a pinhole is placed in front of the photodetector to allow measurement of the light intensity at the beam center. However, the noise can be reduced by measuring a one-dimensional beam profile with an image sensor. The precision, i.e., the sensitivity, is improved 3.8 times by using this technique. Further improvement is achieved by replacing the pinhole with an optical mask, consisting of  $3 \times 10^4$  pixels, and its density is proportional to a parabolic function, i.e.,  $F(r) = r^2/r_{\max}^2$ . This technique allows two-dimensional reduction of noise without using a digital computation technique. The precision in signal measurement is improved by a factor of 16. Other spectrometric techniques have been developed to improve the enhancement factor, e.g., by using a Mach-Zehnder interferometer or an intracavity quenching effect.

### Single-Beam and Dual-Beam Configurations

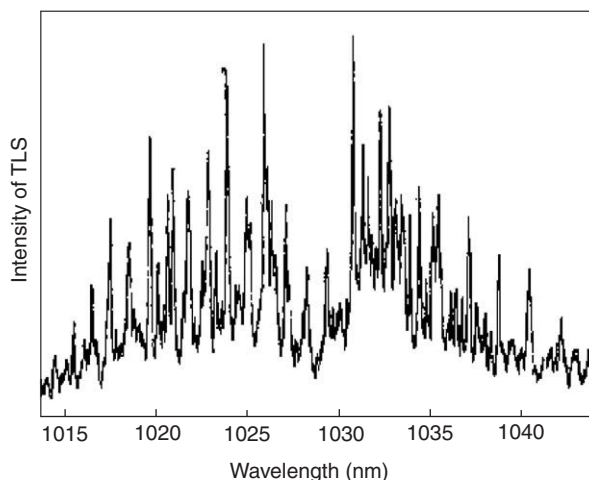
A thermal lens spectrometer is easily constructed using a single laser. However, a waveform digitizer equipped with a microcomputer must be used to analyze the transient curve, as described above. On the other hand, it is possible to use two lasers, i.e., an exciting laser and a probe laser, in thermal lens spectrometry. These laser beams, emitting at different wavelengths, are coaxially aligned and focused into a sample cell. The exciting beam is modulated by a beam chopper to form a periodic thermal lens and is interrupted by a filter placed in front of a pinhole. The signal can then be measured easily by using a lock-in amplifier, cf. **Figure 2B**. However, the beam diameter, i.e., the beam divergence, of the probe laser is usually different from that of the exciting laser. This means that the two laser beams are not focused at the same position. Mode mismatch also occurs due to aberration of the focusing lens – the focal length of the lens depends on the wavelength of the light and is therefore different for the exciting and probe lasers. Mode mismatch results in asymmetry in the dependence on the sample position; the maximum enhancement factor obtained by the thermal focusing effect is different from that obtained by the thermal defocusing effect. It is also possible to conduct a dual-beam experiment using a single laser. In this technique, the laser beam

is split into two parts by a polarization beam splitter. One part is used as the heating beam and the other part as the probe beam.

## Applications

### Gas Phase

**Figure 4** shows a thermal lens spectrum for ammonia recorded by using a dye laser pumped by an excimer laser as an exciting source. The laser wavelength is extended to the near-infrared (IR) region by stimulation due to Raman effect. This spectrum consists of the third overtone vibration of the  $\nu_1$  band for ammonia. A similar spectrum is recorded for atmospheric nitrogen dioxide. The detection limit is  $4 \text{ ng ml}^{-1}$ , which is three orders of magnitude better than that of a conventional absorption spectrometer. The enhancement factor achieved is 16 at a pulse energy of 1.3 mJ, which is lower than the theoretical value by a factor of 33. As pointed out earlier, this is due to the poor beam quality of the dye laser used. Nitrogen dioxide is also measured by using a 700 mW argon ion laser. The linear analytical range extends to  $250 \mu\text{g ml}^{-1}$ , the detection limit being  $5 \text{ ng ml}^{-1}$ . The enhancement factor observed is 7. This small enhancement factor is ascribed to the application of CW thermal lens spectrometry to a gas-phase sample. However, the detection limits achieved in the above experiments are substantially better than detection limits obtained by conventional absorption spectrometry. This is due to efficient reduction of noise by periodic modulation of the exciting beam using boxcar or lock-in amplifiers. A thermal lens spectrum is also measured in the IR region by using a CW line-tunable carbon dioxide laser (7 W), in which differentiation of several organic gases is



**Figure 4** Thermal lens spectrum (TLS) for ammonia.

demonstrated. The detection limit reported is 12 ppb for methanol, which corresponds to an absorbance of  $2.7 \times 10^{-7}$ . Use of a pulsed carbon dioxide laser (100 mJ) increases the enhancement factor for a gas-phase sample. The detection limit reported is  $10 \text{ ng ml}^{-1}$  for dichlorodifluoromethane.

### Liquid Phase

In one analytical application of thermal lensing spectrometry, Cu(II) is determined after color development with ethylenediaminetetraacetic acid using a 4 mW helium–neon laser. The enhancement factor achieved is 2.0, and the minimum detectable absorbance is  $1.0 \times 10^{-3}$  corresponding to 3.3 ng of copper. More sensitive analysis is achieved by using an argon ion laser with a larger output power (600 mW). For example, Fe(II) is determined by chelation with bathophenanthroline disulfonic acid in an aqueous solution. The enhancement factor achieved is 72, with a reported detection limit of  $5 \times 10^{-10} \text{ mol l}^{-1}$  ( $0.03 \text{ ng ml}^{-1}$ ). For an Fe(II)–bathophenanthroline disulfonate complex extracted into chloroform, the detection limit is further improved to  $3 \times 10^{-11} \text{ mol l}^{-1}$ . The enhancement factor is 1040, the minimum detectable absorbance being  $6 \times 10^{-7}$ . By using a CW dye laser pumped by an argon ion laser, phosphorus is determined to  $5 \text{ pg ml}^{-1}$  based on the molybdenum blue method. Phosphorus is also measured by solvent extraction of a molybdophosphate/auramine ion pair into 1:2 isobutanol/hexane. The compound has an absorption band at 440 nm, and a helium–cadmium laser emitting at 442 nm (100 mW) is used for sample excitation. The detection limit reported is  $600 \text{ pg ml}^{-1}$ , which is restricted by variable blank measurement. A water-soluble analyte such as  $\text{Tb}^{3+}$  is measured by dissolving it in a reverse micelle, instead of using solvent extraction.

A pulsed dye laser is also used for measurement of liquid-phase samples, because of its wide tunability. For example, Cu(II) is determined by using Pb(II)-tetrakis (4-*N*-methylpyridyl)porphinetetra-*p*-toluene sulfonate; a sharp Soret band appears at 422 nm ( $\epsilon = 1.5 \times 10^5 \text{ l mol}^{-1} \text{ cm}^{-1}$ ). However, this approach is less successful because of a poor background subtraction capability, which is ascribed to pulse-to-pulse instability of the pulse energy of the pulsed dye laser.

### Practical Thermal Lens Systems

In order to simplify the thermal lens spectrometer, a compact and inexpensive diode laser may be used as an exciting source. For example, phosphorus forms a blue complex with molybdenum, which has an

absorption band in the near-IR region (780–870 nm). Phosphorus is determined in an aqueous solution to 2.2 and  $0.7 \text{ ng ml}^{-1}$  with single- and dual-beam systems, respectively. When the sample is measured after solvent extraction into 2-butanol, the detection limits are improved to  $0.21 \text{ ng ml}^{-1}$  for both methods. The maximum enhancement factor achieved is 61 with a 10 mW laser. The minimum detectability is limited primarily by blank absorption occurring from the solvent. The solvent blank can be reduced by using chloroform, but the detection limit obtained by ion pair extraction of the complex into chloroform using myristyldimethylbenzylammonium chloride (Zephiramine) is unchanged since in this case the minimum detectability is limited by reagent blank.

A similar procedure is followed in the determination of Fe(II) with 2-nitroso-5-diethylaminophenol using ion pair extraction into chloroform. The enhancement factor achieved experimentally is 47. The detection limit for Fe(II) is  $8 \times 10^{-9} \text{ mol l}^{-1}$ , which is three to four times better than that obtained by conventional absorption spectrometry. The detection limit is governed by the background signal originating from the impurity of Fe(II) contained in the reagent.

In the thermal lens system, many optical components are necessary, including a beam splitter, a focusing lens, a sample cell, and a pinhole. The positions of these components must be carefully optimized and be fixed rigidly throughout the experiment. This requirement raises difficulties in the application of thermal lens spectrometry to practical analysis. A simple thermal lens system can be constructed using a pair of optical fibers placed facing each other, as shown in Figure 5. In this system the sample is placed between the fibers; in other words the fibers are immersed in the sample. This configuration removes the sample cell and allows detection of the sample in a high-pressure vessel or a high-temperature reactor. When the laser beam is emitted from the optical fiber, it diverges as if focused at the distal end of the fiber. In order to retain the high

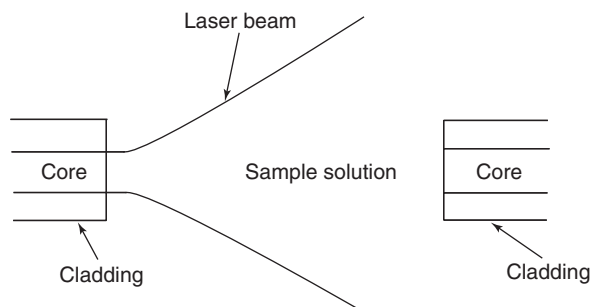


Figure 5 Fiber-coupled thermal lens spectrometer.



beam quality of the laser, a single-mode fiber must be used for light introduction; many transverse modes appear in multimode fibers and therefore the focusing capability is degraded. It should be noted that the single-mode fiber is designed for a specific wavelength and the fiber should be matched to the laser wavelength used. A multimode fiber may be used for light collection. The enhancement factor achieved is 42, but this enhancement factor is only 5% of the theoretical value. This is attributed to the small core diameter (6  $\mu\text{m}$ ) of the fiber, resulting in a short confocal distance (31  $\mu\text{m}$ ). This distance is much shorter than the pathlength (1 cm) assumed in calculating the enhancement factor. In order to extend the effective pathlength, an optical fiber with a larger core diameter would seem to be necessary, but such fibers are not yet commercially available.

### Gas Chromatography Detector

Samples such as dichlorodifluoromethane, chlorotri-fluoromethane, and ethanol are separated by a conventional gas chromatograph column (Carbowax 20M) and are excited at the P32 line of the carbon dioxide laser (933  $\text{cm}^{-1}$ ). The reported detection limit is an absorbance of  $2.5 \times 10^{-7}$ .

### Liquid Chromatography Detector

In early work, nitroaniline isomers were separated by a conventional octadecyl silane (ODS) column and detected by a thermal lens spectrometer consisting of a 190 mW argon ion laser. The detector volume was 10  $\mu\text{l}$ , and the sample pathlength 1 cm. The enhancement factor achieved was 120 and the time constant was 28 ms. The detection limits were from the sub-nanogram to several-nanograms range. In more recent work, the minimum absorbance detected has been improved to  $1.2 \times 10^{-6}$ . Thermal lens spectrometry has also been applied to microcolumn liquid chromatography. The detection volume and the pathlength are 0.5  $\mu\text{l}$  and 1 mm, respectively. The minimum absorbance detected is  $2\text{--}4 \times 10^{-6}$ . When an open tubular column is used for sample separation, the pathlength is further reduced to 100 or 200  $\mu\text{m}$ . The reported baseline drift is approximately  $3 \times 10^{-5}$  absorbance. The detection limit achieved is 30 pg for *o*-nitroaniline. In order to reduce the detection volume further, cross-beam thermal lens spectrometry (thermal refraction spectrometry) is used. Detection limits reported for dimethylaminoazobenzenesulphonyl derivatives are 0.75 fmol, at which only 50 analyte molecules are expected within the probe volume (0.2 pl) at the peak maximum.

### Detector for Electrophoresis

Many biological samples are now separated by polyacrylamide gel electrophoresis. For example, protein is separated on a gel plate and is stained by Coomassie Brilliant Blue R250 or G250. This is placed on a translator, which is electrically driven for recording a chromatogram. Capillary electrophoresis has recently become popular because of its better sample separation and the capability of real-time sample monitoring. Twenty phenylthiohydantoin-amino acids are separated in a capillary (50  $\mu\text{m}$  ID, 50 cm long), and the sample is detected by a cross-beam thermal lens spectrometer using a highly repetitive krypton fluoride excimer laser with a low pulse energy. The detection limit is  $9 \times 10^{-7} \text{ mol l}^{-1}$  of sample injected in a 0.6 nl volume, corresponding to 0.5 fmol (100 fg) of labeled amino acids.

### Limitations

Thermal lens spectrometry is a sensitive analytical technique for measuring small amounts of light absorption. However, lasers with large output power or large pulse energy are necessary in order to obtain a sufficiently large enhancement factor for an aqueous-phase sample. Thus, small lasers such as an air-cooled argon ion laser or a pulsed dye laser pumped by a nitrogen laser are not suitable for trace analysis. To overcome this problem, the sample may be extracted into an organic solvent to improve the enhancement factor. However, the extraction efficiency is sometimes limited at low sample concentrations. Furthermore, background absorption originating from impurities in the solvent is not negligible, usually occurring in the UV region. At the same time, background absorption originating from overtone vibration of solvent molecules is unavoidable in the near-IR region. Thus, it is necessary to subtract the background signal carefully and effectively.

The absorption band is rather broad, especially for a liquid-phase sample, and at present only CW line-tunable lasers are usable in practice. Spectral selectivity is therefore quite limited in thermal lens spectrometry. Analytical selectivity is improved by a chemical reaction before the measurement, e.g., by chelation of ions with a specific ligand. However, chelation is not necessarily completed at ultratrace levels.

In order to obtain an enhancement factor close to the theoretical one, the sample pathlength must be much shorter than the confocal distance. It is possible to increase the pathlength, e.g., to 10 cm, by focusing the laser beam loosely, but this decreases the time constant in signal decay. To compensate, the



modulation frequency must be decreased, resulting in a poor signal-to-noise ratio in signal detection.

It appears that thermal lens spectrometry is truly advantageous only when it is applied to a chromatograph detector, when a small pathlength is required, e.g., 1 mm or less. In the case of liquid chromatography, the background signal is effectively subtracted by using a stable CW laser. Note that there is no appreciable background signal in gas chromatography using an ultrapure carrier gas. In such chromatography, additional selectivity is given by a separation procedure. However, laser fluorometry is generally more sensitive, and therefore to make the best use of thermal lens spectrometry its application should be restricted to nonfluorescent samples. Many biological samples, e.g., deoxyribonucleic acid, have absorption bands in the visible or UV region, and they are nonfluorescent at room temperature. Moreover, most chelates containing heavy metals are known to be nonfluorescent. Therefore, thermal lens spectrometry may be useful for the measurement of such biological and chemical species. Finally, it is emphasized that the background

originating from the reagent, solvent, and, sometimes, the sample species itself included as an impurity in the reagent must be carefully reduced for ultratrace analysis using thermal lens spectrometry, as it is in other ultrasensitive spectrometric techniques.

*See also:* **Capillary Electrophoresis:** Overview. **Electrophoresis:** Polyacrylamide Gels. **Gas Chromatography:** Detectors. **Laser-Based Techniques.** **Liquid Chromatography:** Instrumentation.

## Further Reading

- Snook RD (1997) Laser techniques for chemical analysis. *Chemical Society Reviews* 26: 319–326.
- Snook RD and Lowe RD (1995) Thermal lens spectrometry – a review. *Analyst* 120: 2051–2068.
- Tokeshi M, Minagawa T, Uchiyama K, *et al.* (2002) Continuous-flow chemical processing on a microchip by combining microunit operations and a multiphase flow network. *Analytical Chemistry* 74: 1565–1571.
- Tran CD (1993) Photothermal detectors for high-performance liquid chromatography. In: Patonay G (ed.) *HPLC Detection: Newer Methods*. New York: Wiley.

# THERMOGRAVIMETRIC ANALYSIS

See **THERMAL ANALYSIS: Overview**

# THIN-LAYER CHROMATOGRAPHY

Contents

**Overview**

**Principles**

**Plate Technology**

**Method Development**

**Instrumentation**

## Overview

**E Reich and A Blatter**, CAMAG Laboratory, Muttenz, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

This article serves as an overview covering definition, historical development, basic principles, and advantages

of thin-layer chromatography (TLC) in comparison to column chromatography. It concludes with a summary of common applications. Details on theory, instrumentation, plate technology, and method development are covered in subsequent articles.

## Definition and General Description

TLC is a form of liquid chromatography. It is used for rapid separation of samples into components based on differences in their retention behavior. The

modulation frequency must be decreased, resulting in a poor signal-to-noise ratio in signal detection.

It appears that thermal lens spectrometry is truly advantageous only when it is applied to a chromatograph detector, when a small pathlength is required, e.g., 1 mm or less. In the case of liquid chromatography, the background signal is effectively subtracted by using a stable CW laser. Note that there is no appreciable background signal in gas chromatography using an ultrapure carrier gas. In such chromatography, additional selectivity is given by a separation procedure. However, laser fluorometry is generally more sensitive, and therefore to make the best use of thermal lens spectrometry its application should be restricted to nonfluorescent samples. Many biological samples, e.g., deoxyribonucleic acid, have absorption bands in the visible or UV region, and they are nonfluorescent at room temperature. Moreover, most chelates containing heavy metals are known to be nonfluorescent. Therefore, thermal lens spectrometry may be useful for the measurement of such biological and chemical species. Finally, it is emphasized that the background

originating from the reagent, solvent, and, sometimes, the sample species itself included as an impurity in the reagent must be carefully reduced for ultratrace analysis using thermal lens spectrometry, as it is in other ultrasensitive spectrometric techniques.

*See also:* **Capillary Electrophoresis:** Overview. **Electrophoresis:** Polyacrylamide Gels. **Gas Chromatography:** Detectors. **Laser-Based Techniques.** **Liquid Chromatography:** Instrumentation.

## Further Reading

- Snook RD (1997) Laser techniques for chemical analysis. *Chemical Society Reviews* 26: 319–326.
- Snook RD and Lowe RD (1995) Thermal lens spectrometry – a review. *Analyst* 120: 2051–2068.
- Tokeshi M, Minagawa T, Uchiyama K, *et al.* (2002) Continuous-flow chemical processing on a microchip by combining microunit operations and a multiphase flow network. *Analytical Chemistry* 74: 1565–1571.
- Tran CD (1993) Photothermal detectors for high-performance liquid chromatography. In: Patonay G (ed.) *HPLC Detection: Newer Methods*. New York: Wiley.

# THERMOGRAVIMETRIC ANALYSIS

See **THERMAL ANALYSIS: Overview**

# THIN-LAYER CHROMATOGRAPHY

Contents

**Overview**

**Principles**

**Plate Technology**

**Method Development**

**Instrumentation**

## Overview

**E Reich and A Blatter**, CAMAG Laboratory, Muttenz, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

This article serves as an overview covering definition, historical development, basic principles, and advantages

of thin-layer chromatography (TLC) in comparison to column chromatography. It concludes with a summary of common applications. Details on theory, instrumentation, plate technology, and method development are covered in subsequent articles.

## Definition and General Description

TLC is a form of liquid chromatography. It is used for rapid separation of samples into components based on differences in their retention behavior. The

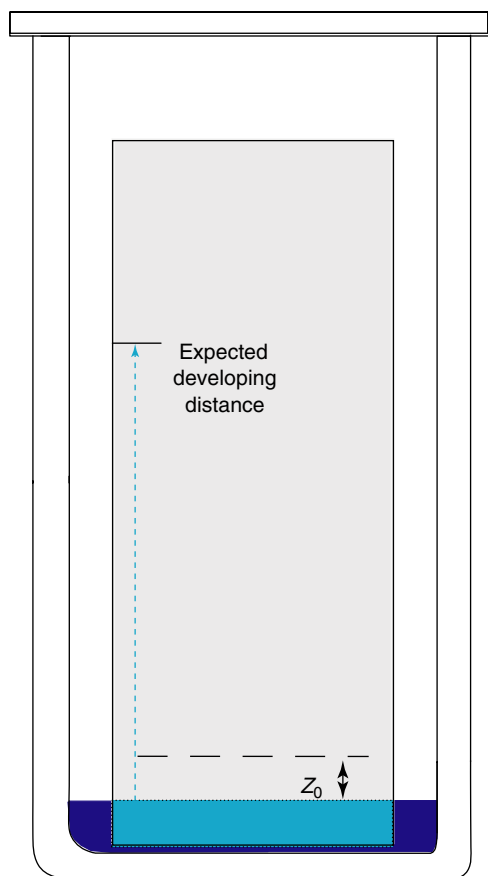
stationary phase, usually an adsorbent such as silica gel, is coated onto a rectangular support such as a glass, aluminum, or plastic plate as a thin layer.

Liquid samples are applied as spots or bands onto the dry stationary phase forming a line close to one edge of the plate. The TLC plate is placed vertically in a chromatographic chamber with lid, containing an amount of the liquid mobile phase sufficient to cover the bottom of the chamber, but not reaching up to the application position of samples on the plate (Figure 1). The mobile phase is drawn through the layer by capillary action starting separation when reaching the applied samples (start position). When the front of the mobile phase has moved to a certain predefined height (developing distance) the plate is removed from the chamber, appropriately dried, and, if necessary, derivatized for detection of samples. The resulting chromatogram is evaluated qualitatively by visual or densitometric comparison of the migration distance of the separated components to those of reference standards analyzed simultaneously on the same plate. Often, the behavior of substances during

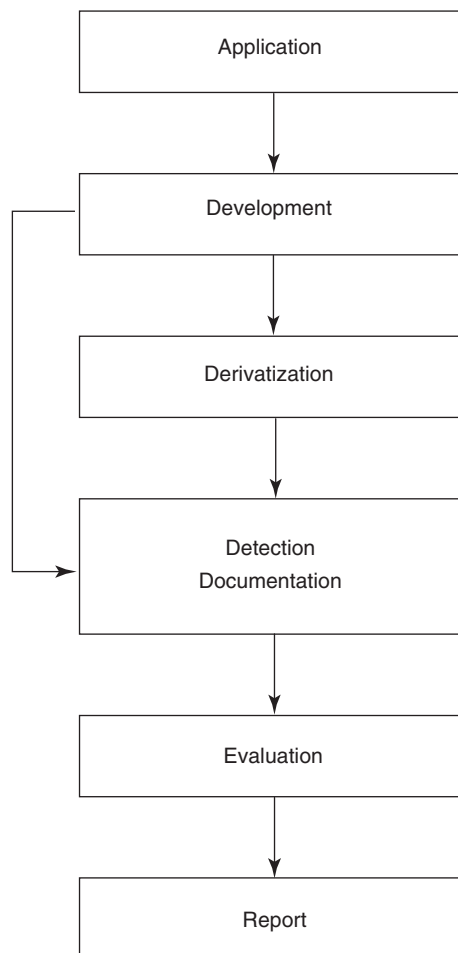
derivatization is also observed. Quantitative determination is possible by comparison of the intensity of the separated zones to those of known standard concentrations.

TLC is a very popular chromatographic technique because it is experimentally simple, enormously flexible, and inexpensive. It can lead rapidly to reliable qualitative and quantitative results primarily based on visual impression.

TLC is an offline technique (Figure 2). An analysis is performed as a sequence of individual steps, which are independent with respect to time and location. Compared to chromatography in columns, which is an online process, this design does not only offer great flexibility for technical solutions but also almost unlimited possibilities for combination of parameters to optimize the chromatographic result. In TLC, a large number of samples can be analyzed and compared to one another simultaneously on the same plate. The actual chromatographic separation provides exactly the same conditions for samples and references and there is no danger of permanently



**Figure 1** Schematic description of the development of a TLC plate.  $Z_0$  is the distance between the application position and the level of the developing solvent. After the mobile phase has reached the desired distance, the plate is removed from the tank.



**Figure 2** Steps of TLC as an offline technique.

contaminating the system by components of the sample's matrix because each plate is used only once. After separation, the TLC plate functions as a storage device for the chromatogram and thus enables multiple detection/evaluation. For all steps of the TLC process modern instruments are on the market, ensuring reproducible results at various levels of sophistication. As an open system the TLC plate is easily affected by environmental factors such as humidity, fumes, light, and mechanical stress during handling. To ensure reproducibility of results this must be taken into account when designing and performing a TLC experiment.

## Historical Development

At the end of the 1930s, adsorption chromatography in columns as introduced by Tswett had become a powerful separation technique for plant extracts and natural products. Simultaneously, the need for a more rapid alternative suitable for identification of separated substances led to the invention of an open chromatographic system. In 1938, Izmailov and Shraiber reported the separation of belladonna alkaloids on a thin adsorbent layer, coated onto microscopic slides. Development of circular chromatograms was achieved by placing small amounts of various solvents to the center of samples previously applied as spots onto the layer. This method was an extremely rapid microtechnique requiring only small amounts of stationary and mobile phases.

In 1944, Consden, Gordon, and Martin took a different approach using the principle of partitioning. This resulted in the invention of paper chromatography, which soon became a universal chromatographic technique. TLC did not advance much further until the 1950s when several significant improvements were made. Kirchner, Miller, and Keller in 1951 incorporated a fluorescent indicator into the stationary phase. Silica gel was becoming the most widely used adsorbent. In 1956, Stahl introduced the term 'thin-layer chromatography'. This, together with Merck developing standardized aluminum oxide, kieselguhr, and silica gel based on Stahl's specifications and Desaga bringing a basic instrument kit to the market, manifested the birth of TLC as a recognized and broadly accepted analytical technique. In 1962, Stahl with his fundamental book, *Thin Layer Chromatography – A Laboratory Handbook*, gave a forum to the leading researchers in the field of TLC and summarized the knowledge of that time. As a standard analytical technique, TLC was soon widely adopted by pharmacopoeias and official method collections. The improvement of

plate materials, culminating in precoated high-performance layers of reproducible quality in the late 1970s, the rapid development of the theoretical foundation, and the appearance of powerful densitometers became the basis for reliable quantitative analyses by TLC. At that time, high-performance liquid chromatography (HPLC) evolved and began to compete with TLC as a quantitative technique. The 1980s saw remarkable improvements particularly in instrumentation, automation of individual TLC steps, and theoretical concepts contributed by Snyder, Kaiser, Ebel, and Geiss. In 1987, Geiss published *Fundamentals of Thin Layer Chromatography*, which is regarded as one of the most influential books on the subject. The *Journal of Planar Chromatography* was founded in 1988 as a platform for discussion of all aspects of modern TLC.

Aside from the traditional way of performing TLC, several other approaches have also been taken. They include forced-flow techniques using centrifugal force, a pump (over-pressured layer chromatography, OPLC), or an electrical field to move the mobile phase. Hyphenated techniques were introduced combining TLC separation with infrared, Raman, or mass spectrometric detection. A technique for analysis of analytes that can be evaporated (IATRO-SCAN) utilizes separation performed on the surface of reusable thin chromatographic rods combined with a flame ionization detector. Also, gradient techniques have been developed for TLC, the most powerful of which is AMD – automatic multiple development – patented by Burger.

Today, TLC is a mature technique that meets all requirements of a modern analytical tool. Although it has lost its dominant role as a routine procedure to HPLC due to the lack of full automation, it is still indispensable for rapid qualitative analyses (identification), as a screening tool, and as a complementary technique in research and development.

## Basic Concepts

### Principles of Separation

The stationary phase in TLC is usually an adsorbent (Table 1) composed of very fine and highly porous particles coated as a thin layer onto a support. Today, the most widely used stationary phase is silica gel 60. With a solvent or solvent mixture as mobile phase, a liquid–solid chromatographic system is formed. Separation of sample components can be primarily described as adsorption chromatography. Sample molecules are retained due to specific interaction on the surface of the stationary phase while

**Table 1** Important inorganic and organic adsorbents

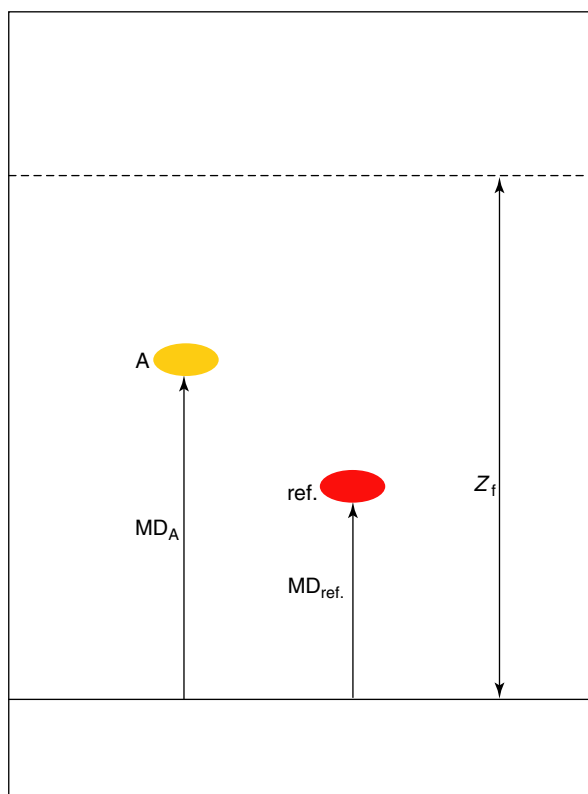
<i>Inorganic</i>	<i>Organic</i>	<i>Bonded phases on silica gel</i>
Silica gel	Cellulose	Octadecyl-, octyl-, dimethyl-, dipenylsilanized
Alumina	Polyamide	Aminopropylsilanized
Kieselguhr (diatomaceous earth)	Sephadex	Cyanopropylsilanized
Magnesium silicate		Diol (propanediol)

migration takes place when the molecules are dissolved in the mobile phase.

A dynamic adsorption equilibrium is established for each sample component. The equilibrium is characterized by a partition coefficient  $K_A$  representing the ratio of concentration of the sample component A in the stationary phase ( $c_s$ ) and in the mobile phase ( $c_m$ );  $K_A = c_s/c_m$ . Separation of two samples A and B can only be achieved if their partition coefficients are different. Typical forces causing adsorption of molecules on the stationary phase include nonspecific dispersion forces, dipole–dipole interactions, acid–base interactions, hydrogen bonding, electrostatic interaction between ions, and others. There are two extreme cases. If the sample is strongly adsorbed and only sparingly soluble in the mobile phase it will remain at the application position. Weakly adsorbed and readily soluble samples will migrate with or close to the solvent front.

A suitable combination of stationary and mobile phase will result in a separation of sample components. They will be located at different migration positions when the development of the TLC plate is interrupted as soon as the mobile phase has reached the specified developing distance. The ability of a solvent to displace an adsorbed sample molecule from the surface of a given stationary phase is called solvent strength. It can be measured as energy that is released when a solvent molecule is adsorbed. Although it is not simply a measure of dipole moment, solvent strength in adsorption chromatography is often expressed as solvent polarity. Adjusting the solvent strength of the mobile phase affects the position of the sample components in the chromatogram. The quality of separation, expressed as resolution between zones of the chromatogram, depends on the selectivity of the TLC system. Adjusting the composition of the mobile phase is the primary way of changing selectivity.

Apart from adsorption phenomena, a variety of other separation mechanisms can be utilized. If chemically bonded phases or liquid stationary phases on an inert support are used as stationary phase, separation is predominantly based on partition



**Figure 3** Schematic presentation of result in TLC. MD, migration distance;  $Z_f$ , distance between application and front;  $R_F = MD/Z_f$ .

equilibria. Examples include octadecyl, aminopropyl, diol, cyanopropyl modified silica gel, impregnated silica gel, and cellulose or kieselguhr. Separations based on ion exchange, complex formation, size exclusion, and even chiral recognition are also performed.

### Describing the Result

In column chromatography, the chromatographic system has a fixed length. Each sample component can be characterized by the time it requires to pass through the column and reach the detector. This is the retention time  $t_R$  and it is measured at the peak maximum. In TLC, the analysis takes a fixed time (development time) during which sample components can migrate. Depending on its retention each sample component will reach a specific migration distance (MD) and remains there as a zone (spot) for detection when the TLC plate is dried (**Figure 3**). A compound with high MD would in comparison have a short retention time in column chromatography, provided the separation mechanism is the same. The MD of a zone is measured in millimeters from the application position to the point of highest

concentration (intensity). When the chromatogram is evaluated densitometrically this point represents the peak maximum. As retention time is dependent on the mobile phase velocity, the migration distance is dependent on the position of the mobile phase front  $z_f$ . Therefore, a relative measure has been introduced. The  $R_F$  value (retardation factor) of a zone is the ratio of its migration distance to that of the mobile phase front.  $R_F = MD/z_f$ .

$R_F$  values are always  $< 1$ . For convenience it is common to multiply the  $R_F$  value by 100 and report  $bR_F$  with two digits. It is possible to compare the migration distance of an unknown (A) to that of a reference compound (ref) to yield the  $R_{rel}$  value.  $R_{rel} = MD_A/MD_{ref}$ .

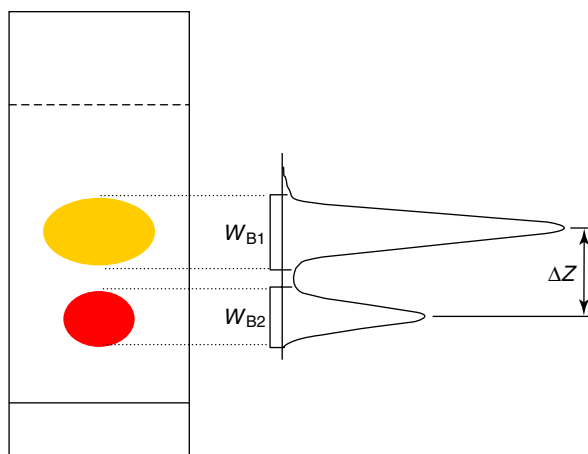
Although the  $R_F$  value is characteristic for a substance in a given TLC system, it must be treated with caution because it is affected by several parameters. In practice, it is often difficult to reproduce  $R_F$  values exactly. The  $R_{rel}$  value has no physical meaning.

The principle of qualitative analysis is a comparison of  $R_F$  values obtained from samples and standards on the same TLC plate. If two substances are the same they will have the same  $R_F$ . However, different substances may also migrate to the same position. For further confirmation of identity a specific derivatization or recording of the ultraviolet (UV) spectrum of the substances on the plate can be utilized. The size and intensity of the zone of the analyte is visually compared for estimation of quantity to that of standards at several known levels on the same plate. A precise quantitation is possible by scanning or video densitometry by which the absorption or fluorescence of separated zones of each chromatogram track is recorded. The resulting analog curves are integrated and evaluated based on peak height or area. The fundamental requirements for reliable quantitation are baseline resolved zones and symmetric peaks for the compounds in question. The resolution between two zones or peaks can be calculated from the chromatogram (Figure 4).

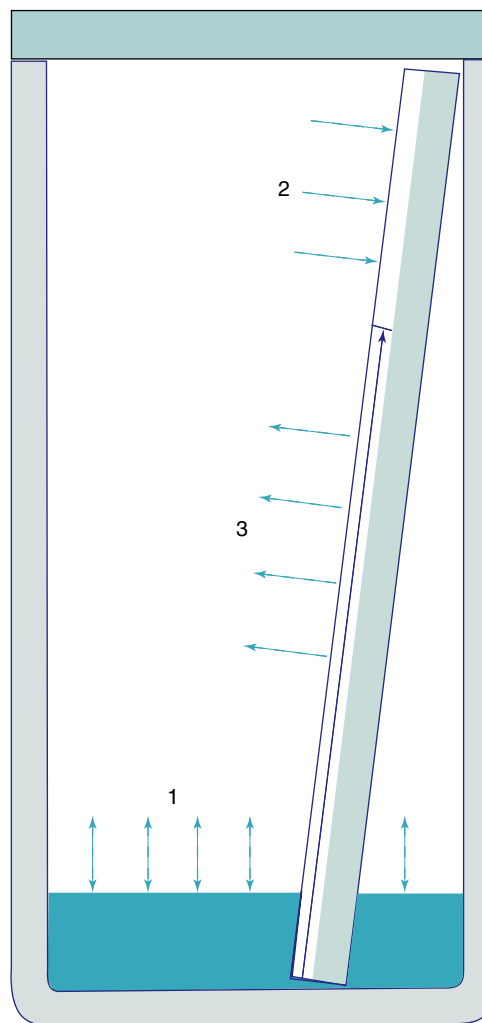
### Influence of the Developing Chamber

In TLC, the chromatographic system consists of a plate coated with the stationary phase and a developing chamber containing the mobile phase. As seen in Figure 5, a gas phase consisting of components of evaporated solvent molecules is also present. It is a matter of time, geometry, and whether fitting the chamber with filter paper enlarges the surface of the developing solvent, but as long as the chamber is tightly closed, saturation is eventually established.

This means that the vapor in the gas phase is in equilibrium (1) with the developing solvent on the



**Figure 4** The resolution ( $R_s$ ) between two separated peaks is calculated from the chromatogram:  $R_s = 2\Delta Z / W_{B1} + W_{B2}$ .



**Figure 5** The composition of the gas phase is affected by: (1) equilibrium of gas phase with developing solvent on the bottom of the chamber (saturation); (2) and (3) equilibrium between gas phase and stationary phase ((2) absorption of gaseous molecules, (3) evaporation of mobile phase).



bottom of the chamber. When the dry TLC plate is placed in the chamber solvent vapor is readily adsorbed onto the stationary phase. If the stationary phase has no contact with the developing solvent, as it is possible in a twin trough chamber, over time another equilibrium, (2) between the gas phase and the stationary phase is established. This is called preconditioning. It leads to adsorptive saturation of the stationary phase. When the mobile phase rises up the stationary phase a third equilibrium (3) is approached. While (1) is dependent on the vapor pressure of the solvent components and (2) is controlled by adsorptive forces, (3) is affected by both parameters.

During chromatography all three equilibria are effectively influencing each other as well as the chromatographic result. Experimental details, type, and geometry of the chamber are therefore important parameters that can be used to optimize separation. Even in so-called saturated chambers chromatography is typically performed in nonequilibrium, which makes it difficult to mathematically describe all processes in exact detail. Generally, there are neither good nor bad chambers; however, results obtained in one may be quite different from those obtained in another chamber. In consequence, it is imperative for obtaining reproducible results that the chamber and all of its parameters are clearly defined and kept constant. As a rule of thumb saturated chambers tend to give lower  $R_F$  values for the same separation than unsaturated chambers. The zones in saturated chambers are somewhat more diffuse, but reproducibility of the result is much higher than in unsaturated chambers. TLC can also be performed in sandwich chambers, where covering the stationary phase with a counterplate limits the gas phase. As a result of this the stationary phase remains active and cannot be preloaded with solvent molecules. The mobile phase is then separated into its components. So-called secondary fronts are formed and can interfere with the separation. In saturated chambers secondary fronts are usually eliminated because the polar components of the mobile phase are preferably preadsorbed through the gas phase. Secondary fronts are one of the principal drawbacks of OPLC where the gas phase is completely eliminated.

## Practical Approaches

### Sample Application

During sample application two principal requirements must be met. The samples must be precisely and correctly positioned in order to identify separated components based on their migration distance/ $R_F$

**Table 2** Typical parameters for sample application on HPTLC plates

Parameter	Band	Spot
Distance from lower edge of plate in mm	8	10
Minimum distance from left and right edge of plate in mm	10	10
Minimum space in mm between bands/spots <sup>a</sup> in mm	2	2
Band length in mm	8	
Maximum diameter of application spot in mm		4

<sup>a</sup>Most instruments do not allow programming of distance between bands/spots. Therefore distance between tracks (center to center) and band length must be chosen in order to meet the minimum distance requirement. For spot application, volume and application speed have to be determined empirically.

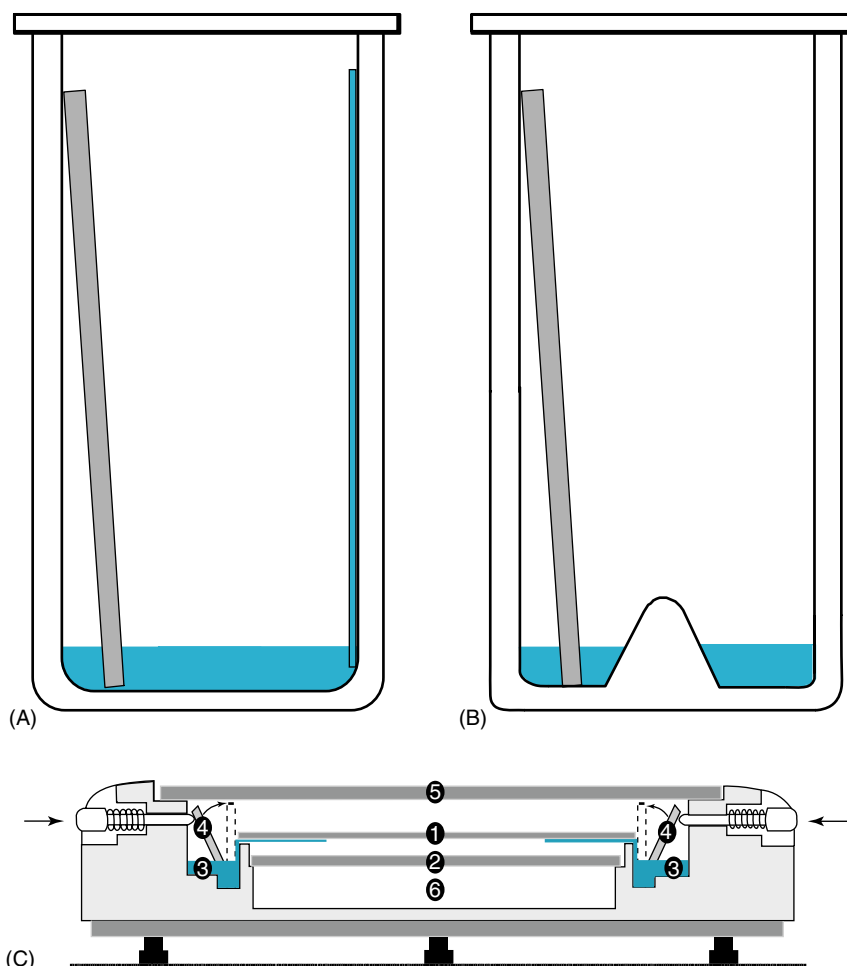
values and the sample volume must be controlled in order to allow quantitative evaluation of the chromatogram. The size of the plate determines the number of samples. Samples should not be applied too close to the edges of the plate and must be evenly and sufficiently spaced for best results. Typical application parameters for high-performance thin-layer chromatography (HPTLC) plates are summarized in **Table 2**. Application of samples can be performed manually or with the help of instruments. For manual application the positions are carefully marked on the TLC plate with soft pencil. Use of a spotting guide helps avoiding any damage to the layer. Samples are transferred with the help of disposable capillaries, graduated micropipettes, or microliter syringes and deposited onto the plate as spots. If larger volumes are applied this is done in portions with intermediate drying of the application position. The smaller the applied spot, the better is the obtainable resolution of the chromatogram. Best results in sample application are obtained with instruments, particularly those that use the spray-on technique. For maximum resolution and best precision in quantitative TLC, samples are sprayed on as narrow bands. Following sample application it is common to mark the desired developing distance on one edge of the plate.

### Chromatogram Development

Chromatogram development is performed in a chromatographic chamber, often called a tank. There are many different kinds of chambers, but flat bottom or twin trough chambers with a rectangular base are most common. The size of the chamber should be appropriate for the size of the plate. A saturated chamber is created, when one or more walls of the chamber are fitted with filter paper, which is thoroughly wetted with the developing solvent. For saturation to establish, the charged chamber is left

closed for a defined period of time prior to introducing the TLC plate and starting development. Unsaturated chambers do not contain filter paper. Chromatography is started immediately after the chamber is charged with developing solvent. The level of solvent in the chamber is chosen in a way that the sample application position on the plate is clearly above it. Most plates are developed in a saturated chamber in vertical position, but horizontal developing chambers are also available (**Figure 6**). With special chambers circular and anticircular development can also be performed. For circular development samples are applied to a square plate around the center forming a circle. The mobile phase is applied to the center of the plate. Anticircular development requires delivery of mobile phase from a ring gap to the plate onto which samples have been applied forming a circle of large diameter.

All developments discussed so far were single developments. Multiple developments are also possible. After intermediate drying of the plate a subsequent development can be performed (1) over the same, (2) over a longer, and (3) over a shorter developing distance. While in case (1) the goal is to improve separation by using the same mobile phase for development, in case (2) the first development over a few millimeters with a strong mobile phase is typically used to concentrate the sample into a narrow application zone. The second development with a suitable mobile phase is used for separation. In case (3), a nonpolar matrix can be removed from the sample and transported to the upper edge of the plate with a weak mobile phase leaving the compounds of interest at the application position. In the second development with a stronger mobile phase the actual separation of the sample is performed.



**Figure 6** Comparison of flat-bottom chamber (A), twin trough chamber (B), and horizontal developing chamber (C): (1) HPTLC plate, (2) glass plate for sandwich configuration, (3) reservoir for developing solvent, (4) glass strip, (5) cover plate, (6) conditioning tray. Examples: The flat-bottom chamber is lined with filter paper soaked with the developing solvent for saturation. Twin trough chambers can be filled with two different liquids and are less solvent-consuming. Horizontal developing chambers are easy to use in sandwich configuration.

When the mobile phase has reached the desired developing distance of typically 6 cm on HPTLC plates and 12 cm on TLC plates development is interrupted, the plate is removed from the chamber, and dried.

### Visualization and Derivatization

One of the most striking features of TLC is the possibility of presenting the chromatographic result visually. Visual detection is possible when the separated sample components are colored, have native fluorescence, or absorb UV light so that they quench the fluorescence of the indicator, which can be built into the layer of the TLC plate. Derivatization can enhance visualization dramatically by improving the detectability of all or selected zones of the chromatogram. Numerous chemical reactions are known to convert the analyte into a colored or fluorescing derivative. Nonspecific reactions such as charring with acid are broadly applicable, whereas specific derivatization such as the reaction of amines with ninhydrin can be used to single out compounds of certain functionality from a separated mixture. The convenient availability of biochemical reactions and *in situ* biological tests add extra flexibility to the detection step in TLC. Examples include inhibition of an enzyme such as choline esterase or measurement of effects on the growth of bacteria or yeast cells.

For chemical derivatization, the developed TLC plate is either sprayed with or immersed into a solution of the reagent. While spraying requires small amounts of reagent and provides great flexibility the advantage of immersion lies in ensuring a homogeneous and reproducible coverage. In many cases derivatization is completed with a heating step using either an oven or a plate heater.

### Densitometric Evaluation

If a separated substance absorbs UV or white light or can be excited to fluoresce, densitometric evaluation of the chromatogram is possible without derivatization. With a densitometer, also called scanner, the tracks of the TLC plate are evaluated by registering the reflected/emitted light when a narrow beam of light of a single wavelength is moved across the plate. A photomultiplier converts the light into an electrical signal, which as a function of migration distance presents the analog curve of the chromatogram. Following integration of the raw data the peak height and/or area can be quantitated. Densitometry is a very sensitive measurement. Typically less than 100 ng of substance per zone can be detected. In some cases, such as aflatoxins, detection limits are in the picogram range. Most modern densitometers allow recording of *in situ* UV spectra, which can be

used to confirm identity or purity of separated compounds.

Densitometry is performed after derivatization, if that improves the detectability of the analyte. Examples include introduction of fluorophoric groups, charring of nonabsorbing compounds, specific color reactions.

### Documentation

Documentation of a TLC chromatogram as an image is conveniently possible. Traditionally photographs have been taken but the availability of electronic devices such as flatbed scanners or digital cameras have simplified the process. The principal advantage of electronic images is their durability. Added benefit is the possibility of qualitative and quantitative densitometric evaluation.

## Special and Hyphenated Techniques

### High-Performance Thin-Layer Chromatography

HPTLC is based on the use of special layers made from narrowly distributed fine particles of  $\sim 5\text{ }\mu\text{m}$ . It is actually not a special technique because all fundamental parameters, theoretical considerations, and practical aspects of classical TLC still apply. However, due to the consequent use of specialized instrumentation HPTLC does not only achieve miniaturization of the chromatogram, but also and more importantly a significant improvement of sensitivity, reproducibility, and separation power. A modern HPTLC workstation including devices for each individual chromatographic step is software controlled and provides good manufacturing or good laboratory practice compliant results similar to that of HPLC and GC instruments.

### Automated Multiple Development

Automated multiple development (AMD) is a gradient technique consisting of a variable number of developments with intermediate drying steps, which are achieved by application of vacuum to the special developing chamber. In a computer-controlled method, each subsequent development proceeds to a higher developing distance using increments of 2–3 mm than the previous step. Furthermore, each run uses a weaker (in adsorption chromatography less polar) mobile phase. As a consequence, the separated zones are repeatedly focused. In combination with the mobile phase gradient this results in a significant increase of the separation power. Results obtained with AMD are predictable and very reproducible.

## Forced-Flow Techniques

The separation efficiency of TLC is limited due to the fact that capillary forces move the mobile phase. Forced-flow techniques are an attempt to solve this problem. One approach, rotational planar chromatography, utilizes centrifugal forces and has become a widely used preparative technique. A circular plate is mounted on a centrifuge and the sample followed by the mobile phase is applied close to the center of the plate. Separated sample components can be collected when they elute from the rotating plate.

Another very flexible forced-flow technique is OPLC introduced by Tyihak and Mincsovcics. The plate is completely sealed on all four edges. Following sample application the plate is covered with a Teflon foil, which is part of a cassette. In a special pressure chamber the cover foil seals the plate in vertical direction under external pressure. With the help of a programmable pump the mobile phase is forced with constant velocity through an inlet in the coversheet of the cassette to travel in a laminar flow through the chromatographic layer toward an outlet located at the opposite edge of the plate. The two principal drawbacks of the technique, which can considerably interfere with the separation and must therefore be minimized in the experimental setup, are solvent demixing resulting in secondary fronts when multicomponent mobile phases are employed and a so-called front of total wetness. This is a disturbance at the region where gas bubbles, which are trapped in the pores of the adsorbent when the mobile phase first enters the layer, are finally displaced by solvent molecules. OPLC has found various analytical, semi-preparative, and preparative applications. The system can be operated in an online mode like HPLC, offline mode like regular TLC, or in a mixed mode.

## Hyphenated Techniques

Several so-called hyphenated techniques have been developed, where the developed TLC plate is transferred to a modified spectrometer to record *in situ* the Fourier transform infrared, surface enhanced Raman, or mass spectra of the separated zones. This way more detailed structural information can be obtained to complement the data from densitometric evaluation. A true hyphenation is the direct application of the eluate from a microbore HPLC column onto an HPTLC plate, which is then developed by AMD.

## Applications

Typical applications of TLC are listed in Table 3. In many areas including herbal, pharmaceutical, and

**Table 3** Major applications of TLC

Sample type	Application
Pharmaceutical	Impurities in synthetic drugs
	Stability tests of synthetic drugs
	Content uniformity
	Pharmacokinetic studies and drug monitoring
	Assay
Industrial	Enantiomeric purity
	Product uniformity
	Impurity determination
	Surfactants, dyes
Herbal	Cleaning validation
	Identity, purity, and stability tests of plant drugs
	Quantitation of marker compounds
Biomedical	Organic acids, lipids
	Carbohydrates
	Porphyryns and bile pigments
	Amino acids, peptides
	Steroids, doping
Nutritional	Mycotoxins (incl. aflatoxins)
	Drug and pesticide residues
	Antioxidants, preservatives
	Pigments, dyes, spices, flavors, vitamins
Environmental	Pesticides residues
	Water and soil analysis
Forensic	Poisons
	Drugs of abuse
	Inks

environmental analyses the flexibility of the technique as a research tool is a great asset. Other applications such as content uniformity, residue, and doping test as well as cleaning validation take advantage of short analysis times due to parallel chromatography of many samples. Qualitatively, TLC is primarily used in general research and in quality control of raw materials for rapid and cost-efficient identification. Quantitative applications include assays, determination of impurities and process monitoring.

See also: **Thin-Layer Chromatography:** Principles; Plate Technology; Method Development; Instrumentation.

## Further Reading

- Frey HP and Zieloff K (1993) *Qualitative und quantitative Dünnschicht-Chromatographie*. Weinheim: VCH Verlagsgesellschaft mbH.
- Geiss F (1987) *Fundamentals of Thin Layer Chromatography*. Heidelberg: Alfred Hüthig Verlag.
- Hahn-Deinstrop E (1998) *Applied Thin-Layer Chromatography*. Weinheim: Wiley-VCH.
- Pachaly P (1991) *DC-Atlas*. Stuttgart: Wissenschaftliche Verlagsgesellschaft.
- Poole CF and Poole SK (2003) *The Essence of Chromatography*. Amsterdam: Elsevier.

Sherma J and Fried B (eds.) (1994) *Thin-Layer Chromatography, Techniques and Applications*, 3rd edn. New York: Dekker.

Sherma J and Fried B (eds.) (2003) *Handbook of Thin-Layer Chromatography*, 3rd edn. New York: Dekker.

Stahl E (1969) *Thin-Layer Chromatography, A Laboratory Handbook*, 2nd edn. Berlin: Springer.

Touchstone JC (1992) *Practice of Thin-Layer Chromatography*. New York: Wiley.

Wagner H and Bladt S (1995) *Plant Drug Analysis, A Thin Layer Chromatography Atlas*, 2nd edn. Berlin: Springer.

## Principles

**C F Poole**, Wayne State University, Detroit, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Thin-layer chromatography (TLC) as we know it today became established in the 1950s with the introduction of standardized materials and procedures that led to improved separation performance, reproducibility, and a rapid growth in applications. During the 1970s, fine-particle layers and associated instrumentation required for their correct use were introduced. Methods that employed advanced instrumentation for separations became known as high-performance thin-layer chromatography (HPTLC), instrumental TLC, or modern TLC to distinguish it from its parent, now generally referred to as conventional TLC, or simply TLC. Both approaches coexist today because of their complementary

features (Table 1). Conventional TLC provides a quick, inexpensive, and portable method for qualitative analysis. It requires minimal and readily available instrumentation and uses easily learned laboratory skills. HPTLC, on the other hand, uses fine-particle layers for faster and more efficient separations and requires sophisticated instruments for automated sample application, development, and detection. HPTLC affords accurate and precise results based on *in situ* measurements and a record of the separation in the form of a chromatogram (see Figure 1). While there are no barriers to the use of conventional TLC in any analytical laboratory, only those laboratories that have invested in the necessary instrumentation have the option of using HPTLC methods.

TLC is a type of liquid chromatography in which the stationary phase is in the form of a uniform thin layer immobilized on a flat surface rather than packed into a tube (column). In the basic experiment, the sample is applied to the layer as a spot or band near to the bottom edge of the layer. Separations are performed in a closed chamber by contacting the bottom edge of the layer with the mobile phase, in a process referred to as development. In traditional methods, the mobile phase advances through the layer by capillary forces. Alternatively, in forced flow and electrochromatography the mobile phase is driven through the layer by a pressure gradient, centrifugal force, or electroosmosis. In all cases, separation results from the different rates of migration of the sample components in the direction traveled by the mobile phase. After development and evaporation of the mobile phase, the sample components are immobilized in the layer and their position and quantity determined by visual evaluation or densitometry.

**Table 1** Characteristic properties of conventional (TLC) and high-performance (HPTLC) thin-layer chromatographic separations

Parameter	TLC	HPTLC
Plate dimensions (cm)	20 × 20	10 × 10 or 10 × 20
Layer thickness (mm)	0.1–0.25	0.1 or 0.2
Nominal particle size (μm)	20	5
Sample volume (μl)	1–5	0.1–0.2
Starting spot diameter (mm)	3–6	1–2
Diameter of separated spots (mm)	6–15	2–6
Solvent front migration distance (cm)	10–15	3–6
Time for development (capillary flow) (min)	20–200	3–20
Favorable detection limits (densitometry)		
Absorption (ng)	1–5	0.1–0.5
Fluorescence (pg)	50–100	5–10
Typical plate number	<600	<5000

### Flow through Porous Layers

TLC is a simple experimental method based on a complex separation mechanism, more complex, in



Sherma J and Fried B (eds.) (1994) *Thin-Layer Chromatography, Techniques and Applications*, 3rd edn. New York: Dekker.

Sherma J and Fried B (eds.) (2003) *Handbook of Thin-Layer Chromatography*, 3rd edn. New York: Dekker.

Stahl E (1969) *Thin-Layer Chromatography, A Laboratory Handbook*, 2nd edn. Berlin: Springer.

Touchstone JC (1992) *Practice of Thin-Layer Chromatography*. New York: Wiley.

Wagner H and Bladt S (1995) *Plant Drug Analysis, A Thin Layer Chromatography Atlas*, 2nd edn. Berlin: Springer.

## Principles

**C F Poole**, Wayne State University, Detroit, MI, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Thin-layer chromatography (TLC) as we know it today became established in the 1950s with the introduction of standardized materials and procedures that led to improved separation performance, reproducibility, and a rapid growth in applications. During the 1970s, fine-particle layers and associated instrumentation required for their correct use were introduced. Methods that employed advanced instrumentation for separations became known as high-performance thin-layer chromatography (HPTLC), instrumental TLC, or modern TLC to distinguish it from its parent, now generally referred to as conventional TLC, or simply TLC. Both approaches coexist today because of their complementary

features (Table 1). Conventional TLC provides a quick, inexpensive, and portable method for qualitative analysis. It requires minimal and readily available instrumentation and uses easily learned laboratory skills. HPTLC, on the other hand, uses fine-particle layers for faster and more efficient separations and requires sophisticated instruments for automated sample application, development, and detection. HPTLC affords accurate and precise results based on *in situ* measurements and a record of the separation in the form of a chromatogram (see Figure 1). While there are no barriers to the use of conventional TLC in any analytical laboratory, only those laboratories that have invested in the necessary instrumentation have the option of using HPTLC methods.

TLC is a type of liquid chromatography in which the stationary phase is in the form of a uniform thin layer immobilized on a flat surface rather than packed into a tube (column). In the basic experiment, the sample is applied to the layer as a spot or band near to the bottom edge of the layer. Separations are performed in a closed chamber by contacting the bottom edge of the layer with the mobile phase, in a process referred to as development. In traditional methods, the mobile phase advances through the layer by capillary forces. Alternatively, in forced flow and electrochromatography the mobile phase is driven through the layer by a pressure gradient, centrifugal force, or electroosmosis. In all cases, separation results from the different rates of migration of the sample components in the direction traveled by the mobile phase. After development and evaporation of the mobile phase, the sample components are immobilized in the layer and their position and quantity determined by visual evaluation or densitometry.

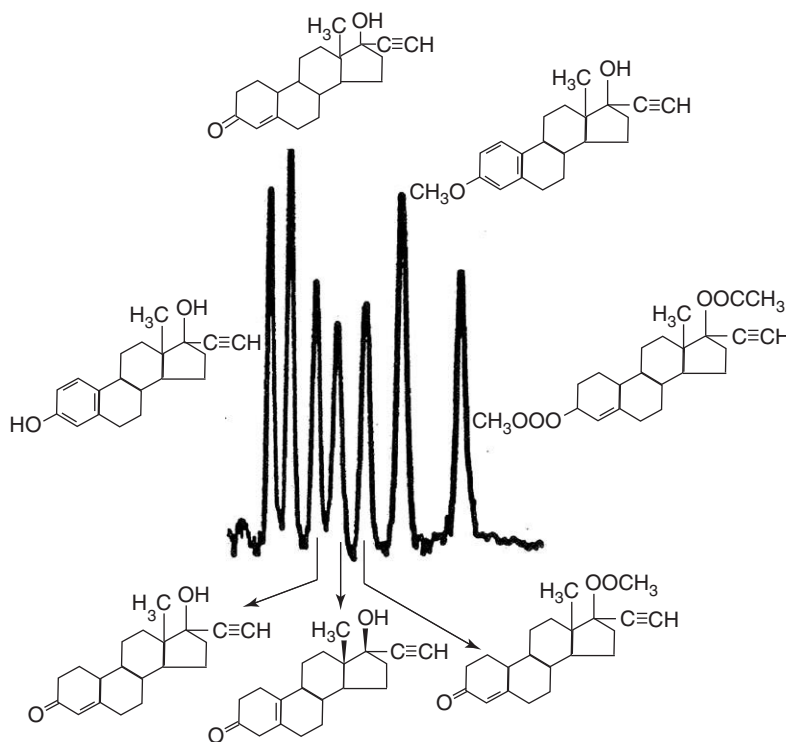
### Flow through Porous Layers

TLC is a simple experimental method based on a complex separation mechanism, more complex, in

**Table 1** Characteristic properties of conventional (TLC) and high-performance (HPTLC) thin-layer chromatographic separations

Parameter	TLC	HPTLC
Plate dimensions (cm)	20 × 20	10 × 10 or 10 × 20
Layer thickness (mm)	0.1–0.25	0.1 or 0.2
Nominal particle size (μm)	20	5
Sample volume (μl)	1–5	0.1–0.2
Starting spot diameter (mm)	3–6	1–2
Diameter of separated spots (mm)	6–15	2–6
Solvent front migration distance (cm)	10–15	3–6
Time for development (capillary flow) (min)	20–200	3–20
Favorable detection limits (densitometry)		
Absorption (ng)	1–5	0.1–0.5
Fluorescence (pg)	50–100	5–10
Typical plate number	<600	<5000





**Figure 1** Separation of ethynyl steroids by HPTLC. (Reprinted with permission from Poole CF (1999) Planar chromatography at the turn of the century. *Journal of Chromatography A* 856: 399–427; © Elsevier.)

fact, than column liquid chromatography. The mobile phase velocity varies with migration distance and is subject to local fluctuations due to the presence of a vapor phase and from solvent demixing. In addition, in TLC the mobile phase normally penetrates a dry layer. The flow of solvent at the developing front is generally unsaturated and the speed with which the front moves depends on the experimental conditions. Capillary forces are stronger in the narrow interparticle channels, leading to a more rapid advance of the mobile phase. Larger pores below the solvent front are filled at a slower rate resulting in an increased thickness of the mobile phase layer. If the vapor phase and mobile phase are not in equilibrium, evaporation will cause a loss of mobile phase from the layer surface and a decrease in the solvent front velocity. On the other hand, the dry layer ahead of the solvent front progressively adsorbs vapor, filling some of the pores and interparticle channels, and increasing the apparent velocity with which the solvent front migrates. Various kinds of sandwich chambers, which either eliminate or minimize contact of the plate surface with the vapor phase, offer reasonable control of the mobile phase velocity.

Separations performed with mixed solvents of different solvent strength may form multiple solvent

fronts. As the mobile phase moves through the layer, it becomes depleted in the solvent with the highest affinity for the stationary phase. Eventually, a secondary front is formed that separates the bulk mobile phase from mobile phase now totally depleted in the solvent selectively adsorbed by the stationary phase. For a mobile phase consisting of  $n$  solvents, as many as  $n + 1$  solvent fronts are possible. This process is referred to as solvent demixing and results in sample components moving in regions of different solvent strength and selectivity. Under these conditions, retention is no longer easily related to the bulk mobile phase composition. Sample zones close to a secondary front are often focused into narrow bands with a different shape to neighboring zones, which allows demixing conditions to be identified. Solvent demixing occurs more commonly in developing chambers with an unsaturated atmosphere.

### Capillary Flow

The common methods of development in TLC employ capillary forces to transport the mobile phase through the layer. These weak forces arise from the decrease in free energy of the solvent as it enters the porous structure of the layer. The velocity at which

the solvent front moves through the layer is a function of the distance of the front from the solvent entry position and declines as this distance increases. There are two consequences of this effect: (1) the mobile phase velocity varies as a function of time (and migration distance) and (2) the mobile phase velocity is established by the system variables and is otherwise beyond experimental control. In the absence of a significant exchange of solvent flux with the vapor phase the position of the solvent front with respect to time is adequately described by eqn [1]

$$(Z_f)^2 = \kappa t \quad [1]$$

and after differentiation, the velocity of the solvent front by

$$u_f = \kappa / 2Z_f \quad [2]$$

where  $Z_f$  is the distance of the solvent front position above the solvent level in the developing chamber,  $\kappa$  the velocity constant ( $\text{cm}^2 \text{s}^{-1}$ ),  $t$  the time from contacting the layer with the solvent, and  $u_f$  the solvent front velocity. In the absence of equilibrium, rather complex correction factors must be applied to eqn [1]. For a given set of experimental conditions the velocity constant is related to the characteristic properties of the mobile and stationary phases by eqn [3]

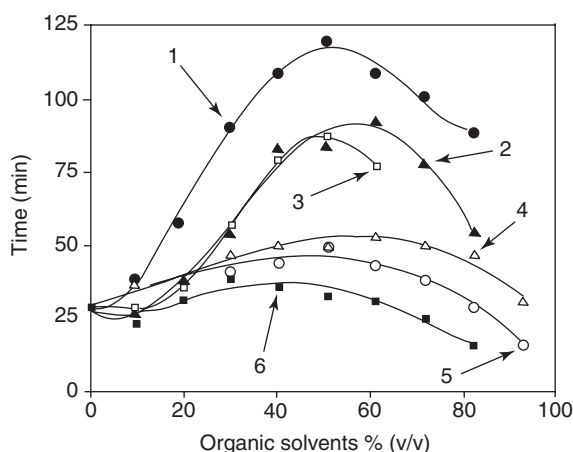
$$\kappa = 2K_0 d_p (\gamma / \eta) \cos \theta \quad [3]$$

where  $K_0$  is the permeability constant of the layer,  $d_p$  the average particle size,  $\gamma$  the surface tension,  $\eta$  the viscosity of the mobile phase, and  $\theta$  the contact angle between the mobile phase and the layer. The permeability constant is dimensionless and takes into account the effect of porosity on the permeability of the layer and the bulk liquid velocity to the solvent front velocity. The permeability constant varies little for well-prepared layers and has typical values of  $1\text{--}2 \times 10^{-3}$ . Higher velocities are achieved by using layers of a relatively large particle size with mobile phases that fully wet the layer and have a favorable surface tension to viscosity ratio. Increasing the average particle size of the layer increases the mobile phase velocity at the expense of broader chromatographic zones requiring that a compromise is struck between the two effects. In addition, the velocity constant varies linearly with the ratio of the surface tension to viscosity of the mobile phase. Consequently, solvents that maximize this ratio (and not just optimize one of the parameters) are required for fast and efficient separations. Virtually all solvents fully (or very nearly so) wet silica gel and

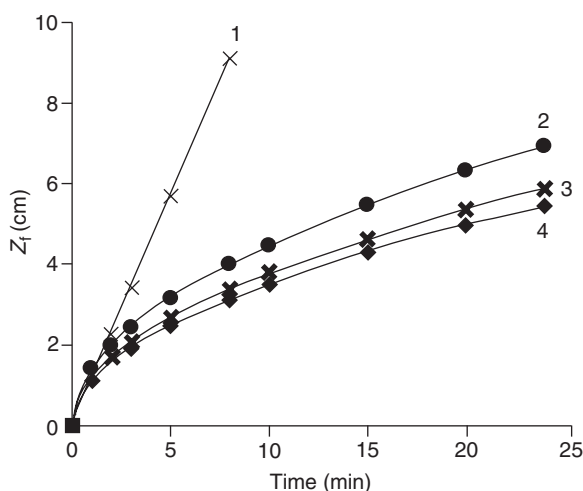
polar chemically bonded layers with a contact angle close to zero ( $\cos \theta = 1$ ). This is not the case for aqueous mobile phases on chemically bonded layers prepared from reagents with long-chain alkyl groups. To make these layers completely compatible with aqueous phases for reversed-phase separations a sorbent of larger particle size with a lower bonding density of alkyl groups is used. Separation times can be relatively long depending on the composition of the mobile phase (Figure 2).

There are two features of importance when using eqns [1] and [2] to represent the flow characteristics of the mobile phase through the layer. The velocity constant depends on the identity of the mobile phase, the characteristic properties of the layer, the layer thickness, and the state of equilibrium between solvent vapors in contact with the layer and the bulk solvent moving through the layer. The velocity constant is, in fact, a system constant, only invariant for a given set of experimental conditions. The bulk mobile phase velocity representing saturated flow through the region occupied by the sample zones is moving at a lower velocity than the solvent front velocity. A value of  $0.8u_f$  is taken as a reasonable estimate of the bulk mobile phase velocity for calculation of the effect of mobile phase velocity on zone broadening.

The system conditions that affect the mobile phase velocity are outlined above, but this does not answer the question of whether capillary forces are strong enough to enable mobile phase velocities close to optimum under typical conditions. The optimum



**Figure 2** Plot of the time required for the solvent front to migrate 5 cm on a Merck HPTLC RP-18 WF254s layer with aqueous mobile phases of different composition. Organic solvents: (1) 2-propanol; (2) *N,N*-dimethylformamide; (3) 2,2,2-trifluoroethanol; (4) methanol; (5) acetone; and (6) acetonitrile. (Reproduced with permission from Kiridena W and Poole CF (1999) *Journal of Planar Chromatography* 12: 13–25.)



**Figure 3** Relationship between the solvent-front migration distance for dichloromethane on an HPTLC silica gel layer as a function of time for different experimental conditions. Identification: (1) forced flow development at  $u_{\text{opt}}$ ; (2) capillary flow in a saturated chamber; (3) capillary flow in a saturated chamber with a covered layer (sandwich chamber); and (4) capillary flow in an unsaturated chamber.

mobile phase velocity for a separation can be established by forced flow development (see next section). This velocity is considerably higher than the mobile phase velocity obtained by use of capillary flow under different experimental conditions (Figure 3). This figure illustrates the three cardinal deficiencies of capillary flow systems: (1) capillary forces are inadequate to achieve the desired optimum mobile phase velocity; (2) the mobile phase velocity declines as the solvent front migration distance increases and is increasingly less favorable for maintaining acceptable chromatographic efficiency as the solvent front moves higher up the layer; and (3) the mobile phase velocity is a complex function of the system conditions and can only be reproduced by careful attention to experimental details.

### Forced Flow

Forced flow separations overcome the principal deficiencies of capillary flow separations by establishing a constant and optimum mobile phase velocity. Forced flow separations require specially designed developing chambers exploiting either centrifugal or pneumatic forces to drive the mobile phase through the layer. Centrifugal methods are more popular for preparative-scale separations and have been little used for analysis. The preferred approach for analytical separations is to seal the open face of the layer by contact with a flexible membrane, under hydraulic pressure, and deliver the mobile phase to

the layer with a constant-volume, reciprocating-piston pump. For as long as the difference between the hydraulic seal and the mobile phase inlet pressure is sufficient, the position of the solvent front at any time after the start of development is described by  $Z_f = u_f t$ . The mobile phase velocity ( $u_f$ ) no longer depends on the contact angle and there are no restrictions on solvent selection for reversed-phase layers, unlike capillary flow systems. A disadvantage of forced flow is the possible appearance of several fronts during the development due to the difficulty of displacing air from the dry layer by the mobile phase and due to solvent demixing. Sample zones moving in the disturbing zone, a region of the mobile phase containing microbubbles of air, or passed over by it, are often distorted and difficult to record by densitometry. Various operating procedures are available to eliminate or minimize problems caused by the disturbing zone. For example, an initial development with a weak mobile phase to displace trapped air from the layer before starting the separation or by using a backpressure regulator to increase the solubility of air in the mobile phase.

### Electroosmotic Flow

Electroosmotic flow has emerged as a viable alternative transport mechanism to pressure-driven flow in column chromatography. Benefits include a plug-flow profile (reduced transaxial contributions to zone broadening) and a mobile phase velocity that is independent of the column length and particle size. The electroosmotic-driven flow is governed by the dielectric constant of the mobile phase, the zeta potential at the stationary phase/mobile phase interface, and the applied electric field. The efficiency obtainable is limited by double layer overlap or radial dispersion induced by inefficient heat dissipation.

The current status of electroosmotic-driven flow in TLC is probably more confusing than reassuring, although recent studies have brought some enlightenment to this approach. Early studies of electroosmotic flow in vertically mounted layers using solvents of low polarity as the mobile phase are now believed to be the result of thermal effects. Enhanced flow results from forced evaporation of the mobile phase from a solvent-deficient region at the top of the layer. Because of drainage in vertically mounted layers, electrical resistance is highest at the top of the layer, and the increase in heat production drives the evaporation of solvent, pulling additional solvent through the layer. True electroosmotic flow was demonstrated in horizontally mounted layers at modest field strengths ( $<1 \text{ kV cm}^{-1}$ ) with mobile

phases of suitable dielectric constant. Movement of the mobile phase toward the layer surface competes with electroosmotic flow along the layer. The voltage and buffer concentration must be optimized to minimize either excessive flooding or drying of the layer to avoid degradation of the separation quality. Many questions remain unanswered at this time. For example, which solvents to use, the necessity to impregnate the layer with solvent containing ions as current carriers, the effect of local heating on zone profiles, the effect of binder chemistry on flow characteristics, and mass transfer properties. Comprehensive studies of zone broadening are conspicuous by their absence and the future of electroosmotic-driven flow in TLC remains a matter for conjecture.

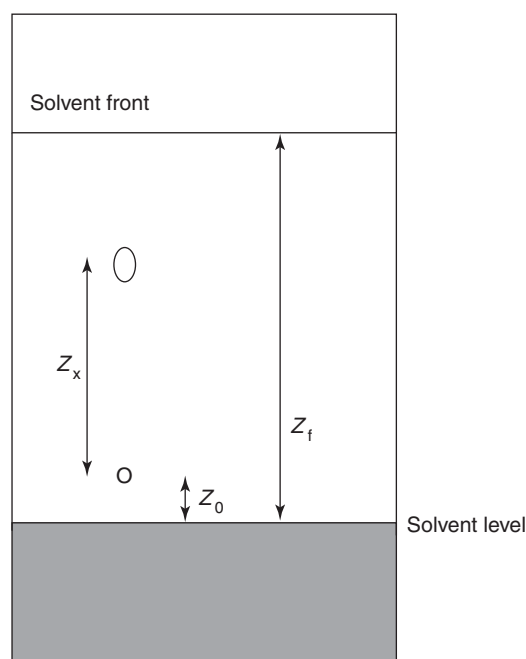
## Retardation Factor

The fundamental parameter used to characterize the position of a sample zone in a thin-layer chromatogram is the retardation factor, or  $R_F$  value. It represents the ratio of the distance migrated by the sample compared to the distance traveled by the solvent front, and for linear development is given by

$$R_F = Z_X / (Z_f - Z_o) \quad [4]$$

where  $Z_X$  is the distance traveled by the sample from its origin,  $(Z_f - Z_o)$  the distance traveled by the mobile phase from the sample origin,  $Z_f$  the distance traveled by the solvent front measured from the mobile phase entry position (the solvent level at the start of the separation), and  $Z_o$  the distance from the sample origin to the position used as the origin for the mobile phase (see Figure 4). The boundary conditions are  $1 \geq R_F \geq 0$ . When  $R_F = 0$ , the sample remains at the origin, and for  $R_F = 1$ , the sample migrates with the solvent front and does not interact with the stationary phase.  $R_F$  values are calculated to two decimal places or expressed as whole numbers, as  $hR_F$  values, equivalent to  $100R_F$ .

Systematic errors in determining  $R_F$  values result from the difficulty in locating the exact position of the solvent front. The thermodynamic solvent front, which may be slightly lower than the visible solvent front for unsaturated flow, can be determined by using an unretained substance as a solvent-front marker. The repeatability of  $R_F$  values can be improved using the saturation grade of the developing chamber to interconvert observed  $R_F$  values between different chamber designs. When the samples are available as standards, it is common practice to separate the standards and samples in



**Figure 4** Illustration of the measurements used to calculate the retardation factor and variation of the plate height with solvent front migration distance.

the same system for identification purposes. In surveillance programs, the simultaneous separation of appropriate standard substances is used to improve the certainty of identification by correcting observed  $R_F$  values to standard  $R_F$  values for automated library searches. Using the mean list method, for example, all substances that migrate in a  $R_F$  window, which might be confused, are ranked and compared across a number of separation systems. If the separation systems are complementary, the list of possible substances that might be confused will become shorter as an increasing number of substances fall outside the identification window for the unknown. Eventually, only a handful of possible substances remain on the list, at which point suitable selective separation and spectroscopic techniques are used to confirm the identification of the unknown.

The  $R_F$  value is not related directly to the distribution properties of the separation system. The  $R_M$  value is used in studies that attempt to correlate migration properties with solute structure. The  $R_M$  value is equivalent to the ratio of the residence time of the solute in the stationary and mobile phases, and is formally equivalent to the retention factor ( $\log k$ ) in column liquid chromatography. It is calculated from the  $R_F$  value by

$$R_M = \log [(1 - R_F)/R_F] \quad [5]$$

## Zone Broadening and the Plate Height Equation

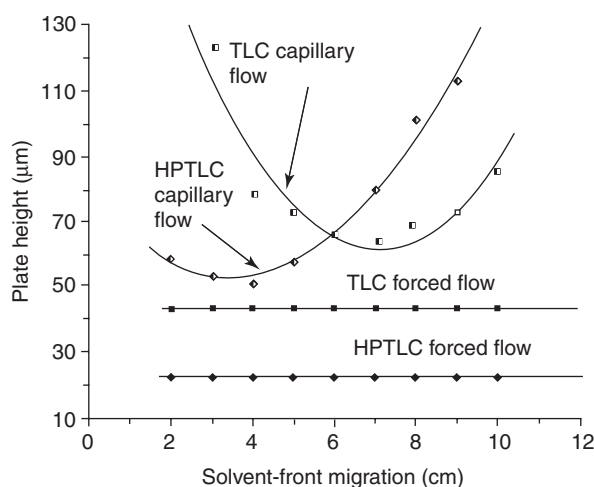
Zone dimensions in the direction of development in modern TLC are usually symmetrical, and their profiles recorded by densitometry are Gaussian to a good approximation. The average plate height ( $H_{\text{obs}}$ ) and plate number ( $N_{\text{obs}}$ ) can be calculated directly from the chromatogram as

$$H_{\text{obs}} = w^2/aZ_X \quad [6]$$

$$N_{\text{obs}} = a(Z_X/w)^2 \quad [7]$$

where  $w$  is the peak width at the base (when  $a = 16$ ) or the peak width at half-height ( $a = 5.54$ ). The zone migration distance,  $Z_X$ , is equivalent to  $R_F(Z_f - Z_o)$ . In TLC each separated zone migrates a different distance, the opposite of separations by column chromatography. For column chromatography, it is reasonable to consider the plate height as a system property with an approximately constant value for all solutes with similar diffusion coefficients. In TLC, individual zones only experience those theoretical plates through which they travel and, therefore, the average plate height and plate number depend on the zone migration distance. In addition, an average value is used to indicate that the mobile phase velocity varies throughout the chromatogram when capillary flow conditions are used and individual zones migrate through regions of different local efficiency. Thus, the average plate height and the observed plate height in the chromatogram are identical. For the same reason, the average plate height depends on the distance between the sample application position and the mobile phase entry position or solvent level for the layer ( $Z_o$ ). Selecting higher values of  $Z_o$  results in zones being forced to migrate with a lower mobile phase velocity, which results in further zone broadening.

A plot of the average plate height as a function of the solvent-front migration distance, **Figure 5**, reveals several important features of zone broadening with capillary flow conditions. First, there is a dominant relationship between the solvent-front migration distance, the average particle size for the layer, and the separation performance. Layers of a small average particle size (HPTLC) afford compact zones if the solvent-front migration distance does not exceed  $\sim 5$ – $6$  cm. At longer solvent-front migration distances, layers of a larger average particle size (TLC) are more efficient. This contrary finding is easily explained by the relative permeability of the layers. The mobile phase velocity for the HPTLC layers declines rapidly with the solvent-front migration distance until eventually zone broadening



**Figure 5** Variation of the average plate height as a function of the solvent front migration distance for HPTLC and TLC layers. (Reprinted with permission from Poole CF and Poole SK (1995) Multidimensionality in planar chromatography. *Journal of Chromatography A* 703: 573–612; © Elsevier.)

exceeds the rate of zone center migration. It is futile to use solvent-front migration distances longer than 5–6 cm for HPTLC layers for this reason. For the more permeable TLC layers the mobile phase velocity is higher for longer distances. As expected, the minimum in the plate height is shifted to longer solvent-front migration distances for the TLC layer, but the intrinsic efficiency, measured by the minimum plate height, is more favorable for the HPTLC layer. When the development distance is optimized the separation performance of conventional and high-performance layers are not very different. The virtues of HPTLC layers are that it requires a shorter migration distance to achieve a given efficiency resulting in faster separations and more compact zones, which are easier to detect by scanning densitometry.

For forced flow separations a constant plate height independent of the solvent-front migration distance is obtained (see **Figure 5**). The minimum plate height for capillary flow is always greater than the minimum for forced flow at the optimum mobile phase velocity. The limited range of mobile phase velocities ( $0.02$ – $0.005 \text{ cm s}^{-1}$ ) for capillary flow prevents the optimum performance of the layers from being realized. At the mobile phase optimum velocity, forced flow affords zones that are more compact and have shorter separation times compared with capillary flow. In addition, for forced flow the intrinsic efficiency increases with a reduction of the average particle size for the layer.

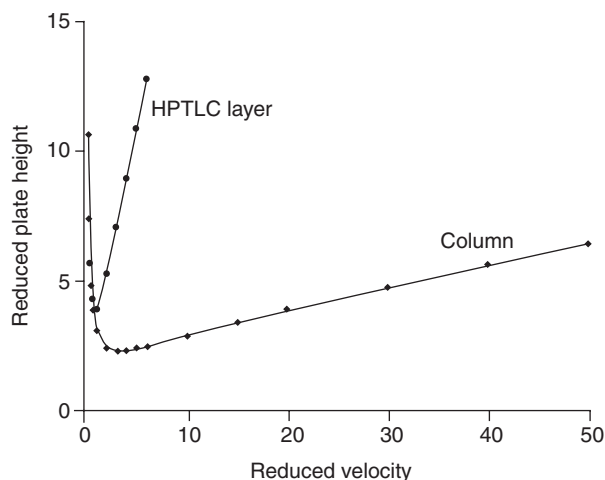
For forced flow conditions a direct comparison with column separations is possible within the framework of the reduced parameter model. The



**Table 2** Common terms used in the reduced parameter model for forced flow separations

Parameter	Description
Reduced plate height	Number of particles to a theoretical plate $h = H_{\text{obs}}/d_p$
Reduced velocity	Rate of flow of the mobile phase relative to the rate of diffusion of the solute over one particle diameter $v = ud_p/D_M$
Flow resistance parameter	Measure of the resistance to flow of the mobile phase $\phi = \Delta P d_p^2 t_M / \eta L$
Separation impedance	Measure of the difficulty of achieving a certain performance ( $E_{\text{min}}$ corresponds to optimum performance) $E = h^2 \phi$
Knox equation	$h = A v^{1/3} + B/v + C v$ $A$ = measure of the uniformity and packing density of the layer; $B$ = measure of the extent to which solute diffusion is hindered by the presence of particles in the layer and the structure of the layer; $C$ = reflects the efficiency of mass transfer between the streaming portion of the mobile phase and the stationary phase

$d_p$  = particle size,  $u$  = mobile phase (constant) velocity,  $D_M$  = mobile phase solute diffusion coefficient,  $\Delta P$  = pressure drop across the layer,  $t_M$  = retention time of an unretained compound (or compound moving with the solvent front),  $\eta$  = mobile phase viscosity,  $L$  = development length.



**Figure 6** Plot of the reduced plate height against the reduced velocity for a typical HPTLC layer and HPLC column. (Reprinted with permission from Poole CF (1999) Planar chromatography at the turn of the century. *Journal of Chromatography A* 856: 399–427; © Elsevier.)

reduced parameters are dimensionless and calculated from characteristic column and layer properties as indicated in Table 2. Figure 6 illustrates the variation of the reduced plate height ( $h$ ) as a function of the reduced mobile phase velocity ( $v$ ) for an HPTLC silica gel layer with forced flow conditions, for a typical column. The optimum reduced mobile phase velocity is shifted to a lower value compared with that for the column and the minimum in the reduced plate height ( $h \approx 3.5$ ) is higher than typical values for a good column ( $h \approx 2.0$ – $2.5$ ). Also, at higher reduced mobile phase velocities the reduced plate height for

the layer is significantly larger than for the column. Consequently, separations by forced flow will be slower than those achieved with typical columns, and fast separations at high flow rates will be less efficient.

The reasons for the difference between columns and layers can be deduced by interrogating the coefficients of the Knox equation (see Table 2) providing the best fit to the experimental data (Table 3). The term ( $B/v$ ) is a measure of the contribution of longitudinal diffusion to the plate height. The magnitude of the  $B$  coefficient is related to the solute diffusion coefficient and is expected to have similar values for columns and layers. The ( $A v^{1/3}$ ) term is a measure of flow anisotropy within the streaming part of the mobile phase and is related to the packing density and homogeneity of the layer. For HPTLC layers the  $A$  coefficients are smaller than for typical columns, indicating that the layers are homogeneously packed and have a good packing structure. TLC layers are not as homogeneously packed, perhaps because of the greater difficulty of preparing layers from silica gel with a wider particle-size distribution. The ( $C v$ ) term is a measure of the resistance to mass transfer between the stationary phase and the streaming portion of the mobile phase. Typical values of the  $C$  coefficient for columns are much smaller than for layers. The large resistance to mass transfer term might arise as a result of restricted diffusion within the porous particles or from partial blocking of the intraparticle channels, owing to the presence of binder.

Some expectations for forced flow separations are summarized in Table 4. For a normal development distance of 18 cm a modest increase in performance



**Table 3** Characteristic properties of silica gel precoated layers and columns

Parameter	Layers		Column
	High performance	Conventional	High performance
Porosity			
Total	0.65–0.70	0.65–0.75	0.8–0.9
Interparticle	0.35–0.45	0.35–0.45	0.4–0.5
Intraparticle	0.28	0.28	0.4–0.5
Flow resistance parameter	875–1500	600–1200	500–1000
Apparent particle size ( $\mu\text{m}$ )	5–7	8–10	$d_p$
Minimum plate height ( $\mu\text{m}$ )	22–25	35–45	2–3 $d_p$
Optimum velocity ( $\text{mm s}^{-1}$ )	0.3–0.5	0.2–0.5	2
Minimum reduced plate height	3.5–4.5	3.5–4.5	1.5–3
Optimum reduced velocity	0.7–1.0	0.6–1.2	3–5
Knox equation coefficients			
Flow anisotropy ( $A$ )	0.4–0.8	1.7–2.8	0.5–1.0
Longitudinal diffusion ( $B$ )	1.2–1.6	1.2–2.0	1–4
Resistance to mass transfer ( $C$ )	1.4–2.4	0.70–0.85	0.05
Separation impedance	10 000–20 000	11 000–13 000	2000–9000
Mean pore diameter (Si 60) (nm)	5.9–7.0	6.1–7.0	

**Table 4** Performance expectations for forced flow separations. Assumptions: viscosity =  $3.5 \times 10^{-4} \text{ N s m}^{-2}$  and solute diffusion coefficient =  $2.5 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$  (other terms defined in **Table 2**)

Development time (min)	Pressure drop (atm)	$N_{\text{max}}$	Development length (cm)
(a) HPTLC (optimum conditions $h = 3.75$ , $\nu = 0.8$ , $\phi = 800$ , and $d_p = 6 \mu\text{m}$ )			
4	2.1	3 550	8
9	4.7	8 000	18
25	12.9	22 200	50
50	25.8	44 400	100
(b) HPTLC (fast development option $h = 9$ and $\nu = 5$ )			
0.6	12.9	1 480	8
1.4	29.1	3 330	18
4.0	80.7	9 250	50
8.0	161.0	18 500	100
(c) HPTLC ( $d_p = 3 \mu\text{m}$ ; other parameters as in (a))			
2.0	16.5	7 610	8
4.5	37.2	17 100	18
12.5	103	47 600	50
25.0	207	95 200	100
(d) Conventional TLC ( $h = 4.5$ and $\nu = 0.8$ , $\phi = 600$ , and $d_p = 9 \mu\text{m}$ )			
6	0.44	1 980	8
13.5	1.03	4 450	18
37.5	2.9	12 350	50
75	5.7	24 700	100

Reproduced with permission from Poole CF and Wilson ID (1997) *Journal of Planar Chromatography* 10: 332–335.

(a maximum plate number of 8000) in a credible time of 9 min is achieved compared with typical results for capillary flow (plate number < 5000 in  $\sim 25$ –45 min). Really significant increases in efficiency are achieved

only by the use of longer layers at the expense of separation time. The optimum mobile-phase velocity is sufficiently low that compared with columns the pressure drop remains modest. Fast separations on HPTLC layers result in low efficiency, and high pressures are required if long development distances are used. These conditions are not useful for most separations, especially because layers enable separations to be performed in parallel, so the relative time per separation is not adversely impaired compared with those of other separation techniques. If a relatively large plate number is to be achieved then high pressures are required. Reducing the particle size from an average of 5 to  $3 \mu\text{m}$  results in significantly improved efficiency with favorable separation times, but is more demanding in terms of operating pressure. No benefits are expected from the use of conventional layers for forced flow separations.

### Layer Characteristics

Layers have a heterogeneous structure consisting of the sorbent particles, held together by added binder, with other possible additives, such as a fluorescence indicator for visualization of ultraviolet-absorbing compounds. Representative values for the kinetic properties of precoated silica gel layers are summarized in **Table 3**. The similar values for the interparticle porosity of columns and layers suggests that their packing density is similar, whereas the significant difference in the total porosity and intraparticle porosity indicates that the intraparticle volume of the layers is substantially smaller. A significant amount of the binder used in stabilizing the layers must be

contained within the pores. The difference in flow resistance properties of the HPTLC and TLC layers might arise from the presence of a greater portion of fine particles in the HPTLC layer compared with the TLC layer, reflecting the fact that HPTLC layers are prepared from particles of a smaller average particle size initially. Layers with a narrower particle size distribution but similar average particle size could lead to a further improvement in capillary flow separations on HPTLC layers. The small difference between apparent average particle sizes for HPTLC and TLC layers is the origin of the small difference between the minimum plate height for modern layers. Historically, conventional layers afforded poorer separations because they were prepared from sorbents with a significantly wider particle size distribution. Contemporary TLC layers are prepared from particles of smaller average size and narrower particle size distribution compared with a decade ago, and this is reflected in their improved separation performance.

### Multiple Development

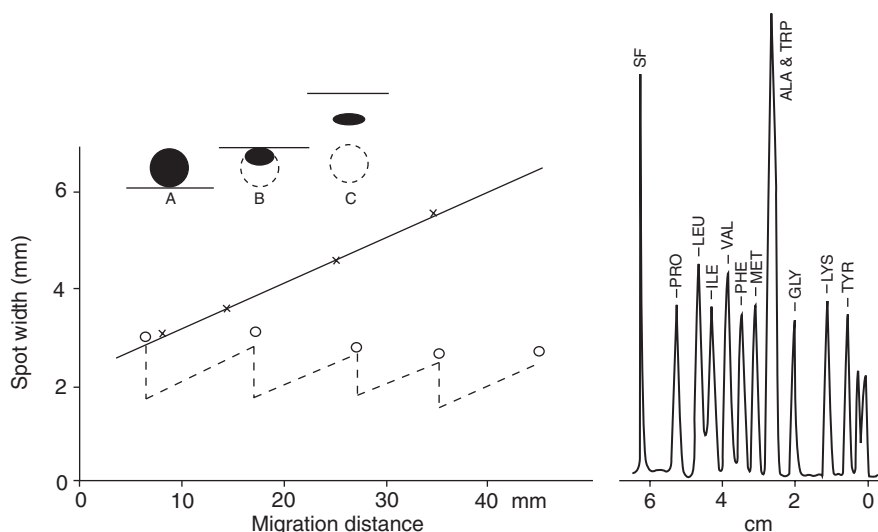
Multiple development techniques (Table 5) provide a complementary approach to forced flow for minimizing zone broadening. All multiple development techniques employ successive repeated development of the layer in the same direction with the removal of mobile phase between developments. Approaches differ in the changes made at each development, such

as the solvent front migration distance, time of development, and mobile phase composition, and in the total number of developments employed for the separation. By varying the mobile phase composition at some or all development steps, a simple means of generating solvent gradients is obtained. The separation can be recorded as a single chromatogram using incremental multiple development with a reverse gradient of solvent strength (e.g., automated multiple development or AMD); with an increasing gradient of solvent strength the mixture composition is inferred by scanning intermediate steps in the development sequence where the resolution of each group of components is optimized, even if at a later development step they are merged again.

A unique feature of multiple development techniques is the zone focusing mechanism that is exploited to counteract the normal zone broadening that occurs in each successive development. Each time the solvent front traverses the stationary sample zone, the zone is compressed in the direction of development (Figure 7). The compression occurs because the mobile phase contacts the bottom edge of the zone first; here the sample molecules start to move forward before those molecules still ahead of the solvent front. When the solvent front has reached beyond the zone, the focused zone migrates and is subject to the normal zone broadening mechanisms. Both theory and experiment indicate that beyond a minimum number of development steps zone widths converge to a constant value that is roughly

**Table 5** Multiple development techniques

<i>Method</i>	<i>Features</i>
Multiple chromatography	<ul style="list-style-type: none"> <li>• Fixed development length</li> <li>• Same mobile phase for each development</li> <li>• The number of developments can be varied</li> </ul>
Incremental multiple development	<ul style="list-style-type: none"> <li>• Variable development length               <ol style="list-style-type: none"> <li>(a) First development is the shortest</li> <li>(b) Each subsequent development is increased by a fixed distance</li> <li>(c) Last development length corresponds to the maximum useful development distance</li> </ol> </li> <li>• Same mobile phase for each development</li> <li>• The number of developments can be varied</li> </ul>
Two-dimensional chromatography	<ul style="list-style-type: none"> <li>• Sample applied at corner of layer</li> <li>• Two developments used, with the direction of the second being orthogonal to the first</li> <li>• Requires that the retention mechanisms for the two developments are complementary</li> </ul>
Increasing solvent strength gradients	<ul style="list-style-type: none"> <li>• Fractionates sample into manageable subsets</li> <li>• Optimizes separation of each subset</li> <li>• Complete separation of all components is not achieved at any segment in the development sequence</li> </ul>
Decreasing solvent strength gradients (AMD)	<ul style="list-style-type: none"> <li>• Uses incremental multiple development</li> <li>• First development employs the strongest solvent with a weaker solvent for each subsequent step</li> <li>• Final separation recorded as a single chromatogram</li> </ul>



**Figure 7** Illustration of the zone focusing mechanism (left) and its application to the separation of a mixture of PTH-amino acid derivatives (right). The broken line on the left-hand side of the figure represents the change in spot size due to the expansion and contraction stages in multiple development and the solid line depicts the expected zone width for a zone migrating the same distance in a single development. (Adapted with permission from Poole CF, Poole SK, Fernando WPN, *et al.* (1989) *Journal of Planar Chromatography* 2: 336–345.)

independent of migration distance. The position of a zone in multiple chromatography (constant development length) is given by

$$R_{F,p} = 1 - (1 - R_F)^p \quad [8]$$

and in incremental multiple development (each successive development distance is increased by a constant amount) by

$$R_{F,p} = 1 - (1 - R_F)[\{1 - (1 - R_F)^p\}/pR_F] \quad [9]$$

where  $R_{F,p}$  is the apparent  $R_F$  value after  $p$  successive developments and  $R_F$  is the  $R_F$  value in the first development. The prediction of zone widths is more complicated than for zone positions. Inexact phenomenological models and computer simulations are used for this purpose. The theory developed so far is for isocratic mobile phases and is not applicable to solvent gradients, although the same general phenomenon of a roughly constant zone width is observed in optimized separations by automated multiple development. For difficult separations by capillary flow, multiple development is the general strategy used to increase the zone capacity.

## Resolution

The resolution,  $R_S$ , between two separated zones in TLC is defined as the ratio between the separation of

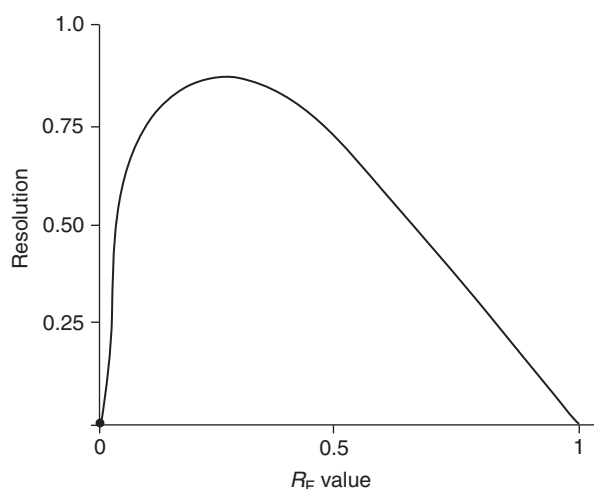
the zone centers and the average zone widths

$$R_S = 2(Z_{X2} - Z_{X1})/(w_{b1} + w_{b2}) \quad [10]$$

where  $w_b$  is the width of the zone at its base and the subscripts 1 and 2 refer to the positions of the zones in the chromatogram ( $Z_{X2} > Z_{X1}$ ). To achieve a certain separation it is necessary to know how the resolution varies with the experimental parameters, such as the layer efficiency, the ratio of the equilibrium constants for the separation process (or selectivity), and the position of the zones in the chromatogram. There is no exact relationship for this purpose, but qualitatively eqn [11] and Figure 8 provide an adequate qualitative description of the separation process for capillary flow conditions

$$R_S = [\sqrt{(N_1 R_{F2})/4}][(k_1/k_2) - 1][1 - R_{F2}] \quad [11]$$

where  $N_1$  is the plate number the zones would have passed over if they migrated to the solvent front position and  $k$  is the retention factor. Resolution increases with efficiency in a manner that depends linearly on the  $R_F$  value. Relatively small changes in selectivity have a large impact on the ease of obtaining a separation, since the total plate number available is never very large. Separations by TLC are easy when  $(R_{F2} - R_{F1}) > 0.1$ , and very difficult or impossible for  $(R_{F2} - R_{F1}) < 0.05$ , in the region of the optimum  $R_F$  value for the separation. The effect of zone location on resolution shows the opposite behavior to layer



**Figure 8** Variation of the resolution of two closely migrating zones as a function of the  $R_F$  value of the faster moving zone.

quality. At large values of  $R_{F2}$ , the term  $(1 - R_{F2})$  is small and resolution falls to zero at  $R_{F2} = 1$ . Differentiation of eqn [11] indicates that the maximum resolution of two difficult to separate zones will occur at an  $R_F$  value of  $\sim 0.3$ . From **Figure 8**, it can be seen that resolution does not change significantly for  $R_F$  values between 0.2 and 0.5; and within this range, the resolution is greater than 92% of the maximum value (75% between  $R_F = 0.1$  and 0.6). This is the target  $R_F$  range for those components most difficult to separate in a mixture. Of course, to create sufficient separation capacity to separate complex mixtures, only the sample components most difficult to separate will be positioned in the optimum range, with substances easier to separate outside of it, since the optimum resolution region has a limited zone capacity.

Resolution in forced flow separations is not restricted by the same factors that apply to capillary flow. Resolution increases almost linearly with the solvent-front migration distance and is highest at the optimum mobile phase velocity. There is no theoretical limit to resolution for forced flow separations; the upper bounds are established by practical constraints, such as plate length, separation time, and inlet pressure.

## Zone Capacity

The potential of a chromatographic system to provide a separation can be estimated from its zone capacity, sometimes referred to as the separation number or spot capacity in TLC. The zone capacity envisages the arrangement of the zones in a chromatogram as being similar to a string of beads,

**Table 6** Experimental or theoretical zone capacity for different separation conditions

Separation mode	Dimensions	Zone capacity
<i>(i) Predictions from theory</i>		
Capillary flow	1	$< 25$
Forced flow	1	$< 80$ (up to 150 depending on pressure limit)
Capillary flow	2	$< 400$
Forced flow	2	Several thousand
<i>(ii) Experimental observations</i>		
Capillary flow	1	12–14
Forced flow	1	30–40
Capillary flow (AMD)	1	30–40
Capillary flow	2	$\sim 100$
<i>(iii) Predictions based on (ii)</i>		
Forced flow	2	$\sim 1500$
Capillary flow (AMD)	2	$\sim 1500$

each bead touching its neighbor, with no unoccupied space between the beads. In practice, the zone capacity leads to an inflated estimate of the real separation capacity, since real chromatograms do not consist of equally spaced zones. It does provide, however, a plausible guide for comparison of the separation potential of different thin-layer techniques. In addition, it affords an indication of the possibility of separating a given mixture, for unless the zone capacity exceeds the number of sample components by a significant amount, the separation will be difficult to impossible to achieve.

The zone capacity is defined as the number of spots completely separated with  $R_s = 1$  between  $R_F = 0$  and  $R_F = 1$ . Most methods of calculation, however, use approximate models for either the form of the chromatogram, such as a geometric increase in zone size throughout the chromatogram, or attempt to estimate the increase in zone size by a theoretical model. No approach is completely satisfactory. Some represented results from theory and experiment are summarized in **Table 6**. Results from theory are probably too high and represent an upper bound to the zone capacity. Experiment indicates a zone capacity of  $\sim 12$ –14 for a single development with capillary flow. This rises to  $\sim 30$ –40 for forced flow conditions. Automated multiple development with capillary flow provides a similar zone capacity to forced flow separations. Two-dimensional TLC employing different retention mechanisms for each orthogonal separation provides a significantly larger zone capacity than any of the one-dimensional development methods. This accounts for the continuing interest in this technique. In theory, the zone capacity becomes very large for two-dimensional TLC if forced flow or multiple

development is used for the orthogonal separations, but this has not been confirmed by experiment.

*See also: Thin-Layer Chromatography: Overview; Plate Technology; Method Development; Instrumentation.*

## Further Reading

- Geiss F (1987) *Fundamentals of Thin-Layer Chromatography*. Heidelberg: Huthig.
- Nyiredy Sz (ed.) (2001) *Planar Chromatography. A Retrospective View for the Third Millennium*. Budapest: Springer Scientific Publisher.
- Poole CF (2003) *The Essence of Chromatography*. Amsterdam: Elsevier.
- Poole CF (2003) Thin-layer chromatography: Opportunities and challenges. *Journal of Chromatography A* 1000: 963–984.
- Poole CF and Dias NC (2001) Practioner's guide to method development in thin-layer chromatography. *Journal of Chromatography A* 892: 123–142.
- Poole CF and Fernando WPN (1993) Comparison of the kinetic properties of commercially available precoated silica gel plates. *Journal of Planar Chromatography* 6: 357–361.
- Sherma J and Fried B (eds.) (1996) *Handbook of Thin-Layer Chromatography*. New York: Marcel Dekker.
- Wilson ID, Adlard ER, Cooke M, and Poole CF (eds.) (2000) *Encyclopedia of Separation Science*. London: Academic Press.

## Plate Technology

**F Rabel**, EMD Chemicals, Inc., Gibbstown, NJ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

One of the most important tools when performing thin-layer chromatography (TLC) is the thin-layer plate itself. It is often overlooked as not important since it is just removed from the box as a prepared TLC plate, spotted and developed. But considerable technology has gone into the manufacturing of the TLC sorbents, the binders, the supports, and the plating machinery to make a reproducible product suitable for the market. If the layer is not well made, it would be possible to see thickness differences, bubbles in the dried layer, crooked solvent fronts, and impurities at the solvent front – all of which can contribute to poor, nonreproducible results. What is desired is an even, flat layer, free of imperfections and impurities. This article will describe some of the aspects of the plate-making process including details of the sorbents, the binders, and the supports. Details of the manufacture, although of interest, are proprietary, and as such cannot be discussed here.

## Before the Prepared TLC Plate

Prior to the availability of the prepared thin-layer plates, anyone wanting to do TLC had to make their own plates. This is a laborious task, involving cleaning glass plates, putting them on a special holder, making up a slurry of the sorbent with water (of a

definite consistency), putting this slurry in an applicator, and then drawing this applicator across the plates at a controlled, even rate. A gate could be set on the applicator to allow different thicknesses of thin layers to be deposited on the glass plates.

After a few minutes the slurry began to set, turning from a wet, watery look to a matt finish as the binder (discussed below) begins to hydrate. The plate was then allowed to air dry, and finally dried in an oven at 100°C for 30 min to remove remaining moisture.

Although time consuming, it was the only way people interested in TLC could get the thin-layer plates for their work. Fortunately, the manufacturers of the sorbents realized that supplying a precoated plate would be more cost-efficient and more reproducible for the users. In the early 1960s, the first prepared TLC plates came onto the market.

## The Sorbents Used in TLC

### Silica Gel (Silica)

The sorbent used the most for TLC work is silica gel. It is an amorphous, porous matrix made by the addition of acid to a sodium silicate solution. With proper control of the hydration process, the silica gel will cross-link to form a hydrogel, which is then washed and dried, yielding a product with pore sizes of  $\sim 60 \text{ \AA}$  and a surface area of  $\sim 500 \text{ m}^2 \text{ g}^{-1}$ . Other pore sizes/surface area silica gels can be made by changing the conditions of the hydrolysis, but the  $60 \text{ \AA}$  silica gel has been the standard used in TLC for



development is used for the orthogonal separations, but this has not been confirmed by experiment.

*See also: Thin-Layer Chromatography: Overview; Plate Technology; Method Development; Instrumentation.*

## Further Reading

- Geiss F (1987) *Fundamentals of Thin-Layer Chromatography*. Heidelberg: Huthig.
- Nyiredy Sz (ed.) (2001) *Planar Chromatography. A Retrospective View for the Third Millennium*. Budapest: Springer Scientific Publisher.
- Poole CF (2003) *The Essence of Chromatography*. Amsterdam: Elsevier.
- Poole CF (2003) Thin-layer chromatography: Opportunities and challenges. *Journal of Chromatography A* 1000: 963–984.
- Poole CF and Dias NC (2001) Practioner's guide to method development in thin-layer chromatography. *Journal of Chromatography A* 892: 123–142.
- Poole CF and Fernando WPN (1993) Comparison of the kinetic properties of commercially available precoated silica gel plates. *Journal of Planar Chromatography* 6: 357–361.
- Sherma J and Fried B (eds.) (1996) *Handbook of Thin-Layer Chromatography*. New York: Marcel Dekker.
- Wilson ID, Adlard ER, Cooke M, and Poole CF (eds.) (2000) *Encyclopedia of Separation Science*. London: Academic Press.

## Plate Technology

**F Rabel**, EMD Chemicals, Inc., Gibbstown, NJ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

One of the most important tools when performing thin-layer chromatography (TLC) is the thin-layer plate itself. It is often overlooked as not important since it is just removed from the box as a prepared TLC plate, spotted and developed. But considerable technology has gone into the manufacturing of the TLC sorbents, the binders, the supports, and the plating machinery to make a reproducible product suitable for the market. If the layer is not well made, it would be possible to see thickness differences, bubbles in the dried layer, crooked solvent fronts, and impurities at the solvent front – all of which can contribute to poor, nonreproducible results. What is desired is an even, flat layer, free of imperfections and impurities. This article will describe some of the aspects of the plate-making process including details of the sorbents, the binders, and the supports. Details of the manufacture, although of interest, are proprietary, and as such cannot be discussed here.

## Before the Prepared TLC Plate

Prior to the availability of the prepared thin-layer plates, anyone wanting to do TLC had to make their own plates. This is a laborious task, involving cleaning glass plates, putting them on a special holder, making up a slurry of the sorbent with water (of a

definite consistency), putting this slurry in an applicator, and then drawing this applicator across the plates at a controlled, even rate. A gate could be set on the applicator to allow different thicknesses of thin layers to be deposited on the glass plates.

After a few minutes the slurry began to set, turning from a wet, watery look to a matt finish as the binder (discussed below) begins to hydrate. The plate was then allowed to air dry, and finally dried in an oven at 100°C for 30 min to remove remaining moisture.

Although time consuming, it was the only way people interested in TLC could get the thin-layer plates for their work. Fortunately, the manufacturers of the sorbents realized that supplying a precoated plate would be more cost-efficient and more reproducible for the users. In the early 1960s, the first prepared TLC plates came onto the market.

## The Sorbents Used in TLC

### Silica Gel (Silica)

The sorbent used the most for TLC work is silica gel. It is an amorphous, porous matrix made by the addition of acid to a sodium silicate solution. With proper control of the hydration process, the silica gel will cross-link to form a hydrogel, which is then washed and dried, yielding a product with pore sizes of  $\sim 60 \text{ \AA}$  and a surface area of  $\sim 500 \text{ m}^2 \text{ g}^{-1}$ . Other pore sizes/surface area silica gels can be made by changing the conditions of the hydrolysis, but the  $60 \text{ \AA}$  silica gel has been the standard used in TLC for



many years. The active part of any silica gel surface is the Si–O–H groups (called silanols). They strongly bind polar groups on any structure allowing for differentiation and separation. Water also binds strongly with silanols, so control of moisture in the air and the solvent is important for getting reproducible results when using this type of plate. The surface pH of most silica gels is neutral (pH 7.0), allowing for the separation of neutral, acidic, and basic site compounds; hence, its wide use continues even today. There is probably no class of compounds that have not been successfully separated on silica gel layers, so the literature is easily found for any new attempts by users to applying this layer for their work.

After the initial manufacturing, the resulting large silica gel particles are crushed to a powder-like consistency. This is the fine size range found on all TLC plates today. Depending on the product to be made – whether a standard thin layer (TLC), a high-performance thin layer (HPTLC or HPTLC spherical), or a preparative layer (PLC) – the particle size range is changed to give the best in performance and plate-making ability. This is done by sieving or air classifying the crushed or spherical materials. The UTLC (ultra-thin layer) is made differently than the others and will be discussed separately. These are shown in Table 1. The cross-section of a standard TLC plate is shown in Figure 1.

### Aluminum Oxide (Alumina)

This is another metal oxide used in thin-layer applications. The chemical formula is  $\text{Al}_2\text{O}_3$ . On aluminum oxide thin-layers sorption is based on partial positive and negative charges on the surface and any water sorbed thereon. In the manufacturing process aluminum oxide can be made to have a basic surface (pH 9–10), a neutral surface (pH 7–8), or an acidic surface (pH 4–4.5). This allows for a different type of adsorption separation based solely on surface pH. Other properties of the different aluminas available for TLC are shown in Table 2. The different pore size and surface area types will also impart different separation characteristics to these sorbent layers. Layers of this sorbent are available with and without binders both organic and inorganic. A whole range of

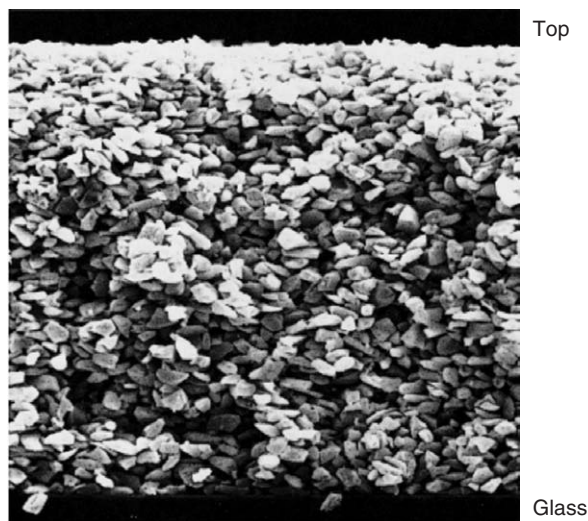
separations have been done on alumina, particularly basic species and polyaromatic hydrocarbons.

### Other Oxides

Theoretically any metal oxide could be used as a thin-layer sorbent, but most have no specific properties that promote a unique separation or selectivity to a thin-layer plate. Infrequently, someone publishes work done on another metal oxide, but only those mentioned above are used frequently and are considered significant.

### Cellulose

This is a sorbent derived from natural sources. It is fibrous in nature, with a polymerization of 400–500 glucose units. Of course, any fiber can be broken to smaller sizes more suitable for TLC. And the native



**Figure 1** Cross-section of prepared silica gel TLC plate.

**Table 2** Aluminum oxide specifications for TLC

Type of aluminum oxide	Pore diameter (Å)	Surface area ( $\text{m}^2 \text{g}^{-1}$ )
60 (type E)	60	180–200
90	90	100–130
150 (type T)	150	70

**Table 1** Silica gel particle sizes for different types of layers

	TLC	HPTLC	Spherical HPTLC	PLC	UTLC
Average particle size	15 $\mu\text{m}$	5 $\mu\text{m}$	4 or 7 $\mu\text{m}$	25 $\mu\text{m}$	Not applicable
Particle size range	5–20 $\mu\text{m}$	4–7 $\mu\text{m}$	3–5 or 6–8 $\mu\text{m}$	20–40 $\mu\text{m}$	Not applicable

fiber can also be hydrolyzed to a smaller fiber, with different properties. This latter type is usually referred to as 'microcrystalline' and is ~40–200 glucose units in length. Of course, because it is the same material used for paper, thin layers of cellulose have replaced any paper chromatography done years previously. Although the solvent migration with TLC work on cellulose is slow compared to that done on silica gel, it has a unique selectivity for many biological compounds.

Derived or coated celluloses are also available. These include acetylated, diethylaminoethyl (DEAE), polyethyleneimine (PEI). More on these will be discussed in the section Ion Exchange Layers below.

Cellulose binds very well to glass because of the fiber overlay and hydrogen bonding with the glass surface thus is often binder free. To prevent cracking when the layer is made, the thickness is usually only 0.1 mm on the analytical layer. This layer has been used to separate amino acids, antibiotics, flavonoids, peptides, and related compounds.

### Polyamide

Two versions of polyamide have found use in TLC. Both – polyamide 6 (polycaprolactam) and polyamide 11 (polyundecanamide) – are synthetic resins, each having slightly different properties. The resins have to be chilled to be effective ground for use on TLC plates. These are very good sorbents for any compounds that can hydrogen bond to the amide groups. These layers are also usually free of binders and are also manufactured with only 0.1 mm thickness.

### Bonded Silica Layers

Over the past 35 years, reversed-phase high-performance liquid chromatography (HPLC) usage has grown tremendously. The manufacturers of TLC plates thought that the chromatographer would also like to switch to various bonded reversed-phase TLC plates. TLC and HPTLC layers with RP18 (an –Si–C18 linear chain), RP8 (an –Si–C8 linear chain), RP2 (–Si–CH<sub>2</sub>–CH<sub>3</sub>), phenyl, and diphenyl (and others like propyl amino (–Si–CH<sub>2</sub>–CH<sub>2</sub>–CH<sub>2</sub>–NH<sub>2</sub>), propyl cyano (–Si–CH<sub>2</sub>–CH<sub>2</sub>–CH<sub>2</sub>–CN), and diol (–Si–CH<sub>2</sub>–CHOHCH<sub>2</sub>OH) for normal-phase separations – alternatives to pure silica gel plates) have been available for 30 years. However, perhaps only 10% of the methods previously done on silica gel have been transferred to bonded reversed-phase plates based upon their sales ratio. The reversed-phase mode does offer unique ionization control with buffers added to the much simpler methanol/water or acetonitrile/water mobile phase combinations.

Although more expensive than silica gel plates, all the bonded phases are available as HPTLC layers, so smaller plates can be used, cutting down the cost of each plate used. One company makes dual layer plates, with part of the plate being silica gel, the other part being an RP18 section. Thus, on one plate you can accomplish two modes of chromatography, first silica gel/adsorption, and then, secondly, RP18/reversed phase (or the reverse order). This is referred to as 'multidimensional' chromatography. Such a plate increases the resolving power of the chromatography because you are doing two separations using two modes rather than using only one separation and one mode.

Any of these reversed-phase plates can be useful in developing solvent systems for similar reversed-phase HPLC work, or can help determine what may not be moving through such a column. It should be remembered that only a TLC plate shows what is left at the origin with a given solvent system and hence would remain at the top of any HPLC column with this same solvent system. It is these 'remains' of samples that drastically affect HPLC column separations and column life.

### Ion Exchange

Although in the past micronized ion-exchange resins had been placed on supports for thin-layer work, these are no longer manufactured. The only ion-exchange layers available today are composed of derivatized celluloses and these are limited in number. Only PEI, DEAE, and acetylated celluloses are available. Many others (carboxymethyl, aminoethyl, and phosphorylated) are no longer manufactured. Most of these plates fell into little use with the development of ion chromatography and ion-exchange separations possible with HPLC systems. Of the bonded silica thin layers, the propylamino mentioned above is also a weak cation exchanger when used with a buffered mobile phase that can protonate the amino group.

### Chiral

Often chromatographers have asked the TLC plate manufacturers if they could make chiral TLC layers. This would allow them to determine which of the five classes of chiral columns would be most useful for their enantiomer separations. Unfortunately, such layers would be very costly since a TLC plate is used only once. Only one plate is available, typically with a 'chir' in its name. It is composed of a layer that is impregnated with proline and copper(II) ion to allow complexation of amino acids and amino alcohols for enantiomer separation. Unfortunately, this

combination only works for these two classes of compounds.

As an alternative, papers have appeared in the thin-layer literature using TLC for chiral separations by putting the chiral discrimination reagent (antibiotic, cyclodextrin, etc.) into the mobile phase. The separations work as well as if the reagents had been coated or bonded to the silica gel (or in some cases, the reversed bonded phase).

### Diatomaceous Earth (Kieselguhr)

This is often part of a thin layer in a special type of plates. Diatomaceous earth has no separation capabilities, but it can be used at the bottom end of a thin-layer plate as a spotting area. Large quantities of a dilute solution can be placed on this sorbent. When dried and the mobile phase is allowed to travel through this part of the thin-layer plate, the solvent only concentrates the components with the solvent front. Only when the sample band has crossed to the silica gel (or other sorbent) does the separation actually begin. It is essentially an on-plate concentration device. Plates with these dual zones are called 'concentration zone' or 'preconcentration' plates. They can be of immense help in placing biological samples where debris can cause problems if spotted on an active layer. This portion of the plate can also be buffered or changed to affect different capture characteristics to the plate for other on-plate clean-up. Some manufacturers use a wide pore silica gel in place of the diatomaceous earth, but the use and effect are the same. The use of this plate is shown in Figure 2.

### Thickness of Layers

The usual thickness of an analytical thin layer is 0.2–0.25 mm when put onto a glass support. Certain sorbents would crack with this layer thickness on a

glass support or if they are on a flexible support (polymer or metal). In these instances, the thickness used is reduced to 0.1 mm. Thinner layers are also available to do more sensitive analyses (by keeping the sample closer to the surface of the thin layer, so the limits of detection are increased), or for special densitometer use (the thinner layer can be penetrated better by the densitometer light to activate the fluorescent indicator or the natural fluorescence of the compounds).

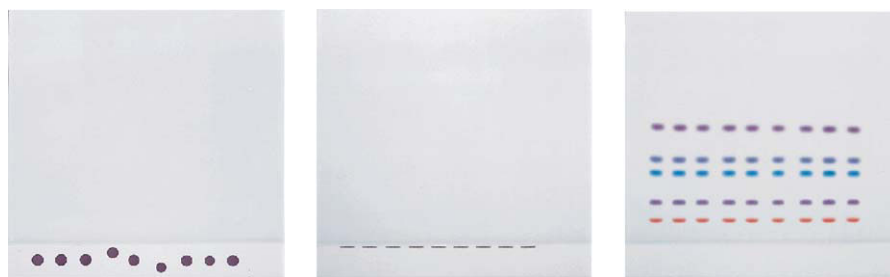
Scale up of a separation is also possible in TLC. In these cases, a thicker layer would be desirable. Preparative layers (PLC) are available from manufacturer's that are 0.5, 1.0, and 2.0 mm in thicknesses. If the separation needing to be scaled up is optimized on a thinner layer, then a thicker layer can be used for the isolation of greater quantities.

Loadability for any scale up work is always related to the best separation obtainable for the critical pair, and the solubility of the compounds in the mobile phase chosen for the separation. Thus, there is no absolute answer to the amounts you can put on a single, thicker TLC plate. This must be worked out by trial and error. This is particularly true since the thicker layers generate more heat of solvation, so the interior of a developing tank will be warmer with a thicker layered plate than with a thinner one. This affects the solubility and mobility of the compounds being separated. It is often necessary to reoptimize the mobile phase for the preparative separation to account for this effect.

### The Binders

#### Calcium Sulfate

No matter what the size of the silica gel, it will not stick to a support without a binder. The binder first used to make a suitable thin layer that could be spotted, developed, and visualized was calcium sulfate hemihydrate ( $\text{CaSO}_4 \cdot \frac{1}{2}\text{H}_2\text{O}$ ). This is more



**Figure 2** Use of concentration zone thin-layer plate. Left, spotting of samples; middle, initial concentration of sample spots; right, developed sample bands.

commonly known as gypsum (the G in the name of thin-layer plates containing this binder) or plaster of Paris. It is added in about a 10% concentration to help the adhesion of the silica gel to the support. This support is most often glass, but other supports will be discussed. With this binder a layer that adheres, but is still soft to the touch and easily abraded, is produced. All operations using a G-type plate have to be done carefully so as not to inadvertently touch and disturb the thin layer. They are shipped in special boxes that separate one plate from another to maintain their integrity.

### Polymeric Binders

Because of the fragile nature of the G-type plate, manufacturers began experimenting with various water-soluble monomers/polymers (like poly(vinyl alcohol) or poly(vinyl pyrrolidone)) as suitable binders. Added to the aqueous slurry during manufacture, they cross-link during drying to form a much harder surface thin-layer plate. The binders are also relative inert to visualizers (except those that strongly char) and have little affect on the separation. The particular binders used by each manufacturer are proprietary. The binder concentrations are usually 1% or less.

As the binders are water soluble, even after they have been cross-linked by the initial heating during manufacture (and any subsequent heating for activation) they can swell when exposed to high concentrations of water in the mobile phase. If this occurs, the solution is to use a  $0.1 \text{ mol l}^{-1}$  sodium chloride solution in place of the water or to use a specially formulated plate for high water content mobile phases. These water-tolerant plates are usually designated by the manufacturers as W-type, Water tolerant, or some such description.

### Other Additives

One of the most used additives to the thin-layer plate is the fluorescent indicator. These are inorganic compounds such as alkali earth tungstate, manganese activated zinc silicate, and others (many are proprietary) and are added in a concentration of 1–3% by weight to the slurry during manufacture. Being insoluble inorganic compounds, they do not migrate with any solvents used as mobile phases on the TLC plates. After development and drying of a thin-layer plate, the layer can be illuminated with the ultraviolet (UV) wavelength that activates that fluorescent indicator. This wavelength is usually specified in the plate name, i.e., Silica gel 60 F<sub>254</sub> – activated by 254 nm UV light.

When illuminated by light of this wavelength, the indicator will fluoresce green, light blue, or white, depending on the fluorescent inorganic used. When the UV light is absorbed by a component in the layer, the indicator does not fluoresce and a dark area can be seen against the fluorescing background. Depending on the concentration of the components in the mixture, the level of fluorescence in the plate, and the intensity of the UV light, some components are more easily seen than others. Certainly, very low level components can be ‘washed out’ by the intensity of the fluorescent indicator.

It is possible to add other chemicals to the slurry when preparing a TLC plate. These might be added to adjust the pH of the layer, to aid in the separation or detection of various compounds. These are not used routinely, so should be purchased as needed to insure they are not degraded in any manner.

Among these chemicals are silver nitrate – added to allow separation of *cis/trans* isomers, particularly for lipid analyses; potassium oxalate – added to allow discrimination of polyphosphoinositides; magnesium acetate – added to help in the separation of phospholipids; ammonium sulfate – added so that the plate is self-charring (after heating); carbomer – added for the analysis of mannitol/sorbitol; and sodium hydroxide – added to improve the separation of bases and organometallics.

If in need of any of these plates, it is also possible to dip or spray these chemicals onto a plain silica gel plate (use the W-type mentioned above), dry, activate, and use. This could be done initially to see how the method works before purchasing a box of them. Often, most manufacturers of TLC plates can be called for sample plates.

## Supports

### Glass

The initial support used for all TLC work was glass. Although standard window glass is readily available and inexpensive, the borosilicate glass used for TLC manufacture has to be carefully selected. The preferred glass used is ‘float’ glass. In this manufacturing method, the molten glass is pushed through an opening onto a bed of molten tin. This produces a sheet of glass with an extremely smooth, level surface. Such a surface insures the best TLC plate will result from its use. This glass is also thinner than the usual window glass, being  $\sim 1.2 \text{ mm}$  thick. This thickness gives a satisfactory thin-layer plate and minimizes the package weight. Before the coating process, the glass is scrupulously cleaned and dried to rid it of any oils



or surface dirt that would cause an imperfect layer to be laid down.

The glass plates come in a number of sizes, and the user chooses the size needed depending on the number of samples and standards to be separated at one time. They are generally made as 20 × 20 cm prepared plates, and then the other sizes are broken down from this larger size. Some glass plates are prescored on the back of the glass, so that the large size can be broken down to smaller sizes with a minimum of effort. Cutting tools for better glass cuts of non-prescored plates are available from hobby or stained glass stores.

### Polymer and Metal

Because of the difficulty of cutting glass plates, the manufacturers came with other supports made from polymers or metals that could be more easily cut. The best know are the plastic (polyterephthalate – thickness, ~200 µm) or aluminum foil (thickness, ~100 µm) supports. As with the glass support mentioned before, the surface characteristics and cleanliness are of equal importance to producing a good thin layer. With plates using these supports, anyone can cut any size from the larger size. In doing this, the layer is moistened a bit with methanol, placed face down on a filter paper, and cut with a straight edge and razor (or X-Acto<sup>®</sup> knife), or rotary cutter. Alternatively, scissors can be used for cutting the plates to size. After cutting, the new edges should be wiped with a paper towel to eliminate fine particles from adhering to the edges of the support. This will prevent crooked solvent fronts.

### Sorbent Particle Size

The speed and resolving power are important in any chromatographic process. The driving force in TLC is capillary action through the fine matrix of the solvent mixture chosen to be best for the separation. If we can reduce the viscosity of the mobile phase, it would improve flow rate. Likewise, if we develop the plate to shorter distances then the development time will be much less. Realizing these factors from chromatographic theory, various particle size ranges shown in **Table 1** were developed over time.

These particle size changes lead to the development of an HPTLC in the mid-1970s. By reducing the average particle size on the thin-layer plate, the analysis time is reduced because a more efficient separation results. Development times are reduced to a few minutes with developing distances of 2–3 cm only. As with any separation on a TLC

plate, the optimum speed and separation are related to the viscosity of the mobile phase, but time savings on the order of three to four times are realized. Another benefit is that the limits of detection are increases because the radial diffusion on the HPTLC plates is reduced.

A summary of the performance differences in the irregular TLC and HPTLC thin-layer plates are given in **Table 3**.

Improvements in thin-layer plate technology over the past 5 years have yielded plates made not with the usual irregular silica gel but of a spherical silica gel. This also improves the speed of analysis by another factor. Some of these thin layers of spherical silica gel have been designed for direct Raman spectroscopy use.

Finally, the silica gel can be made for form as a monolith, or porous honeycomb matrix. This appears as a sheet laid down on the glass plate, so particles of silica gel are not involved. This is a truly unique matrix on which absorption separations can be performed. The layer formed of the monolithic

**Table 3** Differences in TLC and HPTLC plates

	TLC	HPTLC
Average particle size	15 µm	5 µm
Particle distribution	5–20 µm	4–7 µm
Sample size	100 µg–1 mg	100 ng–1 µg
Sample volume	0.5–5 µl	10–200 nl
Optimum spot diameter	3–6 mm	1–1.5 mm
Development distance	10–15 cm	3–7 cm
Development time	10–120 min	3–25 min
Detection limits – standard detection	5 ng	0.5 ng
Detection limits – fluorescence detection	100 pg	10 pg

**Table 4** TLC plate manufacturers

North American/European plate manufacturers	Web URL
Analtech (Wilmington, DE, USA)	www.analtech.com
J.T. Baker (Phillipsburg, NJ, USA)	www.jtbaker.com
Macherey-Nagel (Düren, Germany)	www.mn-net.com
Merck KGaA (Darmstadt, Germany) – available through EMD Chemicals, Inc. (Gibbstown, NJ, USA) in North America	http://chrombook.merck.de/ www.emdchemicals.com
Whatman, Inc. (Maidstone, Kent, UK)	www.whatman.com

Other TLC plate manufacturers may be found around the world, but current publications tend to be found using these manufacturers' plates.

silica is only 0.01 mm thick and can give a very fast, efficient separation. Use of these plates is referred to as 'UTLC'.

Table 4 lists the major manufacturers of thin-layer plates in Europe and North America and their websites. Since product availability changes with usage and demand, their latest offerings are best found on their websites or in their latest catalogs.

See also: **Thin-Layer Chromatography:** Overview; Principles; Instrumentation.

## Further Reading

- Hahn-Deinstrop E (2000) *Applied Thin-Layer Chromatography: Best Practice & Avoidance of Mistakes*. New York: Wiley.
- Poole CF (2003) *The Essence of Chromatography*. New York: Elsevier.
- Sherma J and Fried B (eds.) (1999) *Thin Layer Chromatography – Techniques and Applications*, 4th edn. New York: Dekker.
- Sherma J and Fried B (eds.) (2003) *Handbook of Thin-Layer Chromatography*, 3rd edn. New York: Dekker.

## Method Development

**E Reich and A Blatter**, CAMAG Laboratory, Muttenz, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

This article covers various aspects of development and validation of methods in thin-layer chromatography. A detailed treatment of the underlying theory is given elsewhere in this encyclopedia.

### General Considerations

Thin-layer chromatography (TLC) is an offline process comprising several individual steps, which are independent in time and location. Each step, such as sample application, chromatogram development, detection (including derivatization if necessary), and evaluation is associated with a number of important parameters that can affect the final result of the analysis. The development and validation of a TLC method must not only involve a clear specification of those parameters but also an evaluation of their suitable ranges. For a given analytical task it is usually the separation step that is particularly emphasized during method development; however, the other steps may become equally important if repeatability of the qualitative result or detection limits, accuracy, and precision of quantitation are of importance.

Unlike high-performance liquid chromatography (HPLC) and gas chromatography, where the chromatographic separation generally takes place in an equilibrated system of stationary and mobile phase, TLC usually is a process during which such equilibrium is being approached and in many cases only

partially reached. As a consequence, predictability of the result in TLC is difficult and rather limited. Method development is still a widely empirical process and theoretical models or computer-aided strategies provide satisfactory solutions applicable to 'real' analytical problems only for defined systems with a strictly controlled set of parameters. Nevertheless, from a practical point of view method development in TLC can be quite simple and straight forward, if performed in a structured 'trial and error' approach. The following sections summarize such concepts.

### Defining the Analytical Goal and Experimental Approach

The starting point of any method development is a clear understanding of the analytical goal. Based on information about the nature of the sample, its size, matrix, and number of components some estimations and principal decisions about the experimental setup are made. TLC can analyze many different substances; however, it must be considered that the process includes two drying steps (following sample application and chromatogram development), which may affect the sample. Very volatile samples such as essential oils can partially evaporate from the layer. Qualitative analysis is still possible but reliable quantitation will be difficult. Large biological molecules such as proteins, which have tertiary and quaternary structures, may collapse, become insoluble, and remain at the application position. Light sensitive samples and those sensitive to oxidation require special experimental conditions in order to avoid their decomposition.

The zone capacity (also called separation number) of a chromatographic system is limited, which means that only a finite number of sample components can



silica is only 0.01 mm thick and can give a very fast, efficient separation. Use of these plates is referred to as 'UTLC'.

Table 4 lists the major manufacturers of thin-layer plates in Europe and North America and their websites. Since product availability changes with usage and demand, their latest offerings are best found on their websites or in their latest catalogs.

See also: **Thin-Layer Chromatography:** Overview; Principles; Instrumentation.

## Further Reading

- Hahn-Deinstrop E (2000) *Applied Thin-Layer Chromatography: Best Practice & Avoidance of Mistakes*. New York: Wiley.
- Poole CF (2003) *The Essence of Chromatography*. New York: Elsevier.
- Sherma J and Fried B (eds.) (1999) *Thin Layer Chromatography – Techniques and Applications*, 4th edn. New York: Dekker.
- Sherma J and Fried B (eds.) (2003) *Handbook of Thin-Layer Chromatography*, 3rd edn. New York: Dekker.

## Method Development

**E Reich and A Blatter**, CAMAG Laboratory, Muttenz, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

This article covers various aspects of development and validation of methods in thin-layer chromatography. A detailed treatment of the underlying theory is given elsewhere in this encyclopedia.

### General Considerations

Thin-layer chromatography (TLC) is an offline process comprising several individual steps, which are independent in time and location. Each step, such as sample application, chromatogram development, detection (including derivatization if necessary), and evaluation is associated with a number of important parameters that can affect the final result of the analysis. The development and validation of a TLC method must not only involve a clear specification of those parameters but also an evaluation of their suitable ranges. For a given analytical task it is usually the separation step that is particularly emphasized during method development; however, the other steps may become equally important if repeatability of the qualitative result or detection limits, accuracy, and precision of quantitation are of importance.

Unlike high-performance liquid chromatography (HPLC) and gas chromatography, where the chromatographic separation generally takes place in an equilibrated system of stationary and mobile phase, TLC usually is a process during which such equilibrium is being approached and in many cases only

partially reached. As a consequence, predictability of the result in TLC is difficult and rather limited. Method development is still a widely empirical process and theoretical models or computer-aided strategies provide satisfactory solutions applicable to 'real' analytical problems only for defined systems with a strictly controlled set of parameters. Nevertheless, from a practical point of view method development in TLC can be quite simple and straight forward, if performed in a structured 'trial and error' approach. The following sections summarize such concepts.

### Defining the Analytical Goal and Experimental Approach

The starting point of any method development is a clear understanding of the analytical goal. Based on information about the nature of the sample, its size, matrix, and number of components some estimations and principal decisions about the experimental setup are made. TLC can analyze many different substances; however, it must be considered that the process includes two drying steps (following sample application and chromatogram development), which may affect the sample. Very volatile samples such as essential oils can partially evaporate from the layer. Qualitative analysis is still possible but reliable quantitation will be difficult. Large biological molecules such as proteins, which have tertiary and quaternary structures, may collapse, become insoluble, and remain at the application position. Light sensitive samples and those sensitive to oxidation require special experimental conditions in order to avoid their decomposition.

The zone capacity (also called separation number) of a chromatographic system is limited, which means that only a finite number of sample components can

be baseline resolved. Assuming an optimized chromatographic system, it is generally possible in TLC to achieve baseline resolution for  $\sim 10$  components of a sample during a single development. In certain cases, such as for homologs, this number may increase. In other cases, for instance, where the components fall into extremely different substance classes, the number can decrease. If the sample is more complex, limited resolution might be still acceptable for many qualitative investigations. Fingerprint analysis of herbal drugs, often resulting in a sequence of many colored yet not fully resolved zones, is a typical example. Multiple development such as automated multiple development (AMD) or forced flow techniques such as over-pressured layer chromatography (OPLC) typically offer baseline resolution for  $\sim 25$ – $30$  and in some cases up to  $40$  sample components. If the components cover a wide polarity range, gradient techniques in combination with multiple development (AMD) or forced flow techniques (OPLC) are employed. Still higher zone capacities ( $100$ – $1000$ ) are available through two-dimensional development, where two different separation modes are combined. Although theoretically very appealing this technique is not widely used because only one sample can be analyzed per plate, complex experimental requirements have to be met (HPLC–AMD coupling), and evaluation of the chromatograms is complicated in practice.

An important decision concerns sample preparation, which in many cases can be reduced to a minimum due to the single use of the TLC plate. As long as it does not interfere with the separation the sample matrix does not have to be removed prior to chromatography. Separations can often be setup to either leave the matrix at the application position or move it close to the solvent front. Otherwise general sample treatment such as solid-phase extraction, liquid–liquid extraction must be performed. At this point, also decisions about any prechromatographic derivatization either as part of sample preparation or *in situ* on the plate immediately following sample application are made. Prior to application the sample is dissolved in a suitable solvent. It should be of low polarity when silica gel is used as stationary phase in order to keep the application zone as narrow as possible. The solvent should be of medium to high volatility to ensure proper drying of the application zone prior to chromatogram development, but at the same time avoid volume error due to unwanted evaporation. Solvents such as dimethyl sulfoxide, glycerol, and dimethylformamide or pentane, and diethyl ether are difficult to handle in TLC. For method development samples are typically prepared at an average concentration of  $1 \text{ mg ml}^{-1}$  for each

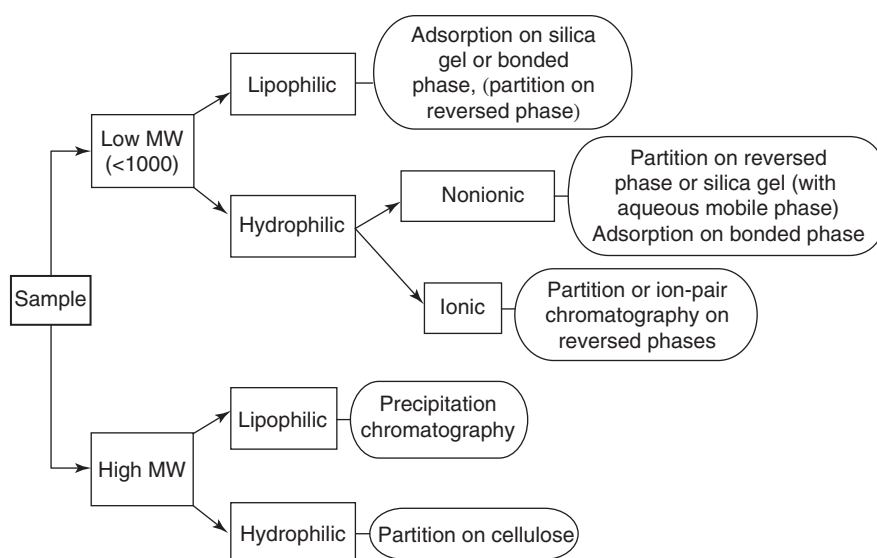
analyte. Application volumes of  $2$ – $5 \mu\text{l}$  as  $8 \text{ mm}$  bands usually ensure good detectability on high-performance thin-layer chromatography (HPTLC) plates.

In most cases methods are developed for single development in a specific chromatographic chamber. Because geometry and mode of operation (degree of saturation) can significantly affect the chromatographic result, the chamber must be selected appropriately. Although such selection generally follows the practical consideration of availability other aspects are also important. Except for initial selection of mobile phase (see below) where special chambers such as the HPTLC VARIO-Chamber (CAMAG) may be used, methods are developed using the same chamber that will later be employed during analysis. While unsaturated chambers generally yield sharper zones, saturated chambers usually give better reproducibility. If the method is intended for analysis of large numbers of samples, horizontal chambers offer the advantage of simultaneously developing HPTLC plates from two opposite edges.

## Selection of the Stationary Phase and the Separation Mode

For TLC a wide range of stationary phases is commercially available and many more materials have been reported in the literature. As in column chromatography separation can be based on different modes of interaction, including adsorption, partition, ion exchange, and complexation. The choices of stationary phases and separation modes are numerous and depend on the type of sample to be analyzed. The more information about the chemical nature of the sample components including molecular weight, functional groups,  $\text{pK}_a$  value, and solubility of all sample components is available, the easier it is to make the proper choice. A flow chart similar to that in **Figure 1** is a suitable tool for linking the separation mode for various types of samples with appropriate stationary phases.

For most samples, which are soluble in organic solvents and of low molecular weight, adsorption chromatography (also called liquid–solid chromatography, LSC) on silica gel, aluminum oxide, or other inorganic oxides is suitable. Separation is based on specific interaction of the functional groups on the surface of the adsorbent with those of the sample. Silica gel is by far the most widely used adsorbent, available in consistent quality and at comparatively low cost. Due to its ionic nature aluminum oxide offers additional selectivity; however, it is more difficult to obtain products with consistent separation properties. Therefore, separations on aluminum



**Figure 1** Selection of separation mode for a sample by thin-layer chromatography.

**Table 1** Typical applications of bonded stationary phases (adapted from C. Poole)

Modification	Functionality on Si	Application
Alkylsiloxane	$-\text{C}_2, \dots, -\text{C}_{18}$	<ul style="list-style-type: none"> <li>RP chromatography of polar water soluble organics</li> <li>Acids and bases after ion suppression or by ion-pair mechanisms</li> <li>Homolog compounds and PAHs</li> </ul>
Cyanopropylsiloxane	$-(\text{CH}_2)_3\text{CN}$	<ul style="list-style-type: none"> <li>NPC similar to low capacity silica</li> <li>RPC similar to chorht chain RP phase (no dipole selectivity)</li> </ul>
Aminopropylsiloxane	$-(\text{CH}_2)_3\text{NH}_2$	<ul style="list-style-type: none"> <li>NPC retention by hydrogen interaction, different from Si</li> <li>IEC – weak anion exchanger in acidic mobile phase</li> </ul>
Spacer bonded propane diol	$-(\text{CH}_2)_3\text{OCH}_2\text{CH}(\text{OH})\text{CH}_2\text{OH}$	<ul style="list-style-type: none"> <li>NPC retention of polar compounds by H-bonds and dipolar interaction</li> <li>Similar retention, different selectivity to short-chain RP phases</li> </ul>

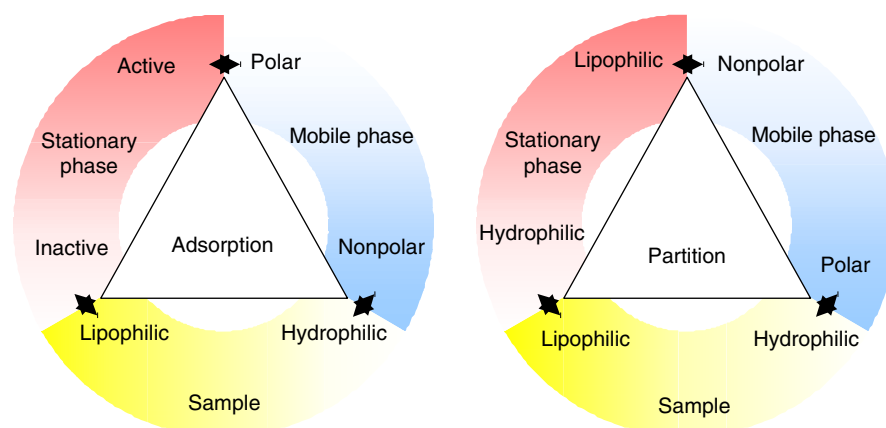
oxide are rather the exception. Chemically bonded phases on silica gel basis having diol, amino, or cyano functionality provide a range of complementary selectivity for separation by LSC and are often considered whenever initial attempts to solve a separation problem on silica gel have failed.

Alternatively samples of low molecular weight, particularly those of very low and very high polarity, can be separated by reversed-phase (RP) chromatography using partitioning processes. Compared to stationary phases for RP-HPLC, TLC material cannot tolerate a high water content of the mobile phase due to the limited wettability. To ensure migration of the mobile phase by capillary action the carbon load on the surface of the silica gel is therefore limited (water-wettable layers). **Table 1** summarizes principal applications of some polar and nonpolar bonded phases.

From a practical point a simple triangle scheme, which goes back to Stahl, is a convenient start for selecting a stationary phase (**Figure 2**). The scheme focuses on the interaction of the sample with both

stationary and mobile phases. For example, a very lipophilic sample requires an active adsorbent and a nonpolar (organic) mobile phase for separation in adsorption mode. If a lipophilic stationary phase such as a  $\text{C}_{18}$  reversed phase is chosen for partition chromatography, the same sample would require a very polar (aqueous) mobile phase.

Whenever possible silica gel will be the first choice as stationary phase during method development. However, silica gels of different brands may exhibit significantly different chromatographic behavior with respect to a particular analyte, mainly due to proprietary aspects of the manufacturing process. Therefore, it is important for routine methods that the continuous availability of the selected silica gel is ensured. The activity of silica gel is affected by humidity. This fact has to be considered during method development, at least to the extent that the relative humidity during chromatography is recorded. If methods are developed for quantitation and if long-term reproducibility is of concern, the TLC plates are typically cleaned prior to use in order to remove any



**Figure 2** Triangle scheme after Stahl. The corners of the triangle connect proper choices for the mobile and stationary phase based on solubility of the sample. (Adapted from Stahl E (1967) *Dünnschicht-Chromatographie – ein Laboratoriumshandbuch*. Berlin: Springer.

contamination from packaging or exposure to lab atmosphere. Cleaning is possible by development with or immersion into polar solvents such as 2-propanol or methanol and subsequent drying and re-equilibration with the relative humidity in the laboratory.

## Selection and Optimization of the Mobile Phase Composition

When a stationary phase has been chosen for a given separation problem the next step of method development is the selection of a suitable mobile phase. The choice of solvents for use as mobile phases in TLC is extremely wide because there is no interference with the detection step. However, from a practical point of view certain limitations apply. The use of toxic, obnoxious, and hazardous solvents should be avoided or at least limited. It is important that only solvents of known purity from reliable sources are used, because certain impurities can significantly affect the chromatographic result. Also, known additives, such as stabilizers in chloroform and ethers or denaturing agents in ethanol, will affect the chromatographic behavior of a solvent. For use as a mobile phase the viscosity and volatility of the liquid must be suitable.

The mobile phase should meet three general requirements: it must be able to dissolve the sample to a certain extent in order to allow its migration; it must be of appropriate 'strength' in order to obtain suitable  $R_F$  values (between 0.3 and 0.8) for the analyte; and it must provide suitable selectivity for separation of the sample components to occur. However, both strength and selectivity are not sole properties of the solvent acting as mobile phase, but also a function of the stationary phase. For example,

water has a high solvent strength on silica gel, but is very weak on a RP layer. Furthermore, the effects of these properties on the actual separation are also dependent on the nature of the sample. Nevertheless, it is common to classify solvents for use as mobile phase with a given stationary phase, such as silica gel, according to their strength and selectivity. The simplest approach is that of an eluotropic series. Solvents are arranged according to their strength, which is either based on empirical results or like  $\varepsilon^\circ$  calculated from physicochemical data such as adsorption energy (Table 2). The principal value of such an arrangement for the practitioner is that it provides guidance for adjusting the  $R_F$  values for the zones of a given chromatogram. A solvent with a higher strength will yield a higher  $R_F$  value for a given substance on a given stationary phase. Isoeluotropic solvents yield comparable  $R_F$  values. Practical selection of the mobile phase based on this concept starts with finding a suitable solvent strength to obtain  $R_F$  values around 0.3 for the substance pair to be separated. With the help of graphical presentations iso-elutotropic solvents or solvent mixtures are then selected to improve separation by changing selectivity of the chromatographic system.

More complex classification systems also involve parameters describing the selectivity of solvents. The most widely used of those is Snyder's solvent selectivity triangle. Based on the model solutes ethanol, dioxane, and nitromethane, solvents are characterized by three selectivity coordinates ( $x_e$ ,  $x_d$ ,  $x_n$ ) describing their interaction with a proton acceptor, a proton donor, and a dipolar solute. When plotted on the surface of a triangle ( $x_e + x_d + x_n = 1$ ) eight clusters, called selectivity groups, can be distinguished. Each solvent is further characterized by a solvent strength parameter  $P'$  (Table 3). Another concept

**Table 2** Properties of some commonly used solvents in TLC

<i>Solvent</i>	<i>Estimated <math>\epsilon^\circ</math> (silica gel)</i>	<i><math>\eta</math> (viscosity (cP))</i>
n-Pentane	0.00	0.23
Hexane	0.01	0.31
Cyclohexane	0.03	1.00
Diisopropyl ether	0.22	0.37
Isopropyl chloride	0.22	0.33
Toluene	0.22	0.59
Benzene	0.25	0.65
Diethyl ether	0.29	0.23
Chloroform	0.31	0.57
Dichloromethane	0.32	0.44
Tetrahydrofuran	0.35	0.47
Methylethyl ketone	0.39	0.44
Acetone	0.43	0.32
Dioxane	0.43	1.54
Ethyl acetate	0.45	0.45
Methyl acetate	0.46	0.37
Dimethyl sulfoxide	0.48	2.24
Nitromethane	0.49	0.67
Acetonitrile	0.50	0.37
Propanol, n- and <i>i</i> -	0.63	2.3
Ethanol	0.68	1.20
Methanol	0.73	0.60
Ethylene glycol	0.86	19.9
Acetic acid	>> Large	1.26
Water	>> Large	1.00

Adapted from J. Miller.

uses the solvatochromic parameters  $\pi_1^*$  as measure of polarizability and dipolarity,  $\alpha_1$  for hydrogen bond acidity, and  $\beta_1$  for hydrogen bond basicity to classify solvents.

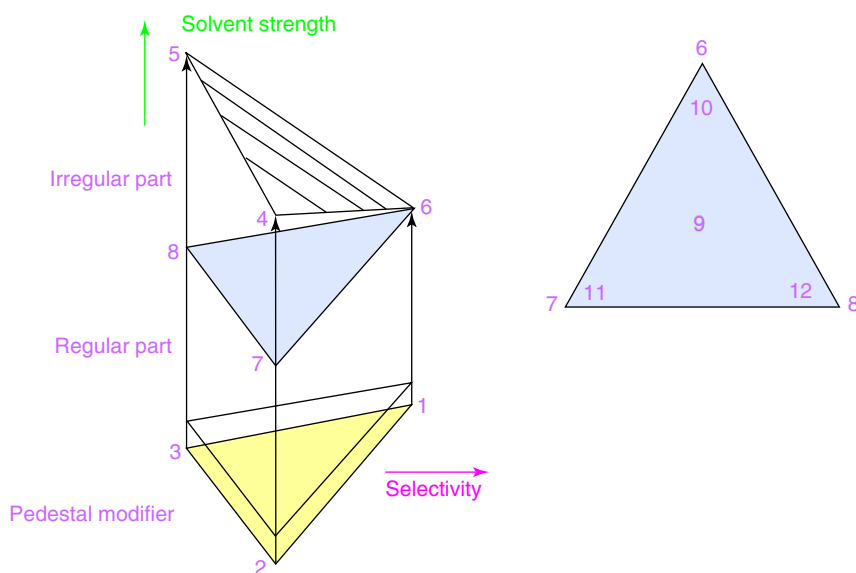
Although evaluation of properties of pure solvents is fairly easy, difficulties arise for accurate description of mixtures, because many parameters are neither independent nor additive, and in many cases not even available.

Regardless of which theoretical basis is adopted, practical approaches have to include a number of different solvents to cover a broad selectivity range. These solvents are initially evaluated and then combined in binary, ternary, or quaternary mixtures, if necessary including 'modifiers' for adjusting the solvent strength or to improve and optimize selectivity. Several concepts have been developed for such optimization including window diagrams, sequential simplex methods, overlapping resolution mapping, principal component analysis, and even numerical taxonomy. Although of great theoretical value most of these strategies lack general applicability for routine use because separation in TLC often relies on mixed modes of interaction and many parameters of the chromatographic system cannot be controlled.

**Table 3** Selectivity groups, solvent strength, and selectivity parameters based on Snyder's selectivity triangle (adapted from C. Poole)

<i>Solvent</i>	<i>Selectivity group</i>	<i>Solvent strength</i>		<i>Solvent selectivity</i>		
		<i>P'</i>	<i>S<sub>T</sub></i>	<i>x<sub>e</sub></i>	<i>x<sub>d</sub></i>	<i>x<sub>n</sub></i>
n-Butyl ether	I	2.1		0.44	0.18	0.38
Diisopropyl ether		2.4		0.48	0.14	0.38
Methyl <i>t</i> -butyl ether		2.7				
Diethyl ether		2.8		0.53	0.13	0.34
n-Butanol	II	3.9		0.59	0.19	0.25
2-Propanol		3.9		0.55	0.19	0.27
Ethanol		4.3	3.6	0.52	0.19	0.29
Methanol		5.1	3.0	0.48	0.22	0.31
Tetrahydrofuran	III	4.0	4.4	0.38	0.20	0.42
Dimethylformamide		6.4		0.39	0.21	0.40
Acetic acid	IV	6.0		0.39	0.31	0.30
Formamide		9.6		0.38	0.33	0.30
Dichloromethane	V	4.3		0.27	0.33	0.40
Ethyl acetate	VI	4.4		0.34	0.23	0.43
Methylethyl ketone		4.7		0.35	0.22	0.43
Dioxane		4.8	3.5	0.36	0.24	0.40
Acetone		5.1	3.4	0.35	0.23	0.42
Acetonitrile		5.8	3.1	0.31	0.27	0.42
Toluene	VII	2.4		0.25	0.28	0.47
Benzene		2.7		0.23	0.32	0.45
Chloroform	VIII	4.3		0.31	0.35	0.34
Water		10.2	0.0	0.37	0.37	0.25





**Figure 3** PRISMA model after Nyiredy *et al.* The corners of the triangular base (1, 2, 3) represent solvents of different selectivities and the height at each corner (4, 5, 6) is a measure of the  $P'$  value of the individual solvents. Stronger solvents are reduced by mixing them with hexane ( $P' = 0$ ) to arrive at the strength of the weakest solvent (7, 8). The selectivity of the mobile phase system is varied by changing the proportions of the adjusted 'corner' solvents (6, 7, 8), beginning with the center point (9 = 3, 3, 3) and then moving to points close to the corners (10 = 3, 1, 1; 11 = 1, 3, 1; 12 = 1, 1, 3).

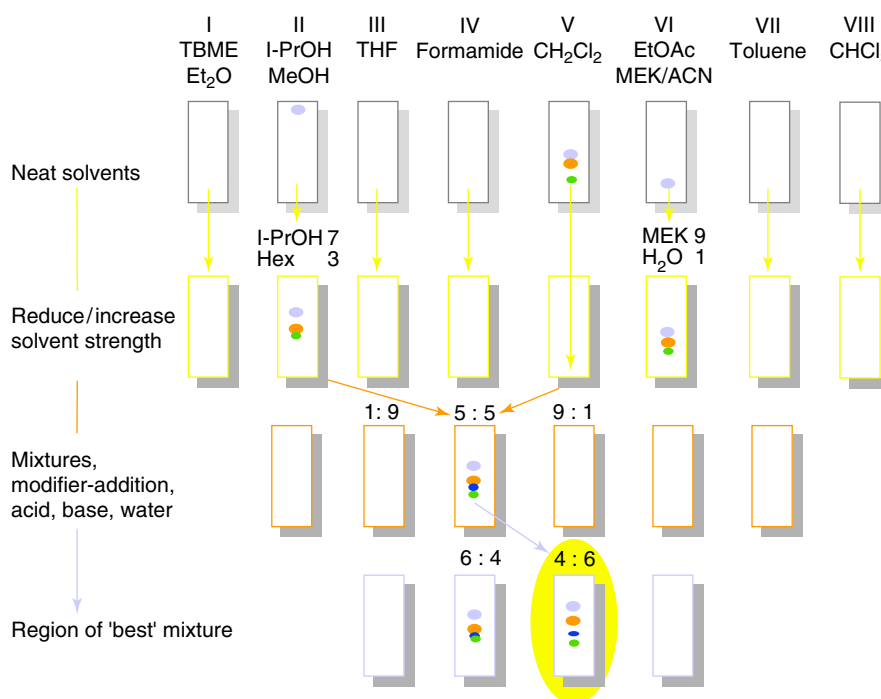
The PRISMA model (Figure 3) developed by Nyiredy and co-workers combines a theoretical basis derived from the selectivity triangle with a guided trial and error approach. In the first step, 8–10 neat solvents, representing all selectivity groups, are selected. Of those in the second step the three solvents offering best selectivity are taken for construction of the prism. The corners of the triangular base (1, 2, and 3 in Figure 3) represent the different selectivities and the height at each corner (4, 5, and 6) is a measure of the  $P'$  value of the individual solvents. For samples of low and medium polarities the regular portion of the prism is used. In any triangle congruent to the base, all solvent mixtures are equi-elutotropic, meaning that the average  $R_F$  values for the sample components obtained with those mixtures will be about the same. In the third step, the stronger solvents are reduced by mixing them with a certain amount of hexane ( $P' = 0$ ) to arrive at the strength of the weakest solvent (7 and 8). For samples of very low polarity the strength of all three solvents may need to be reduced. In the fourth step, selectivity of the mobile phase system is varied to yield satisfactory resolution by changing the proportions of the adjusted 'corner' solvent, beginning with the center point and then moving to points close to the corners. For very polar samples optimization can be performed in the irregular top portion of the prism. In this region, usually only small changes are made in mobile phase

composition because they will also affect the average  $R_F$  value of the sample. If during the first step the solvent strength of the neat solvents is not sufficient to yield  $R_F$  values between 0.2 and 0.8, modifiers such as acids, bases, or water can be used in small proportions to deactivate the stationary phase. The pedestal of the prism represents these modifiers. If no satisfactory separation is obtained during this optimization process a new stationary phase or different solvents have to be chosen. For separations in RP systems the PRISMA model can also be employed. Instead of the  $P'$  value the  $S_T$  value is used and solvent strength is adjusted with water ( $S_T = 0$ ; see Table 3).

There are also other trial and error approaches, the simplest of which is the so-called spot test. The sample is applied as several spots on a TLC plate. Then specified volumes of different solvents are applied to the centers of the spotted samples. The resulting circular chromatograms can give preliminary information about solvent strength and selectivity required for separation of the sample. With modern instruments for sample application this test can be automated. However, actual optimization of the mobile phase must still be performed in a suitable chromatographic chamber.

CAMAG has proposed a practical scheme (Figure 4) for guidance during method development, which is derived from the PRISMA model but omits any





**Figure 4** CAMAG optimization scheme. For details see text. Note: The solvents selected in the scheme serve as examples and may be different for other separation problems.

calculation. Method development can be performed in saturated, unsaturated, or sandwich chambers. Initially, silica gel is selected as the stationary phase. At level 1 a set of 8–12 neat solvents is evaluated. Looking at the  $R_F$  values of the sample in the chromatograms three results are possible. In case (A),  $R_F$  values are between 0.2 and 0.8. These solvents are considered further in level 3. In case (B),  $R_F$  values are below 0.2. In case (C),  $R_F$  values are above 0.8. These solvents are taken to level 2, where either adding a small proportion of polar modifier such as water, acid, or base increases their strength, or the addition of hexane reduces it. On level 3, two solvents of suitable selectivity from level 1 (case A) and/or level 2 are combined in order to achieve improved separation. At this point small portions of acid, base, or water can be added to reduce tailing or improve the shape of the separated zones. If no improvement results two other solvents are combined. In level 4, the proportions of the selected solvents are slightly changed to optimize the result. If no satisfactory result is obtained during the process, another stationary phase has to be used. The approach works similarly and equally well for polar bonded phases. On reversed phases water is used to adjust solvent strength.

According to Nyiredy mobile phases for forced flow separations such as OPLC can also be selected

and optimized using the PRISMA model. In this process also one solvent is usually chosen, in which the sample does not migrate. This solvent can be used in a prerun to eliminate the so-called front of total wetness caused by undissolved gasses in the sealed layer.

For separations using the AMD technique not only solvent strength and selectivity can be varied but also the type of the gradient. Development of a method begins with a standard gradient from a strong solvent such as methanol, acetonitrile, or acetone over a medium polar base solvent such as dichloromethane, chloroform, or an ether to a nonpolar solvent like hexane. Fifteen to twenty developments of increasing distance typically reach 60 mm for the last step. Because the solvent composition at each point of the chromatogram is known the results can easily be optimized. The  $R_F$  values of the separated compounds are lowered by eliminating the steps containing large portions of polar solvent or by making the gradient steeper.  $R_F$  values are increased by eliminating the steps containing hexane, making the gradient less steep, or for very polar samples adding polar modifiers such as water, acid, or base in order to deactivate the stationary phase. Distances between separated zones are affected by changing the slope of the gradient. Exchanging the base solvent modifies selectivity of the separation.

## Optimizing the Detection

Throughout the entire method development process detection is an important step, particularly if samples of unknown composition are to be analyzed. To ensure that all noncolored sample components are detected in a qualitative analysis, the developed plate, which typically contains a so-called fluorescence indicator, is evaluated visually under ultraviolet (UV) at 254 nm. Substances absorbing this wavelength are seen as dark zones on fluorescing background. Evaluation under UV 366 nm detects all components, which are excited by that wavelength to fluoresce. Densitometric evaluation of the chromatogram at multiple wavelengths will detect all substances, which absorb between 190 and 800 nm. Substances without chromophors may be rendered detectable by employing various derivatizing agents. During the selection/optimization of the mobile phase universal reagents are often used for visualization. Examples include mixtures of sulfuric acid and anisaldehyde in methanol, copper(II) sulfate, and phosphoric acid in water, or phosphomolybdic acid. If the sample components are known also specific derivatization can be used to visualize the chromatogram. The visual impression often gives additional information about the quality of separation and the homogeneity of results across the plate.

When the separation is satisfactory the detection step has to be optimized. For densitometric evaluation in absorbance mode without derivatization this optimization includes the selection of the proper scanning wavelength based on spectral or multi-wavelength evaluation and choice of mechanical parameters such as scanning speed and slit dimension. For fluorescence measurements also the excitation wavelength and an appropriate cutoff filter are selected. If detection includes a derivatization step additional parameters must be optimized for reproducible results. Derivatization by immersing the plate into the reagent is a reproducible process. Parameters for optimization are reagent concentration, type of the solvent, and immersion time. Usually derivatization is completed by a heating step, for which temperature and duration have to be selected. In many cases the intensity and color of zones changes with temperature and time. Also, the fluorescence of separated substances or their derivatives is usually not constant over time. It goes through a maximum at some point and then decreases, but it can be stabilized or enhanced by covering the plate with viscous liquids such as paraffin or poly(ethylene glycol) by spraying or dipping. This prevents the excited sample molecules from relaxation and eliminates fluorescence quenching by oxygen.

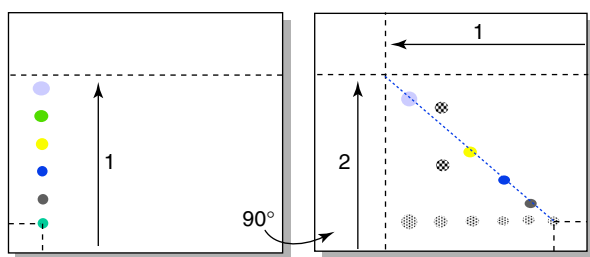
When developing a quantitative TLC method also the amounts of the separated sample components have to be adjusted. It is desirable to work within a linear range of the corresponding calibration curve. Limits of detection (LOD) and quantitation (LOQ) have to be established for each component. These limits and the working range of the method determine the concentration at which the sample has to be prepared and the volume that has to be applied onto the plate.

## Method Validation

Prior to its use a method has to be validated. Validation is the formal proof that the method is suitable for the intended purpose. This requires that all steps and parameters of the method have been clearly specified in a written method description, any necessary equipment was qualified, and acceptance criteria for each validation point have been agreed upon. For quantitative methods the International Conference on Harmonization (ICH) has issued specific guidelines for setting up a validation protocol and for parameters that have to be validated for different applications. These include specificity, accuracy, precision, LOD, LOQ, linearity, and range as well as robustness. The only required validation parameter for qualitative methods is specificity.

For both qualitative and quantitative TLC methods it is useful to perform certain prevalidation experiments. This can be done during method development. Because TLC is an offline process and the TLC plate is an open system, which can be affected by various environmental factors, it is necessary to establish the stability of the analyte in solution, on the plate and during chromatography in order to avoid any artifacts. Furthermore, the stability of the mobile phase must be ensured. Stability during chromatography is investigated by a two-dimensional experiment. The sample is applied in the lower left corner of the TLC plate. After development the plate is dried, turned 90° to the left, and developed with fresh mobile phase according to the specification of the method. The sample is stable during chromatography if after the second development all sample components are on the diagonal connecting the application position with the intersection of the mobile phase fronts (Figure 5).

Because the TLC plate is used only once, intermediate precision (reproducibility of the result from plate to plate) is particularly important for methods intended for routine use. To maximize precision the method must be robust. During validation or during method development the effects of the following



**Figure 5** Evaluation of stability of the analyte in the chromatographic system by two-dimensional chromatography. Development in the first direction separates the mixture into its components. For each stable component the second development gives the same  $R_f$  value as the first development. Any substance that is located off the diagonal is formed during chromatography. For further details see text.

parameters are typically assessed: relative humidity, performance of plates from different batches and manufacturers, developing distance, times for chamber saturation, heating times and temperature during derivatization, and waiting times between the individual TLC steps.

See also: **Thin-Layer Chromatography:** Overview; Principles; Plate Technology; Instrumentation.

## Further Reading

- Ferenczi-Fodor K, Végh Z, Nagy-Turák A, Renger B, and Zeller M (2001) Validation and quality assurance of planar chromatographic procedures in pharmaceutical analysis. *Journal of AOAC International* 84: 1258–1264.
- Geiss F (1987) *Fundamentals of Thin Layer Chromatography*. Heidelberg: Alfred Hüthig Verlag.
- Jork H, Funk W, Fischer W, and Wimmer H (1990) *Thin Layer Chromatography, Volume 1a, Physical and Chemical Detection Methods*. Weinheim: VCH Verlagsgesellschaft mbH.
- Koll K, Reich E, Blatter A, and Veit M (2003) Validation of standardized high-performance thin-layer chromatographic methods for quality control and stability testing of herbals. *Journal of AOAC International* 86(5): 909–915.
- Nyiredy S, Meier B, Erdelmeier CAJ, and Sticher O (1985) 'PRISMA': a geometrical design for solvent optimization in HPLC. *HRC & CC* 8: 186–188.
- Poole CF and Dias NC (2000) Practitioner's guide to method development in thin-layer chromatography. *Journal of Chromatography A* 892: 123–142.
- Sherma J and Fried B (eds.) (2003) *Handbook of Thin-Layer Chromatography*, 3rd edn. New York: Marcel Dekker.
- Snyder LR (1978) Classification of the solvent properties of common liquids. *Journal of Chromatographic Sciences* 16: 223–234.

## Instrumentation

**E Reich and A Blatter**, CAMAG Laboratory, Muttenz, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

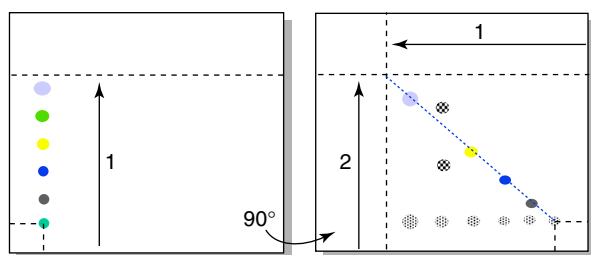
## Introduction

Thin-layer chromatography (TLC) is considered a simple technique and it can be successfully performed without sophisticated instrumentation. However, the quality and reproducibility of results can be significantly improved if instrumentation is utilized. To unlock the full potential of modern high-performance TLC (HPTLC) and to automate its individual steps, state-of-the-art equipment is a necessity. This article describes modern instrumentation for all steps of the TLC process in its functionality.

## Sample Application

Devices for application have to ensure proper positioning of the sample on the TLC plate and precise

dosage of specified sample volumes. Furthermore, any damage to the layer must be avoided. Modern instruments support positioning of samples in  $x$ - and  $y$ -directions either in predefined or freely selectable increments. While manual devices generally utilize a ruler, automated sample applicators include software-controlled mechanics to position the sample reproducibly. For manual sample dosage, disposable fixed volume capillaries, graduated micropipettes, or microsyringes are used. Automated instruments use motor-driven syringes. There are two principal ways of transferring the sample onto the layer, contact spotting and spray-on application. During contact spotting the sample is transferred into the layer by capillary action, when the capillary or needle touches the surface. The duration of contact and the dispensed sample volume determine the size of the applied spot. If the sample is dissolved in a strong (polar) solvent and large volumes are applied, circular chromatography will take place during application resulting in poorly separated and distorted



**Figure 5** Evaluation of stability of the analyte in the chromatographic system by two-dimensional chromatography. Development in the first direction separates the mixture into its components. For each stable component the second development gives the same  $R_f$  value as the first development. Any substance that is located off the diagonal is formed during chromatography. For further details see text.

parameters are typically assessed: relative humidity, performance of plates from different batches and manufacturers, developing distance, times for chamber saturation, heating times and temperature during derivatization, and waiting times between the individual TLC steps.

See also: **Thin-Layer Chromatography:** Overview; Principles; Plate Technology; Instrumentation.

## Further Reading

- Ferenczi-Fodor K, Végh Z, Nagy-Turák A, Renger B, and Zeller M (2001) Validation and quality assurance of planar chromatographic procedures in pharmaceutical analysis. *Journal of AOAC International* 84: 1258–1264.
- Geiss F (1987) *Fundamentals of Thin Layer Chromatography*. Heidelberg: Alfred Hüthig Verlag.
- Jork H, Funk W, Fischer W, and Wimmer H (1990) *Thin Layer Chromatography, Volume 1a, Physical and Chemical Detection Methods*. Weinheim: VCH Verlagsgesellschaft mbH.
- Koll K, Reich E, Blatter A, and Veit M (2003) Validation of standardized high-performance thin-layer chromatographic methods for quality control and stability testing of herbals. *Journal of AOAC International* 86(5): 909–915.
- Nyiredy S, Meier B, Erdelmeier CAJ, and Sticher O (1985) 'PRISMA': a geometrical design for solvent optimization in HPLC. *HRC & CC* 8: 186–188.
- Poole CF and Dias NC (2000) Practitioner's guide to method development in thin-layer chromatography. *Journal of Chromatography A* 892: 123–142.
- Sherma J and Fried B (eds.) (2003) *Handbook of Thin-Layer Chromatography*, 3rd edn. New York: Marcel Dekker.
- Snyder LR (1978) Classification of the solvent properties of common liquids. *Journal of Chromatographic Sciences* 16: 223–234.

## Instrumentation

**E Reich and A Blatter**, CAMAG Laboratory, Muttenz, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Thin-layer chromatography (TLC) is considered a simple technique and it can be successfully performed without sophisticated instrumentation. However, the quality and reproducibility of results can be significantly improved if instrumentation is utilized. To unlock the full potential of modern high-performance TLC (HPTLC) and to automate its individual steps, state-of-the-art equipment is a necessity. This article describes modern instrumentation for all steps of the TLC process in its functionality.

## Sample Application

Devices for application have to ensure proper positioning of the sample on the TLC plate and precise

dosage of specified sample volumes. Furthermore, any damage to the layer must be avoided. Modern instruments support positioning of samples in  $x$ - and  $y$ -directions either in predefined or freely selectable increments. While manual devices generally utilize a ruler, automated sample applicators include software-controlled mechanics to position the sample reproducibly. For manual sample dosage, disposable fixed volume capillaries, graduated micropipettes, or microsyringes are used. Automated instruments use motor-driven syringes. There are two principal ways of transferring the sample onto the layer, contact spotting and spray-on application. During contact spotting the sample is transferred into the layer by capillary action, when the capillary or needle touches the surface. The duration of contact and the dispensed sample volume determine the size of the applied spot. If the sample is dissolved in a strong (polar) solvent and large volumes are applied, circular chromatography will take place during application resulting in poorly separated and distorted



**Figure 1** Nanomat 4 (photograph courtesy of CAMAG.)

chromatogram zones. This has great consequences for the achievable resolution during development of the plate. The spray-on technique overcomes these problems. The sample is atomized with the help of compressed air or inert gas when it leaves the tip of the applicator's syringe. This way a sharply focused starting zone can be achieved. If the spray nozzle is moved relative to the TLC plate during application the sample is sprayed on as a narrow band. By adjusting the dosage speed, wetting of the stationary phase during application can be avoided, thus eliminating any undesired chromatography.

The Nanomat (CAMAG, Switzerland) (Figure 1) is a simple instrument for manual sample application as spots. Using capillaries, 0.5, 1, 2, or 5  $\mu$ l volumes can be applied as spots with distance in multiples of 5 mm. The application position is also selectable in the y-direction. With a capillary holder a disposable capillary is obtained from a dispenser and filled by immersing the tip into the sample. After placing the holder against the applicator head, where it is held by magnetic force, the sample is applied by gently pushing down the applicator head until the capillary makes contact with the layer.

The TLS100 (Desaga, Germany) is designed for automated spot application from a motor-driven syringe. The instrument can be programmed via keypad to apply up to 30 samples and four standards on up to six TLC/HPTLC plates of 20  $\times$  10 cm. The instrument can also generate bands by dividing the specified sample volume into small spots applied side by side. A similar device is the TLC AutoSpotter (Analtech, USA), which can be fitted with up to 18 syringes for simultaneous spot application of multiple samples.

The first semiautomatic instrument utilizing all advantages of the spray-on technique was the (Linomat 5) (Figure 2). Today's latest model is software controlled. Only the y-position of the application is set manually and the user loads the sample into the



**Figure 2** Linomat 5 (photograph courtesy of CAMAG).

syringe. Sample size is adjustable from 100 nl to 2 ml. The selectable band length is 0–195 mm length.

A fully automatic, software-controlled application device is the AS30 (Desaga, Germany). When combined with a regular autosampler up to 30 samples can be applied as spots or bands. The ATS4 (CAMAG, Switzerland) is the most sophisticated and versatile system for sample application. Fully software-controlled, up to 66 samples from vials or 96 samples from well plates can be applied. Volumes and any x- or y-position of the TLC plate are freely selectable for either contact spotting or spray-on application. Large sample volumes containing a high percentage of matrix can also be applied as rectangles for clean-up by predevelopment.

## Chromatogram Development

Although TLC plates can generally be developed in three geometrical modes, linear, circular (radial), and anticircular, today only linear development is of practical relevance. Due to the presence of a gas phase in addition to stationary and mobile phases, TLC differs from all other chromatographic techniques. For ensuring reproducible chromatographic results this gas phase must be carefully controlled because it affects the separation significantly. Chamber form and chamber saturation are playing a dominant role in this regard. Even though there are generally neither good nor bad chambers it must be



expected that all chambers will give somewhat different results.

### Conventional Developing Chambers

The traditional flat-bottom chamber is readily available and comes in many shapes (cylindrical or rectangular) and sizes to fit almost any plate format. The typical 'chromatographic tank' is made of glass and has a rectangular base. The chamber can easily be saturated when one or more walls are lined with filter paper, which is wetted with developing solvent. The flat-bottom chamber has two major disadvantages, the high solvent consumption and the difficulty of securely placing the TLC plate into it. Nevertheless, flat-bottom chambers are widely used.

Twin Trough Chambers (CAMAG, Switzerland) (Figure 3) are also very popular and represent an economic and very flexible alternative. Available in sizes to fit  $10 \times 10$  cm,  $20 \times 10$  cm, and  $20 \times 20$  cm plates these chambers require very little solvent. For example, only 5 ml of developing solvent are needed to develop a  $10 \times 10$  cm HPTLC plate in unsaturated mode. In this mode only the front trough is charged with solvent and development is started immediately. In the saturated mode both troughs are charged and the rear trough is fitted with filter paper, which was wetted with solvent. In a Twin Trough Chamber the TLC plate can be easily preconditioned. While positioned in the empty front trough, having the stationary phase face the inside of the chamber, the plate is exposed to vapor from a conditioning solvent such as acid, base, solution affecting humidity, or developing solvent. Following the conditioning step the mobile phase is carefully introduced into the front trough and chromatography is started.



**Figure 3** Twin trough chambers (photograph courtesy of CAMAG).

### Special Chambers

Several special chambers are on the market, which allow development of TLC plates in a horizontal mode. The Horizontal Separation Chamber (Desaga, Germany) is available for  $5 \times 5$  cm and  $10 \times 10$  cm plates. The Horizontal Developing Chamber (CAMAG, Switzerland) (HDC) (Figure 4) offers additional flexibility. Available in two formats, for  $10 \times 10$  or  $20 \times 10$  cm HPTLC plates, the device allows simultaneous development from opposite edges of a plate. This way up to 72 samples applied as spots are analyzed in parallel. Both horizontal chambers can be operated in saturated, unsaturated, and sandwich modes. For the latter a counter-plate is put in place. The central compartment of the chamber can be charged with a suitable liquid for preconditioning a plate or operation in saturated mode.

A useful tool for method development and optimization of chromatographic parameters is the VARIO Chamber (CAMAG, Switzerland). On  $10 \times 10$  cm HPTLC plates, which have been previously scored to form six parallel tracks, up to six different solvents or six different conditions can be investigated. The chamber can also be operated in saturated, unsaturated, or sandwich mode, thus enabling convenient transfer of the optimized separation conditions to a HDC.

The CycloGraph system (Analtech, USA) (Figure 5) is a centrifugally accelerated device for performing preparative chromatographic separations. The device spins a layer of adsorbent material coated as a flat ring on a glass backing. A solvent pump is used to apply the sample and mobile phase to the center of the spinning adsorbent ring. The centrifugal force accelerates the flow of the mobile phase through the adsorbent, separating the sample components as circular bands. The mobile phase elutes continuously into a collection channel inside the body of the instrument.



**Figure 4** Horizontal developing chamber (photograph courtesy of CAMAG.)





**Figure 5** CycloGraph (photograph courtesy of Analtech, used by permission).

Other special devices, which are of historical interest, include a special chamber for short bed continuous development. In this technique, the plate extends through a slit out of the actual chamber allowing the developing solvent to evaporate. Separation of very similar compounds can thus be achieved at low  $R_f$  values. The U-chamber according to Kaiser was a special device for circular and anticircular HPTLC.

### Automatic Chambers

The primary purpose of automatic chambers is ensuring highest possible reproducibility of the chromatographic result in routine analysis. Therefore, the instrument controls most parameters of the developing step. The Automatic Developing Chamber (CAMAG, Switzerland) (ADC) is designed for classical one-step isocratic development in tank or sandwich configuration. Through a charge-coupled device (CCD) element, the ADC controls the migration distance of the mobile phase. Prior to start of development the TLC plate can be preconditioned through the gas phase. When development is complete the systems can dry the plate automatically. A very similar instrument is the TLC-MAT (Desaga, Germany).

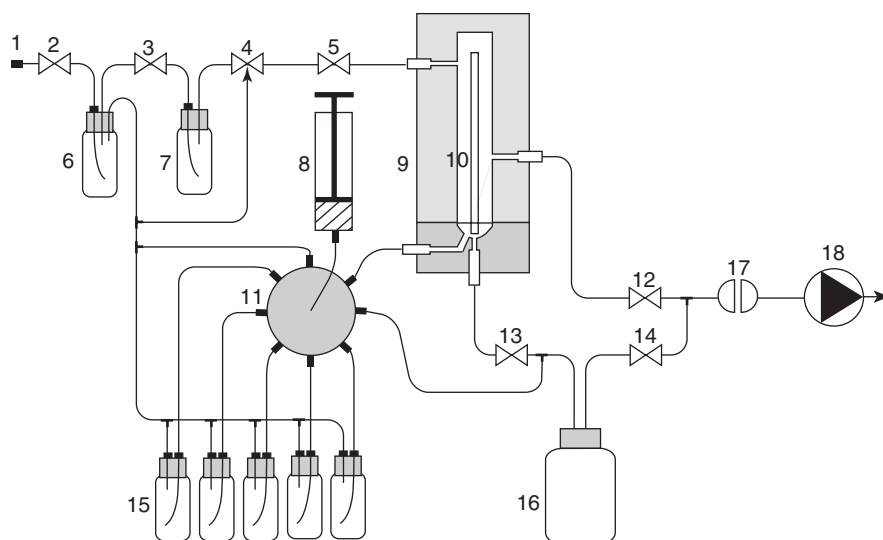
### Automated Multiple Development

A technique that can achieve the maximum attainable resolution in TLC on a given separation distance without forced flow is automated multiple development (AMD). This step-gradient technique was developed by Burger. With respect to peak capacity the technique can be compared to HPLC, but it still maintains all benefits of planar chromatography. The heart of the instrument is a specially designed vacuum-tight chamber. Following sample application

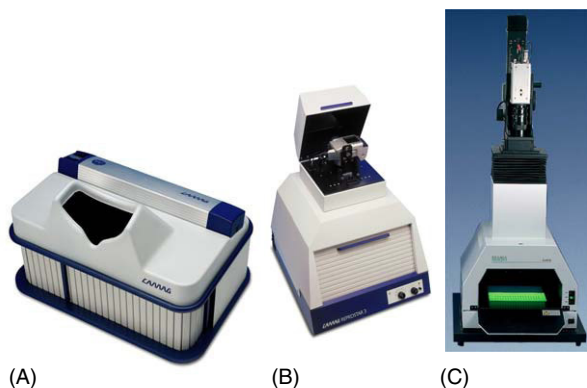
the HPTLC plate of  $20 \times 10$  cm is precisely positioned in the chamber. The mobile phase for each development is prepared automatically by mixing appropriate portions of solvent from up to five different reservoirs. The mobile phase is then introduced into the chamber. Chromatography begins and is monitored by a CCD element, which stops the run if the selected developing distance has been reached. Vacuum from a pump is connected, remaining solvent withdrawn from the chamber, and the plate completely dried. While vacuum is released the plate can be preconditioned via the gas phase, leaving the system ready for the next development step. An AMD gradient always starts over a short distance (2–3 mm) with a mobile phase of high elution power to ensure that the sample, which has been applied as a narrow band, is completely removed from the application position and at the same time further concentrated. Successive runs will be performed with decreasing solvent strength but increasing developing distance. Development of a method for an unknown sample typically begins with a so-called universal or scouting gradient which over 20 to 25 steps and 60 mm developing distance changes from methanol over dichloromethane to hexane. Separation of samples is primarily based on polarity. However, by changing the selectivity of the mobile phase components resolution can easily be changed. Changing the slope of the gradient will affect the distance between zones. One of the advantages of ADM is that desired changes in a given chromatogram can be predictably performed. AMD solvents must be especially pure in order to avoid accumulation of impurities during multiple developments. Because drying of the plate between the individual steps is performed by vacuum AMD solvents must have a sufficient volatility. On the other hand, samples separated by AMD must not be volatile. The AMD 2 (CAMAG, Switzerland) instrument (see scheme **Figure 6**) is fully computer controlled and integrated in a GMP-compliant software. All gradient parameters are conveniently programmed in a spreadsheet. AMD is a highly reproducible but somewhat time-consuming technique. Typical fields of application include analysis of pesticides, lipids, and screening for biological activity.

### Visualization, Detection, and Documentation

The possibility of generating a visual image of the chromatographic result is one of the principal advantages of TCL. It is a common practice to first 'look' at the result before the plate is subjected to further evaluation. Colored substances can be seen



**Figure 6** AMD 2 with a syringe (8) via the motorized valve (1) the selected portions of solvent are taken from the storage bottles (15) and delivered to the chamber (9). When the mobile phase has reached the programmed height on the plate (10) the chamber is drained into the waste bottle (16) and the plate is dried by vacuum from a pump (18). Vacuum is released through conditioning bottles (6, 7) allowing adjustments to the gas phase in the chamber. (photograph courtesy of CAMAG.)



**Figure 7** Instruments for visualization of plates. (A) UV cabinet, (B) documentation system with digital camera, and (C) documentation system with video camera (photograph courtesy of CAMAG and Desaga).

directly. Substances that can be excited to fluoresce by ultraviolet (UV) light can be visualized under a lamp emitting UV at 366 nm. Quenching of a so-called UV indicator can visualize substances that absorb UV light of 254 nm. The UV indicator is excited by UV at 254 nm and emits either green or blue light. Quenching zones of UV absorbing substances are seen as dark zones on fluorescing background when the plate is viewed under the UV lamp. Several UV lamps (Figure 7A) and viewing cabinets are on the market, combining UV 254 nm and UV 366 nm light in reflectance as well as white light in transmission, reflectance, and mixed illumination modes. Substances that are not seen in any of these modes have to be derivatized or otherwise visualized for evaluation and documentation.

For the purpose of documentation modern viewing cabinets such as the Reprostar3 (CAMAG, Switzerland), the VD30 (Desaga, Germany), or the ChromaDoc-it (Analtech, USA) can be fitted with a video or digital camera (Figures 7B and 7C). This software-controlled setup allows instant, reliable, and reproducible capture of electronic images, which are durable, can be conveniently annotated and exchanged, and allow quantitative densitometric evaluation. Documentation in white light is also possible with flatbed scanners. The traditional way of documenting the chromatographic result by photography cannot compete with the advantages offered by electronic systems.

### Derivatization

Prechromatographic derivatization is a useful tool to change the chromatographic behavior of the sample or some of its components in order to increase their stability, improve separation, or reduce matrix effects. It is either performed as part of sample preparation or following sample application *in situ* on the TLC plate. Postchromatographic derivatization is a necessary step in TLC if a sample component is otherwise undetectable and if selectivity or sensitivity of detection can be improved.

**Instruments** From a practical point of view, the principal requirements for derivatization are homogeneity of reagent transfer across the TLC plate and repeatability of the results from plate to plate. Ideally, the amount of reagent involved in derivatization should be controlled as well as the time and temperature of

any subsequent heating step required for completion of the reaction. For qualitative investigations these requirements are typically less stringent.

A very uniform derivatization is possible (yet not generally applicable) if the reagent is transferred via the gas phase using a twin trough chamber or a special conditioning chamber. In the case of a twin trough chamber the developed and solvent-free TLC plate is positioned in one trough, while the other trough is either charged with the reagent (for example, iodine) or with a mixture of chemicals generating the reagent *in situ* (for example, permanganate and hydrochloric acid to generate chlorine).

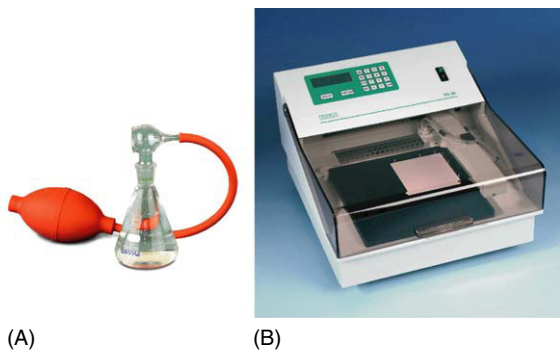
The majority of reagents are used in a solution the TLC plate can either be sprayed with or dipped into. Spraying uses comparatively small volumes of reagent and is very flexible. It is also often the method of choice, if a sequence of reagents has to be applied. The principal disadvantage is the difficulty of ensuring homogeneity and reproducibility. Another problem is the generation of fumes and aerosols, which for safety require spraying to be performed in a hood or special spray cabinet. There are many different reagent sprayers available that utilize the atomizer principle (Figure 8A). They are operated manually with a rubber bulb pump, compressed air, or electropneumatically. The Chromajet (Desaga, Germany) (Figure 8B) is a sophisticated fully automatic spraying device that can solve all problems associated with manual spraying. The exact amount of reagent applied to each of up to 30 defined areas of the plate can be specified. A built-in exhaust fan removes the fumes generated.

Derivatization by immersion offers several advantages. Controlling the concentration of the reagent solution and the immersion time allows increased reproducibility. The coverage of the plate is always uniform and the consumption of reagent per plate is low if the reagent is used repeatedly. Generally, solutions for immersion are less concentrated (only ~20%)

than those for spraying. No fumes are generated and the exposure of the operator to chemicals is limited. The principal disadvantage of immersion is the potential for diffusion of separated zones and even elution from the plate during immersion, if the sample components are readily soluble in the solvent of the reagent. A typical immersion device is shown in Figure 9. It allows selection of immersion speed and immersion time. The plate is smoothly lowered into the reagent thus eliminating any 'tide marks' typically encountered in manual dipping.

In most cases the derivatization reaction is initiated and/or completed by a heating step, which can be performed in a conventional oven or on a plate heater. For optimum reproducibility modern plate heaters ensure homogenous temperature across their surface and allow precise adjustment over a wide temperature range, typically up to 200°C.

**Reagents** Chemical derivatizing agents are generally grouped in specific and nonspecific or universal reagents. The latter include acids such as hydrochloric, phosphoric, and sulfuric, which can char many organic compounds at elevated temperature to yield colored and finally dark brown products. Addition of aldehydes such as anisaldehyde or vanillin to the reagent leads to characteristically colored zones for many natural products. Other universal reagents include antimony(III) or (V) chloride and phosphomolybdic acid.



**Figure 8** (A) Glass sprayer and (B) Chromajet (photograph courtesy of CAMAG and Desaga).



**Figure 9** Immersion device (photograph courtesy of CAMAG).

**Table 1** Common reagents for identification of functional groups

Functional group	Reagent	Comments
Acetylene compounds	Dicobalt octacarbonyl	Formation of colored compounds
Aldehydes	2,4-Dinitrophenylhydrazine	Formation of colored hydrazones and osazones
Alcohols, polyols, sugars	Lead(IV) acetate-dichlorofluorescein	Cleavage of vicinal diols
Amines (primary), amino acids	Ninhydrin	Formation of reddish-bluish compounds
Amines (primary and secondary)	7-Chloro-4-nitrobenzo-2-oxa-1,3-diazole (NBD-chloride)	Formation of fluorescent derivatives
Amines (capable of coupling)	Fast blue salt B and BB, fast black salt K, diazosulfanilic acid, diazosulfanilamide, 4-nitroaniline	Formation of azo dyes
Carboxyl groups	2,6-Dichlorophenolindophenol	Organic acids and undissociated acids become blue and reductants colorless
Ketones	2,4-Dinitrophenylhydrazine	Formation of colored hydrazones and osazones
Peroxides	Potassium iodide with starch	Formation of iodine from iodide, which forms blue complex with starch
Phenols, thiols	7-Chloro-4-nitrobenzo-2-oxa-1,3-diazole (NBD-chloride)	Formation of fluorescent derivatives

A list of reagents specific for certain functional groups is given in **Table 1**.

### Detection of Bioactivity

By focusing on their microbiological or biological and physiological activity a number of compounds can be detected. For bioautography, enzymes or test organisms such as spores, yeast cells, cell organelles, or bacteria in a nutrient medium can be directly transferred onto the plate by dipping. Biological activity is detected either immediately or following incubation typically involving enzymatic reactions or assessment of growth, respectively, their inhibition, linked to certain color changes. A commercial kit for detection of antibiotic activity based on growth inhibition of *Bacillus subtilis* is available from E. Merck. The developed TLC plate is dried and then immersed in a suspension of bacteria. Following incubation the plate is sprayed with MTT tetrazolium salt reagent (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide). Antibiotics inhibiting the growth of the bacteria appear as bright zones on colored background. Another kit, supplied by ChromaDex, is based on instant detection of effects on the chemiluminescence of *Vibrio fischeri*. Following development the TLC plate is immersed in the bacteria solution and immediately evaluated in a dark room using a very sensitive digital camera. Active compounds can either further excite or inhibit the luminescence of the bacteria seen as brighter or darker zones on a weakly glowing background.

### Detection of Radioactivity

Radioactively labeled compounds can be detected by autoradiography. Traditionally, the developed plate is

brought in contact with a special X-ray film suitable for the investigated isotopes. After development this film serves as an image of the plate, which can be densitometrically evaluated. Filmless autoradiography can be performed with so-called storage phosphor screens. The central element of this technology is a reusable phosphor screen, which is sensitive to ionizing radiation. In contact with the TLC plate containing labeled compounds an image is imprinted into the screen, which can be read out during a scanning step. Prior to a subsequent use of the screen the image is erased by exposure to visible light. Most recent approaches are based on electronic autoradiography directly measuring the radioactivity on the TLC plate with a position sensitive multiwire proportional counter chamber (Charpak, Nobel prize for physics 1992), or a proportional gas flow counter chamber. Modern instruments such as RITA or GITA (Raytest, Germany) allow real-time evaluation of chromatograms.

### Densitometry

For qualitative and quantitative evaluation of chromatograms in TLC, densitometry is the preferred technique. In general, a signal obtained from a position of the TLC plate that contains the sample component is compared to the signal of the plate without sample, which serves as baseline. The signal is displayed as a function of position thus generating an analog curve of the chromatogram. Following integration and peak detection the chromatogram can be evaluated quantitatively based on peak area and/or height of the investigated compound against a calibration curve generated from peak data of standards chromatographed on the same plate.



### Scanning Densitometry

In scanning densitometry a beam of light of selectable dimension and wavelength is moved across the plate following the individual tracks in direction of chromatography. In absorbance mode a photomultiplier is used to measure the amount of light, which goes either through the TLC plate (transmission mode) or is diffusely reflected from the surface of the plate (reflectance mode). Transmission measurements are of limited applicability because irregularities of the layer significantly affect the results and only white light can be used. For scanning in reflectance mode the spectral range from 190 to 800 nm can be utilized. Most frequently the absorbance of a substance is measured. Although the response of the photomultiplier is linear to the amount of received light the dependence of the signal on the amount of substance is essentially nonlinear. This phenomenon is caused by the fact that the sample is located not only on the surface of the stationary phase but also in deeper layers and these layers do not contribute to the signal to the same degree. This phenomenon is best described by the Kubelka–Munk equation.

The second type of measurement is that of fluorescence. It can be applied to substances that are excited by UV light. Prior to reaching the photomultiplier (PM) the UV light used for excitation is quantitatively absorbed by a cut-off filter so that no reflected light is measured. The fluorescent light emitted by the sample is able to pass the filter and reach the PM.

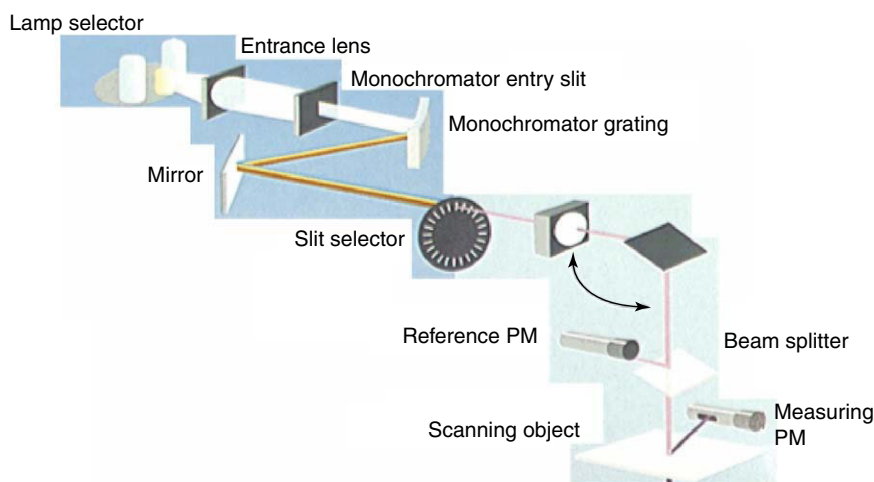
Modern densitometers such as the CD60 (Desaga, Germany), the CS9000 series (Shimadzu, Japan), and the TLC Scanner 3 (CAMAG, Switzerland) are single-beam, single-wavelength instruments controlled by powerful software managing all mechanical

functions and the evaluation of scan data. Figure 10 illustrates the principal design of such instruments using the TLC Scanner 3 as example. The densitometer consists of a compartment for exact positioning of the plate on a stage, electronic parts, and the optical system. Three lamps, a deuterium lamp covering the UV region of the spectral range, a mercury vapor lamp providing a spectrum of several lines of high intensity, and a tungsten lamp producing white light are available to cover the UV and visible range of the spectrum. After passing a lens system the emitted light reaches the monochromator, a holographic grating for selection of the scanning wavelength ( $\sim 20$  nm bandwidth). Through a revolving disk with fixed apertures the appropriate slit dimension is chosen. Finally, a beam splitter directs a part of the light beam to a reference photomultiplier reducing fluctuation and effects of lamp aging. The other part of the beam reaches the plate for measurement.

Another feature of modern densitometers is the option of recording spectra. For that purpose the light beam is positioned on the zone, of which the spectrum is to be obtained. Then the wavelength of the light is changed to cover the selected spectral range. After applying corrections for the lamp and the background the resulting spectrum can be used for identification of the separated substance.

### Video Densitometry

Video densitometry is usually performed on electronic images of the visualized chromatogram. Software such as VideoScan (CAMAG, Switzerland) and ProResult (Desaga, Germany) evaluates the pixels of each line of the image, which are within the user-selected tracks, according to their intensity on a 256 level gray scale. The resulting analog curve can



**Figure 10** Scheme of the light pass of the TLC Scanner 3 (photograph courtesy of CAMAG).

be used for quantitation by comparison to curves generated with calibration standards. Advantages of video densitometry are simplicity and speed. The major disadvantage compared to scanning densitometry is the absence of spectral selectivity. Only substances that can be visualized are accessible for the technique.

## TLC Software

As an offline technique the TLC process consists of several individual steps. Traditionally, instruments for each step are controlled by separate more or less sophisticated software. winCATS-Planar Chromatography Manager (CAMAG, Switzerland) is the first integrated software that can communicate with all instruments involved in the TLC process. A winCATS method file can include information about the stationary phase, samples and their components, standards and their preparation, application parameters, prechromatographic derivatization, development conditions, postchromatographic derivatization, all parameters of densitometric and spectral evaluation, and details of electronic documentation using a

digital or video camera. While the first step of a method is executed an analysis file including an analysis log is generated. The user is prompted to initiate the next step after the TLC plate has been manually transferred into the next instrument. The software is fully compliant with cGMP/cGLP and ready for use in an environment compliant to US CFR 21 rule 11.

*See also:* **Thin-Layer Chromatography:** Overview; Principles; Plate Technology; Method Development.

## Further Reading

- Jork H, Funk W, Fischer W, and Wimmer H (1990) *Thin Layer Chromatography, Volume 1a, Physical and Chemical Detection Methods*. Weinheim: VCH Verlagsgesellschaft mbH.
- Sherma J and Fried B (eds.) (2003) *Handbook of Thin-Layer Chromatography*, 3rd edn. New York: Dekker.
- Sherma J and Fried B (eds.) (1994) *Thin-Layer Chromatography, Techniques and Applications*, 3rd edn. New York: Dekker.
- Stahl E (1969) *Thin-Layer Chromatography, A Laboratory Handbook*, 2nd edn. Berlin: Springer.

# TIME-OF-FLIGHT MASS SPECTROMETRY

*See* **MASS SPECTROMETRY: Time-of-Flight**

## TIN

**L G Mackay**, National Analytical Reference Laboratory, Pymble, NSW, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Tin is a group IV element of atomic number 50 and atomic weight 118.71. Of all the elements it has the largest number of stable isotopes (10) with relative atomic masses ranging from 112 to 124. Tin normally exists in the oxidation states of +2 or +4, with tin(IV) being the more stable.

The toxicity of an element such as tin is highly dependent upon its chemical form. It is of relatively low toxicity in its inorganic form whereas the tri-substituted organotin forms of  $R_3SnX$  (where R is an aryl or alkyl group) are of very high toxicity. The ability to determine the levels of each species

of tin is thus highly important in determining the toxic impact on our environment. Government regulations have come into existence over the last decade requiring the monitoring of different forms of tin, in particular as specific organotins. A range of methods has been applied to organotin speciation analysis and some very elegant hyphenated techniques such as gas chromatography inductively coupled plasma mass spectrometry (GC-ICP-MS) have been employed.

## Total Tin Determination

Currently, many routine testing laboratories still determine only total tin concentrations in samples. This analysis typically involves complete digestion of the sample in a mixed acid media or alkali fusion of the material. Alkali fusion with compounds such as



be used for quantitation by comparison to curves generated with calibration standards. Advantages of video densitometry are simplicity and speed. The major disadvantage compared to scanning densitometry is the absence of spectral selectivity. Only substances that can be visualized are accessible for the technique.

## TLC Software

As an offline technique the TLC process consists of several individual steps. Traditionally, instruments for each step are controlled by separate more or less sophisticated software. winCATS-Planar Chromatography Manager (CAMAG, Switzerland) is the first integrated software that can communicate with all instruments involved in the TLC process. A winCATS method file can include information about the stationary phase, samples and their components, standards and their preparation, application parameters, prechromatographic derivatization, development conditions, postchromatographic derivatization, all parameters of densitometric and spectral evaluation, and details of electronic documentation using a

digital or video camera. While the first step of a method is executed an analysis file including an analysis log is generated. The user is prompted to initiate the next step after the TLC plate has been manually transferred into the next instrument. The software is fully compliant with cGMP/cGLP and ready for use in an environment compliant to US CFR 21 rule 11.

*See also: Thin-Layer Chromatography: Overview; Principles; Plate Technology; Method Development.*

## Further Reading

- Jork H, Funk W, Fischer W, and Wimmer H (1990) *Thin Layer Chromatography, Volume 1a, Physical and Chemical Detection Methods*. Weinheim: VCH Verlagsgesellschaft mbH.
- Sherma J and Fried B (eds.) (2003) *Handbook of Thin-Layer Chromatography*, 3rd edn. New York: Dekker.
- Sherma J and Fried B (eds.) (1994) *Thin-Layer Chromatography, Techniques and Applications*, 3rd edn. New York: Dekker.
- Stahl E (1969) *Thin-Layer Chromatography, A Laboratory Handbook*, 2nd edn. Berlin: Springer.

# TIME-OF-FLIGHT MASS SPECTROMETRY

*See* MASS SPECTROMETRY: Time-of-Flight

## TIN

**L G Mackay**, National Analytical Reference Laboratory, Pymble, NSW, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Tin is a group IV element of atomic number 50 and atomic weight 118.71. Of all the elements it has the largest number of stable isotopes (10) with relative atomic masses ranging from 112 to 124. Tin normally exists in the oxidation states of +2 or +4, with tin(IV) being the more stable.

The toxicity of an element such as tin is highly dependent upon its chemical form. It is of relatively low toxicity in its inorganic form whereas the tri-substituted organotin forms of  $R_3SnX$  (where R is an aryl or alkyl group) are of very high toxicity. The ability to determine the levels of each species

of tin is thus highly important in determining the toxic impact on our environment. Government regulations have come into existence over the last decade requiring the monitoring of different forms of tin, in particular as specific organotins. A range of methods has been applied to organotin speciation analysis and some very elegant hyphenated techniques such as gas chromatography inductively coupled plasma mass spectrometry (GC-ICP-MS) have been employed.

## Total Tin Determination

Currently, many routine testing laboratories still determine only total tin concentrations in samples. This analysis typically involves complete digestion of the sample in a mixed acid media or alkali fusion of the material. Alkali fusion with compounds such as

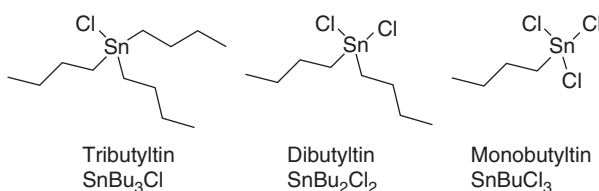
$\text{LiBO}_2$  can be carried out in a platinum crucible at  $1000^\circ\text{C}$ . Acid digestion typically involves acid mixtures such as concentrated nitric and hydrochloric acid heated to temperatures of  $100^\circ\text{C}$  or more. The aim of the digestion in this case is to completely break down the matrix. Total tin analysis is routinely carried out using atomic absorption spectrometry (AAS) and inductively coupled plasma (ICP) coupled with atomic emission spectrometry (AES) or with mass spectrometry (MS). Hydride generation is commonly used to reduce detection limits. This technique involves the addition of a reductant such as sodium borohydride to form tin hydride. Hydride generation has been used commonly with ICP-AES, ICP-MS, and AAS. Other techniques employed for total tin determination are instrumental neutron activation and X-ray fluorescence spectrometry.

## Organotin Compounds

Tin forms a wide range of organic compounds as a result of the formation of covalent bonds to  $\text{Sn(IV)}$ . These compounds are commonly referred to as organotins. Organotins have a wide range of industrial applications and hence they are now widely found in the environment. Analytical methods have been developed to carry out tin speciation to determine the different forms of organotins present in the environment as opposed to the determination of total tin levels. Inorganic tin is considered to be harmless, however, trisubstituted organotins have high toxicity and environmental guidelines now are internationally starting to include specific requirements for maximum allowed levels of triorganotins.

Two of the major organotin environmental contaminants are tributyltin and triphenyltin. Both are used as antifouling agents on marine vessels where they are applied to reduce the buildup of organisms on ships hulls. Triphenyltin has also been used as a fungicide and pesticide. Tricyclohexyltin is less common but is used as an acaricide to kill mites and ticks. Once these trisubstituted species are present in the environment, they degrade to less substituted species following a path of loss of the organic groups via the di- and monosubstituted species down to inorganic tin finally. **Figure 1** shows the tri-, di-, and monosubstituted forms for the butyltin chloride species.

Unfortunately this degradation process is slow, particularly once the trisubstituted species are bound to sediments in the environment. Dibutyltin and monobutyltin, the decomposition products of tributyltin, are also used industrially as stabilizers in chlorinated polymers such as PVC and hence their contamination has been found in a wide range of



**Figure 1** Structures of tri-, di-, and monobutyltin chloride species, three commonly determined environmental contaminants.

different matrices. Octyltin species have also been reported in environmental samples. Methylated tin species have been observed, which can form as a result of the methylation of inorganic tin in the environment. The tetrasubstituted species such as tetrabutyltin are of low toxicity and are mainly used as synthetic precursors for other organotins. They do have a delayed toxic effect as a result of their degradation to less substituted species.

One of the major environmental problems with organotins is that they accumulate in sediments, which are the ultimate environmental sinks for contaminants such as this. Many aquatic organisms, such as oysters and mussels, suffer from buildup of toxic levels of these contaminants because they are filter feeders and thus bioaccumulate these contaminants as a result of this. Significant malformation and growth retardation has been observed in aquatic organisms following exposure to organotin species and compounds such as tributyltin are suspected endocrine disruptors.

The use of tributyltin, the major environmental contaminant, as a biocide is now strictly regulated in most countries, however, its levels in waters and coastal sediments remain high enough to pose a significant toxicological risk. Many early analytical methods for organotins concentrated only on the trisubstituted species due to their high toxicity. However, as the demand to understand more about the biogeochemical pathways and degradation patterns of the trisubstituted species has grown, methods have been extended to include the less substituted species. Environmental monitoring now typically involves looking at the range of tri-, di-, and monosubstituted species present in sediments, water, and aquatic organisms. This range of testing enables authorities to obtain a true indication of the spread of contamination throughout the environment.

In the environment, organotins are associated with a range of different counterions, for example, they exist as chlorides, hydroxides, or oxides. Quantification of the levels of organotins concentrates on preserving the toxic organic component and measuring this species, irrespective of the type of counterion present. Organotin determination methods are

required to have an effective extraction technique followed by a separation technique coupled with a suitable detector. The various components of such methods are outlined in the following sections. Additional important factors such as sample storage, sample cleanup, and calibration techniques are also discussed.

## Extraction Techniques for Organotin Speciation

Traditional digestion techniques for total tin analysis are inappropriate for organotins as all species information is lost. Therefore, much more refined extraction techniques have to be developed. One of the major concerns in the extraction of organotins is the compromise between efficiently extracting the species while inducing minimal degradation of the specific individual compounds, which might result in distortion of the ratio of species present. The situation is made more complex by the fact that the polarity of the various species covers a wide range, with monosubstituted organotins having reasonably high polarity in comparison with their relatively nonpolar trisubstituted counterparts. To minimize this problem, chelating agents such as tropolone or diethyldithiocarbamate are sometimes employed which chelate with the organotins, particularly monobutyltin, forming less polar complexes which can be extracted with solvents suitable for the more highly substituted tin species.

The commonly used extraction techniques for organotins are mechanical shaking, ultrasonic extraction, microwave-assisted extraction, or pressurized fluid extraction. Simple mechanical shaking has been used by many testing laboratories, but this has been generally shown to be less effective than the latter three techniques, particularly with respect to recoveries for monosubstituted species. One of the advantages of shaking or ultrasonication is that they

have been observed to give lower levels of degradation. By contrast, microwave and pressurized fluid extraction techniques need to be more carefully optimized to avoid degradation but under the right conditions can provide high recoveries with minimal degradation.

It should be noted that it is difficult to find a universal extraction technique, which is applicable to all forms of the various tri-, di-, and monosubstituted organotins. Acids have generally been employed for the extraction of sediments. Where an inorganic acid such as HCl is used, concentrations above about  $1 \text{ mol l}^{-1}$  have been observed to cause degradation of organotins, and therefore their concentrations need to be limited. A combination of halogenated acid and organic solvent (with or without the addition of a chelating agent such as tropolone) has been used in many cases. Recently weaker acid systems such as aqueous acetic acid have been shown to be effective with the advent of extraction systems utilizing microwave or pressurized fluid conditions. For biota samples, basic and enzymic hydrolysis have been used to digest samples (Table 1).

### Mechanical Shaking and Ultrasonic Extraction

Mechanical shaking and ultrasonication of samples are traditional extraction techniques for organotins requiring limited specialized equipment. With appropriate solvents or chelating agents they can effectively extract a range of organotin species, although shaking often requires longer extraction times, typically several hours.

### Microwave Extraction

Microwave-assisted extraction has become one of the most widely applied techniques for organotins in sediments and biota. The important concept for effective speciation is the use of a low power microwave field (e.g. 50–100 W). Microwave digestion is commonly employed for total tin analysis where

**Table 1** Summary of a range of reported extraction procedures for a range of organotin species in environmental matrices

<i>Matrix</i>	<i>Extraction solvent</i>	<i>Species extracted/analyte</i>
Sediment	Hydrochloric acid followed by hexane	Tri-, di-, and monobutyltin
Sediment	Acetic acid followed by tropolone in toluene	Tri-, di-, and monobutyltin, tri-, di-, and monophenyltin, tri- and dicyclohexyltin, di- and monoethyltin
Sediment	Acetic acid and methanol	Tri-, di-, and monobutyltin, tri-, di-, and monophenyltin
Oyster	Hydrochloric acid followed by tropolone in hexane	Tri-, di-, and monobutyltin, tri-, di-, and monophenyltin
Mussel	Acetic acid and methanol	Tri-, di-, and monobutyltin
Fish and mussel	Aqueous tetramethylammonium hydroxide	Tri-, di-, and monobutyltin, tri-, di-, and monoethyltin

complete destruction of the matrix is desired, however, at low microwave powers organotin species can be leached without cleavage of the Sn–C bonds. Solvents such as glacial acetic acid, aqueous acetic acid, or methanolic acetic acid are most commonly used for sediments with extraction times using modern automated microwave systems of only around 5 min. Biota is often digested with tetramethylammonium hydroxide in a short 2–3 min extraction at low power.

### Pressurized Fluid Extraction

Pressurized fluid extraction using automated systems has become a highly popular extraction technique for organotins. The method allows the effective extraction of the species using short extraction times with minimal solvent. Typical conditions are 5–10 min extractions with temperatures  $\sim 100^{\circ}\text{C}$  and pressures  $\sim 2000$  psi. Simple solvent mixes such as methanolic acetic acid have been commonly employed with minimal degradation observed under these conditions.

### Solid Phase Microextraction

Solid phase microextraction (SPME) involves extraction onto a thin fiber and the technique has become more prevalent recently and additionally provides a preconcentration of analytes prior to analysis. The fibers used in the technique can be coated with a range of stationary phases and a non-polar phase such as polydimethylsiloxane (PDMS) is typically used for the extraction of derivatized organotin species. An equilibrium is established between the sample extract solution (or the headspace above the solution) and the stationary phase coating the fiber. The analytes are then typically desorbed from the fiber for analysis, for example, using thermal desorption during GC analysis. The technique allows rapid and solvent-free extraction of the analytes. Very good extraction has been achieved for water samples; however, the technique has been shown to be more variable with more complex matrices.

## Separation Techniques

For the effective analysis of organotin species, hyphenated analysis techniques are generally required involving separation of the individual species followed by an appropriate detection system. The two main chromatography techniques available for separation of these species are liquid chromatography (LC) and GC.

### Liquid Chromatography

Several forms of LC have been used for the separation of organotins. These techniques offer a simple, direct separation, which does not require derivatization of the species. The main disadvantages are the lower resolution available with LC and the limitations on detectors that can be coupled to LC for organotin analysis. Ultraviolet detection is not appropriate due to its very poor sensitivity for tin; however, fluorimetry has been successfully employed. The recent development of the two hyphenated techniques of LC–MS and LC–ICP–MS has provided two sensitive and selective detectors available for LC speciation.

The two main forms of LC, which have been applied to organotin analysis, are ion-exchange chromatography (using silica based cation-exchange columns) or reverse-phase chromatography (using C18 columns). In the former case, the monosubstituted organotins bind much more tightly than the di- and trisubstituted forms to the stationary phase, and hence mobile phase gradients, which vary the levels of complexing agents and pH, have been employed. Reverse-phase columns have become more widely used recently with mobile phases such as methanol/water/acetic acid employed with modifiers such as triethylamine. Under reverse-phase conditions, the various tin species can often be separated under isocratic conditions, which is more amenable to modern detection systems such as ICP–MS.

### Gas Chromatography

GC offers excellent chromatographic resolution and can be coupled to a range of appropriate detectors for organotin detection such as MS, ICP–MS, atomic emission (AED), AAS, and flame photometric detectors (FPD). The main problem with GC is the low volatility of all organotin species except in the tetraorganotin forms. This means that a derivatization step is required prior to GC analysis. This typically involves alkylation of the sample. GC columns, which have been employed for organotin analysis are typically nonpolar PDMS fused silica columns such as HP-1 or the 5% phenyl equivalents such as HP-5.

### Derivatization Techniques for GC Analysis

The two most commonly used derivatization processes used to produce volatile derivatives from sample extracts are Grignard reagents or aqueous phase ethylation procedures with a reagent such as sodium tetraethylborate. Both these reagents convert ionic organotin compounds such as  $\text{R}_n\text{SnCl}_{4-n}$  into non-polar volatile derivatives such as  $\text{R}_n\text{SnEt}_{4-n}$  in the case of ethylation reagents. Sodium tetraethylborate



has become a very common derivatizing reagent for the routine analysis of environmental samples. Grignard derivatizations are still employed by some laboratories utilizing reagents such as pentylmagnesium bromide to transform Sn-Cl groups into Sn-pentyl groups and thus producing a volatile species. Where Grignard derivatization is employed, the sample extract has to be typically dried and reconstituted in a dry aprotic solvent such as diethyl ether before the derivatization can occur. In contrast, sodium tetraethylborate can be used directly on aqueous based extracts with the derivatized species then extracted into a nonpolar solvent such as hexane for GC analysis. Some methods have used hydridization as the derivatization method, however the highly volatile nature of the derivatives makes them more difficult to work with routinely and in general the technique is more applicable to water samples.

## Detection Methods

### Fluorimetric Detection

Organotins are nonfluorescent compounds but fluorescence detection has been applied to the analysis of organotins in conjunction with appropriate derivatization reagents. One of the main problems has been that differently substituted forms of organotins can require different derivatizing reagents and thus generic methods for mono-, di-, and trisubstituted forms of organotins are quite difficult to develop. Despite this, methods for some organotin species in waters, sediments, and biota have been reported.

### Flame Photometric Detector and Pulsed Flame Photometric Detector

Traditional FP detection for organotins typically suffers from interferences from coextracted sulfur and phosphorus compounds, and problems with baseline stability have been reported for more complex matrices such as sediments. The new generation FPDs with a pulsed flame has offered a good method for routine organotin analysis. The higher selectivity available with pulsed flame photometric detector (PFPD) allows the use of a 390 nm optical filter (corresponding to the Sn-C emission) rather than the classical 610 nm (corresponding to the Sn-H emission). This has been observed to significantly reduce the effect of interferences and offers high sensitivity. However, the PFPD has a narrow linear range and samples must often be diluted to allow effective quantification, which can be time consuming.

### Atomic Absorption Spectrometry

AAS has been coupled with both GC and LC separation techniques. With GC separation, AAS is typically used with a quartz furnace following hydride generation. The hydrides are heated in the quartz tube to around 1000°C where they are atomized. Several wavelengths are available for tin determination by AAS using both hollow cathode lamps and electrodeless discharge lamps. Flame, quartz furnace, and electrothermal AAS have all been applied with LC separation, although the interface must cope with a continuous flow of solvent, which is not ideal. In general, AAS has been applied widely to tin speciation but is becoming less popular as it is less able to cope with low level samples and more complex matrices.

### Atomic Emission Spectrometry

AES with a microwave-induced plasma has been a popular technique for organotin analysis allowing very selective detection by utilization of specific Sn emission wavelengths such as 326, 303, or 271 nm. However, the detector can require high maintenance, which can be a drawback for routine use.

### Inductively Coupled Plasma Mass Spectrometry

Mass selective detection does not offer a traditional element-specific detection system as in the above three cases of FPD, AAS, and AES. However, ICP-MS offers excellent sensitivity and selectivity by means of a mass selective filter. Because of the very high selectivity of the detector, sample cleanup can often be minimal. Modern MS instruments have a large dynamic range and can be coupled with either LC or GC systems. Using LC, the commonly used narrow bore columns of 2 mm internal diameter can be easily connected directly to the ICP-MS nebulizer. The major requirement is typically the addition of a significant level of oxygen (~5%) to the argon nebulization gas flow in order to avoid the buildup of carbon deposits within the interface. A cooled spray chamber (~5°C) is also typically employed to minimize the solvent loading on the plasma as high organic solvent levels can cause the plasma to be unstable. Problems with plasma instability mean that isocratic LC mobile phases are preferable, as the organic loading then remains constant.

The large number of tin isotopes (10) means that there is significant choice for ICP-MS analysis, however, a significant number suffer from interferences or low abundances. As a result  $^{120}\text{Sn}$  is the most commonly employed isotope for analysis using  $m/z$  120. The use of ICP-MS detection also allows the primary method of isotope dilution mass spectrometry (IDMS) to be employed. This involves the

addition of an isotopically enriched organotin internal standard, for example,  $^{117}\text{Sn}$ -,  $^{118}\text{Sn}$ -, and  $^{119}\text{Sn}$ -containing organotin can be used for this purpose. This method can provide highly accurate and precise results as the determination depends upon the simple measurement of an isotopic ratio.

GC-ICP-MS is becoming one of the most common techniques for organotin analysis. The required interface to link the GC and ICP is now available commercially and a large number of laboratories have adopted this technique and it is now regularly utilized for organotin analysis. The high resolving power of GC for separation of different species and the excellent selectivity and sensitivity of the ICP-MS detector make this an excellent combination offering a rapid and precise analytical system.

### Other Mass Spectrometry Techniques

LC-MS has been used for organotin analysis, though its application is much less wide spread than ICP-MS. It offers the advantage of providing structural information because of the soft ionization processes employed in the atmospheric pressure interfaces, and molecular ions (or their adducts) are generally observed for organotin species.

High-resolution mass spectrometry has been utilized effectively by coupling a magnetic sector mass spectrometer with GC and an electron impact interface. This provides a highly sensitive and selective detection method but has not been reported for routine use.

### Sampling and Storage

The sampling and storage of collected samples for organotin speciation is highly important, as it is necessary to ensure that minimal species degradation occurs prior to analysis. The sample collection stage is particularly important when low level samples are being analyzed, e.g., at the nanogram per liter or nanogram per kilogram level. Many laboratories have investigated ideal storage conditions. It is recommended that sediments and waters be stored at in the dark at low temperature, typically 4°C. To maximize the long-term stability of water samples, pH adjustment to low pH (e.g., 2) is recommended. Ideally sediments and biota are stored in the freeze-dried state, but this is usually not feasible for routine samples. The filtering of water samples should be avoided as the organotin species are often bound to particulates in water samples. Samples are ideally collected in polytetrafluoroethylene (PTFE), polyethylene, or glass. PVC should be avoided, as dibutyltin is a common stabilizer in PVC materials.

### Sample Cleanup

Many methods involving highly selective detectors do not include a sample cleanup step. Where cleanup is required, techniques such as silica, alumina, or florisil cleanup is often employed to remove coextractives. The extraction of biological tissues can lead to high lipid levels in the extracts, these can be removed using silica cartridges. Silica and alumina have also been used to remove interferents from extracts of organic rich sediments. Water samples are often simply preconcentrated using solid-phase extraction cartridges with stationary phases such as C18.

### Contamination

Contamination of laboratory areas with organotins can be a considerable issue and the appropriate handling and disposal of contaminated articles should be carefully controlled and blank levels should be monitored.

### Standards

Standards for many organotins are available from commercial suppliers with chemical purities of typically 95% or better, however, limited uncertainty information is generally available on these purity values. Care should be taken in the choice of solvent and container which standards are prepared in to minimize any absorption onto the walls. Stock solutions are typically prepared in methanol and refrigerated. Working-level solutions are often prepared regularly, even daily.

### Calibration

The calibration technique of standard addition is often used to minimize matrix effects in the quantification of organotins. External calibration is also used by many groups, however, the use of an internal standard is highly recommended with the nature of the best compound to use dependent on the measurement system involved. Tripropyltin chloride has been used by many groups as a surrogate to monitor the whole method process, and species such as tetra-butyltin and tetrapentyltin have been used as internal standards added just prior to GC analysis.

Several laboratories have prepared isotopically enriched organotin species to use as internal standards in the technique of isotope dilution analysis. This MS technique can provide higher levels of accuracy than other calibration techniques and is currently used in many reference laboratories.



## Certified Reference Materials

In order to validate speciation methods more effectively, a number of certified reference materials (CRMs) have been produced to allow laboratories to measure the accuracy of their techniques. A range of sediment and biota materials are available such as the sediment materials PACS-2 from the National Research Council of Canada, NIES-12 from the National Institute for Environmental Studies, Japan and the CRM-462 from Community Bureau of Reference (BCR), EU. These materials have been rigorously homogeneity and stability tested and the levels of organotins have been certified by a range of techniques utilizing either 'definitive' methods or multiple independent methods.

*See also:* **Atomic Absorption Spectrometry:** Interferences and Background Correction. **Atomic Emission Spectrometry:** Principles and Instrumentation; Interferences and Background Correction; Flame Photometry; Inductively Coupled Plasma; Microwave-Induced Plasma. **Atomic Mass Spectrometry:** Inductively Coupled Plasma; Laser Microprobe. **Countercurrent Chromatography:** Solvent Extraction with a Helical Column. **Derivatization of Analytes.** **Elemental Speciation:** Overview; Practicalities and Instrumentation. **Extraction:** Solvent Extraction Principles; Solvent Extraction: Multistage Countercurrent Distribution; Microwave-Assisted Solvent Extraction; Pressurized Fluid Extraction; Solid-Phase Extraction; Solid-Phase Microextraction. **Gas Chromatography:** Overview. **Isotope Dilution Analysis.** **Liquid Chromatography:** Overview.

## Further Reading

Abalos M, Bayona J-M, Comaño R, *et al.* (1997) Analytical procedures for the determination of organotin

compounds in sediment and biota: A critical review. *Journal of Chromatography A* 788: 1–49.

Alonso JIG, Encinar JR, Rodríguez-González P, and Sanz-Medel A (2002) Determination of butyltin compounds in environmental samples by isotope-dilution GC-ICP-MS. *Analytical and Bioanalytical Chemistry* 373: 432–440.

Davies AG (1997) *Organotin Chemistry*. Weinheim: VCH. Ebdon L, Hill SJ, and Rivas C (1998) Organotin compounds in solid waste: A review of their properties and determination using high-performance liquid chromatography. *Trends in Analytical Chemistry* 17: 277–288.

Ebdon L, Pitts L, Cornelis R *et al.* (eds.) (2001). *Trace Element Speciation for Environment, Food and Health*. Cambridge: Royal Society of Chemistry.

Leroy MJF, Quevauviller P, Donard OFX, and Astruc M (1998) Determination of tin species in environmental samples. *Pure and Applied Chemistry* 70: 2051–2064.

Rajendran RB, Tao H, Nakazato T, and Miyazaki A (2000) A quantitative extraction method for the determination of trace amounts of both butyl- and phenyltin compounds in sediments by gas chromatography-inductively coupled plasma mass spectrometry. *The Analyst* 125: 1757–1763.

Rodríguez-González P, Encinar JR, Alonso JIG, and Sanz-Medel A (2003) Isotope dilution analysis as a definitive tool for the speciation of organotin compounds. *The Analyst* 128: 447–452.

Smedes F, de Jong AS, and Davies IM (2000) Determination of (mono-, di- and) tributyltin in sediments. Analytical methods. *Journal of Environmental Monitoring* 2: 541–549.

Suzuki T, Kondo K, Uchiyama M, and Murayama M (1999) Chemical species of organotin compounds in sediment at a marina. *Journal of Agricultural and Food Chemistry* 47: 3886–3894.

# TITRIMETRY

Contents

**Overview**

**Potentiometric**

**Photometric**

## Overview

**A Townshend**, University of Hull, Hull, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A titration is defined as 'the process of determining the quantity of a substance A by adding measured increments of substance B, the titrant, with which it reacts until exact chemical equivalence is achieved

## Certified Reference Materials

In order to validate speciation methods more effectively, a number of certified reference materials (CRMs) have been produced to allow laboratories to measure the accuracy of their techniques. A range of sediment and biota materials are available such as the sediment materials PACS-2 from the National Research Council of Canada, NIES-12 from the National Institute for Environmental Studies, Japan and the CRM-462 from Community Bureau of Reference (BCR), EU. These materials have been rigorously homogeneity and stability tested and the levels of organotins have been certified by a range of techniques utilizing either 'definitive' methods or multiple independent methods.

See also: **Atomic Absorption Spectrometry**: Interferences and Background Correction. **Atomic Emission Spectrometry**: Principles and Instrumentation; Interferences and Background Correction; Flame Photometry; Inductively Coupled Plasma; Microwave-Induced Plasma. **Atomic Mass Spectrometry**: Inductively Coupled Plasma; Laser Microprobe. **Countercurrent Chromatography**: Solvent Extraction with a Helical Column. **Derivatization of Analytes**. **Elemental Speciation**: Overview; Practicalities and Instrumentation. **Extraction**: Solvent Extraction Principles; Solvent Extraction: Multistage Countercurrent Distribution; Microwave-Assisted Solvent Extraction; Pressurized Fluid Extraction; Solid-Phase Extraction; Solid-Phase Microextraction. **Gas Chromatography**: Overview. **Isotope Dilution Analysis**. **Liquid Chromatography**: Overview.

## Further Reading

Abalos M, Bayona J-M, Comaño R, *et al.* (1997) Analytical procedures for the determination of organotin

compounds in sediment and biota: A critical review. *Journal of Chromatography A* 788: 1–49.

Alonso JIG, Encinar JR, Rodríguez-González P, and Sanz-Medel A (2002) Determination of butyltin compounds in environmental samples by isotope-dilution GC-ICP-MS. *Analytical and Bioanalytical Chemistry* 373: 432–440.

Davies AG (1997) *Organotin Chemistry*. Weinheim: VCH. Ebdon L, Hill SJ, and Rivas C (1998) Organotin compounds in solid waste: A review of their properties and determination using high-performance liquid chromatography. *Trends in Analytical Chemistry* 17: 277–288.

Ebdon L, Pitts L, Cornelis R *et al.* (eds.) (2001). *Trace Element Speciation for Environment, Food and Health*. Cambridge: Royal Society of Chemistry.

Leroy MJF, Quevauviller P, Donard OFX, and Astruc M (1998) Determination of tin species in environmental samples. *Pure and Applied Chemistry* 70: 2051–2064.

Rajendran RB, Tao H, Nakazato T, and Miyazaki A (2000) A quantitative extraction method for the determination of trace amounts of both butyl- and phenyltin compounds in sediments by gas chromatography-inductively coupled plasma mass spectrometry. *The Analyst* 125: 1757–1763.

Rodríguez-González P, Encinar JR, Alonso JIG, and Sanz-Medel A (2003) Isotope dilution analysis as a definitive tool for the speciation of organotin compounds. *The Analyst* 128: 447–452.

Smedes F, de Jong AS, and Davies IM (2000) Determination of (mono-, di- and) tributyltin in sediments. Analytical methods. *Journal of Environmental Monitoring* 2: 541–549.

Suzuki T, Kondo K, Uchiyama M, and Murayama M (1999) Chemical species of organotin compounds in sediment at a marina. *Journal of Agricultural and Food Chemistry* 47: 3886–3894.

# TITRIMETRY

Contents

**Overview**

**Potentiometric**

**Photometric**

## Overview

**A Townshend**, University of Hull, Hull, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

A titration is defined as 'the process of determining the quantity of a substance A by adding measured increments of substance B, the titrant, with which it reacts until exact chemical equivalence is achieved

(the equivalence point)'. The titrant is usually added as a standardized solution, but electrolytic generation, as in coulometric titrations, is also possible.

The achievement of the equivalence point is indicated in one of two ways. The first is visually, for example, by the addition of an indicator, which changes color, or by fluorescence or a similar property at or close to the equivalence point. The change indicates the endpoint of the titration, and the indicator should be chosen so that the endpoint is as close as possible to the equivalence point. The second method is to measure a physical property of the solution being titrated (e.g., conductivity, pH, absorbance) and identify the equivalence point by processing the signals obtained. The latter approach is particularly useful when titrations are automated.

The amount of titrant added is usually measured by volume (by dispensing the solution from a burette), and in this case, titrimetry is an example of volumetric analysis. Occasionally, the titrant is measured by weight (especially if greater accuracy is required) or by amount of electricity (as in coulometric titrations).

Titrimetry is one of the oldest analytical techniques, originating in the middle of the eighteenth century as a rapid means of quality control of industrial processes, such as acid manufacture. Since that time, the equipment has been refined, the procedures have been automated, and the number of chemical reactions utilized greatly increased, but the basic principles are unchanged. Its continued popularity stems from the simplicity of equipment and execution, wide applicability, and high accuracy and precision (greater than most instrumental techniques), all of which make it particularly applicable to the determination of major and minor components of samples. Skilled titrimetric analysis should give results with a precision of  $<0.2\%$  at the  $1 \times 10^{-2} \text{ mol l}^{-1}$  level.

Titrimetry may be classified with respect to the types of reaction that are involved. The major reactions are acid–base reactions (hence acid–base titrimetry), redox reactions (redox titrimetry), complexing reactions (compleximetric titrimetry), and precipitation reactions (precipitation titrimetry), which will be discussed in more detail below.

Titrimetry may also be classified by the nature of the endpoint measurement. The use of electrical measurements gives rise to potentiometric, amperometric, and coulometric titrations. Measurement of heat changes is used in thermometric titrimetry, and of absorbance in photometric and turbidimetric titrations. Radiometric titrations measure changes in radioactivity during the titration. All of these techniques are dealt with in other articles in this Encyclopedia. This article discusses only those titrations that use visual indicators.

## General Manual Titrimetric Technique

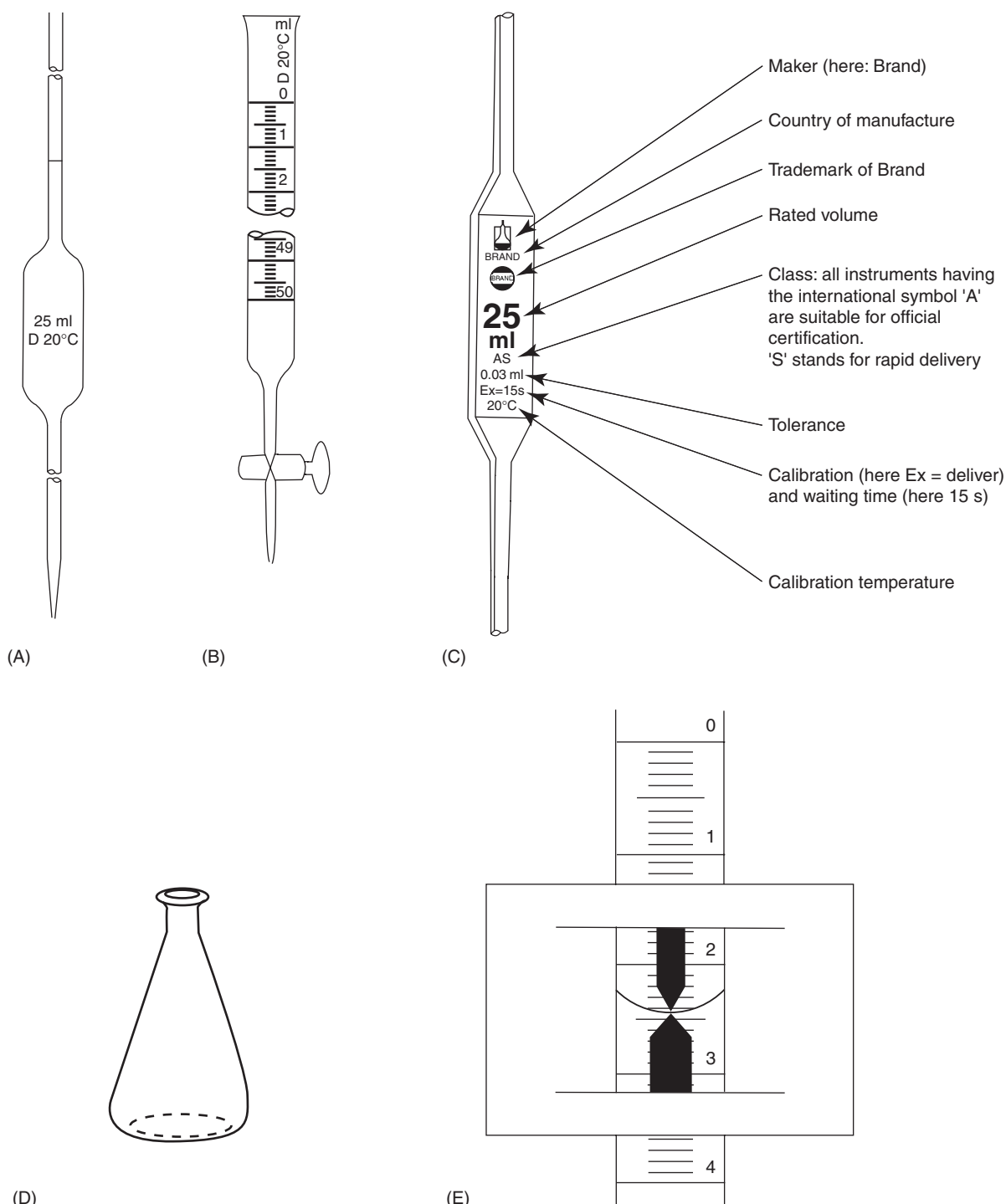
The simple equipment used for a typical titration is shown in **Figure 1**. A sample is measured into the flask from a pipette, or by weighing. The accuracy and precision of manual titrations using visual indicators is critically dependent on the use of correct experimental technique by the analyst. As in most analytical procedures, precise measurement of the amount of sample (or sample aliquot) is necessary, but it is most important in titrimetry if an accuracy of  $<0.2\%$  is to be achieved. Thus, proper use of the pipette, burette, and balance, and a careful sample preparation procedure is crucial. Measurement devices of high quality, such as class A pipettes and burettes, should be used.

When carrying out a titration the burette, holding the titrant solution, is clamped vertically just above the flask. Generally, the flask containing the solution being titrated should be placed on a white tile. Where the color change is somewhat difficult to detect, a reference solution should be used for comparison (this is a solution held in a flask similar to that being used for the titration, has the same volume, contains the same indicator, and has been adjusted to the endpoint). Titrant is added from the burette to the solution in the flask, which is then swirled by hand. As the endpoint approaches (which will be signaled by a transient color change in the portion of the solution where the titrant is added), the titrant should be added dropwise. When very close to the endpoint, fractions of drops can be added by touching the tip of the burette with a partly formed drop on the inside of the flask, and washing down with water from a wash bottle. The detection of the endpoint, when using a 50 or 10 cm<sup>3</sup> burette, should be possible to within 0.02 cm<sup>3</sup>. The volume of titrant run out of the burette is used to calculate the concentration of analyte in the titrated solution.

## Standards

Titrimetric analysis depends upon the availability of solutions of accurately known concentration for use as titrants. Such standard solutions may be prepared and themselves standardized by titration with solutions prepared from materials of guaranteed purity and composition, preferably solutions of chemicals known as primary standards. These materials have the following properties, as listed by Dodge:

1. They are easily obtained in an analytically pure state.



**Figure 1** Equipment used for a manual titration. (A) Transfer pipette; (B) burette; (C) information provided on volumetric glassware; (D) conical or Erlenmeyer flask (normally 250 or 100 cm<sup>3</sup>); (E) burette reader. (Reproduced with permission from Belcher *et al.* (1970) and Rudolf Brand and Co. *Working with Volumetric Instruments*. Wertheim: Brand.)

2. They should be unalterable in air at ambient or moderately high temperatures.
3. They should have a high equivalent weight, thus decreasing the effect of small weighing errors.
4. They should be readily soluble under the conditions of the analysis, thus allowing immediate titration in the cold.
5. On titration, no interfering product should be present.

**Table 1** Some primary titrimetric standards

Compound		Type	Type of titration
Name	Formula		
Anhydrous sodium carbonate	$\text{Na}_2\text{CO}_3$	Weak base	Acid–base
Sodium borate (borax) (recrystallized)	$\text{Na}_2\text{B}_4\text{O}_7 \cdot 10\text{H}_2\text{O}$	Weak base	Acid–base
Sulfamic acid	$\text{NH}_2\text{SO}_3\text{H}$	Strong acid	Acid–base
Potassium hydrogenphthalate	$\text{KHC}_8\text{H}_4\text{O}_4$	Weak acid	Acid–base
Potassium hydrogenbiiodate	$\text{KH}(\text{IO}_3)_2$	Strong acid	Acid–base
Silver nitrate	$\text{AgNO}_3$		Argentimetric
Sodium oxalate	$\text{Na}_2\text{C}_2\text{O}_4$	Reductant	Redox
Arsenic(III) oxide	$\text{As}_2\text{O}_3$	Reductant	Redox
Potassium dichromate	$\text{K}_2\text{Cr}_2\text{O}_7$	Oxidant	Redox
Ammonium hexanitrocerate(IV)	$(\text{NH}_4)_2\text{Ce}(\text{NO}_3)_6$	Oxidant	Redox
Potassium iodate	$\text{KIO}_3$	Oxidant	Redox
Potassium bromide	$\text{KBrO}_3$	Oxidant	Redox
Calcium carbonate	$\text{CaCO}_3$	Source of calcium ions	Compleximetric
Zinc oxide	$\text{ZnO}$	Source of zinc ions	Compleximetric
Ni, Zn, Cu metals	Ni, Zn, Cu	Source of metal ions	Compleximetric
Anhydrous disodium EDTA	$\text{C}_{10}\text{H}_{14}\text{N}_2\text{O}_8\text{Na}_2$	Complexing agent	Compleximetric

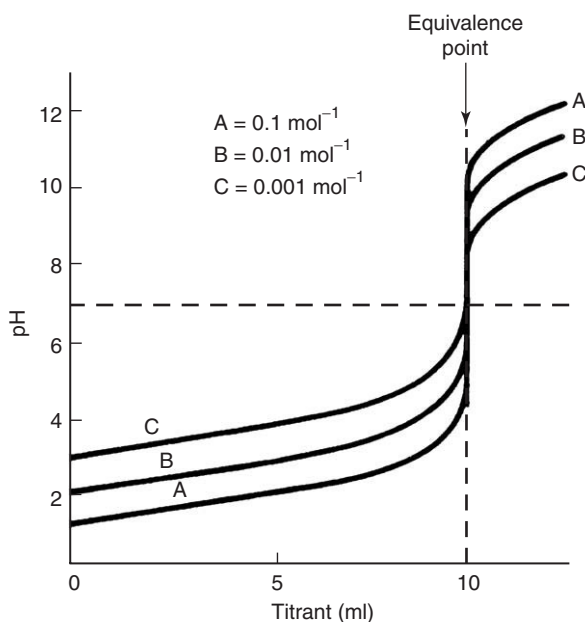
6. They should be colorless, before and after titration, to avoid interference with indicators.

Primary standards for particular titrations are given in Table 1. There are relatively few compounds that satisfy all these conditions.

It is also possible to buy concentrated standard solutions, which, after accurate dilution, can be used for titrimetry.

## Acid–Base Titrations

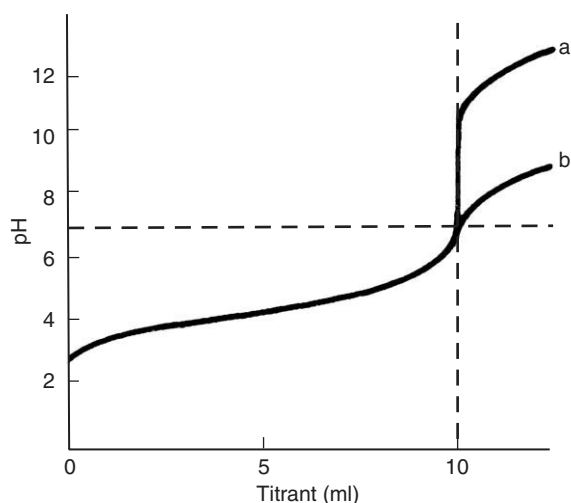
In acid–base titrations, an acid is determined by titration with a base, or vice versa. The essential reaction is between  $\text{H}^+$  and  $\text{OH}^-$ , giving water. The pH at the endpoint depends on the dissociation constants of the reactants and products. Thus, the titration of a strong, i.e., completely dissociated, acid with a strong, almost completely dissociated, base reaches equivalence at pH 7.0. A typical example is the titration of hydrochloric acid with sodium hydroxide solution. If a weaker, i.e., less dissociated base, such as ammonia, is used, the equivalence pH is  $< 7.0$ ; the weaker the base, the lower the equivalence pH. Likewise, if a weak, i.e., less dissociated, acid is titrated with a strong base, the equivalence pH is  $> 7.0$ , the pH increasing with increasing weakness of the acid. For the titration of a weak acid with a weak base, the equivalence pH depends on the relative dissociation constants of the acid and base, but the limited pH change means that the equivalence point is not as sharp as in a strong acid or strong base titration. The change of pH during the course of such titrations is illustrated in Figures 2 and 3.



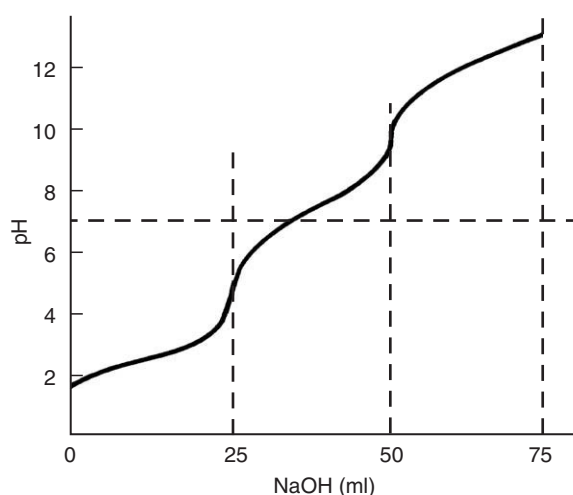
**Figure 2** Titration of 10 ml of hydrochloric acid of various concentrations with sodium hydroxide solution of the same concentration. (Redrawn from Belcher *et al.* 1970.)

The extent of the pH change also depends on the concentration of the analyte and titrant. Figure 2 shows how the pH change decreases with decreasing concentration of HCl and NaOH. The indicators for such titrations are chosen to change color very close to the pH at equivalence, and are described in detail in a separate article.

Some acids are polybasic, i.e., they give rise to more than one hydrogen ion. Phosphoric acid, for



**Figure 3** Titration of 10 ml of  $0.1 \text{ mol l}^{-1}$  acetic acid with (A)  $0.1 \text{ mol l}^{-1}$  sodium hydroxide; (B)  $0.1 \text{ mol l}^{-1}$  ammonia. (Redrawn from Belcher *et al.* 1970.)

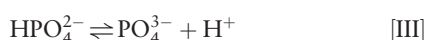
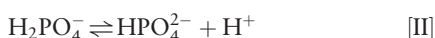
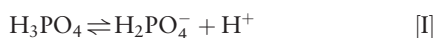


**Figure 4** Titration of 25 ml of  $0.1 \text{ mol l}^{-1}$   $\text{H}_3\text{PO}_4$  with  $0.1 \text{ mol l}^{-1}$  NaOH. (Redrawn from Belcher *et al.* 1970.)

**Table 2** Examples of acid–base titrimetric analyses

Analyte	Titrant	Indicator	Conditions
$\text{CaCO}_3$	HCl	Methyl red or phenolphthalein	–
$\text{H}_3\text{PO}_4$	NaOH	Methyl orange	First endpoint
		Thymolphthalein	Second endpoint
$\text{NH}_4^+$	HCl	Methyl red	Distillation from NaOH solution
Boric acid	NaOH	Phenolphthalein	Binding with mannitol or sorbitol to increase acidity

example, produces three hydrogen ions:



The change in pH on titration with NaOH is shown in **Figure 4**. The first dissociation (reaction [I]) occurs most easily, and titration with NaOH gives an equivalence point at pH 4. Further titration, of the second, more strongly bound hydrogen ion (reaction [II]), also gives rise to an equivalence point, at pH 9. The third hydrogen ion does not give rise to a sharp endpoint. Thus, phosphoric acid may be determined by titration to the first or second endpoints.

Some examples of acid–base titrimetric analyses are given in **Table 2**.

## Precipitation Titrations

These are titrations in which the analyte and titrant react to form a precipitate. The only common titrant used is silver nitrate (argentimetric titrations), and its use is mainly restricted to the determination of

chloride, bromide, iodide, cyanide, and thiocyanate, although in principle any species that is precipitated by silver ions could be determined. Direct titrations involve the use of potassium chromate (Mohr's method) or fluorescein derivatives as indicator. An indirect (back-titration) procedure is also popular, in which excess of precipitant ( $\text{Ag}^+$ ) is added to the sample, and the unreacted  $\text{Ag}^+$  titrated with thiocyanate ions (Volhard's method), using iron(III) as indicator. The mechanisms of the indicator reactions are described in another article.

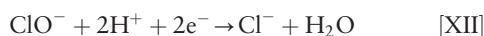
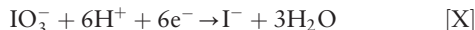
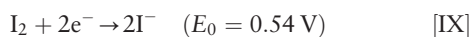
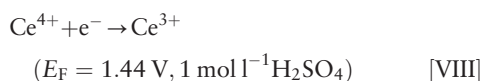
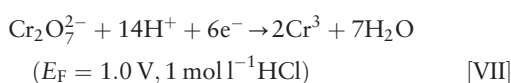
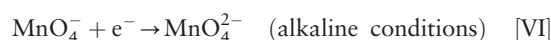
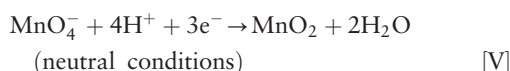
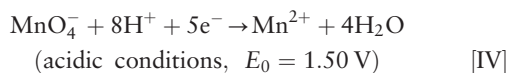
## Redox Titrations

In these titrations, a reducing agent is titrated with an oxidizing agent, or vice versa. The common oxidizing titrants are potassium permanganate ( $\text{KMnO}_4$ ), potassium dichromate ( $\text{K}_2\text{Cr}_2\text{O}_7$ ), cerium(IV) sulfate ( $\text{Ce}(\text{SO}_4)_2$ ), iodine ( $\text{I}_2$ ), potassium iodate ( $\text{KIO}_3$ ), and potassium bromate ( $\text{KBrO}_3$ ), all of which are solids, and sodium hypochlorite ( $\text{NaClO}$ ), which is available as a solution. The most important reducing titrants are iron(II) salts, often ammonium iron(II) sulfate ( $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ ), sodium thiosulfate ( $\text{Na}_2\text{S}_2\text{O}_3 \cdot \text{H}_2\text{O}$ ), and arsenic(III) oxide ( $\text{As}_2\text{O}_3$ ).

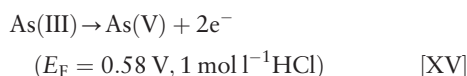
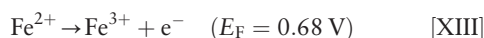


The stoichiometry of the reaction between one of these titrants and a particular analyte is established by combining the appropriate half-reactions. For the titrants above, the half-reactions are as follows:

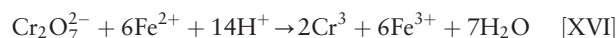
### Oxidants



### Reductants



For example, iron(II) can be determined by titration with dichromate, so combination of the appropriate half-reactions [VII] and [XIII], so as to achieve a charge and mass balance, gives the overall reaction [XVI]:



The driving force for each half-reaction is measured by its oxidation potential,  $E$ , measured in V, which is

given by the Nernst equation:

$$E = E_0 + \frac{0.059}{n} \log \frac{[\text{ox}]}{[\text{red}]} \quad [1]$$

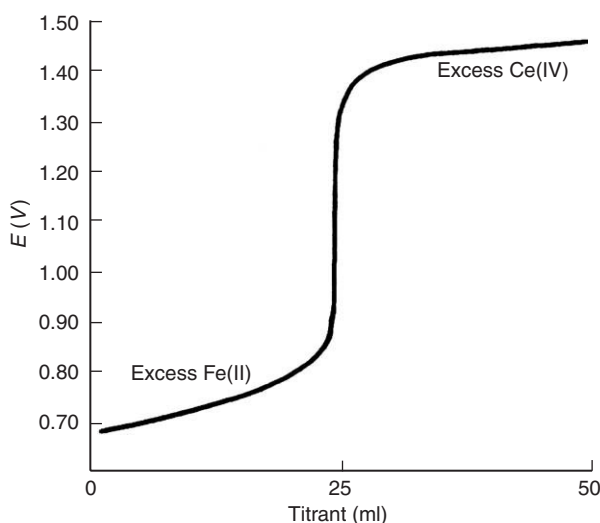
where  $n$  is the number of electrons involved in the above half-reaction, and [ox] and [red] are the concentrations (or better, activities) of the oxidized and reduced forms of the species, respectively.  $E_0$  is a constant known as the standard potential, which is the idealized potential when [ox] = [red]. A similar parameter, but measured under actual experimental conditions, is known as the formal potential ( $E_F$ ), and is more immediately useful for application in redox titrimetry. Some potentials for the above half-reactions are given in parentheses in the above equations.

The change in oxidation potential during the titration of a reductant (iron(II)) by an oxidant (cerium(IV)) is shown in Figure 5. The final oxidation potential increases with the strength of the oxidant used. Sometimes, it is possible to make a simple calculation of the equivalence potential ( $E_{EP}$ ) as follows. For the reaction that can be written as in [XVII]:



then

$$\begin{aligned} E_{EP} &= \frac{E_F(\text{Ce}^{4+}/\text{Ce}^{3+}) + E_F(\text{Fe}^{3+}/\text{Fe}^{2+})}{2} \\ &= \frac{1.44 + 0.68}{2} \\ &= 1.06 \text{ V} \end{aligned} \quad [2]$$

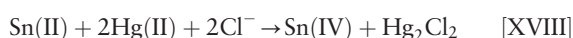


**Figure 5** Titration of 25 ml of  $0.1 \text{ mol l}^{-1}$  iron(II) with  $0.1 \text{ mol l}^{-1}$  cerium(IV). (Redrawn from *Vogel's Textbook of Quantitative Inorganic Analysis*, 4th edn (1978). London: Longman.)

For more complex systems, however, especially those involving oxoanions, such simple calculations are not valid.

Indicators for redox titrations will be chosen to change color reversibly by oxidation or reduction at a potential as close as possible to the equivalence potential (starch indication for iodine is an exception). This aspect is described in detail in another article.

Some analytes may be determined by titration with an oxidant, after their reduction. There are several ways of carrying out such reductions. One commonly used reductant is tin(II) chloride, in which the excess of tin(II) is destroyed by addition of mercury(II) chloride:



The mercury(I) chloride is unaffected by oxidants during the subsequent titration. Sulfite (or  $\text{SO}_2$ ) and hydrogen sulfide are alternative reductants. Metals may also be used. Small pieces of metal (zinc – a Jones reductor, silver – a Walden reductor) are used to fill a column, through which the analyte solution is passed. The effluent is titrated with oxidant. A comparison of the reduction products of the two reductor columns is given in Table 3.

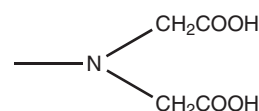
**Table 3** Comparison of the Jones and Walden reducers

Walden (HCl)	Jones ( $\text{H}_2\text{SO}_4$ )
$\text{Fe(III)} \rightarrow \text{Fe(II)}$	$\text{Fe(III)} \rightarrow \text{Fe(II)}$
Ti(IV) not reduced	$\text{Ti(IV)} \rightarrow \text{Ti(III)}$
Cr(III) not reduced	$\text{Cr(III)} \rightarrow \text{Cr(II)}$
$\text{V(V)} \rightarrow \text{V(IV)}$	$\text{V(V)} \rightarrow \text{V(II)}$
$\text{Mo(VI)} \rightarrow \text{Mo(V)}$	$\text{Mo(VI)} \rightarrow \text{Mo(III)}$
$\text{Cu}^{2+} \rightarrow \text{Cu(I)}$	$\text{Cu}^{2+} \rightarrow \text{Cu(0)}$

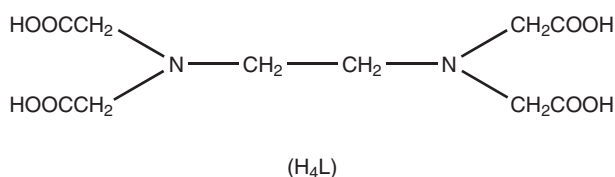
Redox titrations are still widely used. Table 4 summarizes some applications of redox titrations.

## Compleximetric Titrations

Compleximetric titrations are used mainly to determine metal ions by use of complex-forming reactions. Although in theory many complexing agents (cyanide, thiocyanate, fluoride, 1,2-diaminoethane, etc.) could be used for this purpose, in practice the titrants are almost always compounds having the iminodiacetic acid functional group:



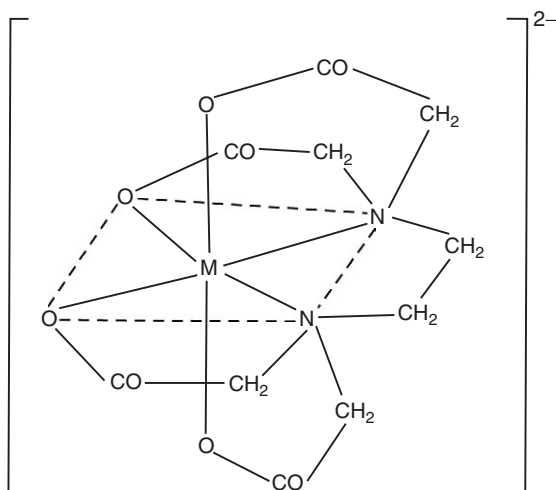
and by far the most popular of these is ethylenediaminetetraacetic acid (EDTA):



This ligand fulfills many of the qualities of a good compleximetric titrant. It forms complexes with most metal ions (those with the alkali metals are too weak to be useful); all the complexes have exact 1:1 stoichiometry, because the ligand is hexadentate, and can therefore occupy up to six coordination positions on the metal ions, as shown in Figure 6. The reaction with most metal ions is rapid ( $\text{Cr}^{3+}$  is a well-known exception), and the complexes are water

**Table 4** Examples of redox titrations

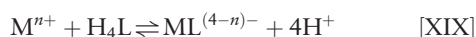
Analyte	Titrant	Indicator	Condition
Iron(II)	$\text{KMnO}_4$	Self-indicating	$\text{SnCl}_2$ reduction
Iron(III)	$\text{KMnO}_4$	Self-indicating	
$\text{H}_2\text{O}_2$	$\text{KMnO}_4$	Self-indicating	
Iron	$\text{K}_2\text{Cr}_2\text{O}_7$	Diphenylamine sulfonic acid	In iron ore, $\text{SnCl}_2$ reduction
Ethanol	$\text{K}_2\text{Cr}_2\text{O}_7$	<i>N</i> -Phenylanthranilic acid	Add excess oxidant, heat, back-titrate with iron(II)
Oxalate	$\text{Ce(SO}_4)_2$	Nitroferroin	Add excess oxidant, heat, back-titrate with iron(II)
Nitrite	$\text{Ce(SO}_4)_2$	Ferroin	Add excess oxidant, back-titrate with iron(II)
Copper(II)	$\text{Na}_2\text{S}_2\text{O}_3$	Starch	Iodide oxidized to iodine
Acids	$\text{Na}_2\text{S}_2\text{O}_3$	Starch	$5\text{I}^- + \text{IO}_3^- + 6\text{H}^+ \rightarrow 3\text{I}_2 + 3\text{H}_2\text{O}$
Available chlorine	$\text{Na}_2\text{S}_2\text{O}_3$	Starch	Oxidation of $\text{I}^-$ to $\text{I}_2$
Antimony(III)	$\text{I}_2$	Starch	Precipitate $\text{Mg}^{2+}$ with 8-quinolinol, add excess $\text{KBrO}_3/\text{KBr}$ to brominate precipitate, determine excess $\text{KBrO}_3$ by oxidation of $\text{I}^- \rightarrow \text{I}_2$
Magnesium(II)	$\text{Na}_2\text{S}_2\text{O}_3$	Starch	
Ascorbic acid	$\text{KIO}_3$	Self-indicating ( $\text{I}_2/\text{CCl}_4$ )	Add $\text{Br}^-$ and excess $\text{NaClO}$
Ammonia	$\text{NaClO}$	Bromothymol blue	



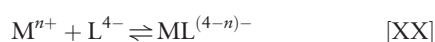
**Figure 6** Structure of chelate of EDTA anion ( $L^{4-}$ ) with a metal ion  $M^{2+}$ . (Redrawn from *Vogel's Textbook of Quantitative Inorganic Analysis*, 4th edn (1978). London: Longman.)

soluble and colorless (unless the metal ion itself is colored).

The reaction between a typical metal ion and EDTA ( $H_4L$ ) can be written as



that is, as a competition between the metal ion and hydrogen ions for binding with  $L^{2-}$ . The stability of binding of  $M^{n+}$  with  $L^{2-}$  is measured by its stability constant  $k_1$ , which is the equilibrium constant for the reaction:



$$k_1 = \frac{[ML^{(4-n)-}]}{[M^{n+}][L^{4-}]} \quad [3]$$

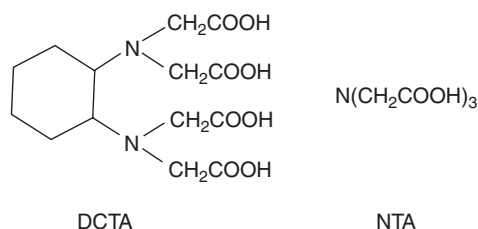
where  $[ ]$  denotes concentrations (better, activities).

Some typical stability constants are given in **Table 5**. Metals forming weaker complexes, therefore, require less acidic (i.e., higher pH) conditions for complex formation. The optimum pH values for the titration of a number of metal ions are included in **Table 5**. Metals forming stronger complexes can be titrated at lower pH values, at which the weaker complexing metals do not react, thus selective titration of, for example, bismuth can be carried out in the presence of lead at pH 1–2. The reaction of metal ions with EDTA (reaction (XIX)) generates  $H^+$ . Thus, to prevent a pH change during the titration, the solution must be adequately buffered.

**Table 5** Stability constants of some EDTA complexes and optimum pH for titration of the metal ions with EDTA

Metal ion	$\log k_1$	Optimum pH for titration
$Mg^{2+}$	8.7	10
$Ca^{2+}$	10.7	7.5
$Ba^{2+}$	7.8	12–13
$Mn^{2+}$	13.8	5.5
$Zn^{2+}$	16.5	4
$Ni^{2+}$	18.6	3
$Co^{2+}$	16.3	4
$Fe^{2+}$	14.3	5
$Fe^{3+}$	25.1	1
$Cu^{2+}$	18.8	3
$Hg^{2+}$	21.8	5.0–5.5 <sup>a</sup>
$Pb^{2+}$	18.0	4
$Bi^{3+}$	27.9	1–3

<sup>a</sup>The behavior of  $Hg^{2+}$  is somewhat anomalous because of hydroxocomplex formation. Values obtained from Pribil (1982).



**Figure 7** Two alternative titrants to EDTA.

Adjustment of pH often will not give sufficient selectivity and, of course, is inappropriate if a weaker-complexing metal ( $M_W$ ) has to be titrated in the presence of a more strongly complexing metal ( $M_S$ ). In such circumstances it is possible to 'mask'  $M_S$  by adding another complexing agent that complexes much more strongly with  $M_S$  than with  $M_W$ , so that  $M_W$  but not  $M_S$  will react with EDTA. The use of such masking agents is widespread in compleximetric titrimetry. A typical example is the use of triethanolamine to mask iron(III) when calcium is titrated with EDTA in alkaline solution.

A wide range of visual indicators is available for compleximetric titrations. These generally function by forming a colored complex with the metal ion being titrated, which causes a color change when the metal ion is removed from the complex by reaction with EDTA and releases the free ligand. These indicators are described in detail in another article.

Other EDTA-type compounds are sometimes used as titrants, including 1,2-diaminocyclohexane- $N,N,N',N'$ -tetraacetic acid, which generally forms stronger complexes than EDTA, and nitrilotriacetic acid, which generally forms weaker complexes than EDTA (**Figure 7**).

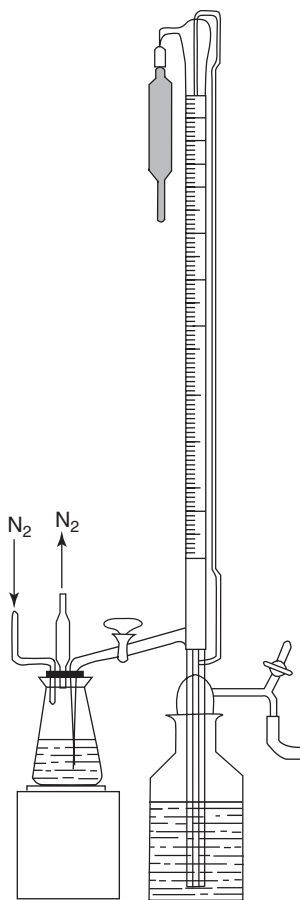
## Nonaqueous Titrimetry

Most titrations are carried out in aqueous solution, including all those described above. In some circumstances, however, it is advantageous to use other solvents, especially organic solvents. Such nonaqueous titrations are normally used for acid–base reactions, but redox reactions may also be applicable. The Karl–Fischer titration of water, in particular, is based upon redox reactions in a nonaqueous medium.

The ionization of a molecule HB in a solvent S is influenced by the solvation of the ions:



The ease of dissociation to form  $\text{HS}^+$  (solvated  $\text{H}^+$ ) increases with increasing basicity of the solvent, i.e.,



**Figure 8** Apparatus for visual titration with nonaqueous solvents. Automatic burette protected with guard tubes; a guard tube and blow-bulb are attached to the bottom outlet of the burette. The nitrogen flow can be omitted for titrations in acidic solvents. (Redrawn from Belcher *et al.* 1970.)

with increasing binding strength between  $\text{H}^+$  and the solvent. Thus, an acid that is very weak in aqueous solution will be stronger in a more basic solvent such as pyridine or dimethylformamide, and will give a bigger ‘pH’ change on titration. Phenols, for example, which are too weak acids to be titrated in aqueous solution, can be titrated in pyridine solution with tetrabutylammonium hydroxide in benzene–methanol (9:1, v/v) as titrant, and thymolphthalein in methanol as indicator.

Similarly, bases that are very weak in aqueous solution (e.g., amines) show increased basicity in solvents of greater acidity, such as anhydrous acetic acid. Perchloric acid in acetic acid may be used as the titrant, with crystal violet in acetic acid as indicator.

Because many of the solvents used are aggressive, volatile, and obnoxious, nonaqueous titrations are normally carried out in a closed environment, which also minimizes the ingress of moisture (**Figure 8**). It is essential to ensure that all apparatus used is dry and, especially for titrations in basic solvents, a stream of nitrogen is used to prevent access of carbon dioxide to the solution being titrated. The titrant is stored in a reservoir connected directly to the burette.

Compounds that may be determined by nonaqueous titrimetry include amines, amino acids, phenols, and Schiff’s bases. Carbonyl compounds (by oxidation and titration of the released  $\text{H}^+$ ) can also be determined. Such titrations are especially useful in the pharmaceutical industry.

**See also:** Indicators: Acid–Base; Redox; Complexometric, Adsorption, and Luminescence Indicators. **pH.** **Quality Assurance:** Internal Standards. **Water Determination.**

## Further Reading

- Belcher R, Nutten AJ, and Macdonald AMG (1970) *Quantitative Inorganic Analysis*, 3rd edn. London: Butterworths.
- Kolthoff IM and Belcher R (1957) *Volumetric Analysis (Redox Titrations)*, vol. III. New York: Interscience.
- Kolthoff IM and Stenger VA (1947) *Volumetric Analysis*, vols. I and II. New York: Interscience.
- Kucharsky J and Safarik L (1965) *Titration in Non-Aqueous Solvents*. Amsterdam: Elsevier.
- Pribil R (1982) *Applied Complexometry*. Oxford: Pergamon.
- Schwarzenbach G and Flaschka H (1969) *Complexometric Titrations*, 2nd English edition. London: Methuen.
- West TS (1969) *Complexometry with EDTA and Related Reagents*. Poole: BDH Chemicals.

## Potentiometric

**A Hulanicki and S Głąb**, University of Warsaw, Warsaw, Poland

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

A potentiometric titration belongs to chemical methods of analysis in which the endpoint of the titration is monitored with an indicator electrode that records the change of the potential as a function of the amount (usually the volume) of the added titrant of exactly known concentration. Potentiometric titrations are especially versatile because indicator electrodes suitable for the study of almost every chemical reaction used in titrimetry are now available. This technique is also frequently used in the study of operational conditions of visual titrimetric indicators proposed for general use in chemical analysis, as well as in the study of numerous reactions, such as protonation and complexation, which find their application not particularly in analytical measurements. The course of the potentiometric titration curve provides information not only about the titration endpoint position, but also the position and shape of the curve that may provide data about the processes accompanying the titration reaction. Another advantage is that the necessary apparatus is generally less expensive, reliable, and readily available in the laboratories.

### Potentiometric Equipment

Applications of potentiometry involve the use of an electrochemical cell consisting of a reference electrode of constant potential and an indicator electrode that responds to the analyte studied and sample composition. The electromotive force (e.m.f.) of this cell can be regarded as the difference of the potentials of the two electrodes (half-cells).

#### Indicator Electrodes

**pH-sensitive electrodes for acid–base titrations** The glass electrode is most often used but occasionally metal–metal oxide (e.g., antimony electrode) and pH-sensitive polymeric membrane electrodes are also used. These electrodes are also useful in investigation of chemical reactions through potentiometric titrations, as in many of them their course can be traced by the changes of hydrogen ion activity (concentration).

**Electrodes for precipitation titrations** The choice of the electrode depends on the character and properties of ions taking part in the titration reaction. Common

is the use of silver or mercury electrodes of the first order to trace the respective metal ions, or as the second-order electrodes, silver–silver halide and mercury–mercury(I) halide electrodes, when the corresponding anions are involved in the reaction. Metal electrodes with sulfide (or other sulfur(II) containing species) as co-anion are used less often. There is also a large group of ion-selective electrodes (ISEs) sensitive to  $F^-$ ,  $Cl^-$ ,  $Br^-$ ,  $I^-$ ,  $S^{2-}$ ,  $Cu^{2+}$ ,  $Pb^{2+}$ , and  $K^+$ , which are used to monitor precipitation titrations of both inorganic and organic ions.

**Electrodes for compleximetric titrations** Electrodes respond mainly to change of metal ion concentration, similarly as electrodes used in precipitation reactions. Occasionally, the electrode responds to the ligand, as is the case of the fluoride ISE. Electrodes with a similar scheme as the electrode of third order ( $Ag/Ag_2C_2O_4/CaC_2O_4/Ca^{2+}$ ) can be used in compleximetric titration, having as an intermediate step the equilibrium with the appropriate ligand (e.g.,  $AgEDTA/MEDTA$ ), where M denotes the metal ion being titrated. There is additionally a large group of ISEs that respond to the metal ions. As complexation with protonated ligands is accompanied by pH changes, pH-sensitive electrodes are occasionally used.

**Electrodes for redox titrations** For direct measurements of changes of the oxidation–reduction potentials inert electrodes such as bright platinum or gold are mainly used. Iridium and tungsten electrodes can also be used for measuring the redox potentials; however, the kinetic effects may influence the response. The practical application of mercury as an inert electrode is limited to regions of negative redox potentials, but in this region it has a definite advantage over the noble metal electrodes. This is connected with the high hydrogen overpotential on mercury.

#### Reference Electrodes

Most commonly used reference electrodes are based on mercury (calomel electrode) and silver, in equilibrium with the saturated solution of the corresponding chlorides. Their potential will be constant if the chloride ion concentration around the electrode is constant. The mercury–mercury(I) sulfate electrode is used instead of the calomel electrode when the presence of chloride is undesirable.

#### Instrumentation

The experimental instrumentation for potentiometric titration consists primarily of a high-impedance

electronic pH or millivolt meter, a beaker and a magnetic stirrer, indicator and reference electrodes, and a burette for titrant delivery, which are needed for manual titrations and point-by-point plotting. Automatic burettes and direct curve recording devices are increasingly used, in particular when the decrease of the scale of the titration is expected to go down below 1 ml of the amount of titrant used up to the endpoint. A variety of instruments are commercially available. In general, pH (or potential) meters can be divided on the basis of price and performance: utility (portability), general purpose, expanded scale, and research grades.

## Potentiometric Titration Curves

Potentiometric titration curves are obtained by plotting the potential of the indicator electrode (in practice the e.m.f. of the cell) against the volume of the titrant. In acid–base titrations, the pH value is usually plotted instead of the potential. The logarithmic shape of the titration curve is a consequence of the logarithmic dependence of activity or activity ratio of the species participating in the titration reaction according to the Nernst equation:

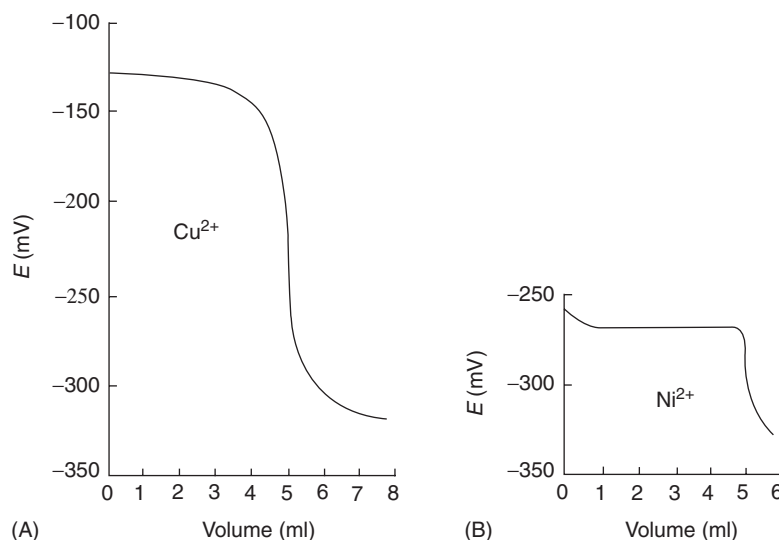
$$E = E^0 + \frac{RT}{zF} \ln \frac{a_{\text{ox}}}{a_{\text{red}}}$$

where  $a_{\text{ox}}$  represents the oxidant activity and  $a_{\text{red}}$  the reductant activity. When the activity of one of those species is constant, the activity ratio is simply replaced by the activity of the ion sensed by the indicator.

From the shape and position of the titration curve along the potential axis, information can be gained

about the parameters of the titration reaction. For example, in alkalimetric titration of an acid the pH value of the half-neutralization of the acid corresponds numerically to the  $\text{p}K_{\text{a}}$  value of the acid. This makes possible the evaluation of the  $\text{p}K_{\text{a}}$  value of the titrated species. Similarly, in other types of titrations (precipitation, compleximetric, and redox) the titration curve may supply information about the relevant titration reaction.

In the case of an ideal indicator electrode, such a curve should correspond to the change of the indicator ion activity. In practice, the curve is often distorted because of changes of liquid junction potential and insufficient selectivity of the electrode (especially for ISEs) and occurrence of processes not described by the Nernst equation. Nevertheless, in most cases it does not influence the evaluation of the endpoint. A potentiometric titration curve should have a symmetrical ‘S’ shape, when the dilution effect is neglected, and the indicator electrode is directly sensitive to the ion titrated. If the course is traced indirectly, e.g., in a compleximetric titration of nickel using a copper-selective electrode, the curve is asymmetric (Figure 1). Another reason for the asymmetry of the curve is the asymmetry of the titration reaction, which mostly occurs in redox titrations when the number of electrons in the reaction of the oxidant (e.g.,  $\text{MnO}_4^- \rightarrow \text{Mn}^{2+}$ ) is not equal to the number of electrons exchanged by the reductant (e.g.,  $\text{Fe}^{2+} \rightarrow \text{Fe}^{3+}$ ). Unusual shapes of the titration curve may occur when the titrated substance is directly inactive but after the addition of the titrant is gradually transformed into an active form in a kinetically restricted reaction. Such situations may occur also



**Figure 1** (A) Symmetric titration curve: compleximetric titration of copper ion, using a copper-selective electrode. (B) Asymmetric titration curve: compleximetric titration of nickel using a copper-selective electrode.



when catalytic processes occur. An improved and rapid potential change is observed when an indicator reaction is catalyzed by the small excess of the titrant. This can be exemplified by silver determination using iodide as titrant. The indicator reaction:  $\text{Ce(IV)} + \text{As(III)} = \text{Ce(III)} + \text{As(V)}$ , followed by a redox electrode, is catalyzed by the presence of free iodide ions, which appear just after the completion of the reaction between silver and iodide.

### Endpoint Evaluation

The endpoint of a titration is an experimentally estimated value, whereas the equivalence point corresponds to the theoretical amount of the titrant that is exactly equivalent to the amount of the analyte. The difference between the endpoint and the equivalence point should be minimized and is referred as the titration error.

The endpoint of the titration is often evaluated on the basis of a graphically represented titration curve. When the potential (or pH) change is large and the curve steep, the endpoint can be easily seen and evaluated. For less obvious cases of symmetrical titrations, several graphical methods of finding the point of maximal slope were used. Another method is titration to a fixed potential. The equivalence-point concentration of the indicator ion may be calculated from the known reaction equilibrium constants, and the corresponding potential values are subsequently evaluated. The equivalence point potential may also be estimated empirically. Titration is continued until this potential is reached, and the volume of the titrant noted. The maximum slope of the titration curve is easily found by plotting the first derivative curve,  $dE/dV$ , or the second derivative curve,  $d^2E/dV^2$ , against the titrant volume,  $V$  (Figure 2). In practice, this can be done for finite increments and the curve is approximated by plotting  $(E_n - E_{n+1})/(V_n - V_{n+1})$  against  $0.5(V_n + V_{n+1})$ , where  $n$  represents the successive additions of titrant. The numerical procedure termed the Hahn method is based on the same

principle. Computerized titrations can achieve the same goal without calculation by the analyst. It must be pointed out that such procedures are exactly valid for symmetrical curves only.

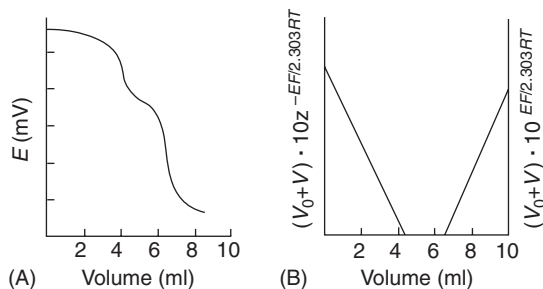
The common procedure of finding the endpoint, which should correspond to the equivalence volume, is based on the Gran method. This procedure is especially useful for flat and unsymmetrical titration curves. The titration data are transformed in such a way that the titration curve consists of two linear parts – before and after the equivalence point. The intersection of those lines or their intersection with the abscissa defines the equivalence volume of the titrant (Figure 3). The linearization is based on the transformation of the Nernst equation into the form

$$\text{antilog}(E/S) = \text{constant} \cdot a_X$$

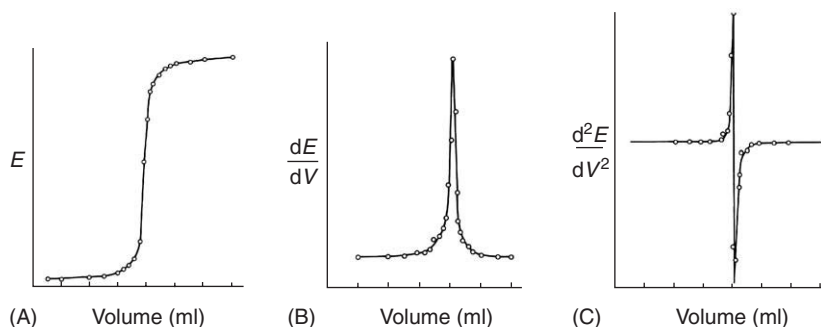
where the constant is  $\text{antilog}(E^0/S)$ , and  $S = RT/2.303F$ . The straight line is obtained when the increase of volume ( $V_0 + V$ ) is taken into account. Any deviation from linearity may indicate the occurrence of another reaction, either preceding or following the main reaction.

### Single Point Titration

This procedure is based on two potential measurements, one directly using the sample solution ( $E_1$ ) and another on the sample solution with addition of a known amount of analyte ion ( $E_2$ ). This is also



**Figure 3** Titration of two-component mixture: (A) titration curve; (B) linearization according to the Gran method.



**Figure 2** (A) Titration curve plotted in coordinates: potential,  $E$ , versus titrant volume,  $V$ . (B) First-derivative curve,  $dE/dV$ , versus  $V$ . (C) Second-derivative curve,  $d^2E/dV^2$ , versus  $V$ .

termed the known (or standard) addition method. For successful application of this method certain conditions are required:

1. The activity coefficient of the indicator ion should not vary significantly between the two measurements.
2. The side-reaction coefficient of the indicator ion, i.e., the ratio of the free determinand to all its chemical forms, should be constant.
3. The calibration slope of the electrode ( $S = RT/F$ ) must be known and constant.

When these conditions are fulfilled,

$$E = E_2 - E_1 = S \log(V_S C_X + V_A C_A) [C_X (V_S + V_A)]^{-1}$$

where  $V_S$  and  $V_A$  are volumes of sample and of the addition respectively.  $C_X$  and  $C_A$  are unknown and addition concentrations, respectively. This enables  $C_X$  to be calculated as

$$C_X = \frac{C_A V_A}{V_A + V_S} \left( 10^{E/S} - \frac{V_A + V_S}{V_A} \right)^{-1}$$

which for negligible volume of addition ( $V_A \gg V_S$ ) simplifies to

$$C_X = C_A (10^{E/S} - 1)^{-1}$$

A reverse method is based on the removal of the determinand and is called the known (or standard) subtraction method. The removal of the determinand is based on its precipitation or complexation. The necessary condition is that the reaction must be as complete as possible.

Potentiometric titrations may be also performed in the flow injection mode. This is attractive because small amounts of sample are needed, the time of determination is short, and the results for low concentration are reliable. Besides, the potentiometric signal depends linearly on the concentration of the analyte, enabling determination in a broad concentration range. There are several types of such titration, e.g., a fixed sample volume is mixed with a titrant solution of constant concentration. The width of the potentiometric signal, measured on the time scale as  $\Delta t$ , is proportional to the logarithm of the analyte concentration. The quantitative evaluation is based on a calibration graph. The important factors are the linear range of determination and the time constant of the potentiometric detector. The latter must be compatible to the flow rate of the solution in the system.

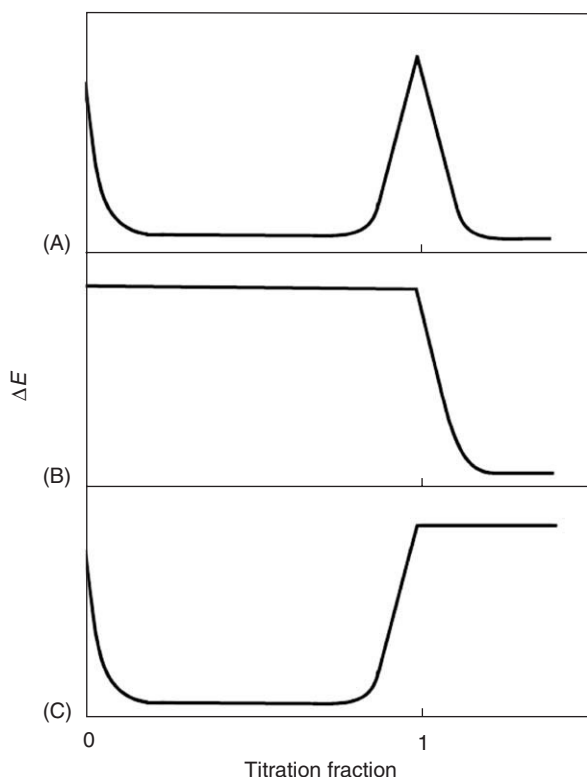
### Potentiometric Titrations with Polarized Indicator Electrodes

A number of oxidation–reduction systems (e.g., as  $\text{Cr}_2\text{O}_7^{2-}/\text{Cr}^{3+}$ ) are slow in establishing constant potentials at the platinum indicator electrode when measurements are made without current flow. It is possible to avoid long waiting times by forcing slight electrolysis to occur at the indicator electrode. Potentiometric titrations with polarized electrodes are divided in two classes, depending on whether one or two electrodes are used. In the former, the potential of a single polarized platinum indicator electrode against a reference electrode is measured. This indicator electrode may be polarized anodically or cathodically. If both couples involved in titration are reversible, the potential change of one polarized indicator electrode at the endpoint will be the same as when the electrode is not polarized. When one or both of the couples are irreversible, the potential change may be much greater in the case of a polarized electrode.

In the case of potentiometry with two polarized electrodes (bipotentiometry, or differential electrolytic potentiometry), the difference of potential between the two electrodes is measured. The shape of the titration curve depends on the reversibility of the couples involved in the titration. If both the titrant and the analyte are reversible couples (e.g.,  $\text{Fe}^{3+} + \text{Ce}^{4+}$  in  $\text{H}_2\text{SO}_4$ ), a curve with a maximum at the endpoint is obtained (Figure 4A). When the analyte forms an irreversible couple and is titrated with a reversible titrant couple (e.g.,  $\text{S}_2\text{O}_3^{2-} + \text{I}_2$ ) the potential decreases at the endpoint (Figure 4B). For the titration of a reversible analyte with an irreversible titrant another curve is obtained (Figure 4C).

### Automated Potentiometric Titrations

The most important component of an automated potentiometric titrator is the potentiometer. In simpler titrimeters, ordinary burettes are used to deliver the titrant and are equipped with automatic valves. Usually piston (syringe) burettes that allow delivery of the titrant with predetermined constant or variable speed are employed. Most automatic titrators have built-in electronic switches to perform various duties and include one or more derivatizing circuits enabling the location of the endpoint. The automated potentiometric titrators available in the market may be divided into four groups: (1) automatic titration curve recording instruments; (2) preset endpoint titrators, in which the flow of the titrant is stopped when the potential of the indicator electrode reaches a preselected value; (3) first- or



**Figure 4** Titration curves for potentiometric titration using two polarized electrodes: (A) analyte and titrant reversible systems; (B) irreversible analyte, reversible titrant; (C) reversible analyte, irreversible titrant.

second-derivative titrators that continue titration until  $dE/dV = \max$  or  $d^2E/dV^2 = 0$ ; and (4) automatic titrimetric analyzers, highly automated and suitable for routine analysis. These last instruments take a sample, dispense the solution, refill the burette, and rinse the cell and electrodes automatically. They also calculate the result and produce a printout or a graph.

Some automatic potentiometric titrators (continuous titrators) carry out the analysis continuously. The flow rate of the titrant varies according to the signal obtained from the electrode. Once a steady state has been attained, the flow rate of the titrant becomes the measure of the concentration of the analyte in the sample. If the potential of the indicator electrode and the flow rate of the titrant are recorded as function of time one obtains the record of the variation of analyte concentration with time.

Most automatic titrators can be utilized as pH-stats or potentiostats, with the purpose of maintaining the pH or redox potential of a solution at a given value.

Automatic potentiometric titrations are widely used nowadays in industrial and research laboratories. Most of the procedures used in normal titration need only slight alterations to make them suitable for automatic work.

Potentiometric endpoint detection is frequently used in automatic constant current coulometric titrators (coulometric titrations).

## Applications

Many different types of indicator electrodes can be used as endpoint indicators in potentiometric titrations. For example, an acid–base titration can be performed with a glass electrode as endpoint detector instead of using colored indicators, or chloride ions can be titrated with silver(I) using a chloride-ion- or silver-ion-selective electrode.

### Acid–Base Titrations

Acid–base potentiometric titrimetry will always play an important role in quantitative analysis because of the extremely large number of inorganic, organic, and biochemical compounds that possess acidic or basic properties. Potentiometric titrations of strong acids and bases with suitable strong titrants are relatively simple, since there is a relatively large and abrupt change in pH at the endpoint. As the acid (or base) titrated becomes weaker, the sharpness and magnitude of the endpoint decrease. A similar situation occurs when the concentration of the substance titrated (and the titrant) is decreased. If 0.1% accuracy is needed in titration of strong acid or base, the concentration of the substance titrated should be greater than  $\sim 3 \times 10^{-4} \text{ mol l}^{-1}$ , but if 1% accuracy is sufficient, the concentration can be decreased by one order of magnitude. For the titration of a weak acid with a strong base, the product of the acid concentration and its dissociation constant,  $K_a$ , should be greater than  $\sim 10^{-7}$  for 0.1% accuracy, or  $10^{-9}$  for 1% accuracy.

When the reaction equilibrium constants differ significantly (e.g.,  $\Delta pK > 4$ ) for comparable concentrations of two species, determination of both is possible from one titration run. As the reaction constants depend on the solvent properties, the change of solvent may be advantageous in subsequent titration of two analytes.

The range of applications of potentiometric titrations for determination of acids and bases is very wide, as illustrated by the following examples. Carbonate, hydrogencarbonate, and hydroxide ions are all bases that can be titrated with a strong acid such as hydrochloric acid. The most popular method for determination of nitrogen, which is found in many important substances such as proteins, fertilizers, drugs, pesticides, natural waters, is the Kjeldahl method, based on the conversion of the bound nitrogen to ammonia, which is then separated by distillation and determined by titration with hydrochloric

acid as the titrant. A large variety of esters can be determined by a potentiometric acid–base back-titration procedure, where an excess of potassium hydroxide added for saponification of the ester is determined. One very interesting acid–base titration is the determination of metal ions by the titration of hydrogen ions released by a complexation reaction.

The apparent acidity or basicity of a compound is strongly dependent on the solvent used. For example, phenol is too weak an acid for titration in aqueous solution. Very strong acids such as hydrochloric and perchloric acids cannot be titrated individually in water. However, such titrations can be performed successfully when appropriate nonaqueous solvents are used. Bases too weak to be titrated in water can be

titrated in protogenic solvents (solvents more acidic than water), for example, acetic acid. In this solvent, amines, medicinal sulfonamides, and most common alkaloids such as caffeine can be titrated with perchloric acid dissolved in the same solvent. Acids too weak to be titrated in water appear much stronger in protophilic solvents (solvents more basic than water). For example, ethylenediamine can be employed as a solvent for titration of phenols. Some solvents permit differentiation (or stepwise titration) of a series of acidic or basic species that, in water, titrate either together or not at all. This group of solvents includes dioxan, ketones, and hydrocarbons. For example, perchloric, hydrochloric, and acetic acids, and phenol, can be titrated in 4-methyl-2-pentanone solvent

**Table 1** Examples of acid–base titrations in nonaqueous solvents

<i>Solvent</i>	<i>Analyte</i>	<i>Titrant</i>
Methanol, ethanol	Inorganic and carboxylic acids	Cyclohexylamine in methanol or 2-propanol, CH <sub>3</sub> OK in methanol
2-Propanol	Carboxylic acids	(C <sub>4</sub> H <sub>9</sub> ) <sub>4</sub> NOH in 2-propanol
2-Methyl-2-propanol	Acid mixtures	(C <sub>4</sub> H <sub>9</sub> ) <sub>4</sub> NOH in 2-methyl-2-propanol
Acetic acid	Amines, aminoacids, salts, weak acids, sulfonamides	HClO <sub>4</sub> in dioxan or acetic acid
Dimethylformamide	Carboxylic acids, phenols, ammonium salts	CH <sub>3</sub> OK in methanol or benzene
Glycol + hydrocarbons	Salts of weak acids, amines	HClO <sub>4</sub> in dioxan
Benzene + methanol	Carboxylic acids, acid anhydrides	KOH in methanol, CH <sub>3</sub> OK in benzene + methanol
Ethylene diamine	Weak acids, phenols	(C <sub>4</sub> H <sub>9</sub> ) <sub>4</sub> NOH in 2-methyl-2-propanol
Acetonitrile	Amines, carboxylic acids	HClO <sub>4</sub> in dioxan, CH <sub>3</sub> OK in acetonitrile
Chloroform	Acids, alkaloids	C <sub>2</sub> H <sub>5</sub> ONa in ethanol, HClO <sub>4</sub> in acetic acid
Nitromethane	Acid amides, urea, amine oxides	HClO <sub>4</sub> in dioxan
4-Methyl-2-pentanone	Acids, acid mixtures	(C <sub>4</sub> H <sub>9</sub> ) <sub>4</sub> NOH in 2-propanol
	Heterocyclic amines	HClO <sub>4</sub> in dioxan

**Table 2** Examples of typical potentiometric precipitation titrations of inorganic and organic compounds

<i>Analyte</i>	<i>Titrant</i>	<i>Indicator electrode</i>	<i>Conditions</i>
AsO <sub>4</sub> <sup>3−</sup>	Ag <sup>+</sup>	Ag	pH 9
Br <sup>−</sup>	Ag <sup>+</sup>	Ag or Ag-ISE	Dil. HNO <sub>3</sub> , 0.1 mol l <sup>−1</sup> KNO <sub>3</sub>
Ce <sup>3+</sup>	MoO <sub>4</sub> <sup>2−</sup>	Pt	In the presence of Ce <sup>4+</sup>
Cl <sup>−</sup>	Ag <sup>+</sup>	Ag or Ag-ISE	Dil. HNO <sub>3</sub> , 0.1 mol l <sup>−1</sup> KNO <sub>3</sub>
Cl <sup>−</sup>	Hg <sub>2</sub> <sup>2+</sup>	Hg	Dil. HNO <sub>3</sub>
F <sup>−</sup>	La <sup>3+</sup>	F-ISE	Neutral soln, 1:1 methanol
Fe(CN) <sub>6</sub> <sup>4−</sup>	Ag <sup>+</sup>	Ag	
Fe(CN) <sub>6</sub> <sup>3−</sup>	Ag <sup>+</sup>	Ag	
I <sup>−</sup>	Ag <sup>+</sup>	Ag or Ag-ISE	0.1 mol l <sup>−1</sup> HNO <sub>3</sub>
K <sup>+</sup>	Tetraphenylborate	K-ISE	
Li <sup>+</sup>	NH <sub>4</sub> F	F-ISE	95% methanol
S <sup>2−</sup>	Ag <sup>+</sup>	S-ISE	Alkaline soln
SO <sub>4</sub> <sup>2−</sup>	Pb <sup>2+</sup>	Pb-ISE	50% dioxan or 1:1 ethanol
Zn <sup>2+</sup>	K <sub>4</sub> Fe(CN) <sub>6</sub>	Pt	pH 3, in the presence of Fe(CN) <sub>6</sub> <sup>3−</sup>
Barbiturates	Ag <sup>+</sup>	Ag or Ag-ISE	Borate buffer, ethanol
Cysteine	HgCl <sub>2</sub>	Hg	pH 2–6
Dithioamide	Ag <sup>+</sup>	Ag or Ag-ISE	
Fatty acids salts	Ag <sup>+</sup>	Ag	Aqueous soln
Mercaptans	Hg(ClO <sub>4</sub> ) <sub>2</sub>	Hg	
Oxalates	Ag <sup>+</sup>	Ag or Ag-ISE	0.03 mol l <sup>−1</sup> NH <sub>3</sub>
Sulfonamides	Ag <sup>+</sup>	Ag	pH 7–8
Thioacetamide	Ag <sup>+</sup>	Ag	pH 8.5–9
Thiourea	Ag <sup>+</sup>	Ag	2 mol l <sup>−1</sup> NH <sub>3</sub>

to obtain an endpoint for each compound, using tetrabutylammonium hydroxide in 2-propanol as titrant. Examples of acid–base titrations in nonaqueous solvents are given in Table 1.

### Precipitation Titrations

The potentiometric detection of the endpoint of precipitation titrations is very often used because not many visual indicators are available, in particular when mixtures of analytes are titrated. Halides, cyanide, sulfide, chromate, mercaptans, and thiols can be titrated with silver nitrate, using the silver sulfide-based ISE. Also complex mixtures, such as sulfide, thiocyanide, and chloride ions, or chloride, bromide, and iodide ions, can be titrated potentiometrically with silver(I) ions. When the solubility of a compound formed during titration is too high, nonaqueous or mixed solvents are used, for example,

when sulfates are titrated with lead(II) using a lead-ion-selective electrode. In this case water–ethanol solvent is used. Table 2 presents some examples of potentiometric precipitation titrations.

### Compleximetric Titrations

Many metal ions (e.g., calcium, cadmium, aluminum, lead) can be titrated with standard ethylenediaminetetraacetic acid (EDTA) or other compleximetric titrants, using an appropriate indicator electrode. When no direct appropriate indicator electrode exists, the addition of indicator metal ions can permit a determination. For example, barium may be titrated with EDTA in the presence of silver–EDTA complex as an indicator reagent using a silver electrode. Examples of compleximetric titrations are given in Table 3.

**Table 3** Examples of typical compleximetric titrations

Indicator electrode	Titrant	Analyte	Conditions
Ag	EDTA	$\text{Ba}^{2+}$ , $\text{Ca}^{2+}$ , $\text{Mg}^{2+}$	pH 9–10.5, in the presence of small amount of $\text{Ag}^+$
Ca-ISE	EDTA	$\text{Ca}^{2+}$	pH 10–12
Cu-ISE	EDTA	$\text{Cd}^{2+}$ , $\text{Co}^{2+}$ , $\text{Mn}^{2+}$ , $\text{Ni}^{2+}$ , $\text{Pb}^{2+}$ , $\text{Zn}^{2+}$	pH 7–8
Cu-ISE	TETREN	$\text{Cu}^{2+}$ , $\text{Ni}^{2+}$ , $\text{Zn}^{2+}$	pH 4–6, acetate buffer
Cu-ISE	$\text{CuSO}_4$	Chelating agents	pH 5–6
F-ISE	NaF	$\text{Al}^{3+}$	pH 3.5–4.0
Hg	$\text{I}^-$	$\text{Hg}_2^{2+}$	
Pt	EDTA	$\text{Bi}^{3+}$ , $\text{In}^{3+}$ , $\text{Th}^{4+}$	0.2 mol l <sup>-1</sup> chloroacetic acid, in the presence of $\text{Fe}^{3+}/\text{Fe}^{2+}$
Pt	EDTA	$\text{Al}^{3+}$ , $\text{Cd}^{2+}$ , $\text{Cu}^{2+}$ , $\text{Ni}^{2+}$ , $\text{Pb}^{2+}$ , $\text{Zn}^{2+}$	Ammonium acetate, pH 5–6

TETREN, tetraethylenepentamine.

**Table 4** Typical potentiometric redox titrations of inorganic and organic compounds (in most cases a Pt electrode was used as the indicator electrode)

Analyte	Titrant	Conditions
$\text{As}^{3+}$	$\text{Ce}^{4+}$	4 mol l <sup>-1</sup> HCl, in the presence of ICl
$\text{C}_2\text{O}_4^{2-}$	$\text{MnO}_4^-$	0.2–1 mol l <sup>-1</sup> $\text{H}_2\text{SO}_4$ , 70°C
$\text{Ce}^{3+}$	$\text{MnO}_4^-$	1 mol l <sup>-1</sup> $\text{H}_2\text{SO}_4$
$\text{Co}^{2+}$	$\text{Fe}^{3+}$	Excess of 1,10-phenanthroline, pH 2–4
$\text{Cr}_2\text{O}_7^{2-}$	$\text{As}^{3+}$	20% $\text{H}_2\text{SO}_4$
$\text{H}_2\text{O}_2$	$\text{MnO}_4^-$	0.02 mol l <sup>-1</sup> $\text{H}_2\text{SO}_4$
$\text{I}^-$	$\text{MnO}_4^-$	0.1 mol l <sup>-1</sup> $\text{H}_2\text{SO}_4$
$\text{NO}_2^-$	$\text{Pb}(\text{CH}_3\text{COO})_4$	1 mol l <sup>-1</sup> NaCl
$\text{Ti}^{4+}$	$\text{Cr}^{2+}$	2 mol l <sup>-1</sup> $\text{H}_2\text{SO}_4$ , Hg electrode, absence of $\text{O}_2$
$\text{Ti}^+$	Chloramine T	3–4 mol l <sup>-1</sup> HCl, in presence of ICl
Aldehydes	$\text{KMnO}_4$	1.5 mol l <sup>-1</sup> $\text{H}_2\text{SO}_4$
Alkenes	$\text{Br}_2$ in $\text{CH}_3\text{COOH}$	In $\text{CH}_3\text{COOH}$
Amines	$\text{NaNO}_2$	Acidic solution
Ascorbic acid	$\text{Br}_2$ in $\text{CH}_3\text{COOH}$	In $\text{CH}_3\text{COOH}$
Carboxylic acids	$\text{KMnO}_4$	0.1 mol l <sup>-1</sup> NaOH
Cysteine	$\text{KIO}_4$	9 mol l <sup>-1</sup> HCl
Hydroquinone	$\text{Pb}(\text{CH}_3\text{COO})_4$ or $\text{Br}_2$ in $\text{CH}_3\text{COOH}$	$\text{Br}_2$ in $\text{CH}_3\text{COOH}$
Nitro-, Azo- comp.	$\text{Cr}(\text{CH}_3\text{COO})_2$	In $\text{CH}_3\text{COOH}$
Phenols	$\text{K}_2\text{Cr}_2\text{O}_7$	$\text{H}_2\text{SO}_4$
Quinone	$\text{Na}_2\text{S}_2\text{O}_3$	$\text{CH}_3\text{COOH} + \text{H}_2\text{O}$
Thiols	$\text{Pb}(\text{CH}_3\text{COO})_4$ or $\text{Br}_2$ in $\text{CH}_3\text{COOH}$	In $\text{CH}_3\text{COOH}$
Thiourea	Chloramine T	HCl + KBr



## Redox Titrations

Potentiometric titrations of oxidizing and reducing substances can be performed potentiometrically using inert electrodes, such as platinum or gold electrodes. Typical examples of applications include the titration of iron(II) ion with permanganate, titration of As(III) with bromine or iodine, and the determination of ascorbic acid with iodine. Many organic compounds such as azo, nitro, nitroso, carbonyl compounds, and quinones may also be determined. Because of their limited solubility in water those compounds are titrated in mixed or nonaqueous solvents. The Karl Fischer method of water determination based on the redox reaction between sulfur dioxide and iodine in the presence of a stoichiometric amount of water may be also mentioned. Water in liquid samples is titrated with titrant solutions containing iodine, sulfur dioxide, pyridine, and methanol as active solvent. Potentiometry, or more often bipotentiometry, is used for the detection of the endpoint of this titration. Some examples of potentiometric redox titrations are given in Table 4.

*See also:* **Coulometry.** **Indicators:** Acid–Base; Redox; Complexometric, Adsorption, and Luminescence Indicators. **Ion-Selective Electrodes:** Overview. **pH.** **Solvents.** **Titrimetry:** Overview.

## Further Reading

- Brett C and Brett AMO (1998) *Electroanalysis*. Oxford: Oxford Sciences.
- Gyenes J (1970) *Titrationen in nichtwässrigen Lösungsmitteln*. Stuttgart: Enke.
- Mascini M (1980) Uses of known addition, Gran's plot and the related methods with ion-selective electrodes. *Ion-Selective Electrode Reviews* 2: 17–71.
- Midgley D (1978) *Potentiometric Water Analysis*. New York: Wiley.
- Serjeant EP (1984) *Potentiometry and Potentiometric Titrations*. New York: Wiley.
- Svehla G (1978) *Automatic Potentiometric Titration*. Oxford: Pergamon Press.
- Vanysek P (1995) *Modern Techniques in Electroanalysis*. Chichester: Wiley.
- Vytras K (1985) Potentiometric titrations based on ion-pair formation. *Ion-Selective Electrode Reviews* 7: 77–164.
- Wang J (1994) *Analytical electrochemistry*. New York: VCH.

## Photometric

**C McGill**, Avecia Ltd, Grangemouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by I Wells, pp. 5336–5339, © 1995, Elsevier Ltd.

## Introduction

Titrimetry is a quantitative analytical technique which determines the volume of a solution of accurately known concentration that is required to react with the species being measured volume of the substance to be determined. The solution of accurately known strength is termed the titrant, and the substance being titrated is called the titrand.

For a reaction to be useful in titrimetry, the following conditions must be fulfilled:

1. The reaction must be simply explained by a chemical equation.
2. The reaction should be relatively fast.
3. There must be a change in the chemical or physical properties of the solution at the equivalence point (or end-point).

4. An indicator should be available that will identify the equivalence point of the titration (e.g., the indicator may undergo a color change or form a precipitation at the equivalence point).

However, there are instances when no visible indicator is available to identify the equivalence point of a titration. Also, some color changes can be very subtle, which can introduce subjectivity (and hence errors) in determining the equivalence point. In these situations, a more reliable detector than the human eye is required. In situations where the analyte, reagent, or titration product absorbs radiation, photometric measurements can be employed to determine the equivalence point of the titration.

## Principles

The end-point of a photometric titration is determined from absorbance measurements of the solution being analyzed. Beer's law for monochromatic light passing through a solution can be written as

$$A = \log(I_0/I) = \epsilon cl$$



## Redox Titrations

Potentiometric titrations of oxidizing and reducing substances can be performed potentiometrically using inert electrodes, such as platinum or gold electrodes. Typical examples of applications include the titration of iron(II) ion with permanganate, titration of As(III) with bromine or iodine, and the determination of ascorbic acid with iodine. Many organic compounds such as azo, nitro, nitroso, carbonyl compounds, and quinones may also be determined. Because of their limited solubility in water those compounds are titrated in mixed or nonaqueous solvents. The Karl Fischer method of water determination based on the redox reaction between sulfur dioxide and iodine in the presence of a stoichiometric amount of water may be also mentioned. Water in liquid samples is titrated with titrant solutions containing iodine, sulfur dioxide, pyridine, and methanol as active solvent. Potentiometry, or more often bipotentiometry, is used for the detection of the endpoint of this titration. Some examples of potentiometric redox titrations are given in Table 4.

*See also:* **Coulometry.** **Indicators:** Acid–Base; Redox; Complexometric, Adsorption, and Luminescence Indicators. **Ion-Selective Electrodes:** Overview. **pH.** **Solvents.** **Titrimetry:** Overview.

## Further Reading

- Brett C and Brett AMO (1998) *Electroanalysis*. Oxford: Oxford Sciences.
- Gyenes J (1970) *Titrationen in nichtwässrigen Lösungsmitteln*. Stuttgart: Enke.
- Mascini M (1980) Uses of known addition, Gran's plot and the related methods with ion-selective electrodes. *Ion-Selective Electrode Reviews* 2: 17–71.
- Midgley D (1978) *Potentiometric Water Analysis*. New York: Wiley.
- Serjeant EP (1984) *Potentiometry and Potentiometric Titrations*. New York: Wiley.
- Svehla G (1978) *Automatic Potentiometric Titration*. Oxford: Pergamon Press.
- Vanysek P (1995) *Modern Techniques in Electroanalysis*. Chichester: Wiley.
- Vytras K (1985) Potentiometric titrations based on ion-pair formation. *Ion-Selective Electrode Reviews* 7: 77–164.
- Wang J (1994) *Analytical electrochemistry*. New York: VCH.

## Photometric

**C McGill**, Avecia Ltd, Grangemouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by I Wells, pp. 5336–5339, © 1995, Elsevier Ltd.

## Introduction

Titrimetry is a quantitative analytical technique which determines the volume of a solution of accurately known concentration that is required to react with the species being measured volume of the substance to be determined. The solution of accurately known strength is termed the titrant, and the substance being titrated is called the titrand.

For a reaction to be useful in titrimetry, the following conditions must be fulfilled:

1. The reaction must be simply explained by a chemical equation.
2. The reaction should be relatively fast.
3. There must be a change in the chemical or physical properties of the solution at the equivalence point (or end-point).

4. An indicator should be available that will identify the equivalence point of the titration (e.g., the indicator may undergo a color change or form a precipitation at the equivalence point).

However, there are instances when no visible indicator is available to identify the equivalence point of a titration. Also, some color changes can be very subtle, which can introduce subjectivity (and hence errors) in determining the equivalence point. In these situations, a more reliable detector than the human eye is required. In situations where the analyte, reagent, or titration product absorbs radiation, photometric measurements can be employed to determine the equivalence point of the titration.

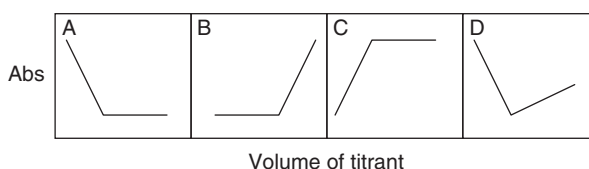
## Principles

The end-point of a photometric titration is determined from absorbance measurements of the solution being analyzed. Beer's law for monochromatic light passing through a solution can be written as

$$A = \log(I_0/I) = \epsilon cl$$

where  $A$  is the absorbance,  $I_0$  is the intensity of the incident light,  $I$  is the intensity of the transmitted light,  $\epsilon$  is the molar absorption coefficient,  $c$  is the concentration of the absorbing species, and  $l$  is the length of the light path through the absorbing medium. In photometric titrations, the path length,  $l$ , remains constant, and so the absorbance is directly proportional to the concentration of the absorbing species. The absorbance obtained at any point during the titration will be the sum of the absorbances of the titrant, titrand, and reaction product. As the end-point is reached, there should be a decreasing amount of titrand, an increasing amount of reaction product, and no titrant present in the titration vessel: therefore the absorbance is dependent on the absorption properties of the titrand and the reaction product. At the end-point of the titration, the concentration of the reaction product will have reached a maximum, and there should be no titrant or titrand present in the titration vessel, therefore the absorption at the end-point will be dependent on the optical properties of the reaction product. As more titrant is added once the end-point has been reached, any change in absorbance will be due solely on the absorption properties of the titrant as there should be no titrand present and the concentration of reaction product should remain constant. Therefore, in a titration in which the titrant, the titrand, or reaction product absorbs radiation, the plot of absorbance versus volume of titrant added will consist of two straight lines intersecting at the end-point (assuming that the volume change is small and the reaction goes to completion).

The shape of the curve obtained from a photometric titration is dependent on the optical properties of the titrant, reactant, and products of the reaction at the wavelength chosen for the measurement. Typical titration plots are shown in **Figure 1**. **Figure 1A** is typical of the curve obtained when a titrand that absorbs is converted into a product that does not absorb. The absorbance decreases as the amount of product increases and the amount of analyte decreases. However, after the end-point the product concentration remains constant, and therefore the absorbance measurement remains constant. **Figure 1B** shows the curve that would be obtained if the substance being titrated and the product were nonabsorbing and the titrant absorbs. The absorbance increases with excess titrant after the end-point. **Figure 1C** shows the curve that would be obtained where the product of the reaction absorbs, but both the titrant and the substance being titrated are non-absorbing. The absorbance increases with increasing product concentration, but after the end-point the product concentration is no longer rising, and so the



**Figure 1** Typical plots obtained from photometric titrations.

**Table 1** Selected examples of photometric titrations

Analyte	Titrant	Analytical wavelength (nm)	Corresponding graph in <b>Figure 1</b>
$\text{Fe}^{2+}$	$\text{K}_2\text{Cr}_2\text{O}_7$	335	B
$\text{Cu}^{2+}$	EDTA	745	C
2-Nitroaniline	$\text{HClO}_4$	464	A
$\text{I}^-$ (2nd end point)	$\text{Ce}^{4+}$	435	D

absorbance remains constant. Finally, **Figure 1D** is obtained when an absorbing titrant and an absorbing reactant produce a colorless product. The absorbance initially decreases as the analyte is consumed by the reaction, giving the nonabsorbing product. The absorbance reaches a minimum at the end-point and then increases due to the increasing amount of excess titrant, which absorbs. **Table 1** gives examples of photometric titrations that would exhibit titration curves similar to those shown in **Figure 1**.

In reality however, there is usually a small curvature at the end-point due to the reaction not going to full completion. This may be caused by the dissociation of reagents or products or the hydrolysis of weak acids or bases. The problem of curvature can be easily overcome by adding an excess of titrant and extrapolating from points remote from the end-point. It is also possible to improve the quality of the end-point determination by changing experimental parameters, e.g., pH and solvent/reagent concentrations. The best accuracy in determining an end-point of a photometric titration is when the extrapolated lines of the absorbance curve meet at an acute angle. Dilution effects caused by the addition of titrant must be corrected for. The absorbance values obtained are corrected by multiplying by  $(V_s + V_t)/V_s$ , where  $V_s$  is the volume of the sample and  $V_t$  is the volume of titrant added. If the correction for dilution effects is not carried out, the lines of the titration curve will bend towards the volume axis, resulting in an inaccurate determination of the volume of titrant added. To minimize dilution effects, a microsyringe can be used to add a relatively concentrated titrant to the sample being analyzed.

## Instrumentation

Photometric titrations can be carried out using a spectrophotometer that has been modified to permit the insertion of a titration vessel in the light path. However, it is much more common nowadays for photometric titrations to be carried out using automated apparatus consisting of a spectrophotometer or photometer, a titration vessel and stirrer, an automated burette, and software to control the titration and present the results. The titration vessels used are often cylindrical, with volumes of up to 100 ml. The titrant is incrementally added to the stirred vessel through the top and absorption readings taken after each titrant addition. Movements of the titration vessel during titrations must be avoided as inaccuracies will result due to deviations from Beer's Law. A diagram of a typical photometric titration cell is shown in **Figure 2**.

It is not essential to calibrate absorbance values to standards as the change in absorbance rather than the absolute absorbance is the important measure. This means that other species present that also absorb at the analytical wavelength or cause scattering of the light do not interfere with the end-point determination. However, interference can still occur if a species other than the analyte reacts with the titrant being used.

The accuracy and precision of photometric titrations are generally better than those obtained for direct absorbance measurements as the titration curve averages the data from a number of measurements. The relative precision of the end-point determination is typically around 0.5%. The accuracy is usually limited by the assumed concentration of the titrant, and therefore it is normal experimental procedure to titrate against a primary standard.

One of the major advantages of photometric titrations is the ease with which the sensitivity of the measurements can be changed simply by changing

the wavelength or the length of the cell path. When a titration reaction does not contain a self-indicating system, an indicator can be added in large amounts to provide a sufficient linear segment on the titration curve beyond the equivalence point.

## Applications

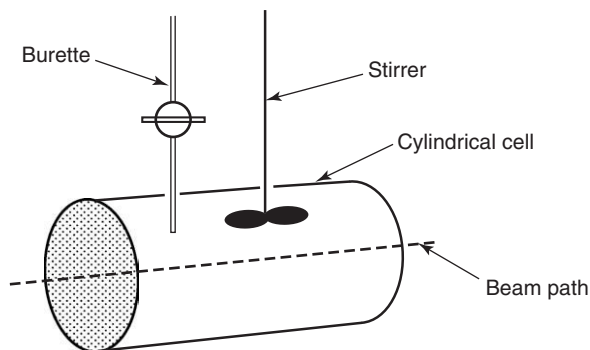
Photometric titrations are advantageous because the data from several measurements are used to determine the end-point, which results in a more accurate analysis. As only a change in absorbance is being measured, the presence of any other absorbing species will not interfere with the result obtained. Photometric titrations are popular for determining the amount of a metal ion in solution.

The concentration of copper(II) in a solution can be determined by titrating with ethylenediaminetetraacetic acid (EDTA). At a wavelength of 745 nm, the copper-EDTA complex has a much higher molar absorption coefficient than Cu(II), which means that a plot similar to **Figure 1C** would be obtained.

The concentration of iron(III) in solution can also be obtained by titration with EDTA. However, in this case salicylic acid is also added to the titration cell. Iron(III) forms a strongly colored complex with salicylic acid; however, the iron-EDTA complex has a much higher stability constant. As the titration is carried out, there is a very sharp disappearance of color at the end-point as the iron-EDTA complex replaces the iron-salicylic acid complex. This titration can be followed photometrically at 525 nm and produces a plot similar to **Figure 1A**.

Photometric titrimetry can be used for simultaneous determination of arsenic(III) and antimony(III) in a mixture. In acidic solutions, arsenic(III) can be oxidized to arsenic(V) and antimony(III) can be oxidized to antimony(V) by titration with a solution of potassium bromate and potassium bromide. The end-point of this titration can be observed by measuring the absorption at 326 nm. Initially, no change in the absorbance is observed until all the arsenic(III) has been oxidized to arsenic(V). The absorbance then decreases to a minimum at the end-point of the antimony titration, and then the absorbance increases again as excess titrant is added.

Similarly, the two end-points in the determination of *meta*- and *para*-nitrophenol can be determined by titrating with sodium hydroxide. This titration can be monitored by measuring the absorption of the phenoxide ion at 545 nm. The *para*-isomer is the stronger of the two acids and is neutralized first. There is a slight change in the slope from slightly positive gradient to a more positive gradient at this first end-point, followed by the absorbance



**Figure 2** A typical cylindrical titration vessel for use in photometric titrimetry.

remaining constant after the end-point of the *meta*-isomer neutralization. This titration would not be possible to follow using a visual indicator as the color change as the first end-point is too subtle.

Oxidizing agents such as permanganate and dichlorate absorb strongly in the visible region of the electromagnetic spectrum. This allows them to be used for photometric titrations. An example is the determination of iron(II) concentration by titrating with dichromate and measuring the absorbance at 335 nm.

Acid/base titrations can be monitored photometrically if a suitable indicator is added to the titration vessel. It is this indicator that causes the sharp change in the absorption profile at the end-point.

Photometric titrations have also been adapted for determining the end-point of precipitation titrations. The suspended solid produced by the titration reaction has the effect of reducing the energy of the radiation due to scattering, and so titrations can be carried out till a state of constant turbidity has been obtained.

**See Also:** **Indicators:** Acid–Base; Redox; Complexometric, Adsorption, and Luminescence Indicators. **Optical Spectroscopy:** Detection Devices. **Spectrophotometry:** Overview; Inorganic Compounds; Organic Compounds; Pharmaceutical Applications. **Titrimetry:** Overview; Potentiometric.

## Further Reading

Fifield FW and Kealey D (1995) *Principles and Practice of Analytical Chemistry*, 4th edn. Glasgow: Blackie Academic and Professional.

Headridge JB (1961) *Photometric Titrations*. Oxford: Pergamon Press.

Jeffery GH, Basset J, Mendham J, and Denney RC (1989) *Vogel's Textbook of Quantitative Chemical Analysis*, 5th edn. Harlow: Longman Scientific & Technical.

Skoog DA, Holler FJ, and Nieman TA (1998) *Principles of Instrumental Analysis*, 5th edn. Philadelphia: Harcourt Brace College Publishers.

Willard HH, Merritt LL, Dean JA, and Settle FA Jr. (1988) *Instrumental Methods of Analysis*, 7th edn. California: Wadsworth Publishing Company.

# TOILETRIES

See **COSMETICS AND TOILETRIES**

# TOTAL ANALYTICAL SYSTEMS, MICROFABRICATION

See **MICRO TOTAL ANALYTICAL SYSTEMS**

# TOXINS

Contents

**Mycotoxins**

**Neurotoxins**

## Mycotoxins

**S Dragacci**, Agence Française de Sécurité Sanitaire des Aliments (French Agency for Food Safety), Maisons-Alfort, France

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by M O Moss, pp. 1490–1499, © 1995, Elsevier Ltd.

## Introduction

Mycotoxins are fairly low relative-molecular-mass secondary metabolites of filamentous fungi that are toxic for humans and domesticated animals, causing economic losses by illness and death of animals when contaminated food or animal feed is consumed. There are many toxic metabolites produced by fungi, but only a fraction of these are currently considered to be important in foods. Similarly, a very broad

of the total volume. The acetone is left for 30 min to destroy any potentially harmful but reactive intermediates formed by the hypochlorite and then washed thoroughly. It is known, for example, that aflatoxin B1 can form a carcinogenic derivative if treated with hypochlorite alone.

A very detailed discussion of laboratory decontamination and destruction of aflatoxins in laboratory wastes has been prepared for the International Agency for Research on Cancer.

**See also:** **Food and Nutritional Analysis:** Mycotoxins. **Immunoassays, Techniques:** Enzyme Immunoassays. **Liquid Chromatography:** Food Applications. **Mass Spectrometry:** Food Applications. **Thin-Layer Chromatography:** Overview; Principles.

## Further Reading

- Betina V (1993) In: Betina V (ed.) *Chromatography of Mycotoxins: Techniques and Application*, pp. 141–233. Amsterdam: Elsevier.
- Chu FS (1995) Mycotoxin analysis. In: *Analyzing Food for Nutrition Labeling and Hazardous Contaminants*, pp. 283–331. New York: Dekker.
- Codex Alimentarius: <http://www.codexalimentarius.net>.
- Dorner JW (1996) Mycotoxins in food: Methods of analysis. In: Nollet LML (ed.) *Handbook of Food Analysis, Residues and other Food Components Analysis*, ch. 26, pp. 1089–1146.
- FAO (Food and Agriculture Organization) (1997) *Worldwide Regulations for Mycotoxins*, Food and

Nutrition Paper, 1995 – a compendium no. 64. Rome: FAO.

- IARC (International Agency for Research on Cancer) Scientific Publications (1980) Laboratory decontamination and destruction of aflatoxins B1, B2, G1, G2 in laboratory wastes. In: Castegnaro M, Hunt DC, Sansone EB, *et al.* (eds.). Lyon: IARC Publications.
- IARC (International Agency for Research on Cancer) Scientific Publications (1991) Laboratory decontamination and destruction of carcinogens in laboratory wastes: Some mycotoxins. In: Castegnaro M, Barek J, Freymy JM, *et al.* (eds.). Lyon: IARC Scientific Publications.
- IARC (International Agency for Research on Cancer) Scientific Publications (1993) *Monograph on the Evaluation of Carcinogenic Risk to Humans. Some Naturally Occurring Substances: Food Items and Constituent Heterocyclic Aromatic Amines and Mycotoxins*, vol. 56, IARC edition. Lyon: IARC.
- ISO (1998) ISO 5725-1 to 5: *Accuracy (Trueness and Precision) of Measurement Methods and Results*.
- Miller JD and Trenholm HL (1994) *Mycotoxins in Grain. Compounds Other than Aflatoxin*. St Paul, MN: Eagan Press.
- Pfohl-Leschkowicz A, Petkova-Bocharova T, Chernozemsky IN, and Castegnaro M (2002) Balkan endemic nephropathy and the associated urinary tract tumours: Review on etiological causes, potential role of mycotoxins. *Food Additives and Contaminants* 19(3): 282–302.
- Samson RA, Hoekstra ES, Frisvad JC, and Filtenborg O (2000) *Introduction to Food-Borne Fungi*. Baarn, The Netherlands: C.B.S.
- Van Egmond HP (1989) *Mycotoxins in Dairy Products*. London: Elsevier.

## Neurotoxins

**M-T Fernández-Sánchez, A Ferrero, A Pérez-Gómez, and A Novelli**, University of Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The term neurotoxin applies to a variety of substances that can affect the nervous system producing a broad variety of effects usually harmful to human health, although in some cases these substances can be of therapeutic use. Neurotoxic compounds are occasionally present in food and beverages, and consumption of contaminated products may therefore contribute to the etiology of certain psychiatric disorders or neurodegenerative diseases in humans. Three different main groups of

food-related neurotoxins may be considered: (1) metal and organometallic compounds (e.g., mercury and methylmercury, lead, cadmium, manganese, aluminum, and alkyltin); (2) marine and freshwater toxins of algal origin; and (3) toxicants purposefully added or resulting from industrial activities and accumulated in food including organophosphates and polychlorinated biphenyls (PCBs). In addition to food-related toxic compounds, poisons from venomous species may also contain potent neurotoxins. The analysis of metals and organometallic compounds as well as of organophosphates and PCBs are covered by other articles within the Encyclopedia. The following is a discussion of a variety of methods for the analysis and detection of different types of marine and freshwater neurotoxins of algal origin, and for some neurotoxic compounds from snake, scorpion, and spider venoms.



## Marine and Freshwater Toxins of Algal Origin

Algal toxins affect human health primarily through the consumption of seafood. Filter-feeding shellfish, zooplankton, and herbivorous fishes serve as vectors, either directly or through food web transfer, of sequestered toxins to higher trophic levels. According to their chemical structure and pathological effects, algal toxins have been classified into five major different groups, causing five different human seafood poisoning syndromes: paralytic shellfish poisoning (PSP), amnesic shellfish poisoning (ASP), ciguatera fish poisoning (CFP), neurotoxic shellfish poisoning (NSP), and diarrhetic shellfish poisoning (DSP). Toxins responsible for PSP, ASP, CFP, and NSP mediate their effects by diverse, highly specific interactions with ion channels involved in neurotransmission. DSP toxins are inhibitors of ser/thr protein phosphatases, critical components of signaling cascades in eukaryotic cells, including neurons. However, DSP toxins cannot be strictly referred as neurotoxins and will not be considered in this discussion.

In addition to seafood neurotoxins, we will also consider cyanobacterial toxins, which primarily impact on human health through drinking water contamination.

**Table 1** summarizes the biological activities of the main marine and freshwater algal neurotoxins, and the analytical methods most frequently used for their determination in shellfish tissue or algal extracts. The regulatory limit established for each toxin group is also indicated.

### Seafood Neurotoxins

The principal regulatory method to protect consumers from marine toxins in food is the mouse bioassay. It measures the time to death after intraperitoneal injection of a toxic extract, and toxicity is expressed as the amount of toxin (usually in micrograms) or toxin equivalents (eq) per gram (or 100 g) of shellfish tissue, standardized against a reference toxin solution. As it is a reliable indicator of human oral toxicity, the mouse bioassay is obligatory prior to commercialization of seafood in many countries. However, this biological test does not allow for specification of toxins or studies on the formation and effect of individual toxins. In addition, it has a high variability ( $\pm 20\%$ ) and low sensitivity, lacks specificity, and often has detection limits too near to the accepted threshold. Yet, the major limitation of the mouse bioassay is of ethical and political nature due to the use of live animals.

*In vitro* biological techniques are very promising approaches to detection of these phycotoxins in

extracts of shellfish tissues and toxic microalgae. Depending on the nature of the measured response, *in vitro* bioassays may be functional or nonfunctional. The latter category includes receptor binding assays and immunoassays, whereas the former includes activation/inhibition of enzymatic and cytotoxicity assays, all performed in live cells. Functional assays facilitate more information about the potential risk of a given sample because toxin analogs that may be not recognized by antibodies, or have too low receptor affinity to be detected by receptor-binding methods, may still be evaluated for their functional activity. Cytotoxicity assays using different cultured cell types have been developed in the last years. In particular, cultured neurons constitute a very suitable system because they are exquisitely sensitive to any compound that alters their physiological equilibrium, and allow any type of biochemical process activated by toxins to be monitored starting from steady-state conditions.

**Paralytic shellfish poisoning** PSP is the most widespread algae-derived shellfish poisoning. PSP toxins block voltage-dependent sodium channels, inhibiting nerve conduction and eventually leading to muscular paralysis and respiratory failure. The same mechanism is shared by a different toxin, tetrodotoxin, one of the most lethal seafood toxins, associated with the consumption of most pufferfish species. Chemically, PSP toxins are a group of alkaloid compounds collectively called saxitoxins (**Figure 1**).

Liquid chromatography (LC) is a powerful tool for the analysis of PSP toxins. These toxins have only a weak natural chromophore and must be modified prior to detection. The postcolumn reaction system is the usual method of choice, in which the effluent from the LC column is mixed with a chemical oxidant (periodic acid, pH 9.0), passed through a reaction coil, and acidified to form a fluorescent product. Oxidation can be also done electrochemically and, if a coulometric electrochemical detector is used, the oxidation results in a signal that can be used for the direct electrochemical detection of the toxins (ECD). However, this method requires long equilibration times to achieve steady background voltages and cannot be used with gradient elution profiles. Therefore, the use of direct ECD is limited to research laboratories working with relatively clean sample matrices.

To exclude inaccuracies, mainly due to interference between saxitoxin congeners, the use of two different LC methods is recommended. Positive findings should be confirmed by application of mass spectrometry (MS) but unfortunately, phosphate, ion-pair formers, as well as periodic acid from the



**Table 1** Summary of characteristics and methods for analytical detection of marine and freshwater neurotoxins of algal origin

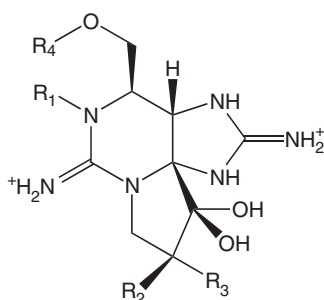
<i>Toxin group and principal toxins</i>	<i>Pharmacology</i>	<i>Method</i>	<i>Approximate detection limit</i>
<b>PSP</b>			Regulatory Level: $0.8 \mu\text{g g}^{-1}$
Saxitoxins	Block of voltage-sensitive sodium channels	Mouse bioassay	$0.35 \mu\text{g g}^{-1}$
Gonyautoxins		Liquid chromatography Capillary electrophoresis Enzyme-linked immunosorbent assay Reverse cytotoxicity assay	$0.04 \mu\text{g g}^{-1}$ $0.5 \mu\text{g g}^{-1}$ $3 \text{ ng g}^{-1}$ $40 \text{ ng g}^{-1}$
<b>ASP</b>			Regulatory Level: $20 \mu\text{g g}^{-1}$
Domoic acid and isomers	Potent selective agonist at the non-NMDA subtype of glutamate receptors	Mouse bioassay	$40 \mu\text{g g}^{-1}$
		Thin-layer chromatography	$10 \mu\text{g g}^{-1}$
		Liquid chromatography	$20\text{--}30 \text{ ng g}^{-1}$ (UV) $15 \text{ pg g}^{-1}$ (fluorescence)
		Capillary electrophoresis	$150 \text{ ng g}^{-1}$
		Mass spectrometry	$1 \mu\text{g g}^{-1}$
		Enzyme-linked immunosorbent assay Direct cytotoxicity assay	$38 \text{ ng g}^{-1}$ $21 \mu\text{g g}^{-1}$
<b>CFP</b>			Regulatory Level: $50 \text{ pg g}^{-1}$
Ciguatoxins	Activation of voltage-sensitive sodium channels	Mouse bioassay	$0.5 \text{ ng g}^{-1}$
		LC tandem mass spectrometry	$0.04 \text{ ng g}^{-1}$ (P-CTX-1) $0.1 \text{ ng g}^{-1}$ (C-CTX-1)
		Immunoassays	$1 \text{ ng g}^{-1}$ (EIA) $0.05 \text{ ng g}^{-1}$ (S-PIA)
<b>NSP</b>			Regulatory Level: $0.8 \mu\text{g g}^{-1}$
Brevetoxins	Activation of voltage-sensitive sodium channels	Liquid chromatography-mass spectrometry	$6\text{--}200 \mu\text{g l}^{-1}$
		Radioimmunoassay	$\text{EC}_{50}$ : $1.2 \pm 0.2 \text{ nmol l}^{-1}$
		Receptor assays	$\text{EC}_{50}$ : $4.3 \pm 1.5 \text{ nmol l}^{-1}$
		Direct cytotoxicity assays	$< 0.5 \mu\text{g g}^{-1}$
<b>CYANOBACTERIAL TOXINS<sup>a</sup></b>			Regulatory Level: $3 \text{ ng g}^{-1}$
Anatoxin-a	Inhibition of the nicotinic and muscarinic acetylcholine receptor	Mouse bioassay	$5 \mu\text{g g}^{-1}$
Homoanatoxin-a		Liquid chromatography Thin-layer chromatography Liquid chromatography-mass spectrometry	$0.02\text{--}20 \text{ ng g}^{-1}$ $10 \mu\text{g g}^{-1}$ $0.02\text{--}0.1 \text{ ng g}^{-1}$
Anatoxin-a (S)	Acetylcholinesterase inhibition	LC tandem mass spectrometry	$50 \text{ pg g}^{-1}$
		Gas chromatography-mass spectrometry	$< 0.5 \text{ ng g}^{-1}$
		Receptor assays	$\text{EC}_{50}$ : $0.2\text{--}0.7 \mu\text{mol l}^{-1}$
		Acetylcholinesterase inhibition assay	$0.17\text{--}1 \text{ ng g}^{-1}$ (anatoxin-a(S))

<sup>a</sup>Cyanobacterial toxins should also include saxitoxins. However, information about saxitoxins has been included in the PSP section (first line of this table).

Abbreviations: UV, ultraviolet; EIA, enzyme immunoassay; S-PIA, solid-phase immunobead assay; P-CTX-1, Pacific ciguatoxin-1; C-CTX-1, Caribbean ciguatoxin-1;  $\text{EC}_{50}$ , half maximally effective concentration.

postcolumn derivatization unit often prevent an efficient application of the LC–MS coupling. An LC method has been developed that allows direct coupling of the liquid chromatograph with the mass

spectrometer. Separation of the PSP toxins was obtained on a weak cation-exchange resin using an aqueous eluent with ammonium acetate as the only additive. In the case of a parallel application of



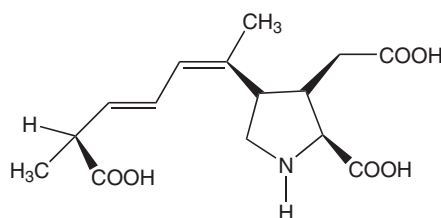
Toxin	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>
<b>Carbamate toxins</b>				
STX	H	H	H	CONH <sub>2</sub>
NeoSTX	OH	H	H	CONH <sub>2</sub>
GTX-1	OH	H	OSO <sub>3</sub> <sup>-</sup>	CONH <sub>2</sub>
GTX-2	H	H	OSO <sub>3</sub> <sup>-</sup>	CONH <sub>2</sub>
GTX-3	H	OSO <sub>3</sub> <sup>-</sup>	H	CONH <sub>2</sub>
GTX-4	OH	OSO <sub>3</sub> <sup>-</sup>	H	CONH <sub>2</sub>
<b><i>n</i>-Desulfocarbamoyl (sulfamate) toxins</b>				
GTX-5 (B1)	H	H	H	CONHSO <sub>3</sub> <sup>-</sup>
GTX-6 (B2)	OH	H	H	CONHSO <sub>3</sub> <sup>-</sup>
C1 (epiGTX-8)	H	H	OSO <sub>3</sub> <sup>-</sup>	CONHSO <sub>3</sub> <sup>-</sup>
C2 (GTX-8)	H	OSO <sub>3</sub> <sup>-</sup>	H	CONHSO <sub>3</sub> <sup>-</sup>
C3	OH	H	OSO <sub>3</sub> <sup>-</sup>	CONHSO <sub>3</sub> <sup>-</sup>
C4	OH	OSO <sub>3</sub> <sup>-</sup>	H	CONHSO <sub>3</sub> <sup>-</sup>
<b>Decarbamoyl toxins</b>				
dcSTX	H	H	H	H
dcneoSTX	OH	H	H	H
dcGTX-1	OH	H	OSO <sub>3</sub> <sup>-</sup>	H
dcGTX-2	H	H	OSO <sub>3</sub> <sup>-</sup>	H
dcGTX-3	H	OSO <sub>3</sub> <sup>-</sup>	H	H
dcGTX-4	OH	OSO <sub>3</sub> <sup>-</sup>	H	H

**Figure 1** Chemical structure of PSP toxins. STX, saxitoxin; GTX, gonyautoxin; dcGTX, decarbamoyl gonyautoxin.

LC-MS and fluorometric detection, electrochemical postcolumn derivation is suggested to avoid contamination of the ion source with chemical oxidation substances. This LC method is also well suited for the production of PSP standards from contaminated algae and mussels.

Capillary electrophoresis (CE) with ultraviolet detection is also a useful method for the separation and determination of PSP toxins. Separation conditions have been developed allowing the identification of products in crude enzymatic digestions using CE combined with tandem MS. However, CE separations demand highly purified extracts to obtain reproducible values and, also, the handling of very small volumes for injection results in limits of detection ~10 times higher compared to LC with fluorescence detection.

Total PSP toxicity after LC or CE separation must be calculated by summing up the individual concentrations of the respective PSP toxins quantified from



**Figure 2** Structure of domoic acid.

their corresponding peaks in the chromatograms. The values obtained by these chemical methods usually agree well with the results of the mouse bioassay.

Immunoassays, including radioimmunoassays (RIA) and enzyme-linked immunoabsorbent assay (ELISA), can be also used for the detection of saxitoxins. However, the presence of cross-reactions with lower binding specificity and the lack of response of all toxins within the saxitoxin family limit the use of these systems.

Based on the ability of PSP toxins to block sodium channels, a reverse toxicological assay has been proposed using cultured cerebellar neurons in which protection by saxitoxins from neurodegeneration induced by the sodium channel activator veratridine is measured. PSP content is calculated by comparison with known concentrations of purified saxitoxin, and values obtained with this method show very good correlation with values from LC and mouse bioassay.

**Amnesic shellfish poisoning** The main toxin associated with ASP is the tricarboxylic amino acid domoic acid (Figure 2), responsible for a human intoxication characterized by a widespread neurological dysfunction, chronic loss of memory, and motor neuropathy. The potent neurotoxic actions of domoic acid are due to the activation of a subtype (non-NMDA) of glutamate neurotransmitter receptors, and the subsequent depolarization leading to seizures and neuronal death.

Domoic acid can be analyzed semiquantitatively using thin-layer chromatography, which allows the analysis of multiple samples in a single run at relatively low cost, advantages particularly useful when applied to routine screening procedures of contaminated shellfish.

For quantitative purposes, the most common method is LC with ultraviolet detection, facilitated by the diene chromophore at 242 nm. When extreme sensitivity is required, ultraviolet absorbance is replaced by fluorescence detection of derivatized domoic acid. Prior to LC analysis, cleaning of sample is recommended. Usual cleaning procedures include aqueous methanol (1:1 v/v) extraction followed by a

solid-phase extraction or by passage of the supernatant through a strong anion-exchange column preconditioned with methanol, water, and extraction solvent. Effective separation can be accomplished using standard C18 analytical columns and isocratic elution. Acidic mobile phases are preferred, and in most cases a mixture of acetonitrile and water acidified with trichloroacetic or phosphoric acid is used.

Ion-spray LC-MS serves as an excellent confirmatory method for domoic acid. The positive-ion mass spectrum is characterized by an abundant protonated molecule ( $[M+H]^+$ ) at  $m/z$  312 with no significant fragmentation and, in order to achieve more structural information, tandem MS may be used to generate fragment ion spectra.

CE has been also successfully used for analysis of domoic acid and its isomers in seafood. For that purpose, crude homogenates should be previously cleaned-up using a strong anion-exchange column followed by a strong cation-exchange column, and then applied to a CE system interfaced to a variable wavelength ultraviolet detector operating at 242 nm. As mentioned for PSP toxins, the small injection volumes used makes the technique less sensitive than LC for shellfish samples. However, advances in technology such as the use of helium/cadmium laser in combination with CE to detect domoic acid that has been derivatized with dansyl chloride have led to unprecedented detection limits in the attomole ( $10^{-18}$  mol) range. Still, the major advantage of CE in routine applications remains the high degree of resolution for domoic acid and its isomers in a variety of shellfish homogenates.

A specific and sensitive direct cytotoxicity assay has been developed for domoic acid using cultured cerebellar neurons, consisting in the determination of neuronal death following 18–24 h exposure to domoic acid or analogs. Quantification is done using purified domoic acid as standard.

**Ciguatera fish poisoning and neurotoxic shellfish poisoning** CFP and NSP define two different syndromes caused by consumption of polyether compounds able to induce persistent activation of voltage-sensitive sodium channels. Interestingly, PSP toxins and tetrodotoxin (see above) antagonize the action of CFP and NSP toxins.

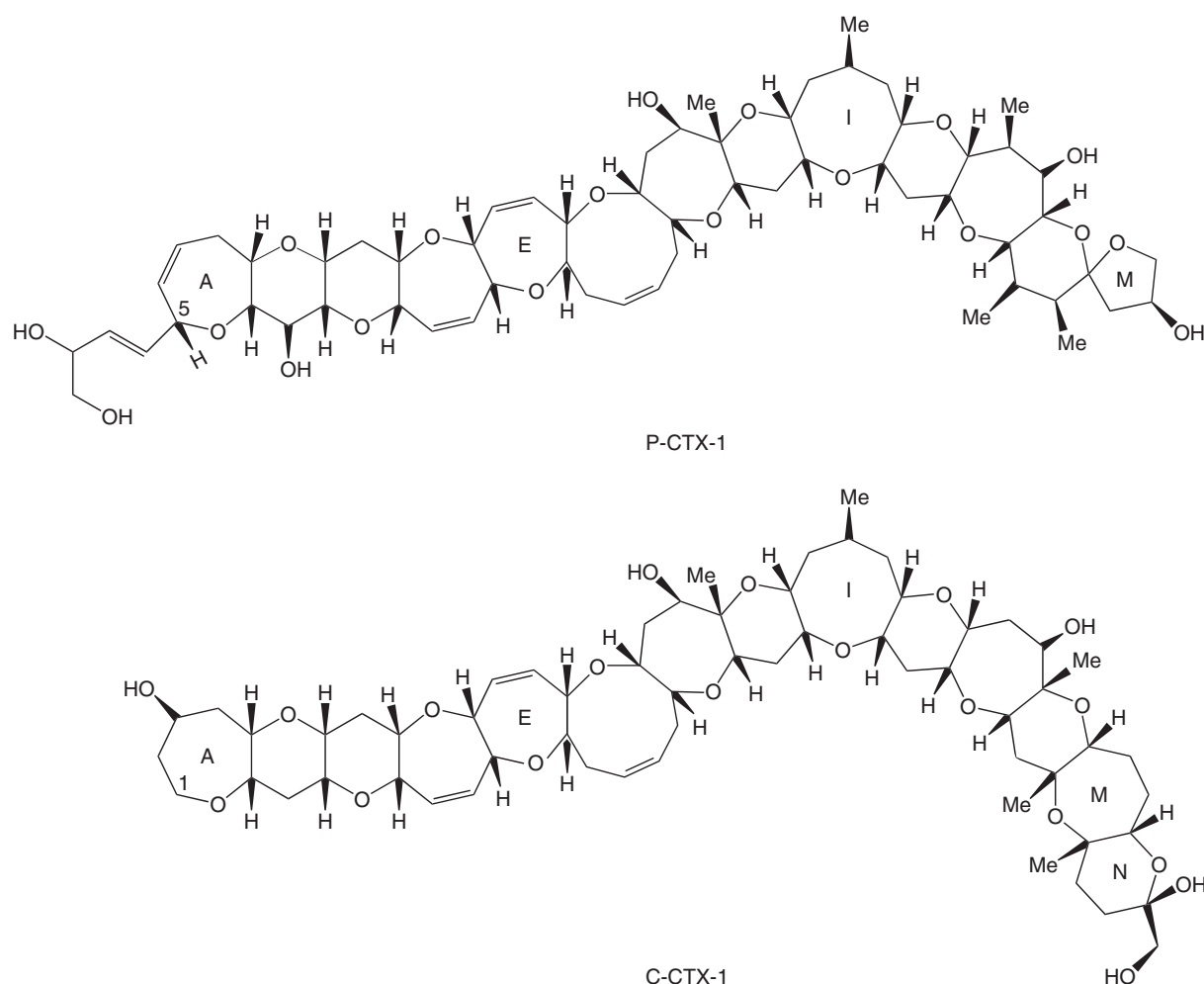
CFP is associated with the consumption of many species of tropical and subtropical fish contaminated by ciguatoxins, a family of lipid-soluble, heat-stable molecules (Figure 3) produced by the benthic dinoflagellate *Gambierdiscus toxicus*. Pacific ciguatoxin-1 (P-CTX-1) and Caribbean ciguatoxin-1 (C-CTX-1) are the major toxins contributing to CFP, and are usually considered the main

targets in any screening program. P-CTX-1 accumulates to levels between 0.1 and 5 ng g<sup>-1</sup> in the flesh of ciguateric fish. Based on a twofold risk factor, to secure public health it is necessary to reliably detect P-CTX-1 in fish flesh at 0.05 ng g<sup>-1</sup> and above. This is not possible using the mouse bioassay for P-CTX-1, which allows the detection of levels above 0.5 ng g<sup>-1</sup>. LC, MS, and LC-MS methods have been used to detect ciguatoxins in highly enriched extracts, but these approaches have insufficient sensitivity to detect clinically relevant toxin levels in crude extracts of fish. A robust gradient reverse-phase LC tandem MS method has been developed to determine CFP toxins using fish extracts spiked with pure P-CTX-1 or C-CTX-1 as internal standards.

Immunological detection of ciguatoxins and related polyethers has received particular attention compared to other marine toxins. The initial RIA and enzyme immunoassay employing a polyclonal sheep anti-ciguatoxin antibody revealed cross-reactivities between ciguatoxins and other polyether toxins, suggesting the need for monoclonal antibodies. The availability of monoclonal antibodies allowed for the development of stick enzyme immunoassay methods and solid-phase immunobead techniques (known as the paddle test), which successfully recognized toxins attached to correction fluid-coated bamboo sticks or paddles previously exposed to toxic fish tissues.

The toxins responsible for NSP are a group of at least 10 natural compounds known as brevetoxins (Figure 4), and produced by the unarmored marine dinoflagellate *Gymnodinium breve*. LC-MS methods have been proved successful for the analysis of brevetoxins in crude extracts of cultured *G. breve*. Indeed, LC-MS is a powerful technique that has been proposed as a 'universal' method for the analysis of most seafood toxins. Among advantages, LC-MS provides high selectivity and sensitivity, separation of complex mixtures, accurate quantitation, automation, legal acceptability for confirmation, and structural information for the identification of new toxins. As a drawback, costly equipment and operators with high level of expertise are required.

All brevetoxins can be detected by immunological techniques, including RIA as well as immunofluorescent tests and immunoabsorbant assays. The specific binding of brevetoxins and ciguatoxins at a particular site of the voltage-dependent sodium channel can be exploited and these receptors can be used in simple and effective receptor-based assays very much like antibody tests. The receptor assay works in membrane preparations, in microtiter plates, and in solubilized form, and can be done with tissue extracts or biological fluids as toxin source. The system should be also approachable by affinity techniques,



**Figure 3** Structure of ciguatoxins.

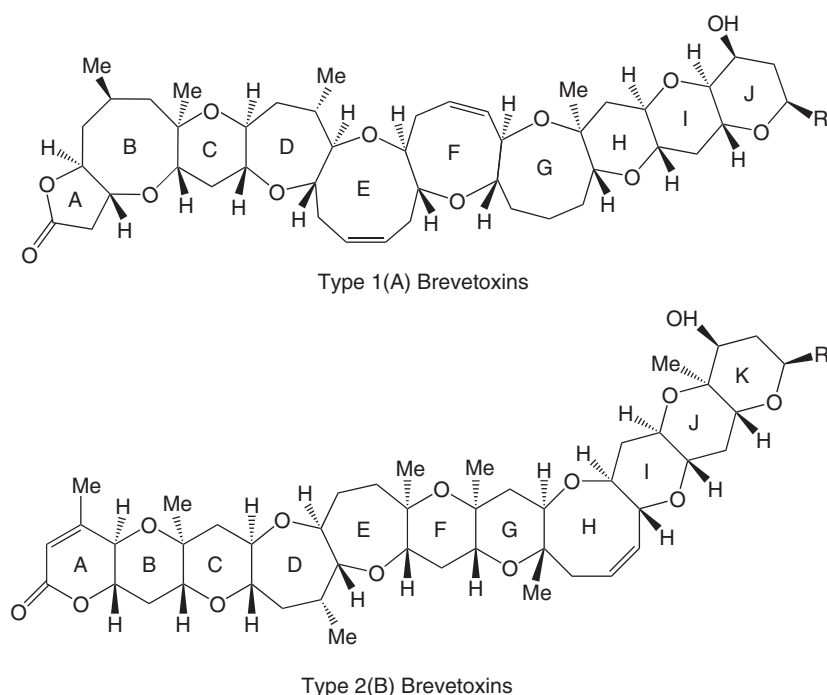
and it is expectable that derivatized toxins in radioactive, colorimetric, or fluorescent forms would be as useful as they are in immunological protocols. Since binding is the essential first step in the onset of toxicity, and more potent forms of the toxin bind more tightly to the receptor, the results obtained in receptor assays should properly reflect the toxicity of a sample.

Citotoxicity assays using murine neuroblastoma cells (Neuro-2A) have been proposed for the assessment of ciguatoxins and brevetoxins. The simplest method is based in the ability of metabolically active cells to reduce the tetrazolium compound MTT. After incubation with serial dilutions of test samples, cells are exposed to MTT and formation of color due to MTT reduction is read using a spectrophotometer. A new cytotoxicity assay is based on a c-fos-luciferase reporter gene that uses luciferase-catalyzed light generation as an endpoint and a microplate luminometer for quantification. After

incubation with toxins, cells are lysed and exposed to a mixture of luciferase and ATP, and the luminescence generated from each well is measured. Quantification of brevetoxins and ciguatoxins is done using known concentrations of the sodium channel activator veratridine as standard.

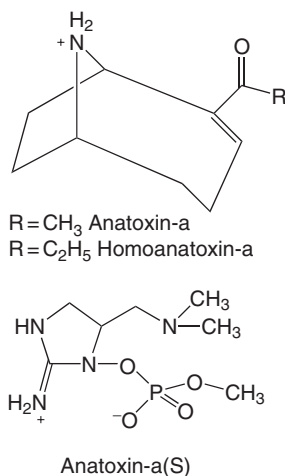
### Freshwater Neurotoxins

In freshwater, harmful algal blooms are mostly attributed to cyanobacteria, also known as blue-green algae. Cyanobacterial neurotoxins fall into three different categories: anatoxin-a and homoanatoxin-a, anatoxin-a(S), and saxitoxins (see **Figure 1** and **Figure 5** for chemical structures). As mentioned above, saxitoxins are produced by marine dinoflagellates and are known to concentrate in seafood and cause PSP in humans. Saxitoxins can be also found in cyanobacteria, where the presence of at least 19 different saxitoxins has been reported.



Type A(1) brevetoxins		Type B(2) brevetoxins	
Toxin	R	Toxin	R
PbTx1	$\text{CH}_2\text{C}(\text{=CH}_2)\text{CHO}$	PbTx2	$\text{CH}_2\text{C}(\text{=CH}_2)\text{CHO}$
PbTx7	$\text{CH}_2\text{C}(\text{=CH}_2)\text{CH}_2\text{OH}$	OxiPbTx2	$\text{CH}_2\text{C}(\text{=CH}_2)\text{COOH}$
PbTx10	$\text{CH}_2\text{CH}(\text{CH}_3)\text{CH}_2\text{OH}$	PbTx3	$\text{CH}_2\text{C}(\text{=CH}_2)\text{CH}_2\text{OH}$
		PbTx5	[PbTx2], $\text{C}_{37}$ O-Acetate
		PbTx6	[PbTx2], $\text{C}_{27-28}$ Epoxide
		PbTx8	$\text{CH}_2\text{COCH}_2\text{Cl}$
		PbTx9	$\text{CH}_2\text{CH}(\text{CH}_3)\text{CH}_2\text{OH}$

**Figure 4** Chemical structure of brevetoxins.



**Figure 5** Chemical structure of anatoxins.

The development of analytical methods for anatoxin-a has been facilitated by the fact that the structure of this toxin was already solved in the late 1970s and chemical synthesis yielded the pure

compound to be used as a standard. Moreover, no variations of the chemical structure of anatoxin-a have been found except for the homoanatoxin-a and nonbioactive degradation products of both of these compounds. The method of choice now depends mainly on the amount of toxin contained in the samples to be analyzed, and the equipment available. Thin-layer chromatography has been used for purification of toxins and also as an inexpensive screening method. LC methods with ultraviolet or fluorimetric detection have been used to quantitate anatoxin-a, and several gas chromatography (GC) protocols using different derivatization procedures have been also used for anatoxin-a and homoanatoxin-a. The combination of LC or GC with MS has greatly improved the identification of these toxins and made it possible to look for the degradation products. In addition to MS, different nuclear magnetic resonance (NMR) techniques have been very useful for solving the unknown structures of cyanobacterial neurotoxins.

The same methods used for analysis of marine samples can be applied to saxitoxins originating from cyanobacteria. As already described, the most common methods to quantitate saxitoxins are based on LC with postcolumn derivatization and fluorimetric detection.

## Neurotoxins from Venomous Animals

Poisoning by venomous animals, particularly snakes, scorpions, and spiders, represents an important health problem, particularly in developing countries, where it has been estimated to cause more than 150 000 deaths per year. On the other hand, these venoms represent an incredible source of biologically active substances, extremely useful as research tools in a variety of biochemical and pharmacological studies. In particular, they contain peptides that are highly specific ligands for selected subtypes of ion channels, and that have been very useful to elucidate the structure and physiological roles of potassium, sodium, and calcium channels in a variety of cell types, including neurons.

**Table 2** summarizes the characteristics and the techniques usually employed for the analytical

detection of the main types of neurotoxins present in snake, scorpion, and spider venoms in body fluids.

### Snake Venoms

Snake venoms include toxins showing high selectivity for subtypes of muscarinic acetylcholine receptors, which control a variety of processes including the modulation of the heart rate, control of motor systems, and the modulation of learning and memory, as well as toxins that block particular subtypes of voltage-dependent sodium or potassium channels in neurons.

Detection of snake toxins and toxin antibodies in body fluids remains very important for the identification of the biting species and the correct management of envenomation. As snake venoms consist of a complex mixture of pharmacologically active peptides and proteins, detection is usually approached by immunological techniques. Radioimmunoassay (RIA) using specific monoclonal antibodies has proved to be highly reliable and sensitive. However, the difficulty of handling radioisotopes and the need of elaborate equipment limit its application in

**Table 2** Biological activity and detection methods for neurotoxins from venomous animals

<i>Neurotoxin source and principal toxins</i>	<i>Pharmacology</i>	<i>Method</i>	<i>Approximate detection limit</i>
<i>Snake</i>	Highly specific ligands for subtypes of acetylcholine receptors	Radioimmunoassay	4 ng ml <sup>-1</sup> (urine)
<i>Peptides</i>	Block of selected subtypes of voltage-dependent potassium and sodium channels	Agglutination Assay Enzyme-linked immunosorbent assay Fluorescence immunoassay Avidin-biotin immunoassay Immunosensor	5 µg ml <sup>-1</sup> (serum) 2 ng ml <sup>-2</sup> –3 µg ml <sup>-1</sup> 1–100 ng ml <sup>-1</sup> 0.1 pg ml <sup>-1</sup> 100 pg ml <sup>-1</sup> 15–30 ng ml <sup>-1</sup>
<i>Scorpion</i>	Block of selected types of voltage-dependent sodium channels	Enzyme-linked immunosorbent assay	0.9–5 ng ml <sup>-1</sup>
<i>Peptides</i>	Block of some types of voltage-dependent potassium channels and chloride channels		
<i>Spider</i>	Selective action on voltage-dependent sodium, potassium and calcium channels	Enzyme-linked immunosorbent assay	10–20 ng ml <sup>-1</sup>
<i>Peptides</i>			
<i>High molecular weight proteins</i>	Induction of neurotransmitter release		
<i>Polyamines</i>	Antagonists of ionotropic glutamate receptors		



routine diagnosis in the field. In contrast, agglutination tests using purified rabbit antivenom IgG are inexpensive, rapid, and simple to perform, making them appropriate for use in rural health centers of developing countries, but because of low level sensitivity and instability of the coupling agents, these tests have received only limited attention. In terms of specificity, sensitivity, rapidity, and simplicity, ELISA methods appear to be the ideal systems for detecting venom and venom antibodies. Approaches to achieve species specificity have included the use of a single venom component as the immunogen and the use of mixtures of monoclonal antibodies or affinity purified venom-specific antibodies as immunoreagents. The incorporation of avidin–biotin system and fluorogenic substrates significantly increased the sensitivity of these techniques allowing the detection of venom concentrations at picogram levels. Since snakebite is a medical emergency, immediate identification of the species is needed in order to administer the specific antivenom. In this respect, optimization of ELISA with unprocessed and undiluted blood and other biological samples reduced the assay time considerably. Optimization of the incubation steps at ambient temperature and visual discrimination of the optical density of test samples compared to controls are the two main approaches followed to achieve simplicity. Only a few commercial detection kits have been developed. Among them, a highly sensitive species-specific avidin–biotin microtiter ELISA (AB-microELISA) kit for the detection of venoms from four common Indian snakes has demonstrated efficacy to detect venom levels up to  $10 \text{ ng ml}^{-1}$  without ambiguity. The assay can be performed using  $600 \mu\text{l}$  of blood collected with an anticoagulant without further manipulation and can be completed in 30 min. Recently, an ion-sensitive field effect transistor based immunosensor has been developed for the detection and quantification of the potent neurotoxin  $\beta$ -bungarotoxin. For analysis, separation, or structural studies of these toxins at the basic research level, a variety of methods have been used including LC, LC–MS, CE, MS, and NMR.

### Scorpion Venoms

Scorpion envenoming remains a serious public health problem in certain regions, causing a considerable number of fatalities especially among children. Venoms contain many small proteins specifically interacting with ionic channels in excitable membranes. The main targets are voltage-sensitive sodium channels, although there is also a big family of peptides mostly active on potassium or chlorine

channels. The usual therapy includes the application of antivenom to neutralize circulating toxins. In order to administer the adequate amount of antivenom, it is important to do a rapid quantification of the circulating toxin concentration and to establish a correlation between the concentration of venom, the clinical symptoms, and the severity of the envenoming. Similarly to snake venoms, ELISA methods are very suitable for scorpion neurotoxin analysis. For example, a sandwich-type ELISA has been developed for the identification of circulating toxic antigens from *Tityus serrulatus* from Brazil, using the toxic fraction from the scorpion venom purified by Sephadex G-50 chromatography, and specific antibodies identified by immunoaffinity techniques to react to that fraction. The assay was specific for *T. serrulatus* venom, but because of its low sensitivity ( $4.8 \text{ ng ml}^{-1}$ ) it failed to distinguish between patients with mild envenoming and the nonenvenomed control group. A more sensitive sandwich-type ELISA test has been reported for the quantification of toxic fraction from North African scorpion venoms in human serum. The test proved to be reproducible, and its high sensitivity (detection limit of  $0.9 \text{ ng ml}^{-1}$ ) allowed for an accurate determination of scorpion venom concentration in human sera even at low levels.

### Spider Venoms

Most spider neurotoxins characterized to date have been found to be low molecular weight organic molecules (inorganic ions and salts, free acids and aminoacids, biogenic amines, neurotransmitters, and acylpolyamines), peptides, or proteins. Polyamines and a variety of polypeptides with a molecular mass ranging from 3000 to 8000 Da appear to represent the main toxic compounds in spiders. More than 60 peptide toxins have been described to date, whose presence can be confirmed by reversed-phase LC. Peptides can be easily distinguished from polyamines because of their different retention times and by ultraviolet detection using a diode-array detector.

The search for novel pharmacological tools in spider venoms requires precise and reproducible methods for the identification of species. Reversed-phase LC and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) analysis of crude venoms offers a powerful systematic tool that, in combination with morphological methods, has been used successfully to discriminate among closely related tarantula species. The method allows for the precise, sensitive, and rapid mass fingerprinting of large numbers of

samples in a reproducible manner. In most cases, the concomitant presence of peptides and polyamines has been detected, although with an overall predominance of peptides as toxic constituents. LC-MS may be also used as an additional venom characterization tool. When applied to the study of several tarantula venom samples, both MALDI-TOF-MS and electrospray ionization MS gave similar overall profiles and allowed fine discrimination of samples. In those studies, coupling of MS with LC allowed for clear discrimination between components of similar molecular weight and provided relevant information about venom composition, number of molecular species, and molecular weight distribution.

Spider venoms may also contain high molecular weight neurotoxic proteins. For example, in the genus *Latrodectus* (black widow) the high neurotoxicity of the venom is caused by a family of 110 kDa proteins called latrotoxins, which induce massive neurotransmitter release. Latrotoxins may be purified to homogeneity using chromatographic techniques and testing for the ability of fractions to deplete nerve terminals of their acetylcholine quantal store and of synaptic vesicles. Electrophoretic techniques may be also a convenient approach for the separation and characterization of venom components. Several proteins from the spider *Selenocosmia huwena* have been identified using two-dimensional gel electrophoresis, with the separation in the first dimension on a wide range of immobilized pH gradients. Proteins were then electrotransferred to a poly(vinylidene difluoride) membrane and analyzed by N-terminal microsequencing and MS peptide mapping.

Specific antigens in the sera of envenomed patients can be detected using ELISA analysis provided that purified specific antibodies are available. Thus, a highly specific sandwich-ELISA has been developed using affinity chromatography purified horse anti-*P. Nigriventer* immunoglobulins. Also, a sensitive ELISA has been developed for the specific detection of *Loxosceles* species venom.

**See also:** **Bioassays:** Overview. **Capillary Electrophoresis:** Overview. **Electrophoresis:** Overview; Principles; Two-Dimensional Gels. **Gas Chromatography:** Overview; Mass Spectrometry. **Immunoassays:** Overview. **Immunoassays, Applications:** Food. **Immunoassays, Techniques:** Radioimmunoassays; Enzyme Immunoassays. **Liquid Chromatography:** Reversed Phase; Liquid Chromatography-Mass Spectrometry; Food Applications. **Mass Spectrometry:** Overview; Principles; Matrix-Assisted Laser Desorption Ionization; Time-of-Flight. **Radiochemical Methods:** Radioreceptor Assays; Food and Environmental Applications. **Thin-Layer Chromatography:** Overview.

## Further Reading

- Botana LM (ed.) (2000) *Seafood and Freshwater Toxins. Pharmacology, Physiology and Detection*. New York: Dekker.
- Chávez-Olórtegui C, Bohórquez K, Alvarenga LM, *et al.* (2001) Sandwich-ELISA detection of venom antigens in envenoming by *Phoneutria nigriventer* spider. *Toxicon* 39: 909–911.
- Chorus I, Bartram J, Isaacson RL, and Jensen KF (1992) In: Isaacson RL and Jensen KF (eds.) *The Vulnerable Brain and Environmental Risks: Toxins in Food*. New York: Plenum Press.
- Escoubat P, Chamot-Rooke J, Stockling R, *et al.* (1999) A comparison of matrix-assisted laser desorption/ionization time-of-flight and liquid chromatography electrospray ionization mass spectrometry methods for the analysis of crude tarantula venoms in the *Pterinochilus* group. *Rapid Communications in Mass Spectrometry* 13: 1861–1868.
- Harada K-I, Kondo F, and Lawton L (1999) Laboratory analysis of cyanotoxins. In: Chorus I and Bartram J (eds.) *Toxic Cyanobacteria in Water. A Guide to Their Public Health Consequences, Monitoring and Management*, pp. 369–405. London: E & FN Spon.
- Itagaki Y, Fujita T, Naoki H, *et al.* (1997) Detection of new spider toxins from a *Nephilengys borbonica* venom gland using on-line mu-column HPLC continuous flow (FRIT) FAB LC/MS and MS/MS. *Natural Toxins* 5: 1–13.
- Lehane L (1999) *Ciguatera Fish Poisoning: a Review in a Risk-Assessment Framework*. Canberra, Australia: National Office of Animal and Plant Health, Agriculture, Fisheries and Forestry.
- Lehane L (2000) *Paralytic Shellfish Poisoning: A Review*. Canberra, Australia: National Office of Animal and Plant Health, Agriculture, Fisheries and Forestry.
- Liang S, Li X, Cao M, *et al.* (2000) Identification of venom proteins of spider *S. Huwena* on two-dimensional electrophoresis gel by N-terminal microsequencing and mass spectrometric peptide mapping. *Journal of Protein Chemistry* 19: 225–229.
- Reguera B, Blanco J, Fernández ML, and Wyatt T (eds.) (1998) *Harmful Algae*. Xunta de Galicia and Intergovernmental Oceanographic Commission of UNESCO.
- Selvanayagam ZE and Gopalakrishnakone P (1999) Test for detection of snake venoms, toxins and venom antibodies: Review on recent trends (1987–1997). *Toxicon* 37: 565–586.
- Shaw G and Smith M (2000) Algal analysis – Organisms and toxins. In: Nollet LML (ed.) *Food Science and Technology* 102 (*Handbook of Water Analysis*), pp. 143–167. New York: Dekker.
- Sivonen K (2000) Freshwater cyanobacterial neurotoxins: ecobiology, chemistry, and detection. In: Botana LM (ed.) *Food Science and Technology* 103 (*Seafood and Freshwater Toxins*), pp. 567–581. New York: Dekker.
- Weinberger R (2001) Capillary electrophoresis of venoms and toxins. *Electrophoresis* 22: 3639–3647.

remaining constant after the end-point of the *meta*-isomer neutralization. This titration would not be possible to follow using a visual indicator as the color change as the first end-point is too subtle.

Oxidizing agents such as permanganate and dichlorate absorb strongly in the visible region of the electromagnetic spectrum. This allows them to be used for photometric titrations. An example is the determination of iron(II) concentration by titrating with dichromate and measuring the absorbance at 335 nm.

Acid/base titrations can be monitored photometrically if a suitable indicator is added to the titration vessel. It is this indicator that causes the sharp change in the absorption profile at the end-point.

Photometric titrations have also been adapted for determining the end-point of precipitation titrations. The suspended solid produced by the titration reaction has the effect of reducing the energy of the radiation due to scattering, and so titrations can be carried out till a state of constant turbidity has been obtained.

**See Also: Indicators:** Acid–Base; Redox; Complexometric, Adsorption, and Luminescence Indicators. **Optical Spectroscopy:** Detection Devices. **Spectrophotometry:** Overview; Inorganic Compounds; Organic Compounds; Pharmaceutical Applications. **Titrimetry:** Overview; Potentiometric.

### Further Reading

Fifield FW and Kealey D (1995) *Principles and Practice of Analytical Chemistry*, 4th edn. Glasgow: Blackie Academic and Professional.

Headridge JB (1961) *Photometric Titrations*. Oxford: Pergamon Press.

Jeffery GH, Basset J, Mendham J, and Denney RC (1989) *Vogel's Textbook of Quantitative Chemical Analysis*, 5th edn. Harlow: Longman Scientific & Technical.

Skoog DA, Holler FJ, and Nieman TA (1998) *Principles of Instrumental Analysis*, 5th edn. Philadelphia: Harcourt Brace College Publishers.

Willard HH, Merrit LL, Dean JA, and Settle FA Jr. (1988) *Instrumental Methods of Analysis*, 7th edn. California: Wadsworth Publishing Company.

## TOILETRIES

See **COSMETICS AND TOILETRIES**

## TOTAL ANALYTICAL SYSTEMS, MICROFABRICATION

See **MICRO TOTAL ANALYTICAL SYSTEMS**

## TOXINS

Contents

**Mycotoxins**

**Neurotoxins**

### Mycotoxins

**S Dragacci**, Agence Française de Sécurité Sanitaire des Aliments (French Agency for Food Safety), Maisons-Alfort, France

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by M O Moss, pp. 1490–1499, © 1995, Elsevier Ltd.

### Introduction

Mycotoxins are fairly low relative-molecular-mass secondary metabolites of filamentous fungi that are toxic for humans and domesticated animals, causing economic losses by illness and death of animals when contaminated food or animal feed is consumed. There are many toxic metabolites produced by fungi, but only a fraction of these are currently considered to be important in foods. Similarly, a very broad

range of fungal species may produce toxic metabolites, but in human food the most important genera are *Aspergillus*, *Penicillium*, and *Fusarium*. Although the presence of a mycotoxin in a food usually implies previous infection by, and growth of, the mold producing the mycotoxin on the food, or a constituent of the food, there are two significant instances when that is not so. Thus aflatoxin M1 is secreted into milk following consumption by a cow of feed contaminated by aflatoxin B1. Ochratoxin A can pass unmetabolized into the kidney and muscle tissue of the pig after feeding on a contaminated feed such as barley.

Toxicogenic fungi (i.e., those able to produce mycotoxins) are hardly ever detectable in (cooked) food or feed. Indeed, in most cases, the fungi have disappeared when mycotoxins, which are very stable molecules due to their chemical structures, are found in food. If still present, mycological analysis based on the taxonomic characteristics of the fungi strains provide some indications but cannot be used to draw conclusions on the occurrence of toxins on food. In fact, mycotoxins are more resistant than the micromycetes that synthesized them, and a treatment as simple as heating can greatly diminish the number of revivable germs without modifying the toxicity.

More than 350 mycotoxins have been identified either from food or from laboratory culture media. The main mycotoxins (Table 1) of concern for the safety of consumers are aflatoxins (aflatoxins B and G, aflatoxin M1), ochratoxin A, patulin, fumonisins, zearalenone, and deoxynivalenol (DON) and other

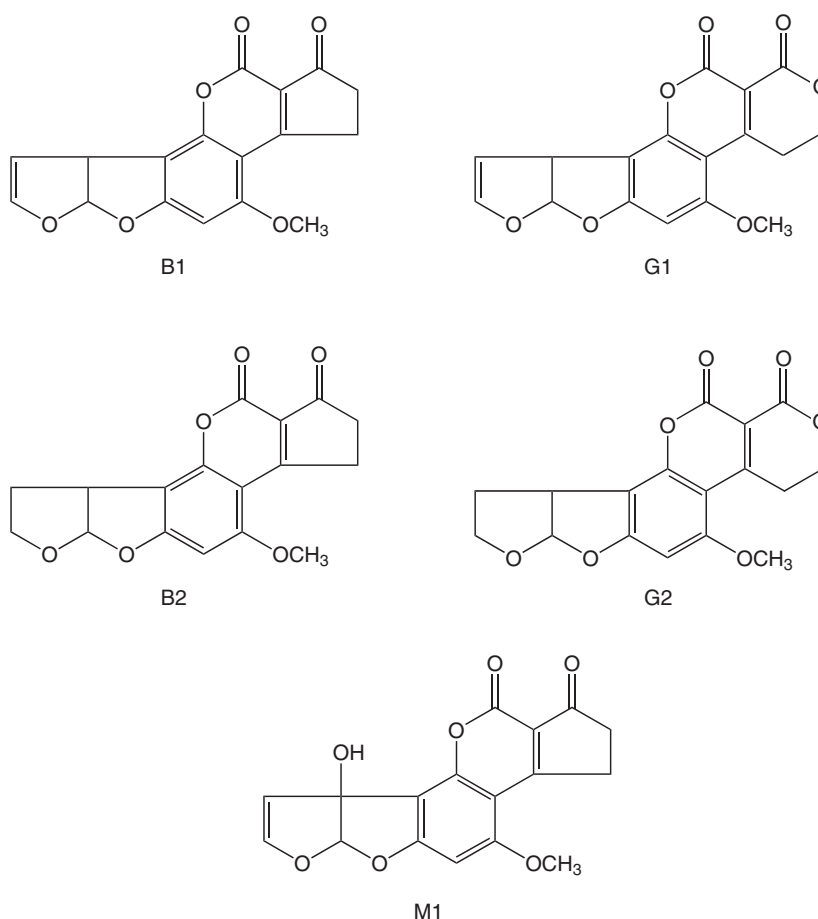
*Fusarium* toxins, also called trichothecenes. Other mycotoxins are also encountered but to a lesser extent, and consecutively they are of less concern for public authorities: cyclotriazonic acid, sterigmatocystin, penitrem A, and tenuazonic acid. The structures of these mycotoxins are shown in Figures 1–10.

## Toxicity of Mycotoxins and Regulations Regarding Mycotoxins in Food

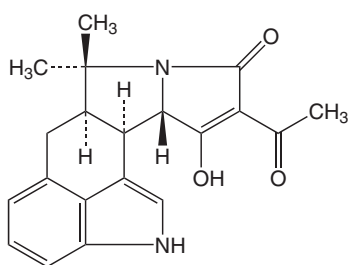
Mycotoxins are highly toxic for human beings and for animals. For instance, aflatoxins, especially aflatoxin B1, are highly hepatotoxic and indicated as potent carcinogens. Ochratoxin A is responsible for nephrotoxicity, and it is suspected of playing a role in the well known Balkan endemic nephropathy, where the incidence of tumors of the bladder or urinary tract is significantly higher than in the healthy population. Fumonisin has been identified as the causative agents for several animal syndromes like equine leucoencephalomalacia, a fatal syndrome in horses consisting of liquefactive necrosis in one or both cerebral hemispheres leading to extensive and fatal damage to tissue brain. Besides, fumonisin B1 may be involved in esophageal cancer as its involvement was suspected in some regions of China and South Africa. DON is also called vomitoxin as it causes food refusal in animals; there is also some evidence of its immunotoxicity effects. Zearalenone has estrogenic effects and is known to produce hyperestrogenism in pigs.

**Table 1** Mycotoxins found in food

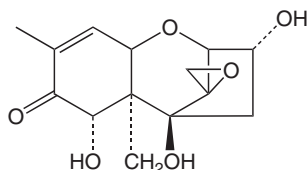
Mycotoxin	Figure	Food	Common fungi species producing the mycotoxin
Aflatoxins B1, B2, G1, G2	1	Groundnuts, peanuts, tree nuts, shell-fruits (pistachios, nuts), maize, figs, dried raisins, spices	<i>Aspergillus flavus</i>
Aflatoxin M1	1	Milk and milk products (cheese)	<i>A. parasiticus</i> Metabolite of aflatoxin B1 transferred into milk
Cyclotriazonic acid	2	Cheese (crust), maize, groundnuts, millet	<i>Aspergillus flavus</i> <i>Penicillium aurantiogriseum</i>
DON	3	Cereals (wheat)	<i>Fusarium graminearum</i> <i>F. culmorum</i>
Fumonisin	4	Maize	<i>Fusarium moniliforme</i>
Ochratoxin A	5	Cereals (barley, oats), maize, coffee and cacao beans, raisins and wine, beer, pork-derived products (offal)	<i>Penicillium verrucosum</i>
Patulin	6	Apple, apple juice, and puree	<i>Aspergillus ochraceus</i> <i>A. carbonarius</i> <i>Penicillium expansum</i>
Penitrem A	7	Walnuts	<i>Penicillium aurantiogriseum</i>
Sterigmatocystin	8	Cereals, coffee beans, cheese	<i>Aspergillus versicolor</i>
Tenuazonic acid	9	Tomato paste	<i>Alternaria tenuis</i>
Zearalenone	10	Maize, barley, wheat	<i>Fusarium graminearum</i>



**Figure 1** Aflatoxins B1, B2, G1, G2, and M1.



**Figure 2** Cyclopiazonic acid.



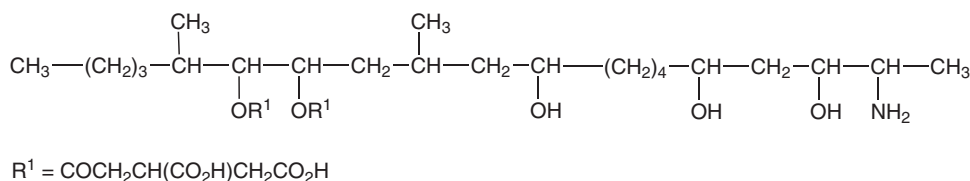
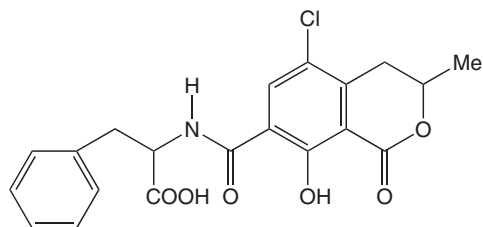
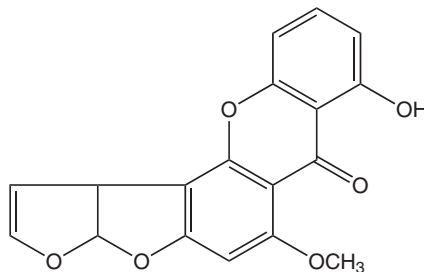
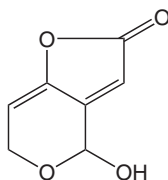
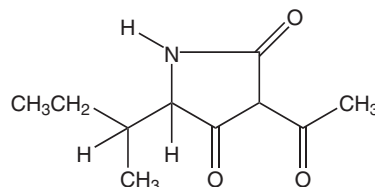
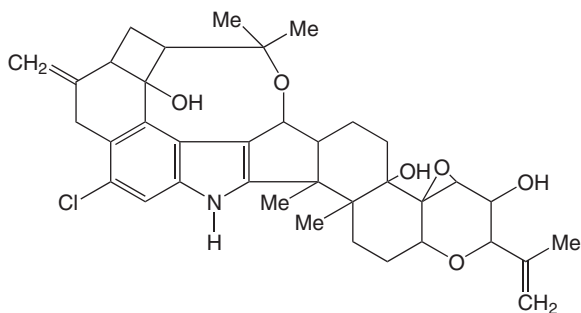
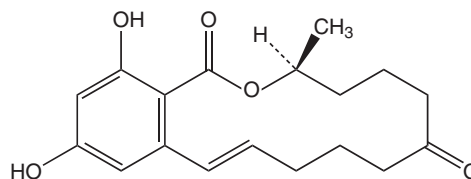
**Figure 3** Deoxynivalenol.

Recognition of the importance of those mycotoxins in public health is reflected in the issuing of statutory and advisory regulations for the maximum tolerable levels in food and animal feed in many countries. The

presence of some mycotoxin molecules has been regulated by public authorities to guarantee food safety for populations (Table 2). A lot of countries have individually enforced regulations for mycotoxins in food and feed, especially for aflatoxins. Many countries have especially stringent levels for aflatoxin M1 in milk and dairy products, ranging from 0.01  $\mu\text{g}$  per kg in pasteurized fresh milk for infants in Austria to 0.5  $\mu\text{g}$  per kg for milk and milk products in the USA. This clearly reflects the major concern regarding exposure of very young and elderly people, who would be more likely to be both more sensitive and exposed to higher levels of toxins.

In the European Union and in the America block, uniform regulations have been put in place. The Codex Alimentarius is also attempting a worldwide uniformity in recommending the maximal tolerable limits for the main mycotoxins of concern (aflatoxins, ochratoxin A, fumonisins, patulin, DON, zearalenone); the Codex Alimentarius is also supervising the writing of codes of practice for farmers to help prevent the production of mycotoxins in crops as the



**Figure 4** Fumonisin B1.**Figure 5** Ochratoxin A.**Figure 8** Sterigmatocystin.**Figure 6** Patulin.**Figure 9** Tenuazonic acid.**Figure 7** Penitrem A.**Figure 10** Zearalenone.

US Department of Agriculture (USDA)/Food and Drug Agency (FDA) have already done.

Many countries have implemented official controls and national monitoring programs to survey food and feed contamination. Also, international trade requires the control of consignments susceptible to contamination by mycotoxins. Due to the carcinogenic potency of most mycotoxins, they should be tracked at trace levels in food, 'trace' referring mainly to the parts per billion level (i.e., micrograms per kilogram). Besides, the diversity of the chemical structures of mycotoxins and the wide range of foods and animal feeds are the main hindrance for developing one single method for many mycotoxins from

any consignment, but it has been achieved for one matrix, namely beer. The establishment and use of agreed analytical methods development are needed to ensure the reliability of analytical data, especially in case of dispute. Nowadays, methods are required by officials to be validated by collaborative studies (i.e., by interlaboratory study) or at least to be in-house validated. For that purpose, international standards like ISO 5725, 1998 or the uniform International Union of Pure and Applied Chemistry–Association of the Official Analytical Chemists (AOAC) protocol for validation of methods by a single laboratory are available that are totally suitable for validation of methods for the determination of mycotoxins in food.



**Table 2** Maximum tolerable limits for mycotoxins: some examples (for more information on worldwide regulations, see [www.fao.org](http://www.fao.org))

<i>Mycotoxins</i>	<i>Maximum tolerable limits (in food as consumed) (<math>\mu\text{g per kg}</math>)</i>	<i>Sources</i>
Aflatoxins B1 (and sum of aflatoxins B and G) in peanuts, nuts, dried fruits, and cereals	2 (4)	European Commission
Aflatoxins B1 (and sum of aflatoxins B and G) in spices	5 (10)	European Commission
Aflatoxin M1 in milk	0.05	European Commission
Ochratoxin A in cereals	0.5	US-FDA and Codex Alimentarius
Ochratoxin A in dried raisins	5	European Commission and Codex Alimentarius
Patulin in apple juice	10	European Commission
Patulin in apple puree	50	European Commission and Codex Alimentarius
DON	25	European Commission
	2	US-FDA and Canada

## Sampling

Before an analysis can be attempted, a sample must be obtained. The analyst is frequently presented with a single sample, adequate in quantity for the analytical methodology but without any guarantee that it is representative of the consignment of food or animal feed from which it came. Many commodities to be analyzed are particulate in structure, and molds do not grow uniformly through such materials. Thus, mycotoxins are not usually distributed evenly throughout a consignment, although exceptions may be mycotoxins in fluids like aflatoxin in milk or patulin in fruit juice, and it is important to agree on a sampling plan that adequately protects the consumer without unnecessarily penalizing the producer. There is not even a normal distribution of mycotoxins within a batch of a commodity. For example, in the contamination of groundnuts, maize, cotton seed, or palm kernels with aflatoxins, it is likely that a relatively small number of kernels or seeds are contaminated with very high levels and many are not contaminated at all, giving a very skewed distribution.

Satisfactory sampling plans, such as those devised for groundnuts by the USDA, the Natural Resources Institute of the UK, or the European Commission should define the size of the sample and how it is to be collected, a method for dividing the sample, and an analytical procedure. The USDA sampling plan for the determination of aflatoxins in groundnuts requires 66 kg of material be collected from at least 100 sacks for each 20 tonne lot. The 66 kg is divided into three subsamples of 22 kg, and the first one of these is grounded, thoroughly mixed, and analyzed. At a time when the maximum tolerated level was 25  $\mu\text{g per kg}$ , if the result of this first analysis was less than 16  $\mu\text{g per kg}$  the lot was accepted and if it was greater than 75  $\mu\text{g per kg}$  it was rejected. If, however,

the first analysis gave a result between these two figures, then a second subsample had to be analyzed in the same way. If the mean of the two was less than 22  $\mu\text{g per kg}$  the lot was accepted and if greater than 38  $\mu\text{g per kg}$  rejected, but again if the result fell between these two figures then the third subsample had to be analyzed. In this example, the mean of all three now needs to be less than or equal to 25  $\mu\text{g per kg}$  for the lot to be accepted.

For many years, sampling was not taken sufficiently seriously, but detailed research on the distribution of mycotoxins in commodities and an appropriate statistical analysis of the results have led to an increased use of acceptable sampling plans. Recognition of the extremely skewed distribution of mycotoxins has also led to the design of methods for recognizing and eliminating the heavily contaminated kernels from consignments, thus ensuring that they do not contribute to the overall contamination of the batch after further treatment such as mixing, grinding, and oil extraction.

## Methodology for Analyzing Mycotoxins in Food

The main difficulties in the analysis of mycotoxins are due to the fact that they should be detected at very low levels (micrograms per kilogram, sometimes at nanogram per kilogram levels) in very complex food matrices. Most mycotoxins (except DON or zearalenone) are not sufficiently UV absorbing to be detected as traces using an UV detector. Hopefully, most of them are fluorescent by themselves or after having a fluorophore coupled to them. This is the case for aflatoxins whose fluorescence is strongly enhanced by iodination or bromination of the mobile phase used in liquid chromatographic systems, where

under an elevated temperature, iodine or bromine is generated and coupled to the aflatoxin, resulting in high emission of fluorescence. Less corrosive reagents such as cyclodextrins have also been demonstrated to be efficient in enhancing the fluorescence of aflatoxins.

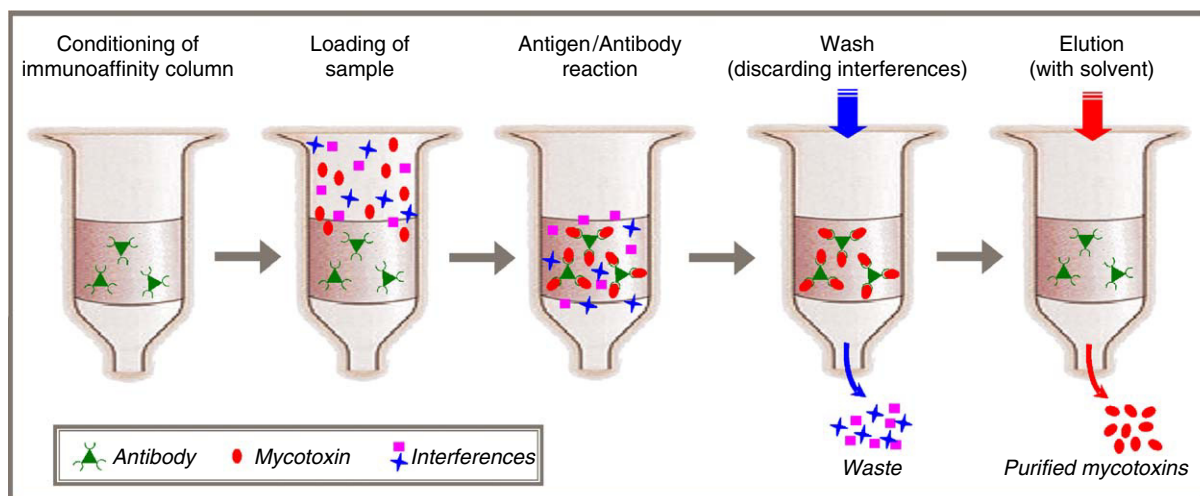
Methods of analysis for mycotoxins mainly include four steps: (1) grinding the sample of food, (2) extracting the mycotoxin using organic solvents or aqueous solvents, sometimes using water, and concentrating the extract, (3) purifying the extract using silica-like cartridges or immunoaffinity columns, and (4) analysis using chromatographic techniques or immunochemical techniques like enzyme-linked immunosorbent assay (ELISA); often, in the case of ELISA techniques, the preparation of samples is easily simplified to the two first steps.

### Extraction Procedures and Purification of Mycotoxins before Analysis

Cleanup procedures vary considerably, depending on both the mycotoxin(s) and the commodity to be analyzed. Although mycotoxins are readily soluble in water-organic solvents, it is not always easy to remove them from the macromolecular food matrix by direct extraction with simple solvents such as chloroform, ethyl acetate, aqueous solvents, or water (as in the case for DON). However chloroform, although one the most efficient solvents for extracting most mycotoxins, is being recommended less frequently because it is an ozone-depleting agent: it has mainly been replaced by acetonitrile, methanol, or a combination of these aqueous solvents. For some

fatty matrices, it may be necessary to remove oils and fats through a preliminary extraction with hexane before extracting the mycotoxin. Some mycotoxins (aflatoxins, ochratoxin A, patulin, etc.) can be adsorbed onto C8 or C18-silica cartridges and/or Florisil<sup>TM</sup> (commercially available) and subsequently eluted. Fumonisin are better extracted using strong anion exchange columns.

Nowadays, the application of immunoaffinity chromatography to mycotoxins has considerably improved the extraction and purification of mycotoxins from a large range of food matrices. Consecutively the recovery and the limit of detection of the methods have been strongly improved. The principle of immunoaffinity chromatography is as follows (**Figure 11**): monoclonal antibodies specifically produced against a mycotoxin are immobilized onto an inert support and packed in a microcolumn (support of 1 cm height or so). The food extract, rediluted in water or phosphate buffered saline-like buffer to be neutral at around pH 7, is passed through the column. When most of the constituents of the extract are washed out of the column, the mycotoxin molecules are specifically retained to the antibodies. Because the mycotoxins are not covalently bound to antibodies, they can be displaced by an organic solvent like methanol or acetonitrile and eluted as a highly pure concentrated extract that could then be analyzed using chromatographic methods. Now, numerous brands of immunoaffinity columns are commercially available for the main mycotoxins analyzed. This method of food extract purification has become increasingly popular despite the price of immunoaffinity columns remaining high. This procedure is now included in most recently validated methods earlier than liquid chromatographic methods.



**Figure 11** Principle of immunoaffinity chromatography.

## Chromatographic Methods

Chromatographic methods used with mycotoxins include thin layer chromatography (TLC), liquid chromatography (LC), gas chromatography (GC), and LC coupled to mass spectrometry (LC-MS).

### TLC Methods

TLC methods were very popular several decades ago when they were used in the discovery of mycotoxins or for the analysis of less known mycotoxins, especially when no standards were available. TLC methods were gradually replaced by LC except in laboratories of developing countries not equipped with LC. TLC is also much cheaper than LC due to the price of the LC apparatus and the cost of the solvents. These methods are also used in analyzing multicontaminated samples (i.e., contaminated by several mycotoxins), which is not possible using LC in that case several kinds of mobile phases are required.

Some mycotoxins (especially aflatoxins) are able to fluoresce under long-wave UV (360 nm) when absorbed on a solid substrate such as the silica gel of a TLC plate. For instance, aflatoxins were divided into two groups, named B and G, after the bluish and greenish fluorescence they produce. The main problem experienced with years of analysis of aflatoxins is that many natural products fluoresce blue or green under UV and many of these compounds have an  $R_F$  value close to, or even identical with, that of one of the four common aflatoxins. These false-positive results need to be confirmed, which could be done by applying to the extract a two-dimensional TLC coupled with the derivatization specific to the mycotoxins. One of these protocols for confirming the presence of aflatoxins uses trifluoroacetic acid to derivatize as hemiacetals the aflatoxins B1 and G1.

The AOAC standardized a lot of procedures based on TLC that were validated by collaborative studies and met the AOAC requirements regarding performance criteria. The AOAC standards provide detailed information on the apparatus, the reagents, the preparation of standards, and the sampling. A selection of these methods and the most recent ones are quoted in the standard procedures section.

### LC Methods

LC methods are now the procedures of choice in the mycotoxin field. Many LC methods have also been collaboratively validated and published as standards (AOAC official methods or ISO standards or EN standards). When standard solutions of mycotoxins are commercially available, any good LC laboratory could implement those kinds of methods; moreover, due to the use of a fluorescence detector, very small

traces can be detected with very good reproducibility. Another major advantage of LC is the possibility of automation and the use of autosamplers, allowing unattended operation of LC equipment. Online cleanup of crude extracts and online derivatization before or after separation are additional advantages. A range of detectors can be used, although the most widespread are based on UV absorption (patulin, zearalenone) or detection of fluorescence (for most mycotoxins), the latter being potentially 30–40 times more sensitive, as has been demonstrated for aflatoxins.

Well-established and interlaboratory validated LC methods are published in the scientific specialized literature or as international standards. Those approved by the AOAC are mentioned in the standard procedures section (see also [www.mycotoxins.org](http://www.mycotoxins.org) for a complete list of references). Their main features are described in Table 3. Most of these protocols include the use of immunoaffinity columns before LC analysis.

### GC Methods

For GC the mycotoxins need to be sufficiently thermostable and volatile or be converted into volatile derivatives. The two most widely used detectors in mycotoxin analysis – flame ionization detectors (FIDs) and electron capture detectors (ECDs) – do not require that a compound show any fluorescence or UV absorption, and the group of mycotoxins for which GC was initially the most widely used was that of the trichothecenes. More recently another group of *Fusarium* mycotoxins, the fumonisins, which also lack the appropriate optical properties, have been analyzed successfully using GC.

Many mycotoxins, including the trichothecenes and fumonisins, have polar hydroxyl groups that can be readily derivatized to trimethylsilyl ethers with such reagents as trimethylsilyl chloride or hexamethyldisilazane and *N*-trimethylsilylimidazole. Alternatively polyhalogenated acyl derivatives can be made, using reagents such as trichloroacetic anhydride or heptafluorobutyric anhydride, which are especially useful in conjunction with ECDs.

There is a large number of different trichothecenes, and ideally their analysis requires confirmation using MS coupled to GC (GC-MS) and analysis of the macrocyclic trichothecenes, which include verrucarins, roridins, and satratoxins, requires considerable experience and a specialized laboratory. However, well-established methods are available for individual trichothecenes of particular importance, such as the T-2 toxin, diacetoxyscirpenol, DON, and neosolaniol. The large number of *Fusarium* trichothecenes are acylated derivatives of a much smaller number of parent alcohols, such as T-2 tetraol,

**Table 3** Some examples of the analytical protocols of collaboratively validated LC methods for mycotoxins

<i>Mycotoxins and matrices</i>	<i>Extraction solution</i>	<i>LC method and detection</i>	<i>Limits of quantification</i>	<i>Recovery (%)</i>
Aflatoxins B and G in peanut butter, pistachios, fig paste, and paprika	Methanol 80% and defatting with pure hexane	Immunoaffinity cleanup prior to LC coupled to fluorescence detector and postcolumn derivatization by using iodination or bromination	0.2 µg per kg of aflatoxin B1  0.4 µg per kg for sum of aflatoxins B and G	85–100
Aflatoxin M1 in milk	No previous extraction	Immunoaffinity prior to LC coupled to fluorescence detector	0.005 µg l <sup>-1</sup>	70–95
Ochratoxin A in barley	Acetonitrile 60%	Immunoaffinity prior to LC coupled to fluorescence detector	0.6 µg kg <sup>-1</sup>	65–110
Ochratoxin A in roasted coffee	Methanol/sodium bicarbonate followed by chromatography on phenylsilane cartridge	Immunoaffinity prior to LC coupled to fluorescence detector	0.3 µg kg <sup>-1</sup>	65–97
Patulin in apple juice and puree	Ethylacetate	LC coupled to UV detector	15 µg kg <sup>-1</sup>	70–100
	Pectinase treatment for puree	LC coupled to UV detector		
Fumonisin in corn and cornflakes	Acetonitrile/methanol/water	Immunoaffinity prior to LC coupled to fluorescence detector	0.4 µg kg <sup>-1</sup>	75–110
DON in wheat	Water	Immunoaffinity prior to LC coupled to UV detector	100 µg kg <sup>-1</sup>	> 80

scirpentriol, DON, and nivalenol, and so another approach to their analysis is to hydrolyze the possibly complex mixture of trichothecenes to the parent alcohols and analyze these as their trimethylsilyl ethers. However, when this is done, there is often a poor correlation between the total trichothecene content and the observed toxicity, reflecting the large differences in the toxicity of different acyl derivatives even of the same parent alcohol.

As for patulin, it is necessary to derivatize it with trimethylsilyl ether or heptafluorobutyrate to make it volatile. Determination of patulin through GC–MS detection and quantification of the heptafluorobutyrate derivative after treatment with heptafluorobutyrylimidazole has also been developed. Trimethylsilyl can also serve as a derivatization agent for the detection of patulin through GC–ECD or GC–MS. However, it should be noted that LC methods are generally preferred over GC methods because of their easiness of application.

### LC–MS Methods

LC–MS is one of the most advanced techniques available for the detection of mycotoxins. They are

especially used for confirmatory purposes; they have very recently become the method of choice for trichothecenes, gradually replacing GC–MS methods, which appear less reliable. Extraction and cleanup techniques have to be applied prior to separation and detection in order to enable well-separated peaks without interferences from matrix components. In brief, the LC effluent enters an ionization chamber via a nebulizer. Ionization is performed through several techniques (electrospray, thermospray, chemical, fast atom bombardment, etc.). In a collision chamber fragmentation takes place and fragments then enter the high-vacuum region of the MS, where detection takes place. Several setups now exist for optimized quantitation and/or identification. Ion trap instruments are generally better suited for identification than are triple quadrupole instruments (higher MS power). Triple quadrupole instruments provide better information for quantification (faster scanning, higher sensitivity). Hybrid instruments exist that provide a linear ion trap in a triple ‘quad’ instrument in an attempt to make the best of both setups.

Several methods for simultaneous determination of mycotoxins have been reported, offering a significant



advantage over conventional techniques. For example, 13 *Penicillium* toxins (ochratoxin A, citrinin, patulin, mycophenolic acid, cyclopiazonic acid, PR-toxin, rubratoxin B, verruculogen, chaetoglobosin B, penitrem A, griseofulvin, roquefortin C, penicillic acid) have been determined in standardized matrix samples, including bread, rice, potatoes, vegetables, fruit, and other food. The LC-MS system used an ion trap instrument with electrospray ionization (ESI) and an atmospheric pressure chemical ionization (APCI) method. An LC-MS/MS method for ochratoxin A (OTA), zearalenone and two of its metabolites, DON and nivalenol, used a triple quad instrument with an ESI source from a pork liver matrix. The limits of quantification were as low as 1 µg per kg.

The use of a variety of ionization methods (fast atom bombardment (FAB) and APCI among them, for trichothecenes has been reported. FAB results in numerous different ions, whereas APCI, which has emerged as the method of choice for the major mycotoxins, provides a suitable spectrum for identification and quantification. Thermospray ionization (TSP) is being used less and less. Derivatization is not necessary, unless needed for additional fragments. The detection limits are in the lower nanogram per gram range. LC-TSP MS has also been used as a multi-mycotoxin method for the determination of several *Fusarium* mycotoxins together with zearalenone and OTA in the 1–40 ng per g range (see [www.mycotoxins.org](http://www.mycotoxins.org)).

### Preparation of Standard Solutions of Mycotoxins

Pure solutions of mycotoxins should be checked for concentration using spectrophotometry. Keep in mind that (1) pure solutions of mycotoxins (especially aflatoxins) should be handled with caution due to their carcinogenicity and (2) whereas stock solutions of mycotoxins ( $\sim 1 \text{ mg ml}^{-1}$ ) are rather stable for up to 1 year in the freezer, diluted solutions are unstable and most degrade with exposure to light. The following example is suitable for preparing stock solutions for aflatoxin B1.

Weigh  $\sim 1 \text{ mg}$  of aflatoxin and dissolve it in 100 ml of acetonitrile. Some suppliers provide mycotoxins as preweighed dry films in a vial; in that case, dissolve the film in the required volume of acetonitrile to get a solution of  $10 \text{ µg ml}^{-1}$ . With a calibrated spectrophotometer, measure the absorbance,  $A$ , at the wavelength of maximal absorption, which depends on the solvent used (355 nm in acetonitrile). Calculate the concentration of the stock solution by using the formula

$$\begin{aligned} &\text{Micrograms of aflatoxin B1 per milliliter} \\ &= A \times \text{molecular weight} \times 100/\epsilon \end{aligned}$$

where the molecular weight of aflatoxin B1 is 312 and the molar absorptivity,  $\epsilon$ , is  $2060 \text{ m}^2 \text{ mol}^{-1}$ .

The coefficient  $\epsilon$  also depends on the solvent used to dissolve mycotoxins (for instance,  $\epsilon = 2180 \text{ m}^2 \text{ mol}^{-1}$  in methanol). It is best to store the stock solution in the freezer as aliquots of 50 or 100 µl and to defreeze one aliquot to prepare working diluted solutions.

### Immunochemical Methods

Among the immunochemical methods, radioimmunoassay (RIA) and ELISA have been used for the determination of mycotoxins in foods and feeds. RIA tests have been developed for the determination of aflatoxins, T-2 toxin, ochratoxin, and zearalenone; however, taking into account the disadvantages of using radioactive substances such as the limited shelf-lives of reagents, the legislative restrictions for exchange or marketing of isotopes, and stringent specifications for facilities, RIA tests have become less and less popular, especially for routine analysis. In contrast, ELISA tests have been widely developed, marketed, and used for routine assays: they are rapid methods, and sometimes when very simplified (by using sticks for instance), they can be used as field methods requiring no laboratory apparatus, so that they could be used in industry or an industrial laboratory.

ELISA tests have been developed as direct and indirect techniques:

The direct technique, where the specific antibody against the mycotoxin is coated on the bottom of the 96 wells of a microtitration plate. The sample solution (extract) and a constant, known level of the toxin-enzyme conjugate are incubated simultaneously. After washing, the toxin-enzyme conjugate, which links to the wells through the antibody, is detected by adding a substrate specific to the enzyme. The resulting color is measured through photometry or visually detected by modification of the color of the medium in the wells.

The indirect technique, where the mycotoxin is fixed onto the 96 wells of the microplate through a toxin-polypeptide conjugate intermediary. The sample solution and a known constant concentration of the antibody are incubated simultaneously. After washing, the quantity of antibody bound to the plate is determined by adding an enzymatic conjugate 'enzyme-immunoglobulin' (specific antibody). The resulting color after adding the specific chromogenic substrate is measured using photometry or judged visually.

A variant of this type of ELISA can also be used: a 'specific antibody–enzyme' complex can also be synthesized and used for competition during the incubation. Thus, the use of a second antibody is no longer necessary, and the specific enzyme substrate can be added. In all these types, a standard curve is plotted using increasing quantities of standard mycotoxins.

The principle of the direct technique is often used in the development of screening tests: the mycotoxin content is performed by visual evaluation with a yes/no response at a specific cutting level. A lack of color indicates contamination of the sample by a mycotoxin above a specific cutting level. Such kits have been developed for instance on Nylon<sup>TM</sup> membrane under specific formats like

Cards such as 'EZ-screen<sup>TM</sup>' from R-Biopharm, which have been developed for total aflatoxins (cutting levels 20 and 5 µg per kg), for ochratoxin A, and for zearalenone (cutting levels of 20 and 100 µg per kg, respectively).

Cups such as 'Afla-20 Cup<sup>TM</sup>', from International Diagnostic System, working for total aflatoxins with cutting levels of 20 µg per kg.

Other formats for other mycotoxins at different cutting levels have also been developed by various research and development companies.

Determination test kits are based on direct but more often on indirect-type ELISA techniques. They use 96-well microtitration plates or beads or sticks. Analysis for mycotoxins can be completed within 3–4 h. This kind of test needs a photometric reader for the quantitation by plotting sample value against standard-curve derived values. Nowadays, a number of companies are developing such tests for aflatoxins, OTA, patulin, zearalenone, T-2 toxin, fumonisins, DON, etc. with limits of quantification as low as 1 µg per kg or even less.

The main advantages of the ELISA techniques are mainly the rapidity of analysis, the high sensitivity of detection (ultrasensitive tests), and the cost per unit sample, but the main drawback is the production of false positives (results being positive when no mycotoxins are present in the extract) and sometimes false negatives (results being negative when mycotoxins occur in the extract). Especially for food or feed complex matrices, such a limitation is very important and explains why ELISA procedures are seldom selected as official methods.

### Standard Procedures

The AOAC prescribes ~50 methods of analysis for mycotoxins. Here is a selection of TLC and LC methods.

AOAC 973.38, sterigmatocystin in barley and wheat – TLC method

AOAC 975.36, aflatoxins in foods and feeds – Romer minicolumn method

AOAC 975.38, ochratoxin A in green coffee – TLC method

AOAC 976.22, zearalenone in corn – TLC method

AOAC 980.21, aflatoxin M1 in milk and cheese – TLC method

AOAC 982.24 aflatoxins B1 and M1 in liver – TLC method

AOAC 985.18, zearalenol and zearalenone in maize – LC method

AOAC 986.17, DON in wheat – TLC method

AOAC 995.15, fumonisins B1, B2, and B3 in corn – LC method

AOAC 999.07, aflatoxin B1 and total aflatoxins in peanut butter, pistachio paste, fig paste, and paprika powder – immunoaffinity–LC method with postcolumn derivatization

AOAC 2000.02, determination of patulin in clear and cloudy apple juices and apple puree using HPLC

AOAC 2000.03, determination of ochratoxin in barley using HPLC and immunoaffinity cleanup

AOAC 2000.08, aflatoxin M1 in liquid milk using immunoaffinity column cleanup and HPLC

AOAC 2000.09, determination of ochratoxin in a roasted coffee using HPLC

AOAC 2000.16, aflatoxin B1 in baby food using an immunoaffinity column and HPLC

AOAC 2001.01, determination of ochratoxin A in wine using immunoaffinity column cleanup and HPLC

AOAC 2001.04, determination of fumonisins B1 and B2 in corn and cornflakes using HPLC and immunoaffinity cleanup

AOAC 2003.02, determination of aflatoxins in animal feed using immunoaffinity column cleanup and HPLC with postcolumn derivatization

## Safety Procedures

Mycotoxins should be handled as toxic substances, and special care should be taken when handling pure toxins in a dry form or samples (including TLC plates) that may be a source of dust. Accidental spillages of material containing toxin can generally be treated with 1% sodium hypochlorite bleach, left for 10 min, swabbed down with 5% aqueous acetone initially and subsequently with plenty of water.

Glassware exposed to aflatoxins can be rinsed with methanol and soaked in 1% sodium hypochlorite solution to which, after 2 h, acetone is added to 5%



of the total volume. The acetone is left for 30 min to destroy any potentially harmful but reactive intermediates formed by the hypochlorite and then washed thoroughly. It is known, for example, that aflatoxin B1 can form a carcinogenic derivative if treated with hypochlorite alone.

A very detailed discussion of laboratory decontamination and destruction of aflatoxins in laboratory wastes has been prepared for the International Agency for Research on Cancer.

**See also:** **Food and Nutritional Analysis:** Mycotoxins. **Immunoassays, Techniques:** Enzyme Immunoassays. **Liquid Chromatography:** Food Applications. **Mass Spectrometry:** Food Applications. **Thin-Layer Chromatography:** Overview; Principles.

## Further Reading

- Betina V (1993) In: Betina V (ed.) *Chromatography of Mycotoxins: Techniques and Application*, pp. 141–233. Amsterdam: Elsevier.
- Chu FS (1995) Mycotoxin analysis. In: *Analyzing Food for Nutrition Labeling and Hazardous Contaminants*, pp. 283–331. New York: Dekker.
- Codex Alimentarius: <http://www.codexalimentarius.net>.
- Dorner JW (1996) Mycotoxins in food: Methods of analysis. In: Nollet LML (ed.) *Handbook of Food Analysis, Residues and other Food Components Analysis*, ch. 26, pp. 1089–1146.
- FAO (Food and Agriculture Organization) (1997) *Worldwide Regulations for Mycotoxins*, Food and

Nutrition Paper, 1995 – a compendium no. 64. Rome: FAO.

- IARC (International Agency for Research on Cancer) Scientific Publications (1980) Laboratory decontamination and destruction of aflatoxins B1, B2, G1, G2 in laboratory wastes. In: Castegnaro M, Hunt DC, Sansone EB, *et al.* (eds.). Lyon: IARC Publications.
- IARC (International Agency for Research on Cancer) Scientific Publications (1991) Laboratory decontamination and destruction of carcinogens in laboratory wastes: Some mycotoxins. In: Castegnaro M, Barek J, Freymy JM, *et al.* (eds.). Lyon: IARC Scientific Publications.
- IARC (International Agency for Research on Cancer) Scientific Publications (1993) *Monograph on the Evaluation of Carcinogenic Risk to Humans. Some Naturally Occurring Substances: Food Items and Constituent Heterocyclic Aromatic Amines and Mycotoxins*, vol. 56, IARC edition. Lyon: IARC.
- ISO (1998) ISO 5725-1 to 5: *Accuracy (Trueness and Precision) of Measurement Methods and Results*.
- Miller JD and Trenholm HL (1994) *Mycotoxins in Grain. Compounds Other than Aflatoxin*. St Paul, MN: Eagan Press.
- Pfohl-Leszkowicz A, Petkova-Bocharova T, Chernozemsky IN, and Castegnaro M (2002) Balkan endemic nephropathy and the associated urinary tract tumours: Review on etiological causes, potential role of mycotoxins. *Food Additives and Contaminants* 19(3): 282–302.
- Samson RA, Hoekstra ES, Frisvad JC, and Filtenborg O (2000) *Introduction to Food-Borne Fungi*. Baarn, The Netherlands: C.B.S.
- Van Egmond HP (1989) *Mycotoxins in Dairy Products*. London: Elsevier.

## Neurotoxins

**M-T Fernández-Sánchez, A Ferrero, A Pérez-Gómez, and A Novelli**, University of Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The term neurotoxin applies to a variety of substances that can affect the nervous system producing a broad variety of effects usually harmful to human health, although in some cases these substances can be of therapeutic use. Neurotoxic compounds are occasionally present in food and beverages, and consumption of contaminated products may therefore contribute to the etiology of certain psychiatric disorders or neurodegenerative diseases in humans. Three different main groups of

food-related neurotoxins may be considered: (1) metal and organometallic compounds (e.g., mercury and methylmercury, lead, cadmium, manganese, aluminum, and alkyltin); (2) marine and freshwater toxins of algal origin; and (3) toxicants purposefully added or resulting from industrial activities and accumulated in food including organophosphates and polychlorinated biphenyls (PCBs). In addition to food-related toxic compounds, poisons from venomous species may also contain potent neurotoxins. The analysis of metals and organometallic compounds as well as of organophosphates and PCBs are covered by other articles within the Encyclopedia. The following is a discussion of a variety of methods for the analysis and detection of different types of marine and freshwater neurotoxins of algal origin, and for some neurotoxic compounds from snake, scorpion, and spider venoms.

# V

## VEGETABLES

See **FOOD AND NUTRITIONAL ANALYSIS: Vegetables and Legumes**

## VITAMINS

Contents

**Overview**

**Fat-Soluble**

**Water-Soluble**

### Overview

**V Spitzer and U Höller**, DSM Nutritional Products (registered as Roche Vitamins Ltd), Basel, Switzerland

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by Gertrud Rehner, pp. 5371–5382, © 1995, Elsevier Ltd.

### Definition of Vitamins and Their Function

Vitamins have been defined as a group of organic substances that, in minute amounts, are essential for normal metabolism of an organism. Like other essential nutrients vitamins have to be supplied by the diet in amounts covering the individual requirements of the organism in question which by definition is unable to synthesize these compounds. For a number of cases, e.g., disease and medication impair micro-nutrient status, pregnancy, people under stress, a supplementation with synthetic vitamins might be necessary to avoid vitamin deficiency and can help to prevent degenerative disorders. The natural source of the essential compounds that have vitamin activity for humans is the metabolism of microorganisms and plants. During evolution the ability to biosynthesize these substances was lost by the higher organisms, including humans. Sporadic vitamins, however, can be synthesized by humans – because all steps of the

biosynthetic pathway have been conserved – but only to an inadequate extent, e.g., niacin from the amino acid tryptophan or vitamin D<sub>3</sub> from its precursor cholecalciferol.

Vitamins are a chemically and functionally inhomogeneous group of biomolecules. As a gross classification distinction is usually made between (1) fat-soluble and (2) water-soluble vitamins. Owing to their insolubility in water the fat-soluble vitamins A, D, E, and K can be accumulated in fat tissue and excessive intake causes hypervitaminoses. The water-soluble vitamins – vitamin B<sub>1</sub>, vitamin B<sub>2</sub>, niacin, vitamin B<sub>6</sub>, folic acid, pantothenic acid, biotin, vitamin B<sub>12</sub>, and ascorbic acid – can generally only be stored in a small amount and intake exceeding actual need is excreted in the urine.

The terms used above for the substances that are actually considered as vitamins are trivial names, mostly group names, used for more than one derivative of a compound with similar biological activity. The function of vitamins in cell metabolism is just as varied as their chemical constitution. By virtue of their lipid solubility, fat-soluble vitamins generally affect physicochemical properties in various cell membranes. Furthermore, they act at the gene level as inductors of protein biosynthesis and as redox agents. The water-soluble vitamins act in many ways as coenzymes and thus enable the catalytic function of hundreds of enzymes.

A brief review of each of the vitamins describes their chemical composition, coenzyme forms, and main functions.

## Vitamin A

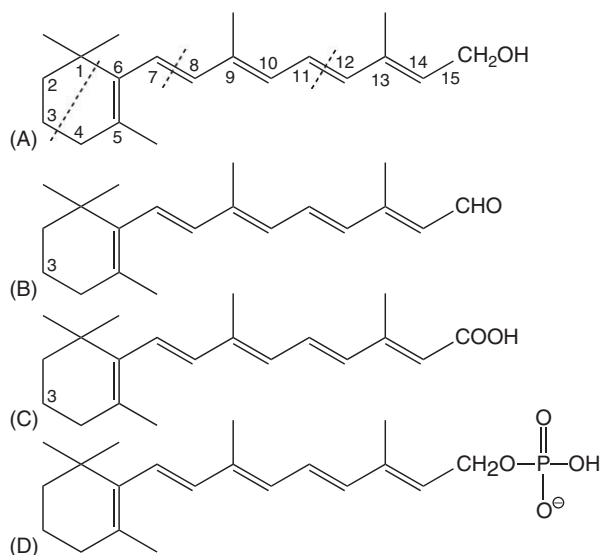
**Nomenclature.** Vitamin A: all retinoids with retinol-like biological activity. Retinoids, all native and synthetic forms of vitamin A, number ~1500 compounds.

**Chemical structure** (Figure 1). Molecules formally composed of four isoprene units; they naturally occur as an alcohol (retinol), an aldehyde (retinal), or as an acid (retinoic acid). From each of the three basic forms two variants exist: vitamin A<sub>1</sub> with a  $\beta$ -ionone ring and vitamin A<sub>2</sub> with a dehydrated  $\beta$ -ionone ring. Native retinoids show *cis-trans* isomerism of the double bonds. Carotenoids are proforms of vitamin A (Figure 2).

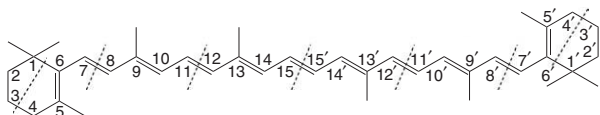
**General metabolic significance.** (1) Somatic functions: growth, development, and differentiation of epithelial and bone tissue (retinol, retinal, retinoic acid). (2) Reproduction: spermatogenesis, development of the placenta, embryonic growth (retinol, retinal). (3) Visual process: 11-*cis*-retinaldehyde is a prosthetic group of the visual pigments.

**Symptoms of deficiency.** (1) Night blindness. (2) Conjunctival and corneal xerosis. (3) Keratomalacia: ulceration and scarring of cornea that leads to loss of vision.

**Typical serum level.** About  $1.1\text{--}2.3\ \mu\text{mol l}^{-1}$ .



**Figure 1** Retinol and natural analogs. (A) All-*trans* retinol, (B) all-*trans* retinaldehyde, (C) all-*trans* retinoic acid, (D) retinylphosphate. Dotted lines in (A) separate the four isoprenoid units composing the molecule.



**Figure 2**  $\beta$ -Carotene.

## Vitamin D

**Nomenclature.** Vitamin D: all steroids with cholecalciferol-like biological activity; cholecalciferol (vitamin D<sub>3</sub>); 25-hydroxycholecalciferol (calcidiol); 1 $\alpha$ ,25-dihydroxycholecalciferol (calcitriol).

**Chemical structure** (Figure 3). 7-Dehydrocholesterol (provitamin D<sub>3</sub>) converted to cholecalciferol by UV irradiation; enzymatic hydroxylation to 25-OH-cholecalciferol in liver; enzymatic hydroxylation to 1 $\alpha$ ,25(OH)<sub>2</sub> colecalciferol in kidney.

**General metabolic significance.** Vitamin D stimulates intestinal absorption of calcium and phosphate, renal reabsorption of these ions, deposition and mobilization of minerals in the hard tissue, controlling normal calcium and phosphate blood level by means of these processes. Molecular mechanism of the vitamin D effects most frequently conform to the effect of steroid hormones (induction of protein biosynthesis).

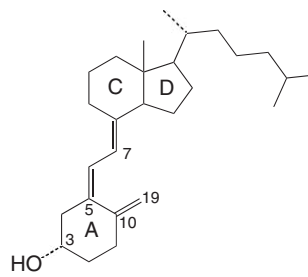
**Symptoms of deficiency.** (1) Rickets in infants; (2) osteomalacia in adults.

**Typical serum level.** About  $75\text{--}175\ \text{pmol l}^{-1}$ .

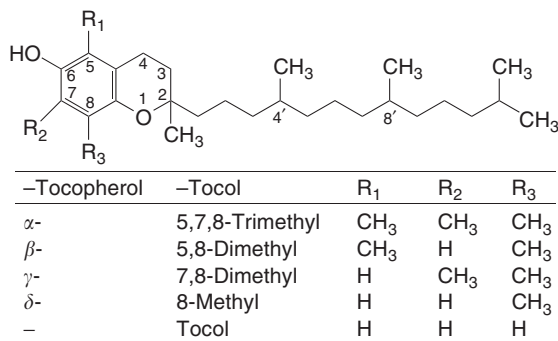
## Vitamin E

**Nomenclature.** Vitamin E: group name for derivatives of 6-chromanol.

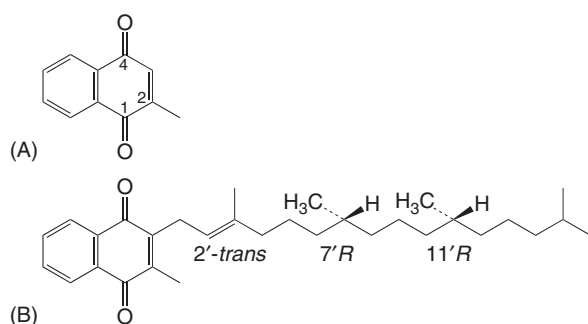
**Chemical structure** (Figure 4). Compounds composed of a chromanol-ring system, with two to four



**Figure 3** Cholecalciferol.



**Figure 4**  $\alpha$ -,  $\beta$ -,  $\gamma$ -, and  $\delta$ -tocopherol.



**Figure 5** Vitamin K: (A) vitamin K<sub>3</sub>, menadione; (B) vitamin K<sub>1</sub>, phyloquinone.

methyl groups in position 2 and a C<sub>16</sub> side chain, saturated in tocopherols and unsaturated in tocotrienols (four variants each).

**General metabolic significance.** Component of and interaction with biological membranes, protection of lipids against peroxidation, preservation of the membrane structure and integrity (exact molecular mechanism of vitamin E action unknown).

**Symptoms of deficiency.** In humans generally not due to dietary insufficiency but due to fat malabsorption syndromes or genetic abnormalities. Symptoms are difficult to categorize, e.g., reproduction disorders, abnormalities of muscles, liver, bone marrow, brain functions. At cellular level: increased oxidation of cell membranes.

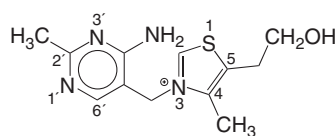
**Typical serum level.** About 12–50  $\mu\text{mol l}^{-1}$ .

## Vitamin K

**Nomenclature.** Vitamin K (menadione, vitamin K<sub>3</sub>): group name for menadione and all of its derivatives with antihemorrhagic activity in vitamin K-deficient animals.

**Chemical structure (Figure 5).** Derivatives of naphthoquinone; natural forms contain an unsaturated isoprenoid side chain on C-3; parent compound 2-methyl-1,4-naphthoquinone (menadione, not thought to occur naturally); green plants contain 2-methyl-3-phytyl-1,4-naphthoquinone (phyloquinone, vitamin K<sub>1</sub>) with one double bond in the side chain.

**General metabolic significance.** Cofactor of a specific microsomal carboxylase, catalyzing  $\gamma$ -carboxylation of specific peptide bound glutamate residues.  $\gamma$ -Carboxyglutamate containing proteins: four of the clotting factors, plasma proteins C, M, S, and Z, osteocalcin, ovocalcin, atherocalcin.  $\gamma$ -Carboxyglutamate containing proteins mediate interaction with Ca<sup>2+</sup> and have specific functions in blood clotting and bone metabolism.



**Figure 6** Thiamine.

**Symptoms of deficiency.** In humans rare and mostly resulting from fat malabsorption syndromes, liver disease, and antibiotic therapy that inhibits microbial vitamin K<sub>2</sub> synthesis in the gut.

**Typical serum level.** About 2–17  $\text{nmol l}^{-1}$ .

## Thiamine

**Nomenclature.** Thiamine, vitamin B<sub>1</sub> (aneurin).

**Chemical structure (Figure 6).** Pyrimidine and thiazole moiety linked by methylene bridge – phosphorylated forms: thiamine monophosphate (TMP), thiamine diphosphate (TDP), thiamine triphosphate (TTP).

**Coenzyme form(s).** TDP.

**Enzymes with TDP as coenzyme in human metabolism.** (1) 2-oxoacid dehydrogenase (pyruvate dehydrogenase (EC 1.2.4.1)); 2-oxoglutarate dehydrogenase (EC 1.2.4.2); branched chain-2-oxoacid dehydrogenase(s) (EC 1.2.4.4); (2) transketolase (EC 2.2.1.1).

**General metabolic significance.** TDP-dependent enzymes in key positions of cell metabolism (carbohydrate catabolism, citrate cycle). Thiamine triphosphate, neurophysiologically active form of thiamine without coenzyme function (biochemical role not completely elucidated).

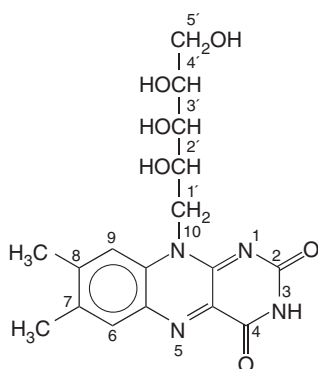
**Symptoms of deficiency.** Fatigue, irritability, lack of concentration due to marginal deficiencies: e.g., during pregnancy, lactation, heavy physical work, regular heavy alcohol intake. Beriberi includes dry beriberi (muscle wasting with heart involvement, tachycardia, etc.) and wet beriberi (edema, anorexia, muscle weakness, ataxia, peripheral paralysis) results from prolonged low dietary intake. Wernicke syndrome is characterized by mental disorder and frequently seen in alcoholics after long periods of alcohol intake without food intake.

**Typical serum level.** Less than 75  $\text{nmol l}^{-1}$ .

## Riboflavin

**Nomenclature.** Riboflavin, vitamin B<sub>2</sub> (lactoflavin, ovoflavin).

**Chemical structure (Figure 7).** 7,8-Dimethyl-10-(1'-D-ribityl)isoalloxazin – different redox states: flavochinon (Fl<sub>ox</sub>), flavosemichinon (Fl-H), flavohydrochinon (Fl<sub>red</sub>H<sub>2</sub>).



**Figure 7** Riboflavin.

*Coenzyme form(s).* FMN (flavin mononucleotide, riboflavin monophosphate), FAD (flavin adenine dinucleotide, riboflavin adenosine diphosphate).

*Enzymes with flavin coenzymes in human metabolism* (examples of ~100 enzymes). Electron-transfer flavo-protein (EC 1.3.99.2-3) FAD; NADPH-cytochrome-P-450-reductase (EC 1.6.2.4) FMN and FAD; succinate dehydrogenase (EC 1.3.99.1) FAD; D-amino acid oxidase (EC 1.4.3.3) FAD; L-amino acid oxidase (EC 1.4.3.2) FMN.

*General metabolic significance.* Flavin coenzyme dependent enzymes catalyze one-electron-transfer reactions, dehydrogenase reactions, reduction of  $O_2$  to  $H_2O_2$ , hydroxylations, etc. – numerous flavin enzymes dealing with biological oxidation.

*Symptoms of deficiency.* Oral, ocular, cutaneous, and genital lesions. Primary deficiency is associated with inadequate consumption of milk and other animal products. Secondary deficiencies are most common in chronic diarrheas, liver disease, chronic alcoholism, and postoperative situations.

*Typical serum level.* 110–640 nmol l<sup>-1</sup>.

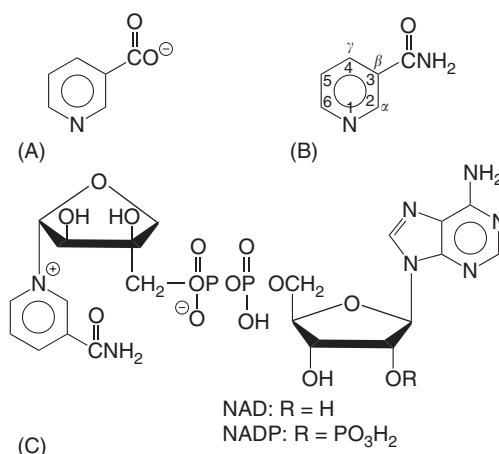
## Niacin

*Nomenclature.* Niacin (vitamin PP): group name for nicotinic acid, nicotinamide, NAD, and NADP.

*Chemical structure* (Figure 8). Nicotinic acid (pyridine-3-carboxylic acid), nicotinamide (pyridine-3-carboxamide), NAD (nicotinamide adenine dinucleotide), NADP (nicotinamide adenine dinucleotide phosphate).

*Coenzyme form(s).* NAD, NADH, NADP, NADPH.

*Enzymes with niacin coenzymes in human metabolism* (examples of ~200 enzymes!). L-Lactate dehydrogenase (EC 1.1.1.27); alcohol dehydrogenase (EC 1.1.1.1); glyceraldehyde-phosphate dehydrogenase (EC 1.2.1.12); NADPH-cytochrome-P-450-reductase (EC 1.6.2.4).



**Figure 8** Niacin: (A) nicotinic acid, (B) nicotinamide, (C) nicotinamide adenine dinucleotide diphosphate.

*General metabolic significance.* Coenzymes participate in most hydrogen transfer reactions in every cell, thus in synthesis and catabolism of carbohydrates, fatty acids, amino acids, etc., as well as in production of metabolic energy. Biosynthesis is possible to a small degree from tryptophan.

*Symptoms of deficiency.* Marginal: multiple symptoms such as insomnia, loss of appetite, weight and strength loss, soreness of the tongue and mouth, burning sensations nervousness, etc. Pellagra: dermatosis, dementia, diarrhea. Risk groups: alcoholics and people from areas where corn is the major diet.

*Typical urine levels.* About 7–10 mg day<sup>-1</sup> (measured as N-methylnicotinamide).

## Pyridoxine

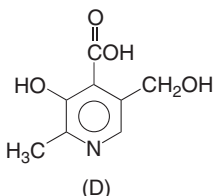
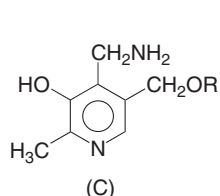
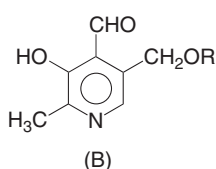
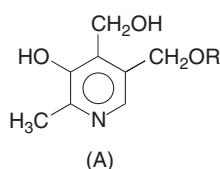
*Nomenclature.* Pyridoxine (vitamin B<sub>6</sub>): group name for pyridoxine (pyridoxol), pyridoxal, pyridoxamine and the 5'-phosphorylated derivatives of these compounds.

*Chemical structure* (Figure 9). Pyridoxine (3-hydroxy-2-methylpyridine) basal compound of the group; substitution (R) is carried out on 5'-C; pyridoxic acid is a catabolite of the compounds.

*Coenzyme(s).* PLP (pyridoxal phosphate), PMP (pyridoxamine phosphate).

*Enzymes with pyridoxine coenzymes in human metabolism* (examples, several hundred enzymes). (1) Transaminases (aspartate-aminotransferase (EC 2.6.1.1), alanine-aminotransferase (EC 2.6.1.2); aromatic-amino acid aminotransferase (EC 2.6.1.57)); (2) decarboxylases (histidine-decarboxylase (EC 4.1.1.22); aromatic-amino acid decarboxylase (EC 4.1.1.28)); (3)  $\alpha,\beta$ -elimination (L-serine-dehydratase (EC 4.2.1.13)).





**Figure 9** Vitamin B<sub>6</sub>: (A) pyridoxine, (B) pyridoxal, (C) pyridoxamine, (D) pyridoxic acid.

*General metabolic significance.* Pyridoxine dependent enzymes catalyze almost exclusively reactions of amino acid metabolism, also participating in heme synthesis and generation of biogenic amines.

*Symptoms of deficiency.* General: weakness, insomnia, nervous disorders, appetite and growth depression, dermatologic disorders. Vascular: arteriosclerosis, anemia, paralysis. Congenital: thrombosis, mental retardation, neuropathies.

*Typical serum level.* About 15–37 nmol l<sup>-1</sup>.

## Folic Acid

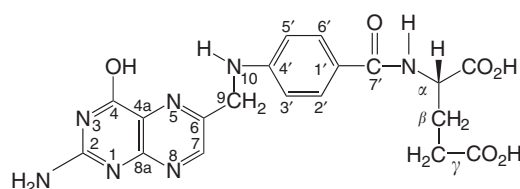
*Nomenclature.* Folic acid, pteroylglutamic acid (PteGlu) ‘folates’ also used as a group name for a family of more than 100 related compounds that have been modified on the pterin moiety, on N-5 and/or N-10, and on glutamyl side-chain.

*Chemical structure (Figure 10).* Pteroylglutamic (PteGlu) is composed of a pterin moiety, 4-aminobenzoic acid and glutamic acid and is the parent substance numerous ‘folates’.

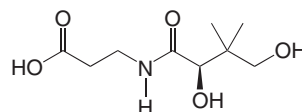
*Coenzyme form(s).* Tetrahydrofolic acid (H<sub>4</sub>PteGlu); 5-methyltetrahydrofolic acid (5-CH<sub>3</sub>-H<sub>4</sub>PteGlu); 5,10-methylenetetrahydrofolic acid (5,10-CH<sub>2</sub>-H<sub>4</sub>PteGlu); 5,10-methenyltetrahydrofolic acid (5,10-CH<sup>+</sup>=H<sub>4</sub>PteGlu); formyltetrahydrofolic acid (5-HCO-H<sub>4</sub>PteGlu); 10-formyltetrahydrofolic acid (10-HCO-H<sub>4</sub>PteGlu); 5-formiminoj tetrahydrofolic acid (5-HCNH-H<sub>4</sub>PteGlu).

*Enzymes with folate coenzymes in human metabolism* (examples, ~20 enzymes). Thymidylate-synthase (EC 2.1.1.45); serine-hydroxymethyl-transferase (EC 2.1.2.1); formiminoglutamate-formiminotransferase (EC 2.1.2.5).

*General metabolic significance.* H<sub>4</sub>PteGlu (main coenzyme) transfers one-carbon units (with or without modification) from an appropriate donor to an



**Figure 10** Folic acid.



**Figure 11** D-(+)-Pantothenic acid.

appropriate acceptor. One-carbon units activated by H<sub>4</sub>PteGlu participate in the biosynthesis of purines, thymidylic acid, and certain amino acids and in activation of tRNA.

*Symptoms of deficiency.* Marginal: tiredness, irritability, loss of appetite. Severe: megaloblastic anemia, abdominal pain, diarrhea, ulcers, skin problems, hair loss, neurological disorders, depression. Deficiency during pregnancy may lead to premature birth or birth defect.

*Typical plasma/serum levels.* About 9–13.5 nmol l<sup>-1</sup>.

## Pantothenic Acid

*Nomenclature.* The term pantothenic acid is used for the ‘free’ acid, i.e., pantothenic acid itself as well as for the physiologically active forms coenzyme A and acyl-carrier protein.

*Chemical structure (Figure 11).* Pantothenic acid is composed of β-alanine and 2,4-dihydroxy-3,3-dimethylbutyric acid (pantoic acid), linked acid-amide-like.

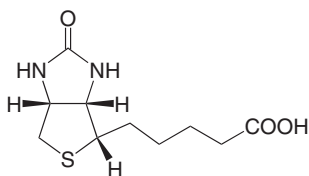
*Coenzyme form(s).* (1) Coenzyme A, composed of cysteamine, pantothenic acid, phosphate, and adenosine-3',5'-diphosphate; (2) acyl-carrier protein.

*Biochemical function in human metabolism.* Activation of metabolites by coenzyme A while a thioester as a high-energy compound is generated. Examples: acetyl CoA, succinyl-CoA, acyl-CoA-derivates. The acyl-carrier protein is a component of the fatty acid-synthetase complex. Both coenzymes transfer acyl groups.

*General metabolic significance.* Acetyl CoA is a central compound of cell metabolism linking catabolic processes with generation of biological energy. Acetyl CoA is also a starting metabolite for biosynthesis of fatty acids, sterols, ketone bodies, etc.

*Symptoms of deficiency.* Headache, fatigue, insomnia, intestinal disturbance, paresthesia. (Only in severely malnourished patients!)





**Figure 12** Biotin.

*Typical urine levels.* About 7–10 mg day<sup>-1</sup> (measured as pantothenic acid).

### Biotin

*Nomenclature.* Biotin (vitamin H).

*Chemical structure* (Figure 12). Biotin (hexahydro-2-oxo-1H-thieno[3,4-d]imidazole-4-pentanoic acid). Biologically active analogs: biocytin ( $\epsilon$ -N-biotinyl-L-lysine), oxybiotin (S substituted by O).

*Enzymes with biotin as coenzyme in human metabolism.* Pyruvate carboxylase (EC 6.4.1.1); acetyl CoA carboxylase (EC 6.4.1.2); propionyl CoA carboxylase (EC 6.4.1.3); 3-methylcrotonyl CoA carboxylase (EC 6.4.1.4).

*General metabolic significance.* Biotin coenzymes participate in gluconeogenesis, biosynthesis of fatty acids, and metabolism of amino acids.

*Symptoms of deficiency.* Anorexia, nausea, vomiting glossitis, pallor, mental depression, alopecia, dry scaly dermatitis, increase in serum cholesterol and bile pigments. Very rare, e.g., after prolonged consumption of raw eggs.

*Typical serum/plasma level.* About 100–400 pmol l<sup>-1</sup>.

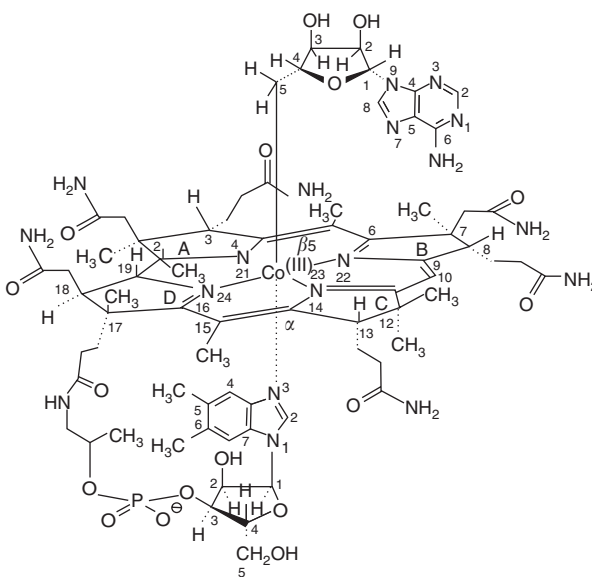
### Vitamin B<sub>12</sub>

*Nomenclature.* Vitamin B<sub>12</sub> (also used as a group name), cyanocobalamine (antipernicious anemia factor).

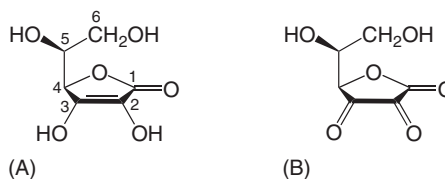
*Chemical structure* (Figure 13). Cyanocobalamine is composed of a corrin nucleus with a cobalt atom in its center. The cobalt atom is coordinately linked to a nitrogen of a 5,6-dimethylbenzimidazole group. Furthermore, the molecule contains ribose with an  $\alpha$ -glycosidic linkage. Substituents at the cobalt atom can be: CN group (cyanocobalamine), 5'-desoxyadenosyl (adenosylcobalamine), CH<sub>3</sub> (methylcobalamine) (numerous synthetic derivatives with different substitutes).

*Coenzyme form(s).* Adenosylcobalamine, methylcobalamine.

*Enzymes with vitamin B<sub>12</sub> as coenzyme in human metabolism.* (1) With adenosylcobalamine: methylmalonyl CoA-mutase (EC 5.4.99.2); leucine-2,3-aminomutase (EC 5.4.3.7); (2) with methylcobalamine: tetrahydropteroyltriglutamate methyltransferase (EC 2.1.1.13).



**Figure 13** Adenosylcobalamine.



**Figure 14** Vitamin C: (A) L-ascorbic acid, (B) dehydro-L-ascorbic acid.

*General metabolic significance.* Vitamin B<sub>12</sub> is essential to avoid certain forms of anemia and neurological disturbances.

*Symptoms of deficiency.* Pernicious anemia: glossitis, weakness, loss of appetite, loss of taste and smell, impotence, irritability, memory impairment, mild depressions, hallucinations. Cause of deficiency is usually not insufficient dietary intake but, e.g., inadequate digestion, lack of binding, and intrinsic factors.

*Typical serum/plasma level.* About 162–694 pmol l<sup>-1</sup>.

### Ascorbic Acid

*Nomenclature.* Ascorbic acid, vitamin C (hexuronic acid, antiscorbutic vitamin).

*Chemical structure* (Figure 14). L-Ascorbic acid (2,3-endi-ol-L-gulonic acid- $\gamma$ -lactone), dehydro-L-ascorbic acid (2-oxo-L-gulonic acid- $\gamma$ -lactone).

*General metabolic significance.* Ascorbic acid participates in numerous biological events concerning electron transport reactions, hydroxylations, oxidative catabolism of aromatic amino acids; one of the most important biologic redox systems (no coenzyme function).

*Symptoms of deficiency.* Scurvy: fatigue, bleeding gums, muscle pain, depression, dry skin, impaired iron absorption, impaired folate metabolism, impaired wound healing.

*Typical serum/plasma level.* About 20–100  $\mu\text{mol l}^{-1}$ .

## Physicochemical Properties of the Vitamins

Table 1 lists some physicochemical characteristics of the vitamins that might be relevant for the analysis of these compounds.

## Generalized Problems of Vitamin Analysis

The analysis of vitamins is a very wide and exciting field for analysts, due to a number of challenges that correspond to the nature of these analytes, e.g., their structural diversity with correspondingly different physical and chemical properties, the wide concentration ranges encountered, the relative instability of many vitamins (Table 1), and the occurrence of vitamers, metabolites, and various bound forms. Chemical diversity ranges from the relatively small and stable nicotinic acid (M 123) to the complex cobalamines, which belong to the largest known non-peptide, nonpolymer natural products (M ~1350). Typical concentration ranges encountered vary from picograms per milliliter, e.g., cobalamin in serum, to pure compounds analyzed in industrial quality control. All these facts pose high demands on methodology and analytical skills. It is thus not very surprising that a large array of techniques, ranging from traditional methods, like microbiology, to state-of-the-art technology, like high-performance liquid chromatography (HPLC) and tandem mass spectroscopy (MS/MS), have been used, and that even today vitamin analysis remains a challenge in many ways.

Methods employed largely vary from sample material to sample material, and for determination of most vitamins many different analytical methodologies can be used. Care must be taken in choosing the best method. Different methods can yield different results, e.g., due to different sample preparation or different specificity of the method toward the analyte. For example, the result of microbiological vitamin determinations is usually a 'biological activity' expressed as an amount of pure vitamin showing the same effect in the test. This assumption is often correct, but the possible presence and influence of, e.g., vitamers or metabolites with vitamin

activity also has to be taken into account. Thus, the same sample analyzed by HPLC–MS may yield a different vitamin content because with this technique chemically defined compounds are determined. The possible differences between results of different methods applied to the same sample should also be taken into account when literature data are examined. In the following, sampling, sample preparation, and different general methodologies for vitamin analysis are discussed.

## Sampling and Sample Preparation

During drawing a representative sample aliquot it has to be taken into account that many vitamins are sensitive to temperature, light, oxygen, or changes in pH, and should be analyzed as soon as possible after sampling. Especially natural samples such as vegetables and fruits are sensitive, and cannot be stored for long, even if kept deep-frozen, without losing significant amounts of their vitamin content. For example, the vitamin C content of vegetables can be used as an indicator of the conditions of their preparation or storage. Also important is the supply of a sample volume large enough for analysis, reflecting both the low vitamin contents in many samples, and a possible inhomogeneity of the material.

After sampling, the sample is prepared for analysis. Usually, each different type of sample material requires specially adapted sample preparation, extracting the vitamins from the matrix. This is often combined with a preliminary concentration and purification step. Roughly three application fields can be differentiated that pose different demands: determination of vitamin content in samples from human subjects and animals, e.g., plasma and tissue; determination of vitamin content in feed, food, plants, processed plants, plant products, and microorganisms; and determination of added vitamin content in fortified products, e.g., tablets, feed, or food.

Sample preparation should liberate the vitamin from the matrix, e.g., tissue or plasma, where it often occurs chemically or physically bound. Many specific transport binding proteins are known, e.g., for retinol. 'Chemical bonding' can include the incorporation of a vitamin into coenzymes, e.g., niacin and pantothenate in NAD and coenzyme A, respectively. Liberating vitamins from industrial product forms (formulations) is also an important issue. Here, vitamins are often encapsulated in small beadlets, e.g., from gelatin, which protects them from oxygen and makes them easier to add during processing.

To protect vitamins from decomposition, antioxidants such as BHT can be added during sample preparation. Saponification using alkaline alcoholic

**Table 1** Some important physicochemical properties of vitamins

Vitamin	Summation formula	Relative molecular mass	Factors influencing stability <sup>b</sup>								
			Absorption maximum <sup>a</sup> wavelength (nm)	Approximate solubility in H <sub>2</sub> O	O <sub>2</sub>	Reducing agents	Acid	Alkaline	Metal ions	100° C	Light/UV
Vitamin A	C <sub>20</sub> H <sub>30</sub> O (retinol)	286.4	325 (ethanol); 328 (isopropanol)	Insoluble	1	st	1	st	1	st	1
Carotenoids	C <sub>40</sub> H <sub>56</sub> (β-Carotene)	536.9	273; 452; 481 (petrol ether)	Insoluble	1	–	1	st	1	st	1
Vitamin D	C <sub>27</sub> H <sub>44</sub> O (vitamin D <sub>3</sub> )	384.6	264 (ethanol)	Insoluble	1	st	1	(st)	1	1	1
Vitamin E	C <sub>29</sub> H <sub>50</sub> O <sub>2</sub> (α-Tocopherol)	430.7	292 (ethanol)	Insoluble	1	st	st	(st)	1	st	(1)
Vitamin K	C <sub>31</sub> H <sub>46</sub> O <sub>2</sub> (vitamin K <sub>1</sub> )	450.7	243; 249; 260; 269; 325 (hexane)	Insoluble	(1)	(1)	st	1	1	st	1
Thiamine	C <sub>12</sub> H <sub>17</sub> ON <sub>4</sub> S × HCl (vitamin B <sub>1</sub> × HCl)	337.3	245 (0.1 mol l <sup>-1</sup> HCl)	1 g ml <sup>-1</sup>	(1)	(1)	st	1	1	(1)	1
Riboflavin	C <sub>17</sub> H <sub>20</sub> O <sub>6</sub> N <sub>4</sub>	376.4	233; 267; 374; 444	0.07 mg ml <sup>-1</sup>	(1)	(1)	st	1	1	st	1
Niacin	C <sub>6</sub> H <sub>5</sub> O <sub>2</sub> N (nicotinic acid)	123.1	261	0.2 g ml <sup>-1</sup>	st	(1)	st	st	st	st	st
Pyridoxine	C <sub>8</sub> H <sub>9</sub> O <sub>3</sub> N (pyridoxal)	167.2	291 (acidic); 254, 324 (neutral); 245, 309 (alkaline)	0–2 g ml <sup>-1</sup>	(1)	(1)	st	(1)	–	st	1
Folic acid	C <sub>19</sub> H <sub>19</sub> O <sub>6</sub> N <sub>7</sub>	441.4	256; 283; 365 (0.1 mol l <sup>-1</sup> NaOH)	Low solubility (higher in diluted alkali metal hydroxide)	1	(1)	1	(1)	1	1	1
Pantothenic acid	C <sub>9</sub> H <sub>16</sub> O <sub>5</sub> NNa (Na salt)	241.2	–	0.4 g ml <sup>-1</sup>	st	st	1	1	–	(1)	–
Biotin	C <sub>10</sub> H <sub>16</sub> O <sub>3</sub> N <sub>2</sub> S	244.3	–	0–2 g ml <sup>-1</sup>	(1)	(1)	1	1	–	st	(1)
Vitamin B <sub>12</sub>	C <sub>63</sub> H <sub>88</sub> O <sub>14</sub> N <sub>14</sub> PCo (CN-cobalamine)	1355.4	278; 361; 550	0.01 g ml <sup>-1</sup>	(1)	(1)	1	1	(1)	(1)	(1)
Ascorbic acid	C <sub>6</sub> H <sub>8</sub> O <sub>6</sub>	176.1	245 (pH 2); 265 (pH 6.4); 300 (pH 14)	0.9 g ml <sup>-1</sup>	1	st	st	1	1	(1)	(1)

<sup>a</sup>The solvent is in brackets (if essential).<sup>b</sup>1 = high lability; (1) = low lability; st = high stability; (st) = low stability.

From different sources (see Further Reading).

solution is used for vitamins such as retinol or tocopherol, which are stable under these conditions. A milder way is enzymatic sample treatment, e.g., with proteases and collagenases. This approach is especially used for plasma and tissue samples, and also for analysis of beadlets. A disadvantage can be a natural vitamin content of commercially available enzymes, which has to be determined and corrected accordingly. A special application is the determination of total pantothenate, including those present in coenzyme A. Liberation requires alkaline phosphatase, and a specific enzyme, pantetheinase, which has to be purified from pigeon liver prior to analysis. Denaturation of binding protein by addition of organic solvents is often used to free vitamins for plasma analysis, e.g., determination of retinol and tocopherol.

### Microbiological Methods

Historically, vitamins have been discovered as growth factors or compounds essential for life of humans, certain animals, and microorganisms. As these findings usually occurred before the exact chemical nature of the underlying active principle was known, experiments with animals or microorganisms were often the only way to test for the presence of vitamins. As soon as their structures were known and reference compounds became available, these tests could also be run in a (semi)-quantitative way. Set on a diet without the vitamin in question, the animal developed typical symptoms that could be cured by vitamin supplementation. These tests had some important disadvantages. It took relatively long before a result could be obtained, and, more important, the result often lacked specificity and accuracy as response varies from species to species. Applied to unknown samples often no chemically defined substance but rather biological activity could be determined.

Thus, of these methods mainly those using microorganisms are still in use, predominately for vitamins that are difficult to determine by classical instrumental analytical methods. This can be due to lack of suitable chromophores (biotin and pantothenate), very low concentrations (e.g., biotin and vitamin B<sub>12</sub>), or the presence of many active forms (folate) that are difficult to determine separately. The microbiological assay involves incubation of a sample extract with a suitable test organism for which the vitamin is essential for growth, often *Lactobacillus* spp. The growth of the bacteria in response to the vitamin can be monitored turbidimetrically and is quantified in relation to incubation with standard solutions. Main disadvantages are the long duration of the assay, usually 24 h, the chemical nonspecificity of the result (determination of

a biological activity), and its susceptibility to antibiotics and other inhibiting substances.

### HPLC

The method of choice for the determination of most vitamins is HPLC due to its high separation capability, its mild analytical conditions, and the possibility to use various specifically adapted detection methods, e.g., UV, fluorescence, or MS detection. All fat-soluble vitamins and most water-soluble vitamins have chromophores suitable for UV detection. Separation of vitamers and stereoisomers can be achieved. If a higher sensitivity is required HPLC with fluorescence detection can be used, either directly (e.g., vitamins A and E) or after derivatization (e.g., thiamine). A further improvement in sensitivity and specificity has been achieved by introducing HPLC with mass spectrometric detection in vitamin analysis. Due to the structural information retrievable, e.g., molecular mass, fragmentation pattern, this is the method of choice for analysis of samples with complex mixtures or low vitamin concentrations. Examples for the use of HPLC–MS in vitamin analysis include the determination of 25-hydroxy-D<sub>3</sub> and pantothenic acid. However, one drawback of mass spectrometry is the need for an isotopically labeled reference compound for reliable quantification. Due to the structural complexity of many vitamins, these reference compounds are often expensive and difficult to synthesize. An interesting unique application is the determination of vitamin B<sub>12</sub> by HPLC–IPC–MS, which is possible due to its cobalt content.

### Enzyme Assays

Enzyme assays are mainly used for determination of body status of vitamins. As vitamins usually function either as coenzymes or building blocks of coenzymes, the activity of the vitamin-dependent enzymes is a measure of vitamin status. Usually, the assay is carried out by determining the enzyme activity with and without activation by added coenzyme. The activity can be monitored by measuring changes in concentration of substrates or products during the reaction. An activation coefficient can be deduced, which reflects the status of the enzyme investigated, and thus the vitamin status. Most assays are conducted with whole blood or the separated erythrocyte fraction. They can be automated with clinical analyzers. Disadvantages include difficulties in assay standardization, instability of the enzymes during storage, and misleading results, e.g., due to conditions other than vitamin deficiency leading to low apoenzyme concentrations.

### Immuno- and Affinity Assays

Immuno- and affinity assays include a diverse range of methods, which have in common that at least one step is based on a selective binding of the vitamin to a binder, e.g., a transport protein. In immunoaffinity chromatography, this binding is used as a purification and/or concentration step. The vitamin, e.g., biotin, is loaded onto a column containing a selective binding protein, e.g., streptavidin. The vitamin is selectively retained on the column. After eluting other nonbinding sample ingredients the vitamin-protein complex is cleaved, and the now purified and concentrated vitamin is eluted.

In quantitative immunoassays, e.g., enzyme-linked immunosorbent assay and radioimmunoassay, a known amount of labeled vitamin is mixed with sample extract in which the vitamin content should be determined. Methods of labeling include radioisotopes (e.g., cobalamine), fluorescence, or luminescence markers (e.g., folate). The mixture is subjected to binding agent, equally forming complexes with both labeled and unlabeled vitamin. This complex is then isolated, and the amount of labeled vitamin present is measured. Sample vitamin concentration can be deduced from the ratio of labeled vitamin added to labeled vitamin measured after isolation. Advantages of immunoassays are short analysis time, and the possibility of automating them on clinical analyzers.

### Other Methods

Capillary electrophoresis can be used instead of HPLC for separation of complex samples, but has not yet found a broader application in vitamin analysis. Gas chromatography is rarely used as most vitamins are not volatile or stable enough for direct analysis. Applications include determination of vitamin D after derivatization.

### Automation of Vitamin Analysis

Automation of vitamin analysis is becoming more and more important, as it potentially offers significant time and resource savings. Automation is best applicable if large numbers of similar samples have to be analyzed. This is the case in clinical chemistry and in industry, for example, for quality control. In clinical chemistry many parameters including vitamins, e.g., ascorbic acid, folate, or cobalamine, can be determined using automated analyzers with a throughput of several hundreds of samples per day. These analyzers have only to be fed with the samples and reagents. Specialized software programs control the fully automated analyses, including calibration, dilutions, derivatizations, quantification, quality control, and interpretation of the results. Automation in

industry focuses on the use of bench-top robotic systems, e.g., for analysis of tablets, formulated vitamins, and product forms. Automated sampling and analyzing is used both for process and quality control.

## Interpretation of Analytical Data in Relation to Body Status

Chronically insufficient vitamin supply is undoubtedly a health risk. Severe vitamin deficiencies are characterized by symptoms that can be diagnosed quite clearly. They usually result from a highly insufficient intake of one or more vitamins over a long period of time and such deficiencies are still widespread in the developing countries. In the industrialized areas clinically distinct vitamin deficiencies are rare but marginal deficiency might be common in certain population groups, e.g., low income groups, teenagers, the elderly, alcoholics, pregnant and lactating women. To protect these groups from the consequences of insufficient vitamin supply, it is of great interest to have suitable methods to estimate the body status of vitamins. The skillful analytical techniques developed in the last decade are adequate for an exact quantification of vitamins in biological material. Less satisfying, however, is the evaluation of the analytical data concerning the true reflection of the 'body status' that mostly means tissue saturation with a certain vitamin. Threshold values concerning 'unsatisfactory supply' versus 'undoubtedly satisfactory supply' are difficult to define and the subject of controversial discussion.

In epidemiological examinations certain methods may also be used that yield statistically significant results at the population level. Nevertheless, the examination of an individual should lead to a clear evaluation of his or her particular nutritional situation.

To judge the value of the analytical data for estimation of the body status, knowledge of the intra-individual variation, i.e., day-to-day variation, is as important as information about the biological variation, i.e., reaction of the values to short and long time changes of intake and to other factors.

### Vitamin A

Concentration of vitamin A in blood serum is very stable and only drops when liver stores are exhausted. Thus, serum level does not reflect vitamin A status of the whole body. In the opinion of the World Health Organization neither intake data nor biochemical or clinical parameters are appropriate to estimate vitamin A status in humans. It recommends the use of more than one parameter to judge the risk of a marginal deficiency: intake records, level of plasma



retinol binding protein, and clinical symptoms. The value of determination of serum carotenoid concentration as an indicator is controversial.

### Vitamin D

The vitamin D plasma level does not adequately reflect its status in the whole body. Therefore, for the diagnosis of, for example, bone mineralization malfunctions, the determination of 25-OH-vitamin D<sub>3</sub> is preferred as a recognized clinical determinant.

### Vitamin E

A generally accepted procedure to estimate vitamin E status in humans does not seem to exist. However, in a number of studies a positive correlation between plasma cholesterol level and the levels of total lipids including cholesterol, triglycerides, etc., was observed. Thus, the plasma level of tocopherol should always be set in relation to the plasma lipids. A usual value of tocopherol is, e.g., 0.8 mg tocopherol per gram total lipids.

Other methods are also used for the determination of plasma/serum levels: *in vitro* lipid peroxidation in erythrocytes (malondialdehyde test); determination of vitamin E concentration in platelets; determination of tocopherols in fat tissue (needle biopsy).

### Vitamin K

Owing to different physical methods developed in the last few years a direct and specific determination of various derivatives of vitamin K in blood plasma is now practicable. These analytical data are considered as reflecting body status adequately. The conventional method of estimating the supply situation is, however, based on determining functions of vitamin-K-dependent plasma factors, especially of the prothrombin time. A more sensitive and more specific method to determine normal and abnormal prothrombin is radioimmunoassay.

### Thiamine

Standard methods for assessment of thiamine status used to be determination of erythrocyte transketolase ( $\alpha$ -ETK) activity (EC 2.2.1.1) with and without stimulation of this enzyme by addition of TDP cofactor (TDP TK effect). A TDP TK effect >15% is considered to show some degree of deficiency, whereas values >22% are considered to indicate severe deficiency. Technical difficulties, including standardization of the assay, instability of the enzyme during storage, and various conditions possibly influencing apoenzyme concentrations led to an increasing use of direct determination of TDP in whole blood, e.g., by HPLC in order to assess thiamine status. The HPLC assay is more robust and easier to perform. Thiamine

status determined by this method is considered to be in good correlation with results from TK activation assay. Usually, whole blood concentrations of 66.5–200 nmol l<sup>-1</sup> are encountered.

### Riboflavin

Riboflavin is the precursor for two important coenzymes, flavine adenine dinucleotide (FAD) and flavin mononucleotide (FMN), which represent the main 'body store' of riboflavin. Body status can be determined by direct and indirect methods. Direct methods include the determination of FAD and FMN in whole blood by HPLC. Usually, whole blood concentrations (FAD) of 175–475 nmol l<sup>-1</sup> are encountered. Another possibility for riboflavin status assessment is monitoring of urinary excretion. Values <27 µg per g creatinine point to deficiency, 27–79 µg per g creatinine are considered to be marginal, and values >80 µg per g creatinine to be normal. Urinary excretion sharply rises after tissue saturation is reached. Indirect methods include determination of erythrocyte glutathion reductase activity (EC 1.6.4.2), an FAD dependent enzyme. Results are often expressed as erythrocyte glutathion reductase activation coefficient (EGRAC), which reflects the ratio of apoenzyme to activated enzyme. Literature values indicating deficiency and normal status vary to some degree. EGRAC coefficients <1.20–1.4 are considered to be normal, whereas values >1.3–1.4 (–1.7) point to riboflavin deficiency.

### Niacin

The status of niacin in relation to most other vitamins is different as it can be synthesized by humans to some extent from tryptophan. Body status determination has been based on the determination of urinary excretion of niacin metabolites, predominately *N*-methyl-2-pyridone-5-carboxylamide and *N*-methyl-nicotinamide. The ratio of these compounds has been used as indicator of niacin status. Recent studies suggest that the determination of the two niacin-derived coenzymes, NAD and NADP, in erythrocytes, and their ratio are more reliable indicators of niacin status. However, a broadly accepted and easy to use determination method does not seem to exist.

### Pyridoxine

Pyridoxine status can be determined both by direct and indirect methods. Direct methods include determination of pyridoxal-5'-phosphate (PLP) in whole blood, and determination of urinary excretion of 4-pyridoxic acid (4-PA). The method of choice for quantification of both compounds is HPLC. Usually, whole blood concentrations of 35–110 nmol l<sup>-1</sup> PLP



are encountered. Concentrations of PLP have been found to be in good correlation with the pyridoxine status determined by indirect methods. Indirect methods measure the stimulated activation of pyridoxine-dependent enzymes in erythrocytes by addition of PLP. Mainly erythrocyte alanine aminotransferase activation coefficient (EALT-AC; EC 2.6.1.2) or erythrocyte aspartate aminotransferase activation coefficient (EAST-AC; EC 2.6.1.1) are determined. The coefficient activity with stimulation to activity without stimulation indicates the pyridoxine status. For EAST-AC, values  $>1.8$  are suggested to show deficiency,  $1.7$ – $1.8$  to be marginal, and  $<1.7$  to be adequate.

### Folic Acid

Folate status seems to be best reflected by whole blood concentration, as erythrocytes have significantly higher folate content than serum. Erythrocyte folate levels are also less susceptible to dietary fluctuations. However, methods for folate analysis in erythrocytes seem to be not yet as reliable as methods for serum analysis. Automated immunoassays that can be run on clinical analyzers have widely replaced the microbiological determination of folate. Serum concentrations  $<5.7 \text{ nmol l}^{-1}$  are suggested to point to deficiency,  $5.7$ – $11.4 \text{ nmol l}^{-1}$  to be marginal, and  $>11.4 \text{ nmol l}^{-1}$  to be adequate. Increased homocysteine levels can also point to folate deficiency.

### Biotin

Whereas severe dietary deficiency of biotin is very rare, except after consuming larger amounts of raw egg white, recent studies indicated that marginal deficiency might be more common than previously thought. Body status can be determined by measuring activity and/or activation of biotin-dependent enzymes, predominately carboxylases, e.g., erythrocyte pyruvate carboxylase (EC 6.4.1.1) by added biotin. More convenient methods are direct determination of biotin in plasma or serum by microbiological methods or avidin binding assays, or determination of excretion of biotin and 3-hydroxyisovaleric acid in urine. Serum concentrations from  $\sim 100$  to  $400 \text{ pmol l}^{-1}$  characterize a normal concentration range.

### Pantothenic Acid

As this vitamin is present in nearly all foods of plant and animal origin, dietary deficiency is practically unknown. Accordingly, little research has been conducted about assays to assess pantothenate status in man. Body status can be deduced from amounts of pantothenate excreted in urine, which equals  $4.5$ – $32 \text{ } \mu\text{mol day}^{-1}$ . A more convenient approach is

determination of pantothenate in serum, or preferably whole blood, by microbiological methods. Whole blood levels typically range from  $0.9$  to  $1.5 \text{ } \mu\text{mol l}^{-1}$ .

### Vitamin B<sub>12</sub>

Major vitamin B<sub>12</sub>-dependent metabolic processes include the formation of methionine from homocysteine, and the formation of succinyl coenzyme A from methylmalonyl coenzyme A. Thus, apart from directly determining vitamin B<sub>12</sub> concentration in serum, elevated levels of both methylmalonic acid and homocysteine may indicate a vitamin B<sub>12</sub> deficiency. Serum cobalamine concentration is often determined by automated immunoassays using an intrinsic factor as binding agent. These assays have mainly replaced the microbiological methods. Literature data about vitamin B<sub>12</sub> concentration in serum varies. Values  $<110$ – $150 \text{ pmol l}^{-1}$  are considered to reflect deficiency, whereas values  $>150$ – $200 \text{ pmol l}^{-1}$  represents an adequate status.

### Vitamin C

Vitamin C status can be assessed by measuring plasma levels or urinary excretion. However, due to practical disadvantages, e.g., quantitative sampling of urine and instability of vitamin C, determination of vitamin C in urine has been mainly replaced by determination of plasma levels. Methods include direct determination by HPLC, or automated assays based on a derivatization of ascorbic acid forming colored or fluorescent derivatives. While a plasma concentration  $<11.4 \text{ } \mu\text{mol l}^{-1}$  is widely used to characterize deficiency, literature data signifying adequate status vary from  $\sim >17$  to  $28.4 \text{ } \mu\text{mol l}^{-1}$ .

## Interpretation of Analytical Data in Relation to Bioavailability

The nutritional value of foods is not synonymous with their content of nutrients as determined by chemical analysis. Obviously, it is not enough to know how much of a nutrient, e.g., a vitamin, is present in the food, but we need to know how much of the biological potential inherent in the analyzed substance can be realized when the food is eaten. Thus, 'bioavailability' has to be considered, which has been defined as 'the portion of a nutrient capable of being absorbed and available for use or storage'.

Sensitive components of foods, like vitamins, are prone to many possible interactions in the food itself and/or in the body. Such interactions have a major influence on the final bioavailability, some being beneficial while others are harmful.

As a first level, problems related to nutrient availability might be located in storage, preparation, and

processing of foodstuffs. Bioavailability, in the widest sense of the term, is also implicated. A few examples may elucidate this.

Heat is the most commonly used agent in the preparation and processing of foodstuffs in order to achieve preservation, favorable texture and flavor changes, increased palatability, etc. Some vitamins are very sensitive to high temperatures (see **Table 1**) and can be entirely inactivated under different time/temperature conditions. But heating may also have advantageous effects by releasing some vitamins from binding proteins, for example, and thus converting them into absorbable form, or by destroying compounds like thiaminase, ascorbic acid oxidase, or avidin, which decompose or inactivate vitamins by binding.

Losses of vitamins also occur as a result of various operations like washing, peeling, crushing, etc., in household preparations as well as in industrial production, oxygen mostly being the deleterious agent. Interaction of some vitamins with heavy metal ions, especially copper, may also cause essential destruction.

These few examples may illustrate that data obtained by analysis of food raw materials have to be weighed critically with regard to their significance concerning real supply with a certain vitamin.

The next process that has to be taken into account is intestinal digestion and absorption. Some interactions with food and nonfood components can be presumed on this level, too. Dietary fiber and some minerals are candidates for such interactions, the significance of which has to be closely examined.

It has been repeatedly shown that intestinal transport of folic acid and of its derivatives has a pH-optimum of ~6. Folic acid bioavailability is reduced in certain clinical situations where human jejunal pH deviates from this value, such as sodium hydrogencarbonate dosing, prolonged administration of

alkaline drugs, some anti-epileptics, achlorhydria in the elderly, acute pancreatitis, etc. Even when intestinal absorption of a vitamin is normal, further problems of bioavailability might arise when the vitamin is transported in the bloodstream in a bound form. For example, retinol deficiency occurs despite a plentiful supply of retinol when protein deficiency limits the production of retinol-binding protein.

Research on bioavailability of vitamins is a modern field of biochemistry with a strong need for non-biological and biological analytical techniques to promote the understanding of the complicated subject of bioavailability.

*See also:* **Vitamins:** Fat-Soluble; Water-Soluble.

## Further Reading

- Ball GFM (1997) *Bioavailability and Analysis of Vitamins in Foods*. London: Chapman & Hall.
- Basu TK and Dickerson JWT (eds.) (1996) *Vitamins in Human Health and Disease*. Wallingford: CAB International.
- Biesalski HG and Grimm P (2002) *Taschenatlas der Ernährung*. Stuttgart: Georg Thieme Verlag.
- Combs GF (ed.) (1998) *The Vitamins: Fundamental Aspects in Nutrition and Health*. San Diego, CA: Academic Press.
- DeLeenheer AP, Lambert WE, and Van Bocxlaer J (eds.) (2000) *Modern Chromatographic Analysis of Vitamins*. New York: Dekker.
- Eitenmiller RR and Landen WO Jr (1998) *Vitamin Analysis for the Health and Food Sciences*. Boca Raton, FL: CRC Press.
- Rucker RB, Suttie JW, McCormick DB, and Machlin LJ (eds.) (2001) *Handbook of Vitamins*. New York: Dekker.
- Song WO, Beecher GR, and Eitenmiller RR (eds.) (2000) *Modern Analytical Methodologies in Fat- and Water-Soluble Vitamins*. New York: Wiley.

## Fat-Soluble

**C J Bates**, Elsie Widdowson Laboratory, Cambridge, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by A De Leenheer, H Nelis, and W Lambert, pp. 5383–5393, © 1995, Elsevier Ltd.

## Introduction

The importance of the fat-soluble vitamins (A, D, E, and K) in both clinical chemistry and nutrition is well

recognized. Fat-soluble vitamin assays in foods are carried out for a variety of reasons including compliance to specifications, possible degradation by processing of the food, and food composition tables. Also, in clinical chemistry the fat-soluble vitamin levels provide useful information with regard to infective-morbidity, skin defects, and night blindness (vitamin A), rickets, osteomalacia and osteoporosis (vitamin D), radical mediated peroxidative processes (vitamin E), and blood coagulation plus other calcium-dependent processes (vitamin K). The four vitamins mentioned above all have an isoprenoid-related

processing of foodstuffs. Bioavailability, in the widest sense of the term, is also implicated. A few examples may elucidate this.

Heat is the most commonly used agent in the preparation and processing of foodstuffs in order to achieve preservation, favorable texture and flavor changes, increased palatability, etc. Some vitamins are very sensitive to high temperatures (see **Table 1**) and can be entirely inactivated under different time/temperature conditions. But heating may also have advantageous effects by releasing some vitamins from binding proteins, for example, and thus converting them into absorbable form, or by destroying compounds like thiaminase, ascorbic acid oxidase, or avidin, which decompose or inactivate vitamins by binding.

Losses of vitamins also occur as a result of various operations like washing, peeling, crushing, etc., in household preparations as well as in industrial production, oxygen mostly being the deleterious agent. Interaction of some vitamins with heavy metal ions, especially copper, may also cause essential destruction.

These few examples may illustrate that data obtained by analysis of food raw materials have to be weighed critically with regard to their significance concerning real supply with a certain vitamin.

The next process that has to be taken into account is intestinal digestion and absorption. Some interactions with food and nonfood components can be presumed on this level, too. Dietary fiber and some minerals are candidates for such interactions, the significance of which has to be closely examined.

It has been repeatedly shown that intestinal transport of folic acid and of its derivatives has a pH-optimum of ~6. Folic acid bioavailability is reduced in certain clinical situations where human jejunal pH deviates from this value, such as sodium hydrogencarbonate dosing, prolonged administration of

alkaline drugs, some anti-epileptics, achlorhydria in the elderly, acute pancreatitis, etc. Even when intestinal absorption of a vitamin is normal, further problems of bioavailability might arise when the vitamin is transported in the bloodstream in a bound form. For example, retinol deficiency occurs despite a plentiful supply of retinol when protein deficiency limits the production of retinol-binding protein.

Research on bioavailability of vitamins is a modern field of biochemistry with a strong need for non-biological and biological analytical techniques to promote the understanding of the complicated subject of bioavailability.

*See also:* **Vitamins:** Fat-Soluble; Water-Soluble.

## Further Reading

- Ball GFM (1997) *Bioavailability and Analysis of Vitamins in Foods*. London: Chapman & Hall.
- Basu TK and Dickerson JWT (eds.) (1996) *Vitamins in Human Health and Disease*. Wallingford: CAB International.
- Biesalski HG and Grimm P (2002) *Taschenatlas der Ernährung*. Stuttgart: Georg Thieme Verlag.
- Combs GF (ed.) (1998) *The Vitamins: Fundamental Aspects in Nutrition and Health*. San Diego, CA: Academic Press.
- DeLeenheer AP, Lambert WE, and Van Bocxlaer J (eds.) (2000) *Modern Chromatographic Analysis of Vitamins*. New York: Dekker.
- Eitenmiller RR and Landen WO Jr (1998) *Vitamin Analysis for the Health and Food Sciences*. Boca Raton, FL: CRC Press.
- Rucker RB, Suttie JW, McCormick DB, and Machlin LJ (eds.) (2001) *Handbook of Vitamins*. New York: Dekker.
- Song WO, Beecher GR, and Eitenmiller RR (eds.) (2000) *Modern Analytical Methodologies in Fat- and Water-Soluble Vitamins*. New York: Wiley.

## Fat-Soluble

**C J Bates**, Elsie Widdowson Laboratory, Cambridge, UK

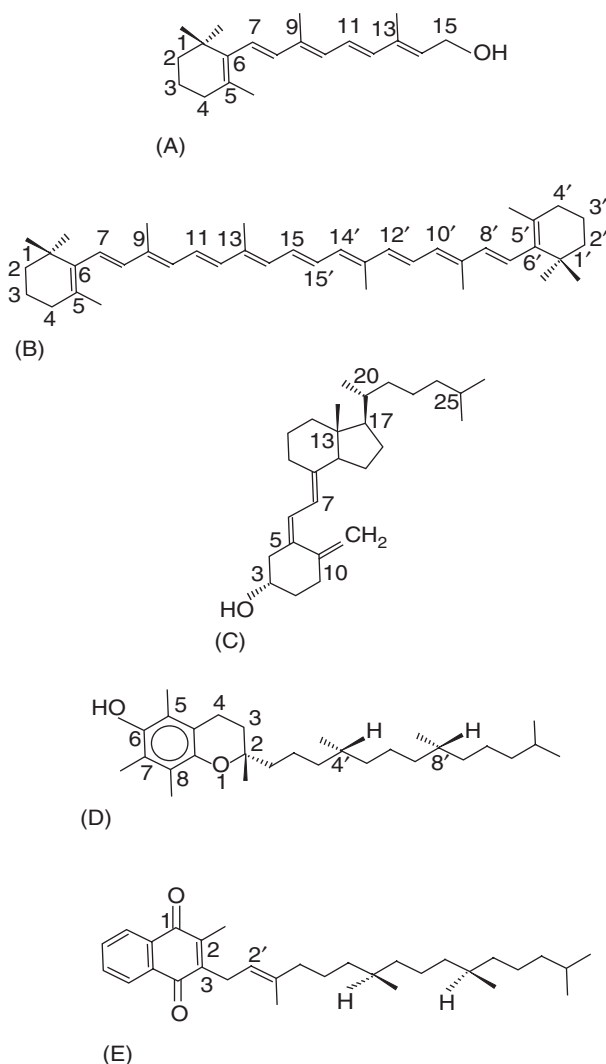
© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by A De Leenheer, H Nelis, and W Lambert, pp. 5383–5393, © 1995, Elsevier Ltd.

## Introduction

The importance of the fat-soluble vitamins (A, D, E, and K) in both clinical chemistry and nutrition is well

recognized. Fat-soluble vitamin assays in foods are carried out for a variety of reasons including compliance to specifications, possible degradation by processing of the food, and food composition tables. Also, in clinical chemistry the fat-soluble vitamin levels provide useful information with regard to infective-morbidity, skin defects, and night blindness (vitamin A), rickets, osteomalacia and osteoporosis (vitamin D), radical mediated peroxidative processes (vitamin E), and blood coagulation plus other calcium-dependent processes (vitamin K). The four vitamins mentioned above all have an isoprenoid-related



**Figure 1** Chemical structure of (A) vitamin A, all-trans-retinol; (B)  $\beta$ -carotene; (C) vitamin D, cholecalciferol; (D) vitamin E,  $\alpha$ -tocopherol; and (E) vitamin K<sub>1(20)</sub>, phyloquinone.

structure (Figure 1) and are liposoluble. This permits similar approaches for extraction. Other similarities include their sensitivity toward isomerization and oxidation by contact with daylight and their degradation by high temperatures. These properties result in a rather limited application of thin-layer chromatography (TLC) and gas chromatography (GC) to the analysis of fat-soluble vitamins in different matrices. On the other hand, in liquid chromatography (LC), the contact with air, light, and high temperatures can be minimized without loss of resolving power, speed, and sensitivity. Consequently, the number of LC applications to the analysis of the fat-soluble vitamins now largely supersedes other techniques (e.g., TLC and GC). A number of reviews dealing with these issues have recently become available.

## Vitamin A

### General Properties

The various forms of vitamin A (retinol, retinyl esters) are present in food of animal origin (liver, milk, meat) while its precursors (certain carotenoids) are present mainly in plant foods.

The retinyl esters from the food are hydrolyzed in the intestine and the resulting retinol is re-esterified in the mucosal cells, mainly with palmitic acid and associated with lymph chylomicrons. After being taken up by the liver parenchymal cells, retinyl esters are hydrolyzed again. Part of the liberated retinol is then transported in the plasma to peripheral tissues by the retinol binding protein–prealbumin complex. Another part is stored mainly as retinyl palmitate in the liver fat storing cells. Vitamin A is metabolized by oxidation of the cyclohexene ring, or by reversible oxidation to retinal and irreversibly to retinoic acid. Glucuronidation and oxidative cleavage of the side chain are further steps in the metabolic process.

Vitamin A is susceptible to oxidation and isomerization and, consequently, all manipulations such as sample collection, extraction, and storage should be carried out at low temperature and in subdued light.

Within cells, vitamin A (retinol, retinal, retinoic acid) functions mainly in vision, cellular differentiation, and embryogenesis. The adverse effects of vitamin A deficiency on complex physiological processes such as reproduction and the immune response result primarily from defective cellular differentiation.

### Methods of Extraction and Cleanup

When total vitamin A is to be measured regardless of whether it is free or esterified, a saponification step can be included. However, degradation of the retinoids occurs even under mild conditions and in the presence of antioxidants. Consequently, direct extraction procedures without any digestion step, allowing the separation and separate quantification of the ester forms and of retinol itself, have gained interest in the last two decades.

After a protein denaturation step with methanol, ethanol, or acetonitrile (in the case of blood samples) or after homogenization or lyophilization (in the case of food samples), extractions are usually performed with *n*-hexane, dichloromethane, or cyclohexane. Supercritical fluid extraction (SFE) has recently opened new perspectives because extraction conditions are milder and can be carried out at lower temperature. In addition, during the extraction the sample is in an inert atmosphere and protected from light, thus reducing further degradation of the vitamin A.



For samples with a high lipid content (e.g., milk), additional cleanup of the extracts may be necessary to eliminate other lipids such as triacylglycerols. In the past, open-column chromatography on magnesia, alumina, or silica gel has been used. More recently, solid-phase extraction (SPE) on disposable prepacked cartridges filled with silica or C<sub>18</sub> reversed-phase packing material has been applied to vitamin A analysis. SPE is a refinement of the earlier open-column chromatographic systems.

### Chemical Methods

Initial bioassay techniques were too time consuming to be employed on a routine basis and were superseded by physicochemical methods based on colorimetry, e.g., after reaction with antimony trichloride (Carr–Price reaction), or trifluoroacetic acid, or based on ultraviolet (UV) spectrophotometry or fluorimetry. However, these procedures cannot readily discriminate molecular species with different vitamin A activities (e.g., *cis-trans* isomers, retinyl esters, or oxidative degradation products).

### Chromatographic Methods

**Thin-layer chromatography** TLC is now rarely used for the analysis of vitamin A. High-performance thin-layer chromatography (HPTLC) together with densitometric scanning has been explored as a possible alternative, but is not widely used.

**Gas chromatography** The instability of conjugated unsaturated compounds such as retinol and retinyl esters, together with their degradation on incompletely inactivated column packing materials, has drastically hindered attempts of retinoid analysis by GC. However, when chromatographed on bonded phase capillary columns only a very low degree of decomposition is observed for retinol. Similarly, short chain retinyl esters have also been successfully chromatographed. By lowering the extent of decomposition of retinol on these capillary GC columns new perspectives are opened also toward gas chromatographic–mass spectrometric (GC–MS) determination of this compound, a technique that has recently been developed for studies of stable isotope-labeled vitamin A and carotenoids and the conversion of certain carotenoids to vitamin A in biological systems.

**Liquid chromatography** If research is focused on the separation of the different isomers of vitamin A, LC adsorption chromatography is the method of choice. Silica packing materials, and to a lesser

extent alumina, are then eluted with a nonpolar eluent based on *n*-hexane and a small percentage of a more polar solvent.

Normal bonded phase packing materials have been used as an alternative to adsorption chromatography because column deactivation due to contamination of polar contaminants injected is less frequently seen in the normal bonded phase than in adsorption chromatographic systems.

Reversed-phase packing materials (C<sub>18</sub> and C<sub>8</sub>) are utilized especially for the separation of the different retinyl esters and for combined vitamin A and carotenoid mixtures such as those in food extracts and clinical samples. Most semiaqueous mobile phases, consisting of mixtures of methanol or acetonitrile and water, result in chromatographic runs of more than 1 h to elute retinyl stearate. Gradient elution shortens this elution time but results in a loss of separation of two predominant esters, retinyl palmitate and retinyl stearate. More recently, isocratic non-aqueous reversed-phase (NARP) conditions with an acetonitrile–dichloromethane mixture resulted in a successful separation of the different esters within 15 min. The big advantage of NARP conditions over semiaqueous systems is the increased solubility of lipids. To compensate for this increased affinity of the compounds of low polarity for the eluent, highly retentive stationary phases are recommended.

A notable application of a polymeric C<sub>18</sub> phase (Vydac 201 TP) in the vitamin analysis was the separation of mono- and di-*cis*-retinols using a water–methanol–*n*-butanol mixture as eluent.

UV absorbance detection has been most widely used for vitamin A analysis. However, because retinol and retinyl esters are highly fluorescent, detection limits of one order of magnitude better than in assays with UV detection can be obtained using fluorescence detection. Also, electrochemical detection is a valuable alternative to UV and fluorescence detection provided the eluent contains water to incorporate essential electrolytes. Another detector for LC is the mass spectrometer. The LC–MS approach has also been applied to the analysis of vitamin A and its metabolites.

**Competitive protein-binding and immunological methods** As far as we know, no competitive protein-binding procedures or immunological methods are in common use for the determination of retinol.

### Examples and Limitations of Current Methodology

The limitations of techniques such as TLC and GC in vitamin A analysis have already been highlighted and

**Table 1** Representative procedures for the determination of vitamin A

Analyte(s)	Matrix	Sample preparation/extraction	Column	Mobile phase (v/v)	Detection
Gas chromatography ROL <sup>a</sup>	Plasma	Ethanol denat./SE: <i>n</i> -hexane	WCOT methyl-silicone, 0.25 $\mu$ m	He	FID
Retinyl esters	Plasma	Ethanol denat./SE: <i>n</i> -hexane	WCOT methyl-silicone, 0.1 $\mu$ m	He	FID
Liquid chromatography 13- <i>cis</i> ; all- <i>trans</i> -ROL	Milk	Cold saponification/SE: hexane– diethyl ether (85:15, v/v)	Apex silica 3 $\mu$ m	1–5% 2-propanol in heptane	ABS
ROL	Serum	ACN denat./SE: <i>n</i> -hexane	Zorbax ODS 5 $\mu$ m	CAN–CH <sub>2</sub> Cl <sub>2</sub> –MeOH (70:15:15)	ABS
ROL	Serum	Ethanol denat./SE: <i>n</i> -hexane	Polygosil 60 5 $\mu$ m	Hexane–CH <sub>2</sub> Cl <sub>2</sub> –2-propanol (90:10:1)	FL
Retinyl esters	Liver	SE: CH <sub>2</sub> Cl <sub>2</sub>	Resolve C <sub>18</sub> 5 $\mu$ m	CAN–CH <sub>2</sub> Cl <sub>2</sub> (80:20)	ABS

<sup>a</sup> Abbreviations used: ROL, retinol; ACN, acetonitrile; SE, solvent extraction; ABS, absorbance detection (nm); FL, fluorescence detection ( $\lambda_{\text{ex}}$ ,  $\lambda_{\text{em}}$ , nm); FID, flame ionization detection; SIM, selected ion monitoring (mass spectrometry); denat., denatured.

that LC is the technique for analysis in that area. Some representative procedures are brought together in Table 1.

## Carotenoids

### General Properties

Carotenoids are yellow to orange-red polyene pigments that strongly absorb ( $\epsilon > 100\,000\text{ mol l}^{-1}\text{ cm}^{-1}$ ) light in the 400–500 nm region of the spectrum. Of the  $\sim 600$  known representatives of this class, roughly 50, including  $\beta$ -carotene,  $\alpha$ -carotene, and  $\beta$ -cryptoxanthin possess vitamin A activity (Figure 1), i.e., can be intestinally cleaved to yield retinal. These compounds are sensitive to light, air, heat, and acid, which catalyze chain cleavage and/or stereoisomerization. The latter process involves a reversible conversion of the all-*trans* to the less stable *cis*-configuration and results in characteristic changes in the absorption spectrum and a decrease in vitamin A activity.

### Methods of Extraction and Cleanup

Carotenoids are commonly extracted from liquid samples (plasma/serum) into lipophilic solvents such as hexane, hexane–ethyl acetate, or diethyl ether, mostly after deproteinization with ethanol or methanol, which also helps to liberate the lipidic substances from protein binding. Extracts should be protected from light and acids and antioxidants may usefully be added. The extract is either used as such or is concentrated under oxygen-free nitrogen. Solid samples, e.g., foods, are either extracted with a solvent miscible with water (acetone, methanol) or, after dehydration of the sample, with a water immiscible solvent. Cleanup of the extract and fractionation of the pigments may involve saponification and/or open-column chromatography.

### Chemical Methods

Without prior fractionation the total concentration of carotenoids ('total carotenes') in a sample extract may be estimated by spectrophotometry at 450 nm (using an average molar absorptivity for calculation). As an improvement, in earlier times, individual carotenoids or carotenoid classes were collected after separation on alumina or magnesia-hyflo supercel columns and the absorbance of the eluted fractions measured. This approach once formed the basis of the Association of Official Analytical Chemists (AOAC) method for foods.



## Chromatographic Methods

**Thin-layer chromatography** The major value of TLC in the carotenoid area lay in prefractionation and identification rather than quantification. Detection was on the basis of color and can be achieved by densitometric scanning at 450 nm. Commonly used sorbents included silica gel and particularly magnesia, which exhibits high selectivity toward geometrical isomers of carotenoids. Octadecyl bonded silica (RP-18) had the advantage of being chemically less aggressive than the polar sorbents.

**Gas chromatography** Only perhydrogenated carotenoids have been found to be sufficiently volatile and thermostable to be amenable to GC. This approach is only valuable in connection with MS for identification purposes.

**Liquid chromatography** Normal-phase systems have the advantage of being directly compatible with extracts in hexane. Silica, alumina, and lime (calcium hydroxide) are all particularly suited to the resolution of carotenoid geometrical isomers (*cis-trans*) and diastereoisomers, but not positional isomers ( $\alpha/\beta$ -carotene). However, silica may cause on-column artifacts, reproducible retention on alumina is strongly dependent on a rigorous control of the water content of the eluent, and lime columns are not commercially available.

By far the majority of methods for the determination of carotenoids in biological materials nowadays are based on reversed phase chromatography. NARP eluents based on acetonitrile-methanol mixtures containing dichloromethane, chloroform, or tetrahydrofuran as a strong modifier are recommended for optimal sample solubility. The overall resolution reportedly benefits from the addition of hexane as a fourth component. Certain monomeric non-end-capped heavily loaded C<sub>18</sub> columns, e.g., Zorbax ODS, permit an isocratic separation of a wide range of carotenoids. Other systems use gradient elution to bridge the large polarity differences between xanthophylls (mainly hydroxyl- and epoxy-carotenes) and carotene hydrocarbons.

The resolution of geometrical isomers is readily accomplished on reversed-phase sorbents based on Vydac-type polymeric silica, unlike on monomeric phases on which all *cis* isomers usually co-elute. The exhaustive profiling of carotenoids in complex samples often requires the complementary use of a reverse phase (C<sub>18</sub>) plus a polar (cyano) sorbent. With modern equipment it is possible to achieve accurate and reproducible separation and quantitation of the major classes of carotenoids, retinoids, and

tocopherols by a 10 min LC run for extracts of plasma or other biological samples. However, the more demanding separation of very similar xanthophylls (e.g., lutein and zeaxanthin) and of lycopene derivatives may require gradient elution and hence longer run times of ~45 min.

SFC is emerging as an interesting tool for the separation of carotenoids, including positional and geometrical isomers, as well as for their identification, i.e., by coupling with MS.

Absorption detection at 450–470 nm is standard in the LC of carotenoids. Photodiode-array detection capable of recording the characteristic absorption spectra of carotenoid peaks forms the backbone of techniques for their identification in biological extracts, often complemented by LC-MS. Various analogs that have been used as internal standards for  $\beta$ -carotene have included apo- $\beta$ -carotenal, nonapreno- $\beta$ -carotene (C<sub>45</sub>), and  $\beta$ -apo-8'-carotenal-ethylxime.

## Competitive Protein Binding and Immunological Methods

Competitive protein binding and immunological methods for carotenoids are not generally available.

## Examples and Limitations of Current Methodology

The only technique that is of major interest for carotenoid analysis is LC. Selected applications to biological samples are presented in Table 2. These methods are mostly satisfactory in terms of selectivity and sensitivity. However, accurate quantification requires precautions because of relative lability of the analytes during sample preparation, their incomplete recovery from LC columns, and the limitations of the available internal standards.

## Vitamin D

### General Properties

The principal biological function of vitamin D is to act as precursor of 1,25-dihydroxyvitamin D, a circulating hormone that regulates calcium transport at several sites in the body, including calcium absorption in the gastrointestinal tract. This hormone, derived by two hydroxylations of the vitamin, firstly in the liver and then in the kidney, performs a wide range of regulatory functions that are implicated in human diseases. The best understood of these are the bone diseases that are classically associated with vitamin D deficiency, namely rickets and osteomalacia, which arise through failure of bone mineralization.

The term vitamin D covers the two natural compounds ergocalciferol (vitamin D<sub>2</sub>) and cholecalciferol

**Table 2** Representative LC procedures for the determination of carotenoids

Analyte(s)	Matrix	Sample preparation/purification	Column	Mobile phase (v/v)	Detection
<i>cis</i> + <i>trans</i> - $\beta$ -carotene, $\alpha$ -carotene, lutein, lycopene	Foods	SE <sup>a</sup> : methanol–chloroform	Vydac 201 TP 5 $\mu$ m	Methanol–chloroform (96:4)	ABS 450
$\beta$ -Carotene, $\alpha$ -carotene, lycopene, $\beta$ -cryptoxanthin (+ retinol, $\alpha$ -tocopherol)	Plasma	SDS pretreatment SE: ethanol + heptane	Spherisorb ODS-2 3 $\mu$ m	Acetonitrile–methanol–chloroform (47:47:6)	ABS 450
$\beta$ -Carotene, $\alpha$ -carotene (+ retinol, $\alpha$ -tocopherol)	Serum	SE: ethanol + hexane	Ultrasphere ODS 5 $\mu$ m	Acetonitrile–methanol– dichloromethane (70:10:20)	ABS 450
18 Carotenoids	Plasma	SE: ethanol + diethyl ether	A: Microsorb C <sub>18</sub> 5 $\mu$ m  B: Nitrile phase 5 $\mu$ m	A: acetonitrile– dichloromethane–hexane–methanol (45:22.5:22.5:10)  B: hexane–dichloromethane– methanol–diisopropylethylamine (74.65:25:0.25:0.10)	ABS 445, 475
$\beta$ -Carotene, $\alpha$ -carotene, lycopene, $\beta$ -cryptoxanthin, lutein	Tissues	Enzymatic digestion, mechanical homogenization; SE: ethanol + acetonitrile + hexane	Ultracarb C <sub>18</sub> 5 $\mu$ m	Acetonitrile–THF–methanol–1% ammonium acetate (65:25:6:4)	ABS 450

<sup>a</sup> Abbreviations used: see also **Table 1**; THF, tetrahydrofuran.

(vitamin D<sub>3</sub>) as well as a number of hydroxylated metabolites. Among the latter the 25-hydroxy- and 1,25-dihydroxyvitamin D are most prominent (**Figure 1**). The D<sub>2</sub> differs from the D<sub>3</sub> form by the presence of a double bond in its isoprenoid side chain. Both vitamins D<sub>2</sub> and D<sub>3</sub> are derived by photoirradiation from their respective diene sterol precursors (provitamin D). Of the 64 theoretically possible steric configurations the triene isomers tachysterol, isotachysterol, and 5,6-*trans*-caldiol are of analytical interest. The triene moiety confers a characteristic UV absorption ( $\lambda_{\text{max}} = 264 \text{ nm}$ ,  $\epsilon = 18\,300 \text{ mol l}^{-1} \text{ cm}^{-1}$ ) upon the molecule. Vitamin D is moderately sensitive to heat and air, which catalyze isomerization to previtamin D and oxidation, respectively.

### Methods for Extraction and Cleanup

Saponification using ethanolic KOH is sometimes useful to eliminate the bulk of neutral fat from lipid-rich samples such as milk, but should preferably be avoided for stability reasons. A sample preparation scheme for vitamin D analysis usually involves a double phase solvent extraction followed by a column chromatographic cleanup step. Depending on the polarity of the solvent system used, this lipid extraction can include either total (vitamin D + metabolites) or be selective (one target metabolite or a group of metabolites). Examples of solvents affording the extraction of ‘total vitamin D’ include chloroform–methanol (Bligh and Dyer) or dichloromethane–methanol. The parent vitamin D partitions into low polarity solvents, e.g., hexane or cyclohexane. Mixtures of these with ethyl acetate or an alcohol are employed for the isolation of metabolites and have the additional advantage of disrupting the vitamin D binding protein complex. Alternative extraction solvents are based on diethyl ether, dichloromethane, benzene/toluene. Column chromatography serves a double purpose, i.e., removal of interfering lipids and prefractionation into classes of vitamin D metabolites. Minicolumns or prepacked cartridges containing Sephadex LH-20, hydroxyalkoxypropylsephadex, alumina, silica, or bonded phases (amino, octadecylsilane) can be used.

### Chemical Methods

Direct UV–visible spectrophotometric determination of vitamin D is limited to pure dosage forms that should not even contain other fat-soluble vitamins. In an older colorimetric AOAC procedure the absorbance of a colored product resulting from the reaction of vitamin D with antimony trichloride was measured at 500 nm. However, the other fat-soluble vitamins, particularly vitamin A, as well as the

vitamin D triene isomers interfered and, hence, had to be removed. Alternative colorimetric procedures employed trifluoroacetic acid, trichloroacetic acid, or  $\alpha$ -dichlorohydrin as dehydrating agents.

### Chromatographic Methods

**Thin-layer chromatography** TLC, particularly in its HPTLC version, can be considered as a (less expensive) alternative to LC for the purification and prefractionation of lipid extracts prior to gas chromatographic or radioligand assays of vitamin D metabolites in biological materials. In foods, vitamin D has been determined by HPTLC-photodensitometry, either at 266 nm or, after spraying with chromogenic reagents, at higher wavelengths. As a technique to separate vitamin D analogs, including irradiation products of provitamin D and the different fat-soluble vitamins, TLC has been replaced by LC with SPE cleanup columns that have replaced TLC for the cleanup process.

**Gas chromatography** GC of vitamin D and its metabolites is subject to a number of unfavorable phenomena, including on-column adsorption, dehydration of 25-hydroxy derivatives, and thermal rearrangement involving B-ring closure, yielding pyro and isopyro isomers. The latter is prevented by derivatization to isotachysterols, using acyl chlorides, acyl anhydrides, or HCl in a chlorinated solvent. With biological samples, flame ionization and electron capture detection lack sufficient selectivity to avoid the interference of other lipids, e.g., cholesterol, so that extensive sample pretreatment is required. Detection of trimethylsilyl- or *t*-butyldimethylsilyl derivatives by electron impact or chemical ionization mass spectrometry results in superior selectivity and sensitivity, thus permitting the determination of vitamin D, 25-hydroxyvitamin D, and several dihydroxylated metabolites in plasma or foods. Deuterium-labeled analogs or dihydrotachysterol are candidate internal standards.

**Liquid chromatography** On silica or polar bonded phases (nitro, cyano), the separation of vitamin D metabolites occurs according to the number and position of hydroxyl groups in the molecule. Binary mobile phases are usually based on hexane–2-propanol mixtures. Improved resolution of normally co-eluting D<sub>2</sub> and D<sub>3</sub> hydroxylated metabolites is afforded by ternary solvent systems containing dichloromethane as a third component. However, the D<sub>2</sub> and D<sub>3</sub> parent compounds remain unresolved in any normal-phase system.

Reversed-phase LC with water–methanol or water–acetonitrile as mobile phases is suitable for the resolution of vitamin D<sub>2</sub> and D<sub>3</sub> and the 25-hydroxy-metabolites, but exhibits inferior selectivity for the dihydroxylated metabolites. When it is to be applied to biosamples a normal-phase LC prefractionation is mostly required for selectivity reasons.

Vitamin D and its metabolites can be detected on the basis of their native UV absorption ( $\lambda_{\text{max}} = 264 \text{ nm}$ ) or after conversion to isotachysterols ( $\lambda_{\text{max}} = 301 \text{ nm}$ ). However, because of the inherent poor selectivity and sensitivity of this approach quantitation is mostly performed offline by a radioligand assay. Several metabolites have also been determined in plasma by LC–thermospray MS.

Vitamin D<sub>2</sub> is a straightforward choice as an internal standard for vitamin D<sub>3</sub>, whereas 5,6-*trans*-caldiol has been used in an LC assay of 25-hydroxyvitamin D<sub>2</sub> and 25-hydroxyvitamin D<sub>3</sub>.

### Competitive Protein Binding and Immunological Methods

A large number of radioligand assays have been reported for the quantitative determination of vitamin D metabolites in biosamples. In radioreceptor assays the target protein is either the vitamin D binding protein from blood or a specific tissue receptor. Antibodies used in radioimmunoassays originate from sheep or rabbits. The traditional one- or two-dimensional LC cleanup prior to radioligand assays of vitamin D metabolites is increasingly replaced by a solid-phase purification on cartridges.

Assays for 25-hydroxyvitamin D (caldiol) and 1,25-dihydroxyvitamin D (calcitriol) in human and other animals' plasma samples are commonly performed nowadays by kit assays that are based on competitive protein-binding or immunoassays, either with or without a preliminary cleanup on SPE cartridges. Although there is some ongoing controversy about details of ultimate accuracy and of possible interferences, these kit assays have proved in practice to be clinically useful, highly sensitive, and are much simpler to use than the alternative LC approach. External quality assurance schemes such as 'DEQAS', operated from Charing Cross Hospital, are invaluable adjuncts to laboratory performance which greatly help to improve interlaboratory harmonization and to detect and resolve any methodological problems at an early stage.

### Examples and Limitations of Current Methodology

A number of representative methods for the determination of vitamin D and its metabolites in biological samples are presented in Table 3. The major

**Table 3** Representative procedures for the determination of vitamin D and its metabolites

<i>Analyte(s)</i>	<i>Matrix</i>	<i>Sample preparation/purification</i>	<i>Column</i>	<i>Mobile phase (v/v)</i>	<i>Detection</i>
<i>Gas chromatography</i>					
D <sub>2</sub> , D <sub>3</sub> 25-OHD <sub>2</sub> 25-OHD <sub>3</sub>	Plasma	Deproteinization: acetonitrile SPE <sup>a</sup> : C <sub>18</sub> -silica, silica RP-LC (only for D <sub>2</sub> /D <sub>3</sub> ) NP-LC	2% OV-1	He	MS 445
1,25(OH) <sub>2</sub> D <sub>3</sub>	Plasma	SE: chloroform–methanol CC: Sephadex LH-20 NP-LC RP-LC	1.5% SE-30	He	MS 445
<i>Liquid chromatography</i>					
D <sub>3</sub>	Fortified milk	SE: ethanol + NH <sub>4</sub> OH + diethyl ether–hexane	Zorbax Sil 10 µm	Hexane–2-propanol (99:1)	ABS 265
25-OHD <sub>2</sub> 25-OHD <sub>3</sub>	Serum	Deproteinization: acetonitrile SPE: C <sub>18</sub> -silica	Zorbax Sil 10 µm	Hexane–2-propanol (98:2)	ABS 254
D <sub>2</sub> , D <sub>3</sub>	Serum	Deproteinization: methanol SE: hexane SPE: silica	Ultrasil ODS 10 µm	Methanol	ABS 264
<i>Radioligand assays</i>					
25-OHD <sub>3</sub>	Plasma	Deproteinization: acetonitrile	Nucleosil-10-NO <sub>2</sub> 10 µm	Gradient-hexane → hexane– 2-propanol–water (28:12:0.42)	CPB
1,25-(OH) <sub>2</sub> D <sub>3</sub> 1,25-(OH) <sub>2</sub> D <sub>3</sub> 25-OHD <sub>3</sub>	Plasma Plasma	SPE: C <sub>18</sub> -silica SE: benzene Deproteinization: methanol–HCl	RSIL-silica 5 µm –	Hexane–2-propanol (9:1) –	CPB or RIA CPB
24,25-(OH) <sub>2</sub> D <sub>3</sub> 1,25-(OH) <sub>2</sub> D <sub>3</sub> 1,25-(OH) <sub>2</sub> D <sub>3</sub>	Serum	SPE: C <sub>18</sub> -silica SPE: NH <sub>2</sub> -silica Saponification SE: dichloromethane SPE: silica	–	–	CPB

<sup>a</sup> Abbreviations used: see also **Tables 1** and **2**; SPE, solid-phase extraction; MS, mass spectrometry (*m/z* value); CPB, competitive protein finding; RIA, radioimmunoassay.

limitation of the LC procedures lies in the lack of highly sensitive detection options for these analytes, thus necessitating the use of radioligand assays for quantification. Although intrinsically selective and sensitive, the latter still require extensive sample purification to remove the bulk of lipids and to differentiate individual metabolites. The obvious alternative is GC-MS, but even with this sophisticated technique the determination of metabolites that occur in the subnanogram range (e.g., 1,25-dihydroxyvitamin D) remains an analytical challenge.

## Vitamin E

### General Properties

There is a considerable body of research that indicates that vitamin E has important antioxidant, or redox-modulatory, functions in cell membranes and other lipidic sites in the body, which include especially the prevention of pro-oxidant (including oxygen free radical) damage by peroxidation of the polyunsaturated lipids. Recent studies suggest that vitamin E also has cell-signaling functions. The different chemical forms of vitamin E have different metabolic fates, and may also have different functions from each other. However, the roles and significance of vitamin E in chronic human diseases have yet to be fully elucidated.

Vitamin E is a collective term for tocopherols and tocotrienols, a series of 6-chromanol derivatives substituted by a saturated (as in tocopherols) or a partially unsaturated (tocotrienols) isoprenoid side chain and one to three methyl groups (**Figure 1**). The principal form in the body is  $\alpha$ -tocopherol (5,7,8-trimethyltocol), which in nature occurs as the 2*R*,4'*R*,8'*R*-diastereomer. Totally synthetic  $\alpha$ -tocopherol (all-rac- $\alpha$ -tocopherol) is a racemate of all possible stereoisomers, whereas a semisynthetic racemic  $\alpha$ -tocopherol (2-ambo- $\alpha$ -tocopherol) is a mixture of the 2*R*,4'*R*,8'*R* and the 2*S*,4'*R*,8'*R* (2-epi) diastereoisomers. The natural dimethyltocols  $\beta$ - and  $\gamma$ -tocopherol are structural variants.  $\gamma$ -Tocopherol is a major component of certain foods. All vitamin E derivatives possess strong reducing properties. They are relatively stable to heat and alkali in anoxic conditions but their oxidation is promoted by alkali and metal ions in the presence of air. Tocopherols and tocotrienols absorb light in the UV region ( $\lambda_{\max} = 292\text{--}295\text{ nm}$ ,  $\epsilon = 3530\text{ mol l}^{-1}\text{ cm}^{-1}$ ) and are natively fluorescent ( $\lambda_{\text{exc}} = 205\text{ and }295\text{ nm}$ ,  $\lambda_{\text{em}} = 330\text{ nm}$ ) and can be electrochemically oxidized. The latter two properties are lost in  $\alpha$ -tocopheryl acetate where the phenolic hydroxyl group is blocked.

### Methods for Extraction and Cleanup

Standard lipid extraction methods may be used for the isolation of vitamin E from a biological matrix. The addition of oxidants (pyrogallol, ascorbic acid, butylated hydroxytoluene) is necessary when saponification is included as purification before extraction. This is useful in connection with the analysis of erythrocytes, tissues, foods, and vegetable oils, but generally does not require plasma. However, saponification may negatively affect the recovery of tocopherols and particularly tocol, a potential internal standard, from lipid-rich samples. Another pretreatment for plasma, erythrocytes, and tissues that reportedly results in cleaner extracts is with sodium dodecyl sulfate. Methods for solvent extraction of tocopherols fall into two categories, i.e., double phase extraction, using hexane, hexane-ethyl acetate, diethyl ether, butanol-ethyl acetate, or chloroform-methanol, and monophasic extractions with water-miscible solvents. Depending on the selectivity of the analytical technique, it may be necessary to remove lipid interferences, particularly cholesterol, from the extract. This can be done by digitonin precipitation, open-column chromatography on silica, florisil, or digitonin-impregnated celite, and by TLC.

### Chemical Methods

In pure samples such as pharmaceutical products tocopherols can sometimes be directly determined by UV spectrophotometry. Improved selectivity is derived from the Emmerie-Engel reaction, based on the reduction by tocopherols of iron(III) ions to iron(II) ions. The latter form a red complex with 2,2'-bipyridyl or bathophenanthroline, which can be measured colorimetrically ( $\lambda_{\max} = 520\text{ nm}$ ). However, as many reducing substances, including carotenoids and sterols, interfere with the reaction, their removal by column chromatography or TLC is imperative. Such prefractionation should also be carried out in the spectrofluorimetric assays of vitamin E, although these are more specific than the colorimetric ones.

### Chromatographic Methods

**Thin-layer chromatography** Historically TLC has had great merits in vitamin E assays as a cleanup step in connection with colorimetry or GC and, for a long time, as the only technique capable of separating  $\beta$ - and  $\gamma$ -tocopherols. This separation was accomplished either by using complex mobile phases in one-dimensional approach or, alternatively, in two-dimensional systems. For quantitative purposes photodensitometric scanning after HPTLC separation has been conducted both in the transmittance and the



reflectance modes, on native tocopherols, after their oxidation to tocopherylquinones or after spraying with chromogenic reagents. The latter sometimes differentiate on a color basis between particular tocopherols.

**Gas chromatography** GC on capillary columns, unlike on packed columns, is capable of separating  $\beta$ - and  $\gamma$ -tocopherols and suffers less from the interference of cholesterol in biological extracts. After prefractionation by LC, GC is capable of resolving all eight  $\alpha$ -tocopherol diastereomers. Although all tocopherols and tocotrienols have been separated in a single run, without derivatization on a WCOT column, a conversion of the phenolic hydroxyl group to TMS ethers, acetate, or butyrate esters is often advantageous for reasons of peak symmetry. The most powerful potential of GC for vitamin E analysis lies in its easy combination with MS. In the selected ion monitoring (SIM) mode using a deuterated internal standard GC-MS has permitted the sensitive quantification of  $\alpha$ -tocopherol in minute amounts of lung and ocular tissues. For the routine determination of vitamin E in biological materials GC with flame ionization detection has been largely superseded by LC.

**Liquid chromatography** The main analytical application of normal-phase chromatography is the resolution of positional isomers of tocopherols and tocotrienols. This is readily accomplished on silica and certain polar bonded phases (amino or aminocyno but not cyano) using hexane-diisopropyl ether or hexane-2-propanol mixtures as eluents.

An advantage of normal-phase systems is their compatibility with extracts in nonpolar solvents. Thus, normal-phase chromatography lends itself very well to the determination of individual tocopherols and tocotrienols in samples of vegetable oils, after a simple dissolution in hexane. As in most other relevant biological samples (plasma/serum, erythrocytes, platelets, feces, tissues, and foods) the content of  $\beta$ -tocopherol is negligible, routine methods for the quantification of the  $\alpha$ -,  $\gamma$ -, and  $\delta$ -homologs are based on reversed-phase chromatography. The isocratic separation of  $\alpha$ -,  $\gamma$ -, and  $\delta$ -tocopherols or tocotrienols is straightforward on any reversed-phase column, but the positional isomers always co-elute. Tocotrienols elute ahead of the corresponding tocopherols. Mobile phases usually consist of methanol with a small percentage of water or sometimes nonaqueous mixtures with strong modifiers. The addition of electrolytes, e.g., perchlorate or acetate, to semiaqueous eluents permits the electrochemical detection of vitamin E. Biological extracts in nonpolar solvents

have to be evaporated to dryness and reconstituted with the mobile phase prior to injection on a reversed-phase column. In contrast, alcoholic extracts resulting from a monophasic extraction can be directly injected without concentration, on condition that sufficient sensitivity and selectivity are provided by the detector (see below).

The highest sensitivity and selectivity in vitamin E LC assays are obtained by using fluorescence or electrochemical detection. In the former, excitation at the low wavelength (205 nm) leads to improved detection limits but at the expense of selectivity, compared with the use of 295 nm. Electrochemical detection in the oxidation mode (amperometry or coulometry) is another factor 20 times more sensitive. In routine practice, however, most vitamin E assays employ the less sensitive absorbance detection at 292–295 nm (variable wavelength instrument) or 280 nm (fixed wavelength detectors). If retinol and carotenoids are included, a programmable multi-channel detector, preferably a diode array instrument, is needed. As noted previously, combined LC assays for vitamins A, E, and carotenoids are now in common use for clinical chemistry and can measure about a dozen components within a  $\sim 10$  min run. The NIST and UK EQAS external quality assurance schemes permit interlaboratory comparisons of performance for these assays.

The choice of a suitable internal standard in quantitative work is practically limited to tocol,  $\alpha$ -tocopheryl acetate (only with absorbance detection), or  $\delta$ -tocopherol (if absent in the sample) in reversed-phase systems and tocol or 2,2,5,7,8-pentamethyl-6-chromanol in normal-phase chromatography.

### Competitive Protein Binding and Immunological Methods

No competitive protein binding or immunological methods have been reported for vitamin E.

### Examples and Limitations of Current Methodology

Table 4 lists a number of recent LC and GC-MS methods for the determination of vitamin E in bio-samples.

Because of the versatility in detection and the relatively high concentration of vitamin E in foods and physiological samples, the available LC methods show few shortcomings in terms of selectivity, sensitivity, and routine applicability. In those cases where extra sensitivity is required GC-MS can be considered. However, there is still room for improved methods to separate the diastereomers.



**Table 4** Representative procedures for the determination of vitamin E

Analyte(s)	Matrix	Sample preparation/purification	Column	Mobile phase (v/v)	Detection <sup>a</sup>
Gas chromatography $\alpha$ -T	Lung tissue	SE <sup>b</sup> : BSTFA-TMCS-pyridine (silylating reagent)	3% Silar 10C	He	MS (502)
Liquid chromatography $\alpha$ -T (+ retinol)	Serum	SE: ethanol + hexane	RSIL C <sub>18</sub> , 10 $\mu$ m	Methanol	ABS (292)
$\alpha$ -T, $\gamma$ -T	Plasma	SE: methanol	$\mu$ -Bondapak C <sub>18</sub>	Methanol-water (98:2)	FL (292, 335)
$\alpha$ -T, $\beta$ -T, $\gamma$ -T, $\delta$ -T	platelets	Saponification	A: RCM-100	A: methanol	FL (295, 330)
	Foods	SE: diethyl ether-petroleum ether	RAD-PAK-C <sub>18</sub> , 5 $\mu$ m B: RCM-100 RAD-PAK	B: hexane-2-propanol (9:1)	
$\alpha$ -T, $\beta$ -T, $\gamma$ -T, $\delta$ -T, $\alpha$ -T <sub>3</sub> , $\beta$ -T <sub>3</sub> , $\gamma$ -T <sub>3</sub> , $\delta$ -T <sub>3</sub>	Vegetable oils	Dissolution in hexane	Silica, 5 $\mu$ m Partisil 5/25 5 $\mu$ m (aminocyanano)	Hexane-THF (96:4)	FL (210, 325)
$\alpha$ -T, $\beta$ + $\gamma$ -T, $\delta$ -T	Feeds	Saponification SE: hexane	Yanapak ODS-T	Methanol + 0.05 mol l <sup>-1</sup> NaClO <sub>4</sub>	ECD (+ 0.8 V)

<sup>a</sup> Wavelength (nm), m/z, etc.<sup>b</sup> Abbreviations used: see also Tables 1-3; ECD, electrochemical detection.

## Vitamin K

### General Properties

Biological samples may contain any or all of three forms of vitamin K: vitamin K<sub>1(20)</sub> (phyloquinone) which is synthesized by green plants and found in chloroplasts of photosynthetic plants and is the major form of vitamin K in plant and most animal foods; vitamin K<sub>2</sub>, a group of vitamers (menaquinones) synthesized by bacteria and ranging from MK-4 to MK-13 according to the number of isoprene units in the side chain, and finally synthetic vitamin K<sub>3</sub> (menadione), which exhibits vitamin K activity by virtue of its *in vivo* conversion to menaquinones, chiefly MK-4. The latter is a pharmaceutical and water-soluble form which does not occur in food and unlike the other forms is toxic at high intakes. Such water-soluble derivatives of menadione are used as animal feed supplement but owing to toxicity to humans menadione cannot be used in human medicine or as a food supplement.

Phylloquinone of either dietary or endogenous (bacterial) origin is rapidly concentrated in the liver where it is further metabolized, mainly to its epoxide. Apart from instances of fat-malabsorption and cases of hemorrhagic disease of early life, especially in some exclusively breast-fed babies, overt and unequivocal vitamin K deficiencies are rare. However, there is increasing evidence to suggest that vitamin K status may be very important for a variety of physiological processes and for health status, especially in older people, by virtue of its effects on certain components and processes of aging bone and of the vasculature. Research activity in these areas of biochemistry and epidemiology is currently intense.

Phylloquinone (vitamin K<sub>1(20)</sub>) and also the other K-vitamins are destroyed in alkaline medium and isomerize in daylight. They are easily reduced but are fairly stable toward oxidizing conditions and heat.

It has been demonstrated that vitamin K<sub>1(20)</sub> is an essential cofactor in the post-translational carboxylation of glutamic acid residues (GLU) to  $\gamma$ -carboxyglutamic acid residues (GLA) in a number of blood clotting factors (II, VII, IX, and X) and in some other proteins such as protein C, protein S, and osteocalcin, the latter being an important component of bone that can also be measured in the blood stream.

### Methods of Extraction and Cleanup

Before vitamin K<sub>1(20)</sub> can be extracted from blood samples, denaturation of its transport proteins (lipoproteins) is necessary. Ethanol, methanol, or 2-propanol is frequently used for this purpose. For

extraction of vitamin  $K_{1(20)}$  from milk samples lipase treatment is to be preferred over alkaline saponification due to the instability of the compound.

Following the hydrolysis and/or denaturation step, vitamin  $K_{1(20)}$  can be extracted either in a monophasic procedure with a mixture of methanol and dichloromethane, or by a biphasic extraction with *n*-hexane. The selectivity, however, of the above-mentioned extractions is rather low and a lot of interfering (lipophilic) compounds are co-extracted. More recently, SFE was also applied to the analysis of vitamin  $K_{1(20)}$  in infant formula. In these procedures the vitamin is adsorbed onto silica gel and eluted before further separation.

### Chemical Methods

In the early days of vitamin  $K_{1(20)}$  discovery, colorimetric and UV-visible spectrophotometric methods were applied for identification and quantitative purposes. Also, the formation of fluorescent derivatives has been described. Other physicochemical methods include titration with cerium(IV) sulfate, polarography, microcoulometry, and phosphorimetry. These methods, however, have now been largely replaced by modern chromatographic techniques such as LC and to a lesser extent GC and TLC.

### Chromatographic Methods

**Thin-layer chromatography** TLC procedures for vitamin  $K_{1(20)}$  can be divided into three main types: (1) adsorption chromatography on silica plates enabling the separation of *cis-trans* isomers, (2) argentation chromatography to separate saturated and unsaturated homologs of vitamin  $K_{1(20)}$ , and (3) reversed-phase chromatography enabling the separation of methylated and demethylated K-vitamer.

These systems are somehow complementary to each other. Despite the introduction of HPTLC and densitometric scanning, the explosive advances of LC in the past decades explain the rapid decline of TLC applications in the vitamin K area.

**Gas chromatography** The high retention times and the on-column degradation of the K-vitamer always hampered the application of GC techniques in this area. However, both vitamin  $K_{1(20)}$  and  $K_1$  epoxide can be chromatographed on OV-17 as well as on CP-Sil 5 columns. Detection is then performed in the flame ionization detection mode or preferably with electron capture detection. The latter procedure results in higher sensitivity and increased selectivity.

**Liquid chromatography** Adsorption chromatography facilitates the separation of *cis*- and *trans*-vitamin

$K_{1(20)}$ . Furthermore, it is the method of choice for a further cleanup of crude extracts of blood or food samples. Under well-defined conditions, vitamin  $K_{1(20)}$  is then separated from different classes of lipids but co-elutes with the internal standard, which is typically a homolog of vitamin  $K_{1(20)}$  that does not occur naturally in the sample. The fraction containing both vitamin  $K_{1(20)}$  and the internal standard is then collected, concentrated to dryness, and used in a further analytical chromatographic step. Modern detection systems can eliminate the evaporation step by increasing the detection sensitivity.

Similar to adsorption chromatography, also on normal bonded phase packing materials, *cis*- and *trans*-vitamin  $K_{1(20)}$  can be separated.

Reversed-phase chromatography is the method of choice for the final step of vitamin K assays. It can easily separate vitamin  $K_{1(20)}$  from lipids with closely related polarities and from structural analogs used as internal standard. Similar to the other fat-soluble vitamins nonaqueous reversed-phased systems are preferable because of the increased solubility of vitamin  $K_{1(20)}$  and co-extracted lipids in the eluents that can be used.

The most interesting aspect of K-vitamer analysis by LC, however, is the detection mode. UV detection, the traditional detection mode, is only suitable for quantification of high levels of vitamin  $K_{1(20)}$ , e.g., after administration of pharmacological doses. Vitamin  $K_{1(20)}$  also exhibits electrochemical activity. Electrochemical detection, however, requires a semi-aqueous mobile phase to support the conducting electrolyte. This restricts the applications to reversed-phase chromatography unless the electrolyte is added postcolumn. The latter procedure however, results in a drastic loss in sensitivity. Reductive electrochemical detection of K-vitamer suffers from oxygen interference and passivation of the working electrode. Dual electrode detection in the redox mode, on the other hand, improves both the sensitivity and the baseline stability.

A third and currently preferred option to detect K vitamer in LC is fluorescence detection. Vitamin  $K_{1(20)}$  does not show native fluorescence, and consequently a number of procedures have been developed to modify vitamin  $K_{1(20)}$  into a fluorescent molecule. A first procedure consists of the use of a coulometric detector as a postcolumn reactor to reduce vitamin  $K_{1(20)}$  to the fluorescent hydroquinone form. Reduction of vitamin  $K_{1(20)}$  can also be obtained by a chemical reaction, either by a reagent that is added postcolumn or one that is incorporated in the eluent itself. In both cases elevated temperature is necessary for the reaction (thermally induced wet-chemical reaction). Alternatively, and currently preferred, the

**Table 5** Representative procedures for the determination of vitamin K<sub>1(20)</sub>

Analyte	Matrix	Sample preparation/extraction/ purification	Analytical column	Mobile phase (v/v)	Detection <sup>a</sup>	
<i>Gas chromatography</i> Menadione	Feeds	SFE/Silica trap CH <sub>2</sub> Cl <sub>2</sub> elution	DB-5, 30 m × 0.25 mm		MSD	
<i>Liquid chromatography</i> K <sub>1(20)</sub>	Plasma	Methanol denat./SE: <i>n</i> -hexane	Ultrasphere ODS 5 μm, 25 × 0.45 cm	ACN <sup>b</sup> :CH <sub>2</sub> Cl <sub>2</sub> (70:30)	ABS	254
K <sub>1(20)</sub>	Milk	Ethanol denat./SE: <i>n</i> -hexane; SPE: Sep-Pak silica Spherisorb CN column	Spherisorb octyl 15 μm	MeOH:0.05 mol l <sup>-1</sup> Acetate Buffer (95:5)	ECD	Redox – 1.3/0 V
K <sub>1(20)</sub>	Serum	Ethanol denat./SE: <i>n</i> -hexane μPorasil column	Zorbax ODS	30 mmol l <sup>-1</sup> Na Acetate pH 6 in EtOH:H <sub>2</sub> O (96.5:3.5) EC Reduction –0.55 V	FL	340/430
K <sub>1(20)</sub>	Plasma	Ethanol denat./SE: <i>n</i> -hexane ROSil column	RSIL C18 HL 5 μm	MeOH:Ethyl Acetate (96:4) + (CH <sub>3</sub> ) <sub>4</sub> NB <sub>3</sub> H <sub>8</sub> (200 mg per 100 ml)	FL	325/430
K <sub>1(20)</sub>	Plasma	Ethanol denat./SE: <i>n</i> -hexane SPE: Sep-Pak Silica	Hypersil ODS 5 μm	2.0 mol l <sup>-1</sup> ZnCl <sub>2</sub> 1.0 mol l <sup>-1</sup> Na Acetate 1.0 mol l <sup>-1</sup> Acetic acid in MeOH:CH <sub>2</sub> Cl <sub>2</sub> (80:20) Reduction on Zn column	FL	248/418
K <sub>1(20)</sub>	Infant formula	SFE CO <sub>2</sub> /Silica trap CH <sub>2</sub> Cl <sub>2</sub> elution	μBondapack C18, 10 μm	ACN:CH <sub>2</sub> Cl <sub>2</sub> :0.025 mol l <sup>-1</sup> NaClO <sub>4</sub> (90:5:5)	ECD	–1.1 V

<sup>a</sup> Wavelength (nm), etc.<sup>b</sup> Abbreviations, see **Tables 1–4**; SFE, supercritical fluid extraction; MSD, mass selective detection.

reduction can also be performed at room temperature on a solid-phase reactor packed with zinc metal particles. Besides the hydroquinone forms, the photodecomposition products of vitamin K<sub>1(20)</sub> are fluorescent. This phenomenon resulted in a third procedure to detect K-vitamins after a high-performance liquid chromatographic separation, namely postcolumn photochemical reaction detection. With modern equipment it is possible to quantitate vitamin K in as little as 0.25 ml human plasma from nonsupplemented subjects, a considerable improvement in sensitivity over previous assays. An external quality assurance scheme (KEQAS) is run from St. Thomas' Hospital, London.

### Competitive Protein Binding and Immunological Methods

As far as we know there is no immunological method available for the determination of the K-vitamins.

### Examples and Limitations of Current Methodology

Some of the most relevant procedures for the analysis of K-vitamins are brought together in Table 5. It should be clear that again LC is the technique of choice, as this technique is applied both to the sample cleanup and to the analytical measurement of vitamin K<sub>1(20)</sub> in different matrices. As with all modern analytical programs, a keystone for success is a sound quality control and assessment strategy, which should always include the exchange of samples between different analytical laboratories and the use of round robin external quality assurance schemes wherever possible.

**See also:** **Amperometry.** **Coulometry.** **Extraction:** Solvent Extraction Principles; Supercritical Fluid Extraction. **Gas Chromatography:** Detectors; Mass Spectrometry. **Liquid Chromatography:** Column Technology; Normal Phase; Reversed Phase; Ion Pair. **Quality Assurance:** Internal Standards. **Spectrophotometry:** Biochemical Applications. **Supercritical Fluid Chromatography:** Applications.

### Further Reading

- Ball GFM (1998) *Bioavailability and Analysis of Vitamins in Foods*. London: Chapman and Hall.
- Bates CJ (1997) Vitamin analysis. *Annals of Clinical Biochemistry* 34: 599–626.
- Bates CJ, Thurnham DI, Bingham SA, Margetts BM, and Nelson M (1997) Biochemical markers of nutrient intake. In: Margetts SA and Nelson M (eds.) *Design Concepts in Nutritional Epidemiology*, 2nd edn. Oxford: Oxford University Press.
- De Leenheer AP, Nelis HJ, and Lambert WE (1993) Lipophilic vitamins. In: Mukherjee KD and Weber N (eds.) *CRC Handbook of Chromatography: Analysis of Lipids*, pp. 261–283. Boca Raton, FL: CRC Press.
- De Leenheer AP, Lambert WE, and Nelis HJ (1992) *Modern Chromatographic Analysis of Vitamins*. New York: Dekker.
- O'Neil CA and Schwartz SJ (1992) Chromatographic analysis of *cis/trans* carotenoid isomers. *Journal of Chromatography* 624: 235–252.
- Rizzolo A and Polesello S (1992) Chromatographic determination of vitamins in foods. *Journal of Chromatography* 624: 103–152.
- Sauberlich HE (1999) *Laboratory Tests for the Assessment of Nutritional Status*, 2nd edn. Boca Raton, FL: CRC Press.

## Water-Soluble

**C J Bates**, Elsie Widdowson Laboratory, Cambridge, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by M T Parviainen, pp. 5393–5401, © 1995, Elsevier Ltd.

### Vitamins and Their Properties

Foods contain ~40 essential nutrients, which can be divided into four main groups: nutrients of energy metabolism, essential amino acids, macro and micro mineral elements, and vitamins. Vitamins are a group of organic substances, of which four are fat-soluble and nine are water-soluble. The latter group consists of vitamin C (ascorbic acid), and eight vitamins of

group B (B<sub>1</sub>, thiamin; B<sub>2</sub>, riboflavin; B<sub>3</sub>, niacin (nicotinic acid); B<sub>5</sub>, pantothenic acid; B<sub>6</sub>, pyridoxine; biotin; folate/folic acid; B<sub>12</sub>, cobalamins). The B-group vitamins are now more commonly referred to directly by name. Designations such as B<sub>3</sub>, B<sub>5</sub>, etc. are only found in the older literature. Each of these actually contain one or more vitamins, in addition to the parent compound. The classical definition of vitamins states that these essential nutrients are not formed in humans, and must be obtained through the diet. Rich food sources of certain water-soluble vitamins include yeast, liver, wheat germ, and vegetables. They are stored only to a limited extent in the body. Most are absorbed efficiently, and occur mainly in the water phase of the body. When this becomes saturated, any excess of the water-soluble vitamins is

reduction can also be performed at room temperature on a solid-phase reactor packed with zinc metal particles. Besides the hydroquinone forms, the photodecomposition products of vitamin K<sub>1(20)</sub> are fluorescent. This phenomenon resulted in a third procedure to detect K-vitamins after a high-performance liquid chromatographic separation, namely postcolumn photochemical reaction detection. With modern equipment it is possible to quantitate vitamin K in as little as 0.25 ml human plasma from nonsupplemented subjects, a considerable improvement in sensitivity over previous assays. An external quality assurance scheme (KEQAS) is run from St. Thomas' Hospital, London.

### Competitive Protein Binding and Immunological Methods

As far as we know there is no immunological method available for the determination of the K-vitamins.

### Examples and Limitations of Current Methodology

Some of the most relevant procedures for the analysis of K-vitamins are brought together in Table 5. It should be clear that again LC is the technique of choice, as this technique is applied both to the sample cleanup and to the analytical measurement of vitamin K<sub>1(20)</sub> in different matrices. As with all modern analytical programs, a keystone for success is a sound quality control and assessment strategy, which should always include the exchange of samples between different analytical laboratories and the use of round robin external quality assurance schemes wherever possible.

**See also:** **Amperometry.** **Coulometry.** **Extraction:** Solvent Extraction Principles; Supercritical Fluid Extraction. **Gas Chromatography:** Detectors; Mass Spectrometry. **Liquid Chromatography:** Column Technology; Normal Phase; Reversed Phase; Ion Pair. **Quality Assurance:** Internal Standards. **Spectrophotometry:** Biochemical Applications. **Supercritical Fluid Chromatography:** Applications.

### Further Reading

- Ball GFM (1998) *Bioavailability and Analysis of Vitamins in Foods*. London: Chapman and Hall.
- Bates CJ (1997) Vitamin analysis. *Annals of Clinical Biochemistry* 34: 599–626.
- Bates CJ, Thurnham DI, Bingham SA, Margetts BM, and Nelson M (1997) Biochemical markers of nutrient intake. In: Margetts SA and Nelson M (eds.) *Design Concepts in Nutritional Epidemiology*, 2nd edn. Oxford: Oxford University Press.
- De Leenheer AP, Nelis HJ, and Lambert WE (1993) Lipophilic vitamins. In: Mukherjee KD and Weber N (eds.) *CRC Handbook of Chromatography: Analysis of Lipids*, pp. 261–283. Boca Raton, FL: CRC Press.
- De Leenheer AP, Lambert WE, and Nelis HJ (1992) *Modern Chromatographic Analysis of Vitamins*. New York: Dekker.
- O'Neil CA and Schwartz SJ (1992) Chromatographic analysis of *cis/trans* carotenoid isomers. *Journal of Chromatography* 624: 235–252.
- Rizzolo A and Polesello S (1992) Chromatographic determination of vitamins in foods. *Journal of Chromatography* 624: 103–152.
- Sauberlich HE (1999) *Laboratory Tests for the Assessment of Nutritional Status*, 2nd edn. Boca Raton, FL: CRC Press.

## Water-Soluble

**C J Bates**, Elsie Widdowson Laboratory, Cambridge, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by M T Parviainen, pp. 5393–5401, © 1995, Elsevier Ltd.

### Vitamins and Their Properties

Foods contain ~40 essential nutrients, which can be divided into four main groups: nutrients of energy metabolism, essential amino acids, macro and micro mineral elements, and vitamins. Vitamins are a group of organic substances, of which four are fat-soluble and nine are water-soluble. The latter group consists of vitamin C (ascorbic acid), and eight vitamins of

group B (B<sub>1</sub>, thiamin; B<sub>2</sub>, riboflavin; B<sub>3</sub>, niacin (nicotinic acid); B<sub>5</sub>, pantothenic acid; B<sub>6</sub>, pyridoxine; biotin; folate/folic acid; B<sub>12</sub>, cobalamins). The B-group vitamins are now more commonly referred to directly by name. Designations such as B<sub>3</sub>, B<sub>5</sub>, etc. are only found in the older literature. Each of these actually contain one or more vitamins, in addition to the parent compound. The classical definition of vitamins states that these essential nutrients are not formed in humans, and must be obtained through the diet. Rich food sources of certain water-soluble vitamins include yeast, liver, wheat germ, and vegetables. They are stored only to a limited extent in the body. Most are absorbed efficiently, and occur mainly in the water phase of the body. When this becomes saturated, any excess of the water-soluble vitamins is



excreted through the kidneys in urine. Water-soluble vitamins are relatively nontoxic when compared to fat-soluble vitamins, and thus relatively high doses are usually well tolerated.

Water-soluble vitamins participate in a vast number of enzymatic reactions, in most cases in the form of co-enzymes. Many of these enzymatic reactions are well characterized. Many biochemical pathways in the body necessitate several different and specific enzymatic reactions, each having its own vitamin-containing co-factors.

The number and chemical variety of vitamin co-enzyme forms, and the instability of many of their reduced forms, have made it difficult to separate and quantify vitamins in complex mixtures. In general, crystalline water-soluble vitamins are relatively stable in the presence of air, whereas in solution some are easily oxidized. They are generally most stable in acidic media, and least stable in alkaline ones (Table 1). Their degradation in solution is further dependent on the temperature, the presence or absence of oxygen, and the presence or absence of metal ions. Biological

**Table 1** Major chemical characteristics of the water-soluble vitamins and their co-enzyme forms

<i>Vitamin or respective active coenzyme form(s)</i>	<i>Major chemical characteristics</i>
AA, DHA	AA is a strongly reducing dibasic acid with a $pK_1$ of 4.1 and a $pK_2$ of 11.8. It is easily oxidized by several agents (e.g., oxygen, halogens, hydrogen peroxide), especially in aqueous solutions, to form DHA. This reaction is initially reversible, since DHA can be reduced to AA by agents such as glutathione, but subsequent ring-opening converts DHA to diketogulonate, which is an irreversible reaction in the body. Oxygen, heavy metal ions ( $Cu^{2+}$ , $Ag^+$ , $Fe^{3+}$ ), and alkaline pH especially favor the oxidation of AA to DHA. Oxidation of DHA is irreversible.
Thiamin, TPP	Thiamin in aqueous solution is most stable between pH 2 and 4, and labile at alkaline pH. It is also heat labile. TPP is stable in aqueous solution at pH 2–6 and at 0°C. In alkaline oxidation these form fluorescent thiochromes.
Riboflavin, FMN, FAD	Riboflavin is a yellow-green naturally fluorescent compound. Riboflavin and FMN are very easily destroyed by UV radiation. When protected from light, riboflavin is rather stable at neutral and slightly acidic pH. FAD is also more stable at acidic pH, and even at elevated temperatures.
Niacin, $NAD^+$ , $NADP^+$ , and reduced forms NADH and NADPH	Nicotinic acid is stable in acid and alkali, even when heated. Nicotinamide is very stable only in neutral solution. The oxidized co-enzymes (NAD, NADP) are most stable in acidic conditions; aqueous and neutral solutions are less stable. In alkaline pH the decay is accelerated. The reduced co-enzymes (NADH, NADPH) are rapidly destroyed in acid. In alkali they are more stable, although they tend to oxidize after long periods of time. Note that in the older literature the niacin co-enzymes that are now abbreviated to NAD, NADP, etc. were formerly known as (i.e., abbreviated to) DPN, TPN, etc.
Pantothenic acid, CoA, CoASH	Pantothenic acid is most stable in a slightly acidic medium (pH 4–5). Both in acid and alkali it is cleaved. The salt form of pantothenic acid is even more stable. CoA is relatively stable in solutions of pH 2–6.
Pyridoxine, PLP	Highly photosensitive. Otherwise relatively stable, although PLP is easily oxidized or hydrolyzed.
Biotin	Biotin is very stable; it is even autoclavable in concentrated sulfuric acid.
Folic acid (PGA), tetrahydrofolate, 10-formyltetrahydrofolate, and 5-methyltetrahydrofolate	Natural folates are easily oxidizable, whereas folic acid (PGA) is not. The use of antioxidants in sample preparation is mandatory. AA and reducing thiols are widely used. Alkaline pH and heating can result in the conversion of 10-formyltetrahydrofolate to other folates and degradation products.
Cyanocobalamin, AdoCbl, MeCbl	Cyanocobalamin is stable in aqueous solution between pH 4 and 7, and can be heated without significant loss. Cyanide can be split off on exposure to light. AdoCbl and MeCbl are extremely photosensitive. They are both stable in neutral aqueous solution in the dark, and withstand heating.

AA, ascorbic acid; AdoCbl, adenosylcobalamin; CoA, co-enzyme A; DHA, dehydroascorbic acid; TPP, thiamin pyrophosphate; FAD, flavin adenine dinucleotide; FMN, flavin mononucleotide; MeCbl, methylcobalamin;  $NAD^+$ , nicotinamide adenine dinucleotide;  $NADP^+$ , nicotinamide adenine dinucleotide phosphate; PLP, pyridoxal phosphate; PGA, pteroylglutamic acid.



samples may contain oxidation products of the vitamins. Therefore, it is essential that samples are suitably preserved immediately after collection. Acidification, protein precipitation, addition of reducing agents (e.g., thiols), and freezing are some of the necessary measures in vitamin analysis. Some vitamins (e.g., flavins) are light sensitive, and this must be taken into account in the analysis for these substances. Addition of chelating agents (to remove metal ion catalysts) and even removal of oxygen (by inert gas) may be necessary in the assays of oxidizable vitamins. However, some water-soluble vitamins, such as nicotinic acid and biotin, are stable even in conditions of extreme pH and heat.

## Methods of Extraction and Cleanup

Water-soluble vitamins have been determined using a wide range of analytical methods: biological (including microbiological), chemical, physicochemical, chromatographic, and immunological. The need for purification and the method of cleanup for analytical samples greatly depends on the vitamin and/or vitamins to be measured, the analytical method that is to be used, and on the sample material. The physicochemical properties of vitamins may place limitations on the methodology. Furthermore, the sample matrix must be taken into consideration. The vitamin/vitamin concentration in a sample may also vary greatly: pharmaceutical vitamin preparations typically have the highest concentrations, and in biological samples their concentration may be very low. In the latter case, concentration and efficient separation from other interfering substances is a paramount requirement for successful analysis.

Manipulation in the analysis of biological samples (plasma, urine, tissue samples) is designed to decrease the amount of protein and other interfering material, and it may be necessary also to control the oxidation/reduction state of the vitamin. The chemical characteristics of the vitamins (Table 1) are of

major importance when choosing the method. Reduction of the protein content may be accomplished by precipitation using acid, heating, or organic solvent. The release of some vitamins from the corresponding esters or conjugates (e.g., folate polyglutamates, thiamin phosphate esters) may demand enzymatic or acid hydrolysis.

The use of solid-phase extraction cartridges is now well established in the analysis of clinical specimens. However, although this method provides efficient purification of the sample, it may lead to a loss of protein-bound vitamins. Direct injection of plasma samples into liquid chromatography (LC) columns is possible in some applications. Dilute filtered or centrifuged urine can be injected in certain LC applications, as is the case in urinary riboflavin assay.

The analysis of foods follows a similar procedure. Liquids (e.g., milk, juice, and infant formula) often require the same process as is applied for plasma or urine, although here lipid removal may also be needed. Solid food is mechanically disrupted or homogenized and hydrolyzed when necessary, followed by extraction or dissolution in a suitable buffer or solvent plus hydrolytic enzyme treatment to release the vitamins from their bound forms, especially those in vitamin-containing enzymes. The analytical method may be directly applied to a clarified supernatant or may require further concentration and purification of the sample, e.g., on a solid-phase extraction cartridge prior to analysis. Table 2 shows an outline of the sample preparation scheme for the determination for thiamin and riboflavin in different foods. This sample preparation procedure efficiently reduces sample matrix interferences and excess oxidizing reagent is flushed off. The extraction cartridge concentrates the vitamins when very low levels are to be analyzed. The subsequent LC analysis is shown in Figure 1 for selected foods.

Pharmaceutical multivitamin preparations can often be analyzed after minimal manipulation: in chromatographic analysis, the finely powdered material is

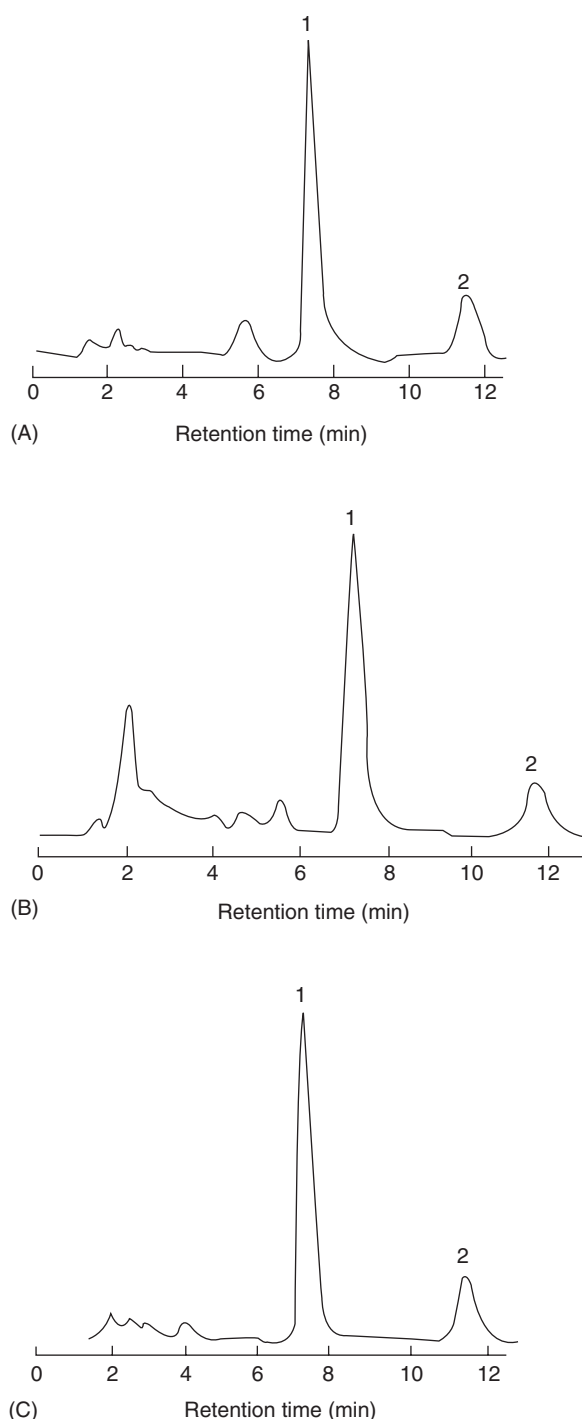
**Table 2** Outline of the sample preparation for determination of thiamin and riboflavin in foods

---

Add 2.5 g of finely ground sample to 50 ml of 0.1 mol l <sup>-1</sup> HCl and sonicate for 5 min.
Autoclave at 121°C for 30 min.
Cool and adjust pH to 4.5 with 2 mol l <sup>-1</sup> sodium acetate.
Add 2.5 ml of 5% diastase and incubate at 45–50°C for 3 h.
Pipette 4.0 ml of a filtered or centrifuged sample extract into a test tube and add 3.0 ml 1% potassium hexacyanoferrate(III) in 15% NaOH solution and vortex for 10 s.
Allow 1 min for oxidation and add 3.0 ml of 3.75 mol l <sup>-1</sup> HCl and vortex for 10 s.
Condition a C <sub>18</sub> Sep-Pak with 5 ml of methanol followed by 5 ml of 5 mmol l <sup>-1</sup> ammonium acetate pH 5.
Load 5 ml of reacted sample onto C <sub>18</sub> Sep-Pak (dropwise), and rinse with 5 ml 5 mmol l <sup>-1</sup> ammonium acetate pH 5.
Elute thiochrome (oxidized thiamin) and riboflavin from the Sep-Pak with 4 ml of methanol/5 mmol l <sup>-1</sup> ammonium acetate, pH 5 (60/40), and inject into the liquid chromatograph.

---

Adapted from Millipore Corp., Milford, MA.



**Figure 1** Reversed-phase LC separation of thiamin and riboflavin from food samples. Column: Waters  $\mu$ Bondapak<sup>TM</sup> C<sub>18</sub> (3.9  $\times$  300 mm). Eluent: 28% methanol in 5 mmol L<sup>-1</sup> ammonium acetate, pH 5.0. Flow rate: 1.5 mL min<sup>-1</sup>. Detector: Waters<sup>TM</sup> 470 fluorescence detector. (A) Flour: 370 nm excitation, 430 nm emission for thiochrome; 370 nm excitation, 520 nm emission for riboflavin. 1, thiamin (measured as thiochrome) 0.704 mg per 100 g. 2, riboflavin 0.442 mg per 100 g. (B) Bread: 1, thiamin (measured as thiochrome) 0.499 mg per 100 g. 2, riboflavin 0.288 mg per 100 g. (C) Cornflakes: 1, thiamin (measured as thiochrome) 3.05 mg per 100 g. 2, riboflavin 1.93 mg per 100 g. (Reproduced with permission from Millipore Corporation, Milford, MA.)

dissolved, or liquid extract is diluted in the mobile phase containing the internal standard. After clarification of the supernatant, appropriate aliquots are injected directly into the LC column. Moreover, the number of vitamers, i.e., different chemical forms used in pharmaceuticals, is generally, rather, limited.

## Methods of Analysis

The earliest methods for vitamin analysis were based on animal bioassays. Later, when it was recognized that vitamins are essential for certain microorganisms, microbiological tests were evolved. The chemical methods for determination of water-soluble vitamins were quite common in the past, before the development of more specific and selective analytical methods.

### Chemical Methods

Chemical determination without separation can be exploited, especially in the analysis of pure or highly concentrated samples such as some foods (e.g., fruit juice and other liquid formula) and pharmaceuticals. Biological samples often contain unknown compounds that can cause nonspecific interference in less sophisticated assays. The chemical methods usually lack the specificity and sensitivity in the quantitative determination of vitamins in biological samples or differentiation between different vitamers.

The physicochemical properties of the water-soluble vitamins are extensively utilized in chemical methods. A method for quantitative vitamin C (ascorbic acid, AA) measurement in food and physiological samples is based on a reaction of the keto groups in dehydroascorbic acid (DHA) with *o*-phenylenediamine (OPD) to give a fluorescent quinoxaline. This method involves the oxidation of AA to DHA, followed by the measurement of 'total AA' in the sample. The reductive capabilities of AA can especially be utilized for direct electrochemical (amperometric or coulometric) measurement when coupled with HPLC separation.

By alkaline oxidation with cyanogen bromide or potassium hexacyanoferrate(III), thiamin can be converted into an inactive product, thiochrome, which is strongly fluorescent in ultraviolet (UV) radiation. The same reaction also occurs with thiamin phosphate esters without affecting the phosphate bond. The cyanogen bromide-based assay is susceptible to interference by some drugs. Riboflavin has a natural brilliant greenish-yellow fluorescence in UV radiation, thus providing a means of quantification pharmaceuticals, and some biological samples.

Alternatively, at alkaline pH riboflavin can be photochemically converted to lumiflavin, which can be specifically extracted with chloroform for subsequent analysis.

Direct fluorimetry of pyridoxal phosphate and pyridoxal is unreliable due to interfering fluorescent compounds, and thus sample cleanup prior to the assay is required. Oxidation of pyridoxal phosphate to pyridoxic acid phosphate by alkaline cyanide, or reaction with bisulfite or semicarbazide can enhance the fluorescence and hence the sensitivity of the assay.

The reactivity of biotin with several reagents can be exploited in some chemical assays like its reaction with diazo derivatives, or the reaction of the ureido ring with *p*-dimethylaminocinnamaldehyde in an acidic medium, but the sensitivity of these assays is relatively poor. Several methods described for biotin assay involve a competitive 'complex' formation between biotin and avidin bound with a chromophore/fluorophore probe. In these assays, biotin, which has a higher affinity for avidin, quantitatively displaces the probe from the complex.

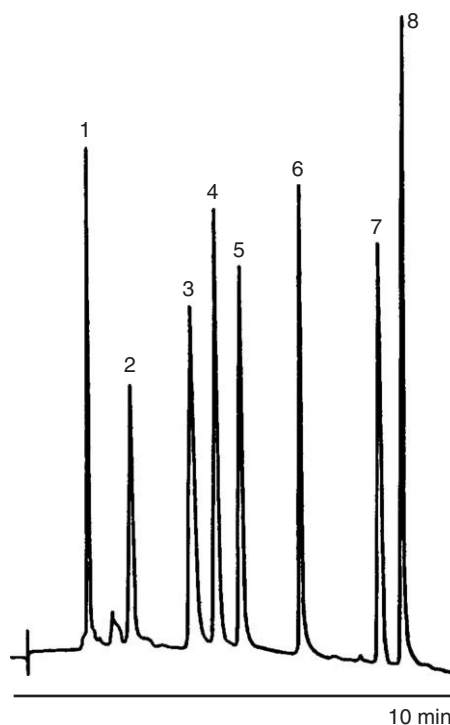
A more specific type of chemical assay is based on enzymatic measurement of vitamin co-enzyme activity. This approach is designed to detect a vitamin deficiency in tissues, and is only feasible for those vitamins that serve as co-enzymes. For instance, thiamin depletion in a subject can be diagnosed by measuring the transketolase activity in red blood cells with and without the addition of thiamin pyrophosphate (TPP) *in vitro*. If TPP increases the activity by more than a given amount, thiamin deficiency is indicated. Similarly, a subnormal level of riboflavin is indicated in tissues if the activity of erythrocyte glutathione reductase is increased after the addition of flavin adenine dinucleotide (FAD). Erythrocyte transaminase activation by pyridoxal-5'-phosphate (PLP) can be measured to establish a deficiency of vitamin B<sub>6</sub>.

### Chromatographic Methods

Chromatographic vitamin assays are now widely used. These methods are often capable of separating and determining different vitamers, and even several vitamins and vitamers simultaneously. LC has proved to be the most powerful analytical technique for several vitamins. Detection can be by light absorption (visible or UV), fluorescence, electrochemical detection, or mass spectrometry (MS). Polarography has been used in some instances. Gas-liquid chromatography (GC), thin-layer chromatography (TLC), capillary electrophoresis, tandem mass spectrometry, and other chromatographic methods are less often used for routine analyses.

LC has proved to be one of the most useful techniques for the analysis of water-soluble vitamins and their vitamers in foods, body fluids, tissues, and pharmaceuticals. One example of LC analysis can be seen in **Figure 2**, which demonstrates the separation of eight different vitamins in a single analysis. The main advantage of LC is its versatility in column technology and detection methods. The basic LC instrumentation is nowadays becoming less expensive. Recent improvements such as column switching technique, robotics in sample preparation, and especially computerized chromatogram processing increase versatility and reduce the 'hands-on' time considerably. One apparent drawback of LC technology is the expertise needed in method development and, to some extent, in its use. Some typical LC conditions applicable to water-soluble vitamin analyses are shown in **Table 3**.

Ion-exchange and reversed-phase media are most commonly used as stationary phases in the LC analysis of water-soluble vitamins. Aminopropyl-bonded silica columns (e.g., LiChrosorb NH<sub>2</sub>,  $\mu$ Bondapak NH<sub>2</sub>) have been used as weak anion exchangers for



**Figure 2** Reversed-phase LC separation of eight water-soluble vitamins. Column: Vydac 201HS54. Eluent: gradient from 2.5% to 50% acetonitrile with 0.1 mol l<sup>-1</sup> acetate buffer, pH 5.2 for 15 min. 1, vitamin C; 2, niacin; 3, pyridoxine (B<sub>6</sub>); 4, thiamin (B<sub>1</sub>); 5, nicotinamide (B<sub>3</sub>); 6, folic acid (M); 7, cyanocobalamin (B<sub>12</sub>); 8, riboflavin, (B<sub>2</sub>). (Reproduced with permission from The Separations Group, Hesperia, CA.)

**Table 3** Determination of water-soluble vitamins by LC

<i>Vitamin</i>	<i>Sample</i>	<i>Sample preparation</i>	<i>LC</i>
Ascorbic acid	Plasma/urine Fruit and vegetables	Extract with 5% <i>m</i> -HPO <sub>4</sub> (1 + 9), centrifuge/filter	Spherisorb NH <sub>2</sub> MeCN/5 mmol l <sup>-1</sup> phosphate (75:25) UV 254 nm
Thiamin	Wheat flour, bread, cornflakes	Extract with 0.1 mol l <sup>-1</sup> HCl, autoclave 30 min, incubate with diastase, precolumn oxidation, cleanup with Sep-Pak C18	μBondapak C18 Methanol/5 mmol l <sup>-1</sup> ammonium acetate pH 5 (28:72) Fluor ex 370 nm, em 430 nm
Riboflavin	Wheat flour, bread, cornflakes	As thiamin	As thiamin Fluor ex 370 nm, em 520 nm
Nicotinamide	Meat, urine, blood tissues	Extract with diethyl ether or methanol	C18/ODS MeCN/10 mmol l <sup>-1</sup> phosphate buffer pH 3 (4.96) UV 260 nm
Pyridoxine, pyridoxal, and pyridoxamine	Pork, beef, eggs, milk, potatoes, frozen peas	Extract and hydrolyze with 0.1 mol l <sup>-1</sup> H <sub>2</sub> SO <sub>4</sub> , filter	Spherisorb ODS 80 mmol l <sup>-1</sup> H <sub>2</sub> SO <sub>4</sub> Fluor ex 290 nm em 395 nm

the simultaneous assay of many of the water-soluble vitamins (thiamin, riboflavin, pyridoxine, cyanocobalamin, AA, folic acid, niacin, niacinamide, and calcium pantothenate) in multivitamin preparations. Relatively few applications are available for more complex samples, e.g., for foods and biological samples using these columns, because better separation is achieved in reversed-phase media, which has the added flexibility of combination with a range of ion-pairing reagents; see below. Amino columns, however, are frequently used for the determination of AA in foods and biological samples. The mobile phase in assays based on amino columns consists of an acidic buffer (phosphate, acetate, or citrate, pH usually between 2.5 and 5) and an organic modifier (e.g., methanol or acetonitrile).

A monomolecular octadecyl layer bound to a porous microparticulate silica support is a basis for modern reversed-phase LC column technology. The popular abbreviations used for such columns include C18, RP, and ODS. Octyl (C8) and phenyl phases are other forms of the reversed-phase medium that have also been used for water-soluble vitamin analysis. The possible ionization of water-soluble vitamins must be taken into consideration (e.g., the carboxyl functions of glutamic acid in folate, the hydroxyl groups of ascorbic acid, etc.) when adjusting the pH of solution. It is usual to mask these ionizable groups to enable interaction of the analyte with the nonpolar stationary phase. This may be accomplished either by using an acidic pH for the mobile phase to cause complete protonation or by using an ion-pairing agent in the mobile phase. Tetrabutylammonium

salts, and pentane-, hexane-, heptane- or octanesulfonic acid are frequently used for the latter purpose. Organic modifiers that are often used in the mobile phase of phosphate or acetate buffers include methanol and acetonitrile. A sufficient chromatographic separation in most cases is achieved by isocratic elution (Figure 1). Gradient elution may be indicated for certain complex samples or multivitamin analyses (Figure 2).

Three methods of detection (UV absorbance, fluorescence, and electrochemical activity) are applicable for the analysis of water-soluble vitamins. Currently, UV absorption detection is used in many LC applications for water-soluble vitamins. A simultaneous assay of several vitamins with several wavelengths along with information of peak purity can be accomplished with a diode array absorbance detector. The detection limit of a UV detector is in the order of 1–10 ng (10–100 pmol), which is poorer than that of fluorescence and electrochemical detectors but often sufficient for analysis for many vitamins in foods and physiological samples (Table 4). The lack of selectivity of UV detection may cause problems with interfering and co-eluting contaminants especially in biological samples, thus necessitating sample purification prior to LC.

Several vitamins are naturally fluorescent, a physical characteristic that can be used in their sensitive and selective detection in LC. The detection limit for riboflavin by fluorescence detection is less than 1 pmol, whereas in UV detection it is ~30 times higher. The detection of an LC assay for thiamin and its phosphate esters using the fluorimetric

**Table 4** Water-soluble vitamin content of selected foods and physiological samples

<i>Vitamin or coenzyme<sup>a</sup></i>	<i>Source</i>	<i>Method<sup>a</sup></i>	<i>Result</i>
Ascorbic acid	Human plasma	LC, UV	20–80 $\mu\text{mol l}^{-1}$
	Human milk (fresh)	LC, UV	54 $\text{mg l}^{-1}$ (mean)
	Human milk (pasteurized)	LC, UV	8.6 $\text{mg l}^{-1}$ (mean)
	Lemon	Chemical	40–60 mg per 100 g
	Potato	Chemical	4–30 mg per 100 g
	Black currant	LC, UV	200–300 mg per 100 g
TPP	Human blood	LC, FL	120 $\text{nmol l}^{-1}$ (mean)
Thiamin (total)	Human blood	LC, FL	80–160 $\text{nmol l}^{-1}$
	Human milk	Chemical	0.01 mg per 100 g
	Cow's milk	Chemical	0.04 mg per 100 g
	Pork fillet	Chemical	1.1 mg per 100 g
	Brewer's yeast	Chemical	15.6 mg per 100 g
Riboflavin (total)	Human urine	LC, FL	40–350 $\mu\text{g}$ per g of creatinine
	Human blood	LC, FL	185 $\text{nmol l}^{-1}$ (mean)
	Cow's milk	LC, FL, or UV	0.1–0.16 mg per 100 g
	Cereals	LC, FL	0.1–1.1 mg per 100 g
	Wheat germ	Chemical	0.6–0.8 mg per 100 g
Nicotinic acid	Coffee	Chemical	10–30 mg per 100 g
Niacin, nicotinamide	Yeast	Chemical	50 mg per 100 g
	Wheat bran	Chemical	9–35 mg per 100 g
MNA	Human urine	LC, UV	31 $\mu\text{mol}$ per 24 h (mean)
2-Pyr	Human urine	LC, UV	60 $\mu\text{mol}$ per 24 h (mean)
Pantothenic acid	Human blood	Radioimmunoassay	1–2 $\text{mg l}^{-1}$
	Human urine	Radioimmunoassay	1–7.5 mg per 24 h
	Baker's yeast	Chemical	5–20 mg per 100 g
	Animal liver	Chemical	4–10 mg per 100 g
	Cow's milk	Chemical	0.3–0.4 mg per 100 g
Pyridoxine	Human serum	LC, FL	19 $\text{nmol l}^{-1}$
Pyridoxal	Human serum	LC, FL	23 $\text{nmol l}^{-1}$
PLP	Human serum	LC, FL	57 $\text{nmol l}^{-1}$
Total B <sub>6</sub>	Yeast	Chemical	1.1 mg per 100 g
	Wheat	Chemical	0.6 mg per 100 g
Biotin	Human plasma	CPB	300 $\text{ng l}^{-1}$ (mean)
	Human urine	Microbiological	18.5–46 $\mu\text{g}$ per 24 h
	Cow's milk	Microbiological	2 $\mu\text{g}$ per 100 g
	Sunflower	Microbiological	119 $\mu\text{g}$ per 100 g
	Soy	Microbiological	25.8 $\mu\text{g}$ per 100 g
Folic acid	Human blood	CPB	100–260 $\text{nmol l}^{-1}$
	Human serum	CPB	4–15 $\text{nmol l}^{-1}$
	Brewer's yeast	Microbiological	1500 $\mu\text{g}$ per 100 g
	Cow's milk	Microbiological	5–12 $\mu\text{g}$ per 100 g
	Animal liver	Microbiological	100–1000 $\mu\text{g}$ per 100 g
Cobalamin	Human serum	CPB	180–660 $\text{pmol l}^{-1}$
	Beef liver	Microbiological	122 $\mu\text{g}$ per 100 g
	Pork	Microbiological	0.56 $\mu\text{g}$ per 100 g
	Swiss cheese	Microbiological	1.71 $\mu\text{g}$ per 100 g

<sup>a</sup>Abbreviations: TPP, thiamin pyrophosphate; MNA, *N*<sup>1</sup>-methylnicotinamide; 2-Pyr, *N*<sup>1</sup>-methyl-2-pyridone-5-carboxamide; PLP, pyridoxal-5'-phosphate; LC, liquid chromatography; UV, ultraviolet detection; FL, fluorescence detection; CPB, competitive protein-binding assay.

thiochrome method is as low as 50 fmol. In the area of folate analysis, tetrahydrofolate, 5-formyltetrahydrofolate, and 5-methyltetrahydrofolate can be determined at picomole levels by detecting their

native fluorescence. Folic acid itself must be converted to a fluorescent pterin derivative by a postcolumn reaction. Similarly, weakly fluorescent B<sub>6</sub> vitamins (pyridoxine, pyridoxal, pyridoxamine) can be oxidized



or converted to their semicarbazone, bisulfite, or hydroxysulfonate derivatives, which are fluorescent in an alkaline and neutral pH, respectively. Fluorescence detection is exemplified in Figure 1 for thiamin (as thiochrome) and riboflavin.

Electrochemical detection in LC provides a sensitive assay method for certain vitamins, such as AA, folates, and flavins. AA may be easily detected with femtomolar sensitivity. Sample preparation and matrix interference problems limit the routine applicability of electrochemistry in the analysis of water-soluble vitamins currently to AA.

Compared to LC, GC is currently less popular in routine vitamin analysis. The polar nature and low volatility of the water-soluble vitamins along with the heat lability of some vitamins hamper analysis by direct GC, making prior derivatization mandatory. GC has been used for the analysis of thiamin, niacin, pyridoxine, cobalamins, biotin, pantothenate, and AA, and for folate derivatives, especially with respect to stable isotope labeled probes. Most GC assays have been applied to relatively simple and pure samples such as multivitamin preparations. Only a few methods have been described for the determination of water-soluble vitamins in complex matrices such as foods and biological samples. A combination of GC with MS has been used in studies on the metabolism of pyridoxine, folate, and some other vitamins. This combination, although expensive in terms of instrumentation, holds promise as a reference method in the analysis of vitamins and their vitamers. The applicability of other analytical techniques could be verified using such a reference method. The main advantages of GC-MS, especially of the isotope dilution MS technique, include high sensitivity, precision, and accuracy of the quantitative measurement. The sample preparation prior to analysis may be more complicated than, for example, LC analysis, and the instrumentation is expensive. Very recently, tandem mass spectrometry and capillary electrophoresis have emerged as useful separation procedures for the highly sensitive analysis of some vitamins in 'difficult' biological matrices.

### Immunological Methods

Immunoassays, specific protein-binding assays, and radioisotope tests are sometimes used for the determination of water-soluble vitamins. These are the only feasible and practical methods for the quantification of certain vitamins in physiological samples. The principle of competitive protein binding (CPB) using labeled radioactive or fluorescent tracer is still routinely applied to quantification of serum vitamin

B<sub>12</sub> and folate, and erythrocyte folate. The assay for folate utilizes binding proteins from milk ( $\beta$ -lactoglobulin) and labeled folic acid (usually [<sup>125</sup>I] or fluorescence-labeled folic acid), which competes with the endogenous folate for the limited amount of binder. In this analysis, endogenous conjugase is allowed to cleave naturally occurring folyl-polyglutamates to monoglutamates, often in the presence of AA, this cleavage being necessary since the polyglutamates have a higher affinity for the binder and can thus degrade the accuracy of the assay. Moreover, the C<sub>1</sub> substituent of folates and the reduction status of the pyrazine ring also have an effect on the affinity for the binder. This assay can quantitatively measure folate at a concentration  $<1 \text{ nmol l}^{-1}$ . In clinical analyses the folate assay in erythrocytes may be favored, since it can reflect long-term folate status better than serum folate status does. Assay using CPB of serum cyanocobalamin is based on competition between the nonlabeled vitamin of the sample and the labeled vitamin (usually [<sup>57</sup>Co]cyanocobalamin or a fluorescent derivative) for a known amount of a specific binder, which usually is the intrinsic factor (IF). The sensitivity of this assay is  $\sim 20 \text{ pmol l}^{-1}$ . These two binding assays have also been combined in the form of a commercially available 'dual assay', in which the radioactive binders have been attached to an easily separable polymer. After incubation, the bound and free radioactivity are separated, and the radioactivity of the sediment is measured using the decay energies of iodine-125 and cobalt-57. With the aid of calibration graphs, the concentrations of these vitamins in serum can be calculated. More recently, fluorescence assays, which avoid the hazards of radioactive tracers, have tended to replace these radioassays.

Pyridoxine has been determined by using an enzymatic immunoassay, in which free vitamin competes with enzyme-labeled vitamin (pyridoxal- $\beta$ -galactosidase) for binding with antipyridoxal antibodies. Binding of the enzyme-labeled vitamin inhibits the enzymatic activity of  $\beta$ -galactosidase, while increasing amounts of the standard pyridoxamine leads to increasing enzyme activity. The vitamin concentration in the sample can then be extrapolated from a calibration graph. The detection limit of assay is  $\sim 50 \text{ nmol l}^{-1}$ . Two apparent problems with the method are cross-reactivity with other B<sub>6</sub> vitamers, the need to develop different antibodies against the various vitamers. Reactivation of tyrosine decarboxylase apoenzyme has been used as an assay for pyridoxal phosphate, and in recent years HPLC-fluorescence assays for vitamin B<sub>6</sub> co-enzymes have become the preferred choice.



Radioimmunoassay has been used for assay of pantothenic acid in physiological samples (sensitivity  $250 \text{ nmol l}^{-1}$ ), and an enzyme-linked immunosorbent assay (ELISA) has been used for this vitamin in food analysis.

## Examples and Limitations of Current Methodology

Table 4 summarizes some of the results obtained by the various methods of the determination of water-soluble vitamins in foods and physiological samples. In the past, biological, microbiological, and physico-chemical methods were widely used. Unfortunately, many of these methods are tedious and cumbersome, and too insensitive and/or nonspecific especially for clinical analyses. One practical problem with biological methods is related to their standardization: cells and microbes are able to utilize vitamins and/or co-enzymes at varying rates. However, these methods are feasible for selected purposes, e.g., in food analysis. Immunological methods are not very appropriate for vitamins. This is partly due to differences found in vitamins in cross-reactions against antibodies. Specific isotope-dilution competitive protein-binding methods, however, are usable at least for biotin, folate, and cyanocobalamin.

Chromatographic methods, especially LC, have offered increased selectivity and sensitivity in vitamin assays. This is reflected in the many methodological publications on the topic over the past decades. Analyses for water-soluble vitamins in physiological samples are now performed routinely using these methods. The problems with these methods are related to sample preparation and to the sensitivity of the detection method that is used. Only a few chromatographic methods enable simultaneous assay of several water-soluble vitamins in physiological samples, and so separate assays are needed. Less complex samples (e.g., pharmaceuticals or fortified foods) are easier to analyze in this respect.

As with all modern analytical programs, an essential keystone for success is a sound analytical quality control and assessment strategy, which should always, where possible, include the exchange of samples between different analytical laboratories and the use of round-robin external quality assurance schemes. External schemes for folate and cobalamins are already in existence in several countries; some other water-soluble vitamin assays are performed regularly by a sufficient number of different laboratories for external sample exchange schemes to be feasible and cost-effective, while others are too specialized and perhaps too rarely performed for regular

exchange schemes to be feasible; hence in these cases, ad hoc arrangements between participating laboratories is the only realistic option. In addition, the analyst should take care to ensure that the biological samples are stored under the optimal conditions for their maximal long-term stability, which usually means deep-freeze temperatures of  $-80^{\circ}\text{C}$  or below, possibly also with the addition of a stabilizing agent like freshly prepared metaphosphoric acid for vitamin C assay samples or AA for folate assay samples. Sample collections and 'banks' are now being prepared in conjunction with large-scale epidemiological studies in several countries for the future analysis of disease-predictive analytes, which may well include vitamins (e.g., EPIC, BIOBANK in Europe, including the UK), and it is critically important to ensure that such priceless sample collections are optimally collected and stored for a robust interpretation of future analytical programs.

*See also:* **Bioassays:** Overview; Microbial Tests; Bioautography. **Blood and Plasma. Fluorescence:** Clinical and Drug Applications; Food Applications. **Food and Nutritional Analysis:** Dairy Products. **Gas Chromatography:** Mass Spectrometry. **Immunoassays, Techniques:** Enzyme Immunoassays. **Liquid Chromatography:** Normal Phase; Reversed Phase; Food Applications. **Microbiological Techniques. Radiochemical Methods:** Food and Environmental Applications. **Vitamins:** Fat-Soluble.

## Further Reading

- Ball GFM (1998) *Bioavailability and Analysis of Vitamins in Foods*. London: Chapman and Hall.
- Bates CJ (1997) Vitamin analysis. *Annals of Clinical Biochemistry* 34: 599–626.
- Bates CJ, Thurnham DI, Bingham SA, Margetts BM, and Nelson M (1997) Biochemical markers of nutrient intake. In: Margetts BM and Nelson M (eds.) *Design Concepts in Nutritional Epidemiology*, 2nd edn. Oxford: OUP.
- De Leenheer AP, Lambert WE, and Nelis HJ (eds.) (1992) *Modern Chromatographic Analysis of Vitamins*, 2nd edn. New York: Dekker.
- FLAIR Concerted Action no. 10 (1993) Status Papers, 'Vitamins'. *International Journal of Vitamin and Nutrition Research* 63: 247–295.
- Macrae R (1990) HPLC determination of vitamins. *Journal of Micronutrient Analysis* 7: 247–260.
- Russell LF (1996) Water-soluble vitamins. In: Nolle L (ed.) *Handbook of Food Analysis*, vol. 1, pp. 649–713. New York: Dekker.
- Sauberlich HE (1999) *Laboratory Tests for the Assessment of Nutritional Status*, 2nd edn. Boca Raton: CRC Press.

# VOLTAMMETRY

Contents

**Overview**

**Linear Sweep and Cyclic**

**Anodic Stripping**

**Cathodic Stripping**

**Inorganic Compounds**

**Organic Compounds**

## Overview

**R J Forster and D A Walsh**, Dublin City University,  
Dublin, Republic of Ireland

© 2005, Published by Elsevier Ltd.

## Introduction

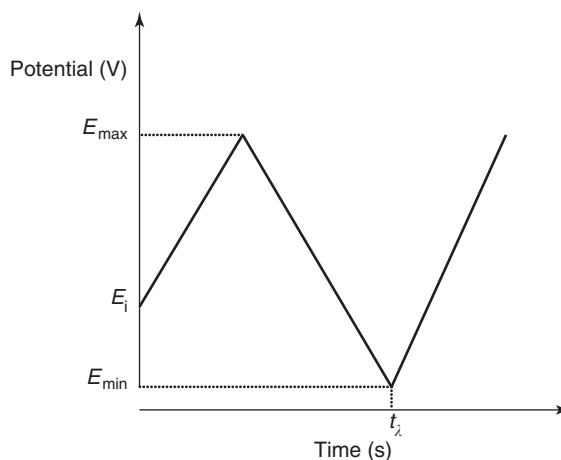
Cyclic voltammetry involves scanning the potential applied to a working electrode according to triangular waveform and monitoring the resulting current flow. The current response provides a powerful and direct insight into the energetics of redox reactions, the dynamics and reversibility of the electron transfer, as well as the rates of coupled chemical reactions. It is the most commonly used electrochemical technique by both the electrochemical specialist and nonspecialist alike. Experimentally, the potential of a working electrode is cycled between two potential limits, driving the successive oxidation and reduction of an electroactive species that can be in solution or adsorbed at the electrode surface. By monitoring the current as a function of potential, current–potential voltammograms can be obtained that are unique to the system under study. Cyclic voltammetry has found applications in areas as diverse as elucidating complex reaction mechanisms and quantifying key environmental, industrial, and medical analytes in solution. In recent years, with the advent of more powerful instruments and microelectrodes, it has been possible to routinely measure the rapid rates of individual electron transfer reactions and even to identify species that exist for just a few nanoseconds using cyclic voltammetry!

In this article, the potential waveform employed to run cyclic voltammetry experiments is first described. The current–voltage waveforms for both reversible and irreversible redox reactions are then presented with an emphasis on the various parameters of

interest that can be extracted from cyclic voltammograms. Finally, some modern applications of the technique are outlined, such as monitoring heterogeneous electron transfer dynamics across the electrode/solution interface and investigation of redox reactions with following chemical reactions.

## Potential Sweep and Current Responses

The triangular potential waveform employed in cyclic voltammetry is shown in **Figure 1**. Typically, the potential is ramped linearly from an initial potential,  $E_i$ , to the switching potential,  $E_{\max}$ . The direction of the potential sweep is then reversed and scanning continues until  $E_{\min}$  is reached. The potential sweep may be terminated at the end of the first cycle or it may continue for an arbitrary number of cycles. The primary experimental parameters are the initial potential, the switching potentials, and the potential sweep rate. Typical sweep rates for cyclic voltammetry, employing electrodes of conventional sizes (e.g.,



**Figure 1** Triangular potential waveform employed in cyclic voltammetry.  $E_{\max}$  and  $E_{\min}$  are the switching potentials;  $t_s$  is the switching time;  $E_i$  is the initial potential.

disk electrodes of radii 1–2 mm) are in the range 1–1000 mV s<sup>-1</sup>. However, the accessible range of experimental sweep rates may be extended to the megavolt per second range by the use of microelectrodes.

The current response for an electroactive species in solution that can be reversibly reduced and reoxidized is shown in **Figure 2**. The initial potential is 0.7 V. Initially, no current flows at the working electrode. As the potential is scanned in a negative potential direction, the potential of the working electrode reaches a sufficiently negative value to cause a reduction of the redox species adjacent to the working electrode surface ( $\sim 0.5$  V). This causes a cathodic current, associated with reduction of the oxidized form of the analyte, to flow which increases rapidly as the surface concentration of oxidized species becomes zero. A diffusion-controlled cathodic current flows at the working electrode as oxidized material diffuses to the electrode surface and becomes reduced. The diffusion-controlled current peaks at  $\sim 0.32$  V and as the potential is scanned to more negative potentials the diffusion layer at the electrode surface becomes depleted of oxidized material. This depletion causes a decrease in the rate of reduction of material at the electrode and a resultant decay in the cathodic current. At potentials negative of 0.32 V, the current decays with  $t^{-1/2}$  according to the Cottrell equation.

At the switching potential, 0 V, the potential sweep is reversed and the potential remains sufficiently negative to cause a reduction reaction at the electrode surface. This small cathodic current continues until the working electrode potential becomes

sufficiently positive to cause oxidation of the reduced species adjacent to the electrode surface ( $\sim 0.25$  V). This causes a diffusion-controlled anodic current to flow, which increases until depletion of reduced material occurs and the anodic current peaks at  $\sim 0.38$  V. The anodic current then decays with  $t^{-1/2}$  as the diffusion layer is depleted of the reduced species. As the potential returns to the initial potential (0.7 V), a small anodic current still flows as the reduced species diffuses to the electrode surface and is oxidized. In the cyclic voltammogram, important quantitative parameters of interest are the cathodic and anodic peak potentials ( $E_{p,c}$  and  $E_{p,a}$ ), the magnitudes of the cathodic and anodic peak currents ( $i_{p,c}$  and  $i_{p,a}$ ), the difference between the cathodic and anodic peak potentials ( $\Delta E_p = |E_{p,a} - E_{p,c}|$ ), the potential at half-height of the peak ( $E_{p/2} = E$  at  $i = i_{p/2}$ ), and the redox formal potential ( $E^{\circ'} = [E_{p,a} + E_{p,c}]/2$ ).

## Reversible (Nernstian) Behavior

For a reversible system in which an oxidized species is reduced with the transfer of  $n$  electrons, the voltammetric peak current (in amperes) at 298 K is given by the Randles–Sevcik equation:

$$i_p = (2.69 \times 10^5) n^{3/2} A D^{1/2} C v^{1/2} \quad [1]$$

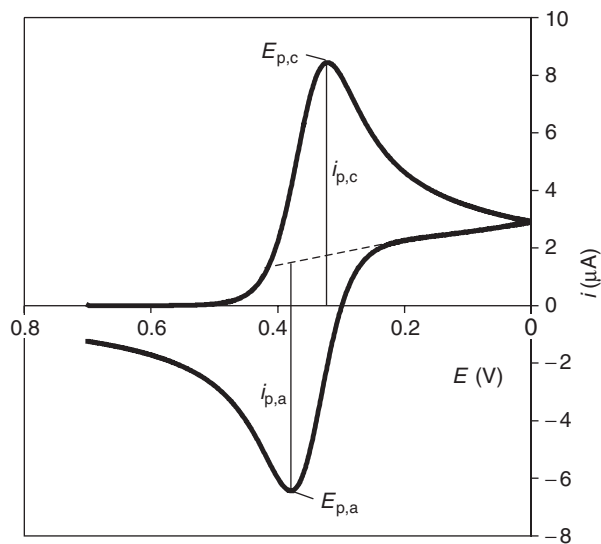
where  $A$  is the electrode area (cm<sup>2</sup>),  $D$  is the diffusion coefficient (cm<sup>2</sup> s<sup>-1</sup>),  $C$  is the bulk concentration of the oxidized species, and  $v$  is the potential sweep rate (V s<sup>-1</sup>).

A common diagnostic of a reversible system is the difference between  $E_p$  and  $E_{p/2}$ :

$$|E_p - E_{p/2}| = 2.2 \frac{RT}{nF} = \frac{56.5}{n} \text{ mV} \quad [2]$$

where  $R$  is the gas constant (8.314 J K<sup>-1</sup> mol<sup>-1</sup>),  $T$  is the temperature (K), and  $F$  is the Faraday constant ( $9.6485 \times 10^4$  C mol<sup>-1</sup>). The peak potential difference,  $\Delta E_p$ , depends on the value of the switching potential relative to the peak potential. When the switching potential exceeds the peak potential by an infinite amount,  $\Delta E_p$  is equal to 57/ $n$  mV.  $\Delta E_p$  increases slightly as the difference between the peak and switching potentials decrease. For example, when the switching potential is 71.5 mV past  $E_p$ ,  $\Delta E_p$  is 60.5 and when the potential is switched 121.5 mV past the peak potential,  $\Delta E_p$  equals 59.2/ $n$  mV.

Thus, for a reversible (Nernstian) process,  $i_p$  is proportional to  $v^{1/2}$ ,  $E_p$  is independent of  $v$ ,  $|E_p - E_{p/2}|$  is 56.5/ $n$  mV,  $\Delta E_p$  is  $\sim 57$  mV, and  $|i_{p,c}/i_{p,a}|$  is 1.



**Figure 2** Cyclic voltammogram for a reversible, one-electron redox reaction.

## Irreversible Behavior

For a completely irreversible system involving the reduction of an oxidized species with the transfer of one electron, the peak current in amperes at 298 K is

$$i_{p,c} = (2.99 \times 10^5) \alpha^{1/2} A C D^{1/2} v^{1/2} \quad [3]$$

where  $\alpha$  is the transfer coefficient and is related to the shape of the energy barrier to electron transfer. As the redox reaction is completely irreversible, the cyclic voltammogram for this system only contains a cathodic wave and linear sweep and cyclic voltammetry will produce the same response.

The difference between the peak potential and the potential at half the peak height is given by

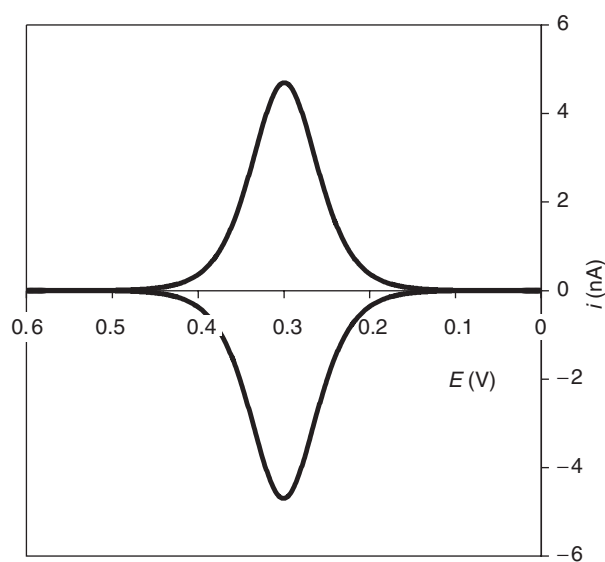
$$|E_p - E_{p/2}| = 1.847 \frac{RT}{\alpha F} = \frac{47.7}{\alpha} \text{ mV at } 25^\circ\text{C} \quad [4]$$

For a completely irreversible wave,  $E_p$  will shift in a negative (for a reduction reaction) potential direction with increasing sweep rate.

## Cyclic Voltammetry of Surface Confined Species

The typical cyclic voltammogram for a redox reaction involving a surface confined, monomolecular layer of a redox species is shown in Figure 3. Both the anodic and cathodic waves are entirely symmetrical across the potential axis and the peak current for the cathodic and anodic reactions is given by

$$i_p = \frac{n^2 F^2}{4RT} v A \Gamma \quad [5]$$



**Figure 3** Cyclic voltammogram for a surface confined, one-electron redox reaction.

where  $\Gamma$  is the surface coverage of electroactive species in moles per square centimeter.

This voltammetric behavior is also observed for thin films (multilayers) of electroactive material at electrode surfaces provided that the experimental timescale is sufficiently long (i.e., slow scan rates). Under these conditions, such a thin layer of redox-active species is exhaustively electrolyzed, finite diffusion predominates, and the voltammetric response will be similar to that shown in Figure 3.

For an immobilized film of electroactive species, the peak current is proportional to  $v$ , rather than the  $v^{1/2}$  dependence observed for freely diffusional species. The area under the wave represents the charge passed during electrolysis of the film and is given by

$$Q = nFA\Gamma \quad [6]$$

A useful insight into the extent of lateral interactions between adsorbates may be obtained from the full-width at half-maximum (FWHM), i.e., the total width of the anodic or cathodic wave at half the peak current. Under Langmuir isotherm conditions (i.e., when there are no interactions between adjacent adsorbates), the FWHM for an ideal Nernstian reaction at 298 K is given by

$$\text{FWHM} = 3.53 \frac{RT}{nF} = \frac{90.6}{n} \text{ mV} \quad [7]$$

Deviations from this value can indicate the presence of attractive ( $< 90.6/n$  mV) or repulsive ( $> 90.6/n$  mV) interactions between adjacent adsorbates.

The formal potential,  $E^{\circ'}$ , contains extremely useful information about the ease of oxidation of redox species within a surface confined film. A shift of  $E^{\circ'}$  of an electroactive species, upon immobilization at an electrode surface, toward more positive potentials indicates that it becomes thermodynamically more difficult to oxidize the film. This suggests a reduced electron density on the redox centre when immobilized.

## Microelectrodes

One of the key barriers to the more widespread adoption of voltammetric techniques was the limited range of media in which analysis could traditionally be performed, i.e., aqueous solutions containing a relatively high concentration of supporting electrolyte. This restriction arose because resistance between the working or sensing electrode and the reference electrode limited the precision with which the applied potential could be controlled. However, microelectrodes, also commonly known as ultra-microelectrodes, whose critical dimension is in the

micrometer range open up new dimensions of space, time, and medium for electroanalysis. The currents observed at microelectrodes typically lie in the picoampere to nanoampere range, which are several orders of magnitude smaller than those observed at conventional millimeter dimensioned macroelectrodes. These small electrolysis currents often completely eliminate ohmic effects. Thus, electrochemical processes can be investigated in high resistance solvents, in solids, in supercritical fluids, and in gases. Moreover, they facilitate measurements at short timescales and at low temperatures.

### Reduced Ohmic Effects

When Faradaic and charging currents flow through a solution, they generate a potential that acts to diminish the applied potential by an amount  $iR_u$ , where  $i$  is the total current and  $R_u$  is the uncompensated cell resistance. This is an undesirable process that leads to distorted voltammetric responses.

The uncompensated solution resistance for a disk-shaped microelectrode is inversely proportional to the electrode radius,

$$R_u = \frac{1}{4\kappa r_0} \quad [8]$$

where  $\kappa$  is the conductivity of the solution and  $r_0$  is the radius of the microdisk. Equation [8] shows that  $R_u$  ‘increases’ as the electrode radius ‘decreases’. However, the currents observed at microelectrodes are typically six orders of magnitude smaller than those observed at macroelectrodes. These small currents often completely eliminate ohmic drop effects even when working in organic solvents. For example, the steady-state current observed at a 5  $\mu\text{m}$  radius microdisk is  $\sim 2 \text{ nA}$  for a  $1.0 \text{ mmol l}^{-1}$  solution of ferrocene in acetonitrile. Taking a reasonable value of  $0.01 \Omega^{-1} \text{ cm}^{-1}$  as the specific conductivity, eqn [8] indicates that the resistance will be of the order of  $50\,000 \Omega$ . This analysis suggests that the  $iR_u$  drop in this organic solvent is a negligible 0.09 mV. In contrast, for a conventional macroelectrode the  $iR_u$  drop would be  $\sim 5\text{--}10 \text{ mV}$ . Under these circumstances, distorted current responses and shifted peak potentials are observed in cyclic voltammetry.

### Capacitive Effects

Altering the potential that is applied to an electrode causes the charge on the metal side of the interface to change and some reorganization of the ions and solvent dipoles in the double layer on the solution side of the interface will occur. This process causes electrons to flow into or out of the surface, giving rise to a charging or capacitive response. The double

layer capacitance for a disk-shaped microelectrode is proportional to the area of the electrode surface and is given by

$$C_{dl} = \pi r_0^2 C_0 \quad [9]$$

where  $C_0$  is the specific double layer capacitance of the electrode. Thus, shrinking the size of the electrode causes the interfacial capacitance to decrease with decreasing  $r_0^2$ . These low capacitive currents are particularly important for analytical applications of cyclic voltammetry where the ability to discriminate a Faradaic signal above a charging current often dictates the limit of detection that can be achieved.

### Mass Transport

A distinct advantage of microelectrodes over larger electrodes is that a range of diffusional mass transport regimes can be observed at timescales that can be easily accessed experimentally. As discussed previously, oxidation or reduction of an electroactive species at the surface of a relatively larger electrode ( $\text{mm}^2$ ) at relatively fast scan rates ( $100 \text{ mVs}^{-1}$ ) causes a diffusional current to flow that peaks at  $E_p$ , after which depletion of electroactive materials occurs and the current decays. However, the timescale of the experiment significantly affects the diffusional mass transport regime and significantly influences the response observed. By examining the time dependence of the diffusional mass transport process it is possible to predict the cyclic voltammetry response and extract quantitative electroanalytical information.

For example, consider a spherical electrode of radius  $r_0$  placed in a solution that contains only supporting electrolyte and a redox-active species at concentration  $C$ . The concentration gradient at the electrode surface after applying a potential step to the electrode is obtained by solving Fick’s second law in spherical coordinates:

$$\frac{\partial C(r, t)}{\partial t} = D \left[ \frac{\partial^2 C(r, t)}{\partial r^2} + \frac{2}{r} \frac{\partial C(r, t)}{\partial r} \right] \quad [10]$$

The boundary conditions for the potential step experiment are

$$\lim_{r \rightarrow \infty} C(r, t) = C^\infty \quad [11]$$

$$C(r, 0) = C^\infty \quad [12]$$

$$C(r_0, t) = 0 \quad [13]$$

where  $r$  is the distance from the center of the sphere,  $r_0$  is the radius of the electrode,  $D$  is the diffusion coefficient for the redox active species, and  $C$  is the concentration as a function of distance  $r$  and time  $t$ .



Equation [10] can be solved using Laplace transform techniques to give the time evolution of the current,  $i(t)$ , subject to the boundary conditions described resulting in eqn [14]:

$$i(t) = \frac{nFADC^\infty}{r_0} + \frac{nFAD^{1/2}C^\infty}{\pi^{1/2}t^{1/2}} \quad [14]$$

Equation [14] reveals that the current contains both time-dependent and -independent terms. The differences in the electrochemical responses observed at macroscopic and microscopic electrodes arise because of the relative importance of these terms at conventional electrochemical timescales. It is possible to distinguish two limiting regimes depending on whether the experimental timescale is short or long.

**Short times** At sufficiently short times, the thickness of the diffusion layer that is depleted of reactant is significantly smaller than the electrode radius and the spherical electrode appears to be planar to a molecule at the edge of this diffusion layer. Under these conditions, the electrode behaves like a macroelectrode and mass transport is dominated by linear diffusion to the electrode surface as illustrated in **Figure 4A**. At short times the first part of eqn [14] becomes insignificant compared to the second part due to the  $t^{-1/2}$  dependence of the current. Therefore, the current decays over time in accordance with the Cottrell equation:

$$i(t) = \frac{nFAD^{1/2}C^\infty}{\pi^{1/2}t^{1/2}} \quad [15]$$

**Long times** At long times (i.e., slow scan rates), the second term in eqn [14] becomes negligible and the current attains a time-independent steady-state value given by eqn [16]:

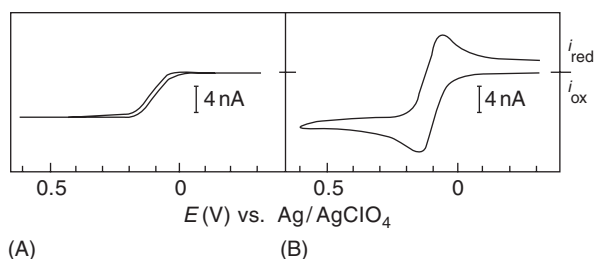
$$i_{ss} = \frac{nFADC^\infty}{r_0} \quad [16]$$

At these long times, the spherical character of the electrode becomes important and the mass transport process is dominated by radial (spherical) diffusion as illustrated in **Figure 4B**.

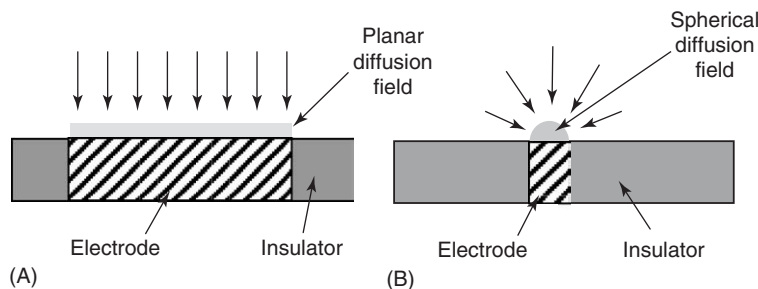
**Figure 5A** shows the sigmoidal-shaped responses that characterize steady-state mass transfer in slow scan-rate cyclic voltammetry. In contrast, as illustrated in **Figure 5B**, at short experimental timescales (high scan rates), peaked responses similar to those observed at conventional macroelectrodes are seen.

The preceding analysis considered a spherical electrode because its surface is uniformly accessible, and a simple closed-form solution to the diffusion equation exists. The microdisk is the most widely used geometry, but derivation of rigorous expressions describing their experimental response is complicated because the surface is not uniformly accessible. For disks, electrolysis at the outer circumference of the disk diminishes the flux of the electroactive material to the center of the electrode. However, microdisk and microring geometries share the advantage of spherical microelectrodes in that quasispherical diffusion fields are established in relatively short periods of time. The steady-state current is given by

$$i_{ss} = \gamma nFDCr_0 \quad [17]$$



**Figure 5** Effect of scan rate on the cyclic voltammetry of  $1.0 \text{ mol l}^{-1}$  ferrocene at a  $6.5 \mu\text{m}$  gold microdisk where the supporting electrolyte is  $0.1 \text{ mol l}^{-1}$  tetrabutyl ammonium perchlorate in acetonitrile: (A) scan rate is  $0.1 \text{ V s}^{-1}$ ; (B) scan rate is  $10 \text{ V s}^{-1}$ .



**Figure 4** Diffusion fields observed at microelectrodes: (A) linear diffusion observed at short times; (B) radial diffusion observed at long times.



where  $\gamma$  is 4 and  $2\pi$  for disk and hemispherical shaped electrodes, respectively.

## Electroanalysis

Voltammetry provides both qualitative and quantitative analytical information about the concentration of redox-active species within samples as diverse as foods, polymers, beverages, drugs, fine chemicals, cosmetics, plating baths, animal feed, raw materials, water, and wastewater. Versatility and simplicity distinguish the technique and the following list highlights some of the main advantages:

- sensitivity,
- speed,
- speciation,
- multicomponent capability,
- versatility,
- response immune to colored or turbid samples, and
- reliability.

As illustrated in Figure 6, a wide range of elements can be analyzed using cyclic voltammetry and this breadth of application is dramatically expanded when one considers the diverse range of inorganic and organic species that is electroactive. Cyclic voltammetry provides routine analysis at low levels using instrumentation and consumables that are widely available at low cost. Also, a wide variety of methods have been developed that can allow complex mixtures to be analyzed often with a minimum

of cleanup. For example, by correlating the formal potentials of unknown components within mixtures with reference compounds it may be possible to identify unknowns within complex mixtures. However, this approach is limited since the experimental response for a single analyte takes up a significant fraction of the available analytical window. While techniques such as differential pulse voltammetry offer significantly greater sensitivity, voltammetry is capable of measuring concentrations at ppb levels.

## Heterogeneous Electron Transfer Dynamics

Beyond an insight into the ease with which compounds can be oxidized or reduced by measuring the formal potential, cyclic voltammetry can be used to determine the rate of electron transfer across the electrode/solution or electrode/film interface. Optimizing the rate of heterogeneous electron transfer is important for technological applications ranging from the analysis of metals in polymers, foods, and cosmetics to the development of biosensors and molecular electronic devices.

Electrolysis (oxidation or reduction) of a freely diffusing electroactive species at an electrode surface during cyclic voltammetry consists of a number of elementary steps. Initially, diffusion of the redox species to the electrode surface under the influence of a concentration gradient must occur. Thermal activation of the redox center and electronic coupling of the activated redox center with the electrode

Periodic table of the elements

1A	2A	3B	4B	5B	6B	7B	8			1B	2B	3A	4A	5A	6A	7A	0
1 H 1.0080																	2 He 4.003
3 Li 6.940	4 Be 9.013											5 B 10.82	6 C 12.010	7 N 14.008	8 O 16.0000	9 F 19.00	10 Ne 20.183
11 Na 22.997	12 Mg 24.32											13 Al 26.97	14 Si 28.06	15 P 30.98	16 S 32.066	17 Cl 35.457	18 Ar 39.944
19 K 39.096	20 Ca 40.08	21 Sc 45.10	22 Ti 47.90	23 V 50.95	24 Cr 52.01	25 Mn 54.93	26 Fe 55.85	27 Co 58.94	28 Ni 58.69	29 Cu 63.54	30 Zn 65.38	31 Ga 69.72	32 Ge 72.60	33 As 74.91	34 Se 78.96	35 Br 79.916	36 Kr 83.7
37 Rb 85.48	38 Sr 87.63	39 Y 88.92	40 Zr 91.22	41 Nb 92.91	42 Mo 95.95	43 Tc 99	44 Ru 101.7	45 Rh 102.91	46 Pd 106.7	47 Ag 107.880	48 Cd 112.41	49 In 114.76	50 Sn 118.70	51 Sb 121.76	52 Te 127.61	53 I 126.92	54 Xe 131.3
55 Cs 132.91	56 Ba 137.36	57 La 138.92	72 Hf 178.6	73 Ta 180.88	74 W 183.92	75 Re 186.31	76 Os 190.2	77 Ir 193.1	78 Pt 195.23	79 Au 197.2	80 Hg 200.61	81 Tl 204.39	82 Pb 207.21	83 Bi 209.00	84 Po (210)	85 At (210)	86 Rn 222
87 Fr (223)	88 Ra 226.05	89 Ac (227)															

**Figure 6** Elements capable of being analyzed using voltammetry at mercury electrodes (elements in the shaded region).

surface must then occur. Finally, the instantaneous elementary electron transfer event occurs. In investigating heterogeneous electron transfer dynamics, one must ensure that the rate of mass transfer to the electrode surface exceeds the rate of electron transfer, for example, by stirring the solution, rotating the electrode, or by using microelectrodes.

Research into this area is dominated by microelectrodes. At short times, the diffusion layer thickness is much smaller than the microelectrode radius and the dominant mass transport mechanism is planar diffusion. Under these conditions, the classical theories, e.g., that of Nicholson and Shain, can be used to extract kinetic parameters from the scan rate dependence of the separation between the anodic and cathodic peak potentials. Using this approach, the standard heterogeneous electron transfer rate constant,  $k^\circ$ , may be determined from the published working curves relating  $\Delta E_p$  to a kinetic parameter  $\Psi$ . The variation of  $\Delta E_p$  with  $\nu$  is determined and, from this,  $\Psi$  is calculated.  $k^\circ$  is then determined by the following equation:

$$k^\circ = \Psi \left[ D_0 \pi \nu \left( \frac{nF}{RT} \right) \right]^{1/2} \left( \frac{D_R}{D_O} \right)^{\alpha/2} \quad [18]$$

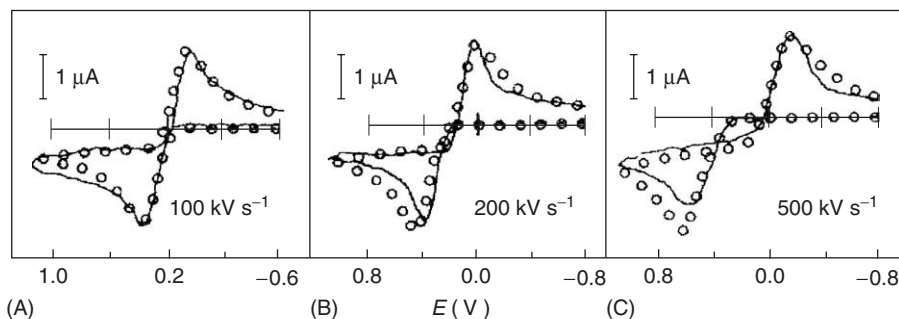
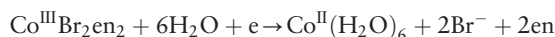
However, this approach can be limited due to factors such as convolution of Faradaic and charging currents at high scan rates. In addition, uncompensated ohmic effects at high scan rates may result in inaccurate measurements of heterogeneous electron transfer kinetics. One particular approach that can overcome these limitations is to simulate the entire cyclic voltammogram. In addition to this, positive feedback circuitry can be used to compensate for Ohmic drop. This approach has been used successfully for a wide range of organic and inorganic compounds and Figure 7 shows an example of a comparison between a simulated and experimental

voltammogram. By simulation of the entire voltammogram, the standard heterogeneous electron transfer rate constant,  $k^\circ$ , has been determined as  $5.1 \text{ cm s}^{-1}$  for the ferrocene/ferrocenium redox couple.

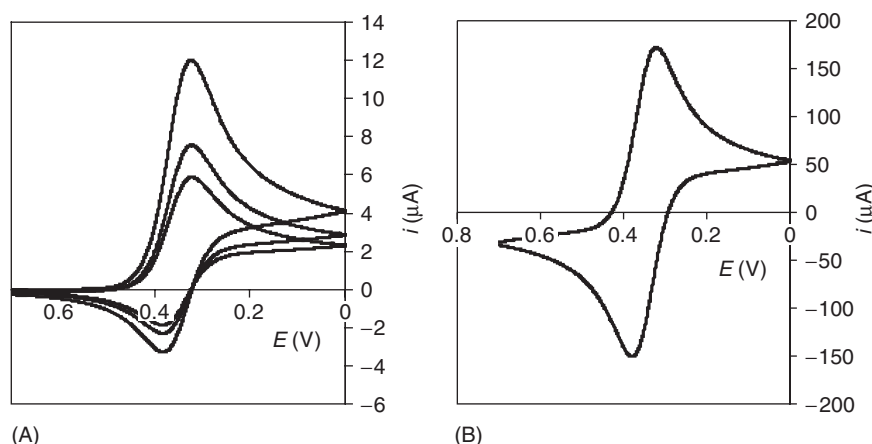
An attractive approach to solving the mass transfer limitations of these investigations is to immobilize the electroactive species at the electrode surface within a monomolecular film. Clearly, if the electroactive species is immobilized at the electrode surface, diffusion of the species to the electrode does not need to occur prior to electron transfer. In addition, immobilization at an electrode surface can preconcentrate the species of interest, resulting in higher currents that are easier to detect. Electroactive adsorbed monolayers have been developed that exhibit close to ideal reversible electrochemical behavior under a wide variety of experimental conditions of timescale, temperature, solvent, and electrolyte. These studies have elucidated the effects of electron transfer distance, tunneling medium, molecular structure, electric fields, and ion pairing on heterogeneous electron transfer dynamics.

## Following Chemical Reactions

Beyond the ability to probe heterogeneous electron transfer reactions that occur on the micro, or even nano timescale, high-speed cyclic voltammetry has found application in the investigation of complex reaction mechanisms. An example of this is the following chemical (EC) reaction. The oxidation or reduction (E reaction) of an electroactive species can produce a chemically unstable product that undergoes a chemical reaction (C reaction) to form an electroinactive product. Electron transfer reactions can also lead to isomerizations and ligand loss in organometallic compounds, such as

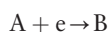


**Figure 7** Background-subtracted experimental (solid lines) and simulated (open circles) cyclic voltammograms for a solution of ferrocene ( $10 \text{ mmol l}^{-1}$ ) in acetonitrile containing  $0.6 \text{ mol l}^{-1}$  TEAP. Sweep rates are 100, 200, and  $500 \text{ kV s}^{-1}$  and the gold microelectrode radius is  $5 \mu\text{m}$ . Simulations allow for ohmic drop and RC constant and  $k^\circ$  is  $3.1 \text{ cm s}^{-1}$ . (Reproduced with the permission of the American Chemical Society from *Analytical Chemistry* 60 (1988) 305.



**Figure 8** (A) First three sweeps of a voltammogram for an EC reaction at a scan rate of  $0.2 \text{ V s}^{-1}$ . (B) Voltammogram for the same system as in (A) at a scan rate of  $50 \text{ V s}^{-1}$ .

where en is ethylenediammine. Figure 8A shows a slow scan rate ( $0.2 \text{ V s}^{-1}$ ) cyclic voltammogram for an EC reaction:



where the reduced product, B, undergoes a reaction to form C, which is electroinactive. The ratio of anodic peak current to cathodic peak current is less than 1. Figure 8B illustrates the same system while using a sweep rate of  $50 \text{ V s}^{-1}$ . The smaller timescale of the voltammogram in Figure 8B illustrates the ability of higher sweep rate cyclic voltammetry to outrun the following chemical reaction.

See also: **Voltammetry**: Linear Sweep and Cyclic; Anodic Stripping; Cathodic Stripping; Inorganic Compounds; Organic Compounds.

## Further Reading

- Bard AJ and Faulkner LR (2001) *Electrochemical Methods: Principles and Applications*, 2nd edn. New York: Wiley.
- Bond AM (2002) *Broadening Electrochemical Horizons: Principles and Illustration of Voltammetric and Related Techniques*. Oxford: Oxford University Press.
- Brett CMA and Brett AMO (1993) *Electrochemistry: Principles, Methods and Applications*. Oxford: Oxford University Press.
- Forster RJ, Vos JG, and Keyes TE (2003) *Interfacial Supramolecular Assemblies*. New York: Wiley.
- Pletcher D (2002) *Instrumental Methods in Electrochemistry*. Albion/Horwood.
- Wang J (2000) *Electroanalytical Techniques*, 2nd ed. New York: Wiley.
- Wightman RM and Wipf DO (1989) In: Bard AJ (ed.) *Electroanalytical Chemistry*, vol. 15. New York: Dekker.

## Linear Sweep and Cyclic

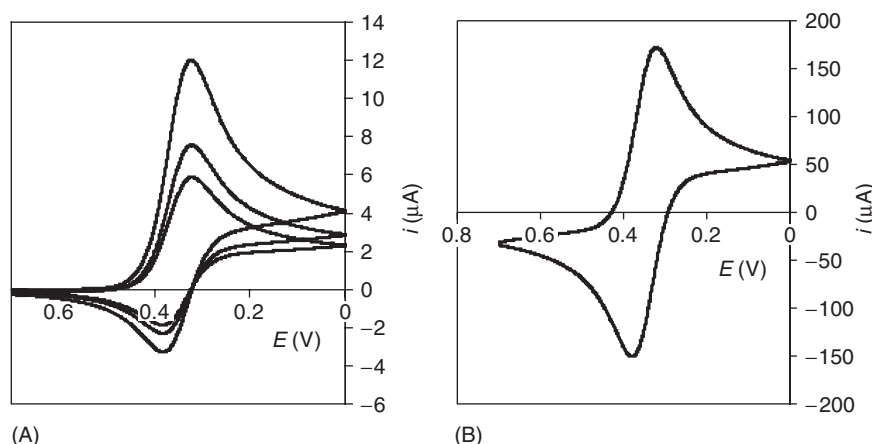
**G Bontempelli and R Toniolo**, Università degli Studi di Udine, Udine, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Linear sweep voltammetry (LSV) and cyclic voltammetry (CV) are the most widely used voltammetric techniques for studying redox reactions of both organic and inorganic compounds because they are unmatched in their ability to provide information on the steps involved in electrochemical processes

with only a modest expenditure of time and effort in the acquisition and interpretation of data. These electroanalytical methods require simple and inexpensive instrumentation and provide not only information on the electrochemical quantities typical of a redox process, but also allow investigations of chemical reactions coupled with charge transfer steps. This is because the electrode can be used as a tool for producing reactive species in a small solution layer surrounding its surface and at the same time to monitor chemical reactions involving these species. Moreover, since the relevant responses can be obtained within a few milliseconds after stimulation of the



**Figure 8** (A) First three sweeps of a voltammogram for an EC reaction at a scan rate of  $0.2 \text{ V s}^{-1}$ . (B) Voltammogram for the same system as in (A) at a scan rate of  $50 \text{ V s}^{-1}$ .

where en is ethylenediammine. Figure 8A shows a slow scan rate ( $0.2 \text{ V s}^{-1}$ ) cyclic voltammogram for an EC reaction:



where the reduced product, B, undergoes a reaction to form C, which is electroinactive. The ratio of anodic peak current to cathodic peak current is less than 1. Figure 8B illustrates the same system while using a sweep rate of  $50 \text{ V s}^{-1}$ . The smaller timescale of the voltammogram in Figure 8B illustrates the ability of higher sweep rate cyclic voltammetry to outrun the following chemical reaction.

See also: **Voltammetry**: Linear Sweep and Cyclic; Anodic Stripping; Cathodic Stripping; Inorganic Compounds; Organic Compounds.

## Further Reading

- Bard AJ and Faulkner LR (2001) *Electrochemical Methods: Principles and Applications*, 2nd edn. New York: Wiley.
- Bond AM (2002) *Broadening Electrochemical Horizons: Principles and Illustration of Voltammetric and Related Techniques*. Oxford: Oxford University Press.
- Brett CMA and Brett AMO (1993) *Electrochemistry: Principles, Methods and Applications*. Oxford: Oxford University Press.
- Forster RJ, Vos JG, and Keyes TE (2003) *Interfacial Supramolecular Assemblies*. New York: Wiley.
- Pletcher D (2002) *Instrumental Methods in Electrochemistry*. Albion/Horwood.
- Wang J (2000) *Electroanalytical Techniques*, 2nd ed. New York: Wiley.
- Wightman RM and Wipf DO (1989) In: Bard AJ (ed.) *Electroanalytical Chemistry*, vol. 15. New York: Dekker.

## Linear Sweep and Cyclic

**G Bontempelli and R Toniolo**, Università degli Studi di Udine, Udine, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Linear sweep voltammetry (LSV) and cyclic voltammetry (CV) are the most widely used voltammetric techniques for studying redox reactions of both organic and inorganic compounds because they are unmatched in their ability to provide information on the steps involved in electrochemical processes

with only a modest expenditure of time and effort in the acquisition and interpretation of data. These electroanalytical methods require simple and inexpensive instrumentation and provide not only information on the electrochemical quantities typical of a redox process, but also allow investigations of chemical reactions coupled with charge transfer steps. This is because the electrode can be used as a tool for producing reactive species in a small solution layer surrounding its surface and at the same time to monitor chemical reactions involving these species. Moreover, since the relevant responses can be obtained within a few milliseconds after stimulation of the

electrode, they may be used for studying mechanisms involving very fast reactions, thus allowing detection of short-lived transient intermediates.

LSV and CV techniques were proposed at the beginning of the 1950s and in those years some theoretical approaches able to rationalize the simplest responses were worked out. However, the use of these electroanalytical methods has received considerable impetus only recently thanks to increased knowledge of more subtle criteria for interpreting the relevant responses and to greater availability of theoretical tools for processing experimental data.

This article will briefly survey the analytical performance of these techniques, as well as the criteria followed to gain information regarding participants in electrochemical processes from the recorded responses.

## Fundamentals

In both LSV and CV, a small ( $<0.1 \text{ cm}^2$ ) stationary working electrode is dipped in an unstirred solution containing an excess of supporting electrolyte to repress migration of charged reactants and products, so that any transfer of electroactive species to and from the electrode surface can occur only through diffusion. In LSV, the potential of the working electrode is changed linearly with time, as shown in the upper plot of Figure 1A, starting from a potential where no

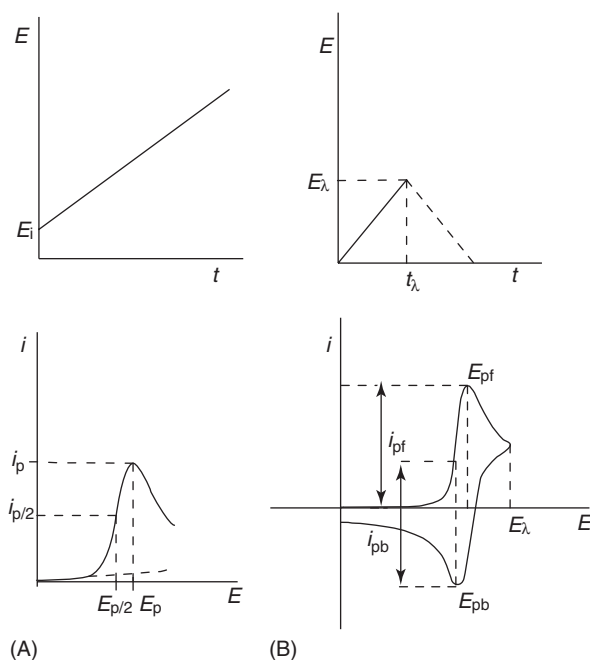
electrode reaction occurs and moving to potential where reduction (more negative values) or oxidation (more positive values) of the investigated analyte takes place, so that the instantaneous potential applied can be written as follows:

$$E_t = E_i \pm vt \quad [1]$$

where  $E_t$  is the electrode potential at time  $t$ ,  $E_i$  the starting potential, and  $v$  the scan (or sweep) rate, i.e., the absolute value of the rate of the potential change,  $dE/dt$ , which is constant within any sweep. The sign depends upon the potential scan direction (positive for anodic sweeps and negative for cathodic sweeps).

The current is measured throughout the experiment and the resulting current–potential curve (voltammogram) displays the typical shape shown in the lower plot in Figure 1A, also reporting main parameters. They are  $i_p$  = peak current, i.e., the maximum current value;  $E_p$  = peak potential, i.e., the potential corresponding to  $i_p$ ;  $E_{p/2}$  = half-peak potential, i.e., the potential at which  $i = i_p/2$ .

Peak position on the potential scale ( $E_p$ ), which is related to the formal potential of the redox process and to its reversibility degree (conditioning also the peak shape  $E_p - E_{p/2}$ ) gives information on the analyte involved. Conversely, peak height ( $i_p$ ) gives information on the analyte concentration, number of electrons involved in the electrochemical process, and possible presence of coupled chemical reactions. As long as the electrode potential is scanned in a region where no charge transfer occurs (from the starting potential  $E_i$  to the rising portion of the peak-shaped response), only a little capacitive current is recorded because the electrode–solution interphase behaves like a capacitor, whose capacitance depends on the electrode potential value because its change causes ions and molecules near the electrode to modify their orientation. On the contrary, when the potential attains the region where the electrochemical reaction occurs, a Faradaic current is recorded which increases sharply until a maximum is reached; thereafter the current decreases steadily. Such a peak shape comes out from the competition of the increase in the rate of the electrode reaction as the potential is made more and more suitable for its occurrence (prevailing in the rising portion of the peak) with the increasing thickness of the diffusion layer due to the progressive depletion of the depolarizer at an increasing distance from the electrode surface (prevailing in the descending branch). Consequently, the peak current is recorded at a potential where these counteracting effects balance each other, i.e., when the analyte concentration at the electrode surface is not yet down to zero, so that  $E_p$  falls at potentials preceding those



**Figure 1** Potential waveforms for linear sweep (A), and cyclic (B) voltammetry and the resulting voltammograms.



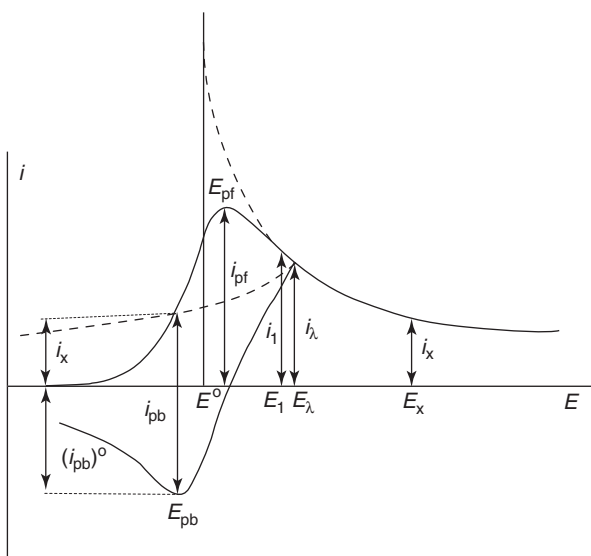
proper for the limiting current region in conventional d.c. voltammetry.

The voltage waveform applied to the working electrode in CV is shown in the upper plot of **Figure 1B**. Before the switching time,  $t_\lambda$ , the electrode potential is varied linearly with time like in LSV, while at  $t_\lambda$  the potential scan direction is reversed, so that the waveform becomes the combination of two joined linear sweeps. The switching potential  $E_\lambda$  is in fact both the final potential for the forward scan and the initial potential for the reverse sweep. Consequently, the potential applied to the working electrode in the forward scan is once again described by eqn [1], while  $E_t$  for the backward scan ( $t_\lambda < t < 2t_\lambda$ ) is given by

$$E_t = E_i \pm vt_\lambda \mp v(t - t_\lambda) = E_i \pm 2vt_\lambda \mp vt \quad [2]$$

where upper signs refer to forward anodic sweeps and lower signs refer to forward cathodic sweeps. This equation applies to symmetrical triangular waveforms that are the most commonly adopted in CV, even though other waveforms have also been proposed. When this triangular waveform is used to investigate electrochemical processes involving oxidized and reduced forms both stable during the time required to record a voltammogram, the electrode reaction occurring in the backward scan involves the species electrogenerated in the forward sweep still present in the solution layer surrounding the electrode surface, so that the electrode reaction symmetrical to the onward reaction takes place. Consequently, as shown in the lower plot in **Figure 1B**, a backward peak is associated to the forward peak due to the renewal of the starting species. In agreement with the notation adopted above,  $i_{pf}$ ,  $i_{pb}$ ,  $E_{pf}$ ,  $E_{pb}$  (f, forward; b, backward) are defined in cyclic voltammograms.

In the rather small potential range where forward and backward reactions occur simultaneously, the current is the algebraic sum of a cathodic contribution and an anodic contribution. Therefore, for a correct evaluation of the backward peak-current a suitable baseline must be used to subtract the contribution of the forward electrode process. This contribution can be experimentally determined by extending the forward scan beyond the switching potential  $E_\lambda$ , until a potential  $E_x$  is reached so that  $E_x - E_\lambda = E_\lambda - E_{pb}$ , as shown in **Figure 2**. The use of this method is, however, precluded when further interfering electrode processes occur beyond  $E_\lambda$ , so that alternative procedures have been proposed, all based on the consideration that the current becomes purely diffusion controlled at a potential beyond enough the peak. Their applicability



**Figure 2** Baseline evaluation for the reverse scan in a cyclic voltammogram.

is, however, restricted to particular electrochemical processes with the exception of the following equation (the meaning of the symbols can be inferred from **Figure 2**) enabling the ratio  $i_{pb}/i_{pf}$  to be calculated directly from a single cyclic voltammogram:

$$\frac{i_{pb}}{i_{pf}} = \frac{(i_{pb})^\circ}{i_{pf}} + \frac{(i_\lambda/i_{pf}) \times \{1 + (E_{pb} - E_\lambda)\}}{[(i_1)^2 - (i_x)^2]/[(i_1)^2(E_1 - E_\lambda)]^{-1/2}} \quad [3]$$

Of course, when the charge transfer is so slow that the associated forward and backward responses are recorded in different potential regions, the baseline for the backward peak is given by the capacitive current alone, as reported above for the forward peak.

## Characterization of Electrochemical Processes

An electrochemical process occurs frequently through a sequence of steps including: (1) the transfer of both the reactant from the bulk of the solution to the electrode surface and the product in the opposite direction, (2) the charge transfer reaction, and (3) possible structural rearrangements and possible chemical reactions preceding, following, or parallel to the charge transfer step, as well as interposed between two subsequent charge transfers. Additionally, adsorption and other surface reactions may need to be considered. Voltammetric responses are affected by each of these steps to an extent depending on the



scan rate adopted, which markedly conditions the timescale of any experiment. Thus, an electron transfer may appear fast or slow and a coupled chemical reaction can or cannot occur, depending upon the potential sweep rate employed. Scan rates varying from few tens of  $\text{mV s}^{-1}$  to hundreds of  $\text{V s}^{-1}$  can be adopted and this favorable opportunity results in timescales (also controlled by the potential range traversed) in the range  $10^{-5}$ – $10^2$  s. In other words, the scan rate defines an experimental time-window through which the electrochemical process is seen, so that chemical reactions occurring significantly at slow enough scan rates become undetectable when  $\nu$  is increased and fast charge transfers can be seen as slow at high enough scan rates.

The effects caused by the main step of those mentioned are shortly reviewed in the following sections, considering first the electrochemical processes in which reactants, intermediates, and products are freely diffusing solution species, not interacting with the electrode surface.

### Uncomplicated Charge Transfer Processes

We first consider electrochemical processes involving sole charge transfers, uncoupled with chemical events or adsorption phenomena, i.e., simple heterogeneous reactions that may occur in the timescale of the experiment with either a very high rate (Nernstian or reversible behavior) or a finite rate (nonreversible behavior). In this last case, the process is defined totally irreversible when forward and backward reactions occur with an appreciable rate only in totally different potential ranges, while it is defined quasireversible when a finite potential range exists in which both forward and backward electrode reactions occur with an appreciable rate, thus affording currents that are the algebraic sum of the two contributions.

In LSV and CV, a redox system may show a Nernstian, quasireversible or totally irreversible behavior depending on the scan rate employed, since  $\nu$  determine the time available for the electrode-solution interphase to attain the equilibrium condition dictated by the Nernst equation. Such a dependence is usually rationalized by the following dimensionless parameter, comparing the standard heterogeneous rate constant  $k_{\text{sh}}$  with the scan rate  $\nu$ :

$$\Psi = \gamma^\alpha k_{\text{sh}} / (\pi a D_{\text{O}})^{1/2} \quad [4]$$

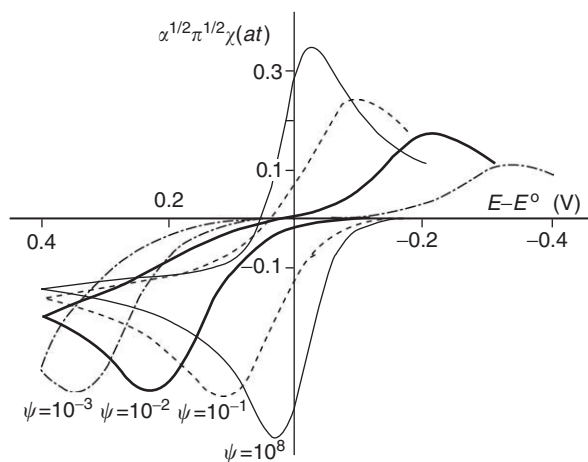
where  $\gamma = (D_{\text{O}}/D_{\text{R}})^{1/2}$  is the square root of the ratio between the diffusion coefficients of the oxidized ( $D_{\text{O}}$ ) and reduced ( $D_{\text{R}}$ ) species,  $\alpha$  is the charge transfer coefficient and  $a = nF\nu/RT$  (where  $n$  is the overall number of electrons involved in a charge transfer,

$F$  is Faraday's constant (charge on 1 mol of electrons) =  $96.487 \text{ (C mol}^{-1}\text{)}$ ,  $R$  is the ideal-gas constant =  $8.31 \text{ J}$  and  $T$  is the absolute temperature (K)) is the normalized scan rate (this definition of  $\Psi$  applies to reduction processes, while for oxidation reactions  $D_{\text{O}}$  is replaced by  $D_{\text{R}}$ ). On these bases, a charge transfer is defined reversible for  $\Psi > 7$ , quasireversible for  $7 > \Psi > 10^{-3}$ , and totally irreversible for  $\Psi < 10^{-3}$ . Consequently, the same redox system displays a reversible behavior for high enough  $\Psi$  values (low values of  $\nu$ , when the system is allowed a long time to attain equilibrium conditions), while it becomes totally irreversible at sufficiently high scan rates when  $\Psi < 10^{-3}$ , on passing through quasireversible conditions for intermediate  $\nu$ . Such a transition is shown in Figure 3 reporting the effect of  $\Psi$  on the response recorded for a one-electron redox system characterized by  $\alpha = 0.5$ .

Indeed, Figure 3 reports normalized current-potential plots, in order to make voltammograms independent of the experimental variables typical of the particular redox process considered. This is accomplished by replacing current with the dimensionless quantity  $\pi^{1/2}\chi(at)$ , i.e., a function of both  $a$  (scan rate) and  $t$  (potential), named current-function, whose generalized form is as follows:

$$\pi^{1/2}\chi(at) = i(at)/nFA(C_{\text{O}})^b(D_{\text{O}})^{1/2}a^{1/2}\alpha^{1/2} \quad [5]$$

where  $i(at)$  is the current,  $A$  is the electrode area,  $(C_{\text{O}})^b$  is the bulk concentration of the oxidized analyte, and other symbols have the meaning defined earlier. For reversible electrode reactions a unity value must be assigned in this equation to the charge



**Figure 3** Effect of the dimensionless parameter  $\psi$  on cyclic voltammetric responses for an uncomplicated one-electron system characterized by  $\alpha = 0.5$ . The starting species is the oxidized one. The x-axis represents the difference of electrode potential ( $E$ ) and standard potential ( $E^\circ$ ).

transfer coefficient  $\alpha$ , even though such a parameter is does not make sense in this particular case.

This current-function exhibits a maximum value (peak-current-function) of 0.4463 for reversible charge transfers and of 0.4958 for totally irreversible redox reactions, but the higher value for this last case must not be misleading, since the following peak-current ratio for these two limiting cases is inferred from eqn [5]:

$$(i_p)_{\text{irr}}/(i_p)_{\text{rev}} = (0.4958/0.4463)\alpha^{1/2} \quad [6]$$

When  $\alpha = 0.5$ , as frequently occurs, this peak-current ratio becomes 0.784, thus pointing out that higher peaks are found under the same experimental conditions for Nernstian charge transfers, in agreement with the progressive peak-lowering and peak-broadening accompanying the decrease of the apparent reversibility degree ( $\psi$ ) shown in Figure 3.

Finally, the linear dependence of the peak current on both the analyte concentration and square root of  $\nu$  must be emphasized. The dependence on the analyte concentration is the basis of the application of LSV to quantitative analysis, while the dependence on  $\nu^{1/2}$  suggests the possibility of modulating suitably the sensitivity of LSV and CV measurements.

Table 1 summarizes the parameters characterizing LSV and CV responses for the different types of uncomplicated charge transfers, as well as their trend with  $\nu$ , whose suitable use enables charge transfers to be characterized even by evaluating their typical quantities ( $k_{\text{sh}}$ ,  $\alpha$ , and  $E_{1/2}$ ).

### Coupled Chemical Reactions

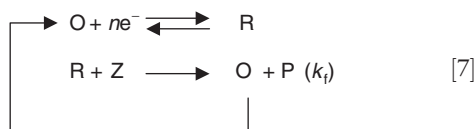
Reactants and products of electrode processes may be involved in chemical reactions following, preceding, or parallel to electron transfers. These reactions affect voltammograms to an extent that depends on both their nature and their rate compared with the timescale of the experiment. In particular, voltammetric responses remain unaltered at scan rates so high as to prevent the occurrence of chemical reactions in the time span of the experiment. Conversely, chemical reactions occurring at higher rates than the potential scan result in potential shifts of voltammetric profiles, at least. When chemical-reaction rates are comparable with the potential sweep rate, voltammograms are affected by both kinetic and thermodynamic characteristics of coupled reactions and it is just under these intermediate conditions that kinetic investigations by LSV and CV are particularly profitable.

**Table 1** Trend of the parameters characterizing linear sweep and cyclic voltammetric responses for uncomplicated reduction processes

Parameter	Process		
	Reversible	Totally irreversible	Quasireversible
$i_p/\nu^{1/2}$	Independent of $\nu$ ; $i_p/\nu^{1/2} = 2.688 \times 108n^{3/2}AD_0^{1/2}C_0^b$ at 25°C	Independent of $\nu$ ; $i_p/\nu^{1/2} = 2.987 \times 108\alpha^{1/2}n^{3/2}AD_0^{1/2}C_0^b$ at 25°C	Slightly dependent on $\nu$ ; it changes from 1.00 to $1.11\alpha^{1/2}$ for about a 108-fold increase in $\nu$
$i_{\text{pb}}/i_{\text{pf}}$	Unity and independent of $\nu$	Lower than one; no relation with $\nu$ has been reported	Unity only for $\alpha = 0.5$
$E_p$	Independent of $\nu$	Dependent on $\nu$ ; it shifts cathodically by $1.15RT/\alpha nF$ mV for a 10-fold increase in $\nu$ (29.5/ $\alpha n$ mV at 25°C)	Dependent on $\nu$ ; for a 10-fold increase in $\nu$ the cathodic shift is smaller than $1.15RT/\alpha nF$ mV
$E_p - E_{p/2}$	Independent of $\nu$ and equal to: $-2.199RT/nF$ mV ( $-56.5/n$ mV at 25°C)	Independent of $\nu$ and equal to: $-1.857RT/\alpha nF$ mV ( $-47.7/\alpha n$ mV at 25°C)	Slightly dependent on $\nu$ ; it changes from $-2.199RT/nF$ mV to $-1.857 \times RT/\alpha nF$ mV (from $-56.5/n$ to $47.7/\alpha n$ mV at 25°C) for about a 108-fold increase in $\nu$
$E_{\text{pb}} - E_{\text{pf}}$	Independent of $\nu$ and equal to: $2.218RT/nF$ mV (57.0/ $n$ mV at 25°C)	Dependent on $\nu$ according to: $\Delta E_p = \{RT/[\alpha(1-\alpha)nF]\}[0.780 - \ln k_{\text{sh}} + \ln(D_0a)^{1/2}] + (RT/nF)[\ln(1-\alpha)^{1/2(1-\alpha)} + \ln \alpha^{(1/2\alpha)}]$	Dependent on $\nu$
$i$ at the foot of the peak	It depends linearly on $\nu^{1/2}$	Independent of $\nu$	It depends on $\nu^{1/2}$ in a nonlinear way

$i$ : current,  $\Delta E_p$ : peak potential separation =  $E_{\text{pb}} - E_{\text{pf}}$  (V).

By way of illustration of the use of LSV and CV to characterize chemical reactions coupled with an electron transfer, the electrochemical pathway relative to electrocatalytic processes (eqn [7]) will be briefly considered. It involves the homogeneous reaction of the reduced product R of a reversible charge transfer with a nonelectroactive substrate Z, allowing regeneration of the electroactive species O, and the concomitant formation of a nonelectroactive product P:

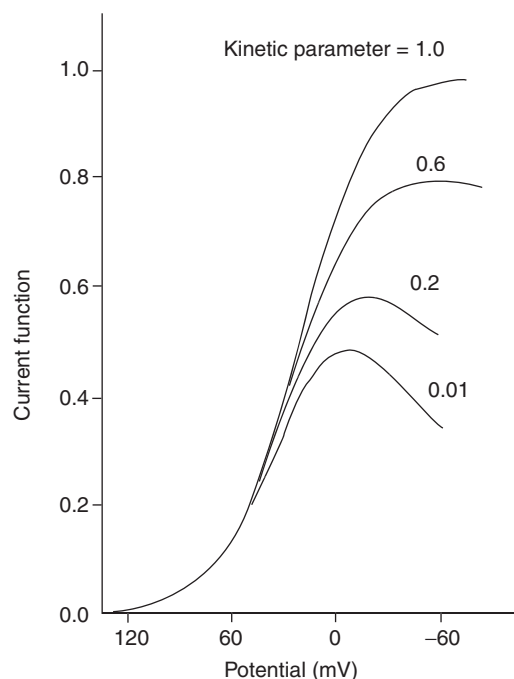


The condition  $(C_Z)^b \gg (C_O)^b$  is usually adopted because it makes relevant responses easier to be interpreted, thanks to the attainment of pseudo-first-order conditions allowing a pseudo-first-order rate constant  $k'_f = k_f(C_Z)^b$  to be adopted.  $(C_Z)^b$  is the bulk concentration of a nonelectroactive analyte ( $\text{mol cm}^{-3}$ ).

To account for the effect of the homogeneous regeneration reaction on the relevant voltammetric profiles, it is convenient to compare this rate constant with the mentioned normalized potential scan rate by adopting the dimensionless kinetic parameter  $k'_f/a$ . Two limiting cases can be recognized. In the former, encountered at scan rates so high as to make  $k'_f/a$  very low, the chemical reaction cannot occur significantly in the timescale of the experiment, thus leading to responses typical for uncomplicated reversible charge transfers. At the other limit, for  $\nu$  so slow as to make high  $k'_f/a$ , the regeneration reaction is allowed to occur quantitatively during the potential sweep and the response is no longer peak-shaped but assumes a sigmoidal shape, as shown in Figure 4. This is because a steady-state for the electroactive species is attained by mutual compensation of its subtraction by the electrode reaction with its chemical regeneration. Under these kinetic conditions, the plateau current attains the following limiting value for high enough overvoltages:

$$i_{\text{lim}} = nFA(C_O)^b(D_O)^{1/2}(k'_f)^{1/2} \quad [8]$$

Such a current, independent of  $\nu$ , enables the homogeneous kinetic constant to be evaluated. To avoid being compelled to know all the experimental quantities accompanying  $k'_f$  in eqn [8], it is convenient to introduce its ratio with the corresponding diffusion controlled current (see eqn [5]), which can be recorded either in the absence of the substrate Z or at scan rates high enough to cancel the effect of the chemical step.



**Figure 4** Theoretical LSV curves for electrocatalytic processes involving reversible charge transfers under pseudo-first-order conditions. Kinetic parameter  $= k'_f/a$ ; potential scale is  $n(E - E_{1/2})$ .

As to CV responses, they display the backward peak typical of uncomplicated reversible processes in the mentioned first limiting case ( $k'_f/a \rightarrow 0$ ), while no backward peak is found for high values of  $k'_f/a$ . Once again, intermediate situations are encountered for intermediate values of  $k'_f/a$ .

By using similar arguments it is possible to anticipate and expound the responses for any type of electrochemical process involving coupled chemical reactions, which are usually classified by using letters indicating the nature of the steps involved. Accordingly, E indicates an electron transfer step and C indicates a chemical step. Thus, a sequence in which a chemical reaction follows a charge transfer is named EC process, while sequences involving a chemical reaction preceding the electron transfer or interposed between two subsequent electrode steps are named CE and ECE processes, respectively. The characteristics displayed by the responses relevant to these types of usual processes are reported in Tables 2–4 summarizing the criteria adopted for characterizing the process involved and for evaluating the relevant kinetic quantities.

### Adsorption Processes

In some electrochemical processes, reactants, intermediates, or products can be either confined onto the electrode surface (adsorbed species, oxide layers,

**Table 2** Characteristics of voltammetric responses for EC processes

Increasing $v$	Case	$O + ne^- \rightleftharpoons R$ $R \xrightarrow{k_f} Z \quad (K = k_f/k_b \gg 1) \quad (EC_{rev})$	$O + ne^- \rightleftharpoons R$ $R \xrightarrow{k_f} Z \quad (EC_{irrev})$
$k_f/a \rightarrow 0$			
Comparable values of $k_f$ and $a$		Uncomplicated reversible process $\left\{ \begin{array}{l} [\pi^{1/2}\chi(at)]_p = 0.4463; \quad n(E_{p/2} - E_p) = 56.5 \text{ mV at } 25^\circ\text{C} \\ \delta E_p / \delta \log v = 0; \quad i_{pb}/i_{pf} = 1 \end{array} \right.$	
$k_f/a \gg 1$ $k_b/a \ll 1$		Superreversible process $\left\{ \begin{array}{l} [\pi^{1/2}\chi(at)]_p : \text{decreases from 0.4958 to 0.4463 at } 25^\circ\text{C} \\ n(E_{p/2} - E_p) : \text{increases from 48.0 to 56.5 mV at } 25^\circ\text{C} \\ \delta E_p / \delta \log v : \text{ranges from } -29.6/n \text{ to } 0 \text{ mV at } 25^\circ\text{C} \\ i_{pb}/i_{pf} = 0 : \text{increases from 0 to 1} \end{array} \right.$	
Comparable values of $k_b$ and $a$		Kinetically uncomplicated reversible process $\left\{ \begin{array}{l} [\pi^{1/2}\chi(at)]_p : \text{increases from 0.446 to 0.4958} \\ n(E_{p/2} - E_p) : \text{decreases from 56.5 to 48 mV at } 25^\circ\text{C} \\ \delta E_p / \delta \log v : \text{ranges from 0 to } -29.6/n \text{ mV at } 25^\circ\text{C} \\ i_{pb}/i_{pf} : \text{decreases from 1 to 0} \end{array} \right.$	Inexistent case
$k_b/a \gg 1$		The response is anodically shifted of an extent : $(RT/nF) \ln(1+K)$ with respect to $E^\circ_{O/R}$ Kinetically uncomplicated reversible process	Inexistent case

covalently attached species, redox polymer films) or even the electrode material itself, as in electrodeposition and electrodisolution. These surface phenomena cause different effects on voltammetric responses, depending on their characteristics (adsorption isotherm, adsorption kinetics, electron transfer coupled to adsorption, etc.), so that a generalized treatment is rather problematic. Consequently, only a brief survey is provided here, aimed at recognizing the presence of surface processes.

The interaction degree with the electrode surface allows a rough distinction to be made, in that weakly adsorbed species cause only enhancement of peak currents. In particular, LSV and CV forward peaks higher than those expected for uncomplicated charge transfers (eqn [5]) are found when the weakly adsorbed species is the reactant, while only enhancement of CV backward peaks is recorded for weak adsorption of the electrode product.

On the contrary, a separate adsorption peak is displayed prior to or after the diffusion-controlled peak when redox products or reactants, respectively, are strongly adsorbed. These adsorption-controlled peaks can be identified because they are symmetrical about  $i_p$ , unlike diffusion-controlled peaks, and a linear dependence of their height with  $(C_0)^b$  is usually observed only in a narrow range of low concentrations, while a constant value for  $i_p$  is attained at higher  $(C_0)^b$ . Moreover,  $i_p$  for adsorption peaks increases linearly with  $v$  instead of  $v^{1/2}$  (eqn [5]), because the electrode-solution interphase, in the presence of adsorbed species, behaves like a capacitor, whose capacitive current ( $i_c$ ) is

$$i_c = dq/dt = C_{dl}(dE/dt) = C_{dl}v \quad [9]$$

where  $C_{dl}$  is the double layer differential capacity depending on both the electrode material and solution composition, as well as on the electrode potential and the analyte concentration, although moderately.

Comparison of eqn [9] with eqn [5] points out that capacitive currents are usually negligible with respect to Faradaic currents at high enough analyte concentrations, but they prevail at very high  $v$ .

## LSV and CV at Microelectrodes

The latest development in LSV and CV has been achieved by the introduction of microelectrodes that are metal or carbon electrodes with diameters usually in the range 0.1–100  $\mu\text{m}$ . Microelectrodes with dimensions down to 0.1  $\mu\text{m}$  are, however, produced by specialists in small numbers worldwide. They were originally developed for *in vivo* biological and

**Table 3** Characteristics of voltammetric responses for CE processes

Increasing $v$	Case	$Z \xrightleftharpoons[k_b]{k_f} O \quad (K = k_f/k_b)$ $O + ne^- \rightleftharpoons R \quad (CE_{rev})$	$Z \xrightleftharpoons[k_b]{k_f} O \quad (K = k_f/k_b)$ $O + ne^- \xrightarrow{k_{sh}} R \quad (CE_{irrev})$
↑	$k_f/a \rightarrow 0$ (or $k_f/\alpha a \rightarrow 0$ )	Uncomplicated reversible process O/R	Uncomplicated totally irreversible process O/R
	Comparable values of $k_f$ and $a$ (or $\alpha a$ ) (Steady state for the concentration of species O)	Pure kinetic process Sigmoidal shaped response with a limiting current independent of $v$ and equal to: $i = nFAD_0^{1/2} C^b K(k_f + k_b)^{1/2}$ $\delta E_{p/2}/\delta \log v = 29.6/n \text{ mV at } 25^\circ\text{C}$ $i_{pb}/i_{pf} > 1$	
	$k_f/a \gg 1$ (or $k_f/\alpha a \gg 1$ )	Uncomplicated reversible process Z/R $(E_{1/2})_{Z/R} = (E_{1/2})_{O/R} - (RT/nF) \ln K/(1+K)$	Uncomplicated totally irreversible process Z/R $(E_{1/2})_{Z/R} = (E_{1/2})_{O/R} - (RT/\alpha nF) \ln K/(1+K)$
		$i_p$ depends linearly on $C^b (= C_0^b + C_2^b)$	

**Table 4** Characteristic of voltammetric responses for pure ECE mechanisms

Increasing $v$	Case	$O + e^- \rightleftharpoons R \quad E_1^0$ $R \xrightleftharpoons[k_b]{k_f} O' \quad K = k_f/k_b \gg 1$ $O' + e^- \rightleftharpoons R' \quad E_2^0 \gg E_1^0$	$O + e^- \rightleftharpoons R \quad E_1^0$ $R \xrightarrow{k_f} O'$ $O' + e^- \rightleftharpoons R' \quad E_2^0 \gg E_1^0$
↑	$k_b/a > 1$ ( $k_f/a \gg 1$ )	Uncomplicated two-electron reversible process; Apparent $E^0 = (E_1^0 + E_2^0)/2 + (RT/2nF) \ln K$ ; $i_{pb}/i_{pf} = 1$	Inexistent case
	Comparable values of $k_b$ and $a$ ( $k_f/a > 1$ )	$[\pi^{1/2}\chi(at)]_p$ decreases from 1.262 to 0.992; $(E_{p/2} - E_p)n$ increases from 29.6 to 47.7 mV at $25^\circ\text{C}$ ; $i_{pb}/i_{pf}$ decreases from 1 to 0	Inexistent case
	$k_b/a < 1$ $k_f/a > 1$	$[\pi^{1/2}\chi(at)]_p = 0.992$ ; $i_{pa}/i_{pb} = 0$ ; $(E_{p/2} - E_p)n = 47.7 \text{ mV at } 25^\circ\text{C}$ ; $\delta E_{p/2}/\delta \log v = -29.6/n \text{ mV at } 25^\circ\text{C}$	
	Comparable values of $k_f$ and $a$ ( $k_b/a < 1$ )	$[\pi^{1/2}\chi(at)]_p$ decreases from 0.992 to 0.4463; $\delta E_{p/2}/\delta \log v$ is at first $< -29.6/n \text{ mV at } 25^\circ\text{C}$ , but it decreases progressively to attain 0; $(E_{p/2} - E_p)n$ increases from 47.7 to 56.5 mV at $25^\circ\text{C}$ on passing through values higher than 56.5; $i_{pb}/i_{pf}$ increases from 0 to 1	
	$k_f/a < 1$ ( $k_b/a \ll 1$ )	Uncomplicated one-electron reversible process $[\pi^{1/2}\chi(at)] = 0.4463$ ; $(E_{p/2} - E_p)n = 56.5 \text{ mV at } 25^\circ\text{C}$ ; $\delta E_{p/2}/\delta \log v = 0$ ; $i_{pb}/i_{pf} = 1$ ; $E_p = E_1^0 - 29/n \text{ mV at } 25^\circ\text{C}$	

medical measurements and attracted initially little attention in the belief that not only the relevant low currents present measurement problems, but also no significant difference would exist between performance of large and small electrodes. Only more recently did it become clear that a decrease in surface area results in improved performance, due to the expansion of the inherently small diffusion layer in the

timescale of the experiment that is greater than the dimensions of the electrode. Consequently, relatively large diffusion layers develop a few seconds after current starts, so that more electroactive particles reach the electrode per unit time and surface area than in the case of conventional electrodes. This means that despite the currents being lower at micro-electrodes, current densities are higher by several



orders of magnitude compared to those at conventional electrodes, so that the Faradaic to capacitive current ratio is drastically improved, allowing lower detection limits ( $10^{-8}$ – $10^{-9}$  mol l $^{-1}$ ) to be attained. Moreover, microelectrodes enable high-quality voltammograms to be recorded even in poorly conducting media (e.g., gas phases or nonpolar solvents), thanks to their small currents (in the nA and pA range) making negligible the relevant ohmic drop. They are virtually nondestructive of the species analyzed, thus being suitable for refined applications, such as *in vivo* measurements, lithography, and electrochemical scanning tunneling microscopy. Nevertheless, microelectrodes require very effective shielding and instrumentation designed suitably to avoid ripple and noise effects, very critical with small currents, which cannot be eliminated with standard equipments.

When CV is conducted at stationary microelectrodes with slow  $\nu$ , in both forward and backward scans sigmoidal current-voltage curves are found which are usually coincident, except for processes involving coupled chemical reactions that display more or less marked hystereses. This sigmoidal shape (steady-state current) can be accounted for by considering the radial diffusion to the edges of ultramicroelectrode surfaces that is very important at slow  $\nu$ , so as to make the diffusion rate of analyte molecules to the electrode surface comparable with the charge transfer rate.

On the contrary, when fast scan rates are used to record cyclic voltammograms at microelectrodes, the radial diffusion to the edges of electrode surfaces becomes negligible and the majority of the diffusion is perpendicular to the electrode surface (usual planar diffusion), thus leading to peak-shaped responses like at conventional electrodes. However, the little currents flowing at these small electrodes result in a very small ohmic drop and this profitable advantage, accompanied by the very favorable ratio between Faradaic and capacitive current mentioned above, permits the use of extremely high scan rates. Thus, while at conventional electrodes  $\nu$  of hundreds of V s $^{-1}$  cannot be overcome, microelectrodes enable sweep rates up to 10<sup>5</sup> V s $^{-1}$  to be used, allowing shorter timescales to be explored and faster reactions to be studied.

## Analytical Applications

Analytical applications of LSV and CV include characterization and quantitative determination of electroactive analytes. Characterization is mainly based on the response position on the potential scale, easily estimated by LSV measurements, even though CV

may provide some additional information about the reversibility of the charge transfer and the presence of adsorptions or coupled chemical reactions. Conversely, quantitative determinations are usually performed solely by LSV since they are based on the response height, so that no additional information is provided by CV.

The detection limit (DL) in LSV is mainly governed by the ratio of the Faradaic signal to charging current background. By restricting our considerations to uncomplicated charge transfers, Faradaic currents are  $i_f = kC\nu^{1/2}$  (eqn [5]), while charging currents are  $i_c = k'\nu$  (eqn [8]), where  $k$  and  $k'$  are constants. Their ratio,  $i_f/i_c = k''C\nu^{-1/2}$ , where  $k'' = k/k'$ , points out that the less favorable DL is achieved at fast  $\nu$  and low concentrations. Typical scan rates used for analytical purposes are in the range 10–1000 mV s $^{-1}$  and offer acceptable compromise for achieving close to optimum performance. Under such conditions, approximately an order of magnitude improvement in DL is gained over conventional d.c. voltammetry, so that concentrations as low as 10 $^{-6}$  mol l $^{-1}$  can be determined by LSV in aqueous media.

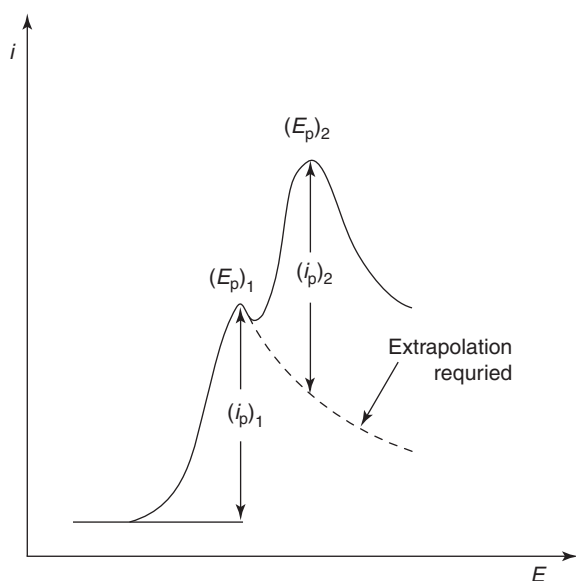
More complicated considerations must be adopted for analytes undergoing adsorption processes. In most of these cases capacitive currents lead, however, to profitable determinations characterized by lower DL (even 10 $^{-8}$  mol l $^{-1}$ ).

Despite the dependence of  $i_p$  on charge-transfer kinetics (eqns [5] and [6]), the sensitivity of LSV is almost independent of the reversibility degree (only a decrease of ~25% is found on passing from a Nernstian to a totally irreversible process with  $\alpha = 0.5$ ). This fact makes LSV the most sensitive voltammetric technique for analytes involved in irreversible processes because pulsed voltammetric methods or alternating current voltammetry provide for these processes very low signals.

In routine analytical work, the main advantages of LSV are: (1) increased sensitivity; quickly recorded current-potential curves, (2) specificity due to both a large available timescale and improved resolution of peak-shaped curves. This last advantage is, however, rather questionable in some cases because the presence of quite close preceding peaks, as shown in Figure 5, makes difficult the correct baseline to be individuated, even when resorting to the procedures mentioned above (Figure 2).

A recent advanced analytical application of LSV and CV is their introduction in electrochemical detectors for flow analysis (HPLC, EC). Fast-scan LSV and CV (20–1000 V s $^{-1}$ ), performed at the rising portion of the typical peaks afforded by these hydrodynamic methods with microelectrodes suitably positioned at the outlet of the flowing system, provide





**Figure 5** Voltammogram relative to the reduction (or oxidation) of the analyte (2) in the presence of another species (1) more easily reduced (or oxidized) when a little difference exists between  $(E_p)_2$  and  $(E_p)_1$ .

in fact additional resolving power. Thus, for instance, peaks separated incompletely in time can be resolved voltammetrically.

See also: **Voltammetry**: Overview; Anodic Stripping; Cathodic Stripping; Inorganic Compounds; Organic Compounds.

## Further Reading

- Bard AJ and Faulkner LR (2000) *Electrochemical Methods; Principles and Applications*, 2nd edn. New York: Wiley.
- Bond AM (1980) *Modern Polarographic Methods in Analytical Chemistry*. New York: Dekker.
- Bontempelli G, Magno F, Mazzocchin GA, and Seeber R (1989) Linear sweep and cyclic voltammetry. *Annali di Chimica (Rome)* 79: 103–216.
- Brett CMA and Oliveira Brett AM (1998) *Electrochemistry; Principles, Methods and Applications*. Oxford: Oxford University Press.
- Brown ER and Sandifer JR (1986) Cyclic voltammetry, AC polarography and related techniques. In: Rossiter BW and Hamilton JF (eds.) *Physical Methods of Chemistry*, 2nd edn. p, pp. 273–432. New York: Wiley.
- Greef R, Peat R, Peter LM, Pletcher D, and Robinson J (1985) *Instrumental Methods in Electrochemistry*. Chichester: Horwood.
- Kissinger PT and Heineman WR (1996) *Laboratory Techniques in Electroanalytical Chemistry*, 2nd edn. Washington: Division of Analytical Chemistry, American Chemical Society.
- Speiser B (2003) Linear sweep and cyclic voltammetry. In: Bard AJ, Stratmann M, and Unwin PR (eds.) *Encyclopedia of Electrochemistry*, vol. 3, pp. 81–124. Weinheim: Wiley-VCH.
- Wang J (1994) *Analytical Electrochemistry*. New York: VCH.
- Wightman RM and Wipf DO (1989) Voltammetry at ultra-microelectrodes. In Bard AJ (ed.) *Electroanalytical Chemistry*, vol. 15, pp. 267–353. New York: Dekker.

## Anodic Stripping

**S Daniele**, University of Venice, Venice, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by J Wang, pp. 5422–5425, © 1995, Elsevier Ltd.

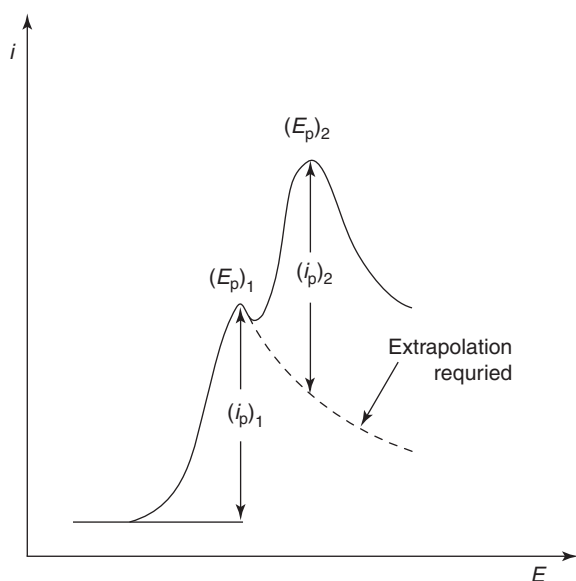
## Introduction

Stripping voltammetry (SV) comprises one of the major families of electroanalytical chemistry. The strength of SV is in its extremely low detection limits ( $10^{-10}$ – $10^{-11}$  mol l $^{-1}$ ). Its remarkable sensitivity is achieved through a preconcentration step in which the target analyte (commonly a metal ion) is accumulated onto working electrodes. This step is followed by the measurement (stripping) step, in which the preconcentrated analyte is stripped back to the solution, a process resulting in a current response

proportional to the analyte concentration in the sample. The different versions of SV depend primarily upon the nature and direction of the preconcentration and measurement steps. Anodic stripping voltammetry (ASV) is the most widely used form of stripping analysis. The extremely low detection of ASV, coupled with its high accuracy, multielement and speciation capabilities, modest cost, and suitability to onsite and online measurements, makes it one of the most powerful techniques for trace metal analysis. This article covers the main aspects of ASV, including past and recent developments of working electrodes, analytical methodologies, instrumentation, and applications in trace metal analysis.

## Principles of ASV

In ASV, the metals are deposited by electrolytic reduction (at a controlled negative potential), and are



**Figure 5** Voltammogram relative to the reduction (or oxidation) of the analyte (2) in the presence of another species (1) more easily reduced (or oxidized) when a little difference exists between  $(E_p)_2$  and  $(E_p)_1$ .

in fact additional resolving power. Thus, for instance, peaks separated incompletely in time can be resolved voltammetrically.

See also: **Voltammetry**: Overview; Anodic Stripping; Cathodic Stripping; Inorganic Compounds; Organic Compounds.

## Further Reading

- Bard AJ and Faulkner LR (2000) *Electrochemical Methods; Principles and Applications*, 2nd edn. New York: Wiley.
- Bond AM (1980) *Modern Polarographic Methods in Analytical Chemistry*. New York: Dekker.
- Bontempelli G, Magno F, Mazzocchin GA, and Seeber R (1989) Linear sweep and cyclic voltammetry. *Annali di Chimica (Rome)* 79: 103–216.
- Brett CMA and Oliveira Brett AM (1998) *Electrochemistry; Principles, Methods and Applications*. Oxford: Oxford University Press.
- Brown ER and Sandifer JR (1986) Cyclic voltammetry, AC polarography and related techniques. In: Rossiter BW and Hamilton JF (eds.) *Physical Methods of Chemistry*, 2nd edn. p, pp. 273–432. New York: Wiley.
- Greef R, Peat R, Peter LM, Pletcher D, and Robinson J (1985) *Instrumental Methods in Electrochemistry*. Chichester: Horwood.
- Kissinger PT and Heineman WR (1996) *Laboratory Techniques in Electroanalytical Chemistry*, 2nd edn. Washington: Division of Analytical Chemistry, American Chemical Society.
- Speiser B (2003) Linear sweep and cyclic voltammetry. In: Bard AJ, Stratmann M, and Unwin PR (eds.) *Encyclopedia of Electrochemistry*, vol. 3, pp. 81–124. Weinheim: Wiley-VCH.
- Wang J (1994) *Analytical Electrochemistry*. New York: VCH.
- Wightman RM and Wipf DO (1989) Voltammetry at ultra-microelectrodes. In Bard AJ (ed.) *Electroanalytical Chemistry*, vol. 15, pp. 267–353. New York: Dekker.

## Anodic Stripping

**S Daniele**, University of Venice, Venice, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by J Wang, pp. 5422–5425, © 1995, Elsevier Ltd.

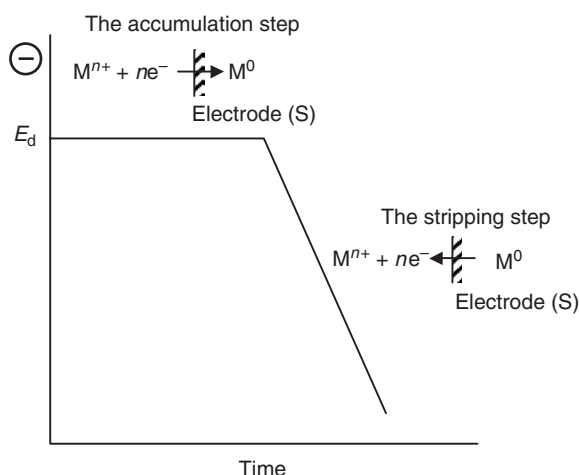
## Introduction

Stripping voltammetry (SV) comprises one of the major families of electroanalytical chemistry. The strength of SV is in its extremely low detection limits ( $10^{-10}$ – $10^{-11}$  mol l $^{-1}$ ). Its remarkable sensitivity is achieved through a preconcentration step in which the target analyte (commonly a metal ion) is accumulated onto working electrodes. This step is followed by the measurement (stripping) step, in which the preconcentrated analyte is stripped back to the solution, a process resulting in a current response

proportional to the analyte concentration in the sample. The different versions of SV depend primarily upon the nature and direction of the preconcentration and measurement steps. Anodic stripping voltammetry (ASV) is the most widely used form of stripping analysis. The extremely low detection of ASV, coupled with its high accuracy, multielement and speciation capabilities, modest cost, and suitability to onsite and online measurements, makes it one of the most powerful techniques for trace metal analysis. This article covers the main aspects of ASV, including past and recent developments of working electrodes, analytical methodologies, instrumentation, and applications in trace metal analysis.

## Principles of ASV

In ASV, the metals are deposited by electrolytic reduction (at a controlled negative potential), and are



**Figure 1** Steps in anodic stripping voltammetry of a metal ion with the reduction–accumulation process at constant potential, and the oxidation of the metal deposited on the electrode.

subsequently being measured during an anodic (positive) potential scan (**Figure 1**)

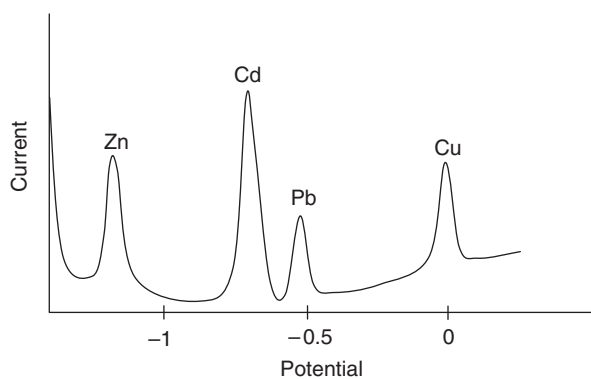


(S is the substrate electrode). The aggregative state of the concentrate on the electrode may be an amalgam or a solid phase. This depends on either the type of the working electrode employed, or the target element investigated.

ASV can be used for trace measurements of more than 20 elements that can be readily reduced to the metallic state and reoxidized. Most successful are ASV measurements of amalgam-forming elements, such as lead, cadmium, copper, zinc, thallium, bismuth, tin, or antimony, at a mercury working electrode. The amalgam-forming preconcentration step is given by



This illustrates that the deposited metal dissolves in the mercury electrode. Other elements, such as selenium, arsenic, mercury, silver, gold, or tellurium, which either do not form amalgam or oxidize at potentials anodic to that of mercury can be measured by ASV using a variety of solid electrodes. The deposition step is usually carried out under conditions of forced convection to facilitate the transport of the metal ions to the electrode surface. The convective transport is achieved mainly by electrode rotation, solution stirring, and flow. The deposition potential ( $E_d$  in **Figure 1**) should be  $\sim 0.3$ – $0.4$  V more negative than the reduction potential of the metal ion to ensure an efficient deposition process. The duration of the preconcentration step is selected according to the



**Figure 2** Typical stripping voltammogram recorded at an HDME for a mixture containing zinc, cadmium, lead, and copper, each at  $1 \times 10^{-8} \text{ mol l}^{-1}$ .

concentration level of the metal ions in question, from less than 1 min at the  $10^{-7} \text{ mol l}^{-1}$  level to  $\sim 5$ – $10$  min for  $10^{-9}$ – $10^{-10} \text{ mol l}^{-1}$  concentrations. Under such conditions, only a small fraction of the metal is actually being deposited, and therefore ASV is practically nondestructive of the sample investigated. The measurement step is carried out in a quiescent solution, by applying an appropriate potential scan. During the sweep, the amalgamated metals, or metal solid phases, are reoxidized to give rise to anodic peaks. The resultant current–potential stripping voltammogram (**Figure 2**) provides both the qualitative identification (from the peak potential) and the quantitative information (from the peak height, or peak area). The latter measurements are usually performed by using classical calibration procedures. However, under given conditions, calibrationless approaches may also be employed.

Dissolved oxygen, which is electrolytically reduced at the mercury or solid electrodes to yield large background currents, is, in general, removed from the solution by purging high-purity nitrogen through the solution for  $\sim 5$ – $10$  min prior to the preconcentration step. A supporting electrolyte (usually a salt or acid over the concentration range  $0.1$ – $0.5 \text{ mol l}^{-1}$ ) is added to the solution to minimize resistance and migration effects. Numerous voltammetric techniques can be used for the stripping step. Linear sweep voltammetry is an effective means and requires only simple instrumentation. Pulse-voltammetric waveforms such as differential pulse and, particularly square wave, are especially useful for the stripping step as they effectively correct for background current contributions. In addition, square wave voltammetry allows minimizing the interference due to dissolved oxygen, and its use therefore eliminates the need for the time-consuming sample de-aeration.

Like most analytical techniques, ASV is subject to various interferences. The major interferences include the formation of intermetallic compounds (e.g., copper–zinc), overlapping peaks (e.g., tin–lead, copper–bismuth), or surface blockage by surfactants. These can all be minimized or eliminated if adequate attention is given to certain key operations. For instance, using mercury electrodes, the formation of intermetallic compounds can be minimized by: reducing the amount of metals within the electrodes; addition in the sample solution of a third element that forms a more stable intermetallic compound with one component of the binary system (e.g., the copper–zinc problem is circumvented by adding an excess of  $\text{Ga}^{3+}$  ions); selective plating of the metal ions. The overlapping peaks problem, in some cases, can be addressed by using in the sample solution a suitable complexing agent that shifts the stripping peak potentials according to the stability of the complexes formed. Alternatively, the medium exchange procedure can also be employed. The latter involves the deposition of metals from the sample medium, followed by stripping them in a different solution containing a more suitable electrolyte. This approach improves the selectivity in the stripping step and may also minimize interference caused by various sample components. The direct determination of heavy metals in real samples is often complicated by a variety of organic compounds that are present in different forms. Organic matter may adsorb on the electrode surface, giving rise to unpredictable effects on the ASV responses. In general, removal or destruction of the surface-active substances is necessary before the ASV measurements are carried out. This is usually accomplished by ultraviolet irradiation and hydrogen peroxide or ozone oxidation procedures.

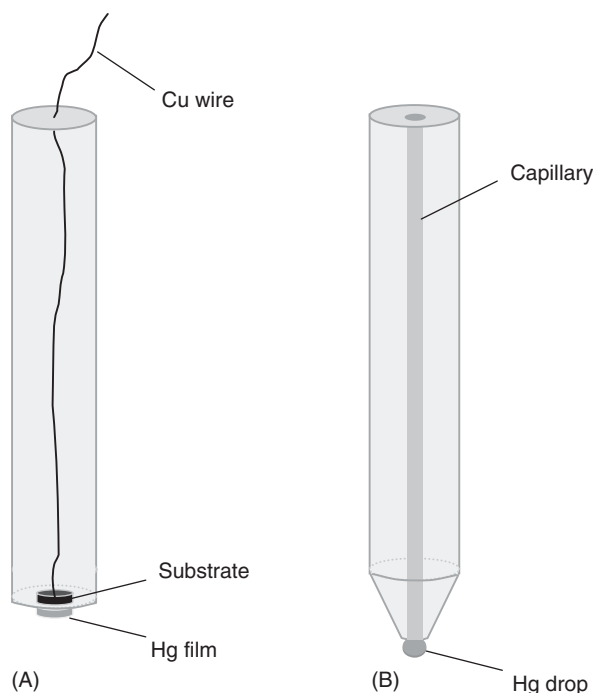
In addition to the above considerations, the analysts should be aware that they are working with extremely dilute solutions and should thus give proper attention for the minimization of contamination risks, adsorption losses, and other alteration during the sampling, storage, and measurement. The reagents used for the preparation of the supporting electrolyte and metal standards should be of the highest purity possible.

## ASV Working Electrodes

Since the advent of ASV, a great effort has been made in the development of suitable working electrodes able to widen the number of elements that can be analyzed, improve sensitivity, accuracy, and reproducibility, lower the detection limits, make the measurements as easier and rapid as possible, and minimizes interference from sample matrices.

The most practical electrodes for ASV are the hanging mercury drop electrode (HMDE) (formed at the tip of a glass capillary), especially in its static mercury drop electrode (SMDE) version, and the thin mercury film (TMF) electrode (Figure 3).

The HMDE consists of a small mercury drop with radius usually not exceeding 1 mm, which hangs by thread in a glassy capillary. The SMDE version utilizes a method of drop formation in which the mercury drop is dispensed rapidly by a solenoid-activated valve, and then allowed to hang stationary at the tip capillary. The drop size in this case can vary over three dimensions and are described as small, medium, and large. TMF is prepared by coating a suitable substrate with a film of metallic mercury. HMDE and TMF are somehow complementary. HMDE has the capability of renewing the electrode surface and therefore minimizes memory effects. However, it is bulky and requires a mercury reservoir and regular maintenance of the capillary. Because of its larger area-to-volume ratio, the mercury film electrode offers higher preconcentration efficiency (i.e., higher sensitivity and lower detection limit) and sharper peaks. In TMF, the mercury film is usually plated electrochemically from solutions containing mercury ions by reduction them under convective transport and negative polarization of the substrate. The mercury film can be preplated before the analysis (*ex situ* plating) or prepared *in situ* (by adding, typically,  $10^{-4}$ – $10^{-6}$  mol l $^{-1}$  mercury ions to the sample



**Figure 3** Schematic of: a hanging drop mercury electrode (A) and a thin mercury film electrode (B).

solution). By far the commonest substrate material for mercury film is carbon due to its chemical inertness and broad useful potential range. Glassy carbon, a vitreous form of isotropic carbon, is the most useful support for the mercury film. Noble metals, or films of noble metals on carbon, such as platinum, gold, silver, and iridium, have also been used as substrates. Deposition of mercury on these substrates does not result in a true homogeneous mercury film, since these metals dissolve to a different extent in mercury, forming metal-amalgam electrodes. This may enhance the interaction of the analyte with the metal substrate, causing interference due to the formation of intermetallic compounds and difficulty in the interpretation of peaks. A more recent method of mercury plating relies on the modification of the electrode surface with a mercury compound (e.g.,  $\text{HgO}$  or  $\text{Hg}^{2+}$  incorporated in Nafion) that, as in the *in situ* plating procedure, produces metallic mercury during the deposition step. Although less used, an alternative to electrochemical plating of mercury onto metal substrates is simply to dip amalgamating noble metals electrodes directly in pure mercury.

The substrate for mercury deposition can be either of millimeters (the so-called macroelectrodes), or micrometers in dimension. The microelectrodes resulting in the latter case have gained wide acceptance in ASV in the last two decades, owing to their favorable analytical properties. In particular, the enhanced mass transfer obviates the need for convective transport during the deposition step and for a rest period before stripping, while makes ASV responses less affected by convective forces in flowing systems. The minimal ohmic drop allows measurements to be carried out in low conducting media without deliberate addition of supporting electrolyte. This largely avoids contamination from external chemicals when ultratrace element analysis has to be performed, while existing chemical equilibria in real samples are not altered. Moreover, due to their small dimensions microelectrodes are suited for assays of extremely small sample volumes. Microelectrode arrays of carbon, iridium, platinum, and gold, fabricated by lithographic processes, offer a further advantage over individual microelectrodes, when ASV assays of very low concentration levels of metal ions have to be performed. The very low current associated to a single microelectrode is amplified, even by several orders of magnitude, at a microelectrode array. In this way, instrumental inability of detection of very low currents is overcome. The developments in the technology of screen-printed electrodes have also impacted on ASV measurements. The screen-printing technology offers large-scale mass production of extremely inexpensive and yet highly reproducible

electrode strips. Such strips rely on planar working and reference electrodes on a plastic or silicon substrate, thus making a self-containing, minuscule electrochemical cell, requiring only sample droplets for ASV measurements. Mercury-coated screen-printed electrodes have been developed for use as low-cost disposable metal-sensing devices. Additional improvements have also been achieved by designing screen-printed arrays of microdisks, the radius of each disk a few micrometers.

A major source for ASV interference arises from adsorption of surface-active compounds on the electrode surface and its consequent fouling. The result is an erratic behavior of the electrode. Coating of the mercury film with a permselective polymeric thin film or a membrane may be useful to address surfactant passivation effects. The modified surface allows analyte species to reach the electrode while excluding any interfering species present in the sample. On the other hand, chemical modification of the mercury film may enhance, for instance, by ion exchange, the selectivity and sensitivity of the ASV analysis toward the analyte of interest. A variety of protective coatings have been employed that range from simple cellulose acetate and Nafion membranes to more complicated polymers. Very attractive electrode systems, which allow ASV measurements of trace metals in real samples, are based on agarose-gel-covered mercury-coated iridium microelectrodes and the corresponding array version. In these cases a mercury microelectrode (or an array) is covered with a gel layer much thicker than the electrode radius, but thin enough to allow quick equilibration with the sample solution. The gel membrane prevents the microelectrode from being fouled by colloids and biopolymers, usually encountered in real samples, and protects the electrode from variable hydrodynamic occurring in the sample investigated. An advance of the array version allows addressing each electrode individually, so that quasi-simultaneous ASV measurements over all the microelectrodes can be performed. This is very useful for real-time, high-spatial resolution concentration profile measurements at interfaces.

In recent years, there has been a growing concern about the use of pure mercury as an electrode material in ASV measurements, owing to its toxicity. Intensive research efforts have been devoted to the development of alternative stripping electrodes, with performance approaching that based on mercury ones. To this purpose different bare solid electrodes of carbon, gold, platinum, silver, and iridium, either of conventional or microscopic dimensions (including microelectrode arrays), have been used directly as substrates for the detection of trace levels of heavy metals, such as copper, lead, cadmium, and zinc.



These mercury-free electrodes, under particular conditions, provide reliable quantification of trace elements. However, multielement analysis is usually complicated by intermetallic interference, multiple ASV peaks, large background contribution due to the relatively low cathodic potential limit (caused by reduction of hydrogen ions, or water).

Bismuth film electrodes (BFEs) have also been introduced as a potentially interesting alternative to TMF. BFE consists of a thin metallic bismuth (rather than mercury) film on substrate materials. Also in this case metallic bismuth is electrodeposited from solutions containing bismuth ions and the substrate of choice is carbon or carbon fibers. Although the bismuth film can be prepared by either *ex situ* or *in situ* procedure, the latter provide better results in terms of reproducibility and detection limits (e.g.,  $2 \times 10^{-9} \text{ mol l}^{-1}$  lead has been detected following 10-min deposition). Although the toxicity of bismuth, as compared to that of mercury, is negligible, which makes these electrodes attractive for routine applications in environmental trace metals analysis, the rather negative oxidation potential of metallic bismuth (less negative of mercury in the absence of strong ligands) poses serious limitations in the anodic range.

Alternative electrode materials, which also address the environmental concerns related to the use and disposal of either liquid mercury or dilute amalgams, are solid metal amalgams such as silver and dental amalgams. These materials are not toxic, cheap, easy to manufacture, and are very stable over a reasonable time (up to several weeks). Dental amalgam electrodes have been found to act similarly to silver electrodes, but with higher hydrogen evolution overvoltage. This makes possible, in addition to other metals like cadmium, lead, copper, thallium, and bismuth, the detection of elements such as zinc, whose reduction step needs to be carried out at very negative potentials. The sensitivity and detection limits of these electrodes compare well with those typical for mercury based electrodes.

In addition to the above types of electrodes, a variety of new and more exotic materials such as boron-doped diamond electrodes, composite electrodes, carbon-film resistors rotograved carbon, with performance comparable with mercury-based electrodes have also been proposed and applied for ASV measurements of trace zinc, cadmium, and lead.

## Anodic Stripping Voltammetry in Flow Systems and Hyphenated Techniques

ASV has been adapted with various flow systems, especially flow injection analysis, sequential injection

analysis, and batch injection analysis. These methods have significant advantages over their batch versions in trace analysis of metal ions, including high sample throughput and simple automated operations, reproducibility, low reagent and sample consumption, less risk of contamination. In addition, operations such as medium exchange or plating of the mercury film online can be accomplished straightforwardly.

To enhance the sensitivity and increase tolerance to surface-active compounds in ASV experiments, different physical phenomena such as ultra- and low-frequency sound, microwave, or controlled heating of the electrode have also been combined with voltammetry. This has provided convenient ASV methodologies for the determination of trace elements in a variety of matrices. The beneficial effects of the physical phenomena rely on the enhanced mass transport during the preconcentration step, and an effective cleaning and activation of the electrode surface (e.g., through cavitation collapse or thermal gradients at the solid-liquid interface) throughout the experiments. By these approaches, ASV of trace lead, copper, and cadmium in highly passivating media have been carried out.

To address problems related to interference due to either organic or inorganic species and to lower detection limits, ASV has also been coupled with other nonelectrochemical analytical techniques. Combination of square-wave ASV and thin-layer chromatography (TLC) has allowed to achieve detection limits down to, respectively,  $5 \times 10^{-11}$ ,  $1.7 \times 10^{-10}$ , and  $7.2 \times 10^{-11} \text{ mol l}^{-1}$  for cadmium, copper, and lead, by performing measurements directly on the TLC plate with minimum pretreatment of the sample. Combination of ASV with atomic spectrometry or inductively coupled plasma-atomic emission spectrometry, inductively coupled plasma-mass spectrometry is capable of analyzing heavy metals present at ultratrace levels (e.g., down to  $1.4 \times 10^{-12} \text{ mol l}^{-1}$  for cadmium and  $1.4 \times 10^{-14} \text{ mol l}^{-1}$  for thallium, following 340 s preconcentration time). These hyphenated techniques represent effective strategies for online operation and convenient medium exchange procedures to eliminate matrix interferences.

## Equipment for ASV Measurements

The equipment requirements for ASV are simple, relatively inexpensive, and commercially available. A typical apparatus includes a three-electrode (or a two electrode, in conjunction with microelectrodes) cell, the voltammetric analyzer, and an X-Y recorder. Reliable and versatile (computerized) instruments,



utilizing various modulation waveforms, are nowadays available from different sources. They are associated with powerful software that allows either generating the various voltammetric waveforms or data acquisition and treatments.

The portable nature and low power demands of electrochemical analyzers satisfy many requirements for onsite and *in situ* measurements. Therefore, a variety of field-based heavy metal analyzers has been developed. They combine advances in sensor technology and hand-held battery powered electrochemical analyzers. Submersible package systems that include several functions such as sampling and mixing the sample with reagents (lab-on-cable) have also been developed for *in situ* natural water analysis. Current efforts focus on the development of miniaturized stripping analytical systems produced by micromachining technology that allows the fabrication of complete laboratory systems on a chip platform (lab-on-chip).

## Applications

The remarkable sensitivity, high accuracy, versatility, and low cost of ASV have led to its application in a large number of analytical problems. The technique has been widely used for the measurements of trace metals in natural waters (including coastal and open ocean or inland waters). The absence of interference from alkali metals makes the technique extremely powerful for analysis of seawaters. Actually, the excess salt fulfills the function of the supporting electrolyte. ASV is especially suited for the study of chemical speciation of metal ions (i.e., determination of different physical-chemical forms of the element), and for their continuous *in situ* monitoring (including ship-board operations). By coupling microelectrode (or microelectrode arrays) with rapid square-wave voltammetry heavy metal ions can be determined at different levels in the water column or at the sediment water interface. Moreover, reliable measurements of trace metals can also be performed directly in low ionic strength aquatic systems such as river and lake waters, rain, and snow. Measurements of complexation capacity, oxidation states, or stability constants represent typical stripping speciation experiments. Other environmental applications include analysis of sediments, rocks, soils, fly ash, and airborne particulate matter. Recent applications are concerned with monitoring of the diffusion processes of heavy metal ions at liquid-liquid and solid-liquid interfaces with steep concentration gradients.

ASV is also widely used for trace metal analysis in body fluids (blood, urine, etc.) and tissues, including

highly reliable blood lead measurements. It is highly suitable for many industrial applications, including assays of food and drinks, gasoline, paint, alloys, or plating baths. High sample throughput, desired for metal ions industrial quality control applications, can be achieved by coupling ASV with high-speed flow injection systems. The use of portable, battery-operated instruments, disposable and microelectrodes, and versatile flow-through stripping systems have greatly facilitated decentralized (onsite) applications of ASV in various fields.

**See also:** **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Flow Injection Analysis:** Principles; Instrumentation; Detection Techniques; Environmental and Agricultural Applications; Clinical and Pharmaceutical Applications; Industrial Applications. **Microelectrodes. Sensors:** Chemically Modified Electrodes. **Thin-Layer Chromatography:** Overview. **Water Analysis:** Freshwater; Seawater – Inorganic Compounds.

## Further Reading

- Achterberg EP and Braungardt C (1999) Stripping voltammetry for the determination of trace metal speciation and *in-situ* measurements of trace metal distributions in marine waters. *Analytica Chimica Acta* 400: 381–397.
- Ashley K (1994) Electroanalytical applications in occupational and environmental health. *Electroanalysis* 6: 805–820.
- Baldo MA, Bragato C, and Daniele S (1997) Determination of lead and copper in wine by anodic stripping voltammetry with mercury microelectrodes: assessment of the influence of sample pretreatment procedures. *Analyst* 122: 1–5.
- Buffle J, Wilkinson J, Tercier M, and Parthasarathy N (1997) *In situ* monitoring of trace metals in natural waters. *Annali di Chimica (Rome)* 87: 67–82.
- Economou A and Fielden PR (2003) Mercury film electrodes: developments, trends and potentialities for electroanalysis. *Analyst* 128: 205–212.
- Ivaska A and Kubiak WW (1997) Application of sequential injection analysis to anodic stripping voltammetry. *Talanta* 44: 713–723.
- Mikkelsen O and Schorder KH (2003) Amalgam electrodes for electroanalysis. *Electroanalysis* 15: 679–687.
- Petrovic SC, King DF, and Dewald HD (1998) Electrochemical detection in thin-layer chromatography (TLC). A review on application of direct on-plate square-wave anodic stripping voltammetry for TLC. *Electroanalysis* 6: 393–398.
- Stripping Analysis* (special issue) (1998) *Electroanalysis* 10 (6).
- Tur'yan YI (1997) Microcelles for voltammetry and stripping voltammetry. *Talanta* 44: 1–13.
- Vydra F, Stulik K, and Julakova E (1976) *Electrochemical Stripping Analysis*. New York: Wiley.

- Wang J (1985) *Stripping Analysis: Principles, Instrumentation and Applications*. New York: VCH Publishers.
- Wang J (2002) Real time electrochemical monitoring: towards green analytical chemistry. *Accounts of Chemical Research* 35: 811–816.

- Zhou F (1996) Stripping analysis combined on-line with inductively coupled plasma-atomic emission spectrometry and with inductively coupled plasma-mass spectrometry. *Electroanalysis* 8: 855–861.

## Cathodic Stripping

**E P Achterberg, J L Barriada, and C B Braungardt,**  
University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by J Wang, pp. 5422–5425, © 1995, Elsevier Ltd.

### Introduction

Voltammetry is based on the measurement of a current response as a function of the potential applied to a voltammetric cell. Stripping voltammetry (SV) forms a subdivision of voltammetry and constitutes one of the most important groups of techniques in electroanalytical chemistry. The major advantage of SV is its extremely low limits of detection ( $10^{-10}$ – $10^{-12}$  mol l<sup>-1</sup>), which are the result of a preconcentration step in which the analyte is accumulated onto or into the working electrode. In the subsequent stripping step, the preconcentrated analyte is stripped back into the solution, whereby the observed current response is proportional to the analyte concentration in the sample. In addition to the excellent sensitivity of SV, this group of techniques also offers multielement and speciation capabilities, is suitable for automated on-line and *in situ* measurements, and utilizes low-cost instrumentation. Anodic stripping voltammetry (ASV) and cathodic stripping voltammetry (CSV) are the two different versions of SV, and differ in the nature of their preconcentration and stripping steps. In ASV, a cathodic preconcentration step is undertaken which is followed by a potential scan toward more positive potentials and the anodic (oxidative) currents are determined. Whereas in CSV, an anodic preconcentration step is undertaken, which is followed by a potential scan towards more negative potentials and the cathodic (reductive) currents are determined. This article discusses the technique of CSV and its applications in natural water analyses.

### Principles of Cathodic Stripping Voltammetry

Cathodic stripping voltammetry is a widely utilized group of electrochemical techniques for the

quantitation of trace and ultratrace analytes in complex environmental, clinical, and industrial samples. The excellent sensitivity resulting from the preconcentration step allows the determination of analytes in very dilute samples. The electrochemical preconcentration process in CSV may consist of anodic precipitation of an insoluble salt (conventional CSV), or adsorption of an appropriate surface-active complex of a metal (adsorptive CSV). The preconcentration requires little or no added reagent, thereby minimizing the risk of sample contamination. The preconcentration step also improves selectivity and stability of the method by isolating the analyte from the sample matrix. Various voltammetric waveforms can be employed during the stripping step, including linear sweep, differential pulse, square wave, staircase, or alternating current operations.

The basic instrumentation for CSV consists of a potentiostat, a three-electrode cell (working electrode, reference electrode, and counter-electrode), and a computer for automated measurements and data acquisition. Modern potentiostats are simple, low-cost, and able to perform a range of scan forms. The Ag/AgCl/KCl reference electrode is commonly used and the counter-electrode may be a platinum wire or a carbon rod. In CSV, it is normal to employ a stationary working electrode and the most popular electrode is the hanging mercury drop electrode (HMDE). The advantage of an HMDE is its reliability. With the formation of each new drop, a new electrode surface is produced. The drops generated by modern HMDEs are very small (e.g., VA Stand 663 from Metrohm (Switzerland) produces drops with an area of 0.52 mm<sup>2</sup>), and safe storage and recycling of the used Hg will ensure minimal environmental and health risks. The solution (typically 5–10 ml) in the voltammetric cell is gently stirred using a polytetrafluoroethylene rod or flue during the preconcentration step. The stirring aids the transfer of analyte to the HMDE surface and decreases the diffusion layer of the mercury drop. The rate of stirring should be reproducible and controlled.

- Wang J (1985) *Stripping Analysis: Principles, Instrumentation and Applications*. New York: VCH Publishers.
- Wang J (2002) Real time electrochemical monitoring: towards green analytical chemistry. *Accounts of Chemical Research* 35: 811–816.

Zhou F (1996) Stripping analysis combined on-line with inductively coupled plasma-atomic emission spectrometry and with inductively coupled plasma-mass spectrometry. *Electroanalysis* 8: 855–861.

## Cathodic Stripping

**E P Achterberg, J L Barriada, and C B Braungardt,**  
University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by J Wang, pp. 5422–5425, © 1995, Elsevier Ltd.

### Introduction

Voltammetry is based on the measurement of a current response as a function of the potential applied to a voltammetric cell. Stripping voltammetry (SV) forms a subdivision of voltammetry and constitutes one of the most important groups of techniques in electroanalytical chemistry. The major advantage of SV is its extremely low limits of detection ( $10^{-10}$ – $10^{-12}$  mol l<sup>-1</sup>), which are the result of a preconcentration step in which the analyte is accumulated onto or into the working electrode. In the subsequent stripping step, the preconcentrated analyte is stripped back into the solution, whereby the observed current response is proportional to the analyte concentration in the sample. In addition to the excellent sensitivity of SV, this group of techniques also offers multielement and speciation capabilities, is suitable for automated on-line and *in situ* measurements, and utilizes low-cost instrumentation. Anodic stripping voltammetry (ASV) and cathodic stripping voltammetry (CSV) are the two different versions of SV, and differ in the nature of their preconcentration and stripping steps. In ASV, a cathodic preconcentration step is undertaken which is followed by a potential scan toward more positive potentials and the anodic (oxidative) currents are determined. Whereas in CSV, an anodic preconcentration step is undertaken, which is followed by a potential scan towards more negative potentials and the cathodic (reductive) currents are determined. This article discusses the technique of CSV and its applications in natural water analyses.

### Principles of Cathodic Stripping Voltammetry

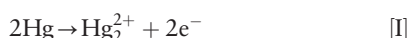
Cathodic stripping voltammetry is a widely utilized group of electrochemical techniques for the

quantitation of trace and ultratrace analytes in complex environmental, clinical, and industrial samples. The excellent sensitivity resulting from the preconcentration step allows the determination of analytes in very dilute samples. The electrochemical preconcentration process in CSV may consist of anodic precipitation of an insoluble salt (conventional CSV), or adsorption of an appropriate surface-active complex of a metal (adsorptive CSV). The preconcentration requires little or no added reagent, thereby minimizing the risk of sample contamination. The preconcentration step also improves selectivity and stability of the method by isolating the analyte from the sample matrix. Various voltammetric waveforms can be employed during the stripping step, including linear sweep, differential pulse, square wave, staircase, or alternating current operations.

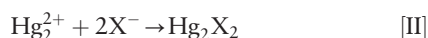
The basic instrumentation for CSV consists of a potentiostat, a three-electrode cell (working electrode, reference electrode, and counter-electrode), and a computer for automated measurements and data acquisition. Modern potentiostats are simple, low-cost, and able to perform a range of scan forms. The Ag/AgCl/KCl reference electrode is commonly used and the counter-electrode may be a platinum wire or a carbon rod. In CSV, it is normal to employ a stationary working electrode and the most popular electrode is the hanging mercury drop electrode (HMDE). The advantage of an HMDE is its reliability. With the formation of each new drop, a new electrode surface is produced. The drops generated by modern HMDEs are very small (e.g., VA Stand 663 from Metrohm (Switzerland) produces drops with an area of 0.52 mm<sup>2</sup>), and safe storage and recycling of the used Hg will ensure minimal environmental and health risks. The solution (typically 5–10 ml) in the voltammetric cell is gently stirred using a polytetrafluoroethylene rod or flue during the preconcentration step. The stirring aids the transfer of analyte to the HMDE surface and decreases the diffusion layer of the mercury drop. The rate of stirring should be reproducible and controlled.

## Conventional Cathodic Stripping Voltammetry

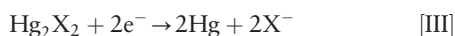
The conventional type of CSV has been used to determine a variety of inorganic and organic compounds that form an insoluble film on the electrode material during the preconcentration step. The most commonly used working electrode is the HMDE. The preconcentration step involves the application of a positive (anodic) deposition potential to the working electrode, with the formation of a sparingly soluble compound with the mercury electrode. The application of the positive potential results in the oxidation of metallic mercury to mercury(I):



and these mercury(I) ions immediately form insoluble salts with, for example, a halide ion ( $\text{X}^-$ ) on the surface of the electrode, following the reaction:



After the preconcentration step, the deposit is stripped from the electrode by application of a scan toward more negative (cathodic) potentials. In the potential scan, the mercury salt is reduced to metallic Hg and  $\text{X}^-$ :



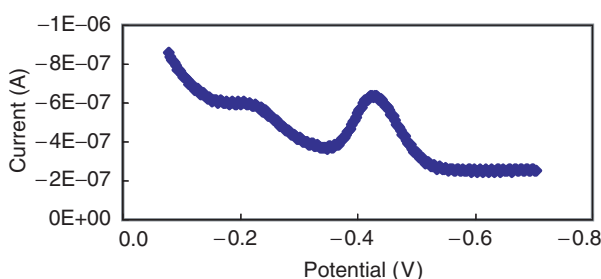
CSV has proved suitable for the determination of halides (iodide, bromide, chloride) and sulfides, and a number of organic compounds, including biological molecules (cystine, glutathione), drugs (penicillin), and pesticides (see Table 1). The organic compounds typically contain sulfur, with the consequent formation of an insoluble mercury salt during the anodic deposition step. The reduction potential of the various compounds determined by CSV is shifted toward more cathodic potentials relative to the uncomplexed mercury(I). The extent of the shift depends on the strength of the formed compound or complex.

An example of the determination of iodide in an aqueous sample using CSV is presented in Figure 1.

The preconcentration potential used was  $-0.05\text{ V}$  applied for 60 s, with the appearance of the reduction peak at a potential of  $-0.41\text{ V}$ .

Although the HMDE is the most commonly used working electrode for CSV, the applications of other types of electrode materials have been reported. For example, silver electrodes have been used for the CSV determination of halides and sulfides. In addition, the use of carbon and platinum electrodes has been reported for metal cation determinations, including iron(II), cerium(III), manganese(II), thallium(I), and lead(II). Such CSV measurements involve the precipitation of insoluble metal hydroxides on the electrode surface during the precipitation step. The drawback of the use of solid electrode materials is the poor reproducibility of the analysis and the low sensitivity, as a result of irregularities on the electrode surface and irreversible reduction of hydroxides (in the case of metals).

CSV can be seen as 'the mirror image' of ASV, but important differences can be noted between the two techniques. In ASV, the deposited metal dissolves in the mercury and forms an amalgam. In CSV, the insoluble salt deposits on the surface of the electrode and a film is formed. For this reason, the optimal electrode geometries for the techniques are different. In ASV, the maximum peak currents are achieved when the electrode volume is as low as possible (i.e., when using a mercury film electrode), because the current intensity depends on the concentration of the



**Figure 1** Scan of iodide in deionized water. Preconcentration potential  $-0.05\text{ V}$  for 60 s. HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid) was used as pH buffer (pH 7.8,  $10\text{ mmol L}^{-1}$ ) and Triton X-100 ( $20\text{ }\mu\text{L}$  of 0.2% v/v) as sensitivity enhancing surfactant. Iodide concentration was  $2\text{ nmol L}^{-1}$ .

**Table 1** List of compounds that can be determined in water by conventional CSV

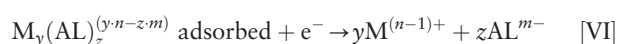
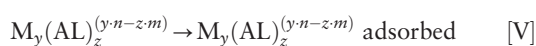
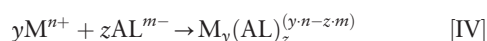
Chloride	Bromide	Iodide	Sulfide/sulfur	Sulfate
Thiols	Cyanide	Ferrocyanide	Ferricyanide	Arsenic
Selenium	Tellurium	Molybdate	Tungstate	Vanadate
Thiocyanate	Oxalate	Chromate	Succinate	Disulfides
Flavins	Flavones	Pterins	Porphyrins	Hemoglobin
Thioamides	Cysteine	Cystine	Penicillins	Iron(II)
Cerium(III)	Manganese(II)	Thallium(I)	Lead(II)	



metal collected as a mercury amalgam. In contrast, in CSV the current intensity depends on the amount of insoluble film formed and hence on the electrode area. Because of the formation of a film, calibration curves in CSV typically show a nonlinear behavior at high concentrations as a result of surface saturation. At low concentrations, curves are reproducible, and can be used for analytical purposes. In general, interfacial complications, i.e., surface saturation or interferences between analytes, and other experimental difficulties have limited the use of the CSV technique. In addition, uncertainties exist with respect to the selectivity of the CSV methods for the analyses of organic compounds.

## Adsorptive Cathodic Stripping Voltammetry

Adsorptive cathodic stripping voltammetry (ACSV) is a very sensitive technique for the analysis of numerous trace metals ( $\sim 40$ ), which cannot be determined using conventional electrolytic stripping procedures (i.e., ASV). ACSV makes use of a specific added ligand (AL), which is added to the water sample and forms an adsorptive complex with the trace metal(s) under investigation (reaction [IV]):



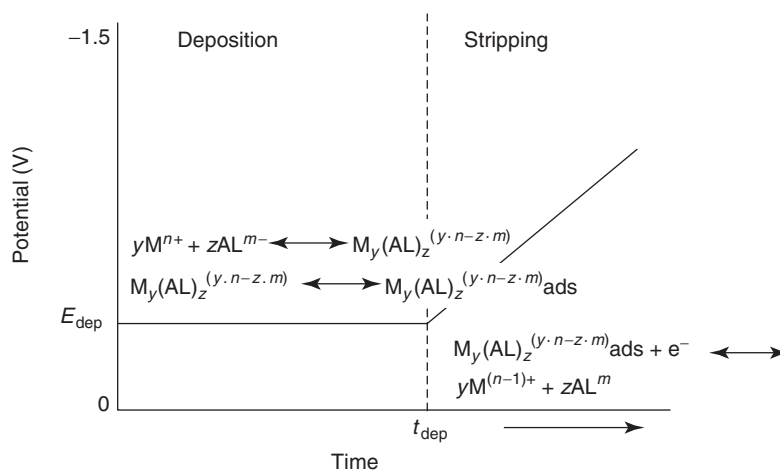
A pH buffer is used to control the pH of the sample, as the formation of the metal–AL complex is pH dependent. Indeed, many of the experimental

conditions used for the ACSV determinations of trace metals are optimized empirically, including pH, deposition potential and time, AL concentration, and waveform. Generally, ACSV is carried out using an HMDE. A minute fraction of the metal–ligand complex is adsorbed on the surface of the Hg drop (reaction [V]) and a potential scan is carried out (see Figure 2). The scan direction is toward more negative potentials and the resulting current is measured.

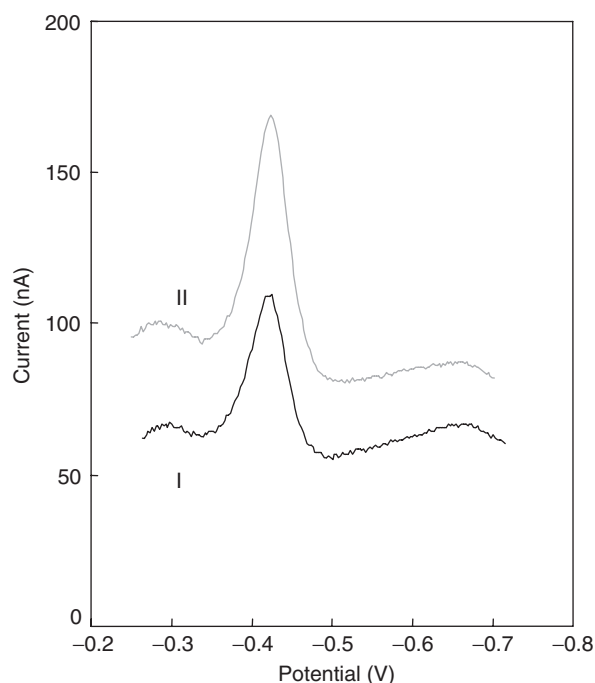
Figure 3 shows an example of the determination of dissolved Cu in seawater with the use of 8-hydroxyquinoline as AL.

The resultant current–potential stripping voltammogram provides (1) quantitative information, that is, the height of the peak is proportional to the metal concentration; and (2) qualitative information, that is, the potential of the peak is an indication for the metal analyzed. Quantification of metal concentrations in samples during voltammetric analysis is commonly undertaken by use of the standard addition method. This is the preferred method as the sensitivity of the stripping voltammetric analysis may vary between samples of different ionic strength and containing different concentrations of surfactants and natural trace metal complexing organic ligands (see ‘Interferences’ section below).

The current produced during the ACSV scan is the result of the reduction of a reducible group on the ligand or of the metal itself in the adsorbed complex (reaction [VI]). The scan forms applied during ACSV include linear sweep, but fast pulse-voltammetric waveforms (e.g., differential pulse and square wave) are also used if the reduction of the metal–AL complex is electrochemically reversible. The advantages of fast scan forms include compensation against the capacitance current contribution, a reduction of interferences from dissolved oxygen, and an improved



**Figure 2** Steps in ACSV measurements of a metal ion. The procedure involves the formation, deposition, and reduction of the surface active metal–added ligand complex.

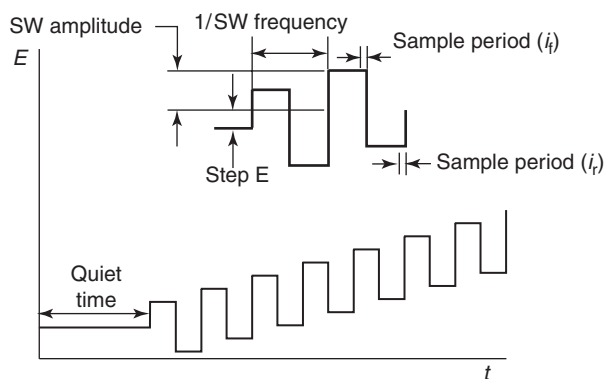


**Figure 3** Voltammetric scan of dissolved Cu in seawater. 8-Hydroxyquinoline ( $0.15 \text{ mmol l}^{-1}$ ) was used as the ACSV ligand and HEPES ( $10 \text{ mmol l}^{-1}$ ) as pH buffer (pH 7.8). Voltammetric conditions were: 20 s deposition at  $-1 \text{ V}$ , 8 s equilibration at  $-0.25 \text{ V}$  and 200 Hz square wave scan toward more negative potentials. Scan I is for a seawater sample, and scan II is for the sample plus standard addition ( $5 \text{ nmol l}^{-1}$ , final concentration).

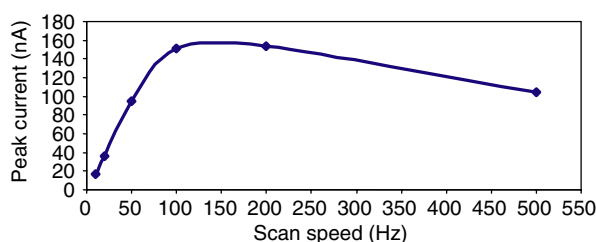
speed of analysis. A detailed schematic of the square wave form is presented in Figure 4.

The optimization of the frequency of the square wave modulation for the determination of Cu using 8-hydroxyquinoline in a low ionic strength solution is presented in Figure 5. This experiment shows that the peak current of the Cu reduction signal increases linearly up to  $\sim 100 \text{ Hz}$ , but decreases gradually at higher frequencies. The linear part of the frequency-signal graph is the result of the efficient reduction of the adsorbed metal-AL complex, which typically occurs as a monolayer on the HMDE surface. At enhanced scan frequencies, the reduction of the metal-AL complex in the monolayer becomes less efficient, and the electron and ion transfer in the low ionic strength medium becomes limiting.

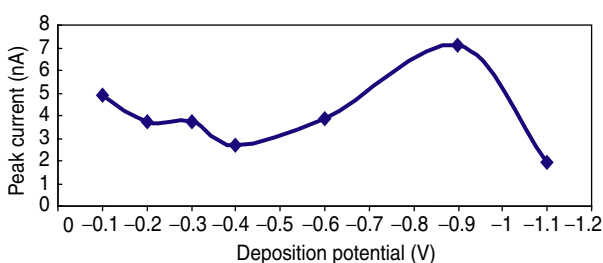
The adsorption step in ACSV is carried out at a carefully controlled potential as it determines the adsorption efficiency. In most cases, an adsorption potential is chosen which is slightly more positive ( $\sim 0.1 \text{ V}$  or more) than the reduction potential of the metal-ligand complex. Figure 6 shows the optimization of the deposition potential for the determination of Zn using ammonium pyrrolidine dithiocarbamate (APDC) as the AL. The optimal deposition potential for this method is  $-0.9 \text{ V}$ , with the reduction



**Figure 4** Detailed schematic of square wave (SW) scan form, indicating the SW frequency and amplitude, potential ( $E$ ) step and sample period.



**Figure 5** Effect of square wave scan speed on reduction peak signal for Cu in low ionic strength solution (deionized water). 8-Hydroxyquinoline ( $0.15 \text{ mmol l}^{-1}$ ) was used as the ACSV ligand, HEPES ( $10 \text{ mmol l}^{-1}$ ) as pH buffer (pH 7.8) and  $10 \text{ nmol l}^{-1}$  added Cu. Voltammetric conditions were: 60 s deposition at  $-0.2 \text{ V}$  and square wave scans toward more negative potentials.



**Figure 6** Effect of deposition potential on reduction peak signal for the determination of Zn using ACSV in deionized water. Conditions:  $0.01 \text{ mmol l}^{-1}$  APDC and  $0.01 \text{ mol l}^{-1}$  borate (pH buffer) with a deposition time of 60 s in the presence of  $10 \text{ nmol l}^{-1}$  added Zn.

potential of Zn appearing at  $-1.05 \text{ V}$  (versus Ag/AgCl/KCl reference electrode).

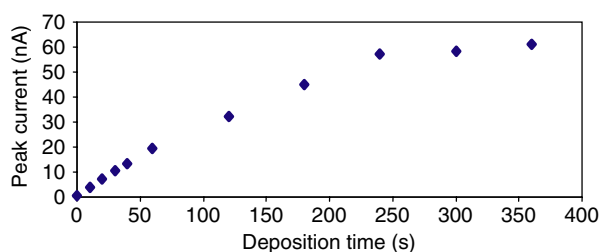
The accumulation of the metal-AL complex on the surface of the HMDE can be described using the Langmuir adsorption isotherm:

$$\Gamma = \Gamma_m \frac{BC}{1 + BC} \quad [1]$$



where  $\Gamma_m$  is the metal–AL surface concentration corresponding to the monolayer cover and  $B$  the adsorption coefficient. Consequently, plots of deposition time versus peak signal typically show a deviation from linearity at long deposition times as a result of surface coverage saturation. This is illustrated in Figure 7 for a deposition time experiment involving Co (with nioxime as AL and  $\text{NO}_2^-$  as catalyst), which shows a deviation from linearity at times  $>240$  s.

The adsorptive accumulation approach results in a very effective preconcentration with short adsorption times, allowing fast and extremely sensitive trace metal measurements. The limit of detection of ACSV for trace metals is typically on the order of  $10^{-9}$ – $10^{-11}$  mol l $^{-1}$ . Even lower metal



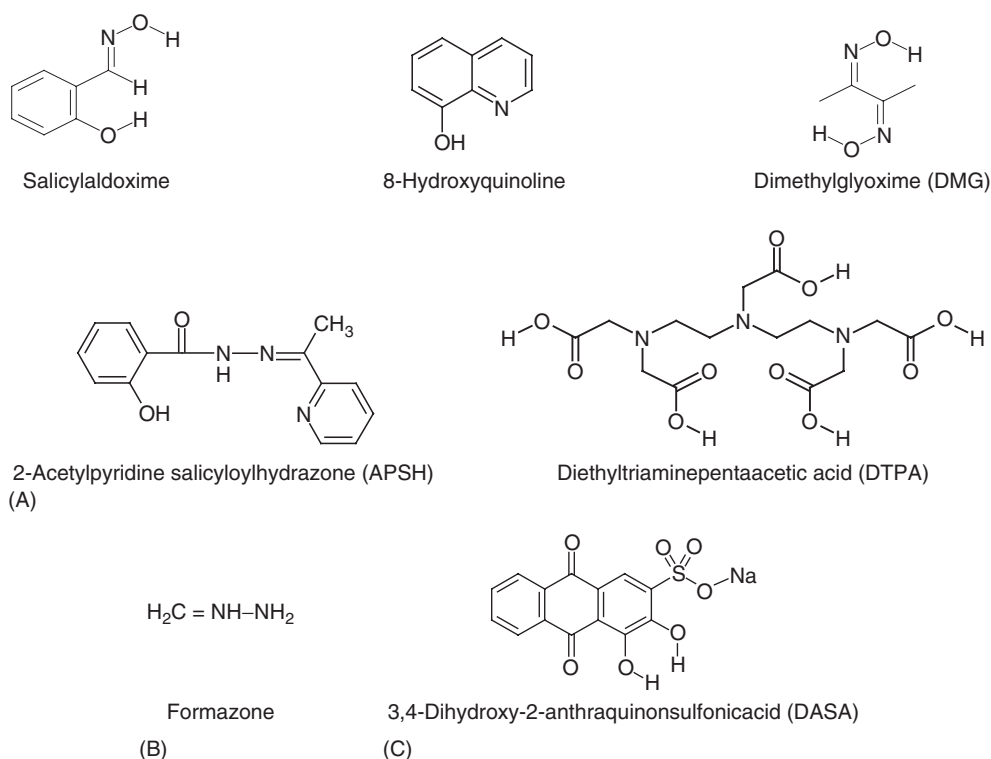
**Figure 7** Effect of deposition time on reduction peak signal for the determination of Co using ACSV in deionized water. Conditions:  $0.01 \text{ mol l}^{-1}$  nioxime and  $0.01 \text{ mol l}^{-1}$  borate (pH buffer) with a deposition time of 60 s in the presence of  $0.5 \text{ mol l}^{-1}$   $\text{NaNO}_2$  and  $1 \text{ nmol l}^{-1}$  added Co.

concentrations can be determined by enhancing the reduction current catalytically through adding an oxidant ( $10^{-12} \text{ mol l}^{-1}$ , e.g., Co ( $\text{NO}_2^-$  added), Pt ( $\text{H}^+$  added), and Ti ( $\text{ClO}_3^-$  added)). ACSV methods have been developed and applied during the last 25 years for a wide range of trace metals. As the formation of a metal amalgam is not a prerequisite for ACSV (in contrast to ASV), methods for any trace metal with a reduction potential within the stability limits of hydrogen evolution and mercury oxidation can potentially be developed. Table 2 lists ACSV methods for direct determination of trace metals in natural waters, together with their limit of detection. Table 2 is not intended as a complete list, neither does it cover all elements for which ACSV methods are available, nor does it include all reported ACSV ligands.

Figure 8 shows the chemical structures of a selected number of ACSV ligands. The organic ligands in Table 2 and Figure 8 contain N and O donor groups (e.g., DMG and 8-hydroxyquinoline), in addition to S donors (e.g., APDC). To be suitable for ACSV, ligands are required to have two properties: (1) the ability to form a complex with the element of interest, and (2) electroactivity (i.e., capability to adsorb onto the surface of the HMDE). Many of the ligands shown in Table 2 and Figure 8 have aromatic ring structures. There are however exceptions, like, for example, DMG, which forms a ring structure on chelation with

**Table 2** Trace elements and added ligands used for ACSV analysis in natural waters

Element	Added ACSV ligand	Limit of detection ( $10^{-9} \text{ mol l}^{-1}$ )
<i>Methods using reduction of element</i>		
As	Pyrrolidine dithiocarbamate	3
Cd	2-Acetylpyridine salicyloylhydrazone	0.060
Co	Dimethylglyoxime	0.1
Cr	Diethylenetriaminepentaacetic acid with nitrate as catalyst	0.05
Cu	8-Hydroxyquinoline; salicylaldehyde	0.2; 0.1
Fe	Catechol; 1-nitroso-2-naphthol	0.2; 0.1
Mo	2,5-Dichloro-3,6-dihydroxy-1,4-benzoquinone	0.2
Ni	DMG	0.1
Sb	Catechol	0.6
Se	$\text{Cu(I)}_2\text{Se}$	0.01
Sn	Tropolone	0.05
U	Mordant Blue; 8-hydroxyquinoline in presence of EDTA	1; 0.2
V	Catechol with bromate as catalyst	0.07
Zn	Ammonium pyrrolidine dithiocarbamate; 8-hydroxyquinoline	0.07; 0.5
<i>Methods using catalytic effect</i>		
Co	Nioxime with nitrite as catalyst; methyl thymol blue, with nitrite as catalyst	0.003; 0.0001
Fe	2,3-Dihydroxynaphthalene with bromate as catalyst	0.013
Mo	Mandelic acid with chlorate as catalyst; methyl thymol blue with chlorate as catalyst	0.002; 0.02
Pt	Formazone with protons as catalyst	0.0004
Ti	Mandelic acid with chlorate as catalyst	0.007
<i>Methods using reduction of ligand</i>		
Al	1,2-Dihydroxyanthraquinone-3-sulfonic acid; Solochrome violet RS	1; 50



**Figure 8** Chemical structures of a selected number of ACSV ligands: (A) reduction of the metal; (B) catalytic hydrogen formation; (C) reduction of ligand.

$\text{Ni}^{2+}$ . The adsorption of the ligands is affected by the deposition potential and it therefore appears that the presence of electrostatic and  $\pi$ -electron interactions are significant for the adsorption process.

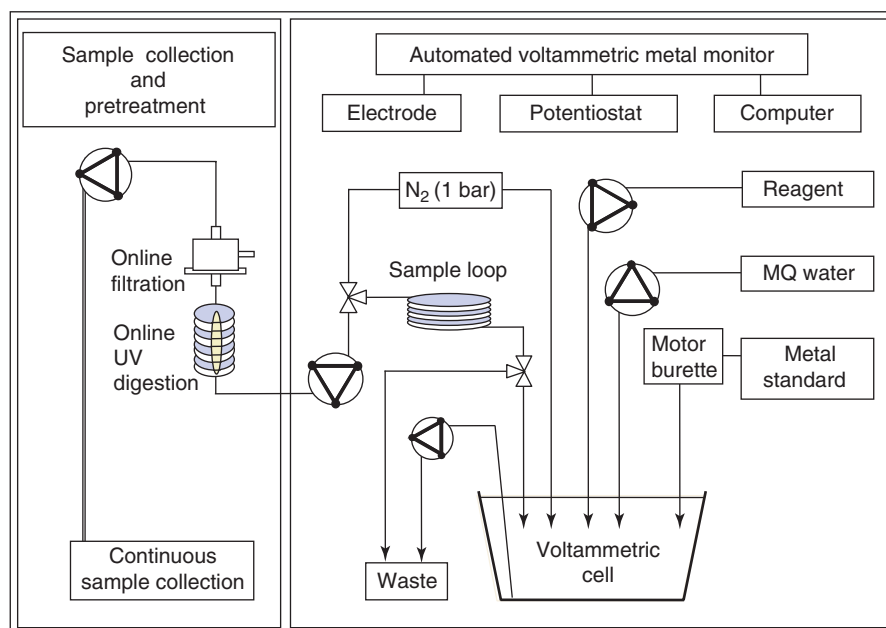
In many cases, ACSV has been utilized as a single-element method. However, a number of ligands allow the simultaneous determination of several elements. For example, catechol allows the determination of Cu, Fe, V, and U in a single measurement, with well-separated reduction peaks. Furthermore, multi-element ACSV methods for HMDE have been developed recently, whereby with the use of a mixture of ACSV ligands up to six trace metals (Cu, Pb, Cd, Ni, Co, and Zn) can be determined simultaneously in coastal and estuarine waters.

### Applications of ACSV

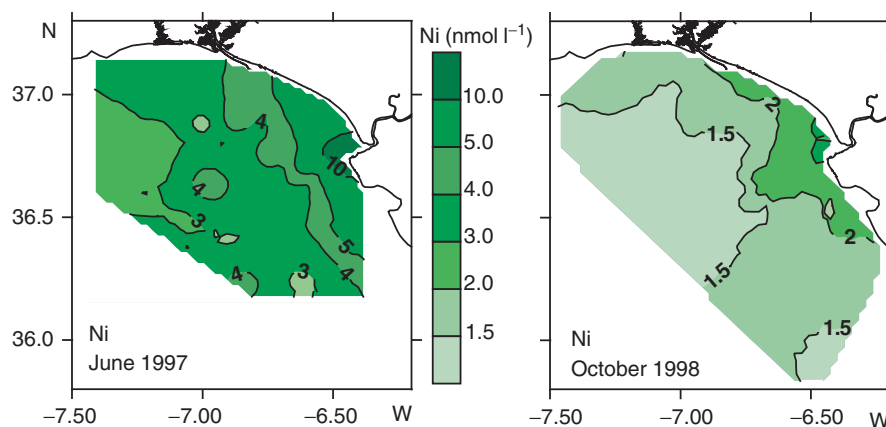
**Shipboard trace metal analyses using ACSV in coastal waters** Trace metal analyses in seawater are challenging because of the low concentrations of the analytes of interest ( $10^{-9}$ – $10^{-12}$  mol l $^{-1}$ ) and the high salt content of the medium. Stripping voltammetric techniques have a major advantage because the analyses can be undertaken directly in seawater, as the preconcentration is performed in the voltammetric cell itself. Alkali metals present in seawater do not interfere with trace metal determinations, but in

many cases actually increase the sensitivity of the voltammetric methods because of their role as electrolytes. The reduction in the sample handling minimizes the risk of sample contamination and allows automation of the instrumentation. Moreover, the portable nature of the instrumentation makes the technique suitable for field deployment. **Figure 9** presents a manifold of a voltammetric system that has been used in the field in an automated batch-mode for trace metal analyses. Aliquots of 10 ml can be analyzed at a rate of one complete measurement every  $\sim 10$ – $20$  min. Each sample is fully calibrated through automated standard additions, resulting in high-quality data required for biogeochemical and pollution studies. A more recent development involves the field deployment of online ACSV systems that use an HMDE in a flow cell arrangement. Such an approach results in very high measurement frequencies with a trace metal analysis every  $\sim 60$ – $90$  s.

The automated voltammetric systems have been used on-board ships for near real-time trace metal measurements of surface seawaters using ACSV. This approach uses underway pumping as a means of sample collection and thereby obviates the need for the vessel to halt for the collection of discrete samples. Sample contamination is prevented by eliminating contact of the seawater with metal components by



**Figure 9** Manifold of an automated voltammetric system for field analysis of trace metals. Sample pickup could be from sample changer, or pumped seawater supply.



**Figure 10** Contour plots of total dissolved Ni in the Gulf of Cadiz, determined on-board ship using automated ACSV with continuous underway sampling during surveys in June 1997 and October 1998 (concentration gradients in  $\text{nmol l}^{-1}$ ). Dissolved Ni was determined using DMG as AL and HEPES as pH buffer.

using inert materials (e.g., Teflon<sup>®</sup>, polyvinyl chloride, polyethylene). Online filtration and ultraviolet digestion is carried out continuously prior to analysis, in order to remove particles and metal complexing organic matter, respectively. **Figure 10** illustrates, with the example of Ni, the advantages of high-resolution monitoring in the coastal waters of the Gulf of Cadiz for surveys in June 1997 and October 1998. The continuous underway sampling and analysis approach resulted in an extensive coverage of the coastal area. Dissolved Ni concentrations were analyzed using a mixed reagent of AL (DMG) and pH buffer (HEPES, pH 7.8). At a ship's speed of 8 knots, and a rate of four to five measurements per hour, the spatial

resolution between samples during automated online metal analyses was between 3.3 and 4.5 km. The distribution of dissolved Ni in **Figure 10** shows enhanced metal levels in the coastal region near the outflow of estuaries. The decrease in Ni concentrations with increasing distance from the coast is the result of mixing of metal-polluted estuarine water with cleaner Atlantic waters.

**Trace metal speciation studies in natural waters using ACSV** The high sensitivity and selectivity of ACSV makes this technique very suitable for trace metal speciation studies. ACSV has consequently been used for this purpose in river, lake, ground, and sea

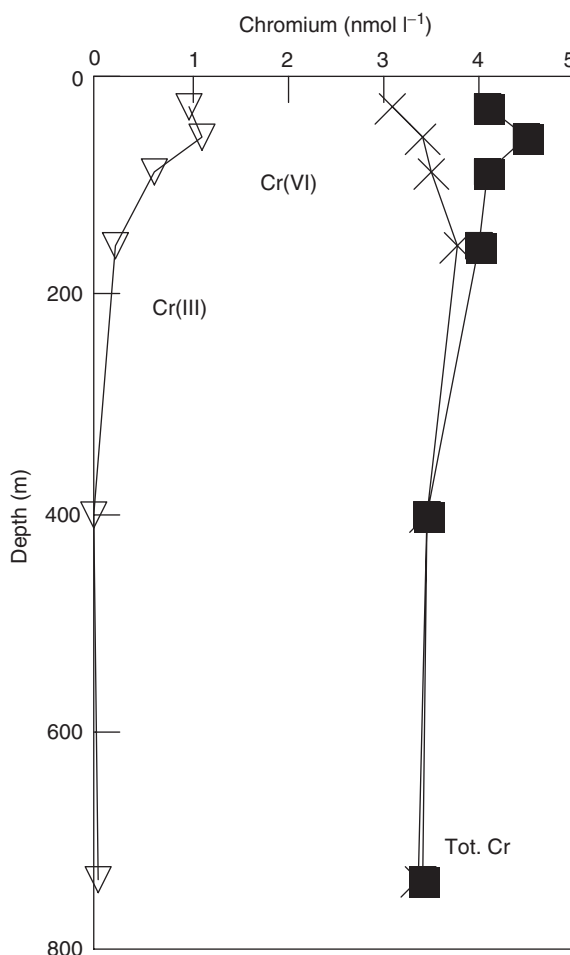
waters. Speciation analysis involves the determination of different physicochemical forms of trace metals. Metal speciation studies have become important because of the recognition that total dissolved metal concentrations do not yield sufficient information about the toxicity, bioavailability, and geochemical behavior of trace metals in natural waters.

Only a few analytical techniques (including SV and chemiluminescence) are sensitive enough to determine labile/free aqueous metal fractions in natural waters. A great number of trace metal speciation studies have been reported for natural waters utilizing ACSV during the last decade. ACSV has been used to investigate (I) redox speciation, and (II) concentration of naturally present metal complexing organic ligands.

(I) ACSV has been used to determine the different redox species of As, Fe, and Cr in natural waters. In the case of Cr, the thermodynamically stable form of Cr in oxygenated seawater is Cr(VI) (as  $\text{CrO}_4^{2-}$ ). However, significant amounts of Cr(III) have been found in natural oxygenated waters. **Figure 11** shows depth profiles of dissolved Cr species in the Mediterranean Sea. Total dissolved Cr concentrations ranged between  $3\text{--}5\text{ nmol l}^{-1}$ , with somewhat lower Cr(VI) concentrations. Dissolved Cr(III) concentrations (determined as the difference between total dissolved Cr and Cr(VI)) ranged between  $0.2$  and  $1\text{ nmol l}^{-1}$ , with maximum concentrations in the surface layer. Photochemical and biochemical reduction of Cr(VI) have been hypothesized to be responsible for the production of Cr(III) in seawater.

(II) Metal titrations are used to determine the concentration of natural metal-complexing ligands (L) and their conditional stability constants ( $K'_{\text{ML}}$ ), utilizing a ligand competition approach. For this purpose, a seawater sample is divided into typically 10 subsamples, to which increasing amounts of metal are added. After addition of the ACSV ligand and a pH buffer, an equilibration period (typically 12 h) is applied during which the naturally present ligands compete with the added ACSV ligand in a controlled manner. Subsequently, the metal concentrations are measured in the subsamples. Linear or nonlinear data transformation allows the determination of L and  $K'_{\text{ML}}$ , in addition to the free aqueous metal concentration. The natural ligand concentration (L) and  $K'_{\text{ML}}$  provide information about the capacity of natural waters to buffer additional inputs of metals, and the binding strength of L for the metal under investigation, respectively. The free aqueous metal concentration, which is calculated from L and  $K'_{\text{ML}}$ , provides information about the possible biological effects of the metal.

**Figure 12** shows a depth profile of concentrations of natural Cu complexing ligands in the Atlantic Ocean

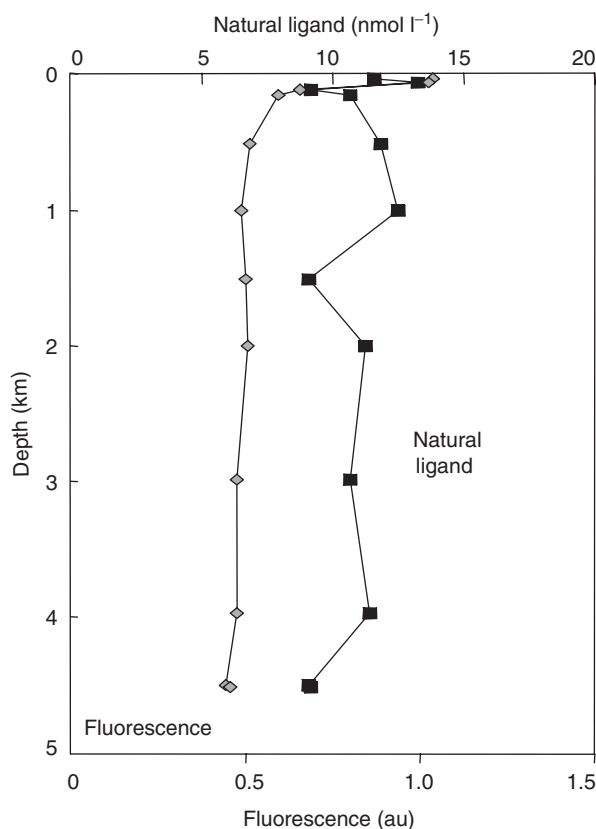


**Figure 11** Depth profiles of total dissolved Cr, Cr(VI), and Cr(III) in the Mediterranean Sea near Gibraltar ( $37.1^\circ\text{N}$ ,  $5.22^\circ\text{W}$ ), analyzed on-board ship using ACSV with DTPA as AL,  $\text{NO}_3^-$  as catalyst, and acetate as pH buffer.

determined using automated ACSV on-board ship. The depth profile of natural Cu complexing ligands showed a close similarity with *in situ* determined fluorescence, which is a measurement of the chlorophyll concentrations. The maximum ligand concentration coincided with a maximum in fluorescence, indicating that the ligands observed in the Atlantic were derived from phytoplankton. The conditional stability constants of the Cu complexing natural ligands ranged between 12 and 13 (log values) and the calculated free cupric ion concentrations  $[\text{Cu}^{2+}]$  ranged between  $7 \times 10^{-14}$  and  $3 \times 10^{-13}\text{ mol l}^{-1}$ .

### Interferences in ACSV

Dissolved oxygen occurs at high concentrations in natural waters ( $\sim 3 \times 10^{-4}\text{ mol l}^{-1}$ ) and interferes with voltammetric measurements. Therefore, oxygen interference is usually removed by deoxygenation of the sample for a period of 3–5 min using an inert gas



**Figure 12** Depth profile of the concentration of natural Cu complexing ligands (analyses on-board ship) and fluorescence in the Northeast Atlantic Ocean (48° N, 20° W).

like N<sub>2</sub>. Furthermore, many CSV methods are affected by the presence of dissolved organic matter in the samples. The organic compounds can interfere with the analyses because of their surface-active properties, and hence adsorb onto the HMDE and block the adsorption of metal-AL complexes. In addition, organic compounds may complex trace metals and out-compete the AL for the analyte of interest. UV digestion is commonly used to remove the surfactants and metal-complexing organic ligands from the sample. Both UV digestion in a batch and

on-line mode have been reported, and the lamps used are usually medium pressure Hg vapor lamps (between 50 and 1000 W). The ultraviolet digestion is to be undertaken in acidified samples, and typically H<sub>2</sub>O<sub>2</sub> is added as an oxidant to improve organic matter breakdown.

**See also:** **Voltammetry:** Overview; Anodic Stripping; Inorganic Compounds; Organic Compounds. **Water Analysis:** Freshwater; Seawater – Inorganic Compounds.

## Further Reading

- Achterberg EP and Braungardt CB (1999) Stripping voltammetry for the determination of trace metal speciation and distribution in marine waters. *Analytica Chimica Acta* 400: 381–397.
- Brett CMA and Brett AMO (1998) *Electroanalysis*. Oxford: Oxford University Press.
- Florence TM (1979) Cathodic stripping voltammetry. Part I. Determination of organic sulfur compounds, flavins and porphyrins at the sub-micromolar level. *Journal of Electroanalytical Chemistry* 97: 219–236.
- Kissinger PT and Heineman WR (1996) *Laboratory Techniques in Electroanalytical Chemistry*. New York: Dekker.
- Monk P (2001) *Fundamentals of Electro-Analytical Chemistry*. New York: Wiley.
- Plambeck JA (1982) *Electroanalytical Chemistry: Basic Principles and Applications*. New York: Wiley.
- Riley T and Watson A (1987) *Polarography and Other Voltammetric Methods*. New York: Wiley.
- Rozan TF and Taillefer M (2002) *Electrochemical Methods for the Environmental Analysis of Trace Element Biogeochemistry*. Washington: American Chemical Society.
- Van den Berg CMG (1989) Adsorptive cathodic stripping voltammetry of trace elements in sea water. *The Analyst* 114: 1527–1530.
- Wang J (2000) *Analytical Electrochemistry*, 2nd edn. New York: Wiley-VCH.
- Wang J (1985) *Stripping Analysis: Principles, Instrumentation, and Applications*. Deerfield Beach: VCH.

## Inorganic Compounds

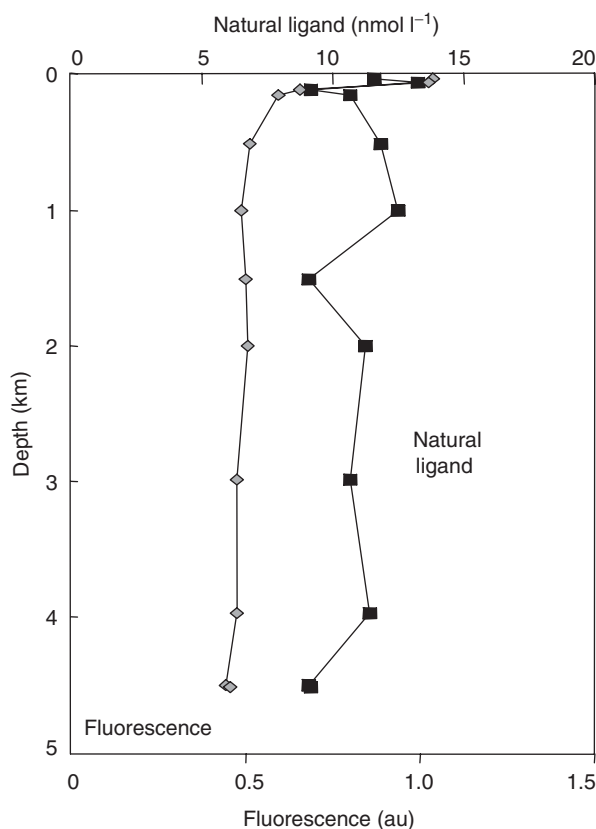
**S B Adeloju**, Monash University, Gippsland, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The analytical application of voltammetry to the determination of inorganic compounds is often achieved by the use of three common voltammetric

techniques: anodic stripping voltammetry (ASV), cathodic stripping voltammetry (CSV), and adsorptive stripping voltammetry (AdSV). The basic principles and applications of ASV and CSV have been described elsewhere. Therefore, the focus of this section is on AdSV, which has gained the most interest of all the voltammetric techniques for the determination of inorganic compounds in the past two to three decades.



**Figure 12** Depth profile of the concentration of natural Cu complexing ligands (analyses on-board ship) and fluorescence in the Northeast Atlantic Ocean (48° N, 20° W).

like N<sub>2</sub>. Furthermore, many CSV methods are affected by the presence of dissolved organic matter in the samples. The organic compounds can interfere with the analyses because of their surface-active properties, and hence adsorb onto the HMDE and block the adsorption of metal-AL complexes. In addition, organic compounds may complex trace metals and out-compete the AL for the analyte of interest. UV digestion is commonly used to remove the surfactants and metal-complexing organic ligands from the sample. Both UV digestion in a batch and

on-line mode have been reported, and the lamps used are usually medium pressure Hg vapor lamps (between 50 and 1000 W). The ultraviolet digestion is to be undertaken in acidified samples, and typically H<sub>2</sub>O<sub>2</sub> is added as an oxidant to improve organic matter breakdown.

**See also:** **Voltammetry:** Overview; Anodic Stripping; Inorganic Compounds; Organic Compounds. **Water Analysis:** Freshwater; Seawater – Inorganic Compounds.

## Further Reading

- Achterberg EP and Braungardt CB (1999) Stripping voltammetry for the determination of trace metal speciation and distribution in marine waters. *Analytica Chimica Acta* 400: 381–397.
- Brett CMA and Brett AMO (1998) *Electroanalysis*. Oxford: Oxford University Press.
- Florence TM (1979) Cathodic stripping voltammetry. Part I. Determination of organic sulfur compounds, flavins and porphyrins at the sub-micromolar level. *Journal of Electroanalytical Chemistry* 97: 219–236.
- Kissinger PT and Heineman WR (1996) *Laboratory Techniques in Electroanalytical Chemistry*. New York: Dekker.
- Monk P (2001) *Fundamentals of Electro-Analytical Chemistry*. New York: Wiley.
- Plambeck JA (1982) *Electroanalytical Chemistry: Basic Principles and Applications*. New York: Wiley.
- Riley T and Watson A (1987) *Polarography and Other Voltammetric Methods*. New York: Wiley.
- Rozan TF and Taillefer M (2002) *Electrochemical Methods for the Environmental Analysis of Trace Element Biogeochemistry*. Washington: American Chemical Society.
- Van den Berg CMG (1989) Adsorptive cathodic stripping voltammetry of trace elements in sea water. *The Analyst* 114: 1527–1530.
- Wang J (2000) *Analytical Electrochemistry*, 2nd edn. New York: Wiley-VCH.
- Wang J (1985) *Stripping Analysis: Principles, Instrumentation, and Applications*. Deerfield Beach: VCH.

## Inorganic Compounds

**S B Adeloju**, Monash University, Gippsland, VIC, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The analytical application of voltammetry to the determination of inorganic compounds is often achieved by the use of three common voltammetric

techniques: anodic stripping voltammetry (ASV), cathodic stripping voltammetry (CSV), and adsorptive stripping voltammetry (AdSV). The basic principles and applications of ASV and CSV have been described elsewhere. Therefore, the focus of this section is on AdSV, which has gained the most interest of all the voltammetric techniques for the determination of inorganic compounds in the past two to three decades.



This technique is remarkably sensitive and, in particular, permits the determination of trace and ultratrace concentrations of numerous substances, many of which cannot be determined by ASV because of their poor solubility in mercury or by CSV because of their inability to form insoluble film with the electrode material. The principles and application of AdSV for the determination of inorganic compounds in various sample materials are described below.

## Adsorptive Stripping Voltammetry

### Background

The merits in the deliberate adsorption of species onto suitable electrode for analytical voltammetric measurement were first evident from Komarek's (a Czech Chemist) work in 1947 when he found that the adsorption of nickel as nickel dimethylglyoximate resulted in substantial enhancement of its polarographic response. The concept was extended by a number of workers in the early 1980s to enable the AdSV determination of nickel and cobalt (as metal dimethylglyoximates) in the picogram to nanogram range with increased sensitivity, improved selectivity, high specificity, and better limits of detection, compared with ASV and CSV.

A major difference between AdSV and other stripping voltammetric techniques is the utilization of a spontaneous adsorption process during the preconcentration step. This involves the preconcentration of the inorganic substance, as a complex, by interfacial accumulation at an open circuit. The adsorbed complex cannot be deposited by electrolysis and significantly larger voltammetric responses are obtained for adsorbed species than for solution-species alone. This has resulted in a substantial improvement in the limits of detection for many inorganic substances by several orders of magnitude compared with the corresponding solution-phase voltammetric measurements, based on the use of ASV and CSV. This substantial improvement in sensitivity, particularly for metals, is due to the formation of metal complex(es) that possess surface-active properties and that can be subsequently reduced during the stripping step with the application of a potential scan. This is particularly beneficial for enhancement of the electrode processes of inorganic substances with extreme irreversible behavior, such as nickel and cobalt. The sensitization factor of this approach is  $\sim 100$ – $1000$  or more, resulting in improved limits of detection for the analytes. AdSV has been successfully applied to the determination of many inorganic substances in various samples, such as biological tissues, blood, foodstuffs, geological materials, groundwater,

industrial plant electrolyte, plant materials, seawater and other natural waters, sediment, sewage, soil, and urine.

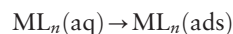
### Principles

The measurement of inorganic substances by AdSV involves three main steps, which are similar to those employed in ASV and CSV, but it is distinctly different in its preconcentration step, which does not involve electrolysis. The specific steps in AdSV are:

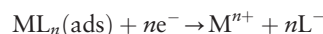
- (a) Interfacial accumulation, which involves the addition of a complexing or chelating agent to enable the formation of an adsorbable metal complex or chelate at an optimum pH:



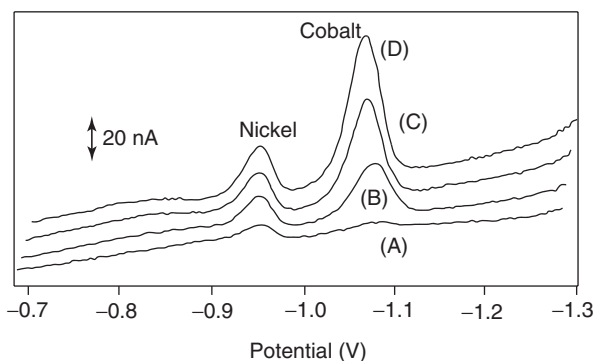
This is followed by preconcentration of the metal complex or chelate by adsorption, as follows:



- (b) Equilibration period, which is a brief rest period (usually 15–30 s) when the stirrer is turned off prior to the measurement and it is the same as employed for ASV and CSV.
- (c) Reductive stripping, which involves an application of a negative potential sweep to enable the reduction of the adsorbed metal complex or chelate, as follows:



It is important to note here that the selectivity achieved in AdSV is governed by four important chemical and electrochemical factors: (1) choice of complexing or chelating agent, (2) solution composition, (3) choice of accumulation or adsorption potential, and (4) the required reduction potential for the adsorbed complex. In practice, steps (a) and (b) described above are carried out by buffering a solution that contained the inorganic analyte(s) to an optimum pH, adding a suitable selective complexing or chelating agent, and then preconcentrating a small fraction of the dissolved complex by adsorption onto the electrode (usually a mercury electrode). A carefully controlled potential is usually employed for the adsorption step to ensure optimum adsorption of the dissolved complex. The final step (c) is intended to remove the adsorbed complex by application of a negative potential scan, which may be either a linear DC potential sweep or a potential sweep with superimposed modulations. During the potential scan, the resulting reduction current is monitored simultaneously until the reduction peak(s) for the analyte(s) is evident. The resulting peak current or peak height is directly proportional to the analyte concentration. **Figure 1** shows the responses obtained



**Figure 1** DPAdSV determination of nickel and cobalt in non-acidified swamp water with an MFE. Standard additions: (A) sample only, (B) + 100, (C) + 300 m, and (D) + 500 ng l<sup>-1</sup>.

for simultaneous determination of nickel and cobalt in a water sample by DPAdSV with a mercury film electrode and quantification is achieved by standard additions method. Evidently, the peak current for each metal increased with increasing concentration and this relationship was used to quantify the concentration of each metal in the water sample.

It is widely accepted that the measured reduction peak current in AdSV results from the reduction of either a reducible group on the complex or of the inorganic substance (e.g., a metal in a metal complex) itself in the adsorbed complex. In either case, a necessary requirement in this measurement is that the reducible component must be close to the electrode surface and, as the adsorbed complex usually occurs as a single molecular layer on the electrode, this ensures that the reduction step is efficient. In effect, all or most adsorbed complex is reduced during the negative potential scan. Another important advantage of AdSV is that any oxidation state with a reduction potential within the stability limits of mercury and hydrogen can be used. This has resulted in the successful utilization of this technique for the determination of over 40 inorganic elements at trace and ultratrace concentrations. **Tables 1–3** list some of the inorganic substances that can be determined by AdSV. A detection limit within 0.01–1 nmol l<sup>-1</sup> is readily accomplished by AdSV due to the high efficiency of the reduction step and of the adsorption of the analyte. Lower detection limits can be accomplished by coupling the very efficient adsorptive accumulation of inorganic complexes on electrode surfaces with a catalytic reaction. This approach will be discussed later in more detail.

#### AdSV Based on Complexing Agent Reduction

Some of the inorganic substances that are difficult to reduce in aqueous media can, in some cases, be

determined indirectly by the reduction of the ligand in the complex. However, unlike the reduction of the inorganic substance, it is sometimes difficult to achieve adequate peak separation of different inorganic substances. The reduction potential is often shifted toward more negative potentials than that of the uncomplexed fraction of the ligand due to the stabilization of the reducible group in the ligand. Consequently, the stability of the complex determines the peak potential of the complex, and its separation from the main ligand peak. So, the separation of the peak potentials of complexes of different inorganic substances with the same complexing agent is only possible if they have different complex stabilities.

**Table 2** lists some of the inorganic substances that have been determined based on the reduction of the ligand in the complex. In general, the achievable sensitivity and detection limit for this approach are usually not as good as those obtained with the direct reduction of the inorganic substances due to competitive co-adsorption of the ligand. Nevertheless, detection limits in the low nanomolar range are often achieved by this approach.

#### Catalytic Adsorptive Stripping Voltammetry

Many complexes of metals that can be adsorbed onto electrode surfaces have also been found to be capable of inducing catalytic reactions with electroinactive oxidizing agents or inducing catalytic hydrogen evolution. Consequently, the addition of an oxidant has been used, in some cases, to achieve significant increase in the sensitivity of AdSV by enabling chemical oxidation of reduction products generated from the electrochemical reduction step. Hence, the resulting voltammetric response obtained for such catalytic adsorption stripping voltammetry (CAdSV) is the product of both the effects of the interfacial accumulation (adsorption) and the catalytic reaction, as expressed for the associated faradaic current ( $iF$ ) as follows:

$$iF = (\text{adsorptive enrichment}) \times (\text{catalytic enhancement})$$

The combination of the enrichment and enhancement effects in CAdSV has been successfully used for reactions involving either a single or a two-electron reduction. As indicated in **Table 3**, the inclusion of oxidants, such as nitrite, nitrate, peroxide, bromate, chlorite, chlorate, and hydrogen ions, in solutions have been shown to result in: (1) considerable amplification of the voltammetric responses obtained with AdSV, (2) a reduction in detection limits, and (3) improved selectivity. This catalytic current generation has been successfully used for the AdSV determination of inorganic substances, such as Co, Cr,

**Table 1** Some of the inorganic substances that can be determined by adsorptive stripping voltammetry

<i>Inorganic element or compound</i>	<i>Complexing agent(s)</i>	<i>Working electrode(s)</i>	<i>Voltammetric mode</i>	<i>pH or electrolyte</i>	<i>Detection limit (nmol l<sup>-1</sup>)</i>
Aluminum	HNB	HMDE	LS	PIPES	1.1
Antimony	Catechol	HMDE, MFE	DP	5.8–6.8	0.6
Arsenic	Copper(II)	HMDE, MFE	DP	HCl	3
	DMTD	HMDE	DP	NaNO <sub>3</sub>	1.3
Cadmium	Oxine	HMDE, MFE	DP	7.5–8.5	0.1
Chromium	Bipyridyl, Cupferron, DTPA, TTHA	HMDE, MFE, DME	DP, SW, DPP, LS	5.2–6.2	4–50
Cobalt	BD, DMG, FD, NO, SCN <sup>-</sup> , BP	HMDE, MFE, DME	DP, SW, DPP, LS	7.4–10	0.1
Copper	Catechol, oxine, salicylaldoxime, tropolone	HMDE, MFE	DP	5–9	0.1–0.4
Germanium	Catechol	HMDE, MFE	DP	H <sub>2</sub> SO <sub>4</sub>	10
Gold	PAN	HMDE	DP	Britton–Robinson buffer	1
Iodide	Copper, mercury(I)	HMDE, MFE	DP, SW	2–5 or 8	0.6–3
Iron	Catechol, NN, SCN <sup>-</sup>	HMDE	DP, LS	6.8–8.0	2
Lead	Oxine	HMDE, MFE	DP	7.0–8.5	0.3
Molybdenum	BO, CA, MA, oxine, tropolone	HMDE, MFE	DP	1.9–3	0.1–20
Nickel	DMG	HMDE, MFE	DP	7–10	0.1
Palladium	DMG	HMDE	DP	5.15	2
Rhodium	Chloride, formaldehyde	HMDE	DP	1.3	50
Selenium	Copper	HMDE	DP	1.6–4.5	0.1–3
Technetium	Thiocyanate	HMDE	DP	2	5
Tellurium	Copper	HMDE	DP	4.5	10
Thorium	Cupferron	HMDE	DP		0.2
Tin	Catechol, tropolone	HMDE	DP	1.5–4.7	0.23–5
Titanium	MA, PC, cupferron	HMDE	DP, LS	1–4.2	0.3–5
Tungsten	CA, ferron, oxine	HMDE	DP		0.1–2
Uranium	Catechol	HMDE	DP	6.3–7.2	0.7
	Oxine		DP	6.5–7.1	0.2
	TTA/TBP		LS	2–3.6	1
Vanadium	Catechol, cupferron, DHN	HMDE, MFE	DP, SW, LS	6.6–7.2	0.6–4
Zinc	APDC, oxine	HMDE	DP	6.2–8.5	0.3

BD =  $\alpha$ -benzildioxime; BO =  $\alpha$ -benzoinooxime; BP = 2,2'-bipyridine; CA = chloranilic acid; DHN = 2,3-dihydroxynaphthalene; DMG = dimethylglyoxime; DMTD = 2,5-dimercapto-1,3,4-thiadiazole; DTPA = diethylenetriaminepentaacetic acid; FD =  $\alpha$ -furildioxime; HNB = *N*-hydroxy-*N*-nitroso benzeneamine; MA = mandelic acid; NO = nioxime; NN = 1-Nitroso-2-naphthol; PAN = 1-(2'-Pyridylazo)-2-naphthol; PC = pyrocatechol; TTHA = triethylenetetraminehexaacetic acid.

Fe, Mo, Pt, Rh, Ti, U, V, and W (Table 3). Some of these systems, such as the AdSV measurement of cobalt with nioxime in presence of nitrite as oxidant (Table 3), have been found to increase the voltammetric response by at least four orders of magnitude. In general, the sensitivity obtained for the AdSV measurement of these inorganic substances in the presence of oxidants (Table 3) are much higher than those obtained without oxidants (Table 1) and has enabled the achievement of detection limits in the subpicomolar range. In addition, the utilization of catalytic systems in AdSV has been found to be useful in widening the analytical concentration range and improving the selectivity of the determination.

The main factors responsible for the improvement of the selectivity of CAdSV are: (1) the increase in the

ratio of the faradaic current of the analyte to the faradaic currents of the other substances present in the solution, and (2) the specificity of the reaction of an ion-catalyst with the chosen oxidant. Consequently, the optimization of the sensitivity and selectivity of CAdSV requires careful investigation of the influence of supporting electrolyte composition (choice of complexing agent, oxidant, pH, buffer/electrolyte concentration, etc.) and instrumental parameters (accumulation time, potential, voltammetric mode, scan rate, etc.). In terms of the choice of voltammetric mode, the use of differential pulse (DP) voltammetric mode is most popular because of the high sensitivity and selectivity commonly achieved with this mode. The two other voltammetric modes that are also commonly used with CAdSV are

**Table 2** Some of the complexing agents that have been used for AdSV determination of inorganic substances based on complexing agent reduction

Complexing agent	Inorganic element or compound	Working electrode(s)	pH	Detection limit (nmol l <sup>-1</sup> )
o-Cresolphthalexone	Cerium	HMDE	9.5	2
	Lanthanum	HMDE	9.5	2
	Praseodymium	HMDE	9.5	2
1,2-Dihydroxyanthraquinone-3-sulfonic acid	Aluminum	HMDE	6–8	1
Eriochrome black-T	Manganese	HMDE	12	5
Molybdate	Silicon	HMDE	1.5	
Mordant blue	Uranium	HMDE	6.5	1
Mordant blue-T	Thorium	HMDE	6.5	2
Solochrome violet-RS	Aluminum	HMDE	4.5	30
	Gallium	HMDE	4.8	2
	Iron	HMDE	5.1	0.7
	Titanium	HMDE	5	5
	Yttrium	HMDE	6.5	
	Zirconium	HMDE	4.5	2

square wave (SW) and linear sweep (LS) modes (Table 3). The successful use of CAdSV for some speciation studies has also been reported and more interest is expected in this area in the future.

## Important Considerations in AdSV

### Choice of Complexing Agent

As indicated previously, the choice of a suitable complexing or chelating agent is a very important consideration for the successful utilization of AdSV for the determination of inorganic compounds in various sample matrices. It is possible that a chosen complexing agent may be capable of forming complexes with the analyte(s) and many other substances, but the determinant factor is whether all of these complexes can be adsorbed onto the electrode at the chosen solution pH and accumulation potential. In most cases only some of these complexes can adsorb under the chosen conditions and the reduction potential for each complex will be specific for each inorganic substance. It is important to note that, due to the formation of a complex, the reduction potential is shifted toward a more negative potential relative to that of the inorganic substance or of the complexing agent depending on whether the inorganic substance (e.g., a metal) or complexing agent is reduced during the negative potential scan. If complexes of more than one inorganic substance is adsorbed onto the electrode and give separate reduction peaks that do not overlap, it is possible to use this for multielement determination in a single sample solution, as demonstrated for nickel and cobalt in Figure 1. A more versatile example of this, which was recently reported, employs the use of two complexing agents for multielement determination of six

metals. The inclusion of oxine enabled the detection of copper, lead, cadmium, and zinc, while the presence of dimethylglyoxime also enabled detection of cobalt and nickel in the same solution. The wide range of complexing agents that have been successfully used for the AdSV determination of numerous inorganic substances are also listed in Tables 1–3.

### Choice of Accumulation Potential

In cases where peak overlap occurs, this problem can usually be overcome by varying the solution composition, such as pH, buffer concentration, or buffer component; or by careful choice of accumulation or adsorption potential. The choice of accumulation potential is influenced mainly by two factors: (1) the charge on the electrode surface, and (2) the oxidation state of the dissolved complex. The choice of a complexing agent which forms complexes with many inorganic substances may result in their adsorption at a chosen applied accumulation potential and leads to competition for sites on the electrode surface. In such instance, the use of a selective deposition strategy that employs an accumulation potential, ~100–300 mV more positive than the reduction potential(s) of the analyte(s), can be effective in minimizing or eliminating the competition.

In other cases, it is possible to take advantage of the possibility of some inorganic substances, specifically metal ions, to be amalgamated with mercury by applying a more negative potential than the reduction potential(s) of the analyte(s) during the adsorption step. This ensures the collection of only the amalgamated metals. After this, the potential is switched back to a more positive potential to enable the amalgamated metals to be reoxidized and complexed by the excess complexing agent in the solution.

**Table 3** Some catalytic agents that have been used for CAdSV determination of inorganic compounds

Catalytic agent	Inorganic element or compound	Complexing agent(s)	Working electrode(s)	Voltammetric mode	Detection Limit (nmol l <sup>-1</sup> )	
BrO <sub>3</sub> <sup>-</sup>	Iron	NN	HMDE	SW	0.08	
	Molybdenum	DHN	HMDE	LS	0.006	
		Ferron	HMDE	DP	94	
	Titanium	MA	HMDE	DP		
	Vanadium	Catechol	HMDE	SW	0.6	
		Cupferron	HMDE	DP	0.0049	
		Cupferron	MFE	DP	0.16	
ClO <sub>2</sub> <sup>-</sup>		DHN	HMDE	SW	0.015	
	Iron	Catechol	HMDE	DP	10	
ClO <sub>3</sub> <sup>-</sup>	Molybdenum	BO	MCPE	LS	20	
		CA	HMDE	DP		
		MA	HMDE	DP	0.0014	
		Oxine	HMDE	LS	0.0017	
	Titanium	MA	HMDE	DP	0.007	
		PC	HMDE	LS	1.25	
	Tungsten	Oxine + HMMA	HMDE	DP	0.025	
	H <sup>+</sup>	Cobalt	8-Quinolinethiol	HMDE	DP	0.5
		Platinum	Formazone	HMDE	DP	0.00004
				HMDE	LS	0.001
MFE				SW	2.5	
		DME	DPP	0.1		
Rhodium	Formazone	HMDE	DP	0.000054		
H <sub>2</sub> O <sub>2</sub>	Tungsten	CA	HMDE	DP	0.02	
	Iron	NN	HMDE	LS	0.16	
		BPH	HMDE	DP	0.18	
NO <sub>2</sub> <sup>-</sup>	Cobalt	BP	HMDE	LS	0.0095	
		DMG	HMDE	DP	0.04	
		Nioxime	HMDE	DP	0.003	
	Chromium	Bipyridyl	HMDE	LS	0.02	
		DTPA	HMDE	DP	0.029	
	Iron	SCN <sup>-</sup>	HMDE	LS	0.716	
	NO <sub>3</sub> <sup>-</sup>	Uranium	DTPA	HMDE	DP	3.9
Chromium		DTPA	HMDE	DP	0.1	
		TTHA	HMDE	SW	0.29	
Molybdenum		Oxine	HMDE	DP	0.7	
		Oxine	MFE	SW	0.15	
		TC	HMDE	LS	0.011	

BO =  $\alpha$ -benzoinooxime; BP = 2,2'-bipyridine; BPH = *N*-benzoyl-*N*-phenylhydroxylamine; CA = chloranilic acid; DHN = 2,3-dihydroxy-naphthalene; DMG = dimethylglyoxime; DTPA = diethylenetriaminepentaacetic acid; HMMA = 4-hydroxy-2-methoxymandelic acid; MA = mandelic acid; NN = 1-Nitroso-2-naphthol; PC = Pyrocatechol; TC = 2(2'-Thiazolulazo)-*p*-crezol; TTHA = triethylenetetramine-nehexaacetic acid.

Then a negative potential scan is applied, thereby minimizing the interference from inorganic substances that do not form an amalgam. So, in this case, the electrode processes involved include adsorptive collection, followed immediately by reduction and amalgamation, reoxidation step, and finally the repetition of the reduction of the adsorbed complex. This approach has been successfully used for the determination of antimony in the presence of iron, vanadium, and uranium, which is usually difficult to achieve, especially in the presence of high uranium concentration.

#### Choice of Electrodes for AdSV

Hanging mercury drop electrode (HMDE) is by far the most commonly used electrode for AdSV, as

indicated in Tables 1 and 3. This is because it has the ability to provide renewed (fresh) mercury surface and, hence, eliminate or minimize fouling caused by adsorption of interfering surface-active substances, or residual-adsorbed complex. The use of dropping mercury electrode (DME), when operating in the polarographic mode, has similar advantages and this electrode is commonly used for adsorptive polarographic determination of inorganic substances at higher concentrations. Also, in recent years, the use of mercury film electrodes (MFE) has also gained more interest in both AdSV and CAdSV (Tables 1 and 3) because of the environmental concern for mercury. Various microelectrodes and ultramicroelectrodes of different sizes and materials (usually plated with



mercury) have also been successfully used for the AdSV determination of some metals. More recently, the use of a bismuth film electrode for AdSV determination of some metals have been reported.

### Interferences in AdSV

The main interferences experienced in AdSV are usually caused by either competitive adsorption of surface-active substances, usually organic but can also be inorganic such as halide ions, or by competitive complexation of other metals by the chosen complexing agent. The latter can result, as indicated previously, in a competition for sites on the electrode surface and, consequently, result in a significant depression of the analytical response. The extent of interference depends on the relative affinities of the analyte and interferant(s) to the electrode surface and the concentration ratio of the analyte to the interferant(s).

Interference from surface-active compounds in AdSV can be removed by destruction of the organic substances prior to the voltammetric measurement by use of ultraviolet irradiation or by appropriate digestion in the presence of reagents, such as perchloric acid or hydrogen peroxide. Other approaches that can be used to reduce or eliminate interferences by inorganic and organic substances in stripping voltammetric measurements include the use of a more selective accumulation potential, use of shorter accumulation times, masking of interfering metals with ethylenediaminetetraacetic acid, more selective or specific choice of complexing agent, use of catalytic systems, medium exchange method, and quantification by standard additions method to compensate for matrix effects.

See also: **Voltammetry**: Anodic Stripping; Cathodic Stripping.

### Further Reading

- Adeloju SB and Hadjichari AH (1999) Simultaneous determination of nickel and cobalt in natural water and sediment samples on an *in-situ* plated mercury film electrode by adsorptive cathodic stripping voltammetry. *Analytical Science* 15: 95–100.
- Adeloju SB and Pablo F (1995) Simultaneous determination of lead and tin in biological and environmental materials by differential pulse cathodic voltammetry on a glassy carbon mercury film electrode. *Electroanalysis* 7: 750–755.
- Bard AJ and Faulkner LR (2001) *Electrochemical Methods: Fundamentals and Applications*. New York: Wiley.
- van den Berg CMG (1991) Potentials and potentialities of cathodic stripping voltammetry to the determination of trace elements in natural waters. *Analytica Chimica Acta* 250: 265–276.
- Bond AM (1980) *Modern Polarographic Techniques in Analytical Chemistry*. New York: Dekker.
- Desimoni E, Brunetti B, and Cattaneo R (2004) Validation of some procedures for quantifying platinum at sub- $\mu\text{g/L}$  level in some real matrices by catalytic adsorptive stripping voltammetry. *Electroanalysis* 16: 304–310.
- Dominguez O and Arcos MJ (2000) Speciation of chromium by adsorptive stripping voltammetry using pyrocatechol violet. *Electroanalysis* 12: 449–458.
- Krolicka A, Bobrowski A, Kalcher K, *et al.* (2003) Study on catalytic adsorptive stripping voltammetry of trace cobalt at bismuth film electrodes. *Electroanalysis* 15: 1859–1863.
- Rubinson KA and Robinson JF (2000) *Contemporary Instrumental Analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Vargas MC, Renedo OD, and Martinez MJA (2003) Simultaneous determination of chromium(VI) and aluminium(III) by adsorptive stripping voltammetry with pyrocatechol violet. *Helvetica Chimica Acta* 86: 2434–2440.
- Wang J (1985) *Stripping Analysis Principles, Instrumentation and Applications*. Deerfield Beach, FL: VCH.
- Wang J and Lu J (2000) Bismuth film electrodes for adsorptive stripping voltammetry of trace nickel. *Electrochemical Communications* 2: 390–393.

## Organic Compounds

**R J Forster and D A Walsh**, Dublin City University, Dublin, Republic of Ireland

© 2005, Published by Elsevier Ltd.

### Introduction

Cyclic voltammetry provides a profound insight into the redox reactions of organic molecules. In this technique, the potential is scanned from an initial

potential to an upper potential before reversing the direction of the scan and usually returning to the initial potential. The timescale of the experiment can be systematically varied by changing the scan rate  $v$ . This potential waveform has the advantage that the product of the electron transfer reaction that occurred in the forward scan can be probed again in the reverse scan. Thus, cyclic voltammetry can be used to determine formal redox potentials, to measure the rate of heterogeneous electron transfer kinetics across the electrode/solution interface and to detect



mercury) have also been successfully used for the AdSV determination of some metals. More recently, the use of a bismuth film electrode for AdSV determination of some metals have been reported.

### Interferences in AdSV

The main interferences experienced in AdSV are usually caused by either competitive adsorption of surface-active substances, usually organic but can also be inorganic such as halide ions, or by competitive complexation of other metals by the chosen complexing agent. The latter can result, as indicated previously, in a competition for sites on the electrode surface and, consequently, result in a significant depression of the analytical response. The extent of interference depends on the relative affinities of the analyte and interferant(s) to the electrode surface and the concentration ratio of the analyte to the interferant(s).

Interference from surface-active compounds in AdSV can be removed by destruction of the organic substances prior to the voltammetric measurement by use of ultraviolet irradiation or by appropriate digestion in the presence of reagents, such as perchloric acid or hydrogen peroxide. Other approaches that can be used to reduce or eliminate interferences by inorganic and organic substances in stripping voltammetric measurements include the use of a more selective accumulation potential, use of shorter accumulation times, masking of interfering metals with ethylenediaminetetraacetic acid, more selective or specific choice of complexing agent, use of catalytic systems, medium exchange method, and quantification by standard additions method to compensate for matrix effects.

See also: **Voltammetry**: Anodic Stripping; Cathodic Stripping.

### Further Reading

- Adeloju SB and Hadjichari AH (1999) Simultaneous determination of nickel and cobalt in natural water and sediment samples on an *in-situ* plated mercury film electrode by adsorptive cathodic stripping voltammetry. *Analytical Science* 15: 95–100.
- Adeloju SB and Pablo F (1995) Simultaneous determination of lead and tin in biological and environmental materials by differential pulse cathodic voltammetry on a glassy carbon mercury film electrode. *Electroanalysis* 7: 750–755.
- Bard AJ and Faulkner LR (2001) *Electrochemical Methods: Fundamentals and Applications*. New York: Wiley.
- van den Berg CMG (1991) Potentials and potentialities of cathodic stripping voltammetry to the determination of trace elements in natural waters. *Analytica Chimica Acta* 250: 265–276.
- Bond AM (1980) *Modern Polarographic Techniques in Analytical Chemistry*. New York: Dekker.
- Desimoni E, Brunetti B, and Cattaneo R (2004) Validation of some procedures for quantifying platinum at sub- $\mu\text{g/L}$  level in some real matrices by catalytic adsorptive stripping voltammetry. *Electroanalysis* 16: 304–310.
- Dominguez O and Arcos MJ (2000) Speciation of chromium by adsorptive stripping voltammetry using pyrocatechol violet. *Electroanalysis* 12: 449–458.
- Krolicka A, Bobrowski A, Kalcher K, *et al.* (2003) Study on catalytic adsorptive stripping voltammetry of trace cobalt at bismuth film electrodes. *Electroanalysis* 15: 1859–1863.
- Rubinson KA and Robinson JF (2000) *Contemporary Instrumental Analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Vargas MC, Renedo OD, and Martinez MJA (2003) Simultaneous determination of chromium(VI) and aluminium(III) by adsorptive stripping voltammetry with pyrocatechol violet. *Helvetica Chimica Acta* 86: 2434–2440.
- Wang J (1985) *Stripping Analysis Principles, Instrumentation and Applications*. Deerfield Beach, FL: VCH.
- Wang J and Lu J (2000) Bismuth film electrodes for adsorptive stripping voltammetry of trace nickel. *Electrochemical Communications* 2: 390–393.

## Organic Compounds

**R J Forster and D A Walsh**, Dublin City University, Dublin, Republic of Ireland

© 2005, Published by Elsevier Ltd.

### Introduction

Cyclic voltammetry provides a profound insight into the redox reactions of organic molecules. In this technique, the potential is scanned from an initial

potential to an upper potential before reversing the direction of the scan and usually returning to the initial potential. The timescale of the experiment can be systematically varied by changing the scan rate  $v$ . This potential waveform has the advantage that the product of the electron transfer reaction that occurred in the forward scan can be probed again in the reverse scan. Thus, cyclic voltammetry can be used to determine formal redox potentials, to measure the rate of heterogeneous electron transfer kinetics across the electrode/solution interface and to detect

chemical reactions that precede or follow electron transfer. Electrochemistry has several advantages over spectroscopy in that it provides 'direct' information about electron transfer and coupled chemical reactions. Moreover, it enables analysis to be performed in colored solutions without the significant problem of a large background absorbance typically found with ultraviolet-visible absorbance spectroscopy. Also, with the development of microelectrodes, voltammetry can now be performed in unusual media such as nonpolar solvents, supercritical fluids, gases, and even solids.

Many significant electrochemical events, such as electron and proton transfers, ligand exchanges, isomerizations, and ejection of leaving groups, occur on the low microsecond and nanosecond time domains. To achieve a meaningful insight into these redox processes, it must be possible to measure rate constants under a wide range of experimental conditions, such as driving force, temperature, etc. However, conventional electrochemical methods cannot fulfill this role since they are restricted to millisecond, or longer, timescales. Thus, while modern laser-based spectroscopy has provided a powerful new insight into chemical processes that occur at picosecond and even femtosecond timescales, it is only recently that electrochemists armed with electrodes of micro- or even nanometer dimensions, have meaningfully probed redox processes occurring on the submicrosecond timescale.

These attractive features of the technique, together with the wide variety of organic compounds shown in Tables 1 and 2 that can be reduced or oxidized, mean that cyclic voltammetry has truly come of age in the electroanalysis of organic compounds.

## Heterogeneous Electron Transfer Dynamics

Measuring the rate of heterogeneous electron transfer across a metal/solution interface is of considerable technological importance in areas ranging from electrosynthesis to the detection of highly reactive organic intermediates. At short times, i.e., large scan rates, the diffusion layer thickness is much smaller than the electrode radius and the dominant mass transport mechanism is planar diffusion. Provided that ohmic drop is absent and that the faradaic and capacitive currents can be separated, under these conditions, kinetic parameters can be extracted from the scan rate dependence of the separation between the anodic and cathodic peak potentials. Figure 1 illustrates a typical example of the effect of increasing scan rate on the voltammetric response for

anthracene. This figure shows that at high scan rates the time constants of the experiment and heterogeneous electron transfer become comparable and the peak-to-peak separation,  $\Delta E_p$ , exceeds the 57/mV expected for an ideal reversible electrochemical reaction. In the case of anthracene, heterogeneous electron transfer is a fast process since there is little difference in the molecular structure of the reduced and oxidized forms. Probing this rapid electron transfer reaction can be accomplished using fast scan rates, typically tens of thousands of volts per second, that is only possible using microelectrodes. The standard heterogeneous electron transfer rate constants for several organic molecules are given in Table 3. These data show that the rate of electron transfer across the electrode/solution interface is very significantly influenced by the identity of the redox couple. These variations reflect differences in the strength of electronic coupling between the electrode and the reactant as well as activation barriers that exist for the individual systems. For example, where the structure of the molecule is significantly distorted by the redox reaction, or where the reaction involves multiple steps, standard heterogeneous electron transfer rate constants as small as  $10^{-9} \text{ cm s}^{-1}$  can be observed, i.e., almost 10 orders of magnitude smaller than that found for the polyaromatic anthracene molecule.

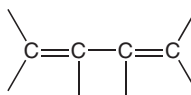
A useful strategy in trying to extend the upper limit of measurable electron transfer rate constants is to perform measurements at lower temperatures. This strategy is successful because even for heterogeneous electron transfers with negligible inner sphere reorganization energies, activation barriers  $\sim 20\text{--}25 \text{ kJ mol}^{-1}$  are expected. Therefore, considerably slower rates of heterogeneous electron transfer are observed even by decreasing the temperature of the electrochemical cell by a few tens of degrees. Measurements of this type are greatly facilitated by microelectrodes since solvents, such as alcohols or nitriles that remain liquid over a wide temperature range can be used without catastrophic ohmic effects and the behavior of organic molecules such as *o*-nitrotoluene and nitrometisylene have been investigated.

## Homogeneous Chemical Kinetics

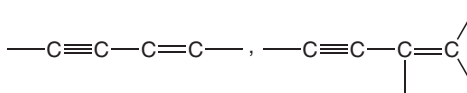
Adding or removing an electron from an organic molecule often creates a radical species that is not indefinitely stable and tends to react either with the parent molecule or other species in solution. For example, electrochemical oxidation of ascorbic acid (vitamin C) produces an oxidized product that

**Table 1** Organic compounds that are capable of undergoing a reduction reaction

Dienes: conjugated double bonds



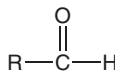
Acetylenes



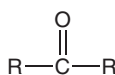
Conjugated aromatics



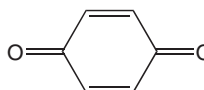
Aldehydes



Ketones



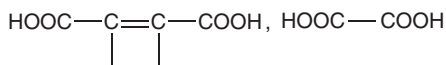
Quinones



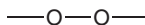
Hydroquinones



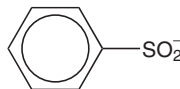
Conjugated carboxylic acids



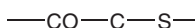
Peroxides



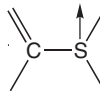
Sulfones



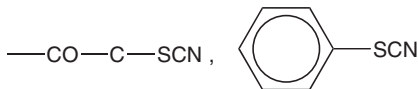
Sulfides



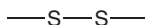
Sulfonium salts



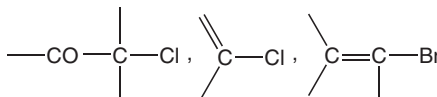
Thiocyanates



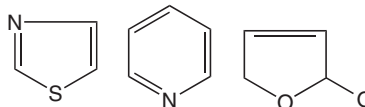
Disulfides



Halides

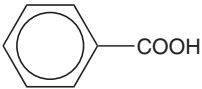


Heterocycles

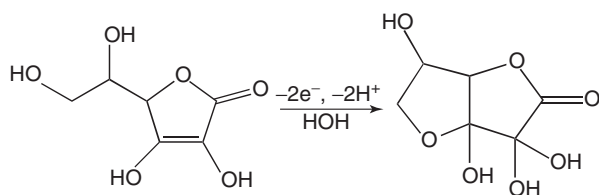


Continued

**Table 1** (Continued)

Organometallics	$\begin{array}{c} \diagup \\ \text{C} \text{---} \text{Hg} \text{---} \\ \diagdown \end{array}, \quad \begin{array}{c} \text{---Se---} \\ \downarrow \\ \text{O} \end{array}$
Aromatic carboxylic acids	
Imines	$\begin{array}{c} \text{---C=N---} \\   \end{array}$
Oximes	$\begin{array}{c} \diagup \\ \text{C=N-OH} \\ \diagdown \end{array}$
Nitriles	$\text{CN---C---N}^+ \begin{array}{l} \diagup \\   \\ \diagdown \end{array}, \quad \text{CO---C---N}^+ \begin{array}{l} \diagup \\   \\ \diagdown \end{array}$
Diazo compounds	$\text{---N}^+ \equiv \text{N}$
Nitroso compounds	$\begin{array}{c} \diagup \\ \text{N=N=O} \\ \diagdown \end{array}$
Nitro compounds	$\text{---NO}_2, \quad \begin{array}{c} \diagup \\ \text{C=C---NO}_2 \\ \diagdown \end{array}, \quad \text{C}_6\text{H}_5\text{NO}_2$

subsequently reacts with water to yield electrochemically inactive dehydroascorbic acid.


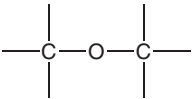
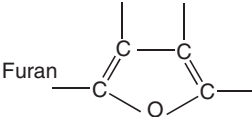
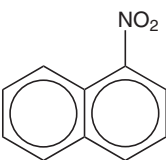

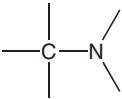
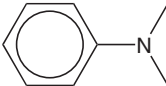
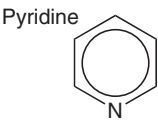
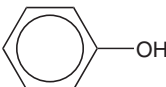

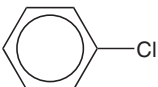
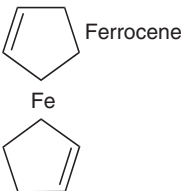


As shown in **Figure 2**, the voltammetric response is significantly altered by the coupled chemical reaction and thus allows the energetics and dynamics of these homogeneous chemical reactions to be probed. This figure shows that as the scan rate is increased, the contribution from the homogeneous reaction becomes less pronounced and the voltammogram approaches the shape of that for an electrochemically reversible process. The fact that ascorbic acid is electroactive means that its concentration in fruit juices can be determined without interferences from the coloration of the sample.

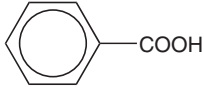
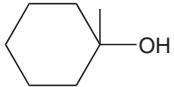
In general terms, coupled mechanistic schemes are described by the letters E (electrochemical) and C (chemical). The order in which they are written denotes the order in which the processes occur. Thus an ECE mechanism describes a process in which an electrochemical step is followed by a chemical step, which is then followed by an electrochemical step. A chemical step is a step where no electron transfer to or from the electrode takes place. Such a step does not by itself produce a charge flow into or out of the electrode and thus is not directly observable by an external measuring circuit. It may, however, influence charge flow because of other steps in the mechanism, which can be detected indirectly. The chemical step is not directly influenced by the electrode potential. An electrochemical step on the other hand involves electron flow to and from the electrode and as such produces a flow of charge that can be monitored by the external measuring circuit.

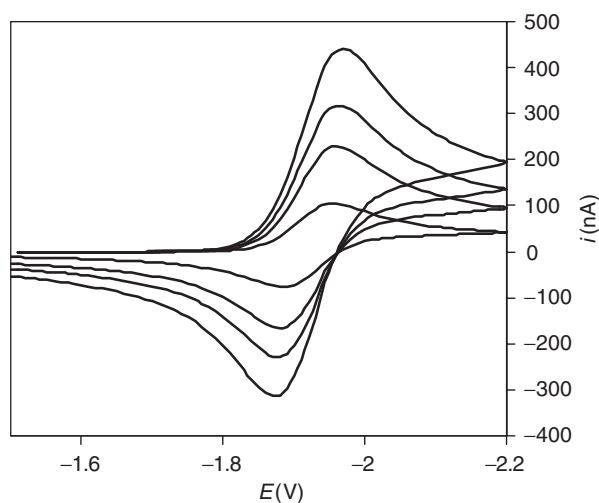
Bimolecular reactions in solution cannot proceed faster than the rate at which molecules come into close contact. Thus, bimolecular rate constants

**Table 2** Organic compounds that are capable of undergoing an oxidation reaction

Aromatics	
Ethers	
Heterocyclics	
Nitroaromatics	
Olefins	
Amines	
Aromatic amines	
Heterocyclic amines	
Phenols	
Aliphatic halides	
Aromatic halides	
Organometallics	

**Table 2** (Continued)

Carboxylic acids	
Alcohols	

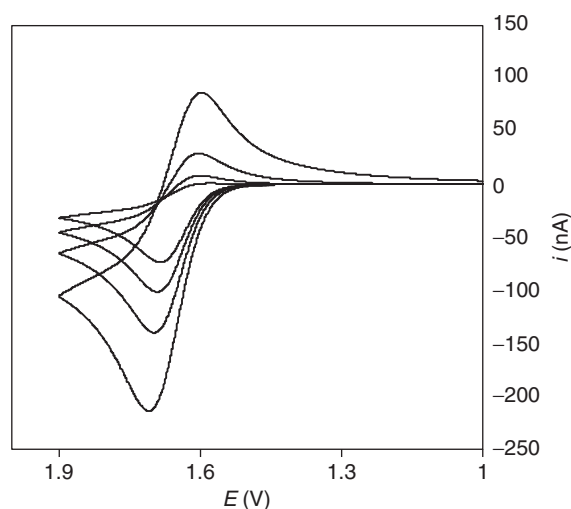
**Figure 1** Dependence of the cyclic voltammetry response for a  $5 \text{ mmol l}^{-1}$  solution of anthracene dissolved in DMF containing  $0.1 \text{ mol l}^{-1}$  tetrabutyl ammonium perchlorate as supporting electrolyte on the scan rate. From top to bottom the scan rates are 20 000, 10 000, 5000, and 1000  $\text{V s}^{-1}$ . The working electrode is a  $5 \mu\text{m}$  radius platinum disk.

cannot exceed the diffusion limited rate constant that is  $\sim 10^9\text{--}10^{10} \text{ mol l}^{-1} \text{ s}^{-1}$  in most organic solvents. Since under semiinfinite linear diffusion conditions, the characteristic time of cyclic voltammetry is  $RT/Fv$ , where  $v$  is the scan rate, experiments performed at  $\text{MV s}^{-1}$  scan rates allow kinetic information, such as lifetimes that are close to the diffusion limit, to be obtained. Thus, by scanning sufficiently fast the following chemical reaction can be outrun, and the irreversible response observed at slow scan rates can be made electrochemically reversible at higher scan rates. This behavior is opposite to that expected when heterogeneous electron transfer is slow where the response would become increasingly irreversible with increasing scan rate. The ability to make the voltammetric response reversible means that the formal potentials of highly reactive species can be accurately measured. As shown in **Table 3**, using approaches of this kind, the dynamics of a wide variety of coupled chemical reactions have been probed.

**Table 3** Rate constants for heterogeneous electron transfer

Analyte	Electrode <sup>a</sup>	$k^0$ (cm s <sup>-1</sup> )
Anthracene	Au ring $\Delta r = 0.09 \mu\text{m}$ , $r = 5 \mu\text{m}$	$3.33 \pm 0.05$
Anthracene oxidation	Pt, $5 \mu\text{m}$	$7.6 \pm 1.0 \times 10^3 \text{ s}^{-1}$
Cytochrome c	C, $6.3 \mu\text{m}$	$> 0.4$
9,10-Diphenylanthracene	Au ring $\Delta r = 90 \mu\text{m}$ , $r = 5 \text{ mm}$	$5.7 \pm 0.1$
MV <sup>2+</sup> (F)	Pt, $22 \text{ \AA}$ to $0.21 \mu\text{m}$	$170 \pm 90$
Naphthalene	Au ring $\Delta r = 0.2 \mu\text{m}$ , $r = 20.5 \mu\text{m}$	$0.88 \pm 0.02$
Tetracyanoethylene	Au ring $\Delta r = 0.2 \mu\text{m}$ , $r = 20.5 \mu\text{m}$	$0.15 \pm 0.01$
Tetracyanoquinodi-ethane	Au ring $\Delta r = 0.2 \mu\text{m}$ , $r = 20.5 \mu\text{m}$	$0.23 \pm 0.01$
Ascorbic acid oxidation	Hg, $5 \mu\text{m}$	$1.4 \times 10^3 \text{ s}^{-1}$
3-Bromoaceto-phenone reduction	Au, $5 \mu\text{m}$	$6.3 \times 10^4 \text{ s}^{-1}$
	Au, $5 \mu\text{m}$	$8 \times 10^3 \text{ s}^{-1}$
Acetophenone oxidation	Au, $5 \mu\text{m}$	$6.3 \times 10^4 \text{ s}^{-1}$
Chlorpromazine + dopamine	Pt, $25 \mu\text{m}$	$10^8 \text{ mol l}^{-1} \text{ s}^{-1}$
<i>N,N</i> -dimethylaniline oxidation	Pt, $12.5 \mu\text{m}$	$6.3 \times 10^5 \text{ mol l}^{-1} \text{ s}^{-1}$
Methylbenzenes	Pt, $5 \mu\text{m}$	$10^9 \text{ mol l}^{-1} \text{ s}^{-1}$

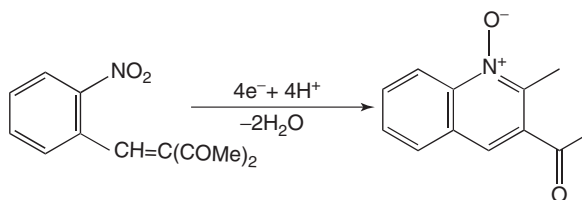
<sup>a</sup> Dimension given is the radius of a microdisk electrode unless otherwise stated.



**Figure 2** Simulated voltammograms showing the behavior observed for an organic molecule, e.g., acetophenone, that undergoes a following chemical reaction upon oxidation. The electrode is a  $5 \mu\text{m}$  radius platinum disk, and the rate constant for the homogeneous chemical reaction is  $7.6 \times 10^3 \text{ s}^{-1}$ .

Reductive coupling reactions are an example of a class of reactions where voltammetry has provided a deep insight into the reaction mechanism and allowed optimum conditions for electrosynthesis to be identified. The reductive formation of nucleophiles followed by an attack on an electrophilic center is also the basis for the synthesis of many heterocyclic compounds. For example, reducing the nitro group in suitable *ortho*-substituted nitro compounds to the

hydroxylamine stage may lead to an *N*-oxide of an *N*-heterocyclic compound. For example, *o*-nitrobenzalacetylacetone is reduced to 3-acetyl-2-methylquinoline-*N*-oxide.



On reduction of aryl halides, the initially formed radical anion decomposes to halide ion and the aryl radical; the aryl radical is generally more easily reduced than the substrate or it may abstract a hydrogen atom from the solvent system, but in suitably substituted compounds the radical may attack a phenyl ring with ring closure. Thus, for example, 5-(2-chlorophenyl)-1-phenylpyrazole is reduced to pyrazolo[1,5-*f*]phen-anthridine.

Oxidation of parent organics has also been extensively investigated using voltammetry leading to both anodic substitution and coupling. For example, the voltammetry of a large number of heterocyclic derivatives has been probed and compounds prepared through the anodic methoxylation of acylated heterocycles, in peptidomimetics, and in the first example of combinatorial electrochemistry.

A second voltammetric approach to probing the dynamics of homogeneous chemical reactions that are coupled to electron transfer is to exploit the



**Table 4** Rate constants for homogeneous chemical reactions involving organic compounds

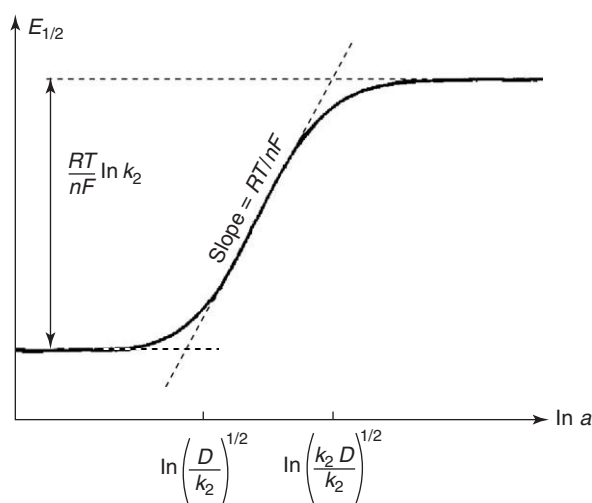
Analyte	Electrode <sup>a</sup>	<i>K</i>
Anthracene oxidation	Pt, 5 μm	$7.6 \pm 1.0 \times 10^3 \text{ s}^{-1}$
9,10-Diphenylanthracene + 4,4-dibromodiphenyl	C, 6–9 μm	$3.9 \pm 0.6 \text{ mol l}^{-1} \text{ s}^{-1}$
Ascorbic acid oxidation	Hg, 5 μm	$1.4 \times 10^3 \text{ s}^{-1}$
Ascorbic acid oxidation at a Prussian blue film	Pt, 2.5–25 μm	$1.3 \times 10^5 \text{ mol l}^{-1} \text{ s}^{-1}$
Acetophenone oxidation	Au, 5 μm	$6.3 \times 10^4 \text{ s}^{-1}$
3-Bromoaceto-phenone reduction	Au, 5 μm	$6.3 \times 10^4 \text{ s}^{-1}$
Chlorpromazine + dopamine	Pt, 25 μm	$10^8 \text{ mol l}^{-1} \text{ s}^{-1}$
<i>N,N</i> -dimethylaniline oxidation	Pt, 12.5 μm	$6.3 \times 10^5 \text{ mol l}^{-1} \text{ s}^{-1}$
Methylbenzenes	Pt, 5 μm	$10^9 \text{ mol l}^{-1} \text{ s}^{-1}$
Anion radicals + alkyl halides	Pt, 0.25 μm	$9 \times 10^{-4} - 1.7 \times 10^4$
Hexamethylbenzene oxidation	Pt, 0.3–25 μm	$720 \pm 100 \text{ s}^{-1}$
1-Naphthylamine oxidation	Pt, 0.5–12.5 μm	$4.1 \times 10^3 \text{ s}^{-1}$
Triphenylamine oxidation	Pt, 0.3–20 μm	$> 3 \times 10^4 \text{ mol l}^{-1} \text{ s}^{-1}$
Thioselenanthrene		$8.87 \pm 1.1 \text{ s}^{-1}$
Dibenzo-1,2-diselenine		$20.7 \pm 2.8 \text{ s}^{-1}$

<sup>a</sup>Dimension given is the radius of a microdisk electrode unless otherwise stated.

steady-state response observed at relatively slow (hundreds of milliseconds) at ultrasmall, i.e., micro- to nanometer, dimensioned electrodes. Steady-state measurements using electrodes of different radii can provide a powerful insight into the kinetics of some classes of homogeneous reactions since the limiting current density depends on the magnitude of the homogeneous rate constant (Table 4). Hence, coupled chemical (C) and electron transfer (E) reactions, e.g., CE mechanisms, catalytic follow-up processes, as well as reactions involving disproportionation have been characterized. It is important to note that reactions, such as chemical reactions that follow electron transfer (EC) mechanisms, cannot be investigated in the same way since the current density is not influenced by the following chemical reaction. In these circumstances, the homogeneous reaction does not affect the height or shape of the reversible steady-state voltammogram. However, as indicated by Figure 3, the position of the wave on the potential axis depends on the homogeneous reaction rate and kinetic information can be obtained by probing how  $E_{1/2}$  depends on the electrode radius.

The technique is best applied to reactions whose rate constants are of the same magnitude as  $D/r^2$ , where  $D$  is the diffusion coefficient of the analyte through solution and  $r$  is the electrode radius. Therefore, given that it is now feasible to fabricate microelectrodes with submicron critical dimensions, the dynamics of first order chemical reactions with rate constants between  $10^2$  and  $10^4 \text{ s}^{-1}$  can be investigated.

The redox properties of the vast majority of organic compounds are investigated at pristine surfaces such as highly polished platinum, gold, and glassy carbon electrodes. However, surfaces may be modified in many ways to get better catalytic effects, chiral

**Figure 3** Dependence of  $E_{1/2}$  for an EC reaction on the logarithm of the electrode radius.

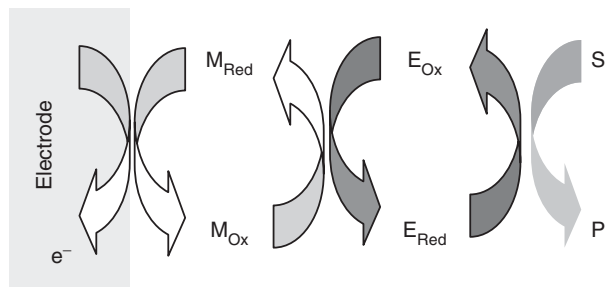
induction, or some other desirable effect not obtainable at the 'classical' electrode materials. Irreversible adsorption, covalent attachment, and electropolymerization have been used. The modified electrodes are usually not stable enough for preparative reactions using high current densities but find common application as sensitive and selective sensors.

## Biologically Relevant Molecules

Voltammetry provides a useful tool for studying redox chemistry of the molecules relevant to biological redox reactions. Many proteins are exclusively involved in intraprotein electron transfer and typically function in ordered structures such as mitochondria. Under these circumstances, the redox

active centers are generally accessible on the outer surface of the protein. In contrast, the redox reactions catalyzed by oxidoreductases involve small molecules and two redox couples, i.e., the substrate and the cofactor or cosubstrate, participate in the reaction. Because the catalytic center of the enzyme is often located deep within the protein, it often undergoes only a very weak electronic interaction with the electrode. This weak coupling results in very slow heterogeneous electron transfer. Driven in part by their ability to selectively catalyze reactions of analytical importance, e.g., oxidation of lactate and glucose, a major objective is to create redox active films on electrodes that are capable of efficiently mediating electron transfer between the active site and the electrode. **Figure 4** illustrates the electron transfer coupling scheme where the enzymatic reaction for the oxidation (or reduction) of the substrate, S, is linked to the electrochemical reduction (or oxidation) of the mediator ( $M_{Ox}/M_{Red}$ ). The enzyme-catalyzed electrochemical oxidation (or reduction) of the substrate is called bioelectrocatalysis. A variety of oxidoreductases including oxidases as well as dehydrogenases can be 'electrochemically wired' to electrodes in this way allowing a wide range of otherwise electrochemically inactive substrates to be detected. This approach of using voltammetry at modified electrodes underpins the development of the disposable glucose sensor.

The challenge in this field is to control both the physical architecture and chemical reactivity of the film so as to promote selected electron transfer reactions while inhibiting others. With polymer-modified electrodes (PMEs), the electrode is conferred with the molecular selectivity and specificity that is lacking at a conventional pristine electrode. For example, poly(4-vinyl)pyridine and poly(*N*-vinylimidazole) can be functionalized with osmium and ruthenium polypyridyl complexes. These synthetic macromolecules act as useful model systems for



**Figure 4** Schematic structure of a bioelectrocatalytic film on an electrode surface. S is the substrate, P the product,  $E_{Ox}$  and  $E_{Red}$  are the oxidized and reduced forms of the enzyme,  $M_{Ox}$  and  $M_{Red}$  are the oxidized and reduced forms of the mediator.

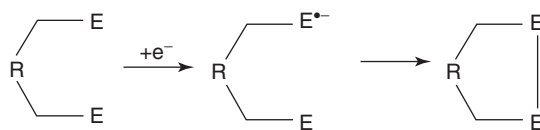
understanding processes that occur in biological systems such as proteins. Proteins regulate processes that occur at reactive centers by controlling the dynamics or energetics of the reactions. For example, while the metal center may control the thermodynamics of a process (e.g., through its redox potential or excited state properties), the protein controls the kinetics.

Redox active polymer films are ideally suited to tackling these issues. The properties of many of these electroactive polymers, e.g., their conductivity, charge distribution, shape, etc., can be changed in a controlled and reproducible way in response to environmental stimuli, e.g., a change in the nature of the contacting solution, an applied voltage, light intensity, or mechanical stress.

## Voltammetry as a Tool for Synthesis

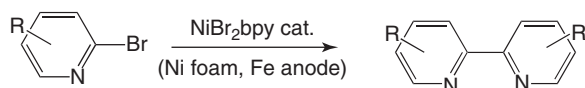
As discussed above, voltammetry allows electrons to be selectively introduced or removed from organic molecules. Therefore, it is an ideal approach for reversing the polarity of known functional groups and triggering reactions that would be difficult to drive thermally with the required degree of selectivity. For example, as illustrated in **Figure 5**, electrons can be added to electron-poor functional groups in order to convert them from electrophiles into nucleophiles. Alternatively, by poisoning the electrode at a sufficiently positive potential, electrons can be removed from electron-rich functional groups in order to convert them from nucleophiles into electrophiles. The ensuing reactive intermediates can then be trapped in order to complete reactions that involve the net coupling of either two electrophiles or two nucleophiles in ways that would otherwise be impossible. Reactions of this kind will play increasingly important roles in the construction of ever more complex 'supermolecules' and supramolecular assemblies.

For example, voltammetry has provided a mechanistic insight into the nickel-catalyzed homocoupling of halopyridines that results in significant improvements in the overall yield and product purity. Also, voltammetry has demonstrated the usefulness of sacrificial iron anode to produce iron ions that

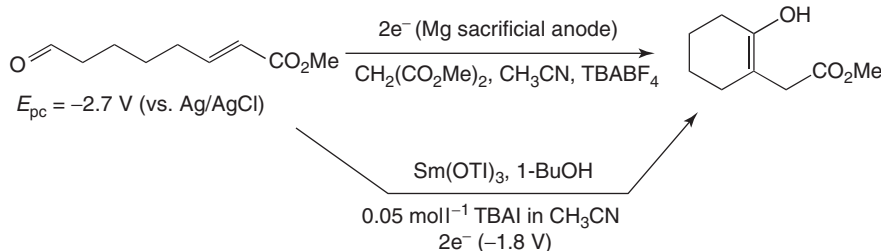


**Figure 5** Creation of nucleophiles from electrophiles by voltammetric injection of electrons followed by reaction to yield desired product.

could complex the product, thereby allowing the nickel to reenter the catalytic cycle.



The development of lanthanides as one-electron reducing agents also depended critically on voltammetry. Electrochemical investigations provided a simple and systematic means to screen the properties of redox active species allowing the optimum reagent, i.e., formal potential and fast electron exchange dynamics, to be identified for specific applications. For example, in the case of samarium(II) diiodide, cyclic voltammetry was used to determine that ligand exchange occurs between triflate and iodide when samarium(III) triflate is mixed with the tetrabutylammonium iodide (TBAI) being used as the supporting electrolyte.

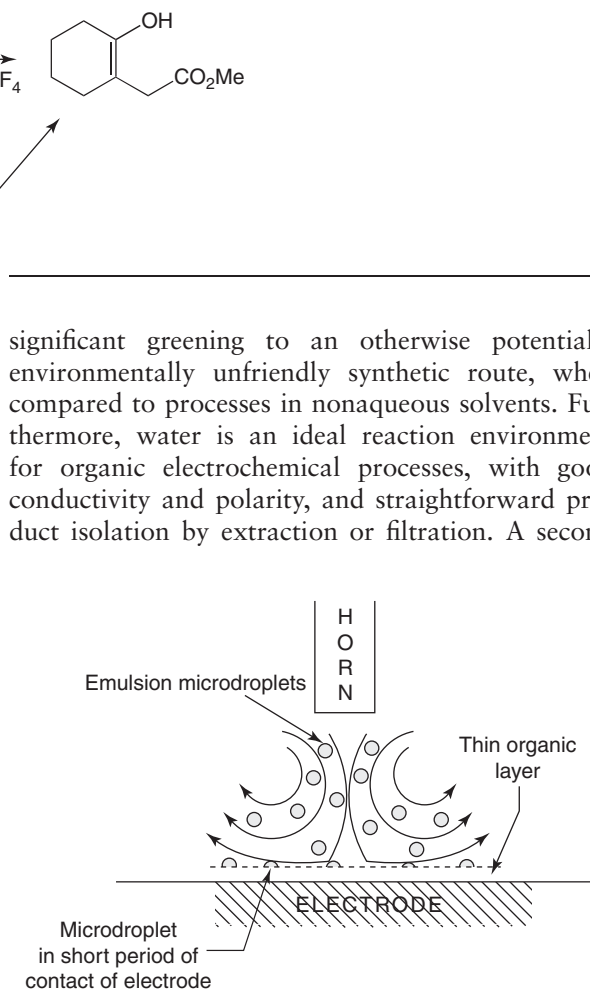


## Solvents and Electrolytic Media

Protic solvents continue to be used and new solvents, with unusually low nucleophilicity, high ionizing power, and high hydrogen bonding strength, making them good solvents for cation radicals, continue to be developed. Aprotic media are also widely used for investigating electrode reactions and synthesis. For oxidations acetonitrile (MeCN) with sodium perchlorate as supporting electrolyte or methylene chloride, sometimes with added trifluoroacetic acid and its anhydride to remove nucleophiles, has been used when MeCN was too nucleophilic and reacted with the intermediates. Liquid sulfur dioxide offers a solvent with very low nucleophilicity, but typical supporting electrolytes have only a limited solubility. For reductions *N,N*-dimethylformamide (DMF), MeCN, and dimethylsulfoxide (DMSO) have been used as media in which to perform voltammetric investigations. MeCN is slightly more acidic than DMF and DMSO, but less likely to lose hydrogen by atom abstraction.

Organic solvents may be undesirable, especially for industrial applications, and emulsions may

present an alternative. Emulsions may be simple emulsions: 'water-in-oil' or 'oil-in-water' or a 'bicontinuous microemulsion'. Such bicontinuous microemulsions are nontoxic, have a large interfacial area facilitating an intimate mixing of polar and nonpolar reagents, and high conductivity. Acoustic emulsification represents a pioneering development in the field of 'clean' technology for electrosynthesis using organics. There are generally two options by which sonoemulsion electrosynthesis may be achieved. The first involves suspending depolarizers as droplets in water, which, as illustrated in **Figure 6**, under sonication are forcefully transported to the electrode surface by acoustic streaming processes. This process may facilitate otherwise impossible voltammetry to be carried out, for example, by achieving dramatically higher mass transport rates or by removing pacifying material from the electrode surface. This approach has the additional advantage that the 'oil-in-water' emulsion that is produced provides a



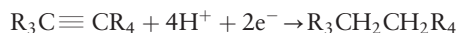
**Figure 6** Schematic representation of the thin film and droplet deposits that govern the sonoelectrochemical response for electrochemical processes in acoustic emulsion systems.

sonoemulsion electrosynthetic approach applies solely for the case of a solid electroactive compound, and consists of dissolving the depolarizer in a nonaqueous solvent, which is then emulsified with an aqueous solution as described previously.

The advantages of utilizing sonoemulsion media in which to undertake voltammetric investigations or indeed perform electrosynthesis include the following. First, the resulting emulsions are extremely stable, precluding the need for surfactants. It is believed that this stability is due to the monodisperse nature of the emulsions and to the formation of an electric double layer around the suspended droplets. Second, the separation of products from the starting materials is facile: removing the acoustic field (by, for example, switching off the ultrasonic transducer) permits the separation of two immiscible phases. The organic phase is likely to contain the products of the electrosynthetic transformation, which can then be extracted and analyzed by conventional techniques. This method permits synthesis in the absence of environmentally unfriendly nonaqueous solvents. Third, there is a large surface area for electrochemical reactions to occur.

The electrochemical reduction of unsaturated organic compounds to paraffins is usually undertaken in nonaqueous solutions. However, the use of ultrasound allows voltammetry to be performed in a more ecofriendly manner by using a sonoemulsion of the substrate in water, so eliminating the need for the nonaqueous solvent. For example, voltammograms for the reduction of diethylmaleate, diethylfumarate, and diethylacetylene dicarboxylate in the form of microscopic droplets generated by applying power

ultrasound to aqueous electrolyte media are sigmoidally shaped, indicating well-behaved responses, even at the very high rates of mass transport applied in the presence of ultrasound. These reactions correspond to the following processes:



This approach offers the possibility of investigating the voltammetric properties of organic molecules in media in which they are essentially insoluble.

*See also:* **Voltammetry:** Overview; Linear Sweep and Cyclic; Anodic Stripping; Cathodic Stripping; Inorganic Compounds.

## Further Reading

- Forster RJ (2000) Ultrafast electrochemical techniques. In: Meyers R (ed.) *Encyclopaedia of Analytical Chemistry*, pp. 10142–10171. New York: Wiley.
- Forster RJ (2003) Microelectrodes – retrospect and prospect. In: Unwin PA and Bard AJ (eds.) *Encyclopaedia of Electrochemistry*, pp. 160–195. New York: Wiley.
- Forster RJ, Vos JG, and Keyes TE (2003) *Interfacial Supramolecular Assemblies*. New York: Wiley.
- Little RD and Moeller KD (2002) *Organic Electrochemistry as a Tool for Synthesis: Umpolung Reactions, Reactive Intermediates, and the Design of New Synthetic Methods*. The Electrochemical Society, Interface.
- Lund H and Hammerich O (eds.) (2001) *Organic Electrochemistry*, 4th edn. NY: Dekker.
- Osha T (ed.) (1998) *New Challenges in Organic Electrochemistry*. Amsterdam: Gordon and Breach.

# VOLUMETRIC GAS MEASUREMENTS

**L Feltl**, Charles University, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 5441–5445, © 1995, Elsevier Ltd.

## Introduction

Gases that react chemically with solutions or solid reagents can be determined by absorption methods. Absorption in specific reagents is the basis for separation of a mixture of gases into its individual components that are then determined by different

methods. Volumetric analysis is one of the oldest techniques that are still in use.

## General Principles

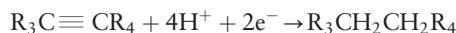
The basic principle of the volumetric system of analysis is that a measured volume of sample, at a constant temperature and pressure, is exposed in a closed vessel to a suitable solid or liquid reagent that absorbs one or more of the constituents of the sample. The extent of the loss in the sample is determined by returning the sample to its original pressure and measuring the decrease in volume. The method is based

sonoemulsion electrosynthetic approach applies solely for the case of a solid electroactive compound, and consists of dissolving the depolarizer in a nonaqueous solvent, which is then emulsified with an aqueous solution as described previously.

The advantages of utilizing sonoemulsion media in which to undertake voltammetric investigations or indeed perform electrosynthesis include the following. First, the resulting emulsions are extremely stable, precluding the need for surfactants. It is believed that this stability is due to the monodisperse nature of the emulsions and to the formation of an electric double layer around the suspended droplets. Second, the separation of products from the starting materials is facile: removing the acoustic field (by, for example, switching off the ultrasonic transducer) permits the separation of two immiscible phases. The organic phase is likely to contain the products of the electrosynthetic transformation, which can then be extracted and analyzed by conventional techniques. This method permits synthesis in the absence of environmentally unfriendly nonaqueous solvents. Third, there is a large surface area for electrochemical reactions to occur.

The electrochemical reduction of unsaturated organic compounds to paraffins is usually undertaken in nonaqueous solutions. However, the use of ultrasound allows voltammetry to be performed in a more ecofriendly manner by using a sonoemulsion of the substrate in water, so eliminating the need for the nonaqueous solvent. For example, voltammograms for the reduction of diethylmaleate, diethylfumarate, and diethylacetylene dicarboxylate in the form of microscopic droplets generated by applying power

ultrasound to aqueous electrolyte media are sigmoidally shaped, indicating well-behaved responses, even at the very high rates of mass transport applied in the presence of ultrasound. These reactions correspond to the following processes:



This approach offers the possibility of investigating the voltammetric properties of organic molecules in media in which they are essentially insoluble.

*See also:* **Voltammetry:** Overview; Linear Sweep and Cyclic; Anodic Stripping; Cathodic Stripping; Inorganic Compounds.

## Further Reading

- Forster RJ (2000) Ultrafast electrochemical techniques. In: Meyers R (ed.) *Encyclopaedia of Analytical Chemistry*, pp. 10142–10171. New York: Wiley.
- Forster RJ (2003) Microelectrodes – retrospect and prospect. In: Unwin PA and Bard AJ (eds.) *Encyclopaedia of Electrochemistry*, pp. 160–195. New York: Wiley.
- Forster RJ, Vos JG, and Keyes TE (2003) *Interfacial Supramolecular Assemblies*. New York: Wiley.
- Little RD and Moeller KD (2002) *Organic Electrochemistry as a Tool for Synthesis: Umpolung Reactions, Reactive Intermediates, and the Design of New Synthetic Methods*. The Electrochemical Society, Interface.
- Lund H and Hammerich O (eds.) (2001) *Organic Electrochemistry*, 4th edn. NY: Dekker.
- Osha T (ed.) (1998) *New Challenges in Organic Electrochemistry*. Amsterdam: Gordon and Breach.

# VOLUMETRIC GAS MEASUREMENTS

**L Feltl**, Charles University, Prague, Czech Republic

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the previous edition, pp. 5441–5445, © 1995, Elsevier Ltd.

## Introduction

Gases that react chemically with solutions or solid reagents can be determined by absorption methods. Absorption in specific reagents is the basis for separation of a mixture of gases into its individual components that are then determined by different

methods. Volumetric analysis is one of the oldest techniques that are still in use.

## General Principles

The basic principle of the volumetric system of analysis is that a measured volume of sample, at a constant temperature and pressure, is exposed in a closed vessel to a suitable solid or liquid reagent that absorbs one or more of the constituents of the sample. The extent of the loss in the sample is determined by returning the sample to its original pressure and measuring the decrease in volume. The method is based



on the Amagat law

$$\frac{V_i}{V} = \frac{n_i}{\sum_i n_i} \quad [1]$$

where  $V$  is the whole volume,  $V_i$  and  $n_i$  are the partial volume and mole fraction of component  $i$ . The partial volume of each component is proportional to its mole fraction.

The gas content in the mixture, expressed as the volume fraction and given in percent ( $p$ ), can be calculated from the relationship

$$p = \frac{V_a - V_b}{V_a} \times 100 \quad [2]$$

where  $V_a$  is the gas volume prior to absorption and  $V_b$  is the volume after absorption.

If the results are obtained at different working conditions and are to form the basis for precise, general predictions, the volume data must be reduced to standard (normal) temperature ( $0^\circ\text{C}$ ) and pressure (101 325 Pa). Thus,

$$V_0 = \frac{p_t V_t}{p_0(1 + t/273.15)} = \frac{p_t V_t \times 273.15}{p_0(t + 273.15)} \quad [3]$$

where  $V_0$  is the corrected volume,  $V_t$  is the observed volume,  $p_t$  is the barometric pressure (Pa),  $p_0$  is the standard pressure (101 325 Pa), and  $t$  is the temperature in degree celsius.

This relationship is valid only for dry gases. If the gas is saturated with water vapor from the confining fluid, the contribution of the water vapor to the total pressure must be taken into account. The relationship for reduction of the gas volume to standard conditions then has the final form

$$V_0 = \frac{p_t - p_{\text{H}_2\text{O}}}{p_0} \times \frac{V_t \times 273.15}{t + 273.15} \quad [4]$$

where  $p_{\text{H}_2\text{O}}$  is the water vapor pressure of the confining fluid at the temperature of measurement ( $t$ ).

In volumetric gas analysis it can be assumed that deviations from the ideal behavior do not introduce significant errors when molecular weights are small (e.g.,  $\text{O}_2$ ,  $\text{N}_2$ ,  $\text{CO}$ ,  $\text{CH}_4$ ). Only for hydrocarbons with more than two carbon atoms and for carbon dioxide must a correction be made according to Table 1.

## Confining Fluids

A gas may be quantitatively transferred from one chamber to another by displacing it with a liquid. There are two basic requirements for a confining liquid. It must not dissolve the gas to be determined or chemically react with it, and its viscosity and

**Table 1** Weights of 1000 ml of certain gases at  $0^\circ\text{C}$  and 101 325 Pa

Gas	$\text{g l}^{-1}$
$\text{CO}_2$	1.9632
$\text{C}_2\text{H}_6$	1.3408
$\text{C}_2\text{H}_4$	1.2507
$\text{C}_2\text{H}_2$	1.1608
$\text{C}_3\text{H}_8$	1.9660
$\text{C}_3\text{H}_6$	1.8665

wetting of glass must be low. The best confining fluid is mercury: its vapor pressure is extremely low and it is inert (although chlorine, bromine, hydrogen sulfide, and sulfur dioxide are exceptions to this rule). Mercury also has drawbacks: it is quite dense, it is poisonous, and it is expensive. Thus, it is often replaced by various salt solutions, even though these solutions do not completely fulfill the above conditions. In most of these aqueous solutions gases are absorbed (although to a limited degree).

Often a 22% (w/v) solution of sodium chloride, acidified with a little sulfuric or hydrochloric acid and colored with methyl orange or methyl red, is used. This procedure greatly decreases the solubility of certain gases (especially carbon dioxide). Before the analysis it is necessary to saturate the confining fluid by the gas to be analyzed. If this precaution is taken the errors caused by solubility losses are less than is commonly supposed.

## Absorption Reagents

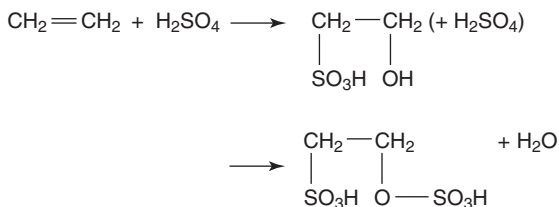
The individual gases are absorbed in absorption reagents, which are solutions or solid substances. The chemical reaction on which the absorption is based must be sufficiently fast, quantitative, and selective, and the absorption capacity of the reagent must be sufficiently large.

For carbon dioxide, a 25–40% solution of potassium hydroxide has been recommended. The 25% solution is least viscous and it absorbs  $\sim 40$  ml of  $\text{CO}_2$  per milliliter. Potassium hydroxide is always preferred to sodium hydroxide, as the potassium carbonate formed is more soluble than sodium carbonate. This reagent dissolves all acid gases, which must be removed before the determination of  $\text{CO}_2$ . In the presence of other water-soluble gases solid reagents must be used over mercury as the confining fluid for this determination. A typical example is ascarite (asbestos soaked with sodium hydroxide).

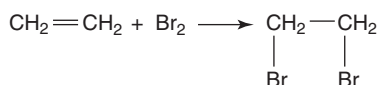
Unsaturated hydrocarbons are mostly absorbed in 'activated' sulfuric acid, i.e., acid with dissolved silver or mercury(II) sulfate, which does not dissolve



saturated hydrocarbons. Often, fuming sulfuric acid is also recommended, but higher alkanes such as propane or butane are also dissolved:

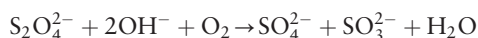


A solution of potassium bromide nearly saturated with bromine dissolves all unsaturated hydrocarbons but it also dissolves traces of saturated hydrocarbons of C<sub>3</sub> and above:



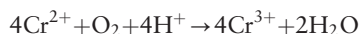
A number of reagents have been recommended for oxygen. The reagent most often used is an alkaline solution of pyrogallol. The solution is usually prepared directly in the absorption pipette with constant bubbling with nitrogen to prevent the entrance of air. The absorption capacity of alkaline pyrogallol solution is ~12 ml of O<sub>2</sub> per milliliter. The absorption is rapid in any mixture where the oxygen content is less than ~20%. It is known that pyrogallol solution has a tendency to form carbon monoxide when the solution is exhausted or if it is insufficiently alkaline.

Oxygen can also be absorbed in an alkaline solution of sodium dithionate:

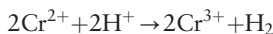


It evolves no gases, but absorption is slower than in an alkaline pyrogallol solution. It can be accelerated by the addition of sodium anthraquinone-2-sulfonate or indigo carmine.

The strong reducing power of chromium(II) chloride solution in an acidic medium is the basis of a reagent for oxygen absorption, based on oxidation of chromium(II) to chromium(III):

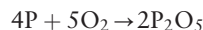


It does not dissolve acidic gases (carbon dioxide, hydrogen sulfide). If the pH is too low, chromium(II) slowly reduces hydrogen ions:



For this reason it is advisable to prepare chromium(II) chloride solution in a weak acid solution (e.g., acetic acid).

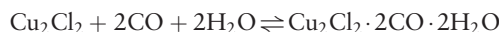
Moist yellow phosphorus is used to determine higher oxygen concentrations. This reagent has enormous capacity and can be used for oxygen contents up to 60%:



As the yellow modification of phosphorus is readily converted to the red form in the presence of light, the reagent must be protected in brown glass pipettes. Small quantities of unsaturated hydrocarbons inhibit the reaction.

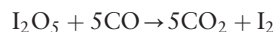
Ammoniacal copper(I) chloride is an excellent absorption reagent and can be used to determine oxygen in concentrations of over 80%, provided carbon monoxide, unsaturated hydrocarbons, and acidic gases are absent.

The usual absorption reagent for carbon monoxide is a solution of copper(I) chloride in acidic, neutral, or ammoniacal medium. The absorption involves the formation of a coordination compound:



This reaction is an equilibrium, so it is impossible to remove the last traces of carbon monoxide with a partially exhausted reagent. Thus, absorption is carried out in several absorption pipettes to ensure quantitative absorption of carbon monoxide. The most commonly used absorbent is an ammoniacal solution of copper(I) chloride that contains ammonium chloride. This solution is stored over metallic copper. Instead of copper(I) chloride, copper(I) sulfate can be used. Its efficiency can be increased by addition of 2-naphthol. The addition compound formed, Cu<sub>2</sub>SO<sub>4</sub>·2CO, is more stable than the corresponding chloride.

Carbon monoxide can also be absorbed in a suspension of iodine pentoxide in fuming sulfuric acid, and oxidized to carbon dioxide:

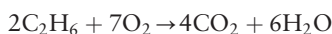
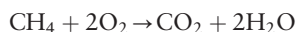
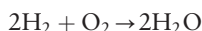


Other gases are absorbed and determined by volumetric methods only in special cases. A solution of palladium(II) chloride or colloidal palladium and picric acid absorbs hydrogen, a solution of potassium dichromate is used for sulfur dioxide, and a solution of cadmium acetate or alkaline solution of hydrogen peroxide for hydrogen sulfide.

## Combustion Analysis

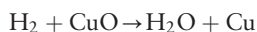
For gases that cause difficulties in absorption volumetric analysis (for which there is no suitable

absorption agent), combustion methods of analysis are employed. They are used primarily to determine hydrogen and saturated hydrocarbons. Combustion methods are based on total oxidation to water and carbon dioxide. The water formed is condensed and has a negligible volume (e.g., 100 ml of hydrogen is burned to form only 0.07 ml of liquid water) and carbon dioxide is determined by absorption in potassium hydroxide. The composition of the sample is calculated on the basis of the change in the volume of the sample and determination of carbon dioxide



The total contraction in volume after combustion of one volume of hydrogen is 1.5, in the case of methane it is 2.0, and for ethane it is 2.5.

The necessary oxygen must be provided either as air or better as pure oxygen or bonded in a metal oxide. Combustion can be carried out in three ways: explosion, slow combustion, and fractional combustion. The simplest technique for total combustion is explosion. A mixture of the gas sample and known amount of oxygen is ignited by an electric discharge. The slow combustion technique utilizes the catalytic effect of a hot platinum or palladium spiral. Today, fractional combustion is usually performed by means of copper(II) oxide or 'Hopcalite' (mixture of silver oxide, mercury(II) oxide, and silver permanganate). The oxygen required is supplied directly by metallic oxides, i.e., solid substances, resulting in considerable simplification in the analysis, both in the experimental arrangement and in the calculation of the contents of the individual gases. The reaction occurs according to



The relatively large differences in the combustion temperatures permits quantitative separation of some gas mixtures (hydrogen, carbon monoxide, methane).

It is evident that absorption reagents are not very selective and then the suitable choice of absorption reagents and, mainly, the order in which the individual gases are to be determined, are important. If complete analysis of common industrial gases by a volumetric method is of interest, then the following procedure must be employed. First, carbon dioxide is determined as it is readily soluble in all alkaline solutions, unsaturated hydrocarbons must be removed as the second component, and then

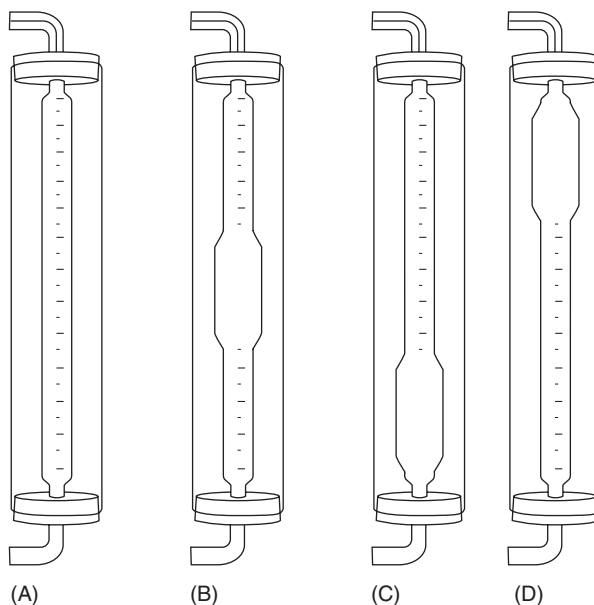
oxygen can be absorbed and determined by absorption of carbon monoxide. The other gases that are not absorbed (hydrogen, saturated hydrocarbons) can be determined by a combustion technique. The nitrogen content is found after removing the combustible components from the difference to 100%.

## Equipment

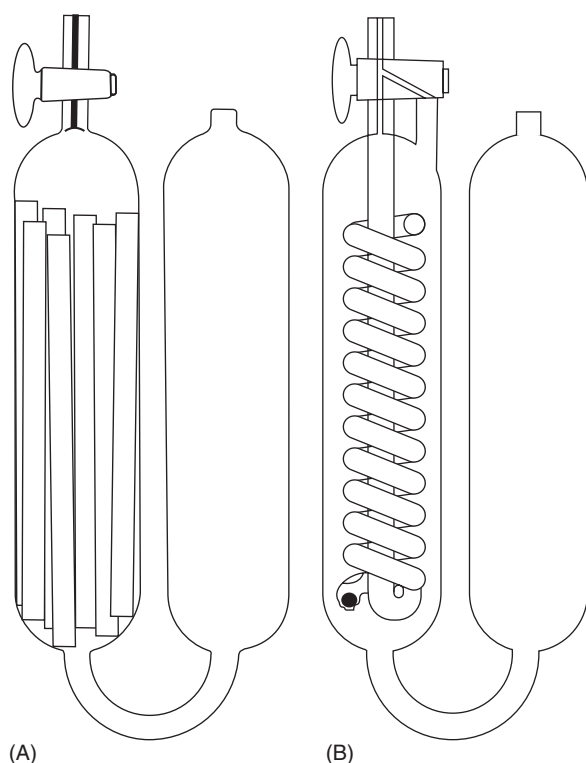
For exact measurement of gas volumes gas burettes can be used. In these burettes volume changes after absorption of gas in a suitable absorption reagent are also measured.

The classical burettes are the Hempel burette and the Winkler, Bunte, and Tutweiler burettes in various modified versions. The most widely used burettes are those employed in Orsat instruments (Figure 1). The burette consists of a glass cylinder graduated upwards from 0 to 100 ml. The water jacket surrounding the burette is generally sufficient to protect the burette from temperature changes. The sample is measured and the gas transferred to the absorption solutions using a confining fluid located in the leveling vessel. At the top is a stopcock so that the burette may be connected to the atmosphere (or to the source of the gas sample) or to any of the absorption pipettes.

Gases are sometimes absorbed directly in the burette (in the case of Bunte and Tutweiler burettes), but usually they are absorbed in absorption pipettes.



**Figure 1** Burettes employed in Orsat instruments: (A) common type; (B) for coal gas; (C) for mixtures of gases with high content of  $\text{CO}_2$ ; (D) for smoke gases.



**Figure 2** Absorption pipettes: (A) bubbler pipette; (B) Heinz pipette.

The prime function of the absorption pipette is to store the absorbent safely, away from contact with the atmosphere. Good absorption pipettes must be constructed so that absorption occurs over a large surface area, and is fast and quantitative. Some types of absorption pipettes employed in Orsat instruments are depicted in **Figure 2**.

For the complete analysis of gas mixtures the Orsat apparatus is commonly used. It consists of a volumetric burette with water jacket, leveling vessel, various numbers of absorption pipettes, and a quartz tube with pieces of copper(II) oxide or combustion pipette with a platinum spiral. This instrument is based on stepwise removal of the individual gases from the total sample volume, measured in a gas burette. Gooderham's modification is very interesting. The gas mixture continuously passes through a system of soap-film meters and absorption pipettes. The volume change after absorption in the individual absorption reagents is found from the change in the flow rate.

## Sources of Errors

The precision and accuracy of determinations by absorption volumetric analysis is influenced by the

system errors. Some of the factors directly affect the reading of the volume, while others are a result of the absorption process. There are three main sources of errors related to:

1. the confining fluid;
2. changes in temperature, atmospheric pressure, and vapor pressure; and
3. absorption reagents.

In the confining fluid, gases are physically dissolved according to Henry's law. The amount of gas dissolved in a given volume of liquid is proportional to the partial pressure of the gas. This is very important in the sampling and analysis. The confining fluid must be saturated with a particular gas mixture of a narrow range of composition. The data in **Table 2** provide a picture of the solubilities of various gases in water and brine solutions. To overcome the solubility of gases in absorption solutions it is also necessary to contact a small portion of the sample with all solutions immediately before the analysis.

Volumetric measurements are greatly dependent on temperature changes. A temperature change of  $1^{\circ}\text{C}$  produces a change of the volume of  $\sim 0.3\%$  (under isobaric conditions). To decrease this effect, the laboratory should at least face north to minimize temperature changes or be thermostatted or air-conditioned, and the measuring burette should be placed in a water jacket. In addition to direct volume changes, changes in the water vapor pressure can result from temperature fluctuations. If the dry gas sample is analyzed, then absorption is accompanied by an increase in the volume concentration of the water vapor corresponding to its vapor pressure at a given temperature, e.g., at  $20^{\circ}\text{C}$  this increase is  $\sim 2.3\%$  (see **Table 3**).

For analysis performed within a short period of time, it is assumed that the atmospheric pressure is constant. If the procedure of the analysis takes several hours or days (a pressure change of 133.32 Pa produces

**Table 2** Solubilities of various gases at  $20^{\circ}\text{C}$  in water and saturated NaCl solution (milliliter of gas in 1000 ml water at a partial pressure of 101 325 Pa)

	$\text{O}_2$	$\text{CO}$	$\text{CO}_2$	$\text{CH}_4$	$\text{C}_2\text{H}_6$	$\text{C}_2\text{H}_4$	$\text{C}_2\text{H}_2$
$\text{H}_2\text{O}$	31.5	23.2	878	33.1	47.2	122	1030
Saturated NaCl			280 <sup>a</sup>				
			140 <sup>b</sup>				
			70 <sup>c</sup>				

<sup>a</sup>From pure  $\text{CO}_2$ .

<sup>b</sup>From a mixture containing 50%  $\text{CO}_2$ .

<sup>c</sup>From a mixture containing 25%  $\text{CO}_2$ .

**Table 3** Vapor pressure above water and some salt solutions

Temperature (°C)	Vapor pressure (Pa)		
	H <sub>2</sub> O	Sat. solution of NaCl	Sat. solution of Na <sub>2</sub> SO <sub>4</sub> ·10H <sub>2</sub> O
15	1700	1270	1600
20	2340	1740	2170
25	3170	2350	2930

a volume change  $\sim 0.13\%$ ), a compensator of the Pettersson type is used. However, it is more common to employ a suitable correction to the measured results using eqn [4].

One important factor, which can also be a source of errors, is the separation process. The absorption can be incomplete, because some absorption reagents are slow in their action. These errors can be avoided by using a suitable instrumental design of the absorption pipettes and repeated absorption in one or two absorption pipettes. In the analysis of a gas mixture it is very important to maintain an exact analysis order.

If the experimental conditions are suitably adjusted and the necessary correction is introduced, the error in determination by absorption volumetric analysis is in the range 0.2–2%.

## Volumetric Analyzers

Volumetric analyzers are basically mechanized Orsat instruments. They work periodically and are mainly used for determining a single component, usually carbon dioxide, sulfur dioxide, ammonia, hydrogen sulfide, hydrochloric acid, or acetylene. In some instruments there are also accessories for combustion, which permit the determination of combustible gases or analysis of mixtures such as  $\text{CO}_2 + \text{CO} + \text{H}_2$  or  $\text{CH}_4 + \text{CO} + \text{H}_2$ . Their sensitivity is relatively poor. They are most often used in plants for control of combustion processes.

See also: **Carbon. Process Analysis:** Overview.

## Further Reading

- Guerin H (1952) *Traité de Manipulation et d'Analyse des Gas*. Paris: Masson.
- Heron AE and Wilson HN (1959) Gas Analysis. In: Wilson CL and Wilson DW (eds.) *Comprehensive Analytical Chemistry*, vol. IA. Amsterdam: Elsevier.
- Jeremina BG (1955) *Gazovyi Analiz*. Leningrad: Goskhimizdat.
- Mullen PW (1955) *Modern Gas Analysis*. New York: Interscience.
- Smolková-Keulemansová E and Feltl L (1991) *Analysis of Substances in the Gaseous Phase*. Amsterdam: Elsevier.

# W

## WATER ANALYSIS

Contents

**Overview**

**Potable Water**

**Freshwater**

**Seawater – Organic Compounds**

**Seawater – Dissolved Organic Carbon**

**Seawater – Inorganic Compounds**

**Industrial Effluents**

**Sewage**

**Organic Compounds**

**Biochemical Oxygen Demand**

**Chemical Oxygen Demand**

**Oil Pollution**

**Microbiological**

**Algal and Microbial Toxins**

**Particle Characterization**

### Overview

**H van der Jagt**, KIWA Water Research, Nieuwegein, The Netherlands

**H F Reijnders**, RIVM, National Institute of Public Health and Environmental Protection, Bilthoven, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by H F Reijnders and R C C Wegman, pp. 5446–5461, © 1995, Elsevier Ltd.

### Introduction

Water is an essential resource for all living species, including man, and it is necessary to analyze a range of natural waters (e.g., seawater, rainwater, surface water, and groundwater), polluted waters (e.g., industrial water and sewage water), and purified waters (e.g., potable water and distilled water). This article presents an overview of sampling and sample treatment procedures and discusses a range of analytical techniques used for the quantification of major, minor, and trace constituents.

### Sampling and Sample Treatment

#### Sampling

The collection of water samples and subsequent sample treatment depend primarily on the type of water that is collected and the aim of the analysis. It is important that samples are collected in such a way that they are representative of the water body from which they are taken. This means that the concentrations of the determinand and interferences in the sample should not differ from the concentrations present in the original water body. The need for representatives may limit the volume of sample that may be collected; for example, to take into account a sharp gradient in analyte concentrations relatively small sample volumes may be taken without disturbing the original gradient.

Sample integrity is also very important. Care must be taken to prevent contamination of the sample or sorption of the determinand. Quality control procedures must be applied to guarantee that contamination or sorption has not occurred. Analysis of samples that have not been taken or stored properly is useless. Results of such analyses may lead to false data



and erroneous conclusions following interpretation of the results.

### Sample Preservation

Ideally samples should be analyzed *in situ* or at the site of sampling in the field. If direct measurements are not possible or are too expensive, samples should be analyzed as soon as possible to avoid the need for preservation. However, samples cannot always be analyzed directly and they may have to be stored for so long that preservation is necessary. The preservation of water samples is covered in ISO standard 5667/3 (1985). The preservation methods described in this standard include time limits for sample storage and analysis; specifications for container material; prevention of exposure to light; temperature control (2–5°C); pH control (addition of sulfuric, nitric, or phosphoric acid, or sodium hydroxide); addition of special reagents (e.g., ethylenediaminetetraacetic acid (EDTA), copper sulfate, zinc acetate, formaldehyde) to retard biological activity, hydrolysis of compounds and complexes; and measures for reducing volatility of compounds and sorption effects.

### Storage

As stated above, preservation and storage of samples should be avoided if possible. When storage is necessary, sample preservation by use of low temperatures is preferred to the addition of chemicals. Low temperature maintains sample integrity while the addition of chemicals may cause interferences or contamination in the sample. However, when samples are cooled, the determinand may crystallize or absorb on any precipitates that are formed; storage at 4°C is therefore preferred. Samples should be stored in chemically inert gas-tight containers of low adsorptivity and kept in the dark to prevent growth of bacteria and algae. Air contact is minimized by filling the sample bottle to the rim before closing. In this way the redox state can be kept stable for a longer period (no oxygen influence). Samples should be uniquely identified and storage conditions (duration, temperature, etc.) recorded to minimize the possibility of errors. Automated systems for laboratory management (e.g., laboratory information and management systems, barcodes for identification) are recommended.

### Pretreatment

Depending on the planned use of the results, pretreatment of the sample may be necessary. As matter initially suspended in the sample may agglomerate as solid matter, the suspended matter must be separated as soon as possible. This separation is most commonly done by filtration, which should ideally be

carried out directly after sampling at the sampling site. A 0.45 µm membrane filter is internationally accepted to achieve a separation between dissolved and suspended matter. Particles smaller than 0.45 µm can pass through the filter and are treated with the dissolved matter. These tiny particles may contain the greatest part of the trace constituents owing to their high adsorptivity and large surface area. Procedures have been designed to correct for the contributions made by small particles to the true dissolved content. The chances of contamination of sample during pretreatment are lower in the laboratory; pretreatment in the field requires some experience. When samples cannot be filtered in the field, they should be filtered on arrival in the laboratory. Addition of chemicals to dissolve a particular determinand from the solid phase generally causes contamination of the sample and increases the variability of the analytical result.

## Analytical Techniques

Water analysis is mostly conducted using physicochemical methods. A limited number of physical parameters (temperature, electrical conductivity, redox potential) and test parameters (such as color, smell, and taste) are measured in water analysis. Chemical/physicochemical methods can be used to analyze for a wide range of components, both inorganic and organic. Spectrochemical or electrochemical methods are mainly used for the analysis of inorganic constituents, while chromatographic methods are applied for the analysis of organic components.

### Physical and Physicochemical Analysis

Formerly, color, smell, and taste were assessed by sensorial testing by panels or experts. Color is now determined by spectrophotometric methods. Following a standard protocol, the extinction is determined at several wavelengths. The color characteristics can be quantified in an automated procedure.

Although it is generally thought that smell can only be estimated qualitatively, recent developments suggest that quantitative measurements are possible. In one method a series of seven dilutions of the sample is prepared with odor-free water. The series of samples, in which each sample is randomly paired with a blank, is offered in ascending order for testing by a panel of at least eight people. For each pair a panelist chooses which sample has the greatest odor. The results of the test are incorporated into a statistical evaluation procedure to determine the threshold dilution number. Interlaboratory variation of less than 20% is possible.

Taste can be determined similarly by organoleptic test panels. As for smell determination, considerable effort is required to obtain accurate results. Again,



interlaboratory variations of less than 20% are possible. As both smell and taste provide qualitative information, valuable indications are derived for analyzing the sample.

Temperature can be measured very easily by negative-temperature-coefficient resistors or temperature-sensitive transistors in small probes. These devices can also be combined with other sensors. Such combined sensors allow the direct correction of the measured parameter for temperature change.

The electrical conductivity can be measured over a wide range of concentrations. The size of the measuring cells is smaller than the apparatus used in the past. A wide range can be accommodated by using three-electrode measuring cells and automatic ionic strength adjustment of the frequency of the alternating current that is applied across the cell for phase-sensitive measurement. Alternating current is used to prevent polarization of the electrode surface. Electronic correction for temperature change is commonly applied in modern apparatus.

Measurement of redox potentials is relatively inaccurate. As the redox potential measurements are easily influenced by adsorption of organic substances such as minute amounts of oil, it is essential to clean the electrode surface. This is best done with a mixture of calcium carbonate, detergent, and trisodium phosphate.

The measurement of pH is a complex subject. Corrections have to be made for the temperature coefficient of the buffer solutions used for calibration of the meter and the difference between the temperature at which the instrument is calibrated and that at which the measurements are made. This is done using microelectronics. Efficient electrodes with high outflow of electrolyte from the reference electrode are used to overcome errors due to liquid junction potentials that may occur, especially in the case of low-ionic-strength water (de-ionized condensate, rainwater, etc.). To avoid the influence of carbon dioxide, ammonia, sulfur dioxide, etc., from the ambient air, a flow-through electrode in a flow-through system can be used for the pH determination. The latest development is the commercial production of ion-selective field effect transistor pH electrodes, which are much less fragile than pH electrodes with a glass membrane that is sensitive to acid. Glass electrodes are still very common. Because of the small size of these electrodes, smaller sample volumes are needed and measurements with probes are possible. This is an advantage, for example, in the *in situ* determination of groundwater or monitoring the pH of rainwater during showers.

Dissolved oxygen is now almost exclusively measured amperometrically with an oxygen electrode.

The apparatus is equipped with a potentiostat and a three-electrode system, which is separated from the sample by a membrane that passes oxygen to the inner compartment of the electrode. In the inner compartment there is a solution with an electroactive substance (usually bromide). A three-electrode measuring cell is used to allow stabilization of the response. A three-electrode system allows determination when an intermediary electroactive substance is used and for subsequent regeneration of this substance. Corrections for ionic strength and temperature can be made automatically. Dissolved organic substances such as phospholipids from microorganisms, detergents, and oil interfere because these substances reduce the permeability of the membrane.

### Groundwater Analysis

Generally, groundwater is sampled from filters in groundwater wells, where it is usually pumped up with an impeller pump. Before sampling, the filters are cleaned with the groundwater to be sampled. The volume of water used for cleaning is about three times the void volume of the filter and connection tube. Applying an under pressure to the well collects samples using this procedure. Often, solute gases like carbon dioxide are expelled, causing the pH of the sample to rise. Next, the sample is filtrated; for example, through a 0.45  $\mu\text{m}$  pore size cellulose nitrate membrane filter. Although it is internationally accepted that a pore size of 0.45  $\mu\text{m}$  can distinguish between suspended and solute substances, colloidal particles have been shown to cause problems here. These particles are solid and can easily slip through the 0.45  $\mu\text{m}$  membrane filters. The sorption of substances is also unpredictable when using filters. Furthermore, the transport of samples to the laboratory takes place under unconditioned circumstances and transport times are often long and variable. The pH shift, filtration, and variable transport times cause drastic changes in the chemical composition of the sample. The changes restrict the representativeness of the sample of groundwater originally present in the well.

Although the representativeness of the original groundwater may be poor, samples are still analyzed to obtain an impression of the condition of the groundwater. Since the chemical composition of groundwater highly resembles the composition of drinking water, groundwater is analyzed in almost the same way as drinking water. However, tritium is usually not measured when drinking water is analyzed. Tritium in groundwater can be used to assess the time groundwater needs to infiltrate the soil from surface level to the filter of a well. Tritium found in

groundwater originates from nuclear tests carried out from 1962 to 1965 (highest in 1963). Having been washed from the atmosphere through precipitation, the tritium ended up in groundwater. As the decay time of tritium is 12.43 years, its activity can be measured over a long time period. Transport times can be calculated from the time series for tritium monitored in a well filter. Although these transport times can be assessed from individual measurements, the accuracy is then much less. Tritium is measured through the decay products, electrons, and helium. Normally, beta radiation is measured to establish the concentration of tritium. Water containing tritium is concentrated to improve the accuracy of the measurements in groundwater. First, a portion of the sample is distilled to remove contaminants. This is followed by enrichment during a 5-day electrolysis of the sample. Sodium hydroxide is added to 150 ml of the sample to increase the conductivity. The sample is then placed in an electrolysis cell, and cooled in a freezer. To prevent absorption of water from the ambient air, the electrolysis cell is trapped on a zeolite absorption column. After electrolysis, 15 ml of the sample remains. This is converted to hydrogen through the reaction with elemental magnesium. The resulting tritium-containing hydrogen gas is catalytically converted with ethene to ethane, using platinum as catalyst. The gas is then brought into a proportional counter and activity counted. On each run, both a background sample and a sample with known activity are analyzed. An alternative way is to add the sample to a portion of degassed water and leave the sample aside for about half a year. Helium is formed and can be measured by mass spectrometry. From the result the amount of tritium original present the tritium concentration is derived.

### Inorganic Chemical Analysis

**Minor constituents** Interest in sensor techniques is growing in environmental analysis. Coupling of sensors with simple telemetric systems facilitates monitoring with remote fast data processing. Moreover, flexible monitoring networks can be realized because sensors are easily transportable. Possible changes in the sample between sampling and analysis and much logistical effort in carrying out analysis in the laboratory are obviated.

One of the most powerful methods available is ion chromatography (IC). In IC, separations are accomplished by passing a small volume (20–250  $\mu\text{l}$ ) of the sample through a column filled with a low-capacity ion-exchange resin. For anion chromatography, colloidal anion-exchange medium is agglomerated as a thin surface shell on sulfonated polystyrene-

divinylbenzene particles. The column is eluted with a sodium hydrogencarbonate/carbonate solution. After elution in a suppressor system, sodium is exchanged with hydrogen, by which hydrogencarbonate/carbonate is converted to carbon dioxide, which has low conductivity. By an analogous reaction, anions are converted to their respective acids, which can be detected conductometrically very sensitively owing to the high ionic mobility of the hydrogen ion in water. Concentrations as low as  $100\mu\text{g l}^{-1}$  can be determined without any treatment of the sample. IC is routinely applied for the determination of chloride, nitrate, and sulfate. Fluoride, nitrite, and phosphate are increasingly determined by IC. IC is less widely applied for the determination of cations, owing to the lower sensitivity of IC compared with atomic absorption spectrometric methods.

A serious problem with IC is insufficient separation. This occurs mostly in the analysis of samples with a complex matrix (e.g., in the presence of complexing organic molecules) like groundwater, percolation water from tipped refuse, potable water after ozonization or chlorination, or when substances are determined in extremely low concentrations. In the first case, sample pretreatment (e.g., addition of eluent, removal of constituents that are present in high concentrations, etc.) is necessary, but this increases the cost of analysis and there is an additional risk of contamination.

Flow-through methods are often applied in water laboratories and can be performed in segmented mode (2–3 mm bore with separation of samples in the tube effected by air bubbles) or in nonsegmented mode (0.3 mm bore). In flow-through systems the determinand is generally derivatized and the colored product is measured photometrically. The segmented system is usually designated as continuous flow (CF). The system working with nonsegmented streams is generally termed flow injection (FI). FI methods allow the samples and reagent zones to be manipulated under thermodynamically nonequilibrium conditions. This makes it possible to use unstable reagents and reaction products. The selectivity is enhanced by kinetic discrimination when the determinand can be measured selectively because of differences in reaction rate. Gradient calibration techniques can also be applied. Moreover, FI enhances the performance of an analytical method through the contamination-free handling environment (addition of reagents inside a closed system), use of small volumes of sample (typically 0.1 ml) and amounts of reagents, and precise sample injection. Apart from these advantages, the introduction of FI into a laboratory with a heavy workload is confronted with difficulties in applying kinetic methods because controlled mixing and

timing is required for the reagent addition and measurement stages. Accordingly, CF rather than FI methods are mainly used in routine laboratories and many methods are already available.

**Discrete analyzer** Discrete analyzer is increasingly applied for practical chemical analysis of water. The analyzers facilitate and speed-up sample processing and reagent handling. Continuous loading of samples increases the efficiency. Cooled reagent storage and the large cuvette allow increased throughput of samples.

The advantage of discrete analyzers is that sample crossover in the system itself is the lowest possible. Volumes of 75  $\mu\text{l}$  of reagent and sample volumes as large as 100  $\mu\text{l}$  are sufficient. In an automated system with a throughput of 200 determinations per hour in the same sample 6 to 10 components (such as ammonium, alkalinity, aluminum, boron, bromide, calcium, chloride, chromium(VI), cyanide, fluoride, iron, magnesium, nitrate, nitrite, phosphate, etc.) can be determined. In discrete analyzers normally conventional spectrophotometric methods are used. These methods are prone to interference of the matrix of the sample. As a good concept for interference studies still is not available, interferences are as yet not sufficiently studied systematically even for routine analyses.

Another promising alternative technique for the determination of anions is capillary zone electrophoresis (CZE). In this technique, a DC voltage (typically 30 kV) is applied over a capillary of typical dimensions 60 cm  $\times$  75  $\mu\text{m}$  filled with a buffer containing (among other things) chromate and an osmotic flow modifier (a proprietary cationic surfactant). The electroosmotic flow and electrophoretic mobility cause the separation of the ions. The separated ions displace chromate locally. When the stream is monitored photometrically, these displacements are registered as peaks whose areas are a measure of the concentration. This is an advantage compared with isotachopheresis (from which CZE is derived), in which the length of the signal plateau is a measure of the concentration. Less time is needed in CZE for a determination and standard chromatographic data-processing computers can be used. CZE can also be used for the determination of cations, in which case another ultraviolet (UV)-absorbing substance with optimal electrophoretic mobility is added and the polarity is switched. The sample volume is  $\sim 50$  nl and only small amounts of reagents are needed for analysis with CZE. The sensitivity is usually a factor of 1000 less than that of IC, but concentration by electromigration is possible. Although CZE is said to be matrix independent, sources of error have not yet been studied in any detail, but the technique seems

promising for efficient analysis of main components in several types of water (time of analysis 5 min; costs US\$20.00 per component).

**Standard methods** Chloride is still frequently determined titrimetrically (argentimetry with potentiometric endpoint detection). But chloride can also be determined in a flow-through system with photometric detection. Chloride releases thiocyanate from mercury(I) thiocyanate, which reacts with Fe(III) to form a colored complex.

Sulfate is still determined in CF systems with barium and a metallochromic indicator such as dimethylsulfonazo(III) (DMSA(III)) or methylthymol blue. As DMSA(III) is a metallochromic indicator, interfering cations should be removed prior to the determination step. Such pretreatment can be done in the flow-through system. The method suffers interference only from high concentrations of potassium, phosphate, hydrogen ion, chloride, and nitrate. As a consequence the method performs favorably when applied to groundwater, wastewater, or percolation water from tipped refuse.

Nitrate is commonly determined in a flow-through system through a Griess (diazo) compound after reduction to nitrite in a cadmium or copperized cadmium reduction column. The diazo-coupling reaction with nitrate is very selective. The reduction step is a major source of error; the reduction efficiency should be repeatable within 5% and be higher than 80% and this must be ascertained by frequent analysis of standards. In a CF system a throughput of 90 samples per hour is possible. The precision is better than 1.5% in the range 1.0–5.0  $\text{mg l}^{-1}$  nitrate. The method suffers interference from dissolved organic substances that produce a high background color in a sulfuric acid medium. Salinity and reactive chlorine compounds remaining after chlorination also affect the signal. Nitrate is also determined spectrophotometrically after reaction with diphenylamine with an excess of sulfuric acid.  $\text{NO}_2^-$ ,  $\text{ClO}_3^-$ ,  $\text{BrO}_3^-$ ,  $\text{IO}_3^-$ , Mn(VII), Cr(VI), V(V), Fe(III), and Mo(VI) interfere. For most types of water, however, the interference is often negligible since these components are not present in sufficiently high concentrations.

Nitrite may be determined through the reaction with sulfanilamide and *N*-(1-naphthyl)ethylenediamine in acidic solution. The lower limit of detection of this method is 30  $\mu\text{g l}^{-1}$  nitrite. Because of its simplicity and reliability, photometric detection is still preferred.

**Kjeldahl nitrogen** For a long time organically bound trivalent nitrogen was measured according to the method of Kjeldahl, where the water sample is

evaporated to dryness and digested with sulfuric acid. Potassium sulfate is added to elevate the destruction temperature. The destruction residue is neutralized and ammonium is measured to establish the Kjeldahl nitrogen content. To calculate 'organic nitrogen', the result of the determination of Kjeldahl nitrogen has to be corrected for the amount of ammonium present in the sample. Kjeldahl used mercuric ions to catalyze the efficient oxidation of protein-like substances. The intoxication caused to humans in Japan in the 1953–60 periods after ingestion of mercury-containing fish from the Minamata Bay gave an impulse to highly reduce the use of mercury in the laboratories. Through this incident, first mercury was replaced by cupric ions, and later, selenium was used as catalyst, although their effectiveness remained controversial. Next the use of catalysts was abandoned in favor of potassium peroxodisulfate. In this procedure sulfuric acid is added to the sample and the sample is evaporated. After cooling potassium peroxodisulfate is added for the destruction of organic matter. The residue then had to be warmed to a temperature of 370°C for completion of the destruction. The determination is completed by the measurement of ammonium. Despite its rationalization, this method remains very laborious. An alternative for the measurement of total nitrogen is the combustion of the sample in an argon–oxygen atmosphere in the presence of ceric oxide at 1000°C. The produced nitrogen oxides are converted to nitrogen oxide, which can be determined electrochemically or by chemoluminescence. In electrochemical detection, nitrogen oxide reacts with the nitrate ion at the working electrode. Chemiluminescent detection is achieved when nitrogen oxide reacts with ozone to form nitrogen dioxide with the emission of light. The amount of light, measured with a photomultiplier, is proportional to the amount of nitrogen present in the sample. This measurement takes about 3 min per sample.

Phosphate is mostly determined photometrically through the reaction with ammonium heptamolybdate in acidic medium in the presence of potassium antimonyl tartrate as a catalyst. 12-Molybdophosphoric acid is formed that upon reduction with, for example, ascorbic acid has a blue color. The lower limit of detection is  $10\text{ }\mu\text{g l}^{-1}$ . Chromium(VI), nitrite, and high concentrations of chloride and iron interfere. The lower recoveries caused by iron are eliminated by addition of hydrogensulfite.

Carbonate and hydrogencarbonate are determined by an acidimetric titration to endpoints of pH 8.35 for carbonate and pH 4.35 for hydrogencarbonate. The indicator electrode is usually a glass electrode that allows the pH to be followed potentiometrically. In one

mode, the titrant ( $0.01\text{ mol l}^{-1}$ ) is added at a constant rate and the pH is monitored continuously. Processing of the recorded data yields the endpoints. Automation of the determination is possible but expensive.

In modern laboratories fluoride is almost exclusively determined with a fluoride-selective electrode after addition of a total ionic strength adjustment buffer with citrate and cyclohexanediaminetetraacetic acid as decomplexing agents. The electrode signal suffers little interference from other ions. When fluoride has to be determined in concentrations as low as  $1\text{ }\mu\text{g l}^{-1}$ , application of the electrode in a CF apparatus is advantageous. Decomposition of metal–fluoride complexes is difficult and high recoveries are required (above 80%). Quality control is necessary. It may even be necessary to carry out the decomposition of the complexes outside the analyzer in order to increase the time available for the reaction. The method suffers interference from extreme pH and multicharged cations such as  $\text{Fe}^{3+}$  and  $\text{Al}^{3+}$ .

Ammonium is determined in many laboratories in a CF system in which the Berthelot reaction is implemented. In the Berthelot reaction, ammonium reacts with chlorine and phenol in the presence of sodium nitroprusside as catalyst in alkaline medium. EDTA is added to prevent interference of calcium and magnesium. Modern systems have been developed that use macroporous polytetrafluoroethylene (PTFE) membranes. In these systems a sample is introduced into a stream to which sodium hydroxide solution is added. Ammonia diffuses through the PTFE membrane into a stream of de-ionized water and the stream is fed through the flow-through cell of a conductivity meter. In this system a minimum of reagents is required and the only interference is from volatile amines.

The determination of hydrogen ions is important for rainwater. Hydrogen ion concentration is often calculated from the pH. However, this may increase the uncertainty (pH-to-hydrogen ion conversion is a logarithmic function) and decrease the accuracy because of the inaccuracy in the value of the activity coefficient (which is affected by temperature and ionic strength) that is required for the conversion from a measured activity to a calculated concentration. pH measurement therefore requires careful attention to the measuring conditions (temperature, ionic strength, avoidance of contact and diffusion potentials, etc.). Under such circumstances procedures such as Gran's plot or titrimetric methods are preferred to direct potentiometry.

Sodium and potassium are mainly determined by flame atomic emission spectrometry (FAES), flame atomic absorption spectrometry (FAAS), and inductively coupled plasma atomic emission spectrometry



(ICP-AES). FAES is widely used for the determination of potassium and sodium, but interference is experienced from stray light. In FAES and FAAS, ethyne (acetylene) and air are generally used as the flame gases. After acidifying the samples with nitric acid, an electron buffer (cesium chloride or lanthanum nitrate) is added to suppress interference by phosphate. When low concentrations (below  $50 \mu\text{l}^{-1}$ ) of potassium (in rainwater or boiler water) have to be determined, atomic absorption spectrometric methods are usually applied. For determinations at such low concentrations, sample containers and the nebulizer system must be scrupulously cleaned by soaking for  $\sim 24$  h in dilute nitric acid. FAAS is therefore preferred for low-concentration measurements. For sodium and potassium measurement in particular, simplified and cheaper AAS instruments can be used that are provided with a light source that is monochromatized with a filter instead of the more sophisticated hollow cathode lamp (HCL) or electrodeless discharge lamp (EDL). In matrix-free water samples such as boiler water, sodium can also be measured down to  $1 \mu\text{g l}^{-1}$  with an ion-selective electrode. The samples are not pretreated in most applications of ICP-AES. Samples with a high salt content can be reproducibly introduced with an ultrasonic nebulizer or must be diluted appropriately.

Calcium and magnesium are often determined by FAAS. Nitrous oxide and ethyne (acetylene) are the best flame gases. Prior to measurement the sample is acidified with nitric acid. Lanthanum nitrate is added to suppress interferences by oxidizable acids, aluminum and organic complexes of aluminum, phosphate, silicate, iron, etc. Strontium chloride can be used instead of lanthanum nitrate. To prevent ionization in the flame, cesium chloride, potassium chloride, or sodium chloride can be added. No pretreatment is used when the determinations are carried out with ICP-AES; because of the high temperature of the flame, all calcium phosphate compounds are decomposed. If a determination is needed only occasionally, calcium and magnesium can be determined titrimetrically with ethyleneglycol-bis (2-aminoethyl)- $N,N,N',N'$ -tetraacetic acid (EGTA) and Eriochrome Black T as indicator in alkaline medium. Magnesium is titrated first. Next, magnesium-EGTA complex is added and calcium displaces magnesium from this complex. The liberated magnesium is titrated as a measure of the calcium in the sample. Cyanide is added to overcome any effect of traces of iron and heavy metals.

Iron and manganese can be determined with FAAS; air and ethyne are used as flame gases. Prior to the determination of total iron and manganese, digestion

with aqua regia in a microwave oven for 1 h is required. This treatment takes care of refractory compounds such as iron oxides. Iron and manganese can also be determined with ICP-AES, in which case no predigestion is required owing to the high temperature of the flame.

Silicate is determined spectrophotometrically with ammonium molybdate and ammonium vanadate. The pH of the sample must be adjusted to 7–8. Potassium cyanide is added to prevent interference of heavy metals. Oxalic acid is added to destroy molybdophosphate and vanadophosphate and to bind aluminum in a complex. As in all spectrophotometric determinations, high and variable optical absorption of the sample (due to color or turbidity) at the wavelengths of investigation causes errors; tannin, iron, and sulfide also interfere. To avoid contamination, all contact surfaces should be of polyethylene. It is also possible to determine silicate using FAAS, in which case nitrous acid and ethyne must be used as flame gases. As silicates are present in colloidal form, the sample must be introduced into the AAS equipment using an ultrasonic nebulizer. Such a nebulizer is also used when silicate is measured by ICP-AES.

*Analysis of bromate in raw and drinking water* In the recent past the analysis of bromate in water has received considerable attention. Not only ozone, but also other strong oxidizing components are applied for the preparation of safe drinking water. By the application of these substances bromate ion may be formed during ozonization of bromide-containing compounds in the waters. Bromate is determined in raw and drinking water with IC. Specific conductimetric detection is combined with the determination of bromate by IC with a concentration of bromide. The most important step is the unselective enrichment of all water constituents by means of a rotator evaporator and a selective removal of bromate in raw and drinking waters is possible reduce to concentrations of at least  $1 \mu\text{g l}^{-1}$ .

In the analysis of raw and drinking water components are of interest that are formed during the chlorination with for instance hypochlorite or chlorine dioxide and are physiologically active. Such components are amongst other bromate- and iodine-containing substances. After concentration of the sample bromate is determined with IC. The limit of the bromate determination is  $1 \mu\text{g l}^{-1}$ .

*Analysis of iodine in drinking water* The determination of iodine concentrations in drinking water is important to evaluate the occurrence iodine deficiency in humans. Iodine deficiency disorders are one of the world's most significant causes of mental retardation

and preventable brain damage. The lowest iodine level for drinking water has been reported to be  $5 \mu\text{g l}^{-1}$  by the World Health Organization (WHO).

For the determination of substances containing iodine, a portion of 10 ml of drinking water is transferred to a 25 ml calibrated flask. About 1.0 ml of  $0.5 \text{ mol l}^{-1} \text{ H}_2\text{SO}_4$ , two drops of bromine water, two drops of 10% (v/v) phenol, 0.5 ml of 0.5% (w/v) KI, and 1.0 ml of 1% (v/v) starch solution is added. The solution filled up to the mark with distilled water and mixed. After 10 min, the absorbance of the solution is measured spectrophotometrically at 590.0 nm against a reagent blank.

**Microcomponents** Inorganic microcomponents cover almost the entire periodic table. Here we discuss only those elements that are determined most frequently: beryllium, vanadium, chromium, cobalt, nickel, copper, zinc, arsenic, selenium, silver, cadmium, antimony, barium, and lead. The techniques of choice are FAAS, graphite furnace-AAS (GF-AAS), ICP, and hydride generation-AAS (HG-AAS). An ultrasonic nebulizer has recently become commercially available for FAAS and ICP-AES, which decreases the lower determination limits. ICP-mass spectrometry (ICP-MS) is a recent development in which ionization is combined with sensitive mass discrimination. In a further development a graphite furnace is used in front of the ICP-MS. Selective evaporation of elements in the graphite furnace reduces the influence of highly interfering matrices. ICP-MS is expensive, which deters its widespread use.

Arsenic, selenium, and antimony can be measured at low concentrations with HG-AAS. Digestion with aqua regia in a microwave oven is necessary prior to the determination. Next, hydroxylammonium chloride is added as a neutral solution; iodide may be added to accelerate the reaction. For arsenic or antimony a reducing agent such as ascorbic acid has to be added to convert all arsenic or antimony to a single oxidation state. A background correction is required; for this purpose apparatus with a deuterium lamp corrector can be used but Zeeman background correction is preferred.

Determination is also possible with GF-AAS. However, this method has a higher limit of detection and suffers interference from many substances.

Currently, the remaining elements listed above are usually determined with GF-AAS equipped with deuterium or preferably Zeeman correction. Matrix modifiers such as  $\text{Mg}(\text{NO}_3)_2$ ,  $\text{Pd}(\text{NO}_3)_2$ ,  $\text{La}(\text{NO}_3)_3$ ,  $(\text{NH}_4)_2\text{HPO}_4$ , or  $(\text{NH}_4)_2\text{SO}_4$  are required. A limit of detection of  $1 \mu\text{g l}^{-1}$  can be realized. Over the largest part of the working range a precision of 5–10% is attainable and there is a range of about two decades between the uppermost and lowest limits of determination. To obtain unambiguous absorption in AAS, the elements have to be in the same oxidation state and therefore a reducing agent is added. Triton X-100 is added to prevent adsorption effects in the AAS itself. Some characteristics of AS methods are summarized in Tables 1–3. In Table 2, the limits of determination for the inorganic microcomponents mentioned above are given for real samples. These limits are up to 2–10 times higher than the limits that are generally specified, which are determined from standard solutions without interferences.

**Table 2** Limits of determination in real samples

Element	Limit of determination ( $\mu\text{g l}^{-1}$ )		
	FAAS	GF-AAS	HG-AAS
Ag		1	
As		1	0.2
Ba		5	
Be		0.05	
Cd		0.1	
Co		1	
Cr		0.5	
Cu		1	
Ni		1	
Pb		0.5	
Sb		1	0.5
Se		1	0.2
V		1	
Zn	5	1	

Limit of determination is the minimum level that can be reliably achieved by a method within specified limits of precision and accuracy.

**Table 1** Characteristics of atomic spectrometric methods

Method	Repeatability (%)	Limit of detection (g)	Sample volume (ml)	Time (min)	Simultaneous analysis
FAAS	2–5	$10^{-6}$ – $10^{-7}$	10–100	6	No
GF-AAS	8–12	$10^{-11}$ – $10^{-12}$	0.1–2	6	No
HG-AAS	5–10	$10^{-8}$ – $10^{-9}$	100	6	No
ICP-AES	5–20	$10^{-5}$ – $10^{-6}$	10–100	6	Yes

Limit of detection is the lowest concentration of an analyte that can be detected at a specific confidence level.



**Table 3** Matrix modifiers for atomic absorption spectrometric determination of some elements, with some figures of merit

Element	Interferents	Matrix modifier	Limit of determination ( $\mu\text{g l}^{-1}$ )	Repeatability (%)
Ag	Various	$(\text{NH}_4)_2\text{HPO}_3$	1	5–10
Ba	Various	$\text{La}(\text{NO}_3)_3$ , $\text{Mg}(\text{NO}_3)_2$	5	5–10
Be	Al, Ca	$\text{La}(\text{NO}_3)_3$	0.05	5–10
Cd	Cl, Sb	$\text{Pd}(\text{NO}_3)_2$ , $\text{Mg}(\text{NO}_3)_2$ , Triton X-100, reductor	0.1	5–10
	Zn, organics	$(\text{NH}_4)_2\text{SO}_4$	0.1	5–10
Co	Volatile with Cl, P, S	$\text{Mg}(\text{NO}_3)_2$	1	5–10
Cr	Desorption	$\text{Pd}(\text{NO}_3)_2$ , reductor, Triton X-100	0.5	5–10
	$\text{Cr}^{3+}/\text{Cr}(\text{VI})$	$\text{Mg}(\text{NO}_3)_2$	0.5	5–15
Cu	Cl, $\text{Cu}_2\text{O}$ , Fe	$\text{Pd}(\text{NO}_3)_2$	1	5–10
Ni	Cr	$\text{Mg}(\text{NO}_3)_2$	1	5–10
		$\text{Pd}(\text{NO}_3)_2$ , reductor, Triton X-100	1	5–10
Pb	Cl	$(\text{NH}_4)_2\text{HPO}_4$	1	5–10
Zn		—	5	5–10
V	Formation of various carbides	$\text{Mg}(\text{NO}_3)_2$	1	5–10

*Monitoring trace elements in waters by ICP-AES*

How to avoid spectroscopic interferences is an important issue during the development of any analytical method. Such interference is also important in the application of ICP-EAS. Elements commonly present in environmental samples give rise to difficulties in the determination of other elements at ultratrace concentrations. For simplicity a list of alternative emission lines based on relative differences between intensities of the measured elements and their interferences are omitted in **Table 4**.

*Monitoring trace elements in waters by ICP-MS*

A simple, rapid, and accurate panoramic method based on ICP-MS has been developed for the determination of ultratrace concentrations of metals in water samples. On the basis of an appropriate selection of the isotopes monitored and combination of internal standardization with correction equations to overcome the matrix effects, the method permits accurate determination of up to 60 elements in a single run of less than 3 min.

Assignments of intensities and calculation of concentrations can be performed by general element survey and specified software. Precision (RSD) below 5% is commonly achieved for elements present at  $25 \mu\text{g l}^{-1}$ . To correct for matrix-induced ion signal variation and instrumental drift, rhodium or indium in combination with panoramic analysis, based on full mass-spectra scan methods, is used as the internal standard (IS). Spectroscopic effects due to Cl, Na, Ca, Mg, S, and C were corrected with interference factors (IF) on the basis of a set of correction equations (see **Table 5**).

Hence, it is less suited for a fast panoramic analysis. In this list the interference factors associated to

the polyatomic species containing one isotope of the previously mentioned interference elements (Cl, Na, Ca, Mg, S, P, and C) in combination with argon, nitrogen, oxygen, or hydrogen demonstrate the daily determination of the interference factors and provide the possibility of increasing the sample throughput considerably.

*Mercury in water by ICP-MS* ICP-MS with addition of gold allows the previously difficult measurement of low concentration of mercury in water.

*Multicollector-ICP-MS* Using MC-ICP-MS offers the possibility of isotope ratios. The development of multicollector ICP systems allows studies of previously immeasurable nonstable transition metals ( $^{13}\text{C}/^{12}\text{C}$  and  $^{16}\text{O}/^{18}\text{O}$ ) isotope ratios.

*Early warning detection and monitoring* A sensor is a device that gives information about the presence or absence of a component of interest. The advantage of sensors is that the sensor measures continuously with the same device. Therefore, the measurements are reliable. This is essential for determinations in the domain of medicinal, chemical, and production industries. Sensors are cost-effective and innovative for analysis of water at point locations.

A new concept sensor is used for the analysis of range 7 compounds in water. The module provides outputs for pH, dissolved oxygen, chlorine, temperature, conductivity, and redox potential, suiting a wide range of applications from laboratory measurements to environmental monitoring. Next, a hybrid sensor is being developed that combines the sensitivity of piezoelectric nanobalance technology with the selectivity of stripping potentiometric analysis for the

**Table 4** Summary of analyte emission lines for ICP-AES measurements used to eliminate interferences

<i>Wavelength (nm)</i>	<i>Rel. intensity</i>	<i>Signal/noise</i>	<i>BEC</i>	<i>Interferences</i>
<i>Arsenic (As)</i>				
193.696 p	5800	100	1.8	Fe, Al
197.197 s1	4300	63.1	2.6	Pb, Pd, Fe, Al
188.979 s2	2100	125.9	4.6	Pd
<i>Cadmium (Cd)</i>				
214.438 s1	170 000	1000	0.08	Fe, Al, Si
228.802 p	150 000	2000	0.09	Co, As
226.502 s2		1600	0.11	Ba, Fe, Ti
361.051		6100	50	Fe
<i>Chromium (Cr)</i>				
205.552 s1		500	0.2	Fe, Ni, Mo
206.149 p		400	0.24	Bi, La, Zn
267.716 s2		1000	0.24	Fe, Mn, V
357.869		1000	0.77	Fe, Tb
<i>Copper (Cu)</i>				
324.754 p		5000	0.10	Fe, Co, Ti, Cr
224.700 s2		1000	0.26	Fe, Pb, Ni
327.396 s1		3200	0.32	Ta, Na, Fe
221.458		158	0.77	Fe
<i>Manganese (Mn)</i>				
257.610 p	580 000	10 000	0.05	Fe, Co, Mo, Cr
260.569 s1		6300	0.07	Fe, Co
294.920 s2	150 000	1600	0.26	Fe, Zr, Ca
279.482	72 000	2000	0.42	Mg, Fe
403.060	44 000	400	1.47	Fe, Al, La
<i>Nickel (Ni)</i>				
221.647 s1		630	0.34	Si
232.003 s2	35 000	250	0.5	Fe, Cr, Ni
231.604 p	27 000	630	0.53	Co, Mo, Sb
341.476	30 000	316	1.61	Co, Ca
<i>Lead (Pb)</i>				
220.353 p		126	1.43	Fe, Al, Co, Ti
216.999 s1	4900	100	3.03	Fe, Cu
261.418		63	4.35	Fe, Co
283.306		50	4.76	Fe, K
<i>Zinc (Zn)</i>				
213.856 s2	240 000	3200	0.06	Cu, Ni, Fe, Ca
202.548 s1	76 000	1000	0.13	Co, Ni, Cr, Co, Mg
206.200 p		1000	0.2	Cr, Fe
334.502	9600	40	4.55	Mo, Zr, Si

Boumans and Wholers show, in an application note, data that indicate relative intensities and sensitivities which can effectively eliminate interferences at specific wavelengths.

Besides these data, background equivalent concentrations (BEC) are presented for the adjustment of the dynamic bandpass tuning parameters.

P.W.J.M. Boumans (relative intensity) and M. Wohlers (signal/noise): Based on background multiplicative correction models where one element changes the emission of another one.

detection of heavy metals, e.g., chromium(VI), chromium(III), and mercury(II) in aqueous samples. Improving the electrochemical circuitry and using suitable electrode modification are extending the

technique for the detection and speciation of additional toxic heavy metals.

Moreover, prototype devices of chemiresistor micro-sensors demonstrated the performance for detecting

**Table 5** ICP-MS (matrix effects: spectroscopic interference corrections) (correction equations based on the experimentally determined interference factors (average,  $n = 5$ ))

<i>Polyatomic ion</i>	<i>Isotope affected</i>	<i>Correction equation</i>
$^{12}\text{C}^{14}\text{N}^1\text{H}$	$^{27}\text{Al}$	$^{27}\text{Al} = ^{27}\text{I} - 0.0040(0.003)^{12}\text{I}$
$^{12}\text{C}^{16}\text{O}^1\text{H}$	$^{29}\text{Si}$	$^{29}\text{Si} = ^{29}\text{I} - 0.013(0.003)^{12}\text{I}$
$\text{C}^{16}\text{O}^{16}\text{O}$	$^{45}\text{Sc}$	$^{45}\text{Sc} = ^{45}\text{I} - 0.030(0.004)^{13}\text{I}$
$^{31}\text{P}^{16}\text{O}$	$^{47}\text{Ti}$	$^{47}\text{Ti} = ^{47}\text{I} - 0.0064(0.0007)^{51}\text{I} - 0.00015(0.0001)^{53}\text{I}$
$^{53}\text{S}^{14}\text{N}$		
$^{33}\text{S}^{16}\text{O}$ $^{34}\text{S}^{16}\text{O}^1\text{H}$ $^{35}\text{Cl}^{16}\text{O}$ $^{36}\text{S}^{15}\text{N}$ $^{37}\text{Cl}^{14}\text{N}$	$^{51}\text{V}$	$^{51}\text{V} = ^{51}\text{I} - 0.00027(0.0002)^{55}\text{I} - 0.0020(0.0008)^{55}\text{I} - 0.00010(0.00086)^{56}\text{I} - 0.0052(0.002)^{57}\text{I}$
$^{12}\text{C}^{40}\text{Ar}$ $^{34}\text{S}^{16}\text{O}$ $^{35}\text{Cl}^{16}\text{O}^1\text{H}$ $^{36}\text{S}^{16}\text{O}$	$^{52}\text{Cr}$	$^{52}\text{Cr} = ^{52}\text{I} - 0.0032(0.001)^{54}\text{I} - 0.00018(0.0001)^{55}\text{I} - 0.0036(0.0022)^{56}\text{I}$
$^{35}\text{Cl}^{16}\text{O}^1\text{H}$ $^{37}\text{Cl}^{16}\text{O}^1\text{H}$	$^{54}\text{Fe}$	$^{54}\text{Fe} = ^{54}\text{I} - 0.00024(0.001)^{35}\text{I} - 0.00064(0.0002)^{37}\text{I}$
$^{23}\text{Na}^{16}\text{O}^{16}\text{O}$ $^{35}\text{Cl}^{18}\text{O}^1\text{H}^1\text{H}$ $^{37}\text{Cl}^{18}\text{O}$	$^{55}\text{Mn}$	$^{55}\text{Mn} = ^{55}\text{I} - 0.00037(0.000004)^{23}\text{I} - 0.00015(0.0002)^{55}\text{I} - 0.00039(0.00005)^{37}\text{I}$
$^{23}\text{Na}^{36}\text{Ar}$ $^{43}\text{Ca}^{16}\text{O}$	$^{59}\text{Co}$	$^{59}\text{Co} = ^{59}\text{I} - 0.0000024(0.0000001)^{35}\text{I} - 0.00086(0.0001)^{43}\text{I}$
$^{12}\text{C}^{16}\text{O}^{16}\text{O}^{16}\text{O}$ $^{24}\text{Mg}^{36}\text{Ar}$ $^{43}\text{Ca}^{16}\text{O}^1\text{H}$ $^{44}\text{Ca}^{18}\text{O}$	$^{60}\text{Ni}$	$^{60}\text{Ni} = ^{60}\text{I} - 0.000018(0.00001)^{12}\text{I} - 0.0000018(0.0002)^{24}\text{I} - 0.0029(0.0003)^{43}\text{I} - 0.00018(0.00003)^{44}\text{I}$
$^{25}\text{Mg}^{38}\text{Ar}$ $^{31}\text{P}^{16}\text{O}^{16}\text{O}$ $^{44}\text{Ca}^{17}\text{O}$	$^{63}\text{Cu}$	$^{63}\text{Cu} = ^{63}\text{I} - 0.000093(0.0001)^{25}\text{I} - 0.00044(0.0006)^{31}\text{I} - 0.000069(0.000004)^{44}\text{I}$
$^{26}\text{Mg}^{40}\text{Ar}$ $^{32}\text{S}^{16}\text{O}^{18}\text{O}$ $^{33}\text{S}^{16}\text{O}^{17}\text{O}$ $^{34}\text{S}^{16}\text{O}$	$^{66}\text{Zn}$	$^{66}\text{Zn} = ^{66}\text{I} - 0.018(0.000002)^{26}\text{I} - 0.000023(0.000004)^{32}\text{I} - 0.0020(0.0007)^{33}\text{I} - 0.00042(0.0001)^{34}\text{I}$
$^{35}\text{Cl}^{40}\text{Ar}$ $^{43}\text{Ca}^{16}\text{O}^{16}\text{O}$	$^{75}\text{As}$	$^{75}\text{As} = ^{75}\text{I} - 0.00020(0.0001)^{35}\text{I} - 0.000017(0.000021)^{43}\text{I}$
$^{34}\text{S}^{16}\text{O}^{16}\text{O}^{16}\text{O}$	$^{82}\text{Se}$	$^{82}\text{Se} = ^{82}\text{I} - 0.0000020(0.000001)^{34}\text{I}$

I = intensity.

Reprinted with permission from Jitaru P, Tirez K, and De Brucker N (2003) *Panoramic Analysis for Monitoring Trace Metals in Natural Waters by ICP-MS*. VITO, B-2400 Mol, Belgium.

volatile organic compounds (e.g., toluene) in the parts per million concentration range. The prototype devices had fast response times, low detection limits, highly reversible behavior, and excellent stability.

### Organic Chemical Analysis

There are many more organic compounds present in water than there are inorganics. Their chemical and physical properties are often very closely related,

which makes it very difficult to discriminate a single compound from others that are also present in the sample. Generally, a method of analysis includes the following steps:

1. Isolation from the water, which usually also includes a preconcentration step.
2. Separation from interfering compounds that are isolated from the water together with the determinant (cleanup).
3. Further concentration in order to enable detection of the compound.
4. Instrumental chromatographic separation and detection.

These steps in the analysis are described below.

The selection of an isolation and/or concentration technique and the instrument for the measurement depends largely on the class of compounds that are to be analyzed. Sometimes it is necessary to analyze simultaneously a wide variety of chemical compounds in the sample. Application of a series of techniques can provide the solution.

For organic residue analysis of samples, several isolation, concentration, cleanup, and detection techniques are available, e.g., liquid-liquid extraction (LLE), use of a sorbent, and headspace analysis (HSA). Purge-and-trap and/or closed-loop stripping analyses (CLSAs) may be used for highly volatile compounds (e.g., chloroform). Often a concentration step is necessary in the procedure in order to reach the concentration range that can be detected by an instrumental technique. It may be necessary to change the chemical nature of the compound to be analyzed by derivatization in order to make it possible to separate the compound from the matrix or to discriminate it from other interfering compounds or to improve the detection.

Instrumental techniques that are most commonly used are gas chromatography (GC) and liquid chromatography (LC), both in combination with a wide range of detectors. With instrumental chromatographic techniques it is possible to make further discrimination between organic compounds before detection. The choice of the column in GC and LC is very important and the detector can be very specific. It is important to set up an optimal combination of steps in the analytical procedure.

The procedure in an analytical scheme for the analysis of a wide range of organic compounds may be a combination of a gas-stripping technique for the volatiles and a number of adsorption or extraction steps. Alternatively, the sample can be passed through an adsorbent resin. The organics that pass the resin can be isolated from the effluent by means

of freeze-drying concentration. An isolation-fractionation scheme can be set up in which organic compounds with different functional groups and sorption parameters are separated and concentrated in fractions with similar behavior, which permits instrumental analysis. However, in practice there are several organic classes that still cannot be efficiently recovered.

**Direct aqueous injection** Direct injection of a sample without any sample treatment into an instrument such as a gas chromatograph or high-performance liquid chromatograph is the simplest and fastest analytical procedure for the analysis of organic pollutants in surface water. However, most columns are not water resistant: support surfaces or stationary phases hydrolyze. Resistant column types are polyglycols, e.g., carbowaxes on a barium carbonate surface. Water is also very strongly retained and inert columns with immobilized silicone coatings (apolar) that have almost no retention for water must be used. However, columns can be deactivated in such a way that direct water injection is possible. The influence of water on a GC detector can be such that the flame ionization detector (FID) flame will be extinguished immediately after the injection or the electron-capture detector (ECD) ground current will effectively be quenched.

On-column injection produces no problems with strongly water-soluble (fully trapped) solutes, few or negligible problems with water-insoluble solutes (non-trapped), and some problems with solutes of intermediate solubility (partially trapped). Splitless GC injection can only be used for strongly soluble solutes.

A very thick film (1–8  $\mu\text{m}$ ) with a high loading capacity can be useful for direct injection. For direct determination of organic compounds in water, a gas chromatographic/concentration apparatus can be used in which organic compounds are retained on a collection precolumn. After venting the water (10–200  $\mu\text{l}$ ) the precolumn is heated rapidly (a few seconds). The method is especially useful for compounds that are difficult to extract or to purge from water. The apparatus can be used in combination with GC-FID, ECD, nitrogen/phosphorus-alkali FID, and photoionization detectors.

Trace substances can be adsorbed from the sample onto a bonded hydrophobic-phase capillary column when it passes through the capillary column at room temperature. These substances are subsequently thermally desorbed, trapped, and analyzed by GC using either the same column or one in series. Utilizing a single column with FID, concentrations can be determined at the microgram per liter level,

while using two columns with ECD allows concentrations to be determined at the nanogram per liter level. A Tenax-GC column in combination with FID for the analysis of organic compounds (e.g., alcohols, acetates, ketones, ethers) gives good results.

In another simple method of direct aqueous injection GC, water is selectively removed by diffusion across a permselective membrane prior to the sample entering the GC column. Injection volumes are 1–20  $\mu\text{l}$ .

Steam can be used as the carrier gas in GC. The sensitivity is  $10\mu\text{g l}^{-1}$  for hydrocarbons. The FID gives a constant and reproducible response, which, however, is lower than under conventional GC conditions.

It is almost always important to avoid direct contact of the water with the GC column. The direct aqueous injection technique is a simple and useful method for determination at relatively high concentrations.

**Xeno-estrogens** It is supposed that the degradation products of penicillin, detergents (polyoxyethylenes), and softening agents have hormonal residual activity in surface water that is adverse for water organisms. At present such emerging substances are measured with liquid chromatography (LC) coupled with MS-MS (LC-MS-MS). The advantages of the LC-MS-MS method is the extended linear range and superior sensitivity, providing a limit of quantification of 0.1 to  $10\text{ ng ml}^{-1}$  for previously difficult monitoring of biofluids and metal-containing compounds in biological matrices. Large-volume injection-LC-ESI-MS-MS in positive or negative ion mode enables the detection of, respectively, negative ion compounds, such as nitrofurantoin, clofibrine acid, bezafibrate, and ibuprofen, and positive compounds such as amoxicillin, ciprofloxacin, cephalixin, metoprolol, paracetamol, sulfamethoxazole, erythromycin, ifosfamide, carbamazepin, and fenofibrate.

### Isolation and preconcentration

**Liquid-liquid extraction** LLE is a simple conventional technique for isolation and concentration. Which classes of compounds can be extracted from samples depends on a number of factors, such as type of sample (presence of particulate matter, ionic strength of the water), pH, type of extraction solvent (polarity of the solvent is the main factor), solvent-to-water ratio, and extraction procedure. Extraction at approximately pH 7 can be inefficient for some classes of compounds, e.g., phenols, which means that in such cases the sample pH has to be adjusted before extraction or consecutive extractions at

different pH values may be advisable. A low solvent-to-water ratio can lead to the formation of an emulsion, and extraction may take a long time because of the slow kinetics involved in reaching equilibrium, especially for high-molecular-mass organic compounds. For extraction, a suitable immiscible solvent has to be used. If there is a high solvent-to-water ratio, a concentration step is usually necessary in order to improve sensitivity. Considerable loss of the more volatile compounds may occur if the extract is concentrated by evaporation. Concentration of the extraction solvent subsequently requires the availability of very pure solvents.

Some LLE techniques that have been in use for sometime are described below.

**Batch LLE (BLLE)** This technique involves shaking or stirring the sample with a small volume of solvent in an extraction vessel. Extraction of a sample with a very low volume of solvent is possible in a vessel with a narrow neck.

**Flow under continuous LLE** In this method, organics are extracted by bringing the sample and solvent into contact without breaking the interface between the two phases. Stirring of the aqueous layer prevents the process from being diffusion-limited. Distillation enables recycling of fresh solvent to extract the sample.

**Continuous LLE** An automated continuous LLE apparatus, based on the mixer-settler principle, can be constructed for use in the field. The apparatus consists of two extractors in series for serial extraction and can be applied at different depths. The units are loaded with  $\sim 200\text{ ml}$  of solvent and have the capacity to extract some  $100\text{ l}$  of water, depending on the solvent used.

Another apparatus for continuous extraction consists of three inverted  $250\text{ ml}$  volumetric flasks. A vibratory mixer is used for vigorous shaking of all three flasks. A  $10\text{ l}$  sample can be siphoned through the apparatus in  $\sim 4\text{ h}$  and extracted, with a total solvent volume of  $10\text{ ml}$ . The advantage is that no concentration step is involved.

Simple batch extraction is most widely used. The extraction efficiencies for semivolatile organic pollutants decrease in the order: pentane  $\geq$  hexane  $>$  hexane saturated with methanol  $>$  isooctane  $>$  15% (v/v) acetone in hexane  $\geq$  benzene. Pentane and hexane show the least interference with standard peaks during GC. Hexane is a more attractive extraction solvent than pentane because the volatility of pentane makes it difficult to handle, particularly when it is used with an automatic injector. The single-step

LLE procedure has some limitations when applied to samples containing large numbers of compounds. Analytical results for such samples are likely to be semiquantitative and will require careful evaluation. The hexane extraction procedure is, however, very useful as a simple, rapid screening method that can be applied to complement other methods. A low solvent-to-water ratio will give problems in the analysis of more foaming samples such as discharge water and the amount of solvent has to be increased. The use of large volumes of organic solvents and the high concentration factor of 100–10 000 subsequently requires solvents of very high purity. Concentration by evaporation gives a loss of volatile compounds and is laborious. A relatively new development is the technique using supercritical fluid carbon dioxide.

*Concentration of the extracts* Common solvent evaporation techniques that are acceptable for use in pollutant analysis include the following:

1. Kuderna–Danish evaporative concentrator equipment with a Snyder or Vigreux column for macro- and microanalysis, depending on the scale of the construction.
2. Rotary evaporation, equipped with a bump trap, under vacuum at ambient temperature. Application is restricted to macroanalysis.
3. Evaporation on a hot plate in an Erlenmeyer flask with or without a modified Snyder column, in which the solution is heated at 30°C. Macro and microanalysis.
4. Heated nitrogen blowdown in an Erlenmeyer flask, in which the solution is heated at 30°C with a gentle stream of nitrogen.

The recoveries improve markedly with the experience of the analyst. Nitrogen blowdown with column is significantly improved using dichloromethane as the solvent.

*Sorbent extraction* Sorbent extraction makes use of the differential distribution of the dissolved compounds between the solid sorbent and water. The use of organic porous polymers is advantageous for the extraction of organic compounds from water because: (1) the partition coefficients of compounds in a polymer–water system reach infinity if the correct polymer is selected for the types of contaminants present; (2) adsorption of water itself on the polymer is minimal; (3) the wettability of the polymer with water makes possible satisfactory transport of the substance toward the polymer surface; and (4) the polymer is inert.

Usually the sample is pumped through a column packed with a sorbent. The adsorbed compounds are then desorbed from the sorbent and analyzed chromatographically. The presence of mineral oil in the water at milligrams per liter levels considerably decreases the adsorption of compounds on the sorbent. This problem can be solved by using an extra layer of inert material at the entrance to the column. When water-containing organic compounds are passed through the column, a chromatographic process occurs and, after a certain time, breakthrough occurs. Under ideal conditions the trapping of organic compounds is 100%. This depends on a number of factors: sorbent, structure of the determinands, composition of the sample, and original concentration of the determinands in the water. Adsorption theory is very complicated, especially when water matrix effects are involved. Water characteristics may vary widely, and recovery studies are difficult to carry out. The water content of particulate matter and the way the organics are stuck to the particles are important. The nature of the active sites determines the ability of the organics to be extracted. When the water is first filtered, colloidal and small particles will pass through the filter and bring part of the organics through the extraction column more or less unaffected. When the water is not filtered, the adsorbent columns are easily clogged, resulting in channeling. Even pores may sometimes be blocked. The necessity to obtain resins of high purity and optimal adsorption capacity is a problem. An extended cleanup procedure for trace analysis is mandatory; this procedure can be automated.

Desorption of the compounds from the sorbent can be performed either with a liquid or by heating. Solvents used include diethyl ether, n-hexane, acetone, isopropanol, methanol, methylisobutyl ketone, and pyridine. The solvent strength is also important; for example, in a stream containing activated carbon the solvent strength increases as follows: water < methanol < ethanol < acetone < propanol < diethylether < butanol < ethyl acetate < n-hexane < benzene. Sometimes the desorbed compounds are led directly into the chromatographic column for LC. In thermal desorption, the column is situated before the analytical column of the GC, similar to HSA. In order to eliminate peak broadening, the initial sections of the chromatographic columns are usually cooled. The temperature must be such that all compounds contained in the sorbent are transferred from it to the GC in a reasonable amount of time. If the temperature is too high, the substances can decompose.



Classes of compounds that can be extracted with the various absorbents include:

XAD-2: PCBs, phthalates, carbamates, alkylbenzenes;  
 XAD-4: Chloroethers, acids, phenols, humic acids;  
 XAD-7: Amino acids, PAHs;  
 Active carbon hydrocarbons: phenols, PAHs;  
 Chromosorb 102: Chlorobenzenes;  
 Tenax-GC: PAHs, chlorobenzenes; and  
 Sep-pack: PAHs, phenols, PCBs, aliphatic carboxylic acids.

A wide range of organic compounds can be extracted using different resins; however, good extraction of highly volatile compounds with a resin is not possible.

**Liquid chromatography** Another possibility is the use of online trace enrichment in LC using a precolumn. The method consists of an online preconcentration step and a postcolumn reaction step. Online preconcentration of mono- and dichlorophenols from aqueous samples has been performed using PRP, a diphenylbenzene–styrene copolymeric sorbent, as packing material for both analytical column and precolumn. Enrichment factors of over 300 can be obtained. The method is suitable for the analysis of effluent samples. A precolumn, when used for preconcentration, serves for sample cleanup and as a guard column at the same time. A guard column protects the valuable analytical LC column against contamination by sample, mobile-phase constituents, and grit released by the chromatographic equipment, including the injector. The requirements that have to be met by a precolumn are summarized as follows:

1. The amount of material should be sufficient for the preconcentration of the desired amount of determinand from a given volume of sample.
2. The precolumn has to be as small as necessary to prevent peak broadening and must have a high-pressure capability. It should not present back-pressure limitations during sampling at high flow rates; it should be easily replaceable and easy to repack, and should be compatible with all eluents generally used in LC.
3. Field sampling and automated sampling handling must be feasible.

Preconcentrations have been performed on both C18 and carbonaceous sorbents.

**Headspace analysis** By analyzing the gas phase that is in equilibrium with the water phase, instead of the water phase, it is possible to eliminate many drawbacks present in the usual methods of

preconcentration (e.g., LLE). Determination of trace organic volatiles in the gas phase above the sample is known as HSA. The techniques of HSA can be divided into two groups, static and dynamic.

**Static headspace analysis** The sample for chromatographic analysis is taken from a closed vessel where the material under study reaches equilibrium with its vapor at a predetermined temperature. Experimental conditions that can influence the results are the temperature and the sample-withdrawal technique. The concentration of the analytes in the gaseous phase can be increased by raising the temperature, adding an electrolyte, and varying the pH. Solutes with low vapor pressures cannot be detected.

**Dynamic headspace analysis** A stream of inert gas (helium or nitrogen) strips the volatile compounds, e.g., organohalogen hydrocarbons, from the water and passes them through a sorbent or cryogenic trap or through a combination of the two. The volatiles are eluted from the sorbent with a solvent or are thermally desorbed and brought into the gas chromatograph or other instrument. Often these techniques are online. The technique can be applied in an open system, where the stripping gas is vented to the atmosphere, or in a closed circuit, where the gas phase is recycled through the water and trap, closed loop sample analysis (CLSA). Internal and external standards can be used to improve reproducibility. Multiple headspace extraction (MHE) can be applied, which is based on a stepwise gas extraction with intermediate HSA. A simple two-step procedure is adequate for practical applications. The optimum static HSA sampling conditions are pH 7.1, 50°C, and  $3.35 \text{ mol l}^{-1}$  sodium sulfate for adjusting ionic strength.

**Purge and trap techniques** With dynamic procedures, the volatile compounds can be enriched to reach the detection limit of the chromatographic system. The trap packed with a solid sorbent can be considered as a chromatographic column. When the volume of the inert purge gas through the column becomes equal to the retention volume of the determinand (at a given temperature), it breaks through and elutes from the column. Knowledge of these breakthrough volumes is necessary. The breakthrough value depends on the form and size of the trap tube, specific area, and particle size of adsorbent, temperature, flow rate of the stripping gas, original concentration of the determinand in the sample, chemical structure of the determinand, and the purge gas. Elution of the volatiles from the trap is performed with small volumes of organic solvents such as methanol, isopropanol, or acetone.

Often, to increase the concentration to reach the detection limit it is necessary to concentrate the eluate by evaporation. Drawbacks of solvent elution are the introduction of artifacts by the solvent, glassware, and gas; the peak in the chromatogram due to the solvent washing the volatile organic compounds; and loss of volatiles by evaporation. Desorption of the volatiles from the sorbent of the trap can also be carried out thermally, in which case the sorbent must have thermal stability and there must be no reactions of the volatiles with the sorbent. For thermal desorption, more complicated equipment is necessary than for elution with a solvent. A suitable choice of sorbent is essential for a successful result. A great number of sorbents are used, such as activated carbon, graphitized carbon black, and porous polymers that exhibit a characteristic low capacity for water and do not show irreversible adsorption or decomposition phenomena, e.g., Porapak, Chromosorb, Amberlite XAD resins, Tenax-GC, Spheron, Synachrom, and Cehachrom. Tenax-GC is widely used owing to its thermal stability, in spite of its lower specific surface area. Compounds determined with purge-and-trap techniques include halomethanes, chlorobenzenes, ethylbenzene, xylenes, and short-chain halogenated hydrocarbons.

**Closed loop sampler** Another type of dynamic HSA is the technique in which a stream of inert gas is passed over the sample and led through a trap. This method is similar to the purge-and-trap method. However, the equilibrium between the water and gas phases depends on the flow of the stripping gas. The surface contact is much smaller than in the purge technique. This technique can be applied to samples with considerable foaming. Stripping gas (0.5 l) is recycled continuously through the sample. The adsorption trap (1.5 mg of activated carbon) is extracted with 10  $\mu$ l of carbon disulfide. The range of compounds that can be measured is somewhat limited. Peaks of highly volatile compounds are hidden by the peak of the extraction solvent and moderately or extremely polar organic series are either poorly purged or not recovered at all. The CLSA technique has been used most often for screening of large numbers of compounds; its outstanding concentration factor without the need for an evaporation step makes the system useful even when only a few compounds are of interest.

The static headspace procedure is the simplest and can be applied to organic compounds with high vapor pressure and low solubility in water. Only simple equipment is necessary and there are no great problems with impurities. However, the method is relatively insensitive because only a part of the vapor

can be injected and there is no enrichment procedure. Sensitivity can be improved by optimizing the conditions and the method can be automated. MHE has higher sensitivity than the static method and compounds with lower vapor pressure can be determined, but special equipment is necessary. The method can be automated. MHE is frequently used in drinking-water analysis for the determination of trihalomethanes. With CLSA, it is possible to determine the widest scope of compounds. However, this technique has some practical problems such as contamination of the filter or system. Another possibility is thermal desorption, which is also semiautomated, but special equipment is necessary for the analysis.

**Steam distillation** A Bleidner apparatus can be used for the extraction of aromatic amines. In this system, a continuous steam distillation or continuous LLE is carried out and good recoveries are obtained (over 90%). The extraction solvent is isooctane and the extraction time is 2–3 h; the technique is not laborious. Steam distillation extraction can be used quantitatively for different types of organic substances at microgram per liter to nanogram per liter levels. Good results can be obtained within 20 min. Steam distillation results in a 'cleaner' extract and reduces the build-up of residue at the GC column. The small volumes of solvent normally used in steam distillation techniques facilitate solvent concentration and minimize interferences due to solvent impurities.

Various commercial extractors are available. The steam codistillation extractor allows simultaneous condensation of a steam distillate and an immiscible extraction solvent, e.g., dichloromethane. The steam distillator extractor passes a sample steam distillate through an immiscible lighter-than-water solvent such as hexane.

**Cleanup** The concentration and/or isolation technique applied has a great influence on the cleanup procedure. HSA techniques such as static headspace and purge-and-trap with thermal desorption have the advantage that no further cleanup is necessary. These techniques are restricted to the analysis of volatile compounds. A high selectivity of the concentration and/or isolation technique decreases the necessity of a further cleanup procedure. By a proper selection of the extraction solvent or the sorbent, the determinands can be extracted and separated simultaneously from the majority of coextractives. Also, a derivatization procedure can increase the selectivity of the method. The following procedures can be used for cleanup.

**Liquid-liquid partitioning** Liquid-liquid partitioning between two immiscible phases will separate

compounds of different solubilities. Thus, if the extract is a nonpolar solvent such as petroleum ether, it is partitioned with a polar solvent such as acetonitrile.

**Liquid-solid chromatography** Liquid-solid chromatography is the technique most widely used to clean up sample extracts. Silica gel, alumina, and Fluorisil are the most commonly used adsorbents. The adsorbent should be free of organic contaminants, inert to the determinands, effective in removing interferences, and uniform in activity from batch to batch. Column chromatography can also be used for preliminary fractionation of sample extracts, e.g., separation of PCBs from organochlorine compounds.

**Gel permeation chromatography** Gel permeation chromatography (size-exclusion chromatography) can be used on humic acids or fulvic acids in water and is frequently applied to the analysis of fatty samples such as fish. Gel permeation chromatography is a very effective method based on differences in molecular mass. Large molecules are excluded from the pores of the gels and are eluted first. This technique can easily be automated. However, after gel permeation chromatography for fractionation, additional liquid-solid column chromatography is still necessary. Thus, there is little advantage in using gel permeation chromatography for the cleanup of less fatty or nonfatty samples such as water.

**Liquid chromatography** LC can be used as a cleanup technique. One of the biggest advantages of LC is the possibility of automation. However, LC analytical columns can be deactivated by sample coextractives and regeneration of the columns is time consuming. A solution is to use small precolumns in series with the analytical column; this minimizes the effect. Good results are obtained for chlorophenols and other aromatics.

**Derivatization** Chemical derivatizations can have advantages such as increase in volatility, thermal stability, sensitivity, greater selectivity, and improvement of peakshape and/or separation behavior. There are several types of derivatization techniques, which are as follows:

1. Chemical derivatization as a prerequisite for the method of analysis, e.g., acylation of chlorophenols: It must meet all requirements associated with a practical, usable analytical method, such as reproducibility, good recovery, no interferences.
2. Chemical derivatization confirmatory test: The demands are less severe since the main criteria are speed, size of operation, and acceptable yield.

3. Cleanup reaction: Chemical reactions to clean up or remove specific interferences.

Examples of applications of derivatization reagents include: diazomethane for acids and phenols; acetic anhydride for phenols; heptafluorobutyric acid for anilines and phenols.

Derivatization can be carried out before instrumental analysis, or is often an integral part of the instrumental analysis such as oncolumn derivatization (GC); precolumn derivatization (LC); postcolumn derivatization (LC).

The lack of suitable detectors in LC for trace and ultratrace analysis of complex matrices has catalyzed this trend. The major advantage of postcolumn reaction detectors is the absence of artifact information, and the only requirement is good reproducibility. A disadvantage is the great influence of the mobile phase on the reaction medium.

Chemical reactions for cleanup are especially interesting in the analysis of organic compounds at very low concentrations in samples contaminated with many interfering compounds. One of the remarkable features of reaction detectors is their ability to fit into automated analytical systems. Their selectivity and sensitivity in the reaction itself will often permit a drastic reduction in the amount of sample preparation or the requirements for a powerful separation. Band broadening and interference of the detection signal with the reagent can be a problem.

**Chromatographic separation and detection** A wide range of analytical detection techniques are available for the determination of organic pollutants. GC is used on a large scale in the analysis of organic pollutants in samples. Capillary columns, selective detectors, and sample introductory techniques are widely used. Detectors for GC are summarized in **Table 6**. With the availability of low-inner-volume detectors, sophisticated sample introduction techniques, high quality of the capillary columns, multidimensional GC, derivatizing reagents, and multidetection systems, very highly efficient and sensitive systems are available for the analysis of trace organic pollutants in samples. The selection of the right chromatographic column and of the instrumental parameters is very important to get the optimal analytical tool. For the choice of a GC column, one should try to separate the particular sample on a small number of standard columns using the following guidelines. Study the table of stationary phases in the catalogs of the column supplier. Select a polar stationary phase to separate polar compounds and an apolar stationary phase for apolar compounds. Use the polarity index. Establish the appropriate internal diameter.

**Table 6** Detectors for gas chromatography

<i>Detector</i>	<i>Elemental or compound information</i>	<i>Minimum detectable amount</i>	<i>Linear dynamic range</i>	<i>Remarks</i>
AASD (atomic absorption spectrometric)	Monofunctional Hg	$10^{-10}$ g Hg		Expensive
AFID (alkali flame ionization)	Nitrogen/phosphorus	$10^{-13}$ – $10^{-14}$ g N or S	$10^4$ – $10^5$	Destructive; can also be used for S, As, Sb, Sn, Pb, Cl, I, F, Br
Coulometric	Cl, Br, I	$2 \times 10^{-13}$ – $10^{-15}$ g P s <sup>-1</sup>		
ECD (electron-capture)	Electron affinity, Cl, Br, I, F, nitrocompounds, conjugated carbonyls	$10^{-9}$ g $10^{-14}$ – $10^{-12}$ g aldrin(s)	$10^7$ $10^4$	Difficult in operation Cleanup of samples important
FID (flame ionization)	All compounds	0.02 coulomb per gram carbon	$10^6$ – $10^7$	Destructive; universal detector
FPD (flame photometric)	P, S	$10^{-12}$ g P s <sup>-1</sup>	$10^4$ – $10^8$	Destructive; also halogens
FTIR (Fourier transform infrared)	Functional group information	$5 \times 10^{-12}$ – $10^{-10}$ g S s <sup>-1</sup> $10^{-9}$ g	$10$ – $10^3$	Moderately expensive; discrimination between isomers
HECD (Hall electrolytic conductivity)	N, S, halogens	$10^{-9}$ g	$10^4$	Destructive; difficult in operation
MS (mass spectrometric)	Elemental and/or molecular information	$10^{-12}$ – $10^{-9}$ g		Destructive; expensive; high specificity
Microwave-induced plasma	Elemental	$10^{-9}$ g		Destructive; expensive; problems with O, P and N
PID (photoionization)	Aromatic compounds	$10^{-12}$ g	$10^7$	High selectivity
TCD (thermal conductivity)		$10^{-7}$	$10^5$	Nonspecific

Reprinted with permission from Wegman RCC and Melis HAM (1986) *Organic Pollutants in Water*, CRC Critical Reviews in Analytical Chemistry, vol. 16, pp. 281–321. Boca Raton, FL: CRC Press.

Choose the appropriate column length and determine the optimum film thickness. The use of chemically bonded columns has many advantages. Rearrangement of the stationary phase film is impossible because of the chemical bonding; therefore, the efficiency of the columns is prolonged. Contaminated columns may be washed with solvents to regain their original performance.

The main benefit of multiple detection is the simultaneous registration of specific and selective items by a single injection, markedly reducing the analysis time. Destructive detectors can be mounted directly on nondestructive ones. Multidimensional GC enhances selectivity. Interesting peak groups from a complex mixture, eluting from a first capillary or packed column, can be transferred to a second column of high selectivity, for example, for organochlorines and PCBs. Capillary GC–MS is routinely used in water analysis, whereas GC–Fourier transform infrared (FTIR) has fewer practical applications. Capillary GC–FTIR has developed into an online technique whereby spectral data can be obtained in a short period of time even for difficult samples. Both techniques have sensitivities compatible

with capillary columns. Generally, capillary columns with higher capacity columns are useful in GC–FTIR. Considerably more sample is necessary for GC–FTIR analysis than for GC–MS. To solve identification problems, use of GC–MS and GC–FTIR has advantages and decreases total analysis time. It is also possible to use a combination of MS and FTIR with a built-in combined MS/IR library search.

LC is used for the analysis for a wide range of compounds too involatile or too thermolabile for analysis by GC. The analytical application of LC depends to a great extent on the availability of detectors with high sensitivity and adequate selectivity. The latter requirement is generally much more important in LC than in GC. In GC analysis of water it is certain that compounds with a high molecular mass will not generate peaks in the chromatogram of determinands with a lower molecular mass. However, interference of high-molecular-mass material is an omnipresent problem in trace analysis with LC. Selectivity can be improved by selective extraction, matrix removal, high efficiency of separation (by the suitable choice of phase system or multicolumn operation), and by selective detection. Important factors

**Table 7** Detectors for liquid chromatography

<i>Detector</i>	<i>Elemental or compound information</i>	<i>Sensitivity (g ml<sup>-1</sup>)</i>	<i>Linear dynamic range</i>	<i>Remarks</i>
UV-visible absorbance	Aromatics	$2 \times 10^{-10}$	$10^5$	Many applications; identification by diode array spectra
Conductivity	Ions	$10^{-5}$	$10^4$	
Amperometric	Oxidizable and reducible compounds	$10^{-12}$	$10^6$	
Fluorescence	PAHs, phenols	$10^{-11}$	$10^3$	High selectivity
Infrared	Macromolecules	$10^{-6}$	$10^4$	High selectivity
Refractive index	Alkanes	$10^{-7}$	$10^4$	Universal

Reprinted with permission from Wegman RCC and Melis HAM (1986) *Organic Pollutants in Water*, CRC Critical Reviews in Analytical Chemistry, vol. 16, pp. 281–321. Boca Raton, FL: CRC Press.

**Table 8** Costs of steps of analysis in routine procedures for determination of organic compounds

<i>Technique</i>	<i>Cost of analysis (US\$)<sup>a</sup></i>
Filtration of water sample	20
Extraction (liquid-liquid)	35
Extraction (sorbent)	35
Purge-and-trap	35
Cleanup with sorbent	55
Derivatization	85
Instrumental detection, e.g., GC	55
GC-MS confirmation of compounds in addition to normal method	105
GC-MS detection, unknown compounds	760
GC-MS screening, known compounds	300

<sup>a</sup>Base year for prices, 1992.

that can accomplish selective LC detection are UV wavelength, simultaneous multiwavelength (photo-diode array), fluorescence, electrochemistry, atomic spectrometry, electron capture, homogeneous post-column reaction, and postcolumn reaction with extraction. Detectors for LC are summarized in Table 7.

LC is a suitable technique for trace analysis of organic micropollutants. However, the separation power of common LC columns is insufficient for very complex mixtures. Online coupling of LC with MS will be an attractive tool for analysis of low-volatility organic compounds in samples. However, cleanup of the sample prior to LC-MS analysis will be very necessary. The sensitivity is less than that obtained by capillary GC-MS combinations and at present there is no universal LC-MS system for all types of compounds.

**Table 9** Total cost of routine analysis of groups of compounds

<i>Compounds</i>	<i>Steps in procedure</i>	<i>Cost of analysis (US\$)<sup>a</sup></i>
Anilines, chlorinated	Bleidner extraction derivatization, GC-ECD	215
Aromatics, volatile and chlorinated hydrocarbons	Purge-and-trap, thermal desorption, GC-FID	90
PAHs 6 of Borneff	Liquid-liquid extraction, cleanup with alumina, LC-UV/fluorescence	130
PAHs 16 of EPA	Liquid-liquid extraction, cleanup with alumina, LC-UV/fluorescence	150
PCBs and organochlorine pesticides	Liquid-liquid extraction, cleanup with alumina and silica, GC-ECD	140
Phenols, chlorinated	Liquid-liquid extraction, acetylation, GC-ECD	170
Phthalates	Liquid-liquid extraction, GC-FID	140
Solvents, e.g., benzene, toluene, xylenes, di- and trichloromethane	Liquid-liquid extraction, GC-FID/ECD	125
Water-soluble solvents, e.g., acetone, alcohols, methyl- and other acetates	Direct injection, GC-FID	175

<sup>a</sup>Base year for prices, 1992.



In order to give an idea of the costs of the steps in the analytical procedure, these costs are summarized in Table 8. As an example, Table 9 gives the total costs of the analysis of different groups of compounds.

## Emerging Techniques

Miniaturization is the most promising development in analytical chemistry. With miniaturized analyzers determinations can be performed when needed and at any location. To date, CZE, miniaturized to chips, has been developed, but only for the analysis of pharmaceutical substances. The field developments in water analysis have been limited to such multiprobe measurements as pH, temperature, and oxygen. The Hydrion 10 is the instrument on the market that has proven capable of performing a complete analysis of a water sample for the main constituents within a matter of minutes. The development of such devices will continue, with the expectation that apparatus for rapid analysis of samples in the field will be both easily available and more readily applicable in future. However, much of the present methods of analysis are time consuming and costly. Moreover, the results are not readily available. Therefore, direct intervention is not possible. This was passed when sensors became available. With sensors major ions can be measured in water in a manner that is fast, inexpensive, and independent of laboratory tests. Because the measurement results are directly available, it is possible to apply them immediately. Next in new developments are early warning detection and monitoring with biomonitors (*Daphnia*, algae, fish, etc.) that measure changes in the functions of organisms in response to pollutants will respond to several toxic compounds. This future activity could reshape new challenges to expand biological early warning systems and cytosensors for possible compounds such as atrazine, malathion, diuron, dieldrin, dioxins, estrogens, mutagens, aldrin, cadmium, etc.).

The monitoring of compounds with dedicated analytical methods is often insufficient to assess and maintain the quality of surface waters. Many unknown compounds can be present, some of them are harmful for the environment and also for humans. For the identification of semipolar and polar target compounds present in water, LC database fingerprints in combination with LC-MS/MS identification or LC database toxic fingerprint in combination with LC-MS/MS identification can improve the efficiency on relevant compounds.

**See also:** **Amperometry.** **Atomic Absorption Spectrometry:** Electrothermal. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Derivatization of Analytes.** **Distillation.** **Electrophoresis:** Principles. **Extraction:** Solvent Extraction Principles. **Flow Injection Analysis:** Environmental and Agricultural Applications. **Gas Chromatography:** Detectors; Mass Spectrometry. **Ion Exchange:** Ion Chromatography Instrumentation. **Liquid Chromatography:** Overview; Size-Exclusion; Liquid Chromatography–Mass Spectrometry. **Nitrogen.** **pH.** **Polychlorinated Biphenyls.** **Sampling:** Theory. **Segmented Flow Analysis.** **Sensors:** Overview. **Sensory Evaluation.** **Sulfur.** **Water Analysis:** Freshwater.

## Further Reading

- Dits JS, van der Jagt H, and Smeenk JGMM (1992) The odour/taste dilution factor and objective measurement for determining water quality. *Water Supply* 10: 11–21.
- International Organization for Standardization (1985) *Water Quality – Sampling, part 3: Guidance on the Preservation and Handling of Samples* (Revision of ISO 5667/3). Geneva: ISO.
- van der Jagt H and Stuyfzand PJ (1991) *Methods of Analysis for Trace Elements in Groundwater; Atomic Spectrometrical Methods. Trace Elements in Groundwater.* Nieuwegein: KIWA, report no. 118.
- van der Jagt H (1999) Evaluation of public health related quality aspects of materials coming into contact with drinking water-odour and flavour tests in particular. *Water Science Technology* 40(6): 239–244.
- van der Jagt H, Noij THM, Ooms PCA, *et al.* (1993) *Analysis and Identification of Bromate in Water by Ion Chromatography and Multiple Detection at the Low Detection Level.* IWSA, Paris, Proceedings Report, Nov. 1993.
- van der Jagt, H, and van Berkel PM, *et al.* (1993) *Comparability of Analytical Results by ICP-EAS and AAS in Different Matrices.* STOWA Report in Dutch, 99-03.
- Koudjonou BK, Müller MC, Van der Jagt H, *et al.* (1995) Bromate ion Analysis by ion chromatography. *Ozone Science and Engineering* 17: 561–573.
- Reijnders HF, Melis HAM, and Griepink B (1983) Durchflussmethoden zur Bestimmung von Hauptkomponenten in Gewässern, Regenwasser und Trinkwasser. *Fresenius Zeitschrift für Analytische Chemie* 314: 627–633.
- Small H (1989) *Ion Chromatography.* New York: Plenum Press.
- Smith CJ, *et al.* (2003) Xeno-oestrogens. *Analytical Chemistry* 75(6): 1463.
- Wegman RCC and Melis HAM (1986) *Organic Pollutants in Water.* CRC Critical Reviews in Analytical Chemistry, vol. 16, pp. 281–321. Boca Raton, FL: CRC Press.



## Potable Water

**C W K Chow**, Australian Water Quality Centre, Salisbury, SA, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

‘Water’ is one of the most abundant substances in the world. However, the conversion of water from various sources to good-quality drinking water can be a very demanding task for drinking water suppliers. The definition of drinking water commonly refers to water that is intended for human consumption and other domestic uses. Ideally, drinking water should be clear, colorless, and without unpalatable taste or odor. In addition, it must not contain harmful chemical substances, either organic or inorganic, and pathogenic microorganisms that may be harmful to human health.

The quality of drinking water in relation to public health is an important issue worldwide. Water may come from groundwater sources or surface waters such as lakes, reservoirs, and rivers with a wide variation in water quality. The treatment required to achieve potability depends upon the characteristics of the source, relevant drinking water standards, and the characteristics of the distribution system. Treatment of some supplies consists of disinfection only while others may use several clarification steps prior to disinfection.

*The Guidelines for Drinking-Water Quality* published by the World Health Organization (WHO) is the key international reference point for standard setting and drinking water safety. The WHO has developed a series of guidelines that present an assessment of the health risks associated with exposure to health hazards through water. The information is essential to assist water suppliers to implement their own safe water acts.

Drinking water suppliers must have a good quality assurance program to ensure that the product (drinking water) they supply to the consumer meets the relevant standards. Analysis of drinking water becomes an important step of the program. It is necessary to monitor key locations in the water supply system with the support of good analytical procedure and quality control to ensure the accuracy of the analysis.

### Drinking Water Guidelines and Standards

In 1984, the term ‘guidelines’ was first introduced by the WHO, which were intended to be used as a basis

for the development of national standards to ensure the safety of drinking water supplied to the consumer. Drinking water standards are regulations that water authorities set to control the level of contaminants in their drinking water. The advantage of using guidelines rather than the adoption of international standards for drinking water quality is the flexibility of the risk–benefit approach in the establishment of local standards and regulations. This could allow standards and regulations to be implemented with different priorities according to the local or national, environmental, social, economic, and cultural conditions. In developing standards, it is important to realize that the standards should be both practical and feasible to implement. They need to take into account the likelihood that the contaminant occurs in a public water system, the levels of concern, and the determination limits of the analytical methods. This can prevent the situation of unnecessary monitoring for relatively minor concern substances.

Chemicals, arising from natural or human activities, can enter the water supply and potentially contaminate the water. Inorganic chemicals such as mineral salts leached from mining sites, organic chemicals such as algal toxins as a result of an algal bloom, man-made chemicals such as pesticides (insecticides and herbicides), pharmaceuticals, and industrial waste products can enter the water supply. Some potentially hazardous chemicals in drinking water are derived directly from treatment chemicals, such as the unreacted monomer from polyelectrolytes used as coagulant aids, and disinfectant reactions with organic matter to form disinfection by-products. Lists of contaminants of concern in drinking water are shown in **Tables 1** and **2** for inorganic and organic contaminants, respectively.

Most water authorities have set standards for drinking water quality. Authorities who play a major role in developing drinking water standards are the United States Environmental Protection Agency (USEPA) and the European Economic Community (EEC). Other countries including Canada, Germany, United Kingdom, Netherlands, Czech Republic, South Africa, China, Taiwan, have also implemented standards to regulate the delivery of safe drinking water to the consumers. In Australia, the National Health and Medical Research Council (NHMRC) in conjunction with the Australian Water Resources Council (AWRC), and more recently with the Agricultural & Resource Management Council of Australia and New Zealand (ARMCANZ) have developed guidelines that are based on the WHO

**Table 1** Contaminant concentrations or values of physical characteristics for inorganic chemicals and radioactive substances in drinking water

	<i>Description and possible sources of contamination</i>	<i>Sam. Loc.</i>	<i>WHO drinking water guideline value</i>	<i>USEPA maximum contamination level (MCL)</i>	<i>USEPA recommended method</i>
Alpha particles (gross alpha activity)	Certain minerals are radioactive and may emit a form of radiation called alpha radiation	RW	0.1 Bq l <sup>-1</sup>	15 pCi l <sup>-1</sup>	Evaporation, co-precipitation
Aluminum	From natural sources and used as coagulant for water treatment	DS	0.2 mg l <sup>-1</sup> *	0.2–0.5 mg l <sup>-1</sup> *	ICP-AES, ICP-MS, AAS
Ammonia (as NH <sub>3</sub> )	Presence may indicate sewage contamination and/or microbial activity	DS	1.5 mg l <sup>-1</sup> *	No value set	ISE, colorimetric (indophenol)
Antimony	From natural sources or discharge from petroleum refineries, fire retardants, electronics, and solder	SC	0.005 mg l <sup>-1</sup> p	0.006 mg l <sup>-1</sup>	ICP-MS, AAS
Arsenic	From natural sources or discharge from mineral/industries/agricultural wastes	RW	0.01 mg l <sup>-1</sup> p	0.01 mg l <sup>-1</sup>	ICP-AES, ICP-MS, AAS
Asbestos	From dissolution of mineral/industrial waste and deterioration of cement water mains	DS	U	7 million fibers per liter	TEM
Barium	From natural sources or discharge from industries	RW	0.7 mg l <sup>-1</sup>	2 mg l <sup>-1</sup>	ICP-AES, ICP-MS
Beryllium	From natural sources or discharge from industries	RW	NAD	0.004 mg l <sup>-1</sup>	ICP-AES, ICP-MS, AAS
Beta-particles (gross beta activity)	Certain minerals are radioactive and may emit a form of radiation called beta-radiation	RW	1 Bq l <sup>-1</sup>	4 millirems per year	Evaporation, co-precipitation
Boron	From mineral leaching and may be associated with seawater intrusion	RW	0.5 mg l <sup>-1</sup> p	No value set	Colorimetric
Bromate	Possible by-product of disinfection using ozone	SC	25 µg l <sup>-1</sup> p	0.01 mg l <sup>-1</sup>	IC
Cadmium	Indicates industrial or agricultural contamination	SC	0.003 mg l <sup>-1</sup>	0.005 mg l <sup>-1</sup>	ICP-AES, ICP-MS, AAS
Chloramine (Monochloramine)	Disinfection agent	SC	3 mg l <sup>-1</sup>	4 mg l <sup>-1</sup> as Cl <sub>2</sub> (MRDL)	Amperometric titration, DPD iron(II) titration, DPD colorimetric
Chloride	From natural mineral salts. No health-based guideline value	TW	250 mg l <sup>-1</sup> *	250 mg l <sup>-1</sup> *	IC
Chlorine	Disinfection agent. Commonly analyzed using DPD (N,N-diethyl-p-phenylenediamine) colorimetric methods	SC	5 mg l <sup>-1</sup> , 0.6–1.0 mg l <sup>-1</sup> *	4 mg l <sup>-1</sup> (MRDL)	Amperometric titration, DPD iron(II) titration, DPD colorimetric
Chlorine dioxide	Disinfection agent	SC	NAD, 0.4 mg l <sup>-1</sup> *	0.8 mg l <sup>-1</sup> as ClO <sub>2</sub> (MRDL)	Amperometric titration, DPD methods
Chlorite	By-product of chlorine dioxide disinfectant	SC	200 µg l <sup>-1</sup> p	1 mg l <sup>-1</sup>	IC
Chromium (as Cr(VI))	Indicates industrial or agricultural contamination	DS	0.05 mg l <sup>-1</sup> p	0.1 mg l <sup>-1</sup>	ICP-AES, ICP-MS, AAS
Copper	Corrosion of copper pipes/fittings by soft, low pH water	SC	2 mg l <sup>-1</sup> p, 1 mg l <sup>-1</sup> *	1.3 mg l <sup>-1</sup> , 1 mg l <sup>-1</sup> *	ICP-AES, ICP-MS, AAS
Cyanide	From industrial waste and some plants and bacteria	DS	0.07 mg l <sup>-1</sup>	0.2 mg l <sup>-1</sup>	Colorimetric
Fluoride	Occurs naturally and water additive which promotes strong teeth	TW	1.5 mg l <sup>-1</sup>	4 mg l <sup>-1</sup> , 2 mg l <sup>-1</sup> *	IC
Hydrogen sulfide	Formed in water by sulfate-reducing microorganism	DS	0.05 mg l <sup>-1</sup> *	No value set	Colorimetric, iodometric, ISE
Iron	Occurs naturally. No health-based guideline value	SC	0.3 mg l <sup>-1</sup> *	0.3 mg l <sup>-1</sup> *	ICP, AAS
Lead	Dissolution from natural sources or household plumbing	SC	0.01 mg l <sup>-1</sup>	0.015 mg l <sup>-1</sup>	ICP-MS, AAS
Manganese	Occurs naturally. No health-based guideline value	DS	0.5 mg l <sup>-1</sup> p, 0.1 mg l <sup>-1</sup> *	0.05 mg l <sup>-1</sup> *	ICP-AES, ICP-MS, AAS

Mercury	From industrial emissions/spills	RW	0.001 mg l <sup>-1</sup>	0.002 mg l <sup>-1</sup>	Cold vapor AAS, ICP-MS
Nickel	From nickel-plated fittings	SC	0.02 mg l <sup>-1</sup>	0.1 mg l <sup>-1</sup>	ICP-AES, ICP-MS, AAS
Nitrate (as NO <sub>3</sub> <sup>-</sup> )	Occurs naturally	DS	50 mg l <sup>-1</sup>	10 mg l <sup>-1</sup>	IC, colorimetric
Nitrite (as NO <sub>2</sub> <sup>-</sup> )	Rapidly oxidized to nitrate	DS	3 mg l <sup>-1</sup> (acute), 0.2 mg l <sup>-1</sup> (chronic) <sup>p</sup>	1 mg l <sup>-1</sup>	IC, colorimetric
pH	<6.5 may cause corrosion, >8 decreases efficiency of chlorination	DS	<8.0*	6.5–8.5*	Electrometric
Radium-226	Radionuclide	RW	No value set	5 pCi l <sup>-1</sup>	Radiochemical, radon emanation
Selenium	Generally very low concentration	RW	0.01 mg l <sup>-1</sup>	0.05 mg l <sup>-1</sup>	ICP-MS, AAS
Silver	Silver and silver salts occasionally used for disinfection. No health-based guideline value	RW	U	0.1 mg l <sup>-1</sup> *	ICP-AES, ICP-MS, AAS
Sodium	Natural component of water. No health-based guideline value	TW	200 mg l <sup>-1</sup> *	No value set	ICP-AES
Sulfate	Natural component of water, may be added via treatment chemicals	TW	250 mg l <sup>-1</sup> *	250 mg l <sup>-1</sup> *	IC, colorimetric
Taste and odor	No health-based guideline value. The threshold odor number (TON) is the greatest dilution of sample with odor-free water yielding definitely perceptible odor	SC	No value set	3 TON*	Flavor profile analysis
Total dissolved solids (TDS)	<500 mg l <sup>-1</sup> is regarded as good-quality drinking water	TW	1000 mg l <sup>-1</sup> *	500 mg l <sup>-1</sup> *	TDS
True color	15 True color unit (TCU) just noticeable in a glass	DS	15 TCU*	15 TCU*	Visual comparison, spectrophotometric
Turbidity	Turbidity is a measure of the cloudiness of water. Nephelometric turbidity unit (NTU) 5 just noticeable in a glass	DS	5 NTU*	1 NTU	Nephelometric
Zinc	Corrosion of galvanized pipes/fittings and brasses	SC	3 mg l <sup>-1</sup> *	5 mg l <sup>-1</sup> *	ICP-AES, ICP-MS

Sam. Loc. (sample location): RW – raw water, TW – treated water, DS – distribution system, and SC – supply to consumer 'consumer taps'.

U: It is unnecessary to recommend a health-based value for these compounds because they are not hazardous to human health at concentrations normally found in drinking water.

NAD: No adequate data to permit recommendation of a health guideline value.

\*Aesthetic guideline value.

p: provisional guideline value.

USEPA MCL: the highest level of a contaminant that is allowed in drinking water.

MRDL: maximum residual disinfection level.

Analytical methods: AAS, atomic absorption spectroscopy; ICP-AES, inductively coupled plasma-atomic emission spectroscopy; ICP-MS, inductively coupled plasma-mass spectrometry; IC, ion chromatography; ISE, ion selective electrode; TEM, transmission electron microscopy.

Standard methods for the examination of water and wastewater (shaded portions): The detailed information of this table is obtained from WHO – *Guidelines for Drinking-Water Quality*, 2nd edn (1993), USEPA – *National Primary Drinking Water Standards* (1995), NHMRC/ARMCANZ – *Australian Drinking Water Guidelines* (1996), USEPA – *Methods and Guidance for Analysis of Water*, Version 2 (1999), and APHA – *Standard Methods for the Examination of Water and Wastewater*, 20th edn (1988). This table is produced to illustrate the contaminant of concerns and their concentrations in drinking water in relation to the analytical techniques used for the determination. For detailed analytical procedure, derivation of guideline/standard values, and health-related concern of contaminant level in drinking water consult the original publication cited and more preferable to obtain the latest version of the publication cited.

**Table 2** Contaminant concentrations for organic compounds in drinking water

	<i>Description and possible sources of contamination</i>	<i>Sam. Loc.</i>	<i>WHO drinking water guideline value</i>	<i>USEPA maximum contamination level (MCL)</i>	<i>USEPA recommended method</i>
Acrylamide	Minor impurity of polyacrylamide (flocculant aid)	TW	0.5 µg l <sup>-1</sup> <sup>p</sup>	Treatment chemical: <0.5 µg l <sup>-1</sup>	Polarography, GC-ECD, HPLC-UV
Benzene	Could occur in drinking water from atmospheric deposition and chemical plant effluent. Human carcinogen	RW	10 µg l <sup>-1</sup> <sup>p</sup>	5 µg l <sup>-1</sup>	GC-PID/ELCD, GC-MS
Carbon tetrachloride	Impurity of chlorine used for disinfection (it is not disinfection by-product) or discharge from chemical plants	TW	2 µg l <sup>-1</sup>	5 µg l <sup>-1</sup>	GC-PID/ELCD, GC-MS, Liq-Liq Ext GC-ECD
Chlorobenzene	From spills/discharges	RW	300 µg l <sup>-1</sup> , 10–120 µg l <sup>-1</sup> <sup>*</sup>	100 µg l <sup>-1</sup>	GC-PID/ELCD, GC-MS
Dichloroethanes (1,2-dichloroethane)	From industrial effluents, spills, discharges	RW	30 µg l <sup>-1</sup>	5 µg l <sup>-1</sup>	GC-PID/ELCD, GC-MS
Trichloroethane (1,1,1-trichloroethane)	Same as dichloroethanes	RW	2000 µg l <sup>-1</sup> <sup>p</sup>	200 µg l <sup>-1</sup>	GC-PID/ELCD, GC-MS, Liq-Liq Ext GC-ECD
Dichloromethane (methylene chloride)	Widely used solvent, commonly found in ground water	RW	20 µg l <sup>-1</sup>	5 µg l <sup>-1</sup>	GC-PID/ELCD, GC-MS
Epichlorohydrin	Used in manufacture of some resins used in water treatment	TW	0.4 µg l <sup>-1</sup> <sup>p</sup>	Treatment chemical: <2 µg l <sup>-1</sup>	GC-MS, GC-ECD, GC-FID
Ethylbenzene	Natural component of petrol and petroleum products	RW	300 µg l <sup>-1</sup> , 2–200 µg l <sup>-1</sup> <sup>*</sup>	700 µg l <sup>-1</sup>	GC-PID/ELCD, GC-MS
Haloacetic acids	By-product of chlorination	DS	NAD	0.06 mg l <sup>-1</sup>	Ion-Ex Liq-Solid Ext GC-ECD, Liq-Liq Ext GC-ECD
Di(2-ethylhexyl) adipate	Plasticizers used in all PVC products and may leach over a long time from linings in water storage tanks and distribution lines. Could also occur in drinking water as spills	SC	80 µg l <sup>-1</sup>	400 µg l <sup>-1</sup>	Liq-Liq Ext GC-PID, Liq-Solid Ext GC-PID, Liq-Solid Ext GC-MS
Di(2-ethylhexyl) phthalate	As above	SC	8 µg l <sup>-1</sup>	6 µg l <sup>-1</sup>	Liq-Liq Ext GC-PID, Liq-Solid Ext GC-PID, Liq-Solid Ext GC-MS
Benzo(a)-pyrene	One of a group of compounds. Polycyclic aromatic hydrocarbons (PAHs). Leaching from linings of water storage tanks	SC	0.7 µg l <sup>-1</sup>	0.2 µg l <sup>-1</sup>	Liq-Solid Ext GC/MS
Pesticides and related compounds	Pesticides should only be authorized for use in water or water catchment areas where necessary. Pesticides not authorized for uses should not be present in drinking water. Contamination by pesticides may occur as a result of accidental spills	RW	Low µg l <sup>-1</sup> range <sup>+</sup>	Low µg l <sup>-1</sup> range <sup>+</sup>	Requires sample preconcentration followed by one of the methods listed: GC-ECD, GC-NPD, GC-MS, LC-UV, LC-fluorescence

Styrene (vinylbenzene)	Industrial contaminant, discharge from rubber and plastic factories. Leaching from landfills	RW	20 µg l <sup>-1</sup> , 4–2600 µg l <sup>-1</sup> *	100 µg l <sup>-1</sup>	GC-PID/ELCD, GC-MS
Trihalomethanes (THMs) (total)	By-product of chlorination, chloramination	DS	No value set	100 µg l <sup>-1</sup>	GC-PID/ELCD, GC-MS, Liq-Liq Ext GC-ECD
Bromoform	By-product of chlorination, chloramination	DS	100 µg l <sup>-1</sup>	No value set	As above
Dibromochloromethane	By-product of chlorination, chloramination	DS	100 µg l <sup>-1</sup>	No value set	As above
Bromodichloromethane	By-product of chlorination, chloramination. Considered to be carcinogenic	DS	60 µg l <sup>-1</sup>	No value set	As above
Chloroform	By-product of chlorination, chloramination. Considered to be carcinogenic	DS	200 µg l <sup>-1</sup>	No value set	As above
Toluene	Occurs naturally in petrol and natural gas, forest fire emissions. Could occur in drinking water from atmospheric deposition, industrial contamination, leaching from protective coatings in storage tanks	DS	700 µg l <sup>-1</sup> , 24–170 µg l <sup>-1</sup> *	1000 µg l <sup>-1</sup>	GC-PID/ELCD, GC-MS
Vinyl chloride	From chemical spills. Used in making PVC pipes. Human carcinogen	RW	5 µg l <sup>-1</sup>	2 µg l <sup>-1</sup>	GC-PID/ELCD, CG-MS
Xylene	Pollutant from solvent used for bonding plastic fittings	DS	500 µg l <sup>-1</sup> , 20–1800 µg l <sup>-1</sup> *	10 mg l <sup>-1</sup> (total xylenes)	GC-PID/ELCD, GC-MS

Sam. Loc. (sample location): RW – raw water, TW – treated water, DS – distribution system, and SC – supply to consumer 'consumer taps'.

NAD: no adequate data to permit recommendation of a health guideline value.

\*Aesthetic guideline value.

p: provisional guideline value.

+: Due to the number of pesticides listed in the original references, a general item as pesticides and related compounds is listed in the table. For individual guideline or standard values refer to WHO – *Guidelines for Drinking-Water Quality* (1993), USEPA – *National Primary Drinking Water Standards* (1999), and NHMRC/ARMCANZ – *Australian Drinking Water Guidelines* (1996).

Generally, organic compounds present in drinking water are in low µg l<sup>-1</sup> concentration. Organic analysis often coupled with a sample pretreatment step as part of the analytical procedure, Liq-Liq Ext: liquid-liquid extraction, Liq-Solid Ext: liquid-solid extraction, Ion-Ex Liq-Solid Ext: ion-exchange liquid-solid extraction.

Analytical techniques: GC, gas chromatography coupled with various detectors; ECD, electron capture detector; ELCD, electrolytic conductivity detector; FID, flame ionization detector; MS, mass spectrometry; NPD, nitrogen phosphorus detector; and PID, photoionization detector; LC, liquid chromatography with UV detector or fluorescence detector.

Methods recommended by WHO – *Guidelines for Drinking-Water Quality*, 2nd edn (1993), vol. 2, Health Criteria and other Support Information (shaded portions).

The detailed information of this table is obtained from WHO – *Guidelines for Drinking-Water Quality*, 2nd edn (1993), USEPA – *National Primary Drinking Water Standards* (1995), NHMRC/ARMCANZ – *Australian Drinking Water Guidelines* (1996), USEPA – *Methods and Guidance for Analysis of Water*, Version 2 (1999), and APHA – *Standards Methods for the Examination of Water and Wastewater*, 20th edn (1988). This table is produced to illustrate the contaminant of concerns and their concentrations in drinking water in relation to the analytical techniques used for the determination. For detailed analytical procedure, derivation of guideline/standard values and health-related concern of contaminant level in drinking water consult the original publication cited and more preferable to obtain the latest version of the publication cited.

guidelines and take into account local conditions. All these authorities use slightly different approaches to establish their standards or acceptance levels. Differences in the values set by authorities are usually due to a local or national problem. A list of contaminants and the maximum contaminant levels (MCL) recommended by the WHO and USEPA is presented in Tables 1 and 2. In addition, to ensure the supply of good-quality drinking water, there are also aesthetic (nonhealth based) guideline values recommended by WHO.

## Water Management Framework

There has been a strong focus on numerical standard values and limits and monitoring of these constituents for compliance to guarantee the safety of water supplies. Monitoring also enables appropriate preventive measures to be taken before failure or contamination occurs. A more proactive approach by using a preventive management concept has been described in the *WHO Guidelines for Drinking-Water Quality* and this strategy was recently developed by NHMRC/ARMCANZ into a water management framework, protection of water from source (catchment) through to the consumer. This approach encourages water authorities to manage their water supply systems by putting in more effort on understanding their entire systems and the development of preventive strategies necessary for ensuring the control of hazards entering the system and, if necessary, the removal of hazards using various treatment options.

### Catchment and Source Water Protection

Water collected from the environment contains trace quantities of dissolved and suspended matter of mineral, plant, and even animal origin, which may create color, taste, and odor. It is also very likely to contain a diverse population of microorganisms and chemicals that may put human health at risk. The protection of water sources from human or animal pollutants should minimize contamination.

### Water Treatment

**Clarification of water** The treatment of drinking water provides multiple barriers to protect public health by removing microorganisms, and natural and man-made chemicals that may cause illness in consumers. Generally, a conventional treatment process employs coagulation/flocculation–sedimentation–filtration, and various modifications of the above configuration can be adopted based on specific needs. Aluminum- and iron-based chemicals, particularly alum ( $\text{Al}_2(\text{SO}_4)_3 \cdot 18\text{H}_2\text{O}$ ) and iron(III) chloride,

are widely used coagulants for drinking water treatment. The use of synthetic organic polymers to destabilize particles has gained acceptance as an alternative coagulant. Positively charged (cationic) polymers, such as polydiallyldimethylammonium chloride, are the most common types used. Other types of polymers are used as coagulant aids to aid the growth of the floc and can be applied to assist the treatment process.

Membrane filtration for drinking water treatment is a relatively new technology. This chemical-free process can be considered as an alternative to the conventional treatment process.

**Disinfection** The biggest risk to public health when managing the supply of municipal potable water arises from inadequate treatment of microbiologically contaminated water. Water should be disinfected before entering the distribution system. Agents and processes that have been used for the disinfection of water include chlorine, chloramines, chlorine dioxide, ozone, bromine, iodine, silver, and ultraviolet radiation. Generally, an adequate concentration of disinfectant should be maintained throughout the distribution system to prevent bacterial recontamination and regrowth.

### Distribution System

Moving water to the customer requires a network of mains, pipes, and, in some cases, service reservoirs for short-term storage. The size and complexity of water distribution systems vary dramatically but it is an important requirement that the distribution system has to be closed to prevent possible recontamination. Maintenance of distribution systems to prevent water quality deterioration by regular cleaning is recommended.

### Monitoring

Drinking water quality monitoring involves a wide range of water quality assessments encompassing the entire water supply system. Careful consideration should be given to the water quality characteristics to be analyzed, including sampling location and frequency, analytical method, recording, evaluation, and reporting, with the emphasis on putting more effort into understanding the entire water supply system. Monitoring can be separated into two categories, operational monitoring and performance monitoring. Effective operational monitoring is critical for confirming that individual barriers for controlling hazards are functioning properly and effectively. Data from operational monitoring are used as triggers for immediate short-term preventive and corrective actions



to operational processes to maintain drinking water quality. Performance monitoring includes regular sampling and testing to demonstrate conformance with guideline values and other regulatory requirements.

## Water Analysis

The main objective of analysis is to ensure the water supplied to the public meets the relevant standards and does not exceed the recommended concentration of hazardous chemicals. The analyses performed by the water laboratory for compliance purposes should be performed in an accredited laboratory and comply with the recognized standard for technical competence of testing laboratories. A complete analytical procedure should include information on sample handling (collection, transport, and storage), sample preparation (concentrate and separate), analysis (methods to identify and quantify components), analytical quality control (criteria), and reporting of analytical results.

### Sampling Locations

Location and frequency of sampling depends on variability, and also the purpose, such as aesthetic or health significance. Locations can include the raw water supplied to the treatment plant, the process streams during water treatment processes, the product water leaving the plant, and the various sampling points in the distribution system. The sampling points in the distribution system should include consumer taps to ensure no deterioration in water quality during transportation in the pipelines. Sample frequency should be high enough to enable the monitoring to provide meaningful information. For characteristics where the concentration does not change greatly within the distribution system, sampling the water at the treatment plant may be sufficient. However, for characteristics that vary in concentration during distribution, sampling should be undertaken throughout the distribution system including the point of supply to the consumer. It should be noted that the behavior of some parameters (e.g., disinfection by-products, residual chlorine, microbial quality) during distribution may vary from one system to another and is likely to require system-specific investigation. A list of sampling locations of the concerning chemicals is presented in **Tables 1** and **2**.

### Types of Sample

Three types of sample, grab, composite, and integrated, are primarily used in drinking water analysis. Grab samples are a single sample collected at a specific location at a specific time. They are often

referred to as a 'snapshot' of a location and only represent the composition of its source at the time and place of collection.

Composite samples are obtained by combining portions of multiple grab samples or by using specially designed automatic sampling devices. This type of sample can reduce sample number and the samples are more representative of heterogeneous matrices. However, the composite samples may lose the analyte relationships in individual samples and increase the potential for analytical interferences.

Integrated samples can be viewed as a different form of composite sample. Instead of combining the grab samples at different sampling times, integrated samples combine grab samples collected at different locations at the same time. It is commonly used to evaluate average composition across the width or depth of a river or reservoir.

### Sample Collection Methods

Samples are generally collected by one of the three distinct sampling systems: a manual sampling system, a system involving sampling lines, or automatic sampling systems. In each it is important that the sample taken for analysis is representative of the whole body of water being sampled, and that the sample is not altered by the sampling system.

Manual sampling systems involve minimal equipment but they are costly and time consuming for routine monitoring. They may simply allow a large (bulk) sample to be obtained, which is then divided into one or more sample containers appropriate to the subsequent analyses. Sometimes samples are required from known depths, for example, in a raw water reservoir, when a depth sampler will be needed. These devices can be lowered to the known depth and then opened and closed before being recovered to the surface. This type of sampling is usually done by sample collecting personnel.

Systems using sampling lines are commonly used in water treatment plants. Often the flow of samples within the line occurs as a result of the pressure differential between the bulk matter and the line outlet. Where this is not possible the sample must be pumped, sucked, or pressurized to the outlet. Each situation must be considered on its merits, so that the system does not affect determinands of interest, e.g., dissolved gases and volatile organics can be lost when using suction pumps and vacuum lines, or, if the sample is to be examined to identify particulate material (such as algae), peristaltic pumps may be inappropriate. In certain critical situations duplicate pumps and sample lines may be necessary, together with sample line backwash facilities, e.g., to provide

sample to online-supply intake-protection water-quality monitors. In these situations, buildup of biological 'slimes' must be avoided as biological activity may alter the chemical composition of the sample before it reaches the analyzer.

Automatic sampling systems usually consist of a sampling device incorporating a sampling line and a unit that automatically controls the collection of a series of samples and stores them either as discrete subsamples or bulked composite samples for subsequent collection. These units can usually be programmed as to the time over which an individual sample is collected, the number of samples to be taken in the period of sample collection, and the overall time of sample collection. These systems can eliminate human errors in manual sampling and can reduce labor costs. Automatic samplers may thus be used to provide samples which will give information on water quality variation with time, to enable sampling at inaccessible and unmanned sites and to allow composite samples to be prepared. Some more sophisticated automatic samplers can accept feedback signals from flow-measuring equipment so the flow-related composite samples can be taken. In all the situations mentioned above it is most important that the automatic sampler does not contaminate the sample and all component parts of the sampling device that come into contact with the sample are inert or are known not to alter the concentration of the determinand(s) of interest. For example, plastic components may be incompatible with certain compounds that react with the plastic parts or that can be contaminated by contact with them.

In any situation involving sampling lines where suspended material may be present, the flow rate in the line should be sufficient to stop it settling out. It is also necessary to allow sample lines to run to waste for sufficient time to change the total volume contained within the line before collecting the sample for analysis.

### Sample Storage and Preservation

Ideally, the collected samples should be returned to a laboratory immediately (within a few hours) for analysis. However, this may not always be possible. The sample is a dynamic system whose composition may change with time and some determinations are more affected by sample storage than others. However, there is no single method of preservation that is entirely satisfactory. Selection of preservation method is based on the determination to be made.

The choice of sample container material may be an important factor for maintaining sample stability and

it is one of the most important components of the analysis. Contamination may be positive with analyte being added to the sample by the material from, or passing through, the container. Sample containers must be free of analytes of interest, especially for very low analyte levels. Glass (both soda and borosilicate), polyethylene (both high and low density), and polypropylene are commonly used materials for sampling containers. More recently, the use of polycarbonate, poly(vinyl chloride) (PVC), poly(ethylene terephthalate) (PET), and polytetrafluoroethylene (PTFE) sample containers have been reported. Only the correct type should be used, as one material may be preferred over the other, e.g., silica, sodium, and boron may be leached from soft glass but not plastic. Absorption is also possible, involving unexpected losses due to sorption onto or reaction with the container or by passage through the container, e.g., trace level of pesticides and metals may sorb onto the walls of glass containers. Thus, hard glass or plastic containers are preferred. The protocol of rinsing the bottle three times with the sample before sampling should be adopted.

Some metals are subject to loss by adsorption, or ion exchange with, the walls of glass containers. These determinands are best collected in separate clean bottles and acidified with nitric acid to a pH below 2. For most metal contaminants listed in **Table 1**, plastic or borosilicate glass sampling bottles can be used with the addition of  $\text{HNO}_3$  to pH <2 for sample preservation. Only refrigeration is recommended as sample preservation method for chromium and quartz or a PTFE sampling bottle for boron with addition of  $\text{HNO}_3$  to pH <2 for sample preservation.

Sample preservation for chloride and fluoride is not required. Fluoride should only be collected using plastic sampling bottles. Sample preservation of cyanide requires the addition of NaOH to pH >12 and refrigeration in the dark. Hydrogen sulfide requires the sampling bottle to be completely filled (no air gap) with minimum aeration. No sample preservation method is recommended for nitrate and nitrite, although it is generally required for the analysis to be carried out as soon as possible (within 1 day). Refrigeration for sulfate is essential.

For taste and odor analysis, only glass bottles can be used and analysis should occur as soon as possible. No sample preservation method is required for pH, although immediate analysis is recommended. Refrigeration is recommended as a sample preservation method for color measurement.

Disinfection agents, such as chlorine and chloramine, should be analyzed immediately. Addition of a quenching agent is required for disinfection by-products analysis. For inorganic disinfection

by-products, such as bromate and chlorite, the addition of 50 mg l<sup>-1</sup> ethylene diamine is recommended. For organic disinfection by-products, such as trihalomethane and haloacetic acid, the addition of sodium sulfite or ascorbic acid and ammonium chloride is recommended, respectively. It is more practical to predispense the quenching agent into the sampling bottle before sample collection to improve sample stability during transportation. In such instances, the sample should simply be filled to the appropriate level and a rinsing protocol should not be used.

Sampling for organic analysis is a rather complex issue as different analytes require different sampling protocols that are specific on sampling container type, sample preservation method, etc. Generally, the plastic sampling bottles are the preferred sampling containers and are required to be completely filled (no air gap), and refrigeration is highly recommended for sample preservation.

In general, when sampling for radioactive elements, concentrated hydrochloric or nitric acid is added at the time of collection to obtain a pH <2 and the sample held at this pH for at least 16 h before analysis. Microbiological samples should be collected using nonreactive borosilicate glass or plastic bottles that have been cleaned, rinsed, and sterilized. It is also very common for some analyses with the addition of a reducing agent, such as sodium thiosulfate, to dechlorinate the sample that prevents the continuation of bactericidal action during sample transit.

As a rule of thumb, during sample transportation, the preferred preservation procedure is to keep the sample cool at ~4°C. If analysis is to be delayed in the laboratory, sample refrigeration is required. The sample preservation methods described above are only a small number of examples to illustrate sample preservation methods with respect to the analyte of interest, for specific details of individual sampling methods consult APHA (1998).

### Sample Pretreatment

There is a common requirement in water analysis to separate the dissolved or suspended component in the sample before analysis. This is usually achieved by a physical separation step. Filtration is the most commonly used sample pretreatment method. Membrane and glass fiber filters are available in a variety of pore sizes and membranes are available in different materials. Water analysts normally use GF/C glass fiber filters for suspended solids determinations and 0.45 µm cellulose-based filters to define the soluble/insoluble fraction in samples. However, all filters and filtration apparatus must be checked to confirm

that they do not introduce contamination during the separation procedure. In trace metal analysis, it may be necessary to use PTFE-coated apparatus.

Four types of analyses can be performed in metal determinations, dissolved, suspended, total, and acid extractable metals, based on different sample pretreatment steps. The term 'total metals' includes all inorganically and organically bound metals and requires a vigorous digestion step of the unfiltered sample before analysis. 'Dissolved metals' determination requires filtration of unacidified sample that passes through a 0.45 µm membrane filter. 'Suspended metals' refer to the metals that are retained by a 0.45 µm filter. 'Acid-extractable metals' are determined after the unfiltered sample is treated with hot dilute mineral acid before analysis.

There are also sample concentration/separation techniques used as pretreatment methods for trace analysis. Commonly used methods include chelating ion-exchange and, for trace organic analysis, solvent extraction, carbon adsorption, and resin adsorption using nonionic macroreticular resins.

Minor variations in microbiological analyses can cause significant changes in results. Extra care is required for sample handling. General laboratory apparatus such as glassware, transfer pipettes, vials, and filtration equipment including disposable syringe filters are generally required to be sterilized before use.

### Analytical Techniques

Inorganic analysis, particularly metal analyses, generally falls into the category of spectrometry, which includes atomic absorption spectrometry (AAS) with flame or electrothermal atomization or the more advanced inductively coupled plasma with atomic emission spectrometry (ICP-AES) or mass spectrometry (ICP-MS). Other techniques such as colorimetry, ion-selective electrode (ISE), and ion chromatography (IC) can be used for anion analysis.

Analytical techniques such as gas chromatography (GC) and liquid chromatography (LC) are used for organic analysis. With the use of sample preconcentration steps such as liquid-solid, liquid-liquid extraction, solid-phase extraction (SPE), etc., or the powerful 'coupled techniques', such as GC-MS, a concentration range down to nanograms per liter can be determined.

A list of USEPA-recommended analytical methods corresponding to the contaminants stated in the WHO guidelines and USEPA standards is presented in **Tables 1 and 2**.

However, there are situations where the guideline value has been chosen for the purpose of determining

the presence of the chemical in the water. The guideline value for some chemicals, such as pesticides, has been set at the detection limit of determination, particularly pesticides that are not authorized to be used in certain areas. It is recognized that there are many experimental uncertainties in measuring extremely low concentrations. The limit of determination is often set as low as five times the detection limit, and is usually verified by interlaboratory tests.

By following good laboratory practice, such as including minimum required quality control for each analysis, the overall accuracy and precision of the test results can be assured. A good analytical quality control program should consist of at least some of the following elements: initial and ongoing demonstration of ability, method detection limit determination, reagent blank, blank spike, matrix spike, duplicate sample, internal standard, calibration, control charts, and quality control acceptance criteria.

*See also:* **Atomic Absorption Spectrometry:** Principles and Instrumentation; Flame; Electrothermal; Vapor Generation. **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Gas Chromatography:** Environmental

Applications. **Liquid Chromatography:** Overview. **Water Analysis:** Organic Compounds; Biochemical Oxygen Demand; Microbiological.

## Further Reading

- APHA (1998) *Standard Methods for the Examination of Water and Wastewater*, 20th edn. Washington, DC: American Public Health Association.
- NHMRC/ARMCANZ (1996) *Australian Drinking Water Guidelines*. National Health and Medical Research Council, Agricultural and Resource Management Council of Australia and New Zealand.
- NHMRC/ARMCANZ (2001) *Framework for Management of Drinking Water Quality – A Preventive Strategy from Catchment to Consumer*. National Health and Medical Research Council, Agricultural and Resource Management Council of Australia and New Zealand.
- USEPA (1995) *National Primary Drinking Water Standards*. Office of Water, Washington, DC: US Environmental Protection Agency, EPA 811-F-95-002-T, EPA 811-F-95-003-T, EPA 811-F-95-004-T.
- USEPA (1999) *Methods and Guidance for Analysis of Water*, Version 2. Office of Water, Washington, DC: US Environmental Protection Agency, EPA-821-C-99-004.
- WHO (1993) *Guidelines for Drinking-Water Quality*, 2nd edn. Geneva: World Health Organization.

## Freshwater

**G Hanrahan**, California State University, Los Angeles, CA, USA

**H Casey**, Institute of Freshwater Ecology, Wareham, UK

**P J Worsfold**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Freshwater systems are highly dynamic and involve complex interactions between a variety of physical, chemical, and biological processes, e.g., water column–particle interactions, redox reactions, and bacterial activity. It is important to have an understanding of these processes for effective watershed management, which includes water quality issues. In this regard, high-quality analytical data, i.e., accurate and precise, are a prerequisite for predictive purposes, establishing water quality guidelines and monitoring compliance with legislation. Requirements vary from country to country but are primarily focused on three broad classes of indicators: physicochemical

(e.g., pH, temperature, turbidity, hardness), chemical (e.g., contaminants such as pesticides and heavy metals and nutrient species such as nitrate and phosphate), and biological (e.g., community structure, abundance, and biomass). Ideally, the chemical composition of the water being analyzed should be measured *in situ* on a frequent (or even continuous) basis. When this is not possible a comprehensive sampling/analysis program with rigorous quality assurance, as shown in **Figure 1**, should be adopted. Minimizing sample contamination and/or degradation during sampling and storage is a key aspect of the program.

The chemical quality of surface water in rivers and streams, lakes, ponds, and wetlands is determined by interactions of the water body with, for example, soil/sediment, suspended solids, rocks, groundwater, and the atmosphere and by in-water biological, chemical, and physical processes. It can also be significantly perturbed by agricultural and industrial inputs that are often classified in terms of point sources (e.g., discharge of effluent from a factory or sewage treatment plant outfall pipe) and diffuse sources (e.g., runoff from agricultural land). There



the presence of the chemical in the water. The guideline value for some chemicals, such as pesticides, has been set at the detection limit of determination, particularly pesticides that are not authorized to be used in certain areas. It is recognized that there are many experimental uncertainties in measuring extremely low concentrations. The limit of determination is often set as low as five times the detection limit, and is usually verified by interlaboratory tests.

By following good laboratory practice, such as including minimum required quality control for each analysis, the overall accuracy and precision of the test results can be assured. A good analytical quality control program should consist of at least some of the following elements: initial and ongoing demonstration of ability, method detection limit determination, reagent blank, blank spike, matrix spike, duplicate sample, internal standard, calibration, control charts, and quality control acceptance criteria.

*See also:* **Atomic Absorption Spectrometry:** Principles and Instrumentation; Flame; Electrothermal; Vapor Generation. **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Gas Chromatography:** Environmental

Applications. **Liquid Chromatography:** Overview. **Water Analysis:** Organic Compounds; Biochemical Oxygen Demand; Microbiological.

## Further Reading

- APHA (1998) *Standard Methods for the Examination of Water and Wastewater*, 20th edn. Washington, DC: American Public Health Association.
- NHMRC/ARMCANZ (1996) *Australian Drinking Water Guidelines*. National Health and Medical Research Council, Agricultural and Resource Management Council of Australia and New Zealand.
- NHMRC/ARMCANZ (2001) *Framework for Management of Drinking Water Quality – A Preventive Strategy from Catchment to Consumer*. National Health and Medical Research Council, Agricultural and Resource Management Council of Australia and New Zealand.
- USEPA (1995) *National Primary Drinking Water Standards*. Office of Water, Washington, DC: US Environmental Protection Agency, EPA 811-F-95-002-T, EPA 811-F-95-003-T, EPA 811-F-95-004-T.
- USEPA (1999) *Methods and Guidance for Analysis of Water*, Version 2. Office of Water, Washington, DC: US Environmental Protection Agency, EPA-821-C-99-004.
- WHO (1993) *Guidelines for Drinking-Water Quality*, 2nd edn. Geneva: World Health Organization.

## Freshwater

**G Hanrahan**, California State University, Los Angeles, CA, USA

**H Casey**, Institute of Freshwater Ecology, Wareham, UK

**P J Worsfold**, University of Plymouth, Plymouth, UK

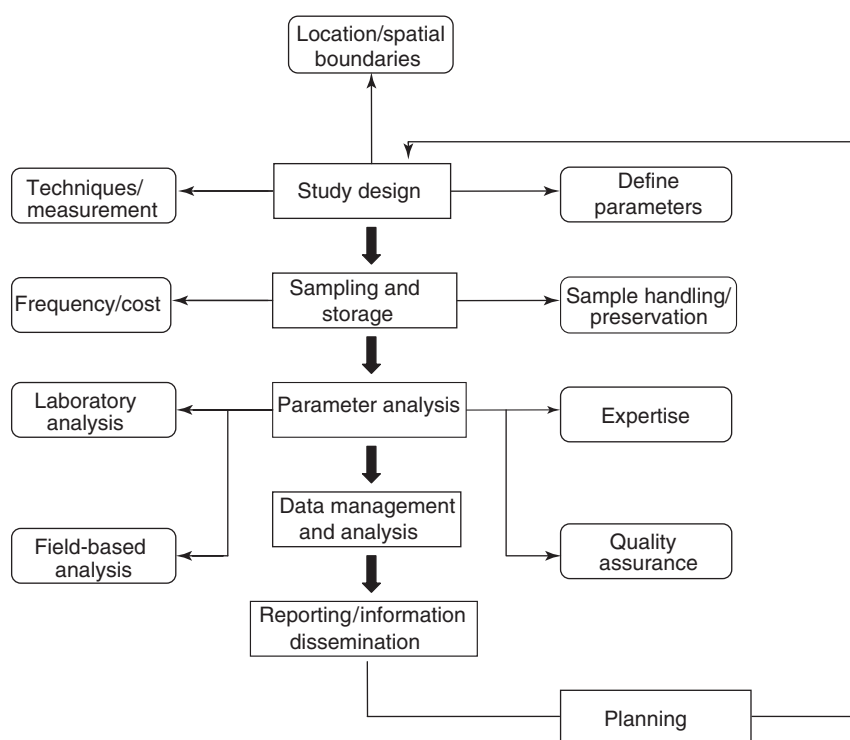
© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Freshwater systems are highly dynamic and involve complex interactions between a variety of physical, chemical, and biological processes, e.g., water column–particle interactions, redox reactions, and bacterial activity. It is important to have an understanding of these processes for effective watershed management, which includes water quality issues. In this regard, high-quality analytical data, i.e., accurate and precise, are a prerequisite for predictive purposes, establishing water quality guidelines and monitoring compliance with legislation. Requirements vary from country to country but are primarily focused on three broad classes of indicators: physicochemical

(e.g., pH, temperature, turbidity, hardness), chemical (e.g., contaminants such as pesticides and heavy metals and nutrient species such as nitrate and phosphate), and biological (e.g., community structure, abundance, and biomass). Ideally, the chemical composition of the water being analyzed should be measured *in situ* on a frequent (or even continuous) basis. When this is not possible a comprehensive sampling/analysis program with rigorous quality assurance, as shown in **Figure 1**, should be adopted. Minimizing sample contamination and/or degradation during sampling and storage is a key aspect of the program.

The chemical quality of surface water in rivers and streams, lakes, ponds, and wetlands is determined by interactions of the water body with, for example, soil/sediment, suspended solids, rocks, groundwater, and the atmosphere and by in-water biological, chemical, and physical processes. It can also be significantly perturbed by agricultural and industrial inputs that are often classified in terms of point sources (e.g., discharge of effluent from a factory or sewage treatment plant outfall pipe) and diffuse sources (e.g., runoff from agricultural land). There



**Figure 1** Schematic of a well-organized sampling/analysis protocol for freshwaters.

can be a strong seasonal trend with some chemical parameters (e.g., nutrients) and major short-term fluctuations due to, for example, rainfall events and pollution incidents. Whilst legislation usually requires concentrations of chemical parameters to be monitored, management issues are often better addressed by determining seasonal or annual fluxes (or loadings), which is also a function of the flow regime for rivers. Determination of the chemical variables to be measured, sampling locations, and frequency of sampling depends on the objectives and economics of the monitoring program. The choice of analytical method will also be a function of detection limit, precision, accuracy, speed, and cost.

## Sampling Protocol

A fundamental requirement of any sampling protocol is that the sample is representative of the body of water from which it originates. Therefore, a well-organized protocol should retain, as closely as possible, the original composition of the freshwater body in the collected sample. Procedures should be kept as simple as possible while minimizing the possibility of contamination or interferences from foreign substances. It is important to collect samples from the water column at a series of depths and cross-sectional locations as manual grab samples or through the use of automatic samplers that can be

programmed to collect samples at specific time intervals or flow-dependent intervals. Commercially available sondes (*in situ* probes) can be deployed in parallel with samplers to log parameters such as pH, dissolved oxygen, temperature, and turbidity and can transmit these data from remote sites via telemetry systems mounted above the water level.

It is also essential to monitor flow rates in rivers so that concentration data can be converted to flux data. This is typically done automatically every 15 min at gauging stations and again these data can be transmitted back to a central laboratory by telemetry. In addition, it is important to avoid boundary areas, for example, confluence of streams or rivers and below sewage treatment works (STWs) unless their impact on the water body is being investigated. Point source contributions from STWs, for example, will affect the overall water quality of freshwater systems, particularly the input of nutrient elements in the form of, for example, phosphate, ammonia, and nitrate.

In lakes and reservoirs, representative sampling is often difficult due to environmental heterogeneity, both spatial and temporal, due to, for example, thermal stratification and ecosystem dynamics. For example, during summer, deep lakes can be subdivided into a relatively uniform, warm, circulating upper region (epilimnion), a deep, cold, relatively undisturbed region (hypolimnion), and a region of



rapid temperature change between the two (thermocline). In addition, localized physical conditions (e.g., isolated bays and inflowing rivers or streams) must be considered to ensure that sampling objectives are not compromised.

Sampling depth is a key parameter when analyzing groundwater and special borehole sampling devices are required, with modified samplers for unstable determinands. There are various pumping mechanisms in use such as suction lift, positive displacement, and those that use inert gas to drive the groundwater to the surface. When considering treated water sources, for example, household taps, care must be taken that the water has not been standing in the pipes for a long period of time. A good flush of the pipes should be given, unless the object of the program is to measure contamination from the pipes, for example, leaching of lead.

Sampling of interstitial waters is particularly challenging due to the potential for the sampling process to perturb the system by, for example, sorption and/or contamination from the sampler and changes in redox state and pH. Five commonly used methods are: centrifugation, centrifugal drainage, basal cup, vacuum filtration, and dialysis.

### Site Selection

The selection of the number and location of sampling sites is crucial in any sampling protocol and will ultimately depend on proper authoritative permission for access and the objectives of the sampling program. It is paramount that sites are safe and accessible for personnel and equipment. In addition, locations must be chosen so that the samples taken are representative of the body of water at the site.

### Safety

Safety is of vital importance as the possibility of bacteriological, virological, and zoological hazards exist in many freshwater locations. Weil's disease, caused by the *Leptospira bacterium*, is one example of an infectious disease that is transmitted by infected animals into freshwater bodies, especially downstream from sewage works or runoff from landfill sites. It is thus recommended that gloves and barrier creams be worn when sampling these locations. Sampling alone, regardless of site location, is not recommended. Details of the sampling site(s) and approximate times visited should be recorded prior to departure so personnel back at the laboratory are aware of your location. In addition, reliable communication should be established between the laboratory and field location in case any emergencies or technical issues arise. Extra care should be taken

in extreme weather conditions such as heavy rain, storms, snow, and ice.

### Frequency and Cost

The frequency of sampling (e.g., one-off, hourly, daily, monthly, or occasional) will ultimately depend on the overall cost-effectiveness of the sampling protocol. In many freshwater bodies, the majority of the annual flux of contaminants (e.g., phosphorus) is transported during intense, short-term events (e.g., storms) when discharge is high. For example, phosphorus loadings in rivers and streams are generally higher during autumn and winter months when flows are highest, and lower in summer months when the flow is low and biological activity is high. Important in-water processes that affect phosphorus concentrations include plant and algal uptake, anthropogenic inputs (e.g., sewage effluent), matrix considerations (e.g., hardness), and resuspension of bottom sediments from increasing discharge.

Diurnal changes must also be taken into consideration. For example, the dissolved oxygen concentration in freshwaters is dependent on temperature (as well as depth, altitude, saline intrusion) and competing autotrophic (photosynthesis) and heterotrophic (respiration) processes, which leads to significant day/night variations. Thus, sampling at frequent intervals over 24 h periods is recommended when investigating these processes.

### Sample Treatment and Storage

Table 1 lists sample storage and treatment criteria for the determination of a number of parameters in freshwater environments. It is important to state that no single method of storage and preservation exists for all parameters, but the effectiveness of such methods depends on factors such as sample matrix, filtration technique, storage container, chemical addition (e.g., acidification), and physical treatment (e.g., refrigeration). When determining nutrients, for example, protocols should take account of the following general guidelines:

1. a rigorous cleaning procedure is essential to avoid contamination;
2. all sample containers should be made of a strong, inert material;
3. filtration (typically 0.45 or 0.2  $\mu\text{m}$ ) should be carried out at time of sampling with a low-pressure gradient (less than 1 atm) to avoid cell lysis; and
4. filtered samples should be analyzed as quickly as possible, ideally within 8 h to minimize nutrient loss.

**Table 1** Sample storage and treatment criteria for various parameters in freshwater samples

<i>Parameter/analyte</i>	<i>Storage container</i>	<i>Preservation method</i>	<i>Storage time</i>
Acidity and alkalinity	Polyethylene or glass	4°C	24 h
Ammonia	Polyethylene or glass	Acidify, filter	1 month
Arsenic	Polyethylene or glass	Acidify	1 month
BOD	Glass	4°C	ASAP, dark
Carbon, organic	Glass	Acidify with H <sub>2</sub> SO <sub>4</sub> , 4°C	ASAP, dark
Chlorides	Polyethylene or glass	4°C	1 month
Chlorophyll	Polyethylene or glass	4°C	24 h
COD	Glass	Acidify with H <sub>2</sub> SO <sub>4</sub> , 4°C	24 h, dark
Conductivity (onsite preferred)	Polyethylene or glass	4°C	24 h
Fluorides	Polyethylene	4°C	1 month
Iodides	Polyethylene or glass	4°C	24 h
Iron	Polytetrafluoroethylene	Acidify, filter	1 month
Lead	Polyethylene or borosilicate	Acidify, filter (do not use H <sub>2</sub> SO <sub>4</sub> )	1 month
Manganese	Polyethylene or borosilicate	Acidify, filter	1 month
Mercury	Polytetrafluoroethylene	Acidify, HNO <sub>3</sub> and K <sub>2</sub> Cr <sub>2</sub> O <sub>7</sub> addition	1 month
Nitrate (matrix dependent)	High-density polyethylene	Filter, 4°C	1–2 months
Phosphates, TP (matrix dependent)	High-density polyethylene	Filter (species dependent), 4°C	1–2 months
Pesticides	Glass (solvent washed)	4°C	24 h
Selenium	Glass or borosilicate	Acidify to pH 1	1 month
Silicates	Polyethylene	Acidify with H <sub>2</sub> SO <sub>4</sub> , filter, 4°C	24 h
Sulfates	Polyethylene or glass	4°C	1 week
Surfactants	Glass	Specific to type	–
Total residue	Polyethylene or glass	4°C	24 h

ASAP = as soon as possible.

Chemical addition (to prevent biological growth) and matrix considerations are also important and are discussed in greater detail below.

### Sample Containers

Sample containers should be made of high-quality materials such as quartz, borosilicate glass, polyethylene, polypropylene, high-density polyethylene (HDPE), or polytetrafluoroethylene (PTFE). Special care must be taken so that the container does not contaminate the sample or remove analytes by adsorption. Glass or borosilicate containers, for example, can increase the silica and sodium concentrations of samples and provide active sites for removal of some metal ions. Fluorides react with glass and pesticides and hydrocarbons adsorb onto polyethylene containers. As shown in **Table 1**, polyethylene and HDPE are recommended for nutrients and trace metals. Glass is recommended for generic parameters such as biological oxygen demand (BOD) and chemical oxygen demand (COD) and for organic compounds such as pesticides.

### Cleaning Protocols

The walls of sample containers are excellent substrates for bacteria, often enhancing bacterial growth, and therefore rigorous cleaning of all laboratory-ware is necessary. For nutrient (e.g., nitrate and phosphate) determinations, containers should be cleaned overnight with a nutrient-free detergent,

rinsed with ultrapure water, soaked in 10% HCl overnight, and then rinsed with ultrapure water. For trace metals, plastic containers should be acid washed with a 1 M solution of nitric acid overnight then rinsed with ultrapure water. For organic compounds, including pesticides and herbicides, chromic acid washing of glass bottles is recommended, followed by ultrapure water and final rinsing with an appropriate organic solvent (e.g., hexane). In all cases, containers should be rinsed at least twice with the water of interest prior to analysis with blank measurements taken to ensure that contamination of samples is not occurring.

### Filtration

The process of sample filtration is often necessary in freshwater analysis, especially when particulate and dissolved fractions are measured separately. Generally, filtration is done using a 0.45 or a 0.2 µm filter, with the fraction passing through it operationally defined as soluble (or dissolved) and matter (both biotic and abiotic) collected on the filter defined as suspended. Filtering removes suspended particulate matter that can cause measurement interferences, for example, scattering of light in spectrophotometry. Filtering through a 0.2 µm filter removes the majority of bacteria and plankton that could otherwise alter analyte concentrations (especially nutrients) during storage.

As with sample containers, the filtration apparatus (including individual filters) must be cleaned prior to

use using similar acid wash/ultrapure water rinse procedures. The filtration procedure can be carried out under positive pressure or vacuum. However, excess pressure gradients should be avoided as rupture of algal cells and the subsequent release of intracellular contents into the sample could occur. Filters vary according to category (depth or screen) structure, pore size, and composition with the correct filter of choice ultimately based on the parameter of interest. For depth filters, the effective pore size changes as the filter becomes more loaded with particles, whereas screen filters are not affected by particle loading. The most widely used filters for water analysis are cellulose-based, glass fiber, polycarbonate, poly(vinyl chloride), and PTFE.

### Storage and Preservation

If immediate analysis is not possible, freshwater samples must be preserved and stored to maintain the original concentration of the analyte within the level of accuracy required by the particular program. A considerable body of literature presents varied opinions on the effectiveness of different preservation methods. However, cooling the sample (i.e., refrigeration and freezing) and acidifying are the most widely used and accepted methods of sample preservation.

For trace metals, sample acidification to a pH of below 2 is the most widely used form of sample preservation. This procedure not only helps eliminate biological growth, but also limits adsorption onto the container walls. For nutrients, it is difficult to select one reliable treatment due to specific matrix characteristics (e.g., hardness, salinity, dissolved organic matter, and bacterial nutrient uptake) of the sampling location. In chalk catchments where calcium concentrations are high, freezing samples is not recommended due to the possibility of inorganic phosphate being co-precipitated with calcite when thawing the samples. Storage at 4°C is highly recommended, often with the addition of chemical additives such as chloroform to prevent biological growth. However, chloroform should not be used in samples with a high organic matter content, as the release of cellular enzymes into the samples is possible.

### Analytical Techniques

There are several alternative methods of analysis available for each parameter in freshwaters, with the method of choice ultimately depending on

1. the range of concentrations that need to be determined;
2. the accuracy and precision required;

3. selectivity required;
4. the maximum time between sampling and analysis; and
5. the cost relative to the analytical objectives.

In view of the above-mentioned challenges associated with sample collection and analysis, it is desirable to monitor each parameter *in situ*. The most important consideration is the accuracy and precision of the data obtained in the most cost-effective manner. Table 2 gives a general overview of common analytical techniques for important chemical parameters in freshwaters.

### Flow Injection Analysis

Flow injection (FI) analysis has become an established tool for sample presentation and online treatment (chemical and physical) in the laboratory environment. It is now being increasingly considered for deployment outside of the laboratory, in both process and environmental locations, for *in situ* analysis. The versatility and adaptability of FI allows it to be used in many different applications with a variety of detection modes.

FI systems have been successfully used with spectrophotometric detection, for example, for the determination of nitrogen and phosphorus species in rivers and streams. It has also been used to determine anionic surfactants (e.g., sodium lauryl sulfate, sodium dodecyl sulfonate), chloride, and organophosphoric insecticides in freshwater bodies. FI also can be coupled with other analytical techniques for sample introduction purposes. Such methods include FI coupled with atomic absorption spectrometry for online preconcentration and separation of chromium species and with mass spectrometry for the determination of polar organic pollutants.

### Atomic Spectrometry

Atomic absorption spectrometry is commonly used to measure a wide range of elements as shown in Table 2. Such techniques as flame, graphite furnace, hydride generation, and cold vapor are employed. Measurements are made separately for each element of interest in turn to achieve a complete analysis; these techniques are relatively slow to use. More sensitive, but also more expensive, multielement analytical techniques such as inductively coupled plasma-atomic emission spectrometry and inductively coupled plasma-mass spectrometry can be used if lower ( $\mu\text{g l}^{-1}$  and below) detection limits are required. These detectors can also be coupled with separation systems if speciation data, e.g., Cr(III) and Cr(VI), are needed.

**Table 2** Common techniques for the analysis of various freshwater determinands

<i>Parameter/analyte</i>	<i>Technique</i>
Alkalinity	Titrimetry
Antimony	Atomic spectrometry
Aluminum	Atomic spectrometry
Barium	Atomic spectrometry
BOD	Measurement of dissolved oxygen after incubation
Bromide	Capillary electrophoresis, flow injection, ion chromatography
Cadmium	Anodic stripping voltammetry, atomic spectrometry
Calcium	Atomic spectrometry
Carbon dioxide	Titrimetry, by calculation (from pH and alkalinity measurements)
Carbon, organic	High-temperature combustion with IR detection, wet digestion with spectrophotometric detection
Chloride	Flow injection, ion chromatography, titrimetry
Chlorinated phenoxyacid herbicides	Gas chromatography with, e.g., MS detection, liquid chromatography
Chlorine	Titrimetry
Chromium	Atomic spectrometry, ion chromatography
COD	Titrimetry
Copper	Atomic spectrometry
Cyanide	Flow injection, ion-selective electrode, titrimetry, ion chromatography
Fluoride	Flow injection, ion chromatography, ion-selective electrode
Hardness, total	Titrimetry
Iodide	Atomic spectrometry, voltammetry, ion chromatography
Iron	Atomic spectrometry, flow injection
Lead	Atomic spectrometry, anodic stripping voltammetry
Magnesium	Atomic spectrometry, ion chromatography
Manganese	Atomic spectrometry
Mercury	Atomic spectrometry
Nickel	Atomic spectrometry
Nitrogen (various forms)	Flow injection, ion chromatography, ion-selective electrode, titrimetry
Organochlorine insecticides	Gas chromatography with, e.g., MS detection
Organophosphorus pesticides	Gas chromatography with, e.g., MS detection
Polynuclear aromatic hydrocarbons	Gas chromatography, liquid chromatography
pH	Ion-selective electrode
Phosphorus (various forms)	Flow injection, ion chromatography
Potassium	Atomic spectrometry, ion chromatography
Selenium	Atomic spectrometry
Silicon	Flow injection analysis
Silver	Atomic spectrometry
Sodium	Atomic spectrometry, ion chromatography
Sulfate	Capillary electrophoresis, flow injection, ion chromatography
Turbidity	Nephelometry, potentiometry
Vanadium	Atomic spectrometry
Zinc	Atomic spectrometry

### Separation Techniques

Chromatography coupled with a suitable detector is the preferred approach for the determination of specific organic compounds in freshwaters. The most common techniques are liquid chromatography (LC) and gas chromatography–mass spectrometry, but LC coupled with mass spectrometry and multi-stage separations are becoming increasingly popular. Ion chromatography is extensively used for the determination of a wide range of cations and anions. Solid-phase techniques using chromatographic stationary phases, ion-exchange resins, and chelating ligands are also used for preconcentration of analytes from freshwaters prior to quantitative detection.

### Ion-Selective Electrodes

Ion-selective electrodes (ISEs) are potentiometric sensors that include a selective membrane to minimize matrix interferences. The most common ISE is the pH electrode, which contains a thin glass membrane that responds to the  $H^+$  concentration in a solution. Other parameters that can be measured include fluoride, bromide, nitrate, and cadmium, and gases in solution such as ammonia, carbon dioxide, nitrogen oxide, and oxygen. ISEs do have their limitations including lack of selectivity and sensitivity and problems connected with conditioning of electrodes. Detection limits for nitrate-N, for example, are typically  $\sim 0.098 \text{ mg l}^{-1}$  for commercial field devices and have chloride as a major interferent.



## Quality Assurance

A wide range of analytical techniques are used to determine freshwater parameters (see Table 2). It is therefore essential that quality assurance protocols are in place to ensure that consistently high-quality data (i.e., accurate and precise) are obtained. Specific techniques used to ensure the quality of measurements in terms of accuracy and precision are given below.

### Controls and Blanks

In-house quality control samples provide a day-to-day check of variability due to factors such as instrumental effects, different operators, blanks, and environmental effects. Laboratories often spike samples with a known amount of analyte, then estimate the recovery, and hence infer the accuracy of the method used. It is assumed that any interference will affect the spike in a similar manner as the unspiked sample.

Method and instrument blanks and, where possible, field blanks should be analyzed with each batch of samples. A method blank uses water, usually high-purity (double-distilled and deionized) water, which is processed through all laboratory steps in the same way as samples. An instrument blank uses the same water directly introduced into the detector. A field blank is high-purity water that has been bottled in the laboratory, shipped with sample bottles to the sampling site, processed, and preserved as a routine sample and returned with the routine samples to the laboratory for analysis. The analysis of a blank should not yield a value higher than that allowed by the acceptance criteria. Blanks are used to determine the limit of detection of a method and to monitor all aspects of the analytical process.

### Certified Reference Materials

Certified reference materials (CRMs) are materials of known analyte concentrations with a similar matrix to the sample being analyzed that are certified by technically valid procedures accompanied by or traceable to a certificate or other documentation issued by a certifying body. CRMs play important roles in helping produce reliable results including calibration and verification of measurement processes, quality control, verification of standardized methods, and the development and validation of new methods.

### Interlaboratory Comparisons

Individual laboratories can compare their results with consensus values generated by several laboratories after analyzing one or more homogeneous and stable materials under designated conditions. Interlaboratory comparisons between expert laboratories

play a major part in the production of CRMs. They can be used to compare the precision and accuracy of results between laboratories, provide impartial evaluation of in-house quality control procedures, identify best practice, and support training needs and provide valuable databases of analytical information.

## Data Management and Reporting

In view of the large amount of data that can be collected when analyzing freshwater samples, it is necessary to properly archive it in order to effectively interpret and report the results. With current computer technologies, databases, and online access, data management has become an easier and more reliable process.

All freshwater monitoring programs must clearly identify the end users' needs, with the reporting system efficiently and accurately transmitting the analyzed data. Time frames to submit the data should be established in advance and allow for any unforeseen complications. In addition, defining the type of transmission or presentation appropriate for each end user is paramount.

*See also:* **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Chemiluminescence:** Overview. **Chromatography:** Overview. **Flow Injection Analysis:** Principles; Instrumentation. **Ion-Selective Electrodes:** Overview. **Quality Assurance:** Quality Control; Reference Materials. **Sample Handling:** Sample Preservation. **Sampling:** Theory. **Water Analysis:** Overview; Organic Compounds. **Water Determination.**

## Further Reading

- APHA/AWWA/WEF (1998) *Standard Methods for the Examination of Water and Wastewater*. Washington, DC: American Public Health Association.
- Bartram J and Balance R (1996) *Water Quality Monitoring: A Practical Guide to the Design and Implementation of Freshwater Quality Studies and Monitoring Programs*. United Nations Environment Programme/World Health Organization. London: E & FN Spon.
- Gardolinski PCFC, Hanrahan G, Achterberg EP, *et al.* (2001) Comparison of sample storage protocols for the determination of nutrients in natural waters. *Water Research* 35: 3670–3678.
- Hewitt CN (1992) *Methods of Environmental Data Analysis*. London: Elsevier Applied Science.
- Hunt DE and Wilson AL (1986) *The Chemical Analysis of Water*. London: Royal Society of Chemistry.
- ISO Guide (1981) *Terms and Definitions and Definitions Used in Connection with Reference Materials*, No. 30-1981. Geneva, Switzerland: ISO.
- Maher WA and Batley G (2002) Design of water quality monitoring programs. In: Burden FR, McKelvie I, Förstner U, and Guenther A (eds.) *Environmental Monitoring Handbook*, pp. 2.1–2.31. New York: McGraw-Hill.

## Seawater – Organic Compounds

M R Preston, University of Liverpool, Liverpool, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The analysis of dissolved organic compounds in seawater is a complex task that poses a considerable number of challenges for the analyst. This article reviews the major problems and describes techniques for overcoming them or circumventing them using recent technological developments. Examples are given of the techniques that have been applied to a number of specific compounds or groups of compounds.

### Basic Considerations

The qualitative and quantitative analysis of organic compounds dissolved in seawater is difficult for at least four main reasons. First, the size and depth of the oceans makes the primary step of accessing the required sample dependent not only on the availability of suitable ships and ancillary equipment but also on such environmental conditions as wind speed and sea state. The quest for *in situ* samplers/sensors or analytical devices has been the focus of considerable efforts in recent years and some of these developments are discussed later. A second problem is that once the sample has been obtained, the high ionic strength of seawater presents other analytical challenges such that many methods applicable to freshwater systems cannot be applied to seawater. Third, the concentrations at which organic compounds normally occur in marine waters are extremely low and frequently fall within the nanomolar or sub-nanomolar range. Finally, the number of individual organic compounds present is effectively infinite because of the presence of the so-called marine uncharacterized material (UCM, also known as the uncharacterized fraction, geopolymer, or Gelbstoff), which is the highly complex persistent material formed by the degradation of plant- and animal-derived material.

The main problems confronting the marine analytical chemist are those of selectivity and sensitivity. In practice, there are comparatively few single techniques that can adequately address both of these requirements. In general, most instrumental methods are applied after some form of preconcentration step and commonly rely on gas (GC) or liquid (LC) chromatography to provide adequate separation of individual components and mass spectrometry (MS) for

detection. MS is not as sensitive a detection method as, for example, fluorescence but its great virtue is that of selectivity, particularly in ion-selective monitoring mode. Only recently have MS systems been made robust enough to take to sea and the very new development of submersible MS devices is opening up new possibilities in this area.

In the following sections aspects of organic component analysis in seawater are considered from sample collection through bulk or individual class analysis to individual component analysis. For interested readers the most useful general manual for seawater analysis is probably that by Grasshoff *et al.* A summary of some analytical strategies for analysis of organic compounds in seawater is given in Table 1.

### Sampling

The procedures used for the collection of samples at sea may be divided into the two categories of discrete and *in situ* procedures. Whilst the former approach is considerably more common, there are issues as to whether representative samples can be collected in this way. Large sample volumes and the greatest possible number of individual samples are the preferred options though both criteria are normally overridden by practicalities.

#### Discrete Sample Collection

Collection of samples for organic component analysis has a number of special difficulties. The sampling platform (i.e., the ship) is a considerable source of organic material, notably hydrocarbons from fuel and lubricating oil residues and CFCs/halons from refrigeration and fire-fighting equipment. Considerable attention must therefore be paid to sample design such that the sampler can be deployed, retrieved, and the sample processed without exposure to external contaminants.

The sampler itself should be constructed from non-contaminating and, equally important, nonadsorptive material. The sampler should also be capable of being lowered through the sea surface closed so as to avoid contamination from the organic-rich sea surface microlayer. Unfortunately, there is no ideal sampler. The modified General Oceanics 'Go-Flo' bottles widely used by workers in the field of oceanic trace inorganic analysis are often the most suitable systems that are commonly available. In some cases (e.g., for halocarbons), special devices have been constructed to allow uncontaminated withdrawal for the sample



**Table 1** Selected analytical strategies for the analysis of various dissolved organic compounds in seawater

Analyte	Examples of methods employed
Alcohols	SPE microextraction, GS-MS
Amides	(1) Amide herbicides – preconcentration using SPMDs then bioassay. (2) Diethanol amides by solid-phase extraction–liquid chromatography–atmospheric pressure chemical ionization/electrospray ionization mass spectrometry (SPE–LC–APCI/ESI-MS)
Amino acids	$\alpha$ -Phthalaldehyde derivatization then reversed-phase LC with fluorescence detection. Chiral derivatization permits separation of optical isomers on standard LC columns
Amino sugars	Anion exchange LC
Aromatic monosulfonic acids	Ion-pair liquid chromatography
Carbohydrates	High-performance anion-exchange chromatography and pulsed amperometric detection (HPIC–PAD)
Dimethyl sulfide	Purge and trap or headspace followed by GC–FPD
Dioxins/dibenzofurans	Solid-phase microextraction, thermal desorption followed by GC–tandem MS
Ethers	SPE microextraction, GS-MS
Halocarbons	Purge and trap GC–ECD or GC–MS
Hydrocarbons	Preconcentration by (1) purge and trap (volatiles) and (2) solvent extraction. Analysis by GC–FID or GC–MS
Neutral sugars	LC (anion) with PAD
Nitriles	SPE microextraction, GS-MS
Nitrophenols	Liquid–liquid extraction using a hollow fiber membrane followed by HPLC
Nucleic acids	Preconcentration on a hydroxyapatite column, derivatization with ethidium bromide dye fluorescence detection
Phthalate esters	Solvent extraction or SPE followed by GC–FID
Phenols	SPE followed by LC with electrochemical detection
Proteins	Preconcentration using cross-flow ultrafiltration. Analysis with $^{13}\text{C}$ NMR pyrolysis GC–MS, LC–MS
Pyridines	SPE microextraction, GS-MS
Uronic acids	Freeze drying, isolation by cation exchange, spectrophotometric determination

from the sampling bottle itself. However, no sampler that is constructed from plastic can be considered entirely suitable for trace organic work. Various other materials have been used for trace organic sample collection including glass though the fragility of this material at sea raises its own problems.

### ***In Situ* or Online Sample Collection and Processing**

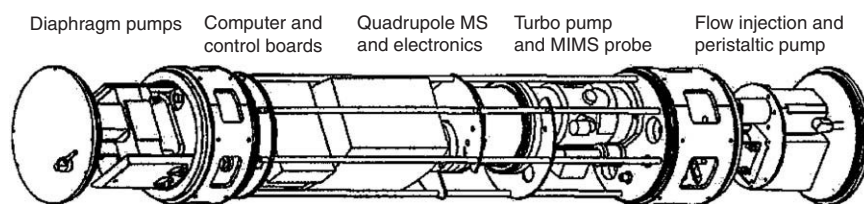
To avoid the problems attendant upon the collection of discrete samples, and also to allow much larger volumes of water to be sampled, a number of *in situ* systems have been devised. These generally consist of metered submersible pumping systems capable of drawing water first through a filter and then an adsorption column. The units can be deployed from ships or attached to remote moorings and often programmed to sample at particular times of day or tidal state. Such units are battery powered and are capable of processing several hundred or even thousands of liters of water. The great advantage of these systems is that the influence of external contamination sources is relatively low and the amount of material adsorbed onto the column is sufficient to reduce the analytical detection problems that would otherwise be encountered in the laboratory. The main disadvantages are cost, the time necessary for deployment, and the fact that each system can only produce a single data point. They are therefore

considerably limited in their ability to produce oceanographically useful data such as depth profiles or time series with the appropriate spatial or temporal resolution.

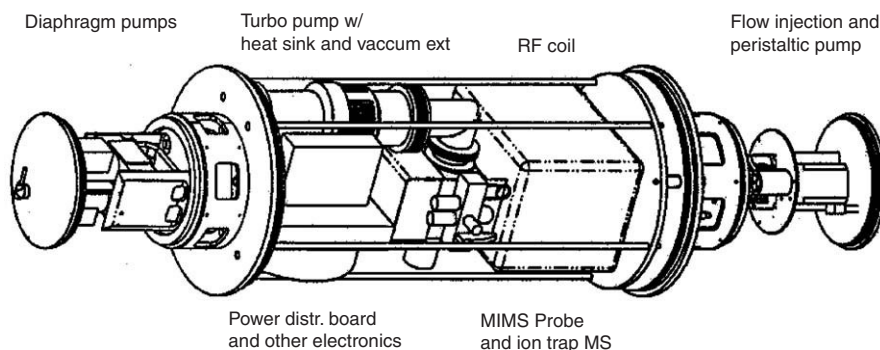
Recent advances in sample collection have included semipermeable membrane devices (SPMDs). These are cheap, passive samplers that rely on diffusion of compounds of interest through a membrane and into an absorbent such as triolein. Whilst only semiquantitative in most applications, they do provide a simulation of biological uptake in that the absorbent mimics the properties of living tissue.

Online collection/concentration of samples through the use of cross (tangential) flow ultrafiltration has also found a number of applications.

***In situ* analysis** The ultimate ambition of marine analysts for *in situ* analysis has now been realized with the advent of underwater MS units. These small, submersible units are presently restricted to relatively shallow depths (<300 m) and low mass ranges (<200 amu) but represent a major step forward in marine chemistry. Such units are available as quadrupole (Figure 1) or ion trap (Figure 2) configurations and can be used in moored or towed systems. Examples of the output from such systems are given in Figure 3. Introduction of samples is through membrane diffusion and the units have a deployment time of around a week at present.



**Figure 1** Schematic of the three-pressure-vessel system for the underwater membrane-introduction quadrupole mass filter system. (Reprinted with permission from Short RT, Fries DP, Kerr ML, Lembke CE, Toler SK, *et al.* (2001) Underwater mass spectrometers for *in situ* chemical analysis of the hydrosphere. *Journal of the American Society for Mass Spectrometry* 12: 676–682; © The American Society for Mass Spectrometry.)



**Figure 2** Schematic of the three-pressure-vessel system for the underwater membrane-introduction ion trap mass spectrometer system. The first and third vessels are identical to those used on the quadrupole mass filter system. (Reprinted with permission from Short RT, Fries DP, Kerr ML, Lembke CE, Toler SK, *et al.* (2001) Underwater mass spectrometers for *in situ* chemical analysis of the hydrosphere. *Journal of the American Society for Mass Spectrometry* 12: 676–682; © The American Society for Mass Spectrometry.)

## Storage

There are no universally satisfactory techniques for the storage of seawater samples prior to analysis for dissolved organic compounds. Acidification and/or refrigeration or freezing have all been used, as have various preservatives such as mercuric chloride. Because of the very low concentrations and (frequently) high biological availability of many dissolved organic compounds, there are considerable dangers of sample loss, degradation, or contamination. There is therefore no substitute for the rapid on-board or preferably *in situ* analysis of samples where possible.

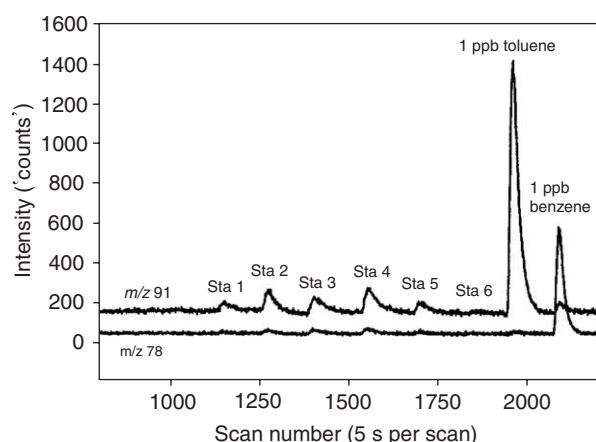
## Filtration

The separation of the organic material in marine water samples into dissolved and particulate phases is not straightforward. The distinction is made on operational rather than theoretical grounds in that organic compounds occur in a wide range of ‘particle sizes’ ranging from truly dissolved through macromolecular species, colloids, bacteria, viruses, and planktonic species and upwards. The choice of filter medium can have a significant impact on the compounds falling into either the ‘particulate’ or

‘dissolved’ categories and the concept of a ‘third’ or ‘colloidal phase’ (i.e., a solid phase that has passed through a filter) has sometimes been invoked to explain the results of adsorption studies.

The problems arising from the filtration process are as follows:

1. It is not normally clear precisely what range of particles sizes are being retained on a particular filter medium and how these may change as the load on the filter increases within a single sample.
2. Some filters (notably membrane ones) may be supplied loaded with wetting/stabilizing agents that are organic in nature and must be removed before the filter is used if contamination of the filtrate is to be avoided. Glass fiber filters must be pre-ashed (normally by heating at ~400–450°C for several hours) to remove organic contaminants but it is probable that this process may lead to the formation/exposure of active sites within the filter capable of adsorbing sample components (see (iii) below). Other filter materials such as polytetrafluoroethylene (PTFE), aluminum, or silver membrane are sometimes used albeit at significant extra expense.
3. The filters used are not inert and may actually retain dissolved organic material on active sites on



**Figure 3** Laboratory data from ion trap MS analysis of water samples that were obtained during towed-deployment of the quadrupole MS system. The  $m/z$  78 ion is diagnostic of benzene and the  $m/z$  91 ion is diagnostic of toluene. Analyses of samples are compared with 1 ppb standards. (Reprinted with permission from Short RT, Fries DP, Kerr ML, Lembke CE, Toler SK, *et al.* (2001) Underwater mass spectrometers for *in situ* chemical analysis of the hydrosphere. *Journal of the American Society for Mass Spectrometry* 12: 676–682; © The American Society for Mass Spectrometry.)

the fiber surfaces. This is particularly true of glass fiber filters though it is a rarely considered issue.

4. The filtration process itself is a mechanically aggressive one for, in particular, delicate unicellular organisms such as the phytoplankton that are abundant in seawater. If performed without due regard for the problems such cells may rupture spilling their contents into the filtrate. In this regard vacuum filtration is probably only satisfactory if performed under weak ( $<0.25$  bar) vacuum and pressure filtration is generally more satisfactory.

5. The low concentrations of both dissolved and particulate species in seawater often requires that large volumes of sample be collected and processed. Special, large reservoir (e.g., 10–20 l) pressure filtration systems (commonly constructed of 316 grade stainless steel) are therefore needed and these are very expensive.

Recent innovations in filtration technology include tangential flow ultrafiltration methods that act as preconcentration methods for studies of marine biopolymers such as proteins (see also the section on Adsorption techniques below).

## Preconcentration

The low concentrations of dissolved organic species in seawater make a preconcentration step an almost mandatory component of the analytical process. Only

in rare cases (e.g., class analysis of primary amines) is such a step not incorporated. Two main approaches have been widely adopted for dissolved species namely solvent extraction and adsorption. Others such as co-precipitation or stripping of volatile species have been used in appropriate cases.

Many preconcentration steps require a subsequent cleanup stage and frequently a fractionation stage as well. Such steps greatly increase the times required for analysis and also considerably enhance the amount of sample handling required and hence the risks of contamination. A recent review of the subject of concentration techniques for isolating polar pesticides in environmental water samples has been produced (see Further Reading section). Many of these techniques have applications for other components of interest.

## Solvent (Liquid–Liquid) Extraction

A fairly limited range of solvents has been utilized for the extraction of trace organic species from seawater. Hexane is a common choice though many others have also been used. Recent trends have been to replace hexane with either pentane or heptane because of the lower toxicity of these alternatives. Strict regulations now apply to safety aspects of the use of solvents at sea and these must not be neglected by the marine analyst when devising new methodologies.

There are considerable practical problems in manipulating the large volumes of water (and solvent) required for seawater analysis. Sample volumes of up to, or exceeding 2–3 l, are required and the difficulties in manipulating such volumes, especially at sea, are considerable. Some large-volume continuous liquid–liquid extractors have been devised which are capable of use at sea though the practical difficulties associated with the operation of any fragile systems on ships should not be minimized. Such units may be capable of handling 10–20 l extractions.

## Adsorption

Preconcentration procedures using adsorption are now generally the preferred options. Adsorption techniques have considerable practical advantages over those utilizing solvent extraction, particularly when working at sea, because they offer the possibility of treating relatively large volumes of water, they are physically robust, and the extracted material can be returned to the laboratory in solid form. Selective adsorption methods have not been widely used in seawater analysis though ion-exchange methods have been used for amino acids in the past and for amino sugars and organic ligands more recently. Most techniques in common use today use nonspecific

adsorption onto materials such as Amberlite XAD-2 resin, Tenax resin, activated charcoal, polyurethane foam, zeolite, or variations on LC packing material such as octa or octadecyl silanized silica.

Traditional adsorption preconcentration techniques are based around the passage of filtered (and sometimes acidified or otherwise pretreated) water through a column of adsorbent. After passage of the sample the compounds of interest are eluted with an appropriate solvent or solvent mixture. Such techniques can be laboratory based, *in situ* (pumped systems or, more rarely, in passive towed units), or, in more recent systems, can be used in online applications.

*In situ* samplers capable of deployment to considerable (although not full oceanic) depths are now available. However, as mentioned above, these are of limited oceanographic usefulness in applications other than those requiring time-averaged concentration data at a very limited number of sites.

One of the main problems with solid adsorbents is that of obtaining satisfactorily low and reproducible blanks. Even if adsorbents can be adequately cleaned in the laboratory, very elaborate precautions must be adopted to prevent recontamination (and in the case of some synthetic resins, degradation) before deployment. Further problems may be encountered with *in situ* systems where the extended deployment times leave the sample (and adsorbent) vulnerable to bacterial attack.

An additional complication with solid adsorbents arises from the considerable variability of breakthrough volumes between different dissolved components. The high 'background' level of marine UCM tends to saturate the adsorption sites on some types of column before adequate amounts of the desired compounds have been extracted. Even within the UCM fraction there is some evidence to show that the adsorbed material changes in character with increasing sample volumes as compounds in the initial part of the sample volume that are only weakly adsorbed onto the column are displaced by more strongly adsorbed compounds arriving later in the process. The tendency in this case is often to displace low relative molecular mass and polar compounds with higher relative molecular mass and more hydrophobic materials.

The development of extraction disks (e.g., Empore<sup>TM</sup>) cartridges and syringe barrels has proved to be a convenient and attractive tool for the enrichment of trace organic materials from natural waters. Such units may consist of membranes of PTFE fibrils in between which are incorporated particles of adsorbent materials such as C<sub>8</sub>, C<sub>18</sub>, ion exchange resins, or polymer-based materials. These disks are in some ways more convenient than columns and may be

used in online or offline modes. Online is best suited to LC-based separation and detection systems whereas offline is generally more convenient with GC-based systems. One major advantage of the disks is that a number of them may be used in series. This feature allows the relatively efficient retention of sufficient quantities of compounds with relatively low breakthrough volumes. Column-based systems are less readily arranged in series and suffer from high back-pressure and/or low flows together with high dead volume which tend to enhance sample mixing during elution. Applications using hollow fiber membrane microextraction have recently been reported and have been reported to compare favorably with US EPA Method 508.

## Extraction of Volatiles

In recent years, there has been a marked increase in interest in the determination of volatile species in seawater. This interest derives from a number of areas such as the use of tracers in ocean water mixing studies, the quantitation of air-sea gas exchange processes, and oil exploration or pollution work.

A variety of techniques have been utilized, nearly all of which are relatively standard. Such methods include stripping techniques with solid phase and/or cryogenic trapping and followed by thermal desorption onto GC systems or passive 'equilibrium' headspace methods also followed by GC or GC-MS analysis. The development of underwater MS devices with membrane introduction of samples and a mass range (so far) up to 200 amu offers great future potential for volatiles work.

Well over 100 compounds have been determined in seawater (and sediments) using a considerable variety of techniques (Table 2). Methods include n- and branched alkanes (up to about C<sub>20</sub>, pristane/phytane), alkenes and aromatic compounds (up to the disubstituted naphthalenes), halocarbons and chlorinated aromatic species, low relative molecular mass alcohols, organic sulfur compounds (notably dimethyl sulfide, a major product of some phytoplankton species, but ranging up to dimethyl trisulfide), and freons (11, 12, and 113 used in studies of oceanic mixing).

## Direct Methods

As indicated above there are considerable problems involved in devising methods suitable for the direct determination of organic species in seawater but a number of techniques have been developed and these are described below.



**Table 2** Concentration ranges for a variety of volatile compounds extracted from estuarine waters<sup>a</sup>

Compound	Concentration range (ng l <sup>-1</sup> )
Methanethiol	10–73
n-Butane	28–60 199
2-Methylpropane	50–60 100
2-Methyl-1,3-butadiene	5–1233
n-Pentane	25–327
Isopentane	25–250
2-Methylbutene-1	50–700
Propanone	< 10–300
Dichloromethane	15–1004
Dimethylsulfide	10–814
Carbon disulfide	10–100
Propanal	30–300
Propanol-1	60–430
Propanol-2	70–300
Freon-113	25–70
Propanethiol	< 10–70
2,2-Diethylbutane	33–94
2,3-Dimethylbutane	38–109
Methyl <i>tert.</i> -butyl ether	15–81
Butanone-2	25–240
Butanone-1	25–200
n-Hexane	47–496
Pentanone-2	35–230
2-Methylpentene-1	130–700
Trichloromethane	10–7502
2-Methylfuran	10–17
1,2-Dichloroethane	15–955
1,1,1-Trichloroethane	< 5–2788
Benzene	100–55 380
2,2,3-Trimethylbutane	85–290
Carbon tetrachloride	< 10–311
n-Butanol	< 10–440
Thiophene	< 10–190
4-Methyl-2,3,-dihydrofuran	< 10–200
3-Methyl-2-butenal	< 10–180
1,2-Dibromoethane	< 10–176
Cyclohexane	30–801
Trichloroethylene	< 10–603
Pentenal	< 10–170
2,2,4-Trimethylpentane	< 10–400
2,4,4-Trimethylpentene-2	< 10–490
2,5-Dimethylfuran	< 10–50
n-Heptane	50–260
Methyl isobutylketone	< 10–347
Dimethyl disulfide	< 10–5250
Methylcyclohexane	< 10–100
1,1,2,2-Tetrachloroethylene	< 10–343
Methylbenzene	10–48 850
3-Ethylhexane	50–110
Chlorodibromomethane	10–2200
3-Methylthiophene	< 10–50
Hexanal	< 10–100
n-Octane	90–290
Octene-1	25–75
1,3-Dimethylbenzene	10–402 070
1,2-Dimethylbenzene	10–400 020
Chlorobenzene	< 10–120
Ethyl benzene	10–312 008
Tribromomethane	10–2597
Styrene	30–296
Dimethyl trisulfide	25–411

**Table 2** Continued

Compound	Concentration range (ng l <sup>-1</sup> )
n-Nonane	95–357
Nonene-1	90–300
Isopropylbenzene	10–47 307
1,5-Cyclooctadiene	< 10–400
Cyclooctene	< 10–380
$\alpha$ -Pinene	25–412
Camphene	25–99
2,4-Dimethyl-4-vinylcyclohexane	31–45
n-Propylbenzene	15–2391
Benzaldehyde	15–569
p-Cymene	35–45
1,3,5-Trimethylbenzene	20–1500
1,2,4-Trimethylbenzene	100–1570
1,2,3-Trimethylbenzene	50–6200
2,3-Dihydroindene	30–400
Limonene	25–633
Indene	55–277
1,2-Dichlorobenzene	35–107
n-Undecane	87–619
1-Methyl-3-propylbenzene	25–550
1-Methyl-2-propylbenzene	25–590
1,4-Dimethyl-3-ethylbenzene	25–200
1,3-Dimethyl-3-ethylbenzene	25–210
2,3-Dimethyl-4-ethylbenzene	30–100
1,3-Dimethyl-4-ethylbenzene	35–120
1,2,3,5-Tetramethylbenzene	25–1430
1,2,3,4-Tetramethylbenzene	150–8767
Naphthalene	45–894
n-Dodecane	95–707
Biphenyl	15–42
2-Methylnaphthalene	15–490
1-Methylnaphthalene	55–450
1,3-Dimethylnaphthalene	55–403
1,2-Dimethylnaphthalene	25–356
n-Hexadecane	30–238
n-Heptadecane	15–209
n-Octadecane	35–300
Pristane	30–155
n-Nonadecane	35–250
Phytane	30–200
n-Eicosane	55–73

<sup>a</sup>Data from Bianchi, Varney, and Phillips (1991) *Journal of Chromatography* 542: 413–450.

### Analysis for Total Dissolved Carbon

The dissolved organic carbon (DOC) reservoir in sea water represents a very significant component of the oceanic carbon cycle but reliable quantitative determination of this parameter has proved to be difficult. For over a decade since the publication of Sugimura and Suzuki's Marine Chemistry paper there has been considerable debate about the most appropriate technique(s) for the determination of DOC in seawater. Techniques can be effectively divided into three categories:

1. (a) Removal of inorganic carbon and organic volatiles before analysis by acidification followed

- by gas stripping. (b) Continuous flow or flow injection methods utilizing UV or UV + chemical wet oxidation of the dissolved organic carbon to CO<sub>2</sub> followed by quantitative measurement using (generally) nondispersive infrared gas analyzers or by chemiluminescence.
2. (a) Removal of inorganic carbon and organic volatiles before analysis by acidification followed by gas stripping. (b) High-temperature catalytic oxidation followed by detection of evolved CO<sub>2</sub> using (mostly) infrared absorption.
  3. One recent study has employed inductively coupled plasma-atomic emission spectrometry methods for nonvolatile dissolved carbon analysis with detection limits of 0.07 and 0.0007 mg l<sup>-1</sup> C in terms of total organic carbon (TOC) and inorganic carbon (IC), respectively. However, this method is not widely used.

The debate over techniques (1) and (2) initially centered on the main component of the analytical technique (i.e., ‘wet-UV’ or ‘dry-high temperature’) but it has become apparent that the most significant difficulties with this measurement are more to do with analytical quality control issues such as the use of appropriate blanks, strict control of day-to-day operating conditions, and great care in calibration procedures. A very recent report by Sharp and co-workers on an extensive intercalibration exercise states that ‘Experienced oceanic analysts, with internal or shared reference materials, can now show reproducibility and comparability at a level closer to 2%’. The development of DOC reference materials is in its early stages and once such materials are reliably available then the DOC issue should settle down as a difficult but essentially routine measurement.

### Dissolved Organic Nitrogen and Phosphorus

Dissolved organic nitrogen (DON) and dissolved organic phosphorus (DOP) are receiving an increasing amount of attention from marine chemists because of the recognition that they represent more active components of their respective nutrient cycles than had previously been thought. Problems relating to the determination of total DON and DOP are very similar to those for DOC outlined above. Destruction of the organic matter by intense UV light to transform organic to inorganic species has been widely used. The nitrate + nitrite or orthophosphate formed may then be determined by classic colorimetric methods such as that involving the formation of either an azo dye by reactions with naphthylethylenediamine and sulfanilamide (for nitrogen) or a molybdenum blue complex (for phosphorus). These systems can be automated.

In the case of DON a high-temperature catalytic oxidation method has also been used for the initial step and whilst some of the same issues arise as for DOC determinations these are generally manageable.

### Analysis of Marine Uncharacterized Material

The importance of the marine uncharacterized material (UCM) fraction is that it represents a large proportion (~90%) of the total marine dissolved organic matter. It is very clear that the material is not a single substance but a highly complex mixture of polymeric (largely polyphenolic) degradation products that may range in relative molecular mass from a few hundreds to many (tens of) thousands with only a few group properties. These include the yellow/orange color from which the original German name (Gelbstoff) derived, a fluorescence emission maximum of ~430 nm and a monotonic absorption spectrum that increases into the UV. Whilst many attempts have been made to utilize the optical characteristics of UCM to assess its composition these have been largely unsuccessful although when used in combination with other methods (e.g., 3D spectrofluorometry, LC, and capillary electrophoresis (CE)) it has been possible, for example, to differentiate between marine and fresh water samples. Stable isotope analysis, <sup>13</sup>C nuclear magnetic resonance (NMR), CE, pyrolysis GC, and GC-MS techniques have also been utilized with some success.

Some recent progress has been made. For example, a recent study employing both proton NMR and monosaccharide analysis has suggested that there are at least two major classes of UCM. The major component is acyl polysaccharide, a biopolymer rich in carbohydrates, acetate, and lipid and that this can account for between 50% and 80% of the total high molecular weight dissolved organic carbon in surface samples.

### Specific Compounds or Classes

Procedures for the determination of biologically active compounds have to be capable of operation at sea. This considerably reduces the number of suitable methods so it is not surprising that such analyses are comparatively rare and restricted to a limited suite of compounds. Some of the available techniques are described below.

#### Amines

Amines form a major fraction of the characterizable marine DON fraction and most attention has been focused on primary amines (notably amino acids) in marine systems. Primary amines may be reacted with *o*-phthaldialdehyde in the presence of mercaptoethanol and at a pH of ~11.5 to yield highly fluorescent



isoindole derivatives. This reaction has been exploited in two different ways to assess either the total primary amine or the individual primary amines in seawater. A real-time, *in situ* total primary amine analyzer based on flow injection techniques has been developed for natural waters though the application has not yet been widely adopted for seawaters. The same reaction has been used as either a postcolumn or, more usually, a precolumn derivatizing agent in liquid chromatography. The precolumn derivatization is followed by separation by gradient elution on a C<sub>18</sub>, reversed-phase column with fluorescence detection and is capable of use at sea. The method is extremely sensitive and detection limits in the low femtomolar range for individual amino acids have been quoted when using large sample volumes. For routine applications detection limits in the picomolar range are probably more realistic. However, the method is quite difficult to perform reproducibly and, with such high sensitivity, contamination is a problem. Additionally, the acetonitrile required to give optimum resolution may not be acceptable aboard research ships. Nevertheless, the method has been successfully employed to determine over 20 amino acids in seawater. Recent developments in chiral chromatography are also beginning to find marine applications.

### Carbohydrates

Carbohydrates represent an important and biologically active fraction of the DOC components within seawater. The classical method in which carbohydrates have been reacted with hot concentrated sulfuric acid (to form furfural derivatives that are determined colorimetrically) has been applied to the analysis of total carbohydrates in seawater. This method is not very specific and has now been largely replaced by better (and safer) procedures. Examples of such methods for dissolved monosaccharides include: (1) Reducing them to the respective sugar alcohols, periodate oxidation to formaldehyde, and detection of this compound with 3-methyl-2-benzothiazoline hydrazone as a colored complex at 635 nm. This method is still relatively indiscriminate and will detect all glycol-containing compounds such as alditols, uronic acids, and amino sugars as well as the monosaccharides. (2) A more specific method involving the reaction of carbohydrates with L-tryptophan/sulfuric acid in the presence of boric acid to form a violet color. Monosaccharides and polysaccharides (hydrolyzed to monosaccharides by the sulfuric acid) may be determined with excellent precision at the mg l<sup>-1</sup> level. (3) Spectrophotometric analysis of a colored product of a reaction with the reagent 2,4,6-tripyridyl-s-triazine.

Individual dissolved monosaccharides have been determined after desalting of the sample by ion-exchange membrane electrodialysis followed by evaporative concentration and ion-exchange LC using colorimetric detection with Cu<sup>2+</sup>-aspartic acid-disodium-2,2' bichinchoninate. More recently a number of papers have described the use of anion-based LC with pulsed amperometric detection for individual monosaccharide analysis though it has been noted that temperature control of the system is particularly important.

### Urea

Recent studies have shown that urea is an important natural constituent of seawater but also derives from human uses of urea as fertilizer and is transported to the oceans primarily through the atmosphere as an anthropogenic source of DON. The most widely used reaction for urea determination is with diacetyl monoxime, semicarbazide, and manganous ions to produce a magenta-colored complex. Chloride and phosphate ions are also included in the reaction to improve sensitivity and reproducibility, respectively. The method has a detection limit of ~100 pmol l<sup>-1</sup>. The method is capable of being automated in a continuous flow system.

### Oil

Oil, and particularly crude oil, is the marine pollutant with the highest public profile and in most cases the most significant coastline contaminant. The need to determine dissolved oil concentrations in water is therefore a common requirement of pollution monitoring schemes. The chemical complexity of oil means that no single analytical method is ideal and the actual technique employed will normally be determined by the need of the analyst for more (or less) detailed information about the oil. In all cases the method selected is a trade-off between required details, time taken to complete an analysis, and the availability of equipment. For routine work, the commonest method is solvent extraction and quantitation by fluorescence measured at 360 nm. This procedure measures not only dissolved oil but also finely dispersed oil such as that commonly found after a spillage has occurred in rough weather. Problems of calibrating the method are addressed by using the source oil (if known and available) and/or chrysene solutions. Chrysene is accepted internationally as a standard for this purpose and oil concentrations are often reported as chrysene equivalents. Multidimensional contour plots of fluorescence excitation/emission spectra are sometimes used as an additional tool to 'fingerprint' dissolved oil samples as an aid to the identification of their sources.

For more detailed study of dissolved oil components GC and/or GC–MS techniques are generally used. The oil is solvent extracted from the sample and after a cleanup procedure may be directly injected into the GC. Interpretation of such data is generally difficult because of the degraded nature of the oil, the compound-selective nature of the dissolution process, and its mixture with other dissolved components. In practice, it is more common to analyze bulk oil samples from a slick rather than to examine the dissolved fraction. Only if there were a need to assess the extent to which the more toxic, water-soluble components were spreading within the water would such a detailed analysis be contemplated.

### Organic Contaminants Other than Oil

It can be assumed that virtually all reasonably persistent organic compounds used in human society will eventually find their way into the oceans through a combination of atmospheric transport and terrestrial runoff. The variety of compounds involved presents significant challenges to the analyst. The techniques for the determination of dissolved organic contaminants in seawater are generally very similar to those utilized for natural compounds. Adsorption onto Amberlite XAD-2 or polyurethane foam or, alternatively, solvent extraction into for example hexane, are common procedures. Cleanup using column chromatography, concentrated H<sub>2</sub>SO<sub>4</sub> (for chlorinated pesticides), or occasionally thin-layer chromatography are then followed by GC using flame ionization detection (FID), electron capture detection (ECD), or mass spectrometric detection. Less commonly LC is used with either UV absorption (including diode array systems) or fluorescence detection. LC–MS is also now finding applications in marine systems though largely in sedimentary biogeochemistry.

## Emerging Techniques

The primary ideal of the marine chemical analyst is to have access to equipment capable of providing real-time, *in situ* data for every chemical of interest, thus permitting the chemical data produced to provide the best possible insight into the myriad processes going on within the oceans. Needless to say this ambition is a very long way from being realized. However, there are one or two analytical areas where there are signs that progress could be made in this direction. The development of small, relatively rugged GC–MS systems for ship or *in situ* use represents the way in which many analyses of dissolved organic compounds in seawater will be performed in the future.

*See also: Water Analysis: Seawater – Inorganic Compounds; Organic Compounds; Biochemical Oxygen Demand; Oil Pollution.*

## Further Reading

- Barcelo D and Hennion MC (1997) Sampling of polar pesticides from water matrices. *Analytica Chimica Acta* 338(1–2): 3–18.
- Grasshoff K, Ehrardt M, and Kremling K (1999) *Methods of Seawater Analysis*, 3rd edn. New York: Wiley-VCH.
- Sharp JH, Carlson CA, Peltzer ET, *et al.* (2002) Final dissolved organic carbon broad community intercalibration and preliminary use of DOC reference materials. *Marine Chemistry* 78(4): 171–184.
- Spyres G, Nimmo M, Worsfold PJ, Achterberg EP, and Miller AEJ (2000) Determination of dissolved organic carbon in seawater using high temperature catalytic oxidation techniques. *Trac-Trends in Analytical Chemistry* 19: 498–506.
- Sugimura Y and Suzuki Y (1988) A high-temperature catalytic-oxidation method for the determination of non-volatile dissolved organic-carbon in seawater by direct injection of a liquid sample. *Marine Chemistry* 24(2): 105–131.

## Seawater – Dissolved Organic Carbon

**G Spyres**, City University of New York, Flushing, NY, USA

**M Nimmo**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The anthropogenically induced increase of greenhouse gases (e.g., CO<sub>2</sub>) in the atmosphere has

triggered what is now widely recognized as the ‘global warming’ effect. Due to the potential adverse climatic conditions associated with global warming, interdisciplinary, international scientific environmental research programs (e.g., International Geosphere Biosphere Programme) in the past decade have increasingly focused on the global carbon cycle and the quantification of fluxes of atmospheric CO<sub>2</sub> at marine and terrestrial system interfaces. In part, these studies are aimed at resolving the discrepancies in the

For more detailed study of dissolved oil components GC and/or GC–MS techniques are generally used. The oil is solvent extracted from the sample and after a cleanup procedure may be directly injected into the GC. Interpretation of such data is generally difficult because of the degraded nature of the oil, the compound-selective nature of the dissolution process, and its mixture with other dissolved components. In practice, it is more common to analyze bulk oil samples from a slick rather than to examine the dissolved fraction. Only if there were a need to assess the extent to which the more toxic, water-soluble components were spreading within the water would such a detailed analysis be contemplated.

### Organic Contaminants Other than Oil

It can be assumed that virtually all reasonably persistent organic compounds used in human society will eventually find their way into the oceans through a combination of atmospheric transport and terrestrial runoff. The variety of compounds involved presents significant challenges to the analyst. The techniques for the determination of dissolved organic contaminants in seawater are generally very similar to those utilized for natural compounds. Adsorption onto Amberlite XAD-2 or polyurethane foam or, alternatively, solvent extraction into for example hexane, are common procedures. Cleanup using column chromatography, concentrated H<sub>2</sub>SO<sub>4</sub> (for chlorinated pesticides), or occasionally thin-layer chromatography are then followed by GC using flame ionization detection (FID), electron capture detection (ECD), or mass spectrometric detection. Less commonly LC is used with either UV absorption (including diode array systems) or fluorescence detection. LC–MS is also now finding applications in marine systems though largely in sedimentary biogeochemistry.

## Emerging Techniques

The primary ideal of the marine chemical analyst is to have access to equipment capable of providing real-time, *in situ* data for every chemical of interest, thus permitting the chemical data produced to provide the best possible insight into the myriad processes going on within the oceans. Needless to say this ambition is a very long way from being realized. However, there are one or two analytical areas where there are signs that progress could be made in this direction. The development of small, relatively rugged GC–MS systems for ship or *in situ* use represents the way in which many analyses of dissolved organic compounds in seawater will be performed in the future.

*See also: Water Analysis: Seawater – Inorganic Compounds; Organic Compounds; Biochemical Oxygen Demand; Oil Pollution.*

## Further Reading

- Barcelo D and Hennion MC (1997) Sampling of polar pesticides from water matrices. *Analytica Chimica Acta* 338(1–2): 3–18.
- Grasshoff K, Ehrardt M, and Kremling K (1999) *Methods of Seawater Analysis*, 3rd edn. New York: Wiley-VCH.
- Sharp JH, Carlson CA, Peltzer ET, *et al.* (2002) Final dissolved organic carbon broad community intercalibration and preliminary use of DOC reference materials. *Marine Chemistry* 78(4): 171–184.
- Spyres G, Nimmo M, Worsfold PJ, Achterberg EP, and Miller AEJ (2000) Determination of dissolved organic carbon in seawater using high temperature catalytic oxidation techniques. *Trac-Trends in Analytical Chemistry* 19: 498–506.
- Sugimura Y and Suzuki Y (1988) A high-temperature catalytic-oxidation method for the determination of non-volatile dissolved organic-carbon in seawater by direct injection of a liquid sample. *Marine Chemistry* 24(2): 105–131.

## Seawater – Dissolved Organic Carbon

**G Spyres**, City University of New York, Flushing, NY, USA

**M Nimmo**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The anthropogenically induced increase of greenhouse gases (e.g., CO<sub>2</sub>) in the atmosphere has

triggered what is now widely recognized as the ‘global warming’ effect. Due to the potential adverse climatic conditions associated with global warming, interdisciplinary, international scientific environmental research programs (e.g., International Geosphere Biosphere Programme) in the past decade have increasingly focused on the global carbon cycle and the quantification of fluxes of atmospheric CO<sub>2</sub> at marine and terrestrial system interfaces. In part, these studies are aimed at resolving the discrepancies in the

estimates of carbon reservoirs and fluxes between reservoirs. Of particular importance is the dissolved organic carbon (DOC) pool in the oceans, being equivalent to that of atmospheric CO<sub>2</sub> (700 pg C). DOC is an important source of carbon for community respiration. Indeed, a net oxidation of only 1% of seawater DOC within a year would be sufficient to generate a CO<sub>2</sub> flux larger than that produced by fossil fuel burning. Hence, its accurate quantification in ocean systems, with both spatial and temporal variability, is one essential requirement in establishing a reliable global C budget. Furthermore, DOC is an essential chemical driver of aquatic biogeochemical cycles. It influences the distribution and transport of key chemical constituents such as inorganic nutrients and trace metals, and can facilitate the vertical export of aquatic colloids and associated material via aggregation processes. The reliable detection of DOC is thus crucial in our understanding of aquatic biogeochemical cycles.

## Historical Overview of Applied Analytical Techniques

Analytical techniques, which have been employed for the measurement of DOC in seawater, are (1) wet chemical oxidation (WCO), (2) ultraviolet (UV) oxidation, (3) dry combustion, and (4) high-temperature catalytic oxidation (HTCO).

Prior to the 1980s, WCO was the most adopted method; its application to DOC seawater analysis was first reported in the 1930s. Oxidants used have included peroxydisulfate, dichromate in sulfuric acid, and potassium peroxide. Prior to oxidation of the sample DOC, inorganic carbon was removed by purging of the acidified sample (pH < 2). The CO<sub>2</sub> produced from the oxidation process is measured by one of the number methods from nondispersive infrared gas analysis (IRGA), colorimetrically, and conductometrically by conversion to CH<sub>4</sub>. However, the limitation of WCO was the progressive weakening of the oxidant (particularly in the presence of Cl<sup>-</sup>), during sample treatment, which may have led to incomplete oxidation. Furthermore, different reagents would have had different oxidation strengths, potentially leading to different oxidation efficiencies.

UV oxidation methods were proposed during the same period as WCO, although they were not adopted until the 1970s. As with WCO, inorganic carbon is removed from the seawater prior to oxidation. Typically, water samples were exposed to a UV radiation source light via a quartz capillary coiled around the light source. Some UV radiation procedures included the addition of a chemical oxidant.

The advantage of the UV oxidation approach compared to that of WCO was that oxidants were generated continuously. However, the main disadvantage with UV oxidation was the deterioration of the UV light source with age. This lowered the efficiency of the oxidation process and potentially led to non-comparability of different research group datasets.

Dry oxidation methods (applied in the 1960s and 1970s) involved initial acidification and dehydration of the sample, followed by high-temperature combustion of the resultant 'sea salt'. Sample dehydration was achieved by heating to dryness at 60°C or by freeze-drying. The main disadvantage of these methods was the high degree of sample manipulation required during the dehydration step, enhancing the risk of sample contamination. The method is also labor intensive and cannot be automated preventing real-time analysis.

Since the late 1980s, HTCO methods, which typically incorporate an Al/Pt catalysis combustion column, have been adopted. The popularity of this technique was, in part, due to the advances in the sensitivity and stability of the IRGA and its real-time direct injection capabilities, requiring lower sample volumes (50–200 µl) compared to those for WCO (5–100 ml). HTCO is rapid and precise ( $\pm 1$ –2% RSD) and can yield equivalent or greater amounts of DOC than WCO methods. In addition, it is effectively adopted for routine analyses and is stable for ship-board determinations. This was first proposed in a paper by Sugimura and Suzuki, whose HTCO method detected 1.5–5 times higher concentrations of DOC (180–280 µmol l<sup>-1</sup> C) in surface oceanic waters than had been previously reported by WCO methods. Underestimation of the DOC concentrations in seawater measured by WCO would have clearly meant an underestimation of the DOC pool in the oceans. The findings of Sugimura and Suzuki, however, were not always reproduced by other researchers, and were subsequently retracted as a result of inaccuracies due to insufficient consideration of the influence of gases other than CO<sub>2</sub> and water/system blanks not subtracted from the initial data. Nevertheless, this work mobilized the scientific community into re-evaluating the accuracy of conventional WCO analytical techniques, and paved the way to the subsequent adoption of HTCO as the preferred analytical method for the determination of seawater DOC. Subsequent to Sugimura and Suzuki, research findings, both supportive and critical of the HTCO method, appeared but there were inconsistencies with the published data. This led to a US National Science Foundation/National Oceanic and Atmospheric Administration/Department of Energy sponsored workshop on the 'Measurement of Dissolved



Organic Carbon and Nitrogen in Natural Waters' (Seattle 1991) in order to critically evaluate the methodological procedures that essentially bear upon generating consistent results. In this workshop, natural seawater samples were distributed to 34 participants prior to the workshop for analysis to allow an intercalibration exercise to be performed. Poor comparability between laboratory analyses was apparent and values were within the mean values  $\pm 40\%$  RSD, although HTCO was confirmed as a more effective method than WCO. The inconsistencies were largely attributed to difficulties in the system blank correction.

In the last decade, HTCO has become the preferred analytical technique for the determination of DOC in natural water samples. With close attention to blank correction and sample treatment, HTCO measurements performed between analytical laboratories are now comparable and environmentally consistent.

This article focuses specifically on the HTCO technique currently in use for DOC analysis in seawater. The technique's hardware, principles, and typical analytical performance are critically presented, along with sample treatment and preservation methodologies.

## Current Procedure for the Measurement of DOC in Seawater

### Sample Collection

Changes in sample DOC composition may result from contamination, sorption onto container walls, and biological and flocculation processes. Due to the low background concentrations of DOC (typically  $40\text{--}50\ \mu\text{mol l}^{-1}\text{C}$ ) in marine waters, seawater samples are easily contaminated via: (1) the atmosphere (e.g., ship's engine exhaust fumes), (2) sampling bottles, and (3) careless handling (i.e., not wearing gloves, dirty laboratory working space). Niskin and Go-Flo bottles are commonly used for sampling the seawater column; they are designed to minimize contamination and they can pass through the air/sea interface closed. This avoids contamination from the organic-rich, microlayer on the ocean surface where surface-active organic materials are present. Glass and plastic containers (e.g., polysulfone, polycarbonate) that have been thoroughly cleaned and 'aged' (i.e., all leachable components removed by soaking in carbon-free water) are used as temporary storage bottles.

### Sample Treatment and Preservation

Filtration should be carried out immediately after sample collection. This is to minimize biological processes taking place in the sample, which may result in either a net increase (i.e., release from any

organisms present) or decrease of DOC (i.e., microbial consumption). Particulate organic carbon (POC) and the majority of organisms present in the sample can be removed by filtration; pore sizes commonly used range from  $0.2$  to  $0.7\ \mu\text{m}$ . Contamination by carbon leaching from the filter may be significant and should be removed by thorough cleaning. Glass fiber filters (GFFs) are the preferred filter media ( $0.7\ \mu\text{m}$ ; Whatman) as they are robust and are easily cleaned in a furnace by combustion at high temperatures (e.g.,  $450^\circ\text{C}$  for  $>4\text{ h}$ ). Negligible contamination from GFF filters has been observed, although their pore size allows passage of some particles (e.g., bacteria and viruses). However, the fraction of POC in the filtrate is generally negligible compared to typical DOC concentrations. Alternative filters with a smaller pore size ( $0.2\ \mu\text{m}$ ) include polycarbonate and polysulfone filters, although these can significantly contaminate the sample by leaching of carbon. Rapid clogging may present a problem when filtering highly productive waters, reducing the effective pore size of the filter, and increasing the back-pressure possibly causing DOC to leach into the sample via cell lysis. In oligotrophic waters, where the particle loading is minimal ( $<1\%$  of total organic carbon, TOC) the filtration step may be omitted.

When samples need to be preserved, acidification combined with freezing is the chosen option. Sample acidification to pH 2–3 is necessary to (1) remove dissolved inorganic carbon (DIC), and (2) to arrest biological processes from occurring. Mercuric chloride had been used until it was suggested that it might deactivate the catalyst in the HTCO process. High-purity orthophosphoric acid ( $\text{H}_3\text{PO}_4$ ) is the preferred choice of acid as it is commercially available and does not contribute significantly to the procedural blank. Several possible interferences can arise from acidification; these include the production of volatile products from the hydrolysis of organic matter and the precipitation of macromolecules. However, no studies on these effects have been reported to date. Once acidified, samples can be stored frozen at  $-20^\circ\text{C}$  or in flame-sealed airtight glass ampoules refrigerated at  $4^\circ\text{C}$ .

### Decarbonation

DIC concentration is at least an order of magnitude greater than DOC concentration in seawater and therefore has to be removed before analysis. This is achieved by subsequent purging of the acidified sample with  $\text{CO}_2$ -free gas (e.g., ultrahigh purity (UHP) nitrogen, oxygen, or helium). Loss of volatile organic carbon during purging is likely to be minor, amounting to  $<1\%$  of total organic carbon. The efficiency

of DIC removal will depend upon gas flow rate, duration, sample volume, and sample pH. For small sample volumes ( $<40$  ml) a gas flow rate of  $100\text{--}120\text{ ml min}^{-1}$  for 6–8 min should be adequate. Coagulation of colloidal matter on the bubbles, flocculation, and precipitation processes during purging could result in heterogeneity of the sample and therefore mixing of the sample prior to injection may be necessary.

### Analysis of DOC by HTCO

Direct injection (typically  $200\text{ }\mu\text{l}$  aliquot volumes) HTCO systems operate on the principle of oxidation of organic carbon compounds in aqueous samples and subsequent quantification of the generated  $\text{CO}_2$  by IRGA. Injection is carried out onto an oxidation column filled with catalyst (e.g., platinum on aluminum oxide, cobalt oxide, copper oxide) at a temperature between  $600^\circ\text{C}$  and  $900^\circ\text{C}$  in a carbon-free, pure gas atmosphere (e.g., oxygen, nitrogen). The organic matter present in the sample is oxidized on the catalyst to  $\text{CO}_2$ . The stream of gas products is dried using a dehumidifier (e.g., electronic dehumidifier, magnesium perchlorate) and purified by means of gas scrubbers (e.g., halogen scrubber) and particle filters before final determination with the IRGA. The signal (voltage) from the IRGA is recorded using a data collection/integration system; the peak area is used to calculate the amount of carbon present in the sample. The system is calibrated using a carbon compound (e.g., potassium hydrogenphthalate, sulfathiazole, sodium hydrogencarbonate, sucrose) diluted

in low-carbon water (LCW). The main components of a HTCO system are highlighted in Figure 1.

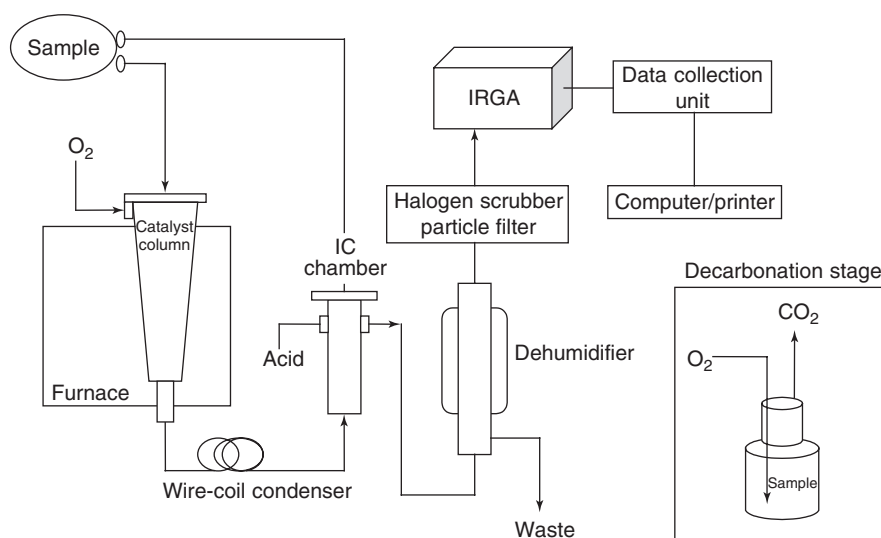
### Analytical Challenges Associated with HTCO

Although current HTCO techniques can provide highly precise ( $\pm 1\text{--}2\%$  RSD) data (Table 1), some analytical problems may be encountered. Quantitatively, the most significant sources of error are: (1) mechanical effects associated with HTCO techniques (e.g., sample injection, salt deposition, memory effects), (2) the estimation of the system blank (i.e., contaminant carbon emissions from the catalyst and components of the analytical system), and (3) the oxidation efficiency.

### Mechanical Effects

Automated injection systems are less labor intensive and less susceptible to contamination leading to improved repeatability.

As the sample is injected onto the catalyst at high temperatures, expansion occurs causing a pressure pulse higher than the void space of the column. A cold zone at the headspace of the column results, and this leads to the deposition of salt/carbonaceous residues. The salt deposits may block the column head resulting in an incomplete combustion of the organic compounds and/or a memory effect (i.e., carryover between samples). Flushing with copious amounts of LCW through the system can remove salt residues but replacement of catalyst may be necessary after extensive salt 'degradation'.



**Figure 1** A schematic diagram of a typical high-temperature catalytic oxidation-discrete injection system for the analysis of DOC (IC, inorganic carbon; IRGA, infrared gas analyzer) (Reprinted with permission from Spyres G, Nimmo M, Miller AEJ, Worsfold PJ, and Achterberg EP (2000) Determination of dissolved organic carbon in seawater using high temperature oxidation techniques. *Trends in Analytical Chemistry* 19(8): 498–506; © Elsevier.)



**Table 1** A summary of analytical figures of merit for HTOC systems ( $r^2$ , regression coefficient of standard calibration; CV%, coefficient of variation)

<i>Analyzer</i>	<i>Detector</i>	<i>Catalyst type</i>	<i>Standard material (<math>r^2</math>)</i>	<i>Oxidation efficiency (% recovery)</i>	<i>DOC range (<math>\mu\text{mol l}^{-1} \text{ C}</math>)</i>	<i>Precision (CV%)</i>
Sumigraph TOC-90	IRGA (Beckman 880)	3% Pt/Al, Cuprox, Sulfix	KHP, glucose, sodium carbonate ( $> 0.99$ )	91–96%	78–148	$\pm 0.2$ –1.3%
Sharp (1973) clone	IRGA	1. 5% Pt on Triton Kaowool/ Co <sub>2</sub> O <sub>3</sub> on Al beads	1. KHP ( $\geq 0.995$ )	91–111%	76–164	1–2%
		2. Pure Pt catalyst	2. Glucose, antipyrine, sulfathiazole, urea, thiourea, glycine, methyl orange		40–400	$\pm 2\%$
1. Sealed tube combustion	1. Capacitance manometer	1. Platinic chloride	1. None quoted	1. Reference method	0–200	$\leq 1.5\%$
2. Homemade	2. IRGA (Licor Li6252)	2. 3% Pt on alumina/CuO/ Sulfix	2. 0.9984	2. 95% average		
3. Shimadzu TOC-5000	3. IRGA	3. Pt on alumina beads	3. 0.9994	3. 98.2–100.6%		
Shimadzu TOC-5000	IRGA (Licor Li6252)	0.5% Pt on alumina	KHP (0.9989)	None quoted	CRM: 1. KHP: 0–417 2. CRM: 43–51	CRM: 1. $\pm 0.4$ –6% 2. $\pm 0.6$ –7%
MQ1001 (modified)	1. IRGA (Licor 7000)	1. Quartz beads/CuO/Sulfix	1. KHP, seawater	1. 106.1%		
	2. IRGA (Licor Li6252)		2. CRM (none quoted)	2. 103.6%		

Reprinted with permission from Spyres G, Nimmo M, Worsfold PJ, Acterberg EP, and Miller AEJ (2000) Determination of dissolved organic carbon in seawater using high temperature catalytic oxidation techniques. *Trends in Analytical Chemistry* 19(8): 498–506; © Elsevier.

### System Blank

In order to accurately measure DOC in seawater, it is necessary to quantify the system blank. The catalyst is considered to be the major source of carbon contamination in the system. Carbon contamination can vary between catalysts (e.g., alumina has a higher adsorption capacity for CO<sub>2</sub> than silica) and there is lack of a completely carbon-free water to calculate the carbon contamination from the system components alone.

It is possible with some commercially available TOC analyzers (e.g., Shimadzu TOC-5000) to condition the catalyst and effectively measure the system blank by running a ‘closed-loop blank-check’ program. This program involves the injection of LCW onto the catalyst where it is combusted and collected downstream as pyrolyzed water (i.e., theoretically carbon-free water). The latter is subsequently reinjected to determine the system blank. An alternative correction approach, when such a program is not available, is the subtraction of an average of LCW injections (i.e., system plus water blank) from the sample measurements. However, this blank-correction method may result in an underestimation of the DOC concentration in seawater samples. The use of reference materials (see below) is important for system accuracy and reliability. Blank concentrations typically quoted in the literature are  $<10 \mu\text{mol l}^{-1}\text{C}$  after rigorous cleaning/conditioning of the catalyst. A common protocol for blank estimation and correction has yet to be accepted; therefore, reporting the applied method of blank estimations is crucial.

### Oxidation Efficiency

Oxidation efficiency of the HTCO technique will be a function of the refractivity of the natural organic material dissolved in the analyzed sample. Since the chemical character of dissolved organic matter is largely unknown, it is not possible to be sure that any standards chosen will be representative of naturally occurring organic matter. Nevertheless, efforts to evaluate the oxidation efficiency of HTCO have been made using various model compounds with different degrees of refractivity (i.e., caffeine, antipyrine, thiourea, graphite, diamond). Recent studies on percentage recovery for current HTCO techniques quote  $>94\%$ ; however, a direct comparison between techniques is not often possible due to the use of different catalysts and reference compounds. A summary of analytical figures of merit from recent studies on the oxidation efficiency of HTCO systems is given in Table 1. Some studies have shown that complete oxidation can be achieved without a catalyst (e.g., using pure quartz beads) and using a carrier gas that does not include oxygen, bringing into question the

combustion mechanism occurring in HTCO systems. It is recommended that a set of compounds of different refractivity, each at the predicted sample range of concentrations, should be injected to evaluate the oxidation efficiency. Oxidation efficiency should be determined on each day of analysis until constant percentage recovery patterns are established, after which periodic assessment is sufficient.

### Analytical Performance

Published literature on measurements of DOC by HTCO has shown that, currently, the most commonly adopted standard material used for analytical accuracy is potassium hydrogen phthalate (KHP). A summary of analytical performance criteria for several HTCO systems using predominantly KHP is provided in Table 1.

### Certified Reference Materials

Analyses of DOC in seawater samples by HTCO techniques can display unreliability on a daily basis. Certified reference materials (CRMs) are therefore essential for checking the performance of the HTCO techniques. Although, official CRMs are not currently available, in the last decade, the international community has seen several efforts to produce reference materials for DOC measurements. Interlaboratory exercises are underway to provide ‘consensus reference materials’ (i.e., samples assigned an agreed concentration by several laboratories) to the analytical community for increased accuracy. Achieving good agreement between analysts (i.e., values that are within the expected concentrations  $\pm 1\text{--}2\%$  RSD) is crucial for the rigorous comparison and interpretation of different environmental datasets.

Accurate measurements of DOC coupled with techniques such as tangential flow ultrafiltration for size distribution of DOM, nuclear magnetic resonance spectroscopy, soft-ionization mass spectrometry, and elemental ratios will allow marine chemists to define major compositional changes of DOM in the marine environment. Such knowledge can provide an understanding of the mechanisms of biogeochemical cycles, as well as the ability to calculate DOM fluxes accurately on regional and global scales within marine systems and at their interface.

*See also:* **Quality Assurance:** Reference Materials; Production of Reference Materials. **Water Analysis:** Overview; Seawater – Organic Compounds; Organic Compounds.

### Further Reading

Hansell DA and Carlson CA (2002) *Biogeochemistry of Marine Dissolved Matter*. San Diego: Academic Press.

*Marine Chemistry* (1993) vol. 41, Special Issue containing reports from the HSF-NOAA-DOE sponsored Workshop on Measurements of Dissolved Organic Carbon and Nitrogen in Natural Waters.

Sharp JH, Carlson CA, Peltzer ET, *et al.* (2002) Final dissolved organic carbon broad community intercalibration

and preliminary use of DOC reference materials. *Marine Chemistry* 77(4): 239–253.

Sugimura Y and Suzuki Y (1988) A high temperature catalytic oxidation method for the determination of non-volatile dissolved organic carbon in seawater by direct injection of a liquid sample. *Marine Chemistry* 24: 105–131.

## Seawater – Inorganic Compounds

**T D Jickells and A R Baker**, University of East Anglia, Norwich, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The oceans cover about two-thirds of the globe and represent most of the the biggest water reservoirs on earth. However, because humans live only on the fringes of the oceans, they are less studied, less well understood, and less perturbed from their pristine state than the terrestrial environment. The chemical analysis of the seawater itself represents a major analytical challenge that for many elements is only now beginning to be successfully accomplished. These problems are in part oceanographic, reflecting the size and complexity of the ocean system, and in part chemical, and these latter are the focus of this brief article. This article covers the inorganic analysis of seawater, excluding radioactive species, in both deep-sea and coastal waters and excludes consideration of gaseous species.

The dissolved ion chemistry of seawater is dominated by eight ions all present at millimolar concentrations or more, with sodium and chloride as the overwhelmingly dominant ions (Table 1). The ratio of these eight ions one to another is very constant in ocean waters because the long residence time of these ions ( $\sim 10^6$  years) allows for complete mixing of the waters at fast rates compared to internal mixing,

input, and removal processes. Seawater is also strongly buffered by the hydrogencarbonate–carbonate–carbon dioxide system at about pH 8. In addition to these dissolved ions, seawater contains both biological and abiological particulate matter, dissolved gases, and dissolved organic matter. The analysis of these phases is the subject of separate articles. While these so-called major ions of seawater are present at concentrations of  $10^{-3} \text{ mol l}^{-1}$  or greater, every stable element in the periodic table is probably present, though at much lower concentrations ranging from  $10^{-6}$  down to  $10^{-12} \text{ mol l}^{-1}$  or possibly lower still. Many of these ions display important variations of concentrations in space and time that are of great interest to marine scientists both in terms of understanding the chemistry of the oceans and in terms of monitoring environmental quality. The challenge for chemical analysis of seawater is therefore to often measure the concentrations of ions (or complexes) present at submicromolar concentrations in the presence of millimolar concentrations of the major ions.

### Sampling

The first task is to collect a water sample that represents the area from which it was collected. There are numerous problems of collecting representative samples associated with the heterogeneity of the biogeochemical processes in the oceans. For the analyst the concerns are related to sample integrity. The collection of surface water samples from a small boat can be relatively straightforward, with the sample storage bottle being opened under water and filled by someone leaning over the front of the boat, taking great care to avoid contaminating the sample with surface films on the water or emissions from the boat. On larger vessels, devices capable of continuously pumping uncontaminated surface water samples into the ship's laboratories are now available. The task is more difficult when considering the collection of samples from water depths of several kilometers. Many early sampling campaigns used inappropriate collection systems, resulting in the

**Table 1** Major ion composition of seawater

<i>Ion</i>	<i>Concentration (<math>10^{-3} \text{ mol kg}^{-1}</math>) at salinity of 35</i>
Sodium	470
Magnesium	53
Calcium	10.3
Potassium	10.2
Chloride	550
Sulfate	28
Hydrogencarbonate	2.3
Bromide	0.84

*Marine Chemistry* (1993) vol. 41, Special Issue containing reports from the HSF-NOAA-DOE sponsored Workshop on Measurements of Dissolved Organic Carbon and Nitrogen in Natural Waters.

Sharp JH, Carlson CA, Peltzer ET, *et al.* (2002) Final dissolved organic carbon broad community intercalibration

and preliminary use of DOC reference materials. *Marine Chemistry* 77(4): 239–253.

Sugimura Y and Suzuki Y (1988) A high temperature catalytic oxidation method for the determination of non-volatile dissolved organic carbon in seawater by direct injection of a liquid sample. *Marine Chemistry* 24: 105–131.

## Seawater – Inorganic Compounds

**T D Jickells and A R Baker**, University of East Anglia, Norwich, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The oceans cover about two-thirds of the globe and represent most of the the biggest water reservoirs on earth. However, because humans live only on the fringes of the oceans, they are less studied, less well understood, and less perturbed from their pristine state than the terrestrial environment. The chemical analysis of the seawater itself represents a major analytical challenge that for many elements is only now beginning to be successfully accomplished. These problems are in part oceanographic, reflecting the size and complexity of the ocean system, and in part chemical, and these latter are the focus of this brief article. This article covers the inorganic analysis of seawater, excluding radioactive species, in both deep-sea and coastal waters and excludes consideration of gaseous species.

The dissolved ion chemistry of seawater is dominated by eight ions all present at millimolar concentrations or more, with sodium and chloride as the overwhelmingly dominant ions (Table 1). The ratio of these eight ions one to another is very constant in ocean waters because the long residence time of these ions ( $\sim 10^6$  years) allows for complete mixing of the waters at fast rates compared to internal mixing,

input, and removal processes. Seawater is also strongly buffered by the hydrogencarbonate–carbonate–carbon dioxide system at about pH 8. In addition to these dissolved ions, seawater contains both biological and abiological particulate matter, dissolved gases, and dissolved organic matter. The analysis of these phases is the subject of separate articles. While these so-called major ions of seawater are present at concentrations of  $10^{-3} \text{ mol l}^{-1}$  or greater, every stable element in the periodic table is probably present, though at much lower concentrations ranging from  $10^{-6}$  down to  $10^{-12} \text{ mol l}^{-1}$  or possibly lower still. Many of these ions display important variations of concentrations in space and time that are of great interest to marine scientists both in terms of understanding the chemistry of the oceans and in terms of monitoring environmental quality. The challenge for chemical analysis of seawater is therefore to often measure the concentrations of ions (or complexes) present at submicromolar concentrations in the presence of millimolar concentrations of the major ions.

### Sampling

The first task is to collect a water sample that represents the area from which it was collected. There are numerous problems of collecting representative samples associated with the heterogeneity of the biogeochemical processes in the oceans. For the analyst the concerns are related to sample integrity. The collection of surface water samples from a small boat can be relatively straightforward, with the sample storage bottle being opened under water and filled by someone leaning over the front of the boat, taking great care to avoid contaminating the sample with surface films on the water or emissions from the boat. On larger vessels, devices capable of continuously pumping uncontaminated surface water samples into the ship's laboratories are now available. The task is more difficult when considering the collection of samples from water depths of several kilometers. Many early sampling campaigns used inappropriate collection systems, resulting in the

**Table 1** Major ion composition of seawater

<i>Ion</i>	<i>Concentration (<math>10^{-3} \text{ mol kg}^{-1}</math>) at salinity of 35</i>
Sodium	470
Magnesium	53
Calcium	10.3
Potassium	10.2
Chloride	550
Sulfate	28
Hydrogencarbonate	2.3
Bromide	0.84



reporting of erroneously high concentrations owing to contamination of the samples. Since the mid-1970s, sampling systems based on nonmetallic or high-grade stainless-steel cables and all-plastic water samplers designed to open below the surface of the water and to then be closed on command at depth in the ocean have become the norm, particularly for the measurement of trace metals at nanomolar concentrations. In deep ocean environments requirements for strength may demand the use of metal sampling frames to support plastic sampling bottles and these may be either of high-grade stainless steel or possibly of a more expensive metal that is less likely to corrode, such as titanium. This has the additional advantage that titanium is not usually an analyte of great interest whereas iron is, so the possibility of contamination from a titanium frame is of less importance than from a stainless steel one.

Once the samples are returned to the sampling platform it is necessary to minimize the risk of contamination from the atmosphere on the research boat. While careful discharge of the samples into sample storage bottles on deck is often satisfactory, the ideal strategy is to carry out all sample handling (at least for trace metals) in specially designed laboratories with positive pressure of filtered air to minimize the risk of contamination. The extent to which these precautions are necessary varies depending on the analyte of interest, its contamination potential, and its ambient concentration. Furthermore, the procedures for one analyte may be incompatible with another, and careful planning and preparation are essential prerequisites for successful sampling. Some marine waters are devoid of oxygen (e.g., deeper parts of the Black Sea) and particular care is necessary in order to sample such waters without introducing oxidation artifacts. In such cases it is usually necessary to conduct all sample processing prior to sample stabilization under an inert (nitrogen) atmosphere (see Figure 1).

## Sample Treatment and Storage

In deep ocean waters particulate matter concentrations are very small and it is usually unnecessary to filter samples. However, in surface waters and coastal areas, particulate matter concentrations are much higher and filtration to remove particulate matter is usually needed. This ideally serves four purposes:

- (1) To remove living organisms whose subsequent growth or death may compromise the sample integrity.
- (2) To remove particulate matter that may interfere with the storage or subsequent analysis of the sample.



**Figure 1** Rosette sampling system containing niskin bottles for water column sampling. (Photograph: Chris Lowe.)

- (3) To allow a rigorous definition of solid and dissolved phases.
- (4) To allow separate analysis of the particulate phase to provide additional geochemical data for research and environmental monitoring.

Unfortunately, no simple filtration procedure can unambiguously achieve all these results and it is therefore normal practice to filter samples through 0.45  $\mu\text{m}$  pore-size filters and define everything in the filtrate as dissolved, although it is recognized that this will include colloidal material. The role of colloids is an important area of current geochemical research and the separate analysis of this phase is therefore sometimes considered. Many analyses cannot be conducted at sea since they require complex instrumentation, and storage is therefore necessary. After filtration further measures to stabilize samples for storage to avoid changes in concentrations as a result of biological processes or adsorption/desorption processes on the sample container walls are needed. As with sampling, careful planning and evaluation are necessary to ensure satisfactory storage, since a suitable storage procedure for one analyte may not be appropriate for another. There are a

number of strategies available to try to stabilize the samples depending on the most likely storage problems. Thus, for metal analysis, adsorption to the walls is of particular concern and samples are routinely acidified to pH  $\sim$ 2, which serves to saturate potential adsorption sites on the container and stabilize ions in solution. However, such acidification will inevitably perturb the chemical speciation of the ions if this is of interest. In the case of nutrient ions (see later), biological processes are of primary concern and two strategies are routinely used: (1) deep freezing the samples to slow all reactions and prevent migration, (2) adding a poison such as mercury(II) chloride to prevent biological activity. In some cases a combination of these approaches is used. However, storage of samples for nutrient analysis from surface waters where concentrations are extremely low and biological cycling is particularly intense is plagued with difficulties and no universally accepted storage procedure is available.

The last point to be noted under sample storage is that it is essential that sample storage be done in containers that have been rigorously precleaned to avoid the leaching of ions from the walls contaminating the samples. This is particularly of concern for trace metals and rigorous acid cleaning of sample bottles is usually necessary. For most analytes, plastic containers (e.g., polytetrafluoroethylene (PTFE), low-density polyethylene (LPDE)) are preferred, but, for mercury in particular, glass is preferred to avoid the possible loss of volatile species.

## Major Ions

Since the major ions are present in seawater at constant ratios to one another, it is normally not necessary to measure the concentrations of all the ions since the concentration of one will allow the prediction of the others. Thus, chloride has traditionally been measured using a silver nitrate titration, and from this the salinity (i.e., total dissolved salt concentrations) can be derived. Now, however, conductivity is the routinely measured parameter and this is converted to salinity by a relationship agreed internationally with interlaboratory agreement ensured by the distribution of standard seawater samples for instrumental calibration. The use of modern inductively coupled conductivity measurements with careful temperature controls allows salinities to be determined with accuracy and precisions of the order of  $\pm 0.01\%$  or better.

If the measurement of individual major ions is required, the metals can readily be determined by atomic absorption spectrometry (AAS) or inductively coupled plasma-atomic emission spectrometry

(ICP-AES) after suitable dilution and with appropriate matrix matching and background correction. It is possible to determine chloride and bromide by classic silver nitrate titration and sulfate by barium chloride precipitation, but the analytical method of choice would now probably be ion chromatography with careful control of eluent concentrations and instrument sensitivity to avoid the very large chloride peak overlapping with the other anions of interest. Ion chromatography can also be used to measure fluoride, iodate, and iodide, though electrochemical and spectrophotometric techniques are more generally used. Hydrogencarbonate (bicarbonate) is rarely determined exclusively; rather the alkalinity is measured traditionally by titration with acid to pH 4, thus determining the sum of the bicarbonate, carbonate, and the small amounts of borate and phosphate present. The alkalinity of the surface ocean and related components of the  $\text{CO}_2\text{--HCO}_3^- \text{--CO}_3^{2-}$  system in the ocean vary with space and time. Current concerns over the rate of carbon dioxide uptake by the oceans as a factor in the greenhouse effect have led to the development of extremely accurate and precise measurements of the various components of the carbon dioxide system.

## Nutrients

Many of the ions in seawater are essential for the growth of phytoplankton, the microscopic floating photosynthetic algae that form the base of the marine food web. However, many of these ions are present in relatively high concentrations (e.g.,  $\text{CO}_3^{2-}$ ,  $\text{SO}_4^{2-}$ ,  $\text{K}^+$ ,  $\text{Mg}^{2+}$ ) or in the case of some metals are required only in very small amounts. Thus, it has traditionally been accepted that only the elements nitrogen, phosphorus, and silicon are required in relatively large amounts and yet they are present at rather low concentrations and hence may limit the rates of phytoplankton growth in seawater. This view is now being challenged and it may be that some trace metals can also limit algal growth; nevertheless, the traditional view of the importance of these elements coupled with concern over eutrophication problems in coastal waters, arising from excessive inputs of these elements, justifies a separate treatment for them. Their biochemical roles are rather different from one another, with nitrogen and phosphorus being used for organic matter and silicon being used for skeleton construction by some species of algae, but all are generally called nutrients.

Dissolved inorganic phosphorus (DIP) is present in seawater as the various dissociation products of



phosphoric acid, and silicon is present mainly as the undissociated silicic acid, though the analytical methods in routine use for these ions measure all the inorganic forms of the elements. Nitrogen gas in seawater is unavailable to most, but not all marine algae, so the inorganic species of interest are nitrate (the thermodynamically stable form under oxygenated conditions), nitrite, and ammonium. In addition to the inorganic forms of these nutrients, organic nitrogen and phosphorus species are known to occur in seawater. The analyses of the individual organic compounds are discussed elsewhere, but it is common practice in some laboratories to determine total dissolved organic nitrogen and phosphorus by measuring the concentrations of DIP, nitrate, ammonium, and nitrite before and after strong chemical or photochemical oxidation and considering the differences between the two measurements to represent organic phosphorus and nitrogen. The efficiency of these oxidation procedures has recently been evaluated and there is now no doubt that there are significant levels of dissolved organic nitrogen and phosphorus in surface water.

Colorimetric methods for the determination of DIP, nitrate, nitrite, ammonium, and silicate are well established, with both manual and automatic procedures well understood and able to provide adequate sensitivity for most purposes. These methods can readily be used at sea and in the future probably *in situ*. International intercalibration exercises have shown that many laboratories now have the analytical expertise to measure nutrient concentrations at ambient levels, but this is not true of all laboratories and the analysis still requires careful analytical procedures that recognize the importance of contamination control, blank correction, and the complications arising from the saltwater matrix. Although there have been a number of such intercalibration studies, there is no widely available standard reference material for nutrient analysis at present.

In surface waters nutrient concentrations are often depleted below the detection limits of the routine colorimetric methods even when using the longest-pathlength spectrophotometer cells routinely available (10 cm). The accurate measurement of concentrations of these nutrients at nanomolar concentrations is now an important analytical goal. Therefore, in recent years, new methods for the analysis of nitrate and ammonium with lower detection limits have been developed based on chemiluminescence for nitrate, fluorescence or concentration of the standard colorimetric complex for ammonium, and preconcentration for DIP. Recently, the development of wave guides has opened up the possibility of using cell

lengths of a meter or more which allow great improvements in the sensitivity of nutrient methods using conventional colorimetric procedures.

## Trace Metals

Over recent years the main analytical challenge in inorganic marine chemistry has been the measurement of the trace metals, i.e., metals and metalloids other than those major ions in seawater. The concentrations of trace metals are universally low ( $10^{-9} \text{ mol l}^{-1}$  or less) and their determination in a seawater matrix represents a major challenge. However, as noted earlier, it was the collection of samples without contamination that prevented, until recently, the generation of oceanic profiles of trace metals that could be readily interpreted in terms of well-understood oceanographic processes. Standard reference materials have now been prepared to improve interlaboratory comparability, and a number of international intercomparison exercises have demonstrated good agreement between experienced laboratories. Furthermore, it has become clear that in deep ocean waters, trace metal concentrations show little short-term variation and show variation in concentration with depth and between ocean basins that can be explained by well-understand biogeochemical and oceanographic processes. This allows repeat sampling at locations to be used as a rigorous test of sampling and analytical procedures. Indeed, the International Oceanographic Commission (IOC) has begun a process of establishing the baseline concentrations of some trace metals in the major ocean basins of the world. Once the extremely rigorous protocols necessary for such sampling were put into force, it became possible to focus on the chemical analysis of these low concentrations.

A wide variety of different methods of analysis have been applied to different trace metals in seawater. All these documents are beyond the scope of this article and the reader is referred to Further Reading. In general, the methods fall into two groups: those that can be used without preconcentration and those that do require such a procedure. In the first group are spectrophotometric, chemiluminescence, and fluorimetric procedures, electrochemical methods, and some ICP-AES and inductively coupled plasma mass spectrometry (ICP-MS) methods. In the latter group, involving preconcentration, a variety of concentration methods have been used followed by AAS, gas chromatography, mass spectrometry, and neutron activation methods, or indeed any of the direct analysis procedures. These methods will now be briefly discussed.

## Techniques

### Electrochemical Methods

Anodic stripping voltammetry has been widely used as an analytical procedure because it has the sensitivity to measure a number of metals of interest (zinc, cadmium, copper, and lead) directly in seawater. Furthermore, the method can be sensitive to the chemical speciation of the metal. Thus, if the metals are present as strong and stable organic complexes, as is now known to be the case, for example, for a number of metals including copper, the labile (weakly complexed) metal can be determined directly and the total metal after suitable pretreatment. Recent advances mean that stability constants can also be determined for these strong complexes. The instrumentation for these analyses is portable and robust, thus allowing determinations to be made at sea, minimizing storage artifacts. In the future *in situ* electrochemical methods will become practical. The applicability of electrochemical methods has been limited to a few metals of appropriate reduction potentials, but more recently techniques involving complexation of a metal and the electroplating of this complex using cathodic stripping voltammetry have extended the range of metals that can be measured and the sensitivity of analysis for other metals.

### Spectrophotometry, Chemiluminescence, and Fluorescence Methods

There are a number of colorimetric methods available for the measurement of trace metals in seawater, though, in general, few have the sensitivity to measure the low ambient concentrations. However, in some environments the ferrozine method for iron and the formaldoxine method for manganese have been used successfully. Recently, these methods have been adapted to allow *in situ* measurements of iron and manganese in hydrothermal plumes at mid-ocean ridges.

Fluorimetry, in principle, offers improved sensitivity, but rather few methods have been developed using this approach, with the notable exception of the lumogallion method for aluminum, which has been widely adopted with considerable success. Chemiluminescence methods have also been described for some metals, most notably for iron, a species of great current interest. Chemiluminescence methods have now been used on remotely operated vehicles operating in the deep sea. At ambient concentrations chemiluminescent measurement of iron requires a preconcentration step which can be setup inline for automatic operation.

### Inductively Coupled Plasma Methods and Electrothermal Atomic Absorption Spectrometry

A disadvantage of direct measurements of trace metals in seawater by atomic spectrometric methods is the very large interferences arising from the major ion matrix. The very high temperatures of plasma systems can act to reduce these interferences and, for barium at least, direct measurement of seawater concentrations by ICP-AES has been reported even at open-ocean concentrations. ICP-AES lacks the sensitivity to measure most other trace metals directly at oceanic concentrations, but may offer possibilities for contaminated estuarine waters. ICP-MS offers much greater sensitivity, though various interference problems limit the range of metals to which this technique can be applied. However, direct measurements of uranium and barium have now been reported using ICP-MS after dilution to minimize matrix effects. The speed and multielement capability of ICP-MS obviously offers real advantages coupled to the possibility of measuring isotopic ratios. Against these advantages must be set the disadvantages of cost, complexity, and the fact that the systems cannot be used at sea.

Direct measurements of several trace metals by electrothermal atomic absorption spectrometry (ETAAS) have been reported. In general, sensitivities are inadequate for open-ocean waters, though in more metal-enriched environments (e.g., coastal waters and sediment pore waters) such analysis is possible; careful corrections for the large and complex salt effects are necessary. The interferences can be minimized by the use of appropriate chemical modifiers, platforms in the graphite tubes, and sophisticated background correction schemes such as Zeeman.

### Preconcentration Procedures

Probably the most widely used procedure for trace metal analysis of seawater over the last 20 years has been preconcentration followed by ETAAS because of the wide availability, good sensitivity, and large range of elements that can be measured by this method. Preconcentration procedures have in general been one of three types. First, co-precipitation, with iron(III) 'hydroxide' or cobalt pyrrolidinedithiocarbamate, being the most widely used co-precipitants. Second, complexation followed by solvent extraction with a number of different complexants, of which dithiocarbamates, 8-quinolinol, and dithizone are particularly popular, and with extraction into a range of solvents. The third approach is extraction on to a chelating column (usually Chelex-100 but recently also other complexants such

as ethylenediaminetetraacetic acid or 8-quinolinol immobilized on a silica gel). These approaches all effect a substantial concentration (10–1000-fold), and separate the metals of interest from the bulk matrix, allowing measurement by ETAAS with minimal interferences. The extraction methods are generally efficient, but extraction is pH dependent and careful validation of a technique is always necessary prior to its use. Most methods will extract several metals. In some instances, the extraction techniques can be used to estimate the speciation of metals in seawater by analyzing samples directly without acidification and after acidification. The difference between the results of the two approaches yields a metal fraction only available after acidification has destroyed complexes. While ETAAS has proved a popular method for analysis of these metals, other techniques have been used for quantification including instrumental neutron activation, thermal ionization mass spectrometry, ICP-AES, and ICP-MS. The mass spectrometric techniques offer the advantage of measuring different isotopes of the metal, a property that has been effectively exploited for lead. The ICP techniques offer the option of simultaneous multi-element analysis of a sample with real advantages in rates of sample analysis. Measurement of iron in seawater by isotope dilution ICP-MS with co-precipitation preconcentration has recently been reported in which the coprecipitant is magnesium hydroxide from seawater itself, offering a simple, low blank procedure.

For some metals including arsenic, selenium, and tin, hydride-generation techniques coupled to atomic absorption instrumentation offer sensitive methods with minimum interference from seawater. Similarly for mercury, cold vapor techniques, usually coupled to gold amalgam as a preconcentration stage, offer suitable sensitivity for measurements in seawater. For all of these metalloids strong organic complexes can be formed in seawater, of which methylmercury and tributyltin complexes have received particular attention because of their high toxicity and tendency to bioaccumulation. While these organic complexes can be estimated by measurements before and after destruction of the organic complexes, such difference techniques are of very limited application if the organic complex forms only a small component of the total element pool. Direct methods of measuring both methylmercury and tributyltin have therefore been developed based on selective volatilization and detection of the organic species.

One last group of techniques that have been developed for trace metals involves complexation of the metal by compounds containing halogens, with subsequent extraction and analysis of these

complexes by electron-capture gas chromatography. The excellent sensitivity of this detection system makes these methods suitable for determination of several trace metals in seawater, limited only by the availability of suitable complexing agents that are element specific or can be separated out by chromatography. These methods can be made species specific, for example, allowing selenium(IV) and (VI) species to be analyzed separately. A further advantage is that gas chromatography is usually sufficiently robust to be used at sea.

**See also:** **Atomic Absorption Spectrometry:** Electrothermal; Vapor Generation. **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Chemiluminescence:** Liquid-Phase. **Fluorescence:** Quantitative Analysis. **Ion Exchange:** Ion Chromatography Applications. **Nitrogen.** **Quality Assurance:** Reference Materials. **Sampling:** Theory; Practice. **Spectrophotometry:** Inorganic Compounds. **Voltammetry:** Anodic Stripping; Cathodic Stripping.

## Further Reading

- Berman SS, Sturgeon RE, Desaulniers JAH, and Mykyfluk AP (1983) Preparation of the sea water references material, NASS-1. *Marine Pollution Bulletin* 14: 69–73.
- Danielsson L-Q, Magnusson B, Westerlund S, and Zhang V (1982) Trace metal determinations in estuarine waters by electrothermal atomic absorption spectrometry after extraction of dithiocarbonate complexes into freon. *Analytica Chimica Acta* 144: 183–188.
- Dehairs F, Neybergh H, and Hoenig M (1989) Direct determination of dissolved barium in seawater by inductively coupled plasma atomic emission spectrometry. *Analytica Chimica Acta* 222: 55–61.
- Howard AG and Statham PJ (1993) *Inorganic Trace Analysis, Philosophy and Practice*. Chichester: Wiley.
- Hydes DJ and Liss PS (1976) Fluorimetric method for the determination of low concentrations of dissolved aluminium in natural waters. *Analyst* 101: 922–931.
- Jones RD (1991) An improved fluorescence method for the determination of nanomolar concentrations of ammonium in natural waters. *Limnology and Oceanography* 36: 814–819.
- Kirkwood D, Aminot A, and Perttola M (1990) ICES (fourth) intercomparison exercise for nutrients in seawater. Marine Chemistry Working Group. *ICES Cooperative Research Report No. 174*. Copenhagen: International Council for the Exploration of the Sea.
- Matthias CL, Bellama JM, Olson GJ, and Brinckman FE (1986) Comprehensive method for the determination of aquatic butyl tin and butylmethyl tin species at ultratrace levels using simultaneous hybridization/extraction with gas chromatography-flame photometric detection. *Environmental Science and Technology* 20: 609–615.
- Measures CI, Grant B, Khadem M, Lee DS, and Edmond JM (1984) Distribution of Be, Al, Se and Bi in the surface

- waters of the western North Atlantic and Caribbean. *Earth and Planetary Science Letters* 71: 1–12.
- Obata H, Karatani H, and Nakayama E (1993) Automated determination of iron in seawater by chelating resin concentration and chemiluminescence detection. *Analytical Chemistry* 65: 1524–1528.
- Okamura K, *et al.* (2001) Development of a deep-sea *in situ* Mn analyser and application for hydrothermal plume observations. *Marine Chemistry* 76: 17–26.
- Parsons TR, Maita Y, and Lalli CM (1984) *A Manual of Chemical and Biological Methods for Seawater Analysis*. London: Pergamon.
- Robinson C and Williams PJ, leB, (1991) Development and assessment of an analytical system for the accurate and continual measurements of total dissolved inorganic carbon. *Marine Chemistry* 34: 157–175.
- Sharp JH, *et al.* (2002) A preliminary methods comparison for measurements of dissolved organic nitrogen in seawater. *Marine Chemistry* 78: 171–184.
- Sturgeon RE, Berman SS, Desaulniers JAH, *et al.* (1980) Comparison of methods for the determination of trace elements in seawater. *Analytical Chemistry* 52: 1585–1588.
- van den Berg CMG (1989) Adsorptive cathodic stripping voltometry of trace elements. *Analyst* 114: 1527–1530.
- Vink S, Boyle EA, Measures CI, and Yuan J (2000) Automated high resolution determination of the trace elements iron and aluminium in the surface ocean using a towed Fish coupled to flow injection analysis. *Deep-Sea Research I* 47: 1141–1156.
- Wu JF, Sunda W, Boyle EA, and Karl DM (2000) Phosphate depletion in the western North Atlantic Ocean. *Science* 289: 759–762.
- Zhang JZ and Chi J (2002) Automated analysis of nanomolar concentrations of phosphate in natural waters with liquid waveguide. *Environmental Science and Technology* 36: 1048–1053.

## Industrial Effluents

R Guerra, University of Bologna, Ravenna, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Industrial effluents result from various types of industrial processes and disposal practices, and may contain pollutants at levels that could affect the quality of receiving waters, as well as the aquatic ecosystem (Table 1). The emission of industrial pollutants in liquid effluents has to comply with stringent regulatory requirements and guidelines, in which chemicals listed should not exceed a given concentration. On the other hand, a chemical company may release a large number of different chemicals, which are not considered by regulatory requirements

and in many cases are unknown. These compounds may be the final products, precursors, or intermediates of the process, or impurities and by-products. One of the distinguishing characteristics of effluents of industrial origin, as compared to municipal wastewaters, is that often they may contain a mixture of different and very toxic substances. Approved analytical methods exist for compliance monitoring of conventional pollutants in industrial effluents; however, because of the complexity of the sample matrix, several analytical methods are required to determine polar and nonpolar organic compounds and new emerging pollutants that may impact water quality. As a consequence, modifications in instrumentation, sampling, and sample preparation techniques have become essential to comply with the regulatory water standards, as well as to achieve a faster speed of analysis.

**Table 1** General types of water pollutants

<i>Class of pollutant</i>	<i>Significance</i>
Heavy metals	Health, toxicity, aquatic biota
Organically bound metals and metalloids	Toxicity, aquatic biota
Inorganic species	Water quality, toxicity, aquatic biota
Trace organic pollutants	Toxicity
Polychlorinated biphenyls	Toxicity, aquatic biota, wildlife
Pesticides	Toxicity, aquatic biota, wildlife
Detergents	Toxicity, aquatic biota
Chemical carcinogens	Incidence of cancer

### General Chemical Analysis of Industrial Effluents

This section provides a brief overview of some of the most important tests that are performed on wastewaters.

#### pH, Hardness, Alkalinity, and Conductivity

General measures of the ionic characteristics of water are pH, hardness, alkalinity, and conductivity. There are several others that could be added to this group, such as redox potential and salinity; however, these



- waters of the western North Atlantic and Caribbean. *Earth and Planetary Science Letters* 71: 1–12.
- Obata H, Karatani H, and Nakayama E (1993) Automated determination of iron in seawater by chelating resin concentration and chemiluminescence detection. *Analytical Chemistry* 65: 1524–1528.
- Okamura K, *et al.* (2001) Development of a deep-sea *in situ* Mn analyser and application for hydrothermal plume observations. *Marine Chemistry* 76: 17–26.
- Parsons TR, Maita Y, and Lalli CM (1984) *A Manual of Chemical and Biological Methods for Seawater Analysis*. London: Pergamon.
- Robinson C and Williams PJ, leB, (1991) Development and assessment of an analytical system for the accurate and continual measurements of total dissolved inorganic carbon. *Marine Chemistry* 34: 157–175.
- Sharp JH, *et al.* (2002) A preliminary methods comparison for measurements of dissolved organic nitrogen in seawater. *Marine Chemistry* 78: 171–184.
- Sturgeon RE, Berman SS, Desaulniers JAH, *et al.* (1980) Comparison of methods for the determination of trace elements in seawater. *Analytical Chemistry* 52: 1585–1588.
- van den Berg CMG (1989) Adsorptive cathodic stripping voltometry of trace elements. *Analyst* 114: 1527–1530.
- Vink S, Boyle EA, Measures CI, and Yuan J (2000) Automated high resolution determination of the trace elements iron and aluminium in the surface ocean using a towed Fish coupled to flow injection analysis. *Deep-Sea Research I* 47: 1141–1156.
- Wu JF, Sunda W, Boyle EA, and Karl DM (2000) Phosphate depletion in the western North Atlantic Ocean. *Science* 289: 759–762.
- Zhang JZ and Chi J (2002) Automated analysis of nanomolar concentrations of phosphate in natural waters with liquid waveguide. *Environmental Science and Technology* 36: 1048–1053.

## Industrial Effluents

R Guerra, University of Bologna, Ravenna, Italy

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Industrial effluents result from various types of industrial processes and disposal practices, and may contain pollutants at levels that could affect the quality of receiving waters, as well as the aquatic ecosystem (Table 1). The emission of industrial pollutants in liquid effluents has to comply with stringent regulatory requirements and guidelines, in which chemicals listed should not exceed a given concentration. On the other hand, a chemical company may release a large number of different chemicals, which are not considered by regulatory requirements

and in many cases are unknown. These compounds may be the final products, precursors, or intermediates of the process, or impurities and by-products. One of the distinguishing characteristics of effluents of industrial origin, as compared to municipal wastewaters, is that often they may contain a mixture of different and very toxic substances. Approved analytical methods exist for compliance monitoring of conventional pollutants in industrial effluents; however, because of the complexity of the sample matrix, several analytical methods are required to determine polar and nonpolar organic compounds and new emerging pollutants that may impact water quality. As a consequence, modifications in instrumentation, sampling, and sample preparation techniques have become essential to comply with the regulatory water standards, as well as to achieve a faster speed of analysis.

**Table 1** General types of water pollutants

<i>Class of pollutant</i>	<i>Significance</i>
Heavy metals	Health, toxicity, aquatic biota
Organically bound metals and metalloids	Toxicity, aquatic biota
Inorganic species	Water quality, toxicity, aquatic biota
Trace organic pollutants	Toxicity
Polychlorinated biphenyls	Toxicity, aquatic biota, wildlife
Pesticides	Toxicity, aquatic biota, wildlife
Detergents	Toxicity, aquatic biota
Chemical carcinogens	Incidence of cancer

### General Chemical Analysis of Industrial Effluents

This section provides a brief overview of some of the most important tests that are performed on wastewaters.

#### pH, Hardness, Alkalinity, and Conductivity

General measures of the ionic characteristics of water are pH, hardness, alkalinity, and conductivity. There are several others that could be added to this group, such as redox potential and salinity; however, these

four are the most important. Generally, pH is measured using a pH meter consisting of a potentiometer, a glass electrode, a reference electrode, and a temperature-compensating device. A circuit is completed through the potentiometer when the electrodes are submersed. For alkalinity the samples are titrated electrometrically with hydrochloric acid or sulfuric acid to an endpoint of pH 4.5, and specific conductance is measured by use of a self-contained conductivity meter.

### **Total Organic Carbon**

Total organic carbon (TOC) analysis is based on the principle that carbon in an aqueous sample is converted to carbon dioxide by catalytic combustion or wet chemical oxidation. The carbon dioxide formed is measured directly by an infrared detector or converted to methane and measured by a flame ionization detector. The amount of carbon dioxide or methane is directly proportional to the concentration of carbonaceous material in the sample. Inorganic carbon is removed by acidification of the sample prior to analysis.

### **Chemical Oxygen Demand**

The chemical oxygen demand (COD) test gives a measure of the total organic content in terms of oxygen by oxidizing all biodegradable and nonbiodegradable organic materials with a strong oxidizing agent such as potassium dichromate. Silver sulfate is used as a catalyst, and mercury(II) sulfate is added to remove chloride interference. Absorbance is measured spectrophotometrically at 600 nm.

### **Biological Oxygen Demand**

The rapid and real-time monitoring of the gross amount of organic matter in municipal and industrial effluents is a major problem for every industry. Amongst all the parameters for which wastewaters are monitored, the organic biodegradable carbon is one of the most important and more frequently used parameters for estimating the level of wastewater pollution. The biological oxygen demand (BOD) test is a bioassay procedure that measures the amount of oxygen consumed by bacteria and other microorganisms from the decomposition of organic matter over a given period of time at a specified temperature. In the classical 5-day BOD test, the sample of wastewater is diluted with well-oxygenated and nutrient-containing water, and microorganisms adapted to the wastewater are introduced. The initial dissolved oxygen concentration is determined, and the sample is stored in darkness at 20°C for 5 days. The difference in oxygen concentration between the beginning and end

of the test period gives the 5-day BOD value. The need for fast, portable, and cost-effective methods for environmental monitoring has stimulated the production of a variety of field analytical tools such as biosensors, which shortened the BOD time analysis from 3–5 days to 5 min.

### **Oil and Grease**

Oil and grease refer to a broad class of organic substances that are recovered from effluent matrices by extraction with a nonpolar solvent under acidic conditions. Oil and grease compounds are substances such as hydrocarbons, vegetable oils, animal fats, waxes, soaps, and greases. Many solvents can dissolve other substances (e.g., sulfur compounds, chlorophyll, etc.). Therefore, oil and grease is operationally defined by the solvent used and the analytical method used to perform the analysis. There are two basic methods used to analyze oil and grease: the gravimetric technique and the infrared method.

### **Ammonia**

Monitoring the nitrogen content is of great importance in many industries for many reasons. In waters and wastewaters the most interesting forms of nitrogen, in order of decreasing oxidation state, are nitrate, nitrite, ammonia, and organic nitrogen. The occurrence of ammonia in wastewaters is primarily attributed to its formation resulting from the reduction of nitrogen-containing organics, deamination of amines, and hydrolysis of urea, and also to its use in water treatment plants for dechlorination. Ammonia may be analyzed after distillation of the wastewater sample over one of the following solutions:

- Boric acid – colorimetric Nesslerization method.
- Boric acid-indicator solution – titrimetric analysis.
- Sulfuric acid – colorimetric phenate method or ammonia electrode method.

The colorimetric Nesslerization method uses alkaline phenol and hypochlorite to react with ammonia to form indophenol blue in an amount proportional to the ammonia concentration. The blue color is intensified with sodium nitroprusside, and the concentration is measured using a calibrated colorimeter. The titration of basic ammonia is accomplished with standard sulfuric acid using a mixed indicator. Determination of ammonia may also be based upon the indophenol reaction adapted to automated gas-segmented continuous flow analysis. Potentiometric determination of ammonia is performed by ion-selective ammonia electrodes.



## Anions

Determination of inorganic ions (nitrate, nitrite, bromide, chloride, fluoride, phosphate, and sulfate) in surface water, ground water, drinking water, and domestic and industrial wastewaters can be determined by ion chromatography. The samples are introduced into an ion chromatograph, where the anions of interest are separated and measured using a system comprising a guard column, an analytical column, a suppressor device, and a conductivity detector. Potentiometric determinations of bromide, chloride, fluoride, and sulfide may be accomplished with ion-selective electrodes. Several methods are available for nitrate and nitrite determinations in wastewaters, including direct ultraviolet spectrophotometry, derivatization with benzoic acid, use of nitrate-selective electrodes, and reduction to nitrite followed by derivatization with sulfanilamide and *N*-1-naphthylethylenediamine dihydrochloride.

## Cyanide

Cyanides are soluble metal salts containing one metal ion, or insoluble complexes containing two metal ions, and the cyanide ion ( $\text{CN}^-$ ). The degree and rate of dissociation of complex cyanides to  $\text{CN}^-$  and HCN depend on several factors, including the nature of the metal, pH of the solution, and dilution. Cyanide in industrial wastewaters may be determined by the following methods:

- colorimetric method;
- silver nitrate titrimetric method;
- ion chromatography; and
- flow injection, ligand exchange, and amperometric detection.

Any cyanide present in the sample is converted to hydrocyanic acid by means of a reflux-distillation operation under acidic conditions, and absorbed in a scrubber containing sodium hydroxide solution. In the colorimetric method, the cyanide is released from cyanide complexes by means of a manual reflux-distillation operation and absorbed in a scrubber containing sodium hydroxide solution. The cyanide ion in the absorbing solution is converted to cyanogen chloride by reactions with chloramine-T, which subsequently reacts with pyridine and barbituric acid to give a red-colored complex. Alternatively, cyanide can be determined by titration using silver nitrate ( $\text{AgNO}_3$ ) with *p*-dimethylaminobenzalrhodanine indicator. Fast detection of cyanide in water and wastewaters may be also accomplished using a flow-injection analysis system, pretreated with ligand-exchange reagents. Cyanide is also quantitated by

means of silver working electrode, silver/silver chloride reference electrode, and platinum/stainless steel counter electrode.

## Total Phenols

The total phenolic compounds (*ortho*- and *meta*-substituted phenols) in an aqueous sample can be determined by a colorimetric method using 4-aminoantipyrine. This reagent reacts with phenolic compounds at pH 8 in the presence of potassium hexacyanoferrate(III) to form a colored antipyrine dye, the absorbance of which is measured at 500 nm. The antipyrine dye may also be extracted from the aqueous solution by chloroform, and the absorbance of the chloroform extract is measured at 460 nm.

## Analysis of Metals in Industrial Effluents

Although many metals are toxic, only some metals including aluminum, antimony, arsenic, beryllium, cadmium, chromium, copper, lead, mercury, nickel, selenium, silver, and zinc are considered as priority pollutants in industrial effluents. Metals in general can be analyzed in industrial effluents by the following techniques:

- spectrophotometry;
- atomic absorption spectrometry; and
- atomic emission spectrometry.

Colorimetric techniques are employed to analyze aluminum, arsenic, beryllium, cadmium, chromium(VI) and total, copper, lead, and zinc. Some metals may also be determined by other methods, including gravimetry, redox titration, ion chromatography, ion-selective electrodes, anodic stripping voltammetry (ASV), differential pulse polarography, etc. Mercury can be determined by oxidation, purge and trap, desorption, and cold-vapor atomic fluorescence spectrometry (CVAFS). Mercury is oxidized to  $\text{Hg(II)}$ , and then reduced to the elemental state and purged with nitrogen onto a gold-coated sand trap. The mercury vapor passes through a cell of a cold-vapor atomic fluorescence spectrometer for detection. Determinations of dissolved organic and inorganic divalent mercury ions and compounds can also be accomplished by ASV.

In atomic absorption spectroscopy, a sample is aspirated and atomized in a flame. A light beam from a hollow cathode lamp or an electrodeless discharge lamp is directed through the flame into a monochromator, and onto a detector that measures the amount of absorbed light. Absorption depends upon

the presence of free unexcited ground-state atoms in the flame. Because the wavelength of the light beam is characteristic of only the metal being determined, the light energy absorbed by the flame is a measure of the concentration of that metal in the sample. An atomic absorption spectrometer equipped with a graphite furnace or an electrically heated atomizer instead of the standard burner head gives better sensitivity and much lower detection limit than what is obtained with the flame technique.

Metals can be conveniently determined by inductively coupled plasma-atomic emission spectrometry (ICP-AES). A great advantage of ICP emission spectroscopy as applied to environmental analysis is that several metals can be determined simultaneously by this method. Another advantage is that, unlike atomic absorption spectroscopy, the chemical interferences in this method are very small. Chemical interferences are generally attributed to the formation of molecular compounds (from the atoms) as well as to ionization and thermochemical effects. The application of this technique has been selected for the analysis of toxic elements in drinking, ground, surface, and wastewater, or sludge and leachate from wastes. With ICP-AES, these samples are nebulized and the resulting aerosol transferred to a plasma torch, where it is decomposed, atomized, and ionized whereby the atoms and ions are excited. Each element emits light at characteristic wavelengths and these lines can be used for quantitative analysis after calibration. More recently, a new and more expensive technique, inductively coupled plasma-mass spectrometry (ICP-MS), has been introduced as a routine tool. ICP-MS offers the advantages of ICP-AES and the detection limit advantages of graphite furnace-atomic absorption spectrometry.

## Analysis of Organic Compounds in Industrial Effluents

In recent years, increasing attention has been paid to the detection, identification, and quantitation of trace organics and new emerging pollutants in wastewater samples. Analytical techniques for measuring organic compounds in industrial effluents require four general steps:

1. isolation (extraction and separation) of the target chemicals from sample matrix (drinking water, surface water, wastewaters, etc.);
2. separation and purification of the target chemical from co-extracted, nontarget chemicals (sample cleanup);
3. sample concentration; and

4. identification and quantitation by highly selective and sensitive analytical equipment.

Chromatography is the most widely used technique for the analysis of organic compounds in wastewaters. Gas chromatography (GC) (where the sample must be volatilized) and liquid chromatography (LC) (where the sample can be determined in the liquid state) are the most common. LC is the technique of choice whenever the sample cannot easily be converted to the gas phase. A mass spectrometer should be used wherever possible to identify the compounds more correctly. Although it has a lower sensitivity than other GC detectors, MS is, by far, the best confirmatory test for compound identification.

### Selection of Extraction Methods

Chemical substances and their solubility in water depend on their physicochemical characteristics, such as their molecular weight, boiling point, and polarity (Figure 1). Methods for isolating chemicals from water based on these broad chemical categories are shown in Figure 2.

The two main methods for the isolation of volatile organic compounds (VOCs) from water are:

1. The purge and trap (P&T) method, which separates volatile compounds from the aqueous matrix by passing an inert gas such as helium or nitrogen through the matrix (purging). The target, volatile compounds are desorbed from the aqueous phase to the gas phase (purged) and are then separated from the stream of gas by trapping (adsorbent) filters (e.g., porous polymer beads, activated charcoal, silica gel, etc.). In general, the P&T technique is applied to analyze substances that have boiling points below 200°C and are insoluble or slightly soluble in water.
2. The headspace method, less sensitive compared to the P&T method, in which the sample is placed in a sealed container, such as a vial, and left at a constant temperature until the gas and liquid phases are in equilibrium. The target substances in the gas phase (headspace) are collected by gas tight syringe, and injected into the gas chromatograph-mass spectrometer.

Methods to extract semivolatile compounds (SVOCs) from water include the following.

1. Liquid-liquid extraction (LLE) is the most common method for the isolation of SVOCs and nonvolatile compounds from water samples. It is based on the partitioning of analytes from an aqueous phase to a liquid organic phase. In LLE,

Volatility $\longrightarrow$																						
	<table><tr><td></td><td><b>Volatile</b></td><td><b>Semivolatile</b></td><td><b>Nonvolatile</b></td></tr><tr><td rowspan="3"><b>Polarity</b> <math>\uparrow</math></td><td><b>Polar</b></td><td>Alcohols, ketones, carboxylic acids etc</td><td>Alcohols, ketones, carboxylic acids</td><td>High molecular electrolytes, carbohydrates, fulvic acid</td></tr><tr><td><b>Semipolar</b></td><td>Ethers, esters, aldehydes</td><td>Ethers, esters, aldehydes, epoxides, heterocyclic compounds</td><td>Protein, carbohydrates, humic acids</td></tr><tr><td><b>Nonpolar</b></td><td>Aliphatic compounds, aromatic compounds</td><td>Aliphatic compounds, aromatic compounds, alicyclic compounds</td><td>Nonionic polymers, lignin</td></tr><tr><td></td><td><b>Small</b></td><td><b>Medium</b></td><td><b>Large</b></td></tr></table>		<b>Volatile</b>	<b>Semivolatile</b>	<b>Nonvolatile</b>	<b>Polarity</b> $\uparrow$	<b>Polar</b>	Alcohols, ketones, carboxylic acids etc	Alcohols, ketones, carboxylic acids	High molecular electrolytes, carbohydrates, fulvic acid	<b>Semipolar</b>	Ethers, esters, aldehydes	Ethers, esters, aldehydes, epoxides, heterocyclic compounds	Protein, carbohydrates, humic acids	<b>Nonpolar</b>	Aliphatic compounds, aromatic compounds	Aliphatic compounds, aromatic compounds, alicyclic compounds	Nonionic polymers, lignin		<b>Small</b>	<b>Medium</b>	<b>Large</b>
	<b>Volatile</b>	<b>Semivolatile</b>	<b>Nonvolatile</b>																			
<b>Polarity</b> $\uparrow$	<b>Polar</b>	Alcohols, ketones, carboxylic acids etc	Alcohols, ketones, carboxylic acids	High molecular electrolytes, carbohydrates, fulvic acid																		
	<b>Semipolar</b>	Ethers, esters, aldehydes	Ethers, esters, aldehydes, epoxides, heterocyclic compounds	Protein, carbohydrates, humic acids																		
	<b>Nonpolar</b>	Aliphatic compounds, aromatic compounds	Aliphatic compounds, aromatic compounds, alicyclic compounds	Nonionic polymers, lignin																		
	<b>Small</b>	<b>Medium</b>	<b>Large</b>																			
<b>Molecular weight</b> $\longrightarrow$																						

**Figure 1** Classification of organic compounds by physicochemical properties.

		Volatility $\longrightarrow$			
Polarity $\uparrow$	<b>Polar</b>	None	Derivatization		Reduced pressure distillation
		Liquid–solid-phase extraction	Liquid–liquid extraction	Liquid–solid extraction	
	<b>Semipolar</b>	Distillation	pH adjustment		Freeze dry
			Liquid–liquid extraction	Liquid–solid extraction	Reversed permeation
	<b>Nonpolar</b>	Purge and trap			Ultrafiltration
		Headspace	Liquid–liquid extraction	Liquid–solid extraction	
		<b>Small</b>	<b>Medium</b>	<b>Large</b>	
		Molecular weight $\longrightarrow$			

**Figure 2** Chemical separation methods for water samples.

hydrophobic analytes are extracted into an organic phase. The water sample and the solvent are placed in a separatory funnel, shaker flask, or bottle, which is shaken or continuously rotated to increase the contact surface area between the solvent and the sample.

2. Solid-phase extraction (SPE), where water sample analytes are passed through and adsorbed onto a

solid support, and then extracted with a solvent or by thermal desorption. The most popular and rapidly emerging method employs commercially available packed columns, rigid disks, speedisks, or extraction disk cartridges, filled with various phases bonded to a silica gel matrix.

3. Solid-phase microextraction (SPME) is a further modification of the SPE technique. Analytes are

isolated from water by passive isolation. SPME provides a solvent-free approach to SPE based on equilibrium sampling. The extraction medium is a fused silica fiber (coated with polymer phase and bonded to a stainless steel plunger) and a holder that has a form of microliter syringe.

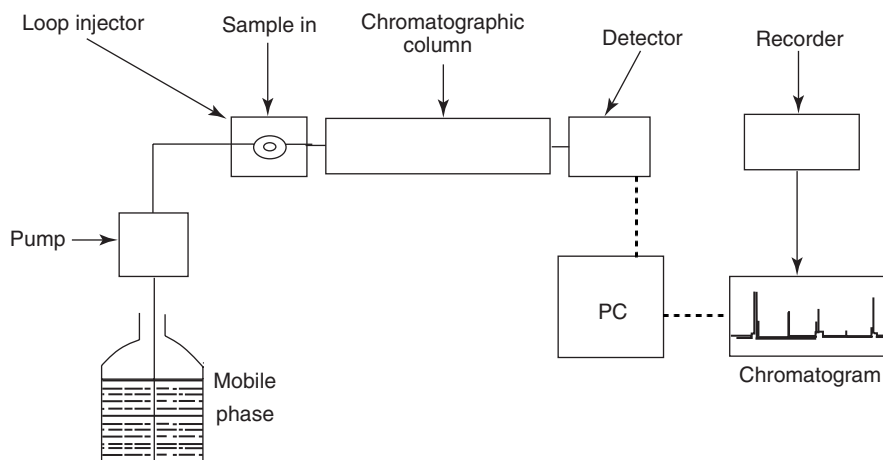
Extracts from industrial effluent samples can be complicated mixtures, with components interfering with GC and LC analysis by giving poor separation because of over separation capacity of a column or by containing compounds that elute at the same time, and interfere with, the target compounds' peaks. The sample extracts may be purified by one or more of the following techniques for different classes of organic compounds.

1. Adsorption chromatography (alumina, silica gel, Florisil): separates analytes of a relatively narrow polarity range away from extraneous, interfering compounds of a different polarity (e.g., organochlorine pesticides, polycyclic aromatic hydrocarbons).
2. Acid-base partitioning: separates acidic or basic organics from neutral organics (e.g., separates chlorophenoxy herbicides from phenols).
3. Gel permeation chromatography: separates high molecular weight, high boiling material from the sample analytes; prevents contamination of GC injection heads and columns (e.g., semivolatile organics and pesticides).
4. Sulfur cleanup: eliminates sulfur from extracts, which may cause chromatographic interference with analytes.
5. Sulfuric acid/permanganate: destroys most organic compounds except polychlorinated biphenyls (PCBs).

## Liquid Chromatography

LC is a sensitive, adaptable, and quantitative analytical technique to separate compounds that are dissolved in solution. **Figure 3** depicts the main components of a modern LC system: (1) a pump with a constant flow control, (2) a high-pressure injection valve, (3) a chromatographic column, (4) a detector, and (5) a data system for measuring peak areas and retention times (**Figure 3**).

LC basically involves the separation of analytes in a mixture by adsorption to a polar or nonpolar support surface or by partition into a stationary liquid phase. These analytes are first dissolved in mobile phase (eluant), and then forced to flow through a chromatographic column (stationary phase) under high pressure. The interaction of the solute with mobile and stationary phases can be manipulated through different choices of both solvents and stationary phases. The stationary phase is defined as the immobile packing material in the column, and which consists of the rigid porous particles, usually silica based, with the specific surface properties. In normal-phase liquid chromatography, a bonded polar surface such as cyano-, diol-, or amino group bound to silica is employed, while the mobile phase is nonpolar. On the other hand, reversed-phase liquid chromatography uses a nonpolar surface such as octadecyl-, octyl-, methyl-, and diphenyl groups bound to silica or polymers, and a polar mobile phase. Water-acetonitrile and water-methanol mixtures are commonly used as polar mobile phase. The most common LC detectors are based on refractive index, ultraviolet/visible, fixed and variable wavelength, diode array, and fluorescence; and the less common but important ones are the conductivity and the evaporative light scattering detectors.



**Figure 3** Diagram of a typical LC setup.

LC has been extensively used in municipal and industrial wastewaters analysis for phenolic compounds, benzidines, diquat, and paraquat, surfactants like alkylphenol and alcohol ethoxylates, agents employed in consumer detergent products like linear alkylbenzene sulfonates (LAS), and other aromatic sulfonates. Recent analytical advances have also been made for priority endocrine disrupting compounds by means LC-MS with the electrospray interface and atmospheric pressure chemical ionization coupled with MS. In the last decade, the significant progress in the development of LC-MS interfacing systems has made unforeseen analytical capabilities available for a wide variety of water pollutants such as LAS, nitro- and chlorophenols, alkylphenol and alcohol ethoxylates, toxins generated during blooms of blue-green algae, cationic surfactants, and antibacterial agent such as dimethylammonium or benzyl alkyl dimethylammonium, organometallic compounds like tributyltin and triphenyltin and related compounds, benzidines and nitrogen-containing pesticides, organophosphorus compounds, chlorinated phenoxyacid compounds and their esters, carbamates, etc.

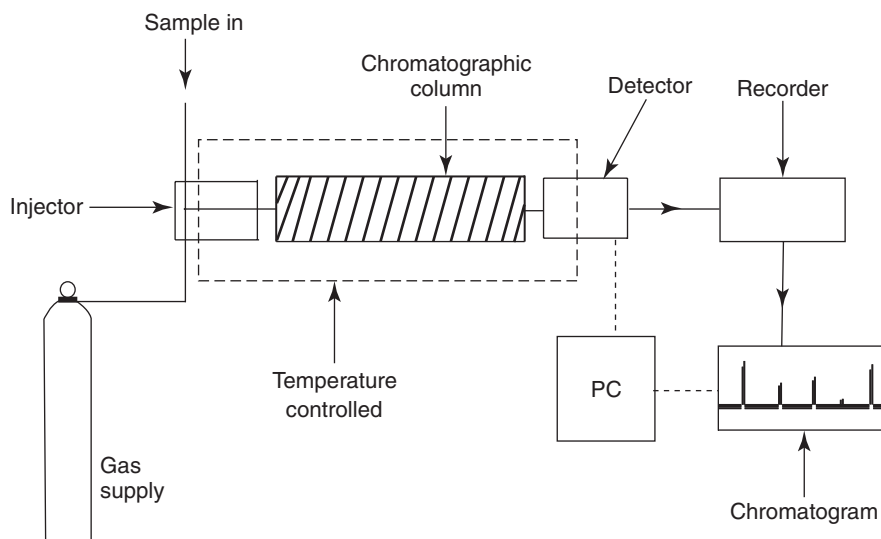
### Gas Chromatography

GC is the most widely used instrumental technique today for quantitative organic chemical analysis for its unparalleled resolving power and sensitivity of detection, combined with relative simplicity of operation. The concept is based on the selective interaction of gaseous or semivolatile compounds (analytes) with the column packing or film as the vaporized

sample is forced through the column by a chemically inert carrier gas (usually nitrogen or helium). Good separation in the shortest period of time is ideal for good GC analysis, and these criteria are related to the liquid phase of columns, film thickness, length, internal diameter, and temperature. Capillary columns are available with many kinds of liquid phases, from nonpolar methylsilicones to highly polar polyethylene glycols. The basic components of a GC system are: (1) a carrier gas cylinder with regulator, (2) a flow controller for the gas, (3) an injection port for introducing the sample, (4) the column, and (5) the detector and the recorder (Figure 4). For GC a number of detectors have been developed; however, the most important for trace organic analysis, besides the mass spectrometer, are:

- flame ionization detector (only detects compounds containing the C-H bond);
- electron capture detector (halogenated hydrocarbons, halogenated pesticide residues, and oxygen-containing groups or nitro groups); and
- nitrogen/phosphorus detector (phosphorus- and nitrogen-containing pesticides, organophosphates, and carbamates).

GC plays a key role in the analysis of VOCs and SVOCs in industrial effluents. VOCs that have boiling points below 200°C can be extracted from water sample matrices using a P&T system. Purged sample components are trapped in a tube containing suitable sorbent materials. When purging is complete, the sorbent tube is heated and backflushed with helium to desorb trapped sample components onto a GC



**Figure 4** Diagram of a typical GC setup.



column. The column separates the analytes, which are identified and quantified using a mass spectrometer. Volatile, water-soluble compounds can be included in this analytical technique by the use of azeotropic distillation or closed-system vacuum distillation. Such compounds include low molecular weight halogenated hydrocarbons, aromatics, ketones, nitriles, acetates, acrylates, ethers, and sulfides. Nonhalogenated VOCs in aqueous samples are introduced into the gas chromatograph using the P&T system, and detection is achieved by a flame ionization detector.

Among semivolatile compounds, organochlorine pesticides are analyzed by GC using single/double fused-silica column (open-tubular and/or capillary columns) with an electron capture detector or an electrolytic conductivity detector. Depending on the nature of interferences and target analytes, a combination of cleanup procedures including alumina, Florisil, silica gel, gel permeation chromatography, and sulfur may be applied prior to instrumental analysis.

PCBs congeners and Aroclor mixtures are also determined either by a single/dual column analysis system with electron capture detectors. Compound identification based on single-column analysis should be confirmed on a second column, or should be supported by a confirmation technique such as GC/MS. Extracts for PCB analysis may undergo a sulfuric acid/potassium permanganate cleanup to remove coeluting organochlorine or organophosphorus pesticides and sulfur cleanup to remove chromatographic interferences due to elemental sulfur.

Herbicides may be determined by specific extraction, esterification, and gas chromatographic conditions. Aqueous samples are extracted with diethyl ether and then esterified with either diazomethane or pentafluorobenzyl bromide. The derivatives are determined by GC with an electron capture detector. Compound identifications should be supported by GC-MS for the qualitative confirmation. Alkaline hydrolysis and subsequent solvent wash removes many chlorinated hydrocarbons and phthalates that might otherwise interfere with the electron capture analysis.

### Gas Chromatography-Mass Spectrometry

GC-MS is the coupling of two analytical techniques into a versatile and probably the best technique to identify a wide array of organic substances in industrial effluents. GC-MS is based on chromatographic separation of the different constituents of the sample on a GC column, and their subsequent analysis and identification by MS. The major components of

the mass selective detector are: an ionization source, mass separator, and ion detector. The two common mass separators commercially available for GC-MS are the quadrupole and the ion trap. As a sample elutes from the GC column, it enters the ionization chamber of the mass spectrometer where the molecules are ionized, typically by electron impact. The mass analyzer scans stepwise through a set range of mass values to evaluate the relative abundance of ions at each mass value. The result is a total ion chromatogram, which is a plot of the total mass eluting from the GC and detected by MS as a function of time. Compounds are identified from the mass spectrum by a computerized comparison of the mass spectra for the sample with spectra library for known compounds.

SVOCs that are soluble in methylene chloride and capable of being eluted, without derivatization, as sharp peaks from a gas chromatographic fused-silica capillary column coated with a slightly polar silicone, can be analyzed by GC/MS. Such compounds include polycyclic aromatic hydrocarbons (PAHs), chlorinated hydrocarbons and pesticides, phthalate esters, organophosphate esters, nitrosamines, haloethers, aldehydes, ethers, ketones, anilines, pyridines, quinolines, nitroaromatic compounds, phenols, and nitrophenols. Detection and quantitative measurements of dioxins (PCDDs) and dibenzofurans (PCDFs) in industrial effluents require the use of high-resolution capillary column gas chromatography/low-resolution mass spectrometry (HRGC/LRMS) techniques. Sample extracts may undergo specified cleanup procedures including back-extraction with acid and/or base, and gel permeation, alumina, silica gel, Florisil, and activated carbon chromatography. LC can be used for further isolation of the 2,3,7,8-isomers or other specific isomers or congeners. Isotope dilution HRGC/HRMS may also be employed for PCDD/PCDFs analysis. This technique involves spiking the aqueous sample with stable isotopically labeled analogs prior to sample extraction.

VOCs can also be determined by isotope dilution GC/MS, where stable isotopically labeled analogs of the compounds of interest are available; otherwise, an internal standard method is used. After purging and trapping, the sample is introduced into a gas chromatograph, and the compounds are separated by the GC and detected by a mass spectrometer. The labeled compounds serve to correct the variability of the analytical technique. Identification of a compound (qualitative analysis) is performed by comparing the GC retention time and the background corrected characteristic spectral masses with those of authentic standards.



## Nomenclature

- 2,3,7,8-Isomers = 2,3,7,8-substituted penta-, hexa-, hepta-, and octachlorinated dibenzo-*p*-dioxins and dibenzofurans.
- 4-Aminoantipyrine ( $C_{11}H_{13}N_3O$ ); Synonyms: aminoantipyrine; ampyrone; 4-amino-1,2-dihydro-1,5-dimethyl-2-phenyl-3*H*-pyrazol-3-one; 4-aminophenazone; 4-amino-2,3-dimethyl-1-phenyl-3-pyrazolin-5-one; 4-amino-1,5-dimethyl-2-phenyl-4-pyrazolin-3-one; 3*H*-pyrazol-3-one; 4-amino-1,2-dihydro-1,5-dimethyl-2-phenyl.
- Acetates = (1) Metallic salts derived from acetic acid by interaction of the metallic oxide, hydroxide, or carbonate with the acid; or the esters derived by interaction of alcohols with acetic acid, which include the common esters of ethyl, propyl, isopropyl, butyl, and amyl acetates. (2) A generic name for cellulose acetate plastics, particularly for fibers thereof. When at least 92% of the hydroxyl groups are acetylated, the term triacetate may be used as the generic name of the fiber. (3) A compound containing the acetate group,  $CH_3COO-$ .
- Acrylates = (def) Different acrylic, styrene acrylic, vinyl acrylic products for applications ranging from cement additives to filters (e.g., acrylic acid esters, metallic salts of acrylic acid).
- Alcohol ethoxylates; (C6–C18) Ethoxylates; Synonyms: ethoxylated alcohol, fatty alcohol ethoxylate, alcohol ethoxylate, polyoxyethylene alcohol ether, polyethylene glycol alcohol ether. Surfactant alcohol ethoxylates and their derivatives are used by the detergent manufacturing industry as active ingredients in industrial and consumer products such as soaps, detergents, and other cleaning products.
- Alkylphenol ethoxylates (APEs) have been widely used as part of industrial processes and as detergents in both industrial and household applications for more than 30 years. Octylphenol ethoxylates and nonylphenol ethoxylates are two of the most common surfactants in the marketplace.
- Amines = Compounds formally derived from ammonia by replacing one, two, or three hydrogen atoms by hydrocarbyl groups, and having the general structures  $RNH_2$  (primary amines),  $R_2NH$  (secondary amines),  $R_3N$  (tertiary amines).
- Anilines = Includes aniline and chloro-, bromo-, and/or nitroanilines (benzene amines).
- Nitroaromatic compounds = Nitrogen-containing organic compounds used worldwide as explosives, pesticides, and as precursors for the manufacture of many products, including dyes, pharmaceuticals, and plastics pathway (e.g., 2,4,6-trinitrotoluene (TNT), 1,3,5-trinitrobenzene (TNB), dinitrotoluene (2,4-DNT, 2,6-DNT), dinitrobenzene (DNB), methyl-*N*,2,4,6-tetranitroaniline, and 2,4,6-trinitrophenol).
- Barbituric acid ( $C_4H_4N_2O_3$ ); Synonym(s): 2,4,6 (1*H*,3*H*,5*H*)-Pyrimidinetrione.
- Benzidines = Benzidine and 3,3'-dichlorobenzidine.
- Carbamates = Insecticides that are derivatives of carbamic acid,  $HOC(O)NH_2$ ; widely used against insects, fungi, and weeds, under several commercial names (e.g., Alicarb, Carbaryl, Propoxur).
- Cationic surfactants = (def) Quaternary ammonium compounds that are major components in industrial and commercial formulations such as cosmetics, textile softeners, and pharmaceuticals (e.g., tetramethyl, tetrapropyl, tetrabutyl, dodecyl, cetyltrimethyl, and myristyltrimethylammonium).
- Chlorinated hydrocarbons = (def) (1) Chemicals containing only chlorine, carbon, and hydrogen. These include a class of persistent, broad-spectrum insecticides that linger in the environment and accumulate in the food chain. Among them are DDT, aldrin, dieldrin, heptachlor, chlordane, lindane, endrin, Mirex, hexachloride, and toxaphene. Other examples include TCE, used as an industrial solvent. (2) Any chlorinated organic compound including chlorinated solvents such as dichloromethane, trichloromethylene, chloroform.
- Chlorinated phenoxyacid compounds = One of the most important classes of chlorinated herbicides used to control agricultural and aquatic weeds (e.g., Dalapon, Dicamba, 2,4-D, MCPA, MCPP, 2,4,5-T, etc.). In these compounds, chloro-substituted benzene rings are attached to lower carboxylic acids via an oxygen atom.
- Dibenzofurans (PCDFs) = (def) Polychlorinated dibenzo-*para*-dioxins (PCDDs) and polychlorinated dibenzofurans (PCDFs) are two series of compounds that have a triple-ring structure that consists of two benzene rings connected by a third oxygenated ring. For PCDDs, the benzene rings are connected by a pair of oxygen atoms, whereas in PCDFs they are connected via a single oxygen atom. The number of chlorine atoms can vary between one and eight. PCDDs and PCDFs are not commercially produced but are formed as trace amounts of undesired impurities in the manufacture of other chemicals, such as chlorinated phenols and their derivatives, chlorinated diphenyl ethers, and polychlorinated biphenyls (PCBs). Ongoing or earlier use of pentachlorophenol is considered to be a major source of PCDDs and PCDFs in many industrialized countries. PCDDs and PCDFs are also formed in

combustion processes such as waste incineration and in the production of iron and steel.

- Dioxins (PCDDs) = Dioxin is a generic term used to denote a single compound or mixture of compounds derived from polychlorinated dibenzodioxins. The basic structure of all dioxins consists of two benzene rings joined by two oxygen atoms. There are 75 different polychlorinated dibenzo-*p*-dioxins and 135 dibenzofurans that are classified into groups termed homologs on the basis of the number of chlorine atoms in the molecule.
- Endocrine disrupting compounds = Exogenous agents that interfere with the synthesis, secretion, transport, binding, action, or elimination of natural hormones in the body that are responsible for the maintenance of homeostasis, reproduction, development, and/or behavior (e.g., alkylphenols, alkylphenol polyethoxylates, polyaromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), phthalates, bisphenol-A, polybrominated flame retardants, dioxins, furans, herbicides, pesticides, and steroid hormones).
- Haloethers = Compounds that contain an ether moiety (R–O–R) and halogen atoms attached to the aryl or alkyl groups (e.g., bis(2-chloroethyl) ether; bis(2-chloroethoxy)methane; bis(2-chloroisopropyl) ether; 4-bromophenyl phenyl ether; 4-chlorophenyl phenyl ether).
- Halogenated hydrocarbons = Hydrocarbons containing chlorine, bromine, iodine, fluorine (e.g., chloroform, trichloroethylene, tetrachloroethylene, polychlorinated biphenyls).
- Linear alkylbenzene sulfonates (LAS) = Anionic surfactants with molecules characterized by a hydrophobic (apolar) group and a hydrophilic (polar) group. As a class of chemicals, they are generally mixtures of closely related isomers and homologs. Each molecule contains an aromatic ring sulfonated at the *para* position and attached to either a linear or a branched alkyl chain at any position except the terminal carbons. Chain lengths vary but are predominantly in the range of C10–C14.
- Mixed indicator = Methyl red in ethyl alcohol with methylene blue in ethyl alcohol.
- Nitriles = Compounds having the structure  $RC\equiv N$ ; thus, C-substituted derivatives of hydrocyanic acid; the suffix nitrile denotes the triply bound  $\equiv N$  atom, not the carbon atom attached to it.
- Nitrosamines = Nitroso derivatives of amines in which a nitroso (NO) group is attached to the nitrogen atom of the amine (e.g., *N*-nitrosodimethylamine, *N*-nitrosodiphenylamine, *N*-nitrosodi-*n*-propylamine).
- Organochlorine pesticides = A unique class of pesticides because of their cyclic structure, number of chlorine atoms, and low volatility. They can be classified into four categories: dichlorophenylethanes (e.g., DDT), cyclodienes, chlorinated benzenes (e.g., hexachlorobenzene (HCB)), and cyclohexanes (e.g., hexachlorocyclohexane (HCH)).
- Organometallic compounds = Compounds having bonds between one or more metal atoms and one or more carbon atoms of an organyl group. Organometallic compounds are classified by prefixing the metal with organo group, e.g., organopalladium compounds.
- Organophosphate esters are used in a wide variety of applications including hydraulic fluids, plasticizers, and antiwear additives to hydraulic fluids and engine oils. Some are referred to as 'natural' phosphate esters because the cresols and xylenols used as raw materials are derived from petroleum oil or coal tar. The first commercial trialkyl phosphate esters (TAP) were tricresyl phosphate (TCP) and trixylenyl phosphate (TXP); others are known as 'synthetic' phosphate esters, which are derived from specific synthetic reactions to produce triaryl, trialkyl, and alkylaryl esters.
- Organophosphorus pesticides = Most organophosphorus pesticides are vinyl ester derivatives of phosphates. Four classes of organophosphorus pesticides are in use: phosphates ( $R_1, R_2 = O$ ), ( $R_1 = S, R_2 = O$ ), phosphorothioates ( $R_1 = S, R_2 = O$ ), phosphorothiolothionates ( $R_1, R_2 = S$ ), and phosphonates (with a P–C bond). Examples are Parathion, Diazinon, etc.
- Phthalate are the esters of phthalic acid used as plasticizers of synthetic polymers such as poly(vinyl chloride) and cellulose acetate.
- Polychlorinated biphenyls (PCB or PCBs) are a class of chlorosubstituted biphenyl compounds that were once widely used as additives in transformer oils, lubricating oils, and hydraulic fluids. They are formed by the addition of chlorine ( $Cl_2$ ) to biphenyl ( $C_{12}H_{10}$ ), which is a dual-ring structure comprising two 6-carbon benzene rings linked by a single carbon–carbon bond. The nature of an aromatic ring allows a single attachment to each carbon. This means that there are 10 possible positions for chlorine substitution (replacing the hydrogens in the original biphenyl). PCBs have been used as coolants and lubricants in transformers, capacitors, and other electrical equipment because they do not burn easily and are good insulators.
- Polycyclic aromatic hydrocarbons (PAHs) are a group of over 100 chemicals containing

two or more benzene rings fused together, which are formed during the incomplete burning of coal, oil and gas, garbage, or other organic substances like tobacco or charbroiled meat.

- Quinoline ( $C_9H_7N$ ); Synonyms: 1-azanaphthalene, 1-benzazine, benzo(*b*)pyridine. Quinoline is a hygroscopic, colorless liquid at room temperature, with a penetrating amine-like odor; is a constituent of creosote, coal tar, and certain other products derived from fossil fuels, and it is also produced by combustion of a number of substances including tobacco. It is used as a solvent and a decarboxylation reagent, and as a raw material for manufacture of dyes, antiseptics, fungicides, niacin, pharmaceuticals, and 8-hydroxyquinoline sulfate. Quinoline is a base that combines with strong acids to form salts, e.g., quinoline hydrochloride.
- Sulfanilamide ( $C_6H_8N_2O_2S$ ) = White, odorless crystalline sulfonamide, used in the treatment of various bacterial infections. Synonyms: Albexan; albosol; ambeside; *p*-aminobenzene sulfamide; sulfanilamide purified; bacteramid; 4-aminobenzenesulfonamide.
- Tributyltin (TBT) ( $(C_4H_9)_3Sn^+$ ) is an organotin compound used primarily as a biocide in anti-fouling paints. It is extremely toxic to aquatic life and is an endocrine-disrupting chemical that causes severe reproductive effects in aquatic organisms. Organotins are compounds consisting of one to four organic components attached to a tin atom via carbon–tin covalent bonds. When there are fewer than four carbon–tin bonds, the organotin cation can combine with an anion such as acetate, carbonate, chloride, fluoride, hydroxide, oxide, or sulfide.
- Triphenyltin (TPTH) ( $C_{18}H_{16}OSn$ ); Synonyms: triphenyltin hydroxide.

See also: **Atomic Absorption Spectrometry:** Flame. **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Derivatization of Analytes.** **Extraction:** Solvent Extraction Principles; Solid-Phase Microextraction; Solid-Phase Extraction. **Gas Chromatography:** Column Technology; Detectors; Mass Spectrometry. **Gravimetry.** **Herbicides.** **Ion-Selective Electrodes:** Water Applications. **Liquid Chromatography:** Column Technology; Mobile Phase Selection. **Polychlorinated Biphenyls.** **Polycyclic Aromatic Hydrocarbons:** Determination; Environmental Applications. **Spectrophotometry:** Inorganic Compounds. **Surfactants and Detergents.** **Titrimetry:** Potentiometric. **Voltammetry:** Inorganic Compounds. **Water Analysis:** Organic Compounds; Biochemical

Oxygen Demand; Chemical Oxygen Demand. **Water Determination.**

## Further Reading

- American Public Health Association (APHA) (1998) *Standard Methods for the Examination of Water and Wastewater*, 20th edn. Washington DC: American Public Health Association, American Waterworks Association, Water Environment Federation.
- Handley AJ (1999) *Extraction Methods in Organic Analysis Separations*, Sheffield Analytical Chemistry Series, vol. 2. Boca Raton, FL: CRC Press.
- Kiewiet AT and de Voogt E (1996) Chromatographic tools for analyzing and tracking non-ionic surfactants in the aquatic environment, review. *Journal of Chromatography A* 733: 185–192.
- Liška I (2000) Fifty years of solid-phase extraction in water analysis – historical development and overview, review. *Journal of Chromatography A* 885: 3–16.
- Patnaik P (1997) *Handbook of Environmental Analysis: Chemical Pollutants in Air, Water, Soil, and Solid Wastes*. Boca Raton, FL: Lewis Publishers.
- Rastogi S, Rathee P, Saxena TK, Mehra NK, and Kumar R (2003) BOD analysis of industrial effluents: 5 days to 5 min. *Current Applied Physics* 3: 191–194.
- Reemtsma T (1996) Methods of analysis of polar aromatic sulfonates from aquatic environments, review. *Journal of Chromatography A* 733: 473–489.
- Reemtsma T (2003) Liquid chromatography–mass spectrometry and strategies for trace-level analysis of polar organic pollutants, review. *Journal of Chromatography A* 1000: 477–501.
- Skoog DA, Holler FJ, and Nieman TA (1997) *Principles of Instrumental Analysis*, 5th edn. Pacific Grove, CA: Brooks/Cole.
- USEPA (1986) *Test Methods for Evaluating Solid Waste (SW-846): Physical/Chemical Methods*. Washington DC: US Environmental Protection Agency, Office of Solid Waste.
- USEPA (1995) *Methods for Organic Chemical Analysis of Municipal and Industrial Wastewaters*, EPA/600/4-82/057. Cincinnati, OH: US Environmental Protection Agency, Environmental Monitoring and Support Laboratory.
- USEPA (1997) *Methods for the Determination of Chemical Substances in Marine and Estuarine Environmental Matrices*, EPA/600/R-97/072. Washington, DC: US Environmental Protection Agency, Office of Research and Development.
- USEPA (1999) *Methods and Guidance for Analysis of Water*, CD ROM Version 2.0, EPA 821/C-99-004. Washington DC: US Environmental Protection Agency, Office of Water.
- Voulvoulis N and Scrimshaw MD (2003) Methods for the determination of endocrine disrupters. In: Birkett JW and Lester JN (eds.) *Endocrine Disrupters in Wastewater and Sludge Treatment Processes*, pp. 58–101. Boca Raton, FL: Lewis Publishers.

## Sewage

**C Bendicho and I Lavilla**, Universidad de Vigo, Vigo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Human activity and its associated impact cause important changes over natural water composition leading frequently to pollution of the natural water resources. The pollution risk involved makes necessary the establishment of different quality levels according to the various uses of water (human consumption, industrial and agricultural use, etc.). Different effluents can affect water quality. Domestic effluents usually have fairly uniform characteristics despite being of heterogeneous nature, whereas industrial effluents display a broader diversity due to both between-industry and within-industry variations.

Sewage can be defined as water that has been changed in its characteristics by domestic, industrial, or agriculture effluents.

Urban sewage originates mainly from domestic sources and also from rain, and to a lesser extent from public cleaning-up, watering, wastewaters from industry emptied into municipal collectors, etc. Whereas literature information on the characteristics of urban sewage is abundant, industrial sewage should be addressed in a particular way.

Discharging of sewage to rivers and coasts has been a common practice in the past, hence causing strong changes in biotic conditions. Natural degradation processes occurring in any environmental compartment are slow so that suitable treatment is required for removal, or at least, decreasing the initial loadings up to an acceptable level. Treatment addressed to purification of sewage effluents are carried out in sewage treatment plants following different guidelines (e.g., EPA guidelines in USA, EU directives in Europe).

Analytical measurement of sewage is required to establish the appropriate treatments as well as to provide a measure of treated effluent quality and for optimization of the treatment processes. Apart from sewage itself, solid residues originated in treatment plants require analytical characterization regardless their final purpose (i.e., use as a resource or discharged).

The elimination and fate of sewage sludge originating in sewage treatment plants is assuming great importance. There are two major kinds of sludge generated in a waste treatment plant. The first

of these is organic sludge from activated sludge, trickling filter, or rotating biological reactors. The second is inorganic sludge from the addition of chemicals, like iron chloride or aluminum sulfate as in phosphorus removal. Most commonly, sewage sludge is subjected to anaerobic digestion in a digester designed to allow bacterial action to occur in the absence of air. Following digestion, sludge is generally conditioned and thickened to concentrate and stabilize it and make it more dewaterable.

Some of the alternatives for the ultimate disposal of sludge include land spreading or incineration. Each of these choices has disadvantages, such as the presence of toxic substances in sludge spread on land, or the high fuel cost of incineration.

### Chemical Pollutants in Sewage

Typical domestic sewage contains oxygen-demanding materials, sediments, grease, oil, scum, pathogenic bacteria, viruses, salts, algal nutrients, pesticides, heavy metals, etc. Several characteristics are used to describe sewage. These include turbidity, suspended solids, total dissolved solids, acidity, and dissolved oxygen.

Generally, early objectives concerning sewage treatment were related to elimination of suspended solids, biodegradable organic matter, pathogen organisms, and nutrients (P, N). Nowadays, sewage treatment does not only focus on environmental issues, but also on removal of toxic compounds at trace level, which may cause health problems in the long term. Main pollutants to be removed in the treatment are shown in **Table 1**, including both organic and inorganic substances. A primary issue concerning pollution by organics is dissolved oxygen decrease occurring as a result of biodegradation.

A brief description of typical parameters to be taken into account in the analytical characterization of sewage is given below.

#### Physicochemical Parameters

**Redox potential** This measurement is important since redox conditions of water affect the environmental behavior of many elements. The potential can be measured electrometrically by using an inert metal electrode.

**pH** This parameter is closely related to the redox properties of the system. In addition it dictates many water processing steps (e.g., acid–base neutralization, softening, precipitation, coagulation, disinfection,



**Table 1** Relevant pollutants in sewage treatment

<i>Pollutants/parameters</i>	<i>Concentration (mg l<sup>-1</sup>)</i>	<i>Remarks</i>
<i>Solid material</i>		
Total solids (TS)	350–1200	Organic matter can give rise to dump and anerobial conditions when discharged
Total dissolved solids (TDS)	250–850	
Suspended solids (SS)	100–350	
Sedimentable substances	5–20	
<i>Biodegradable organic matter</i>		
Biochemical oxygen demand (BOD <sub>5</sub> )	110–400	Biodegradable organic matter can cause an oxygen decrease in water
Total organic carbon (TOC)	80–290	
Chemical oxygen demand (COD)	250–1000	
<i>Nutrients</i>		
Total nitrogen Kjeldahl (NTK)	20–85	Nutrients are essential for life. Excess nutrients can give rise to eutrophication
Ammonium	12–50	
Nitrates	0.1–0.4	
Nitrites	0–0.1	
Total phosphorus	4–15	
Inorganic phosphorus	3–10	
<i>Pathogens (no. per 100 ml)</i>		
Total coli, salmonella, etc.	10 <sup>6</sup> –10 <sup>9</sup>	Disease transmission
<i>Refractory organic matter</i>		
Surfactants	2–6	Organic matter resistant to conventional treatments
Phenols	0.5–1	
<i>Inorganic dissolved solids</i>		
Chlorides	30–100	Salinity of water is increased. This can cause some problems during treatments (e.g., corrosion)
Sulfate	20–50	
Alkalinity (as Na <sub>2</sub> CO <sub>3</sub> )	50–200	
S <sup>2-</sup> , F <sup>-</sup> , Ca, Na, K, Fe, Al, etc.		
<i>Priority pollutants</i>		
Heavy 'metals' (As, Cd, Hg, Cr, Pb, etc.), CN <sup>-</sup> , chlorinated hydrocarbons (CHCs), polychlorinated biphenyls (PCBs), polycyclic aromatic hydrocarbons (PAHs), pesticides, dioxins/furans, etc.	μg l <sup>-1</sup>	Organic and inorganic compounds with carcinogenic, mutagenic, or toxic character

and corrosion control). pH measurements are made potentiometrically.

**Conductivity** This parameter helps to establish water quality and its temporal changes. Also, it allows dissolved solids to be quantified. This parameter is measured by conductimetry and is related to the total ionic concentration.

**Temperature** It influences the solubility of salts and gases. This parameter also affects electric conductivity through dissociation of salts, potential, and pH.

**Decantable water and suspended matter** Decantable water is quantified by allowing a given volume of water to stand for 2 h and then measuring the volume of the decanted water. Suspended matter can be determined by sedimentation, filtration, or centrifugation.

### General Chemical Parameters

**Salinity** This parameter is defined as the total amount of solutes in water after conversion of carbonates into oxides (and bromides and iodides into chlorides) and oxidation of all organic matter. It is often determined from conductivity measurements.

**Hardness** It is the combined concentration of all metal cations present except those of alkali metals and the proton. It includes to a large extent the contributions of Ca<sup>2+</sup> and Mg<sup>2+</sup>. The most usual determination, total hardness, involves titration with ethylenediaminetetraacetic acid (EDTA) in ammonia-ammonium buffer at pH 10 with Net as indicator.

**Acidity and alkalinity** Water acidity is defined as the capacity to react with a strong base up to a given

pH value. Total acidity is determined by titration with NaOH up to the color change in phenolphthalein (pH 8–9) while mineral acidity is obtained by titration in the presence of Methyl Orange as indicator (pH 5). Water alkalinity is defined as the capacity to react with a strong acid up to a preset pH. A titration with a mineral acid using phenolphthalein (free alkali and alkali carbonate contents) or Methyl Orange (total alkalinity) as indicators is performed.

### Heavy Metals and Arsenic

These are among the most harmful pollutants in sewage. Essential elements (e.g., Fe) as well as toxic metals such as Cd, Hg, and Pb are included. Main sources of heavy metals are industrial wastes, mining, fuels, coal, metal plating, etc. Metal determinations in sewage are preferably carried out by atomic spectrometry (flame and electrothermal atomization), atomic emission spectrometry, inductively coupled plasma-mass spectrometry, stripping voltammetry, spectrophotometry, and kinetic methods. Hg is advantageously determined by the cold vapor technique and As by the hydride technique.

### Anionic Species and Gases

These can be classified as nutrients (basically nitrogen- and phosphorus-containing species such as nitrate, nitrite, orthophosphate, polyphosphate, organic phosphate), toxic species (cyanide), and others (sulfide, sulfate, sulfite).

**Cyanide** It is a very toxic species. Cyanide is widely used in industry, especially for metal cleaning and electroplating. It is determined following distillation. Cyanide ion is released on acidification and collected in an NaOH solution where it can be quantified by titration with  $\text{AgNO}_3$ .

**Nitrogen-containing species** The nitrate content in waters is usually low but can rise to appreciable levels in farming wastewater as a result of soil fertilization. Nitrite originates from the microbiological reduction of nitrate or the oxidation of ammonia. Nitrite is usually determined spectrophotometrically using the Griess reaction. Nitrate is determined by the same method after prereduction to nitrite.

Ammonia is the initial product of the decay of nitrogenous organic wastes, its presence being an indication of such wastes. It originates from the deamination of nitrogen-containing organics and the hydrolysis of urea. Ammonia is usually determined by photometric, volumetric, or potentiometric methods.

**Phosphorus-containing species** This is one of the major nutrients in waters. It usually comes from fertilization, detergents, and animal and plant decay. Excess phosphorus and other nutrients in sewage are responsible for eutrophication. The most recommended method for its determination is spectrophotometry after formation of the molybdophosphate heteropolyacid and reduction to molybdenum blue.

**Sulfur-containing species** The most toxic sulfur species in waters are  $\text{S}^{2-}$  and  $\text{SO}_3^{2-}$ . Sulfide ion in sewage comes from sulfur-containing proteins and other organic compounds in feces and from industrial effluents. Sulfide ion can be determined spectrophotometrically by formation of Methylene Blue. Sulfite ion is present in some industrial sewages (textile, cellulose) and can be determined iodometrically. Sulfates in sewage come from some industrial processes and cause an increase in the salt load of the receiving stream. Sulfate ion is determined gravimetrically as  $\text{BaSO}_4$ .

**Dissolved oxygen** It is directly related to the concentration of organic matter, which uses oxygen for its bacterial degradation. Analytical determination of dissolved oxygen can be performed by the Winkler method or using membrane electrodes (polarographic and galvanic).

### Organic Parameters

**Biochemical oxygen demand (BOD)** The degree of oxygen consumption by microbially mediated oxidation of pollutants is called BOD. This parameter is commonly measured by determining the quantity of oxygen required by suitable aquatic microorganisms during a 5-day period ( $\text{BOD}_5$ ).

**Chemical oxygen demand (COD)** It consists of the amount of oxygen required for the organic matter to be oxidized by a chemical oxidant (e.g., potassium dichromate). For many wastewaters, a correlation between COD and  $\text{BOD}_5$  can be established which is advantageous due to the shorter time required to complete COD determinations.

**Total organic carbon** It refers to the amount of organically bound carbon present in sewage. Its measurement is based on the oxidation of organic carbon to  $\text{CO}_2$ , which can be subsequently determined by several methods (e.g., volumetric, gravimetric, conductimetric, coulometric).

**Detergents, soaps, detergent builders** Detergents are the primary source of domestic wastes in addition



to fecal matter and meal leftovers. Traditional detergents containing phosphates are banned in many countries to avoid eutrophication. Nowadays, only biodegradable detergents are allowed.

**Biorefractory organics** These are a group of great importance, including mainly aromatic or chlorinated hydrocarbons. These compounds are not completely removed by the biological treatment, other treatments such as stripping, solvent extraction, ozonation, and carbon adsorption being necessary.

**Pesticides** These are substances with biological action that can be classified into insecticides, fungicides, herbicides, etc. Insecticides in sewage come mainly from agriculture applications. Organophosphorus and carbamate pesticides have replaced organochlorine pesticides, which have been banned due to their high toxicity, persistence, and bioaccumulation in the environment. Gas chromatographic methods are employed for volatile pesticides. Liquid chromatography is recommended for pesticides with low thermal stability.

**Phenols** The term 'phenols' is used to designate a series of hydroxyl derivatives of benzene, with phenol as the parent compound. These compounds come from paper mills and cellulose factories as well as from degradation of pesticides such as carbaryl. Disinfecting treatments of chlorine-containing waters produce chlorophenols, which cause a persistent taste. Analytical control of these compounds is commonly performed by chromatographic or photometric methods.

**Other organic compounds** Other dangerous substances include hydrocarbons (fats and oils), polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), and dioxins. Chromatographic methods (gas or liquid chromatography) are extensively used for separation and quantification of these groups.

## Treatment for Sewage Purification

There is a wide variety of treatment suitable for sewage purification that may be classified according to its nature, i.e., physical, chemical, thermal, or biological. In physical treatment, solid removal processes are involved. Some modification of the chemical structure of pollutants is developed in chemical treatment so that total degradation or separation is feasible. Thermal treatments use high temperatures for decomposition of pollutants, but high costs are normally involved, so they are mainly used for refractory substances, which are difficult to remove by other treatments. In biological treatment, microorganisms are applied to the sewage for degradation of organics and formation of biological flocs that can be easily separated. Typical sewage treatments are shown in Table 2. The suitability of treatments to the different pollutants is shown in Table 3.

According to the size of solids originally present in sewage, current processes for the treatment of wastewater may be divided into the following main categories of pretreatment, primary treatment, secondary treatment, and tertiary treatment.

During the pretreatment, solid particles are removed, physical procedures being mostly involved. Primary treatment of wastewater consists of the removal of insoluble matter such as grit, grease, and scum from water. Secondary wastewater treatment is designed to remove BOD, but also involves removal of nitrogen by denitrification, usually by taking advantage of the same kind of biological processes that would otherwise consume oxygen in water receiving the wastewater. Tertiary waste treatment (sometimes called advanced waste treatment) is a term used to describe a variety of processes performed on the effluent from secondary waste treatment. The contaminants removed by tertiary waste treatment fall into the general categories: (1) suspended solids, (2) dissolved organic compounds, and (3) dissolved inorganic materials, including the important class of algal nutrients. In addition to these chemical contaminants, secondary sewage

**Table 2** Main treatments applied to sewage purification

<i>Physical</i>	<i>Chemical</i>	<i>Thermal</i>	<i>Biological</i>
Sedimentation	Coagulation–flocculation	Wet oxidation	Bacterial filters
Flotation	Precipitation	Supercritical oxidation	Activated sludges
Filtration	Oxidation	Incineration	Biodisks
Evaporation	Electrolysis		Anaerobic degradation
Adsorption	Ion exchange		
Stripping	Osmosis		
Extraction			

**Table 3** Suitability of treatments for relevant pollutant groups in sewage

	Biodegradable			Inorganic			
	Solids	Matter	Nutrients	Substances	Metals	Volatiles	Refractory matter
<i>Physical</i>							
Sedimentation	A	—	—	—	—	—	—
Flotation	A	A	—	—	—	—	—
Filtration	A	A	—	—	—	—	—
Evaporation	—	—	—	—	—	A	—
Adsorption	—	A	—	P	—	—	P
Stripping	—	—	A	—	—	A	—
Extraction	—	A	—	P	A	—	—
<i>Chemical</i>							
Coagul-floc.	A	—	—	—	—	—	—
Precipitation	—	—	P	A	A	—	—
Oxidation	—	A	—	P	—	—	—
Electrolysis	—	A	—	A	A	—	—
Ion exchange	—	A	A	A	A	—	—
Osmosis	—	A	—	A	A	—	—
<i>Thermal</i>							
Wet oxidation	—	A	—	—	—	—	A
Supercritical ox.	—	A	—	—	—	—	A
Incineration	—	A	—	—	—	—	A
<i>Biological</i>							
Bacterial filters	—	A	A	P	—	—	—
Activated sludge	—	A	A	P	—	—	—
Pools	—	A	A	P	—	—	—
Biodisks	—	A	A	P	—	—	—
Anerobic degr.	—	A	A	A	—	—	A

A, appropriate; P, possible; —, inappropriate.

effluent often contains a number of disease-causing microorganisms, requiring disinfection in cases where humans may later come into contact with the water.

## Analytical Control in Sewage Treatment Plants

Before considering parameters to monitor in a sewage treatment plant, it is useful to briefly describe the different pathways involved during purification.

Figure 1 shows a block diagram of a typical sewage treatment plant. The sewage pathway includes different treatments addressed to water purification (Table 2); the sludge pathway includes a diversity of treatments for the solid residues (i.e., sewage sludge) generated during sewage purification. These treatments include sludge conditioning, concentration, dehydration, and digestion and depend on the sludge properties and final disposal, leading to volume reduction and lower fermentation power. The gas pathway is related to the anaerobic digestion process, which generates gases such as methane and carbon dioxide; finally, the residue pathway includes

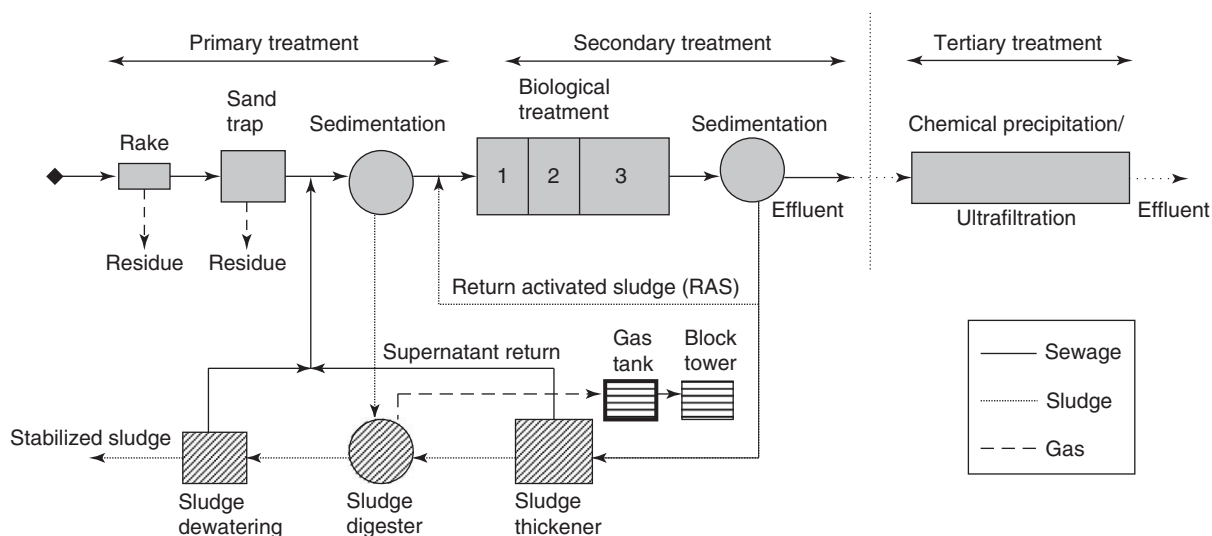
those pollutants separated during the pretreatment process.

### Analytical Control in the Sewage Pathway

Analytical determinations of relevant parameters in sewage are performed following reference methods, widely applied to both potable waters and wastewater (Table 4).

It is important to emphasize that representative samples must be collected before any analytical determination is performed. Factors such as sampling point, sample volume, sampling times, number of samples should be considered.

Typical analytical determinations are carried out at the sewage input of the treatment plant, outlet of the mechanical treatment, biological treatment stage, and sewage outlet into the waters. Temperature, pH, conductivity, sedimentable substances, suspended solids, COD, and BOD<sub>5</sub> are the parameters determined in all the points mentioned above. Alkalinity, NTK (total nitrogen Kjeldhal), ammonium, nitrite, orthophosphate, total phosphorus, sulfate, phenols, tenside, and lipophilic substances are primary parameters at the input, biological treatment, and outlet. Generally, analysis at both the input and outlet is



**Figure 1** Block diagram of a sewage biological treatment plant: (1) anaerobic conditions, biological phosphorus elimination, (2) denitrifying stage, and (3) aerobic conditions, oxidation of organic compounds.

important for the determination of substances that either affects treatment (e.g., aluminum and iron in case of chemical elimination of phosphorus) or are priority pollutants. In the latter, it is important to establish the purification degree taking into account the industrial activity in the area (e.g., heavy metals, cyanides, PCBs).

For control purposes as well as for modeling treatment plants, the most important parameter may be the characterization of organic matter. Dissolved oxygen levels determine water quality, but also indicate the possibility of bad odor.  $BOD_5$  is always monitored in sewages, its importance lying in three different features concerning treatment plants: (1) estimation of the oxygen required for biological stabilization of the organic matter present, (2) estimation of the size of facilities for the biological treatment of wastewater, and (3) assessing the treatment efficiencies.

Analytical control of a large number of routine parameters has driven developments in two areas: (1) automation and (2) sample handling. Thus, multiparameter probes allow determination of different parameters (e.g., total solids, COD, TOC, and nitrate) in different points of the treatment plant thereby avoiding sample handling and sampling.

Monitoring water pollution in a continuous manner provides more thorough information. Online analyzers are especially valuable in analytical control of many parameters in sewage treatment plants. Thus, parameters such as alkalinity, hardness, pH, COD,  $NH_3$ ,  $Ca^{2+}$ ,  $Cl^-$ ,  $F^-$ ,  $NO_3^-$ ,  $PO_4^{3-}$ ,  $SiO_3^{2-}$ ,  $SO_4^{2-}$ , and various metals can be determined using single-pollutant photometric analyzers. There are also automated methods for dissolved oxygen, conductivity, redox potential, and turbidity. Multipollutant

analyzers are of greater interest than those for single pollutants (e.g., Inductively coupled plasma-atomic emission spectroscopy (ICP-AES) and stripping voltammetric detectors for determination of multiple metals in waters).

With regard to the secondary treatment, monitoring of temperature, pH, dissolved oxygen, organic matter, and nutrients is needed to maintain the micro-organism activity under control. Thus, the  $BOD_5/COD$  ratio is employed as an estimation of the degradation capacity, being used for checking micro-organism activity.

On the other hand, it is necessary to monitor certain substances that can inhibit enzyme activity (e.g., insoluble salts, metal ions, etc.). These inhibiting substances can destroy or reduce the biodegradation capabilities of the activated sludge bacteria in the treatment plant, giving rise to an increase in treatment time and consequently to increase in cost. Toxicity can be estimated by a toxicity-direct analysis based on classical bioassays, measuring the inhibition percentage of respiration rate in activated sludge. This test is too laborious, and the use of biological test systems based on living cells or sub-organisms, e.g., inhibition of the bioluminescence of luminescent bacteria, is recommended. Also, the toxicity test is employed to assess environmental risk due to specific pollutant groups.

### Analytical Control of Sewage Sludge

A large variability exists in the sewage sludge composition depending on the origin and treatment of sewage. Sewage sludge of municipal origin contains high amounts of organic matter (i.e., >60%), apart

from significant inorganic fraction formed by  $\text{SiO}_2$ ,  $\text{Fe}_2\text{O}_3$ ,  $\text{CaO}$ ,  $\text{P}_2\text{O}_5$ ,  $\text{Al}_2\text{O}_3$ , and  $\text{K}_2\text{O}$ . Sewage sludge is rich in nutrients, containing  $\sim 5\%$  N,  $3\%$  P, and

$0.5\%$  K on a dry-weight basis, and can be used to fertilize and condition soil. Among the factors limiting this application of sewage sludge are excess

**Table 4** Standard methods for analysis of sewage

Contaminants/parameters	Method
<i>Physical and aggregate properties</i>	
Color	Visual comparison, spectrophotometric tristimulus filter
Turbidity	Nephelometric
Acidity/alkalinity	Titration
Hardness	EDTA titration
pH	Electrometric method
Salinity	Electrical conductivity
Flotables	Trichlorotrifluoroethane-soluble floatable oil and grease
Solids	Total solids dried at $103\text{--}105^\circ\text{C}$ , total dissolved solids dried at $180^\circ\text{C}$ , total suspended solids dried at $103\text{--}105^\circ\text{C}$ , fixed and volatile solids ignited at $550^\circ\text{C}$ , settleable solids (volumetric or gravimetric test); total, fixed, and volatile in solid and semisolid samples
<i>Tests on sludges (uniquely applicable to sludges or slurries)</i>	
Oxygen-consumption rate	Oxygen probe
Settled sludge volume	Settling columns, sludge volume index (SVI)
Zone settling rate	Settling vessel
Specific gravity (SG)	Weight of sample and weight of equal volume of water
Capillary suction time	The CST meter characterizes the sludge dewatering processes
Anaerobic sludge	Volumetric method, gas chromatographic, dissolved gas supersaturation
Digester gas analysis	(direct-sensing membrane diffusion method)
<i>Metals</i>	
	Acid digestion or extraction, atomic absorption spectrometry (flame, cold vapor, hydride generation, and electrothermal), emission spectroscopy (plasma and flame), spectrophotometry, anodic stripping voltametry
<i>Inorganic nonmetallic constituents</i>	
Anions	Ion chromatography
Boron	Spectrophotometry
Cyanide	After distillation, spectrophotometry, ion-selective electrode
Chlorine	Titrimetric, potentiometric, amperometric, colorimetric
Fluoride	Distillation, ion-selective electrode, colorimetric
Iodine	Colorimetric, amperometric titration
Nitrogen	Ammonia (distillation, titrimetric, ammonia-selective electrode, phenate method); nitrite (ion chromatography, colorimetric); nitrate (ultraviolet spectrophotometric screening, nitrate electrode, reduction); organic (Kjeldahl, persulfate method)
Dissolved oxygen	Iodimetric and modifications, membrane electrode
Phosphorus	Acid digestion, colorimetric methods
Silicon	Gravimetric, atomic absorption spectrometry, colorimetric
Sulfide	Colorimetry, iodometric, ion-selective electrode
Sulfate	Gravimetric, turbidimetric, colorimetric
<i>Aggregate organic constituents</i>	
Biochemical oxygen demand (BOD)	Five-Day BOD test, ultimate BOD test, respirometric method
Chemical oxygen demand (COD)	Reflux method, titrimetric, colorimetric
Total organic carbon (TOC)	Combustion-infrared, persulfate-ultraviolet oxidation, wet oxidation
Dissolved organic halogen (DOX)	Adsorption-pyrolysis-titrimetric method
Oil and grease	Partition-gravimetric, partition-infrared, extraction
Phenols	Extraction (total phenols), distillation (volatiles), colorimetry
Surfactants	Separation by sublation, methylene blue active substances (MBAS), anionics and cobalt thiocyanate active substances (CTAS), nonionic surfactants
Organic and volatile acids	Chromatographic, distillation
<i>Individual organic compounds</i>	
Volatile organic compounds, polychlorinated biphenyls (PCBs), pesticides, etc.	Purge and trap, extraction, gas chromatography, high-performance liquid chromatography, mass spectrometry

nitrogen pollution of runoff water and groundwater, survival of pathogens, and the presence of heavy metals in the sludge. The presence of heavy metals such as Pb, Cd, Cr, Cu, Hg, Ni, and Zn can preclude its application as a source of nutrients for crop land, mainly for prolonged use on soil.

Apart from heavy metals, other undesirable components typically found in sewage sludge are some organics (e.g., PCBs, pesticides, phenols, etc.) and pathogenic microorganisms (e.g., salmonella, fecal coliforms, etc.). While microorganisms can be destroyed by several treatments (aerobic and anaerobic digestion, composting), accumulation of heavy metals is of greatest concern for the use of sewage sludge in agriculture.

It has been widely recognized that prolonged application of sewage sludge on land can cause soil to be enriched with heavy metals, which, in turn, can reach the food chain. Different countries have enacted regulations to limit the heavy metal (Hg, Cd, Pb, Cr, Cu, Ni, and Zn) content in sewage sludge used for agriculture purpose.

Analysis of sewage sludge for metal determination must be performed at least every 6 months, with atomic spectrometry techniques mostly recommended (e.g., flame-atomic absorption spectrometry; electrothermal atomic absorption spectrometry; inductively coupled plasma-atomic emission spectrometry). Despite the need for only total metal amount, mobility and bioavailability studies must include some metal partitioning by sequential extraction. For this purpose, a number of extractable fractions with environmental significance are obtained using sequential extraction schemes such as that of BCR (The Community Bureau of Reference, now Standards, Measurements and Testing Programme) (fractions: acid soluble, reducible, and oxidizable) and Tessier (fractions: exchangeable, carbonate-bound, Fe–Mn oxides-bound and organic matter-bound).

See also: **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Principles and Instrumentation; Inductively Coupled

Plasma. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Bioassays:** Overview. **Flow Injection Analysis:** Environmental and Agricultural Applications. **Gas Chromatography:** Overview. **Liquid Chromatography:** Overview. **Mass Spectrometry:** Environmental Applications. **Nitrogen.** **Pesticides.** **pH.** **Phosphorus.** **Polarography:** Overview. **Potentiometric Stripping Analysis.** **Sampling:** Theory. **Segmented Flow Analysis.** **Spectrophotometry:** Overview. **Sulfur.** **Surfactants and Detergents.** **Titrimetry:** Overview. **Voltammetry:** Overview. **Water Analysis:** Overview; Industrial Effluents; Organic Compounds; Biochemical Oxygen Demand; Chemical Oxygen Demand. **Water Determination.**

## Further Reading

- Droste RL (1997) *Theory and Practice of Water and Wastewater Treatment*. New York: Wiley.
- Eaton DA, Clesceri LS, and Greenberg AE (1998) *Standard Methods for the Examination of Water and Wastewater*. Washington: American Public Health Association.
- Ewards JD (1995) *Industrial Wastewater Treatment: A Guidebook*. Boca Raton: Lewis Publishers.
- Farré M and Barceló D (2003) Toxicity testing of wastewater and sewage sludge by biosensors, bioassays and chemical analysis. *Trends in Analytical Chemistry* 22: 299–310.
- Filgueiras AV, Lavilla I, and Bendicho C (2002) Chemical sequential extraction for metal partitioning in environmental solid samples. Tutorial review. *Journal of Environmental Monitoring* 4: 823–857.
- Hammer MJ (2001) *Water and Wastewater Technology*. Englewood Cliffs, NJ: Prentice Hall.
- Katebi R, Johnson MA, and Wilkie J (1999) *Control and Instrumentation for Wastewater Treatment Plants*. London: Springer.
- Keith LH (1996) *Compilation of EPA's Sampling and Analysis Methods*. New York: CRC-Lewis.
- Manahan SE (1994) *Environmental Chemistry*. Boca Raton: Lewis Publishers.
- Pérez-Bendito MD and Rubio S (1999) *Environmental Analytical Chemistry*. Amsterdam: Elsevier.
- Smith RK (1999) *Handbook of Environmental Analysis*. New York: Genium Publishing Corporation.

## Organic Compounds

M R Preston, University of Liverpool, Liverpool, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The low concentrations at which most organic compounds appear in water (typically  $1:10^{-10}$ – $10^{-12}$ )

pose a number of significant challenges to the analytical scientist. These challenges relate to both sensitivity and selectivity of the analytical method and have generally required isolation/preconcentration steps to address the former and fractionation procedures to achieve the latter. Clearly, the ultimate ambition for analytical chemists is to achieve (near)



nitrogen pollution of runoff water and groundwater, survival of pathogens, and the presence of heavy metals in the sludge. The presence of heavy metals such as Pb, Cd, Cr, Cu, Hg, Ni, and Zn can preclude its application as a source of nutrients for crop land, mainly for prolonged use on soil.

Apart from heavy metals, other undesirable components typically found in sewage sludge are some organics (e.g., PCBs, pesticides, phenols, etc.) and pathogenic microorganisms (e.g., salmonella, fecal coliforms, etc.). While microorganisms can be destroyed by several treatments (aerobic and anaerobic digestion, composting), accumulation of heavy metals is of greatest concern for the use of sewage sludge in agriculture.

It has been widely recognized that prolonged application of sewage sludge on land can cause soil to be enriched with heavy metals, which, in turn, can reach the food chain. Different countries have enacted regulations to limit the heavy metal (Hg, Cd, Pb, Cr, Cu, Ni, and Zn) content in sewage sludge used for agriculture purpose.

Analysis of sewage sludge for metal determination must be performed at least every 6 months, with atomic spectrometry techniques mostly recommended (e.g., flame-atomic absorption spectrometry; electrothermal atomic absorption spectrometry; inductively coupled plasma-atomic emission spectrometry). Despite the need for only total metal amount, mobility and bioavailability studies must include some metal partitioning by sequential extraction. For this purpose, a number of extractable fractions with environmental significance are obtained using sequential extraction schemes such as that of BCR (The Community Bureau of Reference, now Standards, Measurements and Testing Programme) (fractions: acid soluble, reducible, and oxidizable) and Tessier (fractions: exchangeable, carbonate-bound, Fe–Mn oxides-bound and organic matter-bound).

See also: **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Principles and Instrumentation; Inductively Coupled

Plasma. **Atomic Mass Spectrometry:** Inductively Coupled Plasma. **Bioassays:** Overview. **Flow Injection Analysis:** Environmental and Agricultural Applications. **Gas Chromatography:** Overview. **Liquid Chromatography:** Overview. **Mass Spectrometry:** Environmental Applications. **Nitrogen. Pesticides. pH. Phosphorus. Polarography:** Overview. **Potentiometric Stripping Analysis. Sampling:** Theory. **Segmented Flow Analysis. Spectrophotometry:** Overview. **Sulfur. Surfactants and Detergents. Titrimetry:** Overview. **Voltammetry:** Overview. **Water Analysis:** Overview; Industrial Effluents; Organic Compounds; Biochemical Oxygen Demand; Chemical Oxygen Demand. **Water Determination.**

## Further Reading

- Droste RL (1997) *Theory and Practice of Water and Wastewater Treatment*. New York: Wiley.
- Eaton DA, Clesceri LS, and Greenberg AE (1998) *Standard Methods for the Examination of Water and Wastewater*. Washington: American Public Health Association.
- Ewards JD (1995) *Industrial Wastewater Treatment: A Guidebook*. Boca Raton: Lewis Publishers.
- Farré M and Barceló D (2003) Toxicity testing of wastewater and sewage sludge by biosensors, bioassays and chemical analysis. *Trends in Analytical Chemistry* 22: 299–310.
- Filgueiras AV, Lavilla I, and Bendicho C (2002) Chemical sequential extraction for metal partitioning in environmental solid samples. Tutorial review. *Journal of Environmental Monitoring* 4: 823–857.
- Hammer MJ (2001) *Water and Wastewater Technology*. Englewood Cliffs, NJ: Prentice Hall.
- Katebi R, Johnson MA, and Wilkie J (1999) *Control and Instrumentation for Wastewater Treatment Plants*. London: Springer.
- Keith LH (1996) *Compilation of EPA's Sampling and Analysis Methods*. New York: CRC-Lewis.
- Manahan SE (1994) *Environmental Chemistry*. Boca Raton: Lewis Publishers.
- Pérez-Bendito MD and Rubio S (1999) *Environmental Analytical Chemistry*. Amsterdam: Elsevier.
- Smith RK (1999) *Handbook of Environmental Analysis*. New York: Genium Publishing Corporation.

## Organic Compounds

M R Preston, University of Liverpool, Liverpool, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The low concentrations at which most organic compounds appear in water (typically  $1:10^{-10}$ – $10^{-12}$ )

pose a number of significant challenges to the analytical scientist. These challenges relate to both sensitivity and selectivity of the analytical method and have generally required isolation/preconcentration steps to address the former and fractionation procedures to achieve the latter. Clearly, the ultimate ambition for analytical chemists is to achieve (near)



real-time determination of all the compounds of interest in a particular situation with a minimum number of steps in the procedure. Multicompound, *in situ* probes are still some way off in most cases. Nevertheless, it is noticeable that in the past few years there has been significant progress in this direction through the exploitation of advances in sample extraction and detection (primarily hyphenated mass spectrometric techniques).

It must be remembered that the analysis itself is rarely the primary goal. The need to measure the concentration of a compound or group of compounds is usually determined by its real or potential impact on the system in which it occurs. Many of the developments in recent analytical procedures relating to the analysis of organic compounds in water therefore derive from the presence of the compounds on lists of Priority Pollutants (e.g., US Environment Protection Agency list of 307 compounds or the UK Department of the Environment, Fisheries and Rural Affairs list of 'Chemicals of Concern'), quality standards for drinking waters (e.g., disinfection by-products), or because they have suspected deleterious effects (e.g., endocrine disrupting chemicals, EDCs). Modern developments have, in many cases, led toward online or automated systems that not only allow a greater sample throughput but also reduce sample handling and thus the risks of sample contamination. However, the difficulties of obtaining reliable quantitative data of trace organic compounds in water are not to be underestimated and great attention must always be paid to the provision of appropriate quality assurance and quality control procedures.

It is worth noting that the analysis of organic contaminants in freshwaters presents different analytical challenges to those encountered in seawater. River and lake water concentrations of both natural and contaminant organic compounds are generally higher than seawater (as also may be the suspended particle loadings) but the background concentrations of humic and fulvic acids in the former may cause particular difficulties by, for example, overloading extraction devices. The presence of colloidal species in the freshwater environment may also be of greater import in determining the speciation of organic molecules than in marine systems. Potable water systems are very largely free of these problems but may contain higher proportions of volatile halomethane compounds created during water treatment processes. The high mineral content of groundwater may also provide sampling and analytical challenges where concentrations of inorganic species are high (and often variable) whilst those of organic compounds are very low. As a

general rule, it is probably true to say that analytical methods developed for seawater analysis will often be satisfactory in the freshwater environment but the converse is not true.

## Dissolved versus Particulate Organic Matter

The issue of separating organic compounds in water into solid and particulate phases is nontrivial. Organic species may occupy a size distribution that encompasses a spectrum of physical states from the truly dissolved to the unambiguously particulate. Between these states lie the complications. Traditionally, filtration through a preweighed membrane (typically 0.45  $\mu\text{m}$ ) or glass fiber (0.7–1.0  $\mu\text{m}$ ) filter has been the method of choice. However, this is a purely operational distinction and the use of other filter media such as polytetrafluoroethylene, silver, or aluminum filters can have a significant impact on the determined solid/liquid speciation (and expense). A number of experiments in recent years have identified a so-called third phase. This is essentially colloidal material that passes through a conventional filter and yet can have a significant impact on the apparent partitioning coefficients of organic species in natural systems. To add complications to an apparently simple procedure, the act of filtration itself may alter partitioning and worse, the degree of this effect may alter as the filter clogs and its filtration characteristics change. A further complication is that some filters may be supplied loaded with wetting/stabilizing agents that are organic in nature and which must be removed if contamination of the filtrate is to be avoided. Glass fiber filters are normally pre-ashed (by heating to 400–420°C for several hours), but this can lead to the formation/exposure of active sites within the filter capable of adsorbing sample components. Other filter types may also exhibit sorptive properties though these are rarely taken into account.

The filtration process itself is a mechanically aggressive one, particularly for delicate unicellular organisms such as the phytoplankton that are abundant in natural waters. If filtration is performed without due regard for this problem, cells may rupture spilling their contents into the filtrate. To avoid this problem vacuum filtration is probably only satisfactory under weak (<0.25 bar) vacuum and pressure filtration is generally more satisfactory.

Therefore, to obtain truly valid comparisons within the same sample set, let alone between different sets or different scientists and laboratories, is a very challenging task. One resolution of the problem is to avoid the filtration step entirely and analyze the whole sample. This approach has value

in some circumstance where, for example, variations in sample composition are small but it is not without its limitations and is potentially seriously flawed where the detailed effects of the presence of a specific compound are to be determined or modeled.

### Analysis of Total Dissolved Carbon

The dissolved organic carbon reservoir in natural waters represents a very significant component of the carbon cycle but reliable quantitative determination of this parameter has proved to be difficult particularly in seawater. Commercially available systems are better suited to freshwaters though obtaining accurate results is still not a trivial exercise and great care must be taken to ensure that blanks are satisfactorily and consistently low and that precision remains consistent.

### Preconcentration/Extraction Strategies

With the exception of highly sensitive methods such as the determination of amino acids in water using precolumn derivatization followed by liquid chromatography (LC)-fluorescence, virtually all other determinations require a preliminary preconcentration stage. Traditional procedures such as lyophilization and vacuum distillation are now rarely used because they are time consuming and unsuitable for compounds with significant volatility. Whilst reverse osmosis and ultrafiltration, or, its more recent variant tangential flow ultrafiltration, have found applications, they are again slow and susceptible to considerable unpredictability in their performance. This section therefore focuses on those techniques that have seen increased usage in recent years.

### Extraction of Particulate Organic Compounds

Traditional Soxhlet extractions are still widely used though increasingly being supplanted by faster technologies. Supercritical fluid extractions have not always lived up to their initial promise and the modern trends are toward microwave or pressurized liquid extractions (PLEs) using conventional solvents. Water itself can be used as the extracting solvent in subcritical water extraction (SWE). Analytical recoveries using PLE and SWE are often superior to those obtained by traditional methods as well as being much more rapid.

These techniques are not necessarily used alone. Some recent reports combine PLE with solid-phase extraction (SPE) technologies that include the use of

restricted access materials and liquid chromatography-mass spectrometry (LC-MS) to measure alkyl-phenolic compounds and steroid sex hormones in sediments.

### Extraction of Dissolved Organic Compounds

Classical methods such as liquid-liquid extraction (LLE) are comparatively rarely employed now because of their twin disadvantages of being slow and unwieldy where large sample volumes are to be extracted. The past decade has seen major advances in sorptive extraction techniques such as SPE or solid-phase microextraction (SPME), semipermeable membrane devices (SPMDs), and even stir bar sorptive extraction (SBSE). Such devices rely on the equilibrium (sometimes nonequilibrium) partitioning of analytes between aqueous and solid phases.

SPE is presently the most important technology for the isolation of organic compounds from water and SPE cartridges and disks have found a large number of applications. Numerous SPE systems that may be general purpose (e.g., alkyl silica or polymer) or class specific (molecularly imprinted polymers or immuno-affinity solid phases) are on the market. When these are coupled with the increasing availability of LC-mass spectrometry (MS) systems, sometimes in in-line format (i.e., SPE-LC-MS), the number and range of published applications is increasing rapidly. The recent development of molecularly imprinted solid-phase extraction (MISPE) is leading to an increasing range of applications including the extraction of triazine herbicides, sulfonyl urea, and chlorinated phenoxyacid herbicides from surface water samples and 4-nitrophenol from river water using online MISPE-reversed phase LC.

SPME technology is normally based on hollow fibers internally coated with a very small (typically 500 nl) quantity of suitable adsorbent. The great advantages of this technology are its relatively low cost, compatibility with both aqueous and gaseous samples, and its ease of integration into gas chromatographic (GC) systems. Numerous types of SPME coatings are now on the market giving a rapidly increasing range of applications. Some organometallic species have also been extracted using SPME. One significant advantage of SPME extraction is that the quantity of analyte removed from the system is normally a negligible proportion of the whole. This so-called 'negligible depletion solid-phase microextraction' does not alter the fundamental equilibrium within the system under study, so allowing estimates of true bioavailability to be made; a result that is very difficult to achieve using conventional systems.

Further recent developments in fiber SPME have extended applications to compounds with low volatilities and/or low thermal stability. Thus, fiber- (or tube-) based SPME–LC has a considerable future potential, particularly as it has been recognized that GC capillary columns (available with a very wide range of internal coatings) can be used for this purpose. Samples are pumped through the tube using a micropump and then eluted onto the LC column using appropriate solvents. Applications of in-tube SPME (in combination with LC) include phthalates, chlorinate phenoxy acid herbicides, tributyl tin compounds, polyaromatic hydrocarbons (PAHs), and polar aromatic compounds in water.

SBSE is a variation of SPME where a polydimethylsiloxane-coated stirrer bar is used in place of a sorbent-coated fiber. The extraction is conducted by simply placing the bar in the sample solution and using a magnetic stirrer to stir it. Because the bar is relatively large in size the quantity of sorbent on its surface is also large, giving it a high sorption capacity yielding improved analyte recovery and sensitivity. To retrieve the sorbed materials from the bar it is placed in a thermal desorption unit (for GC analysis) or simply washed in solvent (for LC analysis). See Table 1 for a summary of applications of SPE variants.

### Semipermeable Membrane Devices

The underlying concept behind SPMDs is that an adsorptive material is enclosed within a semipermeable membrane thereby allowing compounds of interest to passively diffuse into the bulk adsorbent over time. To an extent this mimics the natural situation where chemicals passively diffuse into the fatty

tissues of exposed organisms. Compounds suitable for SPMD extraction are hydrophobic ( $\log K_{ow} > 2$ ) and nonionic. Whilst a variety of sorbents have been tried, the commonest is triolein. Applications include extraction of PAH, polychlorinated biphenyl (PCB), polybrominated diphenyl ether (PBDE – flame retardants), and methyl triclosan. SPMDs have had some applicability in the long-term integration of contaminant levels in natural waters and have also found a variety of applications in atmospheric sampling.

### Integrated Online Extraction and Detection Systems

The need for maximum sample throughput and minimal human interaction within analytical procedures has provided considerable impetus to the development of integrated systems. SPE–LC in-tube SPME followed by ultraviolet (UV) or MS detection and membrane introduction mass spectrometry (MIMS) have both been used to this end. Submersible MIMS systems capable of extended underwater deployment down to  $\sim 200$  m and with a mass range of up to 200 amu have recently come onto the market. Flow injection coupled with MIMS allows fast, near-real-time determination of, for example, phenols in water. Derivatization of the phenols with acetic anhydride can be used to enhance both the selectivity and sensitivity of this method. Other online derivatization procedures are under development with a view to increasing the scope for rapid determination of highly polar compounds that have previously proved difficult to analyze. Large volume injection techniques and developments in enzyme-linked immunosorbent assay (ELISA) technologies

**Table 1** Examples of compounds extractable using solid-phase extraction (SPE), solid-phase microextraction (SPME), and stir bar solid extraction (SBSE)

SPE	SPME	SBSE
Alkylphenols	Chlorinated paraffins	Organotin compounds
Alkylphenol ethoxylates	Chlorophenols, methyl <i>tert</i> -butyl ether	PAHs
Antimicrobials and antibiotics such as sulfonamides, tetracyclines, fluoroquinolone, and triclosan	Organochlorine and organophosphorus pesticides, biocides	PCBs
Bisphenol A		
Caffeine		
Dimethylbenzylammonium chloride		
Fragrance materials		
Perfluorinated surfactants		
Pesticides such as those in the triazine, phenylurea, phoxy acid, acetanilide, carbamate, organophosphorus, and pyridylium classes and their metabolites		
Pharmaceuticals such as ibuprofen, codeine and estrogens		
Phthalate esters		

are also helping to improve sensitivity and reduce analytical times. Other strategies, including a cryo-trapping step prior to MIMS, have been used for organohalogens in water with detection limits as low as parts per trillion.

One recent development that offers scope for direct injection of aqueous samples into an MS system is the high-field, asymmetric waveform ion mobility spectrometer (FAIMS). By interfacing the FAIMS technology to an ESI-MS instrument a significant reduction in detection limits has been achieved for molecules of <300 Da. Preconcentration is unnecessary and analytical times are significantly shortened. Recent applications of this technique include naphthenic and haloacetic acids in water and amphetamines, morphine and codeine in urine.

## Separation and Detection Methods

Numerous combinations of separation and detection methods have been reported in the literature and only some of the more important examples are discussed below. An overview of analytical strategies as applied to some important chemicals is given in Table 2.

### Gas Chromatography

GC remains the technique of choice for very many volatile or semivolatile analytes or samples that can be made volatile through derivatization. GC offers excellent sensitivity and selectivity and an increasingly wide range of detectors. Amongst the most significant developments in column technology in recent years has been the introduction of chiral columns (many based on  $\beta$ -cyclodextrin). The ability to separate enantiomers of, for example, semivolatile organic pollutants is a major new tool in environmental studies allowing significant insights into the mode of action of chemicals and differences in their degradation behavior. The use of enantiomer fractions to describe compounds such as chiral hexachlorobornane,  $\alpha$ -HCH, *trans*-chlordan, PCBs 95 and 136 has led to new insights into their behavior in biotic and abiotic environmental systems. Chiral phases in LC are as yet less well developed but it is to be anticipated that the availability and uses of such materials will become more important in the relatively near future.

Efforts to decrease analytical times have largely depended on improved response times in detectors. Fast GC, where complete separations are achieved in a few minutes, has a number of reported applications. Time-of-flight mass spectrometry (TOFMS) is often

the chosen detection system though some quadrupole systems also have sufficiently fast scan times to be of service. Rapid online MIMS systems have permitted parts per trillion analysis of several trihalomethanes in water with a sample throughput of  $\sim 20$  per hour. Fast GC in combination with TOFMS has been used for PCB analysis, as has GC with pulsed electron capture detection. Fast GC with inductively coupled plasma MS (ICP-MS) has been used for the determination of polybrominated diphenyl ethers in sewage sludge with low ppb detection limits achievable in an analysis time of  $\sim 10$  min.

Efforts to improve selectivity have introduced a trend from single GC columns to the use of two (or more) columns (two-dimensional GC). The advantages of such systems are in a huge ( $10 \times$ ) improvement in peak capacity over conventional GC allowing much more complex mixtures to be handled. Such systems are more complex than simply connecting two columns together and have only been successful because of improvements in hardware such as heated sweepers, cryofocussed modulators, and valve switching. Such systems have enormous potential to identify compounds of major environmental importance from an extremely complicated background matrix.

Improvements in column technology have led to the ready commercial availability of columns with chiral stationary phases that allow enantiomers to be separated. A similar result can be achieved through the use of chiral derivatization reagents. The ability to separate different enantiomers, which may have very different biological reactivities, has opened major avenues of research. In particular, it has allowed very significant reassessments of biological degradation rates of contaminant chemicals where one enantiomer may exhibit very different behavior to its mirror image.

GC detector technology has also developed in areas other than MS. Fourier-transform infrared spectrometry (FTIR) and atomic emission detection (AED) can both provide structural insights into eluted compounds. AED, in particular, allows analysts to move away from the carbon-based detection capabilities of the standard flame ionization detector to focus on other elements such as N, Cl, S, P, Br, and F allowing molecules with particular elemental constituents to be readily identified.

### Liquid Chromatography

Applications of LC and, in particular, LC-MS continue to grow. Much recent emphasis has also been placed on combined extraction/analytical systems with SPE and automated in-tube SPE-LC

**Table 2** Examples of analytical strategies applied to chemicals in water

<i>Compound</i>	<i>Sample preparation method</i>	<i>Analyte detection technique</i>
Acidic compounds	Goulden extractor versus SPE	GC-MS
Amines, aliphatic	Derivatization with 2,4-dinitrofluorobenzene or benzenesulfonyl chloride, solvent extraction with CH <sub>2</sub> Cl <sub>2</sub>	GC-MS
Atrazine, diuron, simazine, terbutylazine	SPE	LC-ESI-MS-MS
Atrazine	Direct injection	ESI-ion mobility spectrometry
Azo dyes, sulfonated	Direct injection	IC electrochemical detection
Bromocyclen	LLE, SPE	Chiral GC
Caffeine	SPE	GC-MS, GC-MS-MS
Carbendazim	Protein G immunoaffinity column	GC-MS-MS
2,4-Dinitrotoluene	None	Surface-enhanced Raman spectroscopy
2,4-Dinitrotoluene	Custom vapor sampler	Sensory polymers attached to fiber optics
Chemical warfare agents		Flow injection amperometric enzyme biosensor
Chloro-bispropyl ethers		Enantioselective GC-MS
Chloroethers	SPME	GC-MS
Chlorolignosulfonic acids	SPE	Curie point pyrolysis-GC-MS-MS
Clofibric acid	SPE, methylation with diazomethane	HRGC-MS-MS
Dinitroanilines	SPME	GC-ECD
1,3-Dioxanes, 1,3-dioxolanes	Closed loop stripping	GC-CI-MS or GC-EI-MS-MS
Diquat, paraquat, difenzoquat	SPE with graphitized carbon black	LC
EDTA	Lyophilization, esterification with ethanol	GC-NPD
Endocrine disrupting chemicals	SPE	GC-MS Microemulsion electrokinetic chromatography, capillary zone electrophoresis, micellar electokinetic chromatography (MEKC), cyclodextrin modified MEKC
Fluocofuron, sulcofuron	LLE, SPE	LC-ESI-MS-MS
Fluoroquinolone	SPE with mixed-phase cation exchange resin	LC/UV
Fragrance materials	SPE	Isotope dilution GC-MS
Geosmin (methylisoborneol)	Hexane microextraction	GC-MS
Haloacetic acids	Many methods, SPE, derivatization, headspace sampling	GC-MS, negative ion GC-MS
Ibuprofen	SPE Biobeads SM-2, methylation with diazomethane	Chiral GC-MS-MS-SIM or SRM
MTBE	Headspace SPME, purge and trap	GC-MS, MIMS
Organonitrogen pesticides	SPE with graphitized carbon black or LLE	GC-NPD, LC-APCI-MS
Organotin compounds	SBSE, derivatization	TD-GC-ICPMS
Organotin compounds	Ethylation, SPME	Multicapillary GC-AES
Organotin compounds	Ethylation, SPME	Isotope dilution GC-MS
Organotin compounds	Ethylation, LLE, PLE	Isotope dilution GC-MS
PCBs	Many methods, SPE polyurethane columns	GC-ECD, positive and negative ion GC-MS
Perfluorinated surfactants	SPE	LC-MS-MS, <sup>19</sup> F-NMR
Pharmaceuticals – hormones and other consumer products	SPE, continuous liquid extraction	LC-MS, GC-MS with and without derivatization
Pharmaceuticals (acid and neutral)	SPE (Oasis HLB sorbent) derivatization with diazomethane	GC-MS
Phthalate esters	LLE	GC-ECD, GC-SIM
Sulfonylurea, imadazolinone, sulfonamide	Spe with RP-102	LC-ESI-MS
Sulfonamide and tetracycline microbials	SPE	LC-MS
Triazines, acetamindes phenoxy acids	SPE with graphitized carbon black, derivatization with diazomethane	GC-MS
Triclosan	SPE, PLE	GC-MS, GC-MS-MS
Triclosan and methyl triclosan	SPE, dialysis, GPC derivatization	GC-MS
Triclosan and methyl triclosan	SPE, SFE, derivatization	GC-MS, GC-HRMS
Triclosan	SPE	GC-MS
Trifluoroacetate	SPE, extracted with MTBE, derivatized with pentafluorophenyldiazoethane	GC-MS (negative ionization)



systems being the subject of recent reviews. Important applications include analysis of estrogens and progestogens, pesticides, nonionic surfactants, phenols, and organotin and organophosphorus compounds in aqueous samples. Developments in micro-ELISA systems have in one application allowed the sample volumes for atrazine analysis to be reduced to 30  $\mu\text{l}$  permitting the determination of atrazine at levels of 0.022–2.90  $\mu\text{g l}^{-1}$ .

As in GC, the uses of LC for the separation of chiral species have significantly increased. Column materials now include chiral phases that may, for example, be based on monolithic silica columns with chemically bonded beta-cyclodextrin, teicoplanin, or cellulose tris(3,5-dimethylphenylcarbamate). Hydrophobic amino compounds have been separated by LC using a crown ether dynamically coated chiral stationary phase.

Chiral derivatization agents have found an increasing number of applications in drug analysis. Examples of such reagents include (–)-1-(9-fluorenyl)ethyl chloroformate or *o*-phthalaldehyde in combination with the chiral thiol *N*-acetyl-L-cysteine for amphetamine analysis, benzofurazan-bearing chiral derivatization reagents (drugs and other biologically important compounds), subsequent chiral derivatization with (–)-alpha-methoxy-alpha-(trifluoromethyl)phenylacetyl chloride (*R*-(+)- and *S*-(–)-metoprolol). Applications in environmental samples are more limited but include the separation of amino acid enantiomers using *o*-phthalaldehyde/*N*-isobutyryl-L-cysteine to yield fluorescent diastereomeric derivatives of chiral primary amino acids.

## Mass Spectrometry

The use of mass spectrometry in environmental analysis is now ubiquitous and has recently been reviewed by William L. Budde and (separately) by Susan Richardson. Whilst such a detailed review is clearly beyond the scope of this article, the reader is referred to other articles in this volume and the present section is restricted to highlighting a few relevant modern developments and they relate to analysis of organic compounds in water.

Developments in LC–MS have probably been greater than in GC–MS though that too has seen significant advances. LC–atmospheric pressure chemical ionization (APCI)–mass spectrometry has become particularly important and has found numerous applications in the determination of, for example, industrial mass chemicals such as dyes, aromatic sulfonates, surfactants, and complexing agents, and also trace compounds such as drugs, endocrine-disrupting compounds, toxins, phenols,

and haloacetic acids. LC–MS has, however, only had limited success in the characterization of natural organic matter though it is beginning to develop the potential for the study of organometallic compounds. LC–APCI–MS is, however, not without its difficulties. Both qualitative and quantitative analyses may be inhibited by the limited resolution of quadrupole and ion-trap MS and the limited fragmentation created through collision-induced dissociation. Matrix effects can create difficulties in quantitation, particularly when electrospray technologies are used. These problems are not impossible to solve and improvements in instrumentation such as tandem MS or TOFMS both offer considerable potential for improvements particularly when coupled with better chromatographic (or other) separations before introduction into the mass analyzer. Ion-mobility mass spectrometry has been developed into a field instrument initially for vapor-phase organic chemicals and chemical warfare agents. However, when coupled with electrospray ionization MS, nonvolatile and biological compounds can also be measured in aqueous field samples. This technique has been used for the determination of atrazine in natural water samples. There are some calibration problems with this technology but regardless of analyte properties and upper limit of  $10^{-5} \text{ mol l}^{-1}$  is commonly observed in most electrospray ionization systems.

New developments in analysis utilizing GC–MS have included GC–FTIR coupled with high-resolution EI and CI–MS to identify drinking water contaminants and GC–IR–MS to examine contaminated water, clay, and soil samples. GC–MS, LC–nuclear magnetic resonance (NMR), and LC–MS have been used to identify contaminants in industrial landfill leachate. Online systems have also been developed such as the measurement of organic contaminants in water samples. MIMS has allowed the development of commercial *in situ* devices that are now on the market (see above). Almost all combinations of preconcentration, separation, and mass detection have seen significant advances.

Isotope ratio MS (IRMS) is also receiving increasing attention as a tool for the identification of biomarkers and the tracking of chemicals from particular sources. Such systems generally employ a combustion stage between the GC and IRMS and are, therefore, sometimes designated GC–C–IRMS systems. Such a system has, for example, been used to track the isotopic enrichment of tetrachloroethene and trichloroethene (TCE) in groundwater. Down-gradient wells showed  $\delta^{13}\text{C}$  values as enriched as –18‰ as compared to delta C-13 values for TCE in the source zone of –25.0 to –26.0‰.



## Emerging Contaminants and Techniques

As a very rough rule of thumb it is probably true to say that for every order of magnitude increase in analytical sensitivity the number of detectable analytes also increases by a similar factor. Thus, chemists are, in principle, able to identify and quantify ever-smaller amounts of individual chemicals, be they natural products or contaminants. This raises considerable problems in a number of areas but particularly the time taken for analysis, data management, and interpretation. Analysts are, therefore, increasingly being forced to focus on a restricted number of target compounds within a complex matrix and these targets tend to be set by legal priorities as much as environmental ones. Priority pollutant lists enshrined in law are examples of such factors.

For example, much attention is being paid to organic compounds generated inadvertently as a result of water treatment processes. Disinfection by-products (DBPs) are a major area of analytical activity because of their potential toxicity to humans. Examples of such chemicals are given in Table 3. Analysis of DBPs includes the use of techniques such as LLE-GC-ECD (with or without prior derivatization), SPE-GC-MS, LC-APCI-MS, headspace-SPME-GC-ECD, SPE-ion pair-LC-ESI-MS, and purge-and-trap-GC-MS. It is worth noting that some of the DBPs are thermally unstable requiring that suitable precautions to be employed to avoid decomposition of the target compound(s) during analysis.

Another area of active research is in the field of pharmaceuticals, hormones, and EDCs. Unmetabolized or partially metabolized pharmaceuticals may

be introduced into wastewater systems from domestic or medical sources such as hospitals. The presence of pharmaceuticals in drinking water has been known for a number of years and reports continue to be published, the great majority since 2000. The range of pharmaceuticals detected in ground waters is considerable and includes carbamazepine, clofibrac acid, dehydroerythromycin, dichlorfenac, famethazine, ibuprofen, phenazone and metabolites, primidone, sotalol, sulfamethazine, and sulfamethoxazole. Lists for drinking water occurrences are similarly extensive. SPE followed by GC-MS and/or GC-MS-MS are the most commonly adopted methods though LC-ESI-MS-MS has also been utilized.

Concerns over the endocrine disrupting group of chemicals extend back a considerable number of years and arose over such problems as the occurrence of imposex in mollusks, mismatches between the displayed physical characteristics of fish and their genetic sex, hermaphroditism in alligators and polar bears. Recognition that many biochemically active chemicals can pass unchanged through water treatment plants has led to increased monitoring of both wastewater and drinking water. The list of chemicals with so-called endocrine disrupting properties continues to grow and includes alkyl phenols (e.g., nonyl phenol and bisphenol A), brominated fire retardants, certain chlorinated pesticides (e.g., *p,p'*-DDE, lindane) and PCBs (particularly those with *para*-hydroxy groups), dibenzodioxins and dibenzofurans, polybrominated diphenyl ethers, parabens, and phthalate esters (notably diethylhexyl phthalate). Concentrations of these compounds in waters can be ecotoxicologically significant. For example, diethylhexyl phthalate has been detected in concentrations as high as  $3.5 \text{ mg l}^{-1}$  in drinking water in Holland.

Such a variety of compounds requires the application of the full panoply of analytical techniques including ELISA, LC, LC-MS, GC-MS (negative and positive ion), GC-MS-MS. Newer methods may include online SPE systems combined with MS or UV or electrochemical detection systems. One important consideration is that analysis of a particular EDC should also include its degradation and/or metabolic products because these too may remain active. So for alkyl phenols the concurrent measurement of the appropriate alkyl phenol ethoxylates and alkyl phenoxycarboxylic acids is environmentally important and a significant analytical challenge.

Developments in chiral separations are also rapidly occurring and were the subject of a recent review by Ward. Whilst many applications utilize capillary electrophoresis, applications are also found using GC, LC, and supercritical fluid chromatography.

**Table 3** High-priority disinfection by-products (DBPs)

### General categories

Brominated DBPs

Chlorinated DBPs

Iodinated DBPs

### Specific compounds or compound groups

3-Chloro-4-(dichloromethyl)-5-hydroxy-2 (5 *H*)-furanone (MX) and its analogues ( $\geq 10$ )

Haloacetates ( $\geq 1$ )

Haloacetoneitriles ( $\geq 5$ )

Haloaldehydes ( $\geq 4$ )

Haloacids (e.g., 3,3-dichloropropionic acid)

Haloamides ( $\geq 5$ )

Haloketones

Halomethanes ( $\geq 10$ )

Halonitromethanes ( $\geq 8$ )

Nitrosodimethylamine

Nonhalogenated aldehydes and ketones ( $\geq 6$ )

Volatile organic compounds and other DBPs ( $\geq 4$ )

Chiral separations are particularly important for biologically active chiral compounds (e.g., pesticides) where the active ingredient is only one enantiomer. Chiral separations have been used to track biological degradation rates and changes in source characteristics. GC-based applications include chiral chordane and  $\alpha$ -HCH in the polar bear food chain and levetiracetam and its enantiomer in dog plasma and urine whilst LC applications include amino acids in water and chiral aroma compounds in alcoholic drinks.

There are many potential avenues of development in the field of analysis of organic compounds in water. Primary requirements will always be in the directions of greater selectivity and sensitivity and decreased analytical times. It is already clear that developments in MS technology are increasing the opportunities of near-real-time *in situ* analysis, thus providing high-quality field instruments and reducing (though probably never excluding) the need for high-quality laboratory techniques and instrumentation.

See also: **Extraction:** Solid-Phase Extraction; Solid-Phase Microextraction. **Gas chromatography:** Overview. **Liquid Chromatography:** Overview. **Mass Spectrometry:** Overview. **Water Analysis:** Seawater – Organic Compounds; Seawater – Dissolved Organic Carbon; Biochemical Oxygen Demand; Chemical Oxygen Demand; Particle Characterization.

## Further Reading

Budde WL (2001) *Analytical Mass Spectrometry: Strategies for Environmental and Related Applications*. Washington, DC: Oxford University Press.

- Clement RE and Yang PW (2001) Environmental analysis. *Analytical Chemistry* 73: 2761–2790.
- Kataoka H (2002) Automated sample preparation using in-tube solid-phase microextraction and its application—a review. *Analytical Bioanalytical Chemistry* 373: 31–45.
- Koester CJ, Simonich SL, and Esser BK (2003) Environmental analysis. *Analytical Chemistry* 75: 2813–2829.
- Patnaik P (1997) *Handbook of Environmental Analysis. Chemical Pollutants in Air, Water, Soil and Solid Wastes*. Boca Raton, FL: CRC Press Inc., Lewis Publishers.
- Reemtsma T (2001) The use of liquid chromatography-atmospheric pressure ionization-mass spectrometry in water analysis—Part I: achievements. *Trac—Trends in Analytical Chemistry* 20(9): 500–517.
- Reemtsma T (2001) The use of liquid chromatography-atmospheric pressure ionization-mass spectrometry in water analysis—Part II: obstacles. *Trac—Trends in Analytical Chemistry* 20(10): 533–542.
- Richardson SD (1999) Water analysis. *Analytical Chemistry* 71: 181R–215R.
- Richardson SD (2001) Mass spectrometry in environmental sciences. *Chemical Review* 101: 211–254.
- Richardson SD (2003) Water analysis: emerging contaminants and current issues. *Analytical Chemistry* 75: 2831–2857.
- Snyder SA, *et al.* (2003) Pharmaceuticals, personal care products, and endocrine disruptors in water: implications for the water industry. *Environmental Engineering Science* 20(5): 449–469.
- US-EPA (1999) *Introduction to Water Quality Standards*. Office of Water 4305 EPA-823-F-99-020, 41pp.
- Ward TJ (2002) Chiral separations. *Analytical Chemistry* 74: 2863–2872.
- Zamboni CG (2003) Coupling solid-phase microextraction to liquid chromatography. A review. *Analytical Bioanalytical Chemistry* 375: 73–80.

## Biochemical Oxygen Demand

**R Kumar and A Kumar**, Institute of Genomics and Integrative Biology, New Delhi, India

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by C P Way, pp. 5527–5533, © 1995, Elsevier Ltd.

## Introduction

The biochemical oxygen demand (BOD) is a crucial environmental index for determining the relative oxygen requirements of wastewater, effluents, and polluted water. It refers to the quantity of oxygen required by bacteria and other microorganisms in the biochemical degradation and transformation of organic matter under aerobic conditions. The BOD

is also interpreted as a measure of the concentration of organic material that can serve as a substrate to support the growth of microorganisms.

The BOD test, as used for assessing the efficiency of wastewater treatment, is intended to measure some fraction of the carbonaceous oxygen demand, i.e., the oxygen consumed by heterotrophic microorganisms that utilize the organic matter of the waste in their metabolism, and not the oxygen demand exerted by autotrophic nitrifying bacteria. Since ammonia is usually present in wastewater, nitrification inhibitors must be used to suppress the exertion of nitrogenous oxygen demand. Carbonaceous oxygen demand is called first-stage BOD and nitrogenous oxygen demand is called second-stage BOD.

Chiral separations are particularly important for biologically active chiral compounds (e.g., pesticides) where the active ingredient is only one enantiomer. Chiral separations have been used to track biological degradation rates and changes in source characteristics. GC-based applications include chiral chordane and  $\alpha$ -HCH in the polar bear food chain and levetiracetam and its enantiomer in dog plasma and urine whilst LC applications include amino acids in water and chiral aroma compounds in alcoholic drinks.

There are many potential avenues of development in the field of analysis of organic compounds in water. Primary requirements will always be in the directions of greater selectivity and sensitivity and decreased analytical times. It is already clear that developments in MS technology are increasing the opportunities of near-real-time *in situ* analysis, thus providing high-quality field instruments and reducing (though probably never excluding) the need for high-quality laboratory techniques and instrumentation.

See also: **Extraction:** Solid-Phase Extraction; Solid-Phase Microextraction. **Gas chromatography:** Overview. **Liquid Chromatography:** Overview. **Mass Spectrometry:** Overview. **Water Analysis:** Seawater – Organic Compounds; Seawater – Dissolved Organic Carbon; Biochemical Oxygen Demand; Chemical Oxygen Demand; Particle Characterization.

## Further Reading

Budde WL (2001) *Analytical Mass Spectrometry: Strategies for Environmental and Related Applications*. Washington, DC: Oxford University Press.

- Clement RE and Yang PW (2001) Environmental analysis. *Analytical Chemistry* 73: 2761–2790.
- Kataoka H (2002) Automated sample preparation using in-tube solid-phase microextraction and its application—a review. *Analytical Bioanalytical Chemistry* 373: 31–45.
- Koester CJ, Simonich SL, and Esser BK (2003) Environmental analysis. *Analytical Chemistry* 75: 2813–2829.
- Patnaik P (1997) *Handbook of Environmental Analysis. Chemical Pollutants in Air, Water, Soil and Solid Wastes*. Boca Raton, FL: CRC Press Inc., Lewis Publishers.
- Reemtsma T (2001) The use of liquid chromatography-atmospheric pressure ionization-mass spectrometry in water analysis—Part I: achievements. *Trac—Trends in Analytical Chemistry* 20(9): 500–517.
- Reemtsma T (2001) The use of liquid chromatography-atmospheric pressure ionization-mass spectrometry in water analysis—Part II: obstacles. *Trac—Trends in Analytical Chemistry* 20(10): 533–542.
- Richardson SD (1999) Water analysis. *Analytical Chemistry* 71: 181R–215R.
- Richardson SD (2001) Mass spectrometry in environmental sciences. *Chemical Review* 101: 211–254.
- Richardson SD (2003) Water analysis: emerging contaminants and current issues. *Analytical Chemistry* 75: 2831–2857.
- Snyder SA, *et al.* (2003) Pharmaceuticals, personal care products, and endocrine disruptors in water: implications for the water industry. *Environmental Engineering Science* 20(5): 449–469.
- US-EPA (1999) *Introduction to Water Quality Standards*. Office of Water 4305 EPA-823-F-99-020, 41pp.
- Ward TJ (2002) Chiral separations. *Analytical Chemistry* 74: 2863–2872.
- Zamboni CG (2003) Coupling solid-phase microextraction to liquid chromatography. A review. *Analytical Bioanalytical Chemistry* 375: 73–80.

## Biochemical Oxygen Demand

**R Kumar and A Kumar**, Institute of Genomics and Integrative Biology, New Delhi, India

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by C P Way, pp. 5527–5533, © 1995, Elsevier Ltd.

## Introduction

The biochemical oxygen demand (BOD) is a crucial environmental index for determining the relative oxygen requirements of wastewater, effluents, and polluted water. It refers to the quantity of oxygen required by bacteria and other microorganisms in the biochemical degradation and transformation of organic matter under aerobic conditions. The BOD

is also interpreted as a measure of the concentration of organic material that can serve as a substrate to support the growth of microorganisms.

The BOD test, as used for assessing the efficiency of wastewater treatment, is intended to measure some fraction of the carbonaceous oxygen demand, i.e., the oxygen consumed by heterotrophic microorganisms that utilize the organic matter of the waste in their metabolism, and not the oxygen demand exerted by autotrophic nitrifying bacteria. Since ammonia is usually present in wastewater, nitrification inhibitors must be used to suppress the exertion of nitrogenous oxygen demand. Carbonaceous oxygen demand is called first-stage BOD and nitrogenous oxygen demand is called second-stage BOD.

Some other methods have also been developed to estimate the water pollution potential. The chemical oxygen demand (COD) test was developed because a BOD<sub>5</sub> test requires 5 days for completion and therefore is not suitable either for real-time evaluation of the efficiency of wastewater treatment or for operational control of the treatment processes. The total organic carbon (TOC), dissolved organic carbon (DOC), and the ultraviolet spectrophotometric absorption (at 254 nm) methods for determining dissolved organic material and volatile suspended solids of particulate organic material were developed as alternative methods for measuring the strength of wastewater on the basis of an assumption that the primary purpose of biological treatment is to reduce the concentration of organic material in wastewater.

## History

The development of the field of environmental engineering science has been dependent on the development of the BOD concept and test. The names of Phelps, Streeter, and Butterfield are prominent in the development of the BOD concept. The primary interest of these early workers was to determine the dissolved oxygen (DO) profile in the receiving stream in response to a known dilution of wastewater. The interest was in determining the capacity of the receiving stream to assimilate organic wastes.

The modern concept of the BOD test originated in 1884 when Dupre showed that microorganisms, in metabolizing organic matter, caused a decrease in the DO levels of incubated water samples. Established in 1898, the United Kingdom Royal Commission on Sewage Disposal issued nine reports, the eighth of which formalized BOD to a standardized test in 1913. It recommended 5 days as the standard incubation period for microorganisms to act on organic matter for several reasons.

First, this period was long enough to allow a significant amount of organic material to oxidize but shorter than the time required to initiate significant oxygen consumption by oxidation of ammonia to nitrate (nitrification). Second, as it reflected the likely maximum time for water in any British river to flow to the sea, the effect of organic material on oxygen levels would be limited to that period before the water underwent high dilutions with seawater.

The present standard condition of 20°C was incorporated into the BOD method by the American Public Health Association Standard Method Committee in 1936 in the eighth edition of *Standard Methods for the Examination of Water and Wastewater*.

## Concept and Definition

Prediction of the rate of consumption of DO by microorganisms in a body of water receiving effluents is an essential part of assessing the potential for deoxygenation of the receiving water. The original approach to making this assertion (developed ~100 years ago) was to measure the DO concentration in the receiving water, incubate a sample of the water sealed off from contact with sources of oxygen for a period of time, and then measure the DO remaining to find out how much had been consumed. This approach evolved into the BOD<sub>5</sub> test, which has been in use in more or less its present form for more than 60 years.

The BOD test is a bioassay run under standardized laboratory conditions of time, and temperature and controlled conditions of pH, dilution, nutrients and microorganisms. Complete oxidation of organic matter can take anything from 20 days to several months, but because quantification of the amount of oxygen involved is impractical as a standard laboratory test, the shorter time span of 5 days was chosen (though some analysts use 7, 3, or even 2 days). The number of days over which the BOD test is carried out is often designated by using a subscript, e.g., BOD<sub>5</sub>. Some seeding material (bacteria) from the waste treatment plant in the form of settled sewage or ready-made commercially available material is used in the test.

## Expected Variance

This method involves microorganisms, which might not be specific to a particular wastewater. One might well ask whether a low value of BOD would mean that the organic substrate had been eaten up or simply that the seed cells could not metabolize it.

'Standard Methods', using results from a series of interlaboratory studies, indicates as acceptable a range of  $\pm 15\%$  at the  $200 \text{ mg l}^{-1}$  level for the reference glucose-glutamic acid solution. While variances quoted for real samples range up to 50%, replication of real samples by a single analyst should in general produce results within 15% around the mean.

## Sampling and Sample Storage

Samples for BOD analysis may degrade significantly during storage between collection and analysis. One must consider the following for proper sampling and storage.

- Any source of organic contamination that could contribute a significant fraction to the measured



BOD load, e.g., detergent residues in a sample container, especially for low-BOD samples.

- Any loss of organic material that would be a significant fraction of the BOD load: e.g., where the use of plastic containers could result in the diffusion of significant amounts of organic material into the walls of the container, a glass container should be used instead.
- Collection of a representative sample; as BOD tests are often carried out on sewage effluents and wastewater that are likely to be inhomogeneous, care should be taken to ensure the sample is as representative as possible.
- Sample volume for low-BOD samples (e.g., 0–6 mg l<sup>-1</sup> level); a sample volume of at least 1 l is required for the standard test.

No chemical preservation processes can be used on samples requiring a BOD test as these would affect the microorganisms involved. Since any bacterial action on organic waste can be expected to continue after collection of a sample, there could be a significant decrease in BOD values between collection and analysis, depending on the time involved, the organic substrate, and the vigor of the organisms. Hence it is generally recommended that analysis should begin within 2 h of collection. Chilling the sample to 4°C or lower (but not freezing) will slow bacterial action. With chilling, analysis should begin within 6 h and not exceed 24 h. Results with length and temperature of storage are then reported.

For composite samples (e.g., from automatic samplers), additional requirements include limiting the composting period to 24 h and noting the storage time, starting from the end of the composting period, while following the above recommendations for starting analysis and recording temperature and time of storage.

These recommendations are designed to produce a standardized test. It should be noted, however, that there may be a significant decrease in BOD for some very easily degraded wastes (e.g., those containing a high proportion of the organic material as sugar may degrade significantly in 24 h even with chilling). Conversely, waste with a BOD value in the high hundreds or thousands of milligrams per liter in a completely filled airtight container is likely to degrade very little over several days because it has become anaerobic.

Freezing has been proposed as a preservation technique for BOD samples to kill or inactivate microorganisms. While some researchers have found that this gives lower BOD values when compared with BOD analyses carried out soon after sampling, others have noted no significant change. The adaptation of

microorganisms to the organic material and the number of organisms are the factors that influence BOD results, and freezing and reseeded after thawing may change either or both of these.

## Techniques – Standard Procedures

Standard laboratory safety procedures must be followed since many of the chemicals are toxic or hazardous.

Briefly, any chlorine is removed from a water sample to be tested and the sample or a diluted version of it is adjusted to approximately pH 7. The pH buffer and nutrients are added along with a seed and/or nitrification inhibitor if necessary. Two BOD bottles are filled to an extent such that insertion of the stoppers displaces all the air. The initial DO level (DO<sub>i</sub>) of the first bottle is determined using the Winkler titration method (a titrimetric determination) or using one of the modifications of this method (most commonly the azide modification). The second bottle is then incubated for 5 days, at which time the DO level (DO<sub>f</sub>) is determined:

$$\text{BOD} = (\text{DO}_i - \text{DO}_f - \text{seed contribution}) \times \text{dilution factor} \quad [1]$$

where seed contribution = ((volume (ml) of seed added in BOD bottle)/(volume (ml) of BOD bottle)) × BOD of seed.

Only one bottle will be needed for DO<sub>i</sub> and DO<sub>f</sub> if a DO electrode is used. If replicate determinations are carried out, it is customary to carry out a single determination on the initial DO bottle and have replicates for the final DO determinations, which are subject to greater variation; i.e., for Winkler titration determination of DO, three bottles are prepared, one used immediately for DO<sub>i</sub> and the other two incubated and used for DO<sub>f</sub> determination.

Common sources of error are discussed below.

Dirty BOD bottles produce unreliable results, and so residues such as protein, oil, and grease from previous samples and detergent from cleaning of the BOD bottle must be removed.

Water used in the dilution of samples can cause error when copper, forming part of a water distillation unit, contaminates the distilled water, poisoning the microorganisms.

Another possible source of error is organic leachates from deionizing resins. Theoretically these could lead to a positive error in BOD results, but in practice they can lead to low results by poisoning the bacteria. Leachate problems are most likely to occur when

new resins have been installed in a deionizing unit. High-purity water units with organic scrubber cartridges will help solve this problem, but contamination from this source cannot be ruled out completely as cartridges can be overloaded. Standard Methods sets a limit of less than  $0.2 \text{ mg l}^{-1}$  oxygen demand for dilution of water over the 5 day incubation period.

Some likely sources of contamination are biological growths in plastic tubing attached to distillation or deionizing resin units; the deionizing resin units themselves; nutrient containers, especially the buffer solution container; and the storage containers. As far as possible, plastic tubing and storage containers should be replaced by glass equivalents for dilution water.

The presence of chlorine in a sample will cause an error by killing BOD microorganisms. Chlorine can be detected by odor, or by using a sample chlorine water chemistry test kit ( $0.1 \text{ mg l}^{-1}$  reporting limit). Samples will require seeding after dechlorination, as will any samples disinfected by other means.

In filling the BOD bottle with the sample, care is needed to exclude air bubbles, which contribute a nonreproducible source of oxygen to the test; these can be dislodged by tapping the bottle carefully. As 1 ml of air in a 300 ml BOD bottle contributes  $\sim 1 \text{ mg l}^{-1}$  to a DO determination, tiny air bubbles that may be difficult to dislodge and have total volume less than 0.1 ml may be ignored. The pointed end of the ground-glass stopper on a standard BOD bottle minimizes the risk of trapped air.

Samples with DO levels above  $9.2 \text{ mg l}^{-1}$  at  $20^\circ\text{C}$  (i.e., supersaturated with air) may be found. These high levels can be caused by higher DO levels in cold waters that are brought up to  $20^\circ\text{C}$  or through photosynthesis. The samples should be adjusted to  $\sim 9 \text{ mg l}^{-1}$  by aeration at  $20^\circ\text{C}$ .

Some preliminary treatment of the sample may be required.

Where large particulate material is present, it is often difficult to obtain a representative sample. A decision on whether a high-speed blender should be used to produce a more homogeneous sample can only be made after due consideration of the degree of change that the blending will have on the BOD result. Organic particulates contribute to BOD but generally to a smaller degree than dissolved organic material. Thus, while finer organic particles produced by blending are likely to lead to a higher BOD, the degree of increase is likely to be small when viewed in the context of BOD contributions of dissolved and particulate organics. But this may not be true in some cases.

One preliminary treatment of the sample is the addition of a nitrification inhibitor. That suggested

by Standard Methods is 3 mg of 2-chloro-6-(trichloromethyl)pyridine per 300 ml BOD bottle. An alternative inhibitor is *N*-allyl thiourea, added at the rate of  $0.5 \text{ mg l}^{-1}$ .

It is often necessary to add a seed to samples, and seed microorganisms are available from a number of sources, e.g., settled sewage, the wastewater itself, or soil. A new commercial product is dehydrated microbial seed, which contains a mixture of microorganisms and is suitable for most types of domestic and industrial wastewater. As the general recommendation is that fresh seed be prepared for each working day, allowance must be made for the rehydration time. Some products have a recommended rehydration time of 1 h, while others require overnight rehydration. The advantages of the manufactured dehydrated seed over other sources are convenience, greater reproducibility of BOD results, less variance of seed quality with time, and greater likelihood of a high concentration of viable bacterial species.

For all sources of seed (including the manufactured seed) the possibility exists that some wastes will cause poisoning of the microorganisms. Some wastes will have developed microorganisms adapted to the toxic conditions and hence give expected BOD results, but in other wastes the microorganisms will adapt over the period of the BOD test. Because of the lag time involved in adaptation, a lower BOD is obtained than might be expected. If the toxicity is sufficiently acute, a zero or close-to-zero result is obtained.

In BOD dilutions, the following factors must be considered.

Since air-saturated freshwater at sea level and at a temperature of  $20^\circ\text{C}$  contains  $9.1\text{--}9.2 \text{ mg l}^{-1}$  oxygen, samples with BOD levels above  $6\text{--}7 \text{ mg l}^{-1}$  will require dilution (this assumes that the  $\text{DO}_i$  is  $8 \text{ mg l}^{-1}$  or higher).

Dilutions resulting in a residual DO level of at least  $1 \text{ mg l}^{-1}$  and a DO uptake of at least  $2 \text{ mg l}^{-1}$  after the 5 day incubation produce the most reliable results. Hence a single dilution ( $n$  times) will cover a BOD range of  $n$  times ( $2\text{--}7.5 \text{ mg l}^{-1}$ ), if  $\text{DO}_i$  level is  $8.5 \text{ mg l}^{-1}$ .

Dilutions are based on past experience, visual inspection and odor, published suggestions (Table 1), and they are made when using a more rapid technique such as COD. In almost all cases several dilutions are made. Dilution increments of two to three times (e.g., a sample is diluted 8 times, 20 times, 50 times, 125 times) ensure that the various BOD ranges covered by the different dilutions will overlap. Some laboratories estimate the BOD of a given sample and make four, five, or more dilutions to cover BOD levels lower and higher than their estimate. Higher COD results (greater than  $200 \text{ mg l}^{-1}$ ) can be used reliably for calculating appropriate dilutions for the



**Table 1** Suggested dilutions for various waters

	<i>Expected BOD level (mg l<sup>-1</sup>)</i>	<i>Suggested dilution range</i>
Unpolluted river water	0–3	Undiluted
Polluted river waters	5–10	Undiluted to four times
Biologically treated effluent	15–30	4–20 times
Raw and settled wastewater	80–400	20–100 times
Strong industrial wastes	Highly variable <sup>a</sup>	100 times and higher

<sup>a</sup>A COD or other indicative determination is recommended to establish a possible range.

OD test. Two dilutions will be sufficient:

$$\text{higher dilution} = \text{COD}/7 \quad [2]$$

(e.g., COD = 700 mg l<sup>-1</sup>, higher dilution = 100×)

$$\text{lower dilution} = \text{COD}/20 \quad [3]$$

(e.g., lower dilution = (700 mg l<sup>-1</sup>)/20 = 35×).

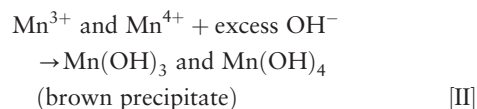
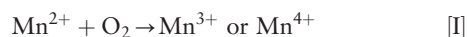
An appreciation that COD can oxidize some inorganic materials and microbially resistant organic materials as well as material oxidized in the BOD process is needed for successful application of the calculation; e.g., coal particles will give a high COD but low BOD.

If a low dilution of less than 10 times is made from a high- or moderately high-BOD sample, the initial DO is likely to be low (less than 8 mg l<sup>-1</sup>). This results from an almost zero sample DO with dilution water having ~9 mg l<sup>-1</sup> DO. A quick calculation with the above assumptions gives some indication of the DO level to be expected. The diluted sample should be aerated to raise the initial DO. The above implies that a higher dilution may be needed since a zero or near-zero DO in the sample suggests the likelihood of a moderately high BOD (though the presence of a chemical reductant or other causes cannot be ruled out).

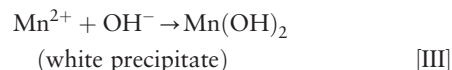
### The Winkler Titration Method

The classical method for DO determination is the Winkler titration method, which is still frequently used as a reference procedure for calibration of DO meters. It is inexpensive in capital cost outlay, requiring only common laboratory glassware (burettes, pipettes, measuring cylinders, conical flasks, volumetric flasks), burette stands, wash bottles, and chemicals. The principal reactions involved are

shown in reactions [I] and [II]:

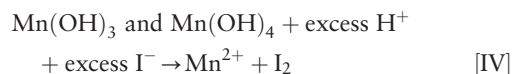


Excess Mn<sup>2+</sup> undergoes reaction [III]:



If the sample is devoid of oxygen at the end of the incubation period, this step in the determination will show a white precipitate. Precipitates ranging from light brown to dark brown indicate low to high oxygen levels. Manganese hydroxide precipitates settle out much more slowly in seawater than in lower-density freshwater.

The second stage of the determination is shown in reaction [IV]:



Iodine is produced in stoichiometric amounts dependent on the amount of Mn(OH)<sub>3</sub> and Mn(OH)<sub>4</sub> present.

Iodine is then titrated with sodium thiosulfate (or the more stable, but poisonous, phenylarsenic(III) oxide (phenylarsine oxide)) using starch as the indicator (reaction [V]):



The starch and iodine form a blue complex. The solution becomes colorless when all the iodine has been reduced to iodide by the thiosulfate. The iodine formed in reaction [IV] will range from a strong yellow color (indicating a high level of oxygen) to a dark brown precipitate from reaction [II] to no iodine (colorless solution), corresponding to zero oxygen.

Aids for BOD determination include syringe-type dispensers for the manganese(II) salt and other solutions, and commercially available digital burettes, which are more convenient than the conventional glass-type ones.

### The Oxygen Membrane Electrode

This is a well-established standard procedure for measuring DO levels. Generally the determination depends on oxygen diffusing through a membrane to an electrode, where it undergoes an electrolytic reduction (by galvanic action or by polarographic

action). The potential or current involved in this is proportional to the rate of oxygen diffusing through the membrane, which in turn depends on the dissolved oxygen concentration in the sample.

There are some practical considerations here. The most suitable DO meters have oxygen probes shaped to fit a standard BOD bottle, or include an adaptor (a funnel) that prevents loss of the sample when the oxygen probe is inserted into the BOD bottle.

In general, electrodes consume oxygen as part of the measuring procedure, but as long as the measuring time does not exceed 1 or 2 min in a 300 ml bottle, this will not affect the results significantly. The rate of oxygen depletion depends on the design and the type of electrode (galvanic or polarographic).

Since most electrodes rely on the rate of oxygen diffusing to the inside of the membrane being proportional to the concentration of oxygen on the sample side of the membrane, anything fouling the membrane, such as oil and grease, will cause a low reading. An air bubble swirling on the membrane or damage to the membrane will also produce an erratic reading. In practice, if the BOD dilutions are close to correct, fouling is rarely a problem.

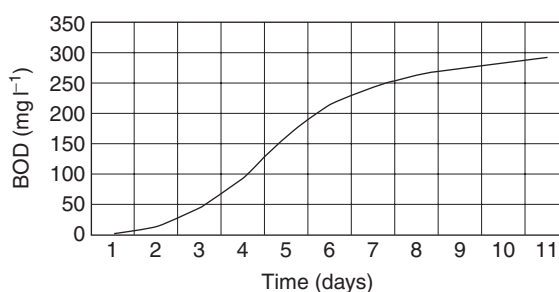
Since the electrodes consume oxygen, localized depletion of oxygen occurs near the membrane. To overcome this, some BOD oxygen probes have built-in stirrers. Other probes require a stirrer bar and an external magnetic motor.

The sensor design consumes oxygen at the cathode like conventional oxygen probes but produces an equivalent amount of oxygen electrolytically at the anode. When these concentrations are equal there is no net consumption of oxygen, but electrolysis proportional to oxygen concentration is still occurring and hence is a measure of the DO. Hence the results are not affected even over prolonged testing. Anything short of complete fouling of the membrane does not affect the final DO reading but does lengthen the time required to reach it.

An advantage of the oxygen probe is that it enables the DO levels to be read before the end of the 5 day period.

For both these methods – Winkler titration and the DO meter method – suspended and settled particulate organic material also contribute significantly to the BOD, though not as readily as dissolved organic matter. In cases where settled material forms a thick layer, dilution allows greater access of the bacteria to the settled material by making more surface area of the material available.

The ratio of the various species of bacteria normally encountered can change in a 5 day period. The BOD is the result of a summation of the oxygen demand of these microorganisms, whose contribution



**Figure 1** Example of S-type curve indicating moderate toxicity, insufficient bacteria, or bacteria acclimatizing to sample.

to the oxygen demand will change with time because of the changing population (implying a possible change in overall reaction rate) and changing feedstock (both in amount and material as it is acted on by bacteria). A well-documented example of the bacterial population changing with time is the increasing presence of nitrifying bacteria in sewage samples.

Material toxic to microorganisms will cause zero BOD results. At sufficiently high dilutions, toxic materials will no longer inhibit microorganisms. The intermediate situation produces increasing BOD results for increasing dilutions. If this is observed, it is often taken to mean that there is toxic material present. There can, however, be other explanations, such as more readily available feedstock or a better food-to-bacteria ratio. Where there is a dramatic change in results with increasing dilutions, such as a 100% increase, the most likely explanation will be the presence of toxic materials.

Another indication of toxic effects can be seen if the oxygen demand is monitored on a daily or more frequent basis. An S-type curve could indicate toxicity (Figure 1). A similar curve also results from a low population of suitable microorganisms for a particular sample. In both cases the 'S' shape is derived from the organisms adapting and increasing in population over a period of 1–3 days.

## Techniques – BOD Biosensor

The conventional BOD test has certain benefits such as being a universal method of measuring most wastewater samples, and furthermore, no expensive equipment is needed. It has, however, the limitation of being time consuming, and consequently it is not suitable for online process monitoring. Thus, it is necessary to develop a measurement method that could circumvent the weaknesses of the conventional method. The need for fast, portable, and cost-effective methods for environmental monitoring has stimulated the development of a variety of field analytical tools such as biosensors. Biosensors are

devices that transduce a selective biochemical response to a measurable signal. Several biosensor methods for BOD measurement have been developed. The first report of a BOD biosensor was published by Karube *et al.* in 1977. After that, several kinds of microbial BOD sensor have been developed, and various modifications have been carried out (Table 2). Most of the BOD sensors referred to above consisted of a synthetic membrane with a single or a random combination of immobilized microorganisms serving as a biocatalyst. A rapid and reliable BOD sensor should be capable of analyzing a sample of complex constituents with relatively low selectivity. Thus, the sensor should respond to all kinds of biodegradable organic solutes in the samples. It is also important that the sensor should give results comparable with those obtained using the conventional BOD method.

Most of the previously reported BOD sensors are biofilm-type whole cell-based microbial sensors, which rely on measuring the bacterial respiration rate in close proximity to a suitable transducer. A common feature of these sensors is that they consist of a microbial film sandwiched between a porous cellulose membrane and a permeable membrane as the biological recognition element. This microbial film is an immobilized microbial population that can biooxidize the organic substrate to be quantified. The response is usually a change in concentration of the DO or other phenomena such as light emission. A physical transducer is used to monitor this process. The result is a change in an electrical or optical signal. The signal is amplified and correlated to the content of biodegradable material present.

A biofilm sensor for BOD consists of a DO probe such as an oxygen-permeable membrane, with another membrane containing bacteria (the biofilm) between the membrane and the sample (Figure 2). Organic material diffuses into the biofilm where the bacteria act on it, causing a drop in oxygen levels that is measured by the probe. The probe is calibrated by setting a baseline level ('zeroing') with a solution containing no organic material, followed by measuring the steady-state response in a solution of known BOD – normally 150 mg glucose plus 150 mg glutamic acid in 1 l of distilled water taken as  $\sim 200 \text{ mg l}^{-1}$  BOD. The response of unknown samples is compared with the standard to obtain the BOD result. The times quoted to reach a steady reading vary from under 10 to 30 min.

A related technique utilizing a fluidized bed (in place of the biofilm) with DO probes has been incorporated into a commercially available instrument.

Another method uses a biofilm as in the previous method, but instead of waiting for a steady-state

reading, it measures the rate of change of oxygen uptake by the bacteria, obtaining results in  $\sim 1$  min.

These methods are similar to a BOD test, using bacterial activity for their measurement and consequently responding strongly to easily biodegradable material. Many researchers use a monoculture; how well this approximates a mixed culture's response will depend mainly on the sample. Some researchers have had trouble obtaining reproducible results from sensors that use mixed microorganisms from activated sludge.

Apart from the effects of dilution on BOD, biofilm sensors cannot include degradation of organic particulate matter to any significant extent since the bacteria are immobilized within the sensor system. It has been shown, however, that biofilms can be acclimatized to different substrates (wastewater as well as pure compounds) to give improved responses.

For low-BOD samples, the kinetic mode has not been shown to be as sensitive as either of the conventional BOD methods or the steady-state biofilm method. Both biosensor techniques share with oxygen electrodes the danger of fouling of the biosensor membrane with oil. Because biofilm techniques allow rapid process control, they are most likely to find application in monitoring a consistent organic process waste.

Because of the importance of biofilm technology in waste processing, it is likely that the research into it will carry over to BOD biofilm sensors.

A comparison of the processes involved in the Winkler titration and DO meter methods with alternative methods illustrates situations where they may not produce equivalent results. For example, where oil and grease are present in water, dilution allows greater accessibility by bacteria owing to the presence of finer oil droplets. In this case a higher BOD result is obtained using the Winkler and DO methods than is obtained by the lower-dilution manometric methods.

## Techniques – Other Procedures

How well a particular alternative technique approximates the BOD test for a given sample will determine how successful the alternative technique is in providing BOD results. And although an alternative technique may not give the same result as a conventional BOD test, it can provide information that may be more appropriate.

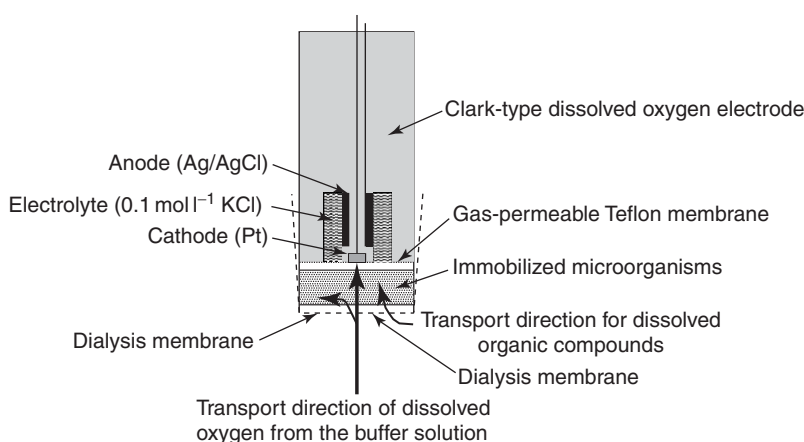
While the above methods are the most commonly used for BOD determination, there are two commercially available alternatives based on supplying extra oxygen from an airspace above a sample or oxygen

**Table 2** Commercially available BOD instruments

Model	BOD-2000 BOD-3000	DKK <sup>TM</sup> BOD sensor 7842	BOD-point	BSB-model	ARAS	BIOX-1010 <sup>a</sup>	ROD TOX 2000	BOD-biomonitor	QBOD metre and EZ- BOD metre <sup>b</sup>	RACOD <sup>TM</sup> meter
Manufacturer	Nisshin Denki & Central Kagaku Co. Ltd., Tokyo, Japan	DKK Corporation, Japan	Aucoteam FmbH, Berlin, Germany	Prufgeratewerk FmbH, Berlin, Germany	Dr. Lange GmbH, Berlin, Germany	STIP Isco GmbH, Groß-Umstadt, Germany	Kelma, Belgium	LAR Analytik & Umweltmesste- chnik GmbH, Berlin, Germany	Bioscience, Inc., Bethlehem, USA	Usfilter, Vineland, NJ, USA
System configurations	Biofilm type, flow- through system	Biofilm type, flow injection with 3 ml measuring chamber	Biofilm type, flow-through system	Biofilm type, flow-through system	Biofilm type, with 2 ml stirred measuring chamber	A bioreactor combined with a dilution system	Respirometer type (BOD and toxicity analyzer)	Respirometer type	Respirometer type (bioreactor)	Respirometer type
Microbial species	<i>Trischosporum cutaneum</i>	<i>T. cutaneum</i>	<i>T. cutaneum</i> , <i>Canelida parapsilosis</i>	<i>T. cutaneum</i> , <i>C. parapsilosis</i>	<i>Rhodococcus erythropolis</i> + <i>Issatchenkia orientalis</i>	Bacteria isolated from wastewater	Activated sludge	Activated sludge	Activated sludge	Activated sludge
Measuring time (min)	20–40	5	< 1	3 or < 1	1–3	3–15	20–40	3–5	20 <sup>c</sup> , 15–60 <sup>a</sup>	10 (high range) or 30
Measuring ranges (mg l <sup>-1</sup> BOD)	0–100, 0–200, 0–500 <sup>d</sup> , 3–1000 <sup>e</sup>	0–60	5–500	0–22 or 2–33	2–300	5–1500, 20–1500, 20–100 000	0–500 000	0–50, 0–200 000	0.5–300 or 0.5–5000 <sup>c</sup>	100–4000 (higher range) or 0–100 (low range)
Working stability (days)	30 <sup>d,e</sup>		30	30	30		14		60–90 <sup>d</sup>	
Precision (± %)	3 <sup>d,e</sup>	5	< 10	< 10	< 5	3	< 5		10 (low range) <sup>c</sup> or 5 (high range) <sup>c</sup>	
Calibration standard	Gga	Gga	Glucose	Glucose	Glycerol		Stabilized wastewater of the plant being monitored			

<sup>a</sup> Measuring principle based on dynamic dilution of two gear pumps, depending on the O<sub>2</sub> respiration of microbes.<sup>b</sup> EZ-BOD meter.<sup>c</sup> QBOD meter.<sup>d</sup> BOD-2000.<sup>e</sup> BOD-3000.

Gga, glucose/glutamic acid



**Figure 2** Schematic presentation of a BOD biosensor.

produced electrolytically, and there are two biofilm techniques that are not commercially available. Two other developments based on the conventional techniques – robotics and a graphical method for determining BOD – are discussed.

#### **Manometric Method – Oxygen from Airspace above Sample**

In this method a manometer reads the pressure of the airspace, isolated from the atmosphere, above the sample or diluted sample to be tested. Agitation keeps the sample supplied with oxygen from the airspace, and carbon dioxide is removed by absorption in potassium hydroxide solution. The manometer measures the pressure drop in the vessel caused by the consumption of oxygen in the BOD process. As the amount of oxygen available from the airspace is considerably higher than that available from the water, the BOD range covered by a single dilution is commensurately higher (one manufacturer quotes a range of 0–350 mg l<sup>-1</sup> without dilution).

Different BOD ranges can easily be set by changing the amount of sample (and hence the sample-to-air ratio) in the vessel. The manufacturers' recommended minimum and maximum volumes of a diluted sample set the BOD range. The minimum sample volume (giving the highest range) is not always appropriate as it causes a loss in resolution.

An advantage of this system over the conventional BOD system (Winkler titration method) and the DO probe is that it can be used to monitor changes in DO levels with time. The procedures for ensuring that sufficient nutrients are available for the 5 days of the tests are not as rigorous as with the standard procedures, however, and this could lead to low BOD results. Additionally, air could leak into the manometer system, causing errors.

A further disadvantage is the limitation on the number of samples that can be processed at any time unless multiple units are purchased.

#### **Manometric Method – Oxygen Supplied Amperometrically**

This is similar in essence to the manometric method discussed above. A sample vessel is fitted with a manometric switch and an electrolytic oxygen generator. As oxygen is consumed by microorganisms and carbon dioxide is removed by potassium hydroxide, a drop in pressure occurs and the manometer triggers production of oxygen to replace the oxygen consumed. The amount of electricity used to generate the oxygen is measured and is proportional to the BOD. The advantages of this method over all the preceding methods are that an even greater BOD range can be achieved before dilution is needed and that the system can be adapted easily for continuous monitoring or readings at fixed time intervals. For one of the commercially available instruments a specification of 60 mg O<sub>2</sub> per hour maximum demand is given. On the assumption that the oxygen demand follows an approximately first order-type curve for the 5 days of the test, the above specification would mean that dilutions would be needed when the BOD<sub>5</sub> levels reach the thousands of milligrams per liter range.

This system has the same advantages and disadvantages as the first manometer method discussed. For both, the comments relating to the effect that dilution can have on a BOD result are appropriate.

#### **Robotics**

The most persistent criticisms of the BOD test are that it is time consuming and labor intensive. In response to the latter claim, laboratories and instrument manufacturers have developed BOD robotic



systems using DO meters for reading the initial and final DO levels. The most likely sources of error from the robotic system are nonrepresentative aliquots of samples and failure to detect bubbles trapped beneath the sensor. As the robotic systems are not yet as reliable as autosamplers, some manual intervention is generally needed.

### Graphical Method

The BOD can be determined graphically by having a close range of dilutions to produce a graph of DO ( $\text{mg l}^{-1}$ ) remaining versus volume of sample in the BOD bottle, i.e., volume of sample (ml) diluted to 300 ml:

$$\text{BOD} = -300 \times \text{slope} - y \text{ intercept} + S \quad [4]$$

where 'slope' is the slope of the best-fitting straight line of the above plot, the  $y$  intercept is the  $y$  intercept of the best-fitting straight line of the above plot, and  $S$  is the DO ( $\text{mg l}^{-1}$ ) of the undiluted sample.

With good quality data the method is capable of a better blank and seed correction than that of the calculation procedure of Standard Methods. It can also make redundant the determination of  $\text{DO}_i$ . With this method dilution water that does not meet the Standard Methods criterion of having a demand not more than  $0.2 \text{ mg l}^{-1}$  over the 5 day incubation period can still be used. With the close dilutions used acting as surrogate replicates, this offers reassurance that BOD results are correct. Similar assurance is, however, gained from two or more dilutions and replication in the conventional BOD procedure.

The graphical method does not take into account that there may be oxygen-consuming compounds present that give zero sample DO and consume oxygen from the dilution water, i.e.,  $S$  may effectively be less than zero.

If there is a scatter of points in the plot, some bias in the slope may occur, producing bias in the BOD result. A large number of dilutions are required for the graphical BOD procedure.

### Applications

The BOD test has traditionally been used to measure waste loadings ahead of and after a sewage treatment works or waste-processing plant to determine the efficiency of their biological treatment. It is one of the important criteria used in the design of organic waste treatment plants and is used in the assessment of polluted waters and wastewater. More recently it has been used in government regulatory control of pollution and disposal of water.

As a biological test it is likely to give a more meaningful indication of the impact of waste on oxygen levels in a stream or river than will a purely

chemical test and is therefore an important water quality parameter.

Although it has been used in process control, the time delay makes it less efficient than alternative methods.

**See also:** **Water Analysis:** Overview; Freshwater; Sewage; Organic Compounds; Chemical Oxygen Demand; Microbiological.

### Further Reading

- American Public Health Association, American Water Works Association, and Water Environment Federation (1998) *Standard Methods for the Examination of Water and Wastewater*, 20th edn. Baltimore: American Public Health Association.
- Analytical Quality Control (Harmonised Monitoring) Committee (1985) Accuracy of determination of biochemical oxygen demand in river waters: Analytical quality control in the Harmonised Monitoring Scheme. *Analyst* 110: 97–102.
- Fitzmaurice GD and Gray NF (1989) Evaluation of manufactured inocula for use in the BOD test. *Water Research* 23: 655–657.
- Fresenius W, Quentin KE, and Schneider W (1988) *Water Analysis*. Berlin: Springer.
- Hawkins J and Lloyd A (1989) The application of robotics in the water industry and the development of a BOD robot. *Water Science and Technology* 21: 937–945.
- Karube I, Matsunga T, and Mitsuda S (1977) Microbial electrode bod sensors. *Biotechnology and Bioengineering* 19: 1535–1547.
- Kumar R (2003) BOD analysis of industrial effluents: 5 days to 5 min. *Current Applied Physics* 3: 191–194.
- Liu J and Mattiasson B (2002) Microbial BOD sensors for wastewater analysis. *Water Research* 36: 3786–3802.
- LeBlanc PJ (1974) Review of rapid BOD test methods. *Journal of the Water Pollution Control Federation* 46: 2202–2208.
- Phelan DM, Taylor RM, and Fricke S (1982) A maintenance-free dissolved oxygen monitor. *International Laboratory* 12(7): 60–75.
- Ranchet J, Pescheux F, and Menissier F (1981) Influence du mode et du temps de conservation des échantillons d'eaux sur les déterminations de  $\text{DBO}_5$  DOC et MES. *Techniques et Sciences Municipales – Eau* 76: 547–551.
- Riedel K, Lange K-P, Stein H-J, et al. (1990) A microbial sensor for BOD. *Water Research* 24: 547–551.
- Rastogi S, Rathee P, Saxena TK, Mehra NK, and Kumar R (2003) BOD analysis of industrial effluents: 5 days to 5 min. *Current Applied Physics* 3: 191–194.
- Stand SE and Carlson DA (1984) Rapid BOD measurement for municipal wastewaters samples using a biofilm electrode. *Journal of the Water Pollution Control Federation* 56: 464–467.
- Woodring SL and Clifford DA (1988) A comparison of the graphical and standard methods for the determinations of biochemical oxygen demand. *Journal of the Water Pollution Control Federation* 60: 537–542.



## Chemical Oxygen Demand

**Z Hu**, Cornell University, Ithaca, NY, USA

**D Grasso**, Smith College, Northampton, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

In addition to biochemical oxygen demand (BOD), chemical oxygen demand (COD) is widely used as a surrogate measure for carbon bioavailability. Compared to the BOD test, which requires days, the COD test is designed to yield results in a much shorter time (within hours). The COD test uses strong oxidants to oxidize organic matter that microorganisms may oxidize only partially or not at all. The sample COD is finally calculated by comparing with standards that are typically used, e.g., phthalate, oxalate, and glucose. In this article, standard reflux methods with dichromate digestion are reported, although determination of COD based on permanganate index is still in practice for natural waters. Due to some limitations and disadvantages of the standard methods, other modified methods are also presented. These include: the replacement of hexavalent chromium, mercury, and silver metals; microwave digestion; automation and online COD measurement; and electrochemical oxidation to measure the sample COD. Determination of the colloidal COD fraction is also discussed and applications of COD tests are presented in light of water and wastewater analysis.

### Background

COD is defined as the amount of oxygen equivalents consumed in the chemical oxidation of organic matter by strong oxidant (e.g., potassium dichromate). The COD test consists of refluxing a sample for 2 h in the presence of a known amount of oxidant. The concentration of organic matter in terms of oxygen equivalents can be determined from the difference of initial and remaining oxidant concentrations in the sample.

Most organic matter can be oxidized by strong oxidants, although straight chain carboxylic acids may not be oxidized in the absence of a silver sulfate catalyst. Silver sulfate is, therefore, added during COD tests to facilitate the complete oxidation. However, the chloride ion, a common aqueous constituent, reacts with silver ions to precipitate silver chloride, and thus eliminates the catalytic activity of silver yielding a negative interference. Alternatively, chloride, bromide, or iodide can react with dichromate

to produce the elemental form of the halogen, yielding an overestimate of COD. Hence, mercury sulfate is added to minimize reaction interference.

Both organic and inorganic components are oxidized during COD tests. If only the COD associated with the organic component is desired, provisions need to be made to eliminate contributions from oxidation of inorganic components. For instance, chloride interference is removed by  $\text{Hg}^{2+}$  complexing of  $\text{Cl}^-$ . Corrections for chloride interference vary as elemental chlorine may react with ammonia and its derivatives in the sample even though ammonia and its derivatives are not oxidized. Nitrite ( $\text{NO}_2^-$ ) has a COD of  $1.1 \text{ mg O}_2 \text{ mg}^{-1} \text{ NO}_2^- \text{-N}$ . To eliminate its interference, nitrite is converted by sulfamic acid with the addition of 10 mg sulfamic acid for each milligram  $\text{NO}_2^- \text{-N}$  present in the sample. Finally, separate determinations of other reduced inorganic species (e.g., ferrous iron and sulfide) are needed if the samples contain significant levels of these ions. Stoichiometric oxidation of the known reduced inorganic species can be assumed and corresponding corrections are made.

### Sampling

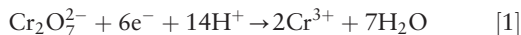
Samples are commonly stored in glass bottles, although it is still possible for a trace of organic matter to sorb to the glassware. A standard sample preparation practice to avoid trace organic contamination may, therefore, apply by baking out glass bottles at  $400^\circ\text{C}$  for at least 1 h. If a sample contains significant amount of particulates, it should be blended first in order to obtain a representative aliquot. All samples should be analyzed as soon as possible. If delay is unavoidable, samples may be stored by acidification to  $\text{pH} \leq 2$  with concentrated sulfuric acid.

If only the soluble COD is desired, the sample should be coagulated with lanthanum chloride ( $\text{LaCl}_3$ ) and filtered through a  $0.45 \mu\text{m}$  membrane to eliminate colloidal and particulate fractions. The filtrate can then be subjected to standard COD analyses.

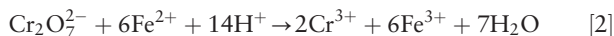
### Standard Methods

Most types of organic matter (electron donor) are oxidized to carbon dioxide and water by boiling a mixture of dichromate and sulfuric acid. Samples are refluxed in a strong acid with a known excess of potassium dichromate ( $\text{K}_2\text{Cr}_2\text{O}_7$ ). Dichromate ( $\text{Cr}^{6+}$ , orange color) is reduced to chromate ( $\text{Cr}^{3+}$ , green

color) via the following reaction:



By using a titrimetric method, the remaining dichromate is back-titrated with ammonium iron(II) sulfate (ferrous ammonium sulfate (FAS)) to determine the amount of  $\text{Cr}_2\text{O}_7^{2-}$  consumed:



The amount of oxidized organic matter is calculated in terms of oxygen equivalents (1 equivalent weight electron equals 8 g of oxygen) with the following equation:

$$\text{COD as mg O}_2 \text{ l}^{-1} = (A - B) \times M \times 8000/V \quad [3]$$

where A = FAS used for blank (ml), B = FAS used for sample (ml), M = molarity of FAS, and V = sample volume (ml).

By using a colorimetric method, the formation of  $\text{Cr}^{3+}$  (green color) is quantified as the oxidation proceeds and the absorbance is read at 600 nm. Alternatively, the disappearance of the orange  $\text{Cr}_2\text{O}_7^{2-}$  can be determined at 420 nm. The sample COD is determined colorimetrically at 600 and 420 nm, respectively, based on standard COD calibration curves.

Most organic compounds are oxidized with the dichromate reflux method under standard procedure, but ammonia and organic nitrogen are not oxidized by dichromate in the absence of significant concentration of elemental chlorine. Hence, the dichromate reflux method is preferred over procedures using other oxidants because of its superior oxidation ability. There are, however, several potential sources of limitation in the test including the following:

- Pyridine and related aromatic compounds resist oxidation.
- Straight chain aliphatic compounds are not oxidized in the absence of  $\text{AgSO}_4$  and may not be completely destroyed even in the presence of  $\text{AgSO}_4$ .
- Volatile organic compounds may escape and are oxidized only to the extent that they remain in contact with the oxidant.

### Open Reflux Method

For an open reflux method, the standard procedure follows: pipette 50 ml sample (for samples with COD higher than  $900 \text{ mg O}_2 \text{ l}^{-1}$ , use a small sample portion diluted to 50 ml) in a 500 ml refluxing flask. Add 1 g  $\text{HgSO}_4$ , and glass beads. Slowly add 5 ml sulfuric acid–silver sulfate mixture (10–15 g  $\text{Ag}_2\text{SO}_4$  per

milliliter concentrated  $\text{H}_2\text{SO}_4$ ) while gently mixing to dissolve use  $\text{HgSO}_4$  and avoid possible loss of volatile organic components. Add 25.00 ml of  $0.0417 \text{ mol l}^{-1} \text{ K}_2\text{Cr}_2\text{O}_7$  and mix. Attach the flask to a condenser and turn on cooling water. While continuously stirring, add the remaining 70 ml sulfuric acid–silver sulfate mixture through the open end of the condenser. Heat the sample for 2 h in the flask under reflux conditions. Wash down the condenser with de-ionized (DI) water into the flask and cool to room temperature. Titrate excess  $\text{K}_2\text{Cr}_2\text{O}_7$  with standard  $0.25 \text{ mol l}^{-1}$  FAS using two to three drops ferroin indicator to determine the volume of FAS utilized. Acquire a blank sample (50 ml DI water) through the same procedure to determine the value of A. For quality assurance and quality control purpose, conduct a test on a standard potassium hydrogenphthalate (KHP) solution ( $500 \text{ mg O}_2 \text{ l}^{-1}$ ).

The above procedure is suitable for a sample with a COD in the range of 0–900  $\text{mg O}_2 \text{ l}^{-1}$ . For high COD ( $> 1000 \text{ mg O}_2 \text{ l}^{-1}$ ) samples, dilution may be necessary. Selection of a dilution ratio can be determined from the color after mixing the sample with dichromate and sulfuric acid. If the sample solution becomes bluish green or green within a few minutes, a 10-fold dilution is suggested. Solutions showing predominantly orange color with little change within minutes should not require dilution.

For low COD ( $< 50 \text{ mg O}_2 \text{ l}^{-1}$ ) samples, the standard procedure is still valid by using dichromate and FAS at 1/10 the concentration given above (i.e.,  $0.00417 \text{ mol l}^{-1} \text{ K}_2\text{Cr}_2\text{O}_7$  and  $0.025 \text{ mol l}^{-1}$  FAS). Great care must be taken with this procedure because even slight contamination of organic compounds on glassware can lead to significant error. If a further increase of sensitivity is required, a large volume of sample may be concentrated to 150 ml (equivalent to the total volume in the standard procedure) after the addition of all reagents by boiling in the refluxing flask open to the atmosphere without the condenser attached. Before refluxing, the amount of chloride must be determined in order to add sufficient  $\text{HgSO}_4$  on the basis of a weight ratio of 10:1 for  $\text{HgSO}_4:\text{Cl}^-$ . A blank should be carried out through the same procedure. Compared to ordinary evaporative concentration methods, this technique has the advantage of reducing losses of easily digested volatile organic compounds.

### Closed Reflux Methods

Compared to the open reflux method, volatile organic compounds are more easily oxidized with closed reflux methods, because those compounds remain in contact with the oxidant in closed vessels.

Furthermore, closed reflux methods are preferred because fewer reagents are used, and therefore less hazardous wastes are generated. The methods are equivalent to ISO 15705–Determination of the chemical oxygen demand index (ST-COD) – Small-scale sealed-tube method.

**Closed reflux: titrimetric method** The procedure is a scaled-down adaptation from the one described above with the same chemistry, and is applicable to samples with COD between 40 and 400 mg O<sub>2</sub> l<sup>-1</sup>. Samples with higher COD values (>400 mg O<sub>2</sub> l<sup>-1</sup>) must be diluted while those with lower COD values (<40 mg O<sub>2</sub> l<sup>-1</sup>) may be determined by using more dilute dichromate and FAS solutions. Since small sample volumes are used, diluted dichromate (0.0167 mol l<sup>-1</sup>) and FAS (0.10 mol l<sup>-1</sup>) are prepared, and it is critical to obtain accurate volumes for samples and digestion solutions. Digestion vessels with premixed reagents and other accessories are commercially available. Otherwise, sample volumes and reagent quantities for various digestion vessels are suggested below. The digestion vessels are placed on a block heater preheated to 150°C for 2 h. The remaining K<sub>2</sub>Cr<sub>2</sub>O<sub>7</sub> is determined by titrating with standard 0.10 mol l<sup>-1</sup> FAS. A blank containing the water reagents and DI is examined the same way for samples and the COD is calculated using eqn [3].

**Closed reflux: colorimetric method** The colorimetric method is another standard method to quantify COD values based on the change of chromate and dichromate concentrations. The chromate ion absorbs strongly in the 600 nm region while the dichromate ion has almost zero absorption. Therefore, samples with high COD (100–900 mg O<sub>2</sub> l<sup>-1</sup>) values convert sufficient amount of Cr<sup>6+</sup> to Cr<sup>3+</sup>, which can be determined in the 600 nm region. On the other hand, the dichromate ion absorbs strongly in the 420 nm region while the chromate ion absorbs weakly at this wavelength. Hence, samples with low COD

(<90 mg O<sub>2</sub> l<sup>-1</sup>) values can be determined by measuring the decrease of absorbance at 420 nm, which directly correlates with Cr<sup>6+</sup> concentration.

A calibration curve must be prepared before COD can be determined with this colorimetric method. At least 5 standards of KHP solution should be prepared to cover a range of sample COD values. Aliquots of standard KHP stock are diluted with DI water to the same volume for samples. A reference solution (blank) contains DI water only. The same reagent volumes, tube, and digestion procedure as for samples are followed (Table 1). Digestion vessels containing a blank, standards, and samples are placed on a block heater preheated to 150°C for 2 h. Samples are cooled to room temperature and their absorbances are measured against the blank. A linear calibration curve is created by measuring the difference between absorbencies of digested standards and the digested blank. COD values are determined based on calibration curves.

## Comparison of Chemical and Theoretical Oxygen Demand

Ideally, COD values obtained from the standard reflux methods should equal the solution's theoretical oxygen demand (ThOD), which is the amount of oxygen required to stoichiometrically oxidize compounds to end products, including CO<sub>2</sub>, NH<sub>3</sub>, and H<sub>2</sub>O. Examination of a database of 565 organic compounds has recently showed, however, that an average 85% of the ThOD value is obtained from COD tests due to the incomplete oxidation of some organic compounds by dichromate. Some halogenated aromatic and aliphatic hydrocarbons have especially low COD/ThOD ratios (as low as 0.03). The low COD/ThOD ratios also apply to many aliphatic alkane compounds with the lowest value down to 0.01. If a water or wastewater sample contains a large fraction of such refractory organic compounds, standard COD tests may be an inappropriate surrogate to

**Table 1** Suggested digestion vessels and sample and reagent volumes in COD tests

Digestion vessel	Sample volume (ml)	H <sub>2</sub> SO <sub>4</sub> volume (ml)	K <sub>2</sub> Cr <sub>2</sub> O <sub>7</sub> volume (ml)	H <sub>9</sub> SO <sub>4</sub> (g)
500- or 250-ml Erlenmeyer flasks	50	75	25	1
16 × 100 mm capped tube	2.5	3.5	1.5	0.05
20 × 150 mm capped tube	5.0	7.0	3.0	0.1
25 × 150 mm capped tube	10.0	14	6.0	0.2

represent the degree of contamination. Alternatively, total organic carbon analyses may be performed to measure the total organic contents in the samples.

## Modifications of Standard Procedures

In addition to the incomplete oxidation of some chemicals with the reflux methods, laboratory wastes generated from the standard COD tests contain hexavalent chromium, mercury, and silver metals, all of which are classified as hazardous wastes by the US Environmental Protection Agency and their disposal is regulated under Resource Conservation and Recovery Act. Consequently, various modifications of the standard procedures or alternative methods have been reported for the COD test. These include: the replacement of hexavalent chromium, mercury, and silver metals; microwave digestion; automation and online COD measurement; and electrochemical oxidation to measure the sample COD.

### Elimination of Mercury and Chromate

Mercury is used in a standard COD test to suppress the chloride interference. Mercury sulfate can be used up for samples with chloride values as high as  $2000 \text{ mg l}^{-1}$ . For samples containing more than  $2000 \text{ mg l}^{-1}$ , however, various chloride removal methods may be considered. These include precipitation as silver chloride, ion exchange, addition of  $\text{Cr}^{3+}$  as a complexing agent, and oxidation of  $\text{Cl}^-$  to chlorine by pentavalent bismuth ( $\text{Bi}^{5+}$ ). Of these methods, the removal of chloride by  $\text{Bi}^{5+}$  proposed by Miller *et al.* is novel and promising. In this method, solid sodium bismuthate containing pentavalent bismuth oxidizes the chloride rapidly without affecting sample organics according to the following reaction:



A special chloride removal cartridge containing a glass-fiber filter (upper cartridge, for removal of particulates) and a column packed with a mixture of solid sodium bismuthate and an inert, free-flow agent (lower cartridge) is used. The acidified sample is drawn through the chloride removal cartridge under carefully controlled conditions of contact time, flow rate, and acid strength to allow chloride and bismuthate to react effectively without destroying sample organics. Solid sodium bismuthate is not soluble in the acidified solution and remains in the cartridge. The by-product, trivalent bismuth ( $\text{Bi}^{3+}$ ), dissolves in the acidified sample solution but has no effect on COD results. The glass-fiber filter containing particulates is removed from the top of the cartridge and combined with the treated liquid component in a COD reagent vial for

total COD measurement. Alternatively, soluble and particulate COD can be determined separately by digesting only the sample component of interest.

Permanganate ( $\text{Mn}^{7+}$ ) and trivalent manganese ( $\text{Mn}^{3+}$ ) have been proposed to replace hexavalent chromium as strong oxidants in COD analysis. In fact, the permanganate method is widely used in Japan to monitor organic pollution. Oxidation potentials of  $\text{Mn}^{3+}-\text{Mn}^{2+}$  and  $\text{Mn}^{7+}-\text{Mn}^{2+}$  half-reactions are 1.54 and 1.51  $E^\circ$  (V), respectively, both of which are greater than that of  $\text{Cr}^{6+}-\text{Cr}^{3+}$  half-reaction (1.35  $E^\circ$  (V)). Trivalent manganese ( $\text{Mn}^{3+}$ ) is a strong, non-carcinogenic chemical oxidant that changes quantitatively from purple to faint pink when it reacts with organic matter. Sample COD values are determined colorimetrically, and the color intensity is inversely proportional to the amount of COD in the sample. Unfortunately, the permanganate and trivalent manganese methods have relatively low oxidation power and poor reproducibility. Manganese ( $\text{Mn}^{3+}$ ) in sulfuric acid can oxidize  $\sim 80\%$  of synthetic and naturally occurring organic compounds.

### Microwave Digestion

The standard reflux methods require a long time (2 h) for the digestion step in the COD analysis. A rapid COD determination technique using microwave digestion has been developed. One of the advantages of using microwave digestion over the standard reflux method is that compounds such as pyridine and aromatic organic compounds are oxidized better. A special microwave with six individual magnetrons is designed to focus the microwave radiation on each sample in a 250 ml glass digestion tube. Each digestion tube is connected to a condenser circulating with cooling water to avoid any loss of volatile organic compounds in the sample.

An aliquot (20 ml) of a sample is added in the glass digestion tube, followed by the addition of 0.5 g of mercuric sulfate, 10 ml of dichromate solution, and 5 ml of sulfuric acid. The tube is placed into the microwave and connected to a condenser. An additional volume (25 ml) of sulfuric acid is added to the top of the condenser. The mixture is digested in the microwave at  $150^\circ\text{C}$  for 8 min. Excess dichromate in the digestion tube is determined with the standard titrimetric or colorimetric procedure. COD determination of real water and wastewater samples indicates that the results obtained from both microwave digestion and closed reflux digestion are consistent.

### Automation and Online COD Measurement

Standard reflux methods require 2 h of digestion, a length of time that would eliminate the benefits of



online COD measurement. With the development of microwave digestion and innovative thermal oxidation techniques, attempts to automate COD analysis in continuous flow systems (also called flow-injection analysis, FIA) have shortened the analysis period to 15 min.

One flow-injection method based on microwave digestion techniques is as follows. A sample is taken with an autosampler and mixed with reagents from a fast loop arrangement into a microwave heated chamber, where it is digested in an acid solution containing mercuric sulfate, dichromate solution, and sulfuric acid for 5–15 min. The process and system design ensures the constant sample volume and the reduced chromate ( $\text{Cr}^{3+}$ ) with green color is determined colorimetrically to give accurate COD values.

Other FIA methods have been developed by using innovative thermal oxidation techniques. The addition of a strong oxidizing agent, cerium(IV) sulfate, results in a high degree of sample oxidation. Potassium permanganate may also be used as both an oxidant and a colorimetric reagent during a segmented FIA to determine COD in aqueous environmental samples. But, it may not apply for analysis of samples containing abundant recalcitrant organic matter due to its limited oxidation power.

### COD Measurement with Electrochemical Oxidation

COD may also be determined with electrochemical methods. An electrochemical cell with the capacity for electrochemical oxidation of organic matter into water and carbon dioxide might enable a fast method to measure the sample COD based on coulometry (the number of electrons consumed in the degradation). The system includes a copper electrode, a potentiostat, and a personal computer. Copper in alkaline media can act as an electrocatalyst to oxidize organic matter. The potentiostat is used to maintain constant potentials during electrolysis. The data are recorded with a data acquisition system. The net charge, corresponding to the number of electrons consumed in the electrochemical oxidation, is correlated to the COD determined by standard reflux methods. The time required for a single sample COD measurement is ~30 min, which is much less than the 2–4 h required by the standard methods. However, electrochemical methods may not provide sufficient oxidation power to completely electrolyze all the organic matter. Furthermore, the electrodes may be easily contaminated with significant fouling, especially in a sample containing high concentrations of humic acid.

### Applications

COD is often used as a measurement of pollutants in water, wastewater, and aqueous hazardous wastes.

One application of the COD test is to measure soluble COD in wastewater, since characterization of total COD in wastewater is critical for accurate modeling of biotransformation in wastewater treatment processes.

Another application is to rapidly infer biodegradability of samples from the COD tests. This is commonly accomplished by establishing a correlation between COD and BOD. This method has been found successful when the proportions and types of materials in a wastewater remain relatively constant. For example, BOD in domestic wastewater samples from 5-day BOD tests can be approximated as  $\text{BOD}_5 = 0.476 \text{ COD}$ .

**See also:** **Sample Dissolution for Elemental Analysis:** Microwave Digestion. **Sampling:** Theory. **Water Analysis:** Overview; Seawater – Dissolved Organic Carbon; Industrial Effluents; Sewage; Biochemical Oxygen Demand.

### Further Reading

- APHA, AWWA, WEF (1998) *Standard Methods for the Examination of Water and Wastewater*. Washington, DC.
- American Society for Testing and Materials (1995) *Standard Methods for Chemical Oxygen Demand (Dichromate Oxygen Demand) of Water*. Philadelphia, PA.
- Baker JR, Milke MW, and Mihelcic JR (1999) Relationship between chemical and theoretical oxygen demand for specific classes of organic chemicals. *Water Research* 33: 327–334.
- Balconi ML, Borgarello M, Ferraroli R, and Realini F (1992) Chemical oxygen demand determination in well and river waters by flow-injection analysis using a microwave oven during the oxidation step. *Analytica Chimica Acta* 261: 295–299.
- Baumann FI (1974) Dichromate reflux chemical oxygen demand: A proposed method for chloride correction in highly saline waters. *Analytical Chemistry* 46: 1336–1338.
- Chen S, Tzeng J, Tien Y, and Wu L (2001) Rapid determination of chemical oxygen demand (COD) using focused microwave digestion followed by a titrimetric method. *Analytical Sciences (Japan)* 17: 551–553.
- Hu Z, Chandran K, Smets BF, and Grasso D (2002) Evaluation of a rapid physical-chemical method for the determination of extant soluble COD. *Water Research* 36: 617–624.
- ISO 6060 (1989) Water Quality – Determination of the Chemical Oxygen Demand.
- ISO 8467 (1993) Water Quality – Determination of Permanganate Index.
- ISO 15705 (2002) Water Quality – Determination of the Chemical Oxygen Demand Index (ST-COD) – Small-scale Sealed-tube Method.
- Korenga T, Zhou X, Okada K, Moriwake T, and Shinoda S (1993) Determination of chemical oxygen demand by a flow injection method using cerium(IV) sulfate as oxidizing agent. *Analytica Chimica Acta* 272: 237–244.

Lee K-H, Ishikawa T, McNiven SJ, *et al.* (1999) Evaluation of chemical oxygen demand (COD) based on coulometric determination of electrochemical oxygen demand (EOD) using a surface oxidized copper electrode. *Analytica Chimica Acta* 398: 161–171.

Messenger AL (1981) Comparison of sealed chamber and standard method COD tests. *Journal of the Water Pollution Control Federation* 53: 232–236.

Miller DG, Brayton SV, and Boyles WT (2001) Chemical oxygen demand analysis of wastewater using trivalent manganese oxidant with chloride removal by sodium bismuthate pretreatment. *Water Environment Research* 73: 63–71.

Westbroek P and Temmerman E (2001) In line measurement of chemical oxygen demand by means of multipulse amperometry at a rotating Pt ring–Pt/PbO<sub>2</sub> disc electrode. *Analytica Chimica Acta* 437: 95–105.

## Oil Pollution

**E Smith**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Oils are very complex mixtures of hydrocarbons in which the boiling points of components can vary from a few to several hundred degrees. Crude oils vary in their physical and chemical composition depending on their geochemical derivation, but all crude oils consist of a complex mixture of compounds comprised mainly of hydrocarbons. The hydrocarbon components of crude oil consist of straight and branched chain alkanes, cycloalkanes, and aromatics, and the relative content of these groups of compounds varies from oil to oil. Compounds containing oxygen, nitrogen, and sulfur, and various metals (Ni, V, Fe, Zn, Cu, U) are also present in crude oil.

The pollution of water and other matrices caused by the accidental leakage or chronic release of crude oil and refined products into the environment occurs each year. As the use and transportation of crude oil occurs on a large scale throughout the world, contamination of the environment with oil is an issue of concern. The amount of total hydrocarbons entering the oceans from all sources has been estimated at 2.35 million tons per year. Oil spills and hydrocarbons from land-based sources are usually limited to the coastal zone, but can be found even in the open oceans.

The characterization of a spilled oil in a contaminated environmental sample can be important for the assessment of environmental damage, and also in the selection of appropriate response and cleanup measures. The identification of an oil spill source is also extremely important for settling any dispute relating to liability and compensation. Petroleum hydrocarbon analysis may also be required to

determine the gradient of concentration around a point source, e.g., an oil platform, or provide baseline/benchmark and distribution concentrations.

Oil or petroleum products spilled on water undergo a series of biotic and abiotic processes that in combination are termed weathering, and cause changes in the physical and chemical properties of the oil. Weathering processes occur at very different rates but begin immediately after oil is released into the environment. These processes include evaporation, dissolution, dispersion, photochemical oxidation, water–oil emulsification, microbial degradation, and adsorption onto suspended particles. The changes in chemical composition of spilled oil can affect the toxicity of the oil and its biological impact over time, and also further complicate the identification of residual oil in the impacted environment. These factors make it difficult to select the most appropriate analytical methods for evaluating environmental samples.

A wide variety of instrumental and noninstrumental techniques are currently used in the analysis of oil hydrocarbons. These include gas chromatography (GC), gas chromatography coupled with mass spectrometry (GC–MS), infrared (IR) spectrometry, ultraviolet fluorescence spectroscopy (UVF) and gravimetry. Accurate and reliable analytical measurements are extremely important in order to unambiguously characterize spilled oil, to understand its fate and behavior, and to predict its long-term impact in environment. Despite advances in recent years in analytical technology, and in our understanding of the environmental fate of spilled oil, the complexity of oil means that there is no one method that can ‘do it all’ for the whole spectrum of oils and petroleum products that may be polluting an ecosystem. Each analytical technique measures only a subset of the complex mixture spilled or subsequently formed.

This article gives a brief survey of several analytical techniques currently used for the analysis



Lee K-H, Ishikawa T, McNiven SJ, *et al.* (1999) Evaluation of chemical oxygen demand (COD) based on coulometric determination of electrochemical oxygen demand (EOD) using a surface oxidized copper electrode. *Analytica Chimica Acta* 398: 161–171.

Messenger AL (1981) Comparison of sealed chamber and standard method COD tests. *Journal of the Water Pollution Control Federation* 53: 232–236.

Miller DG, Brayton SV, and Boyles WT (2001) Chemical oxygen demand analysis of wastewater using trivalent manganese oxidant with chloride removal by sodium bismuthate pretreatment. *Water Environment Research* 73: 63–71.

Westbroek P and Temmerman E (2001) In line measurement of chemical oxygen demand by means of multipulse amperometry at a rotating Pt ring–Pt/PbO<sub>2</sub> disc electrode. *Analytica Chimica Acta* 437: 95–105.

## Oil Pollution

**E Smith**, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Oils are very complex mixtures of hydrocarbons in which the boiling points of components can vary from a few to several hundred degrees. Crude oils vary in their physical and chemical composition depending on their geochemical derivation, but all crude oils consist of a complex mixture of compounds comprised mainly of hydrocarbons. The hydrocarbon components of crude oil consist of straight and branched chain alkanes, cycloalkanes, and aromatics, and the relative content of these groups of compounds varies from oil to oil. Compounds containing oxygen, nitrogen, and sulfur, and various metals (Ni, V, Fe, Zn, Cu, U) are also present in crude oil.

The pollution of water and other matrices caused by the accidental leakage or chronic release of crude oil and refined products into the environment occurs each year. As the use and transportation of crude oil occurs on a large scale throughout the world, contamination of the environment with oil is an issue of concern. The amount of total hydrocarbons entering the oceans from all sources has been estimated at 2.35 million tons per year. Oil spills and hydrocarbons from land-based sources are usually limited to the coastal zone, but can be found even in the open oceans.

The characterization of a spilled oil in a contaminated environmental sample can be important for the assessment of environmental damage, and also in the selection of appropriate response and cleanup measures. The identification of an oil spill source is also extremely important for settling any dispute relating to liability and compensation. Petroleum hydrocarbon analysis may also be required to

determine the gradient of concentration around a point source, e.g., an oil platform, or provide baseline/benchmark and distribution concentrations.

Oil or petroleum products spilled on water undergo a series of biotic and abiotic processes that in combination are termed weathering, and cause changes in the physical and chemical properties of the oil. Weathering processes occur at very different rates but begin immediately after oil is released into the environment. These processes include evaporation, dissolution, dispersion, photochemical oxidation, water–oil emulsification, microbial degradation, and adsorption onto suspended particles. The changes in chemical composition of spilled oil can affect the toxicity of the oil and its biological impact over time, and also further complicate the identification of residual oil in the impacted environment. These factors make it difficult to select the most appropriate analytical methods for evaluating environmental samples.

A wide variety of instrumental and noninstrumental techniques are currently used in the analysis of oil hydrocarbons. These include gas chromatography (GC), gas chromatography coupled with mass spectrometry (GC–MS), infrared (IR) spectrometry, ultraviolet fluorescence spectroscopy (UVF) and gravimetry. Accurate and reliable analytical measurements are extremely important in order to unambiguously characterize spilled oil, to understand its fate and behavior, and to predict its long-term impact in environment. Despite advances in recent years in analytical technology, and in our understanding of the environmental fate of spilled oil, the complexity of oil means that there is no one method that can ‘do it all’ for the whole spectrum of oils and petroleum products that may be polluting an ecosystem. Each analytical technique measures only a subset of the complex mixture spilled or subsequently formed.

This article gives a brief survey of several analytical techniques currently used for the analysis

of crude oil and petroleum products, and oil spill related environmental samples. The analytical method chosen will depend on the information required and a tiered approach can be used starting with a nonspecific method to identify samples that require more specific, lengthy, and expensive analysis. Methods may quantify total petroleum hydrocarbons (TPH) or specific individual components.

TPH methods generate a single number that represents the combined concentration of all petroleum hydrocarbons in a sample that are measurable by a particular method. The use of TPH concentrations to establish cleanup levels for soil or water is a common approach implemented by regulatory agencies. TPH is sometimes referred to as mineral oil, hydrocarbons, oil, extractable hydrocarbon, and oil and grease depending on the method used. There are many analytical techniques available that measure TPH concentrations in the environment. No single method measures the entire range of petroleum-derived hydrocarbons and because techniques can vary in the way hydrocarbons are extracted, cleaned up, and detected, they each measure a slightly different subset of the petroleum-derived hydrocarbons present in a sample.

Analytical measurements can be direct, i.e., the water samples are measured directly by the instrumentation, for example, UVF and immunoassay techniques can measure oil contamination directly in water samples. However, most methods measure the contaminated sample indirectly and therefore require sample treatment. The methods employed in sample treatment are common to all the indirect measurements.

## Sampling and Sample Treatment

### Collection and Preservation

Water samples are collected and stored in glass containers to avoid interference from plastics. For example, shallow water samplers utilizing 2.7l Winchester bottles have been used for more than 15 years up to depths of 50 m, although larger or smaller volumes (1–10 l) may be collected when the expected concentrations are very high or low. Samples should be stored in the dark and at low temperatures, in a fridge or freezer to minimize any loss of volatile components, and prevent photodegradation and microbial biodegradation of the sample. Also, samples should not be divided for analysis as this may introduce errors due to the adsorption of hydrocarbons onto the container walls. Adsorption of components to container walls can be minimized by carrying out the analysis rapidly, and it

has been recommended that water samples are extracted within 7 days of sampling, and analyzed within 40 days. If analysis is not carried out within 24 h the sample should be acidified to pH 2, or the solvent to be used for extraction be added at a ratio of 1:40. It is possible that shipboard contamination may occur when sampling water from leaks, spills, and lubricants during normal operations; it is therefore advisable to analyze fuel oil and bilge oil as part of a sampling strategy.

### Extraction

The extraction process separates the analytes of interest from the matrix, in this case the petroleum hydrocarbons from the aqueous phase. This is necessary for all nondirect methods of analysis.

The most commonly used method is extraction into a solvent using a separating funnel. The water sample is placed in a separating funnel, solvent added (e.g.,  $2 \times 50$  ml pentane) and the mixture shaken vigorously. The first aliquot of solvent may be used to rinse the sampling container to take into account any hydrocarbons adhered to the container walls. After layer separation the solvent layer (extract) is dried with desiccant (anhydrous sodium sulfate). Multiple extractions will increase recovery and the extract may then be concentrated prior to analysis.

Methylene chloride has been the solvent of choice for many semivolatile analyses due to its high extraction efficiency, low cost, and specification by many regulatory methods. In the past chlorofluorocarbon solvents, such as trichlorofluoroethane, have been used for oil and grease analyses, because of their low human toxicity, and spectral qualities (i.e., no absorption at  $2930\text{ cm}^{-1}$ , the wavelength generally used for sample measurement). However, owing to the detrimental effects of chlorofluorocarbons on stratospheric ozone, their use is being phased out.

Other methods in common use include purge and trap and headspace analysis for volatile components, and continuous liquid–liquid extraction and solid-phase extraction (SPE) for semivolatile compounds. Volatile compounds in water may be extracted by purging with an inert gas and trapping the compounds on a sorbent. These may then be analyzed by heating the sorbent to release the volatile compounds into a gas chromatograph. In headspace analysis the water sample is placed in a closed vessel with a headspace, and heated to drive the volatiles into the gas phase. Samples containing heavy oils or high hydrocarbon concentrations may contaminate purge and trap instrumentation, whilst this is minimized in headspace analysis as only the volatile components enter the instrument. In continuous liquid–liquid

extraction, rather than shaking solvent with the water sample, the solvent is continuously heated and sprayed on top of the water. This method is more time consuming than using a separating funnel but useful when samples contain emulsion forming solids. SPE involves passing the water sample through a cartridge or disk containing adsorbent such as silica or alumina. After extraction the analytes are eluted with solvent.

### Concentration

Concentration of the sample extract increases the capability to detect the compounds of interest. It is commonplace that analytes are present at low concentrations even once extracted. Concentrating the sample extract lowers the detection limits, and this may be performed for volatiles by sorbent trapping or cryogenic trapping, and for semivolatiles by methods such as rotary evaporation, a Kuderna Danish concentrator, and/or nitrogen evaporation. Care must be taken not to reduce the samples to dryness when removing the extracting solvent; otherwise losses of the lower molecular weight components (e.g., naphthalene) may occur.

### Cleanup

The aim of the cleanup step is to remove interfering compounds prior to analysis. This step may also be used to isolate a particular fraction containing analytes of interest. Cleanup is usually based on passing the extract through a glass column containing a sorbent such as silica, and eluting the petroleum hydrocarbons with solvent whilst leaving interfering compounds on the column. This is an important element in gravimetric and IR-based analyses as they are very susceptible to nonpetroleum hydrocarbon interferences; for example, polar compounds such as animal and plant fats may be inappropriately identified as petroleum constituents. Prior to analysis by GC the sample can also be fractionated during cleanup into aliphatic and aromatic components, which can be analyzed separately.

## Measurement and Identification

### Gravimetry

Gravimetric methods may be used to quantify TPHs from water samples. The water is extracted with a solvent (e.g., carbon tetrachloride), the solvent extract is passed through a florisil column to remove biogenic interferences, and the eluate is evaporated, dried at 105°C, and weighed. The advantage of this method is that it is simple, fast, and inexpensive. The range of application is typically up to

1000 mg kg<sup>-1</sup> with detection limits of 5–10 mg l<sup>-1</sup> in water. However, this method is not suitable for low molecular weight compounds and provides no detailed information on the nature of the oil.

### Infrared Spectrometry

IR spectroscopy can be used to measure the total hydrocarbon content from a sample. The hydrocarbons are extracted from acidified water using a nonhydrocarbon solvent, e.g., carbon tetrachloride and the absorption measured. IR is normally used to measure the saturated hydrocarbon content of a solution and not the aromatic hydrocarbon content. For example, the absorbance of the extracted material is measured at 2930 cm<sup>-1</sup> corresponding to the C–H stretching frequency of CH<sub>2</sub> groups, although some methods use multiple frequencies including 2960 cm<sup>-1</sup> for C–H stretching in CH<sub>3</sub> groups and 3030 cm<sup>-1</sup> for C–H stretching in aromatic rings. Quantification is based on the response of solutions of oil but problems may be encountered with the preparation of standards and samples. The extracting solvent must not contribute to the response; hence, carbon tetrachloride is commonly used, and biogenic interferences are removed by passing the extract through a florisil or silica column. This method may be used for the analysis of natural and saline waters, sewage and effluents with a limit of detection of 1 mg l<sup>-1</sup> in water.

### Immunoassay

Immunoassay methods have been developed to correlate TPH with the response of antibodies to specific petroleum components. A number of different testing kits based on immunoassay technology are available for rapid determination of TPH. These kits are self-contained portable systems designed to conduct analysis in the field, including sample preparation and instrumentation to read the assay results. Water samples are added directly so there is no lengthy sample preparation. A known amount of labeled analyte is added after the sample, typically an enzyme with an affinity for the antibody, and the sample analyte competes with enzyme labeled analyte on the antibodies. A color development reagent reacts with the enzyme and the absorbance is measured, which is inversely proportional to the concentration. Test kits are available for TPH, BTEX (benzenes, toluenes, ethylbenzenes, and xylenes – low molecular weight volatile compounds), and polycyclic aromatic hydrocarbon (PAH) analysis, but as their precision and accuracy are lower than standard laboratory methods immunoassays are only useful as a screening technique. Typical detection

limits for TPH range from 10 to 500 mg kg<sup>-1</sup> in soil and 200–500 µg l<sup>-1</sup> in water.

### Ultraviolet Fluorescence

The determination of petroleum hydrocarbons in seawater by fluorescence is a well-established technique and has been used in routine analysis in many parts of the world. It can be used to monitor very low concentrations of oil in seawater owing to its greater sensitivity than absorption techniques and has also shown good results for waters in the vicinity of oil spills. The simplicity, speed, and low limit of detection of this technique, coupled with its low cost when compared to other techniques, make it suitable for large-scale monitoring. Several intercalibration exercises have shown good performance and reproducibility of the UVF method, and it has been adopted for many national and international monitoring programs. The main disadvantage of this technique is that due to the complex and varying composition of hydrocarbons in oil and seawater it is not possible to determine actual concentrations as it is not known what is being measured. However, its strength lies as a screening tool that may justify spending resources for further in-depth analysis.

As there is no absolute measure of fluorescence emission, the spectrophotometer is first calibrated with solutions of a reference oil standard. The hydrocarbon concentration is determined by measurement of the fluorescence emitted at 360 nm whilst the sample is excited at 310 nm. This is compared with the fluorescence of standard solutions of oil (e.g., Ecofisk crude oil) under the same conditions. For diesel oil standards the fixed wavelengths used are excitation at 270 nm and emission at 330 nm. The method relies on fluorescence of the more water-soluble aromatic components of the oil; whilst sensitive it is very dependent on oil composition and suitable calibrants. Synchronous excitation-emission spectra (230–500 nm) are run for blanks, standards, and samples and comparison of the spectra indicate whether or not the standard contains aromatic hydrocarbons in a similar boiling point range to the samples, i.e., whether an appropriate standard was used. It is often impractical to make up standards for complex petroleum hydrocarbon samples; consequently, UVF may be measured in equivalents of a specific compound such as chrysene, which has been recommended by the World Health Organization. This technique also enables oil to be measured directly in water thus avoiding the need to extract the water sample with organic solvent. All measurements are carried out within the linear range of the instrument to avoid the effects of quenching, which

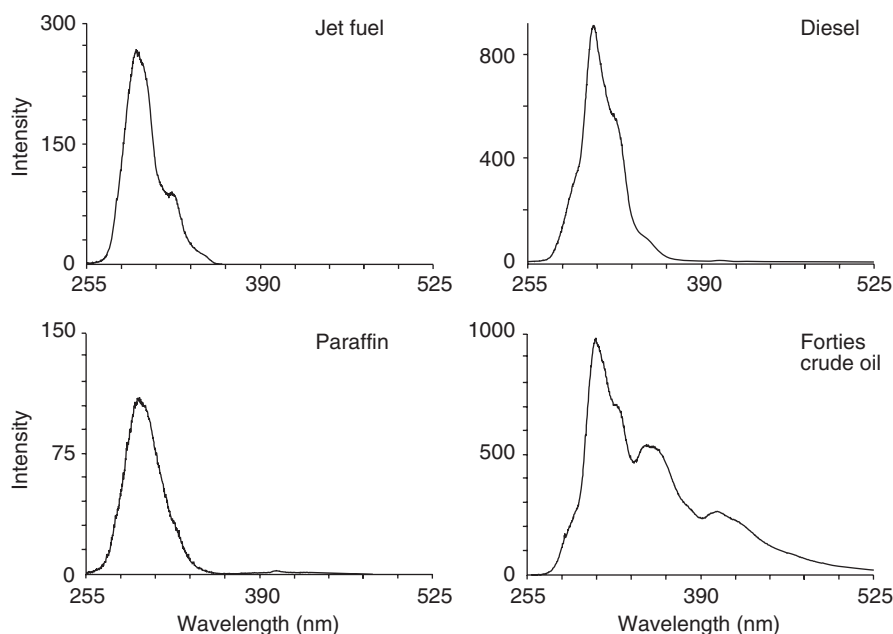
in practical terms is taken to be 0–10 mg l<sup>-1</sup> with the limit of detection at ~0.24 µg l<sup>-1</sup>. If the purpose of the analysis is to look at extent of oil pollution, the presence of combustion or naturally derived PAH would represent an interference. The use of synchronous excitation emission spectroscopy can sometimes highlight such interferences but more detail is gained by GC analysis. Representative synchronous excitation–emission spectra for jet fuel, diesel oil, paraffin, and Forties crude oil are shown in Figure 1.

### Gas Chromatography and Gas Chromatography–Mass Spectrometry

A major shortcoming of the nonspecific methods is that the information provided from using such methods does not afford details of the individual components present and thus petroleum source specific information. If oil pollution is found at a site, one of the first questions will be what is it, followed by where did it come from. GC, in addition to providing quantitative measurements, can provide details of the components present at a contaminated site and thus indicate where the pollution has come from.

Most GC-based methods are applied to sample previously separated into aliphatic and aromatic fractions using liquid chromatography or conventional column chromatography. The two fractions are then analysed on a gas chromatograph with a flame ionization detector (GC-FID). The TPHs are usually estimated by integrating the areas of the resolved and unresolved components of both fractions. For GC-based methods TPH is defined by anything extractable by a solvent or purge gas and detectable by GC-FID within a selected carbon range. Cryogenic cooling can also be employed to analyze the low molecular weight compounds. Detection limits for GC-FID TPH can be as low as 50 µg l<sup>-1</sup> whilst individual analytes may be detected at approximately one order of magnitude lower.

Oil and oil products can be readily identified from their GC traces during the early stages of an oil spill especially where the spilled oil is heavy and background hydrocarbon levels are low. A simple technique is to produce a chromatogram under standard conditions (column packing, flow rate, temperature), which can be compared to the trace of a sample of material suspected to have been discharged. If the type of contaminant is unknown a 'fingerprint' analysis can help in identification by comparison of chromatograms to reference materials. Certain fuels can be identified by characteristic reproducible chromatographic patterns. Hydrocarbon fuels give chromatograms with regularly spaced peaks (consecutive members of homologous series of compounds) whilst



**Figure 1** Synchronous excitation/emission fluorescence spectra. (Reproduced with permission from Kelly CA, Law RJ, and Emerson HS (2000) Methods for the analysis of hydrocarbons and polycyclic aromatic hydrocarbons (PAH) in marine samples. *Aquatic Environmental Protection Analytical Methods*, CEFAS Lowestoft 12, 18pp. © British Crown.)

lubricating oils have fewer resolved peaks. Correspondence of retention times and the overall envelope shape of the chromatogram may be sufficient to characterize the contamination. Total ion chromatograms for jet fuel, diesel fuel, paraffin, and Forties crude oil are shown in **Figure 2**. Complications can occur in interpretation if the pollution is not sampled immediately after discharge due to the loss of volatile components and degradation of the oil.

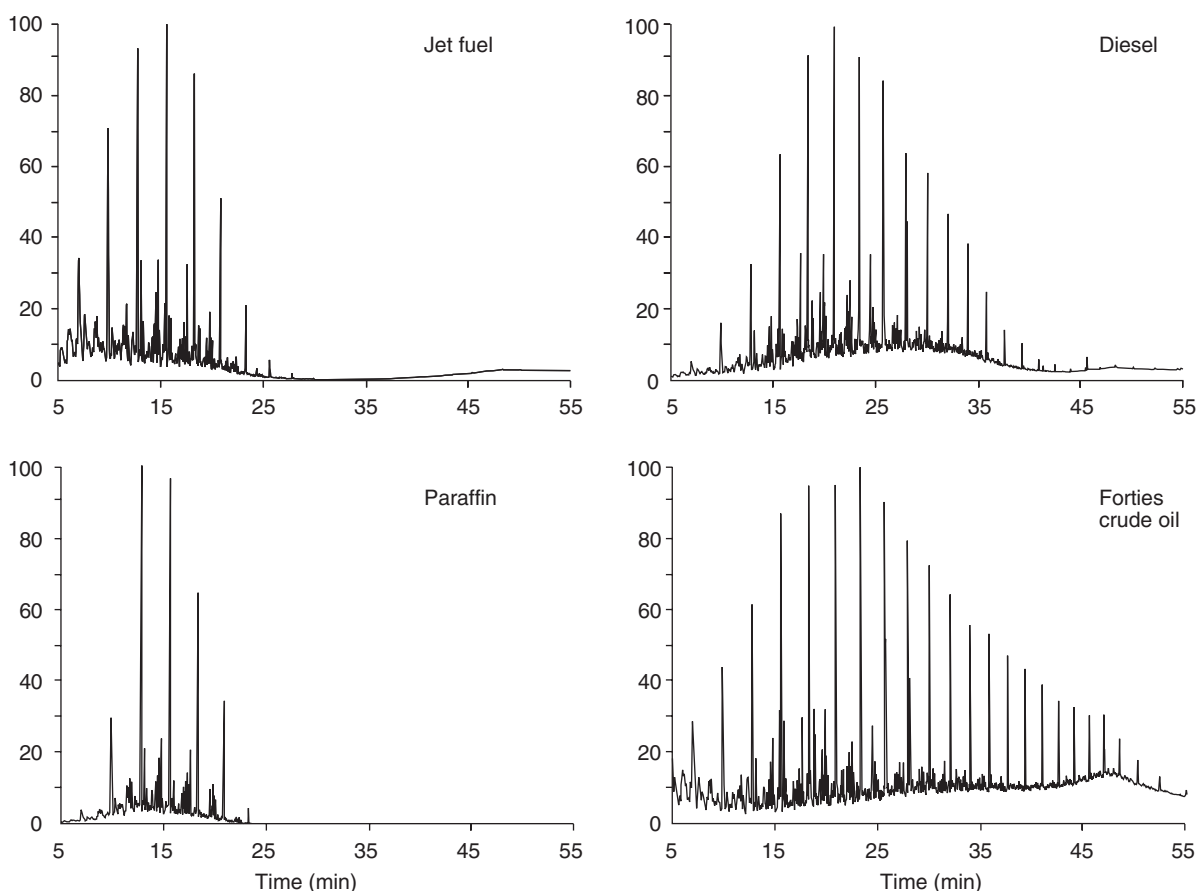
The combination of chemical separation by GC coupled with the spectral resolution afforded by MS permits the identification of specific compounds. The selection of appropriate analytes is dependent mainly on the type of oil spilled and the particular environmental compartment being assessed. The aliphatic fraction contains the majority of the compounds in petroleum, i.e., n-alkanes, branched alkanes, isoprenoids, cycloalkanes including steranes and triterpanes, as well as an unresolved complex mixture of saturated hydrocarbons. The aromatic fraction generally contains more toxic and more persistent compounds than the aliphatic fraction including PAHs. Specific compound determination is especially important for individual PAHs and biomarker compounds at the parts per billion concentration. To further enhance detection of target compounds the mass spectrometer may be run in selected ion monitoring mode, which increases the limit of detection by approximately one order of magnitude over that of full scan operation.

A variety of diagnostic ratios, especially for PAH and biomarker compounds, for interpretation of chemical data from oil spill have been proposed for oil source identification and monitoring of weathering and biodegradation processes. Weathering processes in the environment cause changes in the physical and chemical characteristics of the spilled oil. Significant losses occur due to evaporation of the low molecular weight alkanes but the  $nC_{17}$ /pristane and  $nC_{18}$ /phytane ratios remain virtually unaltered from the original oil and so can be used to determine the source of the oil. However, if the oil is heavily weathered or biodegraded these traditional diagnostic ratios become less reliable, but can be used to indicate microbial degradation at a site. The ratio of phenanthrenes to dibenzothiophenes has been used to discriminate and identify different refined oils, whilst priority pollutant PAH and their alkylated homologs may be used to differentiate between petrogenic and combustion (pyrogenic) sources. The contribution by biogenic hydrocarbons can also be determined.

## Standard Methods

There are many standard methods available for the analysis of oil pollution in water by regulatory agencies. The main techniques used have been summarized within this article, whilst further details





**Figure 2** Total ion chromatograms. (Reproduced with permission from Kelly CA, Law RJ, and Emerson HS (2000) Methods for the analysis of hydrocarbons and polycyclic aromatic hydrocarbons (PAH) in marine samples. *Aquatic Environmental Protection Analytical Methods*, CEFAS Lowestoft 12, 18pp. © British Crown.)

can be found in the references supplied in the Further Reading section.

**See also:** **Extraction:** Solvent Extraction Principles; Solid-Phase Extraction. **Fluorescence:** Environmental Applications. **Gas Chromatography:** Mass Spectrometry; Environmental Applications. **Gravimetry.** **Headspace Analysis:** Static; Purge and Trap. **Immunoassays:** Overview. **Infrared Spectroscopy:** Overview. **Liquid Chromatography:** Overview. **Sampling:** Theory.

## Further Reading

- American Society for Testing and Materials (2003) *Standard Test Method for Oil and Grease and Petroleum Hydrocarbons in Water*. D 3921-96. Philadelphia, PA.
- Butt JA, Duckworth DF, and Perry SG (1986) *Characterisation of Spilled Oil Samples: Purpose, Sampling, Analysis, and Interpretation*. Chichester: Wiley.
- National Academy of Sciences (1985) *Oil in the Sea: Inputs, Fates and Effects*. Washington DC.

Standing Committee of Analysts (1983) *The Determination of Hydrocarbon Oils in Water by Solvent Extraction, Infrared Absorption and Gravimetry*. London: HMSO.

Standing Committee of Analysts (1989) *Analysis of Hydrocarbons in Waters – A Review, and an Ultraviolet Fluorescence Spectrometric Method*. London: HMSO.

USEPA (1999) *Method 1664: Guidelines Establishing Test Procedures for the Analysis of Oil and Grease and Non-polar Material*. Washington DC.

Wang Z and Fingas M (1997) Developments in the analysis of petroleum hydrocarbons in oils, petroleum products and oil-spill-related environmental samples by gas chromatography. *Journal of Chromatography A* 774: 51–78.

Wang Z, Fingas M, and Page D (1999) Oil spill identification. *Journal of Chromatography A* 843: 369–411.

Weisman W (1998) *Analysis of Petroleum Hydrocarbons in Environmental Media*. Total Petroleum Hydrocarbon Criteria Working Group Series, vol. 1. Massachusetts: Amherst Scientific Publishers.



## Microbiological

**F M Schets, A M de Roda Husman, and  
A H Havelaar**, National Institute for Public Health and  
the Environment, Bilthoven, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The principles of the microbiological analysis of water have evolved over nearly a century and are important tools for the protection of public health all over the world. This article will focus on the principles and applications of detection and enumeration methods for health-related microorganisms, as they are used for the sanitary control of drinking water, and fresh and marine waters used for recreation and the growing of shellfish. It will not deal with methods for ecological research, which may be very specific, although the general outline described here may well be applicable. Likewise, methods for the examination of sludges and sediments will not be discussed. They will usually require specific sample pretreatment methods; the final stages of the analysis may again be very similar to those described here.

The microbiological analysis of water aims at a great variety of microorganisms, including viruses, bacteria, and protozoa. The target organisms may be pathogenic to man or they may be harmless model organisms. The different purposes of the sanitary microbiological analysis of water are described in the section on applications.

Microbiological analyses are extremely sensitive: just a single infectious particle may be detected in volumes of up to 1000 l of water. The weight of a bacterium is in the order of  $10^{-12}$  g; the weight of a virus may be as low as  $10^{-17}$  g. Hence, the limit of detection is between  $10^{-15}$  and  $10^{-20}$  g l<sup>-1</sup>. Many microbiological test systems are based on amplification of single particles to such concentrations or size that they can be detected by the naked eye. Consequently, the precision of these tests is governed by the behavior of individual particles in the object to be sampled. This requires statistical models for data analysis that differ from the normal distribution. Ideally, the Poisson distribution can describe microbial counts, but in many practical cases the distribution of microorganisms in the sample shows a greater degree of dispersion. Different proposals for adequate statistical models have been made, such as the 'extra-Poisson' model, the negative binomial distribution, or the lognormal distribution. This last model is often used because it can easily be adapted

to standard statistical methods by simple log-transformation of the data, and because interpretation of the final results is relatively straightforward. Using this approach, the standard principles for describing precision of analytical methods in terms of repeatability and reproducibility can also be used for microbiological analysis.

It must be realized that owing to the detection of single particles, there is a lower limit to precision that cannot be improved. In fact, if better repeatability is reported, this is suspect. The trueness of microbiological methods can only be expressed in relative terms, because there is no 'absolute' method for determining the number of microorganisms of a particular type in a sample. Nevertheless, the concept of accuracy can be used to describe the performance of methods relative to each other. Important aspects of accuracy are the recovery, the interference by competitive microorganisms, and the ability to differentiate the target organisms from other organisms.

### Sampling

#### Sampling Apparatus

Most microbiological samples are collected in borosilicate glass bottles that are sterilized before use by heat treatment (autoclaving for 15 min at 121°C or hot air for 1 h at 170–175°C). Polypropylene bottles that withstand repeated sterilization can also be used. Commercially available, presterilized sample containers are available and may also be suitable. Sample containers can be supplemented with preservative agents before or after sterilization, depending on the type of water to be sampled and the heat stability of the agents. For most practical purposes in health-related water microbiology, sample containers need not be pretreated to prevent adsorption of the target organisms to the wall of the container (this may be quite different in ecological research). However, free-living amoebas adsorb strongly to glass surfaces and pretreatment with silicon compounds is necessary.

Sample volumes greater than 1 l are usually collected by pumps. It is imperative that these pumps can easily be cleaned and disinfected. Pumping equipment can be disinfected by heat treatment or by using a sodium hypochlorite solution (75 mg l<sup>-1</sup> free available chlorine at a contact time of 5 min is usually sufficient), allowing repeated use in the field.

## Sampling Methods

Grab sampling is the most commonly employed method for sample volumes up to 10 l, the actual technique depending strongly on the water body to be sampled (e.g., sampling from taps for drinking water in mains, or bottles with attached weight for sampling surface water at a specific depth). Proper aseptic techniques should be used to prevent contamination of the sample. When sampling drinking water in open reservoirs, great care should also be taken to ensure that the sampling process does not contaminate the water under study. Larger sample volumes, which are usually necessary for the analysis of viruses and protozoa, may take several hours to collect and are therefore usually taken by continuous pumping. If standard bacteriological analyses are made in parallel, the use of the standard grab sampling technique is only appropriate if the temporal variations at the sampling station are known to be small. In all other cases, a continuous sampling method must also be used for these determinants, for example, by filling a 1–5 l bottle with a tube pump and subsampling in the laboratory after careful homogenization.

## Sampling Plan

Microbial counts may vary significantly in time and in place, and this aspect is often not correctly accounted for in sampling plans. It has been shown that effects of between-sample variation may influence the outcome of studies on the relationship between water quality and health and that significant correlations can only be obtained with samples from well-defined sampling points. In general, repeated samples of relatively small volumes should be collected and eventually pooled in the laboratory, rather than collecting a few large-volume samples and repeatedly subsampling in the laboratory. Expert statistical assistance should be sought when designing sampling plans.

## Sample Treatment and Storage

### Samples for Bacteriological Examination

The bacterial content of a water sample is highly unstable and cannot be preserved for long periods by simple methods. Ideally, the analysis should begin within a few hours after sampling and, during this period, the sample should be protected from light and temperature changes by placing it in an insulating container. If samples cannot be examined directly, they should be cooled rapidly by placing them on melting ice or in a high-capacity refrigerator. Ideally, the temperature of the sample should drop below 4°C within 4 h, but the sample should not freeze. Under these conditions, most determinants will not change

significantly in a period up to 24 h, whereas, for example, counts of spores of *Clostridium perfringens* may be stable for periods up to a week. Because local conditions may vary, it is recommended that studies of the effect of sample storage on bacteriological determinants be carried out at least on a national level and preferably by individual or central laboratories. In many cases, transport of samples on melting ice or examination within 24 h is only possible at high cost or is not possible at all. Many solutions to this problem are used, such as simple transport at ambient temperatures, addition of preservatives (e.g., boric acid), prefiltration, and shipment of filters on transport media. Any procedure will most probably result in changes in the test result and they should be assessed and reported. The acceptability of erratic test results should be weighed against the financial or logistic gains, and possible changes should be taken into account when interpreting the data.

Two types of water constituents may have a strong effect on the bacterial counts of water samples and should be neutralized during the sampling process. These are disinfectant residuals and heavy metals in concentrations exceeding  $10 \mu\text{g l}^{-1}$  (notably copper). Disinfectant residuals are conventionally neutralized by sodium thiosulfate to a final concentration of  $30 \text{ mg l}^{-1}$  (this is sufficient to neutralize chlorine residuals up to  $5 \text{ mg l}^{-1}$ ). Heavy metals are 'neutralized' by adding chelating agents such as ethylenediaminetetraacetic acid (EDTA) or nitrilotriacetic acid (NTA) to a final concentration of  $50 \text{ mg l}^{-1}$ . NTA is recently gaining preference because it is easily biodegradable. Neutralizing agents should be added to the sample bottle before filling, so that instantaneous action is guaranteed. In order to prevent recontamination, it is advisable to add a solution of these agents to the sample containers before sterilization. Sodium thiosulfate withstands both autoclaving and dry sterilization, whereas EDTA and NTA will only withstand autoclaving. It is also possible to combine sodium thiosulfate and EDTA or NTA in the wet sterilization process if the pH is above 9.

### Samples for Virological Analysis

Virological analysis usually involves processing of volumes between 10 and 1000 l. It is possible to collect sample volumes up to 10 l in sterile containers and to transport them to the laboratory. These samples should be treated as bacteriological samples (but no neutralization of heavy metals is usually necessary). Again, a procedure that does not permit rapid cooling and limited time of transport before analysis should be validated before use. For larger volumes, it is advisable to perform a first sample processing step

in the field. This usually involves adsorption of the viruses in the water to some type of filter, such as ultrafilters, negatively or positively charged membranes, or glass wool. Depending on the nature of the filter material and the water quality (pH, ionic strength, humic acids, etc.), it may be necessary to promote adsorption by decreasing the pH, using in-line injection of hydrochloric acid or addition of cations such as  $Mg^{2+}$  or  $Al^{3+}$ , or both. If the water has been disinfected, it may also be necessary to neutralize any residuals by injection of sodium thiosulfate. After adsorption, the viruses are usually eluted from the filter by a highly proteinaceous solution at a defined high pH (e.g., beef extract). This elution step is preferably performed in the field or it should be done in the laboratory as soon as possible. Primary eluates can be further treated directly or can be stored for long periods at  $-70^{\circ}C$  or below. Because many variables affect the recovery of the virus concentration procedure, it is advisable to evaluate the recovery of a particular method before use. This can be done by seeding simulated samples with stock virus suspensions or with waters with a high degree of natural pollution, such as sewage. Ideally, recovery studies should be carried out for every type of field sample, but this may add considerably to the cost of virological analysis. A suitable alternative is to evaluate the recovery of the concentration procedure for bacteriophages and to assume that constant performance for bacterial viruses also indicates constant performance for human viruses.

### Samples for Protozoological Analysis

Two types of protozoa are of interest in health-related water microbiology. Although not taxonomically correct, they are usually described by reference to characteristics of their life cycle as the 'free-living amoebas' and the 'parasitic protozoa'. Free-living amoebas are able to complete their entire life cycle in the inanimate environment. Whereas most species are harmless to man, some may be pathogenic. *Naegleria fowleri* is the causative agent of a lethal form of meningoencephalitis, while certain species (e.g., *Hartmanella* spp.) are of interest because they live in symbiosis with the pathogenic bacterium *Legionella pneumophila*. Parasitic protozoa can only complete their life cycle in a suitable animal host and they are excreted in the feces as highly resistant (oo)cysts. The two most important species for water hygiene are *Giardia lamblia* and *Cryptosporidium parvum*.

**Free-living amoebas** Because these organisms are actively multiplying in the environment, there is usually no need to examine very large volumes of sample: 1–3 l are commonly examined. The amoebas are filtered

from the sample by conventional membrane filtration (pore size 1–3  $\mu m$ ) with a low vacuum to prevent passage of the flexible amoebic cell through the pores.

**Parasitic protozoa** As for viruses, grab samples for parasitic protozoa can be transported to the laboratory, but when volumes up to 100–1000 l need to be examined the sample is filtered directly in the field. Protozoan (oo)cysts are retained in microporous (1  $\mu m$ ) filters on the basis of size; hence, it is not necessary to add acids or cations. Because of the high resistance against disinfectants, it is normally not necessary to neutralize any residuals, but this is however done when, e.g., swimming pools are sampled. In samples of this kind, the relatively high chlorine residual concentrations can interfere with staining of the (oo)cysts wall with fluorescent-labeled monoclonal antibodies and thus hamper microscopic detection. The sampling apparatus must be clean and free from protozoan (oo)cysts but need not be sterile. High chlorine concentrations to clean the apparatus should be avoided and the apparatus should be thoroughly rinsed before sampling.

Recovery of (oo)cysts from filters is normally done in the laboratory and filters are transported at 2–5 $^{\circ}C$  in the dark. A maximum period of 48 h between sampling and sample processing is generally considered acceptable. Retained particulates are eluted from the filter with a detergent solution and are further concentrated by centrifugation. *Giardia* cysts and *Cryptosporidium* oocysts are separated from other particulate debris in the concentrate by an immunomagnetic separation (IMS) technique. Magnetic latex beads coated with antibodies directed toward the (oo)cyst walls are gently mixed with the water concentrate allowing beads–(oo)cyst complexes to form. These complexes are recovered from the water concentrate by using a magnet, thus separating the (oo)cysts from other debris. Finally, the (oo)cysts are retrieved from the beads for microscopic detection. Other methods of sample purification, like flotation or flow cytometry, are also being used, but IMS is presently the most commonly applied method around the world. The recovery of this filtration–concentration method depends on many factors and requires experienced analysts. It is therefore important to validate all steps of the procedure before applying it to actual samples. This can be done by seeding water samples with known numbers of (oo)cysts from stock suspensions.

### Techniques

A great variety of detection methods for microorganisms in water is available. These are briefly described below.

## Bacteria

**Culture methods** Culture methods for bacteria can roughly be divided into two categories: liquid enrichment methods and plate count methods. In liquid enrichment methods, a test portion is inoculated into a growth medium that has been formulated to stimulate growth of the target organisms and to suppress growth of all other organisms ('background flora'). Choosing an appropriate incubation temperature and time enhances the selective nature of the growth medium. If the target organism is present in the test portion, this will result in a positive signal, irrespective of the original number. In their simplest form, liquid enrichment methods therefore give presence/absence type of information. In order to obtain semiquantitative information, it is usual to examine a series of different volumes (e.g., 100, 10, 1, and 0.1 ml) to produce an endpoint type of result. If a series of different volumes is examined in replicate, e.g., three- or fivefold, it is possible to use a statistical method known as the 'most probable number' (MPN) technique to estimate the original concentration of the target organism. The precision of this estimate is low (e.g., the 95% confidence interval of a fivefold MPN estimate is roughly between one-third and three times the analytical result). To improve precision it is sometimes recommended performing 96-well MPN estimates in microtiter trays, but with this technique only a limited volume of sample can be examined.

For more precise results the plate count method is more appropriate. The precision of this method is basically determined by the distribution of the target organisms in the sample (see above). In the plate count method, a test portion is inoculated onto the surface of a growth medium that has been solidified by addition of agar (spread-plate method). Each individual cell of the target organism will multiply into a colony that is visible to the naked eye. If several cells of the target organism are physically 'connected' (e.g., by adsorption to a particle of suspended matter), this will also result in one colony. The results of the plate count technique are therefore expressed as 'colony-forming particles' (cfp). Each cfp represents one or more cells of the target organism in the original sample. The pour-plate method, in which the test portion is mixed with the liquefied agar medium, poured into Petri dishes, and incubated after solidification, and the membrane filtration method, in which the test portion is filtered through a membrane filter (usually of 0.45  $\mu\text{m}$  pore size) that is placed on the growth medium, are variations.

It is generally not sufficient simply to inoculate a test portion in or on a selective growth medium to

obtain accurate results. Often, cells of target organisms will be damaged to some degree by physical or chemical stress and it may be necessary to revive them before placing them in the selective environment of the growth medium. This resuscitation procedure is an integral part of many culture methods and usually involves incubation in a less selective medium and/or at a less restrictive temperature. The selective growth medium is usually supplemented with some kind of specific detection system to differentiate growth of target organisms from that of background organisms. The detection system can be based on fermentation of specific sugars, enzymatic degradation of specific substrates, motility, reduction of hydrogen acceptors, etc., and will usually result in recognizable color changes, gas production, and so on. Routine methods should preferably be designed to produce results with an acceptable degree of selectivity at this point. For more critical examinations (e.g., detection of specific pathogens), it may be necessary to examine further the 'presumed' positive results by one or a series of confirmatory tests or even to proceed to identification at the genus, species, or type level. The objectives of the analysis determine the necessary level of confirmation and identification.

**Microscopic methods** Direct enumeration of bacteria in water by microscopic methods has only limited applications in health-related water microbiology because the detection limit is relatively high and because the microscopic image provides only a marginal clue to the identity of the bacteria. This latter fact can theoretically be overcome by specific immunofluorescence methods, but only few selective antibody preparations are readily available. Furthermore, microscopic methods do not commonly differentiate between living and dead cells, which makes interpretation of analytical results in terms of health risks impossible. Several methods have been developed to assess the viability of single cells by microscopic methods, such as exclusion of certain dyes (indicating integrity of the cell wall), reduction of tetrazolium salts (indicating active respiratory metabolism), and cell elongation in the presence of nalidixic acid (indicating active biosynthesis). These methods are time consuming and require the expertise of a research laboratory. It has been demonstrated that in a bacterial population under stress the detectability by culture methods is lost more readily than the viability using microscopic methods and claims have been made that bacteria in this 'viable, nonculturable stage' are infectious to man and experimental animals. However, these claims have not been supported to a great extent by well-published material and solid contradictory evidence has been



published. The infectivity of viable, nonculturable bacteria presently remains the subject of considerable debate.

**Molecular methods** Molecular methods have been explored in many research laboratories in recent years and their use has increased tremendously. Molecular methods include hybridization, restriction enzyme analysis, amplification, cloning, and sequencing. Hybridization and restriction, alone or in combination, are suitable techniques for identification of isolates that are obtained in pure culture. Fluorescence *in situ* hybridization involves direct detection of organisms after concentration on membrane filters with species-specific oligonucleotide probes labeled with fluorescent dyes, without the need for preculture. Amplification and cloning (i.e., amplification in a biological vector) produce nucleic acid fragments that can be compared to nucleotide sequences in databases (e.g., available on the Internet), which provide information on (sub)types, genetic relations, origin, etc., of the target organism. The most widely used amplification method is the polymerase chain reaction (PCR), in which single (c)DNA fragments are serially amplified by *in vitro* enzyme reactions. PCR can be used for direct detection of microorganisms in water samples, but can also be applied to identify isolates. Although PCR is a sensitive method, concentration of water samples is required. It enables detection of nonculturable organisms, but it cannot distinguish between live and dead organisms and may also detect dead organisms or past contamination because DNA is very stable in the environment. mRNA molecules, on the contrary, are very labile and have a rapid turnover. With reverse transcriptase-PCR (RT-PCR) mRNA can be amplified and the presence of this RNA indicates the presence of living or recently dead cells in a sample. Differentiation between live and dead cells can also be accomplished by preceding the PCR by a short enrichment step allowing culturable bacteria to multiply. PCR is basically a presence/absence technique but it can be used in an MPN-type assay to produce semiquantitative results. In recent years development of real-time quantitative PCR systems that are very promising for direct molecular detection and quantification of microorganisms in water samples have made progress towards quantification of PCR.

## Viruses

**Culture methods** Culture methods are most widely employed to detect viruses in concentrates from water samples. Bacterial viruses (bacteriophages) are

increasingly being used as an easily detectable surrogate for human viruses. The most frequently used (continuous) cell lines for supporting the multiplication of human viruses are derived from humans (tumor cells) or from primates (such as monkey kidney cells). The actual virus assay can be performed by a great variety of methods including the plaque assay, which is a quantitative method, or assays with cell cultures in liquid medium, which are presence/absence methods.

Detection of virus multiplication is usually done by observation of the host cell under the microscope for changes in structure or size (cytopathic effect). There are more than 100 types of human viruses that can be present in a water sample and for only a few types a cell-culture method is available that will result in readily observable cytopathic effects. Other viruses will be able to undergo only a limited degree of replication in cell culture and the products of this multiplication (proteins and/or nucleic acids) can be detected in the cell cultures by immunological or molecular methods. These techniques have expanded the range of detectable viruses and they have improved the sensitivity of the assay but, nevertheless, no cell-culture method is presently available for many important waterborne viruses.

**Immunological methods** Methods such as the enzyme-linked immunoassays and radio immunoassays have wide application in the clinical field but are not suitable for water analysis because of their high detection limit.

**Molecular methods** Direct detection of viral DNA or RNA is limited in use because of the high detection limit and amplification by the PCR assay is necessary to produce a detectable signal. Any virus with a known genome sequence may be detected by molecular methods since a set of probes or primers can be selected. Methods for cleaning up environmental concentrates are necessary for the removal of inhibitors of the amplification reaction. PCR has been successfully used to detect viruses at levels similar to or even lower than those detected by culture methods. Some important waterborne viruses cannot be cultured and their detection by molecular methods does not discriminate between infectious and noninfectious particles. In case viruses can be cultured the sensitivity of the assay may be improved by molecular detection of virus offspring. Quantification of virus DNA and RNA containing particles is rapidly improving with the development of real-time PCR assays but so far has only been available for part of the relevant waterborne pathogenic viruses.

## Protozoa

**Culture methods** Free-living amoebas are readily culturable on a lawn of living or killed bacteria that are used as a food source. The amoebas can be washed from filters used for concentration, after which drops of concentrate are placed on the assay medium to produce a semiquantitative result. It is also possible to place the filter directly (facedown) on the assay medium and to observe the margins of the filter for the emergence of amoebas, which will occur in due time. By using elevated incubation temperatures (37–44°C), it is possible to preliminarily differentiate between pathogenic and nonpathogenic *Naegleria* species, but complete identification and virulence testing require specialized laboratory facilities (e.g., inoculation in cell cultures or in mice).

*In vitro* cell culture assays have been developed for demonstration of *Cryptosporidium* oocyst infectivity. A variety of cell lines is in use (e.g., Caco-2, HCT-8, MBDK) and most of the life cycle of *Cryptosporidium* can be completed *in vitro*. The assays have been successfully applied to environmental samples, but detection of infectivity of environmental oocysts is hampered by lack of sensitivity of the assays in some situations. Cell culture assays can however satisfactorily be used in disinfection studies.

**Microscopic methods** The detection of (oo)cysts of the parasitic protozoa *Giardia* and *Cryptosporidium* is usually done by immunofluorescence microscopy. Various preparations of mono- or polyclonal antibodies are available but none of them is fully specific for the human pathogenic species. Confirmation of the microscopic image involves measuring the size and studying the shape of the fluorescent objects, and examination of internal structures using phase-contrast microscopy or interference-contrast microscopy. It is unusual for an (oo)cyst from an environmental sample to reveal all ideal attributes, so that a certain degree of subjective interpretation is presently unavoidable. Microscopic assays that differentiate between living and dead (oo)cysts are based on inclusion or exclusion, or active uptake of certain fluorescent dyes.

**Molecular methods** These methods and their applications are still under development, but much progress has been made in recent years. Many PCR-based methods for detection of *Cryptosporidium* and *Giardia* in water samples have been described. Detection of mRNA with RT-PCR allows discrimination between viable and nonviable organisms, whereas sequence analysis of RT-PCR fragments results in differentiation on species or genotype level

and enables confirmation of the presence of human pathogenic species. Genetic variation (in environmental samples) can be studied with the aid of a PCR based on sequences of the 18S ribosomal RNA gene of *Cryptosporidium*.

Direct quantitative detection of *Cryptosporidium* and *Giardia* has become within reach with the widespread application of real-time PCR and holds great promise for the near future, but is, however, in the early phases of development.

## Applications

### Model Organisms

The majority of microbiological analyses of water is not aimed at directly detecting pathogens but seek the presence/absence or concentration of model organisms. These can serve two functions:

1. 'Index organisms', to assess potentially dangerous contamination (such as *Escherichia coli* to indicate fecal pollution and consequently that a vast array of potential pathogens might be present).
2. 'Indicator organisms', to assess the efficiency of a treatment process (such as the heterotrophic plate count to indicate the efficiency of disinfection of swimming pool water).

Water quality standards are usually formulated on the basis of model organisms and tests for these organisms also play a major role in the surveillance of water quality and control of treatment processes. The major advantage of model organisms is that they can be detected and enumerated using simple, inexpensive methods that give results in a relatively short period of time. The major drawback is that there is not a single model organism that will give the right answer in all cases; hence, designing monitoring programs and interpretation of data always requires experience and knowledge about local situations. The most important model organisms are those occurring in large numbers in human or animal feces because this is the most important source of pathogens in water.

In practice, not all requirements for an ideal model organism can be fulfilled by one single species, but *E. coli* is the first choice in most cases. This bacterium occurs universally in feces of warm-blooded animals (including humans) and can be detected to a fairly high degree of selectivity by simple, rapid procedures. Lactose fermentation, indole production, growth at 44°C, and  $\beta$ -glucuronidase activity are the key properties for detection and confirmation of this species. In practice, many routine methods aim



only for lactose fermentation at 44°C (the thermotolerant coliform group) or at 37°C (the total coliform group) in culture media containing a selective surfactant. Thermotolerant coliforms are an acceptable alternative index of fecal pollution in most cases, albeit less specific than *E. coli*. The total coliform group comprises members that can occur in soil or on plant material or may even multiply in water. Their use is therefore restricted, e.g., as a sensitive means of detecting failures in system integrity in drinking-water distribution networks.

Because Enteroviruses and (oo)cysts of parasitic protozoa are known to be more resistant than *E. coli* and other coliforms, the absence of these organisms in surface or treated water does not necessarily mean absence of pathogens. Depending on the purpose of the examination and the type of water under study, additional model organisms need to be assessed. These include the fecal streptococci (which have a generally higher resistance than coliforms), the bacteriophages (in particular the F-specific RNA bacteriophages are good models of human enteric viruses), and *C. perfringens* (whose spores withstand disinfection much better than bacterial vegetative cells).

### Pathogens

The direct detection of pathogenic microorganisms, in particular viruses and protozoa, is usually more expensive and time consuming than detection of model organisms and also requires more sophisticated laboratory facilities. For this reason, application of these methods has been limited to research laboratories and to the monitoring programs of larger water quality management organizations in the industrialized world. However, the price and required expertise for analyzing viruses and protozoa is similar to that of analyzing chemical contaminants such as pesticides and chlorination by-products. In view of the known and acute health risks of pathogenic microorganisms in water, there does not seem to be a risk-based justification for not carrying out such analyses.

A wide range of pathogenic bacteria can be detected in water samples, usually by culture methods. Most commonly, quantitative tests are carried out that require several steps: preenrichment (= resuscitation), selective enrichment, isolation by plate culture, identification by biochemical and/or serological tests, and eventually epidemiological typing.

Pathogenic bacteria are increasingly being referred to in international water quality standards, whereas other pathogens that are not of fecal origin (e.g., *Legionella*, *Pseudomonas aeruginosa*) can actively multiply in the water environment and because no

adequate model organisms exist direct detection is required. In Europe, new drinking water legislation requires assessing the risk of infection for consumers of drinking water, based on the number of pathogenic microorganisms in the source water and the elimination capacity of the drinking water treatment processes. Therefore, direct detection of pathogens is increasingly being included in monitoring programs. It may also be of great value in the following situations:

- testing the efficiency of treatment methods and identifying the most appropriate model organism for routine surveillance;
- assessment of survival and transport phenomena in the environment;
- prospective or retrospective epidemiological studies; and
- after correction of calamities.

For these different purposes a great variety of methods is available, but only few have been introduced in laboratories and the use is widespread, whereas the majority is restrictively used by specialized laboratories.

See also: **Immunoassays, Techniques:** Enzyme Immunoassays. **Microscopy Applications:** Environmental. **Polymerase Chain Reaction. Sampling:** Theory.

### Further Reading

- American Public Health Association (1998) *Standard Methods for the Examination of Water and Wastewater*, 20th edn. Washington DC: American Public Health Association.
- Dahling DR (1991) Detection and enumeration of enteric viruses in cell culture. *Critical Reviews in Environmental Control* 21: 237–263.
- Department of the Environment: Standing Committee of Analysts (1990) *Isolation and Identification of Giardia cysts, Cryptosporidium oocysts and Free-Living Pathogenic Amoebae in Water*. London: HMSO.
- Grabow WOK, Botma KL, de Villiers JC, Clay CG, and Erasmus B (1999) Assessment of cell culture and polymerase chain reaction procedures for the detection of polioviruses in wastewater. *Bulletin of the World Health Organization* 77(12): 973–980.
- Havelaar AH, Heisterkamp SH, Hoekstra JA, and Mooijman KA (1993) Performance characteristics of methods for the bacteriological examination of water. *Water Science and Technology* 27: 1–13.
- Hunter PR (1997) *Waterborne Disease – Epidemiology and Ecology*. Storrington: Wiley.
- Hurst CJ, Benton WH, and Stetler RE (1989) Detecting viruses in water. *Journal of the American Water Works Association* 89(9): 71–80.

- Monis PT, Andrews RH, and Saint CP (2002) Molecular biology techniques in parasite ecology. *International Journal for Parasitology* 32: 551–562.
- Organization for Economic Co-operation and Development (2003) *Assessing Microbial Safety of Drinking Water – Improving Approaches and Methods*. Geneva: World Health Organization.
- Quintero-Betancourt W, Peele ER, and Rose JB (2002) *Cryptosporidium parvum* and *Cyclospora cayetanensis*: a review of laboratory methods for detection of these waterborne parasites. *Journal of Microbiological Methods* 49: 209–224.
- United States Environmental Protection Agency (2001) *Method 1623: Cryptosporidium and Giardia in water by filtration/IMS/FA (EPA-821-R-025)*. Office of Water 4603, Washington DC: USEPA.
- World Health Organization (1993–2003) *Guidelines for Drinking Water Quality (plus Addenda)*. Geneva: World Health Organization.

## Algal and Microbial Toxins

**C Robillot and L E Llewellyn**, Australian Institute of Marine Science, Townsville, QLD, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Water enters our daily lives through drinking, irrigation, recreation, food manufacture, in medicines, and medicine manufacture. Water itself cannot be manufactured and the world relies on natural supply. When conditions suit, freshwater populations of microbes and microalgae can explode and threaten public and environmental health if toxic species are present in these blooms. Consequently, the World Health Organization, drinking water authorities, and many environmental agencies have incorporated water quality guidelines for nuisance freshwater algae and microbes and the toxins they produce. Analytical methods for these toxins that support these quality guidelines can be divided into those that identify the toxin through their chemical properties such as molecular weight or chemical structure, or by their biological properties such as toxicity.

### Structure and Properties of Water-Borne Toxins

#### Microcystins

Microcystins (Figure 1A) are cyclic heptapeptides of low molecular weight quite frequently implicated in intoxication events and synthesized by the cyanobacteria genera such as *Microcystis*, *Anabaena*, *Planktothrix*, and *Nostoc*. They contain two unusual amino-acids: N-methyl-dehydroalanine (Mdha) and 3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-4,6-dienoic acid (ADDA). The most common sequence is D-Ala<sup>1</sup>-L-X<sup>2</sup>-D-Masp<sup>3</sup>-L-Z<sup>4</sup>-Adda<sup>5</sup>-D-Glu<sup>6</sup>-Mdha<sup>7</sup>, where Masp stands for methylaspartic acid and X and Z for two variable

amino acids giving its name to the molecule under the XZ form (such as microcystin-LR where X = leucine and Z = arginine). More than 70 microcystin variants are known to date, most discovered in the last decade using mass spectrometry, especially electrospray ionization (ESI) coupled to tandem mass analyzers or matrix-assisted laser desorption/ionization coupled to time-of-flight analyzers (MALDI-TOF). These techniques allow the determination of structural variations from very low concentrations. These variations can occur on each amino acid but some occur far more frequently in nature, such as the methylation–demethylation of Masp<sup>3</sup> and Mdha<sup>7</sup> acids, or the esterification of glutamic acid in position 6. About half of the known variants are produced by *Microcystis*.

Microcystins generally present two carboxylic acid functions and sometimes some additional basic functions on the variable amino acids (for example, the guanidine function on the arginine of microcystin-LR). Their ionization will therefore vary as a function of the pH. Microcystins are considered water soluble although few solubility data are actually available. Their polarity is variable with octanol–water partition coefficients (log K<sub>ow</sub>) values ranging from 2.2 to 4.4 as estimated by chromatographic analysis, showing that they are rather hydrophobic.

Microcystins potently inhibit serine/threonine protein phosphatases that dephosphorylate other proteins, an action akin to an on/off switch. The novel hydrophobic amino acid ADDA is essential to its bioactivity. Its action also promotes the formation of tumors. The liver is particularly susceptible to these toxins because unlike many other cell types, these peptides readily penetrate liver cells and are specifically taken up through the bile acid transport pathway. These toxins damage the liver by affecting the maintenance by these phosphatases of the cytoskeleton, a network of protein filaments. Protein phosphatases degrade the colorless *p*-nitrophenyl phosphate into a yellow product and this has been

- Monis PT, Andrews RH, and Saint CP (2002) Molecular biology techniques in parasite ecology. *International Journal for Parasitology* 32: 551–562.
- Organization for Economic Co-operation and Development (2003) *Assessing Microbial Safety of Drinking Water – Improving Approaches and Methods*. Geneva: World Health Organization.
- Quintero-Betancourt W, Peele ER, and Rose JB (2002) *Cryptosporidium parvum* and *Cyclospora cayetanensis*: a review of laboratory methods for detection of these waterborne parasites. *Journal of Microbiological Methods* 49: 209–224.
- United States Environmental Protection Agency (2001) *Method 1623: Cryptosporidium and Giardia in water by filtration/IMS/FA (EPA-821-R-025)*. Office of Water 4603, Washington DC: USEPA.
- World Health Organization (1993–2003) *Guidelines for Drinking Water Quality (plus Addenda)*. Geneva: World Health Organization.

## Algal and Microbial Toxins

**C Robillot and L E Llewellyn**, Australian Institute of Marine Science, Townsville, QLD, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Water enters our daily lives through drinking, irrigation, recreation, food manufacture, in medicines, and medicine manufacture. Water itself cannot be manufactured and the world relies on natural supply. When conditions suit, freshwater populations of microbes and microalgae can explode and threaten public and environmental health if toxic species are present in these blooms. Consequently, the World Health Organization, drinking water authorities, and many environmental agencies have incorporated water quality guidelines for nuisance freshwater algae and microbes and the toxins they produce. Analytical methods for these toxins that support these quality guidelines can be divided into those that identify the toxin through their chemical properties such as molecular weight or chemical structure, or by their biological properties such as toxicity.

### Structure and Properties of Water-Borne Toxins

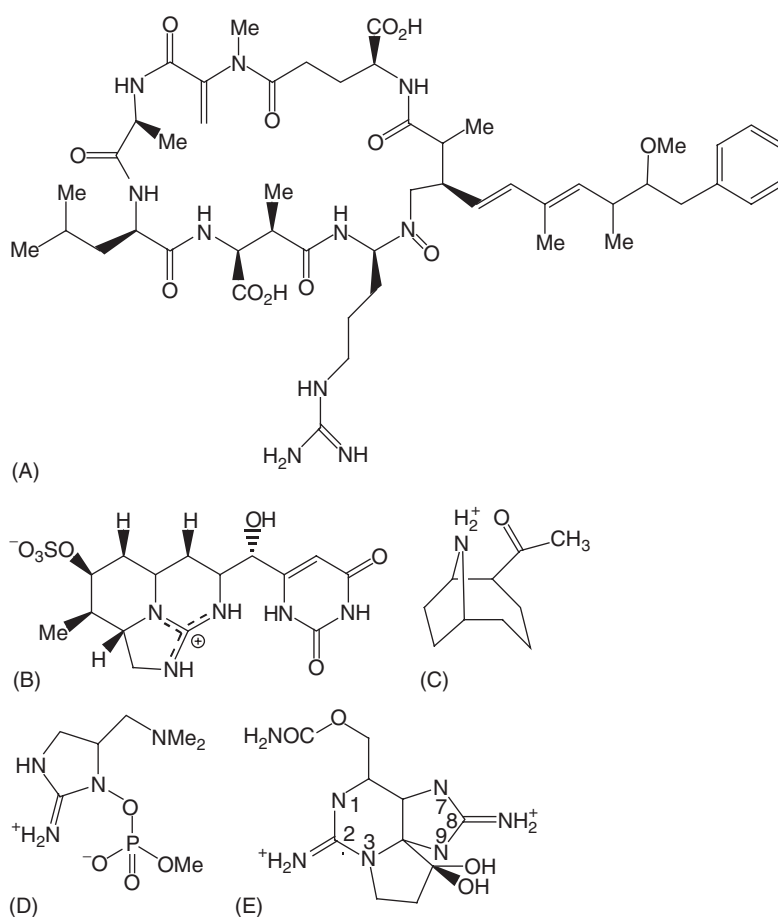
#### Microcystins

Microcystins (Figure 1A) are cyclic heptapeptides of low molecular weight quite frequently implicated in intoxication events and synthesized by the cyanobacteria genera such as *Microcystis*, *Anabaena*, *Planktothrix*, and *Nostoc*. They contain two unusual amino-acids: *N*-methyl-dehydroalanine (Mdha) and 3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-4,6-dienoic acid (ADDA). The most common sequence is D-Ala<sup>1</sup>-L-X<sup>2</sup>-D-Masp<sup>3</sup>-L-Z<sup>4</sup>-Adda<sup>5</sup>-D-Glu<sup>6</sup>-Mdha<sup>7</sup>, where Masp stands for methylaspartic acid and X and Z for two variable

amino acids giving its name to the molecule under the XZ form (such as microcystin-LR where X = leucine and Z = arginine). More than 70 microcystin variants are known to date, most discovered in the last decade using mass spectrometry, especially electrospray ionization (ESI) coupled to tandem mass analyzers or matrix-assisted laser desorption/ionization coupled to time-of-flight analyzers (MALDI-TOF). These techniques allow the determination of structural variations from very low concentrations. These variations can occur on each amino acid but some occur far more frequently in nature, such as the methylation–demethylation of Masp<sup>3</sup> and Mdha<sup>7</sup> acids, or the esterification of glutamic acid in position 6. About half of the known variants are produced by *Microcystis*.

Microcystins generally present two carboxylic acid functions and sometimes some additional basic functions on the variable amino acids (for example, the guanidine function on the arginine of microcystin-LR). Their ionization will therefore vary as a function of the pH. Microcystins are considered water soluble although few solubility data are actually available. Their polarity is variable with octanol–water partition coefficients (log *K*<sub>ow</sub>) values ranging from 2.2 to 4.4 as estimated by chromatographic analysis, showing that they are rather hydrophobic.

Microcystins potently inhibit serine/threonine protein phosphatases that dephosphorylate other proteins, an action akin to an on/off switch. The novel hydrophobic amino acid ADDA is essential to its bioactivity. Its action also promotes the formation of tumors. The liver is particularly susceptible to these toxins because unlike many other cell types, these peptides readily penetrate liver cells and are specifically taken up through the bile acid transport pathway. These toxins damage the liver by affecting the maintenance by these phosphatases of the cytoskeleton, a network of protein filaments. Protein phosphatases degrade the colorless *p*-nitrophenyl phosphate into a yellow product and this has been



**Figure 1** Structures of (A) microcystin-LR, (B) cylindrospermopsin, (C) anatoxin-a, (D) anatoxin-a(S), and (E) saxitoxin.

used to develop an enzyme inhibition assay for these toxins. Fluorescent substrates for protein phosphatases also exist now. Protein phosphatase isoforms sensitive to the microcystins are readily available from both commercial and natural sources. A variety of antibodies have been generated to microcystins that have been incorporated into enzyme-linked immunosorbent assays. This very large family of toxins and the subsequent variety in their chemical properties is problematic, however, in that antibodies targeting the peptidic portion of their structure are not likely to possess significant cross-reactivity to the multitude of microcystin variants. This has been circumvented in part by producing antibodies targeting the structurally invariant ADDA functionality common to all microcystins.

### Cylindrospermopsin

Cylindrospermopsin ( $C_{15}H_{21}N_5O_7S$ ; MW = 415.4), synthesized by the blue-green algae, *Cylindrospermopsis raciborskii*, *Umezakia natans*, and *Aphanizomenon ovalisporum*, presents a chemical

structure radically different from peptidic hepatotoxins (Figure 1B). It is an alkaloid consisting of a tricyclic guanidine moiety combined with hydroxymethyluracil. Zwitterionic, highly water soluble, and stable at acidic pH, does not degrade at 100°C for 15 min. An epimeric form (hydroxy group) can also be detected as well as a mildly toxic desoxycylindrospermopsin variant, as measured in mice.

Cylindrospermopsin's existence was first realized after 149 people were poisoned on Palm Island off the north-east Australian coastline when their fresh-water dam became contaminated by *C. raciborskii*. The molecular mechanism by which this toxin exerts its toxicity is yet to be elucidated. Liver toxicity, however, is the main toxic manifestation but lesions in the kidney, heart, and thymus occur when tested on mice and rats. Orally, cylindrospermopsin has a median lethal dose to mice of  $6 \text{ mg kg}^{-1}$  whereas intraperitoneal injection of mice yields an  $LD_{50}$  of  $0.2 \text{ mg kg}^{-1}$ . Other whole organism assays used to measure cylindrospermopsin toxicity include the brine shrimp *Artemia salina* ( $LD_{50} = 0.7 \mu\text{g ml}^{-1}$  after 72 h) and the crustacean *Thamnocephalus platyurus*,



used in the commercially available Thamnotox kit. Plants too have been used to test for cylindrospermopsin as it inhibits the growth and metabolism of the mustard plant *Sinapis alba*.

Not surprisingly, the toxin's liver toxicity is reflected by the ability of micromolar concentrations of cylindrospermopsin to kill *in vitro* liver cells such as rat hepatocytes and the human hepatoblastoma cell line HEP-G2. Cylindrospermopsin is known to potentially inhibit cellular protein synthesis that can be measured *in vitro* using rabbit reticulocyte lysate.

## Anatoxins

Anatoxins are alkaloids produced exclusively by the cyanobacterial genera *Anabaena*, *Planktothrix*, and *Aphanizomenon*. Anatoxin-a (Figure 1C), the first cyanobacterial toxin to be chemically and functionally characterized in 1972, is a secondary amine whose ionization state varies with the pH ( $pK_a = 9.4$ ). This compound is highly polar and fully soluble in water. Its methylated derivative homoanatoxin-a presents very similar properties. At elevated pH (above 10), both toxins become unstable and degrade very quickly into nontoxic analogs.

Anatoxin-a is a potent nerve toxin that acts postsynaptically. Synapses are a physical gap between nerves, and nerve and muscles, which must be traversed by any nerve impulse. This is achieved by the neurochemical acetylcholine moving across the synaptic junction. On the receiving side of a synapse, one finds acetylcholine receptors in the cell membrane that open when they bind acetylcholine to allow the entry of many ions into the nerve or muscle. Increased intracellular cations depolarize the receiving cell triggering a cascade of events that continue the action potential. Anatoxin-a is a powerful agonist of acetylcholine receptors being many times better than acetylcholine itself in stimulating the receptor to open. Anatoxin-a therefore outcompetes and displaces acetylcholine overstimulating the receptor, impairing its function and incapacitating nerve and muscle.

Acetylcholine receptors can be prepared from a variety of excitable tissues from animals. An example is the electric organ of electric eels, a very dense concentration of excitable tissue and a source of many neuronal receptors. A number of other chemicals competitively bind with anatoxin-a for the acetylcholine receptor (e.g., nicotine, acetylcholine itself,  $\alpha$ -bungarotoxin) and have been radiolabeled and used in receptor binding assays for anatoxin-a.

To date, anatoxin-a(S) is the only known natural organophosphate inhibitor of the neuronal enzyme acetylcholinesterase. The freshwater environment

can also be contaminated by insecticides such as parathion and malathion, which are also organophosphates that act in the same manner as anatoxin-a(S). With a structure (Figure 1D) very different from anatoxin-a, anatoxin-a(S) is rarely encountered and very little is known about its properties owing to strong chemical instability.

As described above, acetylcholine carries the signal across the gap between nerves and nerve to muscle. If acetylcholine lingered in the synaptic gap it would continually stimulate the acetylcholine receptor and the enzyme acetylcholinesterase degrades the acetylcholine. Anatoxin-a(S) inhibits acetylcholinesterase preventing the removal of acetylcholine and nerve and muscle cells become hyperexcited. Mice injected intraperitoneally with anatoxin-a(S) excessively salivate and exhibit lacrimation, urinary incontinence, defecation, convulsion, fasciculation, and die by respiratory arrest. By this route, the toxin has an  $LD_{50}$  approaching  $50 \mu\text{g kg}^{-1}$ .

Anatoxin-a(S) can therefore be measured by its inhibition of acetylcholinesterase whose enzyme activity can be measured by several ways. One example is its degradation of the acetylcholine analog, acetylthiocholine, and subsequent measurement of the released thiocholine by the sulfur reacting chemical Ellman's reagent. Acetylcholinesterase has been cloned and its mutation can increase the enzyme's sensitivity for anatoxin-a(S). Combining different acetylcholinesterase mutants with divergent specificity for anatoxin-a(S) and the above-mentioned organophosphate insecticides has enabled better analyte discrimination. This multiple enzyme method has been implemented in a biosensor that carries several of the acetylcholinesterase mutants.

## Saxitoxins

Saxitoxin ( $\text{C}_{10}\text{H}_{17}\text{N}_7\text{O}_4$ ;  $\text{MW} = 299$ ) was first isolated from the marine environment, namely from the clam *Saxidomus giganteus* from whence it derives its name. Saxitoxin (Figure 1E) has more than 25 naturally occurring chemical variants, all of which are tricyclic molecules with the 1,2,3- and 7,8,9-guandinio groups of STX itself possessing  $pK_a$ s of 11.3 and 8.2, respectively. They are all substituted tetrahydropuric bases, water soluble, thermostable and stable in acidic conditions. However, they are unstable in alkaline conditions and sensitive to oxidative conditions. Most derivatives present a common backbone, saxitoxin or *N*-hydroxysaxitoxin, and only differ on the location and number of sulfate substitutions. However, a new class of toxins has recently been discovered, with a phenolic group in place of the typical sulfate substitution. In the freshwater

environment, they can be produced by the cyanobacteria *Aphanizomenon flos-aquae*, *Anabaena circinalis*, *Lyngbya wollei*, and *Cylindrospermopsis raciborskii*. As expected from their chemical structure, these toxins have very different polarities and their lack of chromophore complicates their analysis.

The longest used bioassay for the saxitoxin family is intraperitoneal injection of mice, with toxin quantitated by comparing the time to death from respiratory paralysis to a standard series of mouse units. Saxitoxin is highly toxic, being lethal to guinea pigs at only  $5 \mu\text{g kg}^{-1}$  when injected intramuscularly and at similar doses when injected intraperitoneally into mice. Saxitoxin blocks sodium ion entry into nerves and muscle by occlusion of the voltage gated sodium channel. This prevents the conduction of a cellular action potential, paralyzing any victim, sometimes fatally.

A variety of *in vitro* assays have been developed that minimize or avoid live animal experimentation. Several capitalize on the sodium channel's affinity for these toxins. Neuronal cell lines lyse in the presence of veratridine, a sodium channel activator, and ouabain, which prevents removal of the excessive sodium ions allowed in by veratridine. In the presence of both these drugs, a sodium channel blocker such as saxitoxin rescues the cells. Cellular viability can then be measured by adding tetrazolium salts that are metabolized by living cells to a colored product. Alternatively, isolated cellular membranes, typically from brain tissue, are used to bind radiolabeled saxitoxin. After incubating receptor and radioligand in the presence of a test sample, any radiolabeled saxitoxin bound to the cell membranes are deposited onto filters by vacuum pressure. Radioactivity from the labeled saxitoxin is then measured with a signal reduction indicating the presence of saxitoxin.

Saxitoxin is also bound by receptors unrelated to the sodium channel. The first is saxiphilin, a protein found in the circulatory fluid of many animals used in a receptor binding assay much like that described above for the sodium channel. An alternative to natural receptors is to generate antibodies to saxitoxin. Saxitoxin itself does not elicit an immune response from mammals and one does not want the animal used to generate the antibodies to die from the toxin. The toxin then is linked to a carrier protein to create an antigenic epitope and reduce its toxicity. An example carrier is the enzyme horseradish peroxidase, which can also be used to generate an assay signal from the enzyme converting a colorless substrate to a visible product. Saxitoxin will interfere with the binding between the saxitoxin-horseradish peroxidase conjugate and the antibody generated to the toxin-protein complex and reduce any assay signal. Such

assays, however, are servant to the fact that antibodies raised toward one member of the saxitoxin family may not detect other chemical relatives.

## Extraction, Concentration, and Purification of Water-Borne Toxins

Most cyanotoxins are secondary metabolites that stay inside the cells during the growth phase and are released in the extracellular medium following cell lysis. In the environment, the concentration of toxins in the water is generally very low and the determination of the toxic potency is often facilitated by the analysis of the cells themselves or by preconcentrating the dissolved toxins by solid-phase extraction (SPE).

### Intracellular Toxins

Because cells are usually rich in proteins, minimizing their co-extraction using acetic acid at 5% (v/v) can be valuable. However, the choice of the extraction solvent will first depend on the polarity and solubility of the target toxins. The absolute efficiency of the extraction process can only be calculated using certified matrices, which are only available for a few toxin types. In the absence of such reference materials, several extraction methods can be tested and compared to identify the most efficient technique.

Microcystins have been extracted from cyanobacterial cells using aqueous acetic acid, various hydroalcoholic mixtures including the very popular *n*-butanol-methanol-water (5/20/75; v/v), but also methanol, pure or acidified with trifluoroacetic acid. If the extraction solvent was to contain too much water, the most hydrophobic toxins would be poorly extracted as a result of their low solubility in such a solution, and vice versa. Supercritical fluid extraction has also been applied to microcystins using 90% CO<sub>2</sub>, 9% methanol, and 1% water as the supercritical fluid.

Cylindrospermopsin is readily extracted with methanol or aqueous acetic acid.

As a result of its high polarity, anatoxin-a is better extracted in aqueous solutions. Aqueous solutions with or without acid have been employed, as well as water-alcohol mixtures. Anatoxin-a is volatile and can be lost during evaporation phases. Minimizing the extraction volumes and shortening evaporation steps can therefore increase extraction yields.

Anatoxin-a(S) has been extracted into acidified ethanol or a mixture of ethanol-aqueous acetic acid  $0.05 \text{ mol l}^{-1}$ .

Saxitoxins present a similar polarity to anatoxin-a and are extracted with aqueous acetic acid or



hydrochloric acid (pH 3) mixed with some alcohol, using ~50 ml of extraction solvent per gram of cells.

### Toxins Dissolved in the Water

When dissolved in the aqueous medium, toxins can be isolated by liquid–liquid extraction or liquid–solid extraction also called SPE. Liquid–liquid extraction consumes relatively large amounts of organic solvent and is not suitable for the extraction of water-soluble toxins and will not be discussed here.

The mechanism of SPE is directly related to liquid chromatography: toxins have to be retained by the adsorbent and not eluted by the water in the sample. Extraction sorbents are therefore similar to those used in liquid chromatography, reversed-phase (*n*-alkyl silica, apolar copolymers, graphitized carbon) for aqueous samples, or ion exchangers for charged compounds. Extraction parameters can be directly extrapolated from the retention times of the analytes as measured by liquid chromatography with water as the mobile phase.

In order to be able to detect very low concentrations of toxins in the water, volumes of about a liter of water can be concentrated employing disposable SPE cartridges. *N*-Octadecyl silica (C18) is by far the most common sorbent for the extraction of neutral, mildly polar, or hydrophobic organic compounds, and is excellent for microcystins concentration. However, it does not allow for an efficient extraction of polar toxins such as anatoxin-a, saxitoxins, or even cylindrospermopsin.

In contrast to intracellular toxins, estimating the extraction yields in water is made easier by our ability to spike the samples with known concentrations of standard toxins. As far as microcystins are concerned, the most adopted extraction protocol involves the acidification of the samples with trifluoroacetic acid before percolating them through a disposable C18 SPE cartridge. Polar variants such as microcystins RR, YR, or LR are still hydrophobic enough for 500 ml of aqueous sample to be percolated through a 500 mg C18 SPE cartridge without decreasing the overall recovery, even at neutral pH. Adding ~5% of methanol to the samples also prevents the adsorption of more hydrophobic microcystins to the plastic walls and help maximizing their recovery. The elution of microcystins from the SPE cartridges can be carried out with a few milliliters of acidified methanol (1% of trifluoroacetic acid, for example). A washing step can also be introduced just before the elution using a few milliliters of 20% methanol to eliminate weakly retained interferants. Detection limits of 0.05 and 0.1  $\mu\text{g l}^{-1}$  can be

reached for individual microcystins in drinking water and surface water, respectively.

Extraction of anatoxin-a is made more difficult by its strong polarity and protonation at pH below 9.4. The retention properties and as a consequence the volume of sample percolated can be improved by working in alkaline conditions where the toxin becomes neutral. An alternative is to use XAD2 resin after adjusting the toxins to pH 10.5. The use of extraction cartridges filled with styrene divinylbenzene (SDB) copolymers could drastically improve the SPE of anatoxin-a. As a matter of fact, these materials present a very large activated surface area and it has been shown that they extract many polar pesticides, even in their ionized form, although those same molecules were not extractable using typical C18 SPE cartridges. It has been found that up to 500 ml of water spiked with anatoxin-a (without pH adjustment) could be percolated onto a 200 mg SDB cartridge without observing any recovery loss. Elution was performed with 5 ml of acidified methanol (0.01% HCl), before evaporating the sample under nitrogen flux to a 100  $\mu\text{l}$  residual volume (avoiding a complete drying of the sample helps maintain high recoveries). Using this protocol, limits of detection of 0.1 and 0.3  $\mu\text{g l}^{-1}$  could be achieved starting with a 500 ml sample of drinking water and surface water, respectively.

Extraction of saxitoxins or cylindrospermopsin from aqueous samples has not been much described. Regarding their strong polarity, copolymers described earlier for anatoxin-a or graphitized carbon should give good results. The latter material is commonly used to desalt saxitoxin samples, but has also been shown to allow the extraction of cylindrospermopsin from lake waters spiked at 1  $\mu\text{g l}^{-1}$ .

### Immunoextraction

The SPE methods described above are all based on nonspecific interactions. Immunoextraction, on the other hand, is based on the high affinity of antibodies for a class of antigens, allowing their highly selective extraction from even very complex matrices. The covalent linkage of antibodies to various solid supports has allowed the development of SPE cartridges for a number of compounds. Some of these cartridges are commercially available such as those used for mycotoxins purification: aflatoxins, ochratoxins, and zearalenone. Antimicrocystin-LR monoclonal antibodies have been used to extract microcystins from environmental matrices, providing simultaneous concentration and purification of water samples.

## Chromatographic Analysis of Toxins

Chromatographic methods are routinely applied to the detection and quantification of toxins. Gas-phase chromatography (GPC) is not well adapted to non-volatile microcystins and requires chemical derivatization steps for other toxins; however, its ability to be directly coupled to mass spectrometers makes it a powerful tool. Capillary electrophoresis coupled to mass spectrometry has also been applied to the analysis of microcystins and saxitoxins. Although it seems to present several advantages over other techniques, it remains at a developmental stage as far as toxins are concerned.

### Gas-Phase Chromatography

Microcystins are not volatile and have to be chemically fragmented to be analyzed by GPC, for example, by oxidizing the ADDA fragment common to all microcystins and forming the 2-methyl-3-methoxy-4-phenylbutyric acid. When coupled to a mass spectrometer, this technique allows the measure of nanogram amounts of total microcystin in less than 30 min.

Anatoxin-a has to be derivatized using chemicals such as *t*-butyldimethylsilyl or pentafluorobenzyl bromide and 0.1–0.5 ng amounts can then be detected.

### Liquid-Phase Chromatography

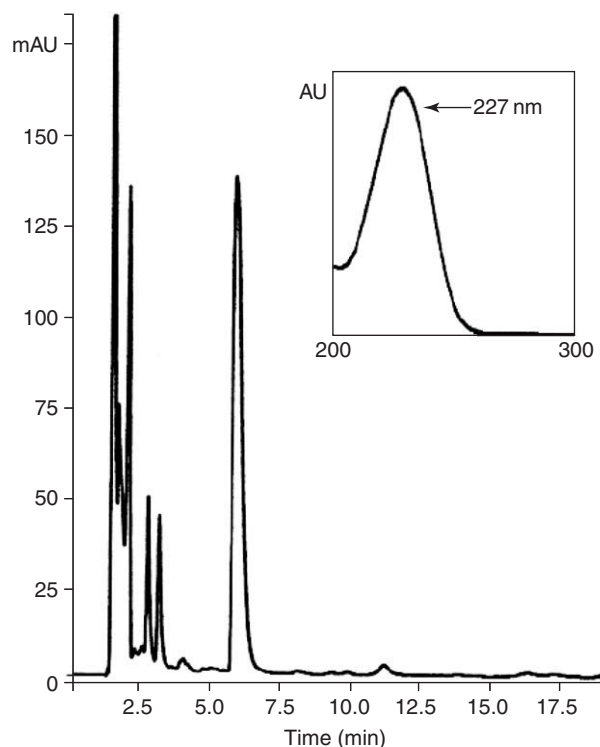
As discussed previously, microcystins are quite hydrophobic and most analyses are performed by reversed-phase liquid chromatography (LC), and most often using octadecylsilica as a stationary phase and a mixture of water and acetonitrile as mobile phase. The latter can be acidified (e.g., with 0.03% trifluoroacetic acid) and a 30–80% acetonitrile gradient allows the elution of most microcystins of various polarities. Separations are usually followed by ultraviolet (UV) detection, since all microcystins can absorb at 238 nm. Some additional maxima at 223 and 232 nm for tryptophan- and tyrosine-containing microcystins can be observed, justifying the use of a photodiode array UV detector. Additional coupling of LC to mass spectrometry has greatly enhanced the rate of microcystins discovery, especially since the ESI interface was introduced and shown to be ideally suited to microcystins analysis. MALDI-TOF mass spectrometry has been used to screen for and characterize microcystins from various environmental samples showing very promising results.

Cylindrospermopsin is a polar compound, retained sufficiently however on C18 phases to be analyzed by reversed-phase LC. As an example, its retention time is ~16 min using a Spherisorb ODS-2 column with a 10 min linear gradient from 0 to 5% methanol

followed by 10 min at 5% methanol. Common detection techniques are UV absorbance (262 nm) or electrospray mass spectrometry (ion at  $m/z = 416$  Da).

Anatoxin-a can be analyzed by reversed-phase LC using a C18 stationary phase and an isocratic mobile phase containing only small amounts of acetonitrile or methanol. Common detection techniques are UV (227 nm) or electrospray mass spectrometry (ion at  $m/z = 166$  Da) (Figure 2).

Saxitoxins do not carry any chromophore and are not detectable by measuring UV absorbance. The most reliable analytical method so far is based on an ion-pairing chromatographic separation followed by postcolumn chemical oxidation and fluorescence detection. The mobile phase is a mixture of water and acetonitrile containing sodium *n*-alkylsulfonate as an ion-pairing agent, and three different isocratic runs are usually performed to reliably separate and quantify saxitoxins from all three classes: saxitoxin, gonyautoxins, and C-toxins. Considering the large number of saxitoxins variants to be detected (>25), systematic coupling of LC to electrospray mass



**Figure 2** An example of the analysis of a water-borne toxin combining several of the procedures outlined in the text. Analysis of culture medium of the freshwater cyanobacterium *Anabaena* sp. 86 by LC–photodiode array detection following a concentration step on SDB SPE cartridges (inset: UV spectrum of anatoxin-a). Experimental conditions: C18 column; mobile phase: acetonitrile (2.5%)/potassium phosphate 5 mmol l<sup>-1</sup>, pH 3.5 (97.5%).

spectrometry is the most efficient technique. However, the mobile phases described above are not compatible with mass spectrometry interfaces and a new protocol has been developed involving the separation of saxitoxins using hydrophilic interaction chromatography, which can then be coupled to a mass spectrometer. The complete range of saxitoxin variants can be detected and quantified in a single chromatographic run.

## Bacterial Endotoxin

Primarily of interest to those requiring high-quality water for the manufacture of pharmaceuticals and biologicals, but emerging as an issue for natural water supplies is bacterial endotoxin, a lipopolysaccharide released when the cell walls of Gram-negative bacteria lyse. In humans, endotoxin release can lead to sepsis and septic shock. Fortunately, there exists a reliable and validated method based on the *Limulus* amoebocyte lysate assay. *Limulus polyphemus* is a horseshoe crab whose circulatory fluid rapidly coagulates when endotoxin is present. Coagulation

is caused by a protein in the amoebocytes in the horseshoe crab blood. Endotoxin quantitation is achieved by measuring various parameters of the clotting reaction such as simple measurement of turbidity or the more sophisticated use of chromogenic substrates that react to the responsible enzyme in the presence of endotoxin.

See also: **Extraction:** Solid-Phase Extraction. **Water Analysis:** Overview; Freshwater; Microbiological.

## Further Reading

- Anderson DM (1994) Red tides. *Scientific American* 271: 62–68.
- Carmichael WW (1994) The toxins of cyanobacteria. *Scientific American* 270: 78–86.
- Falconer IR (1999) An overview of problems caused by toxic blue-green algae (Cyanobacteria) in drinking and recreational water. *Environmental Toxicology* 14: 5–12.
- Hallegraeff GM, Anderson DM, and Cembella AD (2003) *Manual on Harmful Marine Microalgae*. Paris: UNESCO.

## Particle Characterization

**G G Leppard**, National Water Research Institute, Burlington, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

There is a well-established literature on particle analysis with respect to stable nonliving particles and living biological entities that can be considered as particles in the analysis of natural waters. The gross characteristics of aquatic particles larger than 1  $\mu\text{m}$  are relatively well known in terms of chemical nature and size distribution. A literature of similar quality and extent for unstable aquatic particles, especially those in the colloidal size range (1 nm to 1  $\mu\text{m}$ ), is currently being developed. The unstable particles of greatest interest in this endeavor are colloid systems (organic, inorganic, and mixed) that change their physical nature (size, shape, porosity) upon perturbation, especially after perturbations that disturb water molecules intimately associated with the structural integrity of a colloid system. Examples are macromolecular gels, growing flocs, and aggregating surface-active molecules. The focus

of this article will be on such unstable particulate materials.

The emerging literature on unstable aquatic particles stresses minimally perturbing analyses of minimally perturbed sample preparations. It treats native colloids and colloid systems as materials possessing their own intrinsic properties in the complex aquatic media from which they are sampled. Through the use of multimethod approaches, such minimally perturbed materials can be characterized according to their native state, as opposed to degraded (dehydrated, oxidized, coagulated, extracted, shrunken, or altered by microbes) versions of them. Concomitant with improvements in descriptive techniques have come refinements in the systematic means for detecting, minimizing, and assessing artifacts of colloid instability. Essential to producing the most detailed characterizations is analytical electron microscopy (AEM) used in concert with the particle analysis methodology of physics and analytical chemistry, and with novel advances in sample stabilization. With AEM, different kinds of microscopes and accessories are used in a correlative manner and transmission electron microscopy (TEM) is the principal instrument for structural analysis.

spectrometry is the most efficient technique. However, the mobile phases described above are not compatible with mass spectrometry interfaces and a new protocol has been developed involving the separation of saxitoxins using hydrophilic interaction chromatography, which can then be coupled to a mass spectrometer. The complete range of saxitoxin variants can be detected and quantified in a single chromatographic run.

## Bacterial Endotoxin

Primarily of interest to those requiring high-quality water for the manufacture of pharmaceuticals and biologicals, but emerging as an issue for natural water supplies is bacterial endotoxin, a lipopolysaccharide released when the cell walls of Gram-negative bacteria lyse. In humans, endotoxin release can lead to sepsis and septic shock. Fortunately, there exists a reliable and validated method based on the *Limulus* amoebocyte lysate assay. *Limulus polyphemus* is a horseshoe crab whose circulatory fluid rapidly coagulates when endotoxin is present. Coagulation

is caused by a protein in the amoebocytes in the horseshoe crab blood. Endotoxin quantitation is achieved by measuring various parameters of the clotting reaction such as simple measurement of turbidity or the more sophisticated use of chromogenic substrates that react to the responsible enzyme in the presence of endotoxin.

See also: **Extraction:** Solid-Phase Extraction. **Water Analysis:** Overview; Freshwater; Microbiological.

## Further Reading

- Anderson DM (1994) Red tides. *Scientific American* 271: 62–68.
- Carmichael WW (1994) The toxins of cyanobacteria. *Scientific American* 270: 78–86.
- Falconer IR (1999) An overview of problems caused by toxic blue-green algae (Cyanobacteria) in drinking and recreational water. *Environmental Toxicology* 14: 5–12.
- Hallegraeff GM, Anderson DM, and Cembella AD (2003) *Manual on Harmful Marine Microalgae*. Paris: UNESCO.

## Particle Characterization

**G G Leppard**, National Water Research Institute, Burlington, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

There is a well-established literature on particle analysis with respect to stable nonliving particles and living biological entities that can be considered as particles in the analysis of natural waters. The gross characteristics of aquatic particles larger than 1  $\mu\text{m}$  are relatively well known in terms of chemical nature and size distribution. A literature of similar quality and extent for unstable aquatic particles, especially those in the colloidal size range (1 nm to 1  $\mu\text{m}$ ), is currently being developed. The unstable particles of greatest interest in this endeavor are colloid systems (organic, inorganic, and mixed) that change their physical nature (size, shape, porosity) upon perturbation, especially after perturbations that disturb water molecules intimately associated with the structural integrity of a colloid system. Examples are macromolecular gels, growing flocs, and aggregating surface-active molecules. The focus

of this article will be on such unstable particulate materials.

The emerging literature on unstable aquatic particles stresses minimally perturbing analyses of minimally perturbed sample preparations. It treats native colloids and colloid systems as materials possessing their own intrinsic properties in the complex aquatic media from which they are sampled. Through the use of multimethod approaches, such minimally perturbed materials can be characterized according to their native state, as opposed to degraded (dehydrated, oxidized, coagulated, extracted, shrunken, or altered by microbes) versions of them. Concomitant with improvements in descriptive techniques have come refinements in the systematic means for detecting, minimizing, and assessing artifacts of colloid instability. Essential to producing the most detailed characterizations is analytical electron microscopy (AEM) used in concert with the particle analysis methodology of physics and analytical chemistry, and with novel advances in sample stabilization. With AEM, different kinds of microscopes and accessories are used in a correlative manner and transmission electron microscopy (TEM) is the principal instrument for structural analysis.



## Colloids

### General Aspects

In the context of aquatic environments, there is no clear definition of a colloid. For practical purposes, a colloid is considered to be any particle whose least dimension is within the colloidal size range, with the upper limit often being defined by a filtration operation. As a result of their very small size, colloids have properties intermediate between solutes and conventional particles, tending not to settle out of 'solution' unless they have become part of a settling aggregate much larger than 1  $\mu\text{m}$ . While the smallest colloids tend to aggregate strongly, individual colloids can persist as suspended units. When individual units associate into a colloid system, the resultant material takes on properties intrinsic to the material *per se*. Such a colloid system can aggregate or disaggregate in response to relatively minor perturbations of its aquatic milieu. Thus, an individual colloid can alternate from being a solute to being a component of a settling aggregate (a conventional particle). Because aquatic colloids are implicated strongly in the dispersion and fate of environmental contaminants, there is an urgent need to understand better their roles and behavior. As a result of their physical instability, however, they present challenges to attempts at realistic (native state) characterizations.

Colloids common to aquatic ecosystems are quite varied. They can be living (unicellular bacteria) or nonliving; the nonliving types can be organic (secreted fibrils and other refractory extracellular materials, fragmented cell parts, humic substances), inorganic (clay minerals, iron and manganese oxyhydroxides, silicates), or mixed (organic coatings on mineral colloids, mineral precipitates on organic fibrils or bacterial cell walls). Different combinations of the common colloids can coaggregate, sometimes under biological influence, to give rise to mixed materials such as microbe-rich flocs, mixed mineral sediments, and these two classes 'blended' together.

An example of a relatively simple colloid system (natural association of colloids), prepared for quantitative TEM analysis in a minimally perturbing way, is shown in **Figure 1**. It is rich in partially dispersed fibrils (elongated organic structures rich in acid polysaccharide) that tend to aggregate, so as to convert spontaneously into densely packed sheets, when their hydration status is perturbed by prolonged storage in water or by freeze-drying prior to receipt by an analyst. When the system's pore structure collapses as a result of sample mishandling, the fibrils and associated colloids tend to be incorporated into an (artifactual) dense gel, with concomitant reorientation of native associations (via uneven shrinkage) and the occlusion

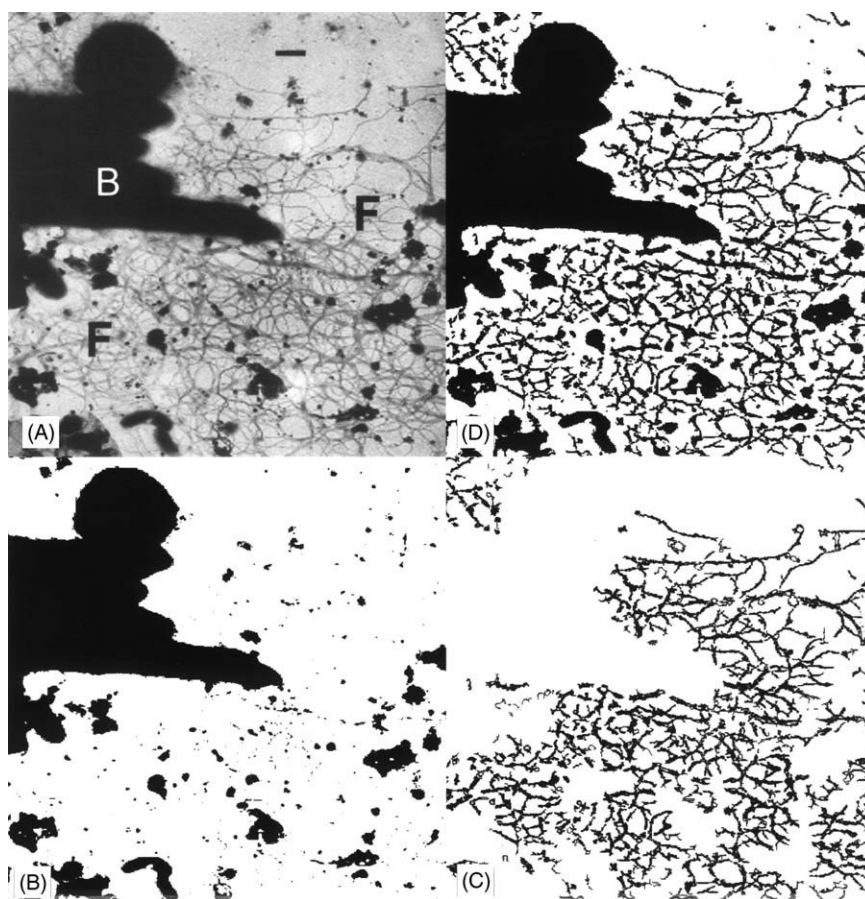
of some colloids deep within a near-impermeable organic matrix (impermeable to some extractants used by water quality analysts). **Figure 1** is chosen as an example of an interesting colloid system because of two features. Because it is a relatively large association of colloids, it reveals the flattening artifact caused by the capture of a heterogeneous, highly hydrated, aquatic particle onto a (transparent film mounted on a) TEM grid. Secondly, despite the flattening, this 'whole mount' preparation was amenable to a quantitative image analysis of the relative frequency of colloid morphotypes.

Artifactual aggregation of many kinds of colloids can occur at surfaces within apparatus used for water fractionation and fraction concentration. In addition to the artifacts resulting from mild mishandling, one must consider the artifacts produced by deliberate harsh treatments devised in the past to allow a high degree of purification of a given component of a natural water (such as clay minerals, biogenic minerals, and undenatured humic substances). Such harsh treatments as exposure to corrosives and exposures to multiple extractants destroy native associations and reduce complex materials to their most refractory components (e.g., 'cleaned' minerals, 'operationally defined' humic substances). Minimal sample perturbation in preparation for AEM is a necessity for characterizing native colloid associations.

### The Nature of the Data

To the greatest extent feasible, a characterization should be based on hydrated samples or samples stabilized in a hydrated configuration. While developing the data collection, one must take into account the dynamic properties of colloids that dominate their behavior in nature and that provoke those artifacts whose quality and extent should be known by the analyst. To the greatest extent feasible there must be a strong emphasis on analyses that are done on a 'per colloid' basis. The nature of the potentially available data for 'aggregated' colloids (and for particles in general) from research centered on TEM analyses is as follows: shapes; sizes and size distributions per aggregate type; porosity and density of packing, including patchiness; relative frequency of colloid types; three-dimensional arrangement of colloid types; fractal dimensions; and quantification of frequent natural associations.

Such data are derived by correlating morphological analyses (on a 'per colloid' basis), obtained by TEM, with information from the following techniques: elemental composition analysis using scanning transmission electron microscopy coupled to energy dispersive spectroscopy (STEM-EDS); TEM-based



**Figure 1** A transmission electron micrograph of a whole mount prepared, by ultracentrifugation, from a water sample rich in colloids, accompanied by images derived from computerized image analysis. (A) The original electron micrograph showing fibrils F, a microcolony of bacteria B, small mineral colloids associated with the meshwork of fibrils, and organic debris (electron-translucent structures at extreme lower left). (B) An image showing all colloids more electron-dense than the fibrils. (C) An image showing only fibrils. (D) A composite of the two images from 'B' and 'C'. The scale bar corresponds to 0.45  $\mu\text{m}$ . (Reproduced with permission from Leppard GG and Arsenault AL (2003) Quantification of individual native biocolloids in natural waters: Assessing indicators of aquatic events, by using transmission electron microscopy with a modified approach to image analysis. *Archiv fuer Hydrobiologie*, 156: 565–573, <http://www.schweizerbart.de>.)

selected area electron diffraction (SAED); electron energy loss spectroscopy (EELS) used with energy filtered TEM; atomic force microscopy (AFM); microchemistry (certain techniques modified by biomedical scientists for use with EM preparations); and highly selective probes modified to make them electron-opaque (e.g., by attaching distinctive gold colloids to them). Correlating standard TEM information with that of (synchrotron-based) scanning transmission X-ray microscopy (STXM), applied to the same particle, has begun; STXM provides information on the atomic environment of a given element and permits mapping of chemical species based on bonding structure. Elemental composition analyses by STEM–EDS can be done on both a colloid system basis (e.g., dot mapping) and on a per colloid basis (for individuals in the mid-size range and larger).

Crystallographic ‘fingerprinting’ of even small colloids can be achieved with SAED of crystalline colloids placed in the specimen plane of a TEM. To a limited extent, microchemistry can reveal the heterogeneous distribution of families of molecules, particularly organic macromolecules. The selective staining of such molecules with distinctive molecular markers has been developed into a fine art to differentiate subcomponents of heterogeneous particles. Information from EELS can complement that from EDS, especially for identifying light elements with a high spatial resolution; EELS can also provide qualitative information about the molecular environment of an element. AFM permits very high resolution of hydrated macromolecules.

Particle analysis techniques independent of EM-based techniques can be used to verify and extend



some of the EM data in multimethod correlative experiments. They can, of course, be employed profitably with no reference to EM in the analysis of aquatic colloids where limited kinds of information (e.g., size distributions of near spheres) are all that one requires. Among the most informative of the technologies are field-flow fractionation (FFF), fluorescence correlation spectroscopy (FCS), photon correlation spectroscopy, ultracentrifugation, and (when 'tuned' by trial-and-error experimentation to minimize colloid aggregation artifact) filtration/ultrafiltration.

The nature of the data required by aquatic scientists in the past tended to be that of particle size and size distributions, approximate particle volumes and shapes, and compositional analyses of a preliminary nature; the information needs of today require much more than this. Techniques unrelated to EM can obtain their data much faster than EM can produce it, concomitantly producing rapid assessments of statistical significance. This past slowness of data production by EM, however, has been overcome partially by specialized image analysis techniques and (for the largest colloids and smallest conventionally defined particles) automated data collection. Use of EM with selected rapid techniques in a correlative manner can accelerate EM data collection while providing necessary details on spatial and chemical heterogeneity unobtainable by rapid techniques. The reluctance on the part of some physical scientists to use EM as a primary tool is based on past unpleasant experiences with a complex, unfamiliar preparative technology whose simplest versions led to uncontrolled artifacts of dehydration. Given the great refinements in EM cryotechnology now at one's disposal and the adaptation of hydrophilic embedding resins for field use, such reluctance is counterproductive. Given a sound and systematic approach to minimal perturbation, the finest details supplied by EM analyses are more amenable to interpretation.

## Large Macromolecules

Macromolecules are defined here in terms of colloids rather than as large solutes. They can be considered generally as small colloids ( $\sim 1$ – $10$  nm), although it is accepted that some mineral colloids and extensively cross-linked biological debris could be considered as macromolecules whose size extends from the small colloid range into the range of conventionally defined particles. FFF in its flow modification can produce particle sizing data ( $1$  to  $>100\,000$  nm) over the entire colloidal size range and extending well into the 'true' particle size range. It can also yield relative molecular mass distributions for macromolecules.

The sedimentation modification of FFF is currently finding success in the separation and size characterization of aquatic colloids and macromolecules in the range of  $10$ – $1000$  nm. These and other rapid sizing techniques for macromolecules can be sufficient in themselves for many research purposes and take a leading role, with EM providing accessory information.

For small aggregates of macromolecules (organic fibrils, some iron polymers), the shape and least diameter of near-nanometer fibrous and acicular structures might best be measured by TEM (although AFM should also be considered). For samples in which one must ascertain native associations within a suspended mixed aggregate, especially when the aggregate is a physically unstable floc, TEM can become an analytical necessity. **Figure 1** illuminates the utility of TEM, and associated image analysis, for recording and quantifying an organic-rich association of colloids that has been captured and stabilized on a support film. The finest components of this porous aggregate (fibrils of  $5$  nm diameter) extend into the macromolecular size range; the fibrils themselves can be linear aggregates of macromolecules.

From a quantitative standpoint, the most important macromolecules suspended and/or dissolved in natural waters are humic acids, fulvic acids, polysaccharides, proteins, natural breakdown products of lignins (and lignocellulose), and cross-linked macromolecular mixtures (of various combinations of the above in differing proportions). Associated with, or bonded to, these mixtures are varying proportions of mineral particles that, in the macromolecular size range, may be either erosion products or the early results of mineral genesis. The humic and fulvic acids of interest to aquatic scientists are, increasingly, the undegraded native versions and not the 'simplified' molecules taken from a chemist's bottle for model studies. The polysaccharides have not yet received the attention that their activities merit, but it is clear that several families of them are (at least on an episodic basis) of quantitative significance. Depending on site and temporal considerations, surface waters can receive important contributions from various acid polysaccharide types (carboxylated or sulphated), cellulose, chitin, and readily decomposed food reserves released from damaged cells (starches, laminarans). Additionally, in localized portions of water bodies, one may find significant quantities of decomposing biota that are releasing other macromolecular substances; among these are organic phosphates (nucleic acids), lipids (lipoprotein complexes), and tannins (tannin-protein complexes).

## Sampling and Sample Preparation

To characterize a sample in a state as close to the native as is possible, one must manipulate the sample as little as is feasible; this usually requires developing a carefully considered sampling strategy that must be linked to all other sample handling steps whether they be in the field or in the laboratory (or even within the analytical instrument). A minimal perturbation approach consists of close adherence to the following general strategy for sampling and sample stabilization/preservation.

1. Take the sample from nature as gently as possible.
2. Avoid sample storage by analyzing and/or stabilizing the sample immediately in the field.
3. Avoid or at least minimize any concentration of sample components.
4. When fractionation is required, isolate the materials/substances as gently as is feasible while minimizing the number of processing steps.
5. When colloid instability artifacts are evident, use TEM to detect their nature and assess their extent, then use TEM as a monitor to aid method modifications designed to minimize the artifacts (tuning the method to suit the sample).
6. Use extraordinary means (when cost–benefit considerations permit) to take samples that are unusually sensitive to perturbation – means such as *in situ* cascade filtration and the stabilization of near-anoxic samples under their natural oxygen tension.

For characterizing readily perturbed colloid systems, certain preparative activities should be avoided completely (such as freeze-drying, sonication, uncontrolled degassing, uncontrolled/excessively high filtration flow rates, exposure to corrosive fluids, and, last but definitely not least, air drying). The method of sample stabilization must be rapid enough to minimize biological activities leading to undesired changes. Sometimes, sample concentration (by way of specialized apparatus) can be made unnecessary by a judicious use of information on the specific properties of the sample site. Some colloids are concentrated naturally by water bodies into layers (iron oxyhydroxides at the oxic-anoxic interface of a lake) or on an episodic basis into zones ('clouds' of organic fibrils being released by a collapsing bloom of algae). The cost of site monitoring to take advantage of such phenomena can be repaid manifold in terms of reducing the cost of adapting to apparatus-induced artifacts of concentration. Many protocols have been established for the nonmicroscopical rapid techniques of particle analysis. To the extent that a given protocol will introduce preparatory artifacts prior to

measurement, the rapid technique will add information on particle characteristics that deviate from those of native particles. The minimal perturbation approach to sample handling can be summarized succinctly in few words – 'do as little as possible'.

Recently, the literature has received many new methods for capturing and/or stabilizing aquatic colloidal materials to permit high-resolution AEM investigations of them and their natural aggregates. In addition to permitting the resolution of nanoscale ultrastructure, these methods permit element composition analysis (EDS and EELS), crystallographic 'fingerprint' analysis (SAED), and sometimes an analysis of the average 'atomic environment' of a given element (EELS, STXM). Attention is currently focused on methods that employ hydrophilic embedding resins, especially the melamine resin Nanoplast, whose usage has been modeled after existing methods in the biomedical literature. The basic Nanoplast method consists of placing hydrated aquatic materials in the water-rich water-miscible resin, which polymerizes so that the three-dimensional relationships of colloids in the material are fixed by resin polymer bridges prior to the loss of all bulk water in the resin-hardening stage of embedding. Thus, the (previously) physically unstable material cannot shrink or distort in response to the final stages of dehydration; it is presented in the specimen plane of a TEM or STEM as a suspended aggregate held in the plane by an electron-transparent resin matrix. Presentation in the specimen plane is accomplished through the use of either ultrathin sections or an ultrathin film. Nanoplast film techniques have evolved well and present certain advantages over sections; sample preparation is faster and the images (see **Figure 1**) are more amenable to quantitative image analysis in the case of films.

Visualization methods based on the ultramicrotomy of samples embedded in the standard hydrophobic resins permit optimal use of biomedical stain technology for selective staining (selective contrast enhancement) of the individual components within (ultrathin sections of) a colloid system. At a structural resolution near 3 nm, one can relate images from aquatic samples (whose contrast has been enhanced in a standard way) to images from the biological literature, to identify many colloid 'species' (fibrils, viruses, bacteria, refractory cell wall fragments, algal scales, eukaryotic cell debris, biogenic minerals). These standard images, despite the artifacts inherent in the preparatory techniques requisite for hydrophobic embedding, present most entities in the context of known artifacts and, usually, known artifacts whose extent has been either minimized or defined for purposes of interpretation; the highest quality portions of this literature are

invaluable as aids for generating characterizations of aquatic colloids. The images derived from the colloids and colloid systems embedded in hydrophobic resins can be correlated with related images derived from the use of a hydrophilic resin with portions of the same sample; the latter approach provides better structural resolution ( $\sim 1$  nm practical resolution) and better structural fidelity, but suffers from the lack of a properly developed stain technology. This latter defect, however, does not impede the analysis of colloidal minerals embedded in a hydrophilic resin; minerals provide sufficient native contrast for preliminary identification via a TEM-based literature that describes the morphology of various mineral colloids; morphological indicators are readily supplemented by spectra from STEM-EDS and 'fingerprinting' by SAED.

## Analysis

### Interfacing EM with Rapid Techniques

There are many techniques of particle analysis that can provide near-nanometer resolution of particle size. Some of these, however, can be inappropriate for a given sample type; examples are samples requiring an undesirable degree of preconcentration and samples with major components that are elongated to the extent that the concept of equivalent spherical diameter becomes meaningless. When used appropriately, high-resolution sizing techniques are increasingly amenable for routine use, especially those techniques that provide data rapidly. They can, however, be underutilized or misleadingly utilized if they are employed with a sample that lacks the integrity necessary for the analyst to do descriptive work at high resolution. Hence, the earlier emphasis on ascertaining the nature of colloid instability artifact with respect to a given type of surface water sample, and the strong argument put forward for a minimal perturbation approach to sample handling, including sampling strategy. The consequent trial-and-error testing requires the analyst to detect, assess, and minimize artifacts of sample manipulation.

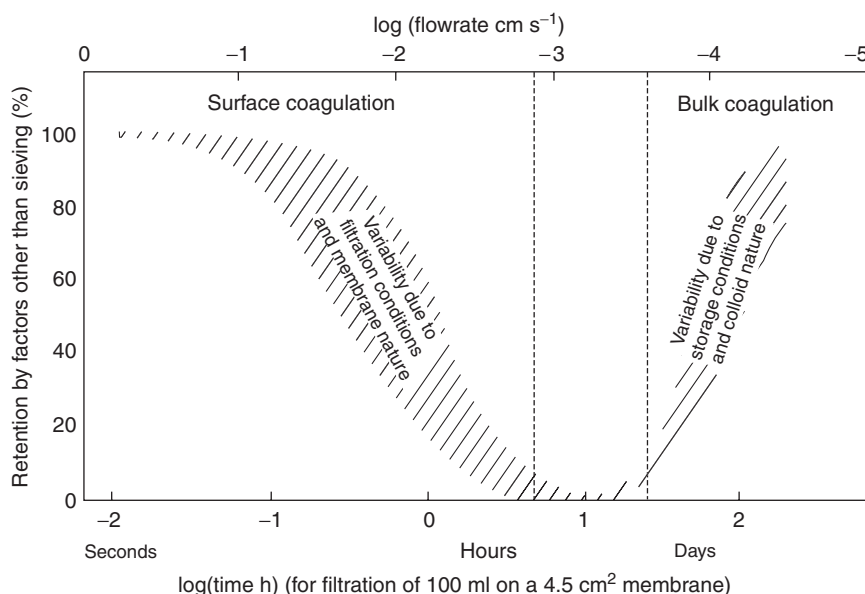
Once the artifact problems have been brought under control, the most informative characterizations will be made with a multimethod approach to sizing and chemistry (one that should include EM analyses of particle shape, internal heterogeneity, porosity, and native associations). The multimethod approach can borrow from chemistry, physics, and biology and make use of existing technology transfer entering limnology and oceanography.

Five kinds of rapid sizing techniques were noted earlier in relation to EM characterizations. Three are

discussed below: they are FFF and FCS on account of their expanding potential, and filtration-based technology because its many artifacts have been so thoroughly addressed. Depending on the research goals, selected standard quantitative techniques of chemistry should be applied to aliquots of both raw samples and fractionated samples, whenever feasible, to provide baseline information on composition. Early compositional analyses via wet chemistry have great utility for interpreting semiquantitative EM data and for selecting tests from the large suite provided by the literature of EM-based microchemistry. In general, where polydispersity confounds the interpretation of rapid analytical techniques, TEM should be employed correlatively as a detective's tool to clarify the problem and add details to existing data. Ultracentrifugation, used effectively for the whole mount preparation displayed in **Figure 1**, has not been developed as much as its potential would indicate for aquatic particle sizing.

To couple EM descriptions (on a per colloid basis) with size distribution data derived from filtration operations, it is necessary to employ a filtration apparatus in a highly controlled manner, so as to minimize operation-induced aggregation. The main contribution to this aggregation artifact is, often, an inappropriate flow rate. As revealed by **Figure 2**, there is an intermediate flow rate window that can be used for size fractionations yielding realistic data on the sizes of polydisperse, unstable colloid systems. In the light of current research needs, it is necessary to delineate this window for a given water type; TEM is essential for doing this well and for characterizing the unperturbed fractions. As a broad generalization, filtration/ultrafiltration is (potentially) a treacherous means for obtaining sizing information on aquatic colloid systems and should not be employed in a straightforward, unmonitored manner. That warning having been voiced, it must also be stated that impressive efforts have been made to identify and then minimize the artifacts of filtration/ultrafiltration, and that much valuable research has come from the application of this technology to natural waters.

FFF is a suite of elution techniques with considerable capacity to separate (by mass, size, charge) an extremely broad range of molecules, macromolecules, colloids, and sedimenting particles. Since the elution patterns are usually recorded by a nondestructive detection method, fractions can be collected for further analyses (chemical, biological, physical) when quantities are sufficient. As a tool for size distribution analyses, it offers many advantages over more conventional techniques for separating substances and materials, one of which is its breadth of applicability (particles may be charged, uncharged,



**Figure 2** A semiquantitative representation of the change in retention of a membrane filter with flow rate due to concentration polarization (high flow-rate domain) and aggregation in the filtration cell (low flow-rate domain). Only an intermediate flow-rate window is usable for size fractionation with a minimal artifact. (Reprinted with permission from Buffle J, Perret D, and Newman M (1992) The use of filtration and ultrafiltration for size fractionation of aquatic particles, colloids and macromolecules. In: Buffle J and van Leeuwen HP (eds.) *Environmental Particles*. IUPAC Environmental Chemistry Series, vol. 1, pp. 171–230. Chelsea, MI: Lewis Publishers; © Lewis Publishers, an imprint of CRC Press, Boca Raton, FL.)

random coil, globular, or fragile). Another advantage is its versatility which derives from the fact that many parameters controlling the separation can be varied precisely over wide limits, allowing one to optimize the range, speed, and power of separation. The range of particle mass to which FFF is applicable is very large, from  $10^3$  to more than  $10^{18}$  Da. As has long been the case for EM analyses of aquatic colloids in ultrathin sections, FFF size analyses were slowed by technical difficulties inherent in working with substances subject to colloid instability. Separations of aquatic colloids and macromolecules are carried out in a thin ribbon-like flow channel by applying a field (or gradient) across the thin dimension of the ribbon-like space and then initiating flow in the channel. Separations are gentle and the interactions of particles with surfaces are minimized. Particle–particle interactions can be minimized when prior sample concentration can be kept near a realistic level.

FCS has been applied only rarely to environmentally relevant aquatic systems, but the technique has a promising future. A major advantage is a high sensitivity that allows the study of native colloids *in situ* at realistic concentrations. It is also selective for specifically labeled molecules, which presents a major advantage for the analysis of chemically heterogeneous media, potentially allowing the study

of formation kinetics and the structure of complex heteroaggregates; a limitation is imposed on its potential by the paucity of naturally fluorescent aquatic colloids and macromolecules. Potential applications include: (1) the study of aggregates and the formation of flocs, known to be significant in the dispersion and transformation of contaminants; (2) analysis of the conformation of fibrillar biopolymers, abundant and active in most surface waters; and (3) the determination of diffusion in complex gels, such as those found at the surface of microorganisms.

### EM Apparatus and Its Correlative Use

While high resolution of structure is essential for many particle analyses, taking full advantage of it can impose a requirement in addition to that of minimal sample perturbation. Individual ultrathin sections and films, under examination by TEM and STEM instruments, pose few problems beyond destabilization of the embedding matrix in the path of the electron beam. However, for heterogeneous particles/materials in the upper portion of the colloidal size range and larger, there comes a requirement for a three-dimensional orientation within the particle/material. This is especially true when data on the heterogeneity within a colloid system must be assessed by statistical analyses for colloid–colloid associations



and intrasystem gradients. Measurements sufficient to obtain representative data often impose on the analyst a need for large numbers of ultrathin sections (the use of thicker sections not being an option for conventional EM). Detailed analyses of heterogeneous colloid systems in the true particle size range, especially comparative analyses requiring serial sectioning ( $< 50$  nm thick), are possible in theory but not possible in practice because of prohibitive cost (e.g., the cost of analyzing  $10^5$  sections). To minimize the number of sections to be analyzed while achieving the intraparticle orientation necessary to understand what is representative, the analyst makes use of correlative microscopy. This is defined as a strategy for utilizing several different kinds of microscopes and accessory techniques in a multi-method context to analyze a given type of specimen for different kinds of information. Examples wherein a systematic approach to orientation (by means of correlative microscopy) can prove invaluable in reducing the time and cost of analysis include the following:

- Sedimentary materials whose irregular porosity and/or low degree of crystallinity must be quantified or at least assessed.
- Coatings whose degree of patchiness in the coverage of solid surfaces must be assessed.
- Biofilms whose activities under study relate to the disposition of individual colloid species (including biological species) with respect to the interface between colloid system and aquatic milieu.
- Growing flocs whose mechanism of aggregation is sought, in time-series experiments, through morphological correlates.
- Natural associations of colloids whose frequency of occurrence must be quantified for the interpretation of some biogeochemical processes.

A general correlative strategy for microscope analyses is outlined below; detailed considerations have been published in recent reviews.

Descriptions of entire fresh particles in their natural water, by various techniques of light microscopy (LM) including confocal laser scanning microscopy, organized in concert with research goals, should serve as a starting point. This effort can be augmented by LM descriptions of particles (as whole mounts or in thick sections) preserved for purposes of substance-specific staining. The resolution limit of classical LM techniques extends down to  $\sim 0.2 \mu\text{m}$ . As a result of a revolution in LM methodology, however, it has become possible to detect structural features many times smaller than implied by the resolution limit. The main contribution to this advanced

methodology is the use of computers for the digital processing and analysis of images, as is the case for video-enhanced-contrast microscopy. Morphological and microchemical information from LM, on a positional basis, can be related to the high-resolution images of TEM through the intermediary use of many specialized biomedical techniques, one of which is conventional scanning electron microscopy (SEM). SEM provides a higher level of resolution than LM while permitting a varied selection of topographical views (of and through particles) that relate to the entire volume of a colloid system. Like LM, it has a great depth of focus relative to TEM; unlike LM, its capacity to resolve the finest details of ultrastructure can approach that of TEM. Additionally, the latest generation of ESEM instruments (environmental scanning electron microscopy) allows one to obtain topographical views almost as detailed as those provided by SEM, but with (fully or partially) hydrated samples.

When properly oriented (using LM and SEM in sequence) with respect to 'particle volume' and the heterogeneity contained therein, the analyst can then employ a selective sectioning approach to provide the ultrathin (essentially planar) specimens for TEM analyses of the finest details at a practical resolution near 1 nm. As STEM instruments continue to improve in quality, versatility, and convenience, the distinctions between TEM and SEM will become blurred. As the potential of the high-voltage EM (or HVEM) becomes more developed for use with the weakly ordered colloid systems taken from natural waters, the need for ultrathin sections should diminish. An HVEM instrument operated at  $10^3$  keV can achieve the same resolution as a TEM instrument operated at  $10^2$  keV when using sections 20 times the thickness permitted by TEM.

EM accessories and adjunct techniques have become increasingly useful on a routine basis to produce information on compositional differences within a colloid system on a 'per colloid' basis. EDS is readily applied to embedded particles (both sections and films), with greatly improved spatial resolution over the whole mount measures of the past. Conventional STEM-EDS provides qualitative elemental composition analyses for elements of atomic number greater than 10. When standards can be created with realistic matrix properties, good semi-quantitative measures can be made (relative amounts of major elements). Some specialized instruments can detect some elements of  $Z < 10$ , as can a related accessory technology, that of EELS. When precautions are taken to minimize instrument destruction of sample during irradiation, diffraction patterns can be obtained from embedded crystals in the near-nanometer

size range (least dimension), although the smallest crystals tend to provide a pattern useful only for fingerprinting purposes (when applied to samples whose gross chemical composition has been ascertained previously). In addition to selective staining and classical microchemistry, both radioautography and enzyme-site visualization techniques can be adapted from LM applications for use with TEM and STEM. One may also add to the list of analytical adjuncts the experimental tools of immunochemistry and selective extraction. Multimethod approaches to EM preparation, when used judiciously, will allow one to assess various kinds of artifact contributions when a given adjunct technique does not permit the use of a minimally perturbing technique of EM preparation.

**See also:** **Centrifugation:** Analytical Ultracentrifugation. **Field-Flow Fractionation.** **Membrane Techniques:** Ultrafiltration. **Microscopy Applications:** Environmental; Light Microscopy; Electron Microscopy; Specimen Preparation for Electron Microscopy; Scanning Electron Microscopy; Atomic Force and Scanning Tunneling Microscopy. **Particle Size Analysis.** **Sampling:** Theory.

## Further Reading

- Buffle J, Wilkinson KJ, Stoll S, Filella M, and Zhang J (1998) A generalized description of aquatic colloidal interactions: the three-colloidal component approach. *Environmental Science and Technology* 32: 2887–2899.
- Egerton RF (1996) *Electron Energy-Loss Spectroscopy in the Electron Microscope*, 2nd edn. New York: Plenum Press.
- Frosch D and Westphal C (1989) Melamine resins and their application in electron microscopy. *Electron Microscopy Reviews* 2: 231–255.
- Giddings JC (1997) Factors influencing accuracy of colloidal and macromolecular properties measured by field-flow fractionation. *Analytical Chemistry* 69: 552–557.
- Hayat MA (ed.) (1987) *Correlative Microscopy in Biology – Instrumentation and Methods*. Orlando, FL: Academic Press.
- Jackson TA and Leppard GG (2002) Energy dispersive X-ray microanalysis and its applications in biogeochemical research. In: Violante A, Huang PM, Bollag J-M, and Gianfreda L (eds.) *Soil Mineral–Organic Matter–Microorganism Interactions and Ecosystem Health*, pp. 219–260. Amsterdam: Elsevier.
- Leppard GG (1992) Evaluation of electron microscope techniques for the description of aquatic colloids. In: Buffle J and van Leeuwen HP (eds.) *Environmental Particles*. IUPAC Environmental Chemistry Series, vol. 1, pp. 231–289. Chelsea, MI: Lewis Publishers.
- Leppard GG (1992) Size, morphology and composition of particulates in aquatic ecosystems: Solving speciation problems by correlative electron microscopy. *Analyst* 117: 595–603.
- Leppard GG and Buffle J (1998) Aquatic colloids and macromolecules: effects on analysis. In: Meyers RA (ed.) *The Encyclopedia of Environmental Analysis and Remediation*, vol. 1, pp. 349–377. New York: Wiley.
- Perret D, Leppard GG, Muller M, *et al.* (1991) Electron microscopy of aquatic colloids: Non-perturbing preparation of specimens in the field. *Water Research* 25: 1333–1343.
- Sarchev K, Wilkinson KJ, and Buffle J (2001) Application of FCS to the study of environmental systems. In: Rigler R and Elson ES (eds.) *Fluorescence Correlation Spectroscopy – Theory and Applications*, pp. 251–275. Berlin: Springer.

# WATER DETERMINATION

**S A Margolis and P H Huang**, National Institute of Standards and Technology, Gaithersburg, MD, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Many physical and chemical methods have been described for the measurement of water. However, a universal method has not been found to work for this widely distributed, highly polar compound that occurs in a wide range of concentrations. For example, water forms weak bonds with other polar compounds, can be chemically bound in crystals or simply

trapped in spaces inside them, may evaporate as an azeotrope, and can range from ~100% in biomatrices to trace levels in hydrocarbons. Further, due to the different chemistries possible in a single material, the total amount of water present may not be accessible to measurement. Many methods are not specific for water and other materials with similar physical or chemical properties can bias the measurement, e.g., evaporative methods that measure total volatile material. Many of the common methods cannot measure accurately less than one milligram of water. When greater sensitivity is needed, the coulometric Karl Fischer (KF) method can measure as little as one microgram of water. Some of the methods for water in gases also have similar limits of detection.



size range (least dimension), although the smallest crystals tend to provide a pattern useful only for fingerprinting purposes (when applied to samples whose gross chemical composition has been ascertained previously). In addition to selective staining and classical microchemistry, both radioautography and enzyme-site visualization techniques can be adapted from LM applications for use with TEM and STEM. One may also add to the list of analytical adjuncts the experimental tools of immunochemistry and selective extraction. Multimethod approaches to EM preparation, when used judiciously, will allow one to assess various kinds of artifact contributions when a given adjunct technique does not permit the use of a minimally perturbing technique of EM preparation.

**See also:** **Centrifugation:** Analytical Ultracentrifugation. **Field-Flow Fractionation.** **Membrane Techniques:** Ultrafiltration. **Microscopy Applications:** Environmental; Light Microscopy; Electron Microscopy; Specimen Preparation for Electron Microscopy; Scanning Electron Microscopy; Atomic Force and Scanning Tunneling Microscopy. **Particle Size Analysis.** **Sampling:** Theory.

## Further Reading

- Buffle J, Wilkinson KJ, Stoll S, Filella M, and Zhang J (1998) A generalized description of aquatic colloidal interactions: the three-colloidal component approach. *Environmental Science and Technology* 32: 2887–2899.
- Egerton RF (1996) *Electron Energy-Loss Spectroscopy in the Electron Microscope*, 2nd edn. New York: Plenum Press.
- Frosch D and Westphal C (1989) Melamine resins and their application in electron microscopy. *Electron Microscopy Reviews* 2: 231–255.
- Giddings JC (1997) Factors influencing accuracy of colloidal and macromolecular properties measured by field-flow fractionation. *Analytical Chemistry* 69: 552–557.
- Hayat MA (ed.) (1987) *Correlative Microscopy in Biology – Instrumentation and Methods*. Orlando, FL: Academic Press.
- Jackson TA and Leppard GG (2002) Energy dispersive X-ray microanalysis and its applications in biogeochemical research. In: Violante A, Huang PM, Bollag J-M, and Gianfreda L (eds.) *Soil Mineral–Organic Matter–Microorganism Interactions and Ecosystem Health*, pp. 219–260. Amsterdam: Elsevier.
- Leppard GG (1992) Evaluation of electron microscope techniques for the description of aquatic colloids. In: Buffle J and van Leeuwen HP (eds.) *Environmental Particles*. IUPAC Environmental Chemistry Series, vol. 1, pp. 231–289. Chelsea, MI: Lewis Publishers.
- Leppard GG (1992) Size, morphology and composition of particulates in aquatic ecosystems: Solving speciation problems by correlative electron microscopy. *Analyst* 117: 595–603.
- Leppard GG and Buffle J (1998) Aquatic colloids and macromolecules: effects on analysis. In: Meyers RA (ed.) *The Encyclopedia of Environmental Analysis and Remediation*, vol. 1, pp. 349–377. New York: Wiley.
- Perret D, Leppard GG, Muller M, *et al.* (1991) Electron microscopy of aquatic colloids: Non-perturbing preparation of specimens in the field. *Water Research* 25: 1333–1343.
- Sarchev K, Wilkinson KJ, and Buffle J (2001) Application of FCS to the study of environmental systems. In: Rigler R and Elson ES (eds.) *Fluorescence Correlation Spectroscopy – Theory and Applications*, pp. 251–275. Berlin: Springer.

# WATER DETERMINATION

**S A Margolis and P H Huang**, National Institute of Standards and Technology, Gaithersburg, MD, USA

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Many physical and chemical methods have been described for the measurement of water. However, a universal method has not been found to work for this widely distributed, highly polar compound that occurs in a wide range of concentrations. For example, water forms weak bonds with other polar compounds, can be chemically bound in crystals or simply

trapped in spaces inside them, may evaporate as an azeotrope, and can range from ~100% in biomatrices to trace levels in hydrocarbons. Further, due to the different chemistries possible in a single material, the total amount of water present may not be accessible to measurement. Many methods are not specific for water and other materials with similar physical or chemical properties can bias the measurement, e.g., evaporative methods that measure total volatile material. Many of the common methods cannot measure accurately less than one milligram of water. When greater sensitivity is needed, the coulometric Karl Fischer (KF) method can measure as little as one microgram of water. Some of the methods for water in gases also have similar limits of detection.

Therefore, some knowledge of the nature and range of content of the water in a sample is desirable in choosing a method. Publications of national and international committees and of other organizations that promulgate standard methods for the measurement of water in specific matrices provide widely accepted methods for specific applications. These publications should be consulted early in deciding on an analytical strategy. The methods noted here are described in more general terms.

## Analytical Methods

The choice of a method for the analysis of water in a given sample is dependent on the nature of the sample and the manner in which the water is associated with that sample. The factors that should be considered are listed in **Table 1**.

Both physical and chemical methods are used for the measurement of water. Every chemical method is combined with a physical method that functions as a detector of the product of the chemical reaction. Physical methods usually are independent of chemical reactions and depend only on the absence of signals interfering with the water signal. **Table 2** summarizes the physical methods that have been used for the determination of water and **Table 3** the chemical methods. These tables also list the major detection methods, the limits of detection, and the compounds or groups of compounds that interfere with each type of method.

## Calibration and Standardization

The calibration and standardization of each method are essential for accurate measurements and many

materials are available for calibration. These include liquids containing known amounts of water such as water-saturated 1-octanol ( $39.2 \text{ mg ml}^{-1}$  solution), solids such as oxalic acid  $\cdot 2\text{H}_2\text{O}$  or sodium tartrate  $\cdot 2\text{H}_2\text{O}$ , and gases such as humidity fixed points of binary saturated aqueous solutions or constant-humidity gas generators. Every instrument should be calibrated at the beginning and end of each day with at least three different levels of water that bracket the range of water measurements in the test samples. This procedure will confirm the linearity of the response and the accuracy of the instrument. In addition, a matrix standard, which is similar in composition to the sample, should be used when available to insure that the instrument is responding to the test material in the expected manner. This is particularly important when complex matrices such as pharmaceuticals, foods, plants, paper, oils, or soils are the test materials.

## Sample Considerations

Knowledge of the nature of the sample and the manner in which the water is associated with the sample is essential for the accurate measurement of the water. If the sample is hygroscopic, it may absorb water from the environment and this must be independently measured or a dry box must be used. Occluded water and included water are often incompletely released if the sample is not completely dissolved in a suitable solvent. Some compounds such as copper sulfate bind their water of crystallization so strongly that it is not measured by most methods. Many of the physical methods respond to properties not unique to water (e.g., the OH signal in the infrared (IR) or nuclear magnetic resonance (NMR) methods) or

**Table 1** Factors determining method selection

Factor	Examples
Sensitivity of the method	10 mg per g to 1 $\mu\text{g}$ per g of sample
Type of water	Surface water Water occluded in spaces in crystals Water of crystallization, e.g., water in the oxalic acid crystal structure Azeotropic water, e.g., ethanol/water
Substances that bias the measurement method	Nonaqueous volatile components, e.g., chemicals that evaporate below the boiling point of water Nonaqueous materials to which the detector responds, e.g., hydroxyl compounds that overlap with IR signal for water Nonaqueous materials that interfere with a critical chemical reaction, e.g., ketones and aldehydes that react with the Karl Fischer reagent Water generated or consumed by chemical side reactions, e.g., esterification or amide formation
Type of sample	Gas, liquid, or solid
Sample solubility	Samples must dissolve in chemical assay systems to insure that all the water is available for analysis, e.g., liquids and solids in Karl Fischer reaction

**Table 2** Physical methods for the measurement of water

<i>Method</i>	<i>Detection</i>	<i>Limitations and interferences</i>
Gravimetric (GR)		
Oven drying	Weight loss	Substances volatile $\leq 100^{\circ}\text{C}$ , azeotropes, bound water
Dessication	Weight loss	Bound or occluded water
Thermogravimetric (TGR)	Weight loss	Substances volatile $\leq 100^{\circ}\text{C}$ , azeotropes, bound water
Lyophilization	Weight loss	Bound or occluded water, sublimates
Condensation	Weight of condensate	Substances volatile $\leq 100^{\circ}\text{C}$ , azeotropes, bound water
Distillation	Volume or weight	Substances volatile $\leq 100^{\circ}\text{C}$ , azeotropes, bound water
Chromatography		
Gas (GC)	GC detectors	Column selectivity, azeotropes
Head space	GC detectors	Column selectivity, partition coefficient accuracy, azeotropes
Ion exclusion	UV detector	Requires very dry solvent, chemical reaction not totally specific
Spectroscopic		
Infrared (IR)	Various IR detectors	Water signal overlaps amine and hydroxyl signals, influenced by hydrogen bonding of water, trimer, and dimer forms of water have different signals
NMR	NMR spectrometer	Limited to samples with $> 1\%$ water, exchange of water protons with other protons, nonwater protons
Visible	Spectrophotometer	Requires dry organic polar reagents and is useful for compounds that do not react with the indicator reagent such as those listed in <b>Table 3</b>
Ultraviolet	Spectrophotometer	Requires dry organic polar reagents and is useful for compounds that do not react with the indicator reagent such as those listed in <b>Table 3</b>
Mass spectrometry (MS)	Mass spectrometer	Requires complex calibration, interferences ammonia and atmospheric water
Microwave	Microwave spectrometer	Dependent on the properties of the material, limited range of measurement
Electrical		
Dielectric	AC bridge system	Most materials do not behave ideally, each individual material must be calibrated separately
Conductimetry	Conductivity meter	Requires the presence of a low concentration of a background electrolyte, and that the other components be nonconductive
Coulometry	Coulometer	Requires the quantitative removal of water from sample by an electrochemical reaction
Neutron scattering	Neutron collector	Requires neutron source, nonaqueous hydrogen sources interfere
Differential scanning calorimetry (DSC)	Heat sensor	Requires that the heat of dissociation of the water is compensated by the heat of transition of the sample, useful for determining water in some hydrates
Evolved gas	IR, MS, GC, GR	Substances volatile $\leq 100^{\circ}\text{C}$ , azeotropes, bound water
Density	Densitometer	Requires that only water be extracted into analytical solvent
Refractive index (RI)	RI meter	Requires that only water be extracted into analytical solvent
Water (humidity) sensors		
Thin film polymer/metal oxide	Capacitance or resistance meter	Measures water in gases; range $10^{-3}$ –20% water
Oscillating crystal	Microbalance	Measures water in gases; range $10^{-5}$ –20% water
P <sub>2</sub> O <sub>5</sub> hygrometer	Coulometric	Measures water in gases; range $10^{-5}$ –20% water
Dew/frost-point hygrometer	Thermometer	Measures water in gases; range $10^{-5}$ –90% water

gravimetric evaporative methods that measure all compounds that volatilize at or below  $100^{\circ}\text{C}$ . Thus, the selected method must take into account the properties, composition, and structure of the sample.

### Method Selection

The methods for the measurement of water utilize one of the fundamental physical properties of the water molecule such as a phase change (evaporation) or the absorption of energy (spectroscopy). The

**Table 3** Chemical methods for the measurement of water

Method <sup>a</sup>	Fundamental reaction	Detection	Interferences and limitations
Karl Fischer <sup>b</sup>	$\text{H}_2\text{O} + \text{I}_2 + (\text{RNH})\text{SO}_3\text{CH}_3 + 2\text{RN} \rightarrow (\text{RNH})\text{SO}_4\text{CH}_3 + 2(\text{RNH})\text{I}$	Colorimetric titration, coulometry, or amperometric titration	Substances that react with methanol or iodine, insoluble compounds, range $\geq 1 \mu\text{g}$ water
Acyl halides	$\text{C}_5\text{H}_5\text{NCH}_3\text{COCl} + \text{H}_2\text{O} \rightarrow \text{C}_5\text{H}_5\text{NHCl} + \text{CH}_3\text{COOH}$	Titration with alkali	Range varies from 0.02% to 3% water depending upon the sample
Anhydrides <sup>c</sup>	$(\text{R}^1\text{CO})_2\text{O} + \text{H}_2\text{O} \rightarrow 2\text{R}_2^1\text{COOH}$	Conductimetric, polarimetric, thermometric, or UV	Alkenes, furans, aromatic, ketones, formic acid, aldehydes, amides, and oximes; range 0.2–10% water depending upon the sample
Calcium carbide	$\text{CaC}_2 + \text{H}_2\text{O} \rightarrow \text{C}_2\text{H}_2 + \text{Ca(OH)}_2$	Acid titration	Not always quantitative, range $\geq 1\%$ water depending upon the sample
Hydrides	$\text{CaH}_2 + \text{H}_2\text{O} \rightarrow 2\text{H}_2 + \text{Ca(OH)}_2$	Acid titration or thermometry	Acetone, alcohols, ammonia, and other compounds; range 0.2–90% water depending upon the sample
Magnesium nitride	$\text{Mg}_3\text{N}_2 + 6\text{H}_2\text{O} \rightarrow 3\text{Mg(OH)}_2 + 2\text{NH}_3$	Kjeldahl method for nitrogen	Methanol and compounds that release ammonia, range $\geq 1\%$ water
Lead tetraacetate	$\text{Pb(OOCCH}_3)_4 + \text{H}_2\text{O} \rightarrow \text{PbO}_2 + 4\text{CH}_3\text{COOH}$	Iodine titration	Used for water in acetone, pyridine, and diethylether
Tertiary butyl orthovanadate	$[\text{CH}_3)_3\text{CO}]_3\text{VO} + \text{H}_2\text{O} + \text{NH}_3 \rightarrow \text{NH}_4\text{VO}_3 + 3(\text{CH}_3)_3\text{COH}$	Titration with ammonium iron(II) sulfate	Range $\geq 0.5\%$ water

<sup>a</sup>For maximum accuracy, sensitivity, and precision all of the chemical methods require that the reaction be performed in a closed system or a dry box, that the solvents be dried, and that the samples be completely soluble in the solvent.

<sup>b</sup>RN = pyridine or an alkyl amine.

<sup>c</sup>R<sup>1</sup> = CH<sub>3</sub>(CH<sub>2</sub>)<sub>x</sub>.

accuracy and precision of any given method depend upon the ability of a method using a basic physical principle to detect and distinguish between the compound of interest and other compounds whose properties are similar to the compound of interest (water as in Table 2 or a chemical compound derived there from as in Table 3).

At the outset it is necessary to define two terms used in water measurement, humidity, and moisture. Humidity is defined as a relatively high mass fraction of water or water vapor. Moisture is defined as condensed or diffused liquid. For example, a distillation method measures moisture, i.e., a mixture of compounds (which may include water) that undergoes a phase transition in a specific temperature range.

The physical properties of water that are used for its measurement, e.g., the absorption of wave motion or magnetic energy, or the conduction of electricity, require separation of water from compounds that may interfere with the measurement (Table 2). Usually water is separated by a phase transition from a liquid or solid to a gas. Other methods utilize chemical properties (Table 3) for water determination. Chemical measurements are achieved either by

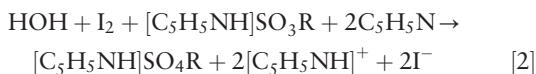
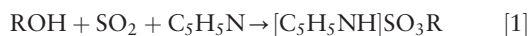
measuring the water *in situ* or by removing the water from the sample and then quantifying it.

The nature and purity of the sample are also a major consideration. If the sample is primarily a single chemical compound (as opposed to a complex sample like soil or a biological tissue) then many of the methods summarized in Tables 2 and 3 can be utilized to measure the water with a relatively high degree of accuracy. Alternatively, if the material is heterogeneous, the water can be removed from the sample and the amount of sample remaining can be measured gravimetrically, or the water in the moisture that has been removed from the sample can be measured. The water can also be measured directly in the intact sample. The methods that measure the water in intact heterogeneous samples are subject to multiple interferences (see Tables 2 and 3) and are not very sensitive, i.e., they are limited to water amounts that are greater than 0.1%. These methods that measure water in intact insoluble samples (e.g., microwave or IR absorption, or dielectric properties) are frequently used to assess product quality during or after processing and play a very significant role in water measurement. (For applications see Further Reading.)

## Measurement of Water in Liquids and Solids

### Karl Fischer Method

The most widely used method for the measurement of water in liquids and solids is the KF method. The fundamental reaction with water is



(ROH = alcohol, other amines can replace pyridine).

The  $2\text{I}^-$  formed in the second reaction is determined either by visual chemical titration with a reagent such as sodium thiosulfate in the presence of a suitable endpoint indicator or by amperometric, coulometric, or photometric titration methods. The most sensitive KF methods for the measurement of iodine are coulometric. For both the volumetric–amperometric and coulometric methods the endpoint is detected by a pair of platinum electrodes called the indicator electrodes. An electrical potential (100–400 mV) is applied across the electrodes to balance the circuit and the endpoint is reached when the concentration of  $\text{I}_2$  ( $\approx 50 \mu\text{mol l}^{-1}$ ) depolarizes the cathode deflecting a galvanometer. The volumetric method measures the amount of standardized reagent necessary to depolarize the platinum electrodes. The coulometric method utilizes, in addition to the indicator electrodes, a second pair of platinum electrodes (generator electrodes) that electrolytically convert the  $\text{I}^-$  to  $\text{I}_2$ . The current consumed in this process is used to calculate the amount of water using the equation that describes Faraday's laws of electrolysis.

The advantages of the coulometric method are its simplicity of operation, its increased sensitivity, and the fact that it does not require a standardized reagent, such as water saturated 1-octanol, to calculate the water content but only to assess the accuracy of the instrument. The advantage of the volumetric method is that it permits the use of a wider range of very polar solvents as well as nonpolar solvents for dissolving the sample in the titration vessel. Furthermore, the volumetric titration vessel can be heated to enhance the dissolution of slightly soluble samples. Either the volumetric or the coulometric titration instrument can be joined with an oven (evaporation) or a distillation apparatus (azeotropic distillation). In this configuration, the moisture that is volatilized from the sample can be transported to the titrator with a dry gas and the evaporated water measured either coulometrically or volumetrically. The recent

addition of an oven with an autosampler to a coulometric titrator has the added advantage of automated analysis of small amounts of water in insoluble samples with increased sensitivity ( $\geq 50 \mu\text{g}$ ) and with good precision.

### Gravimetric Methods

Gravimetric methods have been used widely because of their apparent simplicity. Their utility is limited by the accuracy of the balance that is used to make measurements. More importantly, these methods measure moisture, of which water may only be one component. If there are no other volatile compounds in the material of interest and the water is not bound, this method is satisfactory but these are two very important caveats.

### Spectroscopic Methods

All spectroscopic methods are based upon the absorption of energy of a specific wavelength, a group of wavelengths, or range of wavelengths. The fundamental process of each of these spectroscopic methods is described under specific headings in this Encyclopedia. Specific wavelengths in the ultraviolet (UV) or visible regions of the spectrum are used for the detection and quantification of light-absorbing compounds that result from the chemical reactions of specific compounds with water (see Table 3). The NMR and mass spectrometry instruments are expensive and complex to operate. The NMR method consists of the magnetic perturbation of molecules caused by the absorption of electromagnetic energy. Two basic types of NMR instruments are used. One is based on the continuous-wave or wide-line method and the other on the pulsed or transient-response method. These methods are limited to compounds that do not exchange OH hydrogen with other exchangeable hydrogen and do not have OH signals that overlap with the water signal. This method is not very sensitive (i.e., water concentrations  $> 1\%$ ). Mass spectrometry consists of electron bombardment of water molecules yielding a spectrum consisting primarily of the peak due to the  $\text{H}^{16}\text{OH}^+$  ion. Usually, the contribution from other ions is negligible. This method can give reliable results from 10 to 0.01% water, when properly calibrated and applied. Microwave analysis is based on the fact that the permittivity of the water is much higher than that of most dry substances. A small amount of water can change the permittivity of the material and can be detected by microwave measuring equipment. This method has a very limited range because the efficiency of these instruments depends on the fluctuation of the density of the sample. However, the sensitivity can be increased



by minimizing the effect of the density fluctuation by the simultaneous measurement of the attenuation and the phase shift of the electromagnetic radiation. This method is capable of distinguishing between bound and free water because the relaxation frequencies for the two forms are significantly different.

The measurement of water by IR methods has become more feasible with the application of laser and Fourier transform infrared (FTIR) technologies. The spectral bands in which water can be determined include both the fundamental IR region ( $4000\text{--}1600\text{ cm}^{-1}$ ) and the near IR region of the combination and overtone absorption bands ( $143\,000\text{--}5000\text{ cm}^{-1}$ ). The IR absorption bands are due to OH stretching, asymmetric OH stretching, and HOH bending. The major problem with this method is that the location of these bands shifts with the matrix and the molecular interaction of water, which forms monomers, dimers, and trimers. Furthermore, the absorption bands of hydroxyl and amine groups of other molecules may overlap with the bands produced by water making quantification difficult or impossible particularly for heterogeneous samples. FTIR is a sensitive method that detects microgram quantities of water when there are no interfering signals. It is applicable to gases, liquids, and solids. However, since absorption bands are prone to shift for each new sample it is necessary to standardize and calibrate the instrument with a matrix standard or by spiking the sample.

## Measurement of Water in Gases

In gases, water measurement capabilities range from parts per billion (ppb) to 100% at temperatures from  $-100^{\circ}\text{C}$  to  $100^{\circ}\text{C}$  and at pressures from 4000 to  $10^7\text{ Pa}$ . A frost-point temperature of  $-100^{\circ}\text{C}$  corresponds to water vapor concentration of 14 ppb ( $14 \times 10^{-7}\%$ ) by volume or 8 ppb ( $8 \times 10^{-7}\%$ ) by weight in air at standard atmospheric pressure. Smaller trace quantities do not appear to have any significant technological, commercial, or scientific importance except in industry where ultrahigh purity of gases are required in the processing of semiconductor and microelectronic devices. The major industrial application for utilizing saturated water vapor is in steam technology.

Five primary technologies are utilized to determine the water vapor content in gases:

1. Dew/frost-point hygrometer – utilizing a chilled mirror surface. The mirror is maintained at the dew-point temperature by automatically controlling the current through a thermoelectric cooler. A  $100\text{-}\Omega$  platinum resistance thermometer is embedded beneath the surface of the mirror to measure the temperature at which dew or frost forms.
2. Coulometric hygrometer – utilizing a thin coating of phosphorous pentoxide ( $\text{P}_2\text{O}_5$ ). Water is completely absorbed by the  $\text{P}_2\text{O}_5$ , forming phosphoric acid, which is electrolyzed to oxygen and hydrogen. The electrolytic current generated is proportional to the amount of water vapor contained within the sample.
3. Electric hygrometer – utilizing a substrate, such as a thin film of polymer or aluminum oxide. Water from a sample is absorbed by the film that causes changes in the electrical properties of the film, such as capacitance or resistance. The change of the electrical quantity is related to the amount of water vapor contained in the gas sample and the amount of water is calculated by use of a calibration curve derived from standards.
4. Oscillating crystal – is a microbalance that utilizes a piezoelectric crystal, such as quartz, pulsed by high-frequency electrical energy that causes it to oscillate. Water from a sample is absorbed on a thin film of polymer or metal oxide deposited on the crystal, causing a change in the oscillating frequency. This change is proportional to the amount of water vapor contained within the sample. This is a gravimetric method.
5. Spectroscopic analyzer – utilizing electromagnetic radiation, such as IR or UV radiation. Water from a sample absorbs the radiation resulting in a decrease of radiation emanating from the gas that is proportional to the amount of water vapor contained within the gas sample.

Differences exist between the capabilities of measurements by the different hygrometers. The spectroscopic analyzer has a limited range of capability ( $10^{-4}\text{--}10^{-7}\%$  water), lacks simplicity of operation, and requires significant maintenance, but it has a fast response time. The oscillating crystal and electrical hygrometers have a broad range of capability ( $10^{-4}\text{--}1\%$ ), oscillating crystal hygrometer, the coulometric hygrometer, and the spectroscopic analyzer have high sensitivity in the range to  $10^{-4}\text{--}10^{-7}\%$  water. The chilled-mirror hygrometer is capable of measuring the dew-point temperature from  $0^{\circ}\text{C}$  to  $90^{\circ}\text{C}$ . With cooling it can measure a broad range of frost-point temperature range from  $0^{\circ}\text{C}$  to  $-90^{\circ}\text{C}$ . The full measurement range of the chilled-mirror hygrometer is  $10^{-5}\%$  to 90% water.

In addition to measurements of the humidity of atmospheric air, which is the most frequent, measurements are also made in such technologically important gases as argon, helium, neon, nitrogen, oxygen, methane, sulfur dioxide, refrigerants, and natural gas.



The above mentioned five primary technologies have been utilized to conduct the measurements for these gases with limited accuracy. The best of the commercial hygrometers have uncertainties  $\sim 1\text{--}2\%$  of the reading. Most hygrometers tend to have uncertainties  $\sim 3\text{--}10\%$  at  $10^{-7}\%$  water levels of concentration.

*See also:* **Conductimetry and Oscillometry.** **Coulometry.** **Distillation.** **Gas Chromatography:** Overview. **Gravimetry.** **Infrared Spectroscopy:** Overview; Near-Infrared. **Ion Exchange:** Ion Chromatography Applications. **Mass Spectrometry:** Overview. **Microwave Spectroscopy.** **Nuclear Magnetic Resonance Spectroscopy:** Principles. **Titrimetry:** Potentiometric.

### Further Reading

- ASTM (2004) *Annual Book of ASTM Standards*. Pennsylvania: ASTM International.
- Charsley EL and Warrington SB (1992) *Thermal Analysis – Techniques and Applications*. Special Publication No. 117. Cambridge: Royal Society of Chemistry.
- Fritz JS (1991) Principles and applications of ion-exclusion chromatography. *Journal of Chromatography* 546: 111–118.
- Horwitz W (ed.) (2000) *Official Methods of Analysis of AOAC International*. Maryland: AOAC International.
- Huang PH (1997) Technical characteristics of selected hygrometers for humidity measurement and control. ISA Italy Section, 1: 9–26.
- Khankari RK, Law, and Grant DJW (1992) Determination of water in pharmaceutical hydrates by differential scanning calorimetry. *International Journal of Pharmaceutics* 82: 117–127.
- Kraszewski A (ed.) (1996) *Microwave Aquametry*. New Jersey: IEEE Press.
- Mitchell J Jr. and Smith DM (1977) *Aquametry Part I, A Treatise on Methods for the Determination of Water*. New York: Wiley.
- Mitchell J Jr. and Smith DM (1980) *Aquametry Part III, the Karl Fischer Reagent; A Treatise on Methods for the Determination of Water*, 2nd edn. New York: Wiley.
- Pande A (1974) *Handbook of Moisture Determination and Control: Principles, Techniques, and Applications*, vol. 1. New York: Dekker.
- Pande A (1975) *Handbook of Moisture Determination and Control: Principles, Techniques, and Applications*, vols. 2–4. New York: Dekker.
- Pyper JW (1985) The determination of moisture in solids, a selected review. *Analytica Chimica Acta* 170: 159–175.
- Reidel-de-Haen (1988) *Hydrenal Manual, Eugen Scholz Reagents for Karl Fischer Titration*. Dordrecht: Reidel-de-Haen.
- Rubinson J and Robinson K (1998) *Contemporary Chemical Analysis*. NJ: Prentice-Hall.
- Schmidt SJ (1991) Determination of moisture content by pulsed nuclear magnetic resonance spectroscopy. In: Levine H and Slade L (eds.) *Water Relationships in Foods*. New York: Plenum.
- Smith DM and Mitchell J (1984) *Electrical and Electronic Methods; A Treatise on Methods for the Determination of Water*, 2nd edn. New York: Wiley.
- Willard HH, Merritt L, Dean J, and Settle F (1988) *Instrumental Methods of Analysis*, 7 edn. California: Wadsworth Publishing.

## WET DIGESTION

*See* SAMPLE DISSOLUTION FOR ELEMENTAL ANALYSIS: Wet Digestion

## WINE

*See* FOOD AND NUTRITIONAL ANALYSIS: Wine

## WORKPLACE AIR

*See* AIR ANALYSIS: Workplace Air



# X-RAY ABSORPTION AND DIFFRACTION

Contents

**Overview**

**X-Ray Absorption**

**X-Ray Diffraction – Powder**

**X-Ray Diffraction – Single Crystal**

## Overview

**F Adams**, University of Antwerp, Antwerp, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

X-rays constitute the part of the electromagnetic spectrum with wavelengths between  $\sim 0.01$  and  $10\text{ nm}$  ( $0.1\text{--}100\text{ keV}$ ), which is between gamma radiation at the short-wavelength/high-energy side and the vacuum ultraviolet region at the long-wavelength/low-energy side; these boundaries are not clearly defined. Both wavelength units and energy units are used for their identification. With  $hc/\lambda = \text{energy}$ , the conversion between them is  $\lambda\text{ (}\text{\AA}\text{)} = 12.36/E\text{ (keV)}$ .

X-rays were discovered in 1895 by Röntgen (Nobel prize, 1901) when electrons (cathode rays) were decelerated on the walls of a discharge tube. This new type of radiation showed characteristics similar to those of light but, due to limited experimental facilities, Röntgen nor his contemporaries could find evidence for the fundamental properties of this light: polarization, diffraction, reflection, and refraction. Hence, they were described as 'X'-rays (for unknown). In 1896, Perrin was the first to be able to measure the intensity of X-rays by means of an air ionization chamber. In the period 1906–1911, Barkla discovered that X-rays could be polarized and that there was evidence of element-specific absorption edges and emission line series, labeled K, L, M, etc. In 1912, von Laue, Friedrich, and Knipping established the wave character of X-rays by showing that they could be diffracted. The next year, Coolidge introduced the hot filament, high-vacuum X-ray tube, which is still used today. By then, the

experiments of Bragg had shown that the X-ray spectrum emerging from such a device consists of a continuum, on which a specific line spectrum characteristic of the anode material is superimposed. In 1913, Moseley established the relationship between the wavelength of the characteristic lines and the atomic number, thus providing the basis of qualitative and quantitative X-ray spectrochemical analysis. In the same period, Chadwick observed that X-rays can also be produced by bombardment of a material with  $\alpha$ -particles, thus laying the foundation of particle- or ion-induced X-ray emission (PIXE) analysis. Between 1913 and 1923, Siegbahn systematically measured the wavelengths of the X-ray spectrum of the elements. In 1920, Bergengren observed the effect of the chemical state of the absorber on the X-ray absorption spectra of materials and, in 1923, Glocker and Frohnmeyer applied X-ray absorption (XAS) spectrometry for the first time. In 1922, Hadding initiated the use of the X-ray spectra of minerals for chemical analysis. In the period 1923–1932, Hevesy, Coster, Nishina, Glocker, and Schreiber investigated the possibilities of X-ray fluorescence (XRF) spectroscopy as a means for qualitative and quantitative elemental analysis.

The earliest commercially available X-ray spectrometer appeared on the market in 1938. Ten years later, Friedman and Birks built the prototype of the first commercial X-ray secondary emission instrument. Somewhat later, Castaing and Guinier built the first electron probe microanalyzer using a focused electron beam to induce X-ray emission of the elements present in microscopic samples using electron probe X-ray microanalysis (EPXMA). The use of protons in PIXE for chemical analysis was first demonstrated by Sterk in 1964. From the 1960s until the present day, the use of large-particle accelerators (synchrotron rings) has resulted in a dramatic

increase in the variety, scope, and applicability of X-ray-based methods for the characterization and analysis of materials using X-ray fluorescence analysis and related methods.

## Properties of X-Rays

On their passage through matter, X-rays undergo different modes of interaction, reflection, refraction, diffraction (e.g., by slits, gratings, or crystal structures), polarization, elastic and inelastic scattering, photoelectric absorption, and electron–positron pair production (above twice the rest mass energy of the electron; 1.02 MeV, i.e., for very hard X-rays and  $\gamma$ -rays). Diffraction forms the basis of the structural investigation of crystalline materials; it is also used in wavelength dispersive modes of X-ray detection for XRF. The reflective and refractive properties of X-rays are not directly employed in analytical methods but can sometimes improve the analytical qualities of a method, as in the case of total-reflection X-ray fluorescence analysis (TXRF) and in the use of X-ray optics for focusing X-rays. Photoelectric absorption of X-ray photons is employed in many methods such as XRF, XAS, and XPS (X-ray photoelectron spectroscopy, also called electron spectroscopy for chemical analysis, ESCA). Scatter phenomena are usually not employed directly in analytical applications but may be used to influence the polarization of primary X-rays, or to estimate additional properties of materials under investigation (e.g., the thickness and/or mean atomic number of materials).

### Interaction of X-Rays with Matter

A narrow, monochromatic beam of X-rays of energy  $E$  and intensity  $I_0$ , which is transmitted through a foil of thickness  $d$  is attenuated to a value  $I(d)$ . The attenuation is the result of the interaction of the X-ray photons with matter and varies with the foil thickness according to

$$I(d) = I_0 e^{-\mu_L d} = I_0 e^{-\mu \rho d} \quad [1]$$

$\mu_L$  and  $\mu$  are, respectively, the linear and mass absorption coefficients of the material and depend on the photon energy and on the composition of the foil;  $\mu = \mu_L / \rho$  where  $\rho$  is the density of the material. For a complex material consisting of  $n$  elements, each element present with a weight fraction  $w_i$  ( $1 < i < n$ ), the mass absorption coefficients can be approximated as a sum of the atomic components:

$$\mu(E) = \sum_i w_i \mu_i(E) \quad [2]$$

where  $\mu_i$  represents the mass absorption coefficient of the  $i$ th atomic species. In its turn,  $\mu_i$  can be written as

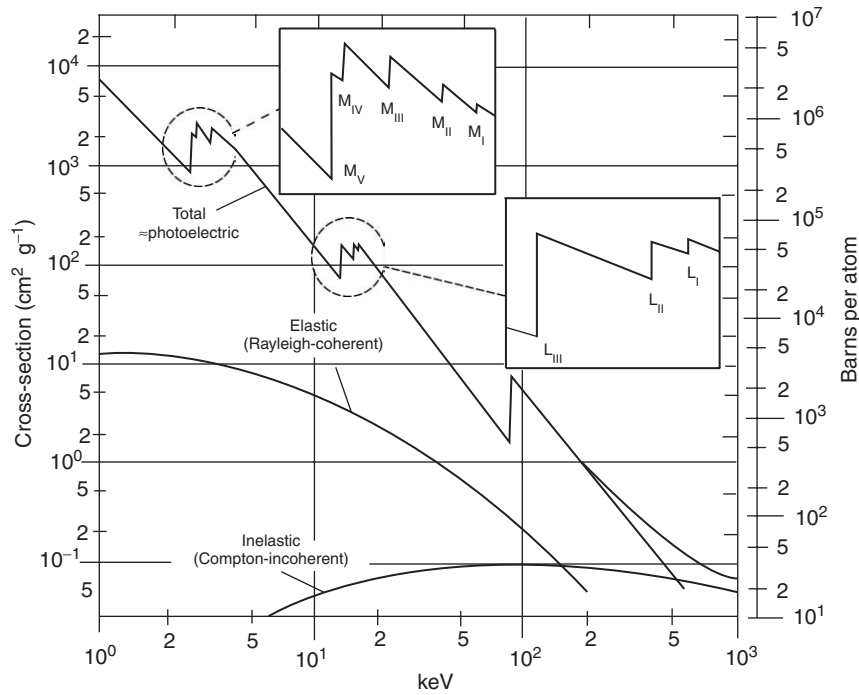
a sum of three contributions:

$$\mu_i(E) = \tau_i(E) + \sigma_i(E) + \pi_i(E) \quad [3]$$

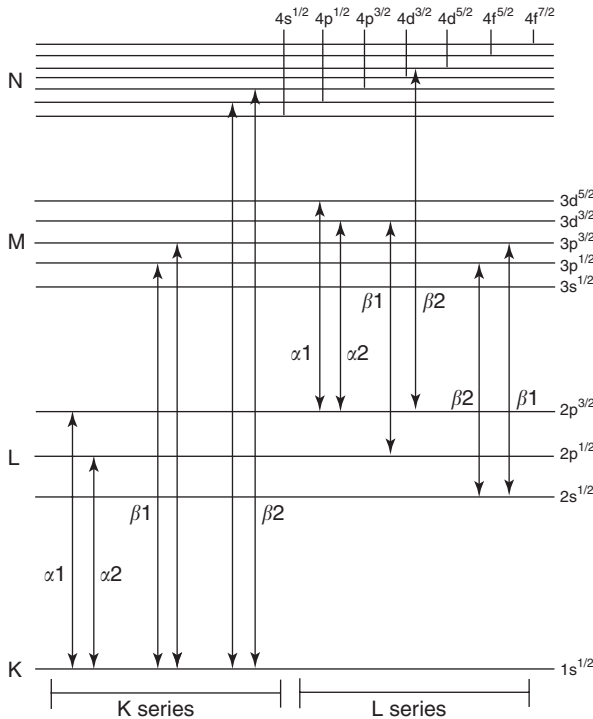
reflecting the three distinct modes through which X-rays interact with matter.  $\tau$ ,  $\sigma$ , and  $\pi$  (usually expressed in  $\text{cm}^2 \text{g}^{-1}$ ), respectively, represent the total cross-section for photoelectric absorption, for scattering, and pair production. Figure 1 shows the variation of the cross-sections as a function of photon energy for lead. Pair production, which only occurs for X-rays with energy above 1.022 MeV, is not discussed further. Overall, in the energy region up to 100 keV the photoelectric effect dominates.

**Photoelectric interaction** In a photoelectric interaction, X-ray photons are absorbed, i.e., their energy is expended imparting kinetic energy to orbital electrons of absorber atoms. The direct measurement of the photoelectrons is used in XPS/ESCA. As the energy of the original photon must be larger than the binding energy of the ejected electron, the photoelectric cross-section  $\tau(E)$  features abrupt discontinuities (called ‘absorption edges’) at photon energies corresponding to those of the various electronic levels in the atom. These edges are denoted with the letters K, L<sub>I</sub>, L<sub>II</sub>, L<sub>III</sub>, M<sub>I</sub>, ..., M<sub>V</sub>, ... and correspond to the ejection of electrons from the  $1s^{1/2}$ ,  $2s^{1/2}$ ,  $2p^{1/2}$ ,  $2p^{3/2}$ ,  $3s^{1/2}$ , ...,  $3d^{5/2}$ , ... orbitals, respectively. As a result of the creation of inner-shell vacancies, the atom is left in an excited state. A relaxation process follows within  $< 10^{-16}$  s involving the transition of an electron of a higher shell (and thus of higher energy) to the shell where the vacancy was created (e.g., from the L<sub>III</sub> to the K level). A schematic overview of some transitions is presented in Figure 2. The excess energy freed during the transition is either emitted as an X-ray photon (a radiative relaxation) or employed in an internal conversion process in which a second electron from a higher shell is ejected from the atom. Such electrons are called Auger electrons (see Figure 3) and are used in Auger emission spectrometry (AES, not to be confused with atomic emission spectroscopy). In the de-excitation there is competition between X-ray (radiative) emission and Auger emission, the first effect being prominent for the high atomic number elements, the latter for low  $Z$ -elements. The relative fraction of both is the radiative and the Auger yield.

Where radiative relaxation takes place, each transition shown in Figure 2 induces the emission of radiation of specific wavelength/energy, characteristic for the atomic species involved. In Figure 2, the terminology adopted by Siegbahn for the characteristic X-ray lines is used, showing that in an iron



**Figure 1** Mass absorption coefficient of lead as a function of energy. The contribution of various interaction modes is indicated. The inserts show details of the variation in the vicinity of absorption edges. 1 barn =  $10^{-28} \text{ m}^2$ .



**Figure 2** Commonly used terminology of energy levels and X-ray lines.

atom, for example, the radiative  $L_{III} \rightarrow K$  transition corresponds to the emission of a characteristic Fe  $K_{\alpha}$  X-ray photon. The IUPAC notation for this line is Fe  $K-L_3$ . Similar to the characteristic X-rays, Auger

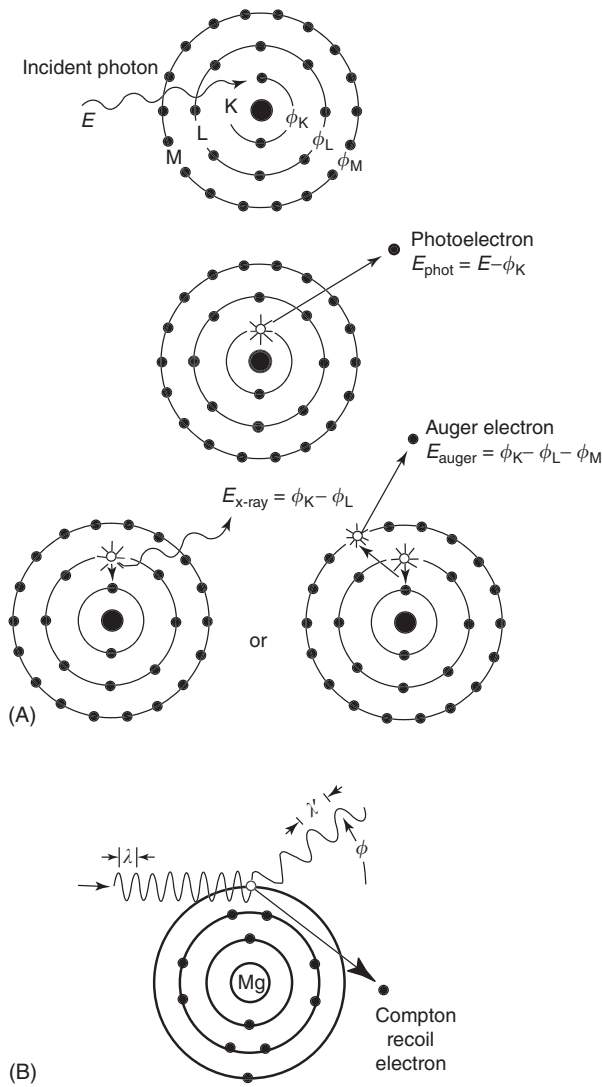
electrons feature discrete energy levels, specific for the atomic species they were ejected from. For the case illustrated in **Figure 3A**, the Auger electron energy  $E_{KLM} = (\phi_K - \phi_{L(III)}) - \phi_{M(V)}$ , where  $(\phi_K - \phi_{L(III)})$  is the energy freed when the vacancy in the K shell is filled and  $\phi_{M(V)}$  the energy required to promote the Auger electron from the  $M_V$  level out of the atom.

Inner-shell vacancies can also be a result of interactions with energetic particles (electrons, protons, etc.); thus, emission of characteristic X-rays and of Auger electrons also occurs when matter is irradiated with electron, proton, or heavier ion beams. These effects are employed in various methods for elemental analysis including PIXE mentioned earlier.

**Scattering of X-rays** Two types of photon scattering can be distinguished, depending on whether the scattered photon loses energy during the process: Compton (or inelastic) scattering and Rayleigh (or elastic) scattering. In terms of their total cross-sections ( $\sigma_C$  and  $\sigma_R$ , respectively), this can be expressed as

$$\sigma(E) = \sigma_C(E) + \sigma_R(E) \quad [4]$$

The principle of Compton scattering is shown in **Figure 3B** and involves the interaction of a photon with a weakly bound atomic electron. Depending on the scattering angle  $\theta$ , the final photon energy  $E = h\nu$



**Figure 3** (A) Schematic representation of the photoelectron absorption followed by characteristic photon emission or Auger electron emission. (B) Compton scattering of an X-ray photon by a magnesium atom.

is given by

$$h\nu = \frac{h\nu_0}{1 + (h\nu_0/m_0c^2)(1 - \cos \theta)} \quad [5]$$

where  $h\nu_0$  is the original photon energy,  $c$  the speed of light, and  $m_0$  the electron rest mass. The inelastically scattered photon has a longer wavelength than before the interaction; the energy difference  $h(\nu_0 - \nu)$  being transferred to the electron, which may be ejected from the atom. The original momentum of the electron also influences the energy transfer, an effect known as Doppler broadening of the Compton lines.

In the case of unpolarized radiation, the probability  $P_c(\theta) d\theta$  for a photon to be inelastically scattered between an angle  $\theta$  and  $\theta + d\theta$  is proportional to the Klein–Nishina differential scattering cross-section

$d\sigma_{\text{KN}}/d\Omega$ :

$$\begin{aligned} P(\theta) d\theta &= \frac{2\pi d}{\sigma_C} \frac{d\sigma_{\text{KN}}}{d\Omega} S(x, Z) \sin \theta d\theta \\ &= \frac{\pi r_0^2}{\sigma_C} \left( \frac{h\nu}{h\nu_0} \right)^2 \left( \frac{h\nu_0}{h\nu} + \frac{h\nu}{h\nu_0} - \sin^2 \theta \right) \\ &\quad \times S(x, Z) \sin \theta d\theta \end{aligned} \quad [6]$$

where  $S(x, Z)$  is the incoherent scattering function,  $Z$  the atomic number of the scattering atom, and  $x = \sin(\theta/2)/\lambda$  the momentum transfer parameter. In the case of plane-polarized radiation, the Klein–Nishina differential cross-section also depends on the angle  $\alpha$  the scattered photon makes with the plane of the electrical field vector of the incident photon, measured perpendicular to the original direction of motion of the photon.

The  $\cos^2 \alpha$  factor causes the differential cross-section to be small for small values of  $\alpha$  so that plane-polarized radiation is preferentially scattered out of the plane of polarization:

$$\frac{d\sigma_{\text{KN}}}{d\Omega} = \frac{r_0^2}{2} \left( \frac{h\nu}{h\nu_0} \right)^2 \left( \frac{h\nu_0}{h\nu} + \frac{h\nu}{h\nu_0} - 2 \sin^2 \theta \cos^2 \alpha \right) \quad [7]$$

In the case of Compton scattering, there is no phase relationship between the incident and scattered X-rays; therefore, this type of interaction is often called incoherent scattering. Not all incoherently scattered radiation, however, has been inelastically scattered. In the case of Rayleigh scattering, the energy of the photon remains unchanged and there exists a definite phase relationship between incident and scattered waves; this type of scattering therefore always occurs in a coherent fashion. The intensity of the radiation elastically scattered by an atom is obtained as the sum of contributions of the  $Z$  bound electrons. The interference of these coherent waves is always constructive provided the phase change over an atomic diameter corresponds to less than  $\lambda/2$ , i.e.,  $4\pi r_a x < 1$  where  $r_a$  is the effective atom radius and  $x$  the momentum transfer parameter.

The probability  $P_R(\theta) d\theta$  for Rayleigh scattering of unpolarized radiation within an angle  $\theta$ ,  $\theta + d\theta$  is given by

$$\begin{aligned} P_R(\theta) d\theta &= \frac{2\pi d}{\sigma_R} \frac{d\sigma_R}{d\Omega} \sin \theta d\theta \\ &= \frac{\pi r_0^2}{\sigma_R} (1 + \cos^2 \theta) |F(x, Z)|^2 \sin \theta d\theta \end{aligned} \quad [8]$$

where  $d\sigma/d\Omega_R$  is called the Rayleigh differential cross-section and  $F(x, Z)$  is the atomic form factor:

$$F(x, Z) = 4\pi \int_0^\infty \rho(r) \frac{\sin 2\pi r x}{2\pi r x} r^2 dr \quad [9]$$

with  $\rho(r)$  the radial atomic electron density function and  $r$  the distance from the nucleus.

For plane-polarized radiation, the differential cross-section is given by

$$\frac{d\sigma_r}{d\Omega} = r_0^2 (1 - \sin^2 \theta \cos^2 \alpha |F(x, Z)|^2) \quad [10]$$

Rayleigh scattering occurs mostly at low energies and for high- $Z$  material (see Figure 1), whereas Compton scattering shows a complementary behavior.

## X-Ray Sources

X-rays are emitted on bombardment of matter by sufficiently energetic electrons, protons, or heavier ions and by irradiation of matter with sufficiently energetic electromagnetic radiation (X- or  $\gamma$ -rays). The former type of radiation is termed *bremsstrahlung*, from the German *bremsen* (to brake) and *strahlung* (radiation). Certain radioactive materials are another type of (mostly) weaker X-ray sources. Very intense beams of X-rays are generated in accelerators where elementary particles (electrons, positrons) at relativistic speeds are forced to alter their direction of motion by means of suitable magnetic fields in a storage ring. This type of radiation is also called synchrotron radiation.

## X-Ray Tubes

The most widely applied source of X-rays is the X-ray tube. Electrons, generated by heating tungsten

or other refractive metal filament, are accelerated toward a cooled anode by means of a 10–100 kV potential difference. The anode is usually made from a high- $Z$  metal (Cu, Mo, Rh, Ag, etc.) with good thermal conductivity and mechanical strength. Bremsstrahlung is generated as a result of the sudden deceleration of the energetic electrons entering the anode. The energy distribution of the continuous (white) X-ray spectrum generated is shown in Figure 4. Its overall intensity is proportional to the tube current  $i$ , the square of the accelerating voltage  $V$ , and the atomic number  $Z$  of the anode.

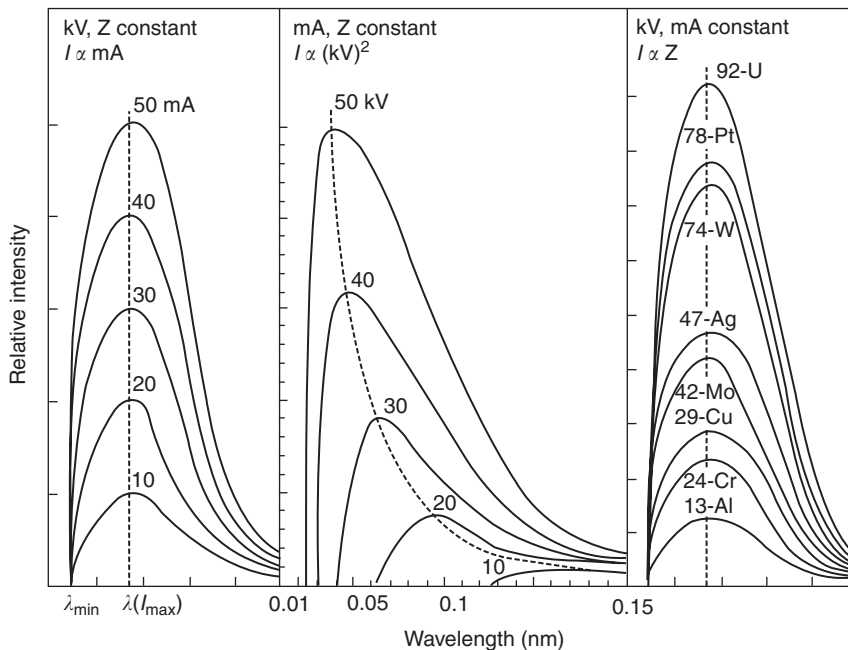
The continuum extends from a shortest wavelength  $\lambda_{\min}$  up to longer wavelengths. The maximum of the distribution is situated near  $\lambda(I_{\max}) \sim 1.5\lambda_{\min}$ . By changing the tube voltage  $V$ ,  $\lambda_{\min}$ , and

$$I(\lambda) \propto \frac{iZ}{\lambda^2} \left( \frac{1}{\lambda_{\min}} - \frac{1}{\lambda} \right) \quad [11]$$

$$\lambda_{\min}(\text{nm}) = \frac{1.2396}{V(\text{kV})} \quad [12]$$

$\lambda(I_{\max})$  can be varied. The high  $\lambda$  cut-off of the spectrum is determined by absorption in the exit window, usually a 5  $\mu\text{m}$  beryllium foil.

Since penetrating electrons in the anode of energy  $E$  higher than the binding energy  $E_0$  also eject core-level electrons from the anode atoms, characteristic lines of the anode material are superimposed on the Bremsstrahlung continuum. The line-to-continuum intensity ratio is a function of atomic number of the



**Figure 4** Effect of the X-ray tube current  $i$  (mA), potential  $V$  (kV), and target atomic number  $Z$  on the output spectrum of an X-ray tube. The dotted line indicates the maximum of the distribution.



anode material and the overvoltage  $E/E_0$ . The relative importance of the continuum and line contributions to the overall X-ray intensity delivered is strongly dependent on the anode materials.

### Types of X-Ray Tubes

There are several types of X-ray tubes but since Coolidge introduced the (side-window) X-ray tube in 1913, a variety of modifications to his basic design have been proposed. The ideal X-ray tube produces sufficient photon flux over a wide spectral range with a stability better than 0.1%. Via a switchable tube potential (usually in the 10–100 kV range), it is possible to optimize the shape of the tube spectrum for the intended application.

The X-ray tube is characterized by the anode element, its input power (0.2–5 kW for conventional high-power tubes), the anode–cathode voltage (10–100 kV), the tube current (0.1–100 mA), and the exit window thickness and composition. Other important features are the angle of incidence of the electrons with the anode and the take-off angle of the X-rays. The radiation output power is of the order of 1% of the input power and (in the side-window tube (SWT)) roughly varies with  $I/t^2$ ,  $t$  being the anode-window distance. In the SWT, the reduction of this distance is limited by the increase in electron bombardment and heating of the window with decreasing  $t$ . An arrangement where this is avoided (permitting smaller  $t$  values) is the end-window tube (EWT). Whereas in the SWT the window and anode are at ground potential, in the EWT the filament and window are grounded while the anode is at a positive potential. In the EWT, electrons are directed to the anode by means of suitable electron optics. Because of the smaller anode-window distances and the fact that the tube voltage is applied between anode and window, EWT tubes voltages are limited to  $\leq 60$  kV.

A configuration in which  $t$  is reduced to zero is called the target transmission tube (TTT). The anode material is evaporated directly on the beryllium window. The anode is irradiated from the back while a significant part of the bremsstrahlung escapes from the front and exits the tube. This geometry requires a much lower input power (200 W compared with 3 kW for the conventional EWT) and, as a result, does not normally require cooling with water.

In order to remove the limitation in the input power caused by excessive heating/melting of the anode in the SWT geometry, the rotating anode tube was developed. By employing a cylindrical anode head that rotates rapidly through the electron beam, an output power gain by a factor of 5–10 relative to

conventional designs can be achieved. The anode head must be perfectly aligned so that the spot on the anode surface where the X-rays originate from is a fixed point in space. The anode head itself is equipped with an integrated high-power cooling system.

The TTT and rotating anode tubes are relatively recent developments for which the application field is still growing; the most popular type of X-ray tube is still the SWT.

Nowadays small, compact tubes are also popular for various applications. For many of these, the brilliance is the vital parameter rather than the total delivered X-ray intensity. Thus, low-power tubes are interesting tube configurations for  $\mu$ -XRF. For a side looking, microfocus tube with 50–100  $\mu$ m spot size, the brilliance is comparable to the high power tubes, although the total power does not exceed more than a few watts.

### Radioactive Sources

X-rays of discrete energy are also generated as a result of various types of radioactive decay processes. In such processes harder radiation ( $\gamma$ -rays) is usually produced as well. Table 1 gives an overview of radioisotopes most commonly used as sources of X-rays, together with some of their characteristics.

Radioactive sources have the advantage that they produce monochromatic radiation, are very compact and easy to use, and straightforwardly produced. As a consequence, they are frequently used in portable (and/or hand-held) XRF spectrometers which are used, for example, for *in situ* analysis of soil, waste materials, etc. In other applications, the absorption or

**Table 1** Radioisotope sources and some of their properties

Radioisotope	Half-life (years)	Photons per disintegration	X-ray and $\gamma$ -ray lines (keV)
$^{55}\text{Fe}$	2.7	0.28	5.9, 6.5 (Mn K)
$^{238}\text{Pu}$	88	0.13	13–20 (U L)
$^{244}\text{Cm}$	17.8	0.08	14–21 (Pu L)
$^{109}\text{Cd}$	1.3	1.07	22, 25 (Ag K)
		0.04	88
$^{125}\text{I}$	0.16	0.07	35
		1.38	41–48 (Te K)
$^{241}\text{Am}$	433	0.36	59.5
$^{153}\text{Gd}$	0.66	1.10	41–48 (Eu K)
		0.30	97
		0.23	103
$^{57}\text{Co}$	0.74	0.86	122
		0.11	136
$^{133}\text{Ba}$	10.3	0.34	81
		0.07	276
		0.18	303
		0.62	356
		0.09	384
$^{137}\text{Cs}$	30.2	0.85	662

scattering of the primary X-rays is employed, e.g., to estimate the mass or mean atomic number of the material irradiated. A disadvantage of radioactive sources is the limited lifetime and relatively low flux of primary radiation.

### Synchrotron Sources

Synchrotron radiation produced in storage rings are now employed on a routine basis as radiation sources for X-rays for various X-ray applications exploiting the very high intensity and the possibility to obtain monochromatic radiation on a microscopic spot.

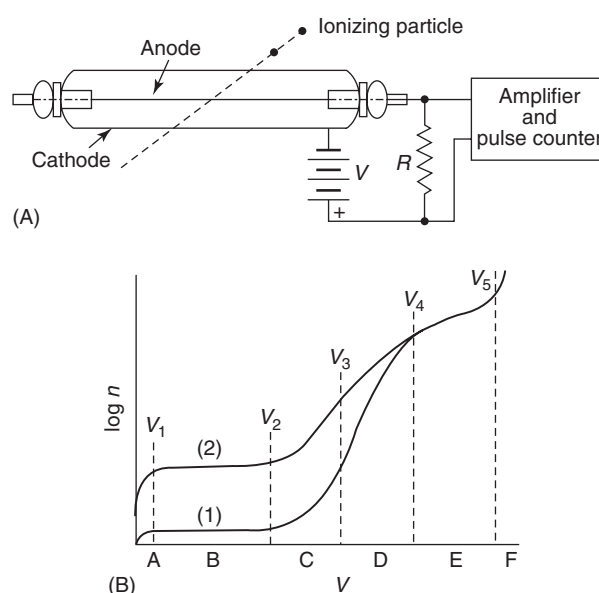
## Detection of X-Rays

X-rays cause ionization in gases, liquids, and solids, thereby altering the electrical properties of these materials. Gas ionization is the basis of gas-filled detectors such as ionization chambers, Geiger-Müller counters and proportional counters are used in conjunction with X-ray dispersion. The ionization of silicon atoms in the solid state forms the basis of the lithium-drifted (Si(Li)) semiconductor detector, making energy-dispersive detection of X-rays possible.

The fact that photographic materials are sensitive to X-rays allows for the photographic recording of X-ray intensities and spectra and is the basis of film-badge dosimetry. The fact that X-rays can induce visible and/or ultraviolet luminescence in certain materials is used in scintillation counters and fluorescent screens.

### Gas-Filled Detectors: Ionization Chambers, Proportional Counters, and Geiger Counters

The principle behind these radiation detectors is the photoionization of gas molecules and/or atoms followed by the separation and collection of the resulting ions by means of an electric field. The differences between the various types can be explained by means of Figure 5, which shows a cylindrical conducting chamber containing an axial anode wire insulated from the chamber walls. The chamber is filled with an inert gas (e.g., argon or helium) at a pressure of 1 atm or less; the anode is at a positive potential  $V$  relative to the chamber walls. When radiation passes through the chamber, ion-electron pairs are formed due to photoionization of the gas atoms/molecules. Even at small potential differences (e.g., 10 V), all electrons produced in this way are immediately accelerated toward the anode, resulting in a current pulse in the external circuit. For a range of potentials (the ionization chamber region, region B in Figure 5) the number of ion pairs  $n$  is independent of the potential  $V$ , although this number is proportional to the energy



**Figure 5** (A) Gas-filled counter with associated electronics. (B) Number of ion pairs  $n$  collected at central electrode as a function of voltage ( $V$ ) for X-ray photons of low (curve 1) and high energy (curve 2). Regions A–F are described in the text.

of the incident photon. Ionization chambers are employed to measure the overall intensity of X-ray radiation and are often employed for normalizing signals collected using more sophisticated detectors.

Above region B, an increase in potential difference also causes an increase in  $n$  as a result of gas multiplication or gas amplification. This phenomenon occurs when the electrons produced by photoionization acquire sufficient energy to produce additional ionizations before they reach the anode wire. As a result,  $n$  increases exponentially with  $V$ , each photoelectron producing an avalanche of secondary electrons. Region C, in which individual electron avalanches are not affected by the presence of others, is called the 'proportional counter' region. Although amplified by a factor  $10^2$ – $10^6$ , the detected signal is still proportional to the photon energy. Proportional counters therefore can be used for energy measurements and to count single events. A maximum count rate of  $\sim 2 \times 10^6$  counts  $s^{-1}$  can be accommodated. Proportional counters are mainly used for detecting soft X-rays ( $\lambda > 0.2$  nm;  $E < 6$  keV). Several types exist, depending on wavelength range of the radiation that needs to be measured. At higher voltages, the number of electron avalanches becomes so large that the remaining positive space charge (formed by the slower moving ions) of one avalanche inhibits the development of the next. As a result, in region D, only a limited proportionality is retained.

Finally, in the Geiger-Müller counter regime (region E), the collected signal becomes independent

of the energy of the original photon. At even higher voltages (region F), the discharge tends to propagate itself indefinitely. Geiger counters are employed for counting (relatively low doses) of radiation ( $<2000$  counts  $s^{-1}$ ); they are usually filled with an argon-ethanol mixture and are operated at 800–1000 V. Alcohol serves to quench discharges in the chamber resulting from the bombardment of the chamber walls by  $Ar^+$  ions.

The energy resolution of the proportional counter is determined by the statistical fluctuations of the number of ion–electron pairs generated by a photon of specific energy. Full-width at half-maximum (FWHM) of the distribution determines the resolution of the detector, which is of the order of 900 eV for a proportional counter. For this reason in X-ray spectrometers this type of detector is always used in conjunction with a dispersive crystal or equivalent.

### Scintillation Counters

This type of detector is only efficient in the hard X-ray range ( $E > 6$  keV,  $\lambda < 0.2$  nm). The scintillation counter consists of a Tl-doped NaI scintillator crystal affixed to the end of a photomultiplier tube. X-ray photons with energy  $E = h\nu$  eject an outer electron from an iodide ion (this process requires  $\sim 30$  eV), thereby transferring most of their energy to the photoelectron. The latter dissipates its energy into the crystal by promoting valence band electrons to an excited state 3 eV above ground level. Upon de-excitation, these electrons emit a 3 eV (410 nm) photon. The intensity of the emitted light pulses is proportional to the number of electrons  $n$  excited by the original X-ray. The NaI (Tl) crystal in which these phenomena occur is aluminized on the outside to reflect the scintillation light toward the photomultiplier tube. In the photomultiplier, the light that strikes the photocathode releases a few photoelectrons. These electrons are then accelerated between the cathode and the first dynode, producing a higher number of electrons as they strike it. In a photomultiplier tube, this process can be repeated up to 10 times until the avalanche of electrons reaches the collector anode, at which the resulting electron current can be measured. Depending on the voltages applied between successive dynodes, an amplification factor of the order of  $10^6$  can be achieved. The amplitude of the final current pulse is proportional to the light intensity entering the tube and thus to the energy of the X-ray photon that originally entered the scintillator crystal.

### Wavelength Dispersion of X-Rays

In view of their complementary sensitivity characteristics, proportional and scintillation counters are

often used in tandem. In such an arrangement, X-ray photons first enter the proportional counter through one window. Those with low energy ( $E < 6$  keV) will mostly be absorbed in the counter gas and detected accordingly; the harder X-rays ( $E > 6$  keV) leave the ionization chamber through a second window to enter the scintillation crystal.

The energy resolution of both detectors is insufficient to allow them to be used directly in X-ray spectrometers. Rather, they are most often used in conjunction with a monochromator so that only photons within a well-defined energy/wavelength band can enter the detector. This monochromatization can be done either by means of dispersive crystals or by means of multilayer systems.

When a parallel monochromatic beam of wavelength  $\lambda$  impinges on a crystal (or multilayer) under an angle  $\theta$ , the photons will be scattered by the atoms in the material. Only in certain directions, in which the path difference between two rays is equal to an integral number of wavelengths ( $n\lambda$ ), constructive interference between the scattered rays occur. The angles  $\theta$  for which this diffraction phenomenon occurs for a given wavelength  $\lambda$  and crystal spacing  $d$  are given by Bragg's law:

$$n\lambda = 2d \sin \theta \quad [13]$$

For  $n=1$ , the path difference between two rays is only one wavelength and the diffraction is said to be of the first order. Second order corresponds to  $n=2$ , etc. By selecting a certain value of  $\theta$  from the diffraction pattern of an incident polychromatic beam, a specific wavelength  $\lambda$  can be selected with a resolution:

$$\frac{d\theta}{d\lambda} = \frac{n}{2d \cos \theta} \quad [14]$$

Depending on the energy/wavelength range the dispersive element operates in, various types of crystals and orientations ( $d$ -spacings) are employed. Table 2 lists a number of currently available natural and synthetic crystals and multilayer systems; the elements whose characteristic lines fall within the operating region of each dispersive element are also indicated. Multilayer systems, with their  $d$ -spacings of the order of 2–12 nm (i.e., much larger than possible in crystals), are used for dispersion of very low energy X-rays ( $0.2 < E < 1.2$  keV;  $1 < \lambda < 6$  nm). In general, for crystal monochromatization, the energy bandwidth of the resulting monochromatic beam is of the order of 1%.

As an alternative to the use of planar crystals and multilayers, in a number of X-ray spectrometers (especially those employing synchrotron radiation or

**Table 2** Overview of dispersive crystals and multilayer systems, their  $2d$ -spacing and the chemical elements whose characteristic lines are situated in the operating wavelength range of each material

<i>Crystal</i>	<i>2d (nm)</i>	<i>Element range</i>
LiF (420)	0.180	Ni–U
LiF (220)	0.285	V–U
LiF (200)	0.402	K–U
Si (111)	0.626	P, S, Cl
Ge (111)	0.653	P, S, Cl
PG <sup>a</sup> (002)	0.671	P, S, Cl
InSb (111)	0.784	Si
PE <sup>b</sup> (002)	0.874	Al–Cl
EddT <sup>c</sup> (020)	0.880	Al–Cl
ADP <sup>d</sup> (101)	1.064	Mg
TIAP <sup>e</sup> (100)	2.575	O–Mg
PX-1 <sup>f</sup>	5.1	O–Mg
PX-2 <sup>f</sup>	12.0	B–C
PX-3 <sup>f</sup>	20.0	B
PX-4 <sup>f</sup>	12.0	C–(N, O)
PbSt <sup>f</sup>	10.0	F, C
OVO 55 <sup>f</sup>	5.5	Mg, Na, F
OVO 100 <sup>f</sup>	10.0	C, O
OVO 160 <sup>f</sup>	16.0	B, C

<sup>a</sup>Pyrolytic graphite.

<sup>b</sup>Pentaerythritol.

<sup>c</sup>Ethylenediamine-*d*-tartrate.

<sup>d</sup>Ammonium dihydrogenphosphate.

<sup>e</sup>Thallium hydrogenphosphate.

<sup>f</sup>Commercial names of multilayer systems.

electron beam excitation) curved crystals are used. These can simultaneously monochromate and focus the X-ray beam.

### Solid-State Detectors

Solid-state energy-dispersive detectors are based on the simple semiconductor diode structure made from, for example, lithium-drifted silicon (Si(Li)) detector or high-purity germanium (HPGe detector). A Si(Li) detector consists of a single-crystalline disk of semiconductor grade silicon consisting of a compensated intrinsic (i-type) region sandwiched between p- and n-type regions, thus forming a p–i–n type diode structure. The central area is made (quasi)intrinsic by diffusing ('drifting') lithium into the p-type silicon to compensate for impurities and dopants present in the crystal. As a result, the Si(Li) crystal features a very low concentration of free charge carriers. The thin p-type layer (typically 0.3  $\mu\text{m}$ ) on the surface of the crystal does not contribute to the detection process and is called the 'dead layer'. Gold film electrodes ( $\sim 20$  nm thick) are evaporated onto the faces of the silicon wafer between which a reverse bias voltage can be applied, creating an active depletion area. Under such

conditions and in the absence of a radiation field the only current flowing between the electrodes is due to thermally generated charge carriers. To minimize this dark current, Si(Li) detectors are operated at liquid nitrogen ( $\text{LN}_2$ ) temperatures (77 K) in a vacuum cryostat at all times. The cooling reduces the noise and ensures optimal energy resolution. It also minimizes diffusion of the highly mobile lithium under the influence of the applied electrical field.

X-ray photons can enter the cryostat and reach the Si(Li) crystal through a thin beryllium window of typically 5–10  $\mu\text{m}$  thickness. The vacuum enclosure protects the detector from surface contamination and moisture condensation (at  $\text{LN}_2$  temperatures), light and scattered electrons (in electron microprobes). For measuring very low energy/long-wavelength X-rays, the beryllium window can be replaced by ultrathin diamond or aluminum coated polyimide windows or (in vacuum instruments) removed completely.

When an X-ray photon enters the active volume of the detector, its energy is transferred to a photoelectron, which in its turn expends its energy by producing electron–hole pairs in the semiconducting silicon crystal; this requires on average 3.76 eV per electron–hole pair. The energy associated with the inner-shell vacancy created by the photoelectric absorption is also absorbed in the crystal, either via the emission by the ionized atom of Auger electrons or low-energy X-rays that are reabsorbed. The total number of electron–hole pairs created is proportional to the energy of the original photons. These free charge carriers are collected by the applied field as a charge pulse that is integrated and amplified to a voltage pulse that can be digitized and stored in a multichannel pulse height analyzer or computer memory, thus collecting the energy distribution of the individually detected photons. Details on the exact functioning of this type of detectors can be found elsewhere. Because of its high-energy resolution (typically  $\sim 150$  eV FWHM for Fe  $K_\alpha$ ), its small size, and high efficiency over a wide spectral region, this type of detector is very useful for the detection of fluorescent X-rays in the 1–40 keV energy range in various types of applications (see below).

### X-Ray Analysis

There are several ways in which the analytical methods that employ X-rays can be categorized. One can make a distinction between techniques that use X-rays to excite sample atoms (as in the case of XPS) and methods that employ the information contained in the energy and number of X-rays emitted by

a material under investigation. In the latter case, the emission of the X-rays can be induced by X-rays themselves (as in the case of XRF and XRD) or by employing beams of charged particles (as in the case of EPXMA or PIXE).

Another subdivision can be made on the basis of the type of information the method provides: structural, elemental, or chemical. Moreover, the information can be derived from a fairly large sample volume, representative of the whole of the material (bulk analysis), or originate from a well-defined (and microscopically small) location on the sample surface (microanalysis). Additionally, the information obtained by an X-ray method may either originate from the topmost layer (analysis of the surface) or emerge from greater depths. For many methods such as XRD, XRF, XPS, AES, the use of synchrotron rings as powerful X-ray sources has prompted the development and maturation of microscopic equivalents.

A short description of each of the most important X-ray techniques will now be given. It is not the intention to provide the reader with a comprehensive overview of all analytical methods that employ X-rays in one form or another, but rather to present a concise discussion of the main points of the most important techniques. Most of these techniques are more exhaustively described in other entries of this work.

### X-Ray Diffraction

Diffraction of X-rays is the basic technique for obtaining information on the atomic structure of crystalline solids and is one of the key standard laboratory techniques. XRD is based on the interference of X-ray waves elastically scattered by a series of atoms orientated along a particular direction in a crystal characterized by a vector  $\mathbf{A}_b$ . The waves scattered by two atoms  $a$  and  $b$  interfere constructively with each other when the path difference PQR is equal to an integer number of wavelengths:  $\text{PQR} = h\lambda$ . This condition is valid for orientations  $\mathbf{K}$  of the scattered waves which satisfy the Laue condition:

$$(\mathbf{K}_l - \mathbf{K}_0) \cdot \mathbf{A}_b = h \quad [15]$$

where  $\mathbf{K}_0$  and  $\mathbf{K}_l$  are the wave vectors of the original and scattered waves. Therefore, by measuring the diffracted intensity as a function of all possible orientations  $\mathbf{K}_l$  relative to  $\mathbf{K}_0$ , data on all possible orientations  $\mathbf{A}_b$  of series of atoms in the crystal can be obtained. Maxima in the diffracted intensity (called reflexes) obtained from a three-dimensional crystal correspond to orientations of  $\mathbf{K}_l$  relative to  $\mathbf{K}_0$  which

fulfill three of the above conditions simultaneously (atoms orientated along directions  $\mathbf{A}_b, \mathbf{A}_k, \mathbf{A}_l$ ). The integer numbers ( $h\ k\ l$ ) are called the Miller indices of the reflex. Since the scattered intensity is a function of the number of electrons around each atomic nucleus, information on which atoms constitute a unit cell of the crystal is also contained in the diffracted intensities. In a single-crystal diffractometer, monochromatic radiation is employed and the crystal and detector system are mounted on goniometer arms, permitting the diffracted intensity along all orientations of  $\mathbf{K}_l$  relative to  $\mathbf{K}_0$  to be accurately measured. More simple setups are appropriate for determining the interplanar distances of polycrystalline or powdered materials (Debye–Scherrer recording).

For laboratory-scale X-ray diffraction, X-ray tubes with copper up to molybdenum and silver anodes are typically employed. By making use of the high-intensity and tunable character of synchrotron radiation, for single-crystal diffraction of more ‘difficult’ systems (very small crystals, weakly scattering systems, radiation-sensitive materials, crystals of very large/complex molecules (proteins, DNA fragments), materials under very high pressure and/or temperature), dramatic improvements relative to the laboratory-based variants have been achieved. Using the related techniques of SAXS and WAXS (small- and wide-angle scattering, respectively), the monitoring of structural changes in materials that are dependent on time, temperature, and/or pressure has also become possible.

### X-Ray Absorption Spectroscopy

As described earlier, in the region 1–40 keV, the photoelectric cross-section  $\tau(E)$  is characterized by discontinuities at the binding energies  $E_b$  of the core level electrons, called absorption edges. Between these jumps, for isolated atoms,  $\log \tau$  versus  $\log E$  curves are nearly straight lines. In condensed matter, however, oscillations on the absorption profiles can be observed close to absorption edges. These features are called the X-ray absorption fine structure (XAFS).

The region from 0 to 40 eV above the actual edge is called the XANES (X-ray absorption near-edge structure) or NEXAFS (near-edge XAFS) region; above 40 eV, the EXAFS (extended XAFS) region commences. Also, pre-edge features may be observed. The fine structure is caused by the interference of the outgoing photoelectric wave front with the waves backscattered from neighboring atoms. For a particular chemical environment of the central atom and for a given photoelectron energy  $E$ , the amplitude of the outgoing photoelectric waves is amplified

(constructive interference) or damped (destructive interference), resulting, respectively, in higher or lower values of  $\tau(E)$  relative to the isolated atom case. From the fine structure, the interatomic distances and coordination numbers around the absorbing atom can be determined. Since the absorption fine structure is caused by the immediate chemical environment of the atomic species involved, the material under study need not show a periodic repetition of this structure in order for the XAFS to be observed. As such, XAS is a very important structural investigation method for studying noncrystalline materials.

The direct method involves measuring the intensity of a highly monochromatic beam of radiation ( $\Delta E/E \leq 10^{-4}$ ) before and after passing through the material under study. From the ratio  $I/I_0$  ( $I_0$  = original intensity,  $I$  = transmitted intensity), the absorption coefficient of the material  $\mu(E)$  at that energy can be calculated (see eqn [1]). The energy of the beam can be continuously varied by changing the Bragg angle  $\theta$  of the double-crystal monochromator. Because of the requirement for such a highly monoenergetic beam, XAS measurements are almost exclusively performed with synchrotron X-ray sources. Since the direct method is confined to samples of limited thickness containing appreciable amounts of the atomic species under study ( $>1\%$ ), for thicker samples (or samples with a lower analyte concentration), instead of measuring the photoelectric absorption directly, the resulting production of fluorescent X-rays is measured. The latter method is known as fluorescent XAS and usually employs solid-state detectors. The two methods are functionally equivalent. Another method involves the measurement of the electron yield. An important variant of the method is SEXAFS (surface EXAFS), in which the photon beam impinges on the sample under grazing incidence, thereby only probing the top few nanometers or less of the sample surface. Applications of XAS are found in the field of metal catalyst research where the chemical environment and oxidation state of the catalytically active metal atoms can be probed, in the geosciences, in electrochemistry, and in materials analysis, where SEXAFS in particular can provide chemical information on the surface of materials as a function of depth.

### X-Ray Photoelectron Spectrometry

Photoelectron spectrometry is a surface analysis technique based on the detection of photoelectrons emitted from a solid sample upon irradiation with a UV (ultraviolet) discharge lamp, an X-ray tube, or a synchrotron source. By means of an electron

analyzer, the energy spectrum of the emitted photoelectrons is recorded. In the XPS spectra, at certain energies  $E_i$ , peaks can be observed, called core electron lines, which correspond to the binding energy levels  $E_{b,i}$  (relative to the Fermi level) of the emitted photoelectrons in the solid under investigation:

$$E_{b,i} = h\nu - E_i - \phi \quad [16]$$

where  $h\nu$  is the energy of the exciting radiation and  $\phi$  the work function of the spectrometer (usually an empirically determined constant). Next to the core electron lines, other bands of lower intensity (close to zero binding energy) are visible which reflect the distribution of electronic states in the valence band of the material. The photoelectrons ejected from the atom in the solid have different kinetic energies  $E_i$  according to the electronic level and type of atom they originate from. XPS therefore permits an elemental analysis of the material under study. All elements except hydrogen can be detected. Furthermore, since the kinetic energy of the photoelectrons also depends on the chemical environment of the ionized atom, the  $E_i$  values shift to lower or higher values as the electronic charge of the atom increases or decreases. This 'chemical shift' allows elemental data about a material to be complemented with chemical information. This phenomenon is also observed in XAS studies.

In laboratory setups, aluminum or magnesium anode X-ray tubes are usually used; the radiation is often monochromatized by means of, e.g., a quartz crystal. In most spectrometers a hemispherical analyzer equipped with electrostatic lenses for collection, acceleration, and focusing of photoelectrons is used; a channeltron or channel plate serves to amplify the primary signal. Although the penetration depth of, e.g., Al  $K_\alpha$  photons is relatively large in solids (typically a few micrometers), the analyzed photoelectrons (having kinetic energies in the 0–1500 eV range) only come from the top 0–10 nm of the material. As a result, XPS is very useful for investigating the extreme surface of materials. XPS detection limits are found at the level of  $10^{12}$ – $10^{13}$  atoms on the surface, corresponding to a relative detection power of 0.01–0.1  $\mu\text{g per g}$  relative to the bulk of the material. XPS experiments must therefore be conducted in an ultrahigh vacuum chamber ( $<10^{-9}$  Torr). A common accessory in the XPS chamber itself is an ion gun for depth profiling purposes. By rastering an area of the surface with a beam of heavy ions (e.g.,  $\text{Ar}^+$ ), thereby gradually removing layer after layer of the surface of the material, XPS spectra of layers of increasing depth with respect to the original sample surface can be recorded. As well as such depth



information, laterally resolved information can also be obtained by using a primary X-ray beam that is focused to a spot 5–20  $\mu\text{m}$  in diameter (small-spot XPS) or by collecting photoelectrons from a microscopically small area on the surface (small-area XPS). In the area of micro-XPS, for soft X-ray excitation, a few types of laboratory-scale instruments are available; for higher energies, the use of synchrotron radiation sources is more appropriate. The tunability, polarization, and brightness of synchrotron light are major advantages in this respect, making it possible, e.g., to optimize the sensitivity for a particular element or to fine-tune the escape depth of the photoelectrons, to perform constant initial and final state experiments, to study the symmetry of electron states at the surface, or to achieve very small spots.

Applications of XPS are found in the broad domain of materials science, involving quantitative surface analysis and stoichiometry determinations, the study of layer growth phenomena, the determination of the oxidation state and the chemical environment of elements, and the investigation of interface structures and compositions.

### X-Ray Fluorescence Analysis

Together with XRD, XRF analysis is one of the most widely applied analytical techniques based on X-rays. The method is aimed at obtaining qualitative and quantitative information on the composition of materials on the basis of the energy and intensity distribution of the characteristic radiation emitted by this material when irradiated with an X-ray beam.

Requiring only relatively simple equipment, XRF spectroscopy is a powerful and convenient nondestructive multielement technique for bulk analysis of a variety of materials that features detection limits in the  $\mu\text{g per kg}$  range for conventional spectrometer types. By employing special excitation conditions (see below), sub- $\mu\text{g per g}$  detection limits can be achieved for most elements heavier than sodium.

Traditionally, for conventional bulk-analysis XRF, a distinction is made between wavelength-dispersive (WD) and energy-dispersive (ED) XRF. Associated primarily with the ED method of detection, several variants of the method exist which either employ alternative excitation forms and/or geometries, use different ways of inducing the emission of characteristic radiation, or constitute a microanalytical version of the bulk technique.

### Wavelength-Dispersive X-Ray Fluorescence

In WD-XRF, an essentially sequential way of collecting the spectrum of characteristic radiation

emerging from the sample is employed. Before reaching the detector, the fluorescent radiation impinges on a suitable monochromator crystal so that only photons of a specific wavelength are diffracted toward the detector. The latter is usually a scintillation counter, optionally preceded by a proportional counter. By rotating the monochromator crystal, the wavelength of the photons entering the detector can be varied and a WD-XRF spectrum recorded. The most important advantage is the wavelength resolution, making accurate and interference-free quantitative analysis possible. Because of the intensity loss that occurs at the monochromatization step, WD-XRF spectrometers are equipped with high-power X-ray tubes. In view of the sequential nature of the data collection process, WD-XRF analysis is well suited for routine analyses (e.g., in industrial quality control processes), whereas ED-XRF is more appropriate for exploratory types of analysis. There also exist parallel variants of WD-XRF spectrometers in which an array of detectors is employed, each measuring the fluorescent intensity at a particular wavelength.

### Energy-Dispersive X-Ray Fluorescence

An ED-XRF spectrometer essentially consists of an X-ray source, a sample holder, and an Si(Li) detector orientated at  $90^\circ$  to the primary radiation. All fluorescent radiation emitted by the sample may be collected in parallel, i.e., without sequential scanning of the energy or wavelength scale. As no significant losses of radiation occur between the sample and the detector, this type of detector is ideally suited for measuring relatively low doses of radiation ( $<10^4 \text{ counts s}^{-1}$ ). An important disadvantage is the relatively low resolving power of the solid-state detectors (typically 120–180 eV at 5.9 keV). This results in considerable peak overlap in ED-XRF spectra, making a reliable quantitative processing of the raw data not straightforward. By employing suitable data evaluation software (fitting of experimental spectra to mathematical models), this disadvantage can almost completely be eliminated.

Because the mechanical requirements of the ED detection geometry are moderate in comparison to the WD equivalent, in the more recently developed analogs or variants of the XRF method solid-state detectors are almost exclusively employed. Such related methods are electron probe X-ray microanalysis (EPXMA) and (microscopic) proton-induced X-ray emission ( $\mu\text{-PIXE}$ ). It should be noted, however, that many EPXMA instruments are also equipped with WD detectors (e.g., for low-Z element analysis). Important variants of the conventional XRF technique that have achieved

considerable success in recent years are TXRF and micro-XRF, using laboratory or synchrotron X-ray sources.

### Electron- and Proton-Induced X-Ray Emission

As described earlier, X-rays can be used to create inner-shell vacancies in sample atoms, but charged particles (in practice, electrons in the 5–100 keV range and protons in the 2–3 MeV range) can also be employed. EPXMA can be performed in most (scanning) electron microscopes equipped with an Si(Li) detector, or in electron microprobe analyzers featuring an Si(Li) detector and one or more WD detection systems. If the electron beam is slowly moved from spot to spot on the sample, while at every point the emerging element-specific radiation is detected, elemental maps can be recorded. Because of the more intense bremsstrahlung, detection limits of EPXMA are in the range 0.1–1%.

If protons are used instead of electrons, a much higher sensitivity can be achieved ( $\mu\text{g per g}$  level for elements lighter than calcium). The PIXE method was originally employed as a bulk analytical method (macro-PIXE) and is performed in a dedicated measuring chamber, installed at a particle accelerator (Van der Graaf or cyclotron); since proton beams can be generated with much higher brilliance (i.e., particle density in the beam) than that of photon beams produced by X-ray tubes, macro-PIXE is ideally suited for analyzing very small amounts of material. More recently, the microscopic variant ( $\mu\text{-PIXE}$ ) has achieved considerable success.  $\mu\text{-PIXE}$  employs a focused proton beam of (sub)-micrometer dimensions for recording maps of elemental species in materials of different kinds.

### Micro X-Ray Fluorescence

$\mu\text{-PIXE}$  and EPXMA have the disadvantage that considerable energy deposition takes place in the sample during irradiation. Another disadvantage is the limited sensitivity of both methods for elements heavier than calcium. Both of these limitations can be overcome by employing a focused beam of X-ray photons instead of charged particles. As X-ray tubes produce a widely divergent cone of primary radiation, in the past such microbeams were not realizable in practice. Recent breakthroughs in X-ray optics and the increasing availability of synchrotron sources provide X-ray beams of sufficient intensity and of micrometer-size diameters can be generated and the number of application fields in which micro-XRF is used is increasing (geosciences, environmental analysis, materials characterization, archeometry).

### Total Reflection of X-Rays

One of the optical properties used in X-ray analysis is total reflection of X-rays at an optically flat surface. This phenomenon occurs when photons impinge on a surface below the critical angle  $\theta_c$  of total reflection. Defined in practical units  $\theta_c$  (in minutes of arc) can be expressed as a function of atomic number  $Z$ :

$$\theta_c = 28.8 \times \frac{1}{E} \sqrt{\frac{Z\rho}{A}} \quad [17]$$

where  $\rho$  is the density ( $\text{g cm}^{-3}$ ) and  $A$  is atomic mass. For X-rays of 10 keV on silicon,  $\theta_c$  is  $\sim 3$  mrad or  $0.2^\circ$ . Total reflection of X-rays is of interest as the X-rays only penetrate very shallowly into the reflector surface. This effect can be employed advantageously in TXRF. The method is able to detect extremely small concentrations of contaminants at the surface of pure materials such as semiconductors. For performing very sensitive multielement analysis of solutions, a microvolume of the solution is pipetted onto an optically flat sample support (e.g., a quartz disk) and allowed to dry. TXRF absolute detection limits are in the 5–50 pg range (without preconcentration) for analysis of trace elements in rain water.

**See also:** **Radiochemical Methods:** Gamma-Ray Spectrometry. **Surface Analysis:** X-Ray Photoelectron Spectroscopy. **X-Ray Absorption and Diffraction:** X-Ray Diffraction – Single Crystal. **X-Ray Fluorescence and Emission:** X-Ray Fluorescence Theory; Wavelength Dispersive X-Ray Fluorescence; Energy Dispersive X-Ray Fluorescence; Total Reflection X-Ray Fluorescence; Particle-Induced X-Ray Emission; Synchrotron X-Ray Fluorescence.

### Further Reading

- Agarwal BK (1991) *X-Ray Spectrometry*. Berlin: Springer.
- Bertin EP (1978) *Introduction to X-Ray Spectrometric Analysis*. New York: Plenum Press.
- Jauch JM and Rohrlich F (1976) *The Theory of Photons and Electrons*. Berlin: Springer.
- Janssens KH, Adams FC, and Rindby A (eds.) (2000) *Microscopic X-Ray Fluorescence Analysis*. Chichester: Wiley.
- Johansson SAE, Campbell J, and Malmqvist KL (1995) *Particle Induced X-Ray Emission Spectrometry (PIXE)*. New York: Wiley.
- Van Grieken RE and Markowicz AM (eds.) (2002) *Handbook of X-Ray Spectrometry*, 2nd edn. New York: Dekker.

## X-Ray Absorption

**A Markowicz and D Wegrzynek**, International Atomic Energy Agency, Seibersdorf, Austria

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

X-rays are characterized by the following features: they are invisible; propagate in straight lines with the velocity of light (in vacuum:  $3 \times 10^8 \text{ m s}^{-1}$ ); unaffected by electric and magnetic fields; differentially absorbed in passing through matter of varying composition, density, or thickness; reflected, diffracted, refracted and polarized; capable of ionizing gases; capable of blackening a photographic plate; capable of producing biological reactions, e.g., to damage or kill living cells and to produce genetic mutations; and able to liberate photoelectrons and recoil electrons. This article describes the interaction of X-rays with matter and provides basic information about the diagnostic use of X-rays, tomography, computed tomography, and nondestructive testing. Qualitative

and quantitative measurements based on the absorption of X-rays are discussed and a brief review of selected applications of X-ray absorption techniques is also given.

### Interaction with Matter

Interactions of photons (including both X-rays and  $\gamma$ -radiation) with matter may be classified according to (a) the kind of target, such as electrons, atoms or nuclei, with which the photons interact, and (b) the type of event, such as absorption, scattering or pair production, that takes place. Possible interactions are summarized in Table 1, where  $\tau$  is the total photoelectric absorption cross-section per atom,  $\sigma_R$  and  $\sigma_C$  are the Rayleigh and Compton collision cross-sections, respectively, and  $K$  is the cross-section for pair production. The sum of all these cross-sections, normalized to a per atom basis, is the probability  $\sigma_{\text{tot}}$  that the incident photon will have an interaction of some kind while passing through a very thin absorber that contains one atom per unit of area normal to the path of the

**Table 1** Classification of photon interactions

Type of interaction	Absorption	Scattering		
		Elastic (coherent)	Inelastic (incoherent)	Multiphoton effects
Interactions with atomic electrons	Photoelectric effect $\tau \propto Z^4$ (low energy) $\tau \propto Z^5$ (high energy)	Rayleigh scattering $\sigma_R \propto Z^2$	Compton scattering $\sigma_C \propto Z$	Two-photon Compton scattering $\propto Z$
Interactions with nucleus or bound nucleons	Nuclear photoelectric effects: reactions ( $\gamma, n$ ), ( $\gamma, p$ ) photofission $\propto Z$ ( $E \geq 10 \text{ MeV}$ )	Nuclear coherent scattering ( $\gamma, \gamma$ ) $\propto Z^2$	Nuclear Compton scattering ( $\gamma, \gamma$ ) $\propto Z$	
Interactions with electrical field surrounding charged particles	(1) Electron–positron pair production in field of nucleus $K \propto Z^2$ ( $E \geq 1.02 \text{ MeV}$ ) (2) Electron–positron pair production in electron field $\propto Z$ ( $E \geq 2.04 \text{ MeV}$ ) (3) Nucleon–antinucleon pair production ( $E \geq 3 \text{ GeV}$ )	Delbrück scattering $\propto Z^4$		
Interactions with mesons	Photomeson production ( $E \geq 150 \text{ MeV}$ )	Coherent resonant scattering ( $\gamma, \gamma$ )		

Reproduced with permission from Hubbell JH (1969) *Photon cross-sections, attenuation coefficients, and energy absorption coefficients from 10 keV to 100 GeV*, NSRDS-NBS 29. Washington, DC: National Bureau of Standards.

**Table 2** The relationship between the attenuation coefficients

Coefficient	Relationship	Units	Units in which thickness is measured
Linear ( $\mu^*$ )		$\text{m}^{-1}$	m
Mass ( $\mu$ )	$\mu^*/\rho$	$\text{m}^2 \text{kg}^{-1}$	$\text{kg m}^{-2}$
Atomic	$(\mu^*/\rho) Z/N_0$	$\text{m}^2 \text{atom}^{-1}$	$\text{atom m}^{-2}$
Electronic	$(\mu^*/\rho) 1/N_0$	$\text{m}^2 \text{electron}^{-1}$	$\text{electron m}^{-2}$

$N_0$  is the number of electrons per kilogram,  $Z$  is the atomic number, and  $\rho$  is the density of the absorber ( $\text{kg m}^{-3}$ ).

incident photon

$$\sigma_{\text{tot}} = \tau + \sigma_{\text{R}} + \sigma_{\text{C}} + K \quad [1]$$

Table 2 shows the relationship between various kinds of attenuation coefficients applied in practical calculations.

It should be emphasized that the absorption coefficient is a much more restricted concept than the attenuation coefficient. Attenuation also includes the purely elastic process in which the photon is merely deflected and does not give up any of its initial energy to the absorber. In a photoelectric interaction, the entire energy of the incident photon is absorbed by an atom of the medium, while in the Compton effect, some energy is absorbed and appears in the medium as the kinetic energy of a Compton recoil electron; the balance of the incident energy is not absorbed and is present as a Compton scattered photon.

The primary effect of the interaction of X-rays with matter includes the production of high-energy electrons, which are the main agents through which all the effects of X-rays arise.

For narrow, parallel, and monochromatic photon beams, the attenuation in homogeneous matter is given by the exponential law

$$I = I_0 e^{-\mu^* t} \quad [2]$$

where  $I$  is the transmitted intensity,  $I_0$  is the incident intensity, and  $t$  is the absorber thickness (in meters).

The linear attenuation coefficient  $\mu^*$  is closely related to the so-called half-value thickness  $D_{1/2}$ , which reduces the intensity of a narrow X-ray beam to one-half of its original value ( $\mu^* = 0.693/D_{1/2}$ ).

If the absorber is a mixture or a chemical compound, its mass attenuation coefficient  $\mu$  can be calculated from the 'mixture rule'

$$\mu = \sum_{i=1}^n W_i \mu_i \quad [3]$$

where  $W_i$  and  $\mu_i$  are the weight fraction and mass attenuation coefficient for the  $i$ th element, respectively, and  $n$  is the total number of the elements in the absorber.

For broad photon beams, the attenuation is no longer exponential and is given by

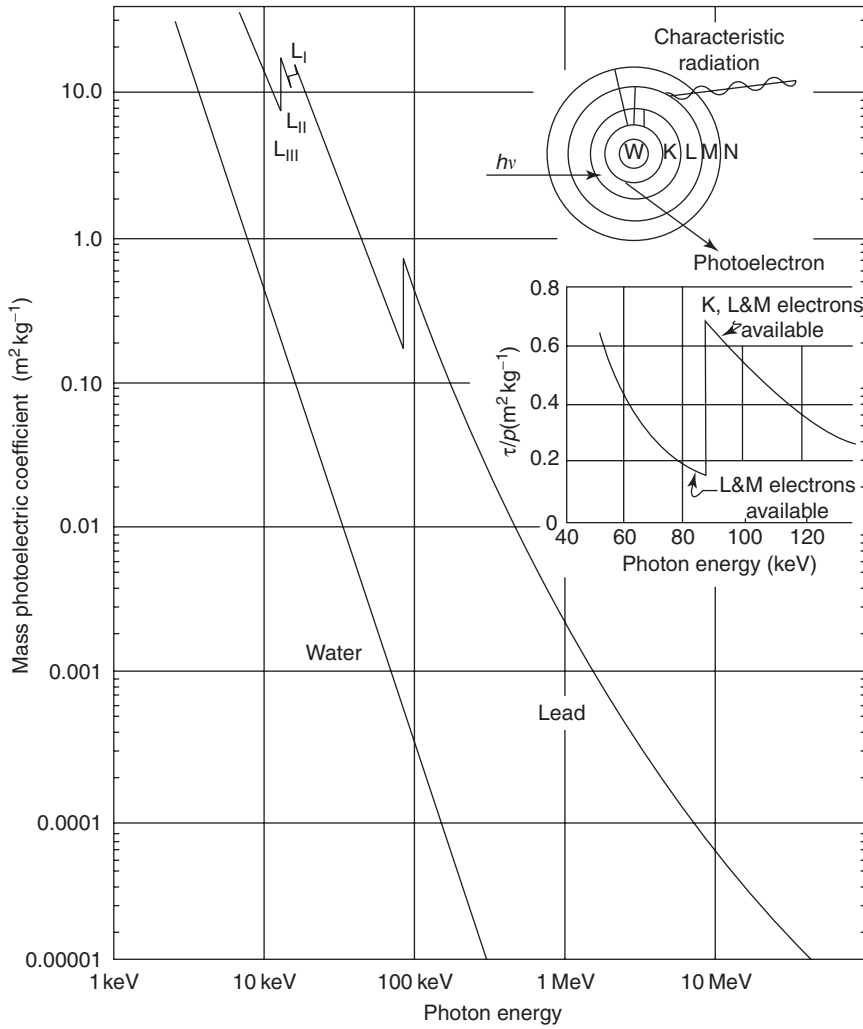
$$I = I_0 e^{-\mu^* t} B(t, h\nu, A, L) \quad [4]$$

where  $B(t, h\nu, A, L)$  is the so-called build-up factor, which depends on the thickness, photon energy, area of the absorber, and the absorber-detector distance. The factor takes into account scattered photons and is defined as the ratio of the observed effect to the effect produced by the primary radiation alone. It is usually obtained experimentally and its value can range from 1 to over 100.

### Photoelectric Absorption

This effect involves the disappearance of a radiation photon and the photoelectric ejection of one bound electron from the absorbing atom, leaving the atom in an excited state. The kinetic energy of the ejected photoelectron is given by the difference between the photon energy  $h\nu$  and the atomic binding energy of the electron  $E_c$  (called ionization energy). Figure 1 shows the mass attenuation coefficient for the photoelectric process as a function of photon energy for water (representing a low atomic number material) and for lead (representing a high- $Z$  material). The discontinuities appear at the energies corresponding to the energies of the absorption edges. The probability of ejection is a maximum if the photon has just enough energy to knock the electron from its shell. The photoelectric cross-section varies with photon energy approximately as  $(h\nu)^{-3}$ . In tissue, the average energy transferred per interaction  $\bar{E}_{\text{tr}}$  is equal to the average energy absorbed per interaction  $\bar{E}_{\text{ab}}$  and nearly equal to the energy of the incident photon  $h\nu$ . In consequence, the transfer ( $\mu_{\text{tr}}^* = \mu^* \bar{E}_{\text{tr}}/h\nu$ ), absorption ( $\mu_{\text{ab}}^* = \mu^* \bar{E}_{\text{ab}}/h\nu$ ), and the attenuation ( $\mu^*$ ) coefficients are nearly equal.

When the apparently sharp X-ray absorption discontinuities are examined at high resolution, they are found to contain a fine structure that extends, in some cases, to about a few hundred electron volts above the absorption edge. The fine structure very close to an absorption edge (less than  $\sim 30$  eV above the edge) is generally referred to as a Kossel structure, while the so-called extended fine structure (Kronig structure) extends to  $\sim 300$  eV above the absorption edge (occasionally to  $\sim 1$  keV above an edge). The modulations of the absorption coefficient in the energy range above an absorption edge can be described theoretically in terms of the electronic parameters and through a Fourier transform relationship are closely related to the radial distribution function around the nucleus of the element of interest.



**Figure 1** Relationship of the mass attenuation coefficient for the photoelectric process versus photon energy. The lower inset shows in greater detail the change in the attenuation coefficient at the K absorption edge for lead. (Reproduced with permission from Johns HE and Cunningham JR (1983) *The Physics of Radiology*. Courtesy of Charles C. Thomas Publisher, Ltd., Springfield, Illinois.)

### Compton Scattering

Compton (incoherent) scattering is the interaction of a photon with a free electron that is considered to be at rest. The process is the most important interaction mechanism in tissue-like materials. From conservation of momentum and energy, the following equations are derived:

$$h\nu' = \frac{h\nu}{1 + \gamma(1 - \cos \theta)} \quad [5]$$

$$T = h\nu - h\nu' = h\nu \frac{\gamma(1 - \cos \theta)}{1 + \gamma(1 - \cos \theta)} \quad [6]$$

$$\tan \phi = \frac{1}{1 + \gamma} \cot(\theta/2) \quad [7]$$

with

$$\gamma = \frac{h\nu}{m_0 c^2} \quad [8]$$

where  $h\nu$  and  $h\nu'$  are the energies of the incident and scattered photon, respectively,  $\theta$  is the angle between the directions of travel of the photon before and following a scattering interaction,  $T$  and  $\phi$  are the kinetic energy and scattering angle of the Compton recoil electron, respectively, and  $m_0$  is the rest mass of the electron.

### Probability of Compton Collision

The differential Klein–Nishina collision cross-section  $d\sigma_{KN}/d\Omega$  is defined as the ratio of the number of photons scattered in a particular direction to the number of incident photons. For unpolarized photons

scattered on unbound, randomly oriented electrons,  $d\sigma_{KN}/d\Omega$  is given by

$$\frac{d\sigma_{KN}}{d\Omega} = \frac{r_0^2}{2}(1 + \cos^2 \theta) F_{KN} \left[ \frac{\text{cm}^2}{\text{electron sr}} \right] \quad [9a]$$

with

$$F_{KN} = \left( \frac{1}{1 + \gamma(1 - \cos \theta)} \right)^2 \times \left( 1 + \frac{\gamma^2(1 - \cos \theta)^2}{(1 + \cos^2 \theta)[1 + \gamma(1 - \cos \theta)]} \right) \quad [9b]$$

where  $r_0$  is the classical radius of the electron equal to  $2.818 \times 10^{-15}$  m. For low energies of incident photons (less than a few tens of kilo electron volts), the angular distribution of Compton-scattered photons is symmetrical  $\sim \theta = 90^\circ$ ; at higher energies, the Compton scattering becomes predominantly in the forward direction.

The fraction of energy transferred to the kinetic energy of a recoil electron in each Compton interaction can be calculated from

$$\frac{d\sigma_{tr}}{d\Omega} = \frac{r_0^2}{2}(1 + \cos^2 \theta) F_{KN} \frac{\gamma(1 - \cos \theta)}{1 + \gamma(1 - \cos \theta)} \quad [10]$$

The energy transfer coefficient  $\sigma_{tr}$  is obtained by integrating eqn [10] over all scattering angles.

The total incoherent (Compton) collision cross-section per atom  $\sigma_C$  involving the so-called electron binding corrections is given by

$$\sigma_C = \frac{r_0^2}{2} \int_0^\pi (1 + \cos^2 \theta) F_{KN} \times S(x, Z) 2\pi \sin \theta d\theta \quad [11]$$

where  $S(x, Z)$  is the incoherent scattering function, which depends upon  $x = \sin(\theta/2)/\lambda$  and atomic number  $Z$ ; the values of  $S(x, Z)$  and  $\sigma_C$  are available in the literature.

In summary, it should be emphasized that (a) the Compton process is almost independent of atomic number; (b) its probability decreases with increase of the photon energy; (c) the fraction of the energy transferred per collision to the kinetic energy of a recoil electron increases with increase in photon energy; and (d) in soft tissue, the Compton process is much more important than other interaction processes for photons in the range 100 keV to 10 MeV.

### Rayleigh Scattering

This is a process by which photons are scattered by bound atomic electrons and in which the atom is neither ionized nor excited. The incident photons are

scattered with unchanged frequency and the scattered waves from electrons within the atom combine with each other. The interference is always constructive, provided the phase change over the diameter of the atom is less than one-half a wavelength, i.e., whenever

$$(4\pi/\lambda)r_a \sin(\theta/2) < 1 \quad [12]$$

where  $r_a$  is the effective radius of the atom. Rayleigh scattering is thus a cooperative phenomenon and is hence called coherent scattering. The process mainly occurs in the forward direction, at low energies, for large  $Z$  values, and is described by the differential cross-section  $d\sigma_R/d\Omega$ , i.e.,

$$\frac{d\sigma_R}{d\Omega} = \frac{r_0^2}{2}(1 + \cos^2 \theta) [F(x, Z)]^2 \quad [13]$$

where  $F(x, Z)$  is the atomic form factor.

The total coherent (Rayleigh) scattering cross-section per atom  $\sigma_R$  can be calculated from

$$\sigma_R = \frac{r_0^2}{2} \int_{-1}^1 (1 + \cos^2 \theta) [F(x, Z)]^2 2\pi d(\cos \theta) \quad [14]$$

The values of  $F(x, Z)$  and  $\sigma_R$  are tabulated in the literature.

### Pair Production

When the energy of the incident photon is greater than 1.02 MeV, the photon may be absorbed through the mechanism of pair production. When the photon passes near the nucleus of the atom and is subjected to the strong field of the nucleus, it may suddenly disappear as a photon and be transformed into a positive and negative electron pair with kinetic energies equal to  $h\nu - 2m_0c^2$ . The positron excites and ionizes atoms as it travels through matter in the same way as an electron until it is finally brought to rest. Then the positron combines with one of the free electrons abundant in matter and is annihilated; two annihilation photons, each of 0.511 MeV, are finally produced. The pair production cross-section increases rapidly with increase in energy above the threshold of 1.022 MeV.

### Total Attenuation Coefficient

As already mentioned, the total attenuation coefficient is the sum of the attenuation coefficients for each interaction process (eqn [1]). The percentage contribution of each process to the total attenuation changes with photon energy and depends on the atomic number of a single element absorber or the so-called effective atomic number in case of complex materials.



## Diagnostic Use of X-Rays

### Attenuation of X-Rays in the Patient

The different tissues in the patient attenuate the X-rays in different ways and therefore the information carried by the transmitted X-rays can be used to visualize the internal structures. From the standpoint of opacity to X-rays, all anatomical structures encountered in clinical radiology can be classified according to their effective atomic numbers into three categories: adipose tissues (fat); soft tissues (with the exclusion of fat) and body fluids; and bone.

The effective atomic number is a convenient parameter for defining the X-rays attenuation properties of a complex medium as a biological tissue, and particularly for the calculation of dose in radiography. The concept of the effective atomic number is based on a proportional relation of the elemental cross-section per atom to  $Z^m$  where  $m$  depends on the process considered. For a specific interaction the atomic cross-section of an element is generally expressed as

$$\sigma = K(E)Z^m \quad [15]$$

The  $m$  values for a tissue vary in the range 4–5 for the photoelectric effect; 2–3 for the coherent (Rayleigh) scattering;  $\sim 1$  for the incoherent (Compton) scattering.  $K(E)$  is a function of photon energy for each partial process. For a specific interaction the mass attenuation coefficient for the individual element is given by

$$\begin{aligned} \frac{\mu^*}{\rho} &= \sigma \left( \frac{N_A}{A} \right) = K(E)Z^m \left( \frac{N_A}{A} \right) \\ &= K(E)N_0Z^{m-1} \end{aligned} \quad [16]$$

$$N_0 = Z \frac{N_A}{A} \quad [17]$$

where  $N_A$  and  $A$  are the Avogadro's constant ( $6.02214199 \times 10^{23} \text{ mol}^{-1}$ ) and the atomic weight, respectively;  $N_0$  is the number of electrons per unit mass of the element of atomic number  $Z$ . Thus for a mixture with an effective atomic number  $\bar{Z}$ , the following relation is fulfilled:

$$\left( \frac{\mu^*}{\rho} \right)_{\text{mix}} = K(E)N_{\text{mix}}\bar{Z}^{m-1} \quad [18]$$

where  $N_{\text{mix}}$  is the number of electrons per unit mass of the mixture. Considering the three main interaction processes which contribute to the total attenuation of X-rays, that is the photoelectric effect, coherent and incoherent scattering, the total mass

attenuation coefficient for a complex medium is given by

$$\begin{aligned} \left( \frac{\mu^*}{\rho} \right)_{\text{mix}} &= N_{\text{mix}} [K_a(E)\bar{Z}_a^{m_a-1} + K_R(E)\bar{Z}_R^{m_R-1} \\ &\quad + K_C(E)\bar{Z}_C^{m_C-1}] \end{aligned} \quad [19]$$

where  $K_a$ ,  $K_R$ , and  $K_C$  are energy-dependent functions for photoelectric effect, coherent, and incoherent scattering, respectively,  $m_a$ ,  $m_R$ , and  $m_C$  are the individual exponents for these processes, and  $Z_a$ ,  $Z_R$ , and  $Z_C$  are the effective partial atomic numbers for the corresponding processes given by

$$\bar{Z}_a = \left[ \sum_i \alpha_i Z_i^{m_a-1} \right]^{1/(m_a-1)} \quad [20a]$$

$$\bar{Z}_R = \left[ \sum_i \alpha_i Z_i^{m_R-1} \right]^{1/(m_R-1)} \quad [20b]$$

$$\bar{Z}_C = \left[ \sum_i \alpha_i Z_i^{m_C-1} \right]^{1/(m_C-1)} \quad [20c]$$

where  $\alpha_i$  is the fractional electron content of the  $i$ th element in the mixture. The total effective atomic number of the mixture  $\bar{Z}$  is obtained from the following relation:

$$\begin{aligned} K_a(E)\bar{Z}_a^{m_a-1} + K_R(E)\bar{Z}_R^{m_R-1} + K_C(E)\bar{Z}_C^{m_C-1} \\ = K_a(E)\bar{Z}_a^{m_a-1} + K_R(E)\bar{Z}_R^{m_R-1} + K_C(E)\bar{Z}_C^{m_C-1} \end{aligned} \quad [21]$$

The effective atomic numbers for bone, muscle, and water are shown in Table 3. In the soft X-ray energy region,  $E < 150 \text{ keV}$ , the total effective atomic number practically does not depend on the photon energy.

### Radiological Image

Figure 2 explains how the primary radiological image is produced. Since the eye is insensitive to X-rays, this image is converted to a visible image by a fluorescent screen, image intensifier, or film (see Figure 3). The conversion of X-ray energy to light is based on the fluorescence and phosphorescence of certain crystals, such as  $\text{CaWO}_4$ ,  $\text{CsI:Na}$ ,  $\text{BaSrSO}_4$ ,  $\text{GdO}_2\text{S:Tb}$ , etc. In computed radiography the image receptor is a storage screen, coated with europium-activated barium halide crystals. Radiation increases the energy of some of the electrons in the crystal lattice, which are held in 'traps'. This creates a latent image that can be read out later in a readout unit by scanning the storage screen line-by-line with a laser beam. The resulting luminescence is detected by

**Table 3** Total and partial effective atomic numbers for biological materials

Energy (keV)	Bone				Muscle				Water			
	Photo, $\bar{Z}_a$	Coherent, $\bar{Z}_R$	Incoherent, $\bar{Z}_C$	Total, $\bar{Z}$	Photo, $\bar{Z}_a$	Coherent, $\bar{Z}_R$	Incoherent, $\bar{Z}_C$	Total, $\bar{Z}$	Photo, $\bar{Z}_a$	Coherent, $\bar{Z}_R$	Incoherent, $\bar{Z}_C$	Total, $\bar{Z}$
10	13.79	11.19	7.95	13.77	7.67	6.62	5.14	7.65	7.60	6.67	5.09	7.58
15	13.98	11.63	8.77	13.95	7.70	6.61	5.57	7.66	7.60	6.64	5.54	7.57
20	14.09	11.78	9.12	14.05	7.71	6.69	5.85	7.66	7.60	6.72	5.86	7.56
30	14.23	11.81	9.43	14.15	7.72	6.77	6.08	7.65	7.60	6.80	6.11	7.54
40	14.33	11.85	9.66	14.21	7.73	6.79	6.13	7.64	7.60	6.82	6.16	7.53
50	14.39	11.93	9.94	14.23	7.74	6.81	6.17	7.63	7.60	6.84	6.20	7.52
60	14.44	12.00	10.10	14.24	7.74	6.83	6.15	7.63	7.60	6.86	6.17	7.52
80	14.51	12.08	10.43	14.24	7.75	6.86	6.25	7.61	7.60	6.89	6.28	7.51
100	14.56	12.12	10.61	14.20	7.76	6.87	6.30	7.59	7.59	6.90	6.33	7.49
150	14.64	12.15	10.90	14.07	7.77	6.89	6.31	7.56	7.59	6.91	6.33	7.25

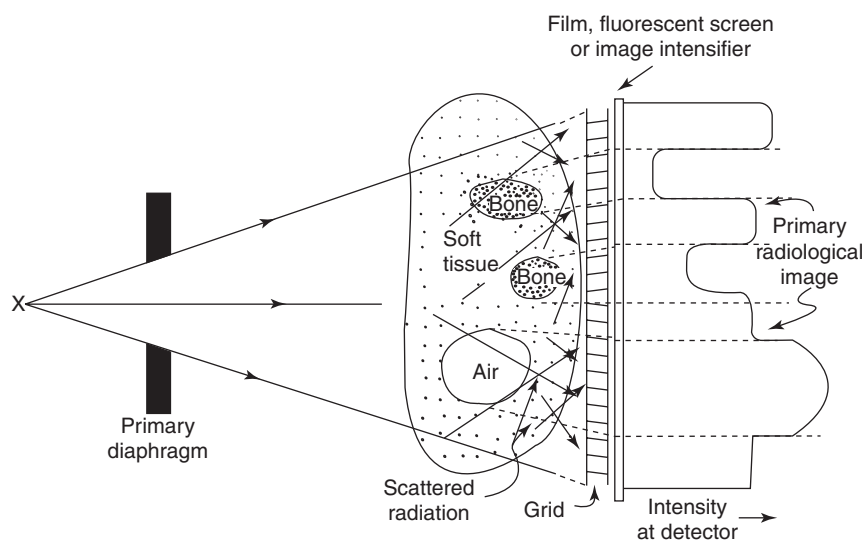
Reprinted with permission from Koc N and Özyol H (2000)  $\bar{Z}$ -dependence of partial and total photon interactions in some biological samples. *Radiation Physics and Chemistry* 59: 339–345; © Elsevier.

a photomultiplier, amplified electronically, and converted into digital value.

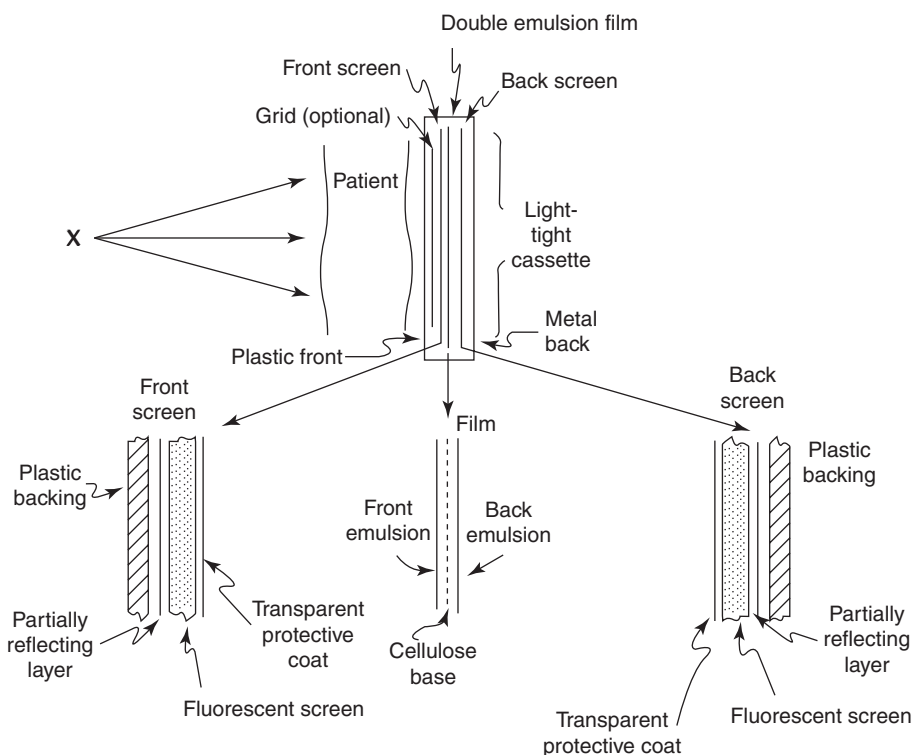
The new imaging systems use a direct digital radiography method based on flat panel detectors. The flat panel detector consists of a scintillation layer mounted on a pixel readout photodiode matrix made of amorphous silicon. Each pixel of the matrix is connected to a row line for driving voltages and to a column line for readout via an active switching element, which may be either a thin-film diode or a thin-film transistor. The electric charges are read out in parallel for one row. The signals are then multiplexed and converted to digital format inside the detector housing. The data are transmitted via a fiber optic link to the acquisition system, where digital processing is performed. A schematic view of the internal construction of the flat panel detector is shown in **Figure 4**. The flat-panel exposures are significantly better than those obtained by conventional screen–film combinations.

To visualize many organs in the body that are characterized by the same attenuation of X-rays, it is necessary to introduce into the patient a contrast medium that is deposited in the organ of interest and which absorbs X-rays either more or less than the surrounding tissues. The sensitivity to contrast agents such as iodine can be enhanced by two or more orders of magnitude by noting the difference between two exposures made with monochromatic X-ray beams, one beam with energy just above the absorption edge and the other with energy just below it. Such beams can be readily produced using synchrotron radiation. For example taking the difference between the two images enhances the contrast between the blood vessels carrying the iodine and the other tissues of the body. This dual-energy subtraction technique may make it possible to image the coronary arteries in less time and with a much smaller iodine concentration than is needed for conventional angiography. The dual-energy X-ray absorptiometry (DEXA) is widely used for measuring bone mineral density in osteoporosis diagnosis. DEXA uses two X-ray beams of different energy to estimate the bone density based on the differences in the absorption of X-rays in bone and soft tissue. This fast method uses very low doses of radiation and can measure as little as 2% of bone loss per year.

Certain radiographic examinations require the visualization of soft, sometimes calcified, tissue lesions that are surrounded by soft tissue. A typical example of such examination is mammography, where the breast is examined radiographically mainly to reveal neoplastic and benign lesions. Since the radiographic examinations require high soft-tissue contrast and high absorption of X-rays in calcium, low-energy X-ray beams are used, e.g., produced by synchrotron facilities (digital mammography).



**Figure 2** Production of primary radiological image. (Reproduced with permission from Johns HE and Cunningham JR (1983) *The Physics of Radiology*. Courtesy of Charles C. Thomas Publisher, Ltd., Springfield, Illinois.)

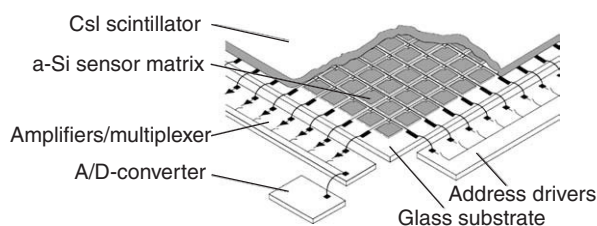


**Figure 3** Diagram to illustrate how films and fluorescent screens are used in diagnostic radiology. (Reproduced with permission from Johns HE and Cunningham JR (1983) *The Physics of Radiology*. Courtesy of Charles C. Thomas Publisher, Ltd., Springfield, Illinois.)

## Tomography, Nondestructive Testing, and Detection of Illicit Materials

In conventional tomography (from the Greek *tomos* meaning section), the confusing shadows cast by the overlying and underlying structures are blurred out, whilst those of the structures of interest are left

sharply defined and therefore visible. To achieve this selective blurring it is necessary to have a controlled, accurate, relative movement of the X-ray tube, film, and patient during the exposure. The simplest tomographic movement is that due to Twyman and is what is known as a linear tomographic movement. The X-ray tube is moving along a line and the film in



**Figure 4** Construction of the flat-panel detector. The X-ray sensitive layer consists of cesium iodide (CsI). The amorphous silicon sensor array comprises  $3000 \times 3000$  pixels. (Reproduced with permission from Hamers S and Freyschmidt J (1998) Digital radiography with an electronic flat-panel detector: First clinical experience in skeletal diagnostics. *Medicamundi* 42: 2–6; © Medicamundi.)

the opposite direction along a parallel straight line. The tomographic apparatus is constructed so that this exact relative movement of film and X-ray tube occurs. The shadows of all objects in the plane of cut (also called the in-focus plane) parallel to the plane of movement of the film will be free from tomographic blurring, while objects outside the plane will have blurred shadows. The tomographic apparatus is designed in such a way that it is possible to place the plane of cut at any desired level in the patient.

More effective blurring out of the shadows of overlying and underlying objects and a possibility of obtaining thinner cuts than those provided by linear tomographic movement can be achieved by using more complex movements, such as circular, elliptical, or hypocycloidal.

In computed tomography (CT), the information contained in the absorption measurements is fully utilized and the sensitivity is improved as compared to the conventional tomography. In CT imaging, the cross-sectional distribution of the linear attenuation coefficient within investigated object is reconstructed mathematically from its projections obtained by X-ray attenuation measurements. This is illustrated in **Figure 5**. From a mathematical standpoint, the solution to the problem of how to reconstruct a function from its projections dates back to 1917; however, the first tomographic scanner was constructed by Hounsfield in 1971, acknowledged by a Nobel prize in 1979, together with Allan Cormack who independently discovered some algorithms used in CT image reconstruction.

There are several types or generations of X-ray tomographic scanners operating in different acquisition geometries:

1. First generation parallel-beam geometry scanners: the measurements of X-ray transmission are obtained using a single highly collimated pencil beam and detector. The beam is translated in a linear motion across the patient to obtain a projection profile. The

source and detector are then rotated around the patient by  $\sim 1^\circ$ , and another projection profile is obtained. This translate–rotate scanning motion is repeated until the source and detector have been rotated by  $180^\circ$ . The highly collimated beam provides excellent rejection of radiation scattered in the patient; however, the complex scanning motion results in long ( $\sim 5$  min) scan time. This geometry was used by Hounsfield in his original experiments, but is not used in modern scanners.

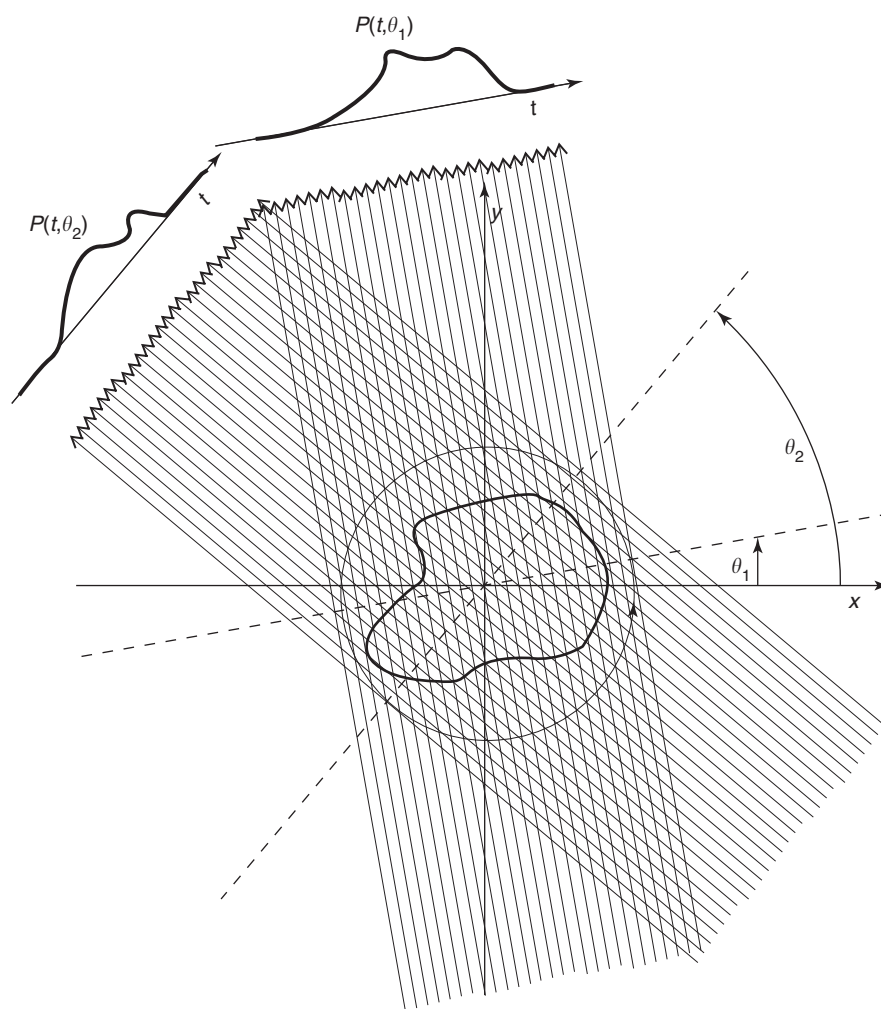
2. The second generation scanners: a narrow fan beam of X-rays and linear array of detectors are used. A translate–rotate scanning motion is still employed; however, a large rotation increment is used, which results in shorter scan times.

3. The third generation scanners: a fan beam of X-rays is rotated  $360^\circ$  around the patient. A curved detector array consisting of several hundred of independent detectors is mechanically coupled opposite to the X-ray source, and both rotate together. No translation motion is used. The fan beam must be wide enough to completely contain the patient. The projection data for a single image are acquired in as little as 1 s.

4. The fourth generation scanner: the fan beam X-ray source rotates around the patient, while the detector array remains stationary. The detector array consists of several hundred up to several thousand independent detectors mounted on a circle that completely surrounds the patient. Scan times are similar to those of third generation scanners. The detectors are no longer coupled to the X-ray source and hence cannot make use of focusing to reject scattered radiation. However, detectors are calibrated twice during each rotation of X-ray source, providing a self-calibrating system.

5. Cine CT scanners: the X-ray source is an integral part of the system design. The detector array remains stationary, while a high-energy electron beam is electronically swept along a semicircular tungsten strip anode. X-rays are produced at the point where the electron beam hits the anode, resulting in a source of X-rays that rotates about the patient with no moving parts. Projection data can be acquired in  $\sim 50$  ms, which is fast enough to image the beating heart without significant motion artifact.

6. Helical (spiral) CT scanners: they allow for fast multiple scans for obtaining three-dimensional images. This is achieved by acquiring projections while the patient is translated through the geometry in a smooth continuous motion rather than stopping for each image. Projection data for multiple images covering a volume of the patient can be acquired in a single breath hold at rates of approximately one slice per second.



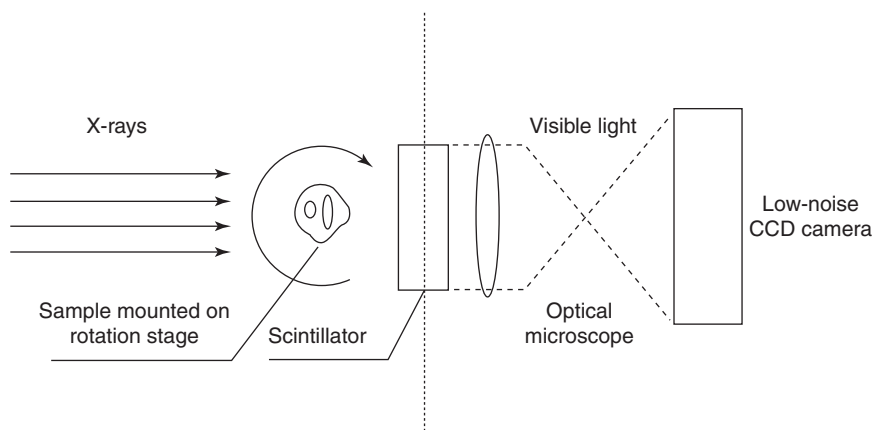
**Figure 5** Projections,  $P(t, \theta)$ , obtained from X-ray attenuation measurements are used in computed tomography to reconstruct the cross-sectional distribution of the linear attenuation coefficient within investigated object.

The thickness of the selected section of a tissue examined by a CT device is usually between 2 mm and 1 cm and varies depending on the requirements of each clinical study. To enhance the contrast of CT various commercially available oral or intravenous contrast media can also be applied.

Recently, computed microtomography has been developed using either conventional X-ray tubes or a synchrotron source. The latter gives superior spatial resolution and shorter data acquisition times because of its high brilliance and continuous X-ray spectrum. The micro tomographic imaging at synchrotron sources can be combined with other techniques such as X-ray absorption near-edge spectrometry, X-ray fluorescence, X-ray diffraction to obtain high resolution three-dimensional images of the object. A typical synchrotron based micro tomographic imaging setup is shown in **Figure 6**.

Industrial radiography (or microradiography) is one of the best known and most common practical

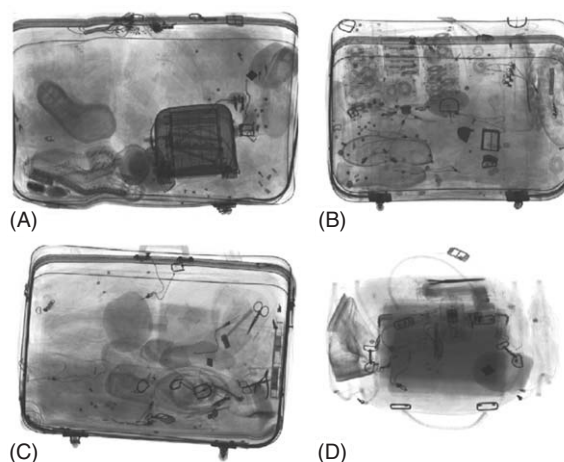
applications of X-rays. The principles of the technique are the same as for medical diagnostic radiography and are based on the direct consequences of differential absorption of X-rays by matter of varying homogeneity, thickness, and density. Photographic film or digital flat-panel detector applied for the detection translates the amount of radiation that penetrates a sample into a two-dimensional picture of the interior of the opaque material. It is worth mentioning that in radiography other kinds of radiation can also be used, such as  $\gamma$ -rays,  $\beta$ -particles,  $\alpha$ -particles, and neutrons. Radiography gives information about the entire sample body, showing diverse phases, segregations, inclusions, and internal voids. It can be used in the inspection of process equipment for gross conditions, such as the position of a frozen valve, for determining the extent of corrosion or scale build-up inside piping or metal containers, and for detection and location of flaws in metal welding and the casting of metal parts. In the last application radiography can



**Figure 6** Schematic view of a typical synchrotron-based computed microtomography setup.

be used as a tool for the development of proper welding or casting techniques or in the control of weld or casting quality. Radiography can also be performed for dynamically changing systems. This is achieved by high-speed digital X-ray radiography systems based on a microcolumnar CsI(Tl) scintillator optically coupled to a charge coupled device. Such systems can take radiographic images with a resolution of  $\sim 250 \times 250$  and a speed up to  $10^5$  frames per second.

Recently, the X-rays have widely been used for detecting explosives and illicit materials to prevent terrorism and smuggling. The scanners based on X-ray attenuation are the most commonly used means to inspect luggage at the airports. They provide information on the object density and effective atomic number. Several types of scanners are in use. Among them the most common are the conventional transmission X-ray scanners. The scanner generates a fan-shaped X-ray beam whose attenuation in the object is measured by line array of detectors. From the absorption data a high quality image is produced. However, such scanners cannot distinguish between thin sheet of a strong absorber and thick slab of a weak absorber. In **Figure 7** some example images obtained with a conventional X-ray scanner are shown. Another type is a dual-energy X-ray scanner in which the X-rays of energy above 100 keV and  $\sim 80$  keV are used. The absorption of the lower energy X-rays depends mainly on the effective atomic number as well as on the thickness of the material. The absorption of the higher energy beam depends mainly on the density of the material. By combining both images a distinction between organic (low  $Z$ ) and inorganic/metal (high  $Z$ ) objects can be made. For detecting explosives scatter imaging scanners are employed. Image data are collected from backscattered X-rays. The backscattered image is sensitive for low- $Z$  high-density materials such as plastic explosives. **Figure 8** shows a relation between the effective atomic number and the



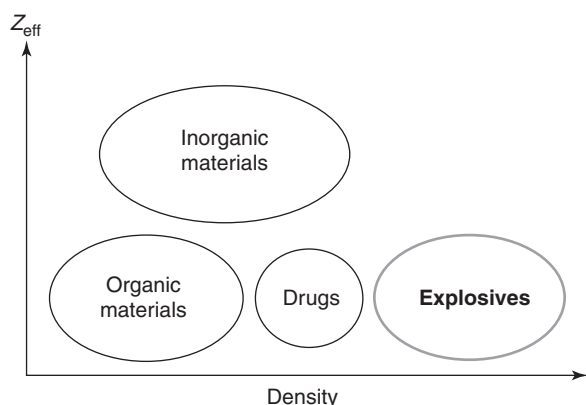
**Figure 7** Some example images of the conventional X-ray scans of luggage at airports. (Reprinted with permission from Singh S and Singh M (2003) Review. Explosives detection systems (EDS) for aviation security. *Signal Processing* 83: 31–55; © Elsevier.)

density for explosives and several other materials. The most advanced systems are the three-dimensional imaging scanners based on CT principle.

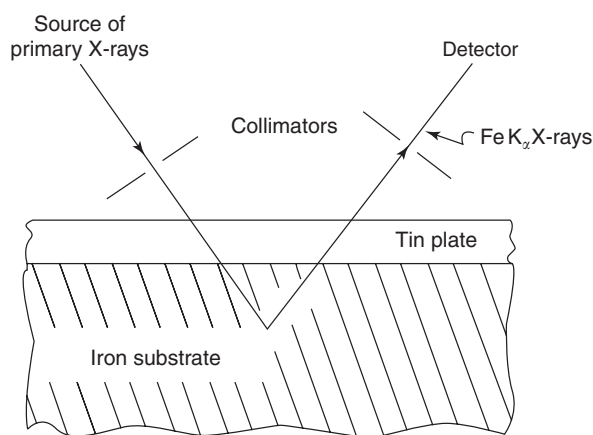
Another example of the practical application of X-ray absorption is the determination of porosity, density, and thickness, including also the control of thin coatings. **Figure 9** shows the principle of a coating thickness measurement. In this case the thickness of tin plate upon steel is not measured by the attenuation of the incident polychromatic beam, but this beam excites the characteristic X-ray lines of the substrate after the beam has gone through film, and finally the attenuation of the iron characteristic X-rays leaving the substrate is used to determine the thickness of the tin film.

The absorption of X-rays is also used in food industry to decrease the population of or prevent the





**Figure 8** Relationship between the effective atomic number and density for different materials. (Reprinted with permission from Singh S and Singh M (2003) Review. Explosives detection systems (EDS) for aviation security. *Signal Processing* 83: 31–55; © Elsevier.)



**Figure 9** Determination of the thickness of tin plate on steel substrate.

growth of undesirable biological organisms in food. This is achieved by depositing the X-ray energy in the irradiated product. The X-rays are generated by using accelerators in which kinetic energy of electrons is converted into well penetrating X-rays (bremsstrahlung). X-rays provide a better dose deposition uniformity in the products with larger areal densities as compared to a direct electron beam.

## Qualitative and Quantitative Measurements

Immediately after discovering X-rays, Röntgen laid the foundation for X-ray absorptiometry with polychromatic beams; this can be compared to colorimetry or photometry with white light. As all atoms absorb X-rays, the method is not specific for any element or group of elements. Polychromatic X-ray beams are

strong and complex and can be used with relatively simple equipment for instantaneous intensity measurements. Through the determination of the energies for which the minimum intensities of transmitted X-rays are observed, it is possible to identify the absorption edges of the elements present. To use absorptiometry with polychromatic beams for the quantitative characterization of materials it is necessary to recourse to the so-called effective wavelength of a polychromatic beam, a concept involving some difficulties.

In X-ray absorptiometry with monochromatic beams, the attenuation at various wavelengths can be measured on both sides of an absorption edge. The results obtained enable (1) the location of the absorption edge and thus the identification of the element, and (2) the calculation of the height of the 'jump' and thus the determination of the concentration of the element of interest. The absorption edge method has been used, for example, for the determination of bromine in liquid samples.

## X-Ray Preferential Absorption Analysis

The method is based on the measurement of the transmitted X-ray intensities at one or more energies. The sensitivity of the analysis depends on the selective absorption of X-rays by the element of interest compared with the absorption by the sample matrix. Assuming that the sample consists of a matrix and an analyte, the concentration of the wanted element  $W_i$  is given by

$$W_i = \frac{[\ln(I_0/I)/\rho t] - \mu_M}{\mu_i - \mu_M} \quad [22]$$

where  $\mu_i$  and  $\mu_M$  are mass attenuation coefficients for the wanted element and the matrix, respectively. The concentration of the analyte can thus be determined if the mass per unit area ( $\rho t$ ) of the sample is known, and  $\mu_M$  is approximately known or the composition of the matrix is constant. To eliminate  $\rho t$  from eqn [22], it is necessary to measure the transmission of narrow X-ray beams at two different energies. In the dual-energy X-ray (or  $\gamma$ -ray) transmission method, the concentration of the analyte is given by

$$W_i = \frac{-(\mu'_M - R\mu''_M)}{[(\mu'_i - \mu'_M) - R(\mu''_i - \mu''_M)]} \quad [23a]$$

where

$$R = \frac{\ln(I_0/I)'}{\ln(I_0/I)''} = \frac{(\mu'_i - \mu'_M)W_i + \mu'_M}{(\mu''_i - \mu''_M)W_i + \mu''_M} \quad [23b]$$

and prime and double prime refer to the first and second energies, respectively. The analysis is accurate when the ratio  $\mu'_M/\mu''_M$  is constant for all possible compositions of sample matrix, which occurs when the X-ray energies are just above and

below the K shell absorption edge of the element of interest. In general, to perform quantitative analysis of a multielement sample by using the X-ray preferential absorption technique it is necessary to use either the same number of sources as unknowns or one less source than unknowns plus a measurement of the sample density thickness.

A similar dual-measurement approach is used in tomography, where by subtraction of two tomograms obtained for X-rays of energy above and below an elemental absorption edge, a map of concentration of that element is produced. The differential tomography technique is valuable for study of major and minor elements to the 0.1% level.

## Other Applications

Although it is impossible to be fully comprehensive and moreover, some examples of applications of X-ray absorption techniques have already been mentioned to illustrate a specific method described, a few additional examples merit mention. In the field of thickness determination, X-ray absorption techniques are used in a variety of applications: such as study of zinc coating on steel, cladding thickness of nuclear fuel elements, plate metals on plated wires, electrolytic silver, iron–nickel films, chromium plating on molybdenum, titanium on Kovar, wall thickness of aircraft propellers, control of thickness and weight of inorganic coatings on paper. Other examples of application of X-ray absorptiometry include: sulfur content of crude and refined oil, tetraethyl lead content of gasoline, additives in heavy-duty lubricating oils, ash content and quality of coal, heavy-metal content of glass, chlorine and fluorine content of polymers and plastics, bromine content in gaseous brominated derivatives, determination of formula of organic compounds (C, H, O), concentration of fillers and impregnants in wood, paper, cloth, rubber, carbon, etc., mineral content and efficiency of softening of water, soil

composition, absorbed contents of charcoals and silica gels, concentration of reagent solutions containing metallic ions, alloy analysis, and histochemical analysis.

Medical diagnostic X-ray radiography and tomography are extensively used in many areas of medicine as excellent diagnostic tools, while industrial radiography has contributed tremendously to our knowledge of metallurgy, metal fabrication, and the internal structure of various materials.

See also: **X-Ray Absorption and Diffraction: Overview.**

## Further Reading

- Kak AC and Slaney M (1999) *Principles of Computerized Tomographic Imaging*. New York: IEEE Press.
- Kereiakes JG and Rosenstein M (1980) *Handbook of Radiation Doses in Nuclear Medicine and Diagnostic X-ray*. Boca Raton, FL: CRC Press.
- Lee JKT, Sagel SS, and Stanley RJ (1983) *Computed Body Tomography*. New York: Raven Press.
- Liebhaufsky HA, Schweikert EA, and Myers EA (1986) The nature of X-rays, spectrochemical analysis by conventional X-ray methods, and neutron diffraction and absorption. In: Elving PJ, Meehan EJ, and Kolthoff IM (eds.) *Treatise on Analytical Chemistry, Part I, Theory and Practice*, 2nd edn, vol. 8, ch. 13, pp. 209–309. New York: Wiley.
- Molins RA (ed.) (2001) *Food Irradiation Principles and Applications*. New York: Wiley-Interscience.
- Nagarkar VV, Tipnis SV, Gaysinskiy V, Miller SR, and Shestakova I (2004) High-speed digital radiography using structured CsI screens. *Nuclear Instruments and Methods in Physics Research B* 213: 476–480.
- Pettersson H, Allison D, von Schulthess GK, and Smith HJ (eds.) (1998) *The Encyclopaedia of Medical Imaging, vol. 1, Physics, Techniques and Procedures*. Lund: NICER Institute/ISIS Medical Media. Published online (2003) Amersham Health's Medcyclopaedia: <http://www.amershamhealth.com/medcyclopaedia>.
- Van Grieken RE and Markowicz AA (eds.) (2002) *Handbook of X-ray Spectrometry*, 2nd edn. New York: Dekker.

## X-Ray Diffraction – Powder

**S Garcia-Granda and J M Montejo-Bernardo,**  
Universidad de Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The diffraction of X-rays by crystalline materials provides a wide spectrum of tools for the study of solids, ranging from qualitative analysis to the study

of internal defects in the atomic arrangement. The techniques addressed in this article may be described collectively as ‘powder’ methods. These techniques apply to the study of polycrystalline materials whether they are in powder form, as supplied by nature or process, or in compacted form, as pellets or metallic mass.

X-ray powder-diffraction experiments involve the interpretation of the powder pattern considering the position (Bragg angle  $\theta$  or interplanar spacing  $d_{hkl}$ )

below the K shell absorption edge of the element of interest. In general, to perform quantitative analysis of a multielement sample by using the X-ray preferential absorption technique it is necessary to use either the same number of sources as unknowns or one less source than unknowns plus a measurement of the sample density thickness.

A similar dual-measurement approach is used in tomography, where by subtraction of two tomograms obtained for X-rays of energy above and below an elemental absorption edge, a map of concentration of that element is produced. The differential tomography technique is valuable for study of major and minor elements to the 0.1% level.

## Other Applications

Although it is impossible to be fully comprehensive and moreover, some examples of applications of X-ray absorption techniques have already been mentioned to illustrate a specific method described, a few additional examples merit mention. In the field of thickness determination, X-ray absorption techniques are used in a variety of applications: such as study of zinc coating on steel, cladding thickness of nuclear fuel elements, plate metals on plated wires, electrolytic silver, iron–nickel films, chromium plating on molybdenum, titanium on Kovar, wall thickness of aircraft propellers, control of thickness and weight of inorganic coatings on paper. Other examples of application of X-ray absorptiometry include: sulfur content of crude and refined oil, tetraethyl lead content of gasoline, additives in heavy-duty lubricating oils, ash content and quality of coal, heavy-metal content of glass, chlorine and fluorine content of polymers and plastics, bromine content in gaseous brominated derivatives, determination of formula of organic compounds (C, H, O), concentration of fillers and impregnants in wood, paper, cloth, rubber, carbon, etc., mineral content and efficiency of softening of water, soil

composition, absorbed contents of charcoals and silica gels, concentration of reagent solutions containing metallic ions, alloy analysis, and histochemical analysis.

Medical diagnostic X-ray radiography and tomography are extensively used in many areas of medicine as excellent diagnostic tools, while industrial radiography has contributed tremendously to our knowledge of metallurgy, metal fabrication, and the internal structure of various materials.

See also: **X-Ray Absorption and Diffraction: Overview.**

## Further Reading

- Kak AC and Slaney M (1999) *Principles of Computerized Tomographic Imaging*. New York: IEEE Press.
- Kereiakes JG and Rosenstein M (1980) *Handbook of Radiation Doses in Nuclear Medicine and Diagnostic X-ray*. Boca Raton, FL: CRC Press.
- Lee JKT, Sagel SS, and Stanley RJ (1983) *Computed Body Tomography*. New York: Raven Press.
- Liebhaufsky HA, Schweikert EA, and Myers EA (1986) The nature of X-rays, spectrochemical analysis by conventional X-ray methods, and neutron diffraction and absorption. In: Elving PJ, Meehan EJ, and Kolthoff IM (eds.) *Treatise on Analytical Chemistry, Part I, Theory and Practice*, 2nd edn, vol. 8, ch. 13, pp. 209–309. New York: Wiley.
- Molins RA (ed.) (2001) *Food Irradiation Principles and Applications*. New York: Wiley-Interscience.
- Nagarkar VV, Tipnis SV, Gaysinskiy V, Miller SR, and Shestakova I (2004) High-speed digital radiography using structured CsI screens. *Nuclear Instruments and Methods in Physics Research B* 213: 476–480.
- Pettersson H, Allison D, von Schulthess GK, and Smith HJ (eds.) (1998) *The Encyclopaedia of Medical Imaging, vol. 1, Physics, Techniques and Procedures*. Lund: NICER Institute/ISIS Medical Media. Published online (2003) Amersham Health's Medcyclopaedia: <http://www.amershamhealth.com/medcyclopaedia>.
- Van Grieken RE and Markowicz AA (eds.) (2002) *Handbook of X-ray Spectrometry*, 2nd edn. New York: Dekker.

## X-Ray Diffraction – Powder

**S Garcia-Granda and J M Montejo-Bernardo,**  
Universidad de Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

The diffraction of X-rays by crystalline materials provides a wide spectrum of tools for the study of solids, ranging from qualitative analysis to the study

of internal defects in the atomic arrangement. The techniques addressed in this article may be described collectively as ‘powder’ methods. These techniques apply to the study of polycrystalline materials whether they are in powder form, as supplied by nature or process, or in compacted form, as pellets or metallic mass.

X-ray powder-diffraction experiments involve the interpretation of the powder pattern considering the position (Bragg angle  $\theta$  or interplanar spacing  $d_{hkl}$ )

and the intensities of the diffraction maxima, the so-called lines or peaks.

## Basic Theory and Instrumental Design

### Angle-Dispersive Technique

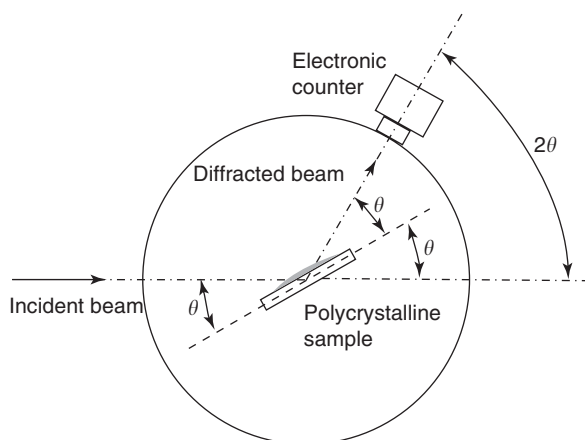
A single crystal, when exposed to monochromatic X-rays, produces diffraction maxima according to the Bragg relationship  $n\lambda = 2d \sin \theta$ . The Bragg angle,  $\theta$ , is the angle between the primary X-ray beam (with  $\lambda$  wavelength) and the family of lattice planes, with interplanar spacing  $d$ ;  $n$  is an integer. In this process, the incident beam, normal to the diffracting plane, and the diffracted beam lie in a plane. If the crystal is rotated about the axis of the X-ray beam, a cone of diffracted rays is generated, the apex of the cone being at the crystal and its solid angle being equal to  $4\theta$ . The same cone of diffracted beams might be achieved by dividing the crystal into small segments and arranging them so that various increments of rotation about the X-ray beam axis are represented. By continuous subdivision of the crystal (within limits) one may arrive at a ‘powdered’ state that statistically furnishes all increments of rotation. Furthermore, the random positioning may be augmented by a rotation of the powdered specimen, usually about an axis normal to the beam. The cones of diffracted X-rays may be recorded by camera techniques using X-ray sensitive films or by diffractometer techniques using a suitable detector and its associated electronic equipment.

A flat photographic film placed perpendicular to the X-ray beam will record intersections of the diffraction cones, producing a pattern of concentric rings, i.e., a powder photograph. A strip of film placed about the sample will record arcs of these concentric rings, i.e., a Debye–Scherrer diagram. The powder diffractometer essentially adds an electronic device for counting the intensity on the arcs of the diffraction rings.

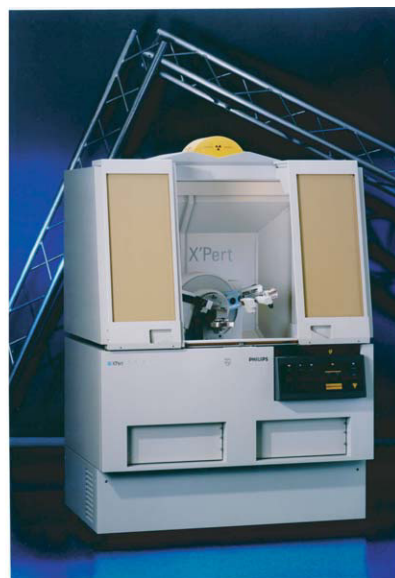
Figure 1 shows, in a simplified scheme, the geometry of a conventional diffractometer while Figure 2 shows a photograph of one of the many commercial diffractometers.

### Energy-Dispersive Technique

X-ray energy-dispersive diffraction (XED), introduced in the late 1960s, uses a primary X-ray beam of polychromatic (‘white’) radiation. In the case of powdered samples the photon energy (or wavelength) spectrum of the X-rays scattered through a fixed optimized angle is recorded using a semiconductor detector connected to a multichannel pulse-height analyzer. In XED powder work the incident- and



**Figure 1** Schematic representation of the geometry of a conventional powder diffractometer.



**Figure 2** Overall view and goniometer detail of a commercial powder diffractometer. (Courtesy of PANalytical.)

scattered-beam directions are determined by slits. The Bragg equation may be written in this case as  $2d \sin \theta_0 = \lambda = hc/E$  where  $E$  is the photon energy

associated with the Bragg reflection,  $h$  is Planck's constant, and  $c$  the speed of light.

## Sample Introduction Techniques

The shape of the average specimen is such that only rarely can it be directly placed into the diffractometer for analysis, so that special care has to be taken to achieve the random distribution of the crystallites to have meaningful peak intensities. The problems arising in sample preparation are: (1) Particle size. The powdered sample must consist of particles smaller than 5  $\mu\text{m}$ . Collection and separation of particles of these dimensions can be effected by sieving, sedimentation, or elutriation. (2) Surface flatness. Special precautions are needed to make the surface smooth and flat, with its plane including the diffractometer axis. If required, common binders, for instance, collodion, paraffin wax, or silicone grease, are applicable. (3) Preferred orientation. When it is necessary to ensure that the particles do not show preferred orientation, mix crushed glass or other amorphous medium with the powder or coat the plane surface of the sample carrier with a film of adhesive that dries at a moderate rate and then dust a layer of powder on the adhesive, poly(vinyl chloride), after it has become tacky.

The method of mounting the sample is of major importance. The 'standard' specimen holders furnished by the manufacturers are rectangular metal or plastic plates of about  $\sim 2$  mm thickness containing a rectangular window to hold the powdered sample. Specimen holders of this kind can be designed for loading samples from the front, edge, or back. Much accumulated experience tends to favor edge or back loading when preferred orientation needs to be minimized.

## Analytical Methodology

### Qualitative Analysis

Every crystalline powder produces a characteristic diffraction pattern. This is the basis of qualitative analysis by powder diffraction. Identification is usually accomplished by systematic comparison of an unknown pattern with a catalog of standard data such as the *Powder Diffraction File* published by the International Centre for Diffraction Data (ICDD). Specialized techniques have been designed for certain situations.

Diffraction patterns of mixtures consist of the superimposed patterns of the individual components. Therefore, powder diffraction is useful in analyzing mixtures as well as pure materials. As the number of

components increases, interpretation becomes more complicated because of the occurrence of peak superposition and success in interpretation depends on the amount of prior information available about the mixture.

As is true with all analytical procedures, X-ray powder diffraction is most powerful when used in conjunction with other techniques, such as emission spectroscopy, X-ray fluorescence spectroscopy, and chemical analysis. An important advantage of X-ray diffraction over these and other techniques is that the results obtained are in terms of the materials as they occur in the sample, not in terms of the elements or ions present. It is often the only satisfactory method of distinguishing among polymorphs or detecting a compound in the presence of others containing the same elements.

Furthermore, the sample is not consumed in X-ray diffraction analysis, so that it may be saved or subsequently analyzed by other techniques. The term 'nondestructive', frequently applied to X-ray analysis, is not always appropriate here because crushing, grinding, or other manipulations are frequently required to prepare the sample for powder diffraction.

### The X-Ray Powder Data File

The powder diffraction file (PDF) is a collection of single-phase X-ray powder diffraction patterns in the form of tables of the interplanar spacings ( $d$ ) and relative intensities ( $I/I'$ ) characteristic of each compound. Other information such as Miller indices, cell constants, and physical properties are included, when available, along with the references of the information source. The PDF has been used for more than six decades and the PDF-4 2003 release contains  $\sim 300\,000$  experimental and calculated patterns. The distribution and maintenance of the PDF are carried out by the ICDD, at Swarthmore, Pennsylvania, USA.

Each pattern is uniquely identified with a PDF number located on the upper left hand portion of the card. The set number is indicated by a two-digit field located to the left of the hyphen, and the digits located at the right identify the particular pattern. For instance, the pattern number 4-587 is the 587th pattern in set 4. The set number can also be used as an indication of the age of the set by adding 1950 to the set number. For the example given, the year of publication is  $(1950 + 4) = 1954$ .

Data on the card image contain the actual  $d$  (interplanar spacing)/ $I$  (relative intensity expressed as percentage of the strongest line) values and, where indexation has been done, the appropriate Miller indices. Additional data such as the formula and name

and mineral name of the compound, along with the experimental procedure and crystallographic and physical information, is also present.

### Search Procedures

Identification of samples can be made using the PDF or similar compilations. A search procedure commonly utilized is the Hanawalt method. The search manual is organized into 'groups' and 'subgroups'. Two lines from the table of '*d*' values and intensities serve to locate the 'entry' of the pattern in the manual. The *d* value of the strongest line of the pattern determines the group into which the entry falls, while the *d* value of the second strongest line determines the subgroup, that is, the location within the group. The entry also includes the next six strongest lines of the pattern in order of decreasing intensities.

Thus, to identify an unknown, we choose the first and second strongest lines of its diffraction pattern as the line-pair with which to look for a matching entry in the search manual. If the other six lines of such a qualifying entry also match lines in the unknown pattern, the identification is probably correct. The final step in the procedure is to confirm the identification by going to the PDF to check the complete pattern. In the more general case in which the pattern is that of a mixture of two or more phases, the manual solution can be tedious and unless carried out systematically, individual mistakes are likely. In the 'Hanawalt Work form' we choose from the table of diffraction pairs of lines to enter on the search manual. We then examine the remainder of the eight lines of any qualifying entry to check whether these lines also match lines of the unknown. We begin with the strongest line available and enter the search manual with line-pairs chosen in order of decreasing intensity until a qualifying entry is found. Once a phase is detected all its lines are 'deleted' and we continue the solution of the problem beginning with the strongest of the unaccounted lines. This work progresses until all lines of an unknown sample are satisfactorily matched.

Although the basic procedures for identifying phases by hand are reasonably well established, the inaccuracies in both the measured pattern and reference database generally make human judgment a necessary ingredient and this has prevented computerized algorithms from successfully automating the identification process. Most algorithms provide a restricted ordered list of candidate matches, but the experimenter must make the final selections and do the identification by hand. Recently, algorithms have been developed in an attempt to translate some of the pattern recognition capabilities commonly found in

human beings into a form convenient for automation. Much of the laborious work of manual searching, matching, and phase identification can now be performed routinely by computer.

### Quantitative Analysis

Powder X-ray diffraction has been applied to quantitative phase analysis (hereafter called QPA) for over 80 years. It seems that the first mention of QPA was made by Hull in the paper 'A new method of chemical analysis', published in 1919 in the *Journal of the American Chemical Society*. At this early stage, quantitative X-ray diffraction (QXRD) was based on a comparison of diffraction line intensities for the analyzed sample with a series of mixtures containing known amounts of the analyte.

The first use of an internal standard technique was reported in 1936 by Clark and Reynolds. Studies in the 1930s and the 1940s by Brentano, Shäffer, Taylor, and Brindley provided the basis for the appropriate corrections in the experimental intensities, particularly for absorption, and the microabsorption phenomena. (Grain-related effect occurs in polyphase samples. It occurs when large crystals preferentially interact with the beam, causing an appreciable reduction in the ratio of the diffracted intensities.)

The development and extension of QPA were facilitated by obtaining a large number of standard diffraction patterns of minerals and chemical compounds, thanks to the development of the powder camera by Debye and Scherrer (1916, 1917). Nevertheless, the take-off of phase quantification started in the late 1940s with the design and industrial manufacture of powder diffractometers and the formulation of a theoretical basis for quantitative analysis by Klug and Alexander in 1948.

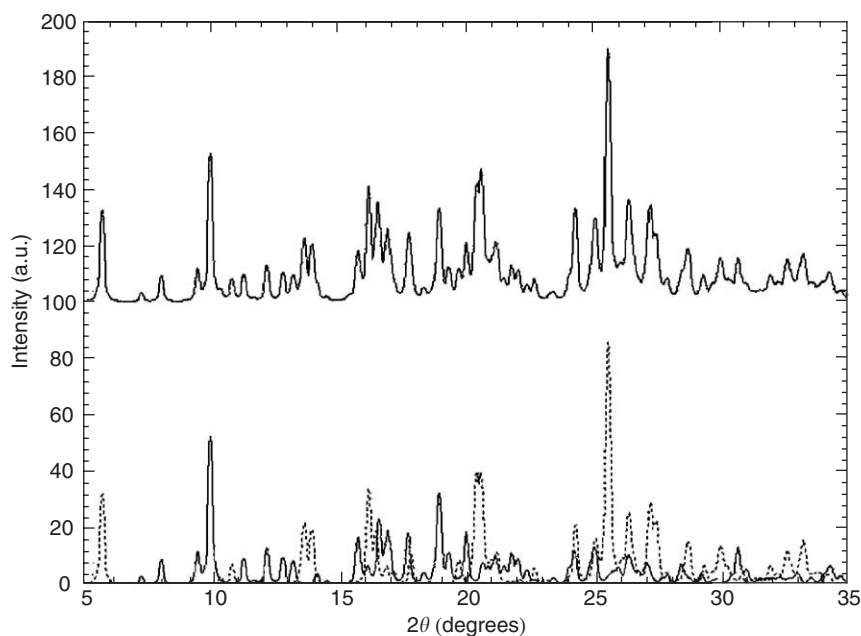
In the 1980s, powder diffraction methods experienced a 'new age', due to the development of high intensity sources, better optics, and computerized diffractometers. This 'revolution' was also caused by the application of the Rietveld method to X-ray diffraction data.

QXRD analysis consists of determining the amount of a compound from measurements of intensity as the X-ray diffraction pattern of a mixture consists of the superposition of the diffraction patterns of its constituent compounds. As an example, the X-ray powder pattern of a binary mixture of crystalline antibiotics, along with the independent patterns of the two components, is shown in **Figure 3**.

### Basic Principles

In QPA, the basic equation that relates the X-ray intensity (of a diffraction peak *i*) of a phase *j* in a





**Figure 3** Powder pattern for a 1:1 binary mixture of two organic pharmaceutical samples showing the contribution of each individual pattern.

sample is

$$I_{ij} = \frac{K'_{ij}x_j}{\rho_j\mu^*} = \frac{K'_{ij}x_j}{\rho_j \sum_{k=1}^n \mu_k^* x_k} \quad [1]$$

where  $K'_{ij}$  is known as the calibration coefficient or calibration constant (and includes composition-independent terms). If the density of phase  $j$  is included in the calibration constant,  $K'_{ij}$ ,

$$I_{ij} = \frac{K_{ij}x_j}{\mu^*} = \frac{K_{ij}x_j}{\sum_{k=1}^n \mu_k^* x_k} \quad [2]$$

where  $x_j$  and  $\rho_j$  are the weight fraction and density of phase  $j$ ,  $\mu^*$  is the mass absorption coefficient of the sample, and  $x_k$  and  $\mu_k^*$  are, respectively, the weight fractions and absorption coefficients of the phases present in the sample.  $K'_{ij}$  is a product of several phase-related, instrument-related, and scattering-angle-related terms.

From determination of the  $x_j$ , the main objective of QPA, it is therefore necessary to:

1. measure the intensity  $I_{ij}$  (calculate the integrated intensity);
2. determine the absorption coefficient  $\mu^*$ ; and
3. determine the calibration constant  $K_{ij}$ .

Calibration is usually performed by preparation and measurement of analytical standards. There are several techniques to treat the dependence of the intensity on the absorption characteristics of the

sample. Absorption correction and calibration are often closely connected, and should be treated together.

### Diffraction–Absorption Technique

In this technique, it is assumed that the mass absorption coefficient  $\mu^*$  has been obtained (is known) by some other technique. Doing measurements on the pure phase  $j$  ( $x_j = 1$  and  $\mu_j^* = \mu^*$ ), we may obtain the calibration constant  $K_{ij}$ , and also  $x_j$ , by eqn [2]:

$$(I_{ij})_0 = K_{ij}/\mu_j^* \rightarrow K_{ij} = (I_{ij})_0\mu_j^* \rightarrow x_j = I_{ij}/(I_{ij})_0 \quad [3]$$

where  $(I_{ij})_0$  is the intensity of peak  $i$  of the pure phase  $j$ . In a multiphase sample, the weight fraction of a determined phase  $j$  is obtained combining eqns [2] and [3]:

$$x_j = \frac{I_{ij}\mu^*}{(I_{ij})_0\mu_j^*} \quad [4]$$

The pure phases must be available for the measurement of their  $(I_{ij})_0$  values.

### Internal-Standard Method

In this method, the sample analyzed is doped with a known amount of reference material, called the internal standard. The ratio of the intensities for the peak  $i$  of phase  $j$  and the peak  $h$  of the standard phase based on eqn [2] is

$$\frac{I_{ij}}{I_{hs}} = \frac{K_{ij}x_j}{K_{hs}x_s} \quad [5]$$

where  $x_j$  and  $x_s$  are the weight fractions in the doped sample. The weight fraction in the original sample ( $x_{j,0}$ ) is given by

$$x_{j,0} = x_j / (1 - x_s) \quad [6]$$

Equation [5] is the basic equation to all internal standard methods of X-ray quantitative analyses. It shows that the curve of the intensity ratio  $I_{ij}/I_{hs}$  obtained from mixtures of components  $j$  and  $s$  plotted against the weight ratio  $x_j/x_s$  is a straight line passing through the origin. The slope of this plot is a unique constant characteristic of components  $j$  and  $s$ , which is independent of the absorption characteristics of the specimen, or even the other components that may be present. An example of linear calibration can be seen in Figure 4.

### Reference Intensity Ratio (RIR)

In general, the RIR ( $I/I_c$ ) is defined as the ratio of the strongest peak of any phase  $j$  to the strongest corundum ( $\alpha$ -Al<sub>2</sub>O<sub>3</sub>) peak (113) in a mixture 1:1 by weight. RIR values are defined for Cu K $\alpha$  radiation and for Bragg–Brentano geometry, with a constant divergence slit. This ratio is ( $x_j = x_s = 0.5$  in eqn [5]):

$$(RIR)_j = I_{ij}/I_{hs} = K_{ij}/K_{hs} \quad [7]$$

Here, the subscript  $s$  refers to corundum, and the subscripts  $i$  and  $h$  refer to strongest peak of the phase  $j$  and corundum (the standard), respectively.

The most usual application of the RIR method is in the internal standard technique. Considering eqns [5]–[7]:

$$x_{j,0} = \frac{x_s}{1 - x_s} (RIR)_j \frac{I_{ij}}{I_{hs}} \quad [8]$$

If the internal standard is not corundum, but any other phase  $t$  (not present originally in the sample),

the relationship becomes

$$x_{j,0} = \frac{x_t}{1 - x_t} \frac{(RIR)_j}{(RIR)_t} \frac{I_{ij}}{I_{ht}} \quad [9]$$

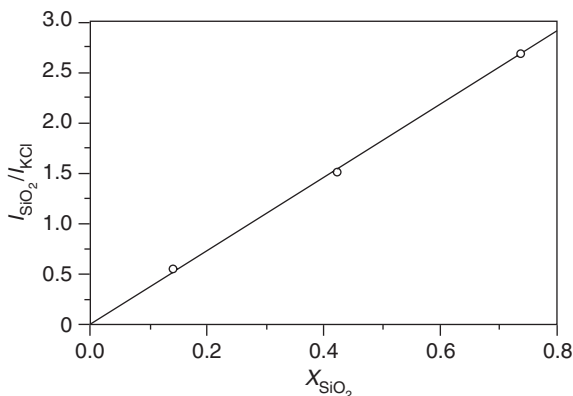
which is the basis of the so-called matrix flushing method (the  $t$  phase is called the flushing agent, and in this context it works as an internal standard) and illustrates the application of the internal standard technique when information on matrix absorption is not available.

The application of this method to three-component samples is shown in Table 1.

### Other Complementary Methods

There are at least two other methods for quantitation of samples that are complementary: the dilution method and the addition method (also called the doping method). In the first, the sample is diluted with any inert material, preferably one with a known mass absorption coefficient. We can know the weight fraction of a phase  $j$  ( $x_j$ ) by measuring the intensity on the peak  $i$  in three samples: the original, the diluted, and the pure phase  $j$ . In this method, the pure phase is used also to obtain the calibration constant  $K_{ij}$ . This technique should be used for analysis of phases with medium or high abundance.

The addition method, however, is more adequate for the analysis of minor phases. This method is similar to the internal standard method but the sample to be analyzed is doped with a known amount of the same phase that is to be determined.



**Figure 4** Calibration line for determination of SiO<sub>2</sub> using KCl as internal standard.  $X_{SiO_2}$  represent the weight fraction of quartz in the sample.

**Table 1** Quantitative analysis using the relative intensity method

	Intensity	RIR <sup>a</sup>	Weight fraction
<b>Sample 1</b>			
ZnO	6266	4.35	0.41
KCl	2987	3.87	0.22
LiF	850	1.32	0.18
Al <sub>2</sub> O <sub>3</sub> (0.1976) <sup>b</sup>	629	1.00	0.18
<b>Sample 2</b>			
ZnO	2999	4.35	0.19
KCl	1733	3.87	0.12
LiF	803	1.32	0.19
Al <sub>2</sub> O <sub>3</sub> (0.5) <sup>b</sup>	1805	1.00	0.50
<b>Sample 3</b>			
ZnO	126	4.35	0.013
TiO <sub>2</sub> (Rutilo)	146	2.62	0.026
CaCO <sub>3</sub>	4994	2.98	0.77
Al <sub>2</sub> O <sub>3</sub> (0.1706) <sup>b</sup>	370	1.00	0.17

<sup>a</sup>Reference intensity ratio.

<sup>b</sup>Weight fraction of corundum added to the sample.

### Full Patterns Methods

All the methods previously cited (and others like the external standard or standardless methods), based on intensity-concentration for single lines, are the traditional techniques used for QPA in the recent past; i.e., the RIR method (using either internal or external standards) is the most widely used. But today, with the availability of digitized diffraction information, full pattern fitting techniques have become more popular. The use of all reflections in a pattern minimizes the effects of preferred orientation, primary extinction, peak overlap, and nonlinear detection systems. The so-called whole-powder-pattern-fitting techniques have been the subject of great development in the last few years, and several procedures using whole-powder-pattern-fitting have been proposed for QPA:

1. using observed digitized whole-pattern traces (information preserved in the PDF database);
2. using whole-powder-pattern decomposition (WPPD); and
3. the Rietveld method.

The WPPD uses observed integrated intensities of standard reference patterns (these patterns have been previously decomposed into individual integrated intensities of Bragg reflections) to fit a determined sample or mixture of phases, refining the integrated intensity parameters, together with unit-cell and profile parameters, by least-squares or iteration, and also adjusting individual scale factors to obtain an overall fit. This procedure does not require knowledge of crystal structure parameters (site occupancy, positional, and thermal parameters) of phases in the mixture, and the weight fractions are obtained as a function of their scale factors and the mass-absorption coefficients of the respective phases (calculated from the chemical formula).

The Rietveld refinement was originally conceived as a method of refining crystal structures using powder neutron diffraction data. Later, it was extended to XRD patterns and phase abundance analysis in the 1980s and the 1990s. The method is based on the assumption that the measured diffraction pattern can be approximated by an analytical expression that contains instrumental and structural parameters (requires knowledge of site occupancy, positional, and thermal parameters), and on the idea that the XRD patterns should be analyzed using each point of the profile as a data point, rather than the integrated intensities of the individual peaks. This application has produced great advances in QPA, especially in accuracy and detection limits of weight fraction of phases (in fact, the detection of an analyte is possible

when its concentration is  $>0.5\%$  w/w, for minor concentrations, and a set of guidelines for the use of Rietveld refinement has been presented by the international Union of Crystallography Commission on Powder Diffraction.

In a sample with  $n$  phases, the weight fraction of phase  $k$  ( $w_k$ ), with a molecular weight  $M_k$  and with  $Z_k$  molecules in a cell of unit volume  $V_k$  is given by

$$W_k = \frac{S_k(ZMV)_k}{\sum_{i=1}^n S_i(ZMV)_i} \quad [10]$$

When an amorphous component is present in the sample, the values of the weight fractions (of crystalline phases) are overestimated to satisfy the normalization condition ( $x_i = 1$ ). So, an internal standard must be added ( $\leq 10\%$  in weight) to determine the absolute weight percentages of amorphous and crystalline phases. The amorphous content  $A(\%)$  is obtained by the small overestimation of the internal standard:

$$A(\%) = \frac{1 - (\omega_s/\varphi_s)}{100 - \omega_s} \times 10^4 \quad [11]$$

where  $\omega_s$  and  $\varphi_s$  are the real and the Rietveld analyzed weight fractions of the internal standard, respectively.

To analyze mixtures either with an amorphous phase or with phases for which structural parameters are unknown, in the last decade, several new techniques have been developed based on known methods.

In the combined Rietveld–RIR method, a known amount of corundum is added to the mixture, and the refined values of the Rietveld phase fractions are converted into weight fractions and rescaled to absolute values with respect to the amount of standard added. Another technique, combining the Rietveld and addition methods, can determine the absolute quantities of the phases, crystalline and amorphous. Here, the added standard is part of the original mixture.

For quantitative analysis of largely amorphous composites, an analysis based on the diffraction–absorption method was also developed. In this approach, the intensity scale factor and the microabsorption parameter are obtained by a calibration procedure based on the diffraction data of different samples of pure crystalline and mixtures. This information is used to calculate the crystalline weight fraction.

For quantifying individual phases in a mixture, a whole-pattern-profile-stripping method, based on the scale of standard profiles to strip them of the experimental pattern, may be used. The method produces a correct quantification for mixtures of components with similar linear absorption coefficients, and the

scale parameter of every phase is taken to represent its volume proportion in the mixture.

An alternative approach to mixtures that contain phases with unknown or imperfectly known crystal structures combines the whole-powder-pattern-decomposition method with the Rietveld analysis. The method includes a modification of the standard Rietveld approach.

Finally, a new method to determining the polymorphic composition in intact compacts, using parallel-beam X-ray powder diffractometry, has also been developed.

Today, QXRD methods are used for organic and inorganic compounds, macromolecules, polymers, drugs, zeolites, cements, catalysts, metals, ceramics, atmospheric aerosols (airborne particulates), fly ash, and minerals. The physical states of the materials can be powders, thin films, polycrystalline, and bulk materials.

### Sources of Error

A preferred orientation that favors the intensity of some line or lines of the diffraction spectrum due to deviations of the ideal random orientation of crystallites is perhaps the most serious problem limiting the quantitative analysis of complex mixtures. Although it is possible to accommodate preferred orientation during data analysis, it is preferable to utilize a procedure that minimizes or eliminates preferred orientation during sample preparation.

Choosing a representative standard is an important part of quantitative analysis. Although there are numerous factors that induce variability in standard data, the most important are composition and order/disorder variations. It is obviously very important to be certain that standards contain 100% crystalline material and no amorphous material.

Numerous authors have emphasized the effects of crystallite size on diffraction data and these effects include primary extinction for well-crystallized phases such as quartz and calcite, particle statistics, microabsorption, and preferred orientation.

The presence of broad, overlapping reflections represents a problem in most quantitative analyses of complex mixtures. The overlap may be of multiple reflections from one phase, making the use of a single reflection difficult, or it may more commonly be the overlap of two important reflections of different phases.

Amorphous materials yield broad, asymmetric peaks. Furthermore, different amorphous materials yield different broad diffraction bands. The difficulty of detecting even significant amounts of amorphous material by the presence of this broad peak and the large errors associated with measuring the integrated

intensity of these broad, asymmetric peaks influence the methods used to quantify amorphous materials.

Reasonable count times can yield detection limits of 0.05 wt% (500  $\mu\text{g}$  per g) for ZnO with an RIR of 4.35. Detection limits <500  $\mu\text{g}$  per g can easily be achieved for materials with RIRs of  $\sim 1.0$  if short scan ranges and long count times are employed. Clearly, it is straightforward to achieve detection limits less than 1 wt% with routine data collection parameters for materials whose reflections are not completely overlapped by stronger adjacent reflections.

### Safety Considerations

The careful, alert, and informed worker need to feel no hesitancy whatsoever in operating X-ray diffraction equipment, but the use of such equipment by careless or uninformed persons is dangerous. The hazards involved come from two sources, the X-rays themselves and the high voltages used to generate them. Workers must never, knowingly or otherwise, expose themselves to direct or secondary radiation. The effects of exposure to X-radiation are cumulative, and may lead to serious and permanent injury. The voltages normally used in X-ray diffraction lie in the range 25–55 kV, and, needless to say, all contact with equipment at such potentials must be avoided.

X-ray installations must be controlled periodically with Geiger counters and personnel must wear badges and have periodical blood analysis.

### Acknowledgments

The authors are indebted to Dr M.A. Salvadó for his contribution, in many senses, to this paper. Thanks to MCyT (MCT-03-BQU-05093) and FICyT (PR-01-GE-4) for financial support.

*See also:* **Air Analysis:** Outdoor Air. **Cement. Ceramics. Fluorescence:** Quantitative Analysis. **Fourier Transform Techniques. Infrared Spectroscopy:** Near-Infrared. **Microscopy Techniques:** X-Ray Microscopy. **Particle Size Analysis. Pharmaceutical Analysis:** Drug Purity Determination. **Qualitative Analysis. Structural Elucidation. Thermal Analysis:** Overview. **X-Ray Absorption and Diffraction:** Overview; X-Ray Absorption; X-Ray Diffraction – Single Crystal. **X-Ray Fluorescence and Emission:** X-Ray Fluorescence Theory; Wavelength Dispersive X-Ray Fluorescence; Energy Dispersive X-Ray Fluorescence; Total Reflection X-Ray Fluorescence; Particle-Induced X-Ray Emission.

### Further Reading

Bish DL and Chipera SJ (1988) Problems and solutions in quantitative analysis of complex mixtures by X-ray powder diffraction. *Advances in X-ray Analysis* 31: 295–308.

- Chung FH (1974) Quantitative interpretation of X-ray diffraction patterns of mixtures. I, II and III. *Journal of Applied Crystallography* 7: 519 and 526; see also 8: 17.
- David WIF (2002) Structure determination from powder diffraction data. In: *International Union of Crystallography Monographs on Crystallography*, No. 13. Oxford: Oxford University Press.
- Klug HP and Alexander LE (1974) *X-ray Diffraction Procedures for Polycrystalline and Amorphous Materials*. New York: Wiley.
- McClune WF (ed.) (1989) *Powder Diffraction File*. Swathmore, PA: International Centre for Diffraction Data.
- Snyder RL and Bish DL (1989) Quantitative analysis. *Reviews in Mineralogy* 20: 101–144.
- Young RA (1995) The Rietveld method. In: *International Union of Crystallography Monographs on Crystallography*, No. 12. Oxford: Oxford University Press.
- Zevin LS and Kimell G (1995) *Quantitative X-ray Diffractometry*. New York: Springer.

## X-Ray Diffraction – Single Crystal

**S Garcia-Granda and J M Montejo-Bernardo,**  
Universidad de Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

To ‘look’ inside a crystal, the wavelength of the incident beam must be of the same order of magnitude as the distance between the units of the crystal ( $\sim 1 \text{ \AA}$ ). X-rays, in the wavelength range  $0.7\text{--}2.0 \text{ \AA}$ , are the ideal radiation and provide a powerful tool for investigating crystal structures and obtaining an image of objects (molecules, ions, etc.) with atomic dimensions.

The electric field of X-rays exert forces on electrons of atoms, and they are accelerated. In terms of a classical model, an accelerated charge radiates, and this radiation has the same frequency as the primary beam, but  $\pi$  out of phase with it, and it is called scattered radiation.

After constructive interference, this radiation produces experimental intensities, and from these experimental magnitudes the structure factor modulus,  $|F(hkl)|$ , is obtained. A Fourier analysis based on structure factors and a calculated set of phases determines with precision the atomic arrangement in space and provides an image of the internal structure of the crystal.

Nowadays, very complex molecules can be investigated; and the structures of proteins, enzymes, nucleic acids, and even viruses are being revealed by new powerful X-ray diffraction techniques.

On the other hand, new models for the electron density, application of topological analysis techniques, and use of synchrotron radiation allow us a quantitative study of several molecular properties in crystalline solids.

### Concepts of Crystallography

A crystal may be defined as a homogeneous solid in which a motif, made of molecules or groups of molecules (or ions, for inorganic crystals) with a definite chemical composition, shows a regular repetition in three-dimensional space and is bounded by plane faces.

Every crystal has a lattice as its geometrical basis. A lattice may be described as a regular, infinite arrangement of points (or motifs) in which every point has the same environment as any other point; the motif is repeated periodically at intervals  $a$ ,  $b$ , and  $c$  along three noncoplanar basic vectors  $a$ ,  $b$ , and  $c$ .

Any point in the lattice may be chosen as an origin, and the position of any other point of the lattice is defined uniquely by the vector:

$$\mathbf{t} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c} \quad [1]$$

where  $u$ ,  $v$ , and  $w$  are positive or negative integers or zero (in some cases, they can have rational values). The three basis vectors ( $a$ ,  $b$ , and  $c$ ) define a parallelepiped, known as the unit cell (the smallest three-dimensional volume element from which the crystal can be constructed). If the cell is primitive, then  $u$ ,  $v$ , and  $w$  in eqn [1] are integers, but if the cell is multiple, then  $u$ ,  $v$ , and  $w$  will have rational values.

The directions of the vectors  $a$ ,  $b$ , and  $c$  are the three axes of the unit cell (indicated by ‘ $a$ ’, ‘ $b$ ’, and ‘ $c$ ’), and the angles between them are indicated by ‘ $\alpha$ ’, ‘ $\beta$ ’, and ‘ $\gamma$ ’ (Figure 1). The values of these six parameters define the type of cell, and so  $a \neq b \neq c$  and  $\alpha \neq \beta \neq \gamma$  is triclinic;  $a \neq b \neq c$  and  $\alpha = \gamma = 90^\circ$ ,  $\beta \neq 90^\circ$  is monoclinic;  $a \neq b \neq c$  and  $\alpha = \beta = \gamma = 90^\circ$  is orthorhombic;  $a = b \neq c$  and  $\alpha = \beta = \gamma = 90^\circ$  is tetragonal;  $a = b = c$  and  $\alpha = \beta = \gamma \neq 90^\circ$  is trigonal;  $a = b = c$  and  $\alpha = \beta = \gamma = 90^\circ$  is cubic;  $a = b \neq c$  and  $\alpha = \beta = 90^\circ$ ,  $\gamma = 120^\circ$  are hexagonal. Considering primitive and

- Chung FH (1974) Quantitative interpretation of X-ray diffraction patterns of mixtures. I, II and III. *Journal of Applied Crystallography* 7: 519 and 526; see also 8: 17.
- David WIF (2002) Structure determination from powder diffraction data. In: *International Union of Crystallography Monographs on Crystallography*, No. 13. Oxford: Oxford University Press.
- Klug HP and Alexander LE (1974) *X-ray Diffraction Procedures for Polycrystalline and Amorphous Materials*. New York: Wiley.
- McClune WF (ed.) (1989) *Powder Diffraction File*. Swathmore, PA: International Centre for Diffraction Data.
- Snyder RL and Bish DL (1989) Quantitative analysis. *Reviews in Mineralogy* 20: 101–144.
- Young RA (1995) The Rietveld method. In: *International Union of Crystallography Monographs on Crystallography*, No. 12. Oxford: Oxford University Press.
- Zevin LS and Kimell G (1995) *Quantitative X-ray Diffractometry*. New York: Springer.

## X-Ray Diffraction – Single Crystal

**S Garcia-Granda and J M Montejo-Bernardo,**  
Universidad de Oviedo, Oviedo, Spain

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

To ‘look’ inside a crystal, the wavelength of the incident beam must be of the same order of magnitude as the distance between the units of the crystal ( $\sim 1 \text{ \AA}$ ). X-rays, in the wavelength range  $0.7\text{--}2.0 \text{ \AA}$ , are the ideal radiation and provide a powerful tool for investigating crystal structures and obtaining an image of objects (molecules, ions, etc.) with atomic dimensions.

The electric field of X-rays exert forces on electrons of atoms, and they are accelerated. In terms of a classical model, an accelerated charge radiates, and this radiation has the same frequency as the primary beam, but  $\pi$  out of phase with it, and it is called scattered radiation.

After constructive interference, this radiation produces experimental intensities, and from these experimental magnitudes the structure factor modulus,  $|F(hkl)|$ , is obtained. A Fourier analysis based on structure factors and a calculated set of phases determines with precision the atomic arrangement in space and provides an image of the internal structure of the crystal.

Nowadays, very complex molecules can be investigated; and the structures of proteins, enzymes, nucleic acids, and even viruses are being revealed by new powerful X-ray diffraction techniques.

On the other hand, new models for the electron density, application of topological analysis techniques, and use of synchrotron radiation allow us a quantitative study of several molecular properties in crystalline solids.

### Concepts of Crystallography

A crystal may be defined as a homogeneous solid in which a motif, made of molecules or groups of molecules (or ions, for inorganic crystals) with a definite chemical composition, shows a regular repetition in three-dimensional space and is bounded by plane faces.

Every crystal has a lattice as its geometrical basis. A lattice may be described as a regular, infinite arrangement of points (or motifs) in which every point has the same environment as any other point; the motif is repeated periodically at intervals  $a$ ,  $b$ , and  $c$  along three noncoplanar basic vectors  $a$ ,  $b$ , and  $c$ .

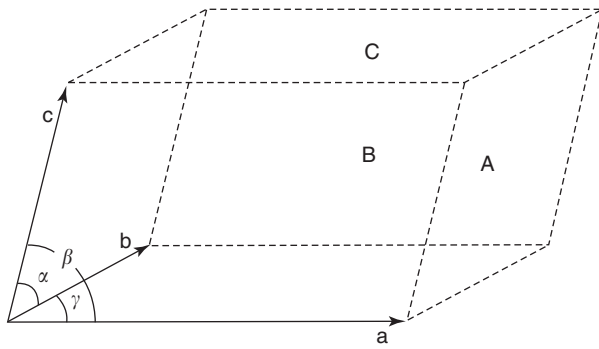
Any point in the lattice may be chosen as an origin, and the position of any other point of the lattice is defined uniquely by the vector:

$$\mathbf{t} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c} \quad [1]$$

where  $u$ ,  $v$ , and  $w$  are positive or negative integers or zero (in some cases, they can have rational values). The three basis vectors ( $a$ ,  $b$ , and  $c$ ) define a parallelepiped, known as the unit cell (the smallest three-dimensional volume element from which the crystal can be constructed). If the cell is primitive, then  $u$ ,  $v$ , and  $w$  in eqn [1] are integers, but if the cell is multiple, then  $u$ ,  $v$ , and  $w$  will have rational values.

The directions of the vectors  $a$ ,  $b$ , and  $c$  are the three axes of the unit cell (indicated by ‘ $a$ ’, ‘ $b$ ’, and ‘ $c$ ’), and the angles between them are indicated by ‘ $\alpha$ ’, ‘ $\beta$ ’, and ‘ $\gamma$ ’ (Figure 1). The values of these six parameters define the type of cell, and so  $a \neq b \neq c$  and  $\alpha \neq \beta \neq \gamma$  is triclinic;  $a \neq b \neq c$  and  $\alpha = \gamma = 90^\circ$ ,  $\beta \neq 90^\circ$  is monoclinic;  $a \neq b \neq c$  and  $\alpha = \beta = \gamma = 90^\circ$  is orthorhombic;  $a = b \neq c$  and  $\alpha = \beta = \gamma = 90^\circ$  is tetragonal;  $a = b = c$  and  $\alpha = \beta = \gamma \neq 90^\circ$  is trigonal;  $a = b = c$  and  $\alpha = \beta = \gamma = 90^\circ$  is cubic;  $a = b \neq c$  and  $\alpha = \beta = 90^\circ$ ,  $\gamma = 120^\circ$  are hexagonal. Considering primitive and





**Figure 1** Representation of a general unit cell, introducing the notation.

nonprimitive cells (centered on a pair of opposite faces, *C* (*A*, *B*); body-centered *I*; and centered on all faces, *F*), we obtain the 14 possible space lattices called Bravais lattices (Auguste Bravais first listed them in 1850): triclinic *P*, monoclinic *P*, *C*, orthorhombic *P*, *C*, *I*, *F*, tetragonal *P*, *I*, cubic *P*, *I*, *F*, hexagonal *P* and trigonal *P* (*R*, described with ‘hexagonal axes’).

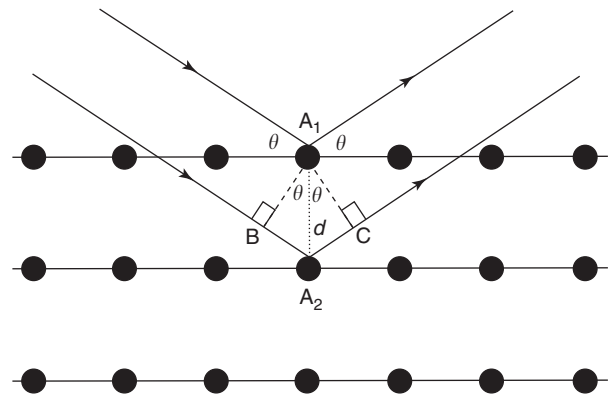
In 1912 M. Von Laue suggested that a crystal could act as a three-dimensional diffraction grating for X-rays with a wavelength comparable with interatomic distances, and in the same year W.L. Bragg described a simple method for obtaining the conditions for diffraction. His interpretation, although an oversimplification, gives a clear representation of the features of the complete process. In the crystal, all scatterers (the motif) forming a three-dimensional array would fall on sets of equispaced planes. Bragg considered that each of these planes would act as a partial reflector of X-rays.

Let *AB* be an incident beam corresponding to a plane wavefront (with  $\lambda$  wavelength), and let  $\theta$  be the angle between this incident beam and the family of planes with interplanar spacing *d*. From **Figure 2** it is clear that a difference exists in the path crossed between the waves scattered at the points *A*<sub>1</sub> and *A*<sub>2</sub> that is equal to  $BA_2 + A_2C = 2d \sin \theta$  ( $BA_2 = A_2C$  and  $\sin \theta = BA_2/d$ ). The two waves will combine with maximum positive interference (will be in phase) if this difference in the path is a multiple of  $\lambda$ :

$$2d \sin \theta = n\lambda \quad [2]$$

Equation [2] is known as the Bragg law and it gives an angle of reflection  $\theta$  in terms of the interplanar spacing, *d*, and the X-ray wavelength  $\lambda$ .

A more realistic treatment, considering the diffraction from a net of scattering centers, gives rise to the conditions known as the Laue equations. These two approaches to X-ray scattering by crystals can be shown to be equivalent.



**Figure 2** Simple scheme of the Bragg reflection process, showing the parallel waves being in phase after reflection and how the path difference equals  $2d \sin \theta$ .

## The Reciprocal Lattice

For each direct lattice, a corresponding reciprocal lattice can be postulated. It has the same symmetry as the direct lattice and may be derived from it both mathematically and graphically. This concept was introduced by P. Ewald in 1921. The parameters are designed as  $a^*$ ,  $b^*$ , and  $c^*$ . By definition they are perpendicular to the planes defined by ‘*bc*’, ‘*ac*’, and ‘*ab*’, respectively, and satisfy the following two conditions:

$$a^* \cdot b = a^* \cdot c = b^* \cdot a = b^* \cdot c = c^* \cdot a = c^* \cdot b = 0$$

$$a^* \cdot a = b^* \cdot b = c^* \cdot c = 1.$$

Also, the volume of the reciprocal lattice is  $V^* = V^{-1}$ . Accordingly,  $a^*$ ,  $b^*$ , and  $c^*$  may be written as

$$a^* = (b \wedge c)/V; \quad b^* = (c \wedge a)/V; \quad c^* = (a \wedge b)/V$$

or, in moduli,

$$a^* = (bc \sin \alpha)/V;$$

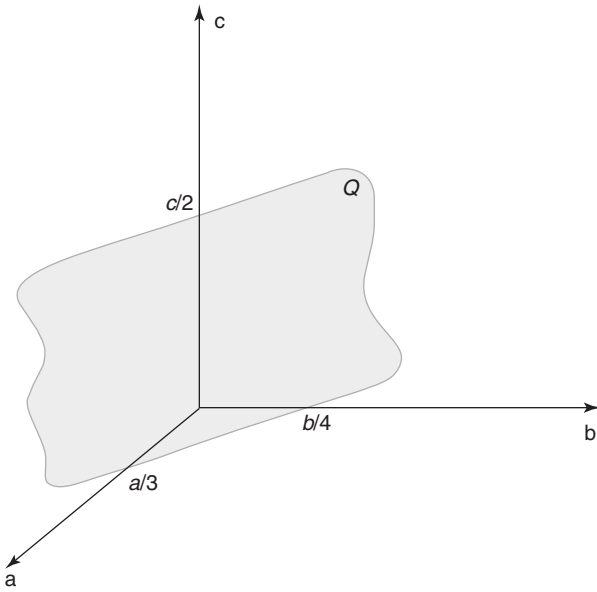
$$b^* = (ca \sin \beta)/V;$$

$$c^* = (ab \sin \gamma)/V.$$

Mathematically, it is more convenient to use the reciprocal lattice since any family of planes in the direct lattice is represented by a point or node in the reciprocal space.

## Miller Indices

The important role that crystallographic planes play in the diffraction process, being necessary to define properly each crystallographic plane, has been shown. Given the crystallographic axes ‘*a*’, ‘*b*’, and ‘*c*’, a plane can be defined by three integers. This notation was introduced first by Miller in 1839, and the integers *h*, *k*, and *l* are called Miller indices. The



**Figure 3** Intercepts of the plane  $Q$ , in three-dimensional space, on the crystallographic axes at lengths  $a/h$ ,  $b/k$ , and  $c/l$ .

plane intersects the three crystallographic axes at lengths  $a/h$ ,  $b/k$ , and  $c/l$ , dividing the crystallographic axes in  $h$ ,  $k$ , and  $l$  parts, respectively. For instance, for the plane  $Q$  in **Figure 3**,  $a/h = a/3$ ,  $b/k = b/4$ , and  $c/l = c/2$ , and the Miller indices for  $Q$  are (342).

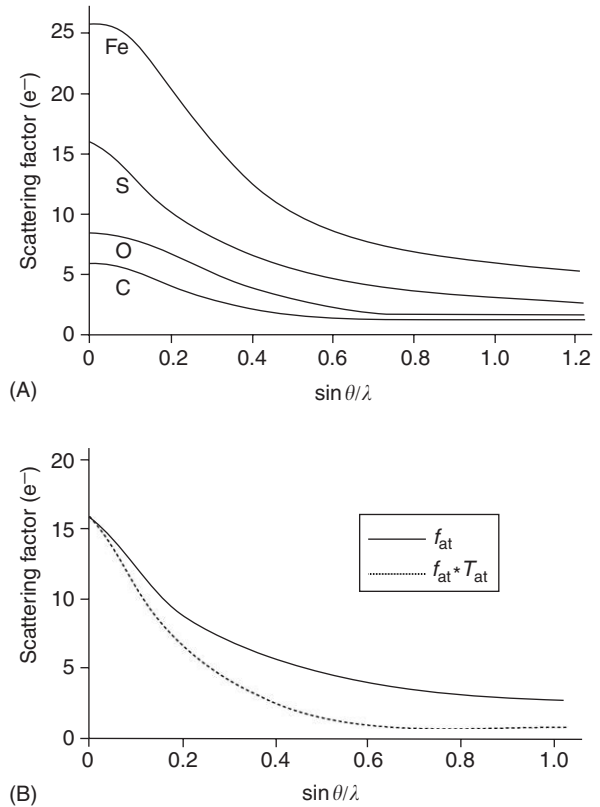
If a plane is parallel to a crystallographic axis, the spacing in that direction will be  $\infty$ , and therefore the Miller index will be 0. The crystallographic planes parallel to one axis are defined by  $(0kl)$ ,  $(b0l)$ , and  $(hk0)$ , respectively. Planes parallel to faces A, B, and C of the unit cell will be  $(h00)$ ,  $(0k0)$ , and  $(00l)$ , respectively.

This notation for the crystallographic planes will be used later in the description of the structure factor.

## Scattering Factor and Structure Factor

The total scattering of the unit cell will be the result of the amplitude of the waves scattered by the electrons of every individual atom. The total electron density of an atom can be regarded as a spherically symmetric region of charge around the nucleus. This distribution is defined by  $\rho(r)$ , the electron density at distance  $r$  from the nucleus.

The Fourier transform of  $\rho(r)$  is called the atomic scattering factor and is denoted by  $f_{\text{at}}$ . It represents the amplitude of scattering expressed in units of the scattering of a single point electron at the



**Figure 4** (A) Atomic scattering factors,  $f_{\text{at}}$ , as a function of  $\sin \theta/\lambda$  for different neutral atoms, showing how  $f_{\text{at}} = Z$  at  $\sin \theta/\lambda = 0.0$ . (B) Influence of the thermal vibration on the atomic scattering factor.

origin:

$$f_{\text{at}} = \int_{\text{all space}} \rho(r) \exp(2\pi i \mathbf{r} \cdot \mathbf{s}) dV \quad [3]$$

where  $\mathbf{s}$  is called the scattering vector, with  $s = 2 \sin \theta/\lambda$ . The electron density,  $\rho(r)$ , is known for practically all neutral atoms and ions from quantum mechanical calculations, via Hartree–Fock methods for lighter atoms and via the Thomas–Fermi approximation for heavier atoms, with high accuracy. The scattering factors are tabulated as a function of  $\sin \theta/\lambda$  in the *International Tables for Crystallography*.

The atomic scattering factor depends on several factors: the nature of the atom, the wavelength of the incident radiation, the scattering direction, and the thermal vibrations of the atom. **Figure 4A** shows how  $f_{\text{at}}$  behaves with  $\sin \theta/\lambda$ , being  $Z$  (the atomic number) at  $\sin \theta/\lambda = 0$ , decreasing at higher values of  $\sin \theta/\lambda$ .

The thermal vibrations produce a more diffused distribution of the electron density than does a stationary model, increasing the atomic effective volume, making the interference within the atom more noticeable, and making consequently the decrease in the atomic scattering factor more rapid (see **Figure 4B**).

Assuming an isotropic model for thermal vibrations, the temperature correction factor for an atom is

$$T_j = \exp[B_j(\sin^2 \theta)/\lambda^2] \quad [4]$$

$B_j$  is usually known as the atomic temperature factor, given by  $B_j = 8\pi^2 U_j^2$  ( $\text{\AA}^2$ ), where  $U_j^2$  is the vibration mean-square amplitude of the atom.

The temperature-corrected atomic scattering factor will be

$$f_j = f_{at,j} \cdot T_j \quad [5]$$

The structure factor,  $F(hkl)$ , which expresses the combined scattering of X-rays for all atoms in the unit cell for the reflection  $(hkl)$

$$F(hkl) = \sum_{j=1}^N f_j \exp\{2\pi i(hx_j + ky_j + lz_j)\} \quad [6]$$

where  $N$  is the number of atoms in the unit cell,  $f_j$  is the atomic scattering factor of the  $j$ th atom (including the temperature factor correction), and  $x_j$ ,  $y_j$ , and  $z_j$  are fractional coordinates ( $x_j = x/a$ ;  $y_j = y/b$ ;  $z_j = z/c$ ).

In the general case, the structure factor has a complex magnitude, and we may write

$$f(hkl) = |F(hkl)|e^{i\Phi(hkl)} \quad [7]$$

where  $|F(hkl)|$  is the structure amplitude and  $\Phi(hkl)$  the phase of the reflection.

In an X-ray diffraction experiment, the intensity of every reflection may be measured by integrating the peak over a suitable angular range around the ideal Bragg angle (a 'real' crystal has defects and mosaic structure, and there exist lattice distortions). This experimental intensity,  $I_{(hkl)}$ , with certain necessary corrections, is directly related to the structure through the structure amplitude,  $|F(hkl)|$ ,

$$I_{(hkl)} = k_1 k_2 I_0 LPTE |F(hkl)|^2 \quad [8]$$

where

$I_0$ : intensity of the incident beam

$k_1$ : collects three universal constants,  $e^4/m^2c^4$

$k_2$ : is a constant for a given diffraction experiment,  $\lambda^3 \Omega/V^2$  ( $\Omega$  is the volume of the crystal and  $V$  is the volume of the unit cell)

$L$ : is the Lorentz factor. Depends on the diffraction technique and is related to the different times-of-reflection opportunity of the different points of the reciprocal lattice. It is a simple function of  $\theta$ , as  $L = (\sin 2\theta)^{-1}$ .

$P$ : is the polarization factor. Depends on the Bragg angle and takes into account the polarization of the X-ray beam after the scattering process. Also a function of  $\theta$ , as  $P = (1 + \cos^2 2\theta)/2$ . Additional

polarization correction may be applied to the monochromator.

$T$ : is the transmission factor. Depends on the capacity of the crystal to absorb the X-rays. Is possible to correct for absorption if the shape of the crystal is known. Since an accurate description of the crystal is very difficult to obtain, the absorption remains one of the most serious (if not the most) sources of error in the experimental determination of intensities.

$E$ : is the extinction coefficient. It depends on the mosaic structure of the crystal and has two components. The secondary extinction (the most important) takes into account that a fraction of the incident beam is reflected by the planes. The primary extinction takes into account the loss of intensity due to multiple reflections from different lattice planes.

## Experimental Structure Factors and Intensities

The structure factor plays a very important role in the determination of the crystal structure because it is the only factor that gives us information about the atomic positions. The main problem in a structure analysis is just the inability to fully determine in an X-ray diffraction experiment the structure factor.

Although the structure factor modulus (structure amplitude) can be obtained from the intensity data (eqn [8]), the corresponding phase cannot be measured directly experimentally. The amplitude values alone cannot be used to determine directly the atomic positions. Incorrect phase values will lead to an incorrect calculated electron density and an erroneous final structure. This is referred to as the phase problem in crystallography and is the central problem in X-ray structure analysis.

Despite this 'limitation', the structure factor amplitude, without knowledge of the phases, has several applications in preliminary structure analysis.

The experimental diffraction pattern has to be described with the proper cell parameters to assign the correct  $hkl$  indices to each observed reflection. This process is called indexing, and essentially it is achieved by solving the following quadratic equation:

$$\begin{aligned} (1/d_{hkl})^2 &= (2 \sin \theta / \lambda)^2 = h^2 a^{*2} + k^2 b^{*2} + l^2 c^{*2} \\ &\quad + 2hka^*b^* \cos \gamma^* + 2klb^*c^* \cos \alpha^* \\ &\quad + 2lhc^*a^* \cos \beta^* \end{aligned} \quad [9]$$

where  $h$ ,  $k$ , and  $l$  must be integers.

**Table 1** Limiting conditions for screw-axes

Screw-axis	Limiting conditions	Orientation	Translation
$2_1$	$h00: h=2n$	$\parallel a$	$a/2$
$2_1$	$0k0: k=2n$	$\parallel b$	$b/2$
$2_1$	$00l: l=2n$	$\parallel c$	$c/2$
$4_1$ or $4_3$	$00l: l=4n$	$\parallel c$	$c/4, 3c/4$

**Table 2** Limiting conditions for glide-planes

Glide-plane	Limiting conditions	Orientation	Translation
$a$	$h0l: h=2n$		$a/2$
$a$	$hk0: h=2n$	$\perp b$	$a/2$
$b$	$0kl: k=2n$	$\perp c$	$b/2$
$c$	$h0l: l=2n$	$\perp b$	$c/2$
$n$	$0kl: k+l=2n$	$\perp a$	$b/2 + c/2$

The experimental X-ray diffraction pattern allows us also to determine the unit cell type. With the help of systematic absences, the unit cell centering may be assigned and the cell classified as primitive, (*P* type), side-centered (*A*, *B*, or *C* type), face-centered (*F* type), or body-centered (*I* type).

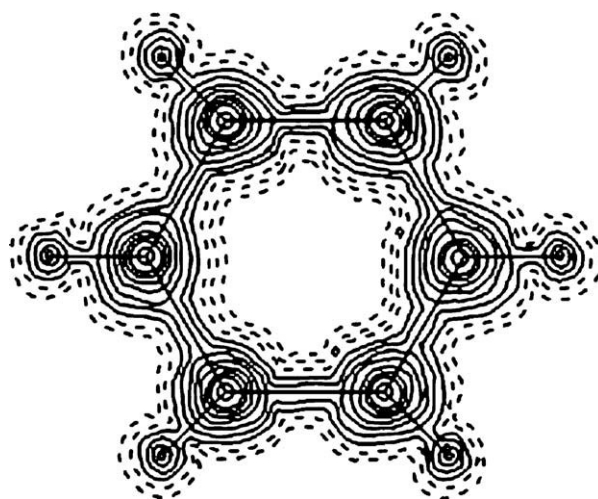
No systematic absences are present for a primitive unit cell. The indices are unmixed (all indices odd or all even; otherwise is called 'mixed') for a side-centered unit cell; showing only the  $k+l=2n$  reflections for the *A* type;  $h+l=2n$  for *B*; and  $h+k=2n$  for *C*. All reflections with mixed indices are missing in a face-centered unit cell, while for a body-centered unit cell all  $h+k+l=\text{odd}$  reflections are missing.

The diffraction pattern can also reveal translation-symmetry elements by considering the systematic absences. These elements are the screw-axes (the action of a  $R_n$  screw axis along the '*a*' direction is a rotation  $2\pi/R$  followed by a translation  $(n/R) \cdot a$ ) and the glide planes (a combination of reflection and translation). **Tables 1** and **2** show the limiting conditions for some screw-axes and glide-planes. Full information on crystallographic symmetry groups is compiled in the *International Tables for Crystallography*, vol. A.

Using this information, the proper space group may be assigned to the experimental data, although the existence of a symmetry centre may remain uncertain in some cases.

## Electron Density and the Phase Problem

The electron density function,  $\rho(r)$  (measured in  $e^- \text{\AA}^{-3}$ ), allows us to know the molecular structure



**Figure 5** Projection of the electron density contour map for benzene on the plane of the molecule. The positive and zero isocontours are represented by solid and dashed lines, respectively.

from which a crystal is made; the electron density is concentrated in the vicinity of the nucleus, showing peaks at the electron density maxima (atomic positions) and taking relatively low values out of the atomic positions.

The electron density is usually computed by Fourier synthesis at the points (*x*, *y*, *z*) of a three-dimensional grid. A grid spacing of  $\sim 0.2\text{--}0.4 \text{\AA}$  in the three directions is satisfactory for most electron density maps.

In order to facilitate interpretation of these three-dimensional maps, the electron density is presented in sections (with a constant value of *x*, *y*, or *z*) that may be contoured by lines passing through points of equal electron density. For illustration, **Figure 5** shows the features of a typical two-dimensional electron density map. To a first approximation, the heights of the peaks are proportional to the atomic number.

The process of finding the electron density is not straightforward. This function can be expressed as a three-dimensional Fourier series,

$$\rho(x, y, z) = \frac{1}{V_{\text{cell}}} \sum_h \sum_k \sum_l F(hkl) e^{-2\pi i(hx + ky + lz)} \quad [10]$$

where  $F(hkl)$  is the structure factor, as was shown in eqns [6] and [7]. The number of reflections used in practice in this expression is over several thousands for the small unit cells typical of common organic/inorganic compounds.

Alternative expressions, considering the Friedel's law ( $|F(hkl)| = |F(-h - k - l)|$ ) and the definition of

complex function, lead to

$$\rho(x, y, z) = \frac{2}{V_{\text{cell}}} \sum_{h=0} \sum_k \sum_l |F(hkl)| \times \cos[2\pi(hx + ky + lz) - \Phi(hkl)] \quad [11]$$

that shows clearly the dependence of  $\rho(x, y, z)$  on the phase  $\Phi(hkl)$ . This is the essence of the crystallographic problem referred to as the phase problem in crystallography: in eqn [11] the values of the moduli  $|F(hkl)|$  can be obtained immediately from the observed intensities (according to eqn [8]), but the phase information is lost in the experiment and cannot be found directly from the experimental data.

However it is possible in general to solve the phase problem; because the number of equations is greater than the number of unknowns, the system is said to be overdetermined.

## Jumping the Phase Problem to Solve the Crystal Structure

In order to solve the structure (have a model of the structure to be refined), methods of extracting phase information are needed, for which the set of  $|F_0(hkl)|$  data (subscript o means ‘observed’ or ‘experimental’) constitute the starting point for the methods of structure analysis.

A general solution to the phase problem has not been found yet, but there are several methods that can be applied successfully. They all need a considerable amount of computation, but with modern computers and software, this process is relatively easily performed. There are two classical approaches to solving the phase problem.

### Patterson Methods

Based on the three-dimensional function proposed by Patterson in 1934, a new Fourier series that could be calculated directly from the measured intensities. This function is defined as the self-convolution of the electron density,  $\rho(r)$ , and has the same periodicity as the electron density:

$$P(u, v, w) = \frac{2}{V_{\text{cell}}} \sum_h \sum_k \sum_l |F(hkl)|^2 \times \cos[2\pi(hu + kv + lw)] \quad [12]$$

The Patterson function gives information about interatomic vectors (distances) in the structure. The height of each peak is proportional to the product of the atomic numbers of the two atoms connected ( $Z_i, Z_j$ ) multiplied by the multiplicity of the vector.

Patterson methods are specially oriented to solving structures containing heavy atoms, where it becomes easy to derive the projected positions of the heavy atoms, and their coordinates represent a good initial model for solving the structure. The calculated phases,  $\Phi_c(hkl)$ , are a good approximation of the true phase for calculating an electron density map by means of eqn [11], using the observed amplitudes,  $|F_0(hkl)|$ , and the calculated phases to find new atomic positions and proceed by successive Fourier cycles until the structure is completed.

Nowadays, procedures for the automatic localization of the heavy atom positions from the Patterson function have been developed both for small molecules and for biological macromolecules.

Alternative ways of exploiting the information in a Patterson map are known as molecular replacement methods, being able to locate structural fragments and using this initial phase information to proceed as in the case of heavy atoms.

All these procedures are nicely implemented in fully automated packages such as DirDif, SHELXS97, or AMORE (for proteins).

### Direct Methods

This term is used to indicate those methods that try to determine the values of the structure factor phases *ab initio* from the amplitudes of the structure factors through mathematical relationships. Particularly useful is their ability to yield good phase information for structures without heavy atoms.

The first mathematical relationships were obtained, starting about 1950, between the sign of phases in centrosymmetric structures in the form of inequalities, later extended to noncentrosymmetric structures, using unitary structure factors.

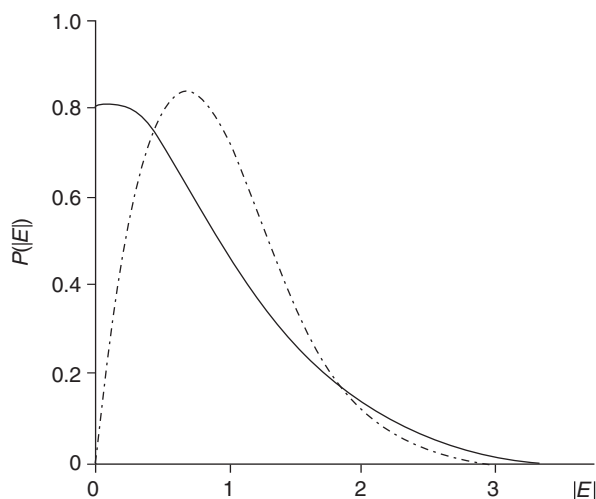
$$U(hkl) = F(hkl) / \sum_{j=1}^N f_j \quad [13]$$

The inequalities have limited practical interest because it requires very large values of  $|U(hkl)|$  to give useful information for solving a structure.

It was in 1953 that Hauptman and Karle established the basic concepts and the probabilistic foundations of direct methods. Of major importance was the introduction of the normalized structure factor, given by the equation

$$E(hkl) = F(hkl) / (\varepsilon \Sigma)^{1/2} \quad \text{with} \quad \Sigma = \sum_{j=1}^N f_j^2 \quad [14]$$

where  $\Sigma$  represents the average value of  $|F|^2$  for the scattering angle corresponding to  $(hkl)$ , and  $\varepsilon$  the enhancement factor, depending on the space group and the reflection itself.



**Figure 6** Probability distributions of  $|E(hkl)|$  for noncentrosymmetric (---) and centrosymmetric (—) structures.

The structure factor,  $E(hkl)$  (and  $U(hkl)$ ), is independent of the scattering angle,  $\theta$ . This is very important in the application of direct methods since high-order reflections with small  $|F(hkl)|$  values can have quite large  $|F(hkl)|$  values.

Other theoretical advantage of the structure factor  $E(hkl)$  is that the distribution of  $|F(hkl)|$  depends only on the space group and not on the complexity of the structure; in this way the form of its probability distributions shows the presence of an inversion center in the crystal (centric distribution) or not (acentric distribution), see **Figure 6**.

In general, the direct methods require initially a few reflections with phases known, either uniquely or symbolically. Some basic features of direct methods are shown briefly below; a detailed description of procedures and applications is beyond the scope of this article and can be found in the work listed in the further reading section.

Structure invariants are quantities independent of the chosen reference system, their value depending only on the structure. The product  $E(h_1k_1l_1)E(h_2k_2l_2) \dots E(h_nk_nl_n) = |E(h_1k_1l_1)E(h_2k_2l_2) \dots E(h_nk_nl_n)| \exp[i(\Phi_1\Phi_2 \dots \Phi_n)]$  is a structure invariant when  $(h_1k_1l_1) + (h_2k_2l_2) + \dots + (h_nk_nl_n) = 0$ .

The most important structure invariants are the triplet invariants, represented by  $[F(-h_1 - k_1 - l_1)F(h_2k_2l_2)F(h_1 - h_2k_1 - k_2l_1 - l_2)]$ , and the quartet invariants  $[F(-h_1 - k_1 - l_1)F(h_2k_2l_2)F(h_3k_3l_3)F(h_1 - h_2 - h_3k_1 - k_2 - k_3l_1 - l_2 - l_3)]$ . The terms ‘triplet’ and ‘quartet’ refer to the sums of the phases since the moduli are origin independent.

The structure semi-invariants are single phases, or linear combinations of phases, which are invariant only for a permitted change of origin. The origin

is restricted in the cell to the sites with the same point symmetry (called permissible origins). A semi-invariant can be transformed into an invariant by adding one or more pairs of symmetry-equivalent phases.

Probability methods are applied to the estimation of the reliability of the relationships between the phases and magnitudes, and their use has proved to be the most important approach in the practical use of direct methods.

The tangent formula provides a powerful equation for establishing relations between structure factors and phases and for refining a probabilistic set of phases till self-consistence: this process is usually known as convergence. With the knowledge of a sufficiently large subset of phases, an electron density map can be calculated using eqn [11] to reveal the structure. These maps are called  $E$ -maps because they use the normalized structure factor,  $E(hkl)$ , as coefficients of the Fourier series rather than the structure factor,  $F(hkl)$ .

Direct methods may be automatically applied using program packages like SIR2004.

## Completing and Refining the Structure

Frequently the initial set of phases produces incomplete structural models. These models have to be completed and refined to obtain the final crystal structure.

The complete structure may be obtained using the difference Fourier synthesis method, based on the difference between the electron density observed,  $\rho_o(x, y, z)$ , and the electron density calculated,  $\rho_c(x, y, z)$ .

$$\Delta\rho(x, y, z) = \rho_o(x, y, z) - \rho_c(x, y, z) \quad [15]$$

with both magnitudes calculated as in eqn [10], but the values of  $\Phi_o$  are unknown, making it necessary to make the approximation  $\Phi_o \approx \Phi_c$ . The electron density difference then becomes

$$\Delta\rho(x, y, z) = \frac{1}{V_{\text{cell}}} \sum_b \sum_k \sum_l (|F_o(hkl)| - |F_c(hkl)|) \exp[-2\pi i(hx + ky + lz) + i\phi_c(hkl)] \quad [16]$$

If the structural model is incomplete, the missing atomic positions appear as positive maxima in a difference Fourier synthesis ( $\Delta\rho$ ).

For refining the structure, the most widely used method is the least-squares method. In essence, the process consists of adjusting a series of parameters



defining the structural model to minimize the difference between the experimental values,  $|F_o(hkl)|$ , and the calculated values,  $|F_c(hkl)|$ , derived from the structure model. Among these parameters are the overall scale factor, the coordinates for each atom, and their isotropic or anisotropic thermal parameters.

The refinement is carried out using  $F^2$  instead of  $F(hkl)$  for all independent measured reflections. This refinement process requires several cycles until convergence is reached to conclude, as a final step, by computing a difference Fourier synthesis in order to locate the hydrogen atoms. Sometimes, when H atoms are difficult to locate, it is possible to assign geometrically calculated positions for the missing hydrogen atoms.

The most widely used measure of convergence is the residual index or  $R$ -factor, expressing the disagreement between the  $|F_o(hkl)|$  and the  $|F_c(hkl)|$  values. The most common  $R$ -indices are based on  $F^2$  and  $F$ .

$$wR2 = \left\{ \sum_{hkl} [w(F_o^2 - F_c^2)]^2 / \sum_{hkl} w(F_o^2)^2 \right\}^{1/2} \quad [17]$$

$$R1 = \sum_{hkl} (|F_o| - |F_c|) / \sum_{hkl} |F_o| \quad [18]$$

and the goodness of fit, always based on  $F^2$ .

$$\text{GOF} = S = \left\{ \sum_{hkl} [w(F_o^2 - F_c^2)]^2 / n - p \right\}^{1/2} \quad [19]$$

where  $w$  is a weighting scheme, important for associating a proper weight to each reflection;  $n$  is the number of reflections; and  $p$  is the total number of parameters refined. Low values of  $wR2$  ( $<15\%$ ),  $R1$  ( $<5\%$ ), and  $S$  ( $\sim 1.0$ ) are indications of the correctness of the crystal structure.

One complementary criterion of model quality during the refinement is the free- $R$  factor, which is computed with a small set of randomly chosen intensities, the test set, which are not used in the refinement.

The structural model used in general for the structural refinement assumes that the atoms are neutral and the electron density is spherically distributed about the atomic nuclei, with a radial dependence. This atomic model is adequate for structure determination but ignores the real effect of chemical bonding.

In the last three decades, a more flexible formalism was carried out, incorporating expansion and contraction of the atomic valence shell, variation of its

occupancy (loss of neutrality), and nonspherical density functions (multipolar spherical harmonic expansion).

This multipole refinement leads to a significant decrease in the residual indices and a reduction of features in the residual maps. Also, a topological study of the electron density described in this form allows us to obtain more quantitative information on electron distributions than can be derived from electron density maps, the bond order and its force, dipolar moments, covalent character, etc.

The limitations of the multipolar refinement are the exceptional demands it places on the X-ray data (in terms of quantity and quality) and the large number of additional parameters per atom that must be included in the least-squares refinement procedures. Nevertheless, charge density studies are becoming more and more popular due to the new experimental possibilities offered by the new area detectors and the more intense sources as provided by the synchrotrons.

The high-intensity, finely collimated polychromatic X-radiation from a synchrotron source can be used in a great variety of applications in modern crystallographic research.

## Macromolecular Crystallography

Macromolecular crystallography has had an extraordinary impact in the progress of major areas of research such as genomics and proteomics. The structure of proteins and other macromolecules of biological interest are best determined using X-ray crystallography.

Some of the so-called physical methods are among the most commonly applied in protein crystallography as the isomorphous replacement method and the anomalous scattering method.

The isomorphous replacement method is based on the experimental data sets of two or more structures differing only in the presence/absence of some atoms or in the presence of some atom of a different kind.

Such structures are called isomorphs, and the comparison of their intensity data can lead to a determination of their structures.

For proteins, the structures selected are the native protein and, in the most favorable cases, several isomorphous derivatives with different heavy atoms, every one bound at a certain site in the protein.

The structure factors are related as

$$|F_P| \exp(i\Phi_P) = |F_{PH}| \exp(i\Phi_{PH}) - F_H \quad [20]$$

where P stands for the native protein, PH for the heavy atom derivative, and H for the heavy atom.  $F_H$

can be determined using different sets of isomorphous derivatives; it is possible to obtain approximate phases capable of yielding interpretable electron density maps.

The anomalous scattering method uses the information provided by the anomalous signal. Anomalous scattering occurs when the frequency of the primary beam is close to the natural absorption frequency of some of the atoms in the structure. The anomalous signal is negligible for light atoms (C, N, O, ...) and only important when some heavy atoms are present.

In this case, the scattering factor is written in the form

$$f = f^o + f' + if'' \quad [21]$$

where  $f^o$  is the normal scattering, and  $f'$  and  $f''$  are the real and imaginary component corrections, respectively. The structure factor equation is made of three contributions taking the form

$$F(hkl) = F^o(hkl) + F'(hkl) + F''(hkl) \quad [22]$$

The most important consequence of this effect is the breaking of Friedel's law (the observation of the effects of the violation of Friedel's law is conditioned by the wavelength and the atoms in the crystal), and Friedel pairs are more correctly called Bijvoet pairs.

This breaking can be used to determine unambiguously the absolute configuration of the structure (in chiral crystals); calculating  $|F_c(hkl)|$  for a number of Friedel (Bijvoet) pairs, one of the enantiomorphs will fit better, and this will be the correct absolute configuration.

Using the anomalous scattering effect, we can locate heavy atom sites (very important in protein studies), and knowing their positions, it is possible to get phase information, although with the ambiguity

$$\Phi(hkl) = \Phi''(hkl) \pm \Delta\Phi(hkl) \quad [23]$$

$$|\Delta\Phi(hkl)| \approx \left| \cos^{-1} \left[ (|F^+| - |F^-|) / 2F'' \right] \right| \quad [24]$$

where  $F^+$  and  $F^-$  are structure factors of a Friedel pair.

The resolution of this equation leads to a solution of the structure. There are various ways of trying to solve this ambiguity, some of which are based on the use of a single wavelength (most protein crystals are sensitive to X-ray irradiation, and isomorphous derivatives are difficult to prepare), and others are based on a combination of different sources of information: (1) combining single isomorphous replacement and one-wavelength anomalous scattering; (2) introducing a second wavelength, generalized in recent years for

measuring data from the same species at different wavelengths, a technique called MAD (multiwavelength anomalous dispersion); (3) using algebraic methods based on simple rules that permit estimation of three-phase structure invariants; (4) using probability formulae integrating direct methods with one-wavelength anomalous scattering; (5) using Wang's method; (6) using direct methods; (7) resolving the phase ambiguity using Wilson statistics.

Finally, the other type of physical method, in an active state of development during the last few years, is that of multibeam reflection methods. These methods are based on the  $N$  beam diffraction effect ( $N > 1$ , integer), which occurs when  $N$  nodes of the reciprocal lattice diffract for a unique incident beam. These simultaneous reflections are very well known processes and can be treated in accordance with dynamic diffraction theory.

For instance, three-beam diffraction (one transmitted and two diffracted beams) is used for determination of triplet phases (see Direct Methods) and provides also a means of resolving the enantiomorphism problem without the need for anomalous scattering.

Experimentally,  $N$  beam diffraction can be achieved by rotating the crystal on a specially designed goniometer. This technique has been proved to be able to provide phase information for macromolecular crystals.

## Databases

There are three major databases storing all the crystal structures determined so far: the CSD (Cambridge Structural Database) contains all the organic and metal-organic compounds; the ICSD (Inorganic Crystal Structure Database) contains the inorganic compounds; and the PDB (Protein Data Bank) contains macromolecular structure data on proteins, nucleic acids, protein-nucleic acid complexes, and viruses.

The CSD, <http://www.ccdc.cam.ac.uk/>, contains more than 270 000 crystal structures determined using X-ray or neutron techniques. For each structure, the data stored include bibliographic information and experimental data, the chemical connectivity, and the crystal and molecular structures.

The ICSD, <http://www.fiz-informationsdienste.de/en/DB/icsd/index.html>, has complete structural information on more than 70 000 inorganic compounds.

The PDB, <http://www.rcsb.org/pdb/>, is freely available worldwide and stores sequence details, atomic coordinates, crystallization conditions, and three-dimensional structure and neighbors computed using various methods, derived geometric data, structure factors, and three-dimensional images.

## Instrumentation

Most of the single crystal structure determination experiments are based on the use of automated diffractometers.

Diffractometers are either used as laboratory instruments, based on the radiation produced by sealed or

rotating anode tubes and more recently in micro-sources, or as instruments set at the synchrotron sources. Apart from the X-ray source, the goniometer and detector are the fundamental parts of a diffractometer. A picture of a modern diffractometer is shown in **Figure 7**.

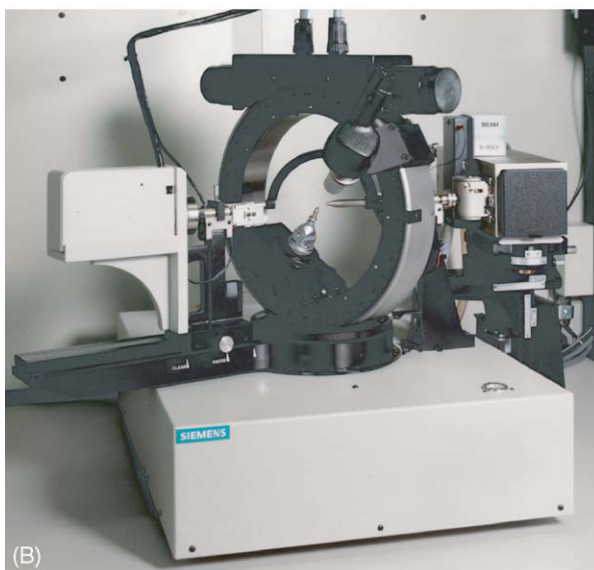
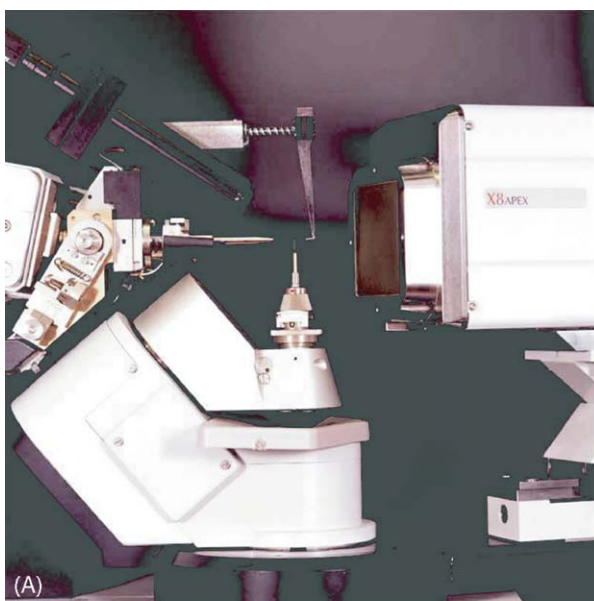
Nowadays most diffractometers use area detectors based on charge coupled device (CCD) technology (see **Figure 8**) although image plate devices are still used, particularly in macromolecular crystallography. Area detectors can collect a complete set of data in a few hours.



**Figure 7** One of the recently designed single crystal diffractometers equipped with a CCD detector and Kappa geometry. (Reproduced with permission from Brüker AXS.)



**Figure 8** View of one CCD detector. (Reproduced with permission from Brüker AXS.)



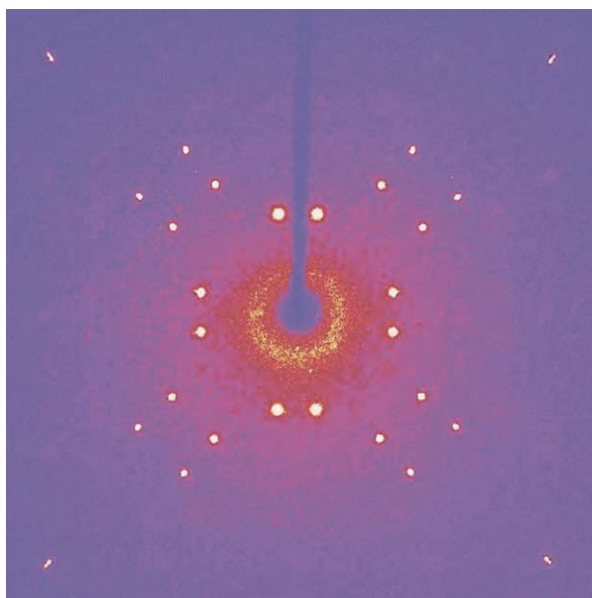
**Figure 9** The two most commonly used four-circle geometries, Kappa (A) and Euler (B). (Reproduced with permission from Brüker AXS.)

The present goniometers have been inspired by Euler or Kappa four circle geometries (see **Figure 9**) used in combination with point detectors; nevertheless the use of area detectors allows us to simplify the goniometer geometries by reducing the number of circles.

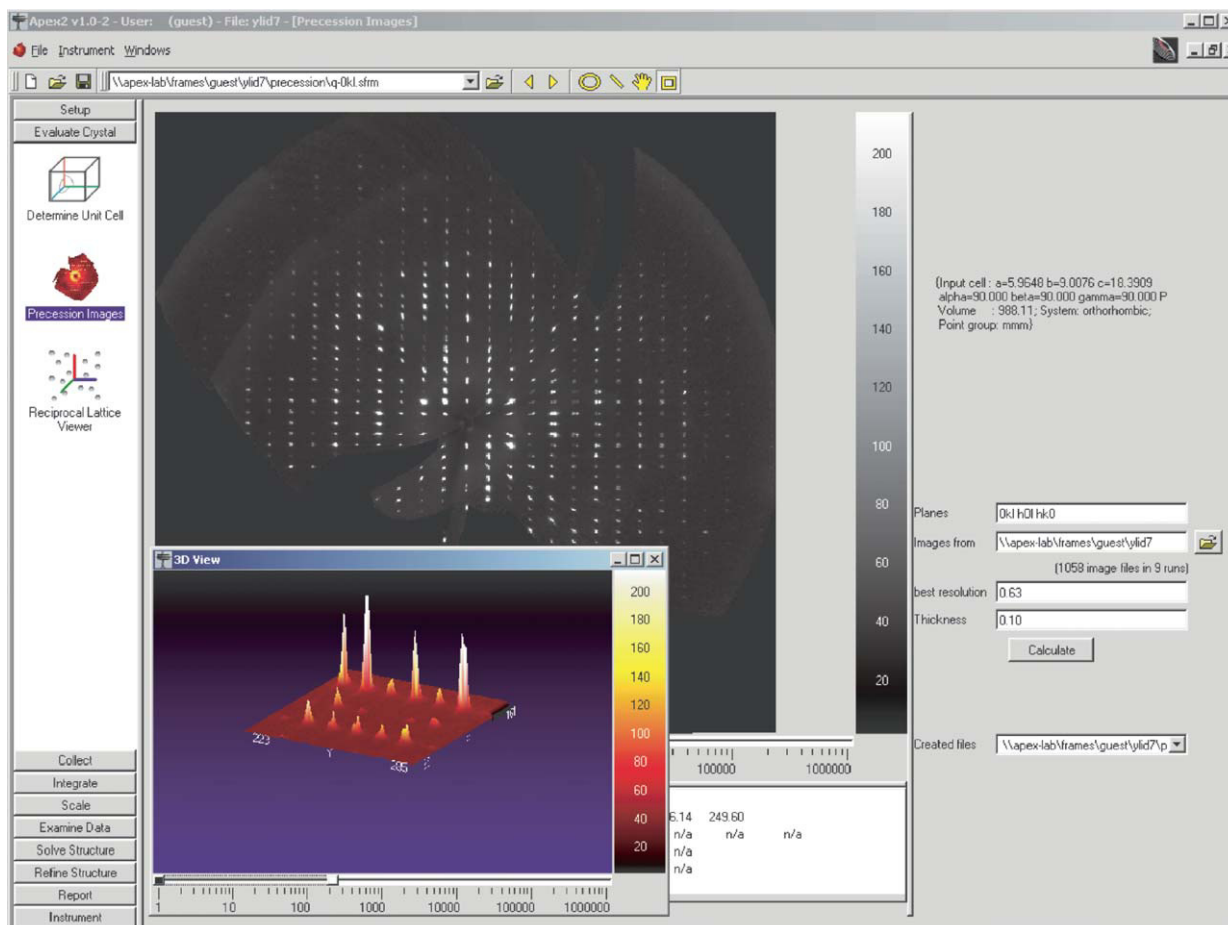
The output of an area detector is a set of frames containing a three-dimensional picture of the reciprocal space obtained by successive rotations of the crystal following a convenient, previously calculated strategy. **Figure 10** shows a frame as detected by a CCD for one small molecule/unit-cell.

The diffraction spots are integrated and scaled using automated procedures to get the intensities and structure factors necessary for solving and refining the crystal structure. **Figure 11** shows a screen of the computer when processing a collected set of data.

The advent of synchrotron radiation as a highly collimated and tunable source of X-rays has opened new prospects for the application of anomalous scattering. Several different data sets can be collected. In multiple-beam scattering methods, synchrotron radiation also offers advantages.



**Figure 10** Diffraction spots, for a small molecule/unit-cell, on a frame detected using a CCD.



**Figure 11** Computer screen showing the processing of a set of collected frames.

Tapping this X-ray source involves highly sophisticated technology tailored for particular applications, which include rapid data collection from unstable samples (e.g., proteins), poorly diffracting samples such as fibrous polymers or extremely small crystals, time-resolved studies of solid-state reactions or transformations including enzyme-catalyzed processes, and studying of crystals defects using X-ray topography.

## Safety

The specific hazards of analytical X-ray equipment can include exposure to an intense, localized primary X-ray beam, exposure to diffracted and/or scattered portions of the primary X-ray beam (includes X-ray leakage), and the high voltages used in generating the X-ray themselves.

There are three general principles of radiation protection: time, distance, and shielding. Decreasing the amount of time spent in the vicinity of the source of radiation, increasing the distance from a source of radiation, and increasing the amount of shielding around a source of radiation will decrease the amount of radiation exposure. Shielding for analytical X-ray units can range from the use of leaded glass to enclosures constructed of tin-impregnated polycarbonate.

The careful, alert, and informed worker need feel no hesitancy whatsoever in operating X-ray diffraction

equipment, but the use of such equipment by careless or uninformed persons is dangerous.

X-ray installations must be controlled periodically using Geiger counters, and personnel must wear badges and do periodical blood analysis.

**See also:** **Fourier Transform Techniques.** **Infrared Spectroscopy:** Near-Infrared. **Qualitative Analysis.** **Structural Elucidation.** **Thermal Analysis:** Overview. **X-Ray Absorption and Diffraction:** Overview.

## Further Reading

Giacovazzo C *et al.* (2002) Fundamentals of crystallography. In: *International Union of Crystallography Texts on Crystallography*, no. 7, 2nd edn. Oxford: Oxford University Press.

Ladd MFC and Palmer RA (1994) *Structure Determination by X-Ray Crystallography*, 3rd edn. New York: Plenum Press.

Rhodes G (2000) *Crystallography Made Crystal Clear*, 2nd edn. New York: Academic Press.

Taylor G (2003) The phase problem. *Acta Crystallographica* D59: 1881–1890.

Warren BE (1990) *X-Ray Diffraction*. New York: Dover Publications Inc.

Weckert E and Hümmer K (1997) Multiple-beam X-ray Diffraction for physical determination of reflection phases and its applications. *Acta Crystallographica* A53: 108–143.

Woolfson M and Hai-Fu F (1995). *Physical and Non-Physical Methods of Solving Crystal Structures*. Cambridge: Cambridge University Press.

# X-RAY ABSORPTION FINE STRUCTURE SPECTROSCOPY

See X-RAY ABSORPTION AND DIFFRACTION: Overview; X-Ray Absorption

# X-RAY FLUORESCENCE AND EMISSION

## Contents

### X-Ray Fluorescence Theory

### Wavelength Dispersive X-Ray Fluorescence

### Energy Dispersive X-Ray Fluorescence

### Total Reflection X-Ray Fluorescence

### Particle-Induced X-Ray Emission

### Synchrotron X-Ray Fluorescence

Tapping this X-ray source involves highly sophisticated technology tailored for particular applications, which include rapid data collection from unstable samples (e.g., proteins), poorly diffracting samples such as fibrous polymers or extremely small crystals, time-resolved studies of solid-state reactions or transformations including enzyme-catalyzed processes, and studying of crystals defects using X-ray topography.

## Safety

The specific hazards of analytical X-ray equipment can include exposure to an intense, localized primary X-ray beam, exposure to diffracted and/or scattered portions of the primary X-ray beam (includes X-ray leakage), and the high voltages used in generating the X-ray themselves.

There are three general principles of radiation protection: time, distance, and shielding. Decreasing the amount of time spent in the vicinity of the source of radiation, increasing the distance from a source of radiation, and increasing the amount of shielding around a source of radiation will decrease the amount of radiation exposure. Shielding for analytical X-ray units can range from the use of leaded glass to enclosures constructed of tin-impregnated polycarbonate.

The careful, alert, and informed worker need feel no hesitancy whatsoever in operating X-ray diffraction

equipment, but the use of such equipment by careless or uninformed persons is dangerous.

X-ray installations must be controlled periodically using Geiger counters, and personnel must wear badges and do periodical blood analysis.

**See also:** **Fourier Transform Techniques.** **Infrared Spectroscopy:** Near-Infrared. **Qualitative Analysis.** **Structural Elucidation.** **Thermal Analysis:** Overview. **X-Ray Absorption and Diffraction:** Overview.

## Further Reading

- Giacovazzo C *et al.* (2002) Fundamentals of crystallography. In: *International Union of Crystallography Texts on Crystallography*, no. 7, 2nd edn. Oxford: Oxford University Press.
- Ladd MFC and Palmer RA (1994) *Structure Determination by X-Ray Crystallography*, 3rd edn. New York: Plenum Press.
- Rhodes G (2000) *Crystallography Made Crystal Clear*, 2nd edn. New York: Academic Press.
- Taylor G (2003) The phase problem. *Acta Crystallographica* D59: 1881–1890.
- Warren BE (1990) *X-Ray Diffraction*. New York: Dover Publications Inc.
- Weckert E and Hümmer K (1997) Multiple-beam X-ray Diffraction for physical determination of reflection phases and its applications. *Acta Crystallographica* A53: 108–143.
- Woolfson M and Hai-Fu F (1995). *Physical and Non-Physical Methods of Solving Crystal Structures*. Cambridge: Cambridge University Press.

# X-RAY ABSORPTION FINE STRUCTURE SPECTROSCOPY

See X-RAY ABSORPTION AND DIFFRACTION: Overview; X-Ray Absorption

# X-RAY FLUORESCENCE AND EMISSION

## Contents

### X-Ray Fluorescence Theory

### Wavelength Dispersive X-Ray Fluorescence

### Energy Dispersive X-Ray Fluorescence

### Total Reflection X-Ray Fluorescence

### Particle-Induced X-Ray Emission

### Synchrotron X-Ray Fluorescence



## X-Ray Fluorescence Theory

P J Potts, The Open University, Milton Keynes, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

X-ray fluorescence (XRF) spectrometry is an analytical technique used widely for elemental determinations in both industrial, and research and development laboratories. It belongs to a family of atomic spectrometry techniques but enjoys some distinctive analytical advantages compared with rival techniques, such as atomic absorption spectrometry (AAS) and inductively coupled plasma–atomic emission spectrometry (ICP–AES). XRF spectrometry is capable of determining almost all the elements in the periodic table, conventionally between sodium and uranium (including the electronegative elements, such as chlorine, phosphorus, sulfur, bromine, and iodine) to detection limits, often in the 1–10 mg kg<sup>−1</sup> range. Determinations can be extended down to boron, although at reduced sensitivity and with significant limitations in terms of the very small critical penetration depth of the corresponding fluorescence X-rays (see later). Unusually for an atomic spectrometry technique, samples are normally analyzed in solid (rather than solution) form, for example, as cast billets of metal, powdered material prepared as a pellet, or glass disks formed after fusion of the sample powder with a suitable flux. However, oils, liquors, brines, and other samples in the form of a liquid can also be analyzed directly, with minor modification to the analytical procedure (principally substituting a helium flush in the vacuum chamber). One important analytical characteristic of XRF is the inherently high precision of measurements that can be achieved in routine operation, generally superior to many other atomic spectrometry techniques other than those based on isotope dilution mass spectrometry. In comparison with ICP–mass spectrometry (MS), another multielement technique that is widely used for elemental determinations, XRF cannot match the low detection limits for the higher-atomic number elements, although ICP–MS itself suffers interferences on a number of lower-atomic number elements and cannot match the precision in their determination using XRF, so that in some routine applications XRF and ICP–MS are considered to be complementary techniques in laboratories undertaking comprehensive multielement analysis.

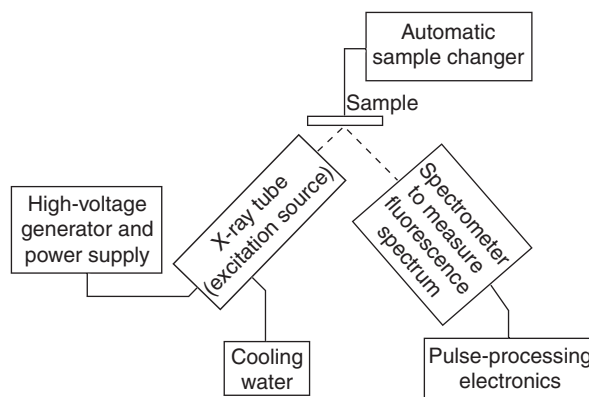
Taking account of its analytical capabilities, XRF has found widespread applications in the analysis of

cements, ceramics, refractories, iron, steel, nonferrous metals, plastics (for inorganic fillers), lubricating oils (for wear metals), fuel (for sulfur), geological samples, soils, sediments, air filters, archaeological, and forensic and biological samples. In addition, special forms of portable instrumentation have been developed for *in situ* measurements of contaminated soil, lead in paint, and alloy sorting (scrap metal recycling). *In vivo* assessment of lead in bone may be made using XRF as well as special forms of instrumentation for characterizing semiconductor materials. Because of the advantages XRF brings to these applications, XRF continues to play a dominant role in the relevant application laboratories.

### Forms of Instrumentation

Conventional XRF analyzers used for routine measurements in applications laboratories comprise an excitation source, sample presentation device, X-ray spectrometer, and associated control and signal-processing electronics (Figure 1). Two categories of instrumentation are available, one based on the use of wavelength dispersive (WD) spectrometers, the other on energy dispersive (ED) X-ray detectors. WD-XRF instruments are used widely for routine elemental analysis in production, quality control, and research laboratories, where speed of analysis and analytical precision are important requirements. Such instrumentation uses an X-ray tube as the excitation source and is available with one or two scanning goniometers (for sequential elemental analysis) or, alternatively, with a number of multiple fixed goniometer channels (for rapid simultaneous elemental detection).

ED-XRF instrumentation incorporating solid state semiconductor detectors offers different detection



**Figure 1** Schematic 'compartment' model of an XRF analyzer system showing the major instrumental components.

characteristics with parallel detection of the full X-ray spectrum. In addition to quantitative elemental analysis, such instrumentation offers unrivalled opportunities for rapid qualitative analysis, of value in forensic and alloy sorting applications, for example. Adequate excitation is provided by a lower-power X-ray tube arranged for direct or secondary target excitation, or alternatively a radioisotope excitation source may be used.

Despite these differences in analytical performance and operation, WD- and ED-XRF instrumentation are based on common principles, which are considered further in this section.

## Principles of Excitation

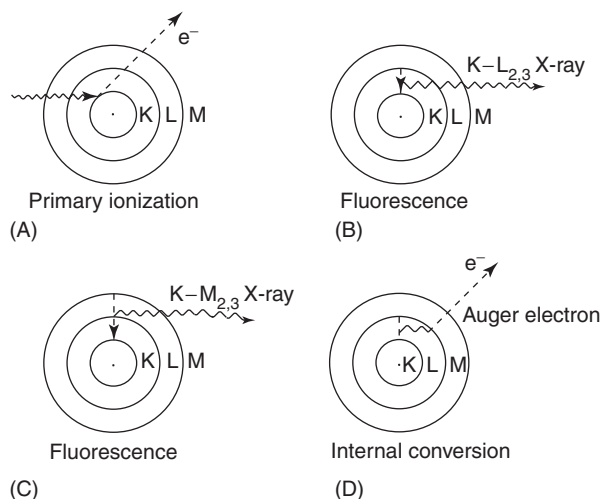
XRF involves the excitation of atoms in a sample using a source of X-ray photons, which are sufficiently energetic to cause ionization of inner-shell electrons. The resultant atoms are intrinsically unstable since they have a vacancy in an inner electron orbital and decay immediately to a more stable electronic state. This decay involves the transition of an electron from an outer orbital to fill this vacancy (Figure 2). As the electron falls down the potential energy gradient to an orbital closer to the nucleus, it loses energy by emitting an X-ray photon, so regaining a stable energetic configuration. The energy of

this photon is characteristic of the difference in energy between the two electronic states involved in the transition. Three analytically important series of X-ray lines can be emitted as a result of this excitation mechanism. These line series originate from the ionization of the K, L, or M orbitals. This excitation mechanism contrasts with that relevant to other atomic spectrometry techniques such as AAS and ICP-AES, which involve absorption and emission of radiation in the ultraviolet-visible region of the spectrum following the excitation of outer orbital electrons of atoms and ions.

## Line Series and Notations

Several alternative electronic transitions can occur following ionization of an inner-shell electron during the XRF process. For example, a vacancy created by ionization of a K-shell electron can be filled by transition of an electron from the L or M orbital level (Figure 2). The probability that an individual transition will occur (and, therefore, the relative intensity of the corresponding fluorescence line) is governed by quantum mechanical selection rules. Under the traditional Siegbahn notation, individual lines are identified using a convention based mainly on relative intensity, the two K-line transitions referred to above being identified as  $K\alpha$  and  $K\beta$  lines, respectively. To account for the fact that L and M orbital levels are not degenerate but are split in energy (an effect that becomes more significant with increased atomic number), corresponding fluorescence lines are subclassified. For example, the  $K\alpha$  line comprises two transitions,  $K\alpha_1$  and  $K\alpha_2$ , to describe the  $K-L_3$  and  $K-L_2$  transitions, both of which are permitted by selection rules. The principal fluorescence lines in the L and M spectra are identified in a similar manner, the more important lines being ( $L\alpha$ ,  $L\beta$ ,  $L\gamma$ ,  $L\eta$ ,  $M\alpha$ , and  $M\beta$ ). Although still widely used, this traditional notation has now been replaced by an International Union of Pure and Applied Chemistry (IUPAC) convention based on the rigorous use of spectroscopic terms. The  $K\alpha_1$  and  $K\alpha_2$  transitions are now denoted as  $K-L_3$  and  $K-L_2$  transitions, respectively, and the equivalence between the IUPAC and Siegbahn notations for other fluorescence lines is listed in Table 1.

One important aspect of XRF is the way in which the energy of each line series increases in a systematic manner with increasing atomic number of elements in the periodic table (Figure 3). This correlation results from the systematic increase in the potential energy of orbitals around an atom with increase in positive charge on the nucleus (and therefore with the atomic number of the element). This phenomenon



**Figure 2** Schematic mechanism of X-ray fluorescence. (A) Interaction of primary X-ray photon with electronic orbital of an atom causing ionization of a K-shell electron. (B) Deexcitation involving an electron transition from the L shell to the K shell accompanied by the emission of a  $K-L_{2,3}$  ( $K\alpha$ ) fluorescence X-ray. (C) Competitive mode of deexcitation involving an electron transition from the M shell to the K shell accompanied by the emission of a  $K-M_{2,3}$  ( $K\beta$ ) fluorescence X-ray. (D) Internal capture of a fluorescence X-ray leading to the emission of an Auger electron.

**Table 1** Correspondence between Siegbahn and IUPAC notation diagram lines

Siegbahn	IUPAC	Siegbahn	IUPAC	Siegbahn	IUPAC	Siegbahn	IUPAC
K $\alpha_1$	K-L <sub>3</sub>	L $\alpha_1$	L <sub>3</sub> -M <sub>5</sub>	L $\gamma_1$	L <sub>2</sub> -N <sub>4</sub>	M $\alpha_1$	M <sub>5</sub> -N <sub>7</sub>
K $\alpha_2$	K-L <sub>2</sub>	L $\alpha_2$	L <sub>3</sub> -M <sub>4</sub>	L $\gamma_2$	L <sub>1</sub> -N <sub>2</sub>	M $\alpha_2$	M <sub>5</sub> -N <sub>6</sub>
K $\beta_1$	K-M <sub>3</sub>	L $\beta_1$	L <sub>2</sub> -M <sub>4</sub>	L $\gamma_3$	L <sub>1</sub> -N <sub>3</sub>	M $\beta$	M <sub>4</sub> -N <sub>6</sub>
K $\beta_2^I$	K-N <sub>3</sub>	L $\beta_2$	L <sub>3</sub> -N <sub>5</sub>	L $\gamma_4$	L <sub>1</sub> -O <sub>3</sub>	M $\gamma$	M <sub>3</sub> -N <sub>5</sub>
K $\beta_2^{II}$	K-N <sub>2</sub>	L $\beta_3$	L <sub>1</sub> -M <sub>3</sub>	L $\gamma_4$	L <sub>1</sub> -O <sub>2</sub>	M $\zeta$	M <sub>4,5</sub> -N <sub>2,3</sub>
K $\beta_3$	K-M <sub>2</sub>	L $\beta_4$	L <sub>1</sub> -M <sub>2</sub>	L $\gamma_5$	L <sub>2</sub> -N <sub>1</sub>		
K $\beta_4^I$	K-N <sub>5</sub>	L $\beta_5$	L <sub>3</sub> -O <sub>4,5</sub>	L $\gamma_6$	L <sub>2</sub> -O <sub>4</sub>		
K $\beta_4^{II}$	K-N <sub>4</sub>	L $\beta_6$	L <sub>3</sub> -N <sub>1</sub>	L $\gamma_8$	L <sub>2</sub> -O <sub>1</sub>		
K $\beta_{4x}$	K-N <sub>4</sub>	L $\beta_7$	L <sub>3</sub> -O <sub>1</sub>	L $\gamma_8$	L <sub>2</sub> -N <sub>6(7)</sub>		
K $\beta_5^I$	K-M <sub>5</sub>	L $\beta_7$	L <sub>3</sub> -N <sub>6,7</sub>	L $\eta$	L <sub>2</sub> -M <sub>1</sub>		
K $\beta_5^{II}$	K-M <sub>4</sub>	L $\beta_9$	L <sub>1</sub> -M <sub>5</sub>	Li	L <sub>3</sub> -M <sub>1</sub>		
		L $\beta_{10}$	L <sub>1</sub> -M <sub>4</sub>	Ls	L <sub>3</sub> -M <sub>3</sub>		
		L $\beta_{15}$	L <sub>3</sub> -N <sub>4</sub>	Lt	L <sub>3</sub> -M <sub>2</sub>		
		L $\beta_{17}$	L <sub>2</sub> -M <sub>3</sub>	Lu	L <sub>3</sub> -N <sub>6,7</sub>		
				Lv	L <sub>2</sub> -N <sub>6(7)</sub>		

Reproduced with permission from IUPAC (1991) Nomenclature system for X-ray spectroscopy. *Pure and Applied Chemistry* 63: 735–746.

was incorporated in the laws propounded by Moseley at the beginning of the second decade of the twentieth century and was used to identify gaps in the periodic table corresponding to the existence of a number of undiscovered elements, including technetium, promethium, and rhenium. One consequence of this systematic variation is that the elemental interpretation of XRF spectra is particularly simple and the prediction of line overlap interferences relatively straightforward, especially when compared with the more complex spectra observed in optical emission techniques.

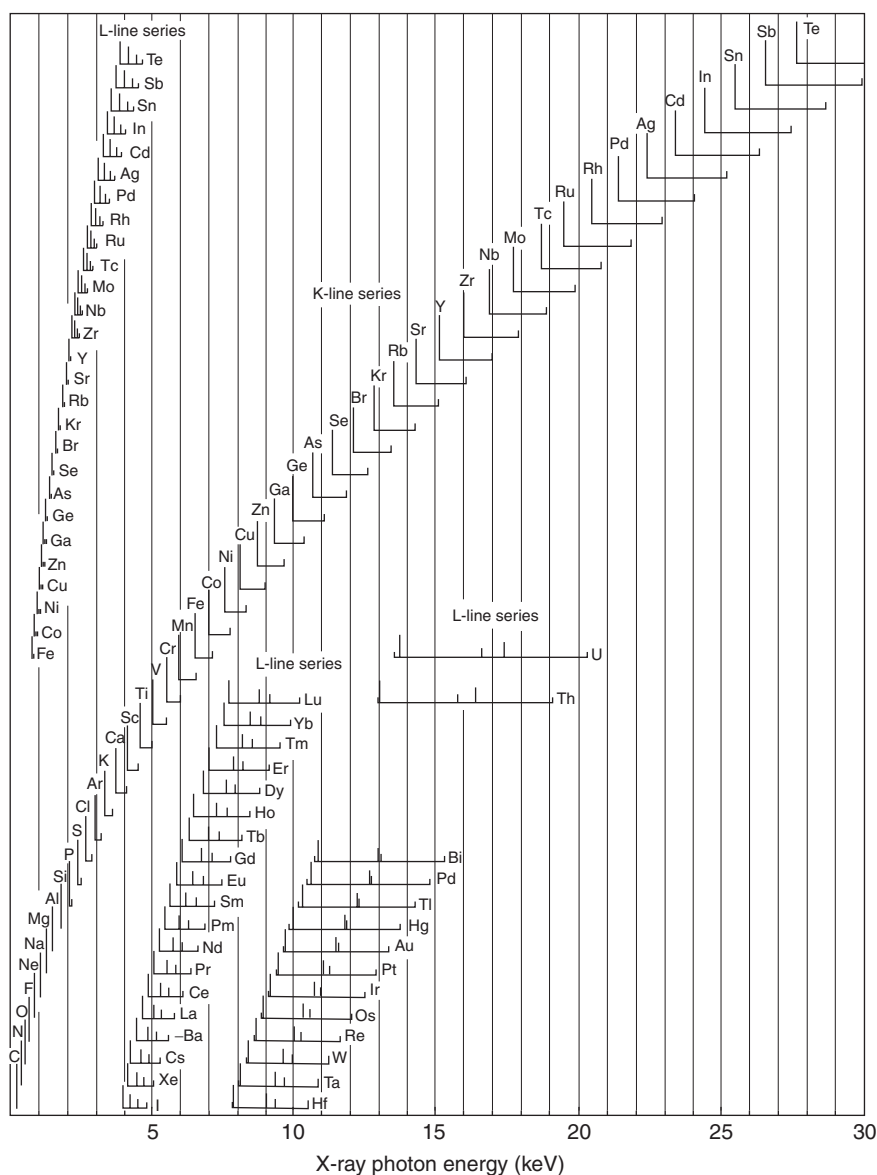
### Fluorescence Yield

The probability that a fluorescence X-ray will be emitted following ionization of an electron from a designated shell is quantified by a parameter called the fluorescence yield ( $\omega$ ). Values of the fluorescence yield (being the fraction of atoms that decay by yielding a fluorescence X-ray after ionization) for selected elements are listed in **Table 2**. The following general trends may be observed in these data. (1) Within a line series, fluorescence yields increase with atomic number. (2) The fluorescence yield of the line series of an individual element decreases in the order  $\omega_K > \omega_L > \omega_M$ , implying that, other factors being equal, the most sensitive determinations will be made from K-line emissions. (3) The fact that fluorescence yields are less than unity implies the existence of a competing deexcitation route for ionized atoms, which becomes increasingly important for the lighter elements. This competing route is the emission of Auger electrons, whereby the fluorescence X-ray is captured internally before it has escaped from the atom, causing the emission of secondary electrons

from outer orbitals (**Figure 2**). This signal forms the basis of a separate analytical technique (Auger spectroscopy) but in the context of XRF serves to reduce the sensitivity of fluorescence measurements.

### Line Selection for Analysis

In analytical applications, the most intense fluorescence line is selected normally for the determination of an element (to maximize sensitivity). Without further consideration, therefore, it might be expected that the appropriate K-L<sub>2,3</sub> (K $\alpha$ ) line would be used for all analytical measurements. However, account must also be taken of the variation in excitation efficiency and detection response of XRF instrumentation as a function of energy. In conventional XRF spectrometers, the upper limit for sensitive XRF measurements is between 20 and 30 keV, this limit being set by the maximum potential that can be applied to the X-ray tube. The lower limit for routine applications is  $\sim 1$  keV, this being set by the substantial increase in attenuation of X-ray photons of lower energy in the windows of X-ray detectors, noting that measurements at reduced sensitivity can be made down to  $\sim 0.2$  keV using more specialized spectrometer components. These instrumental limitations, therefore, dictate that the line selection for the determination of the elements between sodium and molybdenum would normally be the K-L<sub>2,3</sub> (K $\alpha$ ) line and for the elements barium to uranium, the L<sub>3</sub>-M<sub>4,5</sub> (L $\alpha$ ) line. A choice of K or L lines is available for the intermediate elements (ruthenium to barium), depending on the application and/or instrumentation. M lines would not be used normally for



**Figure 3** Systematic variation in the energy of K and L fluorescence emission lines with atomic number in the region 0–30 keV. The main sequence of K-lines for the elements carbon to tellurium is plotted, together with associated L-line emissions (iron–tellurium). Additional L-line emissions found in this region of the spectrum are plotted for convenience in two groups, iodine–lutetium and hafnium–uranium. Line heights correspond approximately to the relative intensities of emission lines, and for clarity, minor L-line emissions have been omitted.

analytical measurements using XRF because of potential spectral overlap problems, although they are used for the heaviest elements in electron microprobe applications.

## Mass Attenuation Effects

To be capable of inducing the ionization of electrons from a particular atomic orbital, X-ray photons from an excitation source must possess energy greater than a threshold value (the ionization energy of the line

series). X-ray photons below this energy threshold are incapable of exciting fluorescence for the line series. This property may be illustrated by examining the degree of attenuation of a monochromatic beam of X-rays as it traverses through a thin foil of an element (**Figure 4**). The degree of attenuation is given by the well known Beer's law,

$$I = I_0 \exp(-\mu x)$$

where  $I$  is the intensity of the transmitted beam,  $I_0$  the intensity of the incident beam,  $x$  the thickness of

**Table 2** Values of fluorescence yield for selected elements

Atomic no.	Element	$\omega_K$	$\omega_L$	$\omega_M$
6	C	0.001	—	—
11	Na	0.013	—	—
14	Si	0.036	—	—
20	Ca	0.142	0.001	—
26	Fe	0.324	0.003	—
30	Zn	0.458	0.007	—
37	Rb	0.653	0.021	0.001
41	Nb	0.732	0.035	0.001
45	Rh	0.792	0.052	0.001
50	Sn	0.845	0.081	0.002
57	La	0.893	0.135	0.004
74	W	0.944	0.304	0.018
78	Pt	0.949	0.345	0.024
82	Pb	0.954	0.386	0.032
92	U	0.960	0.478	0.064

Data abstracted from Colby JW (1968) *Advances in X-ray Analysis* 11: 287–305. The K-lines of the elements lanthanum–uranium are observed at energies above the normal analytical range limit of 30 keV.

the foil, and  $\mu$  the linear attenuation coefficient (which varies according to the composition of the foil and as a function of photon energy). In X-ray applications, it is normal to rearrange this expression as follows:

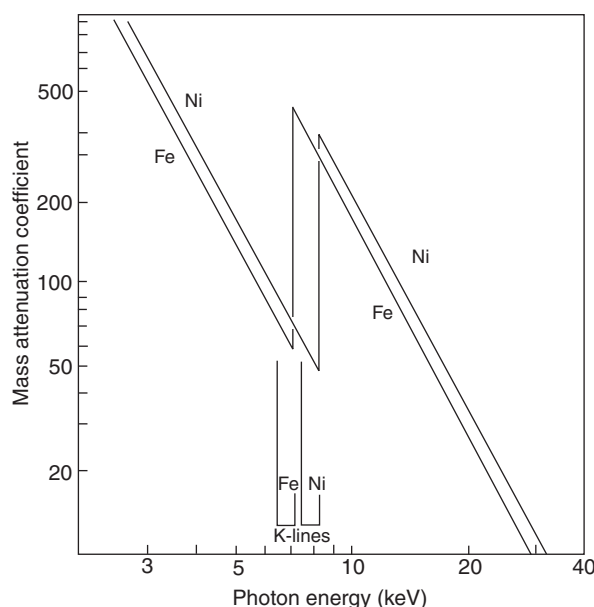
$$I = I_0 \exp ((-\mu/\rho)\rho x)$$

where  $\rho$  is the density and the term  $(\mu/\rho)$  is known as the mass attenuation coefficient. For composite materials, the bulk mass attenuation coefficient ( $\bar{\mu}$ ) may be calculated from the mass fractions of individual elemental components:

$$\bar{\mu} = \sum_i (\mu/p)_i - C_i$$

where  $C_i$  is the weight fraction of element  $i$  and the summation is undertaken over all elements ( $i$ ) present in the sample.

If the attenuation of an X-ray beam  $(\mu/p)$  through a thin foil of a specified element (taking iron and nickel as examples) is plotted as a function of photon energy (Figure 4), the intensity of the transmitted X-ray beam shows a sharp discontinuity called an absorption edge. This absorption edge corresponds to the energy threshold above which XRF phenomena are observed following ionization of the corresponding electron orbital. The energy of the absorption edge varies in a systematic way from one element to another. The form of the graph in Figure 4 also shows that X-ray photons of energy just above the absorption edge are the most efficient at promoting excitation that results in X-ray fluorescence but that this efficiency falls off with increase in photon



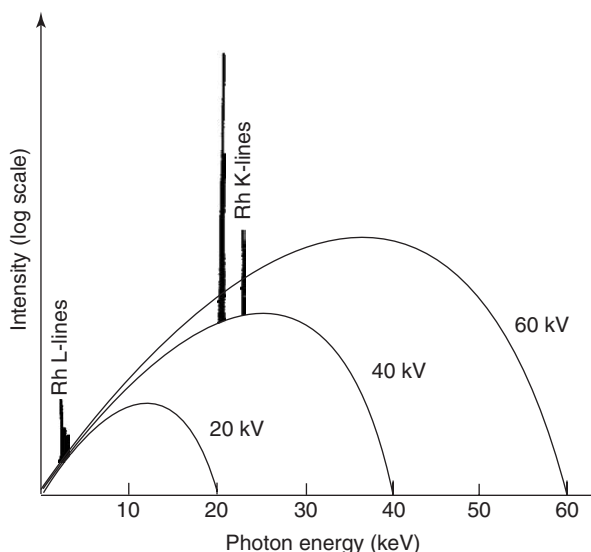
**Figure 4** Variation of mass attenuation coefficient of iron and nickel as a function of photon energy (2–40 keV) showing discontinuities associated with the corresponding K absorption edges. The energies of the corresponding K–L<sub>2,3</sub> (K $\alpha$ ) and K–M<sub>2,3</sub> (K $\beta$ ) fluorescence lines are shown.

energy. This observation contrasts with that relevant to electron and proton excitation, where the excitation function continues to increase with increasing particle energy. An underlying trend can also be seen in Figure 4 in that the attenuation of the X-ray beam generally increases the decreasing photon energy.

In assessing the overall fluorescence emission intensity, account must also be taken of the spectral output of the excitation source. As an illustration of this point, the emission characteristics of a typical X-ray tube are plotted in Figure 5. The tube output comprises two components: characteristic fluorescence lines derived from the tube anode and a broad background continuum component. The excitation efficiency is maximized for elements that have absorption edges just below the energy of the characteristic tube lines, where the maximum tube output intensity is observed.

## Scattering Processes

The excitation mechanism that leads to the emission of fluorescence X-rays is called the photoelectric effect. There are, however, two other important mechanisms of interaction between the primary photon beam and sample that have analytical consequences. The first of these interactions, Rayleigh (or elastic)



**Figure 5** Schematic diagram of spectral output (0–60 keV) from a rhodium anode X-ray tube operated at 20, 40, and 60 kV, showing both the characteristic rhodium fluorescence lines and the continuum components of the tube output.

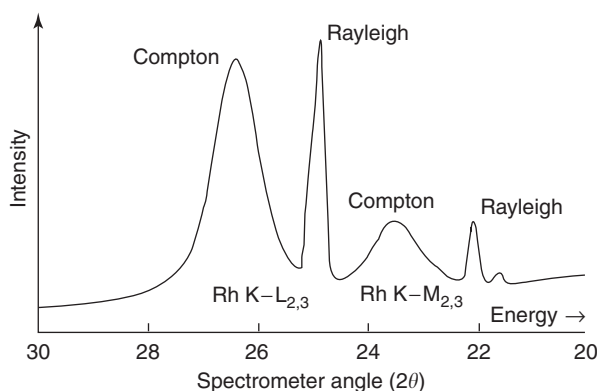
scatter, occurs when the energy of an X-ray photon from the excitation source is absorbed by an atom by interaction with electron orbitals. In Rayleigh scatter, deexcitation results in the emission of an X-ray photon with the same energy as the incident photon but scattered in a randomized direction with respect to the incident photon. The consequence of this phenomenon is that a small fraction of the entire spectrum of the excitation source (**Figure 5**) will be scattered off the sample into the X-ray spectrometer and detected together with the fluorescence X-ray signal. The second important interaction is Compton (or inelastic) scatter. Following excitation of the atom, part of the photon energy is absorbed; the remainder is reirradiated as a ‘Compton scatter’ X-ray photon of reduced energy. The energy ( $E'$ ) (or wavelength,  $\lambda'$ ) of the scattered photon is related to its incident energy ( $E_0$ ) (or wavelength,  $\lambda_0$ ) and the angle ( $\theta$ ) through which scatter occurs by the relationship

$$E' = E_0 / (1 + 0.001957E_0(1 - \cos \theta))$$

or

$$\lambda' - \lambda_0 = 0.0243(1 - \cos \theta)$$

Noting that in an XRF spectrometer the angle  $\theta$  is fixed by the design of the instrument, one analytical consequence of this phenomenon is that a fraction of the entire spectrum of the excitation source will be scattered into the X-ray spectrometer with its energy reduced by a constant proportion (or wavelength increased by a constant value). The most obvious



**Figure 6** Spectral scan recorded from quartz showing the Compton and Rayleigh lines observed by scatter of characteristic rhodium anode X-ray tube lines off the sample. The scan is plotted against decreasing spectrometer  $2\theta$  angle, a parameter that is inversely proportional to increasing X-ray photon energy.

manifestation of these two phenomena can be seen when the X-ray spectrum from a sample, excited by the characteristic X-ray lines from an X-ray tube source, is scanned through the region of these lines. The example in **Figure 6** shows the spectrum observed from a sample excited with characteristic  $K-L_{2,3}$  ( $K\alpha$ ) and  $K-M_{2,3}$  ( $K\beta$ ) radiation from a rhodium anode X-ray tube. The rhodium  $K-L_{2,3}$  ( $K\alpha$ ) and  $K-M_{2,3}$  ( $K\beta$ ) lines are observed in the fluorescence spectrum, caused by Rayleigh scatter of tube radiation. The rhodium  $K-L_{2,3}$  ( $K\alpha$ ) and  $K-M_{2,3}$  ( $K\beta$ ) Compton scatter lines are also observed, shifted to a lower energy compared with the main lines. In addition to the very obvious scatter of the characteristic X-ray tube lines, the background present in the spectrum originates mainly from Compton and Rayleigh scatter of continuum radiation from the X-ray tube.

Two important analytical consequences arise from these scatter phenomena. The first is that scatter results in a background contribution to fluorescence spectra, the magnitude of which dictates ultimately the detection limit of the technique. The second is that the intensity of the Compton scatter peak derived from characteristic tube lines from the X-ray source can, in the appropriate circumstances, be used as a simple and effective means of correcting for matrix effects (see next section).

## Matrix Effects – Interactions in the Sample

In the context of XRF, matrix effects refer specifically to attenuation and enhancement phenomena that influence the intensity of fluorescence X-ray lines observed from the sample. As the magnitude of such



matrix effects varies with elemental composition, the observed fluorescence X-ray intensity is not normally linearly proportional to the concentration of the analyte. A correction must, therefore, be applied to measured intensity data to account for the following:

- (1) Attenuation of the primary X-ray beam from the excitation source as it penetrates into the sample. The intensity and spectral distribution of the X-ray flux available to excite the sample changes with penetration depth.
- (2) Attenuation of fluorescence X-rays as they escape through the sample prior to detection by the X-ray spectrometer.
- (3) Secondary fluorescence or second element effects. In circumstances where the energy of a fluorescence photon (e.g., nickel K–L<sub>2,3</sub> at 7.47 keV) lies immediately above the absorption edge of a second element (e.g., iron K-line edge at 7.11 keV) (Figure 4), the fluorescence intensity of the second element (iron K–L<sub>2,3</sub>) will be enhanced due to preferential excitation by fluorescence radiation (i.e., nickel K–L<sub>2,3</sub>) within the sample. The magnitude of this effect is not always significant except for certain combinations of elements in alloys (e.g., iron–chromium–nickel) and in multilayer thin films.
- (4) Tertiary fluorescence or third-element effects are occasionally significant when secondary fluorescence effects are extended to a third element.

## Critical Penetration Effects

One consequence of the attenuation of fluorescence of X-rays within a sample is that a critical depth below the surface of a sample can be identified, beyond which any emitted photon is essentially totally

absorbed and so will not make a significant contribution to the detected fluorescence spectrum. This critical penetration depth varies as a function of matrix composition as well as of fluorescence photon energy. Values for selected elements in an oxide material (a silicate rock, prepared as a compressed powder pellet) and a metallurgical sample (a steel) are listed in Table 3. These data show that the critical penetration depth of low-energy photons (e.g., the K–L<sub>2,3</sub> (K $\alpha$ ) lines of carbon or sodium) is very small, the analytical results being representative of the surface layers of the sample. Conversely, for more penetrating fluorescence radiation, critical penetration depths can exceed several millimeters. In the direct analysis of some matrix types (e.g., lead in petrol, barium in oil), it is important to ensure that the sample presented for analysis is sufficiently thick to exceed the appropriate critical value; otherwise, the analytical signal will vary as a function of sample thickness as well as of analyte concentration.

## Thin Film Criteria

A complementary way of considering these attenuation phenomena is to identify the limiting thickness of a sample below which absorption/enhancement effects are essentially negligible and so relevant matrix corrections may be ignored. Analysis of such thin films may then be undertaken without applying any matrix corrections. By convention, a 1% attenuation is normally taken as the limit for thin film criteria, and some representative values are also listed in Table 3. Satisfying thin film criteria is particularly important in applications such as direct analysis of dust particles collected on a filter membrane, which then serves directly as the sample substrate for XRF spectrometry.

**Table 3** Critical penetration depth and thin film criteria for selected lines in silicate and steel matrices

Element	X-ray fluorescence K–L <sub>2,3</sub> (K $\alpha$ ) line		X-ray tube	Critical penetration depth ( $\mu\text{m}$ )		Limiting thin film thickness ( $\mu\text{m}$ )	
	keV	$\lambda$ (Å)		Silicate	Steel	Silicate	Steel
C	0.282	44	Cr	–	0.1	–	0.002
Na	1.04	11.9	Cr	4.8	0.4	0.09	0.009
Si	1.74	7.13	Cr	13	1.6	0.2	0.03
Ca	3.69	3.36	Cr	36	9.6	0.7	0.2
Cr	5.41	2.29	Rh	90	30	1.7	0.7
Fe	6.40	1.94	Rh	180	43	3.4	0.9
Rb	13.39	0.927	Rh	900	40	16	0.9
Nb	16.61	0.748	Rh	1400	62	25	1.3
Rh	20.21	0.614	W	3.9 mm	161	72	3.5
La	33.44	0.373	W	10.6 mm	580	190	13
Eu	41.53	0.301	W	15.4 mm	886	280	19

Data for silicates (analyzed as a compressed powder pellet) taken from Potts PJ (ed.) (1987) *A Handbook of Silicate Rock Analysis*. Chs. 8 and 9. Glasgow: Blackie. Data for steel (analyzed as a solid billet) calculated using a similar procedure. 1 Å = 0.1 nm.

## Matrix Correction Procedures

Since the majority of materials analyzed using XRF are prepared as solid samples (of infinite thickness), matrix correction procedures must be applied to fluorescence measurements. Several approaches are described in this section, their use depending on both the application and the range of elements to be determined.

### Fundamental Parameter Approach

An exact mathematical expression, based on the so-called Sherman equations, may be used to calculate the intensity of fluorescence radiation from a given analyte in a specified matrix using a specified excitation source. This expression is based on a physical model of the excitation process and requires accurate values for the various relevant physical parameters. As originally conceived, the spectral distribution (intensity versus energy) of the excitation source must also be known accurately. This mathematical model accounts for both the attenuation of X-ray photons from the excitation source that penetrate into the sample and the attenuation of fluorescence X-ray photons as they traverse out of the sample prior to detection. A subsidiary correction accounts for second-element enhancement effects. All these terms are dependent not only on photon energy but also on the sample composition, which is not known prior to an analysis. Accordingly, the correction must be applied in an iterative manner. Uncorrected concentrations are first used to estimate the matrix correction factors and so apply a first approximation of the matrix effect. Having applied this correction to analytical data, partially corrected concentration values are used to improve the accuracy of the correction during the next cycle of calculations. The correction is programmed to stop after typically five or six iterations, when differences in corrected values between two successive cycles are statistically insignificant. Early application of this model was restricted by the lack of adequate computational power, some uncertainty in the accuracy of fundamental parameters, and limitations in accurate modeling of the spectral distribution emitted by an X-ray tube. However, these difficulties have largely now been overcome. One advantage of the fundamental parameter approach is the option to quantify results without the necessity for making measurements on calibration samples. However, although valuable, in the semiquantitative analysis of unknown sample types, most quantitative applications would use an adaptation of fundamental parameter procedures (described below) in which a number of standard samples are used to calibrate for the major element components of the matrix.

### Empirical Correction Models

Empirical matrix correction procedures do not attempt to model the physical processes of excitation and attenuation of fluorescence X-rays. Rather, the assumption is made that the intensity of fluorescence emission is influenced by the presence of all elements present in the sample and that the magnitude of this effect can be described by the appropriate parameters called influence coefficients. A family of empirical correction models has been developed over the years, each differing slightly in precise formulation and application. However, these correction models may be divided generally into two categories, those based on fluorescence intensities and those based on concentration.

One of the earliest intensity-based models to gain widespread acceptance was that of Lucas-Tooth and Price (which was subsequently amended by Lucas-Tooth and Pyne to reduce the influence of instrumental drift) and has the form

$$C_i = (a_0 + a_i I_i) - (1 + \sum_{ij} \alpha_{ij} I_j)$$

where  $C_i$  is the weight fraction of element  $i$  of measured intensity  $I_i$ ,  $a_0$  and  $a_i$  are the linear calibration constants (the term  $a_0 + a_i I_i$  is, therefore, the apparent concentration),  $\alpha_{ij}$  is the parameter that describes the influence of element  $j$ , of measured intensity  $I_j$ , on element  $i$ , and the summation is performed over all elements,  $j \neq i$ .

This form of correction procedure is still widely used and offers the advantage that it is based on the intensity of matrix elements, a knowledge of concentration data not being required. Its application is usually best restricted to samples that vary over a limited concentration range. An alternative approach, based on a concentration model, is that of Lachance and Traill, who related the concentration (weight fraction) of element  $i$  ( $C_i$ ), to the concentrations of all other elements present ( $C_j$ ,  $C_k$ , ...) through the empirical expression

$$C_i = (a_0 + a_i I_i)(1 + \sum_j \alpha_{ij} C_j)$$

where  $\alpha_{ij}$  is the influence (or  $\alpha$ ) coefficient that describes the effect of element  $j$ , of concentration  $C_j$ , on element  $i$  ( $\alpha_{ij}$  is assumed to be a constant for samples of a particular matrix type prepared in the same form for analysis). This correction must be applied in an iterative manner, and concentration data for all matrix elements must be known.

Correction procedures of this type offer the considerable advantage that they are not limited by uncertainties in the magnitude of fundamental parameters, nor is there any requirement to model

the X-ray source intensity characteristics. However,  $\alpha$  coefficients may only be calculated successfully from measurements made from a large number of samples of accurately known composition that cover an adequate concentration range. If  $n$  elements are to be determined, a minimum of  $2n$  calibration samples is considered desirable to avoid 'trivial' solutions to the resultant set of simultaneous equations. Although the procedures of Lucas-Tooth and Pyne and of Lachance–Traill are two of the better known empirical correction models, other forms of equation have been proposed including those of Rasberry–Heinrich, Claisse–Quentin, and de Jongh.

### Hybrid Correction Models

To offer maximum analytical flexibility, correction models available on modem systems often combine the best characteristics of both fundamental parameter and empirical correction models. The basis for such 'hybrid' correction models (sometimes known as fundamental coefficient or theoretical  $\alpha$  correction models) is that the empirical  $\alpha$  coefficients used in concentration correction models such as those of Lachance–Traill or Claisse–Quentin can, in fact, be calculated from fundamental parameters. A typical approach, therefore, would be to make measurements on well-characterized calibration samples to determine the  $\alpha$  coefficients of the more important major elements (i.e., elements present in the wt% range, which will have the greatest influence on the overall matrix effect). The  $\alpha$  coefficients for minor and trace elements may then be calculated conveniently from fundamental parameters. In this way, accurate and practical matrix corrections may be derived from fewer calibration standards. In theory, it is not necessary to make analytical measurements on all major elements, although to achieve results of the highest accuracy, it would be unusual if this were not undertaken.

### Compton Scatter

The Compton scatter matrix correction is based on the observation, referred to above, that the intensity of the Compton scatter peak is inversely proportional to the bulk matrix attenuation correction factor. Matrix corrections may then be applied by simply normalizing all fluorescence measurements from a sample to the intensity of the Compton scatter line derived from one of the characteristic fluorescence lines from the X-ray source. This procedure is, however, subject to an important restriction. Corrections are only valid providing no significant absorption edge intervenes between the energy of the Compton scatter peak and the fluorescence line

energy. An extension to lower-energy fluorescence lines can be achieved, providing the matrix characteristics of this and any additional absorption edges are applied as part of the correction, possibly using a supplementary fundamental parameter calculation. In practice, the Compton scatter correction is used widely in the determination of the heavier trace elements, particularly in light matrices. Under these circumstances, it is not necessary to analyze any matrix elements, a factor that may be an additional advantage of this correction procedure in some applications.

## Sample Preparation Requirements

### Forms of Sample Presentation

To exploit the quantitative capabilities of XRF spectrometry, samples must normally be placed in a sample cup designed so that the lower surface is flat and can be mounted accurately and reproducibly in the analysis position, the distance between sample and excitation source being particularly important. The appropriate method of sample preparation depends on the application and the form and nature of the sample. Provided small variations in packing density do not influence the reproducibility of results in relation to fitness-for-purpose requirements, granules and powdered materials, as well as oils, bones, and liquors of high salt content, may be analyzed directly after pouring into a suitable sample cup. The lower surface of this cup must be formed from a thin membrane (e.g., 6  $\mu\text{m}$  Mylar) that does not attenuate fluorescence X-rays of interest to an excessive extent. However, it should be noted that samples cannot be analyzed with the spectrometer in its normal state (i.e., under vacuum) – as referred to above, a helium back-flush facility must be used instead. Dust from air samples collected on membrane filters may be mounted in a suitable holder and excited directly. Metallurgical samples can be cast in the form of a cylindrical billet, one surface of which is then ground flat for analysis.

However, for many sample types, including ceramics, refractories, rocks, ores, soils, etc., results of the highest precision can only be obtained after crushing and grinding (to obtain a representative sample) and preparing samples for analysis in a reproducible form as either compressed powder pellets or glass disks.

### Compressed Powder Pellets

Compressed powder pellets represent a simple, rapid, and cheap sample preparation scheme, which is appropriate for a variety of applications involving

analysis of powdered materials. The sample powder (7–10 g) is mixed with a binder, placed in a suitable die, and compressed to form a flat cylindrical disk, which after drying may be analyzed directly. A range of binders and/or backing materials may be used, but one of the more successful ones is a solution of polyvinylpyrrolidone–methyl cellulose. Several drops are mixed with the sample powder to form a ‘dry’ slurry. After compressing, the powder pellet is dried in an oven to remove the solvent, leaving a minimal mass of binder in the sample to be analyzed. This form of sample preparation is appropriate for a range of applications, particularly in the determination of heavier trace elements in oxide or silicate matrices. However, owing to limitations caused by particle size effects in relation to critical penetration depth, powder pellets are not normally suitable for determination of elements from lower-energy fluorescence lines, in particular elements lower in atomic number than potassium or calcium in the periodic table.

### Glass Disks

The sample preparation scheme described above is not satisfactory if (1) differences in the size, shape, or composition of individual particles present in the powdered material cause variations in the attenuation of low-energy fluorescence X-rays (as in the determination of sodium, magnesium, aluminum, and silicon in geological samples) or (b) ‘absolute’ calibration procedures, that is, calibrations based directly on the measurement of high-purity chemical reagents, must be used to satisfy a standard scheme of analysis in the characterization of ores, refractories, and industrial minerals. Under these circumstances the mineralogical structure of the sample must be broken down by fusing 0.1–0.5 g of sample with a suitable flux and pouring the melt into a preheated mould and cooling (or alternatively squashing flat after pouring onto a suitable platten) to form a flat glass disk. The choice of flux depends on the nature of the sample to be analyzed. Many fluxes are based on lithium metaborate and/or lithium tetraborate since this flux has the capability of dissolving a range of oxide and aluminosilicate materials when fused at 1000–1100°C in nonwetting platinum (5%)–gold crucibles. A selection of fluxes and dilution factors appropriate to a range of different materials are listed in Table 4. The particular advantages of this form of sample preparation are that the surface layers of the glass disk are representative of the bulk material, so facilitating accurate determination of elements such as sodium from fluorescence X-rays with short critical penetration depths (see Table 3). Furthermore, it is simple to prepare

**Table 4** Fluxes available for the preparation of glass fusion disks from various sample types

Sample type	Flux composition	Flux: sample ratio (by wt)
Aluminosilicates, calcium silicates (e.g., cement, anorthosite), bone china, phosphate rock, apatite, calcium sulfate, fluorspar, zircon, glasses, glazes, frits	4 LiBO <sub>2</sub> : 1 Li <sub>2</sub> B <sub>4</sub> O <sub>7</sub>	5:1
Chrome ores, chrome magnesite refractories (10–50% Cr <sub>2</sub> O <sub>3</sub> )	5 Li <sub>2</sub> B <sub>4</sub> O <sub>7</sub> : 4 LiBO <sub>2</sub>	22.5:1
Dolomite, limestone	Li <sub>2</sub> B <sub>4</sub> O <sub>7</sub>	5:1
Magnesia, magnesite	Li <sub>2</sub> B <sub>4</sub> O <sub>7</sub>	10:1
Titania (rutile, ilmenite, stabilized zirconium), iron oxides, manganese oxide	Li <sub>2</sub> B <sub>4</sub> O <sub>7</sub>	12:1
Silicon carbide, silicon nitride, silicon boride, elemental silica	Li <sub>2</sub> B <sub>4</sub> O <sub>7</sub> :Li <sub>2</sub> CO <sub>3</sub>	5:1 to 12:1 (various ratios)

Abstracted from Bennett H and Oliver GJ (1992) *XRF Analysis of Ceramics, Minerals and Allied Materials*. Chichester: Wiley.  
LiBO<sub>2</sub>, lithium metaborate; Li<sub>2</sub>B<sub>4</sub>O<sub>7</sub>, lithium tetraborate; Li<sub>2</sub>CO<sub>3</sub>, lithium carbonate.

‘standard’ disks for calibration by fusing appropriate proportions of chemical reagents to match the composition range of samples to be analyzed. An additional advantage is that fusion techniques are effective in producing a homogeneous glass disk from samples containing refractory mineral grains (e.g., zircon) that are difficult to dissolve using acid-based dissolution procedures used in some competitive atomic spectrometry techniques.

See also: **X-Ray Fluorescence and Emission: Wavelength Dispersive X-Ray Fluorescence; Energy Dispersive X-Ray Fluorescence.**

### Further Reading

- Ahmedali S T (ed.) (1989) *X-Ray Fluorescence Analysis in the Geological Sciences, Advances in Methodology*, Geological Association of Canada Short Course, vol. 7.
- Bertin EP (1970) *Principles and Practice of X-Ray Spectrometric Analysis*. New York: Plenum Press.
- Jenkins R (1988) *X-Ray Fluorescence Spectrometry*. Chichester: Wiley.
- Potts PJ (ed.) (1987) *A Handbook of Silicate Rock Analysis*. Chs. 8 and 9. Glasgow: Blackie.
- Terdan R and Claisse P (1982) *Principles of Quantitative X-Ray Fluorescence Analysis*. London: Heyden.
- Van Grieken RE and Markowicz AA (eds.) (1992) *Handbook of X-Ray Spectrometry*. New York: Dekker.

## Wavelength Dispersive X-Ray Fluorescence

P J Potts, The Open University, Milton Keynes, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

The analytical capabilities of X-ray fluorescence (XRF) have been recognized since the work of Moseley, who, in 1913 and 1914, published papers describing the systematic variation in fluorescence wavelength with atomic number and used this observation to characterize the elemental composition of brass. However, the performance of early instrumentation was restricted by various technical difficulties, and it was not until 1948 that the prototype of the first commercial wavelength dispersive (WD)-XRF instrument was described by Friedman and Birks. Commercial instrumentation became available during the mid-1950s and was progressively adopted by laboratories involved in elemental analysis of the range of sample types described in the preceding article. WD-XRF instrumentation offered significant advantages in terms of sensitivity, precision, and speed of analysis in the determination of a wide range of elements compared with competitive techniques (then wet-chemical methods and arc/spark source optical emission spectrometry). Despite the more recent introduction of alternative instrumental techniques of high analytical sensitivity and productivity, WD-XRF spectrometry has withstood the test of time and remains the technique of choice in many applications where its distinctive analytical characteristics can be exploited.

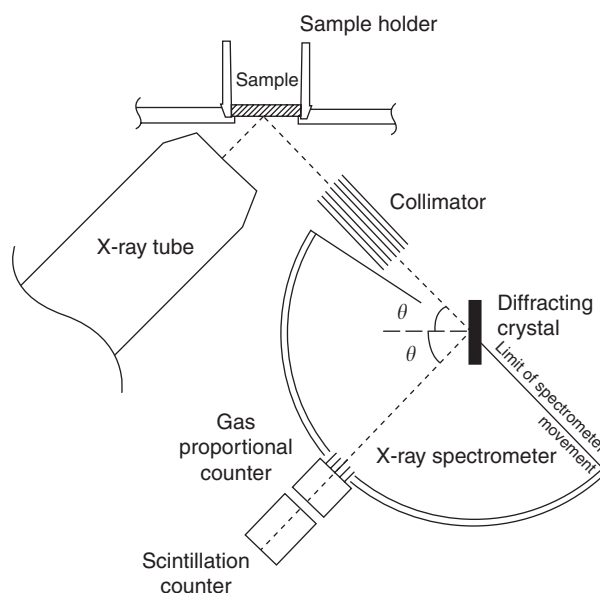
### Categories of Instrumentation

Two principal categories of WD-XRF instrument have been developed. Sequential instruments, shown schematically in **Figure 1**, are designed with one (or more) WD spectrometers. Multielement analysis of samples is undertaken by driving the spectrometer to a preselected set of peak and background angles in a programmed sequence and recording the X-ray count rate at each rest position for a specified count time. Depending on application, a single analysis for 15–20 elements, involving 30–40 separate spectrometer movements may take typically 20 min. If analytical requirements change, the analyst can re-program the instrument to determine an alternative set of elements, so offering maximum analytical flexibility in laboratories involved in research and development applications.

A second category of instrument is the simultaneous WD-XRF analyzer. This instrument is designed with a specified number (usually in the range 8–20) of individual fixed spectrometer ‘channels’, each optimized for the determination of a designated element. Multielement analysis is undertaken very rapidly by acquiring data simultaneously from all channels. Such instruments offer, therefore, high analytical productivity of benefit in production and quality control laboratories that are required to analyze the same sample-type on a routine basis.

Additional flexibility can be obtained by installing a movable goniometer channel (normally substituted for one or two fixed channels), which can be programmed to determine additional elements or scanned to provide qualitative analysis data. The analytical specification of this goniometer is not usually as high as those fitted to sequential instruments.

Elemental determinations by WD-XRF are undertaken by measuring the background-corrected fluorescence intensity of a series of X-ray lines and comparing these data with the appropriate calibration and matrix correction functions. To undertake these measurements, instrumentation comprises: (1) an X-ray tube and accompanying high-voltage generator; (2) a sample holder and mechanism to exchange the sample; (3) a WD X-ray goniometer assembly, including entrance slits, diffracting crystal,



**Figure 1** Schematic diagram of a WD-XRF analyzer showing X-ray tube, sample, and sample holder and components of the X-ray spectrometer.

and X-ray proportional counters; (4) signal-processing and acquisition electronics; and (5) computerized control for instrument operation and data processing. The principal analytical characteristics of these components will now be considered.

## X-Ray Tube

The X-ray tube (Figure 2) is the source of high-energy X-ray photons used to excite characteristic fluorescence radiation in the sample. The X-ray flux is generated when a beam of electrons, emitted from an incandescent filament, is accelerated through a large potential difference (normally in the range 20–100 kV) and directed at a metallic anode. Interactions between the electron and anode result in ionization of inner orbital electrons from atoms of the anode, resulting in the emission characteristic fluorescence X-rays. This phenomenon is very similar to the X-ray fluorescent excitation mechanism, except that atoms are excited directly by energetic electrons, rather than by X-ray photons. Interaction of the electron beam with atoms of the anode results in the emission of a second category of X-ray radiation, a bremsstrahlung or continuous X-ray emission. This emission results from the deceleration effect that occurs when electrons are slowed by the electric fields associated with atoms of the anode. The spectrum of this continuous radiation extends up to an energy equivalent to the accelerating potential applied to the tube. Spectral output from the X-ray tube consists, therefore, of intense fluorescence lines characteristic of the tube anode material, accompanied by a broad continuum radiation.

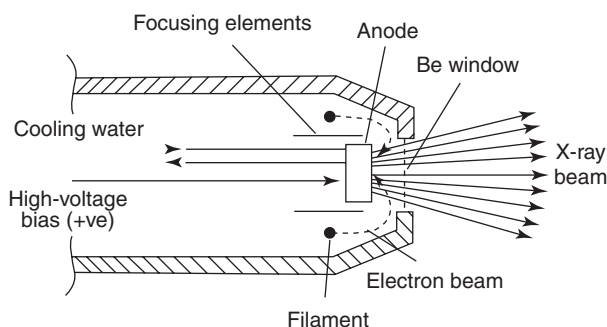
Choice of tube anode influences the analytical capabilities of the instrument, since a specific range of elements that have absorption edges just below the energy of the characteristic tube lines will be excited with maximum sensitivity. In several applications (including geological), the rhodium (Rh) anode tube is considered to be a 'universal' tube because the Rh

K lines efficiently excite a range of heavier trace elements, whereas the Rh L lines are available to boost the excitation of the lightest elements. However, in other applications, molybdenum, chromium, silver, gold, tungsten, scandium, cobalt, and gadolinium anode tubes all offer specific benefits.

To maximize sensitivity, up to 4 kW of power is dissipated by a typical WD-XRF tube. Efficient water cooling is, therefore, essential to prevent damage and premature aging of the tube. Two categories of tube design are available, using either a side- or an end-window configuration. In the side-window design, the filament is held at high negative potential to achieve the required potential difference relative to the anode, at ground potential. This design offers the advantage of being capable of supporting potential differences of up to 100 kV. However, electrons that are backscattered from the anode, impinge onto the beryllium exit window, fitted to maintain high vacuum within the tube. These backscattered electrons heat the window, which must, therefore, be made of a reasonably thick beryllium foil to provide adequate mechanical strength. As a consequence, excitation of the lowest atomic number elements may be reduced because of absorption of the low-energy photon flux within the beryllium window.

In contrast, end-window X-ray tubes (Figure 2) are designed to operate with a reverse bias (i.e., a positive potential) applied to the anode. With this arrangement, the flux of backscattered electrons impinging on the entrance window is reduced significantly, since they are attracted back into the anode, which is held at a high positive potential. Two analytical advantages result. First, the beryllium exit window may be made of thinner foil, reducing the attenuation of low-energy X-ray output with a beneficial effect on the excitation of low atomic number elements. Second, it is easier to design the tube so that it can be positioned closer to the sample, increasing the excitation coupling efficiency. End-window tubes are normally designed to operate at a maximum potential of 60–75 kV (rather than the 100 kV of the side-window design), and are significantly more expensive to construct. A more sophisticated cooling supply is required to prevent breakdown of the high voltage applied to the anode.

The X-ray tube is powered by an appropriate generator capable of being programmed by the operator to apply a potential of typically between 20 and 60–100 kV (depending on tube design) at up to 4 kW. The success of typically modern WD-XRF instrumentation depends in part on the very high degree of stability with which the X-ray output intensity of the tube can be maintained.



**Figure 2** Schematic diagram of an end-window X-ray tube.



## Sample Holder

To achieve high analytical precision, samples must be mounted in the instrument in a highly reproducible manner to avoid discrepancies in excitation efficiency. Because of the small distance between the X-ray tube and the sample, even very small discrepancies in the repositioning height of the sample can have a significant effect on measured fluorescence intensities. Sample cups, 30–50 mm in diameter, are used to hold the analyzed surface in a precise reference plane relative to the X-ray optical path of the instrument (Figure 1). The design of these sample holders varies depending on whether the requirement is to analyze solid disks, loose powders, or liquids, for example. For the latter two categories, the sample must be supported on a thin membrane (e.g., 6  $\mu\text{m}$  Mylar), which minimizes the attenuation of low-energy fluorescence radiation. Samples are normally transferred from an external automatic sample changer into an intermediate chamber, which is evacuated before the sample is finally transferred to the analysis position. During analysis, samples are normally rotated at a few revolutions per second to minimize any variations caused by unevenness in the analyzed surface or inhomogeneity effects. Although the spectrometer chamber would usually be operated under vacuum to avoid attenuation of the lower-energy fluorescence spectrum by the atmosphere, such conditions are incompatible with the analysis of liquids or loose powders, which require the spectrometer to be flushed with helium gas at atmospheric pressure.

## Wavelength Dispersive X-Ray Spectrometer

The WD spectrometer (Figure 1) behaves as an X-ray monochromator. Selected crystalline materials are used to diffract lines in the fluorescence X-ray spectrum. When a beam of X-rays is directed at certain crystalline materials, X-ray photons are reflected off the various atomic layers in the crystal. Destructive interference occurs between almost all reflected photons, except those that satisfy the Bragg equation:

$$n\lambda = 2d \sin \theta$$

where  $n$  is an integer (1 for first-order diffraction, 2 for second-order diffraction, etc.),  $\lambda$  is the wavelength (nm) of the X-ray photon,  $d$  the crystal lattice spacing (nm) of the diffracting medium, and  $\theta$  the angle of incidence subtended at the crystal.

The Bragg crystal spectrometer consists of (1) an entrance slit or collimator, (2) the diffracting element, (3) an X-ray proportional counter, and (4) a

goniometer mechanism to maintain the correct geometric relationship between these components (Figure 1). Most WD-XRF spectrometers are of a semifocusing design and are fitted with parallel-blade collimators and flat diffracting crystals. In the more specialized fully focusing design, a curved crystal is used to diffract the X-ray beam to a line focus at slits placed in front of the proportional counter. Provided certain geometric constraints are satisfied, this second design is more efficient at diffracting fluorescence photons and is often used in fixed channel spectrometers, and electron microprobe instrumentation. These principal components of the X-ray spectrometer are described below.

### Entrance Collimator

As fitted to conventional flat crystal spectrometers, the entrance collimator, comprising a series of parallel blades known as Soller slits, is required to restrict the angular divergence of fluorescence X-rays accepted for diffraction, so avoiding a significant loss in spectrometer resolution. A choice of blade spacings is normally available on modern spectrometers to match the angular dispersion requirements of the diffracting crystal selected for use.

### Diffracting Element

A number of diffracting elements is required to measure the range of fluorescence lines used in most analytical applications (see Table 1 for details). These diffracting elements are selected on the basis of several properties, including: (1) appropriate  $2d$  lattice spacing, given the angular movement that can be achieved on the spectrometer; (2) high reflectivity (i.e., high intensity in the diffracted beam); (3) satisfactory dispersion power and resolution (to give adequate separation between lines in the diffracted spectrum); (4) ability to suppress multiple-order diffractions; (5) physical properties, including low coefficient of thermal expansion (to minimize shifts in the position of diffracted peaks with changes in temperature), stability in air and under vacuum, resistance to moisture, and ability to be attached to a curved substrate (if appropriate) without cracking.

The majority of diffracting elements are crystalline materials (Table 1). However, two other categories of materials have found widespread use in the low-energy (long-wavelength) fluorescence X-ray region (particularly for the elements between boron and magnesium in the periodic table). These are:

(1) *Langmuir–Blodgett ‘pseudocrystals’*. These are usually fabricated from the lead soap compounds (e.g., lead stearate) and are made by coating alternate layers of lead stearate molecules onto a suitable

**Table 1** Details of diffracting crystals commonly used in WD-XRF

Crystal specification		2d spacing (nm)	Typical K-line range	Application and comment
Lithium fluoride	LiF(200)	0.285	K–U	Standard 'heavy-element' crystal
Lithium fluoride	LiF(220)	0.403	V–U	Offers improved resolution but lower intensities than LiF(200)
Lithium fluoride	LiF(420)	0.180	Ni–U	Offers improved resolution but lower intensities than LiF(200) and LiF(220)
Germanium	Ge(111)	0.653	P–Cl	Crystal suppresses second-order interference effects (e.g., from second-order Ca K–L <sub>2,3</sub> on P K–L <sub>2,3</sub> )
Pentaerythritol	PET (002)	0.874	Al–Cl	General-purpose crystal for this range of elements
Indium antimonide	InSb(111)	0.748	Si	Higher intensities in determination of Si than competitive crystals
Thallium hydrogenphthalate	TAP	2.58	O–Mg	General purpose crystal for this range of elements
Layered synthetic microstructure (e.g., W/Si)		5–6	O–Mg	Higher intensities but lower resolution than TAP
Layered synthetic microstructure (e.g., V/C)		11–12	C	Specialized application in the determination of C
Layered synthetic microstructure (e.g., MoB <sub>4</sub> /C)		20–24	B	Specialized application in the determination of B

substrate. In this way, diffracting elements with lattice spacings of typically between 8 and 14 nm can be fabricated. These pseudocrystals are rarely used on modern WD-XRF spectrometers, their role being largely superseded by multilayer dispersion elements.

(2) *Multilayer dispersion elements* (or layered synthetic microstructures). These are artificially manufactured diffraction devices made by depositing alternate layers of a low atomic number element (e.g., silicon, carbon) and a high atomic number element (e.g., tungsten, vanadium, molybdenum, nickel) onto a suitable substrate. The high atomic number layers are designed to reflect the X-ray beam and are separated by the relatively transparent low atomic number layers. By careful control of the deposition process, diffracting elements can be fabricated that have almost any selected 2d spacing, typically in the range 4–22 nm. At the present state of development, these devices show higher reflectivity than alternative diffracting crystals, but with some loss of spectral resolution. Multilayers are normally only capable of diffracting satisfactorily the K lines of one or two adjacent low atomic number elements and so several are required, having optimized 2d values, to cover the full range of low atomic number elements of interest. There is currently considerable interest in the further development of the technology for fabricating these devices.

The more common diffracting elements available and used on modern spectrometers are listed in **Table 1**. Selection for installation on a particular instrument would normally depend on the application (based on

the range of elements to be determined), subject to a maximum of between four and nine diffracting devices that can usually be mounted on the crystal exchange mechanism of a single goniometer.

### X-Ray Proportional Counters

The function of the X-ray counter is to measure the intensity of the diffracted X-ray beam and to provide output pulses that are proportional in height to the energy of the detected photon. This provides a means of discriminating against multiple-order interferences by pulse height analysis. Multiple-order interferences are fluorescence X-rays with 1/2, 1/3, 1/4 of the wavelength of interest in compliance with  $n = 2, 3, 4$  in the Bragg equation. Two types of counter are normally used on WD-XRF spectrometers, gas proportional counters and sodium iodide scintillation detectors.

### Gas Proportional Counters

The argon gas flow proportional counter consists of a cylindrical chamber down the center of which passes a small-diameter wire (the anode), to which a positive potential of between 1 and 2 kV is applied (**Figure 3**). The counter gas, argon, is ionized by incident X-ray photons. Under the influence of fields generated by the applied potential, electrons formed by this ionization process drift toward the wire anode. In the immediate vicinity of the anode, electrons are accelerated by the associated high potential fields, causing further ionization by collision with gas molecules. The resultant avalanche of electrons caused

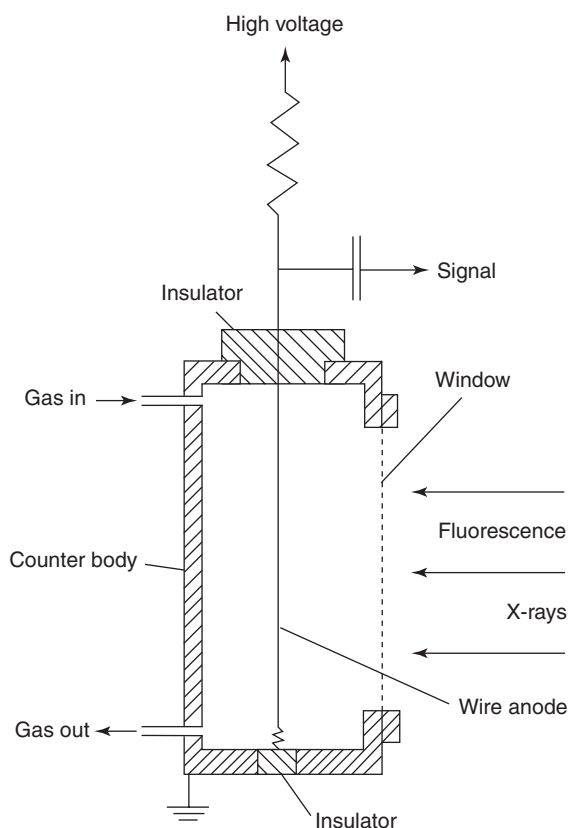
by this 'gas multiplication' effect is discharged at the anode and measured by the counter electronics as an individual X-ray photon signal. Argon ions, also formed by the ionization process, drift to earth at the walls of the counter chamber. To suppress interferences that would be caused by the formation of secondary electrons, generated by the impact of these ions with the wall of the counter, the gas (argon) must contain a small proportion (5–10%) of carbon dioxide or methane as a quench gas. Although related in design to the Geiger–Müller tube, the proportional counter is operated with a lower bias so that

the detected signal is proportional to the number of electrons generated during the primary ionization event (and therefore to the energy of the detected photon). To minimize attenuation of low-energy fluorescence X-rays, a thin membrane (commonly a 1  $\mu\text{m}$  film of polypropylene) is used as the counter entrance window. This window must be strong enough to withstand the spectrometer vacuum against the pressure of the counter gas. However, to compensate for the small rate of diffusion of gas through this window, a continuous flow of argon must be supplied to the counter. Argon flow proportional counters are normally used to detect fluorescence X-rays in the range up to  $\sim 6$  keV. For the higher-energy range, krypton or xenon can be used as counter gas in view of their greater stopping power. In these cases, a thicker window (e.g., 6  $\mu\text{m}$  beryllium foil) can be used, which, being nonporous, allows the counter to be sealed during manufacture ('sealed gas proportional counter'). The applications of these counters are summarized in Table 2.

In simultaneous XRF instruments, individual channels are normally fitted with a gas proportional counter appropriate to the energy of the fluorescence X-ray line to be measured. However, on sequential instruments the argon gas proportional counter is normally paired with a versatile scintillation counter.

### Scintillation Counter

The scintillation counter (Figure 4) is based on a sodium iodide crystal doped with thallium atoms to give it appropriate semiconductor properties ('thallium-drifted sodium iodide scintillation counter'). X-ray photons interact with this crystal, causing it to scintillate, emitting light photons in a band centered at 410 nm in the blue region of the visible spectrum. The detector crystal is optically coupled to the photodiode of a photomultiplier tube. By this means, detected photons are converted into an electronic signal, taking advantage of the large internal amplification factor inherent in the operation of the

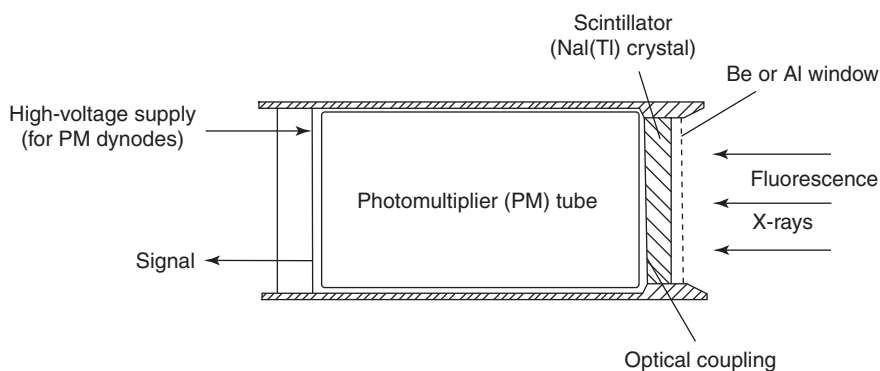


**Figure 3** Schematic diagram of an argon gas flow proportional counter.

**Table 2** Proportional counters used in WD-XRF spectrometers

Photon energy range (keV)	Application (K lines)	Proportional counter details
0.2–0.9	B–F	Argon gas flow counter with ultrathin window supported on a grid (e.g., 0.2 $\mu\text{m}$ collodion)
1–10.5	F–As	Argon gas flow proportional counter with standard window (e.g., 1 $\mu\text{m}$ polypropylene)
3.3–8.0	K–Cu	Krypton sealed counter with 6 $\mu\text{m}$ beryllium window
8.6–18	Zn–Mo	Xenon sealed counter with beryllium window
6–30	Mn–Ba	NaI(Tl) scintillation counter

A combination of the standard argon gas flow proportional counter and the scintillation counter is often used on general-purpose XRF instruments, using tandem operation to compensate for the reduced detection efficiency of each, individually, in the 6–10 keV region.



**Figure 4** Schematic diagram of a scintillation counter.

photomultiplier tube. As with gas counters, the detector is operated so that the magnitude of the photomultiplier signal is proportional to the energy of the detected photon. Scintillation counters are not suitable for the detection of X-ray photons below  $\sim 6$  keV (equivalent to manganese  $K-L_{2,3}$  ( $K\alpha$ )). Conversely, whereas argon gas flow counters show optimum efficiency for lower-energy X-ray photons, they become less efficient in the detection of X-ray photons above the 6–10 keV region (i.e., manganese to arsenic  $K-L_{2,3}$  ( $K\alpha$ )). To optimize detection of fluorescence radiation in this intermediate X-ray region, instruments are normally provided with either (1) an additional krypton sealed proportional counter, which must be mounted on the goniometer independently of the other counters; or (2) a tandem arrangement (Figure 1) of the argon gas flow proportional counter (provided with entrance and exit windows) and scintillation counter. In the intermediate region between  $\sim 6$  and 10 keV, the signal from the gas proportional counter is summed with that from the scintillation counter, the latter deriving its signal from photons that pass straight through the flow counter without interaction.

### Goniometer Mechanism

To satisfy the Bragg equation, a precise angular relationship must be maintained to ensure that for any angle of incidence  $\theta$  at the diffracting crystal, the proportional counter views the diffracted beam through a total angle of diffraction of  $2\theta$ . The simplest way of maintaining this angular relationship is to mount the diffracting element and proportional counters on separate ' $\theta$ ' arms, linked by 2:1 gearing to control their relative movements. However, alternative arrangements have been developed, notably goniometers in which the movement of the diffracting element and proportional counters (each of which may be mounted on individual arms) can be controlled independently. In both cases, great care is taken to ensure that high precision is achieved in the angular

resetting of the goniometer with minimum backlash, and without compromising the speed with which the mechanism can be slewed from one angle to the next.

### Pulse Height Analysis

Interferences arise in wavelength dispersive spectrometers from the diffraction of multiple-order lines (i.e.,  $n = 2, 3$ , etc., in the aforesaid Bragg equation). Although multiple-order diffractions are progressively less intense than those in the first order, some element specific interferences can be troublesome in certain matrices (for example, the fluorine  $K-L_{2,3}$  ( $K\alpha$ ) line suffers a spectrum interference from a third-order diffraction of phosphorus  $K-L_{2,3}$  ( $K\alpha$ )). Furthermore, multiple-order diffraction of higher-energy tube continuum radiation scattered off the sample is likely to contribute to the spectrum background of all analytical measurements. These effects can be minimized by use of pulse height analysis. The signal derived from the X-ray counter (which is proportional to the energy of the detected X-ray photon) is filtered electronically by setting upper and lower discriminator levels, which must normally be optimized by the user. The lower-level discriminator is set to eliminate electronic noise from the counting circuits and the upper-level discriminator set to reject multiple diffraction events that will have a pulse height that is  $\times 2$ ,  $\times 3$ , etc., times the height of the fluorescence X-ray of interest. Only events that pass through this electronic 'window' contribute to the analytical signal. Thus, with the discriminators set to accept pulses corresponding to first-order fluoride  $K-L_{2,3}$  ( $K\alpha$ ) photons, the third-order phosphorus  $K-L_{2,3}$  ( $K\alpha$ ) signal, which will have a pulse height distribution larger by a factor of three, will be effectively suppressed.

### Optimizing Conditions for Analysis

Several factors, often interrelated, are involved in the development of an analytical program. Sample

preparation requirements and choice of X-ray tube in relation to the analytical application have been considered in the preceding entry. More specific features are summarized below.

### Excitation Conditions

A suitable procedure for optimizing the potential applied to the X-ray tube in XRF spectrometry is to analyze a representative sample at a range of kV settings and to choose the setting that maximizes the value of the figure-of-merit ( $\sqrt{I_p} - \sqrt{I_b}$ ) where  $I_p$  is the peak and  $I_b$  the background count rate of a chosen fluorescence line. For multielement determinations, compromise conditions must be selected. It is common practice to select an operating potential of  $\sim 30$ – $40$  kV for the determination of K lines in the region sodium–iron, and between  $60$  and  $100$  kV (depending on the capacity of the generator and design of the X-ray tube) for the determination of the K lines of higher atomic number elements. Tube current is selected to deliver tube power at the maximum (or near-maximum) rating.

### Spectrometer Parameters

Only in unusual circumstances would the fluorescence line selected for analysis not correspond to the most intense line K– $L_{2,3}$  ( $K\alpha$ ) line (or  $L_3$ – $M_{4,5}$  ( $L\alpha$ ) line for higher atomic number elements) available in the analytical region from  $1$  to  $20$  keV. Exceptions to this rule are the occasional necessity to avoid severe spectral overlap interferences. For example, in some applications it would be appropriate to select the lead (Pb)  $L_2$ – $M_4$  ( $L_1\beta_1$ ) rather than the more intense Pb  $L_3$ – $M_5$  ( $L\alpha_1$ ) line to avoid a severe spectral overlap interference from arsenic K– $L_{2,3}$  ( $K\alpha_{1,2}$ ). Other ‘classic’ overlap interferences where similar considerations may apply include titanium K– $L_{2,3}$  ( $K\alpha$ )/barium  $L_3$ – $M_5$  ( $L\alpha_1$ ), and yttrium K– $L_{2,3}$  ( $K\alpha$ )/rubidium K– $M_{2,3}$  ( $K\beta$ ). Accepting these reservations, choice of diffracting crystal, proportional counter, and spectrometer angle are exemplified by those listed in Table 3. Choice of count time would depend on the analytical application and the expected range of concentrations in the samples to be analyzed. A balance must be achieved between adequate precision (or detection limit) and acceptable analysis time. On modern WD-XRF instruments, it would be unusual to select a count time of longer than  $\sim 100$  s (per line), while less than  $10$  s may be sufficient for the most sensitive major elements. One guideline is to select a count time that will result in the accumulation of say  $100\,000$  counts in a measurement.

### Background Corrections and Spectral Interferences

A background correction, necessary in almost all XRF applications, is normally undertaken on sequential WD-XRF instruments by measuring the count rate at a spectrometer angle offset from the peak angle. If the background slopes in this region of the spectrum, measurements on either side of the peak may be necessary, the appropriate value being calculated by interpolation of these data. Background correction may be relatively trivial in the determination of a major element in a high atomic number material (where background scatter off the sample is relatively small), but is significant in the analysis of low atomic number matrices (e.g., oxides and silicates) and is very important in the determination of any trace elements approaching the detection limit level. Background angles must be free of interference from unsuspected fluorescence emissions; those angles selected for one sample type may not all be appropriate for another (Figure 5). More sophisticated approaches are available in which the variation of background intensity is modeled over a range of sample matrices so that curvature of background and mass attenuation effects may be taken into account.

In general, spectral interferences are well characterized in XRF. Where significant overlap occurs, the ratio method is often used to make the appropriate correction. The intensity ratio is measured at the spectrometer angle of the analyte relative to that at an adjacent fluorescence line of the interfering element using a sample that does not contain the analyte. In the analysis of a sample, the magnitude of the interference correction can then be calculated knowing the value of the appropriate interference ratio factor, from measurements of the adjacent fluorescence line of the interfering element.

### Monitoring for Drift

Modern instrumentation offers very high stability so that it is necessary to recalibrate only after extended periods of operation (in the extreme, after replacement of the X-ray tube or components in the spectrometer). However, provision must be made to compensate for small day-to-day variations in instrument response. By normalizing count data to equivalent measurements from one or more drift monitor samples, which are counted periodically during each analytical run, small variations in the slope and offset of the calibration function may be made. The drift monitor samples are normally chosen to be the same samples that were used to normalize count data during the calibration run and will

**Table 3** Typical spectrometer operating parameters and detection limits appropriate for the analysis of silicate rocks by WD-XRF spectrometry. Comparison is made with ED-XRF detection limits, representative of this application

Element data					Typical conditions						Typical 3σ detection limits (element μg g <sup>−1</sup> ) <sup>c</sup>				
Z	Element	Line	Energy (keV)	Wavelength (Å)	Tube anode	kV	Crystal	Counter <sup>a</sup>	Collimator <sup>b</sup>	Normal concentrations in rocks (μg g <sup>−1</sup> )	WD (50s) glass 1:5	WD (50s) glass 1:2	WD (50s) pellet	ED(500s) glass 1:6	ED (800s) pellet
9	F	K <sub>α</sub>	0.677	18.320	Rh, Sc, Cr	40–50	ML <sup>d</sup> , TIAP	fpc	C	10–1000					
11	Na	K <sub>α</sub>	1.041	11.910	Rh, Sc, Cr	40–50	ML <sup>d</sup> , TIAP	fpc	C	0.02–7 oxide%		334		7122	
12	Mg	K <sub>α</sub>	1.253	9.890	Rh, Sc, Cr	40–50	ML <sup>d</sup> , TIAP	fpc	C	0.05–45 oxide%		67		1990	
13	Al	K <sub>α</sub>	1.486	8.340	Rh, Sc, Cr	40–50	PET	fpc	C	0.3–20 oxide%		21		1006	
14	Si	K <sub>α</sub>	1.739	7.126	Rh, Sc, Cr	40–50	PET	fpc	C	35–80 oxide%		18		982	
15	P	K <sub>α</sub>	2.013	6.158	Rh, Sc, Cr	40–50	Ge, PET	fpc	C	0.01–1 oxide%		14		218	
16	S	K <sub>α</sub>	2.307	5.373	Rh, Sc, Cr	40–50	Ge, PET	fpc	C	10–1000					
17	Cl	K <sub>α</sub>	2.621	4.729	Rh, Sc, Cr	40–50	Ge, PET	fpc	C	10–500					
19	K	K <sub>α</sub>	3.312	3.742	Rh, Sc, Cr	40–50	PET, LiF(200)	fpc	F	0.01–15 oxide%		4		332	
20	Ca	K <sub>α</sub>	3.690	3.359	Rh, Sc, Cr	40–50	LiF(200)	fpc	F	0.1–25 oxide%		3		179	
21	Sc	K <sub>α</sub>	4.088	3.032	Rh, Cr, W	40–50	LiF(200)	fpc	F	0.1–50	5.5		4		
22	Ti	K <sub>α</sub>	4.508	2.750	Rh, Cr, W	40–50	LiF(200)	fpc	F	0.01–3 oxide%		1		180	
23	V	K <sub>α</sub>	4.949	2.504	Rh, Cr, W, Au	40–50	LiF(200, 220)	fpc	F	2–600	2.1	9	3.8	125	
24	Cr	K <sub>α</sub>	5.411	2.291	Rh, W, Au	40–50	LiF(200)	fpc	F	3–3000	4.1		3.5	135	
25	Mn	K <sub>α</sub>	5.894	2.103	Rh, Cr, W, Au	60–100	LiF(200, 220)	fpc, sc, st	F	0.01–25 oxide%		21		116	
26	Fe	K <sub>α</sub>	6.398	1.937	Rh, Cr, W, Au	60–100	LiF(200)	fpc, sc, st	F	0.5–15 oxide%		4		175	
27	Co	K <sub>α</sub>	6.924	1.790	Rh, Cr, W, Au	60–100	LiF(200, 220)	fpc, sc, st	F	0.5–200				2.3	
28	Ni	K <sub>α</sub>	7.471	1.659	Rh, Mo, W, Au	60–100	LiF(200)	fpc, sc, st	F	2–3000	3.4	2	1.5	300	12
29	Cu	K <sub>α</sub>	8.040	1.542	Rh, Mo, W, Au	60–100	LiF(200)	fpc, sc, st	F	0.5–150	3.7	2	1.2		8
30	Zn	K <sub>α</sub>	8.630	1.436	Rh, Mo, W, Au	60–100	LiF(200)	fpc, sc, st	F	10–200	2.8	4	1.2		7
31	Ga	K <sub>α</sub>	9.241	1.341	Rh, Mo, Au	60–100	LiF(200)	fpc, sc, st	F	0.5–40	3.3		1.5		6
33	As	K <sub>α1</sub>	10.542	1.176	Rh, Mo, Au	60–100	LiF(200)	St	F	0.1–20	6.6 (Kβ)		1.5		
37	Rb	K <sub>α1</sub>	13.393	0.926	Rh, Mo	60–100	LiF(200, 220)	St	F	0.1–1000	1.6	3	1.6		3
38	Sr	K <sub>α1</sub>	14.163	0.875	Rh, Mo	60–100	LiF(200, 220)	st	F	2–1500	1.7	3	1.8		3
39	Y	K <sub>α1</sub>	14.956	0.829	Rh, Mo	60–100	LiF(200, 220)	st	F	1–100	1.7	3	1.2		3
40	Zr	K <sub>α1</sub>	15.772	0.786	Rh, Mo	60–100	LiF(200, 220)	st	F	1–1000	1.6	4	1.3		3
41	Nb	K <sub>α1</sub>	16.612	0.746	Rh, W	60–100	LiF(200, 220)	st	F	1–100	1.1		1.4		3
42	Mo	K <sub>α1</sub>	17.476	0.709	Rh, W	60–100	LiF(200, 220)	st	F	0.1–5			1.5		
56	Ba	Lβ <sub>1</sub>	4.827	2.658	Rh, W, Au	40–50	LiF(200, 220)	fpc	F	0–2000	20.4	6	17.4		
57	La	Lα <sub>1</sub>	4.650	2.666	Rh, W, Au	40–50	LiF(200, 220)	fpc	F	0.1–100	18.9		8.3		
58	Ce	Lβ <sub>1</sub>	5.261	2.356	Rh, W, Au	40–50	LiF(200, 220)	fpc	F	0.1–200	21.1		5.9		
60	Nd	Lα <sub>1</sub>	5.229	2.370	Rh, W, Au	40–50	LiF(200, 220)	fpc	F	0.05–100	13.2		5.3		
82	Pb	Lβ <sub>1</sub>	12.612	0.983	Rh, Mo	60–100	LiF(200, 220)	st	F	0.5–80	5.2	9	2.2		7
90	Th	Lα <sub>1</sub>	12.967	0.956	Rh, Mo	60–100	LiF(200, 220)	st	F	0.01–50	4.1		2.4		4
92	U	Lα <sub>1</sub>	13.612	0.911	Rh, Mo	60–100	LiF(200, 220)	st	F	0.005–20	4.0		2.7		6

<sup>a</sup>fpc, flow proportional counter; sc, sealed counter; st, scintillation counter. <sup>b</sup>C, coarse collimator; F, fine collimator. <sup>c</sup>WD glass (1:5): Kuiperes: Philips application report, No. 750. Rh tube; WD glass (1:2): Lee and McConchie (1982); Cr and Mo tubes. <sup>d</sup>ML, multilayer.

WD pellet: Typical values for Rh tube; ED glass (1:6): Potts *et al.* (1984); Ag tube: 10 kV; ED pellet: Webb *et al.* (1990); Ag tube, 45 kV, Ag 0.127 mm filter.

This table is taken from Potts and Webb (1992) and is reproduced with permission of Elsevier Science Publishers B.V.

References quoted in the existing footnotes are:

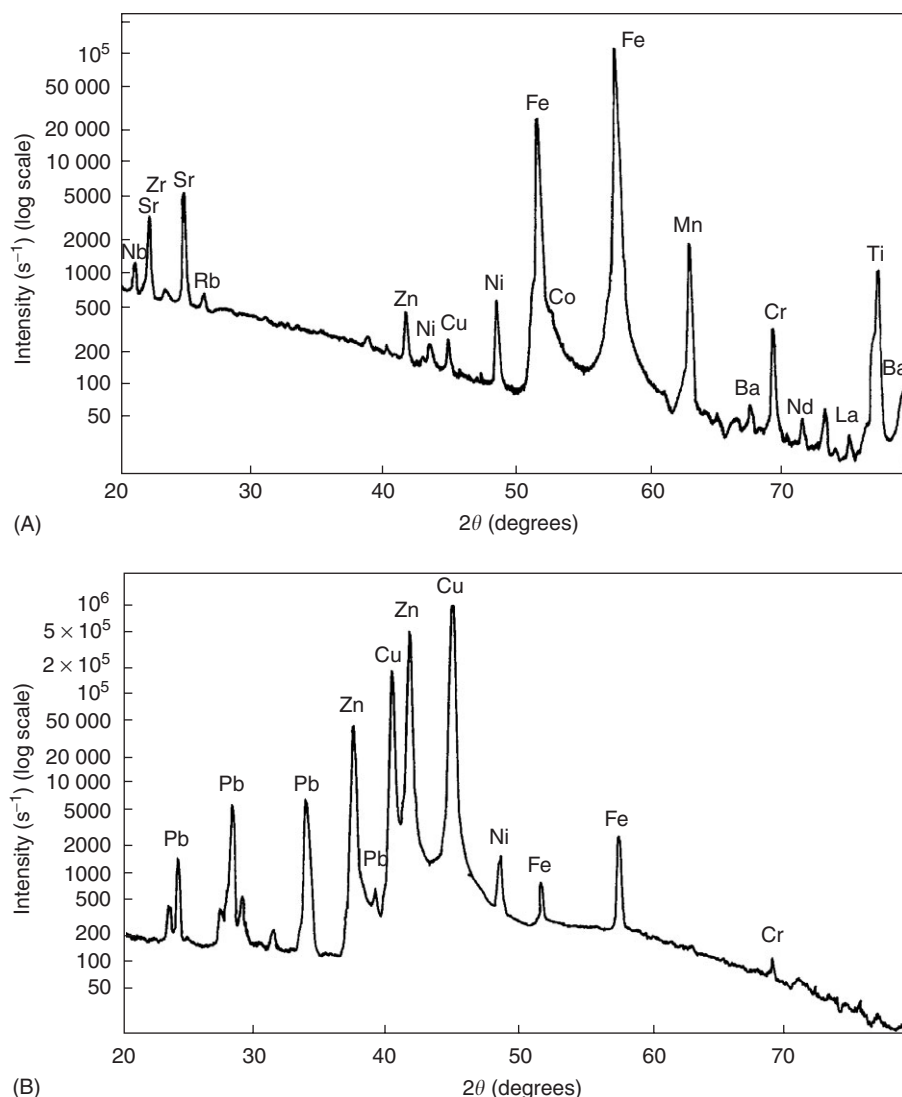
Kuiperes. Philips application report No. 750.

Lee RF and McConchie DM (1982) Comprehensive major and trace element analysis of geological material by X-ray fluorescence using low dilution fusion. *X-ray Spectrometry* 11: 55–63.

Potts PJ, Webb PC, and Watson JS (1984) Energy dispersive X-ray spectrometry fluorescence of silicate rocks for major and trace elements. *X-ray Spectrometry* 13: 2–15.

Webb PC, Potts PJ, and Watson JS (1990) Trace element analysis of geochemical reference samples by energy dispersive X-ray fluorescence spectrometry. *Geostandards Newsletter* 14: 361–372.





**Figure 5** X-ray fluorescence spectra of (A) a silicate rock (basalt) compared with that from a sample of brass (B) recorded with a lithium fluoride diffracting crystal, scanned over the spectrometer range from  $2\theta = 20^\circ$  to  $2\theta = 80^\circ$ . (Spectra by courtesy of Peter Webb, The Open University.)

need replacing periodically to avoid the effects of long-term damage from exposure to X-rays.

### Calibration Procedures

There are several factors that influence the choice of calibration procedure including: (1) selected sample preparation procedure; (2) selected matrix correction procedure; (3) availability of well-characterized reference samples; and (4) requirement to follow a 'standard' scheme of analysis.

Often, the simplest and most convenient calibration procedure is to analyze a range of reference materials, or samples that have been characterized by independent analytical techniques. This procedure is often the only practical method available if samples are to be analyzed in the consolidated state (e.g., as

compressed powder pellets or metallic disks). In these circumstances, it is important that a sufficient number of reference samples are available to cover the range of concentrations and compositions for which determinations are required.

An alternative strategy is appropriate if determinations are to comply with a standard scheme of analysis, such as those developed by the British Standards Institution for the analysis of ceramic and refractory materials. In these schemes, samples are prepared as fused glass disks and calibration samples may then be prepared in the same way using specified proportions of high-purity element oxides. This calibration procedure can, therefore, be considered to be 'absolute' in contrast to the 'relative' measurements made against reference samples.

Before results from a new calibration are accepted, it is essential that the accuracy over the full calibration range is validated independently. Although the intensity response of a WD spectrometer is expected to be linear over six orders of magnitude (subject to corrections for counter dead-time effects), the accuracy of the calibration function is as sensitive as that for any other instrumental technique to the effects of bias, which are likely to be most significant in the analysis of samples at the extremes (low or high) of the calibration range. Discrepancies of this nature are sometimes caused by the use of inaccurate values in reference samples. These discrepancies can only be overcome satisfactorily by a critical evaluation of all calibration data. If the analysis of independently characterized reference materials cannot be used in this evaluation (because, for example, these samples have had to be used as primary calibration samples) then it is possible, though not entirely satisfactory, to evaluate the self-consistency of calibration data in order to identify discrepant points.

### Routine Analysis

Routine XRF analysis typically involves loading a batch of samples together with drift normalization monitors and appropriate quality assurance samples onto an automatic sample changer. Sample names, loss-on-ignition data (if required by the matrix correction procedure), and sample-to-flux ratio (if appropriate) may be required by the analytical program before the instrument commences to analyze the batch in a specified sequence using precalibrated instrument conditions (e.g., spectrometer parameters including count time and the sequence of peak and background angles to be measured, overlap interference correction factors, calibration function, and matrix correction procedures). In calculating elemental concentrations, individual intensity data must be corrected for background contribution, spectral overlap interference (if appropriate), instrumental drift, and matrix effects (including the influence of unanalyzed low atomic number elements assumed by stoichiometry).

### Performance Parameters

The theoretical precision of XRF count data is described by Poisson statistics such that the ideal standard deviation ( $\sigma$ ) in a measurement involving the accumulation of  $n$  counts can be calculated as  $\sqrt{n}$  and the coefficient of variation as  $(\sqrt{n}/n)$ , which equals  $(1/\sqrt{n})$ . In terms of count rate ( $I$ ),  $n$  is calculated from  $n = I \times t$ , where  $t$  is the count time. In assessing the uncertainty in a net peak measurement,

it is necessary to take into account the contribution resulting from subtracting a background measurement. Over the short term, instrumental contributions to the precision of a series of measurements are likely to be very small, perhaps as low as 0.1% relative so that measured precisions will approach theoretical values predicted by Poisson statistics.

Unlike most other atomic spectroscopic techniques, which normally require repetitive measurements on a blank sample, an estimate of XRF detection limit performance can be made from a knowledge of the appropriate background count rate and elemental sensitivity. Using the theoretical basis outlined above, it can be shown that the lower limit of detection (LLD) (i.e., the  $3\sigma$ , detection limit) can be calculated from:

$$\text{LLD} = 3\sqrt{[2(I_b t)]/(I_p t/C)} = 3\sqrt{(2I_b/t)C/I_p}$$

where  $I_b$  is the background count rate counted for time  $t$ ,  $I_p$  is the peak count rate obtained using the same count time from a sample containing the element at concentration  $C$ , and the factor  $\sqrt{2}$  is required to take into account additional uncertainties in peak and background measurements that are subtracted to calculate the net signal. The factor  $(I_p t/C)$ , incorporated into the above expression, is essentially a sensitivity factor to convert the detection limit signal into a concentration.

Representative detection limit data for a range of elements commonly determined by XRF are listed in **Table 3** and show that the technique is capable of achieving detection limits down to the  $\mu\text{g per g}$  level (referring to the solid sample) from the K lines of elements that are most sensitively excited by the X-ray tube. Detection limits increase for low atomic number elements, in part because of the lower excitation efficiency using photons (electron excitation is more effective for elements below about silicon in the periodic table) and in part because of the attenuation effects of counter windows on low-energy X-ray photons.

### Statistical Process Control

XRF analysis is used widely by laboratories that monitor the routine production of materials. In such applications, analyses are required to confirm that the production process is within specified limits. In these (and other) circumstances, it is important to be able to demonstrate that the instrument response has not drifted beyond acceptable limits for day-to-day operation, otherwise the wrong remedial action might be taken, so affecting the quality of the product. Statistical process control procedures require

that selected quality control samples are analyzed periodically in the course of each analytical run. Results from the quality control samples are then compared with previous data, usually plotted on a control chart, to see if any indications of 'out-of-control-data' are exceeded, such as:

1. The current elemental results lie outside predetermined upper or lower control limits, normally set at  $\pm 3\sigma$  of the expected mean.
2. The present and previous six results all lie above or below the center line of the control chart.
3. Six consecutive points all progressively increase or decrease in value.

If any of these criteria are satisfied, action is required to investigate for analytical drift and recalibrate the instrument, if necessary.

## Qualitative Analysis

WD-XRF is not as effective in the qualitative analysis of unknown samples as energy dispersive (ED)-XRF. However, modern instruments can readily be programmed to scan through selected regions of the spectrum and to identify XRF lines, or to 'line hop' from one target line to the next (Figure 5). If it is necessary to evaluate the presence of minor or trace elements, relatively slow scan rates are necessary to improve sensitivity. Furthermore, if it were necessary to evaluate the presence of all XRF-analyzable elements, separate scans using several crystals would be required. As a powerful aid in identifying unknown samples, rapid semiquantitative analysis can be undertaken by programming the instrument to undertake short count measurements on all relevant elements. Fundamental parameter-type

matrix correction procedures can then be applied to data to quantify results without using calibrants. Such procedures often have the option of taking into account the known (or assumed) concentration of unanalyzable elements (e.g., oxygen, carbon, or hydrogen) from knowledge of the sample type (e.g., specifying the sort of plastic that is being measured). Results from such a procedure will not be as accurate as data from a full matrix-matched calibration program. However, the technique is rapid, does not necessarily require any prior knowledge of matrix composition, and the resultant semiquantitative elemental data may be sufficiently successful in fingerprinting the sample to preclude the necessity for further, more detailed analysis.

**See also:** **Process Analysis:** Maintenance, Reliability, and Training. **Quality Assurance:** Quality Control. **X-Ray Fluorescence and Emission:** X-Ray Fluorescence Theory; Energy Dispersive X-Ray Fluorescence.

## Further Reading

- Analytical Methods Committee (1990) Evaluation of analytical instrumentation. Part VI. Wavelength dispersive X-ray spectrometers. *Analytical Proceedings* 27: 324–333.
- Bertin EP (1970) *Principles and Practice of X-Ray Spectrometric Analysis*. New York: Plenum Press.
- Jenkins R (1988) *X-Ray Fluorescence Spectrometry*. Chichester: Wiley.
- Potts PJ (1987) X-ray fluorescence analysis: Principles and practice of wavelength dispersive spectrometry. In: Potts PJ (Ed.) *A Handbook of Silicate Rock Analysis*, ch. 8. Glasgow: Blackie.
- Potts PJ and Webb PC (1992) X-Ray spectrometry. *Journal of Geochemical Exploration* 44: 251–296.
- Williams KL (1987) *Introduction to X-Ray Spectrometry*. London: Allen and Unwin.

# Energy Dispersive X-Ray Fluorescence

**P J Potts**, The Open University, Milton Keynes, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Although the principles and mechanism of excitation for energy dispersive X-ray fluorescence (ED-XRF) are exactly the same as for wavelength dispersive (WD)-XRF, the analytical characteristics of ED-XRF instrumentation demonstrate some significant

differences, largely due to the different response characteristics of an ED detector. Traditionally, ED detectors were popularly associated with the Si(Li) detector (lithium-drifted silicon). However, although Si(Li) remains the standard against which other detector types are compared, technological advances have led to the introduction of alternative detector types, not the least the silicon PIN diode, the silicon drift detector, as well as germanium, mercury(II) iodide, and other semiconductor devices. Even the traditional gas proportional counter can

that selected quality control samples are analyzed periodically in the course of each analytical run. Results from the quality control samples are then compared with previous data, usually plotted on a control chart, to see if any indications of 'out-of-control-data' are exceeded, such as:

1. The current elemental results lie outside predetermined upper or lower control limits, normally set at  $\pm 3\sigma$  of the expected mean.
2. The present and previous six results all lie above or below the center line of the control chart.
3. Six consecutive points all progressively increase or decrease in value.

If any of these criteria are satisfied, action is required to investigate for analytical drift and recalibrate the instrument, if necessary.

## Qualitative Analysis

WD-XRF is not as effective in the qualitative analysis of unknown samples as energy dispersive (ED)-XRF. However, modern instruments can readily be programmed to scan through selected regions of the spectrum and to identify XRF lines, or to 'line hop' from one target line to the next (Figure 5). If it is necessary to evaluate the presence of minor or trace elements, relatively slow scan rates are necessary to improve sensitivity. Furthermore, if it were necessary to evaluate the presence of all XRF-analyzable elements, separate scans using several crystals would be required. As a powerful aid in identifying unknown samples, rapid semiquantitative analysis can be undertaken by programming the instrument to undertake short count measurements on all relevant elements. Fundamental parameter-type

matrix correction procedures can then be applied to data to quantify results without using calibrants. Such procedures often have the option of taking into account the known (or assumed) concentration of unanalyzable elements (e.g., oxygen, carbon, or hydrogen) from knowledge of the sample type (e.g., specifying the sort of plastic that is being measured). Results from such a procedure will not be as accurate as data from a full matrix-matched calibration program. However, the technique is rapid, does not necessarily require any prior knowledge of matrix composition, and the resultant semiquantitative elemental data may be sufficiently successful in fingerprinting the sample to preclude the necessity for further, more detailed analysis.

**See also:** **Process Analysis:** Maintenance, Reliability, and Training. **Quality Assurance:** Quality Control. **X-Ray Fluorescence and Emission:** X-Ray Fluorescence Theory; Energy Dispersive X-Ray Fluorescence.

## Further Reading

- Analytical Methods Committee (1990) Evaluation of analytical instrumentation. Part VI. Wavelength dispersive X-ray spectrometers. *Analytical Proceedings* 27: 324–333.
- Bertin EP (1970) *Principles and Practice of X-Ray Spectrometric Analysis*. New York: Plenum Press.
- Jenkins R (1988) *X-Ray Fluorescence Spectrometry*. Chichester: Wiley.
- Potts PJ (1987) X-ray fluorescence analysis: Principles and practice of wavelength dispersive spectrometry. In: Potts PJ (Ed.) *A Handbook of Silicate Rock Analysis*, ch. 8. Glasgow: Blackie.
- Potts PJ and Webb PC (1992) X-Ray spectrometry. *Journal of Geochemical Exploration* 44: 251–296.
- Williams KL (1987) *Introduction to X-Ray Spectrometry*. London: Allen and Unwin.

# Energy Dispersive X-Ray Fluorescence

**P J Potts**, The Open University, Milton Keynes, UK

© 2005, Elsevier Ltd. All Rights Reserved.

## Introduction

Although the principles and mechanism of excitation for energy dispersive X-ray fluorescence (ED-XRF) are exactly the same as for wavelength dispersive (WD)-XRF, the analytical characteristics of ED-XRF instrumentation demonstrate some significant

differences, largely due to the different response characteristics of an ED detector. Traditionally, ED detectors were popularly associated with the Si(Li) detector (lithium-drifted silicon). However, although Si(Li) remains the standard against which other detector types are compared, technological advances have led to the introduction of alternative detector types, not the least the silicon PIN diode, the silicon drift detector, as well as germanium, mercury(II) iodide, and other semiconductor devices. Even the traditional gas proportional counter can

be used as an ED detector in instruments designed for certain dedicated applications. The key feature of all these devices is the 'parallel' mode of data acquisition, that is, the capability of detecting simultaneously the full X-ray spectrum emanating from a sample using a multichannel analyzer to convert detected X-ray events into a spectrum represented as a histogram plotted as channel number (proportional to X-ray energy) versus counts (i.e., X-ray intensity).

This mode of operation offers unparalleled opportunities for qualitative analysis (from a visual appraisal of spectra, or even quantification in real time whilst spectra are accumulating). However, there are also additional criteria in optimizing excitation conditions to maximize the emission of XRF intensities in the region of analytical interest. The reason for this is that whereas the WD X-ray spectrometer is designed as a monochromator, the entire spectrum emitted from an excited sample is available for detection in an ED-XRF instrument.

ED spectrometers have no moving parts and have benefited extensively from the miniaturization that has taken place in integrated circuit technology. To an increasing extent, ED-XRF instruments can be adapted to XRF applications in a much more flexible manner than WD spectrometers. There are a number of reasons for this flexibility, which may be summarized as follows:

1. The parallel mode of ED spectrum acquisition (mentioned above) means that spectra can be displayed and interrogated during acquisition, permitting qualitative identification, and indeed, an estimate of the quantitative composition to be made in a few seconds.
2. A much larger solid angle of the X-ray spectrum emitted by the sample is detected in ED-XRF than by WD spectrometers (where a high degree of collimation is required to maintain an adequate level of spectrum resolution). Using conventional excitation geometries, ED-XRF excitation sources can be up to 100 times lower in intensity than would normally be required for WD-XRF. For this reason, it is relatively simple to use ED systems in polarized excitation geometries (where excitation intensities are substantially reduced during the polarization process) and so take advantage of the consequential improvement in signal-to-background ratios that result from this method of excitation.
3. The spectral resolution of semiconductor ED detectors is inherently worse than that of WD spectrometers for silicon detectors in the spectral region from 0 to  $\sim 20$  keV. Conversely, this type of detector offers a resolution advantage above 20 keV. The

resolution of ED devices is largely controlled by their electronic characteristics, as explained below. The consequence of the poorer resolution in the 0–20 keV region is that in comparison with WD-XRF measurements of equivalent spectra, signal-to-background ratios, detection limits, and spectrum overlap interferences will all be inherently worse, although there are other opportunities available in ED to compensate for these effects (see below).

4. Because of the electronic noise characteristics of semiconductor detectors, the maximum count rate of an ED system is usually no more than 30 000–50 000 counts per second, compared with in excess of  $10^6$  counts for a WD spectrometer, acting as a single channel analyzer. The reason is the relatively long count time constant that must be selected for ED preamplifier circuits (typically 6  $\mu$ s), compared with  $<1$   $\mu$ s for counter preamplifiers used on WD spectrometers. As a consequence, equivalent analytical precision can only be achieved by extending ED spectrum count times to a significant extent compared to WD.

5. As ED spectra are stored in digital form, more sophisticated procedures are available to correct for background contributions and overlap interferences than WD.

6. Because ED detectors 'see' the entire spectrum, optimization of analytical programs is not just a question of maximizing sensitivity for the elements of interest. Consideration must be given to maximizing detection capabilities to the 'region', which contains the element lines of interest so the finite spectrum processing capacity is not wasted on other spectral regions that contain little or no useful analytical information.

At one time, ED applications were often more limited than those of WD-XRF instruments, and were not used as extensively for high-precision quantitative analysis. However, following developments in semiconductor detector technology and in the speed and stability of digital processing electronics, this distinction is no longer as valid. Because of the versatile way in which ED-XRF can be configured, ED has become the instrument of choice in a large range of quantitative applications, particularly those where state-of-the-art precision or detection limit capability is not a fitness-for-purpose criterion.

## ED-XRF Excitation Configurations

The instrumental configurations most commonly used in ED-XRF spectrometry can be categorized as follows.

### Direct Tube Excitation

In the simplest excitation configuration (comparable to that employed in the majority of WD-XRF instruments), the sample is excited directly by the output of an X-ray tube (Figure 1). Selective excitation can be achieved by judicious choice of tube anode, operating voltage, and the use of a primary beam filter (normally comprising a thin foil of metal or cellulose), placed between tube and sample to modify the energy distribution of the tube spectrum reaching the sample. In this way, effective excitation can be restricted to a particular region of the fluorescence spectrum, so maximizing utilization of the pulse processing capacity for detection of the group of elements of interest. Adequate excitation can normally be achieved using a maximum tube power of only 50 W.

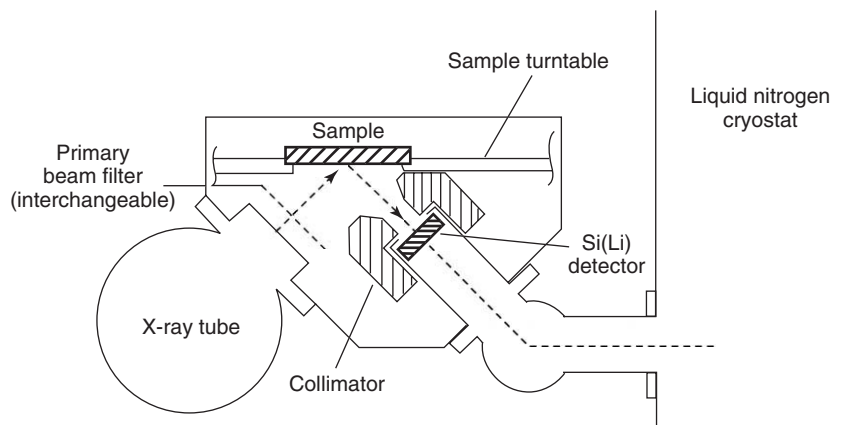
Taking these factors into account, optimum excitation of elements below iron in the periodic table would normally be achieved using a tube potential of 10–15 kV with no primary beam filter (or alternatively a thin cellulose filter if determination of the lightest elements was not required). Conversely, higher atomic number elements determined from fluorescence lines above  $\sim 7$  keV would normally be excited at 30–50 kV using an appropriate metal foil as primary beam filter. As in conventional WD-XRF, the choice of tube anode is important in dictating the elements that may be detected with highest sensitivity. An example of a spectrum acquired for the measurement of the K-lines of trace elements rubidium, strontium, yttrium, zirconium, and niobium in geological samples using a silver anode X-ray tube in conjunction with a silver foil primary beam filter (in a combination sometimes known as a 'regenerative monochromatic' filter) is shown in Figure 2. The silver tube K-lines are efficient in exciting the element lines of interest. At the same time, the silver foil

attenuates the lower-energy tube continuum, so that matrix elements (specifically, K-lines of the elements up to iron in the periodic table) are not excited effectively and do not, therefore, compete for detection capacity (region A in Figure 2). Furthermore, although the attenuation properties of the silver foil allow efficient transmission of the silver tube K lines, lower-energy tube continuum radiation in the region of the trace element lines of interest is attenuated to a significant extent, so reducing the background contribution from scatter off the sample (region B in Figure 2). Using this excitation strategy, together with a count time extended up to 1200 s, detection limits can be achieved for the specified trace elements in the range 1–2  $\mu\text{g per g}$ . In this way, direct tube excitation offers a flexible and sensitive method for the selective excitation of a wide range of elements for quantitative or qualitative analysis.

### Secondary Target Excitation

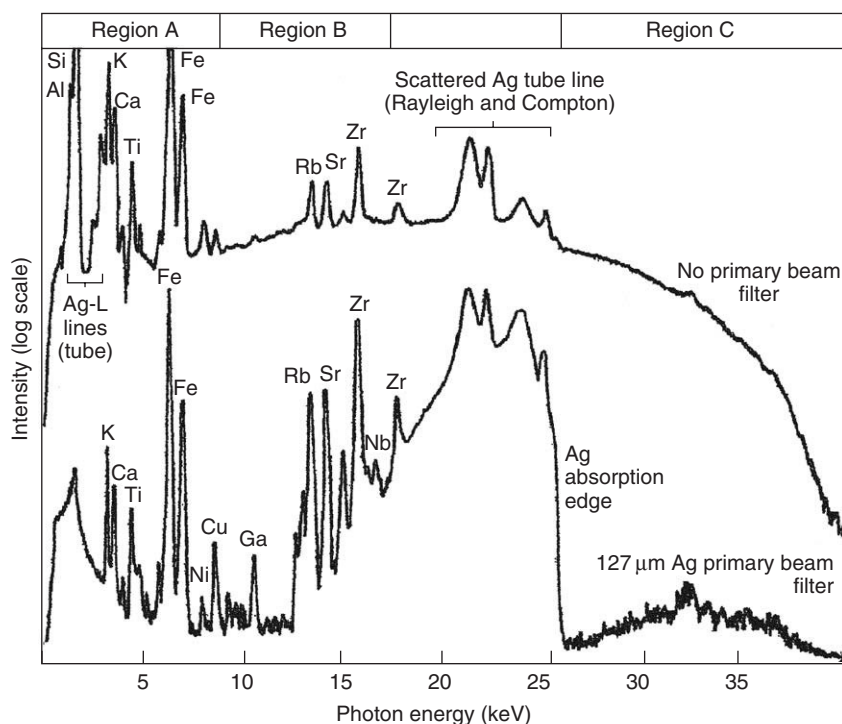
In a secondary target ED-XRF instrument, the sample is excited by characteristic fluorescence radiation from an elemental 'target', which is itself excited by an X-ray tube (Figure 3). This configuration offers the advantage of selective excitation of the group of element lines that have critical excitation energies immediately below the energy of the characteristic fluorescence lines from the secondary target. By exchange of secondary target, the instrument can be 'tuned' to measure different ranges of elements with high sensitivity. Performance may be further enhanced by use of a primary beam filter between X-ray tube and secondary target fulfilling a role similar to that described in the previous section.

The range of secondary targets installed on a typical instrument and their element ranges is shown in Table 1. Although secondary targets of high atomic

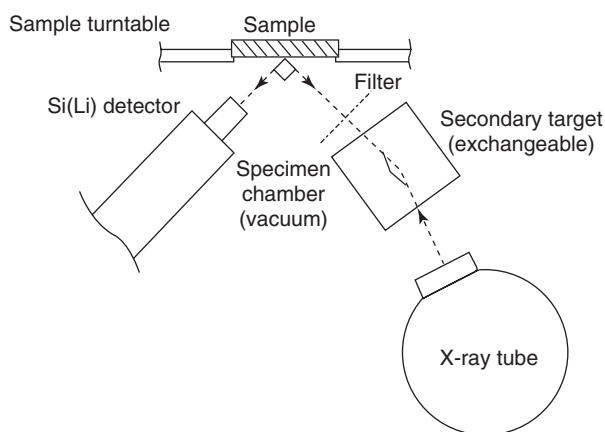


**Figure 1** Schematic view of an ED-XRF instrument incorporating direct tube excitation showing X-ray tube, primary beam filter, sample, and Si(Li) ED detector.





**Figure 2** ED-XRF spectrum of a granitic rock excited with a silver anode X-ray tube at 40 kV, comparing a spectrum recorded by unfiltered direct excitation with one recorded with a 127  $\mu\text{m}$  silver foil filter placed in the primary X-ray beam. Spectra have been offset and are plotted using a logarithmic intensity scale. The effect of the silver filter is as follows: (1) suppression of fluorescence of the major elements (region A); (2) enhancement of peak-to-background ratios for geochemically interesting trace elements (region B); and (3) selective attenuation of the higher-energy tube continuum (region C).



**Figure 3** Schematic view of an ED-XRF instrument configured for secondary target excitation. The tube–secondary target–sample–detector path is designed to have a Cartesian arrangement to suppress scatter of tube radiation into the detector.

number metals (e.g., gadolinium, silver) can be used to excite a relatively wide range of elements, the converse is true of lower atomic number targets (e.g., iron and titanium). These characteristics result from the way in which the mass attenuation properties of elements vary with energy. For the efficient determination of the lowest atomic number elements (in

particular below about potassium in the periodic table), direct tube excitation is often used.

A further advantage of secondary target excitation geometry is the reduction in background that can be achieved if the instrument is designed with a Cartesian X-ray optical configuration between tube–target–sample–detector, as shown in **Figure 3**. After scatter off the secondary target, radiation from the X-ray tube is polarized at right angles to a plane passing through the tube, target, and sample. In theory, this polarized radiation cannot be scattered a second time off the sample in the direction of the detector providing the Cartesian arrangement is employed. As a consequence, the intensity of the ‘Barkla’ scattered background (which is one of the limitations in the detection limit performance of any XRF instrument) is reduced significantly. However, to ensure adequate excitation intensities, most secondary target instruments are designed with close coupling between the various components so that the ideal Cartesian performance is reduced by a solid-angle-effect. The degree of polarization is also reduced by multiple scatter of X-ray photons in the target or sample. Even so, secondary target instruments offer flexible operating characteristics in exciting elements to best detection limits in the

**Table 1** Selection of secondary targets available for comprehensive analysis by secondary target ED-XRF showing principal elemental applications

Target	Principal elemental applications <sup>a</sup>																		
Gd	Sr Y Zr Nb <b>Mo</b> Ru <b>Mo</b> Rh <b>Pd</b> Ag <b>Cd</b> In <b>Sn</b> Sb <b>Te</b>   <b>Cs</b> <b>Ba</b> <b>La</b> <b>Ce</b> <b>Pr</b>																		
Ag	Ni Cu Zn Ga Ge As Se Br <b>Rb</b> Sr Y Zr Nb <b>Mo</b>																		
Zr	Fe Co Ni Cu Zn <b>Ga</b> <b>Ge</b> <b>As</b> <b>Se</b> <b>Br</b> <b>Rb</b>																		
Ge	Sc Ti V <b>Cr</b> <b>Mn</b> <b>Fe</b> <b>Co</b> Ni Cn Zn																		
Fe	K Ca <b>Sc</b> <b>Ti</b> V Cr																		
Ti	S Cl K <b>Ca</b> Sc																		
Direct tube excitation	<b>Na</b> <b>Mg</b> <b>Al</b> <b>Si</b> <b>P</b> <b>S</b> <b>Cl</b>																		

<sup>a</sup> Analytes given in bold type.

1–5 µg per g range. Most efficient use of instrumentation is in the determination of groups of elements that can be excited efficiently by one or two secondary targets, since comprehensive elemental analysis would require sequential measurement of spectra using a number of secondary targets, together with direct tube excitation for the lightest elements. A further development of this geometric arrangement is to substitute targets made of low atomic number material to scatter the tube spectrum, a polarized excitation geometry that is discussed further below.

### Radioisotope Excitation

In view of the efficient detection capabilities of the ED detector, effective XRF measurements can be made with a suitable radioisotope source of moderate activity, despite the lower photon intensity that is likely to be available in comparison with an X-ray tube. To cover the full analytical range, four radioactive isotopes are available: <sup>55</sup>Fe (which emits Mn K-line X-rays at 5.9 and 6.5 keV), <sup>244</sup>Cf (14.3–21.4 keV, Pu L-line X-rays), <sup>109</sup>Cd (22.2–25.5 keV Ag K-line X-rays), and <sup>241</sup>Am (59.5 keV γ-emission). Depending on the application, instruments may be designed with a single source often mounted in an annular configuration to maximize excitation and detection efficiencies (Figure 4), or alternatively with different sources mounted on a carousel to facilitate sequential excitation of samples. In contrast to the instrumentation described in the previous sections, radioisotope excitation offers little scope for changing excitation characteristics (other than by exchange of source). However, with no requirement for a tube power supply, instrumentation is potentially cheap and simple to operate and has found particular use in two areas:

1. Specific applications in the laboratory or factory environment that cannot readily be satisfied by other techniques; for example, the determination of sulfur in oil or chlorine in brines and liquors (competitive techniques based on optical emission, for example, are often insensitive to the electronegative elements).
2. Instruments in which a portable excitation and analysis unit can be used for the field measurement of environmental contamination (e.g., measuring the distribution of heavy metals in soil at an industrial site) or in mineral exploration applications where the ability to analyze a sample directly in the field cannot be rivaled by competitive techniques. In some applications, particularly for the analysis of simple matrices where special overlap interferences do not present a problem, a proportional counter can be used as the ED detector. Although offering poor

spectral resolution, these latter devices are less expensive, robust, and do not require cooling.

### Total Reflection XRF Spectrometry

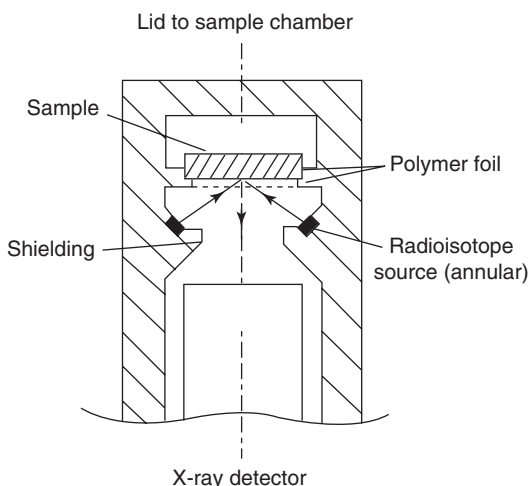
Total reflection XRF instruments (see **Figure 5**) are based on the principle that when an X-ray beam strikes an optically flat surface at grazing incidence (i.e., at below the critical angle), the beam will suffer total reflection from the surface, in theory, without absorption or scatter. In these circumstances, a thin sample deposited on the surface will be selectively excited with considerable suppression of the background component, which would otherwise result from scatter events in the substrate. This specialized XRF configuration incorporates ED

technology and is described in detail elsewhere in this encyclopedia.

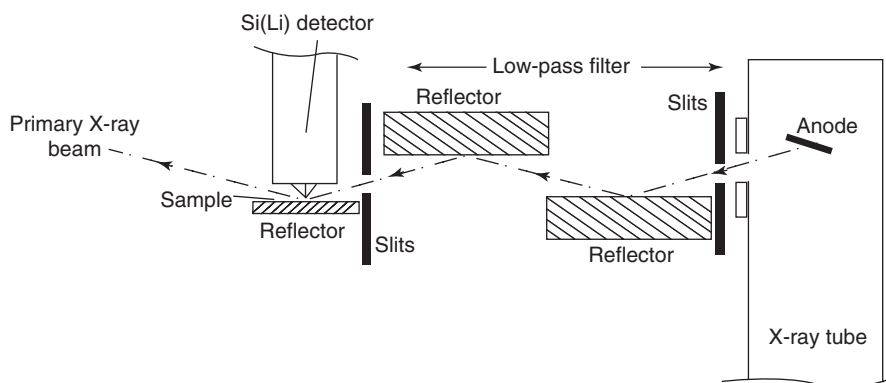
### ED-XRF Using Polarized Excitation

Low atomic number materials are more efficient at scattering X-rays than higher atomic materials. If, therefore, the metal target (in the secondary target ED-XRF configuration, **Figure 3**) is replaced by a low atomic number target such as boron carbide (or corundum), the benefits of polarization can be achieved, without interference from nonpolarized secondary target lines. The entire tube spectrum is polarized by the low atomic number secondary target and is available to excite the sample, but this polarized beam cannot (in an ideal geometric arrangement) be scattered off the sample into the detector. This arrangement offers the benefits of direct tube excitation but using a polarization configuration that suppresses the detection of scattered tube background. Instruments of this design are capable of achieving detection limits in the range 1–5  $\mu\text{g per g}$  in the analysis of environmental samples. To compensate for intensity losses during the excitation process, the primary excitation source must be a higher power X-ray tube (e.g., 1–2 kW).

A second process is capable of achieving a similar effect, but instead based on Bragg diffraction at  $90^\circ$ . A few combinations of X-ray tube lines and crystalline materials cut at a particular orientation display the property of Bragg diffraction through a  $2\theta$  angle of  $90^\circ$ . Polarization is inherent in the diffraction process, so that the overall benefit in this arrangement is similar to the Barkla scatter process. However, diffraction is a more efficient process than scatter, so that the Bragg diffraction of Pd L-lines from a particular orientation of a highly orientated pyrolytic graphite crystal provides a significant boost to the excitation of the low atomic number elements (Na to Cl K-lines).



**Figure 4** Radioisotope ED-XRF instrument showing configuration of annular source, sample, and detector. The X-ray detectors used in this application include Si(Li) and  $\text{HgI}_2$  semiconductor detectors and gas proportional counters, the latter being favored for element-specific laboratory applications.



**Figure 5** Total reflection XRF instrumentation showing the arrangement of X-ray tube, reflectors, and ED detector. The sample is normally prepared by evaporating a sample solution onto the appropriate quartz reflector plate.

### Synchrotron Radiation XRF (SXRF)

The synchrotron is a highly specialized radiation source used to generate high intensity, near parallel, polarized beams of X-rays. These beams can be configured to undertake XRF measurements of samples in a number of ways and a general arrangement is shown in Figure 6. In a more specialized X-ray microprobe configuration, the beam is focused by reflection or diffraction off appropriate X-ray mirrors. Because of the high intensity and polarized nature of this X-ray source, the SXRF microprobe is capable of achieving very low detection limits for some elements in the submicrogram per gram range. This technology and its analytical capabilities are described in more detail elsewhere in this encyclopedia.

### ED-XRF Instrumentation

The particular aspects of ED-XRF instrumentation that facilitated the development of the above forms of instrumentation and give the technique its distinctive analytical characteristics are as follows.

#### Semiconductor Detector

Semiconductor technology is still a rapidly developing field that has seen the introduction of new detector types in recent years. However, the standard detector type on which ED-XRF technology was founded is the lithium-drifted silicon detector, Si(Li). This is the detector type against which the performance of new types of semiconductor detector is evaluated.

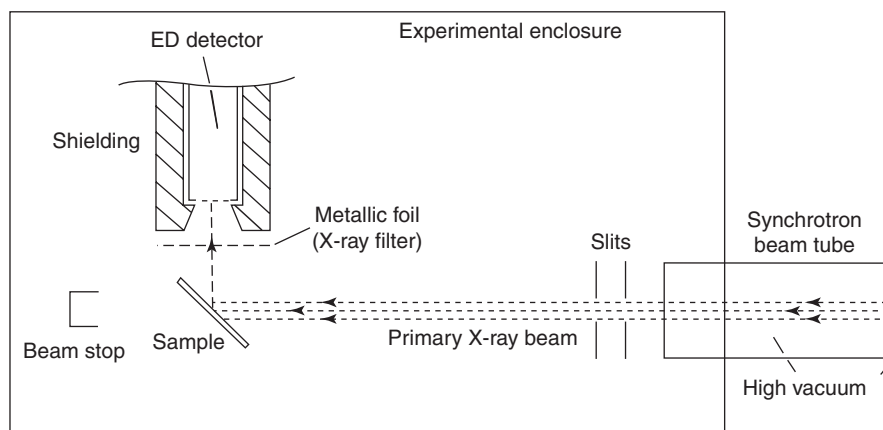
The Si(Li) detector crystal is made from high-purity silicon. Lithium atoms are drifted into its lattice structure to compensate for undesirable electronic properties caused by the presence of impurity

atoms. The detector crystal is normally machined in the form of a 'grooved disk' profile of thickness  $\sim 3\text{--}5\text{ mm}$  and active cross-section  $10\text{--}50\text{ mm}^2$  mounted inside a detector housing, which is maintained under high vacuum and is fitted with an entrance window made of beryllium foil (Figure 7). X-ray photons interact with the crystal and cause ionization from the valence to conduction bands. The small electronic charge formed by this process is proportional to the energy of the detected X-ray photons. This charge is collected by applying a bias of 500–1000 V across the detector crystal and amplifying it using a high-gain field-effect transistor (FET) prior to further amplification and pulse processing.

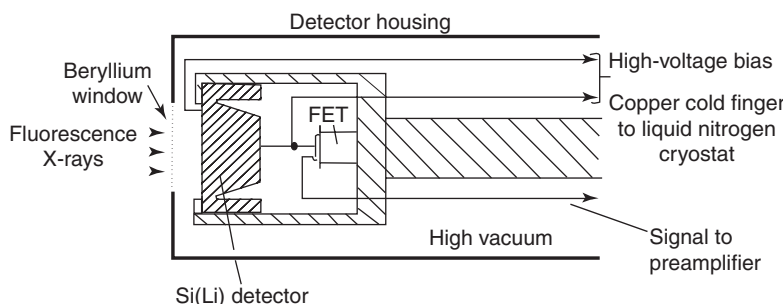
Each detected event creates a relatively small electronic charge that must be distinguished from the inherent electronic noise in the detector crystal. To maximize the signal-to-noise ratio, design features of Si(Li) spectrometers include:

1. Cooling the detector crystal, preferably to liquid nitrogen temperatures, to minimize thermal noise contributions to the detected electronic signal.
2. Use of a low-noise, high-gain FET mounted in the vacuum housing adjacent to detector crystal (to minimize contributions from stray electronic noise) as the first stage of signal amplification.
3. Selection of a relatively long time constant for the integration of the detected signal in the main pulse processing amplifier.

With these features, the overall resolution of a Si(Li) detector and pulse processing system used in XRF applications measured as the full-width at half-maximum (FWHM) of the Mn K- $L_{2,3}$  ( $K\alpha$ ) line at 5.9 keV is typically 140–155 eV (noting that <120 eV FWHM is technically feasible).



**Figure 6** Synchrotron radiation-induced X-ray fluorescence instrumentation showing, schematically, the arrangement of sample and detector in an experimental enclosure placed at the end of a synchrotron beam line.



**Figure 7** Cross-sectional diagram of an ED detector showing an Si(Li) detector crystal and associated FET mounted at the end of a copper cold finger inside an evacuated housing. The detector housing is fitted with an entrance window normally made of beryllium foil.

These detection characteristics apply to other semiconductor detectors, which do, however, possess other response characteristics. For example, germanium X-ray detectors are very similar in design to Si(Li) detectors but theoretically have better energy resolution (down to  $\sim 120$ – $140$  eV FWHM at  $5.9$  keV in XRF applications), but more importantly, higher stopping power to higher-energy fluorescence X-rays (because of the higher atomic number of Ge compared with Si). Germanium detectors offer positive benefits in the detection of the spectrum region above  $20$  keV where the detection efficiency of Si(Li) starts to fall off markedly. Exploitation of this application would normally require the use of a  $100$  kV X-ray tube (or  $^{241}\text{Am}$  excitation source) to excite the K-lines of the heavier trace elements. Mercury(II) iodide is a semiconductor material that has also been developed for this higher-energy fluorescence spectrum application. This detector type has the advantage of operating satisfactorily with a low degree of Peltier cooling (rather than requiring liquid nitrogen cryogenic cooling), and offers an energy resolution of down to  $\sim 185$ – $200$  eV FWHM. Cadmium–zinc–telluride (CZT) has also been developed for this application; however, its use is currently more experimental than for routine applications. The most influential recent developments have been the silicon-PIN diode (Si-PIN) and the silicon drift detector. Both detector types have the advantage of being non-cryogenic (Peltier cooling only). Both are currently fabricated typically as wafers  $300\text{ }\mu\text{m}$  thick and so are optimized for detection of the  $0$ – $20$  keV region of the X-ray spectrum (detector efficiencies of  $300\text{ }\mu\text{m}$  wafers start to drop off significantly above  $15$ – $20$  keV) and both offer resolutions that approach or are better than  $180$  eV (at  $5.9$  keV). The Si-PIN diode operates by detecting the electronic charge created by the interaction of an X-ray photon with the photodiode material. To achieve adequate resolution (currently better than  $\sim 180$ – $200$  eV FWHM at  $5.9$  keV), the detector signal processing electronics must be operated at a relatively large time constant (e.g.,  $6$ – $20\text{ }\mu\text{s}$ ).

This restricts the total count rate capability of the detector system to  $2000$ – $10\,000$  cps. The Si-drift detector has an active area of  $5$ – $10\text{ mm}^2$  and wafer thickness of  $300\text{ }\mu\text{m}$ . The design offers an intrinsically low capacitance, meaning that the detector is capable of retaining its performance even at high count rates ( $10^6$  cps has been achieved with these devices under experimental conditions). Current designs offer resolutions down to  $\sim 140$  eV and, although expensive to fabricate, are also noncryogenic (requiring modest Peltier cooling) and have significant future potential in ED-XRF applications.

In very simple instruments, gas proportional counters can be used as ED detectors. However, their intrinsic resolution of  $\sim 1000$  eV at  $5.9$  keV limits applications to those where significant spectral overlap interferences are not anticipated. At the other extreme, research is currently being undertaken to develop microcalorimeter detectors that must be cooled to a spectacularly low temperature in the millikelvin range to achieve a spectacularly high resolution in the electron-volt range. Not surprisingly, these devices are a long way away from being used in routine applications.

## Pulse Processing Electronics and Dead Time

When considering the operating characteristics of Si(Li) detectors, amplifier time constants of up to  $30\text{ }\mu\text{s}$  are required to distinguish the signal originating from low-energy photon events from the ambient electronic noise component of ED detectors (compared with a value of less than  $1\text{ }\mu\text{s}$  used for WD-XRF proportional counter signal amplifiers). This requirement has the undesirable effect of limiting the maximum data acquisition rate of the ED spectrometer. Furthermore, the electronic design of pulse recognition, shaping, and amplification circuitry requires careful consideration to optimize performance in this application. Each time an X-ray photon is detected, the input circuits of the pulse

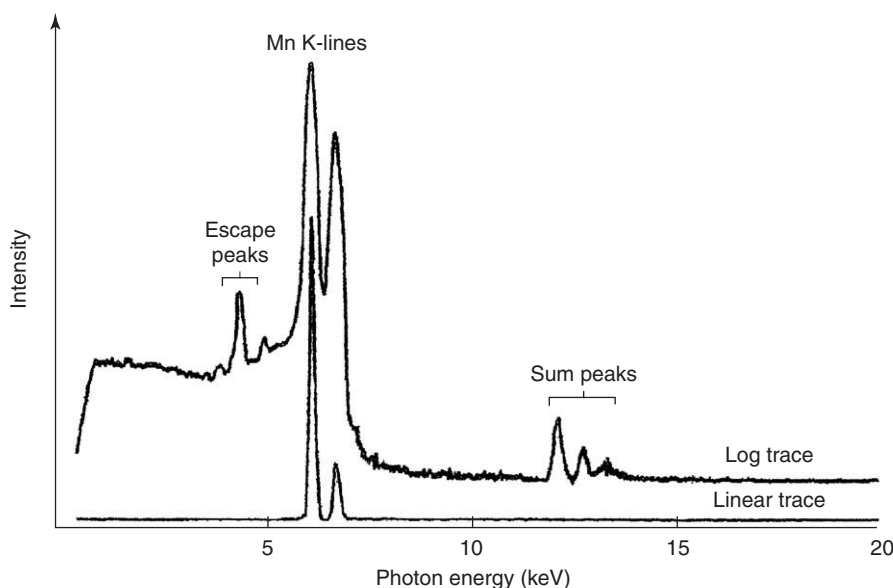
processor must effectively be switched off whilst the current event is being measured. Any events that would have been detected during this 'dead time' interval are essentially unrecorded. To compensate for discrepancies that would occur in the comparison of spectra recorded at different count rates, the spectrum acquisition time must be extended by an amount that will compensate for this 'dead time' effect. In view of the long time constants that must be selected, ED spectrometers cannot normally operate at an output count rate of more than 3000–30 000 cps (depending on circuit design) and would routinely be operated at data acquisition rates corresponding to a dead time of 20–40%. A further problem arises at higher data acquisition rates when there is an increase in the probability of the near-coincident detection of two separate X-ray events. If these events occur within a limiting time interval of a few microseconds, the pulse processor circuits cannot distinguish the separate events. As a result, a sum peak caused by this 'pulse pile-up' event is recorded and appears in the spectral output. This artifact is usually most significant in spectra recorded at high count rates.

## Spectral Artifacts

Spectral artifacts are peaks that are not present in the source fluorescence spectrum but result from idiosyncrasies in the response of the detector and pulse processing system (such as the sum peaks referred to above). Their presence can cause ambiguities in the

qualitative interpretation of ED spectra or errors in quantitative analysis if mistaken for genuine fluorescence lines. Typical artifacts are illustrated in the spectrum recorded using an Si(Li) detector from a radioactive  $^{55}\text{Fe}$  source, which emits Mn K- $L_{2,3}$  ( $K\alpha$ ) and Mn K- $M_{2,3}$  ( $K\beta$ ) radiation, but no continuum (Figure 8). The principal features of this spectrum are as follows:

1. Peaks at 5.9 and 6.5 keV, originating from 'photoelectric detection' of Mn K- $L_{2,3}$  ( $K\alpha$ ) and Mn K- $M_{2,3}$  ( $K\beta$ ), broadened by the statistical factors that affect the detection process as discussed above.
2. 'Escape peaks', which occur when a detected X-ray photon fluoresces silicon atoms in the detector crystal. A small proportion of the resultant Si K-line photons are likely to escape out of the detector rather than be absorbed within the crystal and so represent an apparent loss in detected energy from the parent photon. This phenomenon causes an artifact peak to appear at an energy of 1.78 keV (the energy of the Si K line) below the parent event.
3. 'Sum peaks' corresponding to pulse pile-up of (Mn K- $L_{2,3}$  ( $K\alpha$ ) + Mn K- $L_{2,3}$  ( $K\alpha$ ), (Mn K- $L_{2,3}$  ( $K\alpha$ ) + Mn K- $M_{2,3}$  ( $K\beta$ ), and (Mn K- $M_{2,3}$  ( $K\beta$ ) + Mn K- $M_{2,3}$  ( $K\beta$ )), caused by the inability of the counting system to resolve near-coincident X-ray events, as described in the previous paragraph. Sum peaks are only likely to be of significance when operating at high count rates or when searching for trace elements in a spectrum dominated by major element lines.



**Figure 8** Spectrum of an  $^{55}\text{Fe}$  source (which emits K X-rays) recorded on a Si(Li) ED detector showing spectrum artifacts, including sum peaks and escape peaks. The same spectrum is displayed on both logarithmic and linear intensity scales to show both minor spectral features as well as relative intensities.



4. An 'enhanced background' at photon energies below the fluorescence lines, resulting from incomplete charge collection in the detector crystal. A small fraction of the electronic charge that results from the detection of Mn K-L<sub>2,3</sub> (K $\alpha$ ) and Mn K-M<sub>2,3</sub> (K $\beta$ ) events may be trapped in the crystal lattice so that the detected event contributes in a random manner to the background at energies below the parent peak.

These artifacts must be taken into account in the visual interpretation of ED-XRF spectra and in the spectrum analysis procedures used for quantification of peak areas.

In principle, these detector artifacts are found in other silicon-based detectors (although the relative severity will vary with detector design) and in detector types made of other materials (although some characteristics, such as escape peak energies will be different, depending on the composition of the detector).

## Spectrum Deconvolution

In view of the resolution characteristics of an ED detector, spectral overlap interferences are much more common in ED-XRF than WD-XRF and require a more comprehensive correction procedure than the simple line ratio method often used in the latter technique. However, advantage can be taken of the digital form in which ED spectral data are accumulated in the multichannel analyzer, which offers information about peakshapes as well as intensity. For simple ED spectra, it may be possible to define a series of 'windows' covering each fluorescence peak of interest (the 'region-of-interest' approach). Peak intensities can then be determined by channel integration and a background correction applied by linear interpolation between the lowest and the highest channel of the window. For complex spectra, a more sophisticated approach is required, of which two have found widespread use.

### Background Modeling with Peak Fitting

The background continuum of the fluorescence spectrum can be modeled by fitting a curve derived from Kramer's equation, which has the general form:

$$I = kZ[(E_0 - E)/E]$$

where  $I$  is the continuum intensity at photon energy  $E$ ,  $E_0$  is the energy equivalent to the X-ray tube operating potential,  $Z$  is the mean atomic number of the matrix, and  $k$  a constant. Having corrected for the background continuum using this approach, the area of residual photopeaks (and hence X-ray line intensities) may be calculated by peak fitting, using a

Gaussian or modified Gaussian function. To perform this fitting successfully, the variation of FWHM of the detector as a function of energy must be precalibrated accurately, as must the intensity of the K-M<sub>2,3</sub> (K $\beta$ ) relative to the K-L<sub>2,3</sub> (K $\alpha$ ) line so that the calculated area of the former can be subtracted from the spectrum to compensate for spectral overlap influences.

### Digital Filtering

An alternative approach involves the use of a 'digital filter' to remove the background component of recorded spectra. The digital filter is a simple algebraic function that distinguishes the continuum component of the spectrum from X-ray peaks, normalizing the former to zero. During this transformation, the ED spectrum is convoluted to a form that resembles the negative of the second derivative. Photopeak intensities can then be quantified by fitting, using a least-squares minimization procedure, library spectra of individual elements that have also been convoluted by the same digital filter function. To operate successfully, any drift in spectrum origin and gain must be corrected accurately. This procedure offers a degree of automatic spectral overlap correction and allows quantification of spectral data that show apparently serious overlap interferences involving both K- and L-lines.

## Applications and Operating Practice

Some of the more specialized applications of ED-XRF instrumentation have been summarized in previous sections. Further examples, mainly relevant to direct tube and secondary target excitation using Si(Li) detectors, are summarized here.

### Qualitative Analysis

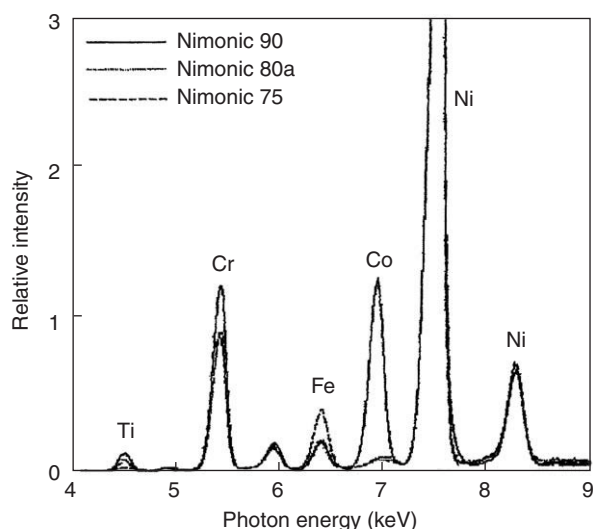
It has already been mentioned that ED-XRF offers unrivalled performance in qualitative analysis, the only real restrictions being that samples should preferably (though not exclusively) be in a solid form and have a size and shape that can be accommodated in the sample chamber. The other main limitation is that the technique is insensitive to elements of atomic number lower than sodium in the periodic table. Since X-ray spectra can normally be interrogated during data acquisition, experienced operators can usually identify the sample matrix from its principal XRF lines within a few seconds. However, if identification depends on the presence of trace elements, ED-XRF data acquisition times may have to be extended substantially and it may then be beneficial to analyze the sample on a WD-XRF instrument to take

advantage of the higher spectral resolution available when scanning the WD spectrometer.

Qualitative ED-XRF analysis is often most effective when the application involves comparing the spectrum of a sample with that from a material of known composition. Differences in composition are then simple to identify visually. This capability, together with the nondestructive nature of ED-XRF, is important in the analysis of museum artifacts and forensic samples. For example, forensic glass fragments can be fingerprinted by ED-XRF on the basis of variations in key elements to distinguish their origin (e.g., vehicle glass versus plate glass versus milk bottle glass). However, such applications are often only fully effective when based on semiquantitative analytical results. The relative intensity of X-ray peaks observed during ED-XRF analysis gives the operator a visual identification of the relative concentration of elements present. However, interpretation cannot be definitive without some form of matrix correction that can be undertaken simply by using the fundamental parameter approach with an appropriate normalization to account for irregular sample geometries. One application of semiquantitative ED-XRF is in alloy sorting, where subtle differences in elemental data from semiquantitative analysis be used to identify the type of alloy (Figure 9).

### Quantitative Analysis

For the highest accuracy, sample presentation must be reproducible, and for this reason, requirements for



**Figure 9** Comparative ED spectra in an alloy sorting application. By rapid ED-XRF analysis for Ni, Cr, Co, Fe, and Ti, these samples can be distinguished as the nickel-based alloys, Nimonic 90, 80, and 75. For convenience, spectra have been normalized to show equivalent Ni K-line intensities. (Spectrum courtesy of AT Ellis, Oxford Instruments.)

the preparation of samples in the form of flat disks of metal, compressed powder pellets, or glass disks are no different from those specified for WD-XRF analysis. However, differences in analytical programs for quantitative analysis by ED-XRF, including (1) instrumental conditions must be optimized to excite selectively regions of the spectrum of analytical interest (as discussed above), and (2) count times for spectrum acquisition must be extended to 500–1200 s to attain adequate precision. In routine schemes of analysis, the precision and detection limits for the lower atomic number elements (particularly between sodium and iron in the periodic table) are not normally as good as those attainable by WD-XRF, because of inferior spectral resolution, additional attenuation of the lower-energy fluorescence spectrum in the beryllium window fitted to the ED detector, and the lower data acquisition rates. However, these restrictions are not as significant in the determination of trace elements in the 10–20 keV region of the spectrum, where ED-XRF results for some elements from an 800–1200 s acquisition time are likely to be comparable with those expected by routine WD-XRF. A comparison of detection limits expected in geological applications is given elsewhere in this encyclopedia.

### Field Analysis

One application where ED-XRF instruments offer exciting potential is in the analysis of samples *in situ* using field-portable instrumentation, based on excitation using radioisotope sources, or more recently miniature X-ray tubes. Such instrumentation facilitates applications that cannot readily be undertaken by other techniques, including the characterization of heavy metal contamination in soils at industrial sites, the nondestructive analysis of large artifacts at archaeological sites, and on-site assessment of environmental hazards from operations such as stripping leaded paint.

*See also:* **Forensic Sciences:** Glass. **Qualitative Analysis. X-Ray Fluorescence and Emission:** Total Reflection X-Ray Fluorescence; Synchrotron X-Ray Fluorescence.

### Further Reading

- Cesareo R (2000) X-ray physics: Interaction with matter, production, detection. *Revista del Nuovo Cimento* 23: 1–23.
- Janssens KH and Adams FC (1989) New trends in elemental analysis using X-ray fluorescence spectrometry. *Journal of Analytical Atomic Spectrometry* 4: 123–135.

- Jenkins R, Gould RW, and Gedke D (1981) *Quantitative X-ray Spectrometry*. New York: Dekker.
- Potts PJ (1987) Energy dispersive X-ray spectrometry. *A Handbook of Silicate Rock Analysis*, ch. 9. Glasgow: Blackie.
- Smith JV (1994) Synchrotron X-ray microanalysis. In: Potts PJ, Bowles JFW, Reed SJB, and Cave MR (eds.)

- Microprobe Techniques in the Earth Sciences*, ch. 5. London: Chapman & Hall.
- Van Grieken RE and Markowicz AA (1992) *Handbook of X-ray Spectrometry*. New York: Dekker.
- Woldseth R (1973) *X-ray Energy Spectrometry*. Burlingame, CA: KeveX.

## Total Reflection X-Ray Fluorescence

**M Schmeling**, Loyola University Chicago, Chicago, IL, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Total reflection X-ray fluorescence (TXRF) spectrometry is a trace elemental microanalysis technique based on conventional energy dispersive X-ray fluorescence. It has become increasingly popular in the last decade and is applied in almost every field of trace elemental analysis where low detection limits and multielement capabilities are required. Like all X-ray techniques, TXRF is nondestructive making it extremely useful and important in areas where samples are precious and/or need to be used for further characterization. New tabletop instruments make this technique affordable and more versatile as it can be used also for field research. In the semiconductor industry, TXRF is now routinely applied to scan wafers for impurities on the surface and in near-surface layers. The following article introduces the basic principle of TXRF and its instrumental features and discusses various applications of this technique.

### Basic Principles

#### Total Reflection, Reflectivity, and Penetration Depth

When an X-ray beam strikes the surface of a solid medium at a certain angle  $\varphi$  it can be refracted or reflected. According to Snell's law the refractive indices  $n$  of two media 1 and 2 are in the following relationship:

$$n_1 \cos \varphi_1 = n_2 \cos \varphi_2 \quad [1]$$

In case absorption has to be considered the refractive index  $n$  consists of a real part  $\delta$  and an imaginary part  $i\beta$ . The real part refers to the decrement of dispersion and the imaginary part to the attenuation:

$$n = 1 - \delta - i\beta \quad [2]$$

For X-rays, both  $\delta$  and  $\beta$  are zero in vacuum or air, but slightly positive for solids. Thus, any solid

medium does have a refractive index  $n$  smaller than one and is optically less dense than air (vacuum). Therefore, the refracted beam will be deflected toward the medium boundary and a critical angle will exist at which total reflection of the X-ray beam occurs. Using Snell's law (1), the critical angle  $\varphi_c$  can be calculated for a medium of known composition by using eqn [3]:

$$\varphi_c = \sqrt{(2\delta)} = (1.65E^{-1})\sqrt{ZA^{-1}\rho} \quad [3]$$

where  $E$  is the energy in keV,  $A$  the atomic mass in  $\text{g mol}^{-1}$ ,  $Z$  the atomic number, and  $\rho$  the density in  $\text{g cm}^{-3}$ . At the critical angle the reflectivity  $R$  of the material increases to almost 100% and the penetration depth  $z_p$  of the primary beam decreases drastically to only a few nanometers. Calculation of the reflectivity  $R$  can be done for any material by using the Fresnel formula. The penetration depth  $z_p$  refers to the depth at which the intensity of the primary beam decreases to  $e^{-1}$  or 37% of its initial value. Below the critical angle the reflectivity remains high and the penetration depth low. However, once the critical angle is exceeded both change, with penetration depth increasing and reflectivity decreases rapidly. The very low penetration depth of the incident beam at and below the critical angle makes TXRF spectrometry well suited for analysis of matter deposited on and embedded in the uppermost layers of a solid medium.

In Table 1, critical angles, reflectivity, and penetration depth of some common materials used in total reflection X-ray analysis are summarized for an excitation energy of 17.44 keV, corresponding to Mo K excitation. The table shows that the reflectivity for materials composed of light elements is high, but decreases for heavier elements. In the latter case, absorption effects become more prominent and affect the reflectivity negatively.

#### Standing Waves

TXRF is intimately connected to a standing wave field created by interference of incident and total

- Jenkins R, Gould RW, and Gedke D (1981) *Quantitative X-ray Spectrometry*. New York: Dekker.
- Potts PJ (1987) Energy dispersive X-ray spectrometry. *A Handbook of Silicate Rock Analysis*, ch. 9. Glasgow: Blackie.
- Smith JV (1994) Synchrotron X-ray microanalysis. In: Potts PJ, Bowles JFW, Reed SJB, and Cave MR (eds.)

- Microprobe Techniques in the Earth Sciences*, ch. 5. London: Chapman & Hall.
- Van Grieken RE and Markowicz AA (1992) *Handbook of X-ray Spectrometry*. New York: Dekker.
- Woldseth R (1973) *X-ray Energy Spectrometry*. Burlingame, CA: KeveX.

## Total Reflection X-Ray Fluorescence

**M Schmeling**, Loyola University Chicago, Chicago, IL, USA

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Total reflection X-ray fluorescence (TXRF) spectrometry is a trace elemental microanalysis technique based on conventional energy dispersive X-ray fluorescence. It has become increasingly popular in the last decade and is applied in almost every field of trace elemental analysis where low detection limits and multielement capabilities are required. Like all X-ray techniques, TXRF is nondestructive making it extremely useful and important in areas where samples are precious and/or need to be used for further characterization. New tabletop instruments make this technique affordable and more versatile as it can be used also for field research. In the semiconductor industry, TXRF is now routinely applied to scan wafers for impurities on the surface and in near-surface layers. The following article introduces the basic principle of TXRF and its instrumental features and discusses various applications of this technique.

### Basic Principles

#### Total Reflection, Reflectivity, and Penetration Depth

When an X-ray beam strikes the surface of a solid medium at a certain angle  $\varphi$  it can be refracted or reflected. According to Snell's law the refractive indices  $n$  of two media 1 and 2 are in the following relationship:

$$n_1 \cos \varphi_1 = n_2 \cos \varphi_2 \quad [1]$$

In case absorption has to be considered the refractive index  $n$  consists of a real part  $\delta$  and an imaginary part  $i\beta$ . The real part refers to the decrement of dispersion and the imaginary part to the attenuation:

$$n = 1 - \delta - i\beta \quad [2]$$

For X-rays, both  $\delta$  and  $\beta$  are zero in vacuum or air, but slightly positive for solids. Thus, any solid

medium does have a refractive index  $n$  smaller than one and is optically less dense than air (vacuum). Therefore, the refracted beam will be deflected toward the medium boundary and a critical angle will exist at which total reflection of the X-ray beam occurs. Using Snell's law (1), the critical angle  $\varphi_c$  can be calculated for a medium of known composition by using eqn [3]:

$$\varphi_c = \sqrt{(2\delta)} = (1.65E^{-1})\sqrt{ZA^{-1}\rho} \quad [3]$$

where  $E$  is the energy in keV,  $A$  the atomic mass in  $\text{g mol}^{-1}$ ,  $Z$  the atomic number, and  $\rho$  the density in  $\text{g cm}^{-3}$ . At the critical angle the reflectivity  $R$  of the material increases to almost 100% and the penetration depth  $z_p$  of the primary beam decreases drastically to only a few nanometers. Calculation of the reflectivity  $R$  can be done for any material by using the Fresnel formula. The penetration depth  $z_p$  refers to the depth at which the intensity of the primary beam decreases to  $e^{-1}$  or 37% of its initial value. Below the critical angle the reflectivity remains high and the penetration depth low. However, once the critical angle is exceeded both change, with penetration depth increasing and reflectivity decreases rapidly. The very low penetration depth of the incident beam at and below the critical angle makes TXRF spectrometry well suited for analysis of matter deposited on and embedded in the uppermost layers of a solid medium.

In Table 1, critical angles, reflectivity, and penetration depth of some common materials used in total reflection X-ray analysis are summarized for an excitation energy of 17.44 keV, corresponding to Mo K excitation. The table shows that the reflectivity for materials composed of light elements is high, but decreases for heavier elements. In the latter case, absorption effects become more prominent and affect the reflectivity negatively.

#### Standing Waves

TXRF is intimately connected to a standing wave field created by interference of incident and total

reflected beam near or at the critical angle. For a thick flat medium, superposition of two or more coherent and monochromatic X-ray beams leads to a triangular standing wave field with minima (nodes) and maxima (antinodes) in front of the medium. **Figure 1** shows such a standing wave field schematically in front of a thick substrate. For a reflectivity of  $\sim 100\%$ , the amplitude of an incoming and a total reflected wave can be subtracted to zero (nodes) or added to a maximum value (antinodes). Minima and maxima follow each other with a period of  $d_{\min/\max} = (\lambda/2)/\varphi$ . Here,  $\lambda$  is the wavelength and  $\varphi$  the angle of the incident beam. A layer or sample deposited on top of the thick flat substrate is subjected to this standing wave field and is excited to fluorescence in proportion to the wave field intensity. It is therefore important that the sample is positioned well inside the standing wave field to gain maximum fluorescence of the sample. In addition, the sample should be granular or uneven, otherwise total reflection might occur inside the deposit. Too large samples should be avoided as well because absorption effects might interfere.

When a thin enough layer is present on a thick substrate, an additional standing wave field can be

observed in the thin layer itself beside the one in front of the substrate. When these conditions are fulfilled it is possible to probe surface layers opening up the field for surface and thin layer characterization.

### Fluorescence Intensity of Samples Deposited on Top of a Substrate and Below the Surface

A sample deposited on a thick flat substrate is exposed to both the incident and total reflected beams. Since the reflectivity  $R$  is  $\sim 100\%$  below and at the critical angle, the fluorescence intensity  $I$  is almost doubled according to eqn [4]:

$$I = I_0(1 + R) \quad [4]$$

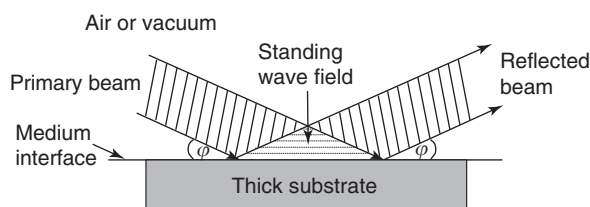
where  $I_0$  is the intensity of the incoming X-ray beam at glancing angles. For a granular residue, the fluorescence intensity is nearly constant below and at the critical angle. **Figure 2A** illustrates that case. The background signal for such a sample is very low as the incident beam only penetrates a few nanometers into the substrate and does not contribute significantly to the background signal. For example, for quartz, silicon, or Plexiglas the background signal is about six orders of magnitude smaller than for conventional X-ray fluorescence. This increase in fluorescence intensity is responsible for the high detection power and correspondingly low detection limits of TXRF.

In **Figure 2B**, the fluorescence intensity at different glancing angles for atoms embedded in a near-surface layer is shown. As seen the fluorescence intensity is clearly angle dependent and experiences a maximum at the critical angle. This feature can be used to probe surface layers at different angles around the critical angle and obtain useful information about their composition and thickness. It is important to note that the layer is nonreflective and homogenous in composition. The angle dependency of the fluorescence intensity permits depth profiling of layered structures as well. The instrumental set-up for these applications is slightly different then for samples deposited on the substrate as it requires a mechanism to change the incident angle in small steps. Surface layer characterization is pertinent in the semiconductor industry where contaminations on wafers have to be determined to avoid malfunction of devices.

**Table 1** Critical angles  $\varphi_c$ , reflectivity  $R$ , and penetration depth  $z_p$  at the critical angle for selected materials calculated for 17.44 keV (Mo K $\alpha$ ) incident beam energy<sup>a</sup>

Material	$\varphi_c$ (°)	$R$ (%)	$z_p$ (nm)
Plexiglas	0.076	93.2	241
Glassy carbon	0.08	93.9	255
Boron nitride	0.1	93.3	188
Quartz Glass	0.1	85.5	83
Aluminum	0.11	82.9	64
Silicon	0.1	81.5	62
Cobalt	0.19	59.1	12.7
Nickel	0.2	58.1	12.1
Copper	0.19	56.1	11.5
Germanium	0.15	51.2	13.1
Gallium arsenide	0.15	51.1	13.0
Platinum	0.28	39.4	4.8
Gold	0.26	38.7	5.0

<sup>a</sup>After Klockenkämper R (1997) *Total Reflection X-Ray Fluorescence Analysis*. New York: Wiley.



**Figure 1** Standing wave field above a thick flat substrate.  $\varphi$  refers to the grazing angle of the incident and total reflected beams.

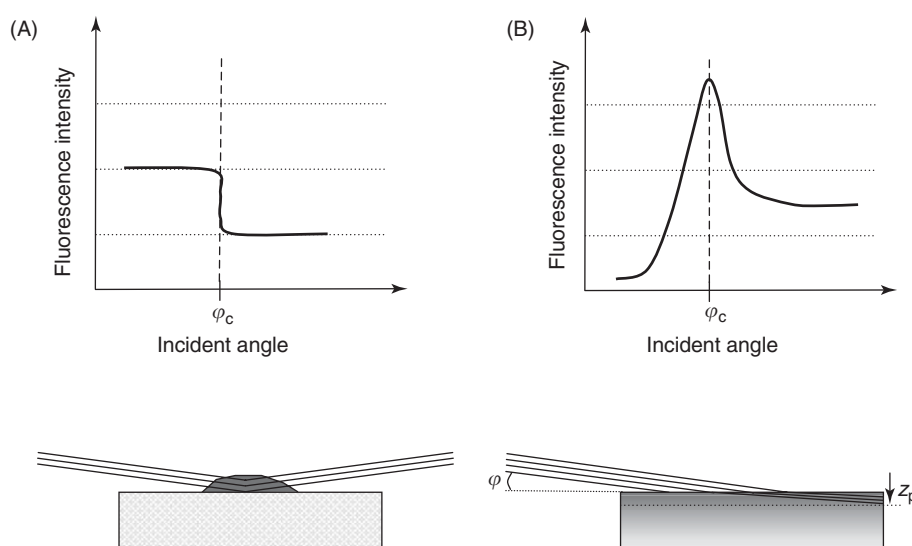
## Instrumentation

The general instrumental set-up for TXRF spectrometry is quite simple with an X-ray source, a reflector or monochromator, a second reflector, acting as sample support or sample itself, and a detector. A low pass filter to reduce the high-energy part of the X-ray spectrum has to be inserted in the path of the

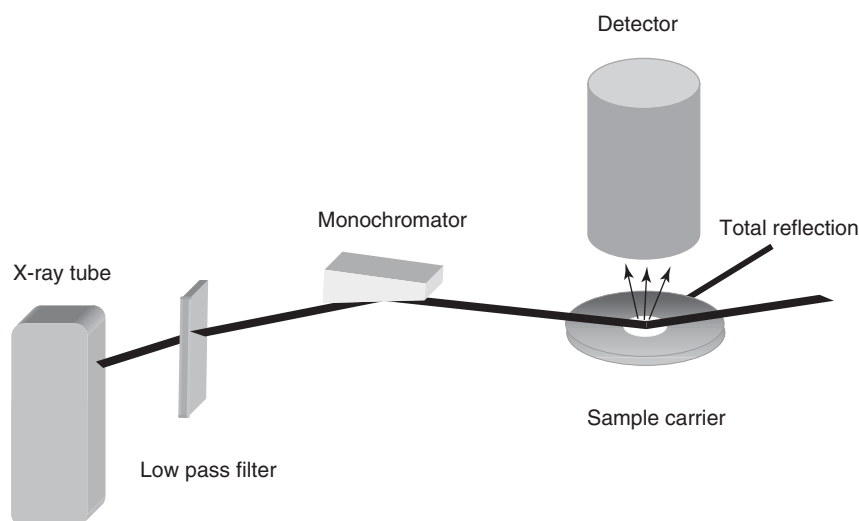
primary beam to avoid a high background signal due to scattered photons. **Figure 3** shows such a basic instrumental set-up. Conditions for successful excitation of the sample under total reflection comprise a monochromatic well-collimated primary X-ray beam with a divergence of less than 0.2 mrad. The beam should be shaped like a strip of paper having a width of  $\sim 10$  mm and a height of  $10\text{ }\mu\text{m}$  to penetrate only a small area of the substrate. This is necessary because the detector has only a small window less than 1 cm in diameter. The beam can be shaped precisely by using slits or diaphragms in the optical path. When analysis of small deposits on a thick substrate is performed the incident angle should be less than the

critical angle of total reflection, i.e.,  $\sim 70\%$  of that angle. In most cases this corresponds to an incidence angle of  $\sim 0.06\text{--}0.07^\circ$  for the primary beam. In some special cases when light elements with atomic numbers less than 12 are of interest the optical path, including sample support (or second reflector), has to be under vacuum conditions. This translates to attaching the X-ray source and the detector with a flange to a vacuum chamber in which the sample and optics will be placed. A similar approach is used when working with synchrotron radiation as excitation source.

For surface analysis the instrument should be equipped with a mechanism permitting the variation



**Figure 2** Fluorescence intensity in dependency of the incident angle  $\varphi$ . For (A) a deposit on a flat thick substrate and (B) atoms present in a surface layer;  $z_p$  refers to the penetration depth of the incident beam.



**Figure 3** Basic set-up for a total reflection X-ray fluorescence spectrometer.



of the sample position in small steps around the critical angle of the incident beam.

### Excitation Devices

A high-intensity X-ray beam is necessary to excite the atoms in the sample to fluorescence. For trace- and microanalysis of deposits fine-focus X-ray tubes are sufficient, limiting already the size of the primary beam. Common anode materials are molybdenum, tungsten, or chromium, with molybdenum providing the widest energy range for elemental detection without change of the tube conditions. Chromium or scandium anodes are also recommended for detection of light elements in combination with a vacuum chamber. When performing surface analysis rotating anodes are required to satisfy the demand for lower detection limits. Rotating anodes show higher brilliance and better intensity, thus improving the detection power. In recent years, synchrotron radiation has become more easily available. Its high brilliance and polarization allows for extremely low detection limits. In addition, the energy of the incident beam can be tuned with a multilayer crystal. The instrumental set-up resembles closely the vacuum version, except that it has to be tilted by  $90^\circ$  owing to the polarization of the synchrotron beam in the horizontal plane. Several beam lines at different synchrotron facilities are now available for TXRF analysis.

### Monochromators

The polychromatic X-ray beam has to be monochromatized to be efficient in excitation of the sample. It is necessary to eliminate the high-energy part of the beam that would be scattered in the sample otherwise causing an increase in the background signal. In trace- and microanalysis a highly polished quartz prism is sufficient, cutting off the high-energy part of the spectrum under grazing incidence by acting as a total reflecting mirror. In addition, a thin metal foil can be inserted into the beam path acting as attenuator for certain energies. For surface and thin layer analysis the X-ray beam has to be strictly monochromatic to obtain good intensity profiles in dependency of the glancing angle. X-ray photons of different energies would interfere with this process and obstruct the angle scan. In this case natural crystals or multilayers are required as monochromatizing units. When natural crystals are used the spectral selectivity is higher but the intensity of the incident beam decreases. To compensate for the loss in intensity rotating anode devices have to be employed. Multilayers, on the other hand, show less spectral selectivity but permit for a higher intensity. One possibility to increase the selectivity for the latter set-up is the use of

two multilayers arranged in parallel position to guide the beam in a zigzag path toward the sample.

### Sample Positioning and Detectors

Utmost care has to be taken with respect to the sample positioning in a TXRF unit. Positioning of the sample needs to be extremely precise as the incident angle is very small, less than  $0.1^\circ$ . For trace- and microanalysis, a fixed sample position is realized with a precise positioning mechanism placing the sample exactly at the same spot for each analysis. In surface and thin layer analysis, sample positioning is even more crucial as the sample has to be moved with respect to the critical angle. For this the sample is deposited on a movable table, which can be rotated, tilted, and adjusted in the horizontal  $x,y$  direction in small steps. The sample support should be optically flat to permit total reflection of the primary beam. The surface roughness should be restricted to less than 5 nm and the waviness should not exceed  $0.001^\circ$ . The sample support has to be free of contamination and chemically inert when performing trace analysis. Materials meeting these requirements and most commonly used are quartz, Plexiglas or polycarbonate, glassy carbon, silicone, and sapphire. Some of them can be cleaned and reused for many analyses without loss in their properties. Detection of the spectra is carried out by Si(Li) detectors located directly on top of the sample at  $90^\circ$  angle and at a distance of a few millimeters. This ensures that the maximum amount of fluorescence radiation is captured by the detector and in addition minimizes the absorption of soft X-rays in the space between the detector and the sample. For analysis of light elements a high-purity germanium detector is recommended, having a better spectral resolution in the low-energy range.

### Analytical Performance

TXRF as a multielement microtechnique is capable of analyzing trace quantities or even ultratrace quantities in a sample. The sample has to be small with a mass of no more than a few micrograms and thin ( $\sim 10\text{--}15\ \mu\text{m}$  thickness) to permit penetration of the primary beam into the substrate for total reflection. Matrix effects are few or nonexistent due to the small sample size, thus eliminating absorption effects in the sample. In most cases, some sample preparation is required to obtain a homogeneous analyte and to restrict it in mass and size. Sometimes, however, a sample can be directly analyzed. This might be the case for homogenous powders of very fine grain sizes or thin sections of biological tissues. Owing to its

surface sensitivity TXRF can be applied for surface and thin layer analyses. This is of particular interest for determination of impurities or contaminations on wafers. In this case, variation of the glancing incident angle provides angle profiles of intensity levels. The detection limits are very low and comparable to most common analytical techniques used in atomic spectroscopy or surface characterization.

### Calibration and Quantification

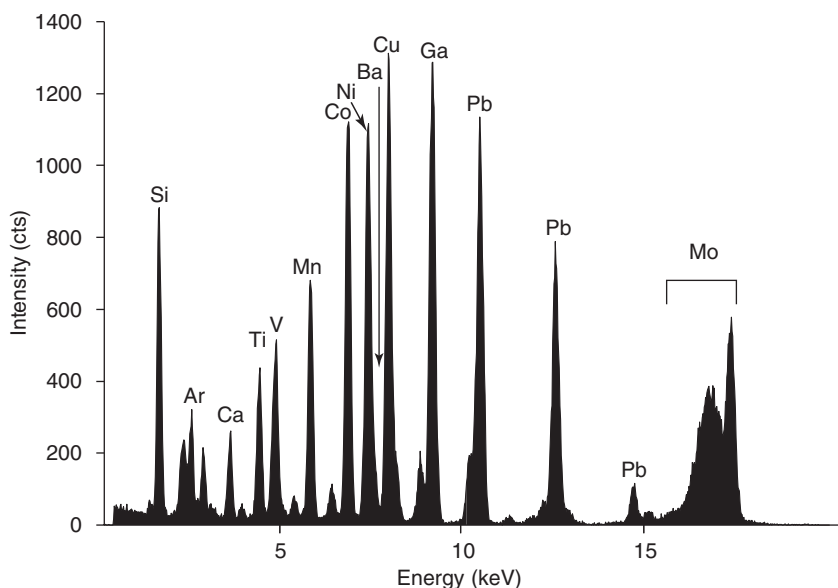
Before quantitative analysis is possible the system has to be calibrated. For this, elemental profiles of each element are measured to obtain their net intensities. This includes preparing of single element standards with a concentration of  $1\text{--}10\ \mu\text{g l}^{-1}$  in ultraclean water and pipetting  $50\text{--}100\ \mu\text{l}$  of this solution on a sample carrier in several steps. The spectrum is recorded and added to the database. Such calibration needs to be done only once for each instrument as long as the components are not altered or exchanged in the equipment. More recent instruments are equipped with a software package enabling the calculation of the net intensities of peaks by using certain algorithms. This has the advantage that only few element profiles need to be measured.

Graphing of elemental concentration versus measured net intensity  $N$  shows a straight line with a slope representing the absolute sensitivity of this element. In Figure 4, intensities of some elements are shown. The relative sensitivity  $S_x$  for an element  $x$  is another important parameter and can be determined by the following equation:

$$S_x = (N_x c_x^{-1})(N_{\text{ref}} c_{\text{ref}}^{-1})^{-1} S_{\text{ref}} \quad [5]$$

where  $N_x$  and  $N_{\text{ref}}$  refer to the net intensity for element  $x$  and reference element (ref), and  $c_x$  and  $c_{\text{ref}}$  to the concentration of these. Since relative sensitivities are determined  $S_{\text{ref}}$  can be set to 1 by simplifying the calculation of the relative sensitivity for each element. It is also possible to obtain relative sensitivities mathematically without intensity measurements. Several factors are incorporated in the equation including fluorescence yield and relative emission rate. When elemental standards are not available for certain elements the relative sensitivity can be extrapolated or calculated by applying the above-mentioned equations.

Quantification is easy by adding an internal standard of known concentration to the sample either before sample preparation or when analyzing powder samples as a droplet on top of the powder. The internal standard consists of an element not present in the sample, often gallium or yttrium, sometimes germanium when the sample is present in basic solution, and should be somewhere in the medium concentration range of the elements to be determined. Care has to be taken when elements have to be quantified at ultratrace levels, which are located in the spectrum adjacent to the internal standard. Here the standard concentration should be as low as possible and in the best case adjusted to the expected element concentration to avoid peak shadowing. A 'blank run' without standard helps to estimate the concentration level of the element. When sample digestion is required the standard should be added before the digestion step to monitor losses. For powder samples the standard ought to be pipetted onto the sample itself and dried for analysis. Elemental concentrations can be



**Figure 4** Intensity profiles for selected elements (Mo K excitation, 20 ng amount). It can be seen that Ba and Ni overlap slightly.

calculated easily by using the following:

$$c_x = (N_x S_x^{-1})(N_{is} S_{is}^{-1})^{-1} c_{is} \quad [6]$$

where  $c_x$  and  $c_{is}$  refer to the concentrations of the element to be determined and the internal standard (known), respectively,  $N_x$  and  $N_{is}$  to their net intensities, and  $S_x$  and  $S_{is}$  to their relative sensitivities.

Calibrations in surface and thin layer analysis have to be performed differently by measuring external standards of pure elements, analyzing dried residues of a standard element, or after spin coating of a support or wafer with a spiked solution. Internal standardization is not suitable for quantification as angle variation of the glancing incident beam does alter the fluorescence intensity. Peak fitting and quantification has to be carried out by using the fundamental parameter method as known for conventional XRF spectrometry.

### Detection Limits

The detection limits for TXRF are significantly lower than for conventional XRF and comparable to most atomic spectrometric methods used for trace analysis or surface characterization. They range from several picograms to a few nanograms for the lighter elements in real samples and are somewhat lower for pure aqueous solutions. Improvement of detection limits is possible when the sample is pretreated and/or the matrix removed. This is particularly important

when analyzing organic, biological, or environmental samples. For surface contaminations the detection limits are  $\sim 10^9$  atoms  $(\text{cm}^2)^{-1}$  for transition metals by direct determination on the surface and can be improved to some  $10^7$  atoms  $(\text{cm}^2)^{-1}$  using special pretreatment techniques like vapor phase deposition. When synchrotron radiation is available as excitation source the detection limits can be decreased further for direct determination and are reported to be  $3 \times 10^8$  atoms  $(\text{cm}^2)^{-1}$  for Ni directly on a wafer. Table 2 shows the range of detection limits for commonly determined elements.

## Applications

The applications of TXRF spectrometry are manifold and frequently more are added. Table 3 summarizes applications in different fields. Generally, there are two types of applications: trace and microanalysis as well as surface and thin layer analysis. In both cases special care has to be taken so that sample preparation and/or analysis is done in very clean environments to avoid possible contamination.

### Trace- and Microanalysis

Almost any sample can be analyzed by TXRF spectrometry for its trace element content either directly or after some sample preparation steps. Such steps might include preconcentration, digestion, or matrix separation. Sometimes the samples do not need to be

**Table 2** Detection limits for commonly analyzed elements

< 2 pg	2–5 pg	5–10 pg	10–100 pg	> 100 pg
Ni, Cu, Zn, Ga, As, Se	Mn, Fe, Co, Br, Rb, Pb, Bi	Ti, V, Cr, Sr, Y, Zr, Mo, Rb	P, S, Ag, Cd, Sn, Ba, La	Al

**Table 3** Scope of applications

Biomedical	Environmental	Industrial	Other
<b>Body fluids</b> Whole blood, serum, urine, amniotic fluids, enzymes	<b>Plant materials</b> Lichen, hay, algae, leaves, moss, needles, wood chips, Daphnia	<b>High-purity chemicals</b> Acids, water, solvents, salts	<b>Forensics</b> Hair, nails, skin, fibers
<b>Tissue</b> Lung, liver, kidney	<b>Water</b> Tap water, sea water, waste water, river water, snow <b>Particulate matter</b> Aerosols, fly ash, dust  <b>Other</b> Mud, sediment, sewage, sludge, peat	<b>Metals and Ores</b> Aluminum, iron, steel, ores, rocks, powders <b>Oils and greases</b> Crude oil, vegetable oil, essential oil, fat, grease, cream <b>Thin deposits</b> Films, foils, layers, residues	<b>Art</b> Pigments, oil paints, inks, varnishes <b>Food</b> Wine, beer, mineral water, fish, mussel tissue, vegetables, fruits, nuts

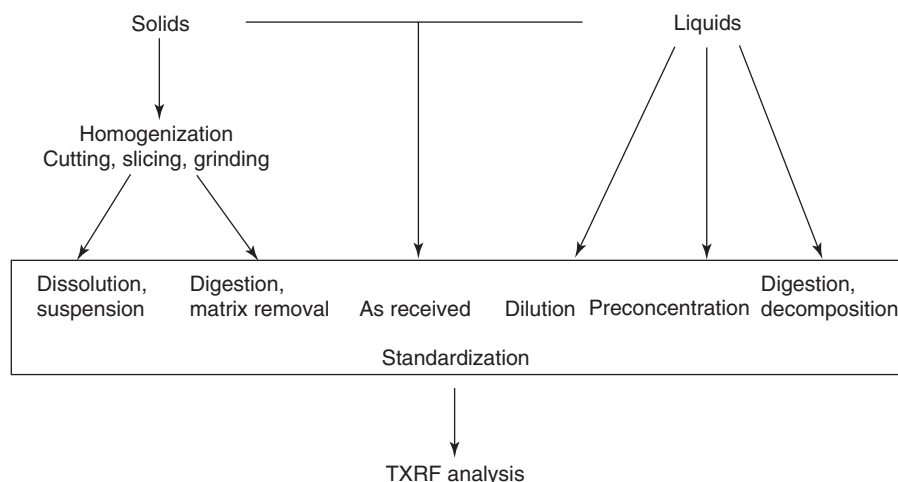
pretreated but can be analyzed directly, which is the case for drinking or tap water, mineral waters, rain water, and even atmospheric particulates, when deposited directly on a reflector. An internal standard ensures quantification. Biological and medicinal materials generally need to be pretreated to remove the organic matrix that would increase the background signal and thus worsen the detection limits. Similar steps should be considered for environmental samples like soil or river and sea water. Microwave digestion with concentrated nitric acid or high-pressure digestion in specially designed polytetrafluoroethylene bombs are the most common pretreatment methods. When silicates are present in a sample some quantity of hydrofluoric acid can be added to ensure total dissolution. Oil samples can be diluted with a solvent or subjected to cold oxygen plasma ashing if necessary. In all cases the internal standard should be added before the sample preparation steps are performed to avoid errors due to element losses. Tissue samples can be analyzed as well when sliced with a microtome into small and thin sections and directly placed on the sample carrier or reflector. The general procedure of trace and ultratrace analysis is been summarized in **Figure 5**. With or without pretreatment the sample is deposited on the sample carrier either as a solid or as liquid in small droplet form of  $\sim 20\text{--}50\text{ }\mu\text{l}$  including the internal standard. For a solid sample the analysis can be performed directly whereas the liquid has to be dried either under infrared light or on a hot plate before analysis. Counting times range between 100 and 1000 s depending on the concentrations to be determined. **Figure 6** displays the spectrum of an atmospheric aerosol sample after microwave digestion with nitric acid. Here gallium was added as internal standard.

### Surface and Thin Layer Analysis

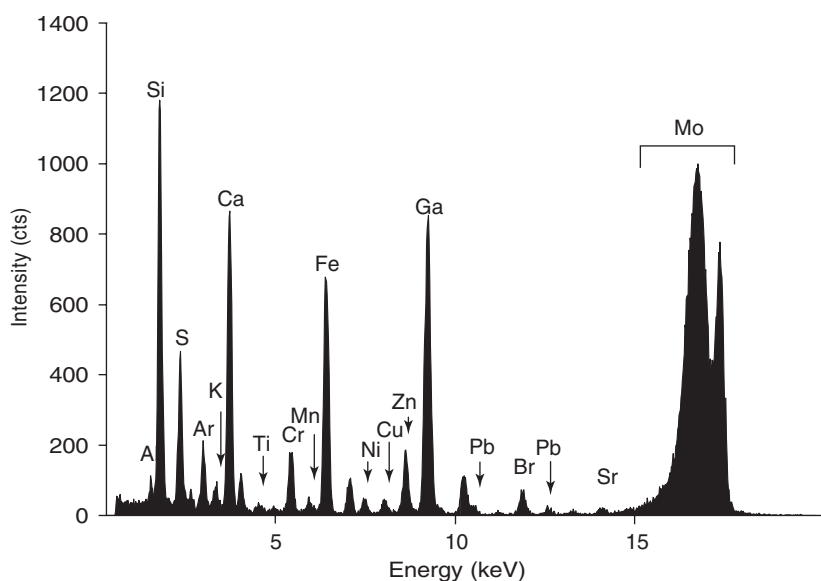
Surface analysis is possible when the samples are even and optically flat. Ideally suited for this are wafers as produced and used in the semiconductor industry. Both uncoated and coated wafers can be characterized. For uncoated wafers surface analysis is performed, whereas thin layer analysis is applied for coated wafers. In both cases intensity profiles in dependency of the incident angle around the critical one are recorded. This provides information about the elemental composition and thickness of the surface layer. When scanning at different locations of the wafer a surface map of the wafer can be obtained. In practice, some spots are selected and angle scans performed. It is possible to distinguish between different types of contamination – bulk, particulate type deposited on the surface, or thin layer type – on the wafer by comparing the shape of the angle scan with calibration samples. **Figure 7** shows these three types of contamination schematically. In thin layer analysis, layer parameter can be evaluated via iterative peak fitting procedures utilizing the angle scan. The detection limits for wafer analysis can be improved when the wafer is subjected to vapor phase decomposition (VPD) treatment. In this case the surface is slightly etched by hydrofluoric acid, and silica ( $\text{SiO}_2$ ) or silicon (Si) is removed as hexafluorosilicic acid ( $\text{H}_2\text{SiF}_6$ ). The reaction also produces water droplets in which the surface contaminants are trapped. These droplets are dried and analyzed by TXRF. Special chambers for VPD are usually employed.

### Recent Developments

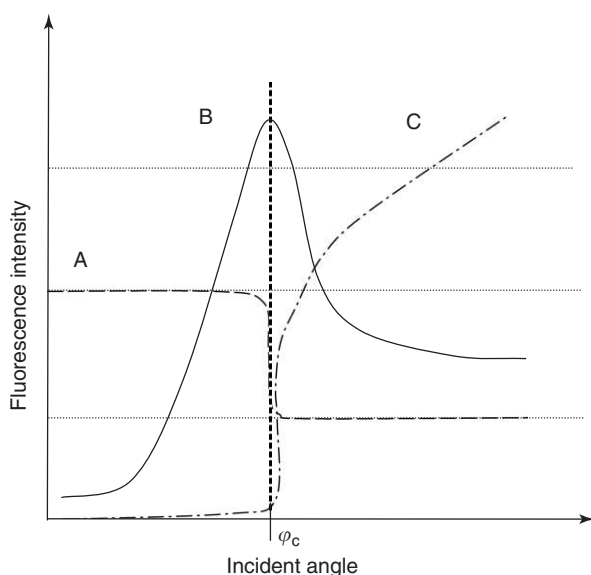
Besides the more and more popular and accessible use of synchrotron radiation for TXRF, two variants



**Figure 5** General sample preparation steps for TXRF analysis.



**Figure 6** Spectrum of an urban aerosol sample taken in Chicago. Gallium was added as internal standard with the concentration of  $100 \mu\text{g l}^{-1}$ .



**Figure 7** Intensity profiles for different types of contaminations: (A) particulate type deposit; (B) thin film type deposit, and (C) bulk substrate.

of the technique have emerged in the recent years. One is grazing exit X-ray fluorescence (GE-XRF) and the other grazing incidence and exit X-ray fluorescence (GIE-XRF). In GE-XRF, the instrumental setup has been flipped, having a  $90^\circ$  incident angle, but employing a grazing take-off or exit angle. In GIE-XRF, both angles, the incident and the take-off angle of the beam, are grazing or close to zero. GE-XRF is often employed for nondestructive surface and thin layer analysis in combination with synchrotron radiation. It has certain advantages

compared to traditional TXRF like permitting a polychromatic incident beam with a lateral resolution of  $\sim 10 \mu\text{m}$ . Surface mapping becomes much faster and easier. In addition, a wavelength dispersive detector can be used which has a higher spectral resolution and thus shows a better peak separation. The disadvantage is that absorption effects can occur and matrix effects make quantification difficult. GIE-XRF is less common and more complicated to set up, but has the advantage that both the incident and the take-off angles can be adjusted. This allows for analysis of thin films and vertical concentration profiles.

**See also:** **Atomic Absorption Spectrometry:** Principles and Instrumentation. **Atomic Emission Spectrometry:** Inductively Coupled Plasma. **Sample Dissolution for Elemental Analysis:** Wet Digestion; Microwave Digestion. **Surface Analysis:** Secondary Ion Mass Spectrometry of Polymers.

## Further Reading

- Born M and Wolf E (1997) *Principles of Optics*, 6th edn. Cambridge: Cambridge University Press.
- Hockett RS (1995) A review of standardization issues for total reflection X-ray fluorescence and vapor phase total reflection X-ray fluorescence. *Analytical Sciences* 11: 511–513.
- Klockenkämper R, Knoth J, Prange A, and Schwenke H (1992) Total-reflection X-ray fluorescence spectroscopy. *Analytical Chemistry* 64: 1115A–1121A.
- Strelcić C (1997) Total reflection X-ray fluorescence analysis of light elements. *Spectrochimica Acta* 52B: 281–293.

- Tölg G and Klockenkämper R (1993) The role of total-reflection X-ray fluorescence in atomic spectroscopy. *Spectrochimica Acta* 48B: 111–127.
- Tsuji K, Sato S, and Hirokawa K (1995) Depth profiling using the glancing-incidence and glancing-take off X-ray fluorescence method. *Review of Scientific Instruments* 66: 4847–4852.

- Van Grieken RE and Markowicz AA (eds.) (2002) *Handbook of X-ray Spectrometry*, 2nd edn. New York: Marcel Dekker.
- Yoneda Y and Horiuchi T (1971) Optical flats for use in X-ray spectrochemical microanalysis. *Review of Scientific Instruments* 42: 1069–1070.

## Particle-Induced X-Ray Emission

**W Maenhaut**, Ghent University, Ghent, Belgium  
**F Adams**, University of Antwerp, Antwerp, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Particle-induced X-ray emission (PIXE) is an elemental analysis technique that employs a beam of energetic heavy charged particles (usually protons of 1–4 MeV) to induce element-specific X-ray emission. Depending on the sample material and thickness and on the conditions of the analysis, the technique offers detection limits at the  $0.1\text{--}10\text{ mg kg}^{-1}$  concentration level for low- $Z$  matrices. Applied in nuclear microprobes, micro-PIXE combines a high sensitivity with the possibility of providing elemental maps with a lateral resolution in the sub-micrometer range.

### Basic Principles

PIXE was introduced in 1970, combining two technical developments into a novel variant of X-ray emission analysis with interesting capabilities. One innovation was the use of lithium-drifted (Si(Li)) detectors, which allow the simultaneous detection of the X-ray intensities of (nearly) all elements in the sample. The other novelty was that X-ray emission was induced by a proton beam of MeV energy instead of by an electron beam or bremsstrahlung radiation from an X-ray tube. The proton beam ejects inner-shell electrons from the atoms in the bombarded specimen (sample), the resulting vacancies are virtually instantaneously filled by outer-shell electrons, and this de-excitation process gives rise to the emission of X-rays whose energies are characteristic for the excited elements. Although the excitation mechanism is analogous to that in the electron microprobe, the much smaller deceleration of the protons (or other ions) in the specimen leads to a much lower bremsstrahlung continuum. The typical PIXE spectrum shown in Figure 1 consists of

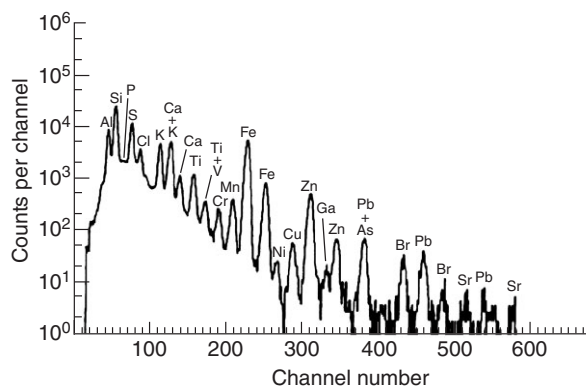
characteristic X-ray peaks superimposed on a continuous background. Whereas the former result from the de-excitation of inner-shell vacancies, the background arises from several processes, especially secondary electron bremsstrahlung (SEB).

When a beam of heavy charged particles of a few MeV penetrates into matter, it loses its energy gradually with depth, until it is finally stopped. The energy loss occurs mainly through inelastic Coulomb interactions with bound electrons and in contrast to the case of electron beams, the direction of travel of an ion beam is scarcely altered during the slowing down process. The stopping power  $S(E)$  of an ion with energy  $E$  is defined as the energy loss per unit mass thickness traversed:

$$S(E) = -\frac{1}{\rho} \frac{dE}{dx} \quad [1]$$

where  $\rho$  is the density of the stopping material and  $x$  the distance. As defined here,  $S(E)$  is expressed in  $\text{keV g}^{-1} \text{cm}^2$ . The total path length  $R$  for an ion with incident energy  $E_0$  is given by:

$$R(E_0) = \int_{E_0}^0 \frac{dE}{dE/dx} \quad [2]$$



**Figure 1** PIXE spectrum of an urban aerosol sample recorded using a 'funny filter'. (Reproduced with permission from Campbell *et al.* (1986) *Nuclear Instruments and Methods in Physics Research B* 14: 204–220; © Elsevier.)



Tölg G and Klockenkämper R (1993) The role of total-reflection X-ray fluorescence in atomic spectroscopy. *Spectrochimica Acta* 48B: 111–127.

Tsuji K, Sato S, and Hirokawa K (1995) Depth profiling using the glancing-incidence and glancing-take off X-ray fluorescence method. *Review of Scientific Instruments* 66: 4847–4852.

Van Grieken RE and Markowicz AA (eds.) (2002) *Handbook of X-ray Spectrometry*, 2nd edn. New York: Marcel Dekker.

Yoneda Y and Horiuchi T (1971) Optical flats for use in X-ray spectrochemical microanalysis. *Review of Scientific Instruments* 42: 1069–1070.

## Particle-Induced X-Ray Emission

**W Maenhaut**, Ghent University, Ghent, Belgium

**F Adams**, University of Antwerp, Antwerp, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Particle-induced X-ray emission (PIXE) is an elemental analysis technique that employs a beam of energetic heavy charged particles (usually protons of 1–4 MeV) to induce element-specific X-ray emission. Depending on the sample material and thickness and on the conditions of the analysis, the technique offers detection limits at the  $0.1\text{--}10\text{ mg kg}^{-1}$  concentration level for low- $Z$  matrices. Applied in nuclear microprobes, micro-PIXE combines a high sensitivity with the possibility of providing elemental maps with a lateral resolution in the sub-micrometer range.

### Basic Principles

PIXE was introduced in 1970, combining two technical developments into a novel variant of X-ray emission analysis with interesting capabilities. One innovation was the use of lithium-drifted (Si(Li)) detectors, which allow the simultaneous detection of the X-ray intensities of (nearly) all elements in the sample. The other novelty was that X-ray emission was induced by a proton beam of MeV energy instead of by an electron beam or bremsstrahlung radiation from an X-ray tube. The proton beam ejects inner-shell electrons from the atoms in the bombarded specimen (sample), the resulting vacancies are virtually instantaneously filled by outer-shell electrons, and this de-excitation process gives rise to the emission of X-rays whose energies are characteristic for the excited elements. Although the excitation mechanism is analogous to that in the electron microprobe, the much smaller deceleration of the protons (or other ions) in the specimen leads to a much lower bremsstrahlung continuum. The typical PIXE spectrum shown in Figure 1 consists of

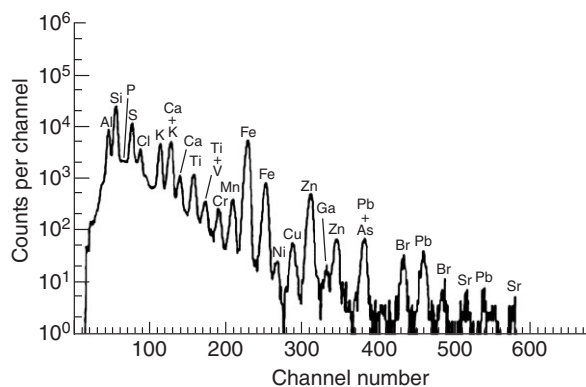
characteristic X-ray peaks superimposed on a continuous background. Whereas the former result from the de-excitation of inner-shell vacancies, the background arises from several processes, especially secondary electron bremsstrahlung (SEB).

When a beam of heavy charged particles of a few MeV penetrates into matter, it loses its energy gradually with depth, until it is finally stopped. The energy loss occurs mainly through inelastic Coulomb interactions with bound electrons and in contrast to the case of electron beams, the direction of travel of an ion beam is scarcely altered during the slowing down process. The stopping power  $S(E)$  of an ion with energy  $E$  is defined as the energy loss per unit mass thickness traversed:

$$S(E) = -\frac{1}{\rho} \frac{dE}{dx} \quad [1]$$

where  $\rho$  is the density of the stopping material and  $x$  the distance. As defined here,  $S(E)$  is expressed in  $\text{keV g}^{-1} \text{cm}^2$ . The total path length  $R$  for an ion with incident energy  $E_0$  is given by:

$$R(E_0) = \int_{E_0}^0 \frac{dE}{dE/dx} \quad [2]$$



**Figure 1** PIXE spectrum of an urban aerosol sample recorded using a 'funny filter'. (Reproduced with permission from Campbell *et al.* (1986) *Nuclear Instruments and Methods in Physics Research B* 14: 204–220; © Elsevier.)

**Table 1** Range  $R$  and stopping powers  $S(E)$  for 2.5 MeV protons in several solids

Sample material	Atomic number	Penetration range, $R$ ( $\mu\text{m}$ )	$S(E)$ ( $\text{keV mg}^{-1} \text{cm}^2$ )
C	6	55	122.9
Si	14	68	98.2
Fe	26	27	74.9
Ag	47	28	56.1
Pb	82	37	40.6

Adapted from Johansson SAE and Campbell JL (1988) *PIXE: A Novel Technique for Elemental Analysis*. Chichester: Wiley.

As an illustration, for 2.5 MeV protons impinging on various materials, Table 1 lists numerical values for  $S(E)$ , and  $R$ .

### Ionization Processes

The primary purpose of bombarding a specimen with protons or heavier ions in PIXE is to eject bound electrons from the K or L atomic shells. The ejection can also be achieved by means of electrons or photons with energy in the 1–100 keV range. The vacancies will de-excite within  $10^{-16}$  s with the emission of characteristic radiation or Auger electrons or both; the probability of the radiative relaxation is the fluorescence yield  $\omega$ . For low- $Z$  elements ( $Z < 30$ ) and for Kvacancies, the production of Auger electrons is more probable than emission of characteristic radiation; in the case of Lshell vacancies, this is always the case. Nevertheless, for elements with  $Z < 20$ , protons are more efficient than photons or electrons for producing characteristic X-ray. For heavier elements, photon-induced X-ray emission (i.e., X-ray fluorescence, XRF) is more efficient.

### Processes Contributing to the Background

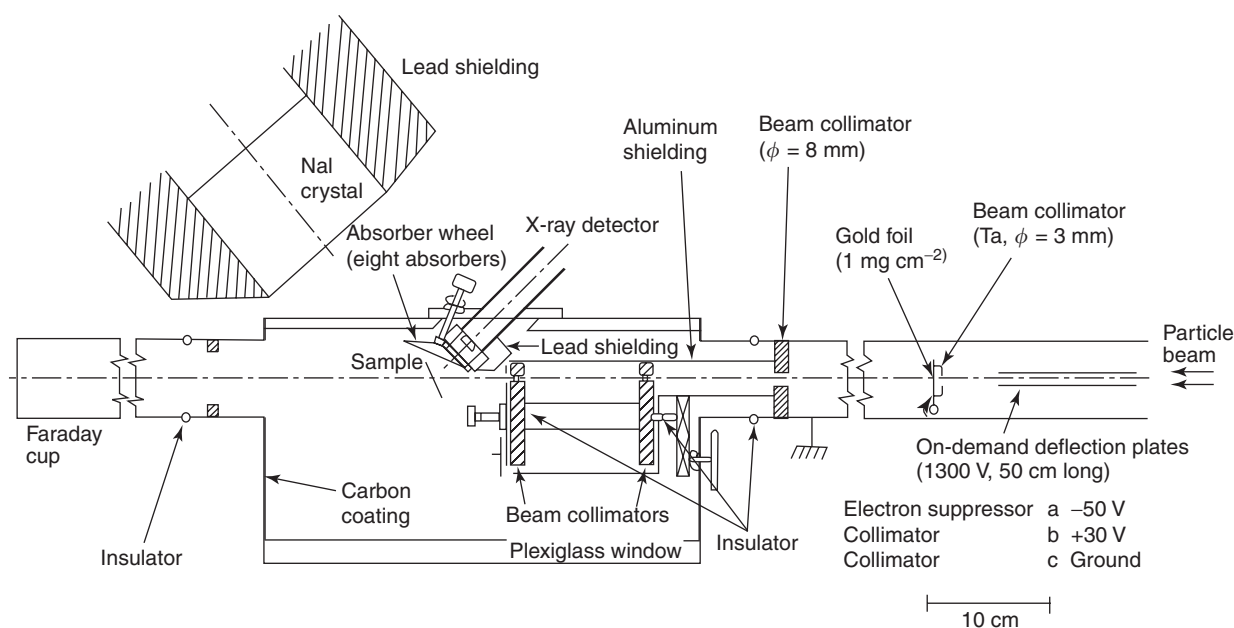
In general, when an electrical charge changes speed or direction, electromagnetic radiation is emitted. The deceleration experienced by the projectiles in PIXE and in electron-induced X-ray emission results in emission of bremsstrahlung photons. In electron-induced X-ray emission, this phenomenon provides the main contribution to the background and has the effect that the limits of detection (LODs) are at the percentage level. Since the bremsstrahlung intensity is inversely proportional to the square of the mass of projectile, the bremsstrahlung background is much smaller with protons than with electrons. As a consequence, other processes prevail over projectile bremsstrahlung in PIXE. Of particular

importance is the SEB; it arises from the electrons that are ejected from the specimen atoms in the slowing down of the protons. The SEB provides a very intense contribution to the background in the PIXE spectrum up to a maximum energy, but it becomes unimportant at higher energies. At high energies, the largest contribution to the background comes from  $\gamma$ -rays. These are produced by nuclear reactions in the specimen; mainly  $(p, \gamma)$ ,  $(p, p'\gamma)$ , and  $(p, \alpha\gamma)$  reactions with light elements ( $Z < 20$ ). The  $\gamma$ -ray background may become quite important when heavier ions (e.g., helium, oxygen) are used as projectiles in PIXE. With protons of a few MeV, the nuclear reaction background prevails over the proton bremsstrahlung in the energy region above 10 keV in the PIXE spectrum. Accordingly, in most PIXE setups, protons are preferred to heavier ions and used at energies between 2 and 3 MeV. Incident energies below 2 MeV are not advisable because of the decrease in ionization cross-sections.

### Instrumentation

Most PIXE work is done with proton beams of 2–3 MeV, which are provided by a small Van de Graaff accelerator. A typical arrangement of a beam line and specimen chamber for conventional broad-beam PIXE (macro-PIXE) is shown in Figure 2. Whereas the beam line is generally assembled from commercially available parts, the specimen chamber usually is not. Consequently, there is a wide variety of configurations. The pressure in the beamline and specimen chamber is typically in the range of  $10^{-6}$ – $10^{-4}$  Pa. High vacuum conditions are not crucial for macro-PIXE, but are very much so in micro-PIXE, where beam defocusing by gas molecules must be avoided.

In macro-PIXE, the proton beam has to be made uniform, defined by collimators (diaphragms), directed onto the specimen and integrated for calibration and monitoring purposes. At the same time, the X-ray and  $\gamma$ -ray background near the X-ray detector must be kept to a minimum. Accordingly, for collimators, specimen holders, Faraday cups, etc., materials with a low  $\gamma$ -ray yield under proton bombardment such as tantalum and graphite are employed. Since fluorine has a high  $\gamma$ -ray yield, the use of fluorine-rich materials such as polytetrafluoroethylene is avoided. On grounds of convenience and cost, the specimen chamber is usually constructed in aluminum, but in order to limit the production of  $\gamma$ -rays, regions of the chamber that are exposed to scattered protons are often lined with graphite foil.



**Figure 2** Schematic overview of a typical PIXE experimental setup. (Reproduced with permission from Malmqvist *et al.* (1982) *Journal of Radioanalytical Chemistry* 74: 125–147; © Elsevier.)

### Beam Handling

In macro-PIXE, one often analyses specimens with a nonuniform distribution of the analyte elements across the size of the beam. For quantitative analysis of such specimens one needs to have a uniform intensity distribution across the beam. The quasi-Gaussian radial intensity distribution of the beam as it emerges from the accelerator can be converted to a rectangular profile in several ways. A first approach is to focus the beam onto a scattering foil (the gold foil in **Figure 2**), which distributes the protons homogeneously over the cross-section selected by the diaphragm. A second method is to create a very narrow beam by means of quadrupole focusing magnets (see below: micro-PIXE) and sweeping this beam rapidly in both  $x$  and  $y$  directions over the specimen area. This does not give a genuinely homogeneous distribution and may therefore produce high instantaneous count rates in the X-ray detector for heterogeneous samples. On the other hand, this method has the advantage of relatively low loss of beam intensity, whereas such losses may be more than 90% in the case of the first method. The third approach consists of defocusing the beam by means of quadrupole magnets and selecting the central, approximately flat-topped portion of the broadened intensity distribution by means of a diaphragm. The beam-defining diaphragm is usually followed by a second one, which does not intersect with the beam, but serves to remove particles that are scattered by the first one.

### X-Ray Detection

For detection of the X-rays in PIXE, one normally uses a Si(Li) detector, although other energy-dispersive detectors (i.e., HPGe detectors) are also occasionally employed, for example, in setups with two X-ray detectors. The preferred location for the X-ray detector is at  $135^\circ$  to the beam direction, because at this angle the intensity of the bremsstrahlung background is significantly less than at  $90^\circ$  to the beam, but in some cases chamber design restrictions result in a  $90^\circ$  orientation.

In contrast to energy-dispersive XRF spectra, PIXE spectra have most of their X-ray intensity in the low-energy region; hence, the spectral shape can be modulated by placing an X-ray absorber between sample and detector. The aim of absorbers is to reduce or eliminate unwanted continuum background and/or intense X-ray peaks and their associated pile-up peaks and at the same time to allow bombardments at higher beam intensities, so that the elements of interest can be measured in shorter bombardment times and with fewer spectral interferences. The absorbers are usually made of organic material (e.g., Mylar) or of light element metal foils (Be and Al). Plain absorbers are employed if complete elimination of the low-energy part of the spectrum is desired. In many cases, however, it is preferable to allow a certain fraction (e.g., a few per cent) of the low-energy X-rays to pass on onto the detector. This can elegantly be realized by resorting to pinhole absorbers (usually called ‘funny filters’ in the PIXE

community). The spectrum of Figure 1 was acquired using a funny filter; elements down to aluminum can be observed.

When measuring minor and trace elements in a high- $Z$  matrix, one can also resort to a critical absorber, which preferentially attenuates the X-rays of the matrix elements. For example, for the analysis of steel a chromium-critical absorber is used. Since the Fe K X-rays lie just above the K edge of chromium, the former will be selectively absorbed, while photons of higher and lower energies will be transmitted.

When one is interested in measuring the light elements ( $Z < 13$ ), only a very thin X-ray absorber may be used, and for the very light elements, a windowless detector or a detector with a very thin polymer window would be advisable. However, the bombardment of a sample with charged particles also gives rise to backscattered particles, and a fraction of these may penetrate into the Si(Li) detector and cause problems. Hence, it is advisable to use an absorber (typically made of Be) that is sufficiently thick to stop all scattered particles, but this hampers or precludes the detection of the light elements. Placing a magnetic deflection trap between the sample and the detector crystal is a viable alternative to an absorber for removing the scattered ions. However, the installation of such a system also results in a lower solid angle of detection.

### Detection of Other Radiation

The bombardment of a specimen with MeV protons does not lead only to the emission of X-rays. Several other interactions, such as elastic and inelastic scattering of the incident particles and the production of prompt  $\gamma$ -rays, also occur. These other interactions form the basis of various other ion beam analysis (IBA) techniques, such as Rutherford backscattering spectrometry (RBS) and particle-induced  $\gamma$ -ray emission (PIGE). These other IBA techniques can often be used in combination with PIXE in order to obtain complementary information (e.g., by extending the elemental coverage down to hydrogen). Consequently, PIXE chambers are usually equipped with appropriate detectors for the complementary IBA techniques. For RBS one normally uses a silicon surface barrier detector, which is placed at a large backward angle ( $\sim 180^\circ$ ) in order to maximize the elemental resolution. For PIGE, one utilizes a Ge or NaI(Tl) detector. Some light elements, such as F and Na, which are difficult (or impossible) to measure by PIXE, have high cross-sections for  $\gamma$ -ray-producing nuclear reactions, so that the combination of PIGE and PIXE allows a more complete elemental analysis.

## Analysis of Thin Specimens

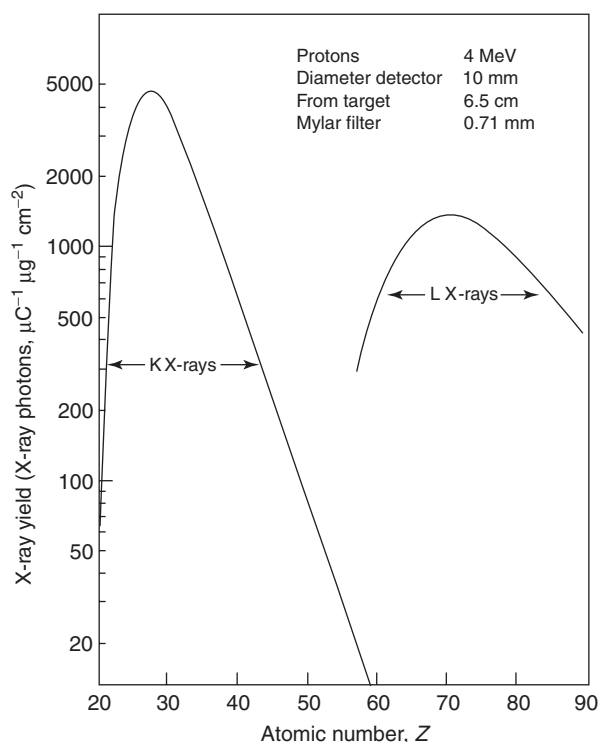
The simplest case of PIXE is that of thin samples. Thin samples do not decrease the energy of the proton beam in an appreciable way during transit and they do not absorb the emitted X-rays on their way to the detector. The 'thinness' criterion is a function of the beam energy and of the sample composition.

### Calibration

When protons of 1–4 MeV are used, elements with  $Z$  up to  $\sim 50$  are generally determined through their K X-rays (typically the  $K\alpha$  line) while the heavier elements are measured through their L X-rays ( $L\alpha$  line). After passage of  $N_p$  protons with incident energy  $E_0$  through a thin specimen, the  $K\alpha$  (or  $L\alpha$ ) line intensity of an analyte element with atomic number  $Z$  and atomic mass  $A_Z$  is given by:

$$Y(Z) = \frac{\sigma_Z(E_0)\omega_Z p_{\alpha,Z}\varepsilon_Z N_A}{A_Z \sin \theta} N_p M_a(Z) \\ = k(Z) N_p M_a(Z) \quad [3]$$

where  $M_a(Z)$ ,  $\sigma_Z(E_0)$ ,  $\omega_Z$ , and  $p_{\alpha,Z}$  apply to the analyte and denote, respectively, the mass per unit area, the ionization cross-section at energy  $E_0$ , the K (or L) fluorescence yield, and the fractional intensity of the  $K\alpha$  (or  $L\alpha$ ) line relative to the total intensity of all the K (or L) lines.  $\varepsilon_Z$  is the absolute detection efficiency (including the solid angle) at the energy of the  $K\alpha$  (or  $L\alpha$ ) line,  $N_A$  is Avogadro's number, and  $\theta$  the angle between the incident proton beam and the specimen surface.  $k(Z)$  is called the thin target sensitivity factor and is expressed in X-ray counts per proton (or unit charge) per  $\mu\text{g cm}^{-2}$ . With regard to the K shell, the compound uncertainty of the fundamental and other parameters that are combined into  $k(Z)$  is of the order of 5–8%. In principle, the use of eqn [3] allows for 'standardless' analysis with an accuracy of better than 10% for elements whose characteristic X-rays are not interfered by others. When L X-rays are used, the potential accuracy is markedly poorer. Instead of relying on fundamental and other parameters (such as the detection efficiency) to obtain the sensitivity factors, most PIXE users prefer to determine the latter experimentally by irradiating uniform standard targets. Generally, these standards are very thin layers ( $M_a = 20\text{--}50 \mu\text{g cm}^{-2}$ ) of elements or compounds evaporated onto Mylar foils or Nuclepore<sup>®</sup> filters. Experimentally determined K and L line sensitivity curves are shown in Figure 3. The accuracy of such a calibration depends almost entirely on the certified areal densities of the standard films; usually they are accurate to within 5%. Once a



**Figure 3** Thin target sensitivity curves obtained with micromatter standards. (Reproduced with permission from Johansson SAE, Campbell JL, and Malmqvist KG (1995) *Particle-Induced X-Ray Emission Analysis. Chemical Analysis Series*, Vol. 133, p. 59. Wiley; © John Wiley & Sons, Inc.)

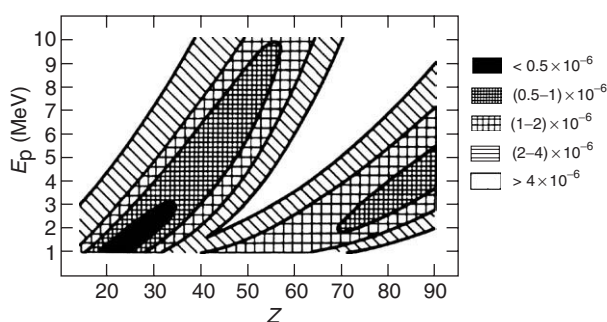
PIXE system has been calibrated with these external standards, the exercise does not need to be repeated for every analysis. Furthermore, in practice, one does not need to measure  $k(Z)$  for each possible analyte element. A complete sensitivity curve can be obtained by interpolation, thereby making use of the fundamental parameters.

### Limits of Detection

As in other spectrometric techniques, the LODs in PIXE are determined by the sensitivity factors and by the spectral background intensity where the analyte signal (X-ray line) is expected. The LOD values in PIXE are proportional to:

$$\frac{I_B^{1/2}}{I_Z} \left( \frac{\text{FWHM}}{Q\Omega} \right)^{1/2} \quad [4]$$

where  $Q$  is the integrated (preset) proton charge,  $I_Z$  and  $I_B$  are the net intensity of a characteristic peak of element  $Z$  and the integrated background below that peak respectively, FWHM its full-width at half-maximum, and  $\Omega$  is the detector solid angle. Of the factors shown in eqn [4], usually the dose  $Q$  and the peak-to-background ratio ( $I_Z/I_B$ ) can be used to improve



**Figure 4** Contour plot of the limit of detection (LOD) as a function of incident proton energy and atomic number of the analyte for a  $0.1 \text{ mg cm}^{-2}$  thick carbon matrix. Experimental conditions: detector FWHM 165 eV, solid angle of detection 38 msr, collected charge  $10 \mu\text{C}$ . The background interval selected for calculating LOD was equal to one FWHM. (Reproduced with permission from Johansson and Johansson (1976) *Nuclear Instruments and Methods* 137: 473–516; © Elsevier.)

the LOD values, since the other factors are geometry and instrument dependent. The preset charge  $Q$  can be increased by increasing measurement time or beam current. With regard to the  $I_Z/I_B$  ratio, this depends on the incident beam energy and on the specimen matrix and possible support (backing) film. It can be improved by using a very thin backing film and by removing the matrix elements from the sample, thereby preconcentrating the analyte elements. Figure 4 shows the variation with proton energy of the LOD for the case of a  $0.1 \text{ mg cm}^{-2}$  carbon matrix. For this matrix, irradiated over  $10 \text{ mm}^2$ , a LOD of  $1 \text{ mg kg}^{-1}$  ( $10^{-6}$ ) corresponds to an absolute analyte mass of 10 pg. It is assumed in Figure 4 that elements with atomic number up to  $\sim 50$  are determined through their  $K\alpha$  X-ray line and the heavier elements through their  $L\alpha$  line. As can be seen in the figure, there is a valley of optimum detection limits for both the K and L cases, with the best K LODs obtained at lower proton energy than the best L LODs. Furthermore, within either the K or L case, the bombarding energy for optimum detection limits depends upon the atomic number of the analyte elements of interest, with higher bombarding energies favoring the heavier elements. Selection of the energy should thus be made with the objective of the analysis in mind, but in practice some compromise is necessary. Finally, it should be noted that the LODs worsen as the atomic number of the matrix increases.

### Preparation of Thin Specimens

The specimen preparation depends upon the type of sample, its composition, the information looked for

(bulk concentrations or spatially resolved data), the elements of interest, and the mode of irradiation (vacuum or nonvacuum). For nuclear microprobe analyses, special requirements apply. For certain samples, the preparation of thin specimens is impossible or even not allowed, e.g., when the sample has to be returned unaltered after the analysis. In the analysis of atmospheric aerosols, a very successful application area of PIXE, the air can simply be drawn through a thin filter that can then be analyzed without further treatment. Biomedical samples are often freeze-dried and powdered. Thin (or semithick) specimens are then prepared by depositing a few milligrams of the powdered material onto a thin backing film and fixing it with a trace-element-free 'glue' (e.g., polystyrene in xylene). The freeze-drying removes the water of the material and can be considered as a preconcentration step, which will lead to improvement in LODs. Greater preconcentration may be obtained by acid digestion (wet ashing) or by dry ashing, so that LODs are even further improved. For dry ashed material, thin specimens may be prepared in the same way as for the powdered freeze-dried material. Acid-digested samples can be pipetted onto a thin backing film and dried. This may also be done with other liquid samples, such as body fluids and environmental surface water samples.

### Precision and Accuracy

The reproducibility (precision) of an entire PIXE analytical procedure (including the contribution from sample processing and specimen preparation) can be examined by preparing several specimens from the same material, subjecting these to PIXE, and calculating a standard deviation from the spread in the results obtained. In the analysis of biological, geological, and atmospheric aerosol samples, a relative standard deviation of 1–2% is often the ultimate practical limit of precision. The accuracy of a PIXE procedure may be evaluated by analyzing (certified) reference materials or through comparison with other analytical techniques. The literature contains many examples of such studies. As an example, for biological materials a total of 18 elements were measured in up to 14 (certified) reference materials; from a comparison of the PIXE results with the reference values it was concluded that the overall accuracy was better than 5%.

### Analysis of Specimens of Intermediate and Infinite Thickness

In many cases, the specimens do not fulfill the 'thinness' criterion, and one has to consider the proton

energy loss within the specimen and the attenuation of the emerging characteristic X-rays. Such specimens are said to be of 'intermediate thickness' or 'infinitely thick' (when they are thicker than the particle range). For both types of specimens, eqn [3] has to be replaced by

$$Y(Z) = \frac{\omega_Z p_{a,Z} \varepsilon_Z N_A}{A_Z} N_P C_Z \int_{E_0}^{E_f} \frac{\sigma_Z(E) T_Z(E)}{S(E)} dE \quad [5]$$

where  $C_Z$  is the concentration of the analyte and  $E_0$  and  $E_f$  are the incident proton energy and the energy of the protons after passage through the target respectively ( $E_f = 0$  for an infinitely thick specimen).  $E$  is the proton energy,  $T_Z(E)$  the transmission of the X-rays of the analyte from successive depths in the specimen, and  $S(E)$  the matrix stopping power.  $T_Z(E)$  is itself given by

$$T_Z(E) = \exp\left(\frac{-(\mu/\rho) \sin \theta}{\sin \phi} \int_{E_0}^{E_f} \frac{dE}{S(E)}\right) \quad [6]$$

with  $(\mu/\rho)$  the mass attenuation coefficient for the analyte X-rays in the sample matrix, and  $\phi$  the angle between the specimen surface and the specimen-detector axis (i.e., the X-ray take-off angle). When specimens of intermediate thickness are analyzed, the transmitted proton energy  $E_f$  (or rather the energy loss  $E_0 - E_f$ ) is needed for evaluating the integral in eqn [5]. This implies knowledge (or determination) of the specimen mass thickness, because the energy loss is related to the latter through the matrix stopping power. Alternatively, the energy loss can be measured experimentally. As for thin specimens, several approaches for quantitative analysis are possible for specimens of intermediate or infinite thickness. One can make use of experimentally determined thin-target sensitivity factors  $k(Z)$ , as defined by eqn [3], so that eqn [5] can be written as:

$$Y(Z) = \frac{k(Z)}{\sigma_Z(E_0)/\sin \theta} N_P C_Z \int_{E_0}^{E_f} \frac{\sigma_Z(E) T_Z(E)}{S(E)} dE \quad [7]$$

In the analysis of infinitely thick specimens, one can also utilize experimental thick-target sensitivity factors instead of relying on the fundamental parameter approach or on experimental thin-target sensitivities. The thick-target sensitivity factors incorporate the integral of eqn [5] and are usually expressed in X-ray counts per  $\mu\text{C}$  and per  $\mu\text{g per g}$ . They are commonly derived from PIXE measurements on standard samples. In a strict sense, the thick-target factors are only valid for the analysis of unknown samples with identical (matrix) composition as the standards, but in practice, some variability in composition can be tolerated or corrected for.



The necessary correction factor is in this case the ratio of the integral of eqn [5] for the standard to the corresponding integral for the unknown.

The LODs for samples of intermediate or infinite thickness are determined by the same factors as in the case of thin specimens. However, whereas thin specimens often consist of a light-element matrix, the matrix of thick specimens is often made up of higher-Z elements, as is the case with geological and metallurgical samples. Consequently, intense X-ray peaks and their associated tails and pile-up peaks show up in the PIXE spectrum, and the heavier matrix also leads to an increased bremsstrahlung background. As a consequence, LOD values are in the  $1\text{--}10\text{ mg kg}^{-1}$  range for thick geological and metallurgical samples, versus  $0.1\text{--}1\text{ mg kg}^{-1}$  for thin (and thick) samples with a light element matrix.

## Nonvacuum PIXE

Proton bombardment in vacuum gives rise to heating of the specimen and this may damage sensitive objects (such as ancient documents) or lead to losses of volatile elements for biomedical samples. These problems can largely be avoided by resorting to atmospheric pressure or moderate vacuum instead of high vacuum during analysis. Because the accelerator requires a high vacuum, the ion beam must be extracted into the moderate vacuum or atmospheric pressure region through a thin exit foil. The beam eventually deteriorates this foil, and therefore its material has to be carefully selected. The best choice is polyimide foil (Kapton), which withstands high intensities and a high radiation dose before mechanical breakdown. The chamber gas is normally helium or nitrogen. The choice of chamber atmosphere is determined by the objectives of the analysis and by the samples to be analyzed. A rather low pressure of helium gas, for example, suffices to improve heat conductivity and to reduce thermal losses of elements or compounds. On the other hand, an external beam in air is needed when large objects are to be analyzed without sectioning or subsampling. This is very important in archaeology and art science, for example. The disadvantages of bombardment in air are that there is a danger of sample oxidation during irradiation and that interfering argon X-ray lines are present in the spectrum. The strong argon lines may also be used to monitor the beam fluence, however, which is otherwise difficult to do in nonvacuum PIXE.

When using external beams, some safety precautions are necessary. Besides the obvious hazard of direct exposure, there is the potential exposure to the

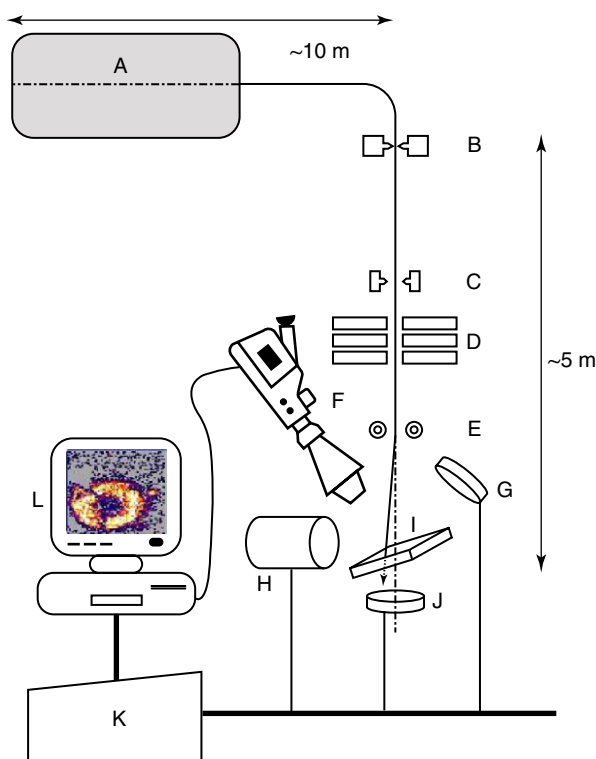
scattered beam and the activation of the air to form positron-emitting radionuclides.

In external beam or nonvacuum PIXE (in air or in a helium or nitrogen atmosphere), poorer detection limits are expected because of the background contribution from interactions in the beam exit foil material and in the air or chamber gas and, for the light elements, also because of the substantial attenuation of their soft X-rays by the same gases. However, practical detection limits in nonvacuum PIXE appear to be comparable to those in vacuum PIXE, at least for analyte elements with atomic number above 25.

## Micro-PIXE

In nuclear microprobe or micro-PIXE analysis, the particle beam is collimated and/or focused down to dimensions in the range of  $1\text{--}50\text{ }\mu\text{m}$ . The best systems currently available are able to produce a spatial resolution of  $\sim 0.5\text{ }\mu\text{m}$  at the specimen while maintaining an ion current that is useful for PIXE analysis ( $>100\text{ pA}$ ). **Figure 5** shows the layout of a nuclear microprobe. Although the exciting particles form the only difference between micro-PIXE and energy-dispersive electron probe microanalysis, equipped with an X-ray detection system, much better peak-to-background ratios and, consequently, lower detection limits are obtainable by micro-PIXE. As shown in **Figure 5**, the normal components of a nuclear microprobe comprise a particle accelerator (normally electrostatic) with a very bright ion source, precision collimators, magnetic or electrostatic quadrupoles for focusing, and a scanning system to raster the beam over the sample, as in a scanning electron microscope. The detection system is, in principle, identical to that for macro-PIXE but usually includes complementary surface barrier detectors for scattered particles and for particles emitted in nuclear reactions to extend the analytical arsenal. Sometimes a detection system for secondary electrons and/or a detection system for near-visible light, iono-luminescence, are also included for imaging of the specimen. The data are acquired in dedicated computer systems that can produce both quantitative results and qualitative elemental maps.

As is the case for macro-PIXE, micro-PIXE may also be carried out under nonvacuum conditions. Because of scattering of the beam by the gas, however, nonvacuum micro-PIXE is only feasible for moderately small beam sizes ( $20\text{--}100\text{ }\mu\text{m}$ ). After collimation and/or focusing, the beam is passed to the nonvacuum region through a pinhole or an exit foil. If the spatial resolution requirements are not too



**Figure 5** Components (not to scale) of a typical nuclear microprobe system: (A) electrostatic particle accelerator; (B) primary object aperture; (C) secondary collimator; (D) focusing system; (E) scanning system; (F) video camera and microscope; (G) surface barrier detector for scattered particles; (H) X-ray detector; (I) specimen; (J) surface barrier detector for transmitted particles (STIM); (K) front-end CAMAC with data bus; (L) main computer and display with elemental map. (Reprinted with permission from Maenhaut W and Malmqvist KG (2001) Particle-induced X-ray emission analysis. In: Van Grieken RE and Markowicz AA (eds.) *Handbook of X-Ray Spectrometry*, 2nd edn. Ch. 12, pp. 719–809. New York: Dekker; © Marcel Dekker Inc.)

high, the nonvacuum micro-PIXE technique is rather straightforward and simple to use. It is very useful when examining large samples or sensitive art objects, such as bronze figures and ancient documents.

Micro-PIXE is related to micro-XRF and especially to synchrotron induced micro-XRF. There are a lot more micro-PIXE installations available than SR micro-XRF.

## Applications

Many of the early applications of PIXE took advantage of its capability to analyze very small specimens (e.g., of mass 0.1–1 mg). The technique found use in many disciplines, such as biology, medicine, environmental and earth sciences, arts and archaeology, and material sciences, achieving detection limits below  $1 \text{ mg kg}^{-1}$  in favorable specimen

types. The samples that can be handled are very diverse. However, considering that the bombardment is normally done in a vacuum, it is evident that the technique is more suitable for analyzing solids than liquids. PIXE analysis of liquids normally involves some preconcentration by drying or some other physical or chemical separation of the analytes from the liquid phase. As far as the analysis of solids is concerned, there are numerous analytical problems for which PIXE is a highly suitable technique. Examples are the multielemental analysis of milligram-sized samples consisting of a light-element matrix (e.g., biomedical and atmospheric aerosol samples), the nondestructive analysis of millimeter-sized areas on a large sample or of thin superficial layers on a bulk sample, and various problems that require sensitive analysis with high spatial resolution.

## Biological and Medical Samples

Most samples of biological origin are composed of essentially organic material, so that its matrix elements are light and do not give rise to characteristic X-ray lines in the spectrum. Furthermore, the electron bremsstrahlung background is lower for an organic matrix than for a matrix of heavy elements. Biomedical samples are therefore well suited for trace element determinations by PIXE. The elements of interest in biological materials are either 'essential' minor or trace elements (e.g., K, Ca, Mn, Fe, Cu, Zn, and Se) or 'toxic' trace elements (e.g., Cd and Pb), and many of these, with some exceptions, such as Cd, can suitably be determined by PIXE in most tissue types. Consequently, ever since the development of PIXE started, biological and medical applications have been prominent areas of application. However, in the last decade, new ultratrace techniques such as inductively coupled plasma-mass spectrometry emerged and the emphasis has shifted from 'total' element determinations toward elemental speciation, for which PIXE is less suitable. In applying PIXE to biomedical problems one should therefore look for cases where one can fully utilize the special advantages of PIXE, such as high spatial resolution, accurate quantitative analysis, and small samples. Particularly micro-PIXE, as used in the nuclear microprobe, is invaluable in biomedical trace element research. It is being applied to single cells and soft and hard tissues of human or animal origin and to samples from the plant sciences field. The relatively high, but still limited, lateral resolution of the nuclear microprobe makes it ideal for studying plant material. The cells are often rather large and it is possible to examine subcellular structures even with the modest beam resolution.

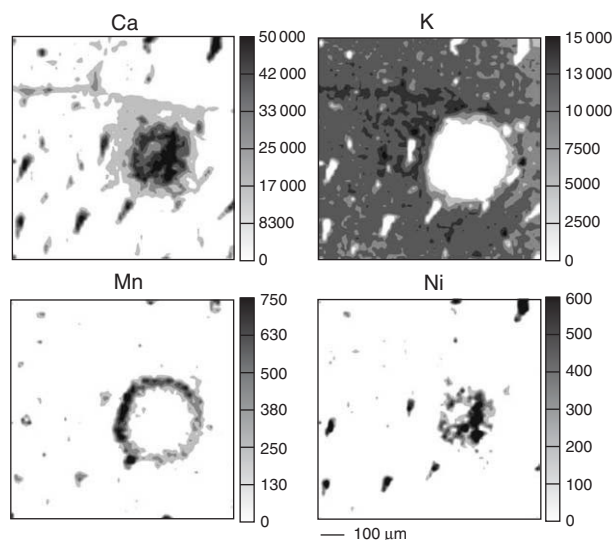
Several reviews (see Further reading section) give examples of micro-PIXE in the plant sciences. One example is a study of plant resistance to pathogen infection. A resistant genotype of a wild plant (*Lagenaria sphaerica* (Sonder) Naudin, Syn. *Lagenaria mascarena* Naudin) belonging to the Cucurbitaceae was inoculated with a foliar fungal pathogen (*Sphaerotheca fuliginea* (Schlecht. ex Fries) Poll. An accumulation of Ni, Cu, Zn, Mn, Fe, Ca, Ti, As, and Sr as well as a drastic depletion of macronutrients (P, S, Cl, K) occurred in the invaded region during the first four days. Six days after inoculation, necrotic cells with high levels of Mn (Figure 6) surrounded the lesions. Si deposition in concentrations of up to 23 wt.% occurred after accumulation of most metals in the fully necrotic lesions. Changes of elemental distribution with time suggest that metal accumulation to phytotoxic levels may contribute to death of infected cells, and that subsequent Si accumulation provides an impermeable, nontoxic barrier protecting the surrounding uninfected tissue.

### Atmospheric Aerosols

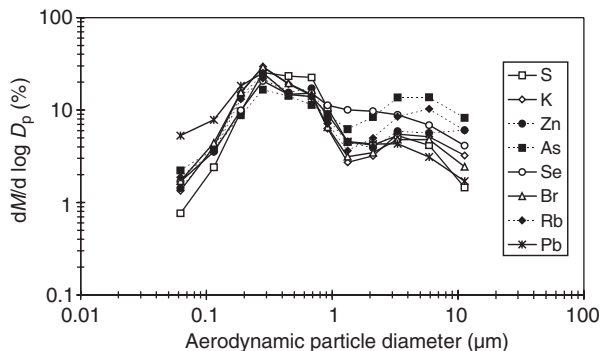
The multielement analysis of airborne particulate material (atmospheric aerosols) has been a very popular and highly successful application of PIXE since the early days of the technique. Considering that atmospheric aerosols are often collected as a thin sample layer on some thin filter or substrate film, that

such samples can be analyzed nondestructively by PIXE without sample preparation, that the sample matrix consists of light elements, and that a 5–10 min bombardment suffices to detect up to 20 elements, including interesting anthropogenic elements, such as S, V, Ni, Cu, Zn, As, and Pb, the analysis of aerosol samples forms almost an ideal application of PIXE. Compared to energy-dispersive XRF, PIXE requires much less sample mass, so that it allows the use of compact samplers with high time and size resolution. Another advantage of PIXE is that it can be complemented with other IBA techniques, so that the light elements (H, C, N, and O) that make up most of the aerosol mass can be measured as well. Motivation for the study of atmospheric aerosols and of their physical and chemical characteristics is that they reduce visibility, lead to radiative forcing on climate, affect human health, contribute to acidification of terrestrial and aquatic ecosystems, and cause damage to structures and buildings.

There is a wealth of examples on the application of PIXE to atmospheric aerosol samples. The aerosol samples analyzed by PIXE range from those collected indoors (e.g., in work environments) or near specific pollution sources to samples dealing with urban, regional, and global air pollution problems, and those collected in areas as remote as the Arctic and Antarctic. As an example, results are presented in Figure 7 from a study that was conducted in the Kruger National Park, South Africa, within the framework of the Southern African Regional Science Initiative (SAFARI 2000). Size-fractionated samples were collected with a 12-stage small deposit area low pressure impactor (SDI), which was developed for subsequent analysis by PIXE, and analyzed for 28 elements. The average mass size distributions for eight selected elements are shown in Figure 7. They



**Figure 6** Quantitative maps of Ca, K, Mn, and Ni in a 6-day lesion caused by *Sphaerotheca fuliginea* on resistant *Lagenaria sphaerica* (Cucurbitaceae) leaf. Scale of intensity in  $\text{mg kg}^{-1}$  dry mass. Scan size  $920 \mu\text{m} \times 920 \mu\text{m}$ . (Reproduced with permission from Mesjasz-Przybyłowicz J and Przybyłowicz WJ (2002) *Nuclear Instruments and Methods in Physics Research B* 189: 470–481; © Elsevier.)



**Figure 7** Average mass size distributions for eight elements, as obtained from PIXE of 64 SDI samples collected during SAFARI 2000. (Reproduced with permission from Maenhaut W *et al.* (2002) *Nuclear Instruments and Methods in Physics Research B* 189: 254–258; © Elsevier.)

have a bimodal or trimodal size distribution, most of their mass (with the exception of As) is in the sub-micrometer size range, and the major fine mode peaks at  $\sim 0.3 \mu\text{m}$  aerodynamic diameter. The fine modes point to high-temperature sources and/or gas-to-particle conversion processes for the eight elements. The likely sources are biomass burning, fossil fuel (coal) combustion, and/or industrial activities. Other elements (i.e., Al, Si, Ca, Ti, Mn, Fe, Sr) had essentially a unimodal coarse size distribution, and over 90% of their mass occurred in super-micrometer-sized particles; these elements result mainly from dispersion of soil dust.

### Earth Science Applications

In the earth sciences, as well as in related economic activities (e.g., mining, prospecting), there is a significant need for elemental analysis. Because of its unique ability to provide quantitative trace element information on a micrometer scale, micro-PIXE offers great potential for mineral prospecting and for improving our understanding of basic geological processes. The nuclear microprobe could serve as an excellent complement to the electron microprobe in these areas.

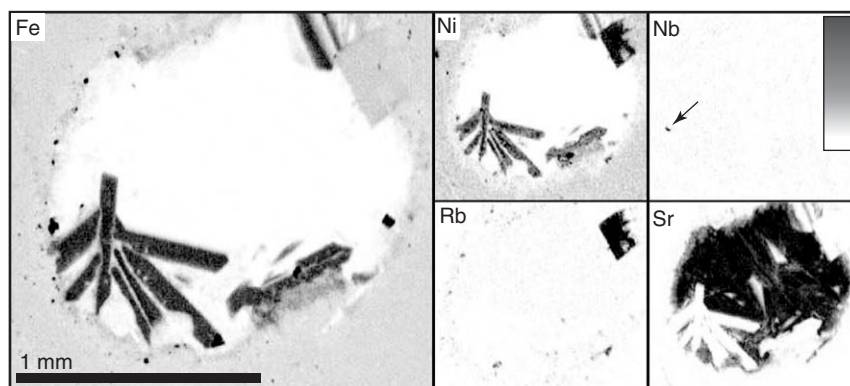
The penetrating, nondestructive nature of MeV proton beams makes the nuclear microprobe ideally suited to *in situ* analysis of individual fluid inclusions in minerals. Fluid inclusions are small samples of fluid trapped in a host mineral during growth or later during healing of fluid filled fractures. They provide samples of the broad spectrum of fluids that have interacted with the earth's crust and upper mantle over geological time. **Figure 8** shows PIXE images of a carbonate-rich melt inclusion in clinopyroxene. The images show spinifex needles of olivine and a large cubic phlogopite (high Rb), set in a matrix of

elongated calcite crystals (high Sr). Nb shows evidence for concentration into rare minor phases (e.g., marked by the arrow in the figure).

### Applications in Art and Archaeology

One application of PIXE and in particular of micro-PIXE that has been unexpectedly successful is the analysis of archaeological and art objects. The aims of such analysis are to provide answers about provenance, authenticity, deterioration, and conservation of the objects. Recognition of the potential of PIXE and other IBA techniques for art investigations has led to the installation of an IBA facility in the scientific laboratory of the Louvre museum in Paris, with a subsequent inclusion of a nuclear microprobe. Similar facilities are now used in several other laboratories specializing in archaeology. Since nondestructiveness is more important here than in other applications, there must often be a trade-off between beam size and intensity in order to avoid damaging the object studied. Larger probe sizes and lower intensities than in other fields of application are normally used. External beam microprobes are particularly useful. Applications of the nuclear microprobe in art and archaeology range from the analysis of antique bronzes and Iron Age tin artifact metals, over gold jewellery, pigments, and paintings, to the inks used in the Gutenberg bible, 'Vinland' map, documents of Galileo, and papyrus from the 'Book of Death' of ancient Egypt.

Besides micro-PIXE, macro-PIXE has also been extensively used in art and archaeology, and found to provide very useful information. Actually, the distinction between micro- and macro-PIXE is difficult to make in this field of application, as a lot of work is carried out with (external) probes with beam



**Figure 8** Dynamic analysis PIXE images of a carbonate-rich melt inclusion in clinopyroxene (in a  $200 \mu\text{m}$  thick polished section). (Reproduced with permission from Ryan *et al.* (2001) *Nuclear Instruments and Methods in Physics Research B* 181: 578–585; © Elsevier.)

sizes ranging from a few tens of a micrometer to  $\sim 1$  mm. Also, one often applies both micro- and macro-PIXE analyses on the same samples. The multielement character of PIXE and the speed of analysis are useful in provenance studies of pottery and earthenware. Obsidians and prehistoric flint tools have been examined by PIXE and other IBA techniques with the aim to trace ancient trade routes or to determine for which purpose the tools were used. Cast iron, ancient iron slag, archaeological bronzes, ancient Greek copper coins, precious metals, and gold artifacts have also been the subject of studies with macro-PIXE and other IBA techniques.

See also: **Activation Analysis:** Charged-Particle Activation. **Archaeometry and Antique Analysis:** Dating of

Artifacts; Art and Conservation. **X-Ray Absorption and Diffraction:** Overview. **X-Ray Fluorescence and Emission:** Energy Dispersive X-Ray Fluorescence; Synchrotron X-Ray Fluorescence.

## Further Reading

Johansson SAE, Campbell JL, and Malmqvist KG (eds.) (1995) *Particle-Induced X-Ray Emission Spectrometry (PIXE)*. Chichester: Wiley.

Maenhaut W, Vandenhoute J, and Duflou H (1987) Applicability of PIXE to the analysis of biological reference materials. *Fresenius Zeitschrift für Analytische Chemie* 326: 736–738.

Tesmer JR, Nastasi M, Barbour JC, Maggiore CJ, and Mayer JW (1995) *Handbook of modern ion beam materials analysis*. Pittsburgh: Materials Research Society.

## Synchrotron X-Ray Fluorescence

**F Adams**, University of Antwerp, Antwerp, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

It has only been since the 1970s that synchrotron radiation (SR) has been employed on a routine basis as a radiation source for X-rays. SR is emitted when charged particles are accelerated by means of a magnetic field perpendicular to their direction of motion. At the heart of a synchrotron light source is an electron (or positron) storage ring, i.e., an evacuated tube through which a densely packed beam of electrons circulates. SR radiation was discovered as a spurious radiation in high-energy particle accelerators (first-generation SR sources), then developed specifically for generating X-rays (second-generation sources). Significant in up-to-date SRs of the present third generation is the systematic use of insertion devices that are placed in the straight sections of the storage ring (wigglers and undulators). Wigglers consist of magnetic structures that create multiple oscillations around the beam path and thus increase both the energy and intensity of the radiation. Undulators are designed to create smaller and more frequent deflections, giving rise to interference effects and in such conditions the coherent radiation is concentrated around specific energies. In addition,

X-ray optics of increased sophistication is used to amplify considerably both flux and brilliance.

A considerable number of storage rings exploit microscopic analysis with microscopic X-ray fluorescence analysis (SR  $\mu$ -XRF) combining the advantages of XRF as an elemental analytical tool with the unique possibilities of the SR. Simultaneously, other analytical uses of intense X-ray beams are being developed. Of special importance are the third-generation storage rings at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, the Advanced Photon Source (APS), Argonne, USA, and SPring-8, Harima, Japan. Compared to earlier second-generation rings and to other recently constructed SR facilities they are characterized by their high energy of 6–8 GeV. Other new SR installations also incorporate insertion devices in smaller, less energetic storage rings.

### Properties of Synchrotron Radiation

Synchrotron X-ray beams have unique properties that make them desirable for use in many disciplines that require intense X-ray beams. First, they have a continuous energy distribution. Via suitable monochromating devices, monoenergetic beams can be produced over a wide energy range. In the horizontal direction, X-rays are emitted in a continuous band, whereas in the vertical plane they are efficiently collimated, making it possible to produce intense

sizes ranging from a few tens of a micrometer to  $\sim 1$  mm. Also, one often applies both micro- and macro-PIXE analyses on the same samples. The multielement character of PIXE and the speed of analysis are useful in provenance studies of pottery and earthenware. Obsidians and prehistoric flint tools have been examined by PIXE and other IBA techniques with the aim to trace ancient trade routes or to determine for which purpose the tools were used. Cast iron, ancient iron slag, archaeological bronzes, ancient Greek copper coins, precious metals, and gold artifacts have also been the subject of studies with macro-PIXE and other IBA techniques.

See also: **Activation Analysis:** Charged-Particle Activation. **Archaeometry and Antique Analysis:** Dating of

Artifacts; Art and Conservation. **X-Ray Absorption and Diffraction:** Overview. **X-Ray Fluorescence and Emission:** Energy Dispersive X-Ray Fluorescence; Synchrotron X-Ray Fluorescence.

## Further Reading

Johansson SAE, Campbell JL, and Malmqvist KG (eds.) (1995) *Particle-Induced X-Ray Emission Spectrometry (PIXE)*. Chichester: Wiley.

Maenhaut W, Vandenhoute J, and Duflou H (1987) Applicability of PIXE to the analysis of biological reference materials. *Fresenius Zeitschrift für Analytische Chemie* 326: 736–738.

Tesmer JR, Nastasi M, Barbour JC, Maggiore CJ, and Mayer JW (1995) *Handbook of modern ion beam materials analysis*. Pittsburgh: Materials Research Society.

## Synchrotron X-Ray Fluorescence

**F Adams**, University of Antwerp, Antwerp, Belgium

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

It has only been since the 1970s that synchrotron radiation (SR) has been employed on a routine basis as a radiation source for X-rays. SR is emitted when charged particles are accelerated by means of a magnetic field perpendicular to their direction of motion. At the heart of a synchrotron light source is an electron (or positron) storage ring, i.e., an evacuated tube through which a densely packed beam of electrons circulates. SR radiation was discovered as a spurious radiation in high-energy particle accelerators (first-generation SR sources), then developed specifically for generating X-rays (second-generation sources). Significant in up-to-date SRs of the present third generation is the systematic use of insertion devices that are placed in the straight sections of the storage ring (wigglers and undulators). Wigglers consist of magnetic structures that create multiple oscillations around the beam path and thus increase both the energy and intensity of the radiation. Undulators are designed to create smaller and more frequent deflections, giving rise to interference effects and in such conditions the coherent radiation is concentrated around specific energies. In addition,

X-ray optics of increased sophistication is used to amplify considerably both flux and brilliance.

A considerable number of storage rings exploit microscopic analysis with microscopic X-ray fluorescence analysis (SR  $\mu$ -XRF) combining the advantages of XRF as an elemental analytical tool with the unique possibilities of the SR. Simultaneously, other analytical uses of intense X-ray beams are being developed. Of special importance are the third-generation storage rings at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, the Advanced Photon Source (APS), Argonne, USA, and SPring-8, Harima, Japan. Compared to earlier second-generation rings and to other recently constructed SR facilities they are characterized by their high energy of 6–8 GeV. Other new SR installations also incorporate insertion devices in smaller, less energetic storage rings.

### Properties of Synchrotron Radiation

Synchrotron X-ray beams have unique properties that make them desirable for use in many disciplines that require intense X-ray beams. First, they have a continuous energy distribution. Via suitable monochromating devices, monoenergetic beams can be produced over a wide energy range. In the horizontal direction, X-rays are emitted in a continuous band, whereas in the vertical plane they are efficiently collimated, making it possible to produce intense



X-ray beams with little angular divergence. As the electrons circulate in equally spaced packets ('bunches') around the ring, a synchrotron is a pulsed X-ray source, with a width of typically 1 ns and, e.g., a 20 ns period. This feature makes synchrotron light very interesting for performing time-resolved studies. In addition, third-generation SR sources provide coherent radiation. The best known feature of synchrotron sources, however, is their high intensity, which may be more than a factor  $10^{12}$  higher than the output of conventional X-ray tubes (see Figure 1).

A particular advantage of SR  $\mu$ -XRF is the extremely high source brilliance (number of photons emitted per unit source area in a unit angle of emission and energy,  $\text{photons s}^{-1} \text{ mrad}^{-2} \text{ mm}^{-2}$  per 0.1% of radiation bandwidth). The high intensity and directionality implies that SR is ideally suited for generating microfocused X-ray beams with very high intensity, now well exceeding  $10^{10}$  photons  $\text{s}^{-1} \mu\text{m}^{-2}$  in a spot size that is presently down to  $\sim 100$  nm.

In addition, in the plane of the storage ring the radiation is linearly polarized with the E-vector parallel and the B-vector normal to the storage ring plane. The radiation is highly collimated along the direction tangential to the movement of the electrons in the ring thus facilitating the delivery of the radiation to a predefined sample area. The polarization of the incident radiation can be used to reduce the

relative contribution of scattered radiation reaching the detector, as scattering cross-sections are dependent on the polarization whereas the photo-absorption cross-sections are not. When performing measurements in the plane of the SR source this increases the signal-to-background ratios by several orders of magnitude (directly dependent on the degree of polarization achieved).

Thanks to the high directionality of the beam, quasi-monochromatic X-ray microbeams can be generated from the white SR spectrum through the use of X-ray monochromators. By tuning the energy over a given energy range, the strong energy dependence of the inner-shell photoelectric cross-sections can be exploited to either increase the specificity of measurements or, else, to obtain speciation information in the XAS application mode (see further).

## Focusing of X-Rays

Limiting the flux throughput through pinholes is insufficient for most practical analytical purposes; hence, techniques for generating intense X-ray microbeams are based on the use of various types of X-ray optics. Refractive lenses that are extensively used in visible or ultraviolet light optics are more difficult to use for X-rays because the refractive index,  $n$ , is very close to and slightly smaller than 1.000 by a factor  $\delta$ . The refractive index can be expressed as

$$n = 1 - \delta \quad \text{with } \delta = \alpha + i\beta \quad [1]$$

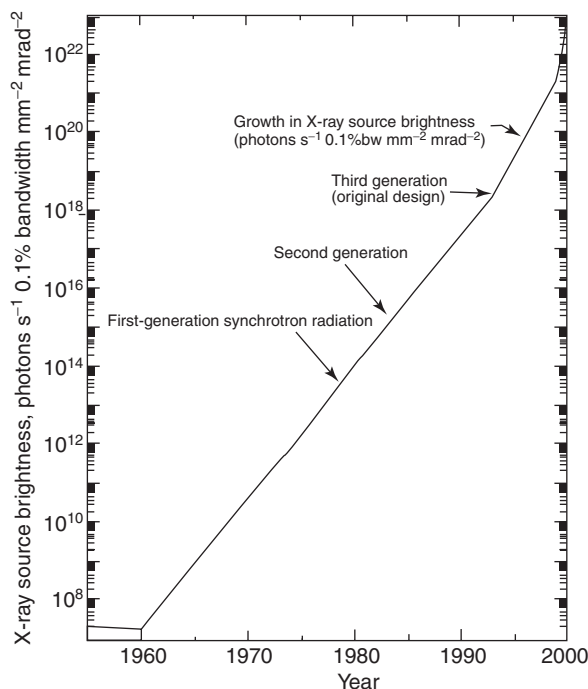
where  $\alpha$  is the absorption index and  $\beta$  the refractive index decrement.  $\delta$  is typically between  $10^{-5}$  and  $10^{-7}$ . In these conditions, the focal distance  $f$  of a single-element refractive lens (a small spherical void in a low  $Z$  material in, e.g., Al) for a parallel beam is very long (of the order of several tens of meters for a  $300 \mu\text{m}$  hole) and given by

$$f = R/2\delta \quad [2]$$

Linear arrays of  $N$  single-element lenses, however, provide reasonable focal distances (of 1–2 m) for parallel SR X-ray beams:

$$f = R/2N\delta \quad [3]$$

Such lenses are cheaply constructed but suffer from the absorption and scattering. Many other techniques for the formation of intense X-ray microbeams are available on the basis of the use of various types of X-ray optics based on refraction, diffraction, or reflection: bent mirrors, crystals and multi-layers, linear and tapered glass mono-capillaries, complex polycapillary lens systems, transmission Fresnel zone



**Figure 1** X-ray source improvements as a function of time, characterized by the source brilliance.

plates, Bragg–Fresnel lenses, one- or two-dimensional (1D, 2D) wave guides. Flux throughput through pinholes for limiting the beam is insufficient for most analytical applications.

In a number of  $\mu$ -XRF installations, capillary optics are used as focusing devices because of their inherent constructional simplicity and the fact that they provide a good lateral resolution and can be used for both polychromatic and monochromatic X-rays. The working distance between capillary tip and sample is short, typically  $<100\mu\text{m}$ . Another popular design is based on Kirkpatrick–Baez focusing mirrors that simultaneously provide a monochromatic and focused beam. Such systems are achromatic (the spot position is independent of the incident energy) and provide a long working distance. They are quite popular for X-ray absorption spectrometry applications (see further), as the beam position remains fixed while scanning over a specific edge.

## Synchrotron X-Ray Microanalysis

SR  $\mu$ -XRF offers a number of advantages compared to other microprobe techniques: it combines high spatial resolution with high sensitivity, can be used in atmospheric conditions, and is relatively insensitive to beam damage to the sample. As we will show further in this section, the simplicity of the method and the quite good understanding of the physics of the processes involved make it more adaptable for quantitative analysis than a number of other beam methods of analysis. The method is now well accepted as an important tool for microscopic elemental analysis, complementing other beam methods of analysis that rely on electron, ion, photon, or laser beams.

The polarization of the incident radiation can be used to reduce the relative contribution of scattered radiation reaching the detector. This increases considerably the signal-to-background ratios and subsequently detection limits. In addition, highly monochromatic X-ray microbeams can be generated through the use of X-ray monochromators. Monochromatic radiation is normally obtained with a double crystal monochromator consisting of a pair of suitable crystals, e.g., in silicon at a particular crystallographic orientation. The energy selection is based on Bragg's law. By tuning the energy of the beam, the strong energy dependence of the inner shell photoelectric cross-sections can then be exploited to increase specificity of measurements, to obtain chemical-state information on elements of interest (speciation) and to make quantification of the detected X-ray intensities easier. Contrary to the

vacuum requirements of most other microprobe methods, samples are normally observed in the ambient environment allowing the measurement of samples in their natural (e.g., in wet) conditions.

The energy resolution of the double crystal monochromator is of the order of  $\Delta E/E = 10^{-4}$  or less, and this is sufficient for absorption edge applications. A direct exploitation mode with broadband polychromatic excitation is also possible. It has the advantage that (nearly) all elements in the sample are excited with quite comparable efficiency providing a more uniform spectrometer response over the range of elements of interest. Since losses in flux due to the monochromator do not occur, the elemental efficiency of polychromatic set-ups is also higher than when monochromatic excitation is used, making it more appropriate for general-purpose materials characterization. In such circumstances, however, quantification of the detected X-ray intensities is more complicated. Also, the signal-to-noise ratios, and hence, detection limits, do not match those obtained with monochromatic excitation. For quantitative analysis the use of monochromatic radiation is, in general, a preferable approach.

In its primary utilization mode as a tool for elemental analysis by XRF analysis a high photon flux rather than such a high-energy resolution is required and an energy resolution of the order of  $\Delta E/E = 10^{-2}$  is sufficient for the purpose. Synthetic multilayers, made by vacuum deposition of alternate thin layers of two materials with different electron densities provide this resolution. On the other hand, through a wide energy band-pass, they yield a photon flux one to two orders of magnitude higher than available with a high-resolution double crystal monochromator.

Excluding the inherent complexity of the SR X-ray source problems involved with monochromatization and focusing of the X-rays, instrumentation for X-ray microanalysis is conceptually simple and consists of a mechanical sample stage with high-precision computer controlled microstepping motors for X, Y, Z sample movement, a semiconductor-type detector for measurement of the fluorescent radiation, visualization tools for observation of the sample, and a range of diagnostic and control tools. One typical configuration is shown in **Figure 2**.

In typical setups available relative detection limits (DL) are  $<0.1\mu\text{g per g}$  for elements of  $Z > 25$  but DLs down to a few ppb are possible for quite a number of elements. Absolute DLs are  $<1\text{ fg}$  for elements on the basis of a flux in the focused beam of  $10^9$ – $10^{10}$  photons  $\text{s}^{-1}$ . For the measurement of elemental distribution maps, spectra are taken as the sample is moved over the beam path. Two-dimensional mapping of the

repartition of elements in larger objects than the X-ray beam is possible with relative detection limits in the ppb region and absolute detection limits for many elements well below the femtogram level.

## Synchrotron Computerized X-Ray Fluorescence Microtomography

As a result of the large penetration and the considerable sampling depth of the X-ray beam, the conventional XY-type of scanning a sample through the microbeam yields 2D images that are actually projections of the 3D compositional structure. This particular drawback can be turned into an advantage by the use of computed X-ray fluorescence microtomography (XFCT), an emerging technique capable of providing 3D, potentially quantitative information on the elemental distributions in the probed sample volume with trace-level detection limits. In XFCT, the sample is irradiated under various observational angles, thus creating different 2D projections of the same 3D structure. The sample is mounted on a goniometer head fixed to a rotation stage. Instead of an XY scan the sample height is kept fixed and an  $X\theta$  scan is performed varying  $\theta$  systematically by small increments (e.g., in steps of  $3^\circ$ ) to obtain the so-called 'sinograms' of the different elements. Through the use of appropriate reconstruction algorithms these can be reconstructed into XZ maps. The procedure can then be related to obtain 3D images of medium-to-heavy elements inside the samples (see Figure 3).

The entire reconstruction in 3D of an object is quite time consuming at present. Techniques are being developed, however, to speed up the measurement process, combining fast detector systems, high-speed data networks, and parallel computing systems.

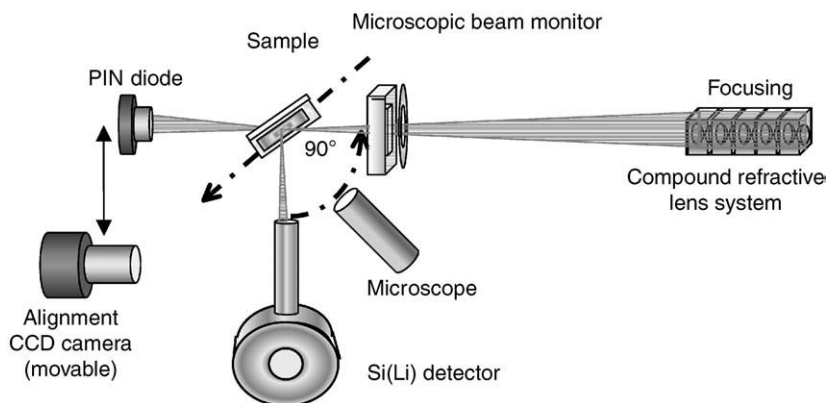
## X-Ray Absorption Methods

In the applications of X-ray absorption spectrometry (XAS, also called X-ray absorption fine structure spectroscopy, XAFS) the energy dependence of the inner shell photoelectric cross is exploited to increase specificity of measurements or to obtain information on the chemical environment (speciation analysis).

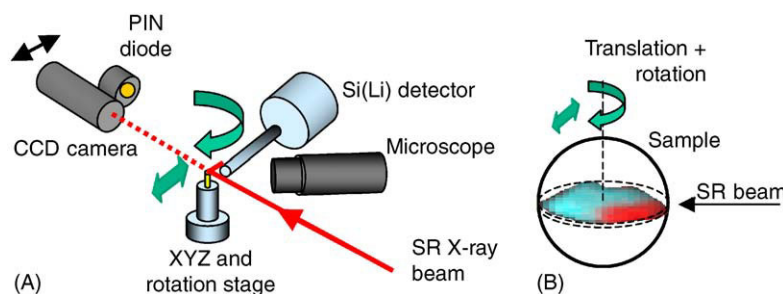
Extended X-ray absorption fine structure analysis provides information on the number, the atomic number, and the distance of neighboring atoms. The method is based on irradiation with a highly monochromatic X-ray beam of tunable energy at ca.  $\Delta E/E = 10^{-4}$  while scanning over a specific absorption edge of an element of interest while recording either the absorption of the beam (absorption XAS), the fluorescence radiation produced (fluorescence XAS), or another shell dependent phenomenon. X-ray absorption near edge structure spectrometry (XANES) measures the position of the edge and characterization can be achieved by exploiting specific features of the X-ray absorption spectrum (coordination number, distances to nearest neighbor atoms). Recently, the combination of  $\mu$ -XRF with spatially resolved XAS became an important tool for speciation in environmental and geological materials and for the study of processes in chemical species transformation. Most  $\mu$ -XAS applications are performed in the XANES fluorescence detection mode serving as a 'fingerprinting' technique on the basis of reference compounds.

## X-Ray Diffraction

X-ray diffraction (XRD) provides the structure of materials. The use of monochromatic excitation combining  $\mu$ -XRF with microscopic XRD ( $\mu$ -XRD) complements localized compositional information with structural information. Mostly, Laue spectra



**Figure 2** Schematic layout of the insertion device 18F instrument at the ESRF; from left to right beam focusing, monitoring equipment for the beam intensity, sample and second beam monitor, different visualization tools.



**Figure 3** (A) Schematic drawing of an X-ray fluorescence microtomography set-up that (B) allows the collection of elemental distributions from a single horizontal slice of the sample. Further slices are obtained by translation.

are taken using a CCD X-ray detector.  $\mu$ -XRD has the distinct information that structural information on a material can be obtained from crystals down to a few (tens) of micrometers in diameter allowing mapping of the local structure/stress of a material.

## Imaging

Several hard X-ray imaging techniques greatly benefit from the coherence of the beams delivered by the modern SR sources. When a coherent beam of X-rays travels through an object both the amplitude and the phase of the wave are modified. Phase imaging is directly related to the small angular size of the source as seen from one point of the sample. Phase radiography and tomography are instrumentally very simple. They are often used in the 'edge detection' regime, where the jumps of density are clearly observed. Recently, a more quantitative approach has been developed, which provides a 3D density mapping of the sample ('holo-tomography'). The combination of diffraction topography and phase-contrast imaging constitutes a powerful tool that can help in monitoring the areas of a sample to be selected for further chemical or structural analysis.

## Applications

It should be clear that with the limited availability of SR sources at the present, the techniques described cannot be applied on a routine basis. One area of applications in analytical chemistry is the use of SR  $\mu$ -XRF as a validation tool for other beam methods for microscopic analysis and assist in providing

reference samples. Microscopic X-ray diffraction ( $\mu$ -XRD) and  $\mu$ -XAS can be applied simultaneously with  $\mu$ -XRF and XFCT to obtain simultaneously as well nondestructive materials characterization providing chemical information and data on crystal-line structure of the sample under investigation as the trace elemental composition and distribution, all this on the microscopic level. The use of the methods described is reserved for solving important problems in environmental sciences, geo-, and cosmochemistry in applications where other microscopic methods fail to provide the necessary sensitivity and specificity.

See also: **X-Ray Absorption and Diffraction: Overview.**

## Further Reading

- Adams F, Janssens K, and Snigirev A (1998) Microscopic X-ray fluorescence analysis and related methods with laboratory and synchrotron radiation sources. *Journal of Analytical Atomic Spectrometry* 13(5): 319–331.
- Bertsch PM and Hunter DB (2001) Applications of synchrotron-based X-ray microprobes. *Chemical Review* 101: 1809–1842.
- Catlow CRA and Greaves GN (eds.) (1989) *Applications of Synchrotron Radiation*. Glasgow: Blackie.
- Haberkorn R and Beck HP (2000) Basics, possibilities, and limitations of the microscopic X-ray fluorescence analysis. *Mikrochimica Acta* 133: 51–58.
- Janssens KH, Adams FC, and Rindby A (eds.) (2000) *Microscopic X-ray Fluorescence Analysis*. Chichester: Wiley.
- Potts PJ, Ellis AT, Holmes M, *et al.* (2000) X-ray fluorescence spectrometry. *Journal of Analytical Atomic Spectrometry* 15: 1417–1442.

# X-RAY PHOTOELECTRON SPECTROSCOPY

See **SURFACE ANALYSIS: X-Ray Photoelectron Spectroscopy**

# ZONE REFINING

**H B Singh**, University of Delhi, New Delhi, India

© 2005, Elsevier Ltd. All Rights Reserved.

This article is reproduced from the first edition, pp. 5668–5673, © 1995, Elsevier Ltd., with an updated Further Reading list supplied by the Editor.

## Introduction

Zone refining is a method for the ultimate purification of substances – element, inorganic compound, or organic compound. It denotes a family of methods for controlling the distribution of soluble impurities or solutes in crystalline materials. A short molten zone travels slowly through a relatively long solid charge and while traveling redistributes the solute in the charge. This article gives a brief account of the principles and application of zone refining.

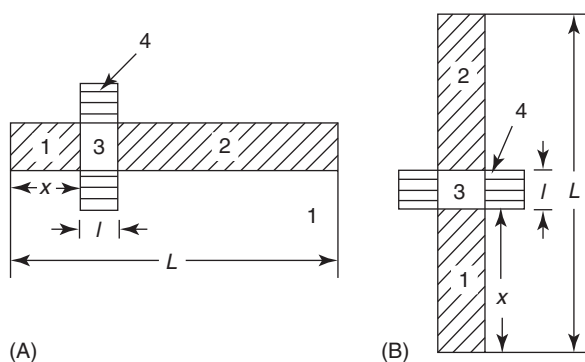
Impurities, even at trace levels, may exert a profound influence upon certain properties of metals and compounds. For instance, the presence of even a few foreign atoms in a metal may affect its deformation and fracture properties, electrical conductivity, or corrosion resistance. This necessitates a study of matter in a high state of purity, which may be achieved following a large number of fractionations or chemical separations using high-purity reagents and guarding against contamination from vessels, laboratory, and even personnel. Handling samples at various stages introduces impurities and this limits the production of high-purity materials. In zone refining, this problem is solved by melting only a small part of the material and slowly passing the molten zone down the solid sample to be purified (**Figures 1A and 1B**). Subsequent passes of the molten zone increase the purity without requiring any handling.

The technique of purification by zone melting is commonly called zone refining. Any substance that can be melted and solidified and exhibits a difference in impurity concentration in the liquid and solid states at the point of solidification can be purified by this method. The impurities redistribute, being displaced toward one end with corresponding purification of the portion being traversed by the molten zone. Though zone melting is largely restricted to

systems exhibiting liquid–solid transformation, successful refining has also been carried out where vapor–solid and even solid–solid transformations take place. In cases where any of these equilibria are not favorable for an impurity, zone refining may be combined with other techniques to achieve the desired purity level.

It is essential to distinguish between zone refining and other common purification techniques. The process in which the whole of the charge is melted and gradually solidified unidirectionally is referred to as normal freezing. This may result in a mass of small crystals that may have entrapped liquid. Therefore, slow growth rates are required to secure a continuous and smooth solid–liquid interface during freezing from solvent or melt. Although many difficult separations, such as isolation of radium and separation of lanthanoids, have been achieved by repeated, fractional crystallization, zone refining is superior as it decreases the effort of making such repetitions.

The main difference between the two techniques is that in zone refining only a part of the charge is melted at any time. This change alone tremendously increases the efficiency of crystallization as a separation and purification method. An impurity travels with or opposite to the movement of the zone depending, respectively, on whether it lowers or raises the melting point of the substance to be purified. In this way the impurities are concentrated in one or



**Figure 1** Schematic diagram of a horizontal (A) and vertical (B) configuration of the zone melting process. 1, Resolidified zone; 2, unmelted zone; 3, molten zone; 4, heater.

the other end of the charge, thereby purifying the remainder. The degree of purification increases with each increase in the number of passes defined as the number of times a single zone passes through the charge. Pfann, working on the purification of germanium for making transistors, was able to produce this element having less than 1 part of impurity in  $10^{10}$  parts of germanium.

Besides purification, zone melting is useful in distributing any impurity uniformly throughout a system and in producing single crystals. Thus, precisely controlled discontinuities in impurity concentration, such as p-n or n-p-n junctions in semiconductors can be made in order to produce basic building blocks for transistors and solid diodes. Since zone refining can remove as well as concentrate the desired impurity, it can be used as a powerful analytical tool for detecting and determining impurities unobservable by conventional means.

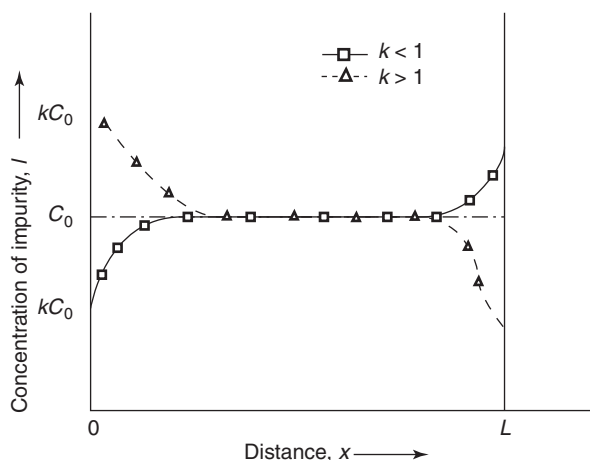
## Theory of Zone Refining

Consider a sample consisting of bulk component B contaminated with impurity I, the concentration of which is  $C_0$ . I is analogous to the solute and B to the solvent. The sample, also called 'the charge', is taken in the form of a bar or rod of length  $L$  ( $\leq 0.5$  m). The heating zone, in the form of a narrow ring of width  $l$  capable of being heated to temperatures above the melting point of the sample, is positioned around the sample rod (Figures 1A and 1B). Generally, the heating zone is mobile and the sample is kept stationary, but the refining process can also be carried out by holding the heating zone stationary and moving the sample. Means of heating include electrical resistance coils, electric arc, electron beam, induction coil, lasers, plasma, radiant energy, and solar energy. Movement of the heating zone ( $4 \text{ cm h}^{-1}$ ) along the sample produces a molten zone at the front and a solid zone at the rear. If the solute lowers the melting point of the solvent, its concentration in the freezing solid is lower than in the liquid. If the solute raises the melting point, the resolidified mass is more impure than the liquid phase. Thus, the freezing interface rejects certain solutes and attracts others.

The measure of concentration of the impurity in the liquid phase is given by the distribution coefficient  $k$ , defined as the ratio of concentration of impurity I in the just-forming solid phase to that in the liquid phase, i.e.,

$$k = C_S/C_L$$

When the solid and liquid phases are in equilibrium, then  $k = k_0$ , where  $k_0$  is the equilibrium distribution



**Figure 2** A schematic impurity distribution for single-pass zone refining.

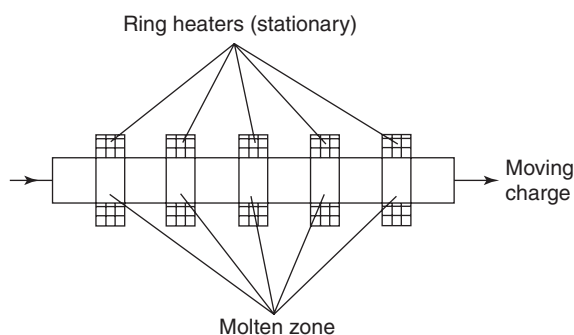
coefficient. When  $k_0 < 1$  addition of solute lowers the melting point; conversely, if the melting point is raised by the addition of solute then  $k_0 > 1$ . Assuming  $k < 1$ , the concentration  $C$  of the impurity I in the resolidified mass at a distance  $x$  from the starting end is given by

$$C = C_0[1 - (k)e^{-kx/l}]$$

where  $C_0$  is the concentration at  $x=0$  (Figure 2). In such a case, moving the heating zone from left to right (Figure 1A) concentrates impurities on the right-hand side. Consequently, a second pass of the heater will result in further purification of the sample by depleting impurities from the left-hand side and accumulating these on the extreme right of the sample rod. This enrichment continues at a decreasing rate until its concentration attains a value  $C_0/k$ . From this point on, the concentrations entering and leaving the zone are equal and the concentration of freezing solid remains  $C_0$ . In the last portion of the charge equal to the zone length  $l$ , the impurity concentration is more than in the original sample (Figure 2).

In the same way as higher purity is achieved by repeated crystallizations or separation and recombination of fractions in fractional crystallization, purity can be increased in zone refining by accumulating impurities in one or other end by multipass zone refining. This can be achieved either by repeated zoning by a single heater or better, by using a multiple heater arrangement (Figure 3). This eliminates all kinds of handling procedures. Another advantage of multipass zoning is the small size of the molten region, which ensures minimal contamination by the container or atmosphere. After several passes, the impurity concentration approaches a steady state or ultimate distribution.





**Figure 3** Multipass zone refining.

## Factors Affecting Zone Refining

### Width of Heating Zone

The width of the heating zone, defined as the length of the solid melted to form the zone (and not the length of the molten zone), is one of the most important operational parameters. It is desirable to express the length of the sample rod  $l$  as a multiple  $n$  of the zone width, i.e.,

$$n = L/l$$

Thus, the greater the value of  $n$ , the greater the number of zone passes required to approach the ultimate distribution. After  $n$  passes, each successive additional pass produces less additional purification. In practice,  $n < 5$  is not very useful. Further, the interzone spacing  $i$  is generally of the order of the zone width  $l$ , though for high thermal conductivity materials like copper, it may be several zone widths.

### Zone Traverse Velocity

Diffusion coefficients of the solid and impurities, and impurity rejection and its transport away from the solidifying face, are governing factors that determine zone travel speed. The rate of solidification consequent upon the speed of zone travel must be much more than the rate of solid diffusion but not so fast as to prevent efficient diffusion of the impurity into the molten zone. Mechanical aids like stirring, to facilitate liquid diffusion, increases the efficiency of the process and permits an increase in zone speed by a factor of 10 or more. However, at high freezing rates, dendrites are formed which entrap liquid, and at somewhat lower rates, prismatic substructures are formed. In fact, for attaining the highest purity in the shortest time, for a given value of  $k$  it is necessary to set the ratio  $n/f$  to its smallest possible value,  $f$  being the heating zone traverse speed. Also, the ratio  $G/R$  (temperature gradient/growth rate) must be less than

the critical value; the former is related to zone width and the latter to zone speed.

### Matter Transport in Zone Melting

The process of zone refining is primarily based on the transport of impurities in one or other direction. The driving force in matter transport is provided by the difference in densities between the solid and liquid phases. If the density of the liquid formed on melting the solid is higher, then the matter is transported in the direction of zone travel. Conversely, if expansion occurs on melting, the matter is transported in the opposite direction to the movement of the heating zone. In situations like this, matter transport is avoided by tilting the sample ingot. The critical angle of tilt is such that the liquid surface at the trailing end is at the same height as that of the original solid before it was molten.

## Factors Influencing the Design of Zone Refining Equipment

These factors include:

1. The nature of the process, whether batch, semi-continuous, or continuous.
2. The yield of purified material to be obtained in relation to time and expense.
3. The chemical and physical properties of the material to be purified and to a lesser extent the properties of the impurities.
4. The properties of the container material with reference to the melting point, thermal conductivity, porosity, and thermal expansion of the material to be purified. Conventional container materials used in zone refining include plastics (e.g., polytetrafluoroethylene), metals, glass, silica, mullite ( $\text{Al}_2\text{Si}_2\text{O}_{13}$ ), alumina, and graphite. In some cases compounds like zirconia ( $\text{ZrO}_2$ ), beryllia ( $\text{BeO}$ ), and various fluorides, silicides, and nitrides are used. Silicon nitride is of special interest in view of its low coefficient of expansion, high temperature strength, resistance to thermal shocks, inertness, and non-wettability by metals. If the properties of the container material are incompatible with those of the sample, refining may be carried out by the unsupported floating zone method (FZM) (see below).
5. The atmosphere in the vessel surrounding the charge.
6. The traversing mechanism for transport of the molten zone(s). The choice lies between moving the charge and moving the heat source. The former has practical limitations where controlled atmosphere and temperature are involved. It is possible to move both the bar and the atmosphere

containment vessel simultaneously. The stirring of the molten zone by mechanical means during zoning, in order to aid diffusion, is more easily achieved if the heating zone is static and the charge is made to move.

## Modifications of Zone Refining

### Continuous Zone Refining

Zone refining, as outlined above, is a multistage batch technique in which successive operations are performed on a single batch of sample. The main limitation of such a procedure is the decrease in efficiency of zone passes as their number increases. Thus, ultimate purity is obtainable only at a prohibitive cost of time and money.

To circumvent the various limitations of the basic technique, several continuous zone refining methods have been developed in which feed enters at one point in the sample while the product and waste leave at other points (Figure 4). The effect of countercurrent movement of solid and liquid phases is achieved by the movement of molten zones. In addition to the horizontal continuous refiner, the zone void vertical refiner and zone transport refiner are other modifications of this class. Cross-flow zone refiners and rotating drum multistage crystallizers based on the above principle are mainly used in growing single crystals rather than in purification of materials.

### Zone Leveling

It is virtually impossible to obtain a uniform concentration of a soluble impurity from one end to the other of a crystal grown from the melt by a conventional zone melting method because the separating crystals differ in composition from that of the melt. Zone leveling denotes a group of zone melting methods that aim to produce a uniform, or level,

distribution of impurity in the charge. In this method, the initial concentration in the first zone length is arranged to be  $C_0/k$  and in the remainder,  $C_0$ . Once this condition is met, the impurity concentrations entering and leaving the heat zone remain constant until the zone reaches the end of the ingot and normal freezing commences. The method is highly valuable in producing homogeneous single crystals having a uniform concentration of impurity, for example, crystals of a germanium-rich germanium-silicon alloy. This alloy is highly prone to segregation;  $k$  for silicon in germanium is  $\sim 3$ .

Another freezing technique has been developed in which the distribution of solute (i.e., impurity) is perturbed during freezing. This technique, called the method of perturbing the concentration, has led to the production of solid electronic devices having p-n and n-p-n junctions.

### Floating Zone Method

The FZM is a melt growth technique that does not require the use of a crucible. Thus, sample contamination by the container is not a problem. A vertical sample is heated with a focused laser light source or another heat source that can produce a molten zone capable of being held in place by surface tension. The molten zone can be seeded to produce oriented or single crystals. A miniaturized version of FZM called a laser-heated pedestal growth method uses a  $\text{CO}_2$  laser operating at  $10.6\text{ }\mu\text{m}$ ; in certain cases several lasers have also been used. The source rod is produced by cold pressing or sintering. To start, a molten region is formed at the top of the source rod. The seed rod, which may be of an inert or refractory material like platinum or indium, or polycrystalline or a single crystal of interest, is introduced into the melt and slowly withdrawn to initiate the crystal growth. The method was first applied to grow crystals of superconducting material of composition  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ . Another version of the FZM applied to incongruently melting materials is called the traveling solvent float zone (TSFZ) method. In this method the composition of the feed rod is that of the desired final product. Single superconducting crystals of  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$  are reported to have been prepared by TSFZ.

### Liquid Encapsulation

Materials with volatile components are taken in an enclosed system so as to maintain the vapor pressure of the volatile constituents equal to or more than the equilibrium dissociation pressure over the melt. The loss of the volatile constituent is prevented simply by maintaining an inert gas pressure greater than

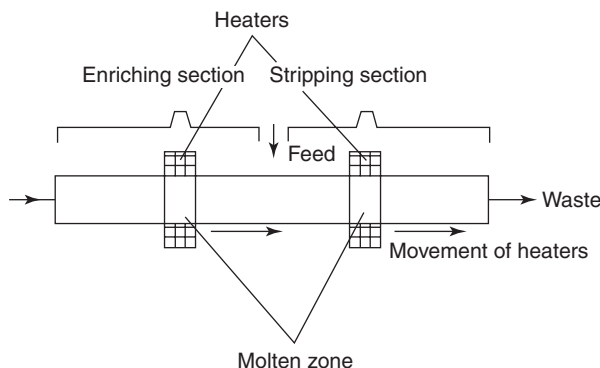


Figure 4 Continuous zone refining process.

the equilibrium dissociation vapor pressure of the volatile component. It is desirable that a liquid encapsulant should be less dense than the melt, be optically transparent, have a low viscosity at the melting point of the sample, and be stable and non-reactive with melt, crucible, and environment.  $B_2O_3$  is the encapsulant of choice above  $1000^\circ\text{C}$ . Below  $800^\circ\text{C}$ ,  $BaCl_2$ ,  $CaCl_2$ , and  $CaCl_2 + KCl$  are used and at still lower temperatures  $LiI + KI$  is used as an encapsulant for zone refining of  $CdSnAs_3$ . Certain organic acids are used for growing bismuth crystals at  $\sim 150^\circ\text{C}$ . On the other hand, high-pressure float zone melting is employed for achieving high purity of dissociating samples like GaP.

Where sample availability is in the range of a few micrograms to a few milligrams, microscale zone refining is very advantageous. Besides processing very small quantities, the operational time is drastically decreased because a large number of heating zones is used and the zone speed can be increased because of the more diffusive mixing achieved.

### Solution Zone Refining

Solution zone refining, also called traveling heater zone refining (THZR), is identical to zone refining. It consists of a liquid zone ( $\sim 1\text{ cm}$  thick) of a concentrated solution containing the phase to be refined as solute; the solvent should be such as to dissolve the solute and the impurities but to be insoluble in the solid solute phase. This process is particularly promising for purification of materials that decompose on melting, have very high melting points, have a high vapor pressure at the melting point, and are highly reactive. By using THZR, purification can be achieved at a much lower temperature than is normally possible.

### Other Methods

Thin alloy zone crystallization (TAZC) and temperature gradient zone melting (TGZM), in which the driving force is, respectively, an electrochemical potential and a temperature gradient, have also been developed based on the principle of zone refining. An alloy of composition 80% indium and 20% antimony (m.p.  $425^\circ\text{C}$ ) for refining thin layers of InSb has been prepared by TAZC whereas TGZM has been used in fabricating semiconductor junctions, and joining and growing single crystals.

## Applications of Zone Refining

Zone refining, an extremely simple technique in principle, provides an excellent means of preparing ultrapure crystalline materials. Handling of the

material and its consequent contamination is kept to an absolute minimum. Further, the use of a controlled atmosphere or vacuum in the system removes the danger of atmospheric contamination or reaction with components of the air. Besides purification, the importance of zone melting methods lies in the uniform distribution of the desired impurity throughout a single crystal, a problem which remained unsolved for a long time. Thus, precisely controlled discontinuities in impurity concentration can be made to produce p-n or n-p-n junctions in semiconductors. Further, joining single crystals and measuring diffusivities in liquids can also be achieved. Some examples of the applications of zone refining are given below.

### Refining of Metals

A large number of metals have been obtained in a state of ultrapurity by zone refining. Quantification of impurities is done by measuring the ratio of resistivities at two different concentrations of the impurity. Uranium is purified by zone refining in a uranium oxide boat and an atmosphere of purified argon; a travel rate of  $6\text{ mm h}^{-1}$  was found to be the most satisfactory. By starting from a relatively pure material, a 90% decrease in iron and carbon content is achieved; oxygen and nitrogen are also removed by vaporization. A final purity  $>99.99\%$  is obtained.

### Refining of Semiconductors

**Elemental semiconductors** Germanium has been widely used for semiconductor diodes and transistors. Six zone passes through an ingot of the purest available germanium sufficed to decrease the concentration of impurities to 1 atom in  $10^{10}$  atoms of germanium.

Ultrapure silicon is not only important for transistors, it is also required for solar batteries. Zone refining of silicon using a quartz boat is unable to remove the last traces of boron except when a chemical purification is simultaneously applied. The reactivity of molten silicon creates problems of contamination with oxygen. For higher purity, the floating zone technique has been used to quantify the concentration of boron, an electron acceptor, and phosphorus, an electron donor. The electrical conductivity measured as a function of distance along the rod is found to be proportional to the difference in concentration of boron and phosphorus; their concentration after refining is of the order of  $10^3\text{ atoms cm}^{-3}$ .

Copper impurity in selenium is related to a trapping center in CdSe. After more than eight passes, copper accumulates at one end. Its concentration at

the other end was below the detectable limit;  $k$  for copper is  $<1$ .

**Compound semiconductors** A (III–V) compound, InSb, is of particular interest in view of its highest electron mobility. In its preparation, zone refining of the compound, of the components, or both, may be utilized. During zone refining, the optical energy gap of InSb decreases from 0.39 eV for relatively impure material to 0.18 eV for an extensively refined sample. The electrical conductivity and Hall coefficient of pure and tellurium-doped InSb were measured as functions of temperature from 90 to 470 K. From the analytical standpoint, it was possible to identify the critical impurities in InSb. Both tellurium and zinc are present. The former, a donor, lowers the melting point and is slower to segregate but is removed by zone refining, while zinc, an acceptor, raises the melting point and is removed only on electroplating.

Thallium bromide semiconductor for X-ray and  $\gamma$ -ray detectors is purified by multipass zone refining. A vacuum-deposited thin film of TlBr is used as a single photon detector and in X-ray imaging including xeroradiography.

#### Preparation of Films and Single Crystals of Superconductors

Several complex compounds have been investigated for their superconducting properties. These usually comprise oxides of a lanthanoid, an alkaline earth metal, and a transition metal. Superconducting properties of such complex oxide systems are very sensitive to the presence of impurities. Films of the Y–Ba–Cu–O system with a critical temperature for superconducting transition of 80 K have been obtained by zone crystallization. Single crystals of  $\text{Bi}_2\text{Sr}_2\text{CaCuO}_8$  are obtained by TSFZ. Measurement of the superconducting properties of various parts of the crystal shows that the later grown part exhibits semiconducting behavior. After 1 day of annealing, homogeneous crystals are obtained. Large single crystals of Bi-2212 superconducting along the  $c$ -axis have been prepared by FZM.

#### Refining of Organic and Inorganic Compounds

Many organic and inorganic compounds have been zone refined. Since the physical properties of compounds differ significantly from metals, these require different apparatus. The low thermal diffusivity requires artificial cooling between zones. Low chemical diffusivity makes agitation of the liquid desirable; agitation also ensures better heat transfer to the liquid.

Ultrapure crystals of nitrobenzene for nonlinear optical devices are prepared by zone refining of the fractionally distilled, purified sample. Single crystals of  $m$ -nitroaniline for use as waveguides in optical second-harmonic generation are produced by zone melting crystallization using a ridge heater. Sodium nitrate has been purified by infrared laser-assisted zone refining. Besides a large number of hydrocarbons, fatty acids, alcohols, nitrogen- and sulfur-containing organic compounds, halides, and nitrates have also been purified by various zone refining methods.

#### Analytical Applications

Zone refining is a powerful tool for separation and concentration. Following preconcentration, conventional analytical techniques are used to identify and quantify the impurity. This has enabled certain observed abnormalities in properties to be correlated with the presence of impurities in ultratrace amounts that had earlier remained undetected. In another version, the distribution coefficient,  $k$ , of a metallic solute is determined by adding its radioisotope under the conditions of the zoning experiment. If the concentration of the nonradioactive form of the solute in the starting material is measurable by an available analytical technique but subsequently falls below the detection limit after refining, its final concentration can still be calculated from the value of  $k$  and the zone refining curves.

See also: **Radiochemical Methods:** Radiotracers.

#### Further Reading

- Brice JC (1973) *The Growth of Crystals from Liquids*. Amsterdam: North-Holland.
- Gatos HC (1979) *Methods for Crystal Growth*. Amsterdam: North-Holland.
- Grubel O (1961) *Metallurgy of Elemental and Compound Semiconductors*. New York: Wiley-Interscience.
- Ito Y and Ma Y (1996) pH zone refining counter-current chromatography. *Journal of Chromatography A* 753(1): 1–36.
- Pamplin BR (ed.) (1980) *International Series in the Science of the Solid State*, vol. 16, *Crystal Growth*, 2nd edn. Oxford: Pergamon.
- Parr NL (1960) *Zone Refining and Allied Techniques*. London: George Newnes.
- Pennanec R, Viron C, Blanchard S, and Lafosse M (2001) Original uses of the pH zone refining principle: adaptation to synthesis imperatives and to ionic compounds. *Journal of Liquid Chromatography and Related Technologies* 24(11–12): 1575–1591.
- Pfann WG (1958) *Zone Melting*. New York: Wiley.

## APPENDIX 1. NOMENCLATURE OF ANALYTICAL SCIENCE

The nomenclature used in Analytical Chemistry, and therefore in this encyclopedia, is that agreed and published by the International Union of Pure and Applied Chemistry (IUPAC) after deliberations by numerous specialist Commissions and Working Parties. The more fundamental definitions and values are to be found in *Quantities, Units and Symbols in Physical Chemistry* (1993), 2nd edn., Oxford: Blackwell Scientific Publications, prepared for publication by Mills I, Cvitas T, Homann K, *et al.* (the so-called Green Book). Some of the information from this publication is included in these appendices, as are the currently accepted values of atomic weights, isotopic, composition, etc., which are the responsibility of the Inorganic Chemistry Division.

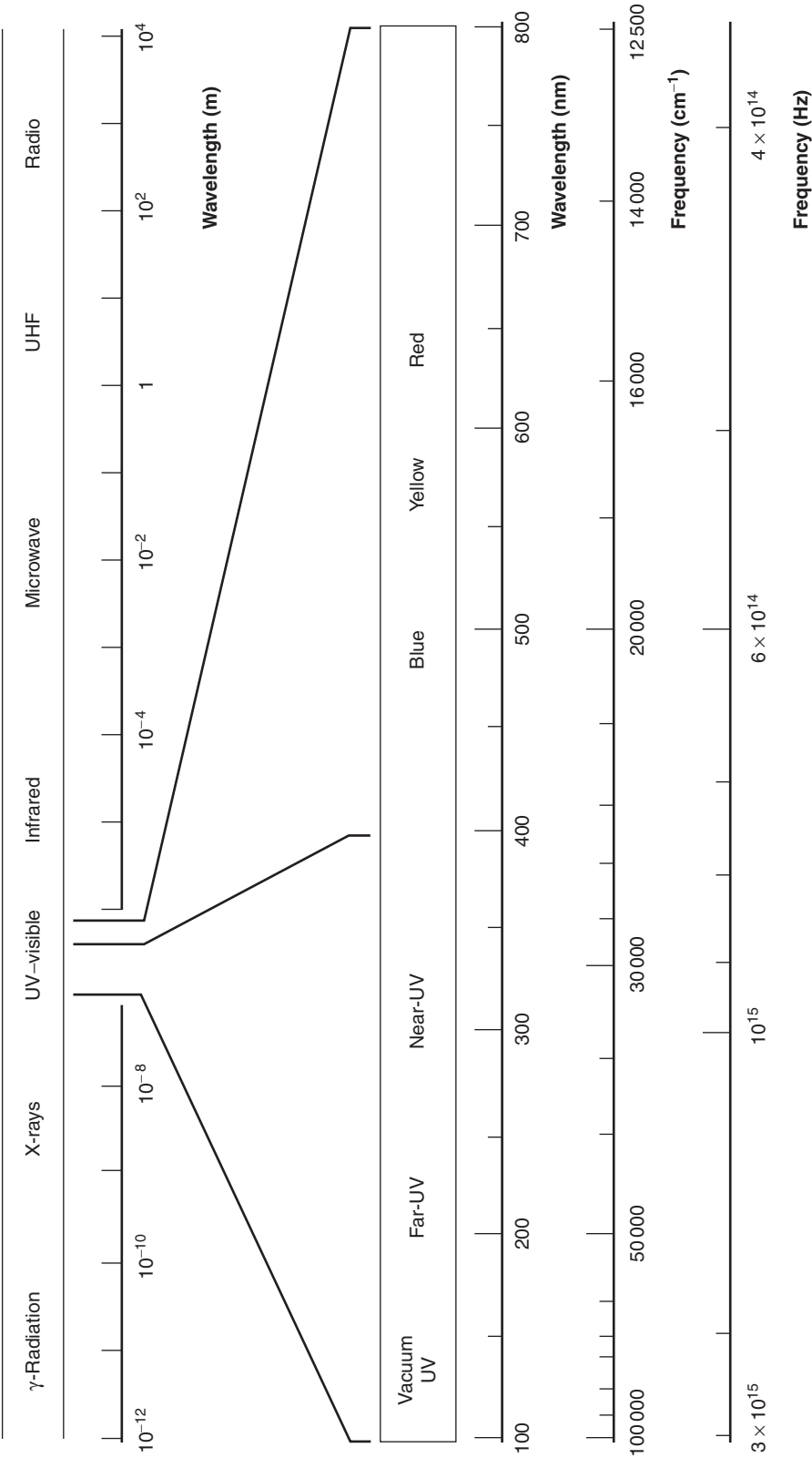
The specialized terminology of Analytical Chemistry has been the enduring responsibility of the Analytical Division of IUPAC. The terminology is to be found in the *Compendium of Analytical Nomenclature. Definitive Rules 1997* (1998), 3rd edn., Oxford: Blackwell Scientific Publications, prepared for publication by Inczédy J, Lengyel T, and Ure AM (the so-called Orange Book). Nomenclature, of course, reflects the relentless advance of science, so new and revised definitions are continually being published, initially for consultation by the international community, in *Pure and Applied Chemistry*.

Proposed IUPAC nomenclature is scrutinized for validity, accuracy and consistency across all branches of Chemistry, and over Science generally, by IUPAC's Interdivisional Committee on Terminology, Nomenclature and Symbols. The committee has representation from other International Scientific Unions: Biochemistry and Molecular Biology (IUBMB), Crystallography (IUCr), Nutritional Sciences (IUNS), Pharmacology (IUPHAR) and Pure and Applied Physics (IUPAP), and from the Bureau International des Poids et Mesures (BIPM) and the International Organization for Standardization (ISO), as well as Divisional and other representatives of IUPAC.

Details of current activities of IUPAC can be found at <http://www.iupac.org>

**Alan Townshend**

APPENDIX 2. WAVELENGTH SCALE



The range of electromagnetic radiation. The lower part is an enlargement of the UV-visible region.



# APPENDIX 3. DEFINITIONS AND SYMBOLS FOR UNITS

© 1993, Blackwell Science. Published by Elsevier Ltd. All rights reserved.

## The International System of Units (SI)

The International System of units (SI) was adopted by the 11th General Conference on Weights and Measures (CGPM) in 1960. It is a coherent system of units built from seven *SI base units*, one for each of the seven dimensionally independent base quantities: the meter, kilogram, second, ampere, kelvin, mole, and candela, for the dimensions length, mass, time, electric current, thermodynamic temperature, amount of substance, and luminous intensity, respectively. The *SI derived units* are expressed as products of powers of the base units, analogous to the corresponding relations between physical quantities but with numerical factors equal to unity.

In the International System there is only one SI unit for each physical quantity. This is either the appropriate SI base unit itself or the appropriate SI derived unit. However, any of the approved decimal prefixes, called *SI prefixes*, may be used to construct decimal multiples or submultiples of SI units.

It is recommended that only SI units be used in science and technology (with SI prefixes where appropriate). Where there are special reasons for making an exception to this rule, it is recommended always to define the units used in terms of SI units.

## Definitions of the SI Base Units

**Meter:** The meter is the length of path travelled by light in vacuum during a time interval of  $1/299\,792\,458$  of a second (17th CGPM, 1983).

**Kilogram:** The kilogram is the unit of mass; it is equal to the mass of the international prototype of the kilogram (3rd CGPM, 1901).

**Second:** The second is the duration of  $9\,192\,631\,770$  periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium-133 atom (13th CGPM, 1967).

**Ampere:** The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 meter apart in vacuum, would produce between these conductors a force equal to  $2 \times 10^{-7}$  newton per meter of length (9th CGPM, 1948).

**Kelvin:** The kelvin, unit of thermodynamic temperature, is the fraction  $1/273.16$  of the thermodynamic temperature of the triple point of water (13th CGPM, 1967).

**Mole:** The mole is the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kilogram of carbon-12. When the mole is used, the elementary entities must be specified and may be atoms, molecules, ions, electrons, other particles, or specified groups of such particles (14th CGPM, 1971).

### Examples of the use of the mole

1 mol of  $\text{H}_2$  contains  $\sim 6.022 \times 10^{23}$   $\text{H}_2$  molecules, or  $12.044 \times 10^{23}$  H atoms

1 mol of  $\text{HgCl}$  has a mass of 236.04 g

1 mol of  $\text{Hg}_2\text{Cl}_2$  has a mass of 472.08 g

1 mol of  $\text{Hg}_2^{2+}$  has a mass of 401.18 g and a charge of 192.97 kC

1 mol of  $\text{Fe}_{0.91}\text{S}$  has a mass of 82.88 g

1 mol of  $\text{e}^-$  has a mass of 548.60  $\mu\text{m}$  and a charge of  $-96.49$  kC

1 mol of photons whose frequency is  $5 \times 10^{14}$  Hz has energy of  $\sim 199.5$  kJ

**Candela:** The candela is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency  $540 \times 10^{12}$  hertz and that has a radiant intensity in that direction of  $(1/683)$  watt per steradian (16th CGPM, 1979).

“Appendix 3. Definitions and Symbols for Units” reproduced with permission from Mills I *et al.* (1993) *Quantities, Units and Symbols in Physical Chemistry*, 2nd edn., Oxford: Blackwell Science.

## Names and Symbols for the SI Base Units

The symbols listed here are internationally agreed and should not be changed in other languages or scripts.

<i>Physical quantity</i>	<i>Name of SI unit</i>	<i>Symbol for SI unit</i>
Length	Meter	m
Mass	Kilogram	kg
Time	Second	s
Electric current	Ampere	A
Thermodynamic temperature	Kelvin	K
Amount of substance	Mole	mol
Luminous intensity	Candela	cd

## SI Derived Units with Special Names and Symbols

<i>Physical quantity</i>	<i>Name of SI unit</i>	<i>Symbol for SI unit</i>	<i>Expression in terms of SI base units</i>
Frequency <sup>a</sup>	Hertz	Hz	$s^{-1}$
Force	Newton	N	$kg\,s^{-2}$
Pressure, stress	Pascal	Pa	$N\,m^{-2} = m^{-1}\,kg\,s^{-2}$
Energy, work, heat	Joule	J	$N\,m = m^2\,kg\,s^{-2}$
Power, radiant flux	Watt	W	$J\,s^{-1} = m^2\,kg\,s^{-3}$
Electric charge	Coulomb	C	$A\,s$
Electric potential, electromotive force	Volt	V	$J\,C^{-1} = m^2\,kg\,s^{-3}\,A^{-1}$
Electric resistance	Ohm	$\Omega$	$V\,A^{-1} = m^2\,kg\,s^{-3}\,A^{-2}$
Electric conductance	Siemens	S	$\Omega^{-1} = m^{-2}\,kg^{-1}\,s^3\,A^2$
Electric capacitance	Farad	F	$C\,V^{-1} = m^{-2}\,kg^{-1}\,s^4\,A^2$
Magnetic flux density	Tesla	T	$V\,s\,m^{-2} = kg\,s^{-2}\,A^{-1}$
Magnetic flux	Weber	Wb	$V\,s = m^2\,kg\,s^{-2}\,A^{-1}$
Inductance	Henry	H	$V\,A^{-1}\,s = m^2\,kg\,s^{-2}\,A^{-2}$
Celsius temperature <sup>b</sup>	Degree Celsius	$^{\circ}C$	K
Luminous flux	Lumen	lm	$cd\,sr$
Illuminance	Lux	lx	$cd\,sr\,m^{-2}$
Activity <sup>c</sup> (radioactive)	Becquerel	Bq	$s^{-1}$
Absorbed dose <sup>c</sup> (of radiation)	Gray	Gy	$J\,kg^{-1} = m^2\,s^{-2}$
Dose equivalent <sup>c</sup> (dose equivalent index)	Sievert	Sv	$J\,kg^{-1} = m^2\,s^{-2}$
Plane angle <sup>d</sup>	Radian	rad	$1 = m\,m^{-1}$
Solid angle <sup>d</sup>	Steradian	sr	$1 = m^2\,m^{-2}$

<sup>a</sup> For radial (angular) frequency and for angular velocity the unit  $rad\,s^{-1}$ , or simply  $s^{-1}$ , should be used, and this may *not* be simplified to Hz. The unit Hz should be used *only* for frequency in the sense of cycles per second.

<sup>b</sup> The Celsius temperature  $\theta$  is defined by the equation

$$\theta/^{\circ}C = T/K - 273.15$$

The SI unit of Celsius temperature is the degree Celsius,  $^{\circ}C$ , which is equal to the kelvin, K.  $^{\circ}C$  should be treated as a single symbol, with no space between the  $^{\circ}$  sign and the letter C. (The symbol  $^{\circ}K$ , and the symbol  $^{\circ}$ , should no longer be used).

<sup>c</sup> The units becquerel, gray and sievert are admitted for reasons of safeguarding human health.

<sup>d</sup> The units radian and steradian are described as 'SI supplementary units'. However, in chemistry, as well as in physics, they are usually treated as dimensionless derived units, and this was recognized by CIPM in 1980. Since they are then of dimension 1, this leaves open the possibility of including them or omitting them in expressions of SI derived units. In practice, this means that rad and sr may be used when appropriate and may be omitted if clarity is not lost thereby.

## SI Derived Units for Other Quantities

This Table gives examples of other SI derived units; the list is merely illustrative.

<i>Physical quantity</i>	<i>Expression in terms of SI base units</i>	
Area	$\text{m}^2$	
Volume	$\text{m}^3$	
Speed, velocity	$\text{m s}^{-1}$	
Angular velocity	$\text{s}^{-1}, \text{rad s}^{-1}$	
Acceleration	$\text{m s}^{-2}$	
Moment of force	$\text{Nm}$	$= \text{m}^2 \text{kg s}^{-2}$
Wavenumber	$\text{m}^{-1}$	
Density, mass density	$\text{kg m}^{-3}$	
Specific volume	$\text{m}^3 \text{kg}^{-1}$	
Amount concentration <sup>a</sup>	$\text{mol m}^{-3}$	
Molar volume	$\text{m}^3 \text{mol}^{-1}$	
Heat capacity, entropy	$\text{J K}^{-1}$	$= \text{m}^2 \text{kg s}^{-2} \text{K}^{-1}$
Molar heat capacity, molar entropy	$\text{J K}^{-1} \text{mol}^{-1}$	$= \text{m}^2 \text{kg s}^{-2} \text{K}^{-1} \text{mol}^{-1}$
Specific heat capacity, specific entropy	$\text{J K}^{-1} \text{kg}^{-1}$	$= \text{m}^2 \text{s}^{-2} \text{K}^{-1}$
Molar energy	$\text{J mol}^{-1}$	$= \text{m}^2 \text{kg s}^{-2} \text{mol}^{-1}$
Specific energy	$\text{J kg}^{-1}$	$= \text{m}^2 \text{s}^{-2}$
Energy density	$\text{J m}^{-3}$	$= \text{m}^{-1} \text{kg s}^{-2}$
Surface tension	$\text{N m}^{-1} = \text{J m}^{-2}$	$= \text{kg s}^{-2}$
Heat flux density, irradiance	$\text{W m}^{-2}$	$= \text{kg s}^{-3}$
Thermal conductivity	$\text{W m}^{-1} \text{K}^{-1}$	$= \text{m kg s}^{-3} \text{K}^{-1}$
Kinematic viscosity, diffusion coefficient	$\text{m}^2 \text{s}^{-1}$	
Dynamic viscosity	$\text{N s m}^{-2} = \text{Pas}$	$= \text{m}^{-1} \text{kg s}^{-1}$
Electric charge density	$\text{C m}^{-3}$	$= \text{m}^{-3} \text{s A}$
Electric current density	$\text{A m}^{-2}$	
Conductivity	$\text{S m}^{-1}$	$= \text{m}^{-3} \text{kg}^{-1} \text{s}^3 \text{A}^2$
Molar conductivity	$\text{S m}^2 \text{mol}^{-1}$	$= \text{kg}^{-1} \text{mol}^{-1} \text{s}^3 \text{A}^2$
Permittivity	$\text{F m}^{-1}$	$= \text{m}^{-3} \text{kg}^{-1} \text{s}^4 \text{A}^2$
Permeability	$\text{H m}^{-1}$	$= \text{m kg s}^{-2} \text{A}^{-2}$
Electric field strength	$\text{V m}^{-1}$	$= \text{m kg s}^{-3} \text{A}^{-1}$
Magnetic field strength	$\text{A m}^{-1}$	
Luminance	$\text{cd m}^2$	
Exposure (X- and $\gamma$ -rays)	$\text{C kg}^{-1}$	$= \text{kg}^{-1} \text{s A}$
Absorbed dose rate	$\text{Gy s}^{-1}$	$= \text{m}^2 \text{s}^{-3}$

<sup>a</sup>The words 'amount concentration' are an abbreviation for 'amount-of-substance concentration'. When there is not likely to be any ambiguity this quantity may be called simply 'concentration'.

## SI Prefixes

To signify decimal multiples and submultiples of SI units the following prefixes may be used.

<i>Submultiple</i>	<i>Prefix</i>	<i>Symbol</i>	<i>Multiple</i>	<i>Prefix</i>	<i>Symbol</i>
$10^{-1}$	deci	d	10	deca	da
$10^{-2}$	centi	c	$10^2$	hecto	h
$10^{-3}$	milli	m	$10^3$	kilo	k
$10^{-6}$	micro	$\mu$	$10^6$	mega	M
$10^{-9}$	nano	n	$10^9$	giga	G
$10^{-12}$	pico	p	$10^{12}$	tera	T
$10^{-15}$	femto	f	$10^{15}$	peta	P
$10^{-18}$	atto	a	$10^{18}$	exa	E
$10^{-21}$	zepto	z	$10^{21}$	zetta	Z
$10^{-24}$	yocto	y	$10^{24}$	yotta	Y

Prefix symbols should be printed in roman (upright) type with no space between the prefix and the unit symbol.

*Example* kilometer, km

When a prefix is used with a unit symbol, the combination is taken as a new symbol that can be raised to any power without the use of parentheses.

*Examples*  $1 \text{ cm}^3 = (0.01 \text{ m})^3 = 10^{-6} \text{ m}^3$   
 $1 \mu\text{s}^{-1} = (10^{-6} \text{ s})^{-1} = 10^{-6} \text{ s}^{-1}$   
 $1 \text{ V cm}^{-1} = 100 \text{ V m}^{-1}$   
 $1 \text{ mmol dm}^{-3} = 1 \text{ mol m}^{-3}$

A prefix should never be used on its own, and prefixes are not to be combined into compound prefixes.

*Example* pm, not  $\mu\mu\text{m}$

The names and symbols of decimal multiples and submultiples of the SI base unit of mass, the kg, which already contains a prefix, are constructed by adding the appropriate prefix to the word gram and symbol g.

*Examples* mg, not  $\mu\text{kg}$ ; Mg, not kkg

The SI prefixes are not to be used with  $^{\circ}\text{C}$ .

ISO has recommended standard representations of the prefix symbols for use with limited character sets.

## Units in Use Together with the SI

These units are not part of the SI, but it is recognized that they will continue to be used in appropriate contexts. SI prefixes may be attached to some of these units, such as milliliter, ml; millibar, mbar; megaelectronvolt, MeV; kilotonne, kt. A more extensive list of non-SI units, with conversion factors to the corresponding SI units, is given in the appendix, Conversion of Units.

<i>Physical quantity</i>	<i>Name of unit</i>	<i>Symbol for unit</i>	<i>Value in SI units</i>
Time	Minute	min	60 s
Time	Hour	h	3 600 s
Time	Day	d	86 400 s
Plane angle	Degree	$^{\circ}$	$(\pi/180) \text{ rad}$
Plane angle	Minute	'	$(\pi/10\,800) \text{ rad}$
Plane angle	Second	"	$(\pi/648\,000) \text{ rad}$
Length	Ångström <sup>a</sup>	Å	$10^{-10} \text{ m}$
Area	Barn	b	$10^{-28} \text{ m}^2$
Volume	Litre	l, L	$\text{dm}^3 = 10^{-3} \text{ m}^3$
Mass	Tonne	t	$\text{Mg} = 10^3 \text{ kg}$
Pressure	Bar <sup>a</sup>	bar	$10^5 \text{ Pa} = 10^5 \text{ N m}^{-2}$
Energy	Electronvolt <sup>b</sup>	eV ( $= e \times V$ )	$\approx 1.602\,18 \times 10^{-19} \text{ J}$
Mass	Unified atomic mass unit <sup>b,c</sup>	u ( $= m_{\text{a}}(^{12}\text{C})/12$ )	$\approx 1.660\,54 \times 10^{-27} \text{ kg}$

<sup>a</sup>The ångström and the bar are approved by CIPM for 'temporary use with SI units', until CIPM makes a further recommendation. However, they should not be introduced where they are not used at present.

<sup>b</sup>The values of these units in terms of the corresponding SI units are not exact, since they depend on the values of the physical constants  $e$  (for the electronvolt) and  $N_{\text{A}}$  (for the unified atomic mass unit), which are determined by experiment. See appendix, Fundamental Constants.

<sup>c</sup>The unified atomic mass unit is also sometimes called the dalton, with symbol Da, although the name and symbol have not been approved by CGPM.

## Atomic Units

For the purposes of quantum mechanical calculations of electronic wavefunctions, it is convenient to regard certain fundamental constants (and combinations of such constants) as though they were units. They are customarily called *atomic units* (abbreviated: au), and they may be regarded as forming a coherent system of units for the calculation of electronic properties in theoretical chemistry, although there is no authority from

CGPM for treating them as units. The first five atomic units in the table below have special names and symbols. Only four of these are independent; all others may be derived by multiplication and division in the usual way, and the table includes a number of examples.

The relation of atomic units to the corresponding SI units involves the values of the fundamental physical constants, and is therefore not exact. The numerical values in the table are based on the estimates of the appendix, Fundamental Constants. The numerical results of calculations in theoretical chemistry are frequently quoted in atomic units, or as numerical values in the form (*physical quantity*)/(*atomic unit*), so that the reader may make the conversion using the current best estimates of the physical constants.

<i>Physical quantity</i>	<i>Name of unit</i>	<i>Symbol for unit</i>	<i>Value of unit in SI</i>
Mass	Electron rest mass	$m_e$	$9.109\,389\,7\,(54) \times 10^{-31} \text{ kg}$
Charge	Elementary charge	$e$	$1.602\,177\,33\,(49) \times 10^{-19} \text{ C}$
Action	Planck constant/ $2\pi^a$	$\hbar$	$1.054\,572\,66\,(63) \times 10^{-34} \text{ J s}$
Length	Bohr <sup>a</sup>	$A_0$	$5.291\,772\,49\,(24) \times 10^{-11} \text{ m}$
Energy	Hartree <sup>a</sup>	$E_h$	$4.359\,748\,2\,(26) \times 10^{-18} \text{ J}$
Time		$\hbar/E_h$	$2.418\,884\,334\,1\,(29) \times 10^{-17} \text{ s}$
Velocity <sup>b</sup>		$a_0 E_h/\hbar$	$2.187\,691\,42\,(10) \times 10^6 \text{ m s}^{-1}$
Force		$E_h/a_0$	$8.238\,729\,5\,(25) \times 10^{-8} \text{ N}$
Momentum, linear		$\hbar/a_0$	$1.992\,853\,4\,(12) \times 10^{-24} \text{ N s}$
Electric current		$e E_h/\hbar$	$6.623\,621\,1\,(20) \times 10^{-3} \text{ A}$
Electric field		$E_h/e a_0$	$5.142\,208\,2\,(15) \times 10^{-11} \text{ V m}^{-1}$
Electric dipole moment		$e a_0$	$8.478\,357\,9\,(26) \times 10^{-30} \text{ C m}$
Magnetic flux density		$\hbar/e a_0^2$	$2.350\,518\,08\,(71) \times 10^5 \text{ T}$
Magnetic dipole moment <sup>c</sup>		$e \hbar/m_e$	$1.854\,803\,08\,(62) \times 10^{-23} \text{ J T}^{-1}$

<sup>a</sup>  $\hbar = h/2\pi$ ;  $a_0 = 4\pi\epsilon_0\hbar^2/m_e e^2$ ;  $E_h = \hbar^2/m^3 a_0^2$ .

<sup>b</sup> The numerical value of the speed of light, when expressed in atomic units, is equal to the reciprocal of the fine structure constant  $\alpha$ ;  $c/(\text{au of velocity}) = c\hbar/a_0 E_h = \alpha^{-1} \approx 137.035\,989\,5\,(61)$ .

<sup>c</sup> The atomic unit of magnetic dipole moment is twice the Bohr magneton,  $\mu_B$ .

## Dimensionless Quantities

Values of dimensionless physical quantities, more properly called ‘quantities of dimension one’, are often expressed in terms of mathematically exactly defined values denoted by special symbols or abbreviations, such as % (percent) and ppm (part per million). These symbols are then treated as units, and are used as such in calculations.

### Fractions (Relative Values, Yields, and Efficiencies)

Fractions such as relative uncertainty, mole fraction  $x$  (also called amount fraction, or number fraction), mass fraction  $w$ , and volume fraction  $\phi$ , are sometimes expressed in terms of the symbols summarized in the table below.

<i>Name</i>	<i>Symbol</i>	<i>Value</i>	<i>Examples</i>
Percent	%	$10^{-2}$	The isotopic abundance of carbon-13 expressed as a mole fraction is $x = 1.1\%$
Part per million	ppm	$10^{-6}$	The relative uncertainty in the Planck constant $h$ ( $= 6.626\,075\,5\,(40) \times 10^{-34} \text{ J s}$ ) is 0.60 ppm The mass fraction of impurities in a sample of copper was found to be less than 3 ppm, $w < 3 \text{ ppm}$

These multiples of the unit one are not part of the SI and ISO recommends that these symbols should never be used. They are also frequently used as units of ‘concentration’ without a clear indication of the type of fraction implied (e.g., mole fraction, mass fraction or volume fraction). To avoid ambiguity they should only be used in a context where the meaning of the quantity is carefully defined. Even then, the use of an appropriate SI unit ratio may be preferred.

**Deprecated Usage**

Adding extra labels to ppm and similar symbols, such as ppmv (meaning ppm by volume) should be avoided. Qualifying labels may be added to symbols for physical quantities, but never to units.

The symbols % and ppm should not be used in combination with other units. In table headings and in labeling the axes of graphs the use of % and ppm in the denominator is to be avoided. Although one would write  $x(^{13}\text{C}) = 1.1\%$ , the notation  $100\ x$  is to be preferred to  $x/\%$  in tables and graphs.

The further symbols listed in the table below are also to be found in the literature, but their use is to be deprecated. Note that the names and symbols for  $10^{-9}$  and  $10^{-12}$  in this table are based on the American system of names. In other parts of the world a billion sometimes stands for  $10^{12}$  and a trillion for  $10^{18}$ . Note also that the symbol ppt is sometimes used for part per thousand, and sometimes for part per trillion.

To avoid ambiguity the symbols ppb, ppt and pphm should not be used.

<i>Name</i>	<i>Symbol</i>	<i>Value</i>	<i>Examples</i>
Part per hundred	pph	$10^{-2}$	(Exactly equivalent to percent, %)
Part per thousand Per mille <sup>a</sup>	ppt ‰	$10^{-3}$	Atmospheric carbon dioxide is depleted in carbon-13 mass fraction by 7‰ (or 7 ppt) relative to ocean water
		$10^{-3}$	
Part per hundred million	pphm	$10^{-8}$	The mass fraction of impurity in the metal was less than 5 pphm
Part per billion	ppb	$10^{-9}$	The air quality standard for ozone is a volume fraction of $\phi = 120$ ppb
Part per trillion	ppt	$10^{-12}$	The natural background volume fraction of NO in air was found to be $\phi = 140$ ppt
Part per quadrillion	ppq	$10^{-15}$	

<sup>a</sup>The permille is also spelled per mille, per mill, permil, or pro mille.



# APPENDIX 5. PROPERTIES OF PARTICLES, ELEMENTS AND NUCLIDES

## Properties of Some Particles

Name	Symbol <sup>a</sup>	Spin <i>I</i>	Charge number <i>z</i>	Rest mass		Magnetic moment $\mu$ ( $\mu_N$ )	Mean life $\tau$ (s)
				<i>m</i> (u)	$mc^2$ (MeV)		
Photon	$\gamma$	1	0	0	0		
Neutrino	$\nu_e$	1/2	0	0	0		
Electron <sup>b</sup>	e	1/2	−1	$5.485\,799\,03\,(13) \times 10^{-4}$	$0.510\,999\,06\,(15)$	$1.001\,159\,652\,193\,(10)^c$	
Muon	$\mu^\pm$	1/2	$\pm 1$	$0.113\,428\,913\,(17)$	$105.658\,389\,(34)$	$1.001\,165\,923\,(8)^d$	$2.1973\,(4) \times 10^{-6}$
Pion	$\pi^\pm$	1	$\pm 1$	$0.149\,832\,3\,(8)$	$139.5679\,(7)$		$2.6030\,(24) \times 10^{-8}$
Pion	$\pi^0$	1	0	$0.144\,900\,8\,(9)$	$134.9743\,(8)$		$8.4\,(6) \times 10^{-17}$
Proton	p	1/2	1	$1.007\,276\,470\,(12)$	$938.272\,31\,(28)$	$2.792\,847\,386\,(63)$	
Neutron	n	1/2	0	$1.008\,664\,904\,(14)$	$939.565\,63\,(28)$	$-1.913\,042\,75\,(45)$	$889.1\,(21)$
Deuteron	d	1	1	$2.013\,553\,214\,(24)$	$1875.613\,39\,(53)$	$0.857\,437\,6\,(1)$	
Triton	t	1/2	1	$3.015\,500\,71\,(4)$	$2808.921\,78\,(85)$	$2.978\,960\,(1)$	
Helion	h	1/2	2	$3.014\,932\,23\,(4)$	$2808.392\,25\,(85)$	$-2.127\,624\,(1)$	
$\alpha$ -Particle	$\alpha$	0	2	$4.001\,506\,170\,(50)$	$3727.380\,3\,(11)$	0	

<sup>a</sup>The Particle Data Group recommends the use of italic symbols for particles and this has been adopted by many physicists.

<sup>b</sup>The electron as  $\beta$ -particle is sometimes denoted by  $\beta$ .

<sup>c</sup>The value is given in Bohr magnetons  $\mu/\mu_B$ ,  $\mu_B = eh/2m_e$ .

<sup>d</sup>The value is given as  $\mu/\mu_\mu$ , where  $\mu_\mu = eh/2m_\mu$ .

In nuclear physics and chemistry the masses of particles are often quoted as their energy equivalents (usually in megaelectronvolts). The unified atomic mass unit corresponds to  $931.494\,32\,(28)$  MeV.

Atom-like pairs of a positive particle and an electron are sometimes sufficiently stable to be treated as individual entities with special names.

### Examples

Positronium ( $e^+ e^-$ )  $m(e^+ e^-) = 1.097\,152\,503\,(26) \times 10^{-3}$  u

Muonium ( $\mu^+ e^-$ ; Mu)  $m(\text{Mu}) = 0.113\,977\,478\,(17)$  u

The positive or negative sign for the magnetic moment of a particle implies that the orientation of the magnetic dipole with respect to the angular momentum corresponds to the rotation of a positive or negative charge, respectively.

## Atomic Weights of the Elements 2001

Values from the 2001 table *Pure Appl. Chem.*, 2003, 75: 1107–1122. The values of zinc, krypton, molybdenum, and dysprosium have been modified. The approved name for element 110 is included; see *Pure Appl. Chem.*, 2003, 75: 1613–1615. The proposed name for element 111 is also included. Reproduced from <http://www.chem.qmul.ac.uk/iupac/AtWt/>. The table is also available from mirror sites in Germany, Japan, South Africa, and USA. World Wide Web version of atomic weight data prepared by G. P. Moss, originally from a file provided by D. R. Lide. Note: The claim (*Phys. Rev. Lett.*, 1999, 83: 1104) for elements 116 and 118 has been withdrawn (see *Chem. Eng. News*, 2001, 79: 10 (6 August 2001); 2002, 80: 12 (22 July 2002); *Phys. Rev. Lett.*, 2002, 89: 039901). Previous values may be consulted from the 1993 table, the 1995 table, the 1997 table or the 1999 table. The original paper should be consulted for details of the half-life of the radioisotopes quoted below. Also there is a report on the different isotopic compositions of some non-terrestrial materials. A number in parentheses indicates the uncertainty in the last digit of the atomic weight.

## List of Elements in Name Order

Atomic no.	Symbol	Name	Atomic wt.	Notes
89	Ac	Actinium	[227]	5
13	Al	Aluminum	26.981 538 (2)	
95	Am	Americium	[243]	5
51	Sb	Antimony	121.760 (1)	1
18	Ar	Argon	39.948 (1)	1, 2
33	As	Arsenic	74.921 60 (2)	
85	At	Astatine	[210]	5
56	Ba	Barium	137.327 (7)	
97	Bk	Berkelium	[247]	5
4	Be	Beryllium	9.012 182 (3)	
83	Bi	Bismuth	208.980 38 (2)	
107	Bh	Bohrium	[264]	5, 6
5	B	Boron	10.811 (7)	1, 2, 3
35	Br	Bromine	79.904 (1)	
48	Cd	Cadmium	112.411 (8)	1
55	Cs	Caesium	132.905 45 (2)	
20	Ca	Calcium	40.078 (4)	1
98	Cf	Californium	[251]	5
6	C	Carbon	12.010 7 (8)	1, 2
58	Ce	Cerium	140.116 (1)	1
17	Cl	Chlorine	35.453 (2)	3
24	Cr	Chromium	51.996 1 (6)	
27	Co	Cobalt	58.933 200 (9)	
29	Cu	Copper	63.546 (3)	2
96	Cm	Curium	[247]	5
110	Ds	Darmstadtium	[281]	5, 6
105	Db	Dubnium	[262]	5, 6
66	Dy	Dysprosium	162.500 (1)	1
99	Es	Einsteinium	[252]	5
68	Er	Erbium	167.259 (3)	1
63	Eu	Europium	151.964 (1)	1
100	Fm	Fermium	[257]	5
9	F	Fluorine	18.998 403 2 (5)	
87	Fr	Francium	[223]	5
64	Gd	Gadolinium	157.25 (3)	1
31	Ga	Gallium	69.723 (1)	
32	Ge	Germanium	72.64 (1)	

Continued

Continued

<i>Atomic no.</i>	<i>Symbol</i>	<i>Name</i>	<i>Atomic wt.</i>	<i>Notes</i>
79	Au	Gold	196.966 55 (2)	
72	Hf	Hafnium	178.49 (2)	
108	Hs	Hassium	[277]	5, 6
2	He	Helium	4.002 602 (2)	1, 2
67	Ho	Holmium	164.930 32 (2)	
1	H	Hydrogen	1.007 94 (7)	1, 2, 3
49	In	Indium	114.818 (3)	
53	I	Iodine	126.904 47 (3)	
77	Ir	Iridium	192.217 (3)	
26	Fe	Iron	55.845 (2)	
36	Kr	Krypton	83.798 (2)	1, 3
57	La	Lanthanum	138.905 5 (2)	1
103	Lr	Lawrencium	[262]	5
82	Pb	Lead	207.2 (1)	1, 2
3	Li	Lithium	[6.941 (2)]	1, 2, 3, 4
71	Lu	Lutetium	174.967 (1)	1
12	Mg	Magnesium	24.305 0 (6)	
25	Mn	Manganese	54.938 049 (9)	
109	Mt	Meitnerium	[268]	5, 6
101	Md	Mendelevium	[258]	5
80	Hg	Mercury	200.59 (2)	
42	Mo	Molybdenum	95.94 (2)	1
60	Nd	Neodymium	144.24 (3)	1
10	Ne	Neon	20.179 7 (6)	1, 3
93	Np	Neptunium	[237]	5
28	Ni	Nickel	58.693 4 (2)	
41	Nb	Niobium	92.906 38 (2)	
7	N	Nitrogen	14.006 7 (2)	1, 2
102	No	Nobelium	[259]	5
76	Os	Osmium	190.23 (3)	1
8	O	Oxygen	15.999 4 (3)	1, 2
46	Pd	Palladium	106.42 (1)	1
15	P	Phosphorus	30.973 761 (2)	
78	Pt	Platinum	195.078 (2)	
94	Pu	Plutonium	[244]	5
84	Po	Polonium	[209]	5
19	K	Potassium	39.098 3 (1)	1
59	Pr	Praseodymium	140.907 65 (2)	
61	Pm	Promethium	[145]	5
91	Pa	Protactinium	231.035 88 (2)	5
88	Ra	Radium	[226]	5
86	Rn	Radon	[222]	5
75	Re	Rhenium	186.207 (1)	
45	Rh	Rhodium	102.905 50 (2)	
111	Rg	Roentgenium	[272]	5, 6
37	Rb	Rubidium	85.467 8 (3)	1
44	Ru	Ruthenium	101.07 (2)	1
104	Rf	Rutherfordium	[261]	5, 6
62	Sm	Samarium	150.36 (3)	1
21	Sc	Scandium	44.955 910 (8)	
106	Sg	Seaborgium	[266]	5, 6
34	Se	Selenium	78.96 (3)	
14	Si	Silicon	28.085 5 (3)	2
47	Ag	Silver	107.868 2 (2)	1
11	Na	Sodium	22.989 770 (2)	
38	Sr	Strontium	87.62 (1)	1, 2
16	S	Sulfur	32.065 (5)	1, 2
73	Ta	Tantalum	180.947 9 (1)	
43	Tc	Technetium	[98]	5
52	Te	Tellurium	127.60 (3)	1
65	Tb	Terbium	158.925 34 (2)	
81	Tl	Thallium	204.383 3 (2)	

Continued

<i>Atomic no.</i>	<i>Symbol</i>	<i>Name</i>	<i>Atomic wt.</i>	<i>Notes</i>
90	Th	Thorium	232.038 1 (1)	1, 5
69	Tm	Thulium	168.934 21 (2)	
50	Sn	Tin	118.710 (7)	1
22	Ti	Titanium	47.867 (1)	
74	W	Tungsten	183.84 (1)	
112	Uub	Ununbium	[285]	5, 6
116	Uuh	Ununhexium		see Note above
118	Uuo	Ununoctium		see Note above
114	Uuq	Ununquadium	[289]	5, 6
92	U	Uranium	238.028 91 (3)	1, 3, 5
23	V	Vanadium	50.941 5 (1)	
54	Xe	Xenon	131.293 (6)	1, 3
70	Yb	Ytterbium	173.04 (3)	1
39	Y	Yttrium	88.905 85 (2)	
30	Zn	Zinc	65.409 (4)	
40	Zr	Zirconium	91.224 (2)	1

1. Geological specimens are known in which the element has an isotopic composition outside the limits for normal material. The difference between the atomic weight of the element in such specimens and that given in the Table may exceed the stated uncertainty.
2. The range in isotopic composition of normal terrestrial material prevents a more precise value being given; the tabulated value should be applicable to any normal material.
3. Modified isotopic compositions may be found in commercially available material because it has been subject to an undisclosed or inadvertent isotopic fractionation. Substantial deviations in atomic weight of the element from that given in the Table can occur.
4. Commercially available Li materials have atomic weights that range between 6.939 and 6.996; if a more accurate value is required, it must be determined for the specific material [range quoted for 1995 table 6.94 and 6.99].
5. Element has no stable nuclides. The value enclosed in brackets, e.g., [209], indicates the mass number of the longest-lived isotope of the element. However three such elements (Th, Pa, and U) do have a characteristic terrestrial isotopic composition, and for these an atomic weight is tabulated.
6. The names and symbols for elements 112–118 are under review. The temporary system recommended by J Chatt (1979) *Pure Appl. Chem.*, 51: 381–384 is used above. The names of elements 101–109 were agreed in 1997 (see *Pure Appl. Chem.*, 1997, 69: 2471–2473) and for element 110 in 2003 (see *Pure Appl. Chem.*, 2003, 75: 1613–1615). The proposed name for element 111 is also included.

**Electronic Configurations of the Elements (Ground States)**

Atomic no.	Element	Shell									
		<i>K</i>		<i>L</i>		<i>M</i>			<i>N</i>		
		<i>1s</i>	<i>2s</i>	<i>2p</i>	<i>3s</i>	<i>3p</i>	<i>3d</i>	<i>4s</i>	<i>4p</i>	<i>4d</i>	<i>4f</i>
1	Hydrogen	1									
2	Helium	2									
3	Lithium	2	1								
4	Beryllium	2	2								
5	Boron	2	2	1							
6	Carbon	2	2	2							
7	Nitrogen	2	2	3							
8	Oxygen	2	2	4							
9	Fluorine	2	2	5							
10	Neon	2	2	6							
11	Sodium	2	2	6	1						
12	Magnesium	2	2	6	2						
13	Aluminum	2	2	6	2	1					
14	Silicon	2	2	6	2	2					
15	Phosphorus	2	2	6	2	3					
16	Sulfur	2	2	6	2	4					
17	Chlorine	2	2	6	2	5					
18	Argon	2	2	6	2	6					
19	Potassium	2	2	6	2	6		1			
20	Calcium	2	2	6	2	6		2			
21	Scandium	2	2	6	2	6	1	2			
22	Titanium	2	2	6	2	6	2	2			
23	Vanadium	2	2	6	2	6	3	2			
24	Chromium	2	2	6	2	6	5	1			
25	Manganese	2	2	6	2	6	5	2			
26	Iron	2	2	6	2	6	6	2			
27	Cobalt	2	2	6	2	6	7	2			
28	Nickel	2	2	6	2	6	8	2			
29	Copper	2	2	6	2	6	10	1			
30	Zinc	2	2	6	2	6	10	2			
31	Gallium	2	2	6	2	6	10	2	1		
32	Germanium	2	2	6	2	6	10	2	2		
33	Arsenic	2	2	6	2	6	10	2	3		
34	Selenium	2	2	6	2	6	10	2	4		
35	Bromine	2	2	6	2	6	10	2	5		
36	Krypton	2	2	6	2	6	10	2	6		

Continued

Continued

Atomic no.	Element	Shell														
		K	L	M	N				O				P			
					4s	4p	4d	4f	5s	5p	5d	5f	6s	6p	6d	
37	Rubidium	2	8	18	2	6		1								
38	Strontium	2	8	18	2	6		2								
39	Yttrium	2	8	18	2	6	1	2								
40	Zirconium	2	8	18	2	6	2	2								
41	Niobium	2	8	18	2	6	4	1								
42	Molybdenum	2	8	18	2	6	5	1								
43	Technetium	2	8	18	2	6	6	1								
44	Ruthenium	2	8	18	2	6	7	1								
45	Rhodium	2	8	18	2	6	8	1								
46	Palladium	2	8	18	2	6	10									
47	Silver	2	8	18	2	6	10	1								
48	Cadmium	2	8	18	2	6	10	2								
49	Indium	2	8	18	2	6	10	2	1							
50	Tin	2	8	18	2	6	10	2	2							
51	Antimony	2	8	18	2	6	10	2	3							
52	Tellurium	2	8	18	2	6	10	2	4							
53	Iodine	2	8	18	2	6	10	2	5							
54	Xenon	2	8	18	2	6	10	2	6							
55	Caesium	2	8	18	2	6	10		2	6				1		
56	Barium	2	8	18	2	6	10		2	6				2		
57	Lanthanum	2	8	18	2	6	10		2	6		1		2		
58	Cerium	2	8	18	2	6	10	2	2	6				2		
59	Praseodymium	2	8	18	2	6	10	3	2	6				2		
60	Neodymium	2	8	18	2	6	10	4	2	6				2		
61	Promethium	2	8	18	2	6	10	5	2	6				2		
62	Samarium	2	8	18	2	6	10	6	2	6				2		
63	Europium	2	8	18	2	6	10	7	2	6				2		
64	Gadolinium	2	8	18	2	6	10	7	2	6		1		2		
65	Terbium	2	8	18	2	6	10	9	2	6				2		
66	Dysprosium	2	8	18	2	6	10	10	2	6				2		
67	Holmium	2	8	18	2	6	10	11	2	6				2		
68	Erbium	2	8	18	2	6	10	12	2	6				2		
69	Thulium	2	8	18	2	6	10	13	2	6				2		
70	Ytterbium	2	8	18	2	6	10	14	2	6				2		
71	Lutetium	2	8	18	2	6	10	14	2	6			1	2		
72	Hafnium	2	8	18	2	6	10	14	2	6			2	2		
73	Tantalum	2	8	18	2	6	10	14	2	6			3	2		
74	Tungsten	2	8	18	2	6	10	14	2	6			4	2		
75	Rhenium	2	8	18	2	6	10	14	2	6			5	2		
76	Osmium	2	8	18	2	6	10	14	2	6			6	2		
77	Iridium	2	8	18	2	6	10	14	2	6			9			
78	Platinum	2	8	18	2	6	10	14	2	6			9	1		
79	Gold	2	8	18	2	6	10	14	2	6			10	1		
80	Mercury	2	8	18	2	6	10	14	2	6			10	2		
81	Thallium	2	8	18	2	6	10	14	2	6			10	2	1	
82	Lead	2	8	18	2	6	10	14	2	6			10	2	2	
83	Bismuth	2	8	18	2	6	10	14	2	6			10	2	3	
84	Polonium	2	8	18	2	6	10	14	2	6			10	2	4	
85	Astatine	2	8	18	2	6	10	14	2	6			10	2	5	
86	Radon	2	8	18	2	6	10	14	2	6			10	2	6	

Continued



Continued

Atomic no.	Element	Shell											
		K	L	M	N	O				P			Q
						5s	5p	5d	5f	6s	6p	6d	
87	Francium	2	8	18	32	2	6	10		2	6		1
88	Radium	2	8	18	32	2	6	10		2	6		2
89	Actinium	2	8	18	32	2	6	10		2	6	1	2
90	Thorium	2	8	18	32	2	6	10		2	6	2	2
91	Protoactinium	2	8	18	32	2	6	10	2	2	6	1	2
92	Uranium	2	8	18	32	2	6	10	3	2	6	1	2
93	Neptunium	2	8	18	32	2	6	10	4	2	6	1	2
94	Plutonium	2	8	18	32	2	6	10	6	2	6		2
95	Americium	2	8	18	32	2	6	10	7	2	6		2
96	Curium	2	8	18	32	2	6	10	7	2	6	1	2
97	Berkelium	2	8	18	32	2	6	10	8	2	6	1	2
98	Californium	2	8	18	32	2	6	10	10	2	6		2
99	Einsteinium	2	8	18	32	2	6	10	11	2	6		2
100	Fermium	2	8	18	32	2	6	10	12	2	6		2
101	Mendelevium	2	8	18	32	2	6	10	13	2	6		2
102	Nobelium	2	8	18	32	2	6	10	14	2	6		2
103	Lawrencium	2	8	18	32	2	6	10	14	2	6	1	2

Reproduced from JE Stark, HG Wallace *Chemistry Data Book* (John Murray, 1982) by permission of John Stark and John Murray Ltd.

## Properties of Nuclides

The table contains the following properties of naturally occurring and some unstable nuclides:

Column

1.  $Z$  is the atomic number (number of protons) of the nuclide.
2. Symbol of the element.
3.  $A$  is the mass number of the nuclide. The “\*” sign denotes an unstable nuclide (for elements without naturally occurring isotopes it is the most stable nuclide) and the “#” sign a nuclide of sufficiently long lifetime to enable the determination of its isotopic abundance.
4. The atomic mass is given in unified atomic mass units,  $u = m_a(^{12}\text{C})/12$ , together with the standard errors in parentheses and applicable to the last digits quoted.
5. Isotopic abundances are given as mole fractions,  $x$ , of the corresponding atoms in percents. They were recommended in 1989 by the IUPAC Commission on Atomic Weights and Isotopic Abundances. The uncertainties given in parentheses are applicable to the last digits quoted and cover the range of probable variations in the materials as well as experimental errors.
6.  $I$  is the nuclear spin quantum number.
7. Under magnetic moment the maximum  $z$ -component expectation value of the magnetic dipole moment,  $m$ , in nuclear magnetons is given. The positive or negative sign implies that the orientation of the magnetic dipole with respect to the angular momentum corresponds to the rotation of a positive or negative charge, respectively. An asterisk indicates that more than one value is given in the original compilation. The value of highest precision or most recent date is given here.
8. Under quadrupole moment, the electric quadrupole moment area is given in units of square femtometres,  $\text{fm}^2 = 10^{-30} \text{m}^2$ , although most of the tables quote them in barns ( $1 \text{ barn} = 10^{-28} \text{m}^2 = 100 \text{ fm}^2$ ). The positive sign implies a prolate nucleus, the negative sign an oblate nucleus. The data for  $Z \leq 20$  were taken from the compilation by P. Pyykkö with values for Cl and Ca corrected by D. Sundholm (private communication), and the others from P. Raghavan. An asterisk indicates that more than one value is given in the original compilation.

<i>Z</i>	<i>Symbol</i>	<i>A</i>	<i>Atomic mass, m<sub>a</sub> (u)</i>	<i>Isotopic abundance, 100 ×</i>	<i>Nuclear spin, I</i>	<i>Magnetic moment, m (μ<sub>N</sub>)</i>	<i>Quadrupole moment, Q (fm<sup>2</sup>)</i>
1	H	1	1.007 825 035 (12)	99.985 (1)	1/2	+ 2.792 847 386 (63)	+ 0.286 0 (15)
		2	2.014 101 779 (24)	0.015 (1)	1	+ 0.857 438 230 (24)	
		3*	3.016 049 27 (4)		1/2	+ 2.978 962 479 (68)	
2	He	3	3.016 029 31 (4)	0.000 137 (3)	1/2	− 2.127 624 848 (66)	
		4	4.002 603 24 (5)	99.999 863 (3)	0	0	
3	Li	6	6.015 121 4 (7)	7.5 (2)	1	+ 0.822 056 67 (26)*	− 0.082 (4)
		7	7.016 003 0 (9)	92.5 (2)	3/2	+ 3.256 462 53 (40)*	− 4.01
4	Be	9	9.012 182 2 (4)	100	3/2	− 1.177 492 (17)*	+ 5.288 (38)
5	B	10	10.012 936 9 (3)	19.9 (2)	3	+ 1.800 644 75 (57)	+ 8.459 (24)
		11	11.009 305 4 (4)	80.1 (2)	3/2	+ 2.688 648 9 (10)	+ 4.059 (10)
6	C	12	12 (by definition)	98.90 (3)	0	0	
		13	13.003 354 826 (17)	1.10 (3)	1/2	+ 0.702 411 8 (14)	
		14*	14.003 241 982 (27)		0	0	
7	N	14	14.003 074 002 (26)	99.634 (9)	1	+ 0.403 761 00 (6)	+ 2.01 (2)
		15	15.000 108 97 (4)	0.366 (9)	1/2	− 0.283 188 842 (45)	
8	O	16	15.994 914 63 (5)	99.762 (15)	0	0	− 2.558 (22)
		17	16.999 131 2 (4)	0.038 (3)	5/2	− 1.893 80	
		18	17.999 160 3 (9)	0.200 (12)	0	0	
9	F	19	18.998 403 22 (15)	100	1/2	+ 2.628 868 (8)	
10	Ne	20	19.992 435 6 (22)	90.48 (3)	0	0	+ 10.155 (75)
		21	20.993 842 8 (21)	0.27 (1)	3/2	− 0.661 797 (5)	
		22	21.991 383 1 (18)	9.25 (3)	0	0	
11	Na	23	22.989 767 7 (10)	100	3/2	+ 2.217 655 6 (6)*	+ 10.06 (20)
12	Mg	24	23.985 042 3 (8)	78.99 (3)	0	0	+ 19.94 (20)
		25	24.985 837 4 (8)	10.00 (1)	5/2	− 0.855 465 (8)	
		26	25.982 593 7 (8)	11.01 (2)	0	0	
13	Al	27	26.981 538 6 (8)	100	5/2	+ 3.641 504 687 (65)	+ 14.03 (10)
14	Si	28	27.976 927 1 (7)	92.23 (1)	0	0	
		29	28.976 494 9 (7)	4.67 (1)	1/2	− 0.555 29 (3)	
		30	29.973 770 7 (7)	3.10 (1)	0	0	
15	P	31	30.973 762 0 (6)	100	1/2	+ 1.131 60 (3)	
16	S	32	31.972 070 70 (25)	95.02 (9)	0	0	− 6.78 (13)
		33	32.971 458 43 (23)	0.75 (1)	3/2	+ 0.643 821 2 (14)	
		34	33.967 866 65 (22)	4.21 (8)	0	0	
		36	35.967 080 62 (27)	0.02 (1)	0	0	
17	Cl	35	34.968 852 721 (69)	75.77 (5)	3/2	+ 0.821 874 3 (4)	− 8.11 (8)
		37	36.965 902 62 (11)	24.23 (5)	3/2	+ 0.684 123 6 (4)	− 6.39 (6)
18	Ar	36	35.967 545 52 (29)	0.337 (3)	0	0	
		38	37.962 732 5 (9)	0.063 (1)	0	0	
		40	39.962 383 7 (14)	99.600 (3)	0	0	
19	K	39	38.963 707 4 (12)	93.258 1 (44)	3/2	+ 0.391 507 31 (12)*	+ 5.9 (6)
		40	39.963 999 2 (12)	0.011 7 (1)	4	− 1.298 100 3 (34)	− 7.3 (7)
		41	40.961 825 4 (12)	6.730 2 (44)	3/2	+ 0.214 870 09 (22)	+ 7.2 (7)

Continued

<i>Z</i>	<i>Symbol</i>	<i>A</i>	<i>Atomic mass,</i> <i>m<sub>a</sub> (u)</i>	<i>Isotopic</i> <i>abundance,</i> <i>100 ×</i>	<i>Nuclear</i> <i>spin,</i> <i>I</i>	<i>Magnetic moment,</i> <i>m (μ<sub>N</sub>)</i>	<i>Quadrupole</i> <i>moment,</i> <i>Q (fm<sup>2</sup>)</i>
20	Ca	40	39.962 590 6 (13)	96.941 (18)	0	0	
		42	41.958 617 6 (13)	0.647 (9)	0	0	
		43	42.958 766 2 (13)	0.135 (6)	7/2	− 1.317 643 (7)	− 4.09 (8)
		44	43.955 480 6 (14)	2.086 (12)	0	0	
		46	45.953 689 (4)	0.004 (4)	0	0	
		48	47.952 533 (4)	0.187 (4)	0	0	
21	Sc	45	44.955 910 0 (14)	100	7/2	+ 4.756 486 6 (18)	− 22 (1)*
22	Ti	46	45.952 629 4 (14)	8.0 (1)	0	0	
		47	46.951 764 0 (11)	7.3 (1)	5/2	− 0.788 48 (1)	+ 29 (1)
		48	47.947 947 3 (11)	73.8 (1)	0	0	
		49	48.947 871 1 (11)	5.5 (1)	7/2	− 1.104 17 (1)	+ 24 (1)
		50	49.944 792 1 (12)	5.4 (1)	0	0	
23	V	50 <sup>#</sup>	49.947 160 9 (17)	0.250 (2)	6	+ 3.345 688 9 (14)	20.9 (40)*
		51	50.943 961 7 (17)	99.750 (2)	7/2	+ 5.148 705 73 (18)	− 5.2 (10)*
24	Cr	50	49.946 046 4 (17)	4.345 (13)	0	0	
		52	51.940 509 8 (17)	83.789 (18)	0	0	
		53	52.940 651 3 (17)	9.501 (17)	3/2	− 0.474 54 (3)	− 15 (5)*
		54	53.938 882 5 (17)	2.365 (7)	0	0	
25	Mn	55	54.938 047 1 (16)	100	5/2	+ 3.468 719 0 (9)	+ 33 (1)*
26	Fe	54	53.939 612 7 (15)	5.8 (1)	0	0	
		56	55.934 939 3 (16)	91.72 (30)	0	0	
		57	56.935 395 8 (16)	2.2 (1)	1/2	+ 0.090 623 00 (9)*	
		58	57.933 277 3 (16)	0.28 (1)	0	0	
27	Co	59	58.933 197 6 (16)	100	7/2	+ 4.627 (9)	+ 40.4 (40)*
28	Ni	58	57.935 346 2 (16)	68.077 (9)	0	0	
		60	59.930 788 4 (16)	26.223 (8)	0	0	
		61	60.931 057 9 (16)	1.140 (1)	3/2	− 0.750 02 (4)	+ 16.2 (15)
		62	61.928 346 1 (16)	3.634 (2)	0	0	
		64	63.927 967 9 (17)	0.926 (1)	0	0	
29	Cu	63	62.929 598 9 (17)	69.17 (3)	3/2	+ 2.2227 345 6 (14)*	− 21.1 (4)*
		65	64.927 792 9 (20)	30.83 (3)	3/2	+ 2.381 61 (19)*	− 19.5 (4)
30	Zn	64	63.929 144 8 (19)	48.6 (3)	0	0	
		66	65.926 034 7 (17)	27.9 (2)	0	0	
		67	66.927 129 1 (17)	4.1 (1)	5/2	+ 0.875 204 9 (11)*	+ 15.0 (15)
		68	67.924 845 9 (18)	18.8 (4)	0	0	
		70	69.925 325 (4)	0.6 (1)	0	0	
31	Ga	69	68.925 580 (3)	60.108 (9)	3/2	+ 2.016 589 (44)	+ 16.8*
		71	70.924 700 5 (25)	39.892 (9)	3/2	+ 2.562 266 (18)	+ 10.6*
32	Ge	70	69.924 249 7 (16)	21.23 (4)	0	0	
		72	71.992 078 9 (16)	27.66 (3)	0	0	
		73	72.923 462 6 (16)	7.73 (1)	9/2	− 0.879 467 7 (2)	− 17.3 (26)
		74	73.921 177 4 (15)	35.94 (2)	0	0	
		76	75.921 401 6 (17)	7.44 (2)	0	0	
33	As	75	74.921 594 2 (17)	100	3/2	+ 1.439 475 (65)	+ 31.4 (6)*

Continued

Continued

<i>Z</i>	<i>Symbol</i>	<i>A</i>	<i>Atomic mass,</i> <i>m<sub>a</sub> (u)</i>	<i>Isotopic</i> <i>abundance,</i> <i>100 ×</i>	<i>Nuclear</i> <i>spin,</i> <i>I</i>	<i>Magnetic moment,</i> <i>m (μ<sub>N</sub>)</i>	<i>Quadrupole</i> <i>moment,</i> <i>Q (fm<sup>2</sup>)</i>
34	Se	74	73.922 474 6 (16)	0.89 (2)	0	0	
		76	75.919 212 0 (16)	9.36 (1)	0	0	
		77	76.919 912 5 (16)	7.63 (6)	1/2	+ 0.535 074 24 (28)*	
		78	77.917 307 6 (16)	23.78 (9)	0	0	
		80	79.916 519 6 (19)	49.61 (10)	0	0	
		82	81.916 697 8 (23)	8.73 (6)	0	0	
35	Br	79	78.918 336 1 (26)	50.69 (7)	3/2	+ 2.106 400 (4)	+ 33.1 (4)
		81	80.916 289 (6)	49.31 (7)	3/2	+ 2.270 562 (4)	+ 27.6 (4)
36	Kr	78	77.920 396 (9)	0.35 (2)	0	0	
		80	79.916 380 (9)	2.25 (2)	0	0	
		82	81.913 482 (6)	11.6 (1)	0	0	
		83	82.914 135 (4)	11.5 (1)	9/2	− 0.970 669 (3)	+ 25.3 (5)
		84	83.911 507 (4)	57.0 (3)	0	0	
		86	85.910 616 (5)	17.3 (2)	0	0	
37	Rb	85	84.911 794 (3)	72.165 (20)	5/2	+ 1.353 351 5 (8)*	+ 22.8 (43)*
		87 <sup>#</sup>	86.909 187 (3)	27.835 (20)	3/2	+ 2.751 818 (2)	+ 13.2 (1)
38	Sr	84	83.913 430 (4)	0.56 (1)	0	0	
		86	85.909 267 2 (28)	9.86 (1)	0	0	
		87	86.908 884 1 (28)	7.00 (1)	9/2	− 1.093 603 0 (13)*	+ 33.5 (20)
		88	87.905 618 8 (28)	82.58 (1)	0	0	
39	Y	89	88.905 849 (3)	100	1/2	− 0.137 415 42 (34)*	
40	Zr	90	89.904 702 6 (26)	51.45 (3)	0	0	
		91	90.905 643 9 (26)	11.22 (4)	5/2	− 1.303 62 (2)	− 20.6 (10)
		92	91.905 038 6 (26)	17.15 (2)	0	0	
		94	93.906 314 8 (28)	17.38 (4)	0	0	
		96	95.908 275 (4)	2.80 (2)	0	0	
41	Nb	93	92.906 377 2 (27)	100	9/2	+ 6.170 5 (3)	− 32 (2)*
42	Mo	92	91.906 809 (4)	14.84 (4)	0	0	
		94	93.905 085 3 (26)	9.25 (3)	0	0	
		95	94.905 841 1 (22)	15.92 (5)	5/2	− 0.914 2 (1)	− 2.2 (1)*
		96	95.904 678 5 (22)	16.68 (5)	0	0	
		97	96.906 020 5 (22)	9.55 (3)	5/2	− 0.933 5 (1)	+ 25.5 (13)*
		98	97.905 407 3 (22)	24.13 (7)	0	0	
		100	99.907 477 (6)	9.63 (3)	0	0	
43	Tc	98*	97.907 215 (4)		6		
44	Ru	96	95.907 599 (8)	5.52 (6)	0	0	
		98	97.905 287 (7)	1.88 (6)	0	0	
		99	98.905 938 9 (23)	12.7 (1)	5/2	− 0.641 3 (51)*	+ 79 (4)
		100	99.904 219 2 (24)	12.6 (1)	0	0	
		101	100.905 581 9 (24)	17.0 (1)	5/2	− 0.718 8 (60)*	+ 45.7 (23)
		102	101.904 348 5 (25)	31.6 (2)	0	0	
		104	103.905 424 (6)	18.7 (2)	0	0	
45	Rh	103	102.905 500 (4)	100	1/2	− 0.088 40 (2)	
46	Pd	102	101.905 634 (5)	1.02 (1)	0	0	
		104	103.904 029 (6)	11.14 (8)	0	0	
		105	104.905 079 (6)	22.33 (8)	5/2	− 0.642 (3)	+ 66.0 (11)*
		106	105.903 478 (6)	27.33 (3)	0	0	
		108	107.903 895 (4)	26.46 (9)	0	0	
		110	109.905 167 (20)	11.72 (9)	0	0	

Continued

<i>Z</i>	<i>Symbol</i>	<i>A</i>	<i>Atomic mass,</i> <i>m<sub>a</sub> (u)</i>	<i>Isotopic</i> <i>abundance,</i> <i>100 ×</i>	<i>Nuclear</i> <i>spin,</i> <i>I</i>	<i>Magnetic moment,</i> <i>m (μ<sub>N</sub>)</i>	<i>Quadrupole</i> <i>moment,</i> <i>Q (fm<sup>2</sup>)</i>
47	Ag	107	106.905 092 (6)	51.839 (7)	1/2	− 0.113 679 65 (15)*	
		109	108.904 756 (4)	48.161 (7)	1/2	− 0.130 690 62 (22)*	
48	Cd	106	105.406 461 (7)	1.25 (4)	0	0	
		108	107.904 176 (6)	0.89 (2)	0	0	
		110	109.903 005 (4)	12.49 (12)	0	0	
		111	110.904 182 (3)	12.80 (8)	1/2	− 0.594 886 07 (84)*	
		112	111.902 757 (3)	24.13 (28)	0	0	
		113 <sup>#</sup>	112.904 400 (3)	12.22 (8)	1/2	− 0.622 300 92 (87)	
		114	113.903 357 (3)	28.73 (28)	0	0	
49	In	113	112.904 061 (4)	4.3 (2)	9/2	+ 5.528 9 (2)	+ 79.9
		115 <sup>#</sup>	114.903 882 (4)	95.7 (2)	9/2	+ 5.540 8 (2)	+ 81.0*
50	Sn	112	111.904 826 (5)	0.97 (1)	0	0	
		114	113.902 784 (4)	0.65 (1)	0	0	
		115	114.903 348 (3)	0.34 (1)	1/2	− 0.918 83 (7)	
		116	115.901 747 (3)	14.53 (11)	0	0	
		117	116.902 956 (3)	7.68 (7)	1/2	− 1.001 04 (7)	
		118	117.901 609 (3)	24.23 (11)	0	0	
		119	118.903 311 (3)	8.59 (4)	1/2	− 1.047 28 (7)	
		120	119.902 199 1 (29)	32.59 (10)	0	0	
		122	121.903 440 4 (30)	4.63 (3)	0	0	
51	Sb	121	120.903 821 2 (29)	57.36 (8)	5/2	+ 3.363 4 (3)	− 36 (4)*
		123	122.904 216 0 (24)	42.64 (8)	7/2	+ 2.549 8 (2)	− 49 (5)
52	Te	120	119.904 048 (21)	0.096 (2)	0	0	
		122	121.903 050 (3)	2.603 (4)	0	0	
		123	122.904 271 0 (22)	0.908 (2)	1/2	− 0.736 947 8 (8)	
		124	123.902 818 0 (18)	4.816 (6)	0	0	
		125	124.904 428 5 (25)	7.139 (6)	1/2	− 0.888 505 13 (43)*	
		126	125.903 309 5 (25)	18.95 (1)	0	0	
		128	127.904 463 (4)	31.69 (1)	0	0	
		130	129.906 229 (5)	33.80 (1)	0	0	
53	I	127	126.904 473 (5)	100	5/2	+ 2.813 273 (84)	− 78.9
54	Xe	124	123.905 894 2 (22)	0.10 (1)	0	0	
		126	125.904 281 (8)	0.09 (1)	0	0	
		128	127.903 531 2 (17)	1.91 (3)	0	0	
		129	128.904 780 1 (21)	26.4 (6)	1/2	− 0.777 976 3 (84)	
		130	129.903 509 4 (17)	4.1 (1)	0	0	
		131	130.905 072 (5)	21.2 (4)	3/2	+ 0.691 861 9 (39)	− 12.0 (12)
		132	131.904 144 (5)	26.9 (5)	0	0	
		134	133.905 395 (8)	10.4 (2)	0	0	
55	Cs	133	132.905 429 (7)	100	7/2	+ 2.582 024 6 (34)*	− 0.371 (14)*
		135	134.905 665 (7)	6.592 (18)	3/2	+ 0.837 943 (17)*	+ 16.0 (3)*
56	Ba	130	129.906 282 (8)	0.106 (2)	0	0	
		132	131.905 042 (9)	0.101 (2)	0	0	
		134	133.904 486 (7)	2.417 (27)	0	0	
		135	134.905 665 (7)	6.592 (18)	3/2	+ 0.837 943 (17)*	+ 16.0 (3)*
		136	135.904 553 (7)	7.854 (36)	0	0	
57	La	137	136.905 812 (6)	11.23 (4)	3/2	+ 0.937 365 (20)*	+ 24.5 (4)*
		138	137.905 232 (6)	71.70 (7)	0	0	

Continued



Continued

<i>Z</i>	<i>Symbol</i>	<i>A</i>	<i>Atomic mass,</i> <i>m<sub>a</sub> (u)</i>	<i>Isotopic</i> <i>abundance,</i> <i>100 ×</i>	<i>Nuclear</i> <i>spin,</i> <i>I</i>	<i>Magnetic moment,</i> <i>m (μ<sub>N</sub>)</i>	<i>Quadrupole</i> <i>moment,</i> <i>Q (fm<sup>2</sup>)</i>
57	La	138 <sup>#</sup>	137.907 105 (6)	0.090 2 (2)	5	+ 3.713 646 (7)	+ 45 (2)*
		139	138.906 347 (5)	99.909 8 (2)	7/2	+ 2.783 045 5 (9)	+ 20 (1)
58	Ce	136	135.907 140 (50)	0.19 (1)	0	0	
		138	137.905 985 (12)	0.25 (1)	0	0	
		140	139.905 433 (4)	88.48 (10)	0	0	
		142	141.909 241 (4)	11.08 (10)	0	0	
59	Pr	141	140.907 647 (4)	100	5/2	+ 4.275 4 (5)	− 5.89 (42)
60	Nd	142	141.907 719 (4)	27.13 (12)	0	0	
		143	142.909 810 (4)	12.18 (6)	7/2	− 1.065 (5)	− 63 (6)
		144	143.910 083 (4)	23.80 (12)	0	0	
		145	144.912 570 (4)	8.30 (6)	7/2	− 0.656 (4)	− 33 (3)
		146	145.013 113 (4)	17.19 (9)	0	0	
		148	147.916 889 (4)	5.76 (3)	0	0	
		150	149.920 887 (4)	5.64 (3)	0	0	
61	Pm	145*	144.912 743 (4)		5/2		
62	Sm	144	143.911 998 (4)	3.1 (1)	0	0	
		147 <sup>#</sup>	146.914 894 (4)	15.0 (2)	7/2	− 0.814 8 (7)	− 25.9 (26)
		148	147.914 819 (4)	11.3 (1)	0	0	
		149	148.917 180 (4)	13.8 (1)	7/2	− 0.671 7 (7)*	+ 7.5 (8)*
		150	149.917 273 (4)	7.4 (1)	0	0	
		152	151.919 728 (4)	26.7 (2)	0	0	
		154	153.922 205 (4)	22.7 (2)	0	0	
63	Eu	151	150.919 702 (8)	47.8 (15)	5/2	+ 3.471 7 (6)	+ 90.3 (10)*
		153	152.921 225 (4)	52.2 (15)	5/2	+ 1.533 0 (8)*	+ 241.2 (21)*
64	Gd	152	151.919 786 (4)	0.20 (1)	0	0	
		154	153.920 861 (4)	2.18 (3)	0	0	
		155	154.922 618 (4)	14.80 (5)	3/2	− 0.257 23 (35)*	+ 130 (2)*
		156	155.922 118 (4)	20.47 (4)	0	0	
		157	156.923 956 (4)	15.65 (3)	3/2	− 0.337 26 (55)*	+ 136 (2)*
		158	157.924 019 (4)	24.84 (12)	0	0	
		160	159.927 049 (4)	21.86 (4)	0	0	
65	Tb	159	158.925 342 (4)	100	3/2	+ 2.014 (4)	+ 143.2 (8)
66	Dy	156	155.924 277 (8)	0.06 (1)	0	0	
		158	157.924 403 (5)	0.10 (1)	0	0	
		160	159.925 193 (4)	2.34 (6)	0	0	
		161	160.926 930 (4)	18.9 (2)	5/2	− 0.480 3 (25)*	+ 250.7 (20)*
		162	161.926 795 (4)	25.5 (2)	0	0	
66	Dy	163	162.928 728 (4)	24.9 (2)	5/2	+ 0.672 6 (35)	+ 264.8 (21)
		164	163.929 171 (4)	28.2 (2)	0	0	
67	Ho	165	164.930 319 (4)	100	7/2	+ 4.173 (27)	+ 349 (3)*
68	Er	162	161.928 775 (4)	0.14 (1)	0	0	
		164	163.929 198 (4)	1.61 (1)	0	0	
		166	165.930 290 (4)	33.6 (2)	0	0	
		167	166.932 046 (4)	22.95 (15)	7/2	− 0.563 85 (12)	+ 356.5 (29)
		168	167.932 368 (4)	26.8 (2)	0	0	
		170	169.935 461 (4)	14.9 (2)	0	0	
69	Tm	169	168.934 212 (4)	100	1/2	− 0.231 6 (15)	

Continued

<i>Z</i>	<i>Symbol</i>	<i>A</i>	<i>Atomic mass,</i> <i>m<sub>a</sub> (u)</i>	<i>Isotopic</i> <i>abundance,</i> <i>100 ×</i>	<i>Nuclear</i> <i>spin,</i> <i>I</i>	<i>Magnetic moment,</i> <i>m (μ<sub>N</sub>)</i>	<i>Quadrupole</i> <i>moment,</i> <i>Q (fm<sup>2</sup>)</i>
70	Yb	168	167.933 894 (5)	0.13 (1)	0	0	
		170	169.934 759 (4)	3.05 (6)	0	0	
		171	170.936 323 (3)	14.3 (2)	1/2	+ 0.493 67 (1)*	
		172	171.936 378 (3)	21.9 (3)	0	0	
		173	172.938 208 (3)	16.12 (21)	5/2	− 0.679 89 (3)*	+ 280 (4)
		174	173.938 859 (3)	31.8 (4)	0	0	
		176	175.942 564 (4)	12.7 (2)	0	0	
71	Lu	175	174.940 770 (3)	97.41 (2)	7/2	+ 2.232 7 (11)*	+ 349 (2)*
		176 <sup>#</sup>	175.942 679 (3)	2.59 (2)	7	+ 3.169 2 (45)*	+ 492 (3)*
72	Hf	174	173.940 044 (4)	0.162 (3)	0	0	
		176	175.941 406 (4)	5.206 (5)	0	0	
		177	176.943 217 (3)	18.606 (4)	7/2	+ 0.793 5 (6)	+ 336.5 (29)*
		178	177.943 696 (3)	27.297 (4)	0	0	
		179	178.945 812 2 (29)	13.629 (6)	9/2	− 0.640 9 (13)	+ 379.3 (33)*
		180	179.946 545 7 (30)	35.100 (7)	0	0	
73	Ta	180	179.947 462 (4)	0.012 (2)	8		
		181	180.947 992 (3)	99.988 (2)	7/2	+ 2.370 5 (7)	+ 328 (6)*
74	W	180	179.946 701 (5)	0.13 (4)	0	0	
		182	181.948 202 (3)	26.3 (2)	0	0	
		183	182.950 220 (3)	14.3 (1)	1/2	+ 0.117 784 76 (9)	
		184	183.950 928 (3)	30.67 (15)	0	0	
		186	185.954 357 (4)	28.6 (2)	0	0	
75	Re	185	184.952 951 (3)	37.40 (2)	5/2	+ 3.187 1 (3)	+ 218 (2)*
		187 <sup>#</sup>	186.955 744 (3)	62.60 (2)	5/2	+ 3.219 7 (3)	+ 207 (2)*
76	Os	184	183.952 488 (4)	0.02 (1)	0	0	
		186	185.953 830 (4)	1.58 (30)	0	0	
		187	186.955 741 (3)	1.6 (3)	1/2	+ 0.064 651 89 (6)	
		188	187.955 830 (3)	13.3 (7)	0	0	
		189	188.958 137 (4)	16.1 (8)	3/2	+ 0.659 933 (4)	+ 85.6 (28)
		190	189.958 436 (4)	26.4 (12)	0	0	
		192	191.961 467 (4)	41.0 (8)	0	0	
77	Ir	191	190.960 584 (4)	37.3 (5)	3/2	+ 0.150 7 (6)*	+ 81.6 (9)*
		193	192.962 917 (4)	62.7 (5)	3/2	+ 0.163 7 (6)*	+ 75.1 (9)*
78	Pt	190	189.959 917 (7)	0.01 (1)	0	0	
		192	191.961 019 (5)	0.79 (6)	0	0	
		194	193.962 655 (4)	32.9 (6)	0	0	
		195	194.964 766 (4)	33.8 (6)	1/2	+ 0.609 52 (6)	
		196	195.964 926 (4)	25.3 (6)	0	0	
		198	197.967 869 (6)	7.2 (2)	0	0	
79	Au	197	196.966 543 (4)	100	3/2	+ 0.148 158 (8)*	+ 54.7 (16)*
80	Hg	196	195.965 807 (5)	0.15 (1)	0	0	
		198	197.966 743 (4)	9.97 (8)	0	0	
		199	198.968 254 (4)	16.87 (10)	1/2	+ 0.505 885 49 (85)	
		200	199.968 300 (4)	23.10 (16)	0	0	
		201	200.970 277 (4)	13.18 (8)	3/2	− 0.560 225 7 (14)*	+ 38.5 (40)*
		202	201.970 617 (4)	29.86 (20)	0	0	
		204	203.973 467 (5)	6.87 (4)	0	0	
81	Tl	203	202.972 320 (5)	29.524 (14)	1/2	+ 1.622 257 87 (12)	
		205	204.974 401 (5)	70.476 (14)	1/2	+ 1.638 214 61 (12)	

Continued

Continued

<i>Z</i>	<i>Symbol</i>	<i>A</i>	<i>Atomic mass,</i> <i>m<sub>a</sub> (u)</i>	<i>Isotopic</i> <i>abundance,</i> <i>100 ×</i>	<i>Nuclear</i> <i>spin,</i> <i>I</i>	<i>Magnetic moment,</i> <i>m (μ<sub>N</sub>)</i>	<i>Quadrupole</i> <i>moment,</i> <i>Q (fm<sup>2</sup>)</i>
82	Pb	204	203.973 020 (5)	1.4 (1)	0	0	
		206	205.974 440 (4)	24.1 (1)	0	0	
		207	206.975 872 (4)	22.1 (1)	1/2	+ 0.582 583 (9)*	
		208	207.976 627 (4)	52.4 (1)	0	0	
83	Bi	209	208.980 374 (5)	100	9/2	+ 4.110 6 (2)	− 37.0 (26)*
84	Po	209*	208.982 404 (5)		1/2		
85	At	210*	209.987 126 (12)				
86	Rn	222*	222.017 571 (3)		0	0	
87	Fr	223*	223.019 733 (4)		3/2	+ 1.17 (2)	+ 117 (1)
88	Ra	226*	226.025 403 (3)		0	0	
89	Ac	227*	227.027 750 (3)		3/2	+ 1.1 (1)	+ 170 (20)
90	Th	232 <sup>#</sup>	232.038 050 8 (23)	100	0	0	
91	Pa	231*	231.035 880 (3)		3/2	2.01 (2)	− 172 (5)
92	U	233*	233.039 628 (3)		5/2	0.59 (5)	+ 366.3 (8)
		234 <sup>#</sup>	234.040 946 8 (24)	0.005 5 (5)	0	0	
		235 <sup>#</sup>	235.043 924 2 (24)	0.720 0 (12)	7/2	− 0.38 (3)*	+ 455 (9)*
		238 <sup>#</sup>	238.050 784 7 (23)	99.274 5 (60)	0	0	
93	Np	237*	237.048 167 8 (23)		5/2	+ 3.14 (4)	+ 388.6 (6)
94	Pu	244*	244.064 199 (5)		0		
95	Am	243*	243.061 375 (3)		5/2	+ 1.61 (4)	+ 420 (130)
96	Cm	247*	247.070 347 (5)				
97	Bk	247*	247.070 300 (6)				
98	Cf	251*	251.079 580 (5)				
99	Es	252*	252.082 944 (23)				
100	Fm	257*	257.095 099 (8)				
101	Md	258*	258.098 57 (22)				
102	No	259*	259.100 931 (12)				
103	Lr	260*	260.105 320 (60)				
104	Unq	261*	261.108 69 (22)				
105	Unp	262*	262.113 76 (16)				
106	Unh	263*	263.118 22 (13)				
107	Uns	262*	263.122 93 (45)				
108	Uno	265*	265.130 16 (99)				
109	Une	266*	266.137 64 (45)				

# APPENDIX 4. FUNDAMENTAL PHYSICAL CONSTANTS

The following values were recommended by the CODATA Task Group on Fundamental Constants in 1986. For each constant the standard deviation uncertainty in the least significant digits is given in parentheses.

## Fundamental constants

Quantity	Symbol	2002 Value (standard uncertainty) <sup>a</sup>	Unit
Speed of light in vacuum	$c_0$	299 792 458	$\text{m s}^{-1}$
Magnetic constant	$\mu_0$	$4\pi \times 10^{-7}$	$\text{H m}^{-1}$
Electric constant	$\epsilon_0 = 1/\mu_0 c_0^2$	8.854 187 817 ...	$\times 10^{-12} \text{ F m}^{-1}$
Planck constant	$h$	6.626 069 3 (11)	$\times 10^{-34} \text{ J s}$
Elementary charge (charge on a proton)	$e$	1.602 176 53 (14)	$\times 10^{-19} \text{ C}$
Electron rest mass	$m_e$	9.109 382 6 (16)	$\times 10^{-31} \text{ kg}$
Proton rest mass	$m_p$	1.672 621 71 (29)	$\times 10^{-27} \text{ kg}$
Atomic mass constant (dalton, or unified atomic mass unit, $m(^{12}\text{C})/12$ )	$m_u$ $= \text{Da} = \text{u}$	1.660 538 86 (28)	$\times 10^{-27} \text{ kg}$
Avogadro constant	$L, N_A$	6.022 141 5 (10)	$\times 10^{23} \text{ mol}^{-1}$
Boltzmann constant	$k, k_B$	1.380 650 5 (24)	$\times 10^{-23} \text{ J K}^{-1}$
Faraday constant	$F$	96 485.33 83 (83)	$\text{C mol}^{-1}$
Gas constant	$R$	8.314 472 (15)	$\text{J mol}^{-1} \text{ K}^{-1}$
Fine structure constant	$\alpha$	7.297 352 568 (24)	$\times 10^{-3}$
Bohr radius	$a_0$	0.529 177 210 8 (18)	$\times 10^{-10} \text{ m}$
Hartree energy	$E_h$	4.359 744 17 (75)	$\times 10^{-18} \text{ J}$
Rydberg constant	$R_\infty$	10 973 731.568 525 (73)	$\text{m}^{-1}$
Bohr magneton	$\mu_B$	9.274 009 49 (80)	$\times 10^{-24} \text{ J T}^{-1}$
Landé $g$ factor for free electron	$g$	2.002 319 304 371 8 (75)	
Nuclear magneton	$\mu_N$	5.050 783 43 (43)	$\times 10^{-27} \text{ J T}^{-1}$
Relative atomic mass of the electron	$A_r(e)$	5.485 799 094 5 (24)	$\times 10^{-4}$
Newtonian constant of gravitation	$G$	6.674 2 (10)	$\times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$

<sup>a</sup>The standard uncertainty given in parenthesis (i.e., the estimated standard deviation of the value quoted), applies to the least significant digits of each constant.

Source: The National Institute of Standards and Technology (NIST) Reference on Constants, Units, and Uncertainties.

## Values of Common Mathematical Constants

Mathematical constant	Symbol	Value
Ratio of circumference to diameter of a circle	$\pi$	3.141 592 653 59
Base of natural logarithms	$e$	2.718 281 828 46
Natural logarithm of 10	$\ln 10$	2.302 585 092 99

# APPENDIX 6. CONVERSION OF UNITS

© 1993, Blackwell Science. Published by Elsevier Ltd. All rights reserved.

“Appendix 6. Conversion of Units” reproduced with permission from Mills I *et al.* (1993) *Quantities, Units and Symbols in Physical Chemistry*, 2nd edn., Oxford: Blackwell Science.

The table below gives conversion factors from a variety of units to the corresponding SI unit. Examples of the use of this table have already been given in the preceding section. For each physical quantity the name is given, followed by the recommended symbol(s). Then the SI unit is given, followed by the esu, emu, Gaussian unit (Gau), atomic unit (au), and other units in common use, with their conversion factors to SI. The constant  $\zeta$  which occurs in some of the electromagnetic conversion factors is the (exact) pure number  $2.997\,924\,58 \times 10^{10} = c_0/(\text{cm s}^{-1})$ .

The inclusion of non-SI units in this table should not be taken to imply that their use is to be encouraged. With some exceptions, SI units are always to be preferred to non-SI units. However, since many of the units below are to be found in the scientific literature, it is convenient to tabulate their relation to the SI.

For convenience units in the esu and Gaussian systems are quoted in terms of the four dimensions *length*, *mass*, *time*, and *electric charge*, by including the franklin (Fr) as an abbreviation for the electrostatic unit of charge and  $4\pi\epsilon_0$  as a constant with dimensions  $(\text{charge})^2/(\text{energy} \times \text{length})$ . This gives each physical quantity the same dimensions in all systems, so that all conversion factors are pure numbers. The factors  $4\pi\epsilon_0$  and the Fr may be eliminated by writing  $\text{Fr} = \text{esu of charge} = \text{erg}^{1/2} \text{cm}^{1/2} = \text{cm}^{3/2} \text{g}^{1/2} \text{s}^{-1}$ , and  $4\pi\epsilon_0 = \epsilon_0^{(\text{ir})} = 1 \text{ Fr}^2 \text{erg}^{-1} \text{cm}^{-1} = 1$ , to recover esu expressions in terms of three base units. The symbol Fr should be regarded as a compact representation of ‘esu of charge’.

Conversion factors are either given exactly (when the = sign is used), or they are given to the approximation that the corresponding physical constants are known (when the  $\approx$  sign is used). In the latter case the uncertainty is always less than  $\pm 5$  in the last digit quoted.

Name	Symbol	Relation to SI
<i>Length, l</i>		
Meter (SI unit)	m	
Centimeter (cgs unit)	cm	$= 10^{-2} \text{ m}$
Bohr (au)	$a_0, b$	$= 4\pi\epsilon_0 \hbar^2 / m_e e^2 \approx 5.291\,77 \times 10^{-11} \text{ m}$
Ångström	Å	$= 10^{-10} \text{ m}$
Micron	$\mu$	$= \mu\text{m} = 10^{-6} \text{ m}$
x unit	X	$\approx 1.002 \times 10^{-13} \text{ m}$
Fermi	f, fm	$= \text{fm} = 10^{-15} \text{ m}$
Inch	in	$= 2.54 \times 10^{-2} \text{ m}$
Foot	ft	$= 12 \text{ in} = 0.3048 \text{ m}$
Yard	yd	$= 3 \text{ ft} = 0.9144 \text{ m}$
Mile	mi	$= 1760 \text{ yd} = 1609.344 \text{ m}$
Nautical mile		$= 1852 \text{ m}$
<i>Area, A</i>		
Square meter (SI unit)	$\text{m}^2$	
Barn	b	$= 10^{-28} \text{ m}^2$
Acre		$\approx 4046.856 \text{ m}^2$
Are	a	$= 100 \text{ m}^2$
Hectare	ha	$= 10^4 \text{ m}^2$
<i>Volume, V</i>		
Cubic meter (SI unit)	$\text{m}^3$	
Liter	l, L	$= \text{dm}^3 = 10^{-3} \text{ m}^3$
Lambda	$\lambda$	$= \mu\text{l} = 10^{-6} \text{ dm}^3$
Barrel (US)		$\approx 158.987 \text{ dm}^3$
Gallon (US)	gal (US)	$= 3.785\,41 \text{ dm}^3$
Gallon (UK)	gal (UK)	$= 4.546\,09 \text{ dm}^3$

Continued

Continued

<i>Name</i>	<i>Symbol</i>	<i>Relation to SI</i>
<i>Mass, m</i>		
Kilogram (SI unit)	kg	
Gram (cgs unit)	g	$= 10^{-3} \text{ kg}$
Electron mass (au)	$m_e$	$\approx 9.109\,39 \times 10^{-31} \text{ kg}$
Unified atomic mass unit, dalton	u, Da	$= m_a(^{12}\text{C})/12 \approx 1.660\,540 \times 10^{-27} \text{ kg}$
Tonne	t	$= \text{Mg} = 10^3 \text{ kg}$
Pound (avoirdupois)	lb	$= 0.453\,592\,37 \text{ kg}$
Ounce (avoirdupois)	oz	$\approx 28.3495 \text{ g}$
Ounce (troy)	oz (troy)	$\approx 31.1035 \text{ g}$
Grain	gr	$= 64.798\,91 \text{ mg}$
<i>Time, t</i>		
Second (SI, cgs unit)	s	
Au of time	$\hbar/E_h$	$\approx 2.418\,88 \times 10^{-17} \text{ s}$
Minute	min	$= 60 \text{ s}$
Hour	h	$= 3600 \text{ s}$
Day <sup>a</sup>	d	$= 86\,400 \text{ s}$
Year <sup>b</sup>	a	$\approx 31\,556\,952 \text{ s}$
Svedberg	Sv	$= 10^{-13} \text{ s}$
<i>Acceleration, a</i>		
SI unit	$\text{m s}^{-2}$	$= 9.806\,65 \text{ m s}^{-2}$
Standard acceleration of free fall	$g_n$	
Gal, galileo	Gal	$= 10^{-2} \text{ m s}^{-2}$
<i>Force, F</i>		
Newton (SI unit) <sup>c</sup>	N	$= \text{kg m s}^{-2}$
Dyne (cgs unit)	dyn	$= \text{g cm s}^{-2} = 10^{-5} \text{ N}$
au of force	$E_h/a_0$	$\approx 8.238\,73 \times 10^{-8} \text{ N}$
Kilogram-force	kgf	$= 9.806\,65 \text{ N}$
<i>Energy, U</i>		
Joule (SI unit)	J	$= \text{kg m}^2 \text{ s}^{-2}$
Erg (cgs unit)	erg	$= \text{g cm}^2 \text{ s}^{-2} = 10^{-7} \text{ J}$
Hartree (au)	$E_h$	$= \hbar^2/m_e a_0^2 \approx 4.359\,75 \times 10^{-18} \text{ J}$
Rydberg	Ry	$= E_h/2 \approx 2.179\,87 \times 10^{-18} \text{ J}$
Electronvolt	eV	$= e \times V \approx 1.602\,18 \times 10^{-19} \text{ J}$
Calorie, thermochemical	cal <sub>th</sub>	$= 4.184 \text{ J}$
Calorie, international	cal <sub>I</sub>	$= 4.1868 \text{ J}$
15°C calorie	cal <sub>15</sub>	$\approx 4.1855 \text{ J}$
Liter atmosphere	1 atm	$= 101.325 \text{ J}$
British thermal unit	Btu	$= 1055.06 \text{ J}$
<i>Pressure, p</i>		
Pascal (SI unit)	Pa	$= \text{N m}^{-2} = \text{kg m}^{-1} \text{ s}^{-2}$
Atmosphere	atm	$= 101\,325 \text{ Pa}$
Bar	bar	$= 10^5 \text{ Pa}$
Torr	Torr	$= (101\,325/760) \text{ Pa} \approx 133.322 \text{ Pa}$
Millimeter of mercury (conventional)	mmHg	$= 13.5951 \times 980.665 \times 10^{-2} \text{ Pa} \approx 133.322 \text{ Pa}$
Pounds per square inch	psi	$\approx 6.894\,757 \times 10^3 \text{ Pa}$
<i>Power, P</i>		
Watt (SI unit)	W	$= \text{kg m}^2 \text{ s}^{-3}$
Horse power	hp	$= 745.7 \text{ W}$
<i>Action, L, J (angular momentum)</i>		
SI unit	J s	$= \text{kg m}^2 \text{ s}^{-1}$
cgs unit	erg s	$= 10^{-7} \text{ J s}$
au of action	$\hbar$	$= \hbar/2\pi \approx 1.054\,57 \times 10^{-34} \text{ J s}$



Continued

Name	Symbol	Relation to SI
<i>Dynamic viscosity, <math>\eta</math></i>		
SI unit	Pa s	$= \text{kg m}^{-1} \text{s}^{-1}$
Poise	P	$= 10^{-1} \text{Pa s}$
Centipoise	cP	$= \text{mPa s}$
<i>Kinematic viscosity, <math>\nu</math></i>		
SI unit	$\text{m}^2 \text{s}^{-1}$	
Stokes	St	$= 10^{-4} \text{m}^2 \text{s}^{-1}$
<i>Thermodynamic temperature, <math>T</math></i>		
Kelvin (SI unit)	K	
Degree Rankine <sup>d</sup>	$^{\circ}\text{R}$	$= (5/9) \text{K}$
<i>Entropy, <math>S</math></i>		
<i>Heat capacity, <math>C</math></i>		
SI unit	$\text{J K}^{-1}$	
Clausius	Cl	$= \text{cal}_{\text{th}}/\text{K} = 4.184 \text{J K}^{-1}$
<i>Molar entropy, <math>S_{\text{m}}</math></i>		
<i>Molar heat capacity, <math>C_{\text{m}}</math></i>		
SI unit	$\text{J K}^{-1} \text{mol}^{-1}$	
Entropy unit	e.u.	$= \text{cal}_{\text{th}} \text{K}^{-1} \text{mol}^{-1} = 4.184 \text{J K}^{-1} \text{mol}^{-1}$
<i>Molar volume, <math>V_{\text{m}}</math></i>		
SI unit	$\text{m}^3 \text{mol}^{-1}$	
Amagat <sup>e</sup>	amagat	$= V_{\text{m}} \text{ of real gas at 1 atm and 273.15 K}$ $\approx 22.4 \times 10^{-3} \text{m}^3 \text{mol}^{-1}$
<i>Amount density, <math>1/V_{\text{m}}</math></i>		
SI unit	$\text{mol m}^{-3}$	
Amagat <sup>e</sup>	amagat	$= 1/V_{\text{m}} \text{ for a real gas at 1 atm and 273.15 K}$ $\approx 44.6 \text{mol m}^{-3}$
<i>Plane angle, <math>\alpha</math></i>		
Radian (SI unit)	rad	
Degree	$^{\circ}$	$= \text{rad} \times 2\pi/360 \approx (1/57.295\ 78) \text{rad}$
Minute	'	$= \text{degree}/60$
Second	"	$= \text{degree}/3600$
Grade	grad	$= \text{rad} \times 2\pi/400 \approx (1/63.661\ 98) \text{rad}$
<i>Radioactivity, <math>A</math></i>		
Becquerel (SI unit)	Bq	$= \text{s}^{-1}$
Curie	Ci	$= 3.7 \times 10^{10} \text{Bq}$
<i>Absorbed dose of radiation<sup>f</sup></i>		
Gray (SI unit)	Gy	$= \text{J kg}^{-1}$
Rad	rad	$= 0.01 \text{Gy}$
<i>Dose equivalent</i>		
Sievert (SI unit)	Sv	$= \text{J kg}^{-1}$
Rem	rem	$\approx 0.01 \text{Sv}$
<i>Electric current, <math>I</math></i>		
Ampere (SI unit)	A	
esu, Gau	$(10/\zeta) \text{A}$	$\approx 3.335\ 64 \times 10^{-10} \text{A}$
Biot (emu)	Bi	$= 10 \text{A}$
au	$eE_{\text{H}}/\hbar$	$\approx 6.623\ 62 \times 10^{-3} \text{A}$
<i>Electric charge, <math>Q</math></i>		
Coulomb (SI unit)	C	$= \text{A s}$
Franklin (esu, Gau)	Fr	$= (10/\zeta) \text{C} \approx 3.335\ 64 \times 10^{-10} \text{C}$
emu (abcoulomb)		$= 10 \text{C}$
Proton charge (au)	$e$	$\approx 1.602\ 18 \times 10^{-19} \text{C} \approx 4.803\ 21 \times 10^{-10} \text{Fr}$

Continued

Continued

Name	Symbol	Relation to SI
<i>Charge density, <math>\rho</math></i>		
SI unit	$\text{C m}^{-3}$	
esu, Gau	$\text{Fr cm}^{-3}$	$= 10^7 \zeta^{-1} \text{C m}^{-3} \approx 3.335\,64 \times 10^{-4} \text{C m}^{-3}$
au	$e a_0^{-3}$	$\approx 1.081\,20 \times 10^{-12} \text{C m}^{-3}$
<i>Electric potential, <math>V, \phi</math></i>		
Volt (SI unit)	V	$= \text{J C}^{-1} = \text{J A}^{-1} \text{s}^{-1}$
esu, Gau	$\text{erg Fr}^{-1}$	$= \text{Fr cm}^{-1} / 4\pi\epsilon_0 = 299.792\,458 \text{ V}$
'cm <sup>-1</sup> g	$e \text{cm}^{-1} / 4\pi\epsilon_0$	$\approx 1.439\,97 \times 10^{-7} \text{ V}$
au	$e / 4\pi\epsilon_0 a_0$	$= E_{\text{H}} / e \approx 27.2114 \text{ V}$
Mean international volt		$= 1.000\,34 \text{ V}$
US international volt		$= 1.000\,330 \text{ V}$
<i>Electric resistance, <math>R</math></i>		
Ohm (SI unit)	$\Omega$	$= \text{V A}^{-1} = \text{m}^2 \text{kg s}^{-3} \text{A}^{-2}$
Mean international ohm		$= 1.100\,049 \Omega$
US international ohm		$= 1.000\,495 \Omega$
<i>Electric field, <math>E</math></i>		
SI unit	$\text{V m}^{-1}$	$= \text{J C}^{-1} \text{m}^{-1}$
esu, Gau	$\text{Fr cm}^{-2} / 4\pi\epsilon_0$	$= 2.997\,924\,58 \times 10^4 \text{ V m}^{-1}$
'cm <sup>-2</sup> g	$e \text{cm}^{-2} / 4\pi\epsilon_0$	$\approx 1.439\,97 \times 10^{-5} \text{ V m}^{-1}$
au	$e / 4\pi\epsilon_0 a_0^2$	$\approx 5.142\,21 \times 10^{11} \text{ V m}^{-1}$
<i>Electric field gradient, <math>E'_{\alpha\beta}, q_{\alpha\beta}</math></i>		
SI unit	$\text{V m}^{-2}$	$= \text{J C}^{-1} \text{m}^{-2}$
esu, Gau	$\text{Fr cm}^{-3} / 4\pi\epsilon_0$	$= 2.997\,924\,58 \times 10^6 \text{ V m}^{-2}$
'cm <sup>-3</sup> g	$e \text{cm}^{-3} / 4\pi\epsilon_0$	$\approx 1.439\,97 \times 10^{-3} \text{ V m}^{-2}$
au	$e / 4\pi\epsilon_0 a_0^3$	$\approx 9.717\,36 \times 10^{21} \text{ V m}^{-2}$
<i>Electric dipole moment, <math>p, \mu</math></i>		
SI unit	C m	
esu, Gau	$\text{Fr cm}$	$\approx 3.335\,64 \times 10^{-12} \text{ C m}$
Debye	D	$= 10^{-18} \text{Fr cm} \approx 3.335\,64 \times 10^{-30} \text{ C m}$
'cm', dipole length <sup>g</sup>	$e \text{cm}$	$\approx 1.602\,18 \times 10^{-21} \text{ C m}$
au	$e a_0$	$\approx 8.478\,36 \times 10^{-30} \text{ C m}$
<i>Electric quadrupole moment, <math>Q_{\alpha\beta}, \Theta_{\alpha\beta}, eQ</math></i>		
SI unit	$\text{C m}^2$	
esu, Gau	$\text{Fr cm}^2$	$\approx 3.335\,64 \times 10^{-14} \text{ C m}^2$
'cm <sup>2</sup> ', quadrupole area <sup>g</sup>	$e \text{cm}^2$	$\approx 1.602\,18 \times 10^{-23} \text{ C m}^2$
au	$e a_0^2$	$\approx 4.486\,55 \times 10^{-40} \text{ C m}^2$
<i>Polarizability, <math>\alpha</math></i>		
SI unit	$\text{J}^{-1} \text{C}^2 \text{m}^2$	$= \text{F m}^2$
esu, Gau, 'cm <sup>3</sup> ', polarizability volume <sup>g</sup>	$4\pi\epsilon_0 \text{cm}^3$	$\approx 1.112\,65 \times 10^{-16} \text{ J}^{-1} \text{C}^2 \text{m}^2$
'Å <sup>3</sup> g	$4\pi\epsilon_0 \text{Å}^3$	$\approx 1.112\,65 \times 10^{-40} \text{ J}^{-1} \text{C}^2 \text{m}^3$
au	$4\pi\epsilon_0 a_0^3$	$\approx 1.648\,78 \times 10^{-41} \text{ J}^{-1} \text{C}^2 \text{m}^2$
<i>Electric displacement, <math>D</math> (Volume) polarization, <math>P</math></i>		
SI unit	$\text{C m}^{-2}$	
esu, Gau	$\text{Fr cm}^{-2}$	$= (10^5/\zeta) \text{C m}^{-2} \approx 3.335\,64 \times 10^{-6} \text{ C m}^{-2}$
(But note: the use of the esu or Gaussian unit for electric displacement usually implies that the irrational displacement is being quoted, $D^{(\text{ir})} = 4\pi D$ .)		
<i>Magnetic flux density, <math>B</math> (magnetic field)</i>		
Tesla (SI unit)	T	$= \text{J A}^{-1} \text{m}^{-2} = \text{V s m}^{-2} = \text{Wb m}^{-2}$
Gauss (emu, Gau)	G	$= 10^{-4} \text{ T}$
au	$h / e a_0^2$	$\approx 2.350\,52 \times 10^5 \text{ T}$
<i>Magnetic flux, <math>\phi</math></i>		
Weber (SI unit)	Wb	$= \text{J A}^{-1} = \text{V s}$
Maxwell (emu, Gau)	Mx	$= \text{G cm}^{-2} = 10^{-8} \text{ Wb}$

Continued

Name	Symbol	Relation to SI
<i>Magnetic field, H (Volume) magnetization, M</i>		
SI unit	$\text{A m}^{-1}$	$= \text{C s}^{-1} \text{m}^{-1}$
Oersted (emu, Gau)	Oe	$= 10^3 \text{A m}^{-1}$
(But note: in practice the oersted, Oe, is only used as a unit for $H^{(\text{ir})} = 4\pi H$ ; thus when $H^{(\text{ir})} = 1 \text{Oe}$ , $H = (10^3/4\pi) \text{A m}^{-1}$ .)		
<i>Magnetic dipole moment, m, <math>\mu</math></i>		
SI unit	$\text{A m}^2$	$= \text{J T}^{-1}$
emu, Gau	$\text{erg G}^{-1}$	$= 10 \text{A cm}^2 = 10^{-3} \text{J T}^{-1}$
Bohr magneton <sup>h</sup>	$\mu_{\text{B}}$	$= e\hbar/2m_{\text{e}} \approx 9.274\,02 \times 10^{-24} \text{J T}^{-1}$
au	$e\hbar/m_{\text{e}}$	$= 2\mu_{\text{B}} \approx 1.854\,80 \times 10^{-23} \text{J T}^{-1}$
Nuclear magneton	$\mu_{\text{N}}$	$= (m_{\text{e}}/m_{\text{p}})\mu_{\text{B}} \approx 5.050\,79 \times 10^{-27} \text{J T}^{-1}$
<i>Magnetizability, <math>\zeta</math></i>		
SI unit	$\text{J T}^{-2}$	$= \text{C}^2 \text{m}^2 \text{kg}^{-1}$
au	$e^2 a_0^2/m_{\text{e}}$	$\approx 7.891\,04 \times 10^{-29} \text{J T}^{-2}$
<i>Magnetic susceptibility, <math>\chi</math>, <math>\kappa</math></i>		
SI unit	1	
emu, Gau	1	
(But note: in practice susceptibilities quoted in the context of emu or Gaussian units are always values for $\chi^{(\text{ir})} = \chi/4\pi$ ; thus when $\chi^{(\text{ir})} = 10^{-6}$ , $\chi = 4\pi \times 10^{-6}$ .)		
<i>Molar magnetic susceptibility, <math>\chi_{\text{m}}</math></i>		
SI unit	$\text{m}^3 \text{mol}^{-1}$	
emu, Gau	$\text{cm}^3 \text{mol}^{-1}$	$= 10^{-6} \text{m}^3 \text{mol}^{-1}$
(But note: in practice the units $\text{cm}^3 \text{mol}^{-1}$ usually imply that the irrational molar susceptibility is being quoted, $\chi_{\text{m}}^{(\text{ir})} = \chi_{\text{m}}/4\pi$ ; thus, for example, if $\chi_{\text{m}}^{(\text{ir})} = -15 \times 10^{-6} \text{cm}^3 \text{mol}^{-1}$ , which is often written as ‘ $-15 \text{cgs ppm}$ ’, then $\chi_{\text{m}} = -1.88 \times 10^{-10} \text{m}^3 \text{mol}^{-1}$ .)		

<sup>a</sup>Note that the day is not exactly defined in terms of the second since so-called leap-seconds are added or subtracted from the day semiannually in order to keep the annual average occurrence of midnight at 24:00 on the clock.

<sup>b</sup>The year is not commensurable with the date and not a constant. Prior to 1967, when the atomic standard was introduced, the tropical year 1900 served as the basis for the definition of the second. For the epoch 1900.0, it amounted to  $365.242\,198\,79 \text{d} \approx 31\,556\,925.975 \text{s}$  and it decreases by 0.530 seconds per century. The calendar years are exactly defined in terms of the day:

$$\text{Julian year} = 365.25 \text{ d}$$

$$\text{Gregorian year} = 365.2425 \text{ d}$$

The definition in the table corresponds to the Gregorian year. This is an average based on a year of length 365 days, with leap years of 366 days; leap years are taken *either* when the year is divisible by 4 but is not divisible by 100, *or* when the year is divisible by 400.

<sup>c</sup>1 N is approximately the force exerted by the earth upon an apple.

<sup>d</sup> $T/^{\circ}\text{R} = (9/5) T/\text{K}$ . Also, Celsius temperature  $\theta$  is related to thermodynamic temperature  $T$  by the equation

$$\theta/^{\circ}\text{C} = T/\text{K} - 273.15$$

Similarly Fahrenheit temperature  $\theta_{\text{F}}$  is related to Celsius temperature  $\theta$  by the equation

$$\theta_{\text{F}}/^{\circ}\text{F} = (9/5)(\theta/^{\circ}\text{C}) + 32$$

<sup>e</sup>The name ‘amagat’ is unfortunately used as a unit for both molar volume and amount density. Its value is slightly different for different gases, reflecting the deviation from ideal behavior for the gas being considered.

<sup>f</sup>The unit röntgen, employed to express exposure to X- or  $\gamma$ -radiations, is equal to:  $\text{R} = 2.58 \times 10^{-4} \text{C kg}^{-1}$ .

<sup>g</sup>The units in quotation marks for electric potential through polarizability may be found in the literature, although they are strictly incorrect; they should be replaced in each case by the units given in the symbol column. Thus, for example, when a quadrupole moment is quoted in ‘ $\text{cm}^2$ ’, the correct unit is  $e \text{cm}^2$ ; and when a polarizability is quoted in ‘ $\text{\AA}^3$ ’, the correct unit is  $4\pi\epsilon_0 \text{\AA}^3$ .

<sup>h</sup>The Bohr magneton  $\mu_{\text{B}}$  is sometimes denoted BM (or B.M.), but this is not recommended.

## APPENDIX 7. STATISTICAL TABLES

The following tables are presented in a format that is compatible with the needs of analytical chemists: the significance level  $P=0.05$  has been used in most cases, and it has been assumed that the number of measurements available is fairly small. Except where stated otherwise, these abbreviated tables have been reproduced from *Elementary Statistics Tables* by Henry R. Neave, published by George Allen & Unwin Ltd. (Tables 1–3 and 6–12). The reader requiring statistical data corresponding to significance levels and/or numbers of measurements not covered in the tables is referred to these sources.

**Table 1** The  $t$ -distribution

Value of $t$ for a confidence interval of:	90%	95%	98%	99%
Critical value of $ t $ for $P$ values of:	0.10	0.05	0.02	0.01
Number of degrees of freedom				
1	6.31	12.71	31.82	63.66
2	2.92	4.30	6.96	9.92
3	2.35	3.18	4.54	5.84
4	2.13	2.78	3.75	4.60
5	2.02	2.57	3.36	4.03
6	1.94	2.45	3.14	3.71
7	1.89	2.36	3.00	3.50
8	1.86	2.31	2.90	3.36
9	1.83	2.26	2.82	3.25
10	1.81	2.23	2.76	3.17
12	1.78	2.18	2.68	3.05
14	1.76	2.14	2.62	2.98
16	1.75	2.12	2.58	2.92
18	1.73	2.10	2.55	2.88
20	1.72	2.09	2.53	2.85
30	1.70	2.04	2.46	2.75
50	1.68	2.01	2.40	2.68
$\infty$	1.64	1.96	2.33	2.58

The critical values of  $|t|$  are appropriate for a *two-tailed* test. For a *one-tailed* test the value is taken from the column for *twice* the desired  $P$ -value, e.g., for a *one-tailed* test,  $P=0.05$ , 5 degrees of freedom, the critical value is read from the  $P=0.10$  column and is equal to 2.02.

**Table 2** Critical values of  $F$  for a *one-tailed* test ( $P=0.05$ )

$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	10	12	15	20
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45
3	10.13	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786	8.745	8.703	8.660
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.912	5.858	5.803
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.678	4.619	4.558
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.000	3.938	3.874
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.575	3.511	3.445
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347	3.284	3.218	3.150
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.073	3.006	2.936
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.913	2.845	2.774
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.788	2.719	2.646
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687	2.617	2.544
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.604	2.533	2.459
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.534	2.463	2.388
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.475	2.403	2.328
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.425	2.352	2.276
17	4.451	3.592	3.197	2.965	2.810	2.699	3.614	2.548	2.494	2.450	2.381	2.308	2.230
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.342	2.269	2.191
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.308	2.234	2.155
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.278	2.203	2.124

$\nu_1$  = number of degrees of freedom of the numerator and  $\nu_2$  = number of degrees of freedom of the denominator.

**Table 3** Critical values of  $F$  for a *two-tailed* test ( $P=0.05$ )

$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	10	12	15	20
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17
4	12.22	10.65	9.979	9.605	9.364	9.197	9.074	8.980	8.905	8.844	8.751	8.657	8.560
5	10.01	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619	6.525	6.428	6.329
6	8.813	7.260	6.599	6.227	5.988	5.820	5.695	5.600	5.523	5.461	5.366	5.269	5.168
7	8.073	6.542	5.890	5.523	5.285	5.119	4.995	4.899	4.823	4.761	4.666	4.568	4.467
8	7.571	6.059	5.416	5.053	4.817	4.652	4.529	4.433	4.357	4.295	4.200	4.101	3.999
9	7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026	3.964	3.868	3.769	3.667
10	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717	3.621	3.522	3.419
11	6.724	5.256	4.630	4.275	4.044	3.881	3.759	3.664	3.588	3.526	3.430	3.330	3.226
12	6.554	5.096	4.474	4.121	3.891	3.728	3.607	3.512	3.436	3.374	3.277	3.177	3.073
13	6.414	4.965	4.347	3.996	3.767	3.604	3.483	3.388	3.312	3.250	3.153	3.053	2.948
14	6.298	4.857	4.242	3.892	3.663	3.501	3.380	3.285	3.209	3.147	3.050	2.949	2.844
15	6.200	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123	3.060	2.963	2.862	2.756
16	6.115	4.687	4.077	3.729	3.502	3.341	3.219	3.125	3.049	2.986	2.889	2.788	2.681
17	6.042	4.619	4.011	3.665	3.438	3.277	3.156	3.061	2.985	2.922	2.825	2.723	2.616
18	5.978	4.560	3.954	3.608	3.382	3.221	3.100	3.005	2.929	2.866	2.769	2.667	2.559
19	5.922	4.508	3.903	3.559	3.333	3.172	3.051	2.956	2.880	2.817	2.720	2.617	2.509
20	5.871	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.837	2.774	2.676	2.573	2.464

$\nu_1$  = number of degrees of freedom of the numerator and  $\nu_2$  = number of degrees of freedom of the denominator.

**Table 4** Critical values of  $Q$  ( $P=0.05$ ) for a two-sided Dixon's test for outliers

Sample size	Critical value
4	0.831
5	0.717
6	0.621
7	0.570

Reproduced from King EP, Journal American Statistical Association.

**Table 5** Critical values of  $G$  ( $P=0.05$ ) for a two-sided Grubbs' test for outliers

Sample size	Critical value
3	1.155
4	1.481
5	1.715
6	1.887
7	2.020
8	2.126
9	2.215
10	2.290

Reproduced from Barnett V and Lewis T (1984) *Outliers in Statistical Data*, 2nd edn. John Wiley & Sons Ltd.

**Table 6** Critical values of  $\chi^2$  ( $P=0.05$ )

Number of degrees of freedom	Critical value
1	3.84
2	5.99
3	7.81
4	9.49
5	11.07
6	12.59
7	14.07
8	15.51
9	16.92
10	18.31

**Table 7** The sign test

$n$	$r=0$	1	2	3	4	5	6	7
4	0.063	0.313	0.688					
5	0.031	0.188	0.500					
6	0.016	0.109	0.344	0.656				
7	0.008	0.063	0.227	0.500				
8	0.004	0.035	0.144	0.363	0.637			
9	0.002	0.020	0.090	0.254	0.500			
10	0.001	0.011	0.055	0.172	0.377	0.623		
11	0.001	0.006	0.033	0.113	0.274	0.500		
12	0.000	0.003	0.019	0.073	0.194	0.387	0.613	
13	0.000	0.002	0.011	0.046	0.133	0.290	0.500	
14	0.000	0.001	0.006	0.029	0.090	0.212	0.395	0.605
15	0.000	0.000	0.004	0.018	0.059	0.151	0.304	0.500

The table uses the binomial distribution with  $P=0.5$  to give the probabilities of  $r$  or less successes for  $n=4-15$ . These values correspond to a *one*-tailed sign test and should be doubled for a *two*-tailed test.

**Table 8** Wilcoxon signed rank test. Critical values for the test statistic at  $P=0.05$ 

$n$	One-tailed test	Two-tailed test
5	0	NA
6	2	0
7	3	2
8	5	3
9	8	5
10	10	8
11	13	10
12	17	13
13	21	17
14	25	21
15	30	25

The null hypothesis can be rejected when the test statistic is  $\leq$  the tabulated value. NA indicates that the test cannot be applied.

**Table 9** Mann-Whitney  $U$ -test. Critical values for  $U$  or the lower of  $T_1$  and  $T_2$  at  $P=0.05$ 

$n_1$	$n_2$	One-tailed test	Two-tailed test
3	3	0	NA
3	4	0	NA
3	5	1	0
3	6	2	1
4	4	1	0
4	5	2	1
4	6	3	2
4	7	4	3
5	5	4	2
5	6	5	3
5	7	6	5
6	6	7	5
6	7	8	6
7	7	11	8

The null hypothesis can be rejected when  $U$  or the lower  $T$  value is  $\leq$  the tabulated value. NA indicates that the test cannot be applied.

**Table 10** The Spearman rank correlation coefficient. Critical values for  $\rho$  at  $P=0.05$ 

$n$	One-tailed test	Two-tailed test
5	0.900	1.000
6	0.829	0.886
7	0.714	0.786
8	0.643	0.738
9	0.600	0.700
10	0.564	0.649
11	0.536	0.618
12	0.504	0.587
13	0.483	0.560
14	0.464	0.538
15	0.446	0.521
16	0.429	0.503
17	0.414	0.488
18	0.401	0.472
19	0.391	0.460
20	0.380	0.447



**Table 11** The Kolmogorov goodness-of-fit test

<i>n</i>	<i>One-tailed test</i>	<i>Two-tailed test</i>
1	0.950	0.975
2	0.776	0.842
3	0.636	0.708
4	0.565	0.624
5	0.509	0.563
6	0.468	0.519
7	0.436	0.483
8	0.410	0.454
9	0.388	0.430
10	0.369	0.409
11	0.352	0.392
12	0.338	0.375
13	0.326	0.361
14	0.314	0.349
15	0.304	0.338
16	0.295	0.327
17	0.286	0.318
18	0.278	0.309
19	0.271	0.301
20	0.265	0.294

Critical values for *one*-tailed and *two*-tailed tests at  $P = 0.05$ . The appropriate value is compared with the maximum difference between the experimental and theoretical cumulative frequency curves.

**Table 12** The Kolmogorov test for normality

<i>n</i>	<i>One-tailed test</i>	<i>Two-tailed test</i>
3	0.367	0.376
4	0.345	0.375
5	0.319	0.343
6	0.297	0.323
7	0.280	0.304
8	0.265	0.288
9	0.252	0.274
10	0.241	0.262
11	0.231	0.251
12	0.222	0.242
13	0.215	0.234
14	0.208	0.226
15	0.201	0.219
16	0.195	0.213
17	0.190	0.207
18	0.185	0.202
19	0.181	0.197
20	0.176	0.192

Critical values for *one*-tailed and *two*-tailed tests at  $P = 0.05$ . The appropriate value is compared with the maximum difference between the experimental and theoretical cumulative frequency curves.

# APPENDIX 8. BIOLOGICAL BUFFERS

**Table 1** This table of frequently used buffers gives the  $pK_a$  value at 25°C and the useful pH range of each buffer. The buffers are listed in order of increasing pH

Acronym	Name	Mol. wt.	$pK_a$	Useful pH range
MES	2-( <i>N</i> -morpholino)ethanesulfonic acid	195.2	6.1	5.5–6.7
BIS TRIS	<i>Bis</i> (2-hydroxyethyl)iminotris(hydroxymethyl)methane	209.2	6.5	5.8–7.2
ADA	<i>N</i> -(2-acetamido)-2-iminodiacetic acid	190.2	6.6	6.0–7.2
ACES	2-[(2-Amino-2-oxoethyl)amino]ethanesulfonic acid	182.2	6.8	6.1–7.5
PIPES	Piperazine- <i>N,N'</i> - <i>bis</i> (2-ethanesulfonic acid)	302.4	6.8	6.1–7.5
MOPSO	3-( <i>N</i> -morpholino)-2-hydroxypropanesulfonic acid	225.3	6.9	6.2–7.6
BIS TRIS PROPANE	1,3- <i>Bis</i> [ <i>tris</i> (hydroxymethyl)methylamino]propane	282.3	6.8 <sup>a</sup>	6.3–9.5
BES	<i>N,N</i> - <i>bis</i> (2-hydroxyethyl)-2-aminoethanesulfonic acid	213.2	7.1	6.4–7.8
MOPS	3-( <i>N</i> -morpholino)propanesulfonic acid	209.3	7.2	6.5–7.9
HEPES	<i>N</i> -(2-hydroxyethyl)piperazine- <i>N'</i> -(2-ethanesulfonic acid)	238.3	7.5	6.8–8.2
TES	<i>N</i> - <i>tris</i> (hydroxymethyl)methyl-2-aminoethanesulfonic acid	229.2	7.5	6.8–8.2
DIPSO	3-[ <i>N,N</i> - <i>bis</i> (2-hydroxyethyl)amino]-2-hydroxypropanesulfonic acid	243.3	7.6	7.0–8.2
TAPSO	3-[ <i>N</i> - <i>tris</i> (hydroxymethyl)methylamino]-2-hydroxypropanesulfonic acid	259.3	7.6	7.0–8.2
TRIZMA	Tris(hydroxymethyl)aminomethane			
HEPPSO	<i>N</i> -2-hydroxyethylpiperazine- <i>N'</i> -(2-hydroxypropanesulfonic acid)	121.1	8.1	7.0–9.1
POPSO	Piperazine- <i>N,N'</i> - <i>bis</i> (2-hydroxypropanesulfonic acid)	268.3	7.8	7.1–8.5
EPPS	<i>N</i> -(2-hydroxyethyl)piperazine- <i>N'</i> -(3-propanesulfonic acid)	362.4	7.8	7.2–8.5
TEA	Triethanolamine	252.3	8.0	7.3–8.7
TRICINE	<i>N</i> - <i>tris</i> (hydroxymethyl)methylglycine	149.2	7.8	7.3–8.3
BICINE	<i>N,N</i> - <i>bis</i> (2-hydroxyethyl)glycine	179.2	8.1	7.4–8.8
TAPS	<i>N</i> - <i>tris</i> (hydroxymethyl)methyl-3-aminopropanesulfonic acid	163.2	8.3	7.6–9.0
AMPSO	3-[(1,1-Dimethyl-2-hydroxyethyl)amino]-2-hydroxypropanesulfonic acid	243.3	8.4	7.7–9.1
		227.3	9.0	8.3–9.7
CHES	2-( <i>N</i> -cyclohexylamino)ethanesulfonic acid	207.3	9.3	8.6–10.0
CAPSO	3-(Cyclohexylamino)-2-hydroxy- <i>L</i> -propanesulfonic acid	237.3	9.6	8.9–10.3
AMP	2-Amino-2-methyl- <i>L</i> -propanol	89.1	9.7	9.0–10.5
CAPS	3-(Cyclohexylamino)- <i>L</i> -propanesulfonic acid	221.3	10.4	9.7–11.1

<sup>a</sup> $pK_a$  = 9.0 for the second dissociation stage.

Reprinted from Sigma Chemical Company, St. Louis, MO.

# APPENDIX 9. pH SCALE FOR AQUEOUS SOLUTIONS

© 1993, CRC Press. Published by Elsevier Ltd. All rights reserved.

Primary ref. standard	Temperature (°C)															
	0	5	10	15	20	25	30	35	37	40	50	60	70	80	90	95
Saturated (at 25°C) potassium hydrogentartrate	—	—	—	—	—	3.557	3.552	3.549	3.548	3.547	3.549	3.560	3.580	3.610	3.650	3.674
0.1 mol kg <sup>−1</sup> Potassium dihydrogencitrate	3.863	3.840	3.820	3.802	3.788	3.776	3.766	3.759	3.756	3.754	3.749	—	—	—	—	—
0.025 mol kg <sup>−1</sup> Disodium hydrogenphosphate	6.984	6.951	6.923	6.900	6.881	6.865	6.853	6.844	6.841	6.838	6.833	6.836	6.845	6.859	6.876	6.886
+ 0.025 mol kg <sup>−1</sup> potassium dihydrogen phosphate	7.534	7.500	7.472	7.448	7.429	7.413	7.400	7.389	7.386	7.380	7.367	—	—	—	—	—
0.03043 mol kg <sup>−1</sup> Disodium hydrogenphosphate	9.464	9.395	9.332	9.276	9.225	9.180	9.139	9.102	9.088	9.068	9.011	8.962	8.921	8.884	8.850	8.833
+ 0.008695 mol kg <sup>−1</sup> potassium dihydrogenphosphate	10.317	10.245	10.179	10.118	10.062	10.012	9.966	9.926	9.910	9.889	9.828	—	—	—	—	—
0.01 mol kg <sup>−1</sup> Disodium tetraborate																
0.025 mol kg <sup>−1</sup> Sodium hydrogencarbonate																
+ 0.025 mol kg <sup>−1</sup> sodium carbonate																

Note: Based on an uncertainty of  $\pm 0.2$  mV in determined ( $E - E^0$ ), the uncertainty is  $\pm 0.003$  in pH in the range 0–50°C.

pH values for operational reference solutions

Operational standard ref. solution	Temperature (°C)															
	0	5	10	15	20	25	30	37	40	50	60	70	80	90	95	
0.1 mol kg <sup>-1</sup> Potassium tetroxalate <sup>a</sup>	–	–	–	–	1.475	1.479	1.483	1.490	1.493	1.503	1.513	1.52	1.53	1.53	1.53	
0.05 mol kg <sup>-1</sup> Potassium tetroxalate <sup>a</sup>	–	–	1.638	1.642	1.644	1.646	1.648	1.649	1.650	1.653	1.660	1.671	1.689	1.72	1.73	
0.05 mol kg <sup>-1</sup> Sodium hydrogendiglycolate <sup>b</sup>	–	3.466	3.470	3.476	3.484	3.492	3.502	3.519	3.527	3.558	3.595	–	–	–	–	
Saturated (at 25°C) Potassium hydrogentartrate	–	–	–	–	–	3.556	3.549	3.544	3.542	3.544	3.553	3.570	3.596	3.627	3.649	
0.05 mol kg <sup>-1</sup> Potassium hydrogentartrate (RVS)	4.000	3.998	3.997	3.998	4.000	4.005	4.011	4.022	4.027	4.050	4.080	4.115	4.159	4.21	4.24	
0.1 mol dm <sup>-3</sup> Acetic acid + 0.1 mol dm <sup>-3</sup> sodium acetate	4.664	4.657	4.652	4.647	4.645	4.644	4.643	4.647	4.650	4.663	4.684	4.713	4.75	4.80	4.83	
0.01 mol dm <sup>-3</sup> Acetic acid + 0.1 mol dm <sup>-3</sup> sodium acetate	4.729	4.722	4.717	4.714	4.712	4.713	4.715	4.722	4.726	4.743	4.768	4.800	4.839	4.88	4.91	
0.02 mol kg <sup>-1</sup> Piperazine phosphate <sup>c</sup>	–	6.477	6.419	6.364	6.310	6.259	6.209	6.143	6.116	6.030	5.952	–	–	–	–	
0.025 mol kg <sup>-1</sup> Disodium hydrogenphosphate + 0.025 mol kg <sup>-1</sup> potassium dihydrogenphosphate	6.961	6.935	6.912	6.891	6.873	6.857	6.843	6.828	6.823	6.814	6.817	6.830	6.85	6.90	6.92	
0.03043 mol kg <sup>-1</sup> Disodium hydrogenphosphate + 0.008695 mol kg <sup>-1</sup> potassium disodium phosphate	7.506	7.482	7.460	7.441	7.423	7.406	7.390	7.369	–	–	–	–	–	–	–	
0.04 mol kg <sup>-1</sup> Disodium hydrogenphosphate + 0.01 mol kg <sup>-1</sup> potassium dihydrogen phosphate	–	7.512	7.488	7.466	7.445	7.428	7.414	7.404	–	–	–	–	–	–	–	
0.05 mol kg <sup>-1</sup> Tris hydrochloride + 0.01667 mol kg <sup>-1</sup> Tris <sup>d</sup>	8.399	8.238	8.083	7.933	7.788	7.648	7.513	7.332	7.257	7.018	6.794	–	–	–	–	
0.05 mol kg <sup>-1</sup> Disodium tetraborate (Na <sub>2</sub> B <sub>4</sub> O <sub>7</sub> )	9.475	9.409	9.347	9.288	9.233	9.182	9.134	9.074	9.051	8.983	8.932	8.898	8.88	8.84	8.89	
0.01 mol kg <sup>-1</sup> Disodium tetraborate (Na <sub>2</sub> B <sub>4</sub> O <sub>7</sub> )	9.451	9.388	9.329	9.275	9.225	9.179	9.138	9.086	9.066	9.009	8.965	8.932	8.91	8.90	8.89	
0.025 mol kg <sup>-1</sup> Sodium hydrogencarbonate + 0.025 mol kg <sup>-1</sup> sodium carbonate	10.273	10.212	10.154	10.098	10.045	9.995	9.948	9.889	9.866	9.800	9.753	9.728	9.725	9.75	9.77	
Saturated (at 20°C) calcium hydroxide	13.360	13.159	12.965	12.780	12.602	12.431	12.267	12.049	11.959	11.678	11.423	11.192	10.984	10.80	10.71	

<sup>a</sup> Potassium trihydrogen dioxalate (KH<sub>3</sub>C<sub>4</sub>O<sub>8</sub>).

<sup>b</sup> Sodium hydrogen 2,2'-oxydiethanoate.

<sup>c</sup> C<sub>4</sub>H<sub>10</sub>N<sub>2</sub> · H<sub>3</sub>PO<sub>4</sub>.

<sup>d</sup> 2-Amino-2(hydroxymethyl)-1,3 propanediol or tris(hydroxymethyl)aminomethane.

Note: Uncertainty is ±0.003 in pH between 0°C and 60°C rising to ±0.01 above 70°C.

Useful data on some standard buffer solutions

	<i>Molecular formula</i>	<i>Molarity (mol kg<sup>-1</sup>)</i>	<i>Relative molar mass</i>	<i>Density at 20° C (g cm<sup>-3</sup>)</i>	<i>Molarity at 20° C (mol l<sup>-1</sup>)</i>	<i>Mass of 1 l at 20° C (g)</i>	<i>Mass tolerance for ± 0.001 pH<sup>a</sup> (g)</i>	<i>Mass tolerance expressed as a percentage (%)</i>
Potassium tetroxalate	KH <sub>3</sub> C <sub>4</sub> O <sub>8</sub> · 2H <sub>2</sub> O	0.1	254.1913	1.0091	0.09875	25.1017	0.07	0.27
Potassium tetroxalate	KH <sub>3</sub> C <sub>4</sub> O <sub>8</sub> · 2H <sub>2</sub> O	0.05	254.1913	1.0038	0.04965	12.6202	0.034	0.26
Disodium hydrogen orthophosphate	Na <sub>2</sub> HPO <sub>4</sub>	0.025	141.9588	1.0038	0.02492	3.5379	0.02	0.56
Potassium dihydrogen orthophosphate	KH <sub>2</sub> PO <sub>4</sub>	0.025	136.0852			3.3912	0.02	0.58
Disodium tetraborate	Na <sub>2</sub> B <sub>4</sub> O <sub>7</sub> · 10H <sub>2</sub> O	0.05	381.367	1.0075	0.04985	19.0117	0.9	4.73
Disodium tetraborate	Na <sub>2</sub> B <sub>4</sub> O <sub>7</sub> · 10H <sub>2</sub> O	0.01	381.367	1.0001	0.009981	3.8064	0.19	0.49
Sodium carbonate	Na <sub>2</sub> CO <sub>3</sub>	0.025	105.9887	1.0021	0.02494	2.6428	0.017	0.064
Sodium hydrogencarbonate	NaHCO <sub>3</sub>	0.025	84.0069			2.0947	0.013	0.62

<sup>a</sup> Calculated from known dilution value of solution."Appendix 9 pH Scale for Aqueous Solutions" is reproduced with permission from Lide DR (editor-in-chief) (1993) *CRC Handbook of Chemistry and Physics*, 72 edn., Boca Raton: CRC Press.

# APPENDIX 10. STANDARD POTENTIALS IN AQUEOUS SOLUTIONS

© 1985 Marcel Dekker. Published by Elsevier Ltd. All rights reserved.

Selected half-reaction potentials are listed in order of increasingly positive (anodic) assignments, providing a convenient guide to redox couples in a given range of standard potentials. The synopsis consists of two tables, one each for acid and basic solutions, which have been advisedly restricted to selected half-reactions.

		Continued	
<i>Couple</i>	<i>E° (V)</i>	<i>Couple</i>	<i>E° (V)</i>
<i>Standard potentials in acid solutions</i>			
$(3/2)\text{N}_2 + \text{H}^+ + \text{e}^- \rightarrow \text{HN}_3$	-3.10	$\text{H}_3\text{BO}_3 + 3\text{H}^+ + 3\text{e}^- \rightarrow \text{B} + 3\text{H}_2\text{O}$	-0.890
$\text{Li}^+ + \text{e}^- \rightarrow \text{Li}$	-3.045	$\text{SiO}_2(\text{vit}) + 4\text{H}^+ + 4\text{e}^- \rightarrow \text{Si} + 2\text{H}_2\text{O}$	-0.888
$\text{K}^+ + \text{e}^- \rightarrow \text{K}$	-2.925	$\text{TiO}^{2+} + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{Ti} + \text{H}_2\text{O}$	-0.882
$\text{Rb}^+ + \text{e}^- \rightarrow \text{Rb}$	-2.925	$\text{Ta}_2\text{O}_5 + 10\text{H}^+ + 10\text{e}^- \rightarrow 2\text{Ta} + 5\text{H}_2\text{O}$	-0.81
$\text{Cs}^+ + \text{e}^- \rightarrow \text{Cs}$	-2.923	$\text{Zn}^{2+} + 2\text{e}^- \rightarrow \text{Zn}$	-0.7626
$\text{Ba}^{2+} + 2\text{e}^- \rightarrow \text{Ba}$	-2.92	$\text{TlI} + \text{e}^- \rightarrow \text{TI} + \text{I}^-$	-0.74
$\text{Ra}^{2+} + 2\text{e}^- \rightarrow \text{Ra}$	-2.916	$\text{Te} + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{H}_2\text{Te}$	-0.740
$\text{Sr}^{2+} + 2\text{e}^- \rightarrow \text{Sr}$	-2.89	$\text{TiBr} + \text{e}^- \rightarrow \text{TI} + \text{Br}^-$	-0.658
$\text{Ca}^{2+} + 2\text{e}^- \rightarrow \text{Ca}$	-2.84	$\text{Nb}_2\text{O}_5 + 10\text{H}^+ + 10\text{e}^- \rightarrow 2\text{Nb} + 5\text{H}_2\text{O}$	-0.65
$\text{Na}^+ + \text{e}^- \rightarrow \text{Na}$	-2.714	$\text{TiCl} + \text{e}^- \rightarrow \text{TI} + \text{Cl}^-$	-0.5568
$\text{No}^{2+} + \text{e}^- \rightarrow \text{No}$	-2.5	$\text{Ga}^{3+} + 3\text{e}^- \rightarrow \text{Ga}$	-0.529
$\text{Md}^{2+} + 2\text{e}^- \rightarrow \text{Md}$	-2.4	$\text{U}^{4+} + \text{e}^- \rightarrow \text{U}^{3+}$	-0.52
$\text{Fm}^{2+} + 2\text{e}^- \rightarrow \text{Fm}$	-2.37	$\text{Sb} + 3\text{H}^+ + 3\text{e}^- \rightarrow \text{SbH}_3(\text{g})$	-0.510
$\text{La}^{3+} + 3\text{e}^- \rightarrow \text{La}$	-2.37	$\text{H}_3\text{PO}_2 + \text{H}^+ + \text{e}^- \rightarrow \text{P}(\text{w}) + 2\text{H}_2\text{O}$	-0.508
$\text{Y}^{3+} + 3\text{e}^- \rightarrow \text{Y}$	-2.37	$\text{H}_3\text{PO}_3 + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{H}_3\text{PO}_2 + \text{H}_2\text{O}$	-0.499
$\text{Ce}^{3+} + 3\text{e}^- \rightarrow \text{Ce}$	-2.34	$\text{Fe}^{2+} + 2\text{e}^- \rightarrow \text{Fe}$	-0.44
$\text{Nd}^{3+} + 3\text{e}^- \rightarrow \text{Nd}$	-2.32	$\text{Cr}^{3+} + \text{e}^- \rightarrow \text{Cr}^{2+}$	-0.424
$\text{Sm}^{3+} + 3\text{e}^- \rightarrow \text{Sm}$	-2.30	$\text{Cd}^{2+} + 2\text{e}^- \rightarrow \text{Cd}$	-0.4025
$\text{Gd}^{3+} + 3\text{e}^- \rightarrow \text{Gd}$	-2.29	$\text{Ti}^{3+} + \text{e}^- \rightarrow \text{Ti}^{2+}$	-0.37
$\text{Mg}^{2+} + 2\text{e}^- \rightarrow \text{Mg}$	-2.356	$\text{PbI}_2 + 2\text{e}^- \rightarrow \text{Pb} + 2\text{I}^-$	-0.365
$\text{Lu}^{3+} + 3\text{e}^- \rightarrow \text{Lu}$	-2.30	$\text{PbSO}_4 + 2\text{e}^- \rightarrow \text{Pb} + \text{SO}_4^{2-}$	-0.3505
$1/2\text{H}_2 + \text{e}^- \rightarrow \text{H}^-$	-2.25	$\text{Eu}^{3+} + \text{e}^- \rightarrow \text{Eu}^{2+}$	-0.35
$\text{Cf}^{2+} + 2\text{e}^- \rightarrow \text{Cf}$	-2.2	$\text{In}^{3+} + 3\text{e}^- \rightarrow \text{In}$	-0.3382
$\text{Es}^{2+} + 2\text{e}^- \rightarrow \text{Es}$	-2.2	$\text{TI}^+ + \text{e}^- \rightarrow \text{TI}$	-0.3363
$\text{Am}^{3+} + 3\text{e}^- \rightarrow \text{Am}$	-2.07	$\text{PbBr}_2 + 2\text{e}^- \rightarrow \text{Pb} + 2\text{Br}^-$	-0.280
$\text{AlF}_6^{3-} + 3\text{e}^- \rightarrow \text{Al} + 6\text{F}^-$	-2.067	$\text{Co}^{2+} + 2\text{e}^- \rightarrow \text{Co}$	-0.277
$\text{Cm}^{3+} + 3\text{e}^- \rightarrow \text{Cm}$	-2.06	$\text{H}_3\text{PO}_4 + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{H}_3\text{PO}_3 + \text{H}_2\text{O}$	-0.276
$\text{Sc}^{3+} + 3\text{e}^- \rightarrow \text{Sc}$	-2.03	$\text{PbCl}_2 + 2\text{e}^- \rightarrow \text{Pb} + 2\text{Cl}^-$	-0.268
$\text{Bk}^{3+} + 3\text{e}^- \rightarrow \text{Bk}$	-2.01	$\text{Ni}^{2+} + 2\text{e}^- \rightarrow \text{Ni}$	-0.257
$\text{Cf}^{3+} + 3\text{e}^- \rightarrow \text{Cf}$	-2.0	$\text{V}^{3+} + \text{e}^- \rightarrow \text{V}^{2+}$	-0.255
$\text{Es}^{3+} + 3\text{e}^- \rightarrow \text{Es}$	-2.0	$2\text{SO}_4^{2-} + 4\text{H}^+ + 4\text{e}^- \rightarrow \text{S}_2\text{O}_6^{2-} + 2\text{H}_2\text{O}$	-0.253
$\text{Be}^{2+} + 2\text{e}^- \rightarrow \text{Be}$	-1.97	$\text{SnF}_6^{2-} + 4\text{e}^- \rightarrow \text{Sn} + 6\text{F}^-$	-0.25
$\text{Fm}^{3+} + 3\text{e}^- \rightarrow \text{Fm}$	-1.96	$\text{N}_2 + 5\text{H}^+ + 4\text{e}^- \rightarrow \text{N}_2\text{H}_5^+$	-0.23
$\text{Th}^{4+} + 4\text{e}^- \rightarrow \text{Th}$	-1.83	$\text{As} + 3\text{H}^+ + 3\text{e}^- \rightarrow \text{AsH}_3$	-0.225
$\text{Np}^{3+} + 3\text{e}^- \rightarrow \text{Np}$	-1.79	$\text{Mo}^{3+} + 3\text{e}^- \rightarrow \text{Mo}$	-0.2
$\text{Md}^{3+} + 3\text{e}^- \rightarrow \text{Md}$	-1.7	$\text{CuI} + \text{e}^- \rightarrow \text{Cu} + \text{I}^-$	-0.182
$\text{Zr}^{4+} + 4\text{e}^- \rightarrow \text{Zr}$	-1.70	$\text{CO}_2 + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{HCOOH}(\text{aq})$	-0.16
$\text{Al}^{3+} + 3\text{e}^- \rightarrow \text{Al}$	-1.67	$\text{AgI} + \text{e}^- \rightarrow \text{Ag} + \text{I}^-$	-0.1522
$\text{U}^{3+} + 3\text{e}^- \rightarrow \text{U}$	-1.66	$\text{Si} + 4\text{H}^+ + 4\text{e}^- \rightarrow \text{SiH}_4$	-0.143
$\text{Ti}^{2+} + 2\text{e}^- \rightarrow \text{Ti}$	-1.63	$\text{Sn}^{2+} + 2\text{e}^- \rightarrow \text{Sn}$	-0.136
$\text{Hf}^{4+} + 4\text{e}^- \rightarrow \text{Hf}$	-1.56	$\text{Pb}^{2+} + 2\text{e}^- \rightarrow \text{Pb}$	-0.1251
$\text{No}^{3+} + 3\text{e}^- \rightarrow \text{No}$	-1.2	$\text{WO}_3(\text{c}) + 6\text{H}^+ + 6\text{e}^- \rightarrow \text{W} + 3\text{H}_2\text{O}$	-0.090
$\text{SiF}_6^{2-} + 4\text{e}^- \rightarrow \text{Si} + 6\text{F}^-$	-1.2	$\text{P}(\text{w}) + 3\text{H}^+ + 3\text{e}^- \rightarrow \text{PH}_3$	-0.063
$\text{TiF}_6^{2-} + 4\text{e}^- \rightarrow \text{Ti} + 6\text{F}^-$	-1.191	$\text{O}_2 + \text{H}^+ + \text{e}^- \rightarrow \text{HO}_2$	-0.046
$\text{Mn}^{2+} + 2\text{e}^- \rightarrow \text{Mn}$	-1.18	$\text{Hg}_2\text{I}_2 + 2\text{e}^- \rightarrow 2\text{Hg} + 2\text{I}^-$	-0.0405
$\text{V}^{2+} + 2\text{e}^- \rightarrow \text{V}$	-1.13	$\text{Se} + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{H}_2\text{Se}$	-0.028
$\text{Nb}^{3+} + 3\text{e}^- \rightarrow \text{Nb}$	-1.1	$\text{GeO}_2 + 4\text{H}^+ + 4\text{e}^- \rightarrow \text{Ge}(\text{hex}) + 2\text{H}_2\text{O}$	-0.009
		$2\text{H}^+ + 2\text{e}^- \rightarrow \text{H}_2$	0.000



Continued

<i>Couple</i>	<i>E° (V)</i>
$\text{CuBr} + \text{e}^- \rightarrow \text{Cu} + \text{Br}^-$	0.033
$\text{HCOOH}(\text{aq}) + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{HCHO}(\text{aq}) + \text{H}_2\text{O}$	0.056
$\text{AgBr} + \text{e}^- \rightarrow \text{Ag} + \text{Br}^-$	0.0711
$\text{TiO}^{2+} + 2\text{H}^+ + \text{e}^- \rightarrow \text{Ti}^{3+} + \text{H}_2\text{O}$	0.100
$\text{CuCl} + \text{e}^- \rightarrow \text{Cu} + \text{Cl}^-$	0.121
$\text{C} + 4\text{H}^+ + 4\text{e}^- \rightarrow \text{CH}_4$	0.132
$\text{Hg}_2\text{Br}_2 + 2\text{e}^- \rightarrow 2\text{Hg} + 2\text{Br}^-$	0.139 20
$\text{S} + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{H}_2\text{S}$	0.144
$\text{Np}^{4+} + \text{e}^- \rightarrow \text{Np}^{3+}$	0.15
$\text{Sn}^{4+} + 2\text{e}^- \rightarrow \text{Sn}^{2+}$	0.15
$\text{Sb}_4\text{O}_6 + 12\text{H}^+ + 12\text{e}^- \rightarrow 4\text{Sb} + 6\text{H}_2\text{O}$	0.1504
$\text{SO}_4^{2-} + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{H}_2\text{SO}_3 + \text{H}_2\text{O}$	0.158
$\text{Cu}^{2+} + \text{e}^- \rightarrow \text{Cu}^+$	0.159
$\text{UO}_2^{2+} + \text{e}^- \rightarrow \text{UO}_2^+$	0.16
$\text{BiOCl} + 2\text{H}^+ + 3\text{e}^- \rightarrow \text{Bi} + \text{H}_2\text{O} + \text{Cl}^-$	0.1697
$2\text{H}_2\text{SO}_3 + 3\text{H}^+ + 2\text{e}^- \rightarrow \text{HS}_2\text{O}_4^- + 2\text{H}_2\text{O}$	0.173
$\text{ReO}_2 + 4\text{H}^+ + 4\text{e}^- \rightarrow \text{Re} + 2\text{H}_2\text{O}$	0.22
$\text{AgCl} + \text{e}^- \rightarrow \text{Ag} + \text{Cl}^-$	0.2223
$\text{HCHO}(\text{aq}) + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{CH}_3\text{OH}(\text{aq})$	0.232
$(\text{CH}_3)_2\text{SO}_2 + 2\text{H}^+ + 2\text{e}^- \rightarrow (\text{CH}_3)_2\text{SO} + 2\text{H}_2\text{O}$	0.238
$\text{HAsO}_2(\text{aq}) + 3\text{H}^+ + 3\text{e}^- \rightarrow \text{As} + 2\text{H}_2\text{O}$	0.248
$\text{UO}_2^{2+} + 4\text{H}^+ + 2\text{e}^- \rightarrow \text{U}^{4+} + 2\text{H}_2\text{O}$	0.27
$\text{HCNO} + \text{H}^+ + \text{e}^- \rightarrow (1/2)\text{C}_2\text{N}_2 + \text{H}_2\text{O}$	0.330
$\text{VO}^{2+} + 2\text{H}^+ + \text{e}^- \rightarrow \text{V}^{3+} + \text{H}_2\text{O}$	0.337
$\text{ReO}_4^- + 8\text{H}^+ + 7\text{e}^- \rightarrow \text{Re} + 4\text{H}_2\text{O}$	0.34
$\text{Cu}^{2+} + 2\text{e}^- \rightarrow \text{Cu}$	0.340
$\text{AgIO}_3 + \text{e}^- \rightarrow \text{Ag} + \text{IO}_3^-$	0.354
$\text{Fe}(\text{CN})_6^{3-} + \text{e}^- \rightarrow \text{Fe}(\text{CN})_6^{4-}$	0.3610
$\text{C}_2\text{N}_2 + 2\text{H}^+ + 2\text{e}^- \rightarrow 2\text{HCN}(\text{aq})$	0.373
$\text{UO}_2^{2+} + 4\text{H}^+ + \text{e}^- \rightarrow \text{U}^{4+} + 2\text{H}_2\text{O}$	0.38
$\text{H}_2\text{N}_2\text{O}_2 + 6\text{H}^+ + 4\text{e}^- \rightarrow 2\text{NH}_3\text{OH}^+$	0.387
$2\text{H}_2\text{SO}_3 + 2\text{H}^+ + 4\text{e}^- \rightarrow \text{S}_2\text{O}_3^{2-} + 3\text{H}_2\text{O}$	0.400
$\text{Ag}_2\text{CrO}_4 + 2\text{e}^- \rightarrow 2\text{Ag} + \text{CrO}_4^{2-}$	0.4491
$\text{Ag}_2\text{MoO}_4 + 2\text{e}^- \rightarrow 2\text{Ag} + \text{MoO}_4^{2-}$	0.486
$\text{PdBr}_4^{2-} + 2\text{e}^- \rightarrow \text{Pd} + 4\text{Br}^-$	0.49
$\text{RhCl}_6 + 3\text{e}^- \rightarrow \text{Rh} + 6\text{Cl}^-$	0.5
$\text{H}_2\text{SO}_3 + 4\text{H}^+ + 4\text{e}^- \rightarrow \text{S} + 3\text{H}_2\text{O}$	0.500
$2\text{H}_2\text{SO}_3 + 4\text{H}^+ + 6\text{e}^- \rightarrow \text{S}_4\text{O}_6^{2-} + 6\text{H}_2\text{O}$	0.507
$\text{ReO}_4^- + 4\text{H}^+ + 3\text{e}^- \rightarrow \text{ReO}_2 + 2\text{H}_2\text{O}$	0.51
$\text{Cu}^+ + \text{e}^- \rightarrow \text{Cu}$	0.520
$\text{TeO}_2(\text{c}) + 4\text{H}^+ + 4\text{e}^- \rightarrow \text{Te} + 2\text{H}_2\text{O}$	0.53
$\text{I}_2 + 2\text{e}^- \rightarrow 2\text{I}^-$	0.5355
$\text{I}_3^- + 2\text{e}^- \rightarrow 3\text{I}^-$	0.536
$\text{AgBrO}_3 + \text{e}^- \rightarrow \text{Ag} + \text{BrO}_3^-$	0.546
$\text{Cu}^{2+} + \text{Cl}^- + \text{e}^- \rightarrow \text{CuCl}$	0.559
$\text{TeOOH}^+ + 3\text{H}^+ + 4\text{e}^- \rightarrow \text{Te} + 2\text{H}_2\text{O}$	0.559
$\text{H}_3\text{AsO}_4 + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{HAsO}_2 + 2\text{H}_2\text{O}$	0.560
$\text{MnO}_4^- + \text{e}^- \rightarrow \text{MnO}_4^{2-}$	0.56
$\text{S}_2\text{O}_6^{2-} + 4\text{H}^+ + 2\text{e}^- \rightarrow 2\text{H}_2\text{SO}_3$	0.569
$\text{CH}_3\text{OH}(\text{aq}) + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{CH}_4 + \text{H}_2\text{O}$	0.59
$\text{Sb}_2\text{O}_5 + 6\text{H}^+ + 4\text{e}^- \rightarrow 2\text{SbO}^+ + 3\text{H}_2\text{O}$	0.605
$\text{Au}(\text{SCN})_4^- + 3\text{e}^- \rightarrow \text{Au} + 4\text{SCN}^-$	0.636
$\text{PdCl}_4^{2-} + 2\text{e}^- \rightarrow \text{Pd} + 4\text{Cl}^-$	0.64
$\text{AgC}_2\text{H}_3\text{O}_2 + \text{e}^- \rightarrow \text{Ag} + \text{C}_2\text{H}_3\text{O}_2^-$	0.643
$\text{Cu}^{2+} + \text{Br}^- + \text{e}^- \rightarrow \text{CuBr}$	0.654
$\text{Ag}_2\text{SO}_4 + 2\text{e}^- \rightarrow 2\text{Ag} + \text{SO}_4^{2-}$	0.654
$\text{NpO}_2^{2+} + 4\text{H}^+ + \text{e}^- \rightarrow \text{Np}^{4+} + 2\text{H}_2\text{O}$	0.66
$\text{O}_2 + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{H}_2\text{O}_2$	0.695
$\text{HN}_3 + 11\text{H}^+ + 8\text{e}^- \rightarrow 3\text{NH}_4^+$	0.695
$\text{PtBr}_4^{2-} + 2\text{e}^- \rightarrow \text{Pt} + 4\text{Br}^-$	0.698
$2\text{NO} + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{H}_2\text{N}_2\text{O}_2$	0.71
$\text{H}_2\text{SeO}_3 + 4\text{H}^+ + 4\text{e}^- \rightarrow \text{Se} + 3\text{H}_2\text{O}$	0.739
$\text{PtCl}_4^{2-} + 2\text{e}^- \rightarrow \text{Pt} + 4\text{Cl}^-$	0.758

Continued

<i>Couple</i>	<i>E° (V)</i>
$\text{Rh}^{3+} + 3\text{e}^- \rightarrow \text{Rh}$	0.76
$(\text{SCN})_2 + 2\text{e}^- \rightarrow 2\text{SCN}^-$	0.77
$\text{Fe}^{3+} + \text{e}^- \rightarrow \text{Fe}^{2+}$	0.771
$\text{Hg}_2^{2+} + 2\text{e}^- \rightarrow 2\text{Hg}$	0.7960
$\text{Ag}^+ + \text{e}^- \rightarrow \text{Ag}$	0.7991
$2\text{NO}_3^- + 4\text{H}^+ + 2\text{e}^- \rightarrow \text{N}_2\text{O}_4 + 2\text{H}_2\text{O}$	0.803
$\text{IrBr}_6^{2-} + \text{e}^- \rightarrow \text{IrBr}_6^{3-}$	0.805
$\text{AmO}_2^{2+} + 4\text{H}^+ + \text{e}^- \rightarrow \text{Am}^{4+} + 2\text{H}_2\text{O}$	0.82
$\text{OsO}_4(\text{c}) + 8\text{H}^+ + 8\text{e}^- \rightarrow \text{Os} + 4\text{H}_2\text{O}$	0.84
$\text{AuBr}_4^- + 3\text{e}^- \rightarrow \text{Au} + 4\text{Br}^-$	0.854
$2\text{HNO}_2 + 4\text{H}^+ + 4\text{e}^- \rightarrow \text{H}_2\text{N}_2\text{O}_2$	0.86
$\text{IrCl}_6^{3-} + 3\text{e}^- \rightarrow \text{Ir} + 6\text{Cl}^-$	0.86
$\text{Cu}^{2+} + \text{I}^- + \text{e}^- \rightarrow \text{CuI}$	0.861
$\text{IrCl}_6^{2-} + \text{e}^- \rightarrow \text{IrCl}_6^{3-}$	0.867
$2\text{Hg}_2^{2+} + 2\text{e}^- \rightarrow \text{Hg}$	0.9110
$\text{Pd}^{2+} + 2\text{e}^- \rightarrow \text{Pd}$	0.915
$\text{NO}_3^- + 3\text{H}^+ + 2\text{e}^- \rightarrow \text{HNO}_2 + \text{H}_2\text{O}$	0.94
$\text{NO}_3^- + 4\text{H}^+ + 3\text{e}^- \rightarrow \text{NO} + 2\text{H}_2\text{O}$	0.957
$\text{AuBr}_2^- + \text{e}^- \rightarrow \text{Au} + 2\text{Br}^-$	0.960
$\text{PtO} + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{Pt} + \text{H}_2\text{O}$	0.980
$\text{HNO}_2 + \text{H}^+ + \text{e}^- \rightarrow \text{NO} + \text{H}_2\text{O}$	0.996
$\text{AuCl}_4^- + 3\text{e}^- \rightarrow \text{Au} + 4\text{Cl}^-$	1.002
$\text{Pu}^{4+} + \text{e}^- \rightarrow \text{Pu}^{3+}$	1.01
$\text{PuO}_2^{2+} + \text{e}^- \rightarrow \text{PuO}_2^+$	1.02
$\text{PuO}_2^{2+} + 4\text{H}^+ + 2\text{e}^- \rightarrow \text{Pu}^{4+} + 2\text{H}_2\text{O}$	1.03
$\text{N}_2\text{O}_4 + 4\text{H}^+ + 4\text{e}^- \rightarrow \text{NO} + 2\text{H}_2\text{O}$	1.039
$\text{PuO}_2^{2+} + 4\text{H}^+ + \text{e}^- \rightarrow \text{Pu}^{4+} + 2\text{H}_2\text{O}$	1.04
$\text{Sb}_2\text{O}_5 + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{Sb}_2\text{O}_4 + \text{H}_2\text{O}$	1.055
$\text{Br}_2(\text{l}) + 2\text{e}^- \rightarrow 2\text{Br}^-$	1.065
$\text{ICl}_2^- + \text{e}^- \rightarrow 2\text{Cl}^- + 1/2\text{I}_2$	1.07
$\text{N}_2\text{O}_4 + 2\text{H}^+ + 2\text{e}^- \rightarrow 2\text{HNO}_3$	1.07
$\text{Cu}^{2+} + 2\text{CN}^- + \text{e}^- \rightarrow \text{Cu}(\text{CN})_2^-$	1.12
$\text{H}_2\text{O}_2 + \text{H}^+ + \text{e}^- \rightarrow \text{OH} + \text{H}_2\text{O}$	1.14
$\text{SeO}_4^{2-} + 4\text{H}^+ + 2\text{e}^- \rightarrow \text{H}_2\text{SeO}_3 + \text{H}_2\text{O}$	1.151
$\text{ClO}_3^- + 3\text{H}^+ + 2\text{e}^- \rightarrow \text{HClO}_2 + \text{H}_2\text{O}$	1.181
$\text{ClO}_2 + \text{H}^+ + \text{e}^- \rightarrow \text{HClO}_2$	1.188
$\text{S}_2\text{Cl}_2 + 2\text{e}^- \rightarrow 2\text{S} + 2\text{Cl}^-$	1.19
$\text{IO}_3^- + 6\text{H}^+ + 5\text{e}^- \rightarrow 1/2\text{I}_2 + 3\text{H}_2\text{O}$	1.195
$\text{ClO}_4^- + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{ClO}_3^- + \text{H}_2\text{O}$	1.201
$\text{O}_2 + 4\text{H}^+ + 4\text{e}^- \rightarrow 2\text{H}_2\text{O}$	1.229
$\text{MnO}_2 + 4\text{H}^+ + 2\text{e}^- \rightarrow \text{Mn}^{2+} + 2\text{H}_2\text{O}$	1.23
$\text{NpO}_2^{2+} + \text{e}^- \rightarrow \text{NpO}_2^+$	1.24
$\text{N}_2\text{H}_5^+ + 3\text{H}^+ + 2\text{e}^- \rightarrow 2\text{NH}_4^+$	1.275
$\text{PdCl}_6^{2-} + 2\text{e}^- \rightarrow \text{PdCl}_4^{2-} + 2\text{Cl}^-$	1.288
$2\text{HNO}_2 + 4\text{H}^+ + 4\text{e}^- \rightarrow \text{N}_2\text{O} + 3\text{H}_2\text{O}$	1.297
$\text{NH}_3\text{OH}^+ + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{NH}_4^+ + \text{H}_2\text{O}$	1.35
$\text{Cl}_2 + 2\text{e}^- \rightarrow 2\text{Cl}^-$	1.3583
$\text{Cr}_2\text{O}_7^{2-} + 14\text{H}^+ + 6\text{e}^- \rightarrow 2\text{Cr}^{3+} + 7\text{H}_2\text{O}$	1.36
$2\text{NH}_3\text{OH}^+ + \text{H}^+ + 2\text{e}^- \rightarrow \text{N}_2\text{H}_5^+ + 2\text{H}_2\text{O}$	1.41
$\text{HO}_2 + \text{H}^+ + \text{e}^- \rightarrow \text{H}_2\text{O}_2$	1.44
$\text{PbO}_2(\text{x}) + 4\text{H}^+ + 2\text{e}^- \rightarrow \text{Pb}^{2+} + 2\text{H}_2\text{O}$	1.468
$\text{BrO}_3^- + 6\text{H}^+ + 5\text{e}^- \rightarrow 1/2\text{Br}_2 + 3\text{H}_2\text{O}$	1.478
$\text{Mn}^{3+} + \text{e}^- \rightarrow \text{Mn}^{2+}$	1.5
$\text{MnO}_4^- + 8\text{H}^+ + 5\text{e}^- \rightarrow \text{Mn}^{2+} + 4\text{H}_2\text{O}$	1.51
$\text{Au}^{3+} + 3\text{e}^- \rightarrow \text{Au}$	1.52
$\text{AmO}_2^{2+} + \text{e}^- \rightarrow \text{AmO}_2^+$	1.59
$\text{NiO}_2 + 4\text{H}^+ + 2\text{e}^- \rightarrow \text{Ni}^{2+} + 2\text{H}_2\text{O}$	1.593
$\text{H}_5\text{IO}_6 + \text{H}^+ + 2\text{e}^- \rightarrow \text{IO}_3^- + 3\text{H}_2\text{O}$	1.603
$\text{HBrO} + \text{H}^+ + \text{e}^- \rightarrow 1/2\text{Br}_2 + \text{H}_2\text{O}$	1.604
$\text{HClO} + \text{H}^+ + \text{e}^- \rightarrow 1/2\text{Cl}_2 + \text{H}_2\text{O}$	1.630
$\text{Bk}^{4+} + \text{e}^- \rightarrow \text{Bk}^{3+}$	1.67
$\text{AmO}_2^{2+} + 4\text{H}^+ + 3\text{e}^- \rightarrow \text{Am}^{3+} + \text{H}_2\text{O}$	1.67
$\text{HClO}_2 + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{HClO} + \text{H}_2\text{O}$	1.674

Continued

Continued

<i>Couple</i>	<i>E° (V)</i>
$\text{PbO}_2(\alpha) + \text{SO}_4^{2-} + 4\text{H}^+ + 2\text{e}^- \rightarrow \text{PbSO}_4 + 2\text{H}_2\text{O}$	1.698
$\text{MnO}_4^- + 4\text{H}^+ + 3\text{e}^- \rightarrow \text{MnO}_2 + 2\text{H}_2\text{O}$	1.70
$\text{Ce}^{4+} + \text{e}^- \rightarrow \text{Ce}^{3+}$	1.72
$\text{AmO}_2^+ + 4\text{H}^+ + 2\text{e}^- \rightarrow \text{Am}^{3+} + 2\text{H}_2\text{O}$	1.72
$\text{H}_2\text{O}_2 + 2\text{H}^+ + 2\text{e}^- \rightarrow 2\text{H}_2\text{O}$	1.763
$\text{Au}^+ + \text{e}^- \rightarrow \text{Au}$	1.83
$\text{Co}^{3+} + \text{e}^- \rightarrow \text{Co}^{2+}$	1.92
$\text{HN}_3 + 3\text{H}^+ + 2\text{e}^- \rightarrow \text{NH}_4^+ + \text{N}_2$	1.96
$\text{S}_2\text{O}_8^{2-} + 2\text{e}^- \rightarrow 2\text{SO}_4^{2-}$	1.96
$\text{Ag}^{2+} + \text{e}^- \rightarrow \text{Ag}^+$	1.980
$\text{O}_3 + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{O}_2 + \text{H}_2\text{O}$	2.075
$\text{F}_2\text{O} + 2\text{H}^+ + 4\text{e}^- \rightarrow 2\text{F}^- + \text{H}_2\text{O}$	2.153
$\text{OH} + \text{H}^+ + \text{e}^- \rightarrow \text{H}_2\text{O}$	2.38
$\text{O}(\text{g}) + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{H}_2\text{O}$	2.430
$\text{Am}^{4+} + \text{e}^- \rightarrow \text{Am}^{3+}$	2.62
$\text{F}_2 + 2\text{e}^- \rightarrow 2\text{F}^-$	2.87
$\text{F}_2 + 2\text{H}^+ + 2\text{e}^- \rightarrow 2\text{HF}(\text{aq})$	3.053

*Standard potentials in basic solutions*

$\text{Ca}(\text{OH})_2 + 2\text{e}^- \rightarrow \text{Ca} + 2\text{OH}^-$	-3.026
$\text{Ba}(\text{OH})_2 + 2\text{e}^- \rightarrow \text{Ba} + 2\text{OH}^-$	-2.99
$\text{Sr}(\text{OH})_2 + 2\text{e}^- \rightarrow \text{Sr} + 2\text{OH}^-$	-2.88
$\text{Y}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Y} + 3\text{OH}^-$	-2.85
$\text{Ho}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Ho} + 3\text{OH}^-$	-2.85
$\text{Er}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Er} + 3\text{OH}^-$	-2.84
$\text{Tm}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Tm} + 3\text{OH}^-$	-2.83
$\text{Lu}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Lu} + 3\text{OH}^-$	-2.83
$\text{Gd}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Gd} + 3\text{OH}^-$	-2.82
$\text{Tb}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Tb} + 3\text{OH}^-$	-2.82
$\text{La}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{La} + 3\text{OH}^-$	-2.80
$\text{Sm}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Sm} + 3\text{OH}^-$	-2.80
$\text{Dy}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Dy} + 3\text{OH}^-$	-2.80
$\text{Pr}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Pr} + 3\text{OH}^-$	-2.79
$\text{Ce}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Ce} + 3\text{OH}^-$	-2.78
$\text{Nd}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Nd} + 3\text{OH}^-$	-2.78
$\text{Pm}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Pm} + 3\text{OH}^-$	-2.76
$\text{Yb}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Yb} + 3\text{OH}^-$	-2.74
$\text{Mg}(\text{OH})_2 + 2\text{e}^- \rightarrow \text{Mg} + 2\text{OH}^-$	-2.687
$\text{Sc}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Sc} + 3\text{OH}^-$	-2.60
$\text{ThO}_2 + 2\text{H}_2\text{O} + 4\text{e}^- \rightarrow \text{Th} + 4\text{OH}^-$	-2.56
$\text{Am}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Am} + 3\text{OH}^-$	-2.53
$\text{Cm}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Cm} + 3\text{OH}^-$	-2.5
$\text{Pu}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Pu} + 3\text{OH}^-$	-2.46
$\text{Al}(\text{OH})_4^- + 3\text{e}^- \rightarrow \text{Al} + 4\text{OH}^-$	-2.310
$\text{Al}(\text{OH})_3(\text{g}) + 3\text{e}^- \rightarrow \text{Al} + 3\text{OH}^-$	-2.300
$\text{Np}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Np} + 3\text{OH}^-$	-2.2
$\text{TiO} + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{Ti} + 2\text{OH}^-$	-2.13
$\text{U}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{U} + 3\text{OH}^-$	-2.10
$\text{H}_2\text{PO}_2^- + \text{e}^- \rightarrow \text{P} + 2\text{OH}^-$	-2.05
$\text{Ti}_2\text{O}_3 + \text{H}_2\text{O} + 2\text{e}^- \rightarrow 2\text{TiO} + 2\text{OH}^-$	-1.95
$\text{B}(\text{OH})_4^- + 3\text{e}^- \rightarrow \text{B} + 4\text{OH}^-$	-1.811
$\text{SiO}_3^{2-} + 3\text{H}_2\text{O} + 4\text{e}^- \rightarrow \text{Si} + 6\text{OH}^-$	-1.7
$\text{HPO}_3^{2-} + 2\text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{H}_2\text{PO}_2^- + 3\text{OH}^-$	-1.57
$\text{Mn}(\text{OH})_2 + 2\text{e}^- \rightarrow \text{Mn} + 2\text{OH}^-$	-1.56
$\text{ZnS} + 2\text{e}^- \rightarrow \text{Zn} + \text{S}^{2-}$	-1.44
$\text{PuO}_2 + 2\text{H}_2\text{O} + \text{e}^- \rightarrow \text{Pu}(\text{OH})_3 + \text{OH}^-$	-1.4
$2\text{TiO}_2 + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{Ti}_2\text{O}_3 + 2\text{OH}^-$	-1.38
$\text{Zn}(\text{CN})_4^{2-} + 2\text{e}^- \rightarrow \text{Zn} + 4\text{CN}^-$	-1.34
$\text{Cr}(\text{OH})_3 + 3\text{e}^- \rightarrow \text{Cr} + 3\text{OH}^-$	-1.33
$\text{Cr}(\text{OH})_4^- + 3\text{e}^- \rightarrow \text{Cr} + 4\text{OH}^-$	-1.33
$\text{Zn}(\text{OH})_4^{2-} + 2\text{e}^- \rightarrow \text{Zn} + 4\text{OH}^-$	-1.285
$\text{CdS} + 2\text{e}^- \rightarrow \text{Cd} + \text{S}^{2-}$	-1.255

Continued

<i>Couple</i>	<i>E° (V)</i>
$\text{Zn}(\text{OH})_2 + 2\text{e}^- \rightarrow \text{Zn} + 2\text{OH}^-$	-1.248
$\text{H}_2\text{GaO}_3 + \text{H}_2\text{O} + 3\text{e}^- \rightarrow \text{Ga} + 4\text{OH}^-$	-1.22
$\text{Te} + 2\text{e}^- \rightarrow \text{Te}^{2-}$	-1.14
$\text{PO}_4^{3-} + 2\text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{HPO}_3^{2-} + 3\text{OH}^-$	-1.12
$\text{WO}_4^{2-} + 4\text{H}_2\text{O} + 6\text{e}^- \rightarrow \text{W} + 8\text{OH}^-$	-1.074
$\text{ZnCO}_3 + 2\text{e}^- \rightarrow \text{Zn} + \text{CO}_3^{2-}$	-1.06
$\text{Zn}(\text{NH}_3)_4^{2+} + 2\text{e}^- \rightarrow \text{Zn} + 4\text{NH}_3$	-1.04
$\text{HGeO}_3 + 2\text{H}_2\text{O} + 4\text{e}^- \rightarrow \text{Ge} + 5\text{OH}^-$	-1.03
$\text{MnO}_2 + 2\text{H}_2\text{O} + 4\text{e}^- \rightarrow \text{Mn} + 4\text{OH}^-$	-0.980
$\text{CNO}^- + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{CN}^- + 2\text{OH}^-$	-0.97
$\text{Cd}(\text{CN})_4^{2-} + 2\text{e}^- \rightarrow \text{Cd} + 4\text{CN}^-$	-0.943
$\text{SO}_4^{2-} + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{SO}_3^{2-} + 2\text{OH}^-$	-0.94
$\text{PbS} + 2\text{e}^- \rightarrow \text{Pb} + \text{S}^{2-}$	-0.923
$\text{MoO}_4^{2-} + 4\text{H}_2\text{O} + 6\text{e}^- \rightarrow \text{Mo} + 8\text{OH}^-$	-0.913
$\text{Sn}(\text{OH})_6^{2-} + 2\text{e}^- \rightarrow \text{HSnO}_2^- + \text{H}_2\text{O} + 3\text{OH}^-$	-0.91
$\text{P} + 3\text{H}_2\text{O} + 3\text{e}^- \rightarrow \text{PH}_3 + 3\text{OH}^-$	-0.89
$2\text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{H}_2 + 2\text{OH}^-$	-0.828
$\text{Cd}(\text{OH})_2 + 2\text{e}^- \rightarrow \text{Cd} + 2\text{OH}^-$	-0.824
$\text{VO} + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{V} + 2\text{OH}^-$	-0.820
$\text{HFeO}_2 + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{Fe} + 3\text{OH}^-$	-0.8
$\text{MoO}_4^{2-} + 2\text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{MoO}_2 + 4\text{OH}^-$	-0.780
$\text{CdCO}_3 + 2\text{e}^- \rightarrow \text{Cd} + \text{CO}_3^{2-}$	-0.734
$\text{Co}(\text{OH})_2 + 2\text{e}^- \rightarrow \text{Co} + 2\text{OH}^-$	-0.733
$\text{Ni}(\text{OH})_2 + 2\text{e}^- \rightarrow \text{Ni} + 2\text{OH}^-$	-0.72
$\text{CrO}_4^{2-} + 4\text{H}_2\text{O} + 3\text{e}^- \rightarrow \text{Cr}(\text{OH})_4^- + 4\text{OH}^-$	-0.72
$\text{Ag}_2\text{S} + 2\text{e}^- \rightarrow \text{Ag} + \text{S}^{2-}$	-0.691
$\text{FeO}_2 + \text{H}_2\text{O} + \text{e}^- \rightarrow \text{HFeO}_2^- + \text{OH}^-$	-0.69
$\text{AsO}_2 + 2\text{H}_2\text{O} + 3\text{e}^- \rightarrow \text{As} + 4\text{OH}^-$	-0.68
$\text{Se} + 2\text{e}^- \rightarrow \text{Se}^{2-}$	-0.67
$\text{AsO}_4^{3-} + 2\text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{AsO}_2^- + 4\text{OH}^-$	-0.67
$\text{Sb}(\text{OH})_4^- + 3\text{e}^- \rightarrow \text{Sb} + 4\text{OH}^-$	-0.639
$\text{Cd}(\text{NH}_3)_4^{2+} + 2\text{e}^- \rightarrow \text{Cd} + 4\text{NH}_3$	-0.622
$\text{ReO}_4^- + 2\text{H}_2\text{O} + 7\text{e}^- \rightarrow \text{Re} + 8\text{OH}^-$	-0.604
$\text{ReO}_4^- + 2\text{H}_2\text{O} + 3\text{e}^- \rightarrow \text{ReO}_2 + 4\text{OH}^-$	-0.594
$2\text{SO}_3^{2-} + 3\text{H}_2\text{O} + 4\text{e}^- \rightarrow \text{S}_2\text{O}_3^{2-} + 6\text{OH}^-$	-0.58
$\text{ReO}_2 + \text{H}_2\text{O} + 4\text{e}^- \rightarrow \text{Re} + 4\text{OH}^-$	-0.564
$\text{Cu}_2\text{S} + 2\text{e}^- \rightarrow 2\text{Cu} + \text{S}^{2-}$	-0.542
$\text{HPbO}_2^- + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{Pb} + 3\text{OH}^-$	-0.502
$\text{Ni}(\text{NH}_3)_6^{2+} + 2\text{e}^- \rightarrow \text{Ni} + 6\text{NH}_3$	-0.476
$\text{Sb}(\text{OH})_6^- + 2\text{e}^- \rightarrow \text{Sb}(\text{OH})_4^- + 2\text{OH}^-$	-0.465
$\text{Bi}_2\text{O}_3 + 3\text{H}_2\text{O} + 6\text{e}^- \rightarrow \text{Bi} + 6\text{OH}^-$	-0.452
$\text{S} + 2\text{e}^- \rightarrow \text{S}^{2-}$	-0.45
$\text{NiCO}_3 + 2\text{e}^- \rightarrow \text{Ni} + \text{CO}_3^{2-}$	-0.45
$\text{Cu}(\text{CN})_2 + \text{e}^- \rightarrow \text{Cu} + 2\text{CN}^-$	-0.44
$\text{TeO}_3^{2-} + 3\text{H}_2\text{O} + 4\text{e}^- \rightarrow \text{Te} + 6\text{OH}^-$	-0.42
$\text{Cu}_2\text{O} + \text{H}_2\text{O} + 2\text{e}^- \rightarrow 2\text{Cu} + 2\text{OH}^-$	-0.365
$\text{SeO}_3^{2-} + 3\text{H}_2\text{O} + 4\text{e}^- \rightarrow \text{Se} + 6\text{OH}^-$	-0.36
$\text{Ti}(\text{OH}) + \text{e}^- \rightarrow \text{Ti} + \text{OH}^-$	-0.343
$\text{O}_2 + \text{e}^- \rightarrow \text{O}_2^-$	-0.33
$\text{Ag}(\text{CN})_2 + \text{e}^- \rightarrow \text{Ag} + 2\text{CN}^-$	-0.31
$\text{Cu}(\text{SCN}) + \text{e}^- \rightarrow \text{Cu} + \text{SCN}^-$	-0.310
$\text{CuO} + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{Cu} + 2\text{OH}^-$	-0.29
$\text{Mn}_2\text{O}_3 + 2\text{H}_2\text{O} + 2\text{e}^- \rightarrow 2\text{Mn}(\text{OH})_2 + 2\text{OH}^-$	-0.25
$2\text{CuO} + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{Cu}_2\text{O} + 2\text{OH}^-$	-0.22
$\text{Cu}(\text{NH}_3)_2^+ + \text{e}^- \rightarrow \text{Cu} + 2\text{NH}_3$	-0.100
$\text{O}_2 + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{HO}_2^- + \text{OH}^-$	-0.0649
$\text{Ti}(\text{OH})_3 + 2\text{e}^- \rightarrow \text{TiOH} + 2\text{OH}^-$	-0.05
$\text{MnO}_2 + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{Mn}(\text{OH})_2 + 2\text{OH}^-$	-0.05
$\text{AgCN} + \text{e}^- \rightarrow \text{Ag} + \text{CN}^-$	-0.017
$\text{NO}_3^- + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{NO}_2^- + 2\text{OH}^-$	0.01
$\text{SeO}_4^{2-} + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{SeO}_3^{2-} + 2\text{OH}^-$	0.03
$\text{Co}(\text{NH}_3)_6^{3+} + \text{e}^- \rightarrow \text{Co}(\text{NH}_3)_6^{2+}$	0.058

Continued

<i>Couple</i>	<i>E</i> <sup>o</sup> (V)
$\text{TeO}_4^{2-} + \text{H}_2\text{O} + \text{e}^- \rightarrow \text{TeO}_3^{2-} + 2\text{OH}^-$	0.07
$\text{HgO}(\text{red}) + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{Hg} + 2\text{OH}^-$	0.0977
$\text{N}_2\text{H}_4 + 2\text{H}_2\text{O} + 2\text{e}^- \rightarrow 2\text{NH}_5(\text{aq}) + 2\text{OH}^-$	0.1
$\text{VO}_4^{3-} + 4\text{H}_2\text{O} + 5\text{e}^- \rightarrow \text{V} + 8\text{OH}^-$	0.120
$\text{Co}(\text{OH})_3 + \text{e}^- \rightarrow \text{Co}(\text{OH})_2 + \text{OH}^-$	0.17
$\text{HO}_2^- + \text{H}_2\text{O} + \text{e}^- \rightarrow \text{OH}^- + 2\text{OH}^-$	0.184
$\text{O}_2^- + \text{H}_2\text{O} + \text{e}^- \rightarrow \text{HO}_2^- + \text{OH}^-$	0.20
$\text{PbO}_2(\beta) + \text{H}_2\text{O} + \text{e}^- \rightarrow \text{HPbO}_2 + \text{OH}^-$	0.208
$\text{PbO}_2(\beta) + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{PbO}(\text{red}) + 2\text{OH}^-$	0.247
$\text{IO}_5^- + 3\text{H}_2\text{O} + 6\text{e}^- \rightarrow \text{I}^- + 6\text{OH}^-$	0.257
$\text{ClO}_3^- + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{ClO}_2^- + 2\text{OH}^-$	0.295
$\text{PuO}_2(\text{OH})_2 + \text{e}^- \rightarrow \text{PuO}_2\text{OH} + \text{OH}^-$	0.3
$\text{PbO}_3^{2-} + 2\text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{HPbO}_2^- + 3\text{OH}^-$	0.330
$\text{Ag}_2\text{O} + \text{H}_2\text{O} + 2\text{e}^- \rightarrow 2\text{Ag} + 2\text{OH}^-$	0.342
$\text{Ag}(\text{NH}_3)_2^+ + \text{e}^- \rightarrow \text{Ag} + \text{NH}_3$	0.373
$\text{ClO}_4^- + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{ClO}_3^- + 2\text{OH}^-$	0.374
$\text{O}_2 + 2\text{H}_2\text{O} + 4\text{e}^- \rightarrow 4\text{OH}^-$	0.401
$\text{NH}_2\text{OH} + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{NH}_3 + 2\text{OH}^-$	0.42
$\text{Ag}(\text{SO}_3)_2^{3-} + \text{e}^- \rightarrow \text{Ag} + 2\text{SO}_3^{2-}$	0.43
$\text{Ag}_2\text{CO}_3 + 2\text{e}^- \rightarrow 2\text{Ag} + \text{CO}_3^{2-}$	0.47
$\text{IO}^- + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{I}^- + 2\text{OH}^-$	0.472
$\text{NiO}_2 + 2\text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{Ni}(\text{OH})_2 + 2\text{OH}^-$	0.490
$\text{FeO}_4^{2-} + 2\text{H}_2\text{O} + 3\text{e}^- \rightarrow \text{FeO}_2^- + 4\text{OH}^-$	0.55
$\text{BrO}_3^- + 3\text{H}_2\text{O} + 6\text{e}^- \rightarrow \text{Br}^- + 6\text{OH}^-$	0.584
$\text{RuO}_4^- + \text{e}^- \rightarrow \text{RuO}_4^{2-}$	0.593
$\text{MnO}_4^{2-} + 2\text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{MnO}_2 + 4\text{OH}^-$	0.62
$2\text{AgO} + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{Ag}_2\text{O} + 2\text{OH}^-$	0.640
$\text{H}_3\text{IO}_6^{2-} + 2\text{e}^- \rightarrow \text{IO}_3^- + 3\text{OH}^-$	0.656
$\text{PbO}_4^{4-} + 3\text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{HPbO}_2^- + 5\text{OH}^-$	0.680
$\text{ClO}_2^- + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{ClO}^- + 2\text{OH}^-$	0.681
$\text{Ag}_2\text{O}_3 + \text{H}_2\text{O} + 2\text{e}^- \rightarrow 2\text{AgO} + 2\text{OH}^-$	0.739
$\text{BrO}^- + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{Br}^- + 2\text{OH}^-$	0.766
$\text{HO}_2^- + \text{H}_2\text{O} + 2\text{e}^- \rightarrow 3\text{OH}^-$	0.867
$\text{ClO}^- + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{Cl}^- + 2\text{OH}^-$	0.890
$\text{ClO}_2 + \text{e}^- \rightarrow \text{ClO}_2^-$	1.041
$\text{O}_3 + \text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{O}_2 + 2\text{OH}^-$	1.246
$\text{OH} + \text{e}^- \rightarrow \text{OH}^-$	1.985

“Appendix 10 Standard Potentials in Aqueous Solutions” is reproduced from Bard AJ, Parsons R, and Jordon R (eds.) (1985) *Standard Potentials in Aqueous Solutions*. New York: Marcel Dekker.

# APPENDIX 11. SOLVENTS FOR ULTRAVIOLET SPECTROPHOTOMETRY

© 1989, CRC Press. Published by Elsevier Ltd. All rights reserved.

<i>Solvent</i>	<i>Cutoff wavelength (nm)</i>	<i>Dielectric constant (20°C)</i>	
Acetic acid	260	6.15	
Acetone	330	20.7	(25°C)
Acetonitrile	190	37.5	
Benzene	280	2.284	
2-Butanol	260	15.8	(25°C)
<i>n</i> -Butyl acetate	254		
Carbon disulphide	380	2.641	
Carbon tetrachloride	265	2.238	
1-Chlorobutane	220	7.39	(25°C)
Chloroform <sup>a</sup>	245	4.806	
Cyclohexane	210	2.023	
1,2-Dichloroethane	226	10.19	(25°C)
1,2-Dimethoxyethane	240		
<i>N,N</i> -Dimethylacetamide	268	59	(83°C)
<i>N,N</i> -Dimethylformamide	270	36.7	
Dimethylsulphoxide	265	4.7	
1,4-Dioxane	215	2.209	(25°C)
Diethyl ether	218	4.335	
Ethanol	210	24.30	(25°C)
2-Ethoxyethanol	210		
Ethyl acetate	255	6.02	(25°C)
Glycerol	207	42.5	(25°C)
<i>n</i> -Hexadecane	200	2.06	(25°C)
<i>n</i> -Hexane	210	1.890	
Methanol	210	32.63	(25°C)
2-Methoxyethanol	210	16.9	
Methyl cyclohexane	210	2.02	(25°C)
Methyl ethyl ketone	330	18.5	
Methyl isobutyl ketone	335		
2-Methyl-1-propanol	230	1	
<i>N</i> -Methyl-2-pyrrolidone	285	32.0	
Pentane	210	1.844	
<i>n</i> -Pentyl acetate	212		
1-Propanol	210	20.1	(25°C)
2-Propanol	210	18.3	(25°C)
Pyridine	330	12.3	(25°C)
Tetrachloroethylene <sup>b</sup>	290		
Tetrahydrofuran	220	7.6	
Toluene	286	2.379	(25°C)
1,1,2-Trichloro-1,2,2-trifluoroethane	231		
2,2,4-Trimethylpentane	215	1.936	(25°C)
<i>o</i> -Xylene	290	2.568	
<i>m</i> -Xylene	290	2.374	
<i>p</i> -Xylene	290	2.270	
Water		78.54	(25°C)

<sup>a</sup> Stabilized with ethanol to avoid phosgene formation.

<sup>b</sup> Stabilized with thymol (isopropyl meta-cresol).

Reprinted from Bruno TJ and Svoronos PDN (1989) *CRC Handbook of Basic Tables for Chemical Analysis*, p. 212. Boca Raton, FL: CRC Press.

# APPENDIX 12. IMPORTANT PEAKS IN THE MASS SPECTRA OF COMMON SOLVENTS

© 1989, CRC Press. Published by Elsevier Ltd. All rights reserved.

“Appendix 12. Important Peaks in Mass Spectra of Common Solvents” is reproduced with permission from Lide DR (editor in chief) (1993) *CRC Handbook of Chemistry and Physics*, 72nd edn., Boca Raton: CRC Press.

The following table gives the most important peaks that appear in the mass spectra of the most common solvents which might occur as an impurity in organic samples. The solvents are classified in ascending order of their  $M^+$  peaks. The highest intensity peaks are indicated with (100%).

Important peaks in the mass spectra of common solvents

<i>Solvents</i>	<i>Formula</i>	<i>M<sup>+</sup></i>	<i>Important peaks (m/z)</i>
Water	H <sub>2</sub> O	18 (100%)	17
Methanol	CH <sub>3</sub> OH	32	31 (100%), 29, 15
Acetonitrile	CH <sub>3</sub> CN	41 (100%)	40, 39, 38, 28, 15
Ethanol	CH <sub>3</sub> CH <sub>2</sub> OH	46	45, 31 (100%), 27, 15
Dimethyl ether	CH <sub>3</sub> OCH <sub>3</sub>	46 (100%)	45, 29, 15
Acetone	CH <sub>3</sub> COCH <sub>3</sub>	58	43 (100%), 42, 39, 27, 15
Acetic acid	CH <sub>3</sub> CO <sub>2</sub> H	60	45, 43, 18, 15
Ethylene glycol	HOCH <sub>2</sub> CH <sub>2</sub> OH	62	43, 33, 31 (100%), 29, 18, 15
Furan	C <sub>4</sub> H <sub>4</sub> O	68 (100%)	42, 39, 38, 37, 29, 18
Tetrahydrofuran	C <sub>4</sub> H <sub>8</sub> O	72	71, 43, 42 (100%), 41, 40, 39, 27, 18, 15
n-Pentane	C <sub>5</sub> H <sub>12</sub>	72	57, 43 (100%), 42, 41, 39, 29, 28, 27, 15
Dimethylformamide (DMF)	HCON(CH <sub>3</sub> ) <sub>2</sub>	73 (100%)	58, 44, 42, 30, 29, 28, 18, 15
Diethylether	(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub> O	74	59, 45, 41, 31 (100%), 29, 27, 15
Methylacetate	CH <sub>3</sub> CO <sub>2</sub> CH <sub>3</sub>	74	59, 43 (100%), 42, 32, 29, 28, 15
Carbon disulphide	CS <sub>2</sub>	76 (100%)	64, 44, 38, 32
Benzene	C <sub>6</sub> H <sub>6</sub>	78 (100%)	77, 52, 51, 50, 39, 28
Pyridine	C <sub>5</sub> H <sub>5</sub> N	79 (100%)	80, 78, 53, 52, 51, 50, 39, 26
Dichloromethane	CH <sub>2</sub> Cl <sub>2</sub>	84	86, 51, 49 (100%), 48, 47, 35, 28
Cyclohexane	C <sub>6</sub> H <sub>12</sub>	84	69, 56, 55, 43, 42, 41, 39, 27
n-Hexane	C <sub>6</sub> H <sub>14</sub>	86	85, 71, 69, 57 (100%), 43, 42, 41, 39, 29, 28, 27
p-Dioxane	C <sub>4</sub> H <sub>8</sub> O <sub>2</sub>	88 (100%)	87, 58, 57, 45, 43, 31, 30, 29, 28
Tetramethylsilane (TMS)	(CH <sub>3</sub> ) <sub>4</sub> Si	88	74, 73, 55, 45, 43, 29
1,2-Dimethoxy ethane	(CH <sub>3</sub> OCH <sub>2</sub> ) <sub>2</sub>	90	60, 58, 45 (100%), 31, 29
Toluene	C <sub>6</sub> H <sub>5</sub> CH <sub>3</sub>	92	91 (100%), 65, 51, 39, 28
Chloroform	CHCl <sub>3</sub>	118	120, 83, 81 (100%), 47, 35, 28
Chloroform-d <sub>1</sub>	CDCl <sub>3</sub>	119	121, 84, 82 (100%), 48, 47, 35, 28
Carbon tetrachloride	CCl <sub>4</sub>	152 (not seen)	121, 119, 117 (100%), 84, 82, 58.5, 47, 35, 28
Tetrachloroethene	CCl <sub>2</sub> = CCl <sub>2</sub>	164 (not seen)	168, 166 (100%), 165, 164, 131, 128, 129, 95, 94, 82, 69, 59, 47, 31, 24

Reprinted from Bruno TJ and Svoronos PDN (1989) *CRC Handbook of Basic Tables for Chemical Analysis*, p. 357. Boca Raton, FL: CRC Press.

## Further Reading

Clerc JT, Pretsch E, and Seibl J (1981) Structural analysis of organic compounds by combined application of spectroscopic methods. In: *Studies in Analytical Chemistry*, vol. I. Amsterdam: Elsevier.

McLafferty FW (1980) *Interpretation of Mass Spectra*. Mill Valley, CA: University Science Books.

Pasto DJ and Johnson CR (1969) *Organic Structure Determination*. Englewood Cliffs, NJ: Prentice-Hall.